Portland Crime Analysis

Ziling Zhen, Nick Starcevich, Omar Fitian

1/26/2024

Introduction

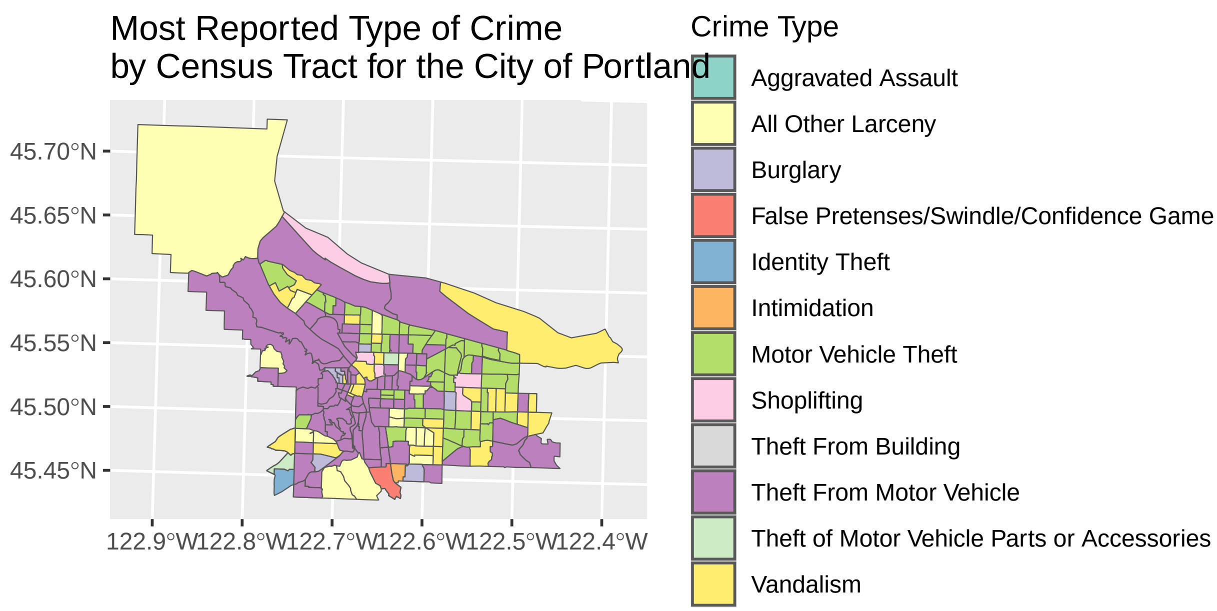
Portland, Oregon? What is so special about this place? Well, other than the fact that Z is from there, not much?… We knew we wanted to show and share the statistical knowledge we learned in this month by looking at crime statistics of a city and if the location of those crimes were affected by other variables of the area. We chose to look at the year 2020, a year of numerous challenges and events that left a lasting impact on the world; for the United States, we have the start of the covid-19 pandemic, lock downs, social distancing, Black Lives Matter protests, and the presidential election. We should be mindful of these events when looking at our crime data.

A little bit about the data that we used, there were a lot of joins used and *some* data manipulation, but the boundary lines we decided to use for the city of Portland are based off the decennial census tracts. From the glossary of Census.gov, "Census tracts are small, relatively permanent statistical subdivisions of a county.' that,"…generally have a population size between 1,200 and 8,000 people, with an optimum size of 4,000 people." For our tracts, there resides 5000 people in one. We also pulled income data from the ACS, which is the american community survey. It's goal is to keep a rough estimate of the population between each decennial census since with the ACS they don't go to everybody's door like the decenial census does. It also does not follow constitutional law made for the decennial census so it can include more information and it does not count toward congressional seats.

Our crime data is from the City of Portland itself, in the year 2020, within our boundaries, there were a total of 52,644 crimes reported. Our overarching goal was to look at the relationship between crimes and location across the city.

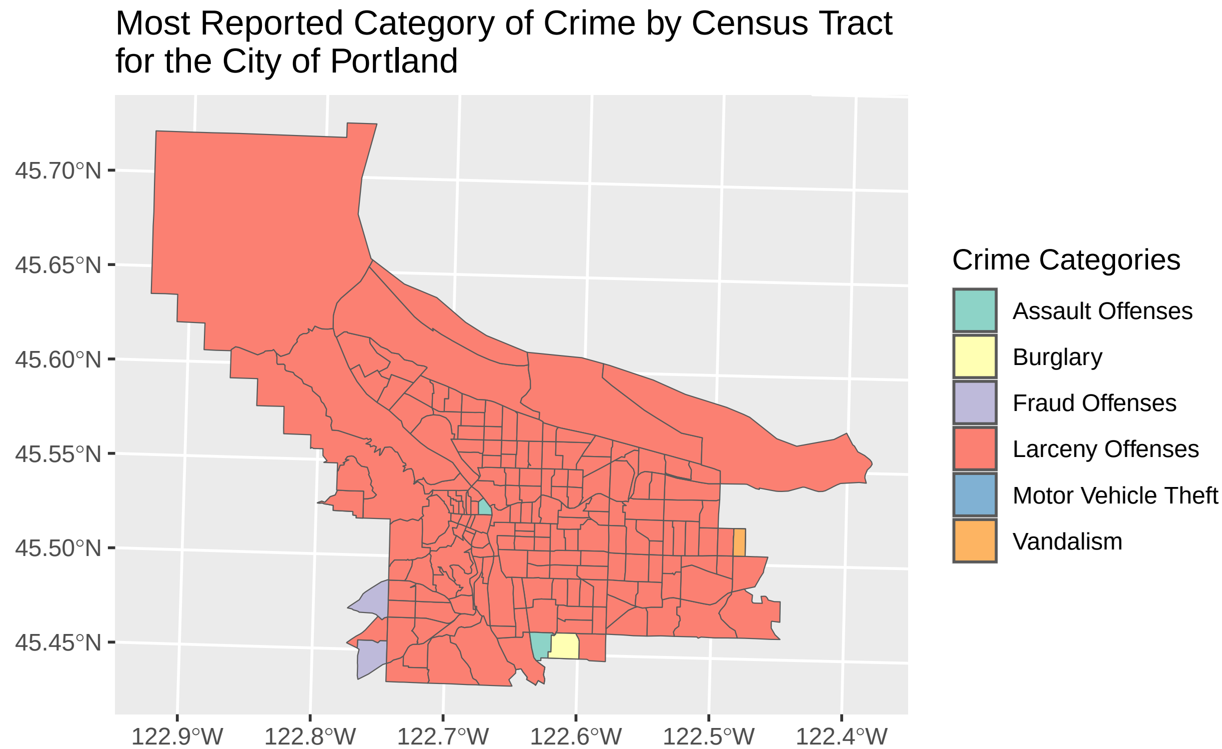
NAME	OffenseType	n
Census Tract 106.02, Multnomah County, Oregon	Vandalism	572
Census Tract 82.01, Multnomah County, Oregon	Shoplifting	254
Census Tract 50.01, Multnomah County, Oregon	Theft From Motor Vehicle	233
Census Tract 11.01, Multnomah County, Oregon	Theft From Motor Vehicle	222
Census Tract 23.03, Multnomah County, Oregon	Vandalism	203
Census Tract 22.03, Multnomah County, Oregon	Shoplifting	181
Census Tract 81, Multnomah County, Oregon	Shoplifting	169
Census Tract 51.03, Multnomah County, Oregon	Vandalism	160
Census Tract 72.01, Multnomah County, Oregon	Shoplifting	159
Census Tract 21.01, Multnomah County, Oregon	Theft From Motor Vehicle	157

Most Reported Type of Crime by Census Tract

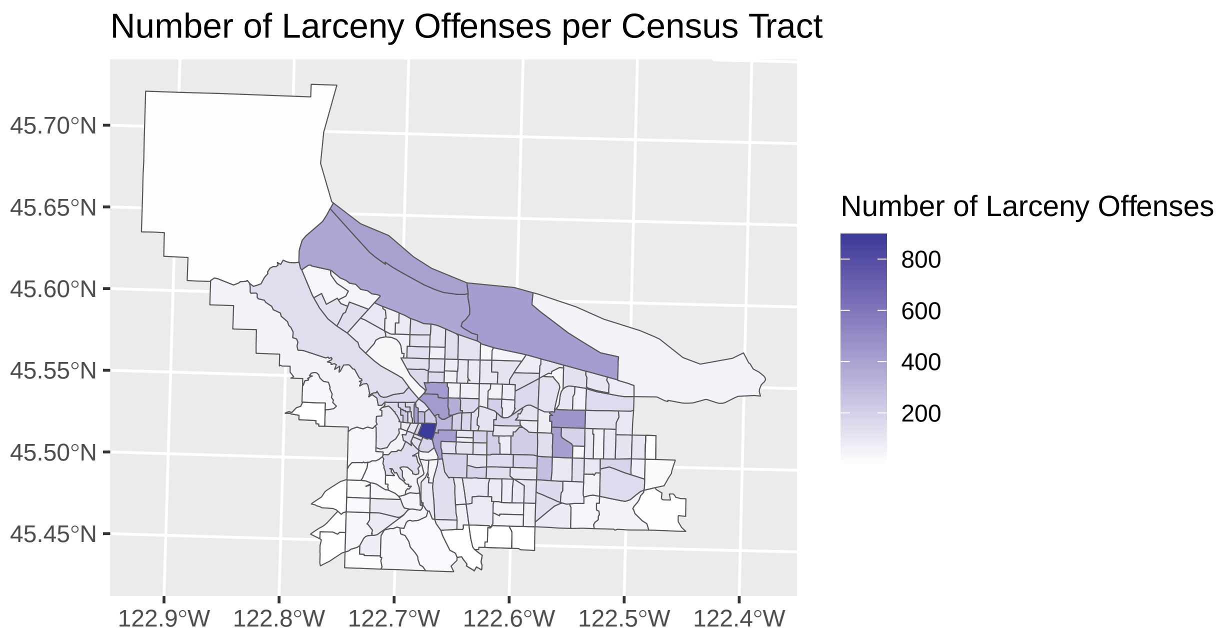


This was a graph we created showing the most reported type of crime per census tract, we can see that, the most common is theft from motor vehicle, which is theft of articles from a motor vehicle, whether locked or unlocked. There doesn't seem to be a lot of variety for the tracts, the most reported type of crime is related to *theft*.

A table for the top ten tracts with the most crimes committed. We can also categorize crime types into larger categories.



From this graph we can see that Larceny Offenses are the most reported category of crime, these offenses consist of pocket-picking, purse-snatching, shoplifting, theft from building, theft from coin-op machine or device, theft from motor vehicle, theft of motor vehicle parts, and all other larceny.



Looking further into just larceny offenses per census tract, we counted the number of larceny crimes and plotted it down in our tracts, at first glance it is apparent that there is spatial autocorrelation of these offenses because we can see that colors on the scale that are close to each other stay by each other. It doesn't look like the colors could be randomly generated on the map, but we cannot say this for sure until we calculate Moran's I.

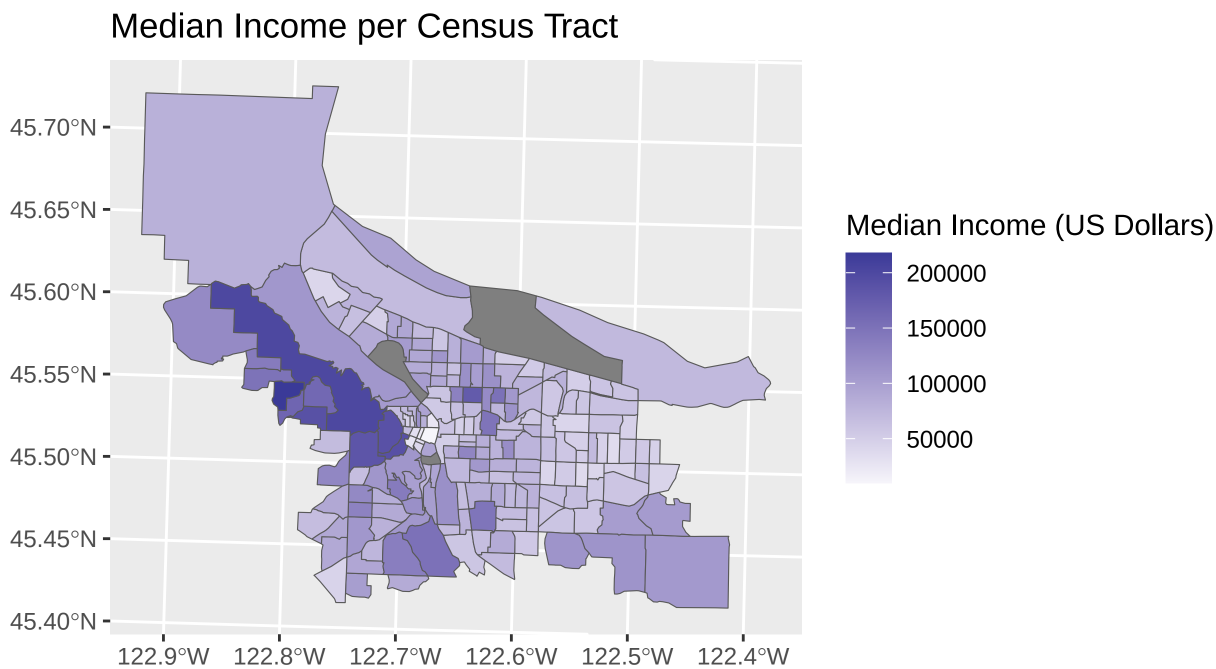
```
##
##  Monte-Carlo simulation of Moran I
##
## data:  pccc_nsf$n
## weights: pccc_nbw
## number of simulations + 1: 500
##
## statistic = 0.37878, observed rank = 500, p-value = 0.002
## alternative hypothesis: greater
```

Our Test

H_O: No spatial autocorrelation, I is close to 0

H_A: Spatial autocorrelation, I $\neq$ 0.

Calculating Moran's I for larceny offenses in using Monte Carlo's simulations we get a Moran's I of .37878 which is positive and moderately strong. We also got a p-value of 0.002. We can reject the null hypothesis and conclude that it is statistically significant that larceny offenses in Portland, Oregon is spatially autocorrelated. Which means a tracts number of larceny offenses reported is similar to its surrounding tracts.



The ACS lacked median income data for three of the census tracts in Portland, which is why in our analysis of Larceny Offenses we also did not include these three tracts. Looking at this map we can also see that there is spatial autocorrelation for the median income of Portland, Oregon, but again, we cannot say this for sure until we calculate Moran's I.

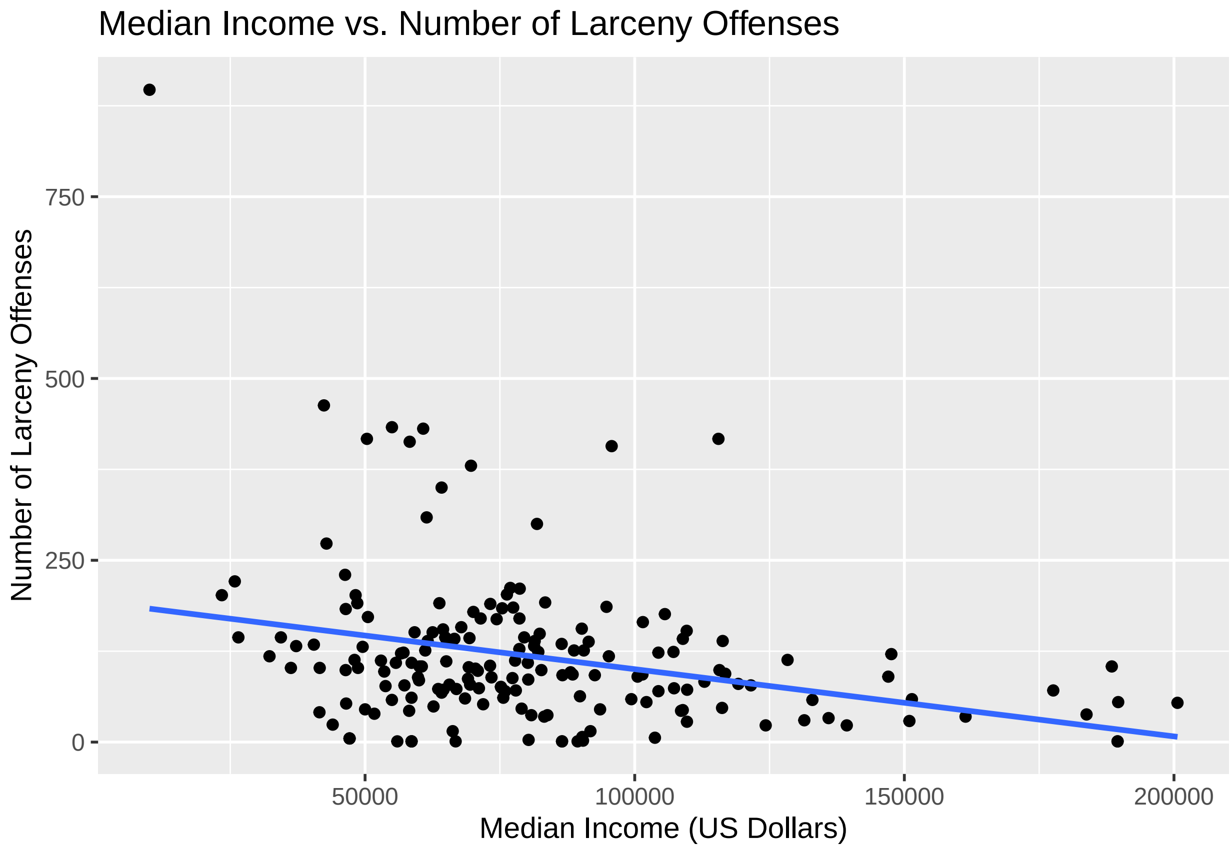
```
##
##  Monte-Carlo simulation of Moran I
##
## data:  pccc_nsf$medincome
## weights: pccc_nbw
## number of simulations + 1: 500
##
## statistic = 0.52218, observed rank = 500, p-value = 0.002
## alternative hypothesis: greater
```

Our Test (Again):

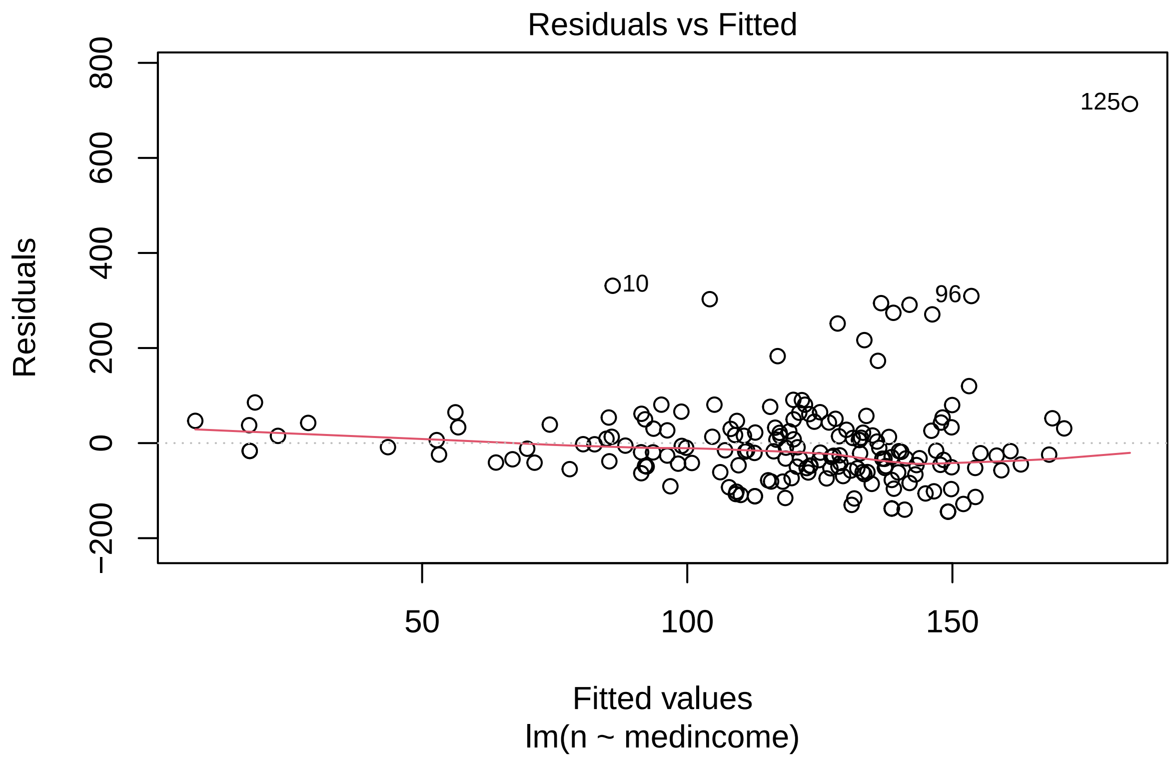
H_O: No spatial autocorrelation, I is close to 0

H_A: Spatial autocorrelation, I $\neq$ 0.

Calculating Moran's I for median income in using Monte Carlo's simulations we get a Moran's I of .52218 which is positive and strong. We also got a p-value of 0.002. We can reject the null hypothesis and conclude that it is statistically significant that median income in Portland, Oregon is spatially autocorrelated. Which means a tracts median income reported is similar to its surrounding tracts.



From this graph we can see a slight negative relationship between median income and larceny offenses so, unlike what some people might expect, as median income goes up, theft slightly goes down. However, it is worth noting that this trend may be effected by the slight leverage point for the downtown area on the left side of the graph.




```
##
## Call:
## lm(formula = n ~ medincome, data = regression_ci)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -144.18  -52.45  -17.61   30.78  713.52
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.928e+02  1.975e+01   9.758  < 2e-16 ***
## medincome   -9.247e-04  2.242e-04  -4.124 5.68e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 103.8 on 180 degrees of freedom
## Multiple R-squared:  0.08632,    Adjusted R-squared:  0.08125
## F-statistic: 17.01 on 1 and 180 DF,  p-value: 5.684e-05
```

The residual plot is reletively flat around the center suggesting linearity. The data appers to be somewhat normal considering there are a large number of points and they are also relatively centered around zero suggesting equal variance. A linear moder can be used, in particular, we will use the simple linear regression (SAR) model to account for the spatial aspect of the data.

```
##
```

```
## Call:lagsarlm(formula = n ~ medincome, data = regression_ci, listw = pccc_nbw)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -174.201  -49.345  -14.237   21.633  673.078
##
## Type: lag
## Coefficients: (asymptotic standard errors)
##                Estimate  Std. Error z value  Pr(>|z|)
## (Intercept)  1.0553e+02  2.1630e+01  4.8788 1.068e-06
## medincome   -6.0611e-04  2.0244e-04 -2.9940  0.002754
##
## Rho: 0.48977, LR test value: 34.913, p-value: 3.4469e-09
## Asymptotic standard error: 0.08277
##     z-value: 5.9172, p-value: 3.2738e-09
## Wald statistic: 35.014, p-value: 3.2738e-09
##
## Log likelihood: -1084.669 for lag model
## ML residual variance (sigma squared): 8310.5, (sigma: 91.162)
## Number of observations: 182
## Number of parameters estimated: 4
## AIC: NA (not available for weighted model), (AIC for lm: 2210.3)
## LM test for residual autocorrelation
## test value: 2.9358, p-value: 0.086635
```

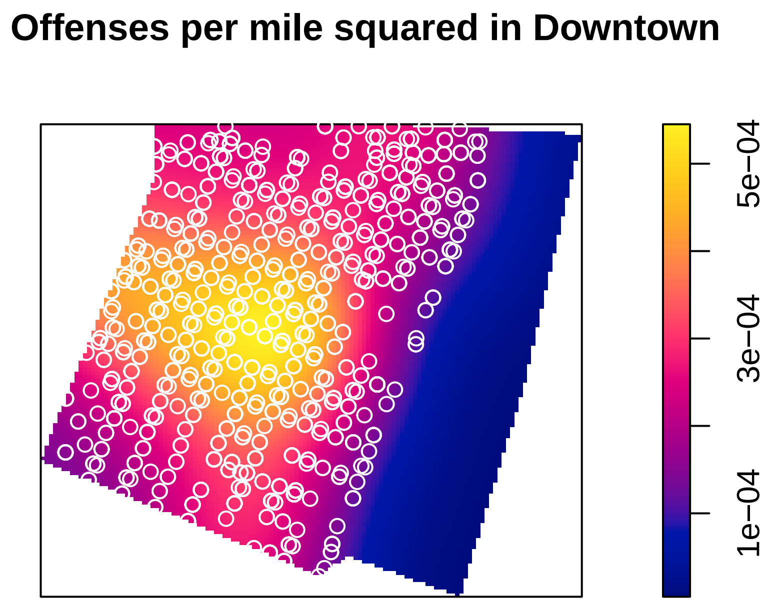
Using a lagged linear model detrended with median income to measure the auto spatial correlation of residuals, a Rho value of 0.487 suggests a positive moderate correlation. A p-value of 0.02 after running Monte Carlo simulations indicates statistical significance.

Tract Point Pattern

For this section, we zoomed in on the census tract that includes the Downtown area of Portland.

The grid layout of the area also produced a regular pattern in the offences.

However, from the density plot, we can clearly see areas of high and low intensity. The discrepancy between the center and other areas of Downtown exhibits an inhomogeneous process.



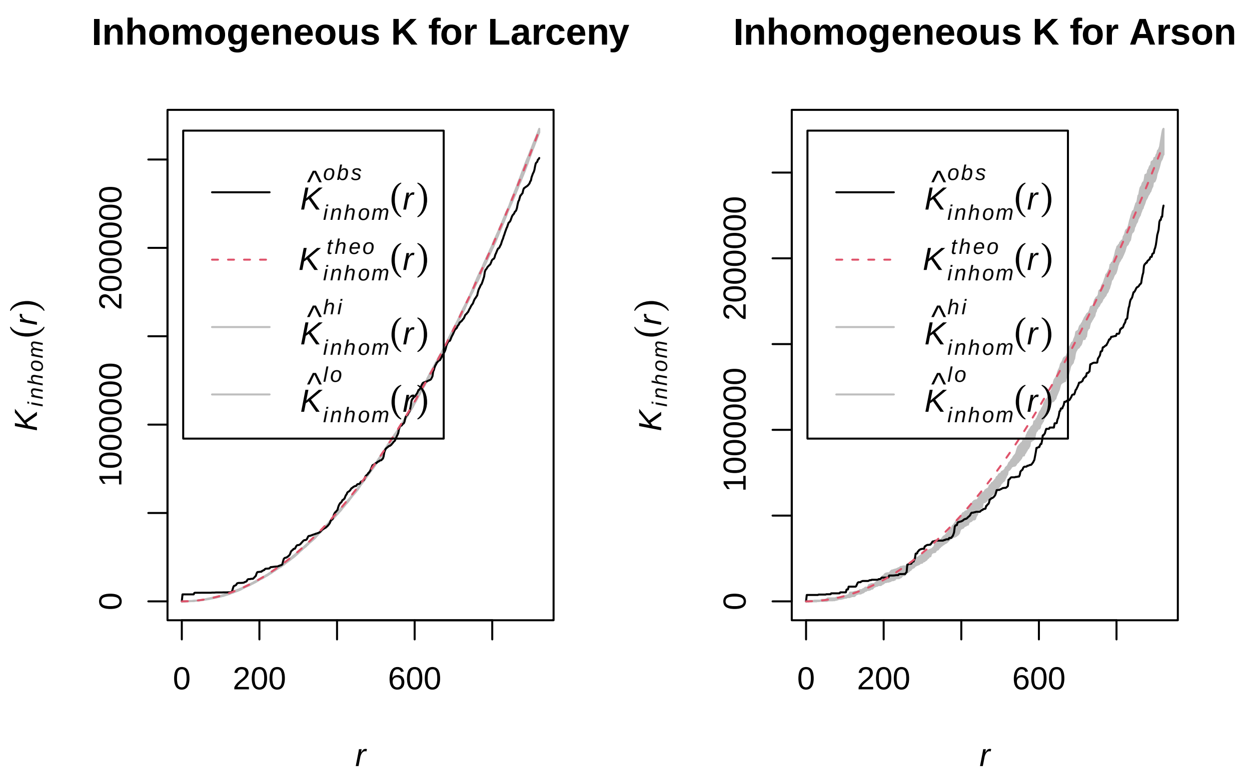
Our group was curious as to how some of the crime locations in 2020 were distributed in downtown Portland. In order to do this, we created a ppp object for the crime locations of different types in the census tract that covers the downtown area. We then performed the following K function analysis on a few different crimes to see how they were distributed.

We used the inhomogenous K function for the analysis since there is a river included on the east side of this tract and to absoluletly nobodys surprise, crimes aren't really happening on this flowing body of water. This is why the crime data for downtown looks homogenous except for that section on the east side containing the river. Since there is this inconsistency in crime locations throughout the census tract, the data is inhomogenous prompting us to use that type of K function for the analysis.

The inhomogeneous K function was used for analysing the point pattern of each of larceny, vandalism, and arson offences at multiple scales.

K Function Analysis for Larceny, Vandalism, and Arson PPP Objects

The first type of crime we examined in the downtown area was larceny because this was the most common crime in downtown Portland along with the majority of the census tracts in the City of Portland. The k function starts out above the envelope for a smaller radius from a given crime while for a larger radius, the k function for larceny in downtown Portland goes below the envelope. This means that on a small scale, the larceny crimes are clustered likely clustering at building locations while on a large scale, the larceny crimes are distributed with regularity. These patterns result in regularly spaced clusters of closely spaced offences. This means these clusters of points are distributed with regularity and that is likely due to the angular shape of the street grid in downtown Portland where each street intersects each other at a 90 degree angle so the buildings are distributed with regularity.

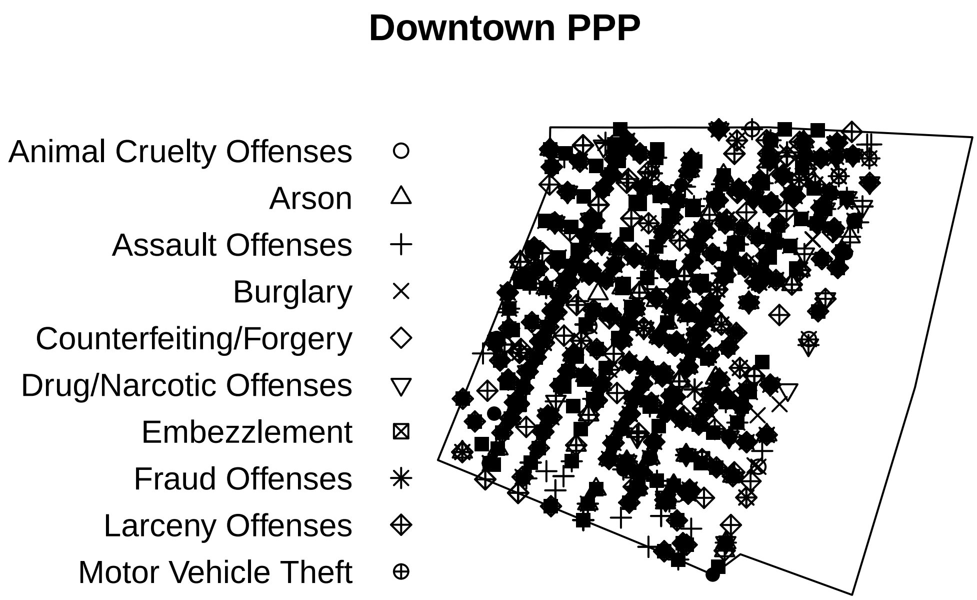


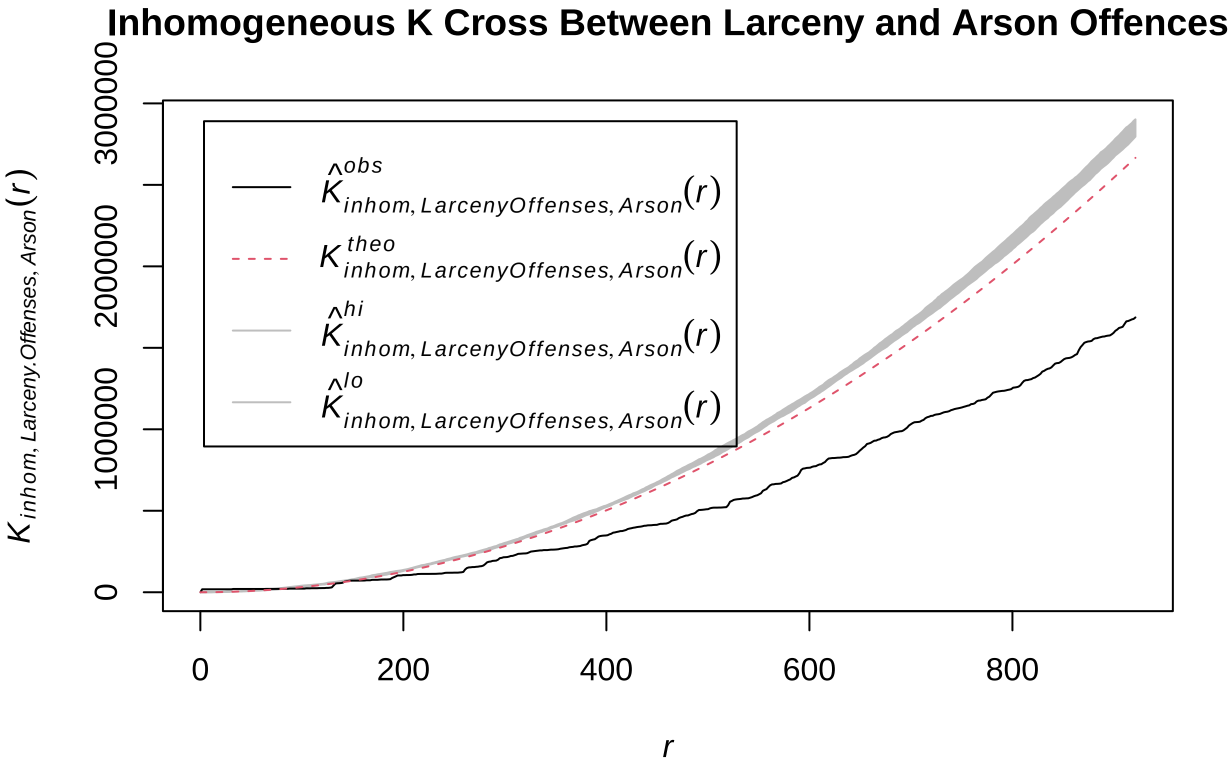
K Function Analysis for Arson PPP object

The next type of crime we examined in the downtown area was vandalism because this was the next most common crime in downtown Portland. We also examined arson since we though that was a more interesting crime type and may have had potential to have a different looking K function. Due to the regularity of buildings across this region due to the street grid though, we did not see very different k functions at all for these crimes where the function once again started above the envelope and finished below.

They aren't completely identical though. The vandalism point pattern exhibits similar behavior to larceny offences. However, its empirical function is closer to the theoretical function than larceny offences. This suggests that although it follows the same trend, it is weaker.

The arson offences have stronger repulsion at greater distances than either offences. This is seen through the larger difference between the empirical function and simulation envelope.


Downtown PPP
Animal Cruelty Offenses
Arson
Assault Offenses
Burglary
Counterfeiting/Forgery
Drug/Narcotic Offenses
Embezzlement
Fraud Offenses
Larceny Offenses
Motor Vehicle Theft



This Kcross function is almost entirely under the envelope suggesting at all scales the locations of arson and larceny crimes repel or in other words are distributed with regularity when combined at all scales. There is strong repulsion upward of a radius of 200, there are significantly less offences with a closest neighbor than expected by the null landscape.

Closure

Overall, although all this analysis was fun and interesting, we are unable to conclude causation for any analysis we've done. It's also worth noting that there were limitations in our analysis where at many time we ran into the issue of having missing values for a few census tracts or a census tract not having complete information on income.

References

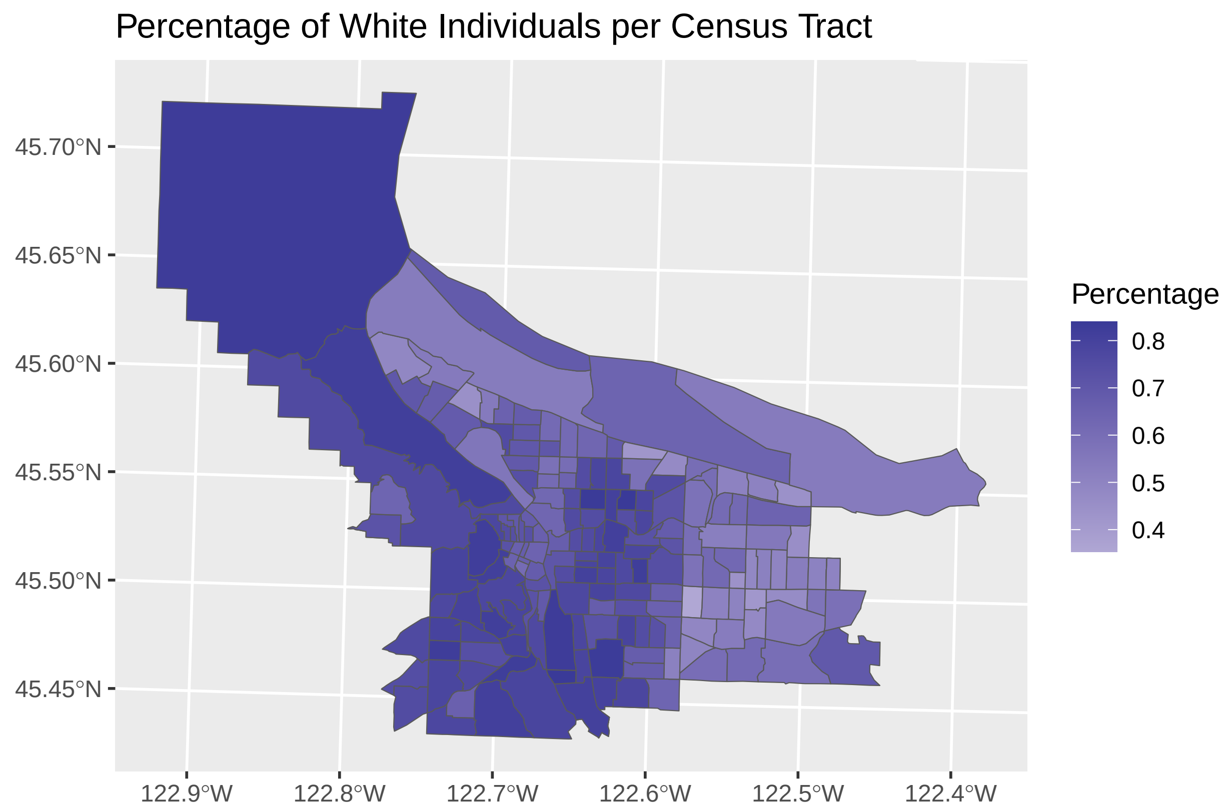
Census.gov Glossary Bureau, U. C. (2022, April 11). Glossary. Census.gov. https://www.census.gov/programs-surveys/geography/about/glossary.html#par_textimage_13

Neighborhood boundaries data: City of Portland. (2024). Portland Neighborhood Boundaries Open Data Shapefile. PortlandMaps Open Data. Retrieved from https://gis-pdx.opendata.arcgis.com/datasets/1e95d9b9076742ed9a71de0535ac255c/explore

Portland crime data: City of Portland. (2020). Portland Police Bureau 2020 Open Data. Tableau Public. Retrieved from https://public.tableau.com/app/profile/portlandpolicebureau/viz/New_Monthly_Neighborhood/MonthlyOffenseTotals

Tidycensus data: Walker, K., & Herman, M. (2023). tidycensus: Load US Census Boundary and Attribute Data as 'tidyverse' and 'sf'-Ready Data Frames (Version 1.5) [Computer software]. Retrieved from https://walker-data.com/tidycensus/

Demographics of Census Tracts




```
##
##  Monte-Carlo simulation of Moran I
##
## data:  pccc_sf$white_per
## weights: pccc_nbw2
## number of simulations + 1: 500
##
## statistic = 0.7077, observed rank = 500, p-value = 0.002
## alternative hypothesis: greater
```

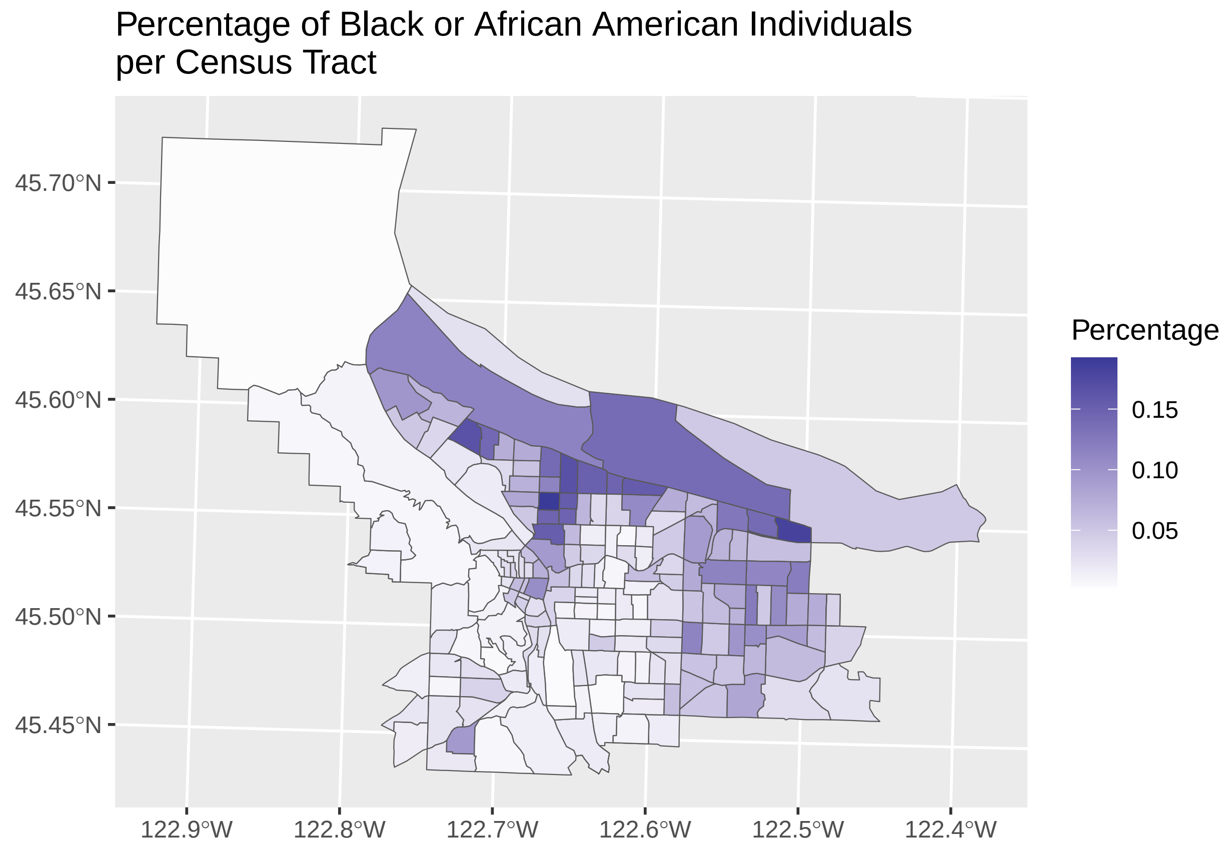
Our Test

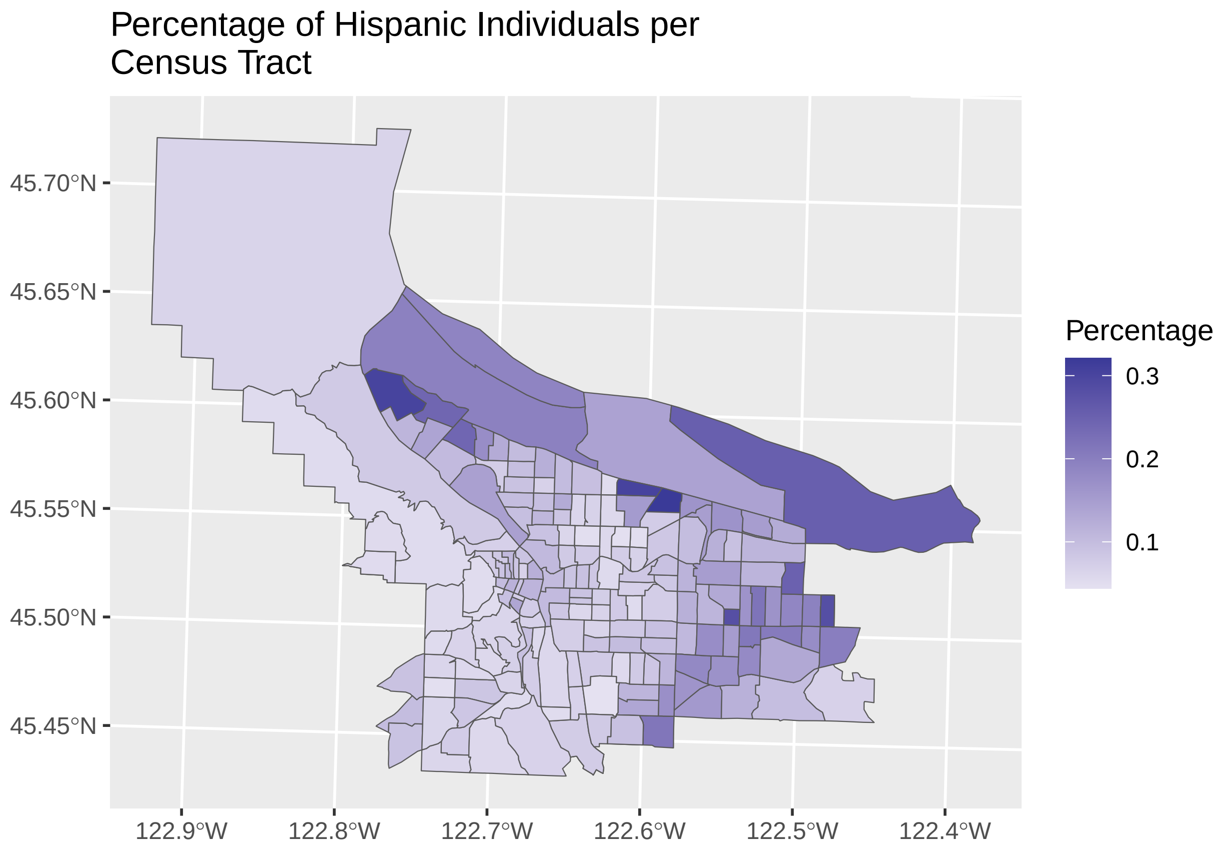
H_O: No spatial autocorrelation, I is close to 0

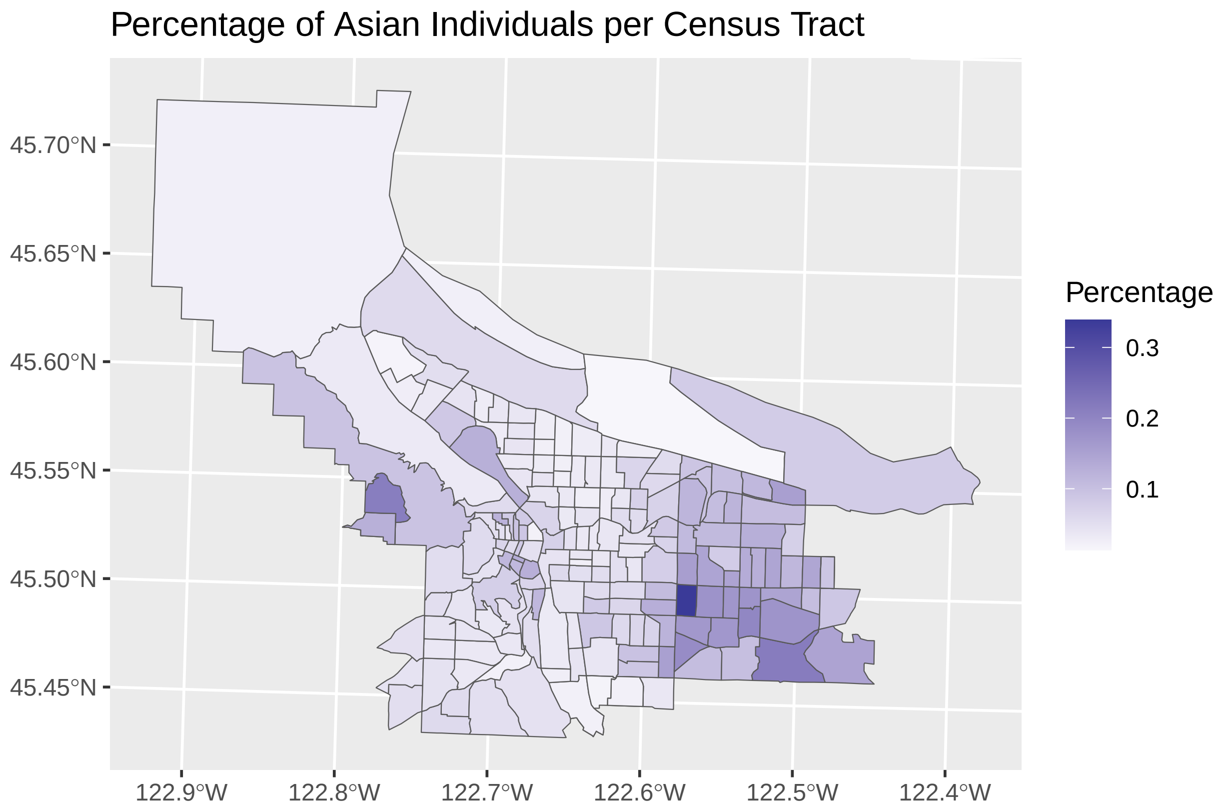
H_A: Spatial autocorrelation, I $\neq$ 0.

Calculating Moran's I for the percentage of white individuals in using Monte Carlo's simulations we get a Moran's I of 0.7077 which is positive and very strong. We also got a p-value of 0.002. We can reject the null hypothesis and conclude that it is statistically significant that the percentage of white individuals in Portland, Oregon is spatially autocorrelated.

More Graphs


Percentage of Black or African American Individuals
per Census Tract
45.70°N
45.65°N
45.60°N
45.55°N
45.50°N
45.45°N
122.9°W
122.8°W
122.7°W
122.6°W
122.5°W
122.4°W
Percentage
0.15
0.10
0.05


Percentage of Hispanic Individuals per
Census Tract
45.70°N
45.65°N
45.60°N
45.55°N
45.50°N
45.45°N
122.9°W
122.8°W
122.7°W
122.6°W
122.5°W
122.4°W
Percentage
0.3
0.2
0.1


Percentage of Asian Individuals per Census Tract
45.70°N
45.65°N
45.60°N
45.55°N
45.50°N
45.45°N
122.9°W
122.8°W
122.7°W
122.6°W
122.5°W
122.4°W
Percentage
0.3
0.2
0.1