In-class exercise 2

LIU Yijun, Edrian

2023-10-05

1: Codebook lookup

1.1 Indicators regarding the quality of education

In P381 of the Codebook, we can see two indicators regarding education:

1. Education 15+ (e_peaveduc), which describes the average years of education among citizens older than 15.
2. Educational inequality, Gini (e_peedgini), which describes the level of inequality of education achieved by the population aged 15 or older.

1.2 Data's coverage

For Education 15+: worldwide coverage, 1820-2022

For Educational inequality, Gini: worldwide coverage, 1850-2010

1.3 Sources

Clio Infra (clio-infra.eu)

2: Subset by columns

2.1 Country-year identifiers and indicators of education quality

Step 1: Loading library and data

```
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("data/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.
```

Step 2: Selecting the identifiers and indicators

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

2.2 Renaming Columns

```
d_edu <- d_edu |>
  rename("Country" = "country_name",
         "ID" = "country_id",
         "Year" = "year",
         "Education_15" = "e_peaveduc",
         "Education_Inequality" = "e_peedgini")

glimpse(d_edu)
```

```
## Rows: 6,789
## Columns: 5
## $ Country              <chr> "Mexico", "Mexico", "Mexico", "Mexico", "Mexico",~
## $ ID                   <dbl> 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3~
## $ Year                 <dbl> 1984, 1985, 1986, 1987, 1988, 1989, 1990, 1991, 1~
## $ Education_15         <dbl> 6.080, 6.220, 6.360, 6.500, 6.640, 6.780, 6.920, ~
## $ Education_Inequality <dbl> 32.732, 32.399, 31.910, 31.443, 31.113, 30.085, 2~
```

3. Subset by rows

3.1 Listing 5 country-years with highest education level

```
d_edu |>
  slice_max(order_by = Education_15, n = 5)
```

```
## # A tibble: 13 x 5
##    Country           ID  Year Education_15 Education_Inequality
##    <chr>          <dbl> <dbl>        <dbl>                <dbl>
##  1 United Kingdom   101  2010         13.3                 6.07
##  2 United Kingdom   101  2011         13.3                NA
```

```
##  3 United Kingdom   101  2012         13.3                NA
##  4 United Kingdom   101  2013         13.3                NA
##  5 United Kingdom   101  2014         13.3                NA
##  6 United Kingdom   101  2015         13.3                NA
##  7 United Kingdom   101  2016         13.3                NA
##  8 United Kingdom   101  2017         13.3                NA
##  9 United Kingdom   101  2018         13.3                NA
## 10 United Kingdom   101  2019         13.3                NA
## 11 United Kingdom   101  2020         13.3                NA
## 12 United Kingdom   101  2021         13.3                NA
## 13 United Kingdom   101  2022         13.3                NA
```

As the United Kingdom has 13 years of identical data (13.3 years), they are all shown in the result.

3.2 Listing 5 country-years with highest education inequality

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

```
## # A tibble: 5 x 5
##   Country         ID  Year Education_15 Education_Inequality
##   <chr>        <dbl> <dbl>        <dbl>                <dbl>
## 1 Burkina Faso    54  1984        0.301                 97.0
## 2 Burkina Faso    54  1985        0.322                 96.9
## 3 Burkina Faso    54  1986        0.343                 96.7
## 4 Burkina Faso    54  1987        0.364                 96.4
## 5 Burkina Faso    54  1988        0.385                 96.1
```

4. Summarize the data

4.1 Check data availability

For `Education_15`:

```
d_edu |>
  mutate(edu_15_missing = as.numeric(is.na(Education_15))) |>
  group_by(Country) |>
  summarize(num_edu_15_missing = sum(edu_15_missing))
```

```
## # A tibble: 181 x 2
##    Country     num_edu_15_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
```

For Education_Inequality:

```
d_edu |>
  mutate(edu_inequality_missing = as.numeric(is.na(Education_Inequality))) |>
  group_by(Country) |>
  summarize(num_edu_inequality_missing = sum(edu_inequality_missing))
```

```
## # A tibble: 181 x 2
##    Country     num_edu_inequality_missing
##    <chr>                            <dbl>
##  1 Afghanistan                         12
##  2 Albania                             39
##  3 Algeria                             12
##  4 Angola                              12
##  5 Argentina                           12
##  6 Armenia                             12
##  7 Australia                           12
##  8 Austria                             12
##  9 Azerbaijan                          12
## 10 Bahrain                             39
## # i 171 more rows
```

4.2 Country-level indicators

4.2.1 Average level of education quality, 1984-2022

For Education_15:

```
# Countries with longest average education years
d_edu |>
  filter(Year >= 1984 & Year <= 2022) |>
  arrange(Year) |>
  group_by(Country) |>
  summarize(Average_Education_15 = mean(Education_15, na.rm = TRUE)) |>
  ungroup() |>
  arrange(desc(Average_Education_15))
```

```
## # A tibble: 181 x 2
##    Country        Average_Education_15
##    <chr>                         <dbl>
##  1 Germany                        12.9
##  2 Australia                      12.9
##  3 United Kingdom                 12.9
##  4 Canada                         12.7
##  5 Switzerland                    12.7
##  6 Japan                          12.6
##  7 Norway                         12.4
##  8 France                         12.0
##  9 South Korea                    12.0
## 10 New Zealand                    11.9
## # i 171 more rows
```

```r
# Countries with shortest average education years
d_edu |>
  filter(Year >= 1984 & Year <= 2022) |>
  arrange(Year) |>
  group_by(Country) |>
  summarize(Average_Education_15 = mean(Education_15, na.rm = TRUE)) |>
  ungroup() |>
  arrange(Average_Education_15)
```

```
## # A tibble: 181 x 2
##    Country      Average_Education_15
##    <chr>                       <dbl>
##  1 Burkina Faso                0.982
##  2 Niger                       1.06
##  3 Mali                        1.25
##  4 Somalia                     1.29
##  5 Burundi                     1.86
##  6 Mozambique                  2.36
##  7 Benin                       2.39
##  8 Angola                      2.46
##  9 Senegal                     2.54
## 10 Guinea                      2.62
## # i 171 more rows
```

For Education_Inequality:

```r
# Countries with least amount of inequality
d_edu |>
  filter(Year >= 1984 & Year <= 2022) |>
  arrange(Year) |>
  group_by(Country) |>
  summarize(Average_Education_Inequality = mean(Education_Inequality, na.rm = TRUE)) |>
  ungroup() |>
  arrange(Average_Education_Inequality)
```

```
## # A tibble: 181 x 2
##    Country        Average_Education_Inequality
##    <chr>                                 <dbl>
##  1 Austria                                6.35
##  2 Barbados                               6.98
##  3 Denmark                                8.17
##  4 Switzerland                            8.28
##  5 United Kingdom                         8.38
##  6 Japan                                  9.33
##  7 Norway                                 9.58
##  8 Australia                              9.60
##  9 Tajikistan                            10.8
## 10 Hungary                               11.2
## # i 171 more rows
```

```r
# Countries with most amount of inequality
d_edu |>
```

```
  filter(Year >= 1984 & Year <= 2022) |>
  arrange(Year) |>
  group_by(Country) |>
  summarize(Average_Education_Inequality = mean(Education_Inequality, na.rm = TRUE)) |>
  ungroup() |>
  arrange(desc(Average_Education_Inequality))
```

```
## # A tibble: 181 x 2
##    Country      Average_Education_Inequality
##    <chr>                               <dbl>
##  1 Burkina Faso                         91.3
##  2 Mali                                 87.9
##  3 Niger                                85.3
##  4 Somalia                              84.7
##  5 Afghanistan                          77.8
##  6 Benin                                76.9
##  7 The Gambia                           76.7
##  8 Guinea                               73.4
##  9 Burundi                              73.0
## 10 Nepal                                69.8
## # i 171 more rows
```

4.2.2 Change of education quality, 1984-2022

For `Education_15`:

```
# Countries with most increase in average education years
d_edu |>
  filter(Year >= 1984 & Year <= 2022) |>
  arrange(Year) |>
  group_by(Country) |>
  summarise(Education_15_Change =
              last(Education_15, na_rm = TRUE) -
              first(Education_15, na_rm = TRUE)) |>
  ungroup() |>
  arrange(desc(Education_15_Change))
```

```
## # A tibble: 181 x 2
##    Country      Education_15_Change
##    <chr>                      <dbl>
##  1 Botswana                    5.17
##  2 Singapore                   4.52
##  3 Libya                       4.07
##  4 Cuba                        3.84
##  5 Chad                        3.82
##  6 Egypt                       3.82
##  7 Jordan                      3.82
##  8 South Korea                 3.54
##  9 Saudi Arabia                3.49
## 10 Algeria                     3.35
## # i 171 more rows
```

```r
# Countries with least increase in education years
d_edu |>
  filter(Year >= 1984 & Year <= 2022) |>
  arrange(Year) |>
  group_by(Country) |>
  summarise(Education_15_Change =
              last(Education_15, na_rm = TRUE) -
              first(Education_15, na_rm = TRUE)) |>
  ungroup() |>
  arrange((Education_15_Change))
```

```
## # A tibble: 181 x 2
##    Country     Education_15_Change
##    <chr>                     <dbl>
##  1 Tajikistan               -0.252
##  2 North Korea               0
##  3 Russia                    0.230
##  4 Azerbaijan                0.252
##  5 Uzbekistan                0.272
##  6 Kyrgyzstan                0.301
##  7 Switzerland               0.328
##  8 Armenia                   0.336
##  9 Germany                   0.350
## 10 Georgia                   0.387
## # i 171 more rows
```

For Education_Inequality:

```r
# Countries with most decline of education inequality
d_edu |>
  filter(Year >= 1984 & Year <= 2022) |>
  arrange(Year) |>
  group_by(Country) |>
  summarise(Education_Inequality_Change =
              last(Education_Inequality, na_rm = TRUE) -
              first(Education_Inequality, na_rm = TRUE)) |>
  ungroup() |>
  arrange(desc(Education_Inequality_Change))
```

```
## # A tibble: 181 x 2
##    Country             Education_Inequality_Change
##    <chr>                                     <dbl>
##  1 Costa Rica                                4.12
##  2 New Zealand                               3.16
##  3 Spain                                     2.30
##  4 Trinidad and Tobago                       2.30
##  5 Switzerland                               1.72
##  6 Lebanon                                   0.718
##  7 Seychelles                                0.696
##  8 France                                   -0.287
##  9 Venezuela                                -0.395
## 10 Jamaica                                  -0.597
## # i 171 more rows
```

```r
# Countries with most increase in education inequality

d_edu |>
  filter(Year >= 1984 & Year <= 2022) |>
  arrange(Year) |>
  group_by(Country) |>
  summarise(Education_Inequality_Change =
              last(Education_Inequality, na_rm = TRUE) -
              first(Education_Inequality, na_rm = TRUE)) |>
  ungroup() |>
  arrange(Education_Inequality_Change)
```

```
## # A tibble: 181 x 2
##    Country  Education_Inequality_Change
##    <chr>                          <dbl>
##  1 Nepal                          -39.8
##  2 Botswana                       -34.0
##  3 Haiti                          -31.5
##  4 Egypt                          -30.8
##  5 Iran                           -30.3
##  6 Angola                         -29.5
##  7 India                          -29.0
##  8 Nigeria                        -27.5
##  9 Malawi                         -27.2
## 10 Uganda                         -26.8
## # i 171 more rows
```

4.3 Discussion

Which countries perform the best and the worst in terms of education quality in the past four decades?

If we look at the **average performance** over the years, then developed countries such as Germany, Australia, and the UK has the best education quality, where an average citizen would have nearly 13 years of education. Meanwhile, developing countries such as Burkina Faso, Niger, and Mali has the worst education quality, where an average citizen only receives about 1 year of education.

In terms of education inequality, countries such as Austria, Barbados, and Denmark are the most equal, while countries such as Burkina Faso, Mali, and Niger have serious inequality.

However, if we look at **improvements**, then countries such as Botswana, Singapore, and Libya see the largest increase in education years of their citizens, while countries such as Tajikistan, North Korea, and Russia see little increase, or even decline, in this regard.

For education inequality, only 7 countries have improvements: Costa Rica, New Zealand, Spain, Trinidad and Tobago, Switzerland, Lebanon, and Seychelles. In other countries, inequality exacerbated, with Nepal, Botswana, and Haiti having the worse situation.