Problem Set 1 - Q1 - Analysis on Education Level and Income for Those who Earn less than $1000000

Chenyang Huan 1004728459

Sep.30.2020

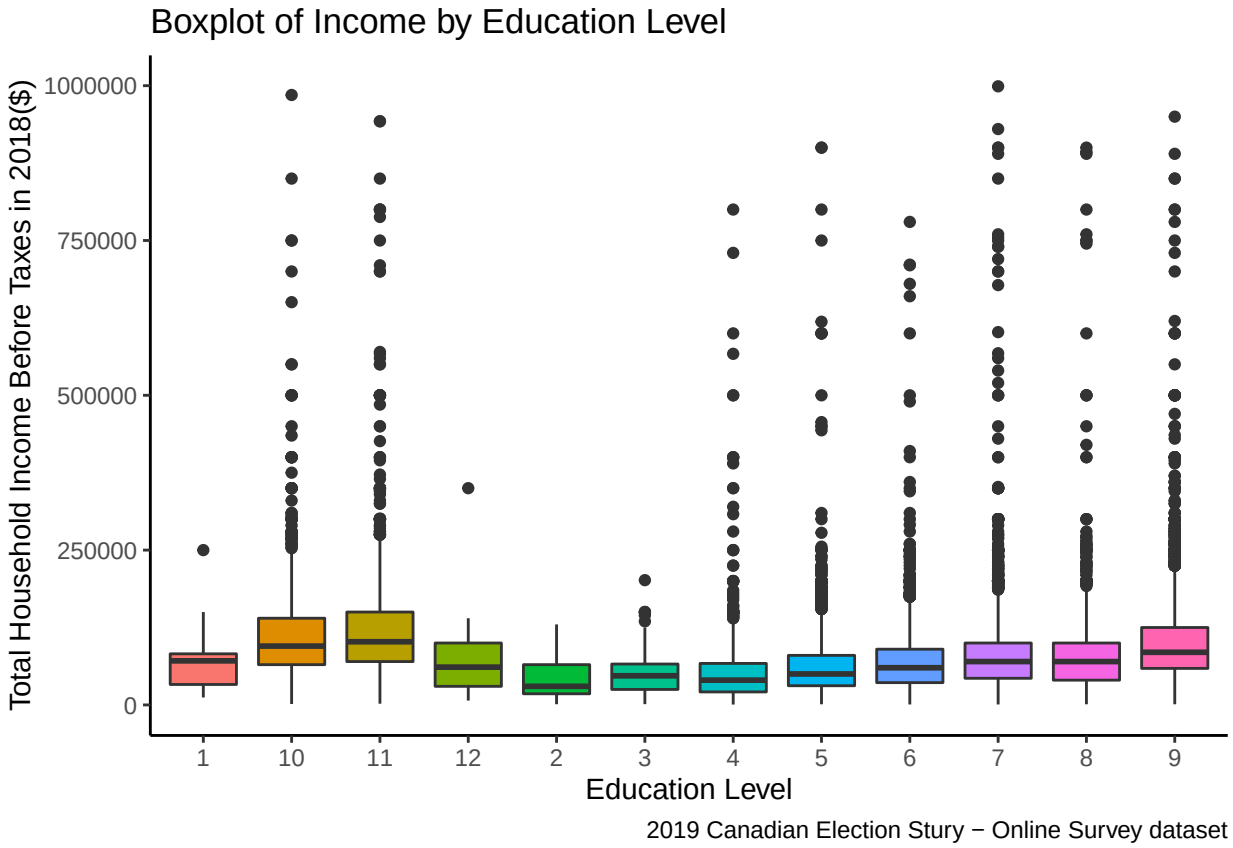
Question 1

Part a

I choose the 2019 Canadian Election Study - Online Survey since this is the most current data that we could use to understand Canadian's political behavior and attitudes. Canadian society and political life could be revealed from this dataset. I will be focusing on the relationship between level of education and household income.The data set includes 37822 observations which is enough to make accurate analysis.

Part b

```
## <labelled<double>[12]>: What is the highest level of education that you have completed?
##  [1] 10  8  5  4  9  6  7 11 12  3  2  1
##
## Labels:
##  value                                                                label
##      1                                                        No schooling
##      2                                             Some elementary school
##      3                                         Completed elementary school
##      4                                         Some secondary/ high school
##      5                                       Completed secondary/ high school
##      6      Some technical, community college, CEGEP, College Classique
##      7 Completed technical, community college, CEGEP, College Classique
##      8                                                  Some university
##      9                                                Bachelor's degree
##     10                                                  Master's degree
##     11                                       Professional degree or doctorate
##     12                                       Don't know/ Prefer not to answer
```



There are lots of outliers in every education level, and they are all right skewed. "12" are people who prefer not to say, beside those people, we could see a steady pattern of increasing in income with the growth of education level. For people have elementary school as their highest education level, there is an approximately $25000 upper limit. The median income for those who have completed college are close to the median income of people completed some university.

Part c

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
## # A tibble: 8 x 6
##   cps19_education1                minimum    mean median maximum     std
##   <chr>                             <dbl>   <dbl>  <dbl>   <dbl>   <dbl>
## 1 College                             700  77435.  68000  999000  61672.
## 2 Doctor                             1980 130823. 102000  942600 109308.
## 3 Don't know/Prefer not to answer    7000  79571.  61000  350000  76102.
## 4 Elementary School                  1100  50930.  45000  201453  34697.
## 5 High School                         600  61060.  50000  900000  56293.
## 6 Master                             1500 108502.  95000  985000  72870.
## 7 No schooling                      12000  74455.  71168  250000  60548.
## 8 University                         1000  94092.  81000  950000  67236.
```

Based on the summary of income, the mean and maximum are abnormally large, so I treat the ones that are greater than $100000 as outlier and filter them out. According to Fraser Institute News Release, Canadians who earn more than $96,000 are in the top 10%, so it's fair to set the upper bound of income as $1,000,000. Notice that the minimum wage is zero for all education levels, a $500 lower bound was set to drop meaningless data. Initially the educational levels were divided into 12 groups and labeled respectively. In order to reduce

the number of rows, I renamed each group and combined similar groups into one category.Surprisingly people who haven't attend school earn more than ones who went to high school. The median for people who have completed high school and the ones who had no schooling are $50000 and $71168 respectively. People who didn't attend post-secondary schools couldn't get more than $250000 income. People completed college could get a maximum of $999000 while the maximum income for a doctor is $942600.

Part d

We investigate on the data of household income before tax and level of education and analyze the data using "summurize" by different education level groups. It's commonly known that the more we learn, the more we earn. By doing the analysis on education levels and income, it's true that with a higher degree we could be offered with more salary. Gaining more in the future could be the motivation of learning hard during school. Studying isn't the only way to earn more and improve living standards, but it's a relatively conservative way with high payback.

Weakness of the analysis: There are other lurking variables that would affect household income. For example, partner's income, years of working experiences. Also, missing data was filtered out.

Next steps for investigating could be making a multivariate model and includes those lurking variables.

Part e

Bibliography:

1. Stephenson, Laura B; Harell, Allison; Rubenson, Daniel; Loewen, Peter John, 2020, '2019 Canadian Election Study - Online Survey',https://doi.org/10.7910/DVN/DUS88V, Harvard Dataverse, V1
2. R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.
3. Hadley Wickham, Romain François, Lionel Henry and Kirill Müller (2020). dplyr: A Grammar of Data Manipulation. R package version 1.0.2. https://CRAN.R-project.org/package=dplyr
4. H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.
5. Hadley Wickham, Jim Hester and Winston Chang (2020). devtools: Tools to Make Developing R Packages Easier. R package version 2.3.2. https://CRAN.R-project.org/package=devtools