231005In_Class_Exercise

Athena Zhang

2023-10-05

In-Class Exercise

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.3     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.3     v tibble    3.2.1
## v lubridate 1.9.2     v tidyr     1.3.0
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
d <- read_csv("_DataPublic_/vdem/1984_2022/vdem_1984_2022_external.csv")
```

```
## Rows: 6789 Columns: 211
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr    (3): country_name, country_text_id, histname
## dbl  (207): country_id, year, project, historical, codingstart, codingend, c...
## date   (1): historical_date
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

1. Codebook lookup

1.1 What indicators regarding the quality of education are available in the V-Dem datasets?

"e_peaveduc" = average years of education among citizens older than 15 "e_peedgini" = How unequal is the level of education achieved by the population aged 15 years and older?

1.2 What are the data's coverage (i.e., for which countries and years do we have data?)

1.2.1 Selecting:

```
d_edu <- d |>
  select(country_name, year, e_peaveduc, e_peedgini)
```

1.2.2 Renaming:

```
d_edu <- d_edu |>
  rename("Country" = "country_name", "Year"="year", "Avg_Edu_Yrs"="e_peaveduc", "Edu_Gini"="e_peedgini")
```

```
d_edu
```

```
## # A tibble: 6,789 x 4
##    Country  Year Avg_Edu_Yrs Edu_Gini
##    <chr>   <dbl>       <dbl>    <dbl>
##  1 Mexico   1984        6.08     32.7
##  2 Mexico   1985        6.22     32.4
##  3 Mexico   1986        6.36     31.9
##  4 Mexico   1987        6.5      31.4
##  5 Mexico   1988        6.64     31.1
##  6 Mexico   1989        6.78     30.1
##  7 Mexico   1990        6.92     30.0
##  8 Mexico   1991        7.03     29.7
##  9 Mexico   1992        7.14     29.5
## 10 Mexico   1993        7.25     29.3
## # i 6,779 more rows
```

1.2.3 Finding how many countries:

```
d_edu |>
  distinct(Country) |>
  count()
```

```
## # A tibble: 1 x 1
##       n
##   <int>
## 1   181
```

We have 181 countries. Below are the individual countries.

```
d_edu |> select(Country) |> distinct()
```

```
## # A tibble: 181 x 1
##    Country
##    <chr>
##  1 Mexico
##  2 Suriname
##  3 Sweden
##  4 Switzerland
```

```
##  5 Ghana
##  6 South Africa
##  7 Japan
##  8 Burma/Myanmar
##  9 Russia
## 10 Albania
## # i 171 more rows
```

1.2.4 **We have data from 1984 to 2022** but some countries have missing data in some years. See below.

```
d_edu|>
  mutate(Avg_edu_yrs_missing = as.numeric(is.na(Avg_Edu_Yrs)), .after = Avg_Edu_Yrs)|>
  group_by(Country)|>
  summarise(N_Year_missing = sum(Avg_edu_yrs_missing))
```

```
## # A tibble: 181 x 2
##    Country     N_Year_missing
##    <chr>                <dbl>
##  1 Afghanistan              0
##  2 Albania                 39
##  3 Algeria                  0
##  4 Angola                   0
##  5 Argentina                0
##  6 Armenia                  0
##  7 Australia                0
##  8 Austria                  0
##  9 Azerbaijan               0
## 10 Bahrain                 39
## # i 171 more rows
```

1.3 What are their sources? Provide the link to at least 1 source

https://clio-infra.eu/Indicators/AverageYearsofEducation.html

^ Average years of Education (from Clio Infra)

https://clio-infra.eu/Indicators/EducationalInequalityGiniCoefficient.html

^ Gini Coefficient for Educational Inequality (from Clio Infra)

2. Subset by columns

2.1 Create a dataset containing only the country-year identifiers and indicators of education quality

See section 1.2.1

2.2 Rename the columns of education quality to make them informative

See section 1.2.2

"Avg_Edu_Yrs"="e_peaveduc"

"Edu_Gini"="e_peedgini"

3. Subset by rows

3.1 List 5 countries-years that have the highest education level among its population

```
d_edu |>
  slice_max(order_by = Avg_Edu_Yrs, n = 5, with_ties = FALSE)
```

```
## # A tibble: 5 x 4
##   Country         Year Avg_Edu_Yrs Edu_Gini
##   <chr>          <dbl>       <dbl>    <dbl>
## 1 United Kingdom  2010        13.3     6.07
## 2 United Kingdom  2011        13.3    NA
## 3 United Kingdom  2012        13.3    NA
## 4 United Kingdom  2013        13.3    NA
## 5 United Kingdom  2014        13.3    NA
```

3.2 List 5 countries-years that suffer from the most severe inequality in education.

```
d_edu |>
  slice_max(order_by = Edu_Gini, n = 5)
```

```
## # A tibble: 5 x 4
##   Country       Year Avg_Edu_Yrs Edu_Gini
##   <chr>        <dbl>       <dbl>    <dbl>
## 1 Burkina Faso  1984       0.301     97.0
## 2 Burkina Faso  1985       0.322     96.9
## 3 Burkina Faso  1986       0.343     96.7
## 4 Burkina Faso  1987       0.364     96.4
## 5 Burkina Faso  1988       0.385     96.1
```

4. Summarise the data

4.1 Check data availability: For which countries and years are the indicators of education quality available?

See section 1.2.4

4.2 Create two types of country-level indicators of education quality

1. Average level of education quality from 1984 to 2022

```
d_edu |>
  group_by(Country) |>
  summarise(Avg_Edu_Level = mean(Avg_Edu_Yrs, na.rm = TRUE), Avg_Gini = mean(Edu_Gini, na.rm = TRUE)) |>
  arrange(Avg_Edu_Level)
```

```
## # A tibble: 181 x 3
##    Country      Avg_Edu_Level Avg_Gini
##    <chr>                <dbl>    <dbl>
##  1 Burkina Faso         0.982     91.3
##  2 Niger                1.06      85.3
##  3 Mali                 1.25      87.9
##  4 Somalia              1.29      84.7
##  5 Burundi              1.86      73.0
##  6 Mozambique           2.36      52.6
##  7 Benin                2.39      76.9
##  8 Angola               2.46      53.9
##  9 Senegal              2.54      66.8
## 10 Guinea               2.62      73.4
## # i 171 more rows
```

2. Change of education quality from 1984 to 2022

```
d_edu |>
  group_by(Country)|>
  arrange(Year)|>
  summarise(Edu_quality_2022_1984 = (last(Avg_Edu_Yrs)-first(Avg_Edu_Yrs)) / first(Avg_Edu_Yrs))|>
  ungroup() |>
  arrange(Edu_quality_2022_1984)
```

```
## # A tibble: 181 x 2
##    Country     Edu_quality_2022_1984
##    <chr>                       <dbl>
##  1 Tajikistan                -0.0262
##  2 North Korea                0
##  3 Azerbaijan                 0.0239
##  4 Russia                     0.0245
##  5 Switzerland                0.0265
##  6 Uzbekistan                 0.0271
##  7 Germany                    0.0277
##  8 Kyrgyzstan                 0.0303
##  9 Armenia                    0.0321
## 10 Georgia                    0.0368
## # i 171 more rows
```

4.3 Examine the data and briefly discuss: Which countries perform the best and the worst in terms of education quality in the past four decades?

On average, Burkina Faso, Niger, Mali, Somalia, and Burundi, from lowest to highest respectively, have the lowest avaerage years of education among citizens above 15.

However, countries like Sweden, Belarus, Australia, Moldova, and Namibia have experienced the least change in average education years over the past four decades.

On the other hand, Germany, Australia, United Kingdom, Canada, and Switzerland have the highest average education years among citizens above 15.