An Exploratory Analysis: EPA Compliance and Safe Drinking Water Facilities in Texas.

Christian Enriquez

INFO-532

Introduction

The Safe Drinking Water Act (SDWA) was originally passed by Congress in 1974 to protect public health by regulating the nation’s public drinking water supply. The law was amended in 1986 and 1996. The SDWA applies to every public water system in the United States. There are currently more than 170,000 public water systems providing water in the United States. What constitutes violations against the safe water drinking act are elevated levels of contaminates as defined by the EPA on their website3. Public water systems are responsible for ensuring that contaminants in tap water do not exceed approved standards. Water systems treat the water, and must test their water frequently for specified contaminants and report the results to states. If a water system is not meeting these standards, it is the water supplier’s responsibility to notify its customers4.

The United States generally ranks high among the world in quality of drinking water. Unfortunately, there are many regions of the country that lag behind acceptable EPA standards. Texas as a state is acknowledged to have the worst water quality relative to its population size1. Texas, according to a report by the Environment Texas Research and Policy Center, was the second biggest polluter of water in the country, but when the toxicity of the water is factored in, becomes the number one polluter of water2.

The aims of this exploratory analysis are as follows.

* **Find where water quality suffers in the state of Texas.**
* **Attempt to identify patterns among areas of the state with bad water quality. Are there demographic similarities among populations served by these water systems?**
* **Conversely, what are the similarities between areas that have good water quality?**

Data and Data Cleaning

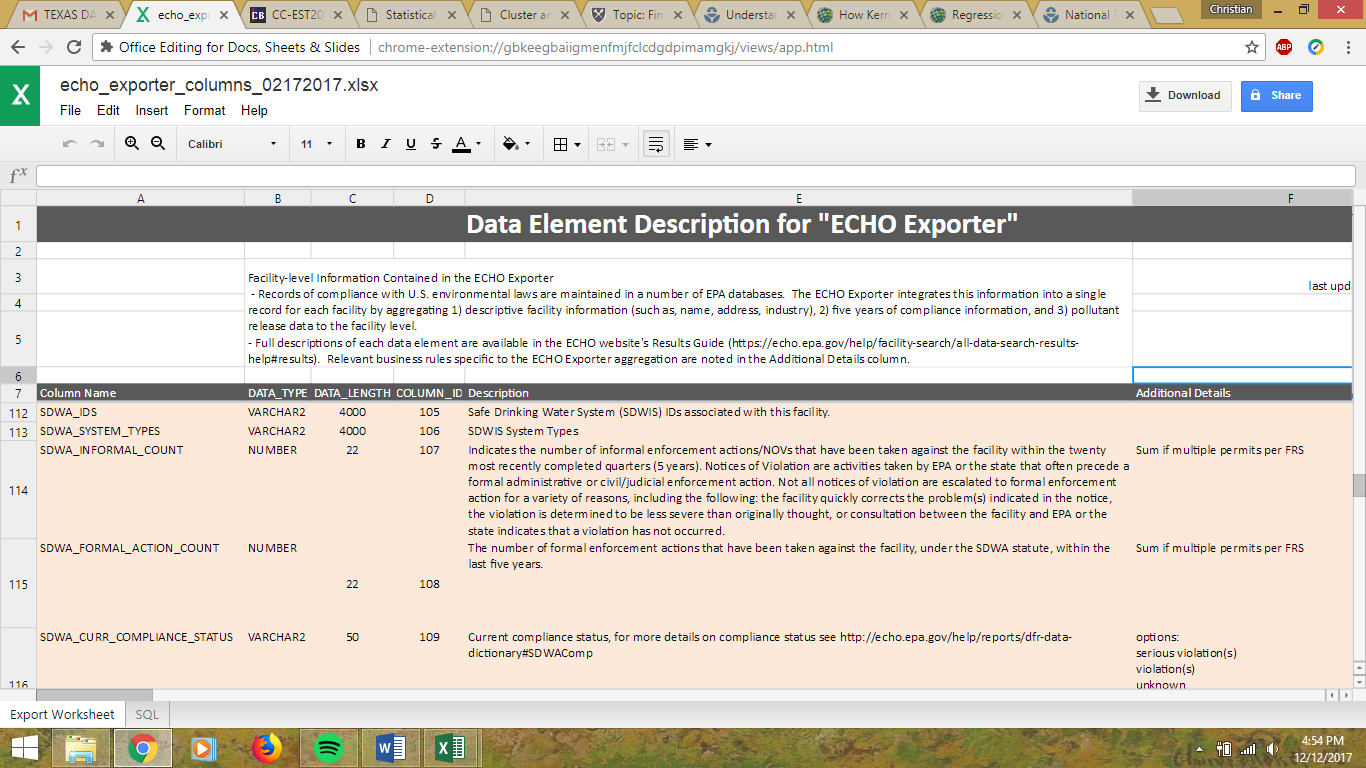
Data Sets

|  |  |  |  |
| --- | --- | --- | --- |
| **Data** | **Summary** | **Importance** | **Source** |
| **ECHO EXPORTER 2.0** | **Data and all facilities regulated by the EPA** | **Contains data on faculties that are regulated by the clean water drinking act** | [**https://echo.epa.gov/tools/data-downloads**](https://echo.epa.gov/tools/data-downloads) |
| **cc-est2014-alldata-48.csv** | **Annual County Resident Population Estimates by Age, Sex, Race, and Hispanic Origin: April 1, 2010 to July 1, 2014** | **Contains up to date demographic population projections for every county in the united states (downloaded Texas subset)** | **https://www2.census.gov/programs-surveys/popest/datasets/2010-2014/counties/asrh/** |
| **Texas Counties: Median Household Income** | Data source: U.S. Census Bureau. Small Area Income & Poverty Estimates. (2017) | **Contains medium income of each county in Texas.** | [**http://www.txcip.org/tac/census/morecountyinfo.php?MORE=1013**](http://www.txcip.org/tac/census/morecountyinfo.php?MORE=1013) |

Water Quality Data

Our water quality data was data downloaded from the EPA. The ECHO exporter dataset “includes facilities regulated as Clean Air Act stationary sources, Clean Water Act direct dischargers, Resource Conservation and Recovery Act hazardous waste handlers, Safe Drinking Water Act public water systems along with other data, such as Toxics Release Inventory releases, industry codes, and permit types. Latitude/longitude information also is included when available”5. The data was handled in Microsoft Access, and a query was made based on the following parameters: That the field included a SWDAID (indicating that it provided drinking water to a population and was regulated as such) and by its state field indicating that the facility was in Texas. This made a file that was 1.3 million records smaller, to approximately 15,000.

Fortunately, all facilities had latitude and longitude coordinates. This following screenshot explains the most important variables from this dataset and come from the data dictionary the EPA provided5.



Demographic data

The demographic data was taken from the United States census website. Specifically, the dataset titled “cc-est-2016-alldata”. This was a data set that contained census data for 2010 as well as population projections for 2011-2014. This dataset was put into Microsoft Access and was queried for data that was specific to counties in Texas and to the 2014 population projections, what was used as demographic data in my analysis. This demographic data was then joined to the county shape file that was worked with in the Hurricane Harvey lab at the beginning of the semester.

Income Data

This data was not modified in anyway, simply copied from a chart and joined to the Harvey shapefile by county name.

Maps

Data Frame Properties

WGS\_1984\_World\_Mercator

WKID: 3395 Authority: EPSG

Projection: Mercator

False\_Easting: 0.0

False\_Northing: 0.0

Central\_Meridian: 0.0

Standard\_Parallel\_1: 0.0

Linear Unit: Meter (1.0)

Geographic Coordinate System: GCS\_WGS\_1984

Angular Unit: Degree (0.0174532925199433)

Prime Meridian: Greenwich (0.0)

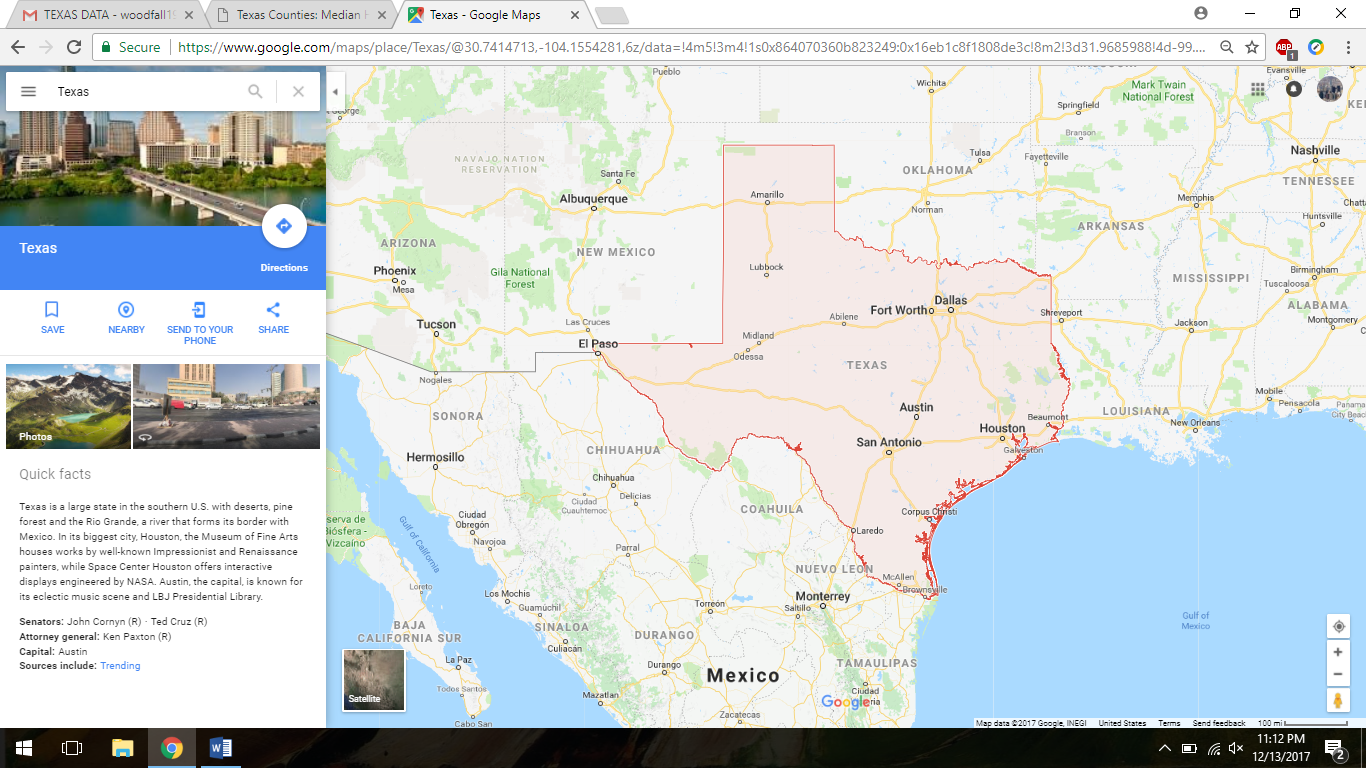
Datum: D\_WGS\_1984

Spheroid: WGS\_1984

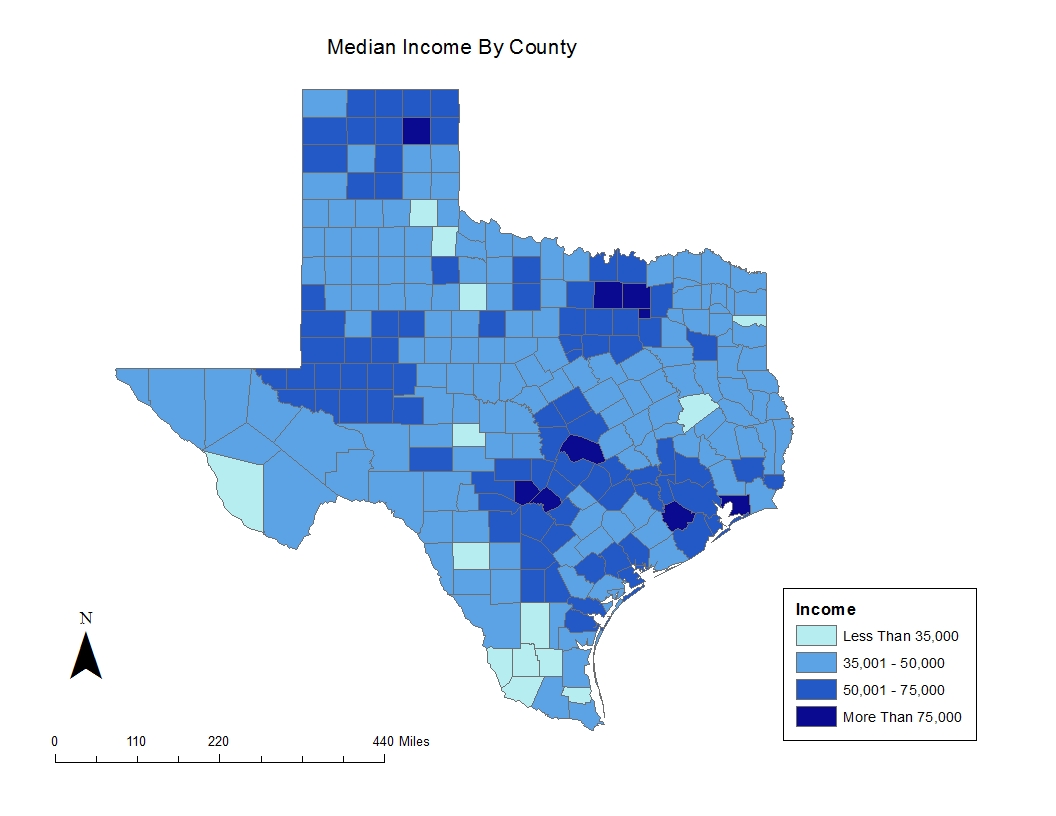
Semimajor Axis: 6378137.0

Semiminor Axis: 6356752.314245179

Inverse Flattening: 298.257223

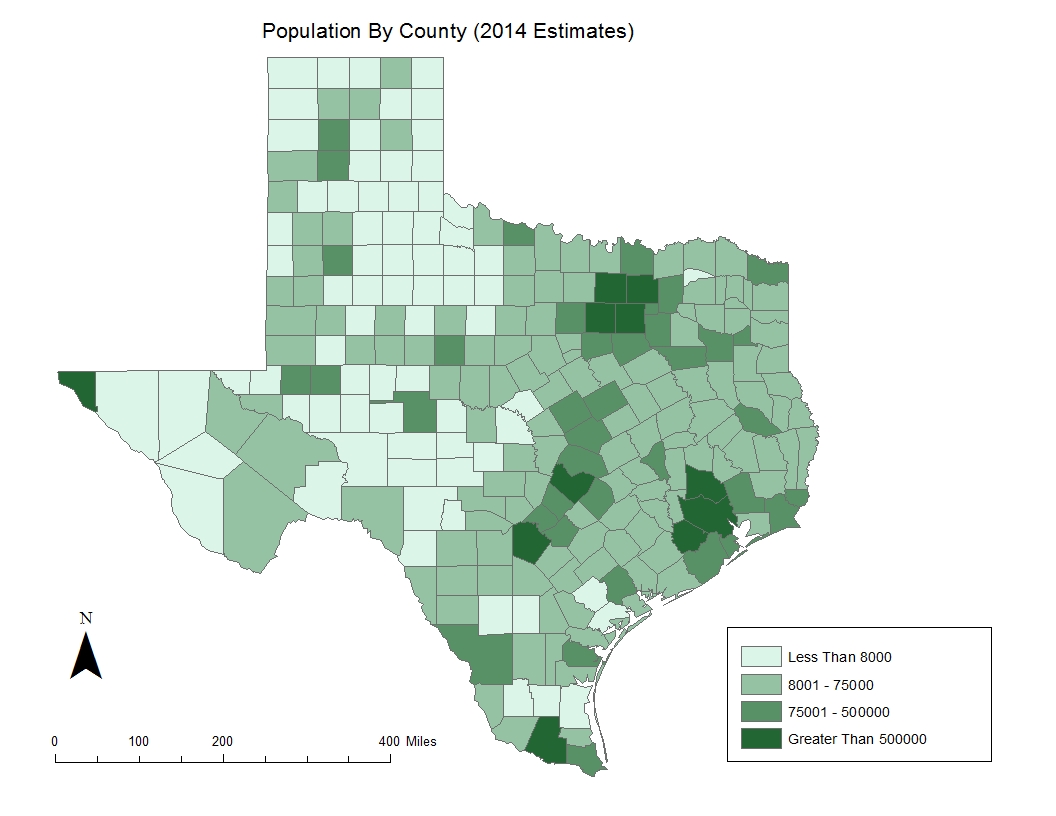
*Map of Texas (From Google Maps*

Mapping the Data: Income



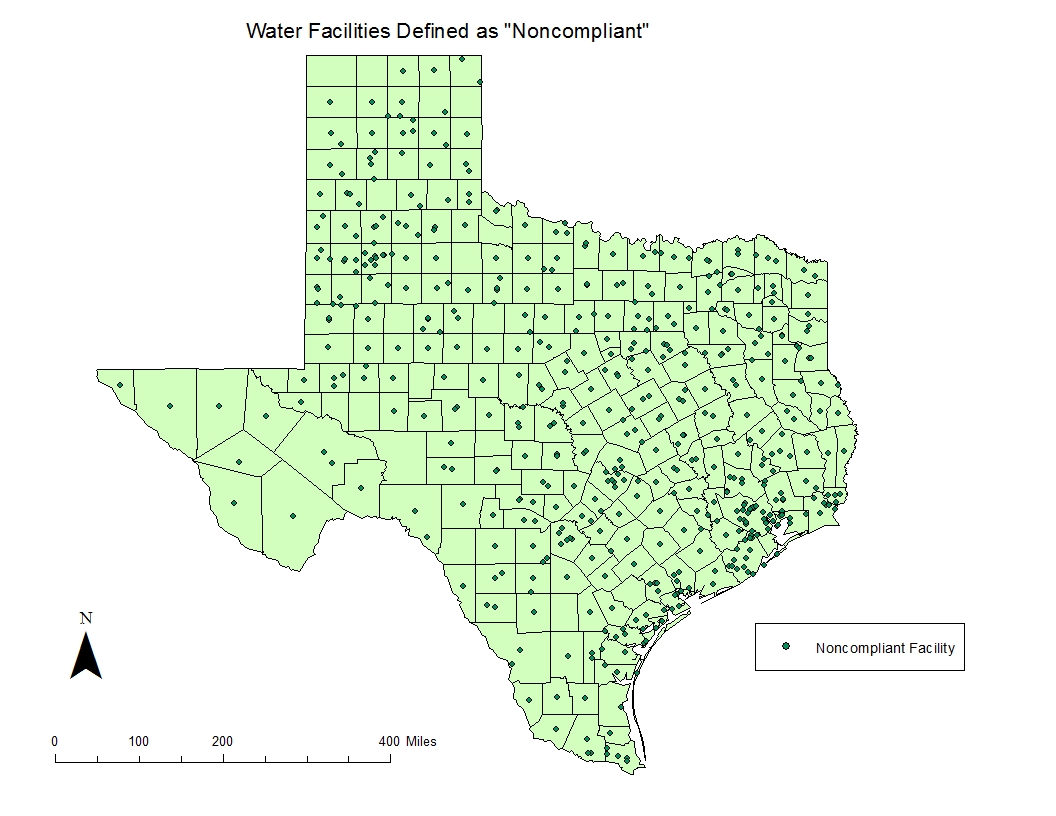
This is a simple choropleth map of median income by county. As can be seen the wealthiest counties also contain or are near major cities.

Mapping the Data: Demographic Map

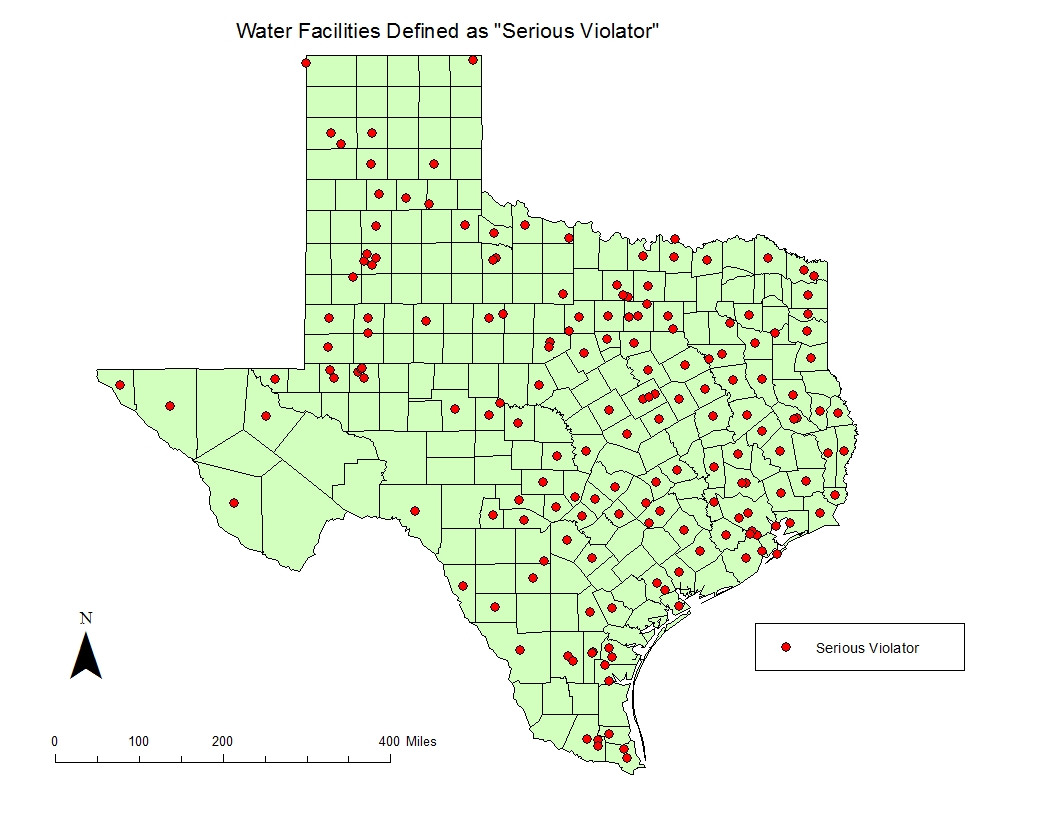


We see in the map above that the largest population centers are in the expected urban centers: Houston, Dallas/Fort Worth, Austin, El Paso to the West. East Texas has more counties with large populations than West Texas.

Mapping the Data: Compliance Status

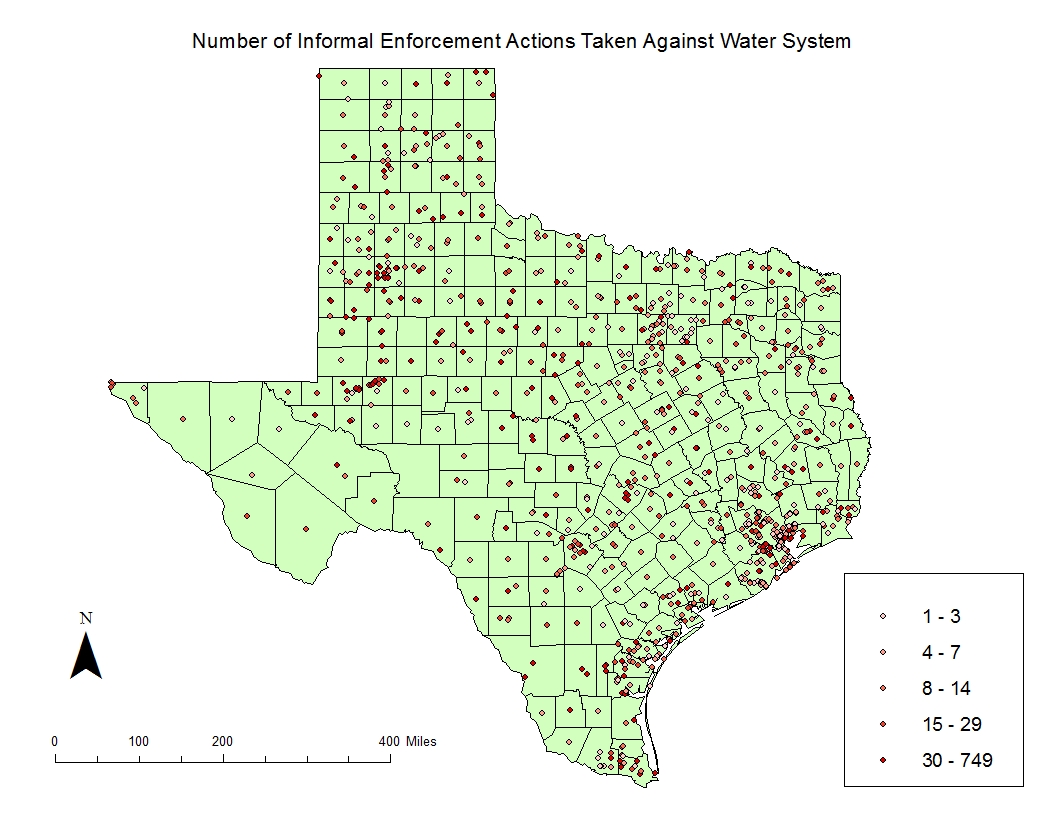


As mentioned earlier, in the dataset that I downloaded from the EPA website, there was a field that gave the classification by the EPA to each SDWA facility as either “complaint”, “noncompliant”, and “serious violator”. I first graphed all the non-complaint facilities to look at the data and see if I could identify any obvious trends. It was interesting to find that almost every county had a noncompliant facility, however there are little trends other than some tight clustering around the major populations centers of the state. Overall this map was not very useful. 2248 out of the 15111 records/facilities were classified as “noncompliant”.

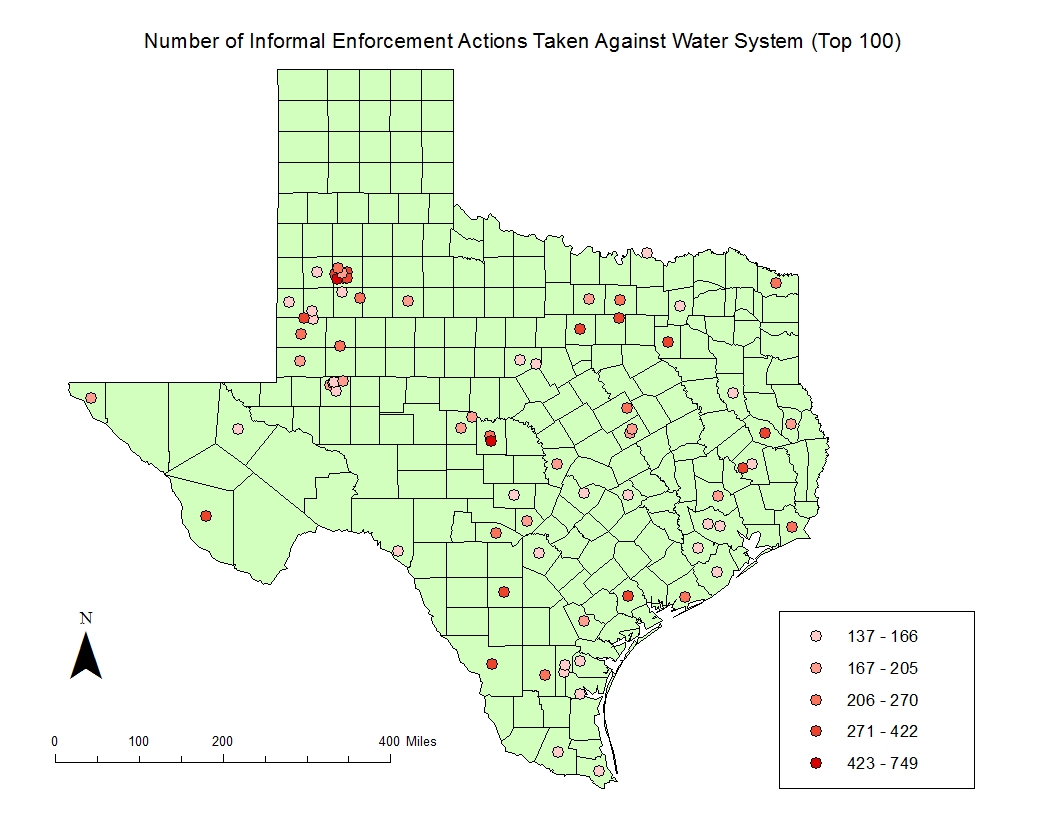


I decided to graph the “serious violators” from the EPA dataset. There are less facilities classified as serious violators than classified as noncompliant. While this map is cleaner than the previous, there is still not a very clear picture of where there are significant clusters. Except for the cluster in the Lubbock area (circled in blue), many of the facilities tend to follow population counts. This makes some sense as population centers would have more water treatment centers. It will take more specific mapping techniques to find any significant trends. I will revisit this data subset later in my analysis. 330 of the 15111 records/facilities were classified as serious violators.

Mapping the Data: Informal Actions Taken

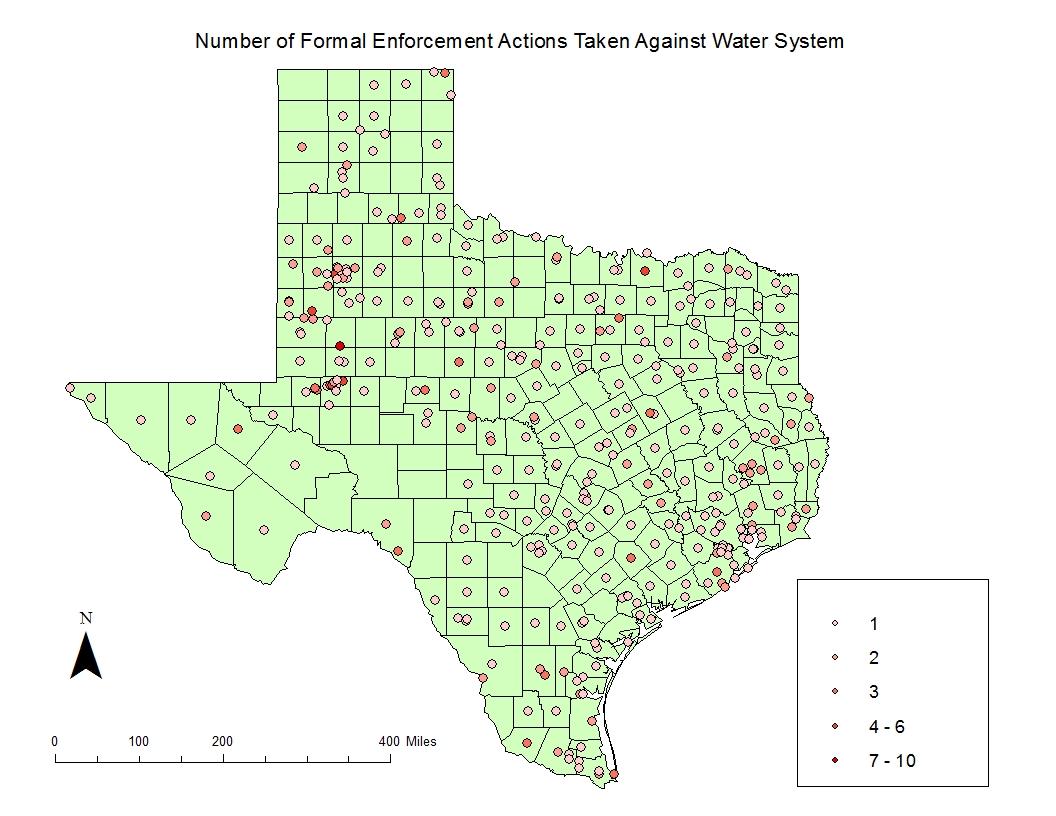


This map was made to get an idea of how many informal actions are distributed through the state. As can be seen, it seems almost every county has had at least one facility get a warning from the EPA. Note the outlier facility with 749 informal actions from the legend. This facility resides in the Lubbock area (circled).



To make sense of the data from the last map, I selected the top 100 sites (of 5,220 sites that had violations). We get a clearer picture than previous maps, of note is the clusters in Odessa and Lubbock area. This was puzzling to me why this area would have so many violations. (Purple circles added to map for emphasis).

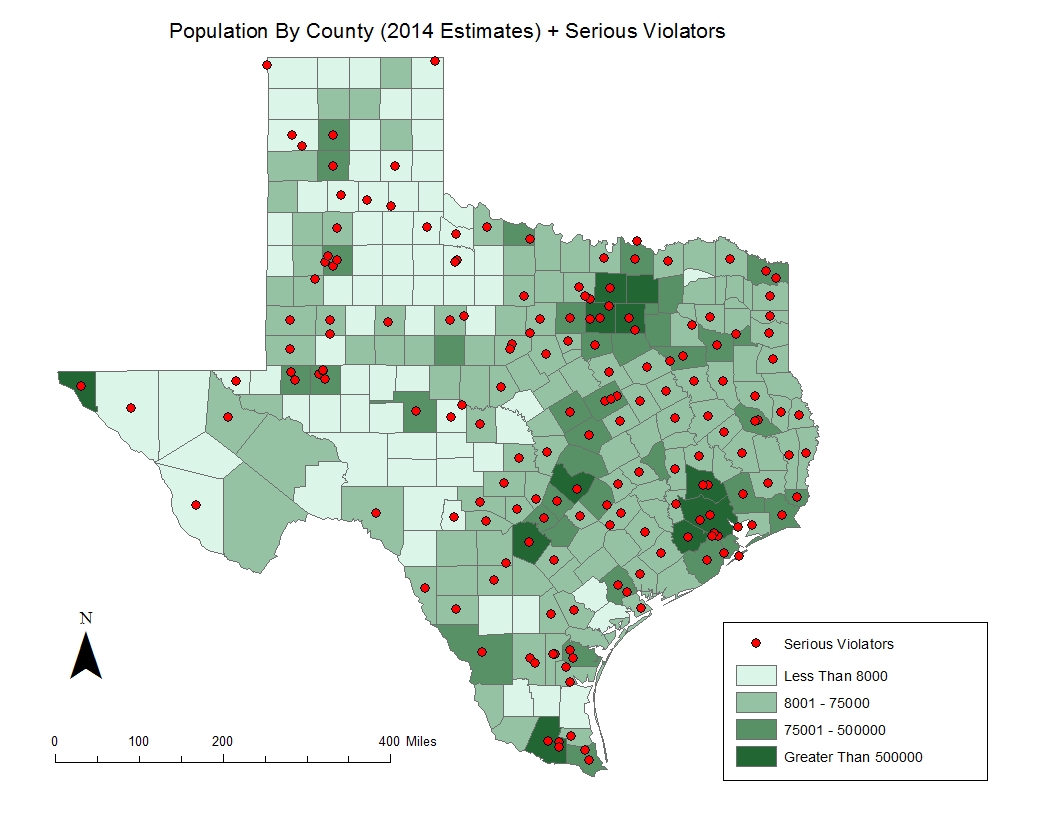
Mapping the Data: Formal Counts



Mapping the *formal* violations against each SDWA provider, we see that the point distributions resembles what was seen with the informal maps. The counts of formal action are generally lower as an informal count generally precedes the formal action from the EPA. Outliers are less common than the informal counts. There are less formal actions as facilities tend to resolve violations after an informal action of enforcement occurs.



Like the informal count maps, I mapped the facilities with top 100 formal counts. Visually it appears that the Houston area has a few facilities that have 4 or more counts. Again, the clusters in the Lubbock and Odessa persist with the eye-catching larger counts being there (circled in blue). A limitation of these simple plots is that they draw attention to mostly the more obvious trends or clusters. Later I will show kernel densities that will do a better job of mapping these counts.

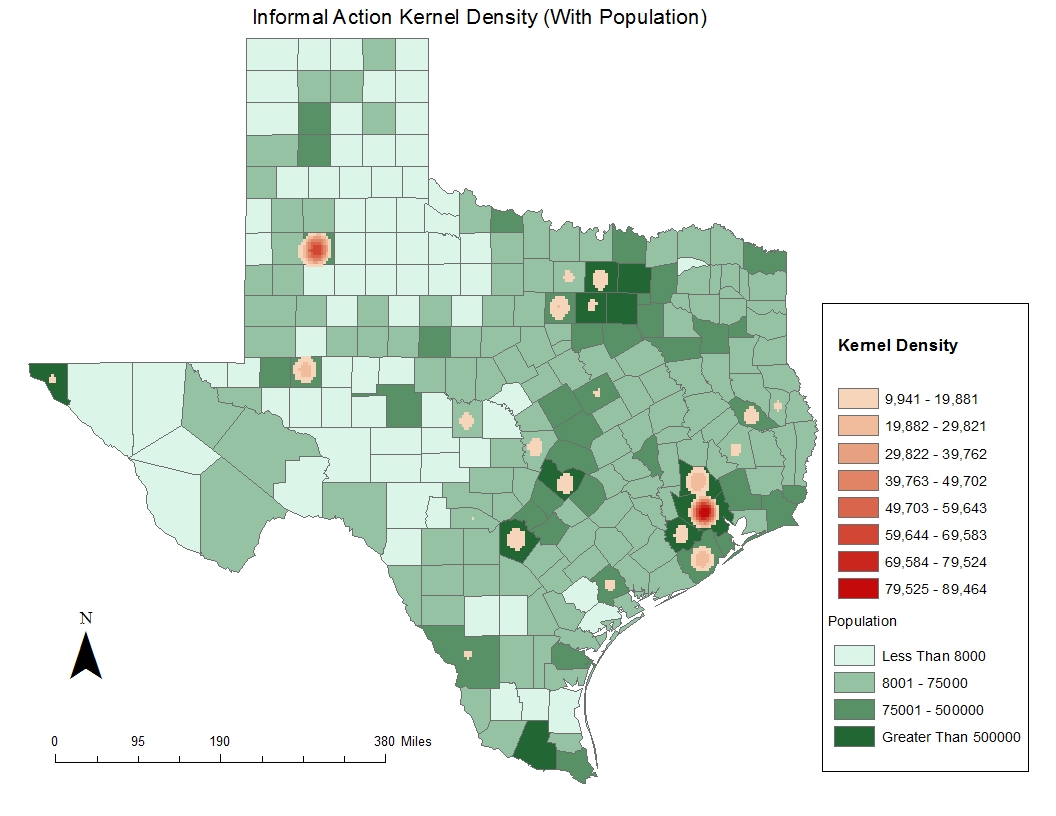


Now we begin exploring the relationship between the population data set and the EPA data set. I chose to begin with the serious violator layer. We see that for the most part the serious violators follow population trends, this map doesn’t tell too much different from when the serious violator layer was the only layer mapped.

Kernel Density Maps



Mapping the kernel density of formal actions counts, we get a much better picture of the hotspots in the state. It becomes obvious that much of the formal counts are in the greater Houston metro area. We still have some hotspots located in Lubbock and Odessa areas. It is surprising that there is a hotspot in the Odessa region, given that this region along with Lubbock are not large population centers like Houston, Dallas, etc.

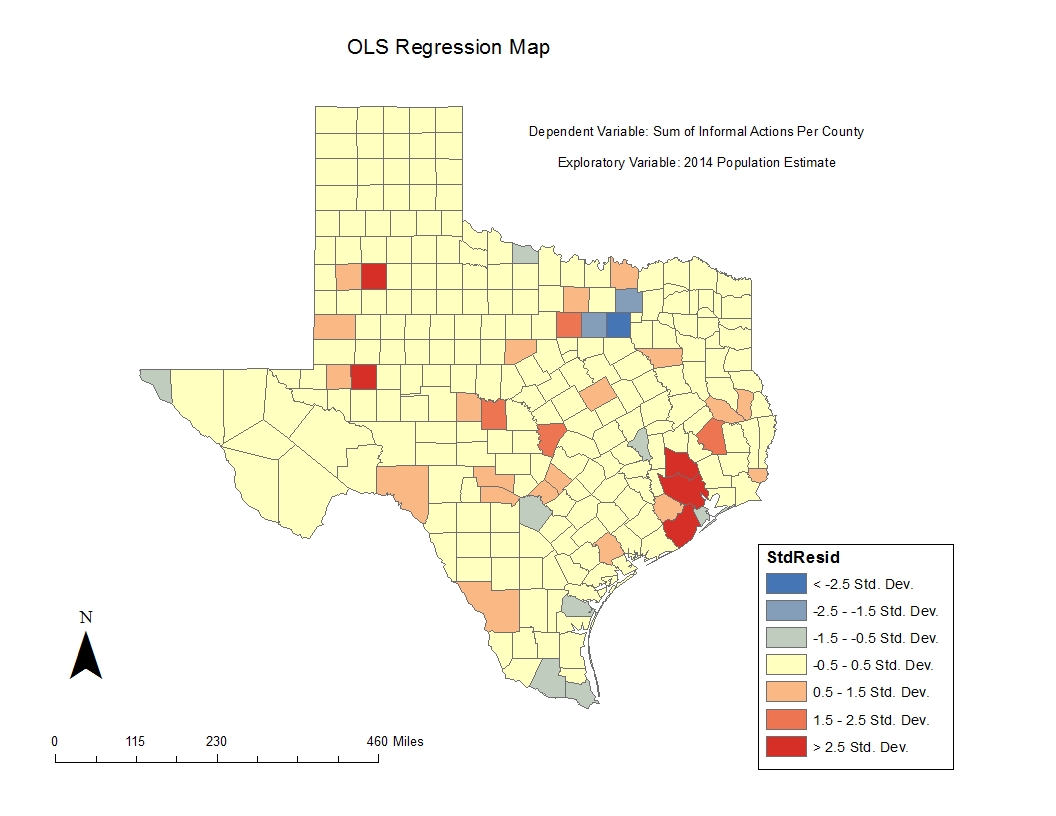


With the informal actions taken we see that large volumes of the informal counts are concentrated in the Houston and Lubbock areas. This is interesting as the formal count kernel density was more evenly distributed around the population centers, but this map depicts tighter clusters.

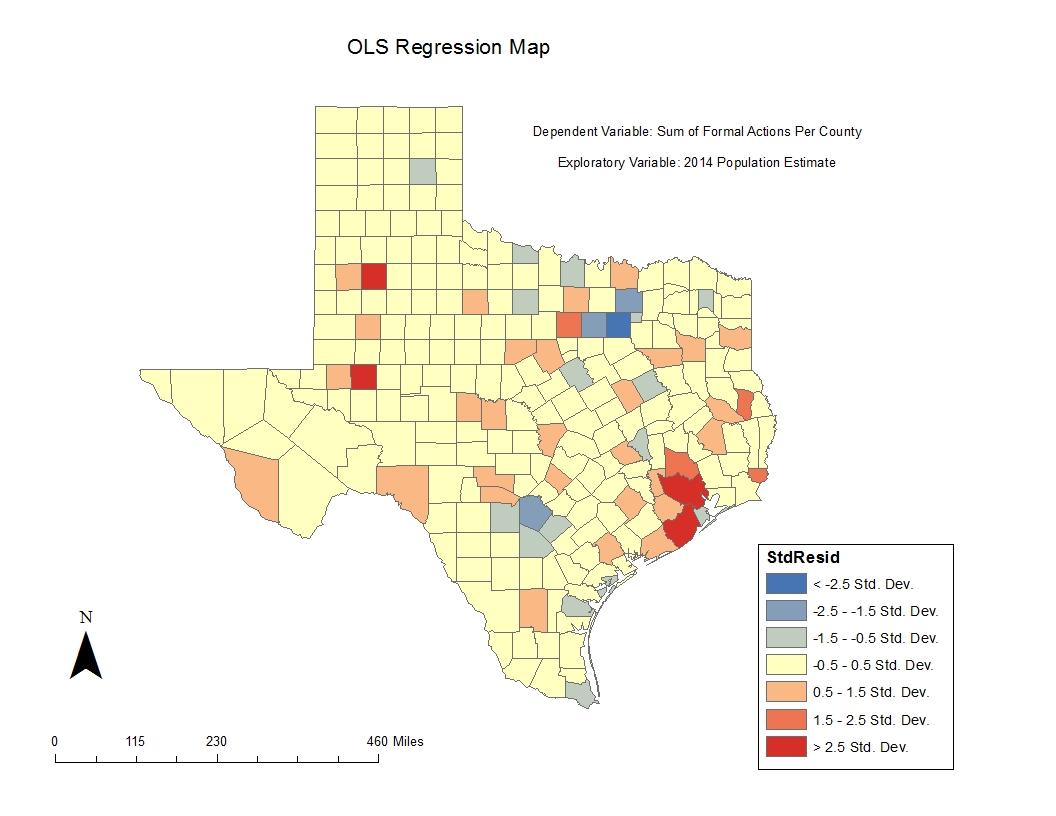
Ordinary Least Squares Regression Models

After looking at the maps that I had created up until this point, I felt confident identifying hot spots. However, I had the suspicion that I was mostly observing a simple positive correlation between population centers and EPA action. In other words, the chance of a facility having an EPA action may be uniform across most of the state and that the hotspots simply corresponded to the large amount of water facilities in the area. To more confidently identify outliers and see if this was in fact the case I used the ordinary least squares regression tool to come up with some maps.

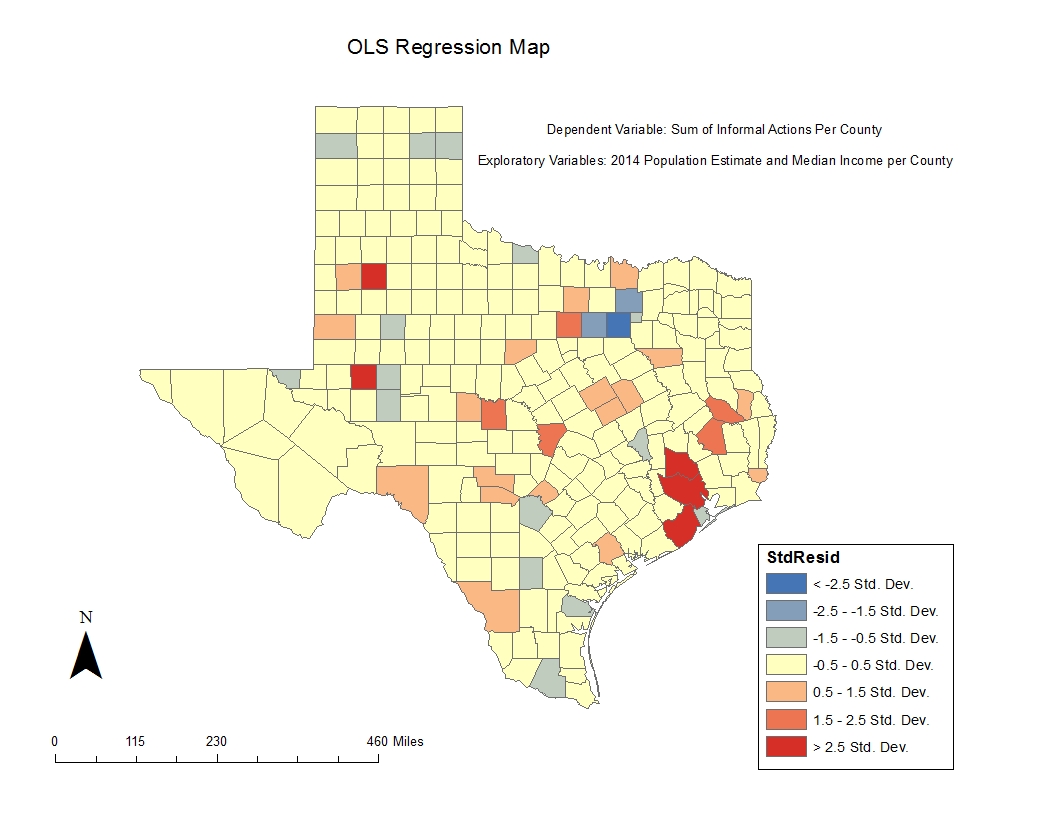
A spatial join was made between individual points (facilities) and the county lines that they fell in. This generated a shapefile with two separate fields for the sum of formal and informal acts of enforcement per county respectively.



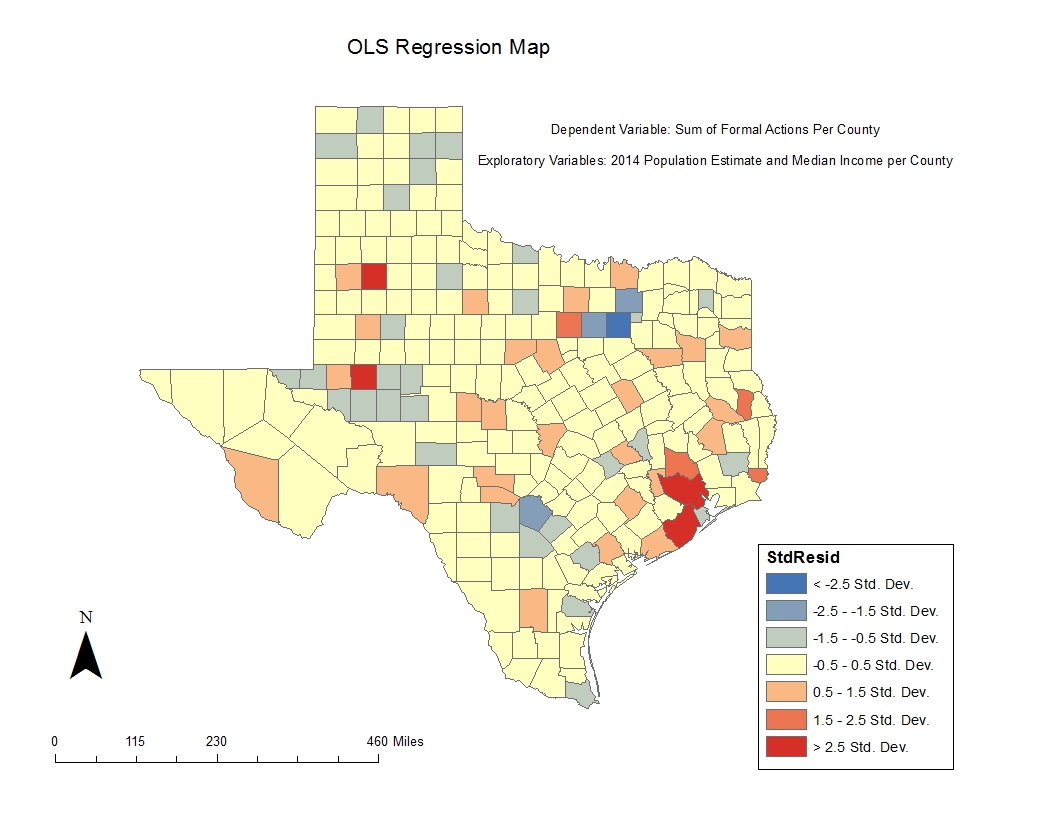
First was a look at a model where the dependent variable was the sum of informal actions per county and the exploratory variable was the county’s 2014 estimated population count. It was interesting to find that the hot spots that were seen in the kernel density map did not fit the linear model, with the Houston, Lubbock and Odessa area all being over 2.5 standard deviations above the linear model. This suggests that despite the population of these counties, they have very high counts of informal actions taken against the facilities within these counties’ borders. The most interesting finding was that the Dallas-Fort Worth metropolitan area, despite its high population count, had much lower informal actions than would be expected in this linear model. The same can be said for the El Paso and Brownsville metro areas (the westernmost and southernmost portions of the Texas respectively).



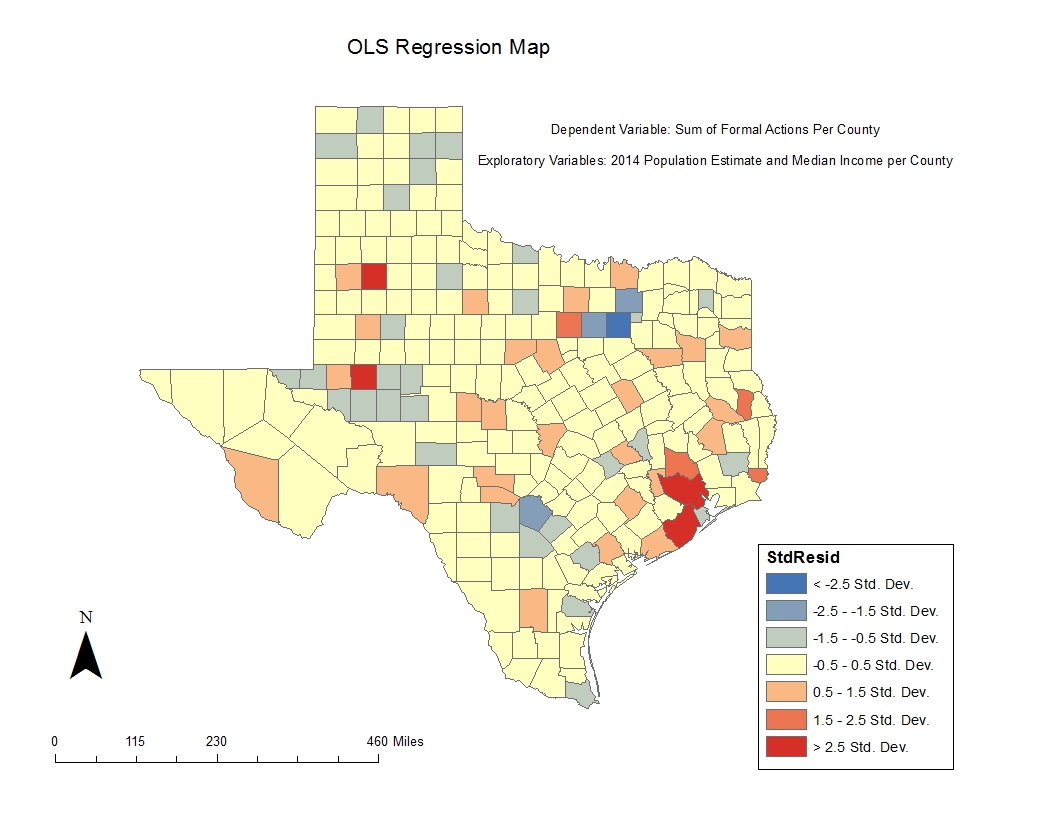
Like the previous map, I wanted to look at formal actions taken as the dependent variable, with the exploratory variable being 2014 population estimates. The outliers remain roughly the same as informal map, which was expected.



For this map and the following map, I repeated the OLS analyses with an important distinction: I added median income per county as an exploratory variable. I understood that this may complicate the model but was interested to see what differences arose. Considering that income and population seemed to be positively correlated, there was little change, with this map bearing similarities to the OLS map that had only informal actions and population as variables.



The trends in the OLS counterpart without income are repeated here. It is worth mentioning that as can be seen with the income map, that the panhandle (the northernmost “handle” shaped portion of Texas) has counties with high income compared to the rest of the state (circled in blue). As a result, they are registered as outliers in this model, given that other counties with high income have much many EPA actions than the panhandle counties.



Like the last map, we have both income and population as exploratory variables for formal actions per county. This map is similar to the previous map.

Analysis

Strengths

One of the strengths of looking at the entire state is that we can get a grasp on where the problem of water quality in Texas resides. Though it isn’t surprising to find that most of the faculties with offenses are in population centers in the state, it was a surprise to find that there was a high density of EPA violations in the Lubbock and Odessa area. This finding would not have been possible with a more specific sample. The large sample size also allows for more accurate regression models and make the OLS maps more useful.

Limitations

There are many limitations to this analysis. First, no detailed regression analysis was performed, so it’s hard to draw any strong statistical conclusions of these maps. However, a detailed regression analysis would be a good second step. We are using EPA regulation as a proxy for water quality, not actually using water data. Using actual water quality data was considered, however it was difficult to differentiate drinking water sources, as the records from the USGS website measure water that has not been treated and may not necessarily used for drinking. Variance in EPA enforcement could have affect the water data we have. We have not looked at data for the rest of the county, so we cannot compare the results of Texas to the rest of the country.

Findings and Conclusion

From this exploratory analysis I was able map a large data set and find where water quality is bad compared to the rest of the state. We found that the Houston metro area and surrounding counties have many safe water drinking facilities that have received many enforcement actions from the EPA. Surprising was that the Houston area was significantly higher than the other large metro area in the state of Texas (Dallas-Ft Worth). A surprise finding was the comparatively smaller cities of Odessa and Lubbock having a large amount of enforcement actions as well.

Houston has many logistical problems in approaching these water quality problems given the aftermath of Hurricane Harvey. However, Lubbock and Odessa having water quality problems suggest a follow-up analysis on why this why this may be the case.

Knowing that water quality in Texas is among the worst in the country, it is discouraging to see that there almost every county in the state has a facility that has received enforcement action by the EPA.

We see that it is difficult to separate median income per county and population. The scope of analyzing every county based on demographic variables was a project of much larger scope than anticipated, and I hope to revisit the topic after becoming more acquainted with the regression tools available within ArcGIS.

Bibliography

1.Fedinick, K., Ph.D., & Wu, M., MPhil, J.D. (n.d.). THREATS ON TAP: WIDESPREAD VIOLATIONS HIGHLIGHT NEED FOR INVESTMENT IN WATER INFRASTRUCTURE AND PROTECTIONS (p. 15, Rep. No. 17-02-A). NRDC.

https://www.nrdc.org/sites/default/files/threats-on-tap-water-infrastructure-protections-report.pdf

2. (2014, July 11). Wasting Our Waterways: Toxic Industrial Pollution and Restoring the Promise of the Clean Water Act. Retrieved December 12, 2017, from <https://www.infrastructureusa.org/wasting-our-waterways-toxic-industrial-pollution-and-restoring-the-promise-of-the-clean-water-act/>

3. Drinking Water Contaminants – Standards and Regulations. (2017, May 22). Retrieved December 12, 2017, from <https://www.epa.gov/dwstandardsregulations>

4. Understanding the Safe Drinking Water Act - US EPA. (n.d.). Retrieved December 12, 2017, from <https://www.epa.gov/sites/production/files/2015-04/documents/epa816f04030.pdf>

5.Data Downloads | ECHO | US EPA. (2017, October 06). Retrieved December 12, 2017, from https://echo.epa.gov/tools/data-downloads