Public Health GIS

# Lab 3 Projecting maps, Geocoding and Symbolizing Point Data, and Spatial Joins

PART 1 – Projecting maps with QGIS

In this exercise, we will project cartographic boundary files in QGIS. The process can be confusing, especially when we are first starting out with a software. We will walk through how to navigate and work with data in different projections in QGIS.

Projecting a map of US States and changing projections
Open QGIS.

1. Open **US_states.shp** file inQGIS.
2. To check the coordinate reference system (CRS), right-click on the layer and select **Properties**. In the Layer Properties window, select the **Source** tab in the left-hand column and check under Assigned Coordinate Reference System (CRS) or **Information** tab and check under Coordinate Reference System (CRS).
3. You can see that our current CRS for this layer is the **EPSG:4326-WGS 84** geographic coordinate system.
4. There are other coordinate reference systems we can use to visualize our data. To change the CRS, click on **EPSG: 4326** at the bottom-right hand corner of your project window.
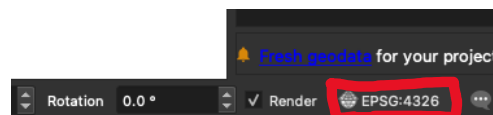
5. One projected CRS we can use is the Pseudo-Mercator projection. In the **Filter** box, we can search mercator and select WGS 84 / Pseudo-Mercator EPSG:3857 from the **Predefined Coordinate Reference Systems** drop-down menu and click **OK**.
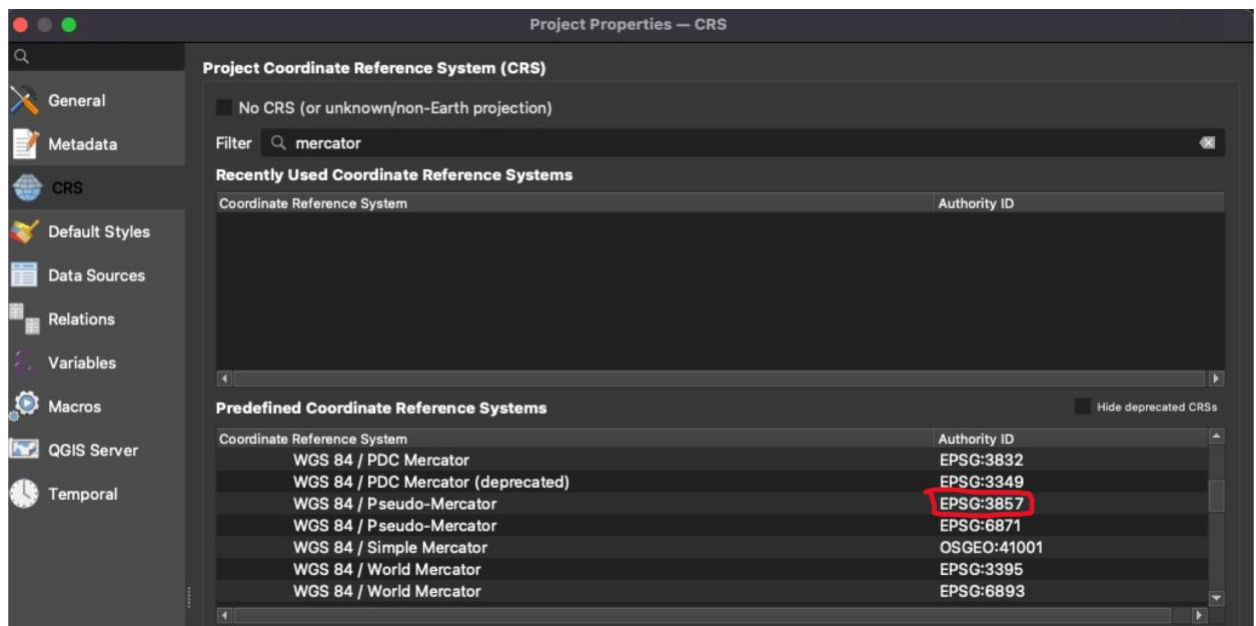
NOTE: QGIS is projecting "on-the-fly." The actual CRS of our layer has not changed from the original **EPSG:4326-WGS 84**. You can check for this by right-clicking on the **US_states** layer, selecting **Properties**, and selecting **Source** or **Information**.

6. Another projected CRS used commonly in the contiguous US is the Albers conic projection.

   Click on **EPSG: 4326** at the bottom-right hand corner of your project window. In the **Filter** box, we can search the EPSG code, 5070 and select NAD83/Conus Albers EPSG:5070 from the **Predefined Coordinate Reference Systems** drop-down menu and click **OK**.

   If prompted to Select Transformation for **US_states**, click **OK**.

7. This projection is still on-the-fly. If we want to permanently change the projection on the layer, we must save out the layer as a specific projection. To do this, right-click on the US_states layer and select **Save Features As...**
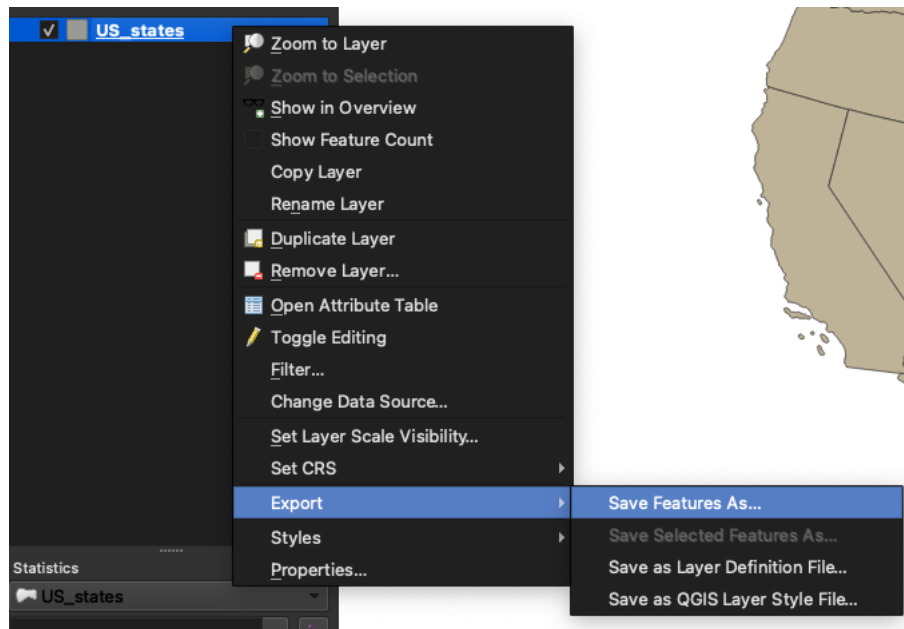
8. Under **File Name**, click the three dots on the right ⬚ and select your **Lab3 folder** as the location to save the file. You can name your file **US_albers** and press **Save**.
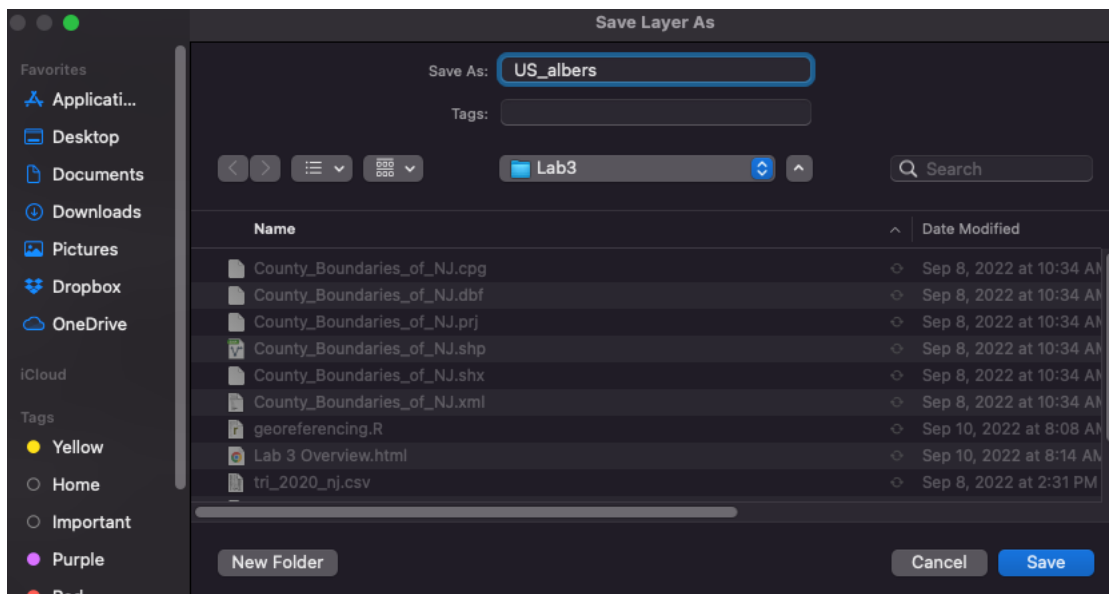
9. Under **CRS**, click the icon on the right 🖼 and select **NAD 83 / Conus Albers** from **Recently Used Coordinate Reference Systems** or search for it in the **Filter** box. Once your CRS is selected, press **OK**.

10. The new projected layer should be saved in your **Lab3** folder and show up as a new layer in your map.

11. You can check the projection of the **US_albers** layer by right-clicking on the layer, selecting **Properties**, and selecting **Source** or **Information**. This confirms that the geographic CRS in the original **US_states** layer has now been changed to a projected CRS in the **US_albers** layer.

Geometry	Polygon (MultiPolygon)
CRS	EPSG:5070 – NAD83 / Conus Albers – Projected
Extent	–2361582.2755981818772852,259071.7189404838427436 : 2263786.1910862838849425,3177424.9655306949280202
Unit	meters

REMINDER: To permanently project a layer to any given CRS, you must _save_ that layer in that specific CRS.

NOTE: QGIS automatically reprojects on-the-fly all layers in reference to CRS of first layer imported. However, the underlying CRS of multiple layers may not be the same, and to confirm you should always double-check the CRS for each layer in Layer Properties.

PART 2 – Geocoding and symbolizing point data in QGIS

In the second part of the lab, we will be using data from the Toxic Release Inventory (TRI) from the EPA to geocode, georeference, and visualize point data in QGIS.

TRI tracks industrial sites that work with and release certain toxic chemicals that may pose a threat to human health and the environment. According to the EPA, a release of a chemical means that it is emitted into air, water, or placed in some type of land disposal. For more details on the TRI, you can visit the TRI EPA website: https://www.epa.gov/toxics-release-inventory-tri-program/tri-data-and-tools

The goal of this portion of the lab is to take the addresses and coordinates of these TRI facilities and plot them in a known coordinate system and symbolize these points so that their sizes are proportional to the amount of total toxic emissions released from each sites.

Using the MMQGIS plugin to geocode addresses

QGIS plugins add additional functionality to what we can do in QGIS. Plugins are developed independently; the MMQGIS plug-in can assist us with many different important functions.

Examining TRI data in MS Excel

Navigate to the **Lab3** folder and open the **tri_2020_nj.csv** file in MS Excel. Scan through the data. Take note of the different columns, which include _facility ID_, _street addresses_, _city_, _county_, _state_, _ZIP code_, _latitude_ and _longitude_, and the _datum_ that displays the coordinates properly, the industry type and total amount of toxic release (lbs). Note that certain facilities have multiple sites.

facility_id	FRS ID	fac_name	st_address	city	county	state	zipcode	lat	long	datum	industry	total_release
07474PSSCR45DEM	1.10E+11	PASSAIC RUBBER CO	45 DEMAREST	WAYNE	PASSAIC	NJ	7470	40.904849	-74.262675	NAD83	Plastics and	1022
08872NWJRSNORTH	1.10E+11	CMC STEEL NEW JERSEY	N CROSSMA	SAYREVILLE	MIDDLESEX	NJ	8872	40.479262	-74.321207	NAD83	Primary Met	3
08872NWJRSNORTH	1.10E+11	CMC STEEL NEW JERSEY	N CROSSMA	SAYREVILLE	MIDDLESEX	NJ	8872	40.479262	-74.321207	NAD83	Primary Met	0.0413
08872NWJRSNORTH	1.10E+11	CMC STEEL NEW JERSEY	N CROSSMA	SAYREVILLE	MIDDLESEX	NJ	8872	40.479262	-74.321207	NAD83	Primary Met	4363
08872NWJRSNORTH	1.10E+11	CMC STEEL NEW JERSEY	N CROSSMA	SAYREVILLE	MIDDLESEX	NJ	8872	40.479262	-74.321207	NAD83	Primary Met	3619
08872NWJRSNORTH	1.10E+11	CMC STEEL NEW JERSEY	N CROSSMA	SAYREVILLE	MIDDLESEX	NJ	8872	40.479262	-74.321207	NAD83	Primary Met	1256
08872NWJRSNORTH	1.10E+11	CMC STEEL NEW JERSEY	N CROSSMA	SAYREVILLE	MIDDLESEX	NJ	8872	40.479262	-74.321207	NAD83	Primary Met	1
08872NWJRSNORTH	1.10E+11	CMC STEEL NEW JERSEY	N CROSSMA	SAYREVILLE	MIDDLESEX	NJ	8872	40.479262	-74.321207	NAD83	Primary Met	17760
08096GLFLL920KI	1.10E+11	GULF OIL LP WOODBURY TERMINAL	920 KINGS H	THOROFARE	GLOUCESTEI	NJ	8086	39.823006	-75.177026	NAD83	Petroleum B	445.99
08096GLFLL920KI	1.10E+11	GULF OIL LP WOODBURY TERMINAL	920 KINGS H	THOROFARE	GLOUCESTEI	NJ	8086	39.823006	-75.177026	NAD83	Petroleum B	5981.96
08096GLFLL920KI	1.10E+11	GULF OIL LP WOODBURY TERMINAL	920 KINGS H	THOROFARE	GLOUCESTEI	NJ	8086	39.823006	-75.177026	NAD83	Petroleum B	476.8
08096GLFLL920KI	1.10E+11	GULF OIL LP WOODBURY TERMINAL	920 KINGS H	THOROFARE	GLOUCESTEI	NJ	8086	39.823006	-75.177026	NAD83	Petroleum B	1074.95
08096GLFLL920KI	1.10E+11	GULF OIL LP WOODBURY TERMINAL	920 KINGS H	THOROFARE	GLOUCESTEI	NJ	8086	39.823006	-75.177026	NAD83	Petroleum B	1491.58
08096GLFLL920KI	1.10E+11	GULF OIL LP WOODBURY TERMINAL	920 KINGS H	THOROFARE	GLOUCESTEI	NJ	8086	39.823006	-75.177026	NAD83	Petroleum B	5951.39
08096GLFLL920KI	1.10E+11	GULF OIL LP WOODBURY TERMINAL	920 KINGS H	THOROFARE	GLOUCESTEI	NJ	8086	39.823006	-75.177026	NAD83	Petroleum B	334.78
08096GLFLL920KI	1.10E+11	GULF OIL LP WOODBURY TERMINAL	920 KINGS H	THOROFARE	GLOUCESTEI	NJ	8086	39.823006	-75.177026	NAD83	Petroleum B	1488.04
08096GLFLL920KI	1.10E+11	GULF OIL LP WOODBURY TERMINAL	920 KINGS H	THOROFARE	GLOUCESTEI	NJ	8086	39.823006	-75.177026	NAD83	Petroleum B	316.01
08096GLFLL920KI	1.10E+11	GULF OIL LP WOODBURY TERMINAL	920 KINGS H	THOROFARE	GLOUCESTEI	NJ	8086	39.823006	-75.177026	NAD83	Petroleum B	35.06
0811WCNNCT13JHN	1.10E+11	CONNECTOR PRODUCTS	1300 JOHN T	PENNSAUKE	CAMDEN	NJ	8110	39.98242	-75.04188	NAD83	Electrical Eqt	0
0811WCNNCT13JHN	1.10E+11	CONNECTOR PRODUCTS	1300 JOHN T	PENNSAUKE	CAMDEN	NJ	8110	39.98242	-75.04188	NAD83	Electrical Eqt	0
0882WLXFRM5BARB	1.10E+11	LUXFER MEL TECHNOLOGIES DBA MEL CHEMICAL-TREATMENT PLANT	500 BARBER	FLEMINGTOI	HUNTERDON	NJ	8822	40.50408	-74.96743	NAD83	Chemicals	250
0882WLXFRM5BARB	1.10E+11	LUXFER MEL TECHNOLOGIES DBA MEL CHEMICAL-TREATMENT PLANT	500 BARBER	FLEMINGTOI	HUNTERDON	NJ	8822	40.50408	-74.96743	NAD83	Chemicals	659
07033SYNRY209NM	1.10E+11	SYNRAY CORP	209 N MICHI	KENILWORTI	UNION	NJ	7033	40.68091	-74.288354	NAD83	Chemicals	458
07033SYNRY209NM	1.10E+11	SYNRAY CORP	209 N MICHI	KENILWORTI	UNION	NJ	7033	40.68091	-74.288354	NAD83	Chemicals	195
07033SYNRY209NM	1.10E+11	SYNRAY CORP	209 N MICHI	KENILWORTI	UNION	NJ	7033	40.68091	-74.288354	NAD83	Chemicals	205
07033SYNRY209NM	1.10E+11	SYNRAY CORP	209 N MICHI	KENILWORTI	UNION	NJ	7033	40.68091	-74.288354	NAD83	Chemicals	17
07033SYNRY209NM	1.10E+11	SYNRAY CORP	209 N MICHI	KENILWORTI	UNION	NJ	7033	40.68091	-74.288354	NAD83	Chemicals	108
07033SYNRY209NM	1.10E+11	SYNRAY CORP	209 N MICHI	KENILWORTI	UNION	NJ	7033	40.68091	-74.288354	NAD83	Chemicals	230
07033SYNRY209NM	1.10E+11	SYNRAY CORP	209 N MICHI	KENILWORTI	UNION	NJ	7033	40.68091	-74.288354	NAD83	Chemicals	106
07073GNRLF13MAN	1.10E+11	FXI INC	13 MANOR F	EAST RUTHE	BERGEN	NJ	7073	40.82558	-74.088563	NAD83	Plastics and	6.07
07033SYNRY209NM	1.10E+11	SYNRAY CORP	209 N MICHI	KENILWORTI	UNION	NJ	7033	40.68091	-74.288354	NAD83	Chemicals	349
07012CMSN7454AL	1.10E+11	COMUS INTERNATIONAL INC	454 ALWOOI	CLIFTON	PASSAIC	NJ	7012	40.84136	-74.15721	NAD83	Electrical Eqt	0.5
0772WRLPHC1CMME	1.10E+11	RALPH CLAYTON & SONS	100 COMME	TINTON FALI	MONMOUTH	NJ	7724	40.30204	-74.06855	NAD83	Nonmetallic	0
0772WRLPHC1CMME	1.10E+11	RALPH CLAYTON & SONS	100 COMME	TINTON FALI	MONMOUTH	NJ	7724	40.30204	-74.06855	NAD83	Nonmetallic	1.28
07066KRNKC330CE	1.10E+11	KARNAK CORP	330 CENTRA	CLARK	UNION	NJ	7066	40.632956	-74.31263	NAD83	Petroleum	34
07066KRNKC330CE	1.10E+11	KARNAK CORP	330 CENTRA	CLARK	UNION	NJ	7066	40.632956	-74.31263	NAD83	Petroleum	295
07066KRNKC330CE	1.10E+11	KARNAK CORP	330 CENTRA	CLARK	UNION	NJ	7066	40.632956	-74.31263	NAD83	Petroleum	34

In this portion of the lab, we will simulate using the addresses listed to geocode the sites in MMQGIS. However, geocoding the sites is quite computationally intensive practically speaking, so afterwards we will plot the points using the latitude and longitude coordinates provided.

Installing MMQGIS
1. Open QGIS and bring in the **County_Boundaries_of_NJ.shp** file into your project.
2. Next, we must install MMQGIS. Navigate to the **Plugins** tab in the top bar and click **Manage and Install Plugins**.
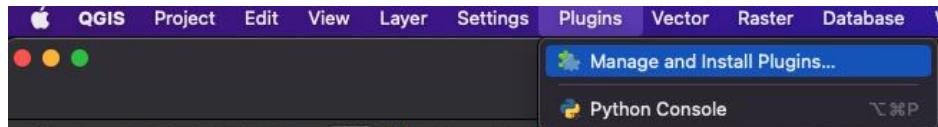
3. Search MMQGIS in the new window and click **Install Plugin**.
4. Once the installation is complete, **MMQGIS** will pop up on the top bar of your QGIS window**.**

Geocoding from Web Service MMQGIS
1. Navigate to the **MMQGIS** tab and click on **Geocode from Web Service**.
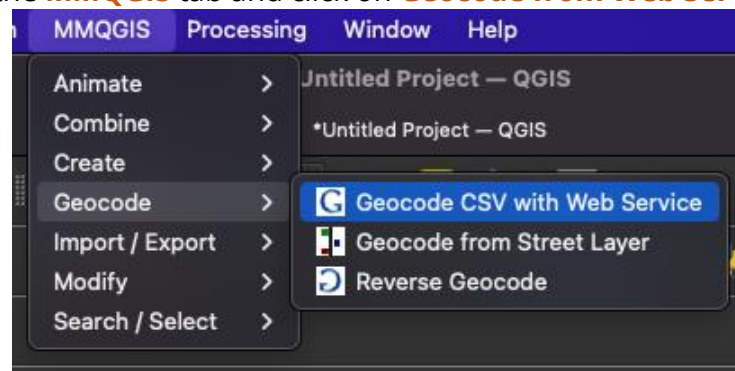
2. In the **Web Service Geocode** window, select the **tri_2020_nj.csv** file as your **Input CSV File**. Check that the other address fields correspond accordingly.
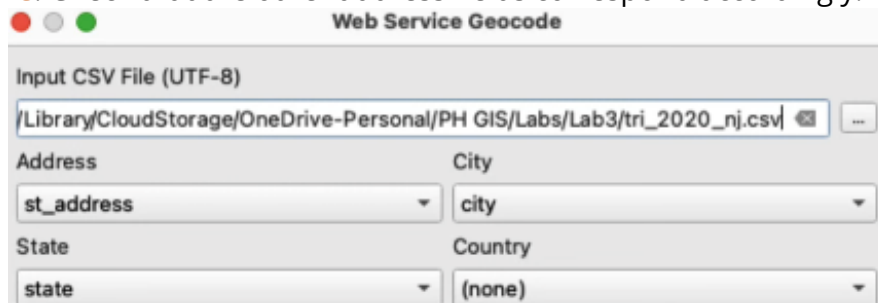
3. Specify **OpenStreetMap** as your **Web Service** (this is a free open-source web service). For the **Duplicate Handling** field, specify **Use Only First Result.**
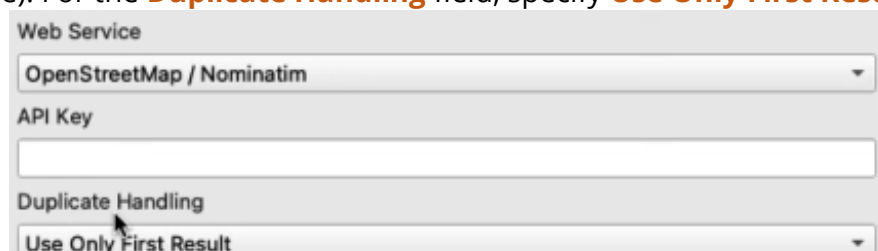
4. Specify **TRI_geocode.shp** (in your **Lab3** folder) as your **Output File Name** and **TRI_not_found.csv** as your **Not Found Output List**.
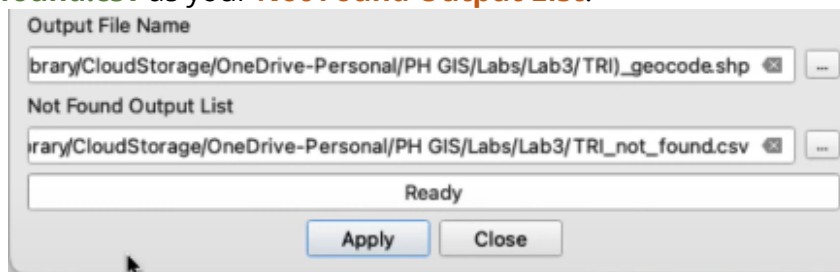

5. Before you click **Apply**, we will not geocode the locations ourselves because we already have the latitude longitude coordinates, as you saw in the Excel sheet.. Close out of the **Web Service Geocode** window.
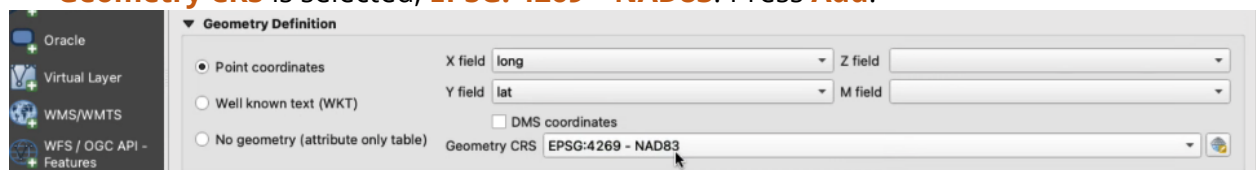
Plotting points from known CRS

1. In the top bar, navigate to the **Layer** tab, **Add Layer**, **Add Delimited Text Layer**.
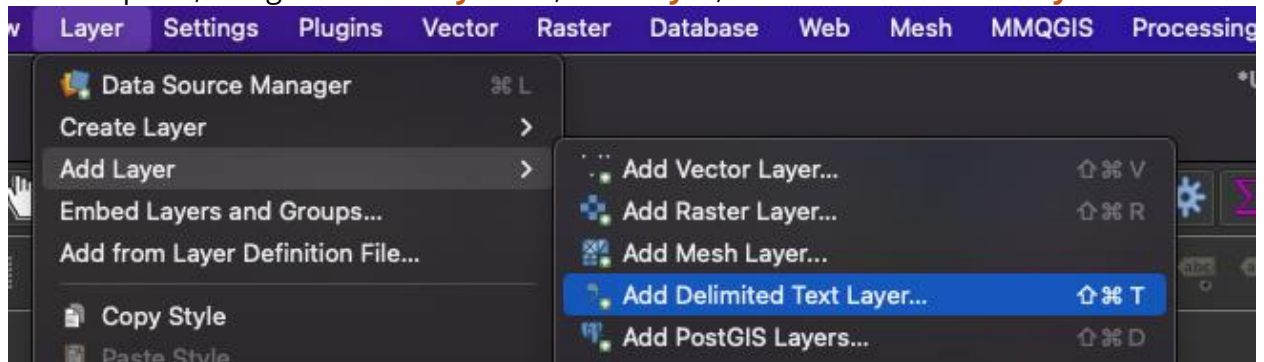
2. In the Data Source Manager pop-up window, specify the **tri_2020_nj.csv** file as your **File name**.
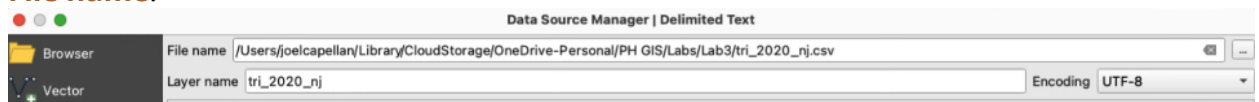
3. Specify **long** as your X field and **lat** as your Y field. Ensure that the correct **Geometry CRS** is selected, **EPSG: 4269 – NAD83**. Press **Add**.


4. Your map will be displayed in the QGIS Canvas:

5. To avoid the cumbersome process of reading in your points from the csv file every time, we will save out the point data as a separate layer. Right-click on the **tri_2020_nj** layer and click on **Save Feature As.**
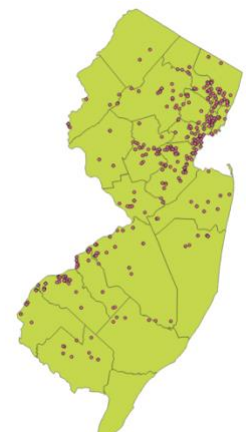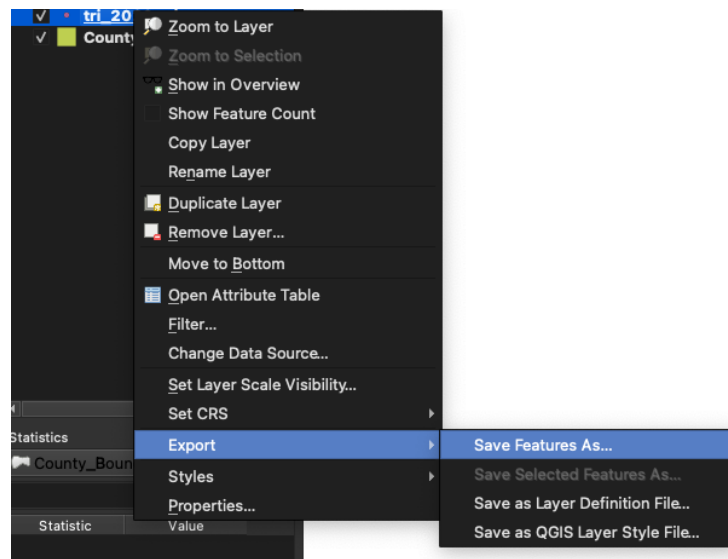
6. In the **File Name** field, click the **...** icon. Navigate to your Lab3 folder and save the layer as **NJ_TRI**. Press **OK** to save out the layer.
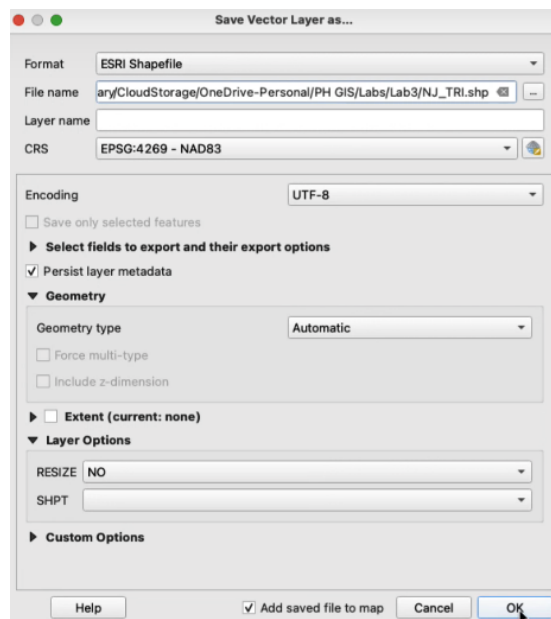
7. Remove the **tri_2020_nj** layer from your map.

Customizing map

We now want to adjust the colors and sizes of our data components to make our map clearer, as well as give it a modern look.

Customizing map—Background

1. Double-click on the **County_Boundaries_of_NJ** layer, go to **Symbology**.
2. Click on **Simple Fill**. For **Fill color**, select a light gray color. For **Stroke color**, let's select a white tone and increase **Stroke width** to 0.66. Click **Apply**.

Customizing map—Points

There are many ways we can customize how we visualize the points: by size, shape, and color. In this case, visualizing the points using a graduated color scale is not ideal because the points are quite small on the map. Instead, we will make the size of the points proportional to the toxic release values for each site.

1. Double-click on the **NJ_TRI** layer, go to **Symbology**.
2. Navigate to the **Size** field and click on the icon. Click on **Assistant**. A new **Symbol size** window should open.
3. Specify the **Source** field to be **total_rele**. Click the icon and the **Values from** and **to** fields should update to the minimum and maximum values from this column.
4. Set the **Size from** and **to** 1.000 and 15.000. Set the **Scale method** to **Radius**. Click **OK**.
5. Customize the color of the points to a darker color, if not already done. Double-click on the **NJ_TRI** layer, go to **Symbology**. There are multiple overlapping sites now due to the different sizes of release. To better visualize all sites, set the **Opacity** to 75%.
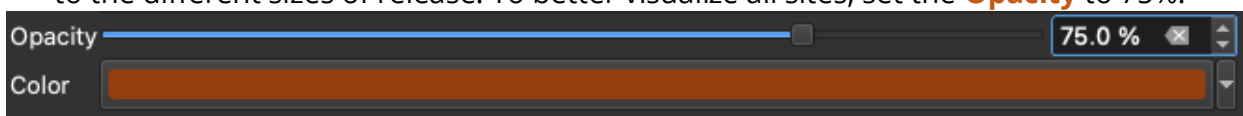
As you can see, most TRI sites reduce relatively little emissions, and it is the few select sites that are responsible for most of the total emissions in New Jersey.

Use New Print Layout to add final elements to the map

1. Click on **New Print Layout** in the upper part of the window.
2. Enter a print layout title *TRI* and click **OK**.

3. In the layout window, right-click on the empty page, and select **Page properties**
4. Under the Item Properties Pane on your right, change orientation to **Portrait**
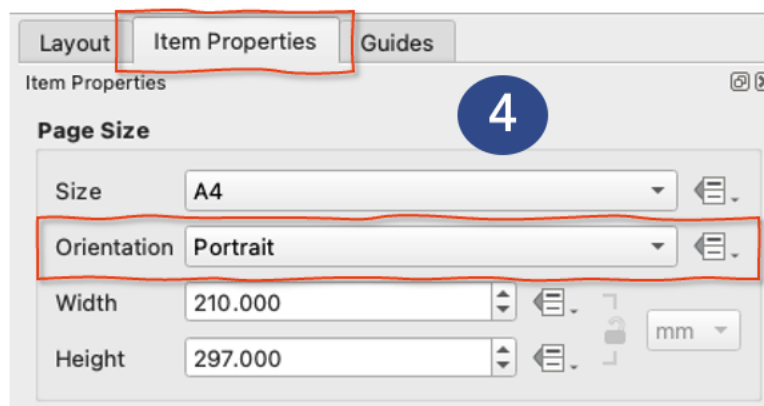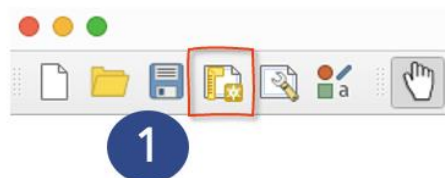
5. To paste our map on the layout, click on **Add Map** bottom**,** left-click-and-hold your mouse to draw the map window.
6. Click on **Select/move item** to move and adjust the size of your map window.
7. Click on **Move item content** to move your layer within the map window.
8. Increase the size of your map by lowering the **Scale** in your **Main Properties Window**.
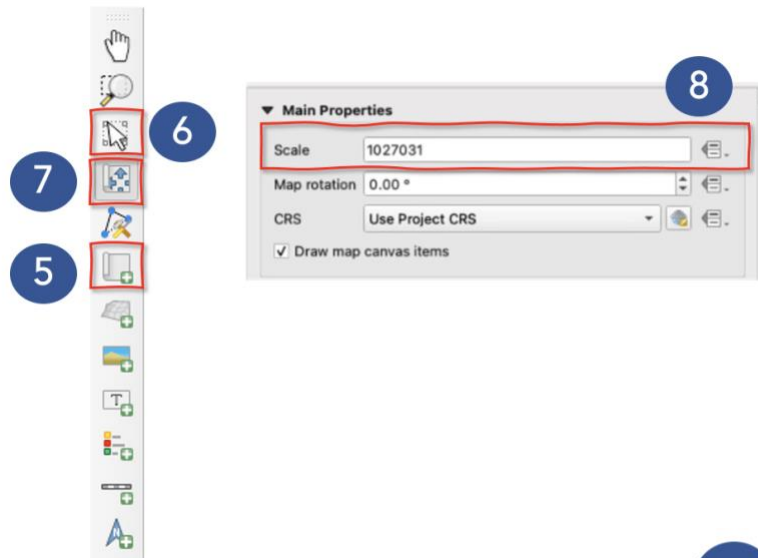
9. Click on the **Add a Scale Bar** to paste a scale bar.
   - Go to Main Properties to change the style, size, and units of the scale bar.
10. Click on **Add North Arrow** to paste a north arrow.
11. Click on **Add Label** to add a text box with details about **Author**, **Data Source**, and **Date**. You can change the font and size to a more appropriate fit.
12. Click on **Add Label** to write a note indicating "*Size of the dots is proportional to the total amount of toxic releases in FY2020*."
13. To add a title at the top of the map, click on **Add Label** and under Main Properties window type in the title "Amount of Toxic Release of Industrial Facilities in New Jersey FY2020." Center the title and increase the font size.
14. Ensure font type is the same among all the text boxes.

Once you are satisfied with the final version of the map, you can export it as a print-ready file, such as a PDF or TIFF, and send it to a printer for production. To export it as an image:
15. Go to the **Layout** menu at the top of the screen, select **Export as an Image**, name the image file "TRI_release", and click **Save**.
16. In the Image Export Options, select your **Export Resolution** (300 dpi is appropriate for most applications) and **Page width** and **Page Height**.
17. Click **Save** to export image.

Amount of Toxic Release of Industrial Facilities in New Jersey, FY2020

Size of the dots is proportional to the total amount of toxic release in FY2020.

0   10   20 mi

Author: Joel a. Capellan
Sources: Toxic Release Inventory, EPA
Date:

PART 3 – Spatial joins

As we saw in Part 2, it is only a handful of TRI facilities that produce most of the total emissions across New Jersey. We can now take our analysis from the individual facility level to the county level. This may be of particular relevance to communicate to government officials and policymakers who are more interested in what influences at the administrative boundary level.

Spatial joins allow for the transfer of information from one spatial feature to another based on geography. We can thus take the total emissions from all facilities within a given county and assign that total amount to each county. Our goal is to create a county-level map of New Jersey of total TRI emissions.

Execute spatial join

1. In the **Lab3** folder, locate and drag in the **County_Boundaries_of_NJ.shp** file into your QGIS window.
2. Locate and drag in the **NJ_TRI.shp** file into your QGIS window.
3. To complete the spatial join, navigate to the **MMQGIS** tab and click on **Spatial Join**.
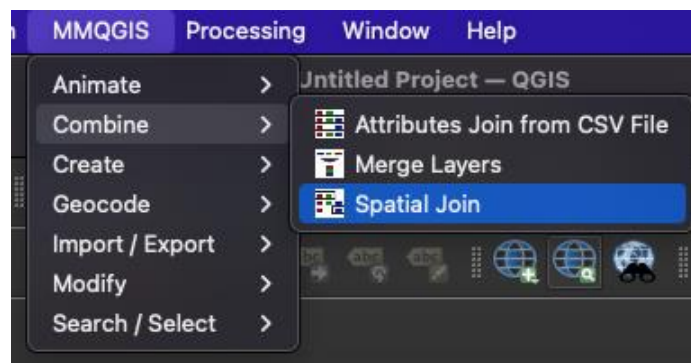

4. Specify **Output Shape (Target) Layer** as **County_Boundaries_of_NJ**.
5. Specify **Spatial Operation** as **Contains**. This is because we want to focus on the points contained in each county.
6. Specify **Data (Join) Layer** as **NJ_TRI**.
7. Highlight **total_rele** in **Fields** and select **Sum** for **Field Operation**. We want to sum the total value of our variable of interest, toxic release emissions.
8. Set your **Output File Name** as **NJ_TRI_joined** in your **Lab3** folder. Click **Apply**.

When we examine the attributes table, we see two columns: **total_rele** and **COUNT**. We can select each row and it will be correspondingly highlighted in the map.
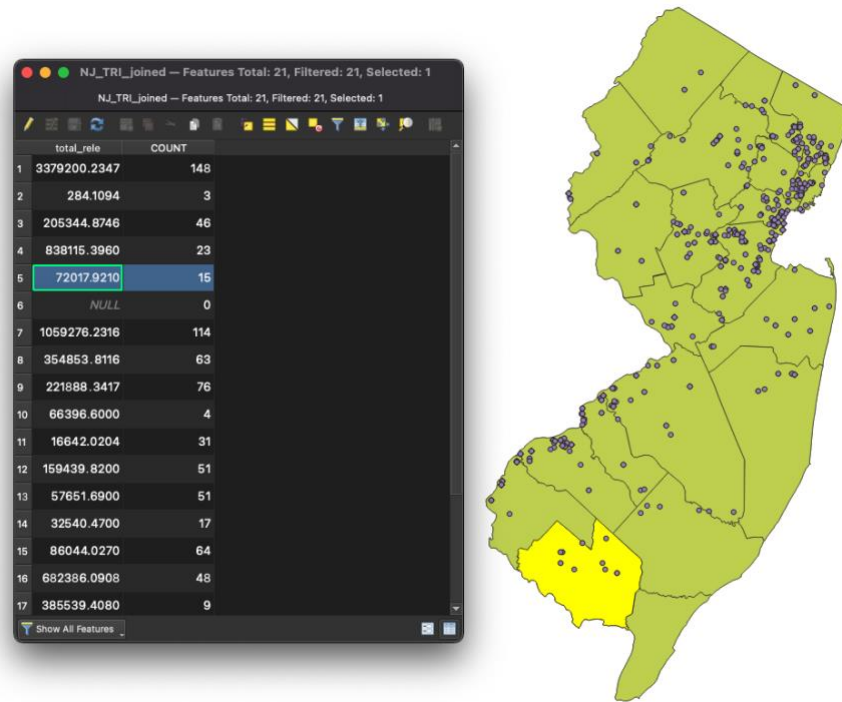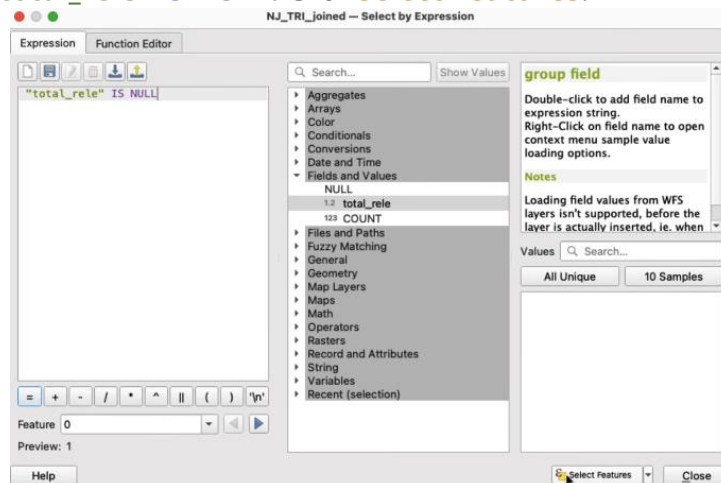
Note: One of the rows shows a *NULL* value in the **total_rele** column. We want to change this value to 0 because QGIS will interpret *NULL* as a missing value.
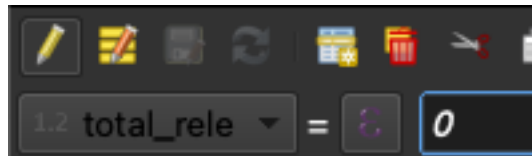
1. In the attribute table, click on **Select features using an expression**.
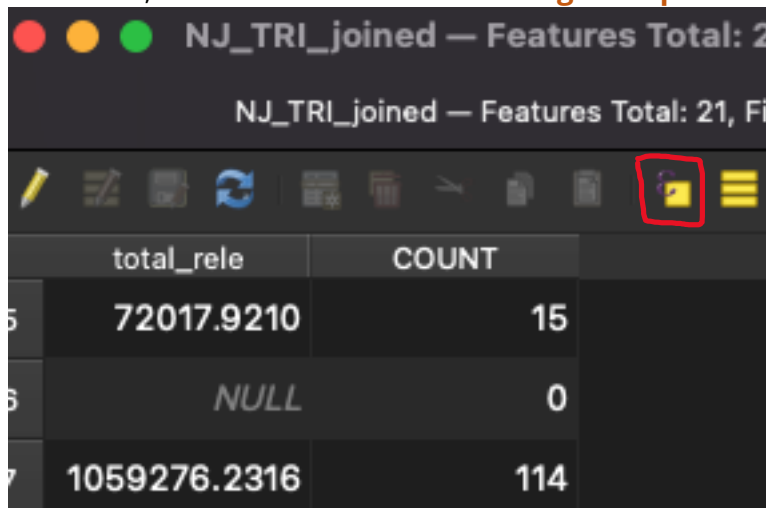

2. Navigate under **Fields and Values** and select **total_rele**. Finish up writing the expression: **"total_rele" IS NULL.** Click **Select Features**.

3. Now that the null value is selected, we can turn on the editing mode ✏️ and set total_rele = 0.


4. Click **Update Selected** so only our selected row have the **total_rele** values changed.

5. Click to turn off editing mode ✏️ and click **Save** when prompted to save the changes to Layer **NJ_TRI_Joined.** Close out and deselect that attribute.
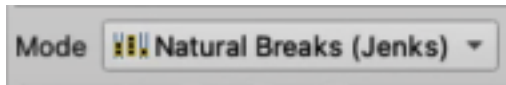
Visualizing county-level emissions

1. Remove the original **County_Boundaries_of_NJ** layer.

2. Double-click on the **NJ_TRI_joined** layer and navigate to **Symbology**. Select a **Graduated** fill type, specify **Value** as **total_rele** and select **Color Ramp** as **YlOrBr**.
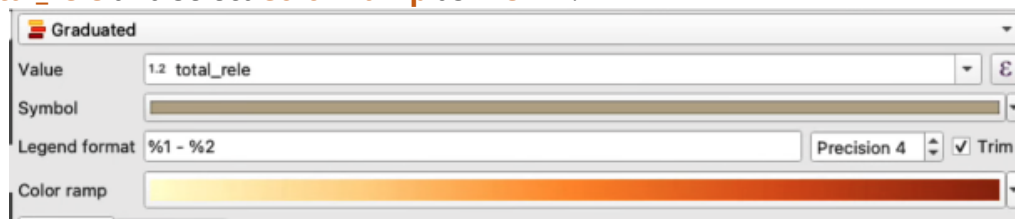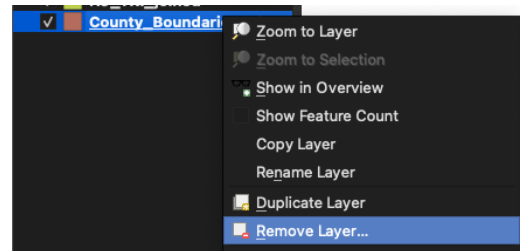

3. Keep **Mode** as **Equal Count (Quantile)** for now. Set **Classes** to **4**. Click **Apply**.

4. However, if we navigate to the Histogram tab, we can see that our distribution is quite skewed.
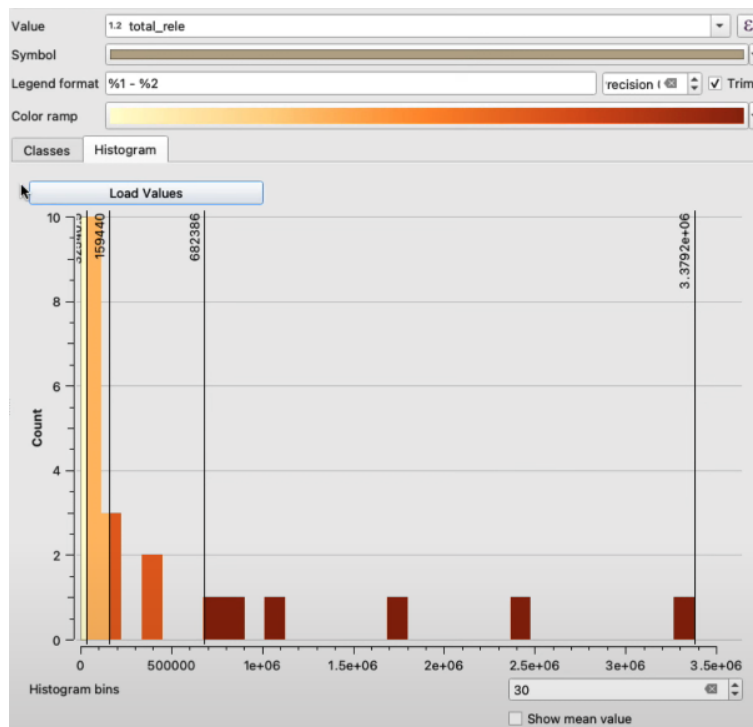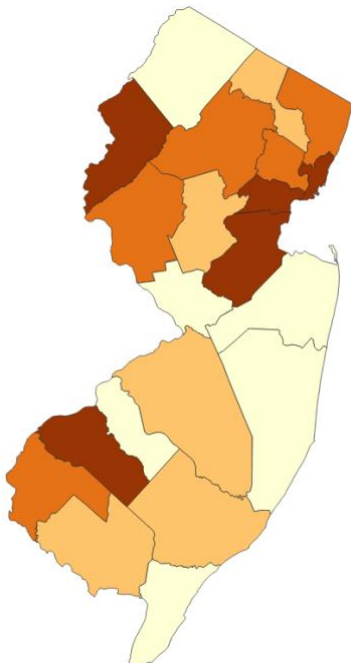
5. In this case, we should select a different mode of classifying. Return to the **Classes** tab and instead select **Mode** as **Natural Breaks (Jenks)**.

Mode ⊞ Natural Breaks (Jenks) ▾

Now the map appears more balanced and representative of the total TRI emissions by county.
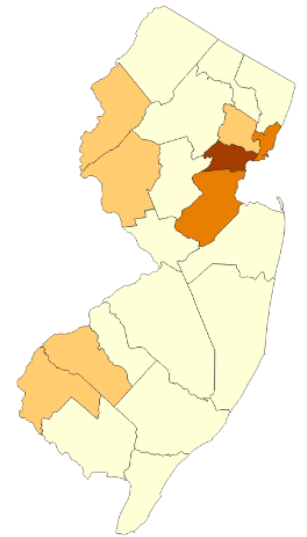
Visualizing emission point values with county-level emissions

We now want to visualize both total amount of toxic emissions by county and the spatial distribution of the individual sites to communicate the number of sites per county as well.

1. Turn back on the **NJ_TRI** layer. We can see upon initial inspection that the point layer can cover and obscure parts of the county-level map below it and we want to avoid this.
2. Double-click on the **NJ_TRI** layer and navigate to **Symbology**.
3. Change the **Fill color** to **black** and minimize the **Size** to **1.3**.
4. We also want to increase transparency of the points to better visualize any clustering. Click on the **Fill color** ribbon and change the **Opacity** to **35%**.

Now we can see that the degree of transparency of the points symbolize the density of sites across a given area.
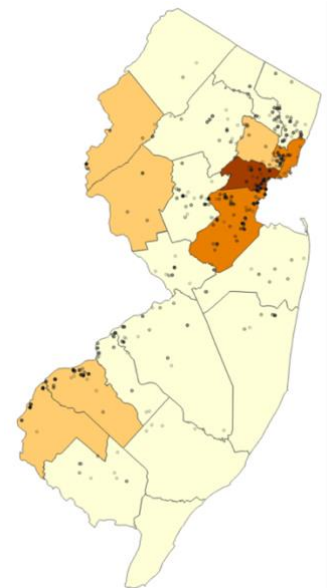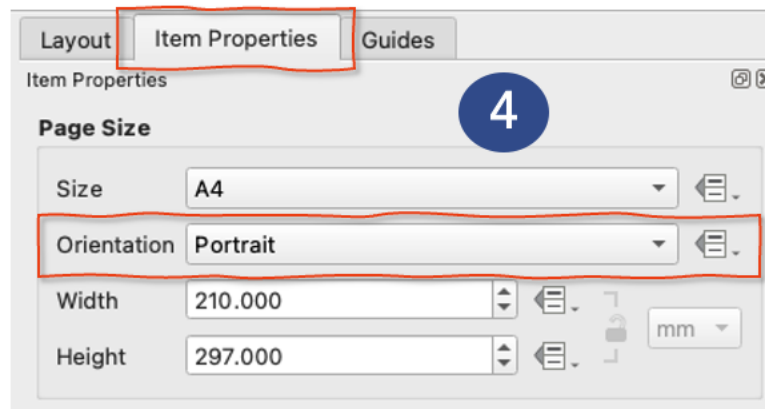
Use New Print Layout to add final elements to the map

1. Click on **New Print Layout** in the upper part of the window.
**2.** Enter a print layout title *TRI_join* and click **OK**.

3. In the layout window, right-click on the empty page, and select **Page properties**

4. Under the Item Properties Pane on your right, change orientation to **Portrait**

5. To paste our map on the layout, click on **Add Map** bottom, click-and-hold your mouse to draw the map window.
6. Click on **Select/move item** to move and adjust the size of your map window.
7. Click on **Move item content** to move your layer within the map window.
8. Increase the size of your map by lowering the **Scale** in your **Main Properties Window**.
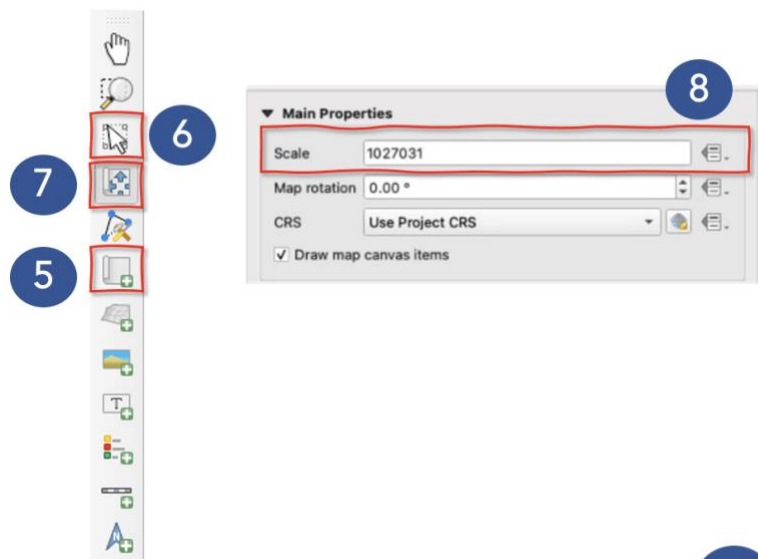
9. Click on the **Add a Scale Bar** to paste a scale bar.
   - Go to Main Properties to change the style, size, and units of the scale bar.
10. Click on **Add North Arrow** to paste a north arrow.
11. Click on **Add Label** to add a text box with details about **Author**, **Data Sources**, and **Date**. You can change the font and size to a more appropriate fit.
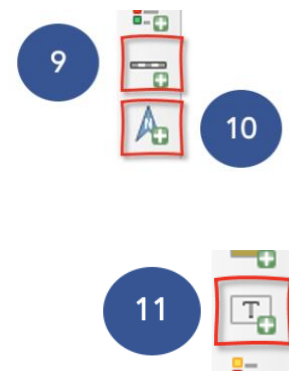
12. To add a title at the top of the map, click on **Add Label** and under Main Properties window type in the title "*Total Pounds of Toxic Emissions across Counties in New Jersey, FY2020*." Center the title and increase the font size.
13. Ensure font type is the same among all the text boxes.
14. Click on **Add Legend** and left-click-and-hold to draw the legend window.
15. Under Legend Items, deselect **Auto update**
16. Double click on the NJ_TRI layer to change title to *TRI Site*. Double click on the NJ_TRI_joined layer to change title to *Total Emissions (lbs)*.

Once you are satisfied with the final version of the map, you can export it as a print-ready file, such as a PDF or TIFF, and send it to a printer for production. To export it as an image:

17. Go to the **Layout** menu at the top of the screen, select **Export as an Image**, name the image file "TRI_join_emissions", and click **Save**.
18. In the Image Export Options, select your **Export Resolution** (300 dpi is appropriate for most applications) and **Page width** and **Page Height**.

Click **Save** to export image.

Total Pounds of Toxic Emissions across Counties in New Jersey, FY2020

TRI Site

Total Emissions (lbs)
- 0 - 221888
- 221888 - 1059276
- 1059276 - 2428504
- 2428504 - 3379200

0   10   20 mi

Author: Joel a. Capellan
Sources: Toxic Release Inventory
Date: