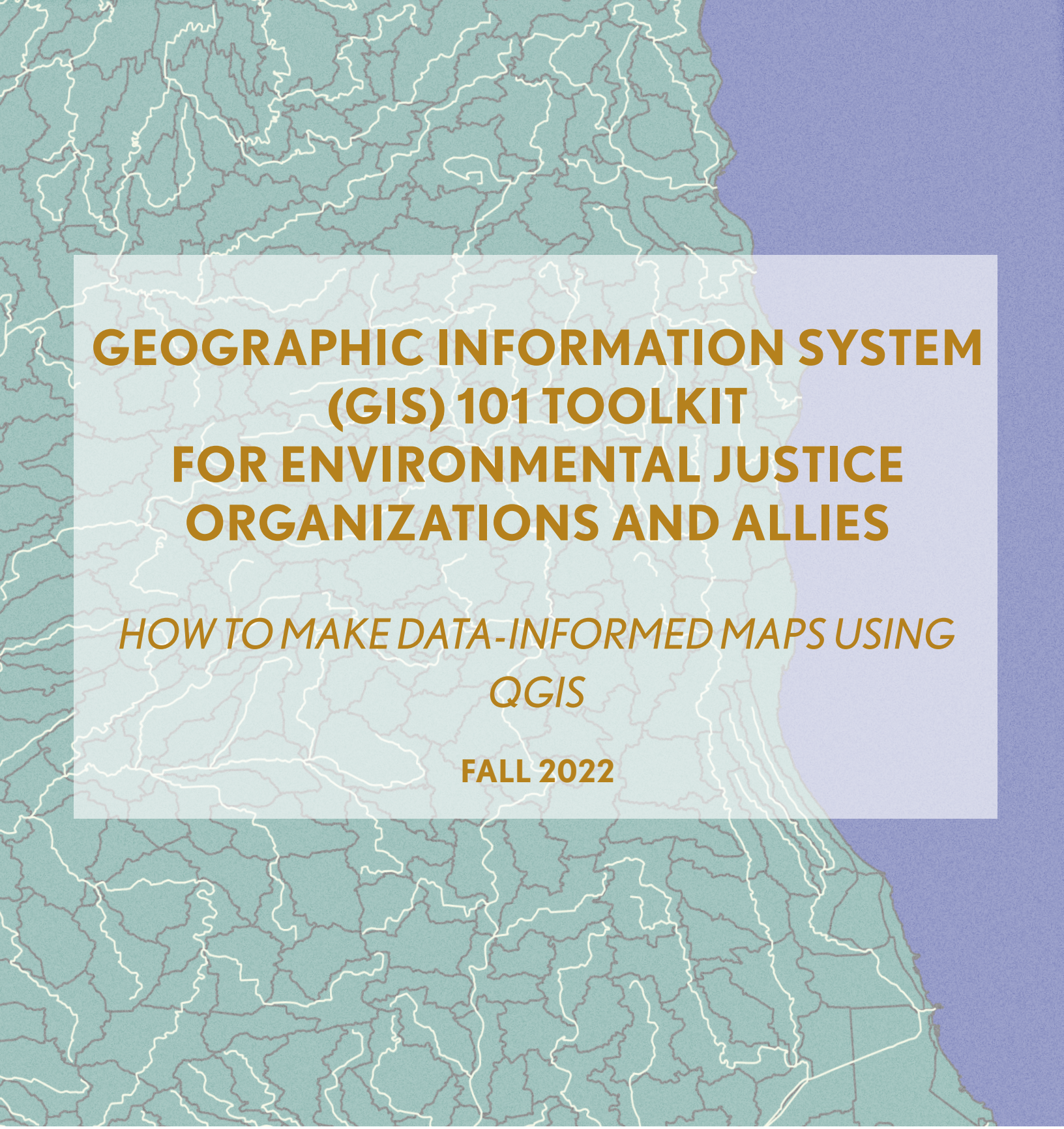


GEOGRAPHIC INFORMATION SYSTEM (GIS) 101 TOOLKIT FOR ENVIRONMENTAL JUSTICE ORGANIZATIONS AND ALLIES

HOW TO MAKE DATA-INFORMED MAPS USING QGIS

FALL 2022

CNT

LVEJO

MILWAUKEE WATER COMMONS

ABOUT THE AUTHORS

Center for Neighborhood Technology (CNT) delivers innovative analysis and solutions that support community-based organizations and local governments to create neighborhoods that are equitable, sustainable, and resilient. Based in Chicago, IL, CNT works on issues of transportation, urban resilience, climate change using urban analytics.

Little Village Environmental Justice Organization (LVEJO) works to organize with our community to accomplish environmental justice in Chicago's Little Village neighborhood and achieve the self-determination of immigrant, low-income, and working-class families. Our vision is to build a sustainable community that promotes the healthy development of youth and families, provides economic justice, and practices participatory democracy and self-determination.

MILWAUKEE WATER COMMONS

Milwaukee Water Commons is a cross-city network that fosters connection, collaboration and broad community leadership on behalf of our common waters. Milwaukee Water Commons promotes stewardship of, equitable access to and shared decision-making for our common waters.

CNT would like to thank the Mott Foundation and Joyce Foundation for their funding support for this project.

INTRODUCTION TO THE TOOLKIT

Executive Summary

Forward

The laws, regulatory proceedings, and incentives of our capitalist society do not promote environmental justice. Industry leaders and political officials actively use quantitative data (numbers) to make the case for industry and pollution instead of the livelihoods of communities and ecosystems. Decisions are often backed by a set of data that feel compelling to the agenda of decision makers. This data is most often quantitative and uses cost-benefit analyses to identify a dollar figure to each cost or benefit. The cultural use of quantitative data as the "be all, end all" makes numbers seem like the absolute truth.

However, the quality of quantitative data is dependent on the methods used to collect the information (raw data) and the methods used to analyze that raw data into something that can be understood by people. For example, the survey methods used to collect data from a small proportion of the population or a small area within a location will not 100% represent the entire geography, even if collected with a more representative sampling method. Some people will always be left out. Further, quantitative data without qualitative data (information conveyed through stories, experiences, art, pictures, etc but not in numbers) provides a distorted picture of reality because qualitative data provides context and nuance that numbers cannot provide on their own.

This toolkit has been created to inform community based organizations how to use spatial data (information tied to locations) in support of their campaigns and technical allies on how to be intentional, collaborative partners for spatial data analysis. Spatial data can be both quantitative and qualitative. By pairing quantitative spatial data with qualitative spatial data, you can identify universality among individual people's experiences.

The authors of this toolkit hope that it can help readers build a foundation with insider tips to critically consider data sources, data collection methods, and data analysis when creating maps. The authors also hope that organizers and advocates feel equipped with an additional tool to critique data from opponents of environmental justice.

Purpose

This GIS (Geographic Information Systems) 101 Toolkit is intended for use by environmental justice organizations or community groups as well as individuals who are in solidarity with these organizations or groups. The toolkit provides an overview of how GIS can be used to supplement organizational campaigns and efforts. It then provides a tutorial of how to create a map that visualizes data from the U.S. Census American Community Survey and U.S. Environmental Protection Agency's EnviroAtlas tool. The skills learned will equip users with the foundation to learn additional map analysis skills and be an informed participant of GIS conversations with future partners and decision-makers.

What is an Environmental Justice Community?

To understand what Environmental Justice (EJ) and an EJ community is we must first understand what environmental racism is. Environmental racism is the disproportionate impact of environmental hazards on Black, Indigenous and people of color. In turn, EJ is the movement's response to environmental racism. Some have used "environmental equity" as a comprehensive and acceptable term; however, that is not environmental justice. Environmental equity is the government's response to the demands of the environmental justice movement. Agencies, such as the U.S. Environmental Protection Agency (U.S. EPA) co-opted the movement by redefining environmental justice as "fair treatment and meaningful involvement," (U.S. EPA) something they consistently fail to accomplish, but also fails to center the goals of the environmental justice movement. EJ seeks to abolish environmental harms in their entirety, not redistribute them (Environmental Justice Network). An EJ community is a disadvantaged and marginalized community where its fundamental rights to political, economic, cultural and environmental self-determination are clearly infringed upon. An EJ community is a community that is disproportionately affected by environmental pollution and other hazards that can lead to adverse health effects, exposure, or environmental degradation.

Intended Audience

Maps can provide powerful visuals to decision-makers and community members about the issues and opportunities a location and its people are experiencing. We hope that environmental justice organizations and community groups feel empowered to create or co-create maps with GIS software. It is helpful for a user to have experience with general navigation on a computer, Excel or another spreadsheet application, and feel comfortable reading data tables and maps. These foundational blocks will be important to navigating this toolkit. If no one on your team has these skill sets, it may help to partner with another organization to collaboratively work through the toolkit.

Use Case

Since this toolkit is intended for environmental justice organizations and community groups, the examples and resources we provide are focused on demographic data and environmental data. Specifically, this tutorial will cover how to work with EPA's EnviroAtlas (Section 5.D) and the American Community Survey (Section 5.E). We also provide additional environmental data sources in Section 5.G.

Additionally, because EJ organizations and community groups campaign around disparate impacts faced by residents, the tutorial map will allow viewers to make preliminary, visual comparisons between areas. These can be further validated with a technical expert who has more experience with spatial statistics. Specifically, the toolkit will focus on data at the census block group level (which are sized to contain about 600 to 3,000 people) within Milwaukee, WI.

Chapter Overview

The toolkit will provide a basic understanding of GIS so that users can visualize data for their intended audience. It will review some data analysis methods and provide links to additional tutorials.

The chapters cover the following topics. To learn what specific GIS skills are taught, refer to page 7 or the beginning of chapters 5 and 6:

Guidelines on resource requirement

Skillsets Needed

In building this toolkit, we assume that the user(s) will have some comfort with data and with computers.

- Basic data literacy skills: GIS is data analysis with the addition of locational information, so it is important to know how to read, analyze, and create data tables. Additionally, the user should have comfortability with reading maps.
- Basic computer skills: Users should feel comfortable using a computer. For example, they should know how to create folders, download programs or files, save projects, navigate folders, navigate the internet, search using a web browser, etc.
- Basic Excel skills: GIS often requires data downloads from multiple different sources or websites, most often this data will be downloaded as a csv (comma separated values) document that can be opened by Excel. In Excel, the user should feel comfortable with creating tables, editing/creating/deleting rows and columns, and using basic formulas.

In addition to all of this, the user(s) should have some expertise in the topic that they are mapping so that they can discern and analyze the GIS results.

If no one within your organization or your community group feels comfortable with the skillsets above, you may want to partner with an academic institution or technical non-profit/organization to conduct GIS analysis. Academic institutions, especially those with degrees or certificates in GIS or data analytics, have classes in which students need to create a map. This may be an opportunity for your organization to partner with a professor, class, or student. Technical non-profits may be interested in working with community-based organizations to expand their portfolio of projects and topics they work in. You can also consider partners within coalitions your organization participates in. With any of these potential partners, we acknowledge that trust must be built, especially when as an environmental justice organization you are fighting for your community's right to live. We suggest having clear rules of engagement so that each individual has a clear understanding of responsibilities and capacities for the project. It may help to have a signed agreement between each organization for clarity.

Hours Needed

This tutorial, if followed step by step, may take 15 - 20 hours. If you are using this tutorial for your own data, additional time may be required.

Hardware/Software Needed

To use this tutorial, we recommend updating your PC or Mac to its most up to date software version.

You will also need to have the following software:

- GIS application (this toolkit provides steps for QGIS, a free software. ArcMap is another commonly used software that requires a paid subscription).
- Spreadsheet application (the most commonly known is Microsoft Excel, which costs money. This toolkit provides steps for Microsoft Excel. If you do not have Excel, it may be worth seeing if your organization can purchase it, perhaps at a discounted rate. Google Sheets could be used as well, though the steps may differ).
- Zipped File Extractor (this usually comes with your computer and is often accessible with a right click on the zipped folder. This toolkit uses 7-ZIP file extractor.)

GIS Skills Learned

If you're interested in learning a specific GIS skill, use the following lists to navigate the toolkit.

Getting Started:

- Setting QGIS up

Data Gathering Skills:

- Download data from TIGER/Line
- Download data from EnviroAtlas
- Use the Text to Columns function in Excel
- Download data from the American Community Survey
- How to Create Survey or Interview Data

Tutorials in QGIS on:

- Open Attribute Table
- Select by Expression
- Select by Location
- Show Selected Features in Attribute Table
- Export Selected Features as a Table (in CSV)
- Export Selected Features as Shapefile
- Join Attributes by Field Value
- Create an alias for attributes
- Projection
- Symbology
- Labels
- Classification of Data using Breakpoints
- Normalization
- Summary Statistics
- Adding Legends

Link to Tutorials on the following QGIS tools:

- Clip
- Intersection
- Buffer
- Merge Vector Layers
- Join Attributes by Location (Spatial Join)

Read an overview of the following GIS analyses:

- Geocode
- Index
- LiDAR Analysis
- Spatial Analysis
- Community Mapping

Little Village Environmental Justice Organization Hilco Implosion 1-year Anniversary Vigil.
Photo by Max Herman

1

DEFINITION OF ENVIRONMENTAL JUSTICE COMMUNITY

To understand what Environmental Justice (EJ) and an EJ community is we must first understand what environmental racism is. Environmental racism is the disproportionate impact of environmental hazards on Black, Indigenous, and people of color. In turn, the environmental justice movement is the response to environmental racism. Some have used "environmental equity" as a comprehensive and acceptable term; however, that is not environmental justice. Environmental equity is the government's response to the demands of the environmental justice movement. Agencies, such as the US Environmental Protection Agency (US EPA) co-opted the movement by redefining environmental justice as "fair treatment and meaningful involvement," (United States Environmental Protection Agency) something they consistently fail to accomplish, but also whose definition fails to address what the environmental justice movement seeks to accomplish. EJ is not about seeking to redistribute environmental harms, but to abolish them in their entirety (Environmental Justice Network).

An EJ community is a disadvantaged and marginalized community where its fundamental rights to political, economic, cultural and environmental self-determination are clearly infringed upon. It is disproportionately affected by environmental pollution and other hazards that can lead to adverse health effects, exposure, or environmental degradation.

To identify EJ communities in maps, several government agencies and organizations have used layering or indices of various data. We will list a few examples of data that has been used below, but want to acknowledge that using data to identify EJ communities without conversations with communities themselves will lead to maps that only partially represent or may misrepresent where these communities are.

Data that is typically used to identify EJ communities often include the following themes:

- **Demographic factors**
  those who identify as a minoritized race/ethnicity and low-income
- **Socio-economic factors**
  age, disability status, housing cost burden, food assistance eligible, education level, etc
- **English language proficiency**
  limited English proficiency or households with those over the age 14 speaking limited or no English
- **Environmental exposure**
  exposure to or close proximity to contaminants or benefits from air, land, or water
- **Access**
  to jobs, education, grocery stores, parks
- **Health**
  hospitalization rates, life expectancy, low birth weights, rates of disease
- **Natural disaster and climate risk exposure**
  drought, flood wildfire

These themes can be narrowed or expanded based on conversations with local EJ communities or the topic of interest. This article shares research on data community groups considered for the Michigan Environmental Justice screen.

Milwaukee Water Commons' We Are Water Event. Photo by Pat Robinson

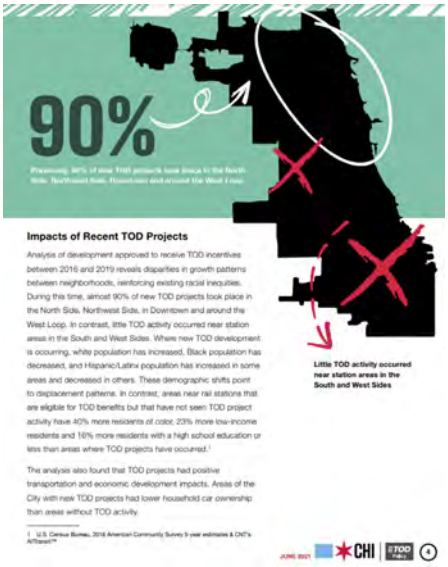
2

WHY SHOULD YOU BE INTERESTED IN USING GIS?

Power of Maps

Maps are a powerful tool when working with communities or decision makers. Maps visualize data presented in tables or written in text, and geographically show where something is or is not happening. Mapping data (either quantitative (numbers) or qualitative (text or pictures)) can show how big or small a problem is, how different areas compare or relate to one another, and can display people's experiences collectively or individually.

For example, in 2021, the City of Chicago published a report on equitable transit oriented development. The beginning of the report uses a map to show where recent transit oriented development occurred to punctuate the case for why policy change was needed. The map is a simple example, using large symbols and numbers to visualize data on different areas of Chicago.



Source: https://www.chicago.gov/content/dam/city/sites/etod/Pdfs/ETOD-Full-Policy-Plan-with-Appendices-6-15-21.pdf

GIS as a Platform for Mapping

GIS (Geographic Information Systems) is a tool typically used in software that allows you to take data from tables and transfer it into a visual map. It is important to note that the accuracy of a map really depends on the quality of the data itself as well as the quality of the methods used to visualize the data within the GIS software.

Maps and their Impact

Environmental justice mapping typically visualizes comparisons between communities to identify disparate burdens or benefits. Though it can also be used to identify assets and strengths within EJ communities themselves.

During the COVID-19 pandemic, maps helped residents and policy makers understand to what intensity different areas were experiencing the pandemic. This informed many of the quarantine policies and decisions for vaccination rollout and outreach. Some organizations paired the COVID-19 pandemic with demographic data (like race and age) to explicitly highlight who was more affected by the pandemic in order to advocate changes in decision-making.

Example

In Chicago, the data were further separated to show how different neighborhoods were experiencing the impacts of COVID-19.

covid19neighborhoods.southsideweekly.com

Example

During the COVID-19 pandemic, GIS was used to visualize COVID-19 infections and deaths in many different ways in the media.

blog.mapbox.com/notable-maps-visualizing-covid-19-and-surrounding-impacts-951724cc4bd8

Example

Little Village Environmental Justice Organization (LVEJO) and Center for Neighborhood Technology worked together to create a report on where water shut-offs, flooding complaints, and COVID-19 infections and deaths occur more frequently in Chicago.

cnt.org/publications/water-health-in-little-village

Maps can reveal cumulative impacts, where people are experiencing multiple stressors at the same time which can impact their quality of life and ability to recover from crises. The LVEJO/CNT map and LVEJO/SETF/SCCBP/NRDC map (see below) show how cumulative burdens are felt most heavily in certain communities. Viewing these maps, one can see how it aligns to residential segregation in Chicago. These maps are particularly powerful for EJ organizations when pressuring policy makers to change, because they visualize how historical and current city planning are related to quality of life impacts. For example, LVEJO has used the NRDC map to support residents' demand to the City of Chicago's Department of Planning and Development for a more thorough and inclusive plan for Little Village's unique industrial corridor that takes into account environmental and socioeconomic factors. The city halted the process to be able to do more analysis based on the recommendations and input. Additionally, LVEJO is using the LVEJO/CNT map to understand the scope of water issues in the Little Village community and inform what policy directions to take. Since publication of the report, LVEJO and CNT have fielded several requests to present and have used the report to support future grant proposals.

LVEJO presents GIS maps from the Water & Health in Little Village report to community members.

Photo by eduardocornejo.com

Example

LVEJO also worked with the Southeast Environmental Task Force, Southeast Side Coalition to Ban Petcoke, and in 2018 to create a map to show the cumulative environmental burdens that occur in neighborhoods of Chicago.

www.nrdc.org/experts/meleah-geertsma/new-map-shows-chicago-needs-environmental-justice-reforms

Example

LVEJO and CNT worked together to create a report on where water shut-offs, flooding complaints, and COVID-19 infections and deaths occur more frequently in Chicago.

cnt.org/publications/water-health-in-little-village

Example

The state of California uses their CalEnviro Map to identify which communities are burdened. They use this map to inform the allocation of some state funds.

oehha.ca.gov/calenviroscreen/report/calenviroscreen-40

Maps can show how past policy decisions impact people's lives today. The New York Times maps illustrating the relationship between redlining and tree canopy show one way that racist policies of the past have a direct effect on people's lives today. These maps can be used to inform policies that correct the harms of the past, by identifying the need for equitable policy solutions and investments. For this reason, some governments, like the State of California, use maps to inform policy decisions and funding allocations.

Example

The New York Times published a piece based on maps that reveal a relationship between redlining and reduced tree investments. This has led to areas with more Black and Brown residents feeling urban heat island effects more intensely than areas where whiter and richer residents live.

https://www.nytimes.com/interactive/2020/08/24/climate/

Example

The state of California uses their CalEnviro Map to identify which communities are burdened. They use this map to inform the allocation of some state funds.

oehha.ca.gov/calenviroscreen/report/calenviroscreen-40

Sometimes maps are used to predict future impacts of an event. The data used in these maps are calculated from mathematical models (equations). This is often used in climate change advocacy to argue why decision-makers need to plan for future changes to ensure community resiliency. For example, some maps use data from models that predict the changes in temperatures or rising sea levels. Models can also be used to predict the health impacts on populations. It is important to note that no map can perfectly reflect reality or predict the future. The validity (accuracy) of a map depends on the research informing the formulas that are used and the quality of the data used in the models. Models cannot fully account for unknown events or factors that happen in the future.

Example

The Great Lakes Integrated Sciences and Assessment group at University of Michigan made maps to predict climate change impacts on the Great Lakes region.

https://glisa.umich.edu/great-lakes-regional-climate-change-maps/

Example

The Georgia Institute of Technology Urban Climate Lab used models (formulas ideally based on validated researchers' work) to show how changing tree canopy and cool roofs affect temperatures.

https://cpn-us-w2.wpmucdn.com/sites.gatech.edu/dist/f/1440/files/2020/09/home-page-graphic.png

Maps do not need to depend only on quantitative (numbers) data. Maps can visualize qualitative (text, images, etc) data like people's stories and other media (photos, audio, etc). This is specifically termed qualitative GIS or counter mapping (discussed in Section 5.I.) By mapping people's experience, we can see how the community views its own assets and can validate or nuance what the numerical data shows.

Example

The UW Milwaukee School of Architecture visualized how the Sherman Park community in Milwaukee, WI sees its neighborhood with videos and audio incorporated into this map.

https://uploads.knightlab.com/storymapjs/68bfe469f701b96f09a8ca90eb690f40/countermapping-sherman-park/index.html

Example

The Wildwoods Foundation, which works on connecting families and educators to Los Angeles' ecologies, is a smaller community-based organization that is using GIS to create a landscape analysis of current parks to identify the current status of parks to then drive infrastructure investments into equitable park creation and maintenance.

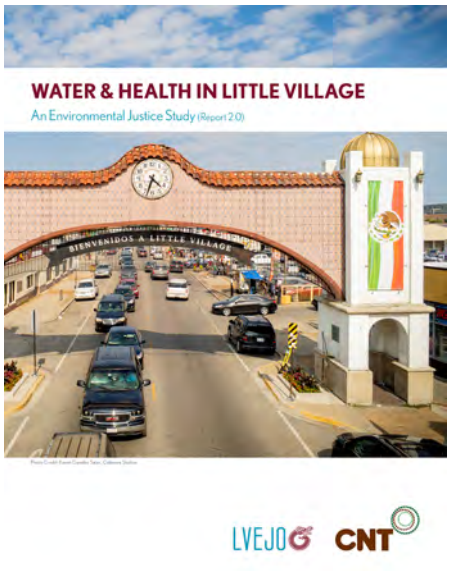
https://www.wildwoodsla.org/los-angeles-cities-connecting-children-to-nature/

3

FRAMEWORK FOR WORKING ON PRODUCT + GIS

Maps that are included in a final product (such as a report, interactive map, poster or a flyer) should be designed to support the product's purpose. The following steps provide a helpful framework to think through what to include in a map and how to develop an impactful product.

To explain each step of the framework, the Water & Health in Little Village report is shared as a case study.



https://cnt.org/sites/default/files/publications/Water-and-Health-in-Little-Village.pdf

A. Clearly identify the issue.

The purpose of this product is to support the organization's work. Your organization may have several issues that you work on. Identify what is a high priority issue or campaign for the organization or something that the organization wants to learn more about to determine future direction.

Case study: When planning for the Water & Health report, LVEJO and CNT first considered the intention for the report. Residents within the Little Village community had expressed drinking water and polluted flooding concerns to LVEJO water team staff. Additionally, the COVID-19 pandemic had just begun and news reports identified that residents of Little Village were facing high rates of infection and death in comparison to other Chicago neighborhoods. With these topics top of mind, LVEJO staff identified that the work product should focus on water and health. (Page 6.)

B. Understand WHY the issue is important.

How does it impact people positively or negatively? Include information from previous organizational research and community comments, new research, and people's experiences.

Case study: LVEJO staff had been informed by their base about water shut-offs and flooding in the past. News reports in Chicago early in the pandemic repeatedly highlighted the devastating impacts of COVID-19 on the Little Village community. At the same time, the Center for Disease Control had published guidance on the importance of handwashing to mitigate spreading of the virus and many cities, including Chicago, announced water shut-off moratoriums. With all of this information, LVEJO and CNT conducted research of journal and newspaper articles to collect information about COVID-19, drinking water, and flooding to understand possible connections. (Page 11-12 reports some of this research.)

C. Identify what question the map will answer.

It is important to identify what questions the maps developed will try to answer. Maps are specifically used to answer questions about a specific location (at one specific point in time or over a period of time) or a comparison of locations (at one specific point in time or over a period of time). Consider the issues identified in step A and the research completed in step B. What are some maps that could help link the WHY to the issues? For an environmental justice analysis, a comparison between neighborhoods is helpful to show disparate burdens.

Some examples of questions include:

Issue	Potential Question
Tree Planting	How do the amount of trees in this neighborhood compare to another neighborhood?
Public Funding Expenditure	Which service locations in the city are frequently used by people in this neighborhood and which are not?
Pollinator Gardens	Where do native pollinators have food sources in this region and how has this changed over the past 50 years?

Case study: Since the issues of interest were drinking water, flooding, and COVID-19 experiences, we focused our question on these issues. Our research found a significant connection between drinking water access and COVID-19. Flooding showed up as a potential issue for COVID-19, when it reduced possibilities of social distancing. Further, LVEJO staff wanted to bring attention to the burden that their residents were feeling, they wanted to understand where else in the city were people experiencing these issues. The questions finally decided upon were: Compared to Little Village, which neighborhoods in Chicago are facing the least burden of these issues? How different are the burden levels between these neighborhoods? (Pages 7-10.)

There might not be data to answer the question you developed. Oftentimes available data does not go into the details that you need. Note this and create a list of data asks you want to advocate for from decision-makers. Consider the data that is available, could it answer a related question that still supports your work? Data often has limitations and being cognizant of them is important.

D. Explore what issue-related geographic data (data related to specific locations) is available related to the issue.

There are federal, state, county, and municipal sources of data. Chapter 5 will provide some initial ideas. You may also be interested in using qualitative data, not just quantitative data. Sometimes the data needed to directly answer your question is not available, so the questions from step C may need to be modified. Think about what data is available and if a similar question or a simpler question can be answered visually with a map. This could be the basis to support future research or data asks.

Case study: CNT was able to request data on drinking water shut-offs and flooding data from various levels of government (Details on how to request data with a FOIA request in 5.H). The Chicago Department of Public Health publicly released and weekly updated COVID-19 infection and death data by zip code. CNT layered the zip code map with the neighborhood boundaries map to identify which zip codes were included in a neighborhood.

E. Map the issue-related data and identify locations with disproportionate burden.

When the issue-related data from step D is mapped, what becomes visible? Where is the maximum occurring? Where is the minimum occurring? How does the quantitative data compare to what has been shared through qualitative data or resident comments? As mentioned in step C, the question for the analysis may look to compare how the data shows up differently between neighborhoods. Looking at the map, is there a concentration of minimum values somewhere and a concentration of maximum values elsewhere? Or does it look pretty random? Using the WHY (information from step B), analyze how the differences in results may affect the people living in those locations. This will help lay out why the data matters and explain your campaign or work.

As you analyze, consider how these results resonate with what you know based on your or your staff's experiences, are the data missing something? See if you can find additional data to fill in gaps or note how the data might be inaccurate (due to what or how something was collected or analyzed).

> ***Case study:*** On pages 7-10 of the report, maps show the number of COVID-19 infections and deaths, vaccination rates, water shut-offs, and flooding complaints in the city. Additionally, there is a table that highlights the values in Little Village and a neighborhood with the least burden (the Near North Side), which were highlighted to compare the clear differences in experiences. The text around each of these maps explains the impact of these data findings on the Little Village community.

F. Find and map demographic data.

Find out WHO is experiencing disproportionate burden and how does the population compare to those who are not feeling burden. For an environmental justice analysis, this can reveal if the burden of the issues are being felt by certain populations more than others. When you map the demographic data, consider how these results resonate with what you know based on your experiences, are the data missing something? See if you can find additional data (both quantitative and qualitative) to fill in gaps or note how the data might be inaccurate due to how data was collected.

> ***Case study:*** On pages 4 & 5 of the report, there are tables showing the demographics between Little Village and the Near North Side. The report focused on these two areas since LVEJO serves Little Village residents and the analysis of the issue-related data highlighted that the Near North Side experienced the least burden among drinking-water, flooding, and COVID-19 issues. The tables in the report reveal that the majority of Little Village residents identify as Latinx, whereas the majority of Near North Side residents identify as white, non-Hispanic. Additionally, the median household income of Little Village is three times lower than that of the Near North Side.

LVEJO highlights policy improvements to consider that the GIS maps helped support.

Photo by eduardocornejo.com

G. Analyze the map findings and relate it to resident stories or history.

Now, lay out for the readers what the maps are showing and what answer they provide to the question from step C. Some of the analysis was completed in step E and F. This is now where steps E, F, and B can be tied together to make a case for your policy/campaign. Some additional questions to consider when analyzing are included below. Also, include community stories in this analysis to pair quantitative data with qualitative data.

- What are the maps saying?
- How do the map findings resonate or counter what the community or decision-makers say?
- How do the map findings fit in with the history of the place?
- How do the map findings fit in with what residents share about the challenges in their lives and in their community?

Case study: The report as a whole (pages 3 - 22) pulls all of the findings together, connecting the maps to one another, so they are not stand alone pieces. The report also grounds the map findings in resident voices through results from a survey on challenges related to drinking water, flooding, pollution, and work conditions (pages 13 - 20). This is how the story is created so that readers can make sense of how the data supports the policy ideas for the final section.

H. Share your policy or action ideas.

The purpose of using GIS is not to just create beautiful maps, it should help the organization make its case for action to the viewer/reader. When you share the policy or action ideas, connect them to the WHY (step B), map findings (step E & F), and resident stories (step G) to the actions and policies that you are advocating for.

Case study: Pages 21 - 22 layout various policies LVEJO is interested in pursuing when it comes to water issues like shut-offs and flooding. Additionally, it explains why these policies would benefit the community and relate back to the issues talked about in the maps.

4

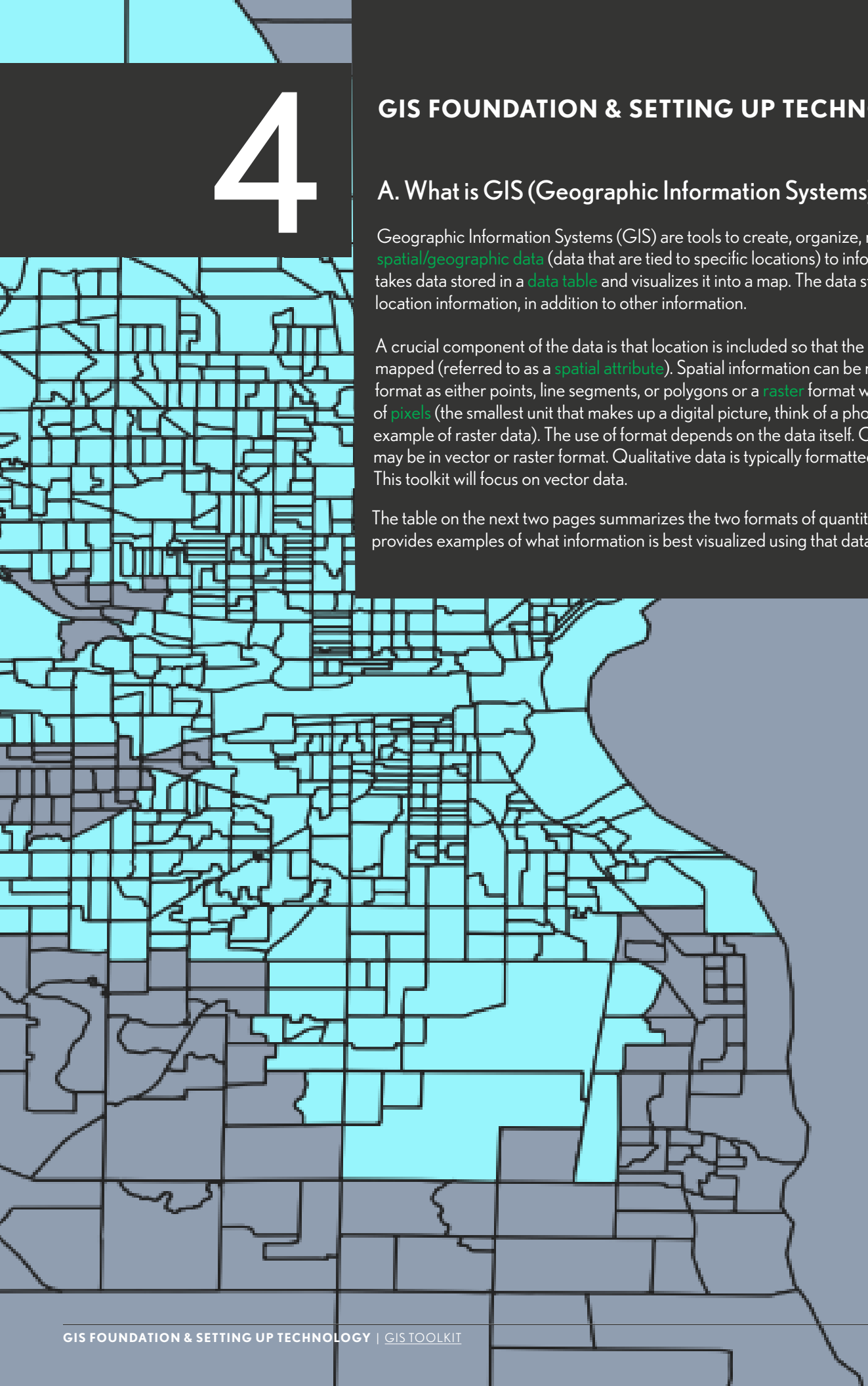
GIS FOUNDATION & SETTING UP TECHNOLOGY

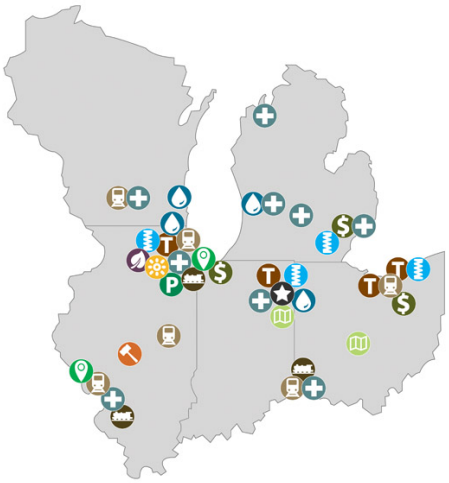
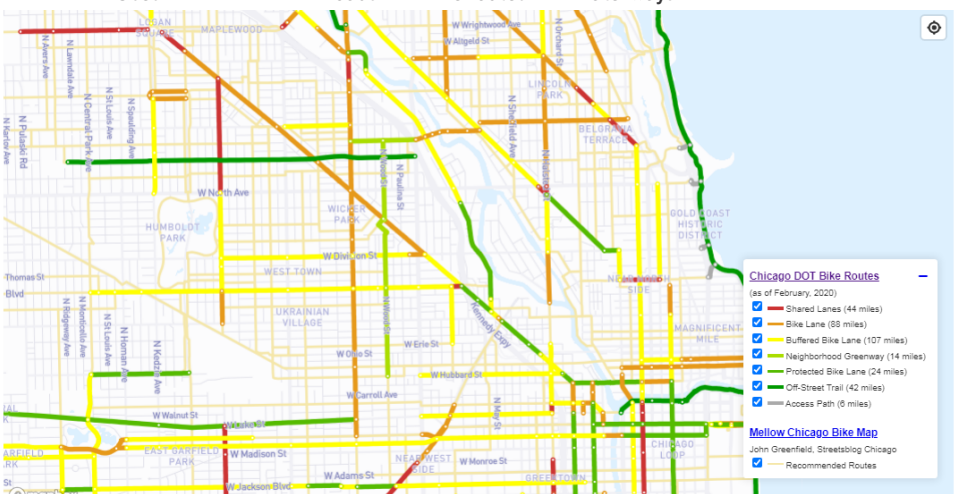
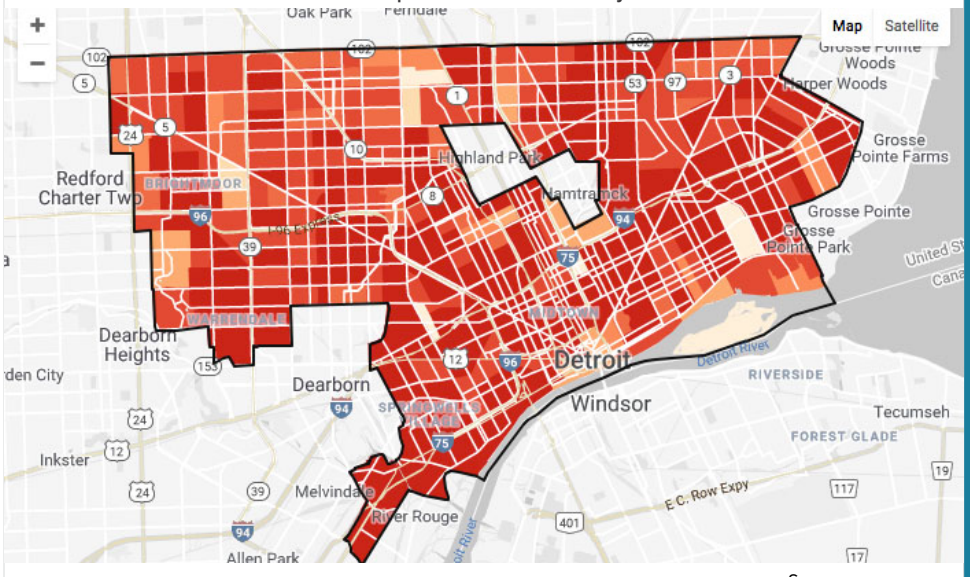
A. What is GIS (Geographic Information Systems)

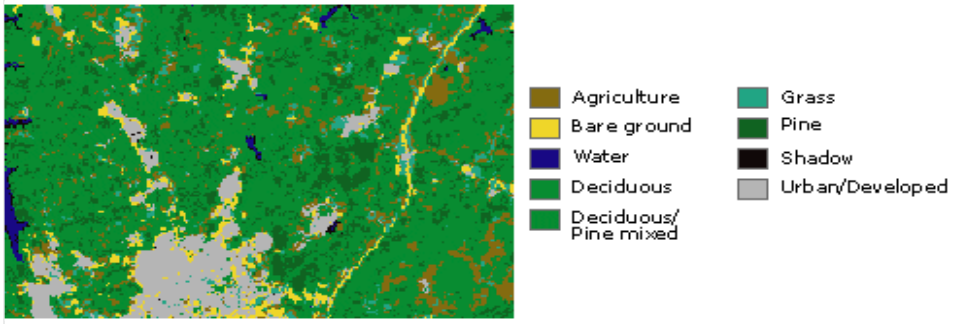
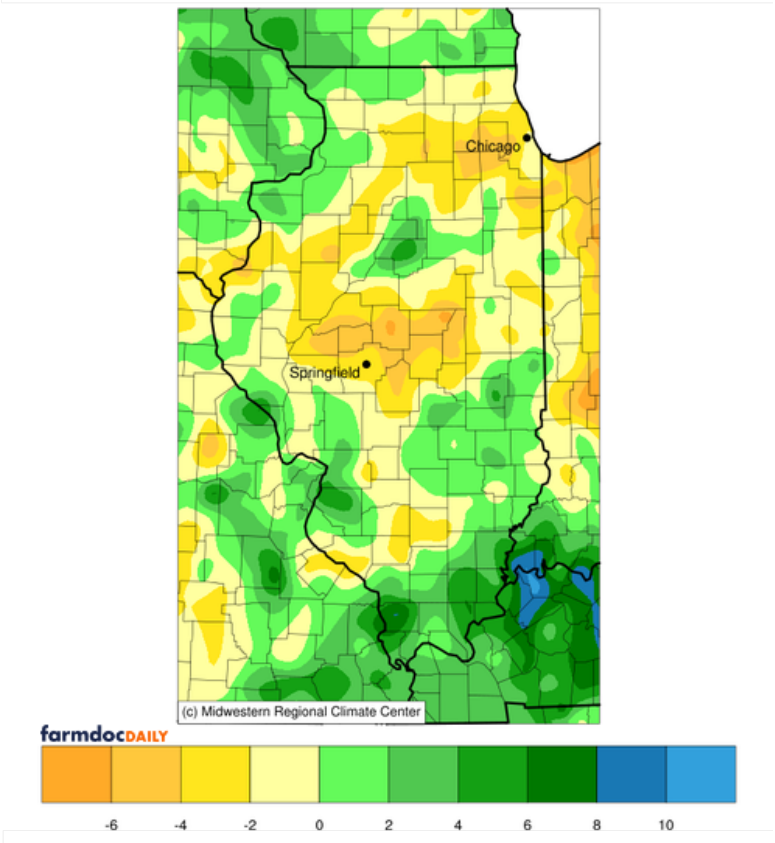
Geographic Information Systems (GIS) are tools to create, organize, map, and analyze spatial/geographic data (data that are tied to specific locations) to inform an audience. It takes data stored in a data table and visualizes it into a map. The data stored will include location information, in addition to other information.

A crucial component of the data is that location is included so that the data can be mapped (referred to as a spatial attribute). Spatial information can be mapped in vector format as either points, line segments, or polygons or a raster format which is made of pixels (the smallest unit that makes up a digital picture, think of a photograph as an example of raster data). The use of format depends on the data itself. Quantitative data may be in vector or raster format. Qualitative data is typically formatted as vector data. This toolkit will focus on vector data.

The table on the next two pages summarizes the two formats of quantitative GIS data and provides examples of what information is best visualized using that data type.



Format	Type	Example
Vector often saved as a Shapefile within GIS software	**Point** Specific longitude/ latitude (x,y)	Uses: Addresses • Points of interest  source: cnt.org
	Line segments	Uses: Roads • Bike routes • Waterways  Source: apps.cnt.org
	Polygons Represents an area	Uses: Census Block Group • State • Country • Parks  Source: uoa.cnt.org

Format	Type	Example
Raster	Discrete specific type	Uses: Land Use - what government designates the land to be used for, like parkland, residential, industrial, or commercial/store uses  source: pro.arcgis.com/en/pro-app/latest/help/data/imagery/introduction-to-raster-data.htm
	Continuous gradual change	Uses: Surface Temperature • Air pollution • Precipitation  Source: farmdocdaily.illinois.edu/2020/10/2020-growing-season-weather-climate-summary.html

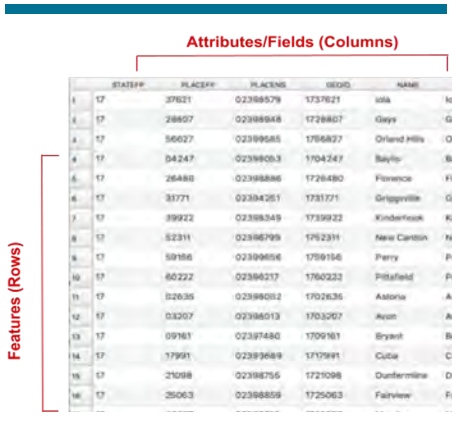
B. When to use GIS

GIS is an appropriate analytical tool if the research question you ask is about the relationship between one or more different attributes (non-spatial information) of a location either at a specific time or over time; or as a comparison between locations. It can use either quantitative or qualitative data, as long as the data relate to a specific location. The following are a few examples:

- Where is the lowest density of tree canopy?
- Do people in certain areas experience more asthma than others?
- What is the relationship between locations with access to healthy food and communities with higher proportions of diabetes?
- How has the number of industrial facilities changed over time in an area?
- Which areas are more vulnerable to floods?
- Which routes do people feel safe to bike on?
- Where do people experience happiness or frustration in the community?

C. Common GIS terms

Term	Definition
Attributes (also called fields)	What topic the information is referring to (like population size), often organized as **columns**
Features	What location the information is referring to (like a specific county), often organized as **rows**
Data Table	Includes the quantitative (numbers) and qualitative (words or pictures) data of the features. Columns of the table refer to the attributes. Rows refer to features.
Shapefile	A format in which GIS data is saved. It maps the data. Can typically be used across GIS software. Once loaded on to a GIS software, we call it a map layer.
Projection	Since maps are flat (two dimensional/2D) but the globe is a three dimensional (3D) object, there are many ways that the globe can be flattened (using mathematical calculations done automatically in GIS software). The mathematical calculations chosen and the resulting map is called a projection. (The next subsection will go into some detail).



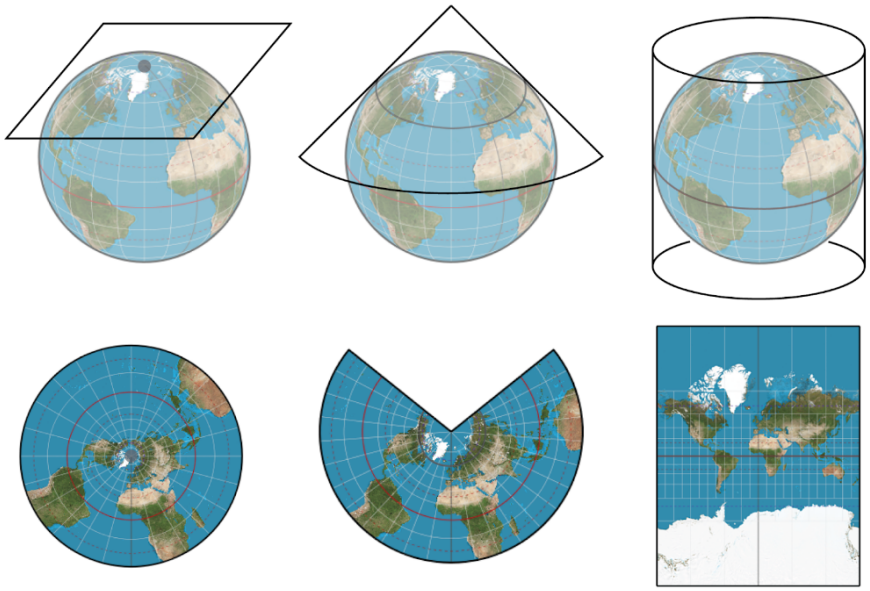
An example datatable

D. Projections

Note: Understanding this section will require some prior knowledge with geographic concepts like longitude and latitude. This video can refresh your memory, note the video is sideways.

Since the Earth is in three dimensions (3D), but maps are two dimensional (2D), GIS uses projections to make a map. Projections are how the coordinate system (longitude and latitude (the x,y location) measured in degrees) of a place on a globe are translated into a specific place on a flat map (measured in a more typical unit for distance like feet or meters). This is done based on the mathematical calculations using a datum (a reference point geographers have identified to do the mathematical calculations). Depending on the datum used and the mathematical calculations used, some projections preserve the surface area of locations (like the relative areas of continents), others preserve the distance between locations. You will need to use projections properly to make sure that the maps look as readers would expect them to, undistorted, and so that calculations and analyses are accurate for the area you are mapping.

Here is a visual example of 3 different projections to map the globe. Notice how the continents look different based on the type of flattening. In the first and second we don't even see Antarctica (white landmass on bottom of the third), but the third shows it. Also notice how the size of Africa changes between first/second and the third. Greenland (the white land mass at the top of each globe) increases in area as you look at each of the different projections.



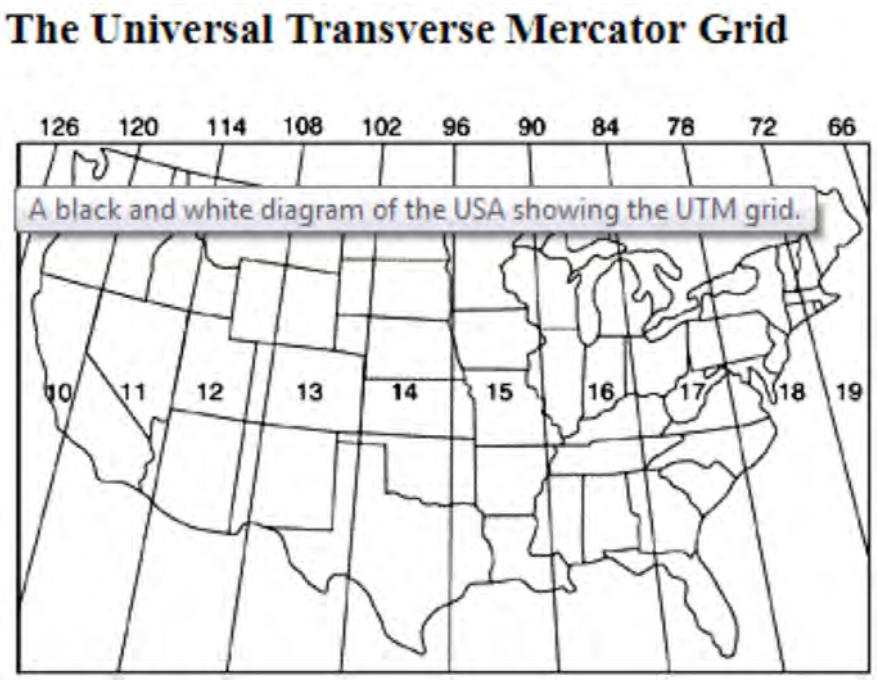
Source: Battersby, S. (2017). Map Projections. The Geographic Information Science & Technology Body of Knowledge (2nd Quarter 2017 Edition), John P. Wilson (ed.). DOI: 10.22224/gistbok/2017.2.7

This example shows how different projections can distort how the map looks. Therefore, it's important to choose a projection that works best for the geographic area you are mapping. To show a more accurate presentation, choose a projection that centers your intended geography area because the geographic area at the edges of the map is often more exaggerated than what is at the center of the map. Geographic locations can have multiple different projections that work well for that specific area.

Most times, the data that is downloaded for a project will be projected already and those who created the data, if they're credible sources, will choose the projection that makes the most sense for the location the data covers. You can use this to inform what projection to use. However, that does not mean that all credible data for a specific geographic location will use the same projection. Therefore, when you are preparing your data for GIS analysis, you'll have to make sure to "reproject" data so that all of your data is in the same projection to accurately do analyses (Steps on how to in Section 5.F).

Some common projection types for use within the United States are:

- State plane provides a projection for each state (and often different parts of a state depending on the size) by considering how vertical or horizontal that state shape is.
- UTM (Universal Transverse Mercator) provides projections based on North to South slices of the globe. Here is a picture of how it looks for the US.



Source: U.S. Geological Survey. (2022). What does the term UTM mean? Is UTM better or more accurate than latitude/longitude? | U.S. Geological Survey. usgs.gov www.usgs.gov/faqs/what-does-term-utm-mean-utm-better-or-more-accurate-latitudelongitude

To make things even more complicated there is something called a geographic coordinate system, also known as "unprojected projection." This is when the GIS software maps the longitude and latitude (x,y) coordinates as degrees. An example of when "unprojected" projection will occur is when a GIS user works with community members to identify specific locations within their community. The GIS user will likely get street intersections, which can be converted to longitude and latitude. To transfer these longitudes and latitudes to GIS software, it will be uploaded in an "unprojected" projection via geocoding (see section 6.F for an example of how to geocode). To analyze this community data with other mapped layers, the GIS user will then "reproject" the community data to match the other data layers' projection.

A helpful resource for further exploration is gistbok.ucgis.org/bok-topics/map-projections, a webpage on projections by the University Consortium for Geographic Information Science.

E. Data Types

Within a datatable, there are different types of data. In the example above, most obviously you can see that some of the columns include text and others include numbers. However, the software reads data differently and stores it with different amounts of memory. If you are working with a lot of data (like thousands of rows), memory will be more of an issue, if not you won't have to worry about it as much. There are several data types and any word or number can be changed into other data types using the software. Knowing what datatype your interested attributes are is important in order to join data together (see section 5.D on how to join data), using various tools in the GIS software, or performing calculations (mentioned briefly in Chapter 6). The following table explains a few key data types. As you can see, the differences can be in how the data is stored, not necessarily the data itself. In other words, the same number (ex:25) may be in two columns but it is stored with more or less memory depending on the data type.

Data Types	
As it's called within the software	Use
Short	typically numeric data, no decimals allowed, lowest amount of memory used
Long	typically numeric data, no decimals allowed, medium amount of memory used
Float recommend using when creating your own datasets in GIS	typically numeric data, decimals allowed, medium amount of memory used
Double recommend using when creating your own datasets in GIS	typically numeric data, decimals allowed, highest amount of memory used - can include most precision (fewest rounding issues)
Text	typically words, numbers can be used. some symbols are okay (like / or _ or -), cannot do mathematical calculations on data when formatted like this, and cannot use graduated symbology (more details in Chapter 6). used often for qualitative data or data that you want to categorize into groups.
Dates	dates when typed as mm/dd/yyyy. useful in certain functions of GIS analysis
BLOBs	when data takes up too much memory to read, such as images, multimedia, you need another application for GIS software to read this

In section 5.D we'll delve more into how to determine what data types you are working with and how to adjust them.

F. GIS Software

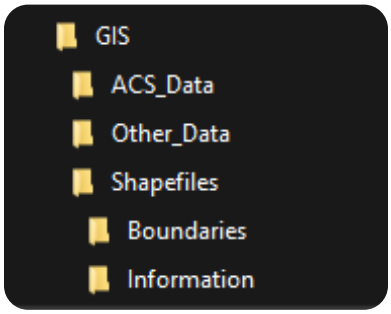
There are several types of GIS software. The simple version of GIS software plots data (puts points on a map) like points on Google Maps. More sophisticated GIS software also incorporates analysis of the data plotted. (Take a look at Chapter 2 for examples of map analysis).

Professionals typically use ESRI ArcGIS software or QGIS. ESRI ArcGIS is the most well known and commonly used software and requires a paid license. QGIS is available for free and is constantly under development so may require updates. ESRI ArcGIS is a more advanced software than QGIS, but QGIS can perform many of the same functions as ESRI ArcGIS. GIS software can provide advanced analysis, especially when you know coding languages like Python, R, and SQL (Structure Query Language) to more directly communicate to the computer. This toolkit will use QGIS as the GIS Software.

Note: Youtube is a great resource for tutorials on conducting different types of GIS analysis for whichever software you have access to.

G. Getting started on GIS

Before you start downloading data it is helpful to create an organized folder on your computer for wherever you want to save your data. Avoid spaces in folder names because that may trigger technical issues with the GIS software. Here is an example of how you might organize your folders. We will be referring to this folder structure throughout the toolkit.



1. Download QGIS from https://www.qgis.org/en/site/. Click "Download now"

2. Choose "Download for..." according to what brand your computer is. We suggest downloading the long term release since it'll be more stable. Downloading QGIS will take a couple of minutes.

3. Once you have downloaded the app, you will need to install it. After installing it, open the app. If your computer tells you it cannot open the app, try to open it by right clicking on the app and then clicking open (rather than by double clicking).

If you run into any other issues in the process of downloading, installing, opening the app, some solutions you may try are:

- Update your computer to the most recent software.
- Clear up storage space on your computer.
- Restart or shut down your computer.
- If your computer admin is handled by an IT department, refer to them.
- Follow this resource about QGIS Installers.

In this toolkit, if we do not cover specific details in QGIS functions, we recommend that you take a look at the user guide. For this toolkit we are using QGIS 3.16 with this as the user guide: https://docs.qgis.org/3.16/pdf/en/QGIS-3.16-DesktopUserGuide-en.pdf.

A newer version of QGIS may be out when you begin this exercise. You will be able to follow along with the toolkit using a different version, but there may be some differences. If you are using a different version, google "QGIS [version number] documentation" for a tailored user guide to your version of QGIS.

Now that you've opened up QGIS, it's important to remember to regularly save your project so as not to lose your work. Depending on the amount of data you are working with, any GIS software can freeze up. So, save early and save often!

5

DATA GATHERING

This chapter reviews how to find and prepare data that relates to the question that the organization is hoping to answer using maps. It will include some conceptual information to understand as well as tutorials on downloading and preparing data for analysis.

Before you can search for data, it is first important to determine your question (section 4.B) and the geography that you will focus on (section 5.A). This will help you search for the data you need to analyze. Additionally, you will learn how to find data that provides the boundaries for your chosen geography. Afterwards, you will learn about data dictionaries, an organizational tool to record what data you use in the map. Then, you will find data that is relevant to your question and prepare the data for analysis.

The specific QGIS tools you'll learn in this chapter include:

- Open Attribute Table
- Select by Expression
- Select by Location
- Show Selected Features in Attribute Table
- Export Selected Features as a Table (in CSV)
- Export Selected Features as Shapefile
- Join Attributes by Field Value
- Create an alias for attributes
- Projection

Additionally, you'll learn to:

- Download data from TIGER/Line
- Download data from EnviroAtlas
- Use the Text to Columns function in Excel
- Download data from the American Community Survey
- How to Create Survey or Interview Data

For the example in this toolkit, the research question we hope to answer is: **Where in Milwaukee would people benefit from additional tree plantings or green infrastructure?**

A. Determine your geography

When using GIS, the data needs to include a location. Data associated with similar geographic sizes are often categorized as a spatial unit, which can vary in size. A spatial unit (the term for the type/size of the location the data is attached to in GIS) could be as large as an ocean, region, or country; as small as a house, block, or section of a stream; or any size in between. Data about the same topic (public health, environment, demographics, etc) can be available at different spatial units (census block group, county, zip code, 1 acre regions, river segment, road etc). When you download a datatable or data set, the spatial unit for each feature is typically the same, so it would be all countries not a mix of countries and states, or all houses not a mix of houses and blocks. However, maps with many data layers can have multiple spatial units (which can become difficult to interpret if there are too many spatial units).

The spatial unit of each data set has an implication for the analysis. Some important things to consider:

*When the spatial unit of a data changes, the visualization changes.

For example, the research question is "Where are more students preparing to graduate in year x?". I have access to data on high school population sizes in a community.

Data Set A	
School	Number of Students
School 1	400
School 2	452

Data Set B		
School	Grade	Number of Students
School 1	9th	130
School 1	10th	100
School 1	11th	95
School 1	12th	75
School 2	9th	105
School 2	10th	125
School 2	11th	160
School 2	12th	62

Data set A shows the population per the whole school. Data set B shows the population per grade per school. Since I am interested in graduation, I want to know how many seniors are at each school. If I only had access to Data set A, since it provides only population for the whole school, one assumption I could make is to divide each school's population by 4 to get an estimate of the senior class (see the math in the new column). Based on Data set A and my calculations, I would assume School 2 has a larger senior class.

Data Set A		Researcher Calculations
School	Number of Students	Estimate of Seniors
School 1	400	100
School 2	452	113

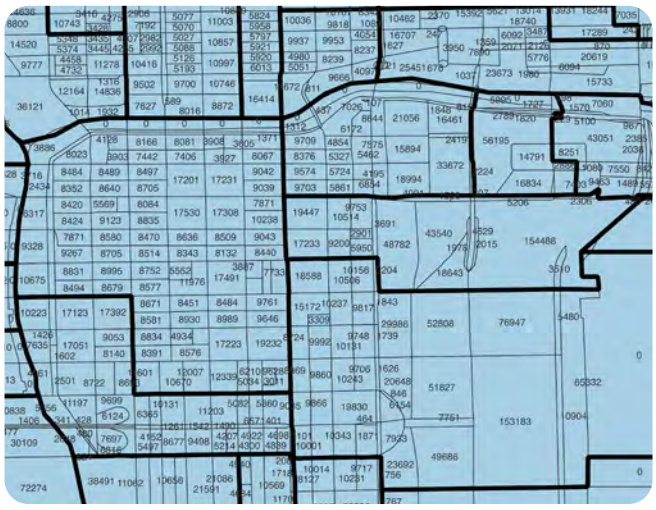
However, with Data set B, I can see the exact number of seniors at each school and I see that School 1 actually has the larger senior class. This is a simplified version of Modifiable Areal Unit Problem (MAUP). The idea of MAUP is that when the spatial unit of a data set changes, the visualization of the map can change depending on how you can analyze the data. A real-life example of MAUP is gerrymandering. By redrawing the lines, the voting district represents a different political opinion. Something similar can happen when you have data layers whose spatial units do not match each other within a map or to your question.

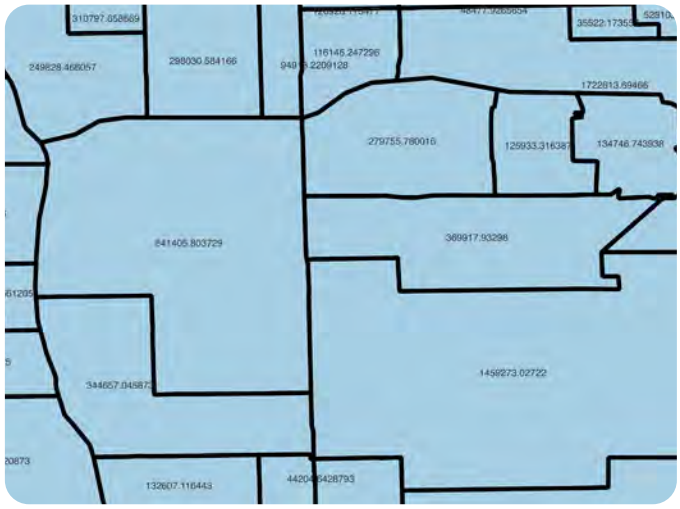
For this reason, try to find data with spatial units that match your question and match each other. If your question is about neighborhoods, find data at the neighborhood level. If your question is about states, find data at the state level. This may not be possible, and you can change the spatial unit, but it's important to be careful about what methods are used to maintain high levels of accuracy.

***Data can be more easily aggregated (added up together) to have an accurate representation, then disaggregated (split up). Therefore, having data with smaller spatial units can be more flexible to work with, than data sets with larger spatial units.**

For example: I know the number of residents in each house, but I want to know the total number of people who live on the block. I can just add the number of residents in each house on the block. However, if I only know the total count of people on the block, I don't know how to accurately find out the number of people in house A vs house B. I could find an average, but it would not be as accurate. GIS researchers use various methodologies to determine more accurate counts. Take a look at chapter 6 for additional details on aggregation and disaggregation.

Visual example: The images below show land area for blocks and block groups. From the blocks, we can easily determine the total area of the block groups by summing them, but we cannot determine the exact area of specific blocks with information on block groups alone.





For maps with environmental justice topics, it is helpful to compare between geographies since a core defining characteristic of an EJ community is one that faces disproportionately more environmental burdens relative to others. It is important to define what the size of an EJ community is for the particular question you are asking. Is it neighborhood to neighborhood, city to city, or something else? Having a clear sense of the comparison you are trying to make will clarify your geography.

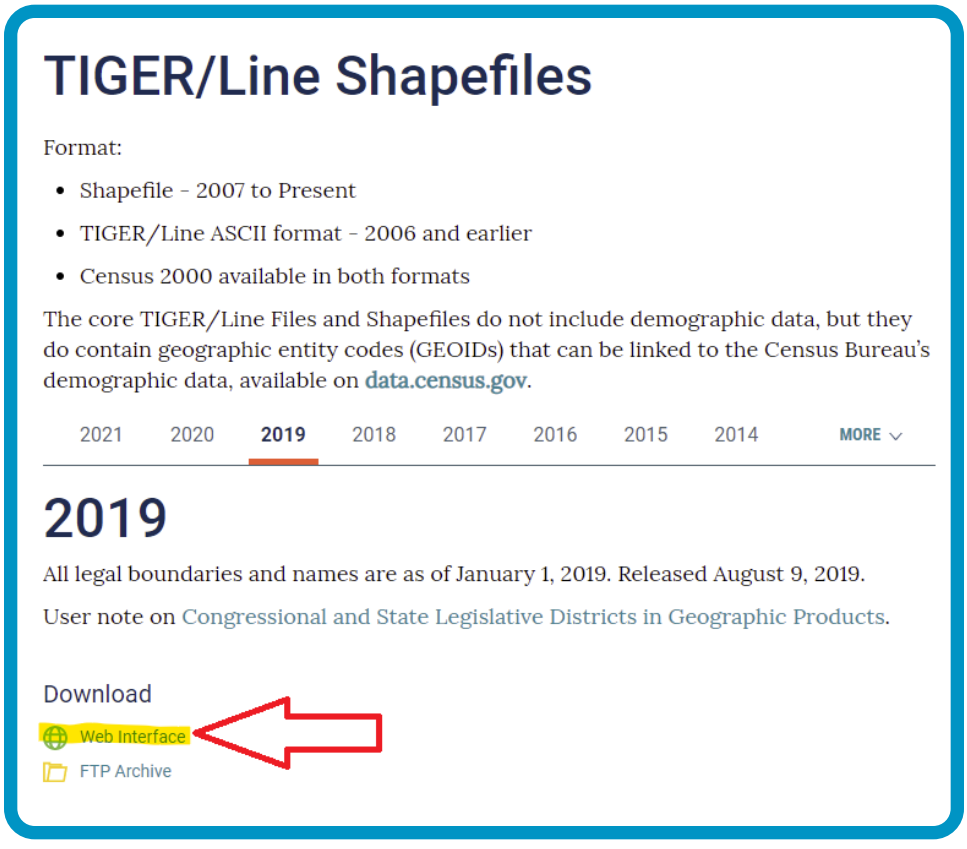
B. Working with TIGER/Line Shapefiles

Later in section 5.E, we will review the American Community Survey (ACS) and where to download that data. The data you download from the ACS will need to be joined (connected) to a shapefile (a spatial data format) that the GIS software reads to map the data visually. To prepare for all of this, you will want to have a shapefile that will display the geography you are interested in. We will download this shapefile from TIGER/Line Shapefiles.

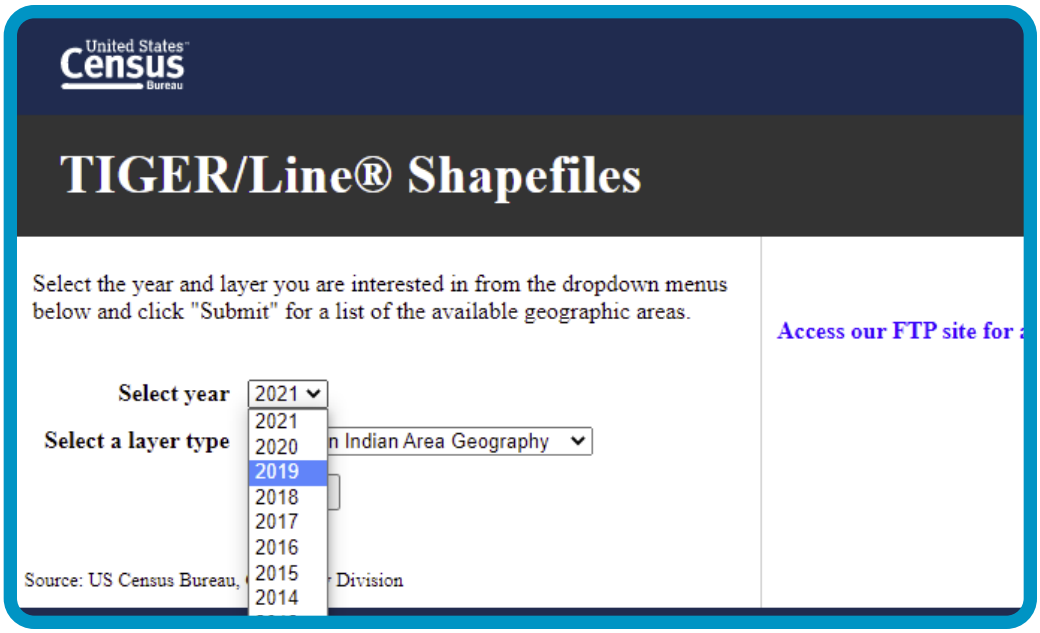
TIGER/Line Shapefiles is a GIS mapping resource maintained by the U.S. Census Bureau. It contains the GIS Shapefiles for the boundaries of various political boundaries like census block groups, municipalities, states, counties, etc and physical boundaries like coastlines, roads, etc. Since political boundaries may change, it provides shapefiles for each year.

Download TIGER/Line Shapefiles

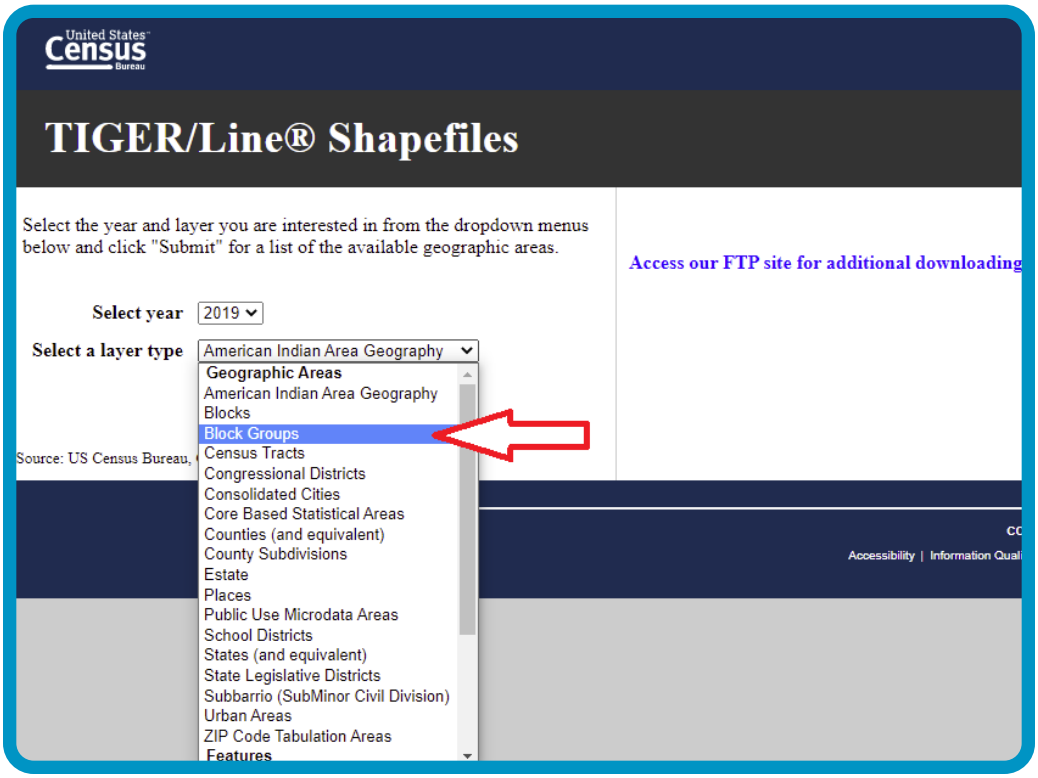
1. Google search "TIGER/Line Shapefiles" or visit www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.html
2. Click on the year you would like to use. We recommend using the year that will match the ACS data you will use. The geographies can change over the years depending on the population changes from the most recent census. As of writing this toolkit, the 2019 ACS data for 5 year estimates is the most recent data accessible.
3. Under 2019 TIGER/Line Shapefiles, and under Download, click on "Web Interface"



4. You'll be directed to a new webpage. Select the year you have decided on by clicking on the drop down arrow near Select year and then which layer type.



5. We will be using block groups (you may want to use a different geography depending on your research question).



6. Click submit

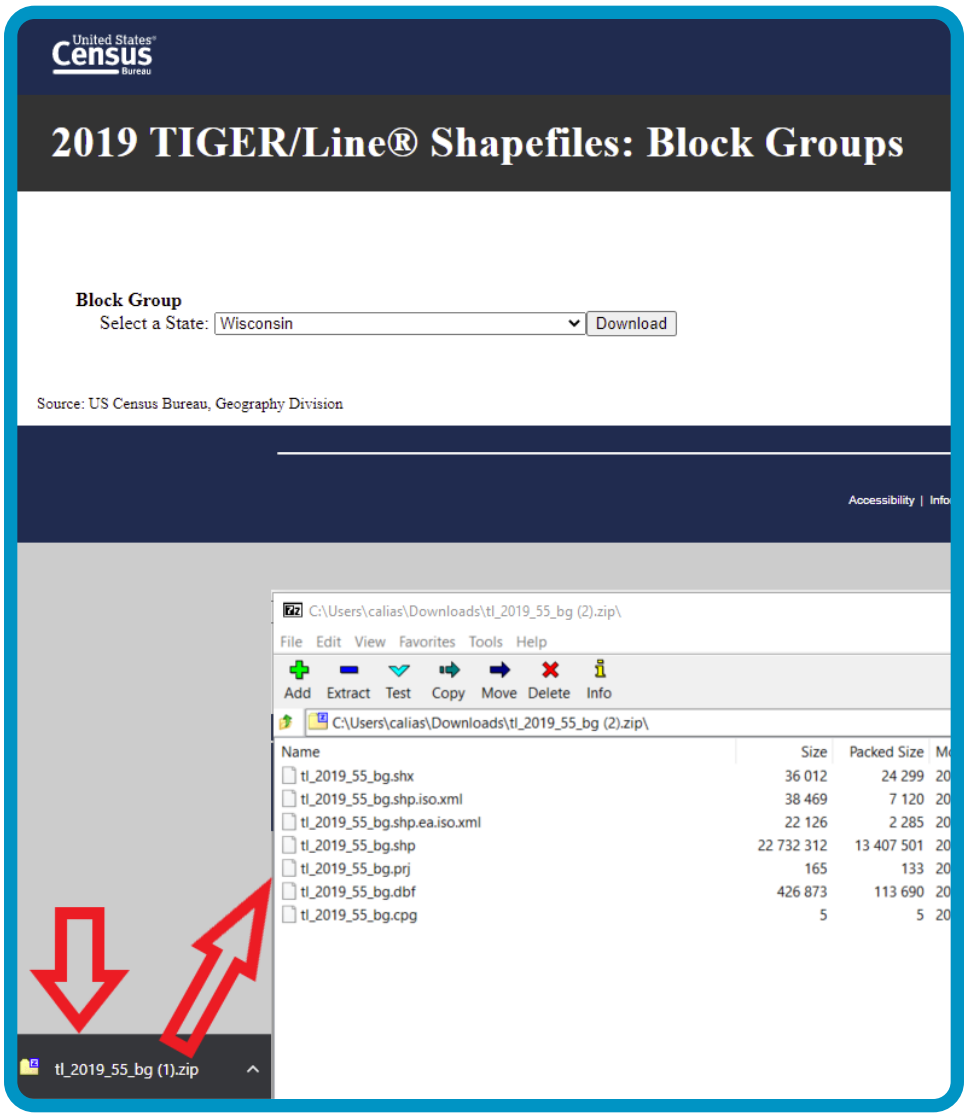
7. Select the state of the municipality(ies) you're interested in

.

8. Click download. You are downloading all of the block groups in that state. After the download is complete, you will see a zipped folder that you will want to extract and save into your boundaries folder. It is a zipped folder because the GIS software connects/relates all of the different types of data within that folder in order to create a visual representation of the location you have chosen.

   The files will be titled tl_[year]_[number]_bg.

   - tl stands for tiger line.
   - year is for the year that you downloaded
   - number refers to which state
   - bg is for block group



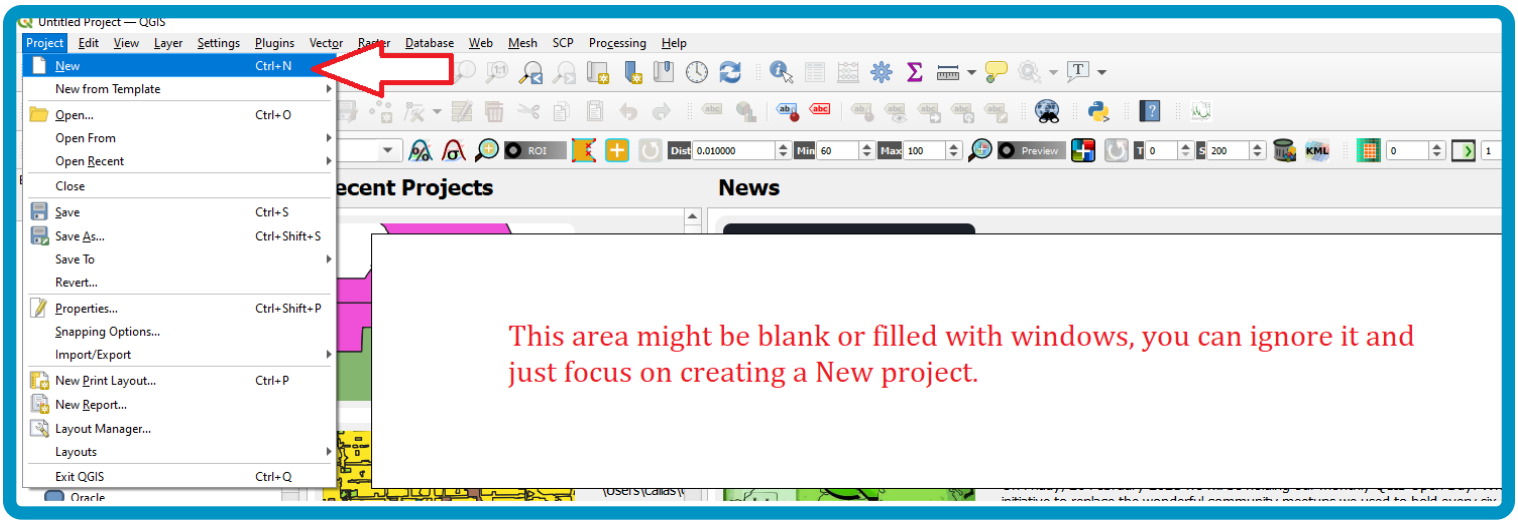
9. Since you are unsure of which block groups are in your municipality, you will also want to download a shapefile of your municipality. You may be able to find this via a Google search (municipality name + boundary + shapefile). If not, you may be able to use the TIGER/Line Shapefiles website. Use steps 4-8, but for "Select a layer type" choose Places and click submit.

10. Then click on the state your municipality(ies) are in and click download. Extract and save to your boundaries folder. You'll notice that these files are titled tl_[year]_[number]_place. Make sure that the number is the same as the number in your block group files to know that they are of the same state. You will see that shapefiles downloaded along with several other files. It is important to keep these files and not delete them, as the shapefile will not work without all of the files on your computer.

Now, we'll open up these files in a GIS Software.

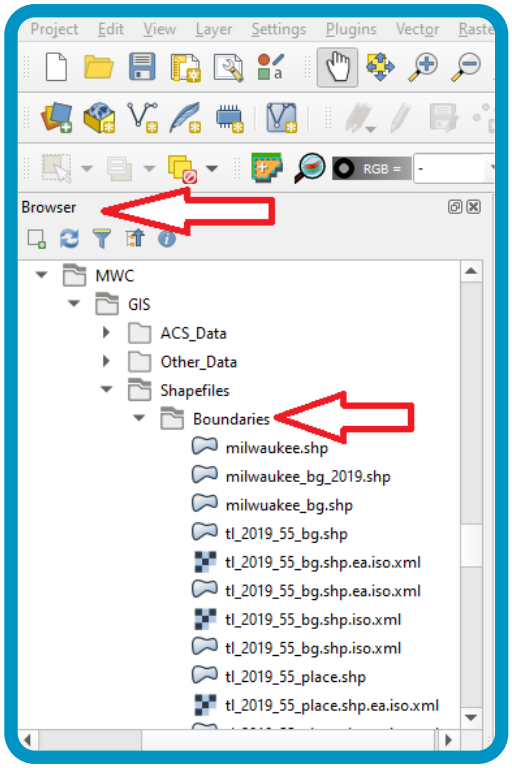
QGIS Setup

1. Open up QGIS and a new project

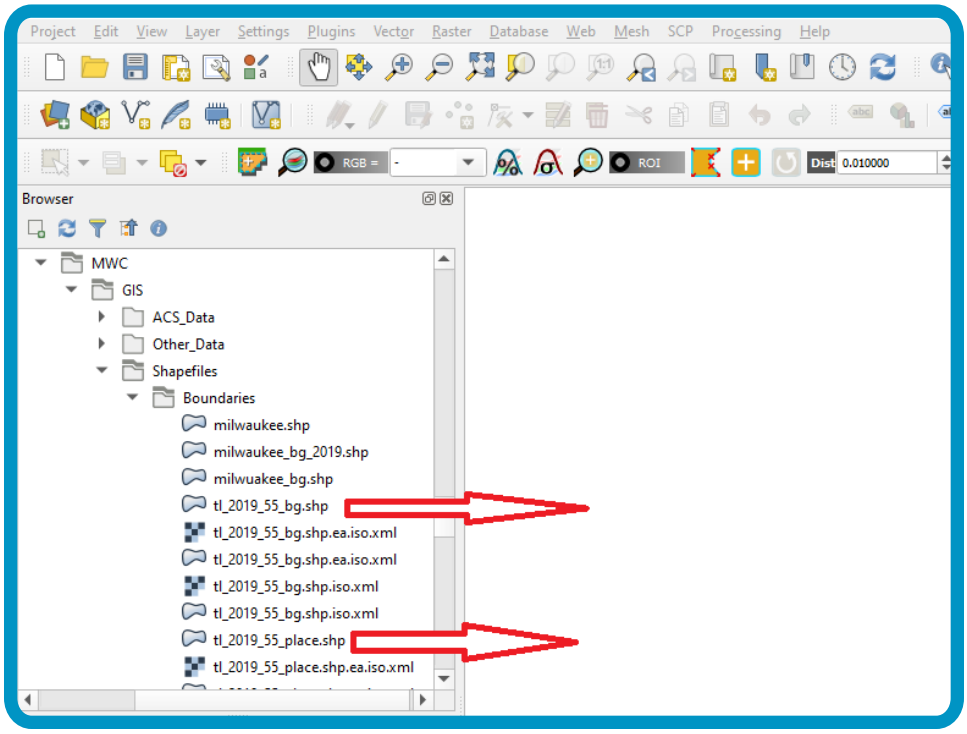


2. Under the project tab, click Save As and name the project whatever you would like. It is important that you frequently Save (Ctrl+S on PC or Command+S on Mac) as you go to avoid losing your work.

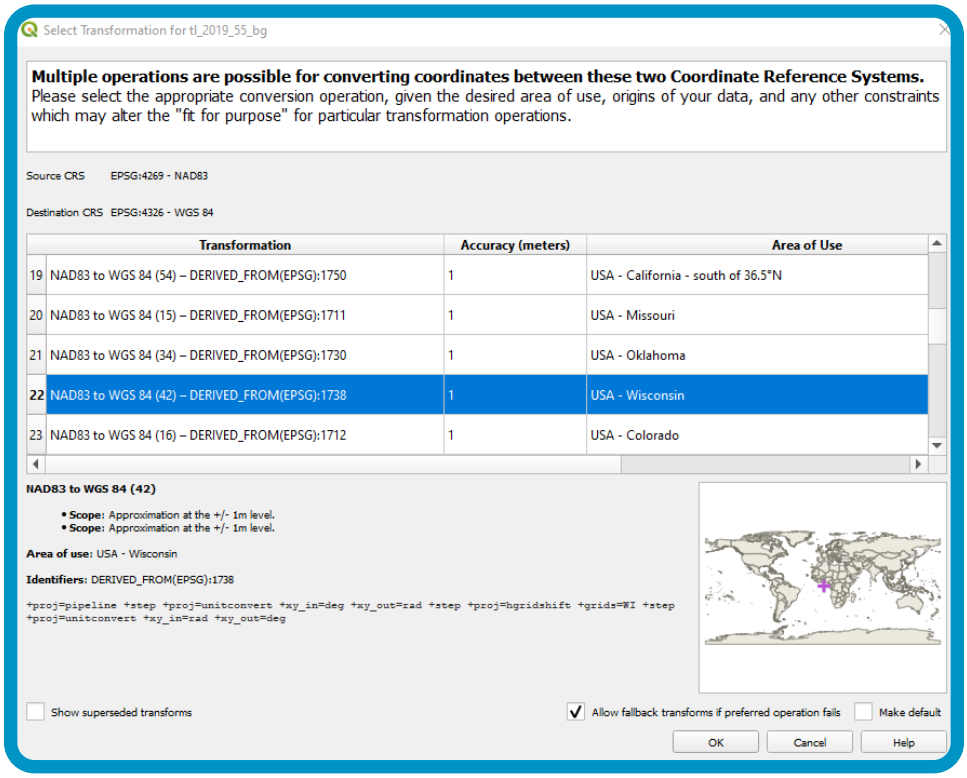
3. On the Left window in Browser, find your boundary data. The Browser tab is a way to locate your data through QGIS, but you may have to dig through the Browser to find the data. Organizing your files in folders will make it easier to locate the data. You can also add your data by dragging it from its location on your computer over the QGIS window and it will automatically load.



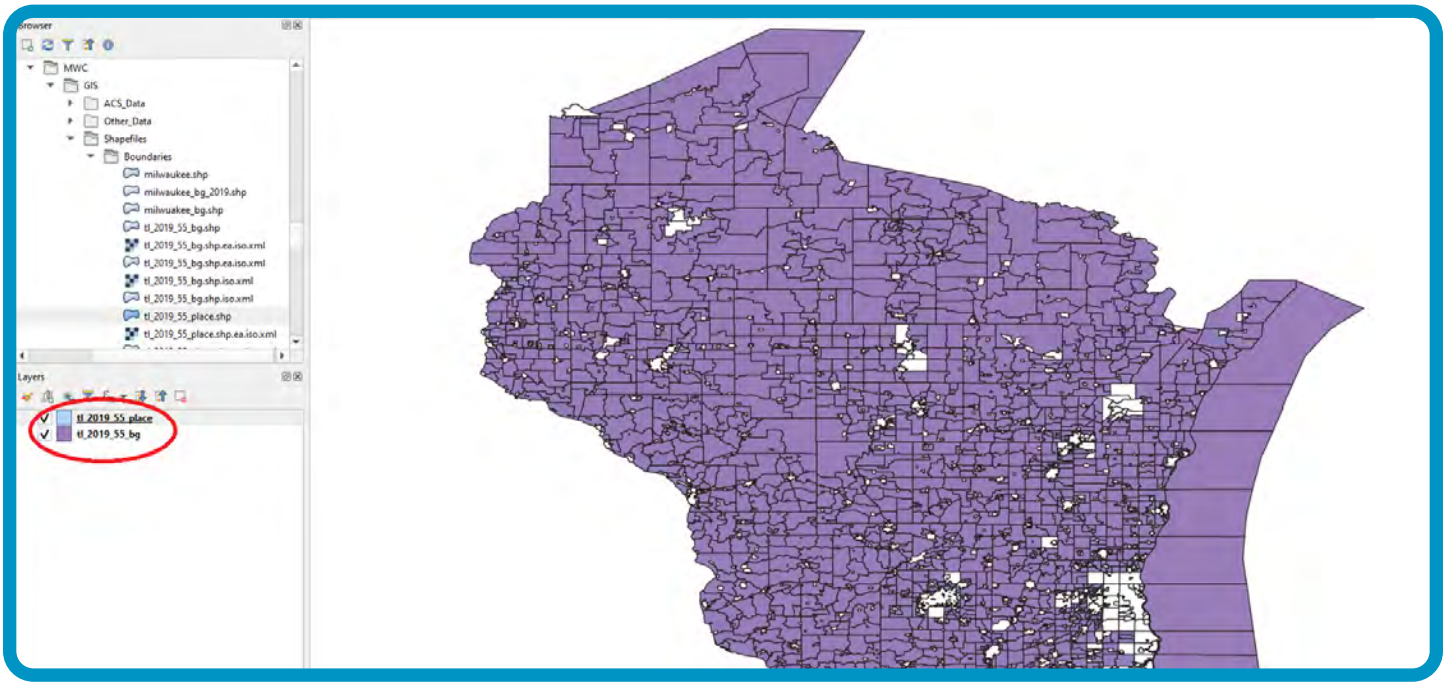
4. Drag the files named tl_[year]_[number]_bg.shp and tl_[year]_[number]_place.shp into the large white empty space. (Your left window may not look like this. If you do not see a Browser tab open, look for "Browser" in the bottom left corner of the QGIS window. You can then drag the tab around and adjust it to your viewing preferences.)



5. It will ask you to select a transformation in order to visualize the data. We suggest looking at the Area of Use (if visible like in the table below) and choosing the transformation that matches the state you're working in. Then click OK.



6. You should now see a map of the block groups and map of the places overlapped on each other. You'll also see both of these shapefiles listed in the layers tab of the left side of QGIS window.

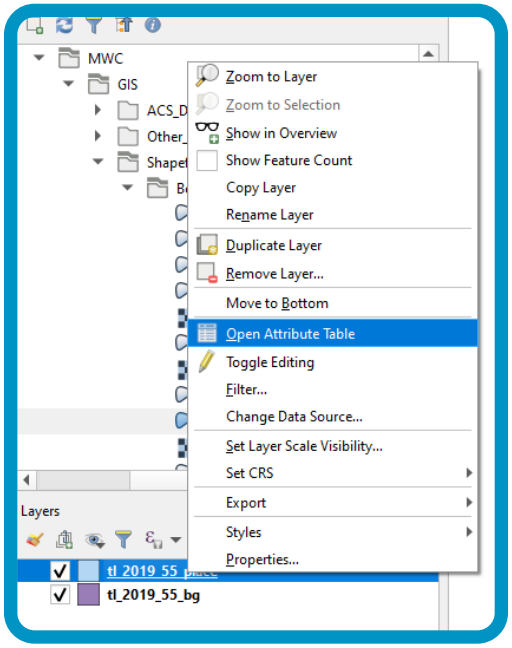


Identify which block groups are in the municipality of interest.

In this example, it is Milwaukee.

Open Attribute Table

1. In the Layers window right click on the tl_[year]_[number]_place and click "Open Attribute Table". This will let you see all of the data that is associated with the map.

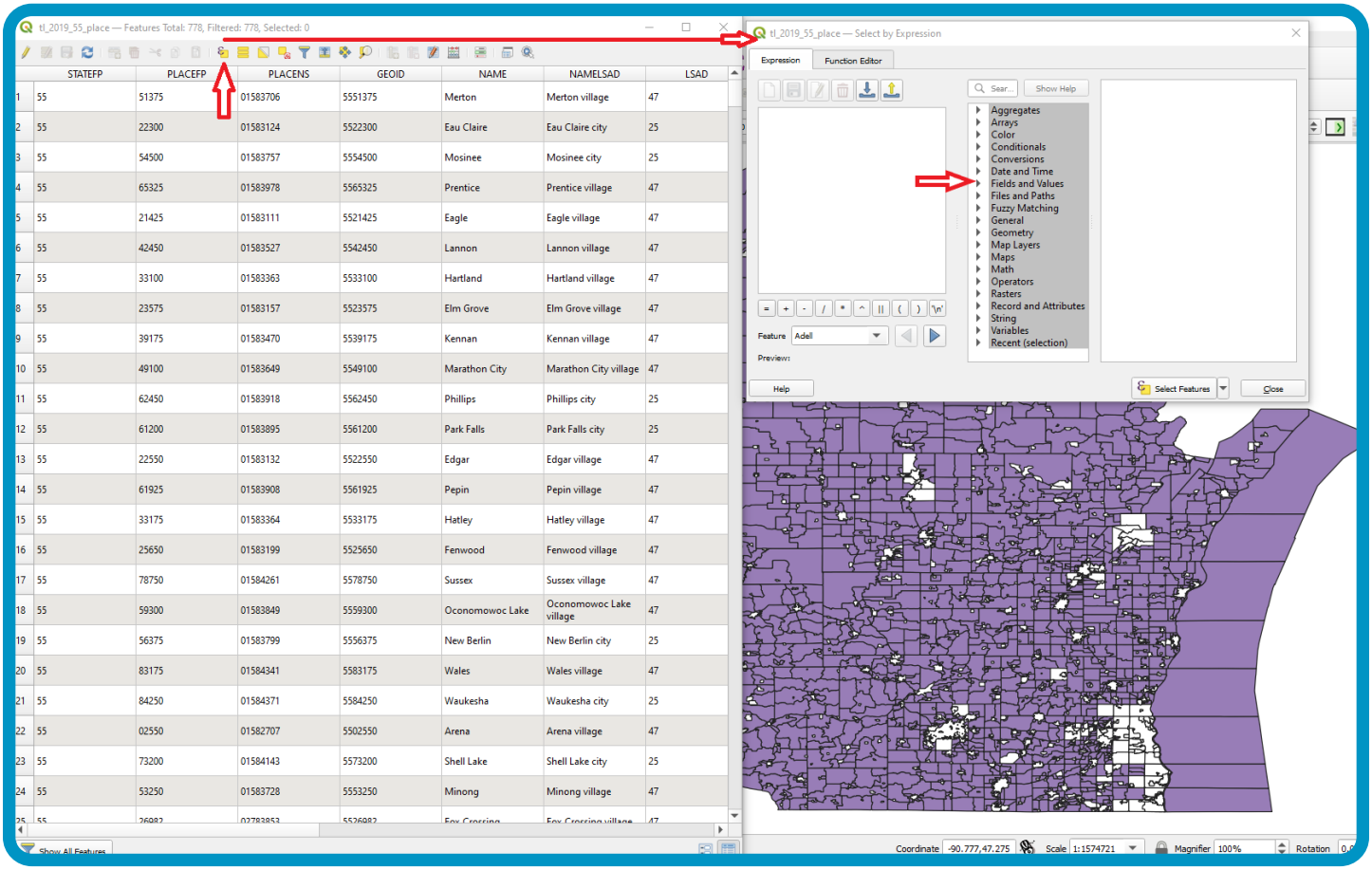


2. When you open up the attribute table you can see all of the attributes (column headings) and all of the different features (each row). Scroll around the table to get a sense of the data available on this map.

Select by Expression

The Select by Expression tool searches through the data table to identify the locations that match up to the expression (math sentence that includes at least one math symbol) you've asked for. For example, you could use it to find locations with a population over 400,000 people (as an expression: >400,000 people). When you've used the tool, the map will visually highlight the locations that match the expression. This will allow you to conduct further analysis on these selected locations.

0. We want to select a specific municipality (Milwaukee). Within the attribute table, click on the "Select by Expression" button at the top of the window which will open another window:

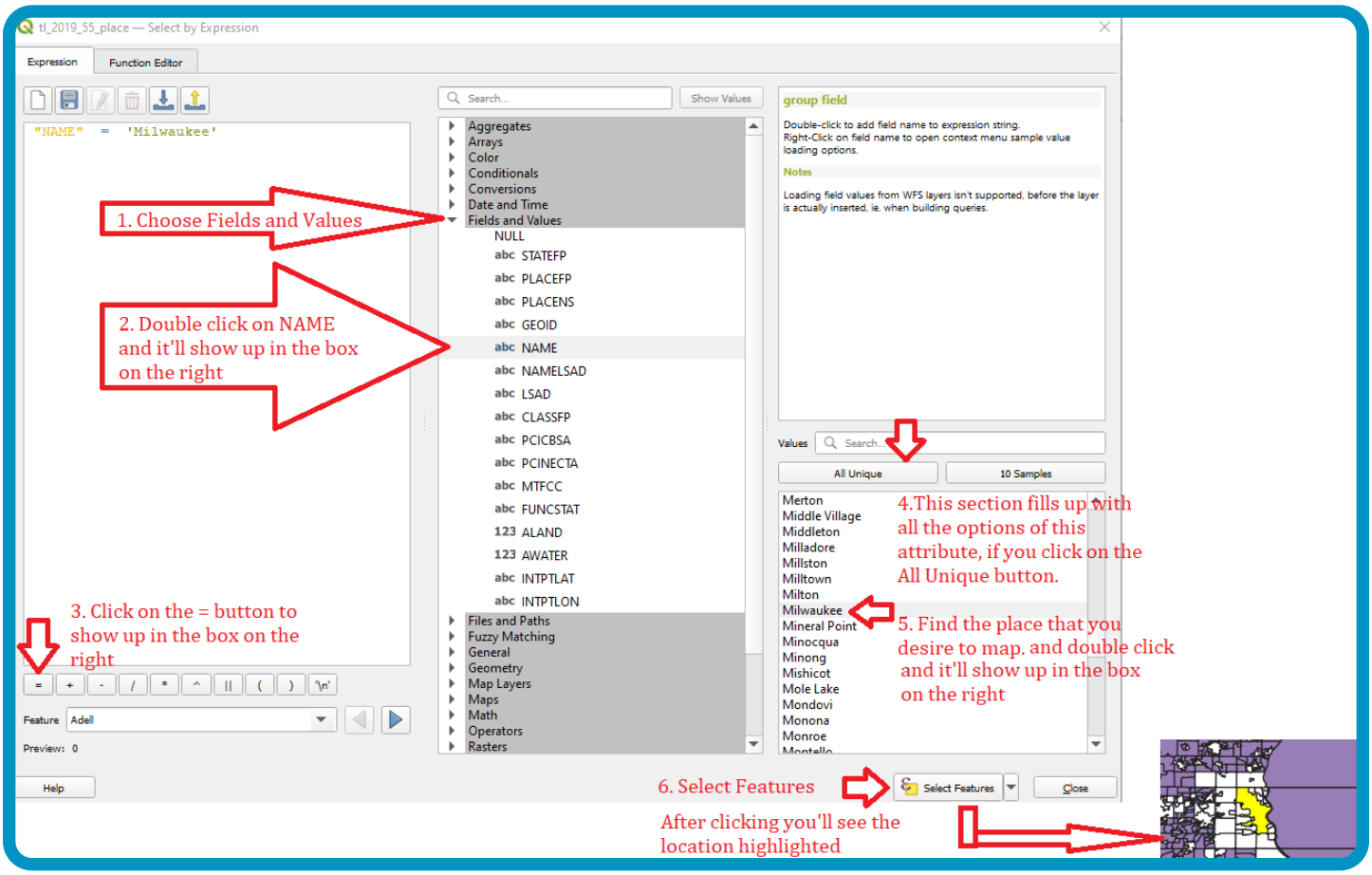


	STATEFP	PLACEFP	PLACENS	GEOID	NAME	NAMELSAD	LSAD
1	55	51375	01583706	5551375	Merton	Merton village	47
2	55	22300	01583124	5522300	Eau Claire	Eau Claire city	25
3	55	54500	01583757	5554500	Mosinee	Mosinee city	25
4	55	65325	01583978	5565325	Prentice	Prentice village	47
5	55	21425	01583111	5521425	Eagle	Eagle village	47
6	55	42450	01583527	5542450	Lannon	Lannon village	47
7	55	33100	01583363	5533100	Hartland	Hartland village	47
8	55	23575	01583157	5523575	Elm Grove	Elm Grove village	47
9	55	39175	01583470	5539175	Kennan	Kennan village	47
10	55	49100	01583649	5549100	Marathon City	Marathon City village	47
11	55	62450	01583918	5562450	Phillips	Phillips city	25
12	55	61200	01583895	5561200	Park Falls	Park Falls city	25
13	55	22550	01583132	5522550	Edgar	Edgar village	47
14	55	61925	01583908	5561925	Pepin	Pepin village	47
15	55	33175	01583364	5533175	Hatley	Hatley village	47
16	55	25650	01583199	5525650	Fenwood	Fenwood village	47
17	55	78750	01584261	5578750	Sussex	Sussex village	47
18	55	59300	01583849	5559300	Oconomowoc Lake	Oconomowoc Lake village	47
19	55	56375	01583799	5556375	New Berlin	New Berlin city	25
20	55	83175	01584341	5583175	Wales	Wales village	47
21	55	84250	01584371	5584250	Waukesha	Waukesha city	25
22	55	02550	01582707	5502550	Arena	Arena village	47
23	55	73200	01584143	5573200	Shell Lake	Shell Lake city	25
24	55	53250	01583728	5553250	Minong	Minong village	47

1. In the middle column, choose “Fields and Values” to find all of the attributes (column headings) from the attribute table.

2. Double click on the column heading that refers to “NAME” and it shows up on the white space to the left.

3. Click the equal button below

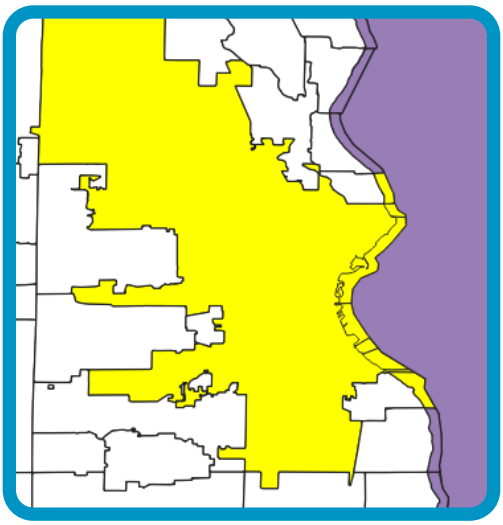


4. To make sure you type the names exactly as they appear in the dataset, the function is case sensitive. You can use the “Value” search to make sure you type the name exactly.

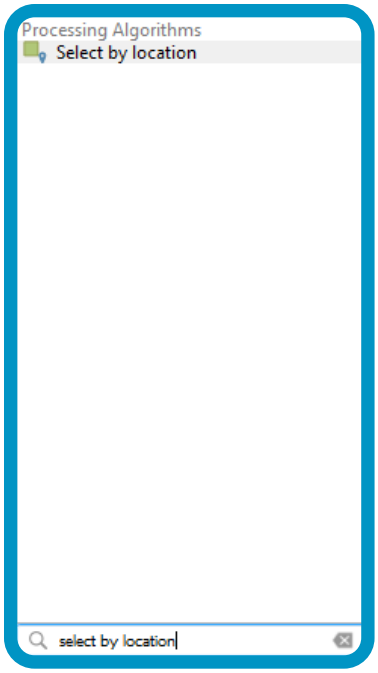
5. Find the place that you desire and double click, so your expression says: “NAME” = ‘Milwaukee’

6. When done, click on “Select Features” and you should see your municipality highlighted you can then click Close on Select by Expression.

7. To get a closer look of your selected municipality, right click the layer and click “Zoom to Selection.”



Select by Location



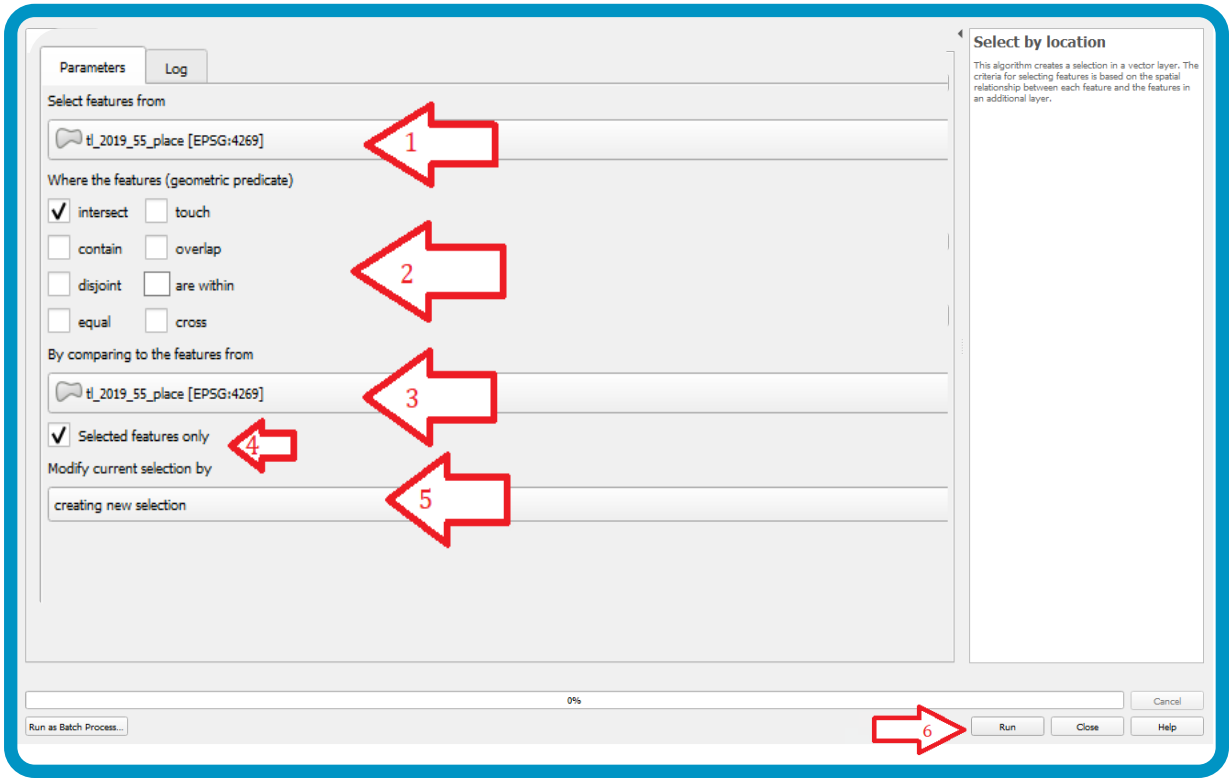
Now we need to also select the block groups that are contained within and intersect your municipality. For this we will use the "Select by Location" tool which chooses locations in one layer based on their spatial relationship to another layer (how do they relate to each other in space - do they overlap, intersect, etc). It's like overlaying a stencil on top of a paper to identify where you want to cut.

When you are working with multiple layers in GIS analysis tools, the software will refer to one layer as the source layer, and the other layer as the target layer. The source layer is the original layer the tool is gathering information from, and the target layer is the layer that will change.

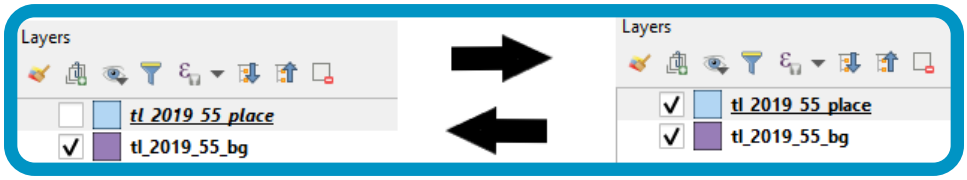
0. At the bottom of the QGIS window, you'll find a search bar. This is how you can search for different tools. In here, search for "Select by Location" and click on it, so that a "Select by Location" window shows up.

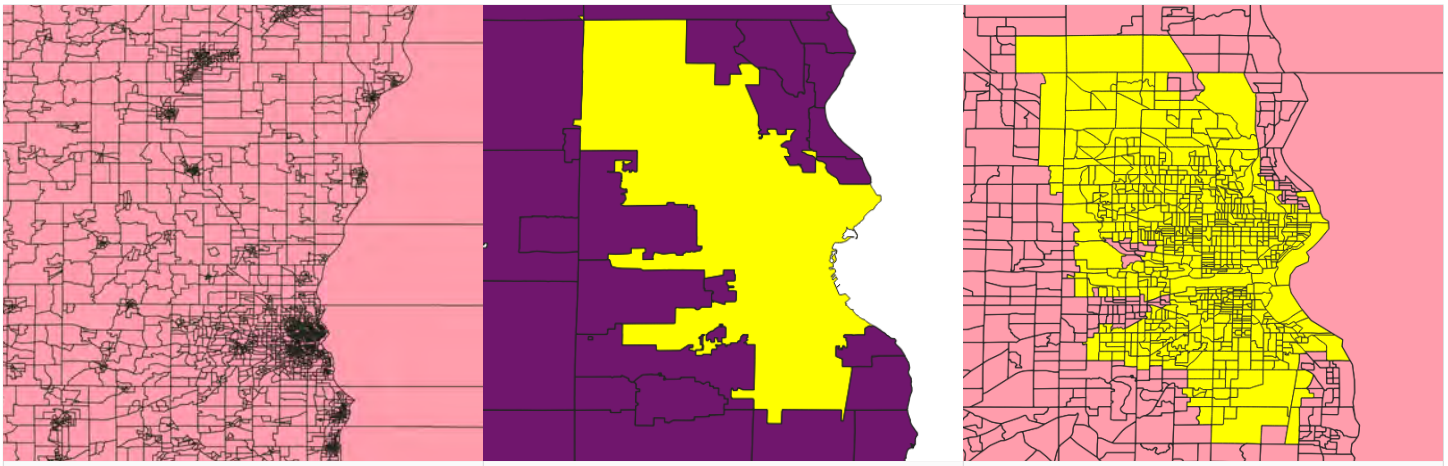
   In the window:

1. Select feature from: tl_[year]_[number]_bg (your block group shapefile)
2. Select "intersect" and "are within" for the section where the features (geometric predicate) which will choose any of the block groups that cross into the municipality's boundary or are within the municipality's boundary.
3. By comparing to the features from: tl_[year]_[number]_place (the shapefile you selected your municipality in)
4. Check mark : selected features only, so it only chooses the block groups that intersect or are within your selected municipality
5. Modify current selection by "creating new selection" so it selects just the block groups
6. Click Run



7. In the Layers window, turn on and off the tl_[year]_[number]_place layer by clicking the checkbox to make sure all the block groups have been included. It may include some extra, but that is okay for the time being.





Before Block Groups Selected

Milwaukee Selected
(Place layer)

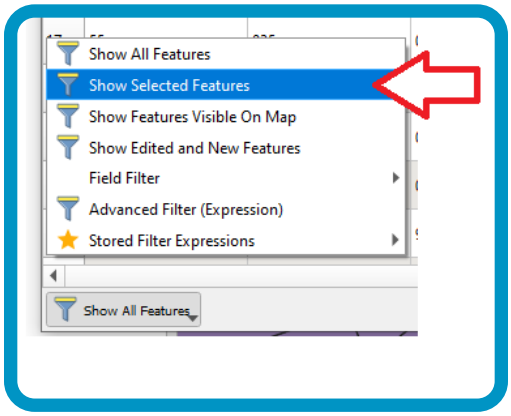
After Block Groups of Milwaukee Selected
(Blockgroup layer)

Remember: It is important that you frequently Save (Ctrl+S on PC or Command+S on Mac) as you go to avoid losing your work.

Show Selected Features in Attribute Table

1. Open up the attribute table for the block group layer. (Refer to section 5.B)

2. At the bottom of the table choose Show Selected Features



3. Now your table only shows the selected block groups

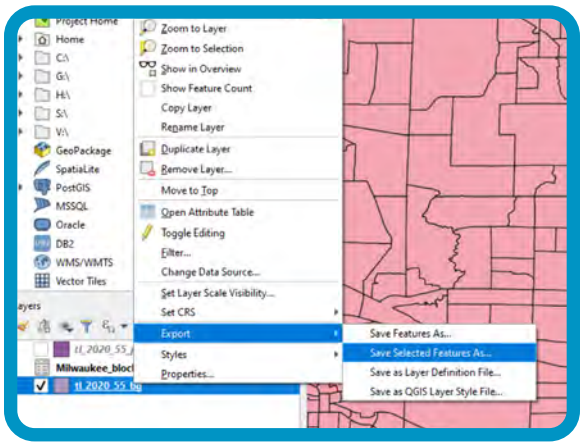
tl_2019_55_bg — Features Total: 4489, Filtered: 685, Selected: 685

	STATEFP	COUNTYFP	TRACTCE	BLKGRPCE	GEOID	NAMELSAD	MTFCC	FUNCSTAT	ALAND	AWATER	INTPTLAT	INTPTLON
9	55	079	013000	2	550790130002	Block Group 2	G5030	S	222534	0	+43.0260187	-087.9890803
0	55	079	008700	1	550790087001	Block Group 1	G5030	S	164043	0	+43.0696480	-087.9399386
1	55	079	091300	1	550790913001	Block Group 1	G5030	S	310226	0	+43.0370681	-087.9990117
2	55	079	060200	1	550790602001	Block Group 1	G5030	S	2342984	39906	+43.1057300	-087.9138805
3	55	133	200202	1	551332002021	Block Group 1	G5030	S	7490642	186166	+43.1708879	-088.0732201
4	55	079	980000	1	550799800001	Block Group 1	G5030	S	2024683	4018925	+43.0444511	-087.8873906
5	55	079	006600	3	550790066003	Block Group 3	G5030	S	209703	0	+43.0732434	-087.9271937
6	55	079	008900	2	550790089002	Block Group 2	G5030	S	484007	0	+43.0664406	-087.9528417
7	55	079	009800	1	550790098001	Block Group 1	G5030	S	440162	0	+43.0597780	-087.9540957
8	55	079	090900	4	550790909004	Block Group 4	G5030	S	384494	0	+43.0663313	-088.0055348
9	55	079	006900	3	550790069003	Block Group 3	G5030	S	244250	0	+43.0800232	-087.9139239
0	55	079	002200	2	550790022002	Block Group 2	G5030	S	593565	0	+43.1148356	-087.9391414
1	55	079	000301	1	550790003011	Block Group 1	G5030	S	1985249	1761	+43.1361484	-088.0617334
2	55	079	006700	2	550790067002	Block Group 2	G5030	S	233319	0	+43.0730178	-087.9204526
3	55	079	160100	3	550791601003	Block Group 3	G5030	S	1172706	0	+42.9233496	-087.9450304
4	55	079	013300	1	550790133001	Block Group 1	G5030	S	596295	0	+43.0348359	-087.9620643
5	55	079	007400	2	550790074002	Block Group 2	G5030	S	432945	476790	+43.0777349	-087.8704971
6	55	079	120400	5	550791204005	Block Group 5	G5030	S	821287	0	+42.9683620	-087.9664557
7	55	079	150100	2	550791501002	Block Group 2	G5030	S	1317879	0	+42.9232275	-087.9548659
8	55	079	120102	3	550791201023	Block Group 3	G5030	S	678821	0	+42.9700414	-088.0330058
9	55	079	120400	2	550791204002	Block Group 2	G5030	S	1981947	0	+42.9655628	-087.9566847
0	55	079	000102	3	550790001023	Block Group 3	G5030	S	1224325	0	+43.1712874	-087.9995252
1	55	079	187000	2	550791870002	Block Group 2	G5030	S	98859	0	+43.0504066	-087.8938680
2	55	079	186900	1	550791869001	Block Group 1	G5030	S	430121	0	+43.0557700	-087.8827615
3	55	079	187000	1	550791870001	Block Group 1	G5030	S	196343	0	+43.0578597	-087.8834148

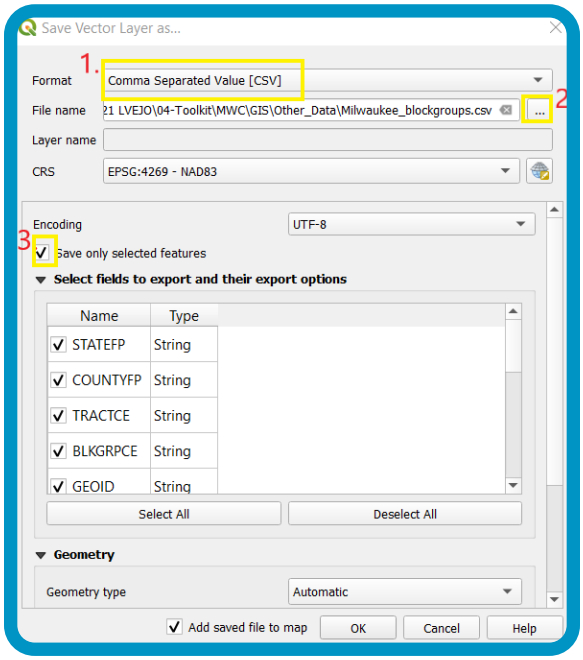
Export Selected Features as a Table (in CSV)

We want to have a list of the block groups that we can use to collect ACS data. You'll want to save this in your folder structure (from section 4.G), likely in the Other_Files folder. You'll save it as a CSV (a comma separated values document) which can be opened up in Excel as a table. It is a way to save large tables that might otherwise have trouble opening in Excel.

0. Right click on the block group layer, go to export, and click on "save selected features as..."



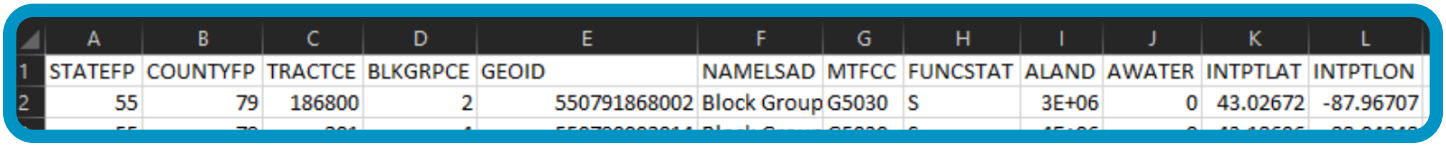
In the window:



1. Format: Comma Separated Values
2. File name: click on the "..." button and save in your Other_Data folder with a title like [municipality]_blockgroups
3. Check on Save only selected features
4. Leave everything else as is. (Take a look at the QGIS user guide, if you're interested in more details.)
5. Click okay to save a table of the block groups you are interested in

6. Open up this CSV table in Excel and you will now have a list of all the block groups

   If you are using a Mac, you will need to open the CSV through Excel or Google Sheets, not "Numbers" (the default for Mac). To open using Excel, right click on the file, go to "open with" and select Microsoft Excel. You can also do this by dragging and dropping the CSV file over the Excel application.



	A	B	C	D	E	F	G	H	I	J	K	L
1	STATEFP	COUNTYFP	TRACTCE	BLKGRPCE	GEOID	NAMELSAD	MTFCC	FUNCSTAT	ALAND	AWATER	INTPTLAT	INTPTLON
2	55	79	186800	2	550791868002	Block Group	G5030	S	3E+06	0	43.02672	-87.96707

An explanation of the column headings:

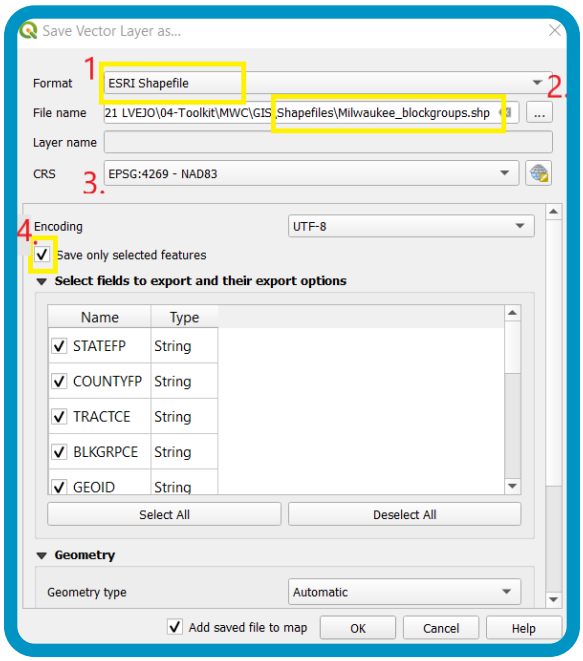
- STATEFP - the state's ID number, you'll notice this is the same as the number in tl_[year]_[number]_bg/place
- COUNTYFP - the county's ID number (also called FIPS - Federal Information Processing Standards)
- TRACTCE - the census tract numbers, this is what you will use when searching in ACS
- BLKGRPCE - the specific block group number
- GEOID - this is a combination of the STATEFP, COUNTYFP, TRACTCE, BLKGRPCE values (we use this ID to join ACS data to the shapefiles, but more on that later in the chapter).
- The other columns name the block group, share how much area is land, area is water, and provides a specific point of that block group. We will not really use these in the toolkit, but the ALAND and AWATER columns may be helpful in some analyses that you might be interested in.

Export Selected Features as Shapefile

At this point, it is also a good idea to save a shapefile of just the selected block groups, so that you do not have to work with the whole state when you want to start joining and visualizing data.

Repeat steps under Exporting Selected Features as CSV except with the following changes, Right click on layer, click export, choose "Save Selected Features As..."

1. Format: ESRI Shapefile
2. File Name: Save in your GIS >> Shapefiles >> Boundaries Folder as "[municipality]_bg"
3. CRS: click on the icon and choose your coordinate reference system (for example, NAD83(2011)/UTM Zone 16N for Milwaukee)
4. Check on Save only selected features
5. Keep everything as is
6. Click okay



You may also want to save your municipal boundary from the place layer.

- Complete steps Select by Expression
- Complete steps Exporting Selected Features as Shapefile

Remember: It is important that you frequently Save (Ctrl+S on PC or Command+S on Mac) as you go to avoid losing your work.

C. Data Dictionary

What is a Data Dictionary?

A data dictionary is a list of the data layers on the map, their attributes, definitions of those attributes, data source (where the data came from), year of data, etc. It informs people about the whole map you are making. This is different from metadata which describes the data for each layer/shapefile. The metadata is created by those who created the data.

Why are data dictionaries necessary?

The data dictionary is used to document the various data used in a project and makes it easier to share data layers with others so they know how to interpret the data. It is also a tool to help the GIS user organize their data layers so they can be aware of and reduce inconsistencies or incorrect use of the data

What should be included?

The data dictionary serves as a reference for future users. Some of the variables often included in the dictionary are:

- **Attribute name** = pct_bipoc
- **Definition** = Percent of total population that identified as Black, Indigenous, People of Color
- **Data type** = number, percent, or text
- **Source** = American Community Survey, 5-year estimates, 2019, Table B03002
- **Other** = depending on the data layer being documented, there might be data saved as codes – for e.g., land use or zoning data. The other column can be used to add descriptions of the codes.

How do you create a data dictionary?

A data dictionary is a table. You can create one in commonly used software such as Microsoft Excel or Microsoft Word. Examples of two different data dictionaries are provided below.

Example 1: Demographic data			
Attribute name	Description	Data Type	Source
Tot_pop	Total number of people residing in a census block group	Number	American Community Survey, 5-year estimates, 2019, Table B01003
Pct_bipoc	Percent of total population that identified as Black, Indigenous, People of Color	Number	American Community Survey, 5-year estimates, 2019, Table B03002

Example 2: Land Use Data				
Attribute name	Description	Source	Other	
Land Use	Characterization of the land based on its usage	Regional Planning Authority	Code	Land Use
			1110	Single-Family Residential
			1130	Multi-Family Residential
			1200	Commercial
Zoning	Characterization of the land based on what the law allows	City	Code	Zoning Class
			RS	Residential Single-Unit District
			RT	Residential Two-Flat, Townhouse and Multi-Unit District

D. Data Gathering for the Issue

When considering data to use for your projects, there are several sources available online. It is recommended to use data from reputable sources that clearly explain and validate their methodology (how did they create the dataset). You can look at the metadata (information about the dataset often saved in another accessible document when data is downloaded) to see where the data came from or how it was created. Another important factor to consider is what time frame the data represents. If you are trying to create a map of the current situation, choose data that is most recent. It is important to know that available data may not be as recent as you wish it to be because funding for data collection is often limited. If you are hoping to show the changes in a spatial area over time, be sure that you are staying as consistent as possible with your datasets or considering the differences in time when drawing your conclusions.

For the rest of this section, we will dive into EnviroAtlas as a reputable source for environmental data. Section 5.G includes additional data resources for other topics.

Introduction to EnviroAtlas

EnviroAtlas is an interactive web-based tool with downloadable data created by the United States Environmental Protection Agency (U.S. EPA) in partnership with many other federal, state, and local governmental agencies. Funded through at least 2022, you can access EnviroAtlas at epa.gov/enviroatlas and it contains over 400 datasets on various environmental topics at varying geographic sizes for the contiguous US as well as some of Alaska, Hawaii, Puerto Rico, and the Virgin Islands. The goal of EnviroAtlas is to allow policy makers, planners, and ordinary citizens to access data that will help them better understand tradeoffs and relationships between nature/ecosystems, well-being/quality of life, public health, and the economy.

Available datasets are organized into two main categories –national and community. Most national datasets in EnviroAtlas are available at the subwatershed level, meaning portions of larger watersheds like the Great Lakes or Mississippi River watersheds. These subwatersheds are identified by a 12-digit hydrologic unit code (HUC) in the US. The subwatershed size is often too large for municipal or community level studies. However, it can be useful in creating regional or large-scale maps. EnviroAtlas also has thirty defined community areas across the US for which the geographic size of the data are available at the census block group size. This toolkit will walk through downloading and visualizing data for two environmental concepts for one of the community areas, Milwaukee.

More information about EnviroAtlas community data can be found in this fact sheet. For more information on the community data, visit this document and check whether EnviroAtlas contains data for your municipality!

It is important to note that EnviroAtlas datasets intentionally include only the most recent data for all its environmental topics and parameters. Therefore, it is not a resource to use for studies to show a change over time.

Other Useful EnviroAtlas Resources

EnviroAtlas contains some helpful resources, fact sheets, and guides on understanding its data and how to use it more. The interactive web tool, which allows users to map and view data layers online, can be found here. Note that not all data that is available for download can be viewed in the web tool. A guide on how to use this tool can be found here.

EnviroAtlas also contains an Eco-Health Relationship Browser that helps users understand and teach about the direct linkages between human health and ecosystem services. This tool can be a great resource on how data from the EnviroAtlas database can be used for EJ purposes. Also useful for EJ organizations is the EnviroAtlas Use Cases page that contains previous reports and studies that have been generated based on analyses using data from EnviroAtlas.

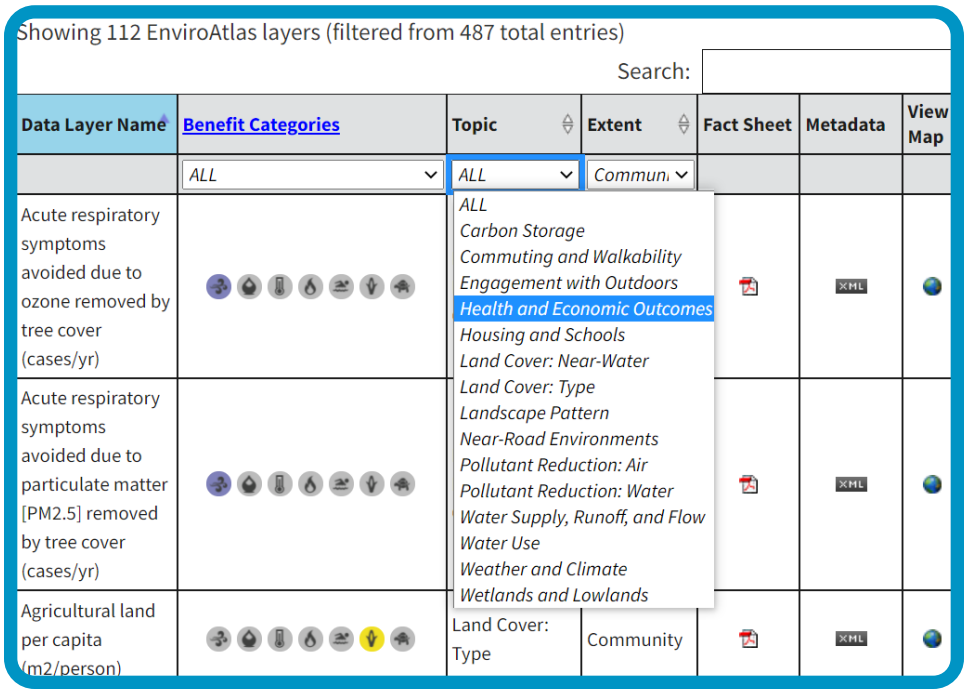
Deciding What Data to Use

It is helpful to access the Dynamic Data Matrix prior to downloading data from EnviroAtlas. This page contains an easy-to-use setup in which all available environmental data can be browsed and read about by name, benefit category, topic, and/or spatial extent. Each column of the Dynamic Data Matrix can be filtered to sift through the eight different benefit categories and thirty-three topics.

1. Click the drop-down menus for these two columns to view some of these. Since this toolkit is focusing on community level data, click the drop-down menu at the top of the "Extent" column and choose "Community". Now we have displayed only datasets available at the community level.

For this tutorial, one question we want to answer is "Are there geographic disparities in experiencing health benefits from trees within Milwaukee?"

2. To find appropriate data, click on the drop-down menu for the "Topic" column and click "Health and Economic Outcomes". You will now see twelve datasets, to choose from that pertain to health and economic outcomes and are available at the community level.

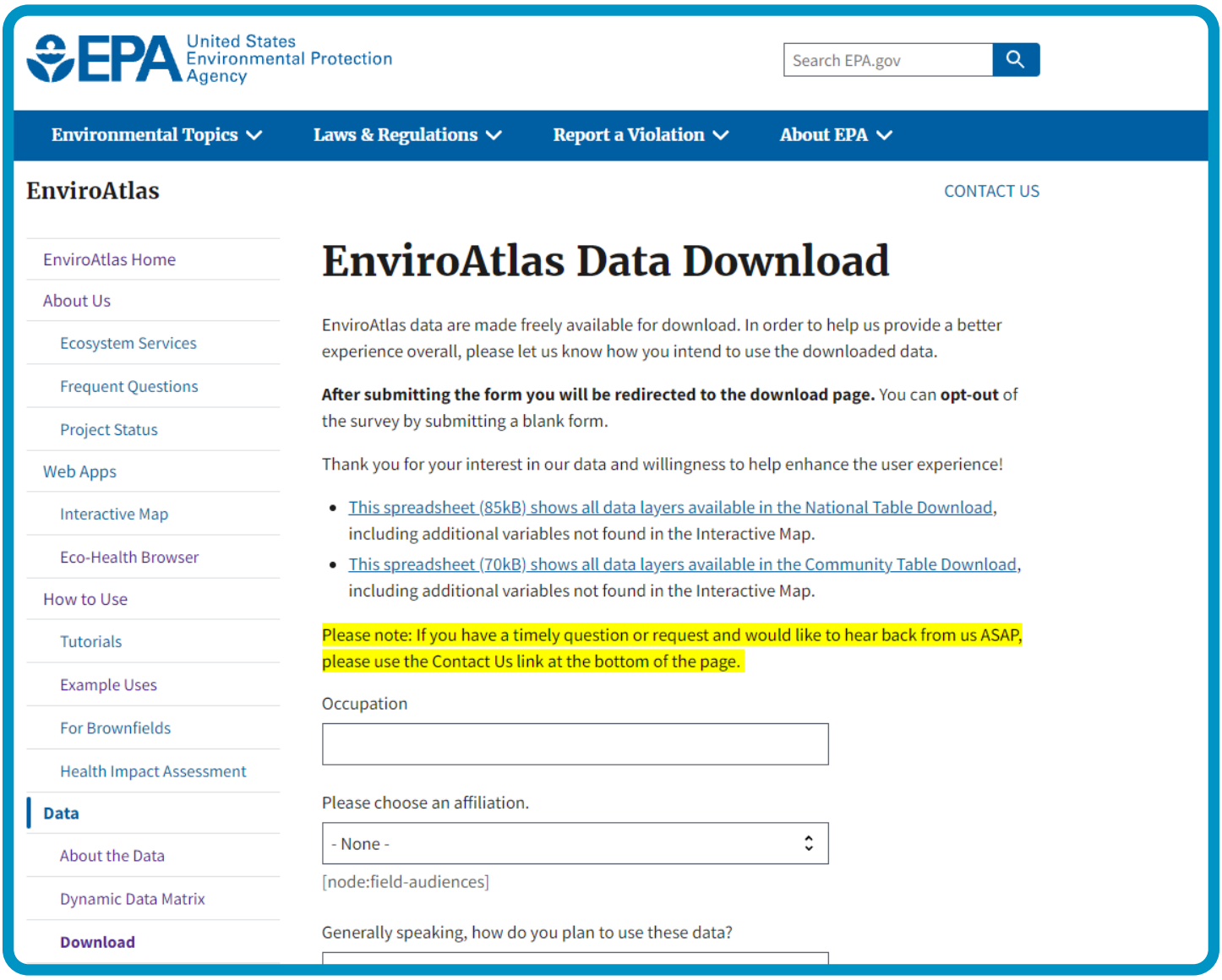


3. Take a moment to browse through the different layers available under "Data Layer Name". We will be using Acute respiratory symptoms avoided due to particulate matter [PM2.5] removed by tree cover (cases/yr) as one of our example layers for download and visualization.

4. To read more about these data or any other layers, click on the PDF icon under the "Fact Sheet" column to open a document containing helpful background, definitions, sources, and more about the layer. To view what a layer may look like before downloading, click the globe icon under the "View Map column" which opens the layer in EnviroAtlas's online web tool.

Downloading EnviroAtlas Data

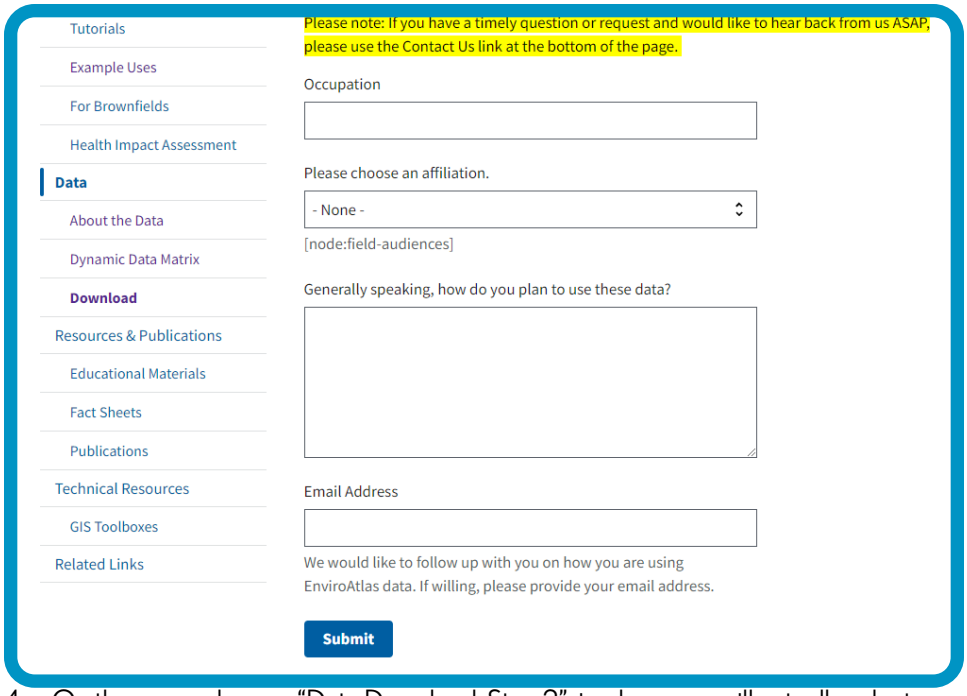
1. Access the EnviroAtlas Data Download page here. This page is not only necessary to navigate to data downloads, but also contains links to two spreadsheets. These spreadsheets are necessary for locating specific data layers of interest within the national data or community data.



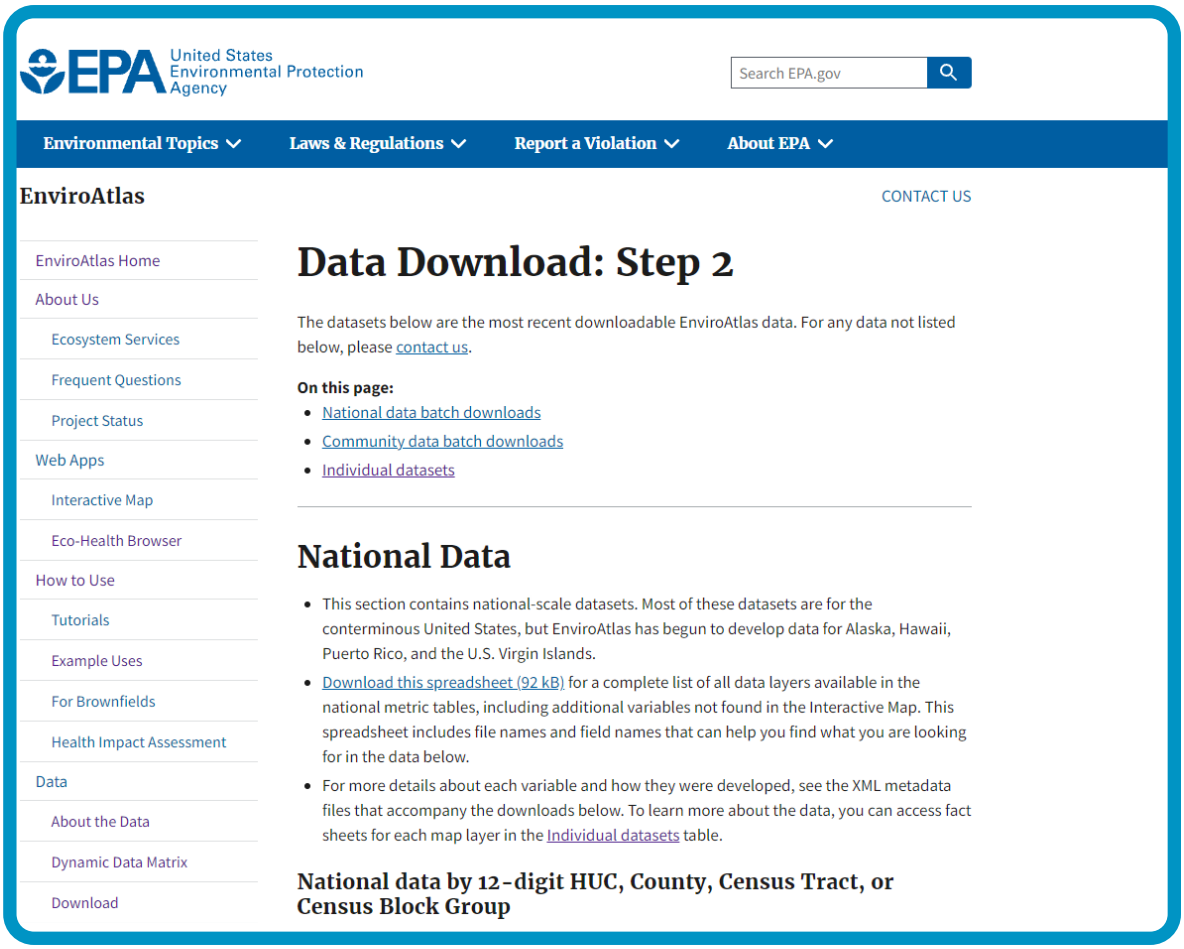
2. Depending on which type(s) of data are desired for your project, click the link(s) to download the spreadsheet for either or both the National Table and the Community Table and save it in the project folder. For this tutorial we will need the community table and throughout the tutorial we will call the downloaded community table: "***the comprehensive Community Data Excel spreadsheet***".

   These spreadsheets help users locate data layers of interest within downloaded files, IDs for the data, the year(s) that the data represent, and other helpful information.

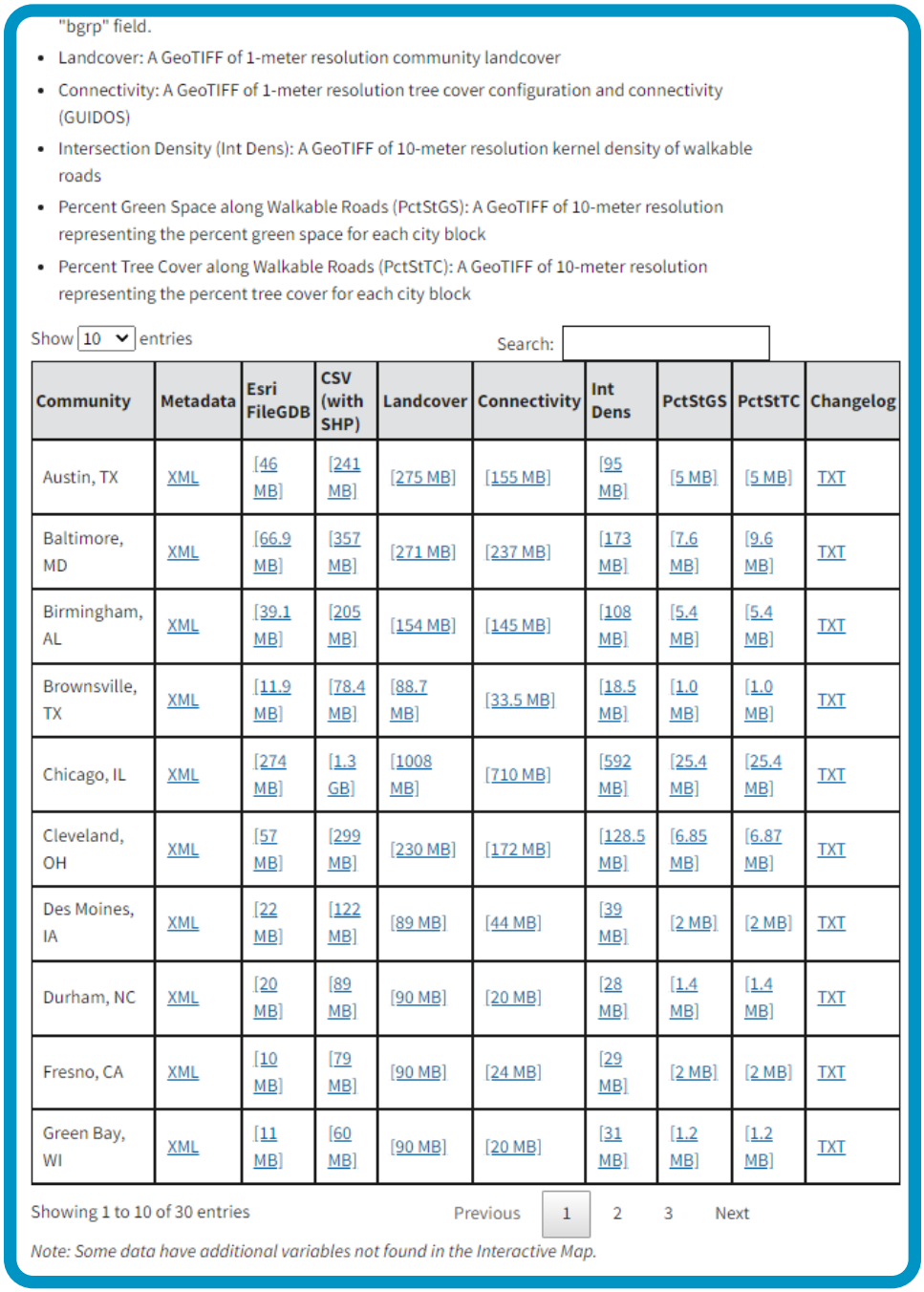
3. Scroll to the bottom of the page. Fill out the optional survey if you wish, then click "Submit."



4. On the new webpage, "Data Download: Step 2", is where you will actually select and download data. The Step 2 page lists the National Data first followed by the Community Data.



5. For this tutorial, scroll to the Community Data section. This section provides basic information about the community data and a table containing download links organized by community area. Each of the columns of this table are explained above the table in bullet points. For the rest of this tutorial we will call this online table: "***the online Community Data table***".



"bgrp" field.

- Landcover: A GeoTIFF of 1-meter resolution community landcover
- Connectivity: A GeoTIFF of 1-meter resolution tree cover configuration and connectivity (GUIDOS)
- Intersection Density (Int Dens): A GeoTIFF of 10-meter resolution kernel density of walkable roads
- Percent Green Space along Walkable Roads (PctStGS): A GeoTIFF of 10-meter resolution representing the percent green space for each city block
- Percent Tree Cover along Walkable Roads (PctStTC): A GeoTIFF of 10-meter resolution representing the percent tree cover for each city block

Show 10 entries Search:

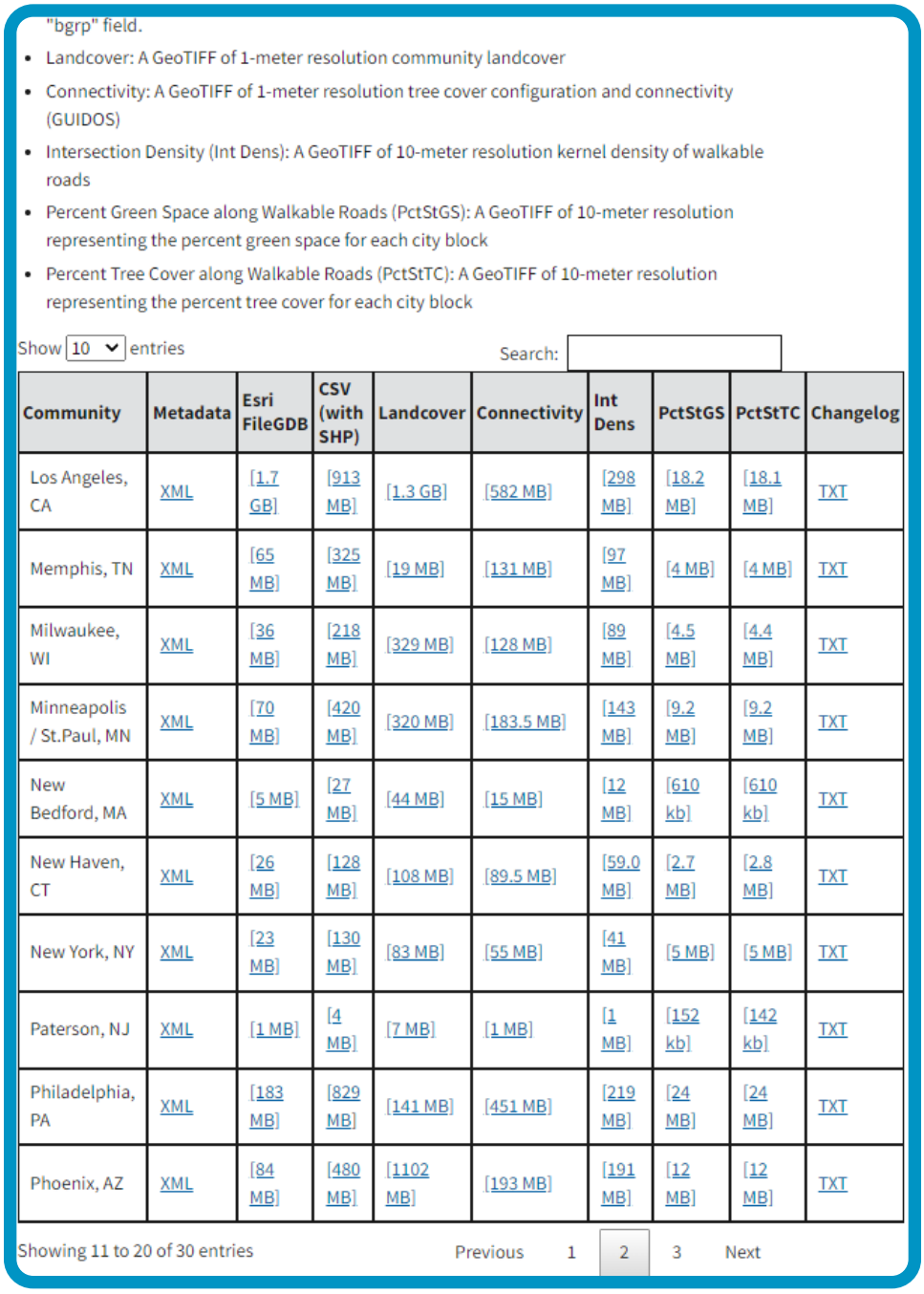
Community	Metadata	Esri FileGDB	CSV (with SHP)	Landcover	Connectivity	Int Dens	PctStGS	PctStTC	Changelog
Austin, TX	XML	[46 MB]	[241 MB]	[275 MB]	[155 MB]	[95 MB]	[5 MB]	[5 MB]	TXT
Baltimore, MD	XML	[66.9 MB]	[357 MB]	[271 MB]	[237 MB]	[173 MB]	[7.6 MB]	[9.6 MB]	TXT
Birmingham, AL	XML	[39.1 MB]	[205 MB]	[154 MB]	[145 MB]	[108 MB]	[5.4 MB]	[5.4 MB]	TXT
Brownsville, TX	XML	[11.9 MB]	[78.4 MB]	[88.7 MB]	[33.5 MB]	[18.5 MB]	[1.0 MB]	[1.0 MB]	TXT
Chicago, IL	XML	[274 MB]	[1.3 GB]	[1008 MB]	[710 MB]	[592 MB]	[25.4 MB]	[25.4 MB]	TXT
Cleveland, OH	XML	[57 MB]	[299 MB]	[230 MB]	[172 MB]	[128.5 MB]	[6.85 MB]	[6.87 MB]	TXT
Des Moines, IA	XML	[22 MB]	[122 MB]	[89 MB]	[44 MB]	[39 MB]	[2 MB]	[2 MB]	TXT
Durham, NC	XML	[20 MB]	[89 MB]	[90 MB]	[20 MB]	[28 MB]	[1.4 MB]	[1.4 MB]	TXT
Fresno, CA	XML	[10 MB]	[79 MB]	[90 MB]	[24 MB]	[29 MB]	[2 MB]	[2 MB]	TXT
Green Bay, WI	XML	[11 MB]	[60 MB]	[90 MB]	[20 MB]	[31 MB]	[1.2 MB]	[1.2 MB]	TXT

Showing 1 to 10 of 30 entries Previous 1 2 3 Next

Note: Some data have additional variables not found in the Interactive Map.

6. Click the "Next" button at the bottom of the matrix to access the Milwaukee, WI data.

7. Draw your attention to the third column of the table, "CSV (with SHP)". This column has a link to download the entire set of available community data layers in vector format. Some of the other columns include data layers that are in raster format and others include data layers in vector format but are formatted differently than a CSV. Keep this webpage up, because we will be referring to it in the next few steps to download the data layers that we are interested in.



"bgrp" field.

- Landcover: A GeoTIFF of 1-meter resolution community landcover
- Connectivity: A GeoTIFF of 1-meter resolution tree cover configuration and connectivity (GUIDOS)
- Intersection Density (Int Dens): A GeoTIFF of 10-meter resolution kernel density of walkable roads
- Percent Green Space along Walkable Roads (PctStGS): A GeoTIFF of 10-meter resolution representing the percent green space for each city block
- Percent Tree Cover along Walkable Roads (PctStTC): A GeoTIFF of 10-meter resolution representing the percent tree cover for each city block

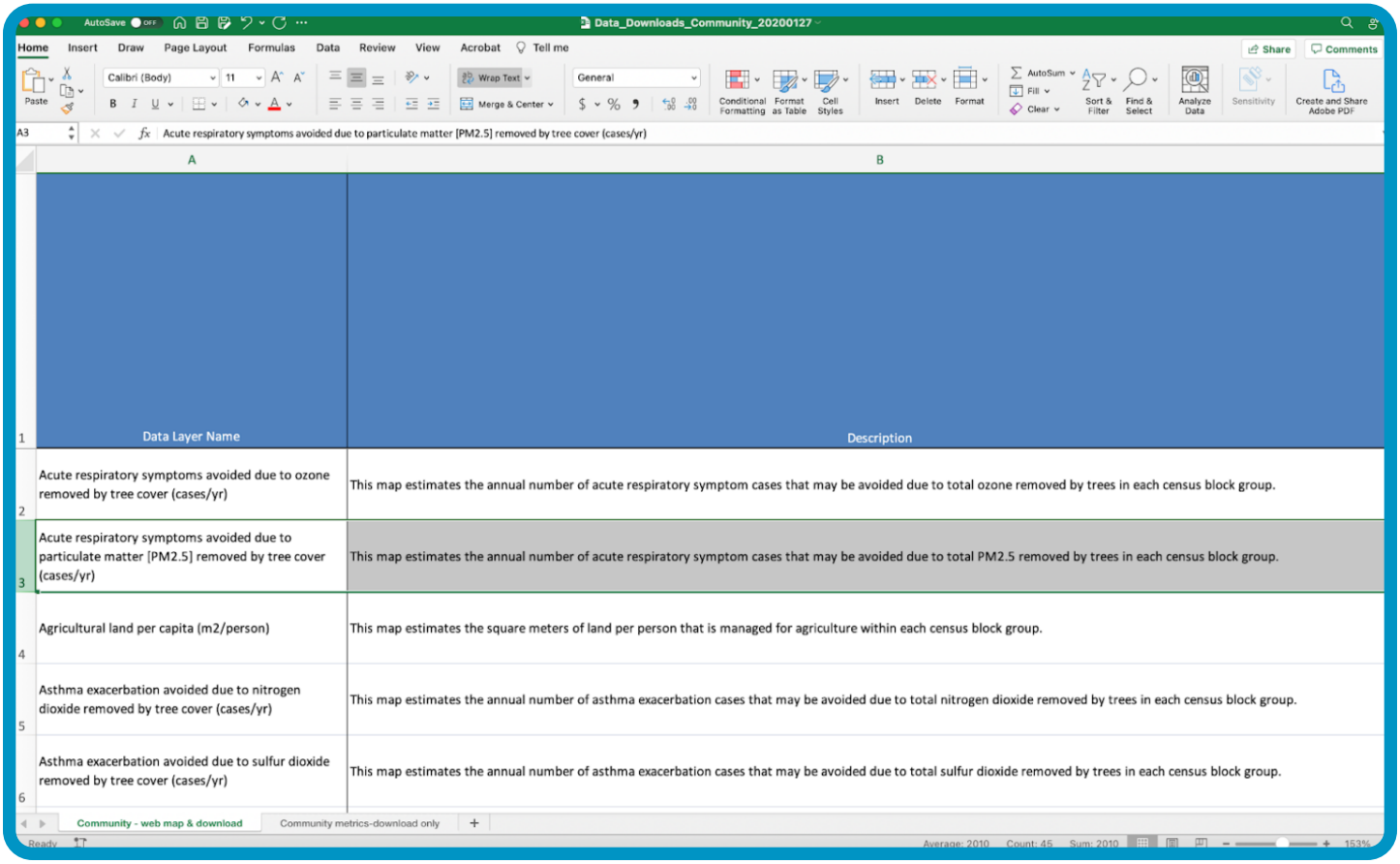
Show 10 entries Search:

Community	Metadata	Esri FileGDB	CSV (with SHP)	Landcover	Connectivity	Int Dens	PctStGS	PctStTC	Changelog
Los Angeles, CA	XML	[1.7 GB]	[913 MB]	[1.3 GB]	[582 MB]	[298 MB]	[18.2 MB]	[18.1 MB]	TXT
Memphis, TN	XML	[65 MB]	[325 MB]	[19 MB]	[131 MB]	[97 MB]	[4 MB]	[4 MB]	TXT
Milwaukee, WI	XML	[36 MB]	[218 MB]	[329 MB]	[128 MB]	[89 MB]	[4.5 MB]	[4.4 MB]	TXT
Minneapolis / St.Paul, MN	XML	[70 MB]	[420 MB]	[320 MB]	[183.5 MB]	[143 MB]	[9.2 MB]	[9.2 MB]	TXT
New Bedford, MA	XML	[5 MB]	[27 MB]	[44 MB]	[15 MB]	[12 MB]	[610 kb]	[610 kb]	TXT
New Haven, CT	XML	[26 MB]	[128 MB]	[108 MB]	[89.5 MB]	[59.0 MB]	[2.7 MB]	[2.8 MB]	TXT
New York, NY	XML	[23 MB]	[130 MB]	[83 MB]	[55 MB]	[41 MB]	[5 MB]	[5 MB]	TXT
Paterson, NJ	XML	[1 MB]	[4 MB]	[7 MB]	[1 MB]	[1 MB]	[152 kb]	[142 kb]	TXT
Philadelphia, PA	XML	[183 MB]	[829 MB]	[141 MB]	[451 MB]	[219 MB]	[24 MB]	[24 MB]	TXT
Phoenix, AZ	XML	[84 MB]	[480 MB]	[1102 MB]	[193 MB]	[191 MB]	[12 MB]	[12 MB]	TXT

Showing 11 to 20 of 30 entries Previous 1 2 3 Next

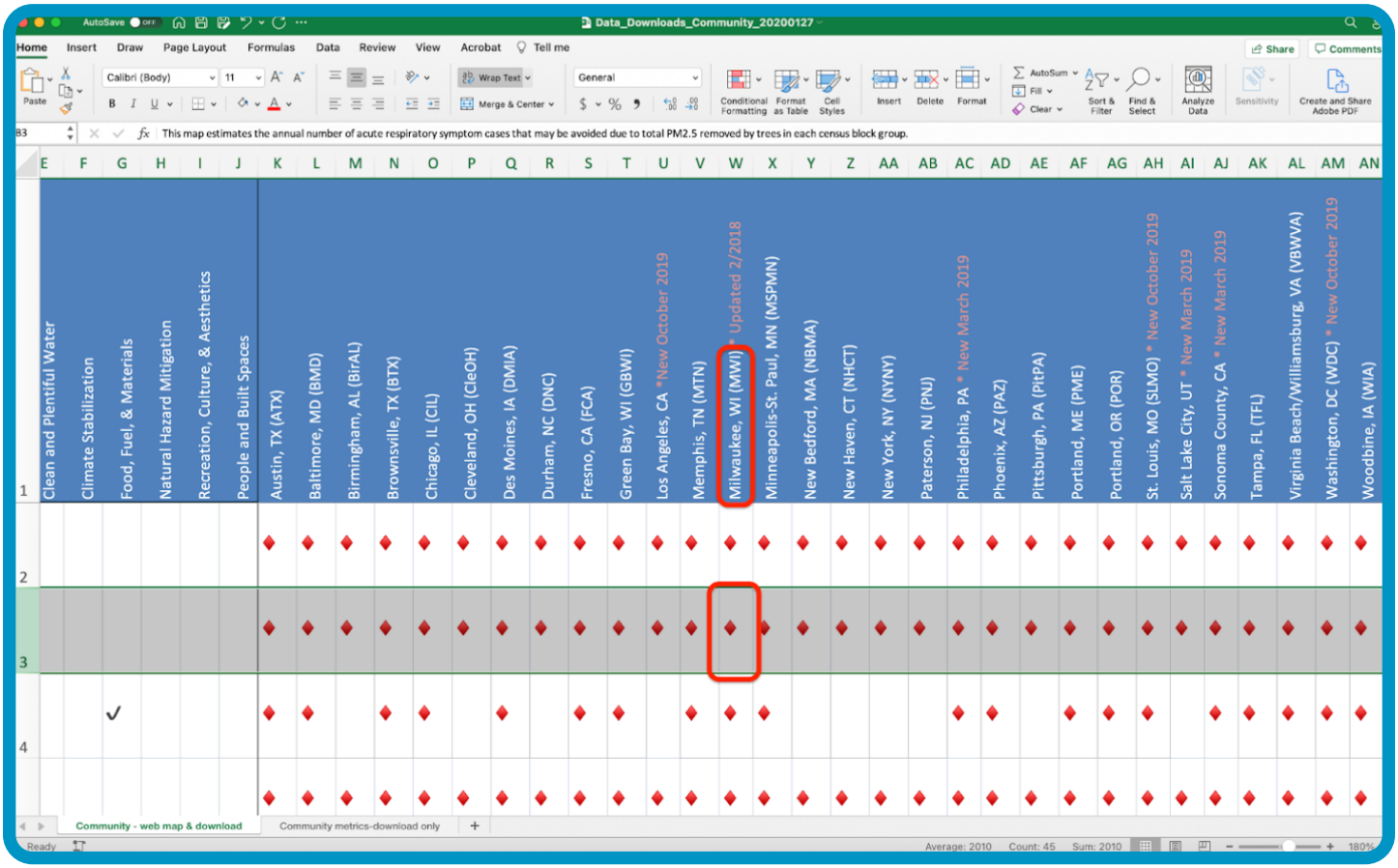
8. Next, open ***the comprehensive Community Data Excel spreadsheet*** that we downloaded in earlier in this section.

    a. Click on the tab "Community metrics- web map & download." Column A of the spreadsheet contains the names of each data layer available for download in EnviroAtlas. Under column A in the Excel spreadsheet, we can find the first data layer we want –Acute respiratory symptoms avoided due to particulate matter [PM2.5] removed by tree cover (cases/year) –in row three.



	Data Layer Name	Description
2	Acute respiratory symptoms avoided due to ozone removed by tree cover (cases/yr)	This map estimates the annual number of acute respiratory symptom cases that may be avoided due to total ozone removed by trees in each census block group.
3	Acute respiratory symptoms avoided due to particulate matter [PM2.5] removed by tree cover (cases/yr)	This map estimates the annual number of acute respiratory symptom cases that may be avoided due to total PM2.5 removed by trees in each census block group.
4	Agricultural land per capita (m2/person)	This map estimates the square meters of land per person that is managed for agriculture within each census block group.
5	Asthma exacerbation avoided due to nitrogen dioxide removed by tree cover (cases/yr)	This map estimates the annual number of asthma exacerbation cases that may be avoided due to total nitrogen dioxide removed by trees in each census block group.
6	Asthma exacerbation avoided due to sulfur dioxide removed by tree cover (cases/yr)	This map estimates the annual number of asthma exacerbation cases that may be avoided due to total sulfur dioxide removed by trees in each census block group.

b. Columns K through AN contain the names of the thirty community areas and their respective community abbreviation in parentheses. The presence of a diamond symbol in that column indicates that the data layer of interest is available for that community area. Locate "Milwaukee, Wisconsin (MWI)" in column W of the spreadsheet and check to see a diamond symbol.



c. Scroll over to column AO to identify which link we need to click on in ***the online Community Data table*** to access this data layer. It says we can download it from the "Esri FileGDB" or "CSV (with SHP)" columns of ***the online Community Data table***.

	AO	AP	AQ	AR	
1	Where to find on the data download page:	Table Name (Replace CBG with community abbreviation)	Field Name	Data type (Replace CBG with community abbreviation)	Fact Sh
2	Community Data: Esri FileGDB or CSV (with SHP)	CBG_BenMAP	O3_Acute_Respiratory_Symptoms_I	Table - joins with feature class CBG_BG	https://enviroatlas.epa.gov/en C/AcuteRespiratorySymptomsA reecover.pdf
3	Community Data: Esri FileGDB or CSV (with SHP)	CBG_BenMAP	PM25_Acute_Respiratory_Symptoms_I	Table - joins with feature class CBG_BG	https://enviroatlas.epa.gov/en C/AcuteRespiratorySymptomsA M25Removedbytreecover.pdf
	Community Data: Esri FileGDB or CSV (with	CBG_LCSum	Ag_PC	Table - joins with feature class	https://enviroatlas.epa.gov/en

9. Repeat the steps in number 8 for our second data layer of interest, Percent impervious area [census block group].

    a. We find it is located in row thirty-eight of the Excel spreadsheet key.

    b. Cell W38 also shows a diamond symbol affirming that it is available for Milwaukee.

    c. Column AO tells us that it will also be found under the same two downloadable sources "Esri FileGDB" and "CSV (with SHP)".

    d. We are now ready to download the appropriate dataset for Milwaukee.

10. Look at ***the online Community Data table***.

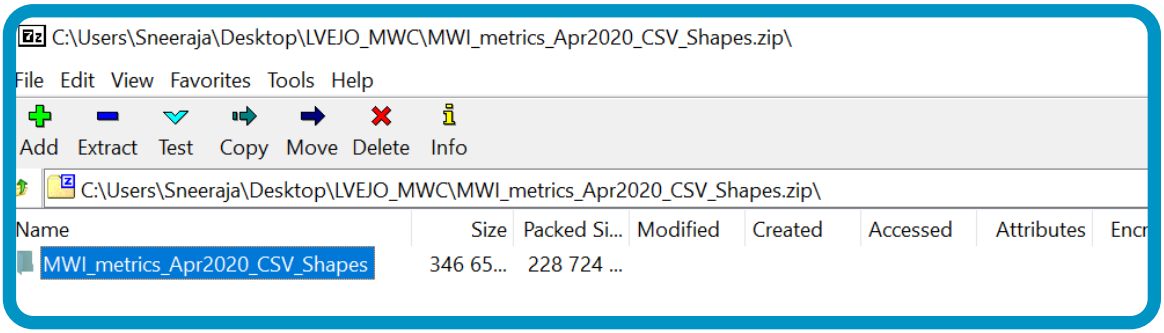
Show 10 entries Search:

Community	Metadata	Esri FileGDB	CSV (with SHP)	Landcover	Connectivity	Int Dens	PctStGS	PctStTC	Changelog
Los Angeles, CA	XML	[1.7 GB]	[913 MB]	[1.3 GB]	[582 MB]	[298 MB]	[18.2 MB]	[18.1 MB]	TXT
Memphis, TN	XML	[65 MB]	[325 MB]	[19 MB]	[131 MB]	[97 MB]	[4 MB]	[4 MB]	TXT
Milwaukee, WI	XML	[36 MB]	[218 MB]	[329 MB]	[128 MB]	[89 MB]	[4.5 MB]	[4.4 MB]	TXT
Minneapolis / St.Paul, MN	XML	[70 MB]	[420 MB]	[320 MB]	[183.5 MB]	[143 MB]	[9.2 MB]	[9.2 MB]	TXT
New Bedford, MA	XML	[5 MB]	[27 MB]	[44 MB]	[15 MB]	[12 MB]	[610 kb]	[610 kb]	TXT

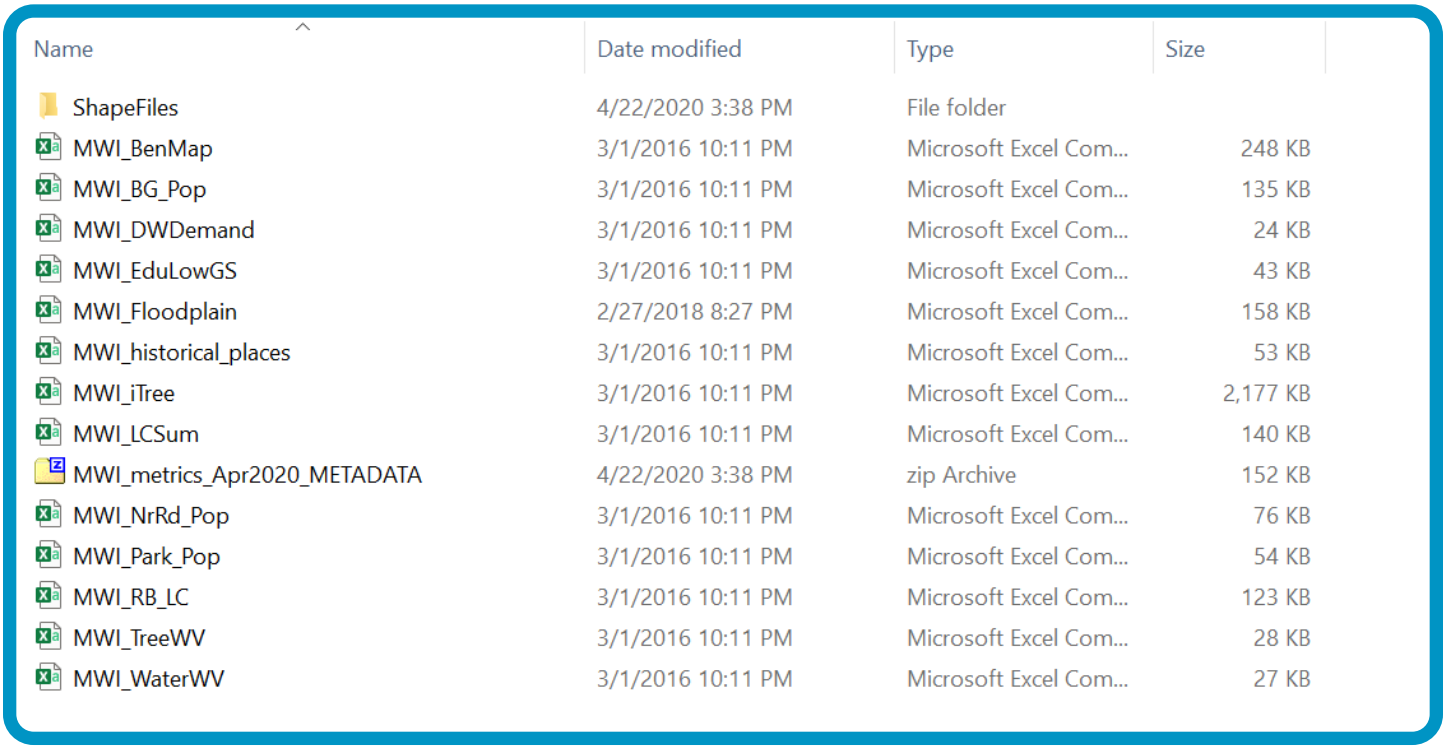
11. Click on the link under the "CSV (with SHP)" column in the Milwaukee, WI row to download the data and save to your project folder. We are not using the Esri FileGDB since this is a tutorial for GIS beginners. As you learn more about GIS, you'll learn about geodatabases (GDB) and may choose to use that in the future.

12. The downloaded file should be a zipped folder. Double click it to open, click the folder inside to select, and click the "Extract" button up top with a blue minus sign to unzip. On a Mac, placed the ZIP file where you want to store the data and simply double click the ZIP file and it will unzip. (This may take a bit of time to download.)

13. Choose the location you want for the unzipped folder - we suggest the project folder. Then click "OK".



14. After extraction, click open the new unzipped folder, called MWI_metrics folder throughout the rest of this tutorial.

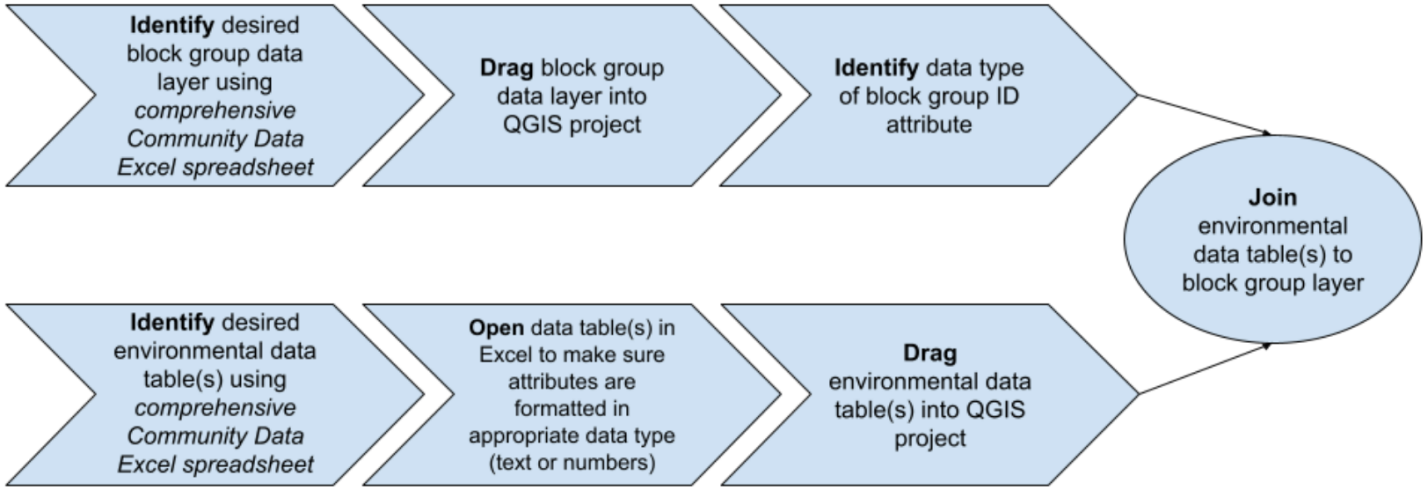


Name	Date modified	Type	Size
ShapeFiles	4/22/2020 3:38 PM	File folder	
MWI_BenMap	3/1/2016 10:11 PM	Microsoft Excel Com...	248 KB
MWI_BG_Pop	3/1/2016 10:11 PM	Microsoft Excel Com...	135 KB
MWI_DWDemand	3/1/2016 10:11 PM	Microsoft Excel Com...	24 KB
MWI_EduLowGS	3/1/2016 10:11 PM	Microsoft Excel Com...	43 KB
MWI_Floodplain	2/27/2018 8:27 PM	Microsoft Excel Com...	158 KB
MWI_historical_places	3/1/2016 10:11 PM	Microsoft Excel Com...	53 KB
MWI_iTree	3/1/2016 10:11 PM	Microsoft Excel Com...	2,177 KB
MWI_LCSum	3/1/2016 10:11 PM	Microsoft Excel Com...	140 KB
MWI_metrics_Apr2020_METADATA	4/22/2020 3:38 PM	zip Archive	152 KB
MWI_NrRd_Pop	3/1/2016 10:11 PM	Microsoft Excel Com...	76 KB
MWI_Park_Pop	3/1/2016 10:11 PM	Microsoft Excel Com...	54 KB
MWI_RB_LC	3/1/2016 10:11 PM	Microsoft Excel Com...	123 KB
MWI_TreeWV	3/1/2016 10:11 PM	Microsoft Excel Com...	28 KB
MWI_WaterWV	3/1/2016 10:11 PM	Microsoft Excel Com...	27 KB

15. Inside MWI_metrics folder is a series of CSVs (Type: Microsoft Excel Com... in the picture above), a folder called "ShapeFiles", and another zipped folder that contains metadata. The next section will use these various files to visualize the data layers of our choice for Milwaukee at the census block group level.
    - Our layers of interest:
        i. Percent impervious area
        ii. Acute respiratory symptoms avoided due to particulate matter [PM2.5] removed by tree cover (cases/year).

Visualizing EnviroAtlas Data

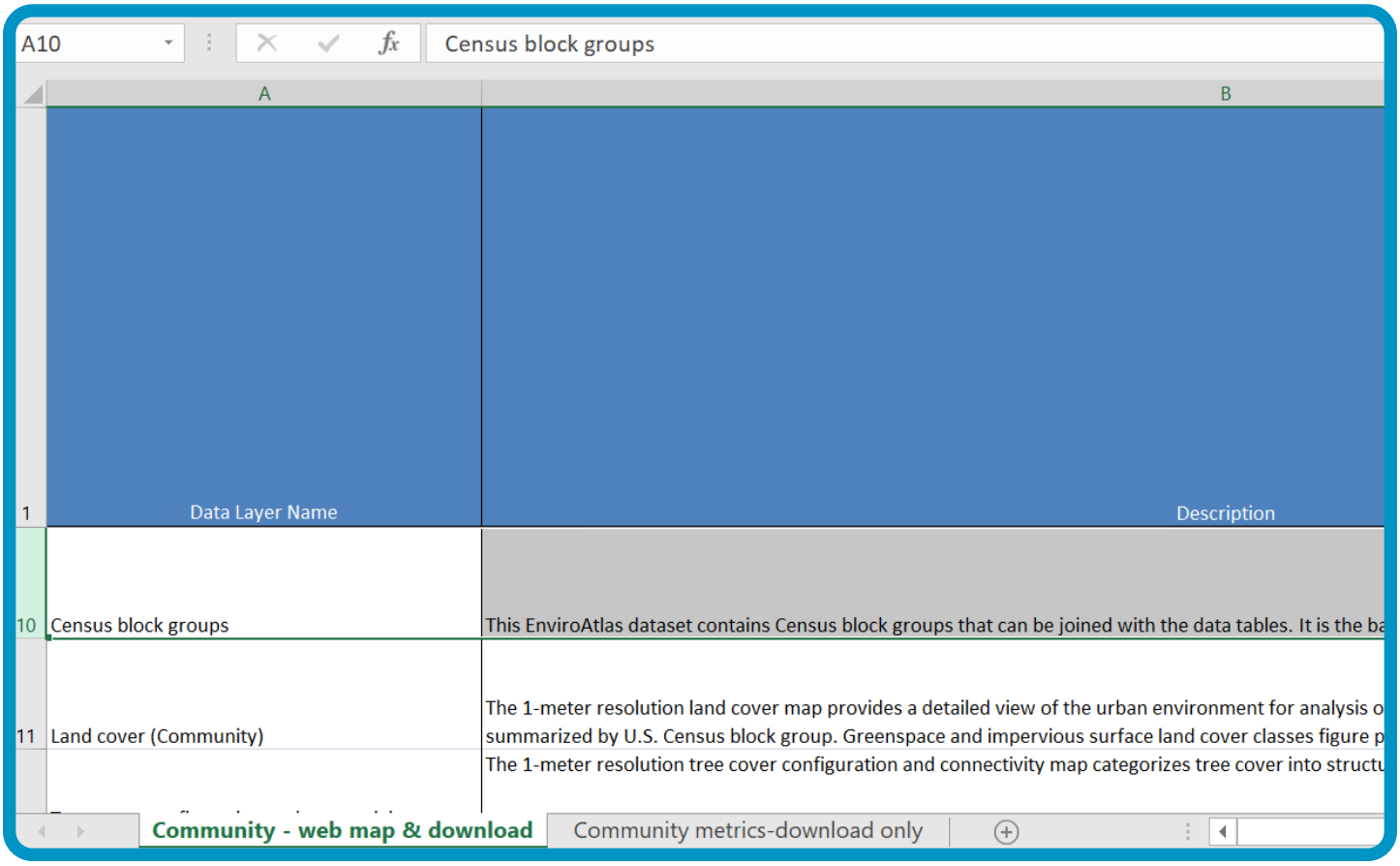
When we visualize the EnviroAtlas data, we will need to work with both Excel and QGIS to make sure that the data is prepared for use in QGIS. The following sections will include a label of (Excel, QGIS, Both) to identify what software will be used.



Find the Appropriate Layers from the Download (Both)

1. Open ***the comprehensive Community Data Excel spreadsheet*** and go to the "Community - web map & download" tab.

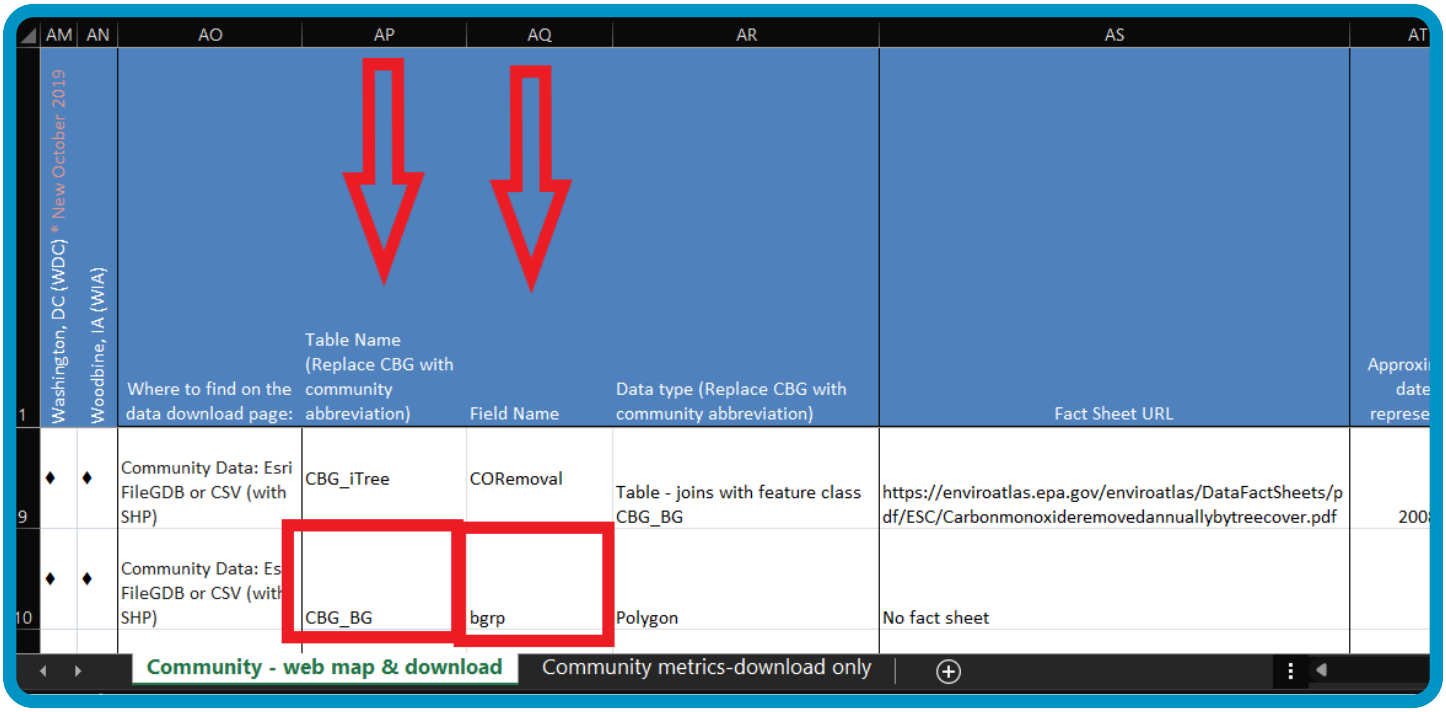
2. Find the data layer name "Census block groups". It is column A, row 10.



3. Scroll to columns AP and AQ in the spreadsheet.

   a. Column AP shows us the file name to find the data in the MWI_metrics folder.

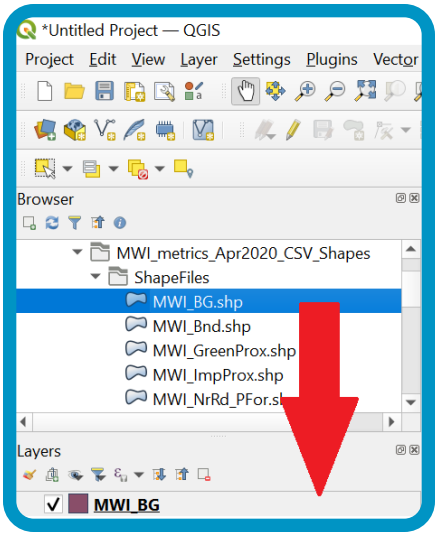
   b. Column AQ tells us the field name (attribute) inside the table to visualize this layer. For census block groups, bgrp refers to the ID of each individual census block group.



   c. If you're interested in knowing the data type take a look at Column AR. Cell AR10 says "Polygon". In section 5, we learned about GIS data formats.

Cheat sheet on how to interpret column AR:

Column AR Value	Format	How we know
Grid	Raster	Column B provides a description stating 1-meter resolution or pixel which are terms used to reference rasters. These layers will need to be downloaded directly from the online Community Data table.
Polygon	Vector (Shapefile)	Polygons are an example of vector formats.
Polyline	Vector (Shapefile)	Polylines - or lines - are an example of vector formats.
Table - joins with feature class CBG_BG	Table	Datatable that will need to join to a shapefile to visualize the data. Steps on how to "join" are provided later in this section.



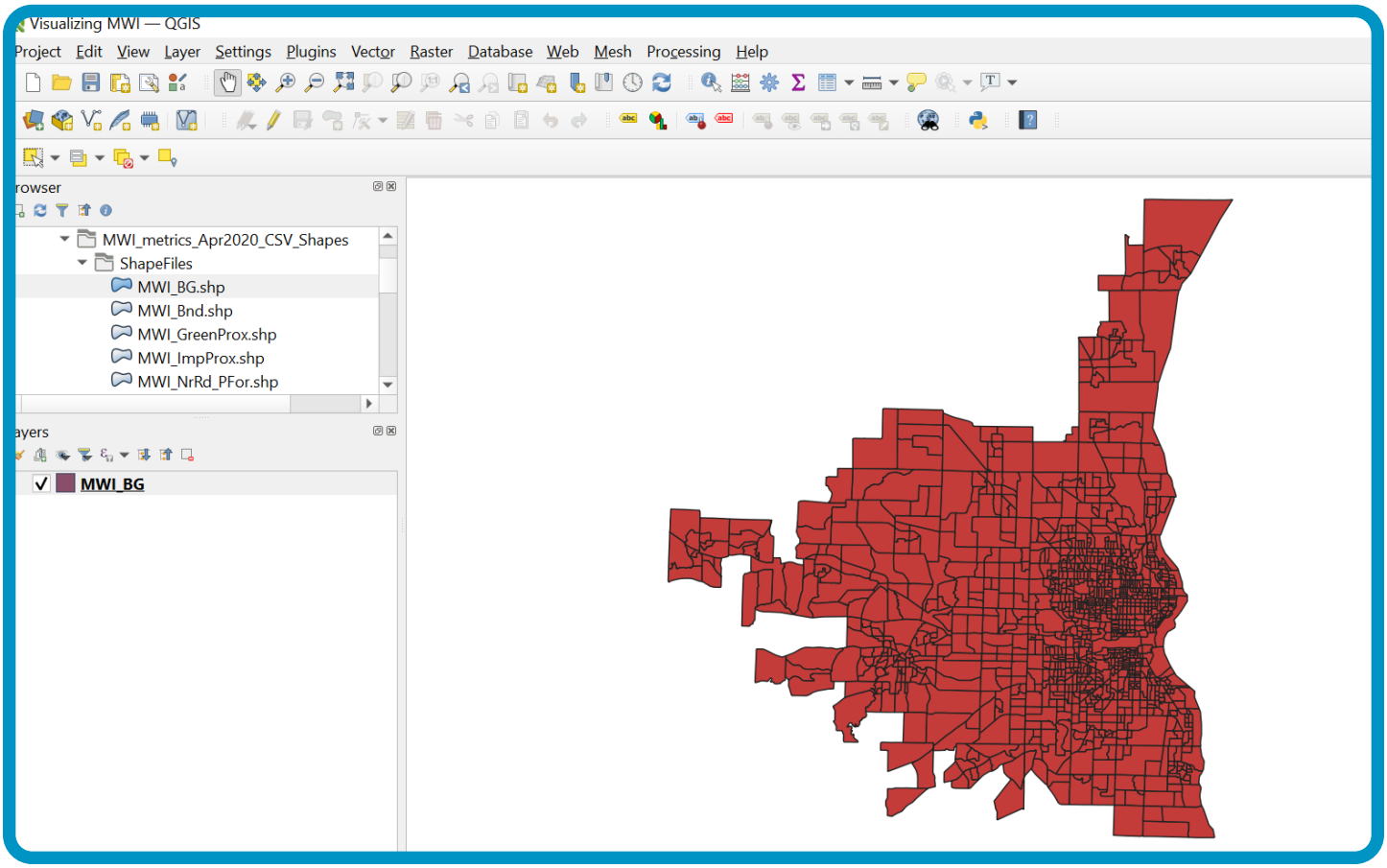
4. Open QGIS.

5. In the QGIS Browser navigate to your project folder. Open the MWI_metrics folder.

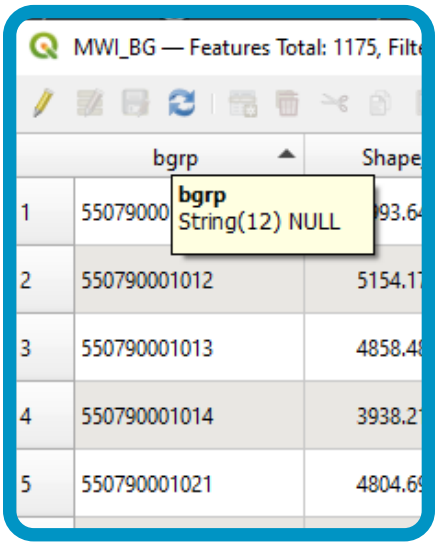
6. Open the "ShapeFiles" folder. Since cell AP10 said that our Table Name is "CBG_BG" and to "replace CBG with community abbreviation", we know we are looking for a shapefile called "MWI_BG".

7. Move the shapefile "MWI_BG.shp" by clicking and dragging the file from the Browser to the Layers section below it.

8. You should now see that all the census blocks come up in the large white area for the Milwaukee community area. Notice that this map consists of a different area than that of the city of Milwaukee layer from the Census. This is because the EnviroAtlas is using the subwatershed geography, which is different from how Census defines the Milwaukee "place" boundaries.

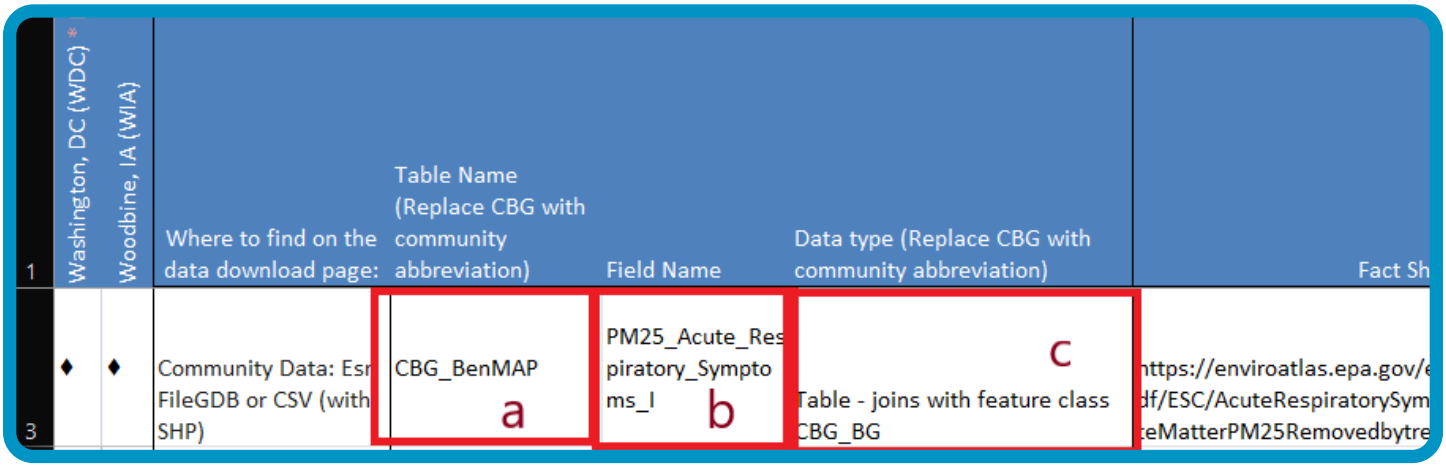


To display the environmental data associated with each of these block groups we will need to join the two data layers based on a similar attribute like block group ID. To successfully perform a join, we first need to determine what data type the attribute block group ID is.



Identifying Attribute Data Type:

1. Right-Click on the Layer "MWI_BG" and click "Open Attribute Table"
2. Identify the attribute or field name that shows the block group ID. (Remember ***the comprehensive Community Data Excel spreadsheet*** noted this). In this case it is "bgrp."
3. Hover over "bgrp" and you'll see a pop-up that says bgrp String(12) NULL. This means that the data type is a string - which means text (as opposed to a number).
    a. If you hover over "Shape_Length" the pop-up says: Shape_Length Real(18,11) NULL, which means that it is stored a real number with 18 digits (field length) and 11 digits after the decimal (precision).
4. Now that we know the block group is identified with a string, we'll know how we need to prepare the EnviroAtlas environmental data so that we can join the environmental data with the block group shapefile.
5. To locate the desired environmental data to join the census block groups, move back to ***the comprehensive Community Data Excel spreadsheet***.
6. Navigate to row three, which is where information for Acute respiratory symptoms avoided due to particulate matter [PM2.5] removed by tree cover (cases/year) is.
7. Scroll to columns AP and AQ –" Table Name (Replace CBG with community abbreviation)" and "Field Name".
    a. Cell AP3 names the table that contains our data layer: CBG_BenMAP.
    b. Cell AQ3 names the attribute within this table that represents the data layer: PM25_Acute_Respirtatory_Symptoms_I.
    c. Cell AR3 says "Table - joins with feature class CBG_BG".



Those cells tell us that to visualize acute respiratory symptoms avoided due to particulate matter [PM2.5] removed by tree cover (cases/year), we will join MWI_BenMAP.csv to the block group shapefile. We can access this file from the MWI_metrics folder.

8. Repeat step 7(a-c) after identifying Percent impervious area [census block group] in column A.

   To visualize percent impervious area in each block group we will join MWI_LCSum.csv to the block group shapefile. This file is also within the MWI_metrics folder.

1	Washington, DC (WDC)	Woodbine, IA (WIA)	Where to find on the data download page:	Table Name (Replace CBG with community abbreviation)	Field Name	Data type (Replace CBG with community abbreviation)	Fact Sh
38	♦	♦	Community Data: Esri FileGDB or CSV (with SHP)	CBG_LCSum a	Imp_P b	Table - joins with feature class CBG_BG c	https://enviroatlas.epa.gov/df/ESC/Percentimperviousar

Each environmental data has an accompanying fact sheet (link in column AS) that contains background information on the data collection and processing methods to create the data layer. This can help you choose which data layers to use and consider how the data may impact your analysis and conclusions. Column AT shows the year(s) that the EnviroAtlas data layer represents. In the case for our two data layers, the respiratory health data we are using is from 2010 (cell AT3) and the impervious area data are based on data from over five years, 2008-2013 (cell AT38).

	A	AR	AS	
1	Data Layer Name	Data type (Replace CBG with community abbreviation)	Fact Sheet URL	Appro da repre
2	Acute respiratory symptoms avoided due to ozone removed by tree cover (cases/yr)	Table - joins with feature class CBG_BG	https://enviroatlas.epa.gov/enviroatlas/DataFactSheets/pdf/ESC/AcuteRespiratorySymptomsAvoidedDueToOzoneRemovedbytreecover.pdf	
3	Acute respiratory symptoms avoided due to particulate matter [PM2.5] removed by tree cover (cases/yr)	Table - joins with feature class CBG_BG	https://enviroatlas.epa.gov/enviroatlas/DataFactSheets/pdf/ESC/AcuteRespiratorySymptomsAvoidedDueToParticulateMatterPM25Removedbytreecover.pdf	

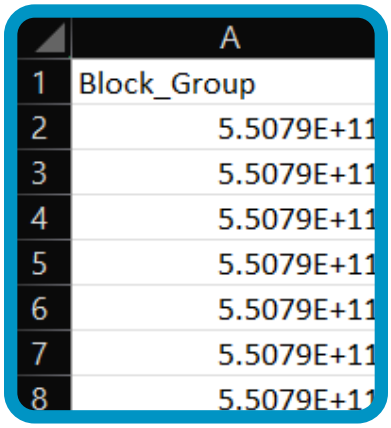
Cleaning EnviroAtlas Data for Use (Excel)

Since ***the comprehensive Community Data Excel spreadsheet*** has noted that the data of interest (PM 2.5 Acute Respiratory Symptoms avoided and Percent Impervious area) are in the tables, it is good practice to take a look at it in Excel before opening it in QGIS. This way, we can make sure that the attributes used to join the table and shapefile are the same data type.

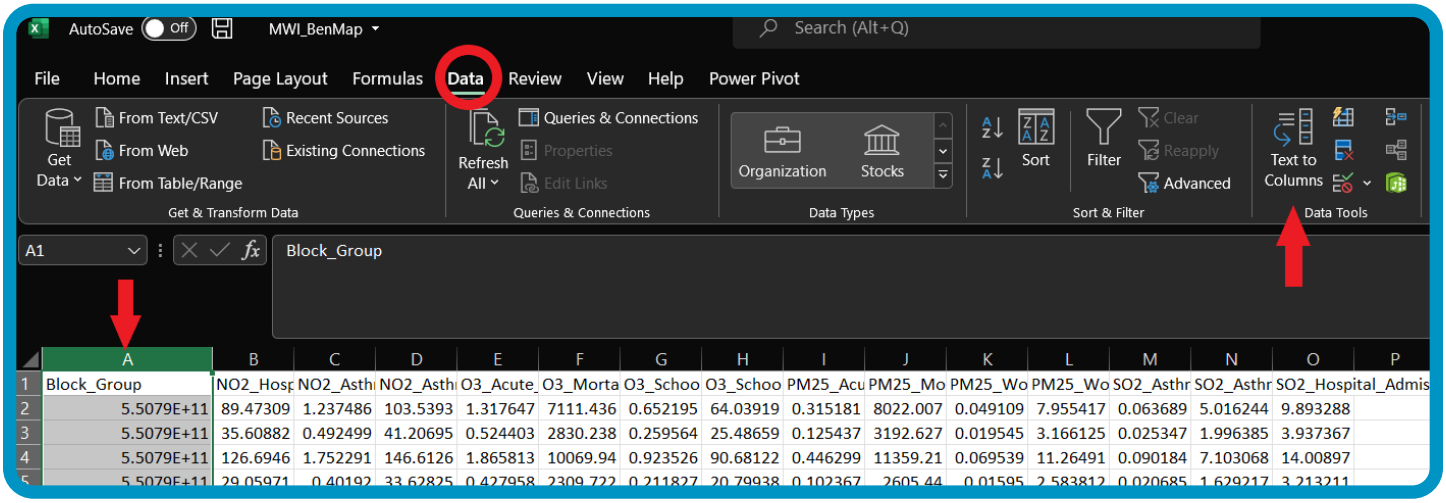
1. Open up the MWI_BenMAP.csv

2. Look at the column labeled Block_Group - this is the attribute we will use to join this table to the block group shapefile.

   Hint: One way to tell if the data type is a number or string/text is widening the column and see if the values in the column are on the left side or right side of the column. If the values are on the left, Excel is reading them as string/text, if the values are on the right, Excel is reading them as numbers.



	A
1	Block_Group
2	5.5079E+11
3	5.5079E+11
4	5.5079E+11
5	5.5079E+11
6	5.5079E+11
7	5.5079E+11
8	5.5079E+11



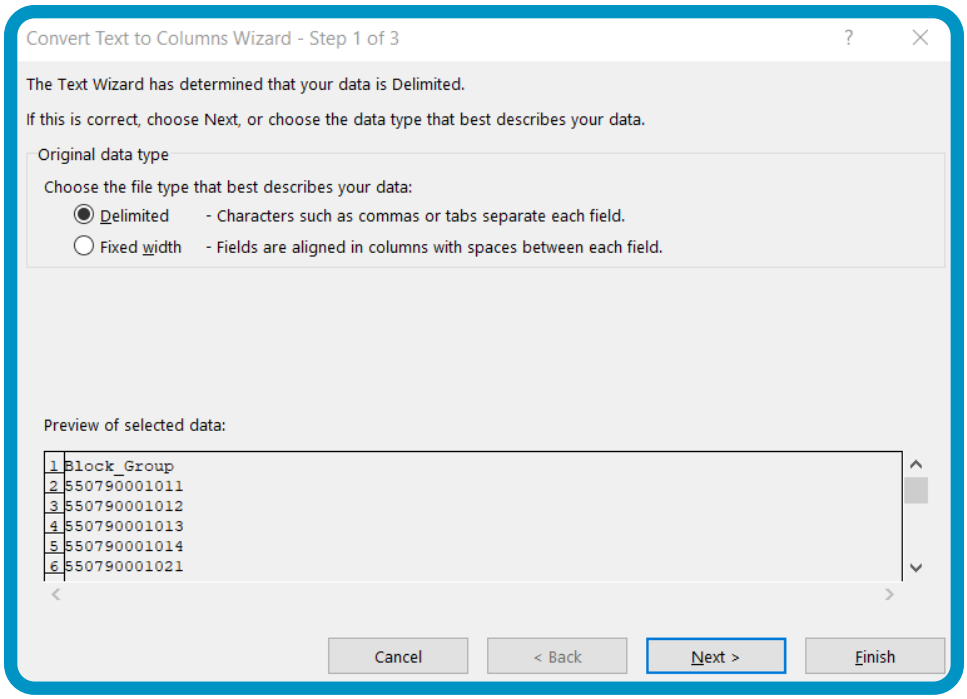
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Block_Group	NO2_Hosp	NO2_Asth	NO2_Asth	O3_Acute_	O3_Morta	O3_Schoo	O3_Schoo	PM25_Acu	PM25_Mo	PM25_Wo	PM25_Wo	SO2_Asthr	SO2_Asthr	SO2_Hospital_Admis	
2	5.5079E+11	89.47309	1.237486	103.5393	1.317647	7111.436	0.652195	64.03919	0.315181	8022.007	0.049109	7.955417	0.063689	5.016244	9.893288	
3	5.5079E+11	35.60882	0.492499	41.20695	0.524403	2830.238	0.259564	25.48659	0.125437	3192.627	0.019545	3.166125	0.025347	1.996385	3.937367	
4	5.5079E+11	126.6946	1.752291	146.6126	1.865813	10069.94	0.923526	90.68122	0.446299	11359.21	0.069539	11.26491	0.090184	7.103068	14.00897	
5	5.5079E+11	29.05971	0.40192	33.62825	0.427958	2309.722	0.211827	20.79938	0.102367	2605.44	0.01595	2.583812	0.020685	1.629217	3.213211	

3. Whether or not you can tell what data type it is, you should always take the following steps.

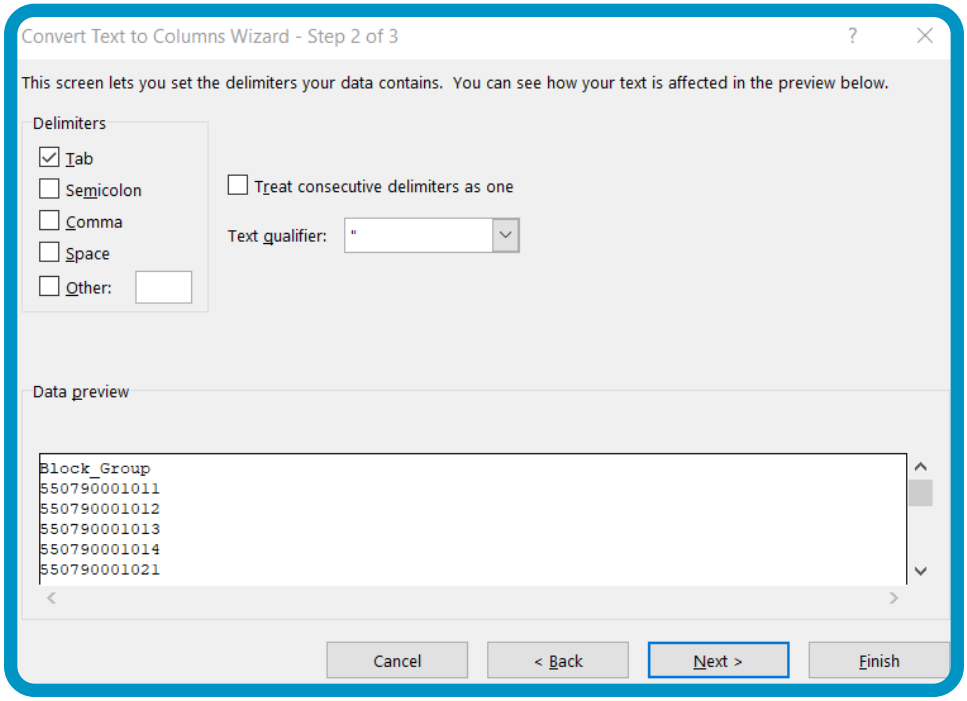
4. Highlight the column of interest by clicking the column heading, in this case Column A.

5. Go to the "Data" toolbar and click on "Text to Columns"

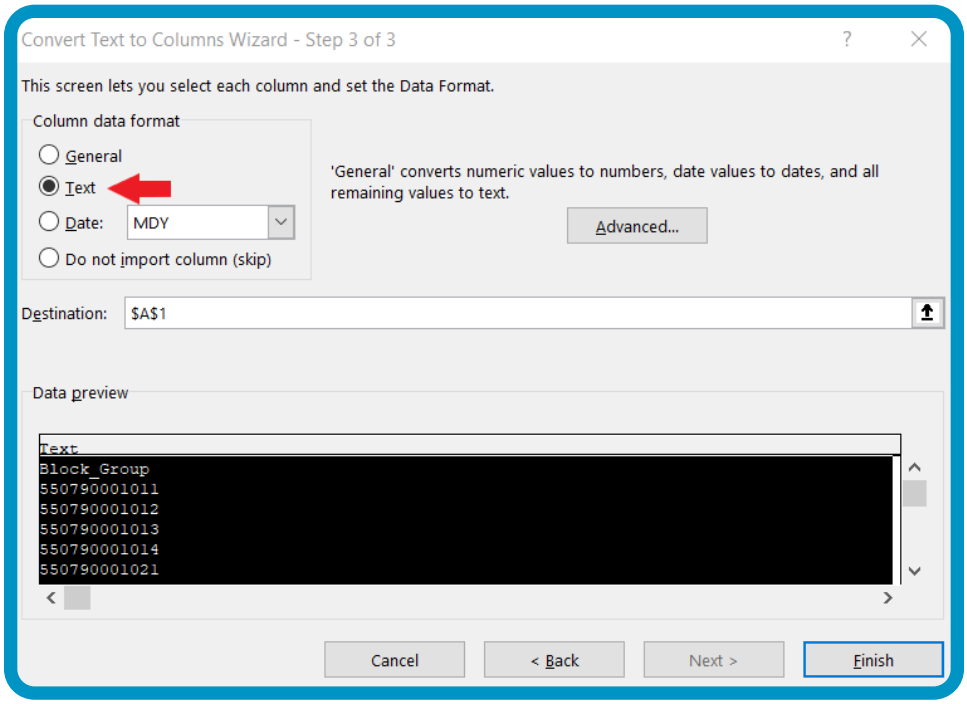
6. In Step 1 of the pop-up window just leave as is and click on "Next"



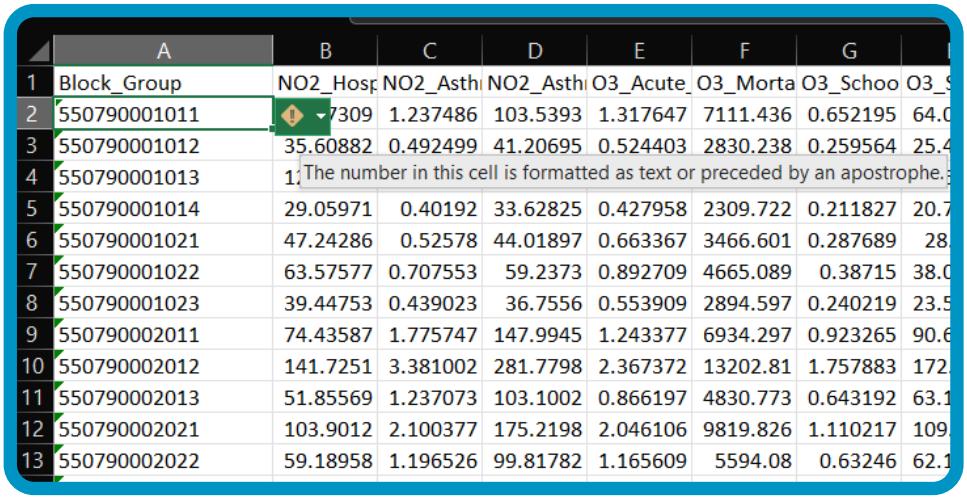
7. In Step 2 of the pop-up window just leave as is and click on "Next"



8. In Step 3 of the pop-up window, it says Column data format. This is where we can change the data type of the values in the column. Since we want the Block_Group to be a String data type to match with the shapefile, we'll click Text. The pop-up window notes that "'General' converts numeric values to numbers, date values to dates, and all remaining values to text". In the Destination section, leave it as is, because we want to replace these values saved as numbers with these values saved as text. Then click Finish.



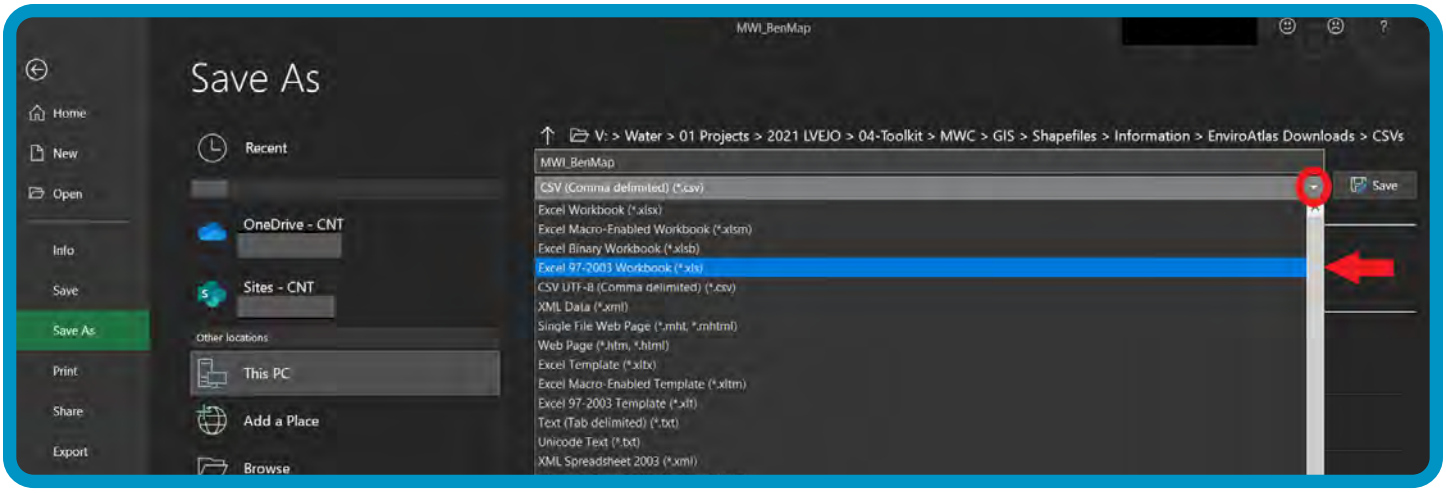
9. You'll notice that immediately, Column A values move over to the left and have green corner flags. If you click on a cell, a warning will pop-up telling you that the number is being saved as text. Excel will typically show this if numbers aren't being saved as numbers.



	A	B	C	D	E	F	G
1	Block_Group	NO2_Hosp	NO2_Asth	NO2_Asth	O3_Acute_	O3_Morta	O3_Schoo
2	550790001011	7309	1.237486	103.5393	1.317647	7111.436	0.652195
3	550790001012	35.60882	0.492499	41.20695	0.524403	2830.238	0.259564
4	550790001013	1					
5	550790001014	29.05971	0.40192	33.62825	0.427958	2309.722	0.211827
6	550790001021	47.24286	0.52578	44.01897	0.663367	3466.601	0.287689
7	550790001022	63.57577	0.707553	59.2373	0.892709	4665.089	0.38715
8	550790001023	39.44753	0.439023	36.7556	0.553909	2894.597	0.240219
9	550790002011	74.43587	1.775747	147.9945	1.243377	6934.297	0.923265
10	550790002012	141.7251	3.381002	281.7798	2.367372	13202.81	1.757883
11	550790002013	51.85569	1.237073	103.1002	0.866197	4830.773	0.643192
12	550790002021	103.9012	2.100377	175.2198	2.046106	9819.826	1.110217
13	550790002022	59.18958	1.196526	99.81782	1.165609	5594.08	0.63246

10. We also want to make sure that the numbers for the environmental data we are interested in are saved as numbers. Based on ***the comprehensive Community Data Excel spreadsheet***, the column we are interested in is "PM25_Acute_Respiratory_Symptoms_I". When looking at this column, we don't see any green flags, but to make sure that it is saved as a number, we'll highlight the appropriate column and follow the same steps for the Text to Columns function except instead of clicking "Text" in the 3rd step of the pop-up window, we'll click on "General" so that the numeric values are saved as numbers.

11. Now we want to save this worksheet in an XLS format. If we keep it in CSV, all of the work we just did may not be saved, because CSV is a simpler format than XLS. Go to File >>Save As and find the appropriate folder to save it in. Then before clicking the Save button, make sure to click on the drop down arrow by "CSV (Comma delimited) (*.csv)" and choose "Excel 97-2003 Workbook (*.xls)" and then click Save.



12. Repeat all of these same steps to prepare the environmental data "percent impervious area" for QGIS. Refer back to ***the comprehensive Community Data Excel spreadsheet*** which informs us that this layer is in the MWI_LCSum.csv and under the attribute "Imp_P". Note, that depending on each file, the block group attribute might be in a different column and might be labeled differently. In MWI_LCsum.csv the block group attribute is labeled bgrp.

    a. Open up MWI_LCSum.csv

    b. Change the data type for bgrp (block groups) to text

        i. Highlight the column "bgrp"

        ii. Go to the "Data" toolbar and click on "Text to Columns"

        iii. Choose "Delimited" and click Next

        iv. Under Delimiters, check "Tab" and click Next

        v. Under Column Data Format, choose "Text" and click Finish.

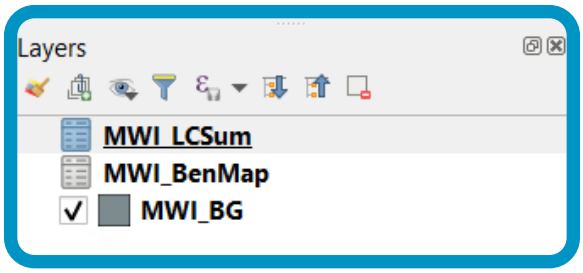
    c. Change the data type for "percent impervious area" to numeric

        i. Highlight the column "Imp_P"

        ii. Repeat the steps from above, but under Column Data Format, choose "General" so that the numeric values are saved as numbers.

Visualizing Cleaned EnviroAtlas Data (QGIS)

1. Move back to your QGIS file. Locate MWI_BenMAP.xls and MWI_LCSum.xls in the Browser within the unzipped community data download folder. Drag these to the Layers section of the window with the block group shapefile.



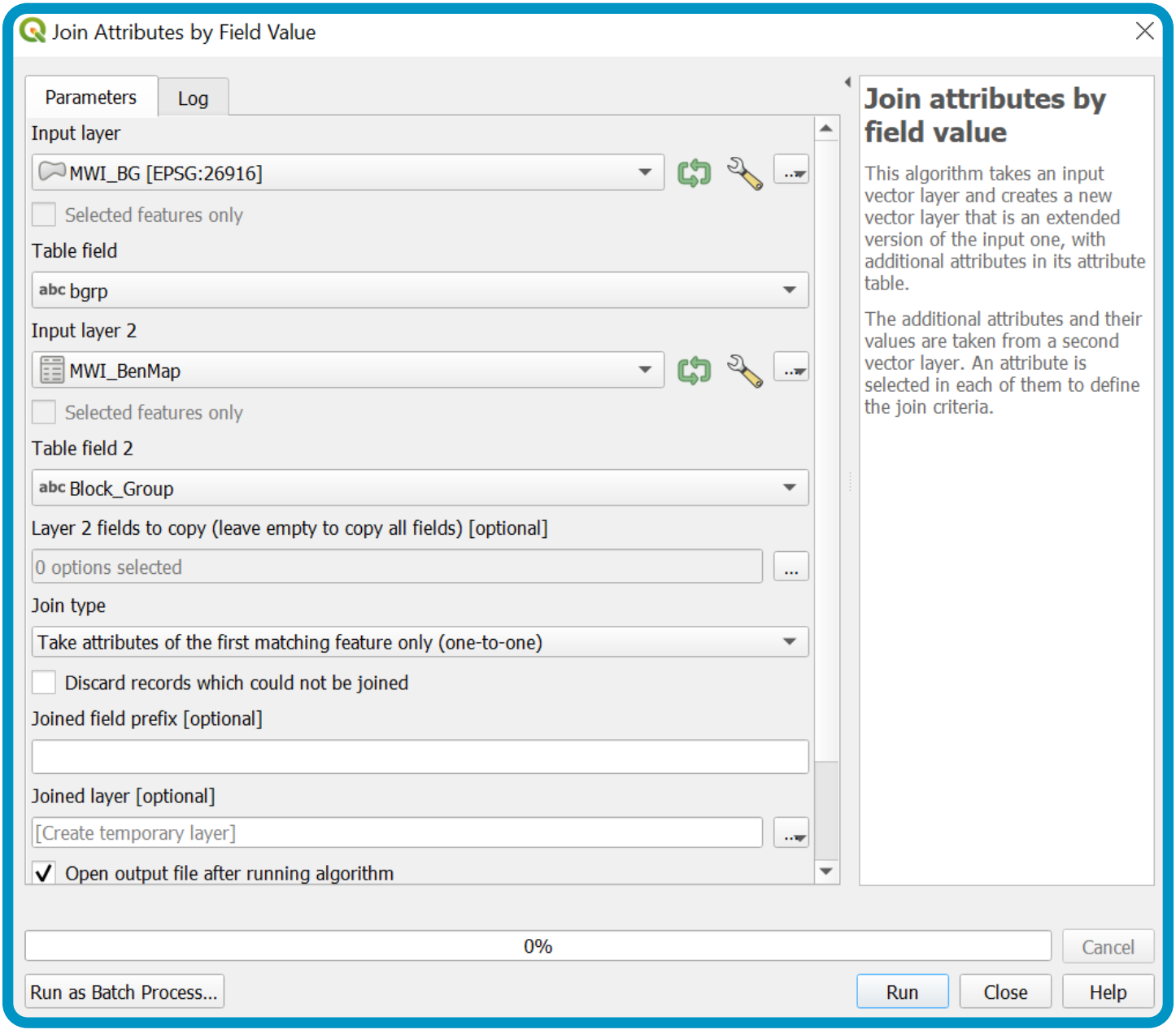
2. Right click on the layers and open the attribute table to explore what attributes are within each of these layers. Go ahead and locate the attribute in each of the three layers that we are interested in: bgrp, PM25_Acute_Respiratory_Symptoms_I, and Imp_P.

Join Attributes by Field Value

1. Next, in the bottom search bar search "Join Attributes by Field Value" and open the tool's window.

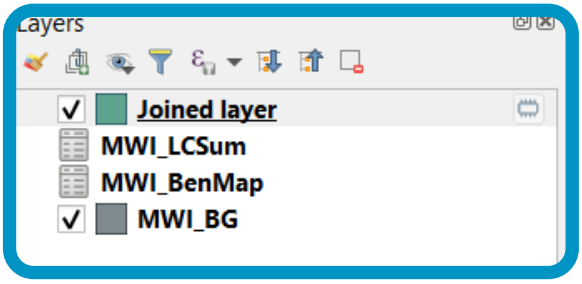
2. Select the MWI_BG.shp as the "Input layer" from the drop down menu and select bgrp from the "Table field". This is the field that contains block group IDs.

3. Choose MWI_BenMap for "Input layer 2" and Block_Group for "Table field 2". This is the name of the attribute that contains block group IDs in this table. Notice how next to "bgrp" and "Block_Group" there is an abc symbol, this tells you that both are in string format.



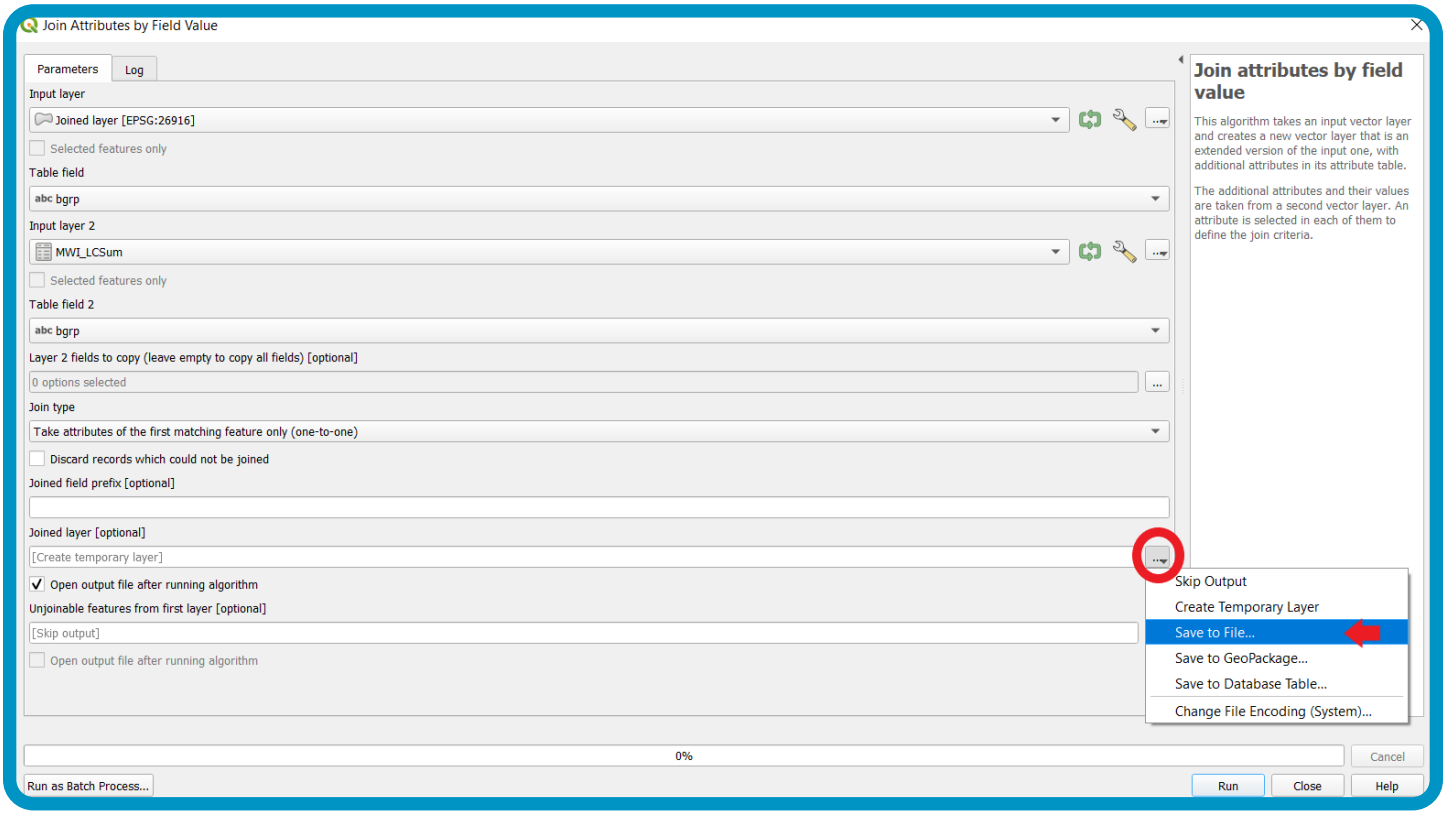
4. Click "Run". A new Joined Layer should now have appeared in the Layers window and an additional map is visualized (on top of the original) in the shade that corresponds with Joined Layer.

5. Join the MWI_LCSum table to the new joined layer. Open the Join Attributes by Field Value tool and choose Joined layer for "Input layer" and MWI_LCSum for "Input layer 2".



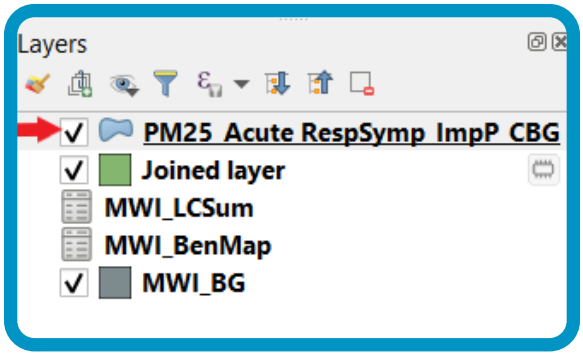
6. Choose the appropriate fields that contain block group IDs for both. There will be two for the "Joined Layer" one that is bgrp from the MWI_BG layer and one that is "Block_Group" from the MWI_BenMap" layer. For consistency, choose the "bgrp" attribute.

7. This time, in the "Joined layer [optional]" click on the downward arrow and choose "Save to File..." and save it in the appropriate folder with a name that makes sense to you. Then click Save.



8. Click "Run". You should see another layer (however you named it) appear in the Layers section and an additional map visualized.

9. If the layer does not appear here, make sure that you are saving the file as a SHP (.shp) file. If the layer appears in the Layers tab but not in the mapping window, right click the layer and click "Zoom to Layer(s)." If this happens, you will want to make sure you are using the appropriate projection.



10. Right click this layer and click Open Attribute Table. Scroll through this table and you should be able to find columns with both the names of the fields we are interested in. In chapter 6, we'll return to this shapefile to analyze it.

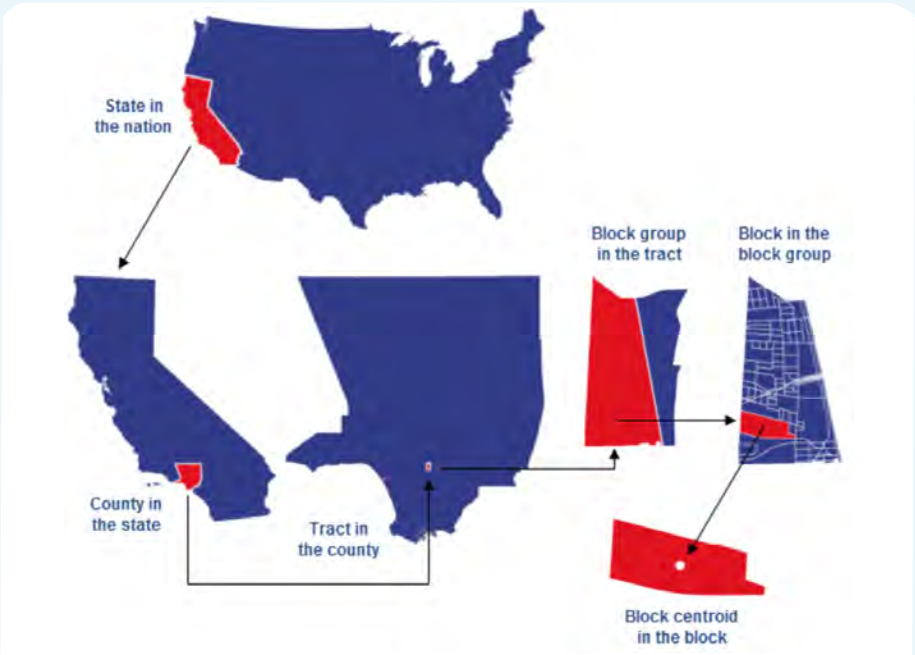
If you are interested in other potential data sources for your topical issue, visit section 5.G.

Remember: It is important that you frequently Save (Ctrl+S on PC or Command+S on Mac) as you go to avoid losing your work.

E. Intro to ACS

Demographic data can be added to a map to show the connections between population information and locations, project areas, or environmental justice issues. Examples of demographic data include age, gender, race, income-level, education level, or employment. Take a look at Chapter 2 to see examples of how demographics are included in maps.

U.S. demographic data are available through the United States Census Bureau. There is the decennial census which is updated every 10 years. The census refers to a count of all the people in the location. There is also the American Community Survey (ACS) which is released every year with 1-year and 5-year estimates. The ACS asks questions related to demographics, education, housing, jobs, and other information to 2-3% of the population every year to represent the entire population. The data collected from these surveys is then analyzed. Though it is meant to represent the entire population, like the Census, it may inaccurately account for certain populations like those who are unhoused, those who are undocumented, and others. Data are updated annually and is typically released in fall the following year. For example, if you want 2020 ACS data, that is released Fall 2021.



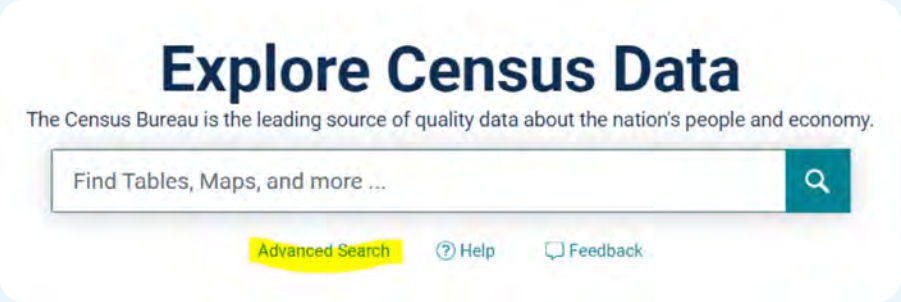
Source: ESRI. (2020). United States census geography—Related Concepts | Documentation. Learn ArcGIS. https://learn.arcgis.com/en/related-concepts/united-states-census-geography.htm

The ACS provides data at various geographies:

- **state**
- **county**
- **census tracts** - sized to include roughly 4,000 people, a combination of the block groups with it, never crosses county lines
- **block groups** - usually sized to include 250 - 500 people, is a combination of the blocks within it, smallest geographic size with published census data
- **blocks** - smallest geographic area that the data of individual persons are combined into. sometimes public data in not available for this size because of privacy concerns

Accessing ACS Data

The American Community Survey which collects demographic data from a sample of households each year is a great source to download data to be used in GIS. ACS data can be accessed from https://data.census.gov/cedsci/. When you get to this website, you'll want to click "Advanced search" to get a feel for the wide variety of information you can find.



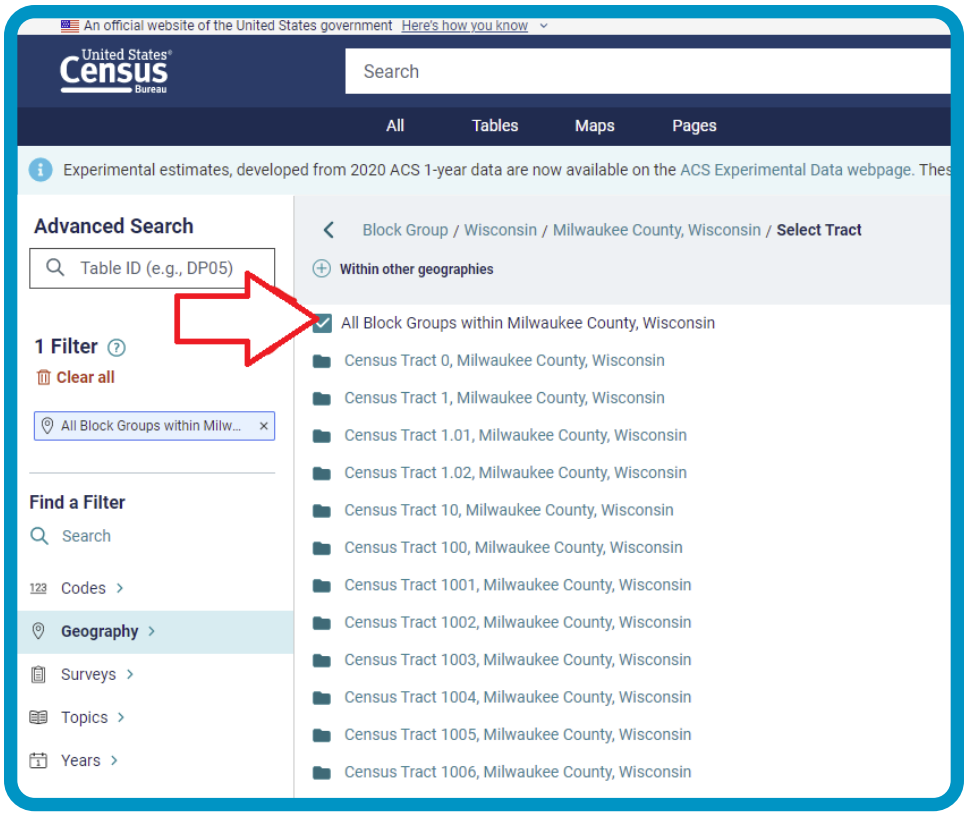
Note, this website has access to more data than just the ACS. Though any of these data are pretty good estimates, none of them are 100% accurate because filling out any of these different surveys is dependent on government outreach to people and people's comfort filling the survey. For example, it will undercount those who are experiencing homelessness, those who are wary of the government due to their documentation status, those who have less experience with reading, or those whose primary language was not available.

In advanced search, you can add filters to help find the data you are looking for. This is really helpful to find the specific geography your location is in and the years you are looking for. If using this method, don't click search until after you made all of the selections on the filters you want. The website automatically saves the selected filters. If you know the table number of the data you are looking for, you can type that straight into the Search box.

Geography

The website provides data at many different geographical levels depending on the topic and survey the data is coming from. For ACS, data is available at Nation, State, County, Tract, Block Group, and Block level. Typically across all ACS topics, the smallest geographical level available is either Tract or Block Group. Note: The ACS boundaries do not align to municipal boundaries and google searches may not return results with a list of census tracts or block groups with a municipality. You will need to take the steps listed in "Determine Your Geography" if this is your first time determining which census tracts or municipalities.

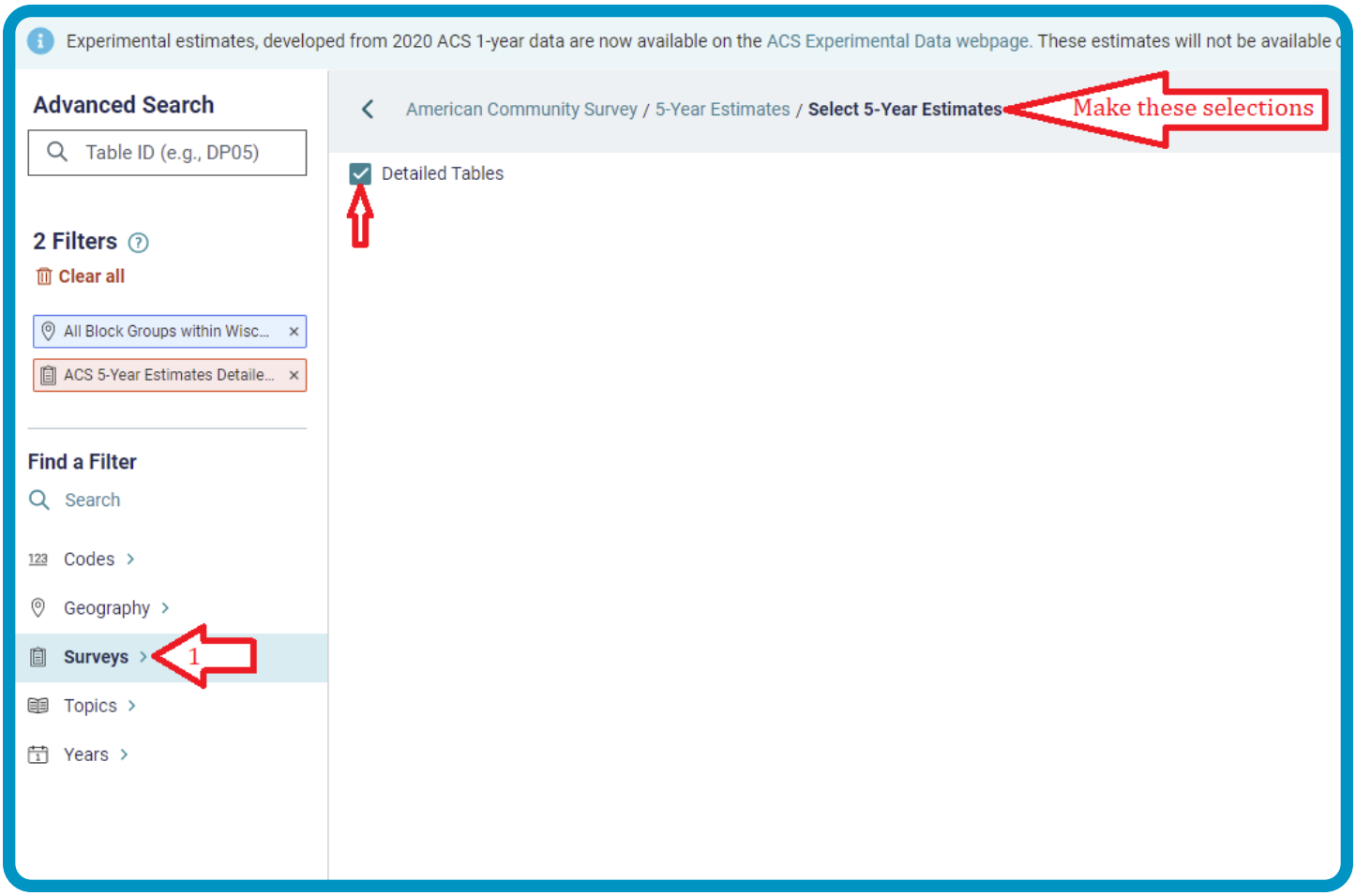
1. Under "Narrow search with filters", click on geography.
2. Click on Block Group
3. Click on the state of interest
4. Click on the county the municipality of interest is in
5. At the column that says "within tract", click on "All Block Groups within [selected county],[state]". This will be more data than you need, but that is okay because when we join (combine/connect) the ACS data to the shapefile you created in "Export Selected Features as Shapefile", that will get cleaned up.
6. You'll see that All Block Groups within the County are now a Selected Filter under Selected Filters on the bottom



Note: The U.S. Census may update its website by the time you view this tutorial so it could look different than what it does below.

Surveys

1. In the left hand column, click on "Surveys." Then, click on the American Community Survey.
2. Choose which year estimate you'd like to use.
    - 5-year estimates have data for all areas and are most reliable, but it is the least current data. We recommend using this data.
    - 1-year estimates are available for areas with a population over 65,000 and the most current data, but it is less reliable. (Guidance on Estimates).
    - Sometimes 1-year estimates will be grayed out or not available because the block groups you requested would not have that data available
3. When you click on 5-year estimates, you can then click on Detailed Tables



Topics

1. Click on Topics and take a look at what is available.
    - Consider what is your definition of an EJ community - what factors need to be measured to identify an EJ community.
    - Some typical data to gather from the ACS:

Demographics		
Race	Age	Sex
U.S. Born or Immigrant (Native and Foreign Born)	**Marital Status**	**Disability Status** (as government programs like SNAP might define it)
Employment	**Educational Attainment**	**Income**

Housing Related		
Household type (how many are living in a household with various family members)	**Living arrangements / Household size**	**Owner vs renter occupied** (Tenure)
Rent Cost	**Age of housing along with other characteristics**	**Vacancy**

Other		
Health Insurance Coverage	**Living arrangements / Household size**	**Commuting** (journey to work by mode, time taken)

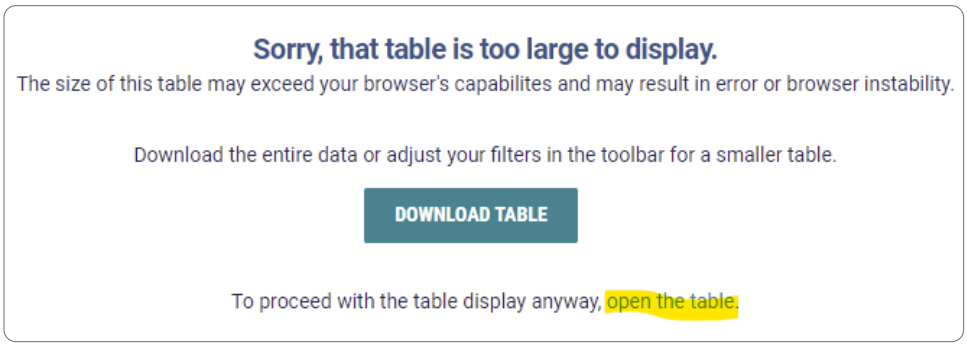
Note: If you choose multiple categories (within or across topics), the table searches will show only those that apply to all of those table categories, so choosing just one category will provide the broadest search.

Years (optional)

1. Click on Years and choose which year you would like your data from. You can only choose to download data one year at a time.

Search

1. At any point you can click Search in the lower right hand corner and take a look at the tables available and what data is available.
    - Some tables might not have information in them for every block group
    - Changing from block group to census tract as your geographic level may increase the amount of data available to you
2. Click on a Table you are interested in to see the data
3. Since there is more data you might not be able to view data within the webpage. You can try by clicking: open the data.



Note: If that makes your webpage crash, then download the data to view or filter to specific census tracts/block groups.

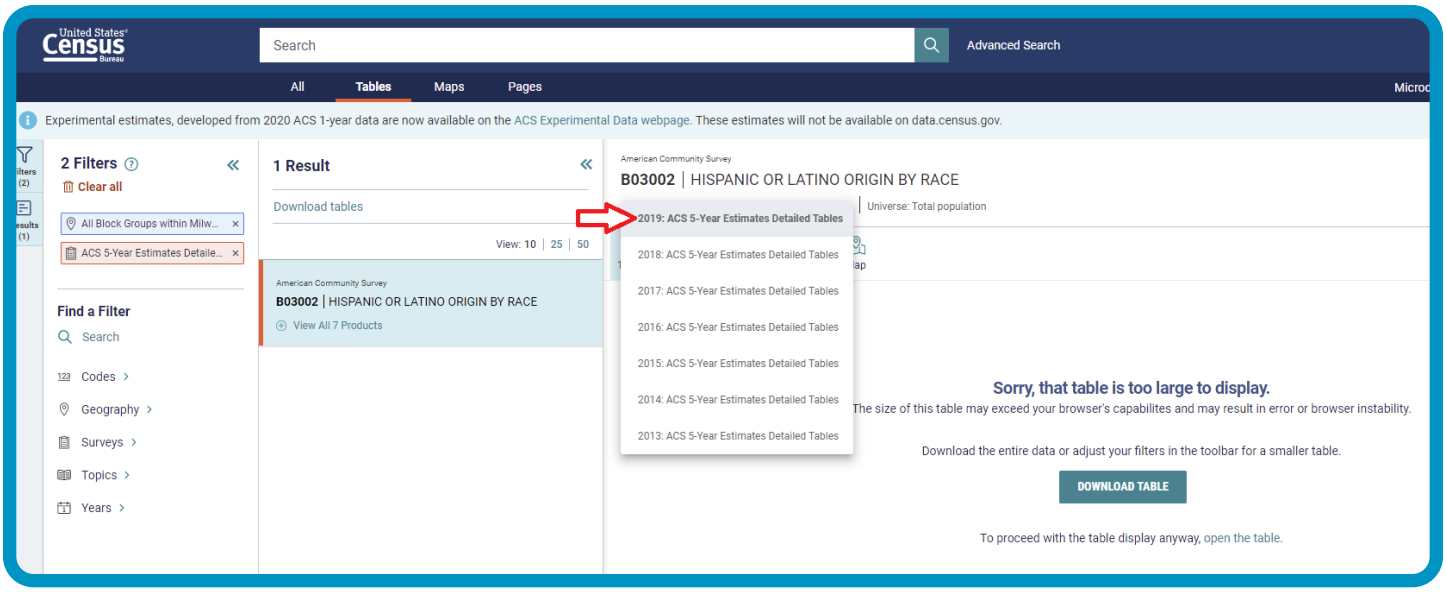
4. If you want to specify specific census tracts you can. You can use the CSV you had exported earlier and look at the "TRACTCE" column. It might help to look at the CSV with the tracts sorted in ascending order. Then click on the desired census tract within the data.census.gov website and click on "All Block Groups within census tract". Again, it's okay if it includes block groups not in the municipality, because it will get cleaned up later when you start joining the data.
5. You can change the year of the data by changing the product

Putting it All Together

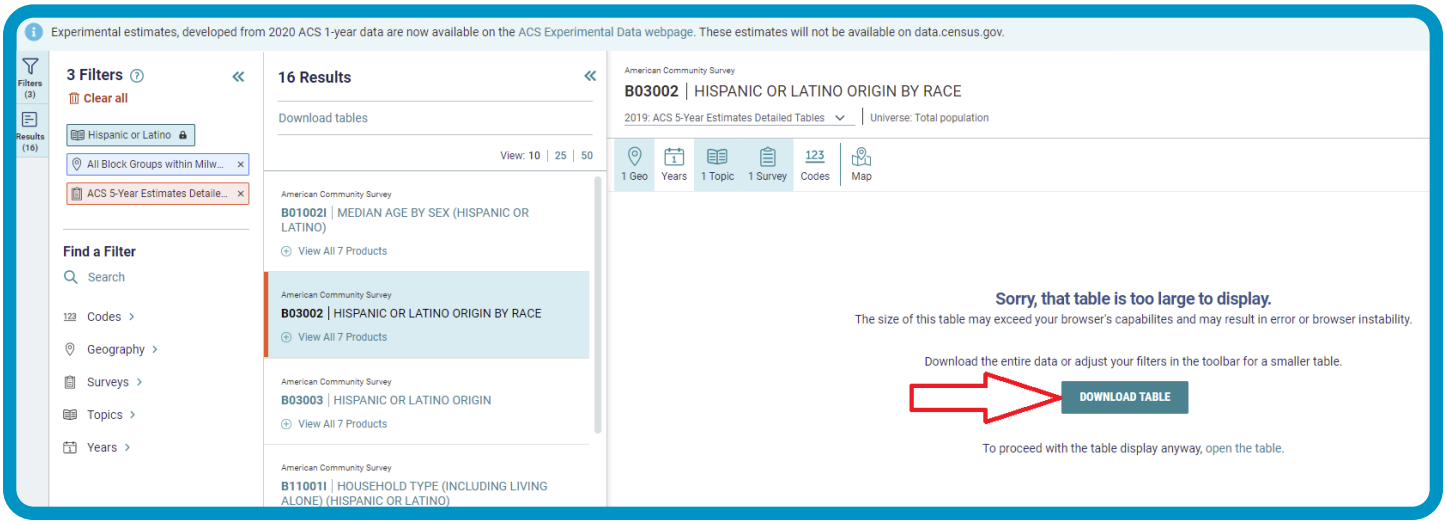
For this tutorial, set the filters to the below:

1. Click "Topics" then click "Hispanic or Latino" then check "Hispanic or Latino" and results will appear to the right.

2. Select Table B03002: Hispanic or Latino Origin by Race

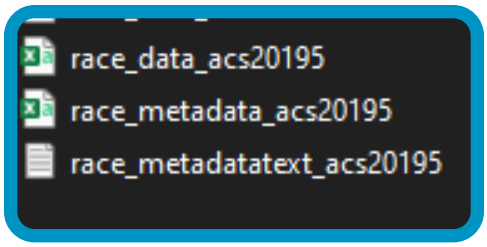


3. To download click on "Download Table" and a pop-up will show you what you will download. You may be given multiple options for what format to download. We recommend downloading as a CSV so that you can clean the datatable for use in QGIS. Then click Download.

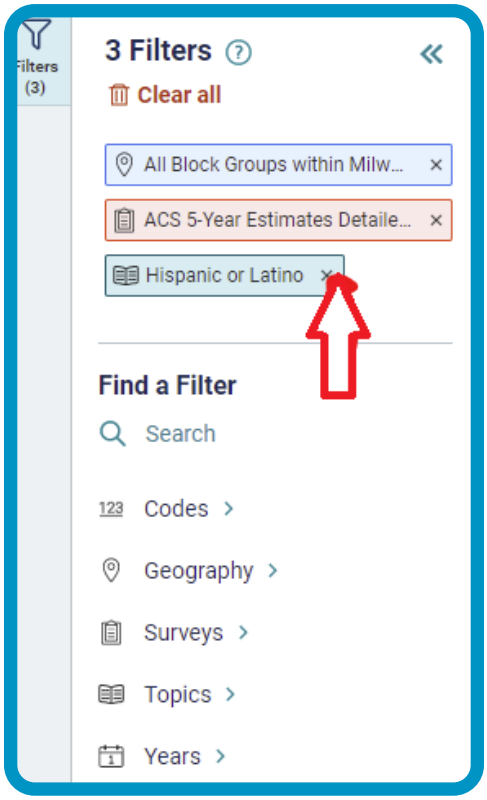


4. You will see a download for a zip file which will contain 3 files. Extract these into your "ACS_Data" folder.

   - Metadata as a csv - as a dictionary to help humans know what each column ID in the data table means. Column ID refers to each of the different attributes from your downloaded data table. The data table uses a column ID so that different computer software (like GIS software) can easily read it.
   - Data as a csv - the data with each geographic unit selected as a row, and each column referring to a specific attribute as labeled by the column ID.
   - Text file - describing the source of the data table, what the table measures, and defining any symbols found in the data
   - You may want to rename each of these files to describe what topic they're about, what year, which survey, and which estimate (ex: race_data_acs20195). Or organize your folder with subfolders about each topic, because if you download several different topics it will be difficult to distinguish just based off the Table ID given by the website. (This method is recommended and used throughout the rest of this toolkit, but you do not have to do this if a different naming method works better for you).



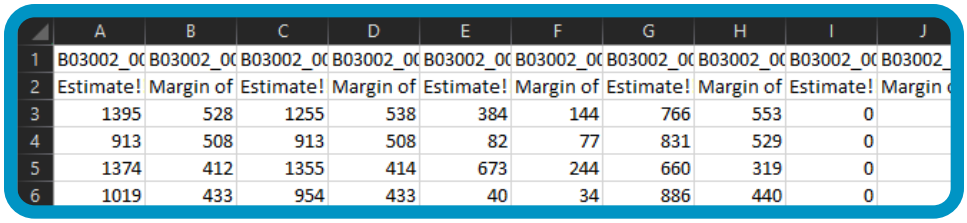
5. Continue searching for each topic you want. Be sure to clear the specific topic you're looking for in the filter before continuing each search by clicking the "x" next to the topic



ACS Data Cleaning

Now that you've downloaded the data, you will need to clean up the data table so that the GIS software can interact with it. It's like when you want to share information with different audiences. You need to format the information so they each can receive and interpret the information.

1. You'll open up the ACS data (that has been saved in a folder as a topic_data_acs[date] .csv file) in Microsoft Excel or Google Sheets and the data table will look something like this.



	A	B	C	D	E	F	G	H	I	J
1	B03002_0(	B03002_0(	B03002_0(	B03002_0(	B03002_0(	B03002_0(	B03002_0(	B03002_0(	B03002_0(	B03002_
2	Estimate!	Margin of	Estimate!	Margin of	Estimate!	Margin of	Estimate!	Margin of	Estimate!	Margin (
3	1395	528	1255	538	384	144	766	553	0	
4	913	508	913	508	82	77	831	529	0	
5	1374	412	1355	414	673	244	660	319	0	
6	1019	433	954	433	40	34	886	440	0	

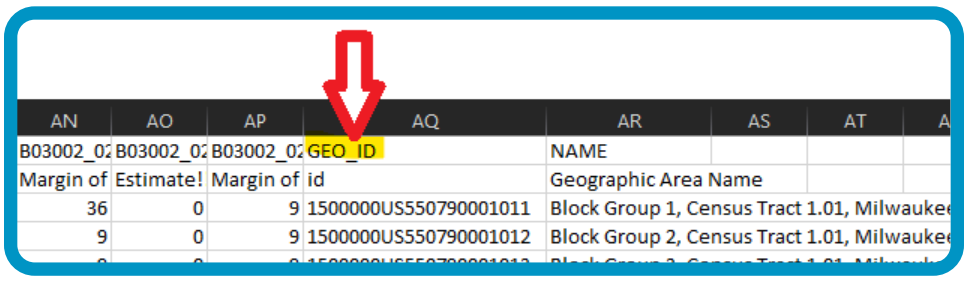
Note: As a reminder of GIS terms rows 3 onwards are the features because each row represents a different location (a census block group in this case). Columns A onwards are the attributes, they are the information/data about each location.

2. Notice Row 1 and Row 2. Row 1 doesn't make much sense to us as readers, but Row 2 does make a little more sense to us. Well, it's the opposite for GIS software. GIS Software can read Row 1, but cannot read Row 2 because it has special characters like!, : and spaces. It will reject the data table if that row is there.

SUPER IMPORTANT TIP: How do we make sure we know what each column is about when we work in GIS software? We need to make sure to refer to the metadata .csv file we saved at the end of "Accessing ACS Data".

3. Delete row 2 by clicking on row 2 to highlight it, and then right click and choose delete.

4. Next we need to make sure that the GIS software can join (connect) the downloaded data table to the block group shapefile. In order to connect it, the label for the location has to be exactly the same. Notice the below two tables. They have a column that refers to the geographic identifier; for the ACS data table, it's one of the last columns. It can be labeled as GEOID/GEO_ID. The location may change depending on when you are using the Census website. Before December 2021, the GEOID/GEO_ID column was the first column, after December 2021 at the time of this toolkit's publication it was one of the last columns, so explore the spreadsheet to find the column.



AN	AO	AP	AQ	AR	AS	AT	A
B03002_02	B03002_02	B03002_02	GEO_ID	NAME			
Margin of	Estimate!	Margin of	id	Geographic Area Name			
36	0	9	1500000US550790001011	Block Group 1, Census Tract 1.01, Milwaukee			
9	0	9	1500000US550790001012	Block Group 2, Census Tract 1.01, Milwaukee			

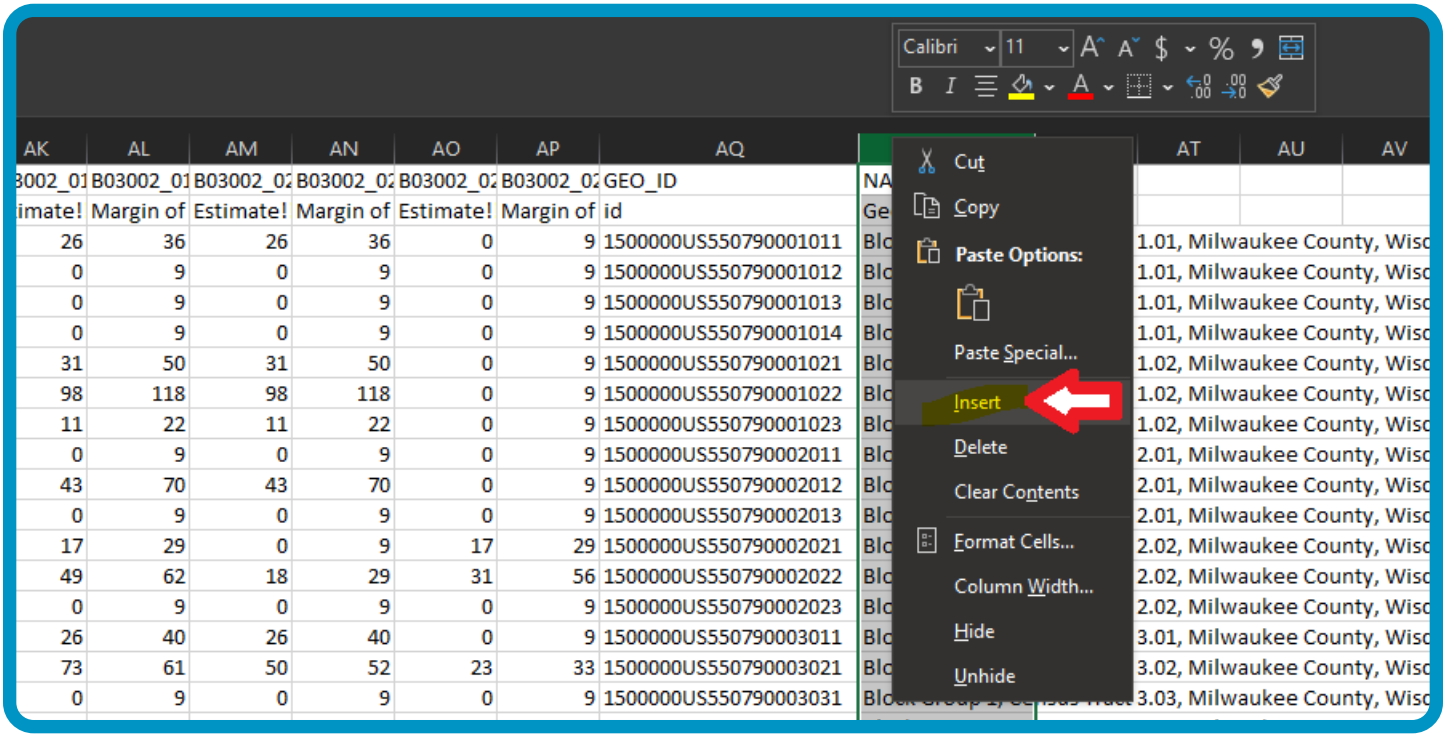
ACS Data Table

5. Open up your QGIS file for this project and open the attribute table for the block group shapefile. At the same time, have your data table from ACS opened up in Excel. When you look at row 1 of both tables (after organizing numerically by using sort & filter in Excel and clicking the column title in QGIS so the black arrow is pointing up) you notice that they look similar, the last 12 numbers are the same. But the ACS data table has more stuff at the beginning. We need to get rid of that.

	STATEFP	COUNTYFP	TRACTCE	BLKGRPCE	GEOID	NAMELSAD
1	55	079	000101	1	550790001011	Block Group 1
2	55	079	000101	2	550790001012	Block Group 2
3	55	079	000101	3	550790001013	Block Group 3
4	55	079	000102	1	550790001021	Block Group 1

Block Group Shapefile

6. To get rid of everything before the list of numbers, an easy way is to split the column, which means to separate one column into multiple columns. This only works because all of the rows are organized the same with 1500000US in front of the 12 digit number. Before we can split the column, we need to create some extra columns that the split column can fill into without deleting any of the data. Right click on the column to the right of GEO_ID (NAME in this case) and click insert. Do this twice so you have 2 blank columns to the right of GEO_ID.



AK	AL	AM	AN	AO	AP	AQ	AT	AU	AV
3002_01	B03002_01	B03002_02	B03002_02	B03002_02	B03002_02	GEO_ID			
imate!	Margin of	Estimate!	Margin of	Estimate!	Margin of	id			
26	36	26	36	0	9	1500000US550790001011	1.01, Milwaukee County, Wisc		
0	9	0	9	0	9	1500000US550790001012	1.01, Milwaukee County, Wisc		
0	9	0	9	0	9	1500000US550790001013	1.01, Milwaukee County, Wisc		
0	9	0	9	0	9	1500000US550790001014	1.01, Milwaukee County, Wisc		
31	50	31	50	0	9	1500000US550790001021	1.02, Milwaukee County, Wisc		
98	118	98	118	0	9	1500000US550790001022	1.02, Milwaukee County, Wisc		
11	22	11	22	0	9	1500000US550790001023	1.02, Milwaukee County, Wisc		
0	9	0	9	0	9	1500000US550790002011	2.01, Milwaukee County, Wisc		
43	70	43	70	0	9	1500000US550790002012	2.01, Milwaukee County, Wisc		
0	9	0	9	0	9	1500000US550790002013	2.01, Milwaukee County, Wisc		
17	29	0	9	17	29	1500000US550790002021	2.02, Milwaukee County, Wisc		
49	62	18	29	31	56	1500000US550790002022	2.02, Milwaukee County, Wisc		
0	9	0	9	0	9	1500000US550790002023	2.02, Milwaukee County, Wisc		
26	40	26	40	0	9	1500000US550790003011	3.01, Milwaukee County, Wisc		
73	61	50	52	23	33	1500000US550790003021	3.02, Milwaukee County, Wisc		
0	9	0	9	0	9	1500000US550790003031	3.03, Milwaukee County, Wisc		

7. It should now look like this:

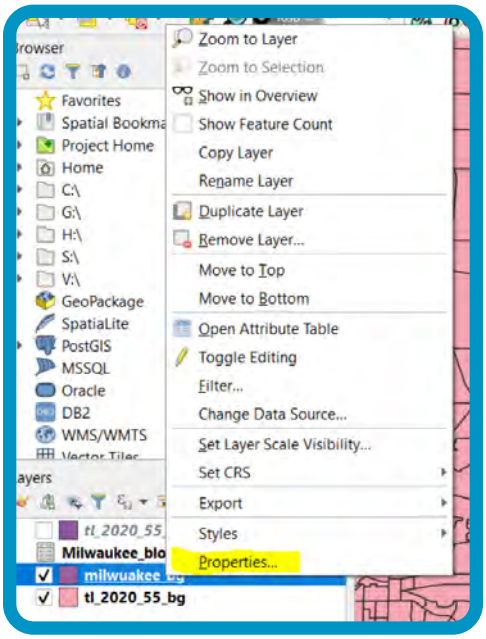
	AP	AQ	AR	AS	AT
2_02	B03002_02	GEO_ID			NAME
ate!	Margin of	id			Geographic Area
0	9	1500000US550790001011			Block Group 1, C
0	9	1500000US550790001012			Block Group 2, C
0	9	1500000US550790001013			Block Group 3, C
0	9	1500000US550790001014			Block Group 4, C

The next step will have us look at data types. Here is some context. Computers store what we type in different ways. Numbers are typically stored in a different way from text or from date/time, etc. Refer to section 4.E for a table on different data types.

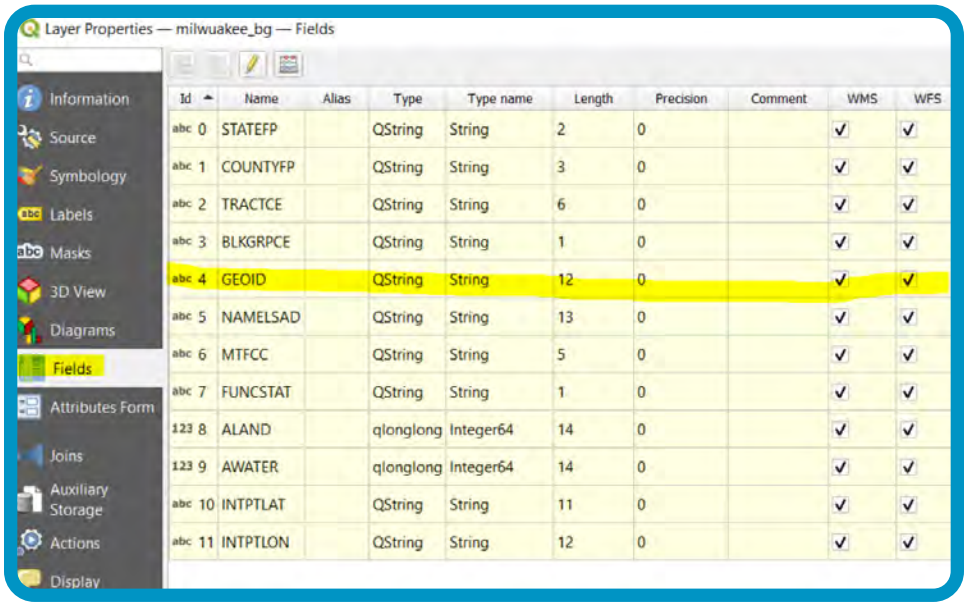
A person can change how the computer is storing the information, for example numbers can be stored as text, which we will do here, so that we can join the tables.

8. Now, before we split the ACS GEO_ID column we need to check what type of data the QGIS GEOID column is to make sure they are exactly the same. To find out in QGIS how the GEOID column is stored, you'll need to right click (or double click) on the block group layer (milwaukee_bg for this example) in the Layers window. Then you'll click on Properties.

   NOTE: this process is different in ESRI ArcMap. You can google "how to find attribute datatype in ArcMap" for steps.



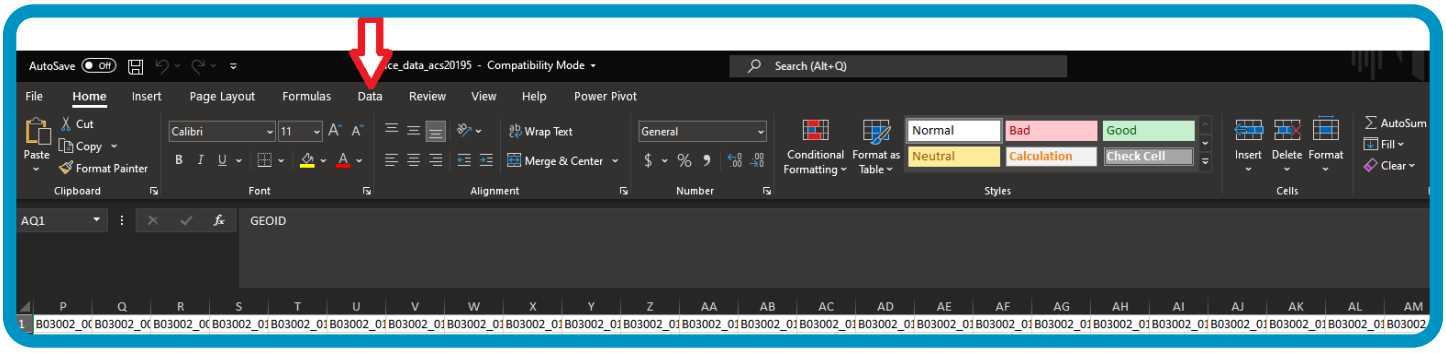
9. Once you're in Properties, click on Fields to show all the attributes in your table. Then find the GEOID attribute. You'll notice in the column "Type Name" it says "String". String data types refer to text.



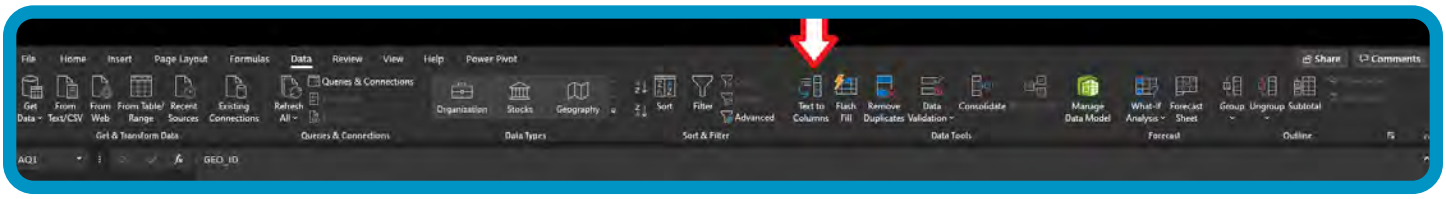
Id	Name	Alias	Type	Type name	Length	Precision	Comment	WMS	WFS
0	STATEFP		QString	String	2	0		✓	✓
1	COUNTYFP		QString	String	3	0		✓	✓
2	TRACTCE		QString	String	6	0		✓	✓
3	BLKGRPCE		QString	String	1	0		✓	✓
4	GEOID		QString	String	12	0		✓	✓
5	NAMELSAD		QString	String	13	0		✓	✓
6	MTFCC		QString	String	5	0		✓	✓
7	FUNCSTAT		QString	String	1	0		✓	✓
8	ALAND		qlonglong	Integer64	14	0		✓	✓
9	AWATER		qlonglong	Integer64	14	0		✓	✓
10	INTPTLAT		QString	String	11	0		✓	✓
11	INTPTLON		QString	String	12	0		✓	✓

So that means, we'll need to make sure that the GEO_ID column in the ACS table is in string or text format, even though it looks like numbers. Changing data types will not change numbers into letters to the human eye; it changes how the computer stores and reads the data.

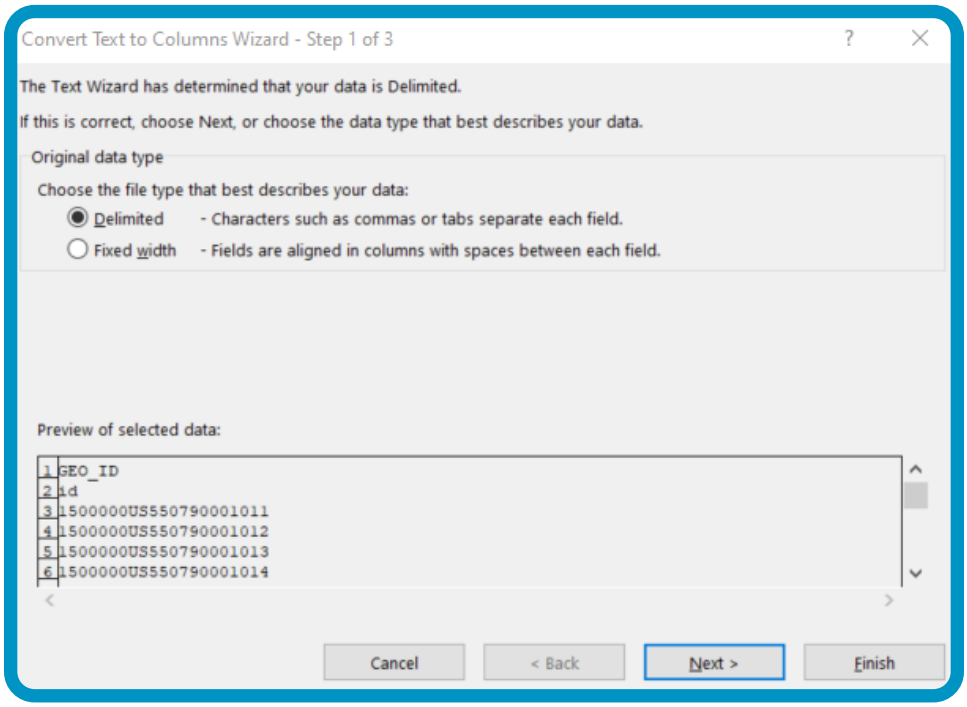
10. Now switch back to the window with the ACS data table. We still need to split the GEO_ID column (and in the process make sure it's the String/Text datatype). Highlight the GEO_ID COLUMN at the top of the Excel window, click on Data.



11. In the Data Tools section, click Text to Columns

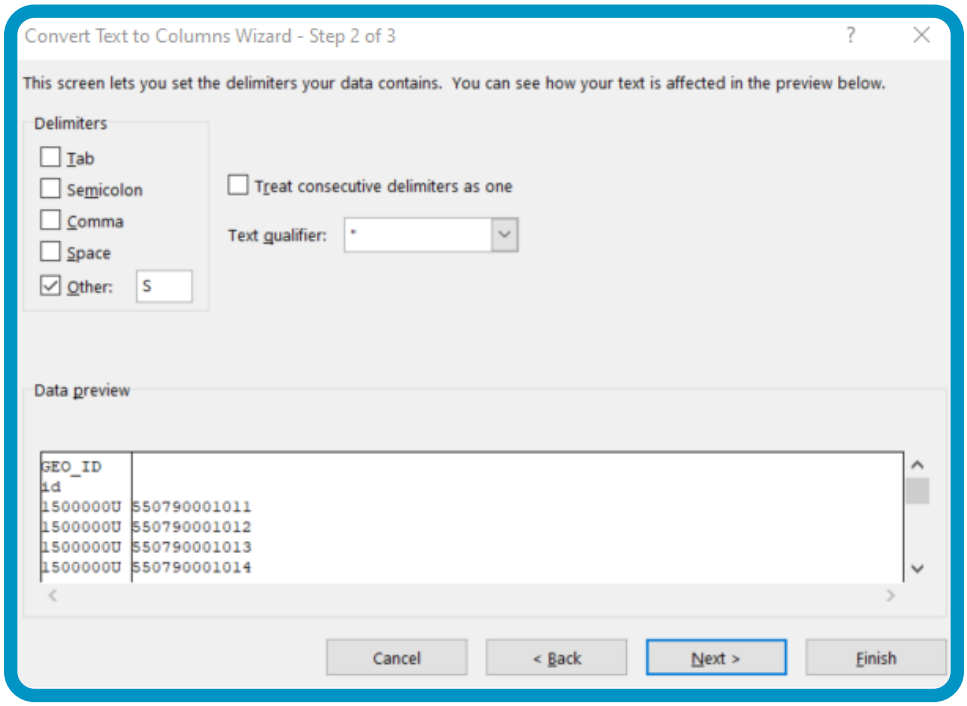


12. You'll get a window like this with the GEO_ID column in the "Preview of the selected data" section



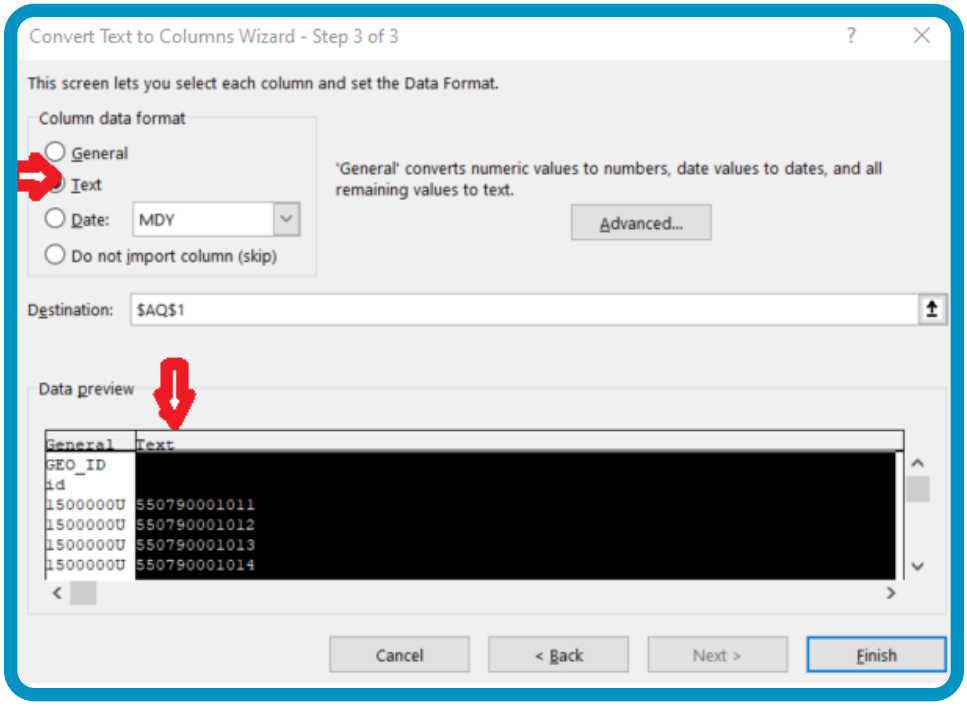
13. Remember, we want to split the columns at the S character, not at a specific width, so we want to choose Delimited in the Original data type section (as shown in the picture above). Then click Next >.

14. You'll come to the window in which you need to choose what Delimiter Excel will use to split the column. For this ACS table, it's the S (in US). So, click "Other:" and type in S. Note, it won't work if in the column, it's an uppercase S but you type in lowercase s. You need to match the case. You can also drag the line behind the "S" to accomplish the same thing.



15. Notice in the Data preview section that there is now a line where the S used to be. This shows where the column will be split. Click Next >.

16. The last step is to choose the correct data type. Remember, we learned that GEOID in the block group layer is a String datatype, otherwise known as Text. So, we need to make sure that the GEOID column is Text. In this last window, make sure that the desired column is highlighted (by clicking on it in the data preview window) and then choose Text in the Column data format section at the top. You'll notice that the top of the desired column switches to Text in the data preview window



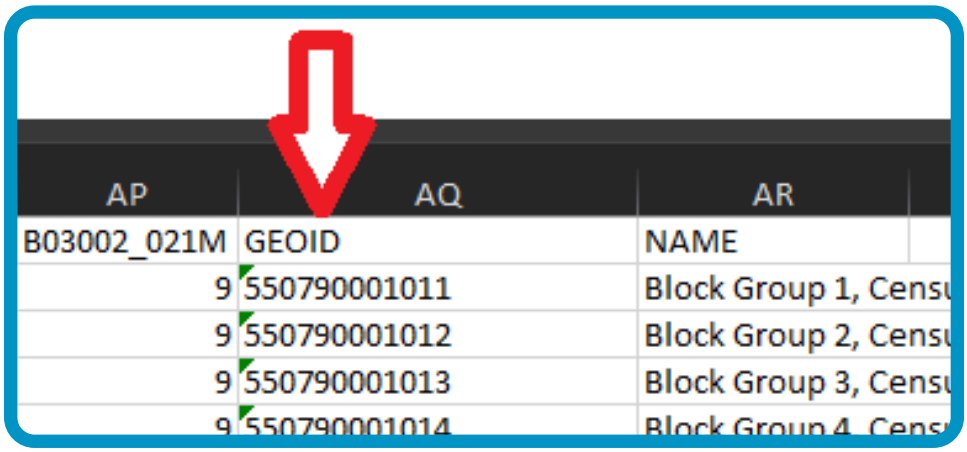
17. Here the incorrect column is highlighted. Make sure to click on the other column.

18. Click finish and you'll see that GEO_ID has been split into 2 columns. The GEO_ID label stayed in Column AQ. Our desired column untitled in Column AR with green little flags. These are green flags because Excel knows that these are numbers, but we decided to store them as Text. Those green flags mean you correctly changed the data type, so don't worry about them. We have an empty column in Column C that acted as a space to prevent data being deleted when we changed the data type. Delete Columns AQ & AS.

	AQ	AR	AS	AT
02	GEO_ID			NAME
of	id			Geographic
9	1500000U	550790001011		Block Group
9	1500000U	550790001012		Block Group
9	1500000U	550790001013		Block Group
9	1500000U	550790001014		Block Group
9	1500000U	550790001021		Block Group
9	1500000U	550790001022		Block Group
9	1500000U	550790001023		Block Group
9	1500000U	550790002011		Block Group

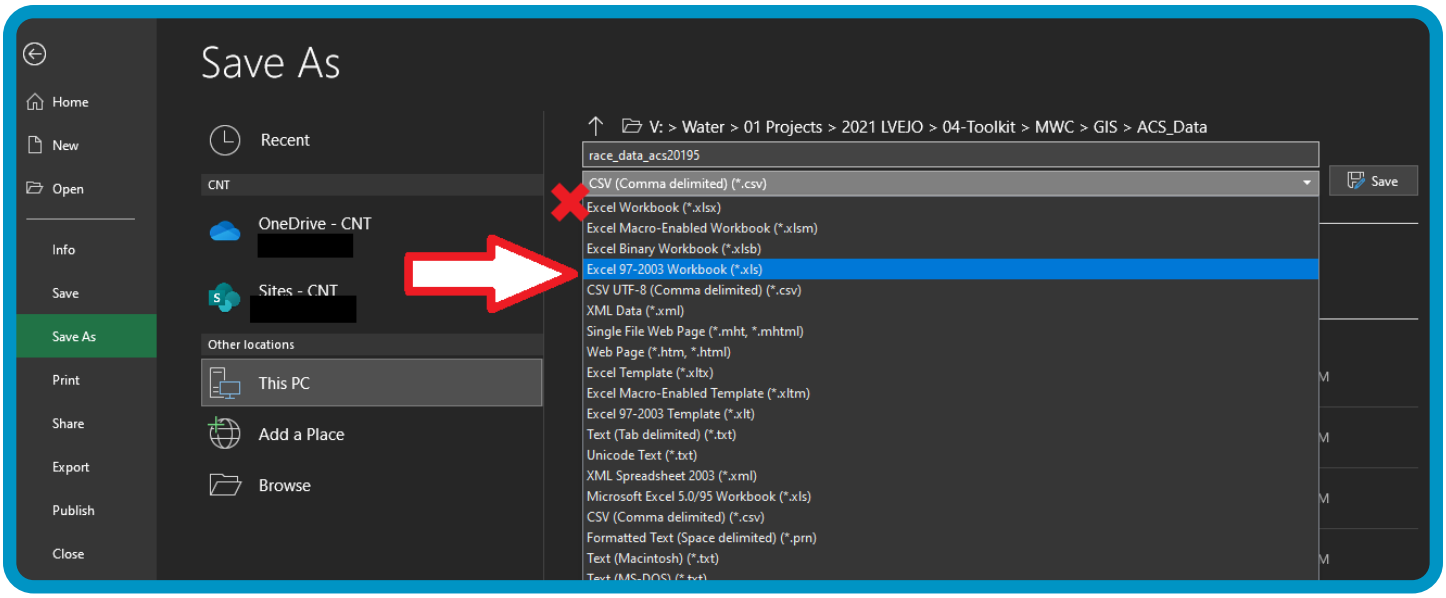
	AQ	AR	AS	AT
M		NAME		
9	550790001011	Block Group 1, Census Tract 1.01, Mil		
9	550790001012	Block Group 2, Census Tract 1.01, Mil		
9	550790001013	Block Group 3, Census Tract 1.01, Mil		

19. Now title the new column AQ as GEOID, so you can easily match this column to the one in QGIS.



AP	AQ	AR
B03002_021M	GEOID	NAME
9	550790001011	Block Group 1, Censu
9	550790001012	Block Group 2, Censu
9	550790001013	Block Group 3, Censu
9	550790001014	Block Group 4, Censu

20. Now Save this data table as either a .CSV or Excel 97-2003 Workbook (.*xls) by clicking on the down arrow. GIS Software typically cannot read the newest Excel files (.xlsx). Close the table.



You'll do the above steps for each ACS data table you downloaded (not for the metadata tables though).

IMPORTANT TIPS

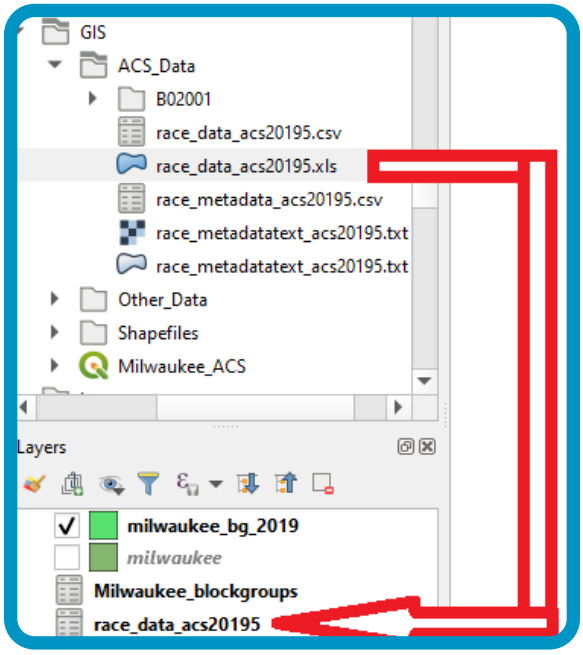
You'll likely need to clean up other data you want to use for your map. Important notes include:

- Title your columns with letters (upper or lower case) or a letter and numbers and underscores.
- Make sure the titles of your columns do not include any special characters (like ! @ # $ % ^ & * ( ) " ' , .) or spaces.
- Don't title your column with only numbers or start with numbers.
- Make sure you have access to a table that explains table columns that don't make sense or are coded (the metadata), otherwise you won't be able to analyze the data.

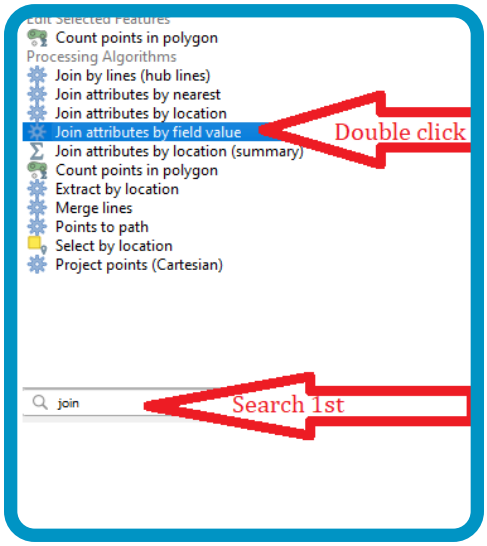
Joining ACS Data to GIS Shapefiles

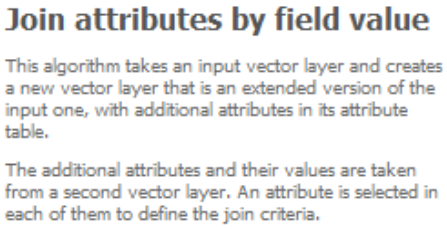
Now that we've prepared the ACS data table, we can attach it to the block group layer. Once you open up the data table in your GIS Software you will not be able to open it up in Excel. So make sure the ACS data table is closed in Excel.

1. In the Browser window, find your cleaned ACS tables and drag it into the layers window.



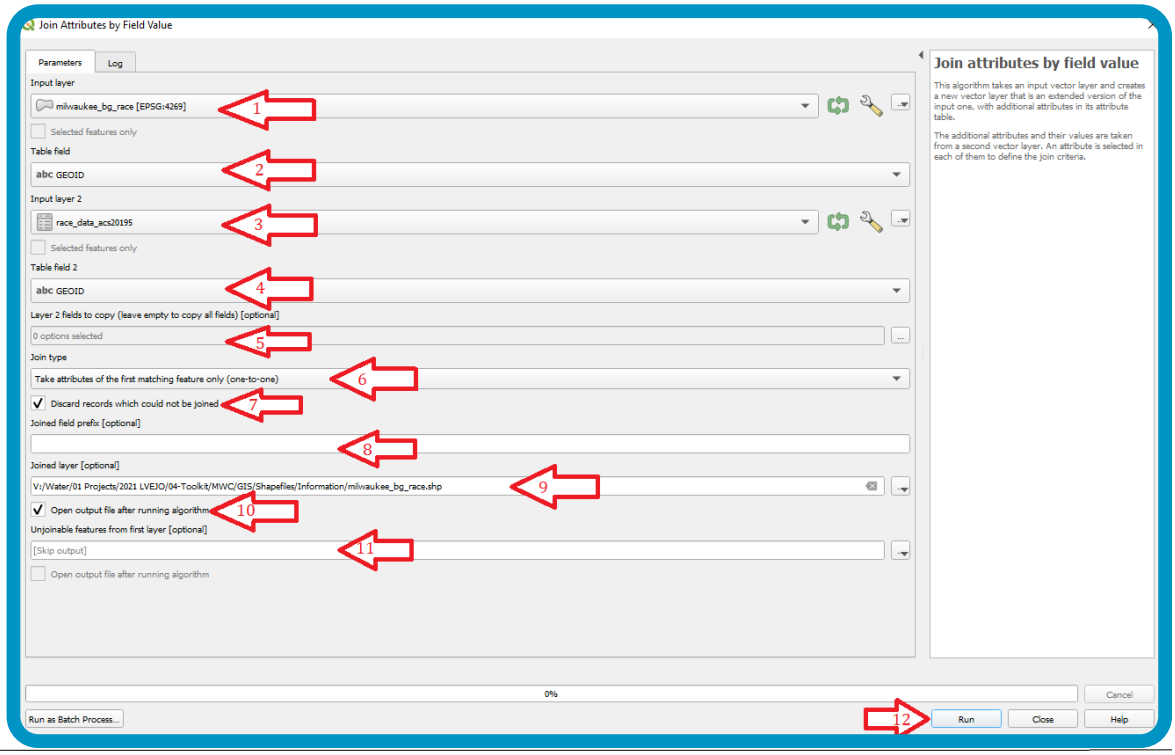
2. You'll search for "Join attributes by Field Value " in the bottom search bar because you are going to join the table to the block group layer where the GEOID labeled- columns have the same value (the list of 12 numbers).



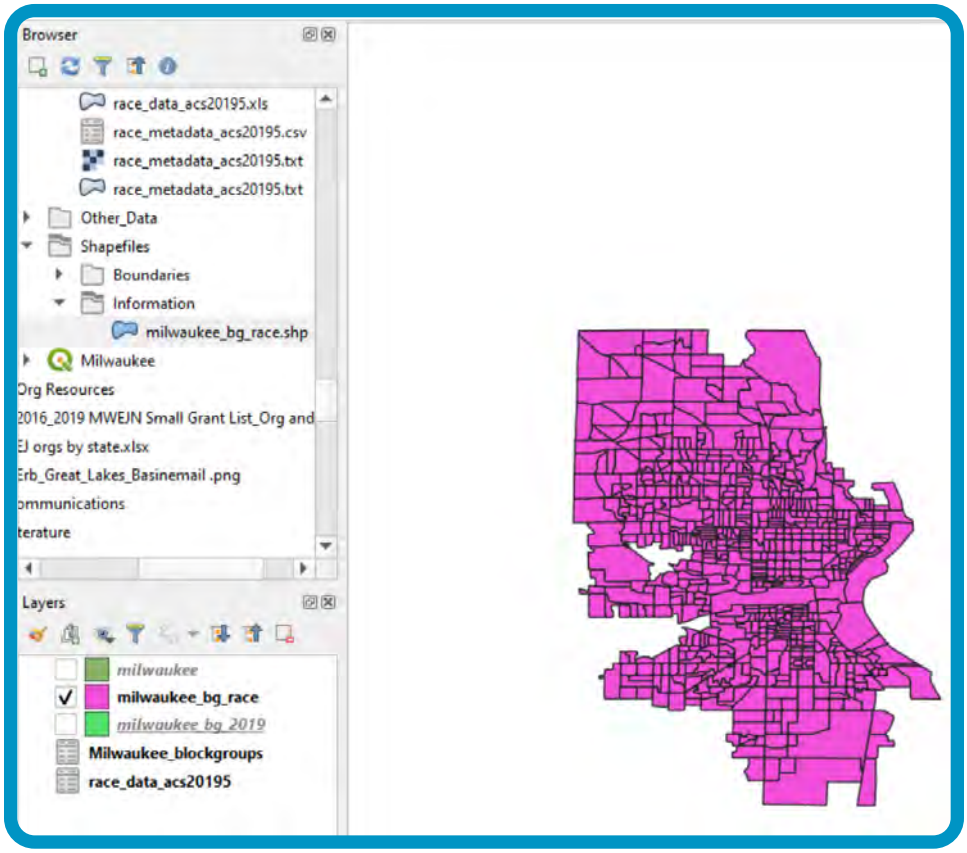


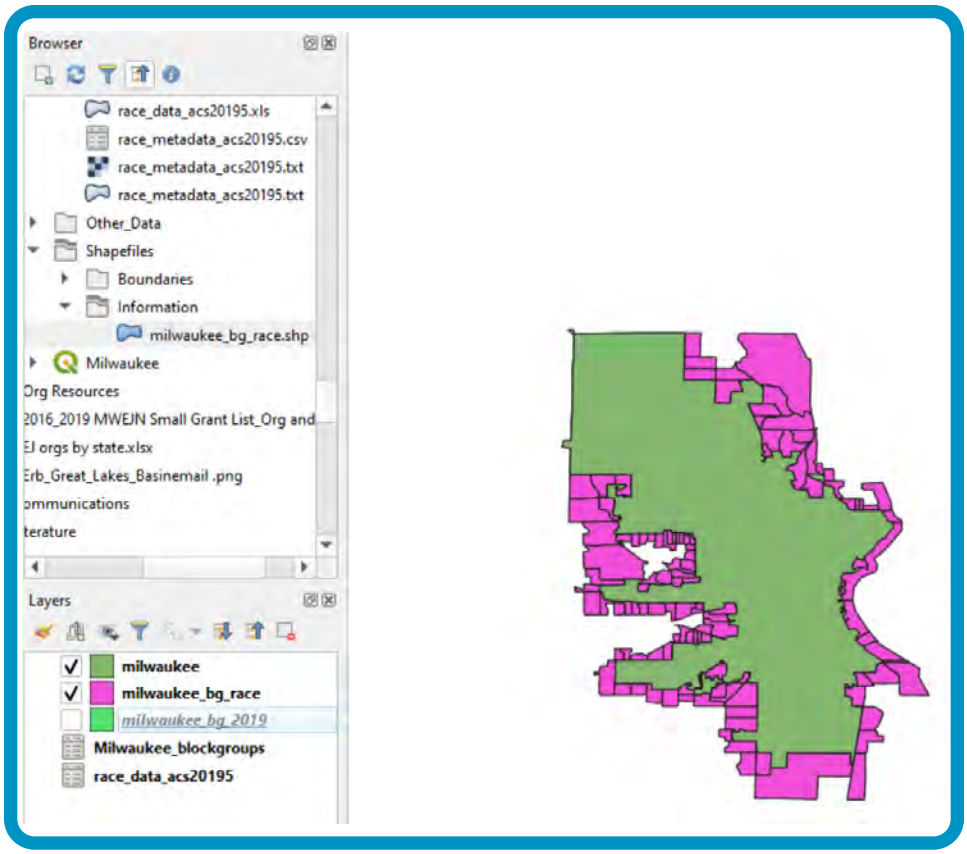
3. In the dialog window you'll want to fill it as follows:

    1. Input layer: block group layer (this is the layer you are going to join the ACS information to)
    2. Table field: GEOID since that is the column that we know matches to the table field in the ACS data table
    3. Input layer 2: ACS data table (this is the datable you want to add to the block group layer)
    4. Table field 2: GEOID since that is the column that we know matches to the table field from the input layer (the block group layer in this case)
    5. Layer 3 fields to copy (leave empty to copy all fields) [optional]: leave empty because you want to copy all fields
    6. Join type: Take attributes of the first matching feature only (one-to-one) because each line in the block group layer is only going to match with one line in the ACS table
    7. You can leave the Discard records which could not be joined to make sure you don't lose a block group you may have had trouble with or if you're block group layer is not completely clean. Check this box if you're sure that the ACS table is completely accurate so that it'll get rid of any block groups that intersected in the first place.
    8. Joined field prefix [optional]: leave blank
    9. Joined layer [optional]: Save to file... and save it in your GIS information folder. If you leave this blank, this join will only be temporary and it won't be saved
    10. Check the checkbox for Open output file after running algorithm
    11. Unjoinable features from first layer [optional]: leave blank
    12. Click run

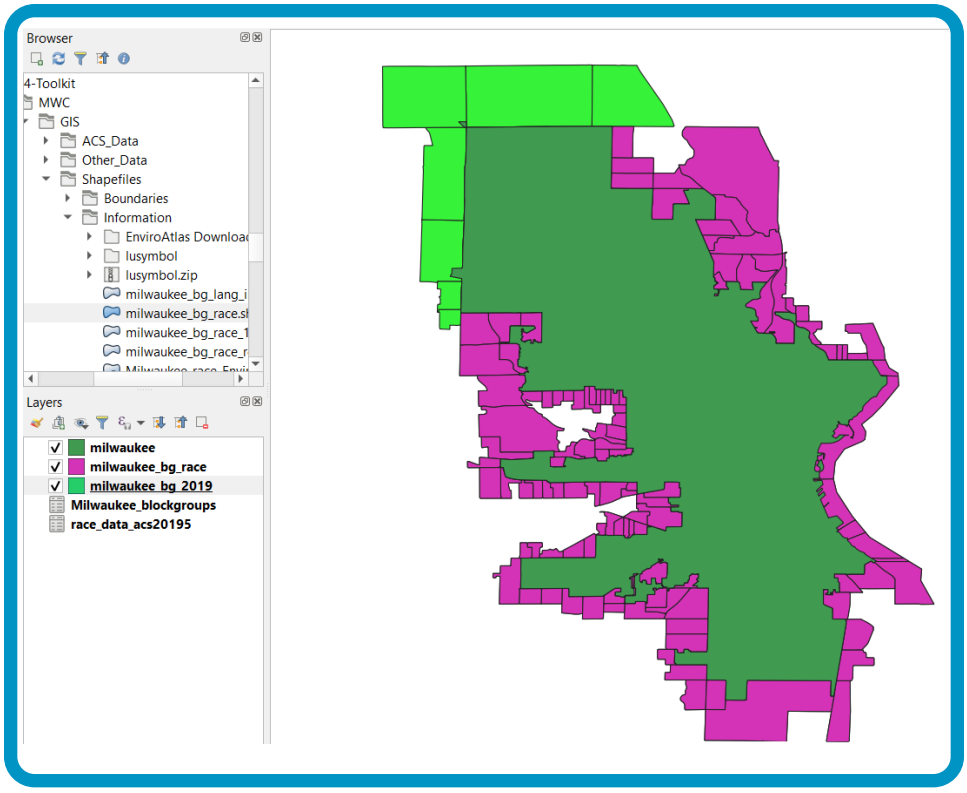


4. The new layer should now show up in your window. This is a good time to see if any data is missing by clicking between layers.



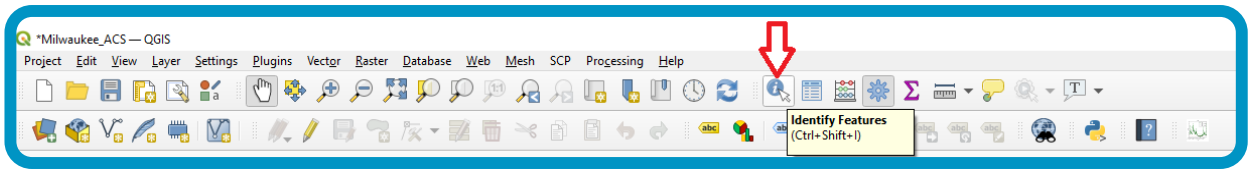


Notice how the pink layer (milwaukee_bg_race) doesn't include two pieces of the milwaukee layer (the triangle in the upper left corner and the squarish-piece on the left boundary).

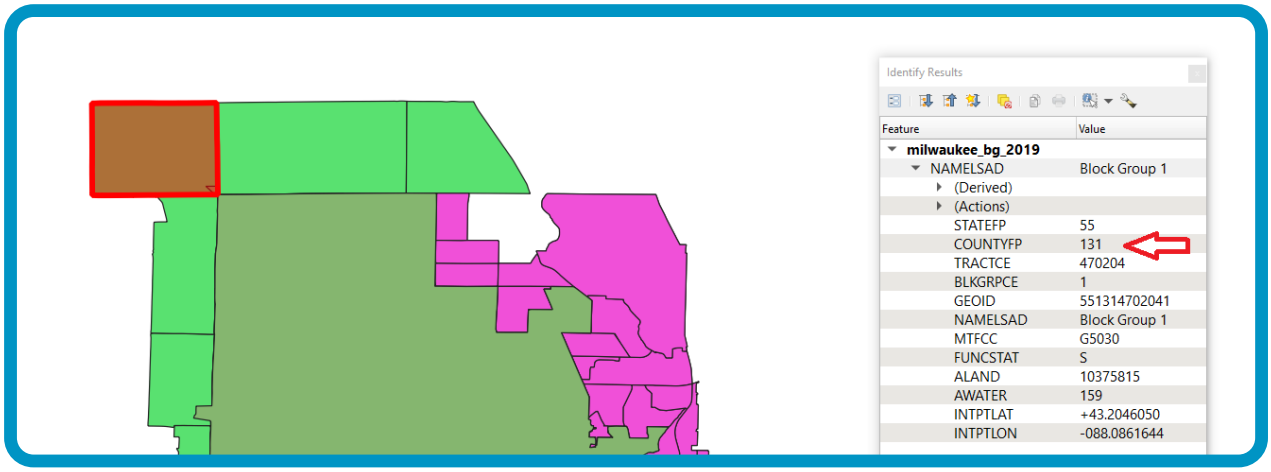


5. Turn on to see which block groups did not get copied over. These did not get copied over because I chose to discard block groups that did not have a match to a row in the datatable race_data_acs20195. Reminder: you have to think critically about whether you want to discard unjoined features.

   a. To check what block groups are specifically missing, in the QGIS toolbar, click on the Identify Features icon.



   b. The cursor/arrow in the map window has a lower-case i next to it. Click on one of the milwaukee_bg_2019 block groups that isn't in the milwaukee_bg_race layer. When clicked on, a window will pop-up with "Identify Results".



c. Notice that this is a different county code than what was for the toolkit's place (Milwaukee which is in Milwaukee County - code 079). Since we only downloaded ACS data for Milwaukee County, this red outlined area did not have a row of information (a feature in the data table) to join to.

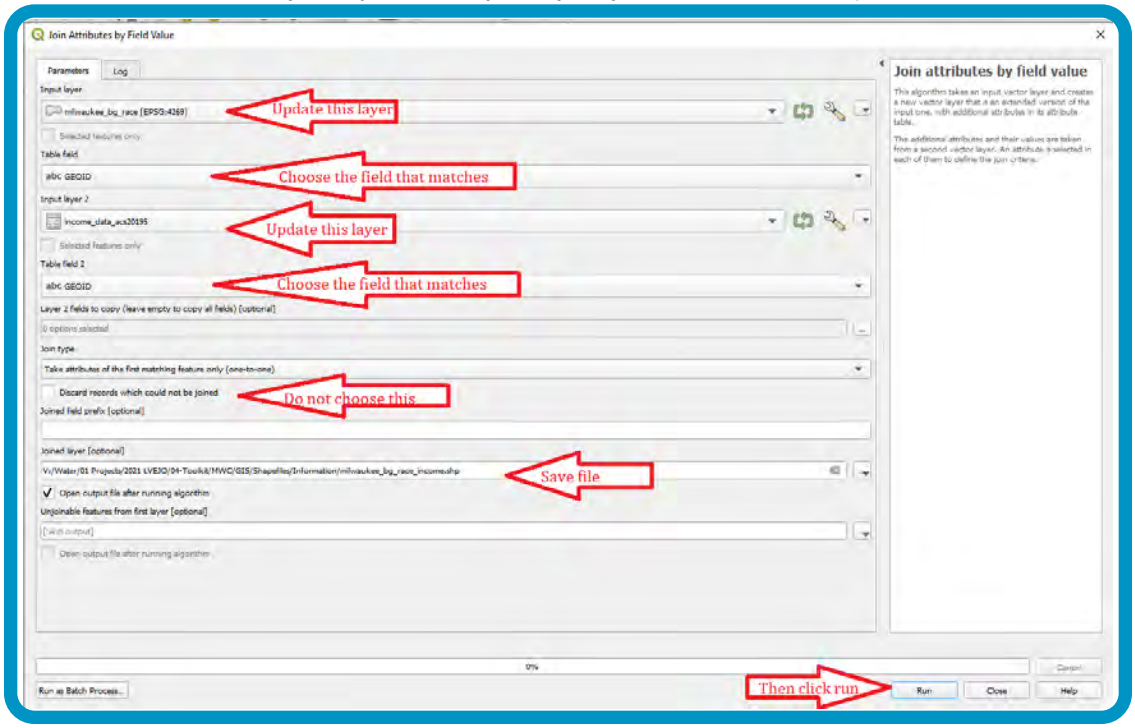
NOTE

If something similar happens to you, talk with your staff if it's important to include this data in the analysis. When trying to decide as staff, it can help to use Google Maps' satellite view to see what is located in the area. If you want to include those other areas in this analysis, you'll want to go through the previous steps of chapter 5 to download data for the additional areas. Then, you need to use the QGIS function: merge the vector layers to have them in the same shapefile to be analyzed together. Alternatively, you can go through previous chapter 5 steps and choose all of the areas you want to analyze.

6. Now that you've finished joining one ACS layer, you can join your other ACS data tables to this same layer. You'll need to complete Accessing ACS Data and ACS Data Cleaning for each data table before joining them.

7. Once your ACS data tables are prepared, you can complete steps "Join Attributes by Field Value" steps 1-3 except make sure that in the input layer, you choose the layer you most recently created (for ex milwaukee_bg_race) and for Input layer 2, choose the newly desired ACS table and do not Discard records which could not be joined, now that the block group shapefile is all cleaned up. For the Joined layer[optional] you name it to include what ACS tables are now joined in the layer. Note: be careful of making the names too long, so you need to think of shorthand for the name of the different layers.

8. Once you've joined all of your layers, you can move onto chapter 6.

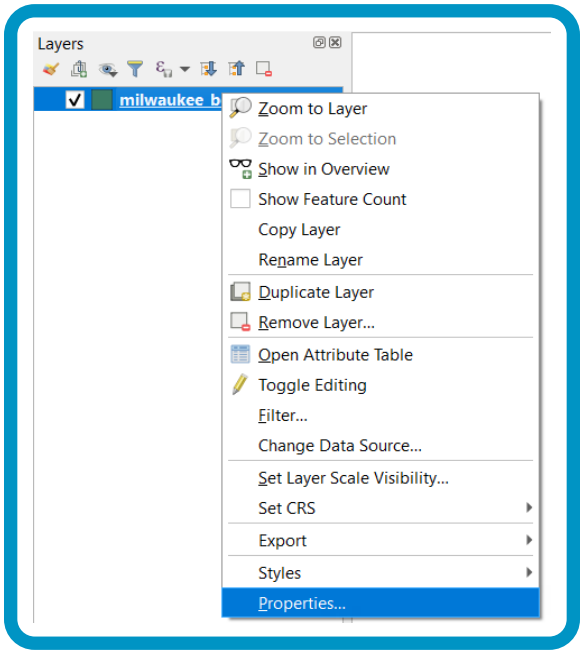


Remember, it will be helpful to have your metadata table up as a reference for the next section as you analyze the data.

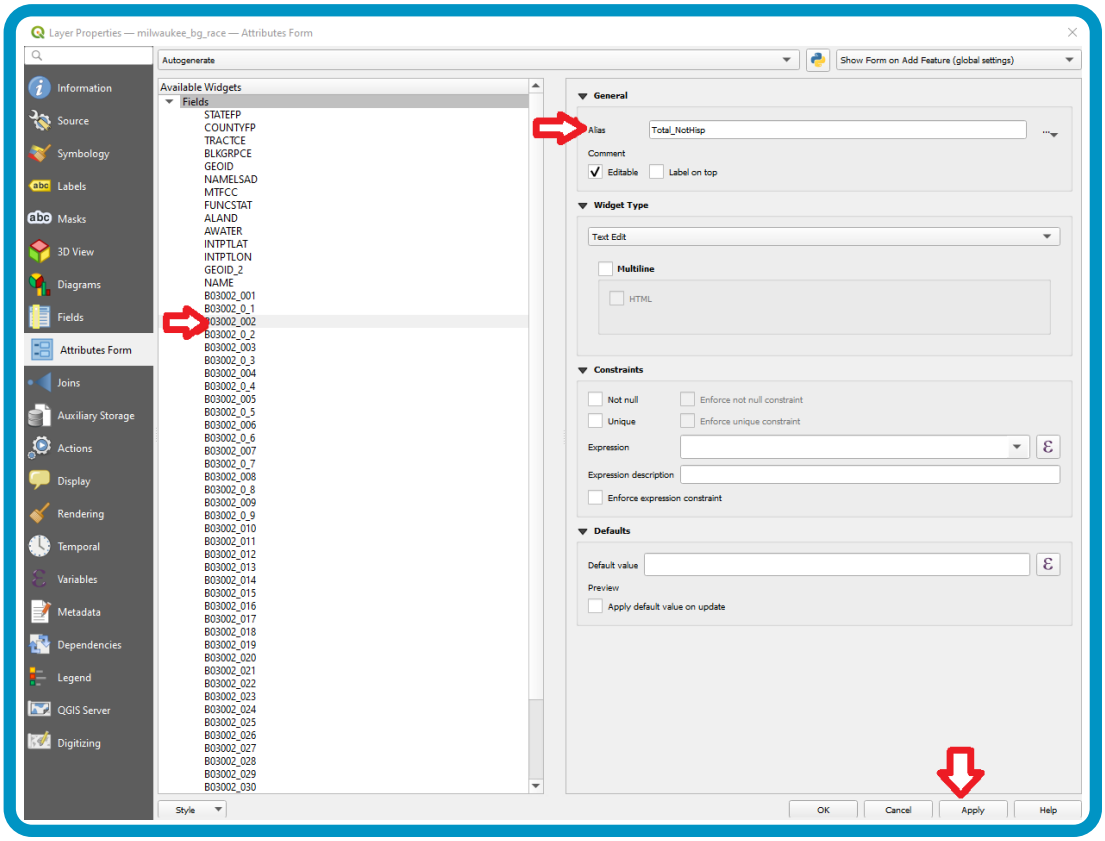
9. [Optional] - You can create an "alias" for your attribute names so that it is more easily understood by a human. This doesn't change the name of the attribute, it just tells the software to show you, the human, a different name. This also means that the alias is only

saved in this project of QGIS, not into the shapefile layer. So, every time you open the shapefile layer into a new QGIS project you'll need to repeat these steps.

a. Navigate to Layer Properties, by right clicking on the layer you want to add aliases

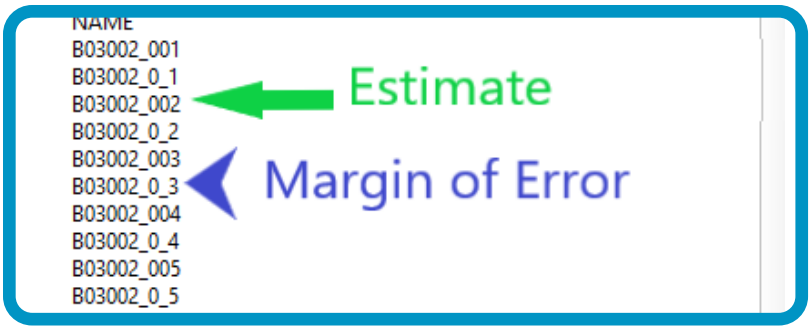


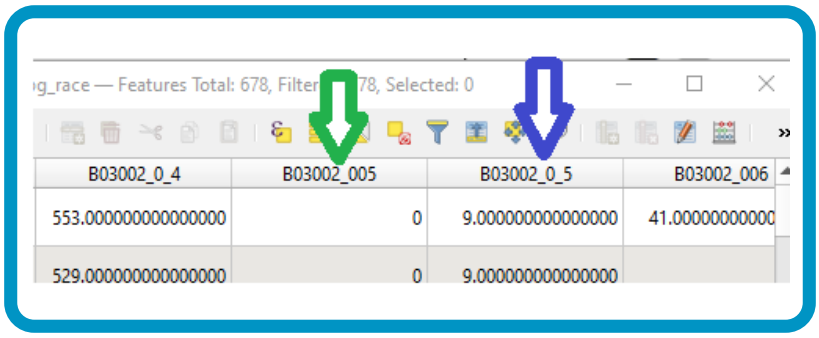
b. Click on "Attributes Form" in the left-hand list



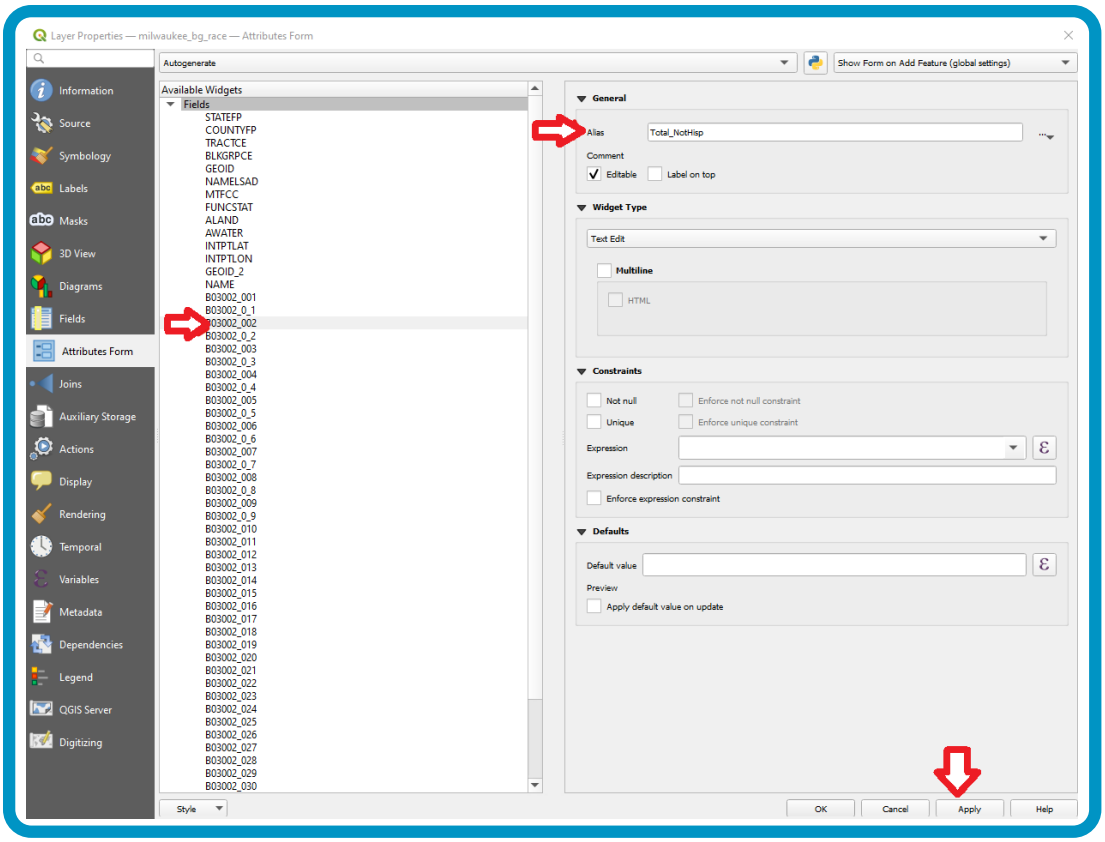
c. Click on the attribute you want to change and you'll see that the gray window on the right fills up with a section called "alias". Refer to your metadata table to find what the attribute represents and type it in. It may not be titled exactly the same, because ACS data includes estimates and margin of error for each topic (so the numbers might even be off once the attributes get to double digits) so be careful and be sure you are creating the right alias. Typically though, the estimate (which

you'll use to map data) is the first one in the set. You can compare the values in columns from your race_data_20195.xls table and the attribute table from the milwaukee_bg_race layer to make sure.

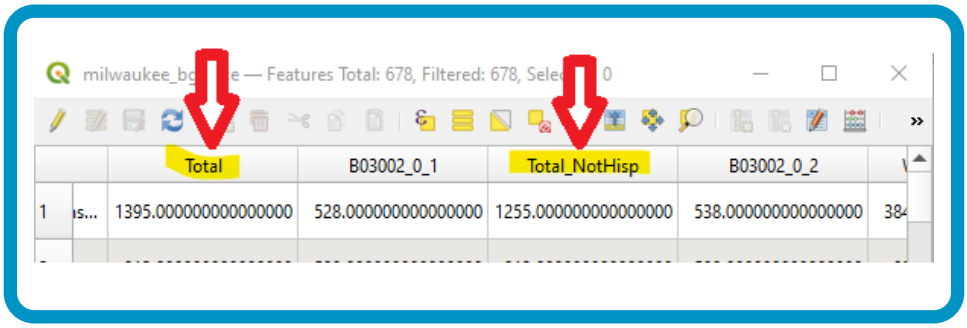




d. Then click "Apply" and go to the next variable you want to provide an "alias". When you are doing that for all of the variables you want, click "Okay".



e. Now when you open the Attribute Table you'll see the aliases.



These steps are optional, if you'd prefer to reference the metadata table without creating an alias, you can.

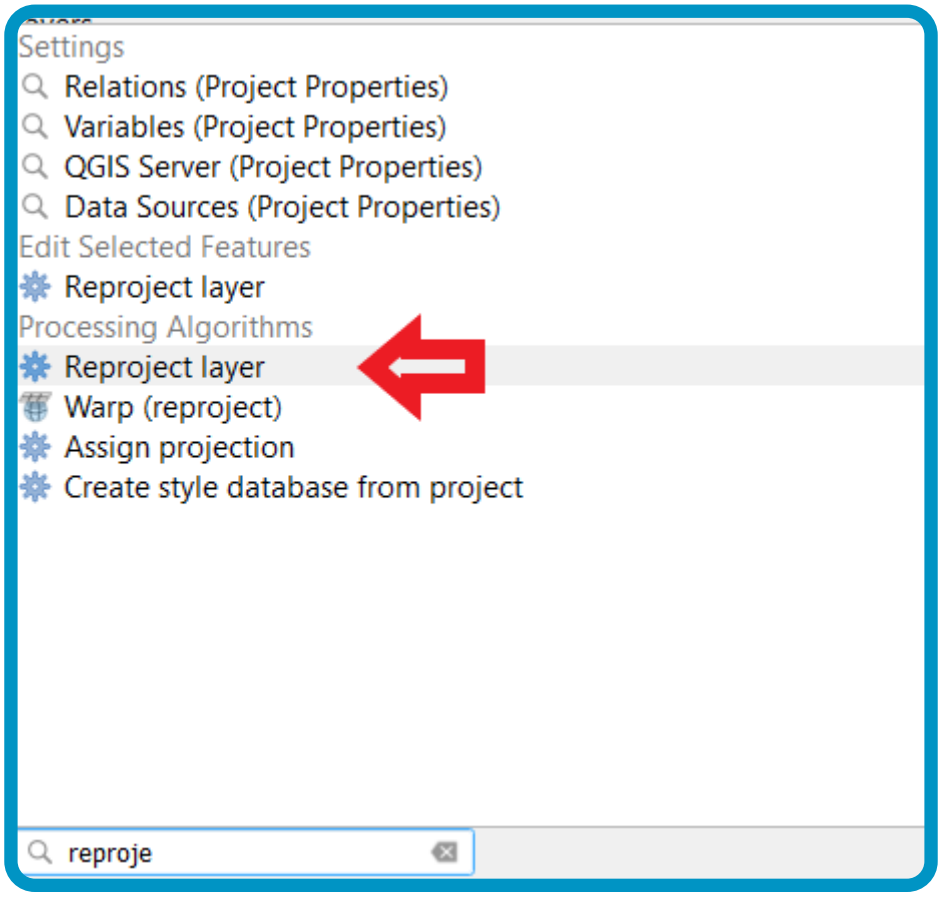
Remember: It is important that you frequently Save (Ctrl+S on PC or Command+S on Mac) as you go to avoid losing your work.

F. Preparing layers for analysis

Projecting all the layers into the same projection

Before we can analyze the maps we have visualized in GIS we need to make sure that they are projected the same, especially since we want to join the EnviroAtlas data map to the ACS data map. This will make sure that both layers are aligned in the final map we may want to make.

1. In the bottom search bar start searching "Reproject layer". You will see that two tools will come up. Click the "Reproject layer" tool under "Processing Algorithms".



2. We are going to change the coordinate system of our ACS shapefile. Therefore, we need to check the coordinate system of the EnviroAtlas shapefile first. Go into the EnviroAtlas shapefile's properties. You will see that it is in NAD83 / UTM zone 16N.

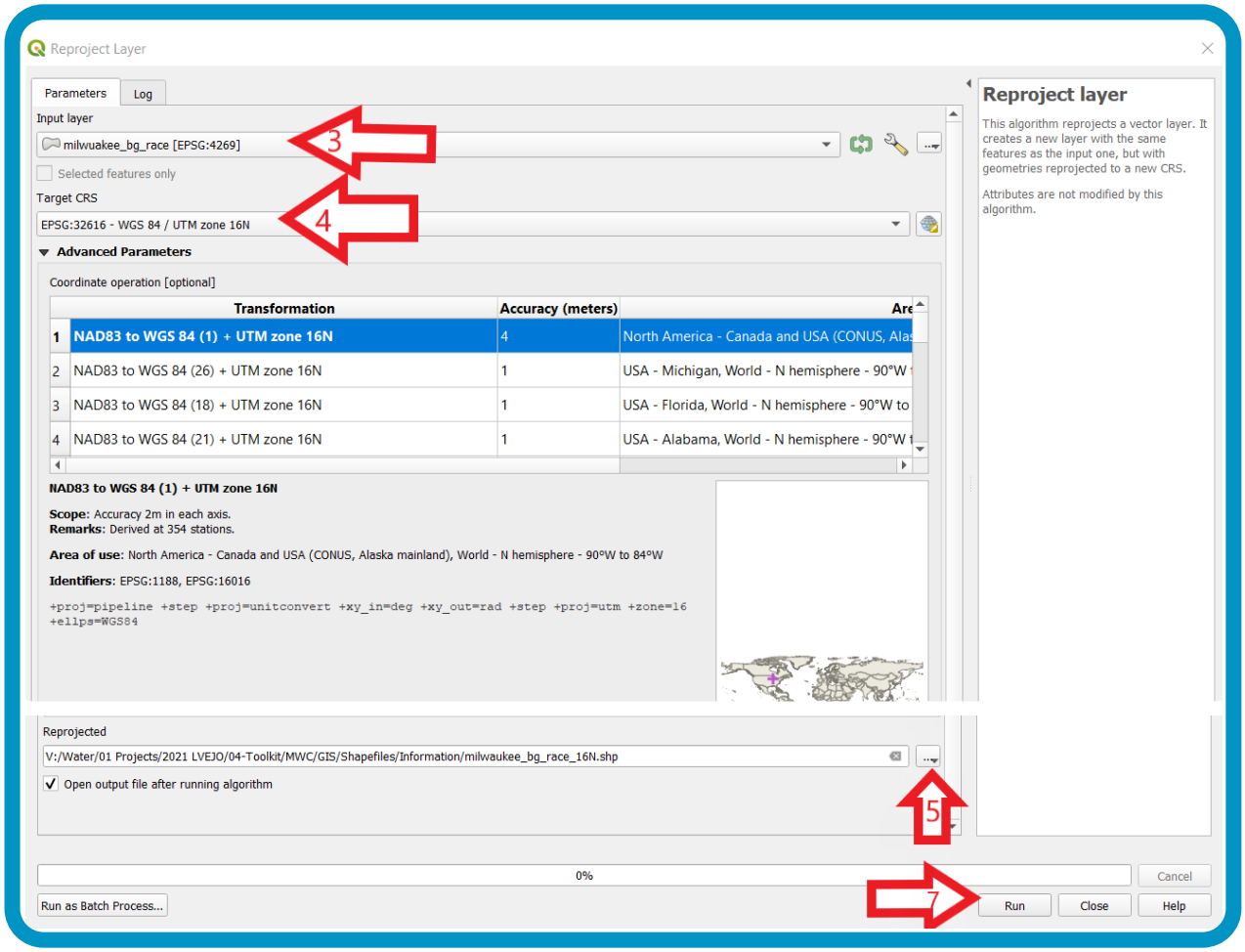
3. In the tool window select the ACS shapefile from the "Input layer" drop-down menu.

4. For "Target CRS" choose NAD83 / UTM zone 16N from the drop-down menu. If you do not see it there, click the "Select CRS" button to the right of the menu and choose it from the window that opens up.

5. For Reprojected, click the button to the right with the three dots and select "Save to File..."

6. In the Save file window that pops up, navigate to the folder where you want to save the reprojected shapefile. Type in the new name for the shapefile you want to give it and select "SHP files (*.shp)" from the Save as type: drop-down menu. Then hit Save.

7. Below is how the Reproject Layer tool window should now look. Leave the "Open output file after running algorithm" option checked. Click Run.



8. You should now see your newly projected ACS shapefile under the Layers window and displayed in the map area of QGIS. It will look the same as before, but QGIS will have given it a new color. We will be working with this version of the ACS shapefile for the rest of this exercise.

Now that we have both of our shapefiles aligned and projected in the same coordinate system, we will join their information so that we can calculate our normalization and visualize the new information.

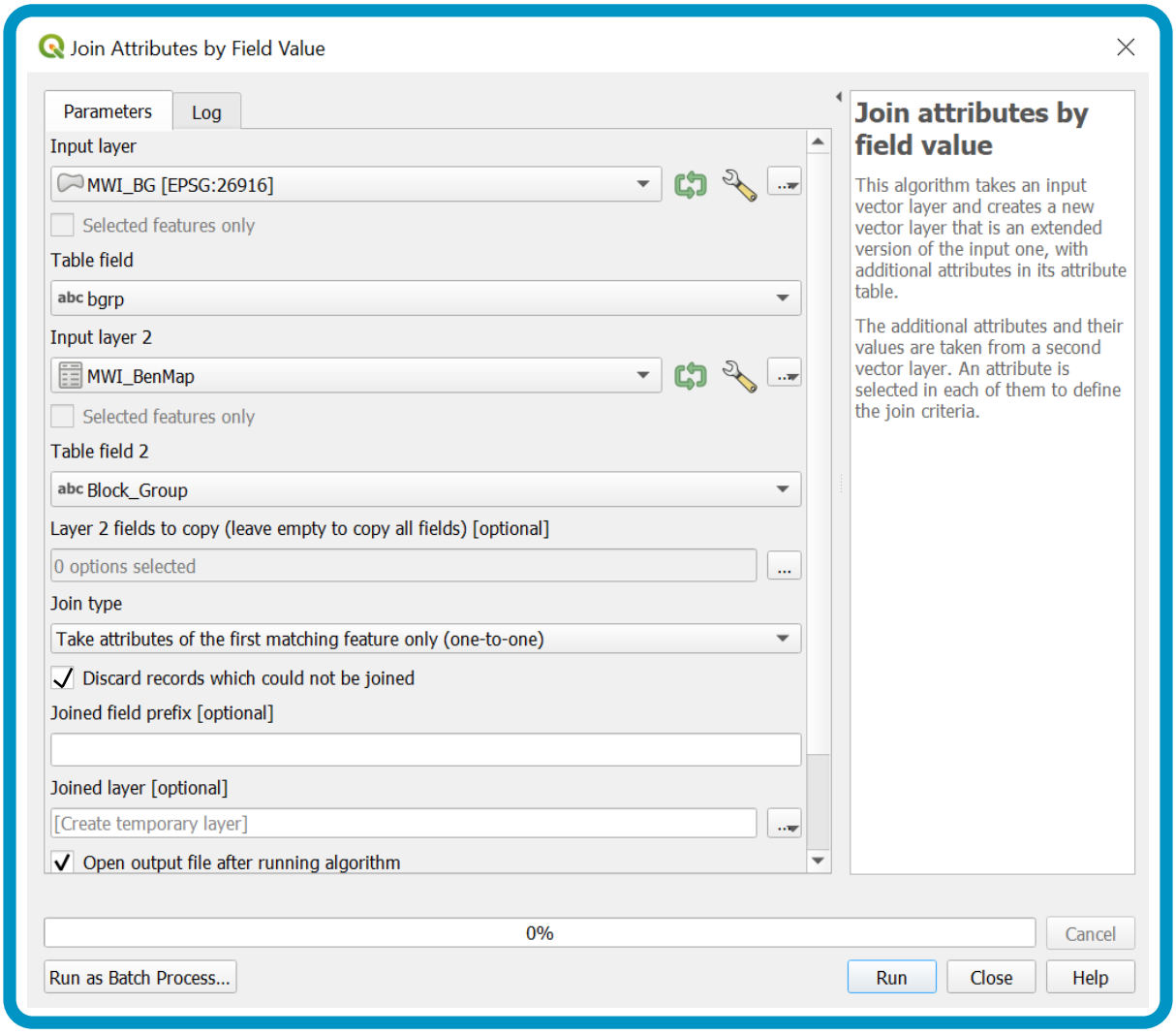
Join the EnviroAtlas Layer to the ACS Layer

1. Similarly to previous examples, we will join our two shapefiles of interest by census block ID. Open the attribute tables of the two shapefiles and note the column name for the block ID in both shapefiles. These should be "GEO_ID" in the ACS shapefile and "bgrp" in the EnviroAtlas shapefile.

2. Search for the "Join Attributes by Field Value" tool in the bottom left search bar again. For a refresher on using this tool, refer to this section.

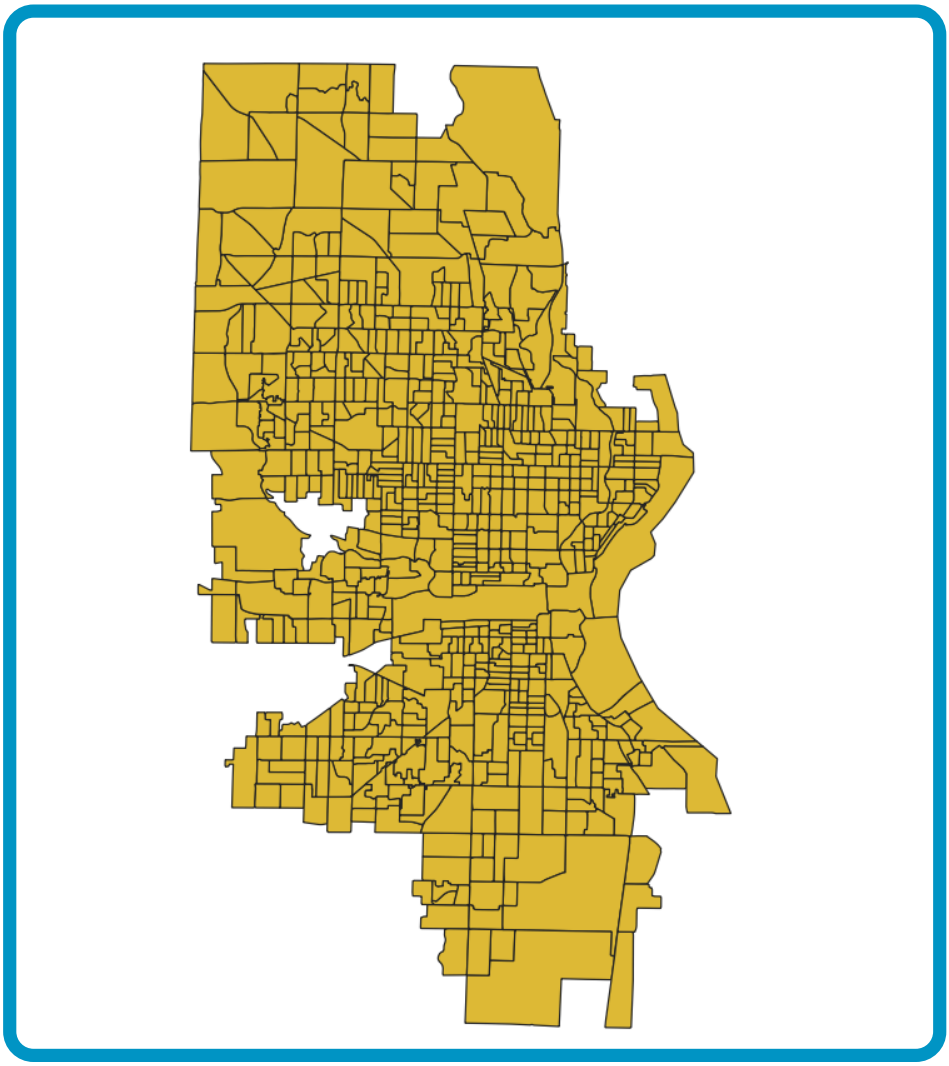
3. Choose the ACS shapefile as the Input layer and the EnviroAtlas shapefile as the join layer. Choose the correct attribute column names for census block ID for each shapefile respectively.

4. [Optional] Create a new shapefile for the newly joined attributes by choosing a location and file name under Joined layer.



5. Check the box for the option "Discard records which could not be joined" so that we can simultaneously clip the EnviroAtlas data to Milwaukee city proper. As usual, check the box for the "Open output file after running algorithm" option.

6. Click Run. Your new shapefile should now look like the example below



Remember: It is important that you frequently Save (Ctrl+S on PC or Command+S on Mac) as you go to avoid losing your work.

G. Other Sources for the Issue

Your research question may require data that cannot be found on EnviroAtlas or on the ACS. The following are some suggestions of resources to explore and find GIS data for different topics. Each of these resources host several different data sets with varying topics, geographic scales, age of data, data format, and have their own instructions on how to download. It is recommended to go through the page's tutorial (if it has one), about page, or help section. Otherwise, please explore the website to see what it can offer in more detail.

Energy

- Open Energy Data Initiative - By the U.S. Department of Energy and includes hundreds of different energy related data sources and even other databases. To find GIS data related to the topic of interest, in the search bar it is recommended to type "GIS + [topic of interest]"
- Low Income Energy Affordability Data - By the U.S. Office of Energy Efficiency and Renewable Energy. It provides data on average energy burden as percent of income, average annual energy costs, and housing costs. The smallest geography available to download is at the census tract level. The platform was updated in 2020 with data from 2016. The data is available as a CSV.

Flooding

- National Flood Insurance Program - By the U.S. Department of Homeland Security's Federal Emergency Management Agency. It provides records of all the flood insurance claims that have been made with the National Flood Insurance Policy (offered to homeowners in certain flood zones) at the census tract level. The data is available as a CSV.

Food Access

- Food Access Research Atlas - By the U.S. Department of Agriculture (USDA). It provides indicators of food access at the census tract level. It appears to be updated every 5 years. The data is available as an excel document or a zipped folder.

Health

- Centers for Disease and Health Promotion Data & Indicators - By the Centers for Disease Control and Prevention (CDC). It includes hundreds of different data sources related to health. To specify for GIS data use the left hand list and click on Maps. Then you can find maps relevant to different health indicators, across various years. Typically, the smallest geographic scale available is at the census tract level.
- PLACES / 500 Cities Project - By the Centers for Disease Control and Prevention (CDC). The 500 Cities Project provides incidence of health issues from 2016 to 2019 at the census tract level. It was rebranded/expanded to the PLACES project in 2020, which also provides a variety of health data available at the census tract, zip code, county, or municipal/regional level.

Pollution

- NATA - By the U.S. Environmental Protection Agency. It includes several different health impacts of air pollution. As of 2021, the most recent data available was collected in 2014. The data is available at the census tract, county, and state level.
- EJ Screen - By U.S. Environmental Protection Agency. It provides measures on demographics, and environmental indicators (covering various data sources on air, water, waste, and lead). The data was collected over various years. The data is available at the census tract level.
- Additional Tools from the EPA - By U.S. Environmental Protection Agency. This list of resources includes additional websites/databases for access to various other environmental and pollution related data.

Worker

- Longitudinal-Employer Household Dynamics (LEHD) - By the U.S. Census Bureau data provides the work and home locations of workers in different sectors. The On the Map application can be used to understand this data, by showing commuting patterns of workers within a certain geography as well. The smallest scale of data is at the census block level. This data may require more advanced comfort with data analysis and software.

Other

- Center for Neighborhood Technology - A non-profit leader in data analytics, CNT provides tools on a variety of topics like housing and transportation costs and green stormwater infrastructure benefits and across multiple geographic areas.

In addition to federal sources, you can look to data sources from the state or the municipality. Sources available from each state vary. The data may be accessed through the various agencies of the state. Commonly found datasets include locations of courthouses, social security offices and other government facilities. Some municipalities have funding to make data available at the local level. The data that is typically available at the municipal level include land use, zoning, and building footprints. To find data at either the state or municipal level, go to the state or municipal website and use the search bar to look up "GIS data."

You may also find sources by searching the internet for "[topic] + GIS data". Remember, make sure that the methodology on how the dataset was created is clear and transparent. If someone is unwilling to share their methodology or their data sources, it likely is not trustworthy or good data.

H. FOIA process

What is FOIA?

FOIA is the Freedom of Information Act and allows the public to access many government records. To make a FOIA request, you need to identify the agency or department that would have the data, provide a written request (in letter or electronic format depending on the agency), and specify details on what data is desired and in what format. FOIA requests can be denied if the person who fills the request believes it fits within an exemption. So, it is important to be really specific about what data you desire.

What data is available via FOIA requests?

FOIA requests can be made of any agency records or documents that already exist. It cannot be asked of data that is not yet created or would require a research methodology. FOIA requests can only ask for the data; no analysis will be provided. Requests may be denied if the data is publicly available on their website or accessible in-person.

FOIA requests may cost money if it takes the agency more than 2 hours to find the data. Check in with the agency if a fee will be required. They should notify you if there is a fee required over $25.

How do I request a FOIA?

- *Identify what data you are looking for*
- *Identify which department would have this data and go to their website to search for a form or call a staff member to find out what their process for FOIA requests are.*
- *When writing a FOIA request, include your name, address, contact information, and organization. Then specify which data is desired (from what agency service - like 311 calls or Vital Records; specific attributes of the data - like time stamps, description; format of data - like CSV;what should be included in the data set - like latitude and longitude, or address with zip code, or block number (and you can provide multiple alternatives if they can't fulfill your top priority).*

EX:

a. *Subject: Vital Records*

b. *Type of Inquiry: General Public*

c. *Type of Request: Requesting copies*

d. *I am requesting datasets in CSV format for low birthweight births at the block group level for the counties of: Cook, DuPage, Kane, Lake, McHenry, and Will from 2013-2017. Preference is for data to be shared as: low birth weights as a percentage of total births for each block group. However, I can also accept if the data is a count of low birth weights and a count of total births for each block group. Dataset is required to include 12-digit FIPS code. If data can't be provided at block group level, please provide at census tract level with FIPS code. Please advise if a data sharing agreement is required.*

Important notes

FOIAs responses do not always go smoothly. The government agency you are requesting from may not use the exact terms or categorization mentioned in the FOIA request because of the political discussions around the topic. It might require that you request email exchanges around a general topic and use those emails to then identify what data they could provide to you. If you have developed a trusted relationship with an agency staff, have a candid conversation with them to see if they can provide the data to you outside of a FOIA or advice on how to write up the FOIA. Additionally, if you have relationships with news reporters, they may be able to give some tips on how to write a FOIA because they may have experience doing so.

I. Own data collection process (and guidance on collecting)

There may be data that you want to collect that is specific to your organization's needs that are not available to download. Instead, you'll have to create it! This data can come from municipal plans of the area (which you might be able to partner with the local department of planning to access data files), surveys you've done with your residents, or interviews you've done with specific community members. Some of this data may be spatial so it can be mapped, others are not and will provide important qualitative or quantitative data to understand additional factors that need to be considered when working on the topic or question of interest. The following section shares two examples of how an organization may create and use that data.

How to create survey or interview data:

The Little Village Environmental Justice Organization (LVEJO) and Center for Neighborhood Technology (CNT) collaborated on a study to analyze and understand the impacts of lack of access to drinking water, urban flooding, and COVID-19 in Chicago's Little Village community. This report highlighted disparities in environmental and health burdens experienced by the community. The collaboration included a Water and Health survey component that informs LVEJO of the direct impacts residents face in the neighborhood and will provide data that can be used for policy recommendations and community organizing strategies.

LVEJO and CNT developed questions to gather specific data from residents living in Little Village to analyze how the lack of clean drinking water and urban flooding impact health. LVEJO and CNT intentionally developed the web survey format and types of questions. To inform development, LVEJO and CNT completed training facilitated by the Digital Coast program at National Oceanic and Atmospheric Administration (NOAA) on survey design and implementation. In this training, NOAA staff shared 25 guidelines to strengthen the ability to critique and improve survey questions (which was adapted from Survey Research and Analysis by Jerry J. Vaske).

The Guidelines are as follows:

1. Identify exactly what kind of information you want respondents to provide.
2. Use fixed-scale or close-ended questions rather than open-ended questions wherever possible.
3. Choose simple words and phrases over more specialized alternatives.
4. Use as few words as possible to ask a question.
5. Use complete sentences; avoid cryptic phrases when asking questions.
6. Avoid vague quantifiers if precise estimates can be obtained.
7. Use an equal number of positive and negative responses for scale
8. Distinguish "neither" from "no opinion" in response options.
9. Avoid double-negatives or asking people to say "yes" in order to mean "no."
10. Avoid double-barreled questions.
11. Make every question count.
12. Use an appropriate timeframe for questions and responses.
13. Reduce impact of sensitive or objectionable questions.
14. Use terminology that makes sense to respondents and define unclear terms.
15. Define abbreviations and acronyms that respondents may not understand.
16. Avoid bias from unequal comparisons.
17. Avoid using leading or loaded introductions and questions.
18. State both sides of scales in question stems or introductions.
19. Minimize the number of "check all that apply" questions.
20. Make response categories mutually exclusive.
21. Make sure that each question or statement is accurate.
22. Select questions and responses that permit comparisons with existing information and data.
23. Provide information instead of assuming too much respondent knowledge.
24. Avoid questions that are too difficult for respondents.
25. Use multiple questions to measure complex concepts.

After drafting the Water and Health survey, internal partners and staff provided feedback on which questions made sense and which did not. This first draft proved to be too long and it included questions that provided us with data that would not be used in the final analysis. In revision, the team took survey length into account to encourage complete responses from residents.

External academic partners also provided feedback on qualitative data versus quantitative data. As the survey was being developed, the team hoped to bring on an academic partner to collect more qualitative data. The suggestion was to host focus

groups. NOAA provided another training on best interview practices.

There are different types of interview structures:

- **Fully Structured Interviews:** Interviewer has a predetermined set of questions with the responses recorded in a standardized way. For this process the interviewer is able to conduct the interview with low level skill training. This structure collects qualitative and quantitative data.
- **Semi-Structured Interview:** Interviewer has a set of questions in advance, but can modify their order based on the course of the conversation; wording can change, can omit or add questions. For this process the interviewer is able to conduct the interview with medium level training. This structure collects primarily qualitative data with very limited quantitative data.
- **Open Structure Interview:** In open interviews, the interviewer has a general area of interest/concern but lets the conversation develop organically. For this process the interviewer is able to conduct the interview with high skill level training. This structure collects qualitative data.

However, due to misalignment with the external academic partner the partnership came to a premature end. Note, it is crucial to have flexibility in data collection projects to allow time to adjust and accommodate any unexpected changes.

The team decided to focus only on quantitative data collected via survey.

Our survey included questions such as:

1. **Demographics:** Age range, ethnicity, gender identity, disability inquiry, renter or homeowner status, children under 5 living at home and prefered language.
2. **Challenges with:** Clean water, flooding, pollution, housing, workplace and other. All of these challenges included subsections that included specific questions pertaining to the timeline of when challenges were observed, where these challenges occurred and level of concern.
3. **Support:** Based on our resources and capacity the question identified what kind of support was available for residents to receive.
4. **Contact preference:** First and last name, street address, email and phone number
5. **Geographic information:** Intersection and/or zip code at which the respondent lived or worked at so that we could map data to see where the responses were coming from

Depending on how the data are collected, this information can be mapped in a variety of ways. If the data collected is zip code, then this data would be viewed as zip code polygons (using a process similar to joining ACS data to GIS shapefiles, but with a zip code shapefile). If data collected provides street intersections or identifies addresses, then this data would be viewed as points using the process of geocoding.

The survey included a description of the partner organizations involved, highlighted the purpose for the survey, identified how responses would be used, assured confidentiality, and assured the option to opt-out of answering any questions. It also included an estimated time for the completion of the survey as well as organizational contact information.

EXAMPLE

The mission of Little Village Environmental Justice Organization (LVEJO) is to organize with our community to accomplish environmental justice in Little Village and achieve the self-determination of immigrant, low-income, and working-class families.

The Little Village Environmental Justice Organization (LVEJO) and Center for Neighborhood Technology (CNT) have partnered together to create this Water and Health survey. This survey will help us understand residents' experiences with urban flooding, access to clean drinking water and health, in the Little Village neighborhood in Chicago. The survey data gathered will support and influence LVEJO's policy work. For example, LVEJO could use survey responses to support advocacy to policy makers. The survey takes approximately 5 minutes to complete. All survey questions are optional and voluntary.

All information shared in the survey will be confidential. None of the individual survey responses will be shared outside of the organization and none of it will be used against anyone.

If you have any questions, please contact _____ at _____ or call _____.

Survey Distribution

Traditionally, LVEJO would hand out these surveys during tabling events, door-to-door canvassing, and community meetings, but these efforts could not be arranged due to the COVID-19 pandemic. The team determined that moving forward with a mobile survey approach would be best.

CNT supported the technical aspects of formatting and created an online survey that residents could take on their computers or mobile devices. CNT used the web platform Alchemer, which requires a paid organizational account. You could also consider using Survey Monkey, Google Forms, or another survey platform. When formatting the survey, CNT considered the types of questions the survey asked and the type of response desired. (Here is a link to the different question types on Alchemer). CNT also thought about what was the flow of questions and answers (also called skip logic). For example if a respondent responded with choice A instead of choice B, they would see different future questions. This way the questions that the respondent saw would directly relate to previous responses to reduce asking irrelevant questions. This improves user experience. Further, questions and answer choices were offered in both Spanish and English since the community includes a large Spanish-speaking population. After putting together the survey, LVEJO beta-tested it with staff and family, and then the survey was revised for distribution.

LVEJO connected with local organizations and health institutions that could share the survey to their bases to support survey distribution. COVID-19, however, proved to be a more significant challenge. We all had to deal with unprecedented events and were unable to conduct outreach in these local spaces. However, we were able to include an incentive component in the survey. All residents who completed the survey and provided their contact information would get mailed a $25 gift card to compensate for their time. We hoped that this would incentivize more residents from Little Village to take the survey. After the initial roll out and low response rate, LVEJO opted in to conduct heavy phone banking outreach to our local base.

Distributing surveys along with other information during community events can help improve the response rate. This image was from a recent community meeting hosted by LVEJO.

Photo by eduardocornejo.com

After the survey was closed, CNT analyzed the data using Excel. The data were cleaned to exclude responses not from within the Little Village community (Alchemer provides information on the nation/state that the respondent filled the information from). Then, CNT analyzed which zip codes were represented. Afterwards, CNT created tables showing the counts in responses to the different questions and created corresponding graphs to identify what issues mattered most for residents. These tables and graphs were used to update the Water and Health report.

Surveys can also be informative for counter mapping, which is detailed below.

Counter mapping

Counter mapping is an approach to using mapping to challenge or contradict the way that we commonly perceive the world around us. The act of creating maps that are counter to common practice and challenge power structures has helped social justice and environmental justice organizations to advocate for new ways of thinking, to visualize community held knowledge, and to promote underrepresented worldviews held throughout culture and history. Counter mapping is a large and evolving field of mapping, in this section we will introduce the concept of counter mapping and some ideas on how to incorporate this tool into the mapping products that you create.

Counter mapping highlights community assets, like this community planting event hosted by Milwaukee Water Commons.

Photo Credit: Pat Robinson

Conventional approaches to mapping are built around established categories for understanding places and interpreting data. For example, a common practice when mapping a neighborhood is to begin by mapping properties and categorizing locations based on ownership. In this approach you might start by identifying commonly understood assets such as public parks, private businesses, residential properties, city owned properties, etc. However, counter mapping would challenge that approach by utilizing an understanding of the neighborhood and its resources built from a grassroots level to categorize different locations. This could be based on things like unique perspectives, cultural norms, or even a moment in time. To build a more grassroots level of understanding of neighborhoods, the section above on surveys is one starting point for getting the community input described below.

Here are some ways that this might look different than traditional data collection for mapping within a neighborhood.

- **Identifying community resources:** Oftentimes while mapping locations researchers will use commonly held categories to define neighborhood resources. For example if you are interested in looking at public health resources available to a neighborhood you may start by mapping the locations of hospitals, medical clinics, fitness centers, and places that sell healthy foods. Approaching this question through counter mapping could mean beginning by asking residents what neighborhood resources they find benefit public health. In this approach you may find that hospitals or medical clinics are avoided, that grocery stores are inaccessible or that there are nontraditional locations like parks, corner stores, or other gathering places that are perceived as public health resources. By mapping these locations you are capturing community held perspectives and knowledge about what resources exist in the neighborhood.

- **Identifying community values:** Traditional mapping approaches often assign categories based on conventional power structures that under-appreciate or entirely leave out community perceptions of the geography being mapped. At a neighborhood level what a researcher might identify as a vacant lot may be viewed as a growing space or a public park, it also may still hold historical relevance for residents in that neighborhood because of what was there before. By amplifying these values you can design a map that is counter to conventional power structures.

In the example of Countermaping Sherman Park, students and community organizers helped to gather stories held by residents in Milwaukee Wisconsin's Sherman Park Neighborhood to identify landmarks that defined the neighborhood for residents currently living there, and mapped those landmarks along with the associated stories held by residents to capture a moment in history, the perception of Sherman Park as it was held by people living in the neighborhood.

6

DATA MAPPING AND ANALYSIS

In the Data Gathering chapter, we cleaned, prepared, and visualized data in QGIS. This chapter will talk about how to analyze the maps in order to help us answer the research question we are interested in.

This chapter includes several subsections:

A. Creating Thematic Maps - We walk through how to prepare your map using symbols, colors, and labels to visualize the differences in data values across locations (Milwaukee in this case).

B. Notes on Scales & Distribution - We talk through how to present the data responsibly, paying attention to techniques for grouping data values into different categories and how you present the data visually.

C. Presenting the Data Lite - We provide a high-level overview of multiple methods that you may try to compare locations within the map and explain additional analysis steps you might want to take to make your point.

D. Creating a Final Map and Interpreting the Map - We walk through how to create a printable map in QGIS, and provide pointers on how to write the text that analyzes and accompanies the map.

E. Additional Tools in GIS - We identify some commonly used tools in GIS analysis and provide links to tutorials on how to do that analysis.

F. What Else Can You Do with GIS - We provide a brief overview of 5 additional types of GIS analysis that your organization may be interested in learning more about in future GIS collaboration projects.

The specific QGIS skills with how-to's you'll learn in this chapter include:

- Symbology
- Labels
- Classification of Data using Breakpoints
- Normalization
- Summary Statistics
- Adding Legends

The following QGIS tools are shared with links to tutorials:

- Clip
- Intersection
- Buffer
- Merge Vector Layers
- Join Attributes by Location (Spatial Join)

The follow GIS analyses are talked about with examples or links to tutorials:

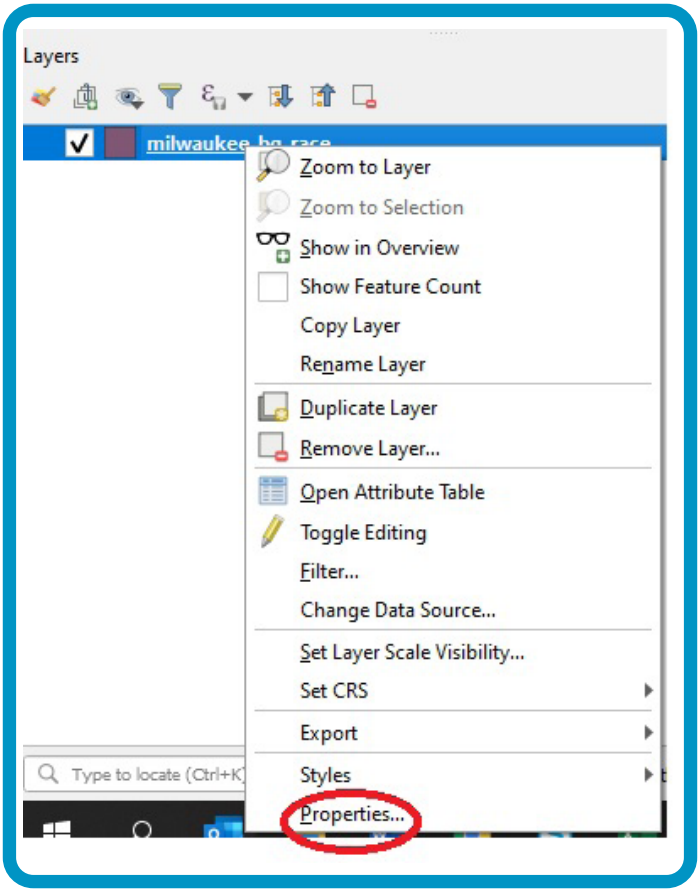
- Geocode
- Index
- LiDAR Analysis
- Spatial Analysis
- Community Mapping

A. Creating thematic maps

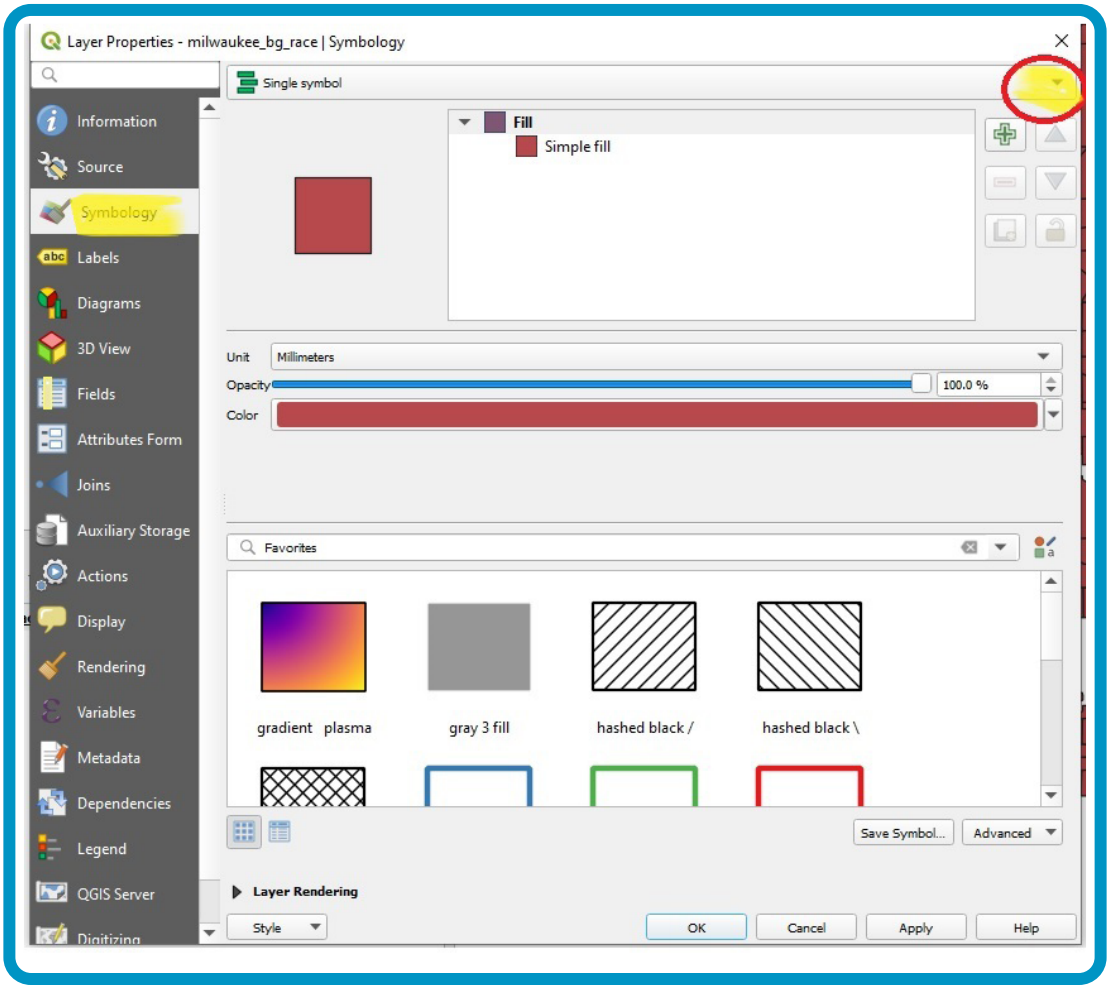
Symbology

Once the data has been joined to the shapefile, the resulting map layer can be visualized. In this section we will discuss how to visually display your data using tools like symbols, colors, sizing and labels. Symbology refers to visually displaying differences in data values between locations through colors, symbols, labels, and more. First, we will explore the available options in the "Symbology" tab. For this illustrative step, we are using race/ethnicity data at the census block group level.

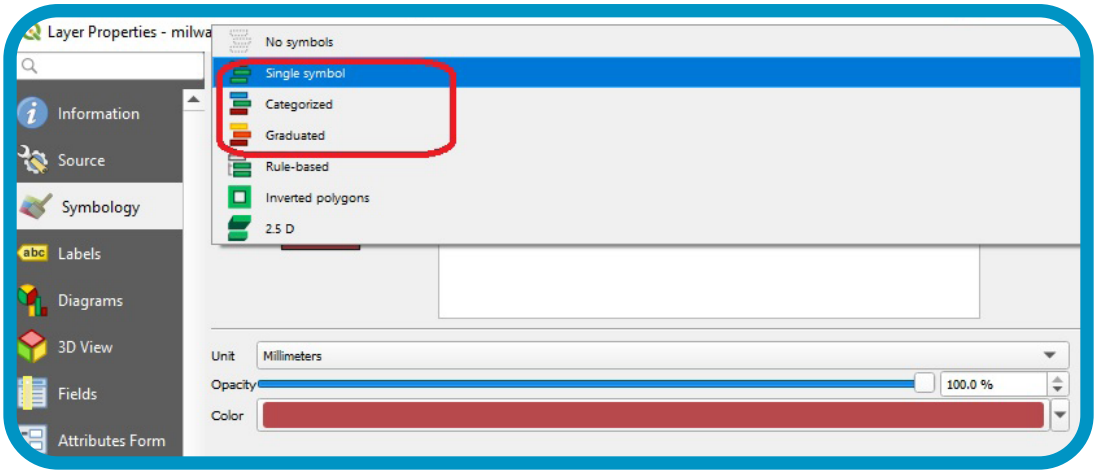
1. To get to the symbology, right click on the layer, and choose "Properties" (or double click the layer).



2. Click on the Symbology tab on the left side of the Properties window (highlighted in below image). By default, a polygon layer is set to a simple fill option, i.e., all the polygons, in this case, census block groups, within the layer have a single color and an outline. Next, click on the drop-down menu on the top right – the one named "Single symbol".



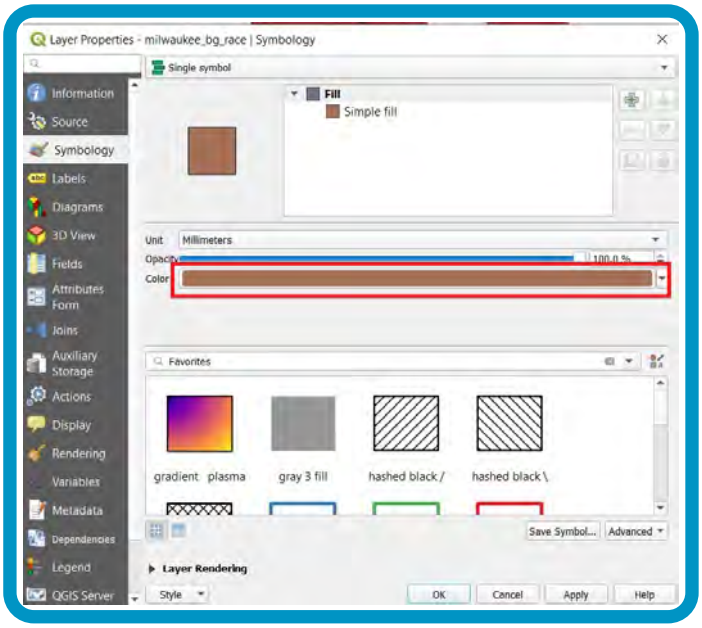
3. The drop-down menu lists the options available to assign themes to the map layer. In this toolkit, we will focus on three options: Single symbol, Categorized, and Graduated. A brief overview of the three symbology types is presented below and a more detailed description follows.

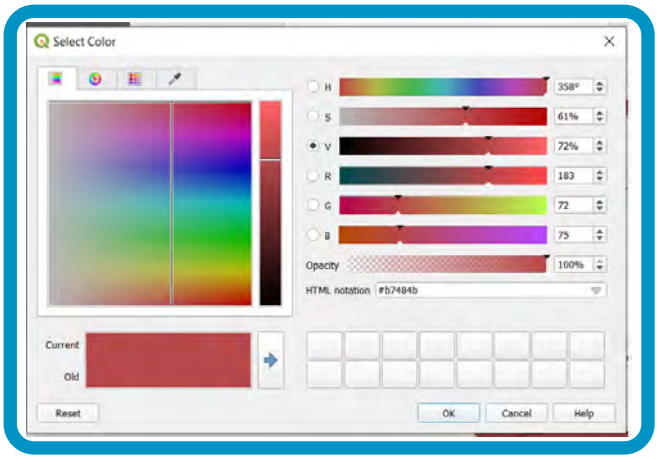


Single Symbol	Categorized	Graduated Colors
All the polygons in the layer are assigned the same color and outline. This is best suited for layers with one polygon (e.g., city boundaries, study area boundaries) or for layers that highlight areas of interest.	Polygons belonging to the same category within an attribute are grouped together. This option works well for qualitative data categories – e.g., land use, zoning	Polygons are grouped together based on quantitative data ranges of a single attribute in the data layer. This symbology option is suitable for demographic data, employment data and other quantitative data sets. **Note:** The attribute's data type must be in a number format, if the attribute is a number but the data type is Text or String, it will not show up as an option.

Single Symbol

Selecting "single symbol" assigns a single color to all the polygons in the data layer. The color is assigned to polygon boundaries and not to specific attributes. The default color choice is a simple fill (same color all over). To change from the default color, click on the color bar (highlighted in the image below). A "Select Color" box pops-up, which will allow you to choose from all available color options. Once the color has been chosen, click on the "Apply" button to finish the symbology selection.



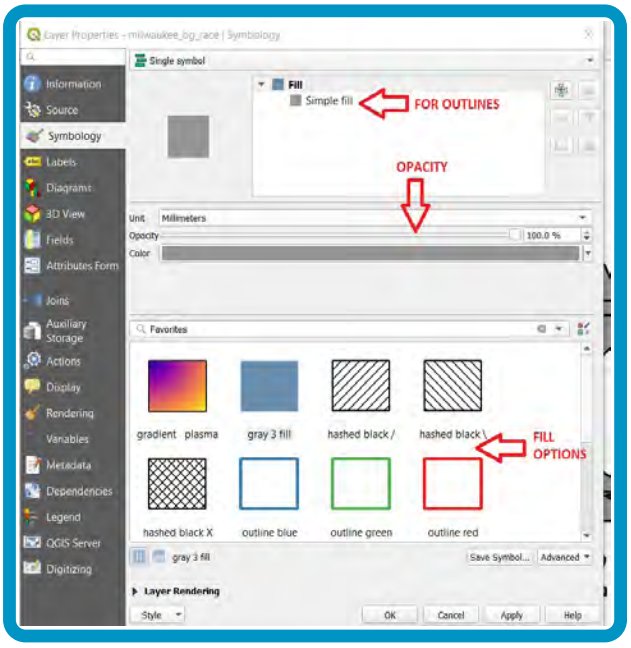


A few additional options to manipulate the data visualized include setting opacity/ transparency, changing outline colors and widths and additional fill options (gradient color, pattern, etc.). They are briefly explained below.

Opacity: By default, opacity is set at 100%. To make layers transparent, decrease the opacity to the suitable level (0% being fully transparent). This option comes in handy when multiple layers are overlaid and makes the map easier to interpret.

Fill options: The default fill option is a simple fill (same color throughout the polygon). Additional options can be found in the bottom half of the Symbology window. These include 1) Gradient 2) Hashed and 3) Patterns.

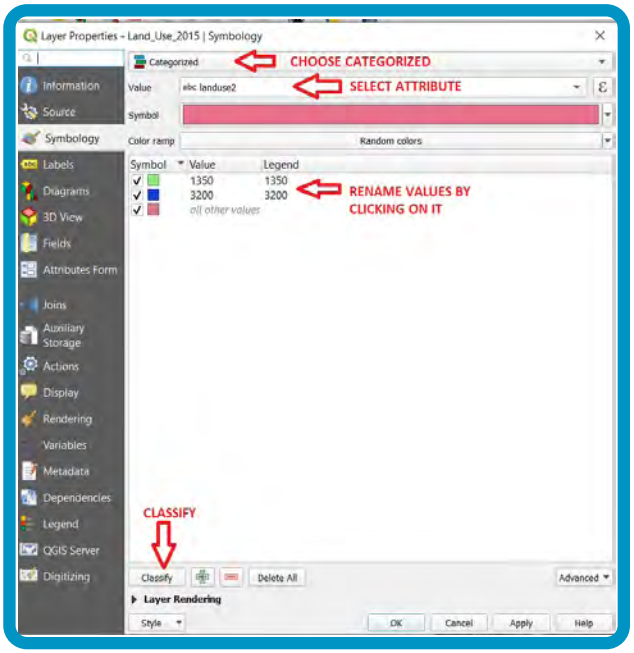
Outlines: In some cases, you may desire to change the appearance of the polygon outline, for example from a solid line to a dotted line. To do this, click on the "Simple fill" button, and select stroke color, stroke width and stroke style. Click "Apply" to retain the changes.



Categorized

The categorized symbology option is suited for datasets that include features in multiple categories that are not numerically ordered, such as land use data. Attributes in land use datasets are typically categorized as residential, commercial, industrial, etc.

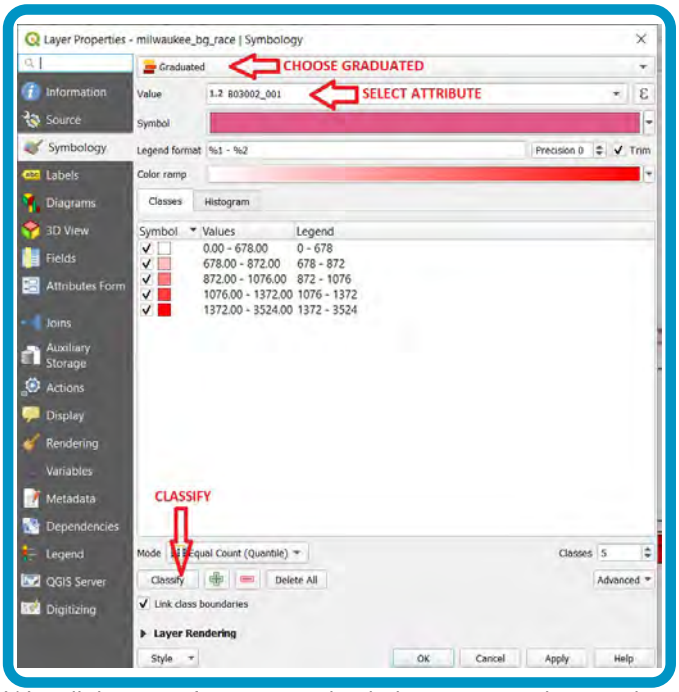
For example, if a land use layer was available, add it to the map, and navigate to "Symbology" in the "Properties" tab. Here, choose the "Categorized" option, and in the "Value" dropdown menu choose the appropriate attribute that contains the land use categories, and then click on "Classify". If certain features do not have categories, keep the "all other values" checked or else uncheck it. You can either click "OK" and be done or rename the values by clicking on it in the legend column. Finally, click "OK" to see the created map.



Graduated Colors

The graduated symbology is appropriate to use for representing quantitative data ranges. Maps using this symbology are also often called choropleth maps.

To assign this symbology to a data layer, navigate to the "Symbology" in the "Properties" tab. Choose the "Graduated" option from the drop-down menu, and then choose the appropriate attribute from the "Value" drop-down menu. Click the "Classify" button on the bottom left of the window and then click "OK". The map is created, and the census block groups are assigned one of the five class colors.



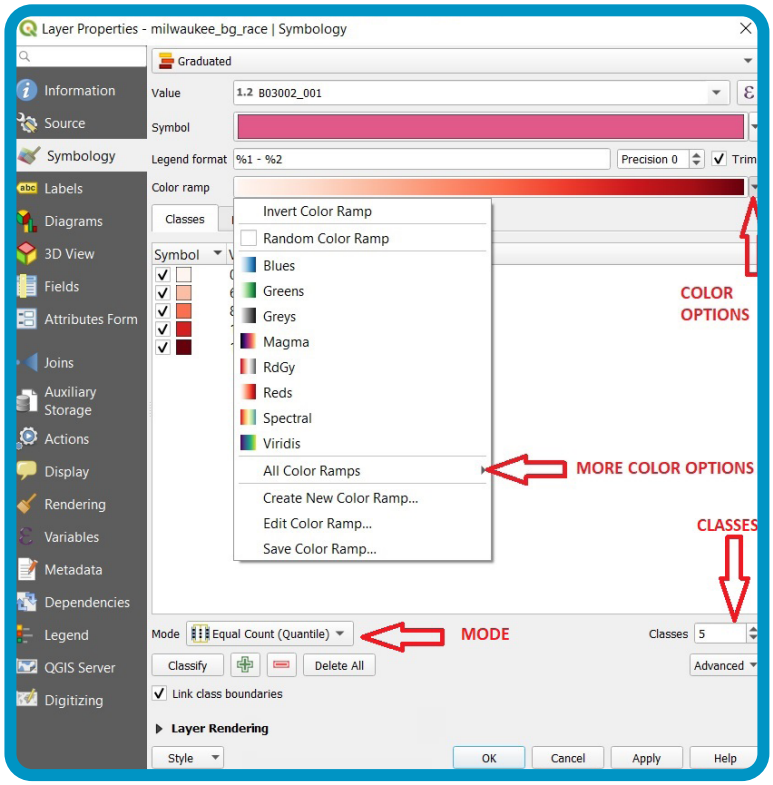
We will dig into a few options that help customize the way data is presented. Navigate back to the "Symbology" tab to explore how to change the colors, modify the classes (or number of groups the data is broken into) and explore different classification methods.

Colors: To change the colors, click on the down arrow next to the color ramp and choose from a preset selection of colors. To choose from a full selection of available options, click on the "All Color Ramps" button to view all.

Classes: The number of classes is equal to the number of groups the data will be broken into. The number of classes are chosen according to how the data is distributed and based on the analysis methods you choose. Typically, between 3-6 classes are most effective, anything larger than 6 can get difficult to interpret visually.

Mode: There are six available methods to classify quantitative (numeric) data in QGIS (To learn more about the classification methods, refer to classification section):

- Equal Count (Quartile)
- Equal Interval
- Logarithmic Scale
- Natural Breaks (Jenks)
- Pretty Breaks
- Standard Deviation

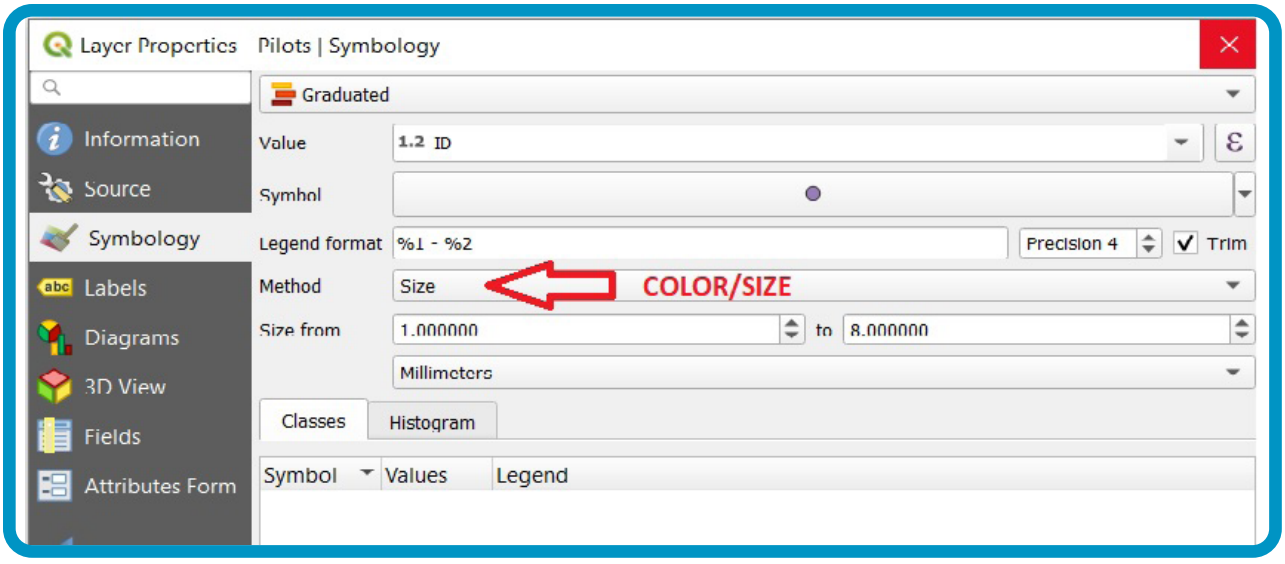


There are three other symbology types available in QGIS that are not covered in this toolkit: rule-based, inverted polygons and 2.5 D. This resource provides more information on these three styles and how to apply them.

Other symbology types

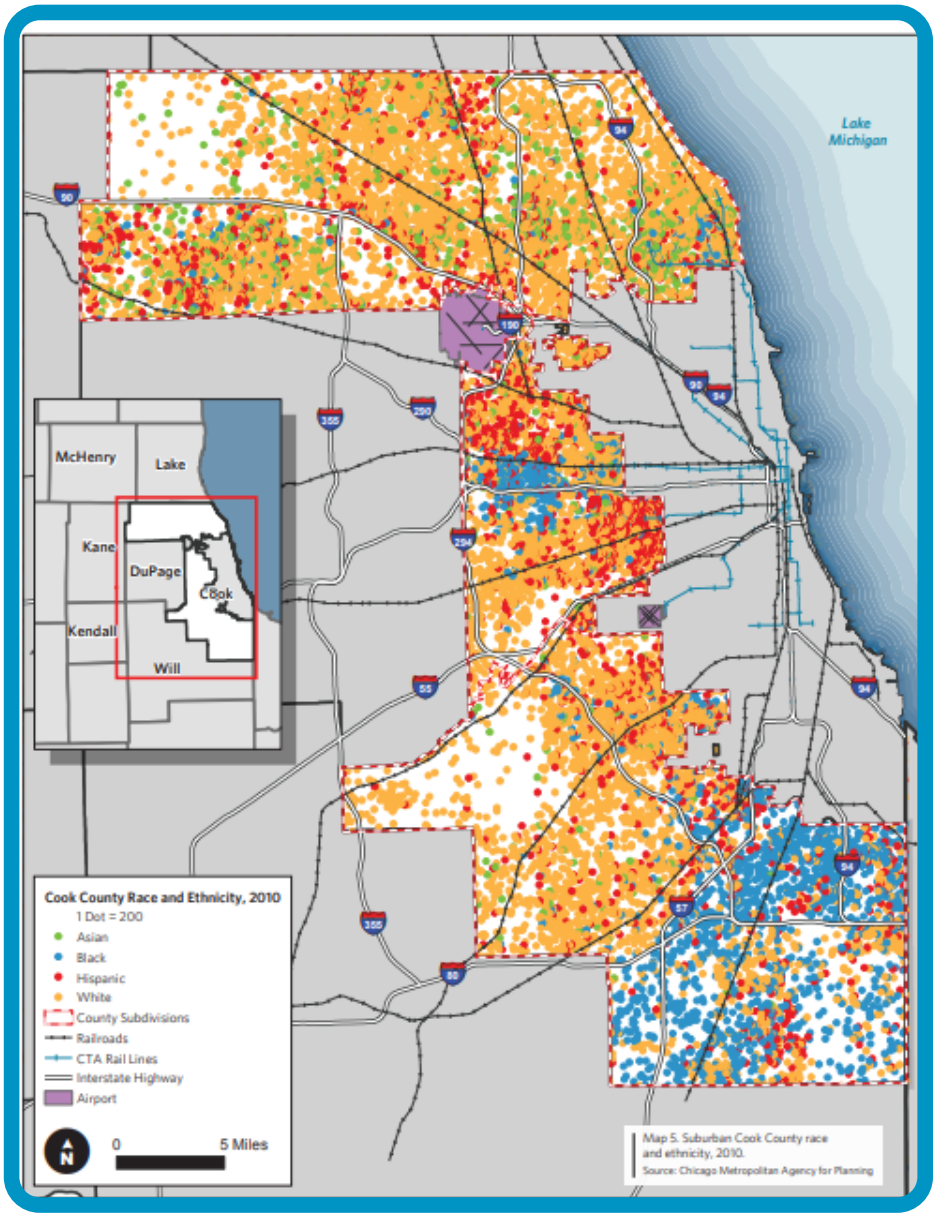
Symbology for Point and Line Layers

The above discussions used polygons for illustrating the different styles available to create thematic maps. For point and line layers, similar options are available and similar steps can be carried out to create maps using single symbols, categorized and graduated symbology. In the "Graduated" symbology type, in addition to colors, the attribute can be displayed by size as well. In the Milwaukee example we have been using for this toolkit, you will not be able to display attributes by size. To practice displaying attributes by size you will need to use a different data set.



Dot Density

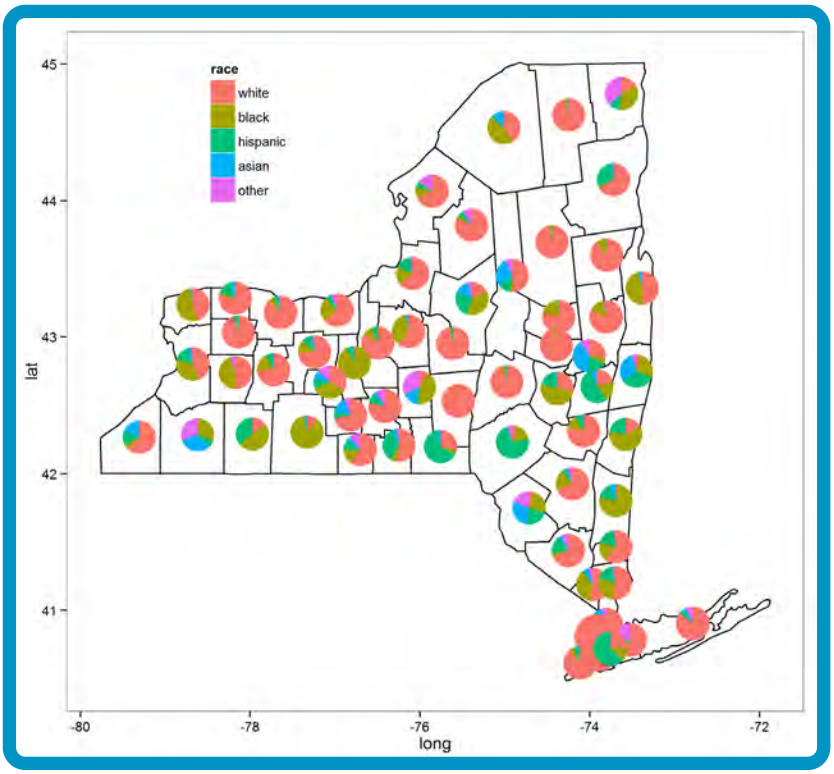
Dot density or dot maps visualize data as single points of uniform size in a map. The point can follow a one-to-one format where one dot represents one event, person or incident, or a one-to-many format wherein one dot represents many people. The location of the dots on the map are distributed based on probability and are not actual locations of the observations. Dot density maps are commonly used to show the racial distribution of population across a geography. A detailed description on how to create a racial dot density map can be accessed here and an example of a finished map is presented below.



Source: Department of Planning and Development. (2015). Planning for Progress: Cook County's Consolidated Plan and Comprehensive Economic Development Strategy, 2015-19. Cook County Bureau of Economic Development. https://www.cookcountyil.gov/sites/g/files/ywwepo161/files/planning_for_progress_plan_0.pdf

Pie charts

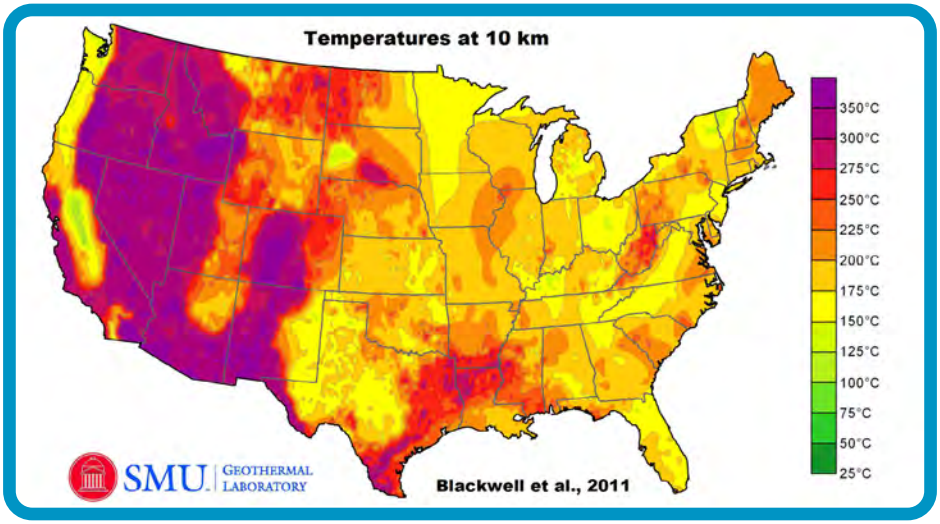
Pie charts can be used to illustrate the numerical proportion of data related to a single attribute. For example, the proportion of tree canopy, paved surfaces, and bare earth in each county in Minnesota. Pie charts are better suited when a small number of geographical components are being mapped. To learn how to create pie charts in QGIS, refer to this resource.



Source: igor. (2017). How to plot haplogroup map with pie charts? Biostars. https://www.biostars.org/p/212408/

Heat maps

Heat maps create hotspots by clustering point data to show trends in the data by location. This symbology is best suited for dense point data and can translate a crowded map into one that groups and emphasizes areas with higher occurrence of the data. This tutorial provides step by step guidance to create heat maps.

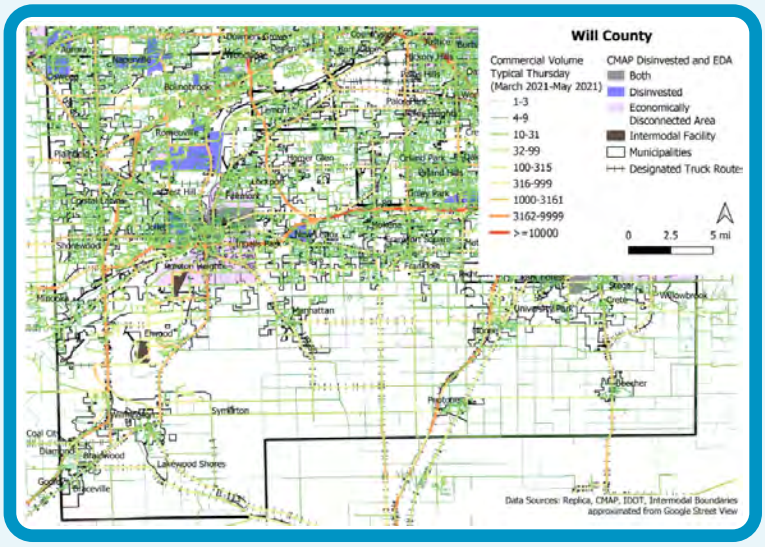


Transactions, 35(GRC1029452). Accessed https://www.smu.edu/Dedman/Academics/Departments/Earth-Sciences/Research/GeothermalLab/DataMaps/TemperatureMaps

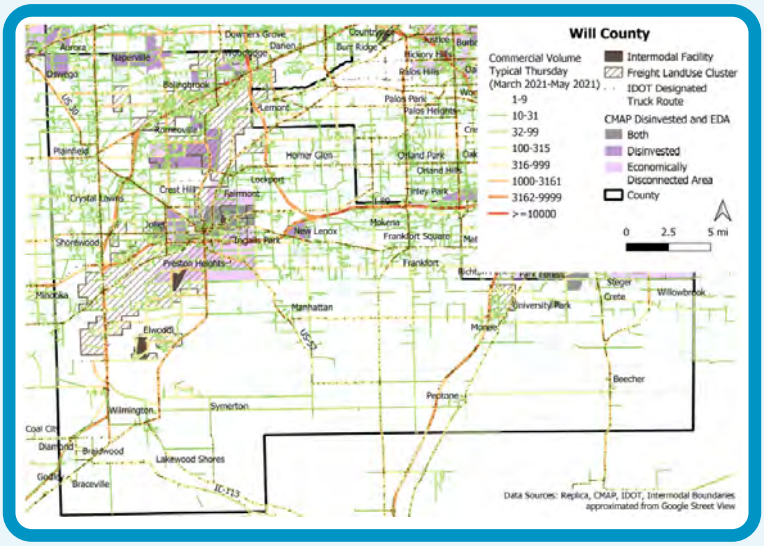
TIPS TO CREATE WELL-DESIGNED THEMATIC MAPS

1. Avoid too many layers: Adding too many layers to a map can make the map difficult to interpret for readers and can fail in conveying the data presented. The context of the map and the data should inform the number of layers to visualize, but more than 3-4 layers for the same geography can make the map difficult to read.

2. Dealing with overlapping layers: If you have multiple overlapping layers, try to change some layer symbologies to outlines only, a hashed fill, or modify the opacity of layers. In the event none of the above visualization techniques work, consider splitting the maps into multiple maps. Below is an example of a crowded map and an easier to read map.



Busy map

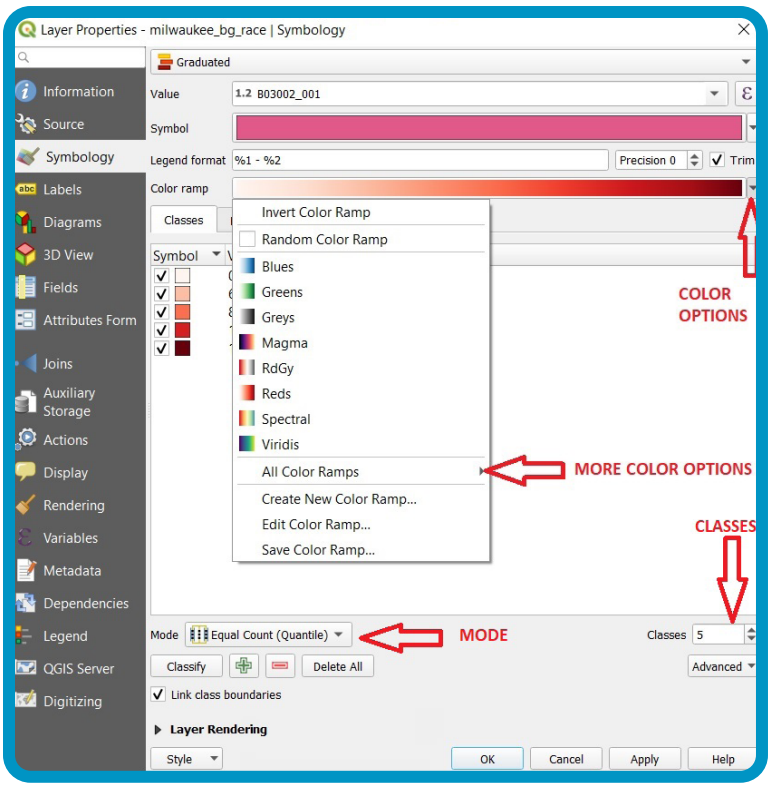


Clearer map

3. Color schemes: The right color schemes can better present the issue and help the reader to make some high-level deductions from the map. QGIS has several preset color options, but resources such as Color Brewer can assist in selecting the right color combinations for your mapping needs. The resource also provides guidance on accommodating colorblind and printer friendly options.

Labels

Labels can be added to maps to show information about data layers and orient readers to the map by providing context. To add labels, right click on the data layer, and navigate to the "Properties" tab. Click on the "Labels" tab on the left and choose "Single labels". Select the corresponding attribute name in the "Values" field and choose desired font and font size. The font color and opacity can also be modified.A. Creating thematic maps



B. Note on Visualizing Data through Maps

How can data be misconstrued

While creating maps, it is important to ensure the data is being presented accurately. In addition to ensuring the correct attribute is being mapped, classification methods, color choices and labeling can influence how a map is interpreted. The list below enumerates a few ways the data visualization can result in misinterpretation.

Classification: How the distribution of numbers within a dataset are classified (or grouped into buckets) makes a huge difference in what the map looks like. It creates different stories about where the impact is. See the next section to learn about different classification methods to avoid misinterpretation.

Color choices: The right color choice bolsters the story being conveyed but poor color choices can have the opposite effect. Pay attention to the range of data and assign darker colors to groups which represent a greater magnitude of the issue.

Labeling: Labels should provide details of the data being mapped. For example, when representing percentages, it is important to clearly state

what the numerator and denominator are like Percent non-Hispanic White residents of total population.

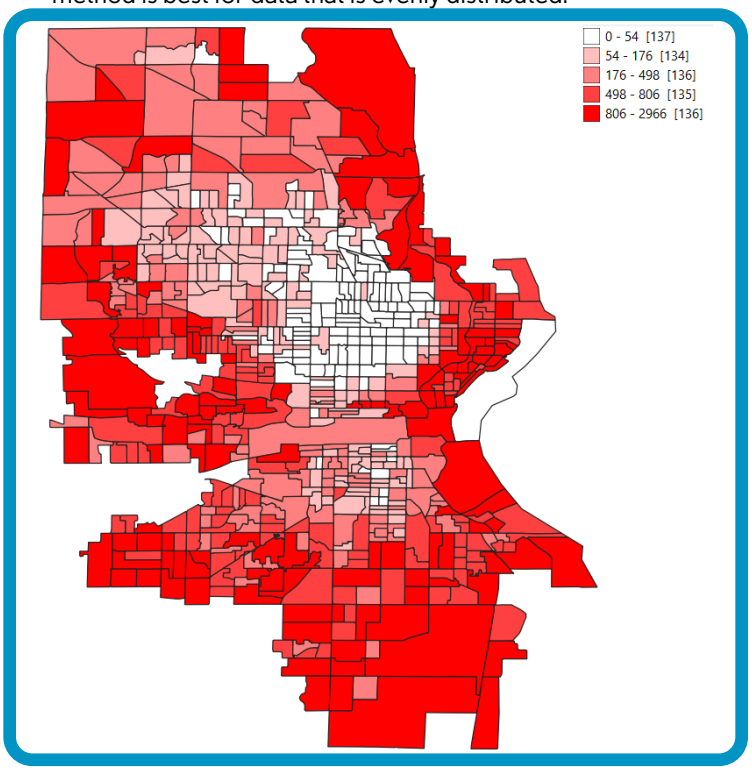
Break points: Inappropriately manually choosing break points (where the map creator determines the different categories not using the preset options) or choosing an inappropriate classification can misrepresent the underlying data by highlighting patterns that are not real, telling a fake story to the map viewers.

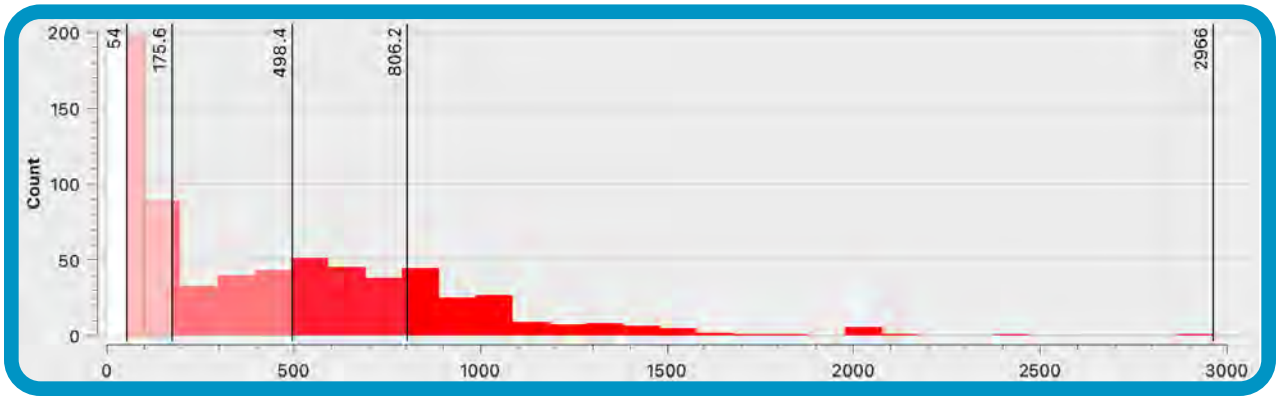
Classification of Data using Breakpoints

Classification of data is the process of grouping similar data features and assigning a symbol/color to each group. Grouping data reduces the number of symbols/colors displayed, making your map easier to comprehend and making patterns in the data distribution more apparent. Breakpoints are the values at which one group ends and another group begins.

In QGIS, there are six classification methods that are discussed briefly below. The classifications are described in order of how they are listed in QGIS. For this illustration, we visualize the count of residents who identify as non-Hispanic White at the census block group level within Milwaukee. Provided are map examples for classifications and histograms (a chart similar to a bar graph) to show how the data is distributed.

1. **Equal Count (Quantile):** The number of features within each group are approximately the same. This classification method arranges the features in ascending order of their values and creates a set number of groups (in this example 5) in a manner that each group has the same number (or count) of features, in this case census block groups. In the map below, there are roughly 136 census block groups in each classification. Even though the number of features (census block groups) is equal within each classification, the total number of non-Hispanic White residents represented in each classification will vary considerably. This classification method is best for data that is evenly distributed.

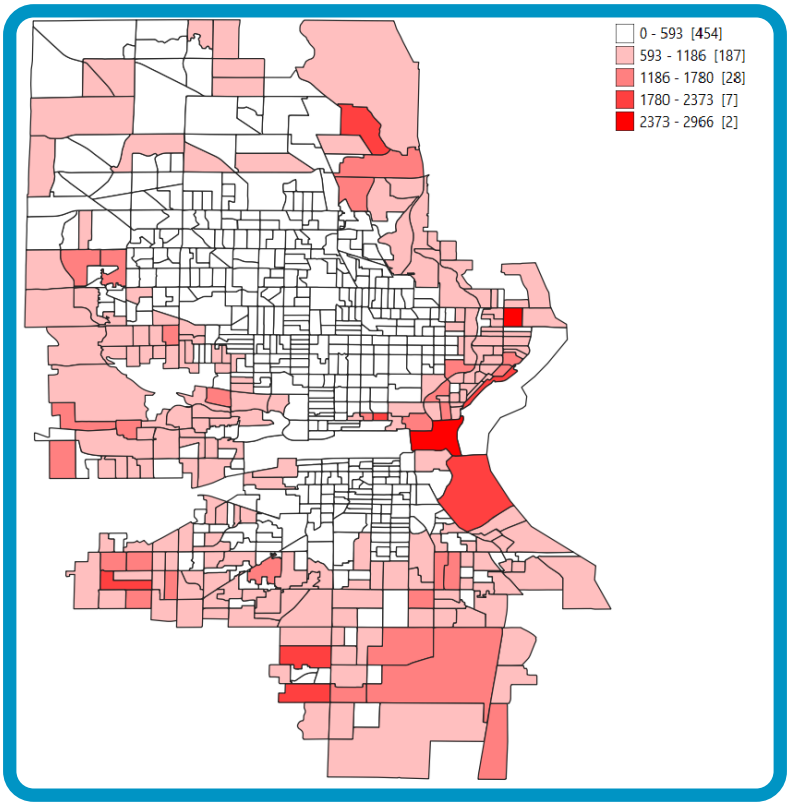


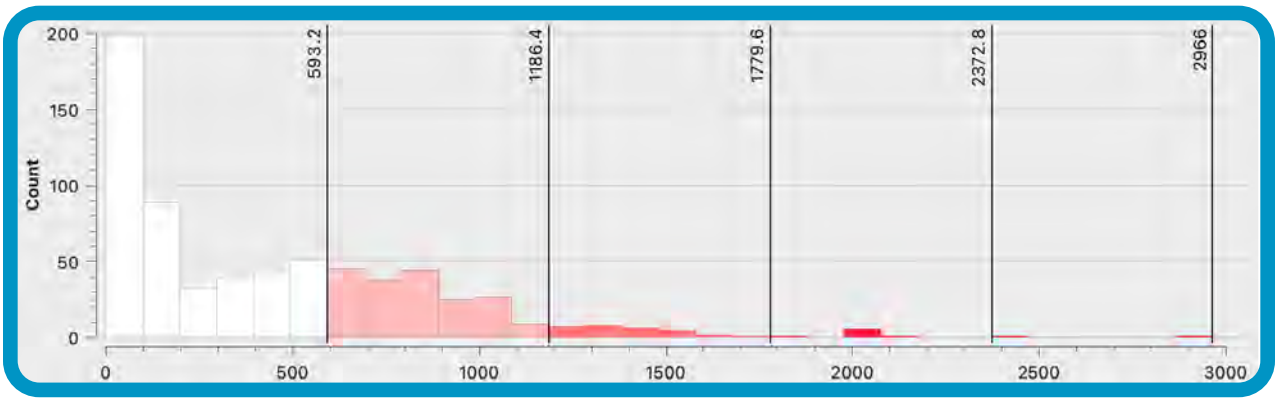


NOTE

How to read a histogram: The histogram above (along with all the histograms that follow in this section) shows how the data was broken up into the different colored groups represented on the map. The x-axis shows the range of data values for the variable of interest, for this example it is the number of non-Hispanic White residents. The y-axis, labeled as ("Count"), shows the number of features that have the value listed on the x-axis. In this example it is the number of block groups that have X amount of non-Hispanic, White residents. The black vertical lines show the boundary values for each color, otherwise known as a "breakpoint." For example, the block groups colored white have between 0-54 non-Hispanic, White people in them, and the block groups colored light pink have between 54-176 non-Hispanic, White people in them. The deepest red, last group, has block groups with between 806 - 2096 non-Hispanic, White people in them. The breakpoints are unevenly distributed across the data range, because the data is broken into groups with an equal number of block groups.

2. **Equal Interval:** The range of numbers for an attribute is divided up into equal portions. This is different from equal count where each group has the same number of features. The extreme classes (higher range groups) will have fewer features as illustrated in the image below. The highest two groups have a total of nine features whereas the first group has 454 features, but the total number of non-Hispanic White residents are about the same in each group. This classification method is best suited when displaying familiar data ranges, such as percentages or temperature.



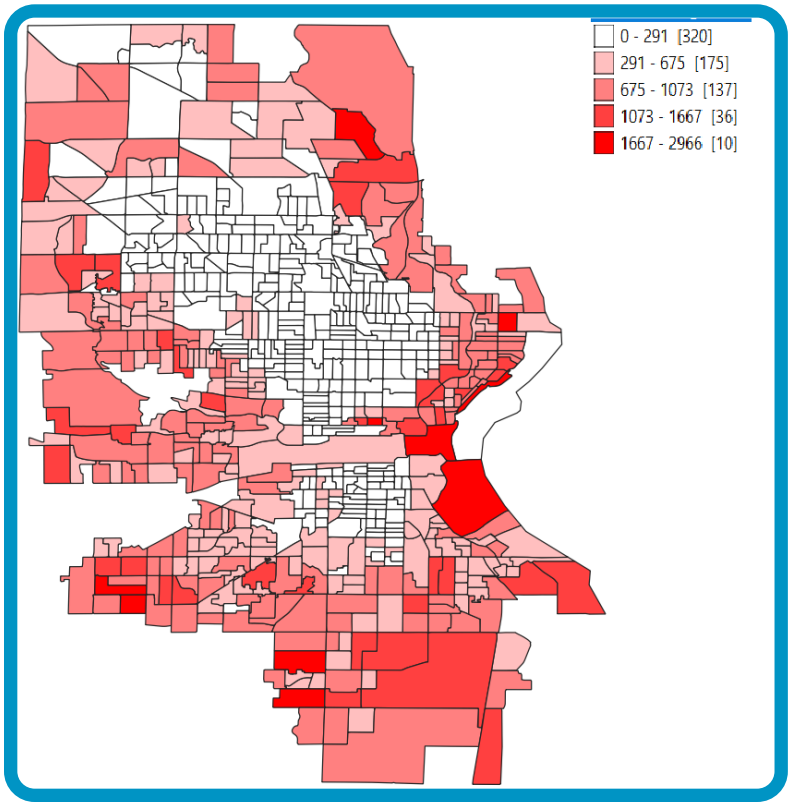


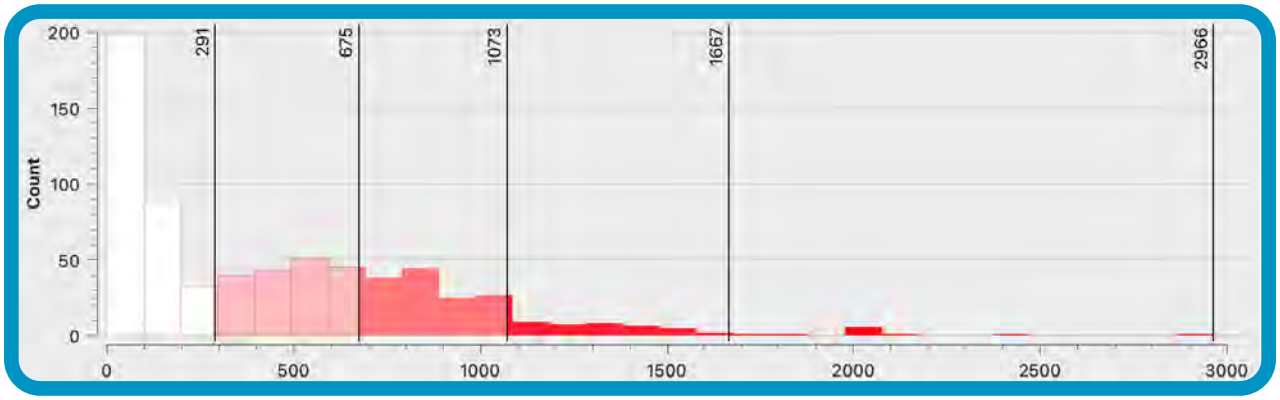
Here, you see that the breaks are spread out evenly across the histogram. This is because for equal interval classification, the range of data is broken up into equal segments. You can see that some groups have many features while others have very few because the groups are split according to the data values rather than the number of block groups.

3. **Logarithmic scale**: The logarithmic scale is useful when displaying data that has a very large range (includes very small data and very large data) and is very skewed (so despite its large range there are a lot of features on one of the extremes) or has lots of zeros. For example, if the data covers population across rural and densely populated cities, logarithmic scale can identify breakpoints that help distinguish different levels of density. We do not share an example using the non-Hispanic White from the Milwaukee data because the range of values within the Milwaukee data would not be accurately represented with a logarithmic scale, given the data is not extremely skewed nor is the data range too large.

4. **Natural Breaks (Jenks)**: The natural breaks classification creates groups wherein the features within a class have values that are more similar to each other and the different classes are more distinct from one another. In more technical terms, the

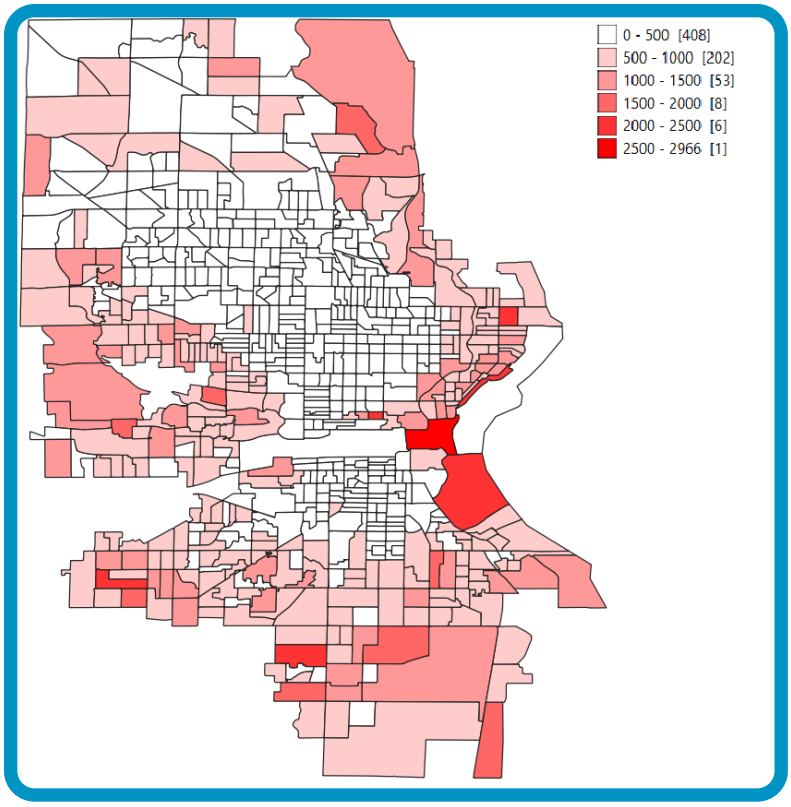
average deviation of variables from the mean within each group is minimized (so they are more similar to one another), but the deviation from the means of other groups is maximized (so each class is most different from one another). This method is not well-suited if there is low variance (not a large range) in the data being visualized. In other words, if the values in a dataset are fairly similar to each other, then natural breaks are not the ideal classification method. This method is also not ideal for comparing across multiple maps as the classification is specific to the dataset. Natural breaks are best for data that is not evenly distributed, such as population density in a city, where there are outliers of much higher density.

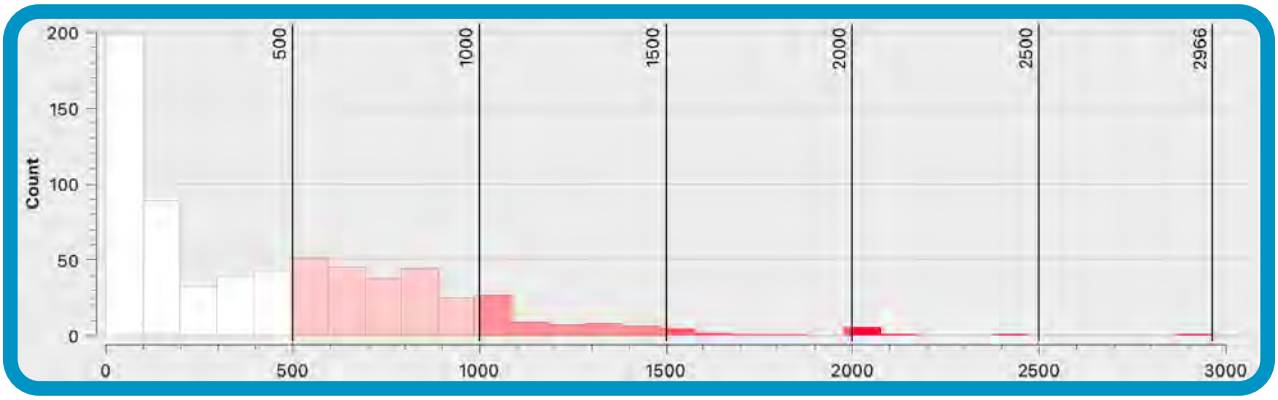




Whereas equal count (quantile) showed breaks that reflected the skewed distribution of the data (more vertical lines to the left) and equal interval showed breaks that were equally spaced across the data range, this histogram shows a combination of those two methods. The breaks somewhat reflect the skew of the data values but they are also spread across the data range to create groups that are not clustered towards lower values only.

5. **Pretty breaks:** This simplified classification method creates equally spaced classes based on rounded numbers. For example, if the data range is 0-100 and it is broken into 4 classes, they would be 0-25, 25-50, 50-75, and 75-100. It is important to note that the number of classes are dependent on the range of data and cannot be set by the user. Pretty breaks may be most helpful when dealing with very large numbers, like population estimates for entire states, for a simpler reader experience.

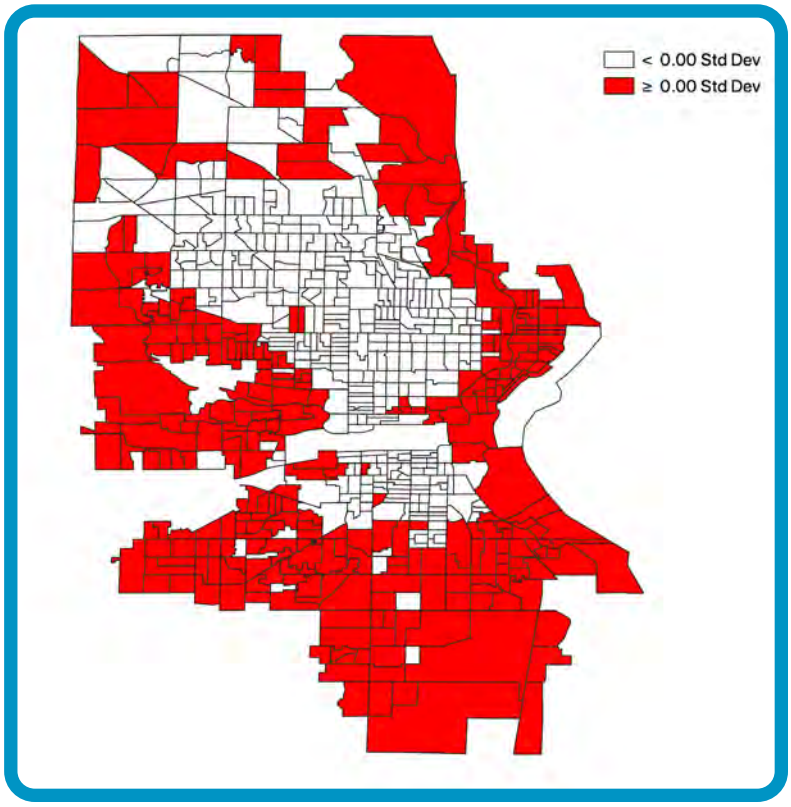


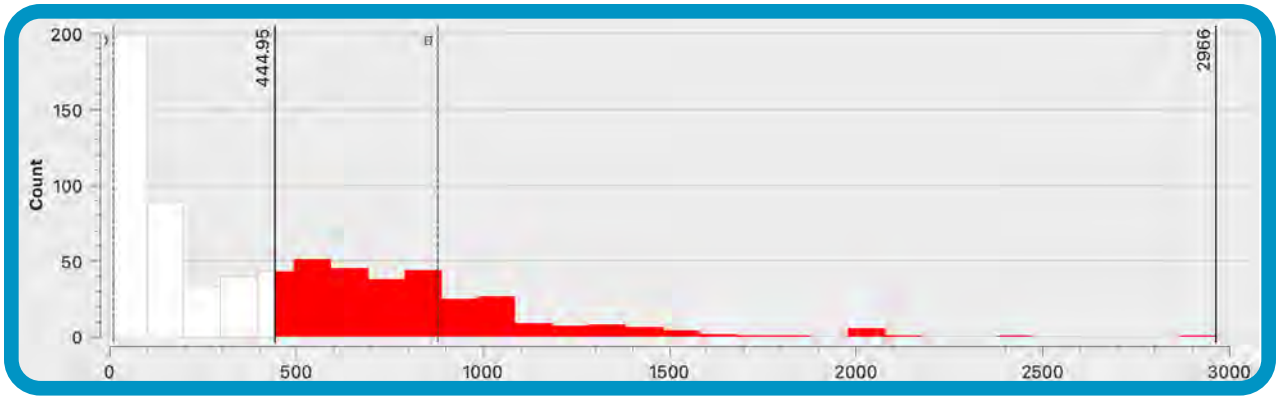


You may notice that the vertical lines are distributed in a similar manner as they are in equal intervals (evenly spaced out across the data range). The largest difference between the histogram for equal intervals and for pretty breaks is that the breaks are determined by rounded numbers, so all of the breaks end in 0. The groups are evenly spaced out as in equal intervals, but in pretty breaks the groups are determined by easy to understand numbers (in this case, spaced out by 500) rather than by the data range itself.

6. **Standard Deviation:** Standard deviation is the measure of how much the data differs from the mean. Using this method groups are created based on the number of standard deviations away from the mean the data within it is. This classification method is best for seeing what features are above or below average for a dataset with a normal distribution (when extreme values will not skew the mean).

The histogram above shows how the data was broken up into the different colored groups represented on the map. In this case, there are only two groups, one that is positive or 0 standard deviations and one that is negative standard deviations. The line shows where these groups split, which is at 445.





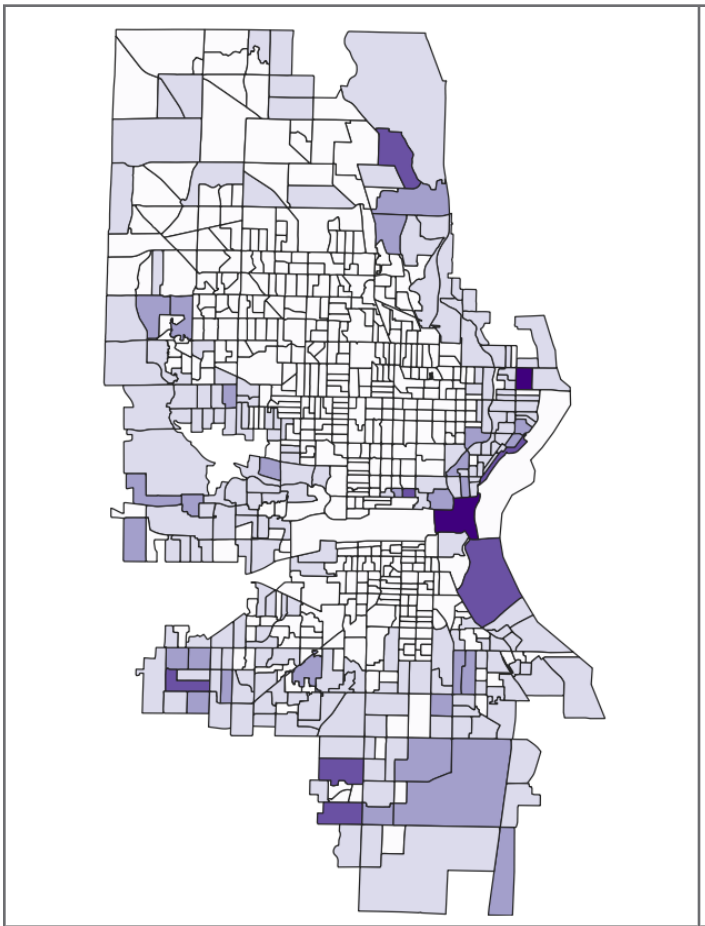
C. Ways to Compare and Present the Data

When presenting data, it is important to make sure your audience is able to make accurate comparisons. In this section, we will provide an overview on normalization, summary statistics, aggregation, weighting, and disaggregation. The methods presented in this section are ways to ensure that the maps created allow readers to compare locations accurately.

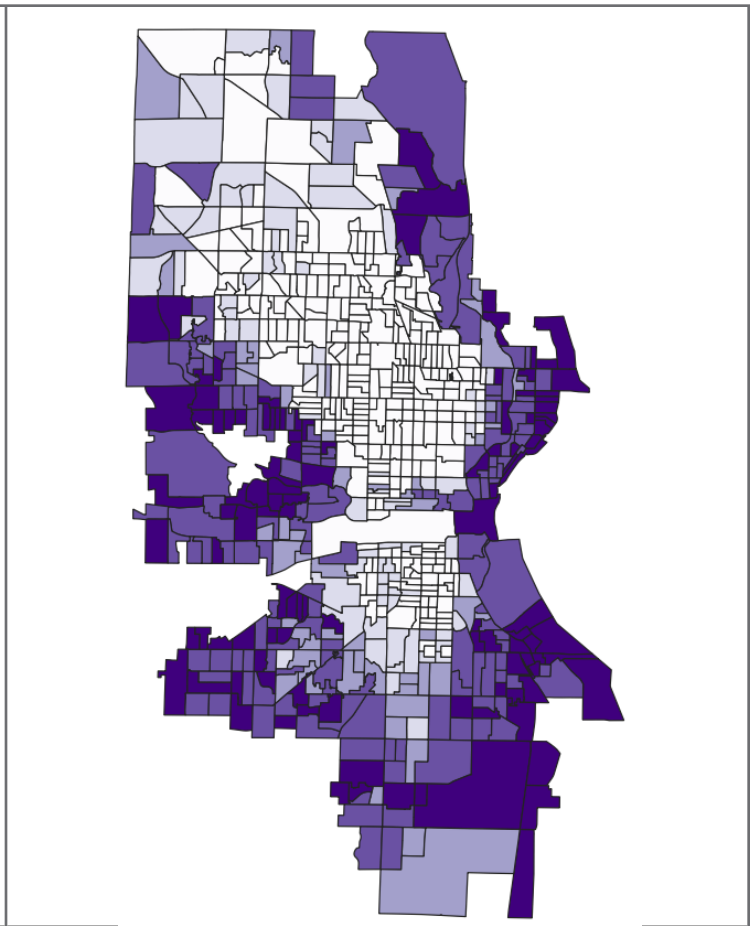
Normalization

Mapping and analyzing data to advocate for environmental justice often involve comparing levels of burden between two or more locations or groups. A core requirement of an accurate comparison between groups is to first have the data values for all locations in the same unit of measurement. Boundaries such as census

block group, zip code boundary, or state boundary may or may not be appropriate for comparing certain metrics because parameters such as area and population vary. For example, when we want to compare the population of white, non-Hispanic people in block groups, if we only provide a count of people who identify as Hispanic per block group, readers may not have an accurate representation. Below are two maps of Milwaukee's white, non-Hispanic block groups. Though they both reflect data about the same thing, they visually look different. The one on the left makes viewers think that there is a high white population in only a very few block groups throughout Milwaukee, whereas the one on the right shows that the east, south, and west boundaries of Milwaukee consist of a high concentration of white people. This is because the one on the left is not normalized, whereas the one on the right is normalized. When normalized as percent white, we see the white population as a share of the area's overall population.



Number of white, non-Hispanic people per block group in Milwaukee, WI.



Percent of white, non-Hispanic people per block group in Milwaukee, WI

Normalization is a useful operation for making accurate comparisons between groups. In data analysis, this process refers to multiplying or dividing units by another unit, or factor, to get a new unit that evenly illustrates the values of all records or groups. A common example of this is to divide a data value for an area by the population of the area so that the trends being represented on a map are not merely a reflection of differences in actual population size.

Below, we will walk through how to normalize the ACS Census data so that instead of the count of people for each racial demographic, we will calculate and visualize the proportion of the total population that identify as a certain racial demographic. Then we will walk through how to normalize the acute respiratory cases avoided per year to

every 1000 people populating each block group. This will present the values of each block group as the number of cases avoided per 1000 people per year, to account for the variation in block group populations. Thus map readers can better compare rates between less densely populated areas and more densely populated areas.

Normalize ACS Data

Create a New Attribute

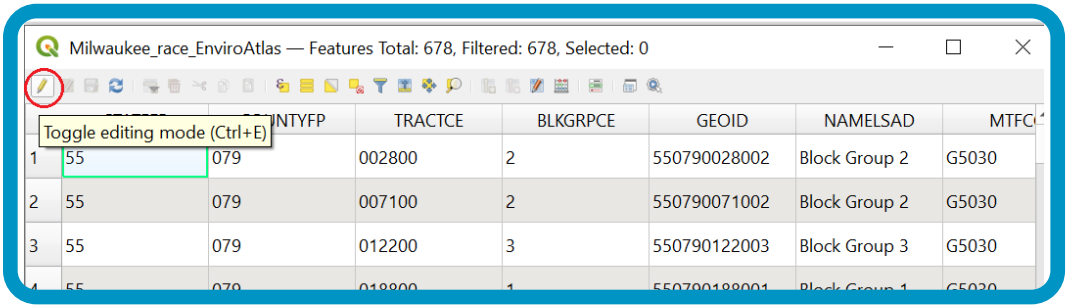
1. To visualize a normalized set of data values, we will need to create a new attribute (field) so that QGIS can use it to display the values in a map, as we did for other attributes earlier in this chapter.

2. In your QGIS project, make sure that you have the shapefile with the EnviroAtlas, and race data joined to the census block group file you created earlier opened up. Additionally, if you did not create an alias (nickname) for the attributes you are interested in, open up the metadata for the ACS table on race so that you can refer to it.

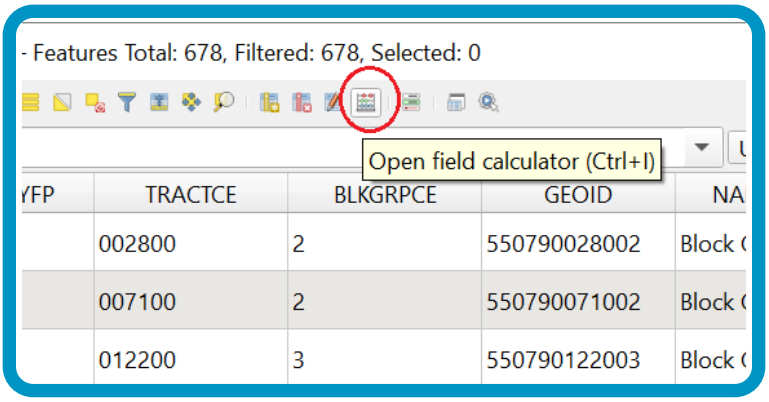
3. In this example we will create the percent of non-Hispanic, white people in the population. So we will need to identify the attribute for non-Hispanic, white, and the attribute for total population from the ACS metadata table. The metadata says that "B03002_001" is total population and "B03002_003" is non-Hispanic, white.

4. Open the attribute table of the shapefile containing both ACS and EnviroAtlas data. Check to make sure that B03002_001 is total population and B03002_003 is non-Hispanic, white by cross-referencing the values for a particular block group from the attribute table in QGIS with the ACS data table you had downloaded originally from the U.S. Census website.

5. Once you've confirmed this, click the Toggle editing mode button in the top left corner (circled in red below).



6. Click the Open field calculator button.

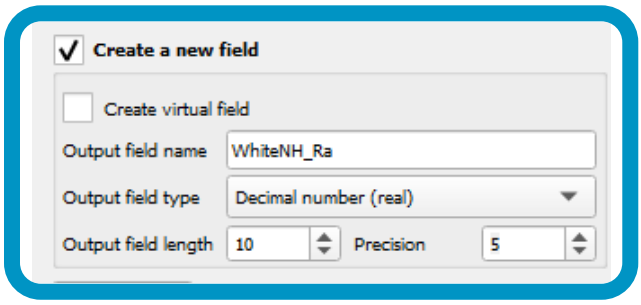


7. In the window that opens up, check the box next to Create a new field.

    a. Output field name box: type "WhiteNH_Ra" which represents White, Non-Hispanic, Rate.

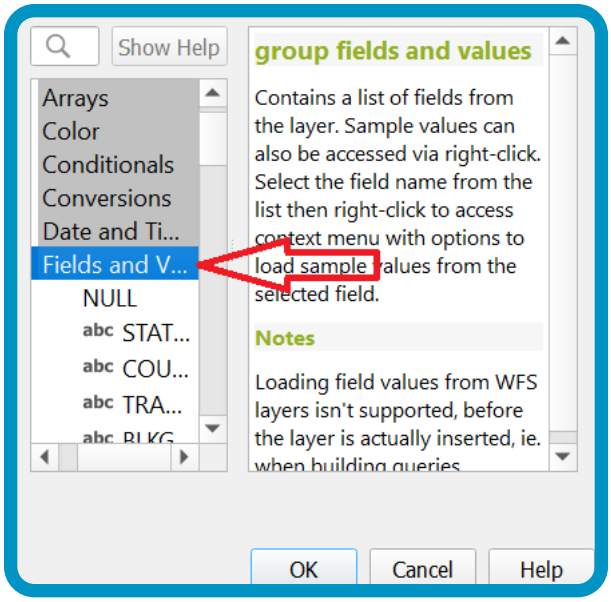
    b. Output field type: select Decimal number (real), from the drop-down menu. Since we are dealing with numbers that are already decimals and dividing them, we want to choose a data type that allows for several decimal points. This helps maintain a higher level of accuracy.

    c. Output field length: 10 and Precision: 5. As currently set, the calculations will show 10 digits max, with 5 digits allowed after the decimal. Depending on your variable you may decide to change these values to impact the accuracy of the value. If you did not set the Output field type to "Decimal number (real)" you may be limited to what numbers you can change and choose.Output field type: select Decimal number (real), from the drop-down menu. Since we are dealing with numbers that are already decimals and dividing them, we want to choose a data type that allows for several decimal points. This helps maintain a higher level of accuracy.



8. We need to create the mathematical expression for our calculation. We will create "non-Hispanic, white" divided by"total population" multiplied by 100 to get a percent value.

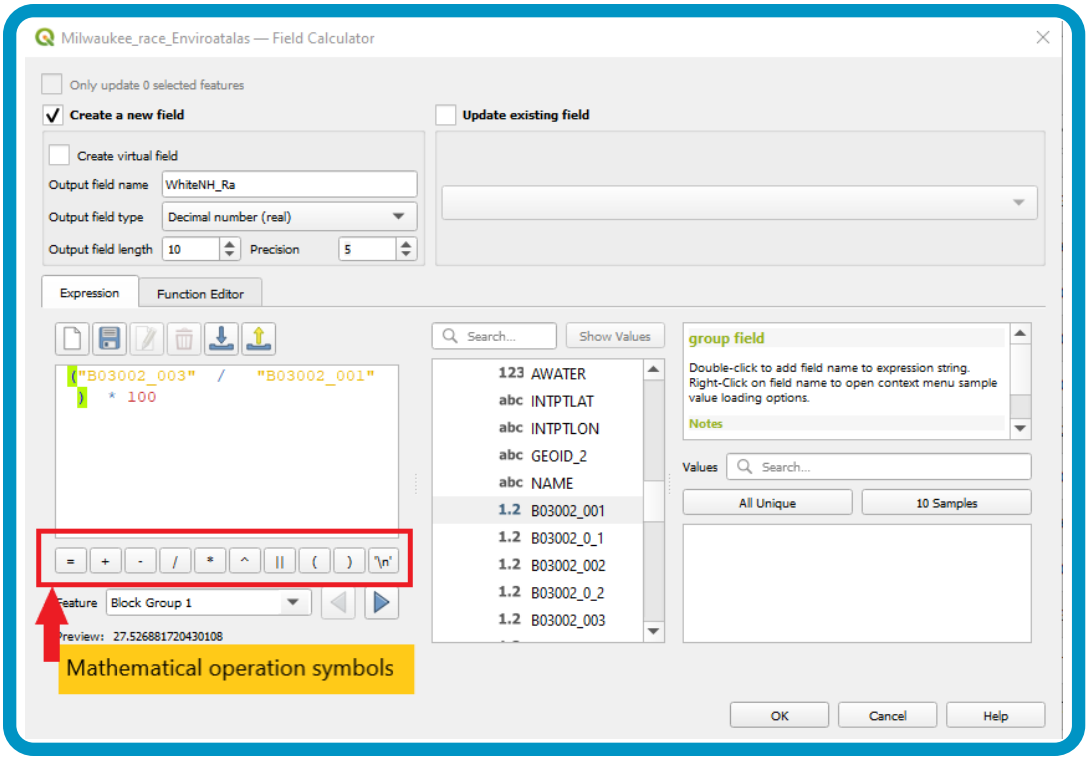
    a. Double click "Fields and Values" in the middle sub-window. You will see all the names of the fields of your shapefile open underneath it.



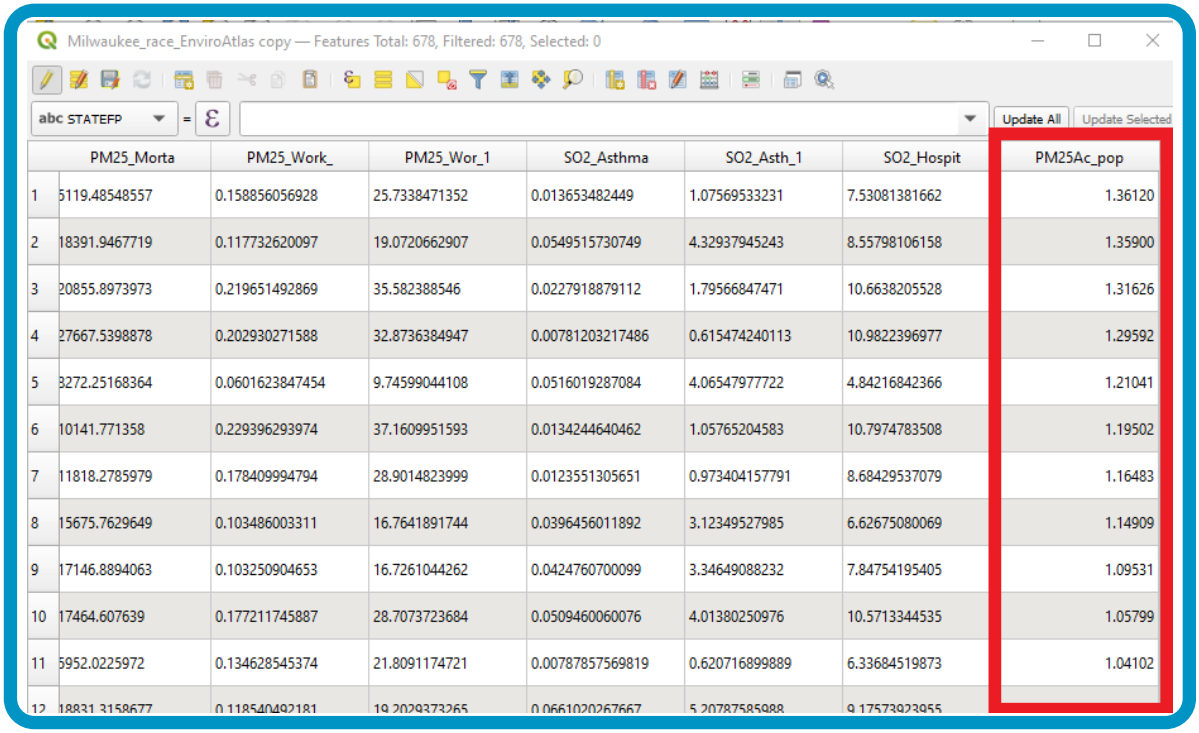
b. Scroll down along the list of shapefile attributes. Double click B03002_003 (for non-Hispanic White). You should see it show up in the Expression box. Do the same for B03002_001 (for total population) and that should also show up in the Expression box.

c. Now that you have both the variables you need, you can complete your expression using the mathematical operation buttons below the Expression box. You can click in the expression box where you want to put each mathematical operation button (it's like typing in a Word document). Make sure to put the expression "B03002_003"/"B03002_001" in parentheses to clarify the mathematical order of operations and then multiply by 100. The equation should be ("B03002_003"/"B03002_001")*100

d. Click Okay.

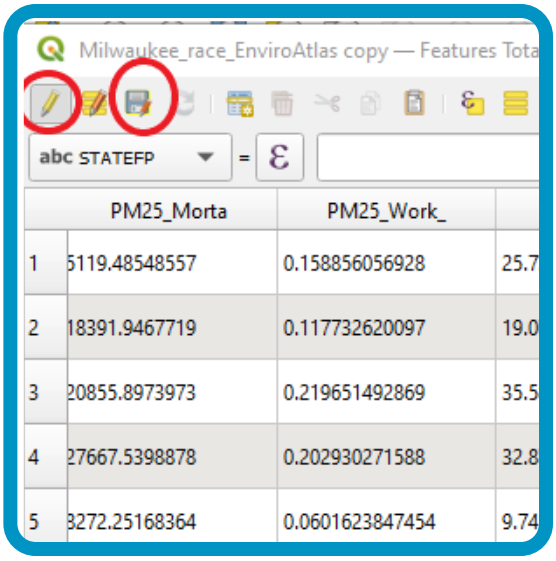


9. When QGIS has finished the calculations, scroll to the last attribute of the attribute table to see the newly created and calculated field.



	PM25_Morta	PM25_Work_	PM25_Wor_1	SO2_Asthma	SO2_Asth_1	SO2_Hospit	PM25Ac_pop
1	5119.48548557	0.158856056928	25.7338471352	0.013653482449	1.07569533231	7.53081381662	1.36120
2	18391.9467719	0.117732620097	19.0720662907	0.0549515730749	4.32937945243	8.55798106158	1.35900
3	20855.8973973	0.219651492869	35.582388546	0.0227918879112	1.79566847471	10.6638205528	1.31626
4	27667.5398878	0.202930271588	32.8736384947	0.00781203217486	0.615474240113	10.9822396977	1.29592
5	3272.25168364	0.0601623847454	9.74599044108	0.0516019287084	4.06547977722	4.84216842366	1.21041
6	10141.771358	0.229396293974	37.1609951593	0.0134244640462	1.05765204583	10.7974783508	1.19502
7	11818.2785979	0.178409994794	28.9014823999	0.0123551305651	0.973404157791	8.68429537079	1.16483
8	15675.7629649	0.103486003311	16.7641891744	0.0396456011892	3.12349527985	6.62675080069	1.14909
9	17146.8894063	0.103250904653	16.7261044262	0.0424760700099	3.34649088232	7.84754195405	1.09531
10	17464.607639	0.177211745887	28.7073723684	0.0509460060076	4.01380250976	10.5713344535	1.05799
11	5952.0225972	0.134628545374	21.8091174721	0.00787857569819	0.620716899889	6.33684519873	1.04102
12	18831.3158677	0.118540492181	19.2029373265	0.0661020267667	5.20787585988	9.17573923955	

10. Click the Save edits button in the top left corner of the window first and then click on the first pencil to toggle off editing. Then close out of the attribute table window.

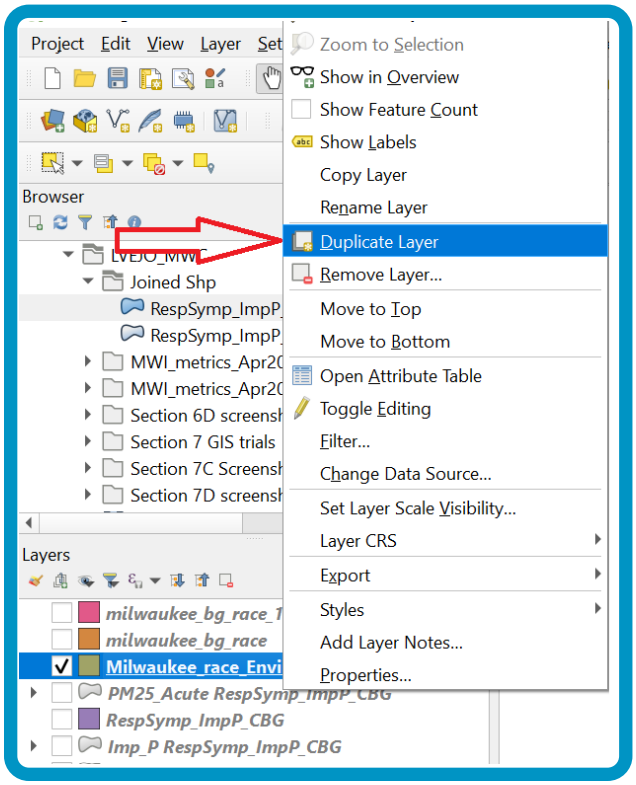


	PM25_Morta	PM25_Work_	
1	5119.48548557	0.158856056928	25.7
2	18391.9467719	0.117732620097	19.0
3	20855.8973973	0.219651492869	35.5
4	27667.5398878	0.202930271588	32.8
5	3272.25168364	0.0601623847454	9.74

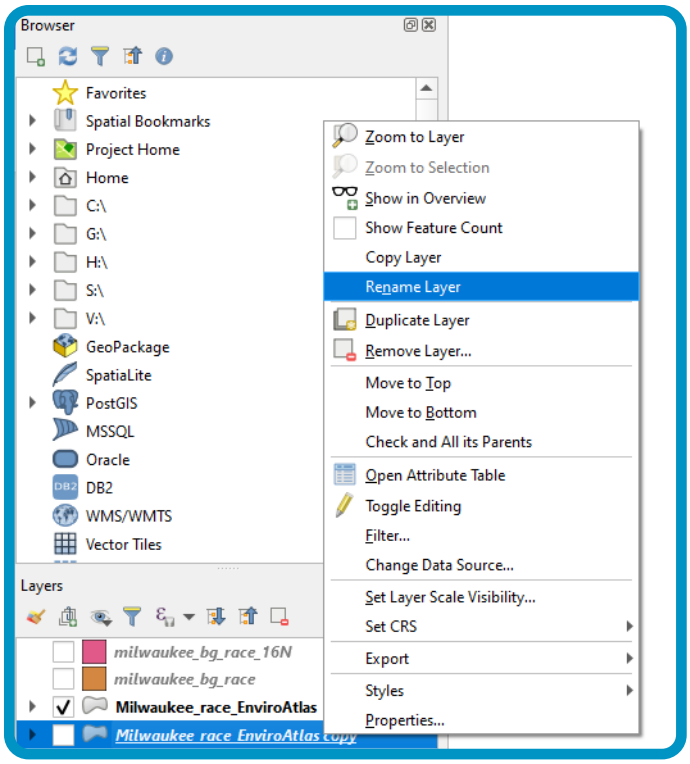
Remember: It is important that you frequently Save (Ctrl+S on PC or Command+S on Mac) as you go to avoid losing your work.

Let's now visualize our newly created attribute. To compare our normalized data to our non-normalized data, let's also copy the layer so that we can visualize two different attributes at once.

1. Right click the ACS + EnviroAtlas shapefile and click Duplicate Layer. You will now see the same layer twice in the Layers window (one will be labeled 'copy').

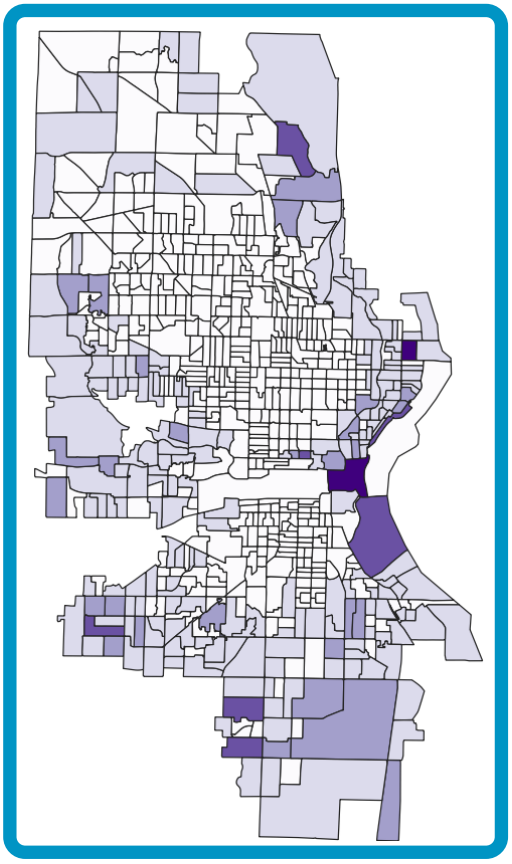


2. Let's rename this layer so we don't get confused. Right click on the copy layer and click "Rename layer". You can rename it whatever you prefer. Note that we are not changing anything about the data within the shapefile as we are simply just visualizing the same file two times in QGIS.

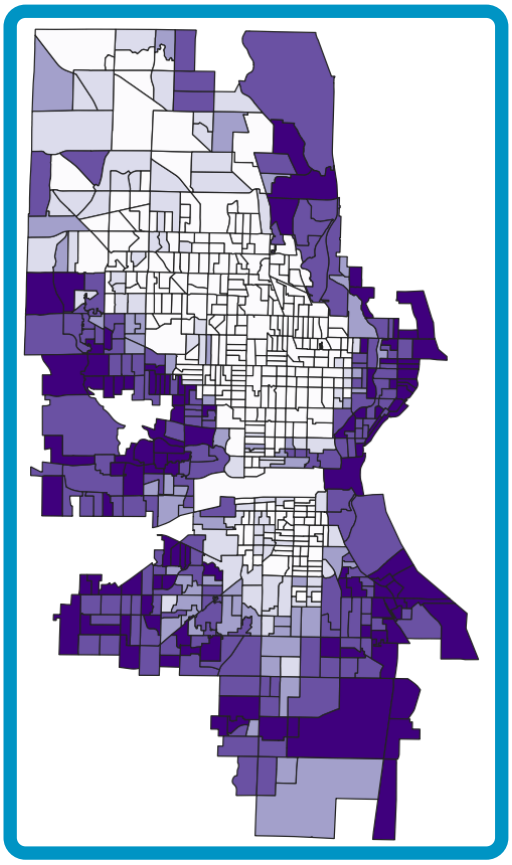


3. In one layer, following the steps from earlier symbolize "B03002_003". In the other layer, symbolize the new attribute "WhiteNH_Ra".

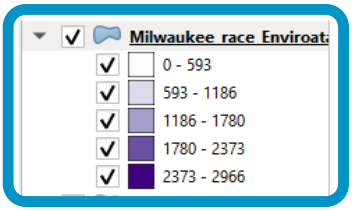
4. Observe how different the two maps look despite both showing ways to visualize the population of white, non-Hispanic people in block groups of Milwaukee.

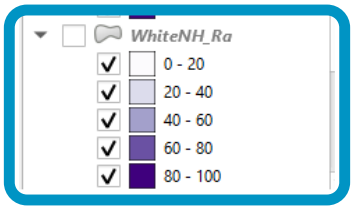


Number of white, non-Hispanic people per block group in Milwaukee, WI.



Percent of white, non-Hispanic people per block group in Milwaukee, WI.



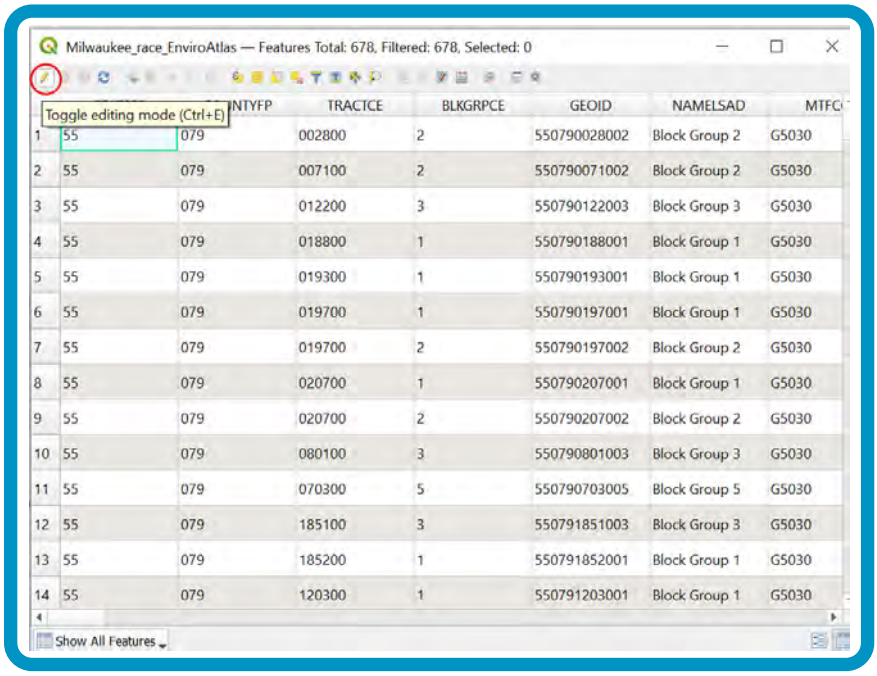


Taking a look at the legends, you also see how the values are different. One is a straight count of the number of white people, which is hard to interpret when comparing between block groups. Whereas the second is by percent which is easier for many map viewers to compare between block groups. For example in the first map, on the west (left) side of the map, it looks like there are not very many white, non-Hispanic individuals in the Milwaukee populations. However, when you take a look at the percentage map, you now learn that though the population count of white, non-Hispanics are low, they make up a majority of the population in that area. Meaning, the area has a lower population overall and would thus show lower values for whatever race is visualized.

Normalize EnviroAtlas Data

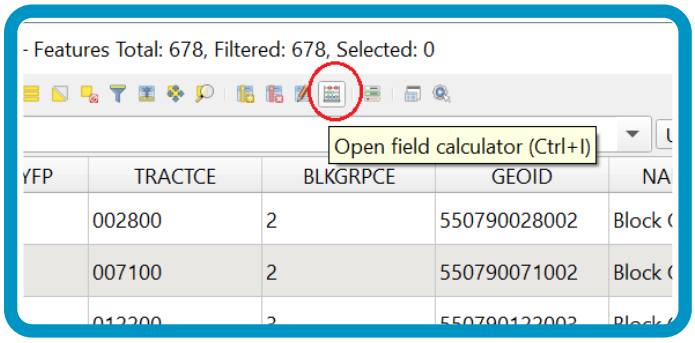
One reason that visualizing the straight number of respiratory cases prevented for each census block does not make it easy for us to compare different census block groups is because the number of cases prevented per year also depends on the population of each census block group. If a block group has a higher total population, then it may prevent more respiratory cases per tree planted than an area with a lower population. Thus, to get a clearer picture of which areas are benefiting more from trees preventing acute respiratory cases, we should control for population size. We will do this by dividing the "PM25_Acute" attribute of the EnviroAtlas shapefile by the total population of each census block group.

1. Open the attribute table of the shapefile containing both ACS and EnviroAtlas data. Click the Toggle editing mode button in the top left corner (circled in red below).



	NTYFP	TRACTCE	BLKGRPCE	GEOID	NAMELSAD	MTFC
1	55	079	002800	2	550790028002	Block Group 2	G5030
2	55	079	007100	2	550790071002	Block Group 2	G5030
3	55	079	012200	3	550790122003	Block Group 3	G5030
4	55	079	018800	1	550790188001	Block Group 1	G5030
5	55	079	019300	1	550790193001	Block Group 1	G5030
6	55	079	019700	1	550790197001	Block Group 1	G5030
7	55	079	019700	2	550790197002	Block Group 2	G5030
8	55	079	020700	1	550790207001	Block Group 1	G5030
9	55	079	020700	2	550790207002	Block Group 2	G5030
10	55	079	080100	3	550790801003	Block Group 3	G5030
11	55	079	070300	5	550790703005	Block Group 5	G5030
12	55	079	185100	3	550791851003	Block Group 3	G5030
13	55	079	185200	1	550791852001	Block Group 1	G5030
14	55	079	120300	1	550791203001	Block Group 1	G5030

2. Click the Open field calculator button.



YFP	TRACTCE	BLKGRPCE	GEOID	NA
	002800	2	550790028002	Block
	007100	2	550790071002	Block
	012200	3	550790122003	Block

3. In the window that opens up, check the box next to Create a new field.
    a. In the Output field name box type "PM25Ac_pop".
    b. In the Output field type drop-down menu, select Decimal number (real). Since we are dealing with numbers that are already decimals and dividing them, we do not want to get our results in whole numbers because we want to keep a higher level of accuracy.
    c. Put 10 for Output field length and 5 for Precision.

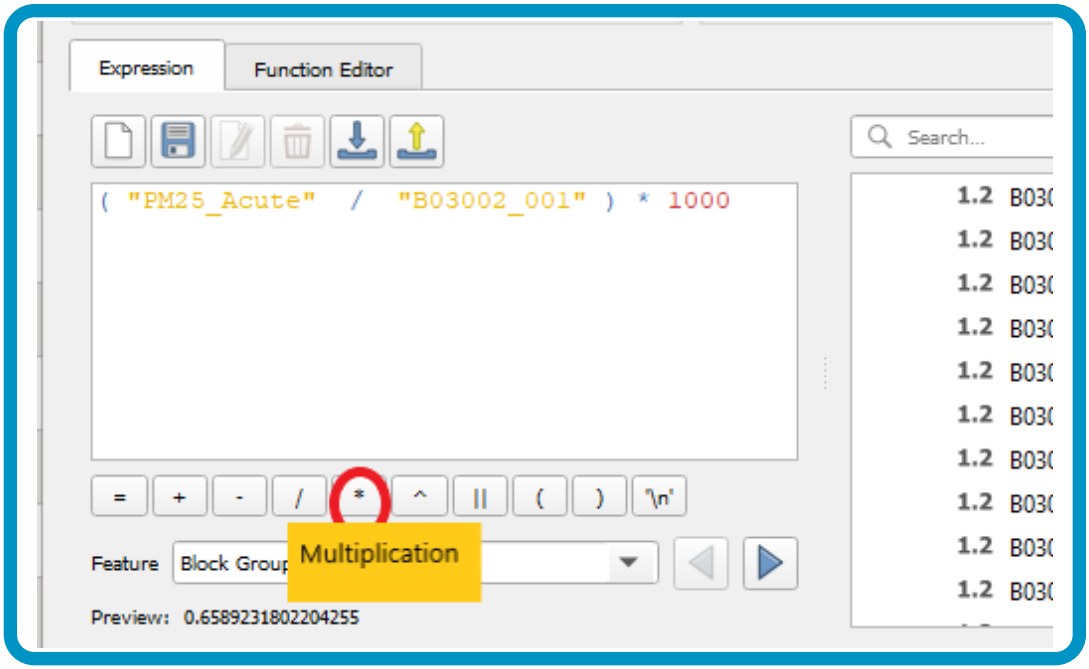
4. We need to create the mathematical expression for our calculation. Double click "Fields and Values" in the middle sub-window. You will see all the names of the fields of your shapefile open underneath it.

    a. Scroll down along the list of shapefile attributes. Double click PM25_Acute. You should see the "PM25_Acute" appear in the expression box.

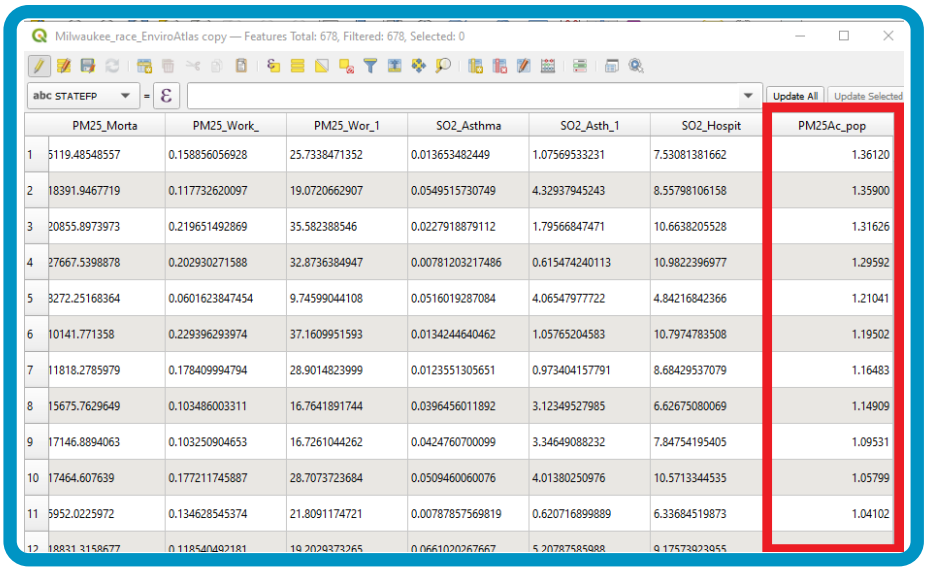
    b. Click the division operator underneath the expression box. This indicates the division sign similar to a calculator.

    c. Back under the Fields and Values category in the middle sub-window, we need to find the attribute that relates to total population. You may need to refer to the ACS metadata table to find out. In this case it is B03002_001. So under "Fields and Values" find B03002_001 and double click it.

    d. Next, because the numbers will be easier for people to read and understand when not in small decimals with many digits, we will express our cases as per 1000 people rather than per one individual person. Thus, we will multiply the expression by 1000. Make sure to put the expression "PM25_Acute"/"B03002_001" in parentheses to clarify the mathematical order of operations and then multiply by 1000. The expression should be: ("PM25_Acute"/"B03002_001")*1000). Then click Okay.

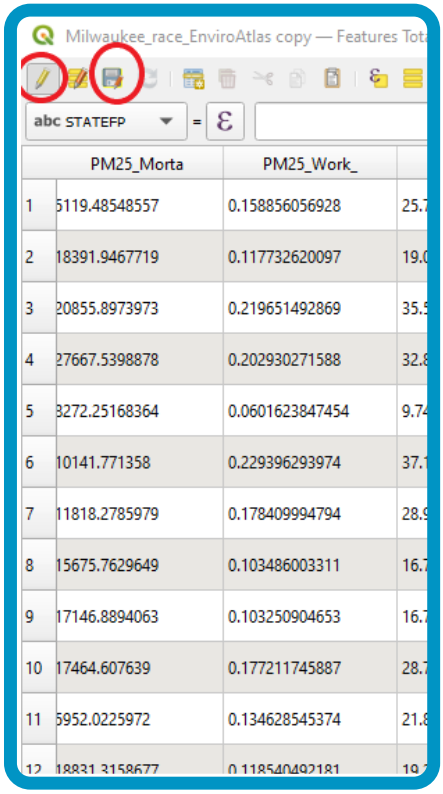


5. When finished running, scroll to the last attribute of the attribute table to see the newly created and calculated field.



	PM25_Morta	PM25_Work_	PM25_Wor_1	SO2_Asthma	SO2_Asth_1	SO2_Hospit	PM25Ac_pop
1	5119.48548557	0.158856056928	25.7338471352	0.013653482449	1.07569533231	7.53081381662	1.36120
2	18391.9467719	0.117732620097	19.0720662907	0.0549515730749	4.32937945243	8.55798106158	1.35900
3	20855.8973973	0.219651492869	35.582388546	0.0227918879112	1.79566847471	10.6638205528	1.31626
4	27667.5398878	0.202930271588	32.8736384947	0.00781203217486	0.615474240113	10.9822396977	1.29592
5	3272.25168364	0.0601623847454	9.74599044108	0.0516019287084	4.06547977722	4.84216842366	1.21041
6	10141.771358	0.229396293974	37.1609951593	0.0134244640462	1.05765204583	10.7974783508	1.19502
7	11818.2785979	0.178409994794	28.9014823999	0.0123551305651	0.973404157791	8.68429537079	1.16483
8	15675.7629649	0.103486003311	16.7641891744	0.0396456011892	3.12349527985	6.62675080069	1.14909
9	17146.8894063	0.103250904653	16.7261044262	0.0424760700099	3.34649088232	7.84754195405	1.09531
10	17464.607639	0.177211745887	28.7073723684	0.0509460060076	4.01380250976	10.5713344535	1.05799
11	5952.0225972	0.134628545374	21.8091174721	0.00787857569819	0.620716899889	6.33684519873	1.04102
12	18831.3158677	0.118540492181	19.2029373265	0.0661020267667	5.20787585988	9.17573923955	

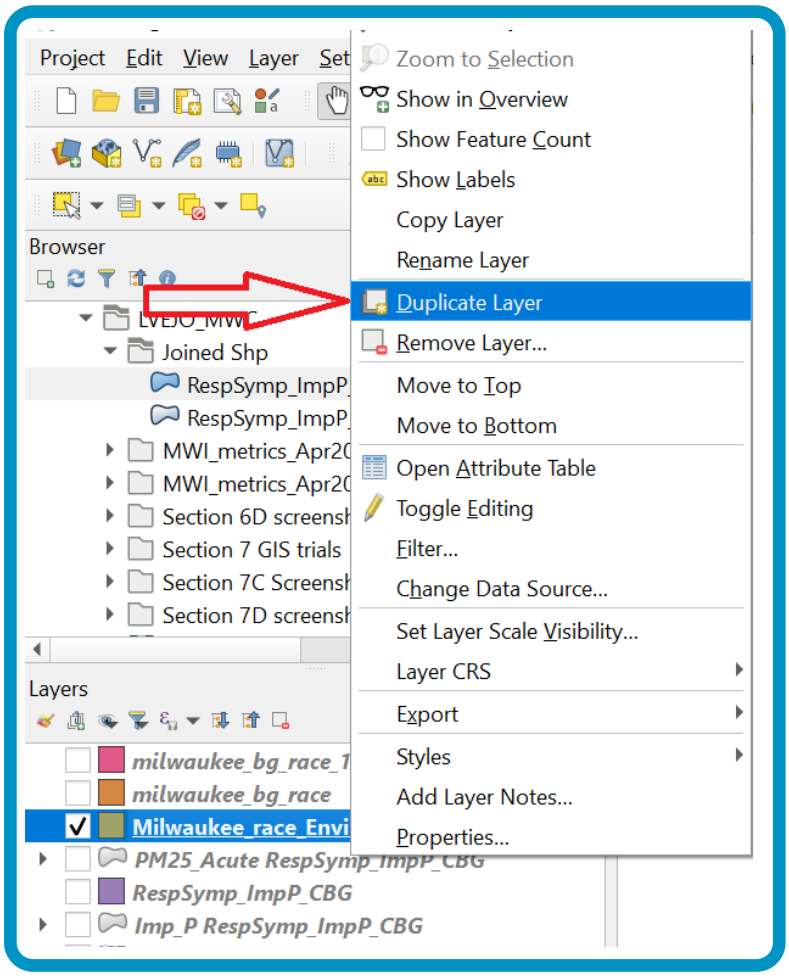
6. Click the Save edits button in the top left corner of the window first and then click on the first pencil to toggle off editing. Then close out of the attribute table window. Remember to save the whole QGIS document.

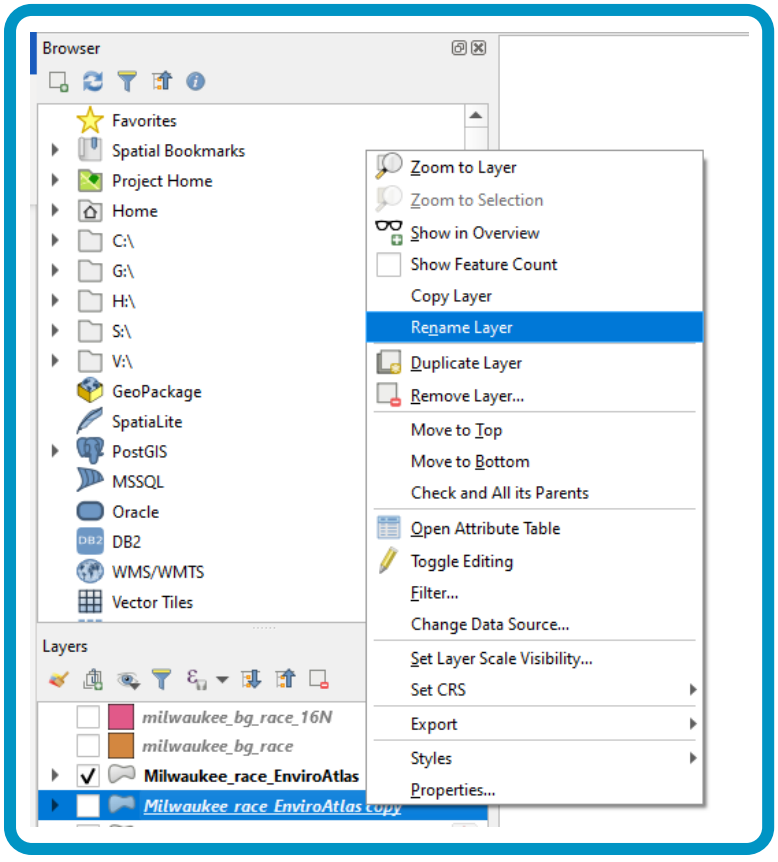


	PM25_Morta	PM25_Work_	
1	5119.48548557	0.158856056928	25.7
2	18391.9467719	0.117732620097	19.0
3	20855.8973973	0.219651492869	35.5
4	27667.5398878	0.202930271588	32.8
5	3272.25168364	0.0601623847454	9.74
6	10141.771358	0.229396293974	37.1
7	11818.2785979	0.178409994794	28.9
8	15675.7629649	0.103486003311	16.7
9	17146.8894063	0.103250904653	16.7
10	17464.607639	0.177211745887	28.7
11	5952.0225972	0.134628545374	21.8
12	18831.3158677	0.118540492181	19.7

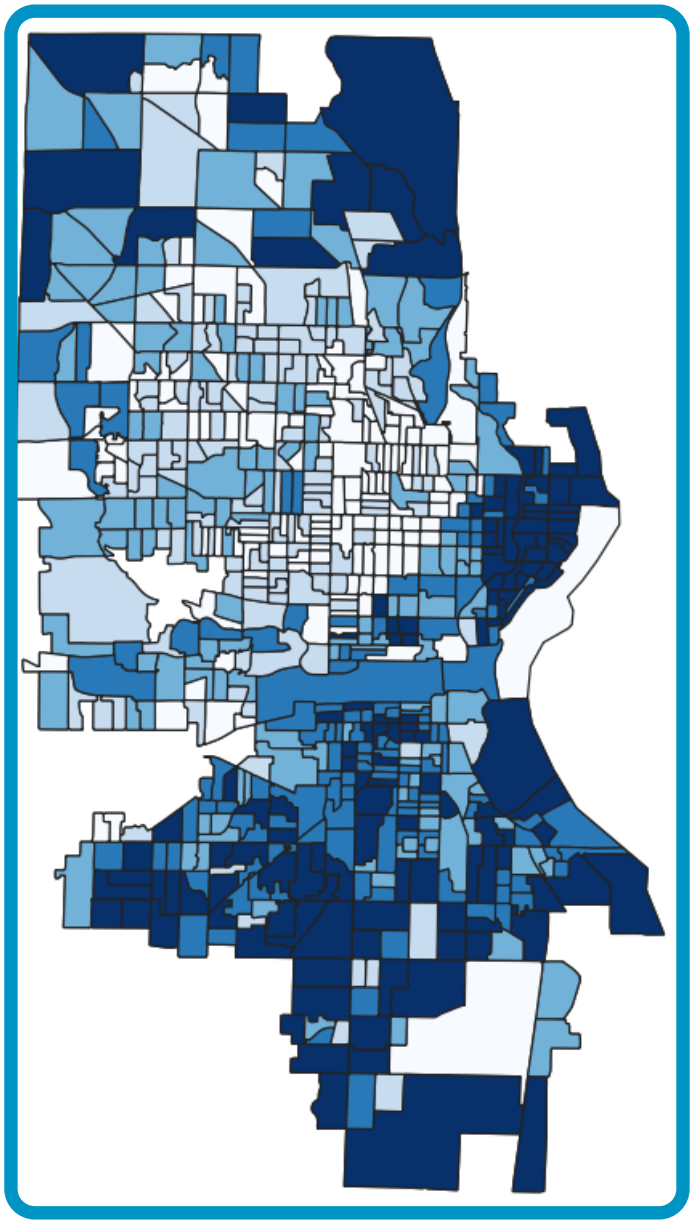
Let's now visualize our newly created attribute. To compare our normalized data to our non-normalized data, let's also copy the layer so that we can visualize two different attributes at once.

1. Right click the ACS + EnviroAtlas shapefile and click Duplicate Layer. You will now see the same layer twice in the Layers window (one will be labeled 'copy'). Let's rename this layer so we don't get confused. Right click on the copy layer and click "Rename layer". Note that we are not changing anything about the data within the shapefile as we are simply just visualizing the same file two times in QGIS.

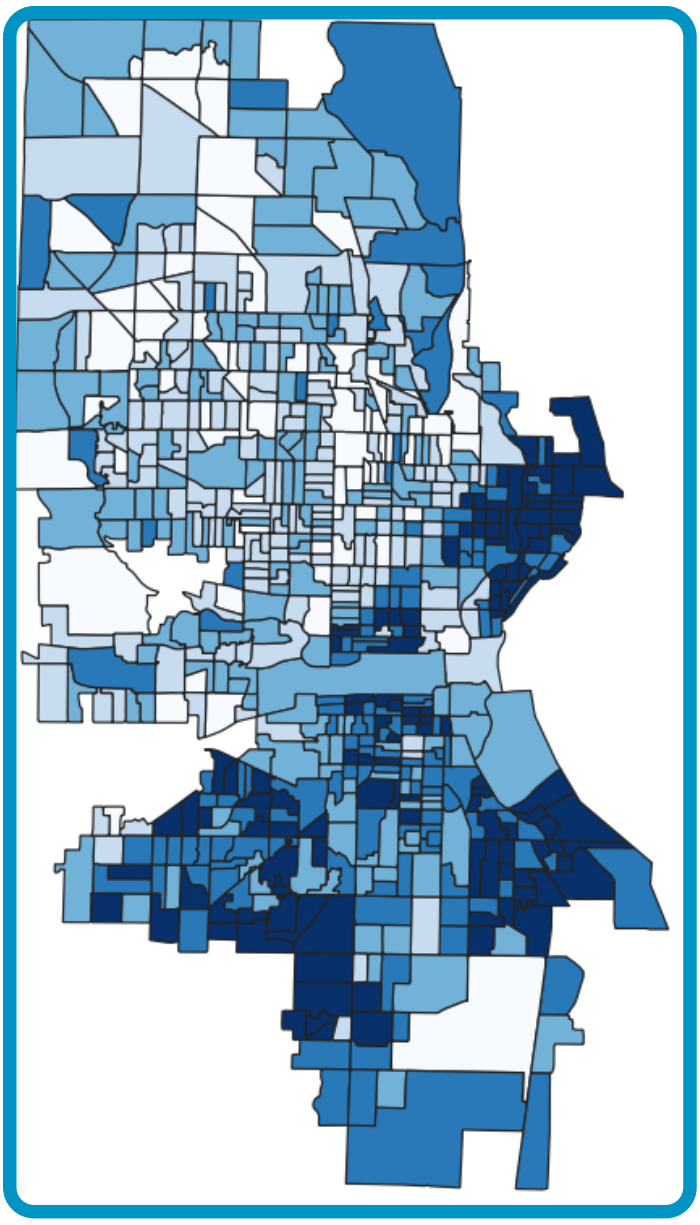

Project Edit View Layer Set
Browser
Joined Shp
Layers
Zoom to Selection
Show in Overview
Show Feature Count
Show Labels
Copy Layer
Rename Layer
Duplicate Layer
Remove Layer...
Move to Top
Move to Bottom
Open Attribute Table
Toggle Editing
Filter...
Change Data Source...
Set Layer Scale Visibility...
Layer CRS
Export
Styles
Add Layer Notes...
Properties...
milwaukee_bg_race_1
milwaukee_bg_race
Milwaukee_race_Envi
PM25_Acute RespSymp_ImpP_CBG
RespSymp_ImpP_CBG
Imp_P RespSymp_ImpP_CBG


Browser
Favorites
Spatial Bookmarks
Project Home
Home
C:\
G:\
H:\
S:\
V:\
GeoPackage
SpatiaLite
PostGIS
MSSQL
Oracle
DB2
WMS/WMTS
Vector Tiles
Layers
milwaukee_bg_race_16N
milwaukee_bg_race
Milwaukee_race_EnviroAtlas
Milwaukee_race_EnviroAtlas copy
Zoom to Layer
Zoom to Selection
Show in Overview
Show Feature Count
Copy Layer
Rename Layer
Duplicate Layer
Remove Layer...
Move to Top
Move to Bottom
Check and All its Parents
Open Attribute Table
Toggle Editing
Filter...
Change Data Source...
Set Layer Scale Visibility...
Set CRS
Export
Styles
Properties...

2. Symbolize one of the layers for PM25_Acute using a gradient color scheme. On the duplicate layer of the same shapefile, symbolize the new attribute "PM25Acu_po" in the same fashion.



Acute respiratory symptoms avoided due to particulate matter [PM2.5] removed by tree cover (cases/year) in Milwaukee, WI.



Acute respiratory symptoms avoided due to particulate matter [PM2.5] removed by tree cover normalized by population (per 1000 people) in Milwaukee, WI.

Observe how different the two maps look from each other despite both being valid ways of visualizing the number of acute respiratory cases avoided due to particulate matter removed by tree cover in Milwaukee.

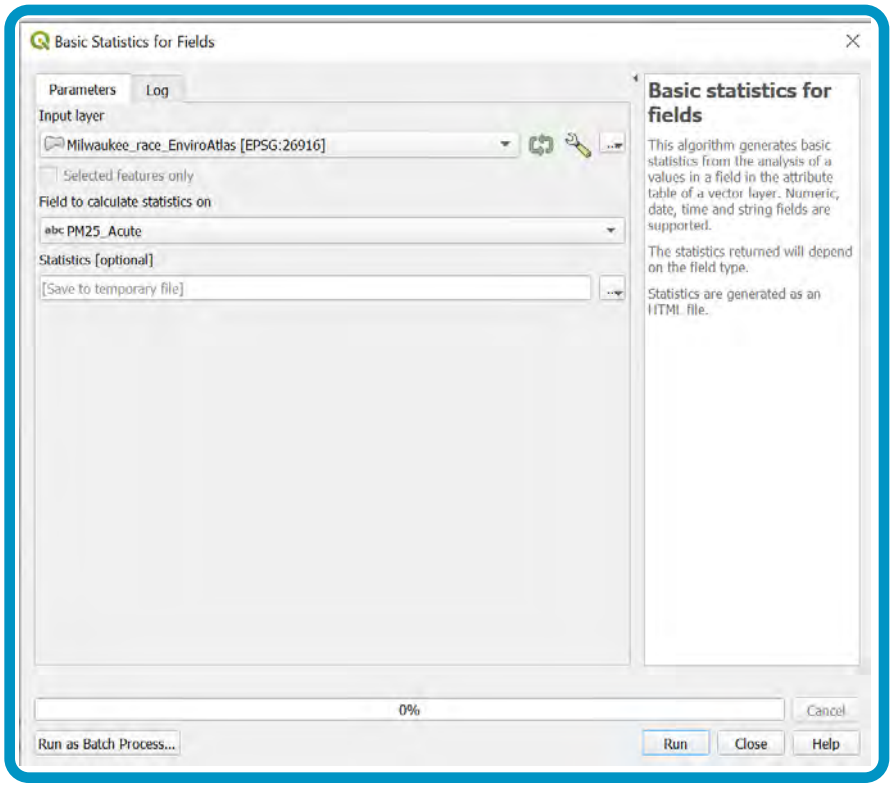
Summary Statistics

Although there are many useful statistical applications in data analysis and mapping, the mathematics involved in these are beyond the scope of this toolkit. However, one quick and easy way to get a few major and useful statistics for an attribute of interest is by accessing the summary statistics of the attribute. Some values you'll calculate include mean, maximum, minimum, sum, median, etc.

How-To

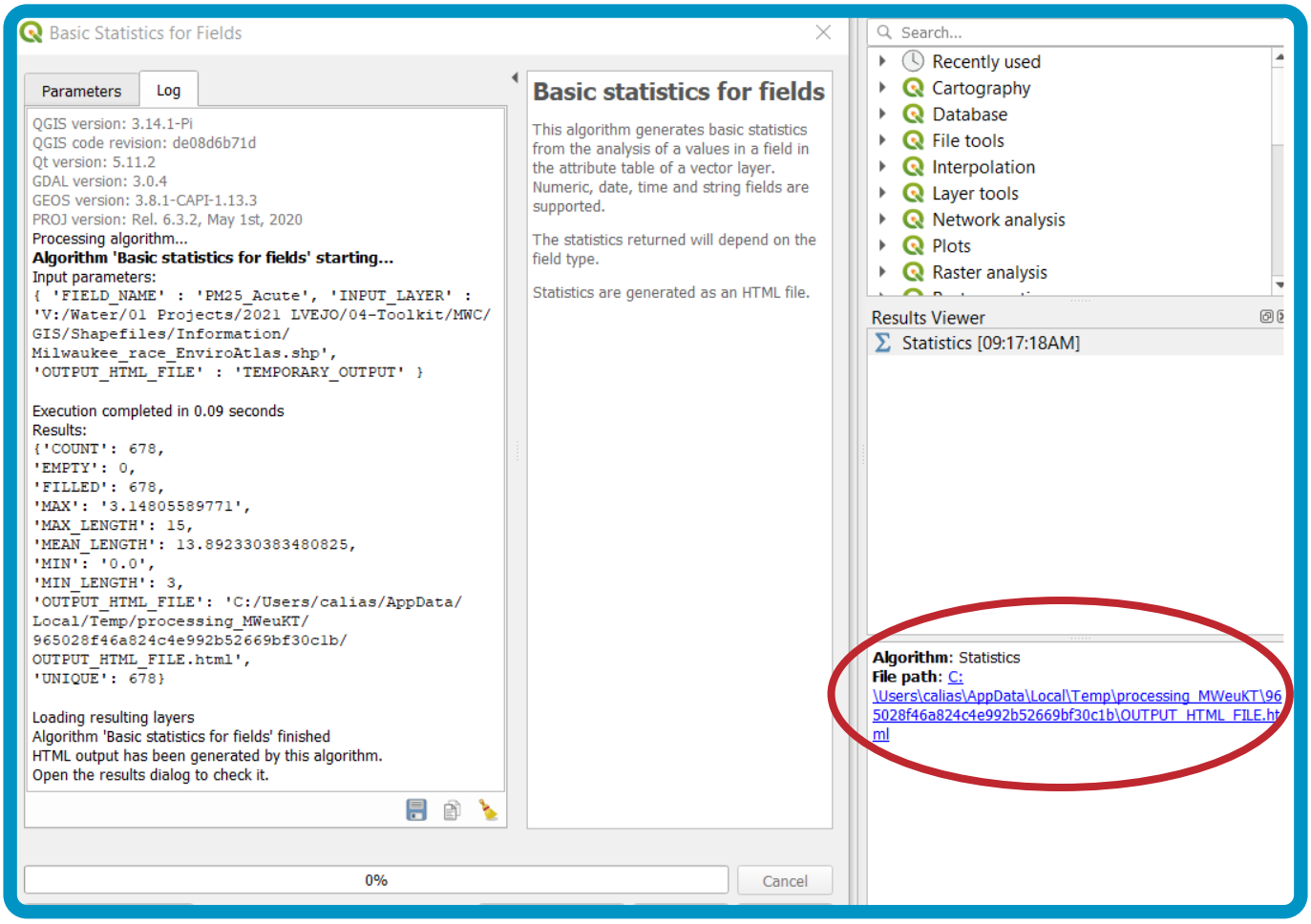
We are going to access basic summary statistics for the shapefile that we created during our last exercise.

1. In the bottom left search engine search for the Basic Statistics for Fields tool.
2. In the tool select the Milwaukee ACS + EnviroAtlas shapefile for the Input layer.
3. Choose the attribute PM25_Acute for Field to calculate statistics on.
4. Fill out the Statistics [optional] if you'd like to save the results somewhere.



5. Click Run. You can view the results by clicking the hyperlink that shows up under Results Viewer (the window on the right in the picture below) called "File Path". When you click the hyperlink, it may open in your web browser or as a document and look like the following picture, providing values for different summary statistics. You can also view the statistics in the "Log" tab of the tool.

Remember: It is important that you frequently Save (Ctrl+S on PC or Command+S on Mac) as you go to avoid losing your work.


Basic Statistics for Fields
Parameters
Log
QGIS version: 3.14.1-Pi
QGIS code revision: de08d6b71d
Qt version: 5.11.2
GDAL version: 3.0.4
GEOS version: 3.8.1-CAPI-1.13.3
PROJ version: Rel. 6.3.2, May 1st, 2020
Processing algorithm...
Algorithm 'Basic statistics for fields' starting...
Input parameters:
{ 'FIELD_NAME' : 'PM25_Acute', 'INPUT_LAYER' : 'V:/Water/01 Projects/2021 LVEJO/04-Toolkit/MWC/GIS/Shapefiles/Information/Milwaukee_race_EnviroAtlas.shp', 'OUTPUT_HTML_FILE' : 'TEMPORARY_OUTPUT' }
Execution completed in 0.09 seconds
Results:
{'COUNT': 678, 'EMPTY': 0, 'FILLED': 678, 'MAX': '3.14805589771', 'MAX_LENGTH': 15, 'MEAN_LENGTH': 13.892330383480825, 'MIN': '0.0', 'MIN_LENGTH': 3, 'OUTPUT_HTML_FILE': 'C:/Users/calias/AppData/Local/Temp/processing_MWeuKT/965028f46a824c4e992b52669bf30c1b/OUTPUT_HTML_FILE.html', 'UNIQUE': 678}
Loading resulting layers
Algorithm 'Basic statistics for fields' finished
HTML output has been generated by this algorithm.
Open the results dialog to check it.
Basic statistics for fields
This algorithm generates basic statistics from the analysis of a values in a field in the attribute table of a vector layer. Numeric, date, time and string fields are supported.
The statistics returned will depend on the field type.
Statistics are generated as an HTML file.
0%
Cancel
Search...
Recently used
Cartography
Database
File tools
Interpolation
Layer tools
Network analysis
Plots
Raster analysis
Results Viewer
Statistics [09:17:18AM]
Algorithm: Statistics
File path: C:\Users\calias\AppData\Local\Temp\processing_MWeuKT\965028f46a824c4e992b52669bf30c1b\OUTPUT_HTML_FILE.html

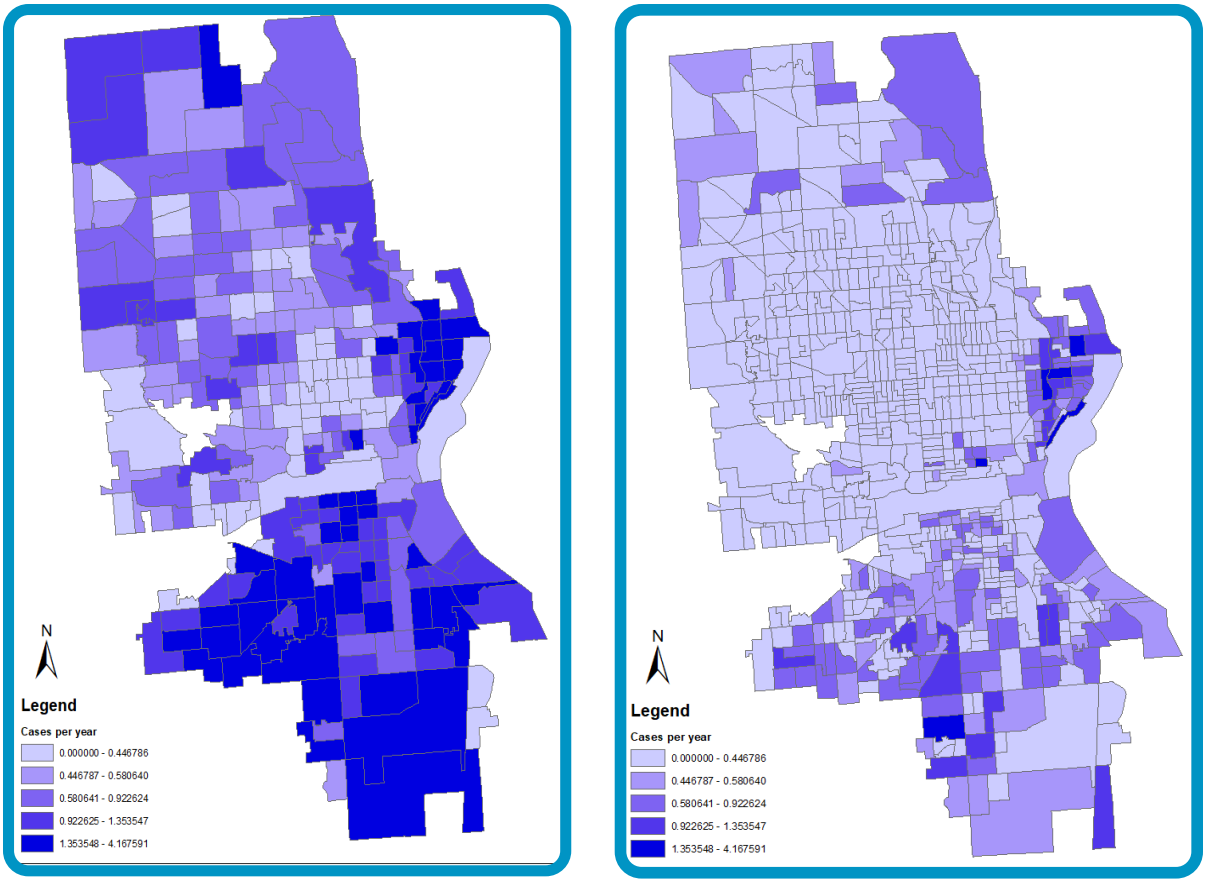

basic statistics for field result qgi
OUTPUT_HTML_FILE.html
File | C:/Users/calias/AppData/Local/Temp/processing_osrp
Analyzed field: PM25_Acute
Count: 678
Unique values: 678
NULL (missing) values: 0
Minimum value: 0.0
Maximum value: 3.14805589771
Range: 3.14805589771
Sum: 240.61435272287954
Mean value: 0.35488842584495506
Median value: 0.282078259172
Standard deviation: 0.2758236057065886
Coefficient of Variation: 0.7772121760518936
Minority (rarest occurring value): 0.0
Majority (most frequently occurring value): 0.0
First quartile: 0.160227877675
Third quartile: 0.469537315461
Interquartile Range (IQR): 0.30930943778600006

Aggregating

Oftentimes, we may want to change the spatial scale of specific data layers for our analysis. There are different spatial operations that you can perform that can change the scale of the data while preserving the quantities of its values. One of the most basic ways to change the scale of a shapefile from a smaller spatial scale (higher resolution e.g. block group) to a larger spatial scale (lower resolution e.g. census tract, county or state) is by aggregating the data. One example of this is changing a dataset from county scale to state scale, or aggregating census blocks to census tracts. You can also aggregate a smaller scale into a larger one that does not perfectly spatially "fit" into each other. For example, you may want to aggregate parcel (a plot of land or a single property) information to the neighborhood, zip code, or city scale. This type of aggregation requires a few more mathematical steps and a process called weighting which will not be covered in a tutorial in this toolkit.

Depending on the units of the data you are trying to aggregate, you will need to choose the appropriate mathematical operation when conducting the aggregation –typically an average or a sum of values being aggregated will best represent the data at a larger spatial scale for many basic analyses. The mathematical operations you choose for data aggregation depend on the units and type of data you have available to you and the goal(s) of your analysis. Oftentimes it will make the most mathematical sense to find the average of all entities while aggregating if the units of the data are a percentage. Other times, an aggregation can be a simple sum of entities, for example, the number of individuals over 65 years old in each county can be added together to represent the total for the state.

Below is an example of a simple aggregation that was done on the same Milwaukee data for respiratory cases avoided due to tree cover. The map on the left shows the data at the census block group scale while the map on the right shows the same data at the census tract scale.



The data in both maps are symbolized with a blue color gradient using the same values for the classification (refer to the legend for this). As you can see, the map on the right has larger sized polygons as all the data for each tract's block groups have been added together. Both maps have pros and cons depending on the question we are asking and what the goal(s) of our analysis is.

Weighting

As mentioned earlier, there are many cases, in which summing or averaging up all the polygons within a desired boundary is not mathematically acceptable because not all of the polygons are evenly spaced, are evenly sized, or align perfectly within the boundaries of the spatial scale we are trying to convert to.

To take account of this, we perform a process called weighting. This process seeks to essentially balance all the data according to the amount of space each feature takes up and then tabulating the values in a way that is proportional to each entity.

The weighting process will not be discussed in-depth in this toolkit. QGIS has developed a tool, overlap analysis, that may be useful to identify what proportion a layer overlaps with another layer, but the tool still has some bugs as of 2021. Visit this webpage, https://docs.qgis.org/3.16/en/docs/user_manual/processing_algs/qgis/vectoranalysis.html and click on Ovelap Analysis to learn about it.

Here is a helpful example of some of its limitations: https://gis.stackexchange.com/questions/336862/overlap-analysis-in-qgis

Disaggregating

In some situations, we may want to downscale the data we have. This means we want to change a dataset from a larger geographic scale to a smaller geographic scale, or increase its resolution. For example, you may be interested in using census-tract level health data for block groups within.

Disaggregation requires making assumptions about the data based on the units of the data, resolutions you are working with, and the goal(s) of your analysis. These assumptions must always be clearly conveyed along with the final results. There are many limitations associated with different disaggregation techniques, and the units of the data will inform whether disaggregation is even possible based on which information you have. For example, disaggregating rates or percentages to smaller areas may be more valid than trying to disaggregate a count of something to smaller areas. This is because there may be many unique factors within each block group that would impact the count of cases each block group has. Exploring these limitations in detail is outside the scope of this toolkit, but as you continue to further your experience with GIS, we recommend you look at educational journal articles or talk to research partners about the topic.

A basic way to disaggregate, that applies when the attribute value is a rate or a percentage, is assigning the same rate of the large scale to the small scale. For example, public health data is often only publicly available at the census tract level, but maybe your analysis is at the block group level. One way to disaggregate is to assign each block group within a census tract the same rate for the specific health attribute.

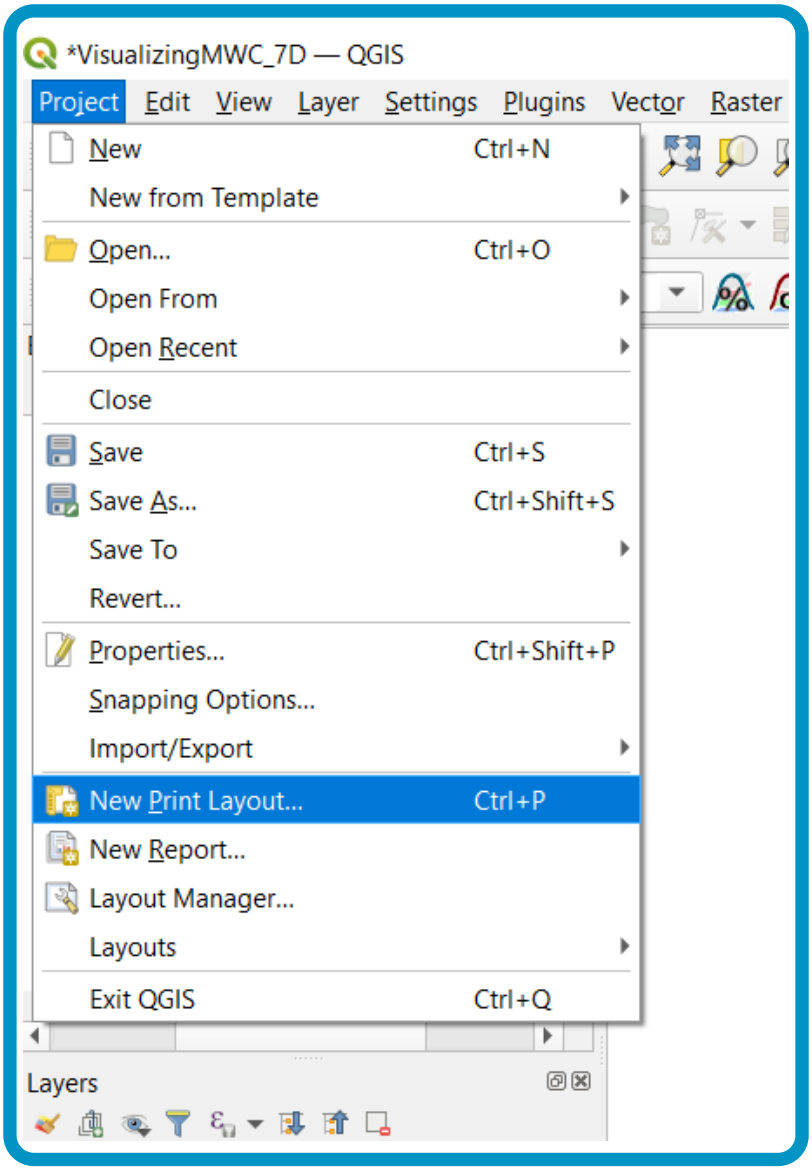
Think carefully about the data to make sure that the disaggregation method makes sense and is valid.

D. Create and Interpret the Final Map

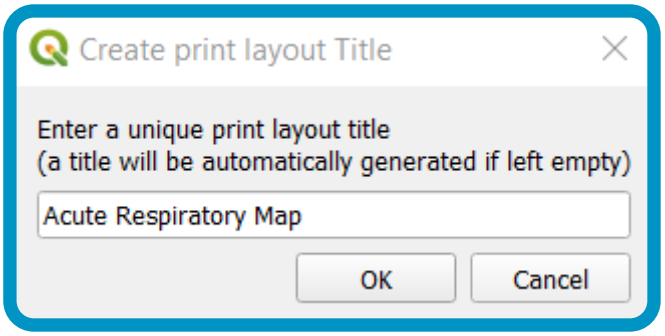
When creating a final map, there are multiple pieces aside from the visualized data that need to be added to help readers understand the map. The following steps layout one way of how to create a print-worthy map for QGIS.

How-To

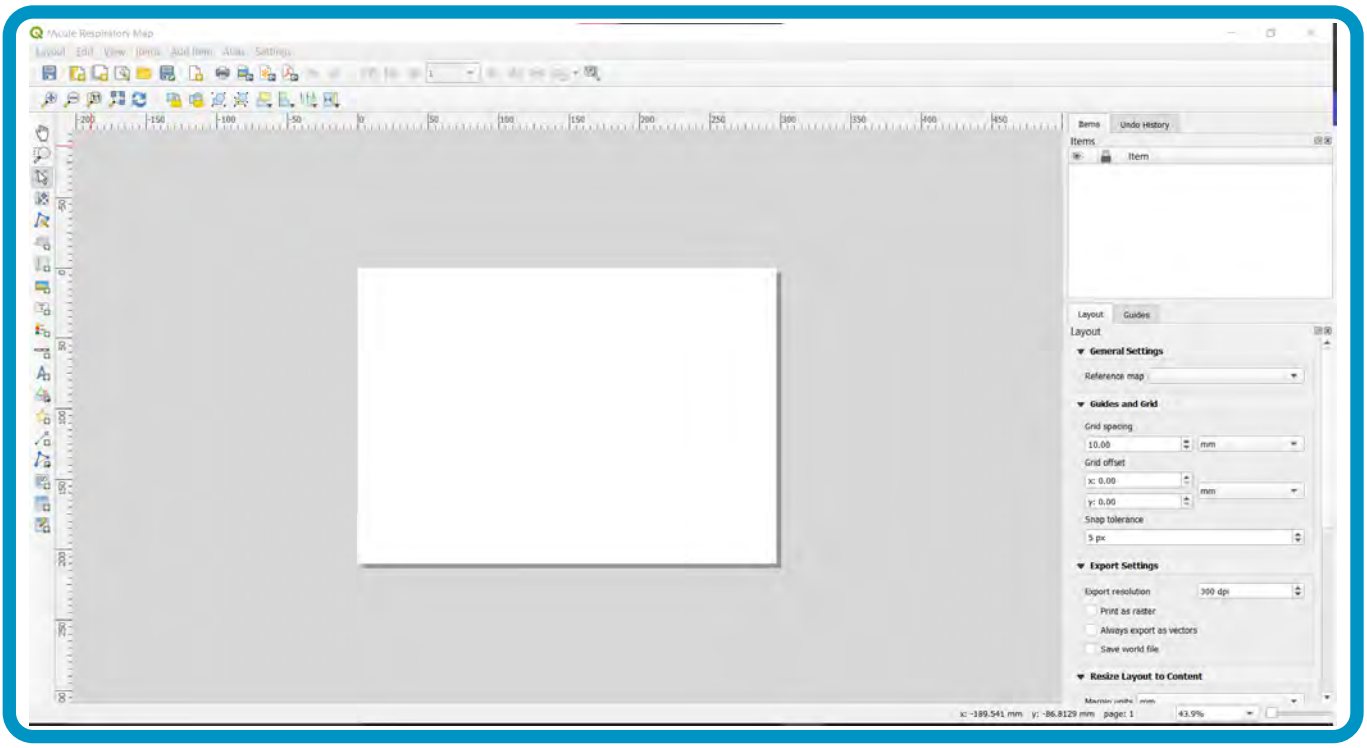
1. Click on project to open a "New Print Layout"



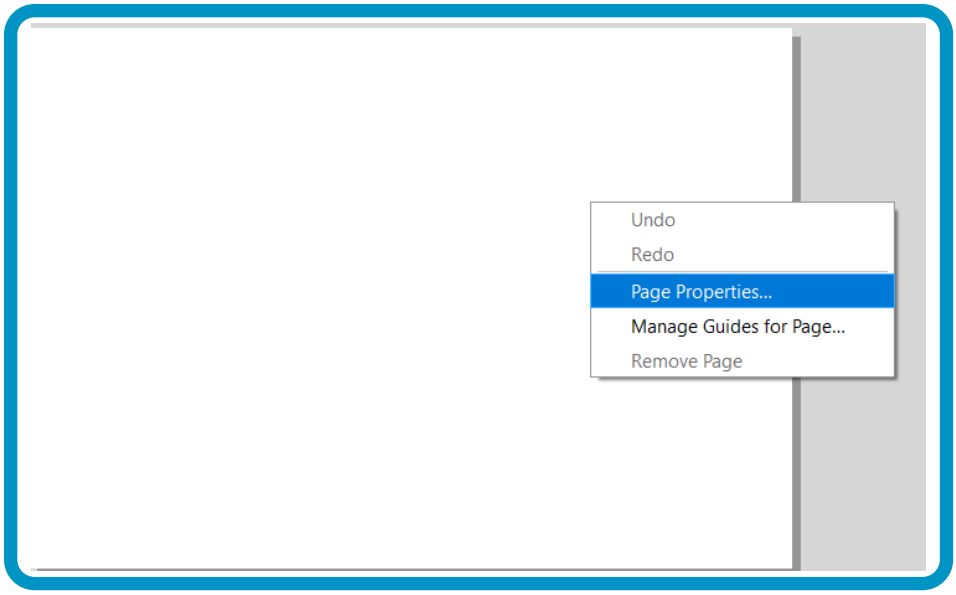
2. A window will pop-up to label the print layout. You can name it whatever you would like. Then click okay.



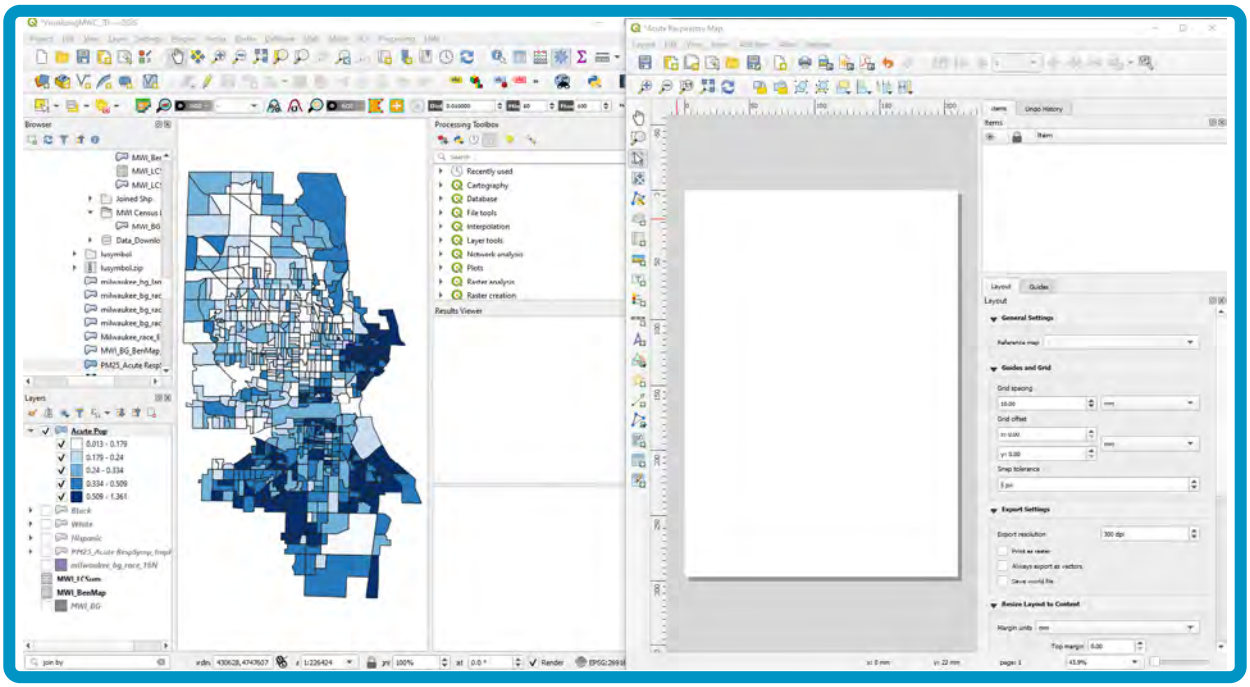
3. A new window will open up with a blank page on it.



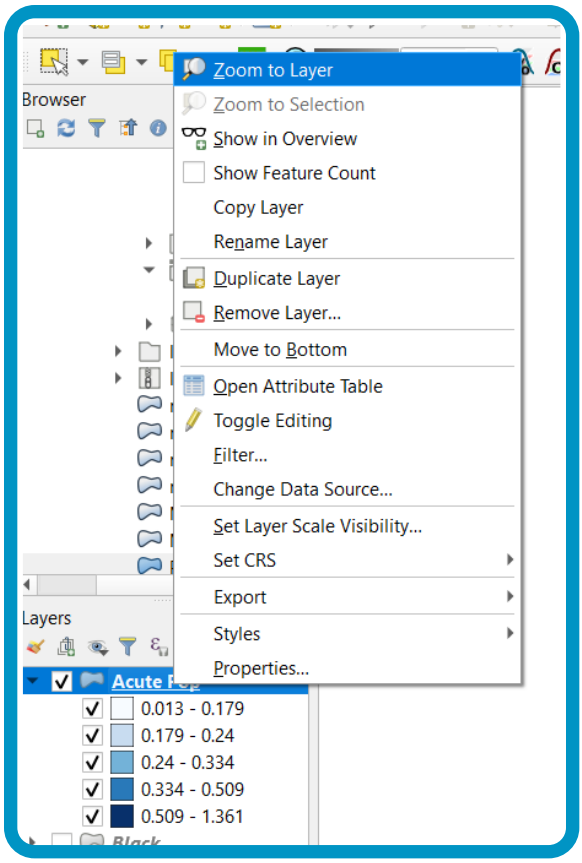
4. You can change the orientation of the blank page by right clicking on the blank page and choosing page properties. You should think about the orientation of the page based on the verticalness or horizontalness of your area of interest. Since Milwaukee is longer than it is wide, we'll change the orientation to portrait.



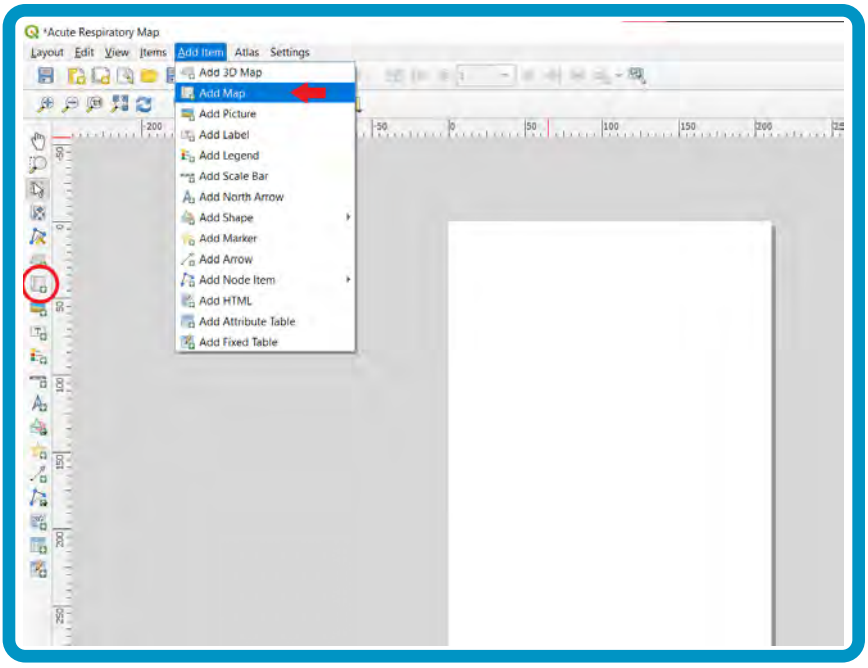
5. You'll need to add your map of interest on the blank page. First, you need to go to your QGIS project where you already see the map (like the window on the right). Make sure that you have created the symbology for the layer you are interested in creating a legend for and make sure that is the only map that is turned "on" in the Layers window by having a black check mark next to it.

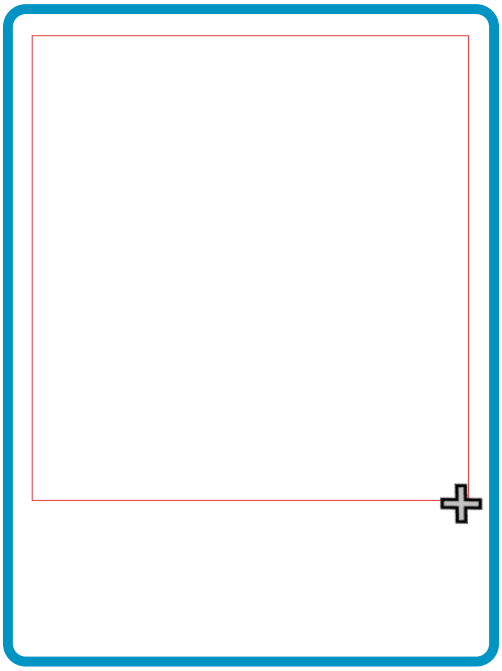


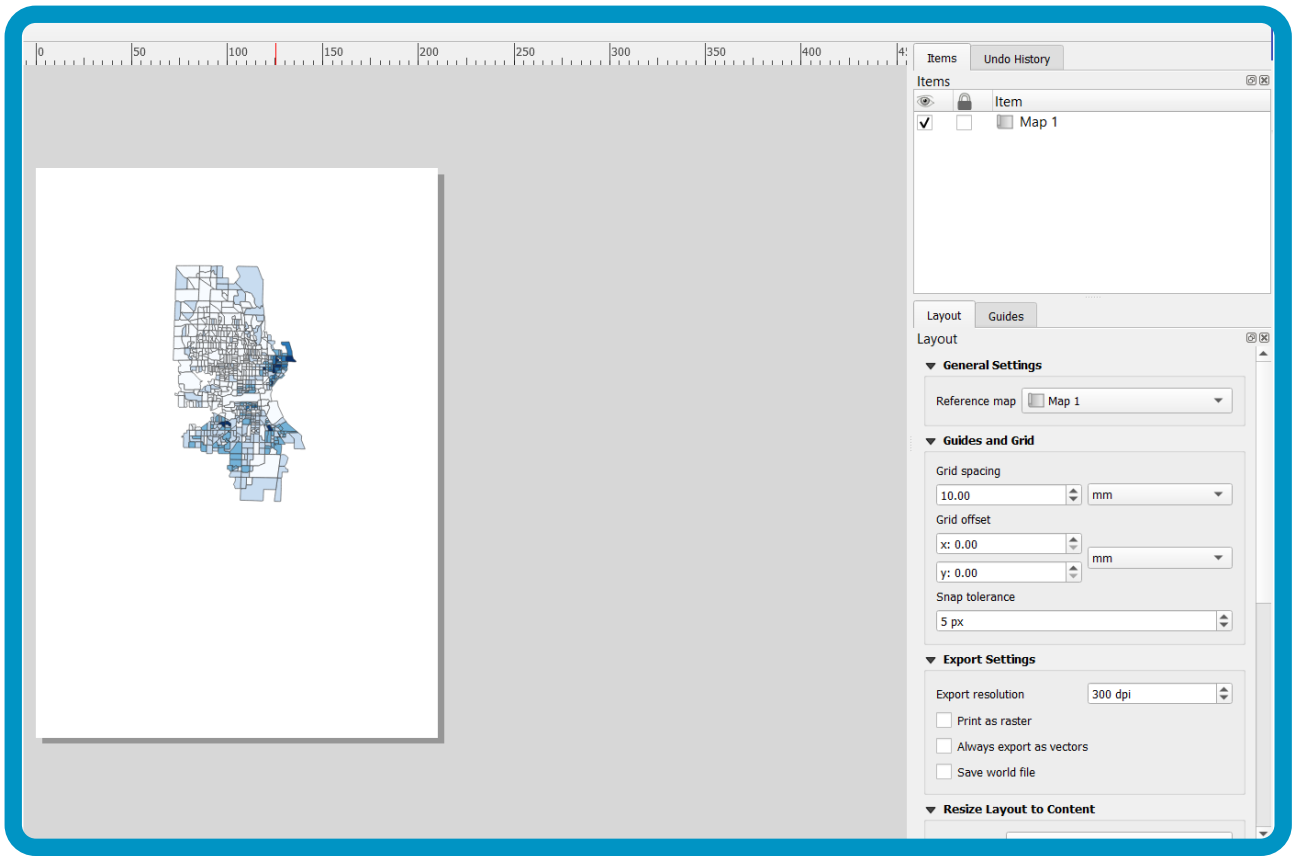
6. Now make sure that the map in the frame is centered. This will make it easier for you to manipulate the map size in the print layout. You can make sure it is centered by right-clicking on the layer and click "Zoom to layer".



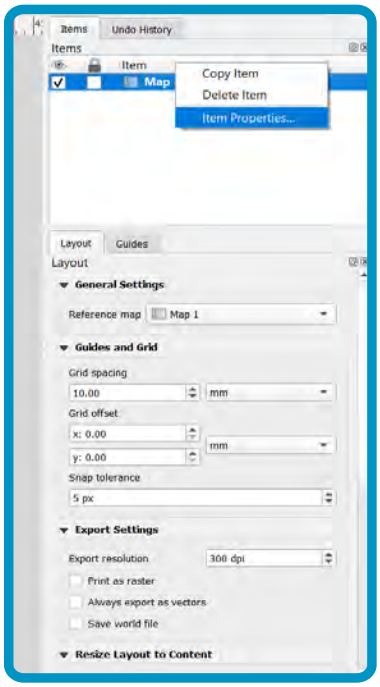
7. Turn to the Blank Print Layout page. You can now add the map. Click on "Add Item" at the top and then "Add Map" or if you have the sidebar visible, click on the "Add Map" button. A little cross will show up on the page and you need to drag it from one corner of the blank page to the other to create the map. Be careful to not drag too fast as it may zoom you too far in or out, making it challenging to maneuver the layout. QGIS may be more sensitive to mouse scrolling than usual and can therefore be finicky.





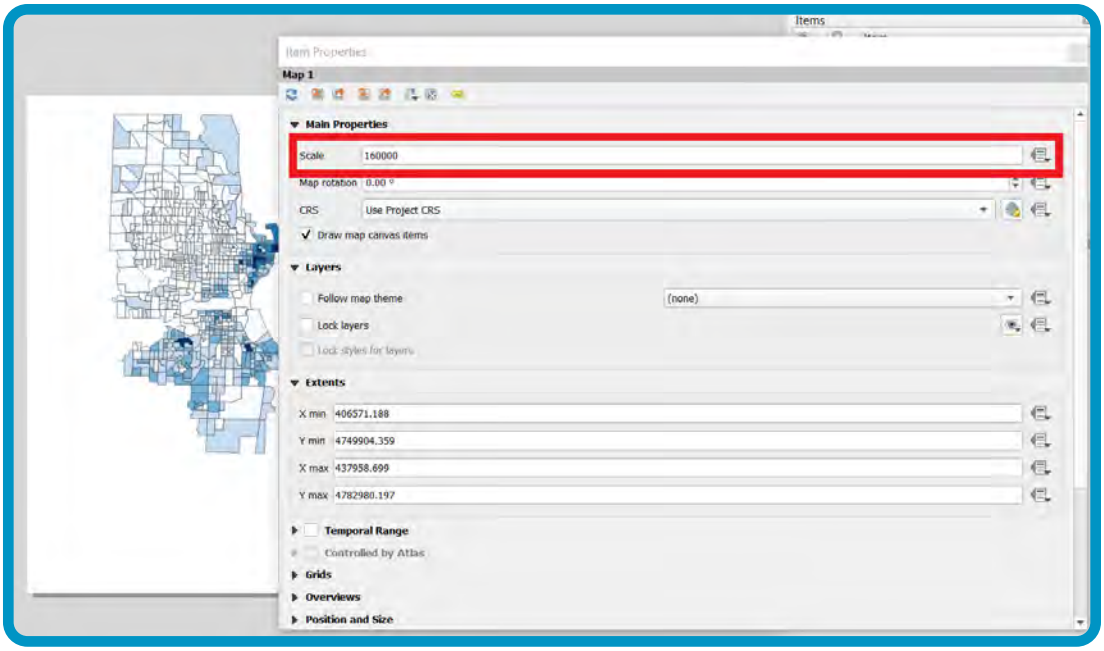


8. You'll likely want to make the map larger (or scale it up in size). In the Items window, right click on Map 1 and choose "Item Properties". *If you are using a Mac, this setting may not apply. Instead, click on the item (the map) in the layout window and the item properties will appear in the right tab. This same step applies to all of the other items you will add in the following steps.



9. A new window pops up and under "Main Properties" you can zoom into and out of the map. Where the window says "Scale" type in a value (try: 160,000) and press the enter button on your keyboard. By putting in a smaller number, you zoom into the map. By putting in a larger number, you zoom out of the map. You can play around with what value works best to fit the page. If you are unhappy with the zoom values you try, you may want to return to the Layers tab of the QGIS window, right click on your layer and click "Zoom to Layer" as we did in step 6. Then you can close the Item Properties window by clicking the x in the corner.

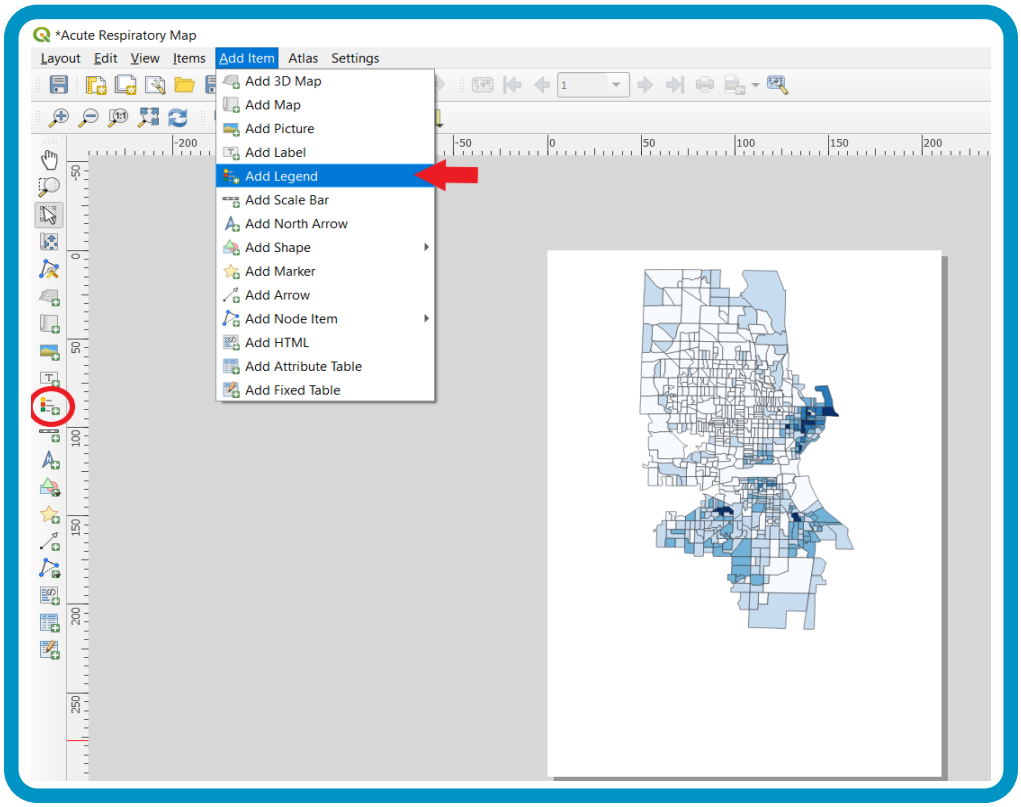
    a. Your map may appear blurry in the Layout, but once exported as an image or PDF the map will become clearer.



Adding legends

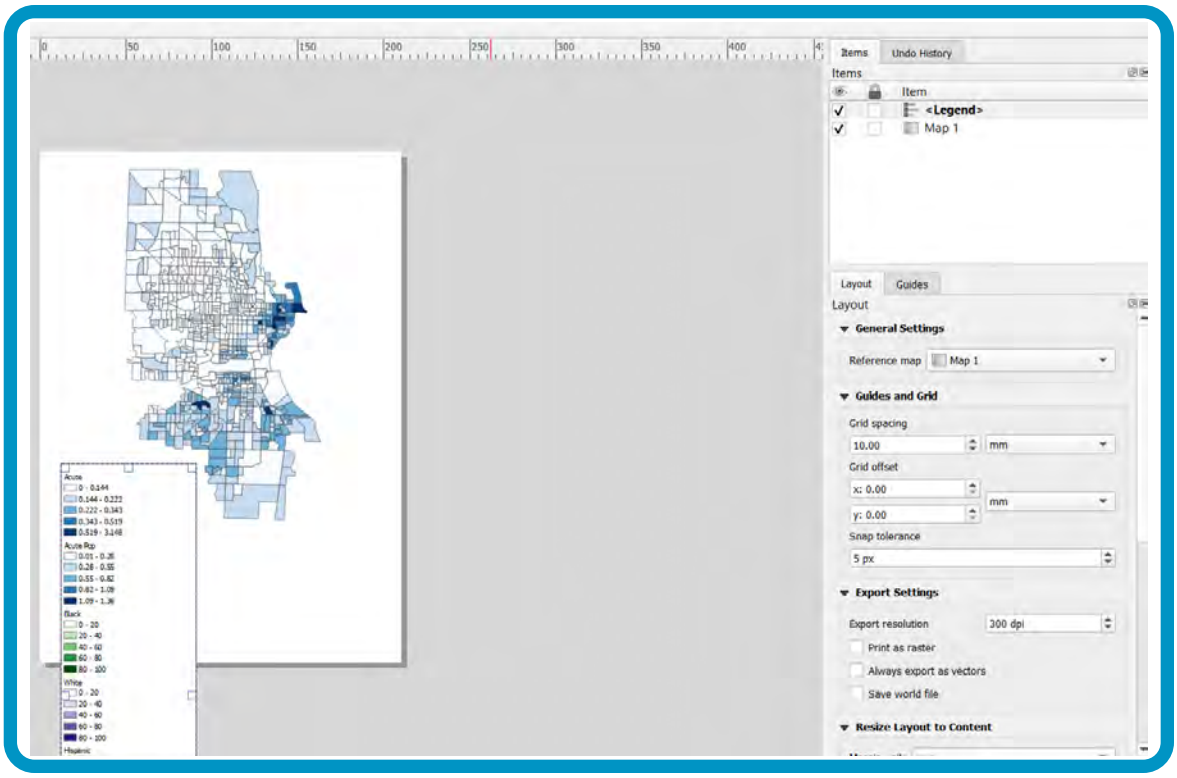
Legends are a necessary component of making a map understandable to an audience. The legend you create and add to the map is based on the symbology that you choose to visualize your data. Each layer you include in your map will be labeled in the legend and it can be organized in a variety of ways. The best maps should be understandable to anyone, even someone who does not have a background in data or the topic at hand. Making your maps legible to the widest range of people is particularly important when advocating for greater procedural justice. Complicated, wonky mapping and data sharing practices can make maps and data difficult to understand and can lead to issues of injustice, making it all the more important that maps intended to advance environmental justice are clear and comprehensible for any reader.

10. To add a legend, you'll go to "Add Item" and click "Add Legend" or choose "Add Legend" from the side bar.

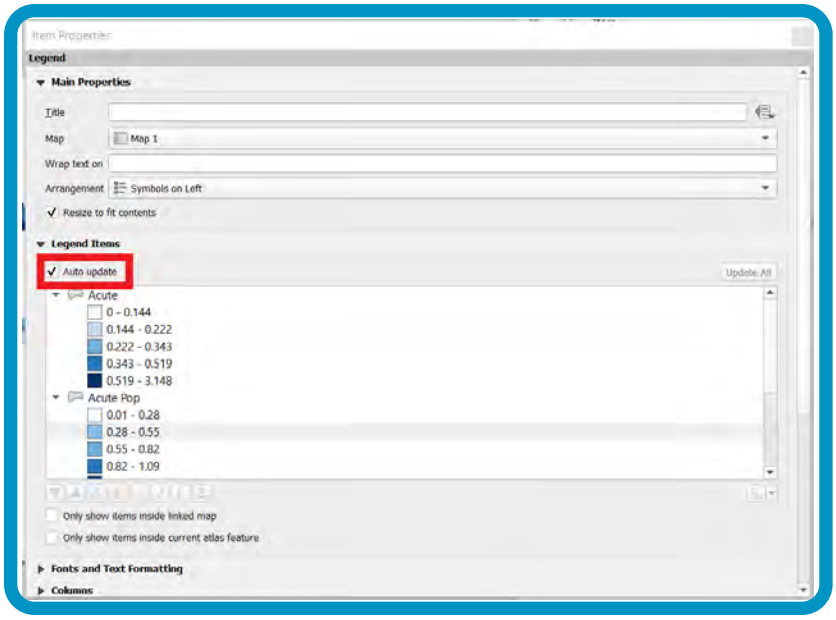


11. Similarly to how you added a map, you drag the mouse to show where you want the legend to go. Then you'll let go and a legend will appear.

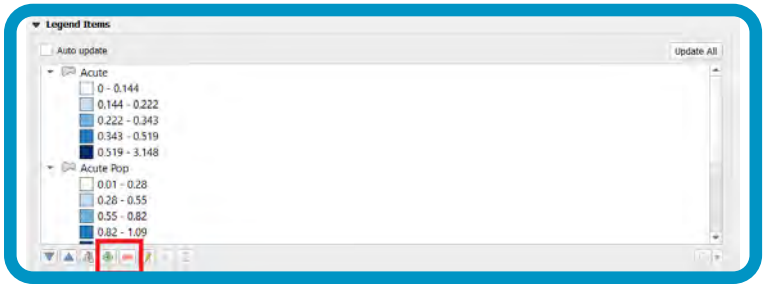
12. You'll notice how the legend in the picture includes the legend for all the layers that were in the QGIS file, but the map on this page only shows one map. So, you need to edit the legend to show just the legends you want to share so people are not confused. To do that, right click on "Legend" in the Items window, and click on "Item Properties"



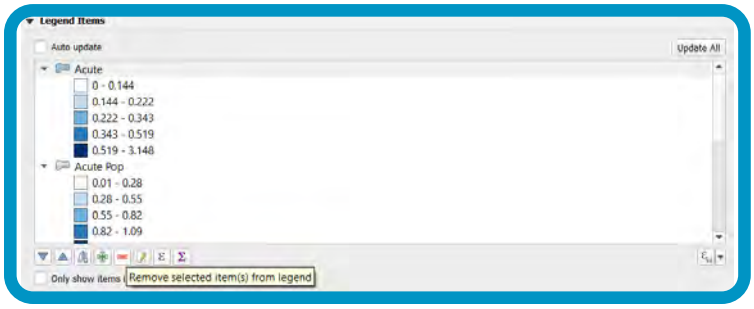
13. In the window that pops up, you'll focus on the Legend Items section. Right now, you'll notice it has the "Auto update" checked on.



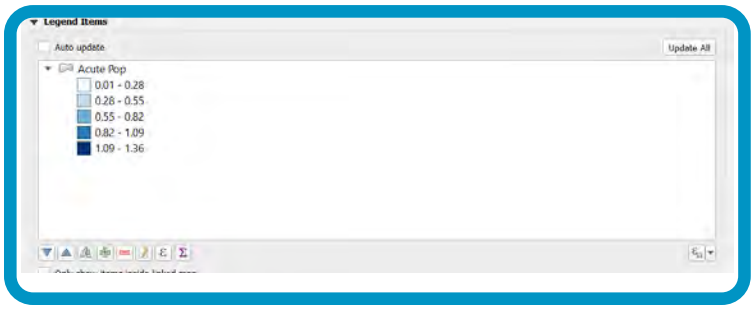
14. First check it off, and then you'll see the buttons below light up. This allows you to edit the legend. We'll focus on the red minus "-" sign to delete legend items. The "+" sign is if you want to add a legend item. A shortcut is also to click "Only show items inside the linked map" to remove them all at once, but for more customized deletion use the red minus button.



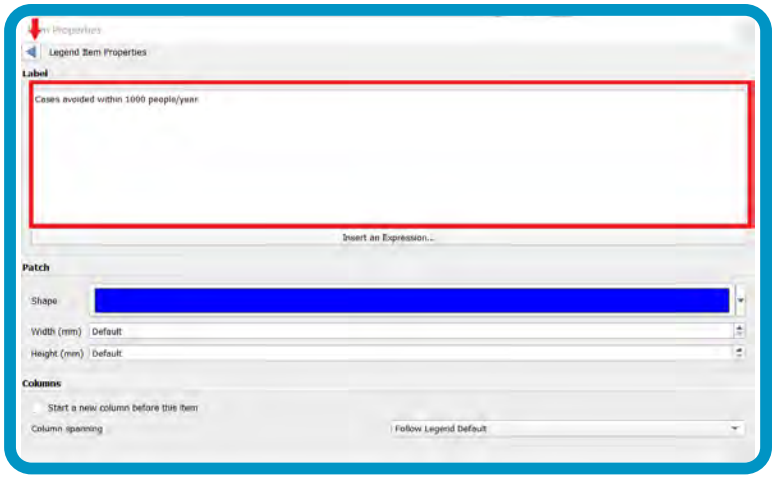
15. To delete a legend item, highlight the item you want to delete by clicking on the name and then click the "-" sign.



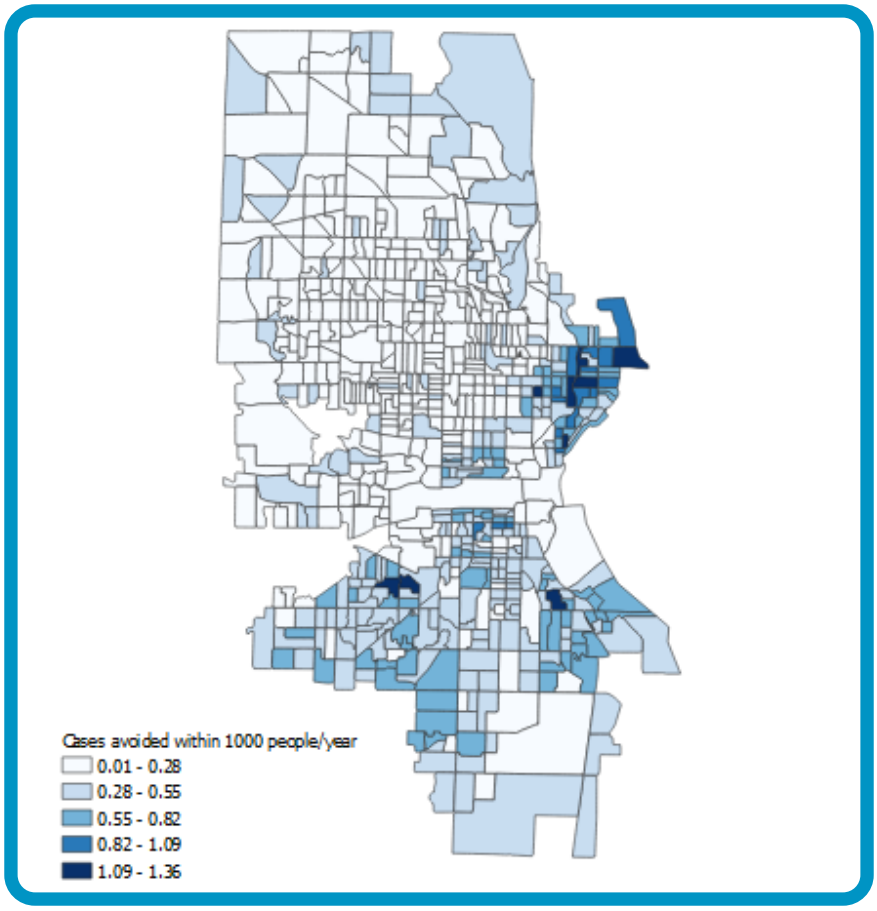
So it looks like this:



16. Now, we want to make sure that people who view the map can understand what the legend is referring to, so we need to change "Acute Pop" to a unit (or label) people will understand. Double click on "Acute Pop." The window will change with a bolded Label at the top. Type in the unit that is appropriate. Then click on the blue triangle in the upper left hand corner to save the change.

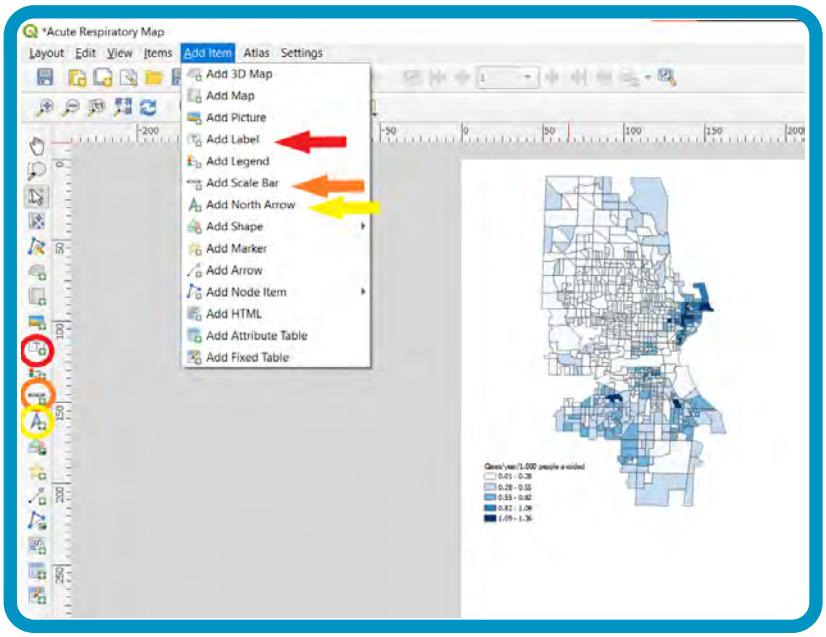


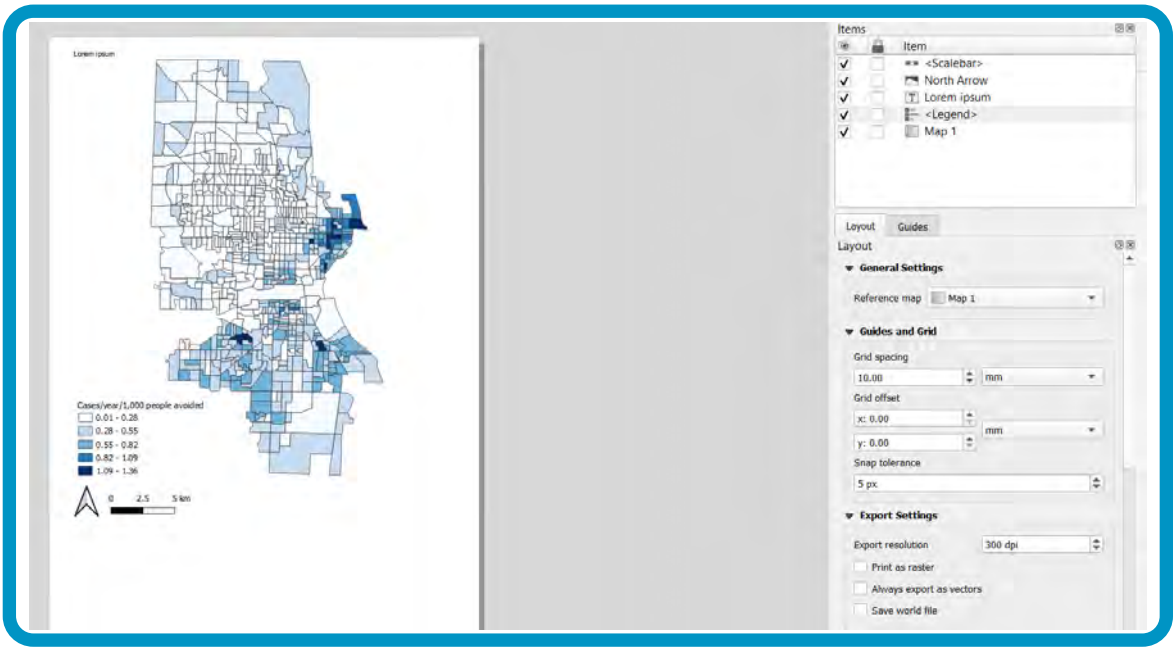
17. Now you can close the "Item Properties" window and you see the map with a more appropriate legend.



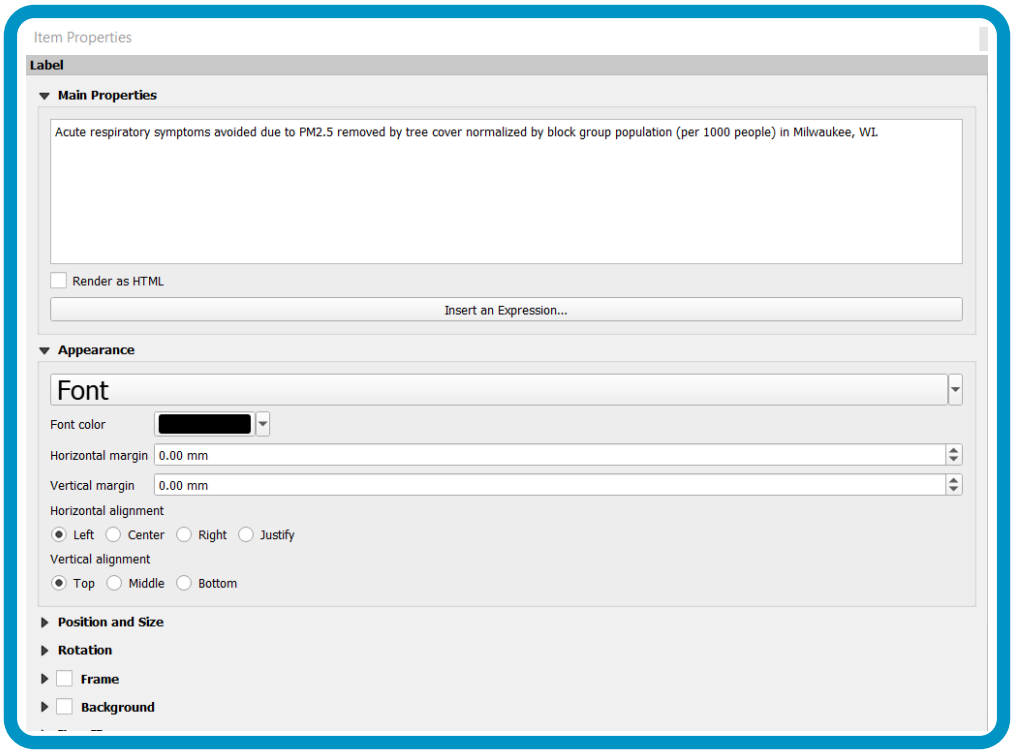
Adding Additional Context to the Final Map

18. Notice how the map still doesn't make a whole lot of sense to the map viewer. The map viewer knows what the colors on the map mean in terms of unit but not what the map is about. We need to add a title. So, go to the "Add Item" drop down again and click "Add Label". Drag the mouse for where you want the title. Additionally, you can "Add North Arrow" and "Add Scale Bar" so that you'll have them ready for the final map. (Many of these items can also be added from the left sidebar.) These are important to orient the map viewer. Similarly, drag the mouse to place the north arrow and scale where you want them.



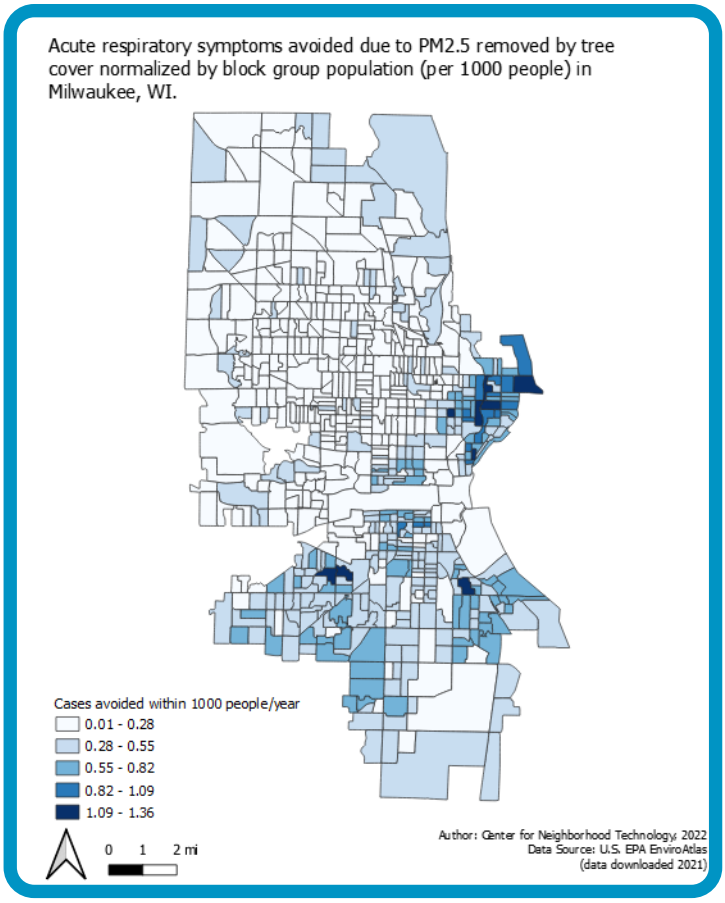


19. To change the title, right click on the Label in the items window and click on item properties (just like you had done for the map and legend). In this window you can change the text, change the font, font color, etc. Once done, click on the "x" in the corner.



20. Once you type your title, adjust the location of everything on the page so that it is easy to read. You can also adjust the scale bar from km to miles by going to the item properties. (Note: for Mac users there has been a bug where the scale bar is not visible but after a few weeks it appears again).

21. It is also helpful to add a note about sources on the map. Add another label and specify the data source and author of the map.



22. Note there are a lot of ways to make maps that orient the map viewers more, explore the different aspects on the Print Layout page and do additional research to find out more. For example, you can add multiple pages to your layout by going to "Layout" in the top toolbar and clicking "Add Pages..."

23. To print the map, click on the print button at the top.

Interpreting

Now that you've created the map, you need to interpret the map so that viewers know what the point is that you are trying to make. To interpret your map, look closely at what sorts of conclusions can be made from the patterns shown on the map and translate those conclusions into text. It is important to be intentional about which groups and calculations you are centering in your story through the data and resulting map.

When interpreting any map, be aware of the uncertainties in the data sources you chose and the limitations of the conclusions you make from the map analysis. The map you have created in this toolkit does not calculate if the differences between locations is statistically significant. Data scientists consider statistical significance to determine what is the likelihood that the difference is because of random chance instead of intentional decisions. Spatial statistics (creating statistically relevant maps) is an advanced level of GIS. The map you've created is a great first way to visualize the issue you are concerned about and can be used in campaigns to make your community, funders, and elected officials aware of the problem. You just need to be careful with how you word your interpretation, using terms like "it appears" or "we see a greater/lesser concentration here" instead of "it is" or "it is greater here or lesser here."

What needs to be included in a map interpretation? Look at the following tips and examples for some basics.

- **Consider how comfortable your audience is with maps.** You may need to break down components of the map. For example, you may need to orient them with north, east, south, west on the map (using the arrow) or how to read the legend. Also, let them know where you got the data from. For example, this is a map of Milwaukee, WI with the top of the page referring to the north. The legend shows that as areas are colored in a darker blue, more acute respiratory cases are avoided per 1000 people per year.

- **Consider the classification method you chose and how to explain that to the reader.** For example, in this map, we used equal intervals. This means that the range of values (from the minimum 0.01 to 1.36) was split into 5 equally sized buckets.

- **Point out where you see the minimum and maximum values.** For example, the areas with very dark blue values (the highest values) are along the coast of Milwaukee in the center. Along with two additional locations in the lower west and east sides. The largest area with the lightest blue values (the lowest values) are in the north center of Milwaukee and south center Milwaukee. Since map viewers may be comfortable with geographies in Milwaukee you can call out specific neighborhoods these block groups represent to help orient map viewers.

- **Explain what different values mean, as it relates to the data and the story you are trying to answer.** For example, the darkest blue values along the center coast of Milwaukee and the block groups in the lower west and lower east sides appear to experience the greatest reduction in acute respiratory symptoms (like nasal congestion, wheezing, chest pain). The areas that appear to experience the least reduction in acute respiratory symptoms appears to be in the north center of Milwaukee and a part of south center Milwaukee. Again, use neighborhood names as appropriate.

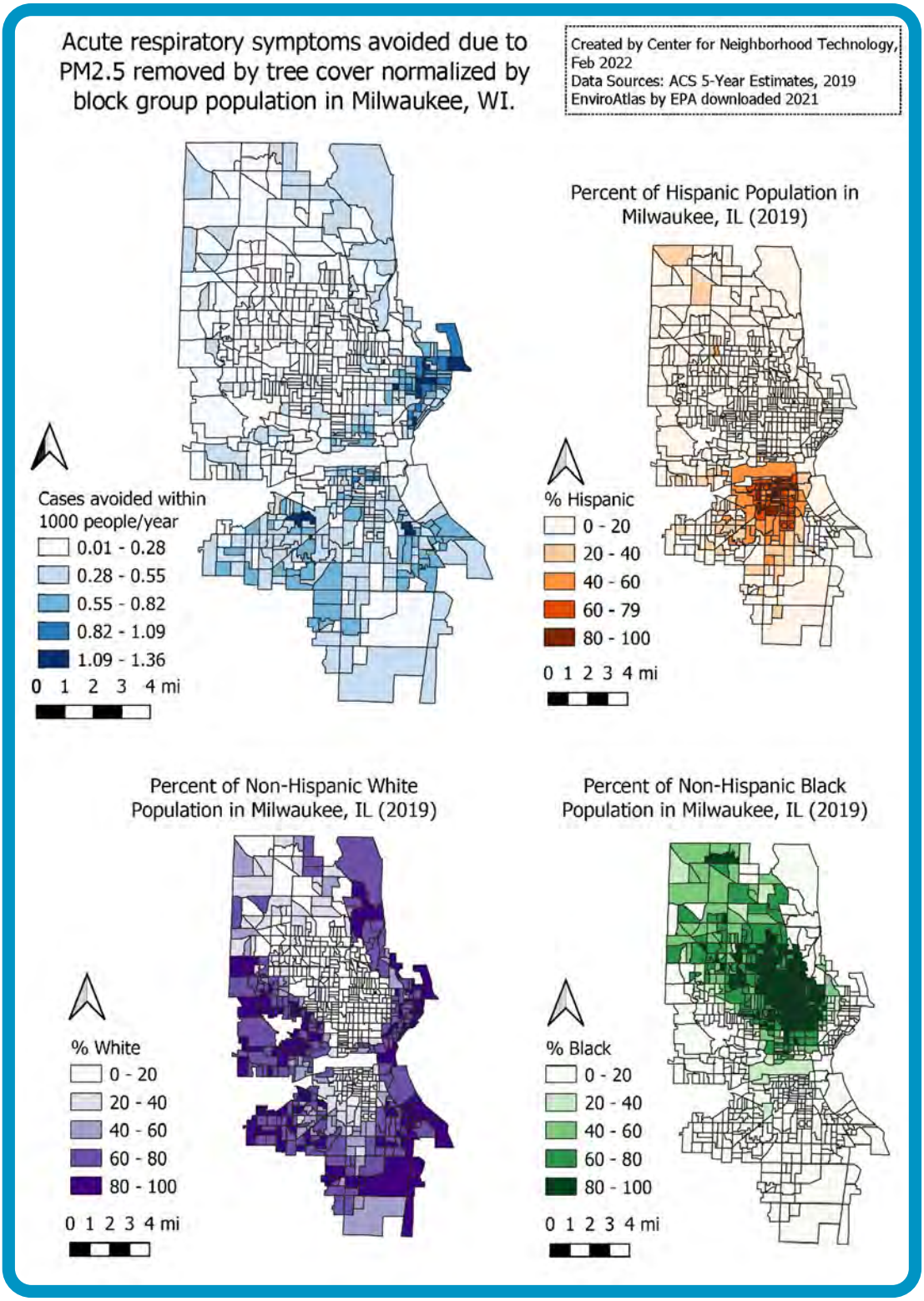
- **Refer back to the ultimate question you are trying to answer.** For this tutorial, we wanted to answer: Where in Milwaukee would people benefit from additional tree plantings or green infrastructure? So you might say, based on this map, it looks like north central Milwaukee and portions of south central Milwaukee would benefit from additional tree plantings.

- **Provide high level focus.** Adding a table with summary values from the attribute table or interesting values from the attribute table can help highlight key points.

- **You can also include qualitative data that you've received from community members that the map either complements or challenges.** For example, you might have pictures of trees in various neighborhoods to add to show the difference in trees that would support the results of the benefits or you have community projects you want to highlight to show the assets being built in the community to improve their experiences. Or maybe, there are stories about pollution in certain communities you want to add to give context to why more benefits need to be created in a certain area to counter the industrial pollution.

- **Consider your phrasing.** When you are writing out the analysis and evidence consider how you phrase the interpretation to empower marginalized communities instead of blaming them (Lakeshia Wright, https://www.lakeshiawright.com/thoughful-data).

Based on what you've learned here. How would you interpret the maps on the following page? First you may want to interpret each map. Then consider how the maps relate to one another. Since these maps talk about race, it'll be especially important to consider how you phrase the interpretation to empower marginalized communities instead of blaming them.



Ex: The maps above show that Milwaukee is a segregated city. Majority, non-hispanic, white populations tend to live near the borders of the west, south, and east block groups (Map C). The majority, non-Hispanic, Black population appears to be clustered towards the north center and northwest area of the city (Map D), and the majority Hispanic population is mostly in the south center of Milwaukee (B). There does not appear to be much overlap of the communities, except for the northwest corner which may have a similar number of non-Hispanic, Black and Hispanics people living.

Map A also shows that more cases of acute respiratory symptoms avoided due to PM2.5 removed by tree cover per 1000 people per year occur along the center east border and some southwest and southeast block groups.

The combination of maps seem to indicate that cases of acute respiratory symptoms due to tree cover are more avoided in areas that are predominantly non-Hispanic, white. Since this layer is based upon the tree cover of the area, these maps indicate more matured and larger trees in non-Hispanic, white communities. This is likely due to the history of higher municipal investment of tree plantings in white, greenlined communities, compared to predominantly Black or brown, redlined communities.[1] Though these are initial findings, these maps display that Black and brown communities of Milwaukee likely experience fewer of trees' benefits than white communities, and so deserve more funding to support investments in tree plantings in conversation with community members.

1 Locke, D. H., Hall, B., Grove, J. M., Pickett, S. T. A., Ogden, L. A., Aoki, C., Boone, C. G., & O'Neil-Dunne, J. P. M. (2021). Residential housing segregation and urban tree canopy in 37 US Cities. Npj Urban Sustainability, 1(1), 15. https://doi.org/10.1038/s42949-021-00022-0

E. Additional Tools in GIS

In addition to the tools and steps mentioned thus far in the toolkit, the following section will provide a high level overview and reference to an appropriate tutorial on how to conduct five additional tools often used in GIS analysis. As with all GIS analysis it is important to make sure all the layers you are using are in the same projection.

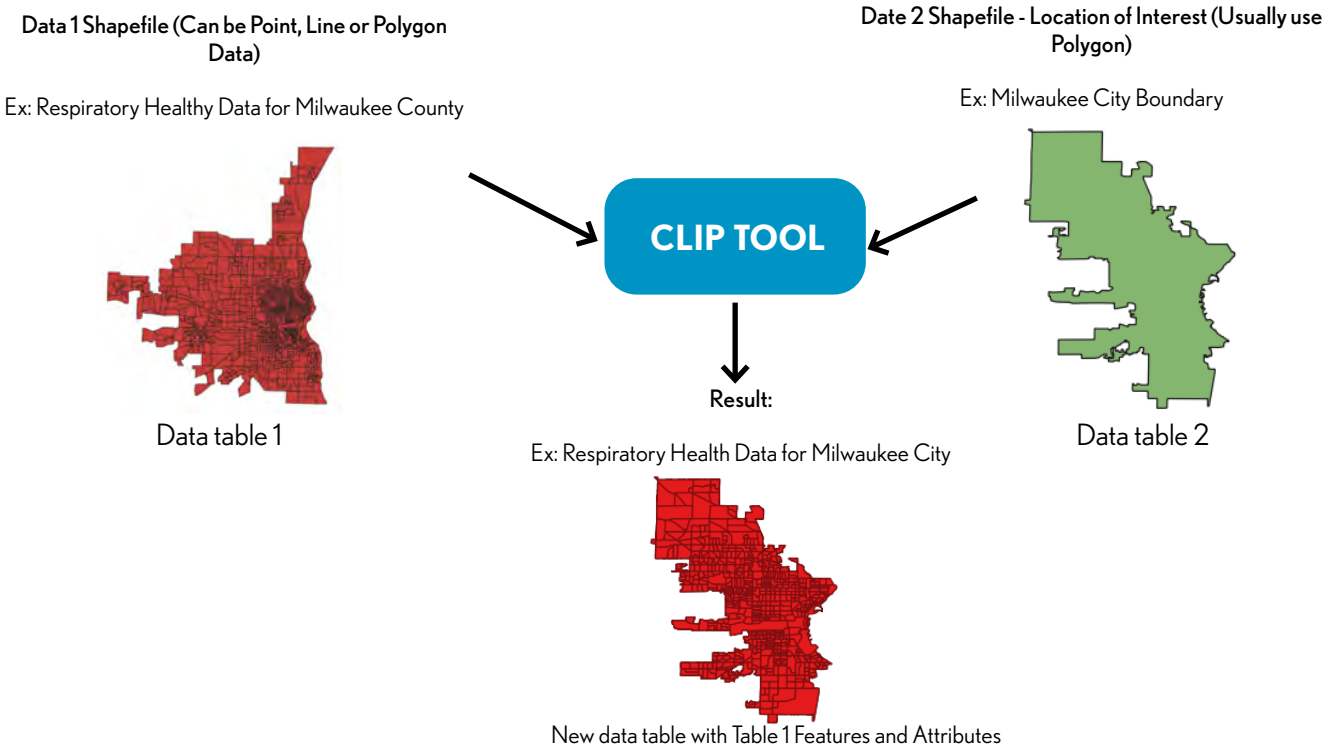
Specifically you'll learn about the following tools and are provided links to external websites for tutorials

- Clip
- Intersection
- Buffer
- Merge Vector Layers
- Join Attributes by Location (Spatial Join)

When working with GIS toolks, remember to frequently save your work.

Clip

Use this tool to focus the data on the location you are interested in, by "cutting" away the areas you are not interested in. This is useful when the data downloaded covers a much larger area than the location that is being focused on for the map. Clip does not join attributes from both layers, it only cuts the original data layer to the shape that is desired by using the extent of another layer. It is important that both layers are in the same projection to have accurate data for the location of interest saved.



Tutorial:

- Video: QGIS 3: Clipping by Eric Brelsford https://youtu.be/Rj8J_6pV620
- Text: QGIS Introduction: Clip Data by Duke University Libraries https://guides.library.duke.edu/QGIS/Clip

Intersection

Use this tool when there are two layers you are interested in and you want to end up with a data layer that includes the attributes of both layers and features that are the intersections of both layers. This means that the features in the final data layer will likely not be the same as either of the first two layers (unless the boundaries of one fit nicely into the other). Rather, the final data layer's features' boundaries are determined by how the two original layers intersect one other. For example if you intersect two data layers of roads, the intersect tool would create a layer of points because when two lines intersect, they intersect at a point. A block group intersected with a watershed would result in a layer that has a single block group broken up along the watershed boundaries.

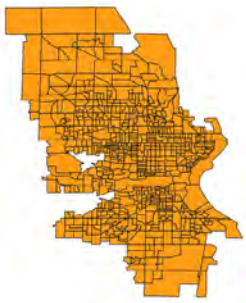
Data 1 Shapefile (Can be Point, Line or Polygon)

Ex: HUC-12 Watersheds Near Milwaukee

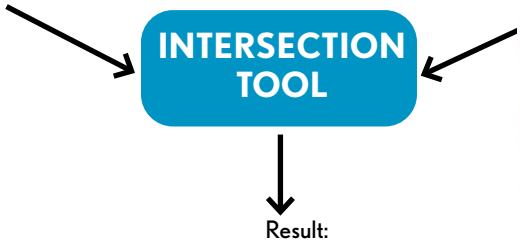
Data table 1

Data 2 Shapefile (Can be Point, Line or Polygon)

Ex: Milwaukee Block Group Boundary

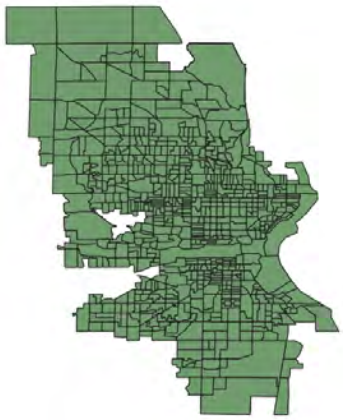


Data table 2



Result:

Ex: Milwaukee Block Groups Intersected by Watersheds



New Data Table with features that match the shape of the intersections. These features may not match Data 1 or Data 2 shapefiles. The features will have attributes from Table 1 and 2.

Tutorial:

- Video: GIS vector overlay tools: clip, intersection, union, (symmetrical) difference by Hans van der Kwast https://youtu.be/QBVv7h2Jhvo (from 2:00-2:55)

Buffer

Use this tool when you have a data layer and create a polygon whose boundary is a certain distance from the boundary of the data layer. For example you may use this if you have river data and want to make sure no development happens within ½ mile of the river or if you have points of air pollutant emitters and want to find out how many people are affected within ½ of the air pollutant emitter. You can dissolve (erase) buffer boundaries so they become connected or you can keep them undissolved so each separate feature has its own buffer. The buffer tool only works if the dataset is projected (not in an unprojected projection).

Data 1 Shapefile (Can be Point, Line or Polygon Data)

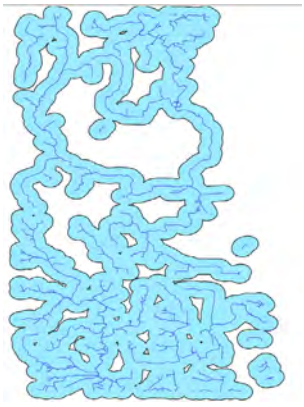
Ex: Named Streams in Milwaukee County

Data table 1

BUFFER TOOL

Final Buffer Data (Dissolved) with Named Streams layer on top

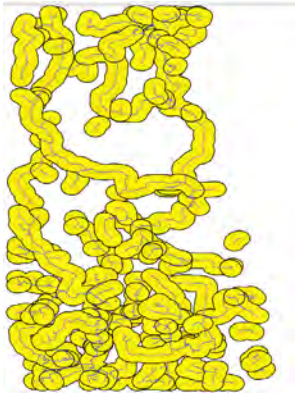
ex: 0.5 mile undissolved buffer of Milwaukee Streams



Data Table with 1 Feature, summing attributes from Table 1

Final Buffer Data (Undissolved) with Named Streams layer on top

ex: 0.5 mile undissolved buffer of Milwaukee Streams



Data Table with Table 1 Attributes

Tutorial:

- Video: Prepare your groundwater project in QGIS by Hans van der Kwast https://youtu.be/9aJdo6vLxeo (from 9:20-10:40)

  In this section of the video, Hans van der Kwast will walk you through the steps and application of the buffer tool in a case study where he is researching groundwater in the Shire Valley Alluvial Aquifer.

- Text: 3.5. Proximity Analysis: Buffers by Health Policy Project (USAID) https://www.healthpolicyproject.com/geoHealth/index.cfm?page=35

  In this tutorial, the Health Policy Project explains how to define catchment areas (the geographic area covered by facilities) for drop-in centers for HIV care in Dhaka using the buffer tool.

- Additional resource: Vector Spatial Analysis (Buffers) by QGIS 2.8 https://docs.qgis.org/2.8/en/docs/gentle_gis_introduction/vector_spatial_analysis_buffers.html

  This resource provides an explanation of buffers, variations in buffering, and different buffering techniques with some practice examples for the reader to try.

Merge Vector Layers

Use this tool when you want to combine GIS shapefiles that are for different locations into one shapefile for analysis. This may happen because you are downloading data from different sources (like different county websites that have streams data), a data source split up the data into individual geographies and you need to analyze multiple geographies together, or if community members are using an app to collect the information for different locations and you want to combine their respective shapefiles into one.

Data 1 Shapefile (Can be Point, Line or Polygon Data)

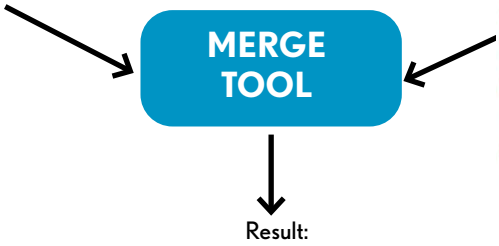
Ex: Named Streams in Milwaukee County

Data table 1

Data 2 Shapefile (Has to match Data 1 vector type)

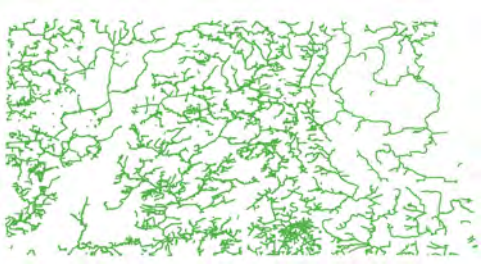
Ex: Named Streams in Waukesha County

Data table 2



Result:

Ex: Streams of Milwaukee and Waukesha County



New data table is created, where Table 1 is listed above Table 2. All features are listed, not all features have all attributes filled out.

Tutorial:

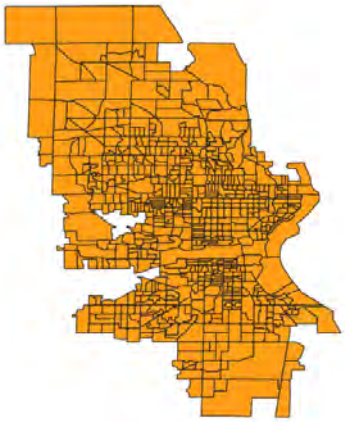
- Video: QGIS 2 Lesson 14b - Merging Vector Data from Different Layers by DMAD Marine Mammals Research Association https://youtu.be/TsormZNF238 (from 0:30-3:10)
- Text: How to Merge Vector Layers on QGIS 3.2 by Free GIS Tutorial https://freegistutorial.com/how-to-merge-vector-layers-on-qgis-3-2/
- Text: Merge Vector Layers on pg.961 of QGIS Desktop 3.16 User Guide (pg. 973 in PDF reader) https://docs.qgis.org/3.16/pdf/en/QGIS-3.16-DesktopUserGuide-en.pdf

Join Attributes by Location (Spatial Join)

Use this tool when working with two shapefiles and you want to add the attributes of one shapefile's table to the other's for further analysis. The end result is similar to the results of "Join Attributes by Field Value." However, you typically use this tool if the geographies of the two shapefiles do not have a field value that can match to each other, but they do geographically overlap. Unlike with Intersection tool, where the output creates new features based on where the two layers intersect, spatial joins keep all the same features as data layer 1 with the addition of attributes from data layer 2.

Data 1 Shapefile (Can be Point, Line or Polygon)

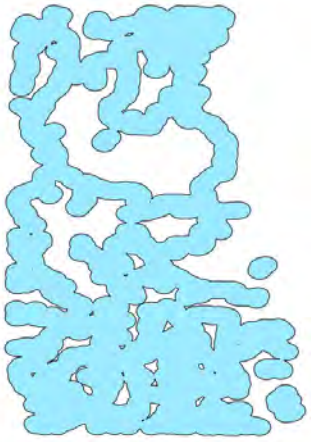
Ex: Milwaukee Block Group Boundary



Data table 1

Data 2 Shapefile (Can be point, line or Polygon Data)

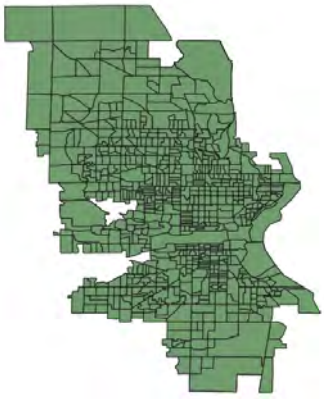
ex: 0.5 mile undissolved buffer of Milwaukee Streams



Data table 2

Result:

Ex: ex: Spatial join of 0.5 undissolved buffer of Milwaukee Streams to Milwaukee Block Groups



New data table is created, with all the features from Table 1 and the attributes of Table 1 and Table 2. Some features will have attributes from both Tables (where the two maps overlapped), some features will only have attributes from Table 1 (where the two maps did not overlap). In this example, block groups that are within the stream buffers will also contain attribute data from the stream buffer dataset, block groups outside of the buffers will not contain information from the stream buffer dataset.

Tutorial:

- Video: QGIS Join attributes by location by Q-tips https://youtu.be/b2KZqx1C0LU
- Text: Performing Spatial Joins (QGIS3) by QGIS Tutorials and Tips https://www.qgistutorials.com/en/docs/3/performing_spatial_joins.html

F. Explore GIS Analyses

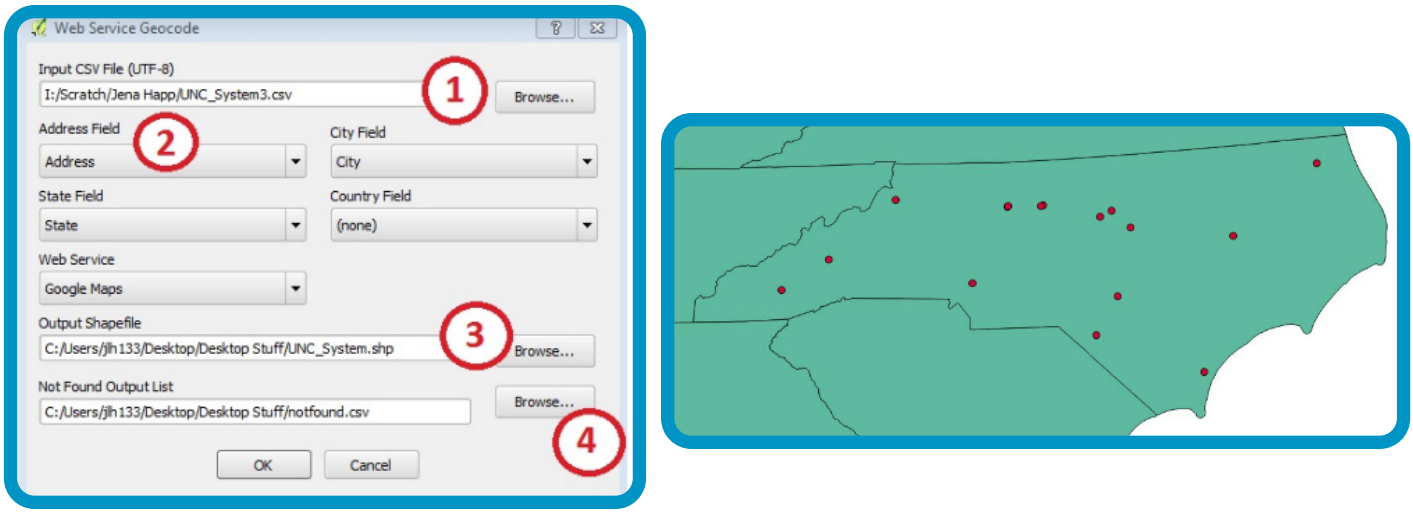
This toolkit introduces users to basic and intermediate techniques that can be undertaken in GIS. The power of GIS can be harnessed for more advanced operations such as creating indices, 3D models, LiDAR, spatial statistics and more. We briefly introduce modeling and indexing methods and a sampling of analytical processes in this section and provide links to tutorials that describe these functions.

Geocoding

Geocoding is the process of converting text-based descriptions (street addresses, zip codes, city name) of a location into geographic coordinates (latitude/longitude). The resulting coordinates identify the location as a point on the map. This can be useful if an EJ organization wanted to denote locations of certain community amenities or plot the location of problem intersections in their neighborhood.

In order to geocode multiple addresses, you will have to create a spreadsheet with columns for the address, city, state, and zip code to enable a geocoder to precisely locate the address on a map. Geocoders are tools through which the coordinates of a location are generated, and these tools are either web-based or desktop-based. Commonly used online geocoding tools, such as ArcGIS Online, US Census Bureau Geocoder and Google Geocoding API (Application Programming Interface that allows two applications to share information with one another) require users to upload a spreadsheet with complete addresses to maximize accuracy in plotting point locations.

Visit the linked resource to get a step-by-step tutorial on how to geocode in QGIS: https://guides.library.duke.edu/QGIS/Geocode



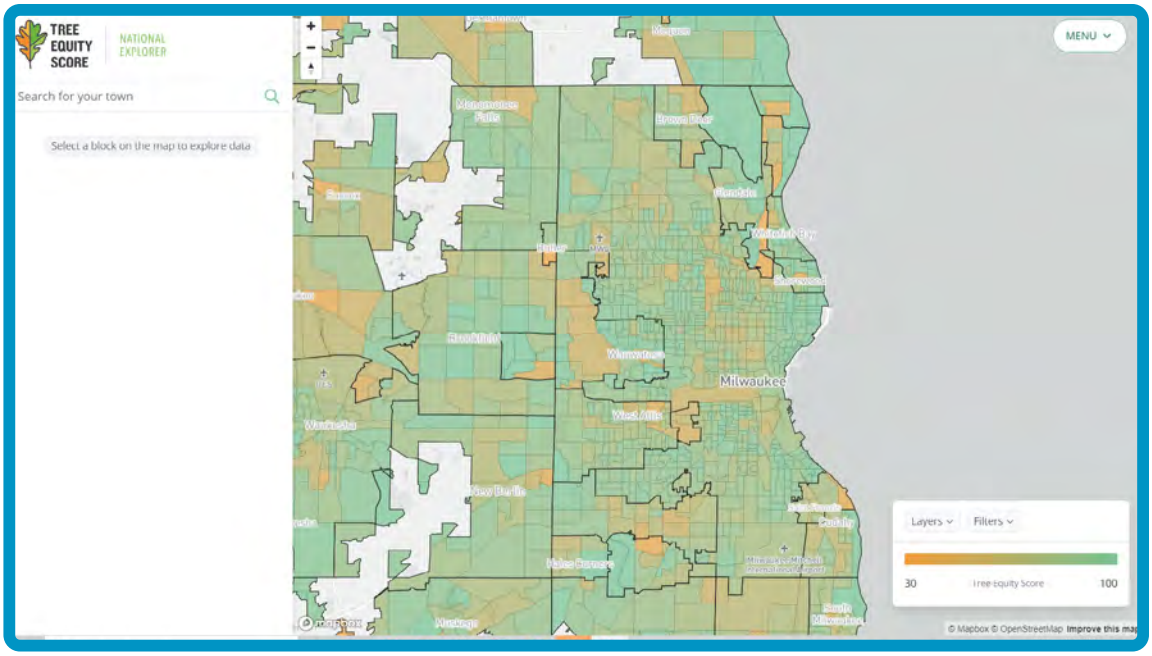
Source: Thomas, M. (2021, September 28). LibGuides: QGIS Introduction: Geocoding. Duke University Libraries. https://guides.library.duke.edu/QGIS/Geocode

Index

An index is a value, created combining several individual data points, that measures a certain concept on a scale (from 0-100, for example). Depending on what the index is measuring, higher values can mean a higher or lower measure of the concept. For example, Walk Score is an index that measures how walkable a neighborhood is in terms of proximity to amenities such as restaurants and stores. A higher value is indicative of a more walkable environment, and a lower value suggests a location has fewer amenities within walking distance.

Indices are useful in comparing how different geographies stack up against each other and provide an easy method for evaluation. To understand how indices are constructed, we will use the example of a tree equity index that measures presence of trees at a neighborhood level. The first step is choosing the data points that will make up the index (percent tree canopy, percent low-income residents, and percent people who are not non-Hispanic Whites). Next, we will select a desirable range for the index (0-10 or 0-100 are commonly used ranges) and assign a weight to each data point. All data points in the index are commonly weighted the same, meaning each data point makes up the same percent of the final index. In some cases, one or more data points in an index might be assigned a double weight because the map creator believes it is more important in the concept. In this example, each of the three data points contribute towards 1/3 of the index. The final step is to create the index - this can be done by summing the three data points, and then rescaling the summed value to the scale of the index.

The index can be visualized on a map by choosing the appropriate attributes and symbology. Refer to this link (Step 3) for a tutorial on how to create an index. The calculation of the index can be done in excel and joined to a shapefile.



Source: American Forests. (2021). Tree Equity Score. Tree Equity Score. https://treeequityscore.org/

LiDAR Analysis

LiDAR (light detection and ranging) is a remote-sensing technique that uses light waves to measure the distance on an object from the earth's surface. The technology has multiple applications and the most relevant use for urban planning is the ability to produce 3D representations of land elevations.

LiDAR sensors are mounted on aircrafts where they emit lasers to the earth's surface and the lasers bounce back when they encounter objects such as trees, shrubs, buildings, or ground. The data captured by the sensors can be processed in GIS to create digital elevation models (DEM) which can be used to identify watersheds, catchment areas, and flow paths. By mapping the flow of water through an area, we can predict the likelihood of flooding and identify mitigation strategies.

In this tutorial, the steps to extract the DEM from a LiDAR file are explained and to explore hydrologic (water flow) tools in QGIS, refer to this tutorial.

Spatial Analysis

Spatial analysis combines many input data layers to create new information that can help identify patterns such as suitability of a place for a certain activity. The analysis examines the relationship between multiple characteristics of a place, detects patterns and makes predictions for the best suited location using various techniques.

Defining the criteria of the analysis would be the starting point of the analysis. For example, our problem statement is to identify suitable locations for parks in the city. Some of the criteria that are important for this analysis could be 1) the location should be within a neighborhood with more than 60% impervious surfaces, 2) is within ¼ mile of a residence and 3) can be accessed by two streets. Using a variety of geoprocessing tools such as buffering, proximity, and intersection, the analysis can create a single layer to identify suitable locations for parks.

To learn how to conduct a multi criteria overlay analysis, refer to this tutorial.

Community mapping/Participatory mapping

Community mapping or participatory mapping processes involve residents of a community to record community experiences and knowledge in a spatial format. A commonly used example is the collective creation of asset maps by community members that identify physical and social resources in a community.

The goal of a participatory mapping process is unique to the community and issue being researched, but methods used to gather this information can be categorized into two large categories. Community-engaged mapping focuses on a particular geography (a neighborhood, a portion of a neighborhood or several neighborhoods) and community input is gathered through focus groups using a map as a way to record responses. Large, poster size maps are printed, and small groups of community members are requested to work together to identify assets and problem areas. A quantitative score (0-10) or grade (A-F) can also be assigned to indicate the quality of the assets. Social investigation research gathers responses from individuals through interviews, surveys, and community walks. These methods are more inclusive in gathering input, especially in communities where members are not comfortable attending large meetings and sharing information in a public setting.

Results from the community mapping engagements can be digitized using GIS and mapped to visualize input from the community. In addition, qualitative data can be analyzed to identify themes and categorize common experiences by community members. The qualitative and quantitative data can be incorporated in reports, presentations, and communication tools such as ArcGIS StoryMaps or Mapbox Storytelling to inform advocacy and organizing efforts by local EJ groups. To learn more about community mapping, please refer to the Participatory Asset Mapping Toolkit.

7

DESIGNING THE FINAL PRODUCT

In chapter 3, we talked about how to think about incorporating maps into a meaningful product. This chapter will use that outline to further expand upon how to develop the final product, which looks to the Water & Health in Little Village report as a template.

Before creating a final product, consider who your intended audience is and how you want them to be able to use the final product. This will determine what type of final product to create, whether it's a report, interactive webpage, article, poster, comic strip, etc. Work with your communications or outreach team to determine what product would make the most impact.

What should you include in your final product?

Context

What basics does your audience need to know in order to understand the gravity of the situation. Content to include in this section include a brief history of the area, especially about demographic and/or land use changes, the issues at hand, and why your organization cares about them (a brief history of your organization). This can be a space to share the assets of the EJ community, making sure that the readers understand how you view your own community and aren't just informed by stereotypes. For context in the Water & Health in Little Village report, see "Community Overview" on pages 4 and 5.

Maps and Discussion

Show-off the maps and create the narrative to make your point. Analyze them for readers. Think about what are the key takeaways you want readers/viewers to leave with and what evidence the maps provide to support this takeaway. Share any interesting points that show up on the map. For maps and discussion in the Water & Health in Little Village report, see "Issues Impacting the Community" on pages 6-11 and "Analysis of Survey Results" on pages 13-20.

> As you write this, refer to section 6.D for more details. Consider how comfortable your audience is with data and maps. If they're pretty new to maps, break down the components of the map. You may also consider meetings with the audience so you can dive into this, and focus on high level findings in the final product.
>
> When you are writing out the analysis and the evidence, consider how you are phrasing the data so that it empowers marginalized populations and puts the responsibility of improving the situation on needed policy changes not on individuals' own behavior. Be explicit about where and who the disparities are between. Lay your key takeaways out for your readers so they do not have to make assumptions. (Lakeshia Wright is a great resource on how to interpret data to empower communities and challenge structural racism).
>
> The discussion is also a space for you to intertwine community stories and history (the qualitative data) with the maps.

Literature Review

Explain the WHY this topic is important in depth. How does the WHY relate to the data results seen in the maps? Connect the historical and current policies that result in the data and how those are connected to environmental racism. This section sets you up to connect the data to the policy changes you propose in the conclusion. For a literature review in the Water & Health in Little Village report, see "The Link Between Health, Jobs, Water, and Economics" on pages 11 and 12.

Policy Recommendations

Based on the findings and the WHY, what are the calls to action in the policy world? Share what the organization is working on and are seeking to advocate. Explain how these changes will make the impact that your organization is looking for. If you have specific calls to action in addition to policy recommendations, include those. This section should consider your audience and what your original intention was for this final product. For policy recommendations in the Water & Health in Little Village report, see "Reducing Community Risks: Building a More Resilient South Lawndale" on pages 21 and 22.

8 COMMUNICATIONS STRATEGY

Now that you've completed the final product, get people talking about it! Each organization has their own communications strategy, the following are some thoughts to consider if you are unsure of where to start.

Developing a communications plan is centered on the audience and their purpose for knowing about the final product. If you've gone through a power mapping exercise, which identifies who has power in your context to create the change your organization desires, this can inform how many different audiences you have. The different audience types may need to know different amounts of information from the final product. Below, we'll cover an example of two.

Audience

Primary audience

These are people who have power. You want them to go through the final product in detail and take action. This audience may include community members, policy makers, or funders. For this audience, ask yourself the following question to help you with communications:

- What are the impactful ways that we've communicated with this audience member in the past?

Some examples of communication modes may include:

- Hosting a community meeting,
- Using the final product at a public comment, or
- Scheduling meetings to share findings from the final product.
- Maps and Discussion

Secondary audience

These are the people and institutions that you want to improve your reach to and may become potential new partners. The goal is to increase your network. The communications for this group may require further summation of your final product. For example:

- Share highlights over social media to be reshared
- A blog linked to the final product to be distributed via newsletter
- A presentation at a summit/symposium or through a webinar. Work through your current networks, technical partners, or funders to identify events to attend

These communications will get the final product out there for the general public to increase engagement with your organization either as volunteers, potential partners like other community organizations or technical institutes, or leads to other opportunities.

Another way to think about communication is to think about it as a journey. Do you want your audience to know about information at different points of the process? The initial conversations of the final product, a midpoint, and then the final product? This could be helpful if you think your process is just as important as the final product.

The most important part of developing communications is having a clear purpose in order to determine how to share the final product.

9

GLOSSARY

Aggregation - Changing the spatial scale of data from a higher resolution (ex: county) to a lower resolution (ex: state) by combining the smaller spatial units to form a single, larger spatial unit.

Attributes - What topic the data is referring to (like population size). In a data table, these are the column headings.

Break points - Where the map creator determines the different categories/classifications, not using the preset options

Comma Separated Values (CSV) - A comma separated values document which can be opened up in Microsoft Excel as a table. It is a way to save large tables that might otherwise have trouble opening in Excel.

Coordinate System - (longitude and latitude (the x,y location) measured in degrees)

Counter mapping - The act of creating maps that are counter to common practice and challenge power structures.

Data Dictionary - an organizational tool to record what data you use in the map

Data Set - Generic term for data which can refer to a map layer or data table that is not mapped

Data Table - Where the quantitative and qualitative information is stored for the map layer. Columns of the table refer to the attributes. Rows refer to features.

Data Type - the format in which data is stored and read by the software

Datum - a reference point geographers have identified to do the mathematical calculations

Demographic Data - Information about a person's/population's age, gender, race, income-level, education, and employment

Deviation - how different is a data point from the mean (average), often mentioned in context as standard deviation

Dissolve - To erase or remove, typically in reference to boundaries for buffers

Expression - Mathematical sentence that uses a math symbol. For example: = 50 or >400 or 12,000

Features - What specific location the information is referring to (like a specific county). In a data table, these are the rows.

Geocoding - process of converting text-based descriptions (street addresses, zip codes, city name) of a location into geographic coordinates (latitude/longitude)

Geographic Coordinate System (also referred to as Coordinate System) - the longitude (x) and latitude (y) coordinates of a location. The units are in degrees.

Geographic Information Systems - GIS; a tool to create, organize, map, and analyze spatial/geographic data

Geometric Predicate - Term QGIS software uses to explain what type of comparison will be used for spatial joins or other tools

Geographic Scale - the size of an area the data represented on a map refers to

Histogram - a graph similar to a bar graph that visualizes the frequency of something within a bucket of classes or values

Join - a GIS function in which you connect two data sets (two map layers or a layer and a data table) together based off a matching attribute in each data set

Land Use Map - a map of how a municipality decided what the land is used for. For example, is it used for parkland, residential, industrial, or commercial/store uses, etc.

Mean - average, add up all of the values and divide by the number of values

Metadata - information about the dataset often saved in another accessible document when data is downloaded

Methodology - decisions made to proccess and analyze the data

Models - using past data and research to create equations that can predict values into the future. Models hypothesize what the future could look like based on past information, so they are not completely accurate because future conditions could change. Models increase in credibility with peer-reviewed research

Modifiable areal unit problem (MAUP) - when the spatial unit of a data set changes, the visualization of the map can change depending on how you can analyze the data

Normalization - process refers to multiplying or dividing units by another unit, or factor, to get a new unit that evenly illustrates the values of all records or groups.

Pixels - the smallest unit that makes up a digital picture

Projection - The result of a geographic location being transformed from the 3D globe to a 2D map.

Qualitative data - data that provides information not using numbers (words, pictures, etc)

Quantitative data - data that provides information using numbers

Range - The span of the data from the smallest to the largest

Raster - when spatial data is made of pixels

Shapefile - A format in which GIS data is saved. It shows the geographic attributes of the data. Can typically be used across GIS softwares

Skew - opposite of even distribution, for example when data shows many features within a specific range

Source layer - terminology used in GIS software to denote the initial layer you are working within the specific analytical tool

Spatial attribute - location information attached to the rest of the data so that the data can be mapped

Spatial/geographic data - data that is tied to specific locations

Spatial relationship - how two layers relate to each other spatially, for example do the map layers intersect at spots, are they within a certain distance of each other, which features of one layer are completely inside of the boundary of the other layer

Spatial unit - the size of the locations within the data. For example, the data is about cities or city blocks, watersheds or lakes, etc.

Symbology - Customization of a map colors, symbols, letters, etc. to show the different data values across the location for the attribute of interest

Target layer - terminology used in GIS software to denote the layer that you are changing by using the source layer within the specific analytical tool

Thematic Map - Map that shows the differences in data values across the chosen geography

Transformation - since Earth is 3D, but we are creating a 2D map, there are multiple ways to change (project) Earth into a map. To change from one way into another is considered a transformation. Different transformations can keep the areas proportional or keep the distances proportional. Here is a video that further explores transformations: https://www.youtube.com/watch?v=wlfLW1j05Dg.

Unprojected Projection - When data is mapped in GIS using longitude and latitude with a unit of degrees, instead of a typical measure of distance like (feet, meters, etc).

Vector - when spatial data are stored as either points, line segments, or polygons

Zipped folder - to make it easier to move and download multiple files at a time, computers can "zip" a folder to store all the files within it. You will then need to "unzip" or extract the files to access each of them individually. On PCs this will require a software like 7-zip, or on Macs, you can usually just double click on it.