



Walkability, Income, and Race in Portland

Troy Ramsey
Rahil Nath
Tony Cipolle
Nnamdi Nwajagu

Motivation & Summary

We looked at walkability scores across Portland Neighborhoods and compared them to demographic and economic mobility factors from 2012 - 2018.

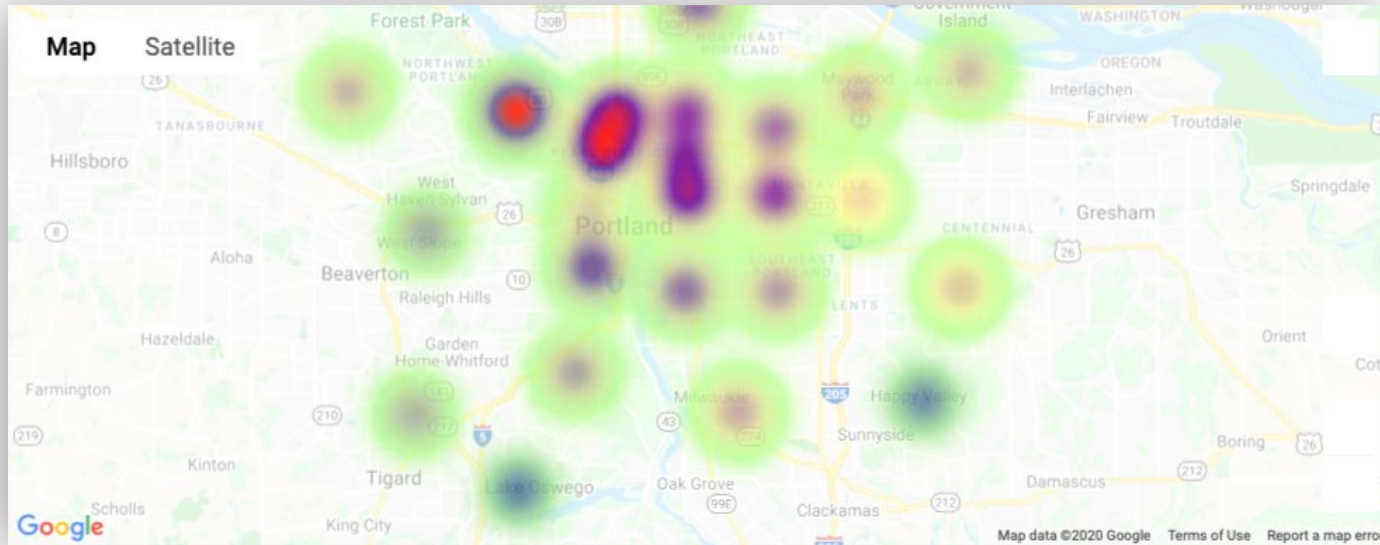
Originally, we wanted to compare walkscores of neighborhoods with *actual* pedestrian foot traffic, but real foot traffic data is expensive.

We decided to look at **demographic** and **economic** changes in neighborhoods, based on their walk scores:

- Median incomes
- General population change
- Population change by race/ethnic communities
- Increases or decreases in crime

Hypotheses:

- A. Higher walk scores correlate to higher increases in median income.
- B. Higher walk scores correlate to higher increases in total population.
- C. Higher walk scores correlate to more significant changes in BIPOC populations.



What is a Walk Score?



A real estate industry metric used to evaluate home value.

Walk Score measures the walkability of any address using a patented system

- Walk Score analyzes walking routes to nearby amenities.
- Points are awarded based on the distance to amenities.
- A decay function is used to give fewer points to more distant amenities.

Walk Score also measures pedestrian friendliness by analyzing:

- Population density
- Block length
- Intersection density

Data sources include: Google, Factual, Great Schools, Open Street Map, the U.S. Census, Localeze, and places added by the Walk Score user community.

Summary of Findings

- A. Zip codes with higher walk scores show a **strong positive correlation** with faster median income growth ($r=0.52$).
- B. There is **little correlation** between walk score and general population change ($r=-0.19$) or aggregate BIPOC population change ($r=-0.14$).
- C. Zip codes with higher walk scores showed a **moderate negative correlation** with changes in Black community population ($r=-.029$) and a **moderate to strong negative correlation** ($r=-0.4$) with the decrease in the ratio of Black residents to all residents.

Data we needed and where we found it:

- Walk Score by Neighborhood (Walkscore.com)
- Latitude/Longitude by Neighborhood (Portland OpenData Arc-GIS “Neighborhoods_Regions” dataset)
- Neighborhood Zip Codes (Google Maps Geocode API)
- US Census API
 - Change in Median Income (2012-2018)
 - Change in Total Population (2012-2018)
 - Change in BIPOC Population (2012-2018)
 - Change in Black Population (2012-2018)
- Crime rate data 2015, 2018 (City of Portland Police Bureau Open Data Source)

Data Cleaning

Jupyter Notebook: Troy Ramsey

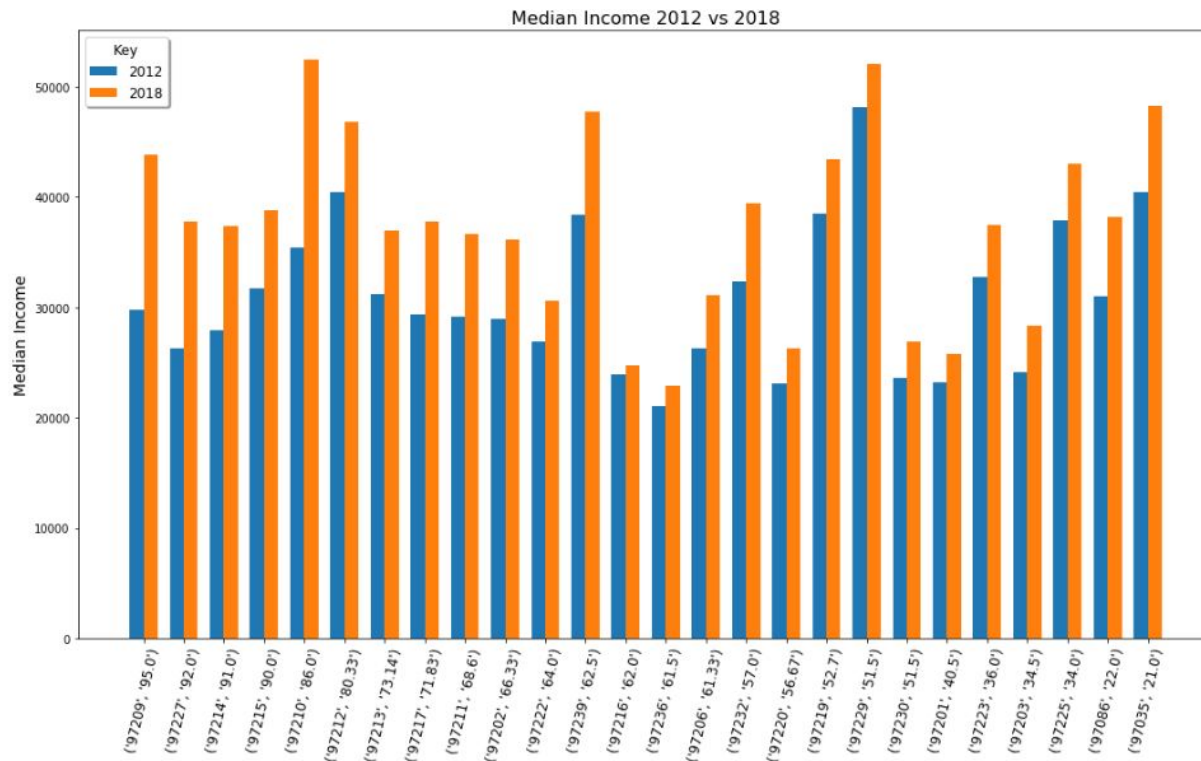
Problems that arose after exploring the data, and how we resolved them

- Most demographic and socioeconomic data is not available by neighborhood, so we had to aggregate our walk scores into zip codes, reducing our available points of plotable data
- Finding supporting datasets as we explored our initial data findings became harder as we narrowed down our scope
- Arc-GIS files use latitude and longitude values to create thousands of anchor points per neighborhood to create vector objects to overlay over maps. It's difficult to determine the most centrally located of these thousands of values, so we defaulted to the second returned value, which results in the northwesternmost pair for each neighborhood
- Some neighborhoods' nomenclature differed between data sets, and there were no easy formatting rules to follow in matching these neighborhood names (eg, "associations," "collectives," or other formal organizational groups). Luckily, this data set is small and each value has a unique index value, so matching wasn't too time consuming, but no script could be written to automate the process

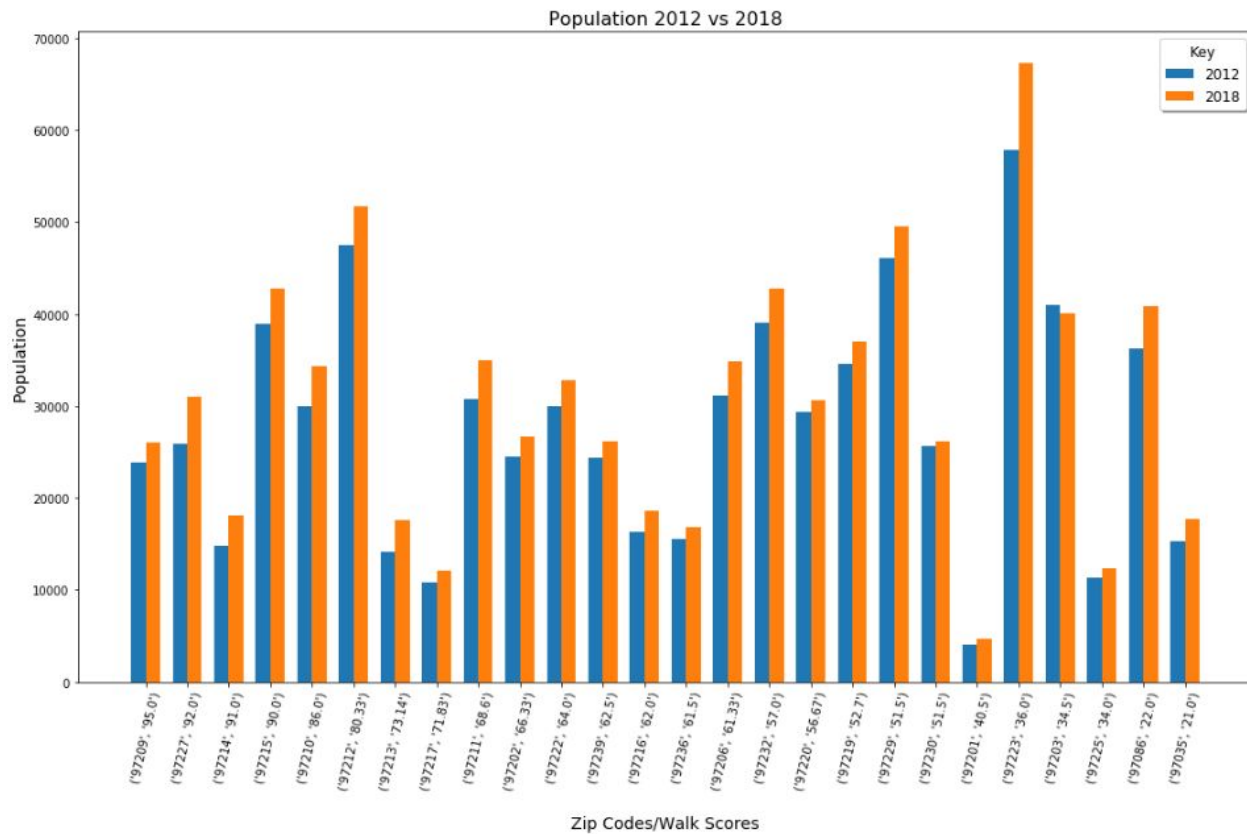
Data Analysis

- Used linear regression to search for relationships between changes over time and walk score
 - Built a boxplot to identify potential outliers in median income data, then presented the same linear regression model with and without median income outliers
- Stacked two sets of Black population ratio data into a shared scatterplot to illustrate change in each pair of ratio values

Data Visualizations



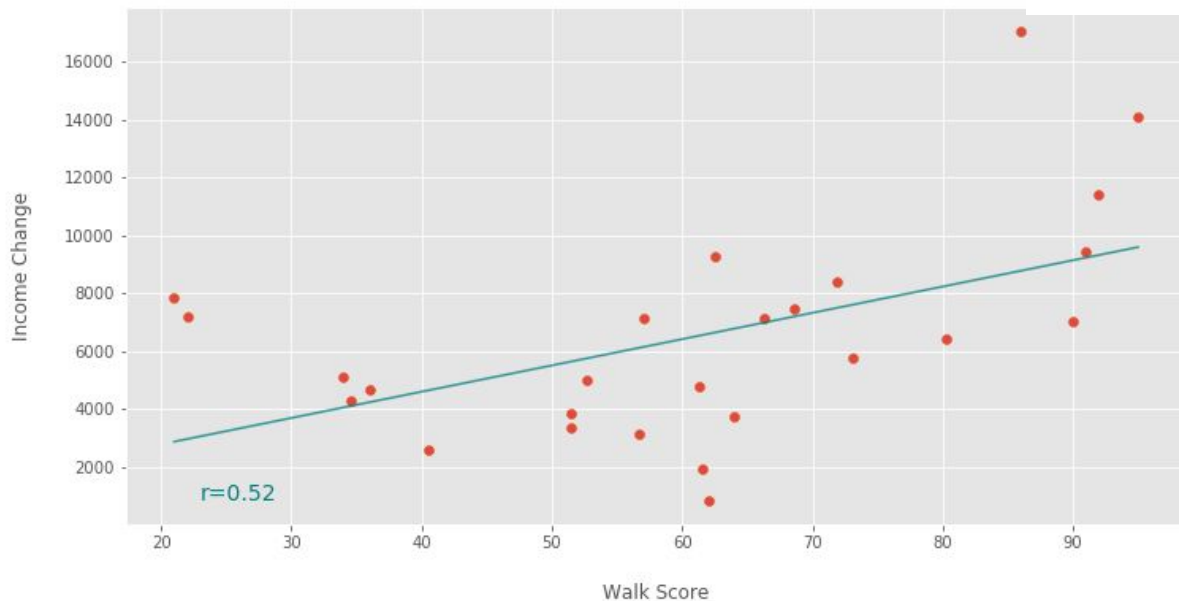
Data Visualizations



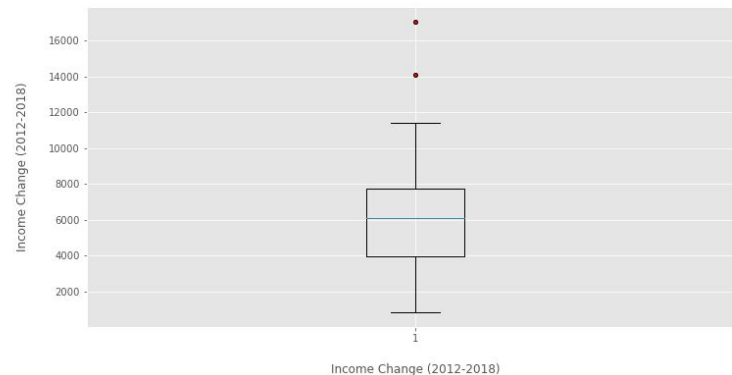
Changes in Median Income

Scatter plots with linear regressions

Changes in Median Income (2012-2018) by Walk Score



Distribution of Income Change Across Portland Zipcodes (2012-2018)

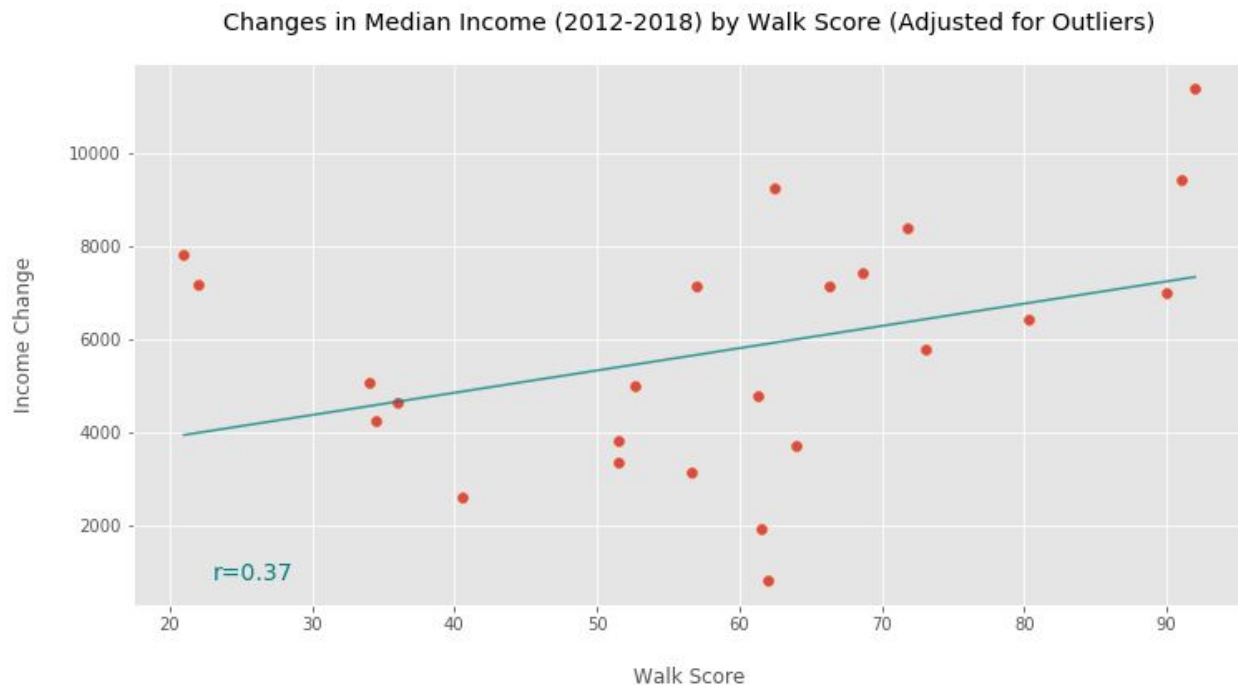


Zip codes with higher walk scores show a **strong positive correlation** with faster median income growth ($r=0.52$).

Median Income (dropped outliers)

Scatter plots with linear regressions

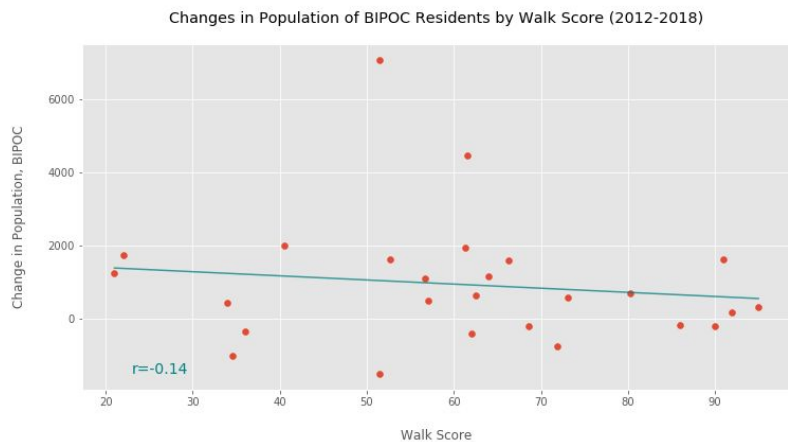
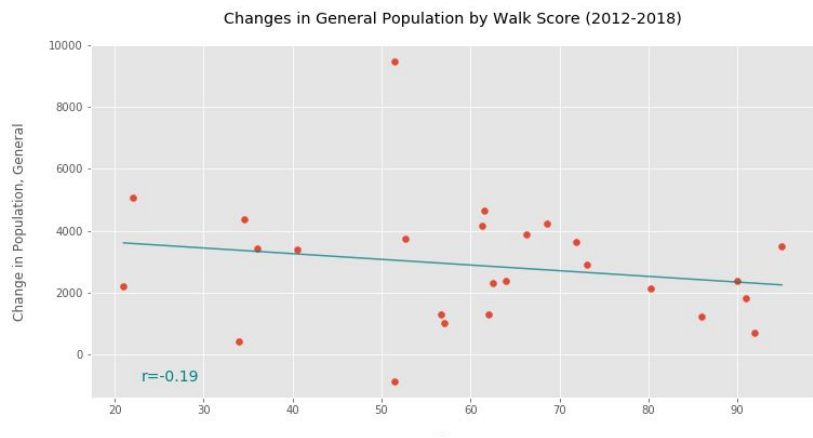
Zip codes with higher walk scores still show a **moderate positive correlation** with faster median income growth adjusted for income outliers ($r=0.37$).



Population Change by Walk Score

Scatter plots with linear regressions

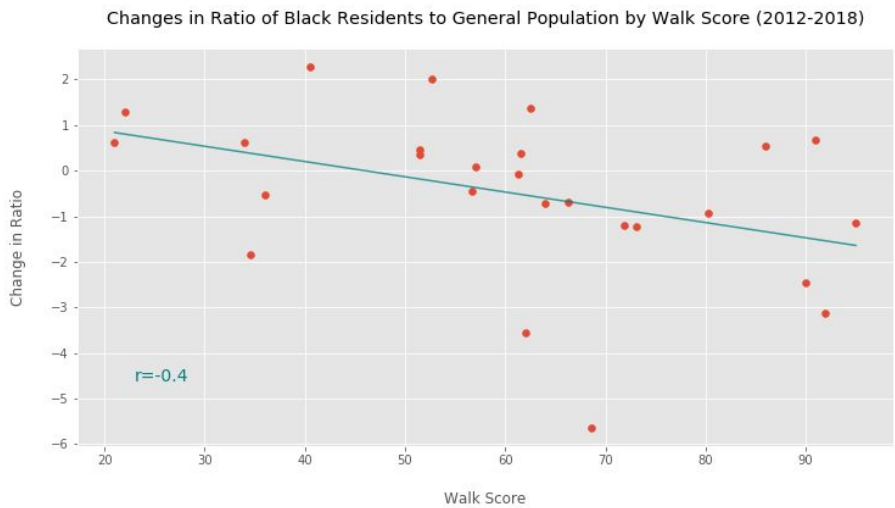
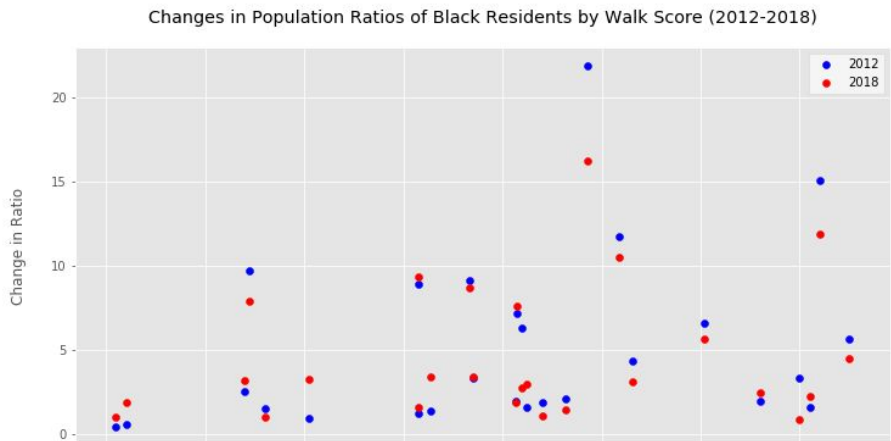
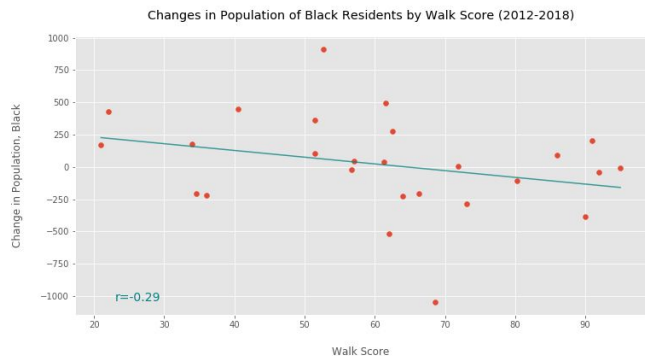
There is **little correlation** between walk score and general population change ($r=-0.19$) or aggregate BIPOC population change ($r=-0.14$).



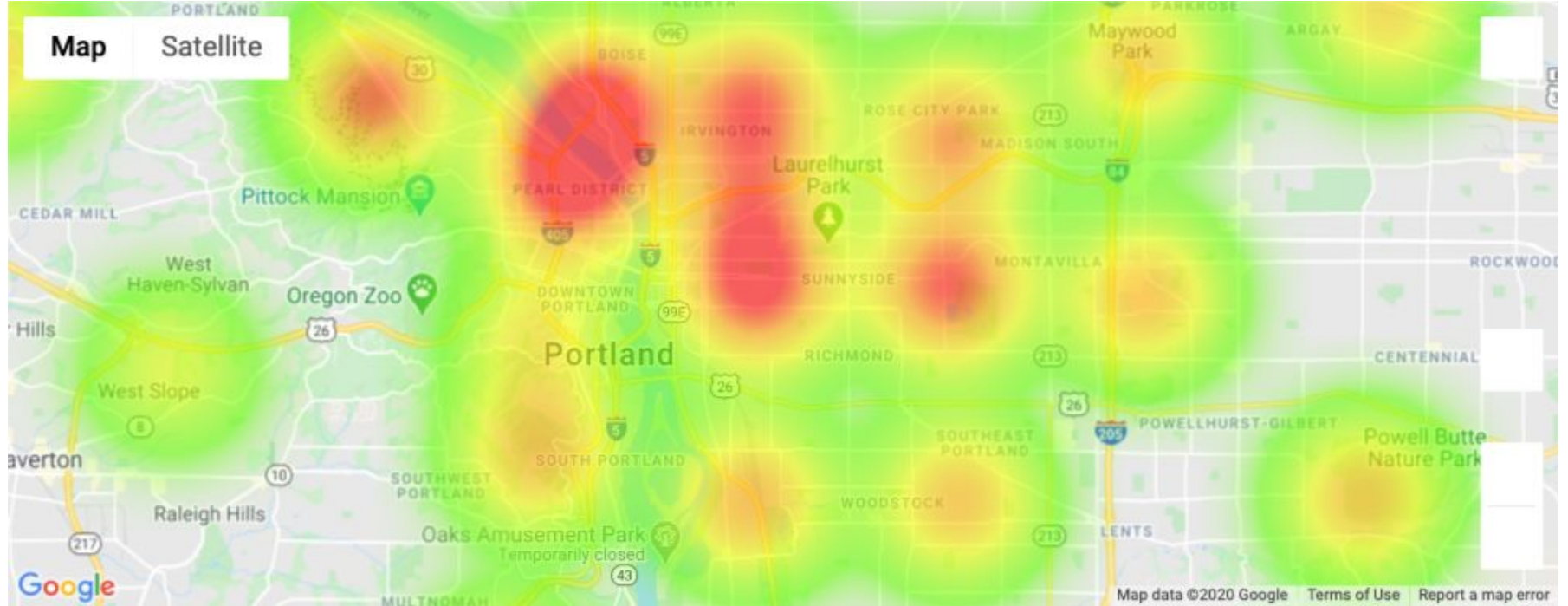
Ratio of Black Residents

Scatter plots with linear regressions

Zip codes with higher walk scores showed a **moderate negative correlation** with changes in Black community population ($r=-.029$) and a **strong negative correlation** ($r=-0.4$) with the decrease in the ratio of Black residents to all residents.



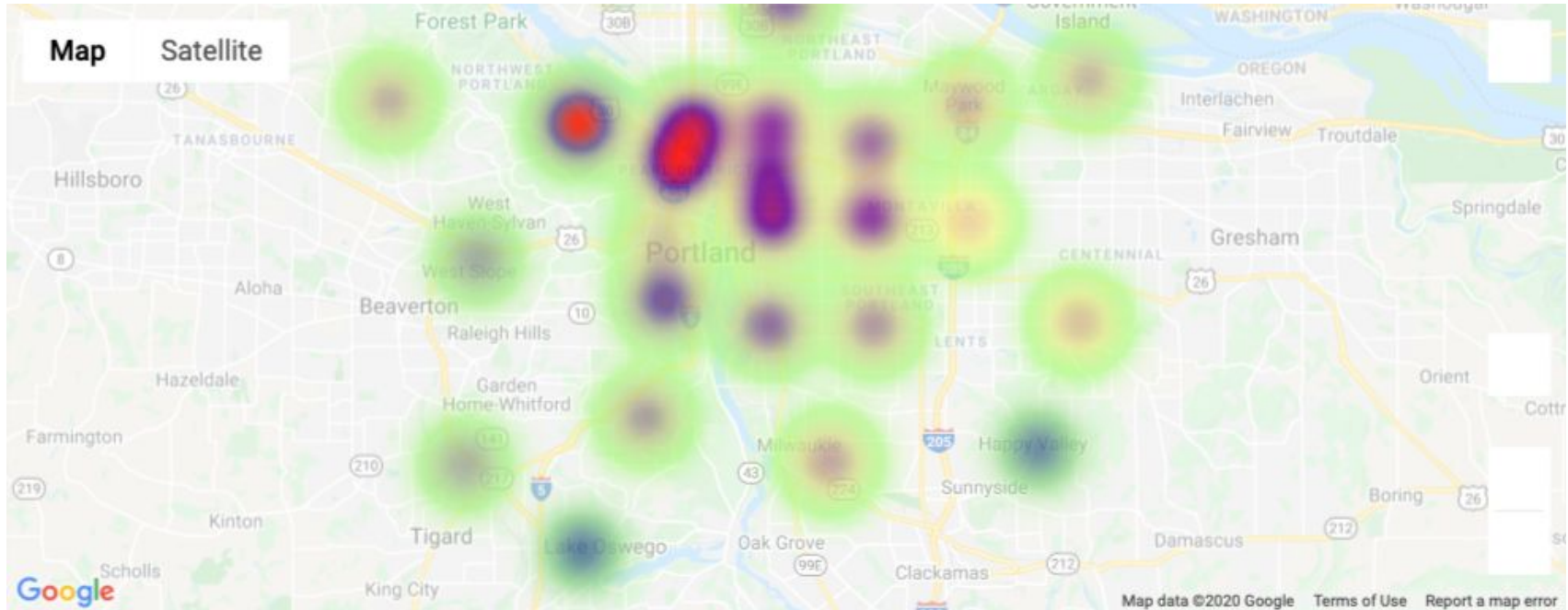
Average Walk Score By Zipcode



Walk Score and Median Income Growth

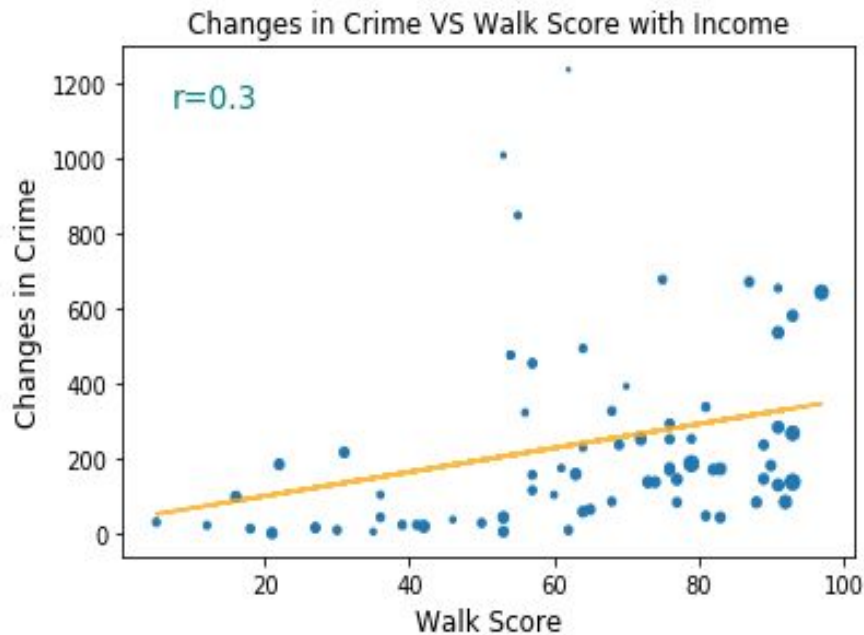
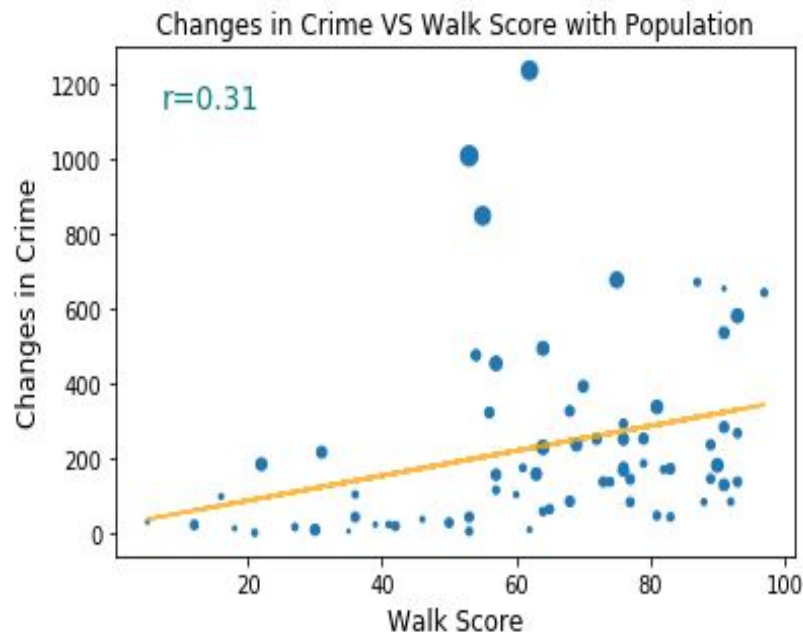
Walk Score = Red

Income Growth = Blue



Crime Data

Scatter plots with linear regressions



Post Mortem

What would we research next, if you had two more weeks?

1. Add a time metric for walk score
2. Look deeper into racial mobility (adjust for births and deaths)
3. Find elusive health metrics and expensive foot traffic data
4. Add heat map of Black population ratio change
5. Integrate crime data analysis into the project more

Data Difficulties

- Issues encountered:
 - Pedestrian data is not easily available, except very limited samples from few cities
 - The census data has over 40,000 poorly coded variables available to call
 - Many census variables exist in name only and contain no data
 - Incorrect data scale for analyzing by neighborhood, aggregating neighborhoods into zip codes to average walk scores
 - Finding supporting datasets as we explored our initial data findings became harder as we narrowed down our scope
 - Arc-GIS and vector data
 - Neighborhood nomenclature across different data sets



Questions?

In Portland Neighborhoods from 2012-2018...

- ★ Higher walk score neighborhoods got richer quicker.
- ★ Higher walk score neighborhoods showed no discernable difference in population change overall.
- ★ Ratios of Black populations in higher walk score neighborhoods dropped significantly.