ArmedConflictEDA

Xinze Yu

Study Objectives and Operationalizations

The **objective** of the current analysis is to study how armed conflicts impact the maternal and child health from a global scope.

Specifically, the primary **exposure variable** of interest is **armed conflict**, as defined by the UCDP, it is a a binary variable indicating the presence of conflict for each country–year observation (0 = no, $<$ 25 battle-related deaths; 1 = yes, $\leq$ 25 battle-related deaths).

The primary **outcome** measures are maternal, under-5, infant, and neonatal **mortality rates**.

A list of covariates is included in the dataset, and will potentially be included in the model: "gdp1000", "OECD", "OECD2023", "popdens", "urban", "agedep", "male_edu", "temp", "rainfall1000", "Drought", "Earthquake".

1. Explore Data Structure and Summary Statistics of Key Variables

We can start from checking the overall structure, data type, and missing values.

```
# import data
acdata <- read.csv(here('data', 'analytical', 'finaldata.csv'), header = TRUE)

# check structure, summary statistics, and missingness
str(acdata)
```

```
'data.frame':   3720 obs. of  21 variables:
 $ country_name  : chr  "Afghanistan" "Afghanistan" "Afghanistan" "Afghanistan" ...
 $ ISO           : chr  "AFG" "AFG" "AFG" "AFG" ...
```

```
$ region        : chr  "Southern Asia" "Southern Asia" "Southern Asia" "Southern Asia" ...
$ Year          : int  2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 ...
$ gdp1000       : num  NA NA 0.184 0.2 0.222 ...
$ OECD          : int  0 0 0 0 0 0 0 0 0 0 ...
$ OECD2023      : int  0 0 0 0 0 0 0 0 0 0 ...
$ popdens       : num  14.1 14.2 14.3 14.4 15.2 ...
$ urban         : num  16.3 16.3 16.4 16.6 16.7 ...
$ agedep        : num  108 109 109 109 109 ...
$ male_edu      : num  2.76 2.86 2.95 3.05 3.16 ...
$ temp          : num  12.7 12.9 12.7 12.2 13 ...
$ rainfall1000  : num  0.276 0.279 0.381 0.429 0.375 ...
$ MatMortality  : int  1450 1390 1300 1240 1180 1140 1120 1090 1030 993 ...
$ InfMortality  : num  90.5 87.9 85.3 82.7 80 77.3 74.6 71.9 69.2 66.7 ...
$ NeoMortality  : num  60.9 59.7 58.5 57.2 55.9 54.6 53.2 51.7 50.3 48.9 ...
$ Und5Mortality : num  129 125 121 117 113 ...
$ totdeath      : int  5065 5394 5553 1157 944 817 1711 4982 7020 5660 ...
$ armed_conflict: int  1 1 1 1 1 1 1 1 1 1 ...
$ Drought       : int  1 0 0 0 0 0 1 0 1 0 ...
$ Earthquake    : int  0 1 1 1 1 1 1 0 0 1 ...
```

```
summary(acdata)
```

```
 country_name           ISO              region              Year
 Length:3720        Length:3720        Length:3720        Min.   :2000
 Class :character   Class :character   Class :character   1st Qu.:2005
 Mode  :character   Mode  :character   Mode  :character   Median :2010
                                                          Mean   :2010
                                                          3rd Qu.:2014
                                                          Max.   :2019

    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      MatMortality     InfMortality     NeoMortality
 Min.   :0.01993   Min.   :   2.0   Min.   :  1.60   Min.   : 0.80
 1st Qu.:0.59146   1st Qu.:  17.0   1st Qu.:  7.60   1st Qu.: 4.90
 Median :1.01288   Median :  66.0   Median : 18.90   Median :12.10
 Mean   :1.20216   Mean   : 210.6   Mean   : 28.90   Mean   :16.18
 3rd Qu.:1.68706   3rd Qu.: 299.8   3rd Qu.: 44.52   3rd Qu.:25.32
 Max.   :4.71081   Max.   :2480.0   Max.   :138.10   Max.   :60.90
 NA's   :20        NA's   :426      NA's   :20       NA's   :20
 Und5Mortality       totdeath       armed_conflict      Drought
 Min.   :  2.00   Min.   :    0.0   Min.   :0.0000   Min.   :0.00000
 1st Qu.:  9.00   1st Qu.:    0.0   1st Qu.:0.0000   1st Qu.:0.00000
 Median : 22.20   Median :    0.0   Median :0.0000   Median :0.00000
 Mean   : 40.50   Mean   :  361.1   Mean   :0.1892   Mean   :0.08737
 3rd Qu.: 61.33   3rd Qu.:    2.0   3rd Qu.:0.0000   3rd Qu.:0.00000
 Max.   :224.90   Max.   :78644.0   Max.   :1.0000   Max.   :1.00000
 NA's   :20
   Earthquake
 Min.   :0.00000
 1st Qu.:0.00000
 Median :0.00000
 Mean   :0.08333
 3rd Qu.:0.00000
 Max.   :1.00000
```

Note that one of the key outcome variables, maternal mortality, has noticeable missingness - 426 (11.5%) observations were missing out of 3720 observations.

2. Visualize the Distributions

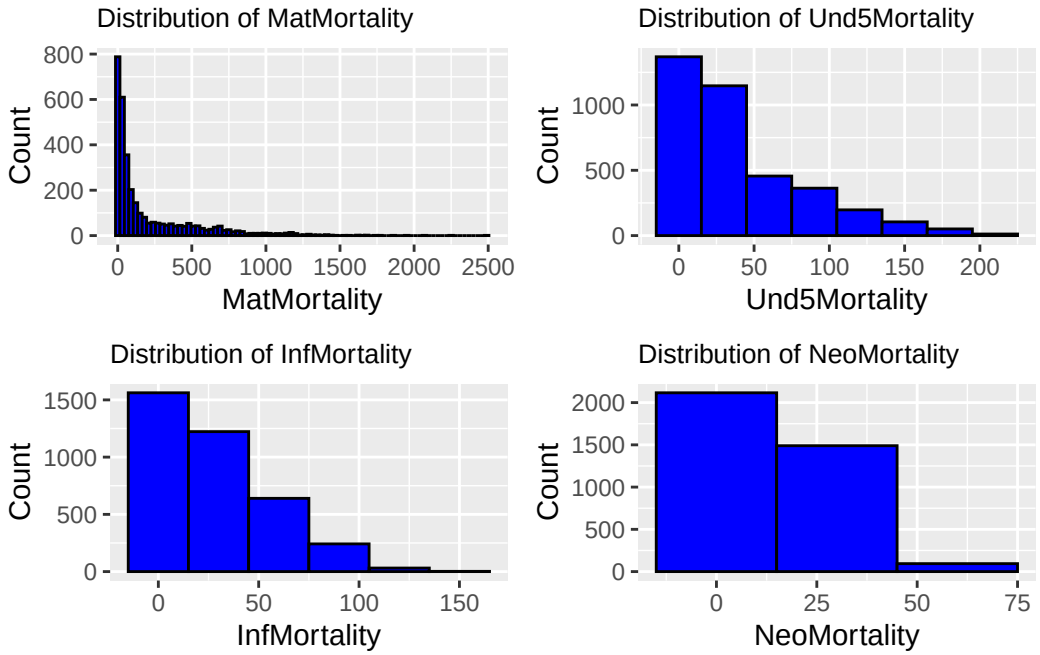
Plot **histograms** for outcome variables and covariates to identify skewness or outliers.

```
# define outcome
outcomes <- c("MatMortality", "Und5Mortality", "InfMortality", "NeoMortality")

# create empty list to store plots
plot_list <- list()
```

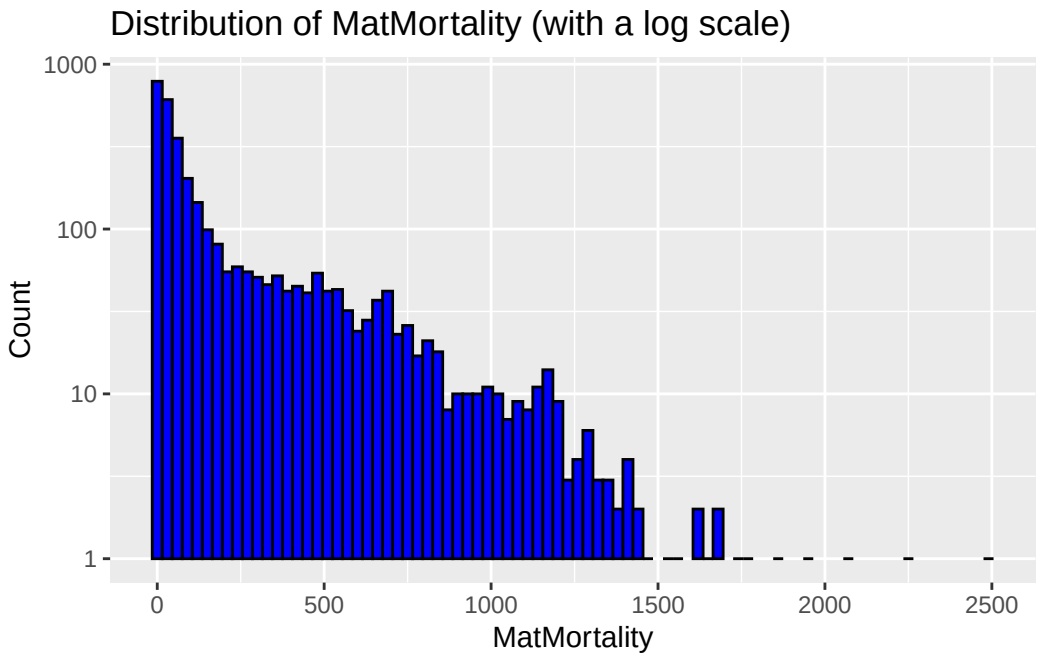
```r
# create histogram for each outcome
for (var in outcomes) {
  p <- ggplot(acdata, aes_string(x = var)) +
    geom_histogram(binwidth = 30, fill = "blue", color = "black") +
    labs(title = paste("Distribution of", var), x = var, y = "Count") +
    theme(plot.title = element_text(size = 10))
  plot_list[[var]] <- p
}

# arrange plots into grid layout
grid.arrange(grobs = plot_list, ncol = 2)
```



The largely empty graph in maternal mortality suggests the presence of outliers. We can proceed to use logarithms.

```r
# create histogram for maternal mortality with log scale
ggplot(acdata, aes_string(x = "MatMortality")) +
    geom_histogram(binwidth = 30, fill = "blue", color = "black") +
    scale_y_log10() +
    labs(title = paste("Distribution of", "MatMortality", "(with a log scale)"),
         x = "MatMortality", y = "Count")
```



To be finished...