Analyzing the Environment's Impact on our Health

Introduction

Lifestyle choices as well as our environment are important factors that affect our quality of life. However, most people live in cities, which have a reputation for being dirtier than rural areas. Since there are more people in cities, more waste is produced in these areas and it must be disposed of safely. In this project, I wanted to analyze how certain environmental factors such as waste sites and green spaces impacted our quality of life. More broadly, I analyzed the impact of crime, income, green spaces, and waste sites on health.

Data

For this project, I retrieved the below datasets to conduct my analysis:

- **Census Bureau**
  - Income and population data
- **Centers for Disease and Control**
  - Chronic health data
- **Massachusetts Government**
  - Waste sites
- **City of Boston**
  - Green spaces, crime and schools

Analysis and Results

I began my analysis by looking at the correlation coefficients. The table below summarizes some of the important findings. For instance, income and health are inversely correlated, indicating that the higher the income, the less health problems there are in that area.

Correlation Coefficients

	Income	Waste	Crime	Open Space
Health	-0.016	0.084	0.427	0.319
Cancer	-0.001	0.147	0.009	0.205

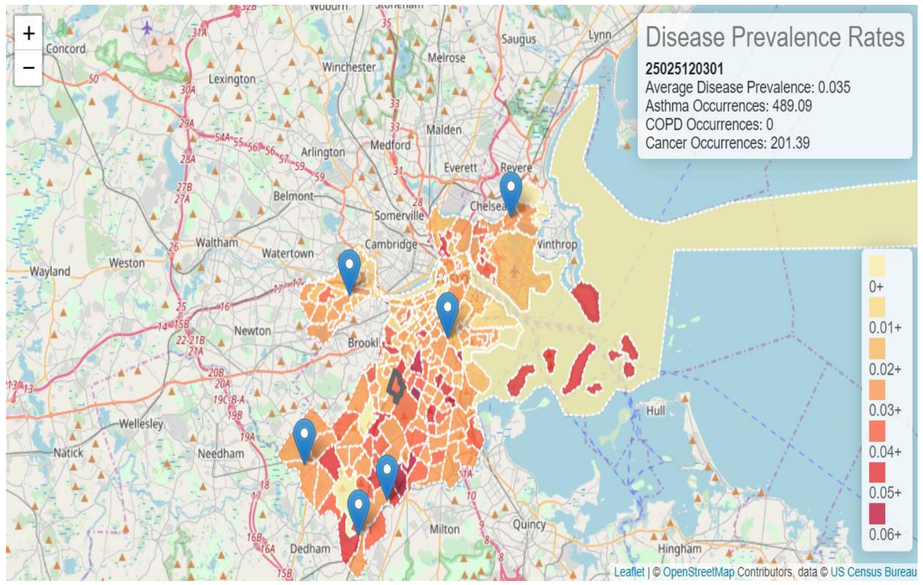
After calculating correlations, I ran the following multiple regression model:

$$y_{health} = \beta_0 + \beta_1 x_{openSpace} + \beta_2 x_{waste} + \beta_3 x_{income} + \beta_4 x_{crime} + \beta_5 x_{population}$$

Ultimately, I found that there was a statistically significant relationship between crime and health, and waste and health at the 5% confidence level. However, the magnitudes of these relationships were pretty small. Based on these results, it is difficult to determine whether there is a causal relationship between these factors. Below, I created a map of Boston with the disease prevalence rate for every census tract and I found the top four most optimal waste sites by solving a constraint satisfaction problem.

Limitations and Future Work

While completing this project, I found it difficult to find reliable and comprehensive health data for Boston. The health data I found were estimates of prevalence rates for certain diseases provided by the Centers for Disease and Control, which is important to keep in mind when interpreting my results. In the future, I would like to conduct a more robust analysis that takes into account factors such as unhealthy behaviors.



Nicole Mis