The Relationship between Urban Green Space and Socioeconomic Factors in the City of Vancouver: A Case of Environmental Injustice *

Zahra Akhtar *400067162*
Anastasiia Kalinina *400050085*
Gabriella Morrone *400085415*
Jillian Walters *400285085*
Jocelyn Wu *400051160*

This paper reports our analysis of opioid prescriptions in Kentucky, USA, and its relationship with income and education at the level of counties. Data were obtained from the Washington Post database and the US Census.

Keywords: opioids, education, income, spatial analysis

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

#Introduction

#Background

#Study Area This study took place at the Census Subdivision level to target the municipality of Vancouver, BC. There are 127 census tracts in Vancouver that were studied.

```
## Reading vectors data from local cache.
```

```
## Reading geo data from local cache.
```

#Data The data used for analysis includes census data for variables of population, population density, visible minority status, non-visible minority status, and income for each census tract in the census subdivision of Vancouver. This data was obtained from Statistics Canada through use of the cancensus package in R. All census data used is from the 2021 census. Point data for each the location of each park, homeless shelter, and street tree in the census subdivision was downloaded in a .csv file format from the City of Vancouver's open data portal. To use the r5r package so that accessibility to parks by both walking and public transit could be determined, a road network dataset of Vancouver, and as well as a public transport feed of the city was needed. The road network dataset was obtained from BBBike, and was stored as a .pbf file. The public transport feed was obtained from Transitland and stored in a GTFS.zip file.

#Methods To determine the relationship between greenspace and other socioeconomic factors in Vancouver, Rstudio was used to analyze and visualize the variables being examined. Choropleth maps for variables of population, population density, visible minority status, non-visible minority status, and income, and accessibility to parks were made to provide a visual representation of the difference between census tracts. Regression analysis was used to determine the relationship between the independent variable studied (number of parks accessible within a 30 minute travel time) and dependent variables (population, population density, visible minority status, non-visible minority status, and income). Scatterplots were used to show a visual representation between independent and dependent variables studied to better understand the results of regression analysis.

*Paper submitted to complete the requirements of ENVSOCTY 4GA3 Applied Spatial Statistics; with additional edits by Antonio Paez for this version.

#Results

To begin our analysis, the centroids for each census tract in the Vancouver CSD were obtained and put into a dataframe with latitude, longitude, and census ID. This allowed us to have point data describing the middle of each CT.

```
## [1] "data.frame"
```

The point data for park in the Vancouver CSD was obtained from a csv file from Vancouver's open data portal and transformed into a data frame with columns describing the X and Y coordinates of each park.

The point data for parks was prepared for use of the r5r package by renaming coulumns for latitude, longitude, and park ID data and to the names required by r5r. Data for latitude and longitude was converted into a numeric format. Lastly, a column to indicate that each park point was one singular park was added to meet the requirement of the opportunites parameter in r5r.

```
## [1] "data.frame"
```

Next, the r5r package was used in order to build a transport network for Vancouver, so that the distances and routing between the centroids of census tracts and park locations can be calculated.

The transport network was built with using a road network dataset of Vancouver, and as well as a public transport feed of the city. The road network dataset was obtained from BBBike, and was stored as a .pbf file. The public transport feed was obtained from Transitland and stored in a GTFS.zip file.

```
## [1] "GTFS_TRANIST.zip"          "network_settings.json"
## [3] "network.dat"               "Vancouver.osm.pbf"
## [5] "Vancouver.osm.pbf.mapdb"   "Vancouver.osm.pbf.mapdb.p"

## No raster .tif files found. Using elevation = 'NONE'.

## Using cached R5 version from /Library/Frameworks/R.framework/Versions/4.2/Resources/library/r5r/jar/

##
## Using cached network.dat from Project data/r5rpath/network.dat

## [1] "data.frame"
```

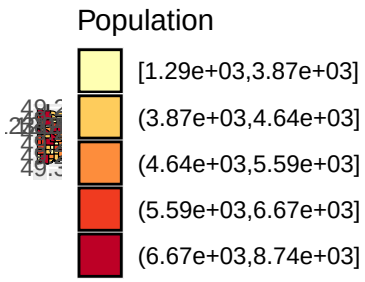
The accessibility function in r5r was used to compute how many parks were accessible within 30 minutes of each census tracts centroid by walking or public transit.

```
## Warning in assign_points_input(destinations, "destinations"): 'destinations$id'
## forcefully cast to character.
```

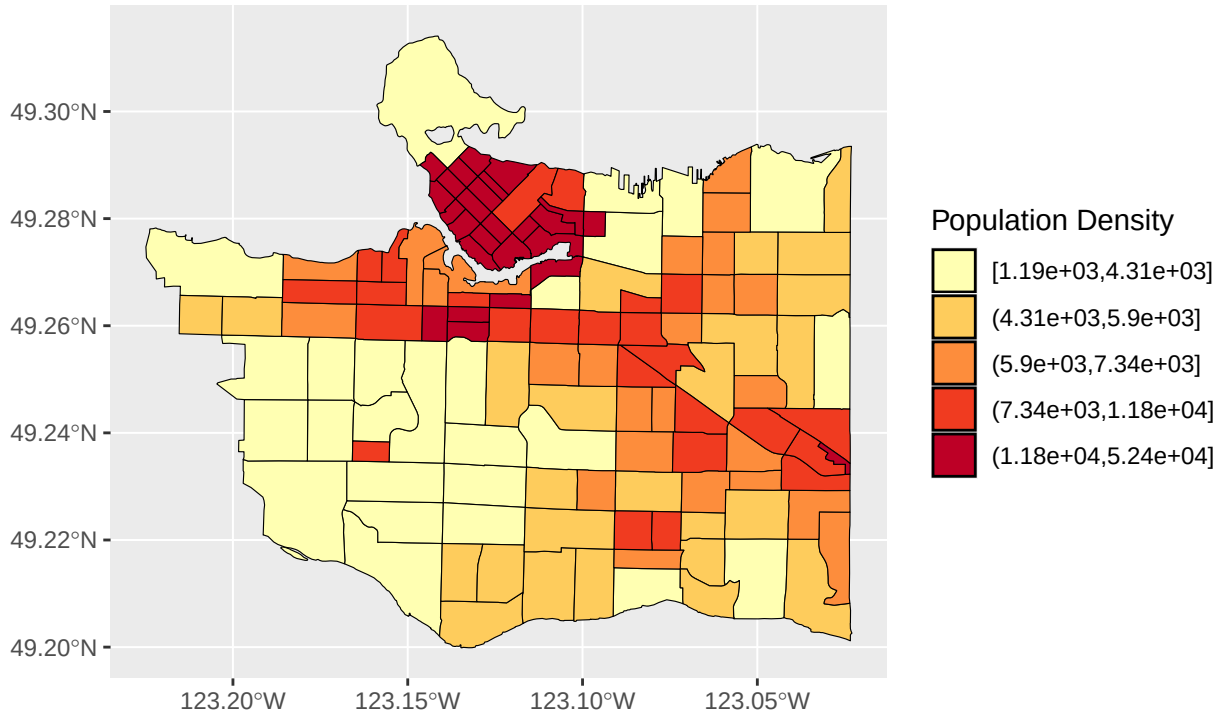
Once the number of accessible parks was calculated, this was added to the data frame displaying census data.

After all variables were in one data frame and prepared for analysis, data was first visualized by creating choropleth maps for each variable of interest. Maps were created for each variable as follows:

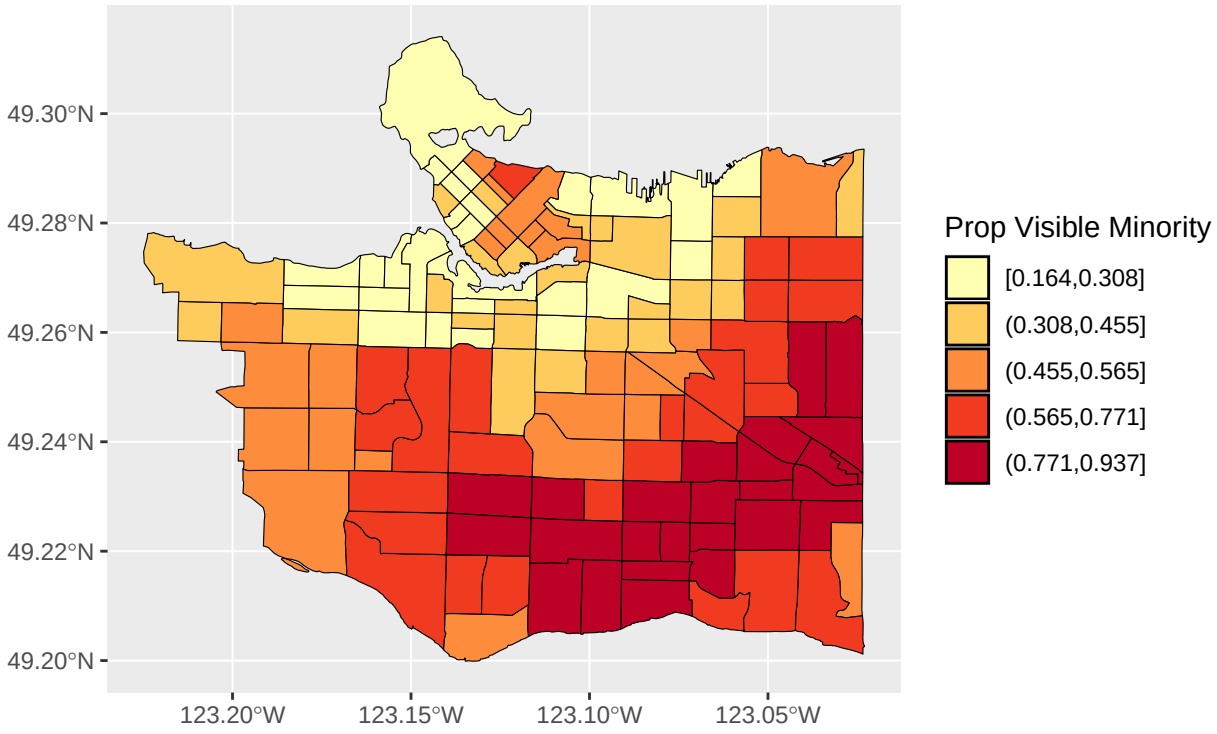
Population of each CT

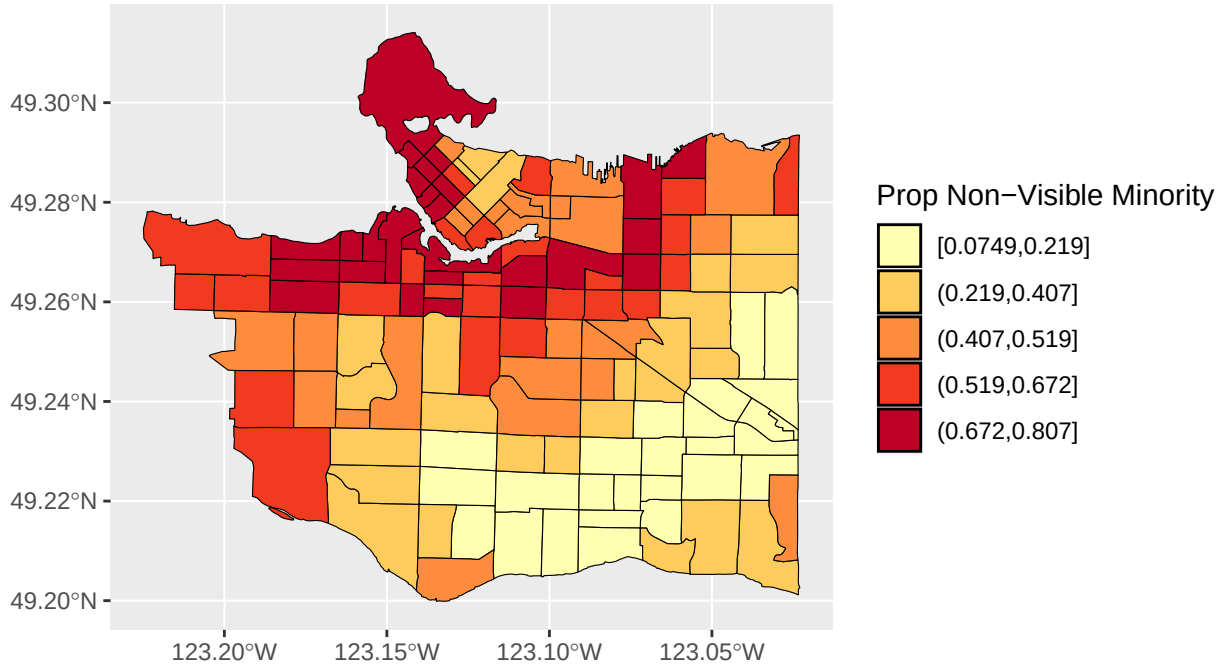


Population density of each CT

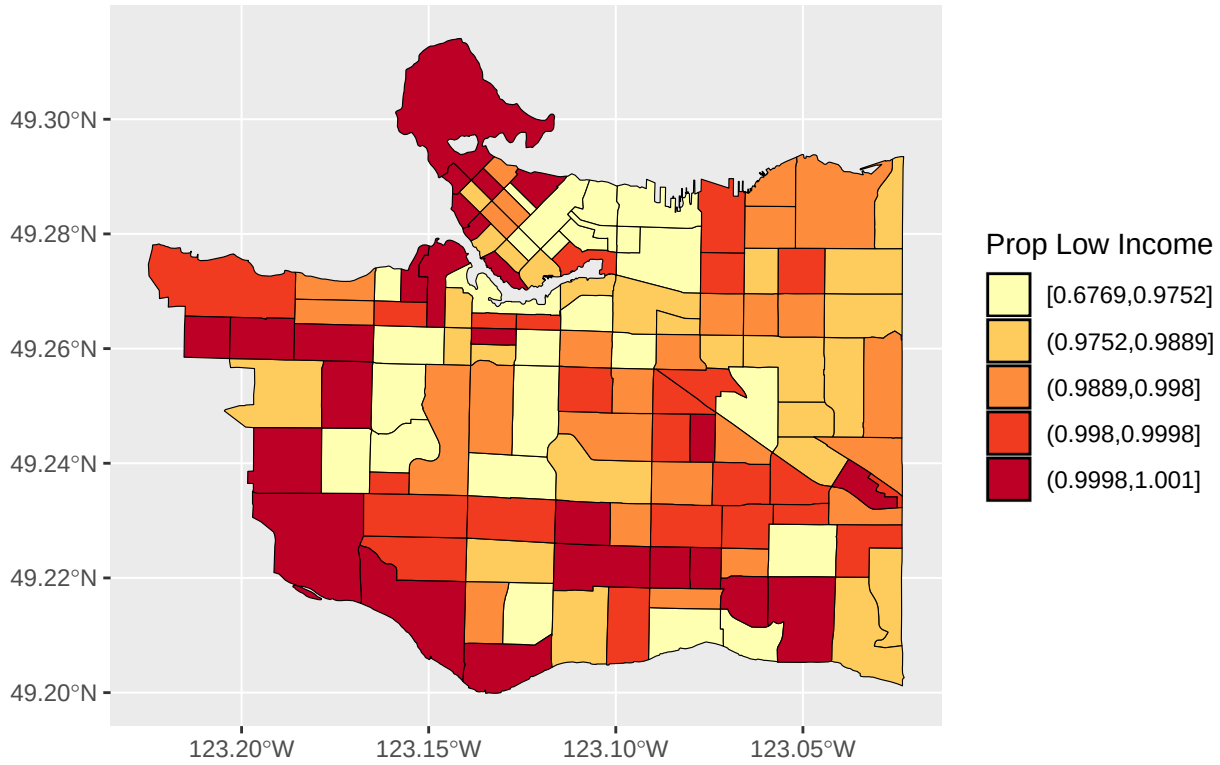


Proportion of population identifying as visible and non-visible minorities

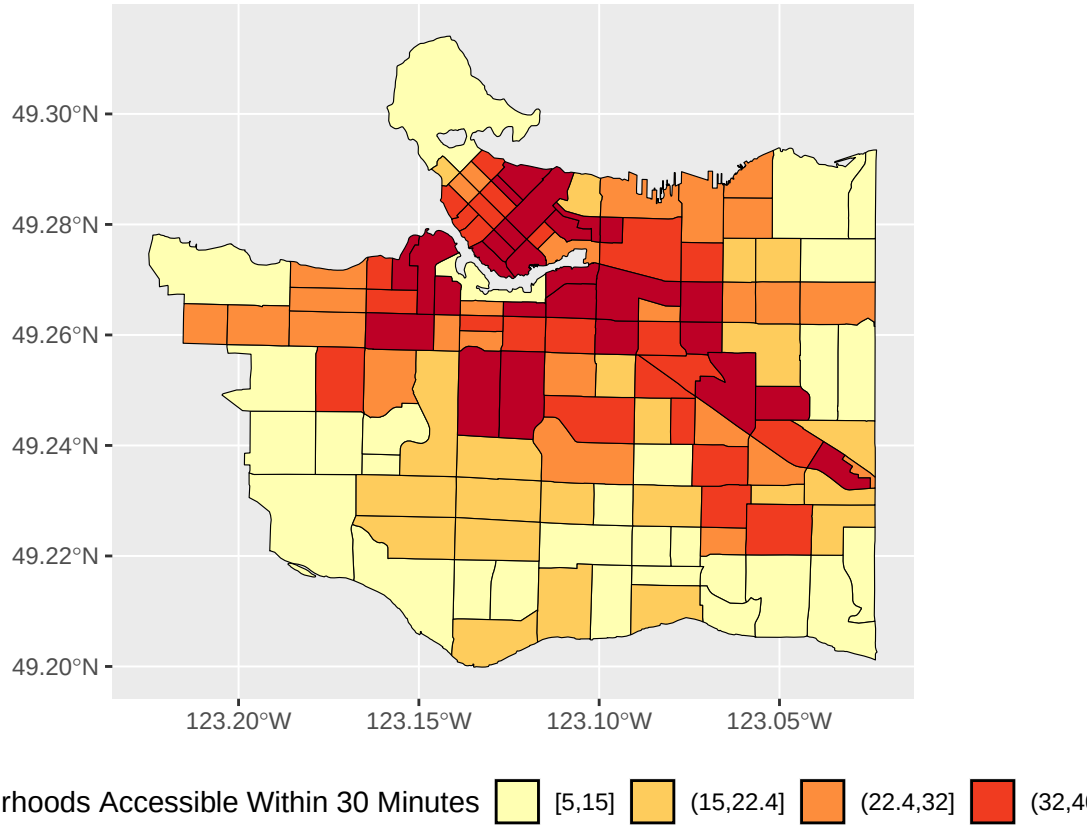

49.30°N
49.28°N
49.26°N
49.24°N
49.22°N
49.20°N
123.20°W
123.15°W
123.10°W
123.05°W
Prop Visible Minority
[0.164,0.308]
(0.308,0.455]
(0.455,0.565]
(0.565,0.771]
(0.771,0.937]



The proportion of the population considered low income for each CT



Finally, a choropleth map was created to display the independent variable, parks accessible within 30 mintues of a centroid.



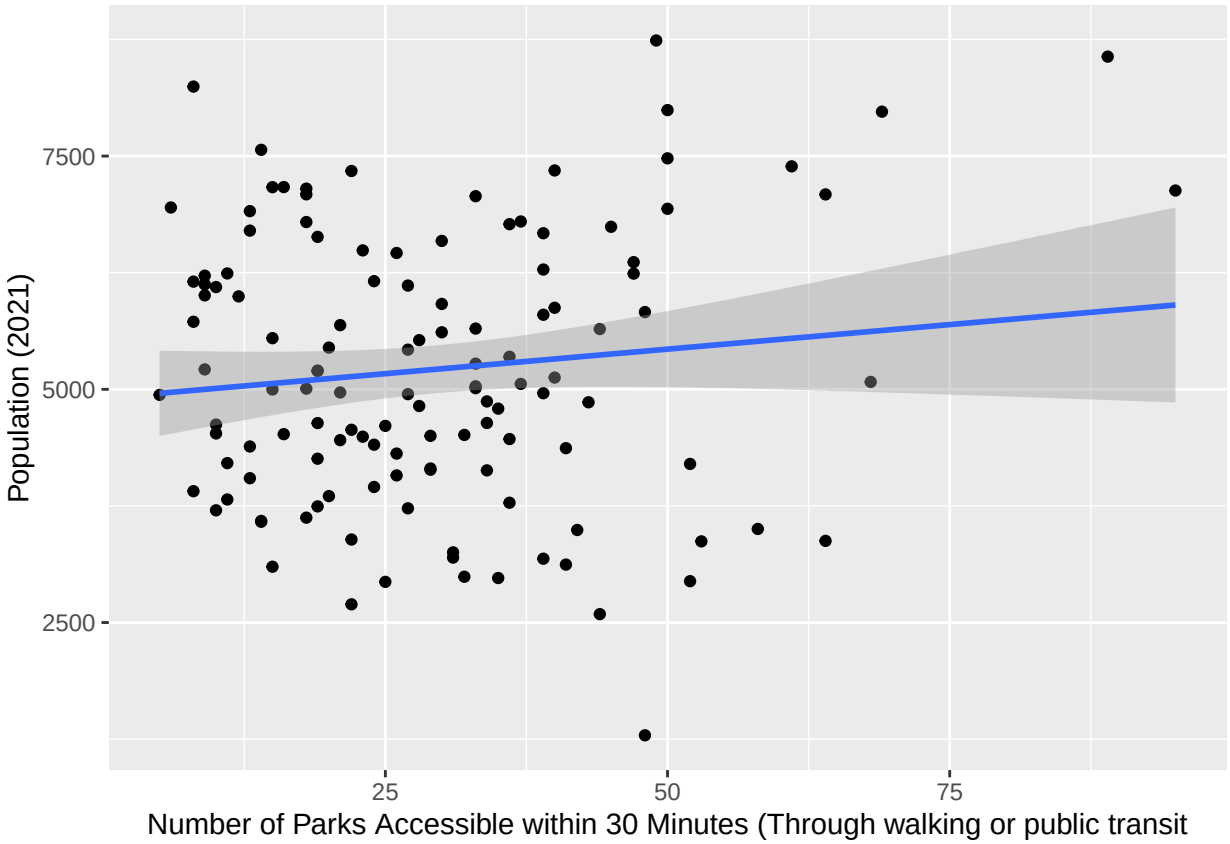
Next, to determine the relationship between the number of accessible parks and each independent variable, the independent variables were regressed to number of accessible parks for each CT.

Table 1: Population of Census Tracts regressed on Number of Parks Accessible

	Dependent variable:
	Population_2021
accessibility	10.518
	(7.816)
Constant	4,904.639***
	(264.208)
Observations	127
R^2	0.014
Adjusted R^2	0.006
Residual Std. Error	1,459.451 (df = 125)
F Statistic	1.811 (df = 1; 125)
Note:	*p<0.1; **p<0.05; ***p<0.01

Call: lm(formula = Population_2021 ~ accessibility, data = census_data)
Residuals: Min 1Q Median 3Q Max -4118.5 -1065.2 -159.5 1192.0 3319.0
Coefficients: Estimate Std. Error t value Pr(>|t|)
(Intercept) 4904.639 264.208 18.564 <2e-16 *** accessibility 10.518 7.816 1.346 0.181
— Signif. codes: 0 ‘’ ***0.001*** ’’ *0.01* ’’ 0.05 ‘.’ 0.1 ’ ’ 1

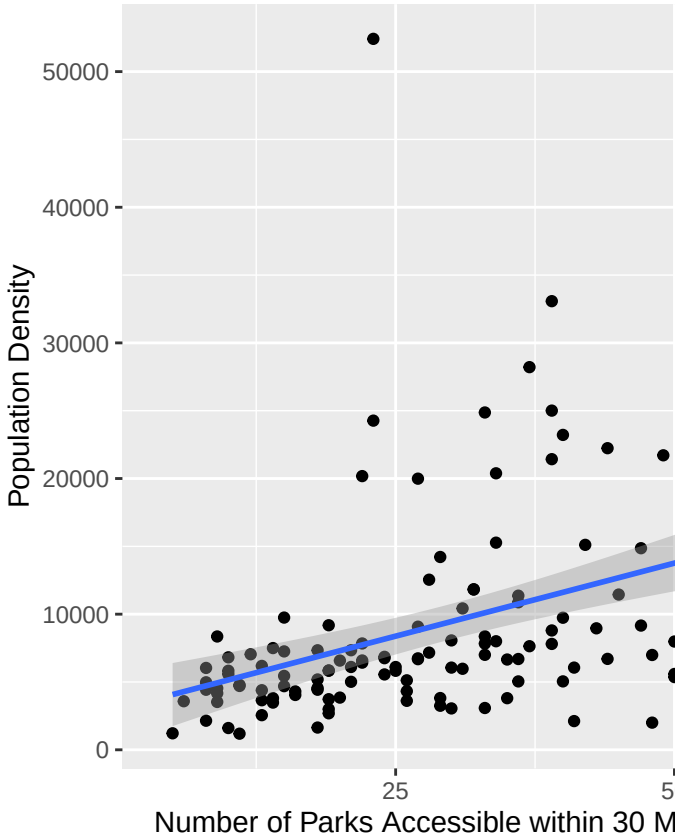
Residual standard error: 1459 on 125 degrees of freedom Multiple R-squared: 0.01428, Adjusted R-squared: 0.006395 F-statistic: 1.811 on 1 and 125 DF, p-value: 0.1808 For each regression, a scatter plot was created to provide a visual representation of the data



Model regressing population density on number of accessible marks

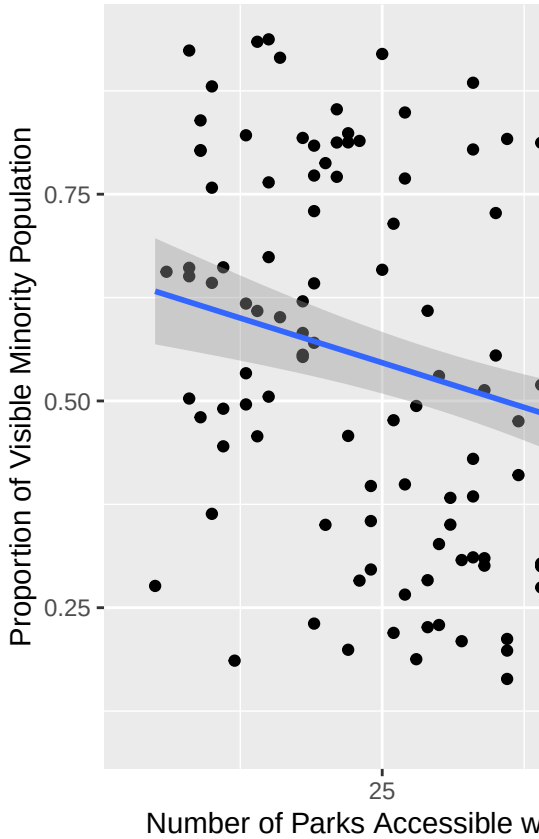
```
##
## \begin{table}[!htbp] \centering
##   \caption{Population Density of Census Tracts regressed on Number of Parks Accessible}
##   \label{}
## \begin{tabular}{@{\extracolsep{5pt}}lc}
## \\[-1.8ex]\hline
## \hline \\[-1.8ex]
##  & \multicolumn{1}{c}{\textit{Dependent variable:}} \\
## \cline{2-2}
## \\[-1.8ex] & Population\_density \\
## \hline \\[-1.8ex]
##  accessibility & 215.382$^{***}$ \\
##   & (39.775) \\
##   & \\
##  Constant & 2,995.220$^{**}$ \\
##   & (1,344.555) \\
##   & \\
## \hline \\[-1.8ex]
## Observations & 127 \\
## R$^{2}$ & 0.190 \\
## Adjusted R$^{2}$ & 0.184 \\
## Residual Std. Error & 7,427.142 (df = 125) \\
```

```
## F Statistic & 29.322$^{***}$ (df = 1; 125) \\
## \hline
## \hline \\[-1.8ex]
## \textit{Note:}  & \multicolumn{1}{r}{$^{*}$p$<$0.1; $^{**}$p$<$0.05; $^{***}$p$<$0.01} \\
## \end{tabular}
## \end{table}

##
## Call:
## lm(formula = Population_density ~ accessibility, data = census_data)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -12200  -3462  -1687    715  44466
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)    2995.22    1344.55   2.228   0.0277 *
## accessibility   215.38      39.78   5.415 3.02e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7427 on 125 degrees of freedom
## Multiple R-squared:   0.19,  Adjusted R-squared:  0.1835
## F-statistic: 29.32 on 1 and 125 DF,  p-value: 3.018e-07
```



Scatterplot of population density vs number of accessible parks
Model regressing proportion of visible minorities on number of accessible parks

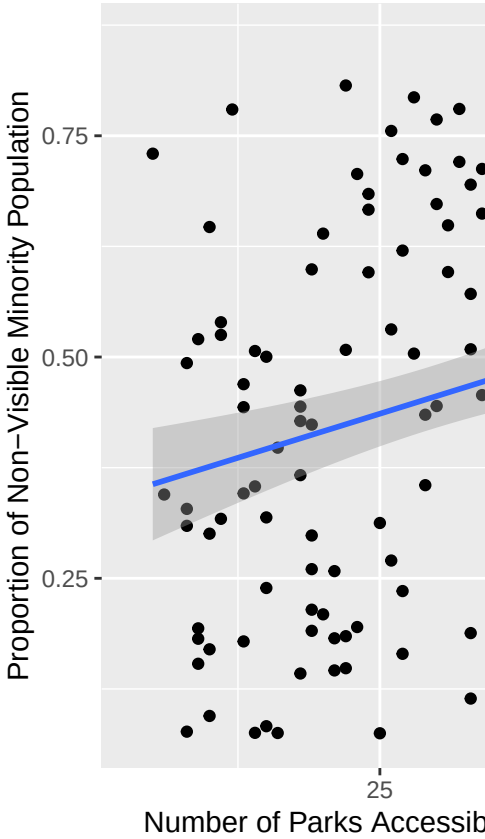
```
##
## \begin{table}[!htbp] \centering
##   \caption{Proportion of Visible Minority Population in Census Tracts regressed on Number of Parks A
##   \label{}
## \begin{tabular}{@{\extracolsep{5pt}}lc}
## \\[-1.8ex]\hline
## \hline \\[-1.8ex]
##  & \multicolumn{1}{c}{\textit{Dependent variable:}} \\
## \cline{2-2}
## \\[-1.8ex] & Proportion\_visible\_minority \\
## \hline \\[-1.8ex]
##  accessibility & $-$0.004$^{***}$ \\
##   & (0.001) \\
##   & \\
##  Constant & 0.654$^{***}$ \\
##   & (0.037) \\
##   & \\
## \hline \\[-1.8ex]
## Observations & 127 \\
## R$^{2}$ & 0.110 \\
## Adjusted R$^{2}$ & 0.103 \\
## Residual Std. Error & 0.205 (df = 125) \\
## F Statistic & 15.511$^{***}$ (df = 1; 125) \\
## \hline
## \hline \\[-1.8ex]
## \textit{Note:}  & \multicolumn{1}{r}{$^{*}$p$<$0.1; $^{**}$p$<$0.05; $^{***}$p$<$0.01} \\
## \end{tabular}
## \end{table}

##
## Call:
## lm(formula = Proportion_visible_minority ~ accessibility, data = census_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.41644 -0.15748  0.00557  0.18028  0.37338
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.654239   0.037080  17.644  < 2e-16 ***
## accessibility -0.004320   0.001097  -3.938 0.000135 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2048 on 125 degrees of freedom
## Multiple R-squared:  0.1104, Adjusted R-squared:  0.1033
## F-statistic: 15.51 on 1 and 125 DF,  p-value: 0.0001355
```



Scatterplot of proportion of visible minorities vs number of accessible parks
Model regressing proportion of non-visible minorities on number of accessible parks

```
##
## \begin{table}[!htbp] \centering
##   \caption{Proportion of Non-Visible Minority Population in Census Tracts regressed on Number of Parl
##   \label{}
## \begin{tabular}{@{\extracolsep{5pt}}lc}
## \\[-1.8ex]\hline
## \hline \\[-1.8ex]
##  & \multicolumn{1}{c}{\textit{Dependent variable:}} \\
## \cline{2-2}
## \\[-1.8ex] & Proportion\_nonvisible\_minority \\
## \hline \\[-1.8ex]
##  accessibility & 0.004$^{***}$ \\
##   & (0.001) \\
##   & \\
##  Constant & 0.336$^{***}$ \\
##   & (0.037) \\
##   & \\
## \hline \\[-1.8ex]
## Observations & 127 \\
## R$^{2}$ & 0.097 \\
## Adjusted R$^{2}$ & 0.090 \\
## Residual Std. Error & 0.203 (df = 125) \\
## F Statistic & 13.473$^{***}$ (df = 1; 125) \\
## \hline
## \hline \\[-1.8ex]
```

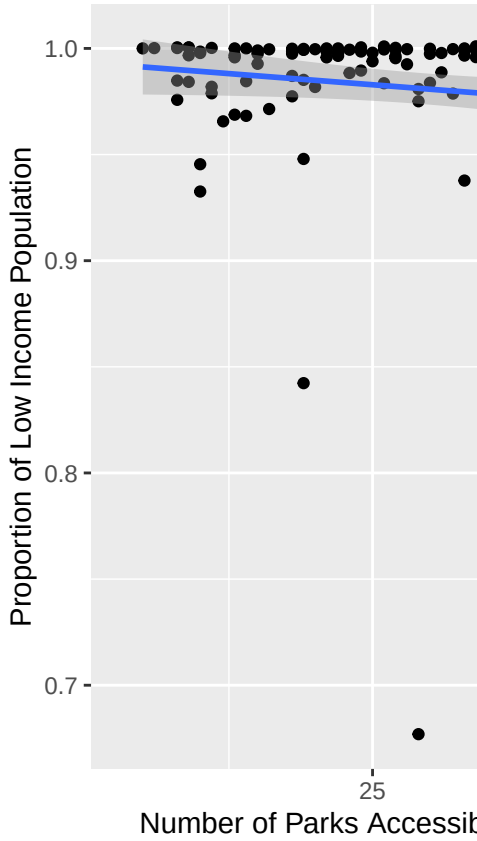
```
## \textit{Note:}  & \multicolumn{1}{r}{$^{*}$p$<$0.1; $^{**}$p$<$0.05; $^{***}$p$<$0.01} \\
## \end{tabular}
## \end{table}

##
## Call:
## lm(formula = Proportion_nonvisible_minority ~ accessibility,
##     data = census_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.36102 -0.18133 -0.00251  0.16906  0.39551
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.336380   0.036681   9.170  1.2e-15 ***
## accessibility 0.003983   0.001085   3.671 0.000357 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2026 on 125 degrees of freedom
## Multiple R-squared:  0.0973, Adjusted R-squared:  0.09007
## F-statistic: 13.47 on 1 and 125 DF,  p-value: 0.0003571
```



Scatterplot of proportion of non-visible minorities vs number of accessible parks
Model regressing proportion of low income population on number of accessible parks

```
##
## \begin{table}[!htbp] \centering
```

```
##   \caption{Proportion of Low Income Population in Census Tracts regressed on Number of Parks Accessil
##   \label{}
## \begin{tabular}{@{\extracolsep{5pt}}lc}
## \\[-1.8ex]\hline
## \hline \\[-1.8ex]
##  & \multicolumn{1}{c}{\textit{Dependent variable:}} \\
## \cline{2-2}
## \\[-1.8ex] & Proportion\_low\_income \\
## \hline \\[-1.8ex]
##  accessibility & $-$0.0004$^{*}$ \\
##   & (0.0002) \\
##   & \\
##  Constant & 0.993$^{***}$ \\
##   & (0.008) \\
##   & \\
## \hline \\[-1.8ex]
## Observations & 127 \\
## R$^{2}$ & 0.028 \\
## Adjusted R$^{2}$ & 0.020 \\
## Residual Std. Error & 0.041 (df = 125) \\
## F Statistic & 3.571$^{*}$ (df = 1; 125) \\
## \hline
## \hline \\[-1.8ex]
## \textit{Note:}  & \multicolumn{1}{r}{$^{*}$p$<$0.1; $^{**}$p$<$0.05; $^{***}$p$<$0.01} \\
## \end{tabular}
## \end{table}

##
## Call:
## lm(formula = Proportion_low_income ~ accessibility, data = census_data)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.304258 -0.001073  0.011271  0.017072  0.032010
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.9933794  0.0075109  132.26   <2e-16 ***
## accessibility -0.0004199  0.0002222   -1.89   0.0611 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.04149 on 125 degrees of freedom
## Multiple R-squared:  0.02777,    Adjusted R-squared:  0.01999
## F-statistic: 3.571 on 1 and 125 DF,  p-value: 0.06112
```



Scatterplot of proportion of low income population vs number of accessible parks
#Analysis
#Conclusion