Carnegie Mellon University

Social Determinants of Health among Children and Youth in Alleghany County

Miguel Antonio Avila Fuentes (mavilafu, mavilafu@andrew.cmu.edu)

Hiba Hassan (hibah, hibah@andrew.cmu.edu)

94-802: Geographic Information Systems

Professor Corey Harper

December 8, 2023

Problem Statement

Social determinants of health (SDOH) are the conditions in which individuals live, work, and engage in recreational activities that influence their health outcomes and quality of life.[1] Children and youth are particularly affected by SDOH, as they provide the foundation for their overall health and well-being in the future.[2] Among others, these factors include safe housing, violence, education, income, access to nutritious food and physical activity, and polluted air and water.[3] With this in mind, we intend to identify some SDOH and health outcomes for people in different neighborhoods in Allegheny County. We hope this work serves as a basis for nonprofits and government agencies, such as the Allegheny County Department of Human Services, to identify and alleviate disparities in SDOH.

Using geospatial analysis and the ArcGisPro tool, we intend to answer the following questions:

1. How does air quality in different neighborhoods affect the respiratory conditions of children and teenagers?
2. Is there a correlation between the number of recreational playgrounds, fast food restaurants, and obesity rates?
3. Can we generate a basic health index for Pittsburgh and identify the significant areas needing intervention to decrease health disparities?
4. Based on the health index built-in #3, can we identify locations where a community center might be built to aid health outcomes?

Methodology and Findings

In this section, we go through a thorough analysis of each of our sections, introducing our analytical theme, our approach, the dataset utilized, our findings, and solutions.

Part 1: Air Quality and Respiratory Conditions

Firstly, we looked into how air quality in different neighborhoods affects the respiratory conditions of children and teenagers in Pittsburgh.

Data Sources:

The primary datasets we utilized for this section were sourced from the following:

1. UPMC Emergency Department Visits: The dataset captures total emergency department visits and primary care visits for the pediatric population aged 0-17. It consists of data related to asthma, injury, low-acuity visits, and respiratory tract infections. [UPMC Emergency Department Visits(wprdc)]

[1] https://health.gov/healthypeople/priority-areas/social-determinants-health
[2] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6996928/#R1
[3] https://health.gov/healthypeople/priority-areas/social-determinants-health

2. Alleghany Air Quality: Data on air quality across Allegheny County was collected from the county health department monitors.
   [Alleghany Air Quality Index(wprdc)]
3. City of Pittsburgh 2010 Census Tracts:
   [2010 Census Tracts - CSV - CKAN (wprdc.org)]

Methodology and Process:

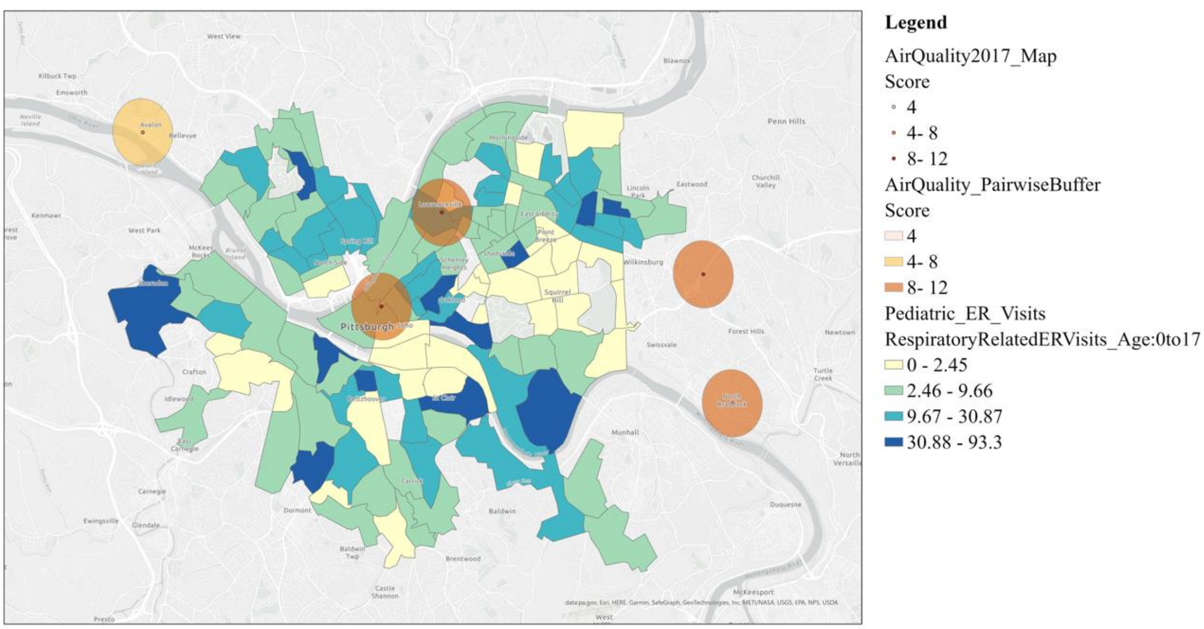
Starting with accessing data from the relevant data sources, we imported them into ArcGIS. We joined the dataset with the Pittsburgh city tracts data based on the GEOID to visualize the ER visits data on the map and mapped the Air Quality data using the XY coordinates.

This was followed by symbology, where we utilized graduated colors for each to identify different concentrations of Acute Respiratory Tract Infections Related Emergency Department visits across Pittsburgh and also to see how air quality levels changed across the city.

Finally, to answer our question more concretely, we used the **Pairwise Buffer** tool to understand the area of influence for each of the mapped air quality data points.

Solutions and Findings:

Figure 1: Air Quality and Respiratory Cases



The map above is the final layout for our analysis. Darker regions on the map, specifically dark blues, represent areas with high respiratory-related ER visits. The top regions with high to moderate respiratory cases are as follows: Glen Hazel, St Clair, parts of Oakland, parts of Shadyside, Sheradon, Downtown, and Lawrenceville.

When we integrate our analysis of the air quality buffers, we see how areas of downtown and Lawrenceville fall under a higher air quality index, representing bad air quality levels overall. Poor air quality could contribute to respiratory issues across these areas. However, it is important to note that due to data limitations, the spread of the air quality analysis may not be as coherent as we would like.

Other factors that may contribute to high respiratory conditions are discussed in the questions ahead. Still, looking at the map, we notice that in most spaces near a highway or areas with more offices/commercial activity, the overall respiratory issues are also more persistent.

Part 2: Obesity Rates and its Influencing Factors

Secondly, we examined whether there is a correlation between the number of recreational playgrounds, fast food restaurants, and obesity rates?

Data sources:

We start by importing our data from the Western Pennsylvania Regional Data Center. The data sources used to answer this question were:

1. Allegheny County Obesity Rates: these include estimates of obesity rates for each Census Tract in Allegheny County from 2006 to 2010
   Allegheny County Obesity Rates - Dataset - CKAN (wprdc.org)
2. Allegheny County Fast Food Establishments: including all chain restaurants without an alcohol permit up to 2016.
   Allegheny County Fast Food Establishments - Dataset - CKAN (wprdc.org)
3. City of Pittsburgh Parks: including city parks as of 2021
   City of Pittsburgh Parks - Dataset - CKAN (wprdc.org)
4. City of Pittsburgh Pools: including city pools and spray parks as of 2017.
   City of Pittsburgh Pools - City Pools Data Dictionary - CKAN (wprdc.org)
5. City of Pittsburgh Playgrounds: city playgrounds as of 2021
   City of Pittsburgh Playgrounds - Dataset - CKAN (wprdc.org)
6. City of Pittsburgh 2010 Census Tracts:
   2010 Census Tracts - CSV - CKAN (wprdc.org)

Methodology and Process:

After downloading the datasets, we imported them into the ArcGis project. Since we only had data from the City of Pittsburgh's parks, pools, and playgrounds, we decided to focus only on the Census tracts from this city. We achieved it using the ‘Select by location’ tool.

Afterward, we decided to locate the parks, pools, and playgrounds and symbolize them as separate layers. Furthermore, we created a choropleth map to identify the areas with the highest ratio of people with obesity.

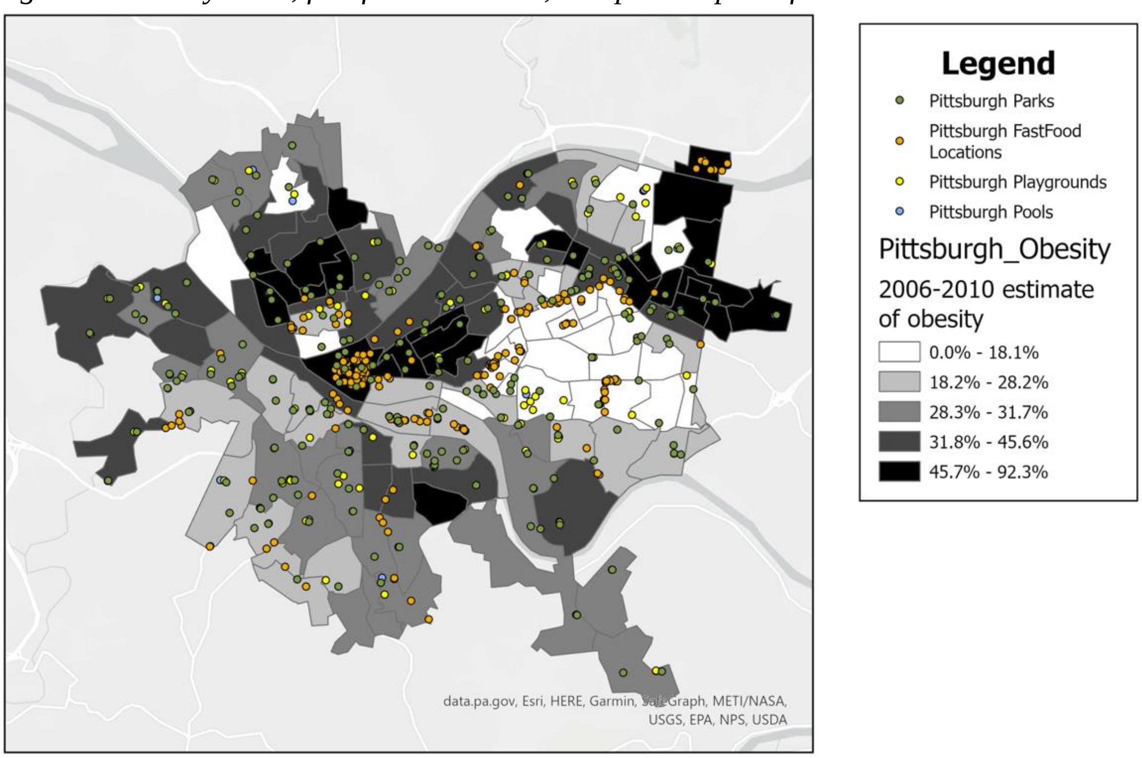
We also used the 'Spatial join' tool to count the number of fast food establishments, parks, pools, and playgrounds per census tract and added a new feature, 'Sports facilities,' with the sum of parks, pools, and playgrounds per census tract.

Finally, to answer our question about the correlation between obesity, fast food, and sports facilities, we created a pair of scatter plots and linear regression equations to show the relationship.

<u>*Solution and findings*</u>

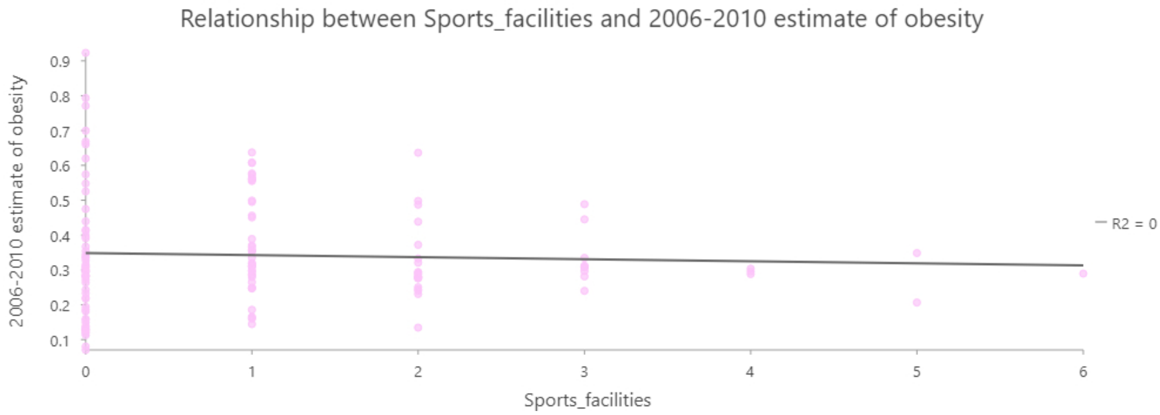
Figure 2 shows three areas with high obesity rates: Downtown, North Shore, and Homewood (northeast). While there is clear evidence of a high concentration of Fast-food restaurants in the downtown area, the other places identified with high obesity rates do not show evidence of it.

Figure 2: Obesity rates, fast food locations, and public sports facilities.



Furthermore, looking at the scatterplots - figures 3 and 5 - and linear regression equations, we can see that even though the coefficients for fast food locations are positive and for sports facilities are negative, there is a lot of variation, and the R-squared is really low. These variables have a slight correlation, but other non-considered variables might affect the outcome more.

Figure 3: Relationship between obesity estimates and sports facilities



With more sports facilities, there is a slight decrease in the estimated ratio for obesity. However, more evidence of the correlation is needed, as the R-squared of the linear regression is low. There is high variation, especially within Census Tracts with few or no sports facilities. The linear regression equation for the graph in Figure 3 is shown in Figure 4

Figure 4: Linear trend and R-squared for the relationship between Obesity (y) and number of sports facilities (x).

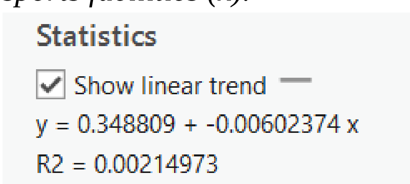
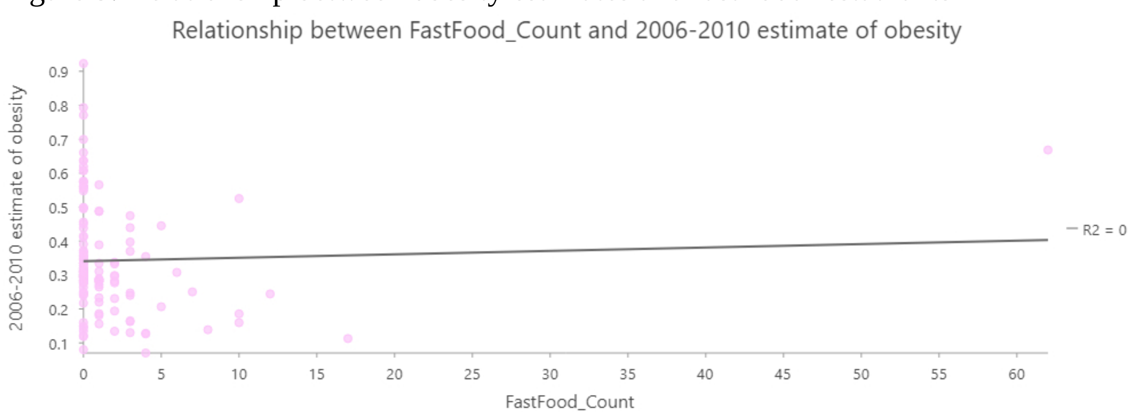


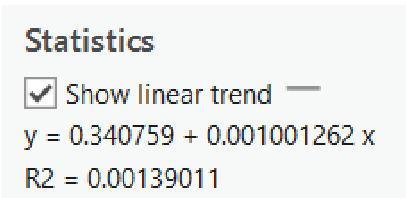
Figure 5: Relationship between obesity estimates and fast food restaurants



With more fast-food restaurants, there is a slight increase in the estimated ratio for obesity. However, more evidence of the correlation is needed, as the R-squared of the linear regression is low. There is high variation in the data, so we cannot conclude at this time that obesity is related to fast food locations.

The linear regression equation for the graph in Figure 5 is shown in Figure 6

Figure 6. Linear trend and R-squared for the relationship between Obesity (y) and number of fast food locations (x).



Part 3: Creating a basic Health Index.

With the previous analysis in mind, we decided to create a basic health index that considers more factors affecting the health of Pittsburgh citizens. For this first iteration, we decided to focus on safety (including the arrests per census tract), obesity rates (using estimated obesity rates per census tract), housing in poor conditions, and children and youth emergency visits due to respiratory tract infections.

Data sources:
We start by importing our data from the Western Pennsylvania Regional Data Center. The data sources used to answer this question were:

- Pittsburgh Police Arrest Data: arrest history data from the City of Pittsburgh
  Pittsburgh Police Arrest Data - Dataset - CKAN (wprdc.org)
- Pediatric Emergency Department and Primary Care Visits (UPMC): from 2016 to 2019, including emergency visits due to respiratory tract infections.
  Pediatric Emergency Department and Primary Care Visits (UPMC) - Dataset - CKAN (wprdc.org)
- Allegheny County Obesity Rates: these include estimates of obesity rates for each Census Tract in Allegheny County from 2006 to 2010
  Allegheny County Obesity Rates - Dataset - CKAN (wprdc.org)
- Allegheny County Poor Housing Conditions: The American Community Survey and Allegheny County Property Assessment database prepared the average estimate of the housing units in each tract as of 2016.

[Allegheny County Poor Housing Conditions - Dataset - CKAN (wprdc.org)](#)

- City of Pittsburgh 2010 Census Tracts:
  [2010 Census Tracts - CSV - CKAN (wprdc.org)](#)

Methodology and Process:

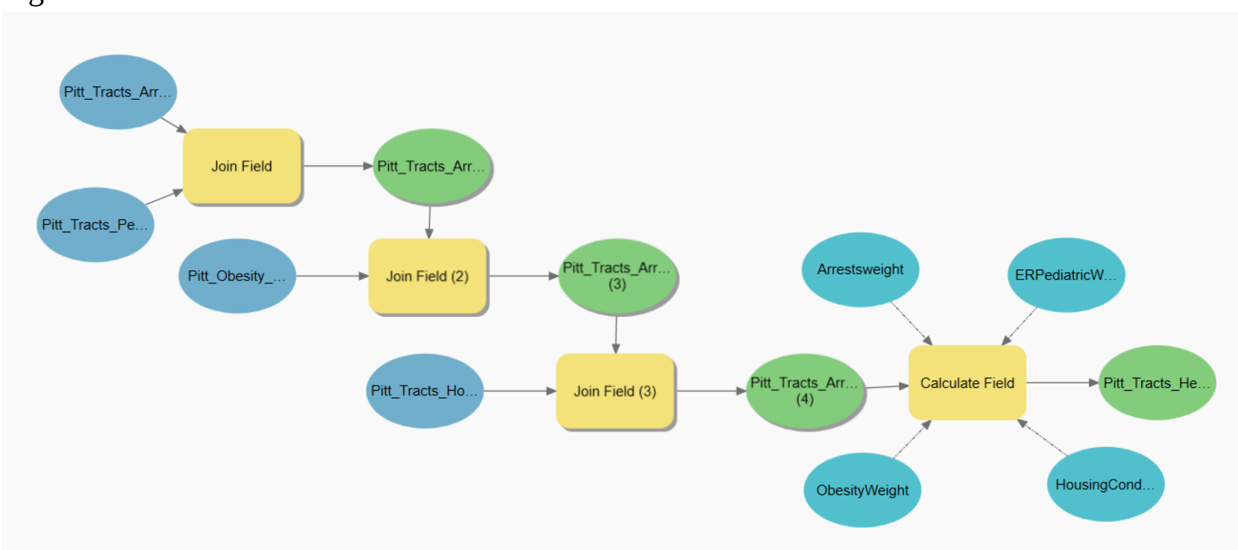
After downloading and importing the datasets specified in the previous section, we made choropleths maps to identify areas where these indicators were particularly high.

Then, we calculated the average values and standard deviation for each of the variables of interest to compute Z-values that will serve as input to the new health index.

Afterward, we created the model shown in Figure 7, where we joined the tables and used the 'Calculate field' tool to generate a new Health Index based on Z-values for each variable and applying a weight. The equation for the health index is the following.

0.25*$feature.ZArrests + 0.25*$feature.Z_ER_Resp_Visits + 0.25*$feature.Z_Obesity + 0.25*$feature.ZPPoorCond

Figure 7. The model used to calculate the Health Index.



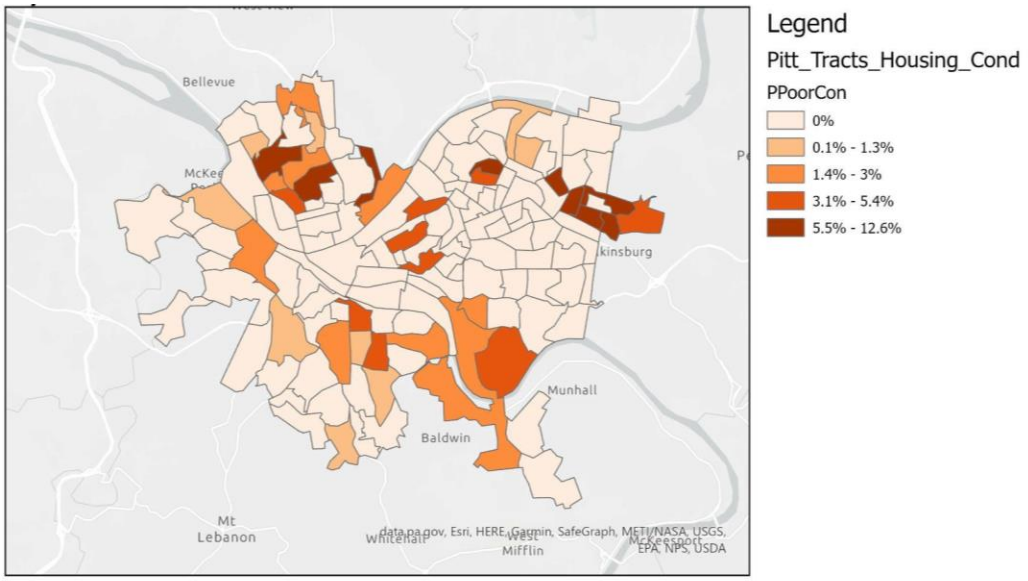
Finally, with the newly calculated health index, we generated a new choropleth map with standard deviations to measure the areas with the highest disparity in health indicators.

Solution and findings

We will first introduce choropleth maps for housing and arrests to identify Census Tracts with the most need for these conditions. Obesity and Emergency Room visits will not be visualized, as they were already covered in the previous sections.

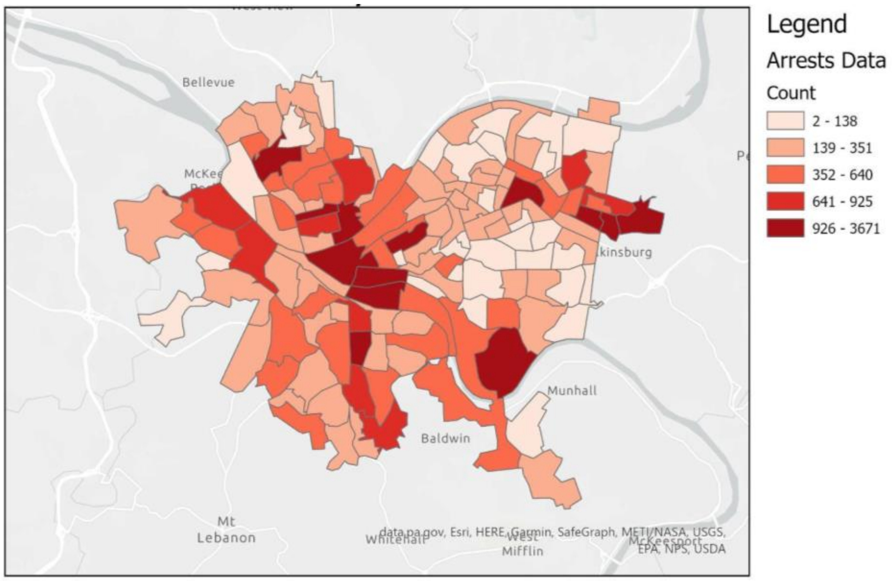
Figure 8 shows four areas with poor housing conditions: Northshore, Hill District, Homewood, and Glen Hazel. It was interesting to see that these areas are almost identical to those identified for obesity and pediatric emergency room visits.

Figure 8. Houses in poor condition by Census Tract



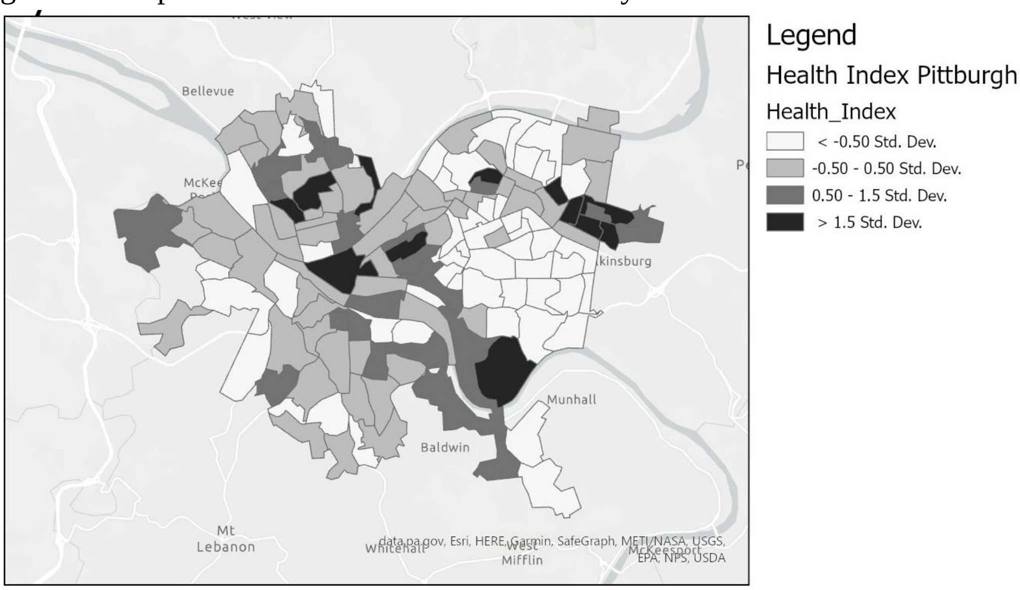
Furthermore, Figure 9 shows the number of arrests by Census Tract. An interesting finding is that the arrests are highest in the specific areas that relate to poor housing quality and the adjacent neighborhoods.

Figure 9. Number of arrests by Census Tract



After running the Health Index model, we corroborated our assumption of the four areas with the highest need: Northshore, Hill District, Homewood, and Glen Hazel. However, it was interesting to identify a fifth possible area with low social determinants of health in Garfield's neighborhood.

Figure 10. Disparities in calculated Health Index by Census Tract



Part 4: Health Index and Community Spaces

Lastly, we utilized our formulated health index(from the previous question) to develop a more targeted approach toward improving the city's social determinants of health status.

Data Sources:
The primary datasets we utilized for this section were sourced from the following:

1. Health Index (data sources were carried on from the previous question)
2. City of Pittsburgh Facilities: The data highlights facilities across the city and categorizes them under different types. Our focus remains on the use of Community Centers. City of Pittsburgh Facilities Data(wprdc)

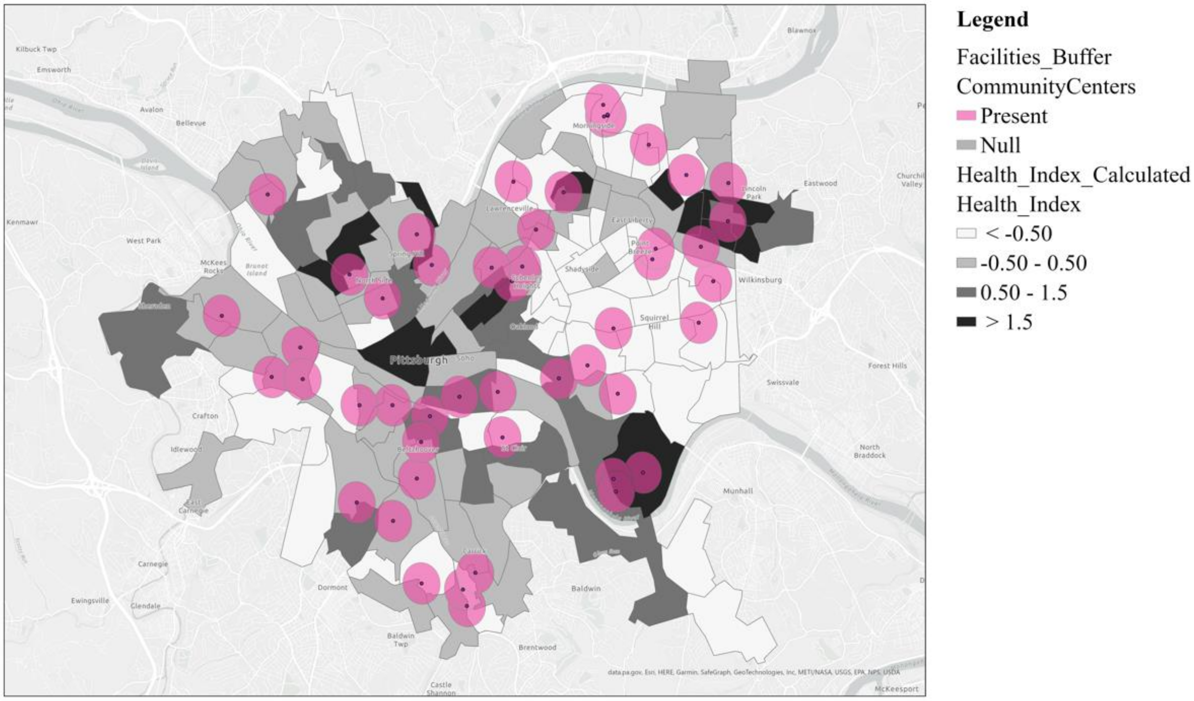
Methodology and Process:
We start by importing the health index from our previous question and integrating the city facilities data utilizing XY coordinates data.

To streamline the analysis, we filtered the types of facilities to focus only on those that contribute as community hubs or centers. Therefore, inculcating facilities like recreational centers, pools, old homes, sports facilities, etc. Once we had this filtered data, we created a new community center column. We calculated it by utilizing the Calculate Field and inputting a function to select only the filtered categories from the column 'type' that was part of the original dataset. Within this we also used select by attribute feature in arcgis to make a new layer of the facilities that only consisted of the selected community center categories.

Following this, using symbology, we utilized the unique values feature to highlight the positions for each community center across Pittsburgh. To understand the reach and impact of these centers, we enhanced our analysis by using the **Pairwise Buffer** tool around the centers.

<u>Solution and findings</u>

Figure 11: The Need for Community Centers, An Analysis



The figure above represents the health index as the base, where darker-colored regions represent poor levels of health in that area. Some of the areas with the worst health levels are Glen Hazel, around Brushton, areas around downtown, around Fowler ground - North Charles Street.

The purple buffer spaces represent the areas where community centers are present. What is interesting to note is that most regions identified with low health index also have limited or no access to community centers. Even with the buffer, we see that the reach of the community centers needs to cover areas most impacted by the multidimensional social and economic factors influencing the health index.

To improve the city's overall SDOH, the map highlights specific areas where there is a need to introduce community centers. Some of these areas are Glen Hazel, around downtown, near Woods Run, etc.

<u>Key Takeaways</u>

1. There is a higher overall prevalence of poor respiratory conditions in areas that have poor air quality across the city.
2. While eating fast food can lead to an increase in body weight, the amount of fast-food restaurants is not very much correlated with obesity rates in Pittsburgh.
3. It is possible to create a Health Index that identifies areas prone to low social determinants of health. However, more data should be considered to have a more robust

index. Additionally, tuning the model with the correct parameters of weight for each attribute used is essential.

4. Community spaces play an integral role in improving the social determinants of health of a region.

Future Work

In this section, we bring together our thoughts from our analysis, secondary research, and introduce a framework for approaching this going forward.

Many factors influence Social determinants of the health of individuals in a region. We focused our analysis on three key factors i.e., health, social environment, and community. However, more attributes should be considered to have a complete Health Index—for instance, truancy rates at schools and the income level of people living in each census tract.

Considering the profound influence of climate change and environmental factors on individuals' lifestyles is crucial. For example, the Breathe Project Pittsburgh stands out for its commendable efforts in improving the city's air quality. The initiative strategically directs its projects towards areas in Pittsburgh that have historically received less attention. As exemplified by this initiative, encouraging partnerships between the public and private sectors is essential for shaping more effective policies related to climate change and air quality. This becomes particularly critical when addressing respiratory and asthma-related concerns among children, emphasizing the necessity of creating spaces conducive to their well-being.

The integration of community hubs serves as a pivotal stepping stone in advancing the city's Social Determinants of Health (SDOH). Beyond merely enhancing SDOH, these hubs present a unique opportunity to discern and understand the diverse needs of the community. Functioning as safe havens, community hubs provide a space where individuals can converge, shedding light on the challenges they encounter. This, in turn, facilitates a collaborative exploration of how both community-based and governmental initiatives can effectively address these issues right from their inception.

References

1. Healthy People 2030. "Social Determinants of Health." U.S. Department of Health and Human Services, https://health.gov/healthypeople/priority-areas/social-determinants-health
2. Sokol R, Austin A, Chandler C, Byrum E, Bousquette J, Lancaster C, Doss G, Dotson A, Urbaeva V, Singichetti B, Brevard K, Wright ST, Lanier P, Shanahan M. Screening Children for Social Determinants of Health: A Systematic Review. Pediatrics. 2019 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6996928/#R1
3. The Breathe Project, Pittsburgh, https://breatheproject.org/about/

Process Log

Part 1:
Air Quality and Respiratory Issues

- Imported data sets "ER_Visits.csv" and "AirQualityData.csv," into the ArcGIS
- Data cleaning
- Joined the cleaned health data with Pittsburgh city tracts data using the unique identifier GEOID.
- Mapped the Air Quality data using XY coordinates
- Implemented graduated color symbology for ER visits, highlighting varying concentrations of Acute Respiratory Tract Infections in different city tracts.
- Utilized symbology to depict changes in air quality levels across Pittsburgh
- Applied the Pairwise Buffer tool to determine the spatial influence of each air quality data point
- Analyzed buffer zones to understand the potential impact areas associated with different air quality levels.

Part 2:
Obesity Rates and its Influencing Factors

- Download data from the WPRDC site.
- Imported csv files into ArcGis Pro
- Used the ‘Display XY data’ option to include Fast Food, Parks, Playgrounds, and Pools locations in the map
- Used ‘Select by location’ to keep data only inside Pittsburgh tracts
    - Used ‘Within’ relationship and manually added the extra tracts not included in the select by location stage.
- Added a new field called GEOidnum with ‘double’ as the data type for join purposes.
- Used Calculate field tool to copy the GEOid to GEOidnum field
- Joined Obesity rates to Pitt tracks using GEOidnum
- Created Cloropeth maps using the quantiles method and 5 classes to map obesity rates
- Used Spatial Join to Count the number of FastFood Restaurants per Tract
    - Created a choropleth map to symbolize
- Used Spatial Join to count the number of Pools, Parks, and playgrounds per Census Tract.
- Sum the number of pools, parks, and playgrounds in a new attribute named Sports_Facilities.
- Finally created two scatter plot charts with obesity rate as the dependent variable and facilities and fast food locations as the independent variable (one for each)

Part 3:
Health Index

- Download data from the WPRDC site.
- Imported csv files into ArcGis Pro
- Used ‘Select by location’ to keep data only inside Pittsburgh tracts
  - Used ‘Intersect’ relationship and manually removed the extra tracts included in the select by location stage.
- Used the summarize tool to get the mean and standard deviations of the variable of interest
- Added a new field and calculated the Z-value for the variables of interest using the mean and standard deviation calculated in the step before.
- Use the model tool to create a model that calculates the Health Index.
- First, we joined the Arrest Data table to the Pediatric Visits table using GEOIDnum and keeping only the Z-value variable of interest.
- Then joined this newly created table to the Obestity table using GEOIDnum and keeping only the Z-value variable of interest.
- Then joined this newly created table to the Housing table using GEOIDnum and keeping only the Z-value variable of interest.
- Added weights as variables and used the Calculated field tool to generate the Health Index
- Finally, symbolized the new Health Index by the Census Tract using standard deviation, 4 classes, and an interval size of 1 standard deviation to generate a choropleth map.

Part 4:
Health Index and Community Centers

- Added the health index layer from the previous question
- Imported the city facilities data using XY coordinates
- Incorporated the filtered data by creating a new column named "Community Centers."
- Utilized the Calculate Field function to determine community centers, applying a function to select specific categories from the "type" column in the original dataset.
- Using symbology, the unique values feature to highlight the positions of each community center within Pittsburgh
- Expanded the analysis by utilizing the Pairwise Buffer tool to comprehend the reach and impact of the identified community centers
- Used the symbology on buffer zones, marking transparency as 40% to showcase the buffers across the map