May Paddor

MGSC 7000

Course Project Executive Summary

I. Introduction

In this project, I worked on datasets from the City of Seattle Public Health department to conduct an exploratory analysis of public health data. I used three datasets from the City of Seattle: 1) the Racial Social Equity Composite Index, the Seattle GeoData hospital dataset, and the Seattle Department of health disparities insurance dataset. I conducted three analyses of the datasets and created short-term and long-term solutions based on the results. In this executive summary, I will walk you through an overview of the datasets, each analysis and its result, and the recommendations for the City of Seattle Department of Public Health.

II. Datasets

The primary dataset I used is the Racial Social Equity Composite Index (RSI), which can be located on data.gov. The RSI was uploaded by the City of Seattle in 2022 and looks at trends in race, socioeconomic status, and health disadvantages in the different neighborhoods in Seattle by census tract. After reviewing the data, I focused my analyses on chronic conditions. RSI contains data on three chronic diseases: diabetes, asthma, and mental health not good. Mental health not good is defined by those experiencing bad mental health for two weeks or more. The dataset illustrates the data in percentages and data. I conducted my initial exploratory analysis using the percentages to illustrate the prevalence of disease in the census tract and switched to percentiles in my further analysis to convey uniformity across census tracts regardless of population. To use the dataset, I omitted a row with NAs because it comprised less than 5% of the 135 rows. [See figure 1 for my initial analysis of the data.]

In addition to the RSI dataset, I used Seattle's GeoData website to compile a list of hospitals in the Seattle area. The dataset is constantly updated with GIS technology and was able to compute a list of hospitals in the Seattle area. I used this dataset to illustrate access to hospital care in Seattle. There are ten hospitals in Seattle, many of which are located in the city's center. [See figure 2 for a map of Seattle's hospitals.]

The last dataset I used was the insurance dataset from the City of Seattle's Race and Social Justice Initiative webpage. The dataset was published to convey Seattle's health disparities. The dataset listed the percentage of people without insurance by census tract. After my initial analysis of the data, I found that Asian populations in Seattle have the lowest average of uninsured populations at 4.5%. American Indians and Native Alaskans had the highest average of uninsured populations at 19.31%. However, the datasets for American Indians and Native Alaskans, Native Hawaiians, and Pacific Islanders are skewed because the City of Seattle could not gather the data for more than 90 census tracts. To compensate for the missing data, I replaced the NAs in Excel with the respective averages of the population. [See figure 3 below to observe the percentages of populations without insurance by race.]

III. Hierarchical Cluster Analysis

To start my analysis, I created a hierarchical cluster and used a heat map to illustrate its findings. I used hierarchical clustering to observe correlations in the data for the percentile of diabetes, asthma, mental health, and overall health percentile. The Health percentile is calculated as an overall measurement of health, with higher values illustrating a higher level of health. The heat map shows a correlation between asthma and mental health percentiles because the shading is similar throughout the map. Overall, there is variation within the census tracts regarding the prevalence of different chronic conditions. [See Figure 4 for the heat map.]

IV. Multiple Linear Regression & Recommendation

After conducting a hierarchical cluster, I created multiple variable regressions for each chronic condition to understand which variables affected the prevalence of the disease. First, I calculated the correlation between the dependent variables and all possible independent variables. Then, I ran regressions and chose variables based on the Adjusted R-Squared to ensure accuracy. Afterward, I ran variance inflation factor (VIF) tests to ensure multicollinearity was not an issue in the models. From the models, I can conclude that obesity is a significant factor in all three chronic conditions. Therefore, I recommend that the City of Seattle invest more resources into implementing campaigns focusing on physical health and incentivize companies to promote physical health through health and wellness stipends or courses. [See Figure 5 for the multiple linear regression significant variable table.]

V. High Burden and Hospitals and results

Next, I compiled a table comprised of tracts with the highest percentiles of each disease. I was able to find eighteen tracts that had the highest prevalence across all three conditions. Then, I compared this data to the tracts with the hospitals. I found that only 1 out of the 18 tracts with a high prevalence of chronic conditions had hospitals, which means that hospitals are not located where the burden of disease is the highest in the city. Therefore, I recommend that the City of Seattle conduct more research into whether these areas are meeting their hospital care needs by analyzing ambulatory care in the area and surveying community members. In terms of long-term solutions, I suggest the City of Seattle invest in healthcare providers in long-term areas and create a program to help community members with care coordination.

VI. High Burden and Insurance

Finally, I looked at the tracts with the highest prevalence of disease and the corresponding percentages of populations without insurance by race in those tracts. I compiled a table to compare the results against the average of populations without insurance by race across Seattle. I found that populations across all races in tracts with a high prevalence of chronic conditions had higher percentages of uninsured populations. I recommend that the City of Seattle partner with community leaders to educate the public on health insurance, enrollment and research barriers to health insurance for minority communities. Lastly, I recommend that the City of Seattle create a program to aid people in finding insurance year-round. [See Figure 6 for the table comparing insurance rate averages against insurance rates in high prevalence tracts.]

Figure 1: Exploratory analysis of RSI dataset.

Measure	Diabetes	Asthma	Mental Health Not Good
Min	1.2%	8.0%	7.1%
Mean	6.84%	8.93%	9.97%
Max	17.77%	11.4%	16.2%

Figure 2: Map of Seattle Hospitals.

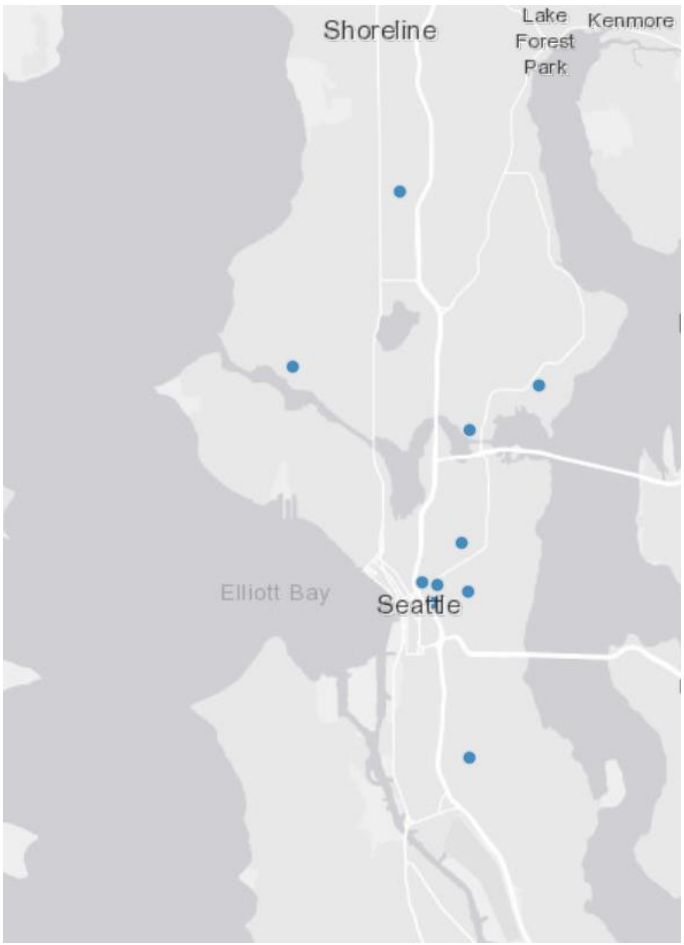


Figure 3: Table of populations without insurance by race from exploratory analysis of insurance data.

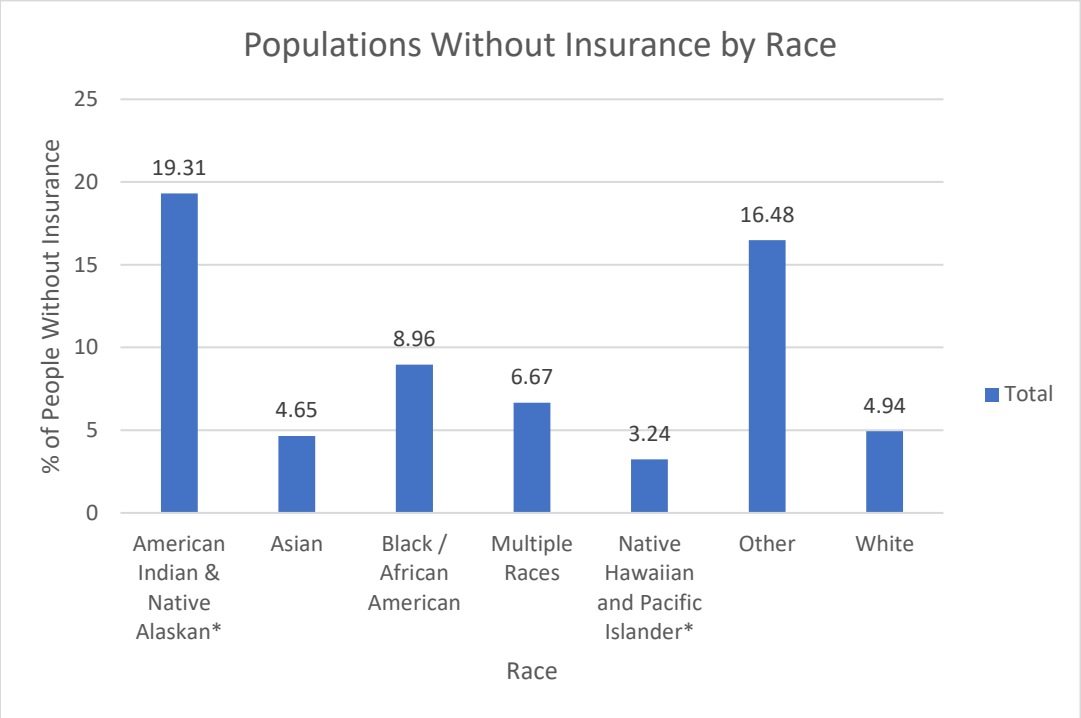


Figure 4: Heat map created from hierarchical cluster analysis.

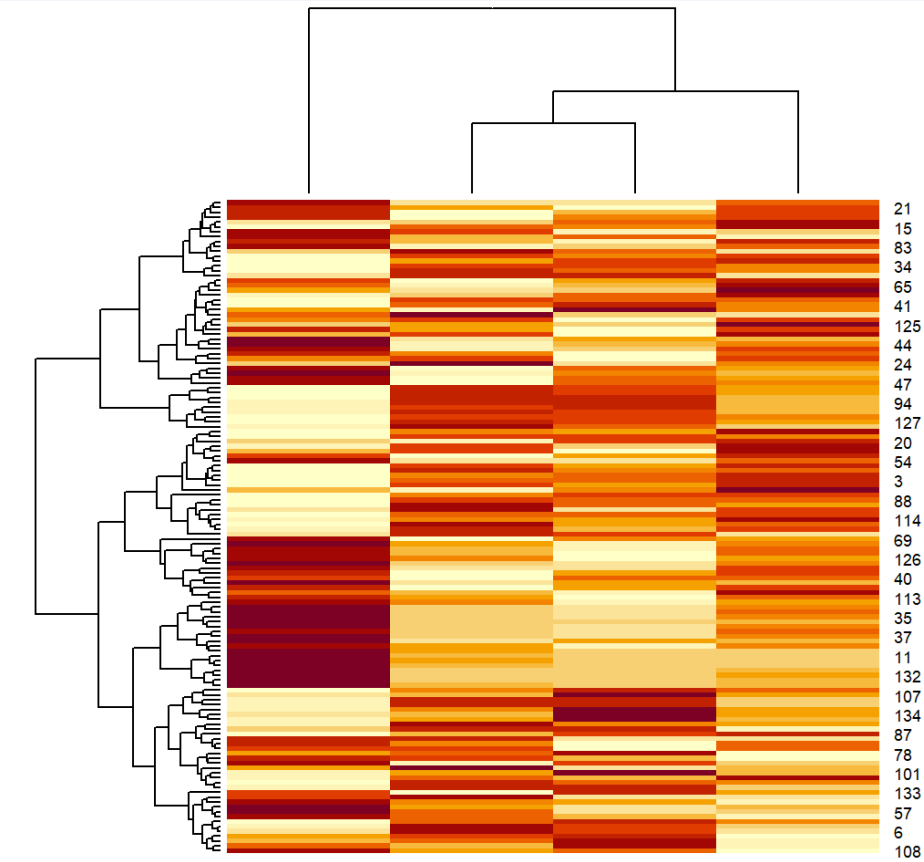
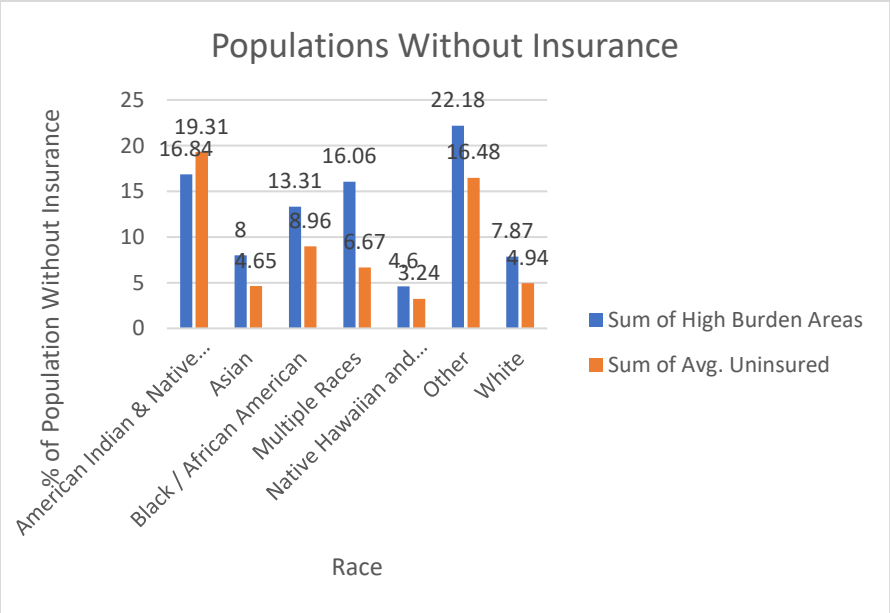


Figure 5: Table showing significant independent variables in multiple linear regressions, denoted by number of stars.

Variable	Asthma	Diabetes	Mental Health
Percent of Foreign Born	**	-	**
No Leisure/Physical Activity	-	***	***
Percentile of Diagnosed Diabetes	NA	NA	*
Percentile of People with Asthma	NA	NA	***
Population in Poverty	-	*	***
Obesity	***	***	*
Mental Health Not Good	***	**	NA
Less than a Bachelor's Degree	*	**	NA

Figure 6: Percentage of people without insurance by race in areas of high prevalence of disease compared to the average percentage of people without insurance by race.



Data Sources

- Insurance: https://www.seattle.gov/rsji/racial-equity-research/health-disparities#socialdeterminantsofhealth
- Racial and Social Equity Composite Index: https://catalog.data.gov/dataset/racial-and-social-equity-composite-index-65384
- Hospital Data: https://data-seattlecitygis.opendata.arcgis.com/datasets/8e0d80152ccb404abc6ac85351e4aed6/explore?location=47.567136%2C-122.280504%2C10.00