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

Daina De Angelis *400240413*
Tim Truong *400264793*
Maeve Nowitsky *400264259*
Jillian Mezenberg *400243427*

R Markdown

Introduction

Vancouver is usually ranked as one the most sustainable as the city tries to move toward a zero-carbon emission city, the role of urban greeny became more important than ever. Urban green space plays an important role in creating liveable spaces and helping to maintain the well-being of the majority of residents. In addition, Urban green spaces can reduce many climate change impacts, support urban ecology, improve property values, and increase social interactions (Nesbitt et al., 2019). The relationship between urban green space and socioeconomic factors in the city of Vancouver can be complex and multifaceted. The distribution of urban green space is inequitable in some cities (Nesbitt & Meitner, 2016). Equitable access to urban green space is often associated with higher levels of well-being, particularly among disadvantaged and lower socioeconomic groups. In this report, Vancouver is chosen because it has high resolution urban green space data and socioeconomic data available. It aims to explore the relationship between accessibility to green space and socioeconomic factors across different census subdivisions in Vancouver, Canada.

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.

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

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

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.

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.

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 columns 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 opportunities parameter in r5r.

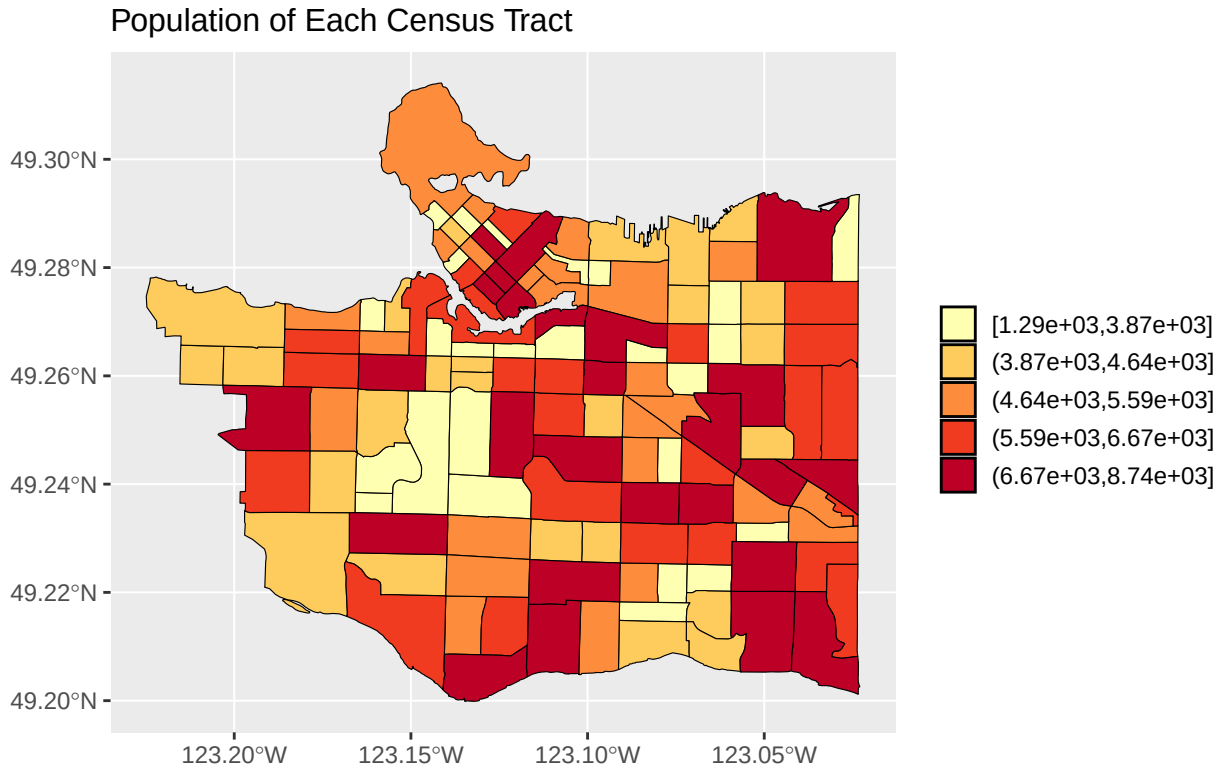
Next, the r5r package was used in order to build a transport network for Vancouver, so that the distances and routing between the centroids of census tracts and park locations can be calculated.

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.

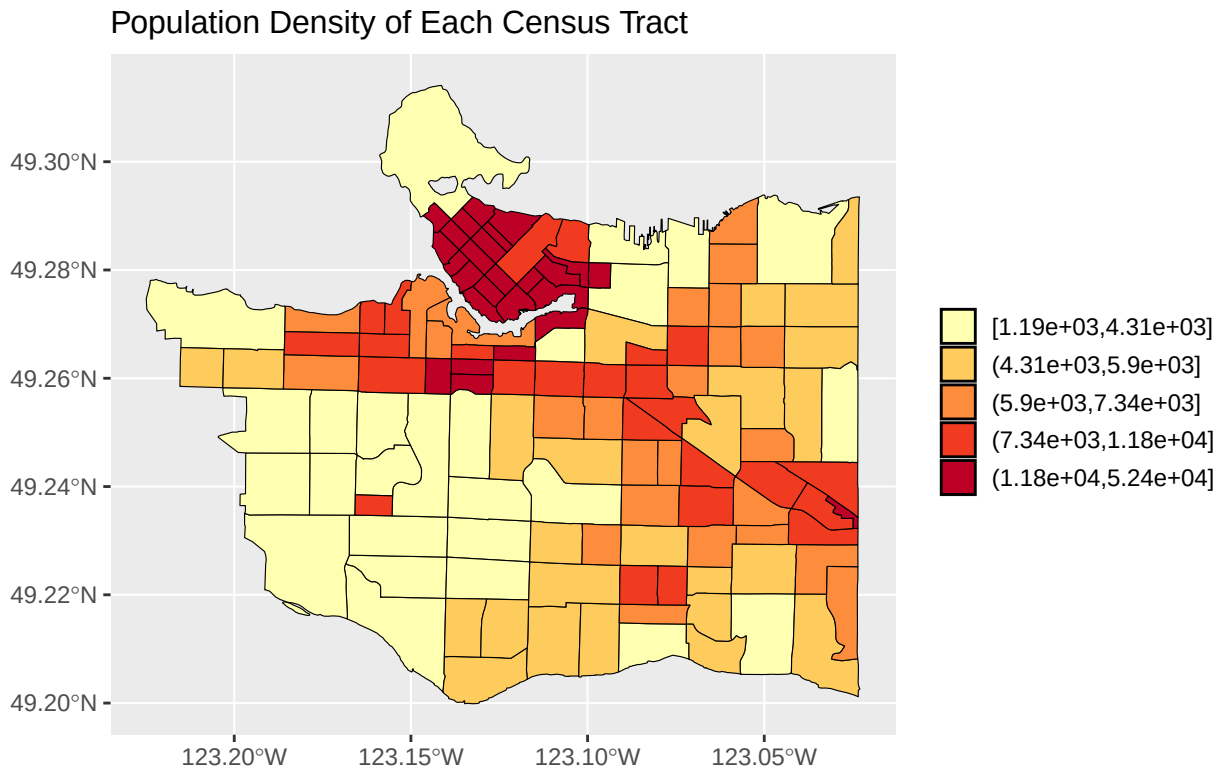
The accessibility function in r5r was used to compute how many parks were accessible within 30 minutes of each census tracts centroid by walking or public transit.

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

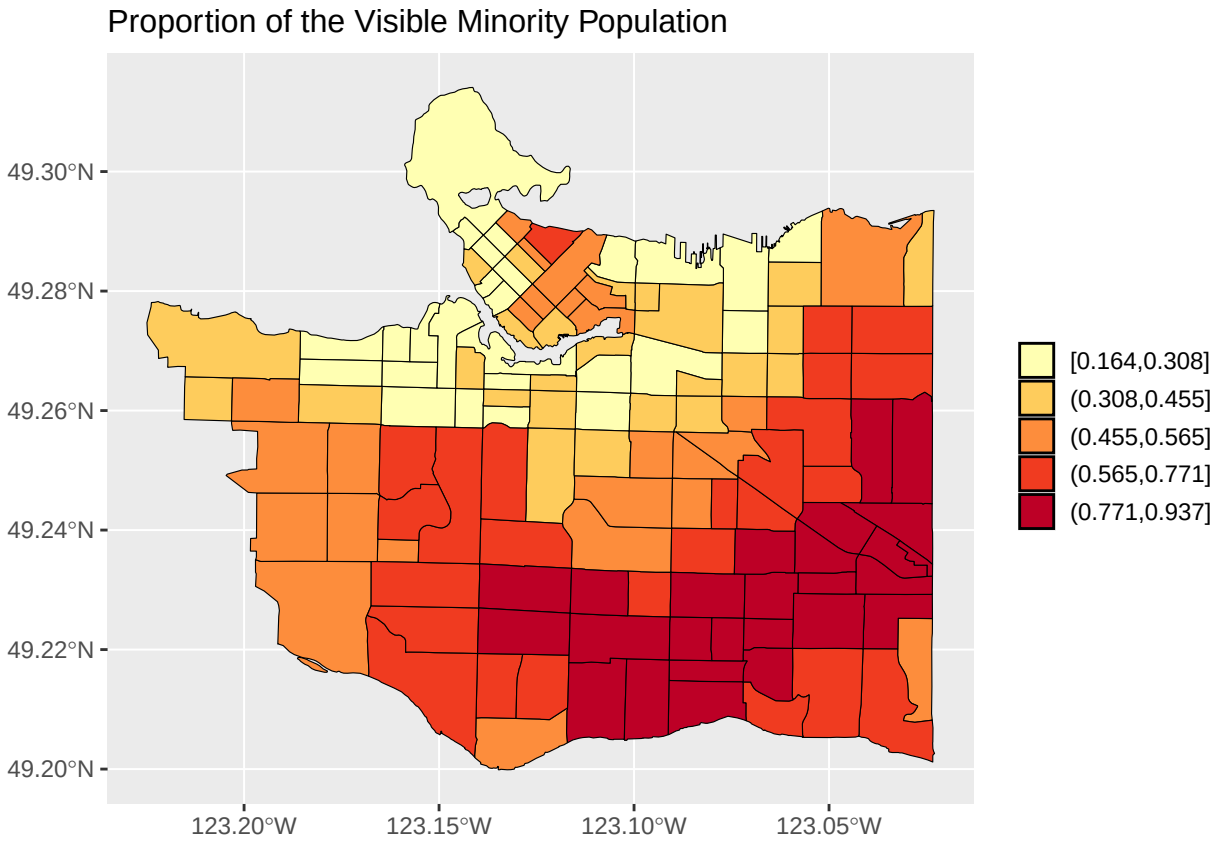
After all variables were in one data frame and prepared for analysis, data was first visualized by creating choropleth maps for each variable of interest. Maps were created for each variable as follows, beginning with creating a choropleth map for the 2021 population:



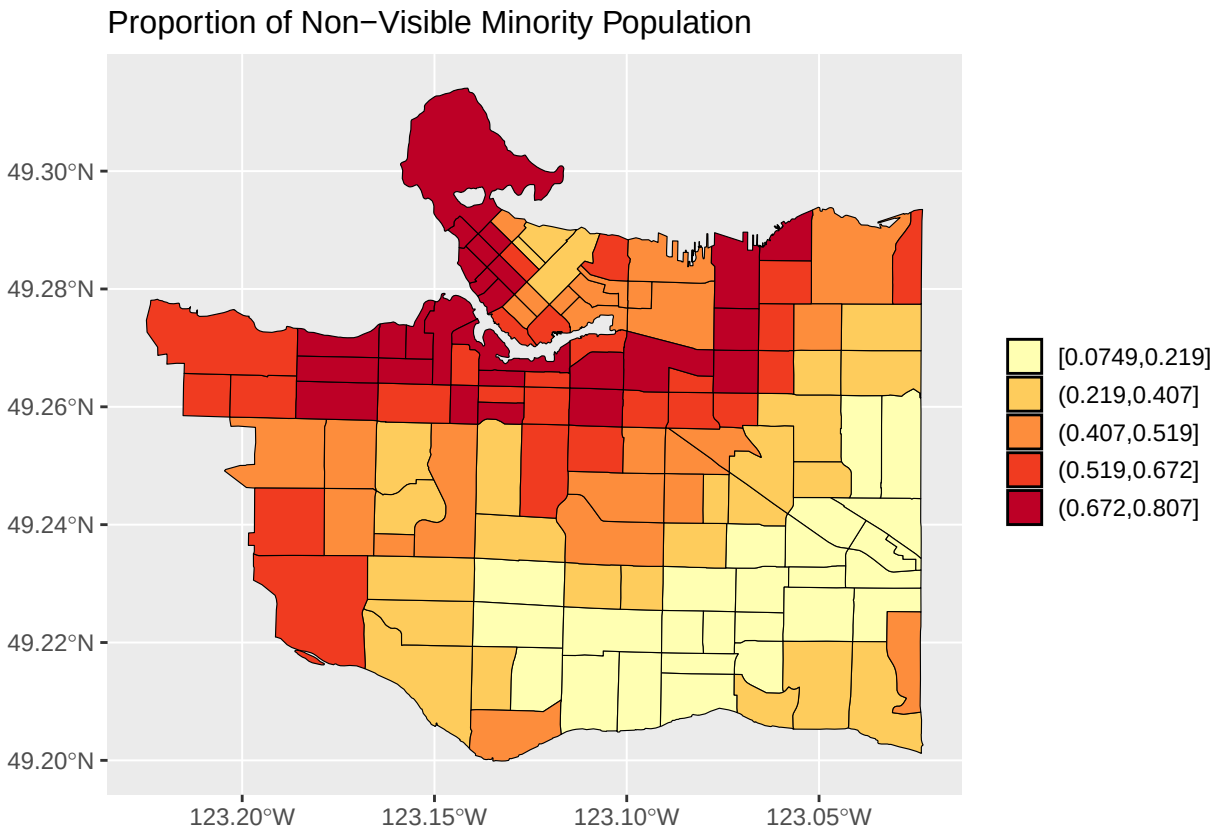
Next, creating a choropleth map for the population density of each census tract.



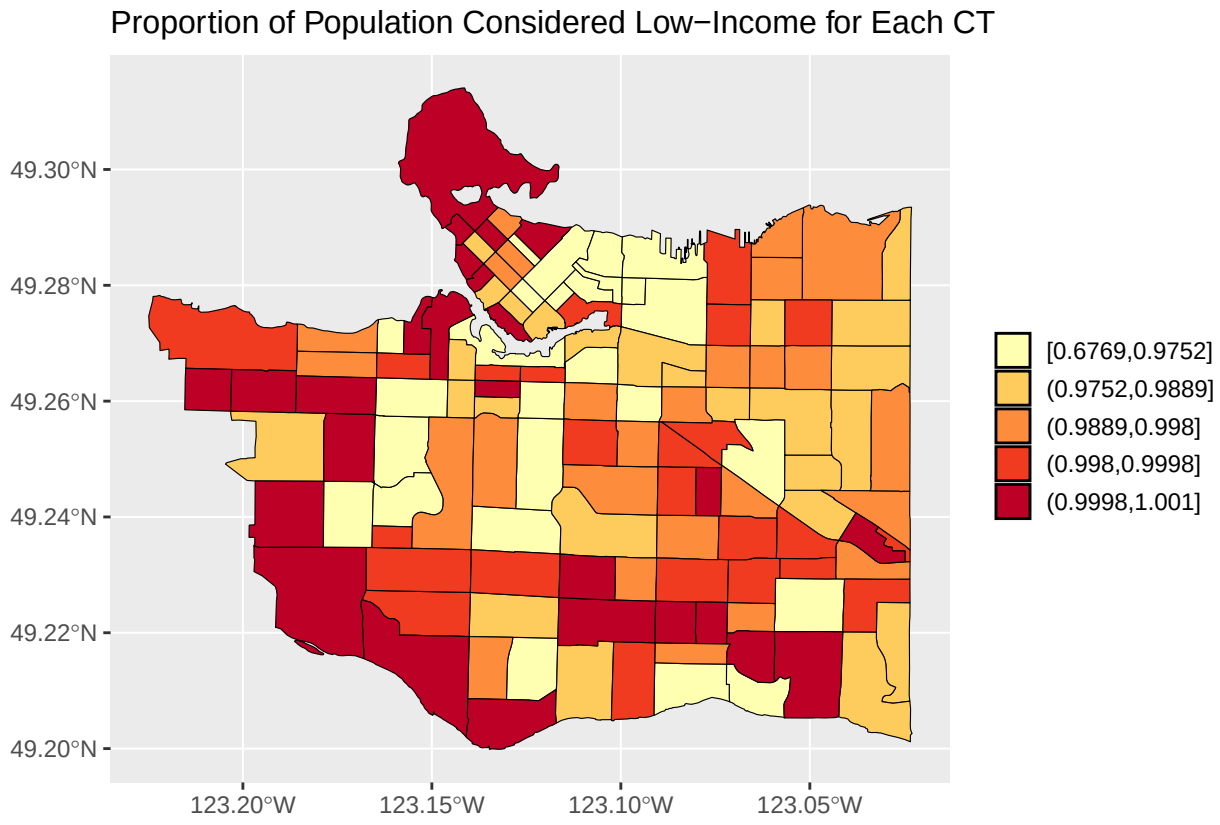
Now creating a choropleth map for the proportion of the visible minority population.



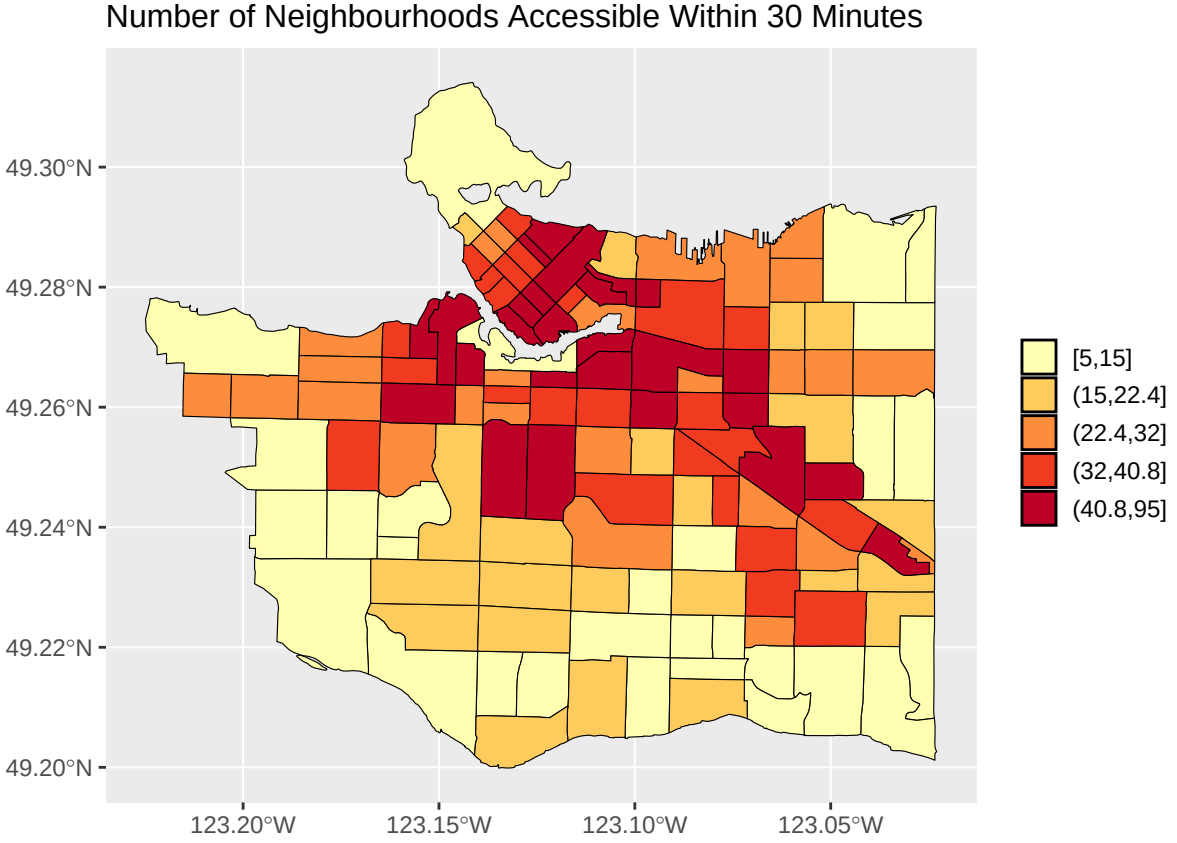
Creating a choropleth map for the proportion of the non-visible minority population.



Creating a choropleth map displaying the proportion of low-income areas.



Finally, a choropleth map was created to display the independent variable, parks accessible within 30 minutes 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

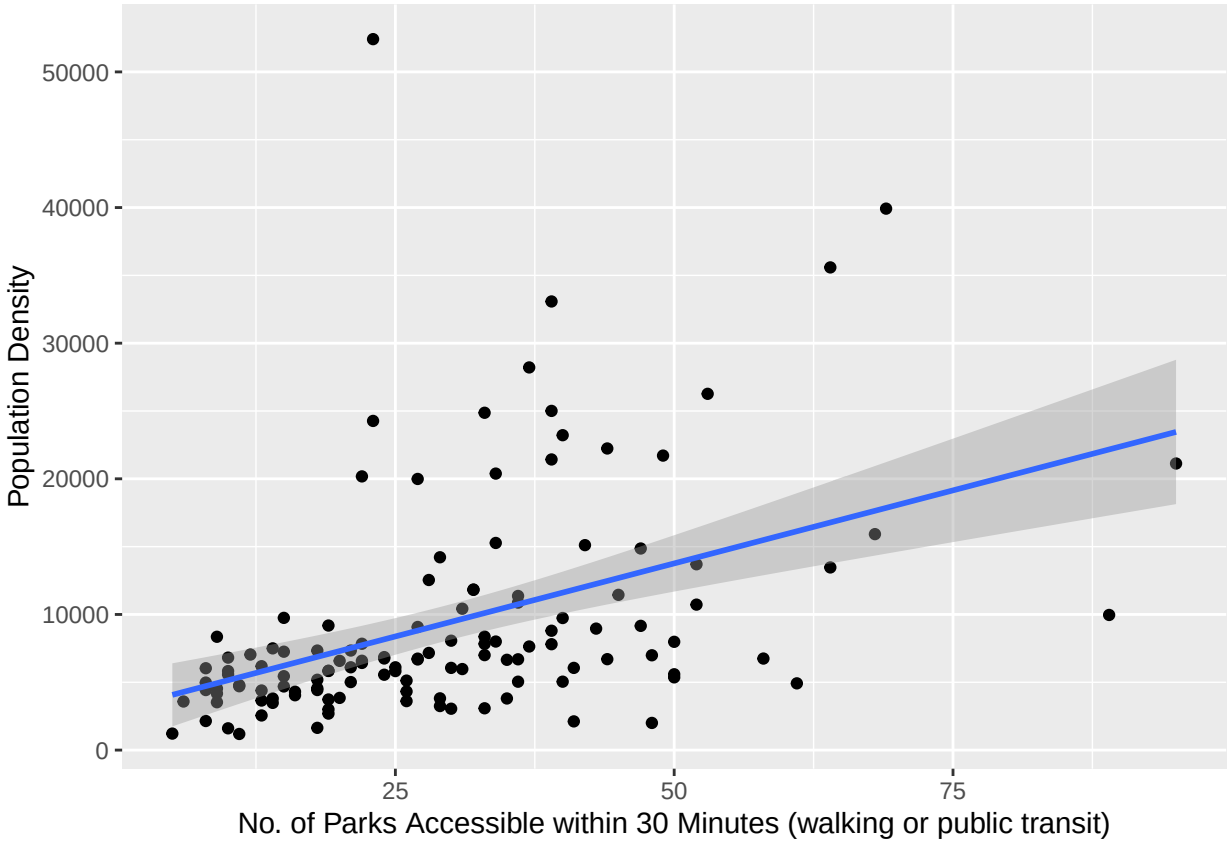
For each regression, a scatter plot was created to provide a visual representation of the data.

Model regressing population density on number of accessible parks.

Table 2: Population Density of Census Tracts regressed on Number of Parks Accessible

	Dependent variable:
	Population_density
accessibility	215.382***
	(39.775)
Constant	2,995.220**
	(1,344.555)
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)
Note:	*p<0.1; **p<0.05; ***p<0.01

Scatterplot of population density vs number of accessible parks.



Model regressing proportion of visible minorities on number of accessible parks.

Table 3: Proportion of Visible Minority Population in Census Tracts regressed on Number of Parks Accessible

	Dependent variable:
	Proportion_visible_minority
accessibility	-0.004^{***}
	(0.001)
Constant	0.654^{***}
	(0.037)
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)
Note:	*p<0.1; **p<0.05; ***p<0.01

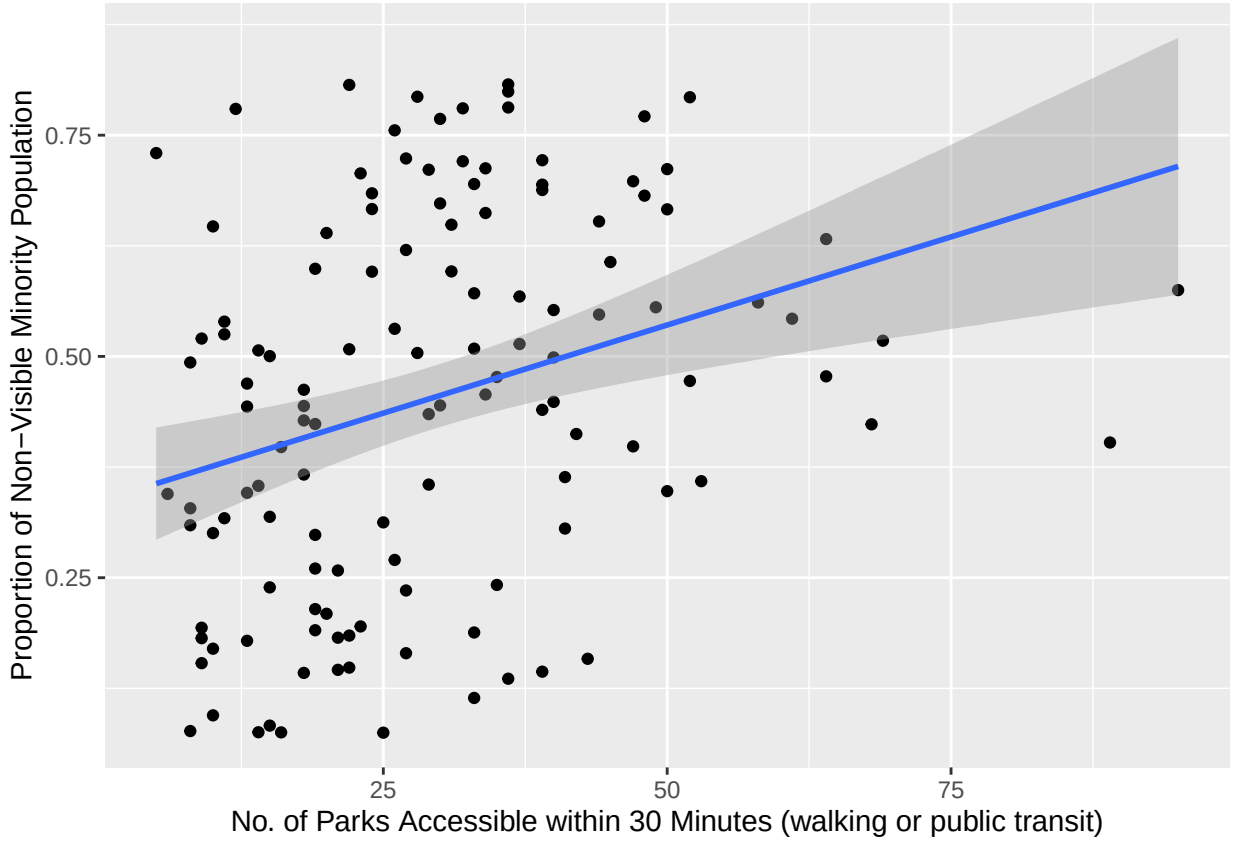
Scatterplot of proportion of visible minorities vs number of accessible parks.

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

Table 4: Proportion of Non-Visible Minority Population in Census Tracts regressed on number of Parks Accessible

	Dependent variable:
	Proportion_nonvisible_minority
accessibility	0.004***
	(0.001)
Constant	0.336***
	(0.037)
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)
Note:	*p<0.1; **p<0.05; ***p<0.01

Scatterplot of proportion of non-visible minorities vs number of accessible parks.

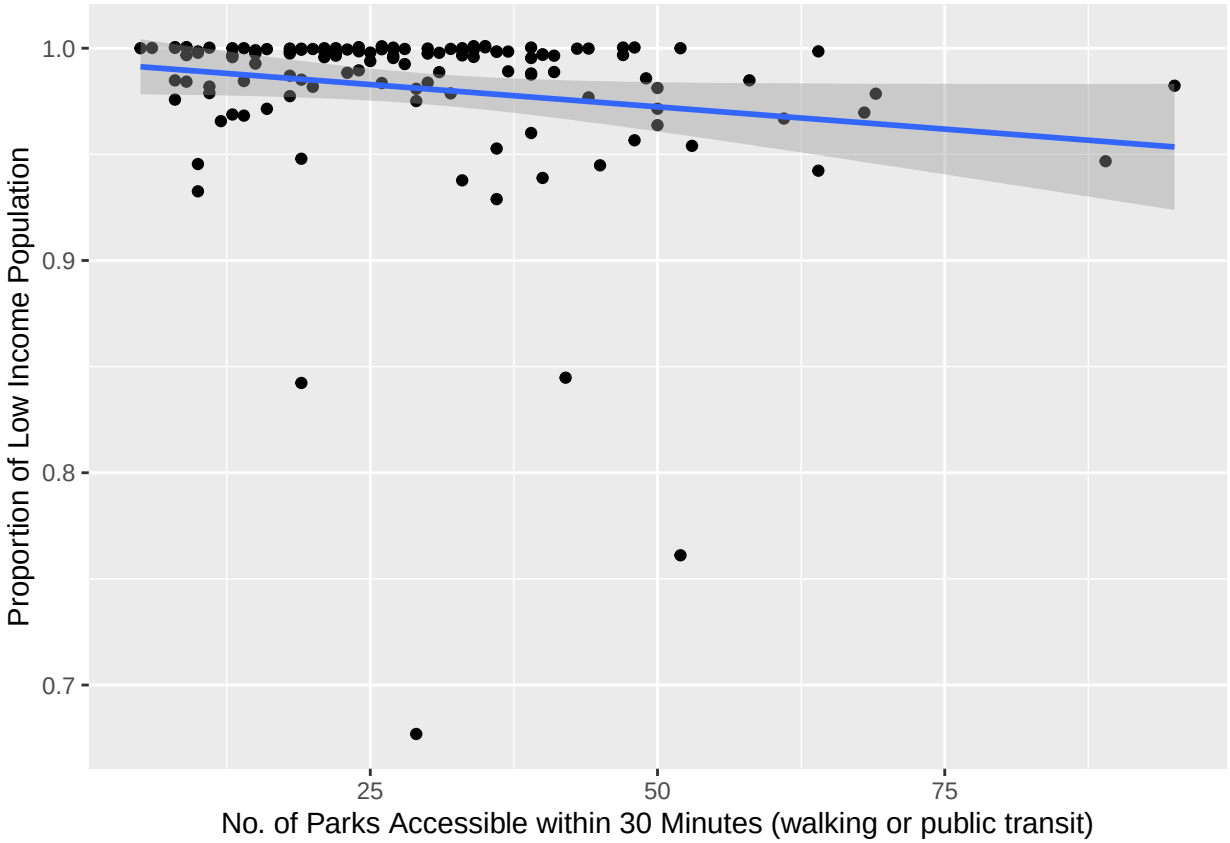


Model regressing proportion of low income population on number of accessible parks.

Table 5: Proportion of Low Income Population in Census Tracts regressed on Number of Parks Accessible

	Dependent variable:
	Proportion_low_income
accessibility	−0.0004*
	(0.0002)
Constant	0.993***
	(0.008)
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)
Note:	*p<0.1; **p<0.05; ***p<0.01

Scatterplot of proportion of low income population vs number of accessible parks.



Analysis

Conclusion