Analysis.Rmd

2024-10-03

```
#renu::snapshot()
#renu::restore()
#renu::hydrate()
#git: different versions of a file
#github: hosting "remotes"
#clone: local copy of remote, remote is "origin"
#fork: copy repo on git
#commit: creating a snapshot of changes
#push: update remote repo with local changes
#pull: update local repo with commits pushed by others
```

Introduction

The Global Health Expenditure Database (GHED) provides time series data from 2000-2022 which covers country-level spending in the world. Healthcare in the USA is notably more expensive than its counter-parts across the world. This dataset will allow us to examine how different countries allocate their funds towards different segments of healthcare. For example, we may be able to explore how much spending is being allocated towards injuries, for example.

```
library(tidyverse)
```

```
## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──
## ✔ dplyr     1.1.4     ✔ readr     2.1.5
## ✔ forcats   1.0.0     ✔ stringr   1.5.1
## ✔ ggplot2   3.5.1     ✔ tibble    3.2.1
## ✔ lubridate 1.9.3     ✔ tidyr     1.3.1
## ✔ purrr     1.0.2
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflic
ts to become errors
```

```
library(readxl)
```

```
GHED_data <- read_excel("GHED_data.XLSX", sheet = 1)

# View the structure of the dataset
str(GHED_data)
```

```
## tibble [4,244 × 3,923] (S3: tbl_df/tbl/data.frame)
##  $ country                 : chr [1:4244] "Algeria" "Algeria" "Algeria" "Algeria"
...
##  $ code                    : chr [1:4244] "DZA" "DZA" "DZA" "DZA" ...
##  $ region                  : chr [1:4244] "AFR" "AFR" "AFR" "AFR" ...
##  $ income                  : chr [1:4244] "Lower-middle" "Lower-middle" "Lower-mid
dle" "Lower-middle" ...
##  $ year                    : num [1:4244] 2000 2001 2002 2003 2004 ...
##  $ che_gdp                 : num [1:4244] 3.49 3.84 3.73 3.6 3.54 ...
##  $ che_pc_usd              : num [1:4244] 62.1 67.3 66.9 76.2 93 ...
##  $ hk_gdp                  : num [1:4244] NA NA NA NA NA NA NA NA NA NA ...
##  $ hk_g_gdp                : num [1:4244] NA NA NA NA NA NA NA NA NA NA ...
##  $ hk_ext_gdp              : num [1:4244] NA NA NA NA NA NA NA NA NA NA ...
##  $ che                     : num [1:4244] 143870 162231 168702 189137 217929 ...
##  $ gghed                   : num [1:4244] 103534 123664 126997 145057 155500 ...
##  $ pvtd                    : num [1:4244] 40261 38492 41630 43985 62327 ...
##  $ ext                     : num [1:4244] 75.1 75.1 75.1 95 102 ...
##  $ dom_che                 : num [1:4244] 99.9 100 100 99.9 100 ...
##  $ gghed_che               : num [1:4244] 72 76.2 75.3 76.7 71.4 ...
##  $ pvtd_che                : num [1:4244] 28 23.7 24.7 23.3 28.6 ...
##  $ oops_che                : num [1:4244] 25.8 21.7 22.5 21.1 26.1 ...
##  $ vpp_che                 : num [1:4244] 0.818 0.848 0.954 0.981 1.445 ...
##  $ ext_che                 : num [1:4244] 0.0522 0.0463 0.0445 0.0502 0.0468 ...
##  $ gghed_gdp               : num [1:4244] 2.51 2.93 2.81 2.76 2.53 ...
##  $ gghed_gge               : num [1:4244] 8.78 9.27 7.97 9.44 8.67 ...
##  $ gghed_pc_usd            : num [1:4244] 44.7 51.3 50.4 58.5 66.4 ...
##  $ pvtd_pc_usd             : num [1:4244] 17.4 16 16.5 17.7 26.6 ...
##  $ oop_pc_usd              : num [1:4244] 16 14.6 15.1 16.1 24.3 ...
##  $ ext_pc_usd              : num [1:4244] 0.0324 0.0312 0.0298 0.0383 0.0435 ...
##  $ tran_shi                : num [1:4244] 0 0 0 0 0 0 0 0 0 0 ...
##  $ shise_shi               : num [1:4244] NA NA NA NA NA NA NA NA NA NA ...
##  $ cfa_che                 : num [1:4244] 72 76.2 75.3 76.7 71.4 ...
##  $ gfa_che                 : num [1:4244] 45.9 50.5 49.3 50.7 47.1 ...
##  $ chi_che                 : num [1:4244] 26.1 25.8 25.9 26 24.2 ...
##  $ shi_che                 : num [1:4244] 26.1 25.8 25.9 26 24.2 ...
##  $ chi_pvt_che             : num [1:4244] 0 0 0 0 0 0 0 0 0 0 ...
##  $ vfa_che                 : num [1:4244] 28 23.8 24.7 23.3 28.6 ...
##  $ vhi_che                 : num [1:4244] 0.818 0.848 0.954 0.981 1.445 ...
##  $ row_che                 : num [1:4244] NA NA NA NA NA NA NA NA NA NA ...
##  $ phc_usd_pc              : num [1:4244] NA NA NA NA NA NA NA NA NA NA ...
##  $ phc_che                 : num [1:4244] NA NA NA NA NA NA NA NA NA NA ...
##  $ phc_gghed_usd_pc        : num [1:4244] NA NA NA NA NA NA NA NA NA NA ...
##  $ gghed_phc_gghed         : num [1:4244] NA NA NA NA NA NA NA NA NA NA ...
##  $ gghed_phc_phc           : num [1:4244] NA NA NA NA NA NA NA NA NA NA ...
##  $ phc_ext_usd_pc          : num [1:4244] NA NA NA NA NA NA NA NA NA NA ...
##  $ ext_phc_ext             : num [1:4244] NA NA NA NA NA NA NA NA NA NA ...
##  $ ext_phc_phc             : num [1:4244] NA NA NA NA NA NA NA NA NA NA ...
##  $ phc_public_gdp          : num [1:4244] NA NA NA NA NA NA NA NA NA NA ...
##  $ phc_pvtd_usd_pc         : num [1:4244] NA NA NA NA NA NA NA NA NA NA ...
##  $ pvtd_phc_pvtd           : num [1:4244] NA NA NA NA NA NA NA NA NA NA ...
##  $ pvtd_phc_phc            : num [1:4244] NA NA NA NA NA NA NA NA NA NA ...
```

```
##  $ hc62_che               : num [1:4244] NA NA NA NA NA NA NA NA NA NA ...
##  $ hc62_g_gghed           : num [1:4244] NA NA NA NA NA NA NA NA NA NA ...
##  $ hc62_ext_ext           : num [1:4244] NA NA NA NA NA NA NA NA NA NA ...
##  $ fp3214_che             : num [1:4244] NA NA NA NA NA NA NA NA NA NA ...
##  $ fp3214_gghed_fp3214    : num [1:4244] NA NA NA NA NA NA NA NA NA NA ...
##  $ fp3214_ext_fp3214      : num [1:4244] NA NA NA NA NA NA NA NA NA NA ...
##  $ dis1_che               : num [1:4244] NA NA NA NA NA NA NA NA NA NA ...
##  $ dis11_che              : num [1:4244] NA NA NA NA NA NA NA NA NA NA ...
##  $ dis12_che              : num [1:4244] NA NA NA NA NA NA NA NA NA NA ...
##  $ dis13_che              : num [1:4244] NA NA NA NA NA NA NA NA NA NA ...
##  $ dis16_che              : num [1:4244] NA NA NA NA NA NA NA NA NA NA ...
##  $ dis2_che               : num [1:4244] NA NA NA NA NA NA NA NA NA NA ...
##  $ dis21_che              : num [1:4244] NA NA NA NA NA NA NA NA NA NA ...
##  $ dis22_che              : num [1:4244] NA NA NA NA NA NA NA NA NA NA ...
##  $ dis23_che              : num [1:4244] NA NA NA NA NA NA NA NA NA NA ...
##  $ dis3_che               : num [1:4244] NA NA NA NA NA NA NA NA NA NA ...
##  $ dis4_che               : num [1:4244] NA NA NA NA NA NA NA NA NA NA ...
##  $ dis5_che               : num [1:4244] NA NA NA NA NA NA NA NA NA NA ...
##  $ disnec_che             : num [1:4244] NA NA NA NA NA NA NA NA NA NA ...
##  $ dis1_g_gghed           : num [1:4244] NA NA NA NA NA NA NA NA NA NA ...
##  $ dis11_g_gghed          : num [1:4244] NA NA NA NA NA NA NA NA NA NA ...
##  $ dis12_g_gghed          : num [1:4244] NA NA NA NA NA NA NA NA NA NA ...
##  $ dis13_g_gghed          : num [1:4244] NA NA NA NA NA NA NA NA NA NA ...
##  $ dis16_g_gghed          : num [1:4244] NA NA NA NA NA NA NA NA NA NA ...
##  $ dis2_g_gghed           : num [1:4244] NA NA NA NA NA NA NA NA NA NA ...
##  $ dis21_g_gghed          : num [1:4244] NA NA NA NA NA NA NA NA NA NA ...
##  $ dis22_g_gghed          : num [1:4244] NA NA NA NA NA NA NA NA NA NA ...
##  $ dis23_g_gghed          : num [1:4244] NA NA NA NA NA NA NA NA NA NA ...
##  $ dis3_g_gghed           : num [1:4244] NA NA NA NA NA NA NA NA NA NA ...
##  $ dis4_g_gghed           : num [1:4244] NA NA NA NA NA NA NA NA NA NA ...
##  $ dis5_g_gghed           : num [1:4244] NA NA NA NA NA NA NA NA NA NA ...
##  $ disnec_g_gghed         : num [1:4244] NA NA NA NA NA NA NA NA NA NA ...
##  $ dis1_ext_ext           : num [1:4244] NA NA NA NA NA NA NA NA NA NA ...
##  $ dis11_ext_ext          : num [1:4244] NA NA NA NA NA NA NA NA NA NA ...
##  $ dis12_ext_ext          : num [1:4244] NA NA NA NA NA NA NA NA NA NA ...
##  $ dis13_ext_ext          : num [1:4244] NA NA NA NA NA NA NA NA NA NA ...
##  $ dis16_ext_ext          : num [1:4244] NA NA NA NA NA NA NA NA NA NA ...
##  $ dis2_ext_ext           : num [1:4244] NA NA NA NA NA NA NA NA NA NA ...
##  $ dis21_ext_ext          : num [1:4244] NA NA NA NA NA NA NA NA NA NA ...
##  $ dis22_ext_ext          : num [1:4244] NA NA NA NA NA NA NA NA NA NA ...
##  $ dis23_ext_ext          : num [1:4244] NA NA NA NA NA NA NA NA NA NA ...
##  $ dis3_ext_ext           : num [1:4244] NA NA NA NA NA NA NA NA NA NA ...
##  $ dis4_ext_ext           : num [1:4244] NA NA NA NA NA NA NA NA NA NA ...
##  $ dis5_ext_ext           : num [1:4244] NA NA NA NA NA NA NA NA NA NA ...
##  $ disnec_ext_ext         : num [1:4244] NA NA NA NA NA NA NA NA NA NA ...
##  $ hccov_che              : num [1:4244] NA NA NA NA NA NA NA NA NA NA ...
##  $ hccov_usd_pc           : num [1:4244] NA NA NA NA NA NA NA NA NA NA ...
##  $ hccov_gghed_gghed      : num [1:4244] NA NA NA NA NA NA NA NA NA NA ...
##  $ hccov_ext_ext          : num [1:4244] NA NA NA NA NA NA NA NA NA NA ...
##  $ gge_gdp                : num [1:4244] 28.6 31.6 35.2 29.3 29.2 ...
##  $ gdp_pc_usd             : num [1:4244] 1780 1755 1795 2117 2625 ...
```

```
##   [list output truncated]
```

Let us do a little bit of data cleaning. Here we will check for missing values. Further, we will use the variable measuring the proportion of national spending on healthcare as our first variable to explore.

```
GHED_data_filtered <- GHED_data %>%
  select(
    country,
    year,
    income,
    che_gdp,
    che_pc_usd,
    gdp_usd2021_pc,
    region,
    oop_pc_usd
  ) %>%
  filter(year == 2021)

missing_values <- colSums(is.na(GHED_data_filtered))
missing_values
```

```
##        country           year         income        che_gdp     che_pc_usd
##              0              0              0              3              4
## gdp_usd2021_pc         region     oop_pc_usd
##              4              0              5
```

Let us find some summary statistics for che_gdp, the proportion of health spending on the basis of GDP.

```
# Calculate summary statistics for 'che_gdp'
summary_stats_che_gdp <- GHED_data_filtered %>%
  select(che_gdp) %>%
  drop_na() %>%
  summarize(
    mean = mean(che_gdp),
    median = median(che_gdp),
    sd = sd(che_gdp)
  )
summary_stats_che_gdp
```

```
## # A tibble: 1 × 3
##    mean median    sd
##   <dbl>  <dbl> <dbl>
## 1  7.33   6.90  3.35
```

The histogram will show how frequently different values of CHE as a percentage of GDP appear in the dataset for 2022.

```
# Histogram of 'che_gdp'
ggplot(GHED_data_filtered %>% select(che_gdp) %>% drop_na(), aes(x = che_gdp)) +
  geom_histogram(binwidth = 1, fill = 'skyblue', color = 'black') +
  labs(
    title = 'Distribution of Current Health Expenditure as % of GDP in 2022',
    x = 'CHE (% of GDP)',
    y = 'Frequency'
  )
```