Age, Marital Status and Related Factors Cannot Explain the Total Number of Children in a Canadian Household*

Jennifer Liu Zhiyi Liu Yiying Chen

20 March 2022

Abstract

Not only is the fertility rate essential to maintain the current population, the number of children in each household also plays an important role. The General Social Survey was conducted by Statistics Canada to collect information on the population. The paper examines potential factors that explain the total number of children per person. We found that respondents' age, age of having a first child, marital status and age of first marriage have little effect on the total number of children per person. The correlation between the age of first birth and the total number of children is weaker among families with fewer children. Understanding the relationships among those variables can help us predict trends and implement policies.

Contents

*Code and data are available at: https://github.com/jenniferliu888/GSS. Survey available at: https://docs.google.com/forms/d/e/1FAIpQLSei8IzSuYwBGC7typll1ohPQQjnE7O0hAtEIile9-sZSLfnEA/formResponse

Introduction

Population growth has always considered as a continuing problem, the Canadian government focused on the population for each family by conducting the General Social Survey (GSS) annually. Various aspects could be extended from this exhaustive database; in this paper, we mainly investigate the factors that could affect the total number of children per household in Canada.

The General Social Survey on the Family (2017) utilizes cross-sectional design, collecting background information on various aspects and including population from all age groups in 10 provinces (*General Social Survey* 2020). It was conducted by Statistics Canada and held through phone interviews; residences in selected regions will be called and instructed to answer questions in surveys. The information was aiming to monitor the general well-being of Canadians and be able to construct a foundation for emerging social issues. Specifically, by estimation, in 2031, there will be close to one to four Canadian citizens who are 65 years or older. As population aging becomes a rising issue for Canadians' society, analyzing the factors behind the trend of total number of children is vital for social study.

This paper was a reproduction based on results of the General Social Survey and emphasized on factors of population growth. Firstly, the original dataset from GSS (2017) was categorized and cleaned, then the select variables are visualized. Then, the methodology of how Statistics Canada was used to collect data are evaluated. Thirdly, the analysis focused on the total number of children per household was conducted. According to our selected variables that could affect the total number of children, we emphasize those variables to investigate their correlation. Exploratory data analysis (EDA), ggplot with categorical data, hypothesis, linear regression was used for further exploration.

These computations and models indicate that age of first child and age at first marriage could positively affect the total number of children in a household to a small extent; inversely, the value of age group for the respondents leads to a negative effect. As their age rises, especially after 30 years old, they are reluctant to keep more than three kids. The lack of significance indicates drawbacks of the survey design; at the end of this paper, a supplementary questionnaire was attached.

Data

This report is analyzed using R (R Core Team 2020), using tidyverse (Wickham et al. 2019) and dplyr packages (Wickham et al. 2020). All the tables and graphs are created using ggplot2 (Wickham 2016) and the file is knitted using knitr (François, Henry, and Müller 2021).

Data Source

The 2017 Canadian General Social Survey (GSS) microdata is retrieved from the University of Toronto Database CHASS (Computing in the Humanities and Social Sciences) and is available to over 25 subscribing universities (*Computing in the Humanities and Social Sciences* 2019). CHASS is a computing facility that aims to promote computing in research.

GSS was established in 1985 and aims to gather data on social trends to compare changes in living conditions and Canadians' well-being. It also provides additional information on social policy issues and emerging trends. The 2017 GSS focuses on family data as it plays an important role in people's lives. Some of the questions it answers include the social-economic conditions, the diversity and the structure of Canadian families. Many changes were made across GSS surveys over the year, which makes it difficult to compare. For instance, in the 2017 survey, income is no longer asked and was rather collected from tax data.

Collection

The 2017 GSS was conducted from February 2nd to November 30th 2017 as a sample survey. The target population is all non-institutionalized people of 15 years old and older, living across the 10 provinces in Canada. Therefore, it excludes residents of the 3 Canadian territories. The data collection was done over the phone, with numbers provided by Statistics Canada's Address Register.

The sampling frame was created using the lists of telephone numbers in use from Statistics Canada and the Address Register. Over 86% of phone numbers were linked to an address, the rest were also included in the frame. When there is more than one phone number linked to an address, the first phone number is considered the best to reach and they will only be put in the poll once. All phone numbers associated with businesses or institutions were removed from the pool.

In order to sample the population, the 10 provinces were separated into strata with Census Metropolitan Areas (CMAs) considered to be separate strata. This was the case for St. John's, Halifax, Saint John, Montreal, Quebec City, Toronto, Ottawa, Hamilton, Winnipeg, Regina, Saskatoon, Calgary, Edmonton and Vancouver. Three more strata are added by regrouping CMAs that are not in the list above and the non-CM areas were then grouped to form 10 more strata. A total of 27 strata were made. A total of 20,626 samples were surveyed.

Data was collected using computer assisted telephone interviews (CATI) with trained interviewers in the official language of their choice. Interviews were done with randomly selected members of the household and those who at-first refused to participate were re-contacted. In the event that no one was at home, numerous calls were made. The response rate was 52.4%.

All survey answers were recorded directly by the computer as the interview progressed. The CATI system also identifies "out-of-range" values, in which the interviewer is mandated to enter it manually and resolve the issue. In 2017, personal income questions were not asked as it was provided through a linkage with tax data.

Data Overview and Cleaning

In terms of the data cleaning process, we have first changed all the variable names, so that it is easier to read and align with the actual data. We have then changed some strings into numbers for variables such as the number of total children so that it is easier to proceed with the data analysis.

Some of the variables are: total_children: total number of children age_at_first_birth: age at the birth of the first child regilion_importance: importance of the religion feelings_life: rated score of feelings about life province: current living province

Exploratory Data Analysis

Table 1: Summary of Respondents' Age

min	Q1	median	Q3	max	mean	standarddeviation
20.3	44.4	58.9	69.8	80	57.29376	15.2113

Table 2: Summary of Respondents' Age at First Birth

min	Q1	median	Q3	max	mean	standarddeviation
18	22.8	26.4	30.3	45	26.86411	5.417533

Table 3: Summary of Respondents'Total Number of Children

min	Q1	median	Q3	max	mean	standarddeviation
1	2	2	3	7	2.36919	1.158505

Table 1, 2 and 3 are summary data for respondents' age, age at first birth and total number of children. In order to only look at respondents who have children, those who do not have children are filtered from this summary data. It is interesting to notice that there is a data spread in the number of children per household, the first quantile is at 1 and the third quantile is at 3, making it interesting to examine possible factors that might contribute to the spread. Similarly, respondents' age at first birth also has a wide spread with a standard deviation of 5.42.

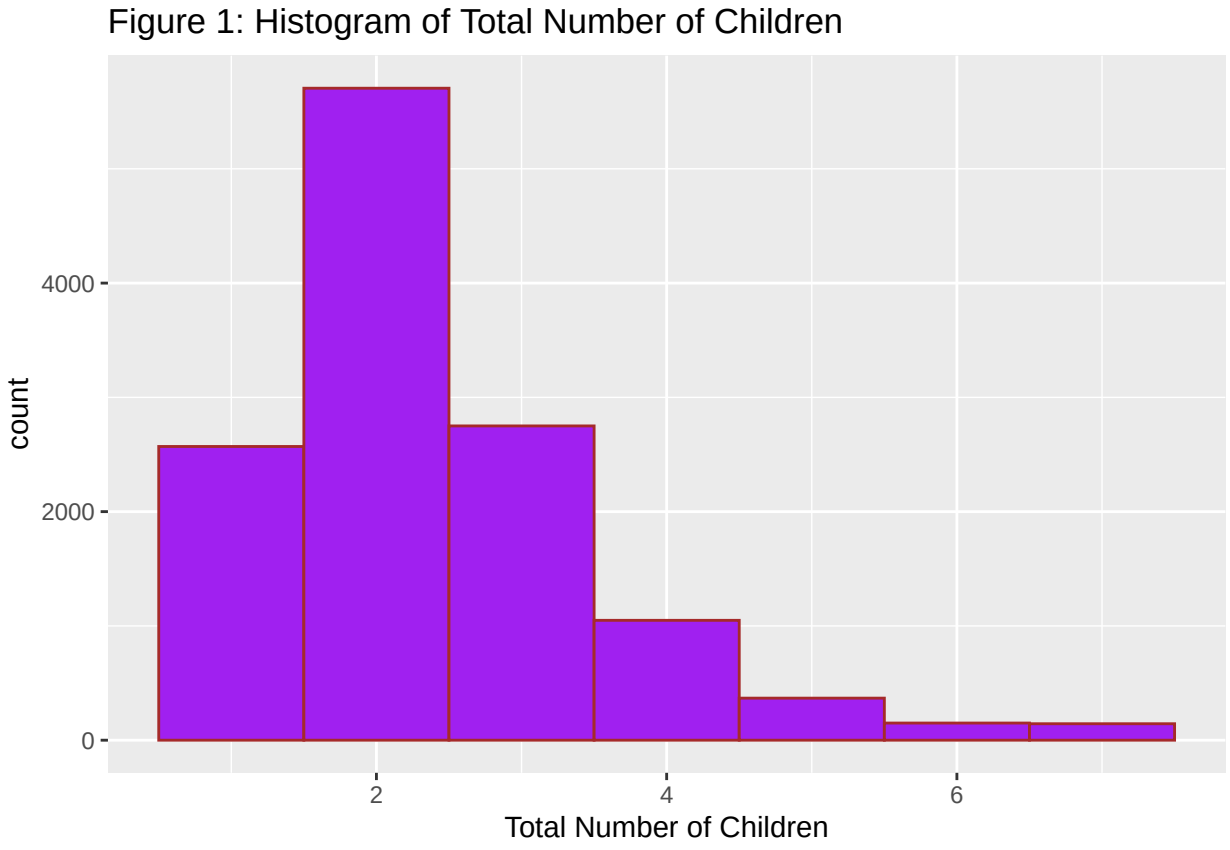


Figure 1 is a right skewed-histogram for a household's total number of children with most data spread between 1 and 3. There is a very significant mode at 2, followed by households with 1 and 3 children. Families with 4 and more children are less and less frequent to see.

Figure 2: Total Number of Children by Age at First Birth

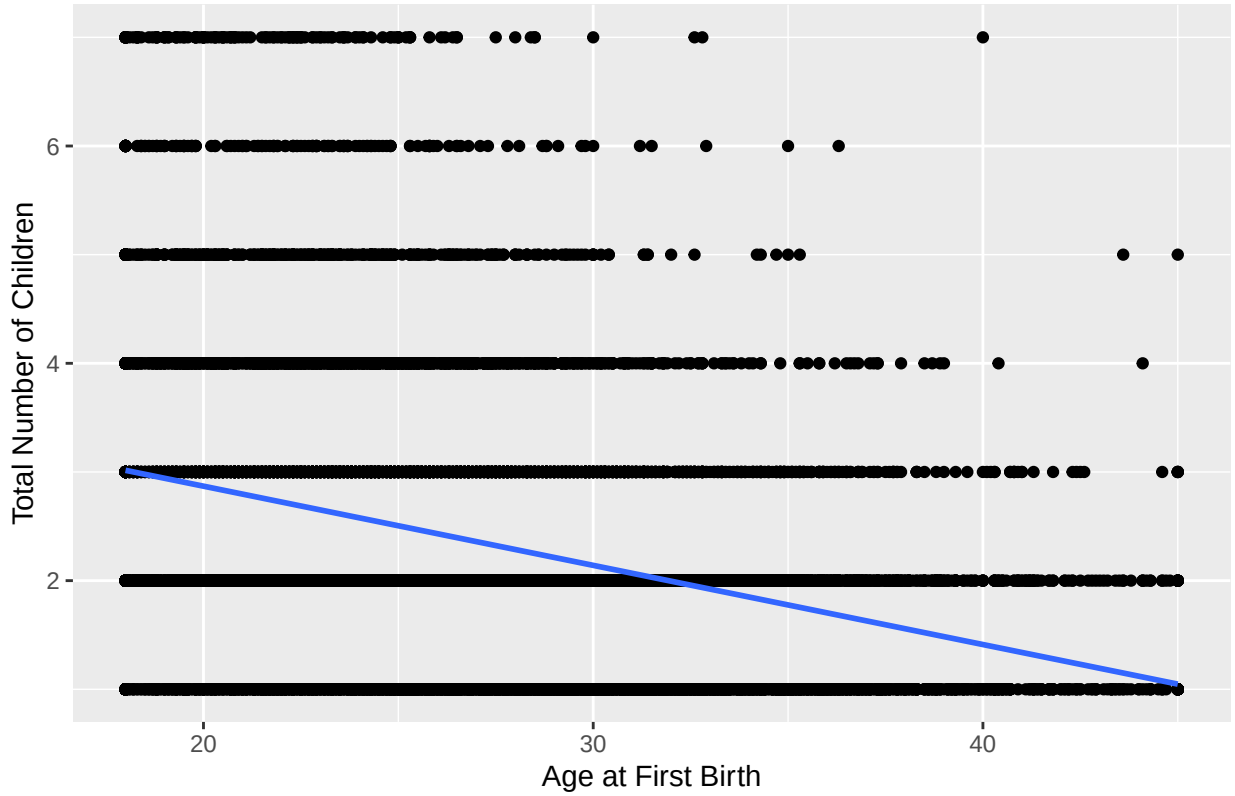
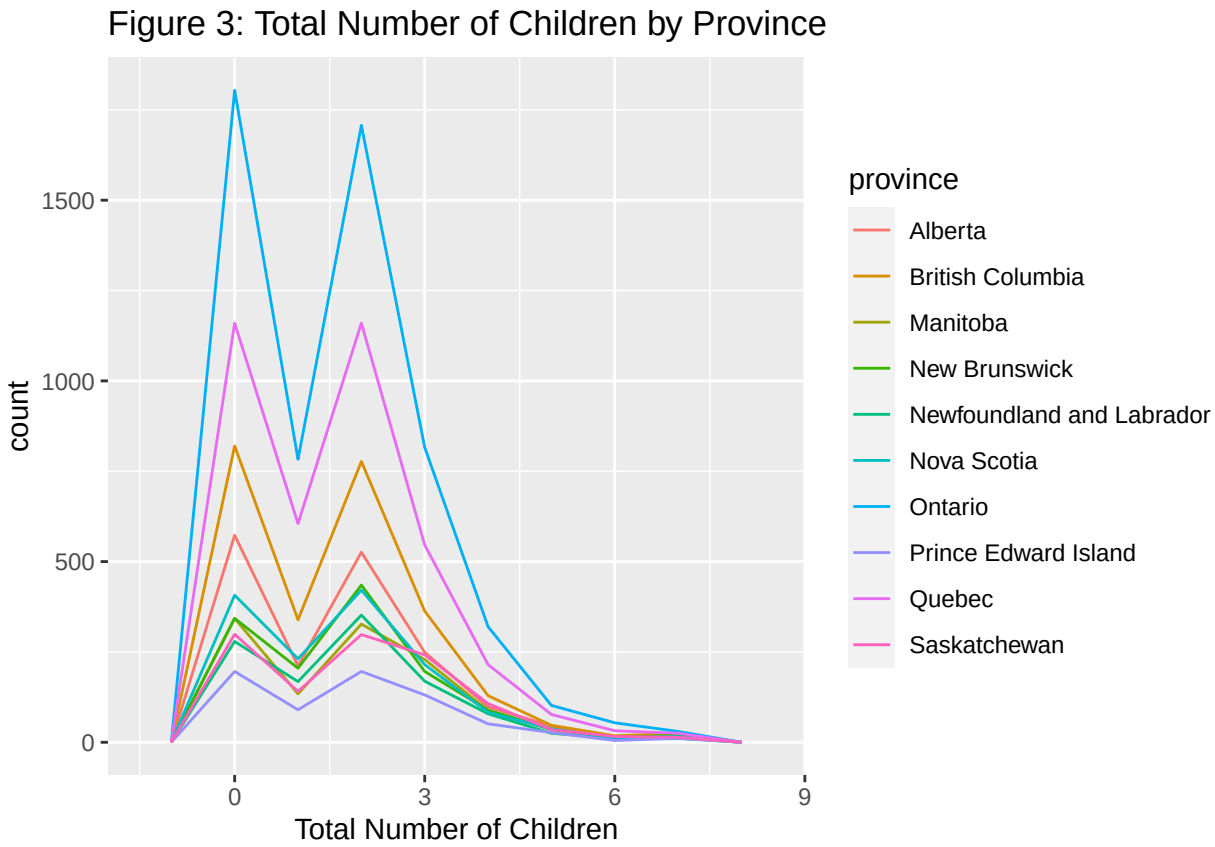
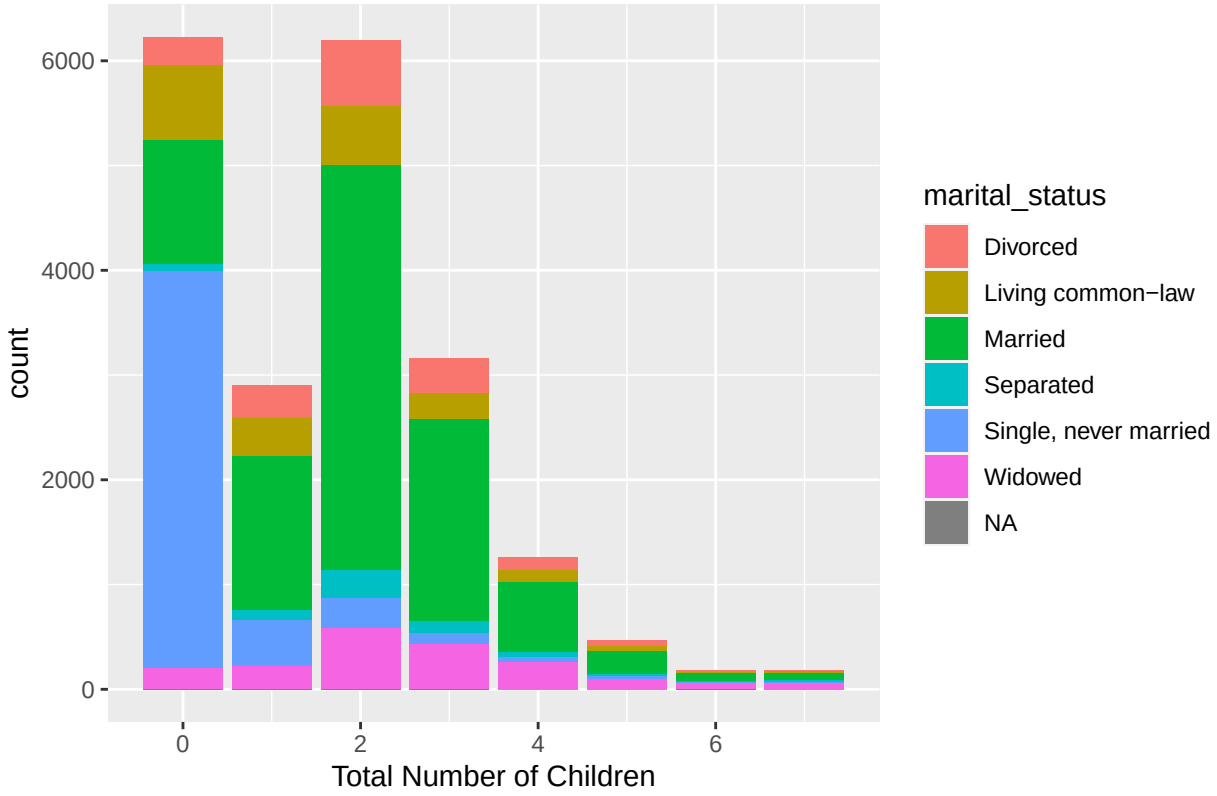


Figure 2 shows a relationship between the total number of children and age at the first children's birth. There is a negative relationship between both, as the age at first birth increases, the total number of children in a family decreases. Although there are minor outliers, the relationship is positive and is illustrated by the blue line. This result entails that most families with a big number of children (above 3) had their first children at a younger age. On the other hand, families who have a smaller number of children (below 3), had their first children across all ages.



In order to analyze the total number of children distributed across provinces, a frequency polygon is created with multiple lines of different colors representing the provinces. The lines are layered one on top of the other because some provinces have a higher number of residents such as Ontario and Quebec, which explains the difference in count. However, all of them have a similar distribution with peaks at 0 and 1.2. This explains that the province where each family is located does not affect their number of children.

Figure 4: Total Number of Children by Marital Status



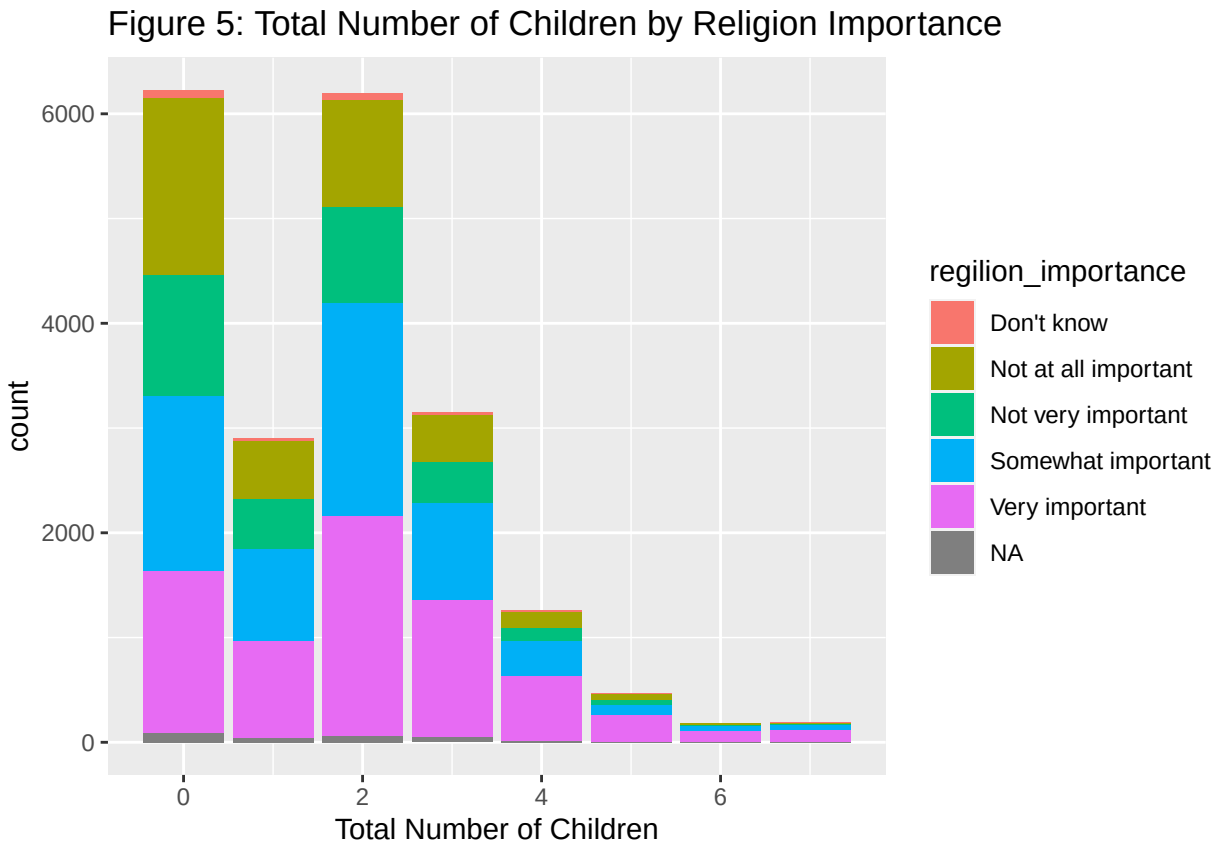


Figure 4 and 5 show histogram of the total number of children by different marital status and religious importance. It is interesting to notice that comparing households who have 0 and 2 children, there is a higher number of living common-law compared to those who are divorced. Most families who have 2 children are married with a significant number of widows. Moreover, the proportion of those who are single compared to separated is much higher for households with 1 child.

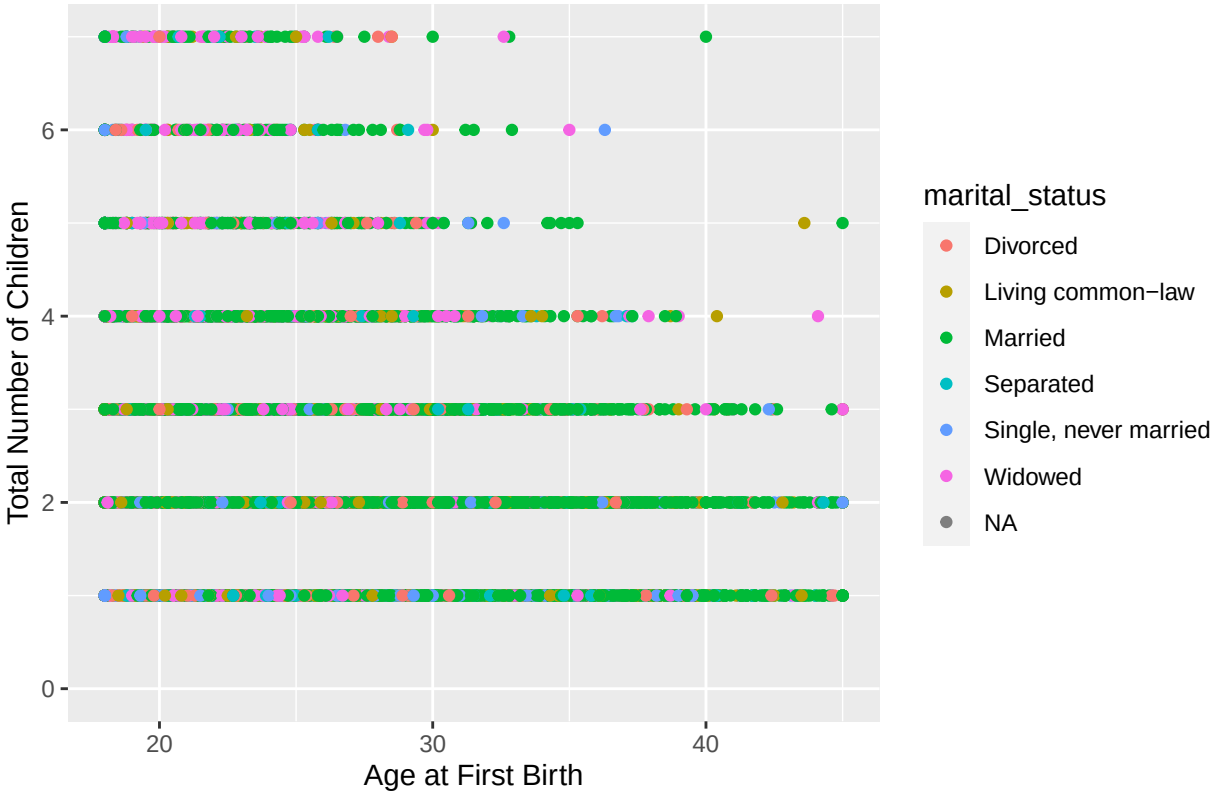
In terms of religion importance, when comparing families with 0 and 2 children, people who believe that religion is not important are higher compared to those who give a high importance in families who do not have any child. For families with 4 children or more, the proportion of those who give a high importance to religion is higher.

Result

By the result of the EDA procedure, we investigate the factors that affect the total number of children per household by different variables, such as age of first birth, marital status and religious importance. This section will draw conclusions between variables and present it by tests and graphs.

1. Scatterplot Analysis with Marital Status

Figure 6: Total Number of Children by Age at First Birth and Marital Status



Firstly, the scatterplot explored the relationship between three variables: total number of children, age of first child and marital status. The last factor was a categorical variable, we present it by the color of the scatterplot. As shown in the graph, when the age at the first child becomes larger, they tend to have a smaller number of total children, especially for people older than 30 years old. For marital status, in Canada, there are divorced, living common-law, married, separated, single (never married) and widowed. As the experimental units tend to become younger when they have their first child, their marital status varies a lot: widowed and divorced occupy a large proportion of the outcome. Conversely, when the age of first birth becomes larger, the majority of the sample are all married.

2. Hypothesis Test

t	df	p-value	95 % confidence interval	mean of x
1.8117	20582	0.03502	1.661732	1.678813

Secondly, the utilization of one sample hypothesis t-test within 95% confidence interval was necessary. In this test, h0: μ=1.6 v.s. H1: μ>1.6. We got a very large sample size in this study; hence it satisfied the assumption of normality. By the research of the Organization for Economic Co-operation and Development (OECD), the average number of children per woman in Canada was 1.6 (Russell 2017). The result suggests that there is strong evidence to support the alternative hypothesis that the average number of children per household in 2017 is more than 1.6. It was strongly against OECD's finding that the average number of children per household was less than 1.6. Therefore, we should expect a larger number of children per household.

3. Constructing Linear Model

	Estimate	Std.Error	t-value	Pr(>t)
Intercept	3.4502	0.1637	21.0704	5.1159
Age of Respondent	-0.0670	0.0044	-15.3170	1.2021e-51
Age of Having First Child	0.08199	0.0039	21.2041	3.9021e-95
Age of First Marriage	0.0124	0.0045	2.7381	6.2048e-03

We then constructed a linear regression model for variables selected from EDA. A mathematical model is useful to explain how each variable affects the total number of children. In this model, β_1 represents age of respondents at the time of the survey interview; β_2 indicates the age of respondents' first child; and β_3 stands for the age they first got married. From the result of the mode, we could observe that the age of respondents has a negative correlation with the total number of children: as respondents' age rises, they will have a smaller number of children, which is also the same as our analysis from the scatterplot session. The age of first child and age at first marriage obtained a positive correlation: as the value of those variables rising, they tended to have more children.

Discussion

In the exploration of the General Social Survey, the paper mainly examines the number of children per person. Based on the analysis of potential factors that might affect the number of children, we can make conclusions about our findings. We found that younger demographics have more varied marital status and most people having their first child after the age of 30 have mostly remained married. Moreover, we further explored the true mean of the total number of children of the population comparing the result of the dataset used and an outside source. We found that the average number of children is higher than 1.6 based on the result of the GSS. Then we conducted a linear model of respondents' age, age of first child and age of marriage (if applicable) with the number of children they have. All variables had weak relationships with the number of children. The age of respondents negatively impacts the number of children they have.

The vast majority of the population give birth between the age of 20-40. The analysis of the linear regression of the total number of children and the age of having first children have shown a slightly negative relationship. People who own more than 3 children are more likely to have their first child below the age of 30. For households who have only 1 or 2 children, the age of first birth is equally likely to be in the range of 20 to 45. We can say that the correlation between the age of first birth and the total number of children is weaker among families with fewer children.

In terms of the marital status of the population, people who have more than 4 children and give birth to their first child at age below 30 tend to have a higher proportion of either divorced or widowed. The marital status of single child families with first child born below age of 30 is most diverse among other groups. In other words, compared to people who had their first child after 30 and have 2 children, people who had 1 child and gave birth before 30 are more likely to not be in a married state.

The results based on the data collected intrigue us to think about whether marital status affects the time to have first child and the total number of children in total. Since we cannot prove causality among the variables, we cannot ignore the possibility that age of first birth might affect families' marital status. As seen in the graph shown in the results section, people who have 1-3 children with their first child delivered after the age of 30 have the most married families. Overall speaking, people who have their first child later tend to be in more married relationships. Assuming people who give birth to their first child are also married early, the Institute for Family Studies have pointed out that young marriage couples are an indicator of higher divorce rates (Wolfinger 2015) . The results of our analysis also echo that point. It can serve as a reference for young couples wanting to get married. It can potentially lead people to make less irrational decisions

and end up in undesirable situations later on in the relationship. Further research can support the study by exploring the potential relationship between divorce rate and the number of children in total.

The average number of children per person is 2.37 among all families who reported to have at least one child. Several sources such as OECD have indicated that the average number of children should be 1.6. The paper examines the claim by setting up a hypothesis test. The GSS data results concluded that the real mean of the number of children per person should be higher than 1.6. This draws attention to whether each source obtains their data and processes them in the same fashion so that they can compare with one another for more accurate results.

The linear model constructed in the results section suggests that the age of first marriage, age of first child and age of the respondent have little effect on the number of children someone has in total. The most significant factor would be age. Each increment of 1 year older in respondents' age contributes to a negative 0.067 children in the results. Further studies can examine whether combinations of two factors can have a better model than the three combined.

Limitation

Upon closer inspection, there are some limitations that are worth mentioning for replication in the future. The General Social Survey was collected through telephone, fax, email and mail (Statistics Canada) (*General Social Survey* 2020). Even though eventually the research reached the desired number of responses, the response rate out of everyone they reached out was 52.4%. The research has identified that the refusal rate was 3.9%, non-response was 43.8% including household and person level. Since the response rate was only half, there are sampling errors in the data. The characteristics in the sample were not directly proportional to the population. For instance, only 3.7% of respondents are males between the ages of 15- 24. The percentage in the entire population is 7.5%. The researchers decided not to reweigh the results. Moreover, the survey also mentioned that a possible reason to explain the non-responses could be due to language difficulties. The data potentially excluded foreign groups or new immigrants in the samples collected.

Reference

Computing in the Humanities and Social Sciences. 2019. http://www.chass.utoronto.ca.

François, Yihui Xiein, Lionel Henry, and Kirill Müller. 2021. *Knitr: A General-Purpose Package for Dynamic Report Generation in R*.

General Social Survey. 2020. Statistics Canada. https://sda-artsci-utoronto-ca.myaccess.library.utoronto.ca/sdaweb/dli2/gss/gss31/gss31/more_doc/GSS31_User_Guide.pdf.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Russell, Andrew. 2017. *Here's Why Canadians Are Having Fewer Children*. Global News. https://globalnews.ca/news/3429950/canada-fewer-children-census-216/.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

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.%0A%20%20%20%2021105/joss.01686.

Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2020. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.

Wolfinger, Nicholas. 2015. *Want to Avoid Divorce? Wait to Get Married, but Not Too Long*. https://ifstudies.org/blog/want-to-avoid-divorce-wait-to-get-married-but-not-too-long.