inclass_exercise_2

Wenxiao HU

Coodbook lookup

1. Indicators include *Eduacation 15+* (e_peaveduc) and *Educational inequality, Gini* (e_peedgini).
2. For *Eduacation 15+* (e_peaveduc), the coverage is from 1820-2022.

   For *Educational inequality, Gini* (e_peedgini), the coverage is from 1850-2010.
3. For *Eduacation 15+* (e_peaveduc), one of the links to the sources is Clio-Infra (2018), 'Clio-Infra Project (Database)'. URL: http://www.clio-infra.eu/

   For *Educational inequality, Gini* (e_peedgini), one of the links to the sources is van Leeuwen, B. V., Van Leeuwen-Li, J. & Földvári, P. (2012a), 'Education as a driver of income inequality in twentieth-century Africa', MPRA Paper 43574, University Library of Munich, Germany . URL: https://mpra.ub.uni-muenchen.de/43574/

Subset by column

```
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 error
```

```
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.
```

```
names(d)
##   [1] "country_name"              "country_text_id"
##   [3] "country_id"                "year"
##   [5] "historical_date"           "project"
##   [7] "historical"                "histname"
##   [9] "codingstart"               "codingend"
##  [11] "codingstart_contemp"       "codingend_contemp"
##  [13] "codingstart_hist"          "codingend_hist"
##  [15] "gapstart1"                 "gapstart2"
##  [17] "gapstart3"                 "gapend1"
##  [19] "gapend2"                   "gapend3"
##  [21] "gap_index"                 "COWcode"
##  [23] "e_v2x_api_3C"              "e_v2x_api_4C"
##  [25] "e_v2x_api_5C"              "e_v2x_civlib_3C"
##  [27] "e_v2x_civlib_4C"           "e_v2x_civlib_5C"
##  [29] "e_v2x_clphy_3C"            "e_v2x_clphy_4C"
##  [31] "e_v2x_clphy_5C"            "e_v2x_clpol_3C"
##  [33] "e_v2x_clpol_4C"            "e_v2x_clpol_5C"
##  [35] "e_v2x_clpriv_3C"           "e_v2x_clpriv_4C"
##  [37] "e_v2x_clpriv_5C"           "e_v2x_corr_3C"
##  [39] "e_v2x_corr_4C"             "e_v2x_corr_5C"
##  [41] "e_v2x_cspart_3C"           "e_v2x_cspart_4C"
##  [43] "e_v2x_cspart_5C"           "e_v2x_delibdem_3C"
##  [45] "e_v2x_delibdem_4C"         "e_v2x_delibdem_5C"
##  [47] "e_v2x_EDcomp_thick_3C"     "e_v2x_EDcomp_thick_4C"
##  [49] "e_v2x_EDcomp_thick_5C"     "e_v2x_egal_3C"
##  [51] "e_v2x_egal_4C"             "e_v2x_egal_5C"
##  [53] "e_v2x_egaldem_3C"          "e_v2x_egaldem_4C"
##  [55] "e_v2x_egaldem_5C"          "e_v2x_elecoff_3C"
##  [57] "e_v2x_elecoff_4C"          "e_v2x_elecoff_5C"
##  [59] "e_v2x_execorr_3C"          "e_v2x_execorr_4C"
##  [61] "e_v2x_execorr_5C"          "e_v2x_feduni_3C"
##  [63] "e_v2x_feduni_4C"           "e_v2x_feduni_5C"
##  [65] "e_v2x_frassoc_thick_3C"    "e_v2x_frassoc_thick_4C"
##  [67] "e_v2x_frassoc_thick_5C"    "e_v2x_freexp_3C"
##  [69] "e_v2x_freexp_4C"           "e_v2x_freexp_5C"
##  [71] "e_v2x_freexp_altinf_3C"    "e_v2x_freexp_altinf_4C"
##  [73] "e_v2x_freexp_altinf_5C"    "e_v2x_gencl_3C"
##  [75] "e_v2x_gencl_4C"            "e_v2x_gencl_5C"
##  [77] "e_v2x_gencs_3C"            "e_v2x_gencs_4C"
##  [79] "e_v2x_gencs_5C"            "e_v2x_gender_3C"
##  [81] "e_v2x_gender_4C"           "e_v2x_gender_5C"
##  [83] "e_v2x_genpp_3C"            "e_v2x_genpp_4C"
##  [85] "e_v2x_genpp_5C"            "e_v2x_jucon_3C"
##  [87] "e_v2x_jucon_4C"            "e_v2x_jucon_5C"
##  [89] "e_v2x_libdem_3C"           "e_v2x_libdem_4C"
##  [91] "e_v2x_libdem_5C"           "e_v2x_liberal_3C"
##  [93] "e_v2x_liberal_4C"          "e_v2x_liberal_5C"
##  [95] "e_v2x_mpi_3C"              "e_v2x_mpi_4C"
##  [97] "e_v2x_mpi_5C"              "e_v2x_partip_3C"
##  [99] "e_v2x_partip_4C"           "e_v2x_partip_5C"
## [101] "e_v2x_partipdem_3C"        "e_v2x_partipdem_4C"
## [103] "e_v2x_partipdem_5C"        "e_v2x_polyarchy_3C"
```

```
## [105] "e_v2x_polyarchy_4C"          "e_v2x_polyarchy_5C"
## [107] "e_v2x_pubcorr_3C"            "e_v2x_pubcorr_4C"
## [109] "e_v2x_pubcorr_5C"            "e_v2x_suffr_3C"
## [111] "e_v2x_suffr_4C"              "e_v2x_suffr_5C"
## [113] "e_v2xcl_rol_3C"              "e_v2xcl_rol_4C"
## [115] "e_v2xcl_rol_5C"              "e_v2xcs_ccsi_3C"
## [117] "e_v2xcs_ccsi_4C"             "e_v2xcs_ccsi_5C"
## [119] "e_v2xdd_dd_3C"               "e_v2xdd_dd_4C"
## [121] "e_v2xdd_dd_5C"               "e_v2xdl_delib_3C"
## [123] "e_v2xdl_delib_4C"            "e_v2xdl_delib_5C"
## [125] "e_v2xeg_eqdr_3C"             "e_v2xeg_eqdr_4C"
## [127] "e_v2xeg_eqdr_5C"             "e_v2xeg_eqprotec_3C"
## [129] "e_v2xeg_eqprotec_4C"         "e_v2xeg_eqprotec_5C"
## [131] "e_v2xel_frefair_3C"          "e_v2xel_frefair_4C"
## [133] "e_v2xel_frefair_5C"          "e_v2xel_locelec_3C"
## [135] "e_v2xel_locelec_4C"          "e_v2xel_locelec_5C"
## [137] "e_v2xel_regelec_3C"          "e_v2xel_regelec_4C"
## [139] "e_v2xel_regelec_5C"          "e_v2xlg_legcon_3C"
## [141] "e_v2xlg_legcon_4C"           "e_v2xlg_legcon_5C"
## [143] "e_v2xme_altinf_3C"           "e_v2xme_altinf_4C"
## [145] "e_v2xme_altinf_5C"           "e_v2xps_party_3C"
## [147] "e_v2xps_party_4C"            "e_v2xps_party_5C"
## [149] "e_boix_regime"               "e_democracy_breakdowns"
## [151] "e_democracy_omitteddata"     "e_democracy_trans"
## [153] "e_fh_cl"                     "e_fh_pr"
## [155] "e_fh_rol"                    "e_fh_status"
## [157] "e_wbgi_cce"                  "e_wbgi_gee"
## [159] "e_wbgi_pve"                  "e_wbgi_rle"
## [161] "e_wbgi_rqe"                  "e_wbgi_vae"
## [163] "e_lexical_index"             "e_uds_median"
## [165] "e_uds_mean"                  "e_uds_pct025"
## [167] "e_uds_pct975"                "e_coups"
## [169] "e_legparty"                  "e_autoc"
## [171] "e_democ"                     "e_p_polity"
## [173] "e_polcomp"                   "e_polity2"
## [175] "e_bnr_dem"                   "e_chga_demo"
## [177] "e_ti_cpi"                    "e_vanhanen"
## [179] "e_peaveduc"                  "e_peedgini"
## [181] "e_area"                      "e_regiongeo"
## [183] "e_regionpol"                 "e_regionpol_6C"
## [185] "e_cow_exports"               "e_cow_imports"
## [187] "e_gdp"                       "e_gdp_sd"
## [189] "e_gdppc"                     "e_gdppc_sd"
## [191] "e_miinflat"                  "e_pop"
## [193] "e_pop_sd"                    "e_total_fuel_income_pc"
## [195] "e_total_oil_income_pc"       "e_total_resources_income_pc"
## [197] "e_radio_n"                   "e_miferrat"
## [199] "e_mipopula"                  "e_miurbani"
## [201] "e_miurbpop"                  "e_pefeliex"
## [203] "e_peinfmor"                  "e_pelifeex"
## [205] "e_pematmor"                  "e_wb_pop"
## [207] "e_civil_war"                 "e_miinteco"
## [209] "e_miinterc"                  "e_pt_coup"
```

```
## [211] "e_pt_coup_attempts"
```

```
d_edu <- d |>
  select(country_name,country_id,year,e_peaveduc,e_peedgini)
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
```

```
d_edu <-d_edu |>
  rename("edu_level"="e_peaveduc","edu_inequality"="e_peedgini","country"="country_name","ID"="country_
d_edu
## # A tibble: 6,789 x 5
##    country    ID  year edu_level edu_inequality
##    <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
```

Subset by rows

```
#coutries-years have the highest education level
d_edu |>slice_max(order_by=edu_level, n=5)
## # A tibble: 13 x 5
##    country           ID  year edu_level edu_inequality
##    <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
```

```
#countries-years that suffer from the most severe inequality
d_edu |>slice_max(order_by=edu_inequality,n=5)
## # A tibble: 5 x 5
##   country         ID  year edu_level edu_inequality
##   <chr>        <dbl> <dbl>     <dbl>          <dbl>
## 1 Burkina Faso    54  1984     0.301           97.0
## 2 Burkina Faso    54  1985     0.322           96.9
## 3 Burkina Faso    54  1986     0.343           96.7
## 4 Burkina Faso    54  1987     0.364           96.4
## 5 Burkina Faso    54  1988     0.385           96.1
```

Summarize the data

```
#check data availablity (edu_level)
d_edu |>
  mutate(edu_level_available=as.numeric(!is.na(edu_level)), .after=edu_level) |>
  group_by(country) |>
  summarise(N_edu_level_available=sum(edu_level_available))
## # A tibble: 181 x 2
##    country     N_edu_level_available
##    <chr>                       <dbl>
##  1 Afghanistan                    39
##  2 Albania                         0
##  3 Algeria                        39
##  4 Angola                         39
##  5 Argentina                      39
##  6 Armenia                        33
##  7 Australia                      39
##  8 Austria                        39
##  9 Azerbaijan                     33
## 10 Bahrain                         0
## # i 171 more rows
```

```
#check data availability(edu_inequality)
d_edu |>
  mutate(edu_inequality_available=as.numeric(!is.na(edu_inequality)), .after=edu_inequality) |>
  group_by(country) |>
  summarise(N_edu_inequality_available=sum(edu_inequality_available))
## # A tibble: 181 x 2
##    country     N_edu_inequality_available
##    <chr>                            <dbl>
##  1 Afghanistan                         27
##  2 Albania                              0
```

```
##  3 Algeria                              27
##  4 Angola                               27
##  5 Argentina                            27
##  6 Armenia                              21
##  7 Australia                            27
##  8 Austria                              27
##  9 Azerbaijan                           21
## 10 Bahrain                               0
## # i 171 more rows
```

```
#average level of education quality
d_edu |>
  group_by(country) |>
  summarize(edu_level_average=mean(edu_level,na.rm=TRUE),
            edu_inequality_average=mean(edu_inequality,na.rm=TRUE))
## # A tibble: 181 x 3
##    country     edu_level_average edu_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
```

```
#change of education quality
#Most countries have edu_inequality records in 2010
d_edu |>
  filter(year>=1984,year<=2010)|>
  group_by(country)|>
  arrange(year)|>
  summarise(edu_level_change_2010_1984=last(edu_level)-first(edu_level),
            edu_inequality_change_2010_1984=last(edu_inequality)-first(edu_inequality))|>
  ungroup()|>
  arrange(country)
## # A tibble: 180 x 3
##    country     edu_level_change_2010_1984 edu_inequality_change_2010_1984
##    <chr>                             <dbl>                           <dbl>
##  1 Afghanistan                       2.52                           -21
##  2 Albania                          NA                               NA
##  3 Algeria                           3.35                           -18.9
##  4 Angola                            1.64                           -29.5
##  5 Argentina                         1.06                            -3.56
##  6 Armenia                           0.336                           -2.87
##  7 Australia                         0.878                           -7.77
##  8 Austria                           1.16                            -5.68
##  9 Azerbaijan                        0.252                           -2.16
## 10 Bahrain                          NA                               NA
## # i 170 more rows
```

Briefly discuss

```
#examine the data
d_edu |>
  group_by(country) |>
  summarize(edu_level_average=mean(edu_level,na.rm=TRUE),
            edu_inequality_average=mean(edu_inequality,na.rm=TRUE))|>
  slice_max(order_by=edu_inequality_average,n=181)
## # A tibble: 181 x 3
##    country      edu_level_average edu_inequality_average
##    <chr>                    <dbl>                  <dbl>
##  1 Burkina Faso             0.982                   91.3
##  2 Mali                     1.25                    87.9
##  3 Niger                    1.06                    85.3
##  4 Somalia                  1.29                    84.7
##  5 Afghanistan              2.80                    77.8
##  6 Benin                    2.39                    76.9
##  7 The Gambia               2.82                    76.7
##  8 Guinea                   2.62                    73.4
##  9 Burundi                  1.86                    73.0
## 10 Nepal                    3.21                    69.8
## # i 171 more rows
```

```
d_edu |>
  group_by(country) |>
  summarize(edu_level_average=mean(edu_level,na.rm=TRUE),
            edu_inequality_average=mean(edu_inequality,na.rm=TRUE))|>
  slice_min(order_by=edu_level_average,n=181)
## # A tibble: 181 x 3
##    country      edu_level_average edu_inequality_average
##    <chr>                    <dbl>                  <dbl>
##  1 Burkina Faso             0.982                   91.3
##  2 Niger                    1.06                    85.3
##  3 Mali                     1.25                    87.9
##  4 Somalia                  1.29                    84.7
##  5 Burundi                  1.86                    73.0
##  6 Mozambique               2.36                    52.6
##  7 Benin                    2.39                    76.9
##  8 Angola                   2.46                    53.9
##  9 Senegal                  2.54                    66.8
## 10 Guinea                   2.62                    73.4
## # i 171 more rows
```

```
d_edu |>
  filter(year>=1984,year<=2010)|>
  group_by(country)|>
  arrange(year)|>
  summarise(edu_level_change_2010_1984=last(edu_level)-first(edu_level),
            edu_inequality_change_2010_1984=last(edu_inequality)-first(edu_inequality))|>
  ungroup()|>
  arrange(country)|>
  slice_min(order_by=edu_level_change_2010_1984,n=181)
## # A tibble: 180 x 3
```

```
##    country     edu_level_change_2010_1984 edu_inequality_change_2010_1984
##    <chr>                            <dbl>                           <dbl>
##  1 Tajikistan                      -0.252                           -6.42
##  2 North Korea                      0                               NA
##  3 Russia                           0.230                          -10.1
##  4 Azerbaijan                       0.252                           -2.16
##  5 Uzbekistan                       0.272                           -2.48
##  6 Kyrgyzstan                       0.301                           -2.73
##  7 Switzerland                      0.328                            1.72
##  8 Armenia                          0.336                           -2.87
##  9 Germany                          0.350                           -2.33
## 10 Georgia                          0.387                           -3.56
## # i 170 more rows
```

In terms of average educational level and average educational inequality index, the countries with the best performance include: Germany, Australia, the United Kingdom, Barbados, Denmark, Sweden, etc. The worst countries include Burkina Faso, Mali, Nigeria, Somalia, etc. Looking from the changes from 1984 to 2010, countries such as Botswana, Singapore, and Libya have made great progress. Tajikistan, North Korea, Russia and other countries have not changed much.