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 exam the education profile in the selected neighbourhoods of Toronto that have polarized household income from the data set package *Wellbeing Toronto - Demographics: NHS Indicators*. We first observe that the income gap between the

Introduction

Education is a contributor to many beneficial socio-economic outcomes. In the previous study where we examined the educational profile of Toronto's neighbourhoods, we found that the neighbourhood with the highest median household income has a much higher attainment rate (45%) of higher education above a bachelor's degree than the compared with the lowest (23%) and the City of Toronto (27%). In this study, we extend the examination of educational profile to the residents' academic majors in the following categories— "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 [the question]. [Broader context: education as social mobility]

Dataset and method descriptiion

City of Toronto Neighbourhood Profiles dataet is sourced from a number of 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.

Exploratory data analysis is conducted to obtain insight from the dataset. The variables we are interested in are median total household income and the academic fields the residents major or majored in. We would like to see if certain academic majors are more common or least common in different neighbourdhoods distinced by income. [Methods: map, lm, comparison…]

Limitation of the dataset and approach

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 neighbourhood is only available in the dataset from 2011, therefore this study is limited to the year of 2011 and is unable to provide comparison between Census years regarding the changes in income and education profile of the neighbourhoods.

Ethical considerations

By revealing the [. . .] Stereotypes Discriminations Transportation Ranking of academic fields by the associated,

```r
### Cleaning geo-location dataset
clean_geo_data <- janitor::clean_names(geo_data)

clean_geo_data <- extract(clean_geo_data, area_name, into = "neighbourhoods" , regex = "([^(0-9)]+)")


clean_geo_data["neighbourhoods"] <-
  janitor::make_clean_names(as.matrix(clean_geo_data["neighbourhoods"]))

clean_geo_data <- janitor::clean_names(clean_geo_data)

clean_geo_data <- select(clean_geo_data, neighbourhoods, longitude, latitude, geometry)


filter(clean_geo_data, neighbourhoods %in% c("mimico","weston_pellam_park"))
```

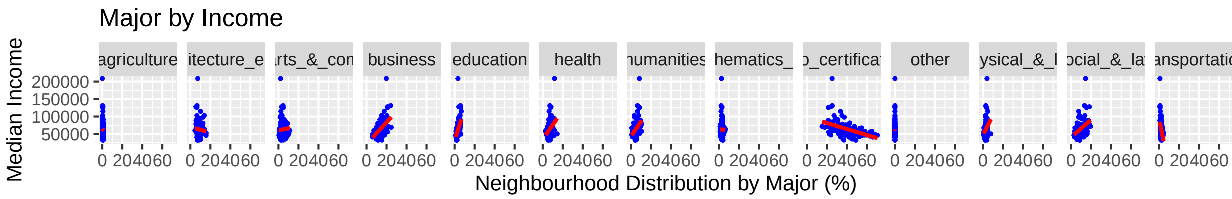
```
## # A tibble: 2 x 4
##   neighbourhoods   longitude latitude                                  geometry
##   <chr>                <dbl>    <dbl>                             <POLYGON [°]>
## 1 mimico               -79.5     43.6 ((-79.4804 43.62107, -79.48033 43.62~
## 2 weston_pellam_p~     -79.5     43.7 ((-79.46005 43.66723, -79.46092 43.6~
```

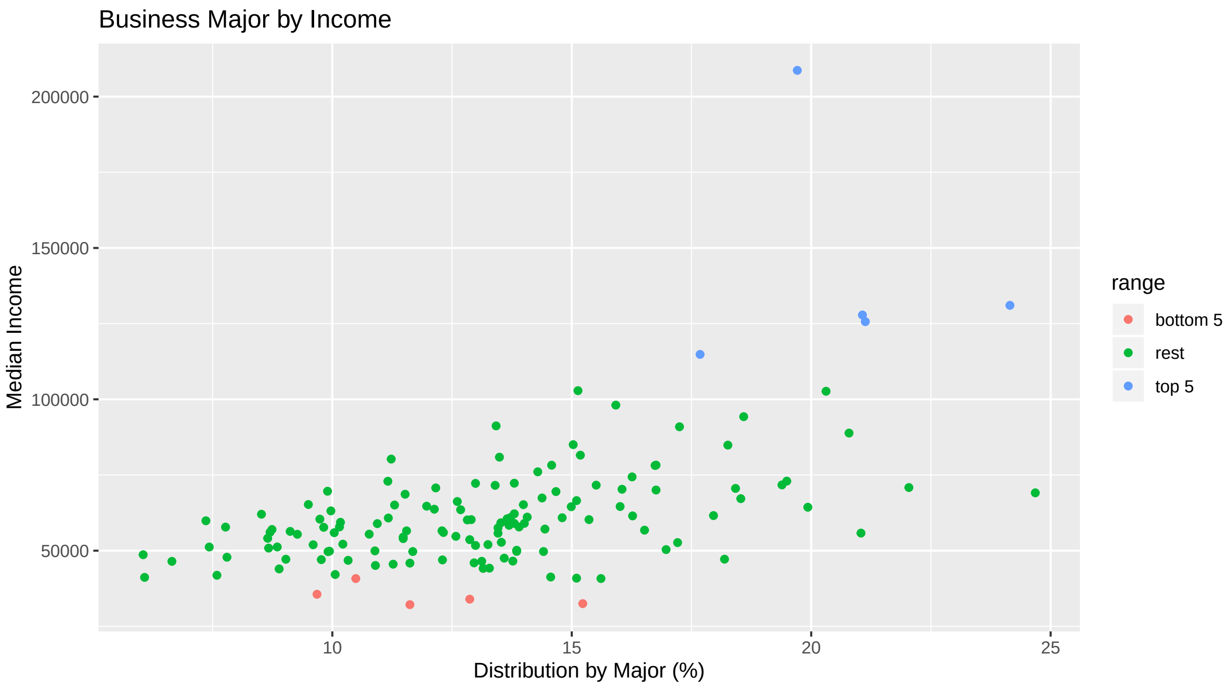
```r
clean_geo_data["neighbourhoods"][c(17,67),] <- c("mimico_includes_humber_bay_shores","weston_pelham_parl

clean_geo_data
```

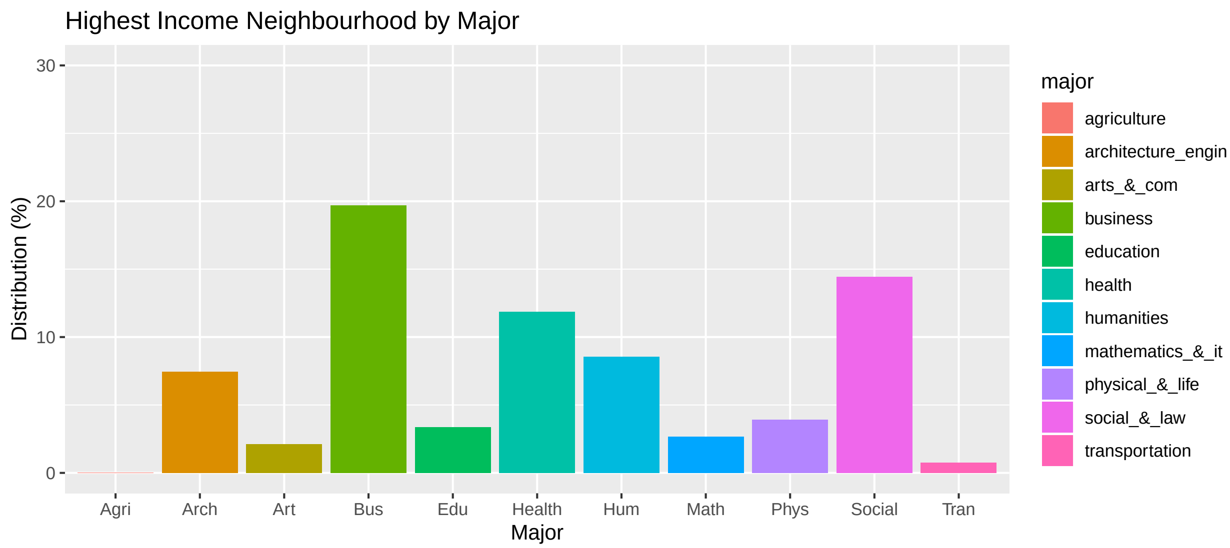
```
## # A tibble: 140 x 4
##    neighbourhoods    longitude latitude                                 geometry
##    <chr>                 <dbl>    <dbl>                            <POLYGON [°]>
##  1 wychwood              -79.4     43.7 ((-79.43592 43.68015, -79.43492 43~
##  2 yonge_eglinton        -79.4     43.7 ((-79.41096 43.70408, -79.40962 43~
##  3 yonge_st_clair        -79.4     43.7 ((-79.39119 43.68108, -79.39141 43~
##  4 york_university_~     -79.5     43.8 ((-79.50529 43.75987, -79.50488 43~
##  5 yorkdale_glen_pa~     -79.5     43.7 ((-79.43969 43.70561, -79.44011 43~
##  6 lambton_baby_poi~     -79.5     43.7 ((-79.50552 43.66281, -79.50577 43~
##  7 lansing_westgate      -79.4     43.8 ((-79.43998 43.76156, -79.44004 43~
##  8 lawrence_park_no~     -79.4     43.7 ((-79.39008 43.72768, -79.39199 43~
##  9 lawrence_park_so~     -79.4     43.7 ((-79.41096 43.70408, -79.41165 43~
## 10 leaside_benningt~     -79.4     43.7 ((-79.37749 43.71309, -79.37762 43~
## # ... with 130 more rows
```

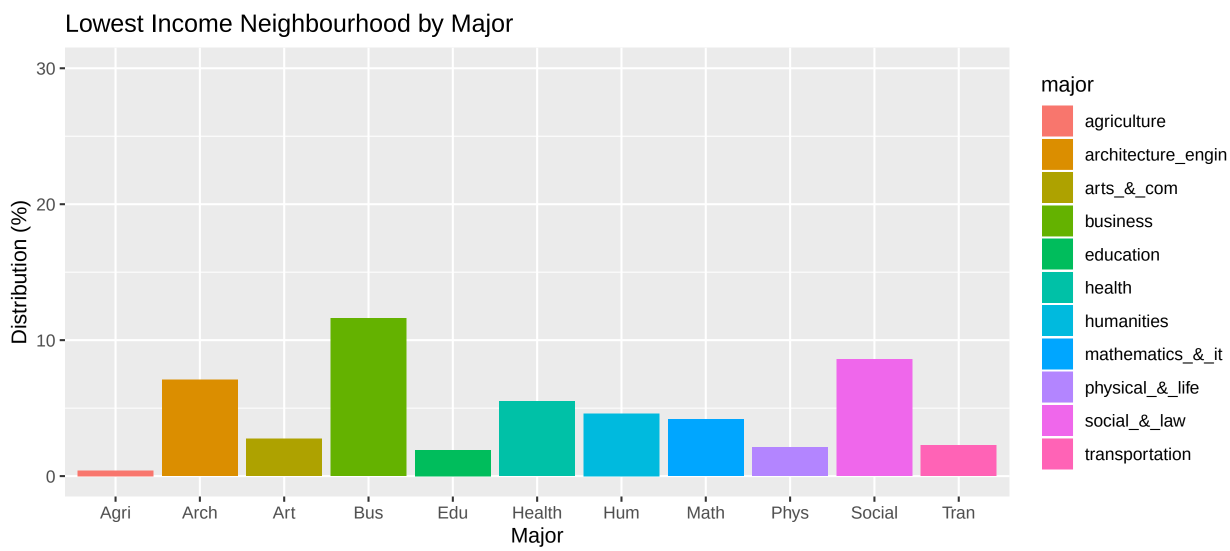
Income and Education by Major 2011

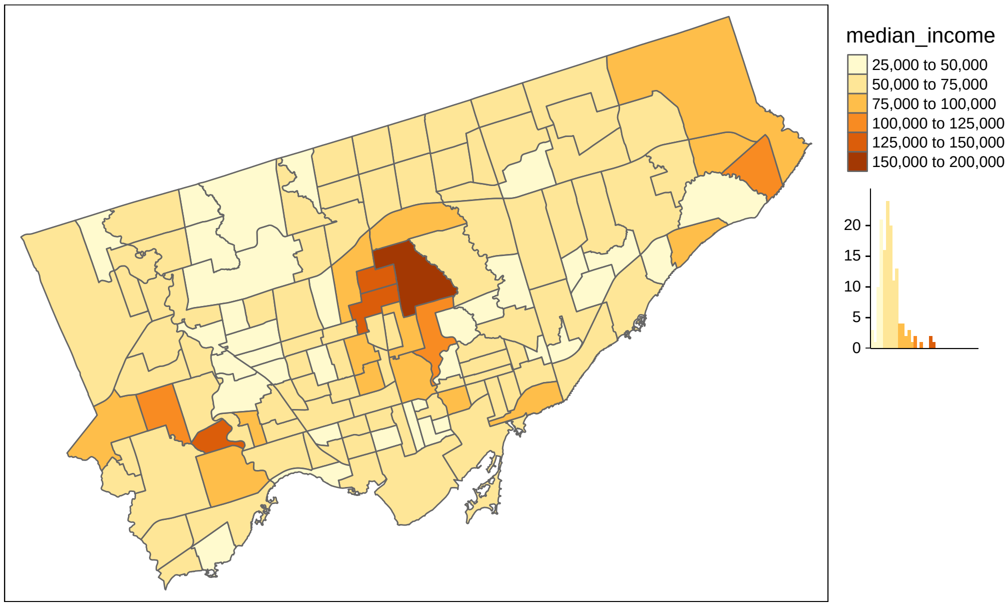


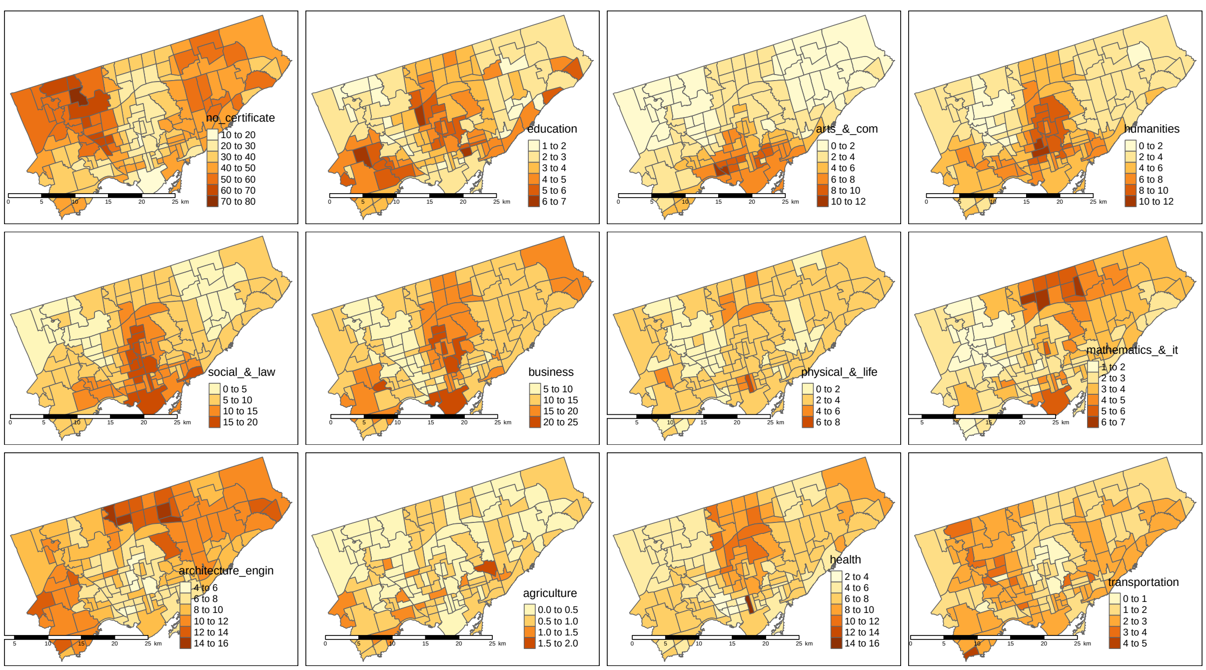



```
##
## Call:
## lm(formula = business ~ median_income, data = education_percentage_only)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.3977 -2.1601 -0.1705  1.6585 10.7048
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   7.473e+00  8.122e-01   9.200  4.8e-16 ***
## median_income 9.409e-05  1.227e-05   7.671  2.7e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.101 on 139 degrees of freedom
## Multiple R-squared:  0.2974, Adjusted R-squared:  0.2924
## F-statistic: 58.84 on 1 and 139 DF,  p-value: 2.699e-12
```


Highest Income Neighbourhood by Major
30
20
10
0
Distribution (%)
Agri
Arch
Art
Bus
Edu
Health
Hum
Math
Phys
Social
Tran
Major
major
agriculture
architecture_engin
arts_&_com
business
education
health
humanities
mathematics_&_it
physical_&_life
social_&_law
transportation


Lowest Income Neighbourhood by Major
30
20
10
0
Distribution (%)
Agri
Arch
Art
Bus
Edu
Health
Hum
Math
Phys
Social
Tran
Major
major
agriculture
architecture_engin
arts_&_com
business
education
health
humanities
mathematics_&_it
physical_&_life
social_&_law
transportation





major	correlation
no_certificate	-0.49
transportation	-0.46
architecture_engin	-0.09
agriculture	-0.05
other	-0.02

major	correlation
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