My title*

My subtitle if needed

First author Another author

17 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 show 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?
- What is distribution of education level of Canadians?
- 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.

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.

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

For the 2017 GSS, 91.8% of the selected telephone numbers reached eligible households. To be eligible, a household had to include at least one person 15 years of age or older. During collection, households that did not meet the eligibility criteria were terminated after an initial set of questions. A respondent was then randomly selected from each household to participate in a telephone interview.

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

Strengths include that there is a large number of variables so there is a lot of opportunity for data exploration and research insights, and there is also a large number of observations for us to draw meaningful conclusions. A weakness includes that response categories for some of the questionnaire questions were collapsed in the code variables provided, for example mother tongue only lists English, French and non-official languages, and race of respondent became visible minority (yes/no), so we cannot comment on the full range of diverse responses.

3 Results

3.1 Household size and family types

Figure 1 shows the distribution of household sizes with family type. The most common household sizes are 1 and 2-person, so Canadians generally live alone or in a romantic relationship. For household sizes greater than 2, they are mostly nuclear families, with 3-4 persons. Only 6.79% of households have 5 or more people living together.

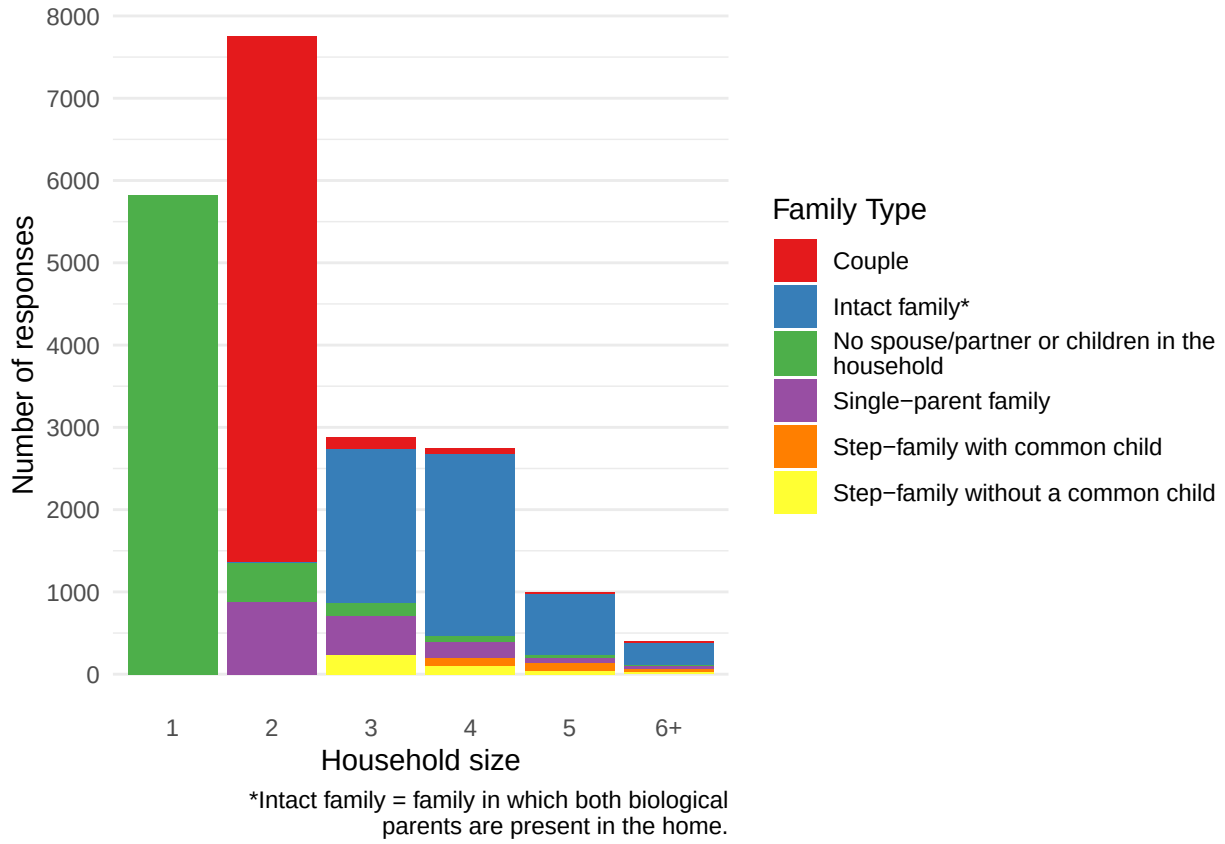

Figure 1: Distribution of family type by household size

3.2 Marriage and Children

Figure 2 shows the distribution of the total number of children per family. Note that this is not *per household* i.e. persons living together, but the number of children ever had by the respondent of the household. The two tallest bars of nearly equal height are for 0 and 2 children, followed by 1 and 3 children also being similar in frequency, but about half as much as for 0 and 2. Thus, 30.21% Canadians do not have any children and those who do usually have 1-3 children.

Figure 3 shows the distributions of the age of respondents when they first got married and had children, as well as the number of years between their first marriage and their first child. For first marriage, a lot of the density of the distribution lies between 20-30 years, with a peak at about 21 years. For first birth, the density mainly lies between 18-35 years, with the peak at 25 years. Both distributions are right-skewed.

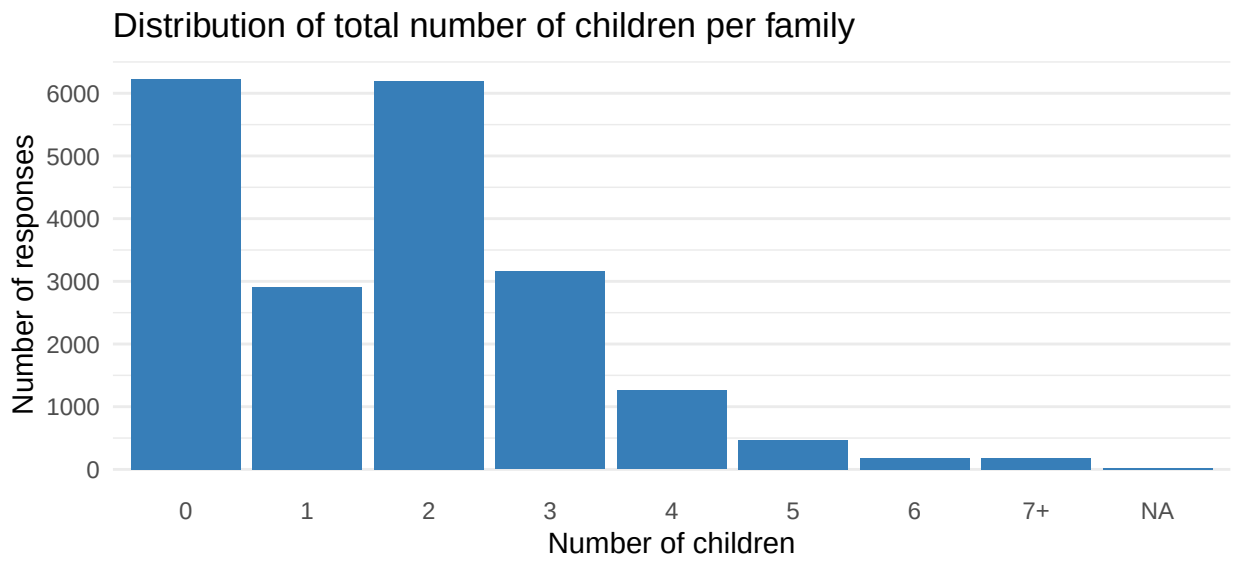

Figure 2: Distribution of total number of children per family

This is expected for first birth since women reach menopause at ages 45-50 (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 3 shows that the majority of Canadians bear children within the first 10 years of their first marriage, with 1 year of marriage being the most frequent occasion for giving birth.

3.3 Place of birth and Immigration

Figure 4 shows the distribution of respondent's province of birth. 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 one of the provinces. 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.

The provinces with the largest second generation families (person born in Canada with one or more parents born outside Canada) are Ontario, British Colombia, Quebec, Alberta, Saskatchewan and Manitoba. In figure 5, we also see that first generation families (person and parents all born outside Canada) mainly settled in the same provinces as the second generation immigrants.

Figure 6 shows the years that respondents who were born outside of Canada reported arriving in Canada and 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 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 3: Distribution of ages at marriage and birth

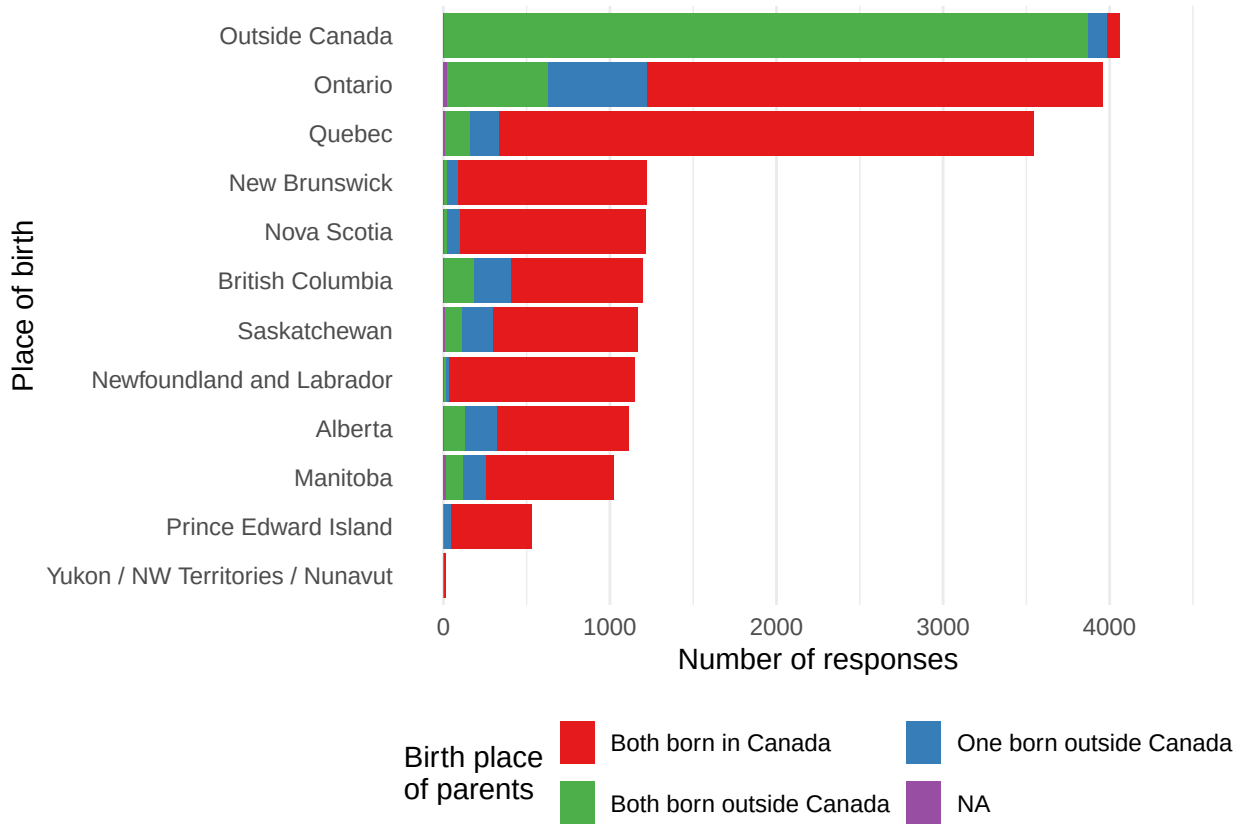

Figure 4: Distribution of persons by place of birth

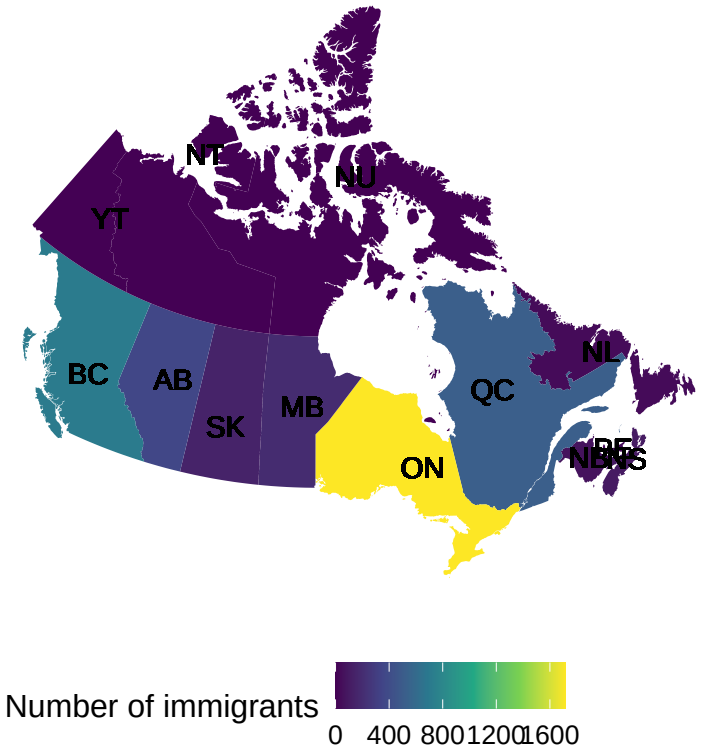

Figure 5: Distribution of immigrants currently residing in Canada by province

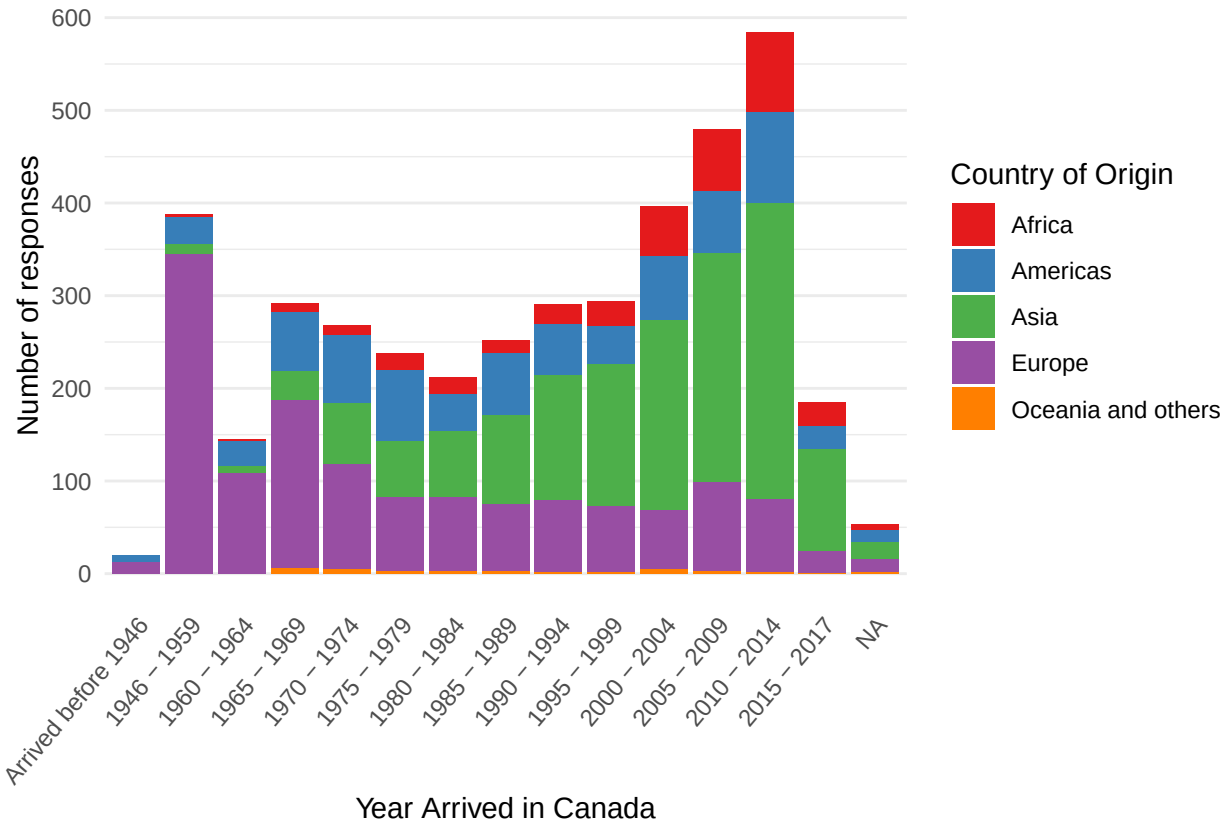

Figure 6: Distribution of immigrants' country of origin by year

4 Discussion

4.1 First discussion point

If my paper were 10 pages, then should be be at least 2.5 pages. The discussion is a chance to show off what you know and what you learnt from all this.

4.2 Second discussion point

4.3 Third discussion point

4.4 Weaknesses and next steps

Weaknesses and next steps should also be included.

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 Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.