Exploration of the Relationship Between Self Rated Mental Health and Family Income

Xu Xinyi

Due Date

Exploration of the Relationship Between Self Rated Mental Health and Family Income

Xu Xinyi

Date

Abstract

The data was from Canadian General Social Survey data done in 2017, I chose to analyze the relation ship between family income and self rated mental health.

Here is where you give a brief (one paragraph overview of your entire paper). This should include some background/introduction, some methodology, results and conclusions.

Introduction

The data set collected a lot of information from residents in Canada, and from all these variables, the income and mental health drew my attention as I have heard constantly from media that people tend to feel bad or even commit suicide while experiencing financial hard times. It is likely that financial difficulties could make an individual rather stressful.

Thus I would like to analyze the relationship between the family income and self rated mental health. I would use a linear regression model to fit as I expect that the more a family earns, the less likely a the individual would evaluate his/her metal health as poor.

The result could be important, as if the hypothesis that an less a family earns the more likely an individual is experiencing an mental illness holds, this may allow researchers who study metal health to find their target faster and easier, as well as give government an idea on what proportion of the population they could offer more aid to.

Here is where you should give insight into the setting and introduce the goal of the analysis. Here you can introduce ideas and basic concepts regarding the study setting and the potential model. Again, this is the introduction, so you should be explaining the importance of the work that is ahead and hopefully build some suspense for the reader. You can also highlight what will be included in the subsequent sections.

Data

The data I chose is the GSS data set from 2017. The data set is very interesting as it listed 20602 observations and gathered information on the individual's income, self rated health, living conditions, religion affinity and lots of other variables. It involves a lot of interesting factors and information that drives my attention. Some variables like religion affinity, self rated mental health, religion importance and income soon caught my eyes, there are lot that we could dive into and do investigation upon.

However, for many variables, even at the first glace we could witness that there are lots of data that is NA, which that the analyze done on those variables could be biased as some extreme values or important sets of data could be missing which could be a serious drawback. Except for that, the remaining values are listed out clearly and are all capable for usage.

Introduce the data, explain why it was selected. Make sure to comment on important features and highlight any potential drawbacks to the data.

Methodology

Model

Data was cleaned up following the instruction of the gss_cleaning.R file.

```r
library(janitor)
library(tidyverse)
# install.packages("readr")
library(readr)

unprocessed <- read_csv("gss.csv")
glimpse(unprocessed)
```

```
## Rows: 20,602
## Columns: 81
## $ caseid                       <dbl> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11,...
## $ age                          <dbl> 52.7, 51.1, 63.6, 80.0, 28.0, 63.0...
## $ age_first_child              <dbl> 27, 33, 40, 56, NA, 37, 40, 59, NA...
## $ age_youngest_child_under_6   <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA...
## $ total_children               <dbl> 1, 5, 5, 1, 0, 2, 2, 7, 0, 1, 0, 0...
## $ age_start_relationship       <dbl> NA, NA, NA, NA, 25.3, NA, NA, NA, ...
## $ age_at_first_marriage        <dbl> NA, NA, NA, NA, NA, NA, NA, 22.1, ...
## $ age_at_first_birth           <dbl> 25.9, NA, 23.2, 27.3, NA, 25.8, 18...
## $ distance_between_houses      <dbl> 30, NA, NA, NA, NA, NA, NA, NA, NA...
## $ age_youngest_child_returned_work <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA...
## $ feelings_life                <dbl> 8, 10, 8, 10, 8, 9, 4, 10, 8, 5, 1...
## $ sex                          <chr> "Female", "Male", "Female", "Femal...
## $ place_birth_canada           <chr> "Born in Canada", "Born in Canada"...
## $ place_birth_father           <chr> "Born in Canada", "Born in Canada"...
## $ place_birth_mother           <chr> "Born in Canada", "Born in Canada"...
## $ place_birth_macro_region     <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA...
## $ place_birth_province         <chr> "Quebec", "Ontario", "Ontario", "A...
## $ year_arrived_canada          <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA...
## $ province                     <chr> "Quebec", "Manitoba", "Ontario", "...
## $ region                       <chr> "Quebec", "Prairie region", "Ontar...
## $ pop_center                   <chr> "Larger urban population centres (...
## $ marital_status               <chr> "Single, never married", "Married"...
## $ aboriginal                   <chr> "No", "No", "No", "No", "No", "No"...
## $ vis_minority                 <chr> "Not a visible minority", "Not a v...
## $ age_immigration              <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA...
## $ landed_immigrant             <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA...
## $ citizenship_status           <chr> "By birth", "By birth", "By birth"...
## $ education                    <chr> "High school diploma or a high sch...
## $ own_rent                     <chr> "Owned by you or a member of this ...
## $ living_arrangement           <chr> "Alone", "Spouse only", "Spouse on...
## $ hh_type                      <chr> "Low-rise apartment (less than 5 s...
## $ hh_size                      <dbl> 1, 2, 2, 2, 2, 2, 1, 1, 1, 6, 5, 1...
```

```
## $ partner_birth_country          <chr> "Canada", "Canada", "Canada", "Can...
## $ partner_birth_province         <chr> "Quebec", "Manitoba", "Ontario", "...
## $ partner_vis_minority           <chr> "Not a visible minority", "Not a v...
## $ partner_sex                    <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA...
## $ partner_education              <chr> "Trade certificate or diploma", "B...
## $ average_hours_worked           <chr> "30.0 to 40.0 hours", "50.1 hours ...
## $ worked_last_week               <chr> "Yes", "Yes", "No", "No", "No", "N...
## $ partner_main_activity          <chr> "Working at a paid job or business...
## $ self_rated_health              <chr> "Excellent", "Good", "Very good", ...
## $ self_rated_mental_health       <chr> "Excellent", "Good", "Good", "Very...
## $ religion_has_affiliation       <chr> "Has religious affiliation", "Don'...
## $ regilion_importance            <chr> "Somewhat important", "Don't know"...
## $ language_home                  <chr> "French", "English", "French", "En...
## $ language_knowledge             <chr> "French only", "English only", "Bo...
## $ income_family                  <chr> "$25,000 to $49,999", "$75,000 to ...
## $ income_respondent              <chr> "$25,000 to $49,999", "Less than $...
## $ occupation                     <chr> "Sales and service occupations", "...
## $ childcare_regular              <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA...
## $ childcare_type                 <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA...
## $ childcare_monthly_cost         <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA...
## $ ever_fathered_child            <chr> NA, "Yes", NA, NA, "No", NA, NA, N...
## $ ever_given_birth               <chr> "Yes", NA, "Yes", "Yes", NA, "Yes"...
## $ number_of_current_union        <chr> NA, NA, NA, NA, "Second union", NA...
## $ lives_with_partner             <chr> "No", "No", "No", "No", "Yes", "No...
## $ children_in_household          <chr> "No child", "No child", "No child"...
## $ number_total_children_intention <dbl> NA, NA, NA, NA, 2, NA, NA, NA, NA,...
## $ has_grandchildren              <chr> "No", "Yes", "Yes", "No", "No", "Y...
## $ grandparents_still_living      <chr> "No", "No", "No", "No", "Yes", "No...
## $ ever_married                   <chr> "No", "Yes", "Yes", "Yes", "No", "...
## $ current_marriage_is_first      <chr> NA, "Yes", "Yes", "Yes", NA, "Yes"...
## $ number_marriages               <dbl> 0, 1, 1, 1, 0, 1, 0, 1, 0, 0, 0, 0...
## $ religion_participation         <chr> "Once or twice a year", "Don't kno...
## $ partner_location_residence     <chr> "In the same province", NA, NA, NA...
## $ full_part_time_work            <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA...
## $ time_off_work_birth            <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA...
## $ reason_no_time_off_birth       <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA...
## $ returned_same_job              <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA...
## $ satisfied_time_children        <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA...
## $ provide_or_receive_fin_supp    <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA...
## $ fin_supp_child_supp            <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA...
## $ fin_supp_child_exp             <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA...
## $ fin_supp_lump                  <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA...
## $ fin_supp_other                 <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA...
## $ fin_supp_agreement             <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA...
## $ future_children_intention      <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA...
## $ is_male                        <dbl> 0, 1, 0, 0, 1, 0, 0, 0, 0, 1, 1, 0...
## $ main_activity                  <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA...
## $ age_diff                       <chr> NA, "Respondent is 4 years older",...
## $ number_total_children_known    <dbl> 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 1, 1...

## tibble [20,496 x 2] (S3: tbl_df/tbl/data.frame)
##  $ income_family           : Factor w/ 6 levels "$100,000 to $ 124,999",..: 3 5 5 1 4 4 6 6 6 6 ...
##  $ self_rated_mental_health: Factor w/ 6 levels "Don't know","Excellent",..: 2 4 4 6 4 6 5 6 6 5 ...
##  - attr(*, "na.action")= 'omit' Named int [1:106] 461 791 1026 1443 1530 1634 1914 1921 2200 2485 ..
```
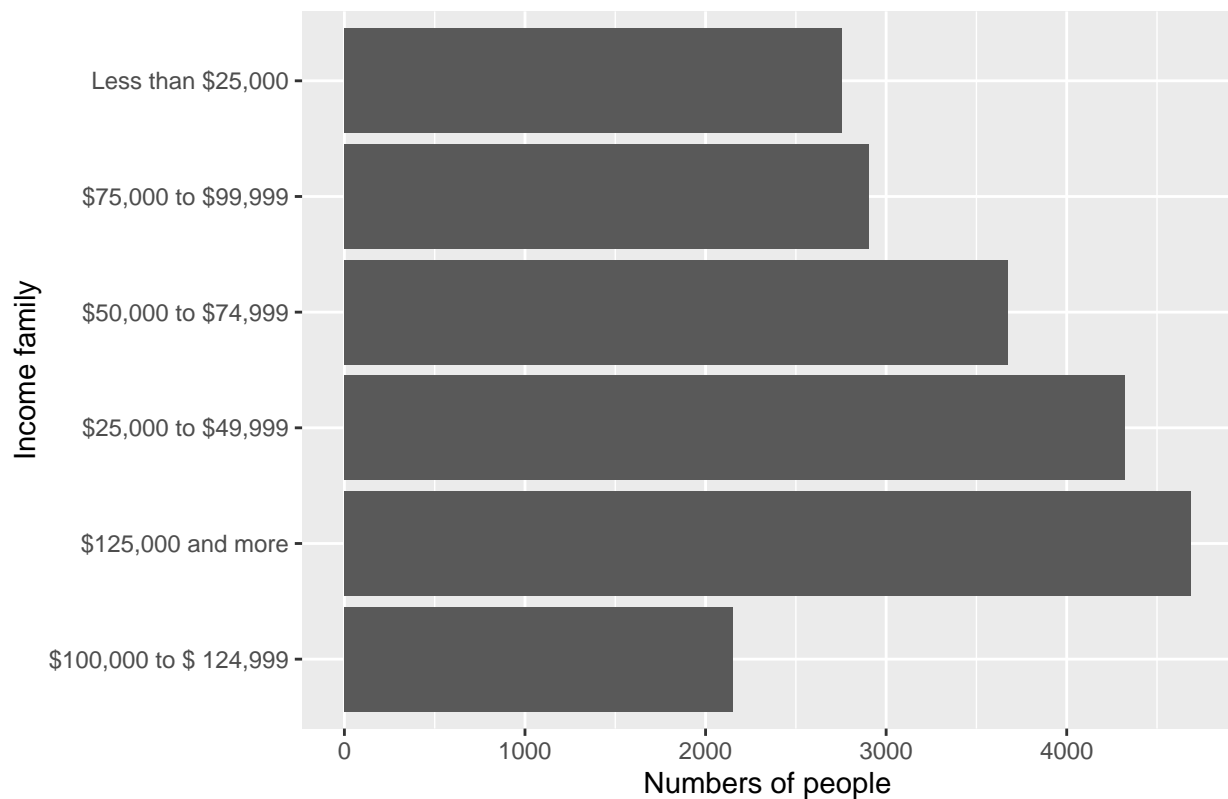
3

```
##   ..- attr(*, "names")= chr [1:106] "461" "791" "1026" "1443" ...
```
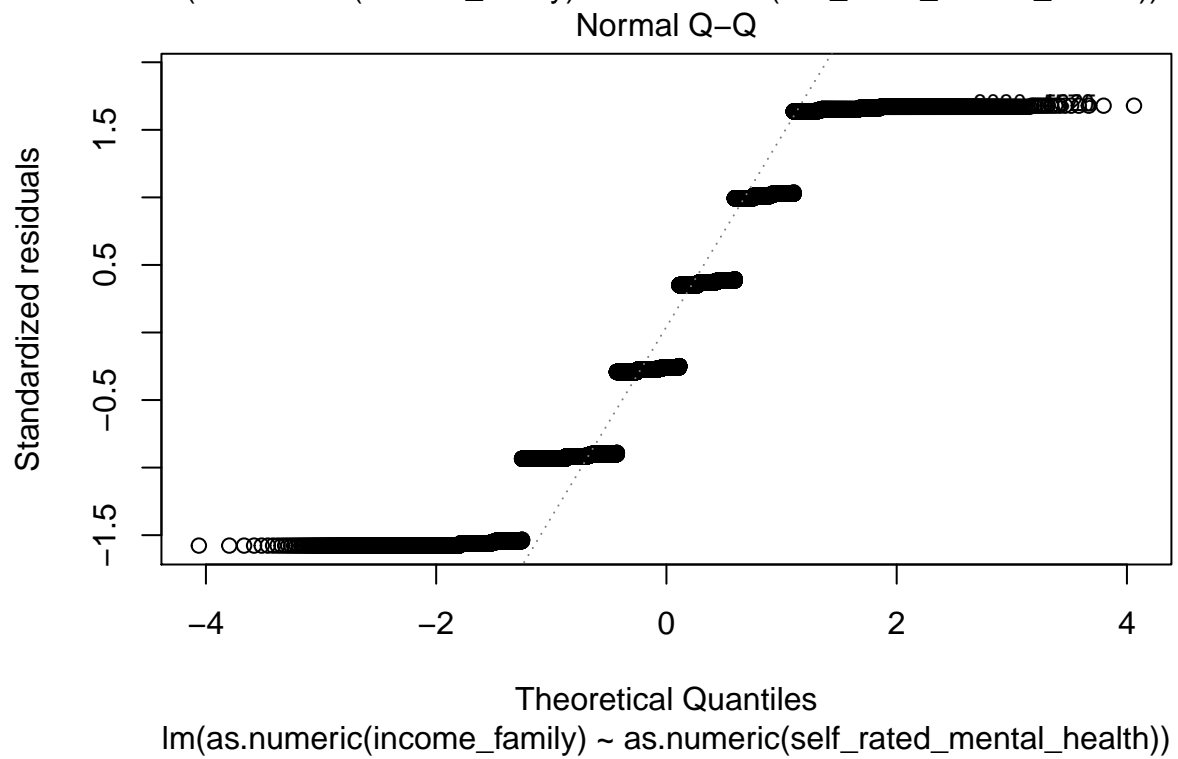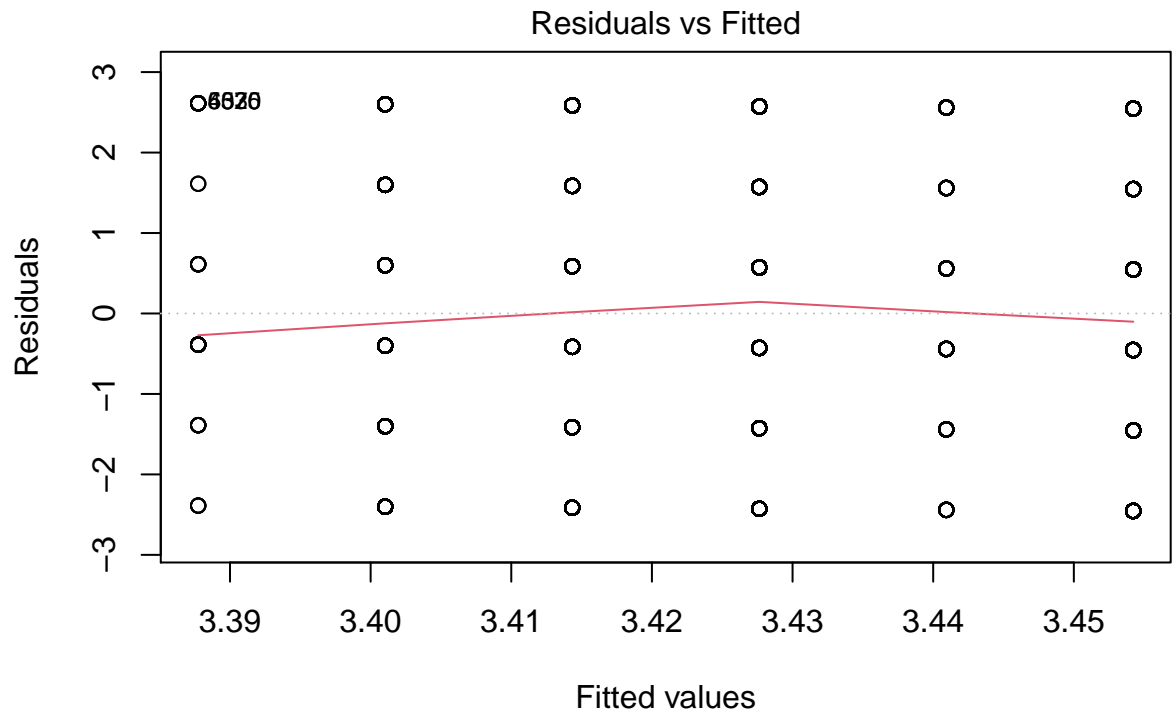
Plot on numbers of people who rate their who rate their mental
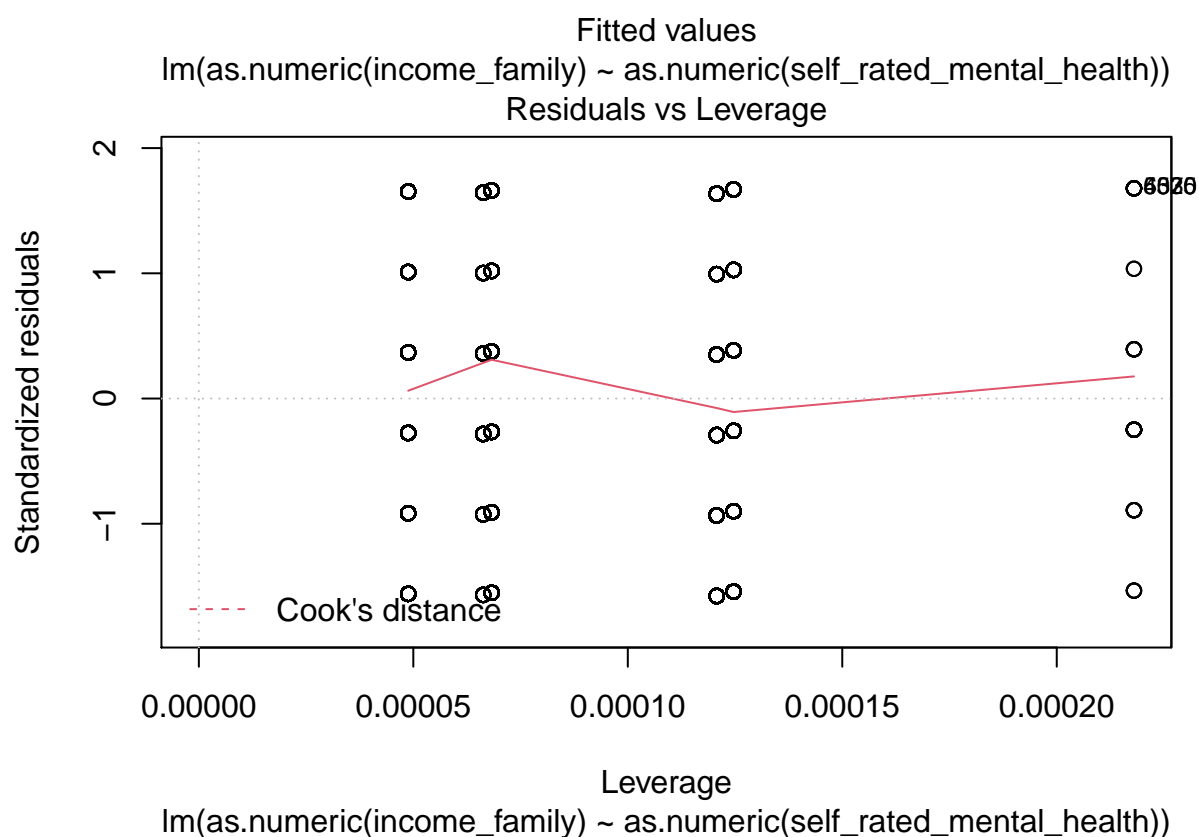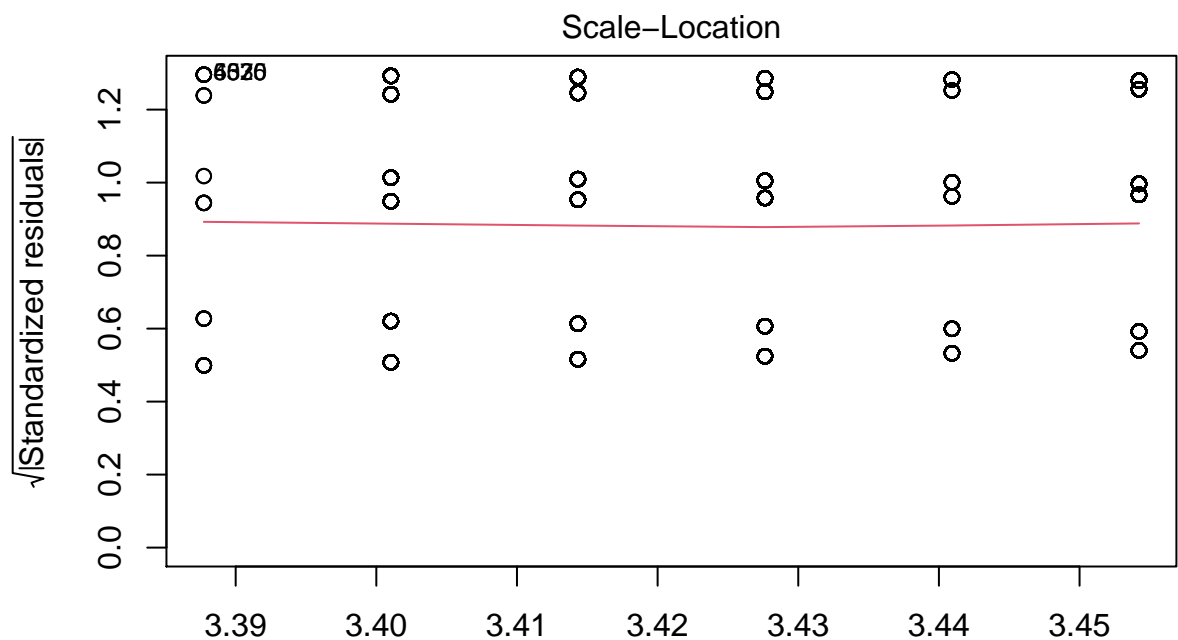health as different level
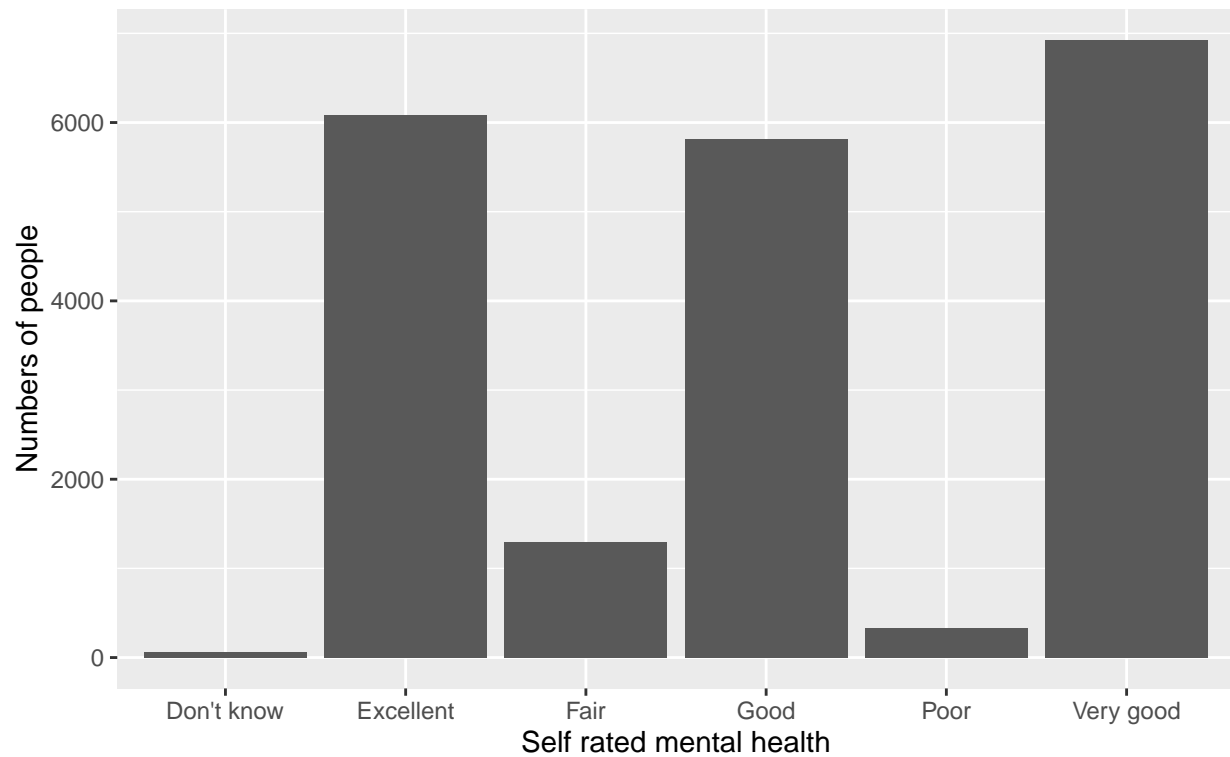
## Plot on numbers of people who has certain amount of family in


```
## 
## Call:
## lm(formula = as.numeric(income_family) ~ as.numeric(self_rated_mental_health),
##     data = income_mental)
## 
## Residuals:
##    Min     1Q Median     3Q    Max
## -2.454 -1.414 -0.401  1.546  2.612
## 
## Coefficients:
##                                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)                          3.374440   0.029043 116.190   <2e-16 ***
## as.numeric(self_rated_mental_health) 0.013298   0.006688   1.988   0.0468 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 1.556 on 20494 degrees of freedom
## Multiple R-squared:  0.0001928,  Adjusted R-squared:  0.0001441
## F-statistic: 3.953 on 1 and 20494 DF,  p-value: 0.0468
```

Residuals vs Fitted

Fitted values
lm(as.numeric(income_family) ~ as.numeric(self_rated_mental_health))

Normal Q–Q

Theoretical Quantiles
lm(as.numeric(income_family) ~ as.numeric(self_rated_mental_health))

Scale–Location

lm(as.numeric(income_family) ~ as.numeric(self_rated_mental_health))

Residuals vs Leverage

lm(as.numeric(income_family) ~ as.numeric(self_rated_mental_health))

```r
attach(income_mental)
plot(income_family, self_rated_mental_health,
     main = "Plot of Self rated mental health against Income family",
     xlab = "Self rated mental health", ylab = "Income family")
```

Plot of Self rated mental health against Income family
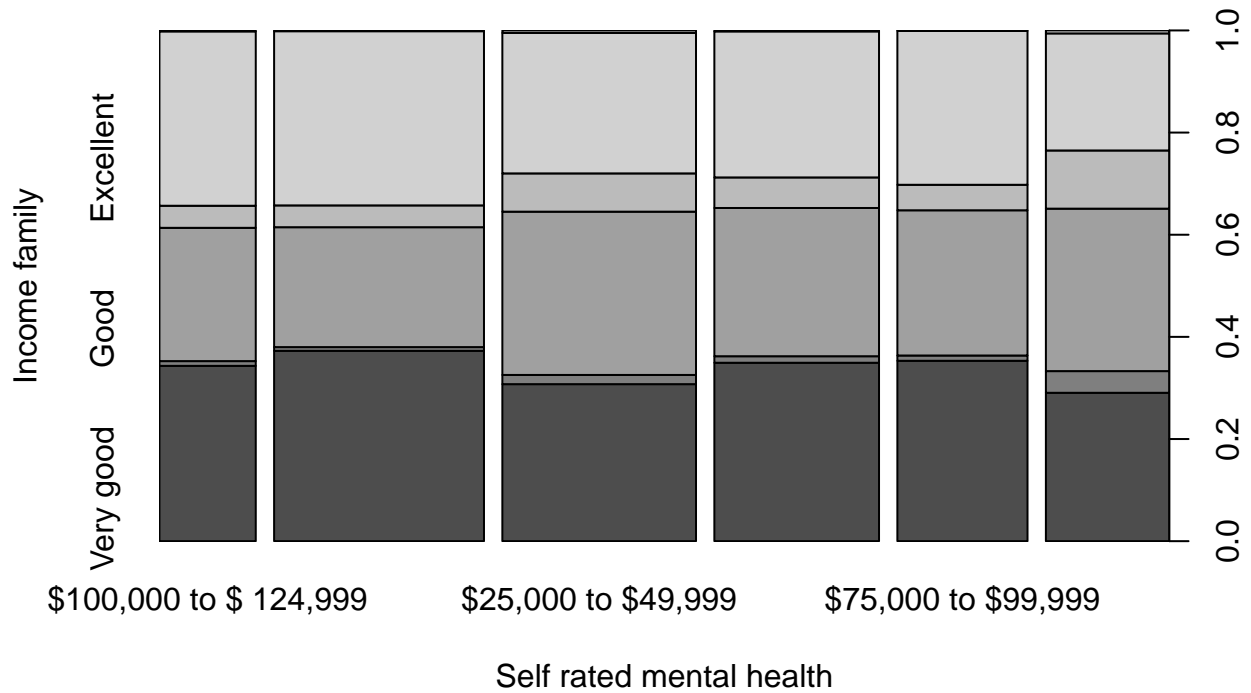

```
plot(as.numeric(income_family), as.numeric(self_rated_mental_health),
     main = "Plot of Self rated mental health against Income family",
     xlab = "Self rated mental health", ylab = "Income family")
abline(lm(income_family~self_rated_mental_health), col="red")
```

Plot of Self rated mental health against Income family
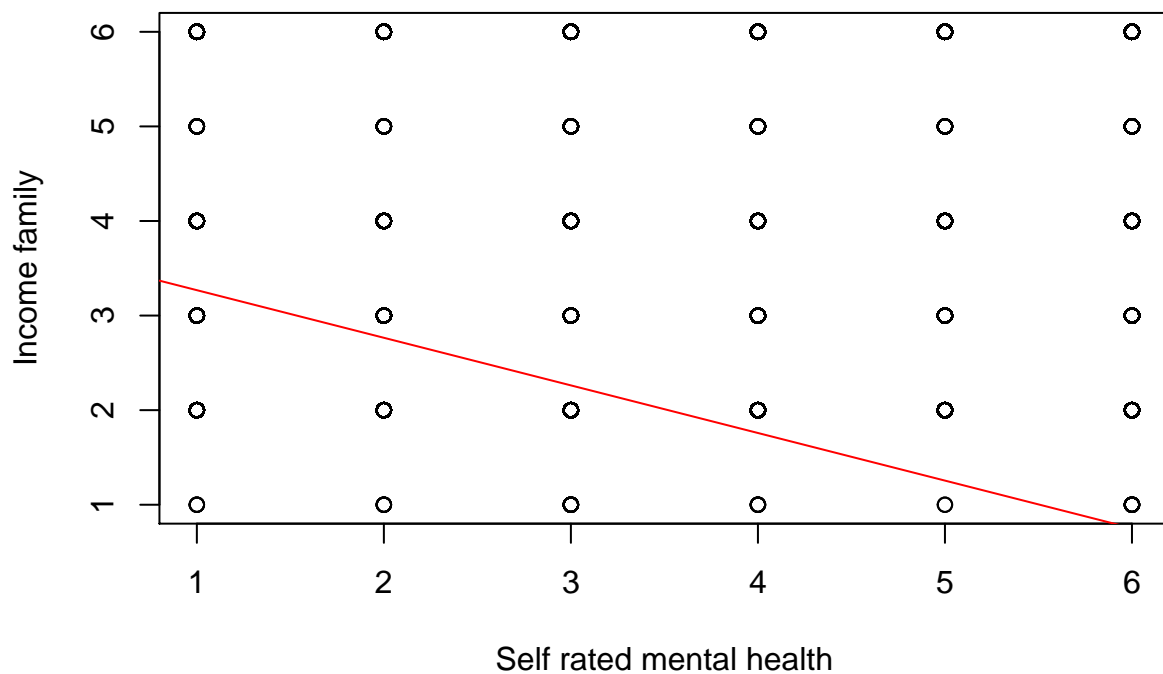
The model I choose was the simple linear regression model.

Introduce the selected model here. It is expected that you will use some mathematical notation here. If you do please ensure that all notation is explained. You may also want to discuss any special (hypothetical) cases of your model here, as well as any caveats.

Results

Here you will include all results. This includes descriptive statistics, graphs, figures, tables, and model results. Please ensure that everything is well formatted and in a report style. You must also provide an explanation of the results in this section. You can overflow to an Appendix if needed.

Please ensure that everything is well labelled. So if you have multiple histograms and plots, calling them Figure 1, 2, 3, etc. and referencing them as Figure 1, Figure 2, etc. in your report will be expected. The reader should not get lost in a sea of information. Make sure to have the results be clean, well formatted and digestible.

Discussion

Methodology:

Here you will discuss conclusions drawn from the results and comment on how it relates to the original goal of the study (which was specified in the Introduction).

Weaknesses

Here we discuss weaknesses of the study, data, analysis, etc. You can also discuss areas for improvement.

Next Steps

Here you discuss subsequent work to be done after this report. This can include next steps in terms of statistical analysis (perhaps there is a more efficient algorithm available, or perhaps there is a caveat in the data that would allow for some new technique). Future steps should also be specified in terms of the study setting (eg. including a follow-up survey on something, or a subsequent study that would complement the conclusions of your report).

References

1. Bruin, J. 2006. newtest: command to compute new test. UCLA: Statistical Consulting Group. https://stats.idre.ucla.edu/stata/ado/analysis/.