

```
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.
```

```
view(d)
```

#1. Codebook lookup

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

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

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

- A) For e_peaveduc, the coverage is 1820 - 2022.
- B) For e_peedgini, the coverage is 1850 - 2010.

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

- A) Source for e_peaveduc: Clio Infra (clio-infra.eu), drawing on Mitchell (1998a, 1998b, 1998c), United States Census Bureau (2021), UNESCO, Földvári and van Leeuwen (2014), Leeuwen, van Leeuwen-Li, Földvári (2011), Leeuwen, van Leeuwen-Li, Földvári (2012a), Leeuwen, van Leeuwen-Li, Földvári (2012b), Didenko, Foldvari, van Leeuwen (2012).
- B) Source for e_peedgini: Clio Infra (clio-infra.eu), drawing on Mitchell (1998a, 1998b, 1998c), United States Census Bureau (2021), UNESCO, Földvári and van Leeuwen (2010a), Leeuwen, van Leeuwen-Li, Földvári (2011), Leeuwen, van Leeuwen-Li, Földvári (2012a), Leeuwen, van Leeuwen-Li, Földvári (2012b), Didenko, Foldvari, van Leeuwen (2012). Link: van Leeuwen, B., van Leeuwen-Li, J. & Földvári, P. (2012b), 'Educational Inequality', Working Paper CLIO-INFRA . URL: https://www.clio-infra.eu/docs/Educational_Gini.docx

#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, year, e_peaveduc, e_peedgini)
```

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

```
d_edu <-  
d_edu |>  
  rename("Country" = "country_name", "Year" = "year", "Edu_above_15" = "e_peaveduc", "Edu_inequality" =  
view(d_edu)
```

#3. Subset by rows. i. List 5 countries-years that have the highest education level among its population.

```
d_edu |>  
  slice_max(order_by = Edu_above_15, n = 5)
```

```
## # A tibble: 13 x 4  
##   Country      Year Edu_above_15 Edu_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
```

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

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

```
## # A tibble: 5 x 4  
##   Country      Year Edu_above_15 Edu_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 i. Check data availability: For which countries and years are the indicators of education quality available?

```
d_edu |>
  mutate(N_15_missing = as.numeric(is.na(Edu_above_15)), .after = Edu_above_15) |>
  group_by(Country) |>
  summarize(N_15_missing = sum(N_15_missing))
```

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

```
d_edu |>
  mutate(N_15_missing = as.numeric(is.na(Edu_above_15)), .after = Edu_above_15) |>
  group_by(Year) |>
  summarize(N_15_missing = sum(N_15_missing))
```

```
## # A tibble: 39 x 2
##   Year N_15_missing
##   <dbl>         <dbl>
## 1 1984          40
## 2 1985          40
## 3 1986          40
## 4 1987          40
## 5 1988          40
## 6 1989          41
## 7 1990          42
## 8 1991          43
## 9 1992          44
## 10 1993          45
## # i 29 more rows
```

```
d_edu |>
  mutate(N_Ineq_missing = as.numeric(is.na(Edu_inequality)), .after = Edu_inequality) |>
  group_by(Country) |>
  summarize(N_Ineq_missing = sum(N_Ineq_missing))
```

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

```
d_edu |>
  mutate(N_Ineq_missing = as.numeric(is.na(Edu_inequality)), .after = Edu_inequality) |>
  group_by(Year) |>
  summarize(N_Ineq_missing = sum(N_Ineq_missing))
```

```
## # A tibble: 39 x 2
##   Year N_Ineq_missing
##   <dbl>         <dbl>
## 1 1984          42
## 2 1985          42
## 3 1986          42
## 4 1987          42
## 5 1988          42
## 6 1989          43
## 7 1990          44
## 8 1991          45
## 9 1992          46
## 10 1993          47
## # i 29 more rows
```

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

a. Average level of education quality from 1984 to 2022

```
Edu_above_15_ave <-
  d_edu |>
  group_by(Country) |>
  summarize(Edu_above_15_ave = mean(Edu_above_15, na.rm = TRUE))
Edu_above_15_ave
```

```
## # A tibble: 181 x 2
##   Country      Edu_above_15_ave
##   <chr>         <dbl>
## 1 Afghanistan 2.80
## 2 Albania     NaN
## 3 Algeria     6.31
## 4 Angola      2.46
## 5 Argentina   8.37
## 6 Armenia     10.7
## 7 Australia   12.9
## 8 Austria     11.2
```

```
## 9 Azerbaijan          10.7
## 10 Bahrain            NaN
## # i 171 more rows
```

```
Edu_inequality_ave <-
  d_edu |>
  group_by(Country) |>
  summarize(Edu_inequality_ave = mean(Edu_inequality, na.rm = TRUE))
Edu_inequality_ave
```

```
## # A tibble: 181 x 2
##   Country      Edu_inequality_ave
##   <chr>          <dbl>
## 1 Afghanistan      77.8
## 2 Albania          NaN
## 3 Algeria          45.8
## 4 Angola           53.9
## 5 Argentina        16.6
## 6 Armenia          16.5
## 7 Australia         9.60
## 8 Austria           6.35
## 9 Azerbaijan       14.5
## 10 Bahrain         NaN
## # i 171 more rows
```

b. Change of education quality from 1984 to 2022

```
d_edu |>
  filter(Year >= 1984, Year <= 2022) |>
  group_by(Country) |>
  arrange(Year) |>
  summarize(Edu_change_above_15 = (last(Edu_above_15) - first(Edu_above_15)) / first(Edu_above_15)) |>
  ungroup() |>
  arrange(Country, Edu_change_above_15)
```

```
## # A tibble: 181 x 2
##   Country      Edu_change_above_15
##   <chr>          <dbl>
## 1 Afghanistan      1.94
## 2 Albania          NA
## 3 Algeria          0.847
## 4 Angola           1.22
## 5 Argentina        0.138
## 6 Armenia          0.0321
## 7 Australia        0.0716
## 8 Austria           0.112
## 9 Azerbaijan       0.0239
## 10 Bahrain         NA
## # i 171 more rows
```

```
d_edu |>
  group_by(Country) |>
  arrange(Year) |>
  mutate (Edu_year_dev_above_15 = (Edu_above_15 - lag(Edu_above_15, n = 1))/lag(Edu_above_15, n = 1))|>
  ungroup() |>
  arrange(Country, Year)
```

```
## # A tibble: 6,789 x 5
##   Country      Year Edu_above_15 Edu_inequality Edu_year_dev_above_15
##   <chr>      <dbl>      <dbl>      <dbl>      <dbl>
## 1 Afghanistan 1984         1.30         85.4         NA
## 2 Afghanistan 1985         1.35         84.8         0.0393
## 3 Afghanistan 1986         1.40         84.8         0.0378
## 4 Afghanistan 1987         1.45         84.6         0.0365
## 5 Afghanistan 1988         1.50         84.5         0.0352
## 6 Afghanistan 1989         1.55         84.1         0.0340
## 7 Afghanistan 1990         1.60         83.8         0.0329
## 8 Afghanistan 1991         1.69         82.8         0.0568
## 9 Afghanistan 1992         1.78         81.9         0.0531
## 10 Afghanistan 1993         1.88         81.0         0.0510
## # i 6,779 more rows
```

```
d_edu |>
  filter(Year >= 1984, Year <= 2010) |>
  group_by(Country) |>
  arrange(Year) |>
  summarize(Edu_change_inequality = (last(Edu_inequality) - first(Edu_inequality))/ first(Edu_inequality))
  ungroup() |>
  arrange(Country, Edu_change_inequality)
```

```
## # A tibble: 180 x 2
##   Country      Edu_change_inequality
##   <chr>      <dbl>
## 1 Afghanistan -0.246
## 2 Albania      NA
## 3 Algeria     -0.335
## 4 Angola     -0.440
## 5 Argentina   -0.185
## 6 Armenia     -0.154
## 7 Australia   -0.551
## 8 Austria     -0.575
## 9 Azerbaijan  -0.132
## 10 Bahrain     NA
## # i 170 more rows
```

```
d_edu |>
  group_by(Country) |>
  arrange(Year) |>
  mutate (Edu_year_dev_inequality = (Edu_inequality - lag(Edu_inequality, n = 1))/lag(Edu_inequality, n = 1))|>
  ungroup() |>
  arrange(Country, Year)
```

```
## # A tibble: 6,789 x 5
##   Country      Year Edu_above_15 Edu_inequality Edu_year_dev_inequality
##   <chr>      <dbl>      <dbl>      <dbl>      <dbl>
## 1 Afghanistan 1984         1.30        85.4          NA
## 2 Afghanistan 1985         1.35        84.8        -0.00642
## 3 Afghanistan 1986         1.40        84.8        -0.000637
## 4 Afghanistan 1987         1.45        84.6        -0.00153
## 5 Afghanistan 1988         1.50        84.5        -0.00143
## 6 Afghanistan 1989         1.55        84.1        -0.00557
## 7 Afghanistan 1990         1.60        83.8        -0.00252
## 8 Afghanistan 1991         1.69        82.8        -0.0119
## 9 Afghanistan 1992         1.78        81.9        -0.0115
## 10 Afghanistan 1993         1.88        81.0        -0.0113
## # i 6,779 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?

Very simple presumption: A) The higher the Edu_above_15, the better. B) The lower the Edu_inequality, the better.

Here, I sort out the best-performing countries based on 1. having the top education above 15 year-length IN AVERAGE;

```
Edu_above_15_ave |>
  slice_max(order_by = Edu_above_15_ave, n = 5)
```

```
## # A tibble: 5 x 2
##   Country      Edu_above_15_ave
##   <chr>      <dbl>
## 1 Germany         12.9
## 2 Australia        12.9
## 3 United Kingdom   12.9
## 4 Canada           12.7
## 5 Switzerland      12.7
```

2. having the most growth in the education above 15 year-length (despite the possible short length at the beginning);

```
Edu_change_above_15 <-
  d_edu |>
  filter(Year >= 1984, Year <= 2022) |>
  group_by(Country) |>
  arrange(Year) |>
  summarize(Edu_change_above_15 = (last(Edu_above_15) - first(Edu_above_15)) / first(Edu_above_15)) |>
  ungroup() |>
  arrange(Country, Edu_change_above_15)
Edu_change_above_15 |>
  slice_max(order_by = Edu_change_above_15, n = 5)
```

```
## # A tibble: 5 x 2
##   Country      Edu_change_above_15
```

```
##   <chr>                <dbl>
## 1 Burkina Faso         3.74
## 2 Nepal                2.78
## 3 Afghanistan         1.94
## 4 The Gambia          1.63
## 5 Somalia             1.62
```

3. having the lowest education inequality IN AVERAGE;

```
Edu_inequality_ave |>
  slice_min(order_by = Edu_inequality_ave, n = 5)
```

```
## # A tibble: 5 x 2
##   Country      Edu_inequality_ave
##   <chr>        <dbl>
## 1 Austria     6.35
## 2 Barbados    6.98
## 3 Denmark     8.17
## 4 Switzerland 8.28
## 5 United Kingdom 8.38
```

and 4. having the most drop in the education inequality over time(despite the possible high inequality at the start).

```
Edu_change_inequality <-
  d_edu |>
  filter(Year >= 1984, Year <= 2010) |>
  group_by(Country) |>
  arrange(Year) |>
  summarize(Edu_change_inequality = (last(Edu_inequality) - first(Edu_inequality)) / first(Edu_inequality))
  ungroup() |>
  arrange(Country, Edu_change_inequality)
Edu_change_inequality |>
  slice_max(order_by = Edu_change_inequality, n = 5)
```

```
## # A tibble: 5 x 2
##   Country      Edu_change_inequality
##   <chr>        <dbl>
## 1 Switzerland 0.225
## 2 New Zealand 0.170
## 3 Trinidad and Tobago 0.165
## 4 Costa Rica  0.146
## 5 Spain       0.112
```

Vice versa, the worst performing countries are selected based on 1. having the shortest education above 15 year-length IN AVERAGE;

```
Edu_above_15_ave |>
  slice_min(order_by = Edu_above_15_ave, n = 5)
```

```
## # A tibble: 5 x 2
```



```
##   Country      Edu_above_15_ave
##   <chr>         <dbl>
## 1 Burkina Faso      0.982
## 2 Niger             1.06
## 3 Mali              1.25
## 4 Somalia          1.29
## 5 Burundi          1.86
```

2. having the least growth in the education above 15 year-length (despite the possible long length at the beginning);

```
Edu_change_above_15 <-
  d_edu |>
  filter(Year >= 1984, Year <= 2022) |>
  group_by(Country) |>
  arrange(Year) |>
  summarize(Edu_change_above_15 = (last(Edu_above_15) - first(Edu_above_15)) / first(Edu_above_15)) |>
  ungroup() |>
  arrange(Country, Edu_change_above_15)
Edu_change_above_15 |>
  slice_min(order_by = Edu_change_above_15, n = 5)
```

```
## # A tibble: 5 x 2
##   Country      Edu_change_above_15
##   <chr>         <dbl>
## 1 Tajikistan    -0.0262
## 2 North Korea      0
## 3 Azerbaijan     0.0239
## 4 Russia         0.0245
## 5 Switzerland    0.0265
```

3. having the highest education inequality IN AVERAGE;

```
Edu_inequality_ave |>
  slice_max(order_by = Edu_inequality_ave, n = 5)
```

```
## # A tibble: 5 x 2
##   Country      Edu_inequality_ave
##   <chr>         <dbl>
## 1 Burkina Faso      91.3
## 2 Mali              87.9
## 3 Niger             85.3
## 4 Somalia           84.7
## 5 Afghanistan       77.8
```

- and 4. having the least drop in the education inequality over time (despite the possible low inequality at the start).

```
Edu_change_inequality <-
  d_edu |>
  filter(Year >= 1984, Year <= 2010) |>
```

```

group_by(Country) |>
  arrange(Year) |>
  summarize(Edu_change_inequality = (last(Edu_inequality) - first(Edu_inequality)) / first(Edu_inequality))
ungroup() |>
  arrange(Country, Edu_change_inequality)
Edu_change_inequality |>
  slice_min(order_by = Edu_change_inequality, n = 5)

```

```

## # A tibble: 5 x 2
##   Country      Edu_change_inequality
##   <chr>          <dbl>
## 1 Botswana      -0.724
## 2 Kenya       -0.624
## 3 Austria       -0.575
## 4 China         -0.572
## 5 Australia     -0.551

```

A very brief conclusion: A) Best performing countries include: Switzerland, Germany, and the United Kingdom. B) Worst performing countries include: Burkina Faso, Niger and Mali.