

WVS: Level of life satisfaction in Chile, Germany, Ethiopia and Russia

Olga Lavrinenko

2022-12-10

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

Loading the last wave of WVS data

Resource: <https://www.worldvaluessurvey.org/WVSDocumentationWV7.jsp>

Let's load the last wave of WVS data and choose the four countries - Germany, Chile, Ethiopia, and Russia. We will also select only those variables that we need in our work. Then, we change the name of our database and make it shorter.

```
setwd("D:/Data Analytics For Politics And Society/Social and Political Attitudes (1-2 modules)/Project 4")
load("WVS_Cross-National_Wave_7_rData_v4_0.rdata", verbose = F)
WVS_DEU_CHL_ETH_RUS <-
  `WVS_Cross-National_Wave_7_rData_v4_0`[`WVS_Cross-National_Wave_7_rData_v4_0`$B_COUNTRY_ALPHA
    %in% c('DEU', 'CHL', 'ETH', 'RUS'), c("B_COUNTRY_ALPHA", "Q288", "Q49",
                                          "Q287", "Q260", "Q262", "Q173")]

WVS_DEU_CHL_ETH_RUS <- na.omit(WVS_DEU_CHL_ETH_RUS)
```

Renaming variables

Next step is changing names of variables that will be used in this project.

```
names(WVS_DEU_CHL_ETH_RUS)[1:7] <- c("Country", "income", "life_sat",
                                     "social_class", "sex", "age", "religiosity")

sapply(WVS_DEU_CHL_ETH_RUS, FUN = class)
```

```
##      Country      income    life_sat social_class      sex      age
## "character"    "integer"    "integer"    "integer"    "integer"    "integer"
## religiosity
##      "integer"
```

Class of variable

Although R shows an integer (numeric) type for all variables, some of them, according to the questionnaire, are categorical. Variables such as social class, sex, and religiosity need to be changed: let's set the values not by numbers, but by text for a more convenient interpretation. Since "social class" is an ordinal variable, let's prefix new values with digits to keep the order.

```
WVS_DEU_CHL_ETH_RUS$social_class <-
  as.factor(ifelse(WVS_DEU_CHL_ETH_RUS$social_class == "1", "1-Upper class",
    ifelse(WVS_DEU_CHL_ETH_RUS$social_class == "2", "2-Upper middle class",
      ifelse(WVS_DEU_CHL_ETH_RUS$social_class == "3", "3-Lower middle class",
        ifelse(WVS_DEU_CHL_ETH_RUS$social_class == "4", "4-Working class",
          ifelse(WVS_DEU_CHL_ETH_RUS$social_class == "5", "5-Lower class", NA))))))

WVS_DEU_CHL_ETH_RUS$sex <-
  as.factor(ifelse(WVS_DEU_CHL_ETH_RUS$sex == "1", "Male",
    ifelse(WVS_DEU_CHL_ETH_RUS$sex == "2", "Female", NA)))

WVS_DEU_CHL_ETH_RUS$religiosity <-
  as.factor(ifelse(WVS_DEU_CHL_ETH_RUS$religiosity == "1", "A religious person",
    ifelse(WVS_DEU_CHL_ETH_RUS$religiosity == "2", "Not a religious person",
      ifelse(WVS_DEU_CHL_ETH_RUS$religiosity == "3", "An atheist", NA))))

sapply(WVS_DEU_CHL_ETH_RUS, FUN = class)
```

```
##      Country      income    life_sat social_class      sex      age
## "character"    "integer"    "integer"    "factor"    "factor"    "integer"
## religiosity
##      "factor"
```

Databases for each country

And finally let's create a database for each country.

```
WVS_DEU <- WVS_DEU_CHL_ETH_RUS[WVS_DEU_CHL_ETH_RUS$Country == 'DEU', ]
WVS_CHL <- WVS_DEU_CHL_ETH_RUS[WVS_DEU_CHL_ETH_RUS$Country == 'CHL', ]
WVS_ETH <- WVS_DEU_CHL_ETH_RUS[WVS_DEU_CHL_ETH_RUS$Country == 'ETH', ]
WVS_RUS <- WVS_DEU_CHL_ETH_RUS[WVS_DEU_CHL_ETH_RUS$Country == 'RUS', ]
```

4. Descriptive statistics of the level of satisfaction with life for the countries. Are there any differences?

- 1) The **range** of values for all countries is the same: 10 points in the level of life satisfaction, because the maximum value for all countries is 10 and the minimum is 1.

- 2) But the middle of the distribution or the central value (**median**) is different for the countries:
 - for Germany the median is the highest among these countries (value is 8).
 - Russia and Chile have the same middle of the distribution (value is 7), which is between the values of Ethiopia and Germany.
 - for Ethiopia the median is the lowest among these countries (value is 6).
- 3) The difference between the median and the **mean** is not very significant. This means that the distribution for all countries is quite “symmetrical”. The largest of all the differences between the median and mean can be observed for Russia (0.5), so there are some low values in the data that are dragging the mean towards them to the left side in the distribution.
 - the highest mean is for Germany and it is equal to 7.74,
 - then for Chile mean is equal 7.29,
 - for Russia mean is equal 6.53,
 - and the lowest mean is for Ethiopia and it is equal to 5.78.
- 4) Ethiopia has the largest **quartile range** (3rd Qu - 1st Qu) in life satisfaction and it is equal to 4. Whereas Germany has the lowest quartile range in life satisfaction and it is equal to 2. Russia and Chile have the same quartile range which is equal to 3.
 - for Germany 25% of all cases are within 7 points of life satisfaction and 75% of cases are no higher than 9,
 - for Chile 25% of all cases are within 6 points of life satisfaction and 75% of cases are no higher than 9,
 - for Russia 25% of all cases are within 5 points of life satisfaction and 75% of cases are no higher than 8,
 - for Ethiopia 25% of all cases are within 4 points of life satisfaction and 75% of cases are no higher than 8.

```
WVS_DEU_CHL_ETH_RUS %>%
  group_by(Country) %>%
  summarize(min = min(life_sat, na.rm = T),
            "1st Qu." = quantile(life_sat, probs = .25, na.rm = T),
            median = median(life_sat, na.rm = T),
            mean = mean(life_sat, na.rm = T),
            "3rd Qu." = quantile(life_sat, probs = .75, na.rm = T),
            max = max(life_sat, na.rm = T),
            sd = sd(life_sat, na.rm = T))
```

```
## # A tibble: 4 x 8
##   Country   min '1st Qu.' median   mean '3rd Qu.'   max    sd
##   <chr>   <int>   <dbl>   <int> <dbl>   <dbl> <int> <dbl>
## 1 CHL         1         6       7  7.20     9    10  1.98
## 2 DEU         1         7       8  7.74     9    10  1.69
## 3 ETH         1         4       6  5.78     8    10  2.84
## 4 RUS         1         5       7  6.53     8    10  2.09
```

Boxplot of the Level of Life Satisfaction in the countries

Boxplots visualize the interpretation of descriptive statistics indicated above, and also show outliers: for Chile, outliers are at a value of 1. For Germany, values of 1, 2, and 3 of the life satisfaction level are outliers.

Summarizing, we can say that the overall level of life satisfaction is the highest in Germany, then slightly lower in Chile, then in Russia and the lowest overall level of life satisfaction is in Ethiopia.

```

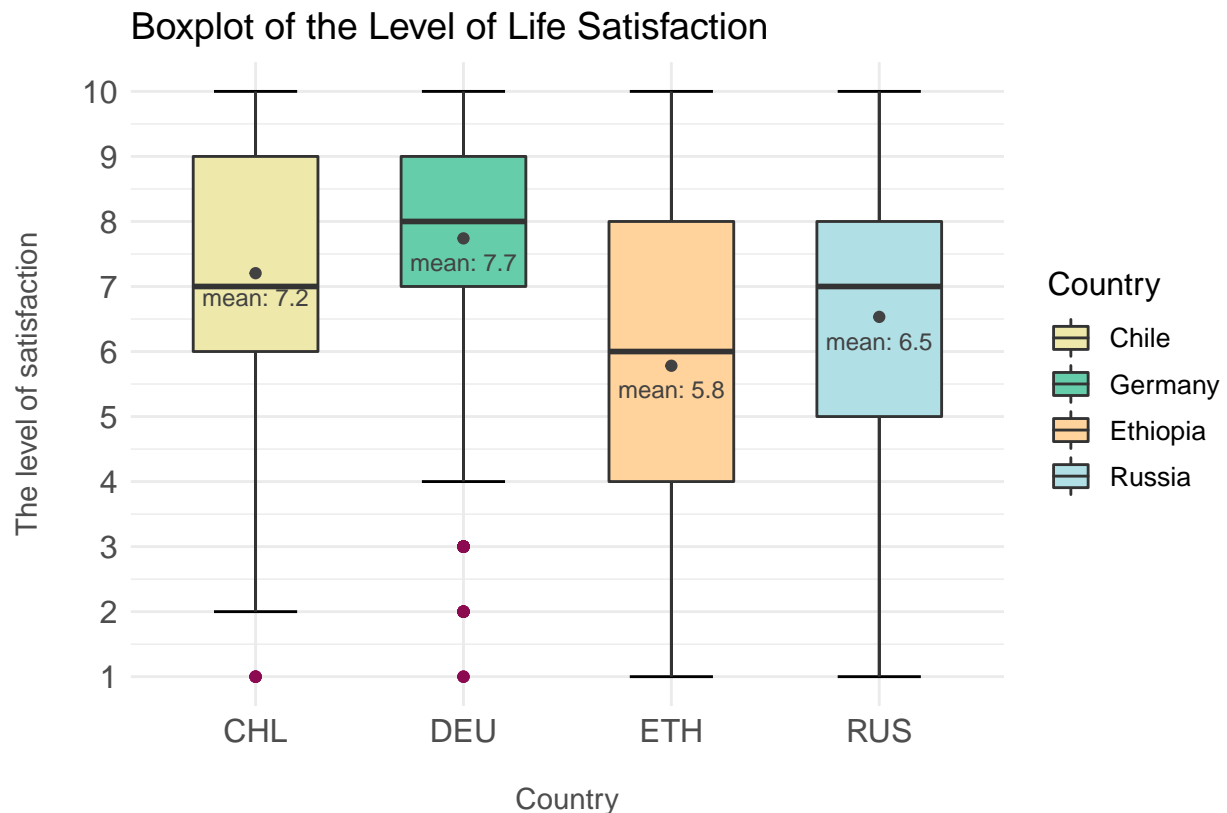
ggplot(WVS_DEU_CHL_ETH_RUS,
      aes(x = Country,
          y = life_sat,
          fill = factor(Country,
                        labels = c("Chile", "Germany", "Ethiopia", "Russia")))) +

stat_boxplot(geom = 'errorbar', width = 0.4) +
geom_boxplot(width = 0.6, outlier.colour = "#8B0A50") +
stat_summary(fun = mean, geom = "point", col = "#424242", show_guide = FALSE) +
stat_summary(fun = mean, geom = "text", col = "#424242", size = 3,
             vjust = 1.9, aes(label = paste("mean:", round(..y.., digits = 1)))) +

scale_y_continuous(breaks = seq(0, 10, 1)) +
scale_fill_manual(values = c("#EEE8AA", "#66CDAA", "#FFD39B", "#BOE0E6")) +
xlab("\nCountry\n") +
ylab("The level of satisfaction\n") +
ggtitle("Boxplot of the Level of Life Satisfaction") +
labs(fill = 'Country') +

theme_minimal()+
theme(text = element_text(size = 12),
      axis.text.x = element_text(size = 12),
      axis.text.y = element_text(size = 12),
      axis.title.x = element_text(size = 11, colour = "#4D4D4D"),
      axis.title.y = element_text(size = 11, colour = "#4D4D4D"))

```



5.1 Descriptive statistics of income for the countries. Are there any differences?

- 1) The **range** of values for all countries is the same: 10 points in the level of income, because the maximum value for all countries is 10 and the minimum is 1.
- 2) The middle of the distribution or the central value (**median**) is the same for all countries as well: value is equal to 5.
- 3) The difference between the median and the **mean** is not very significant. This means that the distribution for all countries is quite “symmetrical”. The largest of all the differences between the median and mean can be observed for Ethiopia (0.6), so there are some low values in the data that are dragging the mean towards them to the left side in the distribution.
 - the highest mean is for Germany and it is equal to 5.20,
 - then for Russia mean is equal 4.79,
 - for Chile mean is equal 4.68,
 - and the lowest mean is for Ethiopia and it is equal to 4.37.
- 4) Ethiopia and Russia have a slightly larger **quartile range** (3rd Qu - 1st Qu) in income and it is equal to 3. Whereas Chile and Germany have quartile range in level of income which is equal to 2.
 - for Chile and Germany and 25% of all cases are within 4 points of income
 - for Ethiopia and Russia 25% of all cases are within 3 points of income
 - for all countries 75% of cases are no higher than 6, so there is no difference in the third quartile among countries.

```
WVS_DEU_CHL_ETH_RUS %>%
  group_by(Country) %>%
  summarize(min = min(income, na.rm = T),
            "1st Qu." = quantile(income, probs = .25, na.rm = T),
            median = median(income, na.rm = T),
            mean = mean(income, na.rm = T),
            "3rd Qu." = quantile(income, probs = .75, na.rm = T),
            max = max(income, na.rm = T),
            sd = sd(income, na.rm = T))
```

```
## # A tibble: 4 x 8
##   Country   min '1st Qu.' median mean '3rd Qu.'   max   sd
##   <chr>   <int>   <dbl>   <int> <dbl>   <dbl> <int> <dbl>
## 1 CHL         1         4         5  4.68         6    10  1.70
## 2 DEU         1         4         5  5.20         6    10  1.68
## 3 ETH         1         3         5  4.37         6    10  2.23
## 4 RUS         1         3         5  4.79         6    10  1.96
```

Boxplot of Income in the countries

Boxplots visualize the interpretation of descriptive statistics indicated above, and also show outliers: for Chile and Germany there are outliers at a value of 10 in level of income.

Summarizing, we can say that the overall level of income is quite similar between countries. By the average level of income, countries can be ordered as follows: first, Germany, second, Russia and Chile, third, Ethiopia.

```

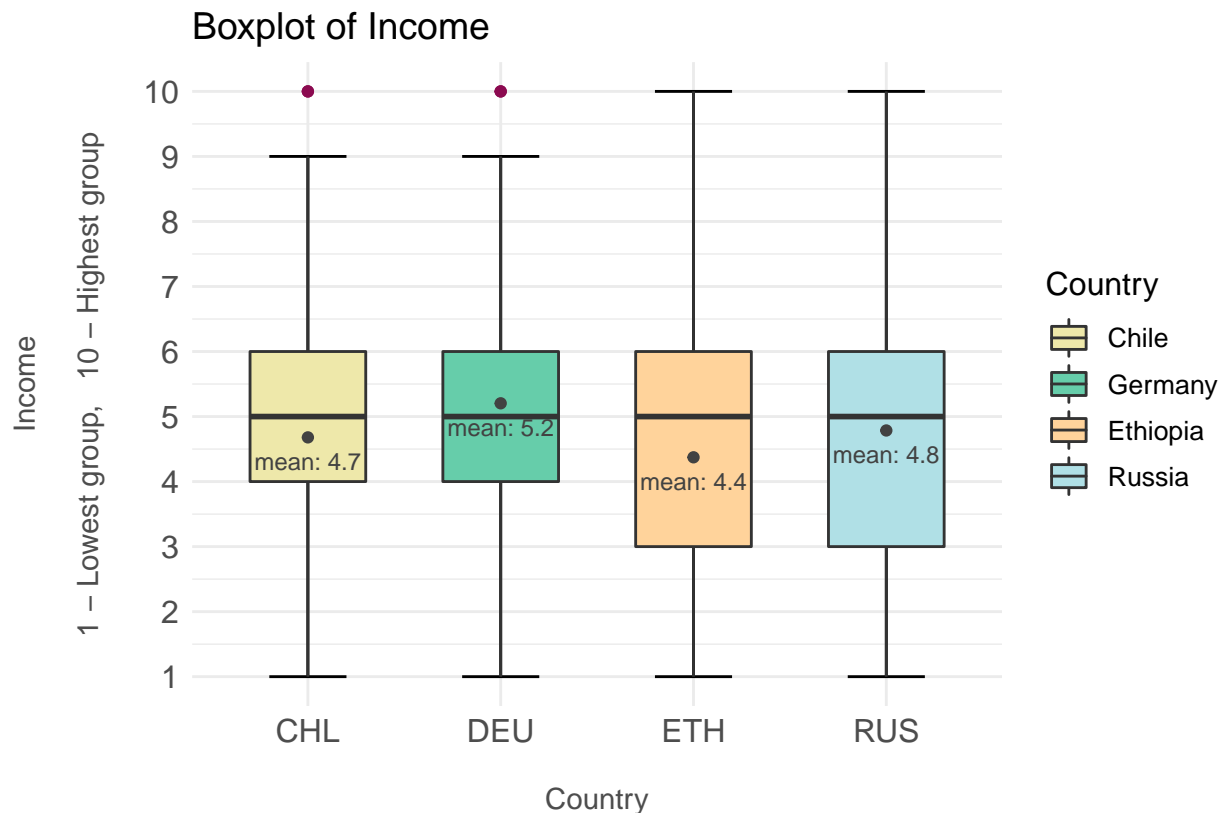
ggplot(WVS_DEU_CHL_ETH_RUS,
      aes(x = Country,
          y = income,
          fill = factor(Country,
                        labels = c("Chile", "Germany", "Ethiopia", "Russia")))) +

stat_boxplot(geom = 'errorbar', width = 0.4) +
geom_boxplot(width = 0.6, outlier.colour = "#8B0A50") +
stat_summary(fun = mean, geom = "point", col = "#424242", show_guide = FALSE) +
stat_summary(fun = mean, geom = "text", col = "#424242", size = 3,
             vjust = 1.9, aes(label = paste("mean:", round(..y.., digits = 1)))) +

scale_y_continuous(breaks = seq(0, 10, 1)) +
scale_fill_manual(values = c("#EEE8AA", "#66CDAA", "#FFD39B", "#B0E0E6")) +
labs(fill = 'Country')+
xlab("\nCountry\n")+
ylab("Income\n\n1 - Lowest group, 10 - Highest group\n") +
ggtitle("Boxplot of Income") +

theme_minimal() +
theme(text = element_text(size=12),
      axis.text.x = element_text(size = 12),
      axis.text.y = element_text(size = 12),
      axis.title.x = element_text(size = 11, colour = "#4D4D4D"),
      axis.title.y = element_text(size = 11, colour = "#4D4D4D"))

```



5.2 Descriptive statistics of subjective social class for the countries. Are there any differences?

Boxplot of Subjective Social Class in the countries

Due to the fact that variable “Social Class” is categorical, we can describe only frequencies and mode.

- 1) **Lower middle class** is the most frequent value for three of the four countries** (from 38% to 43%). For Russia, the most frequent value is the **working class** (38%, whereas for lower middle class the frequency is 35%).
- 2) The least frequent value for all countries is **upper class** (from 1% to 2%).
- 3) The most obvious differences are between the lower class, working class and upper middle class.
 - for Chile, the percentage of people in the **upper middle class** is noticeably lower than in other countries (5%), while in Germany the working class is the the second most frequent after lower middle class (37%). For Ethiopia and Russia the prevalence of the upper middle class is about the same (19%)
 - the prevalence of the **working class** in Germany and Ethiopia is comparatively less (17% - 18%) than in Chile and Russia (38%)
 - the highest prevalence of the **lower class** is observed in Ethiopia (22%), whereas in Chile it is about 12%, even less in Russia (7%) and very little in Germany (2%)

So, we can conclude that the basic social classes in Chile and Russia are lower middle class and working class, in Germany it is middle class (lower middle and upper middle), and in Ethiopia lower middle class is the most frequent but lower class, working class and upper middle class are also significant (22%, 18%, 19%).

```
sosial_class_prop <- round((prop.table(table(WVS_DEU_CHL_ETH_RUS$social_class, WVS_DEU_CHL_ETH_RUS$Country)), 2) * 100)
sosial_class_prop
```

```
##
##           CHL    DEU    ETH    RUS
## 1-Upper class   2.225  1.358  2.211  1.388
## 2-Upper middle class 4.561 37.098 19.328 19.299
## 3-Lower middle class 43.382 42.244 38.165 34.964
## 4-Working class   37.931 17.370 17.854 37.541
## 5-Lower class    11.902  1.930 22.441  6.808
```

```
bp <- barplot(sosial_class_prop,
  beside = T,
  xlab = "Country",
  ylab = "Frequency, %",
  ylim = c(0, 55),
  main = "Barplot of social class",
  col = c("#99DAF7", "#F792C0", "#66CDAA", "#F7BAB7", "#FDF5E6"))

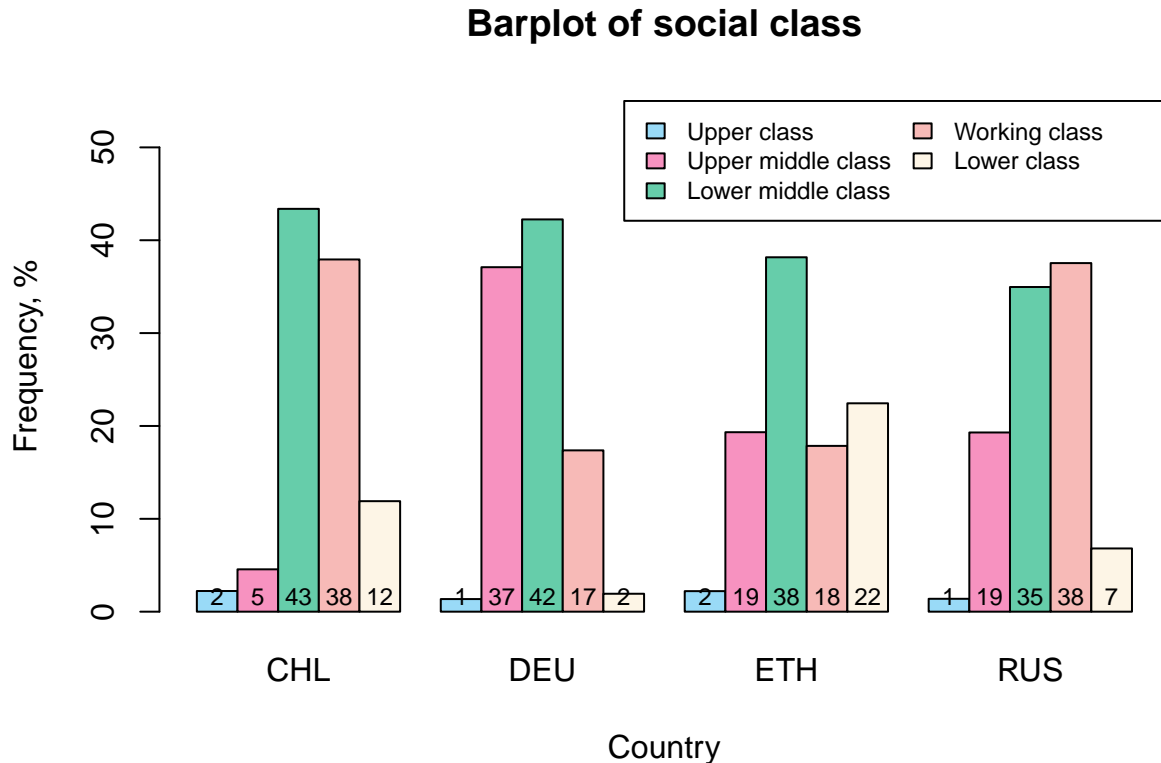
legend("topright",
  legend = c("Upper class",
    "Upper middle class",
    "Lower middle class",
```

```

    "Working class",
    "Lower class"),
  fill = c("#99DAF7", "#F792C0", "#66CDAA", "#F7BAB7", "#FDF5E6"),
  ncol = 2,
  cex = 0.78)

text(bp, -1.4, round(social_class_prop, 0), cex = 0.75, col = "black", pos = 3)

```



6. Create scatterplot with 4 countries (or 4 scatterplots) with a trend line (abline in R) where X is income, Y is satisfaction. Describe the trends you observe.

- For Germany and Russia we can observe a positive trend: the higher the income, the higher the level of life satisfaction.
- For Ethiopia there is a positive level only until the income level reaches 6. Then, as income increases, the level of life satisfaction gradually decreases.
- For Chile there is no association between income and life satisfaction until income reaches level 5. Then, as income increases, the level of life satisfaction drops markedly. But as p-value ($0.98 > 0.05$) shows, this relationship is not statistically significant.

```

ggplot(WVS_DEU_CHL_ETH_RUS,
  aes(x = income,

```



```

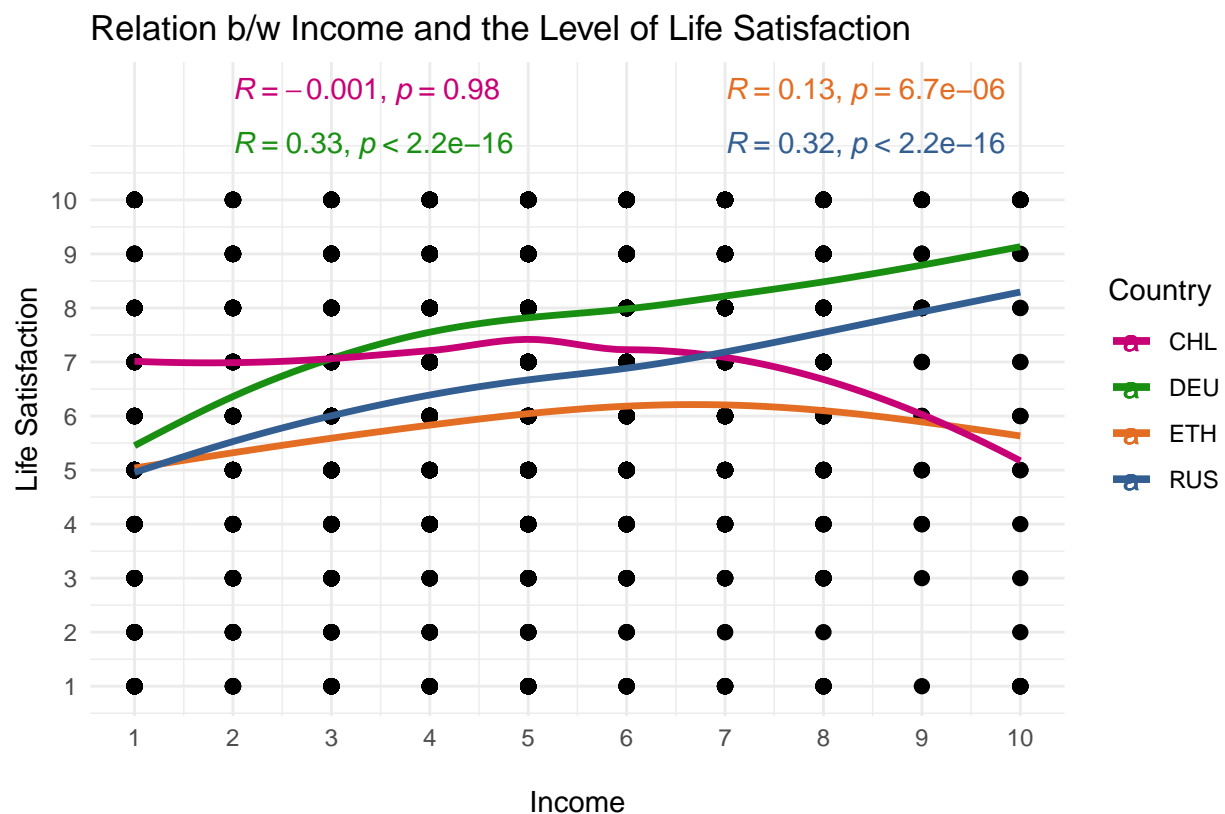
      y = life_sat, col = factor(Country))) +
geom_point(size = 2.2, col = "black") +
labs(x = "\nIncome\n",
     y = "Life Satisfaction",
     col = "Country") +
ggtitle("Relation b/w Income and the Level of Life Satisfaction") +
scale_x_continuous(breaks = seq(0, 10, 1)) +
scale_y_continuous(breaks = seq(0, 10, 1)) +

theme_minimal () +
scale_color_manual(values = c("#C80075", "#198F12", "#E46D24", "#335E92")) +

stat_smooth(data = WVS_DEU, se = F, aes(x = income, y = life_sat), size = 1.2) +
stat_smooth(data = WVS_ETH, se = F, aes(x = income, y = life_sat), size = 1.2) +
stat_smooth(data = WVS_CHL, se = F, aes(x = income, y = life_sat), size = 1.2) +
stat_smooth(data = WVS_RUS, se = F, aes(x = income, y = life_sat), size = 1.2) +

stat_cor(data = WVS_DEU, method = "pearson", label.x = 2, label.y = 11) +
stat_cor(data = WVS_ETH, method = "pearson", label.x = 7, label.y = 12) +
stat_cor(data = WVS_CHL, method = "pearson", label.x = 2, label.y = 12) +
stat_cor(data = WVS_RUS, method = "pearson", label.x = 7, label.y = 11)

```



7. Calculate correlation coefficients (with significance measure) for income and satisfaction.

Germany

Correlation is statistically significant ($p\text{-value} < 2.2\text{e-}16$), positive, and low (0.33): the the higher the income, the higher the level of life satisfaction.

```
cor.test(WVS_DEU$income, WVS_DEU$life_sat)

##
## Pearson's product-moment correlation
##
## data: WVS_DEU$income and WVS_DEU$life_sat
## t = 13.223, df = 1397, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.2861235 0.3793100
## sample estimates:
##          cor
## 0.3335312
```

Chile

Correlation is not statistically significant ($p\text{-value} = 0.976$). We can not establish a correlation between income and level of life satisfaction.

```
cor.test(WVS_CHL$income, WVS_CHL$life_sat)

##
## Pearson's product-moment correlation
##
## data: WVS_CHL$income and WVS_CHL$life_sat
## t = -0.030114, df = 897, p-value = 0.976
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## -0.06638548 0.06438316
## sample estimates:
##          cor
## -0.001005461
```

Ethiopia

Correlation is statistically significant ($p\text{-value} < 2.2\text{e-}16$), positive, and low (0.19): the the higher the income, the higher the level of life satisfaction.

```
cor.test(WVS_ETH$income, WVS_ETH$life_sat)
```

```
##
```

```
## Pearson's product-moment correlation
##
## data: WVS_ETH$income and WVS_ETH$life_sat
## t = 4.5237, df = 1219, p-value = 6.67e-06
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## 0.07291723 0.18327187
## sample estimates:
## cor
## 0.1284923
```

Russia

Correlation is statistically significant (p-value < 2.2e-16), positive, and low (0.32): the the higher the income, the higher the level of life satisfaction.

```
cor.test(WVS_RUS$income, WVS_RUS$life_sat)
```

```
##
## Pearson's product-moment correlation
##
## data: WVS_RUS$income and WVS_RUS$life_sat
## t = 12.978, df = 1511, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## 0.2706026 0.3613087
## sample estimates:
## cor
## 0.3166794
```

For all countries

Correlation is statistically significant (p-value < 2.2e-16), positive, and low (0.24): the the higher the income, the higher the level of life satisfaction.

```
cor.test(WVS_DEU_CHL_ETH_RUS$income, WVS_DEU_CHL_ETH_RUS$life_sat)
```

```
##
## Pearson's product-moment correlation
##
## data: WVS_DEU_CHL_ETH_RUS$income and WVS_DEU_CHL_ETH_RUS$life_sat
## t = 17.203, df = 5030, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## 0.2094600 0.2616534
## sample estimates:
## cor
## 0.2357267
```

8. Create a boxplot with 4 countries (or 4 boxplots) with a trend line (abline in R) where X is social class, Y is satisfaction. Describe the trends you observe.

We can see a difference in the relationship between social class and life satisfaction levels for the countries under consideration:

- for Germany and Russia the trend line is downward: the lower the category of social class, the lower the level of life satisfaction.
- for Ethiopia and Chile the trend line is broken: with a decrease in the categories of social class, the level of life satisfaction can either fall or rise. For instance Chile, category “lower middle class” is below category “upper middle class”, but level of life satisfaction is higher for lower middle class, then for upper middle class. Another example is Ethiopia, for the four categories of social class (upper, upper middle, lower middle and working class) the level of life satisfaction is about the same.

```
ggplot(WVS_DEU_CHL_ETH_RUS,
       aes(x = social_class,
           y = life_sat,
           fill = factor(social_class,
                         labels = c("Upper class", "Upper middle class",
                                   "Lower middle class", "Working class", "Lower class")))) +

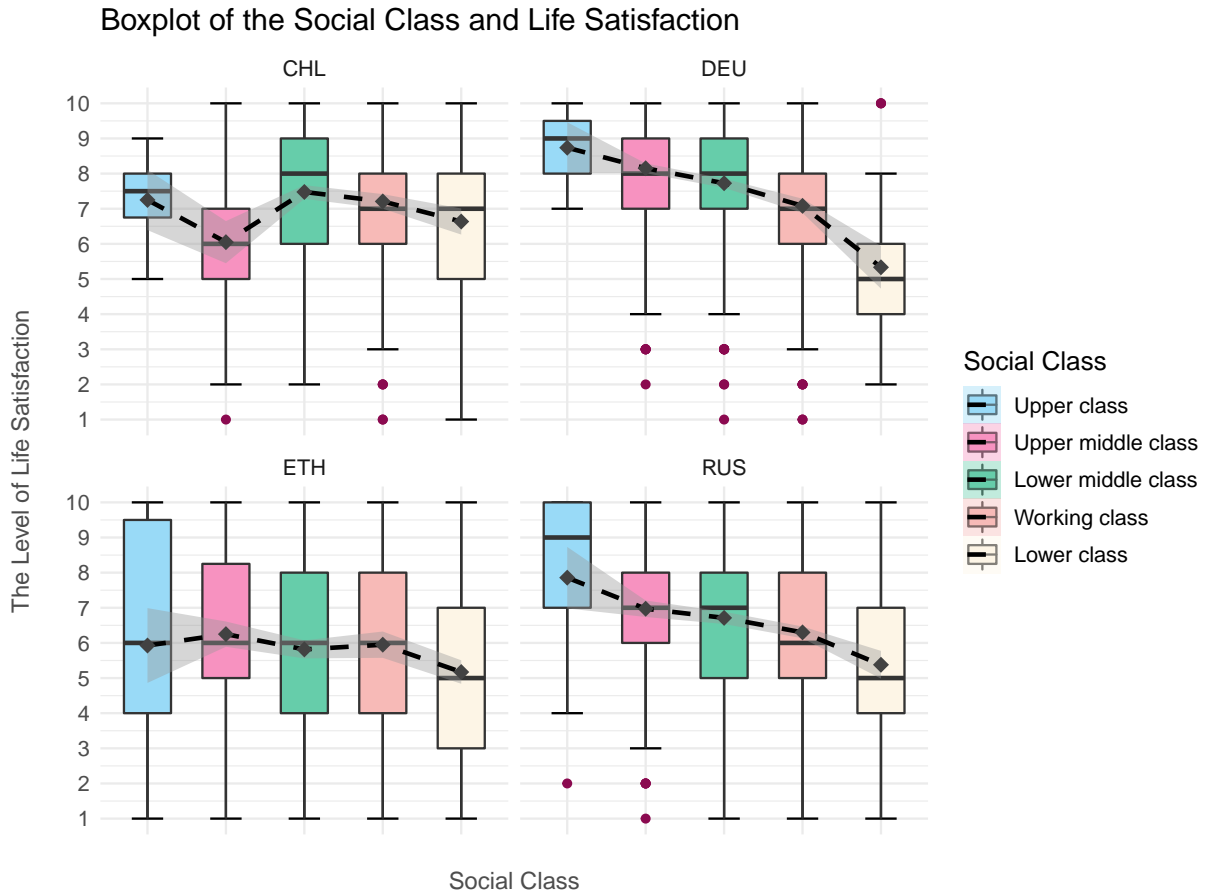
stat_boxplot(geom = 'errorbar', width = 0.4) +
geom_boxplot(width = 0.6, outlier.colour = "#8B0A50") +

scale_y_continuous(breaks = seq(0, 10, 1)) +
scale_fill_manual(values = c("#99DAF7", "#F792C0", "#66CDAA", "#F7BAB7", "#FDF5E6")) +

labs(fill = 'Social Class')+
xlab("\nSocial Class\n")+
ylab("\nThe Level of Life Satisfaction\n")+
ggtitle("Boxplot of the Social Class and Life Satisfaction")+

theme_minimal()+
theme(text = element_text(size=12),
      axis.text.x = element_blank(),
      axis.text.y = element_text(size = 10),
      axis.title.x = element_text(size = 11, colour = "#4D4D4D"),
      axis.title.y = element_text(size = 11, colour = "#4D4D4D")) +

stat_smooth(method = "lm", color = "black",
            lty = 2, aes(group = 1), formula = y ~ poly(x, 5, raw = TRUE))+
stat_summary(fun = mean, size = 0.7, pch = 18, col = "#424242", show_guide = FALSE) +
facet_wrap(Country ~ .)
```



9. Calculate ANOVA (with significance measure) for social class and satisfaction. Describe the trends you observe.

Germany

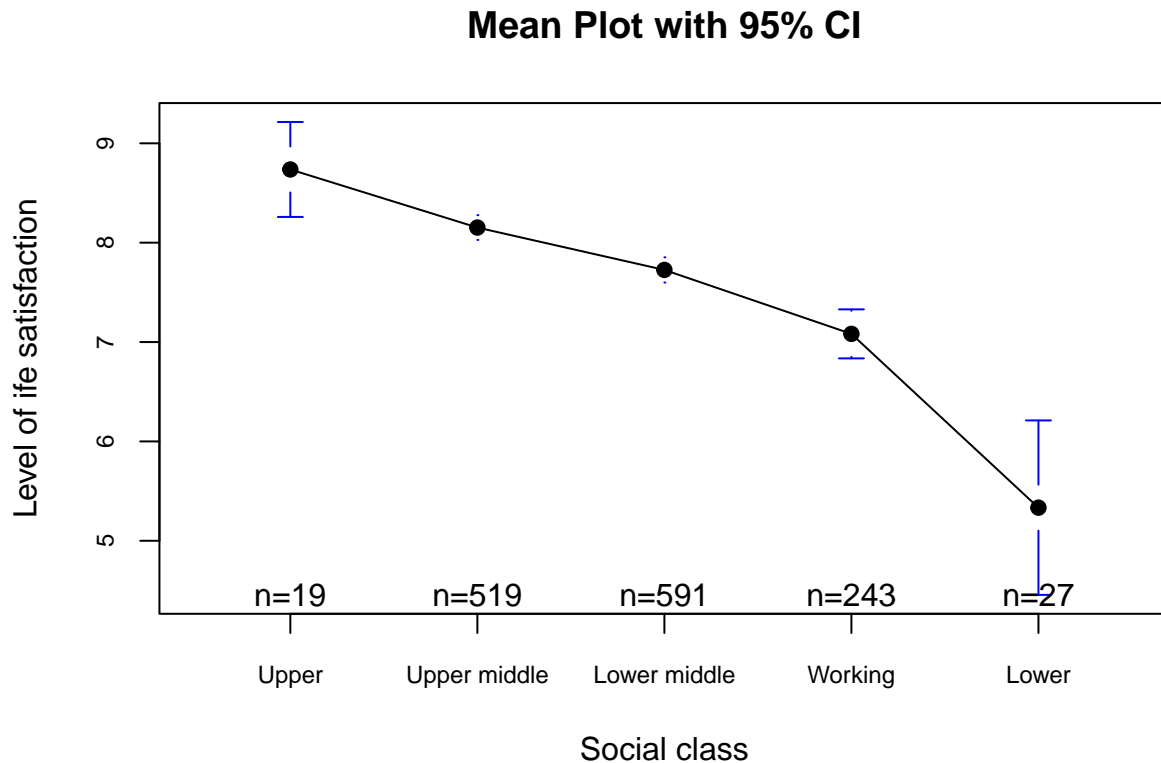
p-value is significant, so we can speak about the association between social class and level of life satisfaction. Upper class is more likely to have a high level of life satisfaction (8.7), while with the increase in social class categories, the level of life satisfaction decreases (8.2 for Upper middle class, 7.7 for Lower middle class, 7.1 for Working class and 5.3 for Lower class).

```
summary(aov(life_sat ~ social_class, data = WVS_DEU))
```

```
##               Df Sum Sq Mean Sq F value Pr(>F)
## social_class    4    369   92.17  35.65 <2e-16 ***
## Residuals  1394   3605    2.59
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
plotmeans(life_sat ~ social_class, data = WVS_DEU,
           xlab = "Social class", ylab = "Level of life satisfaction",
```

```
main="Mean Plot with 95% CI", pch = 19,
legends = c("Upper", "Upper middle", "Lower middle", "Working", "Lower"),
cex.axis = 0.75)
```



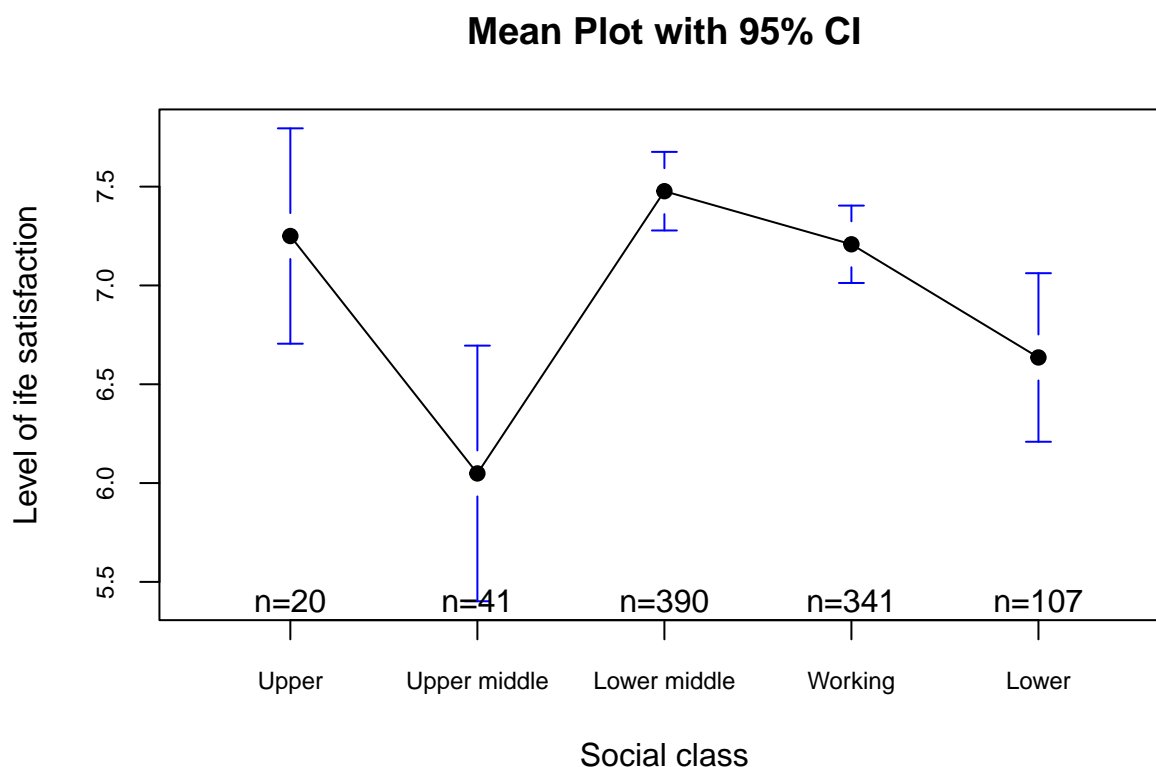
Chile

p-value is significant, so we can speak about the association between social class and level of life satisfaction. But this association is manifested only for such social classes as lower middle class, working class and lower. Wherein the level of life satisfaction of the upper class is on a level with working class (7.2), and upper middle class is likely to have the lowest level of life satisfaction (6.0) among all social classes.

```
summary(aov(life_sat ~ social_class, data = WVS_CHL))
```

```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## social_class  4    118   29.598    7.742 3.91e-06 ***
## Residuals    894   3418    3.823
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
plotmeans(life_sat ~ social_class, data = WVS_CHL,
  xlab = "Social class", ylab = "Level of life satisfaction",
  main="Mean Plot with 95% CI", pch = 19,
  legends = c("Upper", "Upper middle", "Lower middle", "Working", "Lower"),
  cex.axis = 0.75)
```



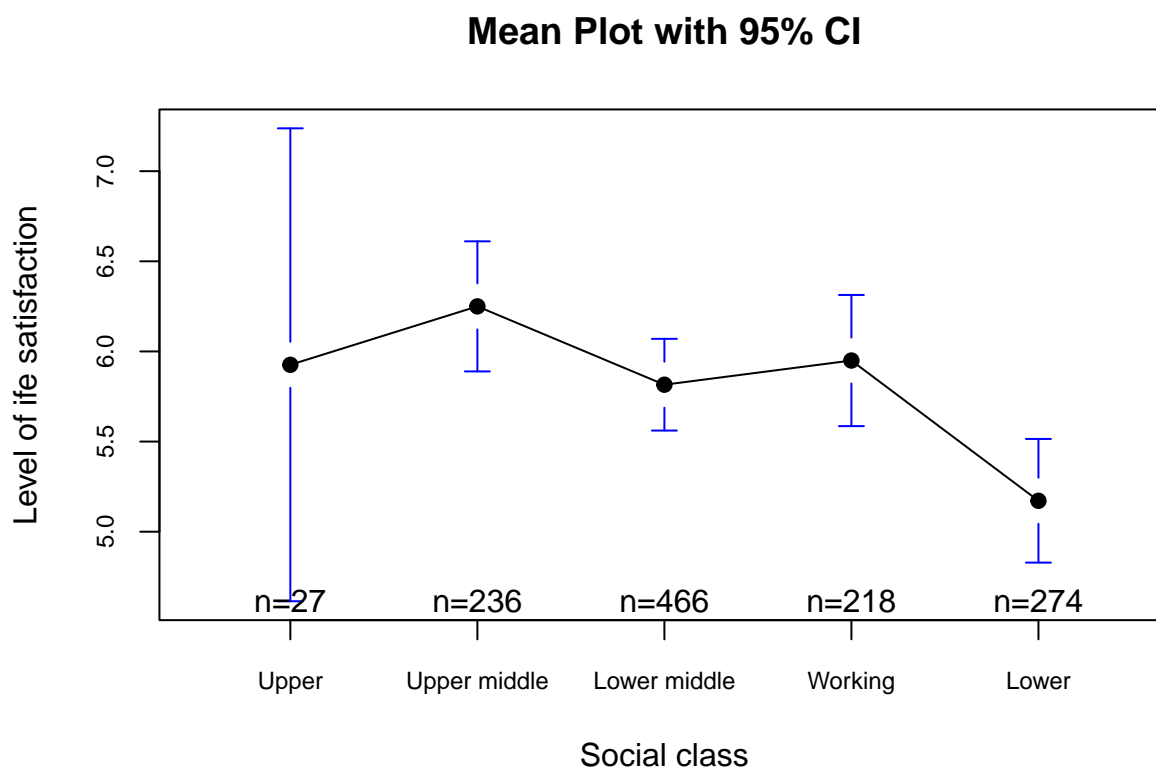
Ethiopia

p-value is significant, so we can speak about the association between social class and level of life satisfaction. But this association is manifested only for such social classes as upper class, lower middle class, and lower. Wherein the level of life satisfaction of the upper middle class is higher (6.2) than for upper class (5.9), and working class is likely to have a little bit higher level of life satisfaction (5.9) than previous category of social class (5.8). Moreover the mean level of life satisfaction of the upper class is on a level with working class (5.9).

```
summary(aov(life_sat ~ social_class, data = WVS_ETH))
```

```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## social_class   4    161   40.25    5.067 0.000474 ***
## Residuals  1216   9660     7.94
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
plotmeans(life_sat ~ social_class, data = WVS_ETH,
  xlab = "Social class", ylab = "Level of life satisfaction",
  main="Mean Plot with 95% CI", pch = 19,
  legends = c("Upper", "Upper middle", "Lower middle", "Working", "Lower"),
  cex.axis = 0.75)
```



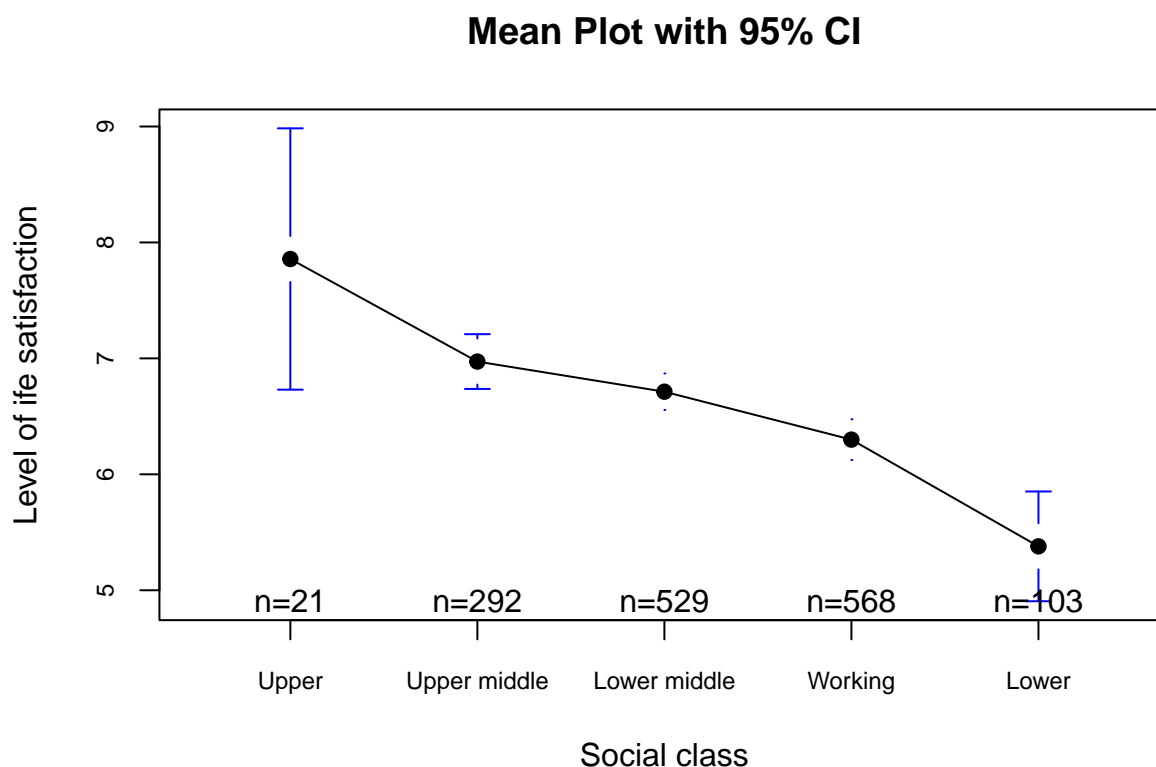
Russia

p-value is significant, so we can speak about the association between social class and level of life satisfaction. Upper class is more likely to have a high level of life satisfaction (7.9), while with the increase in social class categories, the level of life satisfaction decreases (7.0 for Upper middle class, 6.7 for Lower middle class, 6.3 for Working class and 5.4 for Lower class).

```
summary(aov(life_sat ~ social_class, data = WVS_RUS))
```

```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## social_class   4     279    69.65   16.67 2.23e-13 ***
## Residuals  1508    6300     4.18
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
plotmeans(life_sat ~ social_class, data = WVS_RUS,
  xlab = "Social class", ylab = "Level of life satisfaction",
  main="Mean Plot with 95% CI", pch = 19,
  legends = c("Upper", "Upper middle", "Lower middle", "Working", "Lower"),
  cex.axis = 0.75)
```

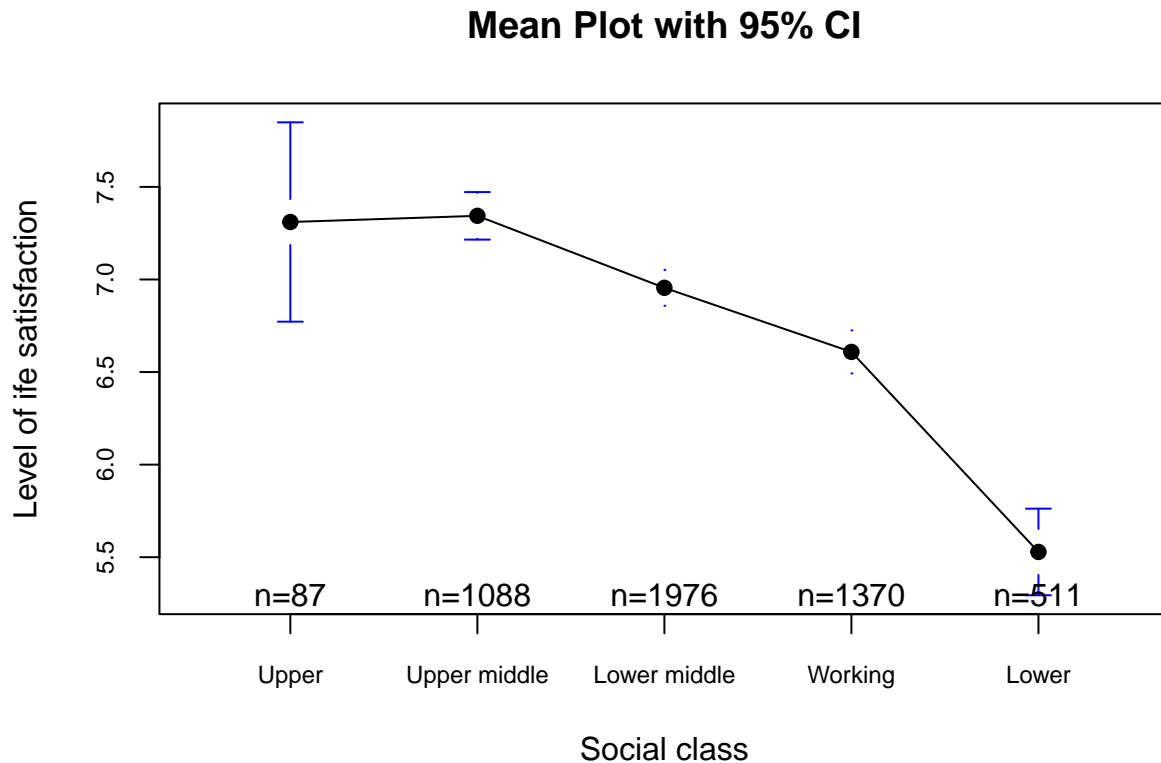
For all countries

p-value is significant, so we can speak about the association between social class and level of life satisfaction. Upper class and Upper middle class are more likely to have a high level of life satisfaction (7.3), while with the increase in social class categories, the level of life satisfaction decreases (7.0 for Lower middle class, 6.6 for Working class and 5.5 for Lower class).

```
summary(aov(life_sat ~ social_class, data = WVS_DEU_CHL_ETH_RUS))
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## social_class   4  1268    317.0   62.74 <2e-16 ***
## Residuals  5027   25399     5.1
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
plotmeans(life_sat ~ social_class, data = WVS_DEU_CHL_ETH_RUS,
  xlab = "Social class", ylab = "Level of life satisfaction",
  main="Mean Plot with 95% CI", pch = 19,
  legends = c("Upper", "Upper middle", "Lower middle", "Working", "Lower"),
  cex.axis = 0.75)
```



10. As a final step, let's create a linear regression model.

1. Outcome: satisfaction. Predictors: income, social class, sex, age, religiosity (is a person religious).
2. There should be a separate regression for each country.
3. Do not forget to calculate standardized coefficients.

Interpret results:

1. Which of the factors are significant?
2. What effect do they have on the outcome?
3. How can you interpret these results?
4. Are there differences in effects between the countries?

Germany

This model is statistically significant (p-value: $< 2.2e-16$) and describes 14% of all cases (R^2 adjusted = 0.139).

1. Such factors as income, social class (working class and lower class), religiosity (an atheist, not a religious person) are significant for Germany. According to values of standardized coefficients, income (0.25), working class (-0.17) and lower class (-0.16) are the strongest factors among all the significant predictors.
2. Effect on the outcome:

- income has a positive effect on the level of life satisfaction: an increase in level of income by 1 point entails an increase in the level life satisfaction by 0.25 points.
- working class relatively to upper class has a negative effect (-0.78), being a working class reduces the level of life satisfaction.
- lower class relatively to upper class has also a negative effect (-1.91), being a lower class reduces the level of life satisfaction.
- religiosity (an atheist) relatively to a religious person has a negative effect (-0.31), being an atheist reduces the level of life satisfaction.
- religiosity (not a religious person) relatively to a religious person has also a negative effect (-0.32), being not a religious person reduces the level of life satisfaction.

```
model_DEU <- lm(life_sat ~ income + social_class + sex + age + religiosity, WVS_DEU)
summary(model_DEU)
```

```
##
## Call:
## lm(formula = life_sat ~ income + social_class + sex + age + religiosity,
##     data = WVS_DEU)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -6.5200 -0.7961  0.1152  1.0882  4.7604
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      6.923782   0.458168  15.112 < 2e-16 ***
## income           0.254715   0.031953   7.972 3.24e-15 ***
## social_class2-Upper middle class -0.311433   0.368752  -0.845 0.398502
## social_class3-Lower middle class -0.371850   0.375491  -0.990 0.322198
## social_class4-Working class      -0.776553   0.390908  -1.987 0.047170 *
## social_class5-Lower class       -1.907568   0.506089  -3.769 0.000171 ***
## sexMale          -0.147732   0.085027  -1.737 0.082527 .
## age              0.003327   0.002359   1.411 0.158602
## religiosityAn atheist -0.313345   0.135991  -2.304 0.021360 *
## religiosityNot a religious person -0.318517   0.090845  -3.506 0.000469 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.564 on 1389 degrees of freedom
## Multiple R-squared:  0.1445, Adjusted R-squared:  0.1389
## F-statistic: 26.06 on 9 and 1389 DF,  p-value: < 2.2e-16
```

```
lm.beta::lm.beta(model_DEU)
```

```
##
## Call:
## lm(formula = life_sat ~ income + social_class + sex + age + religiosity,
##     data = WVS_DEU)
##
## Standardized Coefficients::
##              (Intercept)              income
##                   NA              0.25412648
## social_class2-Upper middle class social_class3-Lower middle class
```

```
##                -0.08927001                -0.10898901
##      social_class4-Working class      social_class5-Lower class
##                -0.17456967                -0.15572376
##                sexMale                age
##                -0.04382753                0.03523339
##      religiosityAn atheist religiosityNot a religious person
##                -0.06043979                -0.09216716
```

Chile

This model is statistically significant (p-value: 1.091e-05) but it describes only 3% of all cases (R^2 adjusted = 0.033).

1. Only such factor social class (upper middle class) is significant for Chile.
2. Effect on the outcome:
 - upper middle class relatively to upper class has a negative effect (-1.37), being an upper middle class reduces the level of life satisfaction.

```
model_CHL <- lm(life_sat ~ income + social_class + sex + age + religiosity, WVS_CHL)
summary(model_CHL)
```

```
##
## Call:
## lm(formula = life_sat ~ income + social_class + sex + age + religiosity,
##     data = WVS_CHL)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -6.3780 -1.3490  0.3601  1.5416  3.8922
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)      8.153331   0.622112  13.106   <2e-16 ***
## income          -0.062983   0.048294  -1.304   0.1925
## social_class2-Upper middle class -1.369337   0.536167  -2.554   0.0108 *
## social_class3-Lower middle class  0.012435   0.460445   0.027   0.9785
## social_class4-Working class     -0.303313   0.477938  -0.635   0.5258
## social_class5-Lower class     -0.928905   0.523778  -1.773   0.0765 .
## sexMale          -0.084512   0.131849  -0.641   0.5217
## age             -0.007591   0.004371  -1.736   0.0828 .
## religiosityAn atheist      0.426926   0.257128   1.660   0.0972 .
## religiosityNot a religious person -0.045259   0.139205  -0.325   0.7452
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.951 on 889 degrees of freedom
## Multiple R-squared:  0.0429, Adjusted R-squared:  0.03321
## F-statistic: 4.427 on 9 and 889 DF, p-value: 1.091e-05
```

```
lm.beta::lm.beta(model_CHL)
```

```
##
## Call:
## lm(formula = life_sat ~ income + social_class + sex + age + religiosity,
##     data = WVS_CHL)
##
## Standardized Coefficients::
##               (Intercept)                income
##                   NA                -0.054106199
## social_class2-Upper middle class social_class3-Lower middle class
##               -0.144041941                0.003107345
## social_class4-Working class      social_class5-Lower class
##               -0.074204071                -0.151659295
##               sexMale                age
##               -0.021283189                -0.059652514
## religiosityAn atheist religiosityNot a religious person
##               0.057679623                -0.011351298
```

Ethiopia

This model is statistically significant (p-value: $< 2.2e-16$) but it describes 2% of all cases (R^2 adjusted = 0.021).

1. Such factors as income, sex (male), and age are significant for Ethiopia. According to values of standardized coefficients, income (0.09) is the strongest factor among all the significant predictors (age (0.07) and sexMale (-0.06)) .
2. Effect on the outcome:
 - income has a positive effect on the level of life satisfaction: an increase in level of income by 1 point entails an increase in the level life satisfaction by 0.11 points.
 - sexMale has a negative on the level of life satisfaction (-0.37), being a male reduces the level of life satisfaction.
 - age has a positive effect (0.02), the older the person, the higher the level of life satisfaction.

```
model_ETH <- lm(life_sat ~ income + social_class + sex + age + religiosity, WVS_ETH)
summary(model_ETH)
```

```
##
## Call:
## lm(formula = life_sat ~ income + social_class + sex + age + religiosity,
##     data = WVS_ETH)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -6.1241 -1.6215 -0.1119  2.1008  5.4111
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)    4.851088   0.674858   7.188 1.14e-12 ***
## income         0.111388   0.048192   2.311  0.0210 *
```

```
## social_class2-Upper middle class    0.390143    0.571428    0.683    0.4949
## social_class3-Lower middle class    0.140630    0.567817    0.248    0.8044
## social_class4-Working class         0.305566    0.585884    0.522    0.6021
## social_class5-Lower class          -0.218501    0.615536   -0.355    0.7227
## sexMale                           -0.368710    0.164229   -2.245    0.0249 *
## age                               0.015996    0.007083    2.258    0.0241 *
## religiosityAn atheist              0.591827    1.989597    0.297    0.7662
## religiosityNot a religious person -0.233679    0.325257   -0.718    0.4726
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.807 on 1211 degrees of freedom
## Multiple R-squared:  0.02815,    Adjusted R-squared:  0.02093
## F-statistic: 3.898 on 9 and 1211 DF,  p-value: 6.777e-05
```

```
lm.beta::lm.beta(model_ETH)
```

```
##
## Call:
## lm(formula = life_sat ~ income + social_class + sex + age + religiosity,
##     data = WVS_ETH)
##
## Standardized Coefficients::
##               (Intercept)                income
##                   NA                0.087540900
## social_class2-Upper middle class social_class3-Lower middle class
##                   0.054321392                0.024088867
## social_class4-Working class      social_class5-Lower class
##                   0.041262556               -0.032142293
## sexMale                age
##                   -0.064994822            0.065301707
## religiosityAn atheist religiosityNot a religious person
##                   0.008438867               -0.020506329
```

Russia

This model is statistically significant (p-value: $< 2.2e-16$) and describes 11% of all cases (R^2 adjusted = 0.105).

1. Such factors as income, social class (lower class), and age are significant for Russia. According to values of standardized coefficients, income (0.28) is the strongest factor among all the significant predictors (lower class (-0.12), age (-0.06)).
2. Effect on the outcome:
 - income has a positive effect on the level of life satisfaction: an increase in level of income by 1 point entails an increase in the level life satisfaction by 0.30 points.
 - lower class relatively to upper class has a negative effect (-1.01), being a lower class reduces the level of life satisfaction.
 - age has a negative effect (-0.01), the older the person, the lower level of life satisfaction.

```
model_RUS <- lm(life_sat ~ income + social_class + sex + age + religiosity, WVS_RUS)
summary(model_RUS)
```

```
##
## Call:
## lm(formula = life_sat ~ income + social_class + sex + age + religiosity,
##     data = WVS_RUS)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -6.4041 -1.3011 -0.0275  1.3413  5.3057
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      6.029239   0.506504   11.904 <2e-16 ***
## income            0.303074   0.031428    9.643 <2e-16 ***
## social_class2-Upper middle class -0.685781   0.446426   -1.536  0.1247
## social_class3-Lower middle class -0.533830   0.443051   -1.205  0.2284
## social_class4-Working class      -0.585399   0.448770   -1.304  0.1923
## social_class5-Lower class       -1.011129   0.492999   -2.051  0.0404 *
## sexMale           0.065921   0.108052    0.610  0.5419
## age              -0.007676   0.003174   -2.418  0.0157 *
## religiosityAn atheist      0.144668   0.211006    0.686  0.4931
## religiosityNot a religious person -0.201613   0.168698   -1.195  0.2322
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.973 on 1503 degrees of freedom
## Multiple R-squared:  0.1105, Adjusted R-squared:  0.1052
## F-statistic: 20.75 on 9 and 1503 DF,  p-value: < 2.2e-16
```

```
lm.beta::lm.beta(model_RUS)
```

```
##
## Call:
## lm(formula = life_sat ~ income + social_class + sex + age + religiosity,
##     data = WVS_RUS)
##
## Standardized Coefficients::
##              (Intercept)              income
##                  NA              0.28478491
## social_class2-Upper middle class social_class3-Lower middle class
##                  -0.12979178                  -0.12207904
## social_class4-Working class      social_class5-Lower class
##                  -0.13594231                  -0.12213706
## sexMale              age
##                  0.01541452                  -0.06274287
## religiosityAn atheist religiosityNot a religious person
##                  0.01699432                  -0.02965101
```

To sum up, income is the strongest factor for Germany, Ethiopia and Russia. Wherein for Germany and Russia the strength of this factor is higher than for Ethiopia. Moreover social class is also a strong factor:

for Germany working class and lower class relatively to upper class have a significant negative effect, for Chile upper middle class has also a significant negative effect, and for Russia lower class have a significant negative effect on level of life satisfaction. In some countries such factors as sex (for Ethiopia), age (it is negative effect for Russia and positive for Ethiopia) and religiosity (for Germany) are also significant.

11. Conclusion

1. In general, what effect do income and social class have on happiness?

- In general, income has a positive effect on happiness among the majority of countries: the higher the income, the higher the level of life satisfaction. This is most clearly observed in Germany and Russia. But there is an exception - Chile, where there is no effect of income on the level of life satisfaction.
- Social class is also a strong factor for majority of countries: for Germany working class and lower class relatively to upper class have a negative effect, for Chile upper middle class has also a negative effect, and for Russia lower class have a negative effect on level of life satisfaction. But there is also an exception - Ethiopia, where there is no effect of social class on the level of life satisfaction.
- Thus, not in all cases, high income or social class entails a high level of life satisfaction.

2. Are there differences between these two variables?

- The difference between the two variables is that while income as a whole can have an effect on life satisfaction, in social class the effect is observed among certain classes, for example, upper middle class, working or lower.

3. Are there differences between the countries? How could you explain them?

- Yes, there are differences between countries, despite the fact that countries are similar in some respects. I think that differences in the assessment of happiness depends on mentality, on how people understand the construct “happiness” and the things they include in it. For instance, people in poor countries can receive joy and happiness from moral satisfaction, which is often not available to citizens of developed countries. At the same time, people in developed countries, in pursuit of increasing their wealth and income, may be in a state of stress, forgetting about happiness and joy in routine matters.