

# ProblemSet2\_\_new

*Ke-li Chiu & Diego Mamanche Castellanos*

*06/02/2020*

## Abstract

Education is a contributor to many beneficial socio-economic outcomes. We intend to examine the relationship between academic majors and household income in Toronto's neighbourhoods—can we find higher or lower percentage population in certain academic majors in higher or lower income neighbourhoods? The result shows that higher income neighbourhoods also have higher percentage of residents with majors in business. The data set is retrieved from the package *City of Toronto Neighbourhood Profiles* from the year 2011. ##### We intend to use the comparison to bring awareness of inequality in income and education existing in the City of Toronto.

## Introduction

Education is a contributor to many beneficial socio-economic outcomes. In this study, we intend to examine the relationship between household income and educational profile in terms of residents' academic majors in Toronto's 140 neighbourhoods. The majors are categorized as follows— “education”, “visual and performing arts and communications technologies”, “humanities”, “social and behavioural sciences and law”, “business\_management\_and\_public\_administration”, “physical and life sciences and technologies”, “mathematics computer and information sciences”, “architecture engineering and related technologies”, “agriculture natural resources and conservation”, “personal protective and transportation services” and “no postsecondary certificate diploma or degree.” The question we want to answer is if we can find a higher population percentage in specific academic majors in higher-income neighbourhoods or vice versa? The question is relevant because the academic major is a significant factor for people's occupations and their associated earning; thus, it is deemed as an essential tool for social mobility.

## Dataset and Method Description

City of Toronto Neighbourhood Profiles data set is sourced from several Census tables released by Statistics Canada every five years. The dataset uses this Census data to provide a portrait of the demographic, social and economic characteristics of the people and households in each City of Toronto neighbourhood. Each data point in this file is presented for the City's 140 neighbourhoods, as well as for the City of Toronto as a whole. The data set consists of 144 columns for neighbourhoods and 1537 rows for characteristics of the neighbourhoods. Data cleaning and reshaping are performed to have a new data frame that has 14 characteristics regarding topics of income as columns and education and the 140 neighbourhoods as rows.

Exploratory data analysis is conducted to obtain insight from the dataset. To get a picture of a neighbourhood's economic status, we look at the median total household income which divides the income distribution into two equal groups, half having income above that amount, and half having income below that amount. As for the education profile, the academic majors of the residents are the variables in which we are interested. Each field for the major is the number of people who took this major in the neighbourhood. To be able to compare the neighbourhood regardless of the difference in population, we turn the value to percentages by dividing the total population between 25 to 64 years old by the population in each major. All codes are available in Appendix A.

## Limitation of the Dataset and Approach

Because we chose to look at the median instead of the mean of household income, we are excluding the outliers in the income distribution. The median value would give us a more accurate picture of the neighbourhoods' economic status but also prohibit us from investigating further the outliers that are skewing the distributions. Moreover, although there are datasets for 2001, 2006, 2011 and 2016, the structure and available information of the datasets are inconsistent. For example, median total household income for all neighbourhoods is only available in the dataset from 2011; therefore, this study is limited to the year 2011 and is unable to provide a comparison between Census years regarding the changes in income and education profile of the neighbourhoods. Finally, we are also not able to find out the income distribution by academic fields. Therefore, earnings differences linking to differences in residents' fields of study are not included in our analysis. This prevents from knowing if specific academic majors are more significant contributors to the neighbourhoods' total household income.

## Ethical Considerations

Population percentage in specific majors have correlations with median household income in the 140 neighbourhoods. For example, higher income neighbourhoods have more percentage of business majors, and lower income neighbourhoods have more percentage of transportation majors. This observation, without further analysis, potentially projects social status and ranking among academic fields.

## Income Distribution Map

Darker areas in the above map indicate high levels of unemployment, while lighter areas indicate low levels of unemployment. As Figure 1 shows, we observe that higher-income neighbourhoods tend to be in the central part of Toronto, neighbouring each other.

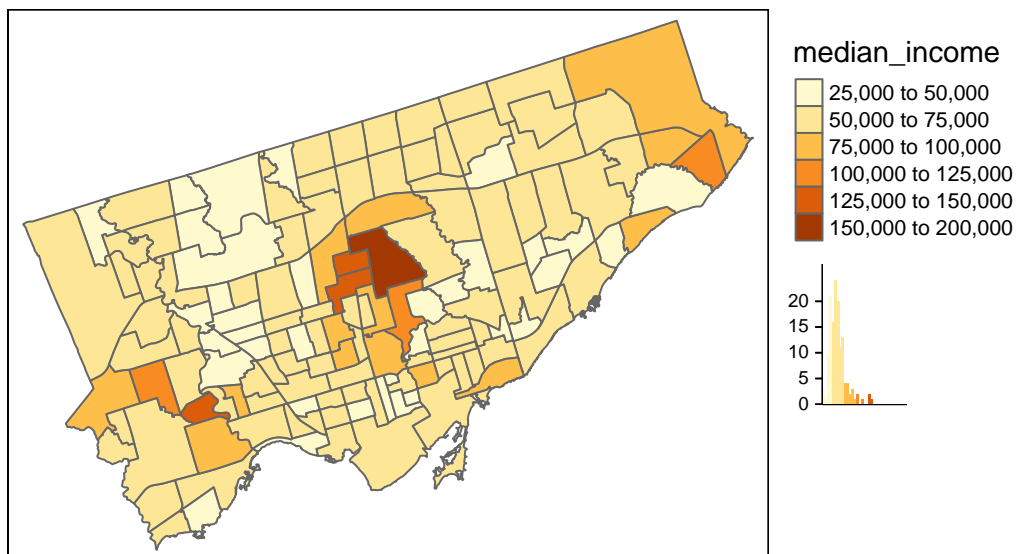


Figure 1: Median household income map

## Correlation Between Household Income and Academic Majors

We conducted a correlation matrix table to see if certain majors correlate with household income. As a result, shown in Table 1, “business management and public administration,” “education,” “social and behavioural sciences and law,” have stronger positive correlations with household income. In contrast, “no certificate” has a strong negative correlation.

major	correlation
no_certificate	-0.49
transportation	-0.46
architecture_engin	-0.09
agriculture	-0.05
mathematics_&_it	-0.01
arts_&_com	0.06
physical_&_life	0.26
health	0.36
humanities	0.44
social_&_law	0.46
education	0.54
business	0.55

Table 1: Correlation score of income and academic majors

## Majors Distribution Maps

In Figure 2, the four categories that have the most significant correlation scores are plotted into choropleth maps to be compared with Figure 1. The distribution of “business management and public administration,” “education,” “social and behavioural sciences and law,” have a similar pattern with Figure 1, where darker colours tend to be around the central part of Toronto. The distribution of “no certificate” is contrasting with Figure 1, where the lighter area locates in the central part of Toronto.

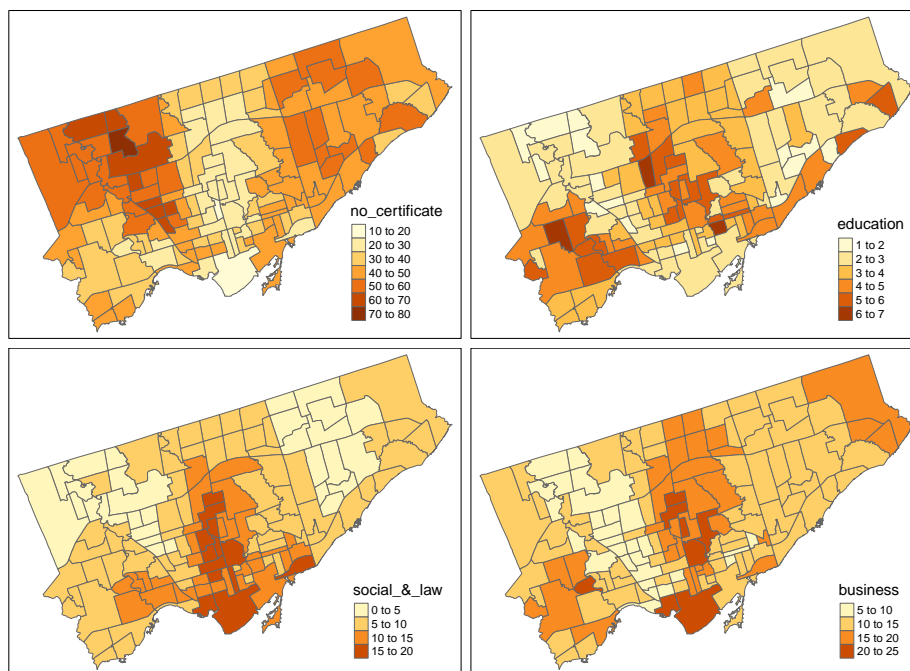


Figure 2: Distribution maps for selected majors

## Comparison Between Polarized Neighbourhoods and the City of Toronto

Finally, we gathered the population percentage in different majors for the neighbourhoods that have the highest income (Bridle Path Sunnybrook York) and the lowest income (Regent Park) and compared them with the City of Toronto. As Figure 3 shows, the shapes of the academic major distributions in the three groups are similar to each other; most population percentage lies in “no certification,” and the second-largest portion lies in “business management and public administration.” However, the scales of population percentages in different majors have apparent differences in the three groups. Bridle Path Sunnybrook York, the percentage of the population who have no certificate is 5.47 % higher than than the percentage of the population who major in “business management and public administration”. In Regent Park, the percentage of the population who have no certificate is drastically higher than the percentage of the population who major in “business management and public administration” by 37.31%.

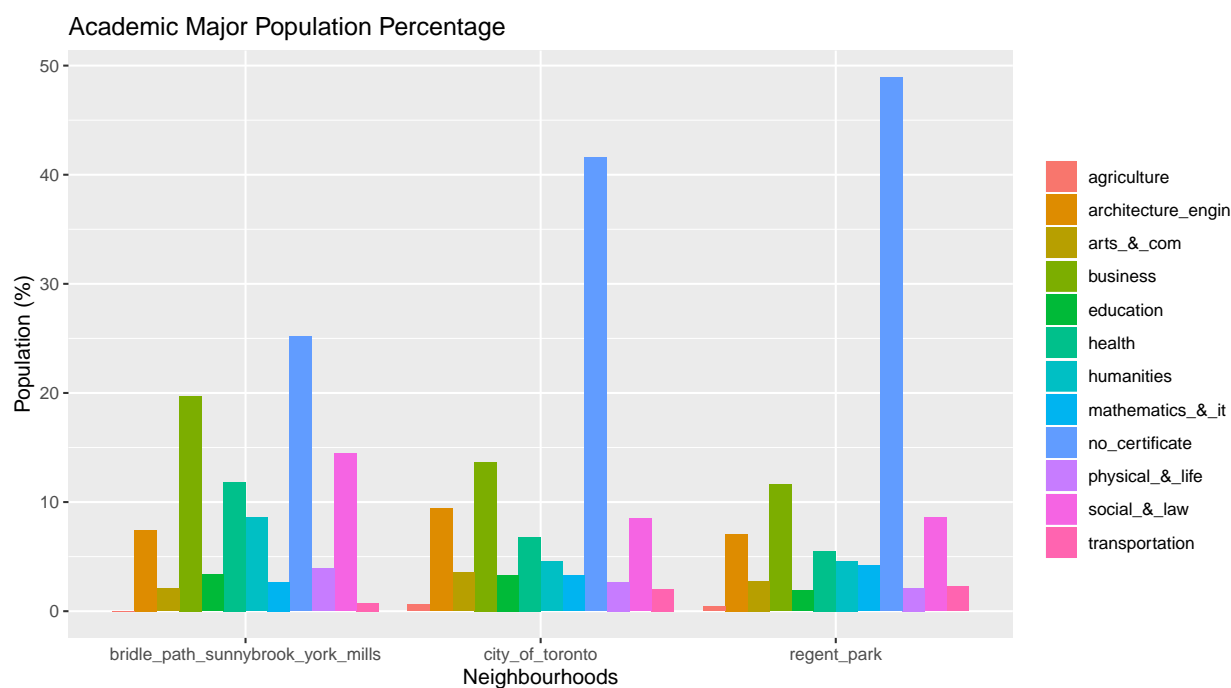


Figure 3: Population percentage in each academic major in different neighbourhoods