Family characteristics and socio-economic conditions in Canada*

Yahui Zhang, Zhiqi Chen

20 March 2022

Abstract

The data discussed in this article comes from the General Social Survey of Canada (GSS), which collects data on various aspects of Canadian households. This paper uses data from the General Social Survey to determine the characteristics and socio-economic conditions of the respondents' families. One study involved analyzing the family characteristics and economic status of Canadian families and what are the stories these data are telling us.

1 Introduction

Canada is one of the world's most diverse countries and is known for its high quality of life. People of all backgrounds are able to find opportunities and succeed in Canada. However, Canada is not perfect. It might be a land- of opportunities for many but a land of sadness for others. This makes us wonder how is the 'average' life in Canada and what are the characteristics and socio-economic conditions like for most Canadians?

This paper will analyze Canadian families' characteristics and socio-economic conditions with the data we gathered from the General Social Survey by Statistics Canada. The microdata of the survey is accessed from the CHASS data center at the University of Toronto. Through an in-depth analysis of the families and respondents' income levels, average working hours, household size, type of housing, primary residence ownership, education level, citizenship status, visible minority status, marital status, place of birth, the region they are from, and what kind of population center they are from, we can put a picture together of what the demographic and quality of life is really like in Canada.

In addition to the data set we have gathered from the General Social Survey by Statistics Canada, we have also put together a short survey to collect more data from the voluntary participants to better our data, thus extending the possibility of improving the depth and breadth of our research paper.

2 Data

To better understand the relationship between family characteristics and their socio-economic relationship in Canada, we have extracted the data generated from the 'General Social Survey, cycle 31 - Families' by Statistics Canada (Canada 2017). The data set we are using is accessed from the computing in the Humanities and Social Sciences (CHASS) data center at the University of Toronto (Centrer n.d.). As the user guide of the survey revealed, the purpose of this survey is to "gather data on social trends in order to monitor changes in the living conditions and well-being of Canadians over time and to provide information on specific social policy issues of current or emerging interest" (Canada 2017). Many important questions

*Code and data are available at: https://github.com/YahuiZhangUofT/Family-characteristics-and-socio-economic-condition-in-canada

were raised to study the demographics of Canadian families and thus, help construct the questionnaires included in the General Social Survey.

The General Social Survey consists of 20,602 observations and 81 variables. Among these 81 variables, we have selected thirteen variables to conduct our analysis of the family characteristics and socio-economic conditions in Canada. Variables include the respondents' birthplace, marital status, visible minority status, citizenship status, education status, residential region, the population center indicator, ownership of the respondents' residence, housing type and size of household, average working hours, family income, and their personal income. By analyzing these data, we can help the audiences grasp the characteristics of Canadian families and their socio-economic conditions.

We analyzed the data set with the help of the R language, and several R packages were used to help visualize data and clean the raw data (R Core Team 2021). For the data cleaning process, we shall thank Dr. Rohan Alexdander for helping our team clean the data (Alexander 2020). The cleaning process was mainly focused on understanding the code that statics Canada uses and transferring them into an understandable form so that we can better comprehend the knowledge, where the package Janitor was used to help examine and clean the raw data (Firke 2021). We have also changed some of the income levels' names so that R and the ggplot function can better rank some elements in an optimal order on graphs. To help plot the graphs we need for our analysis, we used the R package ggplot2 to help construct the graphs effectively and efficiently (Wickham 2016). This package is part of the packed package tidyverse, which we have used to help produce all the graphs and manipulate the data set (Wickham et al. 2019). Moreover, we have applied the readr package to help load and read the dirty data set and the clean data set for further data manipulation (Wickham, Hester, and Bryan 2022). Furthermore, we also used packages like knitr (Xie 2021b), tinytex (Xie 2021c), and bookdown (Xie 2021a) to enable the R markdown function and output this paper in a pdf format.

To visualize our data effectively, we have chosen the bar chart to demonstrate our data. We believe the bar graph is the best fit for this task because the graphs' primary purpose is to show the total number of observations in each answer and illustrate a clear trend in how respondents answered these questions. In our analysis, thirteen bar charts are created to help complete a full picture of the family characteristics and socio-economic conditions of Candian families. By looking at all these data and graphs together, we have discovered that most Canadians are married, born in Canada, from a larger urban population center, not a visible minority, gained citizenship by birth, finished at least high school and with a high chance of completing some higher education, living in a single detached home that the respondent or a family member owns with a smaller household size, work about 30 to 40 hours a week, earns about $25,000 to $49,999 as individual and $50,000 to $74,999 as a family. As for one of the major regions that this 'average' respondent might be residing in, Ontario will be the first choice, then the rest share a similar chance, with British Columbia being the lowest. However, one of the weaknesses of this data set is that a large number of respondents are from Atlantic Canada and the Prairie region of Canada, which raises a concern about the accuracy of the picture that this data set will form as the percentage of the population in these two regions in Canada are much less than the percentage of the observations they are representing in the data set. According to Statics Canada, the Atlantic Canada region only has about 6.3% of the total population in Canada (Canada 2022). Nevertheless, about 22% of the survey responses are from the Atlantic Canada region.

3 Survey

A survey was carried out to identify links with economic status according to the characteristics of various families. The survey was made up of 14 questions, including information about respondents (job, education, ethnicity), children (number of children, age, education investment), and property (ownership, number of properties).

Our survey questions are more based on practical reasons. On the basis of the raw data, we have planned and analyzed more details. For example, the details about children and houses are ignored in the raw data. But in reality, these two account for a large part of the family economy, so in the survey, we analyze the details from more perspectives. As in question 11, we speculate on the family's education level and economic

status by asking about the family's" Education investment." And by asking "age at the time of purchase of the first property" to speculate on the economic level of the respondents.

All possible answers are multiple-choice. The survey also aims to allow respondents to answer within five minutes to avoid survey fatigue.

In conclusion, the survey aims to see the link between economic status according the characteristics of families and provide direct results to the Canadian government.

4 Results

The results of our analysis show a general picture of what a family/individual's characteristics and socioeconomic conditions are like in Canada. They also give us a rough idea about the distribution of the level of socioeconomic conditions among Canadians. These graphs can help audiences understand the demographics of Canada.

In detail, Figure one reveals how many respondents were born in Canada and how many respondents were born outside of the country. As we know, Canada has one of the largest immigration populations in the world, and this graph can present us a gist of the percentage of immigration population in the nation. Furthermore, Figure two shows the major population regions where the respondents reside. This helps us understand the population distribution in Canada across the major population regions. However, as we evidenced previously in the data section, we think a bias might exist in receiving responses in this part of the survey. Therefore, further study or testing should be conducted regarding this issue. Nevertheless, the result indicates that the Ontario region has the largest population in Canada, the Atlantic region comes in second, and the Prairie region and Quebec come in third and fourth, while the British Columbia region comes in the least. Next, the third figure shows that most of the population lives in a larger urban population center, and about 20% of the population lives in rural areas or small population centers. Prince Edward Island accounts for the rest of the population, a tiny percentage of the total population.

Fourth, figure four shows that about half of the total respondents are married, and single or never married people come to the second-largest category, where couples who are currently separated share the smallest percentage of the population. Then, the fifth figure indicates that the majority of the respondents are not a visible minority in Canada. This helps us understand the level of diversity we are at from a wide national perspective. Figures six to eight show the respondents' citizenship status, the highest level of education received, and the ownership of current residence, respectively. The respondents' citizenship status indicates that the majority of the Canadians gained their citizenship by birth, a minority of Canadians are naturalized citizens, and a smaller percentage of the population does not have Canadian citizenship. By looking at the statistics on the highest level of education the respondents have received, we can conclude that most Canadians have had at least some experience in higher education and not many residents in Canada have an education level lower than high school or its equivalent. Next, figure 8 shows that most respondents have declared that their primary residence is either owned by themselves or a member of this household, which indicates that a high number of the population live in a family/self-owned property. Moreover, figures nine and ten further discuss the respondents' housing conditions. Figure nine demonstrates that most people live in single-detached houses in Canada, and the high-rise apartments have the lowest amount of people living there. This chart tells a story that most Canadians prefer or have the luxury to live in a single-detached house as opposed to many countries that mainly have apartment buildings as a primary residence.

Figure ten gives us an idea of what the household size is like in the country. Out of 20,602 respondents, 5,823 live alone, and 7,754 live with a partner or family member. This indicates that most households in Canada are of a smaller size. Figure eleven shows the average working hours of the respondents. Most people work from 30 to 40 hours a week on average, and 7,166 respondents chose not to answer. We believe this might indicate that many people in Canada are not comfortable about revealing how many hours they work. Furthermore, we also noticed that figure eleven is distributed as right-skewed. Therefore, we can put the picture that most respondents are working in the lower edger in terms of working hours, and the curve begins to go down after the peak.

Lastly, figures twelve and thirteen show that the median family and individual income are around $50,000 to $74,999 as family income and $25,000 to $49,999 as personal income. However, figure twelve reveals that the majority of the respondents' family income is in the 'more than $250,000' range, and the rest of the categories follow a right-skewed shape, with the '$25,000 to $49,999' range being the second peak, thus, the second most common income level. This indicates a considerable income gap between the higher-income families in Canada and lower-income families in Canada, which most of the families are in the lower-income category. Additionally, figure thirteen is significantly right-skewed, with the majority of the population earning a lower income. This shows a sign of emergency for the Canadian economy as this means there is not much real middle-class income earner as before. It is concerning that more than half of the population are in the two lowest income levels.

5 Discussion

This paper investigates and shows the overall situation of family or personal characteristics and socioeconomic status in Canada. It also conducts and analyzes the distribution of social and economic conditions of Canadians. The above charts can better help readers understand Canada's demographic data.

The above survey shows us many interesting phenomena and breaks our cognition. For example, Figure 8 shows that most respondents' primary residence is either owned by themselves or as a member of the family, which shows that a large number of people live in the family / own property. It's pretty different from the "housing misery" we saw in Toronto. In addition, figures 9 and 10 further discuss the housing characteristics of respondents. Figure 9 shows that most people in Canada live in independent single houses. Instead, Toronto's ubiquitous high-rise condos have the lowest number of occupants. These surveys shatter our old perception that most Canadians prefer or own single-family homes over apartment buildings.

In addition, the above survey also shows us a lot of economic fragmentation trends. For example, Figure 12 shows that the family income of most respondents is in the range of "more than 250000 dollars," of which the range of "25000 to 49999" is the second peak, so it is the second most common income level. We learned from this that there is a considerable income gap between high-income families in Canada and low-income families in Canada. It is not as small as we imagined. It can be said that most Canadian families belong to the low-income category. Furthermore, Figure 13 also tells us that the majority of the Canadian population has lower incomes. This shows that there is a real emergency in the Canadian economy, and we know that means there aren't many real middle-class sources of income. And to our surprise, more than half of the population is in the two lowest income levels.

The significant weaknesses come from the data set itself. For example, one weakness of the data set is that a large number of respondents come from the Atlantic region of Canada and the prairie region of Canada, which may lead to some deviation from the accuracy of the data set because the population percentage of these two regions is much lower than the percentage of observations they represent in the data set.

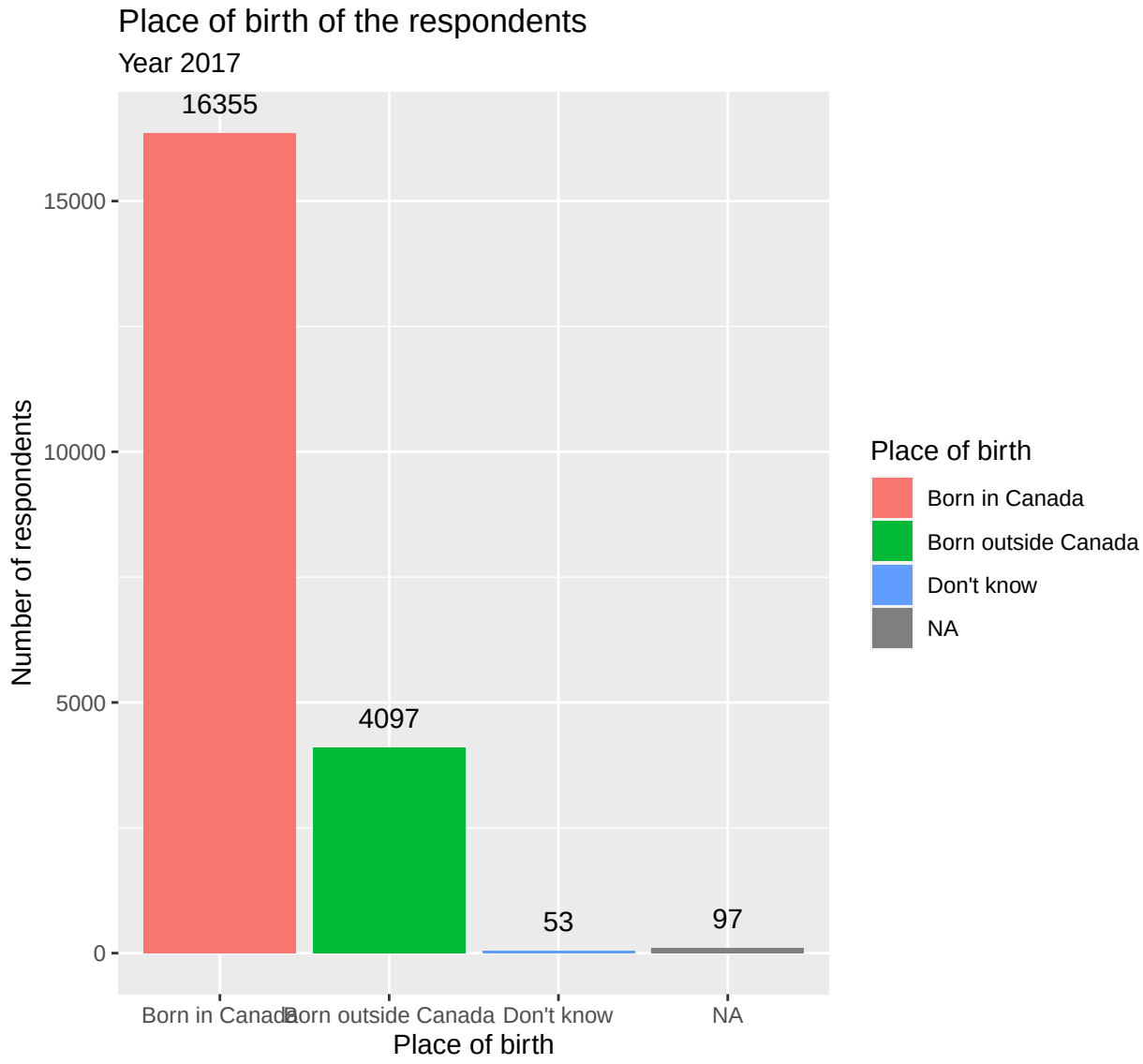
Besides, the survey itself has some limitations. In particular, given the random sampling method, given that our target group comes from all over Canada, the complexity of this data shows that it is impossible to represent all Canadian families.

Secondly, to simplify the analysis, the questionnaire has designed the answer scope and method in advance, making the respondents' answers more limited, and may miss some more detailed and profound information. For complex questions, simple answers can not obtain the required rich information.

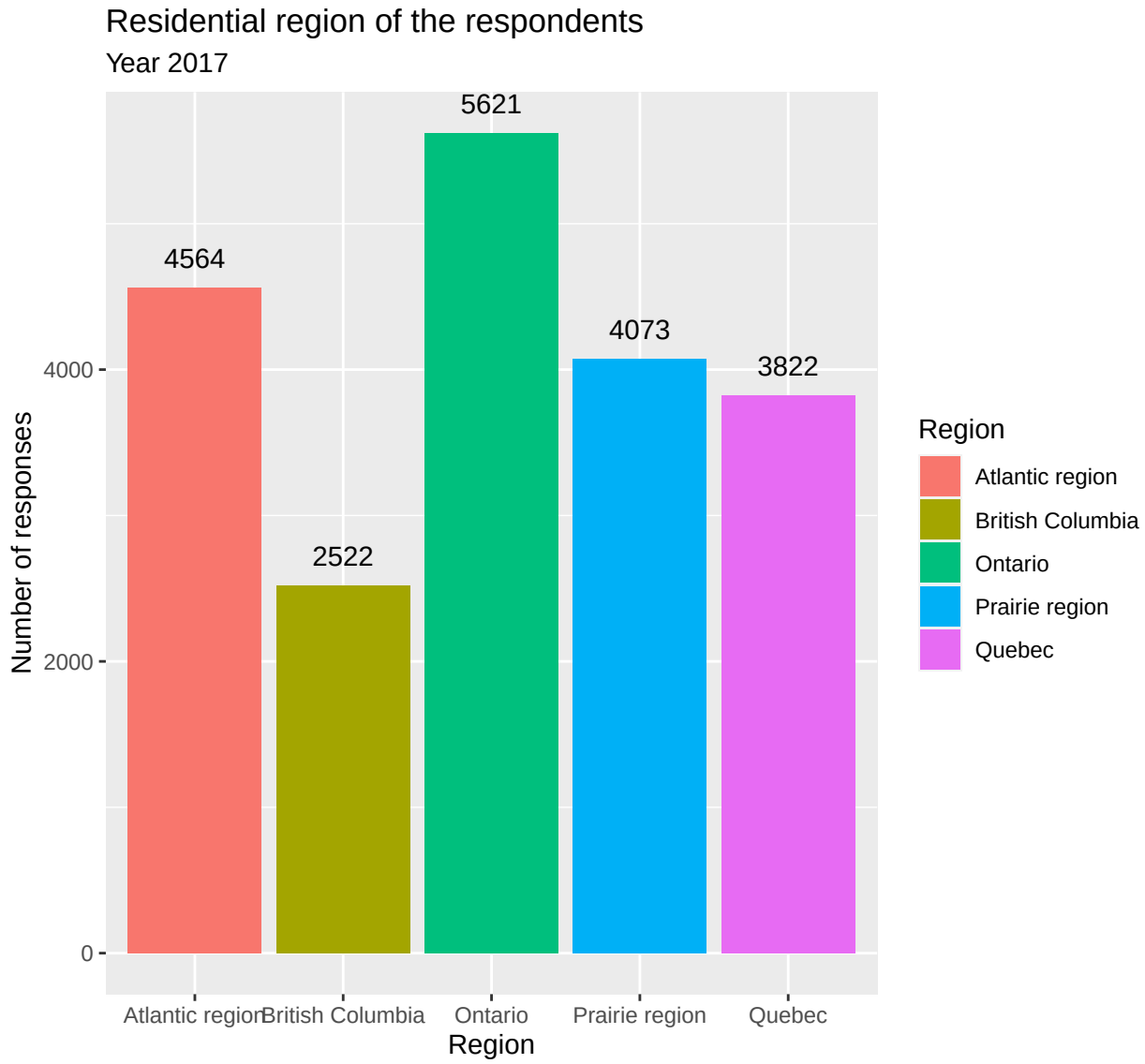
A future version of the survey can consider collecting information according to regions, which may be very helpful given the situation of different provinces and cities and the conditions of other families. Secondly, our survey only used multiple-choice questions. To have a more comprehensive understanding of the respondents' views, we can explore adding more flexible questions and incorporating them into future or separate analyses.

In the potential next step, we can start with the relationship between each variable in terms of data and explore the impact of each variable, such as the relationship between the respondent's occupation and his house purchase. Or perhaps the relationship between the number of children and family income is also very interesting.

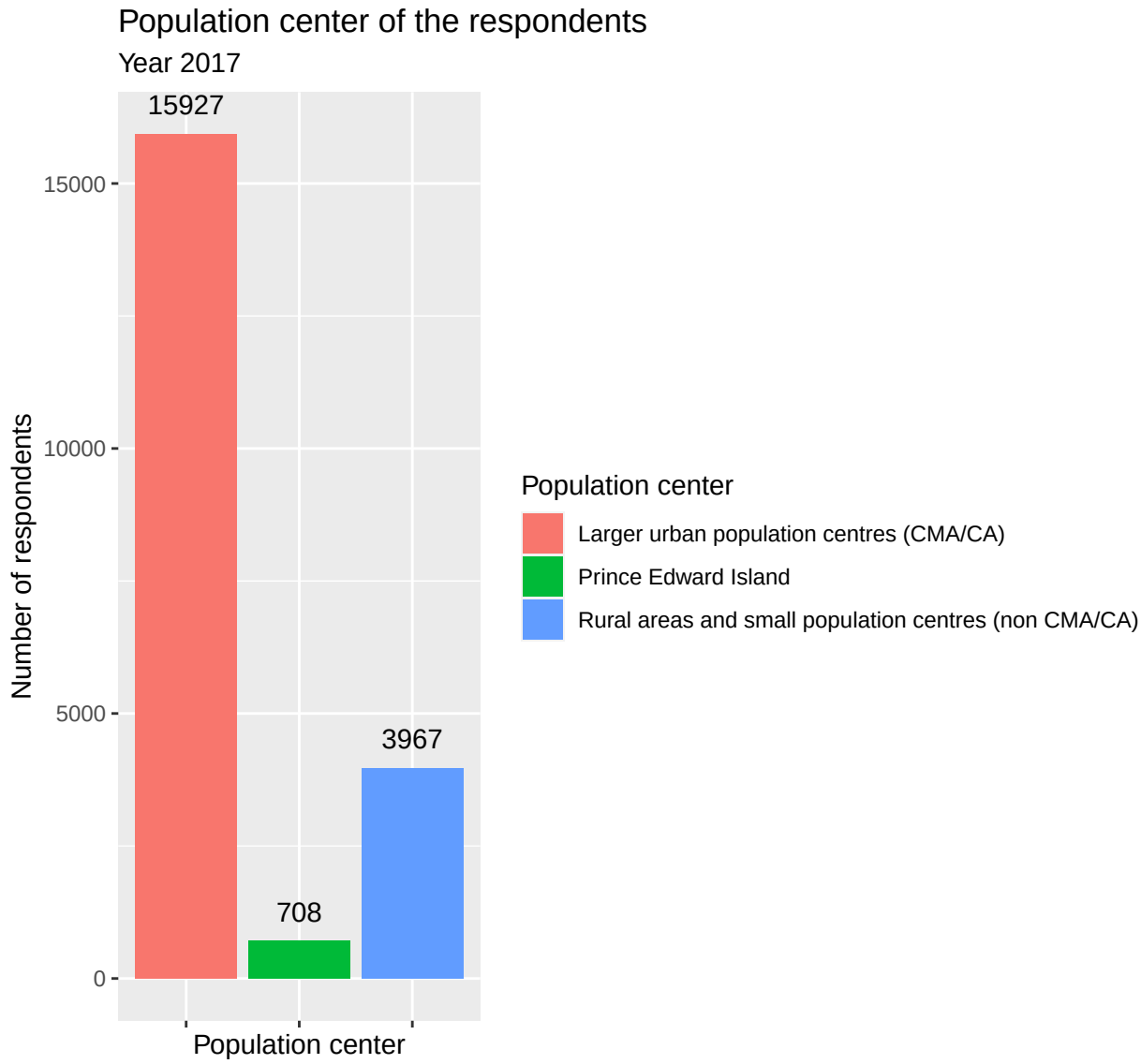
5.1 Figure one



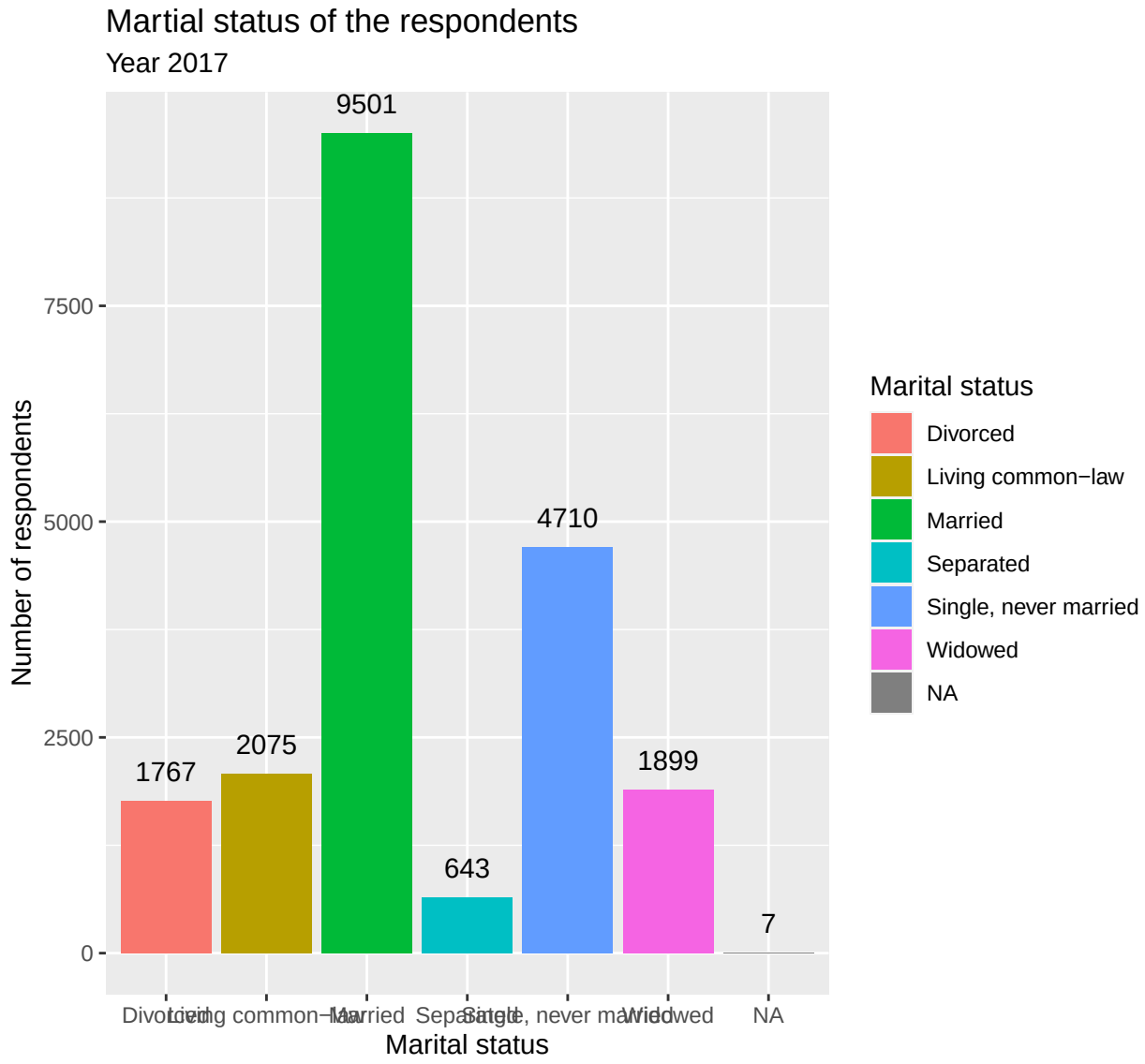
5.2 Figure two



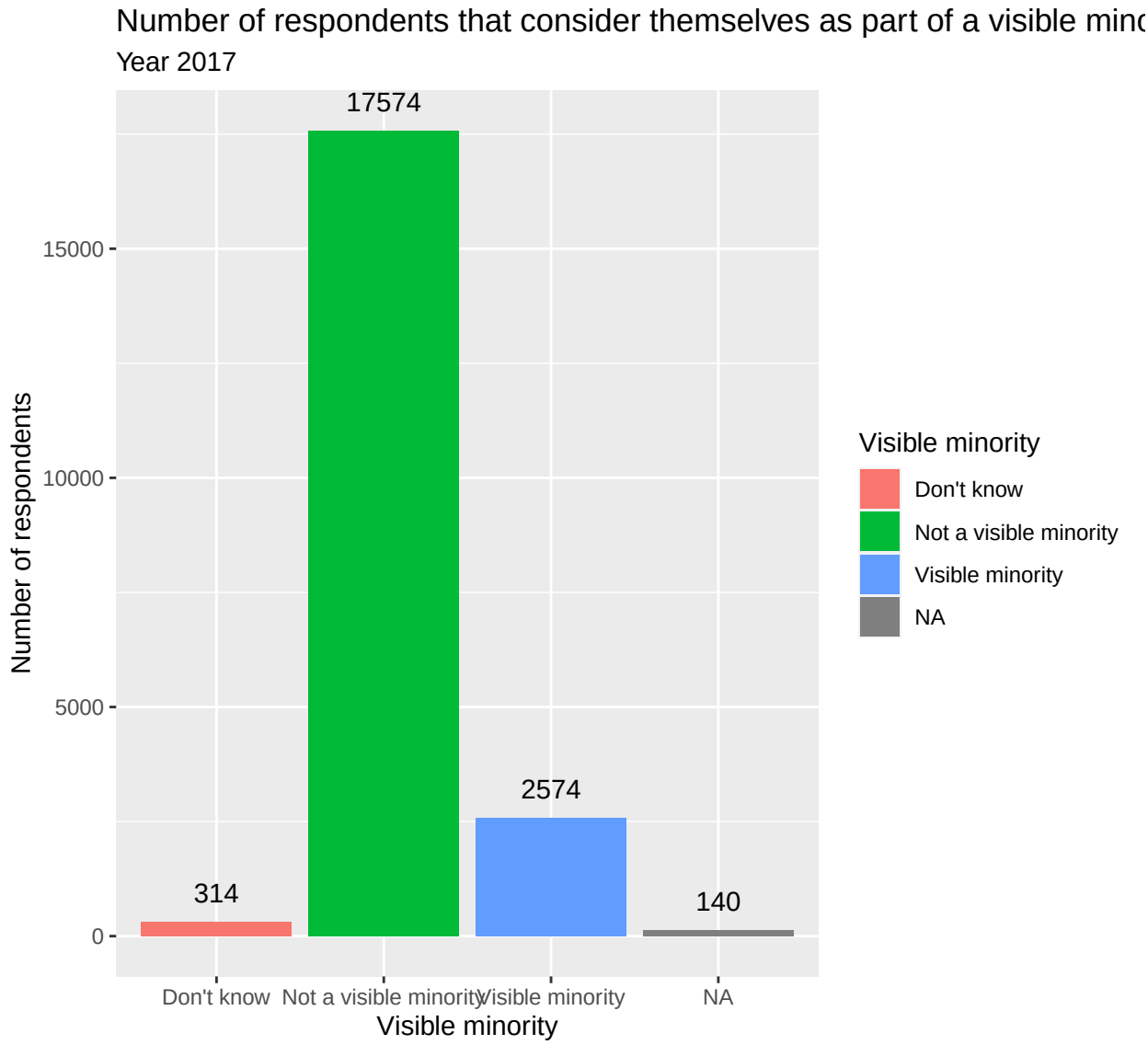
5.3 Figure three



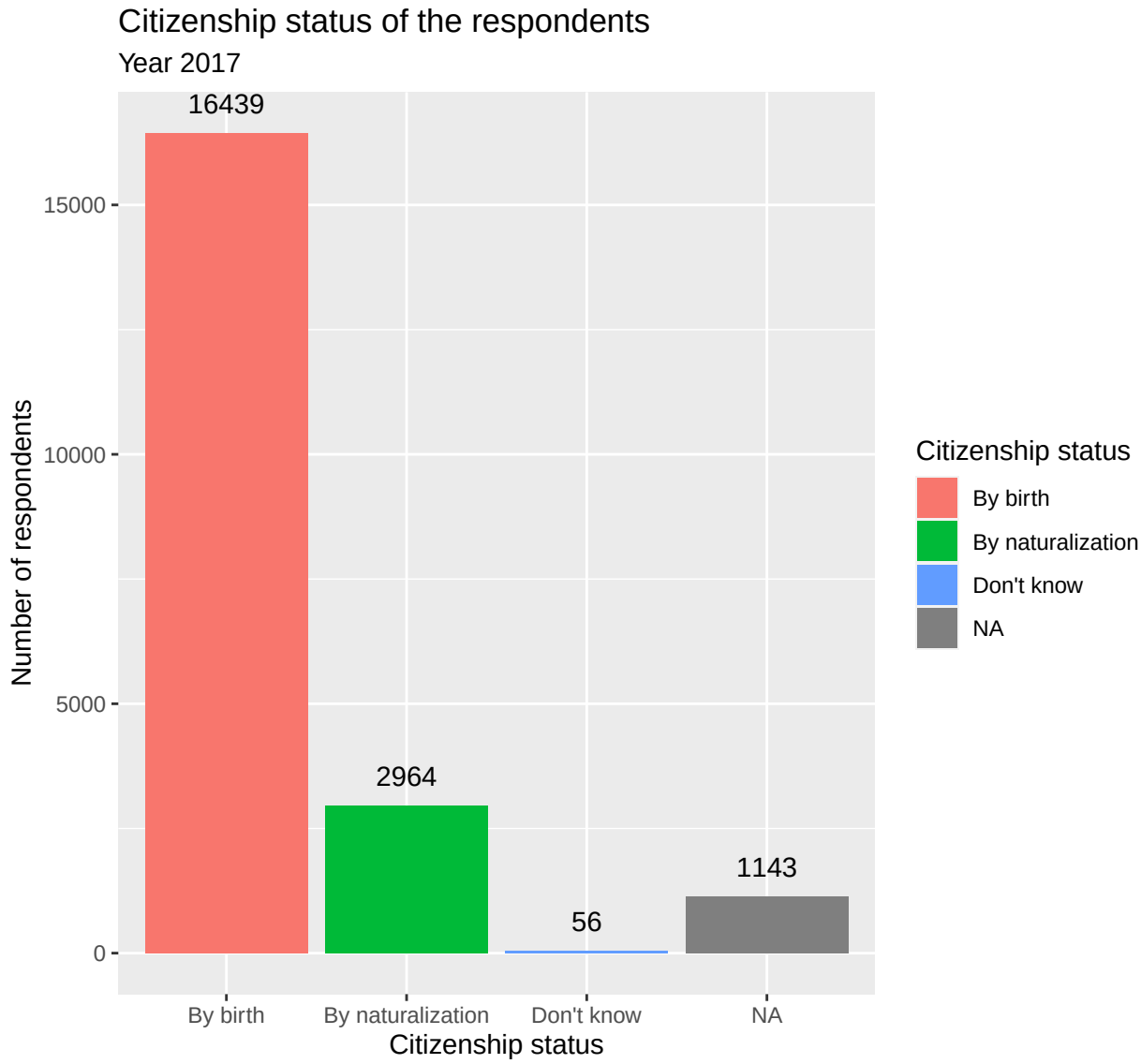
5.4 Figure four



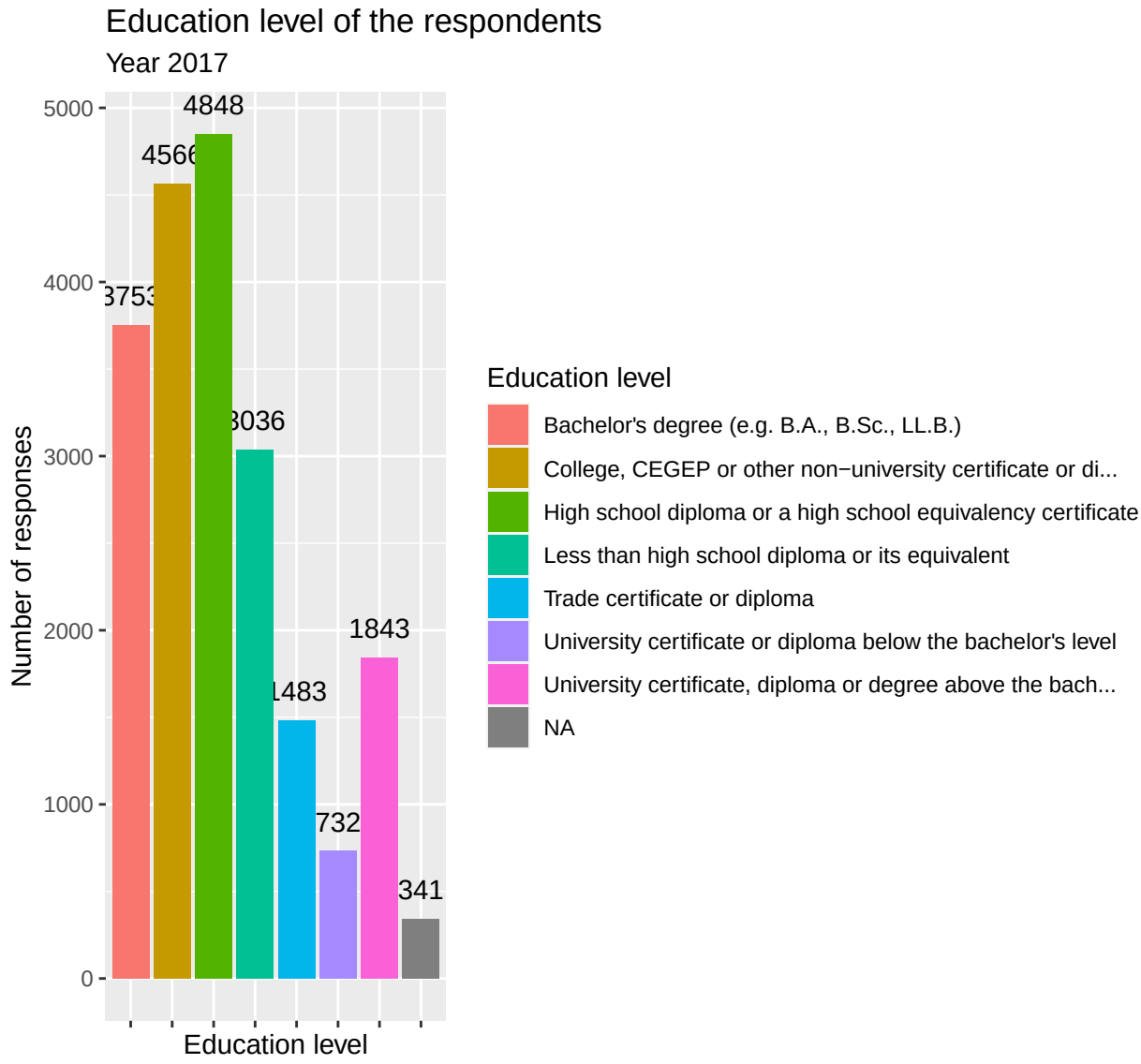
5.5 Figure five



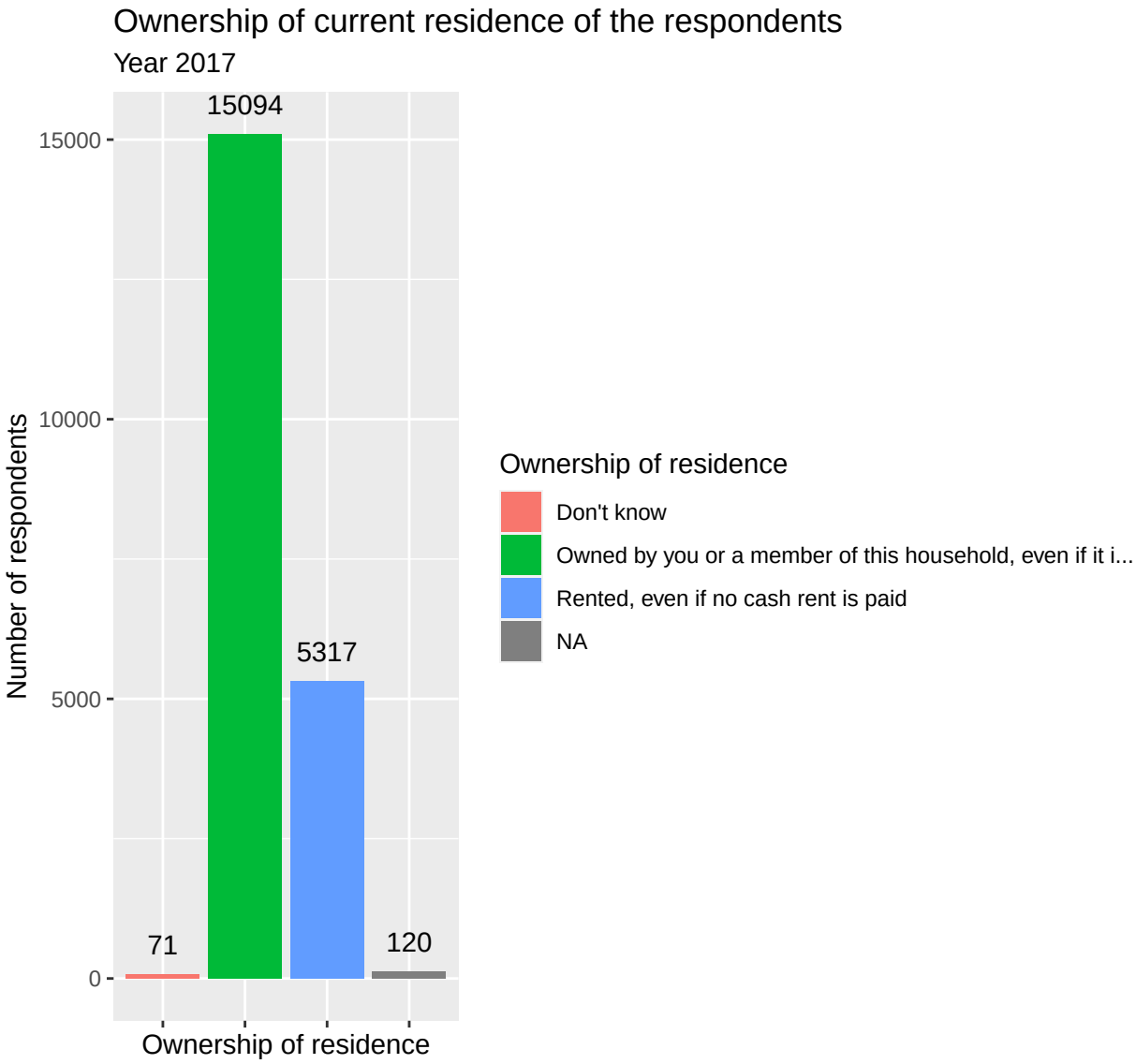
5.6 Figure Six



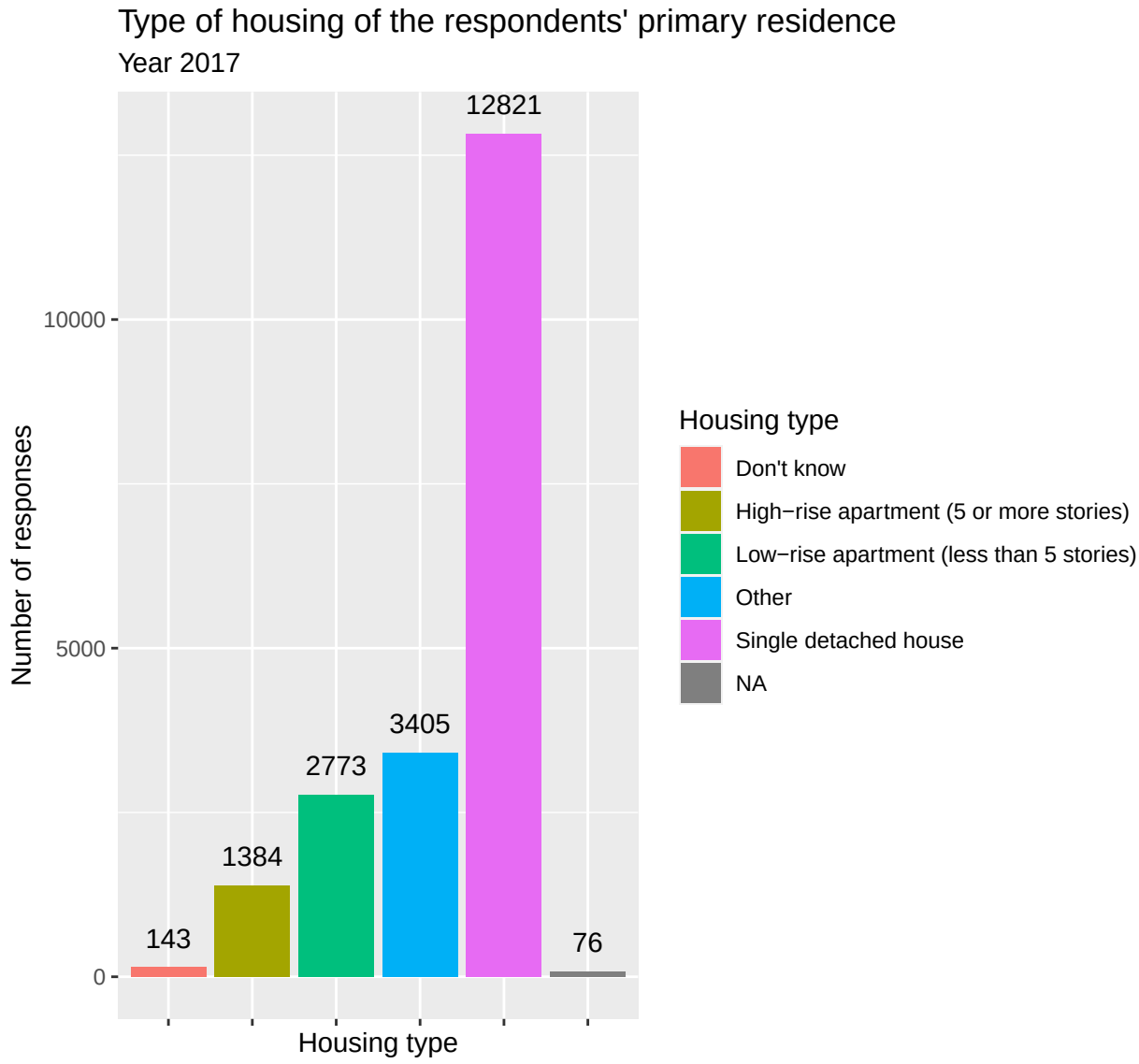
5.7 Figure Seven



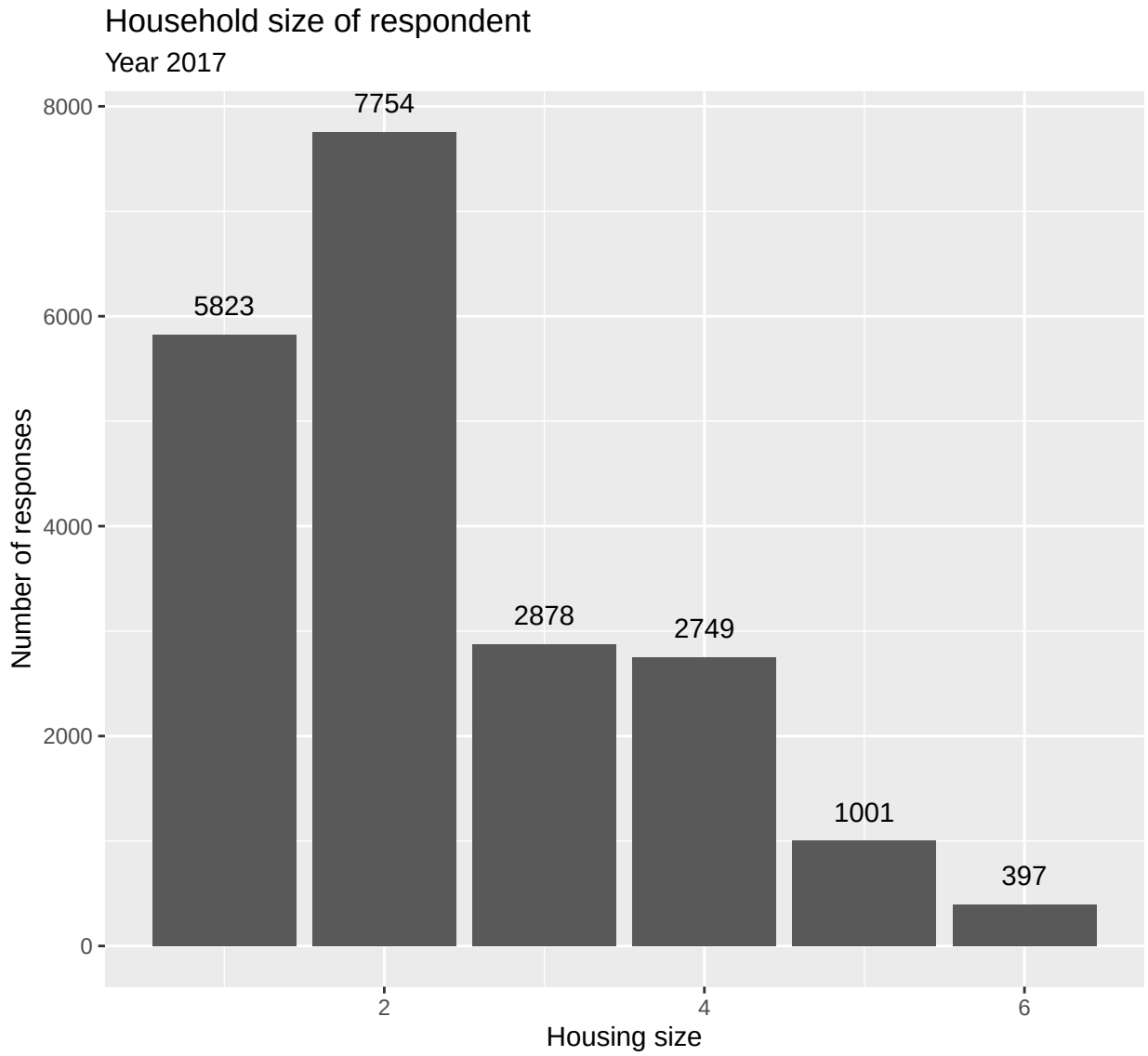
5.8 Figure Eight



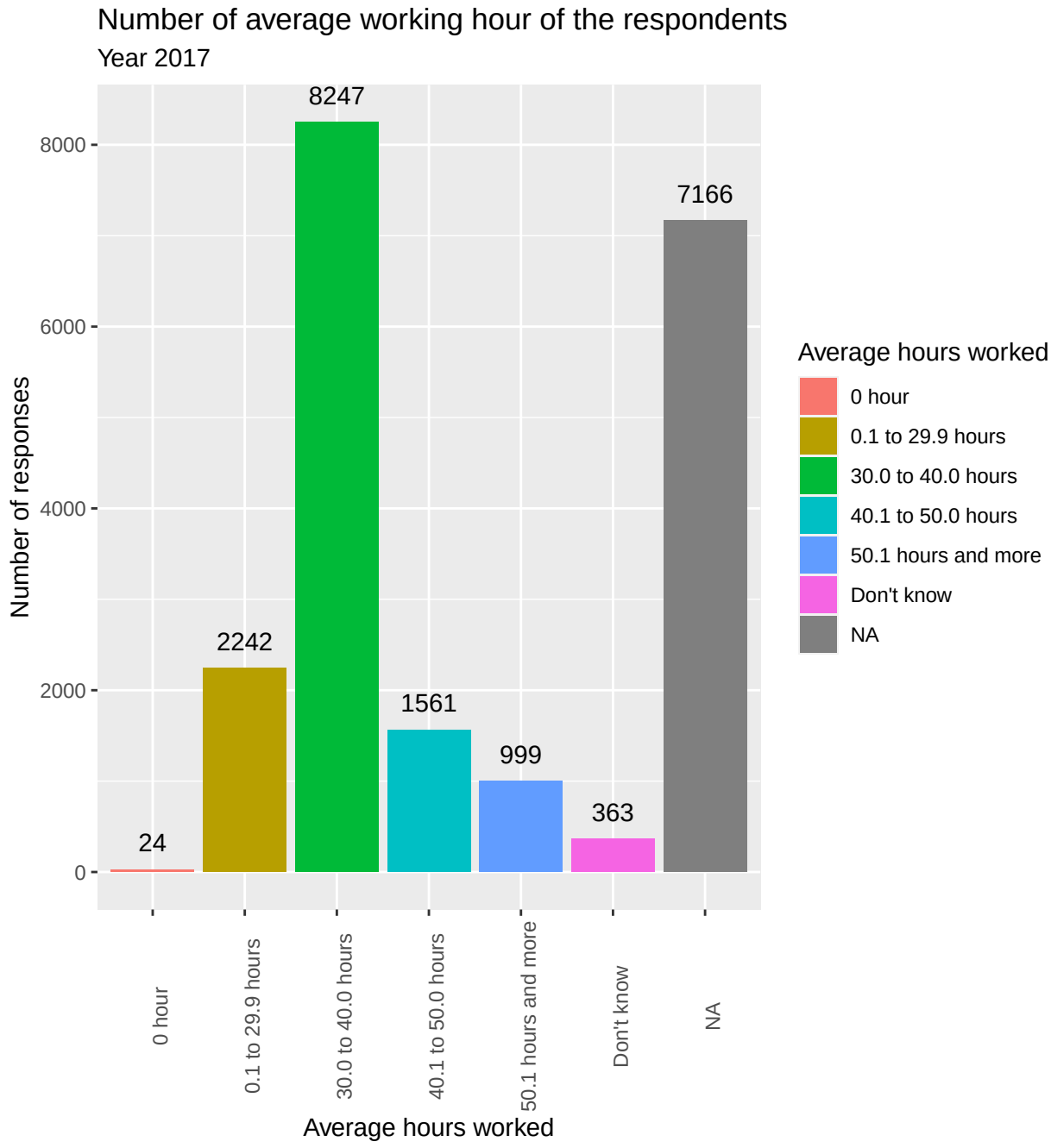
5.9 Figure Nine



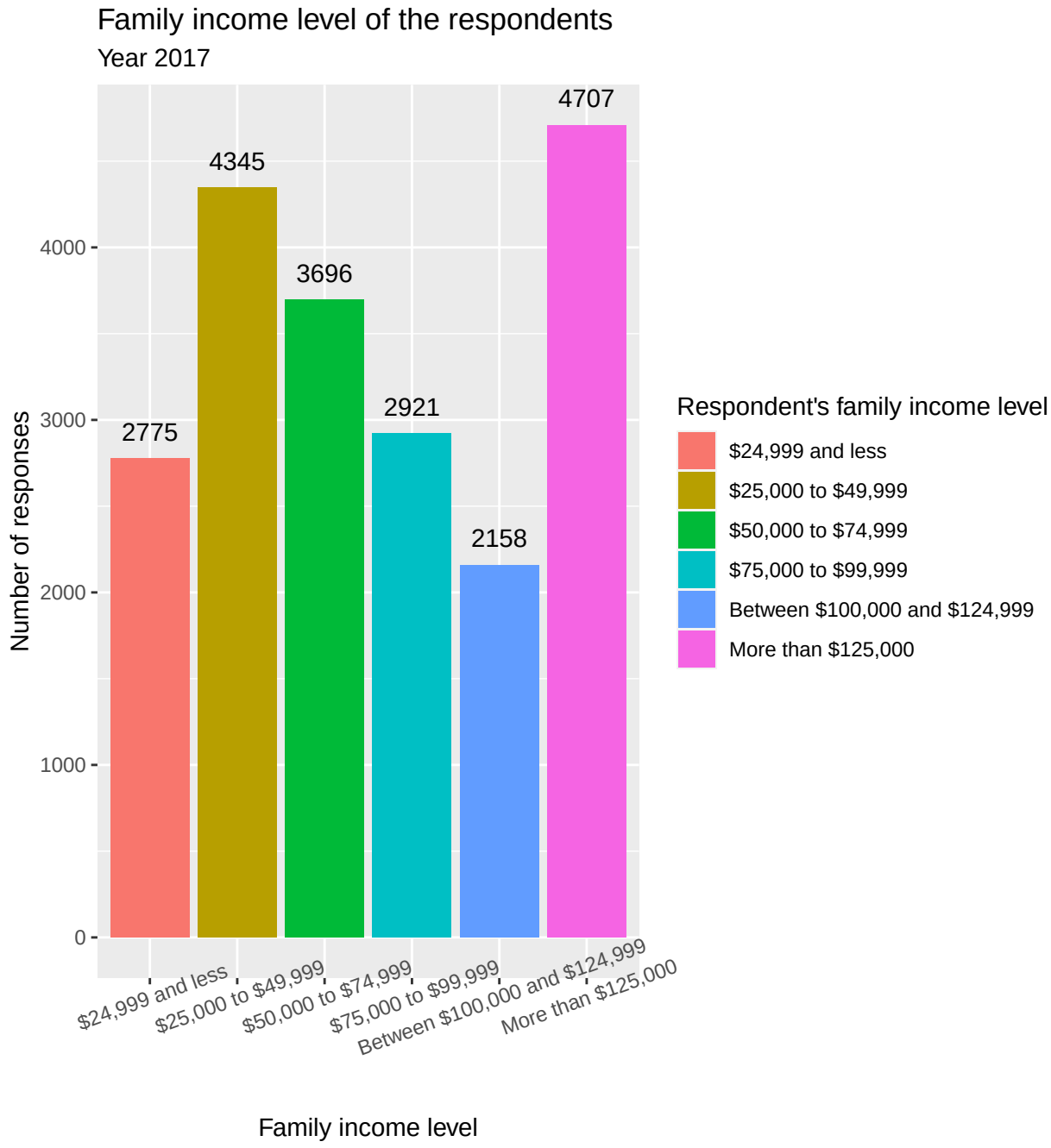
5.10 Figure Ten



5.11 Figure Eleven



5.12 Figure Twelve



5.13 Figure Thirteen

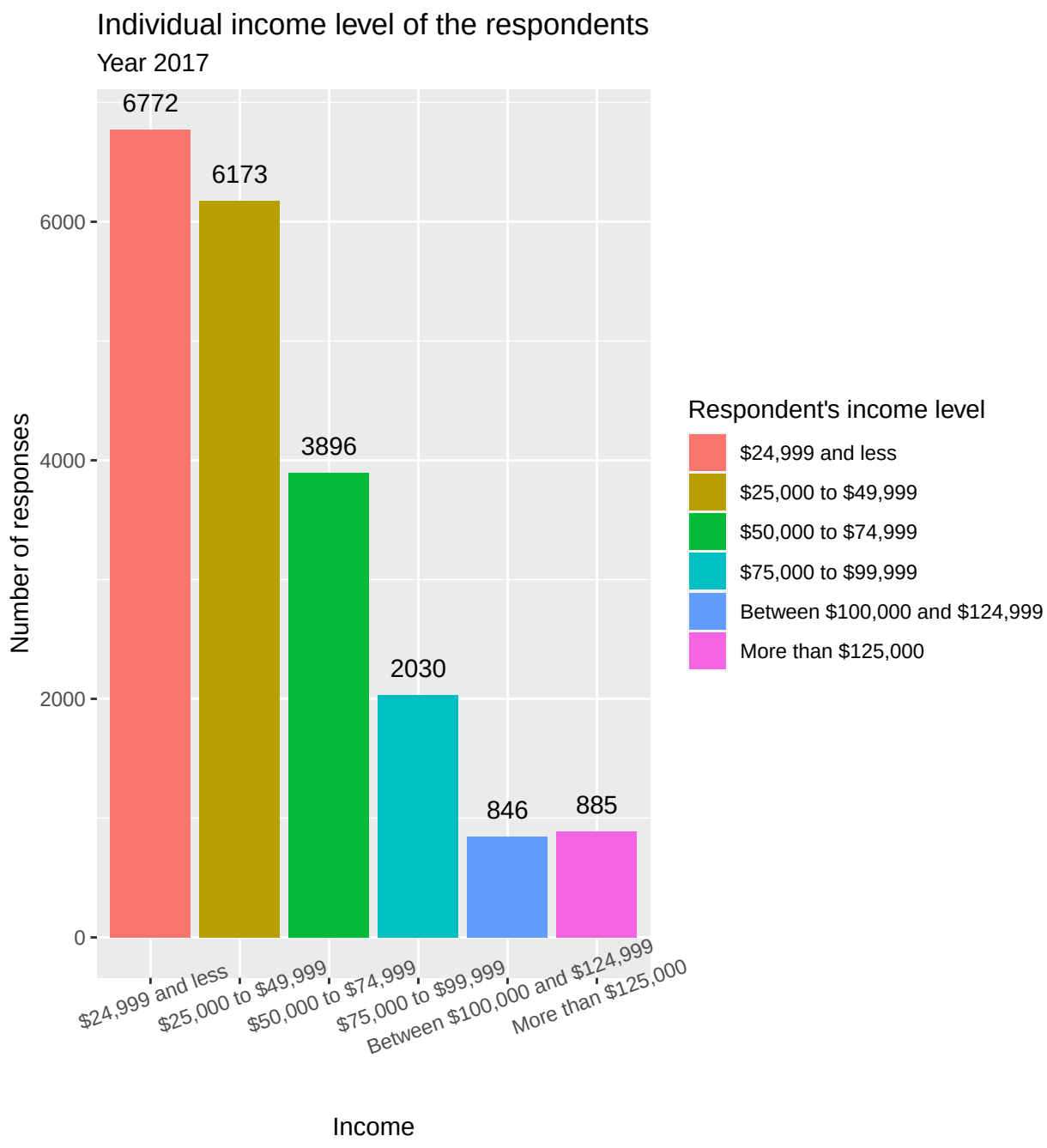


Figure 1: How much do the respondents earn as an individual?

References

Alexander, Rohan. 2020. *Clean-up the 2017 GSS Data*. https://piazza.com/class/kyby6xjrt5p10k?cid=17.

Canada, Statistics. 2017. *General Social Survey - Family (GSS)*. https://doi.org/https://sda-artsci-utoronto-ca.myaccess.library.utoronto.ca/cgi-bin/sda/hsda2?setupfile=harcsda4&datasetname=gss31&ui=2&action=subset.

———. 2022. *Population Estimates, Quarterly*. https://doi.org/10.25318/1710000901-eng.

Centrer, CHASS Data. n.d. *Microdata Accessed from CHASS*. http://dc.chass.utoronto.ca/myaccess.html.

Firke, Sam. 2021. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. https://CRAN.R-project.org/package=janitor.

R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

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

Wickham, Hadley, Jim Hester, and Jennifer Bryan. 2022. *Readr: Read Rectangular Text Data*. https://CRAN.R-project.org/package=readr.

Xie, Yihui. 2021a. *Bookdown: Authoring Books and Technical Documents with r Markdown*. https://github.com/rstudio/bookdown.

———. 2021b. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. https://yihui.org/knitr/.

———. 2021c. *Tinytex: Helper Functions to Install and Maintain TeX Live, and Compile LaTeX Documents*. https://github.com/yihui/tinytex.