II. Data and Methodology → B. Environment

For our analysis of the parks' environmental data, we used three variables: tree canopy coverage, pollution, and sewershed priority. The tree canopy coverage data was sourced from the U.S. Department of Agriculture Forest Service's 2021 Conterminous United States (CONUS) Tree Canopy Cover (TCC) dataset. This report uses satellite imagery to calculate the tree canopy coverage percentage of 30 × 30 meter areas across the Conterminous United States. The pollution data was sourced from a study of annual North American pm2.5 concentrations completed by Washington University of St. Louis. This particular dataset measures the pm2.5 concentrations in $\mu g/m^3$ within 0.01˚ × 0.01˚ areas in the year of 2022. The sewershed priority data was sourced from the 2024 updated dataset describing combined sewershed priority ratings completed by the Pittsburgh Water and Sewer Authority (PWSA). Priority of a sewershed is measured based on the need for intervention in mitigating overflowing of sewers and storm drains. Specific tracts of land within Pittsburgh's city limits were given one of three ratings by the PWSA: high priority, secondary priority, and low priority.

Each dataset was attached to individual parks and their walksheds through the use of geospatial analysis software. The tree canopy and pollution data were loaded as raster files, and therefore a technique called zonal averaging was used to calculate the average tree canopy coverage and pollution levels for each park and walkshed. The sewershed priority rating data was loaded as a vector layer, therefore areal interpolation was used to calculate an aggregated score of priority. Two additional variables were calculated for sewershed priority measuring the percentage of each park contained within a secondary priority and high priority sewershed boundary.

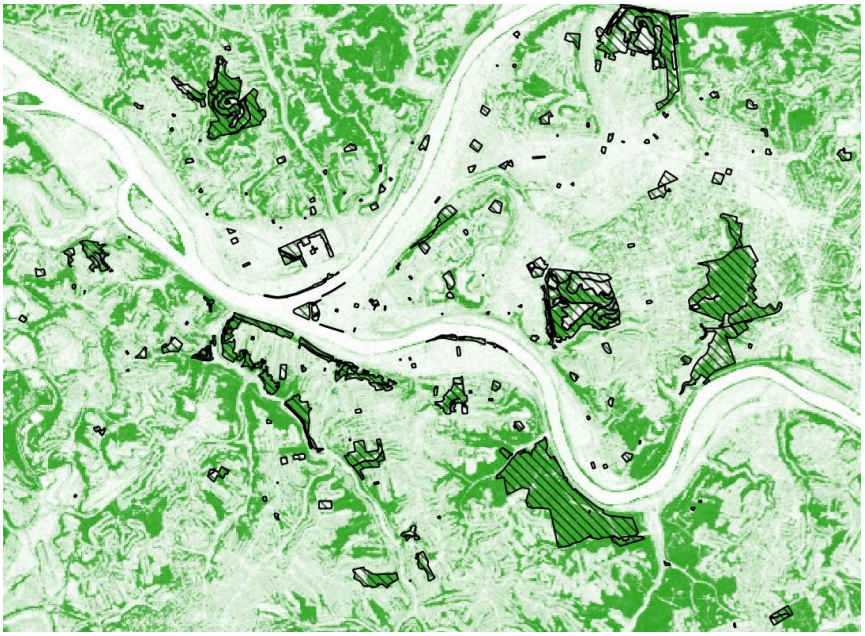
After we calculated tree canopy averages, pm2.5 concentration averages, and aggregated sewershed priority rankings for each park and walkshed, we created additional z-score variables for each. These z-scores variables represent a more standardized form of the data and measure the amount of standard deviations above or below the mean each park and walkshed is for the respective variable.

Using Python packages including seaborn and matplotlib, we visualized the environmental data using bar graphs, scatterplots, and other methods to highlight specific parks with a need for environmental intervention.

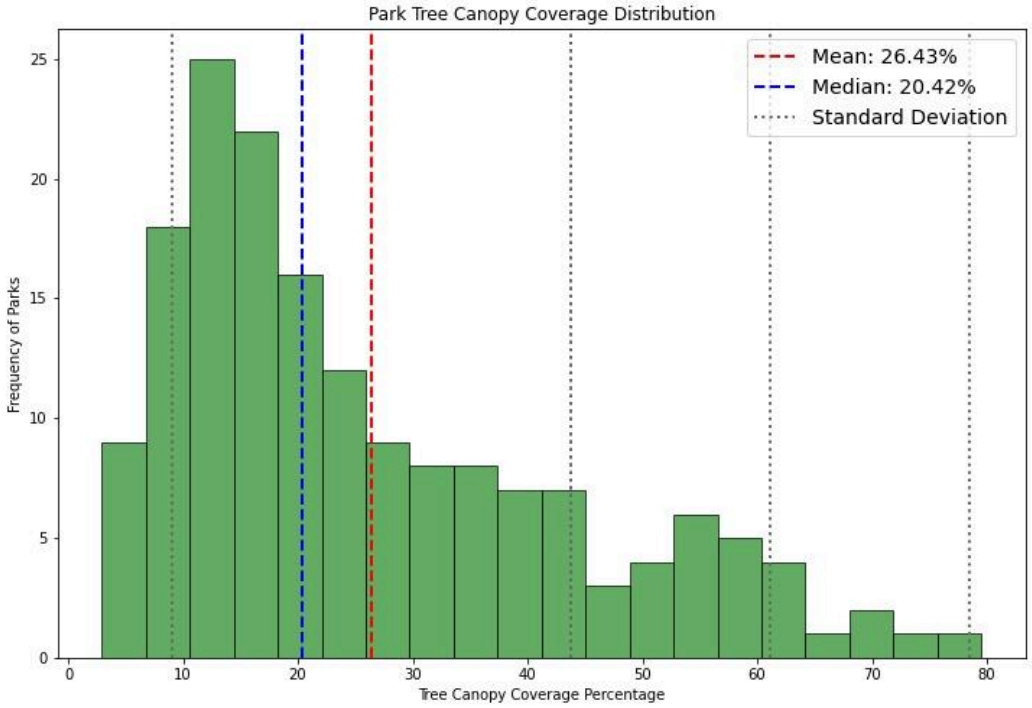
II. Data and Methodology → B. Environment → 1. Tree Canopy

The technique of zonal averaging was used to calculate the average percentage of tree canopy coverage within each park. This technique involves averaging all of the tree canopy coverage percentages (each being calculated in 30 × 30 meter areas) contained within each park and walkshed boundary. The variables of "Tree_Canopy_Park" and "Tree_Canopy_Walkshed" represent average percentages and serve as an accurate indicator for tree canopy distribution among the entire park or walkshed.

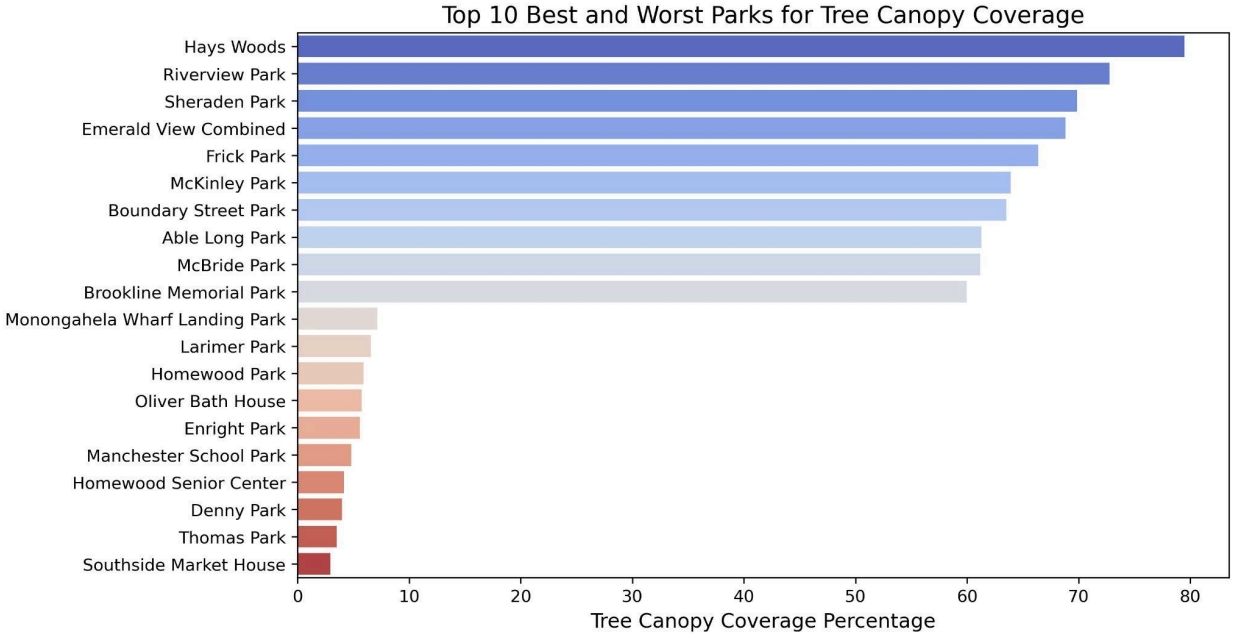
Tree Canopy Coverage of Pittsburgh



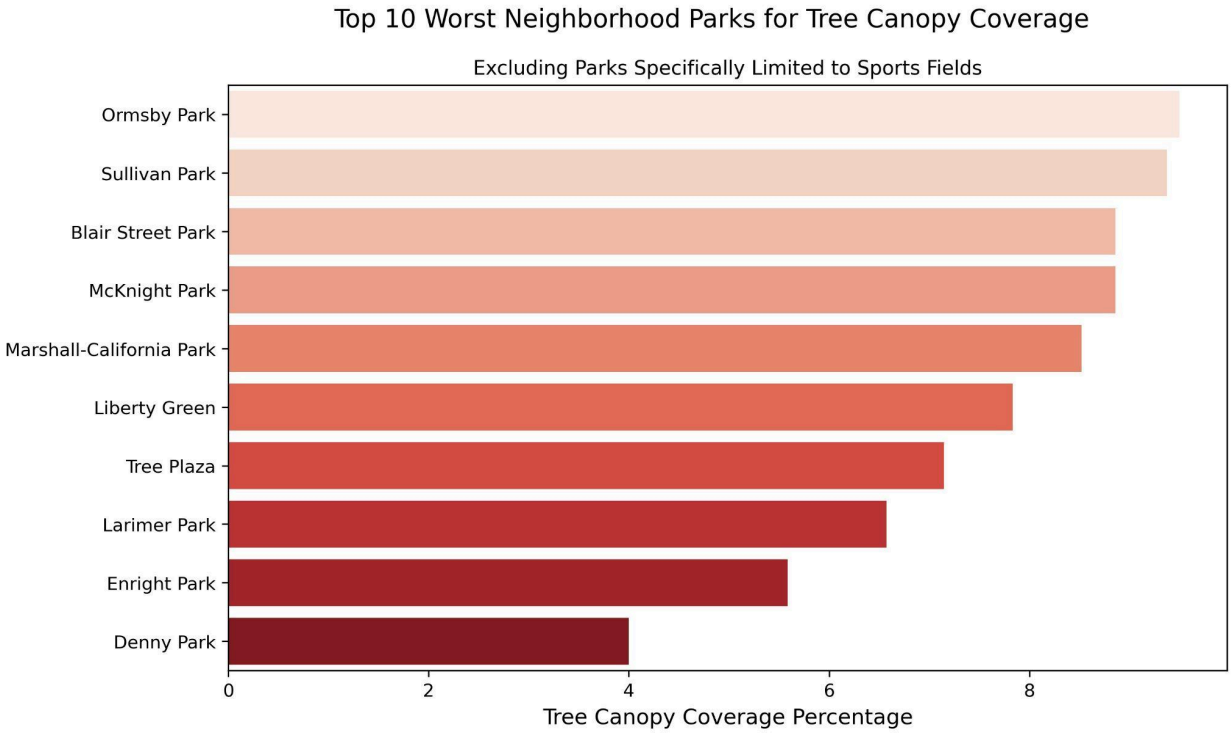
This graphic showcases each Pittsburgh park layered on top of a raster layer sourced from the U.S. Department of Agriculture Forest Service's calculation of tree canopy coverage percentages in the Conterminous United States. Areas with darker green colors represent a higher percentage of tree canopy coverage. Through calculating the zonal tree canopy coverage averages of each park, we were able to compute the mean and median tree canopy coverage for each park in Pittsburgh and rank each park from most to least tree canopy coverage. The mean and median tree canopy coverage percentage across all parks within Pittsburgh is 26.43% and 20.42% respectively. This indicates a right skew in the distribution across all parks and is showcased in the histogram below.



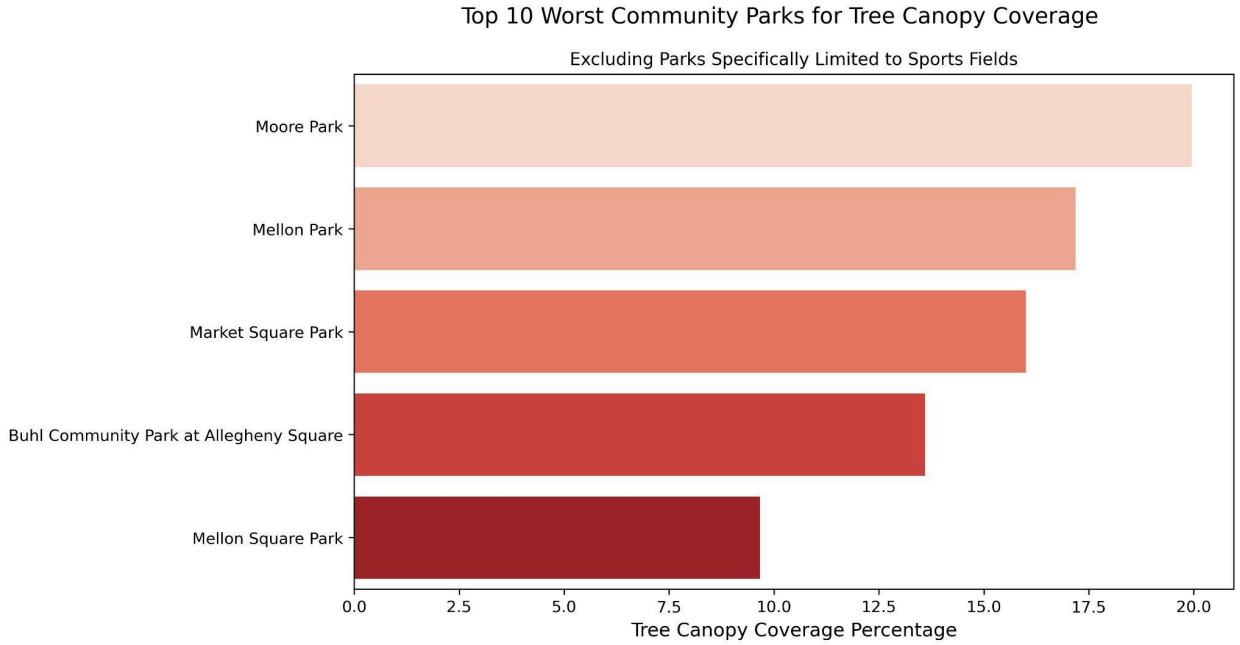
Parks in Pittsburgh significantly vary in their tree canopy coverage with a standard deviation of 17.37% and a range of 76.57%. There is a high concentration of parks below the mean, especially those with around 10%-20% coverage, and a less concentrated and more dispersed collection of parks that exceed the mean. Below is a bar graph that showcases the top 10 best and worst parks for tree canopy coverage.



On the high side of the extreme is Hays Woods, Riverview Park, and Sheraden Park. Parks on the extreme low end include Southside Market House, Thomas Park, and Denny Park. It is important to analyze the underlying reasons of why these parks are on the extreme ends of the dataset. For example, Southside Market House is a historic building that happens to be listed as a park and should have an expected tree canopy coverage of nearly zero. Additionally, parks that are specifically limited to sports fields, like Manchester School Park and Homewood Park, are expected to have low tree canopy. On the other side of the spectrum, Hays Woods is known to be an undeveloped tract of land that lies within city boundaries and it only became a park in 2023, and is therefore expected to have high tree canopy coverage. Three neighborhood parks in the bottom ten that are not buildings and are not specifically limited to sports fields are Denny Park, Enright Park, and Larimer Park. Below is a bar graph representing the bottom 10 parks for tree canopy coverage in parks categorized as neighborhood parks. Neighborhood parks tend to have less tree canopy coverage on average than community parks. All parks specifically limited to sports fields were manually excluded from this bar graph. This includes parks where the entirety of its area comprises a sports field and parking lot. No intervention on tree canopy would be viable in these parks which include Manchester Field and Homewood Park.



While community parks have a higher tree canopy coverage on average than neighborhood parks, some of these parks still struggle with low tree canopy. Below is a bar graph representing the bottom five community parks for tree canopy coverage.



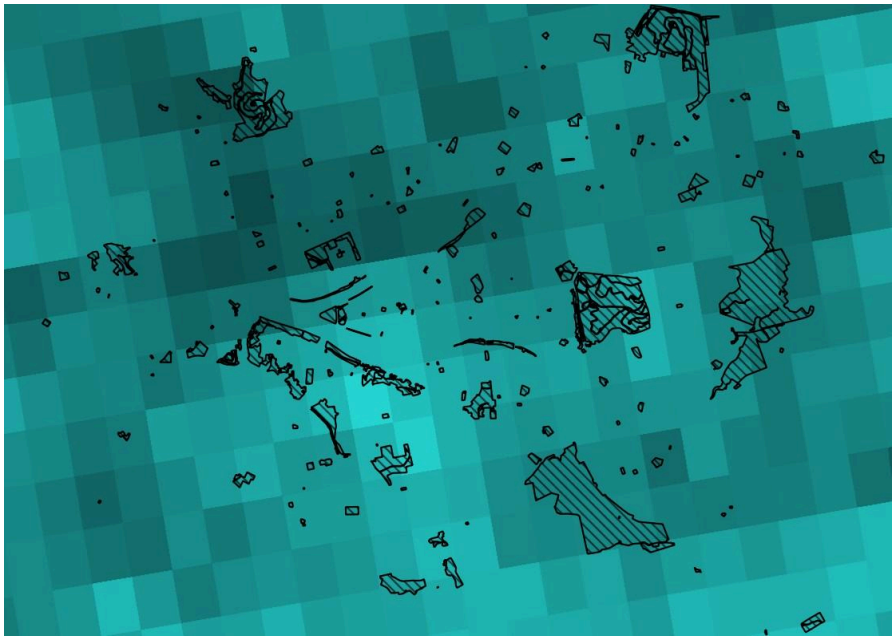
A point of notice is that park acreage is moderately positively correlated with park tree canopy coverage. Smaller parks tend to have lower tree canopy coverage while larger parks have higher tree canopy coverage. This relationship will be explored further in the analysis section of this report. Additionally, relationships between park tree canopy coverage and other factors like depression will be analyzed.

Tree canopy coverage percentages were also calculated for each walkshed. This allowed us to see the differences between a park's tree canopy and its surrounding area's tree canopy. The same method of zonal averaging was used to calculate each walkshed's tree canopy coverage percentage.

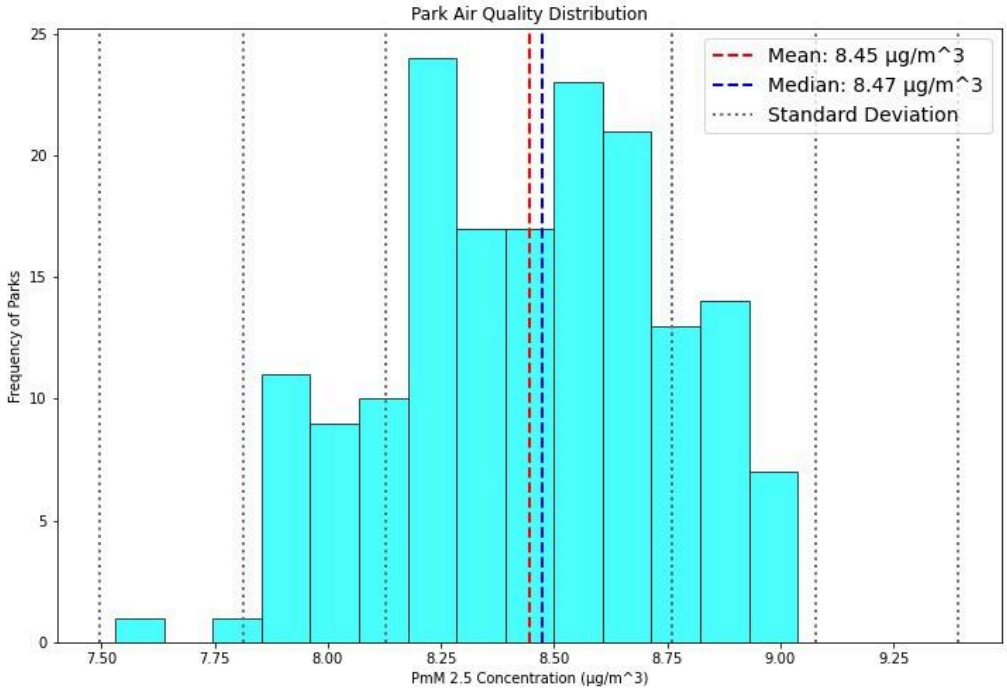
II. Data and Methodology → B. Environment → 2. Pollution

Pollution of a park in this report is measured by the average pm2.5 concentration measured in micrograms per cubic meter (μg/m^3). The technique of zonal averaging was used to calculate the average pollution within each park. This technique involves averaging all of the pm2.5 concentration levels (each being calculated in 0.01° × 0.01° or ~1.11 × .844 km areas) contained within each park and walkshed boundary. The variables of "Pollution_Park" and "Pollution_Walkshed" represent average pm2.5 concentrations and serve as an accurate indicator for pollution levels across the entire park or walkshed.

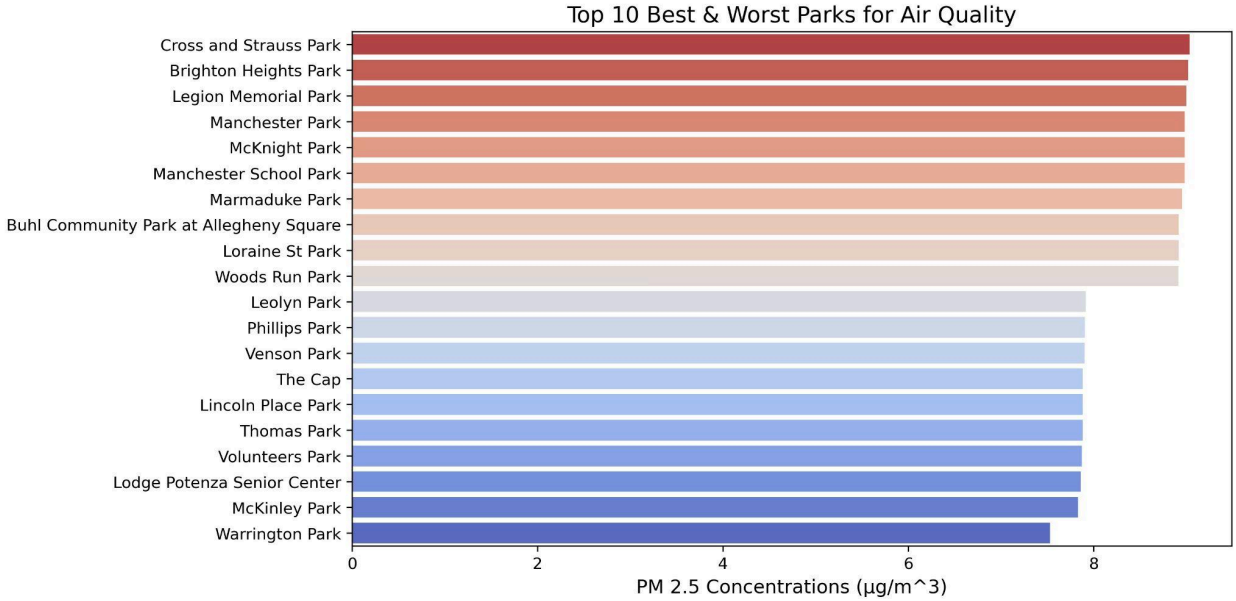
PM2.5 Concentration Levels of Pittsburgh



This graphic showcases each Pittsburgh park layered on top of a raster layer sourced from the Washington University of St. Louis' calculation of pm2.5 concentrations in North America. Areas with darker blue colors represent a higher concentration of pollution. Through calculating the zonal pollution averages of each park, we were able to compute the mean and median pollution levels for each park in Pittsburgh and rank each park from best to worst in terms of their air quality. The mean and median pm2.5 concentration levels across all parks within Pittsburgh is 8.45 μg/m^3 and 8.47 μg/m^3 respectively. This indicates a relatively normal distribution across all parks and is showcased in the histogram below.



Parks in Pittsburgh vary slightly in their pollution levels with a standard deviation of .31 µg/m^3 and a range of 1.51 µg/m^3. All parks are within two standard deviations of the mean with the exception of Warrington Park. Warrington Park is a statistical outlier in low pollution levels. As shown in the raster layer graphic, the area of Mt. Washington has the least amount of pollution and therefore parks close-by like Warrington Park, Lodge Potenza Senior Center, Venson Park, and Mckinley Park are among the parks with the best air quality. Below is a bar graph that showcases the top 10 best and worst parks for air quality.



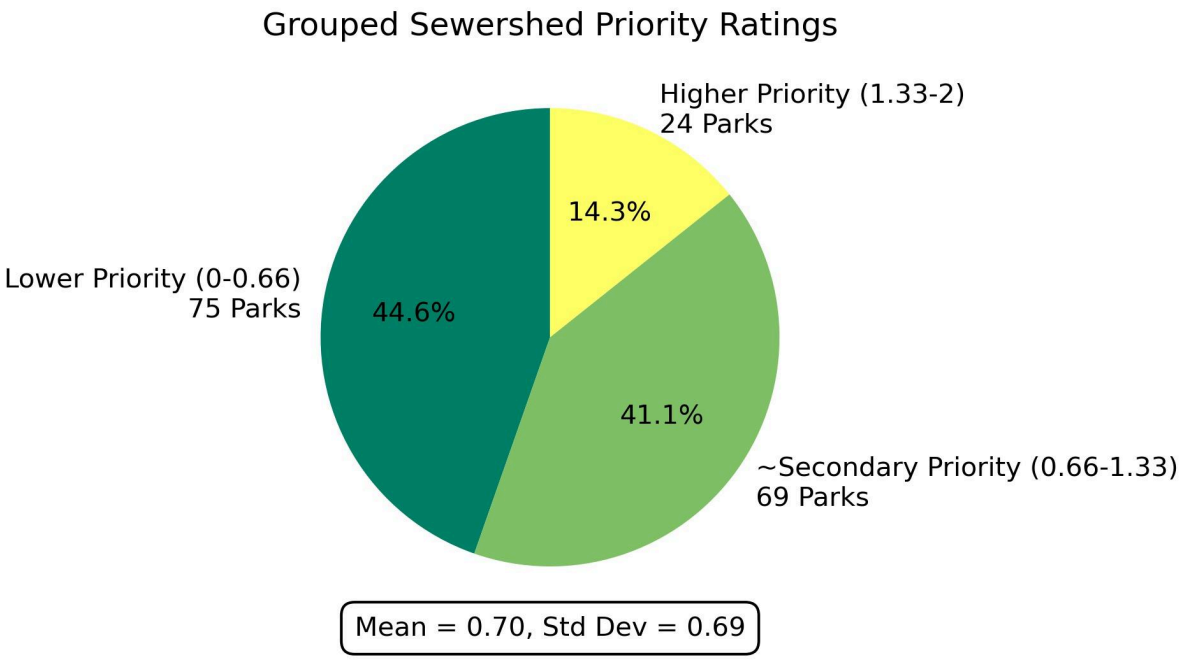
As shown in the raster layer graphic, parks on the North Side are within the area with the worst pollution levels in Pittsburgh's city limits. All ten of the parks with the worst air quality are located north of the Allegheny River.

Pittsburgh as a whole is a city that is below average in terms of their air quality. As of February 7th, 2024, the United States Environmental Protection Agency has set "the level of the primary (health-based) annual $PM_{2.5}$ standard at 9.0 micrograms per cubic meter to provide increased public health protection, consistent with the available health science." The top three parks for worst air quality in Pittsburgh barely exceed this new standard of 9.0 $\mu g/m^3$.

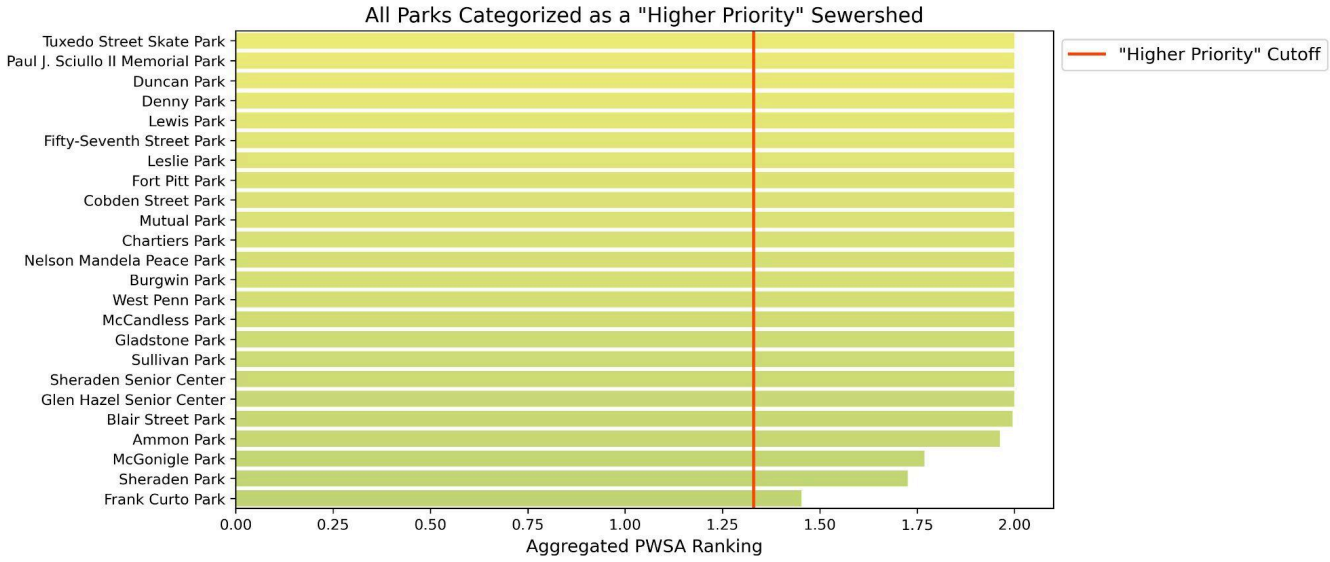
Pollution levels were also calculated for each walkshed. This allowed us to see the differences between a park's pollution level and its surrounding area's pollution level. The same method of zonal averaging was used to calculate each walkshed's pm2.5 concentration average.

II. Data and Methodology → B. Environment → 3. Sewershed Priority

The technique of aerial interpolation was used to calculate the aggregated sewershed priority score within each park. The Pittsburgh Water and Sewer Authority (PWSA) has rated tracts of land within Pittsburgh based on their sewershed's need for intervention. The PWSA assigns a 2 to areas with high priority, a 1 to areas with secondary priority, and a 0 to areas with low priority. Through aerial interpolation, sewershed priority scores are attached to the parks and are weighted depending on how much of the park intersects with the boundaries created by the PWSA. The resulting weighted average of all intersections represent the final sewershed priority score, with the minimum of 0 being low priority and the maximum of 2 being high priority. Below is a pie chart representing three groups of sewershed ratings. Parks with an aggregated priority score of 0-.66 are labelled as "lower priority," .66-1.33 are labelled as "~secondary priority," and 1.33-2 are "higher priority."



The mean of .70 reveals that parks on average in Pittsburgh are located within areas with lower priority ratings. However, there are 24 parks, making up 14.3% of all parks, that exceed an aggregated score of 1.33 and can be labelled as "higher priority." Below is a bar graph that showcases all of the parks that fall within this category of "higher priority."



Among the 24 parks that are categorized as “higher priority,” 19 of them have a priority rating of exactly 2. This means they are located fully within a PWSA sewershed boundary with the “high priority” rating. 5 of the parks have a rating greater than 1.33 but less than 2. This means the park intersects with a “high priority” boundary and one of the lower rated sewershed priority boundaries. However, all of these parks intersect the most with the “high priority” boundaries.

Two additional variables were calculated for explaining the parks’ sewershed priorities. These two variables, labelled as “percent_rank1” and “percent_rank2,” measure the percentage of the area of the park that exists within the sewershed priority boundaries of “secondary priority” and “high priority” respectively. These variables are especially useful in interpreting the aggregated priority rating. They show exactly what comprises each rating for a park by giving exact percentages of the parks’ intersected areas with all three types of sewershed rating boundaries.