

report_with_code

Savannah McNair

2025-03-29

Overview of data: The European Values Study is a comprehensive, cross-national, and longitudinal survey that examines Europeans' views on topics such as family, work, religion, politics, and society. Conducted every nine years, starting in 1981, the survey offers valuable insights into the beliefs, attitudes, preferences, values, and opinions of citizens throughout Europe. The survey has been repeated in the following years: 1990, 1999, 2008, and 2017, with the next wave scheduled for 2026. Each wave includes an increasing number of participating countries, expanding its scope across Europe and beyond.

For this analysis, we have selected a subset of variables from the dataset, including: the age of the respondent ("age"), the education of the respondent (recoded into three categories, "v243_r"), the sex of the respondent ("v225"), and the country (with abbreviation) ("c_abrv"). Additionally, two variables are Likert scales: child suffers with a working mother (1-4 scale, "v72") and jobs are scarce, giving priority to the nation (1-5 scale, "v80").

- clean data for analysis (20%) necessary packages

```
library(haven)
```

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

```
library(stargazer)
```

```
##
```

```
## 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
```

```
pull in data
```

```
location <- "C:/Users/smcnair1/Desktop/MWF_HW2/ZA7500_v5-0-0.sav/ZA7500_v5-0-0.sav"
dataset <- read_sav(location)
```

```
examine data
```

```
names(dataset)
```

```
##      [1] "studyno"      "version"
##      [3] "doi"          "studynoc"
##      [5] "versionc"     "id_cocas"
```

##	[7]	"caseno"	"year"
##	[9]	"fw_start"	"fw_end"
##	[11]	"country"	"c_abrv"
##	[13]	"cntry_y"	"mode"
##	[15]	"mm_select_sample"	"mm_mixed_mode"
##	[17]	"mm_mode_fu"	"mm_matrix_group"
##	[19]	"mm_original_matrix_design_IS"	"mm_original_matrix_group_IS"
##	[21]	"mm_fw_start_fu"	"mm_fw_end_fu"
##	[23]	"mm_year_fu"	"mr_detailed_mode_DE"
##	[25]	"mr_contact_mode_DE"	"mr_incentive_DE"
##	[27]	"fduplicate"	"fmissings"
##	[29]	"v1"	"v2"
##	[31]	"v3"	"v4"
##	[33]	"v5"	"v6"
##	[35]	"v7"	"v8"
##	[37]	"v9"	"v10"
##	[39]	"v11"	"v12"
##	[41]	"v13"	"v14"
##	[43]	"v15"	"v16"
##	[45]	"v17"	"v18"
##	[47]	"v19"	"f20"
##	[49]	"v20"	"v20a"
##	[51]	"v20b"	"v21"
##	[53]	"v22"	"v23"
##	[55]	"v24"	"f24_IT"
##	[57]	"v24a_IT"	"v24b_IT"
##	[59]	"v25"	"v26"
##	[61]	"v27"	"v28"
##	[63]	"v29"	"v30"
##	[65]	"f30a"	"v30a"
##	[67]	"v30b"	"v30c"
##	[69]	"v31"	"v32"
##	[71]	"v33"	"v34"
##	[73]	"v35"	"v36"
##	[75]	"v37"	"v38"
##	[77]	"v39"	"v40"
##	[79]	"v41"	"v42"
##	[81]	"v43"	"v44"
##	[83]	"v45"	"f45a"
##	[85]	"v45a"	"v45b"
##	[87]	"v45c"	"v46"
##	[89]	"f46_IT"	"v47"
##	[91]	"v48"	"v49"
##	[93]	"v50"	"v51"
##	[95]	"v52"	"v52_cs"
##	[97]	"v53"	"v54"
##	[99]	"v55"	"v56"
##	[101]	"v57"	"v58"
##	[103]	"v59"	"v60"
##	[105]	"v61"	"v62"
##	[107]	"v63"	"v64"
##	[109]	"v65"	"v66"
##	[111]	"v67"	"v68"
##	[113]	"v69"	"v70"

## [115]	"v71"	"v72"
## [117]	"v72_DE"	"v73"
## [119]	"v73_DE"	"v74"
## [121]	"v74_DE"	"v75"
## [123]	"v75_DE"	"v76"
## [125]	"v76_DE"	"v77"
## [127]	"v77_DE"	"v78"
## [129]	"v78_DE"	"v79"
## [131]	"v79_DE"	"v80"
## [133]	"v81"	"v82"
## [135]	"v83"	"v84"
## [137]	"f85"	"v85"
## [139]	"v86"	"v87"
## [141]	"v88"	"v89"
## [143]	"v90"	"v91"
## [145]	"v92"	"v93"
## [147]	"v94"	"v95"
## [149]	"f96"	"v96"
## [151]	"v96a"	"v96b"
## [153]	"v97"	"v98"
## [155]	"v99"	"v100"
## [157]	"v101"	"v102"
## [159]	"v103"	"v104"
## [161]	"v105"	"v106"
## [163]	"v107"	"f108"
## [165]	"v108"	"v109"
## [167]	"f110"	"v110"
## [169]	"v111"	"v111_4"
## [171]	"v112"	"f112_SE"
## [173]	"v113"	"v114"
## [175]	"v115"	"v116"
## [177]	"v117"	"v118"
## [179]	"v119"	"v120"
## [181]	"v121"	"v122"
## [183]	"v123"	"v124"
## [185]	"v125"	"v126"
## [187]	"v127"	"v128"
## [189]	"v129"	"v130"
## [191]	"v131"	"v132"
## [193]	"v133"	"v133_11c"
## [195]	"v134"	"v134_11c"
## [197]	"v135"	"v135_11c"
## [199]	"v136"	"v136_11c"
## [201]	"v137"	"v137_11c"
## [203]	"v138"	"v138_11c"
## [205]	"v139"	"v139_11c"
## [207]	"v140"	"v140_11c"
## [209]	"v141"	"v141_11c"
## [211]	"v142"	"v143"
## [213]	"v144"	"v145"
## [215]	"v146"	"v147"
## [217]	"v148"	"v149"
## [219]	"v150"	"v151"
## [221]	"v152"	"v153"

## [223]	"v154"	"v155"
## [225]	"v156"	"v157"
## [227]	"v158"	"v159"
## [229]	"v160"	"v161"
## [231]	"v162"	"v163"
## [233]	"v164"	"v165"
## [235]	"v166"	"v167"
## [237]	"v168"	"v169"
## [239]	"v170"	"v171"
## [241]	"v172"	"v173"
## [243]	"v174_LR"	"v174_cs"
## [245]	"v175_LR"	"v175_cs"
## [247]	"v176"	"v176_DK"
## [249]	"v177"	"v177_DK"
## [251]	"v178"	"v178_DK"
## [253]	"v179"	"v179_DK"
## [255]	"v180"	"v180_DK"
## [257]	"v181"	"v181_DK"
## [259]	"v182"	"v182_DK"
## [261]	"v183"	"v183_DK"
## [263]	"v184"	"v185"
## [265]	"v186"	"v187"
## [267]	"v188"	"v189"
## [269]	"v190"	"v191"
## [271]	"v192"	"v193"
## [273]	"v194"	"v195"
## [275]	"v196"	"v197"
## [277]	"v198"	"v199"
## [279]	"f199_IT"	"v200"
## [281]	"v201"	"v202"
## [283]	"v203"	"v204"
## [285]	"v205"	"v206"
## [287]	"v207"	"v208"
## [289]	"v209"	"v210"
## [291]	"v211"	"v212"
## [293]	"v213"	"v214"
## [295]	"v215"	"v216"
## [297]	"v217"	"v218"
## [299]	"v219"	"v220"
## [301]	"v221"	"v221_DK"
## [303]	"v222"	"v222_DK"
## [305]	"v223"	"v223_DK"
## [307]	"v224"	"v224_DK"
## [309]	"v225"	"v226"
## [311]	"age"	"age_r"
## [313]	"age_r2"	"age_r3"
## [315]	"v227"	"v228b"
## [317]	"v228b_r"	"v229"
## [319]	"v230"	"v231b"
## [321]	"v231b_r"	"v232"
## [323]	"v233b"	"v233b_r"
## [325]	"v234"	"v235"
## [327]	"v236"	"v237"
## [329]	"v238"	"v239_r"

## [331]	"v239a"	"v239b"
## [333]	"v240"	"v241"
## [335]	"v242"	"v242_r"
## [337]	"v243_edulv1b"	"v243_edulv1b_2"
## [339]	"v243_edulv1b_1"	"v243_ISCED_3"
## [341]	"v243_ISCED_2"	"v243_ISCED_2b"
## [343]	"v243_ISCED_1"	"v243_EISCED"
## [345]	"v243_ISCED97"	"v243_8cat"
## [347]	"v243_r"	"v243_cs"
## [349]	"v243_cs_DE1"	"v243_cs_DE2"
## [351]	"v243_cs_DE3"	"v243_cs_GB1"
## [353]	"v243_cs_GB2"	"v244"
## [355]	"v245"	"v246_ISCO_2"
## [357]	"v246_SIOPS"	"v246_ISEI"
## [359]	"v246_ESeC"	"v246_egp"
## [361]	"v247"	"v248"
## [363]	"v248a"	"v249"
## [365]	"v250"	"v251b"
## [367]	"v251b_r"	"v252_edulv1b"
## [369]	"f252_edulv1b_CH"	"v252_edulv1b_2"
## [371]	"v252_edulv1b_1"	"v252_ISCED_3"
## [373]	"v252_ISCED_2"	"v252_ISCED_2b"
## [375]	"v252_ISCED_1"	"v252_EISCED"
## [377]	"v252_ISCED97"	"v252_8cat"
## [379]	"v252_r"	"v252_cs"
## [381]	"v252_cs_DE1"	"v252_cs_DE2"
## [383]	"v252_cs_DE3"	"v252_cs_GB1"
## [385]	"v252_cs_GB2"	"v253"
## [387]	"v254"	"v255_ISCO_2"
## [389]	"v255_SIOPS"	"v255_ISEI"
## [391]	"v255_ESeC"	"v255_egp"
## [393]	"v256"	"v257"
## [395]	"v258"	"v259"
## [397]	"v260"	"v261"
## [399]	"v261_ppp"	"v261_r"
## [401]	"v262_edulv1b"	"v262_edulv1b_2"
## [403]	"v262_edulv1b_1"	"v262_ISCED_3"
## [405]	"v262_ISCED_2"	"v262_ISCED_2b"
## [407]	"v262_ISCED_1"	"v262_EISCED"
## [409]	"v262_ISCED97"	"v262_8cat"
## [411]	"v262_r"	"v262_cs"
## [413]	"v262_cs_DE1"	"v262_cs_DE2"
## [415]	"v262_cs_DE3"	"v262_cs_GB1"
## [417]	"v262_cs_GB2"	"v263_edulv1b"
## [419]	"v263_edulv1b_2"	"v263_edulv1b_1"
## [421]	"v263_ISCED_3"	"v263_ISCED_2"
## [423]	"v263_ISCED_2b"	"v263_ISCED_1"
## [425]	"v263_EISCED"	"v263_ISCED97"
## [427]	"v263_8cat"	"v263_r"
## [429]	"v263_cs"	"v263_cs_DE1"
## [431]	"v263_cs_DE2"	"v263_cs_DE3"
## [433]	"v263_cs_GB1"	"v263_cs_GB2"
## [435]	"v264"	"v265"
## [437]	"v266"	"v267"

```
## [439] "v268"          "v269"
## [441] "v270"          "v271"
## [443] "v272"          "v273"
## [445] "v274"          "v275b_N2"
## [447] "v275b_N1"      "v275c_N2"
## [449] "v275c_N1"      "v276_r"
## [451] "v277"          "mm_v277_fu"
## [453] "v278a"         "v278b"
## [455] "v278c_r"       "v279a"
## [457] "v279b"         "v279c_r"
## [459] "v279d_r"       "mm_v278a_fu"
## [461] "mm_v278b_fu"   "mm_v279a_fu"
## [463] "mm_v279b_fu"   "v280"
## [465] "v281a"         "v281a_r"
## [467] "v282"          "gweight"
## [469] "gweight_no_edu" "dweight"
## [471] "pweight"       "age_r3_weight"
## [473] "v225_weight"   "v243_r_weight"
```

Observations from first glance at data:

- age variable is “age”
- education of respondent recoded into 3 categories is “v243_r”
- sex of respondent “v225”
- country (with abbreviation) is “c_abrv”
- v72 is child suffers with working mother (1-4)
- v80 is jobs are scarce... giving (nation) priority (1-5)
- _DE columns are duplicates with many NAs, I won’t need these for my analyses

remove _DE columns

```
dataset <- dataset[, !grepl('_DE', names(dataset))]
```

remove incomplete questionnaires

```
dataset <- dataset[dataset$fmissings != 1, ]
```

keep only the columns of interest (above are the variables I selected)

```
dataset <- dataset[, c("age", "v243_r", "v225", "c_abrv", "v72", "v80")]
```

remove entries where education level question was refused and remove rows with NAs

```
dataset <- dataset[dataset$v243_r != 66, ]
dataset <- na.omit(dataset)
```

rename columns, keep attributes

```
colnames(dataset) <- c("age", "education", "sex", "country", "child_suffers_mom", "job_scarcity")
```

now we have a clean, relevant, analyzable dataset!

• descriptive tables, graphs and regression tables (20%) necessary packages

```
library(dplyr)
```

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

```
##
```

```
## 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
```

```
#install.packages("summarytools")
```

```
library(summarytools)
```

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

descriptive tables

continuous - given that child suffers and job scarcity variables are bipolar likert scales with numerical values, im making the analytical decision to include them here

```
continuous_summary <- data.frame(  
  Variable = c("age", "child_suffers_mom", "job_scarcity"),  
  Mean = c(mean(dataset$age), mean(dataset$child_suffers_mom), mean(dataset$job_scarcity)),  
  Median = c(median(dataset$age), median(dataset$child_suffers_mom), median(dataset$job_scarcity)),  
  SD = c(sd(dataset$age), sd(dataset$child_suffers_mom), sd(dataset$job_scarcity)),  
  Min = c(min(dataset$age), min(dataset$child_suffers_mom), min(dataset$job_scarcity)),  
  Max = c(max(dataset$age), max(dataset$child_suffers_mom), max(dataset$job_scarcity)),  
  Range = c(max(dataset$age) - min(dataset$age),  
            max(dataset$child_suffers_mom) - min(dataset$child_suffers_mom),  
            max(dataset$job_scarcity) - min(dataset$job_scarcity))  
)
```

```
## Warning: '..1' and '..3' have conflicting value labels.
```

```
## i Labels for these values will be taken from '..1'.
```

```
## x Values: 3 and 4
```

```
## '..1' and '..3' have conflicting value labels.
```

```
## i Labels for these values will be taken from '..1'.
```

```
## x Values: 3 and 4
```

```
print(continuous_summary)
```

```
##           Variable      Mean Median      SD Min Max Range
## 1           age 49.562276      50 17.6102193  18  82   64
## 2 child_suffers_mom 2.712773       3  0.8788531   1   4    3
## 3      job_scarcity 2.312960       2  1.2788007   1   5    4
```

We can learn a number of things from this table. We know that the mean and median age are very close in this population, hovering around 50 with a standard deviation of 17 years (fairly spread out). We know that over time the people answer around 2 or 3 out of 4 (agree->disagree) for child_suffers_mom, with lower variability. We know people answer close to 2 out of 5 (agree->disagree) with higher variability on job_scarcity.

categorical create summary

```
get_categorical_summary <- function(variable, var_name) {
  freq_table <- table(variable) # Frequency table
  prop_table <- prop.table(freq_table) * 100 # Percentage table

  summary_df <- data.frame(
    Variable = var_name,
    Level = names(freq_table),
    Frequency = as.integer(freq_table),
    Percentage = round(as.numeric(prop_table), 2)
  )

  return(summary_df)
}
```

create summary for each and append to one table

```
education_summary <- get_categorical_summary(dataset$education, "education")
sex_summary <- get_categorical_summary(dataset$sex, "sex")
country_summary <- get_categorical_summary(dataset$country, "country")

combined_summary <- rbind(education_summary, sex_summary, country_summary)

print(combined_summary)
```

```
##      Variable Level Frequency Percentage
## 1  education     1     11506      20.31
## 2  education     2     25778      45.50
## 3  education     3     19375      34.20
## 4      sex       1     25036      44.19
## 5      sex       2     31623      55.81
## 6   country    AL      1417       2.50
## 7   country    AM      1480       2.61
## 8   country    AT      1566       2.76
## 9   country    AZ      1726       3.05
## 10  country    BA      1649       2.91
## 11  country    BG      1464       2.58
## 12  country    BY      1494       2.64
```


## 13	country	CH	3061	5.40
## 14	country	CZ	1633	2.88
## 15	country	DE	2062	3.64
## 16	country	DK	3220	5.68
## 17	country	EE	1234	2.18
## 18	country	ES	1173	2.07
## 19	country	FI	1127	1.99
## 20	country	FR	1815	3.20
## 21	country	GB	1721	3.04
## 22	country	GE	2129	3.76
## 23	country	HR	1434	2.53
## 24	country	HU	1464	2.58
## 25	country	IS	1554	2.74
## 26	country	IT	2165	3.82
## 27	country	LT	1357	2.40
## 28	country	LV	1252	2.21
## 29	country	ME	951	1.68
## 30	country	MK	1027	1.81
## 31	country	NL	2284	4.03
## 32	country	NO	1093	1.93
## 33	country	PL	1278	2.26
## 34	country	PT	1170	2.06
## 35	country	RO	1448	2.56
## 36	country	RS	1413	2.49
## 37	country	RU	1711	3.02
## 38	country	SE	1143	2.02
## 39	country	SI	1043	1.84
## 40	country	SK	1393	2.46
## 41	country	UA	1508	2.66

This table tells us a number of things. First, that education is slightly skewed towards higher education (3) over low (1), with most falling in the medium category (2). Less people over all are men (1, 44%) as compared with women (2, 56%). Most countries make up able 2% of the overall sample. The Netherlands (NL) and Denmark (DK) are substantially larger.

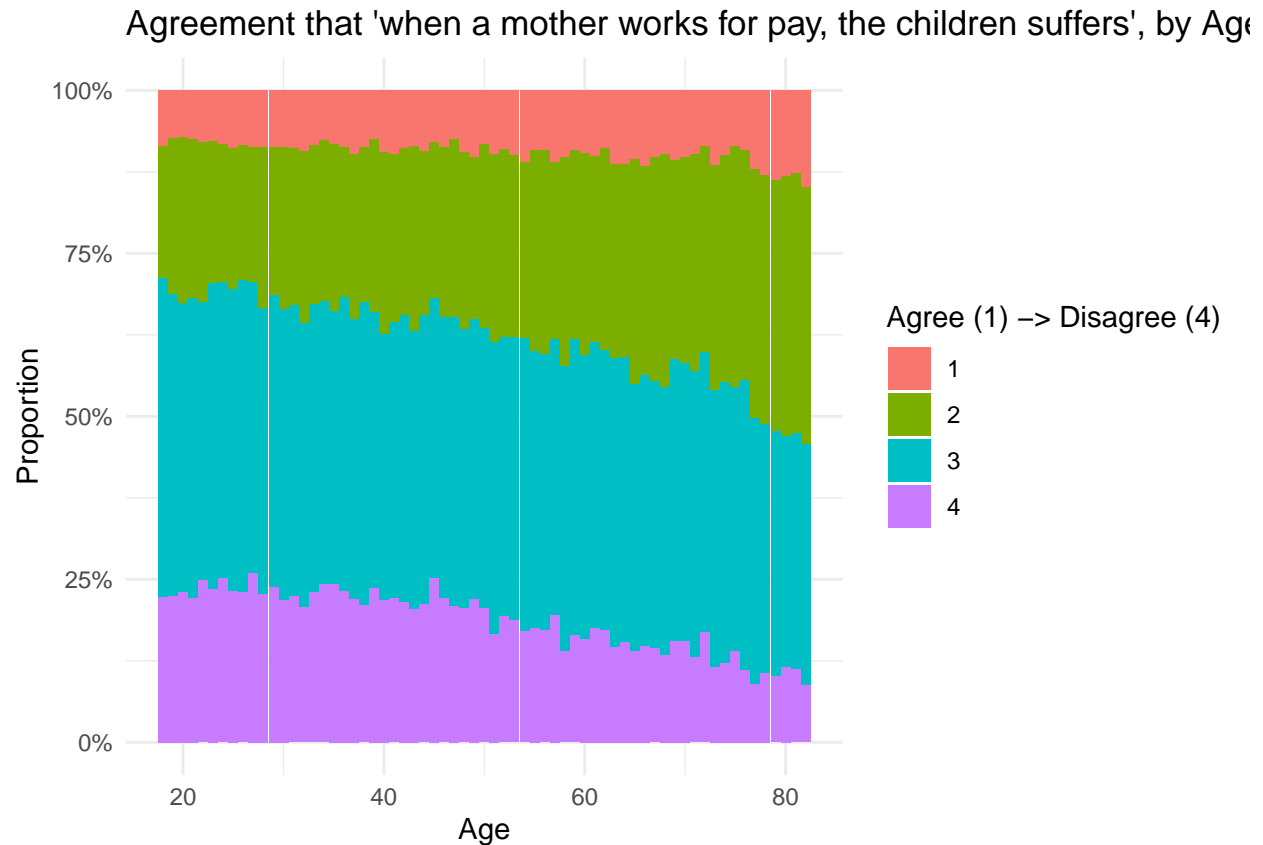
graphs necessary packages

```
library(ggplot2)
```

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

child suffers when mom works

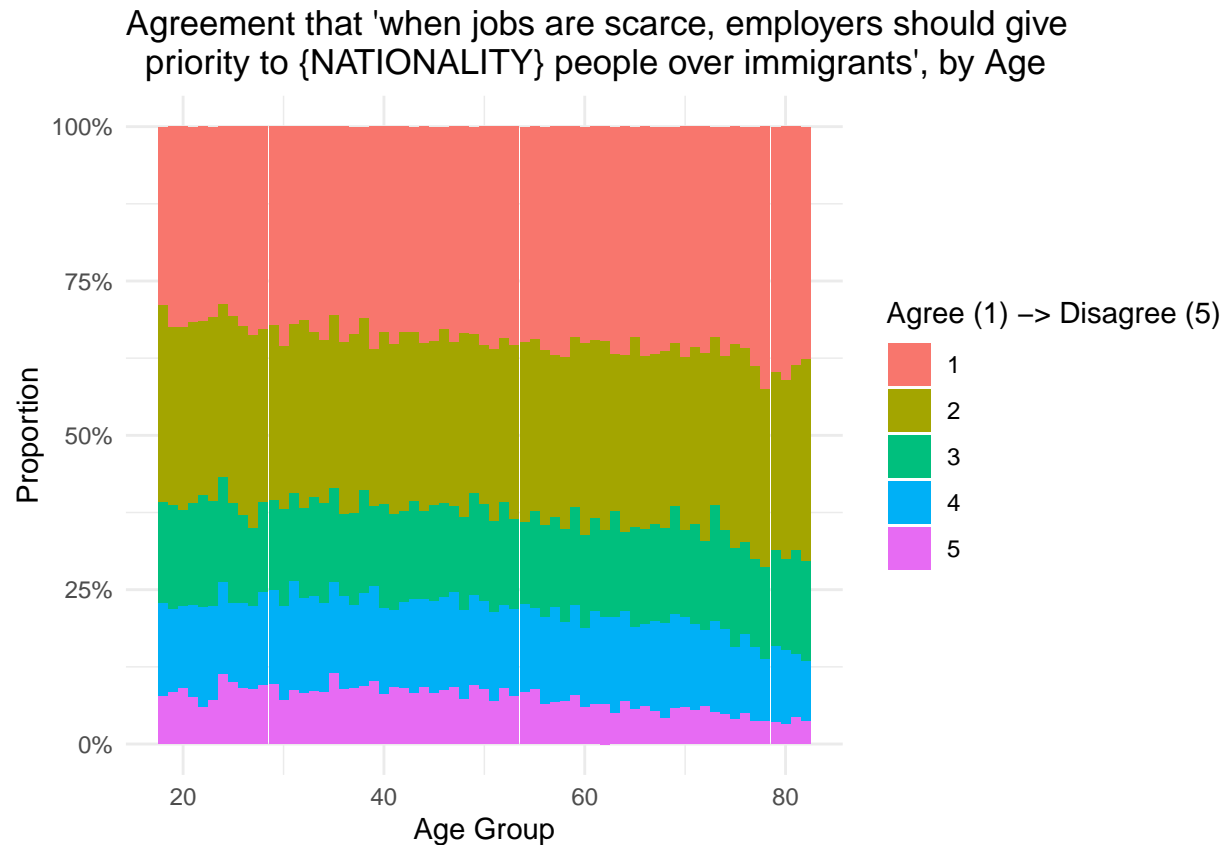
```
ggplot(dataset, aes(x = age, fill = factor(child_suffers_mom))) +
  geom_bar(position = "fill") +
  labs(x = "Age", y = "Proportion", fill = "Agree (1) -> Disagree (4)") +
  ggtitle("Agreement that 'when a mother works for pay, the children suffers', by Age") +
  scale_y_continuous(labels = scales::percent) +
  theme_minimal()
```



With age, people tend to agree with the statement “when a mother works for pay, the children suffers” more. The agreement is more moderate, as most people with age will select “agree” (2) over “agree strongly” (1).

job scarcity

```
ggplot(dataset, aes(x = age, fill = factor(job_scarcity))) +
  geom_bar(position = "fill") +
  labs(x = "Age Group", y = "Proportion", fill = "Agree (1) -> Disagree (5)") +
  ggtitle("Agreement that 'when jobs are scarce, employers should give
priority to {NATIONALITY} people over immigrants', by Age") +
  scale_y_continuous(labels = scales::percent) +
  theme_minimal()
```



With age, people will tend to agree with “when jobs are scarce, employers should give priority to {NATIONALITY} people over immigrants” more. This agreement is more pronounced, as more people will choose “Agree strongly” over “Agree”, and the disagreement is vastly smaller in comparison.

regression tables

create new variables of interest

```
dataset$age_squared <- (dataset$age)^2
```

create models

```
#model 1 child suffers when mom works
model1 <- lm(child_suffers_mom ~ age + age_squared + sex + education, data = dataset)

#model 2 job scarcity impact on immigrants
model2 <- lm(job_scarcity ~ age + age_squared + sex + education, data = dataset)
```

get summaries

```
summary_table <- stargazer(model1, model2, type = "text")
```

```
##
## =====
##                               Dependent variable:
##                               -----
```

```
##                                child_suffers_mom job_scarcity
##                                (1)                (2)
## -----
## age                            -0.003***          -0.002
##                                (0.001)            (0.002)
##
## age_squared                    -0.00001           -0.00001
##                                (0.00001)          (0.00002)
##
## sex                            0.067***           -0.021**
##                                (0.007)            (0.011)
##
## education                      0.210***           0.244***
##                                (0.005)            (0.008)
##
## Constant                      2.362***           1.931***
##                                (0.031)            (0.046)
## -----
## Observations                   56,659            56,659
## R2                             0.046             0.022
## Adjusted R2                    0.046             0.022
## Residual Std. Error (df = 56654) 0.859           1.264
## F Statistic (df = 4; 56654)      679.966***       323.631***
## =====
## Note:                          *p<0.1; **p<0.05; ***p<0.01
```

```
print(summary_table)
```

```
## [1] ""
## [2] "=====
## [3] "                Dependent variable:      "
## [4] "                -----
## [5] "                child_suffers_mom job_scarcity"
## [6] "                (1)                (2)      "
## [7] "-----
## [8] "age            -0.003***          -0.002  "
## [9] "              (0.001)            (0.002)  "
## [10] "              "
## [11] "age_squared    -0.00001           -0.00001 "
## [12] "              (0.00001)          (0.00002) "
## [13] "              "
## [14] "sex            0.067***           -0.021** "
## [15] "              (0.007)            (0.011)  "
## [16] "              "
## [17] "education      0.210***           0.244*** "
## [18] "              (0.005)            (0.008)  "
## [19] "              "
## [20] "Constant       2.362***           1.931*** "
## [21] "              (0.031)            (0.046)  "
## [22] "              "
## [23] "-----
## [24] "Observations   56,659            56,659  "
## [25] "R2              0.046             0.022   "
```

```
## [26] "Adjusted R2                0.046          0.022      "
## [27] "Residual Std. Error (df = 56654) 0.859          1.264      "
## [28] "F Statistic (df = 4; 56654)    679.966***     323.631***  "
## [29] "=====
```

Model 1 (child_suffers_mom) suggests that age, sex, and education are significantly associated with whether people think a child suffers when the mother works, with sex and education being positively related to the outcome and age showing a slight negative relationship.

Model 2 (job_scarcity) finds that education and sex are significant predictors of peoples thoughts that job scarcity should give priority in jobs to nationals over immigrants. Education has a positive association with this variable, and being male slightly reduces this impact.

However, both models have low R^2 values, meaning they only explain a small portion of the variance in the dependent variables, and there may be other factors not captured by these models that influence the outcomes.