

My title*

My subtitle if needed

Kimlin Chin

Jack McKay

Jacob Yoke Hong Si

19 March 2022

Abstract

First sentence. Second sentence. Third sentence. Fourth sentence.

1 Introduction

Canada's General Social Survey (GSS) program is a series of independent, annual, cross sectional surveys on a variety of topics, such as families, caregiving and social identity, with each survey covering one topic in depth. As well as collecting information on the given topic, each survey also collects comprehensive socio-demographic data, recording respondents age, sex, education, religion, etc. Established in 1985, the GSS program was founded with 2 primary objectives: To gather data on social trends as they evolve over time, in order to monitor changes in the living conditions and lifestyle of Canadians, and To provide information on specific social policy issues of current or emerging interest. In these regards, the GSS program has been largely successful, with GSS data often serving as evidence behind key government programs designed to improve the well-being of Canadians. As well, GSS data has also informed research about the social lives of Canadians, and continues to serve as a valuable training tool for quantitative and statistical methods in post-secondary institutions across the country (believe it or not). For these reasons, the GSS program serves an important and foundational role within Canada's national statistical system.

The General Social Survey on Family monitors trends and changes in Canadian families. Family is the single most important influence in a child's life; our family shapes our lives, influencing every aspect of our lives. Indeed, a strong, supportive family is the best and most foundational support a person can have. The GSS on Family aims to provide the government with an accurate picture of the diversity, demographics and trends of Canadian families. The data collected will help guide lawmakers in designing social programs and policies that will best serve the needs of millions of Canadian families. Hence, the GSS on Family is foundational in helping Canadian families, and therefore plays a key role in helping new Canadians have a healthy and functional support system. In this sense, the GSS is helping usher new generations of Canadians into as good a world as possible, and thus plays a critical role in the continued prosperity of Canada.

In this paper, we view a snapshot of what Canadian families look like using data from the 2017 General Social Survey on Family. Specifically, we address the following research questions:

- What are the most common family sizes and types (e.g. married couple, single-parent, etc.)?
- Do most Canadians get married and how many children do they have?
- Are Canadian households mainly low-income, middle-income or high-income?
- Which provinces have the most immigrant families?
- Is religion important in Canadian families?
- What is the mental health of Canadians like, and what factors may be related to this?

*Code and data are available at: <https://github.com/jmacattack27/General-Social-Survey-on-Family-2017-> .

2 Data

2.1 Data Source and Methodology

The data was obtained from Canada's 2017 General Social Survey (GSS) on the Family. The survey was conducted during the period of February 2nd to November 30th 2017 by computer assisted telephone interviews of participants. The target population was all non-institutionalized persons ages 15 years and older in Canada, excluding residents of the Yukon, Northwest Territories, and Nunavut. Participants were chosen using stratified sampling by geographic area within the 10 provinces, with one person representing each household to be interviewed. The response rate was 52.4% with a total of 20,602 respondents.

The age distribution of respondents is shown in figure 1.

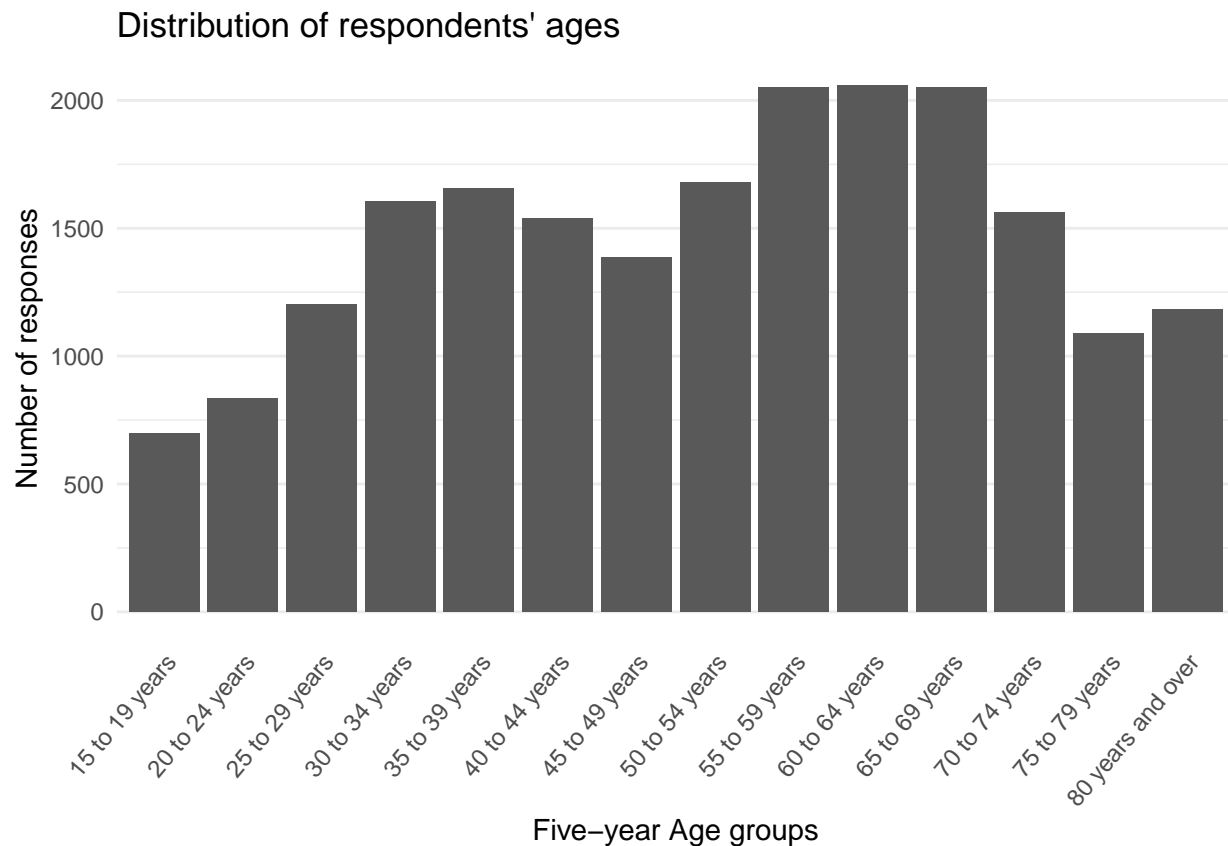


Figure 1: Distribution of Respondents' Ages

Data collected by the GSS is categorized further into 2 components; core content and classification variables. Core content consists of data designed to measure social trends and living conditions, as well as to inform specific policy issues. Examples of core content include questions such as “Are parents satisfied with the child care services they are receiving?” and “How many parents are taking leave for the birth of their child?” Classification variables on the other hand refer to the variables used to collect the aforementioned socio-demographic data. These include age, sex, religion and education of the respondent, and are used to help accurately portray respondent populations, with the goal of helping in the analysis of core content. Below is a brief outline of the questionnaire content on the 2017 GSS on Family:

- Entry component (respondent's date of birth)
- Family origins
- Leaving the parental home

- Conjugal history
- Intentions and reasons to form a union
- Respondent’s children
- Fertility intentions
- Maternity/parental leave
- Organization and decision making within the household
- Arrangements and financial support after a separation/divorce
- Labour market new and education
- Health and subjective well-being
- Characteristics of respondent’s dwelling
- Characteristics of respondent of spouse/partner

2.2 Survey Frame

As previously mentioned, participants of this survey were chosen via stratified sampling wherein each province was divided into strata. Most Census Metropolitan Areas (CMAs) were considered separate strata, while the remaining CMAs were grouped together and divided into 3 separate strata. In total there were 27 strata used for sampling.

The survey frame was then created using 2 main components: Lists of telephone numbers, both home lines and cellular, that were available to Statistics Canada, and The Address Register (AR), a list of all the dwellings within the 10 provinces.

One person was surveyed per household; thus the Address Register was used to group together phone numbers by address. Approximately 86% of telephone numbers were linked to an address via the AR. These numbers, as well as the remaining 14% of unlinked numbers, were then combined to create the sampling frame. When multiple phone numbers were linked to the same address, these numbers were then sorted by type of number, with landlines first and cellular numbers last. The first number was then considered the best number to reach that household; recall that we want to survey households, not individuals, and so landlines were prioritized over cellular numbers. Note that as a result of the cellular surveying method, households without a telephone were excluded from the survey frame.

2.3 Sampling Methods

Each phone number / group of phone numbers (which we will henceforth refer to as “records”) was then sorted into its corresponding geographical strata, at which point a simple random sample without replacement was then performed on each strata. Approximately 91.8% of telephone numbers selected in the sample reached an eligible household (an eligible household being defined as a household that included at least one person 15 years of age or older). Data on households that did not meet the eligibility requirements was not included in the results, and phone calls with such households were terminated.

2.4 Key features

The raw data includes 461 variables which were derived from the questionnaire responses. Some key features included age, sex, place of birth, marriage, ethnic diversity and immigration, education, work, income, children, religion, family structure, geographic location, household size and living arrangements. We used R (R Core Team 2020) and R packages tidyverse (Wickham et al. 2019) and janitor (Firke 2021) to clean the variable names and values. A subset of the key features is shown in table 1.

2.5 Strengths and weaknesses

As with any statistical study, this survey has both strengths and weaknesses. Some of the strengths include the multitude of variables as well as the large sample size. Many variables provide ample opportunity for data exploration and research insights, while the large number of observations allow us to draw meaningful conclusions. One of the weaknesses of the survey is that the response categories for some of the survey questions were collapsed in the provided code; for example, ‘mother tongue’ only lists English, French and

Table 1: Some key features

Age of respondent	Total number of children	Sex	Province	Marital Status	Family Income
52.7	1	Female	Quebec	Single, never married	\$25,000 to \$49,999
51.1	5	Male	Manitoba	Married	\$75,000 to \$99,999
63.6	5	Female	Ontario	Married	\$75,000 to \$99,999
80.0	1	Female	Alberta	Married	\$100,000 to \$ 124,999
28.0	0	Male	Quebec	Living common-law	\$50,000 to \$74,999
63.0	2	Female	Quebec	Married	\$50,000 to \$74,999
58.8	2	Female	Nova Scotia	Single, never married	Less than \$25,000
80.0	7	Female	Quebec	Divorced	Less than \$25,000
63.8	0	Female	British Columbia	Single, never married	Less than \$25,000
25.2	1	Male	Saskatchewan	Single, never married	Less than \$25,000

non-official languages, and ‘race of respondent’ was collapsed to visible minority (yes/no). In doing this, some data was lost, and as such we can make less inferences on the diversity of the responses.

3 Results

3.1 Household Size and Family Types

Figure 2 shows the distribution of Canadian households grouped by household size, based on what type of family they are. Note that an intact family refers to a now-married of common law couple, where all of the children have both parents, biological or adoptive, living in the household. Similarly, a step-family refers to such a couple where at least one of the children is in a step-relationship with at least one of the parents. Here we see that the majority of households consist of only 1 or 2 people. From this we can infer that a large portion of Canadians either live alone or with either a partner or roommate. For household sizes greater than 2, the vast majority consist of intact families, with most families having 1 or 2 children. Only 6.79% of households have 5 or more people living together.

3.2 Marriage and Children

Figure 4 shows the distribution of the total number of children per respondent. Note that this is not *per household* i.e. persons living together, but rather is the number of children had by the respondent of the household throughout their lifetime. Here we see that the majority of respondents either had 0 children or 2, with almost an identical number of respondents with 0 children as there are with 2. Similarly, the number of respondents with 1 child is very close to the number of those with 3. Thus, we note that about 30.21% of Canadians do not have any children, and those who do usually have 1-3 children.

Figure 5 graphs the distributions of the age of respondents when they first got married and when they first had children, as well as the number of years between their first marriage and their first child. For first marriage, the majority of responses lie between 20-30 years, with a peak at around 21 years. For first birth, the bulk of responses were between 20 and 35 years of age, with the peak at 25 years. Notably, the distribution is almost uniform between 20 and 30 years. The distributions for first marriage and first child are both right-skewed. This is to be expected, as the likelihood of getting married would intuitively increase drastically as one approaches 20, and drop off gradually past 30. A similar trend is to be expected with first births, as women first become fertile around puberty, and generally reach menopause at 45-50 years (National Institute on Aging, 2021).

15.3% of people who were ever married do not have any children, and 6.39% of people who have children were never married. For those persons who get married and have kids, plot A of Figure 5 shows that the

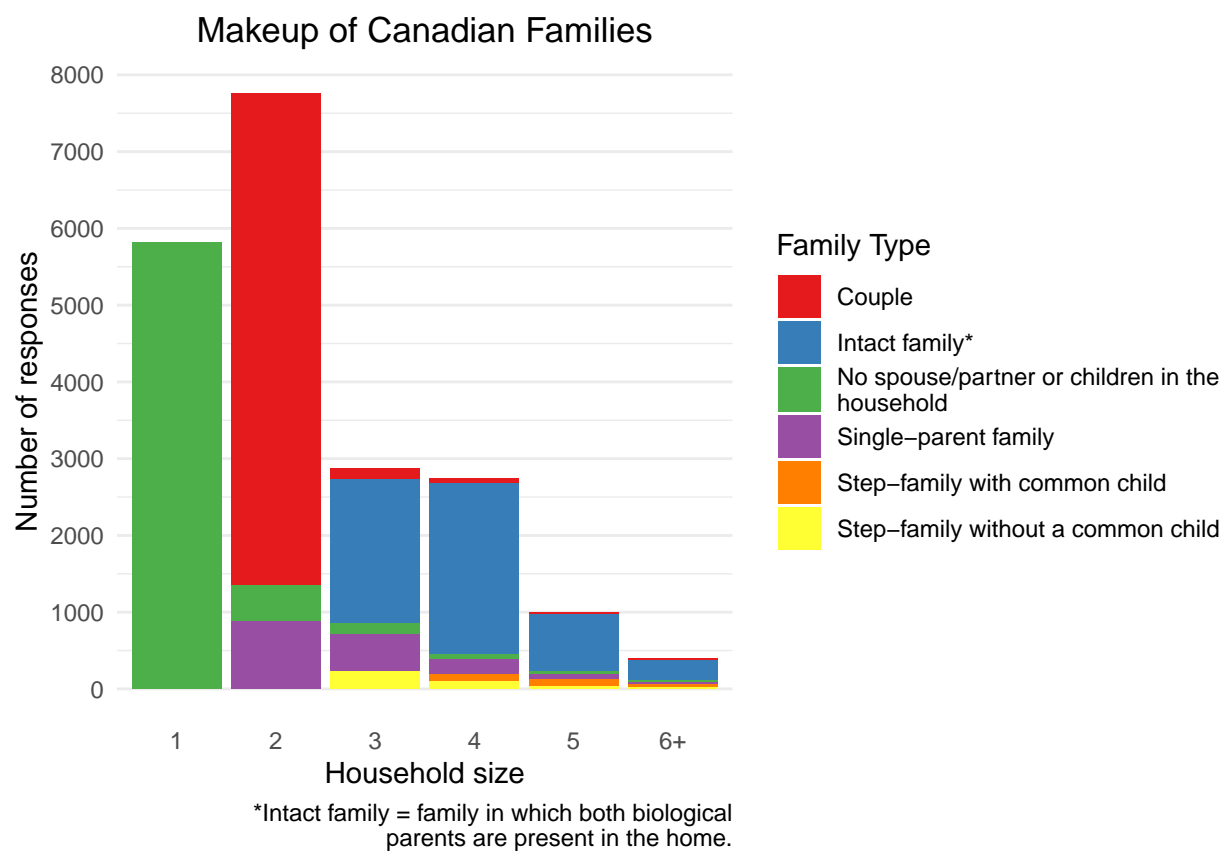


Figure 2: Distribution of Family Type by Household Size

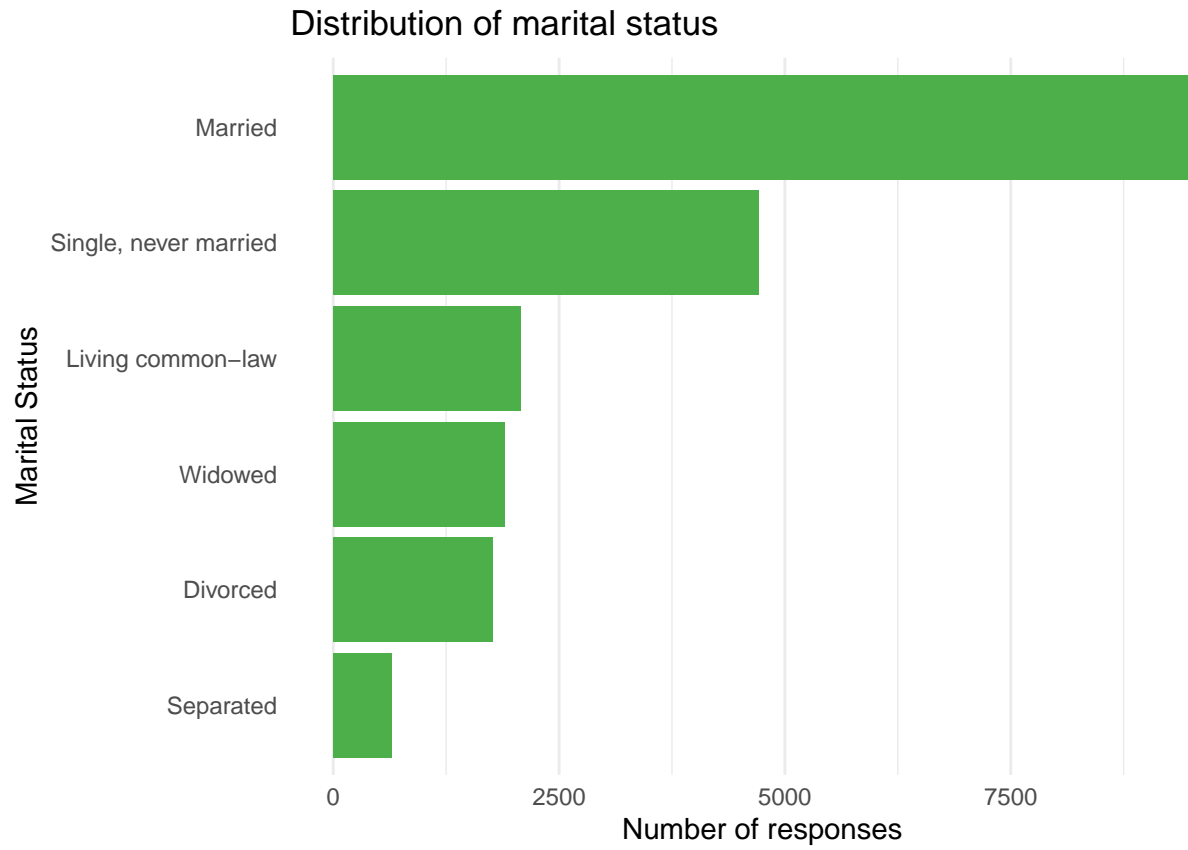


Figure 3: Distribution of marital status

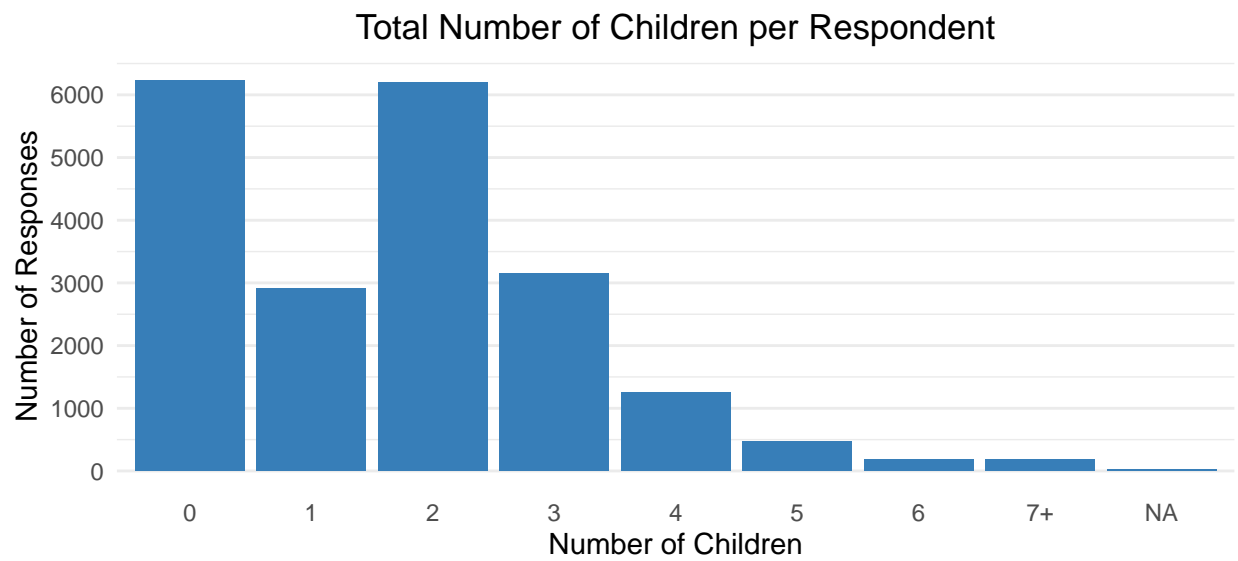


Figure 4: Total Number of Children per Respondent

majority of Canadians bear children within the first 10 years of their first marriage, with by far the most married births occurring within 1 year of marriage.

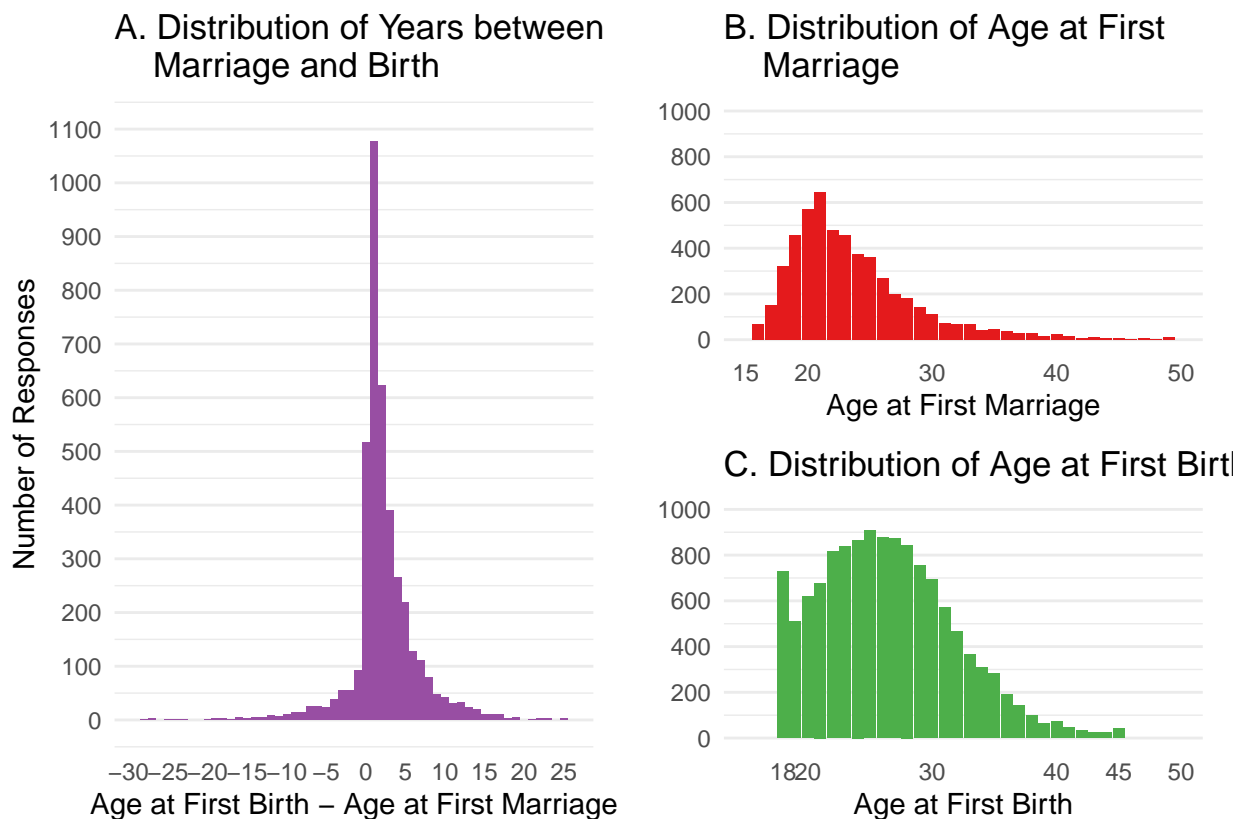


Figure 5: Distribution of Ages at Marriage and Birth

3.3 Place of Birth and Immigration

Figure 6 shows the distribution of respondent's province of birth, as well as distribution of respondents' parents' birth place, the latter distribution only looking at whether the parents were born inside or outside of Canada. Although 79% of Canadians in our sample were born in Canada, the number of individuals born outside Canada is greater than the number of individuals born in any individual province. As expected, the vast majority of these persons' parents were also born outside of Canada. Within Canada, the most common provinces of birth are Ontario and Quebec. Notably, despite a similar number of respondents being born in Ontario and Quebec, respondents born in Ontario are more than twice as likely to have one or more parents born outside of Canada.

The provinces with the largest population of second generation Canadians—defined as a person born in Canada with one or more parents born outside Canada—are Ontario, British Columbia, Quebec, Alberta, Saskatchewan and Manitoba. In figure 7, we see that first generation Canadians—defined as a person with both parents born outside Canada) mainly settled in the same provinces as the second generation immigrants. Notably, provinces on the east coast such as Nova Scotia and New Brunswick are almost entirely comprised of third generation Canadians, or Canadians born in Canada with both parents also born in Canada.

Figure 8 shows the years that respondents who were born outside of Canada reported arriving in Canada, as well as where they hailed from. Notably the population of Europeans made up the largest proportion of immigrants and was at its largest amount during the period 1946-1959, but then decreased to almost

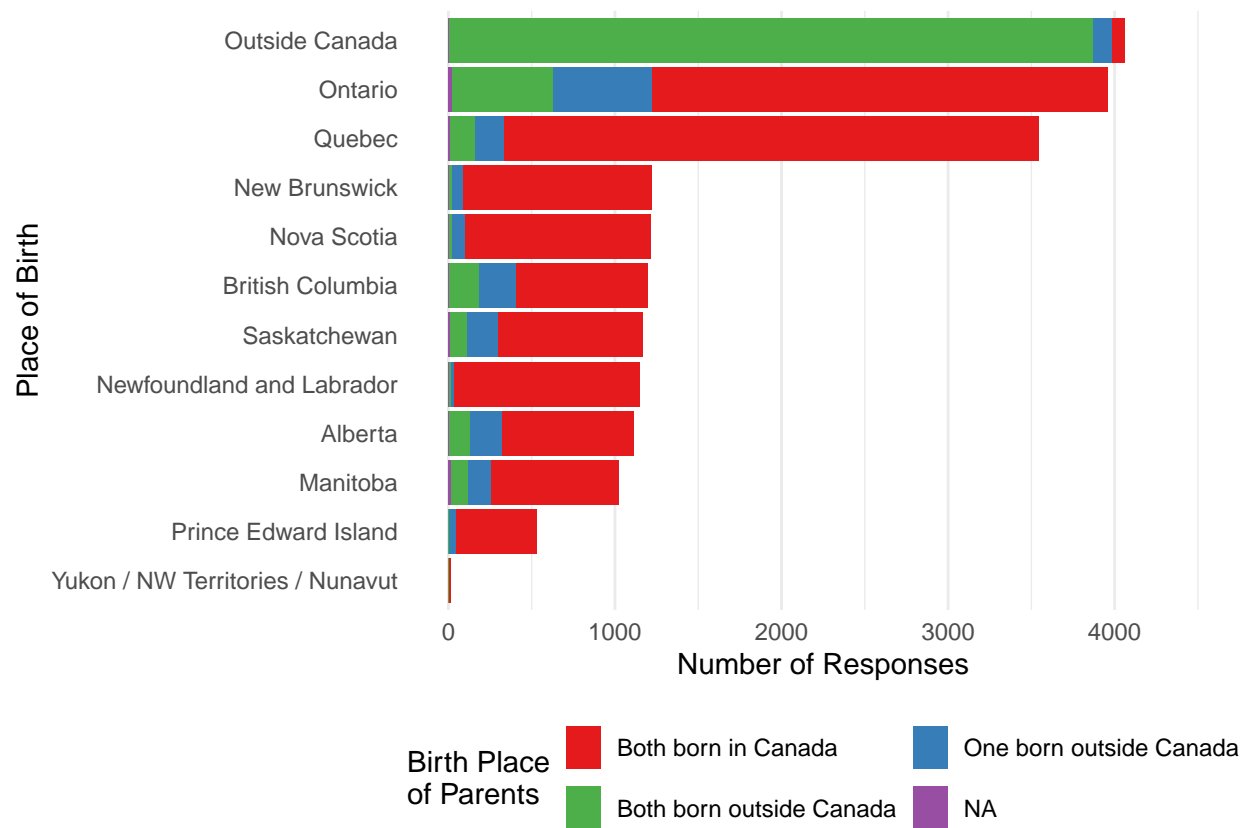


Figure 6: Distribution of Persons by Place of Birth

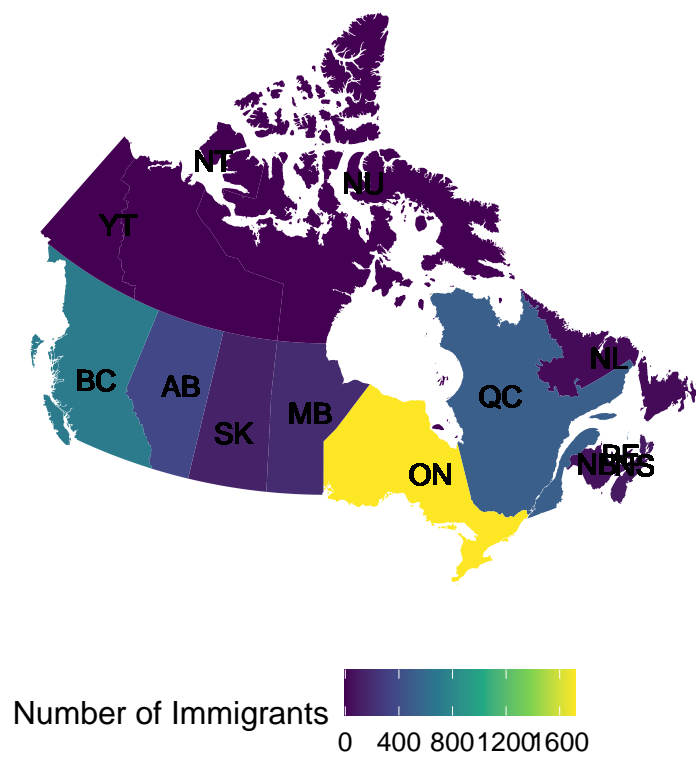


Figure 7: Distribution of Immigrants Currently Residing in Canada by Province

a constant number from about 1970 onwards. In contrast, the population of Asian immigrants has been increasing steadily up to 2014 and it has comprised the largest subsection of immigrants since about 1985. The number of immigrants from Africa and the Americas has not changed too significantly, but the years 2000-2014 saw a boom in immigration from these groups together with Asian immigrants.

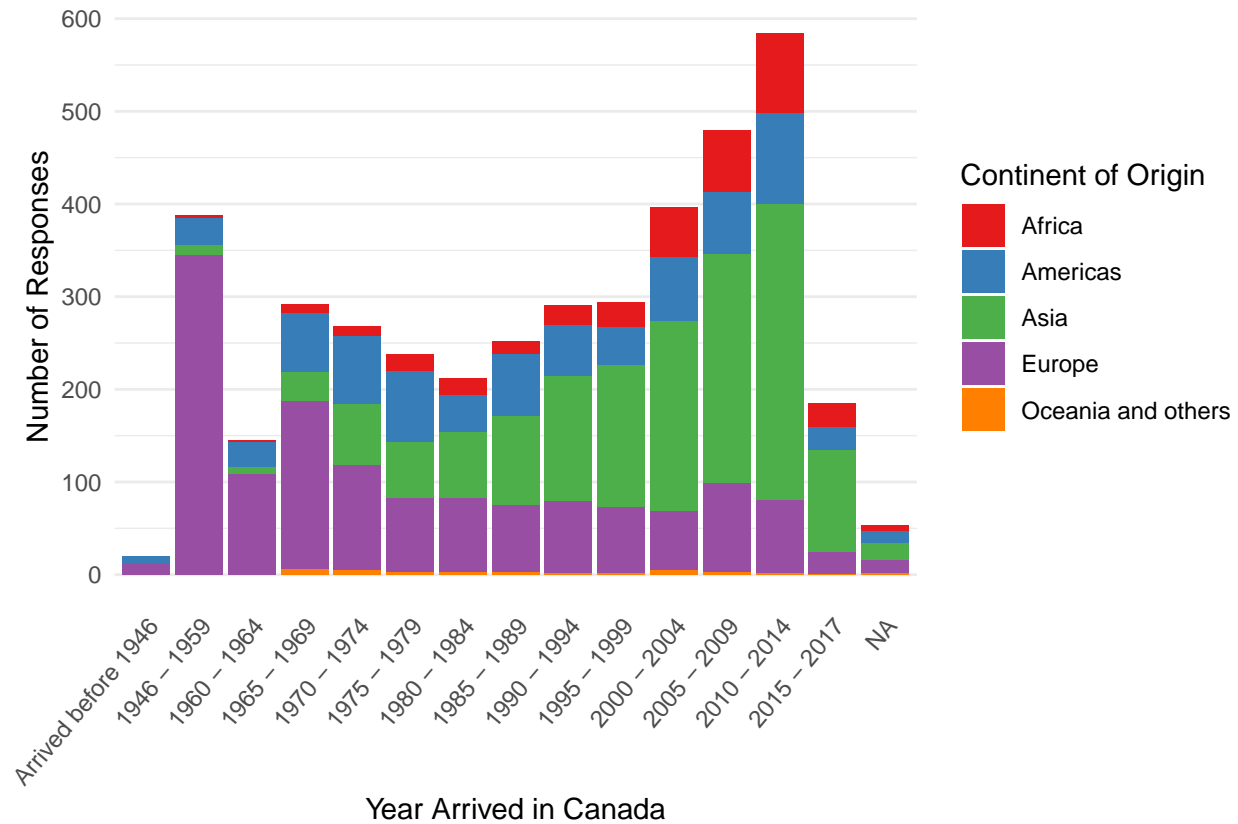


Figure 8: Distribution of Immigrants' Continent of Origin by Year

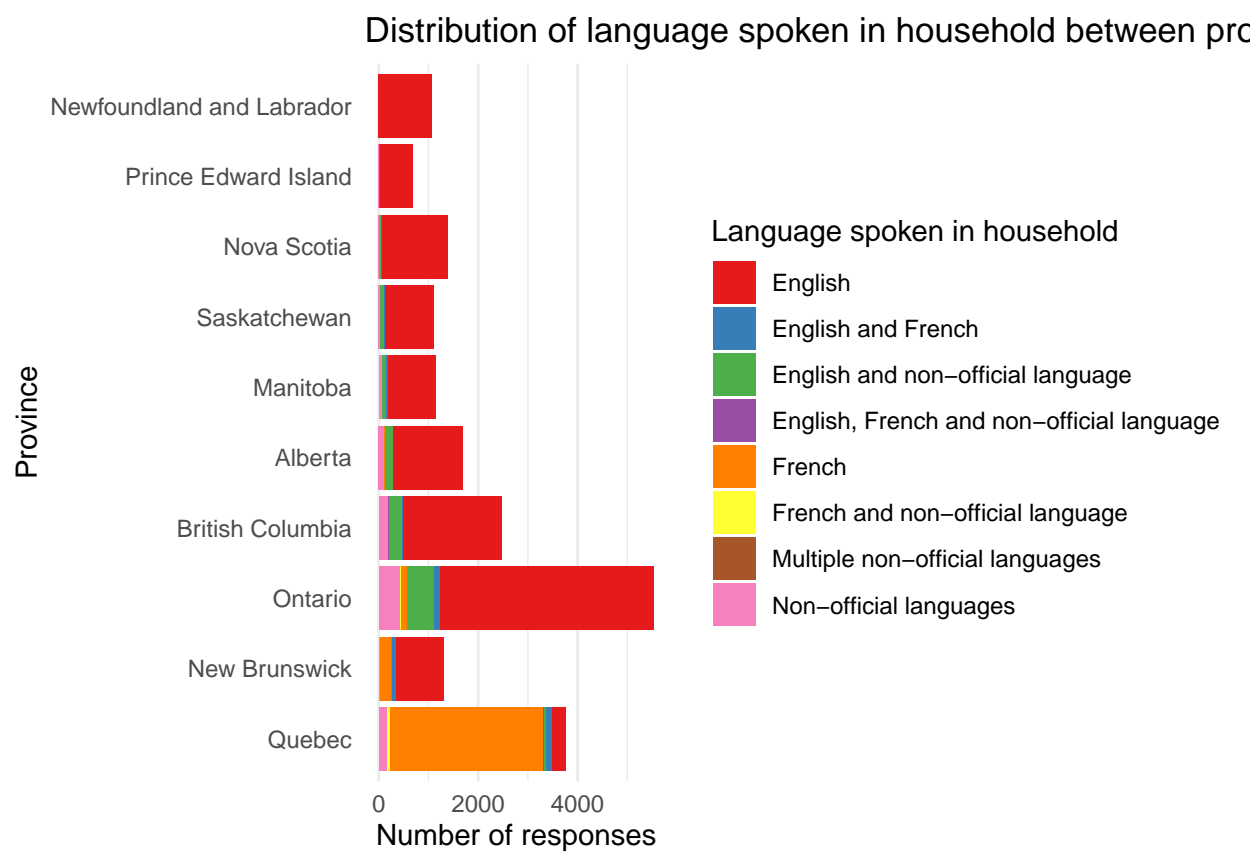


Figure 9: Distribution of languages by province

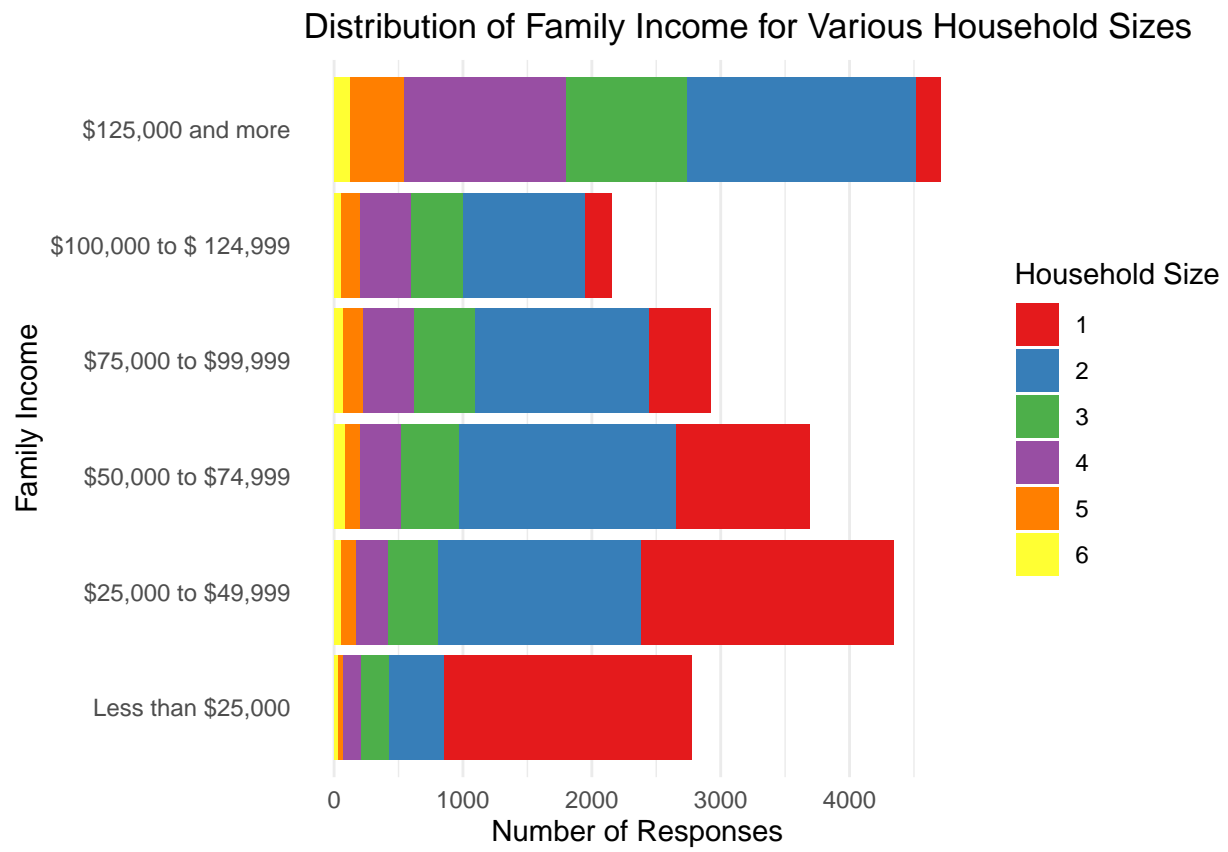


Figure 10: Distribution of Income

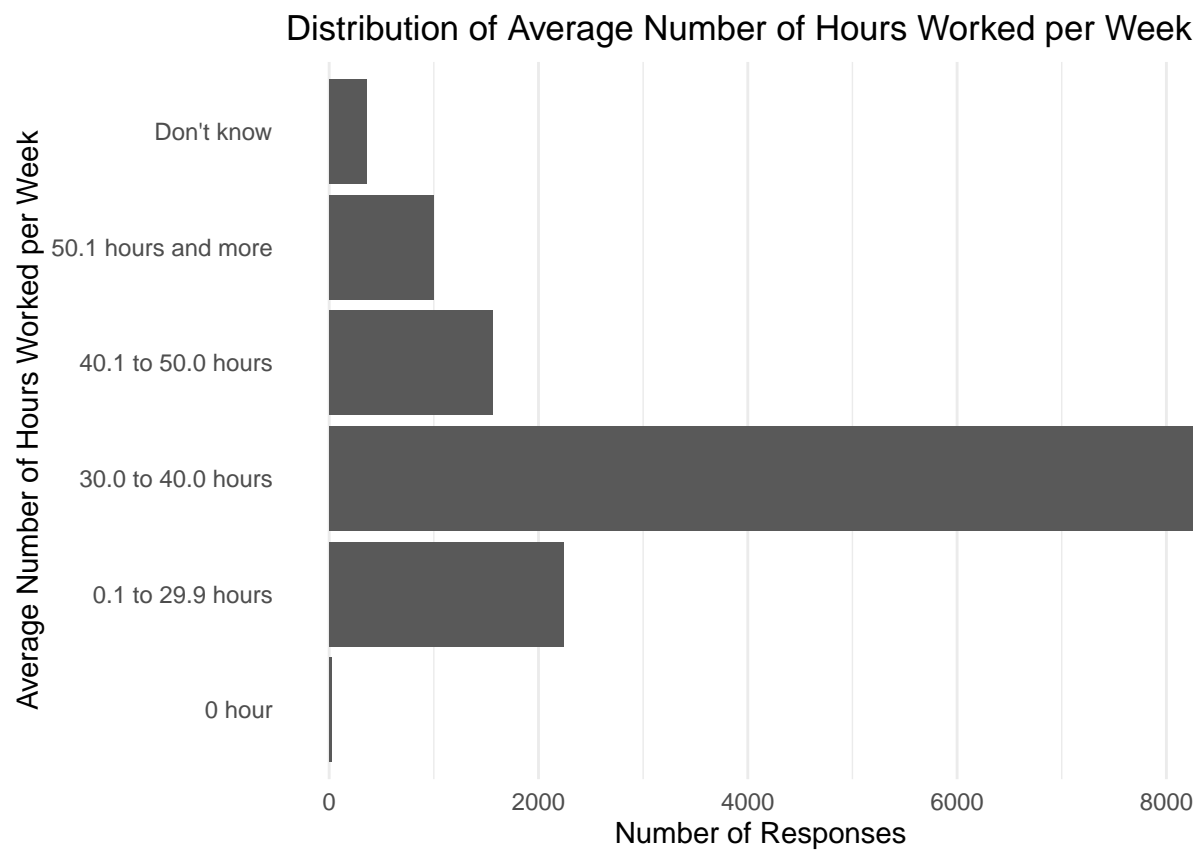


Figure 11: Distribution of Hours Worked

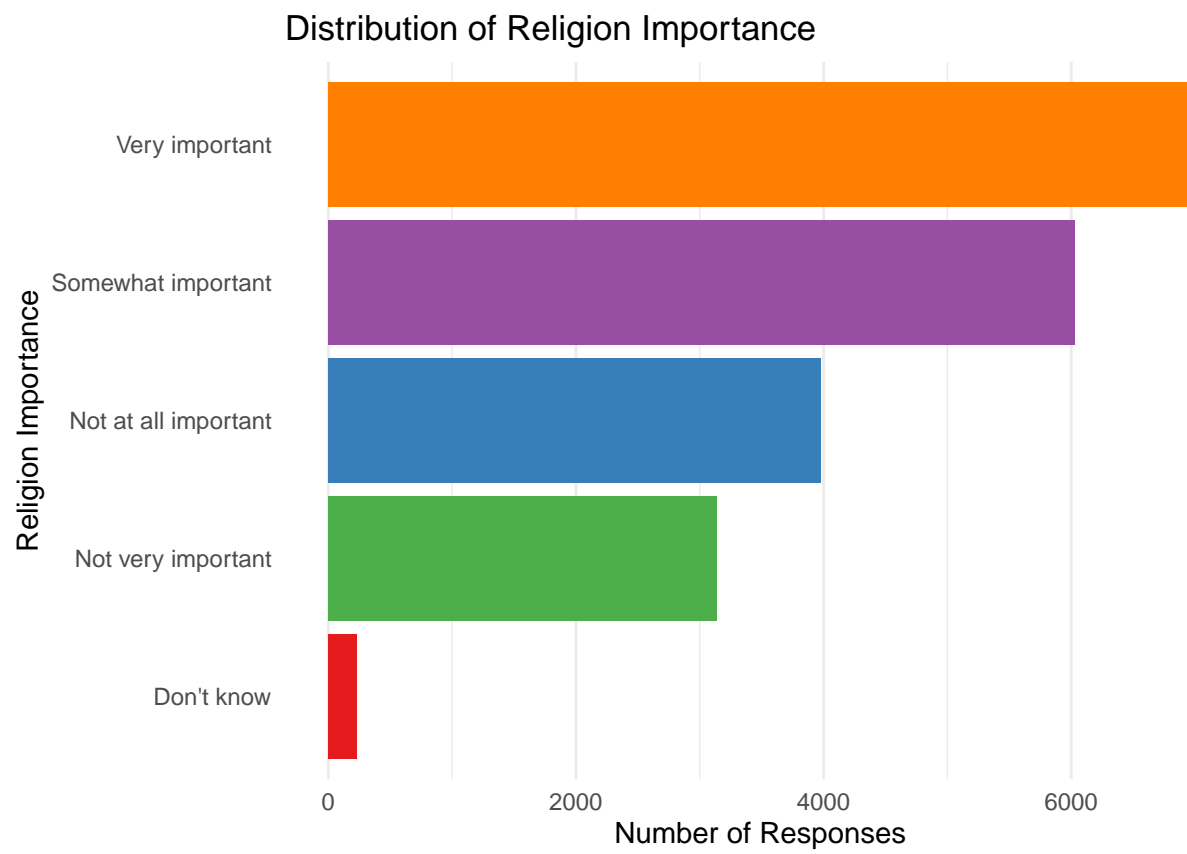
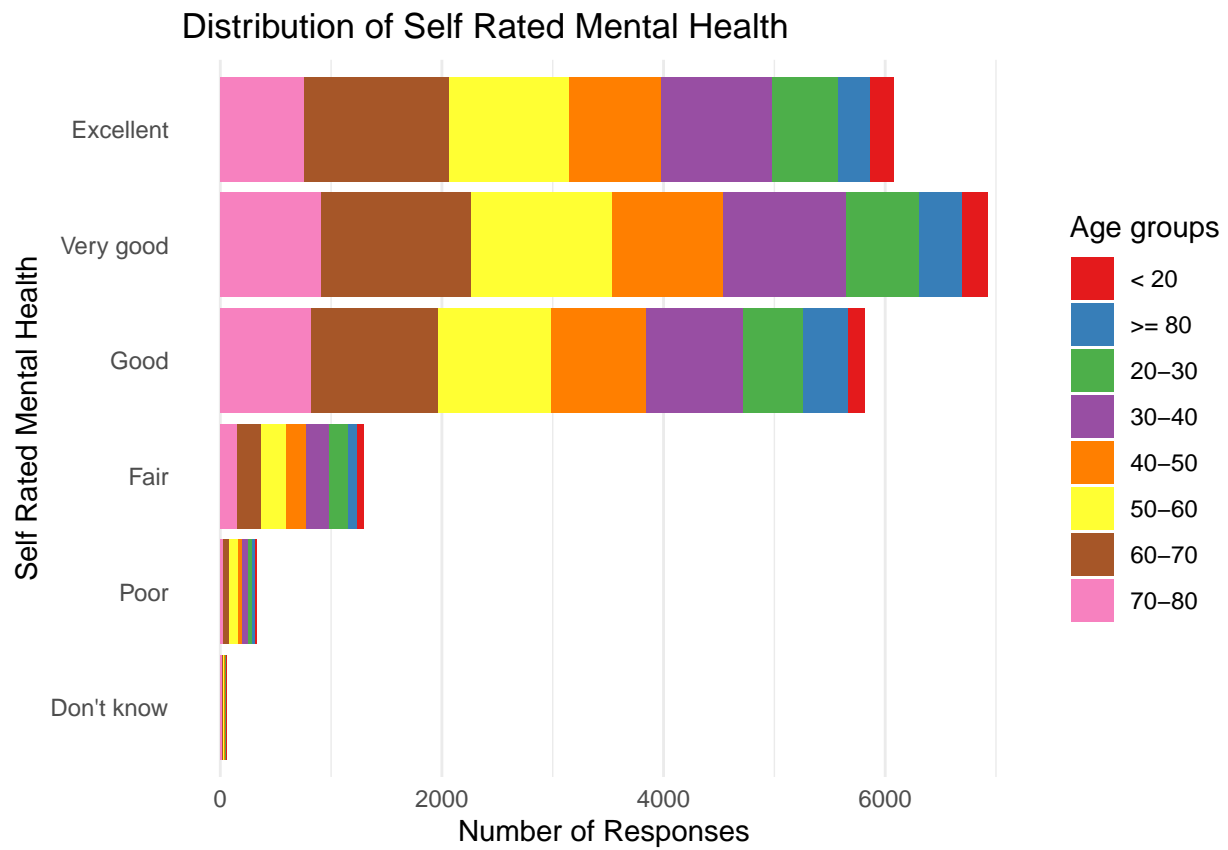


Figure 12: Distribution of Religion

3.4 Family Income

3.5 Religion

3.6 Mental Health



4 Discussion

4.1 Mental Health

The topic of the Canadian General Social Survey in 2017 was family; it aimed to capture as much information as possible about the structure of Canadian families, with the goal of understanding and monitoring the evolution of families over time. This information can then be used to guide social and fiscal policies so that they may best serve the current needs and wants of Canadian families. As well, the survey provides a glimpse at how well current policies are serving Canadians. For example, question SRH_Q115 asked “In general, would you say your mental health is...?” with respondents choosing one of “Don’t Know, Poor, Fair, Good, Very Good and Excellent”. The distribution of these responses is exhibited in Figure (insert the abt mental health). Very notably, the vast majority of respondents rated their mental health as “Good”, “Very Good” or “Excellent”, with an exceedingly small proportion of people answer “Poor”. As well, Figure (insert the abt mental health) also shows the distribution of the age of respondents for each answer, in which we see nearly identical proportions of people of all age groups for each answer. This means that not only are Canadians happy in general, but also that this happiness is not skewed towards certain age groups, as we had originally hypothesized. It must be made note of that this survey was conducted in 2017 and thus conducted pre-pandemic. As such, this graph likely does not provide us with an accurate picture of the current sentiments of Canadians. As well, by virtue of the survey not being mandatory for chosen respondents to complete, there is likely an intrinsic sampling bias towards people with better mental health; some of the defining symptoms of poor mental health include tiredness, lack of energy and a lack of initiative. As such, people with poor mental health would be inherently less likely to voluntarily complete a survey.

4.2 Family Planning

Canada has always been committed to having strong family planning and sexual and reproductive health (SRH) services. In 2017, Global Affairs Canada – a branch of the Canadian government – dedicated CA\$650 million over the next 3 years towards promoting the sexual and reproductive health and rights of women and girls (cite). In 2015, Canada committed CA\$3.5 billion towards maternal, newborn and child health, including improving SRH services and information. This strong foothold in the realm of family planning is reflected in this survey in a number of ways. For example, Figure 2 showcases the distribution of the size of Canadian households, broken down by what type of family lives in that household. Here we see that the majority of households only consist of 1 or 2 people, followed by households containing 3 or 4 people, of which there were a very similar amount. Notably, there is a steep drop-off in frequency of households larger than 4 people, with about a third as many 5-person households as there are 3 or 4-person, and less than a sixth as many households containing 6 or more people. This is likely a reflection of Canada’s ample family planning resources leading couples more likely to opt for 1 or 2 children, thus keeping the average household size low. Figure 4 shows the distribution of the total number of children had by respondents. Interestingly, we see that by far the most frequent responses are 0 children and 2 children, of which there were a nearly identical number of responses. This means that once a couple has one child, they are then more likely to opt to have another child. This seemingly conflicts with Figure 2, which showed that 3-person households were just as common as 4-person households. There could be a couple reasons for these seemingly disjoint results. The firstly, relative to the general population, a high proportion of elderly people were surveyed. As elderly people are unlikely to still be living with their children, this would naturally lower the average household size without affecting the average number of children. As well, members of the same household are not necessarily parents or children, nor even family members. For example, a small proportion of 3-person households consist of couples. As these households are classified as couples and not intact families, this entails that the 3rd member of the household is not one of the couple’s children. In Figure 2 we also see that aside from 0 and 2, 1 and 3 are both by far the next most common responses, with about 2900 respondents having 1 child and 3200 having 3. This also goes to show that Canada’s continued investments into SRH and family planning services have been effective in keeping reproductive rates healthy and sustainable on both an individual and aggregate level.

4.3 Immigration

Canada has a strong track record when it comes to immigration, boasting one of the highest naturalization rates in the world, with about 85% of newcomers eventually becoming citizens. Figure 6 shows the distribution of the birth place of the respondents parents, grouped by the respondent's province or territory of birth. Here we see that for respondents born in Canada, Ontario has by far the greatest amount of second generation Canadians, with about 30% of Canadians born in Ontario having at least one parent born outside of Canada. British Columbia has a similarly high proportion of second generation Canadians, followed by Alberta, Manitoba and Saskatchewan, each of which has a similar proportion. This is to be expected, as the 2 main hotbeds of immigration in Canada, Toronto and Vancouver, are the capital cities of Ontario and BC respectively. Notably, provinces on the east coast such as New Brunswick, Newfoundland, Nova Scotia and Prince Edward Island all have exceedingly low rates of immigration. While there could be cultural and societal causes for this lack of Atlantic immigration, the most obvious cause would be the general rurality of the provinces; in general, immigrants tend to settle down in urban areas, largely due to the economic and societal opportunities that present themselves in big cities. This is evident in Canada, where immigrants flock to large urban centers such as Toronto, Vancouver and Montreal. Provinces on the Atlantic are far less urbanized than Ontario or BC, and as such immigrants are less likely to settle down in them. Figure 7 shows a density map of where immigrants currently reside in Canada, supplementing the previous figure. Here we again see that Ontario has by far the most immigrants of any province, followed by British Columbia. It's worth noting that this map only shows the number of immigrants residing in each province, as opposed to the proportion. So while Ontario has far more immigrants than BC, they in fact have a similar proportion of immigrants in their population. Figure 8 shows how many immigrants arrived in Canada over 5 year periods since 1946, as well as which continent they were immigrating from. Here we see a sharp increase in the number of immigrants arriving in Canada since 2000, with that number continuing to trend upwards. This coincides with the ongoing efforts by the Canadian government to drive immigration rates up, which is one of the key ways that Canada aims to combat the economic decline driven by its aging population. Increasing immigration rates allows for a new inflow of young and eligible workers into the country, bolstering Canada's workforce and stimulating the economy. Another noticeable trend shown in Figure 8 is the increasing rate at which immigrants are hailing from Asia, which has been increasing continuously since 1964 but most noticeably exploded in the early 1980s. This increase coincides with the Early 1980s Recession experienced by Canada, as well as the drastic economic growth China began to experience in the early 1980s and continue to today. Other trends include the increasing rate of African immigrants, and the decreasing rate of European immigrants, with European immigrants making up the smallest proportion of immigrants since 2010. Despite the proportion of European immigrants decreasing, the actual number of European immigrant respondents has remained steady since 1975, meaning that this drop in proportion is more likely related to the drastically increasing amount of Asian immigrants than it is a decreasing propensity for Europeans to immigrate to Canada.

4.4 Weaknesses and Next Steps

There are some weaknesses in our paper that we will discuss. First off, we did not collect the data that we are making inferences from ourselves, and as such, we only know as much about the methodology behind it as we've been told. While this certainly could have been a more substantial issue, Statistics Canada did more than their due diligence in providing information about every aspect of the survey, explaining the sampling procedures and survey methodologies in depth, as well as providing a 264 page codebook explaining every variable in depth, as well as the corresponding question. As such, this is more of an ethical weakness than a practical one. However, there could be biases unaccounted for by Statistics Canada, or other failures in the sampling or surveying process, ethical or otherwise, and we could not know about them. Another possible weakness in our paper is our extensive data cleaning process, wherein we quite literally filtered out hundreds of variables, only keeping the ones we deemed statistically relevant to our paper. In doing so, we sacrificed some of the context that those variables could have lent to the ones we selected. However, this falls back to the ethical question of whether or not data should be left to speak for itself; we opted to extensively clean and filter the data so as to make it as interpretable as possible, but in doing so we did lose information that could have changed our interpretations of the data. As well, some of the variables we filtered out were related to some of the variables that we included, and as such would have almost certainly provided additional context

to our graphs and figures. Despite this loss of context, due to the multitude of variables in the data, this was a sacrifice we deemed worthwhile for the overall interpretability of our paper, as well as to avoid wrangling 50+ variables. The final shortcoming of our paper is that we did not end up conclusively answering all of our research questions. For example, while the data we collected seems to conclusively state that the mental health of Canadians is very good, as previously mentioned there are some inherent biases in the sampling method that make people with poor mental health less likely to take the time and effort to complete the survey, skewing our results. As well, there was a general lack of questions regarding mental health, and as such we were limited in what conclusions we could draw. We account for this lack of questions about mental health in our supplementary survey, which was designed with our research questions in mind. As well, even though we didn't conclusively answer all of our research questions as we had hoped, we still answered the majority of them, if not conclusively than the very least satisfactorily.

Appendix

A Additional details

References

Firke, Sam. 2021. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*.

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

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Golemund, et al. 2019. “Welcome to the tidyverse.” *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.