

Course Report

Francis Biney

May 2022

Contents

1	Introduction	2
2	Exploration of CEJST for EL Paso county.	2
3	Summary of CEJST Tool	3
3.1	4
4	Research Question (What we Did)	4
5	Literature Review	4
6	Methodology	4
7	Statistical Analyses	5
7.1	Data	5
8	Application and Result	10
8.1	Visualization of Texas County and El Paso census Tract	10
8.2	Visualizing Disadvantage Community(tract) by CEJST criterion	11
8.3	Visualizing Communities with water quatlity problem	13
8.4	Effect of high% lead Piant and wastewater Exposure	15
9	Conclusion	16
10	Suggestion	16
	Reference	18

1 Introduction

In this work, we explore the distribution, correlations and effect of lead and wastewater exposure in El Paso communities (census tracts) using geospatial heat maps and data obtained from the Climate and Economic Justice Screening Tool (CEJST) initiative website. The CEJST tool is a Biden-Harris Administration's project to help address inequity and inequality by assisting federal agencies in identifying disadvantaged communities that are marginalized, underserved, and overburdened by pollution (CEJST 2022). We also provided feedback on features that we thought were great, features that we would like to see, and roadblocks we encountered while implementing the Climate and Economic Justice Screening Tool (CEJST) in its beta phase from github.

2 Exploration of CEJST for EL Paso county.

Figure 1 shows the CEJST result for a community in El Paso with zip code 79968. It contains information such as the census tract id, the county, state, population, and whether or not the community is classified as disadvantaged. The community is identified as not disadvantaged because it was not disadvantaged in any of the eight categories defined by CEJST team (CEJST 2022).

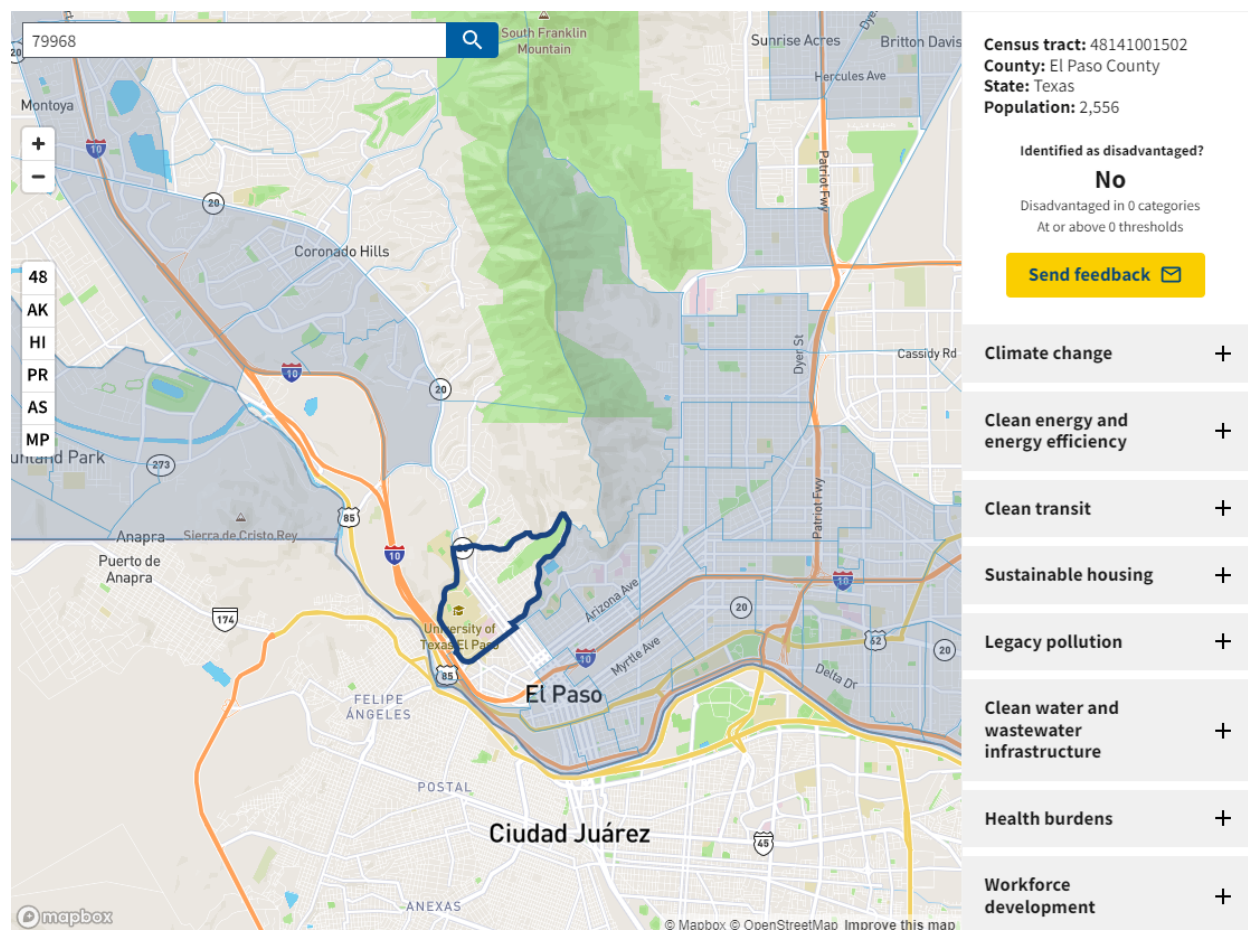


Figure 1: Screening tool output for Zip code 79938 in El Paso

CEJST Tool

3 Summary of CEJST Tool

In this section, we provide a summary of what we would like to see, the features we thought were great, and the roadblocks we encountered during the Climate and Economic Justice Screening Tool implementation process (CEJST).

3.0.1 Feed back on what can be improved

To better analyze, understand and respond to the need of communities, here are the feature we will like to see in the application:

1. A filter that allows users to change the binary scale (disadvantage or not) to a discrete scale $(0, 1, \dots, 8)$, where 0 means not disadvantaged, 1 means least disadvantaged and 8 means most disadvantaged community, that is disadvantaged in all 8 categories. This will allow the user to easily identify highly disadvantaged communities when prioritizing funding.
2. A filter that allows users to filter by categories such as water infrastructure within a county. This will help users to know which communities(census tract) are most affected by a given category. For example, if a user wants to prioritize funding for water infrastructure she/he should be able to filter to see which tracts in the community are affected. This will also help community groups to better understand the different types of challenges they face and the type of funding they should apply for.
3. The data used for the Economic injustice analysis did not include race/ethnicity. I believe there is an association between the dominant race in a community and their health, environmental and socioeconomic needs. The analyses and application will be more informative if the race data is included.
4. To better understand the effect of lead on communities, information on lead pipes use in the communities need to be included and be used in the analysis not only lead paint.
5. A button to display/highlight the effect on health measures and life expectancy of a disadvantaged community given the category. For example, what is the general life expectancy for people in the community given their lead exposure level?
6. Provide code snippets or interactive applets that allow non-tech users to explore their own questions on different data.

3.0.2 Point of Implimenttion

On the implementation of the Climate and Economic Justice Screening Tool (CEJST) in its beta stage from Github, here are our roadblocks:

1. The installation of the application from the Github repository is not user-friendly, and non-tech users will find it hard or almost impossible to get the application to work. it requires a lot of technical know-how and youtube time. It took me about a week to get the application to run on my local computer.
2. To get the application to run on my local computer
 - a. Install git on my windows computer
 - b. Clone (copy) the directory (folder) from Github to my local computer
 - c. Install docker on windows
 - i. Install/Set up WSL2 (windows subsystem for linux 2)
 - ii. Install Ubuntu on Windows
 - iii. Download and install docker

3.1

Finally, here are the features we liked and thought were great. The application was intuitive to use once the installation was done. A good description of the methodology and data used is available on the localhost. This allowed for answering further questions that were not addressed in the application. On the technical end, I appreciated the opportunity to install and run Docker for the first time. This has given me the experience to work in a similar framework in the future

4 Research Question (What we Did)

- Where in El Paso (census tract) has higher percentage of lead paint exposure.
- Where in El Paso (census tract) has higher percentage wastewater discharge exposure.
- What is the effect of higher percentage of lead paint exposure on health of the El Paso community.
- What is the effect of higher percentage of wastewater discharge exposure on health of El Paso community.

5 Literature Review

Lead pollution has been shown to be especially harmful in low-income neighborhoods and communities of color, according to the Center for Disease Control (CDC, 2021). Children who live in lower income neighborhoods or communities of color are more likely to be exposed to lead based paint, dust, or soil than their higher income peers. These exposures have devastating effects on the growth and development of children's brains, and have been linked to behavioral issues such as attention deficit hyperactivity disorder ADHD (Needleman 1982), poor school performance (Kuang et al. 2020), poor memory (Arnvig, Grandjean, and Beckmann 1980), lower IQ scores (Needleman and Gatsonis 1990), and reduced life expectancy in adults (Sandhya, Kumari, et al. 2016).

Paso del Norte community in El Paso, Texas is known to have a high concentration of lead in the air and soil (Darby 2012). This is due to heavy metal contamination caused by the extraction of lead, copper, and zinc from metal ore by the American Smelting and Refining Company (ASARCO) (Marcosson 1949) from 1901 to 1999, which was acquired by Grupo Mexico in 2009 and closed in 2009. (Darby 2012). Lead and arsenic persist in the region's soils (Alvarez 2015) and perhaps in people's bodies, ground, and surface water, despite the fact that the smelter is no longer in operation.

Residents and regulatory agencies have focused on these two pollutants due to their health effects on children and the environment.

Mielke and Reagan (Mielke and Reagan 1998) showed that lead is particularly troubling due to its propensity to settle in the soil, where it can be inhaled or ingested by humans. Long-term lead exposure reduces cognitive functioning, which can influence educational outcomes and economic opportunities (Schell 1997).

6 Methodology

Here we discuss the methods we used to answer these question. We focused on exploring the distribution, correlation and association between water quality related measures such as exposure to lead paint, exposure to waste water discharged and basic demographic measures including education level, socio-economics status, and health related measures such as percent adult diagnose Diabetes, Asthma, Heart disease and life expectancy in the El Pose community. We did this using numerical and graphical summaries such as box plot conditional bar graphs, and geospatial heat maps. The goal was to understand the effect of exposure to high lead pain and wastewater exposure have on disadvantaged communities in El Paso.

Table 1: First eight observations and columns

id	county	state	T_threshld	T_cat_exceded	is_disAdvan	total_population	LILpHE
01001020100	Autauga County	Alabama	0	0	FALSE	1923	FALSE
01001020200	Autauga County	Alabama	0	0	FALSE	2028	TRUE
01001020300	Autauga County	Alabama	0	0	FALSE	3476	FALSE
01001020400	Autauga County	Alabama	0	0	FALSE	3831	FALSE
01001020500	Autauga County	Alabama	0	0	FALSE	9883	FALSE
01001020600	Autauga County	Alabama	0	0	FALSE	3705	TRUE
01001020700	Autauga County	Alabama	0	0	FALSE	4029	TRUE
01001020801	Autauga County	Alabama	0	0	FALSE	2826	FALSE

7 Statistical Analyses

7.1 Data

The data use is available on the Climate and Economic Justice Screening Tool (CEJST) website under methodology and data page, published on March 30,2022. The data contain 74131 observations(rows) and 83 variables(Columns) with information on environmental, climate and socio-economic related measures. The measures include activities that adversely impact human health, climate and the environment such as use of lead paint, Traffic proximity, Proximity to hazardous waste site, diesel particulate matter exposure. It also has data on health related measures such as heart disease, asthma, diabetes and socio-economic measures such as low income, Unemployment, Poverty rate and higher education enrollment rate.The target variable contains information on whether or not a community is disadvantage based on the criterior below. The (CEJST) team defined a community qualifies as disadvantaged if:

1. the U.S. census tract is above the threshold for one or more environmental or climate indicators, and
2. the U.S. census tract is above the threshold for the socioeconomic indicators.

These categories and threshold can be found on the (CEJST) website. Additional goeospatial data was obtained from the *tigris package* in R, which was combined with (CEJST) data to create the El Paso map with the census tract. Table 1 shows the first eight observations and columns in the data.

7.1.1 Data Processin and Feature Enginerring

The original data came from multiple source with long columns name so we gave the variable a shorter name. This was done using the *janitor package* in R. We then explore the 83 columns for water quality and basic demographics related measures such as lead and wasted water, income, Education level. El Paso county data was subset form the 74131 original observations to 161.

Missing values Treatment

Very few observations (3.1%) were missing,so we dropped them from the data. Wastewater discharged, life expectancy and media house hold income have most missing values ranging from 27%,11.4% and 6.2% respectively as shown in the figures 2 and 3.

Feature Engineering

The logical variables *Lead.Paint,Proximity to hazard waste site* and *Proximity to wastewater discharged* were re-coded as binary factor with categories *(yes,no)* for *(True,false)* respectively for modeling purpose.

7.1.2 Data Exploration and Summary

We performed exploratory data analysis on the El Paso data to understand the the distribution, correlations and association between the water quality and demographics variables. This was done by inspecting the data distribution for patterns, trends and also summarize the data into it main characteristics.

Water quality data

From table 2, about 63% of El Paso census tract were classified as disadvantaged communities and 37% not by the CEJST formula (CEJST 2022) for deciding disadvantage communities. Table 2 below show that 5% of El Pasoans live in lead painted house ,24% close to wastewater discharge area and none close to Hazard waste site. This also shown in figure 4. The distribution of the wastewater contain mostly high values see figure 5.

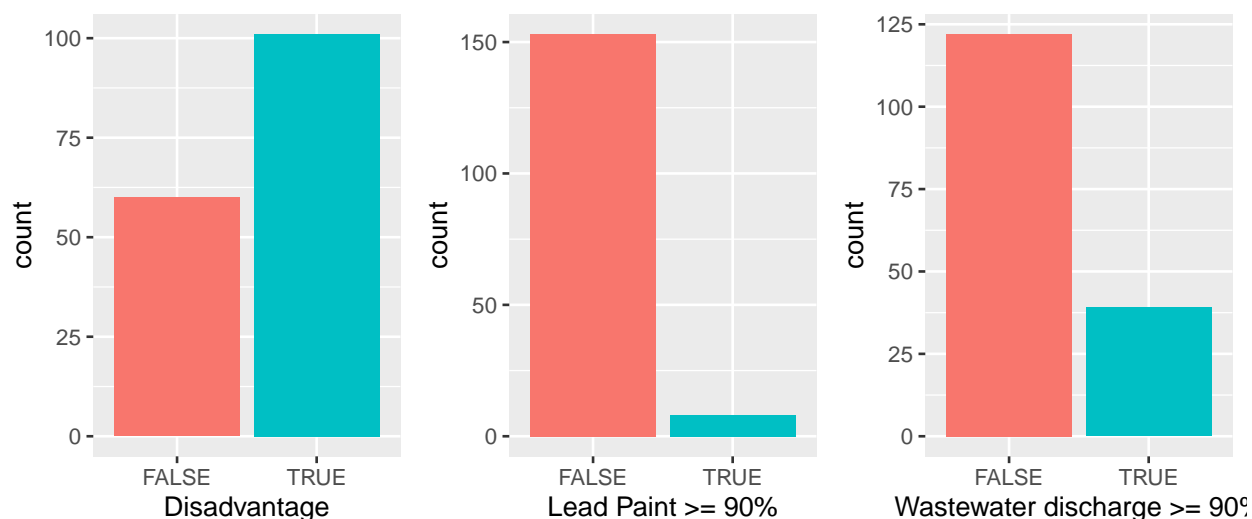


Figure 4: Distribution of disadvantage communities in El Paso, exposure to lead Paint and wastewater

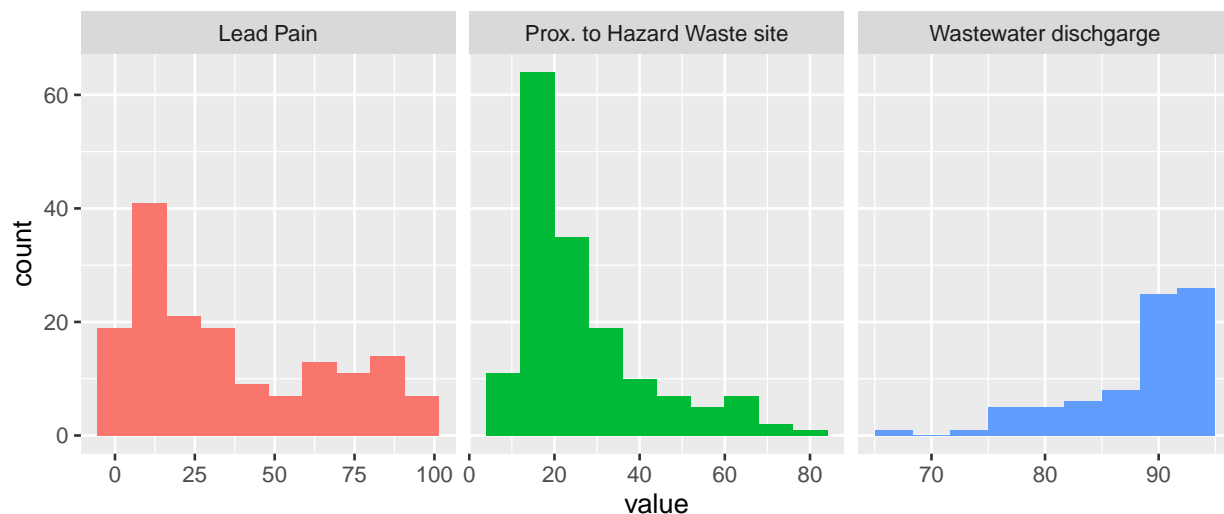


Figure 5: Distribution of the percentile of El Paso communities exposed to lead paint, Harzard waste and wastewater

Table 2: Summary statistics of data distribution

Characteristic	**N**	**n(%)**
___Disadvantage___	161	
FALSE		60 (37%)
TRUE		101 (63%)
___Lead Paint___	161	
FALSE		153 (95%)
TRUE		8 (5.0%)
___Prox. to Harzd. site___	161	
FALSE		161 (100%)
TRUE		0 (0%)
___Wastewater discharge $\geq 90\%$ ___	161	
FALSE		122 (76%)
TRUE		39 (24%)

Table 3: Summary statistics of data distribution

Characteristic	**N**	**Statistics**
___Lead Pain Exp. $> 90\text{th}$ ___	161	
Mean (SD)		36 (30)
Median (IQR)		25 (11, 65)
___Prox. to Hazard Waste site___	161	
Mean (SD)		27 (15)
Median (IQR)		22 (16, 33)
___Wastewater discharge___	77	
Mean (SD)		88.3 (5.8)
Median (IQR)		90.0 (86.0, 93.0)
Unknown		84

Demographics and Health Data

Table 4, figures 7 and 6, shows the summary statistics and distribution of El Paso demographic and health related data. From table 3, about 43% and 14% of the of the adult in the census tract are diagnose with diabetes and heart disease respectively.

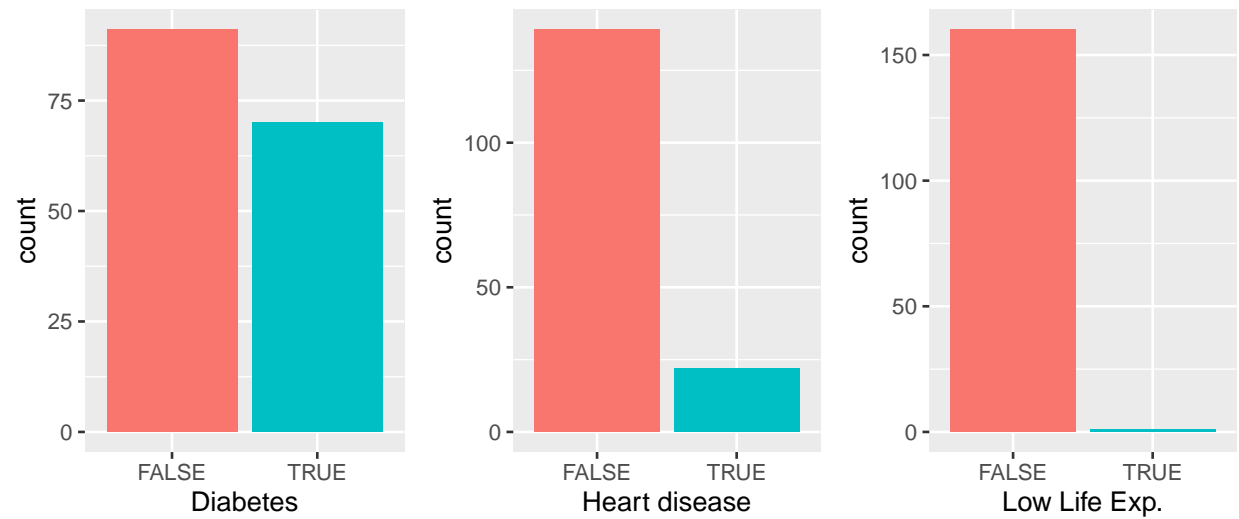


Figure 6: Distribution of Low, Life Expectancy, Diabetes and Heart disease in El Paso community

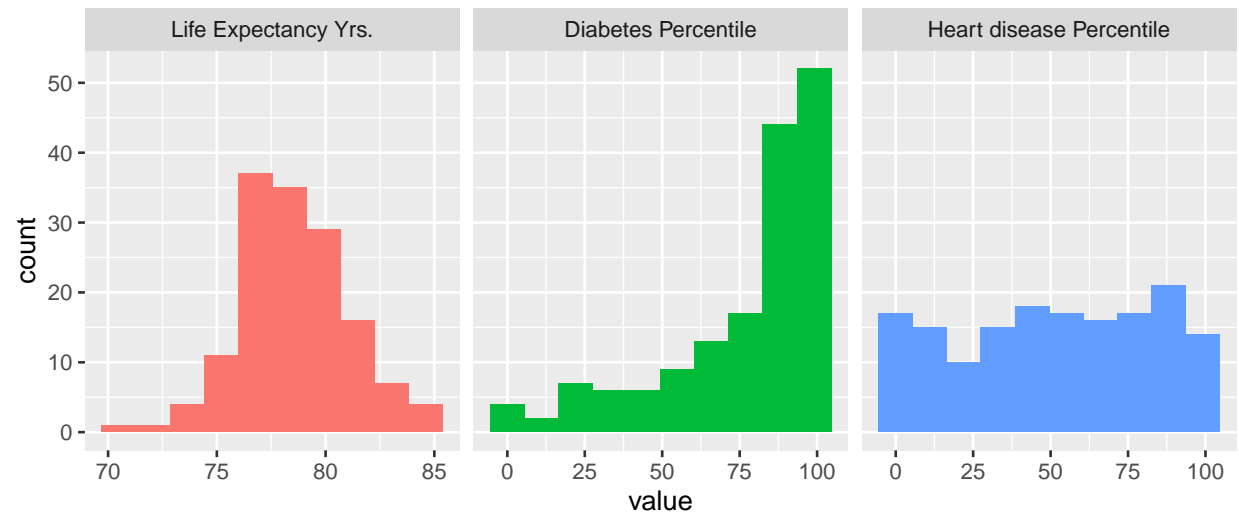


Figure 7: Distribtion of the percentile of El Paso communities for Life Expectancy, Diabetes and Heart disease

8 Application and Result

8.1 Visualization of Texas County and El Paso census Tract

Figure 8 show the 254 county boundaries in Taxes with El Paso highlighted with red arrow. El Paso county and it census tract are shown in figure 9.

Texas County Map



Figure 8: Texas county boundaries

El Paso Census tract

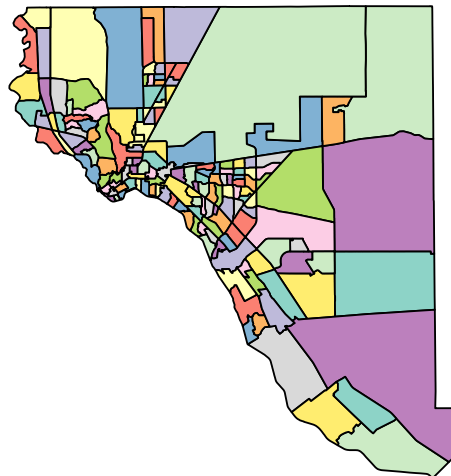


Figure 9: El Paso Census tract

8.2 Visualizing Disadvantage Community(tract) by CEJST criterion

To understand which communities in El Paso are disadvantaged by (CEJST) criterion we made geospatial heat maps for El Paso census tract, showing whether the tract is disadvantaged or not. From table 2, about 63% of El Paso communities are disadvantaged, the blue area in figure 10 below. Here a community is identified as disadvantaged if it is disadvantaged in at least one of the eight categories (CEJST 2022).

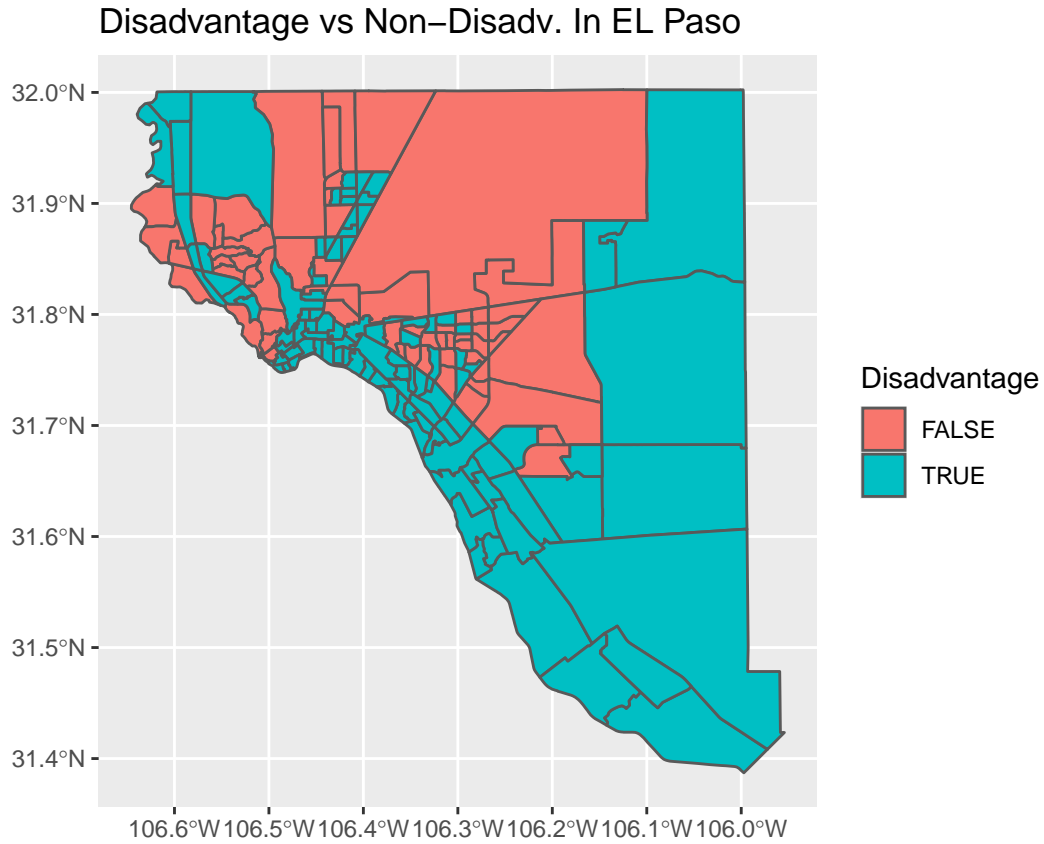


Figure 10: El Paso census trac by Disadvantage Communities, TX

8.2.1 Most disadvantaged communities in El Paso

To highlight El Paso communities that are most disadvantaged we create a heat map with 0 meaning not disadvantaged, 1- least disadvantage and 8 means most disadvantaged community. The areas showing deep red in figure 11 are the most disadvantaged communities.

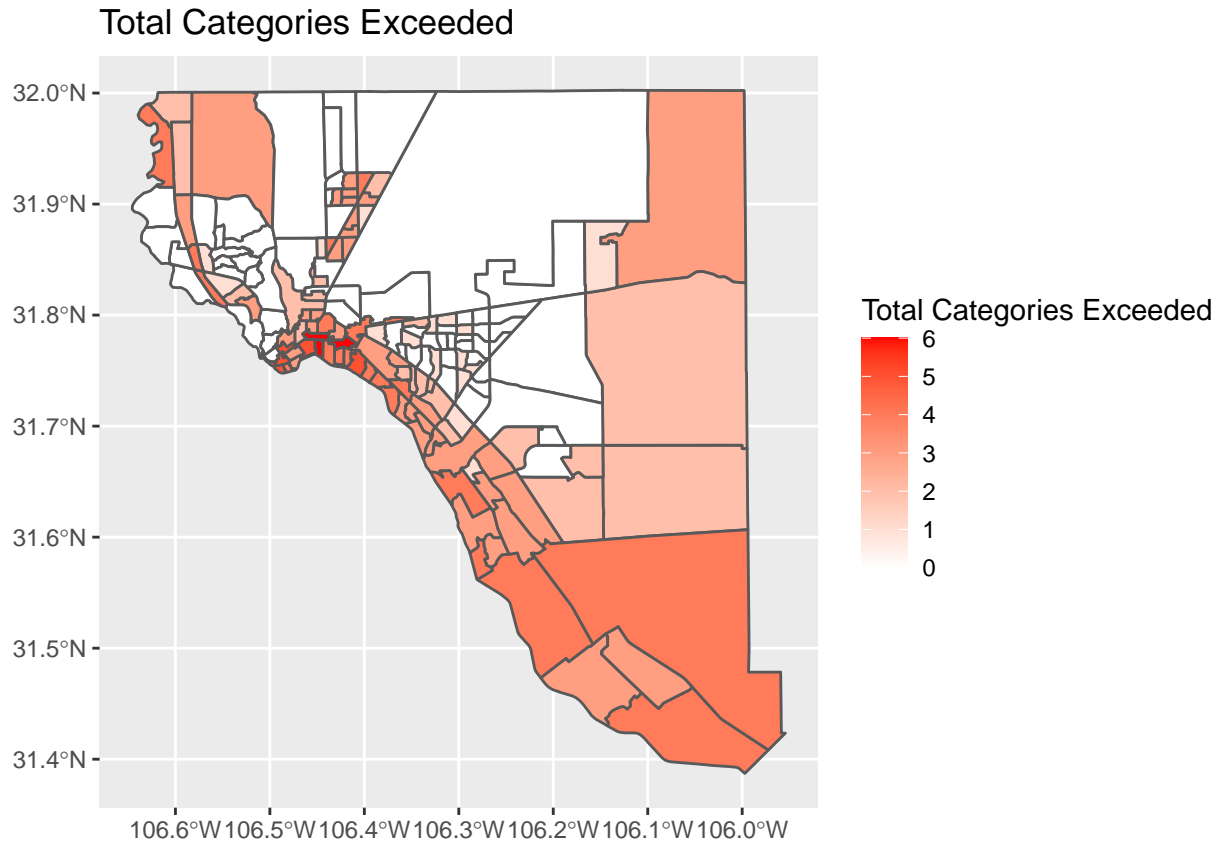


Figure 11: El Paso census trac by Total Categories Exceeded

8.3 Visualizing Communities with water quatlity problem

Here we define community to have water quality problem if they are exposed to lead paint or wastewater.

8.3.1 Lead paint Exposure

From table 2 about 5% of El Paso communities are exposed to lead. These communities live in houses with lead paint. Figure 12 shows these communities in blue, about the same communities that are most disadvantaged in El Paso from figure 11.

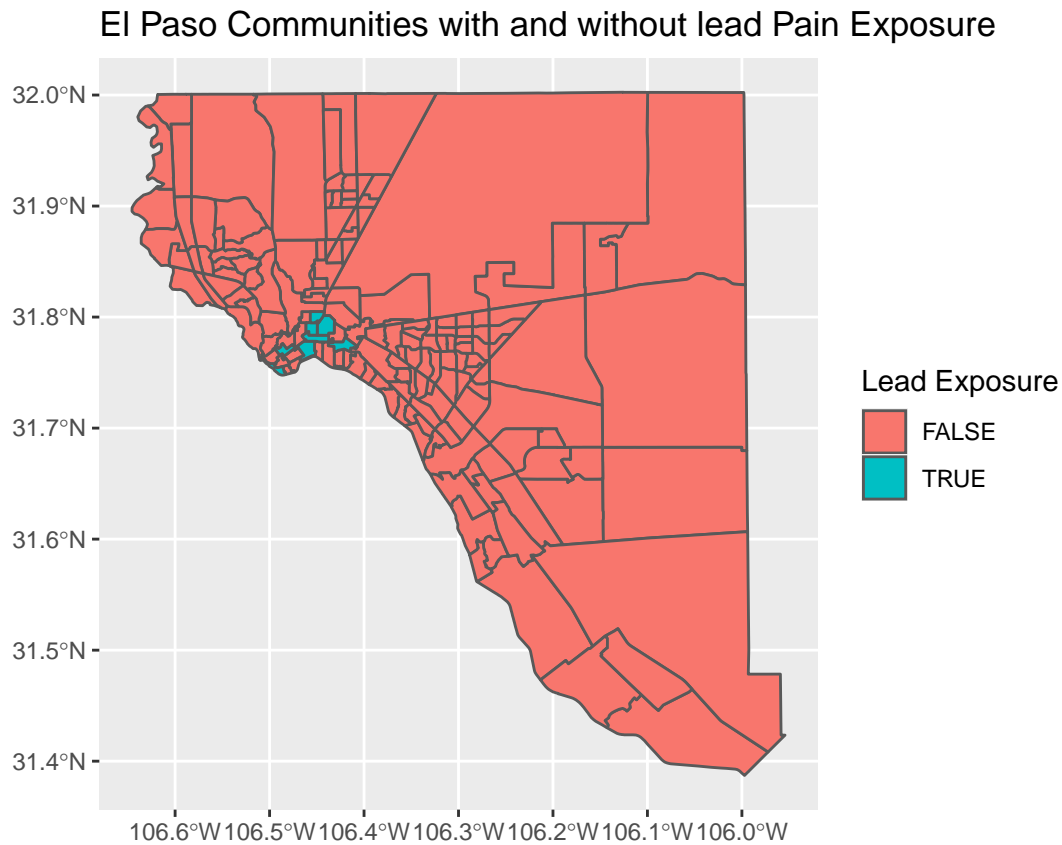


Figure 12: El Paso census trac by Lead Exposure

8.3.2 Wastewater Exposure

Figure 13 shows the heat map of El Paso communities exposed to wastewater discharge. Table 2, about 24% of the El Paso communities are exposed to wastewater discharge, shown in blue in figure 13. This area also covers the communities exposed to lead and the most disadvantaged in figures 11 and 12

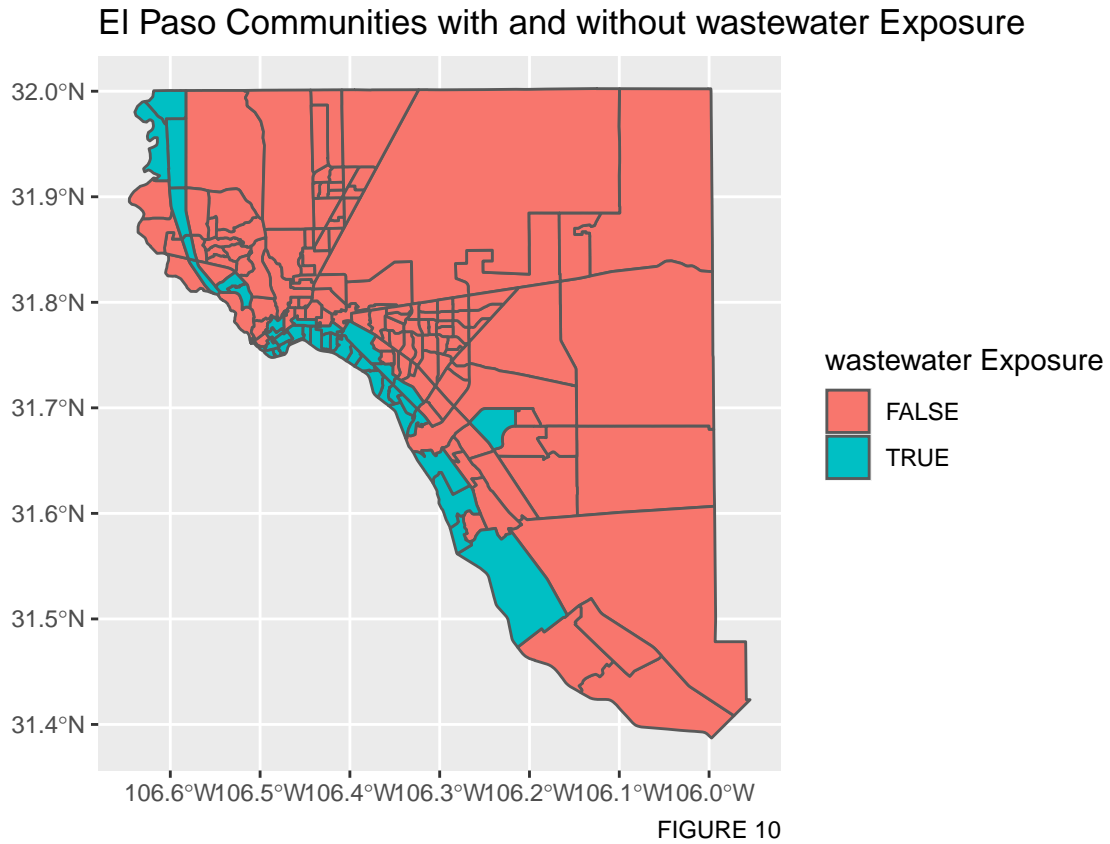


Figure 13: El Paso census trac by wastewater Exposure

8.4 Effect of high% lead Piant and wastewater Exposure

In this section, we explore and visualize whether or not differences exist in medical conditions between communities in El Paso exposed to high percentages of lead and wastewater.

8.4.1 Effect of high% lead paint exposure on life Expectancey, Diabetie and Heart Disease

The box plots in figure 14 below show significant differences in medical conditions between communities that live in lead-painted houses and not. From the box-plots, communities that live in lead-painted houses have a high number of adults diagnosed with diabetes, heart disease and have low life expectancy.

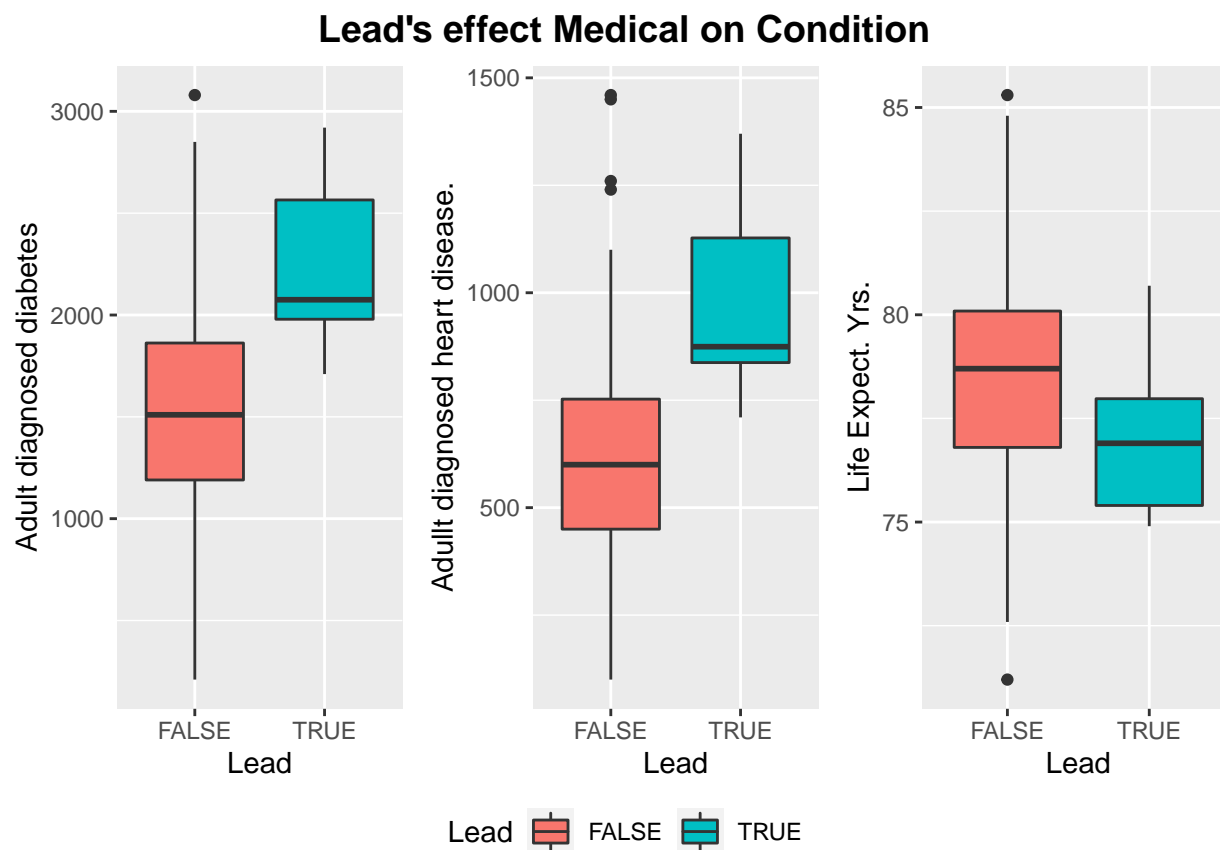


Figure 14: Effect of lead on diabetes,heart diseaseand life expectancy in El Paso

8.4.2 Effect of high% wastewater discharge on life Expectancey, Diabetie and Heart Disease

Figure 15 shows significant differences in medical condition between communities living near and far from wastewater discharge. Communities living near wastewater discharge had a high percentage of adults diagnosed with diabetes, heart disease, and a short life expectancy, according to the box plots.

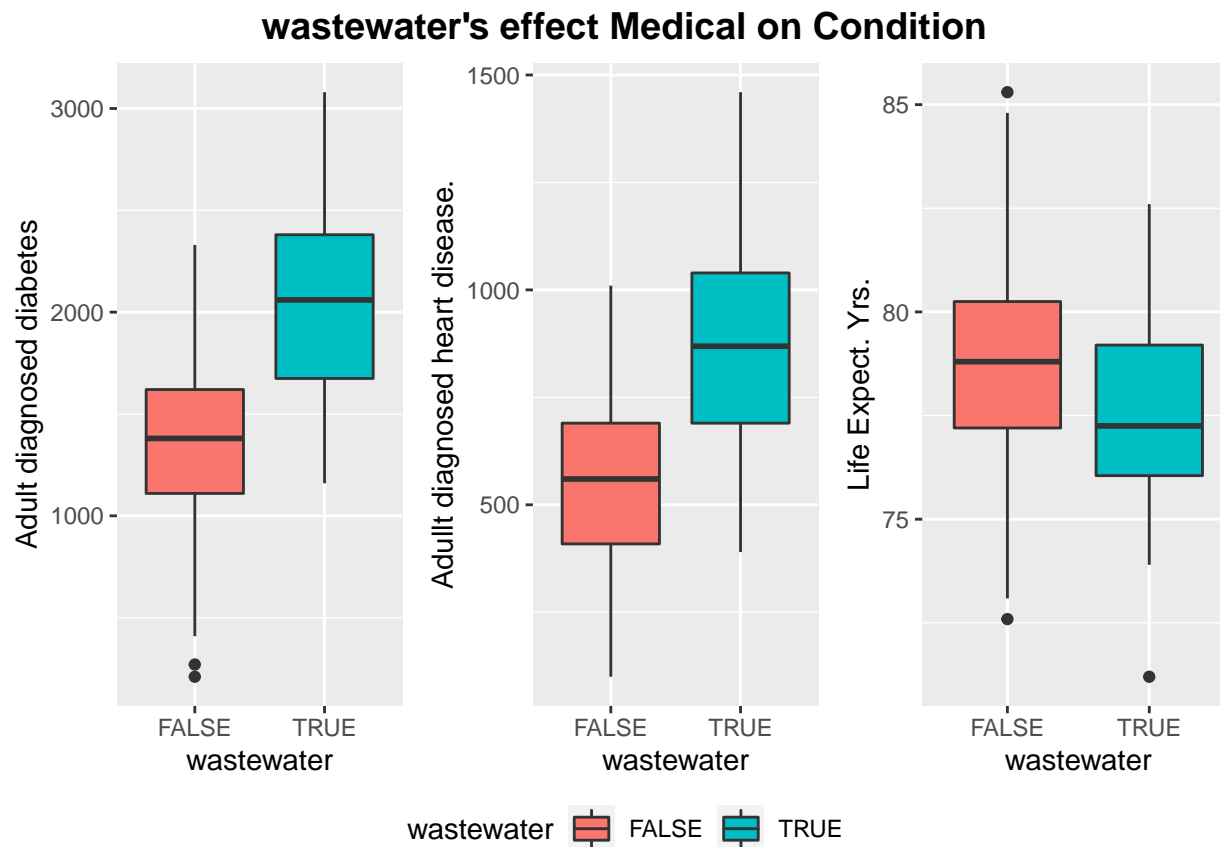


Figure 15: Effect of wastewater on diabetes, heart disease and life expectancy in El Paso

9 Conclusion

This work explores the distribution, correlation, and effect of exposure to lead and wastewater on heart disease, diabetes, and life expectancy in El Paso communities. We accomplished this by using geospatial heat maps, boxplot, data from the CEJST initiative, and the Tigris package for CRAN.

In our exploration and analysis, we observed that communities with high exposure to lead and wastewater tend to have a higher number of adult diagnosed with diabetes and heart disease. These communities also have low life expectancies relative to unexposed communities. We also observed that most communities with high lead exposure also have high wastewater exposure and low life expectancy.

10 Suggestion

The analysis will be more informative if communities with lead pipes are included in the data. Other suggestions include:

- Statistical Analysis to justify whether a significant difference exists between the groups compared
- Model the *above.90th.percentile.PM2.5* using binomial logit model with other columns as predictors.
- Model the *above.90th.percentile.lead.paint.exposure* using binomial logit model with other columns as predictors.
- Find features that are top predictors of disadvantaged communities.
- Model the *Total.categories.exceeded* using the proportional odds model.

Table 4: Summary statistics of data distribution

Characteristic	**N**	**Statistics**
___ Life Expectancy Yrs. ___	145	
Mean (SD)		78.50 (2.46)
Median (IQR)		78.59 (76.80, 80.09)
Unknown		16
___ Diabetes ___	161	
FALSE		91 (57%)
TRUE		70 (43%)
___ Heart disease ___	161	
FALSE		139 (86%)
TRUE		22 (14%)
___ Diabetes Percentile ___	160	
Mean (SD)		77 (26)
Median (IQR)		88 (68, 96)
Unknown		1
___ Heart disease Percentile ___	160	
Mean (SD)		52 (31)
Median (IQR)		53 (26, 80)
Unknown		1

Reference

- Alvarez, Juan M. 2015. *Comparison of Blood Lead Levels Between Children in an Urban Setting and Children in a Rural Setting*. The University of Texas at El Paso.
- Arnvig, Eva, Philippe Grandjean, and Jørn Beckmann. 1980. "Neurotoxic Effects of Heavy Lead Exposure Determined with Psychological Tests." *Toxicology Letters* 5 (6): 399–404.
- Darby, Kate J. 2012. "Lead Astray: Scale, Environmental Justice and the El Paso Smelter." *Local Environment* 17 (8): 797–814.
- Kuang, Wenjie, Zhaofang Chen, Kexin Shi, Hong Sun, Hongbo Li, Lei Huang, and Jun Bi. 2020. "Adverse Health Effects of Lead Exposure on Physical Growth, Erythrocyte Parameters and School Performances for School-Aged Children in Eastern China." *Environment International* 145: 106130.
- Marcosson, Isaac Frederick. 1949. *Metal Magic: The Story of the American Smelting & Refining Company*. New York: Farrar, Straus.
- Mielke, Howard W, and Patrick L Reagan. 1998. "Soil Is an Important Pathway of Human Lead Exposure." *Environmental Health Perspectives* 106 (suppl 1): 217–29.
- Needleman, Herbert L. 1982. "Lead and Impaired Abilities."
- Needleman, Herbert L, and Constantine A Gatsonis. 1990. "Low-Level Lead Exposure and the IQ of Children: A Meta-Analysis of Modern Studies." *Jama* 263 (5): 673–78.
- Sandhya, Kanupuru, JP Kumari, et al. 2016. "Impact of Lead on Environment and Human Health-a Review." *World Journal of Pharmaceutical Research* 5 (4): 531–44.
- Schell, Lawrence M. 1997. "Culture as a Stressor: A Revised Model of Biocultural Interaction." *American Journal of Physical Anthropology: The Official Publication of the American Association of Physical Anthropologists* 102 (1): 67–77.