How education and income effect population structure in Canada*

Benner Feng

April 30 2022

Abstract

With the help of the GSS Cycle 31 (@citegss) survey, This study showed the basic structure of the Canadian population aged 15-80. And the correlation between education, age, income and happiness index was found. This study clearly indicates that Canada is on a journey of population aging and that the demographic structure is highly likely to experience a shortage of young workers in the next fifteen to twenty years.

Introduction

Finding the silver lining in different families can be a daunting task. Every family has different dynamics and is shaped by various reasons/situations. However, as society is formed from countless families and individuals are the essential elements that make up families, it can be argued that the study of individuals can map the structure and future of society to a large extent. Therefore, I decided to follow Canadians at the individual rather than the family level. Many factors play a considerable role for individuals, such as education, marriage, income, and age. Using the data collected by the GSS Family Cycle in 2017, we can begin to track the dynamics of individuals within Canada. I can attempt to create a compelling narrative about Canada's social structure and trends over the next few years and even decades. This research will focus on gender, education, marriage, and income on both individual and societal levels. It will also reflect on the new generation of Canadians born and raised in Canada.

What was found is that families are related to each other in many ways from an abstract group level. However, the inner details of the individuals within the families differ significantly. The graphs and tables generated go into some eye-opening statistics on the families and the trends found among the families who chose to complete the survey. By capturing the new wave of Canadians making their mark on the country

We have the data from the GSS survey on these citizens, but there are no connections between the different columns. To remedy this problem, we had decided to generate a narrative of some of the determinants around fertility (Family Planning); these determinants include income, education, and sex. With this narrative spearheading our analysis of Canadian families, we can begin to provide the government with important information on the future of Canadians. The research done here can also help the people with information about the factors involved with people planning to see if it aligns with their pursuit of happiness.

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.6     v dplyr   1.0.8
## v tidyr   1.2.0     v stringr 1.4.0
## v readr   2.1.2     v forcats 0.5.1
```

*Code and data are available at: LINK.

income_respondent	N	%
$100,000 to $ 124,999	846	4.1
$125,000 and more	885	4.3
$25,000 to $49,999	6173	30.0
$50,000 to $74,999	3896	18.9
$75,000 to $99,999	2030	9.9
Less than $25,000	6772	32.9

```
## -- Conflicts ---------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()

##
## Attaching package: 'janitor'

## The following objects are masked from 'package:stats':
##
##     chisq.test, fisher.test

##
## Attaching package: 'lubridate'

## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union

"

##      sex            feelings_life
##  Length:9399        Min.   : 0.000
##  Class :character   1st Qu.: 7.000
##  Mode  :character   Median : 8.000
##                     Mean   : 8.058
##                     3rd Qu.: 9.000
##                     Max.   :10.000
##                     NA's   :136

##      sex            feelings_life
##  Length:11203       Min.   : 0.000
##  Class :character   1st Qu.: 7.000
##  Mode  :character   Median : 8.000
##                     Mean   : 8.125
##                     3rd Qu.: 9.000
##                     Max.   :10.000
##                     NA's   :135
```

Data

##2.1 Key Features The Family Cycle survey uses families as a bellwether to track the overall conditions of living within the country. Focusing on individuals begins to forecast many aspects of life in Canada(@citegss).

This survey was conducted in 2017, but the data was released in 2019. This survey data has 20,602 respondents and 81 variables. In this survey, we selected a few categories to focus on, such as the intersection of income, education, and fertility that are fundamental to Canadian life. The data here was prepared and shared using the statistical programming language R (@citer). In conjunction with the R to prepare our analysis we also used the 'tidyverse' as the main library for our resources (@citetidyverse).

##2.2 Sources and Methodologies The goal is to looking at different factors to see if there is a correlation between sex, income, and education in Canada. With the help of the GSS Cycle 31 (@citegss) survey, we can get a snapshot of the 2017 life cycle just to shed light on the most recent data on residents in the country. This data is also helpful in pinpointing some trends that the government might have overlooked.

all data is gathered to produce findings of the most recent insights on life cycle dynamics within Canada. From the survey website, we gathered this crucial information: "This survey was entirely voluntary, and the data is collected through the computer-assisted telephone interviewing method directly from the respondents"(@citegss). The overall response rate is 52.4% from the volunteers (@citegss).

As per data collection, the website mentioned a combination of randomization and sampling to get a fair and robust picture of the Canadian population(@citegss). This is important to point out that some regions have more population than others. The survey also looks at residents within Canada's ten provinces and limits the respondents aged at least 15 years and older.

This survey is collected every five years. The family cycle survey results we gathered from the survey at this time will also affect how we interpret the current and emerging families within Canada. The survey is collected every five years because there is little change in family dynamics over year increments, so five years is a good fit. It was also drawn to our attention that the following survey is set to occur in 2022.

The GSS family survey has played a role in the Canadian general social survey since 1984, but it was not until 1990 that the family cycle had its separate category. Although each cycle, the questions asked and the data collected change every so slightly, in the case of the most recent cycle, we are looking at a change that was made to make a note of was the combination of survey data and tax data from the respondents.

"In 2017, personal income questions were not asked as part of the survey. Income information was obtained instead through a linkage to tax data for respondents who did not object to this linkage. Linking to tax data diminishes respondent burden and increases data quality in terms of accuracy and response rates." (@citegss1) This would prove valuable because the survey is trying to move beyond personal income. After all, the self-reporting of income may differ from what is mentioned within tax data, and respondents may become afraid about their information being traced back to them if a false report was made.

Family Cycle Data (Cycle 31) was retrieved from Statistics Canada and initially cleaned by Rohan Alexander and Sam Caetano. We then proceeded to use this data and continue to cook with it.

##2.3 Strengths and Weaknesses The data collected has a solid understanding of the more complicated family dynamics like: "childcare services, childcare arrangements, child custody and financial support, and programs used after a separation or a divorce" (@citegss1). We felt this was a significant change as it also has lots to say about the current relationship trends with an uptake on divorce and common-law living (@citemarriagecan).

A tremendous strong point this survey edition highlighted while reading the many variables is the addition of many columns on immigrant and minority groups as a way to capture Canadian Families. It is best to see all types of families because Canada is a multicultural nation. It makes Canada unique as it tries to be more open with this shift toward multiculturalism.

As of the weaknesses or shortcomings this survey had was the feelings toward life, as we see the intention the government is trying to do here. However, the actual way the answers were provided for the respondents was a bit vague as it is pretty tough to quantify how unhappy someone can be. It is subjective, and everyone has their own criteria, so this information can only be used as a reference.

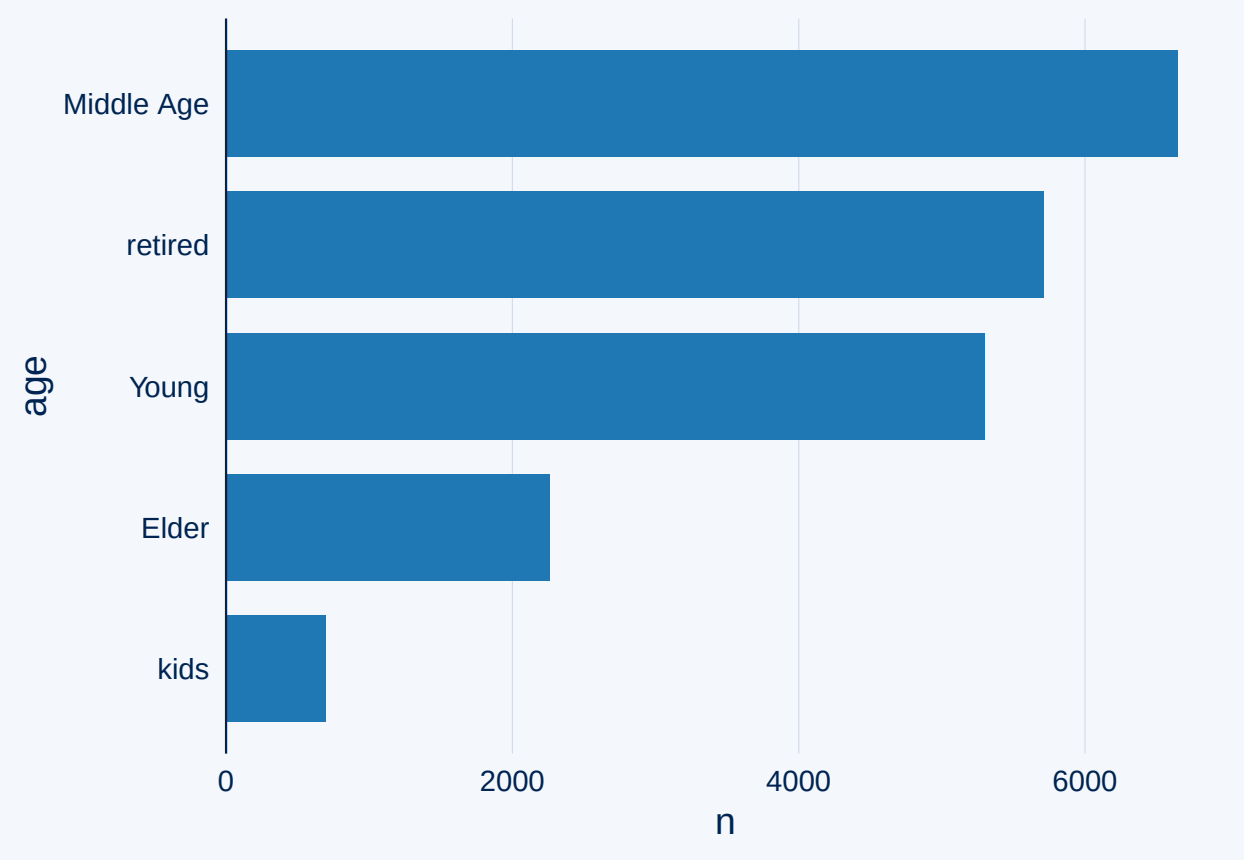
Another weakness picked up from this survey was the worked last week column. We did not understand what relevance knowing that answer would do to enhance this survey. As the responses from this survey were: "Yes," "No," and "Don't Know," the answers themselves were not very descriptive.

##2.4 Ethical Concerns Some of the ethical concerns noticed from the survey were excluding the three provinces. The data only focuses on the ten provinces, but this was a bit worrisome as there are families within the territories that are not being appropriately represented. Another ethical concern is the answers for "birth country" and "partner birth country"; the responses to these questions were either Canada or a country outside Canada. If the purpose of this survey is to capture the family makeup of Canada, would it not be beneficial to input the country you or your Significant other came from? As it could open the door into a more detailed look into families. Nevertheless, we recognize that this request may be met with some backlash as it is information that can be self-identifying, causing some privacy issues. A final ethical concern that needs to be addressed about this survey is the lack of representation for the institutionalized couples. Just because the significant other is institutionalized does not mean the person is not a part of the family. Even though the group of institutionalized individuals may be small, there is still a missing gap. However, the survey asks the respondents for self-rated health and self-rated mental health.

Results

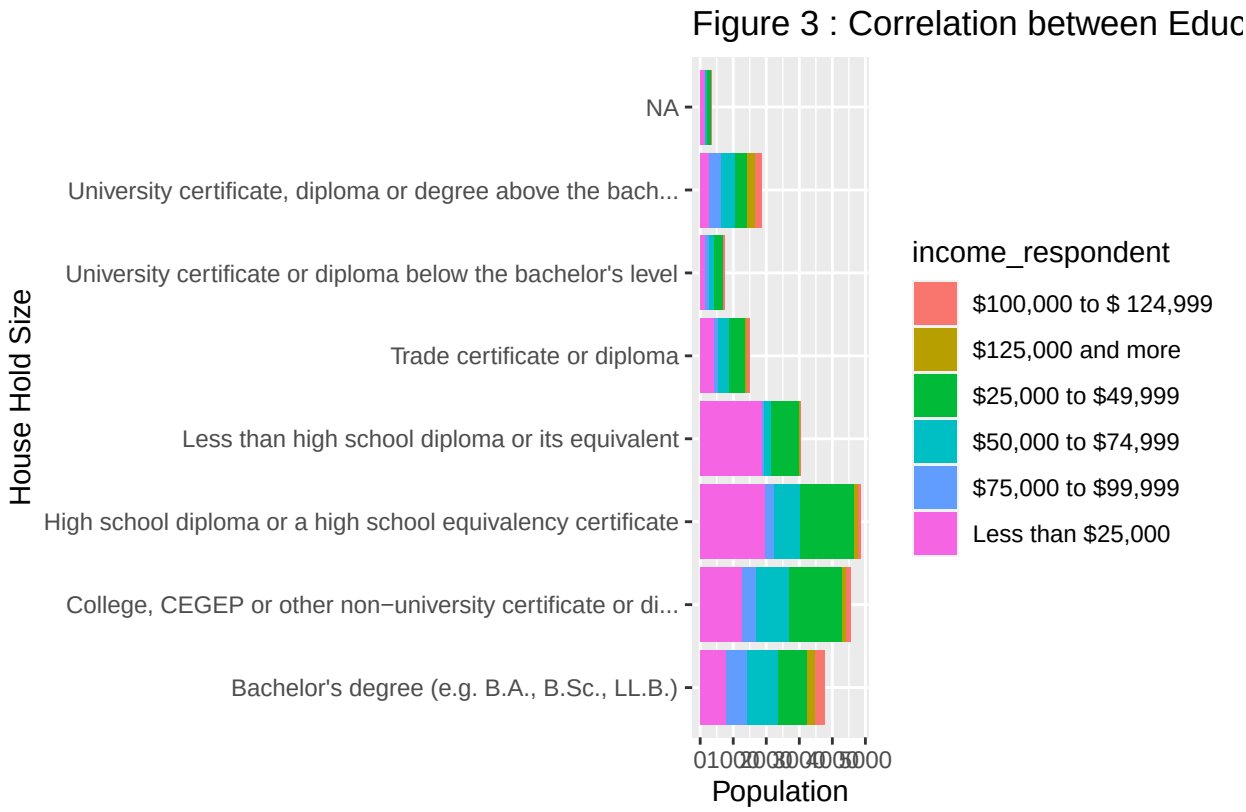
```
##          Number of Children Desire Average
## Ontario                         0.9305673
## Quebec                          1.0207160
## Prairie                         0.9363021
## Atlantic                        0.7421272
## BC                              0.8582996
```

In this table, we can see that only Quebec (1.021) has the highest number of children that people intend to have in the future, followed by Prairie (0.936) and Ontario (0.931). People in the Atlantic region want the least number of children in the future. As the number of desired children increases, it indicates the greater chance more than one kid is expected within the family. If the number of desired kids is less than one, it indicates less of a desire to have kids. With the most region's children desired rate indicates people's willingness to have children decreases in Canada.



According to Figure 2, Canada's population can be divided into five age groups: middle-aged, retired, young, older, and kids. Canada has a high percentage of seniors in the population. According to UN standards, a country or region is considered to have an aging population when 10% of the population is over 60 years old, or 7% of the population is over 65 years old. Middle-aged people are the largest segment of the population, and they are also currently the main contributors to the country's economy. However, as Figure 2 and table in Figure 1 suggested , people's willingness to have children is decreasing, it will leads to a shrinking share of the population of children and adolescents and an increasing share of the elderly population, resulting in a long-term decline in the birth rate. This type of population has a negative future trend of population reproduction and population shrinkage if the fertility level remains unchanged due to the low proportion of people of reproductive age.

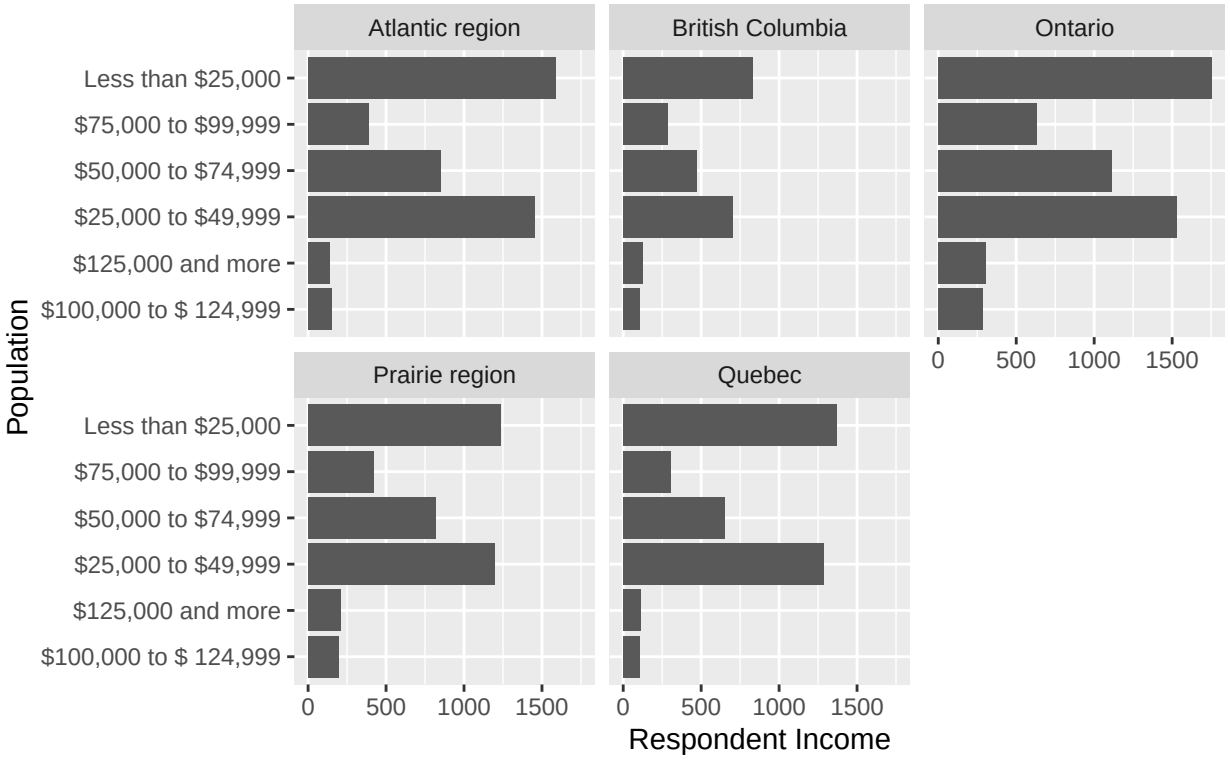
income_respondent	N	%
$100,000 to $ 124,999	846	4.1
$125,000 and more	885	4.3
$25,000 to $49,999	6173	30.0
$50,000 to $74,999	3896	18.9
$75,000 to $99,999	2030	9.9
Less than $25,000	6772	32.9



In Figure 3, I kept the NA values because NA can have different meanings in this situation, did not want it to effect the accuraccy of this data. This graph makes it clear that people who have a high school diploma or a high school equivalency certificate have the highest number of people. The number of people with a university certificate or diploma below the bachelor's level is the lowest. It is worth noting that there is not much difference in the number of people in the stratification of annual income levels for those with university and higher education levels.

As Figure 5 indicates, high level income population($125,000 and more) is 4.1% out of the total population, then followed by (100,000 to 124,999) which is 4.1% of total population, low income(Less than $25,000) have 32.9% of the population, which is the largest group. The income structure of Canadians is typically a pyramid structure, with the lower the income the greater the percentage of the total population.

Figure 6 : Respondent Income Level in Different Region Cana



data from: Canadian General Social Survey

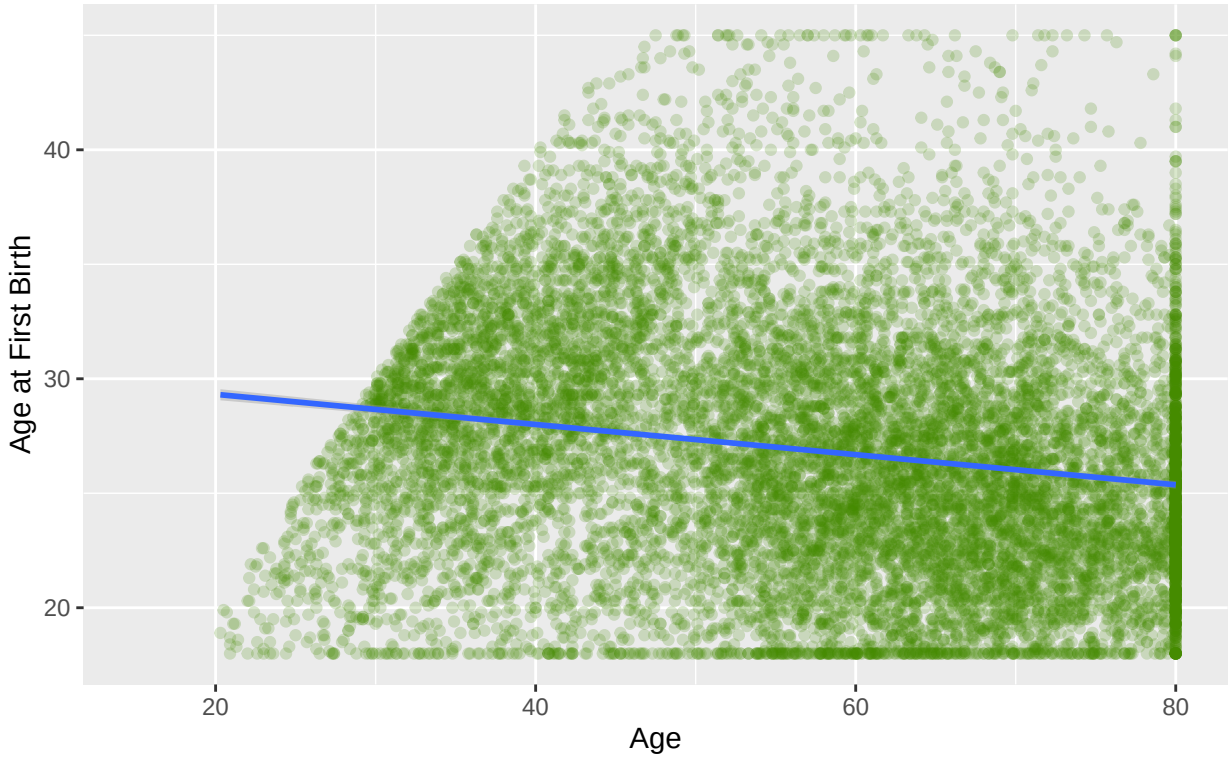
The graph shows that all five regions of Canada follow the same pattern. However, they have different numbers of participants in the survey: Ontario is undoubtedly the province with the highest participation, and a side note is that Ontario has a larger population.As for the data within the faceted graphs* (we should also mention we flipped the x and y-axis to fit the values properly as it makes the data easier to interpret), Most individuals asked in 2 of the 5 regions (Ontario, Prairies) had the highest number of residents in the highest personal income range ($125,000 and above). Meanwhile, the Ontario region had the highest number of individuals within the lowest income range ($25,000 to $49,000). Since all regions share the samiler income pattern.It clearly indicates that the income structure between Canadian regions is relatively small, and there is no abnormal gap between the rich and the poor.

```
## `geom_smooth()` using formula 'y ~ x'
```

```
## Warning: Removed 7865 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 7865 rows containing missing values (geom_point).
```

Figure 7 : correlation between age and age at first birth



data from: Canadian General Social Survey

In figure 7, people of different ages are shown at what age they had their first child. As the blue density line demonstrates, the average age at childbirth increases.

```
##      sex            feelings_life
##  Length:9399        Min.   : 0.000
##  Class :character   1st Qu.: 7.000
##  Mode  :character   Median : 8.000
##                     Mean   : 8.058
##                     3rd Qu.: 9.000
##                     Max.   :10.000
##                     NA's   :136
```

```
## Warning: Use of ‘gss$age‘ is discouraged. Use ‘age‘ instead.
```

```
## Warning: Use of ‘gss$feelings_life‘ is discouraged. Use ‘feelings_life‘ instead.
```

Figure 8 : correlation between age and Happiness

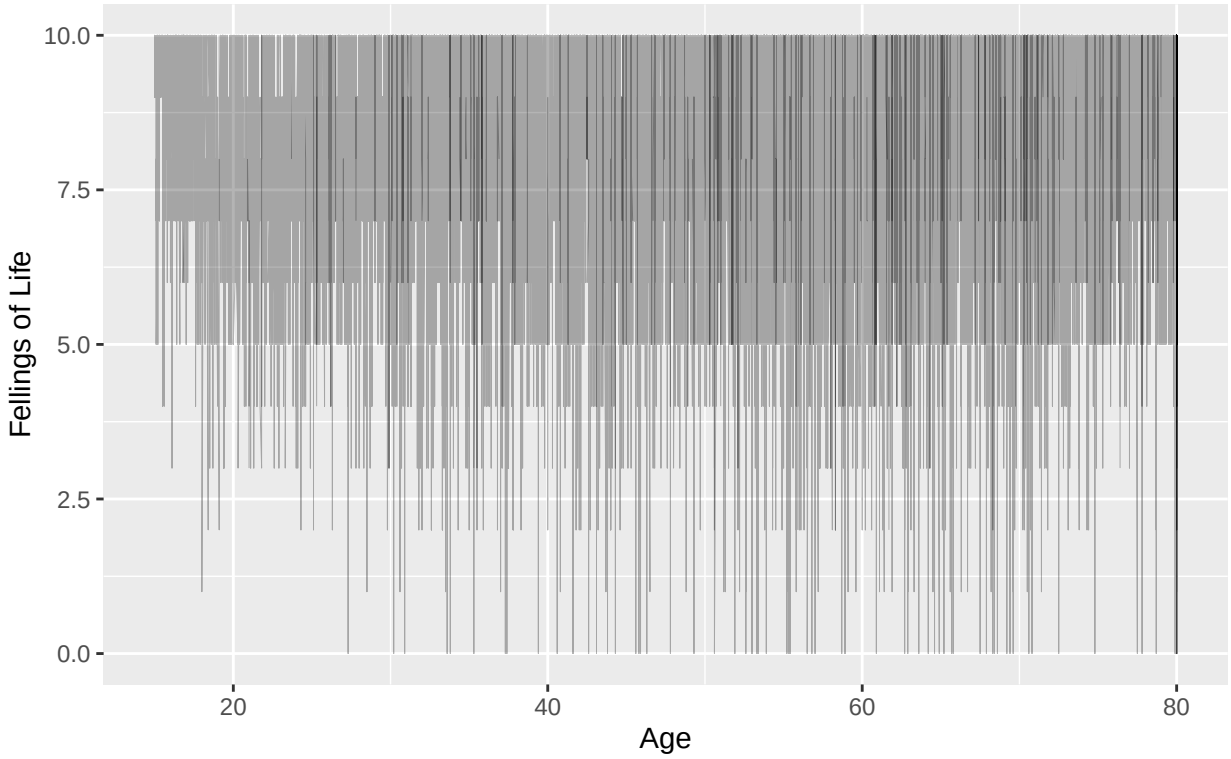


Figure 8 shows how people in different age groups rate their happiness. The people who rated the happiness index the lowest were the older people over 80, while those who rated the happiness index the highest were the young people aged 15-20. From figure8, we can clearly see that as people age, their level of happiness decreases. "'

Discussion

This study showed the basic structure of the Canadian population aged 15-80. And the correlation between education, age, income and happiness index was found. This study clearly indicates that Canada is on a journey of population aging and that the demographic structure is highly likely to experience a shortage of young workers in the next fifteen to twenty years. Together with Figure 7 and Figure 8, it shows that the older population is experiencing not only physical decline but also psychological depression. It is recommended that the government should improve the living conditions of the elderly population. The aging of the population implies a decrease in the share of the working-age population. As a result, first, it will cost a heavier burden of supporting costs on younger generations. The second will increase social security expenditures for the elderly. Third, a decrease in labor supply leads to a decrease in savings and tax revenue. Fourth increase in the fiscal burden on the government.

Population aging is an inevitable product of any socio-economic development to a particular stage, which brings negative factors to the sustainable economic development and brings specific opportunities to the development of the local economic market: The increase in the consumer market of the elderly can bring consumption-driven economic growth. It can be suggested that the government employ the elderly. Their long-term work experience is very beneficial to society, and some consultant positions can be opened in government agencies. Alternatively, charge a fee to help young people raise children and part of the household management work, to reduce the burden of young people and increase their efficiency. so Older people can make younger people more focused on their work.

It is bound to increase. The aging population puts forward higher requirements for social security undertakings and protection facilities, such as the pension insurance system, the elderly service system, and the construction of nursing homes, homes for the elderly, elderly apartments, and elderly service centers, elderly hospitals.

References

Dribe, Martin, J. David Hacker, and Francesco Scalone. 2014. "The Impact of Socio-Economic Status on Net Fertility During the Historical Fertility Decline: A Comparative Analysis of Canada, Iceland, Sweden, Norway, and the USA." Population Studies 68 (2): 135–49. https://doi.org/10.1080/00324728. 2014.889741. Government of Canada, Statistics Canada. 2019. "General Social Survey - Family (GSS)." Surveys and Statistical Programs. https://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&Id=335816#a1. ———. 2021. "Government of Canada, Statistics Canada_2021." Estimates of Population as of July 1st, by Marital Status or Legal Marital Status, Age and Sex. Government of Canada, Statistics Canada. https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1710006001&pickMembers%5B0%5D=1.1&pickMembers%5B1%5D=3.1&pickMembers%5B2%5D=4.1&cubeTimeFrame.startYear=2017&cubeTimeFrame.endYear=2021&referencePeriods=20170101%2C20210101. R Core Team. 2021. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/. Tripathi, Sabyasachi. 2020. "Does Economic Development Reduce Household Size? Evidence from India." Journal of Policy Modeling 42 (5): 982–99. https://doi.org/https://doi.org/10.1016/j.jpolmod.2020.04. 003. Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." Journal of Open Source Software 4 (43): 1686. https://doi.org/10.21105/joss.01686.