

# Group exercise 1: Survey research

## DATA5207: Data Analysis in the Social Sciences

Emily Wong and Sarah Sweeting

### Introduction

In this study, data from The World Value Survey will be explored to explain quality of life. Part 1 will explore potential predictors to predict higher/lower quality of life with supported theory. Part 2 will tests these predictors through the creation of a dependent variable and predictive models.

### Lab 1: Creating Predictors

#### Understanding survey data

```
survey.data <- read.csv("wvs_data.csv")
```

#### Chosen Predictors

Quality of life is difficult to quantify and can be affected by numerous factors in ones life. Another issue is that ‘quality’ for an individual could be determined by numerous things such as satisfaction, health and wealth.

To help guide our choices of predictors, we will use the World Health Organisations definition of quality of life (QoL) as an “individuals’ perceptions of their position in life in the context of the culture and value systems in which they live and in relation to their goals, expectations, standards and concerns”.

We believe the factors that affect an individual’s perception of life spans further than physical health and wellbeing (however this is obviously also a factor to consider), it can include psychological, environmental, societal and spiritual factors.

**Employment - as means of wealth** Q279: Are you employed now or not? If yes, about how many hours a week do you work? If you have more than one job, please tell us about your main job only.

Response Values:

-2 No answer

1 Full time employee (30 hours a week or more)

2 Part time employee (less than 30 hours a week)

3 Self-employed

4 Retired / On a pension

5 Home duties, not otherwise employed

- 6 Student
- 7 Unemployed
- 8 Other (please specify)

**Education - as means of standard of living** Q275: What is your highest educational level that you have attained?

Response Values:

- 2 No answer
- 1 No formal education
- 2 Primary education only
- 3 Lower secondary education (i.e. Year 9 or less)
- 4 Upper secondary education (i.e. between Year 10 and Year 12)
- 5 Post-secondary non-tertiary education (e.g. apprenticeship or certificate)
- 7 Bachelor or equivalent
- 8 Master or equivalent
- 9 Doctoral or equivalent

**Security** Q131: how secure do you feel these days?

Response Values:

- 2 No answer
- 1 Very secure
- 2 Quite secure
- 3 Not very secure
- 4 Not at all secure

**Human Rights** Q253: How much respect is there for individual human rights nowadays in this country? Do you feel there is...

Response Values:

- 2 No answer
- 1 A great deal of respect
- 2 Some respect
- 3 Not much respect
- 4 No respect at all

**Health Levels** Q53: "In the last 12 months, how often have you or your family gone without medicine or medical treatment that you needed?"

Response Values:

- 2 No answer
- 1 Often

- 2 Sometimes
- 3 Rarely
- 4 Never

**Social & Personal Connectivity** Q2. For each of the following, indicate how important it is in your life. How important is... Friends

Response Values:

- 2 No answer
- 1 Very important
- 2 Rather important
- 3 Not very important
- 4 Not at all important

## Theory

Education and employment: When an individual has higher educational attainments and are employed, they are able to have a better quality of life. This is due to higher economic well being and financial security to meet basic needs such as healthcare access and participating in leisure activities. It also can lead to a greater sense of purpose and personal development.

Confounding factor here may be income, since these factors are indicators of what sort of job an individual has and how much they are getting paid as a result (higher education levels = greater income, more hours in employment = higher income)

Security: When an individual is able to live without fear or risk and feels stable, their quality of life will increase. Security can be an impact on the quality of life as it is a measure of both physical safety and economic safety, physical safety affecting individuals health and mental well being, while as economic security to have social safety net to tackle financial challenges and access basic needs.

Human rights: When society upholds and protects human rights, individuals tend to experience better QoL. It is a basic need but also a measure of psychological well-being because of its ability to foster belonging within communities and equality within a just society. It also could promote quality of opportunity, a fair chance for success in society. can you add a blurb for health levels and social and personal connectivity

Health Levels: Health is a foundational element of quality of life. Good health enables individuals to engage actively in various aspects of life including work, social interactions, and leisure activities, thus directly influencing their overall life satisfaction and well-being. Therefore, having a lack of access to necessary medical intervention can lead to decreased physical capabilities, psychological stress, and financial burdens, all of which negatively affect one's quality of life.

Social & Personal Connectivity: Human beings are inherently social creatures, and the quality of our social interactions can significantly impact our mental and emotional well-being. Strong connections with family and friends provide emotional support, reduce stress, and contribute to a sense of belonging and happiness. Furthermore, placing value in personal relationships can influence one's self-esteem and provide opportunities for meaningful engagement in community activities, thereby enhancing an individual's overall quality of life.

## Exploring Predictors

```
glimpse(colnames(survey.data))
```

```
## chr [1:317] "ID" "Mode" "State" "V1" "V2" "V3" "V4" "V5" "V6" "V7" "V8" ...
```

The column names aren't entirely that helpful. We re-code them for convenience.

```
categories = c("Employment", "Education", "Security", "Human Rights", "Treatment Levels", "Friendship  
questions = c("Q279", "Q275", "Q131", "Q253", "Q53", "Q2")  
key = c("V249", "Q275", "V170", "V142", "V190", "V5")
```

```
v249_employment_status <- c(  
  "No answer",  
  "Full time employee (30 hours a week or more)",  
  "Part time employee (less than 30 hours a week)",  
  "Self-employed",  
  "Retired / On a pension",  
  "Home duties, not otherwise employed",  
  "Student",  
  "Unemployed",  
  "Other (please specify)"  
)  
  
values.249 = c(-2,1,2,3,4,5,6,7,8)  
  
q275_education = c("No answer",  
  "No formal education",  
  "Primary education only",  
  "Lower secondary education (i.e. Year 9 or less)",  
  "Upper secondary education (i.e. between Year 10 and Year 12)",  
  "Post-secondary non-tertiary education (e.g. apprenticeship or certificate)",  
  "Bachelor or equivalent",  
  "Master or equivalent",  
  "Doctoral or equivalent")  
  
values.275 = c(-2,1,2,3,4,5,7,8,9)  
  
v170_how_secure_are_you = c(  
  "No answer",  
  "Very secure",  
  "Quite secure",  
  "Not very secure",  
  "Not at all secure"  
)  
  
values.170 = c(-2,1,2,3,4)  
  
v142_respect_human_rights = c(  
  "No answer",  
  "A great deal of respect",  
  "Some respect",  
  "Not much respect",  
  "No respect at all"  
)  
  
values.142 = c(-2,1,2,3,4)
```

```
v190_gone_without_medicine = c(
  "No answer",
  "Often",
  "Sometimes",
  "Rarely",
  "Never"
)
```

```
values.190 = c(-2,1,2,3,4)
```

```
v5_friends = c(
  "No answer",
  "Very important",
  "Rather important",
  "Not very important",
  "Not at all important"
)
```

```
values.5 = c(-2,1,2,3,4)
```

We also will make a dataframe created with the questions, keys and values to be able to make functions later for plotting.

```
vectors <- list(
  list(
    values = values.249,
    chars = v249_employment_status
  ),
  list(
    values = values.275,
    chars = q275_education
  ),
  list(
    values = values.170,
    chars = v170_how_secure_are_you
  ),
  list(
    values = values.142,
    chars = v142_respect_human_rights
  ),
  list(
    values = values.190,
    chars = v190_gone_without_medicine
  ),
  list(
    values = values.5,
    chars = v5_friends
  )
)

# Combine data into a dataframe
Question.key <- bind_rows(lapply(1:length(categories), function(i) {
```

```

data.frame(
  Categories = rep(categories[i], length(vectors[[i]]$values)),
  Question.Number = rep(questions[i], length(vectors[[i]]$values)),
  Codebook = rep(key[i], length(vectors[[i]]$values)),
  Question.values = vectors[[i]]$values,
  Question.chars = vectors[[i]]$chars
)
}))

```

```

survey.predictors <- data.frame(
  education = survey.data$Q275,
  employment = survey.data$V249,
  security = survey.data$V170,
  rights = survey.data$V142,
  health = survey.data$V190,
  social = survey.data$V5
)

```

## Distribution of predictors

```

summary = describe(survey.predictors)
print(summary)

```

```

##          vars      n mean   sd median trimmed  mad min max range  skew kurtosis
## education      1 1813 5.36 2.25      5    5.56 2.97  -2   9    11 -1.14    2.17
## employment      2 1813 2.71 1.91      2    2.59 1.48  -2   8    10  0.23    0.19
## security        3 1813 1.98 0.72      2    1.98 0.00  -2   4     6 -0.97    6.58
## rights          4 1813 1.88 0.80      2    1.85 0.00  -2   4     6 -0.35    3.51
## health          5 1813 3.64 0.88      4    3.87 0.00  -2   4     6 -3.50   15.93
## social          6 1813 1.39 0.80      1    1.40 0.00  -2   4     6 -1.44    6.31
##
##          se
## education 0.05
## employment 0.04
## security  0.02
## rights    0.02
## health    0.02
## social    0.02

```

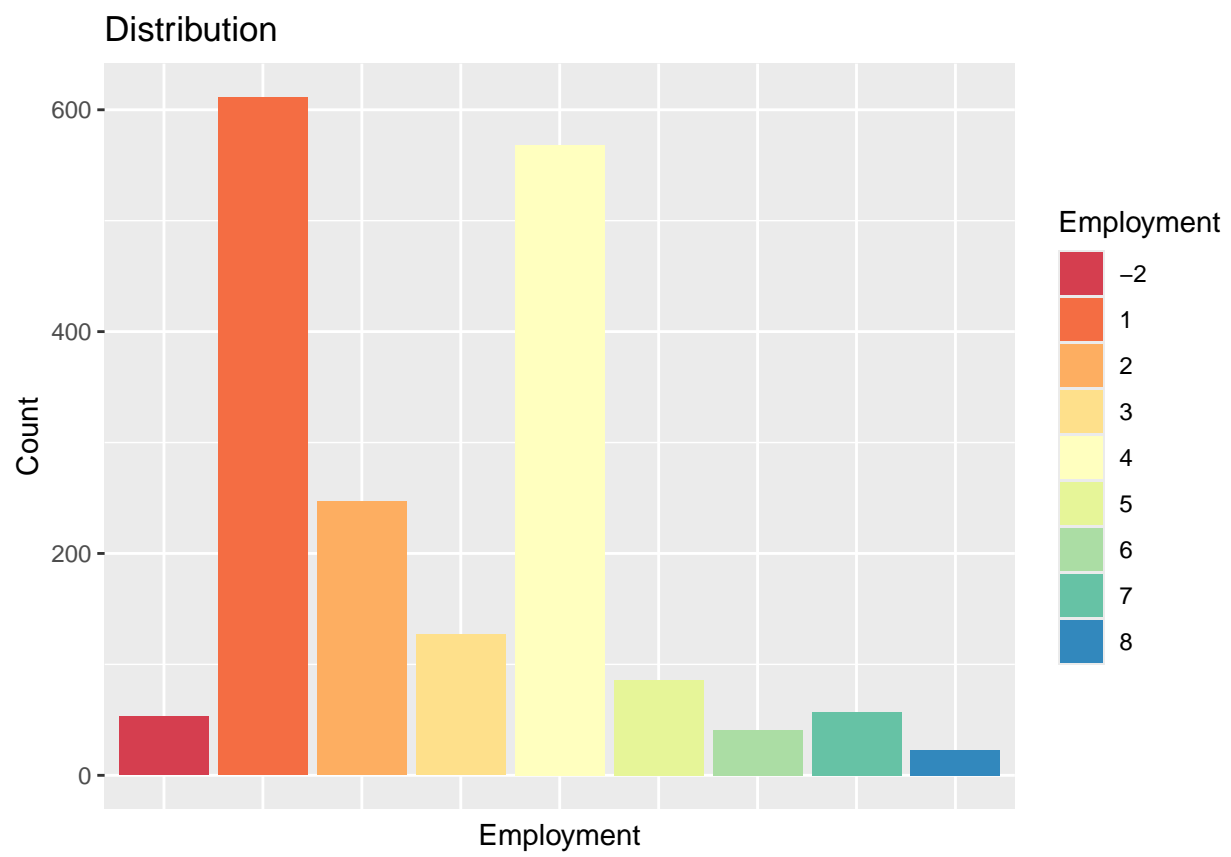
```

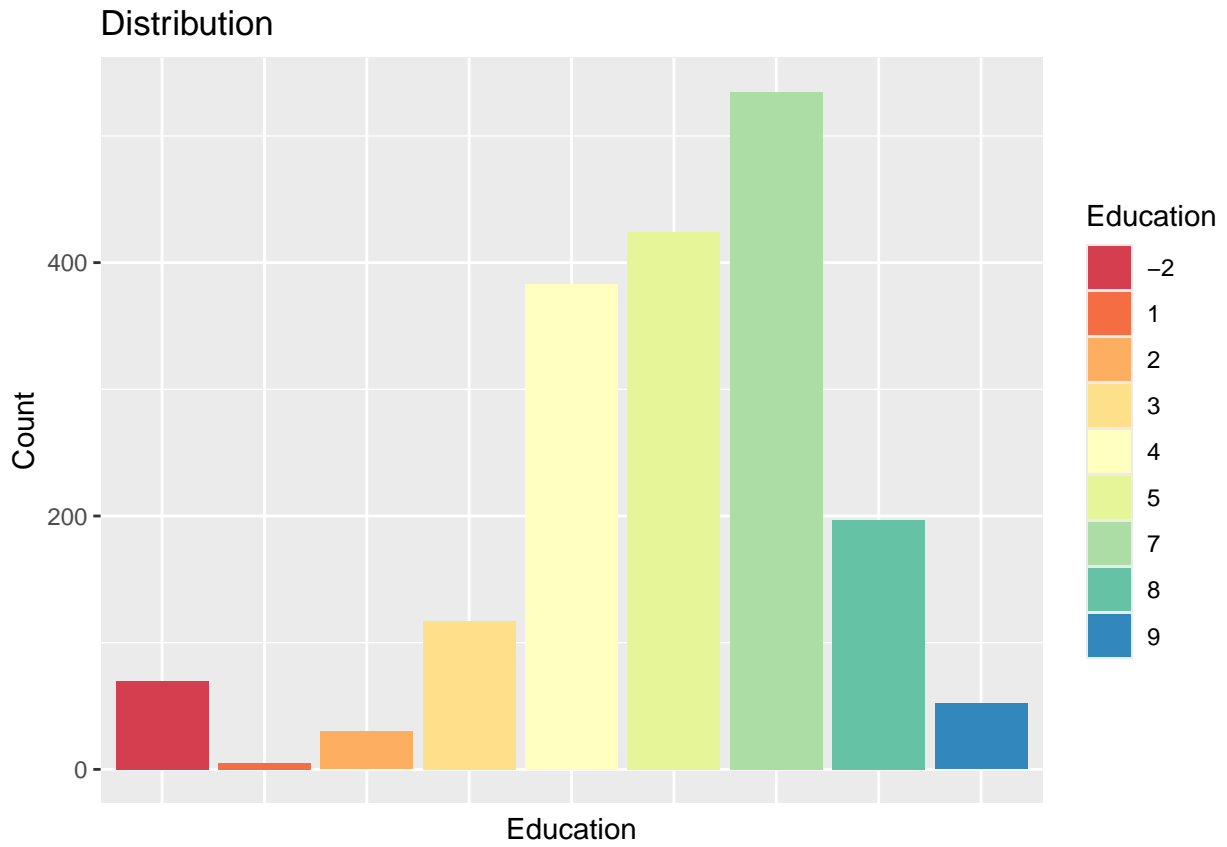
# This function creates bar plot for a predictor and fills in the character values for the labels
plot_barplots <- function(column, name, values_fill) {
  ggplot(survey.data, aes(x = factor(!column), fill = factor(!column))) +
    geom_bar() +
    labs(title = "Distribution", x = name, fill = name, y = "Count") +
    scale_fill_brewer(palette = "Spectral") +
    theme(axis.text.x = element_blank(), axis.ticks.x = element_blank())
}

# This loop goes through the dataframe and applies the function to create bar plots for each predictor
unique_codebooks <- unique(Question.key$Codebook)
for (codebook in unique_codebooks) {
  subset_data <- filter(Question.key, Codebook == codebook)
}

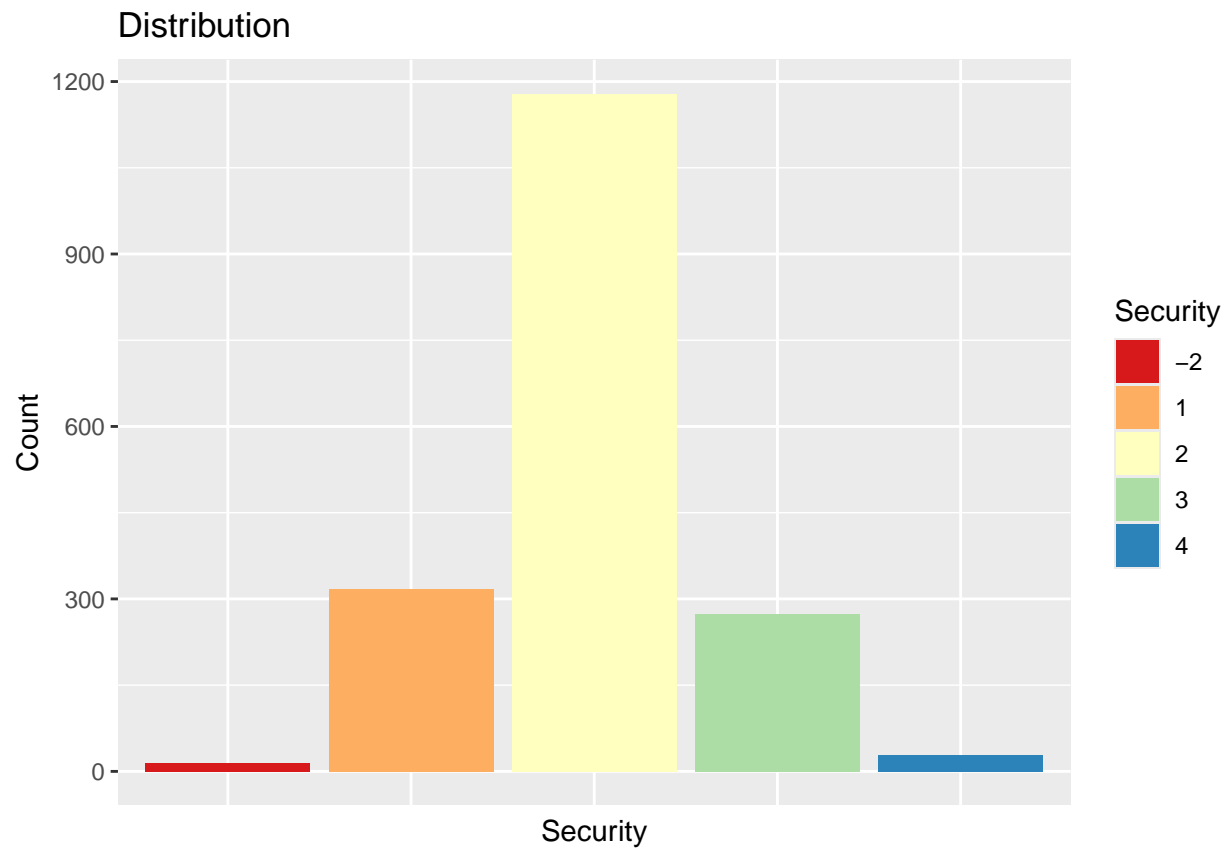
```

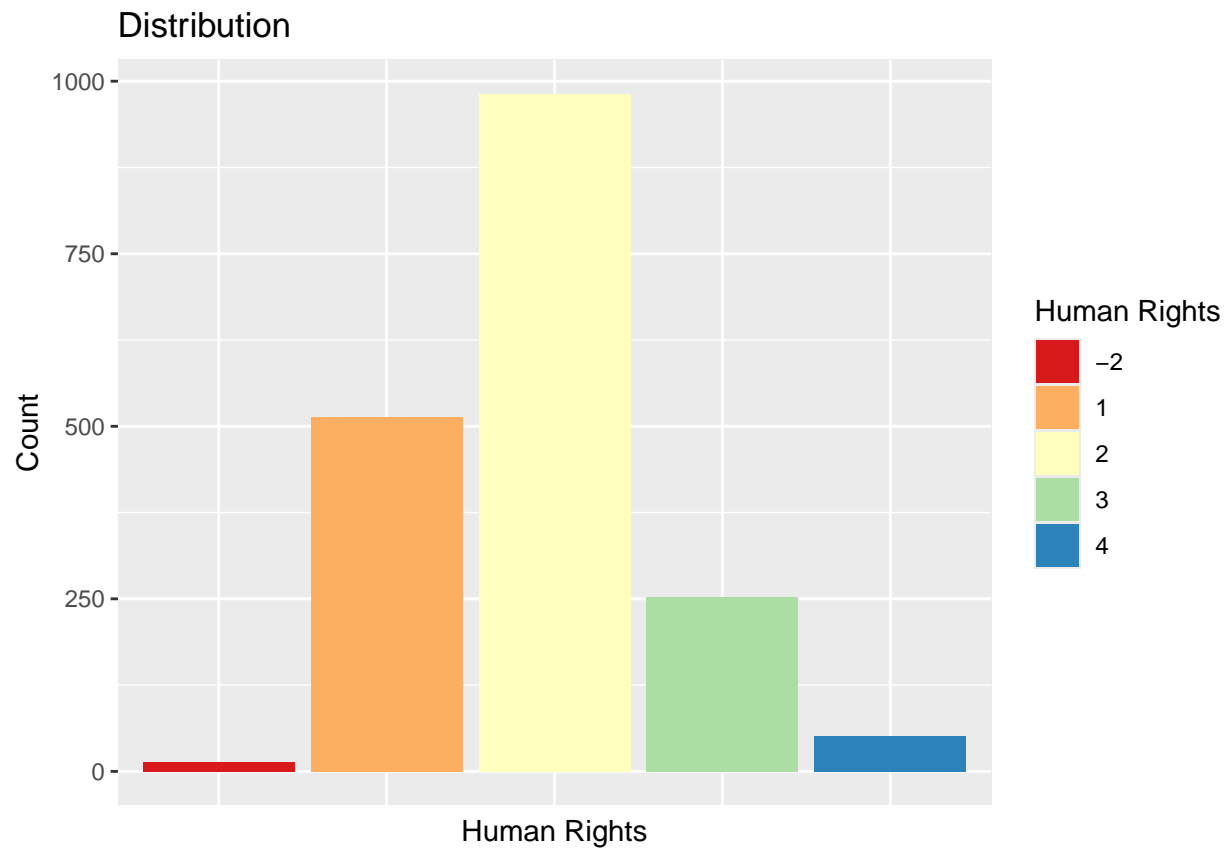
```
print(plot_barplots(column = sym(codebook), name = unique(subset_data$Categories), values_fill = unique(subset_data$Categories))
}
```

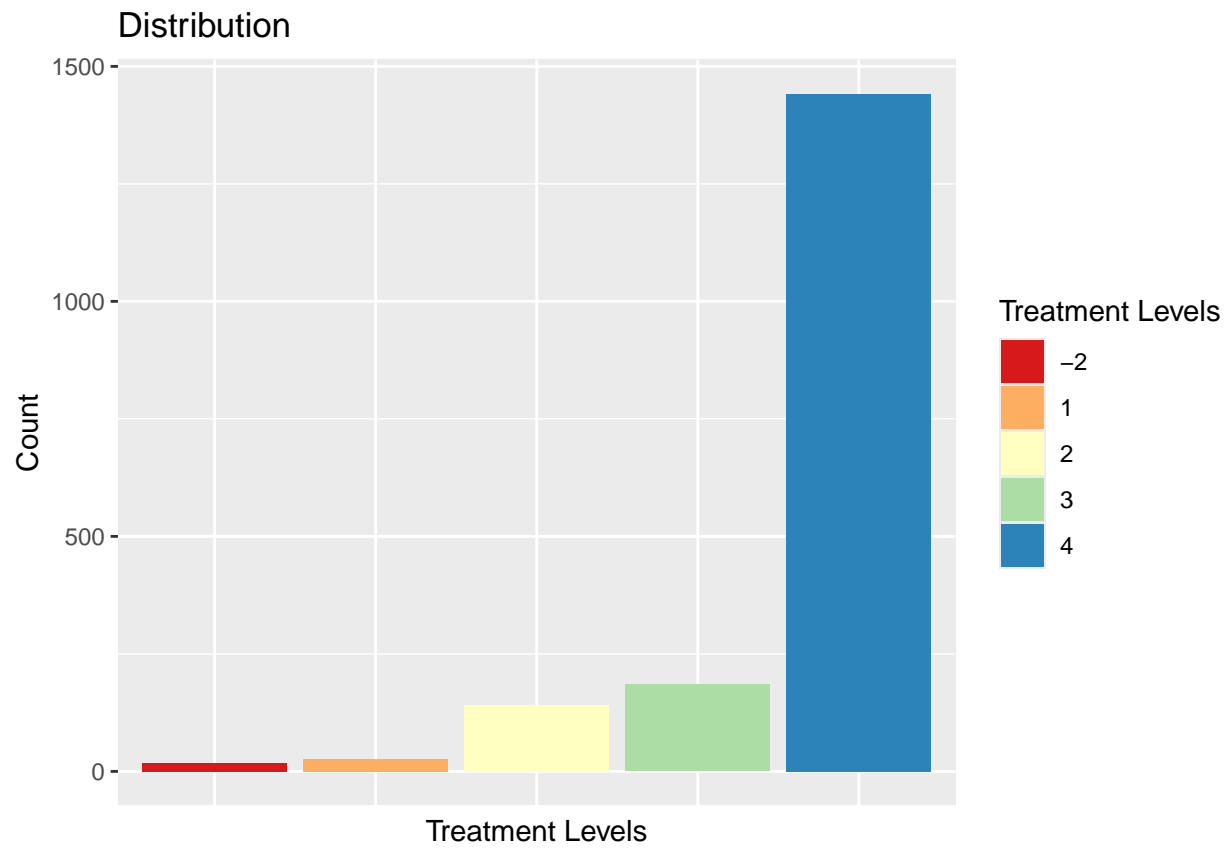


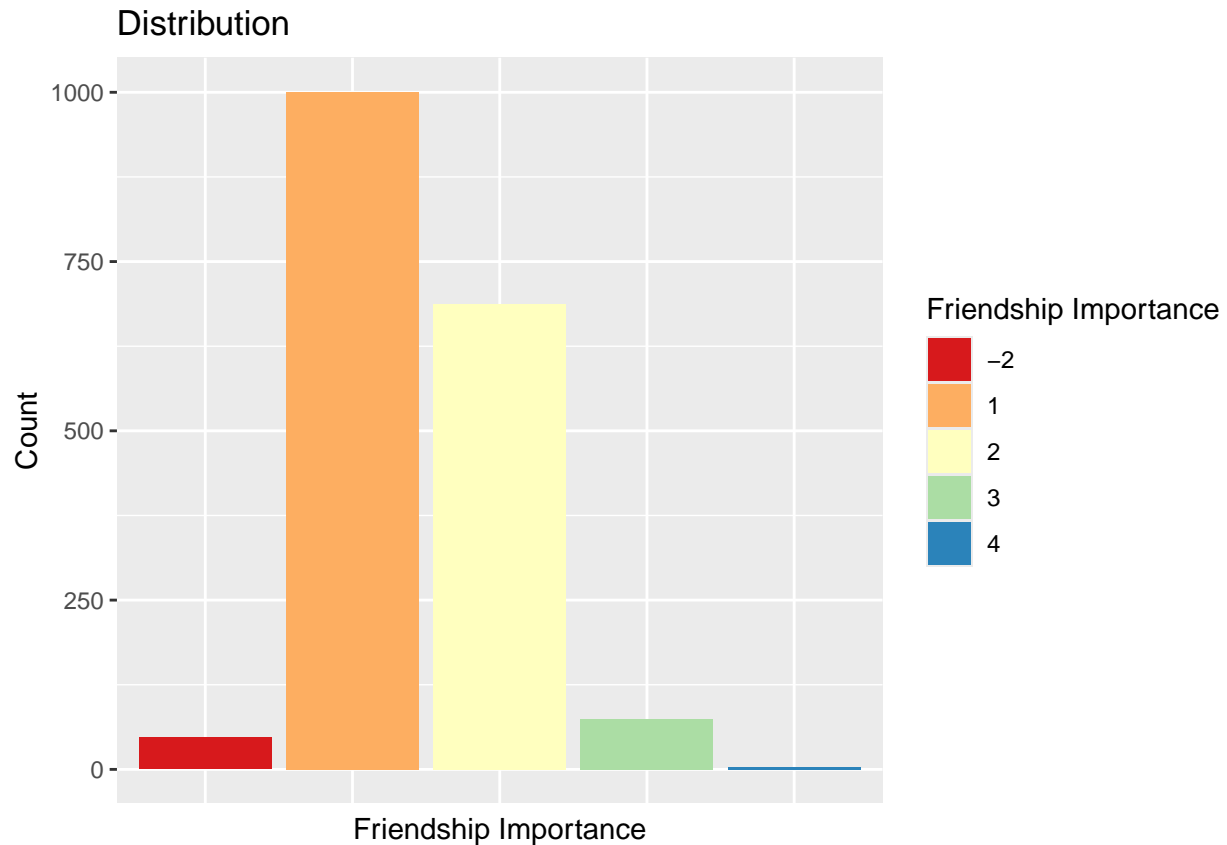












## Employment

### Descriptive Analysis:

- Full-time employment is most common, suggesting stable work schedules for many.
- Significant numbers of students and part-time employees, indicating a mix of work and study.
- Fewer respondents in self-employed, retired, and home duties categories, showing less representation of these demographics.
- Positively skewed distribution with a majority as full-time employees and diminishing counts towards 'Unemployed' and 'Other'.

### Statistical Analysis:

- Mean employment score around 2.709, indicating a skew towards full-time employment.
- Median of 2, reinforcing the prevalence of full-time employment.

## Education

### Descriptive Analysis:

- Upper secondary education is most prevalent, indicating a common level of educational attainment.
- A considerable proportion of respondents with a bachelor's degree, possibly reflecting the target demographic or societal education trends.

- Lesser counts of post-secondary non-tertiary education, master's, or doctoral degrees.
- A distribution with a primary mode at upper secondary education and a secondary mode at the bachelor level, with tapering counts at the lowest and highest education levels.

#### **Statistical Analysis:**

- Mean education level approximately 5.36, showing a skew towards upper secondary education.
- Median value of 5, indicating over half of the respondents completed at least upper secondary education.

## **Security Perception**

#### **Descriptive Analysis:**

- Majority of respondents feel quite secure, which might reflect societal stability or personal circumstances.
- Smaller proportions feel very secure or not secure, suggesting fewer extremes in security perception.
- Negatively skewed distribution where a large majority feels 'Quite secure', and fewer responses are on the 'Not very secure' or 'Not at all secure' end.

#### **Statistical Analysis:**

- Mean close to 2, leaning towards 'Quite secure'.
- Median of 2, confirming 'Quite secure' as a common sentiment.

## **Human Rights Perception**

#### **Descriptive Analysis:**

- Most believe there is 'some respect' for human rights, indicating moderate views.
- Significant perception of 'a great deal of respect', suggesting a positive outlook among many.
- A distribution with a slight negative skew, indicating that most respondents feel there is 'Some respect' for human rights, with a substantial number also feeling there is 'A great deal of respect'.

#### **Statistical Analysis:**

- Mean of approximately 1.882, hinting the average perception is close to 'some respect'.
- Median of 2, aligning with the average perception towards human rights.

## **Health Treatment Levels**

#### **Descriptive Analysis:**

- Predominant 'never' category might suggest good health or barriers to healthcare.
- 'Sometimes' as the next most common response, indicating occasional health concerns.
- A heavily positively skewed distribution, where most respondents 'Never' seek treatment, with progressively fewer responses for more frequent healthcare utilisation.

### Statistical Analysis:

- Mean around 3.638, trending towards infrequently seeking treatment.
- Median of 4, suggesting the middle ground of responses leans towards 'rarely'.

## Friendship Importance

### Descriptive Analysis:

- Friendship considered 'very important' by many, emphasizing the high value on social relationships.
- 'A strongly negatively skewed distribution, showing that 'Very important' is the predominant response, with 'Not at all important' being the least common.

### Statistical Analysis:

- Mean skewed towards 'very important', with a mean value of around 1.839.
- Median value of 1, indicating a majority view friendship as 'very important'.

## Lab 2: Relationship of independent and dependent variables

We now will begin using our chosen independent variables to create a linear regression model and determine the relationship between these and life quality.

First we need to make the dependent variable - life.quality.

### Creating the dependent variable

We recode our variables to provide the variables with more intuitive names, to make our work easier and also recode our predictors to character variables to do descriptive statistics.

```
survey.data <- survey.data %>%  
  mutate(happiness = dplyr::recode(V10,  
                                    '1' = 4,  
                                    '2' = 3,  
                                    '3' = 2,  
                                    '4' = 1,  
                                    '-2' = NULL),  
         health = dplyr::recode(V11,  
                                 '1' = 5,  
                                 '2' = 4,  
                                 '3' = 3,  
                                 '4' = 2,  
                                 '5' = 1,  
                                 '-2' = NULL),  
         finances = dplyr::recode(V59,  
                                   '-2' = NULL,  
                                   .default = V59),  
         satisfaction = dplyr::recode(V23,
```

```

        '-2' = NULL,
        .default = V23),
freedom = dplyr::recode(V55,
        '-2' = NULL,
        .default = V55))

```

## Factor Analysis

Factor Analysis is a measurement model of a latent variable. Latent variable cannot be directly measured. Instead, it is seen through relationships between variables. We assume latent factor drives responses to variables.

```

fa.fit <- fa(survey.data[,c("happiness",
                           "health",
                           "finances",
                           "satisfaction",
                           "freedom")],
            nfactors=1)

survey.data$life.quality <- as.numeric(fa.fit$scores)

```

Factor analysis using fa will calculate the optimal weights as seen below.

```

## Factor Analysis using method = minres
## Call: fa(r = survey.data[, c("happiness", "health", "finances", "satisfaction",
##      "freedom")], nfactors = 1)
## Standardized loadings (pattern matrix) based upon correlation matrix
##           MR1   h2   u2 com
## happiness  0.64 0.41 0.59  1
## health     0.52 0.28 0.72  1
## finances   0.63 0.39 0.61  1
## satisfaction 0.93 0.86 0.14  1
## freedom    0.71 0.50 0.50  1
##
##           MR1
## SS loadings  2.43
## Proportion Var 0.49
##
## Mean item complexity = 1
## Test of the hypothesis that 1 factor is sufficient.
##
## df null model = 10 with the objective function = 1.73 with Chi Square = 3136.16
## df of the model are 5 and the objective function was 0.02
##
## The root mean square of the residuals (RMSR) is 0.02
## The df corrected root mean square of the residuals is 0.04
##
## The harmonic n.obs is 1788 with the empirical chi square 22.35 with prob < 0.00045
## The total n.obs was 1813 with Likelihood Chi Square = 33.28 with prob < 3.3e-06
##
## Tucker Lewis Index of factoring reliability = 0.982

```

```
## RMSEA index = 0.056 and the 90 % confidence intervals are 0.039 0.075
## BIC = -4.23
## Fit based upon off diagonal values = 1
## Measures of factor score adequacy
##
## Correlation of (regression) scores with factors MR1 0.95
## Multiple R square of scores with factors 0.90
## Minimum correlation of possible factor scores 0.79
```

```
describe(survey.data$life.quality)
```

```
## vars n mean sd median trimmed mad min max range skew kurtosis se
## X1 1 1758 0 0.94 0.17 0.08 0.8 -3.93 1.51 5.43 -1.02 1.33 0.02
```

As you can see, these are approximately standardised, with a mean of zero and standard deviation of (almost) one.

## Descriptive Statistics

As seen above, there are 55 NA values which will be difficult to plot. We will remove these.

```
survey.data = drop_na(survey.data, life.quality)
```

```
ggplot(survey.data, aes(life.quality)) +
  geom_histogram(fill = 'black') +
  theme_minimal()
```

Interestingly, life quality is negatively skewed, demonstrating a larger amount of people with positive life qualities than negative with the maximum value being 1.51 and the minimum being -3.93. With average life quality score of 0, this suggests individuals are experiencing moderate level of life quality, but there is a subset of the population with significantly lower life quality than the average due to the negative skewness which brings down the overall distribution.

```
## creating a new data frame that will store the independant variables we are testing.
survey.predictors <- data.frame(
  education = survey.data$Q275,
  employment = survey.data$V249,
  security = survey.data$V170,
  rights = survey.data$V142,
  health = survey.data$V190,
  social = survey.data$V5,
  life.qual = survey.data$life.qual
)
```

```
## including the categorical answers
```

```
#### Employment Q279: Are you employed now or not? If yes, about how many hours a week do you work? If
survey.predictors$emp_cat <- recode(survey.data$V249,
  ` -2 ` = "No answer",
  ` 1 ` = "Full time employee (30 hours a week or more)",
  ` 2 ` = "Part time employee (less than 30 hours a week)",
```



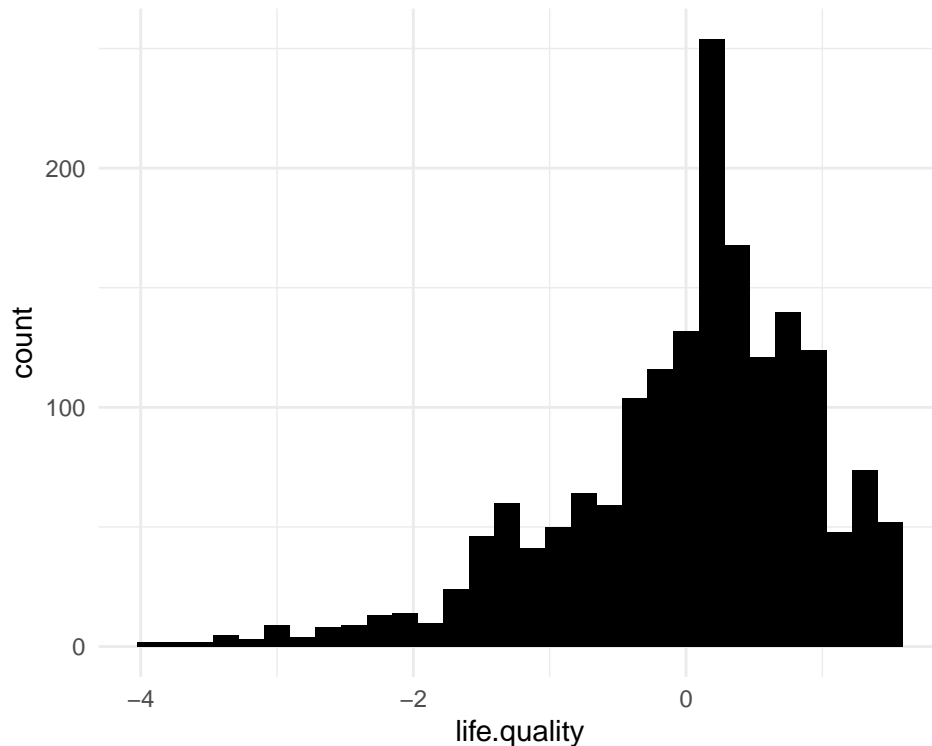


Figure 1: Distribution of quality of life measure produced by factor analysis.

```

`3` = "Self-employed",
`4` = "Retired / On a pension",
`5` = "Home duties, not otherwise employed",
`6` = "Student",
`7` = "Unemployed",
`8` = "Other (please specify)"

#### Education - Q275: What is your highest educational level that you have attained?
survey.predictors$edu_cat <- recode(survey.data$Q275,
  `2` = "No answer",
  `1` = "No formal education",
  `2` = "Primary education only",
  `3` = "Lower secondary education (i.e. Year 9 or less)",
  `4` = "Upper secondary education (i.e. between Year 10 and Year 12)",
  `5` = "Post-secondary non-tertiary education (e.g. apprenticeship or certification)",
  `7` = "Bachelor or equivalent",
  `8` = "Master or equivalent",
  `9` = "Doctoral or equivalent")

#### Security - Q131: how secure do you feel these days?
survey.predictors$secure_cat <- recode(survey.data$V170,
  `2` = "No answer",
  `1` = "Very secure",
  `2` = "Quite secure",
  `3` = "Not very secure",

```

```

`4` = "Not at all secure")

#### Human Rights - Q253: How much respect is there for individual human rights nowadays in this country?
survey.predictors$rights_cat <- recode(survey.data$V142,
  `2` = "No answer",
  `1` = "A great deal of respect",
  `2` = "Some respect",
  `3` = "Not much respect",
  `4` = "No respect at all")

#### Health Levels - Q53: "In the last 12 months, how often have you or your family gone without medication?"
survey.predictors$health_cat <- recode(survey.data$V190,
  `2` = "No answer",
  `1` = "Often",
  `2` = "Sometimes",
  `3` = "Rarely",
  `4` = "Never")

#### Social & Personal Connectivity - Q2. For each of the following, indicate how important it is in your life.
survey.predictors$social_cat <- recode(survey.data$V5,
  `2` = "No answer",
  `1` = "Very important",
  `2` = "Rather important",
  `3` = "Not very important",
  `4` = "Not at all important")

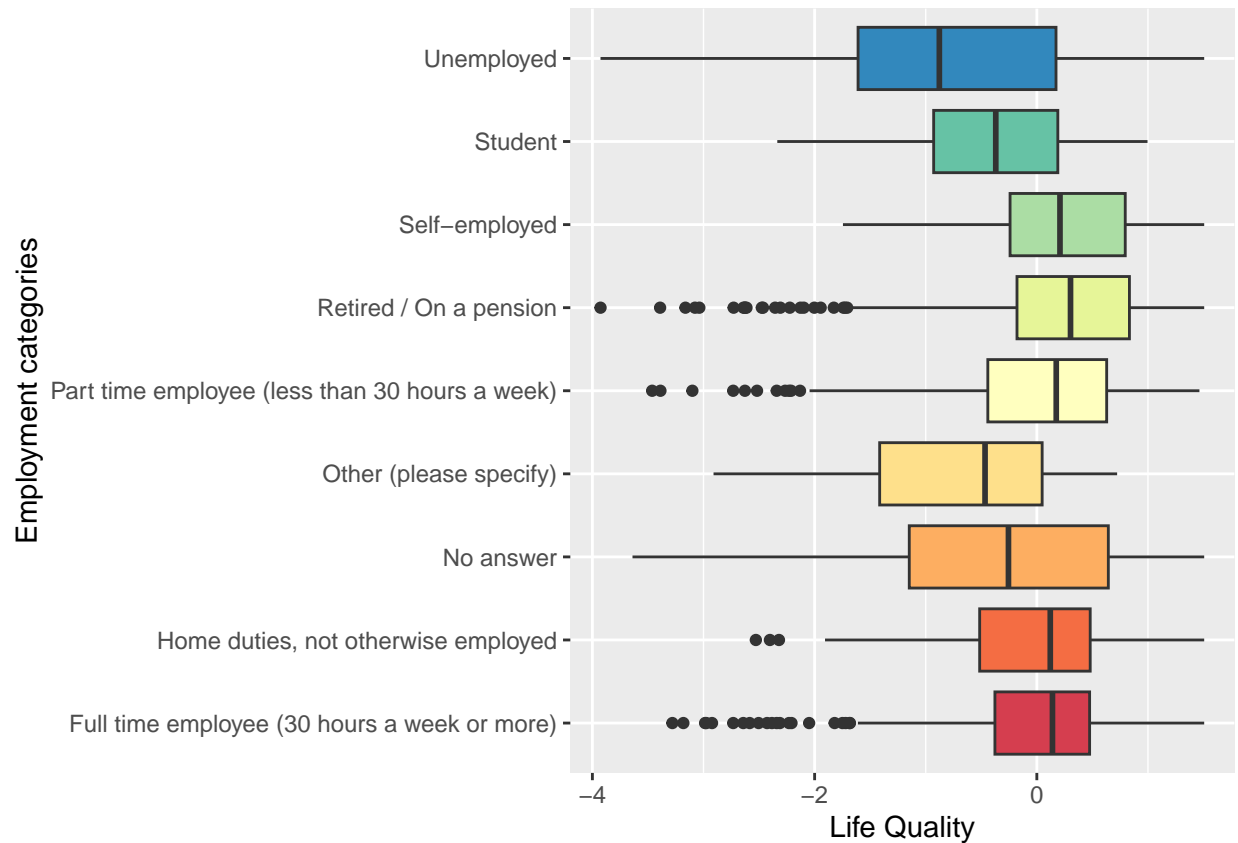
```

Relationship of dependent variable (quality of life) and the chosen predictors

```

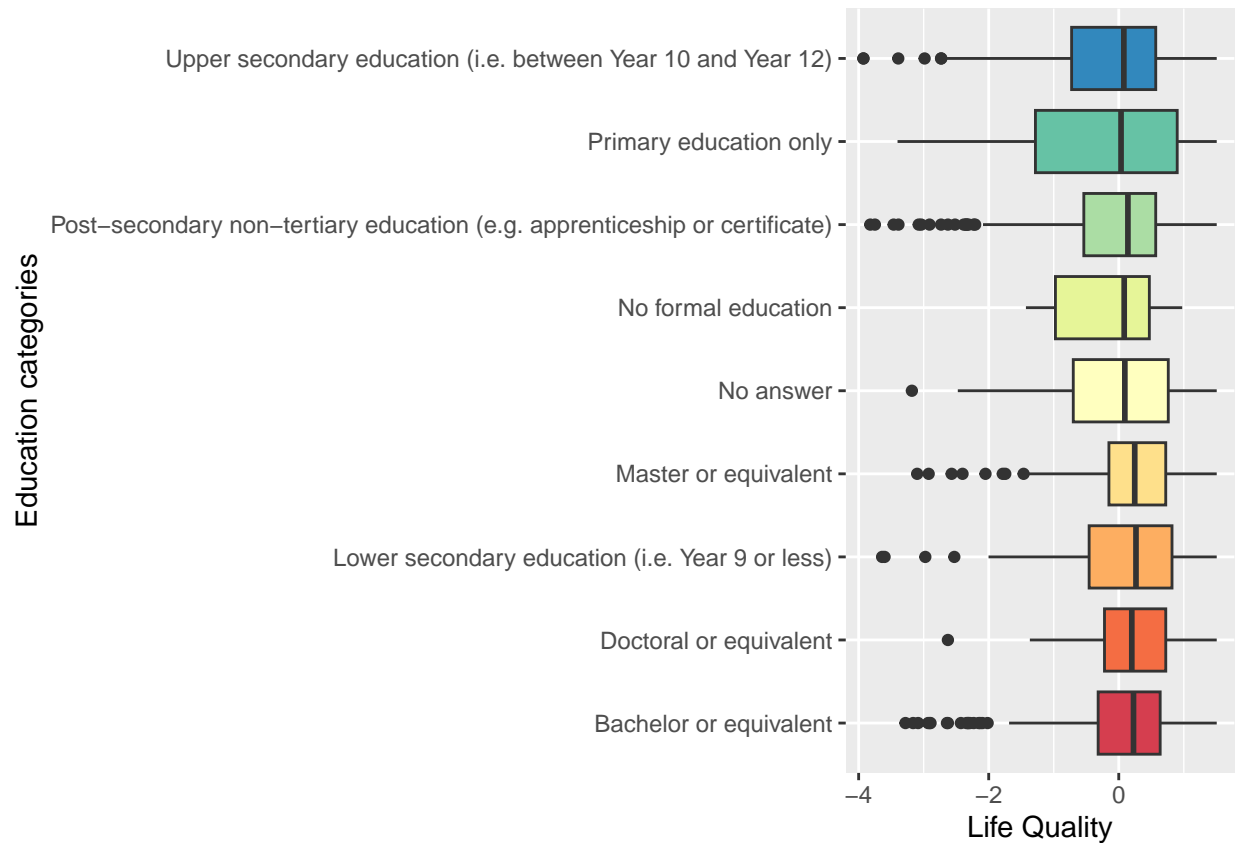
ggplot(survey.data, aes(x=survey.predictors$life.qual, y=survey.predictors$emp_cat, fill = survey.predictors$rights_cat)) +
  geom_boxplot() +
  xlab("Life Quality") +
  ylab("Employment categories") +
  scale_fill_brewer(palette = "Spectral") +
  theme(legend.position = "None")

```



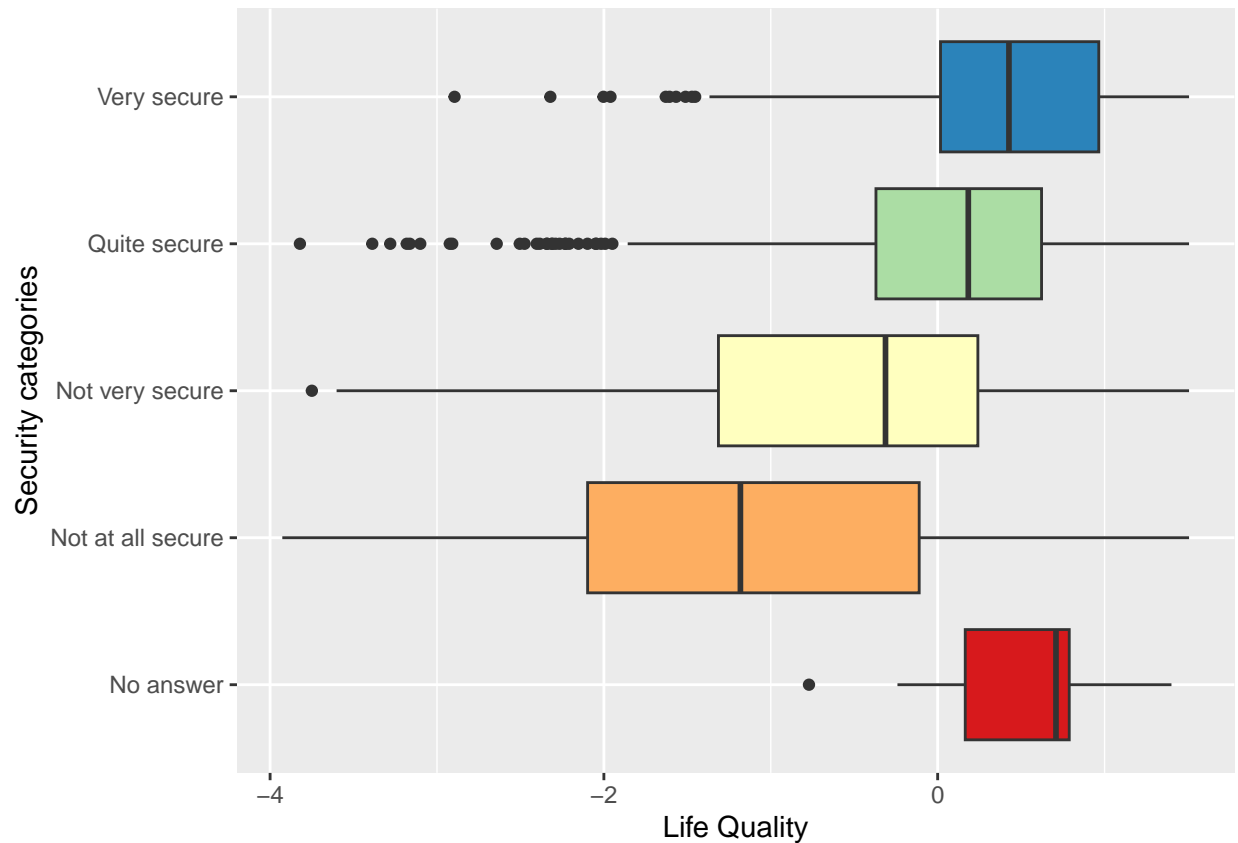
The relationship plotted in this bar plot demonstrates that those who are employed tend to have a higher life quality than those who are not (besides those who are retired). Interestingly, those who are full time employed and retired show both ends quality of life with outliers towards the negative life quality (suggesting people who may struggle to increase life quality even though they have fulltime income or pension). Unemployment shows the largest range of life quality while as self-employed is mainly skewed towards higher values of life quality.

```
ggplot(survey.data, aes(x=survey.predictors$life.qual, y=survey.predictors$edu_cat, fill = survey.predi
  geom_boxplot() +
  xlab("Life Quality") +
  ylab("Education categories") +
  scale_fill_brewer(palette = "Spectral") +
  theme(legend.position = "None")
```



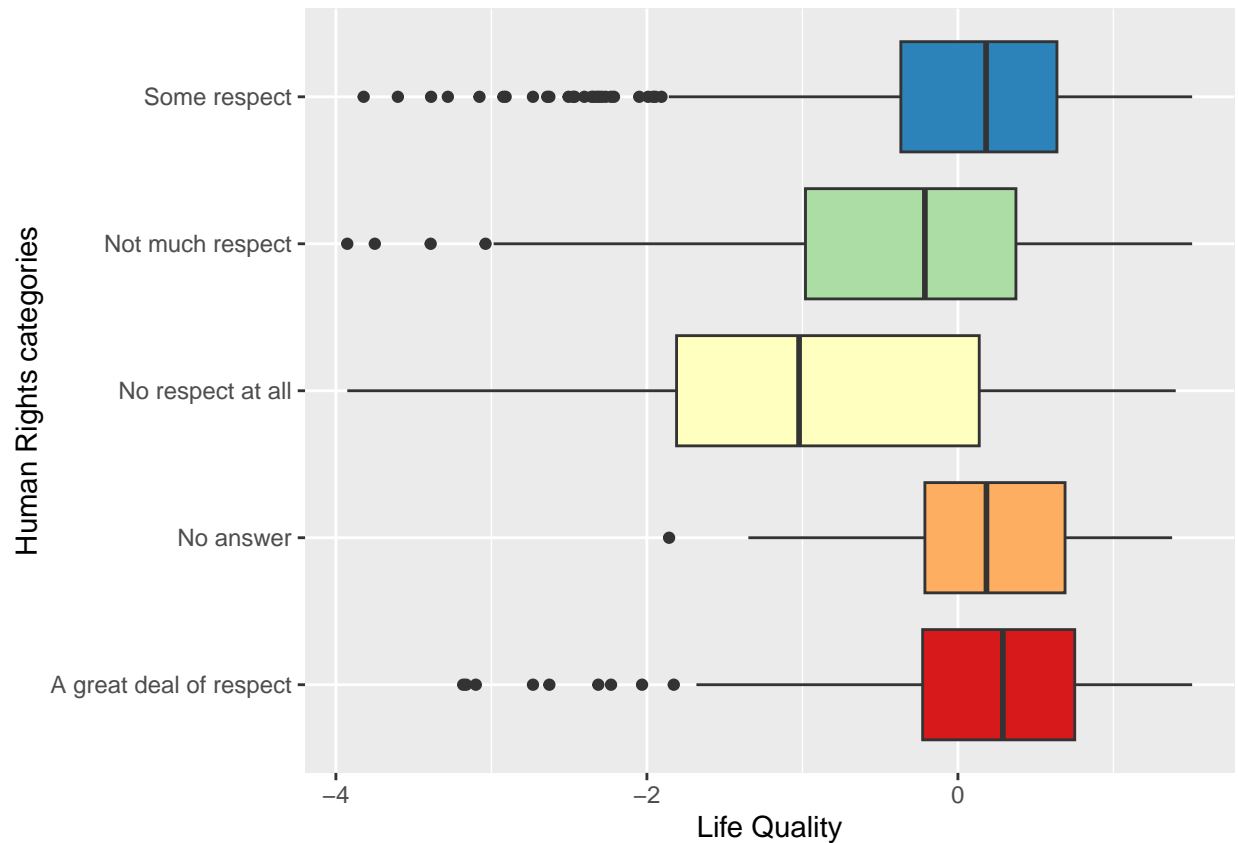
Life quality in relation to education interestingly is always has a median above 0, showing positive median life qualities for all categories. However, the spread of life quality is diverse with those having a doctoral or masters tending to have a smaller IQR and positioned to have a higher life quality while as those with primary education only or no formal education, more prone to a larger IQR spread suggesting more people experiencing lower life qualities in those categories as well as high. Interestingly, there are outliers in individuals with bachelors or post secondary non-tertiary education suggesting that there could be people with degrees however unable to achieve higher life quality maybe due to of lack of access to jobs, not enough specialisation in their degrees.

```
ggplot(survey.data, aes(x=survey.predictors$life.qual, y=survey.predictors$secure_cat, fill = survey.pr
  geom_boxplot() +
  xlab("Life Quality") +
  ylab("Security categories") +
  scale_fill_brewer(palette = "Spectral") +
  theme(legend.position = "None")
```



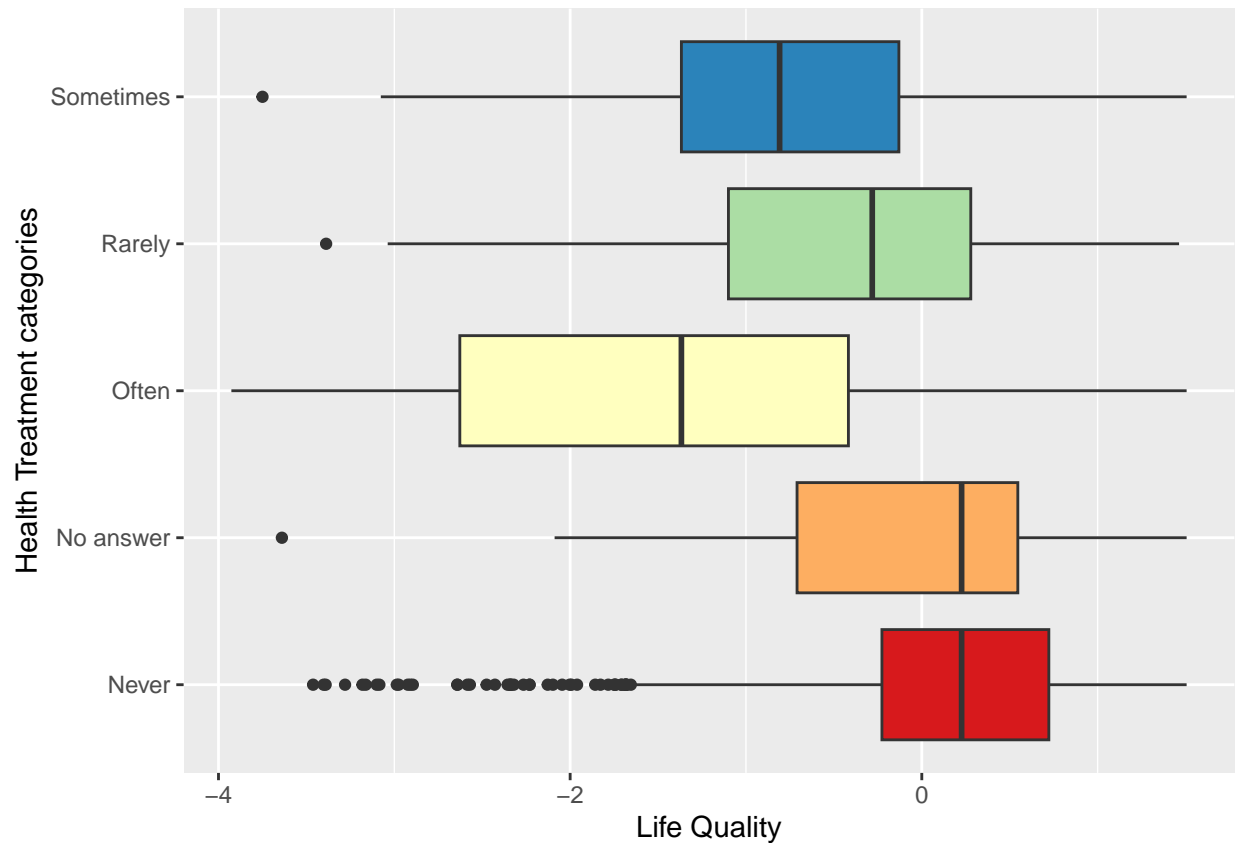
Security demonstrates a strong positive relationship with life quality, those who are not feeling secure tend to have lower life quality while as those who are very secure tend to have higher life quality, aligning with our theory described in lab 1. Interestingly, those who are very secure, have a positive IQR, demonstrating a high proportion with positive life qualities while as those who are not at all secure have a large IQR below 0.

```
ggplot(survey.data, aes(x=survey.predictors$life.qual, y=survey.predictors$rights_cat, fill = survey.pr
  geom_boxplot() +
  xlab("Life Quality") +
  ylab("Human Rights categories") +
  scale_fill_brewer(palette = "Spectral") +
  theme(legend.position = "None")
```



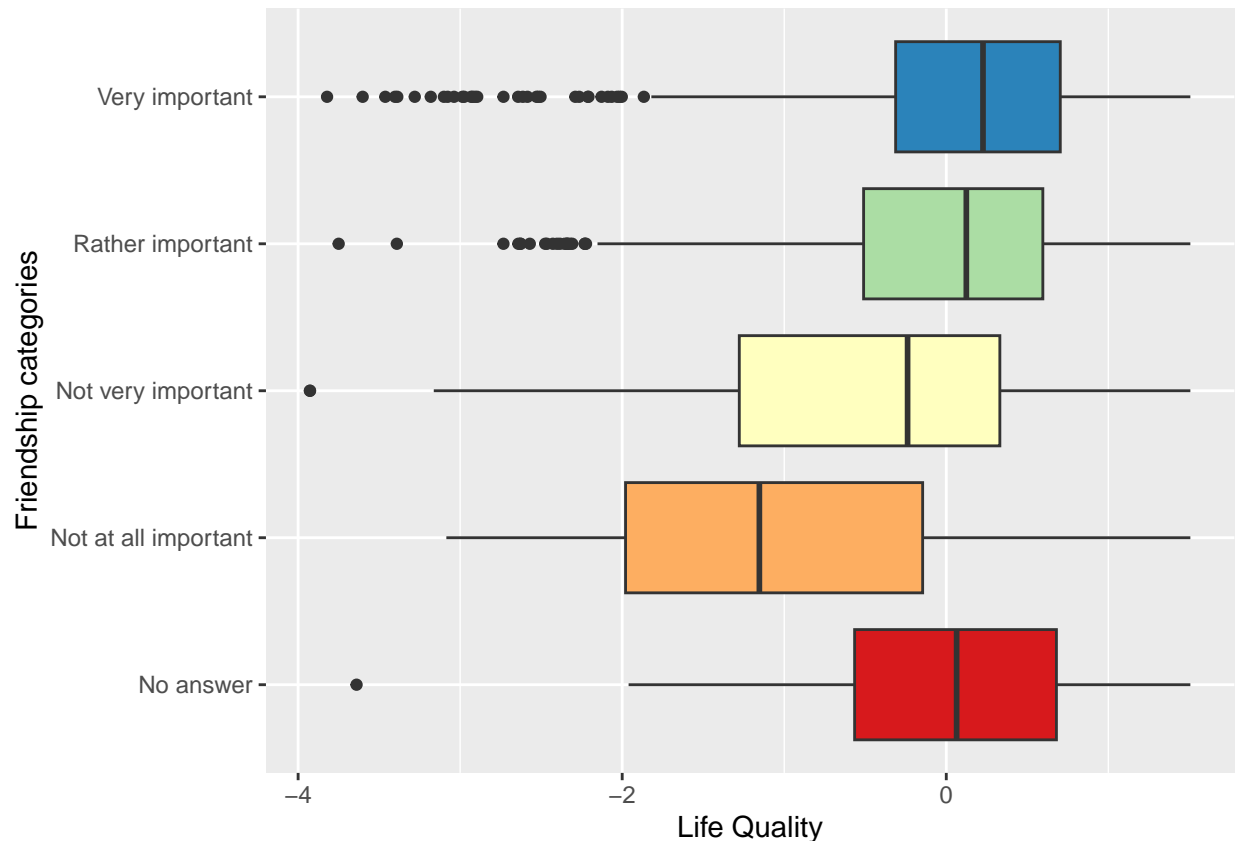
Human rights shows an interesting relationship to life quality. Those with a great deal of respect tend to have higher life quality in comparison to no respect at all, aligning with our theory. It is interesting that with no human respect at all, there can still be values in the positive life quality with Q3 reaching positive life quality values and the whisker on the right still reaching high life quality values.

```
ggplot(survey.data, aes(x=survey.predictors$life.qual, y=survey.predictors$health_cat, fill = survey.pr
  geom_boxplot() +
  xlab("Life Quality") +
  ylab("Health Treatment categories") +
  scale_fill_brewer(palette = "Spectral") +
  theme(legend.position = "None")
```



The positive skew of the Never category makes sense as those with higher life quality will never go without receiving treatment they needed and thus never have to suffer being ill which could dampen life quality. This is further represented in the lowest box plot category of Often, suggesting that those who require health treatment are lacking required treatment and are often ill, having a lower life quality. Similarly, the category of Sometimes has a slightly higher IQR range and then Rarely has an even higher IQR range. This relationship represents that those with better health tend to have a higher life quality and when there is a lack of access to medical resources, health levels could be diminishing leading to lower life quality.

```
ggplot(survey.data, aes(x=survey.predictors$life.qual, y=survey.predictors$social_cat, fill = survey.pr
  geom_boxplot() +
  xlab("Life Quality") +
  ylab("Friendship categories") +
  scale_fill_brewer(palette = "Spectral") +
  theme(legend.position = "None")
```



Friendship is obviously an important factor to life quality with those who deem friendship as Very Important having the greatest positive skew in the boxplot. Rarely Important has not much of a lower IQR but Not very important and Not at all important highlight the decrease on life quality and the significance of this social factor on an individuals life quality.

## Regression Model

In our dataset, certain variables fall into an ordinal category, which means they are ranked on a scale that assesses concepts, such as “perceived security.” These ordinal variables are coded numerically to reflect a spectrum where lower numbers denote a higher sense of security, and higher numbers correspond to a lower sense of security. However, it is difficult to compare variables since the ordinal scale doesn’t have the same impact per response value and some variables have larger amounts of response values than others. We can standardise these variables by subtracting the mean and dividing by the standard deviation.

In contrast, variables that capture education and employment status lack a natural, hierarchical structure; for instance, being self-employed isn’t qualitatively superior to being employed full-time. Given their nominal characteristics, these variables will be incorporated into our regression model as categorical factors, acknowledging the absence of a rank order among the categories.

```
##Dropping the non answers as this will no help our regression model as it provides no information.
survey.predictors2 = drop_na(survey.predictors, life.qual)
survey.predictors2 <- survey.predictors2[!(survey.predictors2$education == -2 |
  survey.predictors2$employment == -2 |
  survey.predictors2$security == -2 |
  survey.predictors2$rights == -2 |
  survey.predictors2$health == -2 |
```



```

survey.predictors2$social == -2 ) ,]

scaled_df <- survey.predictors2 %>%
  mutate(
    scaled_education = scale(education),
    scaled_employment = scale(employment),
    scaled_security = scale(security),
    scaled_rights = scale(rights),
    scaled_health = scale(health) ,
    scaled_social = scale(social) ,
    scaled_life = scale(life.qual)
  )
# Then run the regression model on the new dataframe:
regmodel <- lm(life.qual ~ emp_cat + edu_cat + scaled_security + scaled_rights + scaled_health + scaled_
# Summarise the regression model
summary(regmodel)

```

```

##
## Call:
## lm(formula = life.qual ~ emp_cat + edu_cat + scaled_security +
##     scaled_rights + scaled_health + scaled_social, data = scaled_df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.4911 -0.3882  0.0831  0.5410  2.4398
##
## Coefficients:
##                                     Estimate
## (Intercept)                       -0.055001
## emp_catHome duties, not otherwise employed    0.058519
## emp_catOther (please specify)               -0.397207
## emp_catPart time employee (less than 30 hours a week)    0.076899
## emp_catRetired / On a pension                0.199020
## emp_catSelf-employed                      0.184175
## emp_catStudent                           -0.300322
## emp_catUnemployed                       -0.535204
## edu_catDoctoral or equivalent              0.051841
## edu_catLower secondary education (i.e. Year 9 or less)    0.077258
## edu_catMaster or equivalent                0.050261
## edu_catNo formal education                 0.168186
## edu_catPost-secondary non-tertiary education (e.g. apprenticeship or certificate) -0.002792
## edu_catPrimary education only              -0.160930
## edu_catUpper secondary education (i.e. between Year 10 and Year 12) -0.056069
## scaled_security                        -0.175845
## scaled_rights                         -0.118356
## scaled_health                          0.236538
## scaled_social                         -0.080459
##                                     Std. Error
## (Intercept)                        0.044376
## emp_catHome duties, not otherwise employed    0.098363
## emp_catOther (please specify)                0.181176

```

## emp_catPart time employee (less than 30 hours a week)	0.062766
## emp_catRetired / On a pension	0.052391
## emp_catSelf-employed	0.082495
## emp_catStudent	0.132031
## emp_catUnemployed	0.121466
## edu_catDoctoral or equivalent	0.120766
## edu_catLower secondary education (i.e. Year 9 or less)	0.096018
## edu_catMaster or equivalent	0.069853
## edu_catNo formal education	0.409225
## edu_catPost-secondary non-tertiary education (e.g. apprenticeship or certificate)	0.054922
## edu_catPrimary education only	0.180245
## edu_catUpper secondary education (i.e. between Year 10 and Year 12)	0.057674
## scaled_security	0.022054
## scaled_rights	0.021113
## scaled_health	0.021507
## scaled_social	0.020465
##	t value
## (Intercept)	-1.239
## emp_catHome duties, not otherwise employed	0.595
## emp_catOther (please specify)	-2.192
## emp_catPart time employee (less than 30 hours a week)	1.225
## emp_catRetired / On a pension	3.799
## emp_catSelf-employed	2.233
## emp_catStudent	-2.275
## emp_catUnemployed	-4.406
## edu_catDoctoral or equivalent	0.429
## edu_catLower secondary education (i.e. Year 9 or less)	0.805
## edu_catMaster or equivalent	0.720
## edu_catNo formal education	0.411
## edu_catPost-secondary non-tertiary education (e.g. apprenticeship or certificate)	-0.051
## edu_catPrimary education only	-0.893
## edu_catUpper secondary education (i.e. between Year 10 and Year 12)	-0.972
## scaled_security	-7.974
## scaled_rights	-5.606
## scaled_health	10.998
## scaled_social	-3.932
##	Pr(> t )
## (Intercept)	0.215367
## emp_catHome duties, not otherwise employed	0.551978
## emp_catOther (please specify)	0.028496
## emp_catPart time employee (less than 30 hours a week)	0.220694
## emp_catRetired / On a pension	0.000151
## emp_catSelf-employed	0.025716
## emp_catStudent	0.023060
## emp_catUnemployed	1.12e-05
## edu_catDoctoral or equivalent	0.667785
## edu_catLower secondary education (i.e. Year 9 or less)	0.421157
## edu_catMaster or equivalent	0.471927
## edu_catNo formal education	0.681137
## edu_catPost-secondary non-tertiary education (e.g. apprenticeship or certificate)	0.959456
## edu_catPrimary education only	0.372078
## edu_catUpper secondary education (i.e. between Year 10 and Year 12)	0.331107
## scaled_security	2.91e-15
## scaled_rights	2.44e-08

```

## scaled_health < 2e-16
## scaled_social 8.80e-05
##
## (Intercept)
## emp_catHome duties, not otherwise employed
## emp_catOther (please specify) *
## emp_catPart time employee (less than 30 hours a week)
## emp_catRetired / On a pension ***
## emp_catSelf-employed *
## emp_catStudent *
## emp_catUnemployed ***
## edu_catDoctoral or equivalent
## edu_catLower secondary education (i.e. Year 9 or less)
## edu_catMaster or equivalent
## edu_catNo formal education
## edu_catPost-secondary non-tertiary education (e.g. apprenticeship or certificate)
## edu_catPrimary education only
## edu_catUpper secondary education (i.e. between Year 10 and Year 12)
## scaled_security ***
## scaled_rights ***
## scaled_health ***
## scaled_social ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.812 on 1597 degrees of freedom
## Multiple R-squared:  0.2381, Adjusted R-squared:  0.2295
## F-statistic: 27.72 on 18 and 1597 DF, p-value: < 2.2e-16

```

## Regression Results

**Employment:** Employment status appears to have a varied impact on quality of life. Individuals who are retired or on a pension report a higher quality of life, with a coefficient of 0.19920 and a highly significant p-value of 0.000151, implying that retirement may be associated with increased satisfaction, possibly due to more leisure time and less work-related stress. Part-time workers also report a better quality of life (coefficient: 0.076989) with a significant p-value (0.026904), which might suggest a balance between work and personal time that contributes positively to their overall well-being.

Conversely, unemployment is associated with a lower quality of life, as indicated by a negative coefficient of -0.353024 and a very significant p-value (1.12e-05). This is understandable as unemployment can lead to financial strain, social stigma, and psychological distress, all of which can detract from one's quality of life.

The coefficient for students is negative (-0.300322) with a significant p-value (0.023060). This suggests that being a student is associated with a lower quality of life in the survey population. This finding can open up a discussion about the potential stressors associated with student life, such as academic pressure, financial difficulties due to tuition fees and living expenses, and perhaps a lack of work-life balance. It might also reflect a transitional life stage where students are still establishing their careers and social identities, which could impact their perceived quality of life.

The self-employed category also presents a negative coefficient (-0.184175) with a significant p-value (0.025716). This indicates that self-employment correlates with a somewhat lower quality of life. This result can be discussed in light of the challenges that self-employed individuals often face, such as income variability, lack of employment benefits, and the demands of managing one's own business. While self-employment can offer autonomy and flexibility, it can also come with increased responsibilities and uncertainty, which may adversely affect one's quality of life.

**Education:** The impact of education on quality of life, as per the dataset, does not reach statistical significance. This result is intriguing and could be a subject of further discussion. It might point to a potential ceiling effect where, beyond a certain level, additional education does not translate to improved quality of life. Alternatively, it could indicate that in the UK context, other factors may play more pivotal roles in influencing quality of life than educational attainment alone.

**Security:** Feeling secure is critically linked to quality of life. The dataset shows a strong negative relationship between the lack of security and quality of life, with a coefficient of -0.178545 and a very significant p-value (2.94e-15). As the survey produced results where the lower the number of the response - the more secure they are, hence represented in the negative coefficient.

**Human Rights (Respect):** Respect for human rights is significantly associated with a better quality of life (coefficient: -0.118356, p-value: 5.56e-06) the survey produced results where the lower the number in the response indicates a greater amount of respect for human rights. This association could be discussed in terms of societal factors where individuals who perceive their rights as well-respected may feel more valued and supported within their community, which can enhance their perceived quality of life.

**Health Access:** Access to healthcare is a major determinant of quality of life, as indicated by the positive coefficient of 0.236538 and a highly significant p-value (2e-16). The ability to obtain necessary medical treatment without undue hardship is clearly a key component of overall well-being. This indicated through the higher coefficient, as the survey produced results where the higher the number reported for health care access, the less likely they have been denied or unable to access needed medicine.

**Social - Value of friendships** The dataset reveals a relationship between the value placed on social connections and the quality of life with a coefficient of -0.080459 and p-value of 8.80e-05, as in the response the lower the number reported - the higher value placed on friendships. A high value on social connections, such as relationships with family and friends, is correlated with an individual's quality of life. This is reflective of the well-documented view that robust social ties are essential for psychological well-being, providing support, a sense of belonging, and contributing to an individual's identity and purpose.

## Utilising factors in the regression model

We need to ensure the baseline category is based on the biggest category and as factor to ensure the categories have encoded numerical values (or are treated as 'factors').

```
survey.predictors2$emp_cat = relevel(as.factor(survey.predictors2$emp_cat), ref = "Full time employee")
survey.predictors2$edu_cat = relevel(as.factor(survey.predictors2$edu_cat), ref = "Bachelor or equivalent")
survey.predictors2$secure_cat = relevel(as.factor(survey.predictors2$secure_cat), ref = "Quite secure")
survey.predictors2$rights_cat = relevel(as.factor(survey.predictors2$rights_cat), ref = "Some respect")
survey.predictors2$health_cat = relevel(as.factor(survey.predictors2$health_cat), ref = "Never")
survey.predictors2$social_cat = relevel(as.factor(survey.predictors2$social_cat), ref = "Very important")
```

```
# Then run the regression model on the new dataframe:
```

```
model2 <- lm(life.qual ~ emp_cat + edu_cat + secure_cat + rights_cat + health_cat + social_cat , data =
```

```
# Summarise the regression model
summary(model2)
```

```
##
```

```
## Call:
```

```
## lm(formula = life.qual ~ emp_cat + edu_cat + secure_cat + rights_cat +
##     health_cat + social_cat, data = survey.predictors2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.4385 -0.3902  0.0659  0.5338  2.4725
##
## Coefficients:
##                                     Estimate
## (Intercept)                        0.127230
## emp_catHome duties, not otherwise employed    0.053317
## emp_catOther (please specify)                -0.367373
## emp_catPart time employee (less than 30 hours a week)    0.077015
## emp_catRetired / On a pension                0.206511
## emp_catSelf-employed                        0.187866
## emp_catStudent                             -0.297466
## emp_catUnemployed                          -0.546830
## edu_catDoctoral or equivalent                0.058785
## edu_catLower secondary education (i.e. Year 9 or less)    0.079026
## edu_catMaster or equivalent                  0.045349
## edu_catNo formal education                  0.148448
## edu_catPost-secondary non-tertiary education (e.g. apprenticeship or certificate) -0.001111
## edu_catPrimary education only                -0.154311
## edu_catUpper secondary education (i.e. between Year 10 and Year 12) -0.056702
## secure_catNot at all secure                  -0.649638
## secure_catNot very secure                   -0.284554
## secure_catVery secure                       0.269112
## rights_catA great deal of respect            0.088670
## rights_catNo respect at all                  -0.493962
## rights_catNot much respect                  -0.244446
## health_catOften                            -1.037074
## health_catRarely                           -0.433090
## health_catSometimes                         -0.662609
## social_catNot at all important               -0.158699
## social_catNot very important                 -0.429653
## social_catRather important                  -0.095712
##                                     Std. Error
## (Intercept)                        0.053581
## emp_catHome duties, not otherwise employed    0.098723
## emp_catOther (please specify)                0.181541
## emp_catPart time employee (less than 30 hours a week)    0.062854
## emp_catRetired / On a pension                0.052575
## emp_catSelf-employed                        0.082591
## emp_catStudent                             0.132286
## emp_catUnemployed                          0.122575
## edu_catDoctoral or equivalent                0.120864
## edu_catLower secondary education (i.e. Year 9 or less)    0.096104
## edu_catMaster or equivalent                  0.070019
## edu_catNo formal education                  0.409240
## edu_catPost-secondary non-tertiary education (e.g. apprenticeship or certificate) 0.055009
## edu_catPrimary education only                0.180893
## edu_catUpper secondary education (i.e. between Year 10 and Year 12) 0.057739
## secure_catNot at all secure                  0.189129
## secure_catNot very secure                   0.062678
```

## secure_catVery secure	0.054932
## rights_catA great deal of respect	0.047254
## rights_catNo respect at all	0.140679
## rights_catNot much respect	0.062071
## health_catOften	0.213832
## health_catRarely	0.068805
## health_catSometimes	0.078544
## social_catNot at all important	0.587225
## social_catNot very important	0.103783
## social_catRather important	0.042565
##	t value
## (Intercept)	2.375
## emp_catHome duties, not otherwise employed	0.540
## emp_catOther (please specify)	-2.024
## emp_catPart time employee (less than 30 hours a week)	1.225
## emp_catRetired / On a pension	3.928
## emp_catSelf-employed	2.275
## emp_catStudent	-2.249
## emp_catUnemployed	-4.461
## edu_catDoctoral or equivalent	0.486
## edu_catLower secondary education (i.e. Year 9 or less)	0.822
## edu_catMaster or equivalent	0.648
## edu_catNo formal education	0.363
## edu_catPost-secondary non-tertiary education (e.g. apprenticeship or certificate)	-0.020
## edu_catPrimary education only	-0.853
## edu_catUpper secondary education (i.e. between Year 10 and Year 12)	-0.982
## secure_catNot at all secure	-3.435
## secure_catNot very secure	-4.540
## secure_catVery secure	4.899
## rights_catA great deal of respect	1.876
## rights_catNo respect at all	-3.511
## rights_catNot much respect	-3.938
## health_catOften	-4.850
## health_catRarely	-6.295
## health_catSometimes	-8.436
## social_catNot at all important	-0.270
## social_catNot very important	-4.140
## social_catRather important	-2.249
##	Pr(> t )
## (Intercept)	0.017689
## emp_catHome duties, not otherwise employed	0.589228
## emp_catOther (please specify)	0.043175
## emp_catPart time employee (less than 30 hours a week)	0.220648
## emp_catRetired / On a pension	8.94e-05
## emp_catSelf-employed	0.023060
## emp_catStudent	0.024670
## emp_catUnemployed	8.72e-06
## edu_catDoctoral or equivalent	0.626771
## edu_catLower secondary education (i.e. Year 9 or less)	0.411034
## edu_catMaster or equivalent	0.517300
## edu_catNo formal education	0.716846
## edu_catPost-secondary non-tertiary education (e.g. apprenticeship or certificate)	0.983891
## edu_catPrimary education only	0.393761
## edu_catUpper secondary education (i.e. between Year 10 and Year 12)	0.326231

```

## secure_catNot at all secure                                0.000608
## secure_catNot very secure                                6.05e-06
## secure_catVery secure                                    1.06e-06
## rights_catA great deal of respect                        0.060777
## rights_catNo respect at all                              0.000458
## rights_catNot much respect                               8.57e-05
## health_catOften                                          1.36e-06
## health_catRarely                                         3.98e-10
## health_catSometimes                                     < 2e-16
## social_catNot at all important                           0.787001
## social_catNot very important                             3.66e-05
## social_catRather important                               0.024675
##
## (Intercept)                                             *
## emp_catHome duties, not otherwise employed
## emp_catOther (please specify)                           *
## emp_catPart time employee (less than 30 hours a week)
## emp_catRetired / On a pension                           ***
## emp_catSelf-employed                                    *
## emp_catStudent                                           *
## emp_catUnemployed                                       ***
## edu_catDoctoral or equivalent
## edu_catLower secondary education (i.e. Year 9 or less)
## edu_catMaster or equivalent
## edu_catNo formal education
## edu_catPost-secondary non-tertiary education (e.g. apprenticeship or certificate)
## edu_catPrimary education only
## edu_catUpper secondary education (i.e. between Year 10 and Year 12)
## secure_catNot at all secure                               ***
## secure_catNot very secure                                ***
## secure_catVery secure                                    ***
## rights_catA great deal of respect                        .
## rights_catNo respect at all                              ***
## rights_catNot much respect                               ***
## health_catOften                                          ***
## health_catRarely                                         ***
## health_catSometimes                                     ***
## social_catNot at all important
## social_catNot very important                             ***
## social_catRather important                               *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8114 on 1589 degrees of freedom
## Multiple R-squared:  0.2431, Adjusted R-squared:  0.2307
## F-statistic: 19.63 on 26 and 1589 DF,  p-value: < 2.2e-16

```

The intercept value becomes the mean value when all of our other predictors are held at the baseline category of full time employed, holding a bachelor degree or equivalent, feeling quite secure, having some human rights respected, never having to go without health treatment in the last 12 months and valuing friendship as very important. The estimated life quality is 0.127 at this baseline category.

**Effects** While controlling for other variables in the model the coefficients can now describe the change of a particular category on the life quality of an individual.

The significant variables described by \*\*\* in our model include

Employment, especially retired/on a pension with an associated 0.21 units increase in life quality suggesting that individuals who are retired (rather than fully employed) have a less stressful lifestyle and more leisure time.

Unemployment as well, with decrease 0.55 units in life quality demonstrates the significant financial strain or lack of purpose which can impact life quality.

Security has an overall significant effect to life quality. Being very secure is a significant increase in life quality of 0.27 units, suggesting that individuals may feel less stress on their lives financially or physical wellbeing leading to better life quality. Being not at all secure in comparison to quite secure has a significant decreasing effect on life quality of 0.65 units and indicates that lower life quality is found when individuals don't have security whether financially, physically or emotionally. This could be due to stress or anxiety as a result.

Similarly, not having any respect for rights in comparison to having some human rights show a diminishing effect on life quality of around 0.5 units as individuals may feel oppressed and contribute to feelings of injustice and frustration. Interestingly, both not much respect and no respect at all in are significant measures that impact life quality in comparison to a great deal of respect, demonstrating that even having some respect can not make a great deal of changes in one's life quality.

All health categories in comparison to never going without required health treatment demonstrate decreasing effects on life quality with Often, Rarely, Sometimes demonstrating decreases of 1.04, 0.43 and 0.67 respectively. Health is a fundamental pillar of well being and any challenges in this area significantly impact overall life quality and the ability for one to live a satisfying life.

Finally, friendship is significant to life quality when an individual finds friendship NOT very important in comparison to very important. The significant decrease of 0.43 indicates that individuals who do not find social relationships tend to have lower life quality. Strong social connections are often associated with support and good mental wellbeing so this relationship also makes sense.

From these coefficients, it is interesting to note that the lowest value of any variable's ordinal scale shows significant effects on people's life quality, creating diminishing effects due to lacks of certain factors in people's life styles. It also highlights the importance of factors such as financial security, social connections, physical health and recognition of individual rights, with a lacking of any of these factors bringing a significant effect to quality of life.

**Model fit** The adjusted r-squared demonstrates that this model is perhaps not a good fit with the model only explaining 23% of variance. The F statistic and low p-value indicates that at least one of the predictors is significantly related to the dependent variable as discussed earlier.

## Validating Factor analysis

Factor analysis has assumptions including:

- Underlying latent trait
- Items are continuous measures (or conceptualised as continuous)
- Correlations are linear
- There are no outliers
- There is adequate data



To validate this dependent variable, it is assessed using validity and reliability where reliability is the measure of the latent trait with the least measurement error and validity is whether the measure actually represent what its supposed to.

To measure reliability, we examine the proportion of the variance of the predictor that is account for by variance in the latent variable. In World Values survey data, there is other variables expected to be associated with the latent variable. For example, Q56. Comparing your standard of living with your parent's standard of living when they were about your age, would you say that you are better off, worse off, or about the same?

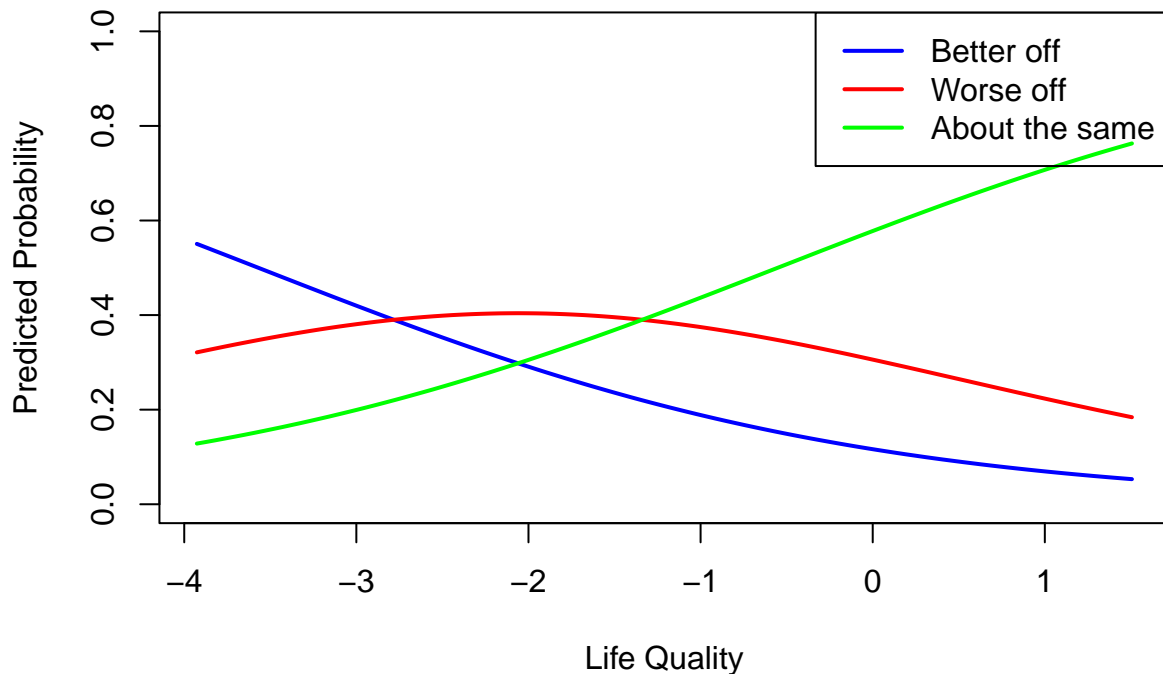
```
result <- polr(formula = Q56 ~ life.quality, data = survey.data %>% mutate(Q56 = factor(Q56, levels = c(
result
```

```
## Call:
## polr(formula = Q56 ~ life.quality, data = survey.data %>% mutate(Q56 = factor(Q56,
##     levels = c("2", "3", "1"))))
##
## Coefficients:
## life.quality
## 0.5678809
##
## Intercepts:
##      2|3      3|1
## -2.0269912 -0.3135848
##
## Residual Deviance: 3166.618
## AIC: 3172.618
## (5 observations deleted due to missingness)
```

The results of the fitted proportional odds log regression model is presented above. The coefficient of 0.5679 is positive and indicates higher values of life quality associated with higher odds of moving to higher categories in Q56.

```
life_quality_seq <- seq(min(survey.data$life.quality), max(survey.data$life.quality), length.out = 100)
predicted_probs <- predict(result, data.frame(life.quality = life_quality_seq), type = "probs")
plot(life_quality_seq, predicted_probs[, 1], type = "l", ylim = c(0, 1), xlab = "Life Quality", ylab = "
lines(life_quality_seq, predicted_probs[, 2], type = "l", col = "red", lwd = 2)
lines(life_quality_seq, predicted_probs[, 3], type = "l", col = "green", lwd = 2)
legend("topright", legend = c("Better off", "Worse off", "About the same"), col = c("blue", "red", "gre
```

## Predicted Probability of Q56



We'd assume people with high quality of life, they should be the same or higher quality than their parents. This relationship can be seen in the graph as the quality of life is better than parents increases dramatically with quality of life. Those with lower life quality demonstrate low predicted probabilities below 20% to say they are living off better than their parents. This strong relationship is a good demonstration of reliability

Inter-item reliability: the consistency between multiple items measuring the same construct, measured using Cronbach's alpha.

```
alpha(survey.data[,c("happiness",
                      "health",
                      "finances",
                      "satisfaction",
                      "freedom")])
```

```
##
## Reliability analysis
## Call: alpha(x = survey.data[, c("happiness", "health", "finances",
##   "satisfaction", "freedom")])
##
##      raw_alpha std.alpha G6(smc) average_r S/N      ase mean   sd median_r
##      0.77      0.81      0.79      0.46 4.3 0.0069  5.8 1.2      0.44
##
##      95% confidence boundaries
##      lower alpha upper
```

```

## Feldt      0.75  0.77  0.78
## Duhachek   0.75  0.77  0.78
##
## Reliability if an item is dropped:
##           raw_alpha std.alpha G6(smc) average_r S/N alpha se  var.r med.r
## happiness      0.76      0.78   0.75      0.47 3.6   0.0079 0.0175  0.46
## health         0.76      0.81   0.78      0.52 4.3   0.0075 0.0122  0.53
## finances       0.74      0.79   0.76      0.48 3.7   0.0074 0.0138  0.44
## satisfaction   0.63      0.72   0.66      0.39 2.5   0.0111 0.0027  0.39
## freedom       0.69      0.77   0.74      0.46 3.3   0.0088 0.0138  0.43
##
## Item statistics
##           n raw.r std.r r.cor r.drop mean  sd
## happiness  1758  0.63  0.74  0.65  0.56  3.2 0.61
## health     1758  0.57  0.67  0.53  0.46  3.9 0.85
## finances   1758  0.81  0.73  0.62  0.58  6.7 2.26
## satisfaction 1758  0.88  0.87  0.87  0.78  7.5 1.74
## freedom    1758  0.81  0.77  0.70  0.63  7.7 1.92
##
## Non missing response frequency for each item
##           1  2  3  4  5  6  7  8  9  10 miss
## happiness  0.01 0.07 0.61 0.31 0.00 0.00 0.00 0.00 0.00 0.00 0
## health     0.01 0.04 0.20 0.49 0.26 0.00 0.00 0.00 0.00 0.00 0
## finances   0.03 0.03 0.05 0.06 0.11 0.12 0.18 0.21 0.13 0.09 0
## satisfaction 0.01 0.01 0.02 0.02 0.08 0.08 0.17 0.32 0.19 0.10 0
## freedom    0.01 0.01 0.02 0.02 0.08 0.07 0.16 0.28 0.15 0.19 0

```

Cronbach's alpha coefficient indicates the reliability with a value closer to 1 indicating higher reliability. Here the raw cronbach alpha is 0.77 (2dp) and the individual raw alphas are all about 0.7 or higher suggesting that there is consistency of each item.