

EVS 2017 Data Review

Chris Rice

2024-03-09

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.3.2
```

```
##
```

```
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
```

```
##
```

```
## filter, lag
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
## intersect, setdiff, setequal, union
```

```
library(purrr) # Make sure to load purrr
```

```
## Warning: package 'purrr' was built under R version 4.3.3
```

```
library(knitr)
```

```
## Warning: package 'knitr' was built under R version 4.3.3
```

```
library(kableExtra)
```

```
## Warning: package 'kableExtra' was built under R version 4.3.3
```

```
##
```

```
## Attaching package: 'kableExtra'
```

```
## The following object is masked from 'package:dplyr':
```

```
##
```

```
## group_rows
```

According to the website: *The European Values Study is a large-scale, cross-national and longitudinal survey research program on how Europeans think about family, work, religion, politics, and society. Repeated every nine years in an increasing number of countries, the survey provides insights into the ideas, beliefs, preferences, attitudes, values, and opinions of citizens all over Europe.*

Beginning steps were to create an account with the aforementioned website in order to extract either the SPSS or STATA file. In this case, we downloaded the STATA file in order to create an EVS 2017 analysis. Specific variables of this project's interest were as follows:

1. Respondent's age (continuous)
2. Respondent's country (categorical)
3. Respondent's education (categorical)
4. Respondent's sex (categorical)
5. Respondent's opinion on if jobs are scarce should national citizens have priority over immigrants (categorical)
6. Respondent's opinion on if a child suffers when the mother is working (categorical)

Table 1 - Categorical Variable Descriptive Statistics

```
load("C:/Users/ricecakes/Desktop/Git1/HW-2/Rcode/df1c.RData")

calculate_frequencies <- function(data, var_name) {
  data %>%
    count(!sym(var_name), name = "Frequency") %>%
    mutate(Percentage = Frequency / sum(Frequency) * 100,
           Category = as.character(!sym(var_name)),
           Variable = var_name) %>%
    select(Variable, Category, Frequency, Percentage)
}

load("C:/Users/ricecakes/Desktop/Git1/HW-2/Rcode/df1c.RData")
# List of categorical variables
categorical_vars <- c("country", "education", "child_suffers", "jobs_are_scarce")

# Calculate frequencies and percentages for each variable and bind rows together
load("C:/Users/ricecakes/Desktop/Git1/HW-2/Rcode/df1c.RData")
categorical_summary <- map_dfr(categorical_vars, ~calculate_frequencies(df1c, .x))

# Display the table
load("C:/Users/ricecakes/Desktop/Git1/HW-2/Rcode/df1c.RData")
knitr::kable(categorical_summary, booktabs = TRUE) %>%
  kable_styling(latex_options = c("striped", "scale_down")) %>%
  column_spec(1, width = "4cm") %>%
  column_spec(2, width = "4cm")
```

```
## Warning in styling_latex_scale(out, table_info, "down"): Longtable cannot be
## resized.
```

Variable	Category	Frequency	Percentage
country	Albania	1435	2.4142804
country	Armenia	1500	2.5236381
country	Austria	1644	2.7659073
country	Azerbaijan	1800	3.0283657
country	Belarus	1548	2.6043945
country	Bosnia and Herzegovina	1724	2.9005014
country	Bulgaria	1558	2.6212187
country	Croatia	1487	2.5017665
country	Czechia	1811	3.0468724
country	Denmark	3362	5.6563141

country	Estonia	1304	2.1938827
country	Finland	1199	2.0172280
country	France	1870	3.1461355
country	Georgia	2194	3.6912413
country	Germany	2170	3.6508631
country	Great Britain	1788	3.0081766
country	Hungary	1514	2.5471920
country	Iceland	1624	2.7322588
country	Italy	2277	3.8308826
country	Latvia	1335	2.2460379
country	Lithuania	1448	2.4361520
country	Montenegro	1003	1.6874727
country	Netherlands	2404	4.0445506
country	North Macedonia	1117	1.8792692
country	Norway	1122	1.8876813
country	Poland	1352	2.2746391
country	Portugal	1215	2.0441468
country	Romania	1613	2.7137521
country	Russia	1825	3.0704263
country	Serbia	1499	2.5219557
country	Spain	1209	2.0340523
country	Sweden	1194	2.0088159
country	Switzerland	3174	5.3400182
country	Ukraine	1612	2.7120697
country	NA	2507	4.2178404
education	Bachelor or equivalent	6508	10.9492244
education	Doctoral or equivalent	555	0.9337461
education	Less than primary	510	0.8580369
education	Lower secondary	8588	14.4486692
education	Master or equivalent	8397	14.1273260
education	Post-secondary non tertiary	2705	4.5509607
education	Primary	3028	5.0943841
education	Short-cycle tertiary	4553	7.6600828
education	Upper secondary	24121	40.5817827
education	dont know	81	0.1362765
education	no answer	308	0.5181870
education	other	84	0.1413237
child_suffers	agree	16484	27.7331000
child_suffers	agree strongly	5556	9.3475554
child_suffers	disagree	25122	42.2658905
child_suffers	disagree strongly	10918	18.3687203
child_suffers	dont know	1158	1.9482486
child_suffers	no answer	200	0.3364851
jobs_are_scarce	agree	16691	28.0813621
jobs_are_scarce	agree strongly	20308	34.1666947
jobs_are_scarce	disagree	8263	13.9018810
jobs_are_scarce	disagree strongly	4309	7.2495710
jobs_are_scarce	dont know	622	1.0464686
jobs_are_scarce	multiple answers Mail	1	0.0016824

jobs_are_scarce	neither agree nor disagree	9062	15.2461388
jobs_are_scarce	no answer	182	0.3062014

Table 2 - Continuous Variable Descriptive Statistics

```
df1c <- df1c %>%
  mutate(age = ifelse(age >= 82, 82, age))

df1c <- df1c %>%
  mutate(age = as.numeric(age), # Attempt to convert age to numeric
         age = ifelse(is.na(age), NA, age)) # Non-numeric become NA

df1c %>%
  summarise(
    Mean = mean(age, na.rm = TRUE),
    SD = sd(age, na.rm = TRUE),
    Min = min(age, na.rm = TRUE),
    `25% Quantile` = quantile(age, 0.25, na.rm = TRUE),
    Median = median(age, na.rm = TRUE),
    `75% Quantile` = quantile(age, 0.75, na.rm = TRUE),
    Max = max(age, na.rm = TRUE)
  ) %>%
  knitr::kable()
```

Mean	SD	Min	25% Quantile	Median	75% Quantile	Max
49.78912	17.79272	18	35	50	64	82

No NAs in “age” confirmed

```
unique(df1c$age)

## [1] 69 55 70 42 23 22 21 20 37 31 65 63 18 68 38 27 62 64 49 28 59 26 34 58 78
## [26] 54 46 19 39 60 36 71 50 57 25 82 47 77 44 24 53 30 29 41 56 66 43 40 51 72
## [51] 32 75 33 35 45 48 73 76 67 74 61 80 81 52 79
```

How the 2 Vars of interest change with age (jobs_are_scare/child_suffers)

jobs_are_scarce with age

Enter words here when ready

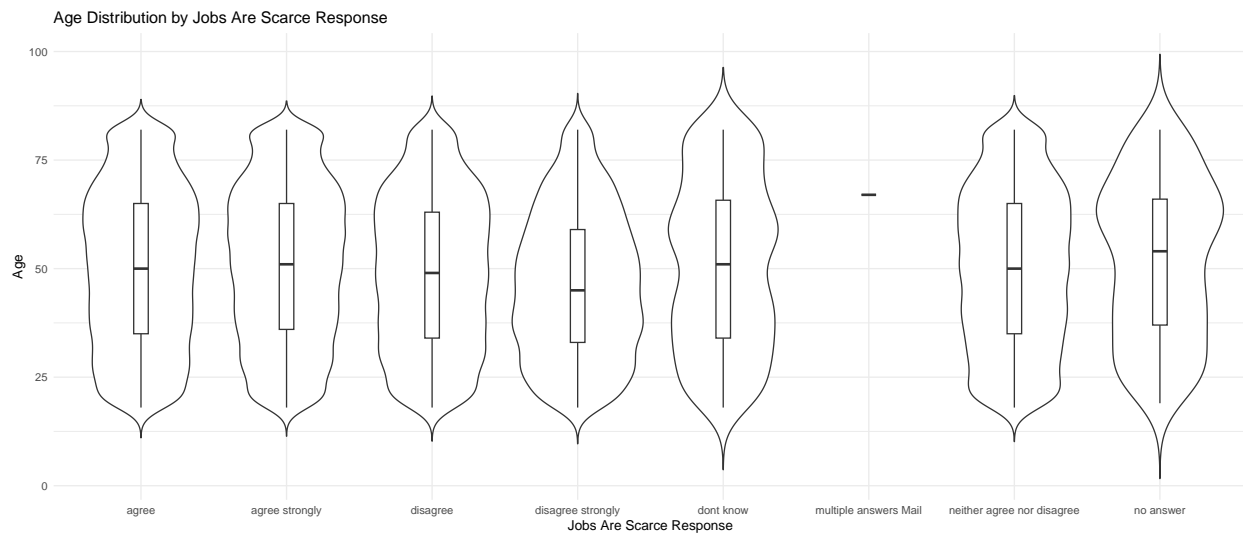
```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.3.2
```

```
df1c <- df1c %>%
  mutate(jobs_are_scarce = as.factor(jobs_are_scarce))

ggplot(df1c, aes(x = jobs_are_scarce, y = age)) +
  geom_violin(trim = FALSE) +
  geom_boxplot(width = 0.1, fill = "white") + # Adding a narrow boxplot inside for median/quartiles
  labs(title = "Age Distribution by Jobs Are Scarce Response",
        x = "Jobs Are Scarce Response",
        y = "Age") +
  theme_minimal()
```

Warning: Groups with fewer than two datapoints have been dropped.
 ## i Set 'drop = FALSE' to consider such groups for position adjustment purposes.



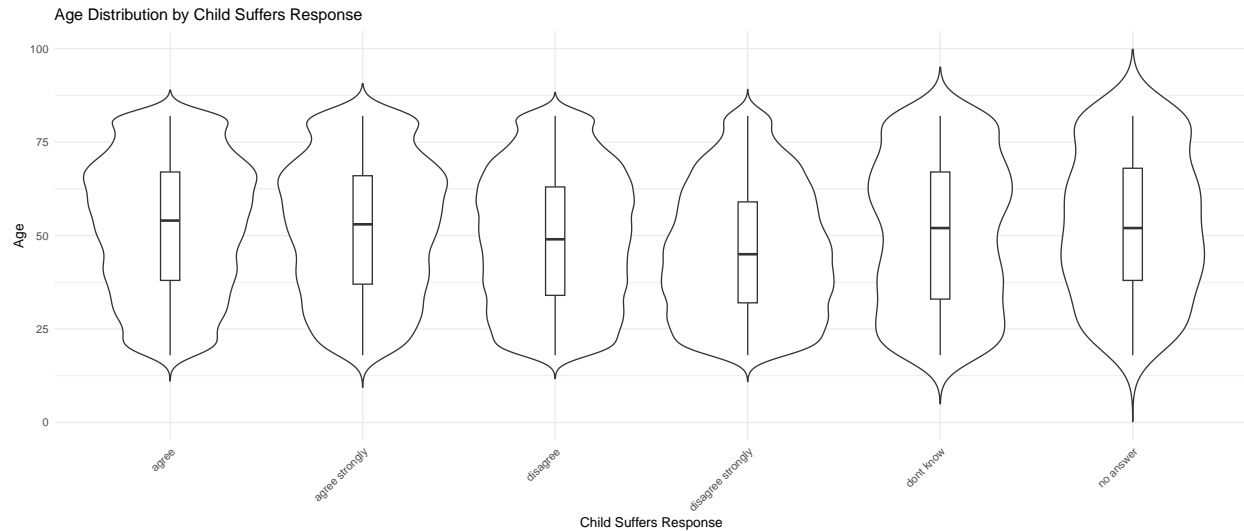
child_suffers with age

Enter words here for dialogue

```
library(ggplot2)

# Ensure 'child_suffers' is treated as a factor
df1c <- df1c %>%
  mutate(child_suffers = as.factor(child_suffers))

# Plotting the relationship between 'child_suffers' responses and 'age'
ggplot(df1c, aes(x = child_suffers, y = age)) +
  geom_violin(trim = FALSE) +
  geom_boxplot(width = 0.1, fill = "white") + # Adding a narrow boxplot inside for median/quartiles
  labs(title = "Age Distribution by Child Suffers Response",
        x = "Child Suffers Response",
        y = "Age") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) #
```



Regression Models

```
library(dplyr)
library(stargazer)
```

```
## Warning: package 'stargazer' was built under R version 4.3.1
```

```
##
```

```
## Please cite as:
```

```
## Hlavac, Marek (2022). stargazer: Well-Formatted Regression and Summary Statistics Tables.
```

```
## R package version 5.2.3. https://CRAN.R-project.org/package=stargazer
```

```
df1c <- df1c %>%
  mutate(age_squared = age^2,
         respondent_sex = as.factor(respondent_sex),
         education = as.factor(education))
```

```
model1 <- lm(jobs_are_scarce ~ age + age_squared + respondent_sex + education, data = df1c)
```

```
## Warning in model.response(mf, "numeric"): using type = "numeric" with a factor
## response will be ignored
```

```
## Warning in Ops.factor(y, z$residuals): '-' not meaningful for factors
```

```
model2 <- lm(child_suffers ~ age + age_squared + respondent_sex + education, data = df1c)
```

```
## Warning in model.response(mf, "numeric"): using type = "numeric" with a factor
## response will be ignored
```

```
## Warning in model.response(mf, "numeric"): '-' not meaningful for factors
```

```
stargazer(model1, model2, type = "text",
  title = "Regression Models Results",
  align = TRUE,
  header = FALSE,
  model.numbers = FALSE)
```

```
##
## Regression Models Results
## =====
##                               Dependent variable:
##                               -----
##                               jobs_are_scarce child_suffers
## -----
## age                          -0.004          -0.009
##
##
## age_squared                  0.00003         0.00003
##
##
## respondent_sexfemale        -0.061          0.091
##
##
## respondent_sexmale          -0.075          0.018
##
##
## respondent_sexno answer      0.135          0.563
##
##
## educationDoctoral or equivalent 0.182          0.223
##
##
## educationdont know          0.124         -0.101
##
##
## educationLess than primary  -0.452         -0.478
##
##
## educationLower secondary    -0.465         -0.487
##
##
## educationMaster or equivalent -0.175         -0.039
##
##
## educationno answer          0.219         -0.075
##
##
## educationother              0.691          0.320
##
##
## educationPost-secondary non tertiary -0.408         -0.240
##
##
## educationPrimary            -0.417         -0.489
```

```

##
##
## educationShort-cycle tertiary          -0.227          -0.143
##
##
## educationUpper secondary              -0.489          -0.375
##
##
## Constant                             3.361           3.169
##
## -----
## Observations                        59,438          59,438
## =====
## Note:                               *p<0.1; **p<0.05; ***p<0.01

```