POLI3148\_In\_class\_exercise\_2

Yujia Tao

2023-10-09

## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

summary(cars)

## speed dist   
## Min. : 4.0 Min. : 2.00   
## 1st Qu.:12.0 1st Qu.: 26.00   
## Median :15.0 Median : 36.00   
## Mean :15.4 Mean : 42.98   
## 3rd Qu.:19.0 3rd Qu.: 56.00   
## Max. :25.0 Max. :120.00

## Including Plots

You can also embed plots, for example:



Note that the echo = FALSE parameter was added to the code chunk to prevent printing of the R code that generated the plot.

# 1. Looking up the codebook

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

#There are two indicators regarding the quality of education available in the V-Dem data sets:

#Education15+(E)(e\_peaveduc), The average years of education in the total population aged 15 years and older.

#Educational inequality,Gini(E)(e\_peedgini), Gini coefficient of educational inequality estimated from average education data.

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 1.9.3 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

## #ii. What are the data’s coverage?

#Education15+

#Educational inequality

d<- read.csv("\_DataPublic\_/vdem/1984\_2022/vdem\_1984\_2022\_external.csv")

#Want: Countries covered in the data set  
d |> select(country\_name) |>   
 distinct()

## country\_name  
## 1 Mexico  
## 2 Suriname  
## 3 Sweden  
## 4 Switzerland  
## 5 Ghana  
## 6 South Africa  
## 7 Japan  
## 8 Burma/Myanmar  
## 9 Russia  
## 10 Albania  
## 11 Egypt  
## 12 Yemen  
## 13 Colombia  
## 14 Poland  
## 15 Brazil  
## 16 United States of America  
## 17 Portugal  
## 18 El Salvador  
## 19 South Yemen  
## 20 Bangladesh  
## 21 Bolivia  
## 22 Haiti  
## 23 Honduras  
## 24 Mali  
## 25 Pakistan  
## 26 Peru  
## 27 Senegal  
## 28 South Sudan  
## 29 Sudan  
## 30 Vietnam  
## 31 Afghanistan  
## 32 Argentina  
## 33 Ethiopia  
## 34 India  
## 35 Kenya  
## 36 North Korea  
## 37 South Korea  
## 38 Kosovo  
## 39 Lebanon  
## 40 Nigeria  
## 41 Philippines  
## 42 Tanzania  
## 43 Taiwan  
## 44 Thailand  
## 45 Uganda  
## 46 Venezuela  
## 47 Benin  
## 48 Bhutan  
## 49 Burkina Faso  
## 50 Cambodia  
## 51 Indonesia  
## 52 Mozambique  
## 53 Nepal  
## 54 Nicaragua  
## 55 Niger  
## 56 Zambia  
## 57 Zimbabwe  
## 58 Guinea  
## 59 Ivory Coast  
## 60 Mauritania  
## 61 Canada  
## 62 Australia  
## 63 Botswana  
## 64 Burundi  
## 65 Cape Verde  
## 66 Central African Republic  
## 67 Chile  
## 68 Costa Rica  
## 69 Timor-Leste  
## 70 Ecuador  
## 71 France  
## 72 Germany  
## 73 Guatemala  
## 74 Iran  
## 75 Iraq  
## 76 Ireland  
## 77 Italy  
## 78 Jordan  
## 79 Latvia  
## 80 Lesotho  
## 81 Liberia  
## 82 Malawi  
## 83 Maldives  
## 84 Mongolia  
## 85 Morocco  
## 86 Netherlands  
## 87 Panama  
## 88 Papua New Guinea  
## 89 Qatar  
## 90 Sierra Leone  
## 91 Spain  
## 92 Syria  
## 93 Tunisia  
## 94 Turkey  
## 95 Ukraine  
## 96 United Kingdom  
## 97 Uruguay  
## 98 Algeria  
## 99 Angola  
## 100 Armenia  
## 101 Azerbaijan  
## 102 Belarus  
## 103 Cameroon  
## 104 Chad  
## 105 China  
## 106 Democratic Republic of the Congo  
## 107 Republic of the Congo  
## 108 Djibouti  
## 109 Dominican Republic  
## 110 Eritrea  
## 111 Gabon  
## 112 The Gambia  
## 113 Georgia  
## 114 Guinea-Bissau  
## 115 Jamaica  
## 116 Kazakhstan  
## 117 Kyrgyzstan  
## 118 Laos  
## 119 Libya  
## 120 Madagascar  
## 121 Moldova  
## 122 Namibia  
## 123 Palestine/West Bank  
## 124 Rwanda  
## 125 Somalia  
## 126 Sri Lanka  
## 127 Eswatini  
## 128 Tajikistan  
## 129 Togo  
## 130 Trinidad and Tobago  
## 131 Turkmenistan  
## 132 German Democratic Republic  
## 133 Palestine/Gaza  
## 134 Somaliland  
## 135 Uzbekistan  
## 136 Austria  
## 137 Bahrain  
## 138 Barbados  
## 139 Belgium  
## 140 Bosnia and Herzegovina  
## 141 Bulgaria  
## 142 Comoros  
## 143 Croatia  
## 144 Cuba  
## 145 Cyprus  
## 146 Czechia  
## 147 Denmark  
## 148 Equatorial Guinea  
## 149 Estonia  
## 150 Fiji  
## 151 Finland  
## 152 Greece  
## 153 Guyana  
## 154 Hong Kong  
## 155 Iceland  
## 156 Israel  
## 157 Kuwait  
## 158 Lithuania  
## 159 Luxembourg  
## 160 North Macedonia  
## 161 Malaysia  
## 162 Malta  
## 163 Mauritius  
## 164 Montenegro  
## 165 New Zealand  
## 166 Norway  
## 167 Oman  
## 168 Paraguay  
## 169 Romania  
## 170 Sao Tome and Principe  
## 171 Saudi Arabia  
## 172 Serbia  
## 173 Seychelles  
## 174 Singapore  
## 175 Slovakia  
## 176 Slovenia  
## 177 Solomon Islands  
## 178 Vanuatu  
## 179 United Arab Emirates  
## 180 Hungary  
## 181 Zanzibar

#Want: Years covered in the data set  
d |> select(year)|> distinct()

## year  
## 1 1984  
## 2 1985  
## 3 1986  
## 4 1987  
## 5 1988  
## 6 1989  
## 7 1990  
## 8 1991  
## 9 1992  
## 10 1993  
## 11 1994  
## 12 1995  
## 13 1996  
## 14 1997  
## 15 1998  
## 16 1999  
## 17 2000  
## 18 2001  
## 19 2002  
## 20 2003  
## 21 2004  
## 22 2005  
## 23 2006  
## 24 2007  
## 25 2008  
## 26 2009  
## 27 2010  
## 28 2011  
## 29 2012  
## 30 2013  
## 31 2014  
## 32 2015  
## 33 2016  
## 34 2017  
## 35 2018  
## 36 2019  
## 37 2020  
## 38 2021  
## 39 2022

## iii. Sources

#Clio-Infra (2018), ‘Clio-Infra Project (Database)’. URL: <http://www.clio-infra.eu/>

#Foldvari,P.&vanLeeuwen,B (2014),‘EducationalandincomeinequalityinEurope,ca.1870-2000’, Cliometrica8(3),271–300. URL:<https://link.springer.com/content/pdf/10.1007%2Fs11698-013-0105-3.pdf>

# 2. Subset by columns

## #i.

#Want: 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)

## #ii.

#Want:Rename the columns of education quality to make them informative.  
d\_edu <- d\_edu |>  
 rename("Edu15" = "e\_peaveduc", "Inequality" = "e\_peedgini", "Country" = "country\_name", "ID" = "country\_id", "Year" = "year")  
view(d\_edu)

# 3.Subset by rows

## #i.

#Want:List 5 countries-years that have the highest education level among its population.  
d\_edu|> slice\_max(order\_by= Edu15, n = 5)

## Country ID Year Edu15 Inequality  
## 1 United Kingdom 101 2010 13.3 6.072  
## 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.

#Want:List 5 countries-years that have suffered from most severe inequality in education.  
d\_edu|> slice\_max(order\_by=Inequality, n = 5)

## Country ID Year Edu15 Inequality  
## 1 Burkina Faso 54 1984 0.301 96.983  
## 2 Burkina Faso 54 1985 0.322 96.876  
## 3 Burkina Faso 54 1986 0.343 96.699  
## 4 Burkina Faso 54 1987 0.364 96.428  
## 5 Burkina Faso 54 1988 0.385 96.076

# 4.Summarize the data

## #i.Check data availability

### #Missing data in Education 15+data set

# Missing data of each country  
d\_edu |>  
 mutate(Edu15\_missing = as.numeric(is.na(Edu15)), .after = Edu15) |>  
 group\_by(Country) |>  
 summarise(N\_Edu15\_missing = sum(Edu15\_missing))

## # A tibble: 181 x 2  
## Country N\_Edu15\_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

# Missing data of each year  
d\_edu|>   
 mutate(Edu15\_missing = as.numeric(is.na(Edu15)), .after = Edu15) |>  
 group\_by(Year) |>  
 summarise(N\_Edu15\_missing = sum(Edu15\_missing))

## # A tibble: 39 x 2  
## Year N\_Edu15\_missing  
## <int> <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

### #Missing data in Educational Inequality data set

# Missing data of each country  
d\_edu |>  
 mutate(Inequality\_missing = as.numeric(is.na(Inequality)), .after = Inequality) |>  
 group\_by(Country) |>  
 summarise(N\_Inequality\_missing = sum(Inequality\_missing))

## # A tibble: 181 x 2  
## Country N\_Inequality\_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

# Missing data of each year  
d\_edu|>   
 mutate(Inequality\_missing = as.numeric(is.na(Inequality)), .after = Inequality) |>  
 group\_by(Year) |>  
 summarise(N\_Inequality\_missing = sum(Inequality\_missing))

## # A tibble: 39 x 2  
## Year N\_Inequality\_missing  
## <int> <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  
d\_edu |>  
 group\_by(Country) |>  
 summarise(Edu15\_average = mean(Edu15, na.rm = TRUE),  
 Inequality\_average = mean(Inequality, na.rm = TRUE))

## # A tibble: 181 x 3  
## Country Edu15\_average Inequality\_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

## #b.

#Change of education quality from 1984 to 2022  
#Year-on-year change of education quality from 1984 to 2022  
#Education 15+  
d\_edu |>  
 group\_by(Country) |>  
 arrange(Year) |>  
 mutate(Edu15\_yoy\_change = Edu15 - lag(Edu15, n = 1)) |>  
 ungroup() |>  
 arrange(Country, Year)

## # A tibble: 6,789 x 6  
## Country ID Year Edu15 Inequality Edu15\_yoy\_change  
## <chr> <int> <int> <dbl> <dbl> <dbl>  
## 1 Afghanistan 36 1984 1.30 85.4 NA   
## 2 Afghanistan 36 1985 1.35 84.8 0.0510  
## 3 Afghanistan 36 1986 1.40 84.8 0.0510  
## 4 Afghanistan 36 1987 1.45 84.6 0.0510  
## 5 Afghanistan 36 1988 1.50 84.5 0.0510  
## 6 Afghanistan 36 1989 1.55 84.1 0.0510  
## 7 Afghanistan 36 1990 1.60 83.8 0.0510  
## 8 Afghanistan 36 1991 1.69 82.8 0.091   
## 9 Afghanistan 36 1992 1.78 81.9 0.0900  
## 10 Afghanistan 36 1993 1.88 81.0 0.091   
## # i 6,779 more rows

#Educational Inequality  
d\_edu |>  
 group\_by(Country) |>  
 arrange(Year) |>  
 mutate(Inequality\_yoy\_change = Inequality - lag(Inequality, n = 1)) |>  
 ungroup() |>  
 arrange(Country, Year)

## # A tibble: 6,789 x 6  
## Country ID Year Edu15 Inequality Inequality\_yoy\_change  
## <chr> <int> <int> <dbl> <dbl> <dbl>  
## 1 Afghanistan 36 1984 1.30 85.4 NA   
## 2 Afghanistan 36 1985 1.35 84.8 -0.548   
## 3 Afghanistan 36 1986 1.40 84.8 -0.0540  
## 4 Afghanistan 36 1987 1.45 84.6 -0.130   
## 5 Afghanistan 36 1988 1.50 84.5 -0.121   
## 6 Afghanistan 36 1989 1.55 84.1 -0.471   
## 7 Afghanistan 36 1990 1.60 83.8 -0.212   
## 8 Afghanistan 36 1991 1.69 82.8 -1   
## 9 Afghanistan 36 1992 1.78 81.9 -0.951   
## 10 Afghanistan 36 1993 1.88 81.0 -0.923   
## # i 6,779 more rows

# Country-year education quality with reference to that of 1984.  
  
#Education 15+  
d\_edu |>  
 group\_by(Country) |>  
 arrange(Year) |>  
 mutate(Edu15\_over\_1984 = Edu15 / first(Edu15)) |>  
 ungroup() |>  
 arrange(Country, Year)

## # A tibble: 6,789 x 6  
## Country ID Year Edu15 Inequality Edu15\_over\_1984  
## <chr> <int> <int> <dbl> <dbl> <dbl>  
## 1 Afghanistan 36 1984 1.30 85.4 1   
## 2 Afghanistan 36 1985 1.35 84.8 1.04  
## 3 Afghanistan 36 1986 1.40 84.8 1.08  
## 4 Afghanistan 36 1987 1.45 84.6 1.12  
## 5 Afghanistan 36 1988 1.50 84.5 1.16  
## 6 Afghanistan 36 1989 1.55 84.1 1.20  
## 7 Afghanistan 36 1990 1.60 83.8 1.24  
## 8 Afghanistan 36 1991 1.69 82.8 1.31  
## 9 Afghanistan 36 1992 1.78 81.9 1.38  
## 10 Afghanistan 36 1993 1.88 81.0 1.45  
## # i 6,779 more rows

#Educational Inequality  
d\_edu |>  
 group\_by(Country) |>  
 arrange(Year) |>  
 mutate(Inequality\_over\_1984 = Inequality / first(Inequality)) |>  
 ungroup() |>  
 arrange(Country, Year)

## # A tibble: 6,789 x 6  
## Country ID Year Edu15 Inequality Inequality\_over\_1984  
## <chr> <int> <int> <dbl> <dbl> <dbl>  
## 1 Afghanistan 36 1984 1.30 85.4 1   
## 2 Afghanistan 36 1985 1.35 84.8 0.994  
## 3 Afghanistan 36 1986 1.40 84.8 0.993  
## 4 Afghanistan 36 1987 1.45 84.6 0.991  
## 5 Afghanistan 36 1988 1.50 84.5 0.990  
## 6 Afghanistan 36 1989 1.55 84.1 0.984  
## 7 Afghanistan 36 1990 1.60 83.8 0.982  
## 8 Afghanistan 36 1991 1.69 82.8 0.970  
## 9 Afghanistan 36 1992 1.78 81.9 0.959  
## 10 Afghanistan 36 1993 1.88 81.0 0.948  
## # i 6,779 more rows

# Education 15+ and Educational Inequality change: comparing 2022 with 1984  
d\_edu |>  
 filter(Year >= 1984 & Year <= 2022) |>  
 group\_by(Country) |>  
 arrange(Year) |>  
 summarise(Edu15\_growth\_2022\_1984 = (last(Edu15) - first(Edu15)) / first(Edu15),  
 Inequality\_2022\_1984 = (last(Inequality) - first(Inequality)) / first(Inequality)) |>  
 ungroup() |>  
 arrange(Country)

## # A tibble: 181 x 3  
## Country Edu15\_growth\_2022\_1984 Inequality\_2022\_1984  
## <chr> <dbl> <dbl>  
## 1 Afghanistan 1.94 NA  
## 2 Albania NA NA  
## 3 Algeria 0.847 NA  
## 4 Angola 1.22 NA  
## 5 Argentina 0.138 NA  
## 6 Armenia 0.0321 NA  
## 7 Australia 0.0716 NA  
## 8 Austria 0.112 NA  
## 9 Azerbaijan 0.0239 NA  
## 10 Bahrain NA 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?

#The country which perform the best in terms of education quality in the past four decades is United Kingdom, because it has 12.88 average years of education in the total population aged 15 years and older, which is the second longest among all countries in the data set. At the same time, Australia has the longest average years of education in the total population aged 15 years and older (12.94), but its average inequality coefficient is around 9.60 which is higher than 8.38.

#In contrast, the country which perform the worst in terms of education quality in the past four decades is Burkina Faso, because it has lowest average level of education in its population, which is around 0.98, in the past four decades among all countries covered in the data set . Also, it suffered from highest average inequality level in education, which the coefficient reached 91.26, above all other countries covered in the data set.