CHL8010: Statistical Programming and Computation in Health Data

Week 4 In-class Assignment

2024-09-30

```
##| label: load-packages
##| include: false
#source("../.Rprofile")
library(here)
```

```
Warning: package 'here' was built under R version 4.3.3
```

```
here() starts at C:/Users/User/OneDrive/Documents/R Projects/Version Control/armed_conflict_
```

```
library(dplyr)
```

```
Warning: package 'dplyr' was built under R version 4.3.2
```

```

Attaching package: 'dplyr'

The following objects are masked from 'package:stats':

    filter, lag

The following objects are masked from 'package:base':

    intersect, setdiff, setequal, union
```

```
library(ggplot2)
```

```
Warning: package 'ggplot2' was built under R version 4.3.2
```

Perfect your GitHub repo

Some of you may still need to organize your GitHub repo. Use this time to do that. When you are confident with your repo, let me know – I will try to reproduce your code.

Your final data should have the following variables (you might have slightly different variable names).

```
finaldata <- read.csv(here("data", "finaldata.csv"), header = TRUE)
names(finaldata)
```

```
 [1] "Country.Name"      "Year"              "MaternalMortality"
 [4] "InfantMortality"   "NeonatalMortality" "Under5Mortality"
 [7] "ISO"               "earthquake"        "drought"
[10] "year.x"            "deaths"            "country_name"
[13] "region"            "year.y"            "gdp1000"
[16] "OECD"              "OECD2023"          "popdens"
[19] "urban"             "agedep"            "male_edu"
[22] "temp"              "rainfall1000"
```

Observations from Canada should look like this...

```
finaldata %>%
  dplyr::filter(country_name == "Canada")
```

```
   Country.Name Year MaternalMortality InfantMortality NeonatalMortality
1        Canada 2000                 9             5.3                3.8
2        Canada 2001                10             5.3                3.8
3        Canada 2002                10             5.3                3.9
4        Canada 2003                10             5.3                3.9
5        Canada 2004                10             5.3                3.9
6        Canada 2005                11             5.2                3.9
7        Canada 2006                11             5.2                3.9
8        Canada 2007                11             5.1                3.8
9        Canada 2008                12             5.1                3.8
10       Canada 2009                12             5.0                3.8
```

```
11       Canada 2010                11            5.0                3.8
12       Canada 2011                11            4.9                3.7
13       Canada 2012                11            4.9                3.7
14       Canada 2013                11            4.8                3.6
15       Canada 2014                11            4.7                3.6
16       Canada 2015                11            4.7                3.6
17       Canada 2016                10            4.6                3.5
18       Canada 2017                10            4.6                3.4
19       Canada 2018                NA            4.5                3.3
20       Canada 2019                NA            4.4                3.3
   Under5Mortality ISO earthquake drought year.x deaths country_name
1              6.2 CAN         NA      NA   1999     11       Canada
2              6.2 CAN         NA      NA   2000     23       Canada
3              6.2 CAN         NA      NA   2001      1       Canada
4              6.2 CAN         NA      NA   2002      0       Canada
5              6.1 CAN         NA      NA   2003      0       Canada
6              6.1 CAN         NA      NA   2004      0       Canada
7              6.0 CAN         NA      NA   2005      0       Canada
8              6.0 CAN         NA      NA   2006      0       Canada
9              5.9 CAN         NA      NA   2007      0       Canada
10             5.8 CAN         NA      NA   2008      0       Canada
11             5.7 CAN         NA      NA   2009      0       Canada
12             5.7 CAN         NA      NA   2010      0       Canada
13             5.6 CAN         NA      NA   2011      0       Canada
14             5.5 CAN         NA      NA   2012      0       Canada
15             5.4 CAN         NA      NA   2013      0       Canada
16             5.4 CAN         NA      NA   2014      0       Canada
17             5.3 CAN         NA      NA   2015      0       Canada
18             5.2 CAN         NA      NA   2016      0       Canada
19             5.1 CAN         NA      NA   2017      0       Canada
20             5.1 CAN         NA      NA   2018      0       Canada
            region year.y  gdp1000 OECD OECD2023  popdens    urban   agedep
1  Northern America   2000 24.27100    1        1 66.19704 56.14335 46.34463
2  Northern America   2001 23.82206    1        1 66.45361 56.40270 45.89632
3  Northern America   2002 24.25534    1        1 66.71112 56.67093 45.46660
4  Northern America   2003 28.30046    1        1 66.96384 56.94365 45.07468
5  Northern America   2004 32.14368    1        1 67.21715 57.20020 44.67374
6  Northern America   2005 36.38251    1        1 67.47283 57.41671 44.26641
7  Northern America   2006 40.50406    1        1 67.73674 57.59143 43.96370
8  Northern America   2007 44.65990    1        1 67.99444 57.75691 43.83612
9  Northern America   2008 46.71051    1        1 68.25765 57.97905 43.85426
10 Northern America   2009 40.87631    1        1 68.53354 58.24228 43.94937
11 Northern America   2010 47.56208    1        1 68.80739 58.52809 44.13587
```

```
12 Northern America   2011 52.22370    1         1 69.04842 58.81437 44.53578
13 Northern America   2012 52.66909    1         1 69.27604 59.05573 45.18393
14 Northern America   2013 52.63517    1         1 69.50772 59.19713 45.95404
15 Northern America   2014 50.95600    1         1 69.76876 59.30361 46.75493
16 Northern America   2015 43.59614    1         1 69.98853 59.42627 47.59164
17 Northern America   2016 42.31560    1         1 70.21484 59.50521 48.41410
18 Northern America   2017 45.12943    1         1 70.40863 59.59325 49.14806
19 Northern America   2018 46.54864    1         1 70.63614 59.68433 49.80166
20 Northern America   2019 46.32867    1         1 70.83794 59.75984 50.47739
   male_edu     temp rainfall1000
1  12.30281 5.486244    0.9971559
2  12.35258 6.469105    0.8644873
3  12.40182 5.979147    0.9460938
4  12.45053 5.416964    1.0189234
5  12.49870 5.556961    1.0008237
6  12.54635 6.187472    1.0367199
7  12.59349 6.895084    1.0917386
8  12.64015 5.900051    1.0134091
9  12.68634 5.650118    1.0693435
10 12.73207 5.398867    0.9928497
11 12.77735 6.781766    1.0379754
12 12.82218 6.269133    1.1343442
13 12.86660 7.249497    0.9747708
14 12.91059 5.954381    1.0282075
15 12.95414 5.584650    1.0377695
16 12.99723 6.436884    0.9632446
17 13.03988 7.184514    0.9677826
18 13.08210 6.539669    1.0995322
19 13.12388 6.539677    1.0991469
20 13.16522 6.539633    1.0987523
```

Observations from Ecuador should look like this...

```
finaldata %>%
  dplyr::filter(country_name == "Ecuador")
```

```
   Country.Name Year MaternalMortality InfantMortality NeonatalMortality
1       Ecuador 2000               122            24.7               14.1
2       Ecuador 2001               117            23.4               13.4
3       Ecuador 2002               110            22.4               12.7
4       Ecuador 2003               100            21.5               12.1
5       Ecuador 2004                94            20.7               11.6
```

```
6       Ecuador 2005                94            19.9              11.1
7       Ecuador 2006                90            19.2              10.6
8       Ecuador 2007                85            18.5              10.2
9       Ecuador 2008                82            17.7               9.7
10      Ecuador 2009                80            17.0               9.3
11      Ecuador 2010                78            16.3               8.9
12      Ecuador 2011                76            15.6               8.5
13      Ecuador 2012                71            14.9               8.1
14      Ecuador 2013                67            14.3               7.8
15      Ecuador 2014                65            13.7               7.5
16      Ecuador 2015                63            13.2               7.3
17      Ecuador 2016                61            12.8               7.1
18      Ecuador 2017                59            12.4               6.9
19      Ecuador 2018                NA            12.0               6.9
20      Ecuador 2019                NA            11.6               6.8
   Under5Mortality ISO earthquake drought year.x deaths country_name
1              29.5 ECU         NA      NA   1999      0      Ecuador
2              28.0 ECU         NA      NA   2000      0      Ecuador
3              26.6 ECU         NA      NA   2001      2      Ecuador
4              25.4 ECU         NA      NA   2002      0      Ecuador
5              24.4 ECU         NA      NA   2003     26      Ecuador
6              23.5 ECU         NA      NA   2004      0      Ecuador
7              22.6 ECU         NA      NA   2005      0      Ecuador
8              21.7 ECU         NA      NA   2006      0      Ecuador
9              20.8 ECU         NA      NA   2007      0      Ecuador
10             19.9 ECU          0       1   2008     25      Ecuador
11             19.0 ECU         NA      NA   2009      0      Ecuador
12             18.1 ECU         NA      NA   2010      0      Ecuador
13             17.3 ECU         NA      NA   2011      0      Ecuador
14             16.6 ECU          0       1   2012      0      Ecuador
15             15.9 ECU          1       0   2013      0      Ecuador
16             15.4 ECU         NA      NA   2014      0      Ecuador
17             14.8 ECU          1       0   2015      0      Ecuador
18             14.4 ECU         NA      NA   2016      0      Ecuador
19             13.9 ECU         NA      NA   2017      0      Ecuador
20             13.4 ECU          1       0   2018      0      Ecuador
                              region year.y  gdp1000 OECD OECD2023  popdens
1  Latin America and the Caribbean   2000 1.451531    0        0 23.27432
2  Latin America and the Caribbean   2001 1.904814    0        0 23.39372
3  Latin America and the Caribbean   2002 2.184209    0        0 23.52087
4  Latin America and the Caribbean   2003 2.438344    0        0 23.58358
5  Latin America and the Caribbean   2004 2.703566    0        0 38.43743
6  Latin America and the Caribbean   2005 3.014310    0        0 38.55361
```

```
7  Latin America and the Caribbean   2006 3.340841    0         0 38.65018
8  Latin America and the Caribbean   2007 3.579032    0         0 38.76505
9  Latin America and the Caribbean   2008 4.260433    0         0 38.83977
10 Latin America and the Caribbean   2009 4.240703    0         0 38.92613
11 Latin America and the Caribbean   2010 4.640246    0         0 39.03066
12 Latin America and the Caribbean   2011 5.202656    0         0 39.09586
13 Latin America and the Caribbean   2012 5.678456    0         0 39.13343
14 Latin America and the Caribbean   2013 6.050355    0         0 39.18619
15 Latin America and the Caribbean   2014 6.374631    0         0 39.27871
16 Latin America and the Caribbean   2015 6.130587    0         0 39.38824
17 Latin America and the Caribbean   2016 6.079089    0         0 39.46201
18 Latin America and the Caribbean   2017 6.246404    0         0 39.53609
19 Latin America and the Caribbean   2018 6.321349    0         0 39.58380
20 Latin America and the Caribbean   2019 6.233258    0         0 39.75109
      urban   agedep male_edu     temp rainfall1000
1  36.19963 67.44216 7.738627 19.54855    1.4201653
2  36.67994 66.57356 7.843942 19.66622    1.1667746
3  37.08903 65.65488 7.949449 20.24695    1.4577981
4  37.23792 64.71472 8.055240 20.05016    1.5781807
5  37.39268 63.78049 8.161433 20.10136    1.0683450
6  37.36968 62.86530 8.268176 19.88163    0.8555447
7  37.47567 61.97042 8.375587 20.07087    1.1114502
8  37.68172 61.11422 8.483729 19.49536    1.0899082
9  37.67445 60.31015 8.592603 19.85711    1.6184816
10 37.39437 59.55262 8.702180 20.39298    1.0870796
11 37.26838 58.83793 8.812409 20.11160    1.7045703
12 37.61553 58.16553 8.923172 19.86633    1.4518388
13 38.00733 57.51051 9.034284 20.19000    1.7520003
14 38.22511 56.84804 9.145523 19.85177    1.3735605
15 38.12421 56.17001 9.256679 20.42252    1.2572257
16 38.15633 55.46511 9.367582 20.95595    1.7284273
17 38.45745 54.73369 9.478071 20.77476    1.3168761
18 38.65993 53.99096 9.587993 20.53262    1.9544485
19 38.87253 53.12249 9.697221 20.53714    1.9573265
20 39.05144 52.29278 9.805670 20.54169    1.9602443
```

Exploratory data analysis

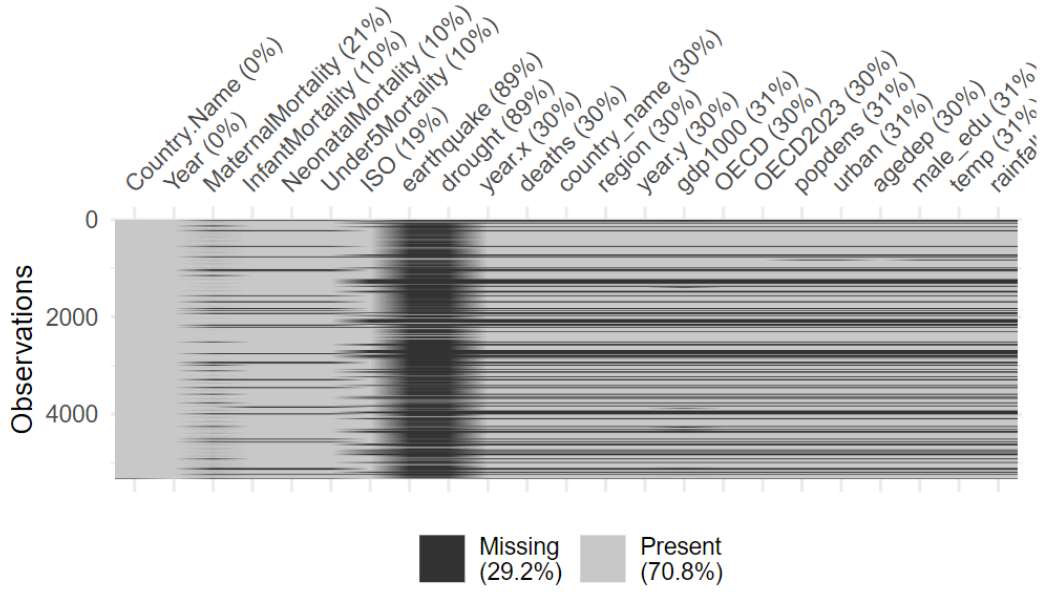
Use the rest of the class time to explore the final data that will be used for analysis starting next week. At the end of the class, write a summary of your findings and push your **Quarto document (pdf)** to your repo.

Checking global view of missing data:

```
library(naniar)
```

```
Warning: package 'naniar' was built under R version 4.3.3
```

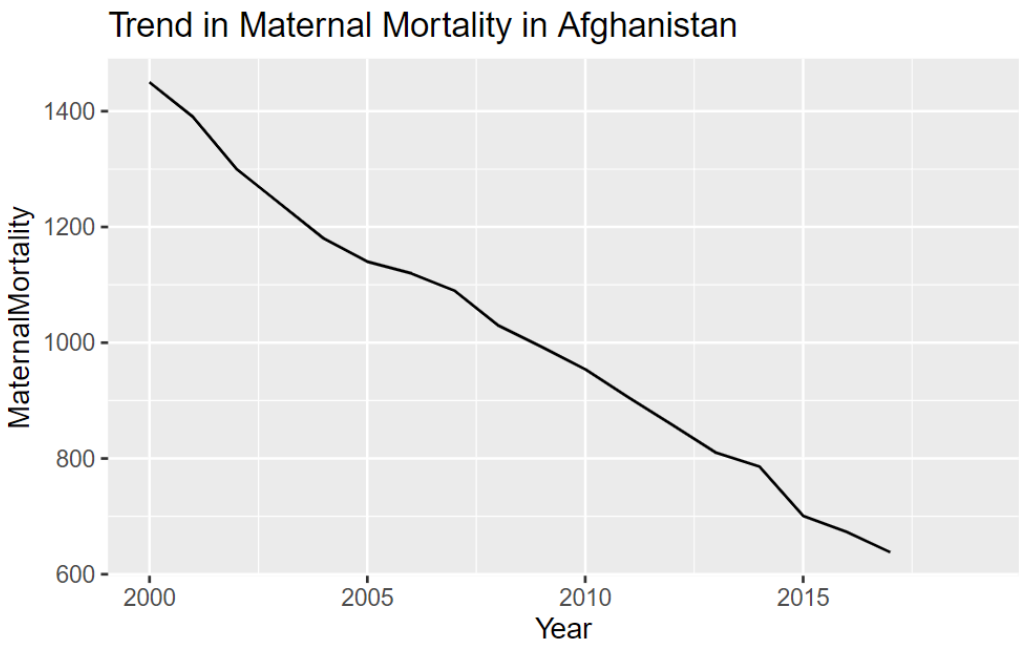
```
finaldata %>% vis_miss()
```



Check trend of a country with known conflicts:

```
finaldata %>%
  filter(Country.Name == "Afghanistan") %>%
  ggplot(aes(x = Year, y = MaternalMortality)) +
  geom_line() +
  ggtitle("Trend in Maternal Mortality in Afghanistan")
```

```
Warning: Removed 2 rows containing missing values (`geom_line()`).
```


Trend in Maternal Mortality in Afghanistan
MaternalMortality
1400
1200
1000
800
600
2000
2005
2010
2015
Year