Bristow Richards & Sara Khoshhal

Geographic Information Systems

Professor Kristen Kurland

12/09/2022

Smell(y) Pittsburgh:

Analysis of Crowd-Sourced Air Quality Data in Pittsburgh

**Introduction**

According to the American Lung Association, Pittsburgh has some of the worst air pollution in the country, ranking 9th worst in the country for long-term particle pollution and receiving an F rating in ozone and particulate matter [1]. There is a wealth of literature in the field of environmental justice that links race and the burden of environmental hazards such as hazardous waste sites or industrial plants [2]. In this report and associated [StoryMaps](https://arcg.is/1b05CH), we investigate data from an open-source community reporting tool to uncover potential links between environmental hazards and underprivileged communities in Pittsburgh, Pennsylvania.

We use the open-source community environmental reporting tool titled Smell Pittsburgh to analyze trends in air pollution and environmental justice in Pittsburgh [3]. This project allows users to create “smell reports” that rate the severity of air quality issues on a scale of 1-5, in addition to the reporter’s geographic coordinates and any qualitative notes the user would like to submit [4]. The data from these reports is anonymized by slightly skewing the coordinates of each user when they submit smell reports. This project helps to capture instances of industrial air pollution that affect residents near polluters and, sometimes, the entire Pittsburgh metropolitan region.

**Problem statement**

Are there any observable trends in the areas that report a high volume of poor odors, specifically pertaining to the demographic or geographic composition of these areas? What areas see the most reporting, and why?

**Approach/ Methodology**

The project aimed to reveal trends in areas with many odor reports and high proportions of severe odor reports. To do so, we used an R script which calls the Smell Pittsburgh API to load the data into memory. In our script, we query Smell Pittsburgh’s global database of community reporting air quality reports, filtering for the Pittsburgh metro region. We then remove text fields from the user and re-categorize the reports by a three-level factor of severity: low for 1 and 2, medium for 3, and high for 4 and 5.

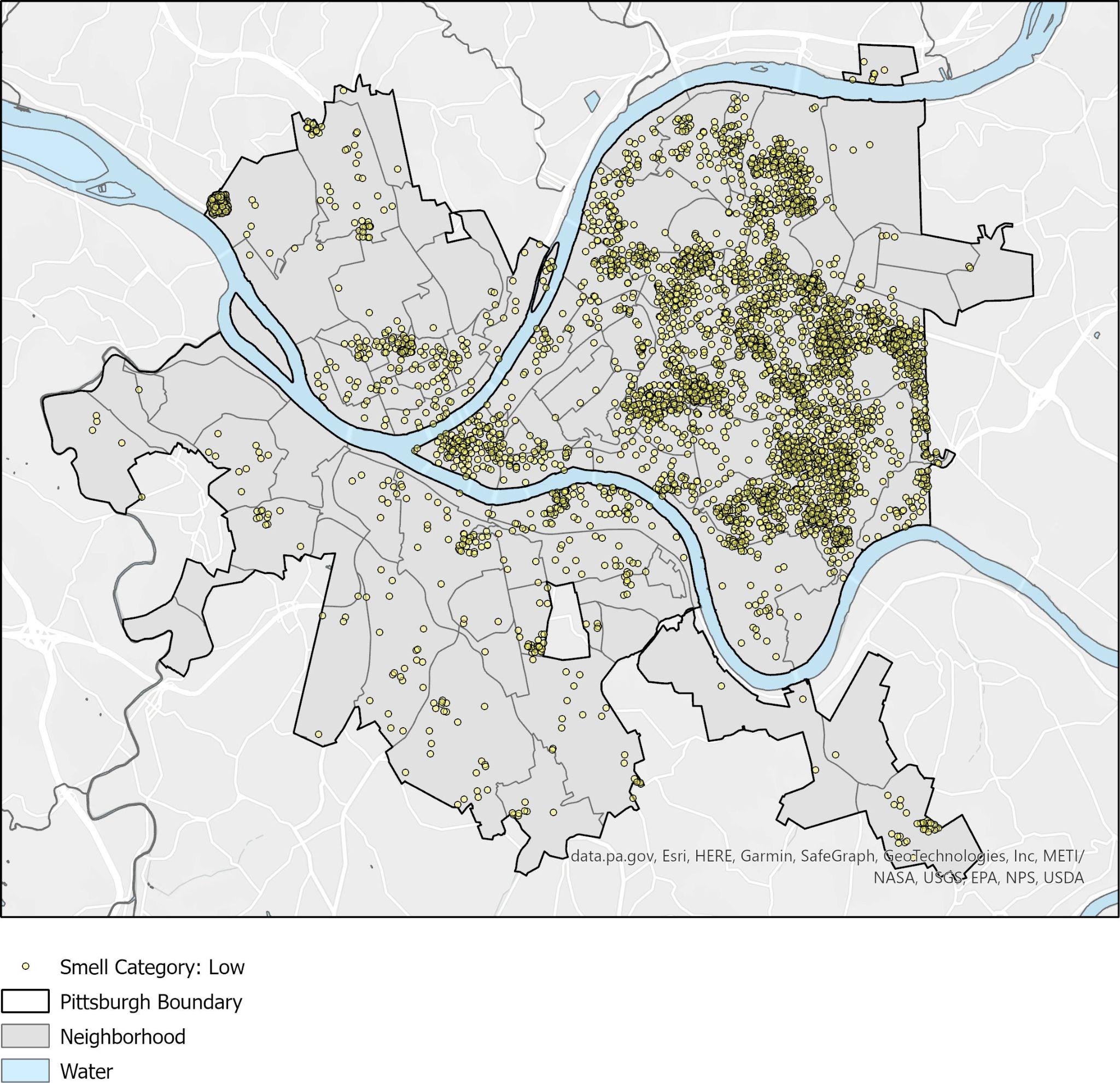
This data is saved and loaded into GIS alongside data from the City of Pittsburgh GIS portal on Neighborhood economic and demographic data [data source 1]. We eliminated points falling outside the city of Pittsburgh, then split the layer of points into three distinct layers for each unique severity value. We aggregated these points into each neighborhood polygon to find each neighborhoods count and percent of each severity level as well as the total count of reports per neighborhood. These counts were displayed as choropleth leyers alongside a choropleth layer representing area median income (AMI) of each neighborhood. Finally, we created a new neighborhood centroid layer which symbolizes the percent of reports in that neighborhood that were high severity. This point layer helps visualize the relationship that sever odor has with neighborhood reporting totals and neighborhood AMI. These layers were exported to a dashboard layout for users to visualize and explore the indicators of smell reporting totals, smell severity proportions, and income for each neighborhood in Pittsburgh.

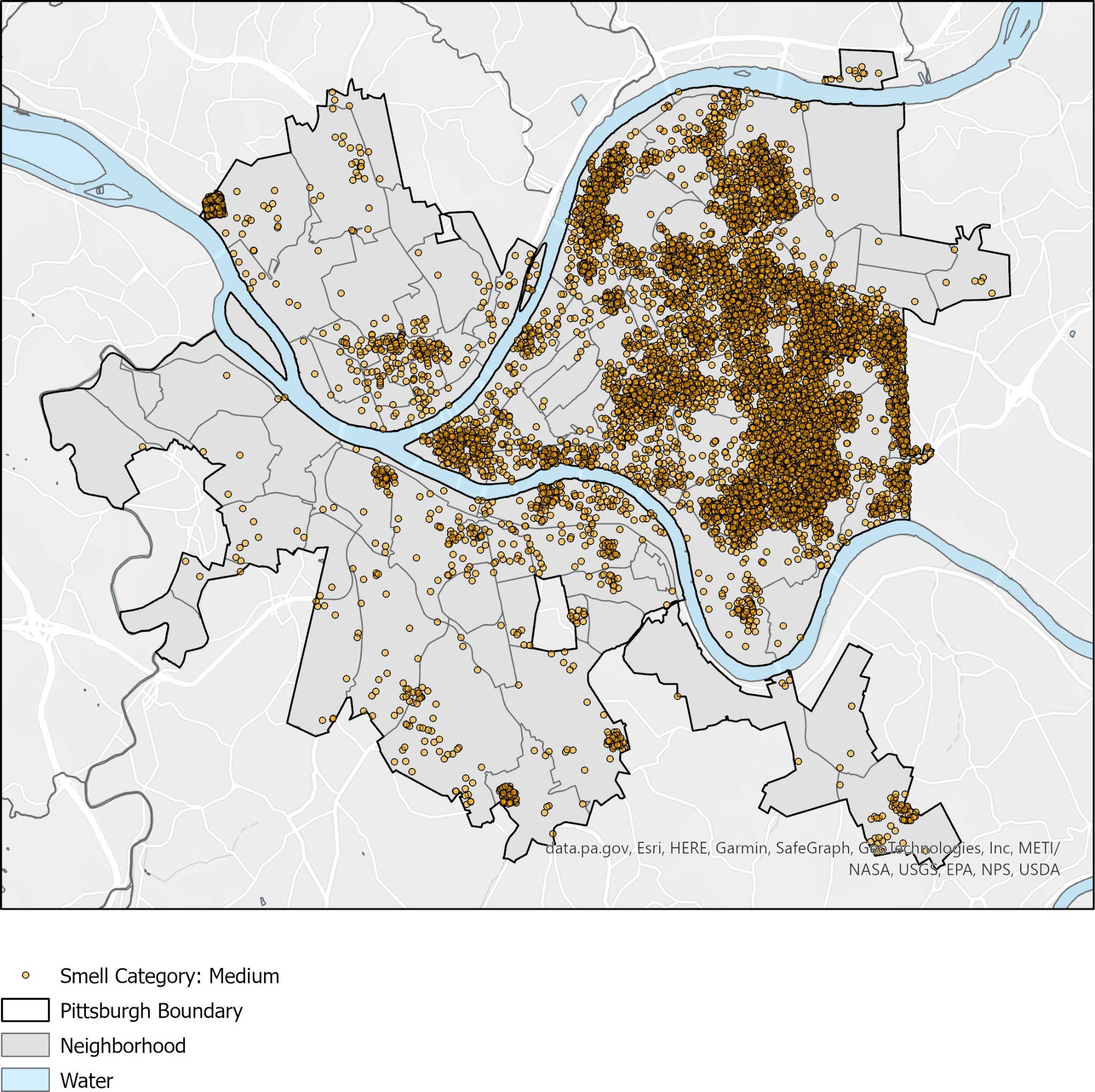
**Results, Solutions, and Findings**

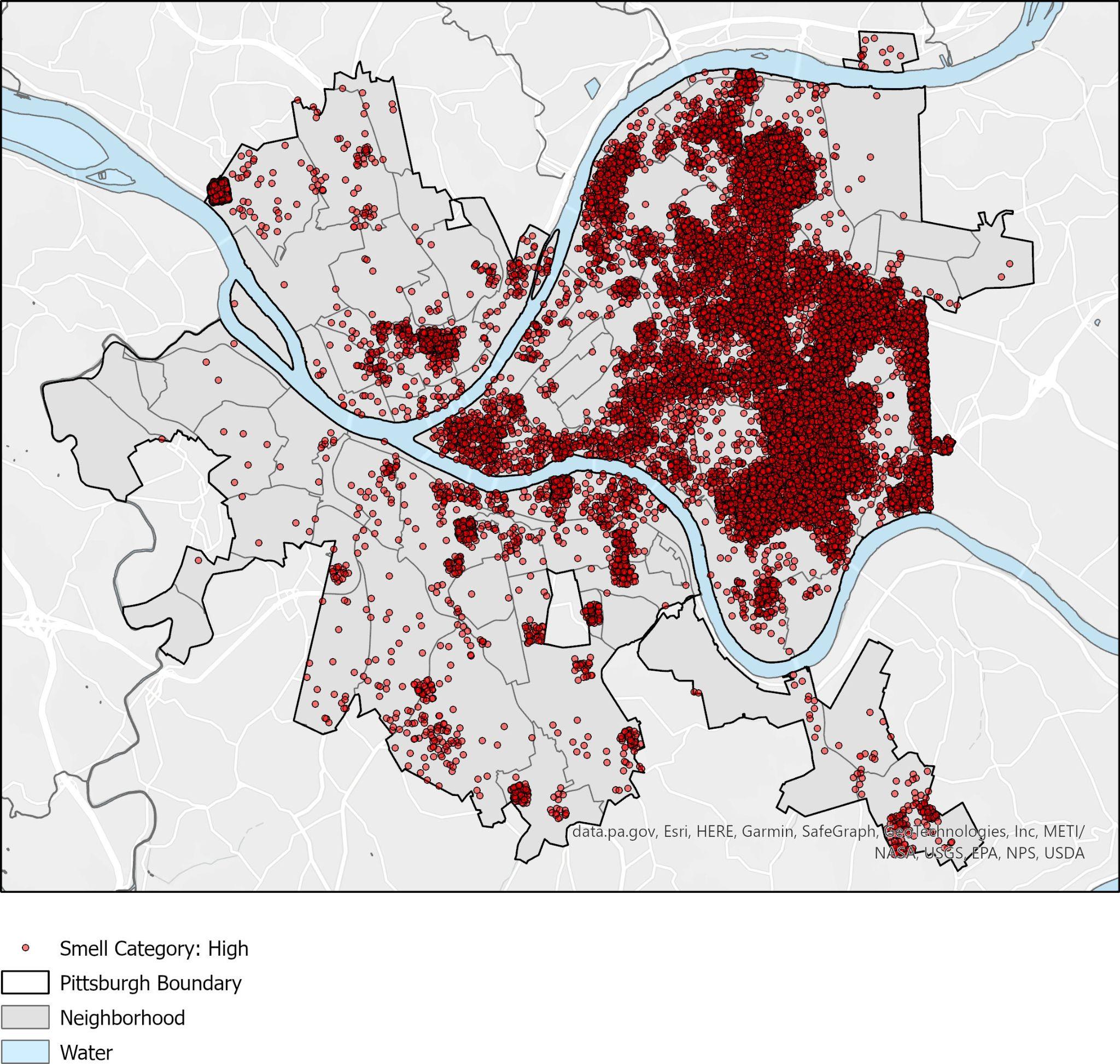
Location of Smell Reports and Their Severity:

In order to better visualize which areas in Pittsburgh have the highest concentration of smell reports, the following maps were made. Each point indicates a smell report, with the yellow points showing low severity smells (Figure 1), the orange points indicating medium severity smells (Figure 2), and the red points indicating high severity smells (Figure 3). These points were plotted in different maps due to the large concentration of reports, but you can see that there seems to be higher volumes of reports of high severity smells, in comparison to the other two categories. This may be due to the fact that it is easier to report anomalies in smells, as no smell/ fine smell is the baseline or norm. This leaves people less likely to report these instances.

Additionally, we can see that reports of all variations/ severities seem to be concentrated in the same areas, which suggests that this is a better indicator of general areas where reporting is heavily concentrated than a visualization of problematic areas.

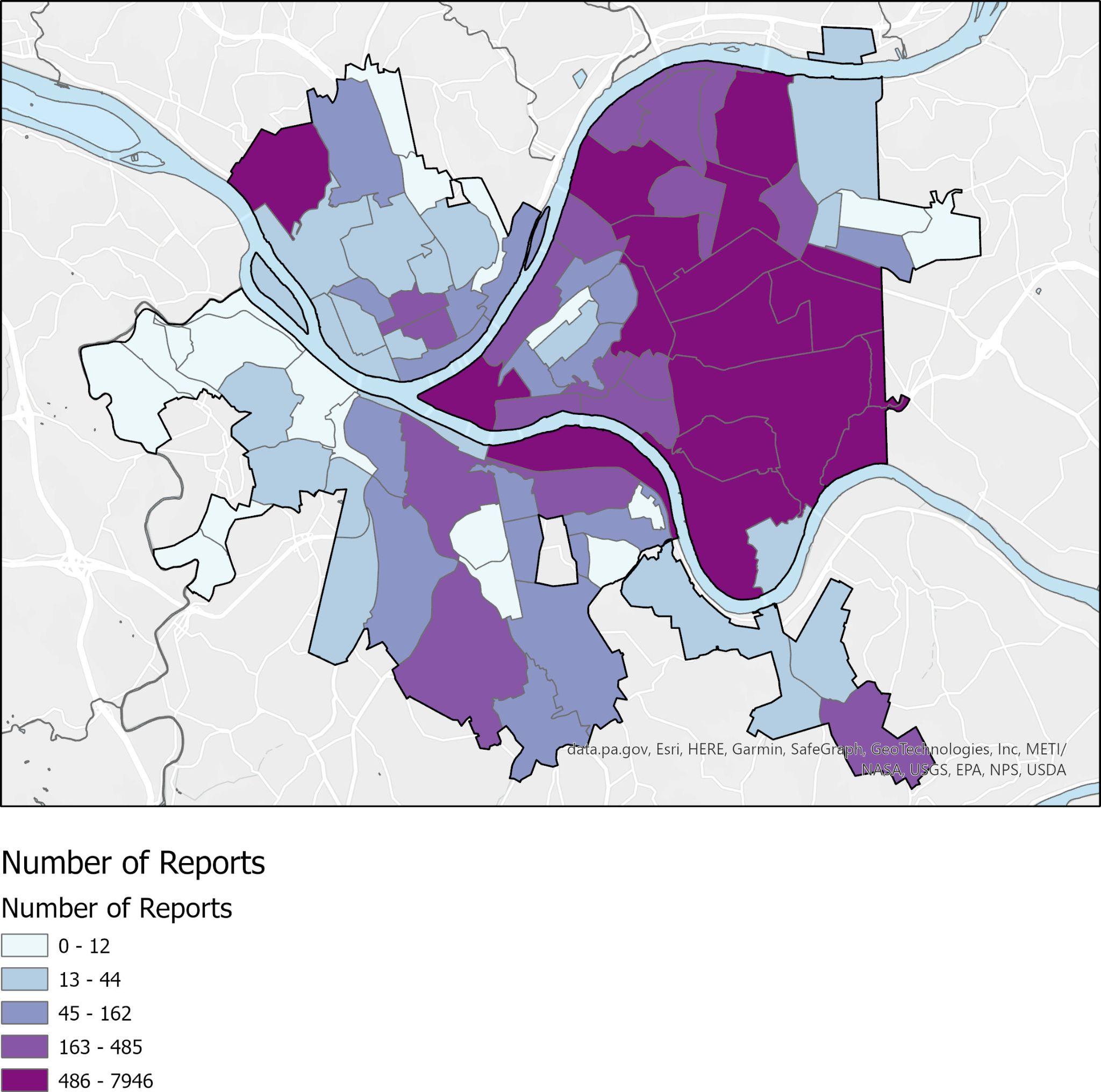
**Figure 1: Reports of Low Severity Smells in Pittsburgh**

**Figure 2: Reports of Medium Severity Smells in Pittsburgh**

**Figure 3: Reports of High Severity Smells in Pittsburgh**

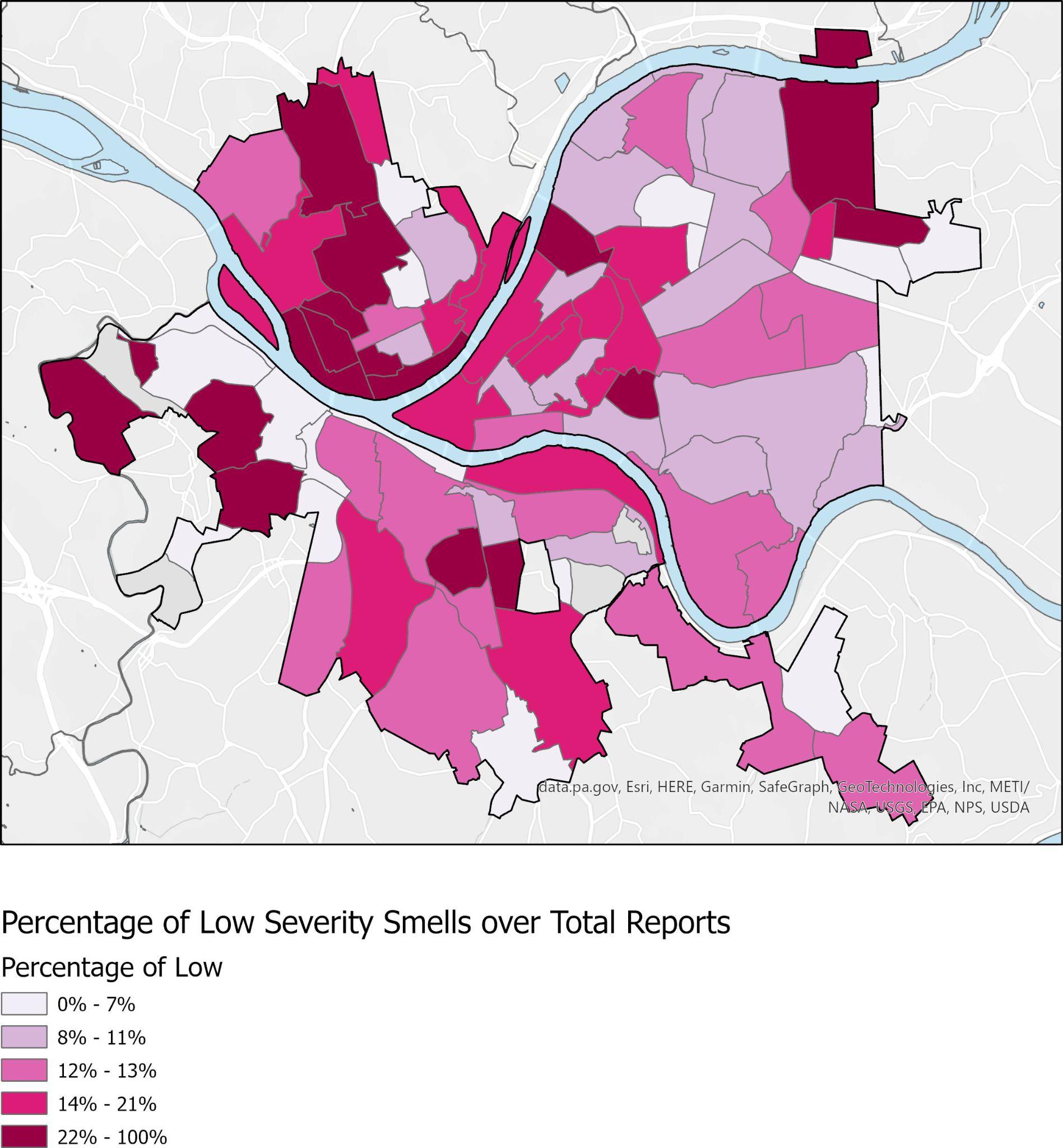
The Distribution of Reporting Across Pittsburgh:

According to Figure 4, the following areas seem to have the largest concentration of smell reporting: Squirrel Hill South, Point Breeze, Squirrel Hill North, and Brighton Heights. Total report numbers for these areas range from 2,000-7,000, and the surrounding neighborhoods also seem to have high volumes of reports. The areas of high reporting match the density of reports across smell severities that we observed in Figures 1, 2, and 3.

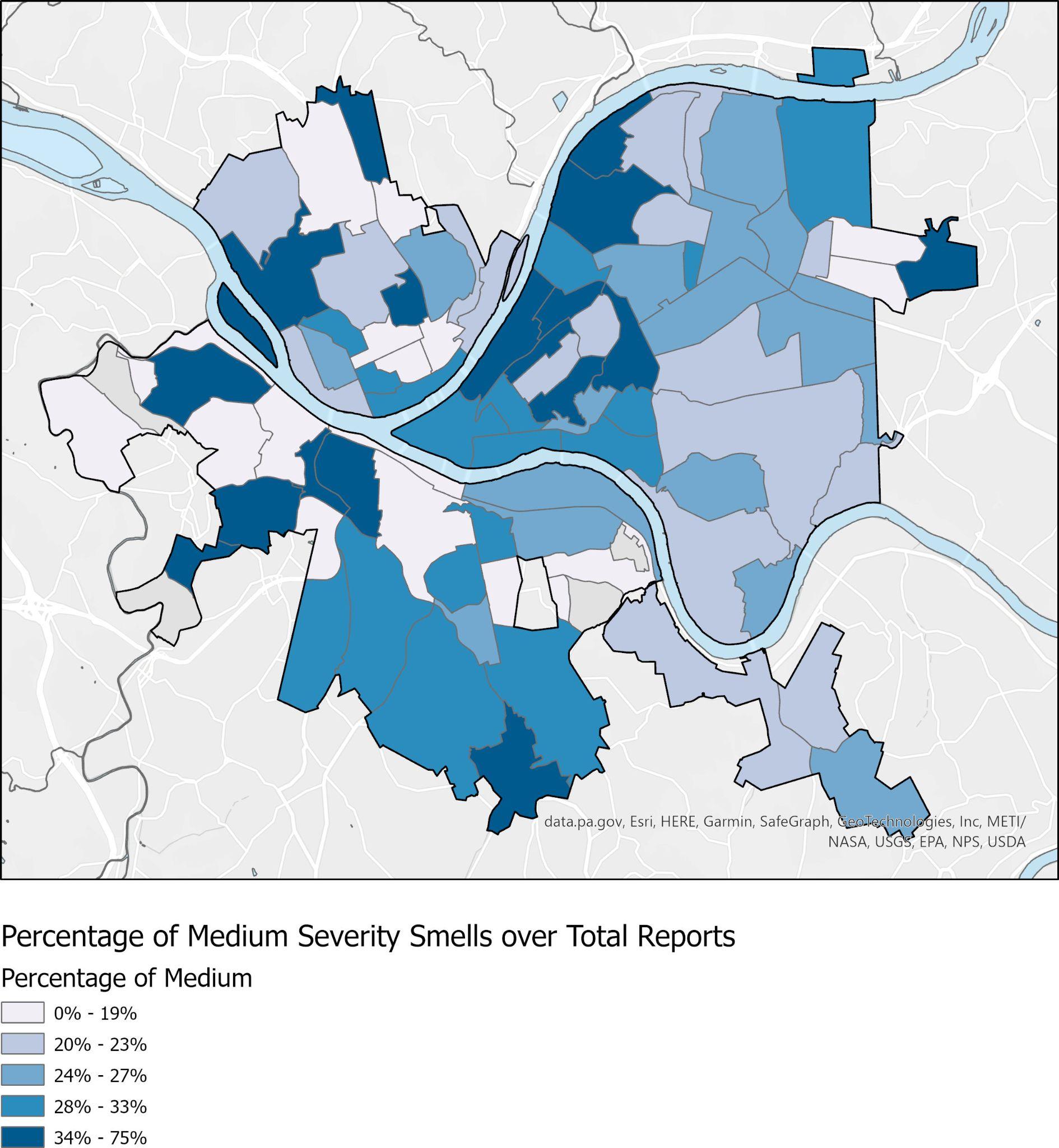
**Figure 4: Number of Smell Reports Per Neighborhood**

The Breakdown of Percentage of Reports that Fall Under Low, Medium, High Smell Severity:

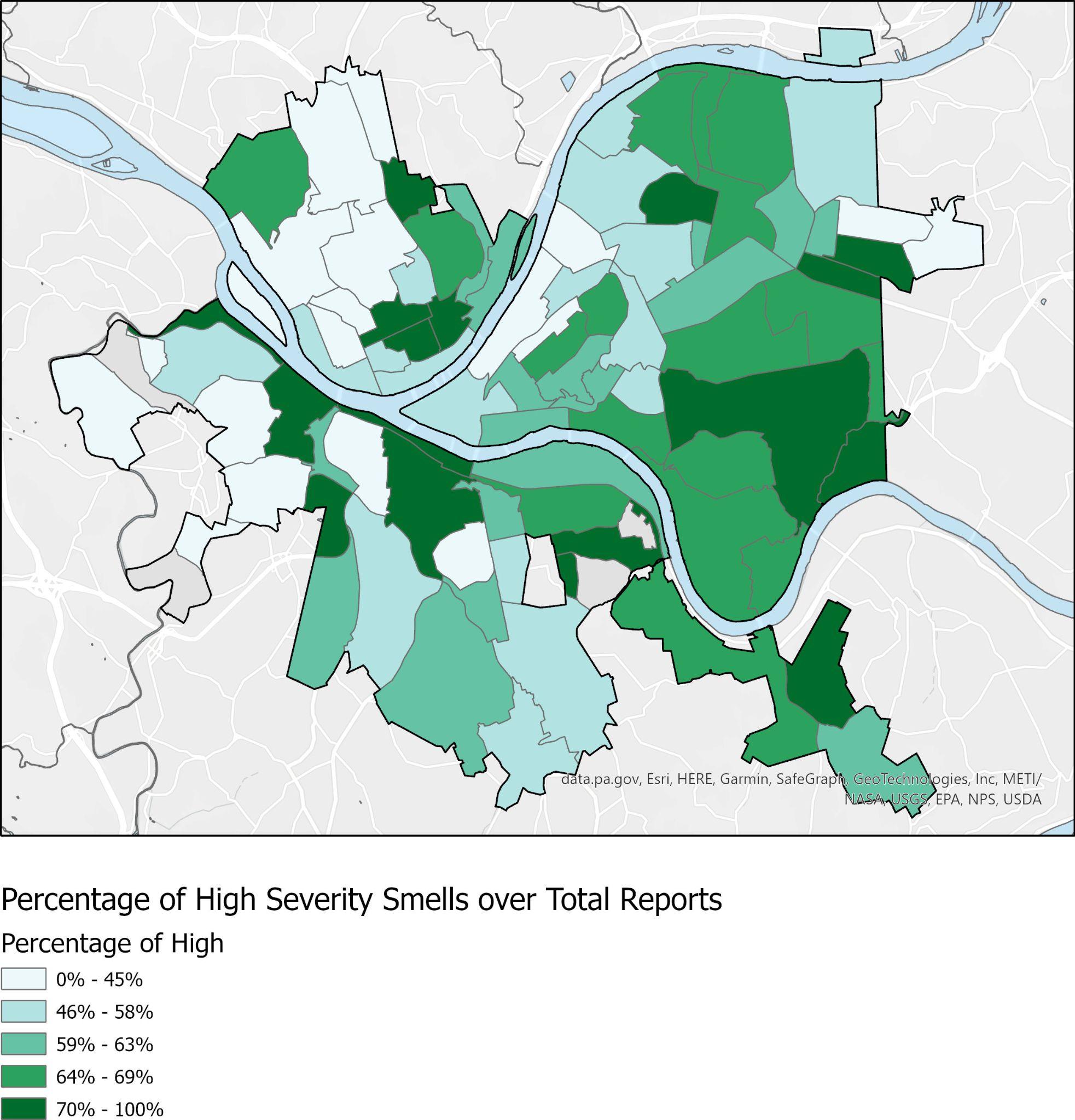
To get a better understanding of what percentage of reports in these neighborhoods are high, medium, or low smell severity reports, Figure 5, 6, and 7 were made to better visualize this breakdown by neighborhood. By looking at the percentage of reports rather than the general number of reports in each category for each neighborhood, we are controlling for unequal reporting of smells in each neighborhood.

**Figure 5: Percentage of Low Severity Smell Reports over Total Reports**

We found that neighborhoods with the highest percentages of Low Severity Smells included: Chartiers City, Homewood North, Fairywood, and Crafton Heights (see Figure 5). However, it is important to note that these neighborhoods also had very low reporting rates, collecting only 1, 7, 7, and 18 smell reports in total, respectively. This means that these results are more indicative of a lack of reporting/ lack of a sufficient sample size than anything else.

**Figure 6: Percentage of Medium Severity Smell Reports over Total Reports**

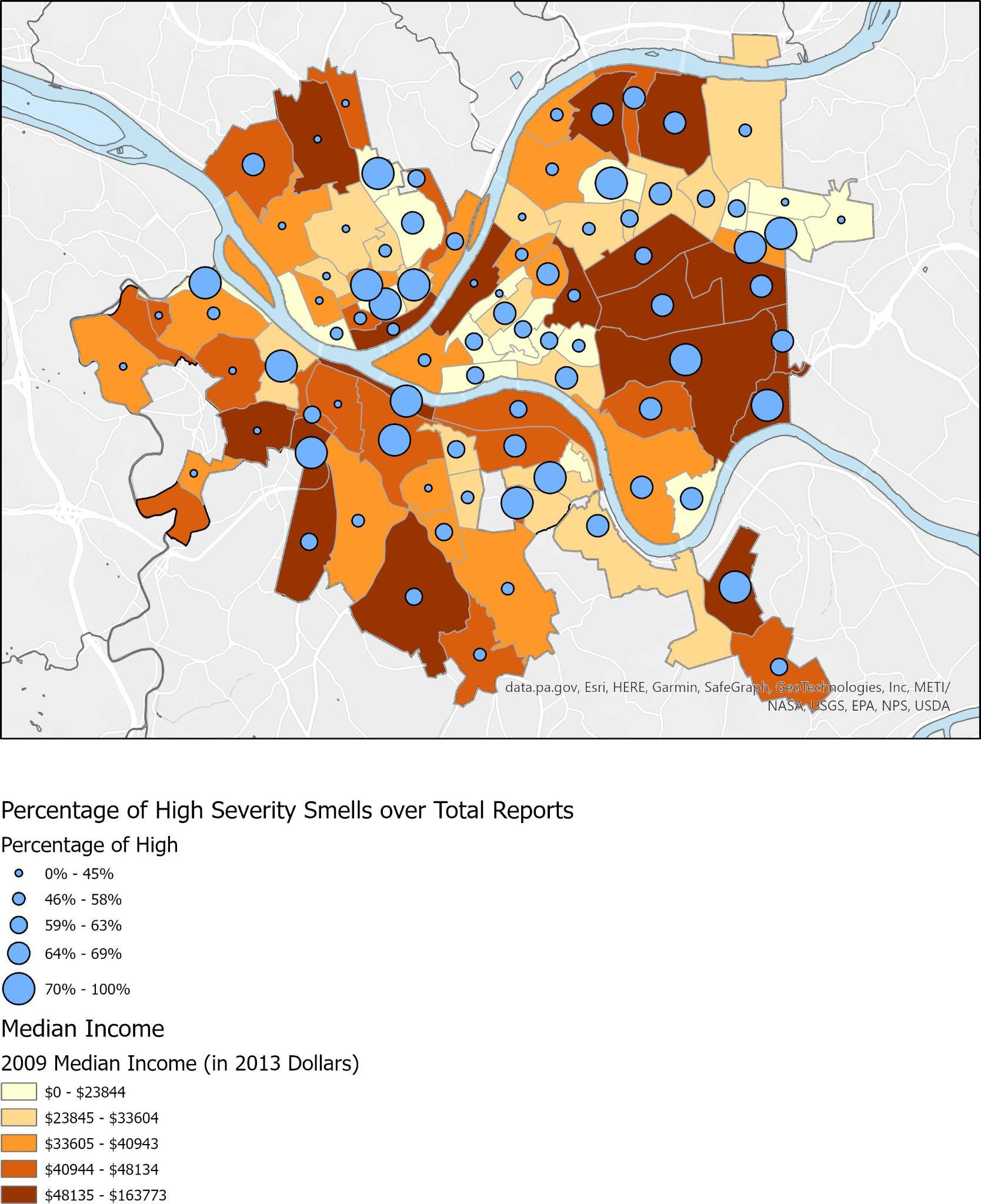
The following neighborhoods had the highest percentage of medium severity smell reports: East Hills, Oakwood, Sheraden, and Bedford Dwellings (see Figure 6). However, we can once again see that the total number of reports in these neighborhoods are once again quite low, totaling to 8, 4, 4, 6 respectively. This once again, highlights more of an issue with the small sample size rather than provide key insights on the percentage of medium severity smells.

**Figure 7: Percentage of High Severity Smell Reports over Total Reports**

The following map shows the neighborhoods with the high percentage of high severity smell reports: Northview Heights, Esplen, Elliot, Ridgemont, and South Shore (Figure 7). These top 5 neighborhoods also suffer from small sample sizes, however, as they only had a total of 1, 1, 10, 24, and 41 total reports. The sample sizes for these neighborhoods are not as small as the ones we observed in Figures 5 and 6, but it still shows that it is more indicative of poor sampling. Interestingly, Squirrel Hill South, which had the highest number of total reports (7,946), still has a significant percentage (70%) of its reports fall into the high severity category.

No Correlation Between Concentration of High Severity Bad Smells and Median Income:

In order to get a better sense of potential correlations between areas that have a high concentration of high severity bad smells and lower income areas, Figure 8 maps the two qualities of each neighborhood simultaneously. According to this figure, there does not appear to be a clear correlation between income and high severity smell neighborhoods, as areas of high and low median income experience both high and low percentages of high severity smell reports.

**Figure 8: Percentage of High Severity Smells And Median Income of Neighborhood**

**Conclusion and Future work**

The majority of our findings indicate that because these smell reports are reported by the community, there is no universal sample size for all neighborhoods. This makes it difficult to find reliable findings, as results tend to be skewed one way or the other depending on how many reports are sent in. For example, we see that with Figures 5, 6, and 7, which showed the percentage of low, medium, and high severity smell reports, yet the highest percentages also had the lowest number of total reports, skewing the results one way.

Future work in this area would benefit from increasing the use of this report tool, particularly in neighborhoods with little to no smell reports. By increasing the number of reports across neighborhoods, we can do more reliable interpretations of the data and the visualization of the data. Additionally, it may be helpful to set thresholds for the minimum number of total smell reports each neighborhood should have in order to be counted in the visualization. Currently, however, too many neighborhoods do not have enough reporting, and that would leave very few data features and points to work with.

**Data source(s)**

1. Pittsburgh PA GIS Open Data Portal - Neighborhoods with SNAP Data: <https://pghgishub-pittsburghpa.opendata.arcgis.com/datasets/pittsburghpa::neighborhoods-with-snap-data/about>
2. Pittsburgh PA GIS Open Data Portal - City of Pittsburgh Boundary: <https://pghgishub-pittsburghpa.opendata.arcgis.com/datasets/pittsburghpa::pittsburgh-city-boundary/explore?location=40.430815%2C-79.979816%2C12.25>
3. Alleghany County GIS Open Data Source: <https://openac-alcogis.opendata.arcgis.com/datasets/AlCoGIS::allegheny-county-hydrology-areas/explore?location=40.434655%2C-80.022402%2C9.65>
4. Smell Pittsburgh Data Documentation and Direct Download:<https://smellpgh.org/data>

**References:**

[1] Pittsburgh’s air improving, but still gets an ‘F’ from American Lung Association: <https://stateimpact.npr.org/pennsylvania/2021/04/21/pittsburghs-air-improving-but-still-gets-an-f-from-american-lung-association/>

[2] Cole, Luke W, and Sheila R Foster. 2001. From the Ground Up: Environmental Racism and the Rise of the Environmental Justice Movement. Critical America. New York: New York University Press. ISBN: 9780814772294

[3] Smell Pittsburgh Website: <https://smellpgh.org/>

[4] Smell Pittsburgh Data Documentation and Direct Download: <https://smellpgh.org/data>

[5] Census Geodatabases with demographic and economic indicators: <https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-data.html>

[6] Smell Pittsburgh API documentation: <https://github.com/CMU-CREATE-Lab/smell-pittsburgh-rails>

[7] Yen-Chia Hsu, Jennifer Cross, Paul Dille, Michael Tasota, Beatrice Dias, Randy Sargent, Ting-Hao (Kenneth) Huang, and Illah Nourbakhsh. 2019. Smell Pittsburgh: Community-Empowered Mobile Smell Reporting System.