in class assignment 2

Yang Xinrui

```
library(tidyverse)
```

```
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.
```

1. Codebook lookup.

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

Education 15+ (e_peaveduc) ; Educational inequality, Gini (e_peedgini)

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

```
# countries
d |> 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
```

```
# years
d |> 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
```

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

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) |> distinct()

d_edu
```

```
## # A tibble: 6,789 x 5
##    country_name country_id  year e_peaveduc e_peedgini
##    <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 6,779 more rows
```

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

```
d_edu |>
  rename("Education" = "e_peaveduc", "Gini" = "e_peedgini",
         "Country" = "country_name", "ID" = "country_id",
         "Year" = "year")
```

```
## # A tibble: 6,789 x 5
##    Country    ID  Year Education  Gini
##    <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 6,779 more rows
```

```
d_edu <- d_edu |>
  rename("Education" = "e_peaveduc", "Gini" = "e_peedgini",
         "Country" = "country_name", "ID" = "country_id",
         "Year" = "year")

d_edu
```

```
## # A tibble: 6,789 x 5
##    Country    ID  Year Education  Gini
##    <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 6,779 more rows
```

3. Subset by rows

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

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

```
## # A tibble: 13 x 5
##    Country           ID  Year Education  Gini
##    <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
```

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

```
d_edu |> slice_min(order_by = Gini, n =5)
```

```
## # A tibble: 5 x 5
##   Country     ID  Year Education  Gini
##   <chr>    <dbl> <dbl>     <dbl> <dbl>
## 1 Barbados   147  2008      9.57  3.77
## 2 Barbados   147  2003      9.32  3.80
## 3 Barbados   147  2007      9.52  4.01
## 4 Austria    144  2007     11.4   4.03
## 5 Austria    144  2008     11.4   4.04
```

4. Summarize the data

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

```
d_edu |>
  mutate(GDP_missing = is.na(Education), .after = Education) |>
  group_by(Country) |>
  summarise(N_GDP_missing = sum(GDP_missing))
```

```
## # A tibble: 181 x 2
##    Country     N_GDP_missing
##    <chr>               <int>
##  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(GDP_missing = is.na(Gini), .after = Gini) |>
  group_by(Country) |>
  summarise(N_GDP_missing = sum(GDP_missing))
```

```
## # A tibble: 181 x 2
##    Country     N_GDP_missing
##    <chr>               <int>
##  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
```

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

a. Average level of education quality from 1984 to 2022

```
d_edu |>
  group_by(Country) |>
  summarise(Education_average = mean(Education, na.rm = TRUE)) |>
  arrange(Country)
```

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

b. Change of education quality from 1984 to 2022

```
d_edu |>
  filter(Year >= 1984 & Year <= 2022) |>
  group_by(Country) |>
  arrange(Year) |>
  summarise(Education_growth_2022_1984 = (last(Education) - first(Education)) / first(Education)) |>
  ungroup() |>
  arrange(Country)
```

```
## # A tibble: 181 x 2
##    Country     Education_growth_2022_1984
##    <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
```

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?

```
d_edu |>
  group_by(Country) |>
  summarise(Education_average = mean(Education, na.rm = TRUE)) |>
  arrange(Education_average)
```

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

```
d_edu |>
  group_by(Country) |>
  summarise(Education_average = mean(Education, na.rm = TRUE)) |>
  arrange(desc(Education_average))
```

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

```
d_edu |>
  filter(Year >= 1984 & Year <= 2022) |>
  group_by(Country) |>
  arrange(Year) |>
  summarise(Education_growth_2022_1984 = (last(Education) - first(Education)) / first(Education)) |>
  ungroup() |>
  arrange(Education_growth_2022_1984)
```

```
## # A tibble: 181 x 2
##    Country     Education_growth_2022_1984
##    <chr>                            <dbl>
##  1 Tajikistan                     -0.0262
##  2 North Korea                     0
##  3 Azerbaijan                      0.0239
##  4 Russia                          0.0245
##  5 Switzerland                     0.0265
##  6 Uzbekistan                      0.0271
##  7 Germany                         0.0277
##  8 Kyrgyzstan                      0.0303
##  9 Armenia                         0.0321
## 10 Georgia                         0.0368
## # i 171 more rows
```

```
d_edu |>
  filter(Year >= 1984 & Year <= 2022) |>
  group_by(Country) |>
  arrange(Year) |>
  summarise(Education_growth_2022_1984 = (last(Education) - first(Education)) / first(Education)) |>
  ungroup() |>
  arrange(desc(Education_growth_2022_1984))
```

```
## # A tibble: 181 x 2
##    Country      Education_growth_2022_1984
##    <chr>                             <dbl>
##  1 Burkina Faso                       3.74
##  2 Nepal                              2.78
##  3 Afghanistan                        1.94
##  4 The Gambia                         1.63
##  5 Somalia                            1.62
##  6 Chad                               1.57
##  7 Niger                              1.43
##  8 Burundi                            1.32
##  9 Nigeria                            1.27
## 10 Liberia                            1.26
## # i 171 more rows
```