Tutorial 2: Query and Select TRI Spatial Data to Study State-Wide Emissions – Quantum GIS

This tutorial will introduce you to the following:

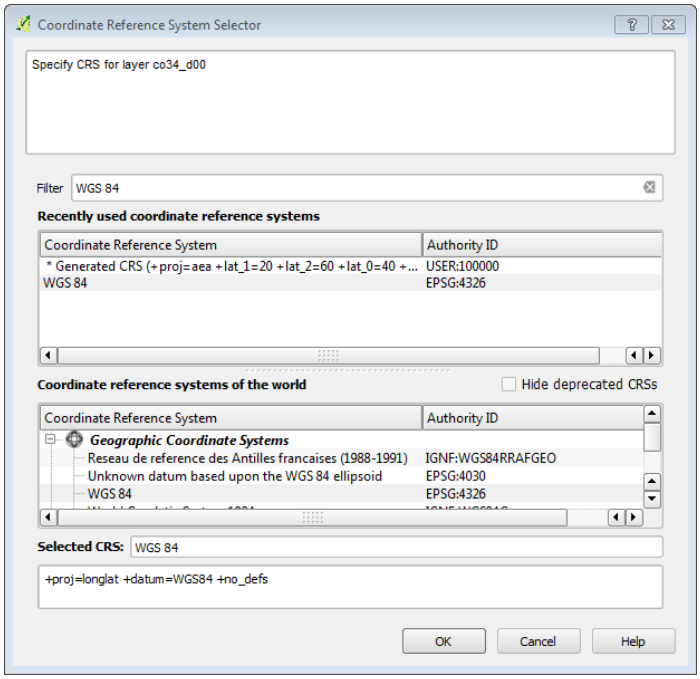
- Identifying Attribute Data Sources – Toxic Release Inventory (TRI) facilities and New Jersey County boundaries
- Adding Spatial References to Vector Data Layers
- Convert Text or Spreadsheet Data Files to Spatial Data Points
- Using the Clip Tool – narrowing down national data to data for one state – New Jersey
- Creating New Fields
- Querying and Selecting Data
- Generating New Data Sets from Existing Data Sets

Part 1. Identifying Attribute Data Sources

This tutorial will expand on the analysis of data used in the previous tutorial: the Toxic Release Inventory (TRI) data set. To begin, download and unzip the facilities data, listed in the Toxic Release Inventory from TOXMAP: http://toxmap.nlm.nih.gov/toxmap/download/facilities.zip . In addition to this data, this tutorial focuses on the analysis of TRI releases in New Jersey. For data on New Jersey county boundaries, visit the US Census website: http://www.census.gov/geo/maps-data/data/cbf/cbf_counties.html. Choose the Census 2000 tab, and select to download New Jersey data (at this point, if you prefer to study a different state, go ahead and pick a different state). As with the TRI data, save the New Jersey data to a known location on your computer. Once you finish downloading state boundary data, unzip the compressed file (right-click or double click the file, and select extract all).

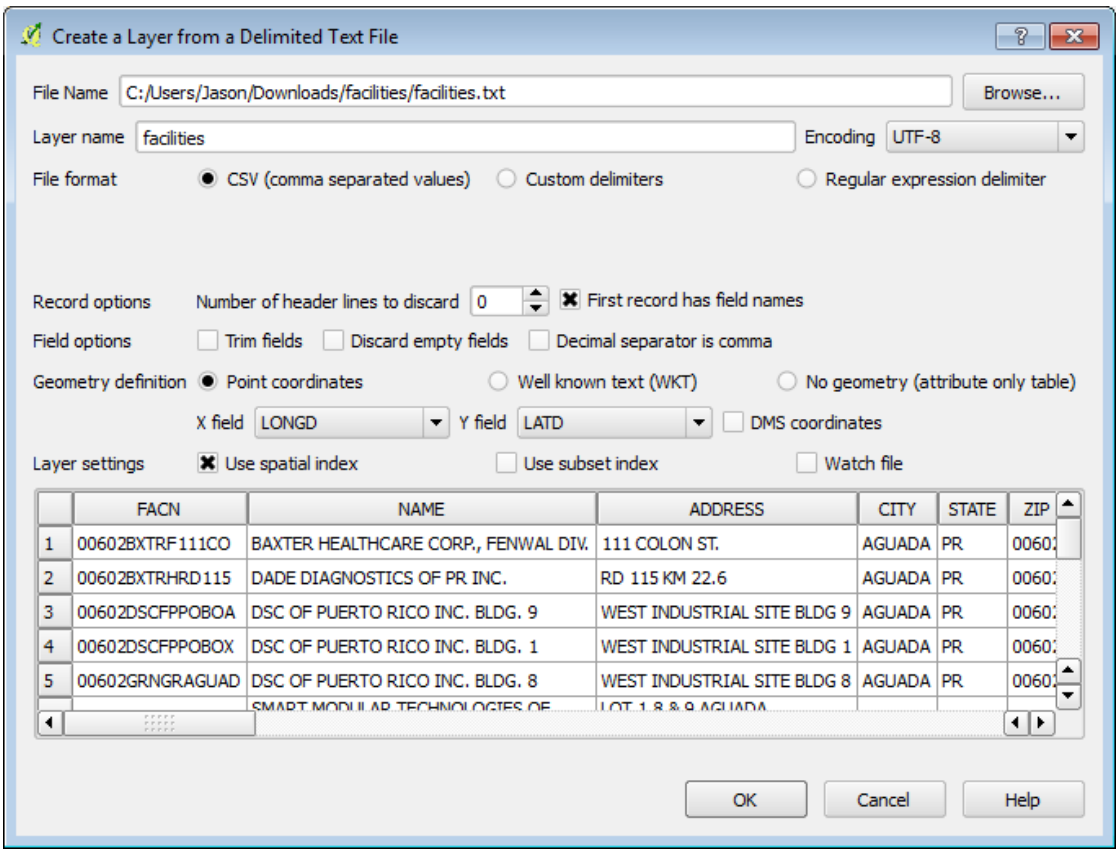
Part 2. Add Spatial References to Vector Data Layers

Open Quantum GIS. Using Layer > Add vector layer … browse to the state boundary folder (for NJ 2000 bounds, the title is co34.d00), and select the shapefile (.SHP). When you add the data, you will be asked to select a spatial reference for the data set. You may use any reference, but I recommend, for this exercise, using the World Geodetic System 1984. WGS 1984 is a good choice for working with, because many other data source (like NASA) and applications (like Google Earth) use this coordinate reference system. Type in WGS 84 into the filter, and pick the top WGS, EPSG: 4326 reference. Then, click OK. A map of the county boundaries should then appear in your map view.



Part 3. Convert Text or Spreadsheet Data Files to Spatial Data Points

Next, click Layer > Add delimited text layer, browse to the facilities folder and facilities.txt file (downloaded from TOXMAP). The facilities.txt folder contains data on each facility in the Toxic Release Inventory (TRI). In the "Create a Layer from a Delimited Text File" dialog box, select CSV for the File format. Also, for Geometry definition, select "Point Coordinates," and choose LONGD for the X field, and LATD for the Y field. Check the box the Use spatial index. An image of the box is shown below. Click OK.

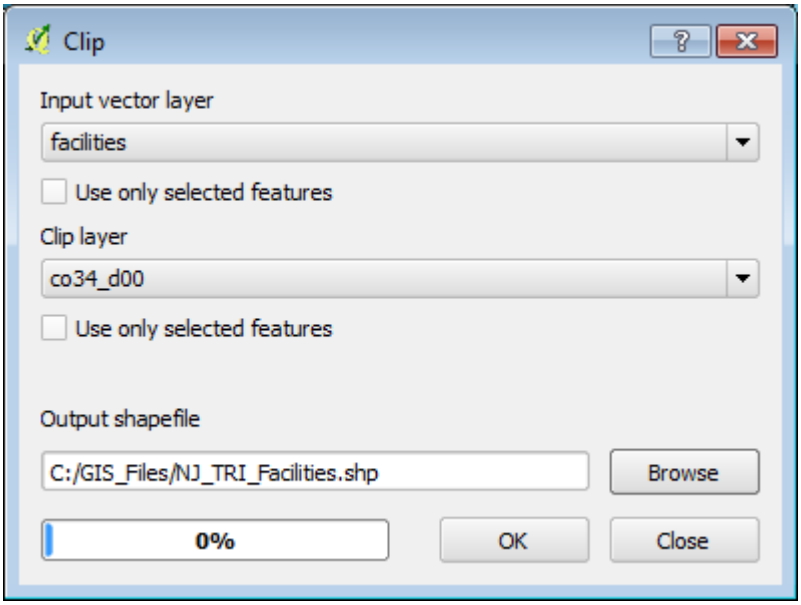


There are some errors in the formatting of a few of the data records, of which Quantum GIS notifies you. Close the error box. Next, you will be asked to specify the coordinate reference system. Choose WGS 84 from the "recently used coordinate reference systems" box, and click OK.

Save your work before continuing.

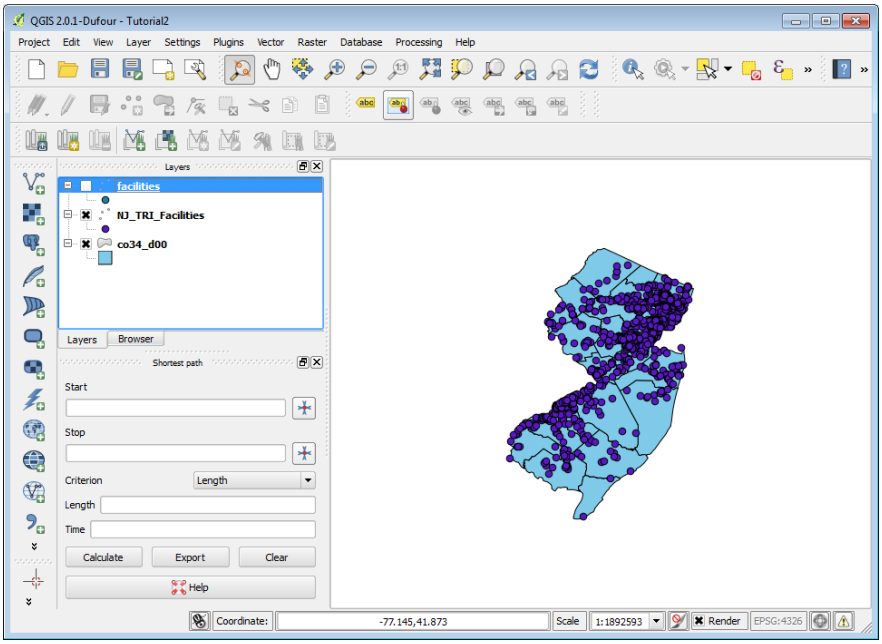
Part 3. Clip Data Tool

Often when working with large data sets, you will want to narrow down the information that you are analyzing. One way to do this is to use the map tools. For example, instead of working with the TRI facilities for the entire US, you may want to work with only facilities that are located in New Jersey. There are several correct ways to do this, but one way is to use the clip tool. To clip data (like a cookie cutter from a sheet of cookie dough), select Vector (from the top tool menu) > Geoprocessing Tools > Clip. Fill out the dialog box as follows (see image below). This will clip away all facilities that are not located in New Jersey. Make sure that you change the Output shapefile to a location that you can find again. Then, click OK. Choose to add the new layer to the TOC.



Before the layer is added, you will be asked to set the coordinate reference again. Again, use WGS 84. Close the Clip box (it will not close automatically).

In the layers menu, you can uncheck the box next to facilities, to reveal the new shapefile that you just created. This shapefile should contain only points that are bounded by the state that you chose.



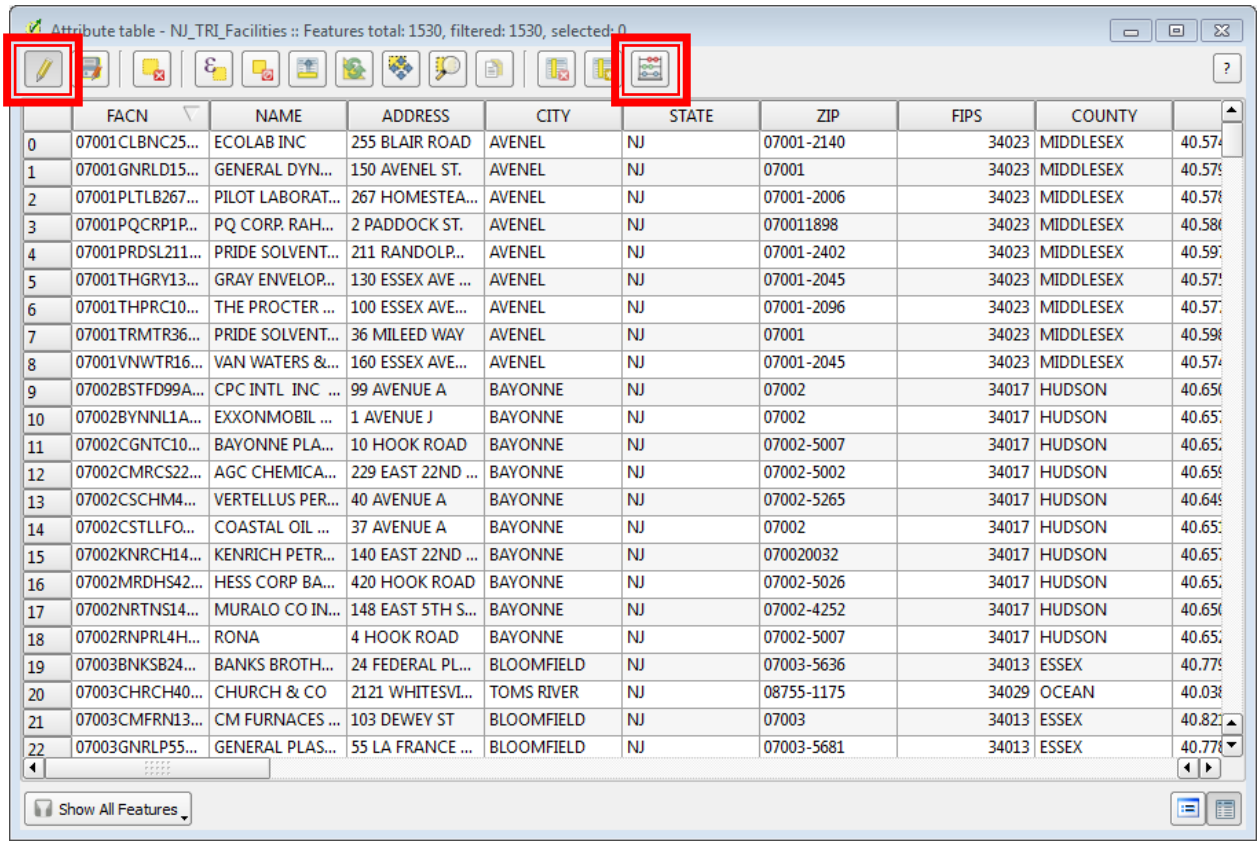
Already you can begin answering some more complex questions with the analysis that you have done. For example, how many EPA regulated facilities are there in the United States (find the layer that you need, e.g. facilities, right-click on it in the table of contents, and select Open Attribute Table – the total number should be reported at the top of the table)? How many of them are in New Jersey (hint: open the attribute table for the New Jersey TRI shapefile)? With that information, you can find out what percentage of all EPA regulated facilities are located in New Jersey. After you compare New Jersey and national totals, you can either remove or uncheck the layer with the national TRI locations.

Part 4. Creating New Fields, Selecting Data and Analyzing Data Using Queries

Existing in the facilities data set are reports of lbs. of toxic releases for each facility, by year, from 1988-2011. But, suppose you would like to find out changes in emissions, over a decade, say from 2000 to 2010, for facilities in New Jersey. You can do that. The following steps will guide you through one way of finding that information.

Step 1. Creating New Fields

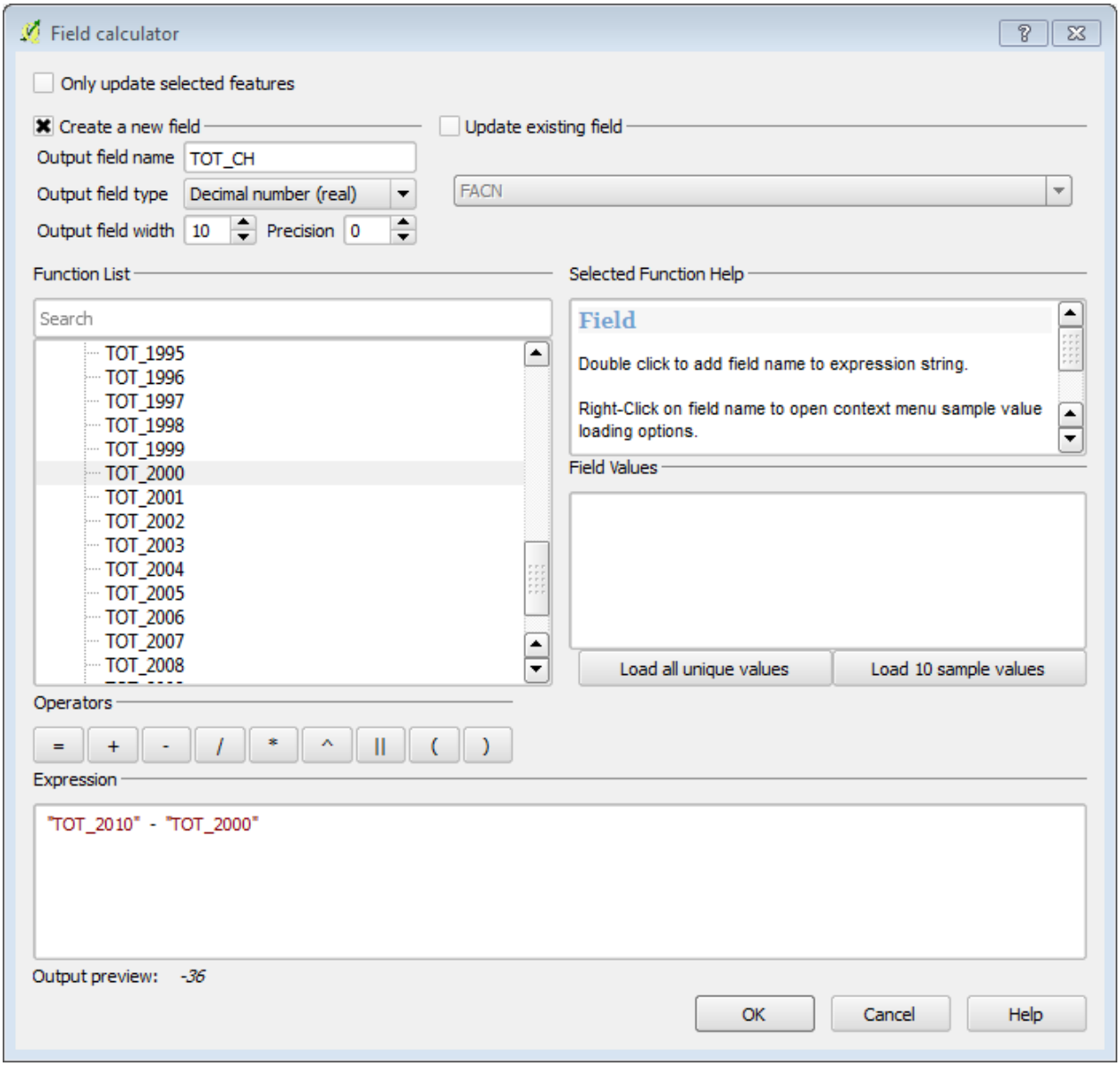
To create a new field, which will contain information on changes in lbs. of releases from 2000 to 2010, first open the attribute table for TRI facilities in New Jersey (right-click on the layer in the table of contents and select Open Attribute Table, make sure that you are using the New Jersey selection, instead of all national data). In the top-left corner of the table view (shown in the image below), click the first button – Toggle Editing Mode – to start editing.



	FACN	NAME	ADDRESS	CITY	STATE	ZIP	FIPS	COUNTY	
0	07001CLBNC25...	ECOLAB INC	255 BLAIR ROAD	AVENEL	NJ	07001-2140	34023	MIDDLESEX	40.574
1	07001GNRLD15...	GENERAL DYN...	150 AVENEL ST.	AVENEL	NJ	07001	34023	MIDDLESEX	40.579
2	07001PLTLB267...	PILOT LABORAT...	267 HOMESTEA...	AVENEL	NJ	07001-2006	34023	MIDDLESEX	40.578
3	07001PQCRP1P...	PQ CORP. RAH...	2 PADDOCK ST.	AVENEL	NJ	070011898	34023	MIDDLESEX	40.586
4	07001PRDSL211...	PRIDE SOLVENT...	211 RANDOLP...	AVENEL	NJ	07001-2402	34023	MIDDLESEX	40.597
5	07001THGRY13...	GRAY ENVELOP...	130 ESSEX AVE ...	AVENEL	NJ	07001-2045	34023	MIDDLESEX	40.575
6	07001THPRC10...	THE PROCTER ...	100 ESSEX AVE...	AVENEL	NJ	07001-2096	34023	MIDDLESEX	40.577
7	07001TRMTR36...	PRIDE SOLVENT...	36 MILEED WAY	AVENEL	NJ	07001	34023	MIDDLESEX	40.598
8	07001VNWTR16...	VAN WATERS &...	160 ESSEX AVE...	AVENEL	NJ	07001-2045	34023	MIDDLESEX	40.574
9	07002BSTFD99A...	CPC INTL INC ...	99 AVENUE A	BAYONNE	NJ	07002	34017	HUDSON	40.650
10	07002BYNNL1A...	EXXONMOBIL ...	1 AVENUE J	BAYONNE	NJ	07002	34017	HUDSON	40.657
11	07002CGNTC10...	BAYONNE PLA...	10 HOOK ROAD	BAYONNE	NJ	07002-5007	34017	HUDSON	40.652
12	07002CMRCS22...	AGC CHEMICA...	229 EAST 22ND ...	BAYONNE	NJ	07002-5002	34017	HUDSON	40.659
13	07002CSCHM4...	VERTELLUS PER...	40 AVENUE A	BAYONNE	NJ	07002-5265	34017	HUDSON	40.649
14	07002CSTLLFO...	COASTAL OIL ...	37 AVENUE A	BAYONNE	NJ	07002	34017	HUDSON	40.651
15	07002KNRCH14...	KENRICH PETR...	140 EAST 22ND ...	BAYONNE	NJ	070020032	34017	HUDSON	40.657
16	07002MRDHS42...	HESS CORP BA...	420 HOOK ROAD	BAYONNE	NJ	07002-5026	34017	HUDSON	40.652
17	07002NRTNS14...	MURALO CO IN...	148 EAST 5TH S...	BAYONNE	NJ	07002-4252	34017	HUDSON	40.650
18	07002RNPRL4H...	RONA	4 HOOK ROAD	BAYONNE	NJ	07002-5007	34017	HUDSON	40.652
19	07003BNKSB24...	BANKS BROTH...	24 FEDERAL PL...	BLOOMFIELD	NJ	07003-5636	34013	ESSEX	40.779
20	07003CHRCH40...	CHURCH & CO	2121 WHITESVI...	TOMS RIVER	NJ	08755-1175	34029	OCEAN	40.038
21	07003CMFRN13...	CM FURNACES ...	103 DEWEY ST	BLOOMFIELD	NJ	07003	34013	ESSEX	40.821
22	07003GNRLP55...	GENERAL PLAS...	55 LA FRANCE ...	BLOOMFIELD	NJ	07003-5681	34013	ESSEX	40.778

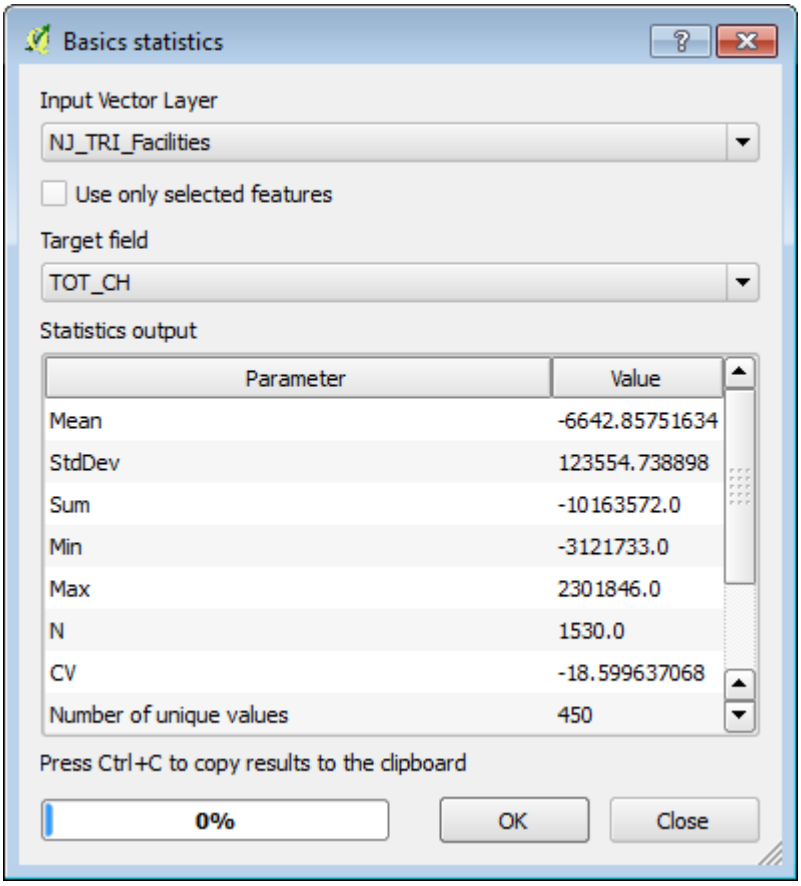
In the attribute table, the last button on the tool bar allows you to open a field calculator. Press this button.

The Field Calculator uses SQL (structured query language) to generate the values for new attributes. Check the box to "Create a new field." Choose an Output field name, like TOT_CH (for total change). Change the Output field type to Decimal number (real), and Output field width to 10 and precision to 0. Double click on existing fields in the "Fields and Values" Function List, and they will appear in the box below, which defines the new change field that you created. Change is calculated as "TOT_2010" – "TOT_2000" . This will mean that the new value will be positive if toxic releases have increased in that decade, and negative, if they have decreased (see dialog box, below). Click OK.



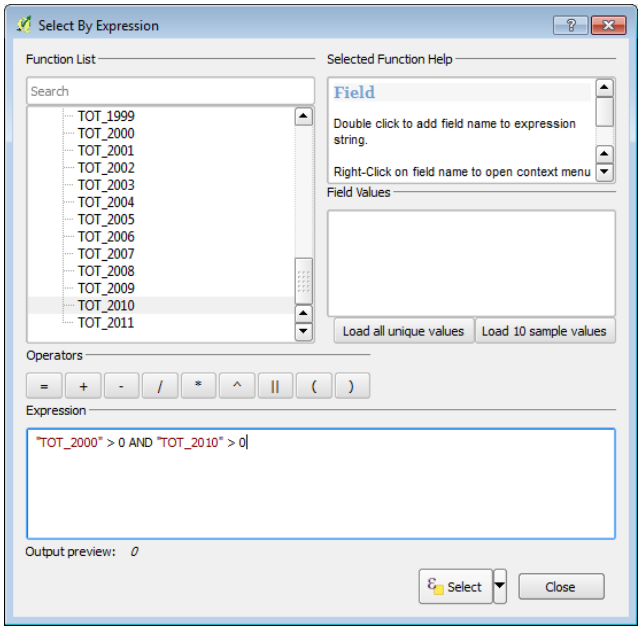
To see the new field that you generated, scroll all the way to the right in the attributes table. There you will find TOT_CH. Close the attributes table. From the top tool menu, choose Vector > Analysis Tool s > Basic Statistics. Scroll to the bottom of the Target field to find TOT_CH, and click OK (box shown below). This will give you a statistical summary of the field that you made.

From the statistical summary, the mean suggests that the total releases have declined for these toxic release facilities. However, there is a catch to working with this data. If you carefully examine the tables, the total releases for many facilities over different years is reported as -1. How can this be? Well, what this really means is that the data are missing or not reported. This could happen when facilities previously reported releases, but no longer have releases to report. Or, where newer companies did not have any releases for previous years in the data set. To make the information that we calculated becomes meaningful, we need to exclude records where the information was missing for either 2000 or 2010. Again, SQL will come in handy. Go ahead and close the basic statistics dialog box.

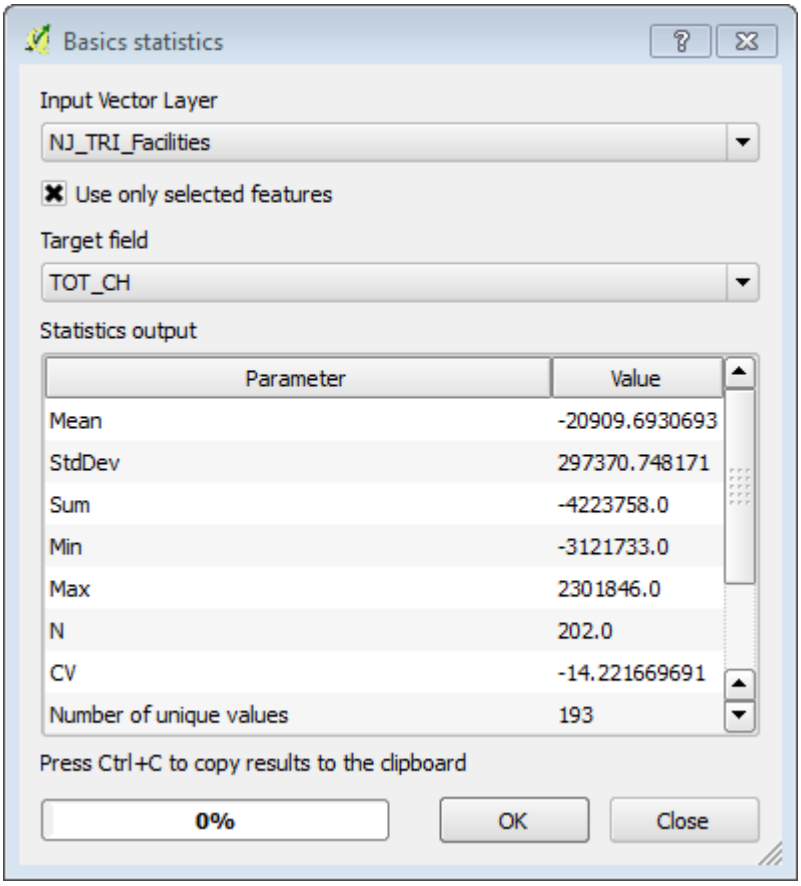


Step 2. Selecting Data Using Queries

To select records, which were not missing for the years 2000 and 2010, reopen the New Jersey TRI attribute table. The fourth button showing an epsilon and a yellow rectangle lets you set up an expression to select data. Click this button. Fill out the "Select by Expression" dialog box as follows (image below), so that "TOT_2000" > 0 AND "TOT_2010" > 0. Then, click Select.



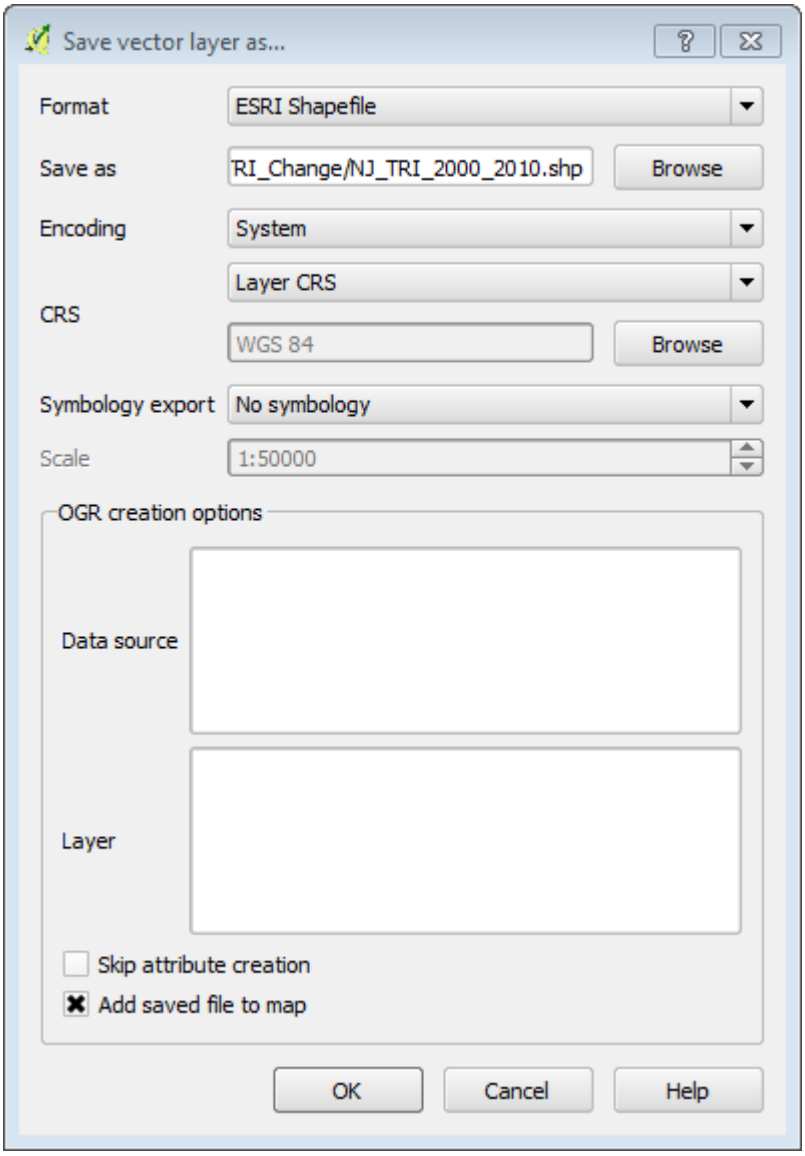
Close the "Select By Expression" dialog box. Now, only 202 of the original 1530 facilities were selected. Click the second button on the left to save edits. Then click the first button to toggle off editing. Close the attributes table. The selected facilities will be highlighted yellow. Recalculate the basic statistics (Vector > Analysis Tools > Basic Statistics, and change the Target field to TOT_CH).



Interestingly, when examining only the facilities that reported releases in 2000 and 2010, the average facility emissions (shown as Mean), declined during this period, more than calculated previously. Close the Basic statistics box.

Step 3. Analyzing Selected Data

From the New Jersey TRI layer, it is possible to make and save a new layer with only the 202 records that had positive releases for both 2000 and 2010. To do this, right-click on the New Jersey TRI layer, and select "Save selection as...." Save the file with a new name and in location that you can remember. Select the box to "Add saved file to map."



Up to this point, you have conducted some great analysis. Now, the hard part, taking the information that you have created and conveying it in a map. Give it your best try! The following steps are suggestions for visualizing the data, so please feel free to choose your own colors and map design along the way.

Part 5. Visualizing Spatial Data

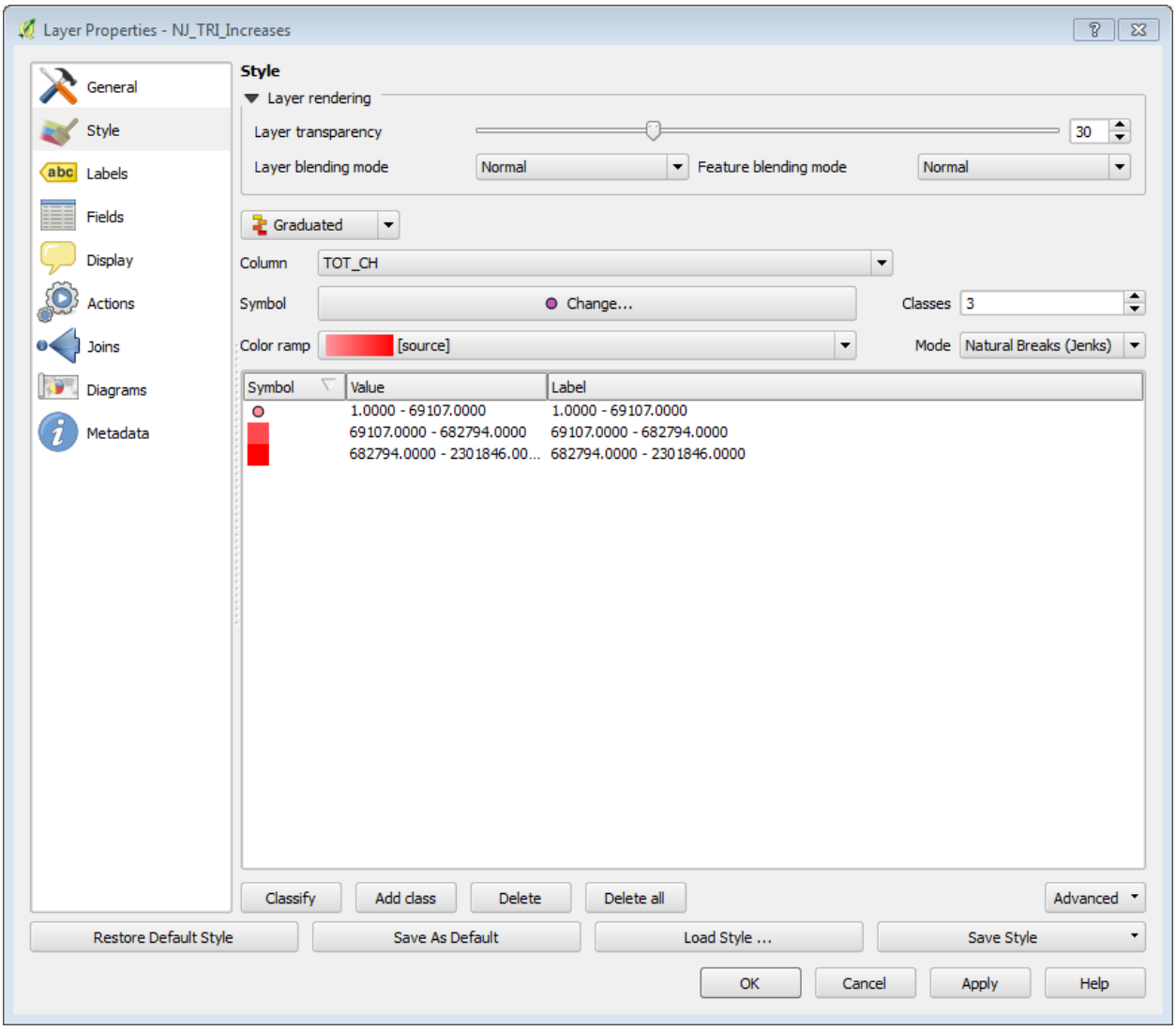
Before I begin working on the data visualization, I often think about what I would like the map to look like. With this particular dataset, I would like to call attention to the facilities with the biggest changes: the largest increases and decreases in emissions. I also want to clearly distinguish between increases and decreases, using different colors. To begin, I created two new selections of the New Jersey TRI data – one with TRI increases and one with TRI decreases. Then, I classified the data with different colors and symbols for each selection. The following steps illustrate the process that I used and the outcome of that process.

Step 1. Create Two TRI Selections

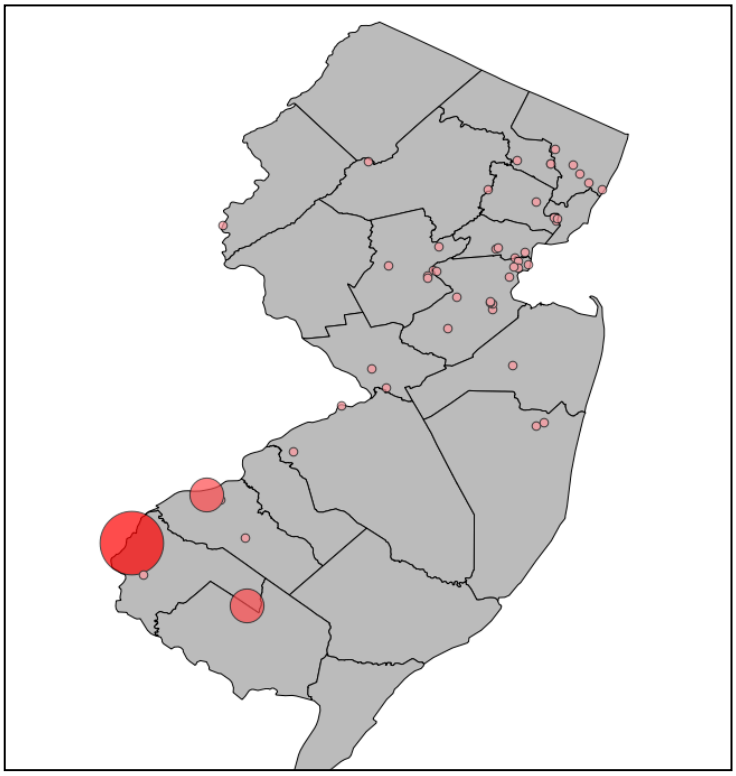
To separate increases and decreases, I opened attributes table for the NJ TRI 2000-2010 dataset that I created. Then, I select the records where TOT_CH > 0 (using the same selection methods described earlier in this document). Again, right-clicking my data layer, I saved the selection as TRI Increases. I used similar steps to create a data layer with TRI Decreases.

Step 2. Change the Symbology for Each Selection

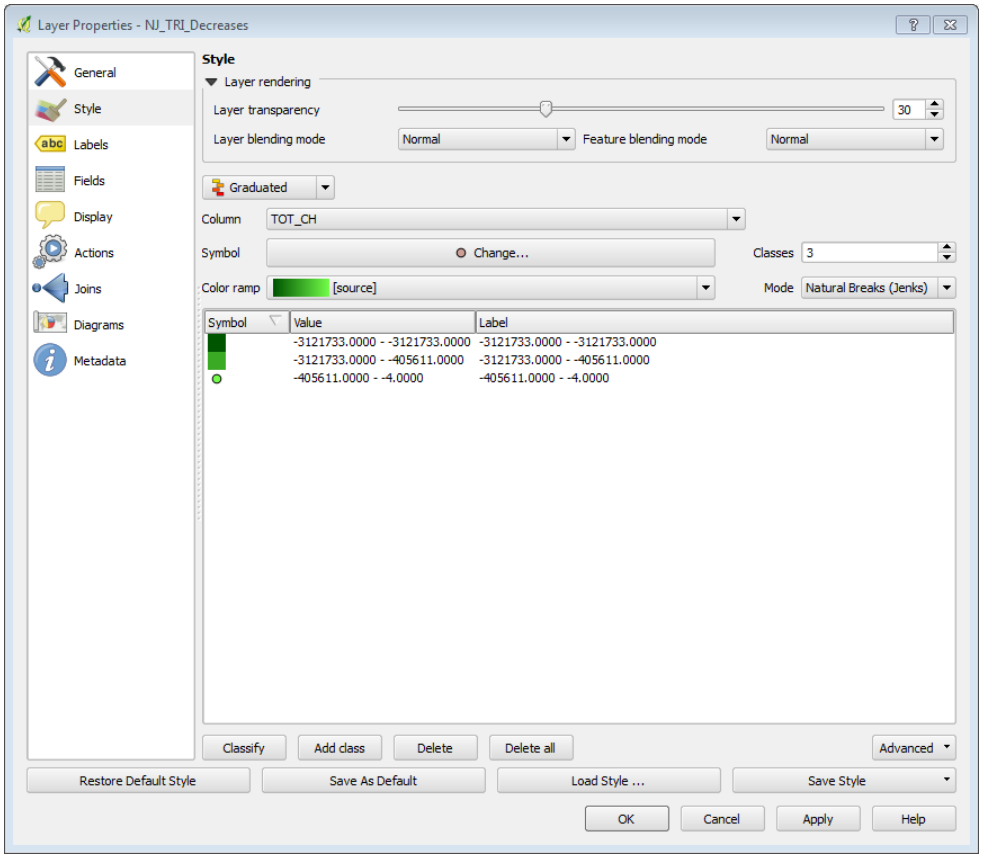
By double clicking the data layer for increases, I opened the Layer Properties menu. In this menu, it is possible to change the symbology. On the left-side of the menu, select Style. The dialog box for the style that I used is shown below. There are many options for displaying spatial data. I chose to create graduated symbols for the TOT_CH variable (column). I changed the classes to 3 and the Mode to Natural Breaks (Jenks). To update changes in the mode or classes, click Classify at the bottom of the menu. I selected the colors for each group, and I changed the symbol size by double-clicking on the symbol (circle) next to each value range, keeping the first one at 1, changing the second on to 8, and the third to 15. I also adjusted the layer transparency to 30.



The results were as follows.

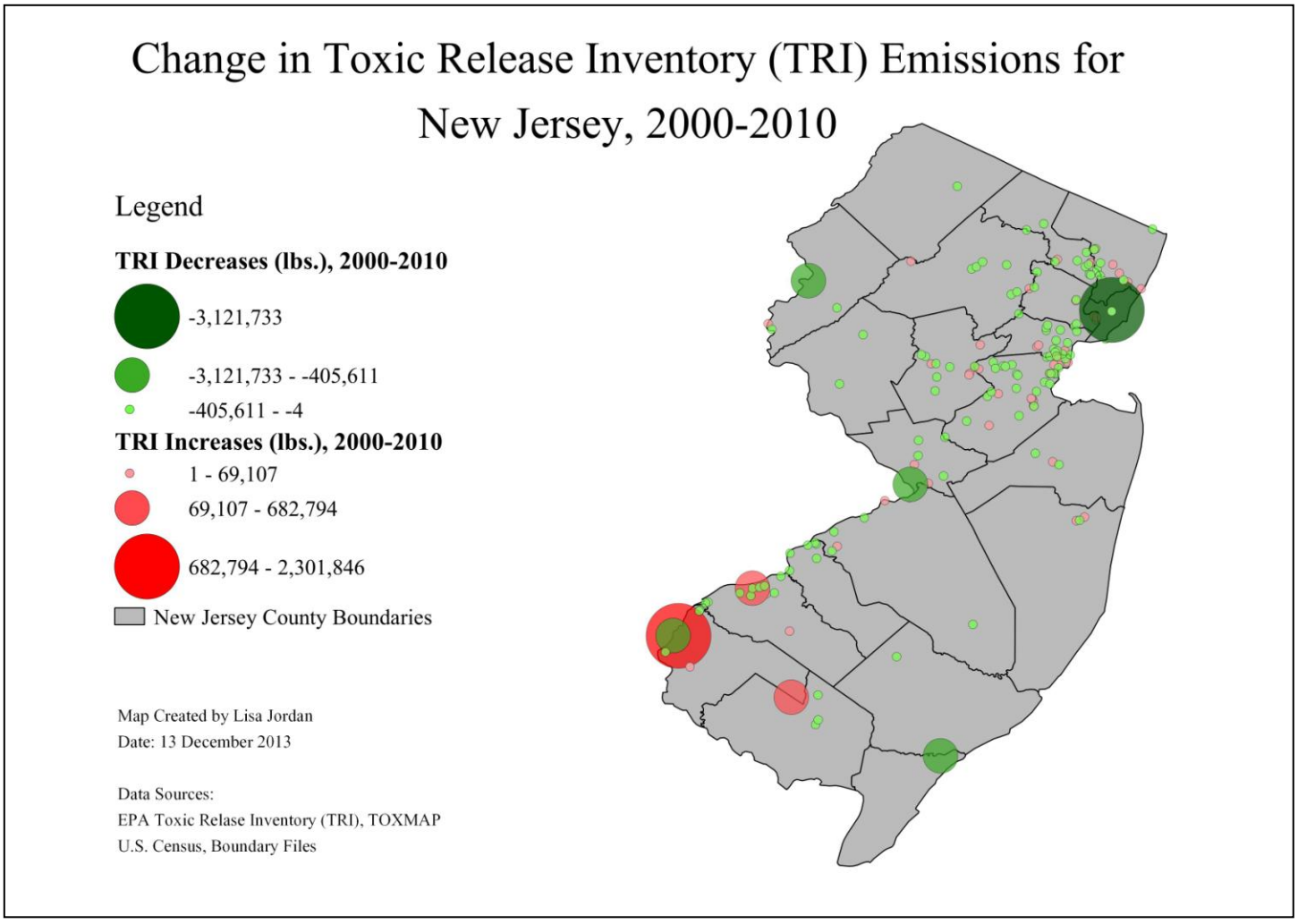


I classified the TRI decreases with the same technique.



Then, following the steps introduced in Tutorial 1, I created a new print composer document to display the changes in TRI emission for the state of New Jersey.

After adding the map, a title, legend, and data credits, I also edited and updated the layer names and the labels for the symbols. The results are below. Remember, cartography is an art, so consider making the presentation of the toxic release data that works best for you. Also, consider how your audience might interpret your map. I hope that viewing the map below would lead the reader to understand that between 2000 and 2010, many facilities experienced decreases in emissions. A few facilities experienced very large increases in emission and a few experienced very large decreases in the pounds of emission released. There are many other issues not discussed here too – what information do you want to portray in your presentation? Have fun experimenting with the options.



As a personal preference, I really like the red and green colors used here, because red or orange works nicely for a hazard symbol, and I associate greens with a more positive environmental influence (e.g. declines in TRI emissions). However, this map is terrible for someone who is red-green color blind. I also created a blue-orange map, shown below. I could choose other colors as well. Again, it's always important to consider your audience, and you can adapt the symbols and colors of your maps to help do this.

Change in Toxic Release Inventory (TRI) Emissions for New Jersey, 2000-2010