Housekeeping

```
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")
```

Codebook lookup:

Indicators regarding the quality of education

Education 15+ (E) (e_peaveduc)

Educational inequality, Gini (E) (e_peedgini)

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

e_peaveduc: 1820-2022

e_peedgini: 1850-2010

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

e_peaveduc:

https://clio-infra.eu/Indicators/AverageYearsofEducation.html

e_peedgini:

https://clio-infra.eu/Indicators/EducationalInequalityGiniCoefficient.html

Subset by columns

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)
```

Rename the columns of education quality to make them informative.

```
d_edu <- d_edu |> rename ("Country"="country_name", "Year"="year", "Edu15P"="e_peaveduc", "EduIne"="e_p
```

Subset by rows

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

```
d_edu |>
  slice_max(order_by = Edu15P, n = 5)
```

```
##            Country Year Edu15P EduIne
## 1  United Kingdom 2010   13.3  6.072
## 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
```

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

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

```
##        Country Year Edu15P EduIne
## 1 Burkina Faso 1984  0.301 96.983
## 2 Burkina Faso 1985  0.322 96.876
## 3 Burkina Faso 1986  0.343 96.699
## 4 Burkina Faso 1987  0.364 96.428
## 5 Burkina Faso 1988  0.385 96.076
```

Summarize the data

Check data availability

For which countries and years are the indicators of education quality available?

```r
# For which countries are the indicators of education quality available
d_edu |>
  mutate(Edu15P_missing = as.numeric(is.na(Edu15P)), EduIne_missing = as.numeric(is.na(EduIne))) |>
  group_by(Country)|>
  summarise(N_Edu15P_missing = sum(Edu15P_missing), N_EduIne_missing = sum(EduIne_missing))
```

```
## # A tibble: 181 x 3
##    Country     N_Edu15P_missing N_EduIne_missing
##    <chr>                  <dbl>            <dbl>
##  1 Afghanistan                0               12
##  2 Albania                   39               39
##  3 Algeria                    0               12
##  4 Angola                     0               12
##  5 Argentina                  0               12
##  6 Armenia                    0               12
##  7 Australia                  0               12
##  8 Austria                    0               12
##  9 Azerbaijan                 0               12
## 10 Bahrain                   39               39
## # i 171 more rows
```

```r
# For which years are the indicators of education quality available
d_edu |>
  mutate(Edu15P_missing = as.numeric(is.na(Edu15P)), .after= Edu15P, EduIne_missing = as.numeric(is.na(
  group_by(Year)|>
  summarise(N_Edu15P_missing = sum(Edu15P_missing), N_EduIne_missing = sum(EduIne_missing))
```

```
## # A tibble: 39 x 3
##     Year N_Edu15P_missing N_EduIne_missing
##    <int>            <dbl>            <dbl>
##  1  1984               40               42
##  2  1985               40               42
##  3  1986               40               42
##  4  1987               40               42
##  5  1988               40               42
##  6  1989               41               43
##  7  1990               42               44
##  8  1991               43               45
##  9  1992               44               46
## 10  1993               45               47
## # i 29 more rows
```

Create two types of country-level indicators of education quality

1. Average level of education quality from 1984 to 2022

2. Change of education quality from 1984 to 2022

```r
# Calculate the average level of education quality
# Edu15P: from 1984-2022
# EduIne: from 1984-2010 (Seen the available time range in codebook)
```

```r
d_edu |>
  group_by(Country)|>
  summarise(Edu15P_average = mean (Edu15P, na.rm = TRUE), EduIne_average = mean (EduIne, na.rm = TRUE))
```

```
## # A tibble: 181 x 3
##    Country     Edu15P_average EduIne_average
##    <chr>                <dbl>          <dbl>
##  1 Afghanistan           2.80          77.8
##  2 Albania             NaN            NaN
##  3 Algeria               6.31          45.8
##  4 Angola                2.46          53.9
##  5 Argentina             8.37          16.6
##  6 Armenia              10.7           16.5
##  7 Australia            12.9            9.60
##  8 Austria              11.2            6.35
##  9 Azerbaijan           10.7           14.5
## 10 Bahrain             NaN            NaN
## # i 171 more rows
```

```r
# Calculate the change of education quality (year-over-year)
d_edu |>
  group_by(Country) |>
  arrange(Year) |>
  mutate (Edu15P_yoy_change = Edu15P - lag(Edu15P, n=1), .after = Edu15P) |>
  mutate (EduIne_yoy_change = EduIne - lag(EduIne, n=1), .after = EduIne) |>
  ungroup() |>
  arrange(Country, Year)
```

```
## # A tibble: 6,789 x 6
##    Country      Year Edu15P Edu15P_yoy_change EduIne EduIne_yoy_change
##    <chr>       <int>  <dbl>             <dbl>  <dbl>             <dbl>
##  1 Afghanistan  1984   1.30           NA        85.4           NA
##  2 Afghanistan  1985   1.35            0.0510   84.8           -0.548
##  3 Afghanistan  1986   1.40            0.0510   84.8           -0.0540
##  4 Afghanistan  1987   1.45            0.0510   84.6           -0.130
##  5 Afghanistan  1988   1.50            0.0510   84.5           -0.121
##  6 Afghanistan  1989   1.55            0.0510   84.1           -0.471
##  7 Afghanistan  1990   1.60            0.0510   83.8           -0.212
##  8 Afghanistan  1991   1.69            0.091    82.8           -1
##  9 Afghanistan  1992   1.78            0.0900   81.9           -0.951
## 10 Afghanistan  1993   1.88            0.091    81.0           -0.923
## # i 6,779 more rows
```

```r
# Calculate the change of education quality (overall)
# Edu15P: from 1984-2022
# EduIne: from 1984-2010 (Seen the available time range in codebook)
d_edu |>
  group_by(Country) |>
  arrange(Year) |>
  summarise(Change_Edu15P = last(Edu15P) - first(Edu15P))
```

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

```
##    Country     Change_Edu15P
##    <chr>               <dbl>
##  1 Afghanistan         2.52
##  2 Albania            NA
##  3 Algeria             3.35
##  4 Angola              1.64
##  5 Argentina           1.06
##  6 Armenia             0.336
##  7 Australia           0.878
##  8 Austria             1.16
##  9 Azerbaijan          0.252
## 10 Bahrain            NA
## # i 171 more rows
```

```
d_edu |>
  filter (Year >= 1984 & Year <=2010) |>
  group_by(Country) |>
  arrange(Year) |>
  summarise(Change_EduIne = last(EduIne) - first(EduIne))
```

```
## # A tibble: 180 x 2
##    Country     Change_EduIne
##    <chr>               <dbl>
##  1 Afghanistan        -21
##  2 Albania             NA
##  3 Algeria            -18.9
##  4 Angola             -29.5
##  5 Argentina           -3.56
##  6 Armenia             -2.87
##  7 Australia           -7.77
##  8 Austria             -5.68
##  9 Azerbaijan          -2.16
## 10 Bahrain             NA
## # i 170 more rows
```

Examine the data and *briefly* discuss

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

```
d_edu |>
  group_by(Country) |>
  summarise (Edu15P_mean = mean (Edu15P)) |>
  arrange(Edu15P_mean)
```

```
## # A tibble: 181 x 2
##    Country      Edu15P_mean
##    <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
```

Burkina Faso performs the worst while Germany performs the best in terms of education level 15+.

```
d_edu |>
  filter (Year >= 1984 & Year <= 2010)|>
  group_by(Country) |>
  summarise (EduIne_mean = mean (EduIne)) |>
  arrange(EduIne_mean, na.rm = TRUE)
```

```
## # A tibble: 180 x 2
##    Country        EduIne_mean
##    <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 170 more rows
```

Burkina Faso performs the worst while Austria performs the best in terms of education inequality.