

Capstone Project 1 – Milestone Report: Geoff Diestel

Q: How significant is the relationship between racial diversity and crime rate for neighborhoods in a largely white city like Portland Oregon?

Q: What other factors tied to a neighborhoods crime rate are affected by racial diversity, if any?

As a resident of Greater Portland Oregon Metro area, I am well aware of the ongoing homeless crisis and dissolving middle class in the U.S. which is easily seen in the Portland area. It is becoming ever more difficult to purchase a home in the greater Portland area and a large population growth has been treated with more and more rental properties instead of affordable housing for purchase.

The above questions are considered to help show that race is less related to crime in Portland than other socio-economic factors related to the availability of home ownership for those in a disappearing middle class. As such, this report is aimed at City officials and planners interested in lowering crime. The analysis will provide some evidence that although race is an important issue, especially in the context of economic opportunity, it is the creation more reasonably priced homes which could make the greatest impact on a cities well-being. You can't keep building only mansions and over-priced apartments and then be surprised crime and staggering homelessness.

DATA WRANGLING

The 2010 census data for the city of Portland given by neighborhood can be found at <https://www.portlandoregon.gov/civic/56897> while 2015 crime data by neighborhoods can be found at <https://www.portlandoregon.gov/police/71978> . Census data is downloadable as Excel spreadsheets while crime data is downloadable as text. Some reformatting of the notebooks is necessary before conversion into CSV files but the crime data is easily converted into CSV without creating any data problems.

The crime data comes by report of crime with a number of location identifiers like neighborhood and latitude and longitude. Also give is the category and type of crime as well as the number of counts tied to each report. A number of identifying codes and identifiers tied to individuals were removed as focus on the report is crime rates per neighborhood in relation to demographic features per neighborhood which come from the census data, already organized by neighborhood.

The first step was to determine a reasonable manner in which to fill in missing neighborhood labels from crime reports. A very small proportion of crime reports were missing both neighborhood names and geographic coordinates. However, there were a reasonable number of reports missing neighborhood names only. These were determined by using the geographic coordinates from each report and the geographic centers of the neighborhoods to determine how to fill in the missing neighborhood name. Once this was done, the total crime counts per neighborhood were aggregated into a pandas data frame.

Capstone Project 1 – Milestone Report: Geoff Diestel

Offense Count	
Neighborhood	
ALAMEDA	104
ARBOR LODGE	328
ARDENWALD	20
ARGAY	398
ARLINGTON HEIGHTS	66

2015 Aggregate Crimes Header.

The census data came with dozens of variables which fit into one of the following general categories related to race, sex, age, types of housing units, and types households types related basic family structures living in a single unit. These were aggregated to produce proportions per neighborhood:

- **Race:** White, African Am., Hispanic, Asian, Other.
- **Sex:** Males
- **Age:** Seniors and Minors.
- **Economic:** Population in Owner Occupied Units and Rental Units.
- **Social:** Single Male/Female(w/wo children) and Couples(w/wo children).

Neighborhood	Crime Rate	White Prop	Affrican Am Prop	Hispanic Prop	Asian Prop	Other Race Prop	Male Prop	Minor Prop
ALAMEDA	0.019946	0.885501	0.021481	0.033372	0.023974	0.035673	0.483506	0.223245
ARBOR LODGE	0.053307	0.764180	0.048594	0.068909	0.049732	0.068584	0.499919	0.145295
ARDENWALD-JOHNSON CREEK	0.004212	0.864575	0.018324	0.057077	0.017481	0.042544	0.483151	0.191660
ARGAY	0.066267	0.532967	0.125042	0.172494	0.102564	0.066933	0.487013	0.215285
ARLINGTON HEIGHTS	0.091922	0.874652	0.009749	0.032033	0.043175	0.040390	0.502786	0.193593

2015 Demographic Proportions(all columns not shown).

To combine the crime and demographic data by neighborhood, a unified list of neighborhood names had to be created. As both data sets used official city designations for neighborhoods, there was a clear one-to-one correspondence between the labels in each data frame with the exception of some neighborhoods sub-divided. All subdivided neighborhoods were combined to create a unified list of labels in both data frames and all labels were capitalized for uniformity. In some cases, a shorter or longer version of the name was used so thes had to be unified as well. All of this was done quite easily using Python.

Capstone Project 1 – Milestone Report: Geoff Diestel

Neighborhood	Crime Rate	White Prop	African Am Prop	Hispanic Prop	Asian Prop	Other Race Prop	Male Prop	Minor Prop
ALAMEDA	0.019946	0.885501	0.021481	0.033372	0.023974	0.035673	0.483506	0.223245
ARBOR LODGE	0.053307	0.764180	0.048594	0.068909	0.049732	0.068584	0.499919	0.145295
ARDENWALD-JOHNSON CREEK	0.004212	0.864575	0.018324	0.057077	0.017481	0.042544	0.483151	0.191660
ARGAY	0.066267	0.532967	0.125042	0.172494	0.102564	0.066933	0.487013	0.215285
ARLINGTON HEIGHTS	0.091922	0.874652	0.009749	0.032033	0.043175	0.040390	0.502786	0.193593

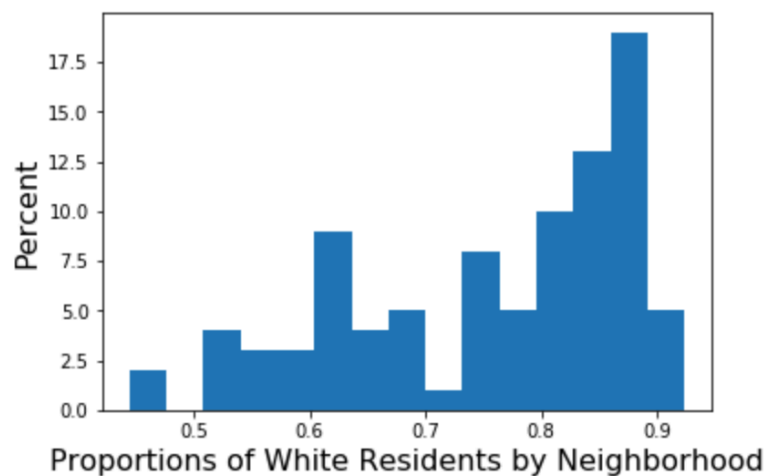
Combined Crime Rate/Capita and Demographic Proportions(all columns not shown).

EDA AND STORYTELLING

Initial EDA revealed a small number of outliers which were removed from the data.

1. **Northwest Industrial Neighborhood:** This non-residential neighborhood, like in many cities, has an extremely large crime rate per capita and must be removed.
2. **Lloyd Neighborhood:** This is the downtown non-residential district which is center of much nightlife and big events in Portland.
3. **Sunderland & Old Town/Chinatown Neighborhoods:** Both had extremely high proportions of male residents, more than 6 standard deviations from the mean. Although these neighborhoods have some residential population they are largely non-residential as well.

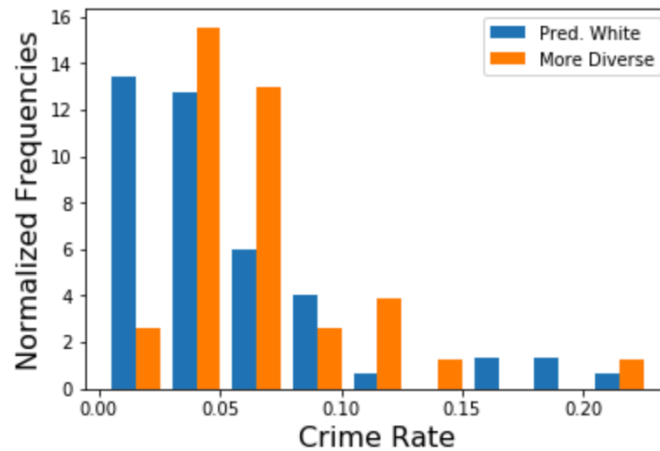
The large proportion of white residents in Portland means defining a useful measure of racial diversity is relative to the proportion of white residents in a neighborhood.



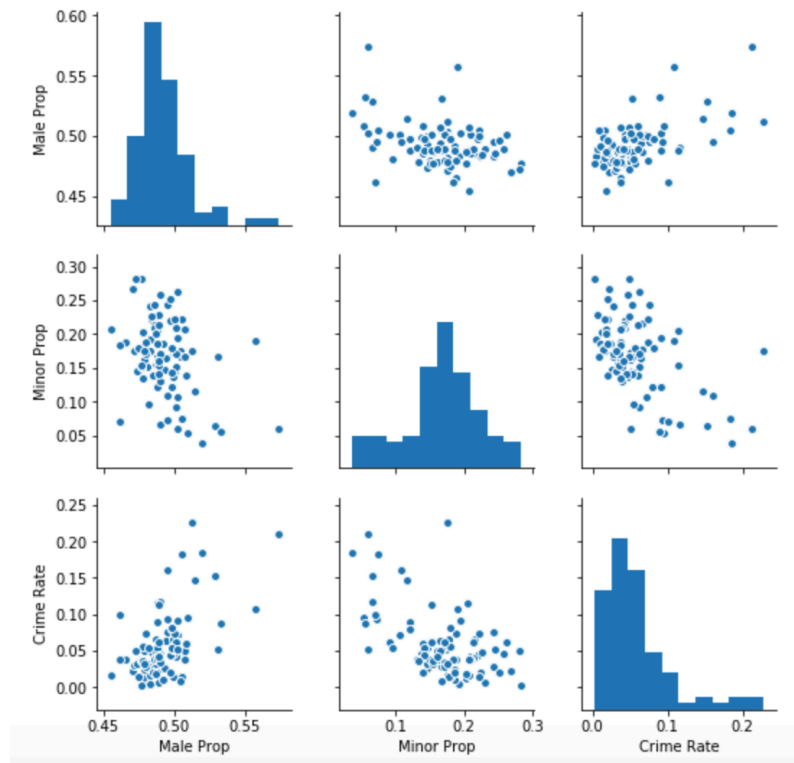
The above histogram provides a somewhat natural Boolean division of neighborhoods by the proportion of white residents. Those neighborhoods with greater than 73% white residents will be called Predominantly White(PW) while the other neighborhoods will be called

Capstone Project 1 – Milestone Report: Geoff Diestel

More Diverse(MD). With this division of racial diversity, it is natural to seek any visual distinction of crime rate between the PW and MD neighborhoods.

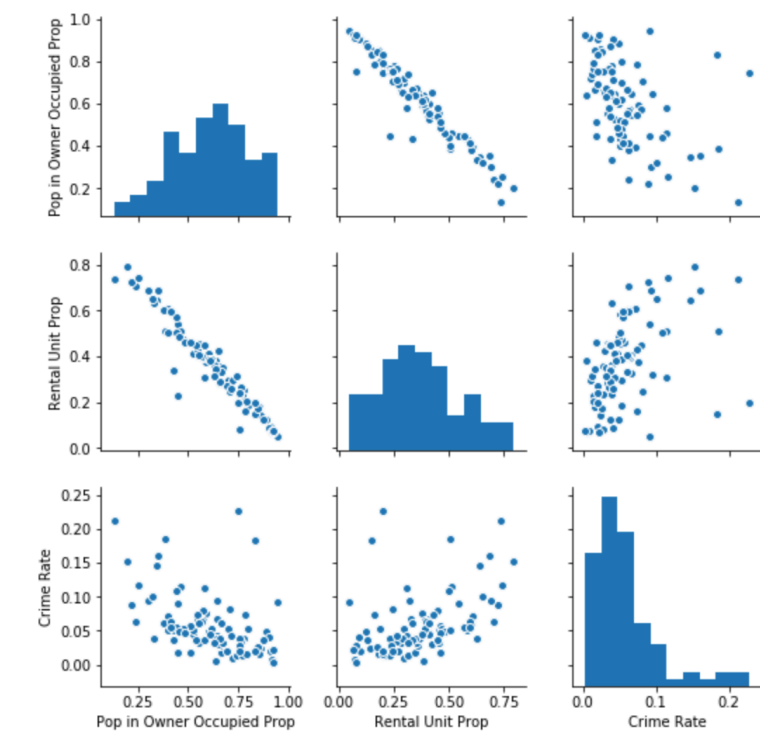


Visually speaking, there appears to be a higher average crime rate in the more diverse neighborhoods but only on the lower end of the crime rate spectrum. However, a visual inspection of these possible differences in mean crime rate appear to be less than 5 crimes per 100 residents in 2015. Visual EDA eliminates a number of less important features and identifies features with potentially more significant relationship to a neighborhoods crime rate.



Capstone Project 1 – Milestone Report: Geoff Diestel

The above scatter-plots suggest a positive correlation between the proportion of male residents and crime as well as a negative correlation between the proportion of minors (age < 18) and crime. There are some natural considerations which can explain these observations without the need of statistics. Testosterone has been clearly shown to be positively correlated with aggressive behavior so larger proportions of males having a positive correlation to crime is not a surprise in general. Moreover, neighborhoods with larger proportions of children have fewer crime reports because the mischief of children may not be reported as criminal behavior so to speak. Also, families with economic means tend to congregate in more economically stable neighborhoods so an economic factor may be another reason behind the negative correlation between the proportion of children and crime.



More non-surprising correlations to crime are the economic factors corresponding to the proportion of a neighborhood population living in owner occupied housing units and the proportion of rental housing units. Respectively, and as expected, these appear negatively and positively correlated to crime. Moreover, as one would expect, these features appear to have a very strong negative correlation with each other as their pairwise scatter-plots indicate most of the data points follow a distinct line. Although this is expected, it should be noted that this need not be the case in general as these proportions are different in nature. One is a proportion of residents while the other is a proportion of housing units. It is conceivable for a neighborhood to have a large number of single individuals renting while many families live in the homes they own. Such neighborhoods would contradict this relationship. Thus, it appears that Portland does not have a significant number of such neighborhoods.

Capstone Project 1 – Milestone Report: Geoff Diestel

Other features worthy of mention relate to the occupancy of housing units. The visual correlations with crime are not as visually obvious but some correlation does exist. Among these categories the number of single occupancy housing units tend to be positively correlated with crime, regardless of the sex of the individuals. Moreover, there appears to be little to no added significance related to the prevalence of single parent homes. Perhaps this is due to the fact that this has become quite common in today's society. Moreover, the proportion of housing units with couples sharing space appears to have a negative correlation with crime while little to no added significance appears to correspond to the proportion of such residents which have or do not have children. Initially, this could be explained by the simple fact that living alone is more of an economic burden than splitting expenses among two or more persons.

STATISTICALLY SIGNIFICANT FEATURES

Based on the visual EDA, various socio-economic features of a neighborhood appeared to have correlation with crime rate. We see which of these correlations have more statistical significance and test whether our definition of racial diversity for Portland neighborhoods has any statistical significance.

Finally, we look at neighborhoods with feature percentile scores in favor of lower crime rates to see if there is any statistical significance in classifying these as a recipe for a lower crime neighborhoods. For these expected low crime (ECL) neighborhoods, racial diversity will be considered to see if the combination of racial diversity and these neighborhoods with expected lower crime rates reveal any additional significance. We will see that there is little statistical significance related to crime and racial diversity in Portland neighborhoods. However, among ECL neighborhoods, predominantly white neighborhoods have a statistically much lower crime rate than the more diverse neighborhoods. However, this difference can be explained by economic factors.

H_0 : Mean PW Crime Rate = Mean MD Crime Rate.

H_a : Mean PW Crime Rate < Mean MD Crime Rate.

Comparing mean crime rates for PW and MD neighborhoods, two-sample, one-sided t-test provides a p-value of 0.08 with a Mean PW Crime Rate of 5.3 crime per 100 residents and a Mean MD Crime Rate of 6.6 crimes per 100 residents. Thus, with 5% significance, the two means do not appear to be different.

Staying with 5% significance, a Pearson test for correlation was run for each feature X with visual correlation to crime rate.

H_0 : X and Crime Rate are Independent.

H_a : X and Crime Rate are Dependent.

The features with statistically significant positive correlation with crime rate were identified as those with positive correlation coefficients and p-values smaller than 0.05.

Capstone Project 1 – Milestone Report: Geoff Diestel

1. Proportion of Males.
2. Proportion of Rental Units.
3. Proportion of Lone Male Households.
4. Proportion of Lone Female Households.

Similarly, the features with statistically significant negative correlation with crime rate are as follows:

1. Proportion of Minors.
2. Proportion of Population in Owner Occupied Homes.
3. Proportion of Households with couples and no children.
4. Proportion of Households with couples and children.

Using the above statistically significant features, percentiles rankings were used to create recipe for a lower crime neighborhood. In order to make sure the total proportion of these neighborhoods was a significant proportion of all neighborhoods, i.e. more than 30% of all neighborhoods, these expected low crime(ECL) neighborhoods are defined as any neighborhood with a percentile ranking smaller than 80th for all features positively correlated with crime and with a percentile ranking of more than 20th for all features with a negative correlation to crime. To determine if such ECL neighborhoods do have lower crime rates, the following test of the mean was conducted.

H_0 : Mean Crime Rate = ECL Mean Crime Rate.

H_a : Mean Crime Rate > ELC Mean Crime Rate.

With a p-value of 0.001, it is reasonable to accept that the ECL neighborhoods have a statistically lower mean crime rate than that of all neighborhoods. The ECL mean crime rate is 3.9 crimes per 100 residents compared to 5.8 crimes per 100 residents. Practically that is about 1 less crime every six months in an ECL neighborhood.

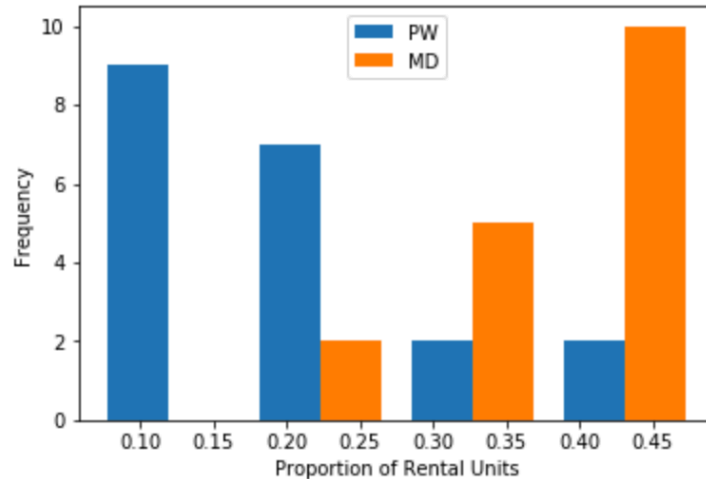
Among ECL neighborhoods, does our definition of racial diversity contribute to a lower crime rate?

H_0 : Mean PW Crime Rate in ECL neigh. = Mean MD Crime Rate in ECL neigh.

H_a : Mean PW Crime Rate in ECL neigh. < Mean MD Crime Rate in ECL neigh.

With a p-value of 0.001, it is reasonable to accept that predominantly white ECL neighborhoods have a lower average crime rate than their more diverse counterparts. In this case, the mean crime rates are 2.8 crimes per 100 residents and 5.2 crimes per resident respectively. However, this is easily explained by the following illustration comparing proportions of rental housing units for PW and MD among the ECL neighborhoods.

Capstone Project 1 – Milestone Report: Geoff Diestel



This is an extremely telling graphic as it shows that there are just about as many PW and MD neighborhoods among the ECL neighborhoods. However, lower proportions of rental units are skewed towards the more predominantly white neighborhoods.

In conclusion, ECL neighborhoods are those with fewer renters, fewer lone residents, and more children. These are neighborhoods designed more economically stable situations.

Where race was not very significant in relation to crime overall, it does appear to have an effect on crime within these more economically stable neighborhoods. However, this too is biased by a large number PW neighborhoods with very low proportions of rental units and no MD neighborhoods with such low proportions.

Of course much more analysis must be done and should be done in many more cities. However, given such analysis is anything like what has been shown above, a potentially key factor in lowering crime requires serious effort on both city planners and a city's population. Our cities need to provide less rental properties and smaller, cheaper homes so people can build more wealth and more stable communities. Moreover, people with less economic means have to take advantage of such housing and accept living within their means. I think many people would be happy to have the chance to buy smaller houses instead of throwing away money on rent. This strategy can reduce economic turmoil by eliminating much economic stress on many people just trying to pursue happiness, which is supposed to be guaranteed by the constitution.