

Predicting People's Intention to Leave The Job Based on Income, Work Fairness and Workload

YUYAN XU, YIFAN XU, YIJIE ZHAO, YUZE KANG

2020/10/16

Abstract

People's planning to leave their current job could be determined by multiple factors. Therefore through using R as a tool of analysis and plotting a fitted regression model based on the data from General social survey(GSS), Cycle 30, 2016: Canadians at work and home, we found that people's personal income, their feelings of their workload and work distribution influence their choice of leaving jobs evidently, which is important for monitoring and estimating when making policies for possible emerging job issues.

Introduction

Work stressors has been a concerned topic among researchers and the general public, it serves as a critical factor that influences the Canadians' life and health conditions. Some research has shown that work stress is a risk factor for major depression episodes (Wang, 2005); some research has discovered that the relationship between work stress and health conditions is stronger among lower social class people (Ibrahim, Smith & Muntaner, 2009); also, there were research suggested that work distribution and schedule affect the employees' health conditions (Jamal, 2004). These research all reinforced the idea that work stress and people's health conditions are correlated in different ways, therefore it is important to investigate how the working environment affects people's mindsets and their other aspects of life.

This report examined the possible relationship between work stressors and people's intention of leaving their current jobs. We used the dataset provided by the 2016 General Social Survey (GSS) on Canadians at work and Home. The 2016 GSS is a sample survey with cross-sectional design, which targeted the population includes all non-institutionalized persons that is or order than 15 years old and living in Canada. This survey aimed to provide information about the working conditions, relationships with coworkers, work-family relationship, and how these factors affect the Canadians' well beings (GSS, 2016). There are many possible reasons for a person to leave their job, we chose to solely consider the factor, work stress. This is because work stressors could be a major cause to job quitting, when a person makes this big decision, it means that the negative effects on their life caused by work stressors might are already intolerable. We aimed to investigate deeply if the Canadians are satisfied with their current working conditions, and how the aspects of work could influence their career choices.

We used generalized linear model (GLM) to measure the relationship between the response variable, intention of leaving jobs, and the possible work stressors. In addition, we used plots to investigate how each variable could potentially influence people's intention of leaving their jobs. The result was, the personal income, the amount of workload, and the fairness of working distribution all serve as closely correlated factors with the Canadians' intentions of leaving their jobs. This important finding shows that the nature of work and the income received can influence a person's level of satisfaction over his, or her job. Therefore, this report is able to provide a considerable insight on the Canadian labour policy, which could potentially affect the employers' decisions on improving the working environment and work benefits.

Data

In this report, We use the dataset from General Social Survey (GSS) on Canadians at Work and Home (cycle 30) for 2016. All the data, codebook and other related files of this GSS can be obtained from the link: <https://sda-artsci-utoronto-ca.myaccess.library.utoronto.ca/cgi-bin/sda/hsda?harcsda4+gss30> (<https://sda-artsci-utoronto-ca.myaccess.library.utoronto.ca/cgi-bin/sda/hsda?harcsda4+gss30>). We downloaded a CSV file of raw data which contains all the variables of this survey. Then we chose some variables for our interests and cleaned the data, in order to make a new clean dataset only with our concerned variables included. In the following data section, the collection process of 2016 GSS and some descriptions of this dataset will be talked about.

Data Collection of GSS

In 2016, Statistics Canada conducted a general social sample survey, to collect information in respect of persons' work and home. The target population of 2016 GSS is all people with age of 15+ who live in the 10 provinces of Canada, except for people in full-time institutions. And for 2016 GSS, the frame of survey had changed to include all the telephone numbers recorded in Statistics Canada and Address Register, which considered lots of households without landline. During the survey process, they used stratified sampling method by dividing population of each province into several strata, which included most Census Metropolitan Areas. Then they used simple random sampling method without replacement in every stratum, by randomly selecting a member of every households for doing the questionnaire electronically or interviewing through phone call. 2016 GSS has a response rate of 50.8%. After the process of survey, conductors got 19609 responses in total as actual sample, out of a target sample of 20000 responses. As for non-response, they had done some adjustment for 3 steps, for complete non-response, non-response with auxiliary information and partial non-response respectively.

Data Selection and Visualization

Out of hundreds of variables in this dataset, we find a variable named `jsr_04` for the preference of leaving a current job is interesting. As we all know, the stability of employees is a really important component for almost every employer. With lower turnover rate of employees, almost every employer can save a considerable sum of costs for training or other fees. So we consider the preference of leaving a current job as our response variable y . We wonder if there are some factors of employees themselves that would affect their mind of leaving their current work. So we select 9 possible x variables from this dataset as follows,

- age;
- sex;
- the size of workplace (small/medium/large businesses);
- workload management;
- fairness of work distribution;
- level of stress in life;
- personal income before taxation;
- under or over qualified for current job;
- level of belongness in current organization.

We plot graphs for each selected variables, including our chosen response variable, the preference of leaving job, to visualize the distribution of raw data for each variable. All the graphs can be found in the appendix section.

In the next part of modeling, We plan to check whether these possible variables could predict if a employee has a strong possibility to leave a current job.

Model

Set Up

We decide to build a logistic regression model for prediction. To start with, we use full model to select relatively better x variables among 9 selected possible x variables, for the purpose of model quality. Then we get three explanatory variables for our model,

- personal income before taxation;
- fairness of work distribution;
- workload management.

Next, we set up a model formula for these three explanatory variables and one response variable in the following,

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_{income_{25,000-49,999}} + \beta_2 x_{income_{50,000-74,999}} + \beta_3 x_{income_{75,000-99,999}} \\ + \beta_4 x_{income_{100,000-124,999}} + \beta_5 x_{fair_{agree}} + \beta_6 x_{fair_{NAND}} + \beta_7 x_{fair_{disagree}} + \beta_8 x_{fair_{StronglyDisagree}} \\ + \beta_9 x_{workload_{ML_{often}}} + \beta_{10} x_{workload_{ML_{some}}} + \beta_{11} x_{workload_{ML_{rarely}}} + \beta_{12} x_{workload_{ML_{never}}}$$

Also, the above formula can be changed into:

$$\frac{p}{1-p} = \exp(\beta_0 + \beta_1 x_{income_{25,000-49,999}} + \beta_2 x_{income_{50,000-74,999}} + \beta_3 x_{income_{75,000-99,999}} \\ + \beta_4 x_{income_{100,000-124,999}} + \beta_5 x_{fair_{agree}} + \beta_6 x_{fair_{NAND}} + \beta_7 x_{fair_{disagree}} + \beta_8 x_{fair_{StronglyDisagree}} \\ + \beta_9 x_{workload_{ML_{often}}} + \beta_{10} x_{workload_{ML_{some}}} + \beta_{11} x_{workload_{ML_{rarely}}} + \beta_{12} x_{workload_{ML_{never}}})$$

In the formula, p represents the probability that the respondent is planning to leave a current job.

Model Interpretation

We then use the `glm()` function in R to build our logistic regression model and view the results.

```
Call:
glm(formula = factor(leaving_job) ~ factor(income) + factor(fair_work) +
    factor(workload_manageable), family = "binomial", data = sample0)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.3866	0.4131	0.4776	0.5570	1.5767

Coefficients:

	Estimate	Std. Error	z value
(Intercept)	1.21358	0.10458	11.605
factor(income)2	1.01825	0.08687	11.722
factor(income)3	1.23840	0.09462	13.088
factor(income)4	1.25829	0.10949	11.492
factor(income)5	1.35771	0.15354	8.843
factor(fair_work)2	-0.33536	0.10308	-3.253
factor(fair_work)3	-0.61563	0.12076	-5.098
factor(fair_work)4	-0.88406	0.12317	-7.177
factor(fair_work)5	-1.38017	0.17945	-7.691
factor(workload_manageable)2	0.21703	0.08165	2.658
factor(workload_manageable)3	-0.03390	0.09455	-0.359
factor(workload_manageable)4	-0.34306	0.14498	-2.366
factor(workload_manageable)5	-0.73591	0.23860	-3.084

	Pr(> z)
(Intercept)	< 2e-16 ***
factor(income)2	< 2e-16 ***
factor(income)3	< 2e-16 ***
factor(income)4	< 2e-16 ***
factor(income)5	< 2e-16 ***
factor(fair_work)2	0.00114 **
factor(fair_work)3	3.44e-07 ***
factor(fair_work)4	7.10e-13 ***
factor(fair_work)5	1.46e-14 ***
factor(workload_manageable)2	0.00786 **
factor(workload_manageable)3	0.71996
factor(workload_manageable)4	0.01797 *
factor(workload_manageable)5	0.00204 **

Table 1

Variable

- income2 represents personal before-tax income in the range of CAD25,000 to CAD49,999;
- income3 represents income in the range of CAD50,000 to CAD74,999;
- income4 represents income in the range of CAD75,000 to CAD99,999;
- income5 represents income in the range of CAD100,000 to CAD124,999;
- fair_work2 represents “agree that work is fairly distributed”;
- fair_work3 represents “neither agree nor disagree that work is fairly distributed”;
- fair_work4 represents “disagree that work is fairly distributed”;
- fair_work5 represents “strongly disagree that work is fairly distributed”;
- workload_manageable2 represents respondent often considers workload manageable;
- workload_manageable3 represents respondent considers workload manageable sometimes;
- workload_manageable2 represents respondent rarely considers workload manageable;
- workload_manageable2 represents respondent never considers workload manageable.

Coefficient

From the above table 1, we can observe that:

[1] The estimate of β_0 is 1.21358, which means the odds that a respondent is planning to leave a current job is equal to $\exp(1.21358)$, when the before-tax income of the respondent is less than \$25,000, and the respondent strongly agree that work is fairly distributed, and the respondent always considers the workload manageable.

[2] The estimate of β_1 is 1.01825, which means the log of odds that a respondent is planning to leave a current job increases by $\exp(1.01825)$ times, for income changing from the range of less than CAD25,000 to the range of CAD25,000 to CAD49,999. The estimate of β_2 is 1.23840, which means the log of odds that a respondent is planning to leave a current job increases by $\exp(1.23840)$ times, for income changing from the range of CAD25,000 to CAD49,999 to the range of CAD50,000 to CAD74,999. The estimate of β_3 is 1.25829, which means the log of odds that a respondent is planning to leave a current job increases by $\exp(1.25829)$ times, for income changing from the range of CAD50,000 to CAD74,999 to the range of CAD75,000 to CAD99,999. The estimate of β_4 is 1.35771, which means the log of odds that a respondent is planning to leave a current job increases by $\exp(1.35771)$ times, for income changing from the range of CAD75,000 to CAD99,999 to the range of CAD100,000 to CAD124,999.

[3] The estimate of β_5 is -0.33536, which means the log of odds that a respondent is planning to leave a current job reduce by $\exp(0.33536)$ times, when a respondent changes mind about whether work is fairly distributed from strongly agree to agree. The estimate of β_6 is -0.61563, which means the log of odds that a respondent is planning to leave a current job reduce by $\exp(0.61563)$ times, when a respondent changes mind about whether work is fairly distributed from agree to neither agree nor disagree. The estimate of β_7 is -0.88406, which means the log of odds that a respondent is planning to leave a current job reduce by $\exp(0.88406)$ times, when a respondent changes mind about whether work is fairly distributed from neither agree nor disagree to disagree. The estimate of β_8 is -1.38017, which means the log of odds that a respondent is planning to leave a current job reduce by $\exp(1.38017)$ times, when a respondent changes mind about whether work is fairly distributed from disagree to strongly disagree.

[4] The estimate of β_9 is 0.21703, which means the log of odds that a respondent is planning to leave a current job increase by $\exp(0.21703)$ times, when a respondent often considers the workload manageable rather than always. The estimate of β_{10} is -0.03390, which means the log of odds that a respondent is planning to leave a current job reduces by $\exp(0.03390)$ times, when a respondent considers the workload manageable sometimes rather than often. The estimate of β_{11} is -0.34306, which means the log of odds that a respondent is planning to leave a current job reduces by $\exp(0.34306)$ times, when a respondent rarely considers the workload manageable rather than sometimes. The estimate of β_{12} is -0.73591, which means the log of odds that a respondent is planning to leave a current job reduces by $\exp(0.73591)$ times, when a respondent never considers the workload manageable rather than rarely.

As for the p-value for each coefficient, all the p-values are less than 0.05, except for β_{10} , which means most coefficients of our model

Model Formula

Based on the estimates of coefficients from the table 1, our model formula is:

$$\left(\frac{p}{1-p}\right) = \exp(1.21358 + 1.01825x_{income25,000-49,999} + 1.23840x_{income50,000-74,999} + 1.25829x_{income75,000-99,999} + 1.35771x_{income100,000-124,999} - 0.33536x_{fair_{agree}} - 0.61563x_{fair_{NAND}} - 0.88406x_{fair_{disagree}} - 1.38017x_{fair_{stronglyDisagree}} + 0.21703x_{workloadML_{often}} - 0.03390x_{workloadML_{some}} - 0.34306x_{workloadML_{rarely}} - 0.73591x_{workloadML_{never}})$$

Model Diagnostic

```

              Df Deviance Resid. Df Resid. Dev Pr(>Chi)
NULL                                7505      6462.8
factor(income)                      4      204.53   7501      6258.3 < 2.2e-16 ***
factor(fair_work)                   4      123.07   7497      6135.2 < 2.2e-16 ***
factor(workload_manageable)         4       30.57   7493      6104.6 3.746e-06 ***
---

```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Table 2

Figure 1: Anova table

From this output, we can see that all three variables had extremely low p-value, which means that they all had significant statistical relationship with the response variable, intention of leaving current jobs. Also, the residual deviance measures the difference between our fitted model and a perfect model, in this case, the residual deviance were 6258.3, 6135.2 and 6104.6 for income, fairness of work distribution and manageable amount of workload respectively. This number is very high, which means that the goodness of fit is bad.

Results

After building and summarizing a fitted model for the data of surveyed respondents, we've calculated the predicted probabilities for each observation and found out that, personal income (before tax), management of workload and fair work distribution are very influential to a respondent's willing to leave his or her job, among variables we were interested in. From the model coefficients in above Table 1, we can tell that most of the p-values for groups of each variable are evidently lower than the significance level($p \leq 0.05$), which represents that the association between these independent variables and the response variable (whether to leave current job) is not to be neglected.

I. Personal Income (before tax)

Figure 1 displays that, among respondents that are willing to leave current job in the next 12 months, people who earn less than \$25000 per year take the greatest proportion, the second greatest proportion is the next higher income level, and so on.

However, according to the predicted probability of leaving current job grouped income levels (in Figure 2), in group 5, the minimum probability has exceeded 60% and maximum is over 90%, but for group 1 the maximum is only over 80%. There is an observable ascending trend of probability as income levels increases.

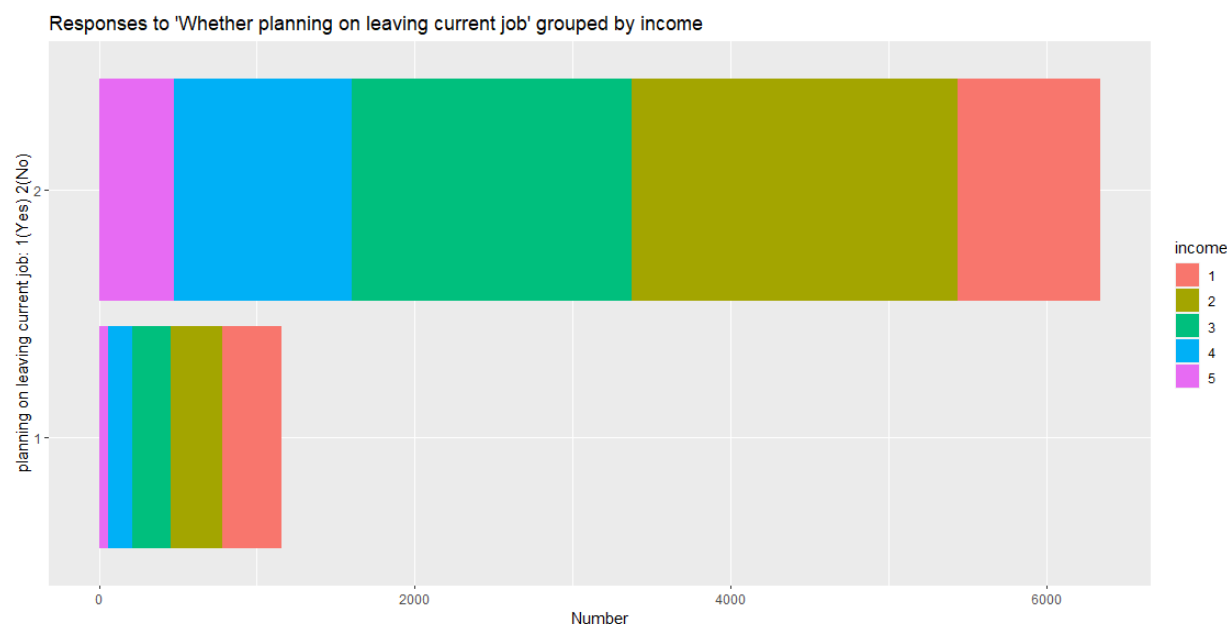


Figure 1

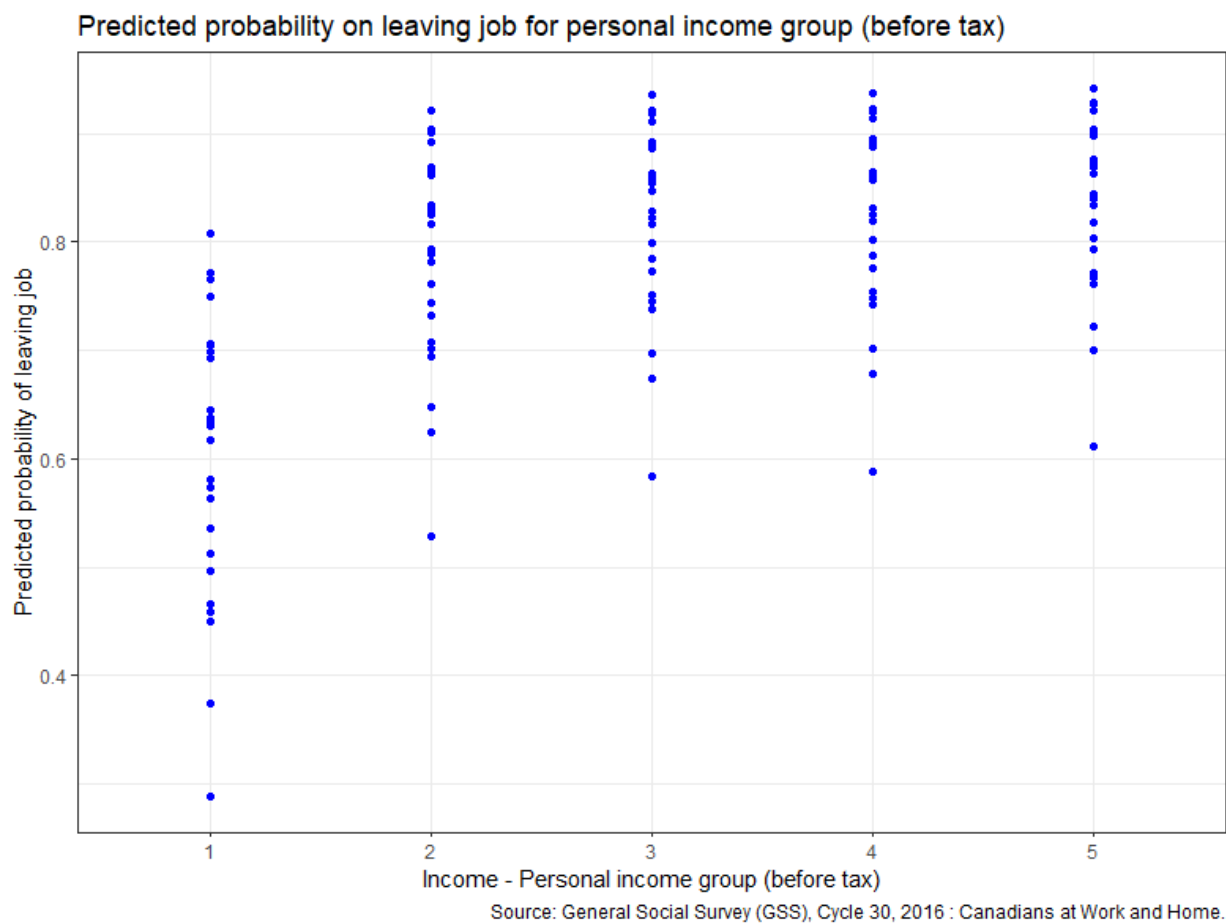


Figure 2

II. Work Load Manageable

Figure 3 showed that among people who plan to to leave their jobs, those who always or often think workload is manageable have the greatest number, even by observing the predicted probabilities (Figure 4) the maximum of these two group have exceeded 90%.



Figure 3

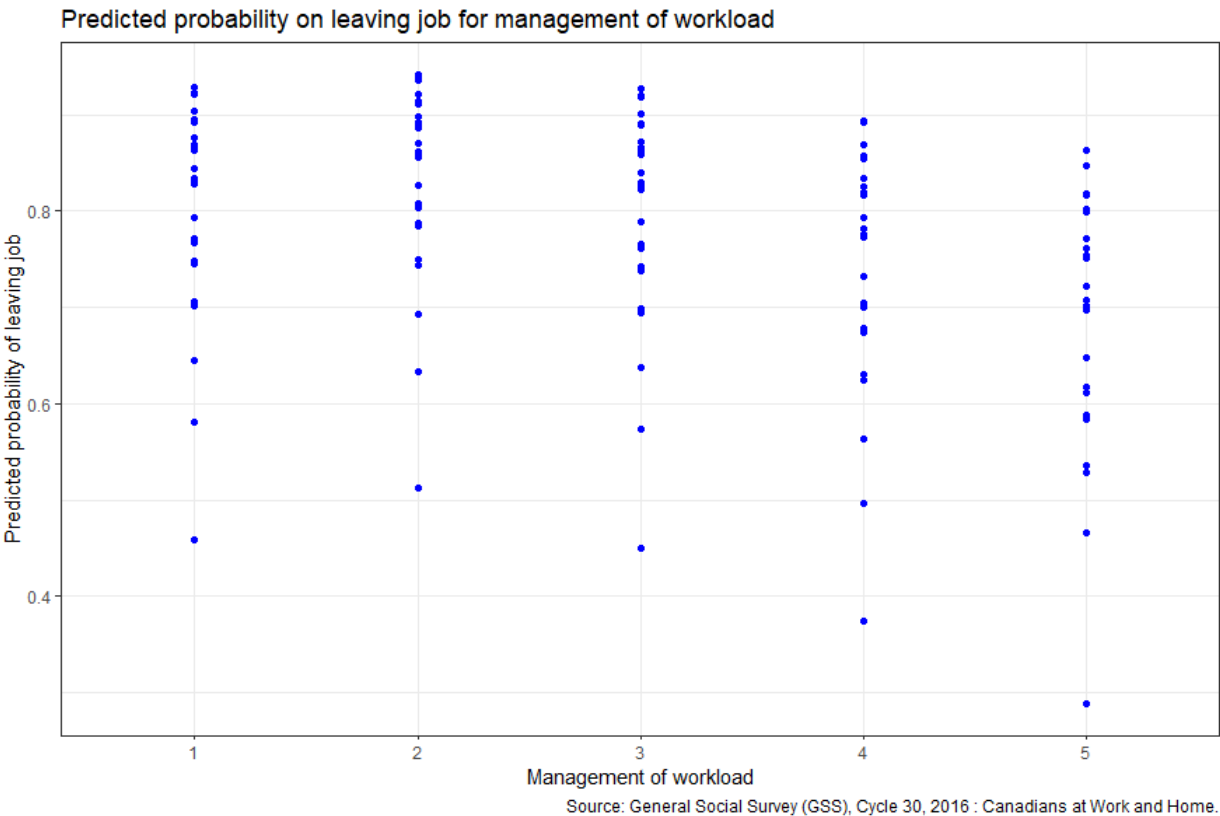


Figure 4

III. Fair Work Distribution

As displayed in Figure 5, respondents agree with that work distribution are fair occupy the greatest percentage of people planning to leave current jobs. The second highest proportion is taken by people disagree with the statement.

By observing the maximum of predicted probability of each group (each level of agreement) shown by figure 6, we can see that as people get more agreed that work distribution is fair, the probability of leaving job also increases.

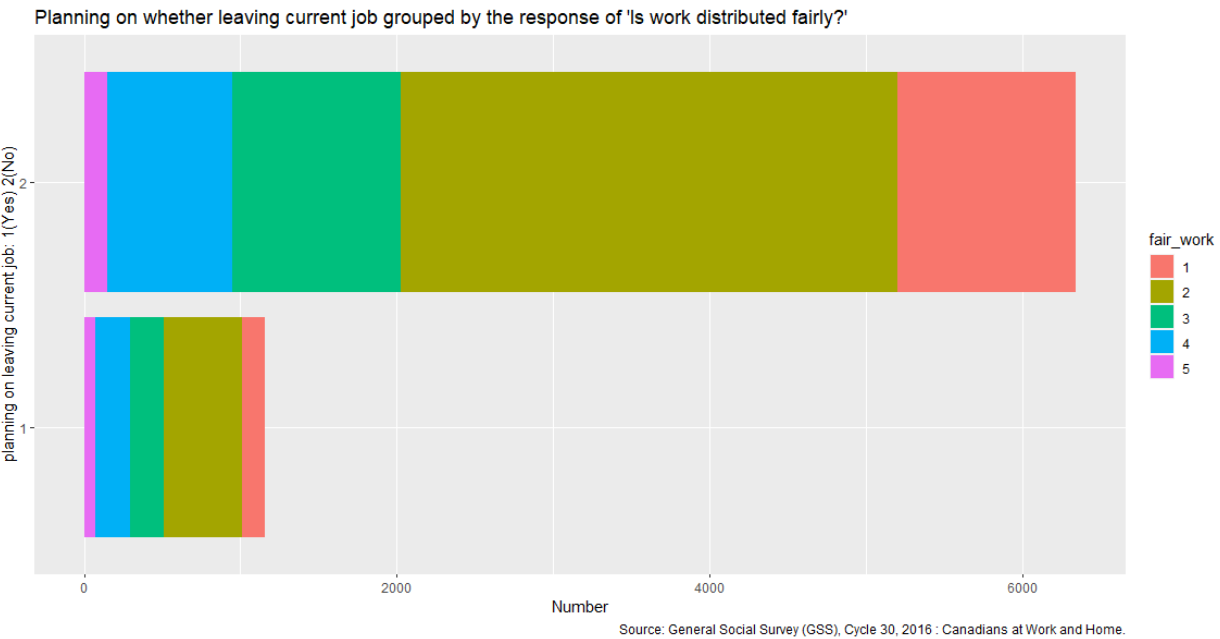
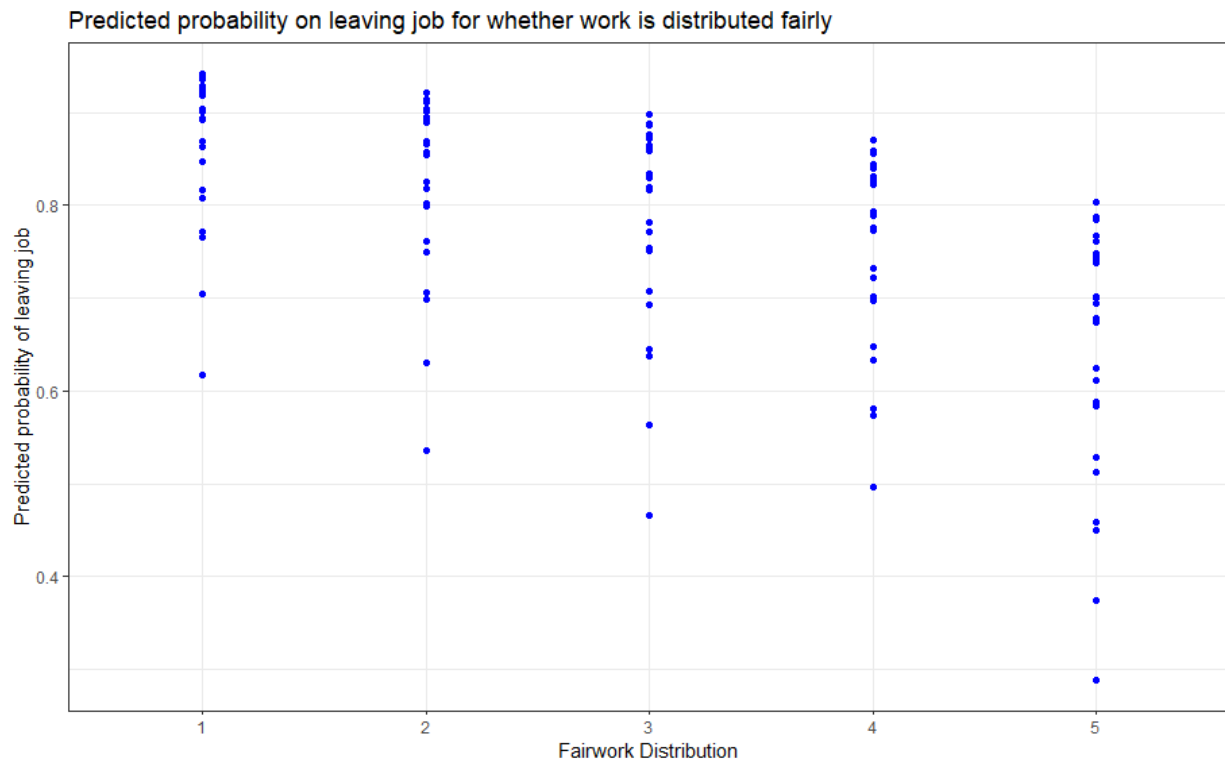


Figure 5



Discussion

The dataset was provided by the 2016 General Social Survey (GSS) on Canadians at work and Home. The survey contained inquiries about the Canadian workers' working environment, work benefits, their life conditions, how they balanced their life and work, the challenges they encountered at work, their mental states, and their intention of leaving the jobs (GSS, 2016). This dataset might have non-response bias, since there was a considerable amount of non-responses in most of the survey questions. Therefore, we set these non-response variables in null, and chose to ignore their effects on the overall result in this research.

The survey can be used to investigate many aspects of a Canadian worker's life, and we chose the aspect of possible work stressors related to their intention of leaving the jobs. We are interested in this topic because the results can contribute to the academic discussion over how policy makers could make better labour policy, in order to best protect the Canadians' health and improve their general life quality. Our finding was interesting yet not surprising that, the personal income, the amount of workload, and the fairness of working distribution all closely correlated with the Canadians' intentions of leaving their jobs. This finding duplicated the results in the Muhammed Jamal's research about how burnout, stress and health of employees is influenced by working schedules and distribution (Jamal, 2004). In addition, the finding also provided support for Jianli Wang's research on work stress serves as a risk factor for major depression among Canadian employees (Wang, 2005). This finding might also provide insights on investigating the issue of increase in unemployment rate among Canadians, although there were other possible factors that could affects the unemployment rate. However, we should also consider the situation which, some people chose to not leave their jobs, even though they were not satisfied with their current working environment, and work stressors already negatively influenced their life. This might due to the reasons that they could not afford a short of income during job switching period, or that their current jobs provided them a high income. This aspect could affects our research findings.

The weakness of this research had two aspects: the survey dataset and our data analysis. First, the original GSS survey only investigated the households with a telephone number; second, the GSS survey asked certain cub-groups such as self-employed, those looking for work, retirees and students a separate set of questions relevant to their specific situations, in an effort to cover as most as possible the Canadians that working with

paid jobs (GSS, 2016). In our research, we did not include the consideration of these sub-groups due to a large amount of data, which could likely provide information that affects the research findings. This aspect should be prevented in future research.

Appendix

Code and Data

- Full code and data supporting this analysis is available at:
<https://github.com/yifanxu33/33333/blob/main/code333.Rmd>
(<https://github.com/yifanxu33/33333/blob/main/code333.Rmd>)

Plots for Data Section

Figure1: Age group of respondent (groups of 10)

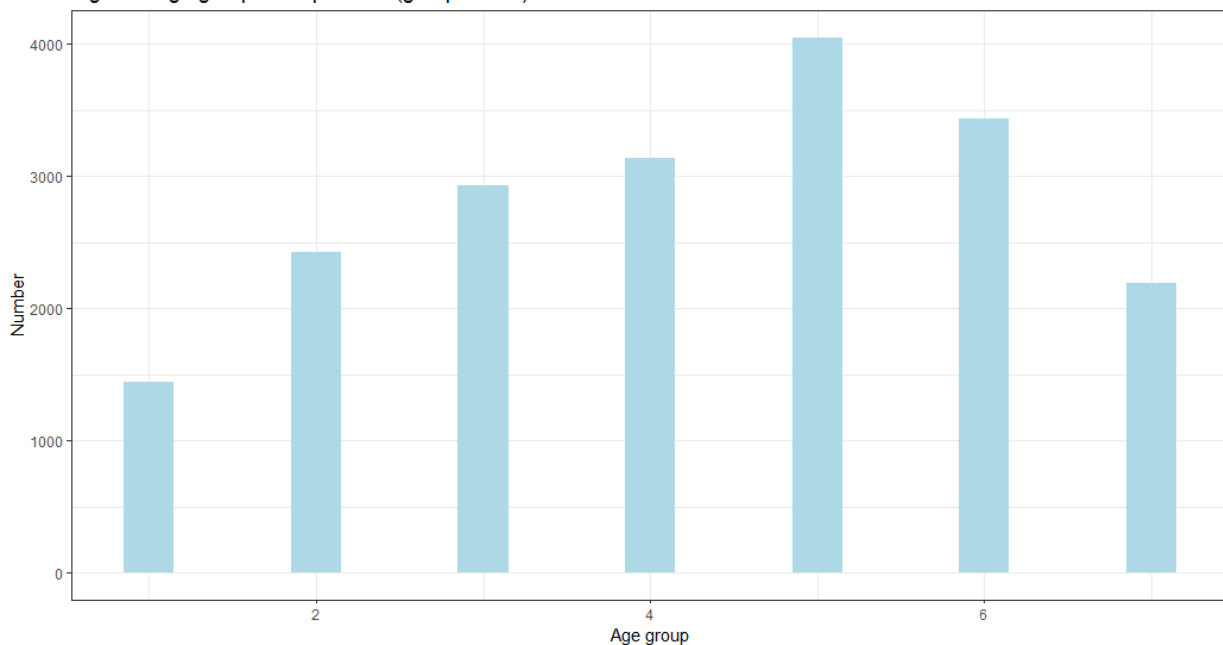
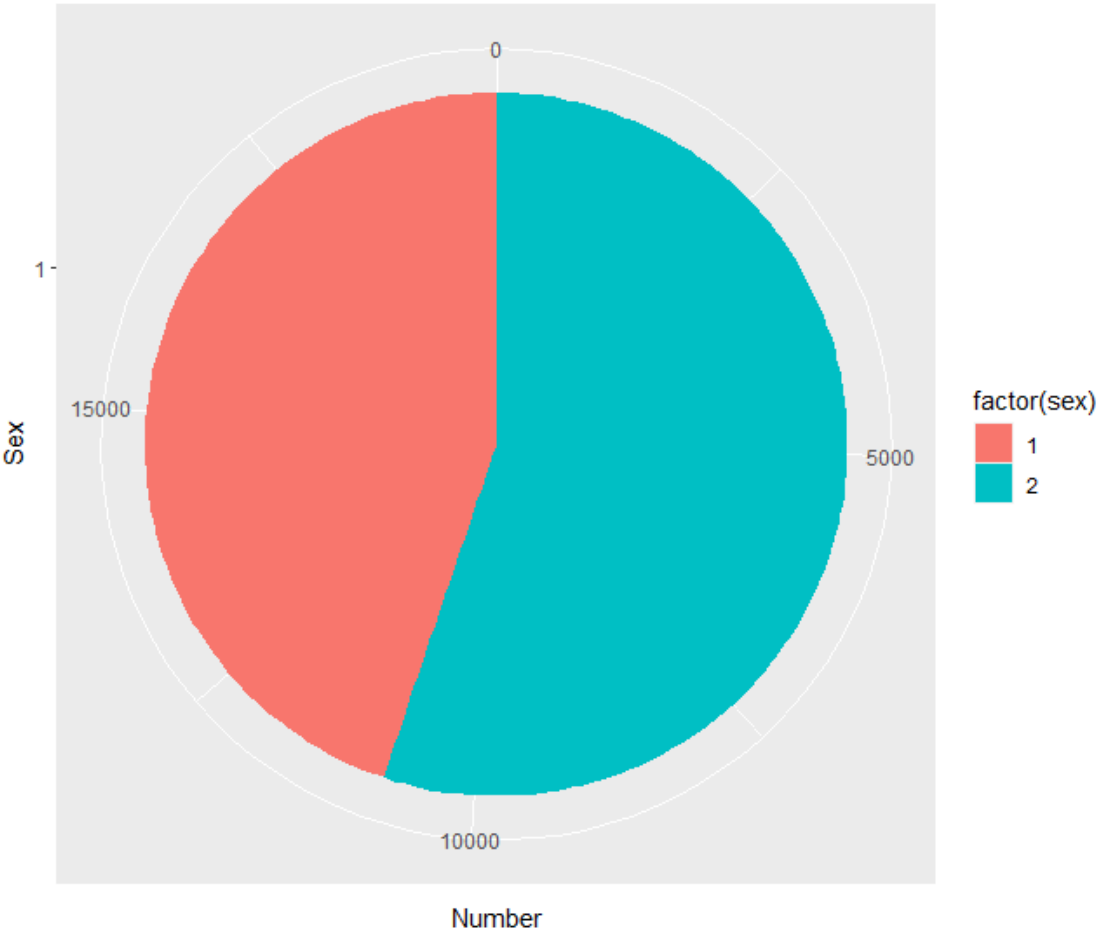
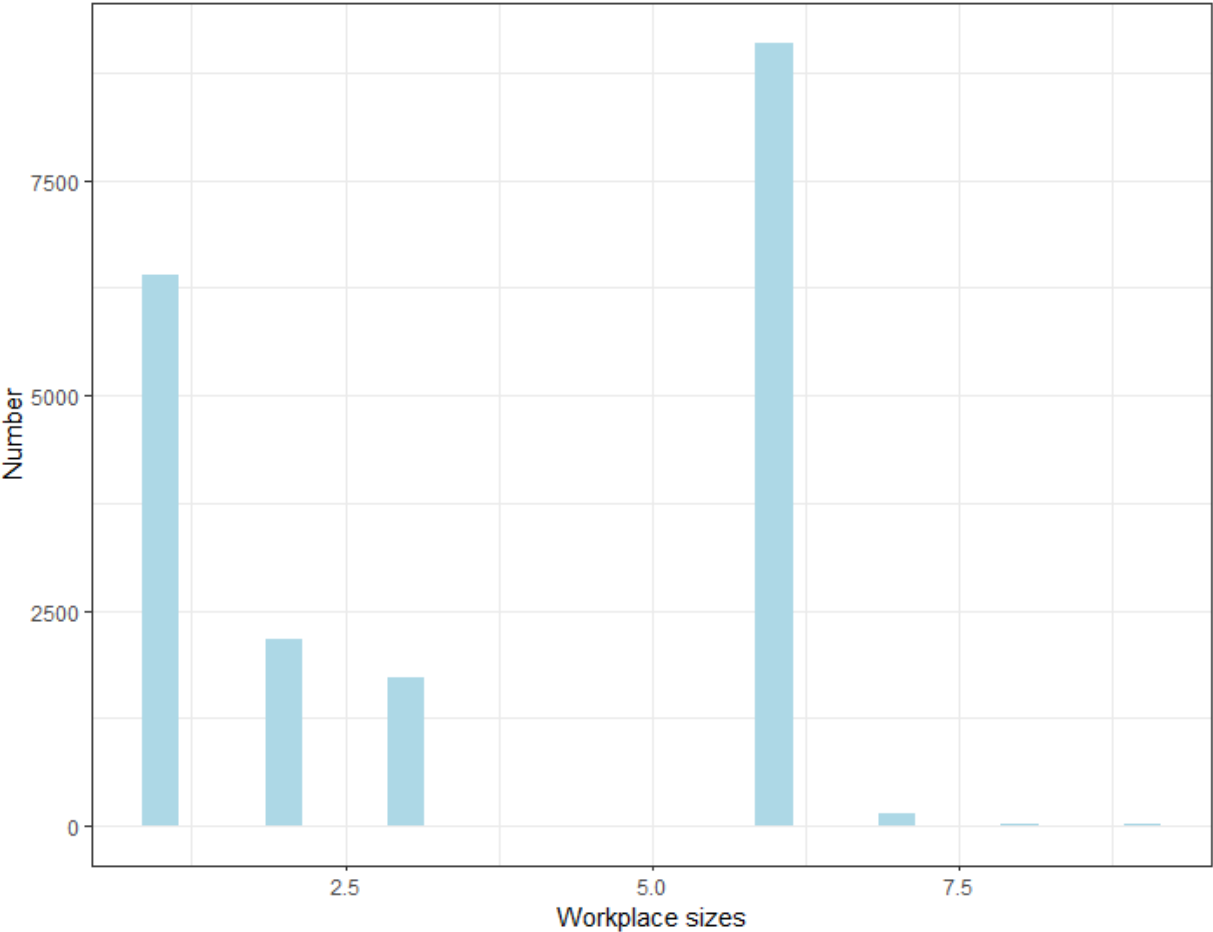


Figure2: Responses to 'What is your first language?'



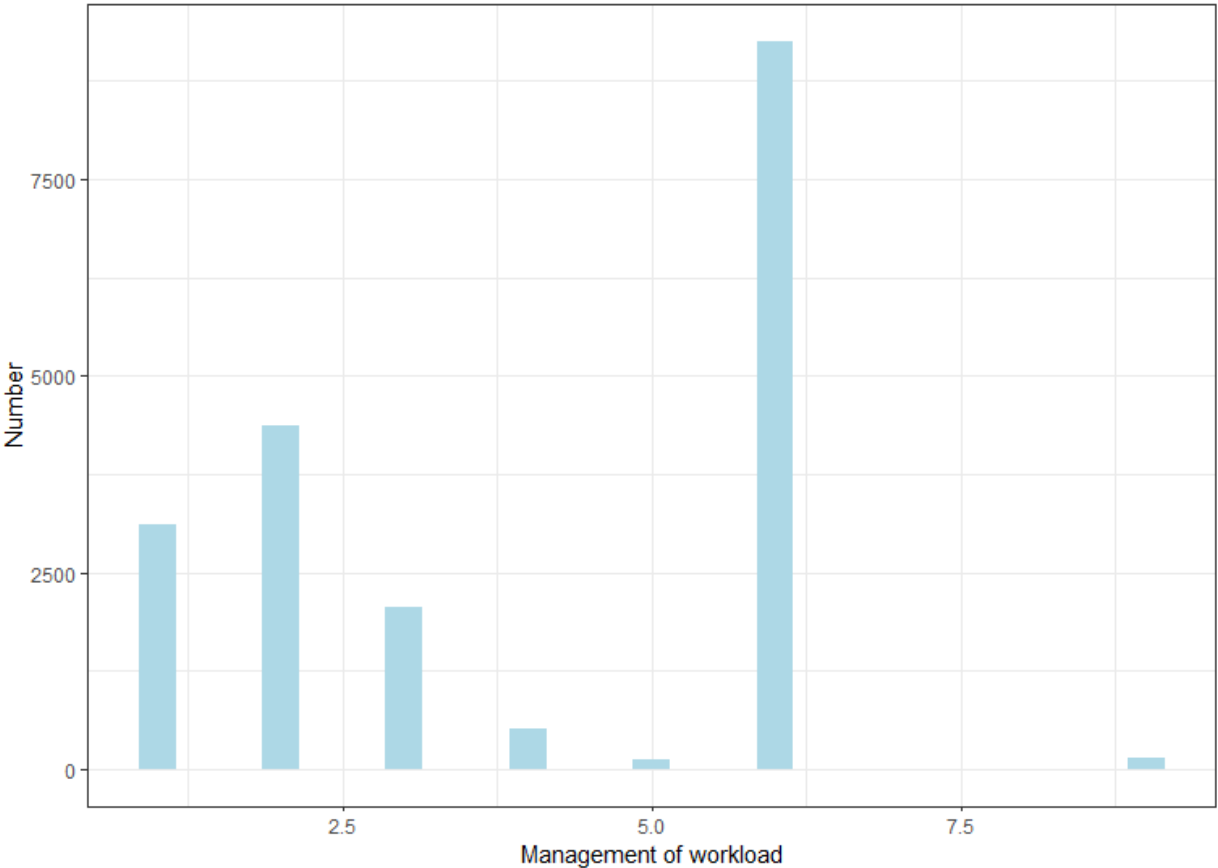
Source: General Social Survey (GSS), Cycle 30, 2016 : Canadians at Work and Home.

Figure3: Respondents' workplace size



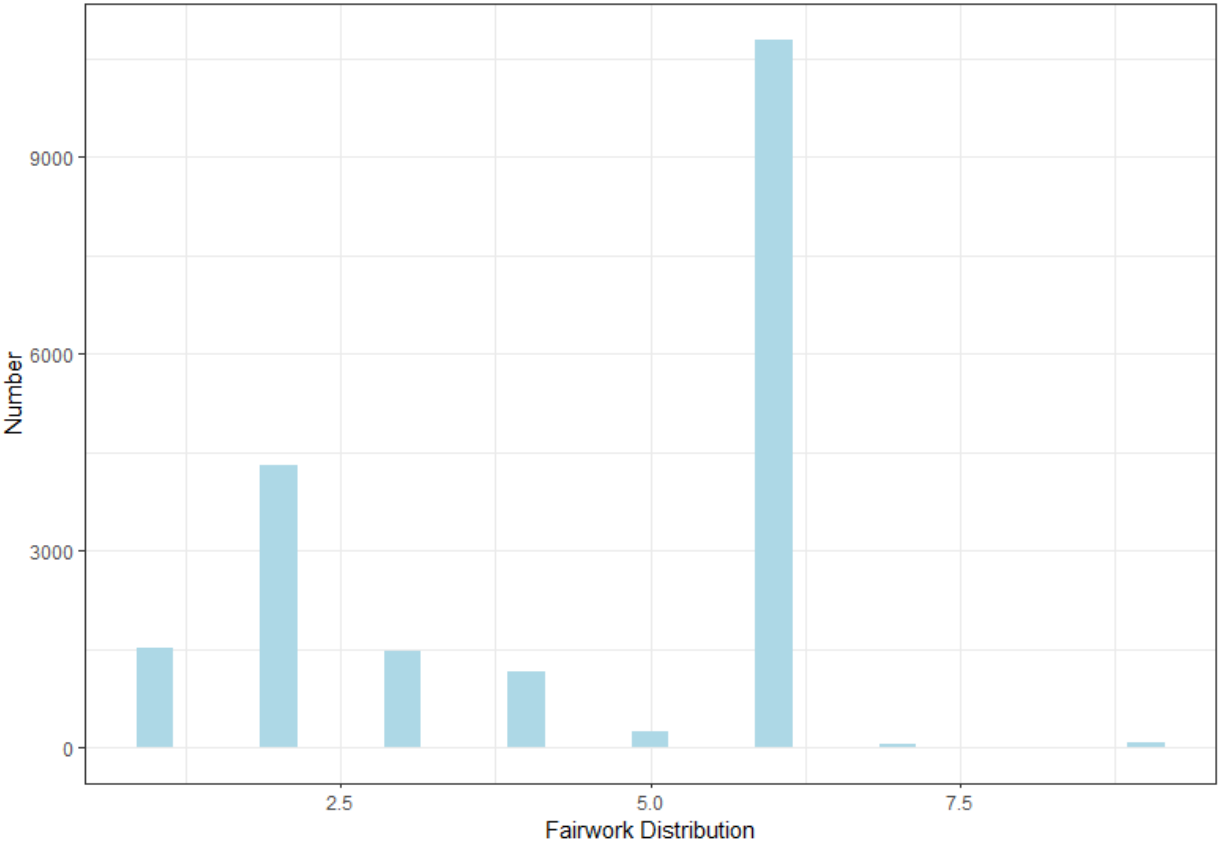
Source: General Social Survey (GSS), Cycle 30, 2016 : Canadians at Work and Home.

Figure4: Response to 'How often do you consider your workload manageable?'



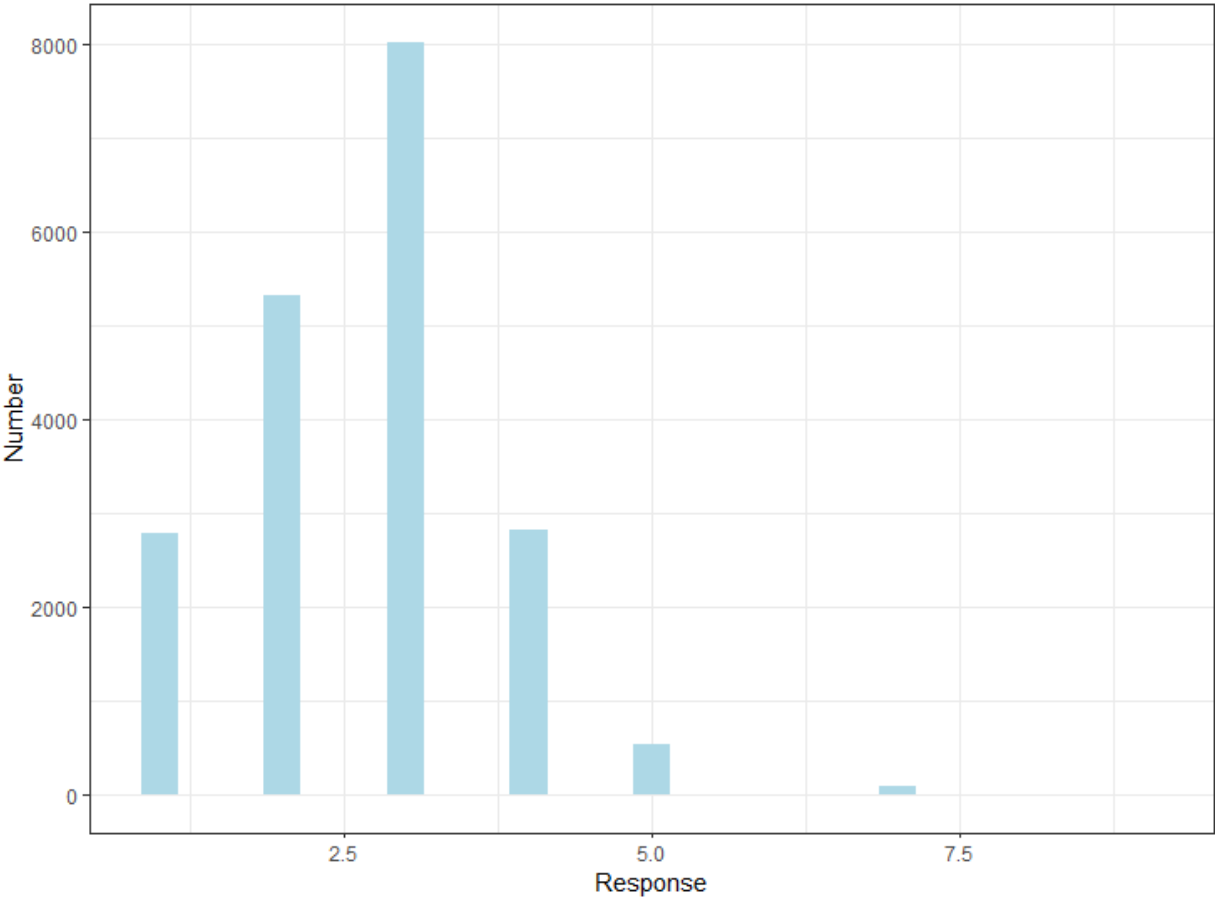
Source: General Social Survey (GSS), Cycle 30, 2016 : Canadians at Work and Home.

Figure5: Response to 'Is work distributed fairly in your workplace?'



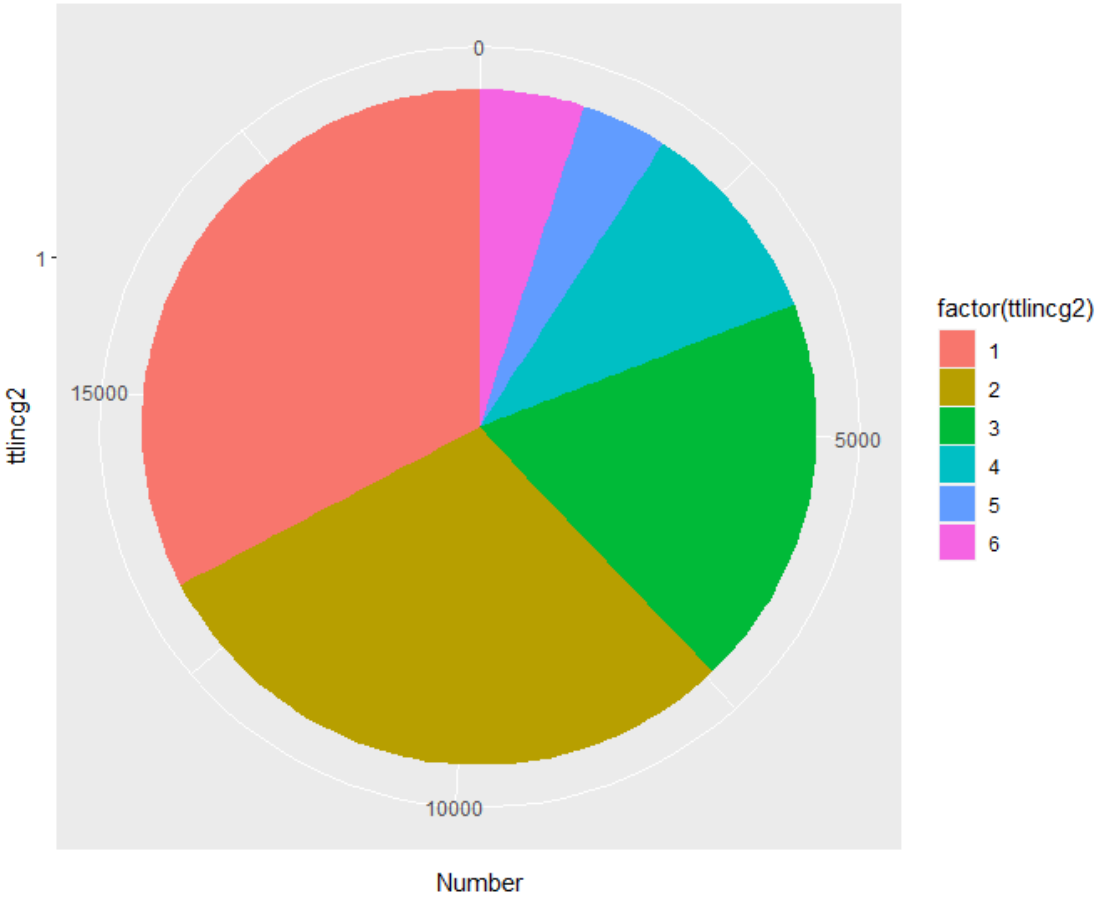
Source: General Social Survey (GSS), Cycle 30, 2016 : Canadians at Work and Home.

Figure6: Response to the amount of stress in your life



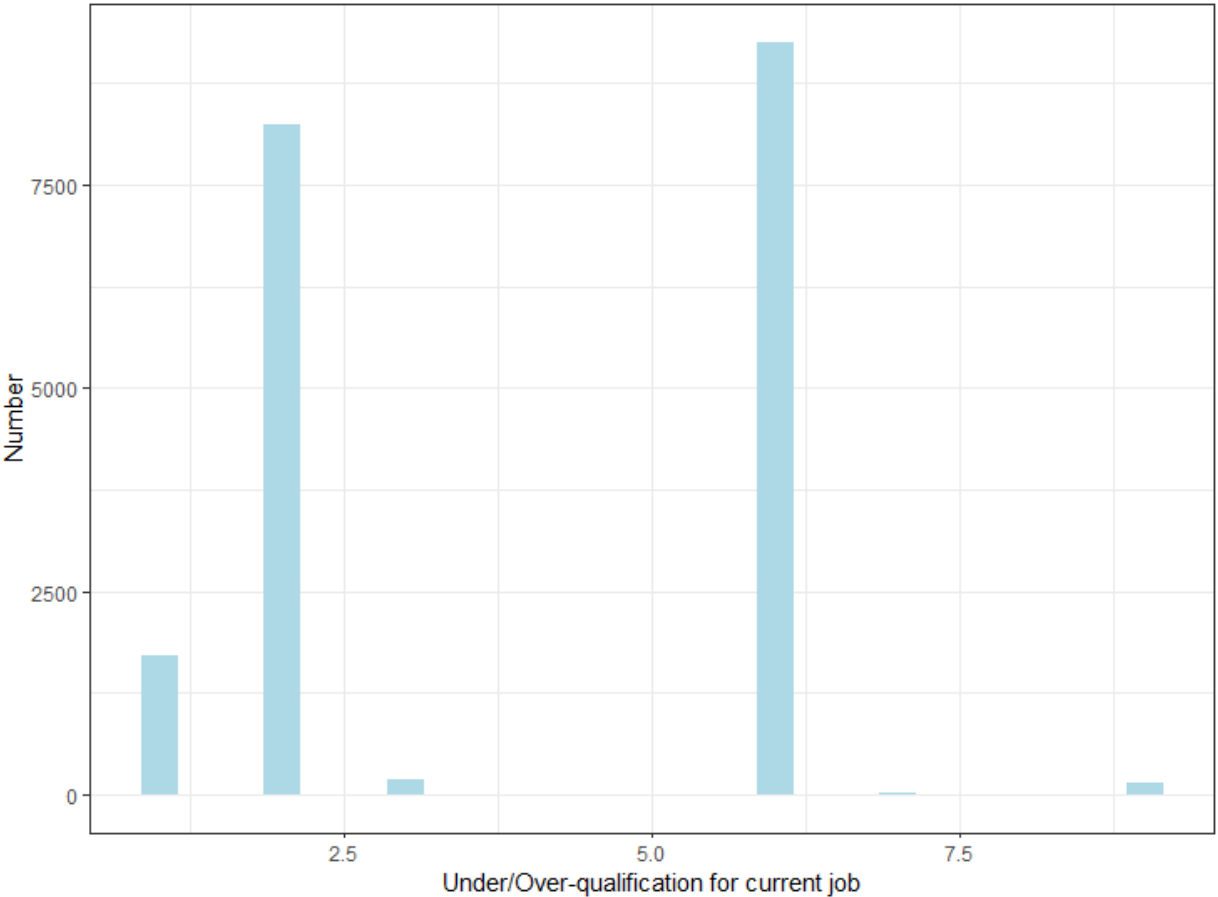
Source: General Social Survey (GSS), Cycle 30, 2016 : Canadians at Work and Home.

Figure7: Response to personal income group (before tax)



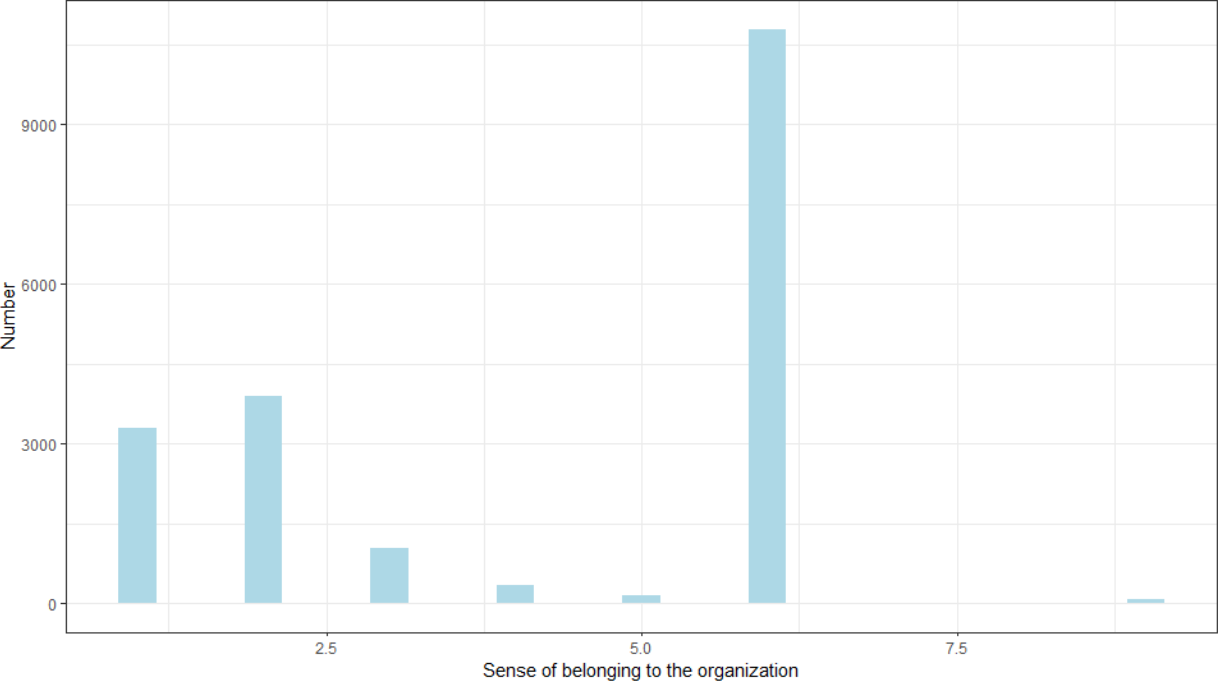
Source: General Social Survey (GSS), Cycle 30, 2016 : Canadians at Work and Home.

Figure8: Response to the degree of qualification for the job



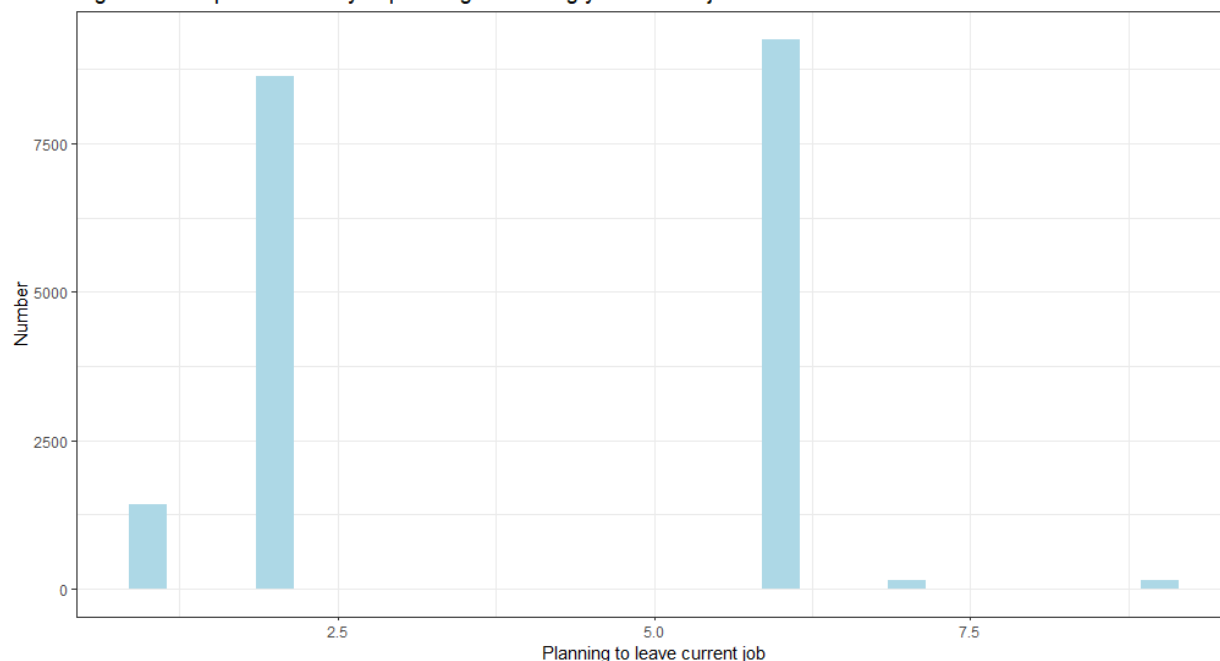
Source: General Social Survey (GSS), Cycle 30, 2016 : Canadians at Work and Home.

Figure9: Response to the degree of belonging to the organization



Source: General Social Survey (GSS), Cycle 30, 2016 : Canadians at Work and Home.

Figure10: Response to 'Are you planning on leaving your current job?'



Source: General Social Survey (GSS), Cycle 30, 2016 : Canadians at Work and Home.

References

- Arnold, Jeffrey B. (2019). ggthemes: Extra Themes, Scales and Geoms for 'ggplot2'. R package version 4.2.0. <https://CRAN.R-project.org/package=ggthemes> (<https://CRAN.R-project.org/package=ggthemes>)
- General Social Survey (GSS), Cycle 30, 2016 : Canadians at Work and Home.
- Hadley Wickham and Evan Miller (2020). haven: Import and Export 'SPSS', 'Stata' and 'SAS' Files. R package version 2.3.1. <https://CRAN.R-project.org/package=haven> (<https://CRAN.R-project.org/package=haven>)
- JJ Allaire and Yihui Xie and Jonathan McPherson and Javier Luraschi and Kevin Ushey and Aron Atkins and Hadley Wickham and Joe Cheng and Winston Chang and Richard Iannone (2020). rmarkdown: Dynamic Documents for R. R package version 2.3. URL <https://rmarkdown.rstudio.com> (<https://rmarkdown.rstudio.com>).
- John Fox and Sanford Weisberg (2019). An {R} Companion to Applied Regression, Third Edition. Thousand Oaks CA: Sage. URL: <https://socialsciences.mcmaster.ca/jfox/Books/Companion/> (<https://socialsciences.mcmaster.ca/jfox/Books/Companion/>)
- John H. Maindonald and W. John Braun (2020). DAAG: Data Analysis and Graphics Data and Functions. R package version 1.24. <https://CRAN.R-project.org/package=DAAG> (<https://CRAN.R-project.org/package=DAAG>)
- R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/> (<https://www.R-project.org/>).
- Sam Firke (2020). janitor: Simple Tools for Examining and Cleaning Dirty Data. R package version 2.0.1. <https://CRAN.R-project.org/package=janitor> (<https://CRAN.R-project.org/package=janitor>)
- Stefan Milton Bache and Hadley Wickham (2014). magrittr: A Forward-Pipe Operator for R. R package version 1.5. <https://CRAN.R-project.org/package=magrittr> (<https://CRAN.R-project.org/package=magrittr>)