Lec4-5 In-Class Exercise

Luc Mekouar

2023-10-04

Extracting the data from local drive:

```
my_data <- read_csv("/Users/lucmekouar/Desktop/POLI 3148/!!class files/_DataPublic_/vdem/1984_2022/vdem
```

```
## 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. Coodbook lookup. Look up the codebook, answer the following questions:

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

- Educational inequality, Gini (e_peedgini)
- Education 15+ (E) (e_peaveduc)

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

The data covers a total of 202 countries, and spans from 1789 to 2022. Find below a detailed list of which countries are included, and for which year.

```
Countries_and_years_clean <- read_csv("/Users/lucmekouar/Desktop/POLI 3148/
REPO_Luc_POLI3148/week 5__data wrangling.2 & data visualization.1/HW/Countries_
and_years_clean.csv") |>
  view()
```

```
## Rows: 2 Columns: 1
## -- Column specification ---------------------------------------------------
## Delimiter: ","
## chr (1): /Users/lucmekouar/Desktop/POLI 3148/
##
## 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.
```

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

Their sources are mentioned in the Bibliographie section of the document called "codebook". Here is the first source they cite:

" Abad, L. A., Davies, E. A. R. & Luiten Van Zanden, J. (2012), 'Prices and wages in Argentina, Bolivia, Chile, Colombia, Mexico and Peru (Data appendix in: Between conquest and independence: Real wages and demographic change in Spanish America, 1530-1820)', Explorations in Economic History 49(2), 149–166. URL: www.iisg.nl/hpw/ ".

2. Subset by columns

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

Now, I will take only a subset of the V-Dem dataset, namely the file called "vdem_1984_2022_external.csv" to do this exercise as adviced in class.

```
educational_quality <- my_data |> select(country_name, year, e_peedgini,
                                          e_peaveduc) |>
  rename("Country" = "country_name", "Year" = "year")
```

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

```
educational_quality <- educational_quality |>
  rename("Educational inequality" = "e_peedgini", "Education 15+" =
           "e_peaveduc" )
```

3. Subset by rows

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

```
educational_quality |> select(Country, Year, `Education 15+`) |>
  slice_max(order_by = `Education 15+`, n = 5)
```

```
## # A tibble: 13 x 3
##    Country        Year `Education 15+`
##    <chr>         <dbl>          <dbl>
##  1 United Kingdom  2010           13.3
##  2 United Kingdom  2011           13.3
##  3 United Kingdom  2012           13.3
##  4 United Kingdom  2013           13.3
##  5 United Kingdom  2014           13.3
##  6 United Kingdom  2015           13.3
##  7 United Kingdom  2016           13.3
##  8 United Kingdom  2017           13.3
##  9 United Kingdom  2018           13.3
## 10 United Kingdom  2019           13.3
## 11 United Kingdom  2020           13.3
## 12 United Kingdom  2021           13.3
## 13 United Kingdom  2022           13.3
```

```
#13 rows are showing as these rows are all tied in terms of "education 15+"
# (a measure of education level)
```

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

```
educational_quality |>
  slice_max(order_by = `Educational inequality`, n = 5)
```

```
## # A tibble: 5 x 4
##   Country      Year `Educational inequality` `Education 15+`
##   <chr>       <dbl>                    <dbl>           <dbl>
## 1 Burkina Faso 1984                     97.0           0.301
## 2 Burkina Faso 1985                     96.9           0.322
## 3 Burkina Faso 1986                     96.7           0.343
## 4 Burkina Faso 1987                     96.4           0.364
## 5 Burkina Faso 1988                     96.1           0.385
```

4. Summarize the data

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

```
educational_quality |>
  na.omit(educational_quality)
```

```
## # A tibble: 3,378 x 4
##    Country  Year `Educational inequality` `Education 15+`
##    <chr>   <dbl>                    <dbl>           <dbl>
##  1 Mexico   1984                     32.7            6.08
##  2 Mexico   1985                     32.4            6.22
##  3 Mexico   1986                     31.9            6.36
##  4 Mexico   1987                     31.4            6.5
##  5 Mexico   1988                     31.1            6.64
##  6 Mexico   1989                     30.1            6.78
##  7 Mexico   1990                     30.0            6.92
##  8 Mexico   1991                     29.7            7.03
##  9 Mexico   1992                     29.5            7.14
## 10 Mexico   1993                     29.3            7.25
## # i 3,368 more rows
```

ii. Create two types of country-level indicators of education quality

a. Average level of education quality from 1984 to 2022

```
# index_1_avg_edu_qual is not a good measure of quality as it uses data on the
# average number years of education (15+ year old), which is arguably different
# form quality, but is the closest the author could find in the dataset. Only
# one variable could be used as the other measure relevant, education
# inequality, has different units and therefore both variables would have to be
# normalized first before any operation can be done to combine them.
educational_quality |>
```

```
  filter(Year >= 1984 & Year <= 2022) |>
  arrange(Year) |>
  group_by(Country) |>
  summarise(index_1_avg_edu_qual = mean(`Education 15+`), na.rm = TRUE) |>
  ungroup() |>
  arrange(-(index_1_avg_edu_qual))
```

```
## # A tibble: 181 x 3
##    Country        index_1_avg_edu_qual na.rm
##    <chr>                         <dbl> <lgl>
##  1 Germany                        12.9 TRUE
##  2 Australia                      12.9 TRUE
##  3 United Kingdom                 12.9 TRUE
##  4 Canada                         12.7 TRUE
##  5 Switzerland                    12.7 TRUE
##  6 Japan                          12.6 TRUE
##  7 Norway                         12.4 TRUE
##  8 France                         12.0 TRUE
##  9 South Korea                    12.0 TRUE
## 10 New Zealand                    11.9 TRUE
## # i 171 more rows
```

b.Change of education quality from 1984 to 2022

```
educational_quality |>
  filter(Year >= 1984 & Year <= 2022) |>
  arrange(Year) |>
  group_by(Country) |>
  summarise(index_2_chg_edu_qual = last(`Education 15+`) -
              first(`Education 15+`), na.rm = TRUE) |>
  ungroup() |>
  arrange(-(index_2_chg_edu_qual))
```

```
## # A tibble: 181 x 3
##    Country        index_2_chg_edu_qual na.rm
##    <chr>                         <dbl> <lgl>
##  1 Botswana                       5.17 TRUE
##  2 Singapore                      4.52 TRUE
##  3 Libya                          4.07 TRUE
##  4 Cuba                           3.84 TRUE
##  5 Chad                           3.82 TRUE
##  6 Egypt                          3.82 TRUE
##  7 Jordan                         3.82 TRUE
##  8 South Korea                    3.54 TRUE
##  9 Saudi Arabia                   3.49 TRUE
## 10 Algeria                        3.35 TRUE
## # i 171 more rows
```

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

Looking at the average national inequality in education, form 1984 to 2022, measured by the number of years of education done on average by citizens at or over the age of 15, Burkina Faso, Niger and Mali had

the worse education quality, according to VDem data. Moreover, one can also inquire into the largest shift education quality from 1984 to 2022, where Botswana, Singapore and Lybia arrive as the top 3 positive change in education quality. Note that only one country decreased in education quality: Tajikistan. As it stands in 2022, the 5 countries with the largest average number of years of education done on average by citizens at or over the age of 15 can be found below:

```r
educational_quality |> select(Country,Year, `Education 15+`) |>
  filter(Year == 2022) |>
  slice_max(order_by = `Education 15+`, n = 5)
```

```
## # A tibble: 5 x 3
##   Country         Year `Education 15+`
##   <chr>          <dbl>         <dbl>
## 1 United Kingdom  2022          13.3
## 2 Japan           2022          13.2
## 3 Australia       2022          13.1
## 4 Canada          2022          13.1
## 5 South Korea     2022          13
```

Furthermore, from the latest data available (2010), the 5 countries with the least inequality of education (potentially aslo a factor of quality of education) can be found below:

```r
educational_quality |> select(Country, Year, `Educational inequality`) |>
  filter(Year == 2010) |>
  filter(`Educational inequality` > 0) |>
  slice_min(order_by = `Educational inequality`, n = 5)
```

```
## # A tibble: 5 x 3
##   Country         Year `Educational inequality`
##   <chr>          <dbl>                   <dbl>
## 1 Austria         2010                    4.20
## 2 Barbados        2010                    5.42
## 3 United Kingdom  2010                    6.07
## 4 Australia       2010                    6.34
## 5 Denmark         2010                    6.49
```

link to github r markdown file