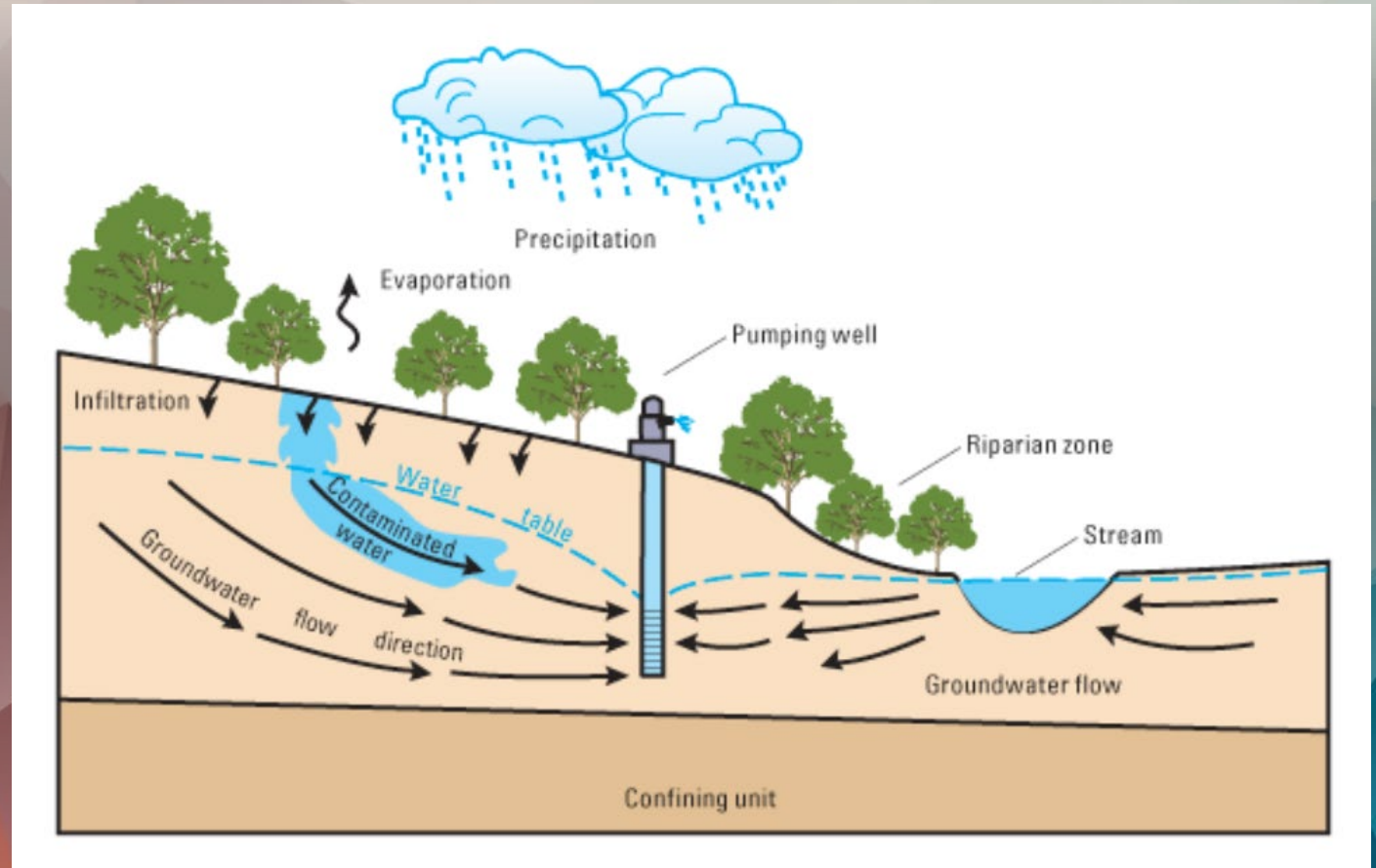# STATISTICAL ANALYSIS OF GROUNDWATER CONTAMINANTS AT HIDALGO COUNTY USING PYTHON
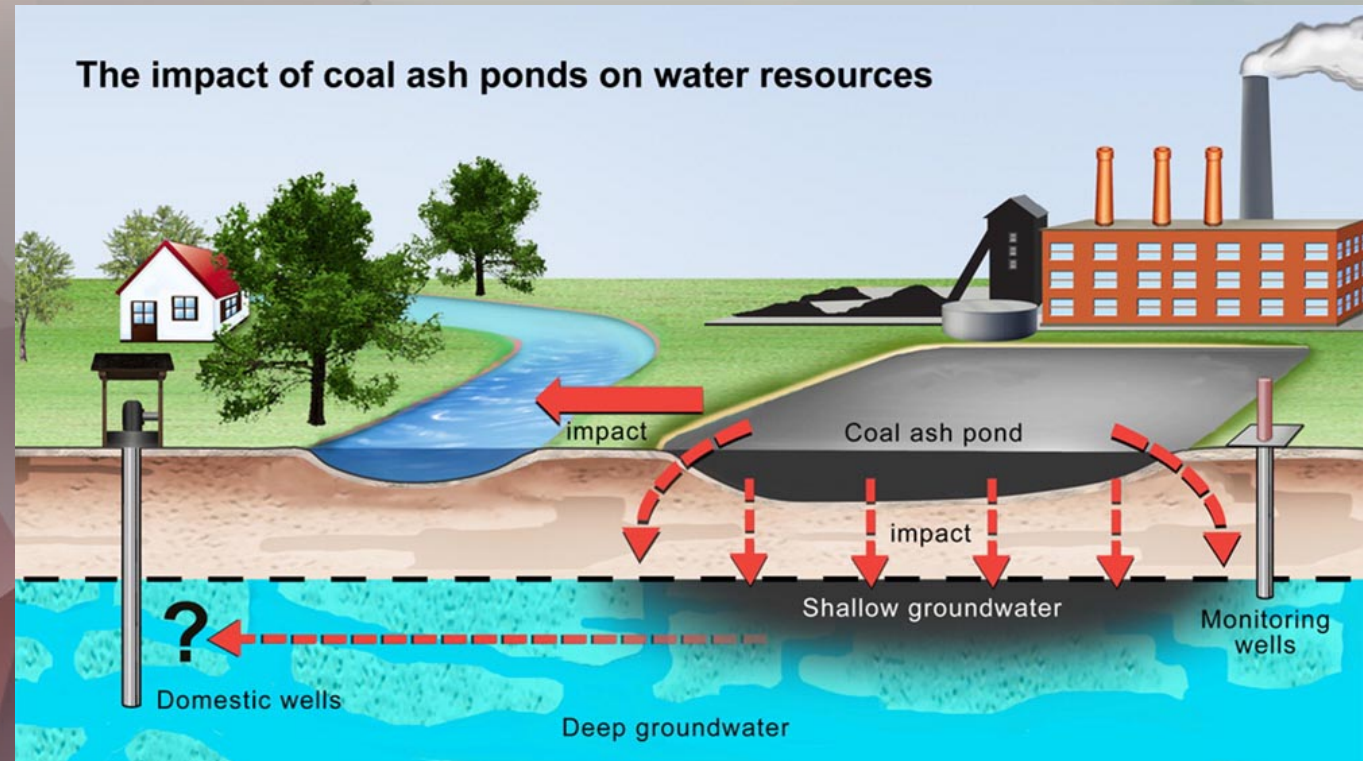
**Dwight Zedric Q. Capus**

## Background

- Texas major source of water (60%)

- 16.1 million acre-feet of water

# Background

-Texas major source of water (60%)

-Groundwater contamination occurs from Human activities

-Research focuses on anthropogenic contamination



The impact of coal ash ponds on water resources

impact

Coal ash pond

impact

Shallow groundwater

Monitoring wells

Domestic wells

Deep groundwater

# Joint Groundwater Monitoring and Contamination Report - 2019

## *Contributing State Agencies and Organizations:*

- Texas Commission on Environmental Quality Texas Water Development Board Railroad Commission of Texas

- Texas Department of State Health Services

- Texas Department of Agriculture

- Texas State Soil and Water Conservation

- Board Texas Alliance of Groundwater

- Districts Texas A&M AgriLife Research

- Bureau of Economic

- Geology of the University of Texas at Austin Texas Department of Licensing and Regulation

## Objective

- Site-Specific visualization analysis of the data from the Contamination report by TCEQ 2019

## Questions asked

- Distribution of contaminants in the area?

- Who are polluters?

- Where are the contaminants?

- What are the status of the sites?

## Methods

- Compile data at Excel worksheet

- Load necessary packages to jupyterlab (dataframe,pandas, cartopy, shapereader,feature)

- Upload data as dataframe

- Manipulate dataframe

- Visualized data and compared variables

- Created a map with plots of contaminants and Bar charts

- Time series Analysis = Not enough Data

Contamination report of Texas published by Texas Commission on Environmental Quality (TCEQ)

County
Division
File number
Location (Address)
Lat/Long
Contamination types
Date

| | A | B | C | D | E | F | G | H | I | J | K | L |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | COUNTY | Division | FILE NAM | File Numbe | LOCATION | LATITUDE | LONGITUL | CONTAMII | DATE | ENF-STATUS | ACT-STATUS | Category |
| 2 | HIDALGO | REM/VCIO | MCDONAL | 1087 | 3318 US B | 26.15282 | -97.9569 | PAH | 3/20/2018 | | 0 | Other Business Establishments |
| 3 | HIDALGO | REM/DCRI | PRIDE'S CL | DC0025 | 2204 WES | 26.24122 | -98.2377 | CHLORINA | 4/29/2005 | 5B | 0 | Other Business Establishments |
| 4 | HIDALGO | REM/PST | STRIPES 96 | 120410 | 721 N MCC | 26.2084 | -98.212 | DIESEL, GA | 12/11/2017 | 2 | 2A | Gas Station |
| 5 | HIDALGO | REM/PST | 7-ELEVEN | 120290 | 822 W US | 26.19157 | -98.1671 | GASOLINE | 8/22/2016 | 2 | 2A | Supermarket/Convenience stores |
| 6 | HIDALGO | REM/PST | 7-ELEVEN | 120511 | 6400 S 23F | 26.15007 | -98.2547 | GASOLINE | 3/7/2018 | 2 | 6 | Supermarket/Convenience stores |
| 7 | HIDALGO | REM/PST | ECONOMY | 116761 | 2015 S MC | 26.28689 | -98.198 | GASOLINE | 10/13/2005 | 1B | 1A | Other Business Establishments |
| 8 | HIDALGO | REM/PST | FFP 297 FORMER ECC | | 1525 N TE | 26.17638 | -97.9912 | GASOLINE | 7/9/1999 | 5B | 6 | Supermarket/Convenience stores |
| 9 | HIDALGO | REM/PST | FORMER G | 120247 | 700 W STA | 26.19726 | -98.1918 | GASOLINE | 4/24/2017 | 2 | 6 | Supermarket/Convenience stores |
| 10 | HIDALGO | REM/PST | HOP SHOF | 113110 | 1417 N CO | 26.22037 | -98.3253 | GASOLINE | 3/25/1998 | 2 | 6 | Supermarket/Convenience stores |
| 11 | HIDALGO | REM/PST | SAN JUAN | 120715 | 723 E UNIV | 26.30065 | -98.1549 | GASOLINE | 11/18/2017 | 2 | 6 | Other Business Establishments |
| 12 | HIDALGO | REM/PST | STRIPES 96 | 119665 | 602 W 2NL | 26.14978 | -97.9161 | GASOLINE | 12/13/2014 | 2 | 2A | Gas Station |
| 13 | HIDALGO | REM/PST | STRIPES 96 | 120448 | 1601 N 10 | 26.21804 | -98.2279 | GASOLINE | 10/20/2017 | 2 | 2A | Gas Station |
| 14 | HIDALGO | REM/PST | STRIPES 96 | 120419 | 621 E NOL | 26.23771 | -98.207 | GASOLINE | 11/3/2017 | 1B | 1A | Gas Station |
| 15 | HIDALGO | REM/PST | SUNRISE 1 | 103228 | FM 1015, \ | 26.22782 | -97.9599 | GASOLINE | 6/3/1992 | 2 | 2A | Gas Station |
| 16 | HIDALGO | REM/PST | TEXACO | 98334 | 1701 S 10 | 26.18876 | -98.2322 | GASOLINE | 3/22/1991 | 5B | 4 | Gas Station |
| 17 | HIDALGO | REM/PST | AZIZ CONV | 119730 | 3000 N W/ | 26.23553 | -98.2564 | GASOLINE | 6/29/2015 | 2 | 6 | Gas Station |
| 18 | HIDALGO | REM/PST | AZIZ CONV | 119727 | 2831 W US | 26.207 | -98.2532 | GASOLINE | 6/29/2015 | 2 | 6 | Gas Station |
| 19 | HIDALGO | REM/PST | JRS XPRES! | 120629 | 3704 N RA | 26.22962 | -98.1489 | GASOLINE | 10/16/2018 | 2 | 6 | Supermarket/Convenience stores |
| 20 | HIDALGO | REM/PST | SUPER OX) | 115176 | FM 1015, F | 26.09042 | -97.9586 | GASOLINE | 5/2/2001 | 2 | 2A | Supermarket/Convenience stores |
| 21 | HIDALGO | REM/VCP | ADOBE RE | 1657 | ABANDON | 26.29863 | -98.036 | METALS, C | 12/8/2003 | 0B | 5 | Oil wells/Refinaries |
| 22 | HIDALGO | REM/VCP | AGRILIAN( | 2417 | 501 E MOI | 26.337 | -98.1507 | OTHER | 7/15/2011 | 0B | 2A | Industrial/Agricutural factory |
| 23 | HIDALGO | REM/VCP | 200 WEST | 2304 | 200 W RAI | 26.16039 | -97.9916 | PESTICIDE | 3/26/2010 | 0B | 2A | Unknown |
| 24 | HIDALGO | REM/CA | WAL-MAR | T2055 | NORTHWE | 26.26775 | -98.2046 | SVOCS | 11/18/2005 | 0A | 0 | Supermarket/Convenience stores |

Methods applied

- Split() method splits a string into a list

- Explode() – transform the list into row
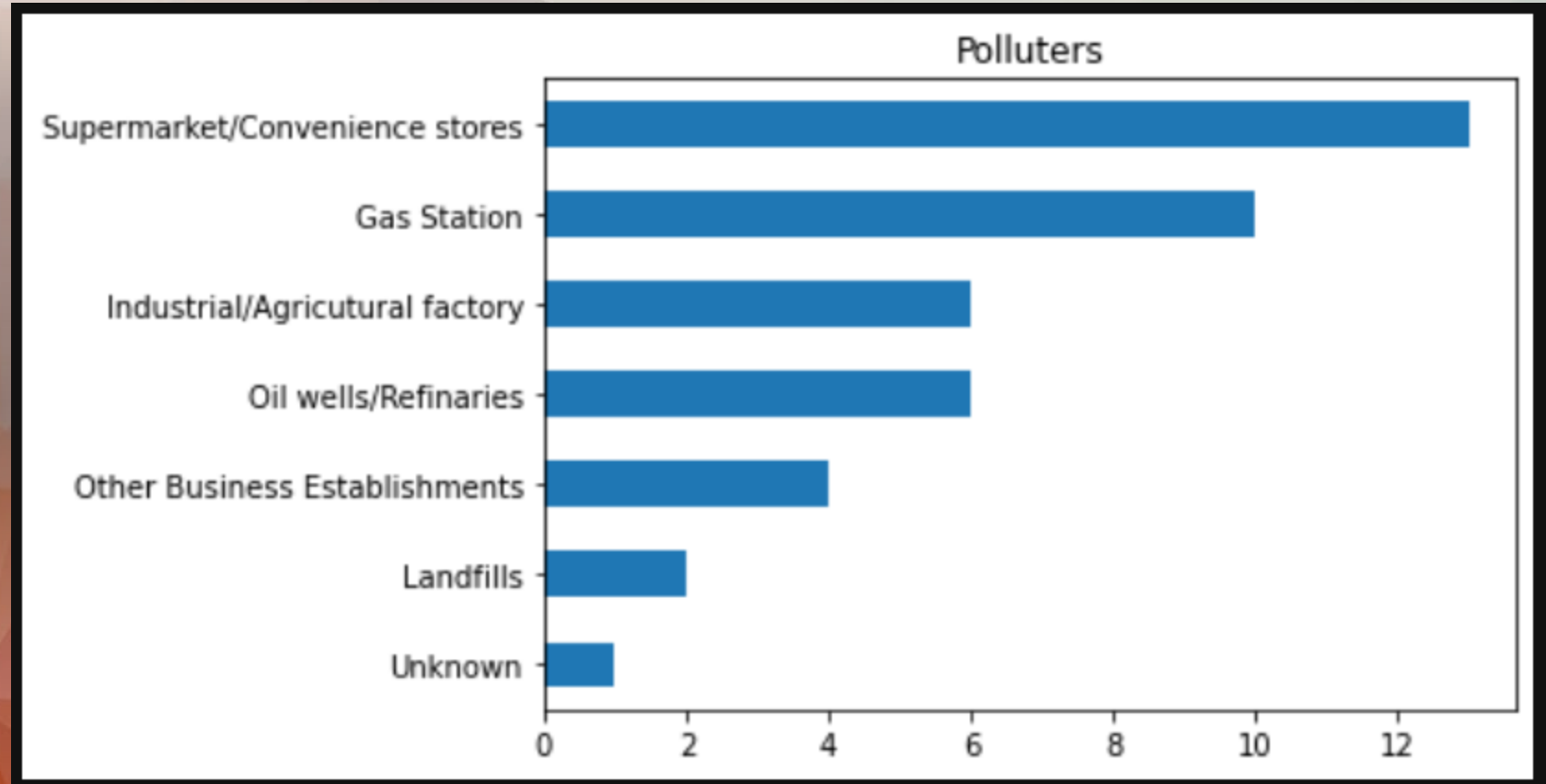
- Value_counts() – To count the values
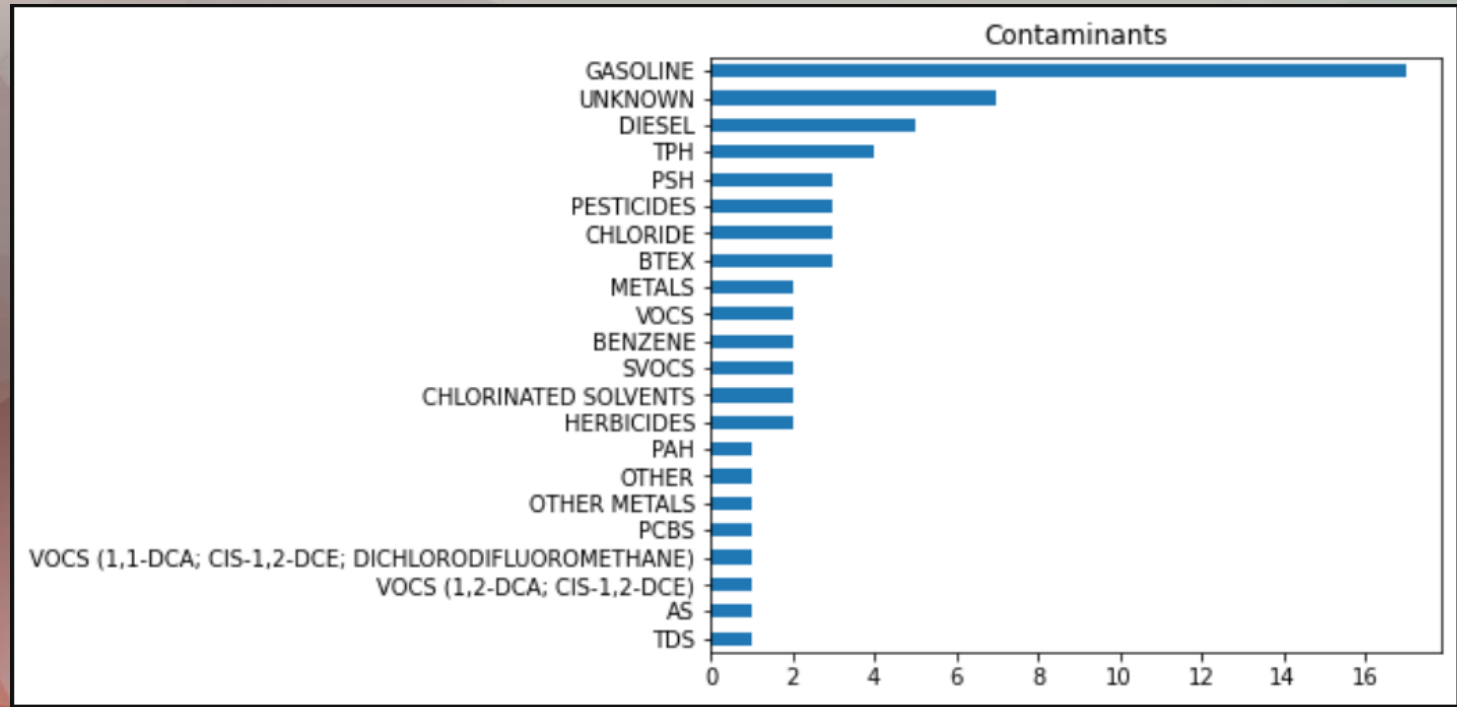

Contaminants

```
#split the  contaminants that has more than 1 types of contaminants in each rows to get the bettercoun
data['CONTAMINANTS'] = data['CONTAMINANTS'].str.split(pat=', ')
data_long = data.explode('CONTAMINANTS')
data_long['CONTAMINANTS'].value_counts()
```

```
data_long['CONTAMINANTS'].value_counts().plot(kind = 'barh',  figsize= (6,5) ).invert_yaxis()
plt.savefig('contaminants.jpeg')
plt.title('Contaminants')
```

Methods applied

Value_counts() – To count the values



Polluters
- Supermarket/Convenience stores
- Gas Station
- Industrial/Agricutural factory
- Oil wells/Refinaries
- Other Business Establishments
- Landfills
- Unknown

Methods

- Convert the dates to datetime

- Extracted the year into a new column using datetime

- Created a dummy column.


Number of Contaminants reported per year vs Closed Reports
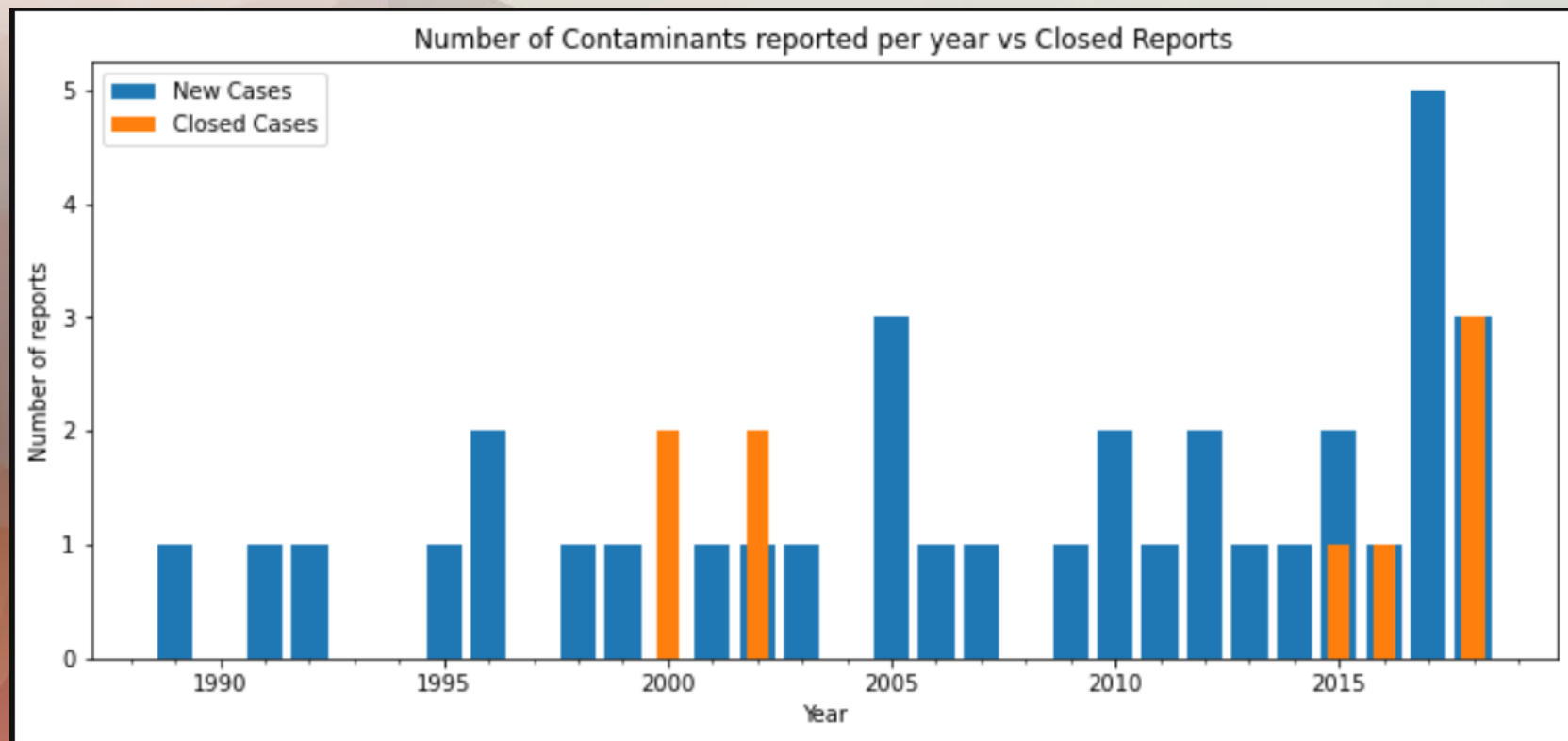
```python
data3=pd.read_csv('Data gathered1.csv')
data2=pd.read_csv('Data gathered1.csv', parse_dates=["DATE"])
#dont delete this
inactive2 = pd.read_csv('inactive.csv', parse_dates=["Year Deleted"])


data2['Date_date'] = pd.to_datetime(data2['DATE'])
data2['YEAR'] =data2['Date_date'].dt.year
data2['COUNTER'] = 1



data3 = data2.groupby('YEAR').count()
data3 = data3.reset_index()
print(data3[['YEAR','COUNTER']])
```

Method

- Used cartopy.io.shapereader package

- Download shapefiles of Counties and cities from TxDot
- Created my own matplotlib features
- Lat/long to my x and y



Hidalgo County
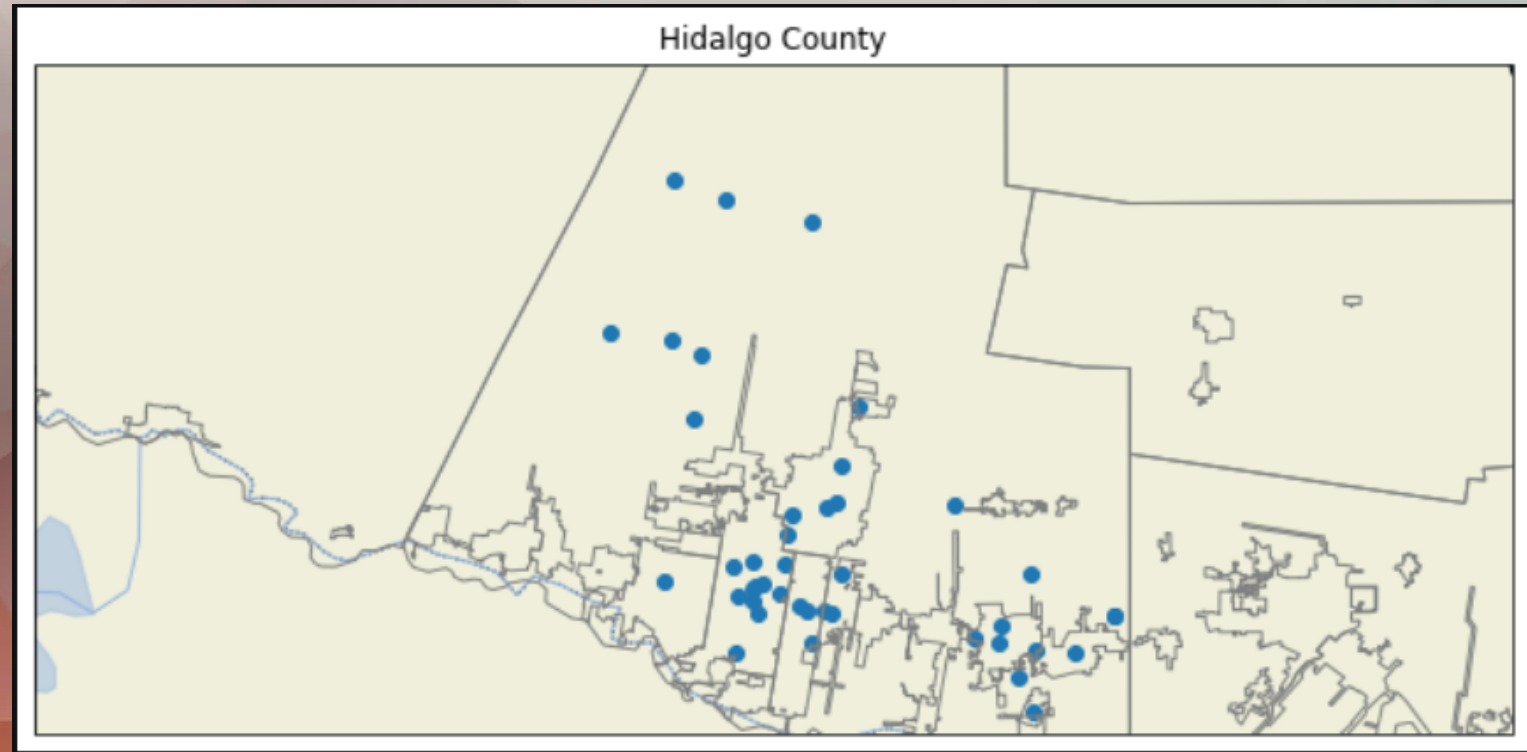
```python
import cartopy.io.shapereader as shpreader

reader = shpreader.Reader('cb_2018_us_county_5m')

counties = list(reader.geometries())

COUNTIES = cfeature.ShapelyFeature(counties, ccrs.PlateCarree())
reader2 = shpreader.Reader('City')

city = list(reader2.geometries())

Cities = cfeature.ShapelyFeature(city, ccrs.PlateCarree())
```

## Method

- OSM Package – OpenStreetMap

- Created a library of status and used apply(lambda [x:] Status[x]

- Repeat method for color

- Subset each status column and plot them as a scatter plot

```
color = { 'No Activity': 'b',
          'Contamination confirmed':'c',
          'Investigation':'r',
          'Corrective Action Planning':'m',
          'Implement Action':'y',
          'Monitor Action':'g',
          'Action Completed':'k'
          }

data_all['COLOR'] = data_all['STATUS'].apply(lambda x: color[x])
```

### Hidalgo County



Legend:
- No Activity
- Investigation
- Action Completed
- Contamination confirmed
- Implement Action
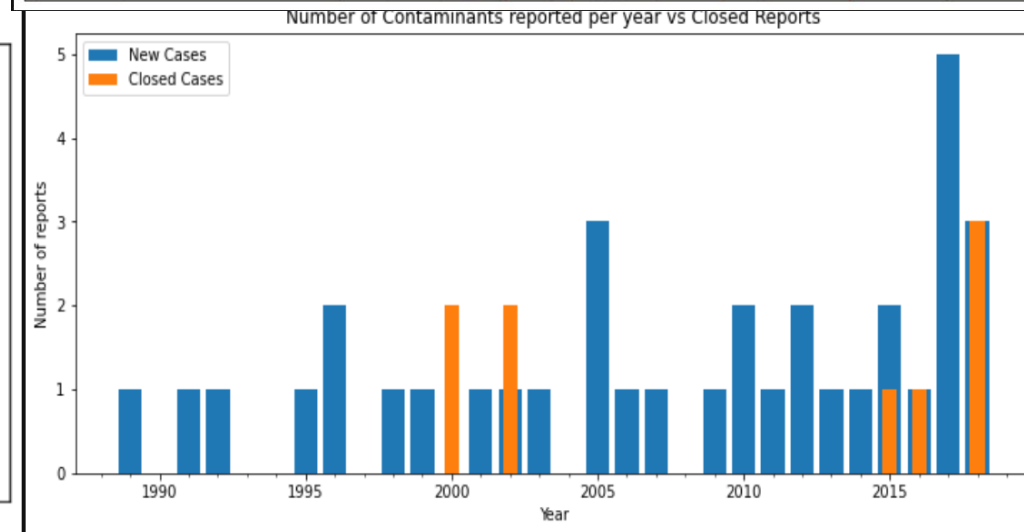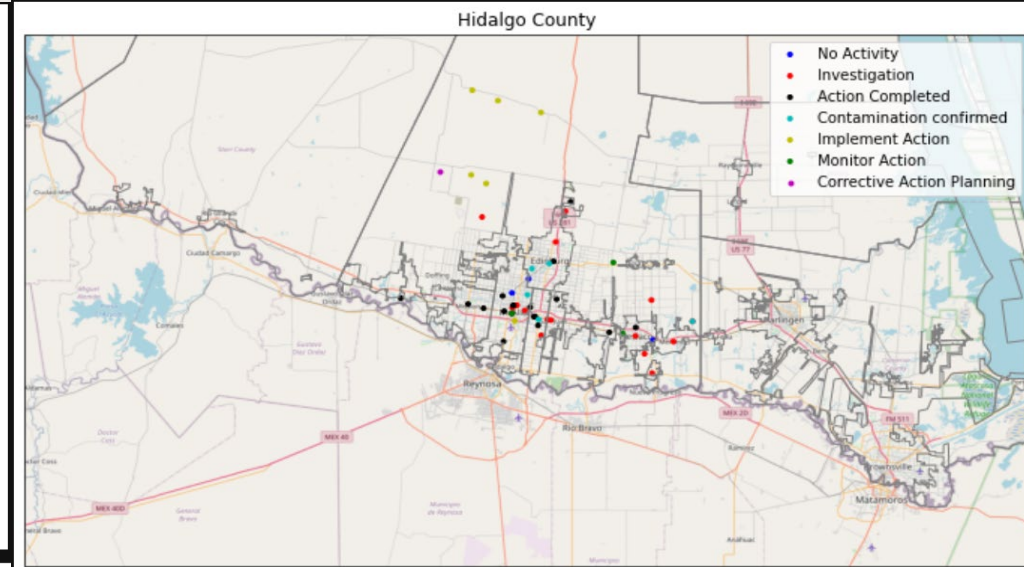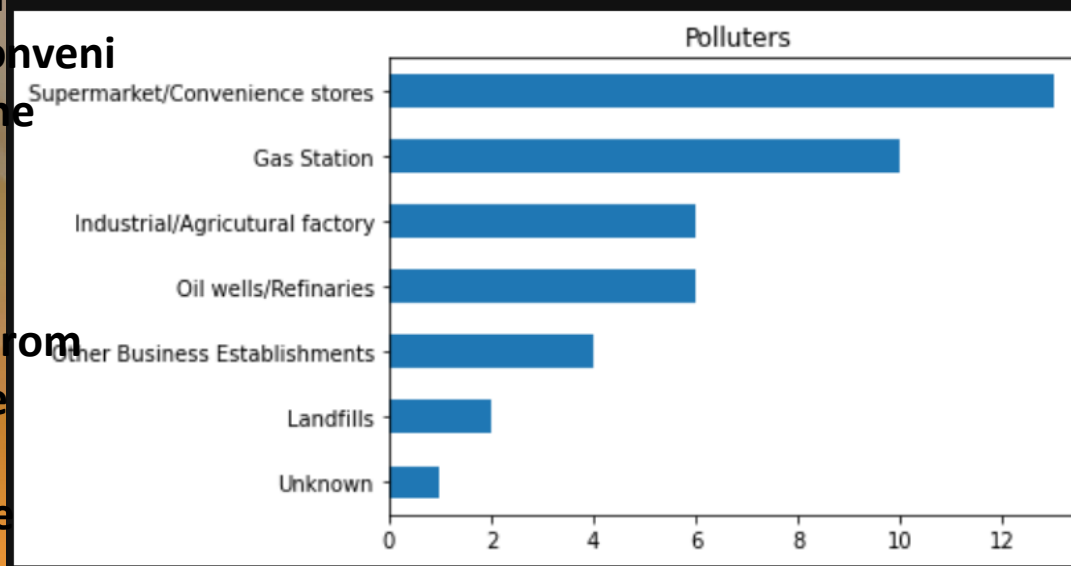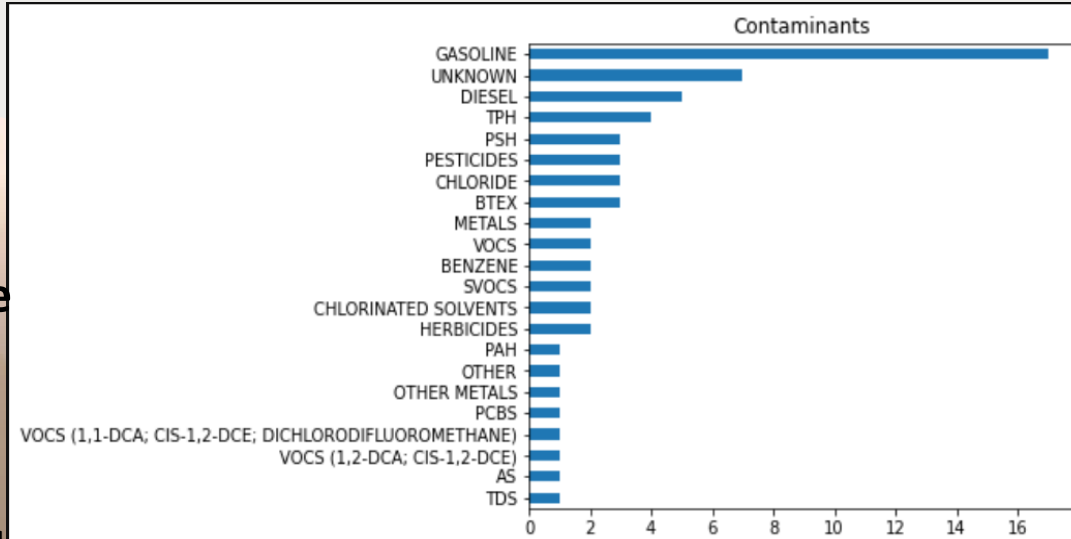- Monitor Action
- Corrective Action Planning
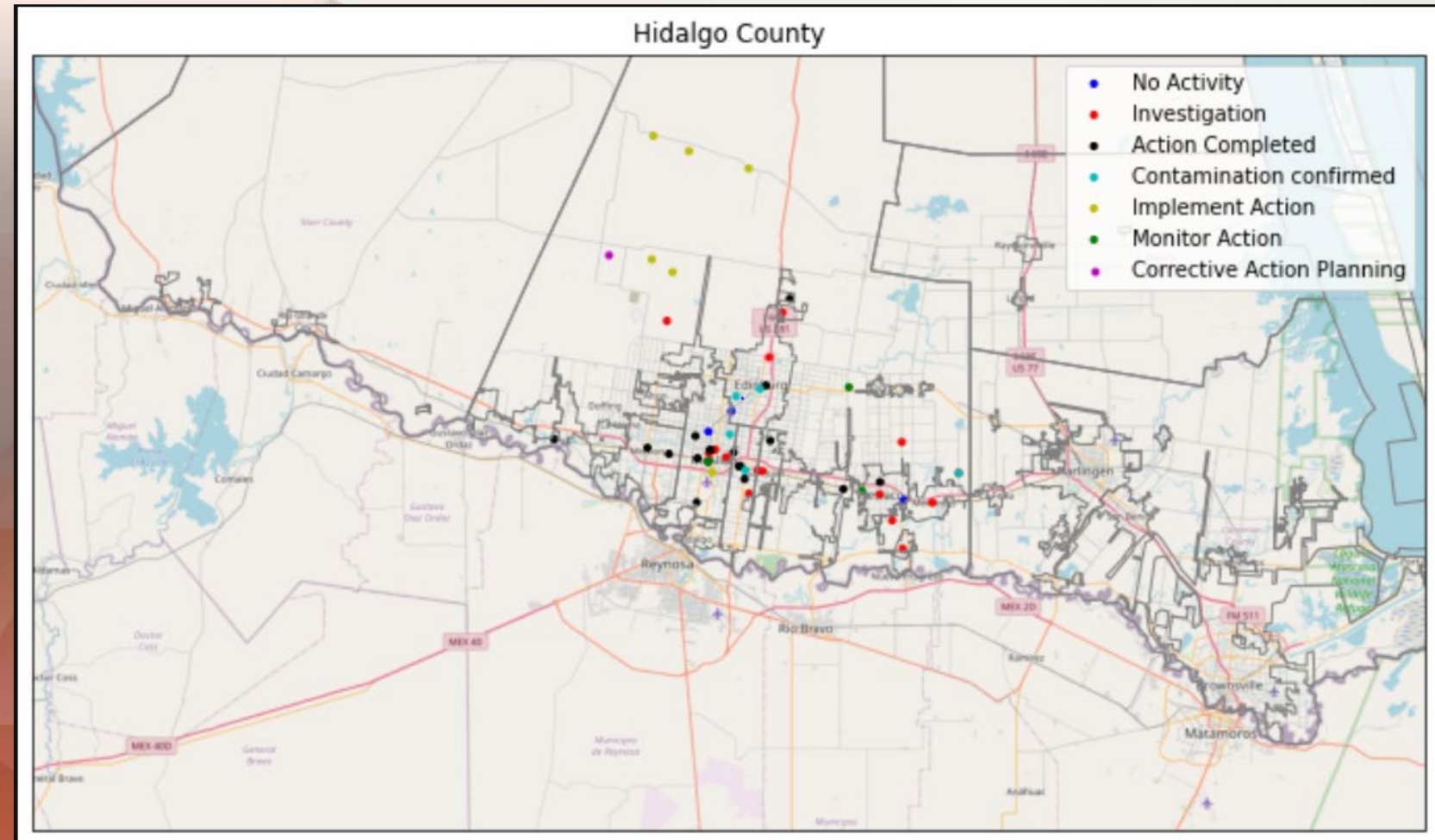
**Status Definition (code #)**

- No Activity (0) – No actions
- Contamination confirmed(1) – Data is being verified
- Investigation (2) – Under study to determine the extent, composition, or other properties
- Corrective Action Planning (3) – Remediation plan is being developed
- Implement Action (4)- Implementation of the remediation
- Monitor Action (5) – Effectiveness of the remedy
- Action completed (6)- Remedy is complete



Hidalgo County

Legend:
- No Activity
- Investigation
- Action Completed
- Contamination confirmed
- Implement Action
- Monitor Action
- Corrective Action Planning
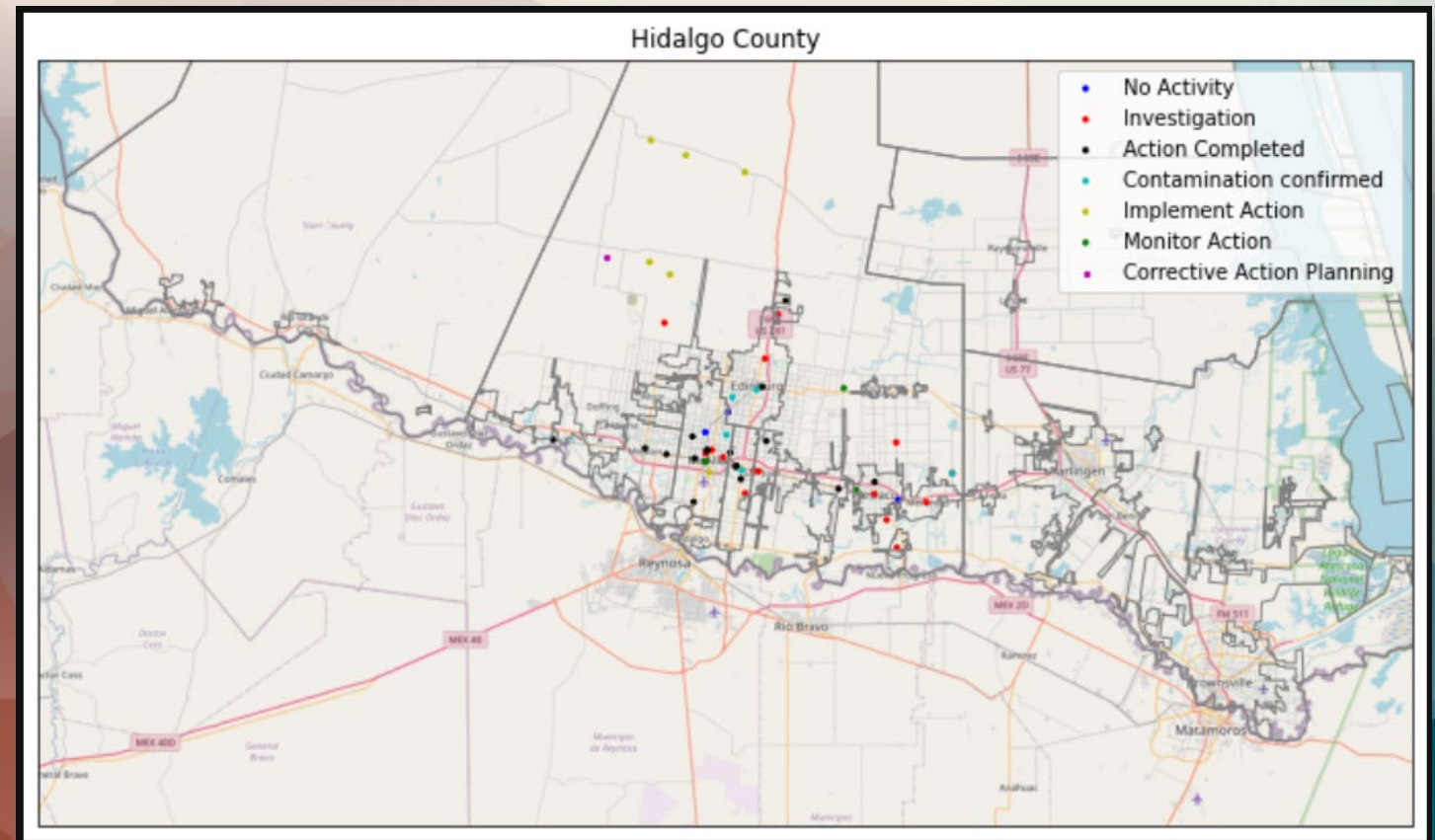
# Results

- **Gasoline is the highest contaminant reported**
- **Gas stations and Supermarket/Convenience store are the highest polluter**

- **Out of 59 Contamination from 1989 only 18 are under "Action completed. 9 are closed.**
- **Action Completed and Investigation.**



Contaminants



Hidalgo County



Polluters



Number of Contaminants reported per year vs Closed Reports

**References**

Chowdhury, A., & Mace, R. E. (2007). *Groundwater Resource Evaluation and Availability Model of the Gulf Coast Aquifer in the Lower Rio Grande Valley of Texas* (No. 368). Texas Water Development Board. https://www.twdb.texas.gov/publications/reports/numbered_reports/doc/R368/R368_GulfCoastGAM.pdf

Gangwar, S., Chauhan, M. S., & Singh, D. (2021). A Supervising Grid Model for Identification of Groundwater Pollute. In M. S. Chauhan & C. S. P. Ojha (Eds.), *The Ganga River Basin: A Hydrometeorological Approach* (pp. 53–66). Springer International Publishing. https://doi.org/10.1007/978-3-030-60869-9_4

*Python Tutorial*. (n.d.). Retrieved May 3, 2021, from https://www.w3schools.com/python/

Texas Groundwater Protection Committee. (2020). *Joint Groundwater Monitoring and Contamination Report—2019 Prepared by the Texas Groundwater Protection Com*. Texas Commission on Environmental Quality.