Summarize statistics across countries grouped by regions using kableExtra

Fan Wang

2024-04-07

Contents

1 Summarize statistics from across countries **1**
1.1 Generating preparing data input 1
1.2 Generate automatic sorter and grouping frame 4
1.3 Formatting inputs with decimal controls and percentage signs 4
1.4 We now generate the table 6

1 Summarize statistics from across countries

Go to the **RMD**, **R**, **PDF**, or **HTML** version of this file. Go back to fan's REconTools research support package, R4Econ examples page, PkgTestR packaging guide, or Stat4Econ course page.

1.1 Generating preparing data input

In this example file, we have the average exposure to air pollution among children from Latin America and the Caribbean in 2010, as well as average temperature, measured in share of hours in the year over UTCI 32 degrees. We will create a table that groups Latin American and Caribbean countries into groupings, show various statistics by each country row by row, including ranking for each country. The output generated is in HTML and tex formats.

Files of this nature generates all the tables in the appendix of Hannum, Kim, and Wang (2024). It is implemented also here for appendix tables in Feng et al (2024).

First, we load the pollution file in. Additionally, we load in a file containing the full names of countries. Clean up the input files so that the country files have the same variable names. Ignore the prior variable names, with the prefix pm10, they are not pm10 measurements.

```
library(readr)
library(dplyr)
library(tidyr)
library(tibble)
library(kableExtra)
```

```
# Parameters
verbose <- TRUE
# Generate paths
spt_path_root <- file.path('C:', 'users', 'fan', 'R4Econ', fsep = .Platform$file.sep)
spt_path_in <- file.path(spt_path_root, 'table', 'bigtable', '_file', fsep = .Platform$file.sep)
spn_aod <- file.path(spt_path_in, 'lac_aod_children.csv', fsep = .Platform$file.sep)
spn_utci <- file.path(spt_path_in, 'lac_utci_children.csv', fsep = .Platform$file.sep)
spn_keys <- file.path(spt_path_in, 'country_code.csv', fsep = .Platform$file.sep)
```

```r
# Load files
df_lac_aod_pollution <- read_csv(spn_aod) %>%
  select(-order)
df_lac_utci_temperature <- read_csv(spn_utci) %>%
  select(
    one_of(colnames(df_lac_aod_pollution)),
    pm10_grp_mean, pm10_overall_mean, pm10_grp_exc_burden
  ) %>%
  rename(
    pm10_grp_mean_x = pm10_grp_mean,
    pm10_grp_mean_y = pm10_overall_mean,
    relative_excess_burden = pm10_grp_exc_burden
  )
df_country_code <- read_csv(spn_keys)
# output tex file path
spn_tex_out <- file.path(
  spt_path_in, 'lac_climate_children_rank.tex', fsep = .Platform$file.sep)
```

Second, we merge the country names into the data file with three letter ISO code, select the subset of variables to be used by the table, and generate some summary stats on the file. We also merge the UTCI temperature information into the pollution file.

```r
# Merge with key
df_lac_aod_pollution <- df_lac_aod_pollution %>%
  left_join(df_country_code, by = "ISOCODE")
# Merge with temperature information
df_lac_aod_pollution <- df_lac_aod_pollution %>%
  left_join(df_lac_utci_temperature %>%
    rename(utcige32c_age0t5 = pm10_grp_mean_x) %>%
    select(ISOCODE, utcige32c_age0t5), by = "ISOCODE")
# Filter region to consider
df_lac_aod_pollution_lac <- df_lac_aod_pollution %>%
  dplyr::filter(region_name == "Americas") %>%
  dplyr::filter(!(ISOCODE %in% c(
    "CAN", "USA", "CUB", "GRL"
    )))
# Select variables
df_lac_aod_pollution_lac <- df_lac_aod_pollution_lac %>%
  rename(
    group_sorter_desc = 'Income group',
    aodavg_age0t5 = pm10_grp_mean_x
    ) %>%
  select(
    ISOCODE, country_name, group_sorter_desc,
    aodavg_age0t5 , utcige32c_age0t5
  )
# Display
str(df_lac_aod_pollution_lac)
```

```
## tibble [36 x 5] (S3: tbl_df/tbl/data.frame)
##  $ ISOCODE          : chr [1:36] "NIC" "PAN" "PER" "PRI" ...
##  $ country_name     : chr [1:36] "Nicaragua" "Panama" "Peru" "Puerto Rico" ...
##  $ group_sorter_desc: chr [1:36] "Lower middle income" "High income" "Upper middle income" "High inc
##  $ aodavg_age0t5    : num [1:36] 227 277 228 244 215 ...
##  $ utcige32c_age0t5 : num [1:36] 0.1827 0.1466 0.0376 0.0624 0.0426 ...
```

Third, for a number of locations that are dependencies and constituent entities, no full country names were included, we include them now.

```
# Identify countries without country_name
ar_st_isocode_dependencies <- df_lac_aod_pollution_lac %>%
  dplyr::filter(is.na(country_name)) %>%
  pull(ISOCODE)
print(ar_st_isocode_dependencies)
```

```
## [1] "MTQ" "MSR" "AIA" "BES" "GUF" "FLK"
```

```
# Add in names
# https://www.britannica.com/topic/list-of-countries-in-Latin-America-2061416
df_lac_aod_pollution_lac <- df_lac_aod_pollution_lac %>%
  mutate(country_name = case_when(
    ISOCODE == "MTQ" ~ 'Martinique',
    ISOCODE == "MSR" ~ 'Montserrat',
    ISOCODE == "AIA" ~ 'Anguilla',
    ISOCODE == "BES" ~ 'St. Eustatius',
    ISOCODE == "GUF" ~ 'French Guiana',
    ISOCODE == "FLK" ~ 'Falkland Islands',
    TRUE ~ country_name
  ))
# Country name and ISO combine
df_lac_aod_pollution_lac <- df_lac_aod_pollution_lac %>%
  mutate(country_name = paste0(
    country_name, " (", ISOCODE , ")"
    ))
```

Fourth, we will use the world bank income groups to categorize countries, and include an additional category for the dependency and constituent entities just renamed above.

```
# Add in categorical name for locations without world bank income group designation
df_lac_aod_pollution_lac <- df_lac_aod_pollution_lac %>%
  mutate(group_sorter_desc= case_when(
    is.na(group_sorter_desc) ~ 'Without World Bank income group designation',
    TRUE ~ group_sorter_desc
  ))
```

Fifth, we generate a ranking variable, for which country has the highest level of AOD, and rescale AOD value to be between 0 and 1.

```
# basic chart with two lines
df_lac_aod_pollution_lac <- df_lac_aod_pollution_lac %>%
  mutate(aodavg_age0t5 = aodavg_age0t5/1000) %>%
  arrange(aodavg_age0t5) %>%
  mutate(aod_rank = row_number()) %>%
  arrange(utcige32c_age0t5) %>%
  mutate(utcige32c_rank = row_number())
```

Sixth, to faciliate group sorting in the way we want groups to be displayed in sequence, we now generate a sorting variable so that richer group comes first.

```
df_lac_aod_pollution_lac <- df_lac_aod_pollution_lac %>%
  mutate(group_sorter = case_when(
    group_sorter_desc == 'High income' ~ 1,
    group_sorter_desc == 'Upper middle income' ~ 2,
    group_sorter_desc == 'Lower middle income' ~ 3,
```

```
    group_sorter_desc == 'Without World Bank income group designation' ~ 4
  ))
# display
kable(df_lac_aod_pollution_lac, caption="Input data") %>% kable_styling_fc()
```

1.2 Generate automatic sorter and grouping frame

Before generate the table using KableExtra, we need to do some final preparation to generating a auxillary sorting file to help us group observations by groups and generate proper headings. We generate `df_lac_aod_pollution_lac`, which counts the number of observations within each group, and the starting and ending row for each group.

```
# Sorted file
df_data_sorted <- df_lac_aod_pollution_lac %>%
  arrange(
    group_sorter, country_name,
  ) %>%
  group_by(group_sorter) %>%
  ungroup()
# Count by group
df_group_counts <- df_data_sorted %>%
  group_by(group_sorter_desc, group_sorter) %>%
  summarize(group_count = n()) %>%
  arrange(group_sorter) %>% ungroup() %>%
  mutate(group_count_start = cumsum(group_count) - group_count + 1) %>%
  mutate(group_count_end = cumsum(group_count)) %>%
  select(group_count_start, group_count_end, everything())
# display
kable(df_group_counts, caption="Group counter") %>% kable_styling_fc()
```

1.3 Formatting inputs with decimal controls and percentage signs

We make a final adjustment to the file so that data values appear with proper formating.

```
# 4. Format columns, decimals, percentage signs, etc
df_data_formated <- df_data_sorted %>%
  arrange(group_sorter, country_name) %>%
  select(-group_sorter_desc, -group_sorter, -ISOCODE) %>%
  select(
    country_name,
    aodavg_age0t5, aod_rank,
    utcige32c_age0t5, utcige32c_rank) %>%
  mutate_at(
    vars(contains("utcige32c_age")),
    list(~ paste0(
      format(round(., 3) * 100,
        nsmall = 1,
        big.mark = ","
      ),
      "%"
    ))
  ) %>%
  mutate_at(
    vars(contains("aodavg_age")),
```

Input data

ISOCODE	country_name	group_sorter_desc	aodavg_age0t5	utcig
FLK	Falkland Islands (FLK)	Without World Bank income group designation	0.1118511	
CHL	Chile (CHL)	High income	0.1064681	
LCA	St. Lucia (LCA)	Upper middle income	0.2160146	
MSR	Montserrat (MSR)	Without World Bank income group designation	0.2109789	
AIA	Anguilla (AIA)	Without World Bank income group designation	0.2343031	
URY	Uruguay (URY)	High income	0.1444593	
PER	Peru (PER)	Upper middle income	0.2280502	
VGB	British Virgin Islands (VGB)	High income	0.2181750	
MTQ	Martinique (MTQ)	Without World Bank income group designation	0.2152106	
TCA	Turks and Caicos Islands (TCA)	High income	0.1793108	
ATG	Antigua and Barbuda (ATG)	High income	0.2208181	
ARG	Argentina (ARG)	Upper middle income	0.1383548	
PRI	Puerto Rico (PRI)	High income	0.2443060	
GRD	Grenada (GRD)	Upper middle income	0.2604545	
BOL	Bolivia (BOL)	Lower middle income	0.2545498	
MEX	Mexico (MEX)	Upper middle income	0.1874433	
ECU	Ecuador (ECU)	Upper middle income	0.2974905	
GTM	Guatemala (GTM)	Upper middle income	0.2226600	
COL	Colombia (COL)	Upper middle income	0.3435167	
JAM	Jamaica (JAM)	Upper middle income	0.2292714	
BHS	Bahamas, The (BHS)	High income	0.2298316	
PRY	Paraguay (PRY)	Upper middle income	0.2549931	
PAN	Panama (PAN)	High income	0.2771060	
BES	St. Eustatius (BES)	Without World Bank income group designation	0.2918912	
BRA	Brazil (BRA)	Upper middle income	0.1942300	
HND	Honduras (HND)	Lower middle income	0.2042283	
TTO	Trinidad and Tobago (TTO)	High income	0.2799472	
CRI	Costa Rica (CRI)	Upper middle income	0.2457463	
DOM	Dominican Republic (DOM)	Upper middle income	0.2253216	
NIC	Nicaragua (NIC)	Lower middle income	0.2267819	
HTI	Haiti (HTI)	Lower middle income	0.2156850	
BLZ	Belize (BLZ)	Upper middle income	0.2737943	
VEN	Venezuela (VEN)	Without World Bank income group designation	0.3142806	
GUF	French Guiana (GUF)	Without World Bank income group designation	0.2485556	
SUR	Suriname (SUR)	Upper middle income	0.2740000	
GUY	Guyana (GUY)	High income	0.2532788	

Group counter

group_count_start	group_count_end	group_sorter_desc	group_sorter	group_count
1	10	High income	1	10
11	25	Upper middle income	2	15
26	29	Lower middle income	3	4
30	36	Without World Bank income group designation	4	7

```
    list(~ ifelse(abs(.) <= 0.0001,
      paste0(
        formatC(. ,
          digits = 1,
          format = "fg",
          drop0trailing = T,
          flag = "#"
        )
      ),
      paste0(
        format(round(., 3),
          nsmall = 1,
          big.mark = ","
        )
      )
    ))
  )
```

1.4 We now generate the table

The output tex file is stored here: lac_climate_children_rank.tex.

```
bl_main_save <- TRUE
ar_st_kableformat <- c("html", "latex")
for (st_kableformat in ar_st_kableformat) {
  # Column names
  ar_st_col_names <- c(
    "Country name",
    "Child (0-5) mean AOD exposure",
    "LAC child AOD rank",
    "Child (0-5) mean annual share of time over 32 UTCI degrees",
    "LAC child temperature rank"
  )
  # Define column groups, grouping the names above
  # =1/3/2 are number of columns group title covers
  ar_st_col_groups1 <- c(
    " " = 1,
    "Air pollution by aerosols (AOD between 0 and 1)" = 2,
    "At least strong heat stress exposure" = 2
  )
  # Second, we construct main table, and add styling.
  st_title <- paste(
    "LAC country child air pollution and heat exposure ranking in 2010"
    )
  bk_tab_a <- kbl(
    df_data_formated,
    format = st_kableformat,
    label = "tab:lac:aod:temp:rank",
    # escape = F,
    linesep = "",
    booktabs = T,
    longtable = T,
    align = "c",
    caption = st_title,
```

```
  col.names = ar_st_col_names
) %>%
  # see https://cran.r-project.org/web/packages/kableExtra/vignettes/awesome_table_in_html.html#Boots
  kable_styling(
    bootstrap_options = c("striped", "hover", "condensed", "responsive"),
    full_width = F, position = "left"
  )
# Third, we add in column groups.
bk_tab_a <- bk_tab_a %>%
  add_header_above(ar_st_col_groups1)
# Fourth, we add in row groups.
for (it_group in seq(1, dim(df_group_counts)[1])) {
  # Reion full name info
  st_loc <- as.character(df_group_counts[[it_group, "group_sorter_desc"]])
  # groups start and end
  it_group_count_start <- df_group_counts[[it_group, "group_count_start"]]
  it_group_count_end <- df_group_counts[[it_group, "group_count_end"]]
  # display text
  st_panel_letter <- base::LETTERS[it_group]
  # Heading group row, year
  st_panel_text <- paste0(
    "Panel ", st_panel_letter, ": ", st_loc
  )
  # Add to table
  bk_tab_a <- bk_tab_a %>%
    pack_rows(
      st_panel_text, it_group_count_start, it_group_count_end,
      latex_gap_space = "0.25em",
      latex_align = "c",
      hline_after = TRUE
    )
}
# Fifth, column formatting.
fl_width_country <- 6
st_width_country <- paste0(fl_width_country, "cm")
bk_tab_a <- bk_tab_a %>%
  column_spec(1, width = st_width_country) %>%
  column_spec(2:dim(df_data_formated)[2], width = "3cm")
# Final adjustments
# Headings on all pages, note use `sub` to replace first midrule
st_headend <- paste0(
  "\\midrule\\endhead\n",
  "\\addlinespace[0.2em]\\hline\\addlinespace[0.2em]\n",
  "\\multicolumn{", dim(df_data_formated)[2], "}{r}{\\emph{Continued on next page}}\\\\\n",
  "\\endfoot\\endlastfoot"
)
bk_tab_a <- sub(bk_tab_a,
  pattern = "\\midrule", replacement = st_headend,
  fixed = TRUE
)

# country-names left-align
bk_tab_a <- gsub(bk_tab_a,
```

```
    pattern = paste0("\\centering\\arraybackslash}p{", st_width_country, "}"),
    replacement = paste0("\\raggedright\\arraybackslash}p{", st_width_country, "}"),
    fixed = TRUE
  )
  bk_tab_a <- gsub(bk_tab_a,
    pattern = paste0("\\$\\textasciicircum{}\\{\\textbackslash{}circ\\}C\\$"),
    replacement = paste0("$^{\\circ}C$"),
    fixed = TRUE
  )
  bk_tab_a <- gsub(bk_tab_a,
    pattern = paste0("\\$\\textbackslash{}ge\\$"),
    replacement = paste0("$\\ge$"),
    fixed = TRUE
  )
  st_text <- ""
  bk_tab_a <- gsub(bk_tab_a,
    pattern = paste0("\\textbackslash{}", st_text, "\\"),
    replacement = paste0("\\", st_text),
    fixed = TRUE
  )
  # midrule replacing hline
  bk_tab_a <- gsub(bk_tab_a,
    pattern = "hline",
    replacement = "midrule", fixed = TRUE
  )

  # 6. Finally, save table content to file
  if (st_kableformat == "latex") {
    if (bl_main_save) {
      fileConn <- file(spn_tex_out)
      writeLines(bk_tab_a, fileConn)
      close(fileConn)
      if (verbose) {
        print(glue::glue("F-815346, S3"))
        print(glue::glue("Latex saved: {spn_tex_out}"))
      }
    }
  } else if (st_kableformat == "html") {
    bk_tab_a_html <- bk_tab_a
  }
}
```

```
## F-815346, S3
## Latex saved: C:/users/fan/R4Econ/table/bigtable/_file/lac_climate_children_rank.tex
```

```
bk_tab_a_html
```

LAC country child air pollution and heat exposure ranking in 2010

Air pollution by aerosols (AOD between 0 and 1)

At least strong heat stress exposure

Country name

Child (0-5) mean AOD exposure

LAC child AOD rank

Child (0-5) mean annual share of time over 32 UTCI degrees

LAC child temperature rank

Panel A: High income

Antigua and Barbuda (ATG)

0.221

14

4.4%

11

Bahamas, The (BHS)

0.230

20

11.0%

21

British Virgin Islands (VGB)

0.218

13

3.9%

8

Chile (CHL)

0.106

1

0.6%

2

Guyana (GUY)

0.253

25

27.5%

36

Panama (PAN)

0.277

31

14.7%

23

Puerto Rico (PRI)

0.244

22

6.2%

13

Trinidad and Tobago (TTO)

0.280

32

15.9%

27

Turks and Caicos Islands (TCA)

0.179

5

4.4%

10

Uruguay (URY)

0.144

4

3.4%

6

Panel B: Upper middle income

Argentina (ARG)

0.138

3

5.0%

12

Belize (BLZ)

0.274

29

19.8%

32

Brazil (BRA)

0.194

7

15.6%

25

Colombia (COL)

0.344

36

10.6%

19

Costa Rica (CRI)

0.246

23

16.6%

28

Dominican Republic (DOM)

0.225

16

16.8%

29

Ecuador (ECU)

0.297

34

7.8%

17

Grenada (GRD)

0.260

28

6.5%

14

Guatemala (GTM)

0.223

15

10.1%

18

Jamaica (JAM)

0.229

19

10.8%

20

Mexico (MEX)

0.187

6

7.0%

16

Paraguay (PRY)

0.255

27

14.4%

22

Peru (PER)

0.228

18

3.8%

7

St. Lucia (LCA)

0.216

12

2.6%

3

Suriname (SUR)

0.274

30

26.3%

35

Panel C: Lower middle income

Bolivia (BOL)

0.255

26

6.6%

15

Haiti (HTI)

0.216

11

19.1%

31

Honduras (HND)

0.204

8

15.9%

26

Nicaragua (NIC)

0.227

17

18.3%

30

Panel D: Without World Bank income group designation

Anguilla (AIA)

0.234

21

3.3%

5

Falkland Islands (FLK)

0.112

2

0.0%

1

French Guiana (GUF)

0.249

24

25.5%

34

Martinique (MTQ)

0.215

10

4.3%

9

Montserrat (MSR)

0.211

9

3.1%

4

St. Eustatius (BES)

0.292

33

15.3%

24

Venezuela (VEN)

0.314

35

22.0%

33