

The attributes that have impact on people's satisfaction of life in Canada

Jiaxin Cheng(1003970241), Shilun Dai(1005414767), Tony Chen(1004265298), Weiqian Ding(1004708426)

2020/10/19

Abstract

We are interested about the living conditions and well-being of Canadians, so we look at the survey data for 20,602 Canadians about their life satisfaction in 2017. We used a given set of attributes to build a multiple linear regression model to analyze Canadians' sense of satisfaction of their lives. Many variables like marital status, religious participation and income level have a significant impact on Canadians' life satisfaction. Canadians can use the result to help them understand what really influences their life satisfaction and thus have the ability to improve their quality of life rationally.

Introduction

Nowadays, people are pursuing a high quality of life. However, people may have no idea about how to improve their level of happiness and satisfaction in life. The goal for this report is to analyze what attributes will have impacts on people's feeling of life. We did some data processing and created a new data set in RStudio. Based on our data set, we built a multiple linear regression model to explain the predictor of feelings_life(life satisfaction) by various explanatory variables. By using the stratified sampling method, we found that religion participation, income, marital status and other seven variables are the main attributes that have impacts on people's feeling of life. Therefore, people may use our conclusion to improve their quality of life. In this report, we will first introduce the data, build and analyse our multiple linear regression model, and then show the result, followed by the discussion of model distributions.

Data

We obtained the data set "General Society Survey"(2017) on the CHASS website. The target population for this data is "all persons 15 years of age and older in Canada, excluding residents of the Yukon, Northwest Territories, and Nunavut, and Full-time residents of institutions"(Beaupre, 2017). The target frame is the "lists of telephone numbers in use (both landline and cellular) available to Statistics Canada from various sources (telephone companies, Census of population, etc.) and the Address Register (AR): List of all dwellings within the ten provinces"(Beaupre, 2017).

People in the target frame will be contacted by telephone and the sample data frame is obtained by the responses. There exist different extents of non-responses. Partial non-responses occur because they refused to answer the questions or they did not understand the question while total non-responses occur because they were unable to contact or no one of the household was able to provide the answer(Beaupre, 2017).

The sampling method for this data set is assigned the survey frame to a stratum within provinces(Beaupre, 2017). And simple random sampling without replacement was done for each stratum(Beaupre, 2017). The total observation of this data is 20,502(Beaupre, 2017). The estimated population size for this survey is 30,302,287(Beaupre, 2017).

The survey uses simple random sampling strategy which is good because the observations that are collected will represent the target population more accurately. Another advantage of the survey is that the data set

was large with a large number of observations and a large number of variables. Large sample size will provide more accurate estimators to estimate the population. However, the data set only contains the information within Canada. Thus, the model we built will only be applied to estimate the life satisfaction for people living in Canada.

We selected `feelings_life` and 10 other variables from the data set to do the analysis (Table 1). Those variables were selected because they contain more valid responses and likely to have relationships with `feelings_life`. We removed 'NA' in '`feelings_life`' and '`religion_participation`'. We also removed 'Don't Know' in '`religion_participation`'. The total observations in our data sample is 20,130. The total variables in our dataframe is 11.

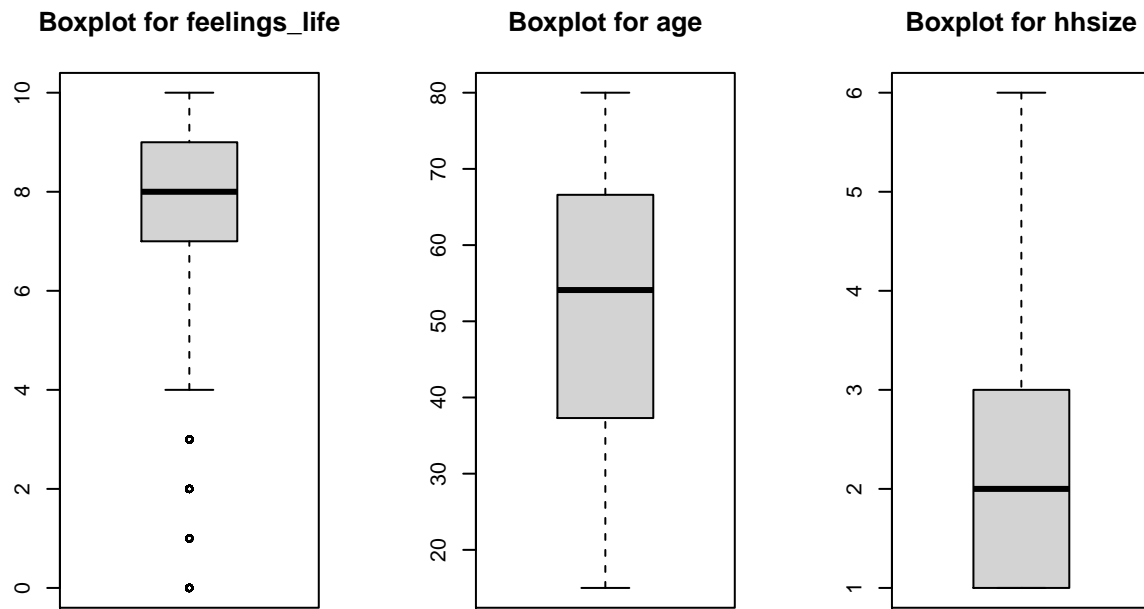
'`feelings_life`' is a numerical variable. Respondents will rate their life satisfaction (`feelings_life`) from 0 to 10. 'age' is a numerical variable. 'pop_center' is a categorical variable. Respondents select the population center that they belong to. 'sex' is a categorical variable. 'religion_participation' is a categorical variable. Respondents will choose the frequency that they participate in their religion events. 'marital_status' is a categorical variable. 'education' is a categorical variable, 'respondents' will choose their level of education. 'province' is a categorical variable, people will choose the province that they are living in right now. 'hh_size' is a numerical variable, it shows the household size of their home. 'average_hours_worked' is a numerical variable, it represents the average hours worked per week for the respondent. 'income_respondent' is a numerical variable. It represents the income for the respondent per year.

Here is the table of first six lines of data set:

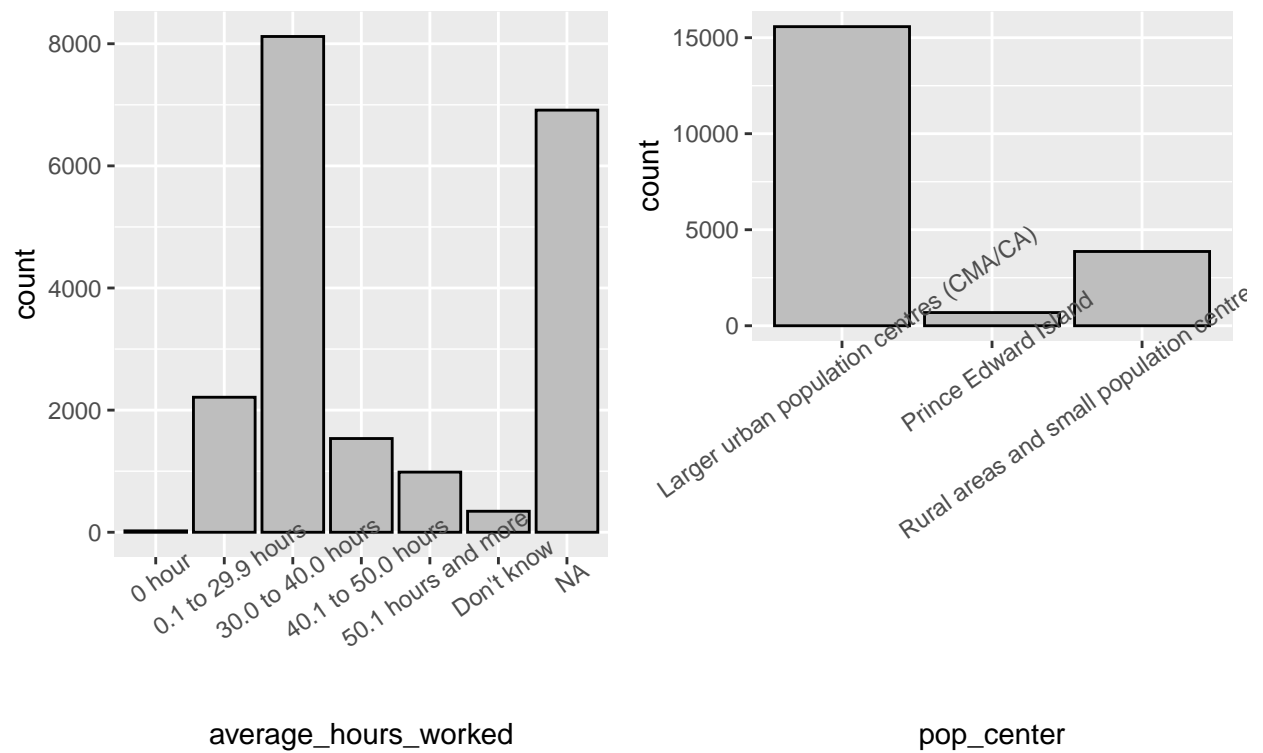
<code>feelings_life</code>	<code>age</code>	<code>pop_center</code>	<code>sex</code>	<code>religion_participation</code>	<code>marital_status</code>
8	52.7	Larger urban population centres (CMA/CA)	Female	Once or twice a year	Single, never married
8	63.6	Rural areas and small population centres (non CMA/CA)	Female	At least once a week	Married
10	80.0	Larger urban population centres (CMA/CA)	Female	Not at all	Married
8	28.0	Larger urban population centres (CMA/CA)	Male	Not at all	Living common-law
9	63.0	Larger urban population centres (CMA/CA)	Female	Not at all	Married
4	58.8	Larger urban population centres (CMA/CA)	Female	Not at all	Single, never married

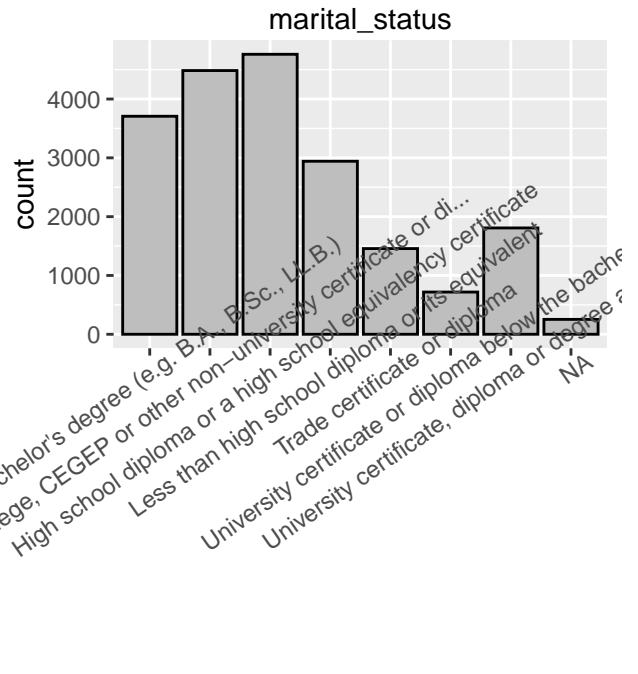
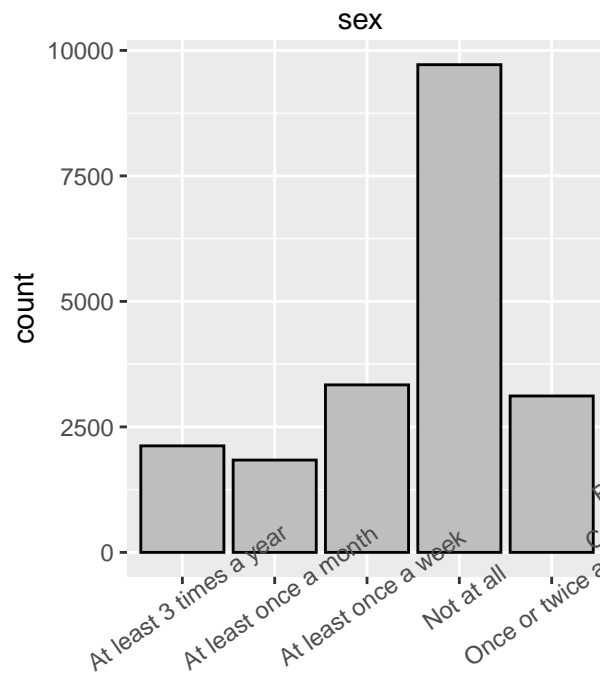
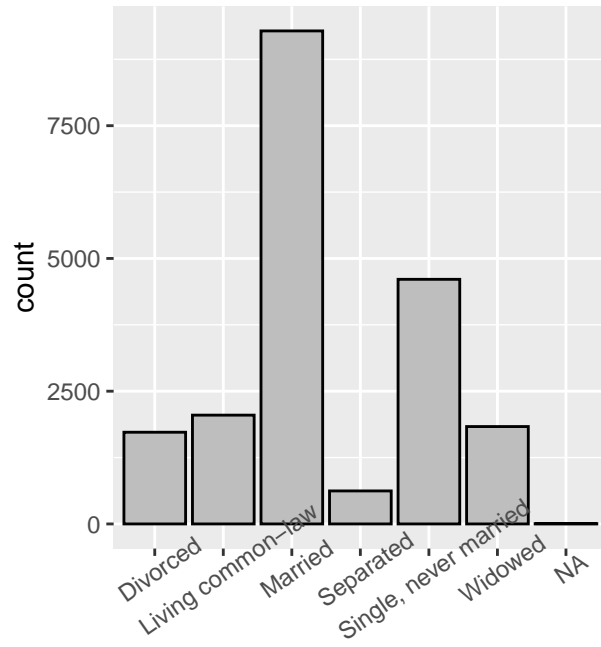
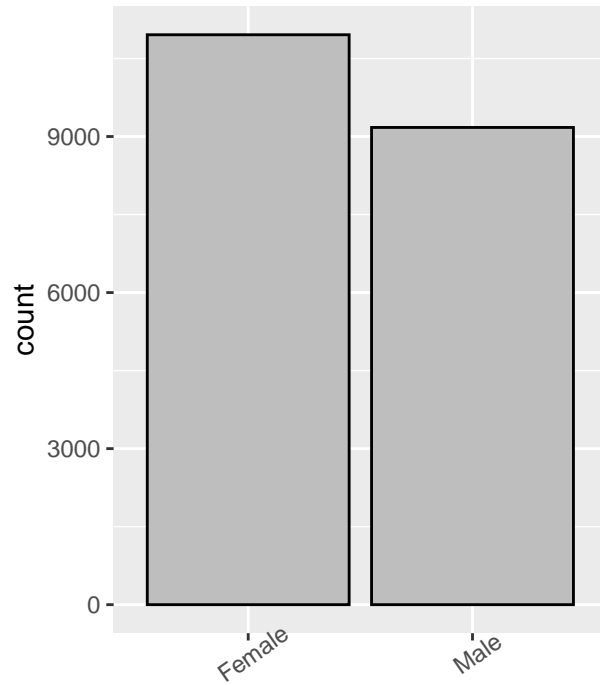
<code>education</code>	<code>province</code>	<code>hh_size</code>	<code>average_hours_worked</code>	<code>income_respondent</code>
High school diploma or a high school equivalency certificate	Quebec	1	30.0 to 40.0 hours	\$25,000 to \$49,999
Bachelor's degree (e.g. B.A., B.Sc., LL.B.)	Ontario	2	NA	\$25,000 to \$49,999
High school diploma or a high school equivalency certificate	Alberta	2	NA	\$50,000 to \$74,999
College, CEGEP or other non-university certificate or di. . .	Quebec	2	30.0 to 40.0 hours	Less than \$25,000
High school diploma or a high school equivalency certificate	Quebec	2	NA	Less than \$25,000
Less than high school diploma or its equivalent	Nova Scotia	1	NA	Less than \$25,000

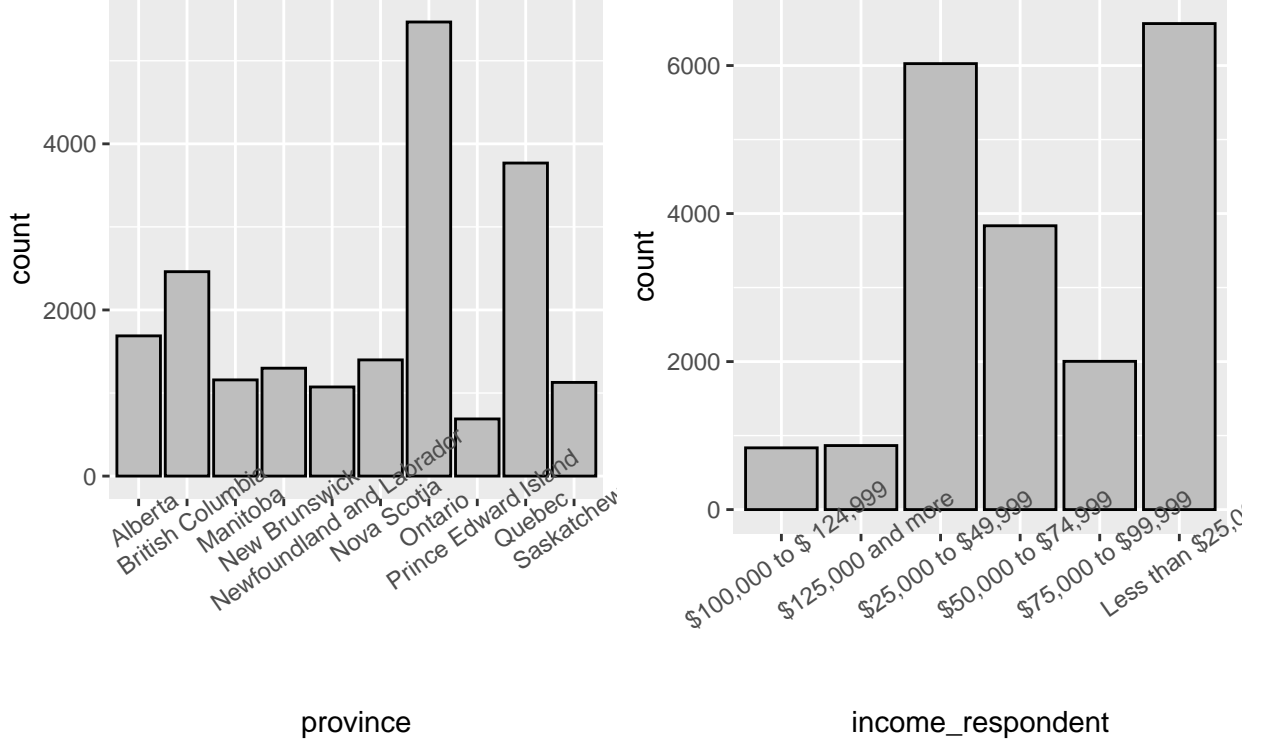
The following are the plots of the raw data:



According to the 'Boxplot for the feelings_life', the trend for the distribution of feelings_life is left-skewed. The following are the bar plot for 8 categorical variables. They are the cumulative frequency graphs.







Based on the data set that we selected with feelings_life and 10 other variables. We created a new variable 'average_income' by using income_respondent divided by 'average_hours_worked' its household size. We did not chose correlated variables(eg. province and region). Our stratified sampling method is conducted by province.

Model

We continue our analysis by building a multiple linear regression model to predict feelings_life using age, pop_center, sex, religion_participation, marital_status and average_income. We choose this model because the predictor feelings_life is numeric and we believe there exists a linear relationship between explanatory variables and feelings_life. The variables of age and average_income are numeric which represents respondents' age and the total income of their family per hour per person. Pop_center, sex, religion_participation, marital_status, education and province are categorical variables which represent if they live in urban city or rural areas, respondents' sex, frequency of participating the religion activities and whether they are married, divorced or single, repectively, while education means the highest level of education respondents get and province is simply the residence of respondents.

Using the data selected, we build a linear regression model with the following formula:

$$\begin{aligned}
 y_i = & \beta_0 + \beta_1 x_{age,i} + \beta_2 x_{pop_center_island,i} + \beta_3 x_{pop_center_rural,i} + \beta_4 x_{marital_status_common,i} \\
 & + \beta_5 x_{marital_status_married,i} + \beta_6 x_{marital_status_separated,i} + \beta_7 x_{marital_status_single,i} + \beta_8 x_{marital_widowed,i} \\
 & + \beta_9 x_{sex_m,i} + \beta_{10} x_{religion_monthly,i} + \beta_{11} x_{religion_weekly,i} + \beta_{12} x_{religion_never,i} + \beta_{13} x_{religion_yearly,i} + \beta_{14} x_{edu_college,i} \\
 & + \beta_{15} x_{edu_highschool,i} + \beta_{16} x_{edu_lowlevel,i} + \beta_{17} x_{edu_trade,i} + \beta_{18} x_{edu_belowbachelor,i} + \beta_{19} x_{edu_abovebachelor,i} \\
 & + \beta_{20} x_{province_BC,i} + \beta_{21} x_{province_Manitoba,i} + \beta_{22} x_{province_NB,i} + \beta_{23} x_{province_NL,i} + \beta_{24} x_{province_NS,i} \\
 & + \beta_{25} x_{province_ON,i} + \beta_{26} x_{province_island,i} + \beta_{27} x_{province_Quebec,i} + \beta_{28} x_{province_Saskatchewan,i} + \beta_{29} x_{average_income,i} + \epsilon_i
 \end{aligned}$$

β_0 is the coefficient of intercept. Other β s are the coefficients of corresponding variables. For numeric variables, $x_{age,i}$ represents the age of respondents and $x_{average_income,i}$ is the average income of respondents'

family per hour per person. For categorical variables, $x_{pop_island,i}$ means whether the respondents live in Prince Edward Island or not. The variable is equal to 1 if the respondent lives there and is 0 otherwise. Same with all other categorical attributes. Finally ϵ_i is the error of our model.

In this model, we choose age and sex as explanatory variable instead of other variables related to age and sex because there are less missing value and NAs in these two variables. Province is a key variable since we do stratified sampling and treat respondents in different provinces as stratas. $Average_income$ is computed as $income_{respondent}/average_hours_worked/hh_size$, which means the average disposable income of respondents' family per work time. We also think that marital status, religion participation and education will influence the quality of life, resulting in the change in the score of feelings life.

The results of code ran by R is:

	Estimate
## (Intercept)	7.641864e+00
## age	3.050938e-03
## pop_centerPrince Edward Island	1.782206e-01
## pop_centerRural areas and small population centres (non CMA/CA)	1.220147e-01
## marital_statusLiving common-law	5.304732e-01
## marital_statusMarried	6.522261e-01
## marital_statusSeparated	-2.739755e-01
## marital_statusSingle, never married	-3.773260e-02
## marital_statusWidowed	-2.325970e-01
## sexMale	-8.993412e-02
## religion_participationAt least once a month	1.008783e-01
## religion_participationAt least once a week	1.235972e-01
## religion_participationNot at all	-1.282438e-01
## religion_participationOnce or twice a year	-9.153288e-02
## educationCollege, CEGEP or other non-university certificate or di...	-2.021141e-02
## educationHigh school diploma or a high school equivalency certificate	-3.681235e-02
## educationLess than high school diploma or its equivalent	1.658869e-01
## educationTrade certificate or diploma	-1.618916e-02
## educationUniversity certificate or diploma below the bachelor's level	7.217443e-02
## educationUniversity certificate, diploma or degree above the bach...	-2.994744e-02
## provinceBritish Columbia	2.121650e-02
## provinceManitoba	2.640103e-02
## provinceNew Brunswick	1.841262e-01
## provinceNewfoundland and Labrador	2.132207e-01
## provinceNova Scotia	7.374044e-02
## provinceOntario	6.142705e-02
## provinceQuebec	1.432766e-01
## provinceSaskatchewan	1.122603e-01
## average_income	-3.275351e-07
##	Std. Error
## (Intercept)	9.576157e-02
## age	1.075961e-03
## pop_centerPrince Edward Island	8.236381e-02
## pop_centerRural areas and small population centres (non CMA/CA)	3.596018e-02
## marital_statusLiving common-law	6.269540e-02
## marital_statusMarried	5.330442e-02
## marital_statusSeparated	8.706360e-02
## marital_statusSingle, never married	5.786634e-02
## marital_statusWidowed	1.001983e-01
## sexMale	2.671910e-02
## religion_participationAt least once a month	5.927306e-02

## religion_participationAt least once a week	5.368389e-02
## religion_participationNot at all	4.339168e-02
## religion_participationOnce or twice a year	4.999637e-02
## educationCollege, CEGEP or other non-university certificate or di...	3.834729e-02
## educationHigh school diploma or a high school equivalency certificate	3.976783e-02
## educationLess than high school diploma or its equivalent	5.480132e-02
## educationTrade certificate or diploma	5.543065e-02
## educationUniversity certificate or diploma below the bachelor's level	7.440149e-02
## educationUniversity certificate, diploma or degree above the bach...	4.898731e-02
## provinceBritish Columbia	5.715657e-02
## provinceManitoba	6.940694e-02
## provinceNew Brunswick	6.900704e-02
## provinceNewfoundland and Labrador	7.230287e-02
## provinceNova Scotia	6.692350e-02
## provinceOntario	4.995471e-02
## provinceQuebec	5.364222e-02
## provinceSaskatchewan	6.912691e-02
## average_income	6.694094e-07
##	t value
## (Intercept)	79.8009510
## age	2.8355480
## pop_centerPrince Edward Island	2.1638222
## pop_centerRural areas and small population centres (non CMA/CA)	3.3930509
## marital_statusLiving common-law	8.4611187
## marital_statusMarried	12.2358709
## marital_statusSeparated	-3.1468428
## marital_statusSingle, never married	-0.6520647
## marital_statusWidowed	-2.3213674
## sexMale	-3.3659117
## religion_participationAt least once a month	1.7019246
## religion_participationAt least once a week	2.3023139
## religion_participationNot at all	-2.9554941
## religion_participationOnce or twice a year	-1.8307907
## educationCollege, CEGEP or other non-university certificate or di...	-0.5270623
## educationHigh school diploma or a high school equivalency certificate	-0.9256817
## educationLess than high school diploma or its equivalent	3.0270595
## educationTrade certificate or diploma	-0.2920615
## educationUniversity certificate or diploma below the bachelor's level	0.9700670
## educationUniversity certificate, diploma or degree above the bach...	-0.6113306
## provinceBritish Columbia	0.3711996
## provinceManitoba	0.3803803
## provinceNew Brunswick	2.6682230
## provinceNewfoundland and Labrador	2.9489938
## provinceNova Scotia	1.1018616
## provinceOntario	1.2296547
## provinceQuebec	2.6709676
## provinceSaskatchewan	1.6239746
## average_income	-0.4892896
##	Pr(> t)
## (Intercept)	0.000000e+00
## age	4.581954e-03
## pop_centerPrince Edward Island	3.049661e-02
## pop_centerRural areas and small population centres (non CMA/CA)	6.933017e-04
## marital_statusLiving common-law	2.936884e-17

```
## marital_statusMarried 3.117901e-34
## marital_statusSeparated 1.654257e-03
## marital_statusSingle, never married 5.143712e-01
## marital_statusWidowed 2.028279e-02
## sexMale 7.651759e-04
## religion_participationAt least once a month 8.879401e-02
## religion_participationAt least once a week 2.133359e-02
## religion_participationNot at all 3.127430e-03
## religion_participationOnce or twice a year 6.715525e-02
## educationCollege, CEGEP or other non-university certificate or di... 5.981596e-01
## educationHigh school diploma or a high school equivalency certificate 3.546290e-01
## educationLess than high school diploma or its equivalent 2.474411e-03
## educationTrade certificate or diploma 7.702444e-01
## educationUniversity certificate or diploma below the bachelor's level 3.320316e-01
## educationUniversity certificate, diploma or degree above the bach... 5.409917e-01
## provinceBritish Columbia 7.104951e-01
## provinceManitoba 7.036695e-01
## provinceNew Brunswick 7.635058e-03
## provinceNewfoundland and Labrador 3.193921e-03
## provinceNova Scotia 2.705428e-01
## provinceOntario 2.188493e-01
## provinceQuebec 7.572948e-03
## provinceSaskatchewan 1.044061e-01
## average_income 6.246451e-01
## [1] 0.06624723
```

From the table above, we notice that most p-values of these variables are small, which means the corresponding coefficients are significant. And from the last line, R-squared is computed as 0.06625, representing that 6.625% of variations can be explained by the model.

Overall, the performance is ok. Even though our r-squared is small, the model can still be well explained because the p-values are significant enough, which means there is no evidence to prove that the coefficients of variables are 0.

We also have an alternative model which is

$$y_i = \beta_0 + \beta_1 x_{age,i} + \beta_2 x_{pop_center_island,i} + \beta_3 x_{pop_center_rural,i} + \beta_4 x_{marital_status_common,i} + \beta_5 x_{marital_status_married,i} + \beta_6 x_{marital_status_separated,i} + \beta_7 x_{marital_status_single,i} + \beta_8 x_{marital_status_widowed,i} + \beta_9 x_{sex_m,i} + \beta_{10} x_{province_BC,i} + \beta_{11} x_{province_Manitoba,i} + \beta_{12} x_{province_NB,i} + \beta_{13} x_{province_NL,i} + \beta_{14} x_{province_NS,i}$$

The explanatory variables are selected from the first model, which are age, pop_center, marital_status, sex and province. We delete those variables with high p values to improve our model.

The results of code ran by R is:

```
## Estimate
## (Intercept) 7.641864e+00
## age 3.050938e-03
## pop_centerPrince Edward Island 1.782206e-01
## pop_centerRural areas and small population centres (non CMA/CA) 1.220147e-01
## marital_statusLiving common-law 5.304732e-01
## marital_statusMarried 6.522261e-01
## marital_statusSeparated -2.739755e-01
## marital_statusSingle, never married -3.773260e-02
## marital_statusWidowed -2.325970e-01
```


## sexMale	-8.993412e-02
## religion_participationAt least once a month	1.008783e-01
## religion_participationAt least once a week	1.235972e-01
## religion_participationNot at all	-1.282438e-01
## religion_participationOnce or twice a year	-9.153288e-02
## educationCollege, CEGEP or other non-university certificate or di...	-2.021141e-02
## educationHigh school diploma or a high school equivalency certificate	-3.681235e-02
## educationLess than high school diploma or its equivalent	1.658869e-01
## educationTrade certificate or diploma	-1.618916e-02
## educationUniversity certificate or diploma below the bachelor's level	7.217443e-02
## educationUniversity certificate, diploma or degree above the bach...	-2.994744e-02
## provinceBritish Columbia	2.121650e-02
## provinceManitoba	2.640103e-02
## provinceNew Brunswick	1.841262e-01
## provinceNewfoundland and Labrador	2.132207e-01
## provinceNova Scotia	7.374044e-02
## provinceOntario	6.142705e-02
## provinceQuebec	1.432766e-01
## provinceSaskatchewan	1.122603e-01
## average_income	-3.275351e-07
##	Std. Error
## (Intercept)	9.576157e-02
## age	1.075961e-03
## pop_centerPrince Edward Island	8.236381e-02
## pop_centerRural areas and small population centres (non CMA/CA)	3.596018e-02
## marital_statusLiving common-law	6.269540e-02
## marital_statusMarried	5.330442e-02
## marital_statusSeparated	8.706360e-02
## marital_statusSingle, never married	5.786634e-02
## marital_statusWidowed	1.001983e-01
## sexMale	2.671910e-02
## religion_participationAt least once a month	5.927306e-02
## religion_participationAt least once a week	5.368389e-02
## religion_participationNot at all	4.339168e-02
## religion_participationOnce or twice a year	4.999637e-02
## educationCollege, CEGEP or other non-university certificate or di...	3.834729e-02
## educationHigh school diploma or a high school equivalency certificate	3.976783e-02
## educationLess than high school diploma or its equivalent	5.480132e-02
## educationTrade certificate or diploma	5.543065e-02
## educationUniversity certificate or diploma below the bachelor's level	7.440149e-02
## educationUniversity certificate, diploma or degree above the bach...	4.898731e-02
## provinceBritish Columbia	5.715657e-02
## provinceManitoba	6.940694e-02
## provinceNew Brunswick	6.900704e-02
## provinceNewfoundland and Labrador	7.230287e-02
## provinceNova Scotia	6.692350e-02
## provinceOntario	4.995471e-02
## provinceQuebec	5.364222e-02
## provinceSaskatchewan	6.912691e-02
## average_income	6.694094e-07
##	t value
## (Intercept)	79.8009510
## age	2.8355480
## pop_centerPrince Edward Island	2.1638222

## pop_centerRural areas and small population centres (non CMA/CA)	3.3930509
## marital_statusLiving common-law	8.4611187
## marital_statusMarried	12.2358709
## marital_statusSeparated	-3.1468428
## marital_statusSingle, never married	-0.6520647
## marital_statusWidowed	-2.3213674
## sexMale	-3.3659117
## religion_participationAt least once a month	1.7019246
## religion_participationAt least once a week	2.3023139
## religion_participationNot at all	-2.9554941
## religion_participationOnce or twice a year	-1.8307907
## educationCollege, CEGEP or other non-university certificate or di...	-0.5270623
## educationHigh school diploma or a high school equivalency certificate	-0.9256817
## educationLess than high school diploma or its equivalent	3.0270595
## educationTrade certificate or diploma	-0.2920615
## educationUniversity certificate or diploma below the bachelor's level	0.9700670
## educationUniversity certificate, diploma or degree above the bach...	-0.6113306
## provinceBritish Columbia	0.3711996
## provinceManitoba	0.3803803
## provinceNew Brunswick	2.6682230
## provinceNewfoundland and Labrador	2.9489938
## provinceNova Scotia	1.1018616
## provinceOntario	1.2296547
## provinceQuebec	2.6709676
## provinceSaskatchewan	1.6239746
## average_income	-0.4892896
##	Pr(> t)
## (Intercept)	0.000000e+00
## age	4.581954e-03
## pop_centerPrince Edward Island	3.049661e-02
## pop_centerRural areas and small population centres (non CMA/CA)	6.933017e-04
## marital_statusLiving common-law	2.936884e-17
## marital_statusMarried	3.117901e-34
## marital_statusSeparated	1.654257e-03
## marital_statusSingle, never married	5.143712e-01
## marital_statusWidowed	2.028279e-02
## sexMale	7.651759e-04
## religion_participationAt least once a month	8.879401e-02
## religion_participationAt least once a week	2.133359e-02
## religion_participationNot at all	3.127430e-03
## religion_participationOnce or twice a year	6.715525e-02
## educationCollege, CEGEP or other non-university certificate or di...	5.981596e-01
## educationHigh school diploma or a high school equivalency certificate	3.546290e-01
## educationLess than high school diploma or its equivalent	2.474411e-03
## educationTrade certificate or diploma	7.702444e-01
## educationUniversity certificate or diploma below the bachelor's level	3.320316e-01
## educationUniversity certificate, diploma or degree above the bach...	5.409917e-01
## provinceBritish Columbia	7.104951e-01
## provinceManitoba	7.036695e-01
## provinceNew Brunswick	7.635058e-03
## provinceNewfoundland and Labrador	3.193921e-03
## provinceNova Scotia	2.705428e-01
## provinceOntario	2.188493e-01
## provinceQuebec	7.572948e-03

```
## provinceSaskatchewan 1.044061e-01
## average_income 6.246451e-01
## [1] 0.06624723
```

Compared with the last model, the p values of coefficients are smaller and the p value for the whole model is small as well. It seems like our model has been improved. However, the r-squared computed in the last line becomes 0.05997 and is smaller than before, which means less variations can be explained by the alternative model. It may happens because of the smaller sample size.

```
## Estimate
## (Intercept) 7.641864e+00
## age 3.050938e-03
## pop_centerPrince Edward Island 1.782206e-01
## pop_centerRural areas and small population centres (non CMA/CA) 1.220147e-01
## marital_statusLiving common-law 5.304732e-01
## marital_statusMarried 6.522261e-01
## marital_statusSeparated -2.739755e-01
## marital_statusSingle, never married -3.773260e-02
## marital_statusWidowed -2.325970e-01
## sexMale -8.993412e-02
## religion_participationAt least once a month 1.008783e-01
## religion_participationAt least once a week 1.235972e-01
## religion_participationNot at all -1.282438e-01
## religion_participationOnce or twice a year -9.153288e-02
## educationCollege, CEGEP or other non-university certificate or di... -2.021141e-02
## educationHigh school diploma or a high school equivalency certificate -3.681235e-02
## educationLess than high school diploma or its equivalent 1.658869e-01
## educationTrade certificate or diploma -1.618916e-02
## educationUniversity certificate or diploma below the bachelor's level 7.217443e-02
## educationUniversity certificate, diploma or degree above the bach... -2.994744e-02
## provinceBritish Columbia 2.121650e-02
## provinceManitoba 2.640103e-02
## provinceNew Brunswick 1.841262e-01
## provinceNewfoundland and Labrador 2.132207e-01
## provinceNova Scotia 7.374044e-02
## provinceOntario 6.142705e-02
## provinceQuebec 1.432766e-01
## provinceSaskatchewan 1.122603e-01
## average_income -3.275351e-07
## Std. Error
## (Intercept) 0
## age 0
## pop_centerPrince Edward Island 0
## pop_centerRural areas and small population centres (non CMA/CA) 0
## marital_statusLiving common-law 0
## marital_statusMarried 0
## marital_statusSeparated 0
## marital_statusSingle, never married 0
## marital_statusWidowed 0
## sexMale 0
## religion_participationAt least once a month 0
## religion_participationAt least once a week 0
## religion_participationNot at all 0
## religion_participationOnce or twice a year 0
```

## educationCollege, CEGEP or other non-university certificate or di...	0
## educationHigh school diploma or a high school equivalency certificate	0
## educationLess than high school diploma or its equivalent	0
## educationTrade certificate or diploma	0
## educationUniversity certificate or diploma below the bachelor's level	0
## educationUniversity certificate, diploma or degree above the bach...	0
## provinceBritish Columbia	0
## provinceManitoba	0
## provinceNew Brunswick	0
## provinceNewfoundland and Labrador	0
## provinceNova Scotia	0
## provinceOntario	0
## provinceQuebec	0
## provinceSaskatchewan	0
## average_income	0
##	
	t value
## (Intercept)	Inf
## age	Inf
## pop_centerPrince Edward Island	Inf
## pop_centerRural areas and small population centres (non CMA/CA)	Inf
## marital_statusLiving common-law	Inf
## marital_statusMarried	Inf
## marital_statusSeparated	-Inf
## marital_statusSingle, never married	-Inf
## marital_statusWidowed	-Inf
## sexMale	-Inf
## religion_participationAt least once a month	Inf
## religion_participationAt least once a week	Inf
## religion_participationNot at all	-Inf
## religion_participationOnce or twice a year	-Inf
## educationCollege, CEGEP or other non-university certificate or di...	-Inf
## educationHigh school diploma or a high school equivalency certificate	-Inf
## educationLess than high school diploma or its equivalent	Inf
## educationTrade certificate or diploma	-Inf
## educationUniversity certificate or diploma below the bachelor's level	Inf
## educationUniversity certificate, diploma or degree above the bach...	-Inf
## provinceBritish Columbia	Inf
## provinceManitoba	Inf
## provinceNew Brunswick	Inf
## provinceNewfoundland and Labrador	Inf
## provinceNova Scotia	Inf
## provinceOntario	Inf
## provinceQuebec	Inf
## provinceSaskatchewan	Inf
## average_income	-Inf
##	Pr(> t)
## (Intercept)	0
## age	0
## pop_centerPrince Edward Island	0
## pop_centerRural areas and small population centres (non CMA/CA)	0
## marital_statusLiving common-law	0
## marital_statusMarried	0
## marital_statusSeparated	0
## marital_statusSingle, never married	0

```

## marital_statusWidowed 0
## sexMale 0
## religion_participationAt least once a month 0
## religion_participationAt least once a week 0
## religion_participationNot at all 0
## religion_participationOnce or twice a year 0
## educationCollege, CEGEP or other non-university certificate or di... 0
## educationHigh school diploma or a high school equivalency certificate 0
## educationLess than high school diploma or its equivalent 0
## educationTrade certificate or diploma 0
## educationUniversity certificate or diploma below the bachelor's level 0
## educationUniversity certificate, diploma or degree above the bach... 0
## provinceBritish Columbia 0
## provinceManitoba 0
## provinceNew Brunswick 0
## provinceNewfoundland and Labrador 0
## provinceNova Scotia 0
## provinceOntario 0
## provinceQuebec 0
## provinceSaskatchewan 0
## average_income 0

```

In this survey, we use a stratified sampling method to divide the sample by province. This sampling method is appropriate because each province has an extremely different sample size. For instance, there are only 453 respondents selected into our model in Prince Edward Island while there are 3535 in Ontario. By dividing sampling into provinces, it ensures a high degree of representativeness of all the strata in the population. When the population size is weighted in each province, the standard error may be reduced and it gives smaller error in estimation. From the table ran by R, we see that the p values of variable coefficients are much smaller, which means the variables are more significant than before.

There are two caveats in this data set. One is that it does not include the respondents outside Canada, which leads to small sample size. Another is that the variable average_income is not accurate. So there will occur some errors in our model.

Results

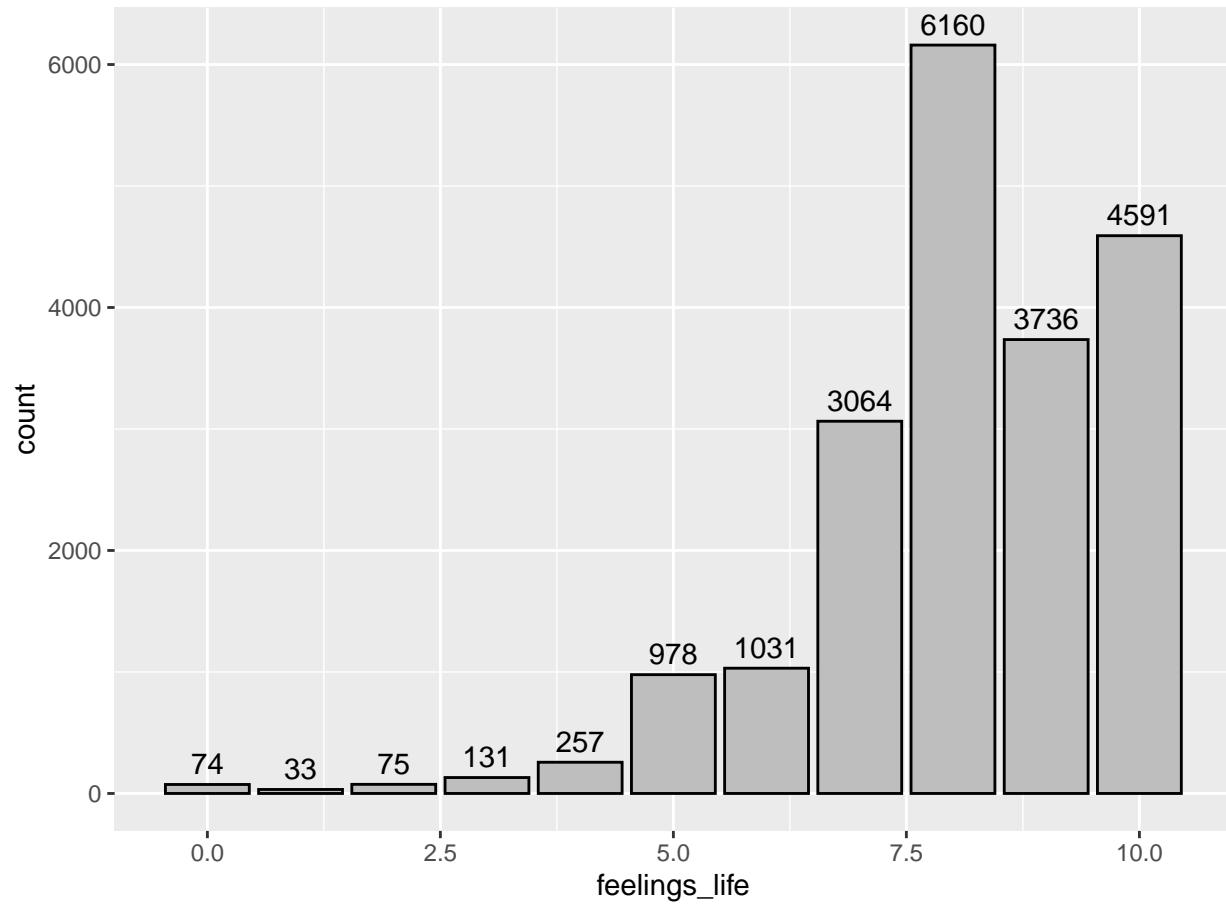


Figure1

By looking at figure1, we find out that most people's life satisfaction value is 8(6160), followed by 10(4591) and 9(3736). (higher value means higher life satisfaction) It means the majority of people have high life satisfaction, and there are only a minority of people who feel so poor about life.

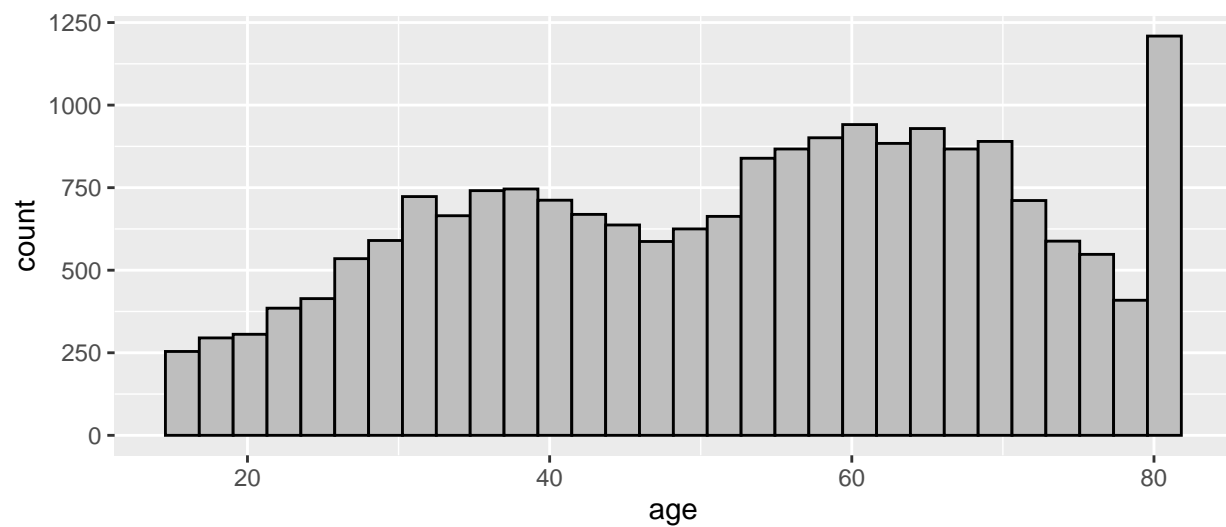
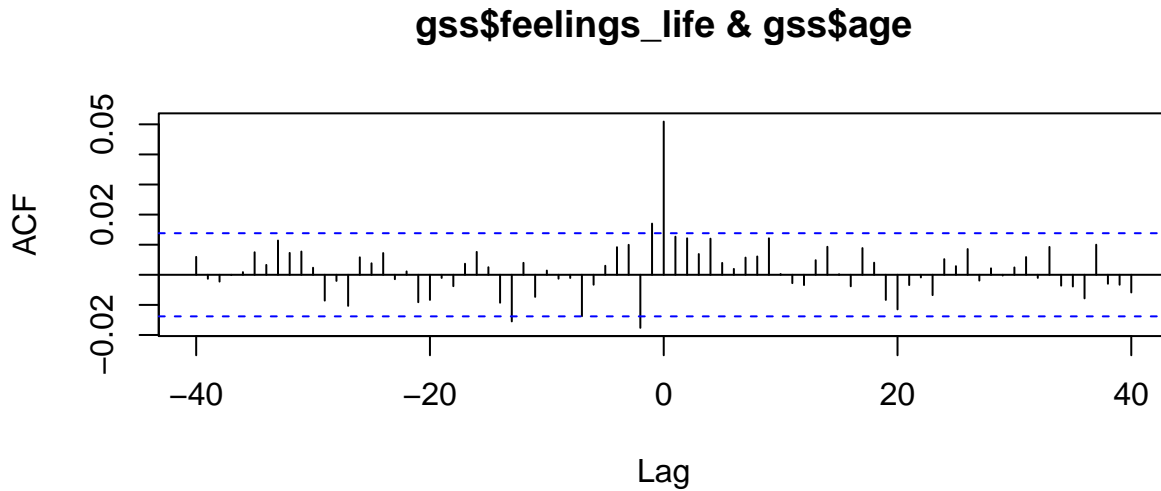


Figure2

Figure 3



Based on figure2 and figure3, we find out that age has no obvious influence on people's life satisfaction since it follows the trend of total population, and they are not highly-correlated.

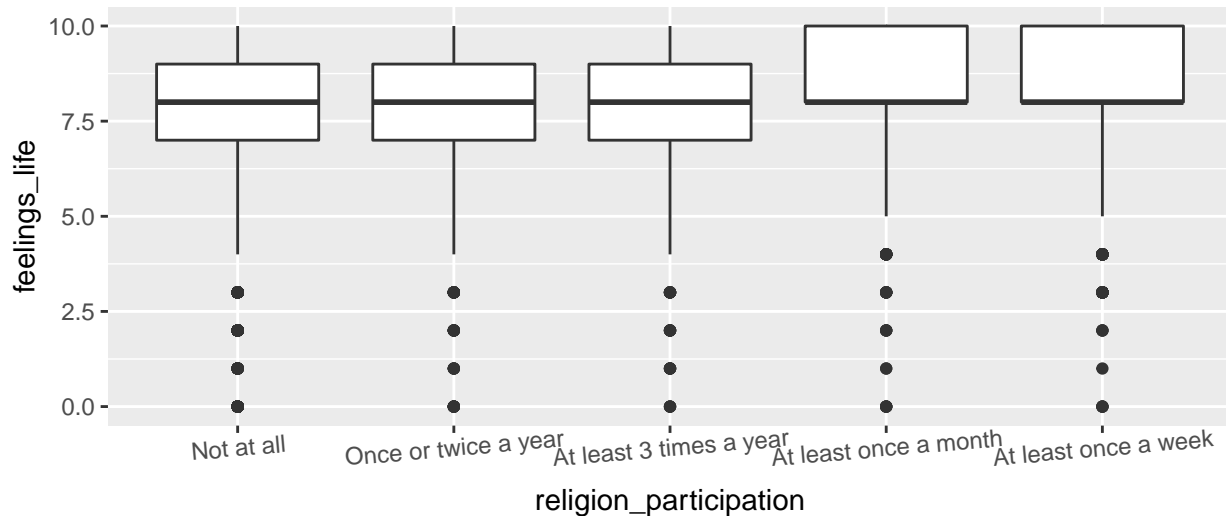


Figure4

By looking at the box plot of religion participation in Figure4, we find out that people with frequent religion participation feel happier about life. It is shown that people with frequent religious participation(at least once a month & at least once a week) have higher life satisfaction due to higher median and mean of feelings_life in box plot.

Table 1

```
## # A tibble: 11 x 2
##   feelings_life mean_income
##   <dbl>         <dbl>
## 1         1      44940.
## 2         0      38904.
## 3         2      33169.
## 4         4      30529.
## 5         3      30282.
```

##	6	6	29313.
##	7	5	28883.
##	8	7	27081.
##	9	8	25042.
##	10	10	23907.
##	11	9	23799.

Based on table1, we are surprised to find that the group of high income have low satisfaction with life. The mean income of people who have the worst life satisfaction(1) is the highest(44940), and the mean income of people who have the best life satisfaction(9) is the lowest(23799). It means the group of high income tend to have more worries that lead to a decrease in their life satisfaction.

Discussion

Figure1 shows the number of people for each level of feelings of life from 0 to 10. By the result, we can see that most people's feeling of life is over 6, while the total number of feelings of life small or equal 6 is 2579, it is the small that any number of other levels of feeling of life. Which 7 is 3060, 8 is 6160, 9 is 3736, 10 is 4591. This can be understood by most people very satisfied with their life, but it also may be because of some misleading survey.

In Figure 3 shows the ACF between feelings_life and age, it shows that age and feelings_life do not have highly-correlated. In other words, age is not a main variable that affects people's feelings of life.

By Figure4, Table 2 and model results, they show there are three main types of variables that have an impact on people's feeling of life. Which are religion participation, income, and marital status. In Figure4, we can easily see that for religious participation, the majority of people who have religious participation at least once a month and at least once a week, have a feeling of life is higher than 7.5. Which is higher than people who have religious participation less or at least 3 times a year. In Table 2, it shows the mean of income for each level of feeling of life. By results, we can see that as feeling_life increases from 0 to 10, the mean decreases from 44940 to 23907. Lastly, model results show the estimate, t-value and p-value for each variable we use. Compared to other variables, marital status has lowest p-value overall, which is less than 0.001. Therefore, we can say that the relationship between people's life satisfaction and marital status or religious participation are acceptable.

For religion participation and marital status, it is very easy to understand. The goal for religion is to help people become optimistic, to thank everything in our life, so people with frequent religious participation of course will feel happier about life. And for marital status, since relationship between people is a big part of our life, specifically marital will affect our life most. Income is another variable that has an impact on people's feeling of life, but different from the other two, the result it shows is surprising to us. Usually, people will believe that higher income will have higher life satisfaction, because they will have more money to buy something they want. But the result shows the opposite way, in which the group of high income have low satisfaction with life. Moreover, when we have those three types of variables, religion participation and marital status is spiritual, income is materially. This tells us what affects happiness is not the material but the mind, and gives us a caveat: "don't be too obsessed with material things".

However, there may exist some bias in our model. For example, our data of income has "less than 25000". Since we do not have a specific number, so as we use is data, we assume their income is \$25000. This may get us an error, since if a person's income is \$0, but based on the data, which we don't know, we assume that is \$25000. This will give us a big deviation. Moreover, when we calculate their income per hour, it has work hours equal to 0. When we use this data, we change it from 0 to 0.01, so it can be used on a denominator. Since the number we use is enough, we believe the model is acceptable, but it may still have errors in the model.

Weaknesses

There are three weaknesses of the data, one is data has a small R-squared, which is 0.06625, another is the small sample population and the last one is the inaccurate average income.

Firstly, r-squared is a goodness-of-fit measure for linear regression models. It measures the strength of the relationship between the predictor and the dependent variables, which means that the model with higher R-squared is more acceptable for the model. However, small R-squared values are not always a problem, and high R-squared values are not necessarily good. Since the p-level in our model is small as well, which means coefficients of each variable are significant, we believe we can still use the linear regression model to do the analysis. Another weakness is limitations of the sample population. Since the GSS program survey is volunteer-based, most people who have low life quality may not be willing to do the survey, which will make our results biased.

Last but not least, we only have the data on the range of income instead of accurate income amounts. So when we use `average_income` to analyze, the mean of income is applied. For example, if the respondent's total income is '\$50,000 to \$74,999', we use $(50000+74999)/2=\$62499.5$ as the respondent's average income. As a result, there may exist errors in our model.

Next Steps

In order to solve our weakness, there are three things we should do, increase the sample size, make income numbers be more specific in the questionnaire, and add more ways to collect data. About increasing the sample size, we want to change our target population from Canadian to people in the whole world. We will choose some other countries that have the economic environment as Canada, so it will increase our R-squared and our result has higher credibility. Also we want to avoid any error caused by deviation of income, we will try to get a more specific number of income or make the range of each level of income become smaller. Lastly, since most people's life satisfaction value is a high number, we are unsure if it is really because the majority of people have high life satisfaction, or we are missing some type of group of people because of our way to collect data. So we will try to add more ways to collect data, to make sure diversity of the crowd.

References

1. Alboukadel Kassambara (2020). `ggpubr`: 'ggplot2' Based Publication Ready Plots. R package version 0.4.0. <https://CRAN.R-project.org/package=ggpubr>
2. Beaupre, P. (2017). General Social Survey. https://sda-artsci-utoronto-ca.myaccess.library.utoronto.ca/sdaweb/dli2/gss/gss31/gss31/more_doc/GSS31_User_Guide.pdf.
3. Hadley Wickham, Romain François, Lionel Henry and Kirill Müller (2020). `dplyr`: A Grammar of Data Manipulation. R package version 1.0.2. <https://CRAN.R-project.org/package=dplyr>
4. Pinheiro J, Bates D, DebRoy S, Sarkar D, R Core Team (2020). *nlme: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-149, <URL: <https://CRAN.R-project.org/package=nlme>>.
5. Technology, A. K. through. (2017). Data Centre. CHASS Data Centre. <http://dc.chass.utoronto.ca/myaccess.html>.
6. T. Lumley (2020) "survey: analysis of complex survey samples". R package version 4.0.
7. Wickham et al., (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686, <https://doi.org/10.21105/joss.01686>
8. Yihui Xie (2020). `knitr`: A General-Purpose Package for Dynamic Report Generation in R. R package version 1.30.

Appendix

The link of repository: 'https://github.com/ShilunDai/problemset2_304'