Week 2 Report

Diana Hilleshein

1/24/2022

General information on the data

Data was downloaded from: https://www.europeansocialsurvey.org/download.html?file=ESS9e03_1&y=
2018.
Codebook fro the data can be found by the following link: chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/
viewer.html?pdfurl=https%3A%2F%2Fwww.europeansocialsurvey.org%2Fdocs%2Fround9%2Fsurvey%
2FESS9_appendix_a7_e03_1.pdf&clen=2111156.

The European Social Survey (ESS) is an academically-driven multi-country survey, which has been admin-
istered in over 35 countries to date. Its three aims are, firstly – to monitor and interpret changing public
attitudes and values within Europe and to investigate how they interact with Europe's changing institutions,
secondly – to advance and consolidate improved methods of cross-national survey measurement in Europe
and beyond, and thirdly – to develop a series of European social indicators, including attitudinal indicators.
For the course we will use round 9 ESS.

Variables of the datases are separated for the following groups:
Country
Weights
Media and social trust
Politics
Subjective well-being, social exclusion, religion, national and ethnic identity
Timing of life
Gender, Year of birth and Household grid
Socio-demographics
Justice and Fairness
Human values
Administrative variables
Sample Design variables
User defined variables

Part 1. Explore the European Social Survey (ESS) data.

I am interested in the following self-made groups of variables:
Involovemt in politics
1) *polintr* -How interested in politics (scale from 1 to 4, 1= Very interested, 1 = Not at all, 4 = Very
interested)
2) *psppipla* -Political system allows people to have influence on politics (scale from 1 to 5, 1 = Not at all
able, 5 = A great deal)
Attitude to immigrants
3) *imwbcnt* -Immigrants make country worse or better place to live (scale from 0 to 10, 0 = Worse place to
live, 10 = Better place to live)
4) *imbgeco* -Immigration bad or good for country's economy (scale from 0 to 10, 0 = Bad for the economy,

10 = Good for the economy) 5) *imueclt* -Country's cultural life undermined or enriched by immigrants (scale from 0 to 10, 0 = Cultural life undermined, 10 = Cultural life enriched)

6) *imsmetn* -Allow many/few immigrants of same race/ethnic group as majority (scale from 1 to 4, 1 = Allow many to come and live here, 4 = Allow none)

7) *imdfetn* -Allow many/few immigrants of different race/ethnic group from majority (scale from 1 to 4, 1 = Allow many to come and live here, 4 = Allow none)

8) *impcntr* -Allow many/few immigrants from poorer countries outside Europe (scale from 1 to 4, 1 = Allow many to come and live here, 4 = Allow none)

General information on a participants

9) *stflife* -How satisfied with life as a whole (from 0 to 10, 0 = extremely dissatisfied, 10 = extremely satisfied)

10) *gndr*-Gender (1 = Male, 2 = Female)

11) *agea* -Age of respondent, calculated (integer)

Comments on the variables: - there is just one continuous variable that is agea, while gender is binary. The rest of the variables are categorical.

I will display descriptive statistics by variable groups: *General information on a participants*

```r
#load necessary libraries
library(ggplot2)
library(GGally)
library(haven)
library(tidyverse)
library(corrplot)
library(tidyverse)
library(dplyr)


#make one dataframe that contains all the cat variables. Dont do the same for age and gender since ther
data <- as.data.frame(data)
interested_var_cat <- select(data, "stflife", "psppipla", "polintr", "imwbcnt", "imbgeco", "imueclt", ":
data$gndr = as.factor(data$gndr)
interested_var <- select(data, "stflife", "psppipla", "polintr", "imwbcnt", "imbgeco", "imueclt", "imsm
# "gndr"
# agea

gather(interested_var_cat) %>% ggplot(aes(value)) + geom_bar(col ="dark red", fill = "dark red", alpha :
```
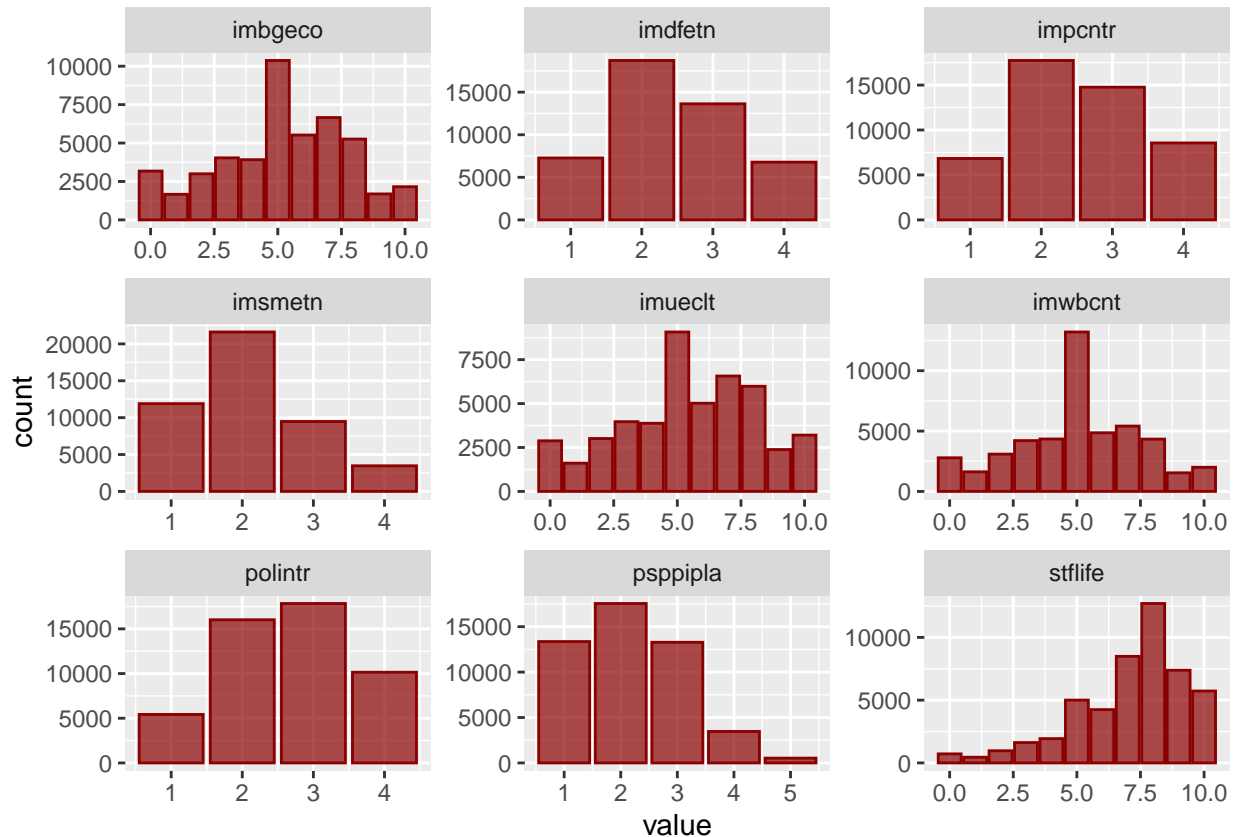
Bar plot indicates that: - there are people who thinks that political system doesnt `allow people to have influence on politics or allows on insufficient level` (*psppipla*). While participants are rather interested in politics that not interested (*polintr*); - in majority people are satisfied with life as whole (*stflife*); - participants tend to be neutral about their opinion regarding immigrants influence on a country in general as well and in economic and cultural seance (*imwbcnt*, *imbgeco*, *imueclt*). However there are slightly more poeple who think of positive influence of immigrants. - participants tend to be more acceptable towards immigrants of similar ethnicity (*imsmetn* vs *imdfetn*), and tend to accept the least immigrants from poorer countries outside Europe (*impcntr*). The level of acceptance can be pictured as following: same ethnicity > different ethnicity > from poorer countries out of Europe.

Next, distinctive statistics for age and gender. As we can see there are slightly more women who participates in the survey.

Average age of participants is 51 years. People from 15 to 90 years old took part in the survey. 50% of all ages are from approximetely 37 to 67 years.

```
#rstudio was complaining that gndr was a list, so I unlisted it
data$gndr <- unlist(data$gndr)
#summary fo rgndr
summary(data$gndr)
```

```
##     1     2
## 23020 26499
```

```
#some vizualization to make the eyes happy
col <- c(rep("dark blue", 1), rep("dark red", 1))
ggplot(data = data, aes(data$gndr)) + geom_histogram(col = col, fill = col, alpha = 0.7, stat="count")
```

3
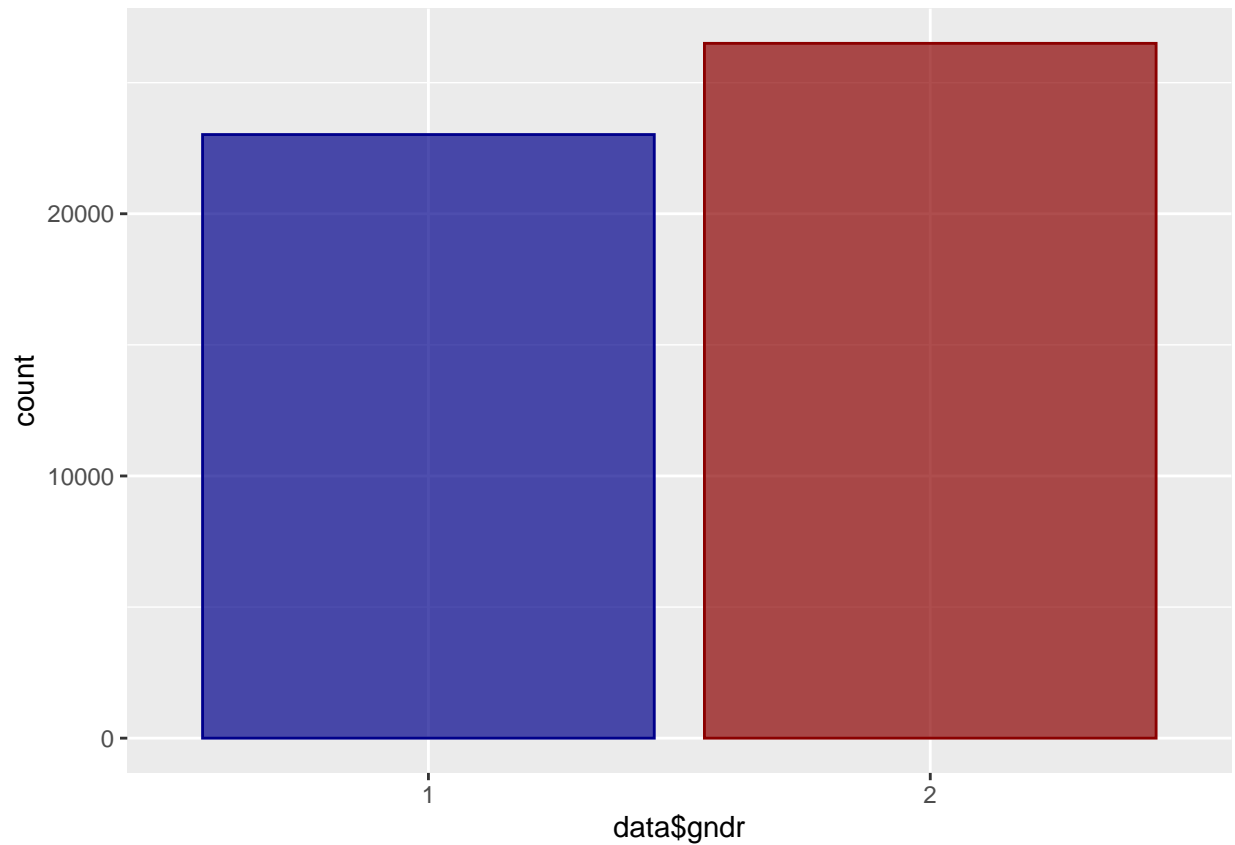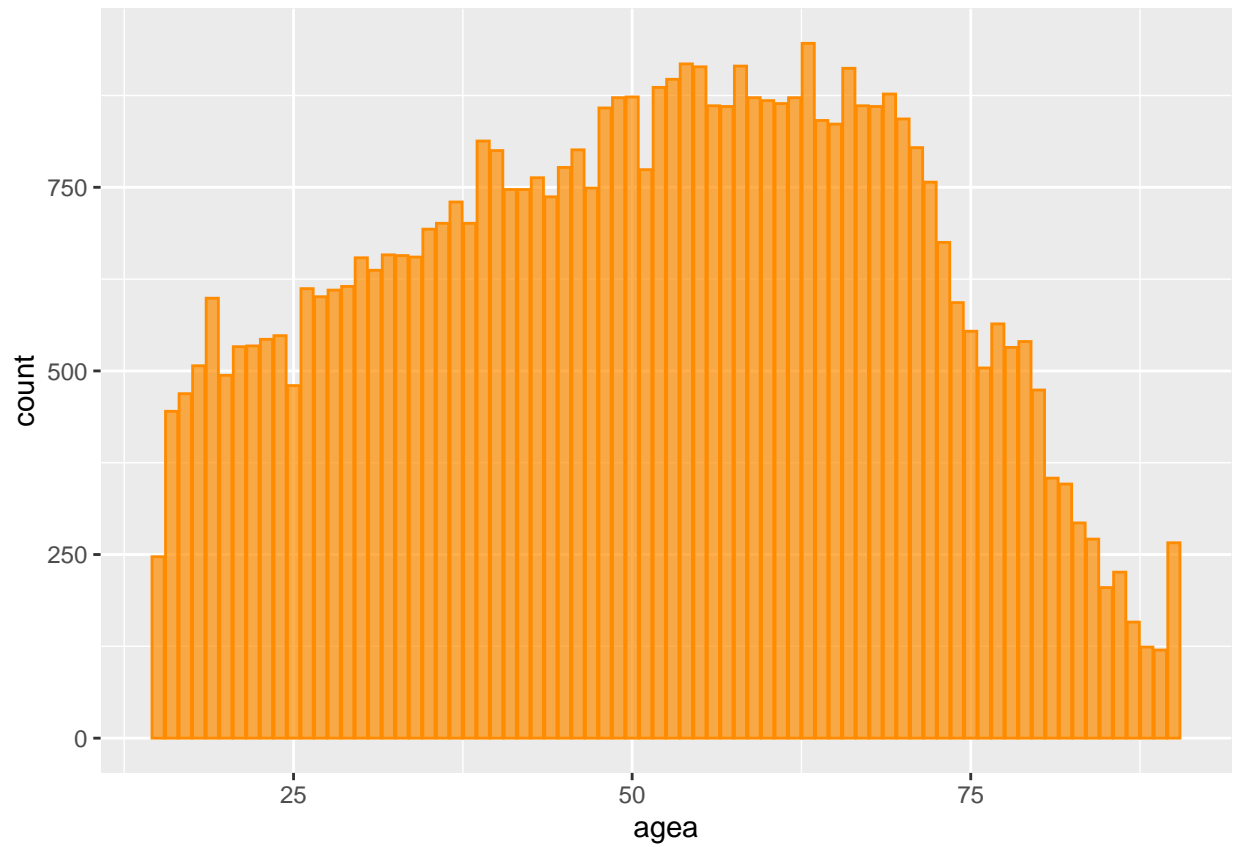
```r
#descriptives for agea
summary(data$agea)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   15.00   36.00   52.00   51.07   66.00   90.00     222
```
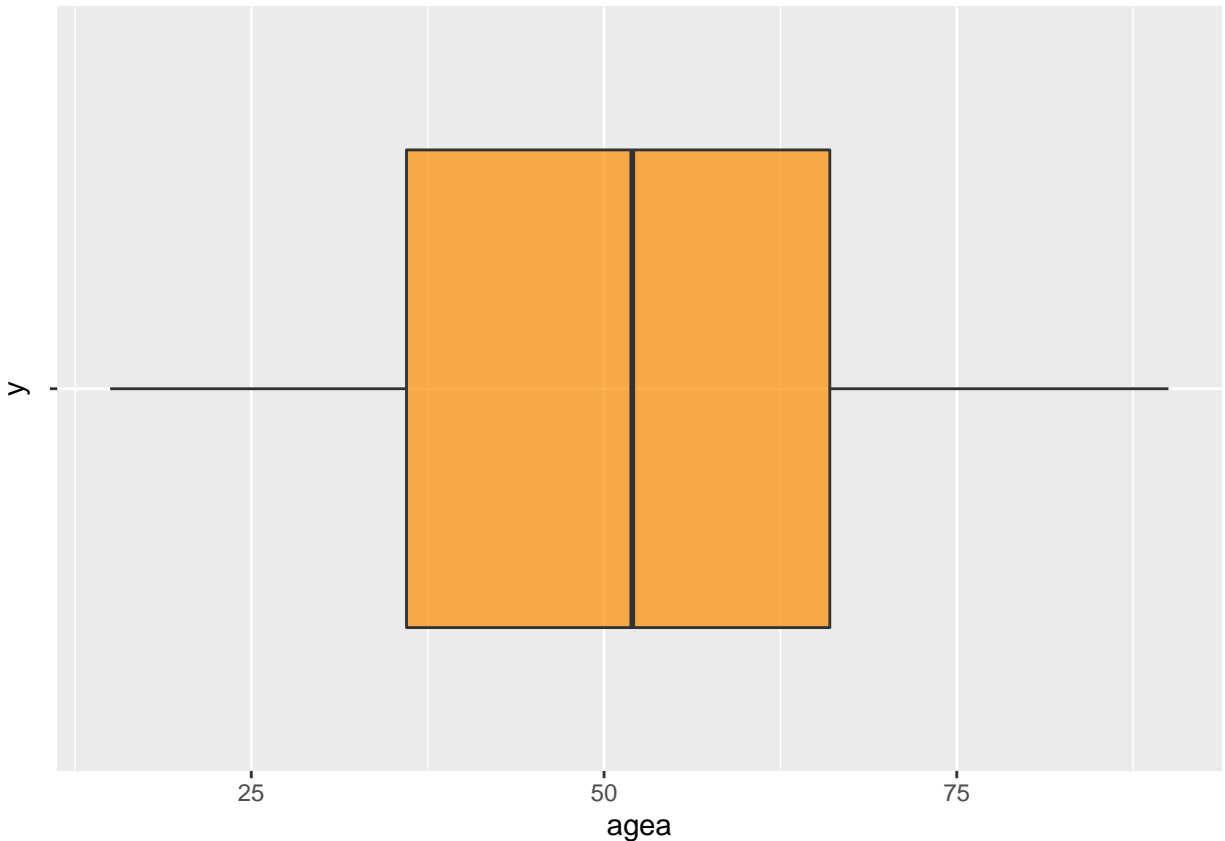
```r
#vizualization to make the eyes happy again
ggplot(data = data, aes(agea)) + geom_histogram(breaks = seq(15,90,1), col = "dark orange", fill = "dar
```

```
col2 = rep("dark orange", alpha = 0.7, stat="count")

ggplot(data = data, aes(agea, y = " ")) + geom_boxplot(fill = "dark orange", alpha = 0.7)
```

Part 2. Copy a small text file data (week2data) from Moodle

```
#load the data for part 2 ex
library(readxl)
part2_data <- read_excel("C:/R scripts/DWA2022/part_two_data.xlsx")

View(part2_data) #good, labes are in the table!
save(part2_data,file="part2_data.csv")
```

Task 1: Add variable labels: id = "ID number", agea = "Age of respondent", q1 = "How often you read newspaper?", q2 ="Government should reduce differences in income levels?"

Try 1

```
#install.packages("expss")
library(labelled)
var_label(part2_data$id) <- "ID number"
var_label(part2_data$agea) <- "Age of respondent"
var_label(part2_data$q1) <- "How often you read newspaper?"
var_label(part2_data$q2) <- "Government should reduce differences in income levels?"
View(part2_data) #labels are in the table
```

Adding variable labels work easy and well!

Task 2: Add value labels to variables q1 and q2
Q1: 1 = "Never", 2 = "Only occasionally", 3 = "A few times a week", 4 = "Most days", 5 = "Every day, 8 ="Don't know", 9 = "No answer"
Q2: 1 = "Agree strongly", 2 = "Agree", 3 = "Neither agree nor disagree", 4 = "Disagree", 5 = "Disagree strongly", -1 = "missing".

```
library(expss)
library(sjlabelled)
part2_data$q1 <- factor(part2_data$q1,
                        levels = c(1, 2, 3, 4, 5, 8, 9),
                        labels = c("Never", "Only occasionally", "A few times a week", "Most days", "E
View(part2_data) #I can see the labels fro q1 in the table

part2_data$q2 <- factor(part2_data$q2,
                        levels = c(1, 2, 3, 4, 5, -1),
                        labels = c("Agree strongly", "Agree", "Neither agree nor disagree", "Disagree"

View(part2_data) # can see the labels for values of q1 and q2 in the table, but there is no labels for

get_label(part2_data$q1) #no label
```

```
## NULL
```

```
get_labels(part2_data$q1)
```

```
## [1] "Never"              "Only occasionally"  "A few times a week"
## [4] "Most days"          "Every day"          "Don't know"
## [7] "No answer"
```

```
get_label(part2_data$q2) #no label
```

```
## NULL
```

```
get_labels(part2_data$q2)
```

```
## [1] "Agree strongly"             "Agree"
## [3] "Neither agree nor disagree" "Disagree"
## [5] "Disagree strongly"          "missing"
```

Here I have got a problem: when I applied labels to values of q1 and q2, the labels of variables q1 and q2 disappeared.

Try 2

One way to sole the problem is to, firstly, name values and after that variables like following:

```
part2_data$q1 <- factor(part2_data$q1,
                        levels = c(1, 2, 3, 4, 5, 8, 9),
                        labels = c("Never", "Only occasionally", "A few times a week", "Most days", "E
View(part2_data) #I can see the labels fro q1 in the table

part2_data$q2 <- factor(part2_data$q2,
                        levels = c(1, 2, 3, 4, 5, -1),
                        labels = c("Agree strongly", "Agree", "Neither agree nor disagree", "Disagree"

library(labelled)
var_label(part2_data$id) <- "ID number"
var_label(part2_data$agea) <- "Age of respondent"
var_label(part2_data$q1) <- "How often you read newspaper?"
var_label(part2_data$q2) <- "Government should reduce differences in income levels?"
View(part2_data) #all labels are in the table
get_label(part2_data$q1)
```

```
## [1] "How often you read newspaper?"
```

```
get_labels(part2_data$q1)
```

```
## [1] "Never"              "Only occasionally"  "A few times a week"
## [4] "Most days"          "Every day"          "Don't know"
## [7] "No answer"
```

```
get_label(part2_data$q2)
```

```
## [1] "Government should reduce differences in income levels?"
```

```
get_labels(part2_data$q2)
```

```
## [1] "Agree strongly"             "Agree"
## [3] "Neither agree nor disagree" "Disagree"
## [5] "Disagree strongly"          "missing"
```

It worked and we can see both labels for variables and values. But this was seems to be not cool.

Try 3

I tried to google other ways. Found this way here: https://cran.r-project.org/web/packages/expss/vignettes/labels-support.html.

```
library(sjlabelled)
#Firstly, add variable labels
var_label(part2_data$id) <- "ID number"
var_label(part2_data$agea) <- "Age of respondent"
var_label(part2_data$q1) <- "How often you read newspaper?"
var_label(part2_data$q2) <- "Government should reduce differences in income levels?"
```

```r
library(expss)
#Secondly, add values of q1 and q2 labels
val_lab(part2_data$q1) = num_lab("
1 Never
2 Only ocassionally
3 A few times a week
4 Most days
5 Every day
8 Don't know
9 No answer
  ")

val_lab(part2_data$q2) = num_lab("
1 Agree strongly
2 Agree
3 Neither agree nor disagree
4 Disagree
5 Disagree strongly
-1 missing
")

View(part2_data) #no values labels
print(val_lab(part2_data$q1)) #labels are here
```

```
##              Never  Only ocassionally A few times a week          Most days
##                  1                  2                  3                  4
##          Every day          Don't know          No answer
##                  5                  8                  9
```

```r
get_label(part2_data$q1)
```

```
## [1] "How often you read newspaper?"
```

```r
get_labels(part2_data$q1)
```

```
## [1] "Never"            "Only ocassionally"  "A few times a week"
## [4] "Most days"        "Every day"          "Don't know"
## [7] "No answer"
```

```r
get_label(part2_data$q2)
```

```
## [1] "Government should reduce differences in income levels?"
```

```r
get_labels(part2_data$q2)
```

```
## [1] "missing"                   "Agree strongly"
## [3] "Agree"                     "Neither agree nor disagree"
## [5] "Disagree"                  "Disagree strongly"
```

Here, I could get labels for variables but not for values. Later, I found out, that labels for values are saved in val_lab(part2_data$q1), but I don't know how to apply it to the dataset. But with "get_lables" function I can see that it is considered a label for the q1 nd q2 variables.

Try 4

So, I used another tip from the same Internet page:

```
library(expss)
library(sjlabelled)
part2_data = apply_labels(part2_data,
                          id = "ID number",
                          agea = "Age of respondent",
                            q1 = "How often you read newspaper?",
                            q1 =c("Never" = 1,
                                    "Only occasionally" = 2,
                                    "A few times a week" =3,
                                    "Most days" = 4,
                                    "Every day" = 5,
                                    "Don't know" = 8,
                                    "No answer" = 9),
                          q2 = "Government should reduce differences in income levels",
                          q2 = c("Agree strongly" = 1,
                                  "Agree" = 2,
                                  "Neither agree nor disagree" = 3,
                                  "Disagree" = 4,
                                  "Disagree strongly" = 5,
                                  "missing" = -1))

View(part2_data)

get_label(part2_data$q1)
```

```
## [1] "How often you read newspaper?"
```

```
get_labels(part2_data$q1)
```

```
## [1] "Never"            "Only occasionally"  "A few times a week"
## [4] "Most days"        "Every day"          "Don't know"
## [7] "No answer"
```

```
get_label(part2_data$q2)
```

```
## [1] "Government should reduce differences in income levels"
```

```
get_labels(part2_data$q2)
```

```
## [1] "missing"              "Agree strongly"
## [3] "Agree"                "Neither agree nor disagree"
## [5] "Disagree"             "Disagree strongly"
```

```
print(part2_data)
```

```
## # A tibble: 10 x 4
##     id        agea       q1         q2
##     <labelled> <labelled> <labelled> <labelled>
##  1  1         51         1          2
##  2  2         74         3          1
##  3  3         37         1          3
##  4  4         44         8          3
##  5  5         19         5          4
##  6  6         88         3         -1
##  7  7         29         9          3
##  8  8         46         4         -1
##  9  9         53         3          4
## 10 10         21         2          2
```

It gace the same result as the previous try: the lables are saved, but not indicated in the table. Not sure if it should be like this or the value lables should be visible in the table.