EDA of Armed Conflict

Desmond Chen

```
library(here)
```

```
here() starts at /Users/apple/Documents/chl8010class2/armed_conflict
```

```
library(tidyverse)
```

```
-- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
v dplyr     1.1.4     v readr     2.1.5
v forcats   1.0.0     v stringr   1.5.1
v ggplot2   3.4.4     v tibble    3.2.1
v lubridate 1.9.3     v tidyr     1.3.1
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 becor
```

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

```
 [1] "ISO"          "year"         "Country.Name" "MatMor"       "InfMor"
 [6] "NeoMor"       "U5Mor"        "total_best"   "conflict"     "drought"
[11] "earthquake"   "region"       "gdp1000"      "OECD"         "OECD2023"
[16] "popdens"      "urban"        "agedep"       "male_edu"     "temp"
[21] "rainfall1000"
```

```r
finaldata$earthquake[is.na(finaldata$earthquake)] <- 0
finaldata$drought[is.na(finaldata$drought)] <- 0
```

```r
finaldata %>%
  dplyr::filter(Country.Name == "Canada")
```

```
   ISO year Country.Name MatMor InfMor NeoMor U5Mor total_best conflict drought
1  CAN 2000       Canada      9    5.3    3.8   6.2         11        0       0
2  CAN 2001       Canada     10    5.3    3.8   6.2         23        0       0
3  CAN 2002       Canada     10    5.3    3.9   6.2          1        0       0
4  CAN 2003       Canada     10    5.3    3.9   6.2          0        0       0
5  CAN 2004       Canada     10    5.3    3.9   6.1          0        0       0
6  CAN 2005       Canada     11    5.2    3.9   6.1          0        0       0
7  CAN 2006       Canada     11    5.2    3.9   6.0          0        0       0
8  CAN 2007       Canada     11    5.1    3.8   6.0          0        0       0
9  CAN 2008       Canada     12    5.1    3.8   5.9          0        0       0
10 CAN 2009       Canada     12    5.0    3.8   5.8          0        0       0
11 CAN 2010       Canada     11    5.0    3.8   5.7          0        0       0
12 CAN 2011       Canada     11    4.9    3.7   5.7          0        0       0
13 CAN 2012       Canada     11    4.9    3.7   5.6          0        0       0
14 CAN 2013       Canada     11    4.8    3.6   5.5          0        0       0
15 CAN 2014       Canada     11    4.7    3.6   5.4          0        0       0
16 CAN 2015       Canada     11    4.7    3.6   5.4          0        0       0
17 CAN 2016       Canada     10    4.6    3.5   5.3          0        0       0
18 CAN 2017       Canada     10    4.6    3.4   5.2          0        0       0
19 CAN 2018       Canada     NA    4.5    3.3   5.1          0        0       0
20 CAN 2019       Canada     NA    4.4    3.3   5.1          0        0       0
   earthquake           region  gdp1000 OECD OECD2023  popdens    urban
1           0 Northern America 24.27100    1        1 66.19704 56.14335
2           0 Northern America 23.82206    1        1 66.45361 56.40270
3           0 Northern America 24.25534    1        1 66.71112 56.67093
4           0 Northern America 28.30046    1        1 66.96384 56.94365
5           0 Northern America 32.14368    1        1 67.21715 57.20020
6           0 Northern America 36.38251    1        1 67.47283 57.41671
7           0 Northern America 40.50406    1        1 67.73674 57.59143
8           0 Northern America 44.65990    1        1 67.99444 57.75691
9           0 Northern America 46.71051    1        1 68.25765 57.97905
10          0 Northern America 40.87631    1        1 68.53354 58.24228
11          0 Northern America 47.56208    1        1 68.80739 58.52809
12          0 Northern America 52.22370    1        1 69.04842 58.81437
13          0 Northern America 52.66909    1        1 69.27604 59.05573
```

```
14          0 Northern America 52.63517    1        1 69.50772 59.19713
15          0 Northern America 50.95600    1        1 69.76876 59.30361
16          0 Northern America 43.59614    1        1 69.98853 59.42627
17          0 Northern America 42.31560    1        1 70.21484 59.50521
18          0 Northern America 45.12943    1        1 70.40863 59.59325
19          0 Northern America 46.54864    1        1 70.63614 59.68433
20          0 Northern America 46.32867    1        1 70.83794 59.75984
     agedep male_edu     temp rainfall1000
1  46.34463 12.30281 5.486244    0.9971559
2  45.89632 12.35258 6.469105    0.8644873
3  45.46660 12.40182 5.979147    0.9460938
4  45.07468 12.45053 5.416964    1.0189234
5  44.67374 12.49870 5.556961    1.0008237
6  44.26641 12.54635 6.187472    1.0367199
7  43.96370 12.59349 6.895084    1.0917386
8  43.83612 12.64015 5.900051    1.0134091
9  43.85426 12.68634 5.650118    1.0693435
10 43.94937 12.73207 5.398867    0.9928497
11 44.13587 12.77735 6.781766    1.0379754
12 44.53578 12.82218 6.269133    1.1343442
13 45.18393 12.86660 7.249497    0.9747708
14 45.95404 12.91059 5.954381    1.0282075
15 46.75493 12.95414 5.584650    1.0377695
16 47.59164 12.99723 6.436884    0.9632446
17 48.41410 13.03988 7.184514    0.9677826
18 49.14806 13.08210 6.539669    1.0995322
19 49.80166 13.12388 6.539677    1.0991469
20 50.47739 13.16522 6.539633    1.0987523
```

```
summary(finaldata)
```

```
     ISO                 year      Country.Name           MatMor       
 Length:3720        Min.   :2000   Length:3720        Min.   :   2.0  
 Class :character   1st Qu.:2005   Class :character   1st Qu.:  17.0  
 Mode  :character   Median :2010   Mode  :character   Median :  66.0  
                    Mean   :2010                      Mean   : 210.6  
                    3rd Qu.:2014                      3rd Qu.: 299.8  
                    Max.   :2019                      Max.   :2480.0  
                                                      NA's   :426     
    InfMor           NeoMor          U5Mor          total_best    
 Min.   :  1.60   Min.   : 0.80   Min.   :  2.00   Min.   :    0.0  
 1st Qu.:  7.60   1st Qu.: 4.90   1st Qu.:  9.00   1st Qu.:    0.0  
```

```
Median : 18.90   Median :12.10   Median : 22.20   Median :    0.0
Mean   : 28.90   Mean   :16.18   Mean   : 40.50   Mean   :  361.1
3rd Qu.: 44.52   3rd Qu.:25.32   3rd Qu.: 61.33   3rd Qu.:    2.0
Max.   :138.10   Max.   :60.90   Max.   :224.90   Max.   :78644.0
NA's   :20       NA's   :20      NA's   :20
   conflict         drought          earthquake         region
Min.   :0.0000   Min.   :0.00000   Min.   :0.00000   Length:3720
1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.00000   Class :character
Median :0.0000   Median :0.00000   Median :0.00000   Mode  :character
Mean   :0.1892   Mean   :0.08737   Mean   :0.08333
3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:0.00000
Max.   :1.0000   Max.   :1.00000   Max.   :1.00000

   gdp1000              OECD          OECD2023         popdens
Min.   :  0.1105   Min.   :0.000   Min.   :0.0000   Min.   : 0.00
1st Qu.:  1.2383   1st Qu.:0.000   1st Qu.:0.0000   1st Qu.:14.79
Median :  4.0719   Median :0.000   Median :0.0000   Median :27.52
Mean   : 11.4917   Mean   :0.171   Mean   :0.1882   Mean   :30.57
3rd Qu.: 13.1531   3rd Qu.:0.000   3rd Qu.:0.0000   3rd Qu.:40.72
Max.   :123.6787   Max.   :1.000   Max.   :1.0000   Max.   :99.86
NA's   :62                                          NA's   :20
    urban             agedep          male_edu           temp
Min.   : 0.1025   Min.   : 16.17   Min.   : 1.067   Min.   :-2.405
1st Qu.:17.2872   1st Qu.: 47.94   1st Qu.: 5.904   1st Qu.:12.928
Median :30.2535   Median : 55.51   Median : 8.368   Median :21.958
Mean   :30.6948   Mean   : 61.94   Mean   : 8.258   Mean   :19.625
3rd Qu.:41.6558   3rd Qu.: 77.11   3rd Qu.:10.849   3rd Qu.:25.869
Max.   :93.4135   Max.   :111.48   Max.   :14.441   Max.   :29.676
NA's   :20                         NA's   :20       NA's   :20
 rainfall1000
Min.   :0.01993
1st Qu.:0.59146
Median :1.01288
Mean   :1.20216
3rd Qu.:1.68706
Max.   :4.71081
NA's   :20
```

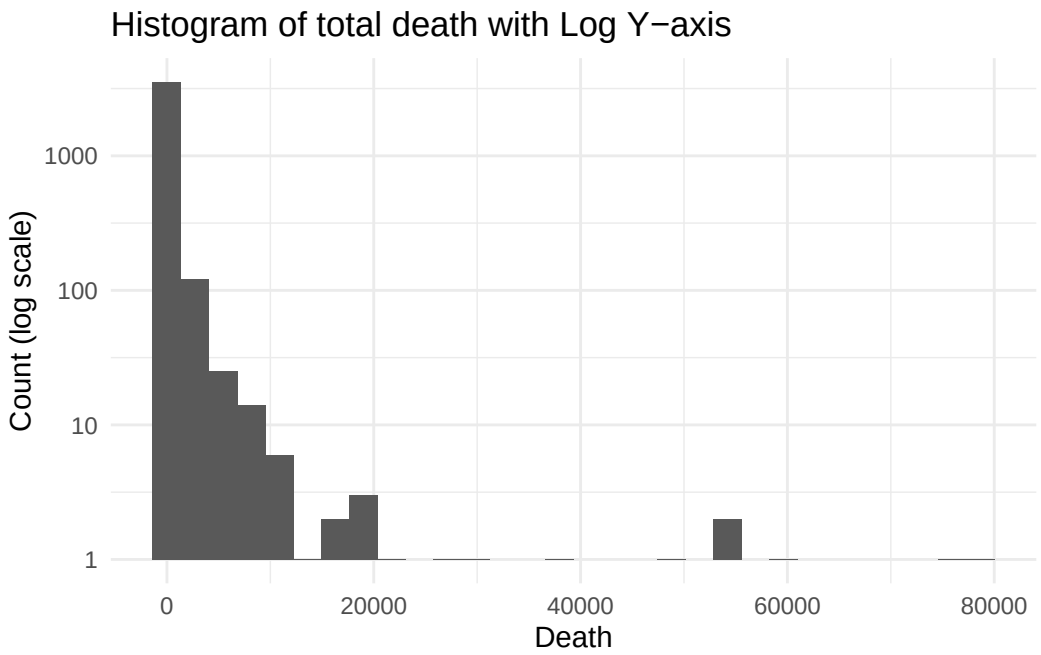
Maternal mortalities is the outcome with the most missingness of 426, followed by gdp with a missing count of 62. There seems to be a country that was not included in the covariates data, hence we see a missing count of 20 across several covariates.

```r
library(ggplot2)
ggplot(data.frame(finaldata$total_best), aes(x = finaldata$total_best)) +
  geom_histogram() +
  scale_y_log10() +
  labs(title = "Histogram of total death with Log Y-axis",
       x = "Death",
       y = "Count (log scale)") +
  theme_minimal()
```

```
`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
Warning: Transformation introduced infinite values in continuous y-axis
```

```
Warning: Removed 13 rows containing missing values (`geom_bar()`).
```



We see a lot outliers at around 55000 death, these could be caused by war happening at certain region.

```r
library(ggplot2)
ggplot(data.frame(finaldata$MatMor), aes(x = finaldata$MatMor)) +
  geom_histogram() +
  scale_y_log10() +
```

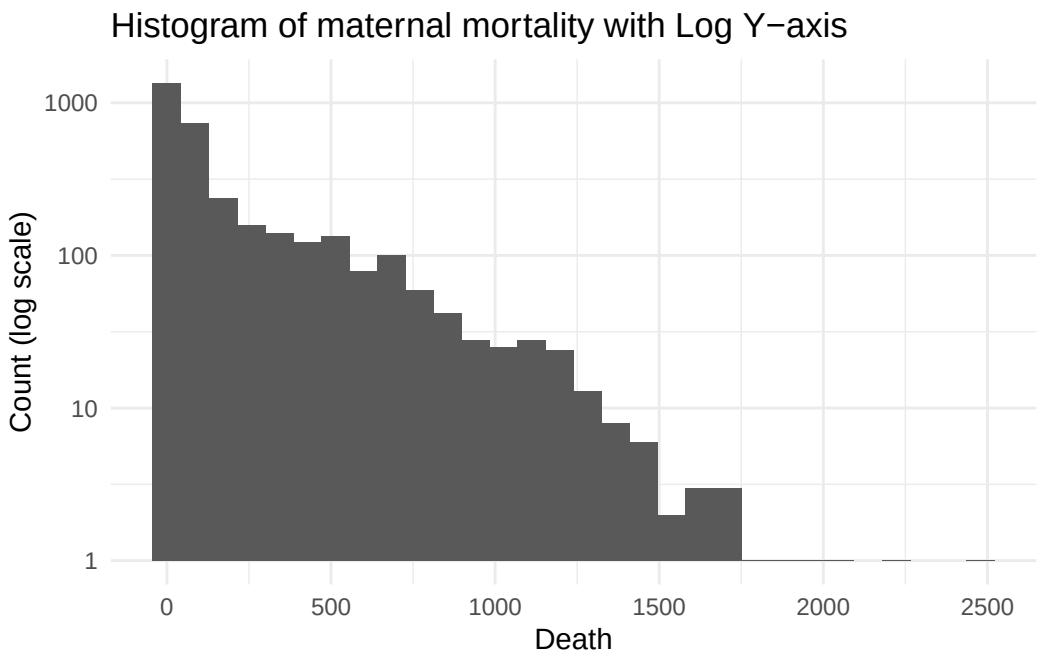
```
  labs(title = "Histogram of maternal mortality with Log Y-axis",
       x = "Death",
       y = "Count (log scale)") +
  theme_minimal()
```

```
`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
Warning: Removed 426 rows containing non-finite values (`stat_bin()`).
```

```
Warning: Transformation introduced infinite values in continuous y-axis
```

```
Warning: Removed 3 rows containing missing values (`geom_bar()`).
```

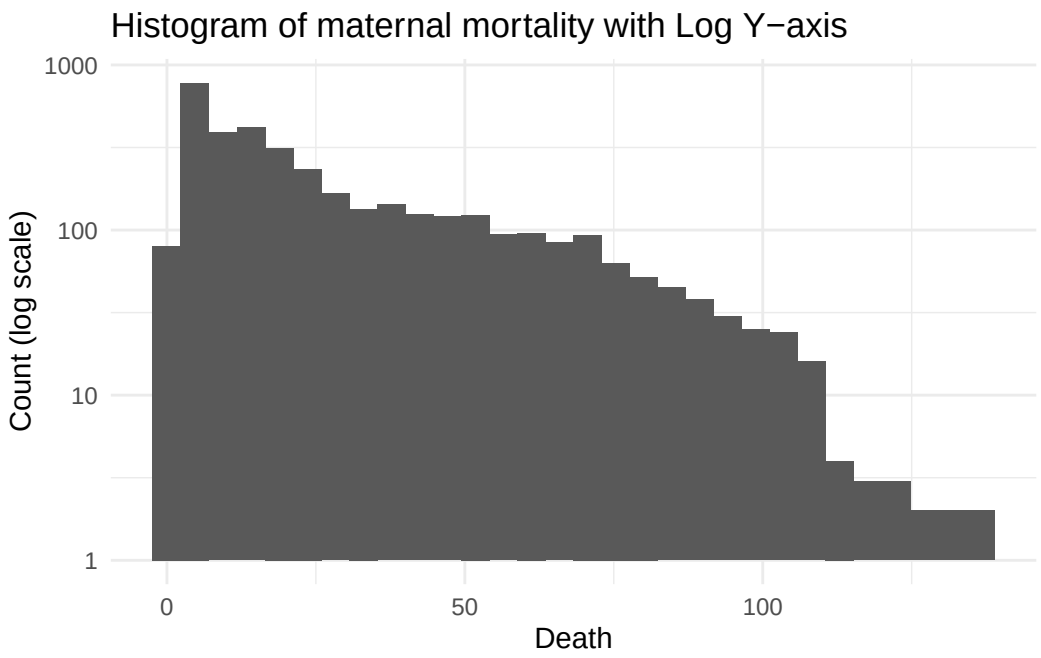


Maternal mortality looks fine, and not any outliers.

```
library(ggplot2)
ggplot(data.frame(finaldata$InfMor), aes(x = finaldata$InfMor)) +
  geom_histogram() +
  scale_y_log10() +
  labs(title = "Histogram of maternal mortality with Log Y-axis",
       x = "Death",
       y = "Count (log scale)") +
  theme_minimal()
```

```
`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
Warning: Removed 20 rows containing non-finite values (`stat_bin()`).
```

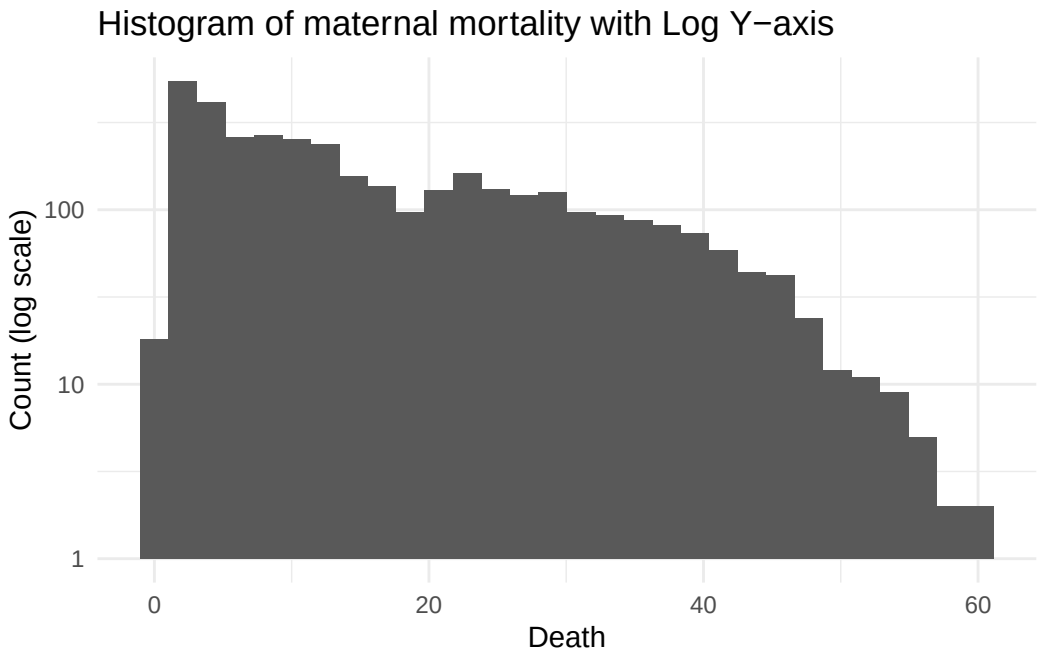


Infant mortalities looks fine as well.

```
library(ggplot2)
ggplot(data.frame(finaldata$NeoMor), aes(x = finaldata$NeoMor)) +
  geom_histogram() +
  scale_y_log10() +
  labs(title = "Histogram of maternal mortality with Log Y-axis",
       x = "Death",
       y = "Count (log scale)") +
  theme_minimal()
```

```
`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

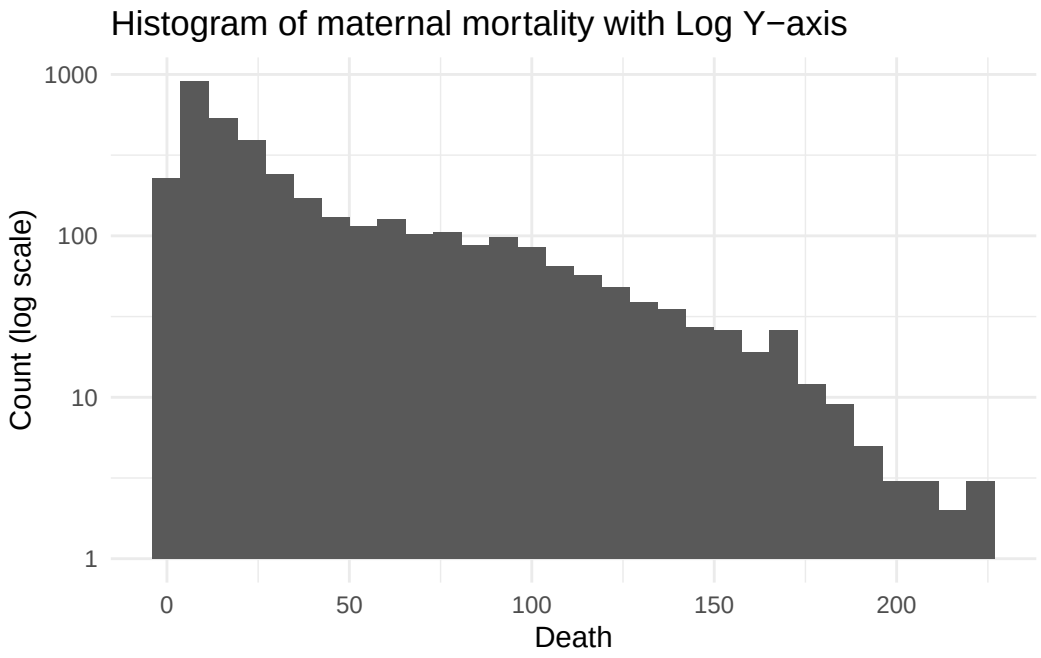
```
Warning: Removed 20 rows containing non-finite values (`stat_bin()`).
```




```
library(ggplot2)
ggplot(data.frame(finaldata$U5Mor), aes(x = finaldata$U5Mor)) +
  geom_histogram() +
  scale_y_log10() +
  labs(title = "Histogram of maternal mortality with Log Y-axis",
       x = "Death",
       y = "Count (log scale)") +
  theme_minimal()
```

```
`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
Warning: Removed 20 rows containing non-finite values (`stat_bin()`).
```




```
cor(finaldata[,c(4:7)], use = "complete.obs")
```

```
          MatMor    InfMor    NeoMor     U5Mor
MatMor 1.0000000 0.8785612 0.8354908 0.8994877
InfMor 0.8785612 1.0000000 0.9590878 0.9861117
NeoMor 0.8354908 0.9590878 1.0000000 0.9278720
U5Mor  0.8994877 0.9861117 0.9278720 1.0000000
```

We can see the main outcomes are highly correlated, so we need to be cautious when running them one by one.