Statistical analysis of groundwater contaminants at Hidalgo county using Python

Dwight Zedric Q. Capus
Ocean, Coastal, and Earth Sciences MS program
University of Texas Rio Grande Valley
Edinburg, TX

May 5, 2021

Abstract

Many living things rely on groundwater as part of their sustenance and therefore, this makes it the most important freshwater asset in the United States. Groundwater has also been a vital factor for water security and expanding nourishment for the population. This research will compile, map and analyze the groundwater contaminants in Hidalgo County, Texas. By using the packages of multiple packages of Python, this study analyze the occurrences of contaminants at Hidalgo county. Bar graphs and a map was constructed to understand the distribution of the contaminants, identify perpetrator, location, and the status of reported sites. This study is significant because it could provide locals with knowledge and awareness about the problem and could also be utilized as additional data for local and regional databases.

Introduction

Groundwater is the most important freshwater asset. Many inhabitants of this nation rely upon this asset for their water needs. Groundwater has been a significant part of expanding nourishment generation and accomplishing water security. Additionally, groundwater has an exceedingly reliable water supply business, households, horticulture, and mechanical needs(Gangwar et al., 2021). The groundwater of the Gulf Coast Aquifer in the Lower Rio Grande Valley of South Texas has limited water supply for domestic, municipal, and agricultural usage. Additionally, in Lower Rio Grande Valley, groundwater in most areas often does not meet drinking water or irrigation standard. Groundwater in the valley is mostly slightly saline, about 1000 to 3000 milligrams per liter of Total Dissolved Solids, with the local occurrence of high boron, chloride, sodium, and nitrate(Chowdhury & Mace, 2007).

Groundwater contamination is generally defined as any harmful alteration of the naturally occurring quality of groundwater. However, in this study, the definitions is limited to the contamination that is associated with activities that fall under the responsibility of the contributing agencies and affecting usable-quality groundwater and does not include naturally occurring groundwater conditions such as high mineralization that may exceed recognized standards for public usage (Texas Groundwater Protection Committee, 2020)
Figure 1 shows an example of groundwater contamination (mainly hydrocarbon contaminants), locally known as “Mcallen Plume”. This study aims to analyze the occurrences and prevalence of groundwater contamination at Hidalgo County, Texas. This paper will use Python’s Matplotlib package to generate the plots and Cartopy for maps. Furthermore, the scope of this study includes closed and open cases reported at *Joint Groundwater Monitoring and Contamination Report* from 2019.

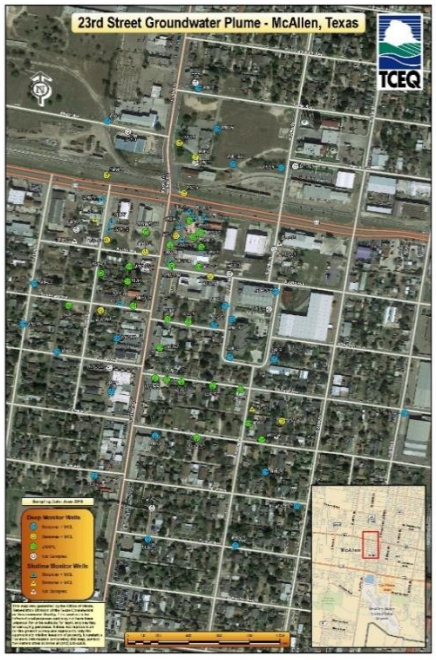

Figure 1. McAllen Plume (TCEQ, n.d.)

Research questions

This paper aims to answer the following questions:

1. Distribution of contaminants in the area?
2. Who are polluters?
3. Where are the contaminants?
4. What are the status of the sites?

Methods

To answer the questions above, the study utilize the Excel worksheet to organize the data and use Python Jupyterlab to plot and analyze the data. The data will be sourced from *Joint Groundwater Monitoring and Contamination Report* 2019, After organizing the data, data was uploaded to the kernel as a DataFrame. Python's Matplotlib package will be loaded into the kernel to create the plots. Different types of contaminants were plotted to understand their distribution using a time-series bar plots. Cases of contaminant will be plotted on the map of the area using python's cartopy package.

COUNTY	Division	FILE NAME	File Number	LOCATION	LATITUDE	LONGITUI	CONTAMINAN	DATE	ENF-STATU	ACT-STATU	Category
HIDALGO	REM/VCIO	MCDONALDS 42	1087	3318 US BUSIN	26.152818	-97.9569	PAH	3/20/2018		0	Other Business Establishments
HIDALGO	REM/DCRI	PRIDE'S CLEANE	DC0025	2204 WEST NO	26.2412221	-98.2377	CHLORINATED	4/29/2005	5B	0	Other Business Establishments
HIDALGO	REM/PST	STRIPES 9625	120410	721 N MCCOLL	26.2084017	-98.212	DIESEL, GASOL	12/11/2017	2	2A	Gas Station
HIDALGO	REM/PST	7-ELEVEN STOR	120290	822 W US HIGH	26.1915684	-98.1671	GASOLINE	8/22/2016	2	2A	Supermarket/Convenience stores
HIDALGO	REM/PST	7-ELEVEN STOR	120511	6400 S 23RD ST	26.1500698	-98.2547	GASOLINE	3/7/2018	2	6	Supermarket/Convenience stores
HIDALGO	REM/PST	ECONOMY DRIV	116761	2015 S MCCOLL	26.2868915	-98.198	GASOLINE	10/13/2005	1B	1A	Other Business Establishments
HIDALGO	REM/PST	FFP 297 FORMER	ECONOMY DR	1525 N TEXAS B	26.17638	-97.9912	GASOLINE	7/9/1999	5B	6	Supermarket/Convenience stores
HIDALGO	REM/PST	FORMER GINOS	120247	700 W STATE A	26.1972605	-98.1918	GASOLINE	4/24/2017	2	6	Supermarket/Convenience stores
HIDALGO	REM/PST	HOP SHOP 1	113110	1417 N CONWA	26.2203685	-98.3253	GASOLINE	3/25/1998	2	6	Supermarket/Convenience stores
HIDALGO	REM/PST	SAN JUANITA TF	120715	723 E UNIVERS	26.3006455	-98.1549	GASOLINE	11/18/2017	2	6	Other Business Establishments
HIDALGO	REM/PST	STRIPES 9634	119665	602 W 2ND ST,	26.1497804	-97.9161	GASOLINE	12/13/2014	2	2A	Gas Station
HIDALGO	REM/PST	STRIPES 9646	120448	1601 N 10TH ST	26.2180371	-98.2279	GASOLINE	10/20/2017	2	2A	Gas Station
HIDALGO	REM/PST	STRIPES 9673	120419	621 E NOLANA	26.2377055	-98.207	GASOLINE	11/3/2017	1B	1A	Gas Station
HIDALGO	REM/PST	SUNRISE 1	103228	FM 1015, WESL	26.22782	-97.9599	GASOLINE	6/3/1992	2	2A	Gas Station
HIDALGO	REM/PST	TEXACO	98334	1701 S 10TH ST	26.1887625	-98.2322	GASOLINE	3/22/1991	5B	4	Gas Station
HIDALGO	REM/PST	AZIZ CONVENIE	119730	3000 N WARE R	26.2355336	-98.2564	GASOLINE, DIE	6/29/2015	2	6	Gas Station
HIDALGO	REM/PST	AZIZ CONVENIE	119727	2831 W US HIG	26.2070009	-98.2532	GASOLINE, DIE	6/29/2015	2	6	Gas Station
HIDALGO	REM/PST	JRS XPRESS	120629	3704 N RAUL LO	26.2296214	-98.1489	GASOLINE, DIE	10/16/2018	2	6	Supermarket/Convenience stores
HIDALGO	REM/PST	SUPER OXXO ST	115176	FM 1015, PROG	26.09042	-97.9586	GASOLINE, DIE	5/2/2001	2	2A	Supermarket/Convenience stores
HIDALGO	REM/VCP	ADOBE REFINER	1657	ABANDONED R	26.2986281	-98.036	METALS, CHLO	12/8/2003	0B	5	Oil wells/Refinaries
HIDALGO	REM/VCP	AGRILIANCE ED	2417	501 E MONTE C	26.3370005	-98.1507	OTHER	7/15/2011	0B	2A	Industrial/Agricutural factory
HIDALGO	REM/VCP	200 WEST RAILF	2304	200 W RAILRD S	26.160388	-97.9916	PESTICIDES, HE	3/26/2010	0B	2A	Unknown
HIDALGO	REM/CA	WAL-MART STO	T2055	NORTHWEST C	26.2677524	-98.2046	SVOCS	11/18/2005	0A	0	Supermarket/Convenience stores
HIDALGO	REM/PST	23RD STREET GV	117954	23RD ST, MCAL	26.2135	-98.2371	UNKNOWN	7/18/1989	5B	2A	Gas Station
HIDALGO	REM/PST	BEST PIC CONVE	119278	1701 E MILE 5 N	26.12675	-97.9738	UNKNOWN	10/14/2013	2	2A	Supermarket/Convenience stores
HIDALGO	REM/PST	ECONOMY FOO	119012	601 S CLOSNER	26.296313	-98.1635	UNKNOWN	8/14/2012	1B	1A	Gas Station
HIDALGO	REM/PST	HERNANDEZ FO	117177	502 S SAN ANT(	26.19015	-98.1602	UNKNOWN	9/29/2006	5B	2A	Supermarket/Convenience stores
HIDALGO	REM/PST	HOP N SHOP 5 (	118036	400 S CAGE BLV	26.1913289	-98.1849	UNKNOWN	3/3/2009	1B	1A	Supermarket/Convenience stores
HIDALGO	REM/PST	QUICK PIC 4	111723	2002 W HWY 83	26.1877779	-97.8782	UNKNOWN	10/8/1996	2	2A	Supermarket/Convenience stores
HIDALGO	REM/PST	QUICK PIC COU	115600	11110 MILE 2 E	26.1877779	-97.8782	UNKNOWN	10/8/2002	1B	1A	Supermarket/Convenience stores
HIDALGO	REM/VCP	ANGLO IRON AN	196	2 S 21ST ST, MC	26.20278	-98.2378	VOCS	3/7/1996	0B	5	Industrial/Agricutural factory
HIDALGO	WPD/MSW	C & T LANDFILL	MSW00151A	S SIDE OF FM 10	26.58	-98.18	VOCS (1,1-DCA	2/13/1995	2B	4	Landfills
HIDALGO	WPD/MSW	CITY OF EDINBU	MSW00956C	900 E ENCINITO	26.395	-98.1316	VOCS (1,2-DCA	7/25/2012	2B	2B	Landfills
HIDALGO	REM/VCP	CROP PRODUCT	2313	WESLACO CITY	26.1655191	-98.0176	VOCS, PESTICIE	5/6/2010	0B	5	Industrial/Agricutural factory
HIDALGO	O&G	EAST MCCOOK	OCP#5026	SHELL	26.4689	-98.381	TPH, BTEX		0	3	Industrial/Agricutural factory
HIDALGO	O&G	FORMER SHELL	OCP#5159	SHELL EXPLORA	26.2177	-98.2351	CHLORIDE, TDS, AS		0	6	Oil wells/Refinaries
HIDALGO	O&G	GANAWAY FACI	OCP#5008	MO-VAC SERVI(	26.3839	-98.2973	CHLORIDE		2	2A	Industrial/Agricutural factory
HIDALGO	O&G	HAMMAN COM	OCP#5045	HILCORP ENER(	26.4468	-98.2892	TPH, BTEX, PSH		0	4	Oil wells/Refinaries
HIDALGO	O&G	MCALLEN RANC	OCP#2613	HILCORP	26.6022	-98.2658	BENZENE, OTHER METALS		0	4	Oil wells/Refinaries
HIDALGO	O&G	MONTE CHRIST	OCP#4286	HILCORP	26.463	-98.319	TPH, BENZENE, PSH, CHLORI		0	4	Oil wells/Refinaries
HIDALGO	O&G	SHELL	OCP#1837	KINDERMORGA	26.6217	-98.3169	TPH, BTEX, PSH		0	4	Oil wells/Refinaries
HIDALGO	REM/CA	PHARR PLANTA1	52175	431 PLANTATIO	26.161431	-98.1797	PESTICIDES, HE	9/5/2007	3B	2A	Industrial/Agricutural factory
HIDALGO	O&G	FORMER SHELL	OCP#5159	SHELL EXPLORA	26.217723	-98.2351	CHLORIDE, TDS	1/1/2018	0	6	Oil wells/Refinaries
HIDALGO	O&G	PENITAS COMP	OCP#4931	ENTERPRISE PR	26.2316	-98.4592	BENZENE	1/1/2016	0	6	Industrial/Agricutural factory
HIDALGO	REM	GUARANTY FED	T0514	1802 N JACKSO	26.214427	-98.2015	METALS	1/1/2002	1B	6	Other Business Establishments
HIDALGO	REM	M-I SWACO FAC	T3280	12303 VICKING	26.4125773	-98.1209	TPH	1/1/2018	1A	6	Oil wells/Refinaries
HIDALGO	REM	SCHLUMBERGEI	33568	1005 N GLASSC	26.212549	-98.2944	TPH, VOCS, ME	1/1/2018	3B	6	Oil wells/Refinaries
HIDALGO	REM	FORMER AZIZ ST	2528	1301 SOUTH CA	26.180259	-98.1859	NITRATE, AMM	1/1/2015	0B	6	Other Business Establishments
HIDALGO	REM	ANGLO IRON AN	196	2 SOUTH 21ST S	26.20278	-98.2378	TPH, TOTAL ME	1/1/2000	0	6	Industrial/Agricutural factory
HIDALGO	REM	CLARK KNAPP M	1107	801 WEST BUSI	26.196382	-98.1939	TPH, METALS	1/1/2002	0	6	Other Business Establishments
HIDALGO	REM	KNAPP- SHERRII	360	307 SOUTH 12T	26.16722	-98.0439	BETX, TPH, ME	1/1/2000	0	6	Other Business Establishments

Table 1. Gathered data from the report. Exported in Excel as csv.

Joint Groundwater Monitoring and Contamination Report

TCEQ published a report on 2020, including extensive data about the groundwater contamination report from the last two decades up to the end of 2019. There are a total of 10 state agencies contributed to the report. The report contains a table of contamination reported from monitoring wells (Figure 2). Hidalgo reports were then exported to Excel. A category of different types of polluters or contaminant sources were manually added. Figure 3 shows the definition of ACT status from the column.

Code (ASC)	Activity	Description
0	NO ACTIVITY (TCEQ ONLY)	No actions have been conducted at the incident site. The code is also given to cases that did not have any activity in the reported calendar year, mainly used by the Voluntary Cleanup Program of TCEQ.
1	CONTAMINATION CONFIRMED	Contamination is being verified, such as with resampling or data quality validation, but has not been fully investigated. Based on the confirmation findings, additional actions such as further investigation may or may not be required: 1A – Action required, or 1B – No action required.
2	INVESTIGATION	The incident is being studied to determine the extent, composition, or other properties and circumstances of the contamination. Based on the findings additional action may or may not be required: 2A – Action required, or 2B – No action required.
3	CORRECTIVE ACTION PLANNING	A remedy (corrective action plan) for the contamination is being developed based on the investigation findings. Examples may include remedies such as a plan to remove the source of contamination, remediate impacted groundwater, disinfect impacted wells, or replace wells.
4	IMPLEMENT ACTION	The planned remedy (corrective action plan) is being carried out and actions to address the contamination are being conducted.
5	MONITOR ACTION	The effectiveness of the remedy is being monitored. This may be a long-term or short-term action and may be performed during and after implementation.
6	ACTION COMPLETED	The remedy is considered complete when the desired result has been achieved. Options in which contamination is no longer present include: 6A - remediation or 6B - source removal. Options in which contamination may still be present but action is complete include: 6C - determination that no further action is needed, 6D - institutional controls (agency action is final; however, contamination still exists under institutional controls such as deed records noting contamination, use and exposure restrictions, or required maintenance of engineering controls), or 6E - engineering controls (agency action is final; however, contamination still exists under required engineered controls). *For RRC, "6C" is the code used when no further action is required/institutional or engineering controls are in effect.*

Table 2. ASC Status

Types of Contaminants

Python packages including pandas, pandas.DataFrame, numpy, matplotlib.pyplot, datetime, matplotlib.dates , cartopy.crs , and cartopy.feature were first coded. The CSV file was then loaded as Dataframe. To create the horizontal bar graph that contains types of contaminants (Figure 2). Split() method was used to separate the strings(contaminants) in the "contaminant" column as a list and used explode() function to transform list into rows. The reason for these methods is because there more than one contaminant per report so one needs to split the strings for all individual contaminants to be counted. Value_counts() were used for total value count and plotted as a series. W3schools (2021) contains all the general information about different methods and functions in Python and other programming languages and their structure and definition (Python Tutorial, n.d.).

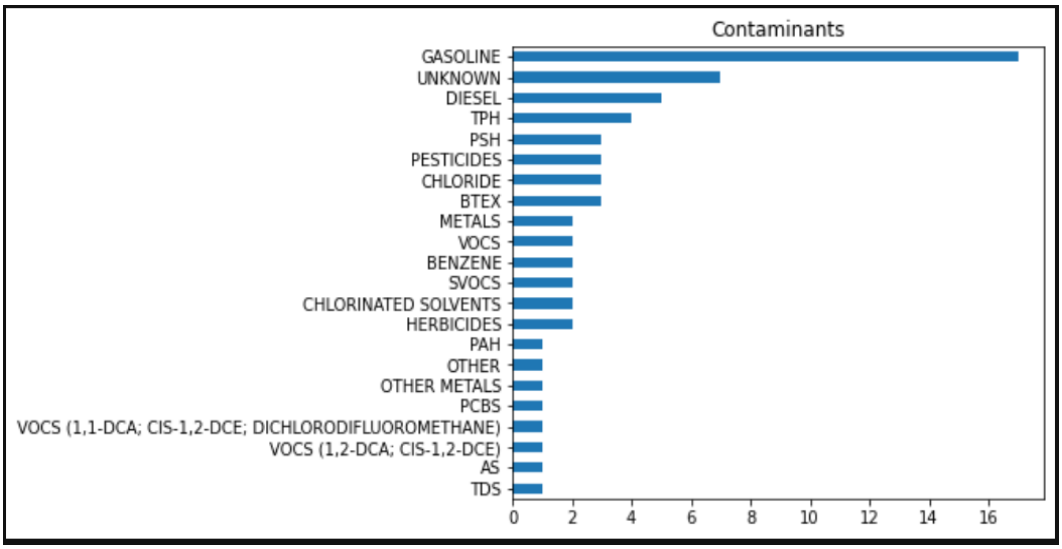

Figure 2. Types of Contaminants

Major polluters

To identify major polluters, a category was manually added in the excel file based on the establishment located in the middle. Value_counts() was used to get the total and it was plotted as a horizontal bar series which is shown in Figure 3.

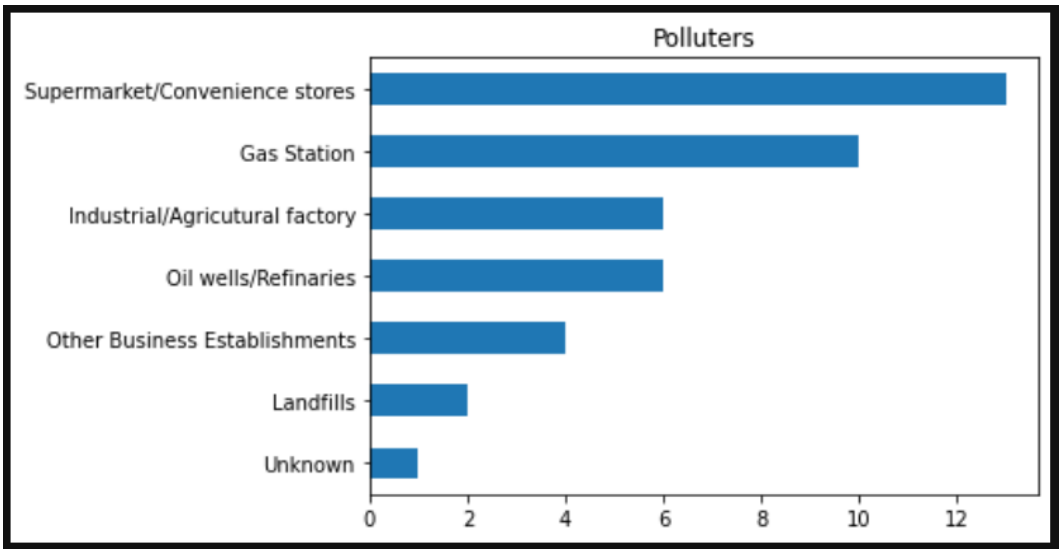

Figure 3. Identified polluters

Number of contaminants per year

Bar plot will be used to correlate Cases per year and Closed Cases per year. Dates column was parsed as dates. Converted to DateTime using pd.DateTime(). The year was extracted and inserted as a new column using df.dt.year(). Created a new column with the only 1s. Year column was group by year using df.groupby(column).count(). The index was reset for the dataframe and then plotted as a stacked bar chart, as shown in figure 4. A new data frame was loaded with just cases that was determined to be no action was needed and repeated the steps above.

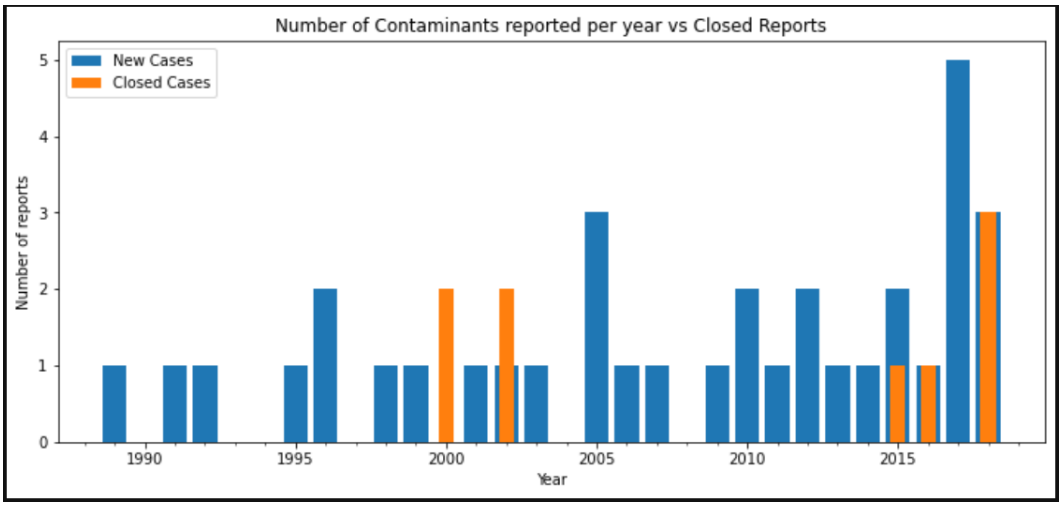

Figure 4. New Cases (Blue) per year and closed cases (orange)

Plotting the location of sites reported with its status

The locations will be plotted as scatter plots on top of a basemap using OpenStreetMap package. First, a library was created with definition of the code from theACT-STATUS column. The library was then inserted as a new column using df.apply(lambda x: [column]) function. This function lets the user take the specific number or arguments (Python Tutorial, n.d.) . The same step was taken to create a new column of color designation base on the status type. Shapefiles of city borders and state borders were added using package cartopy.shapereader, and the geometries were extracted as a list and transformed into a cartopy feature using df.ShapelyFeature(). OpenStreet map was used to render google map images as a base map and overlay with the created features. A new dataframe was created per status to input separately as a scatter plot. This method was the only way the legend would show up. Figure 5 shows the location of the sites, color-coded with specific status.

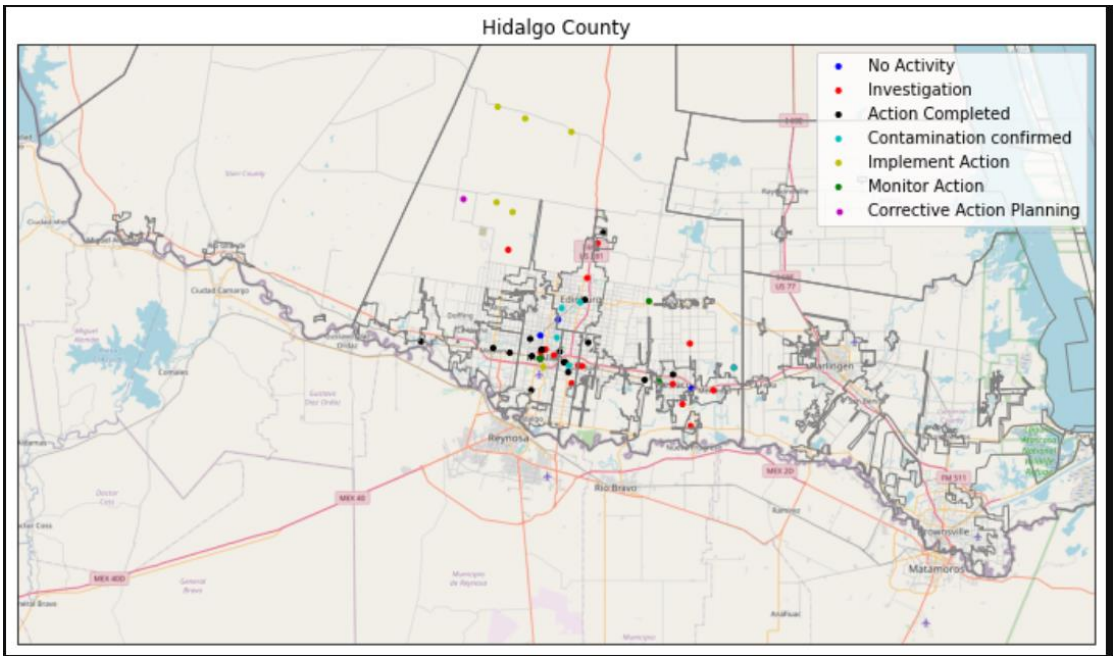

Figure 5. Scatter Map of the contaminants and it's ACT Status

Significance of research

This research provides an illustration of groundwater contaminants occurrences in Hidalgo county that can quickly provide knowledge and awareness to the residents of the area. This paper can be used as additional data for local databases, site-specifc studies and future regional reports.

Results and Discussion

Through Python, we manipulated the data presented from the report by the TCEQ (2020) and visualize the data. Figure 1 showed us the distribution of contaminants in the area. We identify that gasoline has the highest count for contaminants, followed by unknown contaminants. The report did not explain more about the unknown contaminants. For the polluters or perpetrators analysis, convenience stores and supermarkets have the highest count of contaminations followed by the Gas stations. Looking at the excel file, Gas stations and supermarket/convenience stores are mostly the sources of Gasoline contamination. Out of 52 contaminants, only 18 are considered "Action Completed". Based from the time series bar chart. Only 9 are cases that have no further action is needed. The map shows us the location of contaminants and its status. "Action complete" sites are clustering in the Mcallen area, and no other patterns are observed. This clustering are possibly the superfund Donna and the McAllen Plume. Some of the reported cases does not have a date when the action was completed or when did the case got reported. Further research is needed to fully complete the site-specific study.

Conclusion

Groundwater is the primary source of freshwater for most inhabitants of this nation. Groundwater is an essential factor for water security and providing nourishment for the population. This study used Python to analyze the data from the groundwater contamination report by TCEQ. Horizontal Bar plots were used to identify the type of contaminants and the primary perpetrators. Bar plots were used to compare the new cases per year or cases closed per year. Lastly, we plotted the locations of each site using packages of *Cartopy, OpenStreetMap*, and *shapereaders*. Based on the observation, most cases fall under the "Investigation" and "Action completes" status. Python was able to accommodate almost all kinds of analysis that are needed for this study. Further research about the missing information is required to complete the analysis

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.