

A scenic view of a park with a pond, a bridge, and city buildings in the background.

Authors

Lap Pham

Patrick Swain

José de los Ríos

Maxwell Snodgrass

Advisor

Randall Walsh, Ph.D.

MQE Capstone for the Pittsburgh Parks Conservancy: Measuring and Mapping Community Investment Need

Acknowledgements

Throughout this project, **Professor Randall Walsh** was an invaluable resource as an advisor, from broad discussions on the scale of the analysis to guidance on complex technical processes to ensure the robustness of our results. Notably, he recommended the technique of areal interpolation for attaching third-party data to walksheds, which became the cornerstone of our geospatial data analysis. Additionally, **Professor David Huffman** provided counsel in periodic check-ins during the process of creating the project.

We would also like to acknowledge the 2019 report performed by **Interface Studio LLC** for the Pittsburgh Parks Conservancy, from which we derived spatial data including shapefiles of parks and walksheds that became an integral part of our analyses and improved the efficiency of our workflow. Additionally, 2022 data on air quality across North America by the **Atmospheric Composition Analysis Group** at Washington University was a key part of our environmental analysis.

Finally, our primary touchpoint within the Pittsburgh Parks Conservancy was **Ross Chapman**, the Conservancy's Chief of Operations. His breadth of knowledge on Pittsburgh's public spaces from his experience at the Conservancy and as director of the City's Department of Parks and Recreation lent ample expertise to defining the scope of our project and its underlying motivations.

A guide to assist the Pittsburgh Parks Conservancy in their effort to prioritize public investment needs on a park-by-park basis using analyses and visualizations of data characterizing the demographics, environment, public health, crime, and other attributes of Pittsburgh parks and their surrounding walkshed areas.

Demographic
Census Data

Public Health

Environment
and Pollution

Crime

Q: What is a walkshed?

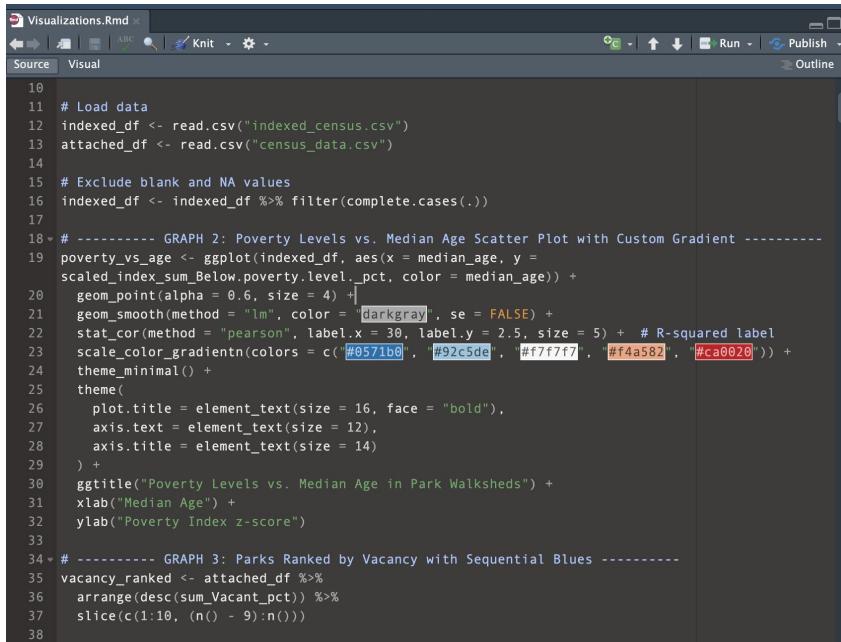
A: The area where someone can walk to a given park from their home in about five minutes. Each park has a walkshed with a measurable shape and population, and we mapped third-party statistics onto them using processes including areal interpolation in R and QGIS.

Tools:

R packages sf, ggplot2, and tmap

Geographic information systems software **QGIS**

Python packages seaborn and matplotlib



A screenshot of the RStudio interface showing an R script named "Visualizations.Rmd". The script contains R code for data manipulation and plotting. The code includes loading data from CSV files, filtering for non-blank and non-NA values, and creating two plots: a scatter plot of poverty levels vs. median age and a plot of parks ranked by vacancy. The code uses ggplot2 for data visualization.

```
10 # Load data
11 indexed_df <- read.csv("indexed_census.csv")
12 attached_df <- read.csv("census_data.csv")
13
14 # Exclude blank and NA values
15 indexed_df <- indexed_df %>% filter(complete.cases(.))
16
17 # ----- GRAPH 2: Poverty Levels vs. Median Age Scatter Plot with Custom Gradient -----
18 poverty_vs_age <- ggplot(indexed_df, aes(x = median.age, y =
19 scaled_index.sum.Below.poverty.level._pct, color = median.age)) +
20 geom_point(alpha = 0.6, size = 4) +
21 geom_smooth(method = "lm", color = "darkgray", se = FALSE) +
22 stat_cor(method = "pearson", label.x = 30, label.y = 2.5, size = 5) + # R-squared label
23 scale_color_gradientn(colors = c("#0571b0", "#92c5de", "#f7f7f7", "#f4a582", "#ca0020)) +
24 theme_minimal() +
25 theme(
26   plot.title = element_text(size = 16, face = "bold"),
27   axis.text = element_text(size = 12),
28   axis.title = element_text(size = 14)
29 ) +
30 ggtitle("Poverty Levels vs. Median Age in Park Walksheds") +
31 xlab("Median Age") +
32 ylab("Poverty Index z-score")
33
34 # ----- GRAPH 3: Parks Ranked by Vacancy with Sequential Blues -----
35 vacancy_ranked <- attached_df %>%
36 arrange(desc(sum_Vacant_pct)) %>%
37 slice(c(1:10, (n() - 9):n()))
38
```

Sources:

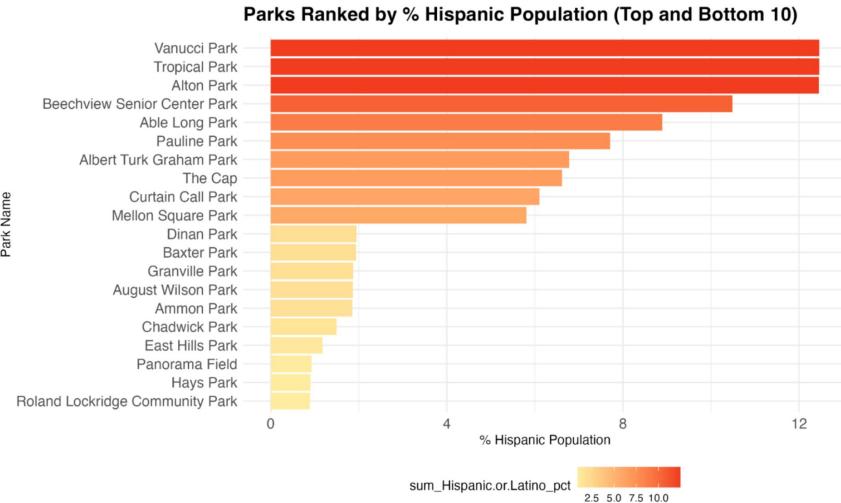
We analyzed data from an array of public sector sources from federal bodies like the **U.S. Census Bureau**, the **Center for Disease Control and Prevention (CDC)**, the **U.S. Department of Agriculture**, the **U.S. Forest Service**, and the **Environmental Protection Agency (EPA)**

Local entities like the **Pittsburgh Water and Sewer Authority (PWSA)** and the **Pittsburgh Police Department (PPD)**

We also cross-referenced private data from the **Pew Research Center** and **Washington University in St. Louis** to corroborate our findings

Race and Ethnicity Demographics

We analyzed **Asian, Black, Hispanic/Latino, and White** populations. Let's use Hispanic/Latino as an example. We drew data from the U.S. Census Bureau, cleaned it, mapped it to parks by interpolating data in the intersection of Census tracts and walksheds, then assembled datasets of raw counts, proportions, index z-scores, and ranks

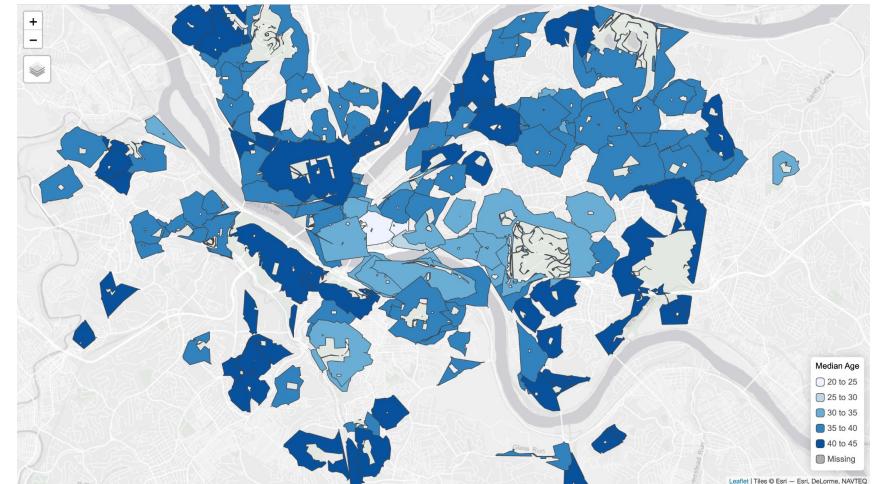
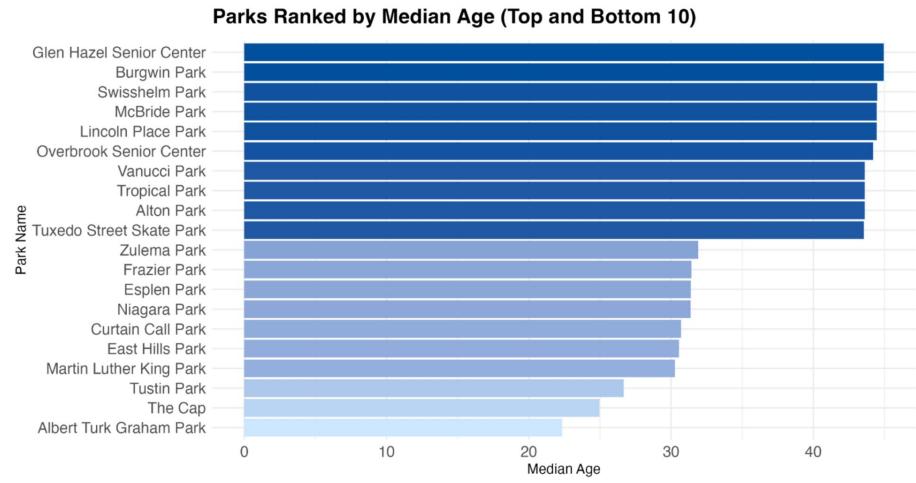


Age

We summarized age columns into a single **median age metric**. This ranged from around 45 for the Glen Hazel Senior Center and Burgwin Park's walksheds in Hazelwood to about 22 for Albert Turk Graham Park in Crawford-Roberts

The locations of **universities** are likely a significant factor in this spatial distribution – many of the youngest walksheds are around the “**studentified**” neighborhoods. The four youngest walksheds are all within half a mile of **Duquesne’s** campus. Parks in Oakland near Pitt also appeared in the top 10 youngest walksheds

Two of the oldest walksheds belonged to **senior centers** – Glen Hazel and Overbrook



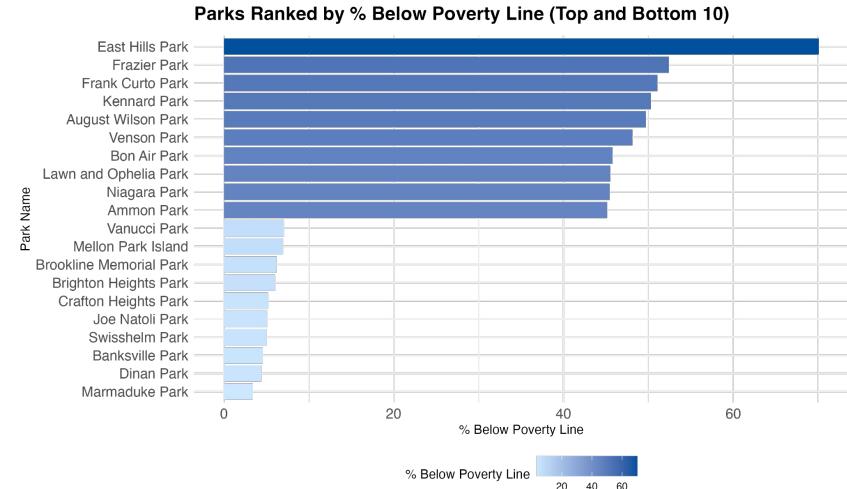
Poverty

The Census Bureau does not have a single threshold for poverty, rather a series of thresholds for each individual based on age, size of household unit, and children. We produced a single column that contained scaled values describing the **relative poverty** of each walkshed

East Hills Park is extremely poor relative to the rest of the walksheds. Poverty tends to be concentrated in walksheds in the **Hill District, Oakland, Knoxville, and Homewood**.

The cluster in **Oakland** could be indicative of the **large student population** in the University of Pittsburgh's footprint, many of whom have little to no annual income as full-time university students

The smallest proportions are in the walkshed of Marmaduke Park, a playground in **Brighton Heights, Downtown, Stanton Heights, Swisshelm Park**, and a few pockets of the **South Hills** made up the rest of the bottom 10 walksheds in poverty

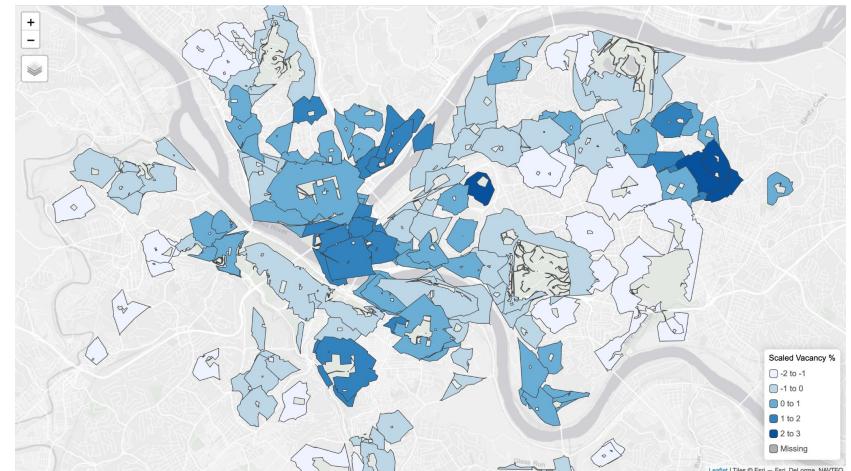
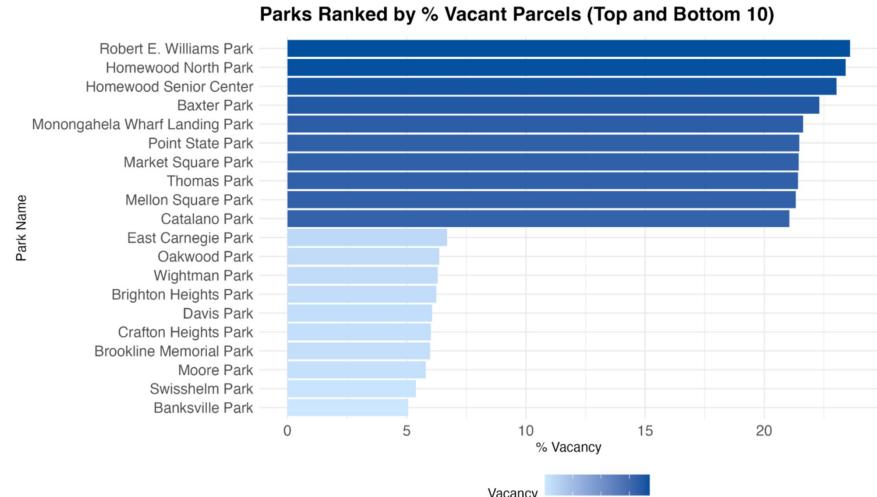


Vacancy

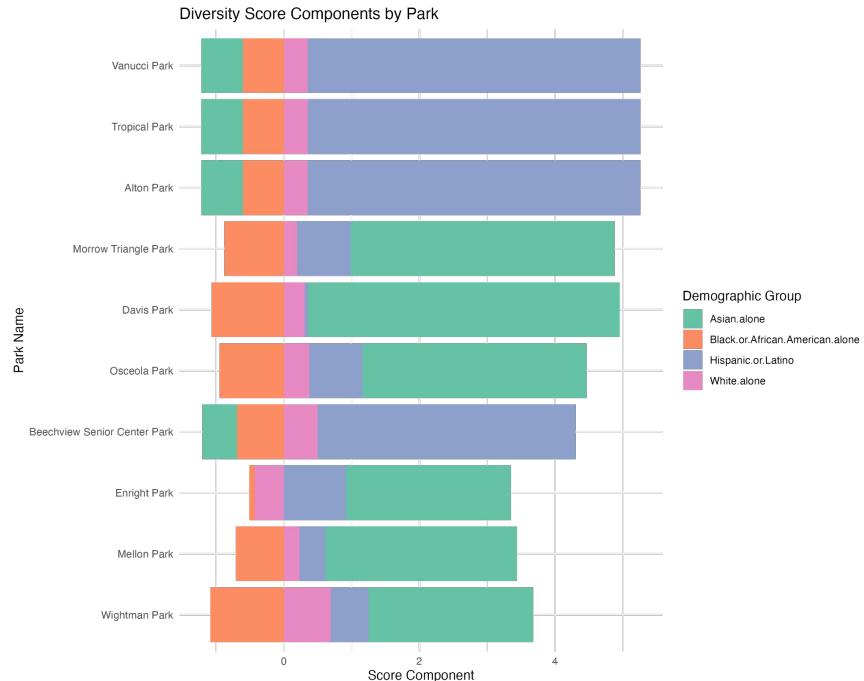
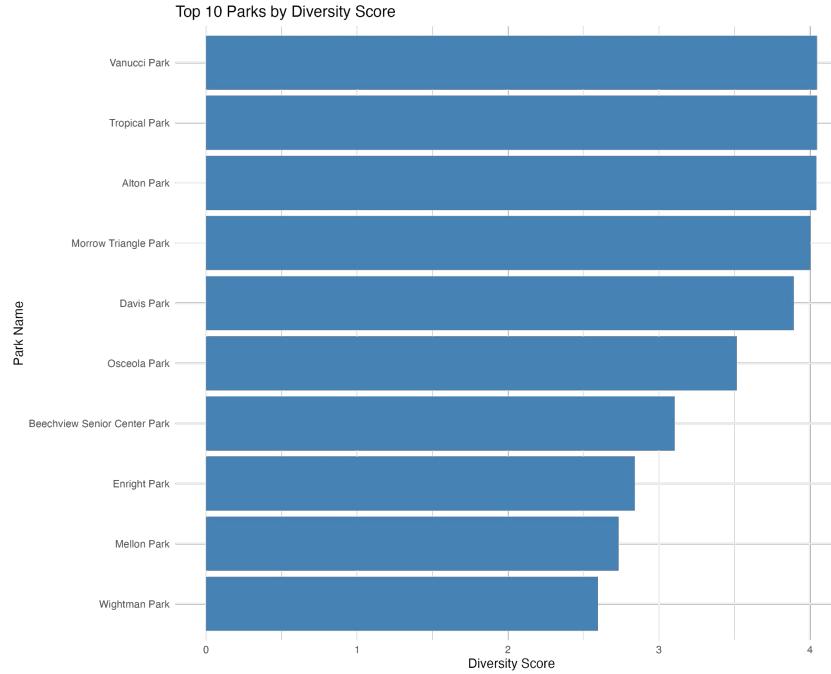
Census Bureau's definition of vacancy – **no permanent residents** at the time of Census recording

The watershed with the highest rate of vacancy was Robert E. Williams Park in the **Upper Hill District**, while the following three were in **Homewood** and the remainder of the top 10, apart from Catalano Parklet in Troy Hill, were in **Downtown Pittsburgh**. The culprit of Downtown's high vacancy rates is likely the high density of **commercially-zoned parcels**, hotels, and office buildings with no legal residents

Vacancy rates were lowest in Banksville Park's watershed in the **South Hills**. Many of the watersheds with the least vacancy are clustered in the corridor from **Friendship to Swisselm Park**. Another noticeable hotspot of low vacancy rates is the heavily residential South Hills



Modeling Diversity



Environmental: Data Preparation

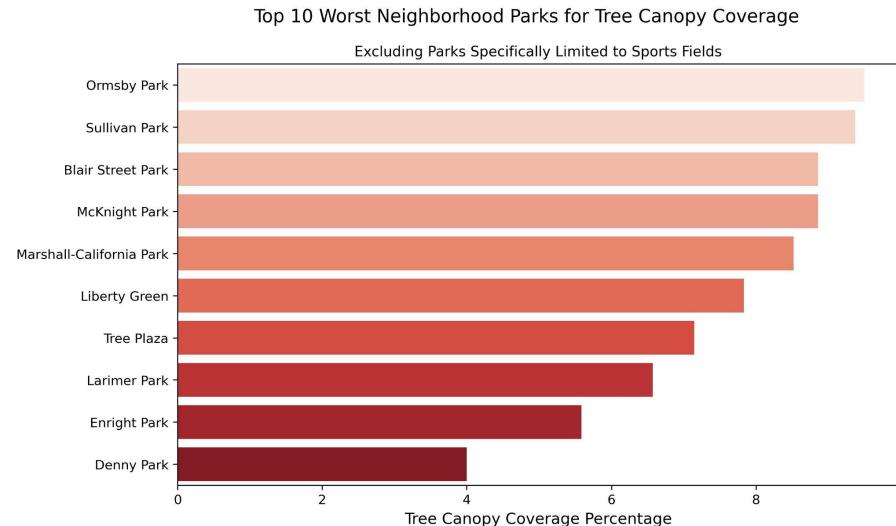
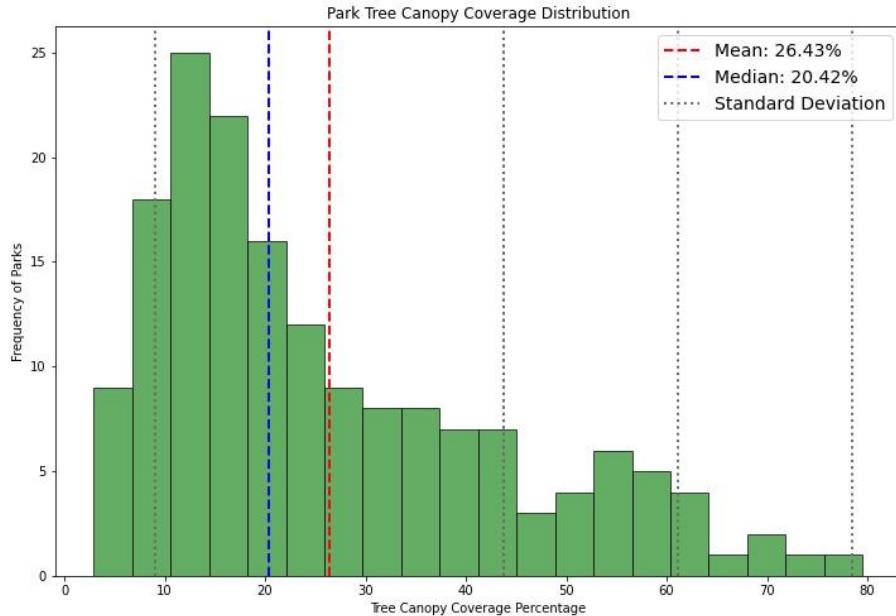
- Tree Canopy Coverage
 - Measures average percentage of tree canopy coverage across entire park
 - Method: Zonal Averaging
- Pollution
 - Measures average PM2.5 concentration levels ($\mu\text{g}/\text{m}^3$) across entire park
 - Method: Zonal Averaging
- Sewersheds Priority
 - Aggregated scores based on PWSA sewersheds tract ratings
 - Measures need for intervention to mitigate overflowing of storm drains, sewers, etc...
 - Method: Areal Interpolation

Zonal Averaging in QGIS



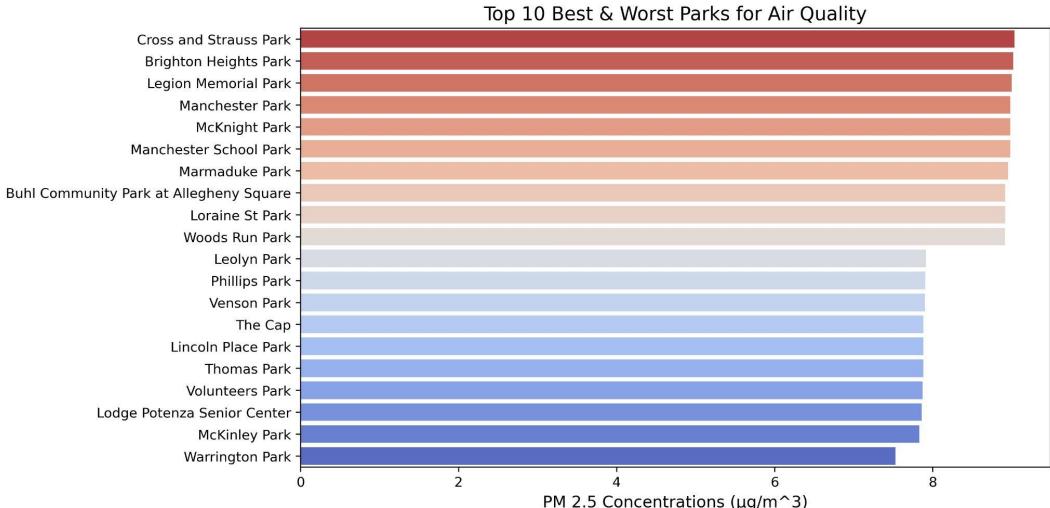
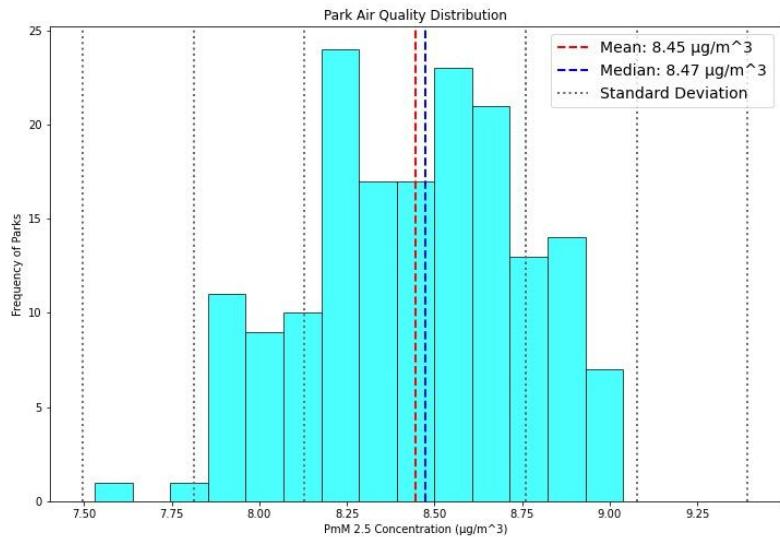
Environmental: Tree Canopy Statistics

- For each variable, summary statistics and data analysis was completed to show the best and worst parks in each. Below is the example for tree canopy coverage.



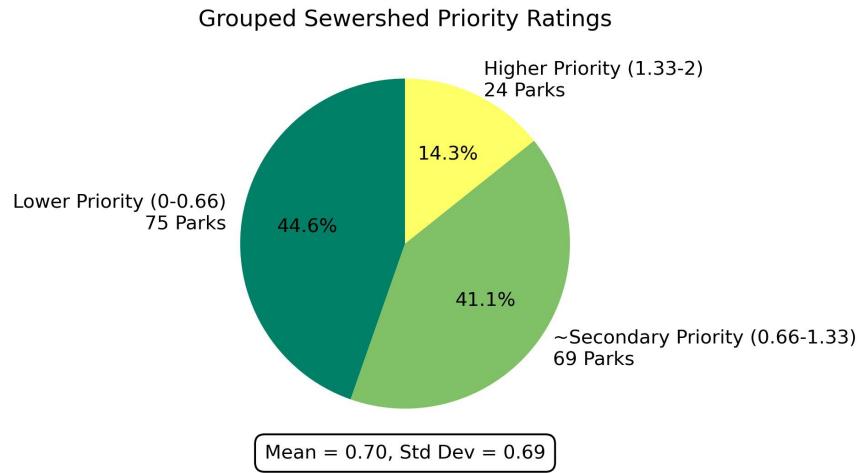
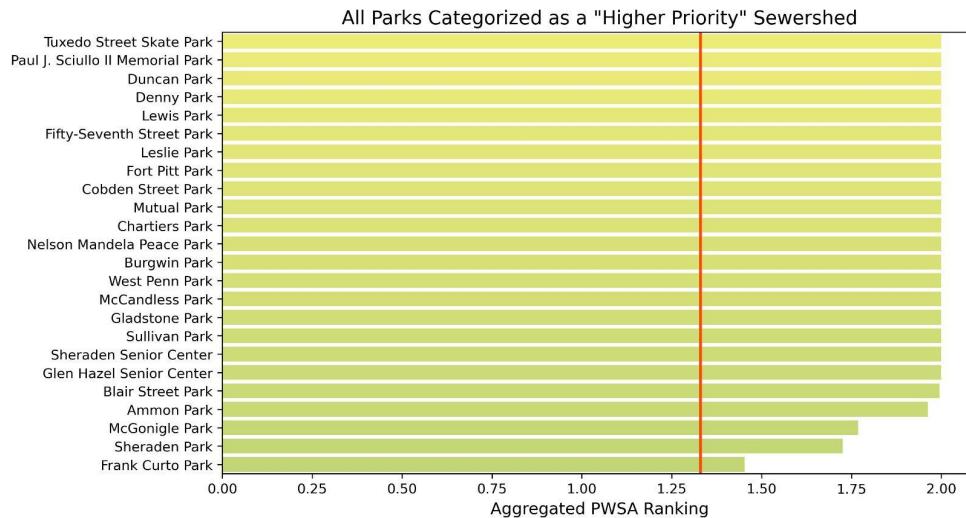
Environmental: Pollution Statistics

- Below are summary statistics for pollution (PM2.5 concentrations)
- Key Takeaway: Parks near Mt. Washington experience the least pollution while parks north of the Allegheny River experience the most pollution



Environmental: Sewersheds Priority Statistics

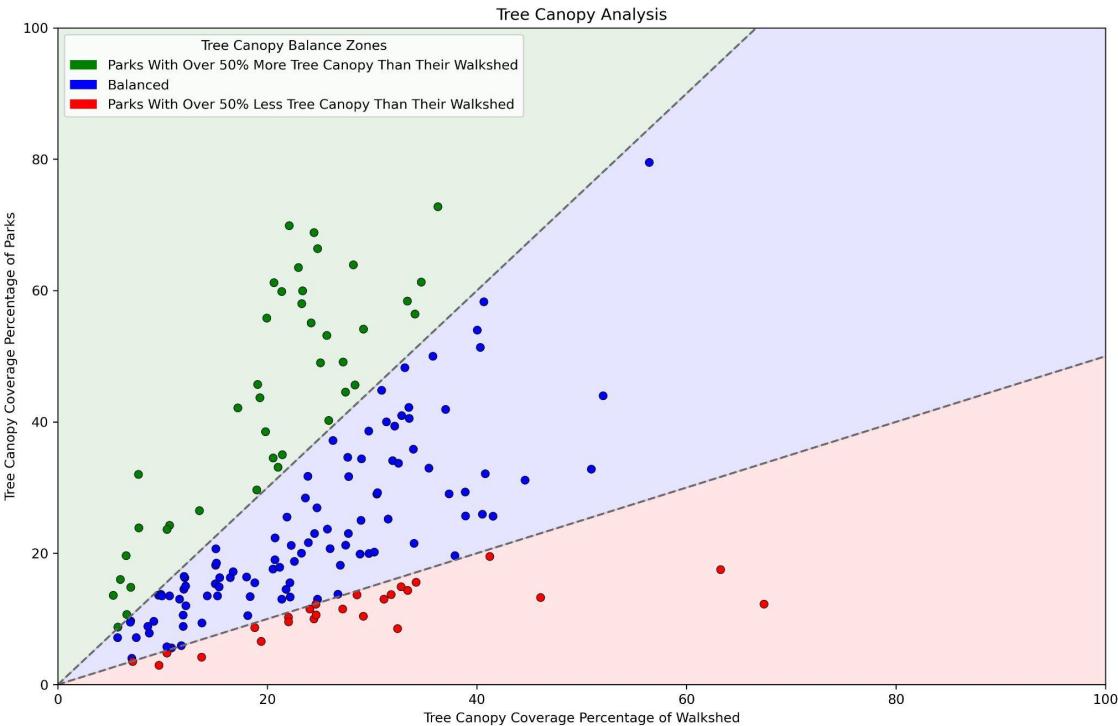
- Parks were given an aggregated rating of 0-2 for sewersheds based off how much they intersect with PWSA rated boundaries.



- 24 parks in the dataset had a rating above 1.33, indicating higher priority for their sewersheds. This makes up 14.3% of all parks.

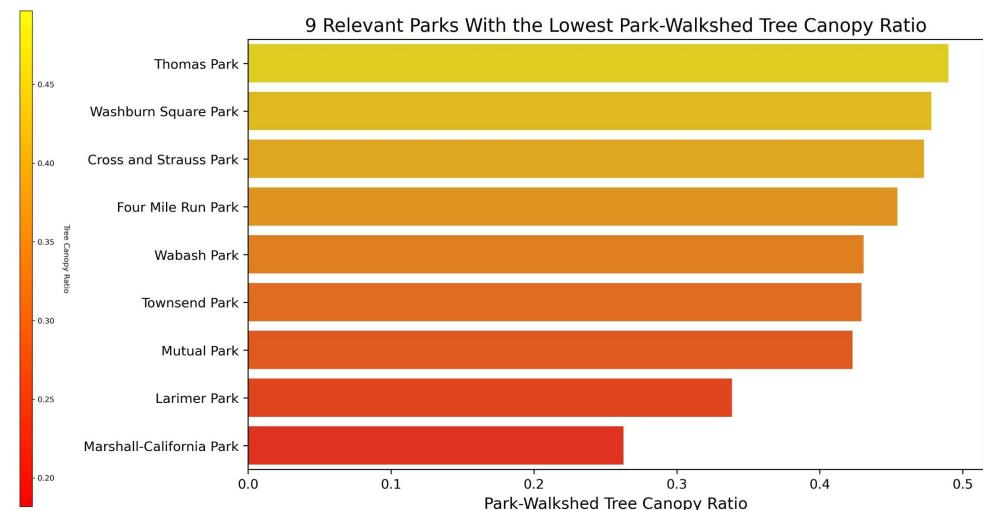
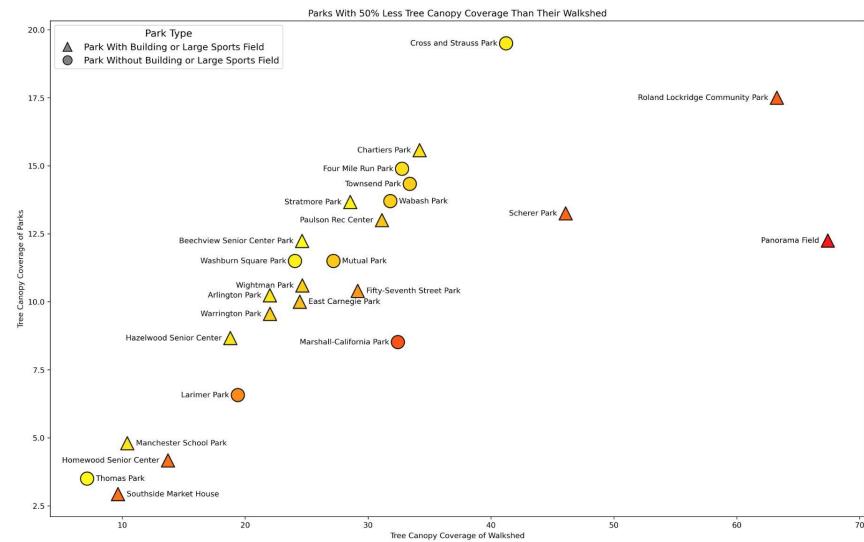
Environmental: Tree Canopy Analysis

- Compare park tree canopy coverage with walkshed tree canopy coverage
- Shows tree canopy coverage relative to the parks' surrounding areas
- Separated into three zones based on park-walkshed tree canopy ratio



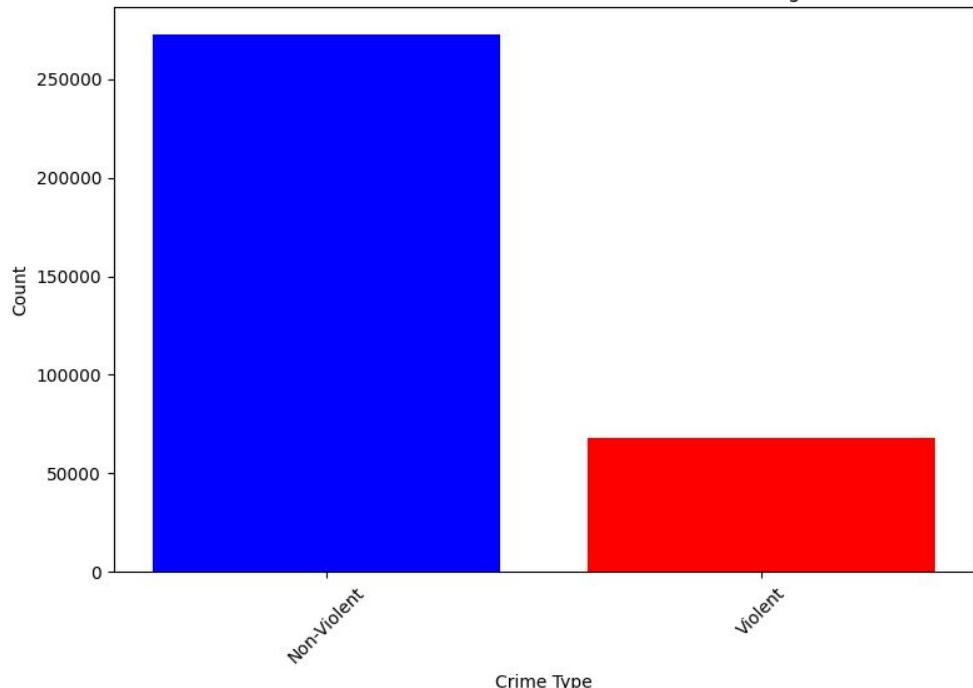
Environmental: Tree Canopy Analysis

- Parks that only comprise of buildings and sports fields are not viable for this analysis. Therefore, they are indicated with a triangle and are excluded from the bar graph.

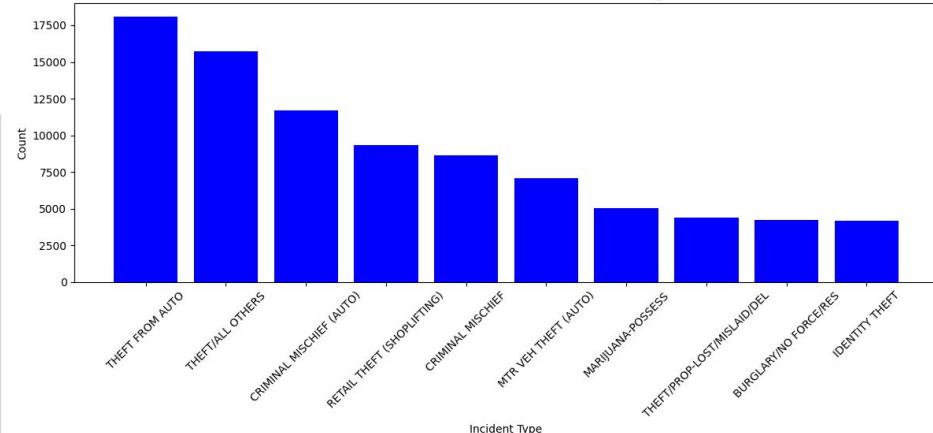


Crime Data Preparation

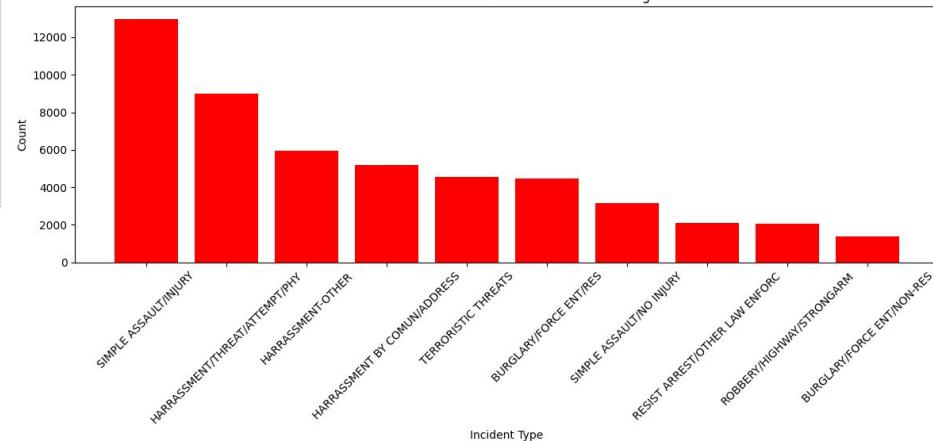
Count of Violent vs Non-Violent Crimes In Pittsburgh



10 Most Common Non-Violent Crimes in Pittsburgh

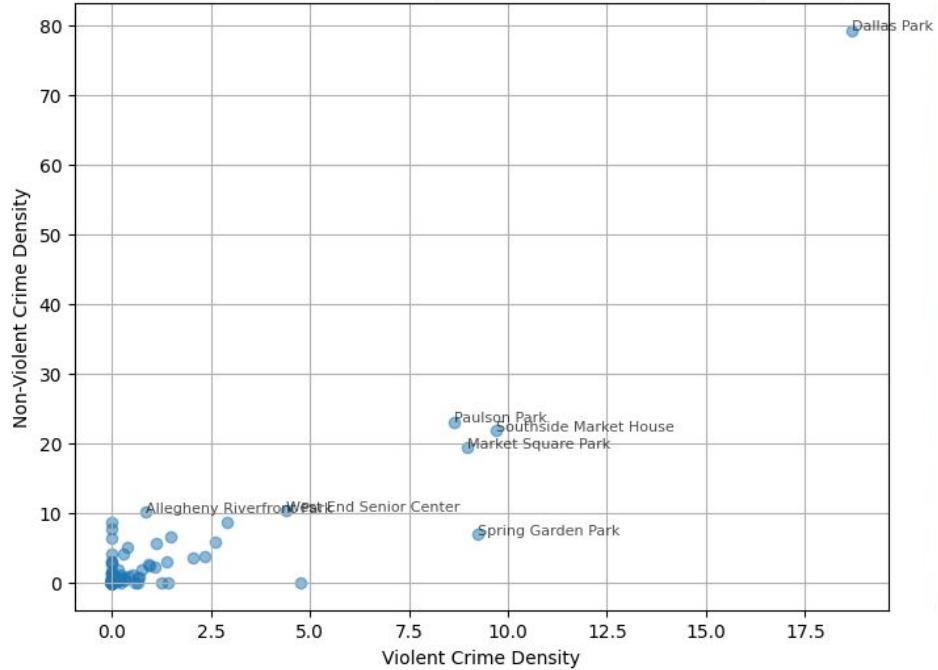


10 Most Common Violent Crimes in Pittsburgh

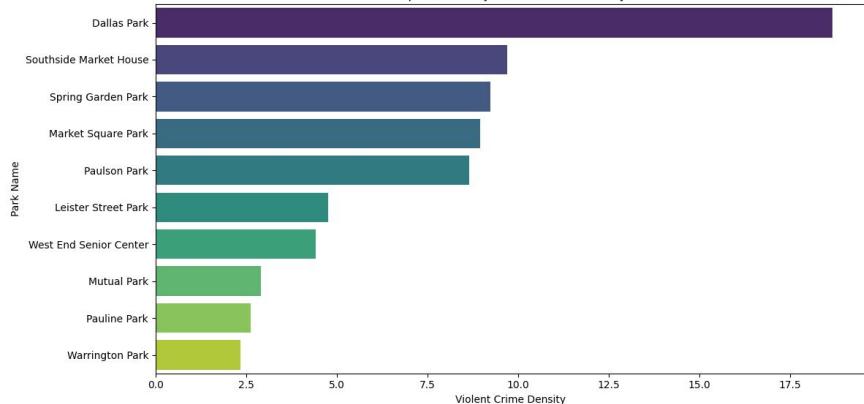


Crime Park Data

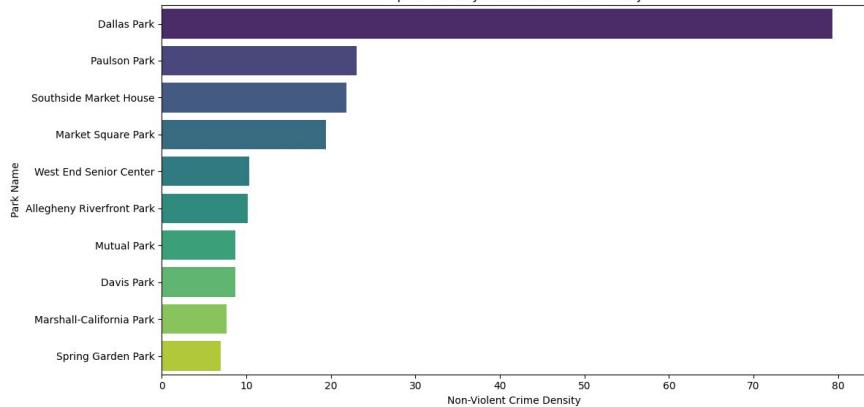
Scatter Plot of Violent Crime Density vs Non-Violent Crime Density Among Parks



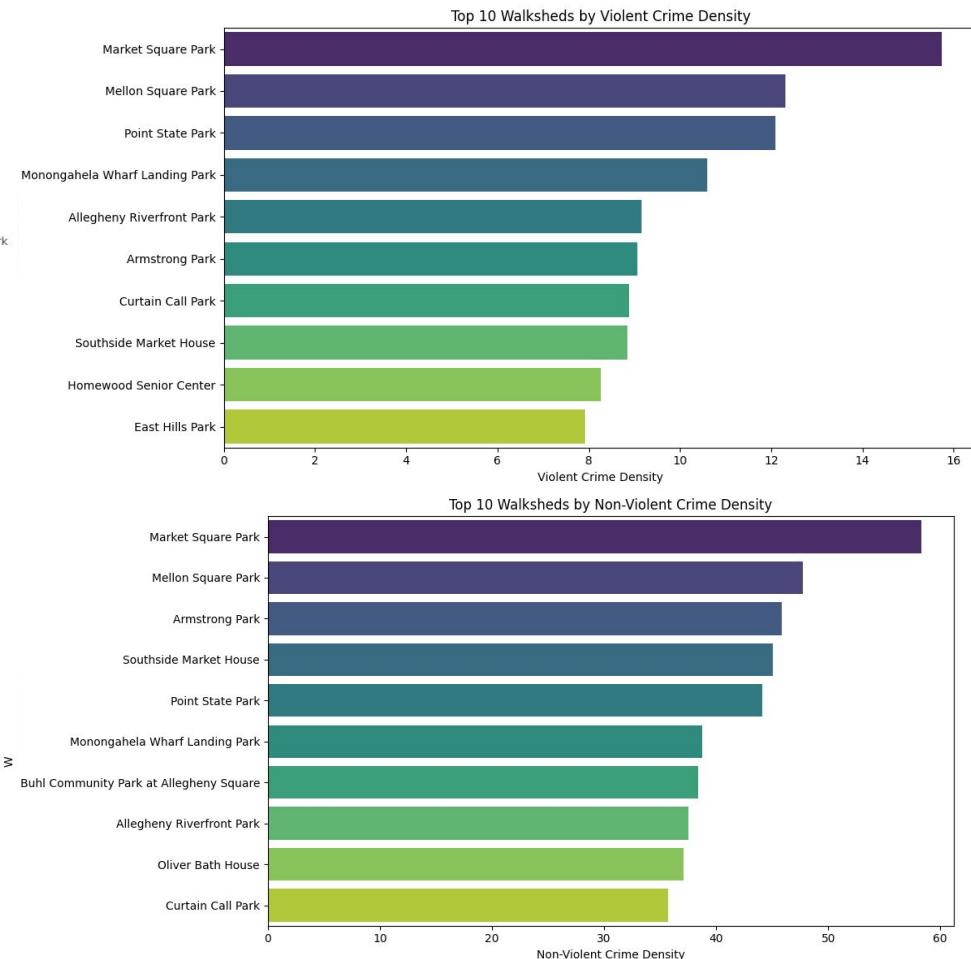
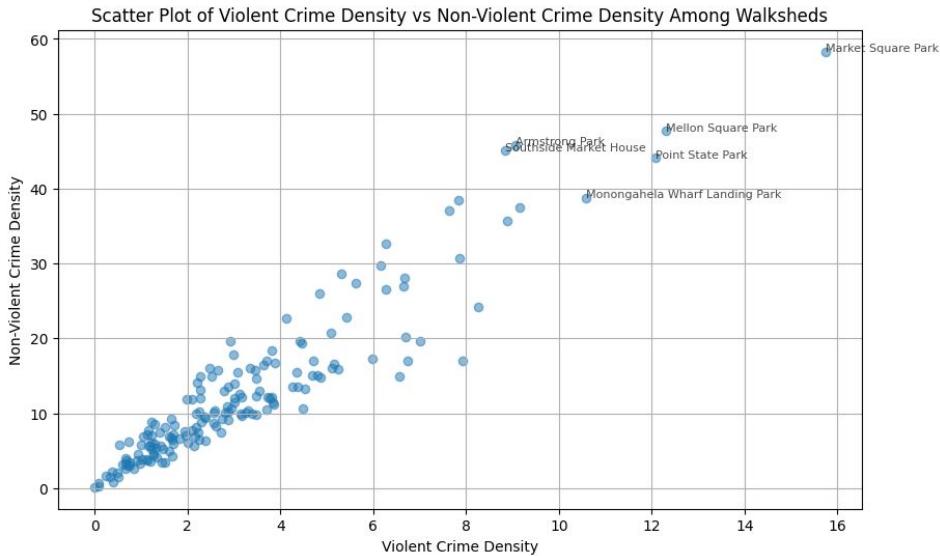
Top 10 Parks by Violent Crime Density



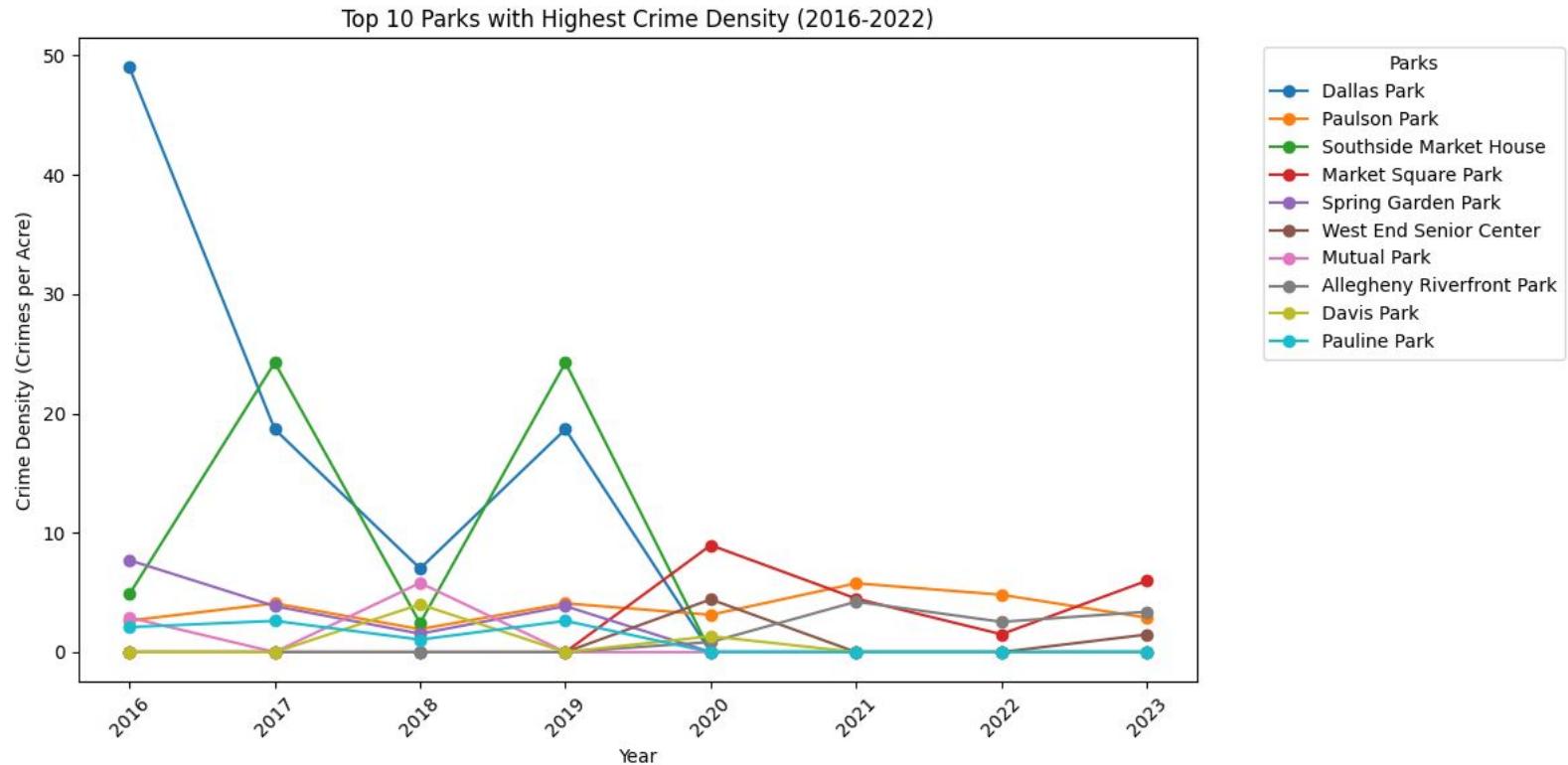
Top 10 Parks by Non-Violent Crime Density



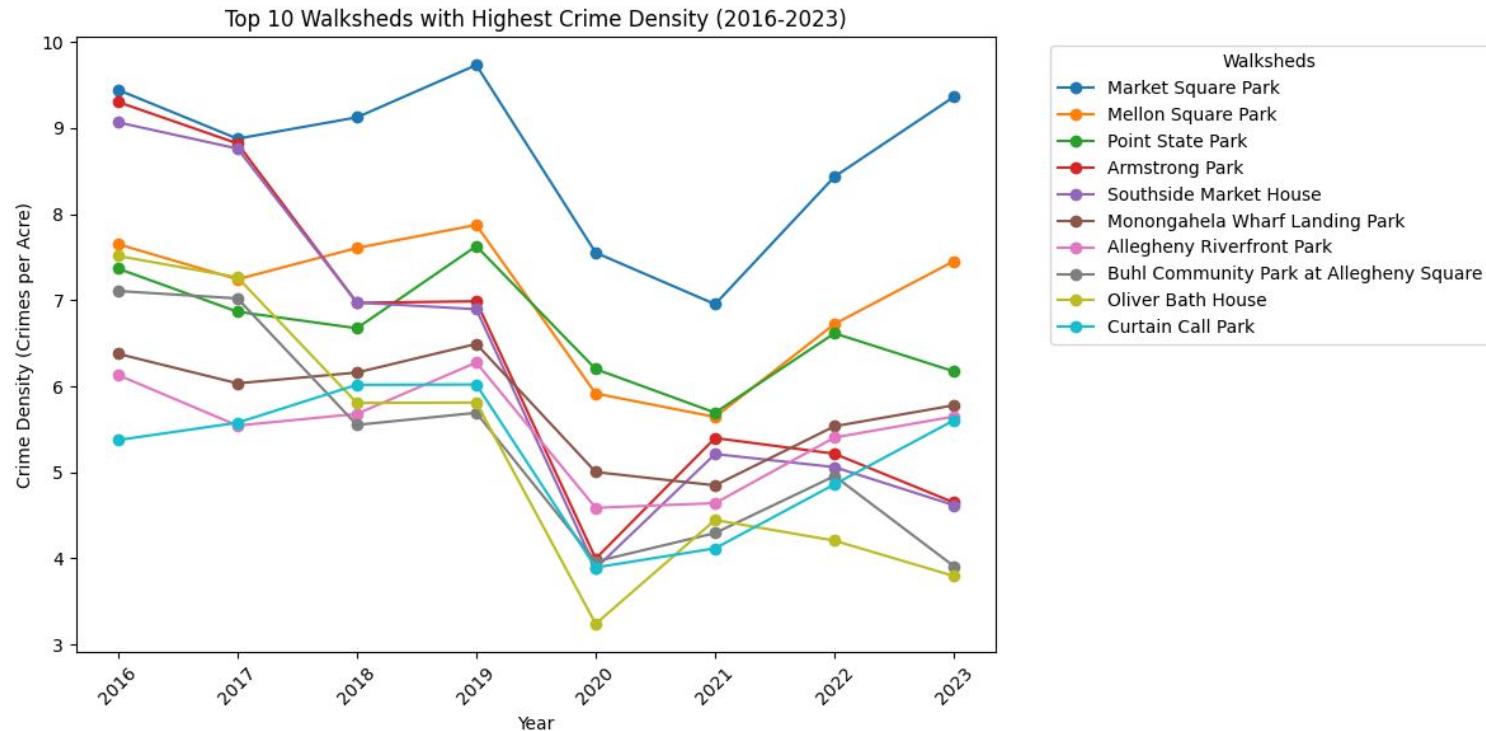
Crime Walkshed Data



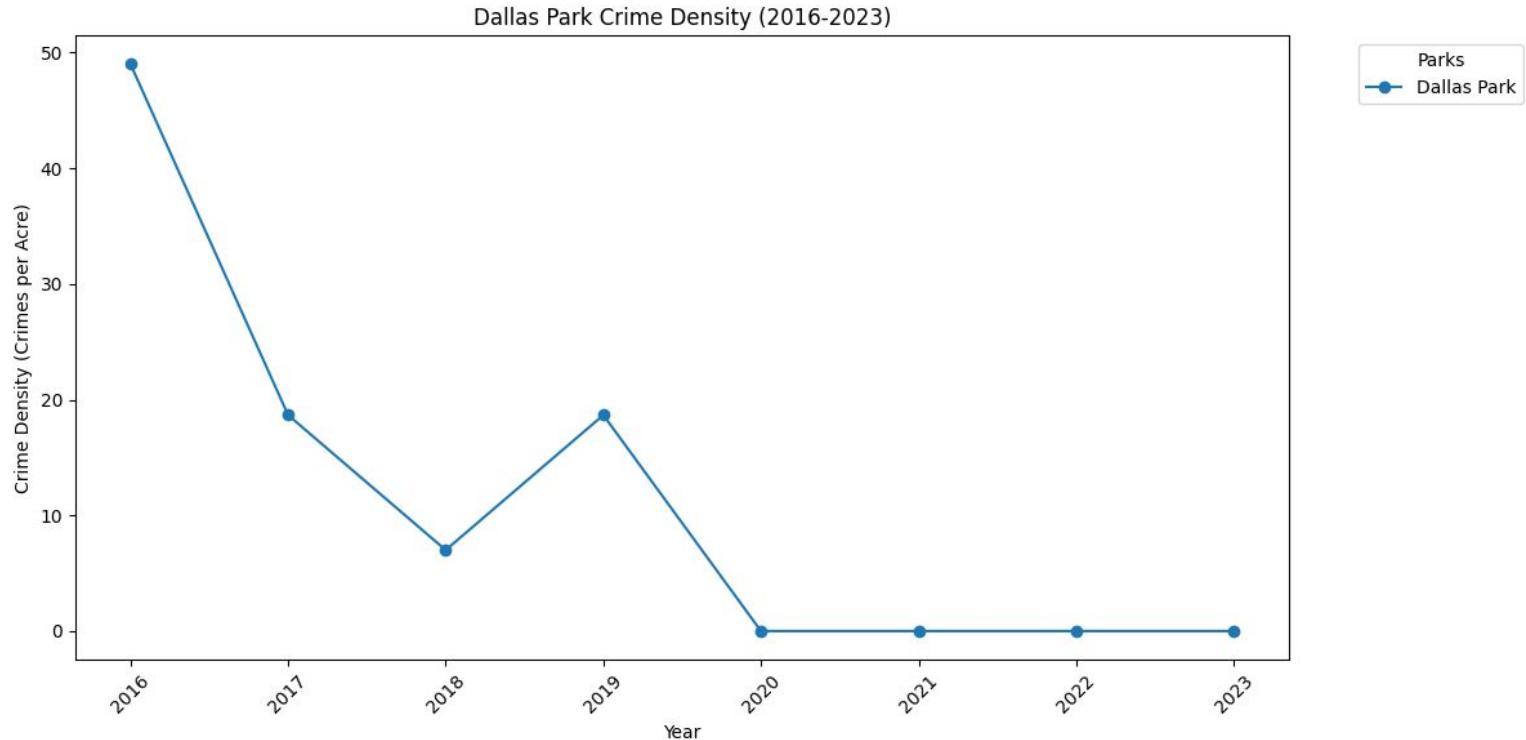
Crime Over Time (Parks)



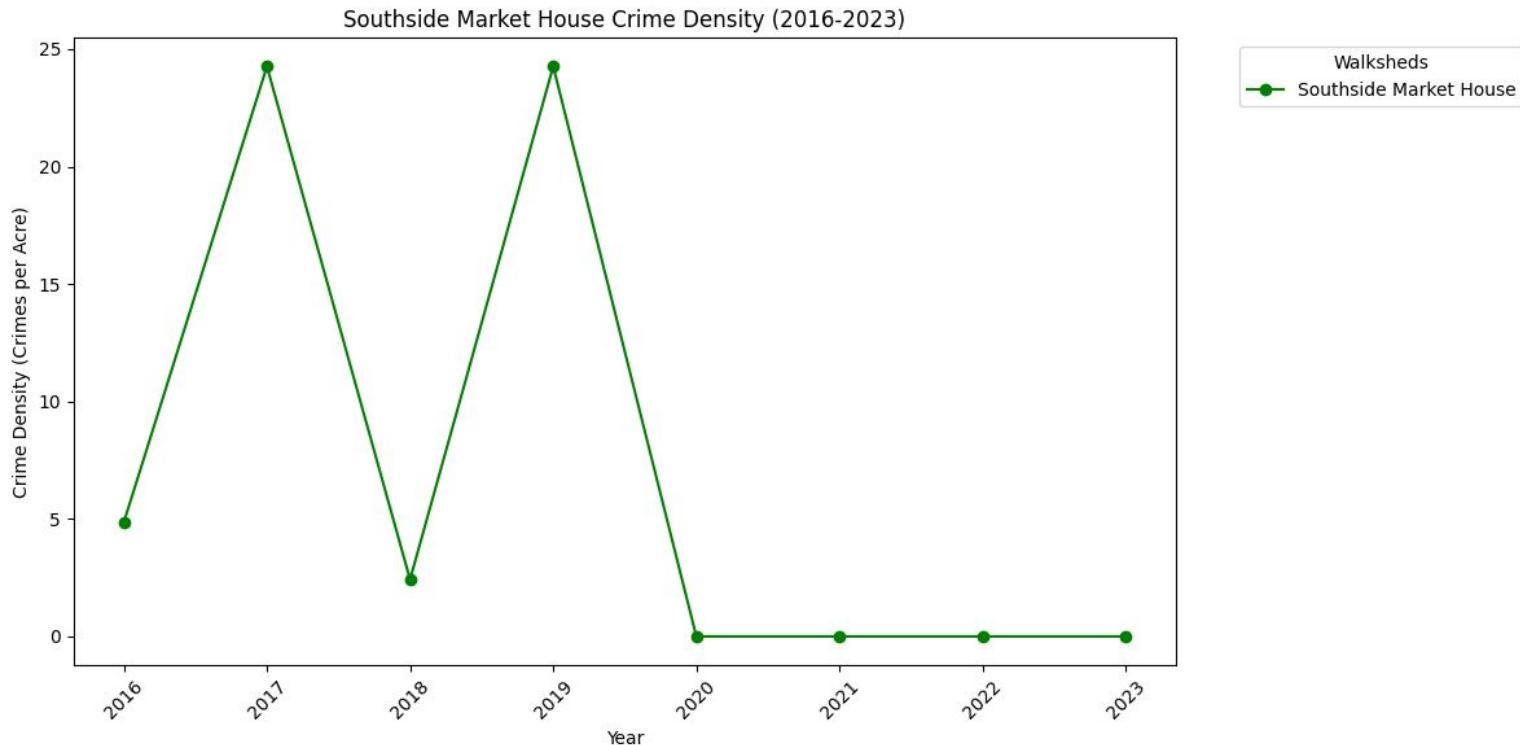
Crime Over Time (Walksheds)



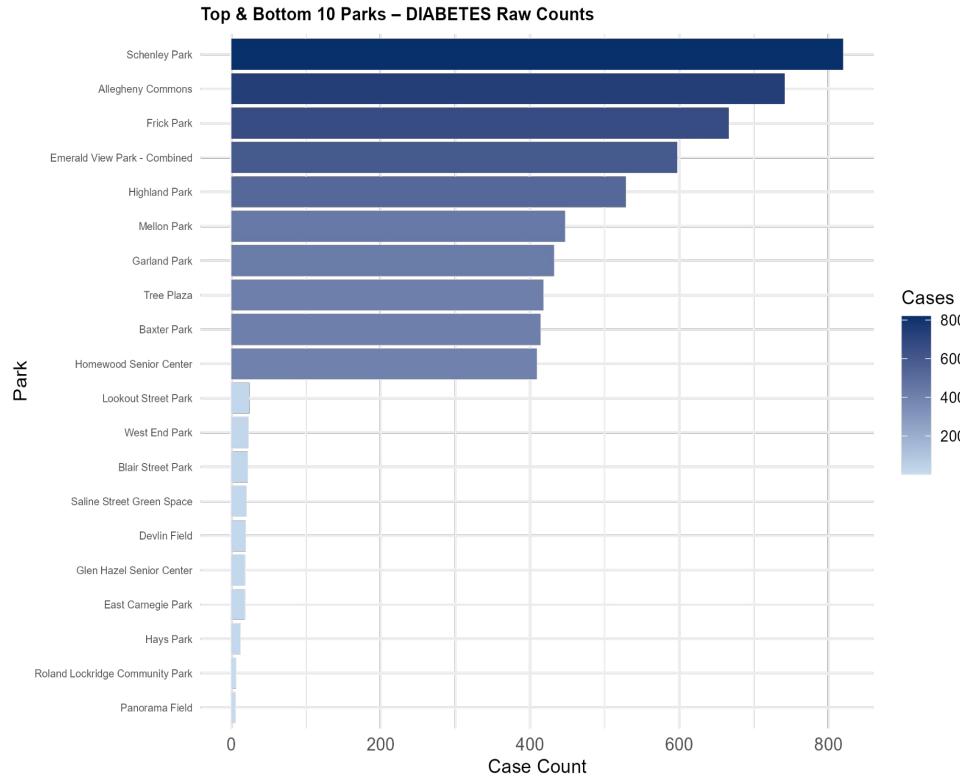
Crime Over Time (Dallas Park)



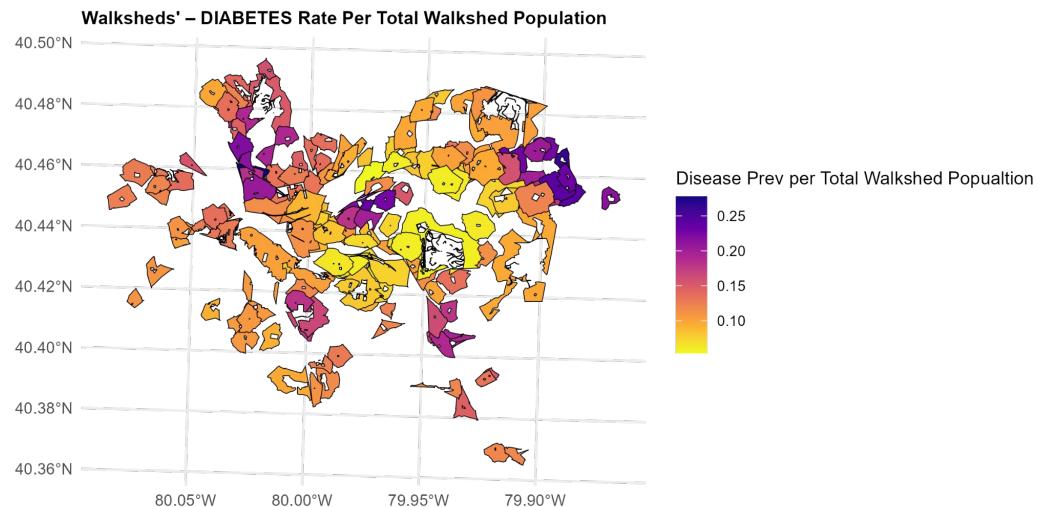
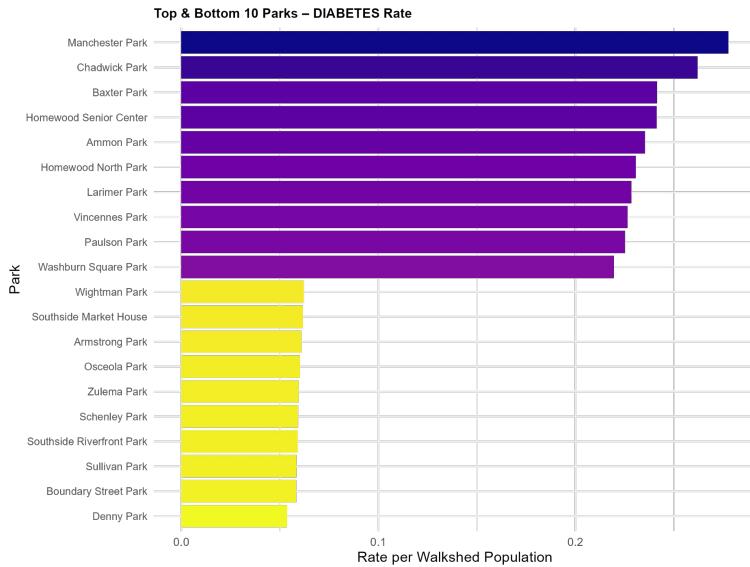
Crime Over Time (Southside Market House)



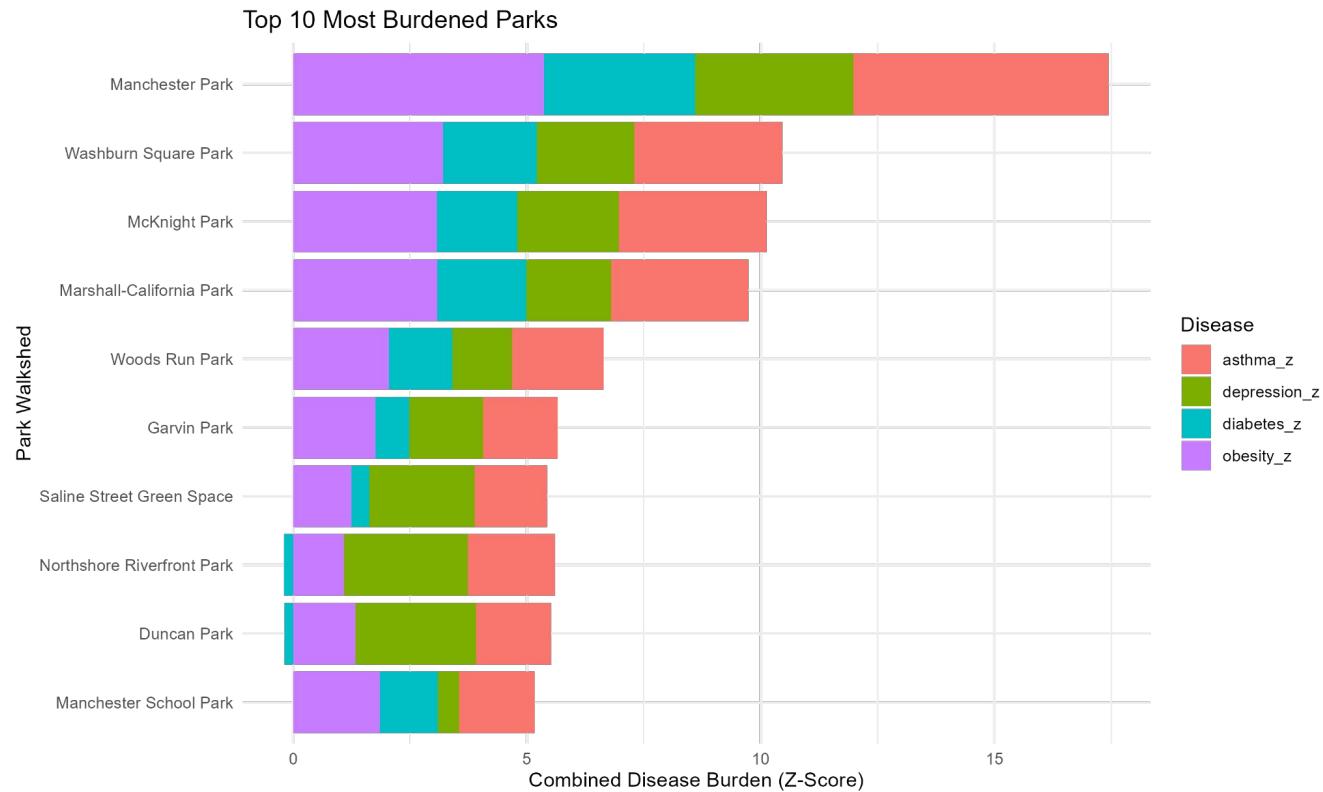
Health Data Preparation



Health



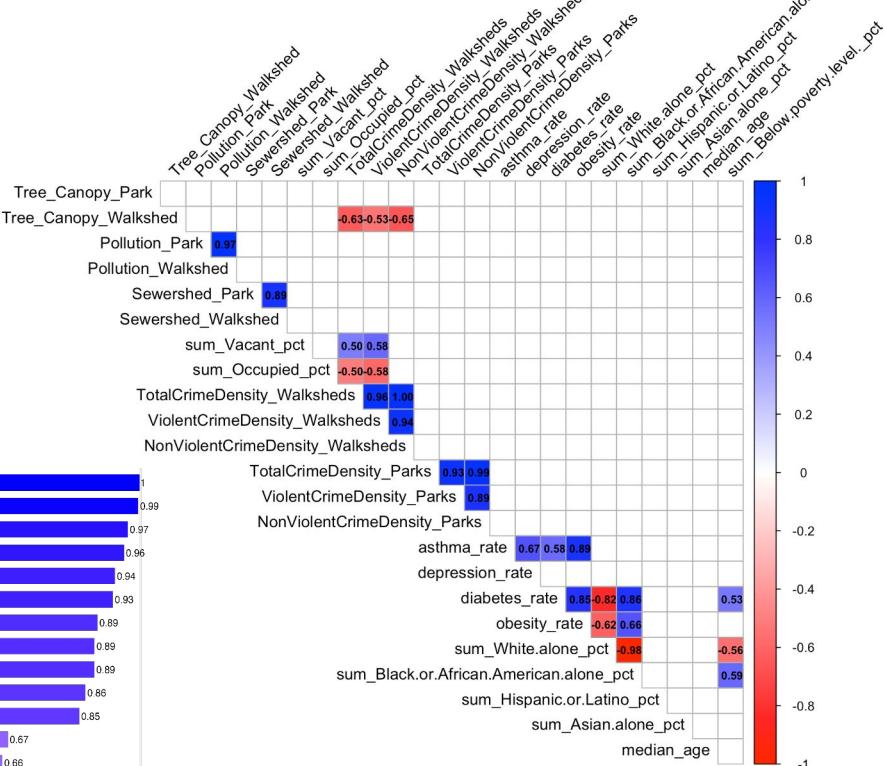
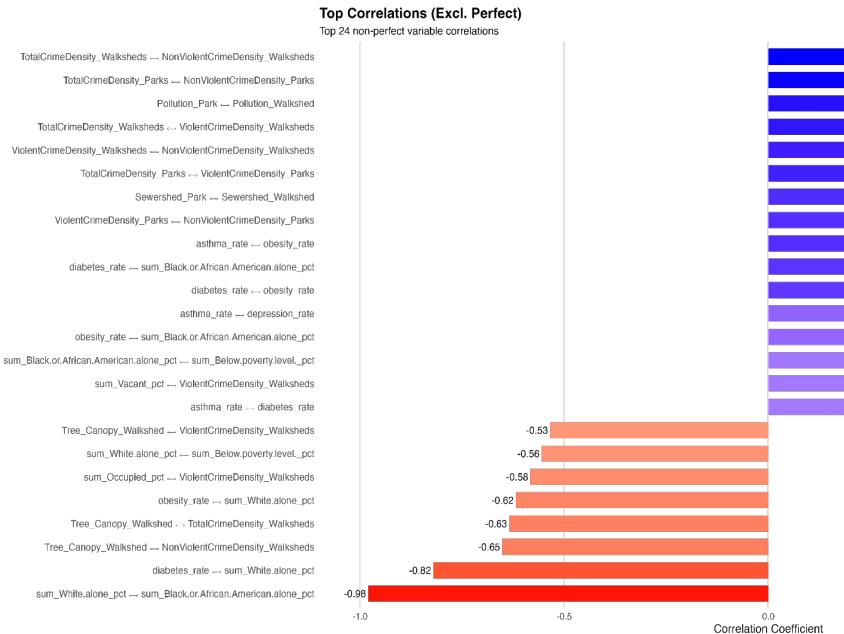
Health



Correlations

Notable relationships:

- Asthma, depression, obesity, and depression
- Black residents proportion and obesity, diabetes, and poverty
- Vacancy and violent crime
- Walkshed tree canopy and crime (negative)
- White residents proportion and poverty, diabetes (negative)



Folders



Files



Deliverable: The document, executive summary, raw and processed datasets, visualizations, R code, and spatial data are available in a shared Google Drive folder, organized in subfolders based on the four data categories, along with a master dataset named **Parks_Master.csv**

This file contains population- and size-scaled information for relevant variables across all parks and walksheds. Description of column names in document appendix along with a full bibliography in APA format

Conclusion

A previous report from Interface included an index that attempted to objectively calculate and rank each park's investment need. **We didn't do this**; instead, we sought to provide the necessary cleaned and scaled data in the master dataset to allow the Conservancy, if they so desire, to construct it themselves

The weighting of different variables is ultimately a **policy decision** – a value-based, subjective judgment reserved for stakeholders, not analysts

Many of these analyses illuminate **systemic disparities** across Pittsburgh's communities, often along racial, economic, and geographic lines, that mirror patterns of **historic segregation and disinvestment**. We hope that the Conservancy can utilize this project to augment their strategic decision-making with data-driven robustness and help forge a more **equitable, sustainable, safe, and healthy** future for Pittsburgh's public spaces