

In-class exercise

Xinhe WANG (Xavier)

2023-10-06

Codebook lookup

What indicators regarding the quality of education are available in the V-Dem datasets? GDP per capita

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

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

```
## # A tibble: 202 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 192 more rows
```

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

```
## # A tibble: 234 x 1
##   year
##   <dbl>
## 1 1789
## 2 1790
## 3 1791
## 4 1792
## 5 1793
## 6 1794
## 7 1795
## 8 1796
## 9 1797
## 10 1798
## # i 224 more rows
```

What are their sources? Provide the link to least 1 source. <https://v-dem.net/data/the-v-dem-dataset/>

Subset by columns

Create a dataset containing only the country-year identifiers and indicators of education quality. Rename the columns of education quality to make them informative.

```
d_gdp <- d |>
  select(country_name, year, e_gdppc)
d_gdp <- d_gdp |>
  rename("Education_level" = "e_gdppc",
         "Country" = "country_name", "Year" = "year")
```

Subset by rows

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

```
d_gdp |> slice_max(order_by = Education_level, n = 5)
```

```
## # A tibble: 5 x 3
##   Country      Year Education_level
##   <chr>      <dbl>      <dbl>
## 1 United Arab Emirates 1974      157.
## 2 United Arab Emirates 1973      156.
## 3 United Arab Emirates 1972      154.
## 4 United Arab Emirates 1980      153.
## 5 United Arab Emirates 1975      153.
```

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

```
d_gdp |> slice_min(order_by = Education_level, n = 5)
```

```
## # A tibble: 5 x 3
##   Country Year Education_level
##   <chr>   <dbl>      <dbl>
## 1 Liberia 1995      0.286
## 2 Liberia 1994      0.307
## 3 Liberia 1996      0.309
## 4 Liberia 1993      0.383
## 5 Romania 1865      0.419
```

Summarize the data

Data availability

```
d_gdp |>
  mutate(Education_level_missing = as.numeric(is.na(Education_level)), .after = Education_level) |>
  group_by(Country) |>
  summarise(N_Education_level_missing = sum(Education_level_missing))
```

```
## # A tibble: 202 x 2
##   Country      N_Education_level_missing
##   <chr>      <dbl>
## 1 Afghanistan      30
## 2 Albania           3
## 3 Algeria           3
## 4 Angola           53
## 5 Argentina        14
## 6 Armenia           4
## 7 Australia         3
```

```
## 8 Austria 3
## 9 Azerbaijan 3
## 10 Baden 27
## # i 192 more rows
```

Create two types of country-level indicators of education quality

a.Average level of education quality from 1984 to 2022 b.Change of education quality from 1984 to 2022

```
d_gdp |> summarise(Education_level_average = mean(Education_level, na.rm = TRUE))
```

```
## # A tibble: 1 x 1
##   Education_level_average
##               <dbl>
## 1                6.42
```

```
d_gdp |>
  filter(Year >= 1984 & Year <= 2022) |>
  group_by(Country) |>
  arrange(Year) |>
  summarise(Education_level_growth_2022_1984 = (last(Education_level) - first(Education_level)) / first
  ungroup() |>
  arrange(Country)
```

```
## # A tibble: 181 x 2
##   Country      Education_level_growth_2022_1984
##   <chr>                <dbl>
## 1 Afghanistan          NA
## 2 Albania              NA
## 3 Algeria              NA
## 4 Angola               NA
## 5 Argentina            NA
## 6 Armenia              NA
## 7 Australia            NA
## 8 Austria              NA
## 9 Azerbaijan           NA
## 10 Bahrain             NA
## # i 171 more rows
```