

WVS: Happiness in Bolivia

Olga Lavrinenko

2023-03-11

```
library(car)
library(dplyr)
library(ggplot2)
library(gplots)
library(ggpubr)
library(lm.beta)
library(QuantPsyc)
library(readr)
library(sjPlot)
library(tidyverse)
library(haven)
```

What has been fixed (11/03/2023):

- Formulation of the research question: “are more significant” - “have stronger effect”.
- Saturation of categories (some categories that are close in meaning were combined, some predictors (family importance, religion status) were removed due to the fact that categories even after coding had very few observations)
- The argument `geom_smooth(method = “lm”)` was removed and a non-linear relationship was found in two cases (age and influence on life).
- The interaction effect model was changed, because after recoding the previous predictor no longer gave a significant interaction effect.

Introduction

LCSR Laboratory Head and professor of sociology Panarin Eduard Dmitrievich (HSE - Saint-Petersburg) in one of his lectures pointed out that the **relatively poor countries of Latin America are one of the happiest regions in the world** (Eduard Ponarin: “In terms of happiness, we have passed the beginning of the 80s”, 2020. URL: <https://yeltsin.ru/news/eduard-ponarin-po-pokazatelyam-schastya-my-pereshli-nachalo-80-h/>). The reason for this is non-material factors of happiness such as close social ties, strong large families, high religiosity, and so on.

This study examines the factors that affect the level of happiness in **Bolivia**. Bolivia was taken for the study as it is one of the poorest countries in the region (South America) in accordance with the IMF assessment (Top 12 Richest Countries in South America (by GDP per capita 2020 - World Bank, PPP INT \$. URL: <https://worldpopulationreview.com/country-rankings/richest-countries-in-south-america>). It is also not in the top of the richest countries in Latin America in general (Report for Selected Countries and Subjects, 2019). This provides an opportunity to study the role of non-material factors of happiness in a country that is not the richest.

Research Question is whether non-material factors have stronger effect on happiness in Bolivia than material ones?

Data description

The Bolivia WVS database dates back to 2017 and includes 2067 observations across 394 variables.

```
setwd("D:/Data Analytics For Politics And Society/Data analysis (2-3 modules)/Project 1")
WVS_Bolivia <- read_stata('WVS_Wave_7_Bolivia_Stata_v5.0.dta')
dim(WVS_Bolivia)
```

```
## [1] 2067 394
```

Hypotheses

- 1) The **income** factor is not statistically significant in determining the level of happiness in Bolivia.
- 2) The factor of **financial situation** is also not statistically significant in determining the level of happiness in Bolivia.
- 3) Respondents who indicate the **importance of family** in their lives are happier than those for whom family is not of particular importance.
- 4) Respondents who do not **trust family** are less happy than those who trust.
- 5) Respondents who indicate the **importance of friends** in their lives are happier than those for whom friendship is not of particular importance.
- 6) Respondents who indicate the **importance of religion** in their lives are happier than those for whom religion is not of particular importance.
- 7) Respondents whose **marital status** is “divorced” are less happy than those who are married.

Analysis with the technical interpretation of the results

I. Preparation:

1. Creating a numeric variable “happiness” by combining variables “happiness” (How happy are you?) and “satisfaction” (How satisfied with your life are you?) into one variable.

```
# replacing negative values with missing ones
table(WVS_Bolivia$Q46)
```

```
##
## -2 -1 1 2 3 4
## 1 5 770 739 517 35
```

```
WVS_Bolivia$happiness_1 <- ifelse(WVS_Bolivia$Q46<0, NA, WVS_Bolivia$Q46)
table(WVS_Bolivia$happiness_1, WVS_Bolivia$Q46)
```

```
##
##      -2  -1   1   2   3   4
##    1   0   0 770   0   0   0
##    2   0   0   0 739   0   0
##    3   0   0   0   0 517   0
##    4   0   0   0   0   0 35
```

```
# replacing negative values with missing ones
table(WVS_Bolivia$Q49)
```

```
##
##  -1   1   2   3   4   5   6   7   8   9  10
##   9  23  18  37  74 256 208 281 462 266 433
```

```
WVS_Bolivia$lifesat_1 <- ifelse(WVS_Bolivia$Q49<0, NA, WVS_Bolivia$Q49)
table(WVS_Bolivia$lifesat_1, WVS_Bolivia$Q49)
```

```
##
##      -1   1   2   3   4   5   6   7   8   9  10
##    1   0  23   0   0   0   0   0   0   0   0
##    2   0   0  18   0   0   0   0   0   0   0
##    3   0   0   0  37   0   0   0   0   0   0
##    4   0   0   0   0  74   0   0   0   0   0
##    5   0   0   0   0   0 256   0   0   0   0
##    6   0   0   0   0   0   0 208   0   0   0
##    7   0   0   0   0   0   0   0 281   0   0
##    8   0   0   0   0   0   0   0   0 462   0
##    9   0   0   0   0   0   0   0   0   0 266   0
##   10   0   0   0   0   0   0   0   0   0   0 433
```

```
# changing the correspondence of numbers to categories of happiness:
# 1 is "Not at all happy", 4 - "Very happy"
WVS_Bolivia$happiness_2 <- (WVS_Bolivia$happiness_1*(-1)+5)
table(WVS_Bolivia$happiness_2)
```

```
##
##    1   2   3   4
##  35 517 739 770
```

```
# to combine variables let's change the range of values from 1 to 4, instead of from 1 to 10
WVS_Bolivia$lifesat_2 <- WVS_Bolivia$lifesat_1/2.5
table(WVS_Bolivia$lifesat_2)
```

```
##
## 0.4 0.8 1.2 1.6   2 2.4 2.8 3.2 3.6   4
## 23  18  37  74 256 208 281 462 266 433
```

```
# Creating a numeric variable "happiness"
WVS_Bolivia$happiness <- ((WVS_Bolivia$lifesat_2 + WVS_Bolivia$happiness_2)/2)
table(WVS_Bolivia$happiness)
```

```
##
## 0.7 0.9 1.1 1.2 1.3 1.4 1.5 1.6 1.7 1.8 1.9 2 2.1 2.2 2.3 2.4 2.5 2.6 2.7 2.8
## 3 2 5 15 7 8 7 21 5 39 9 122 10 86 17 56 77 72 72 50
## 2.9 3 3.1 3.2 3.3 3.4 3.5 3.6 3.8 4
## 128 114 201 50 97 95 133 190 128 234
```

2. Predictors

Since we will not work with all variables in the base, we will select only the necessary ones, reducing the base. We still have 2067 observations, but only 16 variables: 15 predictors and an outcome - happiness.

```
WVS_Bolivia <- WVS_Bolivia[, c("happiness", "Q260", "Q262",
                              "Q288", "Q50", "Q2", "Q58", "Q273",
                              "Q274", "Q6", "Q94", "Q169", "Q47", "Q48")]
base::dim(WVS_Bolivia)
```

```
## [1] 2067 14
```

Let's rename the predictors for more convenience.

```
# demographic factors
WVS_Bolivia$sex <- WVS_Bolivia$Q260
WVS_Bolivia$age <- WVS_Bolivia$Q262

# material factors
WVS_Bolivia$income <- WVS_Bolivia$Q288
WVS_Bolivia$financial_status <- WVS_Bolivia$Q50

# non-material factors: family and friends
WVS_Bolivia$friends_imp <- WVS_Bolivia$Q2
WVS_Bolivia$family_trust <- WVS_Bolivia$Q58
WVS_Bolivia$family_status <- WVS_Bolivia$Q273
WVS_Bolivia$children <- WVS_Bolivia$Q274

# non-material factors: religion
WVS_Bolivia$religion_imp <- WVS_Bolivia$Q6
WVS_Bolivia$religion_participation <- WVS_Bolivia$Q94
WVS_Bolivia$religion_VS_science <- WVS_Bolivia$Q169

# additional factors
WVS_Bolivia$health_status <- WVS_Bolivia$Q47
WVS_Bolivia$influence_on_life <- WVS_Bolivia$Q48
```

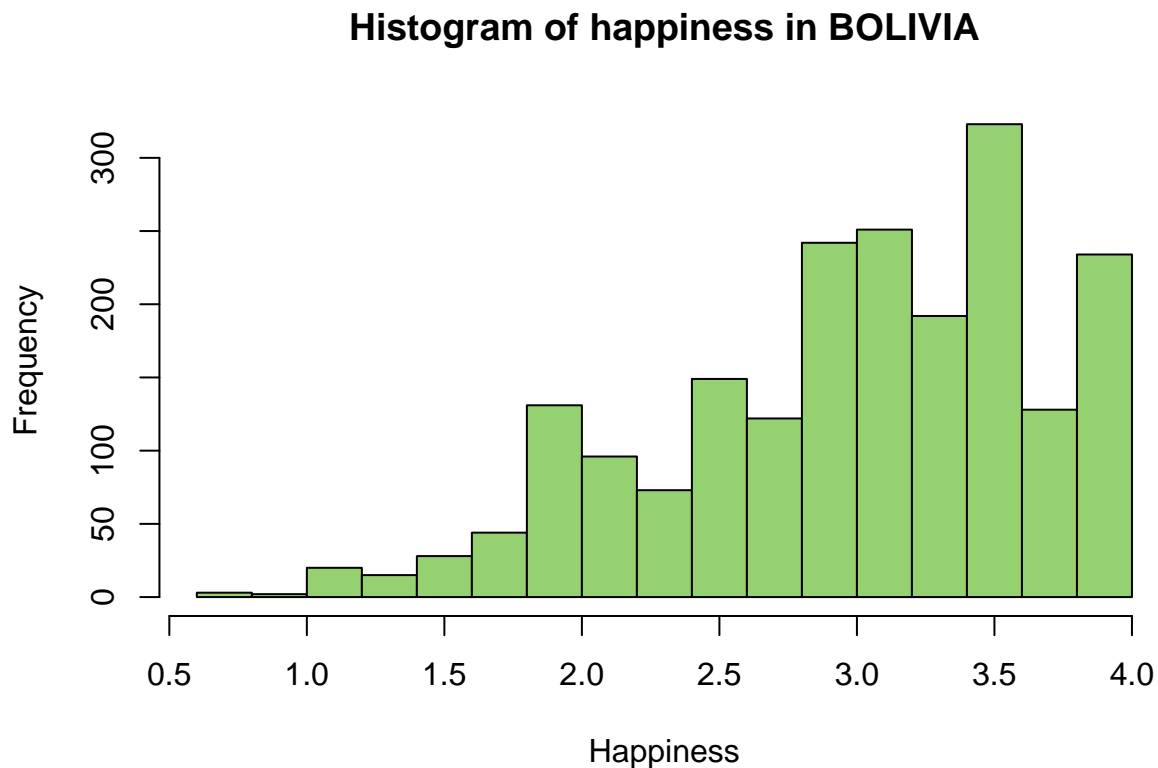
3. Discovering the properties of the data: distributions, character and number of missings, types of the variables.

1. **Outcome - happiness.** As we can see, data is piled up on the right, so it's left-skewed data. Minimal value is 0.70, while maximum is 4.00, and mean is 3.04. In addition, this variable includes 14 missing values. And the type of this variable is numeric.

```
summary(WVS_Bolivia$happiness)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.     NA's  
##      0.70   2.60   3.10   3.04   3.60   4.00       14
```

```
hist(WVS_Bolivia$happiness,  
     main = "Histogram of happiness in BOLIVIA",  
     xlab = "Happiness", col = "#96D171")
```



```
class(WVS_Bolivia$happiness)
```

```
## [1] "numeric"
```

2. Predictors

2.1 Gender: this variable is categorical, it includes two categories - male and female. The variable does not contain missing values. In addition, the data between categories is distributed fairly evenly: 1024 for male and 1043 for female.

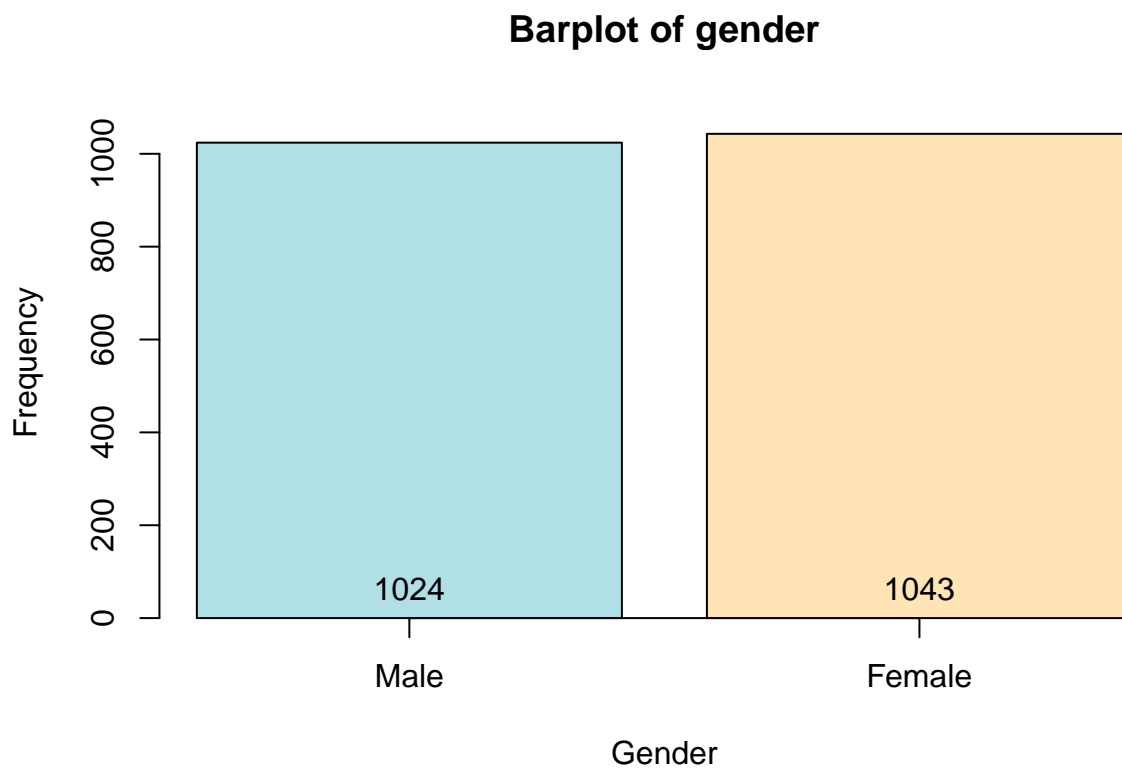
```
WVS_Bolivia$sex <- as.factor(WVS_Bolivia$sex)  
class(WVS_Bolivia$sex)
```

```
## [1] "factor"
```

```
table(WVS_Bolivia$sex)
```

```
##  
##      1      2  
## 1024 1043
```

```
bp_1 <- barplot(table(WVS_Bolivia$sex),  
  beside = T,  
  xlab = "Gender",  
  ylab = "Frequency",  
  ylim = c(0, 1100),  
  xaxt = "n",  
  main = "Barplot of gender",  
  col = c("#B0E0E6", "#FFE4B5"))  
  
axis(side = 1, at = bp_1[c(1,2)], labels = c("Male", "Female"))  
  
text(bp_1, -1.4, table(WVS_Bolivia$sex), col = "black", pos = 3)
```



2.2 Age: data is piled up on the left, so it's right-skewed data. Minimal age is 18, while maximum is 85, and mean is 38.33. In addition, this variable does not contain missing values. And the type of this variable is numeric.

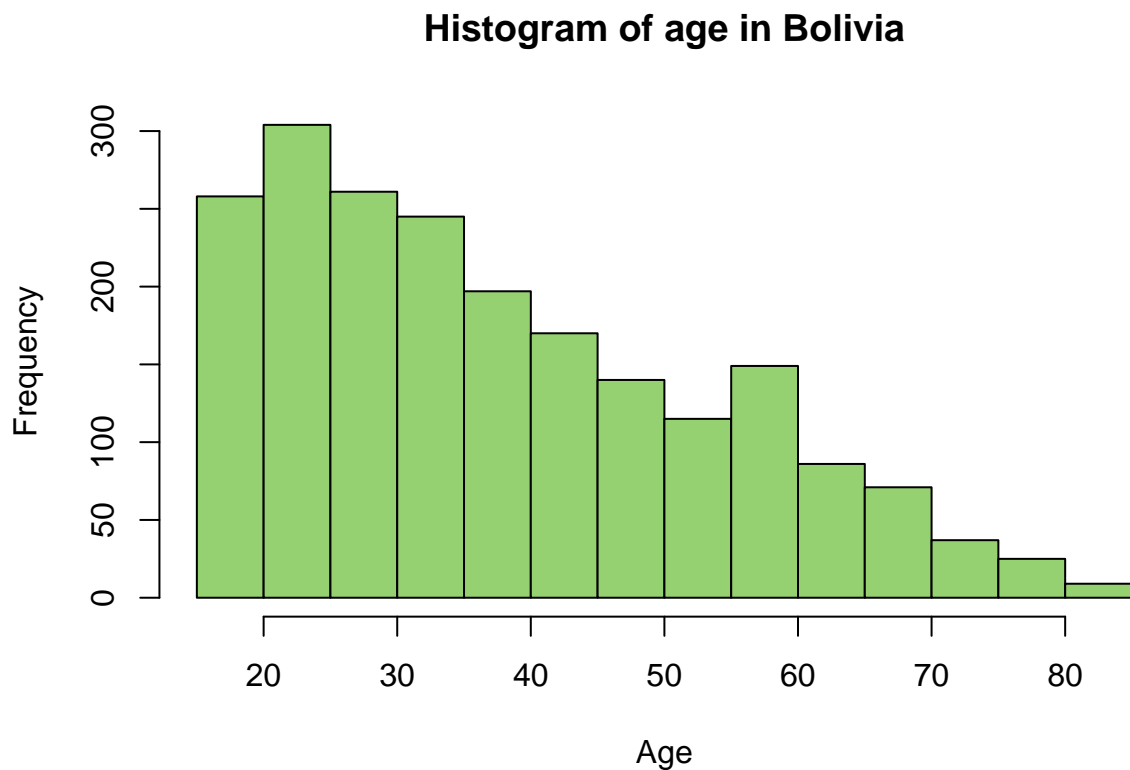
```
WVS_Bolivia$age <- as.numeric(ifelse(WVS_Bolivia$age<0, NA, WVS_Bolivia$age))
class(WVS_Bolivia$age)
```

```
## [1] "numeric"
```

```
summary(WVS_Bolivia$age)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    18.00   25.00   35.00   38.33   50.00   85.00
```

```
hist(WVS_Bolivia$age,
     main = "Histogram of age in Bolivia",
     xlab = "Age", col = "#96D171")
```



2.3 Income: data is almost symmetrical and close to a normal distribution. Minimal value is 1, while maximum is 10, and mean is 5.013. The most the most frequent value is 5 (540 respondents) and lest frequent is 9 (29 respondents). In addition, this variable includes 54 missing values. And the type of this variable is numeric.

```
WVS_Bolivia$income <- as.numeric(ifelse(WVS_Bolivia$income<0, NA, WVS_Bolivia$income))
class(WVS_Bolivia$income)
```

```
## [1] "numeric"
```

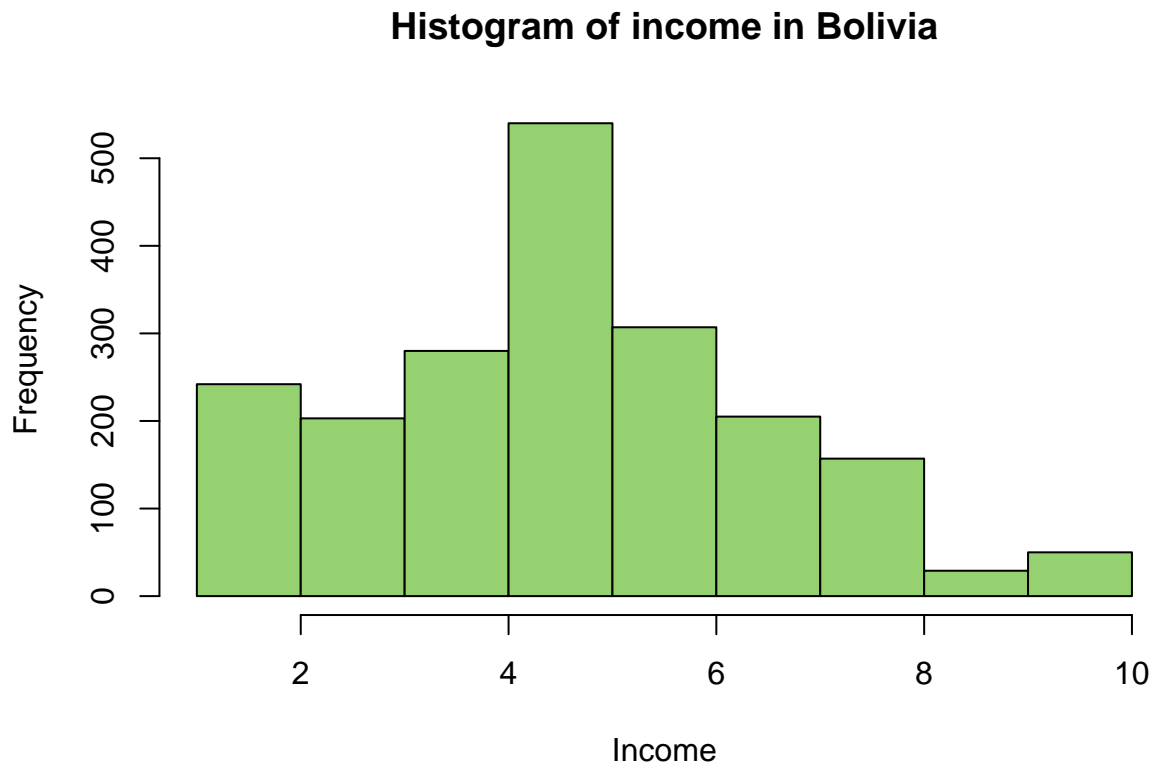
```
summary(WVS_Bolivia$income)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.     NA's  
##      1.000   4.000   5.000   5.013   6.000   10.000      54
```

```
table(WVS_Bolivia$income)
```

```
##  
##      1      2      3      4      5      6      7      8      9     10  
## 115 127 203 280 540 307 205 157  29  50
```

```
hist(WVS_Bolivia$income,  
     main = "Histogram of income in Bolivia",  
     xlab = "Income", col = "#96D171")
```



2.4 Financial status: data is almost symmetrical and close to a normal distribution. Minimal value is 1, while maximum is 10, and mean is 6.454. The most frequent values are 5 and 7 (366 and 367 respondents) and least frequent is 2 (40 respondents). In addition, this variable includes 9 missing values. And the type of this variable is numeric.

```
WVS_Bolivia$financial_status <- as.numeric(ifelse(WVS_Bolivia$financial_status<0,  
                                                  NA, WVS_Bolivia$financial_status))  
class(WVS_Bolivia$financial_status)
```



```
## [1] "numeric"
```

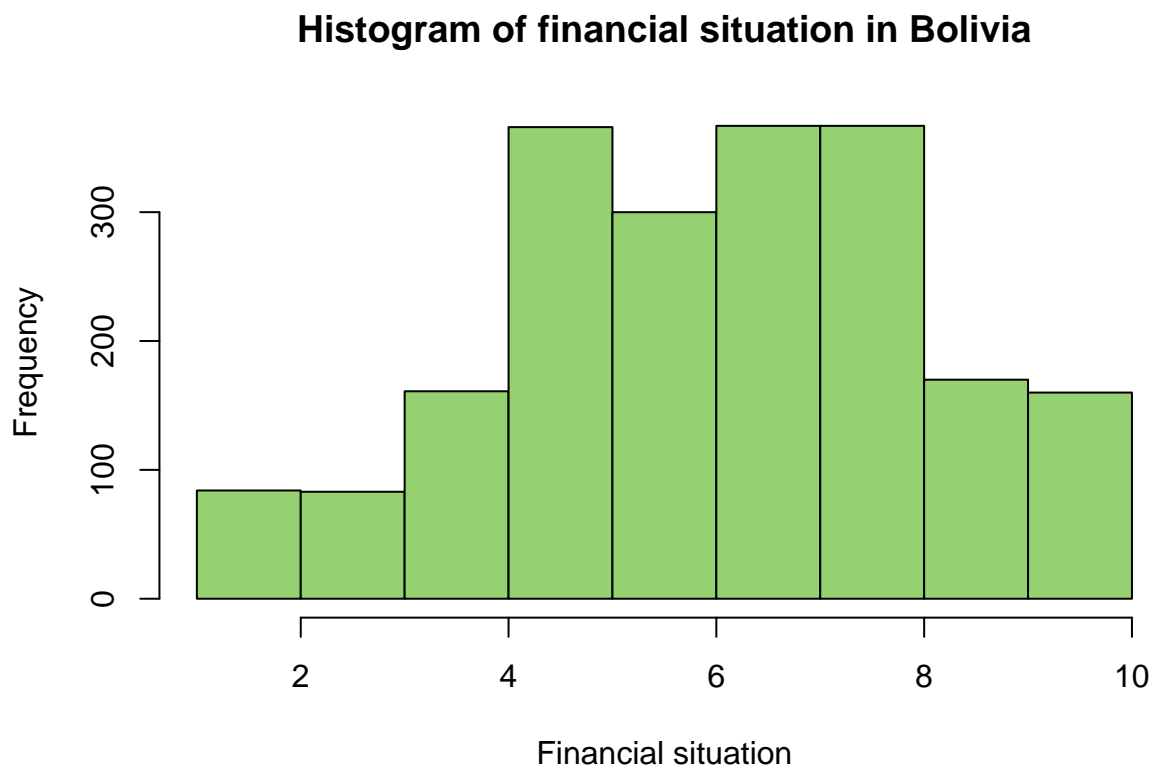
```
summary(WVS_Bolivia$financial_status)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's  
##    1.000   5.000   7.000   6.454   8.000  10.000     9
```

```
table(WVS_Bolivia$financial_status)
```

```
##  
##      1      2      3      4      5      6      7      8      9     10  
##    44     40     83    161    366    300    367    367    170    160
```

```
hist(WVS_Bolivia$financial_status,  
     main = "Histogram of financial situation in Bolivia",  
     xlab = "Financial situation", col = "#96D171",  
     breaks = 10)
```



2.5 Friends importance: this variable is categorical (factor), it includes four categories - 1 - "Very important", 2 - "Rather important", 3 - "Not very important", 4 - "Not at all important". The variable contains 7 missing values. The answers between categories are distributed unevenly:

- 432 for "Very important",
- 575 for "Rather important",
- 918 for "Not very important"
- 135 for "Not at all important"

```
WVS_Bolivia$friends_imp <- as.factor(ifelse(WVS_Bolivia$friends_imp<0,
                                             NA, WVS_Bolivia$friends_imp))
class(WVS_Bolivia$friends_imp)
```

```
## [1] "factor"
```

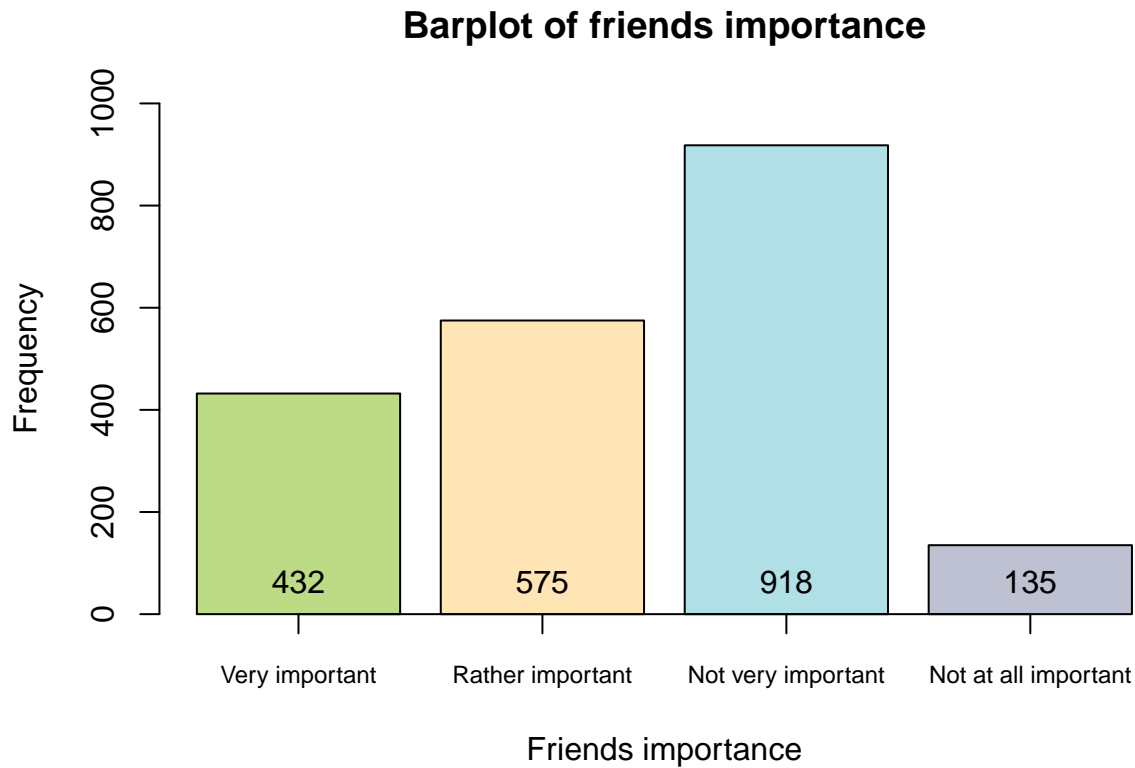
```
summary(WVS_Bolivia$friends_imp)
```

```
##      1      2      3      4 NA's
## 432  575  918  135      7
```

```
bp_2 <- barplot(table(WVS_Bolivia$friends_imp),
                 beside = T,
                 xlab = "Friends importance",
                 ylab = "Frequency",
                 ylim = c(0, 1000),
                 xaxt = "n",
                 main = "Barplot of friends importance",
                 col = c("#BDDA85", "#FFE4B5", "#B0E0E6", "#BEC1D1"))

text(bp_2, 3, table(WVS_Bolivia$friends_imp), col = "black", pos = 3)

axis(side = 1, at = bp_2[c(1:4)], na.rm = T,
     labels = c("Very important", "Rather important",
                "Not very important", "Not at all important"),
     cex.axis = 0.75)
```



2.6 Family trust: this variable is categorical (factor), it includes four categories - 1 - “Trust completely”, 2 - “Trust somewhat”, 3 - “Do not trust very much”, 4 - “Do not trust at all”. The variable contains 1 missing value. The answers between categories are distributed unevenly:

- 1170 for "Trust completely",
- 518 for "Trust somewhat",
- 313 for "Do not trust very much"
- 65 for "Do not trust at all"

```
WVS_Bolivia$family_trust <- as.factor(ifelse(WVS_Bolivia$family_trust<0,
                                             NA, WVS_Bolivia$family_trust))
class(WVS_Bolivia$family_trust)
```

```
## [1] "factor"
```

```
summary(WVS_Bolivia$family_trust)
```

```
##      1      2      3      4 NA's
## 1170   518   313    65     1
```

So, let's recode this variable. After recoding we have 2 categories - “Trust” and “Do not trust”. The unevenness is still present, but the number of observations in the categories is already greater (378 and 1689).

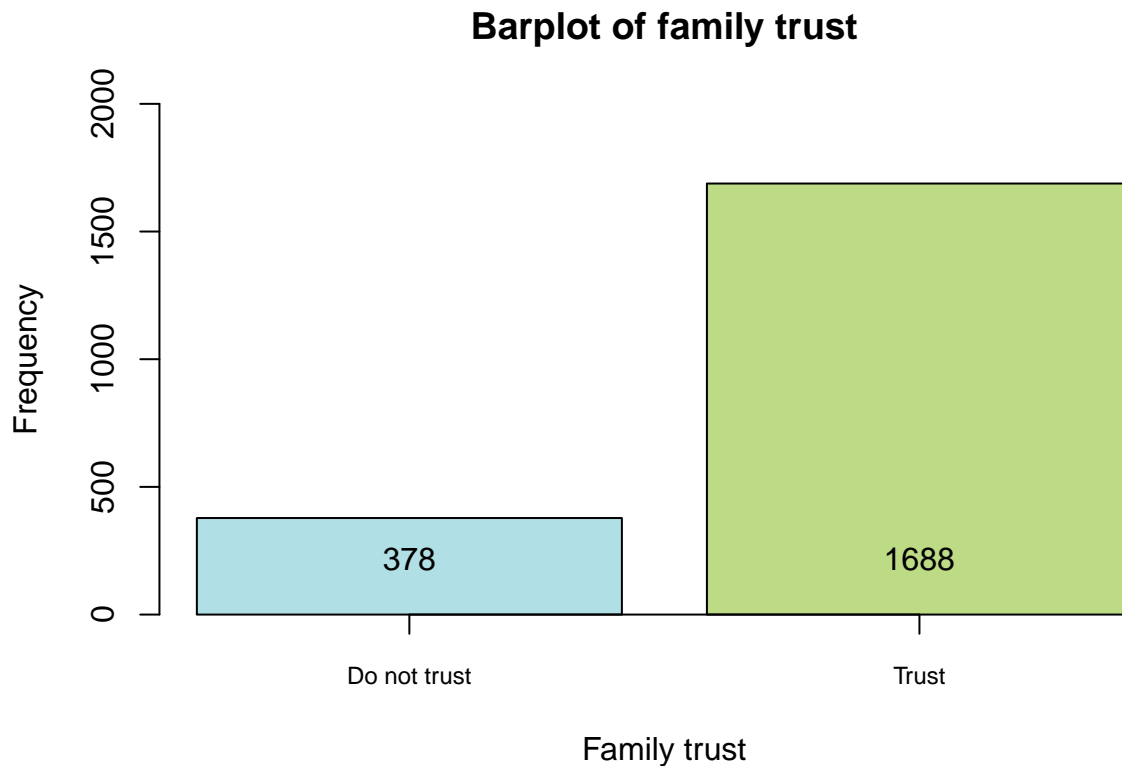
```
WVS_Bolivia$family_trust <- as.factor(ifelse(as.numeric(WVS_Bolivia$family_trust) <= 2, "Trust",
                                             ifelse(as.numeric(WVS_Bolivia$family_trust) >= 3, "Do not trust",
                                                    table(WVS_Bolivia$family_trust)))
```

```
##
## Do not trust      Trust
##           378      1688
```

```
bp_3 <- barplot(table(WVS_Bolivia$family_trust),
  beside = T,
  xlab = "Family trust",
  ylab = "Frequency",
  ylim = c(0, 2000),
  xaxt = "n",
  main = "Barplot of family trust",
  col = c("#BOE0E6", "#BDDA85"))

text(bp_3, 100, table(WVS_Bolivia$family_trust), col = "black", pos = 3)

axis(side = 1, at = bp_3[c(1:2)], na.rm = T,
  labels = c("Do not trust", "Trust"),
  cex.axis = 0.75)
```



2.7 Family status: this variable is categorical (factor), it includes six categories - 1 - “Married”, 2 - “Living together as married”, 3 - “Divorced”, 4 - “Separated”, 5 - “Widowed”, 6 - “Single”. The variable contains 4 missing values. The answers between categories are distributed unevenly:

- 758 for category "Married",
- 425 for "Living together as married",
- 51 for "Divorced"
- 96 for "Separated",
- 105 for "Widowed",
- 628 for "Single".

```
WVS_Bolivia$family_status <- as.factor(ifelse(WVS_Bolivia$family_status<0,
                                              NA, WVS_Bolivia$family_status))
class(WVS_Bolivia$family_status)
```

```
## [1] "factor"
```

```
summary(WVS_Bolivia$family_status)
```

```
##      1      2      3      4      5      6 NA's
## 758 425  51  96 105 628   4
```

Some categories are not saturated with observations, so we need to recode the variable.

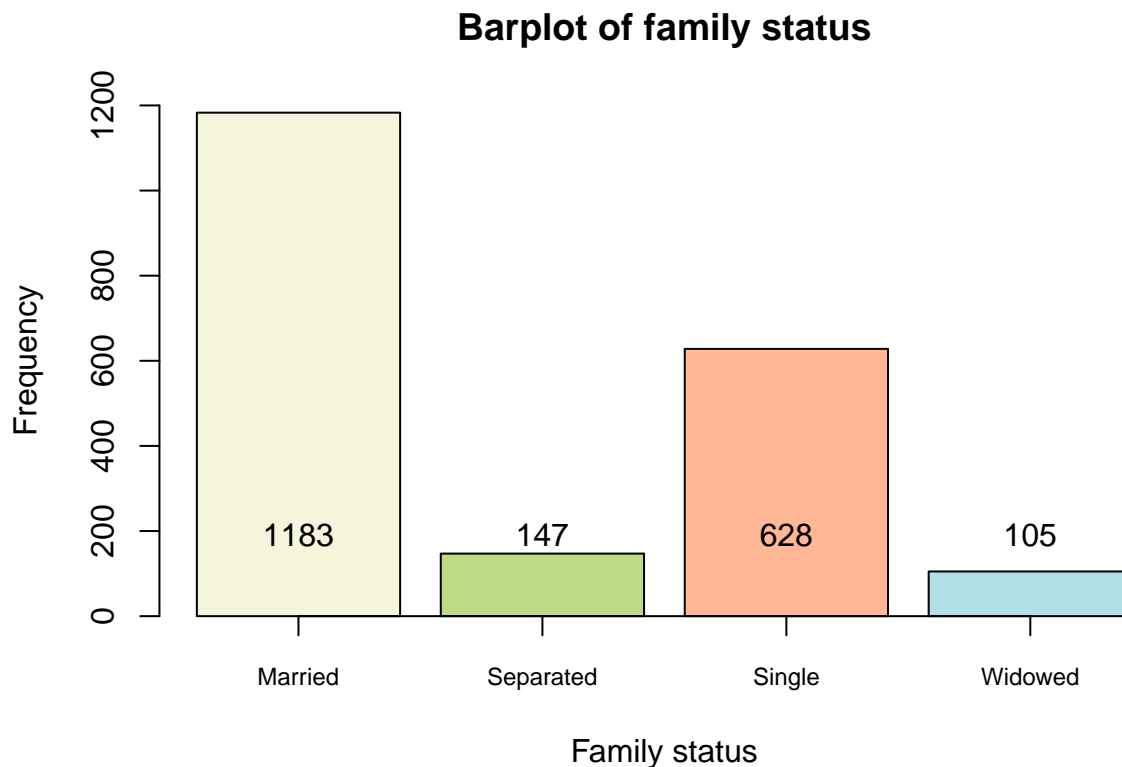
```
WVS_Bolivia$family_status <-
  as.factor(ifelse(as.numeric(WVS_Bolivia$family_status) <= 2, "Married",
                  ifelse(as.numeric(WVS_Bolivia$family_status) >= 3 &
                        as.numeric(WVS_Bolivia$family_status)<5, "Separated",
                        ifelse(as.numeric(WVS_Bolivia$family_status)==5, "Widowed", "Single"))))
table(WVS_Bolivia$family_status)
```

```
##
## Married Separated      Single  Widowed
##    1183      147      628      105
```

```
bp_4 <- barplot(table(WVS_Bolivia$family_status),
  beside = T,
  xlab = "Family status",
  ylab = "Frequency",
  ylim = c(0, 1200),
  xaxt = "n",
  main = "Barplot of family status",
  col = c("#F5F5DC", "#BDDA85", "#FFB796", "#B0E0E6" ))

text(bp_4, 120, table(WVS_Bolivia$family_status), col = "black", pos = 3)

axis(side = 1, at = bp_4[c(1:4)], na.rm = T,
  labels = c("Married", "Separated", "Single", "Widowed"),
  cex.axis = 0.75)
```



2.8 Children: data is piled up on the left, so it's right-skewed data. Minimal value is 0, while maximum is 15, and mean is 2.266. The most frequent value is 0 (559 respondents) and least frequent is 15 (1 respondent). In addition, this variable includes 1 missing value. And the type of this variable is numeric.

```
WVS_Bolivia$children <- as.numeric(ifelse(WVS_Bolivia$children<0,
                                           NA, WVS_Bolivia$children))
class(WVS_Bolivia$children)
```

```
## [1] "numeric"
```

```
summary(WVS_Bolivia$children)
```

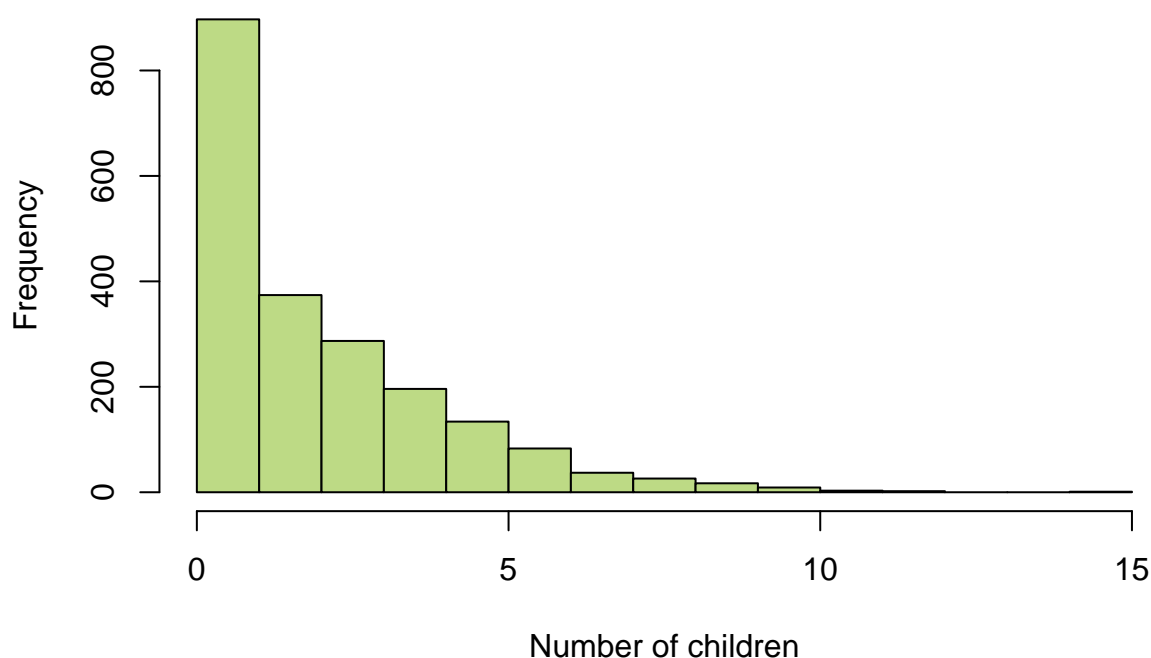
```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.     NA's
##  0.000   0.000    2.000   2.266   3.000   15.000         1
```

```
table(WVS_Bolivia$children)
```

```
##
##  0   1   2   3   4   5   6   7   8   9  10  11  12  15
## 559 338 374 287 196 134  83  37  26  17   9   3   2   1
```

```
hist(WVS_Bolivia$children,
     main = "Histogram of the number of children in the family in Bolivia",
     xlab = "Number of children", col = "#BDDA85")
```

Histogram of the number of children in the family in Bolivia



Let's combine cases where there are more than three children in a family into one category.

2.9 Children_1: this variable is categorical (factor), it includes five categories from 0 to 4. The variable doesn't contain missing values. The answers between categories are distributed almost evenly:

- 559 respondents don't have children
- 338 respondents have 1 child
- 374 respondent have 2 children
- 287 respondent have 3 children
- 508 respondent have more than 3 children

```
WVS_Bolivia$children_1 <- as.factor(ifelse(WVS_Bolivia$children > 3, 4, WVS_Bolivia$children))
table(WVS_Bolivia$children_1)
```

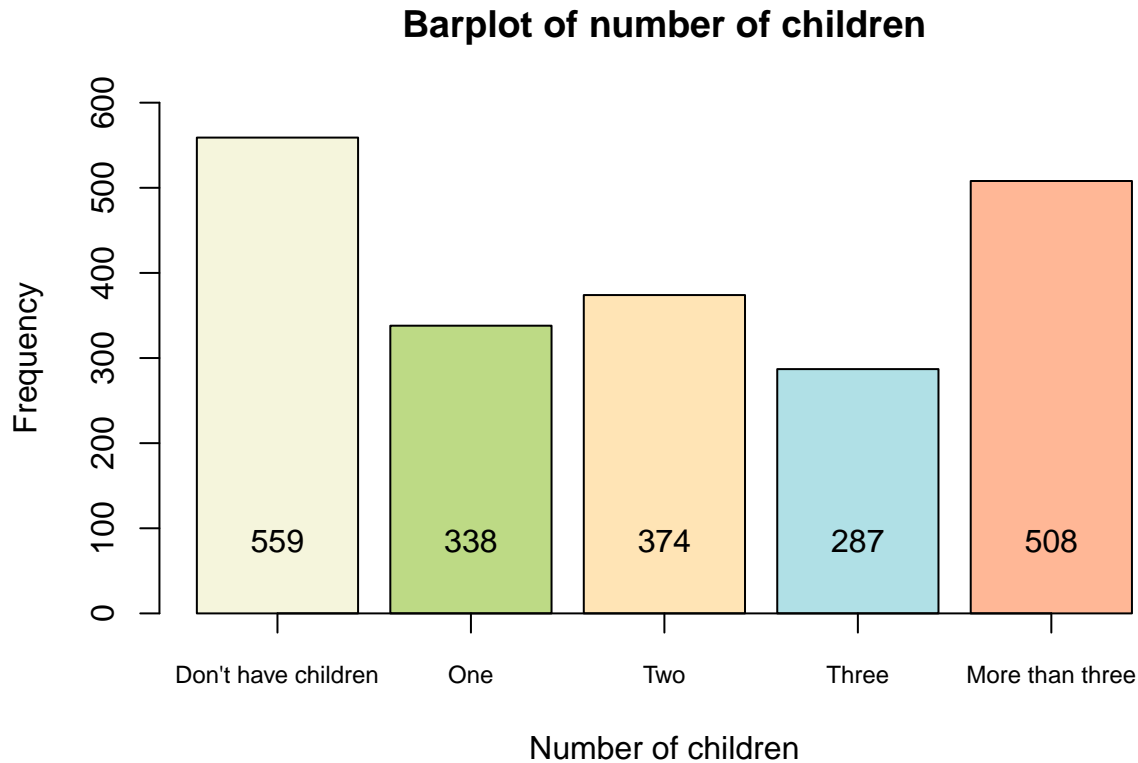
```
##
##  0  1  2  3  4
## 559 338 374 287 508
```

```
bp_6 <- barplot(table(WVS_Bolivia$children_1),
  beside = T,
  xlab = "Number of children",
  ylab = "Frequency",
  ylim = c(0, 600),
  xaxt = "n",
  main = "Barplot of number of children",
```

```
col = c("#F5F5DC", "#BDDA85", "#FFE4B5", "#B0E0E6", "#FFB796"))

text(bp_6, 50, table(WVS_Bolivia$children_1), col = "black", pos = 3)

axis(side = 1, at = bp_6[c(1:5)], na.rm = T,
     labels = c("Don't have children", "One",
                "Two", "Three", "More than three"),
     cex.axis = 0.75)
```



2.10 Religion importance: this variable is categorical (factor), it includes four categories - 1 - “Very important”, 2 - “Rather important”, 3 - “Not very important”, 4 - “Not at all important”. The variable contains 3 missing values. The answers between categories are distributed unevenly:

- 1208 for "Very important",
- 528 for "Rather important",
- 281 for "Not very important",
- 47 for "Not at all important"

```
WVS_Bolivia$religion_imp <- as.factor(ifelse(WVS_Bolivia$religion_imp<0,
                                             NA, WVS_Bolivia$religion_imp))
class(WVS_Bolivia$religion_imp)
```

```
## [1] "factor"
```



```
summary(WVS_Bolivia$religion_imp)
```

```
##      1      2      3      4 NA's
## 1208   528   281   47      3
```

Due to unevenness, we need to recode the variable. After recoding we have 2 categories - “Important” and “Do not trust”. The unevenness is still present, but the number of observations in the categories is already greater (328 and 1739).

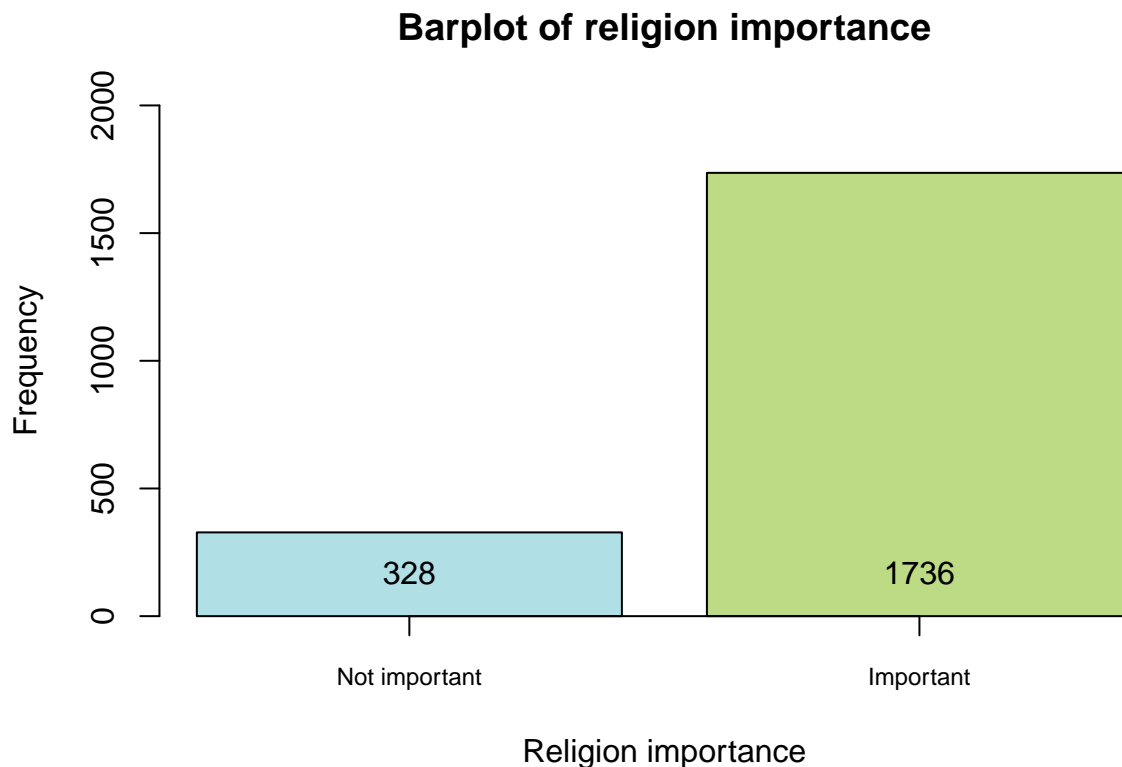
```
WVS_Bolivia$religion_imp <- as.factor(ifelse(as.numeric(WVS_Bolivia$religion_imp) <= 2, "Important",
                                             ifelse(as.numeric(WVS_Bolivia$religion_imp) >= 3, "Do not important",
                                                    table(WVS_Bolivia$religion_imp)))
```

```
##
## Do not important      Important
##              328              1736
```

```
bp_6 <- barplot(table(WVS_Bolivia$religion_imp),
  beside = T,
  xlab = "Religion importance",
  ylab = "Frequency",
  ylim = c(0, 2000),
  xaxt = "n",
  main = "Barplot of religion importance",
  col = c("#B0E0E6", "#BDDA85"))

text(bp_6, 50, table(WVS_Bolivia$religion_imp), col = "black", pos = 3)

axis(side = 1, at = bp_6[c(1:2)], na.rm = T,
  labels = c("Not important", "Important"),
  cex.axis = 0.75)
```



2.11 Religion participation (are you a member of church or religious organization?): this variable is categorical (factor), it includes three categories: 2 - “Active member”, 1 - “Inactive member”, 0 - “Don’t belong”. The variable contains 5 missing values. The answers between categories are distributed almost evenly:

- 640 for "Don't belong",
- 547 for "Inactive member",
- 875 for "Active member".

```
WVS_Bolivia$religion_participation <- as.factor(ifelse(WVS_Bolivia$religion_participation<0,
                                                         NA, WVS_Bolivia$religion_participation))
class(WVS_Bolivia$religion_participation)
```

```
## [1] "factor"
```

```
summary(WVS_Bolivia$religion_participation)
```

```
##    0    1    2 NA's
## 640 547 875    5
```

```
bp_7 <- barplot(table(WVS_Bolivia$religion_participation),
                  beside = T,
                  xlab = "Religion participation",
                  ylab = "Frequency",
```

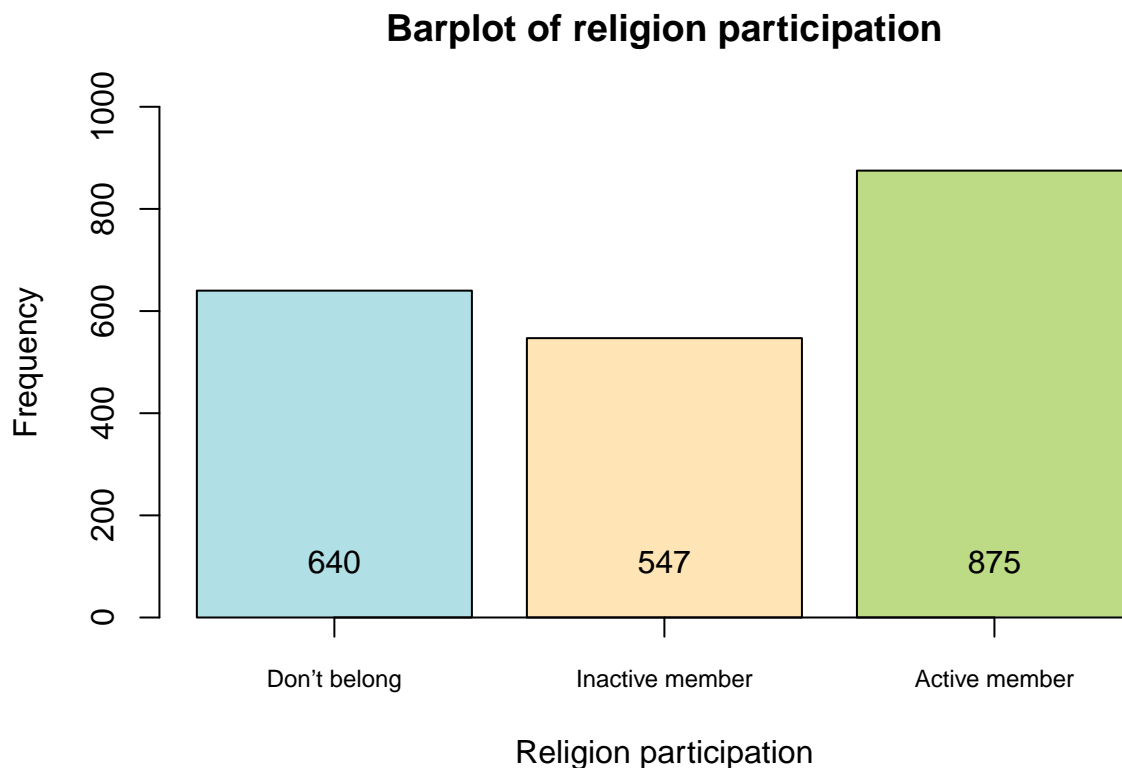
```

ylim = c(0, 1000),
xaxt = "n",
main = "Barplot of religion participation",
col = c( "#B0E0E6", "#FFE4B5", "#BDDA85"))

text(bp_7, 50, table(WVS_Bolivia$religion_participation), col = "black", pos = 3)

axis(side = 1, at = bp_7[c(1:3)], na.rm = T,
     labels = c("Don't belong", "Inactive member", "Active member"),
     cex.axis = 0.75)

```



2.12 Religion_VS_science (Please tell us if you strongly agree, agree, disagree, or strongly disagree with the following statement: “Whenever science and religion conflict, religion is always right”). This variable is categorical (factor), it includes four categories: 1 - “Strongly agree”, 2 - “Agree”, 3 - “Disagree”, 4 - “Strongly disagree”. The variable contains 84 missing values. The answers between categories are distributed unevenly:

- 201 for "Strongly agree"
- 1174 for "Agree"
- 546 for "Disagree"
- 62 for "Strongly disagree"

```

WVS_Bolivia$religion_VS_science <- as.factor(ifelse(WVS_Bolivia$religion_VS_science<0,
                                                    NA, WVS_Bolivia$religion_VS_science))
class(WVS_Bolivia$religion_VS_science)

```

```
## [1] "factor"
```

```
summary(WVS_Bolivia$religion_VS_science)
```

```
##      1      2      3      4 NA's
## 201 1174  546   62   84
```

We again need to recode the variable. After recoding we have 2 categories - “Agree” and “Disagree”. The unevenness is still present, but the number of observations in the categories is already greater (1459 and 608).

```
WVS_Bolivia$religion_VS_science <-
  as.factor(ifelse(as.numeric(WVS_Bolivia$religion_VS_science) <= 2, "Agree",
    ifelse(as.numeric(WVS_Bolivia$religion_VS_science) >= 3, "Disagree",
table(WVS_Bolivia$religion_VS_science)
```

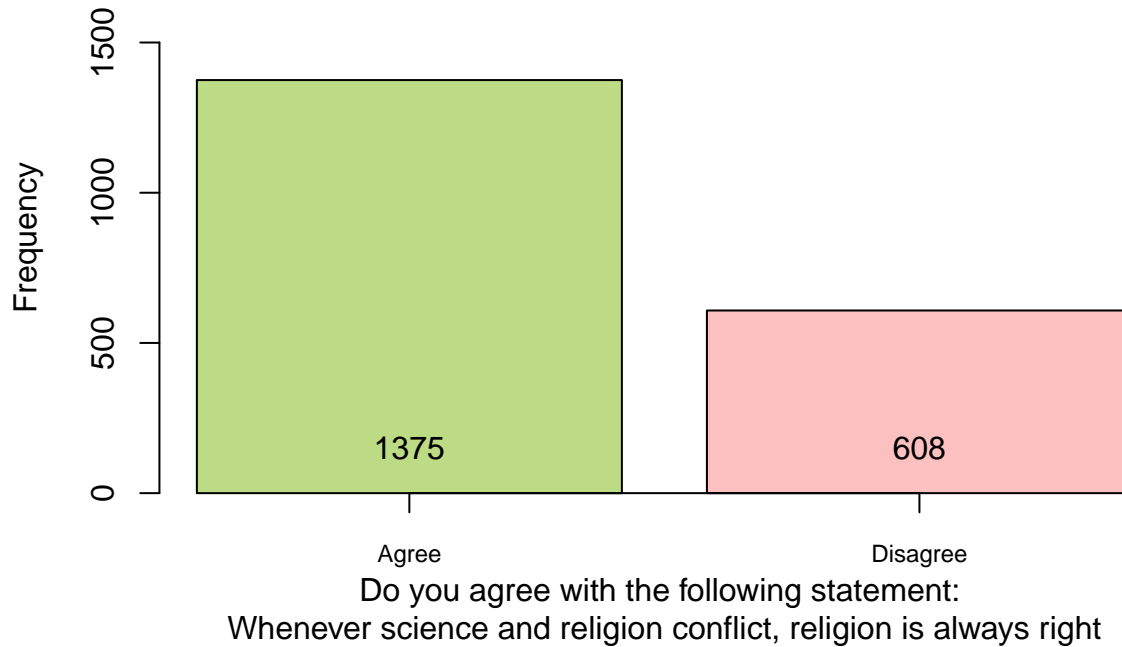
```
##
##      Agree Disagree
##    1375      608
```

```
bp_8 <- barplot(table(WVS_Bolivia$religion_VS_science),
  beside = T,
  xlab = "Do you agree with the following statement: \nWhenever science and religion conflict, re",
  ylab = "Frequency",
  ylim = c(0, 1700),
  xaxt = "n",
  main = "Barplot of the importance of religion over science",
  col = c("#BDDA85", "#FDCOCO"))

text(bp_8, 50, table(WVS_Bolivia$religion_VS_science), col = "black", pos = 3)

axis(side = 1, at = bp_8[c(1:2)], na.rm = T,
  labels = c("Agree", "Disagree"),
  cex.axis = 0.75)
```

Barplot of the importance of religion over science



2.13 Health status (state of health): this variable is categorical (factor), it includes five categories: 1 - “Very good”, 2 - “Good”, 3 - “Fair”, 4 - “Poor”, 5 - “Very poor”. The variable contains 1 missing value. The answers between categories are distributed unevenly:

- 164 for "Very good"
- 817 for "Good"
- 1007 for "Fair"
- 69 for "Poor"
- 9 for "Very poor"

```
WVS_Bolivia$health_status <- as.factor(ifelse(WVS_Bolivia$health_status<0,
                                              NA, WVS_Bolivia$health_status))
class(WVS_Bolivia$health_status)
```

```
## [1] "factor"
```

```
summary(WVS_Bolivia$health_status)
```

```
##      1      2      3      4      5 NA's
## 164  817 1007   69    9    1
```

Let's recode the variable. After recoding we have 2 categories - “Good” and “Not so good”. Now the data is distributed quite evenly: 982 observations for “Good” and 1086 for “Not so good”.

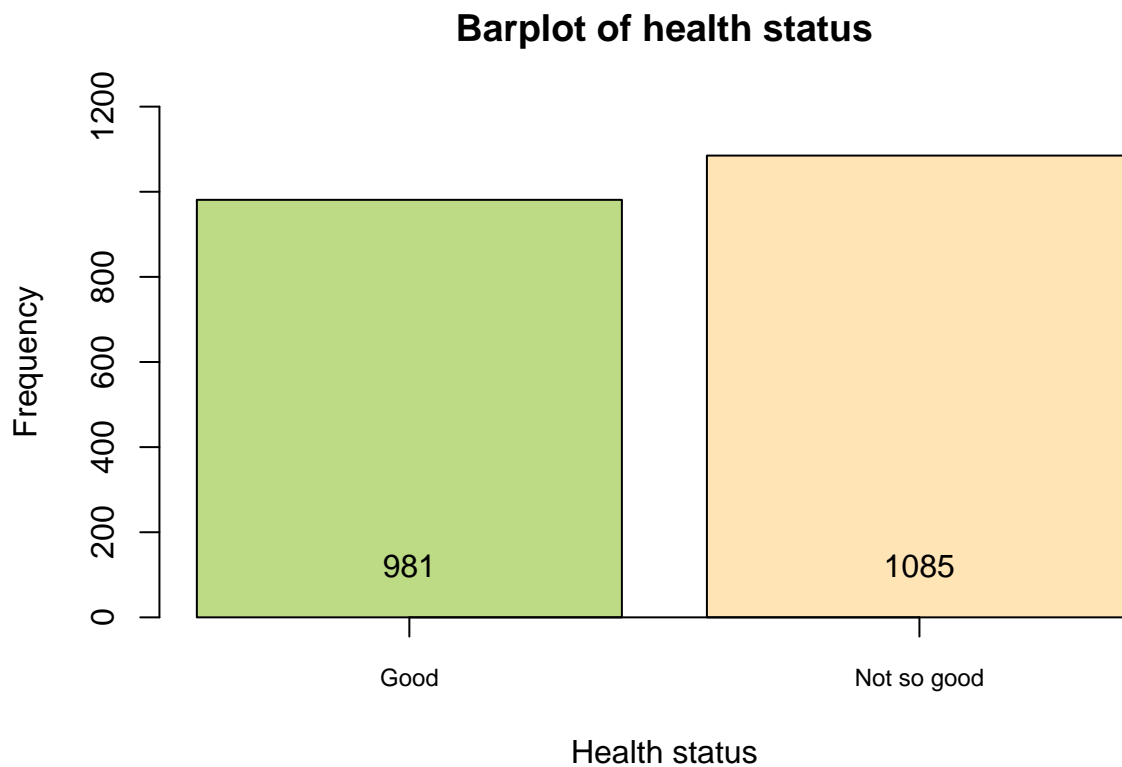
```
WVS_Bolivia$health_status <-
  as.factor(ifelse(as.numeric(WVS_Bolivia$health_status) <= 2, "Good",
                    ifelse(as.numeric(WVS_Bolivia$health_status) >= 3, "Not so good",
                           table(WVS_Bolivia$health_status))
```

```
##
##      Good Not so good
##      981      1085
```

```
bp_10 <- barplot(table(WVS_Bolivia$health_status),
  beside = T,
  xlab = "Health status",
  ylab = "Frequency",
  ylim = c(0, 1200),
  xaxt = "n",
  main = "Barplot of health status",
  col = c("#BDDA85", "#FFE4B5"))

text(bp_10, 50, table(WVS_Bolivia$health_status), col = "black", pos = 3)

axis(side = 1, at = bp_10[c(1:2)], na.rm = T,
  labels = c("Good", "Not so good"),
  cex.axis = 0.75)
```



2.14 Influence on life: As we can see, data is piled up on the right, so it's left-skewed data. Minimal value is 1, while maximum is 10, and mean is 7.477. The most the most frequent values are 8 and 10 (460 and

463 respondents) and least frequent are 1 and 2 (18 and 17 respondents). In addition, this variable includes 40 missing values. And the type of this variable is numeric.

```
WVS_Bolivia$influence_on_life <- as.numeric(ifelse(WVS_Bolivia$influence_on_life<0,
                                                    NA, WVS_Bolivia$influence_on_life))
class(WVS_Bolivia$influence_on_life)
```

```
## [1] "numeric"
```

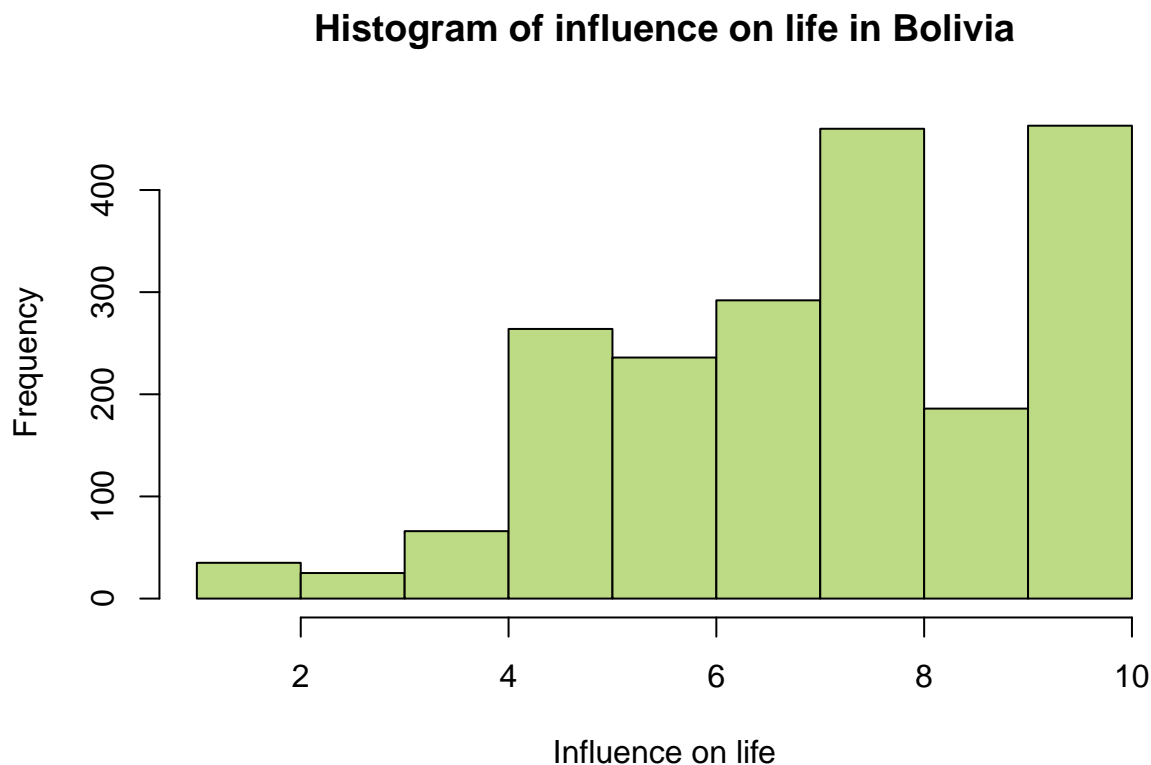
```
summary(WVS_Bolivia$influence_on_life)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.     NA's
##      1.000   6.000   8.000   7.477   9.000  10.000      40
```

```
table(WVS_Bolivia$influence_on_life)
```

```
##
##      1  2  3  4  5  6  7  8  9 10
##     18 17 25 66 264 236 292 460 186 463
```

```
hist(WVS_Bolivia$influence_on_life,
     main = "Histogram of influence on life in Bolivia",
     xlab = "Influence on life", col = "#BDDA85")
```



Let's remove all missed cases for further work. So, now we have 1891 observations.

```
WVS_Bolivia <- na.omit(WVS_Bolivia)
base::dim(WVS_Bolivia)
```

```
## [1] 1891 28
```

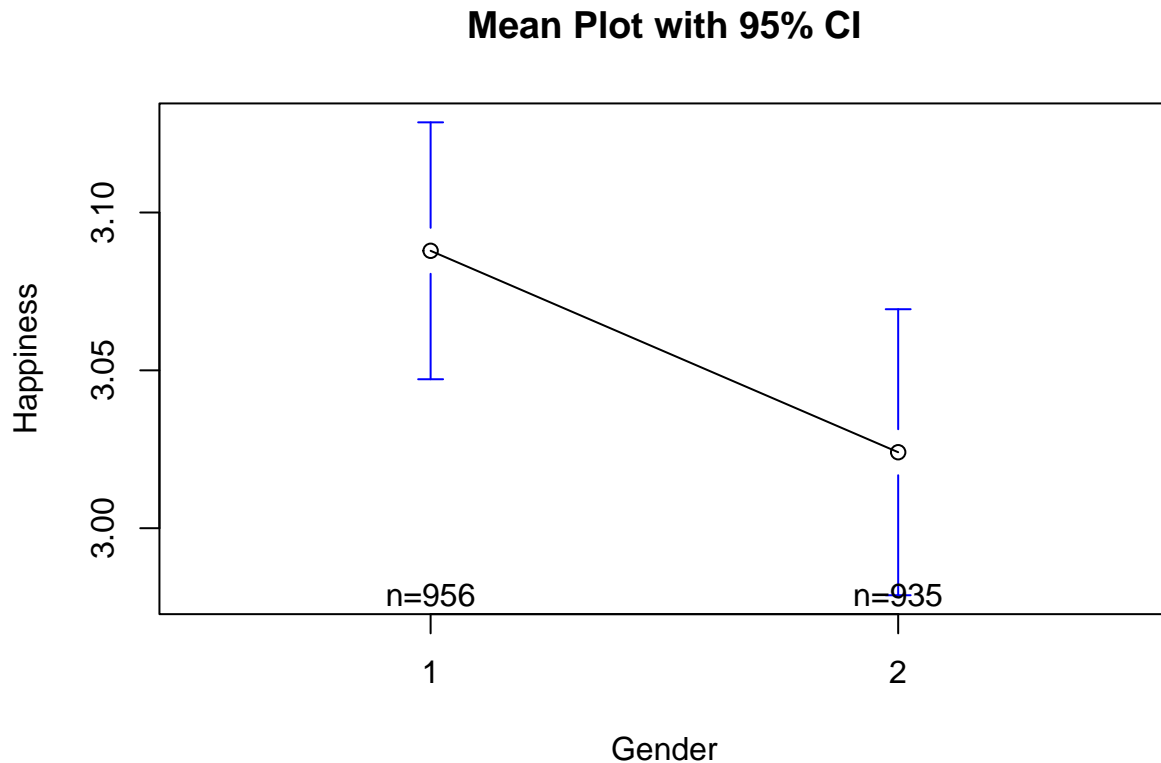
II. Descriptive statistics. The nature of the relations in the data – among the outcome and the predictors (simple bivariate tests).

1. **Happiness and gender:** p-value is significant ($0.04 < 0.05$), so we can speak about the association between these variables. Male respondents are likely to have higher level of happiness (~ 3.09) than female respondents (~ 3.02).

```
t.test(WVS_Bolivia$happiness ~ WVS_Bolivia$sex)
```

```
##
## Welch Two Sample t-test
##
## data: WVS_Bolivia$happiness by WVS_Bolivia$sex
## t = 2.0564, df = 1863.1, p-value = 0.03988
## alternative hypothesis: true difference in means between group 1 and group 2 is not equal to 0
## 95 percent confidence interval:
## 0.002953057 0.124650819
## sample estimates:
## mean in group 1 mean in group 2
## 3.087866 3.024064
```

```
plotmeans(happiness ~ sex, data = WVS_Bolivia,
          xlab = "Gender", ylab = "Happiness",
          main="Mean Plot with 95% CI")
```

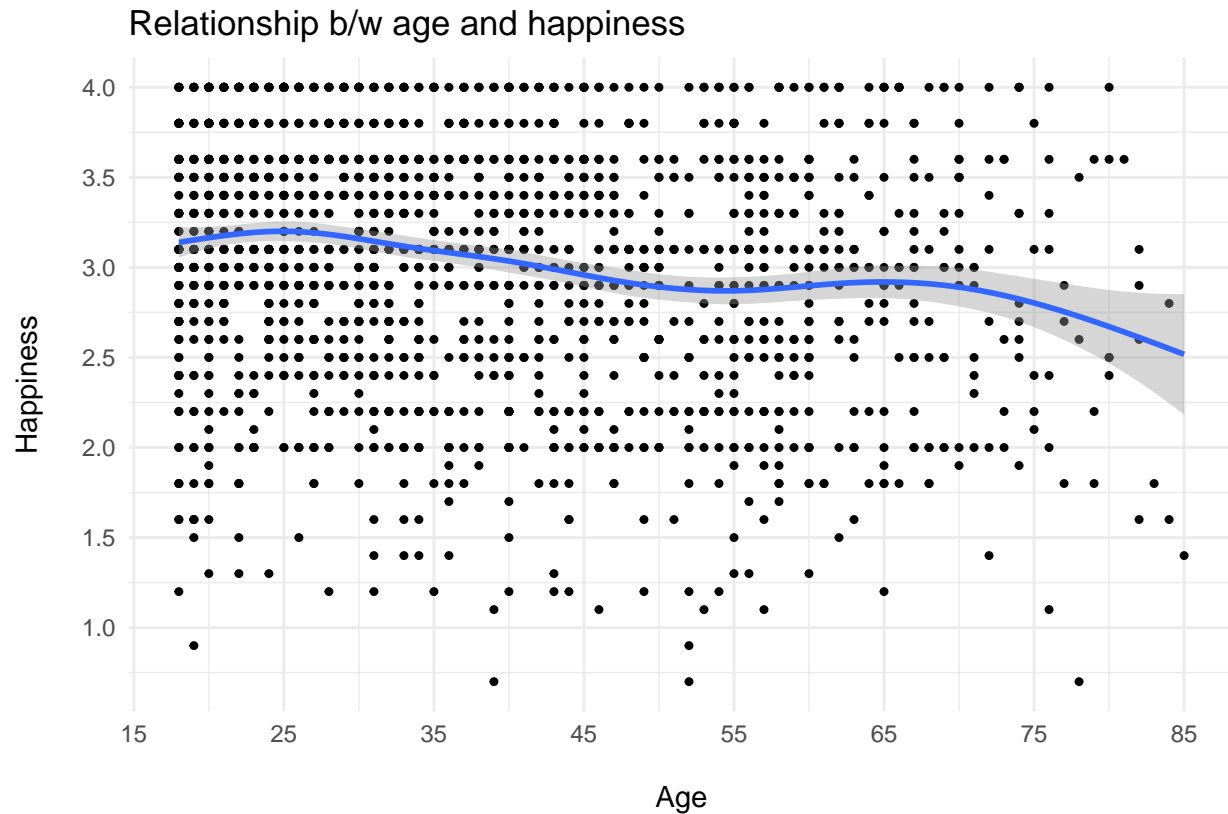
2. Happiness and age: p-value is significant (p-value = 3.265×10^{-15}), so we can speak about the correlation between these variables. Correlation is very low (-0.18) and negative: the older the respondent, the lower their level of happiness is. But if we look at the graph, we will see a non-linear relationship between variables. We can observe that the level of happiness among respondents aged 18 to 25 rises, then begins to gradually decrease until the age of 55, then slightly increases until the age of 70, and then drops quite sharply.

```
cor.test(WVS_Bolivia$happiness, WVS_Bolivia$age)
```

```
##
## Pearson's product-moment correlation
##
## data: WVS_Bolivia$happiness and WVS_Bolivia$age
## t = -7.9466, df = 1889, p-value = 3.265e-15
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## -0.223124 -0.135881
## sample estimates:
## cor
## -0.1798562
```

```
ggplot(data = WVS_Bolivia,
       mapping = aes(x = age, y = happiness)) +
  geom_point(size = 0.9) +
  geom_smooth()+
  ggtitle("Relationship b/w age and happiness")+
```

```
labs(x = "\nAge\n",
     y = "Happiness\n")+
scale_x_continuous(breaks = seq(15, 95, 10))+
scale_y_continuous(breaks = seq(0, 4, 0.5))+
theme_minimal()
```

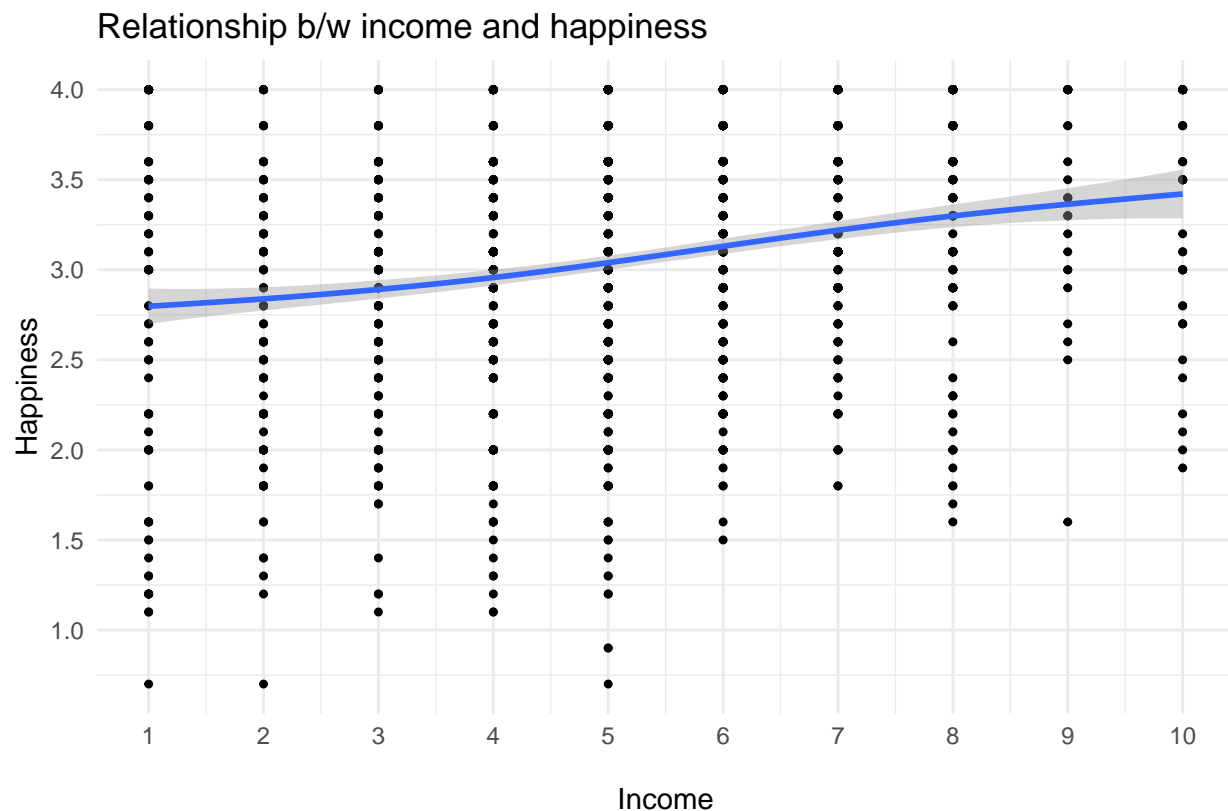


3. Happiness and income: p-value is significant ($p\text{-value} < 2.2e-16$), so we can speak about the correlation between these variables. Correlation is low (0.23) and positive: the higher the income, the higher the level of happiness is.

```
cor.test(WVS_Bolivia$happiness, WVS_Bolivia$income)
```

```
##
## Pearson's product-moment correlation
##
## data: WVS_Bolivia$happiness and WVS_Bolivia$income
## t = 10.236, df = 1889, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## 0.1860892 0.2715141
## sample estimates:
## cor
## 0.229243
```

```
ggplot(data = WVS_Bolivia,
       mapping = aes(x = income, y = happiness)) +
  geom_point(size = 0.9) +
  geom_smooth()+
  ggtitle("Relationship b/w income and happiness")+
  labs(x = "\nIncome\n",
       y = "Happiness")+
  scale_x_continuous(breaks = seq(0, 10, 1))+
  scale_y_continuous(breaks = seq(0, 4, 0.5))+
  theme_minimal()
```



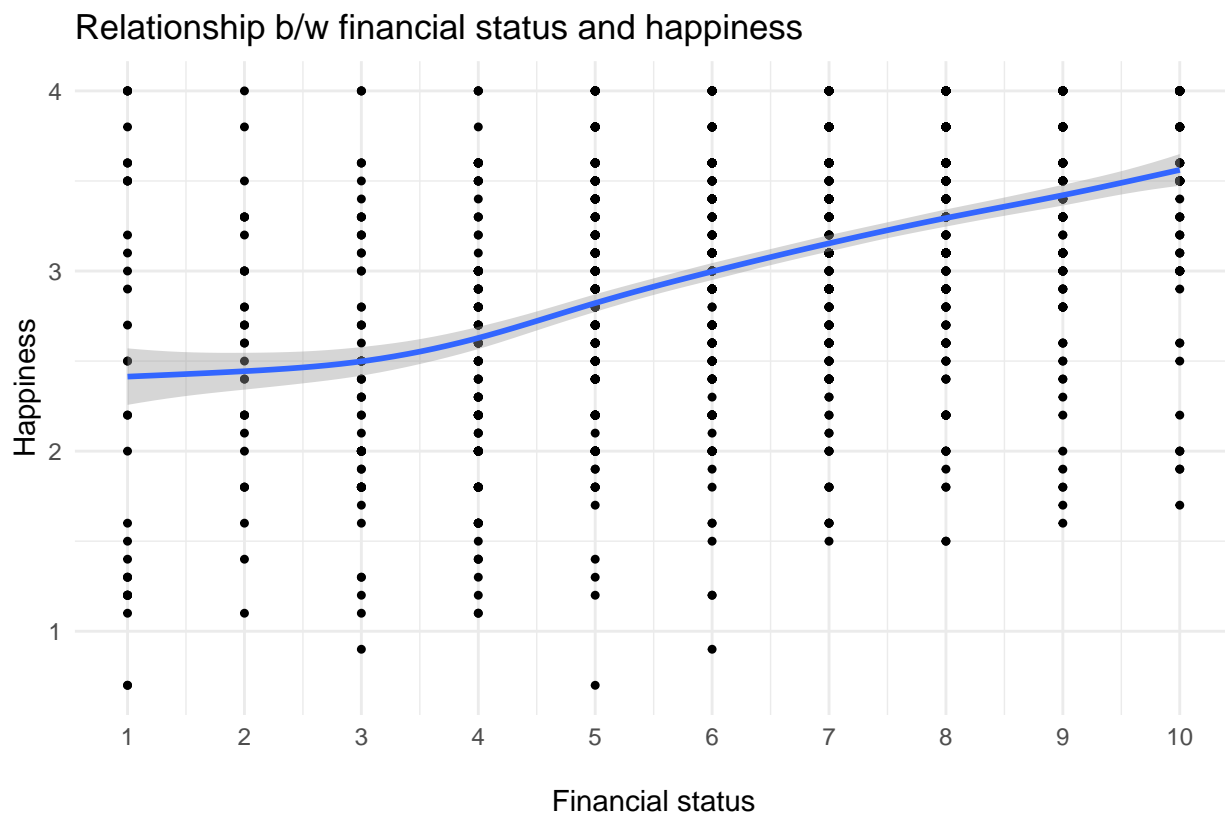
4. Happiness and financial status: p-value is significant ($p\text{-value} < 2.2e-16$), so we can speak about the correlation between these variables. Correlation is medium (0.45) and positive: the better the financial situation, the higher the level of happiness is.

```
cor.test(WVS_Bolivia$happiness, WVS_Bolivia$financial_status)
```

```
##
## Pearson's product-moment correlation
##
## data: WVS_Bolivia$happiness and WVS_Bolivia$financial_status
## t = 22.036, df = 1889, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## 0.4156036 0.4873512
```

```
## sample estimates:
##      cor
## 0.4522087
```

```
ggplot(data = WVS_Bolivia,
       mapping = aes(x = financial_status, y = happiness)) +
  geom_point(size = 0.9) +
  geom_smooth()+
  ggtitle("Relationship b/w financial status and happiness")+
  labs(x = "\nFinancial status\n",
       y = "Happiness")+
  scale_x_continuous(breaks = seq(0, 10, 1))+
  theme_minimal()
```

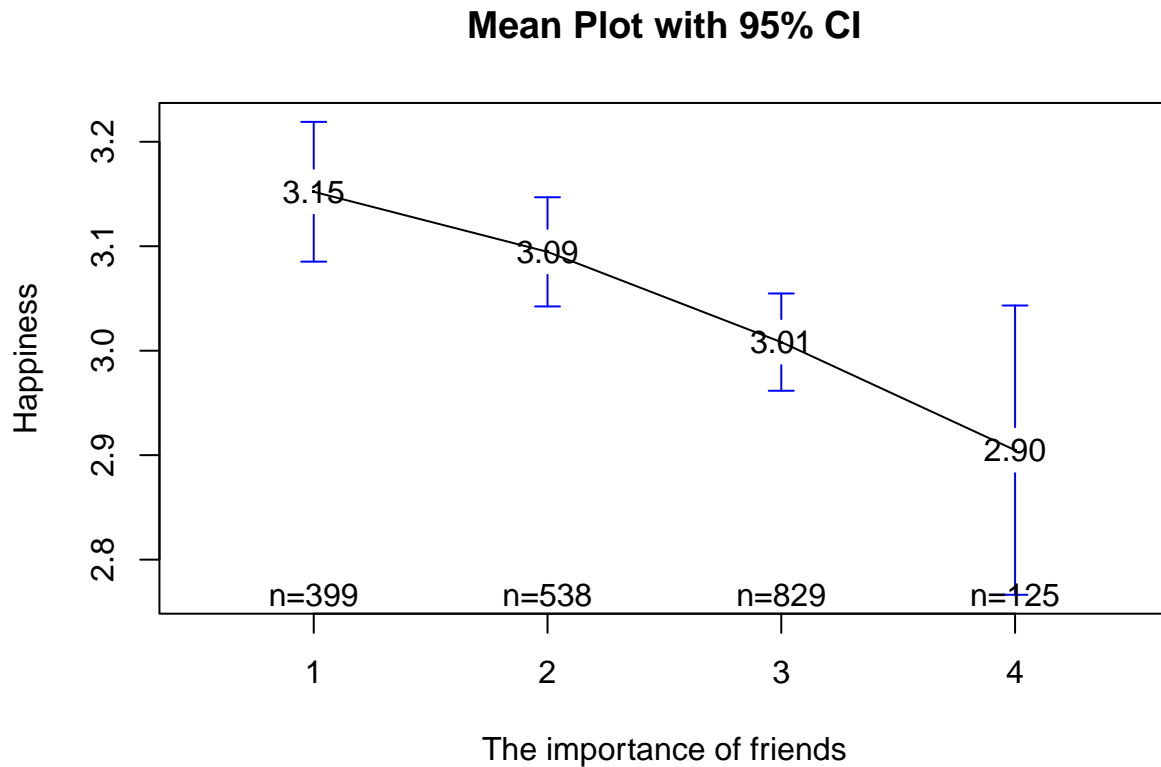


5. Happiness and friends importance: p-value is significant ($0.00014 < 0.05$), so we can speak about the association between these variables. For those respondents for whom friends are not very important or not at all important, the level of happiness is lower (~3.01 and ~2.90 respectively) than for those for whom friends are very important (~3.15).

```
summary(aov(WVS_Bolivia$happiness ~ WVS_Bolivia$friends_imp))
```

```
##              Df Sum Sq Mean Sq F value  Pr(>F)
## WVS_Bolivia$friends_imp    3    9.2   3.0802   6.835 0.00014 ***
## Residuals              1887  850.4   0.4506
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
plotmeans(happiness ~ friends_imp, data = WVS_Bolivia,
          xlab = "The importance of friends", ylab = "Happiness",
          main = "Mean Plot with 95% CI", mean.labels = T, digits = 2, pch=".")
```

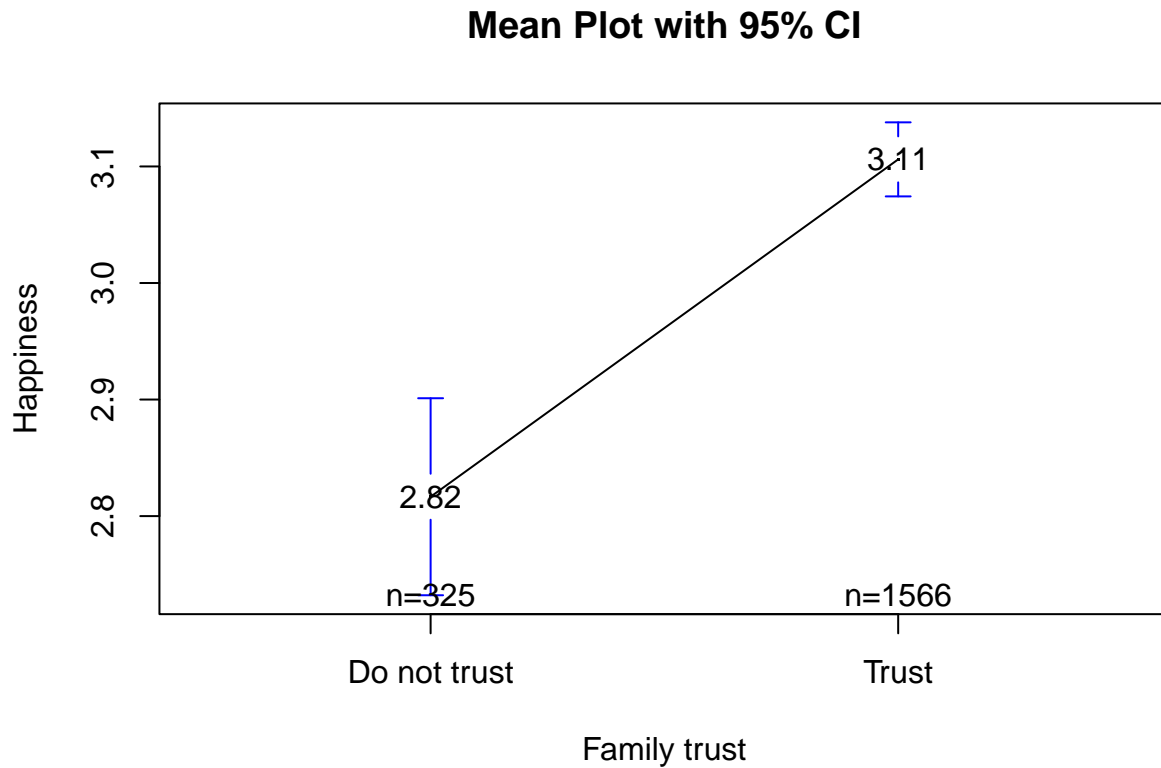


6. Happiness and family trust: p-value is significant (p-value = 1.39e-12), so we can speak about the association between these variables. Respondents who trust family are happier (~3.11) than those who do not trust (~2.82).

```
summary(aov(WVS_Bolivia$happiness ~ WVS_Bolivia$family_trust))
```

```
##               Df Sum Sq Mean Sq F value    Pr(>F)
## WVS_Bolivia$family_trust    1    22.5   22.549    50.89 1.39e-12 ***
## Residuals              1889   837.1    0.443
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
plotmeans(happiness ~ family_trust, data = WVS_Bolivia,
          xlab = "Family trust", ylab = "Happiness",
          main = "Mean Plot with 95% CI", mean.labels = T, digits = 2, pch=".")
```

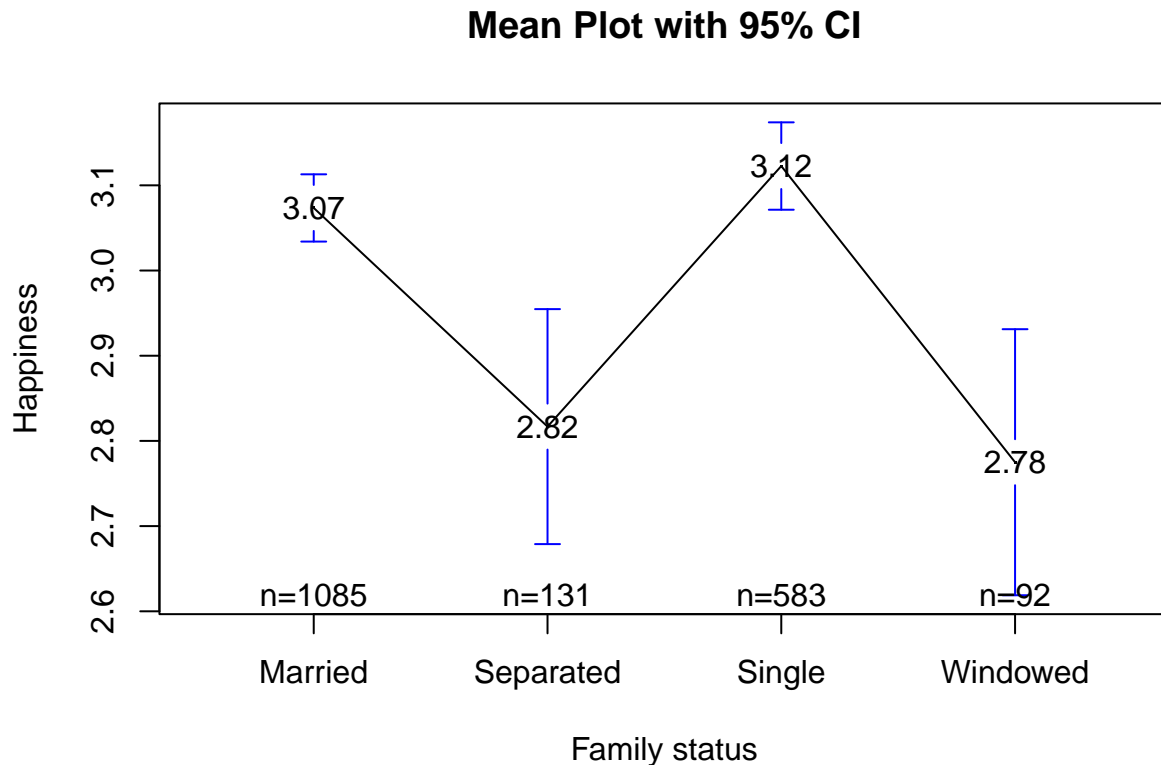


7. Happiness and family status: p-value is significant (p-value = 1.56e-08), so we can speak about the association between these variables. Respondents who are separated or widowed are less happy (~2.82 and ~2.78 respectively) than those who are married (~3.07) or single (~3.12).

```
summary(aov(WVS_Bolivia$happiness ~ WVS_Bolivia$family_status))
```

```
##               Df Sum Sq Mean Sq F value    Pr(>F)
## WVS_Bolivia$family_status    3   17.7    5.893   13.21 1.56e-08 ***
## Residuals              1887  841.9    0.446
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
plotmeans(happiness ~ family_status, data = WVS_Bolivia,
          xlab = "Family status", ylab = "Happiness",
          main = "Mean Plot with 95% CI", mean.labels = T, digits = 2, pch=".")
```



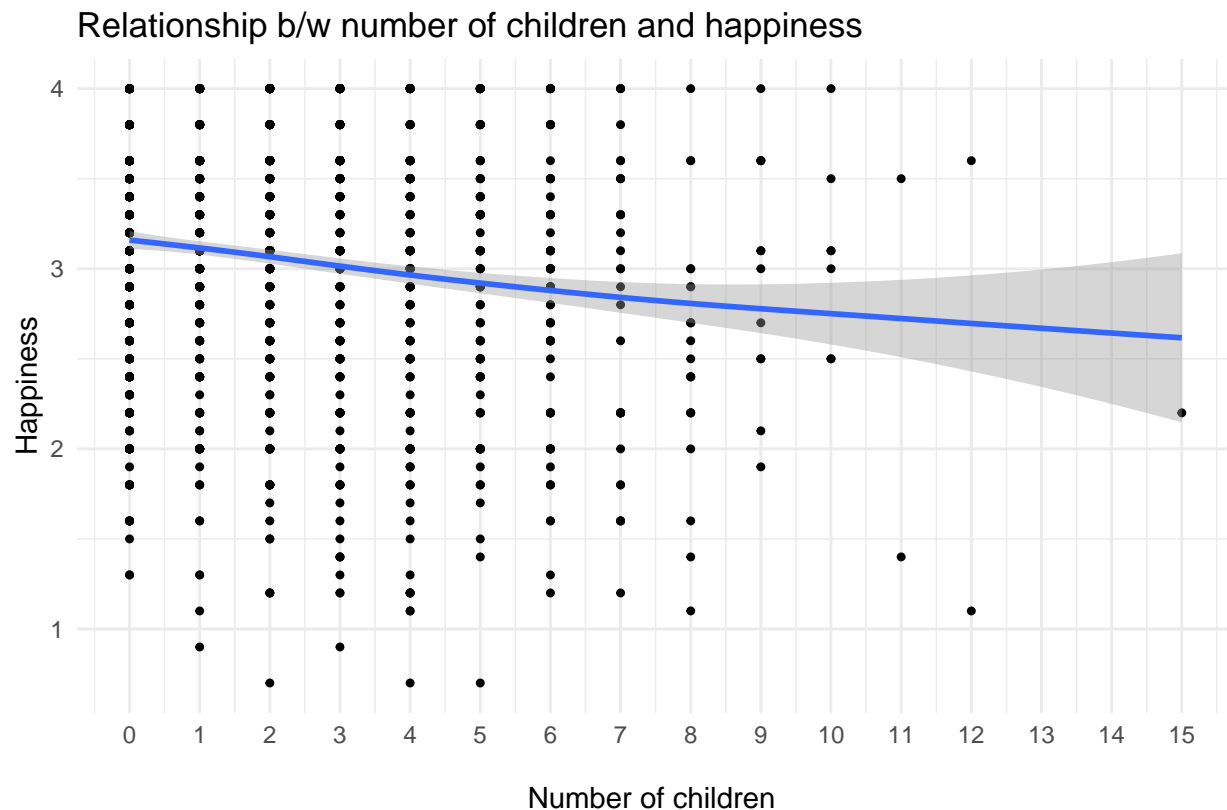
8. Happiness and children: p-value is significant (p-value = 2.672×10^{-11}), so we can speak about the correlation between these variables. Correlation is very low (-0.15) and negative: the more children in the respondent's family, the less happy the respondent is.

```
cor.test(WVS_Bolivia$happiness, WVS_Bolivia$children)
```

```
##
## Pearson's product-moment correlation
##
## data: WVS_Bolivia$happiness and WVS_Bolivia$children
## t = -6.7038, df = 1889, p-value = 2.672e-11
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## -0.1961698 -0.1081071
## sample estimates:
## cor
## -0.152441
```

```
ggplot(data = WVS_Bolivia,
       mapping = aes(x = children, y = happiness)) +
  geom_point(size = 0.9) +
  geom_smooth()+
  ggtitle("Relationship b/w number of children and happiness")+
  labs(x = "\nNumber of children \n",
       y = "Happiness")+
  theme_minimal()
```

```
scale_x_continuous(breaks = seq(0, 15, 1))+
theme_minimal()
```

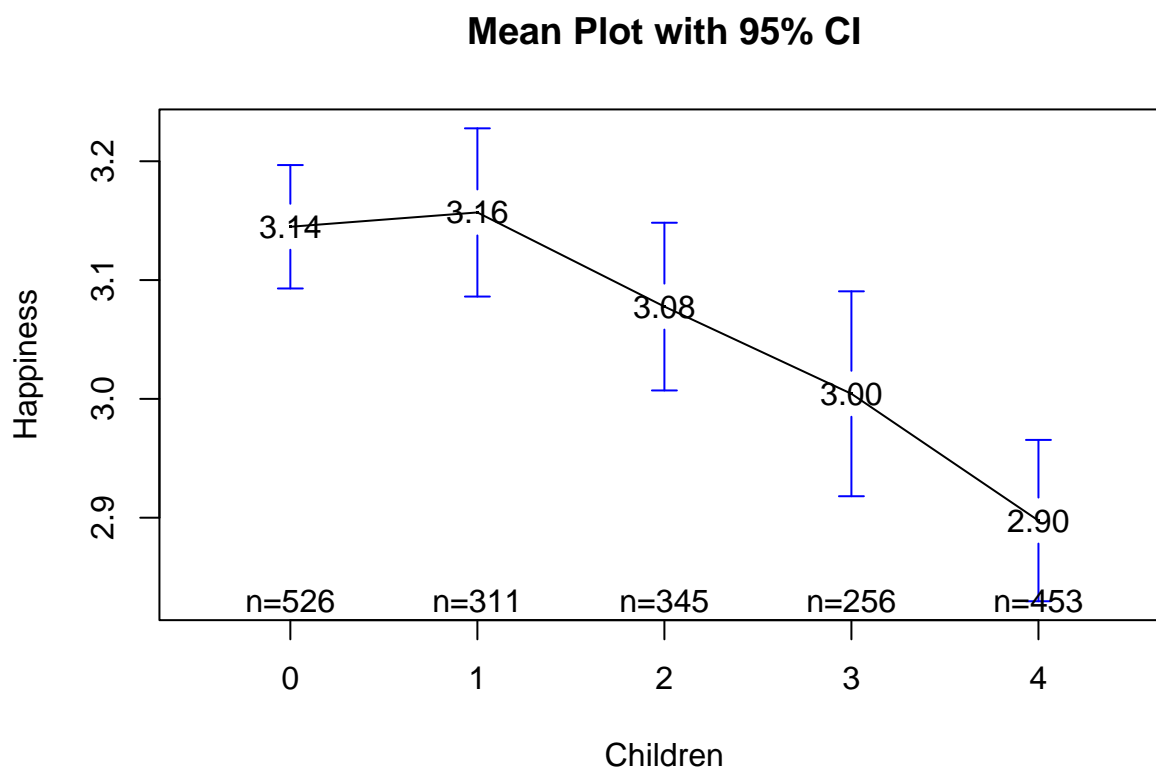


9. Happiness and children: p-value is significant (p-value = 8.62e-09, so we can speak about the association between these variables. Respondents who have three or more than three children in family are less happy (~2.90) than those who have one child (~3.16) or don't have children at all (~3.14).

```
summary(aov(WVS_Bolivia$happiness ~ WVS_Bolivia$children_1))
```

```
##                Df Sum Sq Mean Sq F value    Pr(>F)
## WVS_Bolivia$children_1      4    19.5    4.884    10.97 8.62e-09 ***
## Residuals                1886   840.1    0.445
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
plotmeans(happiness ~ children_1, data = WVS_Bolivia,
          xlab = "Children", ylab = "Happiness",
          main = "Mean Plot with 95% CI", mean.labels = T, digits = 2, pch=".")
```

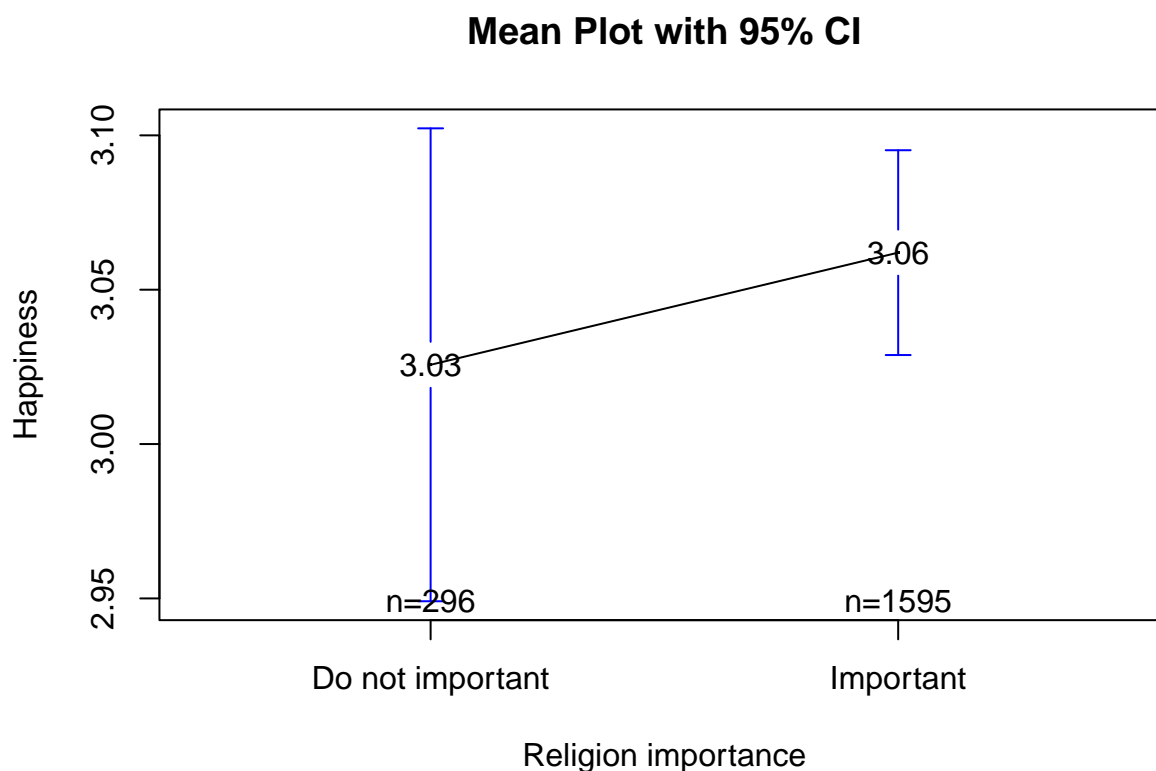



10. Happiness and religion importance: p-value isn't significant (p-value = 0.395), so there are no association between these variables.

```
summary(aov(WVS_Bolivia$happiness ~ WVS_Bolivia$religion_imp))
```

```
##               Df Sum Sq Mean Sq F value Pr(>F)
## WVS_Bolivia$religion_imp    1    0.3   0.3295   0.724  0.395
## Residuals              1889  859.3   0.4549
```

```
plotmeans(happiness ~ religion_imp, data = WVS_Bolivia,
           xlab = "Religion importance", ylab = "Happiness",
           main = "Mean Plot with 95% CI", mean.labels = T, digits = 2, pch=".")
```

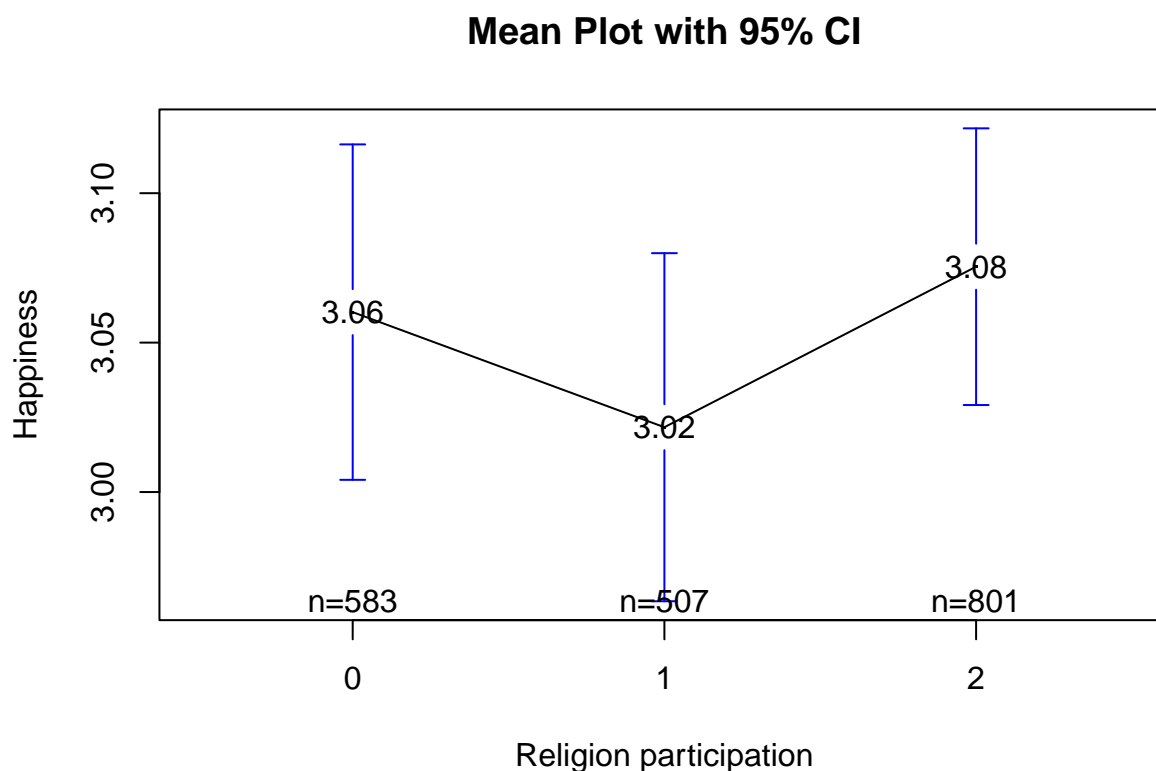


11. Happiness and religion participation: p-value isn't significant (p-value = 0.369), so there are no association between these variables.

```
summary(aov(WVS_Bolivia$happiness ~ WVS_Bolivia$religion_participation))
```

```
##                                Df Sum Sq Mean Sq F value Pr(>F)
## WVS_Bolivia$religion_participation  2    0.9   0.4542   0.999  0.369
## Residuals                      1888  858.7   0.4548
```

```
plotmeans(happiness ~ religion_participation, data = WVS_Bolivia,
           xlab = "Religion participation", ylab = "Happiness",
           main = "Mean Plot with 95% CI", mean.labels = T, digits = 2, pch=".")
```

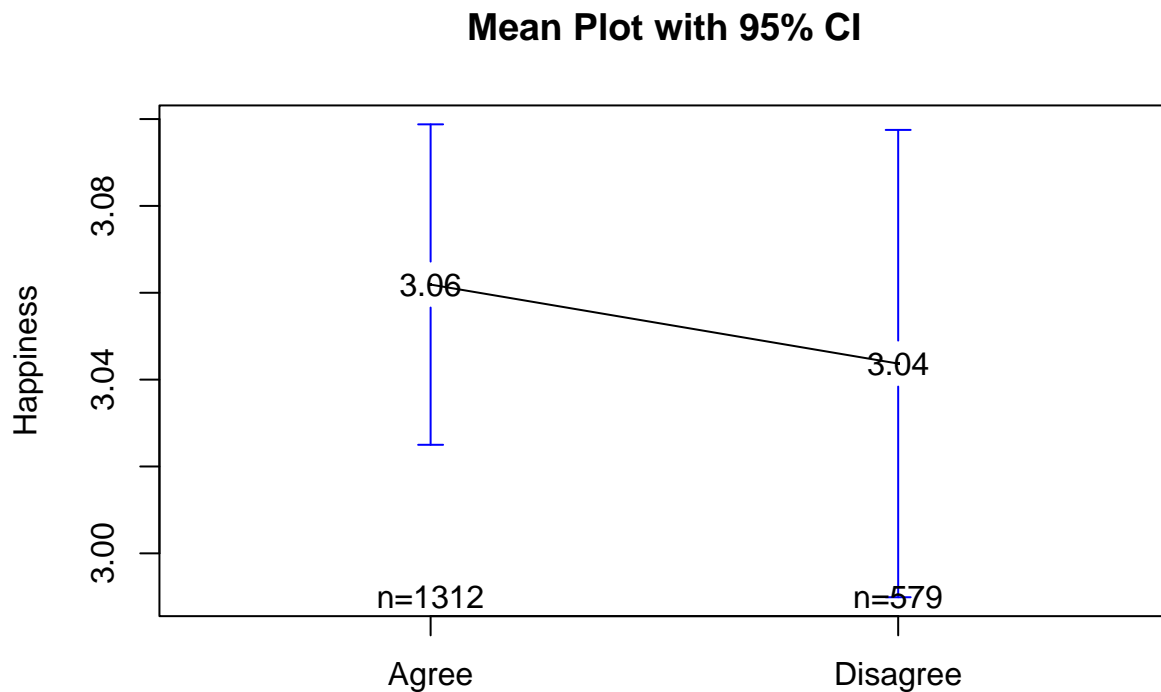


12. Happiness and the importance of religion over science: p-value isn't significant (p-value = 0.589), so there are no association between these variables.

```
summary(aov(WVS_Bolivia$happiness ~ WVS_Bolivia$religion_VS_science))
```

```
##                                Df Sum Sq Mean Sq F value Pr(>F)
## WVS_Bolivia$religion_VS_science  1    0.1    0.133   0.292  0.589
## Residuals                      1889  859.5    0.455
```

```
plotmeans(happiness ~ religion_VS_science, data = WVS_Bolivia,
  xlab = "Whenever science and religion conflict, religion is always right",
  ylab = "Happiness",
  main = "Mean Plot with 95% CI", mean.labels = T, digits = 2, pch=".")
```



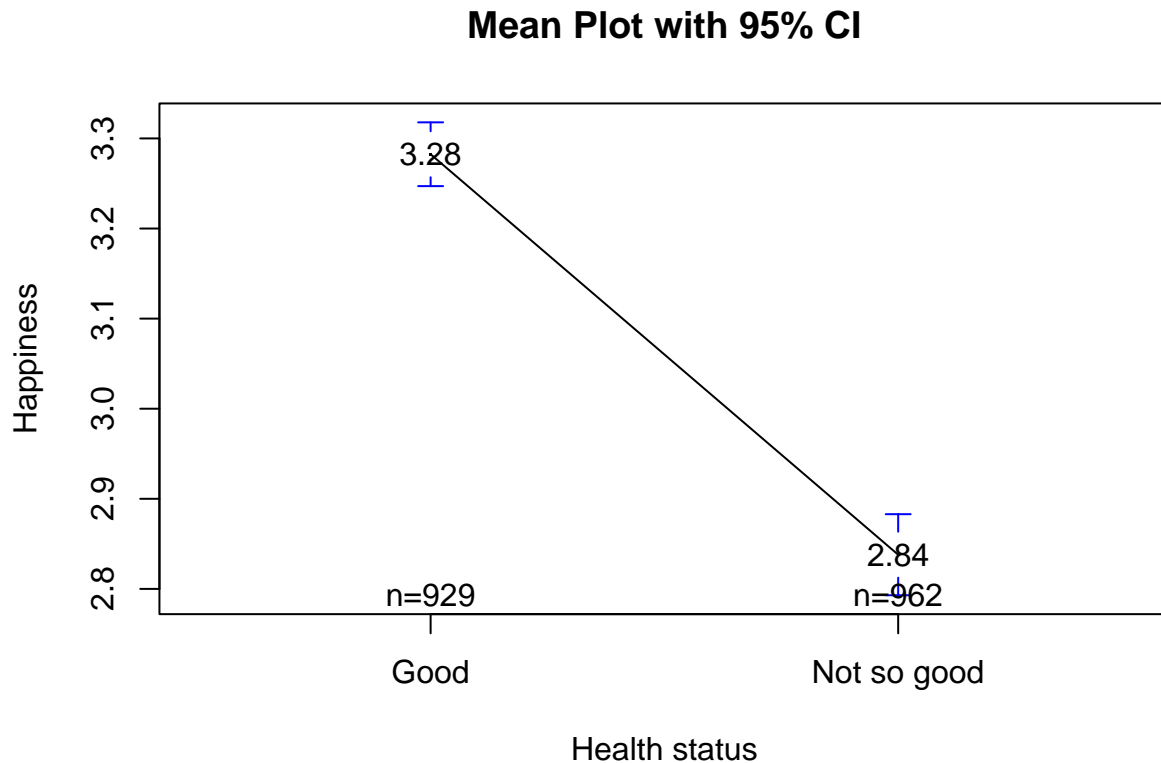
Whenever science and religion conflict, religion is always right

13. Happiness and health status: p-value is significant ($p\text{-value} < 2e-16$), so we can speak about the association between these variables. Respondents whose health state is good are happier (~3.28) than those respondents, whose health state is not so good (~2.84).

```
summary(aov(WVS_Bolivia$happiness ~ WVS_Bolivia$health_status))
```

```
##               Df Sum Sq Mean Sq F value Pr(>F)
## WVS_Bolivia$health_status    1   93.4   93.38   230.2 <2e-16 ***
## Residuals              1889  766.2    0.41
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
plotmeans(happiness ~ health_status, data = WVS_Bolivia,
          xlab = "Health status", ylab = "Happiness",
          main = "Mean Plot with 95% CI", mean.labels = T, digits = 2, pch=".")
```



14. Happiness and influence on life: p-value is significant ($p\text{-value} < 2.2e-16$), so we can speak about the correlation between these variables. Correlation is low (0.342) and positive: the higher the influence on life, the higher the level of happiness is.

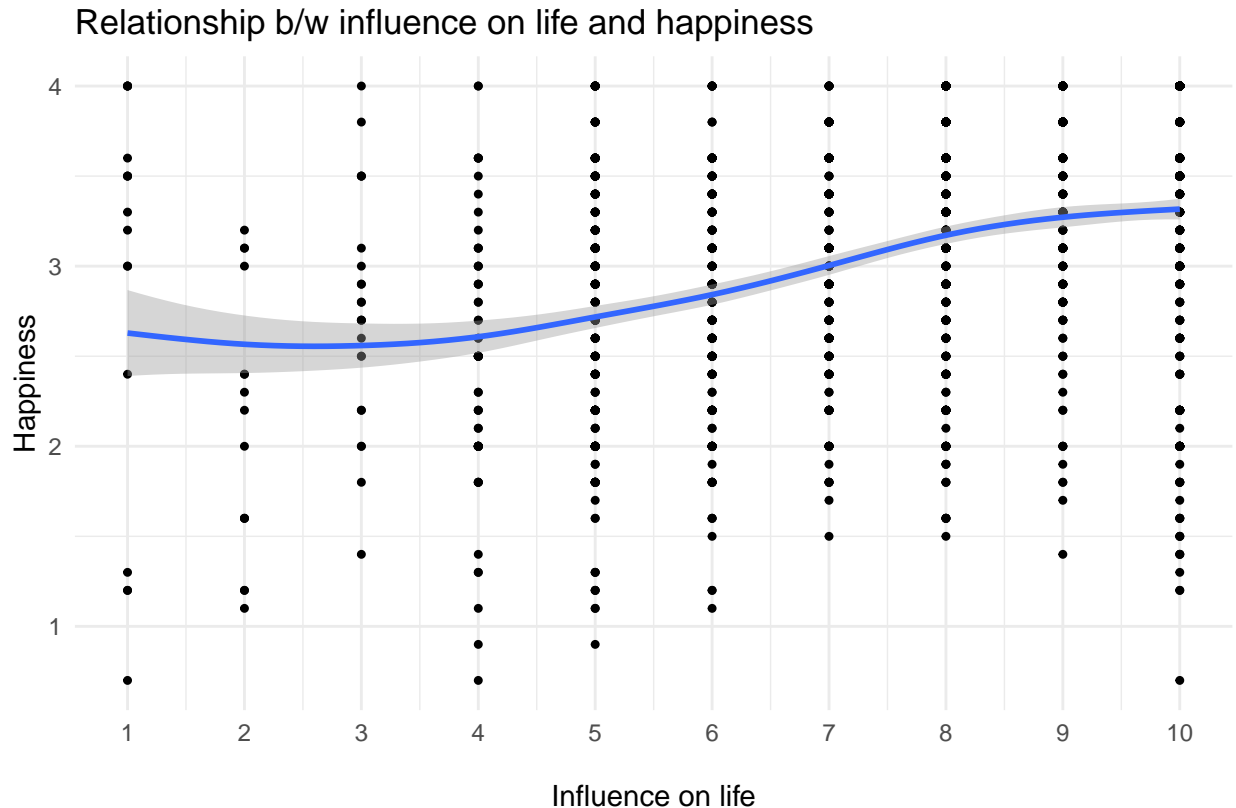
Also, we can notice a non-linear effect. It can be seen that for respondents who have little influence on their lives, the level of happiness is low and remains approximately the same in the range of 1-4 predictor units. Then, as the respondent's influence on life increases, happiness also increases in the range from 4 to 8 units of the predictor. Then the level of happiness stabilizes and almost does not change in the range from 8 to 10 units of influence on life.

```
cor.test(WVS_Bolivia$happiness, WVS_Bolivia$influence_on_life)
```

```
##
## Pearson's product-moment correlation
##
## data: WVS_Bolivia$happiness and WVS_Bolivia$influence_on_life
## t = 15.916, df = 1889, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## 0.3034995 0.3830117
## sample estimates:
## cor
## 0.3438719
```

```
ggplot(data = WVS_Bolivia,
       mapping = aes(x = influence_on_life, y = happiness)) +
```

```
geom_point(size = 0.9) +
geom_smooth()+
ggtitle("Relationship b/w influence on life and happiness")+
labs(x = "\nInfluence on life\n",
     y = "Happiness")+
scale_x_continuous(breaks = seq(0, 10, 1))+
theme_minimal()
```



III. Regression analysis

1. Multiple regression (with comparing models)

Model 1: sex & age

We use the controls variables - sex, age. And firstly, let's include only them in the regression model. We can see that model_1 is statistically significant (p-value = 2.412e-15) as well as both predictors. Anyway, Adjusted R-squared is really low (Adjusted R-squared = 0.034), so the proportion of the variance of the dependent variable explained by the considered model is 3%.

- 1) Gender: we interpret sex2 (female) compared to male. For women, compared to men, the level of happiness is lower by 0.07.

- 2) Age: we can see negative trend: an increase in age by 1 year entails an decrease of happiness by 0.008 points.

```
model_1 <- lm(WVS_Bolivia$happiness ~ WVS_Bolivia$sex + WVS_Bolivia$age)
summary(model_1)

##
## Call:
## lm(formula = WVS_Bolivia$happiness ~ WVS_Bolivia$sex + WVS_Bolivia$age)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.31224 -0.42085  0.06279  0.50243  1.30582
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.3845987   0.0428032   79.074  <2e-16 ***
## WVS_Bolivia$sex2 -0.0698149   0.0304965   -2.289   0.0222 *
## WVS_Bolivia$age -0.0077575   0.0009685   -8.009   2e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6628 on 1888 degrees of freedom
## Multiple R-squared:  0.03503,    Adjusted R-squared:  0.034
## F-statistic: 34.27 on 2 and 1888 DF,  p-value: 2.412e-15
```

Model 2: + income

We add income as a predictor to the previous model_1. So, model_2 is statistically significant (p-value: < 2.2e-16). Adjusted R-squared became higher (0.06812), therefore the proportion of the variance of the dependent variable explained by the considered model is 7%.

- 3) Income: we can see positive trend: an increase in income by 1 point entails an increase of happiness by 0.065. points. *It can be seen that the gender predictor has become insignificant, perhaps this is a sign of multicollinearity.

```
model_2 <- lm(WVS_Bolivia$happiness ~ WVS_Bolivia$sex + WVS_Bolivia$age
              + WVS_Bolivia$income)
summary(model_2)

##
## Call:
## lm(formula = WVS_Bolivia$happiness ~ WVS_Bolivia$sex + WVS_Bolivia$age +
##     WVS_Bolivia$income)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.31692 -0.40547  0.06334  0.49181  1.44658
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.3845987   0.0428032   79.074  <2e-16 ***
## WVS_Bolivia$sex2 -0.0698149   0.0304965   -2.289   0.0222 *
## WVS_Bolivia$age -0.0077575   0.0009685   -8.009   2e-15 ***
## WVS_Bolivia$income  0.0651111   0.0009685   6.754   1e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6628 on 1888 degrees of freedom
## Multiple R-squared:  0.06812,    Adjusted R-squared:  0.064
## F-statistic: 34.27 on 3 and 1888 DF,  p-value: 2.412e-15
```

```
## (Intercept)          2.9685394  0.0650831  45.611  < 2e-16 ***
## WVS_Bolivia$sex2     -0.0582096  0.0299851  -1.941   0.0524 .
## WVS_Bolivia$age      -0.0055459  0.0009873  -5.617  2.23e-08 ***
## WVS_Bolivia$income    0.0645772  0.0077114   8.374  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.651 on 1887 degrees of freedom
## Multiple R-squared:  0.0696, Adjusted R-squared:  0.06812
## F-statistic: 47.06 on 3 and 1887 DF,  p-value: < 2.2e-16
```

Moreover, according to anova, model_2 is better than model_1 (RSS is lower for model_2: 799.78 < 829.50).

```
anova(model_1, model_2)
```

```
## Analysis of Variance Table
##
## Model 1: WVS_Bolivia$happiness ~ WVS_Bolivia$sex + WVS_Bolivia$age
## Model 2: WVS_Bolivia$happiness ~ WVS_Bolivia$sex + WVS_Bolivia$age + WVS_Bolivia$income
##   Res.Df    RSS Df Sum of Sq    F    Pr(>F)
## 1    1888 829.50
## 2    1887 799.78  1    29.723 70.128 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Model 3: + financial status

We add financial_status as a predictor to the previous model_2. So, model_3 is statistically significant (p-value: < 2.2e-16) as well as added predictor. Adjusted R-squared became higher (0.2243), therefore the proportion of the variance of the dependent variable explained by the considered model is 22%.

- 4) Financial status: we can see positive trend: an increase in financial status by 1 point entails an increase of happiness by 0.135 points.

```
model_3 <- lm(WVS_Bolivia$happiness ~ WVS_Bolivia$sex + WVS_Bolivia$age +
              WVS_Bolivia$income + WVS_Bolivia$financial_status)
summary(model_3)
```

```
##
## Call:
## lm(formula = WVS_Bolivia$happiness ~ WVS_Bolivia$sex + WVS_Bolivia$age +
##     WVS_Bolivia$income + WVS_Bolivia$financial_status)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.04965 -0.38051  0.03054  0.42295  1.81765
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    2.2073765   0.0710395   31.073  < 2e-16 ***
```



```
## WVS_Bolivia$sex2          -0.0356173  0.0273815  -1.301    0.193
## WVS_Bolivia$age          -0.0041670  0.0009035  -4.612  4.26e-06 ***
## WVS_Bolivia$income        0.0301888  0.0072528   4.162  3.29e-05 ***
## WVS_Bolivia$financial_status 0.1346688  0.0068996  19.518 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.594 on 1886 degrees of freedom
## Multiple R-squared:  0.226, Adjusted R-squared:  0.2243
## F-statistic: 137.6 on 4 and 1886 DF, p-value: < 2.2e-16
```

Moreover, according to anova, model_3 is better than model_2 (RSS is lower for model_3: 665.37 < 799.78).

```
anova(model_2, model_3)
```

```
## Analysis of Variance Table
##
## Model 1: WVS_Bolivia$happiness ~ WVS_Bolivia$sex + WVS_Bolivia$age + WVS_Bolivia$income
## Model 2: WVS_Bolivia$happiness ~ WVS_Bolivia$sex + WVS_Bolivia$age + WVS_Bolivia$income +
##      WVS_Bolivia$financial_status
##   Res.Df    RSS Df Sum of Sq    F    Pr(>F)
## 1    1887 799.78
## 2    1886 665.37  1    134.41 380.97 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Model 4: + friends importance

We add friends importance as a predictor to the previous model_3. So, model_4 is statistically significant (p-value: <2.2e-16) as well as added predictor (more specifically, the categories 3 - “Not very important” and 4 - “Not at all important”). Adjusted R-squared became higher (0.227), therefore the proportion of the variance of the dependent variable explained by the considered model is 22,7%.

- 5) Friends importance: we interpret friends_imp3 (“Not very important”) and friends_imp4 (“Not at all important”) compared to “Very important”. Respondents for whom friends are not very important, compared to those for whom friends are very important, have the level of happiness lower by 0.101. And respondents for whom friends are not at all important, compared to those for whom friends are very important, have the level of happiness lower by 0.141.

```
model_4 <- lm(WVS_Bolivia$happiness ~ WVS_Bolivia$sex + WVS_Bolivia$age
              + WVS_Bolivia$income + WVS_Bolivia$financial_status
              + WVS_Bolivia$friends_imp)
summary(model_4)
```

```
##
## Call:
## lm(formula = WVS_Bolivia$happiness ~ WVS_Bolivia$sex + WVS_Bolivia$age +
##      WVS_Bolivia$income + WVS_Bolivia$financial_status + WVS_Bolivia$friends_imp)
##
## Residuals:
```

```
##      Min      1Q   Median      3Q      Max
## -2.01641 -0.39020  0.03362  0.41825  1.84546
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      2.2804881   0.0759670   30.019 < 2e-16 ***
## WVS_Bolivia$sex2    -0.0314900   0.0274335   -1.148  0.25117
## WVS_Bolivia$age     -0.0040942   0.0009055   -4.522 6.52e-06 ***
## WVS_Bolivia$income    0.0300866   0.0072403    4.155 3.39e-05 ***
## WVS_Bolivia$financial_status  0.1332775   0.0069079   19.294 < 2e-16 ***
## WVS_Bolivia$friends_imp2 -0.0515745   0.0392467   -1.314  0.18897
## WVS_Bolivia$friends_imp3 -0.1012732   0.0361717   -2.800  0.00517 **
## WVS_Bolivia$friends_imp4 -0.1408874   0.0611404   -2.304  0.02131 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5929 on 1883 degrees of freedom
## Multiple R-squared:  0.2301, Adjusted R-squared:  0.2272
## F-statistic: 80.39 on 7 and 1883 DF,  p-value: < 2.2e-16
```

Moreover, according to anova, model_4 is better than model_3 (RSS is lower for model_4: 661.83 < 665.37).

```
anova(model_3, model_4)
```

```
## Analysis of Variance Table
##
## Model 1: WVS_Bolivia$happiness ~ WVS_Bolivia$sex + WVS_Bolivia$age + WVS_Bolivia$income +
##      WVS_Bolivia$financial_status
## Model 2: WVS_Bolivia$happiness ~ WVS_Bolivia$sex + WVS_Bolivia$age + WVS_Bolivia$income +
##      WVS_Bolivia$financial_status + WVS_Bolivia$friends_imp
##   Res.Df    RSS Df Sum of Sq    F Pr(>F)
## 1    1886 665.37
## 2    1883 661.83  3     3.5419 3.3591 0.01812 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Model 5: + family trust

We add family trust as a predictor to the previous model_4. So, model_5 is statistically significant (p-value: <2.2e-16) as well as added predictor. Adjusted R-squared became higher (0.2344), therefore the proportion of the variance of the dependent variable explained by the considered model is 23,4%.

- 6) Family trust: we interpret family_trust (“Trust”) compared to “Do not trust”. Respondents who trust to family, compared to those who do not trust, have the level of happiness higher by 0.159.

```
model_5 <- lm(WVS_Bolivia$happiness ~ WVS_Bolivia$sex + WVS_Bolivia$age
+ WVS_Bolivia$income + WVS_Bolivia$financial_status
+ WVS_Bolivia$friends_imp + WVS_Bolivia$family_trust)
summary(model_5)
```

```
##
## Call:
## lm(formula = WVS_Bolivia$happiness ~ WVS_Bolivia$sex + WVS_Bolivia$age +
##      WVS_Bolivia$income + WVS_Bolivia$financial_status + WVS_Bolivia$friends_imp +
##      WVS_Bolivia$family_trust)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.04811 -0.39338  0.02972  0.41188  1.92450
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      2.1735753   0.0795364   27.328 < 2e-16 ***
## WVS_Bolivia$sex2    -0.0256130   0.0273387   -0.937  0.34894
## WVS_Bolivia$age     -0.0040941   0.0009012   -4.543 5.90e-06 ***
## WVS_Bolivia$income    0.0293106   0.0072086    4.066 4.98e-05 ***
## WVS_Bolivia$financial_status  0.1293494   0.0069350   18.652 < 2e-16 ***
## WVS_Bolivia$friends_imp2 -0.0572891   0.0390850   -1.466  0.14288
## WVS_Bolivia$friends_imp3 -0.0938792   0.0360426   -2.605  0.00927 **
## WVS_Bolivia$friends_imp4 -0.1356006   0.0608662   -2.228  0.02601 *
## WVS_Bolivia$family_trustTrust  0.1586526   0.0366207    4.332 1.55e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5901 on 1882 degrees of freedom
## Multiple R-squared:  0.2377, Adjusted R-squared:  0.2344
## F-statistic: 73.35 on 8 and 1882 DF, p-value: < 2.2e-16
```

According to anova, model_5 is better than model_4 (RSS is lower for model_5: 655.30 < 661.83).

```
anova(model_4, model_5)
```

```
## Analysis of Variance Table
##
## Model 1: WVS_Bolivia$happiness ~ WVS_Bolivia$sex + WVS_Bolivia$age + WVS_Bolivia$income +
##      WVS_Bolivia$financial_status + WVS_Bolivia$friends_imp
## Model 2: WVS_Bolivia$happiness ~ WVS_Bolivia$sex + WVS_Bolivia$age + WVS_Bolivia$income +
##      WVS_Bolivia$financial_status + WVS_Bolivia$friends_imp +
##      WVS_Bolivia$family_trust
##   Res.Df    RSS Df Sum of Sq    F    Pr(>F)
## 1    1883 661.83
## 2    1882 655.30  1    6.5352 18.769 1.553e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Model 6: + family status

We add family status as a predictor to the previous model_6. So, model_7 is statistically significant (p-value: <2.2e-16) as well as added predictor (more specifically, the categories “Separated” and “Single”). Adjusted R-squared became higher (0.2391), therefore the proportion of the variance of the dependent variable explained by the considered model is 23,9%.

- 7) Family status: we interpret family statuses compared to “Married”. Respondents who are separated or single, compared to those who are married, have the level of happiness lower by 0.11.

```
model_6 <- lm(WVS_Bolivia$happiness ~ WVS_Bolivia$sex + WVS_Bolivia$age
              + WVS_Bolivia$income + WVS_Bolivia$financial_status
              + WVS_Bolivia$friends_imp + WVS_Bolivia$family_trust
              + WVS_Bolivia$family_status)
summary(model_6)
```

```
##
## Call:
## lm(formula = WVS_Bolivia$happiness ~ WVS_Bolivia$sex + WVS_Bolivia$age +
##     WVS_Bolivia$income + WVS_Bolivia$financial_status + WVS_Bolivia$friends_imp +
##     WVS_Bolivia$family_trust + WVS_Bolivia$family_status)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.96926 -0.37496  0.03393  0.41757  2.00534
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      2.262003   0.085312  26.515 < 2e-16 ***
## WVS_Bolivia$sex2    -0.031304   0.027792  -1.126  0.26015
## WVS_Bolivia$age     -0.005010   0.001082  -4.632 3.87e-06 ***
## WVS_Bolivia$income    0.029897   0.007193   4.156 3.38e-05 ***
## WVS_Bolivia$financial_status  0.129149   0.006926  18.647 < 2e-16 ***
## WVS_Bolivia$friends_imp2 -0.053039   0.038995  -1.360  0.17395
## WVS_Bolivia$friends_imp3 -0.099859   0.035981  -2.775  0.00557 **
## WVS_Bolivia$friends_imp4 -0.136400   0.060757  -2.245  0.02488 *
## WVS_Bolivia$family_trustTrust  0.152622   0.036597   4.170 3.18e-05 ***
## WVS_Bolivia$family_statusSeparated -0.109352   0.055048  -1.986  0.04713 *
## WVS_Bolivia$family_statusSingle -0.110545   0.034565  -3.198  0.00141 **
## WVS_Bolivia$family_statusWindowed -0.091920   0.068223  -1.347  0.17803
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5883 on 1879 degrees of freedom
## Multiple R-squared:  0.2435, Adjusted R-squared:  0.2391
## F-statistic: 54.98 on 11 and 1879 DF, p-value: < 2.2e-16
```

According to anova, model_6 is better than model_5 (RSS is lower for model_6: 650.3 < 655.3)

```
anova(model_5, model_6)
```

```
## Analysis of Variance Table
##
## Model 1: WVS_Bolivia$happiness ~ WVS_Bolivia$sex + WVS_Bolivia$age + WVS_Bolivia$income +
##     WVS_Bolivia$financial_status + WVS_Bolivia$friends_imp +
##     WVS_Bolivia$family_trust
## Model 2: WVS_Bolivia$happiness ~ WVS_Bolivia$sex + WVS_Bolivia$age + WVS_Bolivia$income +
##     WVS_Bolivia$financial_status + WVS_Bolivia$friends_imp +
##     WVS_Bolivia$family_trust + WVS_Bolivia$family_status
```

```
##   Res.Df   RSS Df Sum of Sq      F   Pr(>F)
## 1    1882 655.3
## 2    1879 650.3  3     4.9936 4.8096 0.002434 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Model 7: + number of children

We add number of children as a predictor to the previous model_6. So, model_7 is statistically significant (p-value: $<2.2e-16$), but added predictor isn't.

```
model_7 <- lm(WVS_Bolivia$happiness ~ WVS_Bolivia$sex + WVS_Bolivia$age
              + WVS_Bolivia$income + WVS_Bolivia$financial_status
              + WVS_Bolivia$friends_imp + WVS_Bolivia$family_trust
              + WVS_Bolivia$family_status + WVS_Bolivia$children)
summary(model_7)
```

```
##
## Call:
## lm(formula = WVS_Bolivia$happiness ~ WVS_Bolivia$sex + WVS_Bolivia$age +
##     WVS_Bolivia$income + WVS_Bolivia$financial_status + WVS_Bolivia$friends_imp +
##     WVS_Bolivia$family_trust + WVS_Bolivia$family_status + WVS_Bolivia$children)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.97068 -0.37617  0.03725  0.41934  2.02135
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      2.274421   0.085798  26.509 < 2e-16 ***
## WVS_Bolivia$sex2    -0.030454   0.027794  -1.096  0.273338
## WVS_Bolivia$age     -0.004243   0.001224  -3.465  0.000542 ***
## WVS_Bolivia$income    0.029050   0.007219   4.024  5.95e-05 ***
## WVS_Bolivia$financial_status  0.128840   0.006928  18.596 < 2e-16 ***
## WVS_Bolivia$friends_imp2 -0.056504   0.039073  -1.446  0.148306
## WVS_Bolivia$friends_imp3 -0.100381   0.035975  -2.790  0.005319 **
## WVS_Bolivia$friends_imp4 -0.131418   0.060858  -2.159  0.030945 *
## WVS_Bolivia$family_trustTrust  0.151154   0.036606   4.129  3.80e-05 ***
## WVS_Bolivia$family_statusSeparated -0.115523   0.055230  -2.092  0.036600 *
## WVS_Bolivia$family_statusSingle -0.130296   0.037582  -3.467  0.000538 ***
## WVS_Bolivia$family_statusWindowed -0.092076   0.068208  -1.350  0.177206
## WVS_Bolivia$children    -0.012223   0.009139  -1.337  0.181260
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5882 on 1878 degrees of freedom
## Multiple R-squared:  0.2442, Adjusted R-squared:  0.2394
## F-statistic: 50.57 on 12 and 1878 DF,  p-value: < 2.2e-16
```

According to anova, we can't say that model_7 is better than model_6 even though RSS is lower for model_7, because this difference is not statistically significant (p-value = 0.181, >0.05).

```
anova(model_6, model_7)
```

```
## Analysis of Variance Table
##
## Model 1: WVS_Bolivia$happiness ~ WVS_Bolivia$sex + WVS_Bolivia$age + WVS_Bolivia$income +
##      WVS_Bolivia$financial_status + WVS_Bolivia$friends_imp +
##      WVS_Bolivia$family_trust + WVS_Bolivia$family_status
## Model 2: WVS_Bolivia$happiness ~ WVS_Bolivia$sex + WVS_Bolivia$age + WVS_Bolivia$income +
##      WVS_Bolivia$financial_status + WVS_Bolivia$friends_imp +
##      WVS_Bolivia$family_trust + WVS_Bolivia$family_status + WVS_Bolivia$children
##   Res.Df    RSS Df Sum of Sq    F Pr(>F)
## 1    1879 650.30
## 2    1878 649.69  1   0.61875 1.7886 0.1813
```

Model 8: + number of children (categorical)

We add number of children (categorical) as a predictor to the previous model_8. So, model_9 is statistically significant (p-value: $<2.2e-16$), but added predictor isn't.

```
model_8 <- lm(WVS_Bolivia$happiness ~ WVS_Bolivia$sex + WVS_Bolivia$age
              + WVS_Bolivia$income + WVS_Bolivia$financial_status
              + WVS_Bolivia$friends_imp + WVS_Bolivia$family_trust
              + WVS_Bolivia$family_status + WVS_Bolivia$children + WVS_Bolivia$children_1)
summary(model_8)
```

```
##
## Call:
## lm(formula = WVS_Bolivia$happiness ~ WVS_Bolivia$sex + WVS_Bolivia$age +
##      WVS_Bolivia$income + WVS_Bolivia$financial_status + WVS_Bolivia$friends_imp +
##      WVS_Bolivia$family_trust + WVS_Bolivia$family_status + WVS_Bolivia$children +
##      WVS_Bolivia$children_1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.92893 -0.37480  0.02527  0.41974  2.01598
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      2.246824   0.094039  23.892 < 2e-16 ***
## WVS_Bolivia$sex2    -0.033701   0.027985  -1.204 0.228643
## WVS_Bolivia$age     -0.004113   0.001240  -3.317 0.000927 ***
## WVS_Bolivia$income    0.029835   0.007229   4.127 3.83e-05 ***
## WVS_Bolivia$financial_status  0.129373   0.006937  18.649 < 2e-16 ***
## WVS_Bolivia$friends_imp2 -0.056470   0.039066  -1.445 0.148491
## WVS_Bolivia$friends_imp3 -0.099992   0.035986  -2.779 0.005513 **
## WVS_Bolivia$friends_imp4 -0.128042   0.060877  -2.103 0.035574 *
## WVS_Bolivia$family_trustTrust  0.154336   0.036631   4.213 2.64e-05 ***
## WVS_Bolivia$family_statusSeparated -0.113695   0.055228  -2.059 0.039664 *
## WVS_Bolivia$family_statusSingle -0.125998   0.049015  -2.571 0.010228 *
## WVS_Bolivia$family_statusWindowed -0.089260   0.068859  -1.296 0.195043
## WVS_Bolivia$children -0.010922   0.017495  -0.624 0.532524
```

```
## WVS_Bolivia$children_11      0.070813    0.054933    1.289 0.197530
## WVS_Bolivia$children_12     -0.006688    0.067623   -0.099 0.921225
## WVS_Bolivia$children_13     -0.037309    0.080828   -0.462 0.644434
## WVS_Bolivia$children_14      0.016708    0.109047    0.153 0.878246
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.588 on 1874 degrees of freedom
## Multiple R-squared:  0.2464, Adjusted R-squared:  0.2399
## F-statistic: 38.29 on 16 and 1874 DF,  p-value: < 2.2e-16
```

According to anova, we can't say that model_8 is better than model_7 even though RSS is lower for model_8, because this difference is not statistically significant (p-value = 0.253, > 0.05).

```
anova(model_7, model_8)
```

```
## Analysis of Variance Table
##
## Model 1: WVS_Bolivia$happiness ~ WVS_Bolivia$sex + WVS_Bolivia$age + WVS_Bolivia$income +
##   WVS_Bolivia$financial_status + WVS_Bolivia$friends_imp +
##   WVS_Bolivia$family_trust + WVS_Bolivia$family_status + WVS_Bolivia$children
## Model 2: WVS_Bolivia$happiness ~ WVS_Bolivia$sex + WVS_Bolivia$age + WVS_Bolivia$income +
##   WVS_Bolivia$financial_status + WVS_Bolivia$friends_imp +
##   WVS_Bolivia$family_trust + WVS_Bolivia$family_status + WVS_Bolivia$children +
##   WVS_Bolivia$children_1
##   Res.Df    RSS Df Sum of Sq    F Pr(>F)
## 1    1878 649.69
## 2    1874 647.83   4     1.852 1.3394 0.253
```

Model 9: + religion importance

We add religion importance as a predictor to the previous model_8. So, model_9 is statistically significant (p-value: <2.2e-16), as well as added predictor (more specifically, the category “Important”). Adjusted R-squared became a little higher (0.241), therefore the proportion of the variance of the dependent variable explained by the considered model is 24,1%.

- 10) Religion importance: we interpret religion importance compared to “Not important”. Respondents for whom religion is important, compared to those for whom religion is not important, have the level of happiness higher by 0.076.

```
model_9 <- lm(WVS_Bolivia$happiness ~ WVS_Bolivia$sex + WVS_Bolivia$age
+ WVS_Bolivia$income + WVS_Bolivia$financial_status
+ WVS_Bolivia$friends_imp + WVS_Bolivia$family_trust
+ WVS_Bolivia$family_status + WVS_Bolivia$children
+ WVS_Bolivia$children_1 + WVS_Bolivia$religion_imp)
summary(model_9)
```

```
##
## Call:
## lm(formula = WVS_Bolivia$happiness ~ WVS_Bolivia$sex + WVS_Bolivia$age +
```

```
## WVS_Bolivia$income + WVS_Bolivia$financial_status + WVS_Bolivia$friends_imp +
## WVS_Bolivia$family_trust + WVS_Bolivia$family_status + WVS_Bolivia$children +
## WVS_Bolivia$children_1 + WVS_Bolivia$religion_imp)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.93970 -0.37316  0.02682  0.41795  2.00993
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      2.186746    0.098718  22.152 < 2e-16 ***
## WVS_Bolivia$sex2    -0.040987    0.028203  -1.453  0.14631
## WVS_Bolivia$age     -0.004153    0.001239  -3.351  0.00082 ***
## WVS_Bolivia$income    0.029761    0.007223   4.120 3.95e-05 ***
## WVS_Bolivia$financial_status  0.129569    0.006932  18.690 < 2e-16 ***
## WVS_Bolivia$friends_imp2 -0.056521    0.039036  -1.448  0.14781
## WVS_Bolivia$friends_imp3 -0.096139    0.036010  -2.670  0.00766 **
## WVS_Bolivia$friends_imp4 -0.120955    0.060934  -1.985  0.04729 *
## WVS_Bolivia$family_trustTrust  0.151155    0.036638   4.126 3.86e-05 ***
## WVS_Bolivia$family_statusSeparated -0.113585    0.055184  -2.058  0.03970 *
## WVS_Bolivia$family_statusSingle -0.119054    0.049101  -2.425  0.01542 *
## WVS_Bolivia$family_statusWindowed -0.088000    0.068808  -1.279  0.20108
## WVS_Bolivia$children -0.011107    0.017482  -0.635  0.52526
## WVS_Bolivia$children_11  0.072204    0.054895   1.315  0.18856
## WVS_Bolivia$children_12 -0.006883    0.067570  -0.102  0.91888
## WVS_Bolivia$children_13 -0.041506    0.080792  -0.514  0.60750
## WVS_Bolivia$children_14  0.013285    0.108976   0.122  0.90299
## WVS_Bolivia$religion_impImportant  0.076096    0.038328   1.985  0.04725 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5875 on 1873 degrees of freedom
## Multiple R-squared:  0.2479, Adjusted R-squared:  0.2411
## F-statistic: 36.32 on 17 and 1873 DF, p-value: < 2.2e-16
```

According to anova, model_9 is better than model_8: RSS is lower for model_9 ($646.47 < 647.83$).

```
anova(model_8, model_9)
```

```
## Analysis of Variance Table
##
## Model 1: WVS_Bolivia$happiness ~ WVS_Bolivia$sex + WVS_Bolivia$age + WVS_Bolivia$income +
## WVS_Bolivia$financial_status + WVS_Bolivia$friends_imp +
## WVS_Bolivia$family_trust + WVS_Bolivia$family_status + WVS_Bolivia$children +
## WVS_Bolivia$children_1
## Model 2: WVS_Bolivia$happiness ~ WVS_Bolivia$sex + WVS_Bolivia$age + WVS_Bolivia$income +
## WVS_Bolivia$financial_status + WVS_Bolivia$friends_imp +
## WVS_Bolivia$family_trust + WVS_Bolivia$family_status + WVS_Bolivia$children +
## WVS_Bolivia$children_1 + WVS_Bolivia$religion_imp
##   Res.Df    RSS Df Sum of Sq    F Pr(>F)
## 1    1874 647.83
## 2    1873 646.47  1    1.3605 3.9418 0.04725 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```


Model 10: + religion participation

We add religion participation as a predictor to the previous model_9. So, model_10 is statistically significant (p-value: <2.2e-16), but added predictor isn't.

*It can be seen that the religion_imp predictor has become insignificant, perhaps this is a sign of multicollinearity.

```
model_10 <- lm(WVS_Bolivia$happiness ~ WVS_Bolivia$sex + WVS_Bolivia$age
               + WVS_Bolivia$income + WVS_Bolivia$financial_status
               + WVS_Bolivia$friends_imp + WVS_Bolivia$family_trust
               + WVS_Bolivia$family_status + WVS_Bolivia$children
               + WVS_Bolivia$children_1 + WVS_Bolivia$religion_imp
               + WVS_Bolivia$religion_participation)
summary(model_10)
```

```
##
## Call:
## lm(formula = WVS_Bolivia$happiness ~ WVS_Bolivia$sex + WVS_Bolivia$age +
##     WVS_Bolivia$income + WVS_Bolivia$financial_status + WVS_Bolivia$friends_imp +
##     WVS_Bolivia$family_trust + WVS_Bolivia$family_status + WVS_Bolivia$children +
##     WVS_Bolivia$children_1 + WVS_Bolivia$religion_imp + WVS_Bolivia$religion_participation)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.97178 -0.37955  0.03294  0.41876  1.98471
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      2.200567   0.099500  22.116 < 2e-16 ***
## WVS_Bolivia$sex2    -0.041121   0.028201  -1.458 0.144972
## WVS_Bolivia$age     -0.004219   0.001239  -3.406 0.000673 ***
## WVS_Bolivia$income    0.029969   0.007221   4.150 3.47e-05 ***
## WVS_Bolivia$financial_status  0.129206   0.006929  18.647 < 2e-16 ***
## WVS_Bolivia$friends_imp2 -0.053368   0.039034  -1.367 0.171720
## WVS_Bolivia$friends_imp3 -0.094225   0.036030  -2.615 0.008989 **
## WVS_Bolivia$friends_imp4 -0.121975   0.061107  -1.996 0.046069 *
## WVS_Bolivia$family_trustTrust  0.153229   0.036623   4.184 3.00e-05 ***
## WVS_Bolivia$family_statusSeparated -0.108138   0.055201  -1.959 0.050263 .
## WVS_Bolivia$family_statusSingle -0.117953   0.049075  -2.404 0.016334 *
## WVS_Bolivia$family_statusWindowed -0.093019   0.068811  -1.352 0.176602
## WVS_Bolivia$children -0.012482   0.017493  -0.714 0.475587
## WVS_Bolivia$children_11  0.070119   0.054861   1.278 0.201369
## WVS_Bolivia$children_12 -0.003687   0.067542  -0.055 0.956469
## WVS_Bolivia$children_13 -0.037054   0.080763  -0.459 0.646430
## WVS_Bolivia$children_14  0.014419   0.108892   0.132 0.894670
## WVS_Bolivia$religion_impImportant  0.069542   0.039202   1.774 0.076233 .
## WVS_Bolivia$religion_participation1 -0.054747   0.036057  -1.518 0.129089
## WVS_Bolivia$religion_participation2  0.019604   0.033276   0.589 0.555846
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.587 on 1871 degrees of freedom
## Multiple R-squared:  0.2499, Adjusted R-squared:  0.2423
```

```
## F-statistic: 32.81 on 19 and 1871 DF, p-value: < 2.2e-16
```

According to anova, we can't say that model_10 is better than model_9, because this difference is not statistically significant (p-value = 0.086, >0.05).

```
anova(model_9, model_10)
```

```
## Analysis of Variance Table
##
## Model 1: WVS_Bolivia$happiness ~ WVS_Bolivia$sex + WVS_Bolivia$age + WVS_Bolivia$income +
##   WVS_Bolivia$financial_status + WVS_Bolivia$friends_imp +
##   WVS_Bolivia$family_trust + WVS_Bolivia$family_status + WVS_Bolivia$children +
##   WVS_Bolivia$children_1 + WVS_Bolivia$religion_imp
## Model 2: WVS_Bolivia$happiness ~ WVS_Bolivia$sex + WVS_Bolivia$age + WVS_Bolivia$income +
##   WVS_Bolivia$financial_status + WVS_Bolivia$friends_imp +
##   WVS_Bolivia$family_trust + WVS_Bolivia$family_status + WVS_Bolivia$children +
##   WVS_Bolivia$children_1 + WVS_Bolivia$religion_imp + WVS_Bolivia$religion_participation
##   Res.Df    RSS Df Sum of Sq    F Pr(>F)
## 1    1873 646.47
## 2    1871 644.78  2     1.6946 2.4587 0.08582 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Model 11: + importance of religion over science

We add the importance of religion over science as a predictor to the previous model_10. So, model_11 is statistically significant (p-value: <2.2e-16), but added predictor isn't.

```
model_11 <- lm(WVS_Bolivia$happiness ~ WVS_Bolivia$sex + WVS_Bolivia$age
+ WVS_Bolivia$income + WVS_Bolivia$financial_status
+ WVS_Bolivia$friends_imp + WVS_Bolivia$family_trust + WVS_Bolivia$family_status
+ WVS_Bolivia$children + WVS_Bolivia$children_1 + WVS_Bolivia$religion_imp
+ WVS_Bolivia$religion_participation + WVS_Bolivia$religion_VS_science)
summary(model_11)
```

```
##
## Call:
## lm(formula = WVS_Bolivia$happiness ~ WVS_Bolivia$sex + WVS_Bolivia$age +
##   WVS_Bolivia$income + WVS_Bolivia$financial_status + WVS_Bolivia$friends_imp +
##   WVS_Bolivia$family_trust + WVS_Bolivia$family_status + WVS_Bolivia$children +
##   WVS_Bolivia$children_1 + WVS_Bolivia$religion_imp + WVS_Bolivia$religion_participation +
##   WVS_Bolivia$religion_VS_science)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.92945 -0.37588  0.03186  0.41321  1.97981
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    2.232183   0.100987  22.104 < 2e-16
## WVS_Bolivia$sex -0.042419   0.028194  -1.505 0.132605
```

```

## WVS_Bolivia$age -0.004196 0.001238 -3.389 0.000715
## WVS_Bolivia$income 0.030194 0.007218 4.183 3.01e-05
## WVS_Bolivia$financial_status 0.129241 0.006925 18.663 < 2e-16
## WVS_Bolivia$friends_imp2 -0.052183 0.039016 -1.337 0.181235
## WVS_Bolivia$friends_imp3 -0.094453 0.036008 -2.623 0.008785
## WVS_Bolivia$friends_imp4 -0.125098 0.061095 -2.048 0.040741
## WVS_Bolivia$family_trustTrust 0.155100 0.036616 4.236 2.39e-05
## WVS_Bolivia$family_statusSeparated -0.107433 0.055170 -1.947 0.051647
## WVS_Bolivia$family_statusSingle -0.115981 0.049058 -2.364 0.018173
## WVS_Bolivia$family_statusWindowed -0.095790 0.068787 -1.393 0.163919
## WVS_Bolivia$children -0.012684 0.017483 -0.725 0.468237
## WVS_Bolivia$children_11 0.070822 0.054830 1.292 0.196634
## WVS_Bolivia$children_12 -0.003098 0.067503 -0.046 0.963396
## WVS_Bolivia$children_13 -0.036809 0.080715 -0.456 0.648413
## WVS_Bolivia$children_14 0.010211 0.108853 0.094 0.925272
## WVS_Bolivia$religion_impImportant 0.053658 0.040164 1.336 0.181717
## WVS_Bolivia$religion_participation1 -0.057488 0.036067 -1.594 0.111129
## WVS_Bolivia$religion_participation2 0.012630 0.033482 0.377 0.706057
## WVS_Bolivia$religion_VS_scienceDisagree -0.055602 0.030952 -1.796 0.072591
##
## (Intercept) ***
## WVS_Bolivia$sex2
## WVS_Bolivia$age ***
## WVS_Bolivia$income ***
## WVS_Bolivia$financial_status ***
## WVS_Bolivia$friends_imp2
## WVS_Bolivia$friends_imp3 **
## WVS_Bolivia$friends_imp4 *
## WVS_Bolivia$family_trustTrust ***
## WVS_Bolivia$family_statusSeparated .
## WVS_Bolivia$family_statusSingle *
## WVS_Bolivia$family_statusWindowed
## WVS_Bolivia$children
## WVS_Bolivia$children_11
## WVS_Bolivia$children_12
## WVS_Bolivia$children_13
## WVS_Bolivia$children_14
## WVS_Bolivia$religion_impImportant
## WVS_Bolivia$religion_participation1
## WVS_Bolivia$religion_participation2
## WVS_Bolivia$religion_VS_scienceDisagree .
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5867 on 1870 degrees of freedom
## Multiple R-squared: 0.2512, Adjusted R-squared: 0.2432
## F-statistic: 31.37 on 20 and 1870 DF, p-value: < 2.2e-16

```

According to anova, we can't say that model_11 is better than model_10, because this difference is not statistically significant (p-value = 0.073, >0.05).

```
anova(model_10, model_11)
```

```
## Analysis of Variance Table
```

```
##
## Model 1: WVS_Bolivia$happiness ~ WVS_Bolivia$sex + WVS_Bolivia$age + WVS_Bolivia$income +
##   WVS_Bolivia$financial_status + WVS_Bolivia$friends_imp +
##   WVS_Bolivia$family_trust + WVS_Bolivia$family_status + WVS_Bolivia$children +
##   WVS_Bolivia$children_1 + WVS_Bolivia$religion_imp + WVS_Bolivia$religion_participation
## Model 2: WVS_Bolivia$happiness ~ WVS_Bolivia$sex + WVS_Bolivia$age + WVS_Bolivia$income +
##   WVS_Bolivia$financial_status + WVS_Bolivia$friends_imp +
##   WVS_Bolivia$family_trust + WVS_Bolivia$family_status + WVS_Bolivia$children +
##   WVS_Bolivia$children_1 + WVS_Bolivia$religion_imp + WVS_Bolivia$religion_participation +
##   WVS_Bolivia$religion_VS_science
##   Res.Df    RSS Df Sum of Sq    F Pr(>F)
## 1    1871 644.78
## 2    1870 643.67  1    1.1108 3.2271 0.07259 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Model 12: + health status

We add health status a predictor to the previous model_13. So, model_14 is statistically significant (p-value: $<2.2e-16$), as well as each category of the added predictor. Adjusted R-squared became higher (0.2778), therefore the proportion of the variance of the dependent variable explained by the considered model is 27,8%.

- 11) Health status: we interpret category of health status compared to “Good”. Respondents whose health status is “Not so good”, compared to those whose health status “Good”, have the level of happiness lower by 0.269.

*It can be seen that the religion_imp predictor has become insignificant, perhaps this is a sign of multicollinearity.

```
model_12 <- lm(WVS_Bolivia$happiness ~ WVS_Bolivia$sex + WVS_Bolivia$age
+ WVS_Bolivia$income + WVS_Bolivia$financial_status
+ WVS_Bolivia$friends_imp + WVS_Bolivia$family_trust + WVS_Bolivia$family_status
+ WVS_Bolivia$children + WVS_Bolivia$children_1 + WVS_Bolivia$religion_imp
+ WVS_Bolivia$religion_participation + WVS_Bolivia$religion_VS_science
+ WVS_Bolivia$health_status)
summary(model_12)
```

```
##
## Call:
## lm(formula = WVS_Bolivia$happiness ~ WVS_Bolivia$sex + WVS_Bolivia$age +
##   WVS_Bolivia$income + WVS_Bolivia$financial_status + WVS_Bolivia$friends_imp +
##   WVS_Bolivia$family_trust + WVS_Bolivia$family_status + WVS_Bolivia$children +
##   WVS_Bolivia$children_1 + WVS_Bolivia$religion_imp + WVS_Bolivia$religion_participation +
##   WVS_Bolivia$religion_VS_science + WVS_Bolivia$health_status)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.06751 -0.37931  0.01996  0.39382  1.98626
##
## Coefficients:
```

```

##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)      2.433968   0.100906  24.121 < 2e-16
## WVS_Bolivia$sex2 -0.027491   0.027586  -0.997 0.319119
## WVS_Bolivia$age  -0.002953   0.001216  -2.428 0.015296
## WVS_Bolivia$income 0.026997   0.007059   3.824 0.000135
## WVS_Bolivia$financial_status 0.115846   0.006910  16.765 < 2e-16
## WVS_Bolivia$friends_imp2 -0.068848   0.038154  -1.804 0.071323
## WVS_Bolivia$friends_imp3 -0.087853   0.035183  -2.497 0.012609
## WVS_Bolivia$friends_imp4 -0.112970   0.059697  -1.892 0.058593
## WVS_Bolivia$family_trustTrust 0.134317   0.035836   3.748 0.000184
## WVS_Bolivia$family_statusSeparated -0.120731   0.053913  -2.239 0.025248
## WVS_Bolivia$family_statusSingle -0.116648   0.047924  -2.434 0.015025
## WVS_Bolivia$family_statusWindowed -0.108350   0.067210  -1.612 0.107106
## WVS_Bolivia$children -0.008606   0.017084  -0.504 0.614506
## WVS_Bolivia$children_11 0.072389   0.053563   1.351 0.176706
## WVS_Bolivia$children_12 -0.003419   0.065942  -0.052 0.958659
## WVS_Bolivia$children_13 -0.032885   0.078850  -0.417 0.676689
## WVS_Bolivia$children_14 0.013889   0.106337   0.131 0.896094
## WVS_Bolivia$religion_impImportant 0.051224   0.039236   1.306 0.191871
## WVS_Bolivia$religion_participation1 -0.050834   0.035241  -1.442 0.149334
## WVS_Bolivia$religion_participation2 0.001226   0.032730   0.037 0.970122
## WVS_Bolivia$religion_VS_scienceDisagree -0.062455   0.030245  -2.065 0.039063
## WVS_Bolivia$health_statusNot so good -0.268895   0.028259  -9.516 < 2e-16
##
## (Intercept)      ***
## WVS_Bolivia$sex2
## WVS_Bolivia$age      *
## WVS_Bolivia$income    ***
## WVS_Bolivia$financial_status    ***
## WVS_Bolivia$friends_imp2      .
## WVS_Bolivia$friends_imp3      *
## WVS_Bolivia$friends_imp4      .
## WVS_Bolivia$family_trustTrust    ***
## WVS_Bolivia$family_statusSeparated *
## WVS_Bolivia$family_statusSingle *
## WVS_Bolivia$family_statusWindowed
## WVS_Bolivia$children
## WVS_Bolivia$children_11
## WVS_Bolivia$children_12
## WVS_Bolivia$children_13
## WVS_Bolivia$children_14
## WVS_Bolivia$religion_impImportant
## WVS_Bolivia$religion_participation1
## WVS_Bolivia$religion_participation2
## WVS_Bolivia$religion_VS_scienceDisagree *
## WVS_Bolivia$health_statusNot so good    ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5731 on 1869 degrees of freedom
## Multiple R-squared:  0.2858, Adjusted R-squared:  0.2778
## F-statistic: 35.62 on 21 and 1869 DF, p-value: < 2.2e-16

```

According o anova, model_12 is better than model_11 (RSS is lower for model_12: 613.92 < 643.67).

```
anova(model_11, model_12)
```

```
## Analysis of Variance Table
##
## Model 1: WVS_Bolivia$happiness ~ WVS_Bolivia$sex + WVS_Bolivia$age + WVS_Bolivia$income +
##   WVS_Bolivia$financial_status + WVS_Bolivia$friends_imp +
##   WVS_Bolivia$family_trust + WVS_Bolivia$family_status + WVS_Bolivia$children +
##   WVS_Bolivia$children_1 + WVS_Bolivia$religion_imp + WVS_Bolivia$religion_participation +
##   WVS_Bolivia$religion_VS_science
## Model 2: WVS_Bolivia$happiness ~ WVS_Bolivia$sex + WVS_Bolivia$age + WVS_Bolivia$income +
##   WVS_Bolivia$financial_status + WVS_Bolivia$friends_imp +
##   WVS_Bolivia$family_trust + WVS_Bolivia$family_status + WVS_Bolivia$children +
##   WVS_Bolivia$children_1 + WVS_Bolivia$religion_imp + WVS_Bolivia$religion_participation +
##   WVS_Bolivia$religion_VS_science + WVS_Bolivia$health_status
##   Res.Df    RSS Df Sum of Sq    F    Pr(>F)
## 1    1870 643.67
## 2    1869 613.92  1    29.742 90.545 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Model 13: + influence on life

We add influence on life as a predictor to the previous model_12. So, model_13 is statistically significant (p-value: < 2.2e-16) as well added predictor. Adjusted R-squared became higher (0.3198), therefore the proportion of the variance of the dependent variable explained by the considered model is 32%.

- 12) Influence on life: we can see positive trend: an increase in influence on life by 1 point entails an increase of happiness by 0.07 points.

```
model_13 <- lm(WVS_Bolivia$happiness ~ WVS_Bolivia$sex + WVS_Bolivia$age
+ WVS_Bolivia$income + WVS_Bolivia$financial_status
+ WVS_Bolivia$friends_imp + WVS_Bolivia$family_trust + WVS_Bolivia$family_status
+ WVS_Bolivia$children + WVS_Bolivia$children_1 + WVS_Bolivia$religion_imp
+ WVS_Bolivia$religion_participation + WVS_Bolivia$religion_VS_science
+ WVS_Bolivia$health_status + WVS_Bolivia$influence_on_life)
summary(model_13)
```

```
##
## Call:
## lm(formula = WVS_Bolivia$happiness ~ WVS_Bolivia$sex + WVS_Bolivia$age +
##   WVS_Bolivia$income + WVS_Bolivia$financial_status + WVS_Bolivia$friends_imp +
##   WVS_Bolivia$family_trust + WVS_Bolivia$family_status + WVS_Bolivia$children +
##   WVS_Bolivia$children_1 + WVS_Bolivia$religion_imp + WVS_Bolivia$religion_participation +
##   WVS_Bolivia$religion_VS_science + WVS_Bolivia$health_status +
##   WVS_Bolivia$influence_on_life)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.05048 -0.36149  0.02056  0.38197  1.89030
##
## Coefficients:
```

```

##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)      2.044644   0.104365  19.591 < 2e-16
## WVS_Bolivia$sex2 -0.025794   0.026773  -0.963 0.335450
## WVS_Bolivia$age  -0.003695   0.001182  -3.125 0.001807
## WVS_Bolivia$income 0.018536   0.006896   2.688 0.007250
## WVS_Bolivia$financial_status 0.100334   0.006858  14.630 < 2e-16
## WVS_Bolivia$friends_imp2 -0.047687   0.037080  -1.286 0.198588
## WVS_Bolivia$friends_imp3 -0.069103   0.034189  -2.021 0.043398
## WVS_Bolivia$friends_imp4 -0.088727   0.057979  -1.530 0.126102
## WVS_Bolivia$family_trustTrust 0.124725   0.034790   3.585 0.000346
## WVS_Bolivia$family_statusSeparated -0.158374   0.052438  -3.020 0.002560
## WVS_Bolivia$family_statusSingle -0.114426   0.046510  -2.460 0.013975
## WVS_Bolivia$family_statusWindowed -0.116901   0.065231  -1.792 0.073279
## WVS_Bolivia$children -0.001806   0.016592  -0.109 0.913317
## WVS_Bolivia$children_11 0.056479   0.052003   1.086 0.277587
## WVS_Bolivia$children_12 -0.028099   0.064037  -0.439 0.660860
## WVS_Bolivia$children_13 -0.057054   0.076556  -0.745 0.456207
## WVS_Bolivia$children_14 -0.009164   0.103221  -0.089 0.929263
## WVS_Bolivia$religion_impImportant 0.044696   0.038083   1.174 0.240684
## WVS_Bolivia$religion_participation1 -0.040052   0.034215  -1.171 0.241907
## WVS_Bolivia$religion_participation2 0.009591   0.031773   0.302 0.762804
## WVS_Bolivia$religion_VS_scienceDisagree -0.068810   0.029358  -2.344 0.019191
## WVS_Bolivia$health_statusNot so good -0.242184   0.027536  -8.795 < 2e-16
## WVS_Bolivia$influence_on_life 0.072592   0.006728  10.789 < 2e-16
##
## (Intercept)      ***
## WVS_Bolivia$sex2
## WVS_Bolivia$age      **
## WVS_Bolivia$income    **
## WVS_Bolivia$financial_status    ***
## WVS_Bolivia$friends_imp2
## WVS_Bolivia$friends_imp3      *
## WVS_Bolivia$friends_imp4
## WVS_Bolivia$family_trustTrust    ***
## WVS_Bolivia$family_statusSeparated    **
## WVS_Bolivia$family_statusSingle      *
## WVS_Bolivia$family_statusWindowed      .
## WVS_Bolivia$children
## WVS_Bolivia$children_11
## WVS_Bolivia$children_12
## WVS_Bolivia$children_13
## WVS_Bolivia$children_14
## WVS_Bolivia$religion_impImportant
## WVS_Bolivia$religion_participation1
## WVS_Bolivia$religion_participation2
## WVS_Bolivia$religion_VS_scienceDisagree *
## WVS_Bolivia$health_statusNot so good    ***
## WVS_Bolivia$influence_on_life          ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5562 on 1868 degrees of freedom
## Multiple R-squared:  0.3277, Adjusted R-squared:  0.3198
## F-statistic: 41.39 on 22 and 1868 DF, p-value: < 2.2e-16

```

According to anova, model_13 is better than model_12 (RSS is lower for model_13: 577.91 < 613.92).

```
anova(model_12, model_13)
```

```
## Analysis of Variance Table
##
## Model 1: WVS_Bolivia$happiness ~ WVS_Bolivia$sex + WVS_Bolivia$age + WVS_Bolivia$income +
##   WVS_Bolivia$financial_status + WVS_Bolivia$friends_imp +
##   WVS_Bolivia$family_trust + WVS_Bolivia$family_status + WVS_Bolivia$children +
##   WVS_Bolivia$children_1 + WVS_Bolivia$religion_imp + WVS_Bolivia$religion_participation +
##   WVS_Bolivia$religion_VS_science + WVS_Bolivia$health_status
## Model 2: WVS_Bolivia$happiness ~ WVS_Bolivia$sex + WVS_Bolivia$age + WVS_Bolivia$income +
##   WVS_Bolivia$financial_status + WVS_Bolivia$friends_imp +
##   WVS_Bolivia$family_trust + WVS_Bolivia$family_status + WVS_Bolivia$children +
##   WVS_Bolivia$children_1 + WVS_Bolivia$religion_imp + WVS_Bolivia$religion_participation +
##   WVS_Bolivia$religion_VS_science + WVS_Bolivia$health_status +
##   WVS_Bolivia$influence_on_life
##   Res.Df    RSS Df Sum of Sq    F    Pr(>F)
## 1    1869 613.92
## 2    1868 577.91  1    36.012 116.4 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Final model

In the final model, we include only those predictors that are statistically significant and the addition of which to the considered models brought significant improvements according to anova.

The final model, in addition to control variables, includes 8 predictors: income, financial status, friends importance, family trust, family status, religion importance over science, health status and influence on life.

Final model is statistically significant (p-value: < 2.2e-16). Adjusted R-squared is 0.3189, therefore the proportion of the variance of the dependent variable explained by the considered model is 31,9%.

- 1) **Gender:** is not a statistically significant predictor, but we left it in the model as a control variable.
- 2) **Age:** negative trend - an increase in age by 1 year entails an decrease of happiness by 0.004 points.
- 3) **Income:** positive trend - an increase in income by 1 point entails an increase of happiness by 0.18 points.
- 4) **Financial status:** positive trend - an increase in financial status by 1 point entails an increase of happiness by 0.1 points.
- 5) **Friends importance:** respondents for whom friends are not very important, compared to those for whom friends are very important, have the level of happiness lower by 0.07.
- 6) **Family trust:** respondents who trust to family, compared to those who do not trust, have the level of happiness higher by 0.12.
- 7) **Family status:** respondents who are separated, compared to those who are married, have the level of happiness lower by 0.16. Respondents who are single, compared to those who are married, have the level of happiness lower by 0.11.

- 8) **Religion importance over science:** respondents who disagree with the statement “Whenever science and religion conflict, religion is always right”, compared to those who agree with this statement, have the level of happiness lower by 0.08.
- 9) **Health status:** respondents whose health status is “Not so good”, compared to those whose health status is “Good”, have the level of happiness lower by 0.25.
- 10) **Influence on life:** positive trend - an increase in influence on life by 1 point entails an increase of happiness by 0.07 points.

```
model_final <- lm(WVS_Bolivia$happiness ~ WVS_Bolivia$sex + WVS_Bolivia$age
+ WVS_Bolivia$income + WVS_Bolivia$financial_status + WVS_Bolivia$friends_imp
+ WVS_Bolivia$family_trust + WVS_Bolivia$family_status + WVS_Bolivia$religion_VS_scien
+ WVS_Bolivia$health_status + WVS_Bolivia$influence_on_life)
summary(model_final)
```

```
##
## Call:
## lm(formula = WVS_Bolivia$happiness ~ WVS_Bolivia$sex + WVS_Bolivia$age +
##      WVS_Bolivia$income + WVS_Bolivia$financial_status + WVS_Bolivia$friends_imp +
##      WVS_Bolivia$family_trust + WVS_Bolivia$family_status + WVS_Bolivia$religion_VS_science +
##      WVS_Bolivia$health_status + WVS_Bolivia$influence_on_life)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.06245 -0.36806  0.01969  0.37775  1.91481
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      2.087640   0.091424  22.835 < 2e-16
## WVS_Bolivia$sex2 -0.019556   0.026384  -0.741 0.458663
## WVS_Bolivia$age  -0.004059   0.001036  -3.916 9.32e-05
## WVS_Bolivia$income  0.017915   0.006863   2.610 0.009119
## WVS_Bolivia$financial_status  0.099799   0.006851  14.567 < 2e-16
## WVS_Bolivia$friends_imp2 -0.048317   0.036990  -1.306 0.191643
## WVS_Bolivia$friends_imp3 -0.072846   0.034095  -2.137 0.032763
## WVS_Bolivia$friends_imp4 -0.097228   0.057552  -1.689 0.091312
## WVS_Bolivia$family_trustTrust  0.122576   0.034711   3.531 0.000423
## WVS_Bolivia$family_statusSeparated -0.162032   0.052238  -3.102 0.001952
## WVS_Bolivia$family_statusSingle -0.105668   0.032797  -3.222 0.001295
## WVS_Bolivia$family_statusWindowed -0.115327   0.064577  -1.786 0.074280
## WVS_Bolivia$religion_VS_scienceDisagree -0.079402   0.028142  -2.821 0.004830
## WVS_Bolivia$health_statusNot so good -0.245675   0.027476  -8.941 < 2e-16
## WVS_Bolivia$influence_on_life  0.072892   0.006714  10.857 < 2e-16
##
## (Intercept)      ***
## WVS_Bolivia$sex2
## WVS_Bolivia$age      ***
## WVS_Bolivia$income    **
## WVS_Bolivia$financial_status  ***
## WVS_Bolivia$friends_imp2
## WVS_Bolivia$friends_imp3      *
## WVS_Bolivia$friends_imp4      .
## WVS_Bolivia$family_trustTrust  ***
```

```
## WVS_Bolivia$family_statusSeparated      **
## WVS_Bolivia$family_statusSingle         **
## WVS_Bolivia$family_statusWindowed       .
## WVS_Bolivia$religion_VS_scienceDisagree **
## WVS_Bolivia$health_statusNot so good    ***
## WVS_Bolivia$influence_on_life           ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5566 on 1876 degrees of freedom
## Multiple R-squared:  0.3239, Adjusted R-squared:  0.3189
## F-statistic: 64.2 on 14 and 1876 DF, p-value: < 2.2e-16
```

To compare predictor weights, let's look at standardized coefficients (beta coefficients). The most important predictors are financial_status (0.31), influence_on_life (0.22) and health_status (category "Not so good": -0.18).

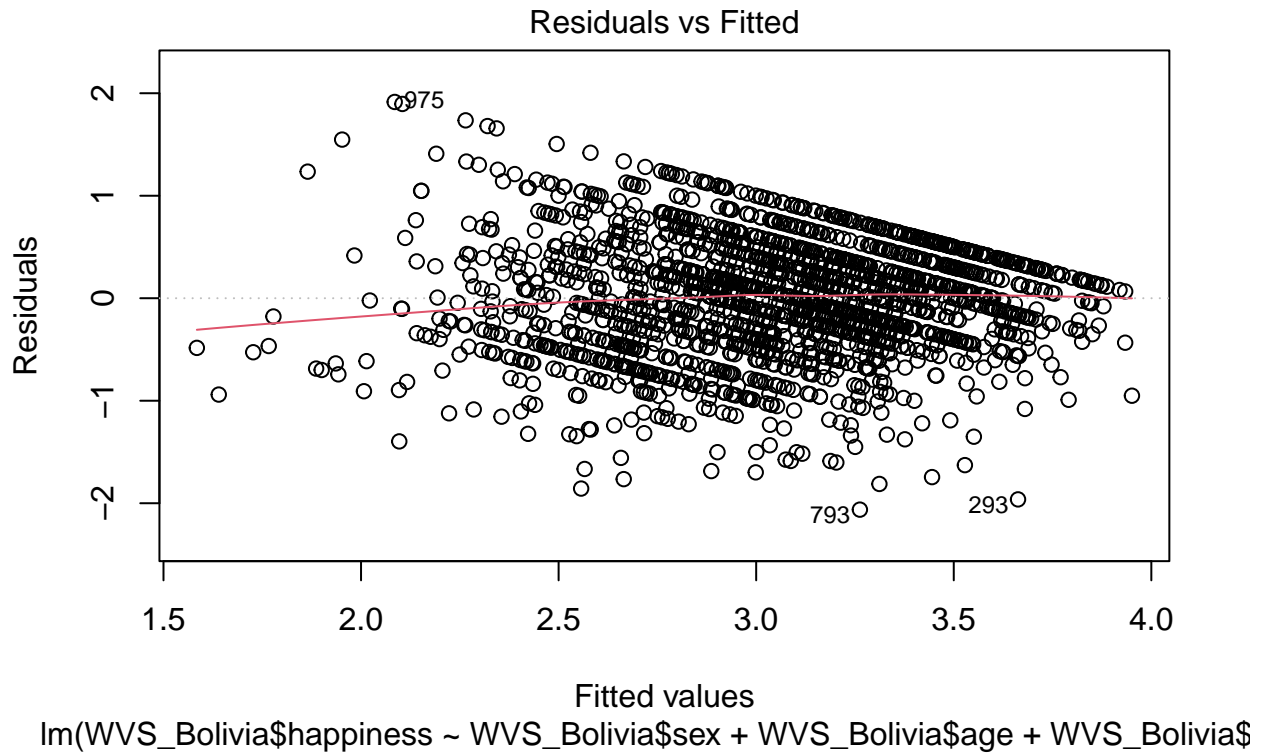
```
lm.beta::lm.beta(model_final)
```

```
##
## Call:
## lm(formula = WVS_Bolivia$happiness ~ WVS_Bolivia$sex + WVS_Bolivia$age +
##     WVS_Bolivia$income + WVS_Bolivia$financial_status + WVS_Bolivia$friends_imp +
##     WVS_Bolivia$family_trust + WVS_Bolivia$family_status + WVS_Bolivia$religion_VS_science +
##     WVS_Bolivia$health_status + WVS_Bolivia$influence_on_life)
##
## Standardized Coefficients::
##              (Intercept)              WVS_Bolivia$sex2
##                   NA                   -0.01450181
##              WVS_Bolivia$age              WVS_Bolivia$income
##             -0.09477200              0.05357682
##      WVS_Bolivia$financial_status      WVS_Bolivia$friends_imp2
##             0.30565571             -0.03233275
##      WVS_Bolivia$friends_imp3      WVS_Bolivia$friends_imp4
##             -0.05361036             -0.03582993
##      WVS_Bolivia$family_trustTrust      WVS_Bolivia$family_statusSeparated
##             0.06858778             -0.06102317
##      WVS_Bolivia$family_statusSingle      WVS_Bolivia$family_statusWindowed
##             -0.07237467             -0.03679959
## WVS_Bolivia$religion_VS_scienceDisagree      WVS_Bolivia$health_statusNot so good
##             -0.05428005             -0.18216238
##      WVS_Bolivia$influence_on_life
##             0.21772598
```

2. Model diagnostics. Diagnostic plots' description (plot(model))

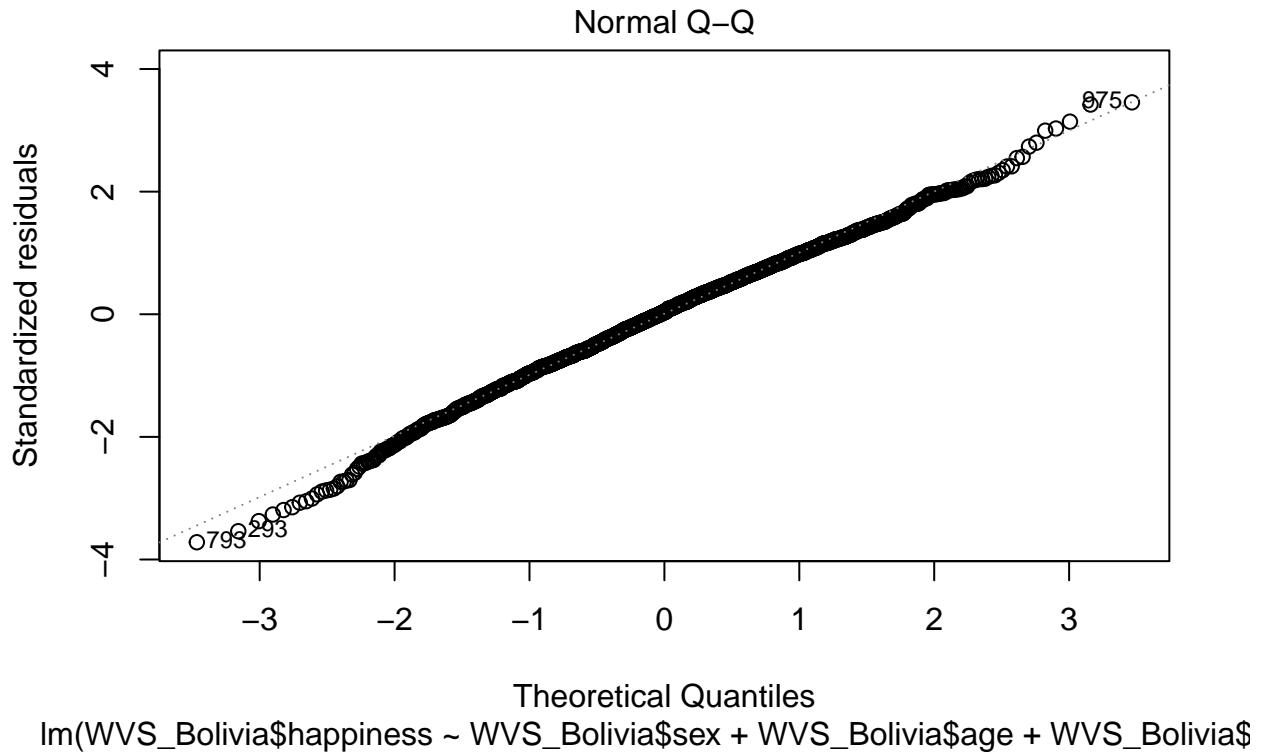
1. Residuals Vs Fitted – red line is close to be horizontal, which indicates a close-to-linear relationship

```
plot(model_final, which = 1)
```



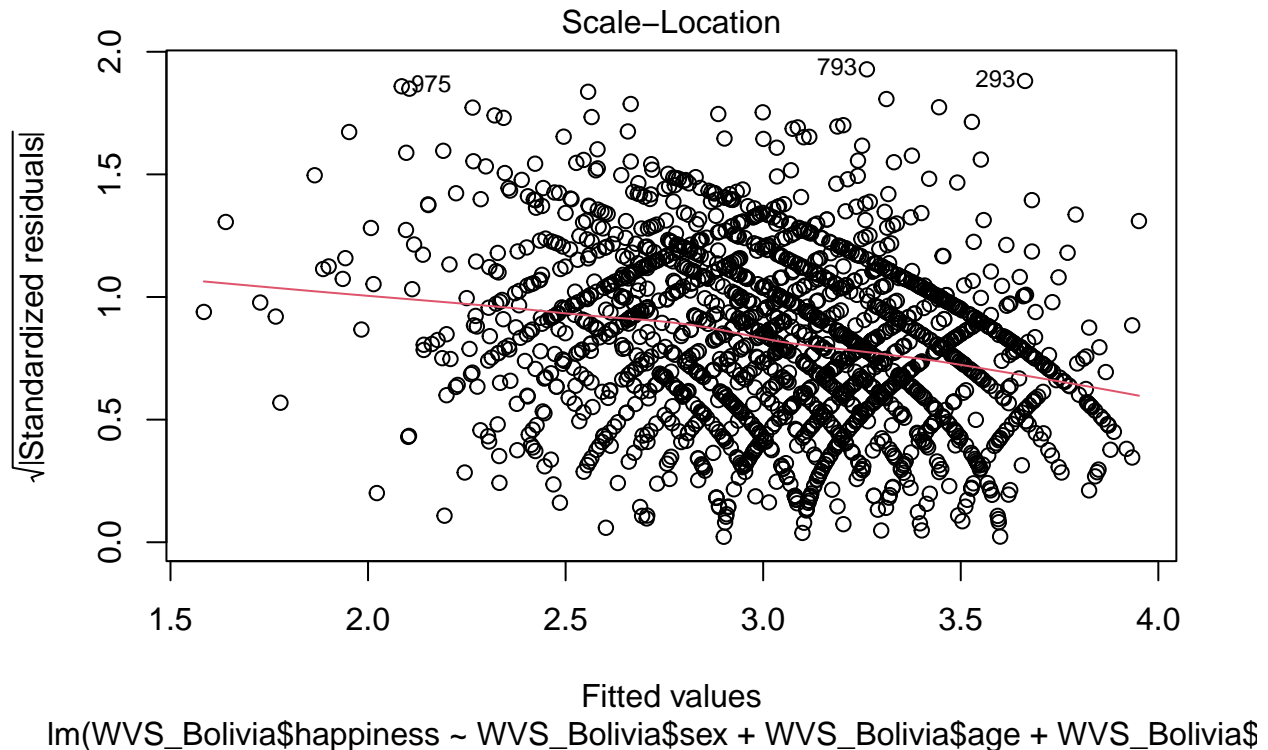
2. Normal QQ – it's almost straight line, so the data is close to a normal distribution.

```
plot(model_final, which = 2)
```



3. Scale-Location (Heteroscedasticity) – the residuals are not randomly located. Also, `ncvTest` is significant ($p = < 2.22e-16$), that is a sign of heteroscedasticity.

```
plot(model_final, which = 3)
```



```
ncvTest(model_final)
```

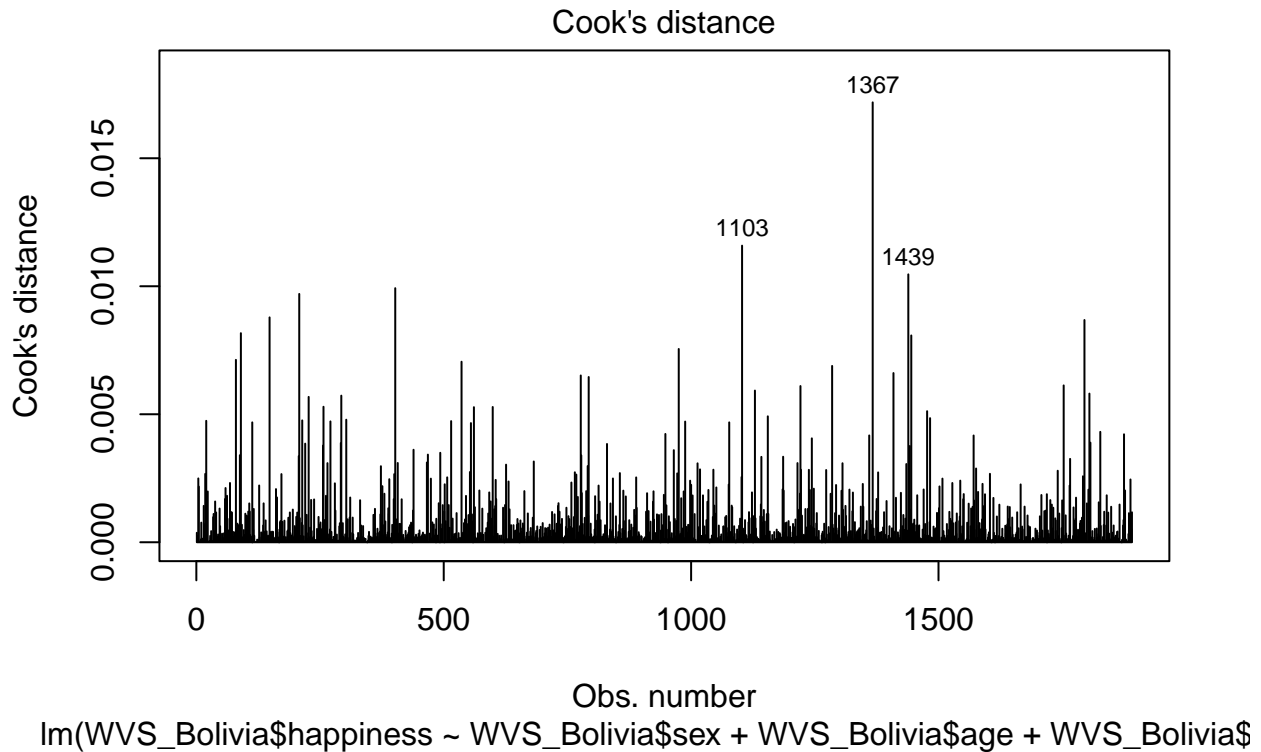
```
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 113.5902, Df = 1, p = < 2.22e-16
```

4. Residuals Vs Leverage (Outliers and leverages). The null for Bonferonni test is that the observation is not an outlier. In this model we have one case that could be outliers and Bonferonni p-value shows that this case is not outlier: Bonferonni p-value is higher than 0.05 (0.371). Therefore, we accept the null that there are no outliers in the model. Well, also on the plot we do not see the red dotted line that would delimit the influential outliers.

```
outlierTest(model_final)
```

```
## No Studentized residuals with Bonferroni p < 0.05
## Largest |rstudent|:
##      rstudent unadjusted p-value Bonferroni p
## 793 -3.731268      0.00019619      0.371
```

```
plot(model_final, which = 4)
```



5. Multicollinearity test

Variance inflation factors are less than 5 for all predictors, therefore there is no multicollinearity in model_final.

```
vif(model_final)
```

##		GVIF	Df	GVIF^(1/(2*Df))
##	WVS_Bolivia\$sex	1.062203	1	1.030632
##	WVS_Bolivia\$age	1.625073	1	1.274784
##	WVS_Bolivia\$income	1.169007	1	1.081206
##	WVS_Bolivia\$financial_status	1.221746	1	1.105326
##	WVS_Bolivia\$friends_imp	1.059100	3	1.009616
##	WVS_Bolivia\$family_trust	1.046769	1	1.023117
##	WVS_Bolivia\$family_status	1.609388	3	1.082539
##	WVS_Bolivia\$religion_VS_science	1.026976	1	1.013398
##	WVS_Bolivia\$health_status	1.151749	1	1.073196
##	WVS_Bolivia\$influence_on_life	1.115987	1	1.056403

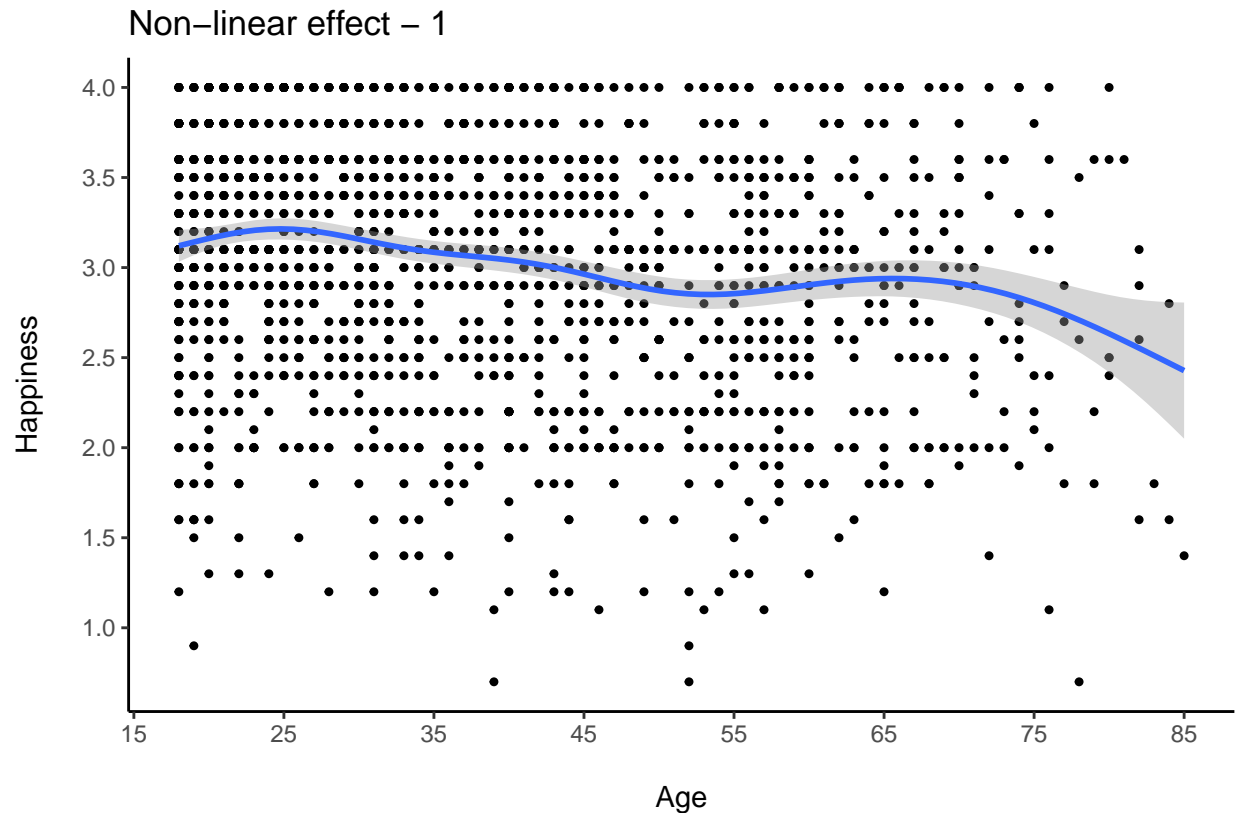
3. Linearity check

To construct a non-linear relationship between variables factor of age as a predictor was taken. This model is statistically significant (p-value < 0.05). We can observe that the level of happiness among respondents aged 18 to 25 rises, then begins to gradually decrease until the age of 55, then slightly increases until the age of 70, and then drops quite sharply.

```
library(gam)
model_non_linear_1 <- gam(WVS_Bolivia$happiness ~ s(WVS_Bolivia$age))
summary(model_non_linear_1)
```

```
##
## Call: gam(formula = WVS_Bolivia$happiness ~ s(WVS_Bolivia$age))
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.34584 -0.39390  0.06709  0.48098  1.28524
##
## (Dispersion Parameter for gaussian family taken to be 0.4379)
##
##      Null Deviance: 859.612 on 1890 degrees of freedom
## Residual Deviance: 825.8384 on 1886 degrees of freedom
## AIC: 3811.804
##
## Number of Local Scoring Iterations: NA
##
## Anova for Parametric Effects
##              Df Sum Sq Mean Sq F value    Pr(>F)
## s(WVS_Bolivia$age)    1  27.81  27.8069   63.504 2.745e-15 ***
## Residuals          1886  825.84   0.4379
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Anova for Nonparametric Effects
##              Npar Df Npar F      Pr(F)
## (Intercept)
## s(WVS_Bolivia$age)      3  4.542 0.003535 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
ggplot(WVS_Bolivia, aes(age, happiness)) +
  geom_point(size = 0.9) +
  labs(x = "\nAge\n",
       y = "Happiness\n") +
  ggtitle("Non-linear effect - 1") +
  scale_x_continuous(breaks = seq(15, 95, 10)) +
  scale_y_continuous(breaks = seq(1, 4, 0.5)) +
  theme_classic()+
  stat_smooth(method = "gam", formula = y ~ s(x))
```



Another non-linear relationship was seen when taking influence on life as a predictor of happiness. This model is statistically significant ($p\text{-value} < 0.05$). It can be seen that for respondents who have little influence on their lives, the level of happiness is low and remains approximately the same in the range of 1-4 predictor units. Then, as the respondent's influence on life increases, happiness also increases in the range from 4 to 8 units of the predictor. Then the level of happiness stabilizes and almost does not change in the range from 8 to 10 units of influence on life.

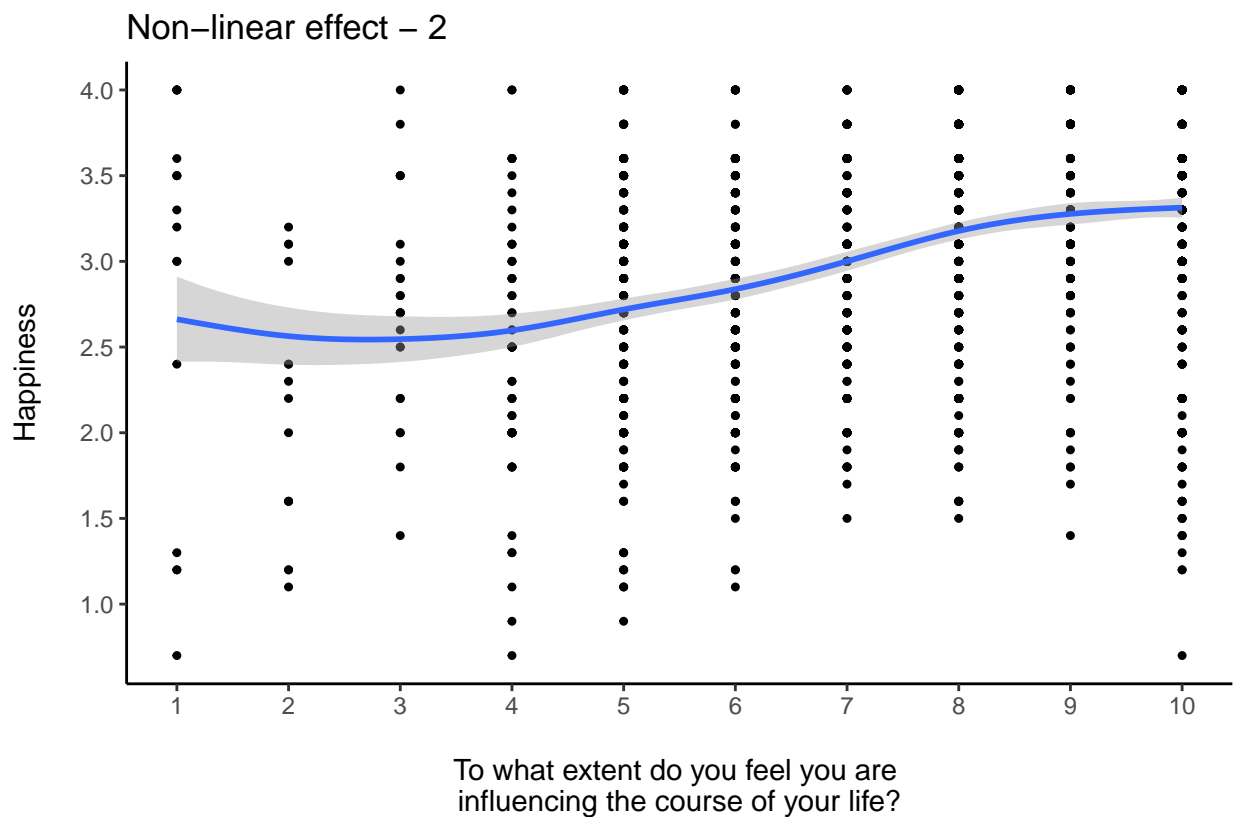
```
library(gam)
model_non_linear_2 <- gam(WVS_Bolivia$happiness ~ s(WVS_Bolivia$influence_on_life))
summary(model_non_linear_2)
```

```
##
## Call: gam(formula = WVS_Bolivia$happiness ~ s(WVS_Bolivia$influence_on_life))
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.6163 -0.3726  0.0966  0.4837  1.4430
##
## (Dispersion Parameter for gaussian family taken to be 0.3974)
##
##      Null Deviance: 859.612 on 1890 degrees of freedom
## Residual Deviance: 749.5646 on 1886 degrees of freedom
## AIC: 3628.553
##
## Number of Local Scoring Iterations: NA
##
## Anova for Parametric Effects
```



```
##                                Df Sum Sq Mean Sq F value    Pr(>F)
## s(WVS_Bolivia$influence_on_life)      1 101.65 101.647   255.76 < 2.2e-16 ***
## Residuals                        1886 749.56    0.397
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Anova for Nonparametric Effects
##                                Npar Df Npar F      Pr(F)
## (Intercept)
## s(WVS_Bolivia$influence_on_life)          3 7.0456 0.0001041 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
ggplot(WVS_Bolivia, aes(influence_on_life, happiness)) +
  geom_point(size = 0.9) +
  labs(x = "\nTo what extent do you feel you are \ninfluencing the course of your life?\n",
       y = "Happiness\n") +
  ggtitle("Non-linear effect - 2") +
  scale_x_continuous(breaks = seq(1, 10, 1)) +
  scale_y_continuous(breaks = seq(1, 4, 0.5)) +
  theme_classic()+
  stat_smooth(method = "gam", formula = y ~ s(x))
```



4. Interaction effect

Interaction effect between religion importance and family status was created and we got the following result: religion importance as a predictor by itself is not statistically significant (p-value > 0.05, p-value = 0.21). But in interaction with family status, it shows that single respondents for whom religion is important are happier by 0.2 than single respondents for whom religion is not important.

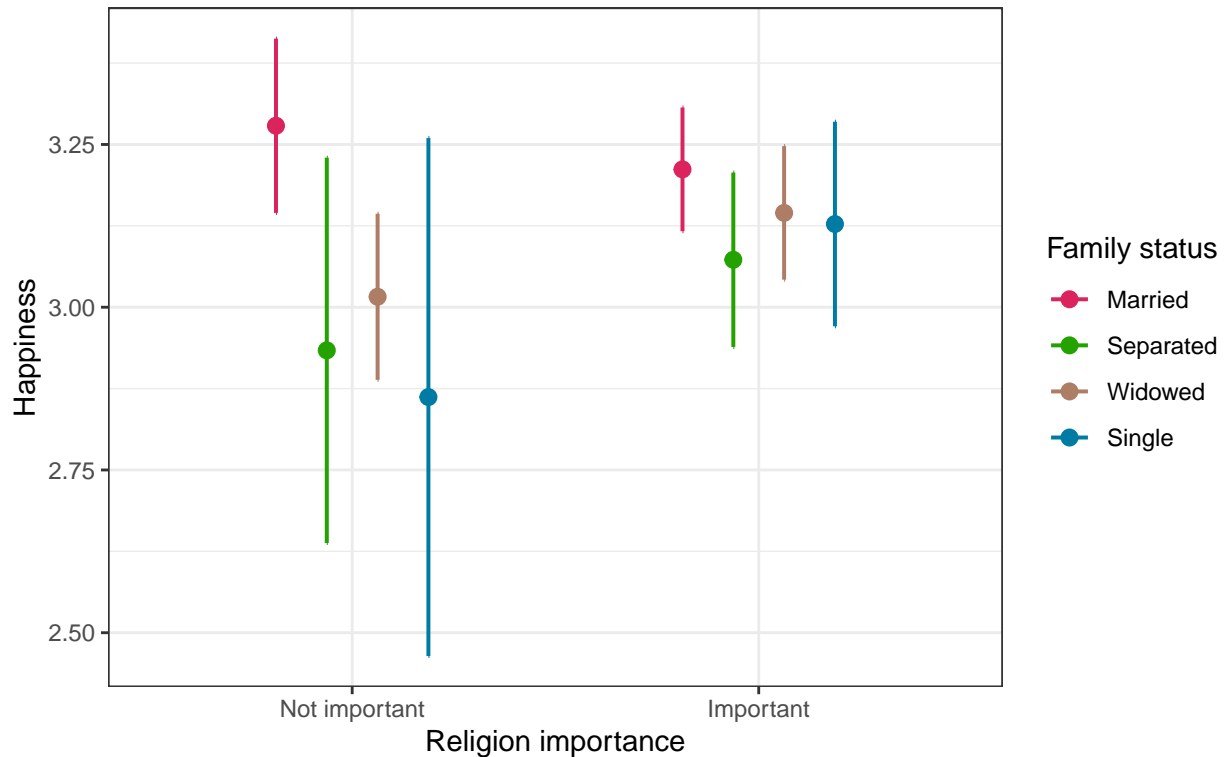
```
model_interaction <- lm(happiness ~ age + sex + income + financial_status + friends_imp
                        + family_trust + religion_imp*family_status + religion_VS_science
                        + health_status + influence_on_life, WVS_Bolivia)
summary(model_interaction)
```

```
##
## Call:
## lm(formula = happiness ~ age + sex + income + financial_status +
##     friends_imp + family_trust + religion_imp * family_status +
##     religion_VS_science + health_status + influence_on_life,
##     data = WVS_Bolivia)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.04931 -0.36112  0.02321  0.37968  1.92934
##
## Coefficients:
##                                Estimate Std. Error t value
## (Intercept)                   2.150951   0.102617  20.961
## age                          -0.004037   0.001037  -3.891
## sex2                         -0.021002   0.026549  -0.791
## income                        0.017370   0.006856   2.534
## financial_status              0.099918   0.006844  14.600
## friends_imp2                 -0.047053   0.036945  -1.274
## friends_imp3                 -0.075089   0.034124  -2.200
## friends_imp4                 -0.104426   0.057706  -1.810
## family_trustTrust            0.116940   0.034756   3.365
## religion_impImportant        -0.067002   0.053841  -1.244
## family_statusSeparated       -0.345015   0.152084  -2.269
## family_statusSingle         -0.262682   0.069432  -3.783
## family_statusWindowed       -0.416771   0.205417  -2.029
## religion_VS_scienceDisagree -0.069131   0.029037  -2.381
## health_statusNot so good    -0.246749   0.027461  -8.985
## influence_on_life            0.073064   0.006705  10.897
## religion_impImportant:family_statusSeparated 0.206171   0.161670   1.275
## religion_impImportant:family_statusSingle   0.195836   0.075067   2.609
## religion_impImportant:family_statusWindowed 0.332824   0.213123   1.562
##
##                                Pr(>|t|)
## (Intercept)                   < 2e-16 ***
## age                          0.000103 ***
## sex2                         0.429005
## income                       0.011373 *
## financial_status              < 2e-16 ***
## friends_imp2                 0.202960
## friends_imp3                 0.027895 *
## friends_imp4                 0.070514 .
## family_trustTrust            0.000782 ***
```

```
## religion_impImportant          0.213491
## family_statusSeparated        0.023407 *
## family_statusSingle          0.000160 ***
## family_statusWidowed         0.042609 *
## religion_VS_scienceDisagree  0.017375 *
## health_statusNot so good      < 2e-16 ***
## influence_on_life             < 2e-16 ***
## religion_impImportant:family_statusSeparated 0.202376
## religion_impImportant:family_statusSingle    0.009158 **
## religion_impImportant:family_statusWidowed   0.118539
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5557 on 1872 degrees of freedom
## Multiple R-squared:  0.3275, Adjusted R-squared:  0.321
## F-statistic: 50.64 on 18 and 1872 DF,  p-value: < 2.2e-16
```

```
plot_model(model_interaction, type = "pred",
            terms = c("religion_imp", "family_status"),
            title = "Predicted values of Happiness: \nInteraction effect - religion importance and famil",
            axis.title = c("Religion importance", "Happiness"),
            dodge = 0.5) +
  theme_bw() +
  scale_x_discrete(limits = c("Not important", "Important")) +
  scale_colour_manual(name = "Family status",
                      values = c("#DB235E", "#24A300", "#AE7D66", "#007BA3"),
                      labels = c("Married", "Separated", "Widowed", "Single"))
```

Predicted values of Happiness:
Interaction effect – religion importance and family status



Meaningful interpretation

The conducted analysis allows both to accept a number of hypotheses, and to reject some of them or clarify them. Thus, despite the assumption, **income** and **financial situation** factors are statistically significant in determining the level of happiness in Bolivia. The higher the income and the better the financial condition, the higher the level of happiness.

However, as expected, **family trust**, **marital status**, **the importance of friendship**, **the importance of religion over science** are significant factors in determining happiness in Bolivia. Family, family trust, friendship allow the respondent to feel happier than the respondent for whom these factors are not important, or who has difficulties with marital status - divorce, widowhood. In addition, another important determinants of happiness were identified - the **state of health** of the respondent and their ability to **influence life**.

Conclusion

To sum up, it cannot be argued that in Bolivia, as one of the poorest countries in Latin America, material factors as happiness predictors are insignificant. Those factors retain an influence on happiness, while the role of non-material factors - family, friends, health, and to a lesser extent - religion is also great. This may be due to the fact that material factors provide an opportunity to improve non-material factors, such as health or the quality of family provision.