

In_Class_Exercise_2

Jonathan Standing

2023-10-02

```
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.
```

```
getwd()
```

```
## [1] "/Users/jonathanrexstanding/Desktop/HKU RStudio/HKU_POLI3148_23Fall"
```

```
#In-Class Exercise 2
```

1 - Codebook lookup

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

Education 15+ (average years of education among citizens older than 15) & Educational inequality, Gini (How unequal is the level of education achieved by the population aged 15 years and older?)

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

Years: 1820 - 2022, Geographical coverage: Worldwide

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

Citation: Clio Infra (clio-infra.eu)

2 - Subset by columns

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

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

summary(d_edu)
```

```
## country_name      country_id      year      e_peaveduc
## Length:6789      Min.   : 3.00      Min.   :1984      Min.   : 0.301
## Class :character  1st Qu.: 51.00      1st Qu.:1994      1st Qu.: 4.840
## Mode  :character  Median : 95.00      Median :2003      Median : 7.489
##                      Mean   : 98.41      Mean   :2003      Mean   : 7.360
##                      3rd Qu.:144.00      3rd Qu.:2013      3rd Qu.:10.118
##                      Max.   :236.00      Max.   :2022      Max.   :13.300
##                      NA's    :1774
## e_peedgini
## Min.   : 3.771
## 1st Qu.:18.726
## Median :27.937
## Mean   :34.298
## 3rd Qu.:46.602
## Max.   :96.983
## NA's   :3411
```

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

```
## # A tibble: 181 x 1
##   country_name
##   <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
```

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

```
## # A tibble: 39 x 1
##   year
##   <dbl>
## 1  1984
## 2  1985
## 3  1986
## 4  1987
## 5  1988
## 6  1989
## 7  1990
## 8  1991
## 9  1992
## 10 1993
## # i 29 more rows
```

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

```
d_edu <- d_edu |>
  rename("Country" = "country_name", "ID" = "country_id",
         "Education_Inequality" = "e_peedgini", "Average_Education_15+" = "e_peaveduc",
         "Year" = "year")
summary(d_edu)
```

```
##   Country          ID          Year  Average_Education_15+
## Length:6789      Min.   : 3.00   Min.   :1984   Min.   : 0.301
## Class :character  1st Qu.: 51.00   1st Qu.:1994   1st Qu.: 4.840
## Mode  :character  Median : 95.00   Median :2003   Median : 7.489
##                               Mean  : 98.41   Mean   :2003   Mean   : 7.360
##                               3rd Qu.:144.00   3rd Qu.:2013   3rd Qu.:10.118
##                               Max.   :236.00   Max.   :2022   Max.   :13.300
##                               NA's    :1774
## Education_Inequality
## Min.   : 3.771
## 1st Qu.:18.726
## Median :27.937
## Mean   :34.298
## 3rd Qu.:46.602
## Max.   :96.983
## NA's    :3411
```

3 - Subset by rows

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

```
d_edu |> arrange(-`Average_Education_15+`)
```

```
## # A tibble: 6,789 x 5
##   Country      ID Year 'Average_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
## # i 6,779 more rows
```

```
d_edu |>
  slice_max(order_by = `Average_Education_15+`, n = 5)
```

```
## # A tibble: 13 x 5
##   Country      ID Year 'Average_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
```

United Kingdom = 2010 - 2022 , Japan = 2010 - 2022, Australia = 2010 - 2022, Canada = 2010 - 2022, Germany = 1990

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

```
d_edu |> arrange(-`Education_Inequality`)
```

```
## # A tibble: 6,789 x 5
##   Country      ID Year 'Average_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
```

```
## 6 Burkina Faso 54 1989 0.405 95.2
## 7 Burkina Faso 54 1990 0.426 94.9
## 8 Burkina Faso 54 1991 0.472 94.7
## 9 Burkina Faso 54 1992 0.518 94.4
## 10 Burkina Faso 54 1993 0.564 94.1
## # i 6,779 more rows
```

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

```
## # A tibble: 5 x 5
##   Country      ID Year 'Average_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
```

Burkina Faso, Niger, Somalia, Mali, Nepal

4 - Summarize the data

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

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

```
## # A tibble: 3,378 x 5
##   Country      ID Year 'Average_Education_15+' Education_Inequality
##   <chr>      <dbl> <dbl>                <dbl>                <dbl>
## 1 Mexico      3 1984 6.08 32.7
## 2 Mexico      3 1985 6.22 32.4
## 3 Mexico      3 1986 6.36 31.9
## 4 Mexico      3 1987 6.5 31.4
## 5 Mexico      3 1988 6.64 31.1
## 6 Mexico      3 1989 6.78 30.1
## 7 Mexico      3 1990 6.92 30.0
## 8 Mexico      3 1991 7.03 29.7
## 9 Mexico      3 1992 7.14 29.5
## 10 Mexico     3 1993 7.25 29.3
## # i 3,368 more rows
```

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

```
#Countries with the highest average education 15+
d_edu |>
  filter(Year >= 1984 & Year <= 2022) |>
  arrange(Year) |>
  group_by(Country) |>
  summarise(Average_Education_15 = mean(`Average_Education_15+`, na.rm = TRUE)) |>
  ungroup() |>
  arrange(desc(Average_Education_15))
```

a. Average level of education quality from 1984 to 2022

```
## # 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
```

```
#Countries with the lowest education inequality rates
d_edu |>
  filter(Year >= 1984 & Year <= 2022) |>
  arrange(Year) |>
  group_by(Country) |>
  summarise(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
```

```
#Education inequality change
d_edu |>
  group_by(Country) |>
  arrange(Year) |>
  mutate(edu_inequality_change = `Average_Education_15+` - lag(`Average_Education_15+`, n = 1)) |>
  ungroup() |>
  arrange(Country, Year)
```

b. Change of education quality from 1984 to 2022

```
## # A tibble: 6,789 x 6
##   Country      ID Year `Average_Education_15+` Education_Inequality
##   <chr>      <dbl> <dbl>          <dbl>          <dbl>
## 1 Afghanistan  36 1984          1.30          85.4
## 2 Afghanistan  36 1985          1.35          84.8
## 3 Afghanistan  36 1986          1.40          84.8
## 4 Afghanistan  36 1987          1.45          84.6
## 5 Afghanistan  36 1988          1.50          84.5
## 6 Afghanistan  36 1989          1.55          84.1
## 7 Afghanistan  36 1990          1.60          83.8
## 8 Afghanistan  36 1991          1.69          82.8
## 9 Afghanistan  36 1992          1.78          81.9
## 10 Afghanistan 36 1993          1.88          81.0
## # i 6,779 more rows
## # i 1 more variable: edu_inequality_change <dbl>
```

```
#Average education change compared to 1984
d_edu |>
  group_by(Country) |>
  arrange(Year) |>
  mutate(Average_Education_15_over_1984 = `Average_Education_15+` / first(`Average_Education_15+`)) |>
  ungroup() |>
  arrange(Country, Year)
```

```
## # A tibble: 6,789 x 6
##   Country      ID Year `Average_Education_15+` Education_Inequality
##   <chr>      <dbl> <dbl>          <dbl>          <dbl>
## 1 Afghanistan  36 1984          1.30          85.4
## 2 Afghanistan  36 1985          1.35          84.8
## 3 Afghanistan  36 1986          1.40          84.8
## 4 Afghanistan  36 1987          1.45          84.6
## 5 Afghanistan  36 1988          1.50          84.5
## 6 Afghanistan  36 1989          1.55          84.1
## 7 Afghanistan  36 1990          1.60          83.8
## 8 Afghanistan  36 1991          1.69          82.8
## 9 Afghanistan  36 1992          1.78          81.9
## 10 Afghanistan 36 1993          1.88          81.0
## # i 6,779 more rows
## # i 1 more variable: Average_Education_15_over_1984 <dbl>
```

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? The data shows:

Highest education inequality = Burkino Faso 1984 ^ Average = Burkino Faso Lowest education inequality = Barbados 2008 ^ Average = Austria

Highest average years of education over 15 = United Kingdom 2010 - 2022 ^ Average = Germany Lowest average years of education over 15 = Burkino Faso 1984 ^ Average = Burkino Faso