# THE NEED FOR INCREASED PUBLIC TRANSPORT DENSITY IN AREAS IN MELBOURNE

Prepared by: Group 73 Jay Cutler

Preparation for COMP20008: Elements of Data Processing

Lecturer: Pauline Lin & Chris Ewin

18th May, 2021

# Table of content

# Section I

## Introduction to the study

### I. Introduction

With a growing population in recent years, it is necessary to investigate whether Melbourne's public transport system provides convenience (access to stations) for higher numbers of travellers. This research subject addresses the liveability and sustainability in Melbourne.

### II. Significance

The results of this study help determine which suburb requires modifications in its current public transport network, by analysing data such as the number of stations, minimum distance to the nearest stations for each suburb. Insights gained can help the government decide on the logistic aspect of building new train/tram stations or expand their railing network, to cover more regions and benefit everyone within.

### III. Methodology

1. *Source of data*

Data for this research were collected from Aurin, the Australian Bureau of Statistics and Spatial Datamart Victoria.

From Aurin, we collected regional and metro stations csv files, with coordinates of each station. This, combined with polygons for different regions in shp files collected on Spatial Datamart Victoria, allowed us to connect the datasets and establish correlations .

We web scraped from the Australian Bureau of Statistics website and exported the data collected into a csv file for ease of manipulation.

2. *Sample selection*

Sample size is narrowed down to be within Greater Melbourne. For each means of transport, we investigated the region bounded by the centroids of the farthest suburbs reached by the transport system in each direction.

3. *Statistical method*
    a. Preprocessing:
        i. We collect the data of car number per household, average income, population of each suburb by web scraping and put it into a csv file of Data Frame

ii. From the .shp file, we use geopandas to read the file and put it into a geographical dataframe that contains all the suburb polygons and postcodes.

b. Data Processing:
   i. To calculate the average distance to a stop taken by traveller
      1. For suburb containing some stops:
         a. Use random lib to generate x and y coordinates for a random point.
         b. If the point is in the suburb, then calculate the distance (Descartes) to the nearest stop. We will then take the average of 100 random points.
      2. For suburb without any stops:
         a. Take the centroid of the suburb
         b. Iterate over all stops to find the distance to the nearest stop.
   ii. For calculating the station density and population density
      1. The number of stations or population was divided by the area of the respective suburb.
   iii. To support heatmap and plotting data, we link the data by putting them into the Data Frame or Geographical Data Frame.
   iv. For numerical data representation, such as the number of train stations in a suburb, plot the data using geopandas on top of a spatial map of Greater Melbourne as a heatmap.
   v. For visualisation of correlation between data such as population we use scatter plots from matplotlib. If some data is seen to have some shape of a 'familiar' equation, fit the data with a line (for example, linear regression or inverse function). Therefore, we can discuss their correlation.

## IV. Scope of the report

Our research focuses primarily on the region that a specific public transportation is to be used (Metro for tram and Greater Melbourne for bus and train). Suburbs containing a station and adjacent areas are investigated. To ensure correctness and consistency, we use data collected from 2016 across all findings.

## V. Limitation of the study

Some information is hard to retrieve, such as the proportion of public transport users by each postcode. Those available are on a national or state level, while our study requires data from each suburb. In such cases, our group used web scraping, where we encountered information of certain suburbs that was either unavailable in 2016 or missing. Other variables that would be beneficial to the research would be how far people from each suburb travel in a day and the timetable of the transportation system.

If such data had been available, we wouldn't have resorted to web scraping, which resulted in several missing entries in the dataset as discussed. Having timetable data would have been especially beneficial in visualizing the frequency of public transport travelling through certain areas.

From the technical aspect, geographical data were in coordinates, meaning the spherical nature of the Earth's surface has to be considered if we want more accuracy. However, as only Greater Melbourne, a small area, was investigated, no significant distance inconsistency was observed. Then we assume that the approximate metric equivalent of 1 degree to 110 km (Approximate Metric Equivalents for Degrees (usna.edu)).

# Section II

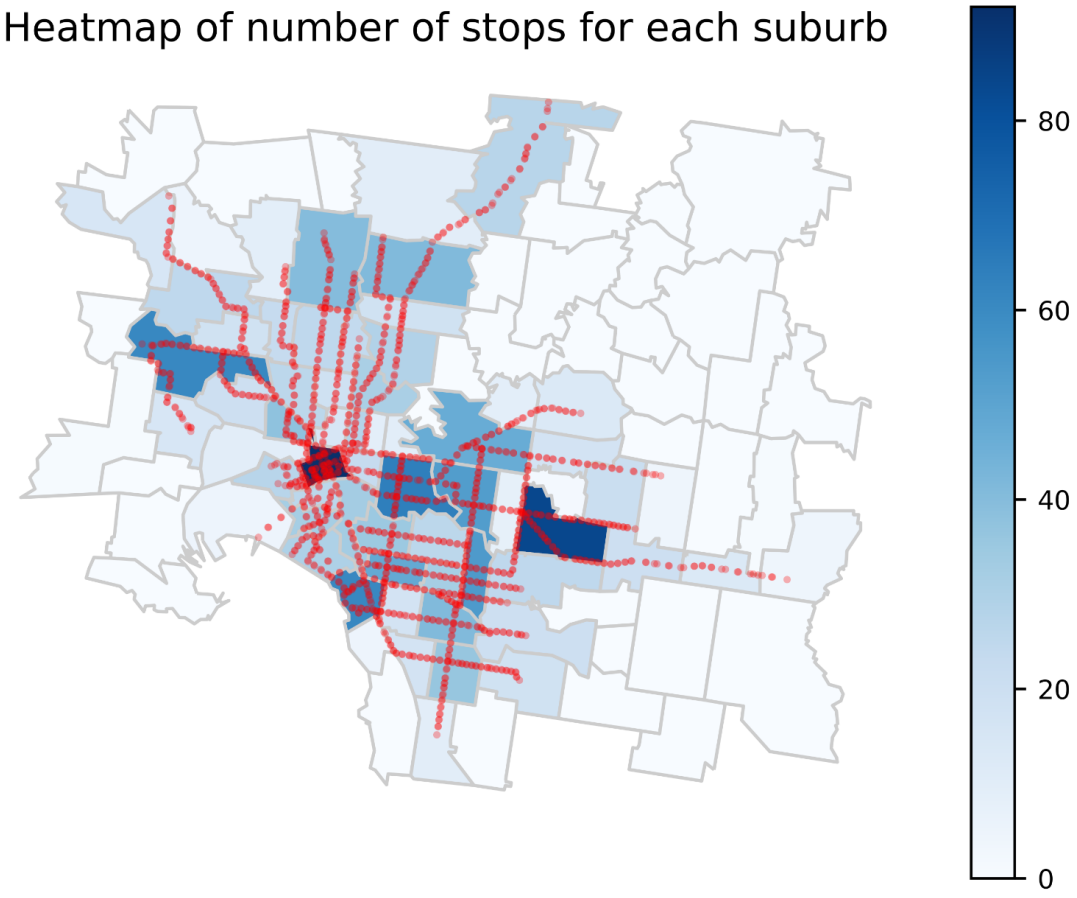
## Findings, conclusions and recommendations

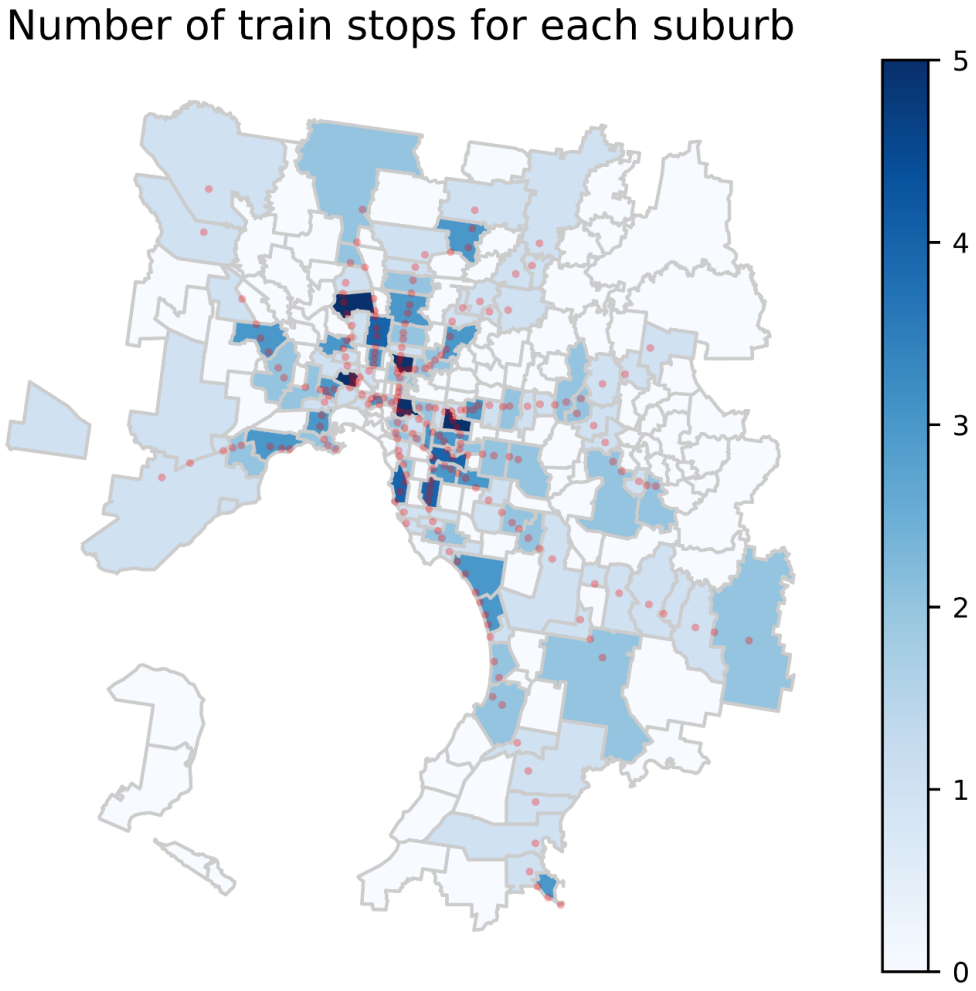
### I. Findings

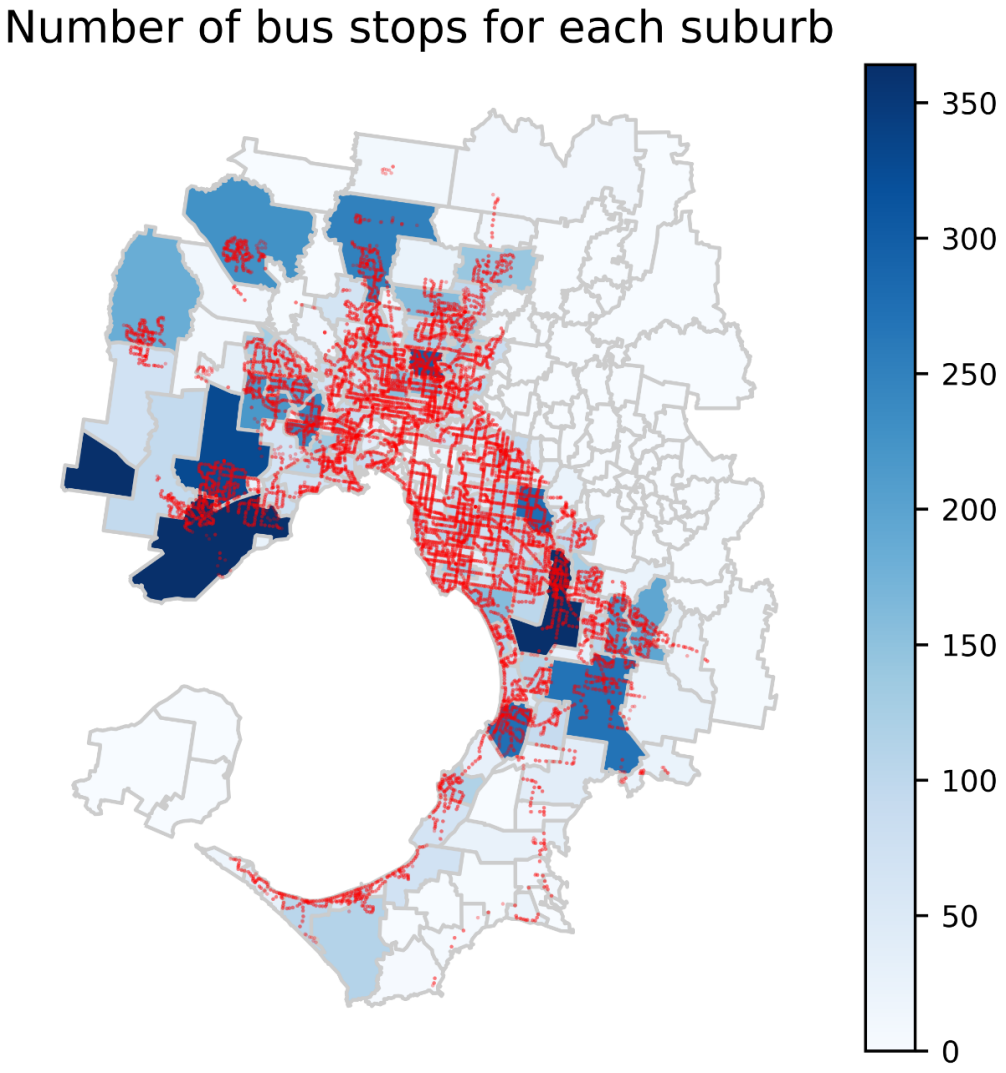
There are a few factors that dictate the distribution of public transport systems in different suburbs.

*1. Coverage*

Coverage is defined as the area one transport network covers. Based on this, bus is the means of transport with highest coverage, followed by train, and finally, tram. This can be explained by the ease at which new bus stops and routes can be constructed, as opposed to those of trains and trams.



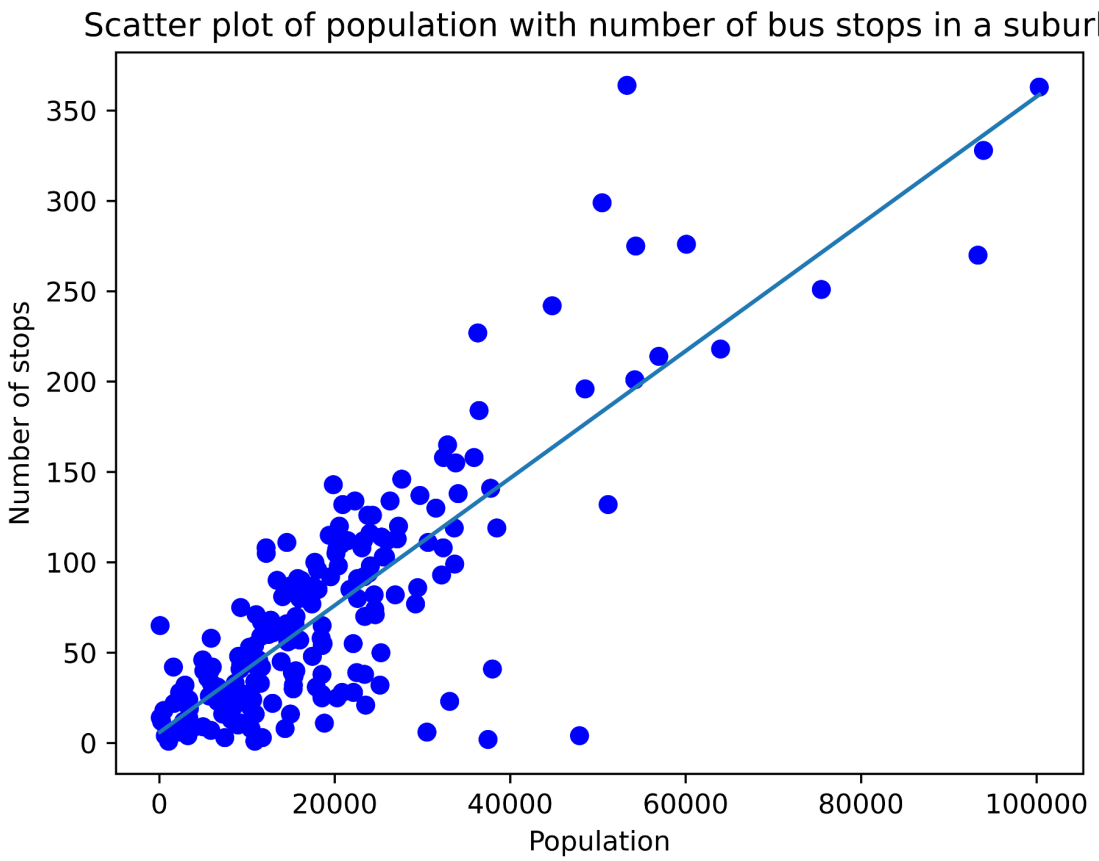

Number of train stops for each suburb
5
4
3
2
1
0



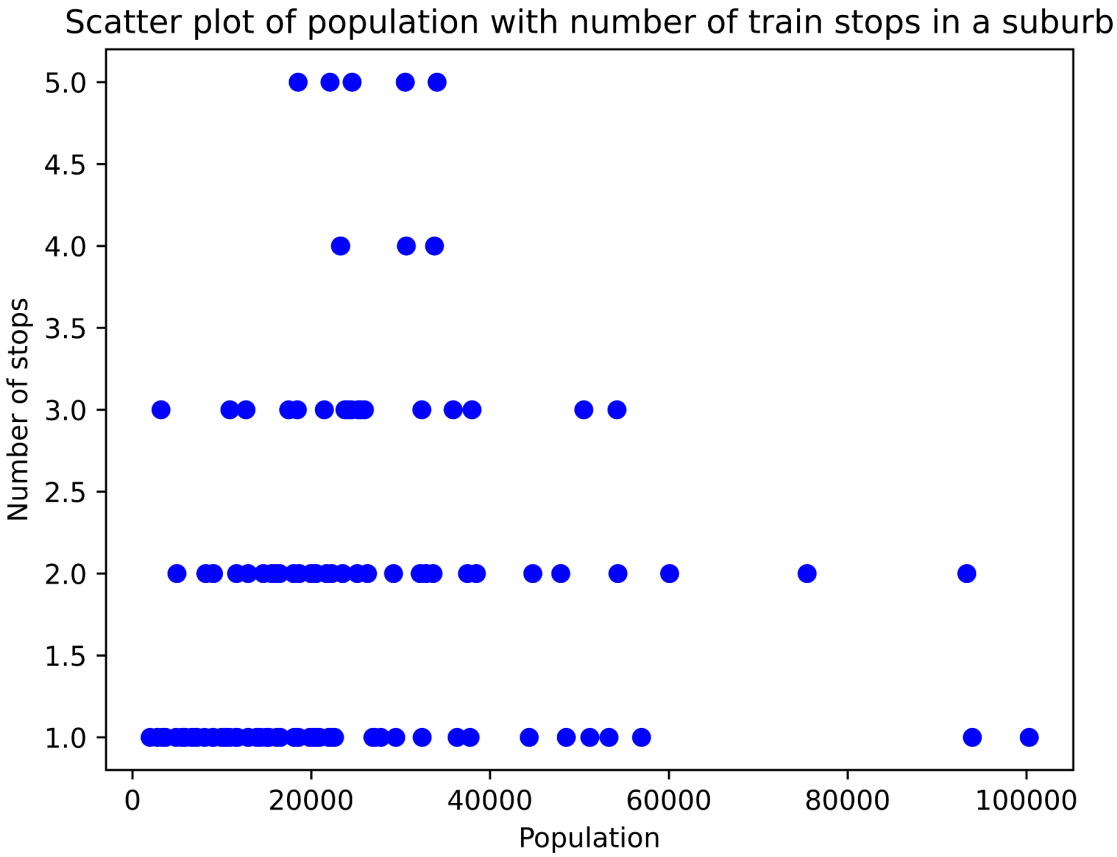

Based on the maps, the Central Business District (CBD) has adequate numbers of stations for all types of public transport, since most routes must travel through the center of Melbourne. Meanwhile, areas far from the CBD record a lower number of stations. However, there are some suburbs that have a significantly higher number of stops. These are the suburb containing transition stations,  therefore having more routes

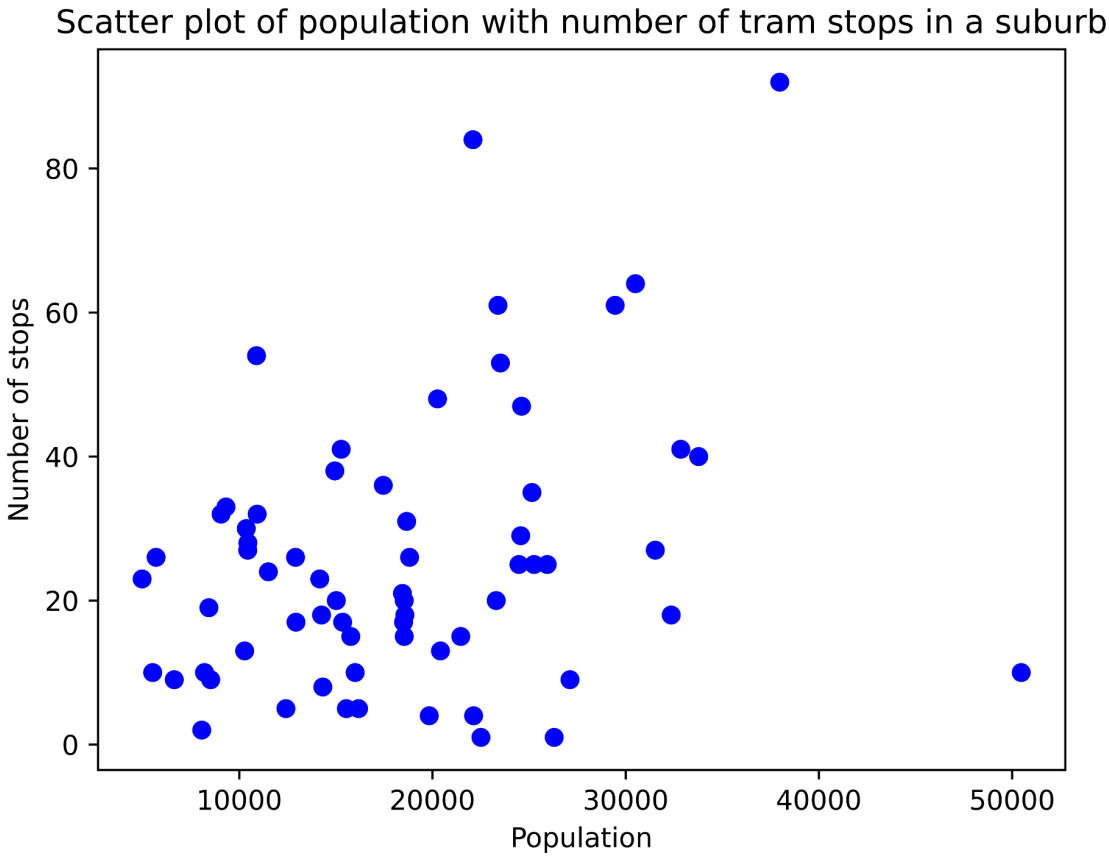
2. *Population:*

The population of each suburb is plotted against the number of stops of one type.




The line of best fit: Number of stops = 5.77467 + 0.00351 * Population with r^2 = 0.69965


Scatter plot of population with number of train stops in a suburb
Number of stops
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0
20000
40000
60000
80000
100000
Population



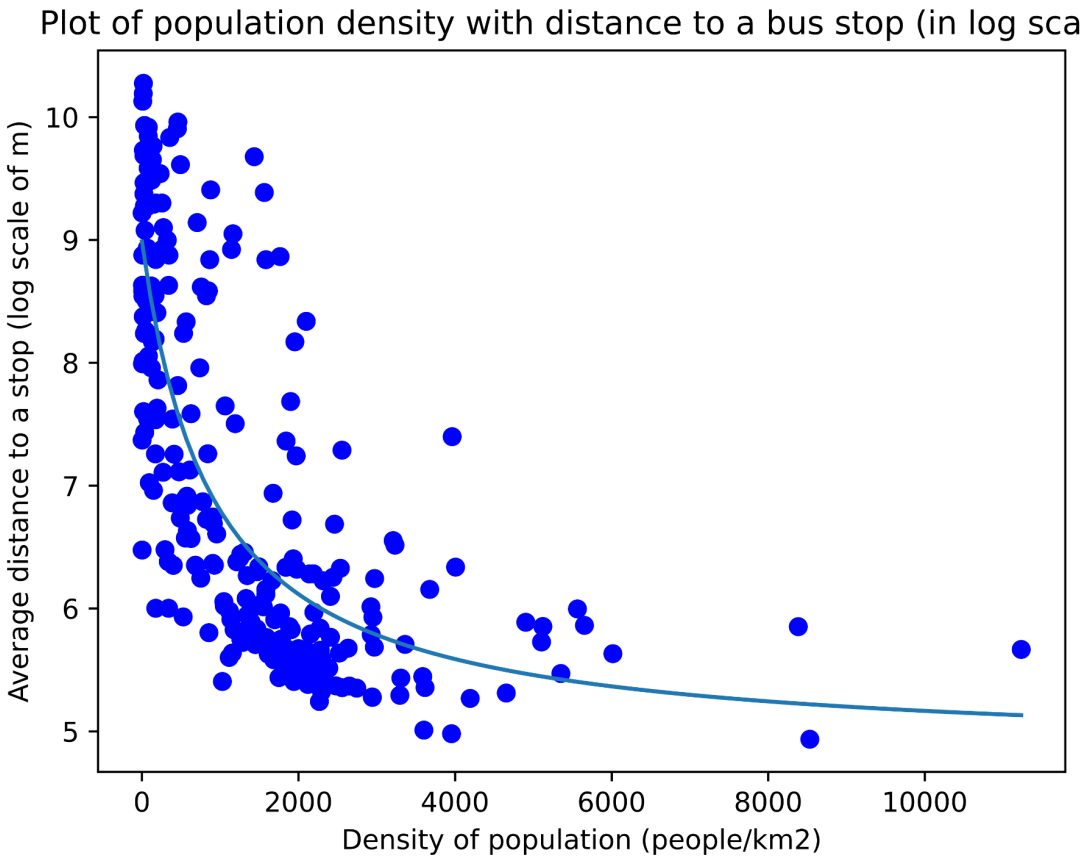

For the number of bus stops, we applied linear regression and observed a highly linear trend. For the other 2 variables, we could not apply linear regression since the points are sparse. The reason is that the tram and train systems tend to follow preconstructed routes with lower number of stops, hence are not really flexible.

Regarding the city's bus system, stations are evenly distributed. A highly positive correlation (fit line provided) can be seen between the number of bus stops and an area's population. A similar pattern, though not as strong, can be observed with the tram network, as its distribution is sparser.

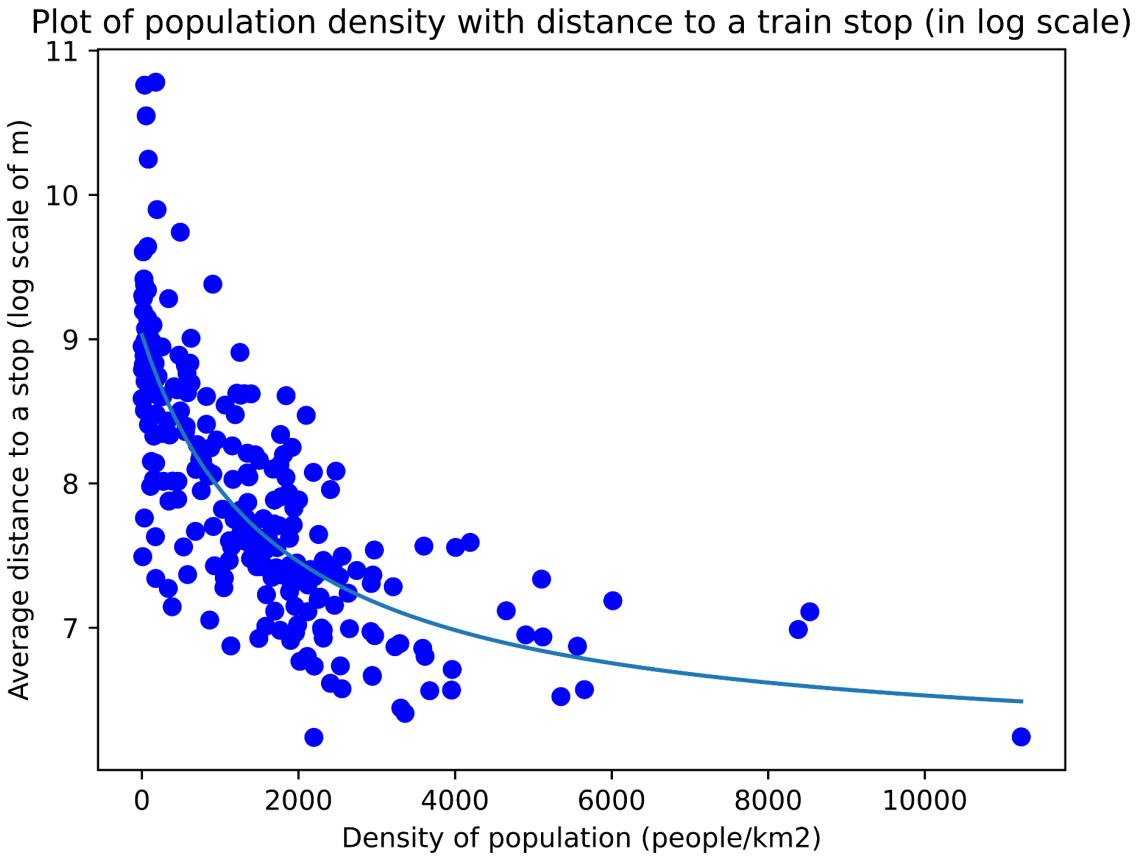
In contrast, the distribution of train stops is the least dense, leading to each suburb only having a small number of stations located within. Therefore, the line of best fit does not provide much insight but an insignificant positive correlation.

3. *Population density:*

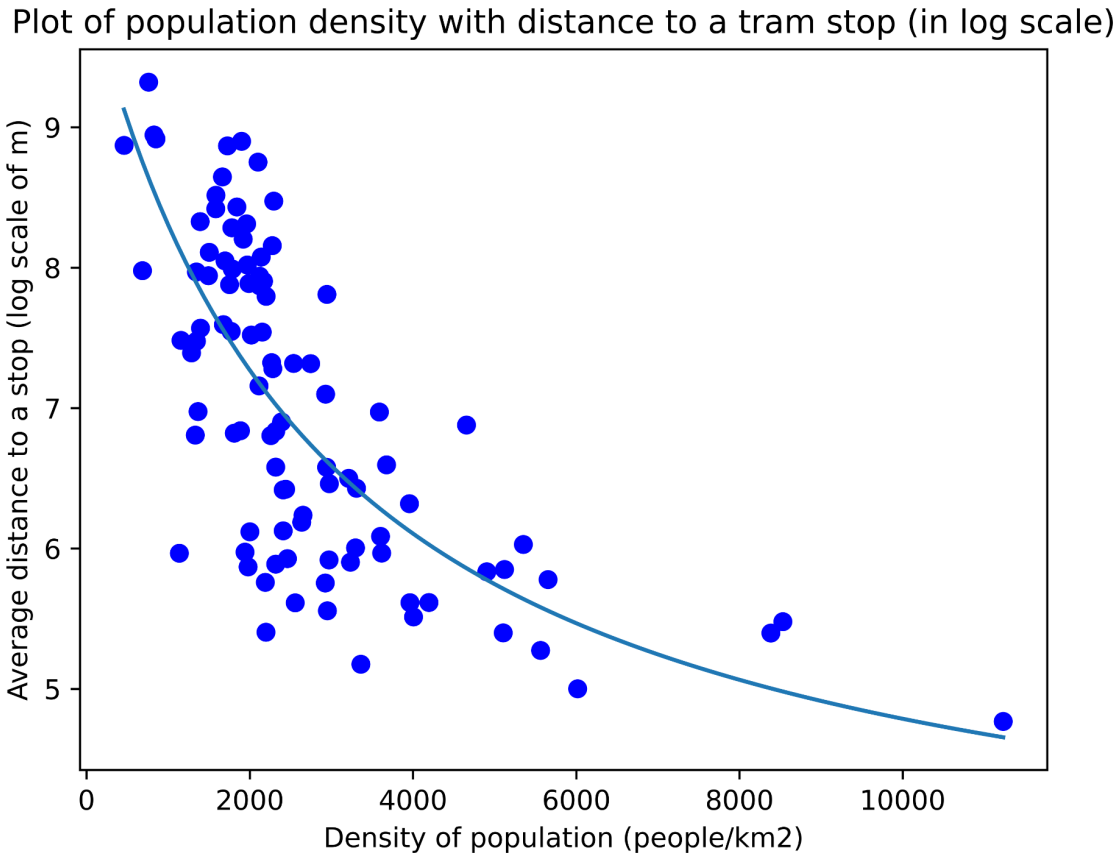
There is a high negative correlation between population density and the distance to travel to a nearest stop of any type. That is, the more densely populated a region, the shorter the average distance to get to a station in that region. Data points follow the fitting line well.




Best fit line: log(avg distance to bus stops) = 4.81415 + 3784.49851/(population density + 906.667424)



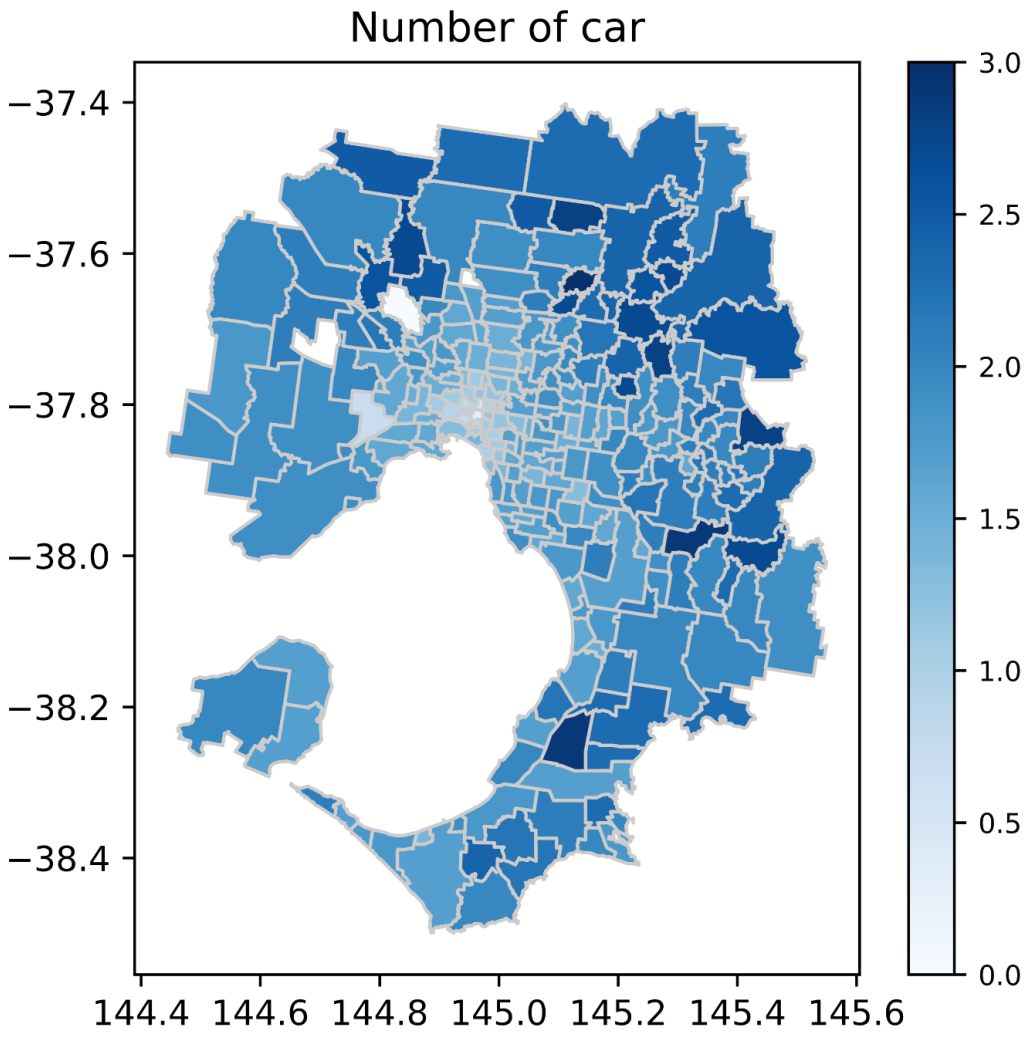

Best fit line: log(avg distance to train stops) = 6.04571 + 5363.78287/(population density + 1797.130438)



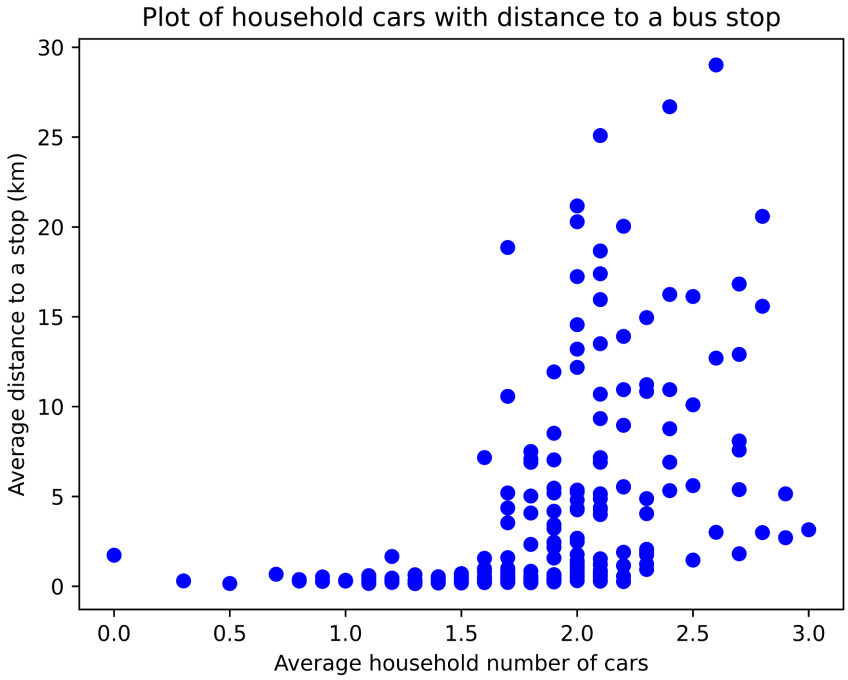

Best fit line: log(avg distance to tram stops) = 3.20249 + 20204.51391/(population density + 2962.54494)

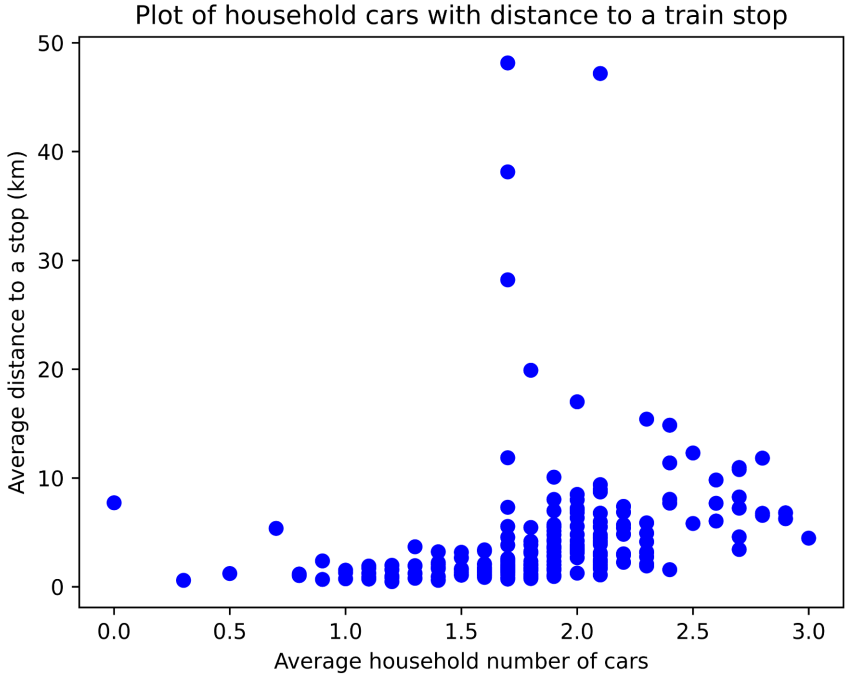
*4. Ownership of personal vehicles*

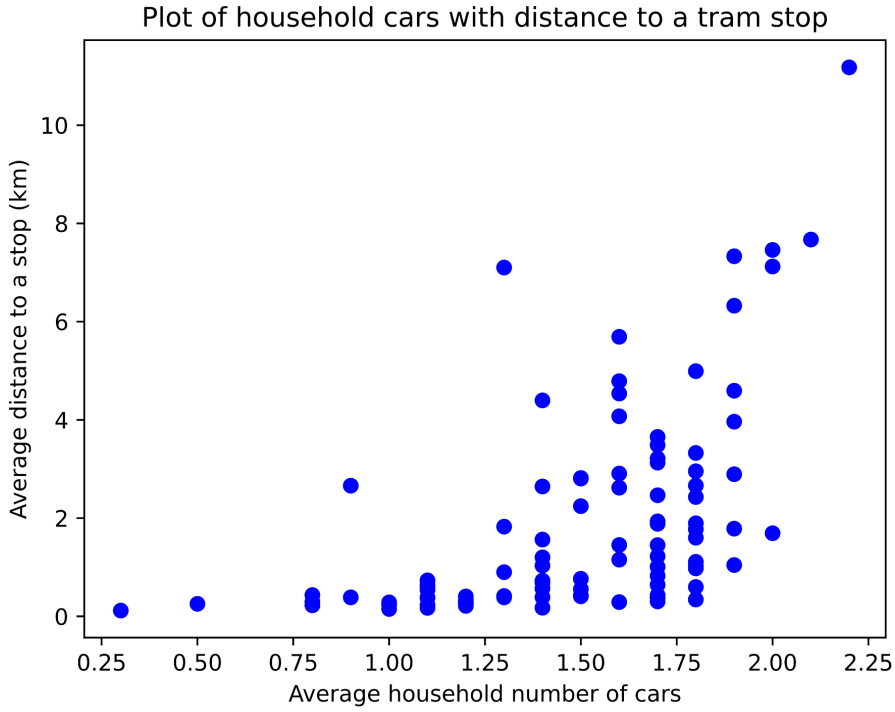
People living closer to the CBD tend to choose against cars, unlike their suburban counterparts. To be specific, on average, citizens in the suburbs own from 1.5 cars per household, while for people in the center, this number is well below 1.



We observed that the further the distance from the CBD, the higher the number of cars each house has, indicating where there is plenty of choice of public transport, people tend to use them instead of cars. The number of cars is positively correlated with the minimum required travel distance to the nearest stop, illustrated as follows:



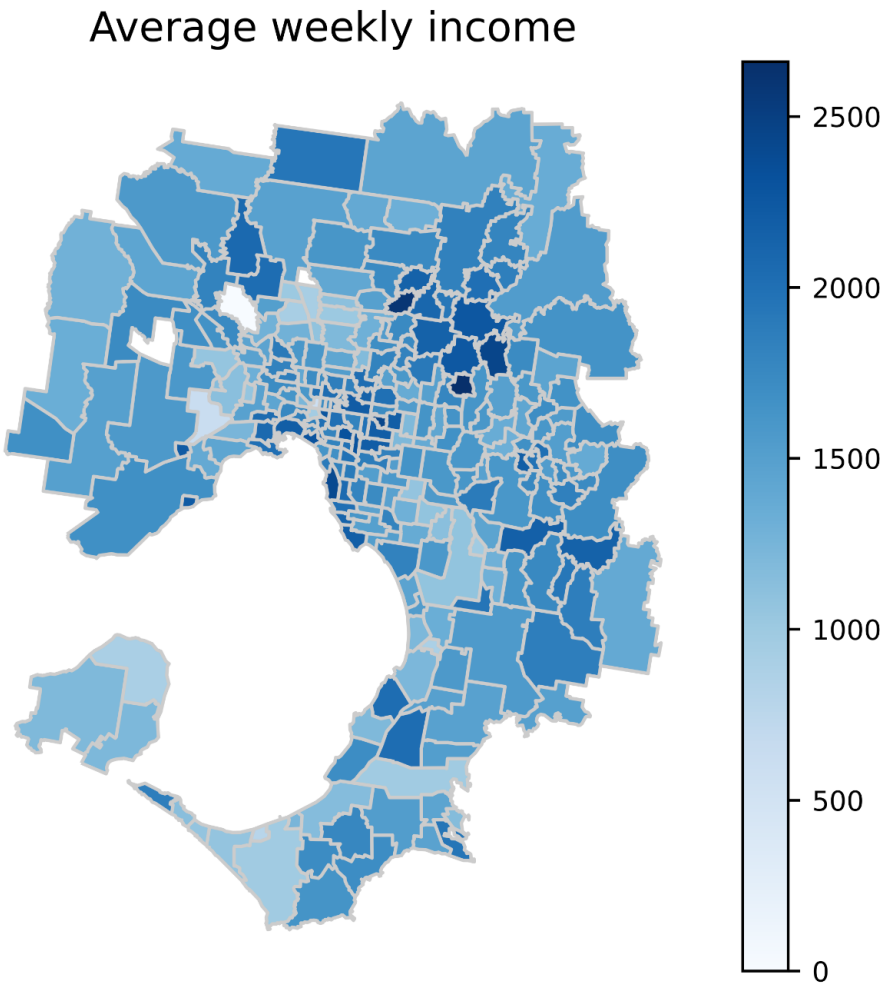





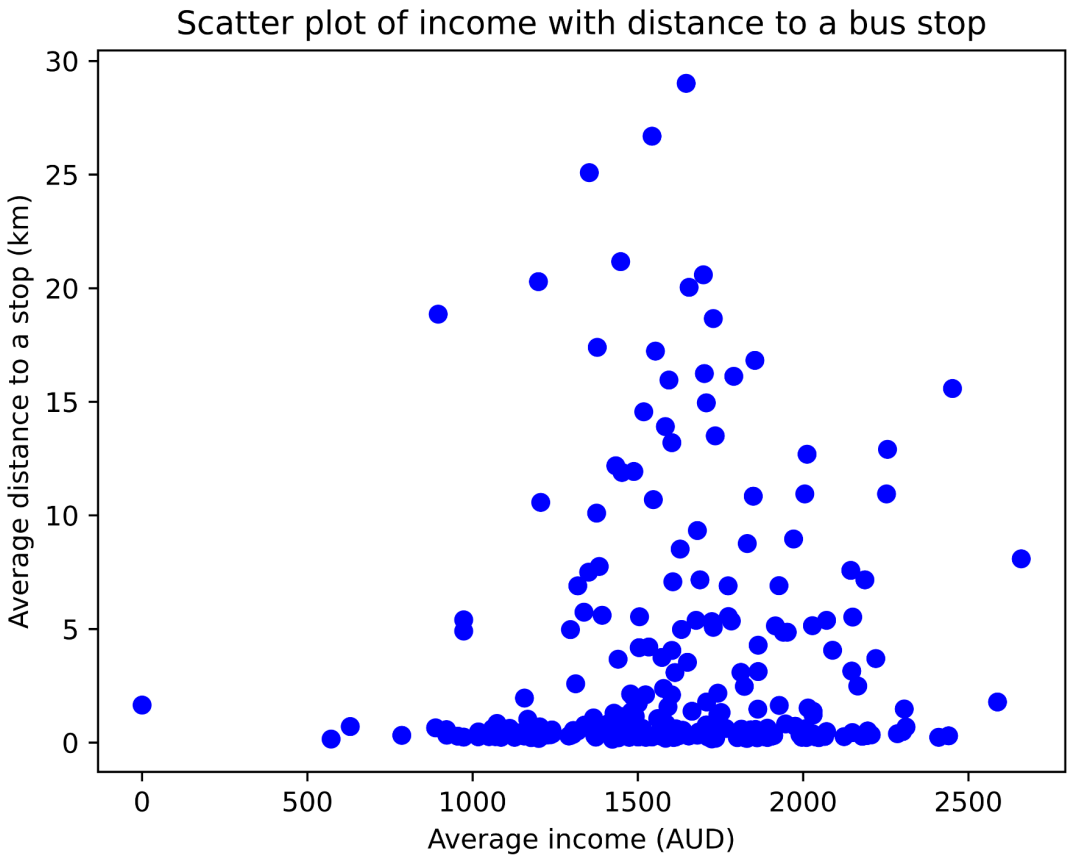

Households with less than 1 car live relatively close to a stop of any kind, while households in suburbs where each family has more cars tend to live the farthest from any stop. The breaking point is about 1.5 cars for all of the 3 plots, as 2 cars are usually enough for a family. However, after the breaking point, distance to stop shoots up and stays relatively unchanged as the number of cars increases. This is because 1.5 cars per household is a reasonable figure to no longer need public transport, or because the farther they live, the more cars they have to own.

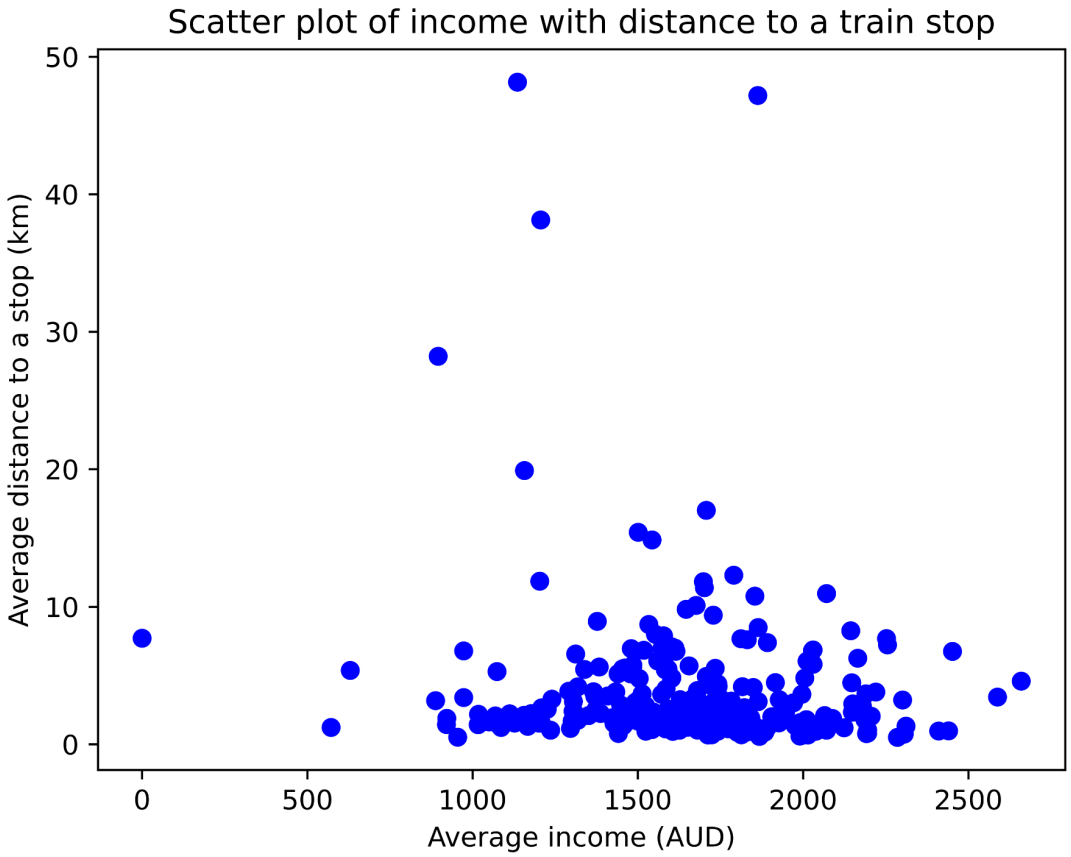
5. *Income*

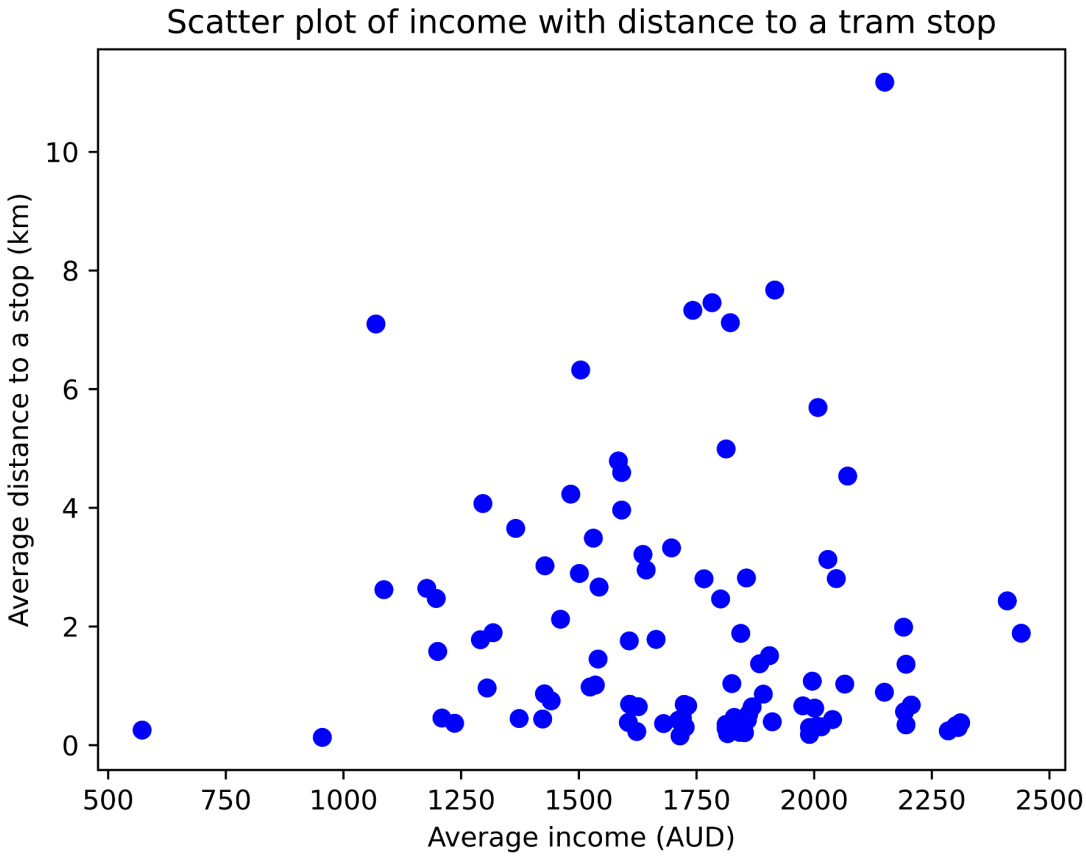
Income distribution does not follow any trend.



For each suburb, average weekly income is then plotted against average distance to the nearest stop:




Scatter plot of income with distance to a train stop
Average distance to a stop (km)
50
40
30
20
10
0
0
500
1000
1500
2000
2500
Average income (AUD)




In all 3 cases, no linear relationship is observed. These plots are mostly for illustration purposes. Points are sparse at the range of overall average income with few outliers. This implies that average weekly income does not correlate with the minimum distance to a stop. However, the lower clusters in these plots emphasizes that in most cases, the distance to a stop is not really high.

## II. Conclusion

The city of Melbourne has done a generally good job in planning and constructing its public transport infrastructure for residents. It does well to address the needs for transportation in suburbs with different characteristics, such as population, population density and the average number of cars a household owns.

## III. Recommendations

Based on the findings and conclusions of the study, the following recommendations are made:

1. Most modifications should be made to the city's bus system, since it is the most flexible with regards to the construction of new routes.
2. There should be an increase in stops and routes to the current bus system going through suburbs where households have on average 1 to 2.5 cars or regions with a low average income.
3. There should be extra stops for suburbs whose population consists of people earning $1000-$1500 per week.
4. Since the system is well provided for the intended regions, the government should explore the system wider in reaction to population growth and redistribution of population in the future.

## IV. Result limitation and possible improvement:

- Our research hasn't come up with a method to determine which specific suburb needed support and additional public transport.
- Additional research needs to be done with collecting timetables of typical routes through each suburb to determine whether public transport frequency needs to be adjusted. This can reflect the commuters' flow and its efficiency, further enforcing the need of point 1.

**References:**


Australian Bureau of Statistics, 2016 - Census QuickStats, Retrieved from https://www.abs.gov.au/websitedbs/D3310114.nsf/Home/2016%20QuickStats on 2021-05-18

Public Transport Victoria, 2019, PTV - Metro Bus Stops (Points), Retrieved from https://portal.aurin.org.au on 2021-05-18

Public Transport Victoria, 2019, PTV - Metro Train Stations (Points), Retrieved from https://portal.aurin.org.au on 2021-05-18

Public Transport Victoria, 2019, PTV - Metro Tram Stops (Points), Retrieved from https://portal.aurin.org.au on 2021-05-18

Public Transport Victoria, 2019, PTV - Regional Bus Stops (Points), Retrieved from https://portal.aurin.org.au on 2021-05-18

Public Transport Victoria, 2019, PTV - Regional Train Stations (Points), Retrieved from https://portal.aurin.org.au on 2021-05-18

Spatial Datamart Victoria, 2021 - Spatial map of Victoria, Retrieved from https://services.land.vic.gov.au/SpatialDatamart on 2021-05-18