POLI3148 In-Class Exercise 2

Bernadette Marie Coronel La'O 3035957329

2023-10-05

```
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("/Users/berni/Desktop/POLI3148 PORTFOLIO/_DataPublic_/vdem/1984_2022/vdem_1984_2022_extern
```

1. Codebook lookup. Look up the codebook, answer the following questions:

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

```
d |> select(e_peaveduc, e_peedgini)
```

```
## # A tibble: 6,789 x 2
##    e_peaveduc e_peedgini
##         <dbl>      <dbl>
##  1       6.08       32.7
##  2       6.22       32.4
##  3       6.36       31.9
##  4       6.5        31.4
##  5       6.64       31.1
##  6       6.78       30.1
##  7       6.92       30.0
##  8       7.03       29.7
##  9       7.14       29.5
## 10       7.25       29.3
## # i 6,779 more rows
```

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

```
d |> select(country_name, year) |> distinct()
```

```
## # A tibble: 6,789 x 2
##    country_name  year
##    <chr>        <dbl>
##  1 Mexico        1984
##  2 Mexico        1985
##  3 Mexico        1986
##  4 Mexico        1987
##  5 Mexico        1988
##  6 Mexico        1989
##  7 Mexico        1990
##  8 Mexico        1991
##  9 Mexico        1992
## 10 Mexico        1993
## # i 6,779 more rows
```

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

```
# Clio Infra (clio-infra.eu), drawing on Mitchell (1998a, 1998b, 1998c), United States Census Bure
```

2. Subset by columns

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

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

```
## # A tibble: 6,789 x 4
##    country_name  year e_peaveduc e_peedgini
##    <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
```

2. Rename the columns of education quality to make them informative.

```
d_eqr <- d_eq |>
  rename("Country" = "country_name", "Year" = "year", "Education" = "e_peaveduc", "Inequality" = "e
d_eqr
```

```
## # A tibble: 6,789 x 4
##    Country  Year Education Inequality
##    <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
```

3. Subset by rows

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

```
d_eqr |>
  slice_max(order_by = Education, n = 5)
```

```
## # A tibble: 13 x 4
##    Country         Year Education Inequality
##    <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
##  6 United Kingdom  2015      13.3      NA
##  7 United Kingdom  2016      13.3      NA
##  8 United Kingdom  2017      13.3      NA
##  9 United Kingdom  2018      13.3      NA
## 10 United Kingdom  2019      13.3      NA
## 11 United Kingdom  2020      13.3      NA
## 12 United Kingdom  2021      13.3      NA
## 13 United Kingdom  2022      13.3      NA
```

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

```
d_eqr |>
  slice_max(order_by = Inequality, n = 5)
```

```
## # A tibble: 5 x 4
##   Country       Year Education Inequality
##   <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. Summarize the data

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

```
d_eqr |>
  group_by(Country) |>
  mutate(eq_missing = as.numeric(is.na(Education))) |>
  arrange(-eq_missing)
```

```
## # A tibble: 6,789 x 5
## # Groups:   Country [181]
##    Country   Year Education Inequality eq_missing
##    <chr>    <dbl>     <dbl>      <dbl>      <dbl>
##  1 Suriname  1984        NA         NA          1
##  2 Suriname  1985        NA         NA          1
##  3 Suriname  1986        NA         NA          1
##  4 Suriname  1987        NA         NA          1
##  5 Suriname  1988        NA         NA          1
##  6 Suriname  1989        NA         NA          1
##  7 Suriname  1990        NA         NA          1
##  8 Suriname  1991        NA         NA          1
##  9 Suriname  1992        NA         NA          1
## 10 Suriname  1993        NA         NA          1
## # i 6,779 more rows
```

2. Create average level of education quality from 1984 to 2022

```
d_eqr |>
  group_by(Country) |>
  summarise(eq_average = mean(Education, na.rm = TRUE)) |>
  arrange(-eq_average)
```

```
## # A tibble: 181 x 2
##    Country        eq_average
##    <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
```

3. Create change of education quality from 1984 to 2022

```r
d_eqr |>
  group_by(Country) |>
  arrange(Year) |>
  mutate(eq_yoy_change = Education - lag(Education, n=1)) |>
  ungroup() |>
  arrange(Country, Year) |>
  arrange(-eq_yoy_change)
```

```
## # A tibble: 6,789 x 5
##    Country    Year Education Inequality eq_yoy_change
##    <chr>     <dbl>     <dbl>      <dbl>         <dbl>
##  1 Botswana   1985      5.12       45.2         0.394
##  2 Botswana   1987      5.90       41.6         0.394
##  3 Botswana   1988      6.30       39.9         0.394
##  4 Botswana   1989      6.69       38.5         0.394
##  5 Botswana   1990      7.08       37.0         0.394
##  6 Botswana   1986      5.51       43.5         0.393
##  7 Singapore  1998      8.75       32.9         0.258
##  8 Singapore  2000      9.26       32.7         0.258
##  9 Singapore  1991      6.94       37.7         0.258
## 10 Singapore  1992      7.20       36.8         0.258
## # i 6,779 more rows
```

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

```r
# Germany, Australia, and the UK have the highest education averages throughout 1984-2022.
# Botswana, Singapore, and Thailand have had the most improvements.

# Burkina Faso, Niger, and Mali have the lowest education averages throughout 1984-2022.
# Namibia, Russia, and Sweden have had the least improvements.
```