Relationship Between Education, Marriage, Self-Rated Mental Health and Life Satisfication

WENYU SHU 1004951082, SHIHAN WANG 1005165063, KEFAN CAI 1004819949

Oct.19, 2020

Abstract:

This research using multiple linear regression to show how education, marriage, and self-rated mental health contribute to life satisfaction scores. The information is selected then analyzed from the GSS2017 data file. The study found that there is a negative relationship between education and feelings of life, and a positive relationships between both marriage and self-rated mental health to feelings of life.

Introduction:

Under the fast pace of nowadays' life, many people are under stress from multiple aspects of their lives. It is significant to know the life satisfaction of people to improve their physical, mental, and social well-being in the future. Through collecting and processing data of people's education, marriage, and self-evaluation of their mental health, which are all critical factors through people's lives, and then relate these factors with the self-scored feelings of lives, it can reveal the relationship between them. By constructing a multiple linear regression model, the result shows that a person's life satisfaction is negatively related to their education level, and positively related to the marriage and self-evaluation of mental health.

Data

```
##                           caseid                              age
##                                0                                0
##                  age_first_child       age_youngest_child_under_6
##                             6835                            18488
##                   total_children           age_start_relationship
##                               19                            18566
##             age_at_first_marriage               age_at_first_birth
##                            15248                             7865
##          distance_between_houses age_youngest_child_returned_work
##                            19476                            19466
##                     feelings_life                              sex
##                              271                                0
##                place_birth_canada               place_birth_father
##                                97                              203
##                place_birth_mother         place_birth_macro_region
##                                47                            16457
##             place_birth_province              year_arrived_canada
##                             4289                            16550
##                         province                           region
##                                0                                0
##                       pop_center                  marital_status
##                                0                                7
##                       aboriginal                     vis_minority
##                             3855                              140
##                  age_immigration                 landed_immigrant
##                            17225                            16450
##              citizenship_status                        education
##                             1143                              341
##                         own_rent               living_arrangement
##                              120                                0
##                          hh_type                          hh_size
##                               76                                0
##             partner_birth_country           partner_birth_province
##                             7697                             7883
##              partner_vis_minority                     partner_sex
##                             7719                            20407
##                 partner_education             average_hours_worked
##                             8259                             7166
##                  worked_last_week            partner_main_activity
##                               23                             7907
##                self_rated_health         self_rated_mental_health
##                               99                              106
##          religion_has_affiliation              regilion_importance
##                              282                              253
##                    language_home               language_knowledge
##                              448                              105
##                    income_family                 income_respondent
##                                0                                0
##                       occupation                childcare_regular
##                             7297                            18756
```

```
##                   childcare_type           childcare_monthly_cost
##                            19365                            19962
##              ever_fathered_child                 ever_given_birth
##                            13604                            12769
##          number_of_current_union              lives_with_partner
##                            18600                                0
##            children_in_household  number_total_children_intention
##                                0                            12202
##                has_grandchildren        grandparents_still_living
##                                4                                9
##                     ever_married        current_marriage_is_first
##                                5                            10416
##                number_marriages           religion_participation
##                                0                              199
##       partner_location_residence              full_part_time_work
##                            18978                            18852
##              time_off_work_birth         reason_no_time_off_birth
##                            18855                            20283
##                returned_same_job          satisfied_time_children
##                            19451                            19691
##      provide_or_receive_fin_supp              fin_supp_child_supp
##                            19578                            20057
##               fin_supp_child_exp                     fin_supp_lump
##                            20057                            20057
##                   fin_supp_other                fin_supp_agreement
##                            20057                            19937
##        future_children_intention                          is_male
##                            13438                                0
##                     main_activity                         age_diff
##                            20602                            10430
##      number_total_children_known
##                                0
```

The data used for this model is from the 2017 General Social Survey (GSS) about family changes in variety aspects of Canadian families. It was published by the authority of the Minister responsible for Statistics Canada, and was constructed from February 2nd to November 30th, 2017. (Beaupré, 2020)

The primary objectives of this data is to montior the changes in the living conditions and well-beings of Canadians over time, and provide information for the social policy issues. The target population of this data was all persons being and above 15 years old in Canada, excluding residents of the Yukon, Northwest Territories, Nunavut and full-time residents of institutions. The actual sample size of this data was 20602 while the expected sample size was 20000.(Beaupré, 2020)

The information was collected via computer-assisted telephone interviews, then use stratification to sampling the data. There are 27 sampling strata in total which were divided geographically by each province, and some of the Census Metropolitan Area was considered as separate strata. The next step was using a simple random sample without replacement in each stratum. All respondents were interviewed through phone calls, therefore the families without telephones were excluded from the frame population. To be eligible for the survey, each household was required to have at least one person being or above 15 years old. Eligible respondents then were randomly selected by each household to participate in the interview. The response rate for this data was 52.4%. For those non-responding telephone numbers, their weights were adjusted to represent by the responded telephone numbers.(Beaupré, 2020)

The data provided by the GSS2017 survey had some postive political impacts on effecting program and policies, such as parental benefits, child care strategies, child custody, and spousal support programs.(General Social Survey - Family (GSS) 2019) However, the method of using computer-assisted telephone interviews had some drawbacks that it neglected some potential respondents who only have cell phones, such as young people. The responsed rate also tends to be lower, compared with in-person interviews, thus there were limitations on the amount and type of information that can be collected.(General Social Survey: An Overview, 2019 2019)

Among all the variables included in the data, this model chooses feelings_life as the dependent Y variable, education, marital_status, self_rated_mental_health as the independent X variables, and then determine whether there exist some relationships between the Y and the Xs . The three X variables are conprehensive and the amount of responded observations are sufficient that they only have a few missing observations, furthermore they are somehow related to the Y variable by common sense. In this model, all the missing observations are removed from the data, and the data is filtered so that it only contains observations in Ontario. To be clearer about the education variable groups, it is classified into three new groups: Bachelor’s degree or higher, Below Bachelor’s degree, and other non-university certificates.

Model

Multiple linear Regression Model studies the relationship between one dependent variable(Y) and multiple independent variables(X).To be more elaborate, Yi stands for the value of the dependent variable taken from the observation unit i, while Xi means the value of independent variables for ith observation. Besides,beta0 can be considered as an intercept(constant) and beta1 to beta 3 represent coefficient with respect to corresponding independent variables Xi. In addition, epsilon means error term for the model, it’s independent and identically distributed(iid) with variance theta squares and mean 0. Basically, beta1 means one unit change in X1 will lead to bata1 units change in Y(The logic is the same for beta 2 and 3.)

Multiple Linear Regression model can be considered as an appropriate model to predict which factors would affect the response variable (feelings of life). Choosing three independent variables including marital status, self-rated mental health, and education is because whether people are happy in their marriages, their own assessment of their mental health and their level of education can lead to different perceptions and definitions of feelings of life. It is worth noting that the classification of people’s educational credentials into three categories is intended to provide a more intuitive analysis of the relationship between people’s level of education and happiness index.

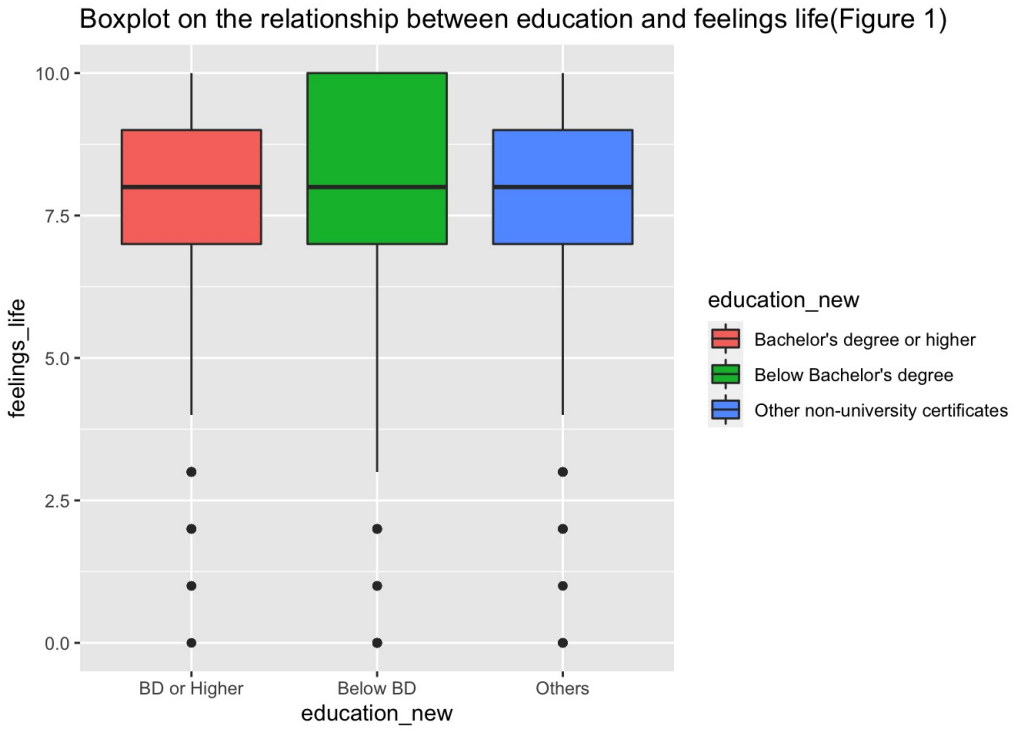
mathematical notation：

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \epsilon_i$$

Compare Multiple linear Regression model with others，it can fit data more accurately and give a more precise association of predictors(Xi) with the outcome(Y)，which is consistent with the goal. In terms of the Single Linear Regression Model, one drawback might include only one X and Y relationship can be studied, but in the reality, it's basically a model of several X's acting on Y together. However, Multivariate models also have disadvantages, such as it's easy to occur that two independent variables can affect each other, resulting in multicollinearity.

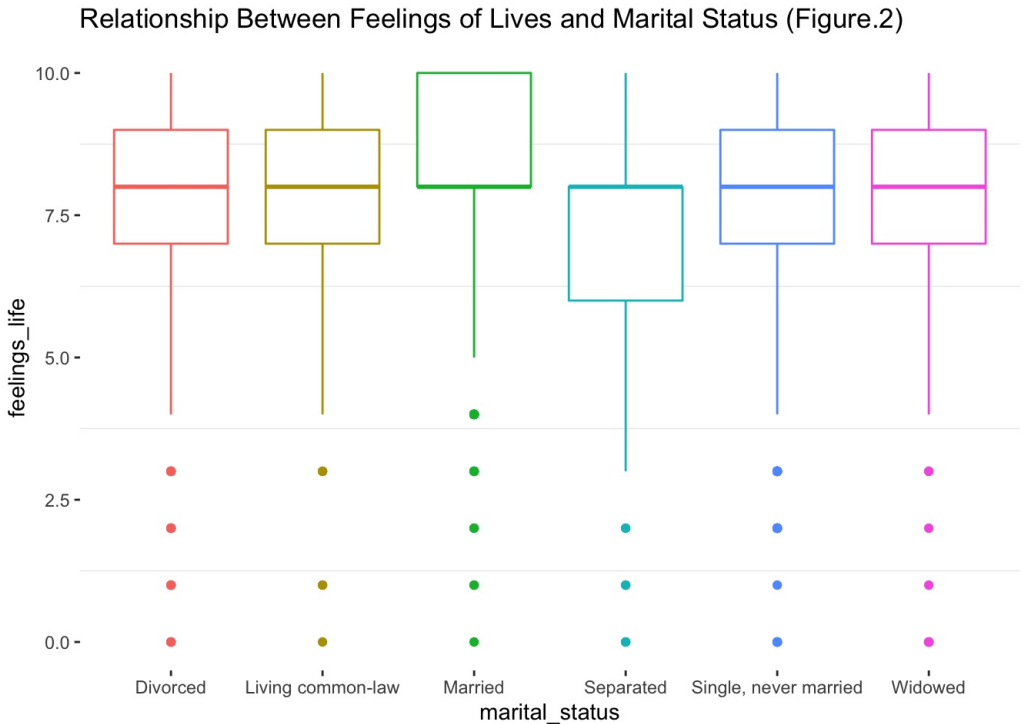
Results

#Relationship between feelings of lives and education:



Firstly, one new column of variable "education" is mutated from the data of Ontario. "education" is reclassified into three groups: Bachelor's degree or higher, Below Bachelor's degree and other non-university certificates. Then, delete the missing data(NA). According to the boxplot above, the mean value for the box plot is about 8. The upper quartile of people who has an education diploma under a Bachelor's degree reaches 10, which is greater than that of the other two(about 8.75). Moreover, this bachelor's degree or higher boxplot is left-skewed which implies most people's feelings life is greater than the median value. In terms of the other two, they have roughly the same boxplots, with four outliers at feelings of life less than 3. Overall, It can be concluded that people with relatively less education are more likely to be satisfied with their lives, Probably because their lives are relatively less stressful.

#Relationship between feelings of lives and marital status of respondents:



From the above Figure.2, it shows the score relationship between respondents' marital status and their feelings of life as a whole. The respondents are separated into six groups based on their different marital statuses in this plot. The mean of the life-feeling score is 8 out of 10, and the range of the score is approximately 2

for all six groups with the minimum score around 4 and the maximum score at 10. The groups of the divorced, the living common-law, the single (never married), and the widowed present similar IQRs that the 25 to 75 percentile of respondents score their lives between 7 to 9 out of 10. The married group appears to be more satisfied with their lives that 25% of them scored their lives 10 out of 10. In contrast, the separated group is less satisfied with their lives compare with the other groups. Only the top 25% of respondents in this group rated their lives at 8 out of 10 or higher, there is 25% of them rated their lives below 3 out of 10. In addition to the graph, there are outliers among each group who rated their lives lower than the minimum scores of each group.

#Relationship between feelings of lives and self-rated mental health:

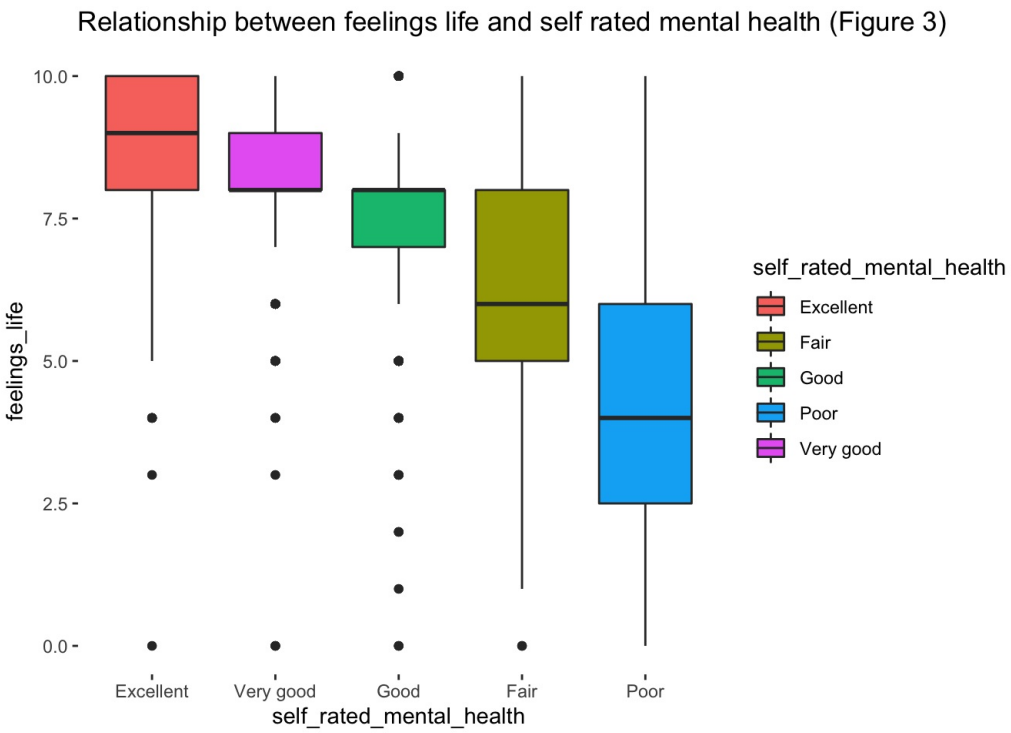
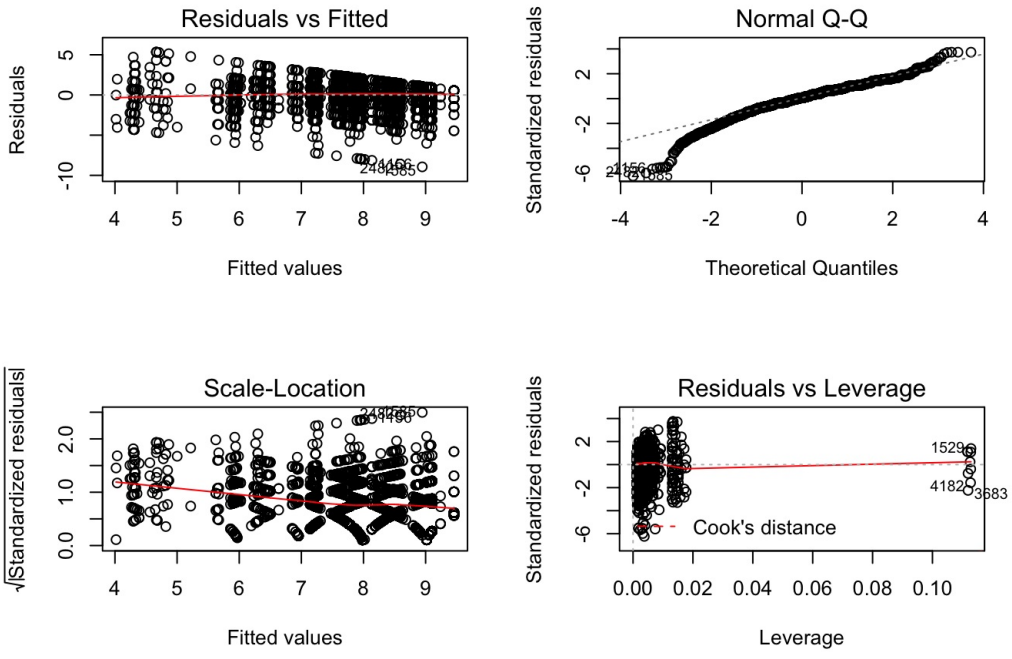


Figure 3 above illustrates the relationship between feelings of life and their self-rated mental health of 5 defined levels, that are excellent, very good, good, fair, and poor. It shows that people who rate their mental health as poor have the lowest median value of feelings of life, about 4, with the largest IQR and whisker. People who rate themselves excellent have the highest median and 75% quantile value of feelings of life. Also, a pattern that the better they rate their mental health, the higher feelings of life value they will have can be concluded from the graph.

#Multiple Linear Regression Model



Summary of Multiple Linear Regression Model (table 1)

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	8.0970	0.4885	16.5759	0.0000
self_rated_mental_healthExcellent	0.4222	0.4823	0.8755	0.3813
self_rated_mental_healthFair	-2.1821	0.4869	-4.4813	0.0000
self_rated_mental_healthGood	0.8990	0.4822	1.8642	0.0624

self_rated_mental_healthGood	-0.8990	0.4822	-1.8642	0.0624
self_rated_mental_healthPoor	-3.8158	0.5055	-7.5485	0.0000
self_rated_mental_healthVery good	-0.1868	0.4820	-0.3875	0.6984
educationCollege, CEGEP or other non-university certificate or di…	0.0471	0.0575	0.8193	0.4127
educationHigh school diploma or a high school equivalency certificate	0.0283	0.0586	0.4833	0.6289
educationLess than high school diploma or its equivalent	0.3997	0.0721	5.5479	0.0000
educationUniversity certificate, diploma or degree above the bach…	-0.0403	0.0703	-0.5725	0.5670
marital_statusLiving common-law	0.3207	0.1038	3.0885	0.0020
marital_statusMarried	0.5377	0.0754	7.1293	0.0000
marital_statusSeparated	-0.2920	0.1260	-2.3170	0.0205
marital_statusSingle, never married	-0.0151	0.0804	-0.1879	0.8510
marital_statusWidowed	0.0268	0.0969	0.2762	0.7824

For the model checking part, there are several assumptions that need to be satisfied. In terms of the linearity(Residuals VS.Fitted plot), it shows that the model follows the linearity assumption since there is a red horizontal line at 0. It's also shown on the second QQ plot that the model is slightly left-skewed, but most of the dots are still on the dotted line.As for Scale-location plot, the overall distribution of points is scattered, meaning the data follows the constant variance assumption. Hence overall the model constructed above can be considered as a good model.

According to the summary of the Multiple linear regression Model, by analyzing the p-values, it shows that people who rate their mental health as fair or poor, people whose education certificates is less than high school diploma, people whose marital status shows married can be considered as significant independent variables contributes to the model.

There are some weakness of the model including not all the independent variables are useful since their p_values are too big, however, the good points of the model is that there is no multicollinearity between predictors which satisfied the model assumptions.

Discussion

This project mainly discusses the relationship between feelings of life and people's marital status，self rated mental health and education level. For marital status, people of six groups, which are divorced, living common-law, married, separated, single(never married), and widowed, have the same median and IQR value of feelings of life. People who married have a relatively higher 75% and 25% quantile value. On the other hand, separated people have a lower 75% and 25% quantile value. For self-rated mental health, the highest median value exists in excellent and the lowest median value is in people who rate low mental health. It also shows a pattern that the better they rate their mental health, the higher feelings of life value they will have. Furthermore, education level has been divided into 3 types, Bachelor's degree or higher, below Bachelor's degree and other non-university certification, to see their influence on feelings of life. A conclusion that generally people with lower education levels have higher score of live feelings can be drawn from the model.

In terms of Multiple Linear Regression Model part, it is been concluded that severel predictors can be used to perdic peoples feelings of life including people rating theri metal health as poor or fair, people whose education is less than high school certificates as well as people who are married.

Weaknesses

One weakness is that the data has many missing data(NA) so a lot of people are removed from some analysis. Moreover, people may be dishonest when taking the survey so that the correctness of the data is questionable. Therefore the reliability of the results is not strong. Since the survey was made by landline phone calls, some young people who don't have landline phones may be ignored. It is also a drawback of the survey.

Next Steps

Some further steps should be taken to make the model more complete and reliable. To improve the data, more survey methods can be used. For example, an online survey by e-mail can be taken for people without landline phones and surveys by mails or door knocks can be used for the elderly who don't know how to use a phone or computer. Another improvement is that the sampling method can be changed to generate data that can lead to a more general and correct result.

#references:

```
##
## To cite ggplot2 in publications, please use:
##
##   H. Wickham. ggplot2: Elegant Graphics for Data Analysis.
##   Springer-Verlag New York, 2016.
##
## A BibTeX entry for LaTeX users is
##
##   @Book{,
##     author = {Hadley Wickham},
##     title = {ggplot2: Elegant Graphics for Data Analysis},
##     publisher = {Springer-Verlag New York},
##     year = {2016},
##     isbn = {978-3-319-24277-4},
##     url = {https://ggplot2.tidyverse.org},
##   }
```

```
##
## To cite package 'dplyr' in publications use:
##
##   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
##
## A BibTeX entry for LaTeX users is
##
##   @Manual{,
##     title = {dplyr: A Grammar of Data Manipulation},
##     author = {Hadley Wickham and Romain François and Lionel {
##               Henry} and Kirill Müller},
##     year = {2020},
##     note = {R package version 1.0.2},
##     url = {https://CRAN.R-project.org/package=dplyr},
##   }
```

```
##
## To cite the 'knitr' package in publications use:
##
##   Yihui Xie (2020). knitr: A General-Purpose Package for Dynamic Report
##   Generation in R. R package version 1.29.
##
##   Yihui Xie (2015) Dynamic Documents with R and knitr. 2nd edition.
##   Chapman and Hall/CRC. ISBN 978-1498716963
##
##   Yihui Xie (2014) knitr: A Comprehensive Tool for Reproducible
##   Research in R. In Victoria Stodden, Friedrich Leisch and Roger D.
##   Peng, editors, Implementing Reproducible Computational Research.
##   Chapman and Hall/CRC. ISBN 978-1466561595
##
## To see these entries in BibTeX format, use 'print(<citation>,
## bibtex=TRUE)', 'toBibtex(.)', or set
## 'options(citation.bibtex.max=999)'.
```

```
##
##   Wickham et al., (2019). Welcome to the tidyverse. Journal of Open
##   Source Software, 4(43), 1686, https://doi.org/10.21105/joss.01686
##
## A BibTeX entry for LaTeX users is
##
##   @Article{,
##     title = {Welcome to the {tidyverse}},
##     author = {Hadley Wickham and Mara Averick and Jennifer Bryan and Winston Chang and Lucy D'Agostino McGowan and Romain François and Garrett Grolemund and Alex Hayes and Lionel Henry and Jim Hester and Max Kuhn and Thomas L in Pedersen and Evan Miller and Stephan Milton Bache and Kirill Müller and Jeroen Ooms and David Robinson and Dana Paige Seidel and Vitalie Spinu and Kohske Takahashi and Davis Vaughan and Claus Wilke and Kara Woo and Hiroaki Yut ani},
##     year = {2019},
##     journal = {Journal of Open Source Software},
##     volume = {4},
##     number = {43},
##     pages = {1686},
##     doi = {10.21105/joss.01686},
##   }
```

```
##
## To cite R in publications use:
##
##   R Core Team (2019). R: A language and environment for statistical
##   computing. R Foundation for Statistical Computing, Vienna, Austria.
##   URL https://www.R-project.org/.
##
## A BibTeX entry for LaTeX users is
##
##   @Manual{,
##     title = {R: A Language and Environment for Statistical Computing},
##     author = {{R Core Team}},
##     organization = {R Foundation for Statistical Computing},
##     address = {Vienna, Austria},
##     year = {2019},
##     url = {https://www.R-project.org/},
##   }
##
## We have invested a lot of time and effort in creating R, please cite it
## when using it for data analysis. See also 'citation("pkgname")' for
## citing R packages.
```


#Data: Beaupré, P. (2020). General Social Survey Cycle 31 : Families Public Use Microdata File Documentation and User's Guide (Vol. 2019001). Ottawa: Authority of the Minister responsible for Statistics Canada. Retrieved October 19, 2020.

General Social Survey - Family (GSS). (2019, February 06). Retrieved October 19, 2020, from https://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey (https://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey)

General Social Survey: An Overview, 2019. (2019, February 20). Retrieved October 19, 2020, from https://www150.statcan.gc.ca/n1/pub/89f0115x/89f0115x2019001-eng.htm (https://www150.statcan.gc.ca/n1/pub/89f0115x/89f0115x2019001-eng.htm)

#FOR GRAPHS：

MrmolejeMrmoleje 39811 gold badge44 silver badges1818 bronze badges, Arg0naut91arg0naut91 12.2k11 gold badge1010 silver badges2929 bronze badges, LyzandeRLyzandeR 32.7k1111 gold badges5959 silver badges7676 bronze badges, & Akrunakrun 600k2121 gold badges334334 silver badges435435 bronze badges. (2019, June 01). Mutate multiple variables based on a given condition. Retrieved October 19, 2020, from https://stackoverflow.com/questions/55260462/mutate-multiple-variables-based-on-a-given-condition (https://stackoverflow.com/questions/55260462/mutate-multiple-variables-based-on-a-given-condition)

Ggplot2 box plot : Quick start guide - R software and data visualization. (n.d.). Retrieved October 19, 2020, from http://www.sthda.com/english/wiki/ggplot2-box-plot-quick-start-guide-r-software-and-data-visualization (http://www.sthda.com/english/wiki/ggplot2-box-plot-quick-start-guide-r-software-and-data-visualization)

Position scales for discrete data - scale_x_discrete. (n.d.). Retrieved October 19, 2020, from https://ggplot2.tidyverse.org/reference/scale_discrete.html (https://ggplot2.tidyverse.org/reference/scale_discrete.html)

R 学习 - 箱线图. (n.d.). Retrieved October 19, 2020, from https://blog.csdn.net/qazplm12_3/article/details/76474663 (https://blog.csdn.net/qazplm12_3/article/details/76474663)

Kenton, W. (2020, September 21). How Multiple Linear Regression Works. Retrieved October 19, 2020, from https://www.investopedia.com/terms/m/mlr.asp (https://www.investopedia.com/terms/m/mlr.asp)