

# Examining Canadians' Feelings of Life in the Context of Children, Age and Income

Arshnoor Gill

Samantha Wong

Zhendong Zhang

2020/10/13

## Abstract

We wanted to investigate the social causes that could affect one's outlook on life, especially during such a tumultuous time. In this paper we investigated general feelings about life, on a scale from 1 to 10, that Canadians harboured in 2017, particularly as it relates to key variables of interest which socially and culturally are considered to impact people's outlooks—income, the number of children one has, and age. We found there is a definite positive relationship between income, age and number of children on one's general outlook on life (though there were more peaks in the age variable, suggesting individual stages in life could affect outlook) through graphical representations of scatterplots and the construction of a linear regression model that took all three variables into consideration with the explanatory variable being the general outlook on life. That being said, much of the data is created from manipulations on discrete information so it is not as accurate as raw continuous data would be, and it is possible that feelings of life could impact conversely how many children one has, for instance—it is ambiguous what is the explanatory and response variable.

## Introduction

Overall world happiness is on a decline, particularly in the US, Latin America, Asia and Sub-Saharan Africa(Newman, 2019). In addition to the global trend of negative thoughts, life satisfaction seems to be highly divided, evidenced in a widening happiness gap and life satisfaction inequality. Studies are showing a rise in negative feelings, including emotions derived from worry, sadness and anger, by 37 percent from 2010 to 2018. Happiness and enjoying life are often compromised by the course of life for many, which is often just accepted as a part of living. So how do certain countries maintain overall happiness and life satisfaction more than others? Adolescents living in the US are reportedly experiencing greater amounts of depression, suicidal thoughts and self-harm since 2010 (Twenge et. Al, 2015). In addition, income inequality has been shown to contribute to the growing happiness gap.

This paper analyzes Canadian data from the Canadian GSS from 2017 to crack down on the cause for inequalities in life satisfaction and outlook based on age, income and number of children. This study has found a positive relationship between income and feelings of life, although most respondents were reported to be within the lowest two income brackets. The number of children and feelings of life were positively correlated from having 0 to 2 children, whereas any more children showed an oscillating pattern but overall did not decrease. The relationship between age and feelings of life were initially high, with a minimum age of 15 years old. However, feelings of life sharply declined right and hit a minimum at around 23 and slowly inclined as age increased. Based on these conclusions, we recognize that there may be different existing factors to explore. Considering that individuals around 20 years old have the lowest reported feeling of life, it may be beneficial to study different aspects of 20 year-olds lives, such as education, occupation, geographic location and debt. To acknowledge the growing issue of happiness inequality, it also may be beneficial to explore information about individuals with the highest reported feelings of life with the lowest reported feelings and life.

## Data

The dataset was obtained from the Canadian General Social Survey (GSS) conducted on an annual basis by Statistics Canada for the purpose of monitoring changes in living conditions and well-being of individuals across Canada as well as providing information on social policy issues of interest during the year the study was conducted. Data from the 2017 iteration of the GSS (Cycle 31) was analyzed. The content of the survey focuses on different aspects over time - the 2017 GSS featured notable changes involving redesigning childcare services, childcare arrangements, child custody and financial support and programs after a separation or a divorce, expanding modules on parents and grandparents and removing module on work history. The targeted population in the survey was all non-institutionalized persons older than the age of 15 living in any of the 10 provinces of Canada. The survey design employed was stratified cross-sectional sampling. Stratification was conducted at the province/census metropolitan area level and the frames were constructed from telephone numbers.

In order to increase response-rates on income, personal income questions were not included in the 2017 GSS. Instead, income information was obtained through tax data for those who chose to respond. Based on given personal tax records, household information, social insurance number and other key variables, linkages were formed to obtain income information.

The variables of interest in this study:

- Age
- Income
- Number of children

The age distribution shown in figure 1.1 of the observations in the dataset are bimodal with two peaks centered at 35 years old and 65 years old. Most of the respondents in the study were around the ages of 35 and 65. The sample mean was 52.2 years and the median was 54.2 years. The range of ages obtained in the study are from 15 years to 80 years. About 2/3rds of the respondents had reported income levels between \$0 - \$50,000, which is shown in Figure 1.2. Median income level was 45,775.29 Canadian dollars where the median was at 37,500 Canadian dollars. The incomes ranged from 0 to 125,000, however, observations in the \$125,000 income bracket had an uncapped amount for their income. Around 1/3rd of individuals reported having 0 children and 1/3rd responded having 2 children. Most of the observations were drawn from individuals who were above the age of 20, where most people start making families and having children. The majority of individuals had equal to or less than 2 children. The age distribution shown in figure 1.1 of the observations in the dataset are bimodal with two peaks centered at 35 years old and 65 years old. Most of the respondents in the study were around the ages of 35 and 65. The sample mean was 52.2 years and the median was 54.2 years. The range of ages obtained in the study are from 15 years to 80 years. About 2/3rds of the respondents had reported income levels between \$0 - \$50,000, which is shown in Figure 1.2. Median income level was 45,775.29 Canadian dollars where the median was at 37,500 Canadian dollars. The incomes ranged from 0 to 125,000, however, observations in the \$125,000 income bracket had an uncapped amount for their income. Around 1/3rd of individuals reported having 0 children and 1/3rd responded having 2 children. Most of the observations were drawn from individuals who were above the age of 20, where most people start making families and having children. The majority of individuals had equal to or less than 2 children.

## Modelling

In terms of modelling, we've decided to use linear regression models in order to attempt to derive a relation between the general feelings of life the respondents of the survey feel, from 1-10, with the other data we've picked as possible explanatory variables: age, income and number of children.

The reasonings are as follows: in terms of age, there is often more uncertainty and fear regarding the future that comes when you are relatively younger, and less so when you are older and more established professionally and interpersonally. We wanted to see if people who are older tend to have a better general feeling of life as described by the survey. In terms of income, it follows that the more financially stable you feel, the greater you might feel towards life in general, and in terms of children, we wanted to know if more kids made life in some ways more fulfilling (or the opposite).

For our preliminary findings, we've decided to examine the distribution of each of the four variables both graphically and also numerically. This is through making histograms of all the variables, which will let us know the general trends the respondents feel towards their general feeling of life, their age, their number of children, and their financial status.

We chose to use a linear regression model given that the feelings of life variable, despite being discrete (in that you cannot pick a value between the integers), is still ordered, and that this way any relation between the two variables could be mapped out graphically and in terms of numerics. We used an alpha of 0.05 to determine if a result is statistically significant—thus, the p-value will test if the null hypothesis in each of these scenarios (that there is in fact no relation), which if it's under 0.05 will be verified.

That being said, in order to conduct this analysis, we had to make several decisions while modelling. First and foremost, we had to take out responses due to the issue of non-response, which decreases our overall sample size. Moreover, though age and number of children can be reported as is, the income was a categorical variable that needed to be redefined in order to fit the context of our analysis. For income, in categories “Less than \$25,000”, “\$25,000 to \$49,999”, “\$50,000 to \$74,999”, “\$75,000” to “99,999” and “\$100,000 to \$124,000”, we replaced each of those results with the midpoint between the ranges of each category. For instance, “\$25,000 to \$49,999” was replaced by 37,499.50. The final category, “\$125,000 and more”, was replaced simply by 125,000.

This choice of course means that the graph is not valid for income greater than \$125,000, but we decided that based on a certain assumption that after a certain level of money perhaps income does not increase quality of life, we decided it was the best way to still represent those respondents in that income bracket without making further assumptions about how much money they have.

## Results

### Exploratory Data

```
gss <- read_csv("./Output/gss.csv")
head(gss)

## # A tibble: 6 x 5
##   caseid    age income_respondent feelings_life total_children
##   <dbl> <dbl>           <dbl>          <dbl>           <dbl>
## 1     1  52.7        37500            8             1
## 2     2  51.1        12500           10             5
## 3     3  63.6        37500            8             5
## 4     4   80          62500           10             1
## 5     5   28          12500            8             0
## 6     6   63          12500            9             2
```

## Histograms of Variables of Interest

Figure1.1: Age Distribution

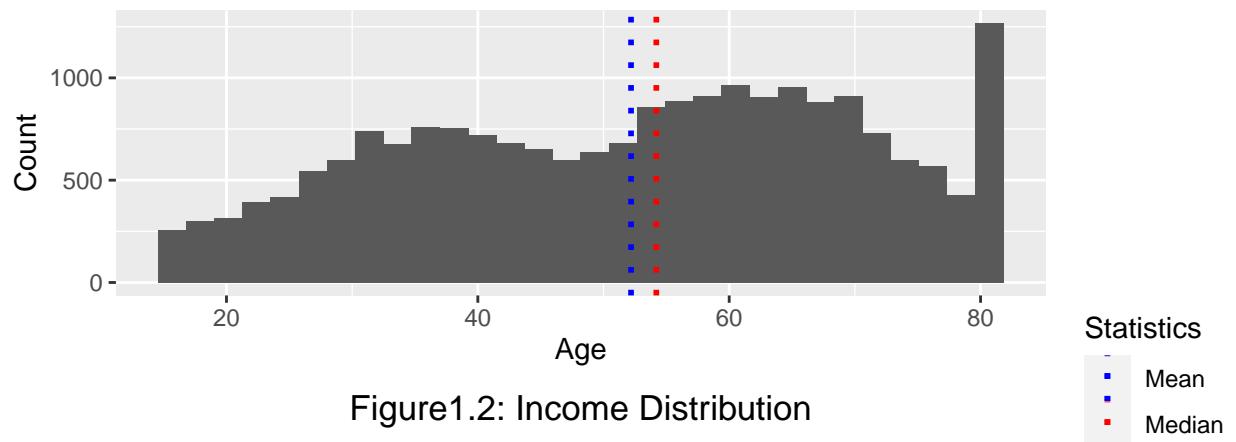


Figure1.2: Income Distribution

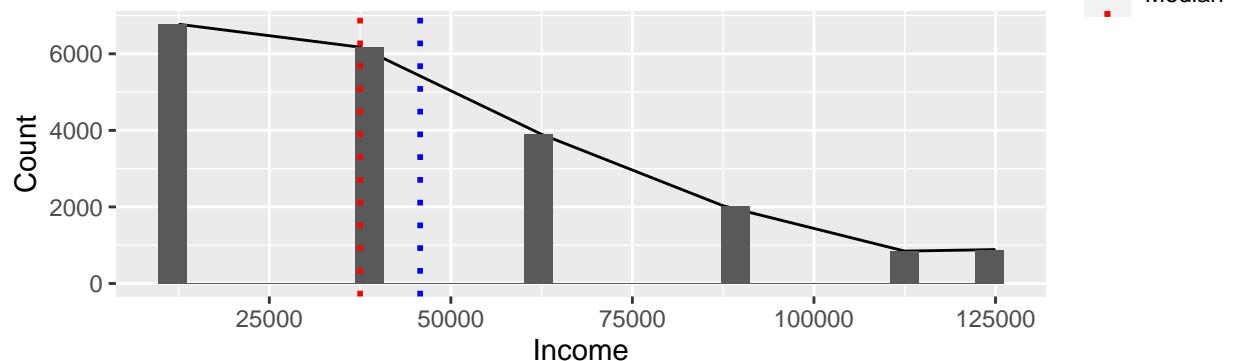
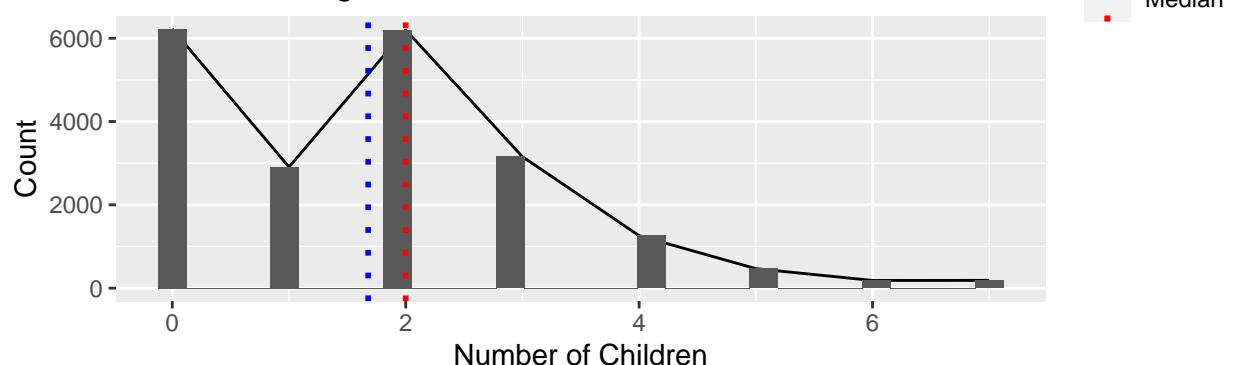


Figure1.3: Feelings of Life Distribution



Figure1.4: Total Number of Children



## Scatterplots

Figure1.5: Income vs. Feeling of Life

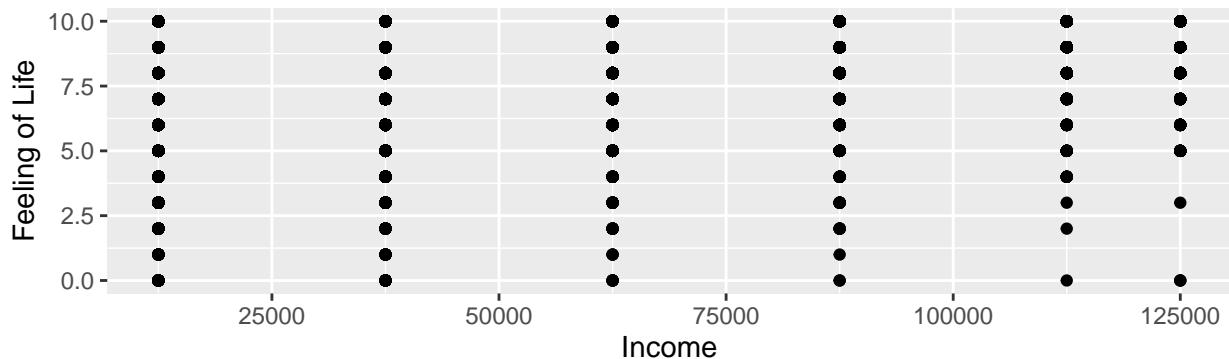
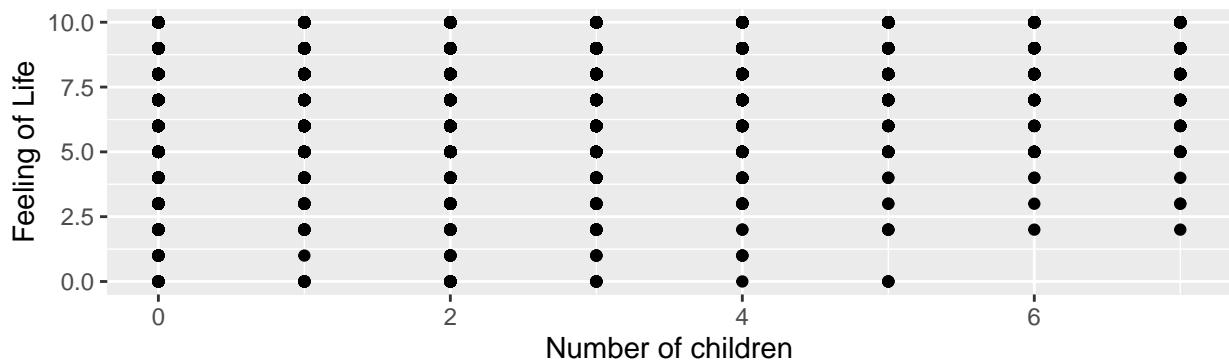


Figure1.6: Number of Children vs. Feeling of Life



## Further Data

### Mean versus Income/Number of Children

```

gss_income <- gss %>%
  group_by(income_respondent) %>%
  summarise(mean_feel = mean(feelings_life, na.rm=T), .groups = 'drop')

gss_child <- gss %>%
  group_by(total_children) %>%
  summarise(mean_feel = mean(feelings_life, na.rm=T), .groups = 'drop')

#Omitted NA values for calculating Means.

fig2.1 <- gss %>%
  ggplot(aes(y = feelings_life, x = age)) + geom_smooth() +
  labs(x = "Age", y = "Feeling of Life", title = "Figure2.1: Age vs. Feeling of Life") +
  theme(plot.title = element_text(hjust = 0.5))

#Age can be plotted without mean, because ages are more consistent. Other variables
#would produce error when plotting smooth line.

#So we plotted these two plot by using mean for each group, might be useful.

```

```

fig2.2 <- gss_income %>%
  ggplot(aes(y = mean_feel, x = income_respondent)) + geom_smooth(stat='identity') +
  labs(x = "Income", y = "Mean", title = "Figure2.2:\n Income vs. Feeling of Life") +
  theme(plot.title = element_text(hjust = 0.5))

fig2.3 <- gss_child %>%
  ggplot(aes(y = mean_feel, x = total_children)) + geom_smooth(stat='identity') +
  labs(x = "Number of Children", y = "Mean",
       title = "Figure2.3:\n # of Children vs. Feeling of Life") +
  theme(plot.title = element_text(hjust = 0.5))

grid.arrange(fig2.1, fig2.2, fig2.3, layout_matrix = rbind(c(1), c(2,3)))

```

Figure2.1: Age vs. Feeling of Life

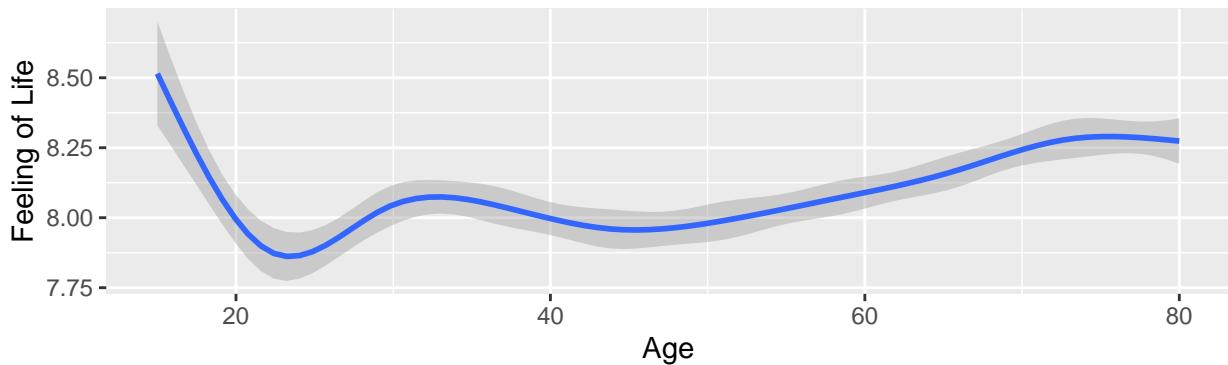


Figure2.2:  
Income vs. Feeling of Life

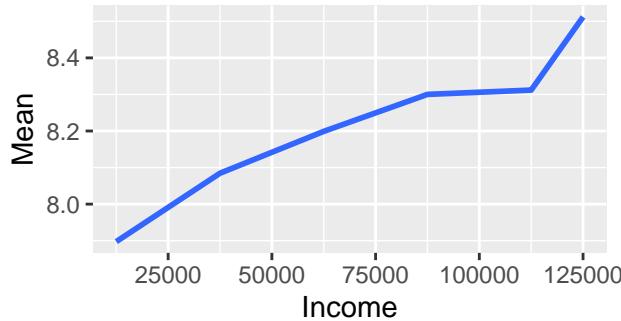
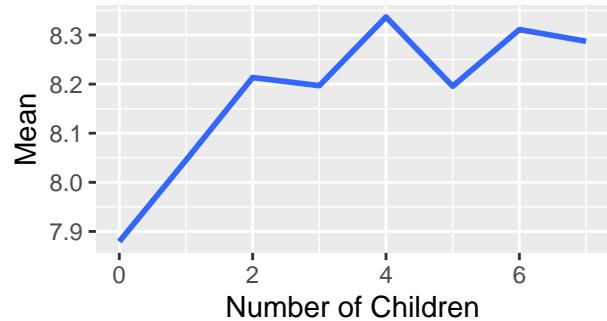
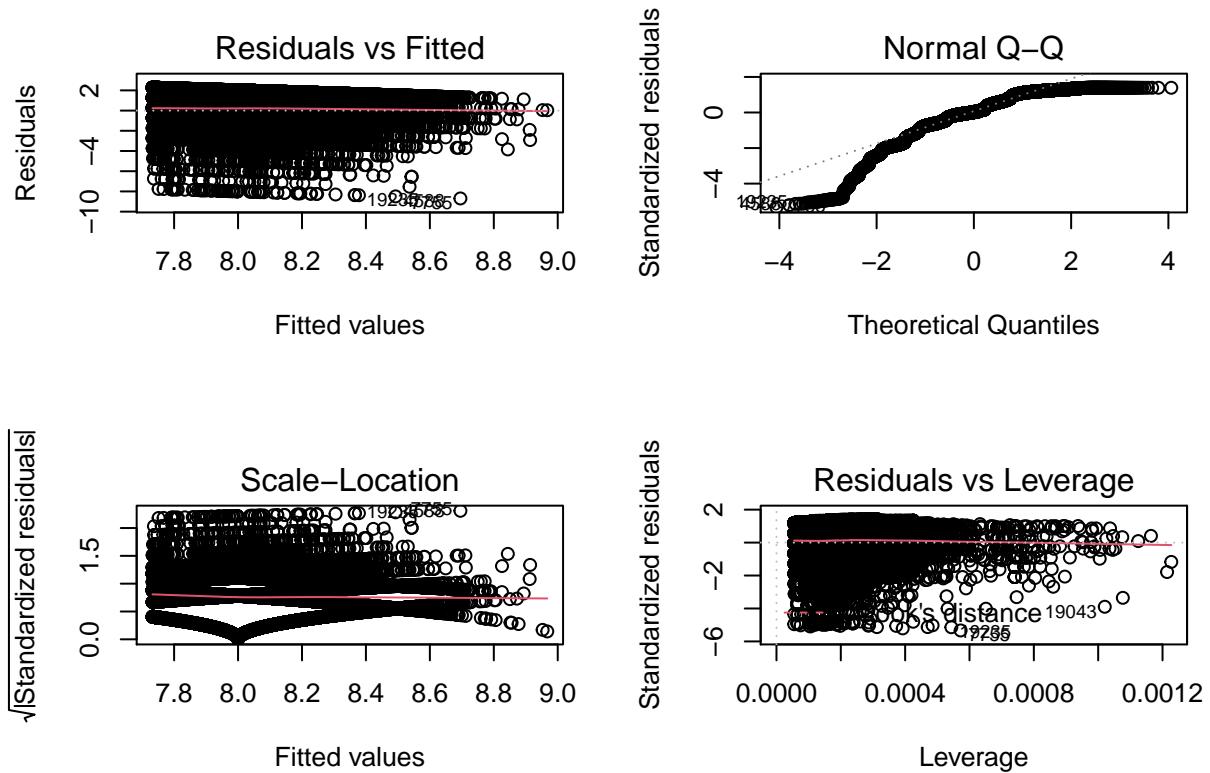


Figure2.3:  
# of Children vs. Feeling of Life



### Linear Regression Model

```
mod <- lm(feelings_life ~ age + income_respondent + total_children, gss)
```



## Discussion

### Histogram and Summary Table Analysis

Before we investigated the specifics in terms of causal relationships between variables, we wanted to example the spread of each of the variables that we decided to investigate, in order to get a better sense of the data we were working with.

Figure 1.1, the distribution of age, is bimodal, with peaks around the 30s and 60s. Outside this spread, there's a huge population of respondents who are in their 80s, strangely not following the general trend of the distribution. The mean and medians are 52.2 and 54.2 years of age respectively, and the data appears to be, in a very broad sense, left-skewed.

Figure 1.2 is the distribution of income, and as you can see the data appears to be like a bar graph despite being a histogram due to the fact that we created this variable by finding the midpoints of each income bracket in the GSS survey data. For this reason our analysis of this section will be pointedly less accurate than say, age, but we can notice from the histogram that income is broadly right-skewed, the great majority of people making less than \$50,000 a year. It was for this reason that we decided it would be okay to relabel the income bracket of people who made more than \$125,000 as \$125,000 because it does not apply to a huge part of our population and most of our analysis would be centred lower in that particular explanatory variable. The mean and medians are \$45,775.29 and \$37,500 respectively, reflecting that right skew.

Figure 1.3 is our chosen response variable, feelings of life. The mean and median are 8.1 and 8, and graphically there appears to a left skew to the data, meaning a great majority of the respondents feel generally good about life. This motivates the question of whether any of the previous variables motivate this distribution—if people with a higher income tend to have a better outlook than those who don't, whether children influence one's influence, and if getting older gives one a more charitable and understanding perspective on life.

Figure 1.4 is the distribution of number of children, and there is also a right skew to this data, reflected in the means and medians being 1.68 and 2 respectively. This does fit with the consensus that in more developed countries people tend to have less children (due to their being a less economical need for a larger family).

## Scatterplot Analysis

One of the first models used to visualize relationships between income and the number of children with feelings of life were scatterplot. However, the input for feelings of life were measured on a scale of 1-10. Since income data was collected through tax linkages, the income observations were categorized by brackets. Therefore each observation plotted were discrete variables. Figure 1.6: Number of Children vs. Feeling of Life show almost uniformly distributed dots, since many plotted points are on top of each other since the variables are discrete. This visually excludes many of the observation points from the data, therefore we decided to use other models for the relationship between number of children and feelings of life. We had attempted to use the mean of feelings of life to plot the scatterplots and found that there was a positive correlation between both income level and number of children on feelings of life.

## LM Model

The linear regression model we found was:

$$\text{Feelings of Life} = 7.647 + 1.559e^{-3} \times \text{age} + 4.987e^{-6} \times \text{income} + 8.223e^{-2} \times \text{number of children}$$

This shows that there is a positive relationship between all three variables, and the response variable of “feelings of life”. We must note that this is only really valid for incomes less than \$125,000 (because we don’t have specific information for income brackets above that one).

In the linear regression models for each variable shown in figures 2.1, 2.2, and 2.3, each variable had p-values below the alpha value of 0.05, providing strong evidence for a relationship between the variables investigated and feelings of life in Canada.

In order to test for normality in the data provided in the sample data, we used a normal QQ-Plot which plots the observations from the data against data found in a normal distribution. A straight line should form if the data roughly follows a normal distribution. In the Normal QQ-Plot, lower values show more variance as the values fall below the dotted line.

## Weaknesses and Next Steps

In general, income information collected through surveys usually receive low response rates to the sensitivity of the subject of income. In the 2017 iteration of the Canadian GSS, income information was collected through linkages from tax survey questions, social insurance information and other key variables to estimate which income bracket a respondent belonged to. Considering that income data was organized into brackets, individuals are placed into categories, rather than having income as a discrete variable. In addition, Canada has been ranked 9th on the world’s top happiest countries (World Happiness Report, 2019). The median feelings of life of the sample was 8 while the mean was 8.1 out of a scale of ten. Out of the observations in the sample, it appears that most of the respondents were on the “happier” side and had greater feelings of life. Therefore, it may be helpful to compare Canada’s results to other countries who are lower on the happiness scale to find more about what causes greater gaps of satisfaction and happiness.

There is also the possibility of many lurking variables—for instance, it has been shown through various studies, such as one by G. Nargund, that more developed countries, which tend to have higher incomes for their populations, tend to have lower children for causes having to do with better access to contraceptives, more education and lifestyles associated with greater economic standing. Thus, it may not be the number of children as an explanatory variable for feelings of happiness, and it may be that income affects both the number of children as well as feelings of happiness.

Supporting code can be found at: <https://github.com/Craymate/GSS>.

## References

- Baptiste Auguie (2017). gridExtra: Miscellaneous Functions for “Grid” Graphics. R package version 2.3. <https://CRAN.R-project.org/package=gridExtra>
- [https://github.com/Craymate/GSS/blob/main/Script/gss\\_cleaning-1.R](https://github.com/Craymate/GSS/blob/main/Script/gss_cleaning-1.R)
- Government of Canada, Statistics Canada. “How’s Life in the City? Life Satisfaction Across Census Metropolitan Areas and Economic Regions in Canada.” Government of Canada, URL <https://www150.statcan.gc.ca/n1/pub/11-626-x/11-626-x2015046-eng.html>
- JJ Allaire and Yihui Xie and Jonathan McPherson and Javier Luraschi and Kevin Ushey and Aron Atkins and Hadley Wickham and Joe Cheng and Winston Chang and Richard Iannone (2020). rmarkdown: Dynamic Documents for R. R package version 2.3. URL <https://rmarkdown.rstudio.com>.
- Luscombe, Belinda. “Do We Need \$75,000 a Year to Be Happy?” Time, Time Inc., 6 Sept. 2010, [content.time.com/time/magazine/article/0,9171,2019628,00.html](http://content.time.com/time/magazine/article/0,9171,2019628,00.html).
- Nargund, G. “Declining birth rate in Developed Countries: A radical policy re-think is required.” Facts, views & vision in ObGyn vol. 1,3 (2009): 191-3.
- R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Statistics Canada, 27 Nov. 2015, [www150.statcan.gc.ca/n1/pub/11-626-x/11-626-x2015046-eng.htm](https://www150.statcan.gc.ca/n1/pub/11-626-x/11-626-x2015046-eng.htm).
- Sachs, Jeffrey D., et al. “World Happiness Report Finds That People Are Feeling Worse.” Greater Good, [greatergood.berkeley.edu/article/item/world\\_happiness\\_report\\_finds\\_that\\_people\\_are\\_feeling\\_worse](http://greatergood.berkeley.edu/article/item/world_happiness_report_finds_that_people_are_feeling_worse).
- Sam Firke (2020). janitor: Simple Tools for Examining and Cleaning Dirty Data. R package version 2.0.1. <https://CRAN.R-project.org/package=janitor>
- Sifferlin, Alexandra. “Here’s How Happy Americans Are Right Now: Harris Poll.” Time, Time, 26 July 2017, [time.com/4871720/how-happy-are-americans/](http://time.com/4871720/how-happy-are-americans/).
- Statistics Canada. “Canadian General Social Survey - Family (GSS) 2017.” Statistics Canada, Sept. 2019, <https://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&Id=335816>.
- Twenge, Jean M., et al. “More Happiness for Young People and Less for Mature Adults: Time Period Differences in Subjective Well-Being in the United States, 1972–2014.” Social Psychological and Personality Science, vol. 7, no. 2, Mar. 2016, pp. 131–141, doi:10.1177/1948550615602933.<https://journals.sagepub.com/doi/10.1177/1948550615602933>
- Wickham et al., (2019). Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686, <https://doi.org/10.21105/joss.01686>
- Yihui Xie and J.J. Allaire and Garrett Grolemund (2018). R Markdown: The Definitive Guide. Chapman and Hall/CRC. ISBN 9781138359338. URL <https://bookdown.org/yihui/rmarkdown>.

## Appendix

### Summary Tables

Table 1: Summary of Variables

	<b>Age</b>	<b>Income</b>	<b>Feelings of Life</b>	<b>Number of Children</b>
<i>Mean</i>	52.2	45775.29	8.1	1.68
<i>Median</i>	54.2	37500	8	2
<i>Minimum</i>	15	12500	0	0
<i>Maximum</i>	80	125000	10	7

Table 2: Regression Statistics

	<b>Df</b>	<b>Sum Sq</b>	<b>Mean Sq</b>	<b>F value</b>	<b>Pr(&gt;F)</b>
<i>age</i>	1	141.38	141.38	53.16	3.18214369163231e-13
<i>income_respondent</i>	1	552.4	552.4	207.72	7.37360433804006e-47
<i>total_children</i>	1	243.11	243.11	91.42	1.2935341689774e-21
<i>Residuals</i>	20308	54007.19	2.66	NA	NA

	<b>Value of Regression Residuals</b>
<i>Min</i>	-8.6956
<i>1st Quantile</i>	-0.879
<i>Median</i>	0.0333
<i>3rd Quantile</i>	1.1291
<i>Max</i>	2.2677