Walkability, Income, and Race in Portland

Team Members:

- Troy Ramsey
- Rahil Nath
- Tony Cipolle
- Nnamdi Nwajagu

Motivation & Summary:

Our project is to look at walkability scores across Portland Neighborhoods and compare them to demographic and economic mobility factors from 2012 - 2018.

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

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

- Do higher walk scores correlate to higher increases in median income?
- Do higher walk scores correlate to greater population change?
- Do population changes vary across race by walk score?

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.

Hypotheses:

- Higher walk scores correlate to higher increases in median income.
- Higher walk scores correlate to higher increases in population.
- Mobility into high walk score neighborhoods differs across racial groups

Summary of findings:

- Zip codes with higher walk scores show a strong positive correlation with faster income growth (r=0.52). Median income grew at a much higher pace in more "walkable" neighborhoods. Even adjusting the data to filter out all outliers, there's still a substantial positive correlation (r=0.37). It would seem, therefore, that neighborhoods with more easily accessible amenities are pulling in people of higher socioeconomic strata.
- There is little correlation between walk score and general population change, as well as aggregate BIPOC population change (r=-0.19 and -0.14 respectively). All zip codes except for one grew during the observed period, but there is no easily apparent dependent dynamic with the variables that increase a walk score.
- Zip codes with higher walk scores showed a moderate negative correlation with changes in Black community population (r=-.029). Higher walk score zip codes either gained Black residents at a much slower pace than lower walk score zip codes or lost Black residents.
    - To support this finding regarding Portland Black communities, we also looked at changes in the ratio of a zip code's general population and it's Black population. Zip codes with higher walk scores showed a strong negative correlation (r=-0.4) with an overall decrease in the ratio of Black residents to all residents by zip code.

Metrics (data source):
- Walk Score by neighborhood (Walkscore.com)
- Neighborhood by Zipcode (Github dataset)
- Lat long by Neighborhood (City of Portland Open Data)
- Lat/Lng by Zipcode (USGS.gov dataset)
- Change in Median Income (US Census API)
- Change in Total Population (US Census API)
- Change in African American Population(US Census API)
- Other metrics (BRFFs SAS dataset)
- Crime rate data 2015, 2018 (City of Portland Police Bureau Open Data Source)

Data Cleanup & Exploration

- Extracted walk score data for each Portland neighborhood from Walkscore
  - https://www.walkscore.com/OR/Portland
- Used Portland OpenData Arc-GIS data to populate latitude and longitude values for each neighborhood
  - http://gis-pdx.opendata.arcgis.com/datasets/neighborhoods-regions
- Made reverse geocoding calls to the Google Geocoding API to assign neighborhoods to zip codes
- Merged our zip code/neighborhood data with our walk score data and dropped non-matched and null values
- Used zip codes to call the 2012 and 2018 US census estimates
  - Pulled median income, general population, BIPOC population, and Black population data for each merged zip code
  - Found differences in values between 2012 and 2018 to measure change over time
  - Aggregated neighborhoods into zip codes to weight each individual zip code with an average walk score to more closely match census data
- Pulled crime statistics open data from https://www.portlandoregon.gov/police/71978 for 2015 (historical only going back 5 years) and 2018
- Used merged zip code/neighborhood data to clean neighborhood names from crime statistics open data
- Merged cleaned crime data with walk score data to plot differences

Data Cleaning:

- Insights we had while exploring the data that we didn't anticipate:
  - Pedestrian data is not easily available, except in very limited samples from a few cities
  - The census data has over 40,000 variables available to call, all of which are not clearly coded, and the accompanying CSV key of those variables is sluggish and difficult to explore
  - Many census variables exist in name only and have no data in them. Occasionally, some variables we wanted to explore returned only null values, and further calls to check into those values showed that ALL values across the United States are null
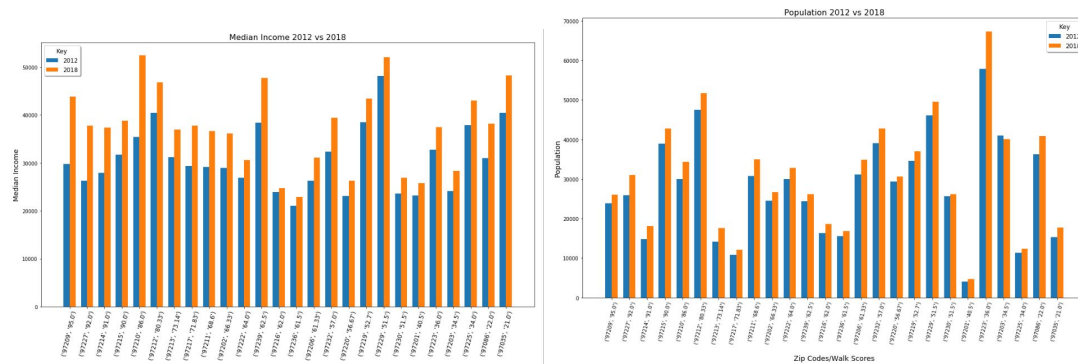
- Problems that arose after exploring the data, and how we resolved them:
    - Most demographic and socioeconomic data is not available by neighborhood, so we had to aggregate our walk scores into zip codes, reducing our available points of plotable data
    - Finding supporting datasets as we explored 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
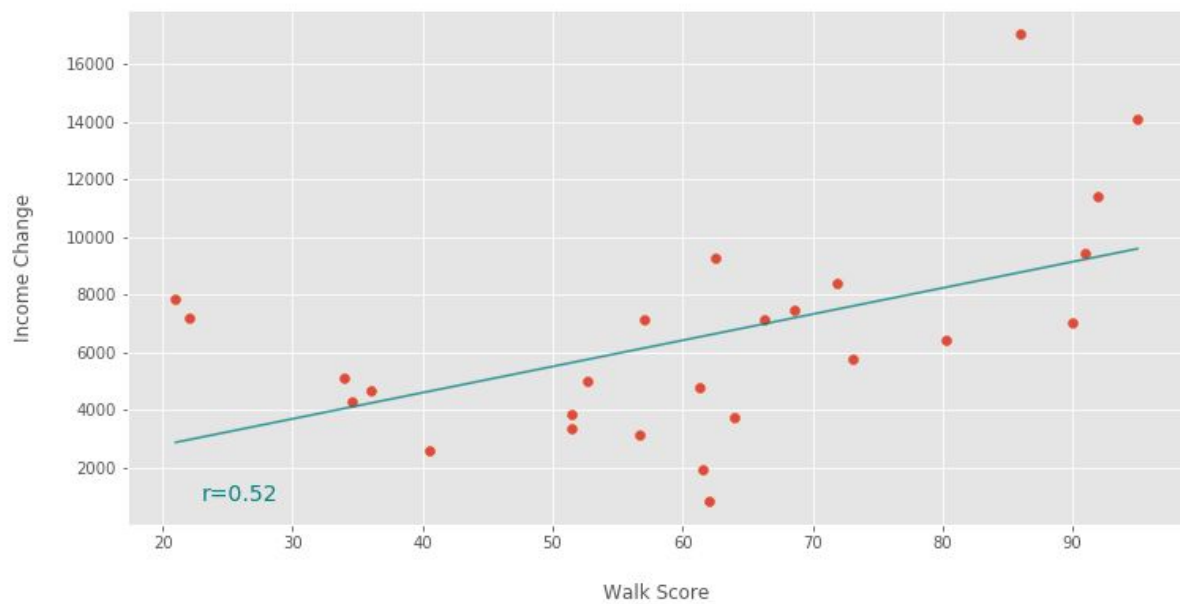
- Built a boxplot to understand potential outliers in median income data, then presented the same linear regression model with and without median income outliers
- Stacked two sets of Black population ratio data into a shared scatterplot to illustrate change in each pair of ratio values
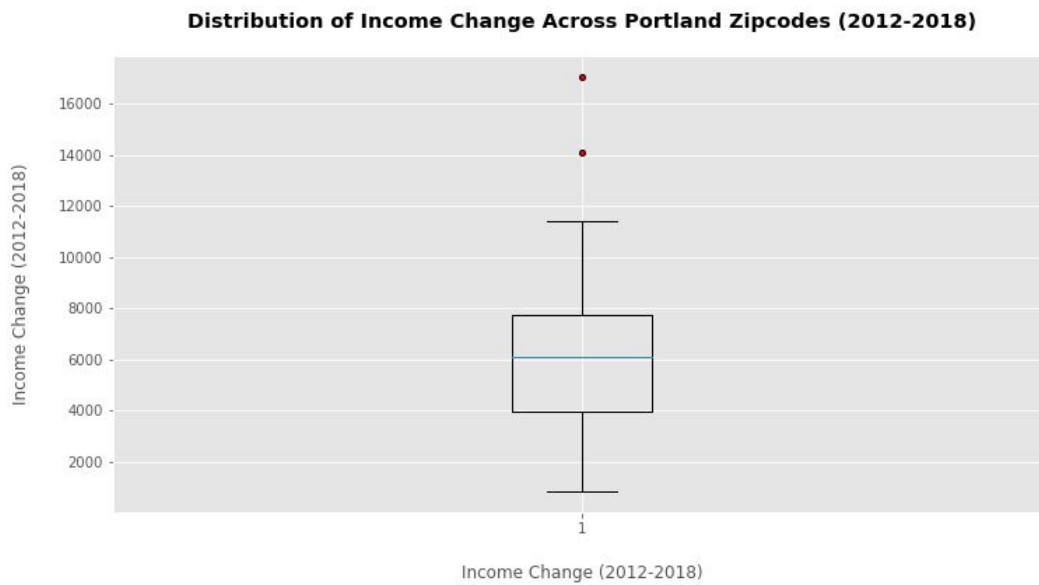
Data Visualizations:

Changes in Median Income

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

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

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


Changes in Median Income (2012-2018) by Walk Score (Adjusted for Outliers)
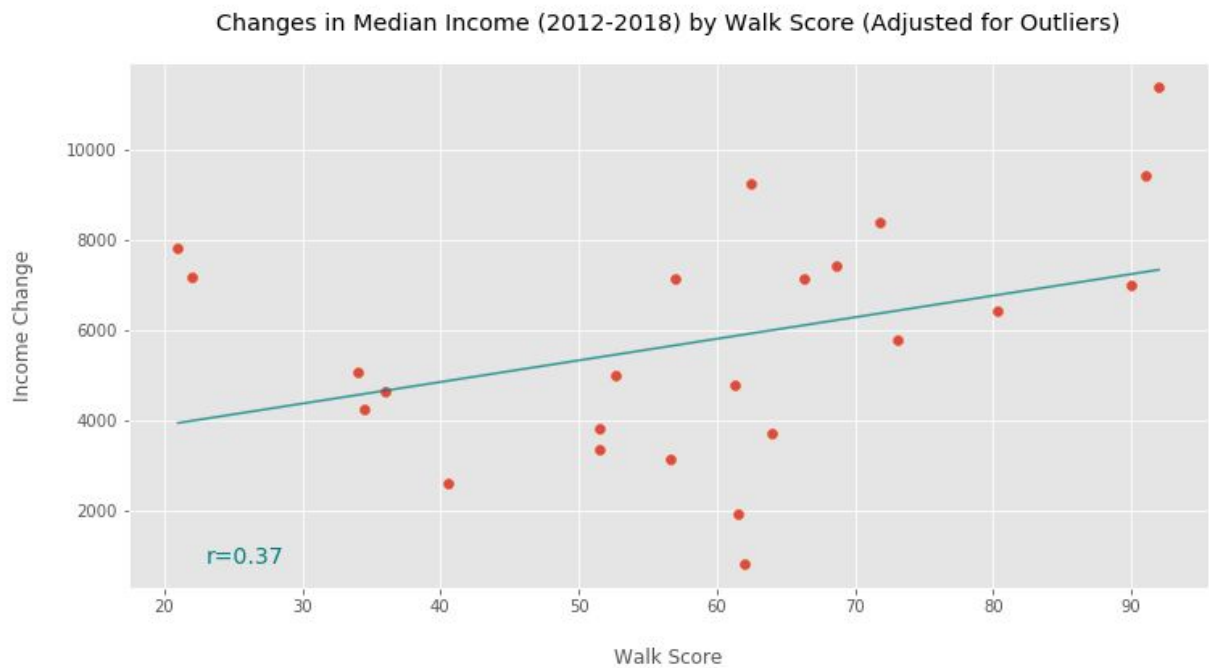
Population Change by Walk Score
Scatter plots with linear regressions


Changes in General Population by Walk Score (2012-2018)
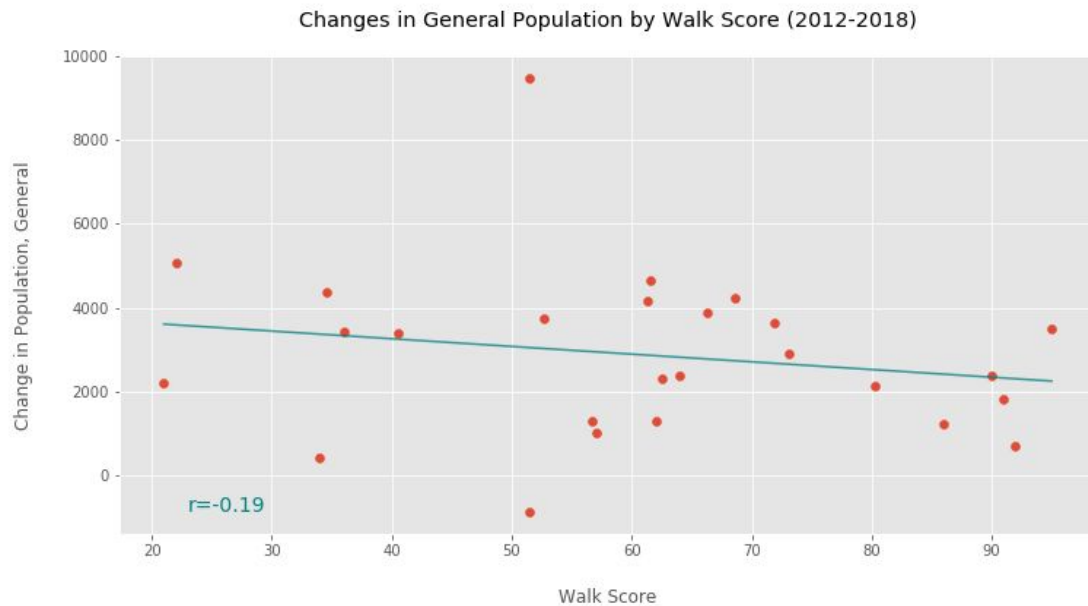
There is **little correlation** between walk score and general population change (r=-0.19) or aggregate BIPOC population change (r=-0.14).

Changes in Population of BIPOC Residents by Walk Score (2012-2018)
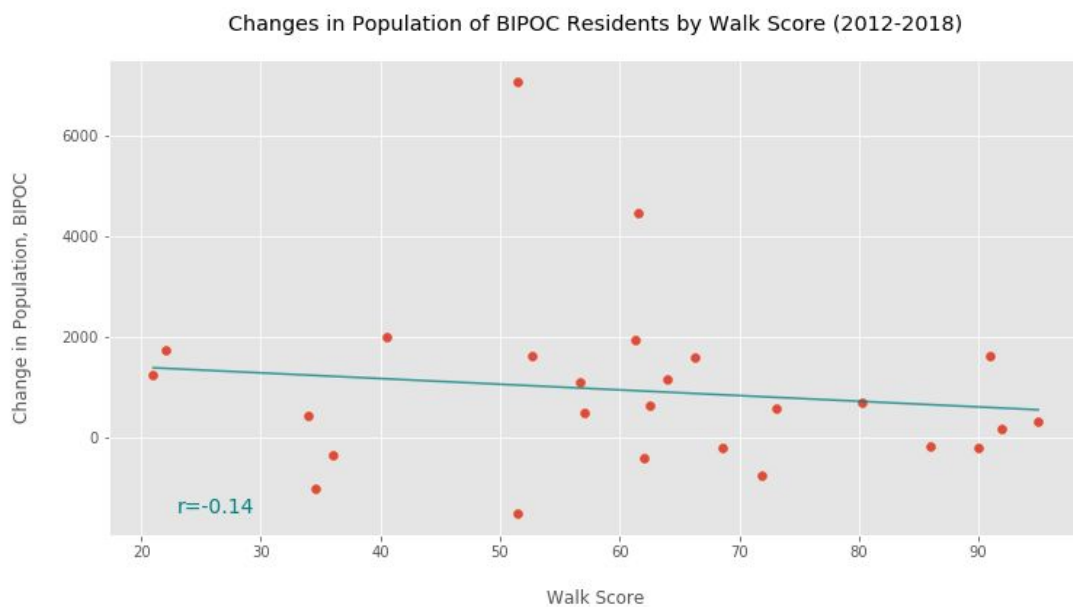
Change of Ratio of Black Residents by Walk Score
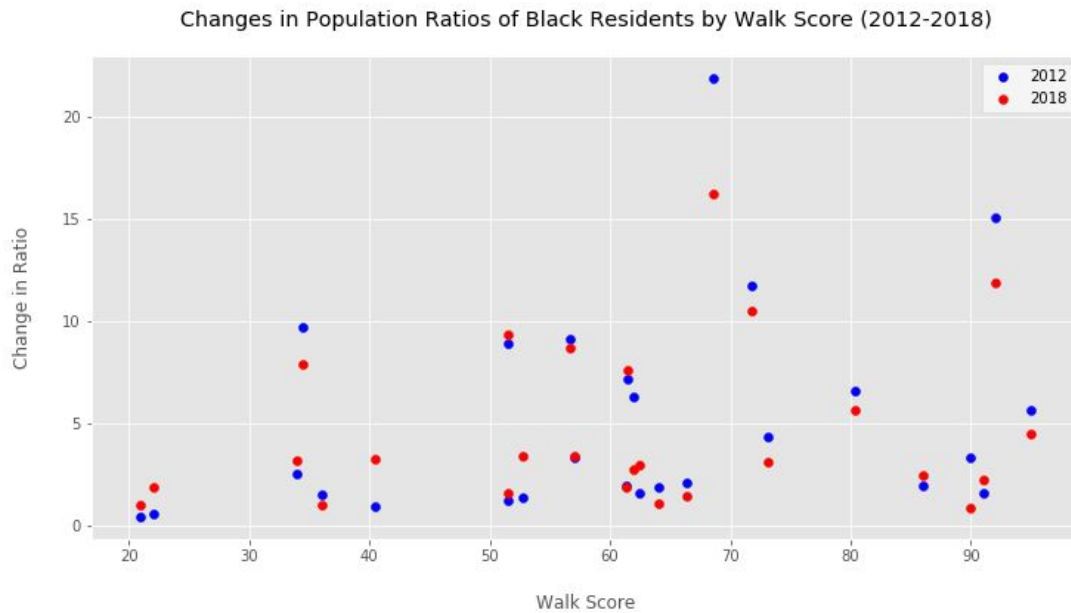
Changes in Population Ratios of Black Residents by Walk Score (2012-2018)

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.

Changes in Population of Black Residents by Walk Score (2012-2018)

Changes in Ratio of Black Residents to General Population by Walk Score (2012-2018)
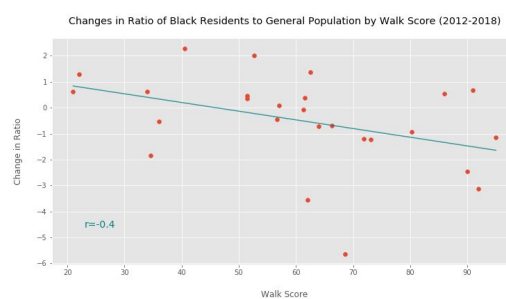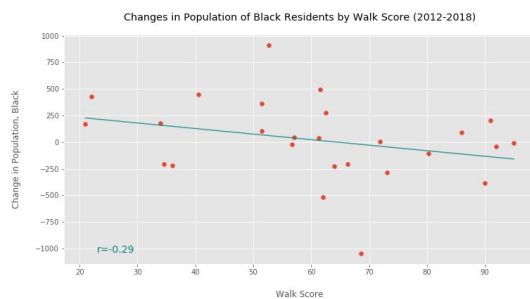
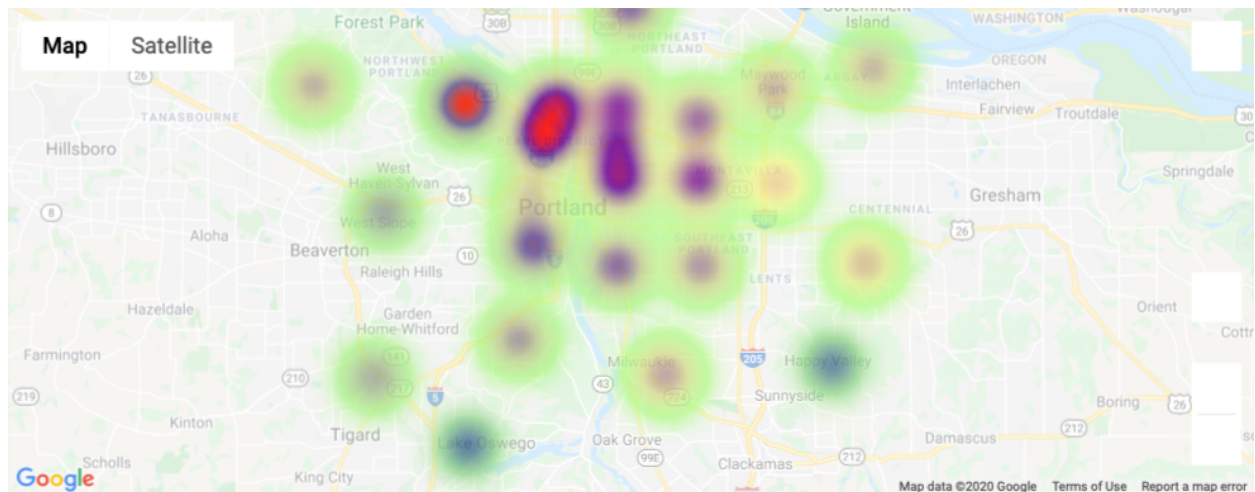## Average Walk Score By Zipcode
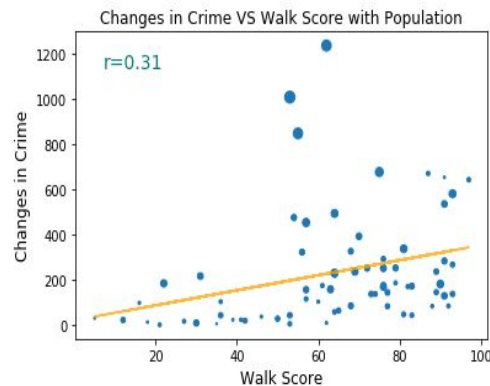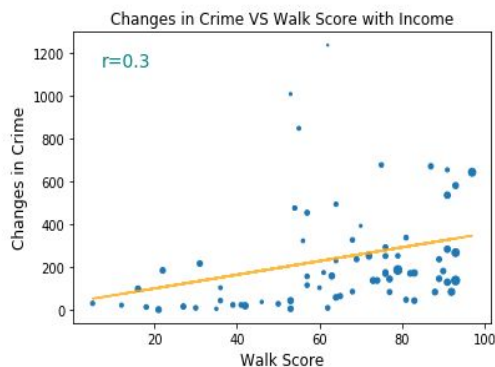
Walk Score = Red
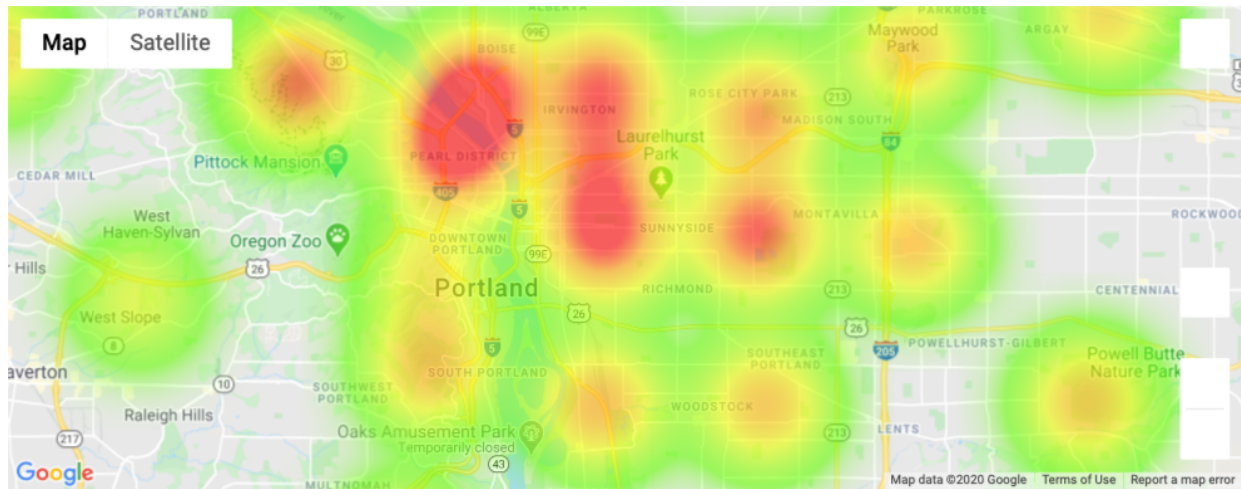
Walk Score and Median Income Growth by Zipcode

Walk Score = Red
Income Growth = Blue

Crime Data by Walk Score

Statistical Analysis:

Higher walk scores strongly correlate (r=.52) with greater increases in median income in Portland between 2012 and 2018.

Walk scores show little correlation with overall population change (r=.17)

Higher walk scores show moderate correlation (r=-.29) with African American population changes

- Zip codes in the Portland area with higher walk scores have trended towards seeing greater increases in median income in the period between 2012-2018.

- There's a small degree of clustering in the middle walk scores, but the income increases in the higher walk scores thoroughly outpace the increases in the central cluster enough to strongly influence our best fit line.

- There is a weak correlation between higher walk scores and lower amounts of general population influx.

- While all zip codes but one showed increases in population, zip code areas with higher walk scores show a slight tendency to grow slower.

- Higher walk scores show a moderate negative correlation with population changes in the Portland African American community.

- Neighborhoods with higher walk scores are more likely to show a loss of Black residents over the 2012-2018 period.

Post Mortem

What would we research next, if we 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

In conclusion

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.