



TOXIC CRUSADERS

CSC 495 Project - Fall 2017

Aaron Denton
Jacob Durham
Swetha Polisetty
Melvin Watlington



Project Goals



- Explore Toxics Release Inventory (TRI) and Zillow Home Value Index (ZHVI) data.
- Find trends in these data that reveal the impact that industrial production has on communities.
- Visualize changes over the last 20-30 years in both industrial chemical emission and house values.
- Compare the benefits and disadvantages of industrial production facilities near populated areas.



Toxics Release Inventory *History*



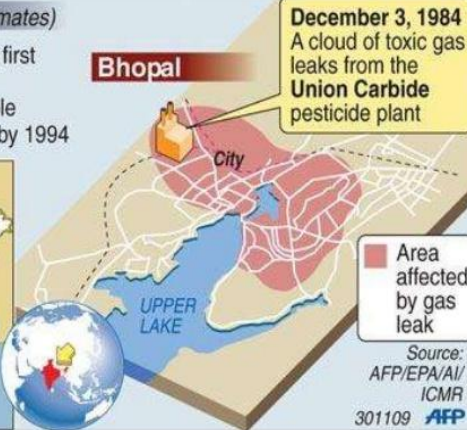
From the EPA:

“On December 4, 1984, a cloud of extremely **toxic methyl isocyanate gas** escaped from a Union Carbide Chemical plant in Bhopal, India. Thousands of people died that night... considered to be the worst industrial disaster in history. Thousands more died later... survivors continue to suffer with permanent disabilities. In 1985, a serious chemical release occurred at a similar plant in West Virginia.

In 1986, Congress passed the Emergency Planning and Community Right-to-Know Act (EPCRA) to support and promote emergency planning and to provide the public with information about releases of toxic chemicals in their community. Section 313 of EPCRA established the **Toxics Release Inventory**.”

The human cost (estimates)

- ▶ Up to 10,000 deaths in first three days
- ▶ Additional 25,000 people died of related injuries by 1994



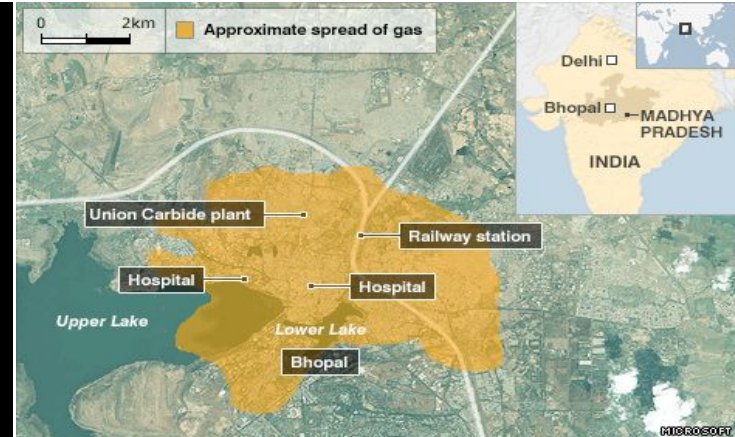
Toxics Release Inventory *History*



From the EPA:

*“On December 4, 1984, a cloud of extremely **toxic methyl isocyanate gas** escaped from a Union Carbide Chemical plant in Bhopal, India. Thousands of people died that night... considered to be the worst industrial disaster in history. Thousands more died later... survivors continue to suffer with permanent disabilities. In 1985, a serious chemical release occurred at a similar plant in West Virginia.*

*In 1986, Congress passed the Emergency Planning and Community Right-to-Know Act (EPCRA) to support and promote emergency planning and to provide the public with information about releases of toxic chemicals in their community. Section 313 of EPCRA established the **Toxics Release Inventory.**”*





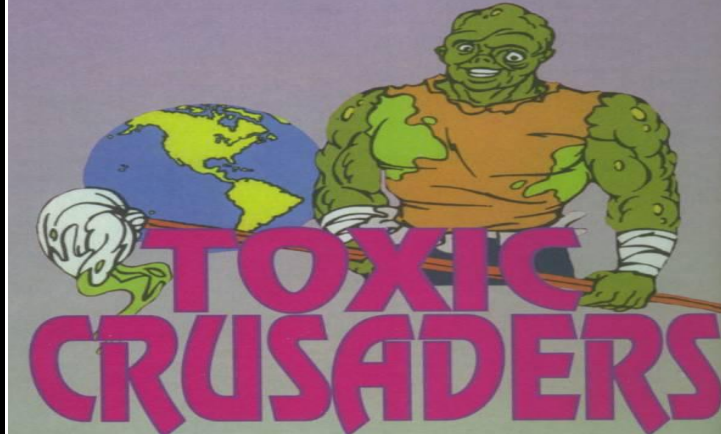
Toxics Release Inventory

Cultural Influence



Interesting side-note:

- Public concern over events like the Bhopal Tragedy motivated not only the creation of the EPCRA, but also created a generation of **cartoon mutants** to make kids aware of these issues (and to sell action figures!).



Toxics Release Inventory *Data Sets*



- The EPA provides data sets containing TRI data for the years 1987 through 2016.
- Facilities across many different industrial sectors must submit annual reports of chemicals released to the environment and/or managed through recycling, energy recovery, and treatment.
- In this context, “release” of a chemical means that it is emitted to the air or water, or placed in some type of land disposal.



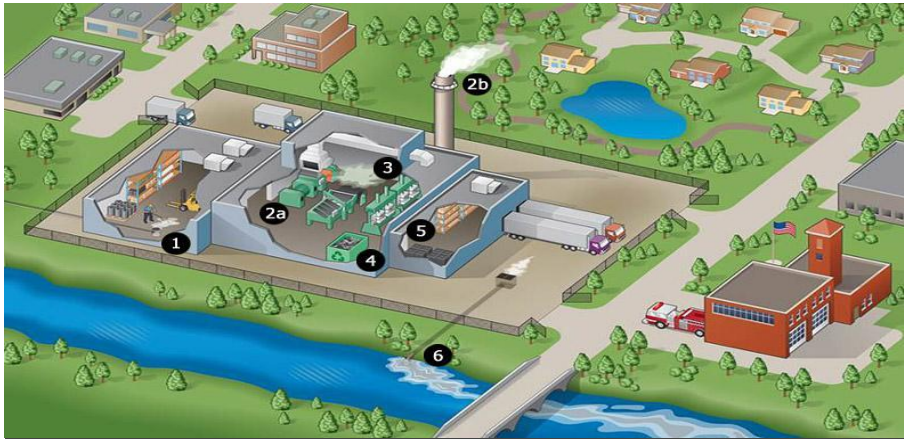
Toxics Release Inventory Data Sets



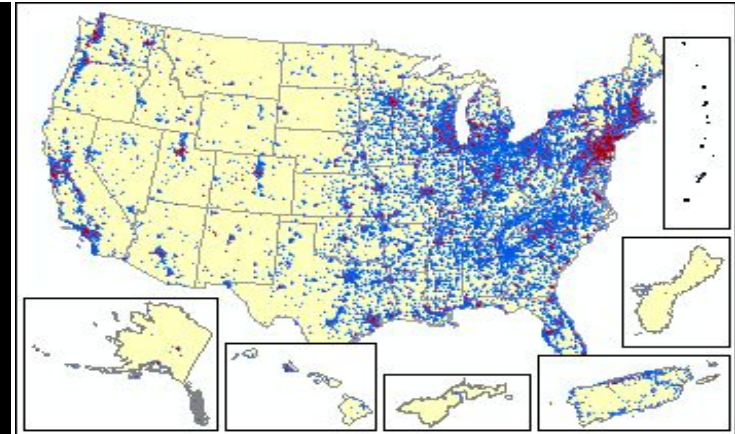
- The EPA site states that the purpose of this data is *“for citizens to better understand possible sources of pollution in their communities.”*
- Our goal in exploring this data, then, is to extract relevant relations and trends that will shed light on understanding how these toxic releases affect the surrounding communities.
- We will accomplish this by comparing the trends in these data to those that we find in the Zillow House Value Index data, and the relations between the two.



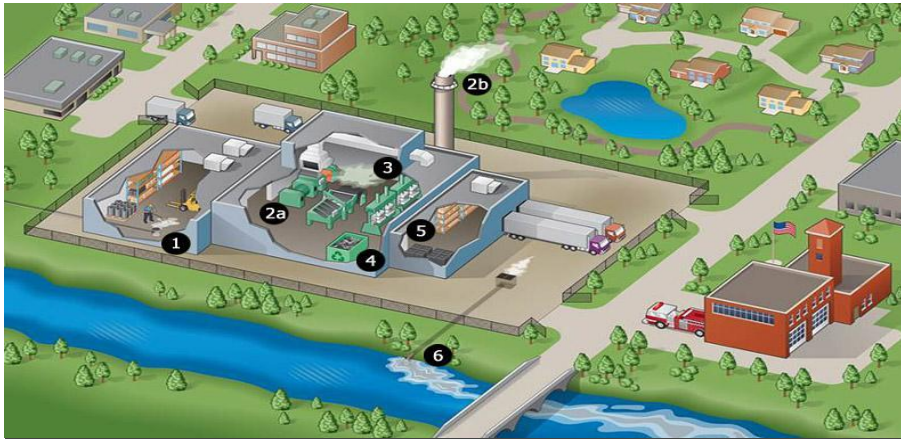
Toxics Release Inventory Data Sets



- We are using the 'National' data files, which contain yearly TRI for all 50 U.S. states and the six districts and territories, and 'State' data files, which contain yearly TRI data for each state individually.
- Each data file (in CSV format) contains 108 attributes, which fall into six categories...



Toxics Release Inventory Data Sets



(1) Facility Information

- Facility Name
- Address
- Latitude & Longitude Coordinates
- SIC codes (4-digit industrial classification)
- NAICS codes (6-digit industrial classification)
- Industry Sector Codes

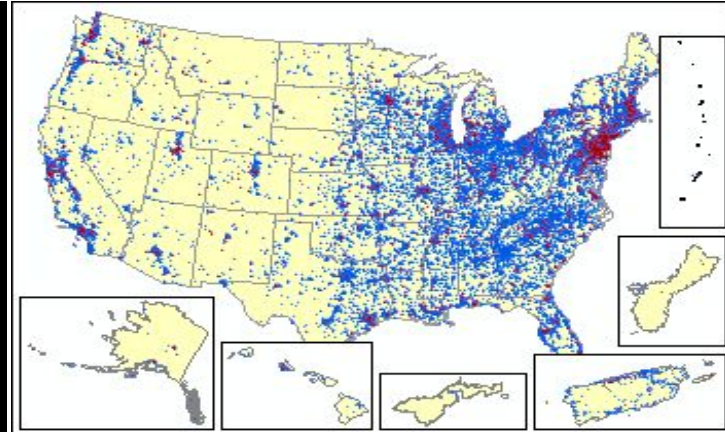
(2) Chemical Identification and Classification Information

(3) On-site Release Quantities

(4) Publicly Owned Treatment Works (POTW) Transfer Quantities

(5) Off-Site Transfer Quantities for Release/Disposal and Further Waste Management

(6) Summary Pollution Prevention Quantities



Missing data...

Toxics Release Inventory *Data Sets*

Zeroes in the TRI Data Sets

Numeric data fields may contain zero (0) for three reasons:

- Release of chemical below threshold for reporting
- Chemical reported as “NA” for that particular industry
- Left blank in hand-written submission of reporting form





Zillow Home Value Index

Background

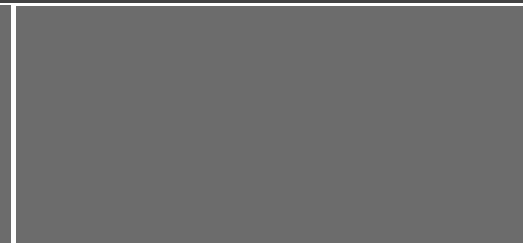


- Zillow is an online real estate database company, founded in 2006.
- They provide many real estate related services, including real estate loan quotes, neighborhood boundary mapping, and home value estimates.
- Their home value estimates, called *Zestimates*, are based on a proprietary algorithm, and serve as the basis of the Zillow data set that we are leveraging for this project.





Zillow Home Value Index Data Sets



- **Zillow Home Value Index (ZHVI) is a collection of Zestimates, split into 12 market segments.**

- All Homes
- Single Family
- Condo
- 0 or “missing” Bedrooms
- 1/2/3/4/5+ Bedrooms
- Top Tier (among homes in same metro area)
- Middle Tier
- Bottom Tier

| Market Segment | National | State | MSA | County | City | Neighborhood | Zip | Total |
|----------------|----------|-------|-----|--------|--------|--------------|--------|--------|
| All Homes | 1 | 51 | 917 | 2,830 | 23,057 | 8,664 | 24,460 | 59,980 |
| Single Family | 1 | 51 | 917 | 2,828 | 22,976 | 8,068 | 24,249 | 59,090 |
| Condo | 1 | 51 | 507 | 895 | 4,189 | 2,916 | 6,629 | 15,188 |
| 0 or missing | 1 | 51 | 868 | 2,464 | 14,023 | 3,304 | 15,097 | 35,808 |
| 1 Bedroom | 1 | 51 | 537 | 1,097 | 2,112 | 1,080 | 3,418 | 8,296 |
| 2 Bedroom | 1 | 51 | 742 | 1,821 | 9,173 | 3,083 | 11,870 | 27,461 |
| 3 Bedroom | 1 | 51 | 817 | 2,105 | 13,310 | 5,523 | 15,796 | 37,603 |
| 4 Bedroom | 1 | 51 | 766 | 1,829 | 8,633 | 3,124 | 11,485 | 25,889 |
| 5+ Bedroom | 1 | 51 | 619 | 1,249 | 3,524 | 1,018 | 5,648 | 12,110 |
| Top Tier | 1 | 51 | 913 | 1,681 | 12,554 | 4,112 | 14,862 | 34,184 |
| Middle Tier | 1 | 51 | 913 | 1,704 | 14,058 | 4,877 | 16,364 | 37,968 |
| Bottom Tier | 1 | 51 | 913 | 1,676 | 12,941 | 5,119 | 15,173 | 35,874 |

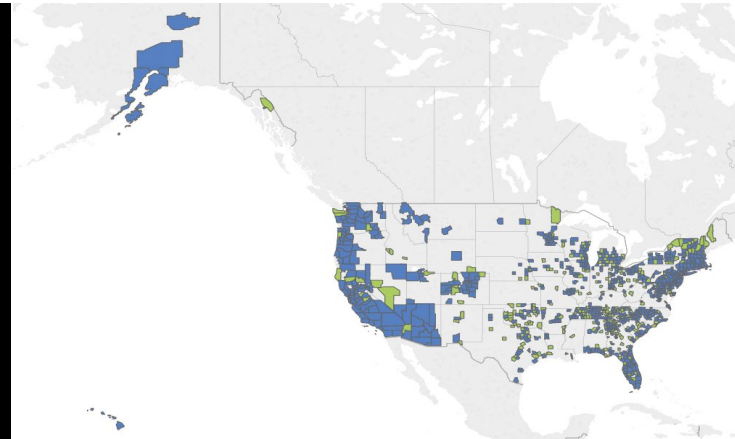




Zillow Home Value Index *Data Sets*



- **Using the Zestimates for residential homes, the ZHVI data is created in a 5 step process:**
 - Calculate Raw Median Zestimates
 - Adjust for Residual Systematic Error
 - Apply Henderson Moving Average Filter
 - Apply Seasonal Adjustment
 - Final Quality Control
- **Zillow provides data sets for each market segment, ranging from 1996-2016 (although this data contains lots of gaps for the first few years)**

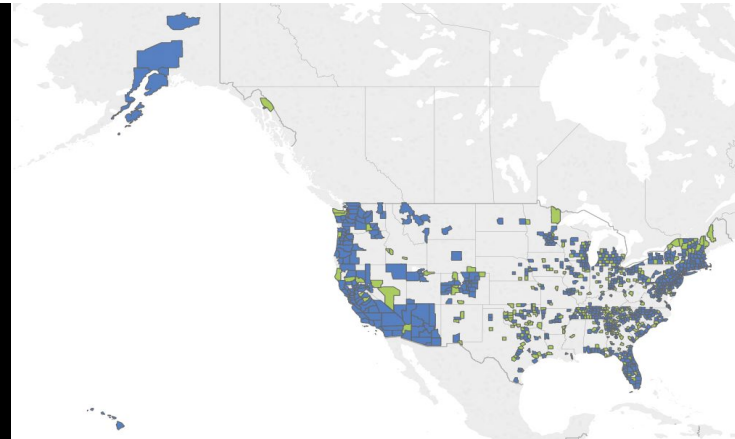


Combining The Data

TRI + ZHVI



- Between the TRI and ZHVI data sets, there is a tremendous amount of information to explore.
- By comparing ZHVI trends to temporal changes in TRI data, we hope to gain insight on the affect that industrial chemical release might have on home values.
- In addition to this combined purpose, there are many interesting questions that we can ask about each data set individually.



Combining The Data

Primary Questions



As of now, the main observations about these data that we are interested in include:

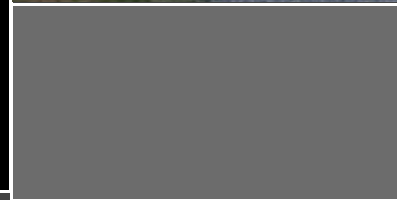
- **Observation of ZHVI variation in different categories of housing (1 bedroom, middle tier, condo, etc...)**
- **ZHVI that is most varied:**
 - In terms of depreciation by industry sector
 - In terms of appreciation by industry sector
- **ZHVI changes depending on:**
 - On-site releases
 - Off-site transfers and POTWs
 - Other waste management



Combining The Data *Game Plan!*



- To start, we will study trends and relationships between the data in a single state, and then expand to the national data by regions, possibly splitting U.S. regions among all team members.
- As we continue, we will identify interesting features of the data, and find ways to visualize our research that will highlight the impact that industrial production has on surrounding communities.
- Another Zillow data set, the Zillow Rental Index (ZRI), might be interesting to explore, as the ratio of rentals versus home purchases is on the rise in recent years, even among families.



To Be Continued...



TOXIC CRUSADERS

CSC 495 Project - Fall 2017

Aaron Denton
Jacob Durham
Swetha Polisetty
Melvin Watlington





TOXIC CRUSADERS

CSC 495 Project - Fall 2017 - Report 2

Aaron Denton
Jacob Durham
Swetha Polisetty
Melvin Watlington



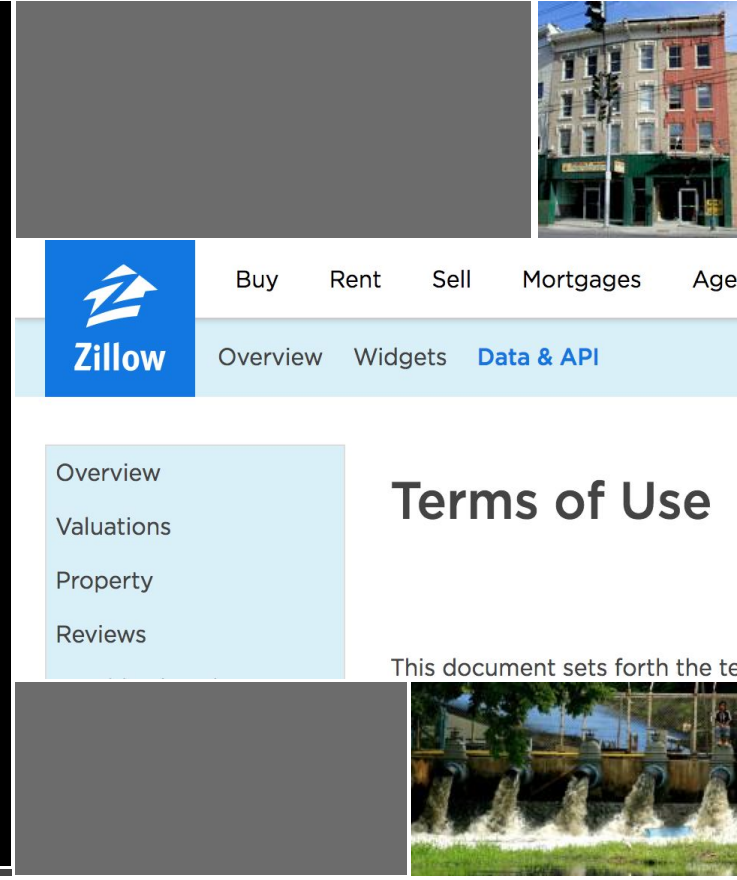
Revised Goals

- Explore Toxics Release Inventory (TRI) data.
- Find trends and relations in these data to characterize facilities, parent companies, and industrial sectors in terms of toxic release and recycling.
- Visualize changes over the last 20-30 years in industrial chemical emissions on national and state levels.



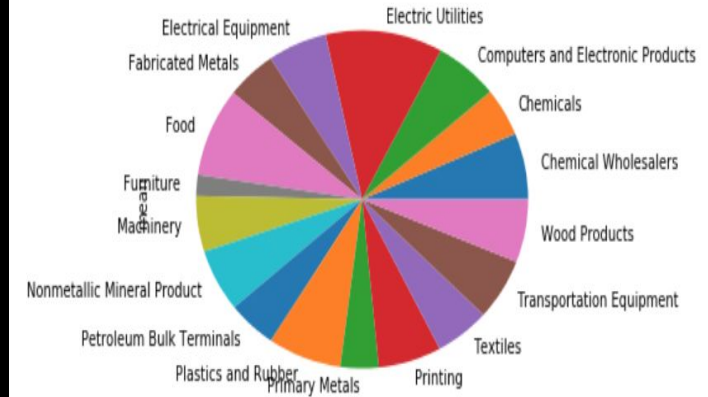
ZHVI Issues

- ZHVI data include ZIP codes and Region IDs, no latitude and longitude data
- Public ZHVI web API has Regions and Subregions, but are scattered and time consuming to match
- Public API also has limited calls
- Neighborhoods are non-standard sizes

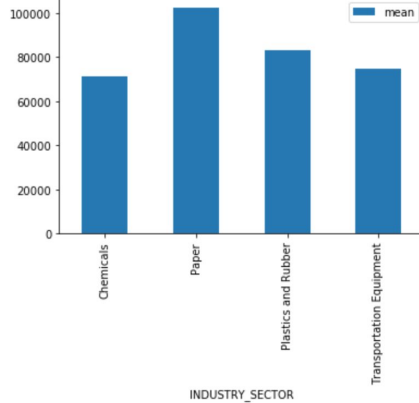
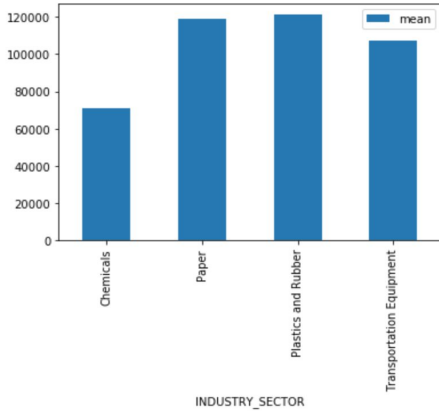


ZHVI Issues

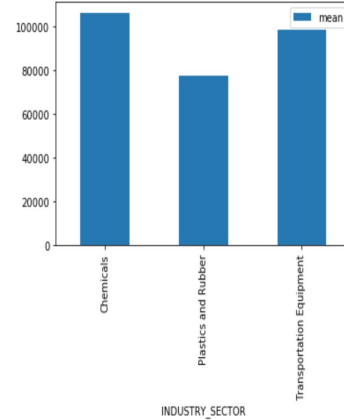
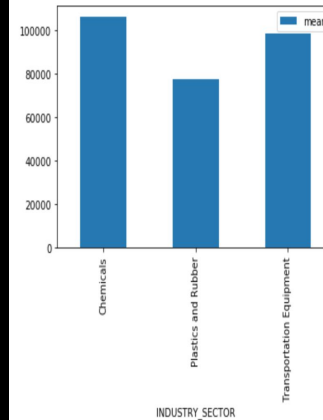
- Potentially use ZIP code centroids, but that has problems
- ZIP code geographic sizes greatly vary, makes it difficult to compare impact of TRI facilities
- Using ZIP code centroids, initial analysis revealed nearly zero impact on home values from close-by facilities



ZHVI: Going Forward



- Although, a certain pattern was observed with the average price of homes(considering the centroid) with respect to their distances from the TRI industries.
- For the Chemical Industry Sector, it was noticed that the average home prices(1-,2-,3-,4- bedrooms,condos) increase as their distance from the industry facility increases.
- Continue to look into these trends going forward, perhaps secondary goal now



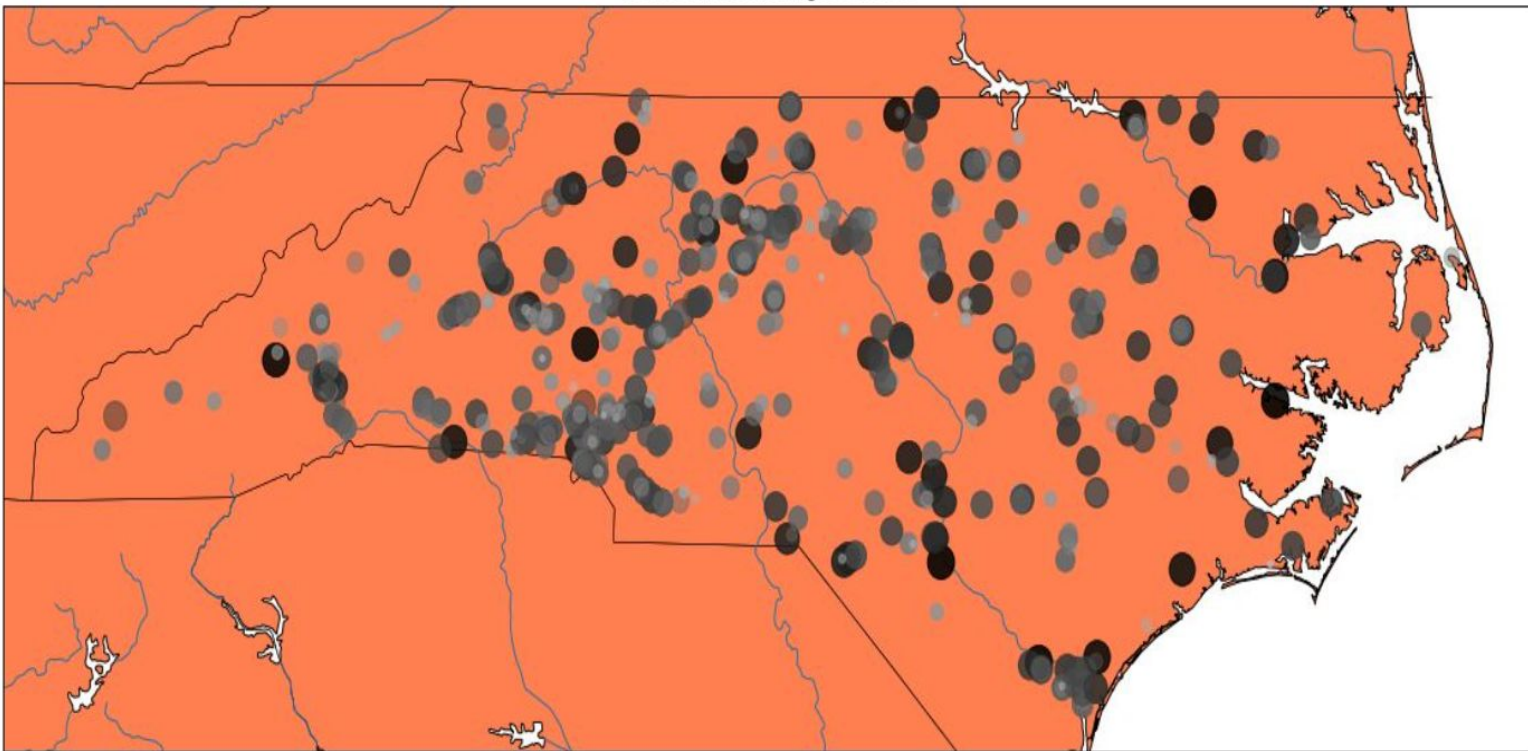
New Goal: Explore TRI Data

- To make the best use of our remaining time, we have chosen to focus on the TRI data sets
- ~30 years of data, over 100 columns
- Many interesting questions to be explored on national and state levels



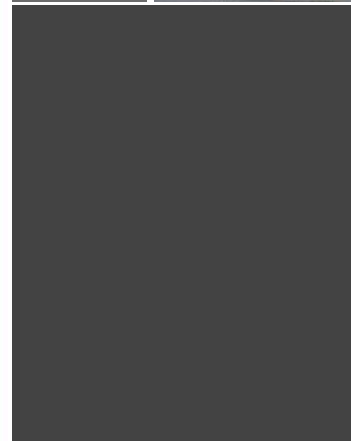
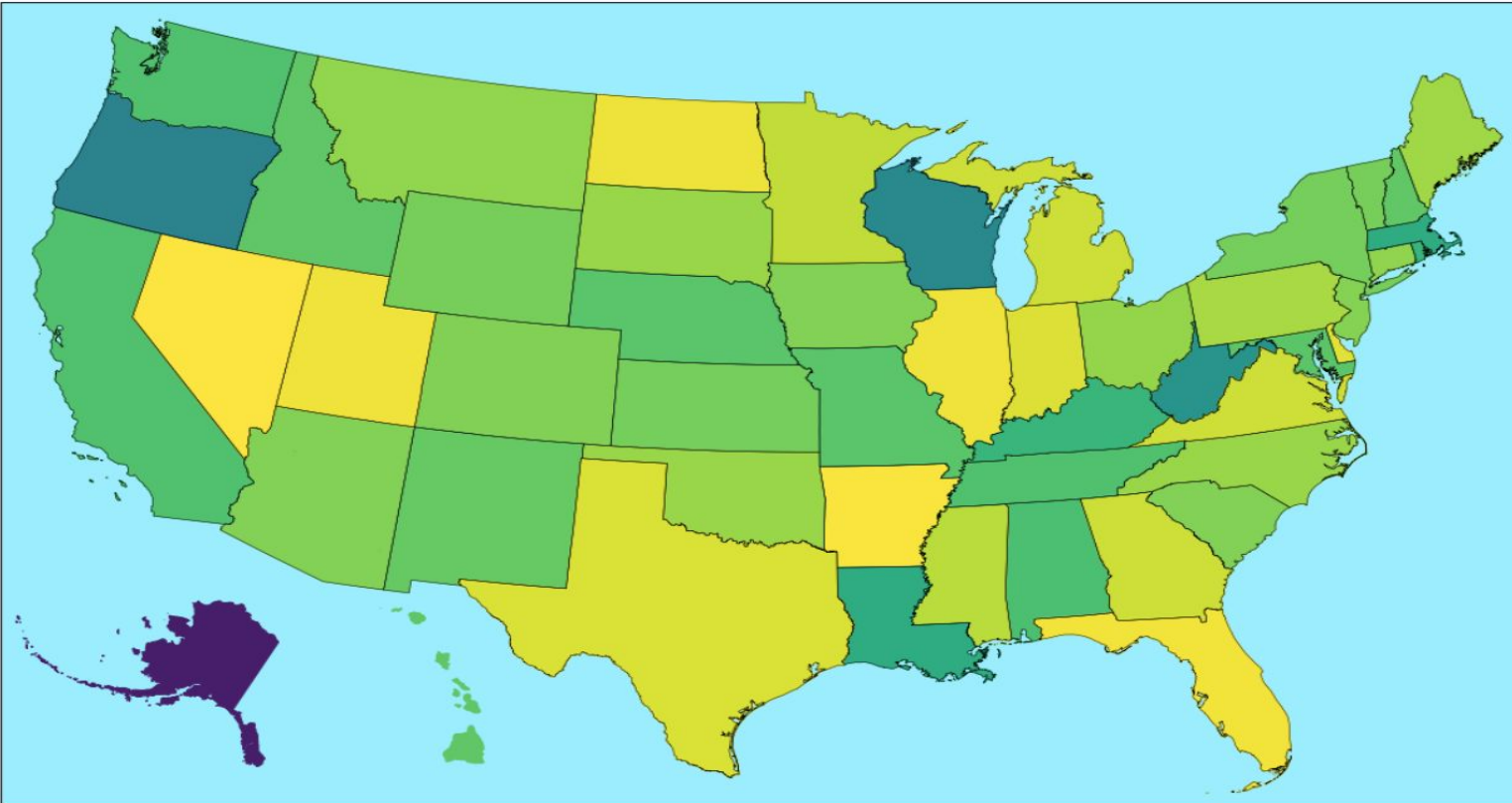
Initial Exploration and Visualization

TRI On-Site Releases 2013-2015
Darker Circles are Larger Releases



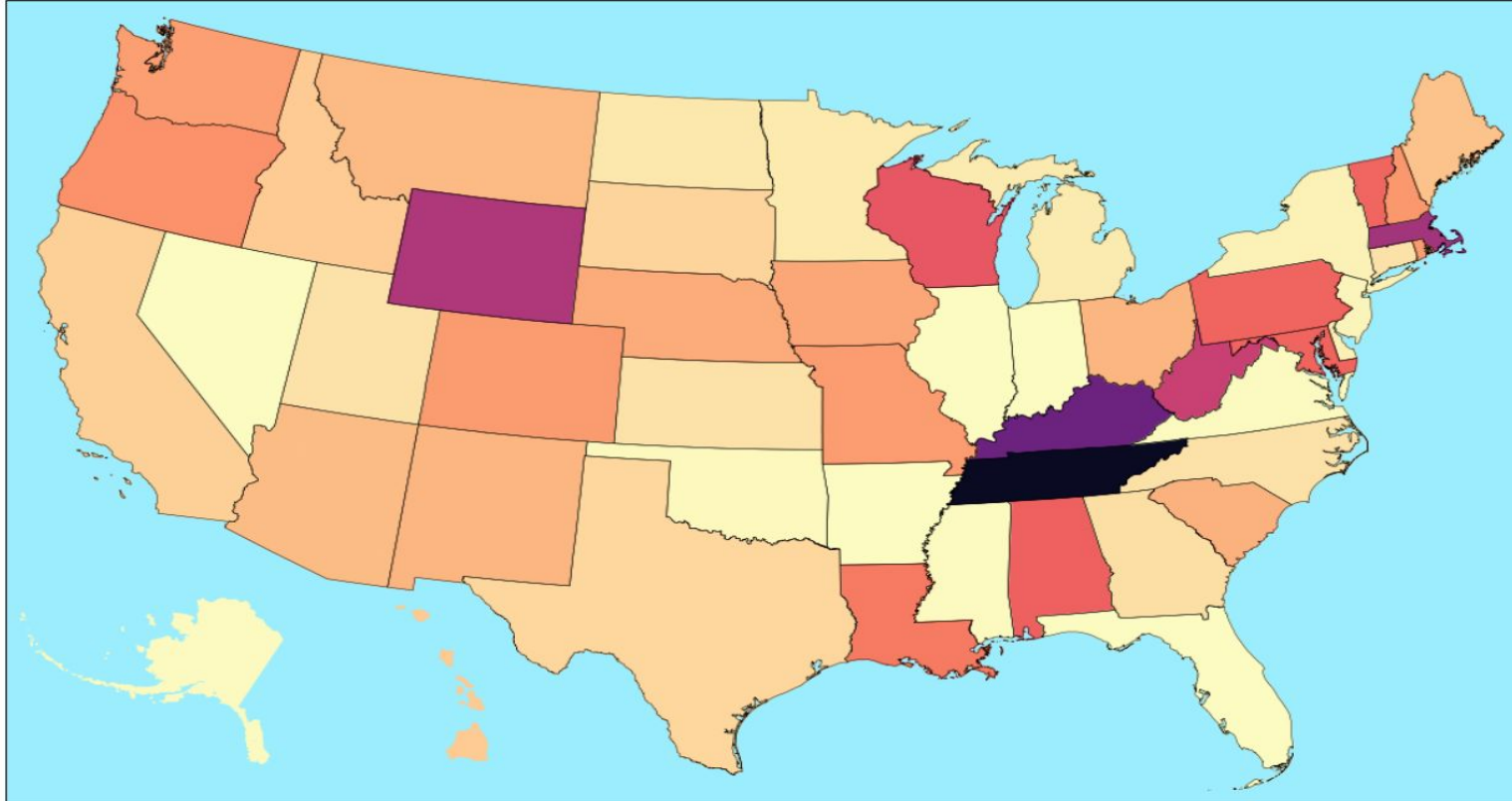
Initial Exploration and Visualization

Total Releases 2012-2016



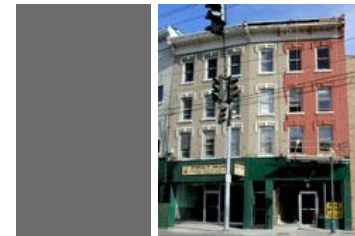
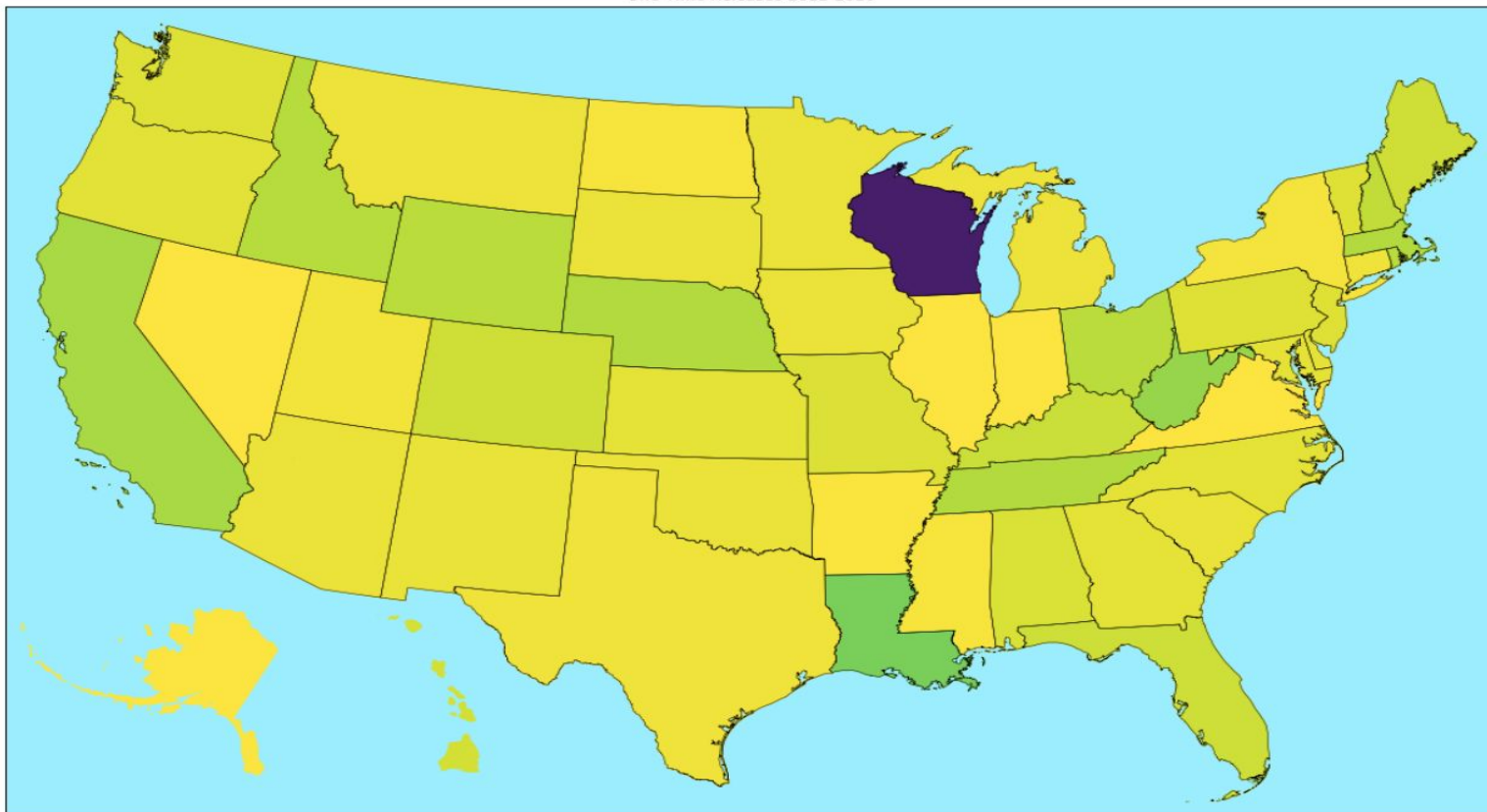
Initial Exploration and Visualization

On-Site Recycled Total 2012-2016



Initial Exploration and Visualization

One-Time Releases 2012-2016



Standardizing Project Data: What Do We Need?


- Over 100 columns on attributes, lots of them are irrelevant for our purposes
- This includes unwanted Government codes, industry compliance codes, etc
- We want location information, chemical information, and release and recycling information
- We identified 21 categories, common to US and State data sets, to use

| | |
|---------------------|-------------------------|
| YEAR | UNIT_OF_MEASURE |
| FACILITY_NAME | CARCINOGEN |
| FEDERAL_FACILITY | CAA_CHEMICAL |
| PARENT_COMPANY_NAME | TOTAL_RELEASES |
| INDUSTRY_SECTOR | ON_SITE_RELEASE_TOTAL |
| ZIP | OFF_SITE_RELEASE_TOTAL |
| STATE | ON_SITE_RECYCLED_TOTAL |
| CITY | OFF_SITE_RECYCLED_TOTAL |
| COUNTY | ONE_TIME_RELEASES |
| LATITUDE | |
| LONGITUDE | |
| CHEMICAL | |




Standardizing Project Data: Cleaning CSV Files

- Created Python script to process US and State CSV files
 - Trims unwanted attributes
 - Renames columns for standard access
 - Handles NaN values in numerical columns
 - Reduces CSV size by ~75%
- Run script on downloaded CSV files to pre-clean for use in ipynb



```
51 # Function to write the DataFrame to csv. For the purposes of this
52 # and the files will be in the format 'TRI_<YEAR>_US_CLEAN.csv'
53 # define out dir to be made -- change this if you want files writt
54 directory = './clean_TRI_Data'
55 def write_cleaned_tri_to_csv(tri_df, tri_type, directory):
56     #create target directory if it doesn't already exist
57     if not os.path.exists(directory):
58         os.makedirs(directory)
59
60     if not tri_type == 'US':
61         #check for state sub-folder, create if none
62         if not os.path.exists('%s/State_TRI/'%directory):
63             os.makedirs('%s/State_TRI/'%directory)
64         tri_df.to_csv('%s/State_TRI/%s'%(directory, 'TRI_%s_%s_CLEAN
65     else:
66         #check for US sub-folder, create if none
67         if not os.path.exists('%s/US_TRI/'%directory):
68             os.makedirs('%s/US_TRI/'%directory)
69         tri_df.to_csv('%s/US_TRI/%s'%(directory, 'TRI_%s_US_CLEAN.c
70
71 if len(sys.argv) == 2:
```



Next Steps

- **Apply graphing, mapping, and statistical functions to US and State TRI data sets, focusing on:**
 - **Nationwide relations and trends**
 - **Regional relations and trends**
 - **Facilities and parent companies' releases**
 - **Industry sector releases**



To Be Continued...



TOXIC CRUSADERS

CSC 495 Project - Fall 2017 - Report 2

Aaron Denton
Jacob Durham
Swetha Polisetty
Melvin Watlington

