Social factors that Affect Income in the United States*

Meha Grewal Danur Mahendra Shenghui Qiao

21 March 2022

Abstract

Canada's General Social Survey is a foundational program for providing information regarding social themes. The survey covers a variety of topics in-depth, specifically social identity which helps inform policy decisions and understand the broader context of Canada's diversity. Particularly, income is a highly significant topic of research as it can be influenced by a variety of social factors. Previous surveys primarily focused on identity and failed to investigate the impacts of identity and socioeconomic conditions. Class is focal to an individual's sense of self, thus a supplementary survey was done that examined income and its determining factors more thoroughly. The survey was distributed in the same manner as the General Social Survey but with the purpose of providing further information. Income is an indicator of economic inequality and information from these surveys allow informed policy decisions. Alongside, understanding the social identity of Canadians and their living conditions.

1 Introduction

The Canadian government's General Social Survey (GSS) (*The General Social Survey: An Overview* 2017) is a collection of annual independent, cross-sectional surveys that began in 1985. The survey examines many topics in detail and the purpose of the program is to observe the well-being and living conditions of Canadians. This contributes to informing the country about social trends and supporting policy decisions. This analysis was performed on R statistical software.

We believe that there are a variety of societal factors that impact an individual's income. This relationship can influence policy and the standard of living for Canadians. Our aim is to understand the existence of this association and what factors have the strongest effect on one's income. In 2013, a GSS survey on Social Identity was conducted to address current issues and provide information for social policy problems. The sample was based around stratified sampling that utilized probability sampling, and only included one person per household to be surveyed. Notably, this was also the first time Statistics Canada offered survey respondents the option to complete the questionnaire on the internet. There was an oversample of youth and immigrant communities to further examine those groups.

Previous GSS surveys pertaining to SI included information on political activities, volunteer work, values, and levels of trust in institutions to name a few. Whereas, the 2013 survey grew to also include matters regarding identity and belonging to national or ethnic culture groups. For future surveys, it would be beneficial to include more than one member of a household, as well as incorporating more questions regarding one's socioeconomic status.

We analyzed how different social factors impact the respondents incomes, and attempted to understand the broader context as to differing amounts of income. With survey data and data visualization we identified the notable social aspects that affect one's salary. These questions explore how identity can form a common set of beliefs that are followed by the majority of people. This survey reveals the societal patterns that dictate a

*Code and data are available at: https://github.com/qiaoshe1/STA304-Paper3.git

culture. This is significant for policy issues on diversity within Canada. The analysis will exhibit the extent to which social identity influences socioeconomic conditions.

Our survey's methodology is constructed and distributed similarly to the GSS. The questionnaire was about social identity, however there was a focus on income. Based on the survey, we conducted an analysis on a variety of factors that impact income. This was done through data visualization to show the influence of specific variables. There are limitations to our survey, such as only multiple choice questions, and only one member of households used as a respondent. This can skew results as one member may not be illustrative of the whole household, and there may have been a wider diversity of answers without limiting answer possibilities. Additionally, respondents may answer questions regarding income dishonestly.

R Core Team (2020)

2 Survery Methodology

Data for the 2013 General Social Survey (GSS) on Social Identity was collected between June 2013 to March 2014. The target population for this survey included all residents 15 years of age in the ten Canadian provinces and were not full-time residents of an institution. Institutional residents are defined by Statistics Canada as persons who, other than staff or family members, reside in an institutional collective dwelling such as hospitals, jails, or nursing homes. The survey used stratified sampling and stratas were created based on geographic location. A simple random sample without replacement would be used on the stratas to obtain samples used for this survey. It however, accounts for the small immigrant and youth populations by oversampling certain geographic areas (strata).

The survey frame consists of two components, a list of telephone numbers in use available to Statistics Canada, and the Address Register which lists all dwellings in the ten provinces. The two would then be used to make a single survey frame. The Address Register was first used to link all telephone numbers associated with a valid address. This constitutes 88% of our survey frame, meaning 12% of telephone numbers in use were not associated with a valid address. They were, however, used in the survey frame with each number accounting for a single record. In cases where more than one telephone number was associated with an address, landline telephone numbers were used as the preferred method to contact survey participants.

The target sample size for this survey was 31,973 however, only 27,695 samples were able to be collected for the purpose of this survey. Furthermore, 161 respondents were removed from the survey due confidentiality issues, bringing the number of samples down to 27,534. Respondents were initially contacted via telephone before being given the option to continue the survey using computer assisted telephone interviewing (CATI) or through an electronic questionnaire (EQ). Respondents who chose to do the questionnaire were given 21 days to complete the survey and those who failed to do so were transferred back to the CATI for a follow-up

3 Results

Based on questions pertaining to individual and family income (before tax) in GSS, we see that a high proportion of survey respondents has an individual pre-tax total income less than $49,000, but most respondents have a total family income of $125,000 and more.

The results of individual income are presented in figure 1, which shows a noticeably positive skew in income distribution. In fact, 32.8705950878555% of total respondents has a individual total "less than $25,000", and 29.9631103776332% has income between $25,000 and $49,999. While this may indicate a relatively low income level among individual survey respondents, figure 1 shows that many of these respondents, in contrast, have a high family income that is over $125,000 at the same time.

Intuitively, we examine possible effects of marital status and children on one's income level. A comparison between figure 3 and figure 4 indicates that while marital status doesn't have a noticeable effect on individual income level, respondents who have married tend to have a higher family income. 64.180790960452% of

respondents who have over $125,000 family income are married. Moreover, figure 5 makes a comparison between income levels based on respondents' number of children, number of children they intend to have in the future, and whether they have grandchildren or not. Most respondents have no children or grandchildren. There is no obvious evidence showing that respondents' choice of having a child or not is affected by their income level, or the number of children (or grandchildren) could be a factor influencing income.

Other social factors and their effects on individual income levels are checked. In figure 6, we see no significant difference in income level among different gender and age groups. The effect of immigration is discussed with figure 7 and figure 8. From these two figures, we first notice that a majority of respondents are born inside Canada, while the others' place of birth are categorized into five geographical regions. Looking at the immigrated respondents only (figure 8), we see that Asia and Europe are the top two regions of immigration to Canada. While most respondents who are born outside Canada earn a relatively low income which is less than $49,000, this fact doesn't give us any information of whether the place of birth provides any additional advantage or disadvantage on earning a high income.

Figure 9 and figure 10 compare respondents' income and their self-rated condition and life satisfaction level, respectively. In figure 9, we notice a weak relation between health and income level. With income less than $25000, between $25,000 and $49,999, and between $50,000 and $74,999, most people rate their health level as 'very good' or 'good', while more people with higher income level consider their physical and mental health level as 'excellence'. As income increases, the proportion of respondents in each income group who chooses 'excellence' for health level also increases, and the proportion of those who chooses 'poor' decreases. For respondents with an income higher than $125,000, 'excellence' in physical and mental health is the most popular response. The level of life satisfaction, as indicated by figure 10, doesn't seem to have a relation with people's income level. Regardless of income, most respondents are satisfied with their life.

Table 1 and figure 11 indicate the effect of education level on income. Based on the table, as income level increases, the number of respondents who have a university diploma below bachelor's degree or less (high school diploma, and less than high school diploma) decreases correspondingly. And it is not surprising to notice that among respondents who earn the highest income ($125,000 or more), the number of people who have a university diploma above bachelor's degree is also the largest.

The effect of occupation is presented by table 2 and figure 12. The income level of respondents, however, doesn't seem to have a large difference among different occupations.

Lastly, we looked at the effect of religion on income level. Figure 13, 14 and 15 shows respondents' income level and their religion affiliation, importance and participation respectively. We see that most people have religious affiliation, but the majority of respondents do not necessarily consider religion important or participate in religious activities a lot. And the religion does not seem to have an obvious effect on the respondents' income level, as indicated by the figures.

Table 1: Individual Income and Education Level

education	Less than $25,000	$25,000 to $49,999	$50,000 to $74,999	$75,000 to $99,999	$125,000 and more	NA
Bachelor's degree (e.g. B.A., B.Sc., LL.B.)	465	654	693	587	229	240
College, CEGEP or other non-university certificate or di...	647	1169	863	351	120	141
High school diploma or a high school equivalency certificate	895	1097	573	248	69	93
Less than high school diploma or its equivalent	526	381	141	48	19	18
Trade certificate or diploma	198	328	261	124	58	65
University certificate or diploma below the bachelor's level	57	145	122	90	24	32
University certificate, diploma or degree above the bach...	155	236	276	307	235	153

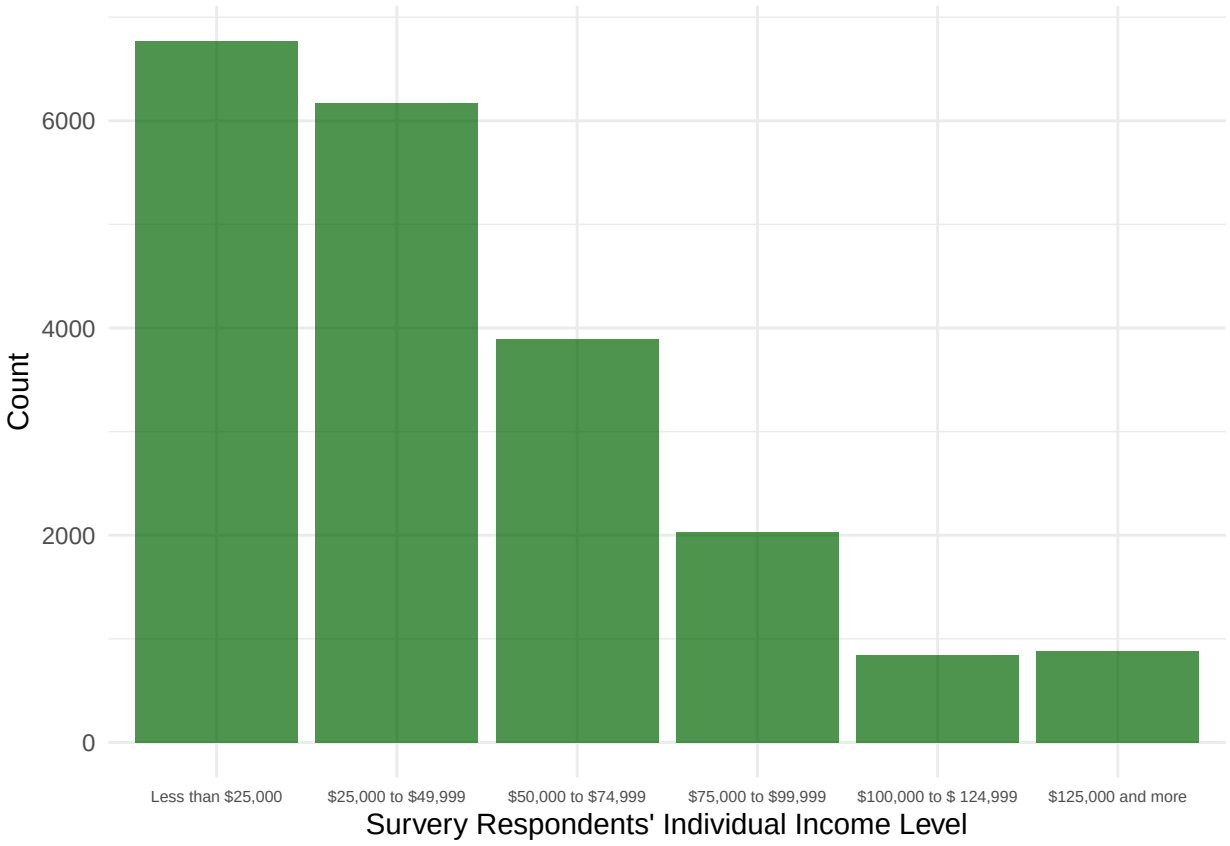

Figure 1: Individual Total Income (Before Tax)

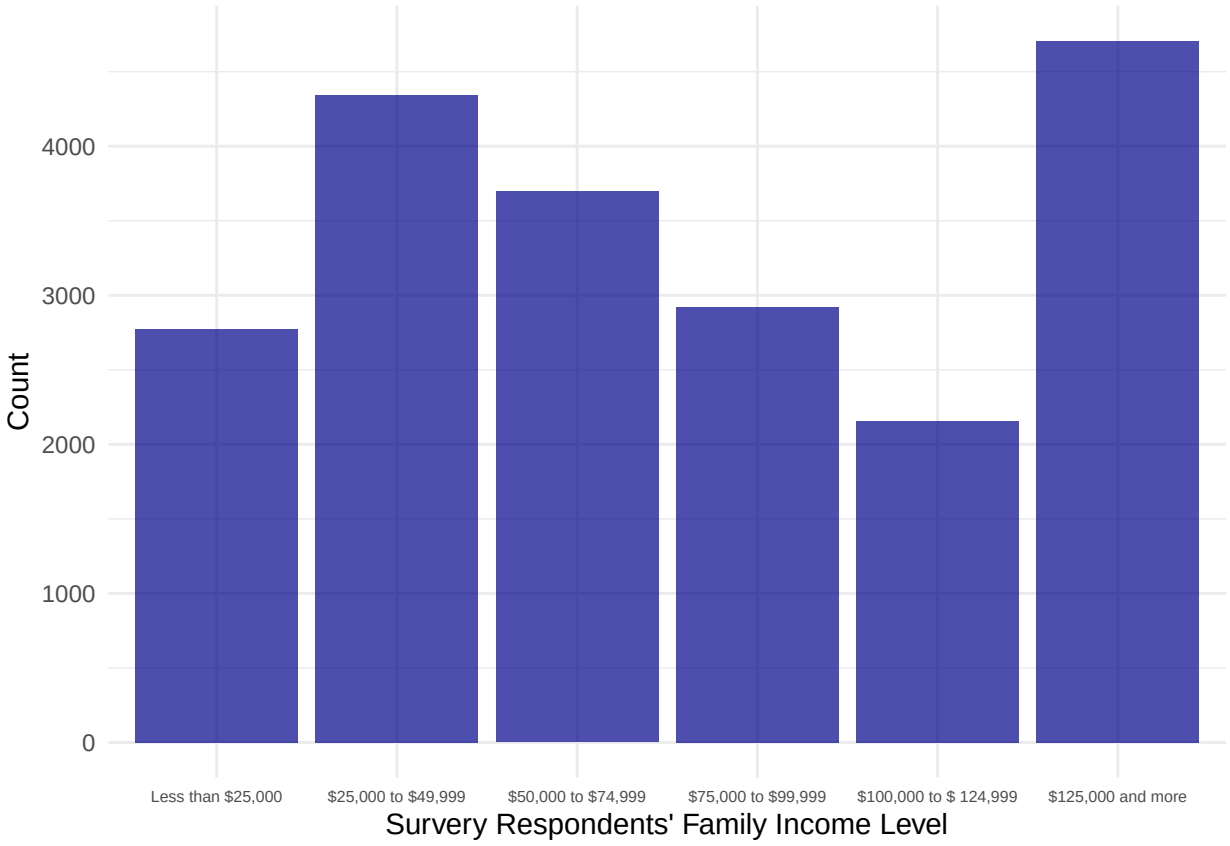

Figure 2: Family Total Income (Before Tax)

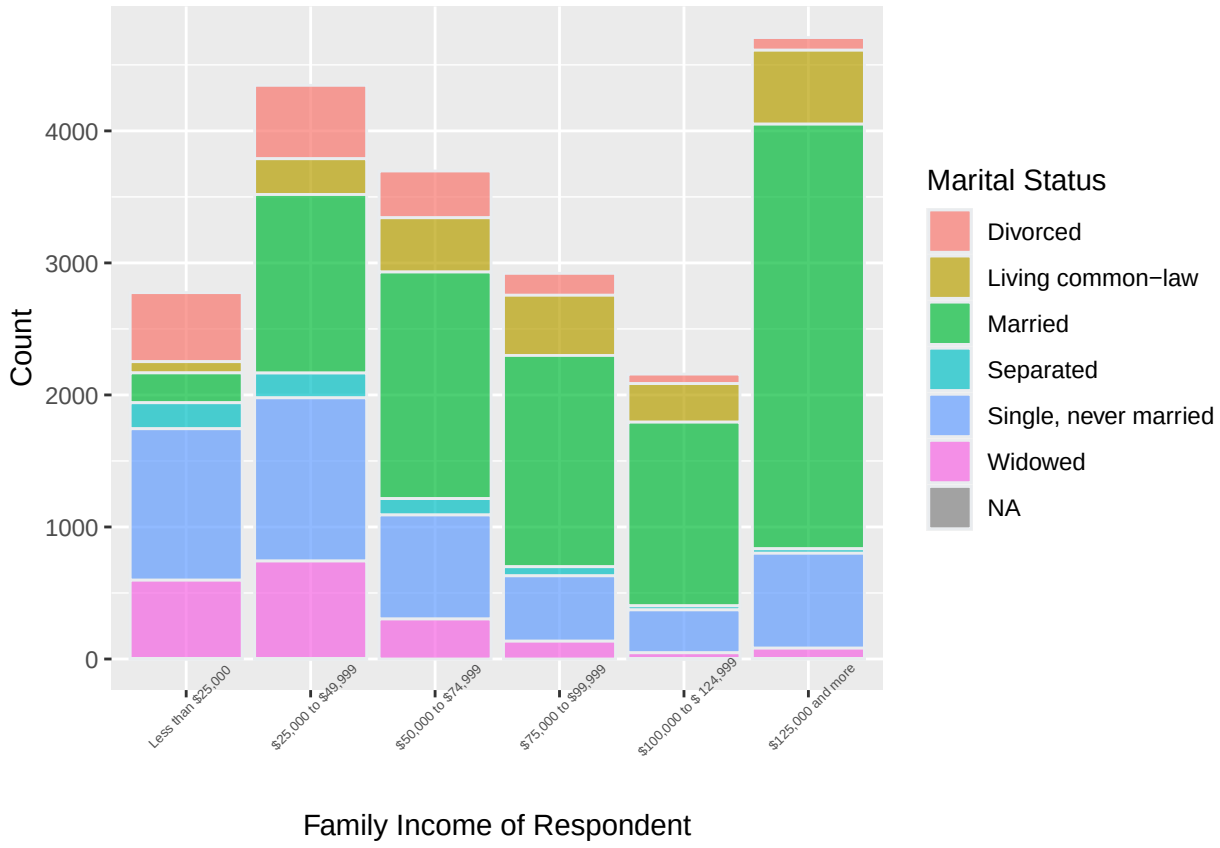

Figure 3: Family Total Income Across Marital Status

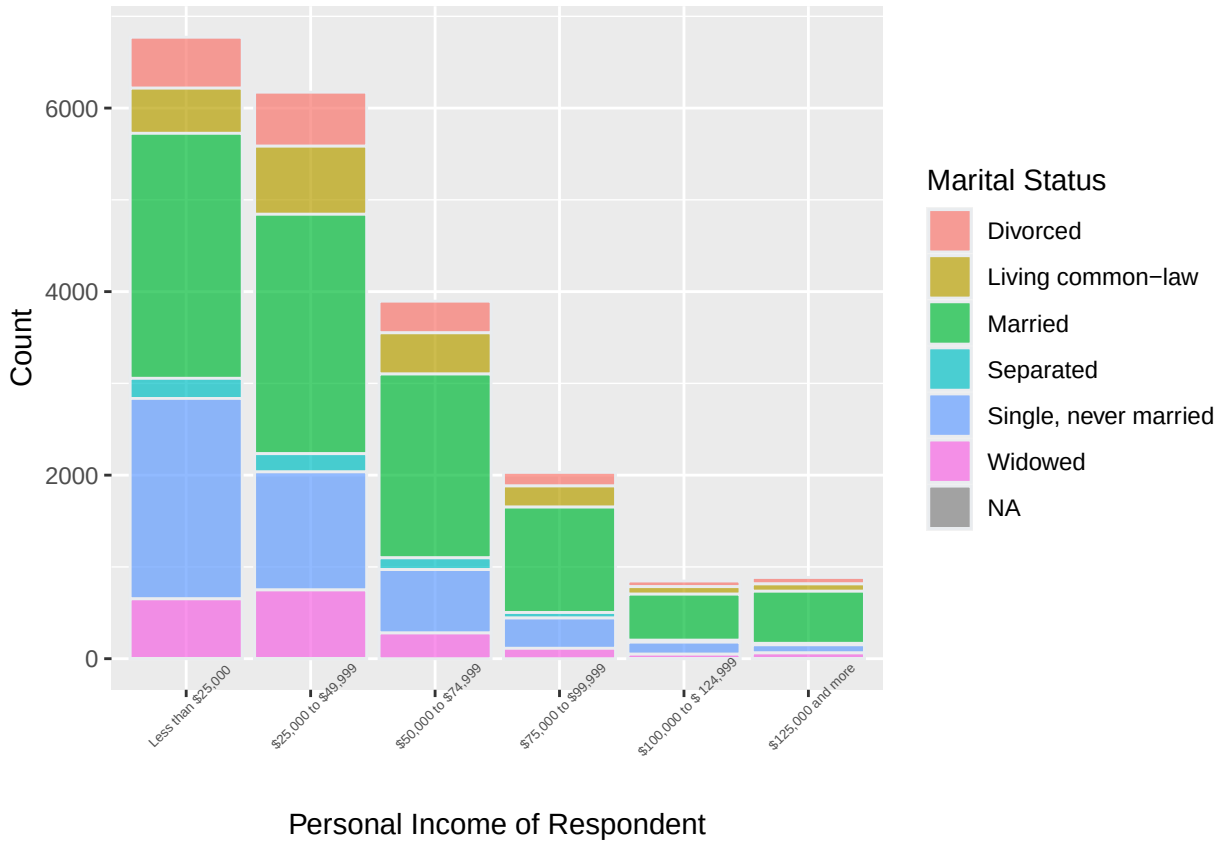

Figure 4: Individual Total Income Across Marital Status

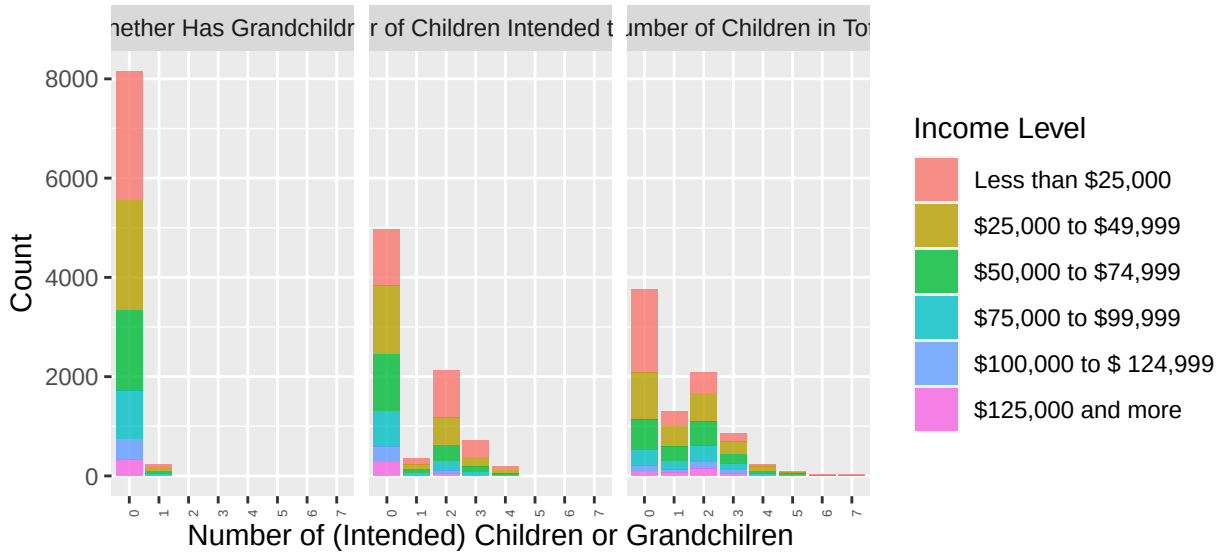

Figure 5: Individual Total Income and Children

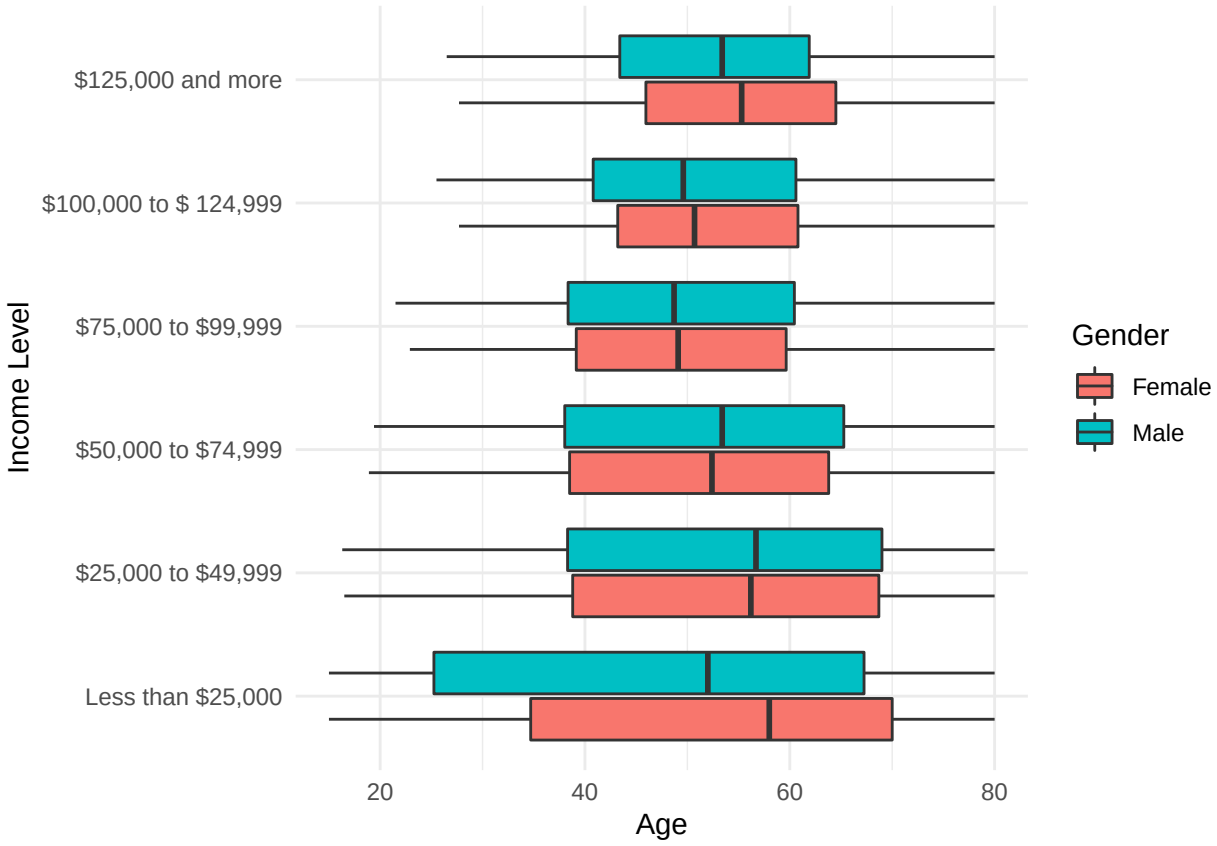

Figure 6: Individual Income, Gender, and Age

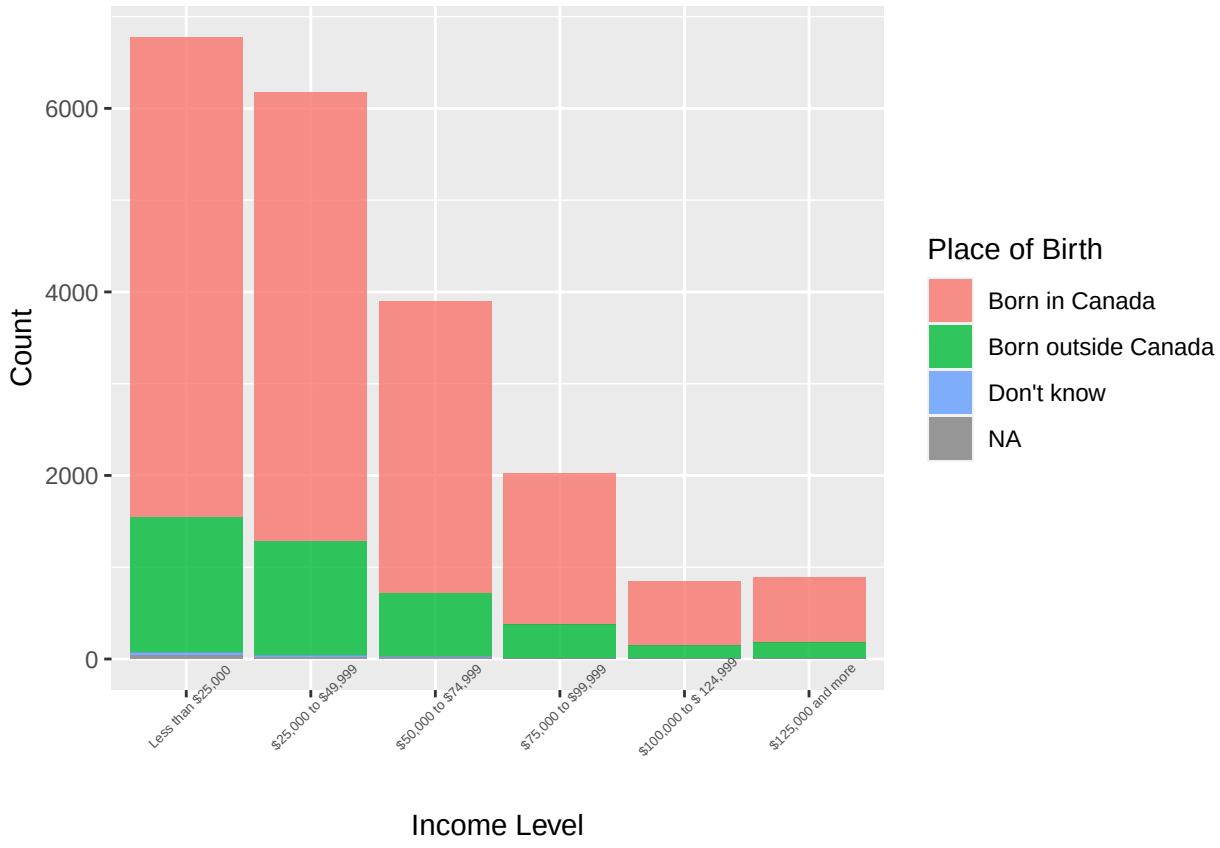

Figure 7: Individual Income and Place of Birth

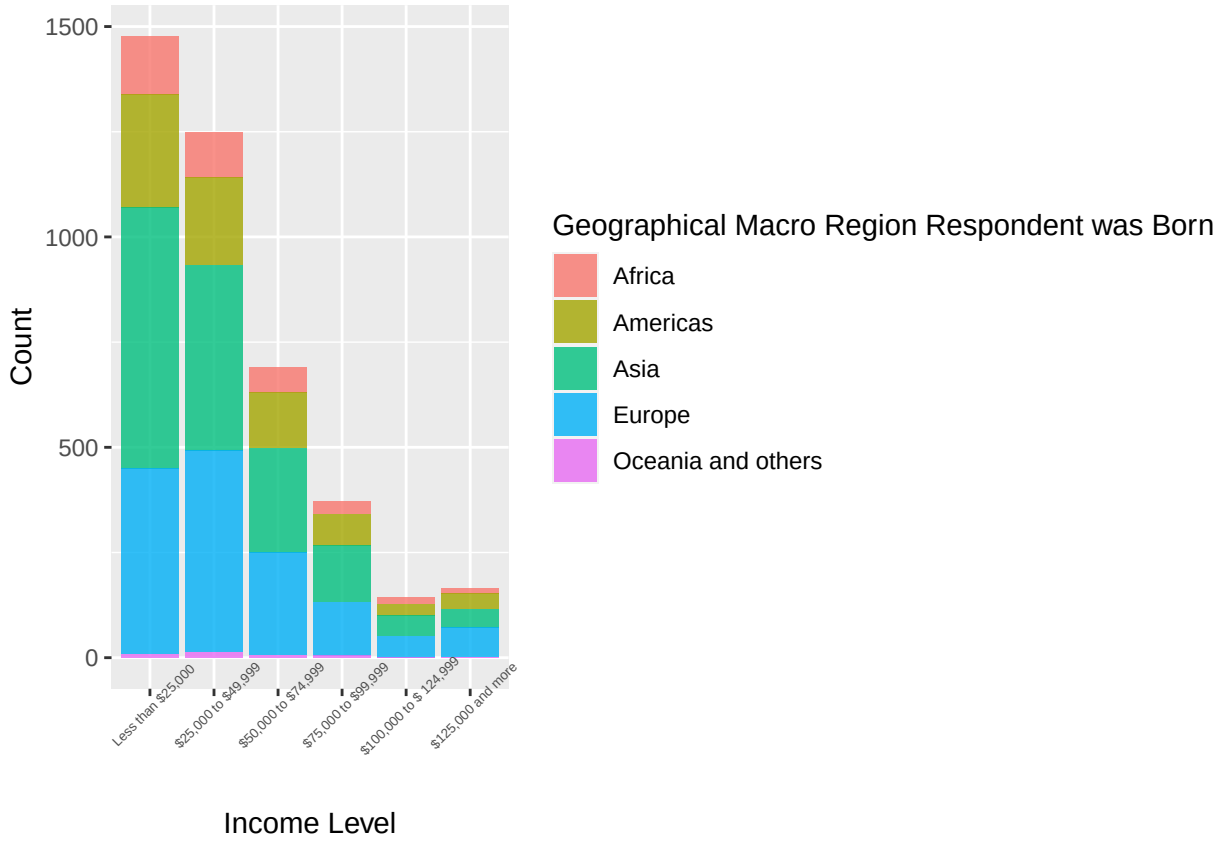

Figure 8: Individual Income of Respondents Born Outside Canada

Figure 9: Individual Income and Self-rated Health Level

Figure 10: Individual Income and Life Satisfaction Level

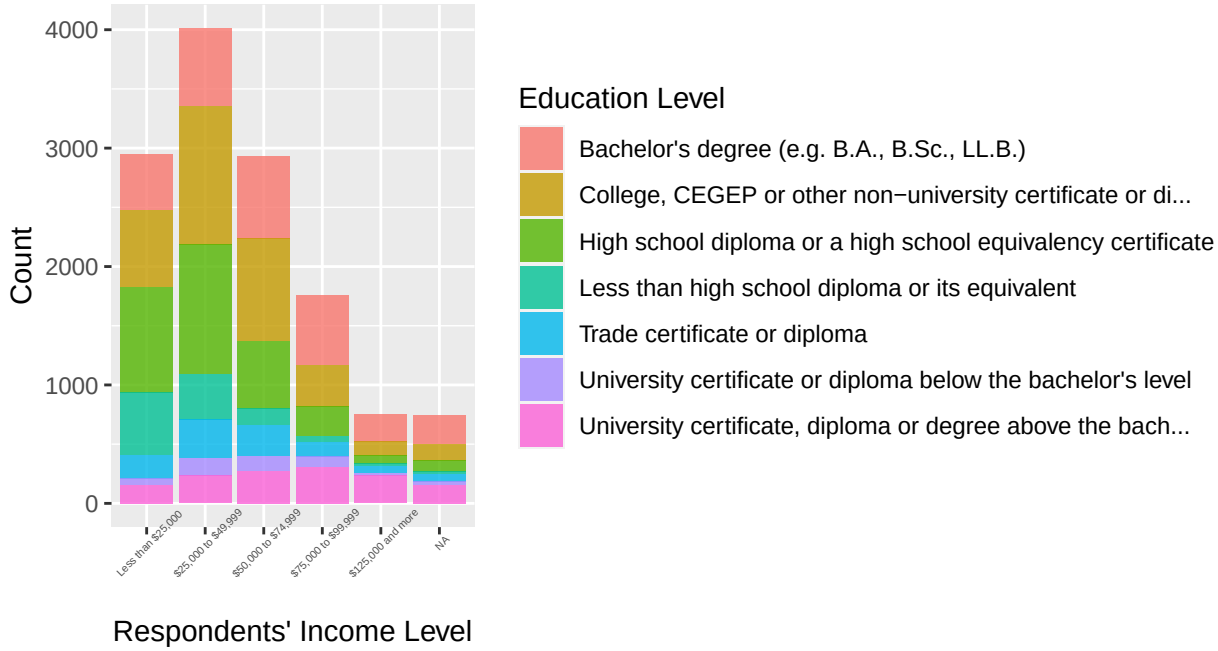

Figure 11: Individual Income and Education Level

Table 2: Individual Income and Occupation

occupation	Less than $25,000	$25,000 to $49,999	$50,000 to $74,999	$75,000 to $99,999	$125,000 and more	NA
Business, finance, and administration occupations	304	771	640	262	118	89
Health occupations	128	288	290	194	58	62
Management occupations	205	272	219	214	203	115
Natural and applied sciences and related occupations	113	187	263	236	109	122
Natural resources, agriculture and related production oc...	97	114	48	30	22	19
Occupations in art, culture, recreation and sport	210	115	74	29	9	8
Occupations in education, law and social, community and ...	312	428	447	384	89	151
Occupations in manufacturing and utilities	80	212	96	38	17	21
Sales and service occupations	1171	1084	378	126	65	50
Trades, transport and equipment operators and related oc...	297	519	459	237	61	103
Uncodable	26	20	15	5	3	2

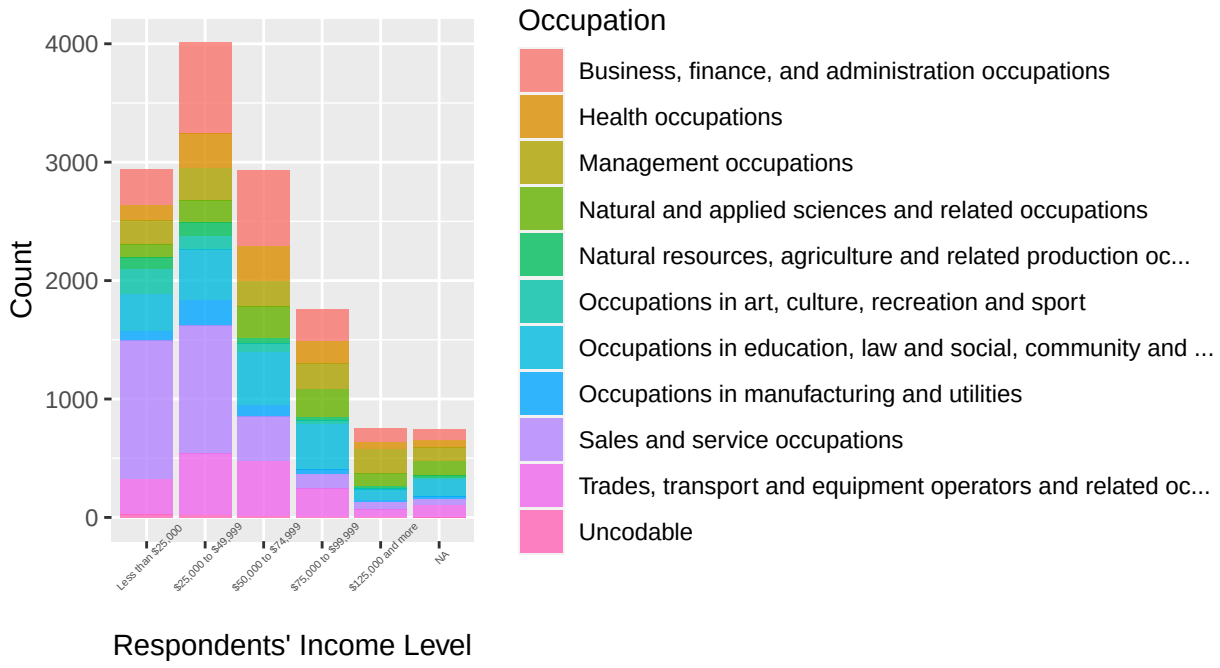

Figure 12: Individual Income and Occupation

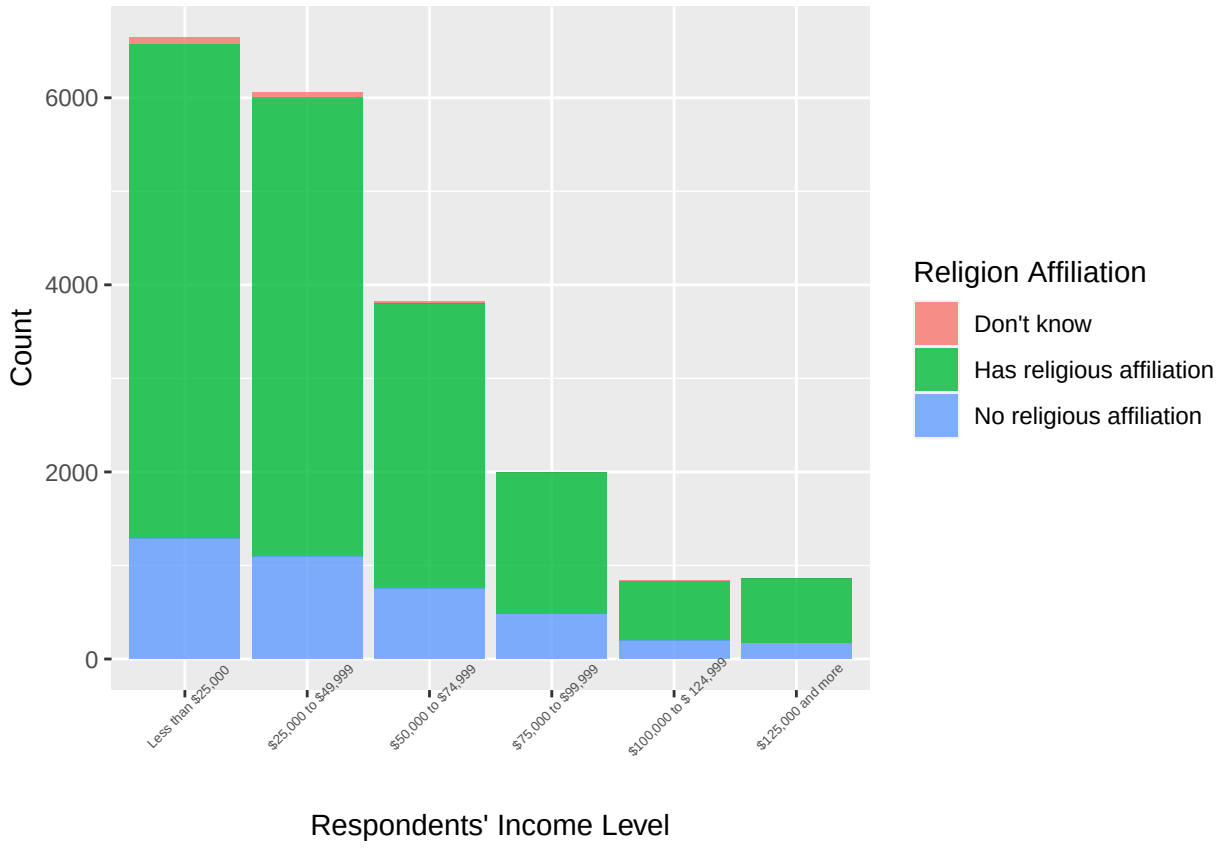

Figure 13: Individual Income and Religion Affiliation

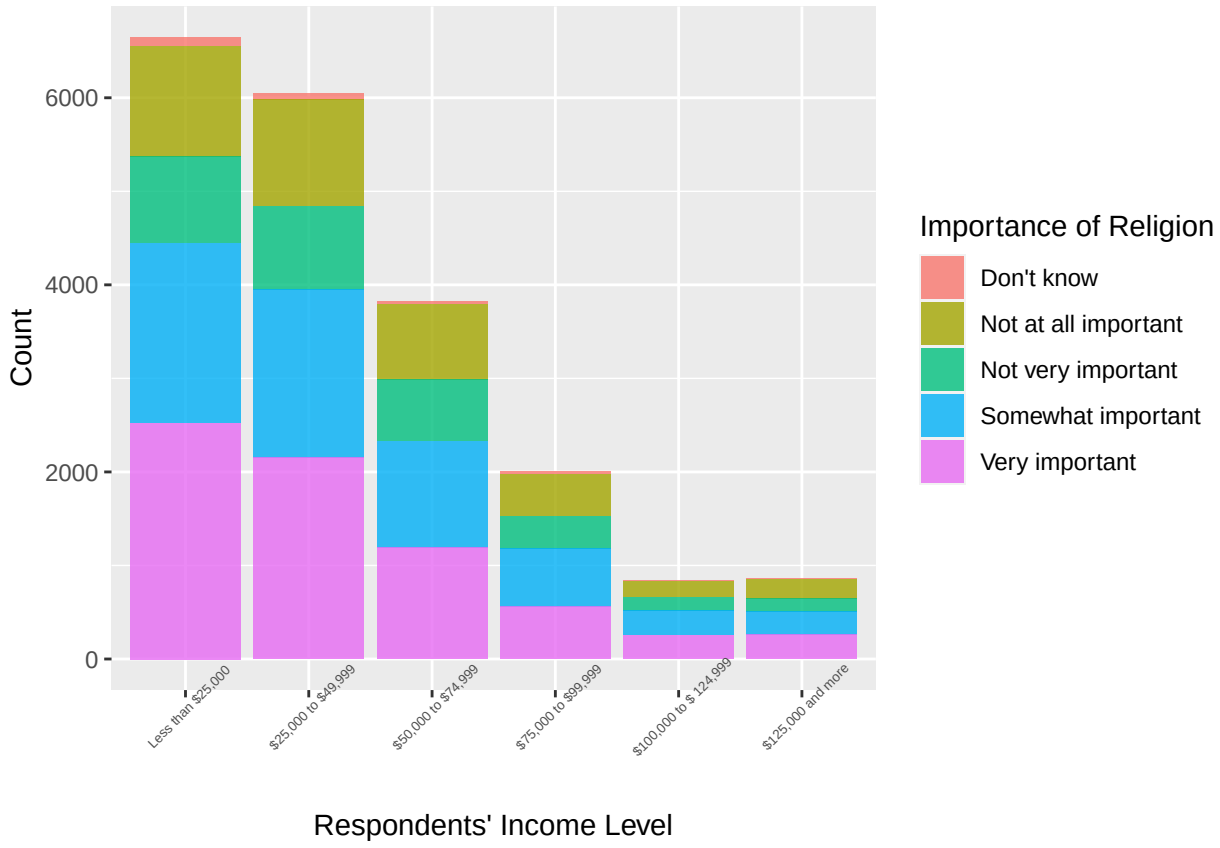

Figure 14: Individual Income and Religion Importance

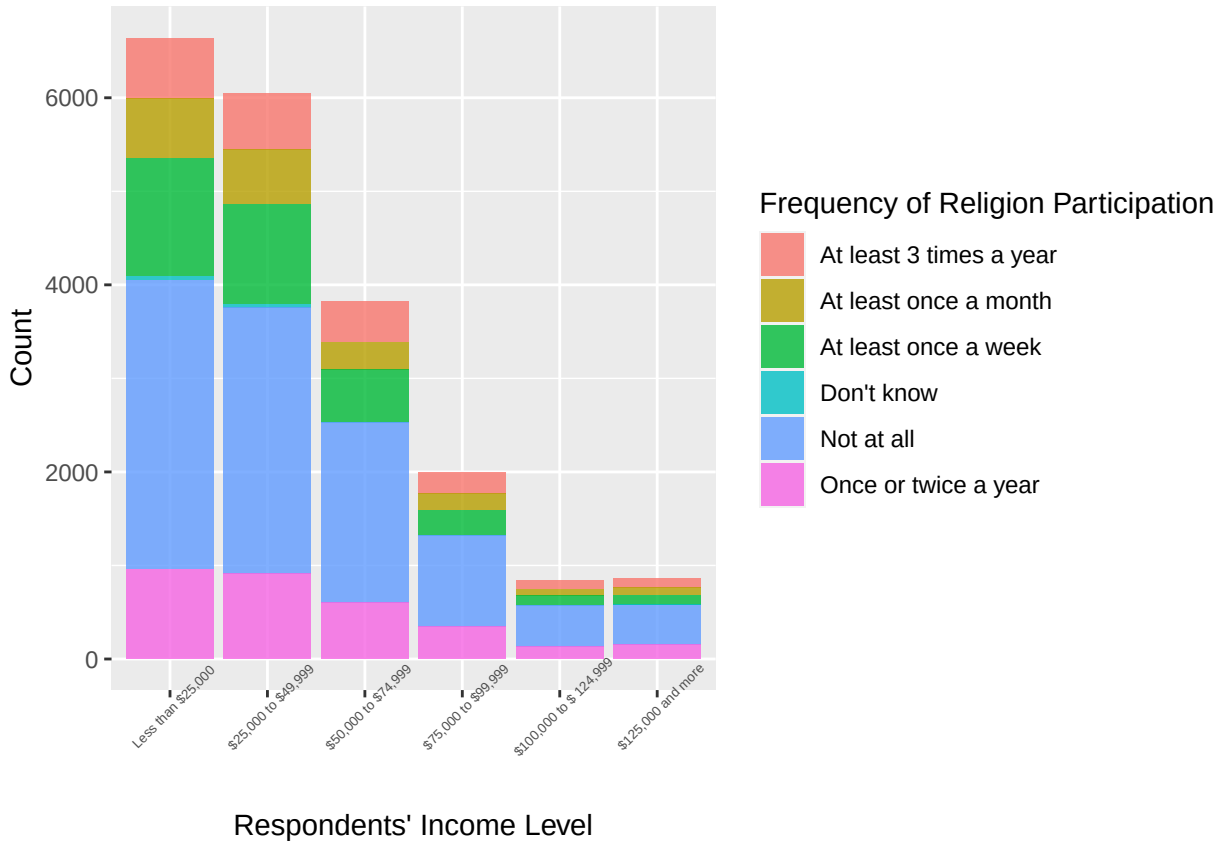

Figure 15: Individual Income and Religion Participation

3.1 Weaknesses and next steps

A weakness with this methodology is that households without a telephone number are excluded with it being the respondent's first point of contact with the conductors. The GSS estimates that roughly 1% of households in 2013 do not have a telephone service. The GSS however, accounts for the 1% of households without service by using weighted estimates to represent them. Another weakness with this survey is that it limits surveys to one taker per household and therefore, may elicit different responses from different members. However, telephone/online surveys are efficient and its easy accessibility allows conductors to cover a wide geographical area needed to cover large populations.

Appendix

Supplementary Survey Questions:

1. What is your age?
2. What was your place of birth?
3. What is your gender?
4. How often do you participate in religious activities?
5. What is the highest level of education you have received?
6. On a scale of 1-10, how satisfied do you feel about life? With 1 being the least satisfied and 10 being the most satisfied.
7. What is your annual income before taxes?
8. What was your income last year before taxes?
9. What is the total income of your family before taxes?
10. Have you faced discrimination in the workforce?

A References

Mercer, Andrew (2020) (*General Social Survey, Cycle 27: Social Identity* 2015) Hadley Wickham et al (2019) Wickham (2016) Mancini, Claudia (2020) al (2018) Y. X. et al (2021) H. Z. et al (2021)

al, Hadley Wickham et. 2018. *Dplyr: A Grammar of Data Manipulation.* https://CRAN.R-project.org/package=dplyr.

al, Hao Zhu et. 2021. *KableExtra: Construct Complex Table with 'Kable' and Pipe Syntax.* https://cran.r-project.org/web/packages/kableExtra/index.html.

al, Yihui Xie et. 2021. *Knitr: A General-Purpose Package for Dynamic Report Generation in R.* https://cran.r-project.org/web/packages/knitr/index.html.

General Social Survey, Cycle 27: Social Identity. 2015. Statistics Canada; Statistics Canada. https://sda-artsci-utoronto-ca.myaccess.library.utoronto.ca/sdaweb/dli2/gss/gss27v2/more_doc/GSS27ENgidV2.pdf.

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

Mancini, Claudia. 2020. *Advantages and Disadvantages of a Phone Survey.* IDSurvey; IDSurvey. https://www.idsurvey.com/en/advantages-and-disadvantages-of-phone-survey/#:~:text=Advantages%20of%20phone%2Dbased%20interviewing&text=Research%20can%20be%20gathered%20quickly,sample%20to%20complete%20the%20survey.

Mercer, Andrew. 2020. *Oversampling Is Used to Study Small Groups, Not Bias Poll Results.* Pew Research Center; Pew Research Center. https://www.pewresearch.org/fact-tank/2016/10/25/oversampling-is-used-to-study-small-groups-not-bias-poll-results/.

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

The General Social Survey: An Overview. 2017. Statistics Canada; Statistics Canada. https://www150.statcan.gc.ca/n1/pub/89f0115x/89f0115x2013001-eng.htm#a1.

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