Exploratory Analysis of Conflict Dataset

Mohsyn Imran Malik

Exploratory Analysis

Here is our final dataset created from last week

`combo_df`

First, we will load the `tidyverse, dplyr and here` package to aid in our exploratory analysis and add our dataset

```
library(tidyverse)
```

```
Warning: package 'tidyverse' was built under R version 4.4.1
```

```
-- 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   3.5.1     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
```

```
library(dplyr)
library(here)
```

```
Warning: package 'here' was built under R version 4.4.1
```

```
here() starts at C:/Users/MoMal/OneDrive/Masters/CHL8010H/Week 2/Disaster Analysis/disaster
```

```
combo_df<- read.csv(here("final_conflict.csv"))
combo_df<- select(combo_df, !X.1)
```

Lets start by looking at head and tail of dataset

```
head(combo_df)
```

```
  Year ISO total_droughts total_earthquakes Total_Best bin_conflict
1 2000 AFG              1                 0          0            0
2 2000 ARM              1                 0          0            0
3 2000 AZE              1                 1          0            0
4 2000 BGD              0                 1          0            0
5 2000 BGR              1                 0          0            0
6 2000 BIH              1                 0          0            0
            country_name          region   gdp1000 OECD OECD2023   popdens
1            Afghanistan   Southern Asia        NA    0        0 14.136539
2                Armenia    Western Asia 0.6032980    0        0 33.862027
3             Azerbaijan    Western Asia 0.6551199    0        0 27.971922
4             Bangladesh   Southern Asia 0.4131002    0        0 57.706342
5               Bulgaria  Eastern Europe 1.6212430    0        0  8.428015
6 Bosnia and Herzegovina Southern Europe 1.3321224    0        0  6.759416
      urban    agedep  male_edu      temp rainfall1000   X Maternal.Mortality
1 16.253241 108.34663  2.762086 12.699593    0.2763704  41               1450
2 41.810238  56.21203 10.781853  7.708524    0.3592467 201                 43
3 25.818459  57.51161 10.572949 13.554278    0.3500783 301                 47
4 27.601130  69.69181  4.380392 25.200782    2.3153057 401                434
5 14.294645  47.96771 11.156618 11.409854    0.4055068 421                 19
6  8.521575  44.86417  9.860110 11.149017    0.7873840 481                 17
  Infant.Mortality Neonatal.Mortality Under.5.Mortality
1             90.5               60.9             129.2
2             27.0               16.1              30.6
3             60.9               34.3              74.4
4             63.1               44.0              86.0
5             14.4                7.9              17.5
6              8.7                6.9               9.9
```

```
tail(combo_df)
```

```
     Year ISO total_droughts total_earthquakes Total_Best bin_conflict
4304 2014 VIR              0                 0          0            0
```

```
4305 2015 VIR              0                 0          0            0
4306 2016 VIR              0                 0          0            0
4307 2017 VIR              0                 0          0            0
4308 2018 VIR              0                 0          0            0
4309 2019 VIR              0                 0          0            0
     country_name region gdp1000 OECD OECD2023 popdens urban agedep male_edu
4304         <NA>   <NA>      NA   NA       NA      NA    NA     NA       NA
4305         <NA>   <NA>      NA   NA       NA      NA    NA     NA       NA
4306         <NA>   <NA>      NA   NA       NA      NA    NA     NA       NA
4307         <NA>   <NA>      NA   NA       NA      NA    NA     NA       NA
4308         <NA>   <NA>      NA   NA       NA      NA    NA     NA       NA
4309         <NA>   <NA>      NA   NA       NA      NA    NA     NA       NA
     temp rainfall1000    X Maternal.Mortality Infant.Mortality
4304   NA            NA 5135                 NA               NA
4305   NA            NA 5136                 NA               NA
4306   NA            NA 5137                 NA               NA
4307   NA            NA 5138                 NA               NA
4308   NA            NA 5139                 NA               NA
4309   NA            NA 5140                 NA               NA
     Neonatal.Mortality Under.5.Mortality
4304                 NA                NA
4305                 NA                NA
4306                 NA                NA
4307                 NA                NA
4308                 NA                NA
4309                 NA                NA
```

We can print summary stats (mean, median and interquartile range) of our dataset and get an idea of the missing data (NA)

```
summary_df <- summary(combo_df)

summary(combo_df)
```

```
      Year          ISO            total_droughts    total_earthquakes
 Min.   :2000   Length:4309        Min.   :0.00000   Min.   : 0.0000
 1st Qu.:2004   Class :character   1st Qu.:0.00000   1st Qu.: 0.0000
 Median :2009   Mode  :character   Median :0.00000   Median : 0.0000
 Mean   :2009                      Mean   :0.07867   Mean   : 0.1279
 3rd Qu.:2014                      3rd Qu.:0.00000   3rd Qu.: 0.0000
 Max.   :2019                      Max.   :3.00000   Max.   :11.0000
```

```
  Total_Best       bin_conflict    country_name          region
Min.   :    0.0   Min.   :0.0000   Length:4309        Length:4309
1st Qu.:    0.0   1st Qu.:0.0000   Class :character   Class :character
Median :    0.0   Median :0.0000   Mode  :character   Mode  :character
Mean   :  288.8   Mean   :0.2165
3rd Qu.:    0.0   3rd Qu.:0.0000
Max.   :78644.0   Max.   :1.0000

   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   :651        NA's   :589     NA's   :589      NA's   :609
    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   :609       NA's   :589      NA's   :609      NA's   :609
 rainfall1000          X        Maternal.Mortality Infant.Mortality
Min.   :0.0199   Min.   :   1   Min.   :   2.0     Min.   :  1.60
1st Qu.:0.5915   1st Qu.:1196   1st Qu.:  17.0     1st Qu.:  7.50
Median :1.0129   Median :2610   Median :  64.0     Median : 18.80
Mean   :1.2022   Mean   :2631   Mean   : 208.6     Mean   : 28.33
3rd Qu.:1.6871   3rd Qu.:3985   3rd Qu.: 292.8     3rd Qu.: 43.00
Max.   :4.7108   Max.   :5320   Max.   :2480.0     Max.   :138.10
NA's   :609      NA's   :9      NA's   :979        NA's   :449
Neonatal.Mortality Under.5.Mortality
Min.   : 0.80      Min.   :  1.800
1st Qu.: 4.80      1st Qu.:  8.875
Median :12.00      Median : 21.950
Mean   :15.95      Mean   : 39.575
3rd Qu.:24.73      3rd Qu.: 58.800
Max.   :60.90      Max.   :224.900
NA's   :449        NA's   :449
```

We can also summarize by country or year (warning: large data output). Given this will produce many stats, we only printed the head results

```r
summary_country <- by(combo_df, combo_df$ISO, summary)
head(summary_country)
```

```
$ABW
      Year          ISO             total_droughts total_earthquakes
 Min.   :2000   Length:20          Min.   :0      Min.   :0
 1st Qu.:2005   Class :character   1st Qu.:0      1st Qu.:0
 Median :2010   Mode  :character   Median :0      Median :0
 Mean   :2010                      Mean   :0      Mean   :0
 3rd Qu.:2014                      3rd Qu.:0      3rd Qu.:0
 Max.   :2019                      Max.   :0      Max.   :0

   Total_Best  bin_conflict country_name          region             gdp1000
 Min.   :0    Min.   :0     Length:20          Length:20          Min.   : NA
 1st Qu.:0    1st Qu.:0     Class :character   Class :character   1st Qu.: NA
 Median :0    Median :0     Mode  :character   Mode  :character   Median : NA
 Mean   :0    Mean   :0                                           Mean   :NaN
 3rd Qu.:0    3rd Qu.:0                                           3rd Qu.: NA
 Max.   :0    Max.   :0                                           Max.   : NA
                                                                  NA's   :20
      OECD         OECD2023      popdens         urban          agedep
 Min.   : NA   Min.   : NA   Min.   : NA   Min.   : NA   Min.   : NA
 1st Qu.: NA   1st Qu.: NA   1st Qu.: NA   1st Qu.: NA   1st Qu.: NA
 Median : NA   Median : NA   Median : NA   Median : NA   Median : NA
 Mean   :NaN   Mean   :NaN   Mean   :NaN   Mean   :NaN   Mean   :NaN
 3rd Qu.: NA   3rd Qu.: NA   3rd Qu.: NA   3rd Qu.: NA   3rd Qu.: NA
 Max.   : NA   Max.   : NA   Max.   : NA   Max.   : NA   Max.   : NA
 NA's   :20    NA's   :20    NA's   :20    NA's   :20    NA's   :20
    male_edu        temp       rainfall1000        X          Maternal.Mortality
 Min.   : NA   Min.   : NA   Min.   : NA   Min.   : 1.00   Min.   : NA
 1st Qu.: NA   1st Qu.: NA   1st Qu.: NA   1st Qu.: 5.75   1st Qu.: NA
 Median : NA   Median : NA   Median : NA   Median :10.50   Median : NA
 Mean   :NaN   Mean   :NaN   Mean   :NaN   Mean   :10.50   Mean   :NaN
 3rd Qu.: NA   3rd Qu.: NA   3rd Qu.: NA   3rd Qu.:15.25   3rd Qu.: NA
 Max.   : NA   Max.   : NA   Max.   : NA   Max.   :20.00   Max.   : NA
 NA's   :20    NA's   :20    NA's   :20                    NA's   :20
 Infant.Mortality Neonatal.Mortality Under.5.Mortality
 Min.   : NA      Min.   : NA        Min.   : NA
 1st Qu.: NA      1st Qu.: NA        1st Qu.: NA
 Median : NA      Median : NA        Median : NA
 Mean   :NaN      Mean   :NaN        Mean   :NaN
 3rd Qu.: NA      3rd Qu.: NA        3rd Qu.: NA
```

```
 Max.   : NA      Max.   : NA        Max.   : NA
 NA's   :20       NA's   :20         NA's   :20

$AFG
      Year          ISO            total_droughts total_earthquakes
 Min.   :2000   Length:20          Min.   :0.00   Min.   :0.0
 1st Qu.:2005   Class :character   1st Qu.:0.00   1st Qu.:0.0
 Median :2010   Mode  :character   Median :0.00   Median :1.0
 Mean   :2010                      Mean   :0.25   Mean   :0.8
 3rd Qu.:2014                      3rd Qu.:0.25   3rd Qu.:1.0
 Max.   :2019                      Max.   :1.00   Max.   :3.0

   Total_Best     bin_conflict  country_name          region
 Min.   :    0   Min.   :0.00   Length:20          Length:20
 1st Qu.: 4164   1st Qu.:1.00   Class :character   Class :character
 Median : 6760   Median :1.00   Mode  :character   Mode  :character
 Mean   : 8316   Mean   :0.95
 3rd Qu.: 9236   3rd Qu.:1.00
 Max.   :26889   Max.   :1.00

    gdp1000            OECD      OECD2023    popdens          urban
 Min.   :0.1835   Min.   :0   Min.   :0   Min.   :14.14   Min.   :16.25
 1st Qu.:0.2993   1st Qu.:0   1st Qu.:0   1st Qu.:15.31   1st Qu.:16.82
 Median :0.5013   Median :0   Median :0   Median :15.87   Median :17.52
 Mean   :0.4504   Mean   :0   Mean   :0   Mean   :15.66   Mean   :17.65
 3rd Qu.:0.5830   3rd Qu.:0   3rd Qu.:0   3rd Qu.:16.22   3rd Qu.:18.44
 Max.   :0.6631   Max.   :0   Max.   :0   Max.   :16.79   Max.   :19.41
 NA's   :2
     agedep          male_edu          temp        rainfall1000
 Min.   : 87.65   Min.   :2.762   Min.   :11.90   Min.   :0.2764
 1st Qu.: 94.29   1st Qu.:3.236   1st Qu.:12.54   1st Qu.:0.3724
 Median :105.88   Median :3.776   Median :12.71   Median :0.4052
 Mean   :101.78   Mean   :3.838   Mean   :12.68   Mean   :0.4095
 3rd Qu.:108.51   3rd Qu.:4.400   3rd Qu.:12.91   3rd Qu.:0.4454
 Max.   :109.45   Max.   :5.124   Max.   :13.23   Max.   :0.5481

       X         Maternal.Mortality Infant.Mortality Neonatal.Mortality
 Min.   :41.00   Min.   : 638       Min.   :46.40    Min.   :36.10
 1st Qu.:45.75   1st Qu.: 822       1st Qu.:54.70    1st Qu.:41.55
 Median :50.50   Median :1012       Median :65.45    Median :48.15
 Mean   :50.50   Mean   :1014       Mean   :66.62    Mean   :48.27
 3rd Qu.:55.25   3rd Qu.:1170       3rd Qu.:77.97    3rd Qu.:54.92
 Max.   :60.00   Max.   :1450       Max.   :90.50    Max.   :60.90
```

```
                NA's   :2
 Under.5.Mortality
 Min.   : 60.10
 1st Qu.: 72.80
 Median : 89.75
 Mean   : 91.62
 3rd Qu.:109.45
 Max.   :129.20


$AGO
      Year          ISO             total_droughts total_earthquakes
 Min.   :2000   Length:20          Min.   :0.0    Min.   :0
 1st Qu.:2005   Class :character   1st Qu.:0.0    1st Qu.:0
 Median :2010   Mode  :character   Median :0.0    Median :0
 Mean   :2010                      Mean   :0.2    Mean   :0
 3rd Qu.:2014                      3rd Qu.:0.0    3rd Qu.:0
 Max.   :2019                      Max.   :1.0    Max.   :0

   Total_Best       bin_conflict  country_name          region
 Min.   :   0.00   Min.   :0.00   Length:20          Length:20
 1st Qu.:   1.75   1st Qu.:1.00   Class :character   Class :character
 Median :  19.50   Median :1.00   Mode  :character   Mode  :character
 Mean   : 211.20   Mean   :0.85
 3rd Qu.:  29.00   3rd Qu.:1.00
 Max.   :2154.00   Max.   :1.00

    gdp1000            OECD      OECD2023    popdens           urban
 Min.   :0.5275   Min.   :0   Min.   :0   Min.   : 1.770   Min.   :31.02
 1st Qu.:1.5958   1st Qu.:0   1st Qu.:0   1st Qu.: 3.664   1st Qu.:31.78
 Median :2.5427   Median :0   Median :0   Median :10.586   Median :33.11
 Mean   :2.6937   Mean   :0   Mean   :0   Mean   :15.189   Mean   :33.61
 3rd Qu.:3.6430   3rd Qu.:0   3rd Qu.:0   3rd Qu.:23.882   3rd Qu.:35.48
 Max.   :5.1020   Max.   :0   Max.   :0   Max.   :42.489   Max.   :37.06

     agedep         male_edu          temp        rainfall1000
 Min.   :92.31   Min.   :4.420   Min.   :21.34   Min.   :0.6939
 1st Qu.:93.49   1st Qu.:5.115   1st Qu.:21.68   1st Qu.:0.8108
 Median :93.58   Median :5.836   Median :21.88   Median :0.9682
 Mean   :93.77   Mean   :5.860   Mean   :22.10   Mean   :0.9200
 3rd Qu.:94.01   3rd Qu.:6.590   3rd Qu.:22.43   3rd Qu.:1.0379
 Max.   :95.58   Max.   :7.380   Max.   :23.19   Max.   :1.0730
```

```
       X          Maternal.Mortality Infant.Mortality Neonatal.Mortality
 Min.   : 81.00   Min.   :241.0      Min.   : 49.90   Min.   :27.90
 1st Qu.: 85.75   1st Qu.:272.0      1st Qu.: 59.80   1st Qu.:31.48
 Median : 90.50   Median :377.0      Median : 78.25   Median :37.00
 Mean   : 90.50   Mean   :435.2      Mean   : 81.55   Mean   :37.90
 3rd Qu.: 95.25   3rd Qu.:560.2      3rd Qu.:102.10   3rd Qu.:43.90
 Max.   :100.00   Max.   :827.0      Max.   :121.50   Max.   :50.30
                  NA's   :2
 Under.5.Mortality
 Min.   : 74.2
 1st Qu.: 91.7
 Median :124.5
 Mean   :131.1
 3rd Qu.:168.3
 Max.   :204.4


$ALB
      Year          ISO             total_droughts total_earthquakes
 Min.   :2000   Length:20          Min.   :0      Min.   :0.0
 1st Qu.:2005   Class :character   1st Qu.:0      1st Qu.:0.0
 Median :2010   Mode  :character   Median :0      Median :0.0
 Mean   :2010                      Mean   :0      Mean   :0.2
 3rd Qu.:2014                      3rd Qu.:0      3rd Qu.:0.0
 Max.   :2019                      Max.   :0      Max.   :3.0

   Total_Best  bin_conflict country_name          region
 Min.   :0    Min.   :0     Length:20          Length:20
 1st Qu.:0    1st Qu.:0     Class :character   Class :character
 Median :0    Median :0     Mode  :character   Mode  :character
 Mean   :0    Mean   :0
 3rd Qu.:0    3rd Qu.:0
 Max.   :0    Max.   :0

    gdp1000           OECD      OECD2023    popdens          urban
 Min.   :1.127   Min.   :0   Min.   :0   Min.   :33.08   Min.   :27.39
 1st Qu.:2.599   1st Qu.:0   1st Qu.:0   1st Qu.:36.41   1st Qu.:30.82
 Median :4.104   Median :0   Median :0   Median :39.34   Median :34.36
 Mean   :3.542   Mean   :0   Mean   :0   Mean   :40.26   Mean   :33.81
 3rd Qu.:4.419   3rd Qu.:0   3rd Qu.:0   3rd Qu.:43.89   3rd Qu.:37.19
 Max.   :5.396   Max.   :0   Max.   :0   Max.   :48.57   Max.   :38.26

     agedep         male_edu          temp        rainfall1000
```

```
 Min.   :46.41   Min.   :8.962   Min.   :12.72   Min.   :0.7942
 1st Qu.:46.87   1st Qu.:9.248   1st Qu.:13.53   1st Qu.:1.0211
 Median :48.87   Median :9.498   Median :13.85   Median :1.0512
 Mean   :50.85   Mean   :9.481   Mean   :13.76   Mean   :1.0675
 3rd Qu.:54.60   3rd Qu.:9.725   3rd Qu.:14.10   3rd Qu.:1.0998
 Max.   :59.66   Max.   :9.941   Max.   :14.21   Max.   :1.3823

       X         Maternal.Mortality Infant.Mortality Neonatal.Mortality
 Min.   :101.0   Min.   :15.00      Min.   : 8.40    Min.   : 5.900
 1st Qu.:105.8   1st Qu.:16.25      1st Qu.: 8.75    1st Qu.: 6.275
 Median :110.5   Median :19.50      Median :12.35    Median : 7.450
 Mean   :110.5   Mean   :19.06      Mean   :13.90    Mean   : 8.355
 3rd Qu.:115.2   3rd Qu.:21.00      3rd Qu.:18.12    3rd Qu.:10.675
 Max.   :120.0   Max.   :23.00      Max.   :24.10    Max.   :12.100
                 NA's   :2
 Under.5.Mortality
 Min.   : 9.40
 1st Qu.: 9.85
 Median :13.90
 Mean   :15.62
 3rd Qu.:20.38
 Max.   :27.20


$AND
      Year          ISO             total_droughts total_earthquakes
 Min.   :2000   Length:20          Min.   :0      Min.   :0
 1st Qu.:2005   Class :character   1st Qu.:0      1st Qu.:0
 Median :2010   Mode  :character   Median :0      Median :0
 Mean   :2010                      Mean   :0      Mean   :0
 3rd Qu.:2014                      3rd Qu.:0      3rd Qu.:0
 Max.   :2019                      Max.   :0      Max.   :0

   Total_Best  bin_conflict country_name          region
 Min.   :0    Min.   :0     Length:20          Length:20
 1st Qu.:0    1st Qu.:0     Class :character   Class :character
 Median :0    Median :0     Mode  :character   Mode  :character
 Mean   :0    Mean   :0
 3rd Qu.:0    3rd Qu.:0
 Max.   :0    Max.   :0

    gdp1000           OECD      OECD2023    popdens          urban
 Min.   :21.62   Min.   :0   Min.   :0   Min.   :20.89   Min.   :37.57
```

```
 1st Qu.:38.57   1st Qu.:0   1st Qu.:0   1st Qu.:23.51   1st Qu.:40.60
 Median :42.12   Median :0   Median :0   Median :28.12   Median :41.59
 Mean   :40.71   Mean   :0   Mean   :0   Mean   :26.47   Mean   :41.17
 3rd Qu.:46.32   3rd Qu.:0   3rd Qu.:0   3rd Qu.:29.22   3rd Qu.:42.35
 Max.   :53.72   Max.   :0   Max.   :0   Max.   :30.78   Max.   :42.77

    agedep         male_edu          temp         rainfall1000
 Min.   :36.62   Min.   :11.68   Min.   : 8.438   Min.   :0.6811
 1st Qu.:37.37   1st Qu.:11.96   1st Qu.: 9.167   1st Qu.:0.7377
 Median :38.03   Median :12.24   Median : 9.552   Median :0.7729
 Mean   :38.00   Mean   :12.23   Mean   : 9.441   Mean   :0.8298
 3rd Qu.:38.71   3rd Qu.:12.51   3rd Qu.: 9.741   3rd Qu.:0.9311
 Max.   :39.09   Max.   :12.77   Max.   :10.131   Max.   :1.0790

      X         Maternal.Mortality Infant.Mortality Neonatal.Mortality
 Min.   :121.0   Min.   : NA        Min.   :2.500    Min.   :1.400
 1st Qu.:125.8   1st Qu.: NA        1st Qu.:3.250    1st Qu.:1.775
 Median :130.5   Median : NA        Median :4.250    Median :2.250
 Mean   :130.5   Mean   :NaN        Mean   :4.510    Mean   :2.375
 3rd Qu.:135.2   3rd Qu.: NA        3rd Qu.:5.675    3rd Qu.:2.925
 Max.   :140.0   Max.   : NA        Max.   :7.300    Max.   :3.800
                 NA's   :20
 Under.5.Mortality
 Min.   :2.70
 1st Qu.:3.45
 Median :4.65
 Mean   :4.94
 3rd Qu.:6.20
 Max.   :8.20


$ARE
     Year          ISO            total_droughts total_earthquakes
 Min.   :2000   Length:20          Min.   :0      Min.   :0
 1st Qu.:2005   Class :character   1st Qu.:0      1st Qu.:0
 Median :2010   Mode  :character   Median :0      Median :0
 Mean   :2010                      Mean   :0      Mean   :0
 3rd Qu.:2014                      3rd Qu.:0      3rd Qu.:0
 Max.   :2019                      Max.   :0      Max.   :0

   Total_Best    bin_conflict  country_name         region
 Min.   :0.00   Min.   :0.00   Length:20          Length:20
 1st Qu.:0.00   1st Qu.:0.00   Class :character   Class :character
```

```
 Median :0.00   Median :0.00   Mode  :character   Mode  :character
 Mean   :0.05   Mean   :0.05
 3rd Qu.:0.00   3rd Qu.:0.00
 Max.   :1.00   Max.   :1.00

   gdp1000           OECD      OECD2023    popdens          urban
 Min.   :29.91   Min.   :0   Min.   :0   Min.   :0.000   Min.   :4.900
 1st Qu.:33.78   1st Qu.:0   1st Qu.:0   1st Qu.:0.000   1st Qu.:5.448
 Median :41.86   Median :0   Median :0   Median :4.802   Median :6.065
 Mean   :39.99   Mean   :0   Mean   :0   Mean   :3.109   Mean   :5.905
 3rd Qu.:45.19   3rd Qu.:0   3rd Qu.:0   3rd Qu.:4.858   3rd Qu.:6.367
 Max.   :46.87   Max.   :0   Max.   :0   Max.   :4.881   Max.   :6.603

    agedep         male_edu           temp        rainfall1000
 Min.   :16.17   Min.   : 8.068   Min.   :27.16   Min.   :0.03444
 1st Qu.:17.44   1st Qu.: 8.924   1st Qu.:27.37   1st Qu.:0.05601
 Median :18.45   Median : 9.772   Median :27.42   Median :0.07221
 Mean   :20.77   Mean   : 9.763   Mean   :27.51   Mean   :0.07553
 3rd Qu.:24.91   3rd Qu.:10.608   3rd Qu.:27.78   3rd Qu.:0.09291
 Max.   :29.99   Max.   :11.423   Max.   :28.03   Max.   :0.12819

      X         Maternal.Mortality Infant.Mortality Neonatal.Mortality
 Min.   :161.0   Min.   :3.000      Min.   :5.80     Min.   :3.700
 1st Qu.:165.8   1st Qu.:3.250      1st Qu.:6.65     1st Qu.:4.175
 Median :170.5   Median :4.000      Median :7.50     Median :4.700
 Mean   :170.5   Mean   :4.167      Mean   :7.58     Mean   :4.725
 3rd Qu.:175.2   3rd Qu.:5.000      3rd Qu.:8.45     3rd Qu.:5.225
 Max.   :180.0   Max.   :6.000      Max.   :9.60     Max.   :5.900
                 NA's   :2
 Under.5.Mortality
 Min.   : 6.800
 1st Qu.: 7.750
 Median : 8.800
 Mean   : 8.850
 3rd Qu.: 9.875
 Max.   :11.100
```

```
summary_year <- by(combo_df, combo_df$Year, summary)
head(summary_year)
```

```
$`2000`
```

```
     Year          ISO             total_droughts   total_earthquakes
Min.   :2000   Length:215         Min.   :0.0000   Min.   :0.0000
1st Qu.:2000   Class :character   1st Qu.:0.0000   1st Qu.:0.0000
Median :2000   Mode  :character   Median :0.0000   Median :0.0000
Mean   :2000                      Mean   :0.1163   Mean   :0.1395
3rd Qu.:2000                      3rd Qu.:0.0000   3rd Qu.:0.0000
Max.   :2000                      Max.   :3.0000   Max.   :5.0000

  Total_Best  bin_conflict country_name          region
Min.   :0    Min.   :0     Length:215         Length:215
1st Qu.:0    1st Qu.:0     Class :character   Class :character
Median :0    Median :0     Mode  :character   Mode  :character
Mean   :0    Mean   :0
3rd Qu.:0    3rd Qu.:0
Max.   :0    Max.   :0

   gdp1000             OECD           OECD2023         popdens
Min.   : 0.1230   Min.   :0.0000   Min.   :0.0000   Min.   : 0.00
1st Qu.: 0.6033   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:13.42
Median : 1.7716   Median :0.0000   Median :0.0000   Median :25.36
Mean   : 6.3931   Mean   :0.1613   Mean   :0.1882   Mean   :28.27
3rd Qu.: 6.1209   3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:37.71
Max.   :48.6596   Max.   :1.0000   Max.   :1.0000   Max.   :99.81
NA's   :34        NA's   :29       NA's   :29       NA's   :30
    urban            agedep          male_edu          temp
Min.   : 0.106   Min.   : 29.99   Min.   : 1.067   Min.   :-1.207
1st Qu.:15.363   1st Qu.: 49.46   1st Qu.: 4.896   1st Qu.:12.998
Median :27.998   Median : 63.49   Median : 7.144   Median :21.437
Mean   :29.128   Mean   : 67.62   Mean   : 7.364   Mean   :19.399
3rd Qu.:39.697   3rd Qu.: 84.68   3rd Qu.:10.116   3rd Qu.:25.446
Max.   :91.551   Max.   :111.47   Max.   :13.957   Max.   :28.576
NA's   :30       NA's   :29       NA's   :30       NA's   :30
 rainfall1000          X        Maternal.Mortality Infant.Mortality
Min.   :0.04797   Min.   :   1   Min.   :   3.0     Min.   :  3.0
1st Qu.:0.59155   1st Qu.:1191   1st Qu.:  24.0     1st Qu.: 10.7
Median :1.00404   Median :2601   Median :  77.0     Median : 26.2
Mean   :1.19906   Mean   :2622   Mean   : 274.4     Mean   : 38.4
3rd Qu.:1.63095   3rd Qu.:3951   3rd Qu.: 437.0     3rd Qu.: 61.6
Max.   :4.71081   Max.   :5301   Max.   :2480.0     Max.   :138.1
NA's   :30                       NA's   :30         NA's   :22
Neonatal.Mortality Under.5.Mortality
Min.   : 1.6       Min.   :  3.90
1st Qu.: 6.9       1st Qu.: 12.60
```

```
 Median :16.6       Median : 30.70
 Mean   :20.1       Mean   : 55.98
 3rd Qu.:30.7       3rd Qu.: 89.60
 Max.   :60.9       Max.   :224.90
 NA's   :22         NA's   :22

$`2001`
      Year          ISO            total_droughts   total_earthquakes
 Min.   :2001   Length:215         Min.   :0.0000   Min.   :0.0000
 1st Qu.:2001   Class :character   1st Qu.:0.0000   1st Qu.:0.0000
 Median :2001   Mode  :character   Median :0.0000   Median :0.0000
 Mean   :2001                      Mean   :0.1023   Mean   :0.1163
 3rd Qu.:2001                      3rd Qu.:0.0000   3rd Qu.:0.0000
 Max.   :2001                      Max.   :2.0000   Max.   :5.0000

   Total_Best       bin_conflict    country_name          region
 Min.   :    0.0   Min.   :0.0000   Length:215         Length:215
 1st Qu.:    0.0   1st Qu.:0.0000   Class :character   Class :character
 Median :    0.0   Median :0.0000   Mode  :character   Mode  :character
 Mean   :  435.8   Mean   :0.2558
 3rd Qu.:    1.0   3rd Qu.:1.0000
 Max.   :48666.0   Max.   :1.0000

    gdp1000             OECD           OECD2023         popdens
 Min.   : 0.1193   Min.   :0.0000   Min.   :0.0000   Min.   : 0.00
 1st Qu.: 0.5710   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:13.64
 Median : 1.8241   Median :0.0000   Median :0.0000   Median :25.28
 Mean   : 6.2987   Mean   :0.1613   Mean   :0.1882   Mean   :28.34
 3rd Qu.: 6.4317   3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:37.63
 Max.   :48.4401   Max.   :1.0000   Max.   :1.0000   Max.   :99.81
 NA's   :33        NA's   :29       NA's   :29       NA's   :30
     urban             agedep          male_edu          temp
 Min.   : 0.1059   Min.   : 28.54   Min.   : 1.110   Min.   :-0.9179
 1st Qu.:15.6496   1st Qu.: 49.34   1st Qu.: 4.994   1st Qu.:13.1622
 Median :28.2383   Median : 62.12   Median : 7.248   Median :21.6022
 Mean   :29.2987   Mean   : 66.78   Mean   : 7.456   Mean   :19.4927
 3rd Qu.:39.8887   3rd Qu.: 83.79   3rd Qu.:10.185   3rd Qu.:25.7861
 Max.   :91.6767   Max.   :111.48   Max.   :13.982   Max.   :29.0397
 NA's   :30        NA's   :29       NA's   :30       NA's   :30
  rainfall1000           X        Maternal.Mortality Infant.Mortality
 Min.   :0.02092   Min.   :   2   Min.   :   3.0     Min.   :  2.80
 1st Qu.:0.62196   1st Qu.:1192   1st Qu.:  23.0     1st Qu.: 10.50
 Median :0.97431   Median :2602   Median :  75.0     Median : 25.30
```

```
 Mean   :1.16535   Mean   :2623   Mean   : 266.2     Mean   : 37.05
 3rd Qu.:1.66361   3rd Qu.:3952   3rd Qu.: 423.0     3rd Qu.: 59.80
 Max.   :3.30938   Max.   :5302   Max.   :2250.0     Max.   :135.60
 NA's   :30                       NA's   :30         NA's   :22
 Neonatal.Mortality Under.5.Mortality
 Min.   : 1.50      Min.   :  3.60
 1st Qu.: 6.50      1st Qu.: 12.20
 Median :16.40      Median : 29.40
 Mean   :19.56      Mean   : 53.84
 3rd Qu.:29.90      3rd Qu.: 87.10
 Max.   :59.70      Max.   :219.40
 NA's   :22         NA's   :22

$`2002`
      Year          ISO            total_droughts   total_earthquakes
 Min.   :2002   Length:217         Min.   :0.0000   Min.   :0.0000
 1st Qu.:2002   Class :character   1st Qu.:0.0000   1st Qu.:0.0000
 Median :2002   Mode  :character   Median :0.0000   Median :0.0000
 Mean   :2002                      Mean   :0.1106   Mean   :0.1705
 3rd Qu.:2002                      3rd Qu.:0.0000   3rd Qu.:0.0000
 Max.   :2002                      Max.   :3.0000   Max.   :5.0000

   Total_Best      bin_conflict   country_name         region
 Min.   :   0.0   Min.   :0.000   Length:217         Length:217
 1st Qu.:   0.0   1st Qu.:0.000   Class :character   Class :character
 Median :   0.0   Median :0.000   Mode  :character   Mode  :character
 Mean   : 170.1   Mean   :0.235
 3rd Qu.:   0.0   3rd Qu.:0.000
 Max.   :5553.0   Max.   :1.000

    gdp1000             OECD           OECD2023         popdens
 Min.   : 0.1105   Min.   :0.0000   Min.   :0.0000   Min.   : 0.00
 1st Qu.: 0.6003   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:13.65
 Median : 2.0040   Median :0.0000   Median :0.0000   Median :25.28
 Mean   : 6.6308   Mean   :0.1613   Mean   :0.1882   Mean   :28.44
 3rd Qu.: 6.4302   3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:38.07
 Max.   :53.0057   Max.   :1.0000   Max.   :1.0000   Max.   :99.83
 NA's   :34        NA's   :31       NA's   :31       NA's   :32
     urban             agedep          male_edu          temp
 Min.   : 0.1057   Min.   : 27.36   Min.   : 1.153   Min.   :-0.405
 1st Qu.:15.9361   1st Qu.: 49.46   1st Qu.: 5.091   1st Qu.:12.926
 Median :28.7748   Median : 60.73   Median : 7.352   Median :21.745
 Mean   :29.4671   Mean   : 65.96   Mean   : 7.549   Mean   :19.607
```

```
 3rd Qu.:40.0837   3rd Qu.: 82.73   3rd Qu.:10.254   3rd Qu.:25.941
 Max.   :91.7872   Max.   :111.20   Max.   :14.007   Max.   :28.986
 NA's   :32        NA's   :31       NA's   :32       NA's   :32
  rainfall1000           X        Maternal.Mortality Infant.Mortality
 Min.   :0.05426   Min.   :   3   Min.   :   3.0     Min.   :  2.60
 1st Qu.:0.57718   1st Qu.:1193   1st Qu.:  23.0     1st Qu.: 10.00
 Median :1.06445   Median :2603   Median :  72.0     Median : 24.70
 Mean   :1.18742   Mean   :2624   Mean   : 259.4     Mean   : 35.71
 3rd Qu.:1.70077   3rd Qu.:3953   3rd Qu.: 410.0     3rd Qu.: 57.00
 Max.   :3.61017   Max.   :5303   Max.   :2080.0     Max.   :132.90
 NA's   :32        NA's   :2      NA's   :32         NA's   :24
 Neonatal.Mortality Under.5.Mortality
 Min.   : 1.40      Min.   :  3.4
 1st Qu.: 6.20      1st Qu.: 11.6
 Median :15.90      Median : 28.5
 Mean   :19.02      Mean   : 51.7
 3rd Qu.:29.30      3rd Qu.: 84.5
 Max.   :58.50      Max.   :213.9
 NA's   :24         NA's   :24

$`2003`
      Year          ISO            total_droughts    total_earthquakes
 Min.   :2003   Length:215         Min.   :0.00000   Min.   : 0.000
 1st Qu.:2003   Class :character   1st Qu.:0.00000   1st Qu.: 0.000
 Median :2003   Mode  :character   Median :0.00000   Median : 0.000
 Mean   :2003                      Mean   :0.06512   Mean   : 0.186
 3rd Qu.:2003                      3rd Qu.:0.00000   3rd Qu.: 0.000
 Max.   :2003                      Max.   :2.00000   Max.   :11.000

   Total_Best      bin_conflict    country_name          region
 Min.   :   0.0   Min.   :0.0000   Length:215         Length:215
 1st Qu.:   0.0   1st Qu.:0.0000   Class :character   Class :character
 Median :   0.0   Median :0.0000   Mode  :character   Mode  :character
 Mean   : 181.5   Mean   :0.2233
 3rd Qu.:   0.0   3rd Qu.:0.0000
 Max.   :7931.0   Max.   :1.0000

    gdp1000            OECD           OECD2023         popdens
 Min.   : 0.1144   Min.   :0.0000   Min.   :0.0000   Min.   : 0.00
 1st Qu.: 0.6904   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:13.66
 Median : 2.2595   Median :0.0000   Median :0.0000   Median :25.54
 Mean   : 7.7371   Mean   :0.1613   Mean   :0.1882   Mean   :29.00
 3rd Qu.: 7.6710   3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:38.43
```

```
 Max.   :65.6893   Max.   :1.0000   Max.   :1.0000   Max.   :99.85
 NA's   :32        NA's   :29       NA's   :29       NA's   :30
     urban            agedep          male_edu          temp
 Min.   : 0.106   Min.   : 26.43   Min.   : 1.197   Min.   :-1.266
 1st Qu.:16.223   1st Qu.: 49.48   1st Qu.: 5.188   1st Qu.:12.941
 Median :29.038   Median : 59.79   Median : 7.460   Median :21.841
 Mean   :29.617   Mean   : 65.14   Mean   : 7.641   Mean   :19.497
 3rd Qu.:40.255   3rd Qu.: 81.00   3rd Qu.:10.321   3rd Qu.:25.903
 Max.   :91.904   Max.   :110.72   Max.   :14.032   Max.   :29.078
 NA's   :30       NA's   :29       NA's   :30       NA's   :30
  rainfall1000          X        Maternal.Mortality Infant.Mortality
 Min.   :0.0480   Min.   :   4   Min.   :   3.0     Min.   :  2.50
 1st Qu.:0.5806   1st Qu.:1194   1st Qu.:  21.0     1st Qu.:  9.70
 Median :1.0064   Median :2604   Median :  70.0     Median : 23.50
 Mean   :1.1797   Mean   :2625   Mean   : 250.8     Mean   : 34.44
 3rd Qu.:1.6474   3rd Qu.:3954   3rd Qu.: 395.0     3rd Qu.: 54.20
 Max.   :3.7505   Max.   :5304   Max.   :1960.0     Max.   :130.20
 NA's   :30                      NA's   :30         NA's   :22
 Neonatal.Mortality Under.5.Mortality
 Min.   : 1.30      Min.   :  3.20
 1st Qu.: 5.90      1st Qu.: 11.30
 Median :15.20      Median : 27.60
 Mean   :18.51      Mean   : 49.65
 3rd Qu.:28.70      3rd Qu.: 78.00
 Max.   :57.20      Max.   :208.10
 NA's   :22         NA's   :22

$`2004`
      Year          ISO            total_droughts    total_earthquakes
 Min.   :2004   Length:216         Min.   :0.00000   Min.   :0.0000
 1st Qu.:2004   Class :character   1st Qu.:0.00000   1st Qu.:0.0000
 Median :2004   Mode  :character   Median :0.00000   Median :0.0000
 Mean   :2004                      Mean   :0.05093   Mean   :0.1944
 3rd Qu.:2004                      3rd Qu.:0.00000   3rd Qu.:0.0000
 Max.   :2004                      Max.   :1.00000   Max.   :6.0000

   Total_Best    bin_conflict   country_name          region
 Min.   :   0   Min.   :0.000   Length:216         Length:216
 1st Qu.:   0   1st Qu.:0.000   Class :character   Class :character
 Median :   0   Median :0.000   Mode  :character   Mode  :character
 Mean   : 171   Mean   :0.213
 3rd Qu.:   0   3rd Qu.:0.000
 Max.   :8021   Max.   :1.000
```

```
    gdp1000              OECD           OECD2023         popdens
 Min.   : 0.1285   Min.   :0.0000   Min.   :0.0000   Min.   : 0.00
 1st Qu.: 0.7933   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:14.07
 Median : 2.6248   Median :0.0000   Median :0.0000   Median :26.46
 Mean   : 8.9156   Mean   :0.1613   Mean   :0.1882   Mean   :29.42
 3rd Qu.: 9.3353   3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:38.67
 Max.   :76.5449   Max.   :1.0000   Max.   :1.0000   Max.   :99.84
 NA's   :33        NA's   :30       NA's   :30       NA's   :31
     urban             agedep          male_edu          temp
 Min.   : 0.1059   Min.   : 25.68   Min.   : 1.241   Min.   :-0.1579
 1st Qu.:16.3131   1st Qu.: 49.40   1st Qu.: 5.285   1st Qu.:13.0464
 Median :29.2236   Median : 58.96   Median : 7.576   Median :21.8446
 Mean   :29.7832   Mean   : 64.33   Mean   : 7.734   Mean   :19.4702
 3rd Qu.:40.4320   3rd Qu.: 79.20   3rd Qu.:10.387   3rd Qu.:25.7555
 Max.   :92.0230   Max.   :110.43   Max.   :14.056   Max.   :29.1929
 NA's   :31        NA's   :30       NA's   :31       NA's   :31
  rainfall1000          X        Maternal.Mortality Infant.Mortality
 Min.   :0.0548   Min.   :   5   Min.   :   3.0     Min.   :  2.40
 1st Qu.:0.6118   1st Qu.:1195   1st Qu.:  19.0     1st Qu.:  9.30
 Median :1.0347   Median :2605   Median :  71.0     Median : 22.40
 Mean   :1.1904   Mean   :2626   Mean   : 241.3     Mean   : 33.25
 3rd Qu.:1.7164   3rd Qu.:3955   3rd Qu.: 372.0     3rd Qu.: 51.20
 Max.   :4.5577   Max.   :5305   Max.   :1850.0     Max.   :127.20
 NA's   :31       NA's   :1      NA's   :31         NA's   :23
 Neonatal.Mortality Under.5.Mortality
 Min.   : 1.30      Min.   :  3.0
 1st Qu.: 5.80      1st Qu.: 11.0
 Median :14.50      Median : 26.5
 Mean   :18.02      Mean   : 47.8
 3rd Qu.:28.20      3rd Qu.: 76.9
 Max.   :55.90      Max.   :202.0
 NA's   :23         NA's   :23

$`2005`
      Year          ISO            total_droughts    total_earthquakes
 Min.   :2005   Length:215         Min.   :0.00000   Min.   :0.0000
 1st Qu.:2005   Class :character   1st Qu.:0.00000   1st Qu.:0.0000
 Median :2005   Mode  :character   Median :0.00000   Median :0.0000
 Mean   :2005                      Mean   :0.09302   Mean   :0.1163
 3rd Qu.:2005                      3rd Qu.:0.00000   3rd Qu.:0.0000
 Max.   :2005                      Max.   :2.00000   Max.   :4.0000
```

```
  Total_Best      bin_conflict    country_name          region
Min.   :   0.0   Min.   :0.0000   Length:215         Length:215
1st Qu.:   0.0   1st Qu.:0.0000   Class :character   Class :character
Median :   0.0   Median :0.0000   Mode  :character   Mode  :character
Mean   : 156.7   Mean   :0.2326
3rd Qu.:   0.0   3rd Qu.:0.0000
Max.   :9714.0   Max.   :1.0000

   gdp1000             OECD           OECD2023         popdens
Min.   : 0.1512   Min.   :0.0000   Min.   :0.0000   Min.   : 0.00
1st Qu.: 0.9602   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:14.13
Median : 3.0727   Median :0.0000   Median :0.0000   Median :26.57
Mean   : 9.7970   Mean   :0.1613   Mean   :0.1882   Mean   :29.64
3rd Qu.:10.5234   3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:38.96
Max.   :80.9881   Max.   :1.0000   Max.   :1.0000   Max.   :99.86
NA's   :32        NA's   :29       NA's   :29       NA's   :30
    urban             agedep          male_edu          temp
Min.   : 0.1057   Min.   : 24.65   Min.   : 1.286   Min.   :-1.495
1st Qu.:16.2738   1st Qu.: 48.72   1st Qu.: 5.368   1st Qu.:12.717
Median :29.4426   Median : 57.50   Median : 7.692   Median :21.982
Mean   :29.9554   Mean   : 63.53   Mean   : 7.827   Mean   :19.469
3rd Qu.:40.6204   3rd Qu.: 78.19   3rd Qu.:10.452   3rd Qu.:25.898
Max.   :92.1499   Max.   :110.10   Max.   :14.080   Max.   :29.422
NA's   :30        NA's   :29       NA's   :30       NA's   :30
 rainfall1000          X        Maternal.Mortality Infant.Mortality
Min.   :0.0472   Min.   :   6   Min.   :   3.0     Min.   :  2.30
1st Qu.:0.5922   1st Qu.:1196   1st Qu.:  19.0     1st Qu.:  8.90
Median :0.9785   Median :2606   Median :  70.0     Median : 21.10
Mean   :1.1999   Mean   :2627   Mean   : 230.2     Mean   : 32.02
3rd Qu.:1.7210   3rd Qu.:3956   3rd Qu.: 348.0     3rd Qu.: 49.60
Max.   :4.2317   Max.   :5306   Max.   :1760.0     Max.   :124.10
NA's   :30                      NA's   :30         NA's   :22
Neonatal.Mortality Under.5.Mortality
Min.   : 1.20      Min.   :  2.90
1st Qu.: 5.60      1st Qu.: 10.50
Median :13.90      Median : 24.90
Mean   :17.56      Mean   : 45.77
3rd Qu.:27.10      3rd Qu.: 73.60
Max.   :54.60      Max.   :195.50
NA's   :22         NA's   :22
```

Note that we have significant amounts of missing data (NA) for a number of variables in the dataset.

Lets compare mean and standard deviation of continuous variables data when we remove all rows with missing data

```
c <- c(3:5, 9, 12, 15:17, 19:22)

mean <- sapply(combo_df[, c], mean, na.rm = TRUE)

mean
```

```
    total_droughts  total_earthquakes         Total_Best             gdp1000
        0.07867255         0.12787190       288.82757020        11.49170787
           popdens            male_edu               temp       rainfall1000
       30.57232405         8.25788555        19.62503908         1.20216333
Maternal.Mortality   Infant.Mortality Neonatal.Mortality  Under.5.Mortality
      208.63753754        28.33264249        15.95038860        39.57450777
```

```
sd <- sapply(combo_df[, c], sd, na.rm = TRUE)

sd
```

```
    total_droughts  total_earthquakes         Total_Best             gdp1000
         0.2810686          0.5933206       2654.9382326         17.3520689
           popdens            male_edu               temp       rainfall1000
        20.7518837          3.0197460          7.3339364          0.8098114
Maternal.Mortality   Infant.Mortality Neonatal.Mortality  Under.5.Mortality
       302.7328968        26.0851106         12.8546394         41.7954393
```

Now lets remove rows with missing data and find new means and sd

```
remove_na <- combo_df[complete.cases(combo_df[ , c]),]

rem_mean <- sapply(remove_na[, c], mean, na.rm = TRUE)

rem_mean
```

```
    total_droughts  total_earthquakes         Total_Best             gdp1000
        0.09141966         0.15163807       320.73447738        11.07575743
           popdens            male_edu               temp       rainfall1000
       30.57721372         8.14653147        19.58812564         1.18518389
Maternal.Mortality   Infant.Mortality Neonatal.Mortality  Under.5.Mortality
      200.94820593        29.12081123        16.20099844        40.91847114
```

```
 rem_sd <- sapply(remove_na[, c], sd, na.rm = TRUE)

rem_sd
```

```
    total_droughts  total_earthquakes         Total_Best             gdp1000
         0.3030299          0.6443609       2897.3491648         17.0752396
           popdens           male_edu               temp       rainfall1000
        20.7486647          3.0168235          7.3334064          0.7910381
Maternal.Mortality   Infant.Mortality Neonatal.Mortality  Under.5.Mortality
       289.9559318         26.4691722         12.9235184         42.5964634
```

We can do the same but this time by using multiple imputations to account for missing data

```
library(mice)
```

```
Warning: package 'mice' was built under R version 4.4.1
```

```

Attaching package: 'mice'

The following object is masked from 'package:stats':

    filter

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

    cbind, rbind
```

```
mult_imp <- mice(combo_df, print = FALSE)
```

```
Warning: Number of logged events: 3
```

```
mice_est <- complete(mult_imp) |> as_tibble()

imp_mean <- sapply(mice_est[, c], mean, na.rm = TRUE)
imp_mean
```

```
    total_droughts  total_earthquakes         Total_Best            gdp1000
         0.07867255         0.12787190       288.82757020        11.28053233
           popdens           male_edu               temp       rainfall1000
        30.41281490         8.30321312        19.63554919         1.22534340
Maternal.Mortality   Infant.Mortality Neonatal.Mortality  Under.5.Mortality
       193.37154792        28.17191924        15.81083778        39.29997679
```

```
imp_sd <- sapply(mice_est[, c], sd, na.rm = TRUE)
imp_sd
```

```
    total_droughts  total_earthquakes         Total_Best            gdp1000
         0.2810686          0.5933206       2654.9382326         16.9188112
           popdens           male_edu               temp       rainfall1000
        20.7154543          2.9673472          7.3134303          0.8181066
Maternal.Mortality   Infant.Mortality Neonatal.Mortality  Under.5.Mortality
       283.7749650         25.5834800         12.6425413         41.0000810
```

Use cbind to visualize the differences in means on method of dealing with missing data in table format

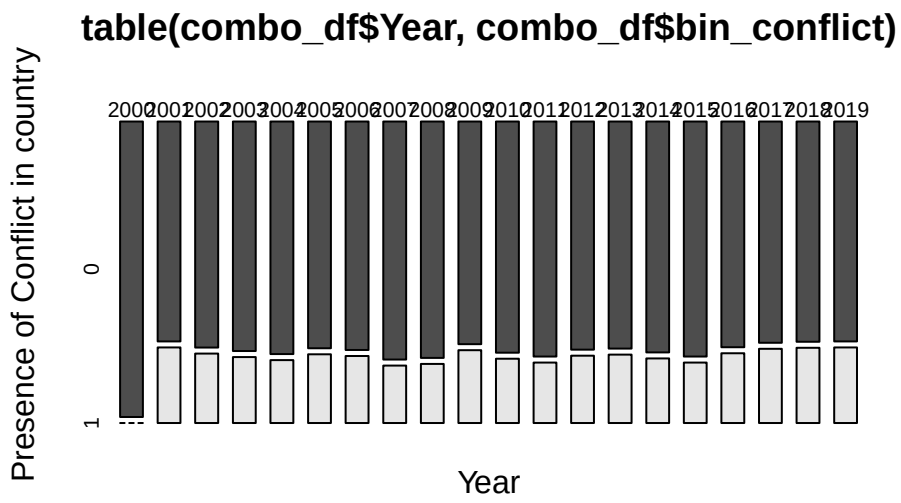
```
md_df <- cbind(mean, rem_mean, imp_mean, sd, rem_sd, imp_sd)

md_df<- as.data.frame(md_df)

md_df<- rownames_to_column(md_df, var = "variables")

md_df
```

```
            variables         mean     rem_mean     imp_mean           sd
1      total_droughts   0.07867255   0.09141966   0.07867255    0.2810686
2   total_earthquakes   0.12787190   0.15163807   0.12787190    0.5933206
3          Total_Best 288.82757020 320.73447738 288.82757020 2654.9382326
4             gdp1000  11.49170787  11.07575743  11.28053233   17.3520689
5             popdens  30.57232405  30.57721372  30.41281490   20.7518837
6            male_edu   8.25788555   8.14653147   8.30321312    3.0197460
7                temp  19.62503908  19.58812564  19.63554919    7.3339364
8        rainfall1000   1.20216333   1.18518389   1.22534340    0.8098114
9  Maternal.Mortality 208.63753754 200.94820593 193.37154792  302.7328968
10   Infant.Mortality  28.33264249  29.12081123  28.17191924   26.0851106
11 Neonatal.Mortality  15.95038860  16.20099844  15.81083778   12.8546394
12  Under.5.Mortality  39.57450777  40.91847114  39.29997679   41.7954393
```

```
         rem_sd       imp_sd
1     0.3030299    0.2810686
2     0.6443609    0.5933206
3  2897.3491648 2654.9382326
4    17.0752396   16.9188112
5    20.7486647   20.7154543
6     3.0168235    2.9673472
7     7.3334064    7.3134303
8     0.7910381    0.8181066
9   289.9559318  283.7749650
10   26.4691722   25.5834800
11   12.9235184   12.6425413
12   42.5964634   41.0000810
```

Lets make a visual plot of contingency tables using mosaicplots.
We can look at year and binary conflict variable.

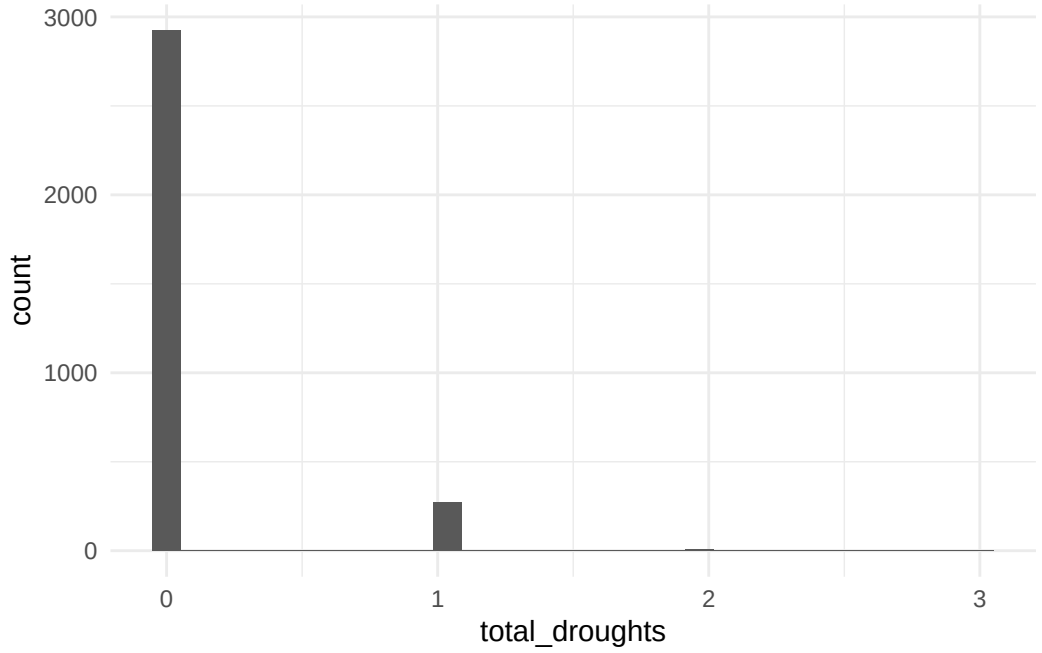
```
mosaicplot(table(combo_df$Year, combo_df$bin_conflict), color = TRUE, xlab = "Year", ylab =
```



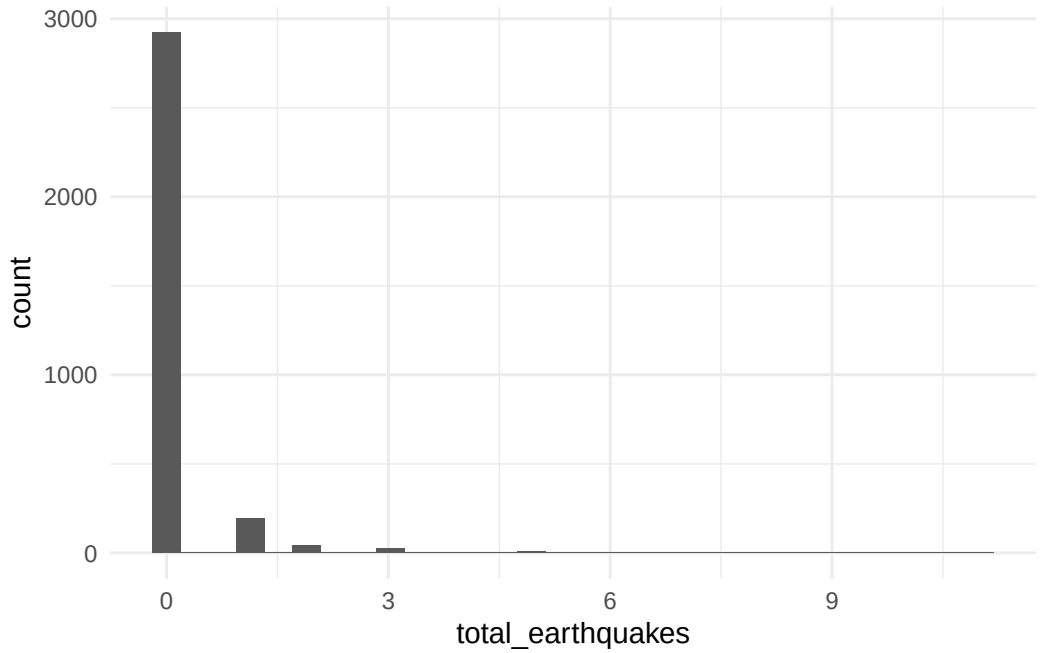
We can visualize the spread of data using histograms for our continuous variables#Lets look at total droughts, total earthquakes, total battle related death, gdp100, popdens, male_edu, temp, rainfall1000, maternal.mortality, infant.mortality, neonatal.mortality, and under 5 mortality

```r
num_combo_df <- sapply(remove_na[,c(3:5, 9, 12, 15:17, 19:22)], as.numeric)
num_combo_df <- as.data.frame((num_combo_df))
for (i in 1:12) {
  print(ggplot(data = num_combo_df, aes(x = num_combo_df[, i])) +
    geom_histogram() + theme_minimal() + xlab(names(num_combo_df[i])))
}
```

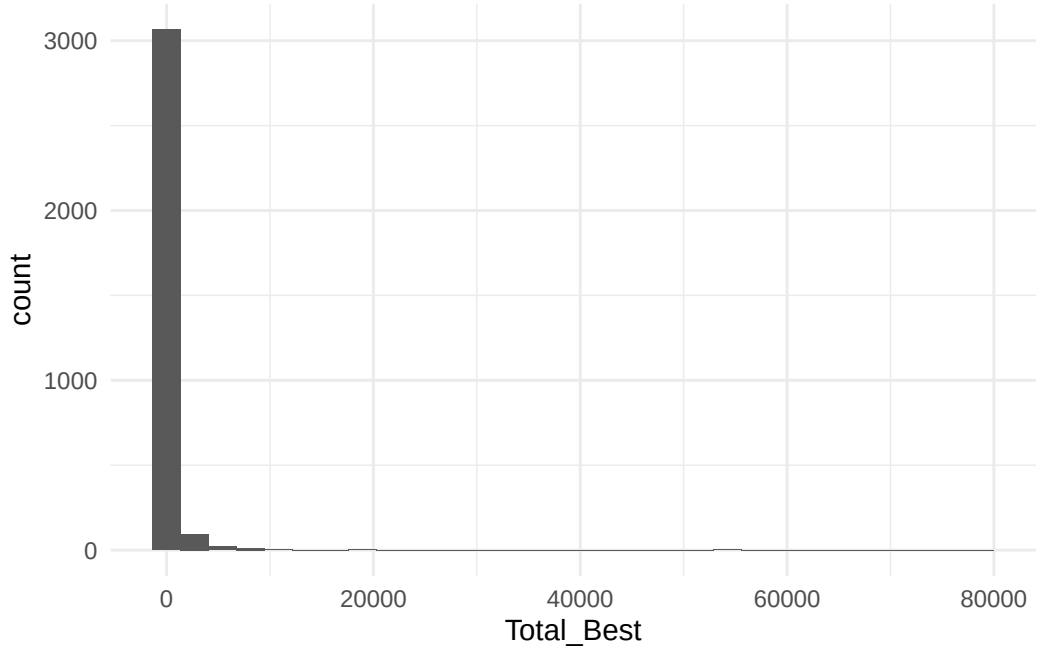
```
`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



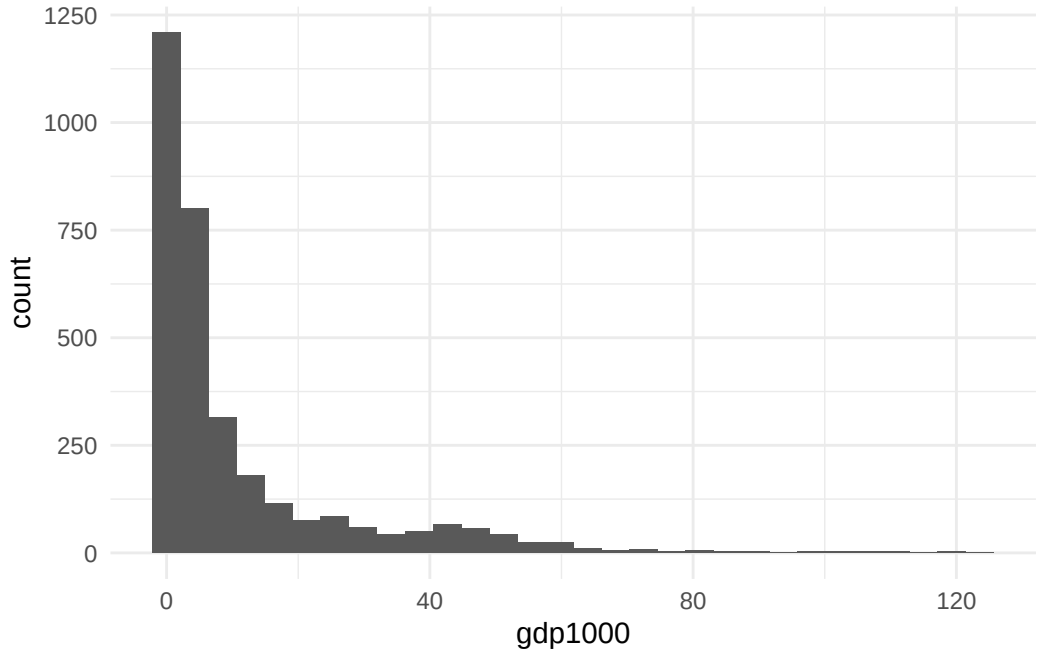

```
`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



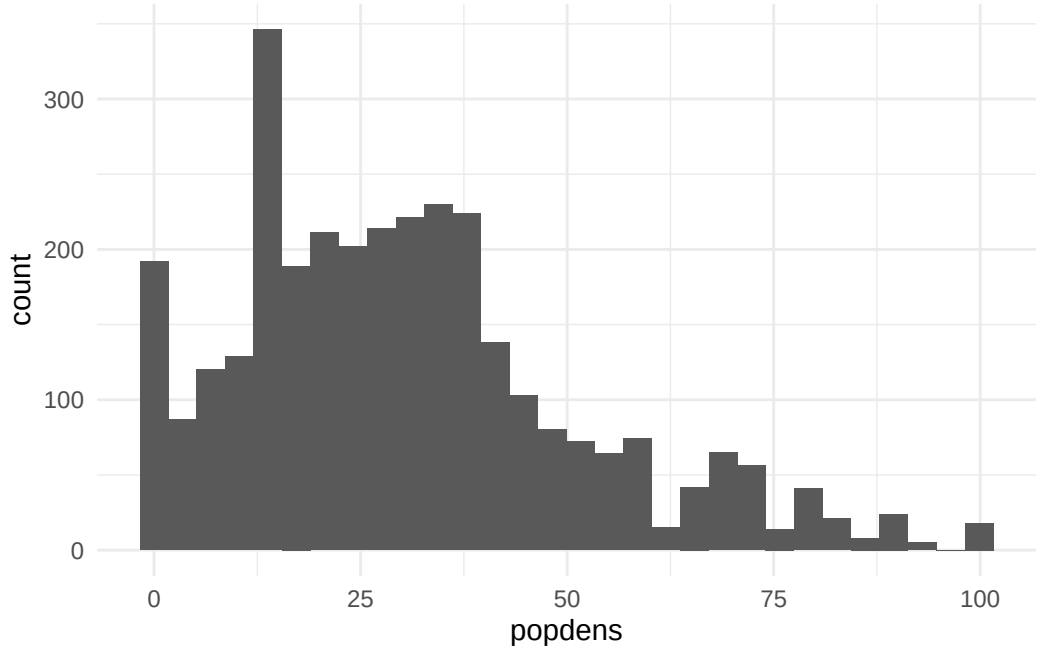

```
`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



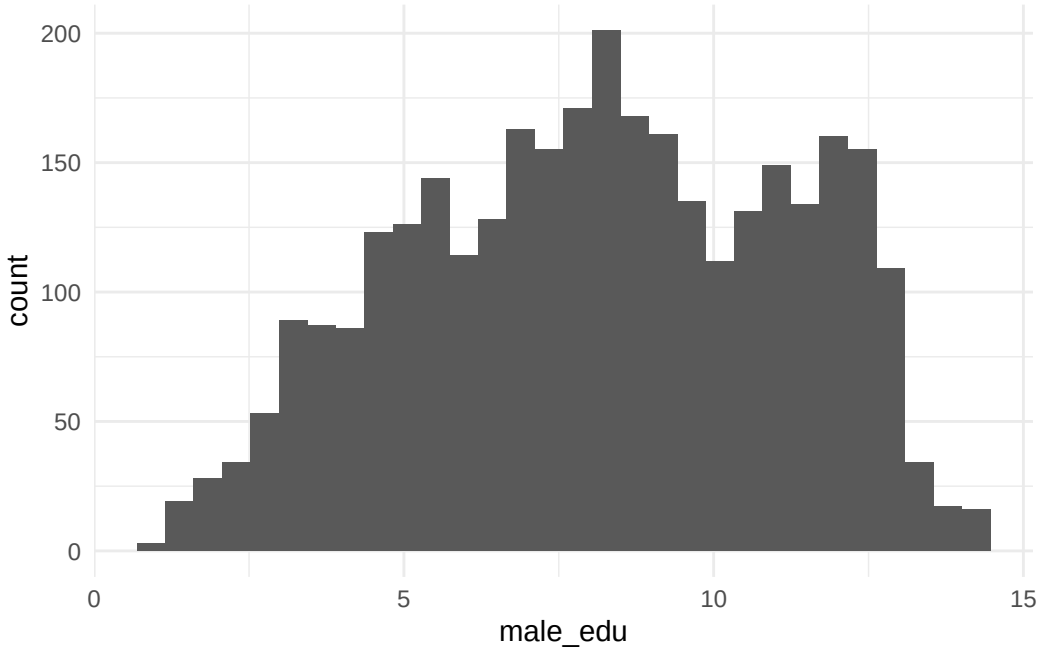

```
`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



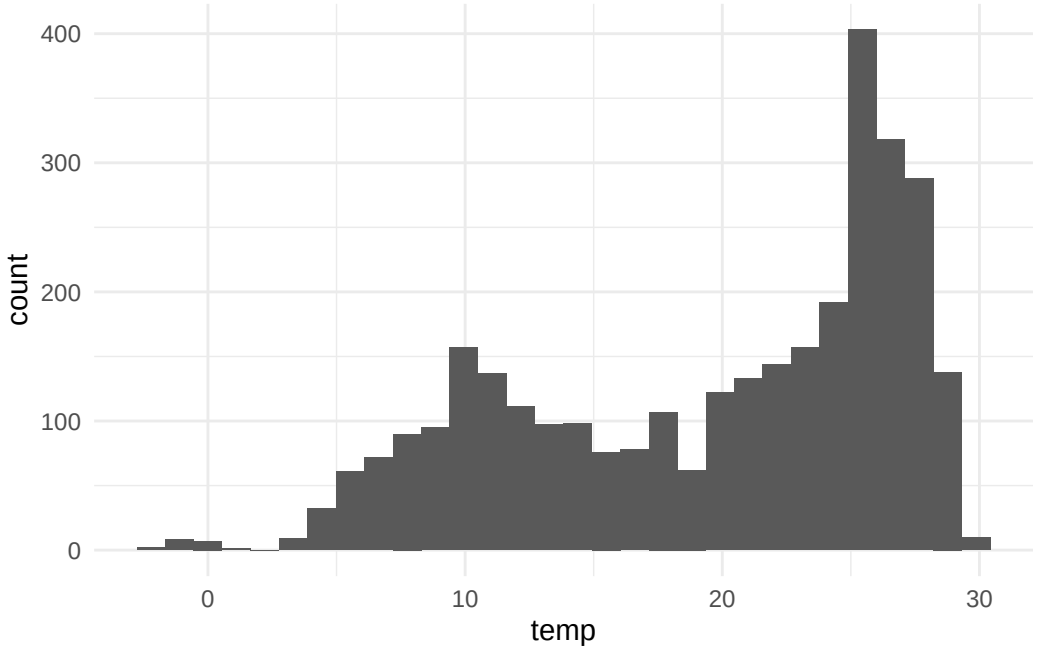

```
`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



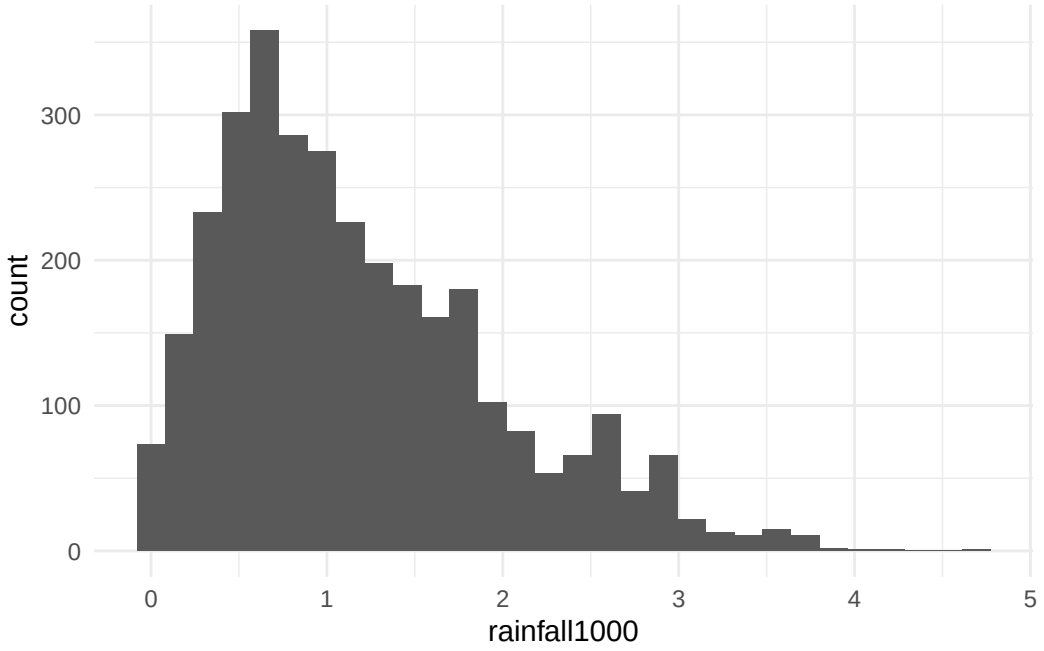

```
`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



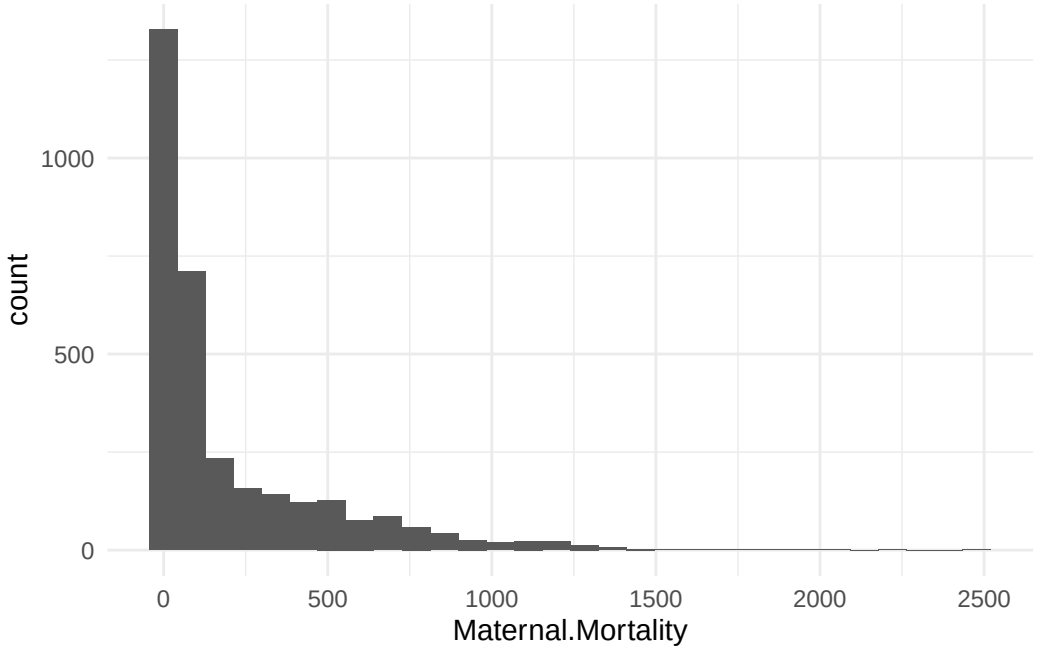

```
`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



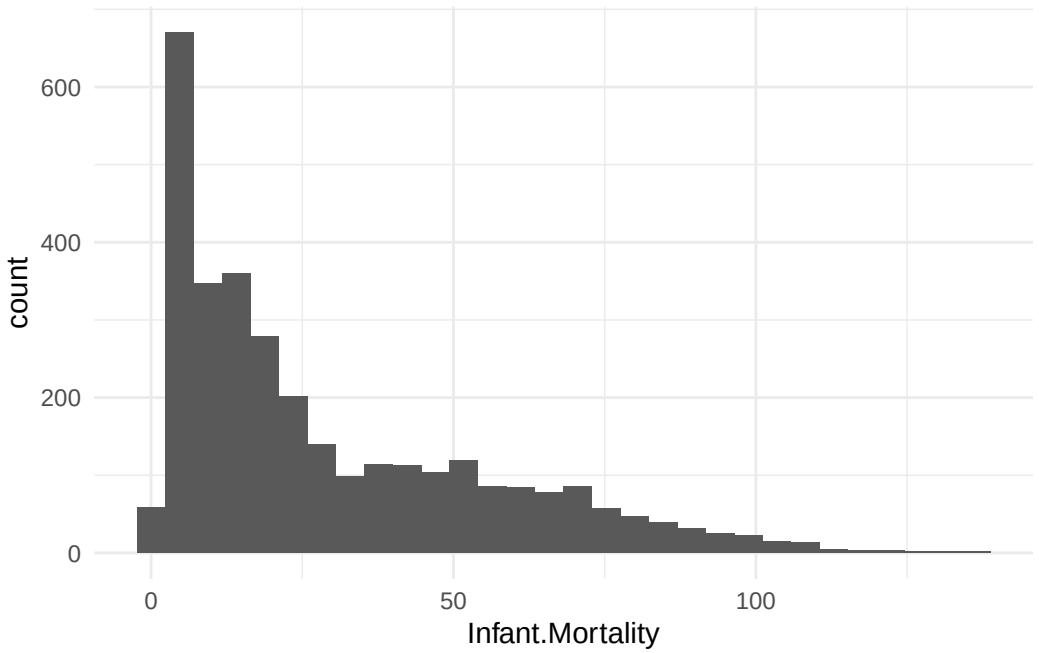

```
`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



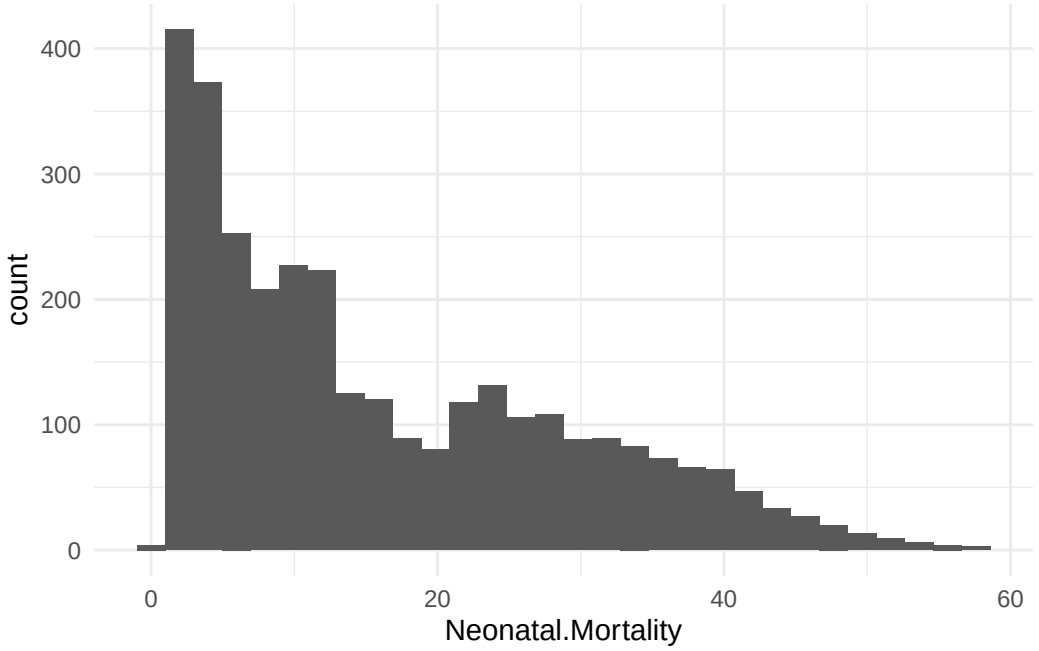

```
`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



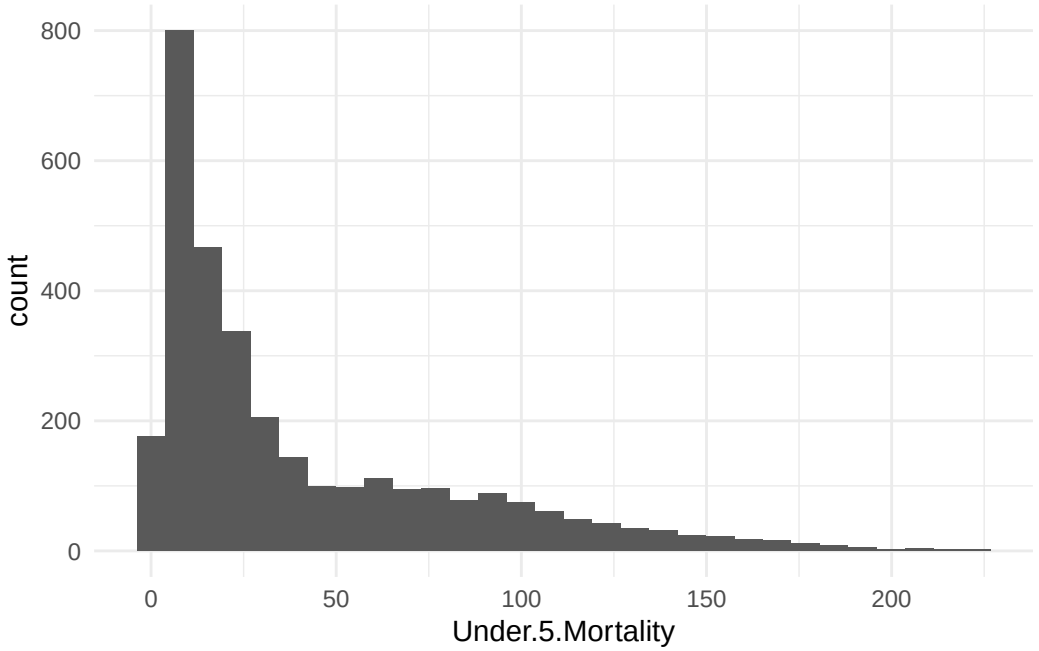

```
`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




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

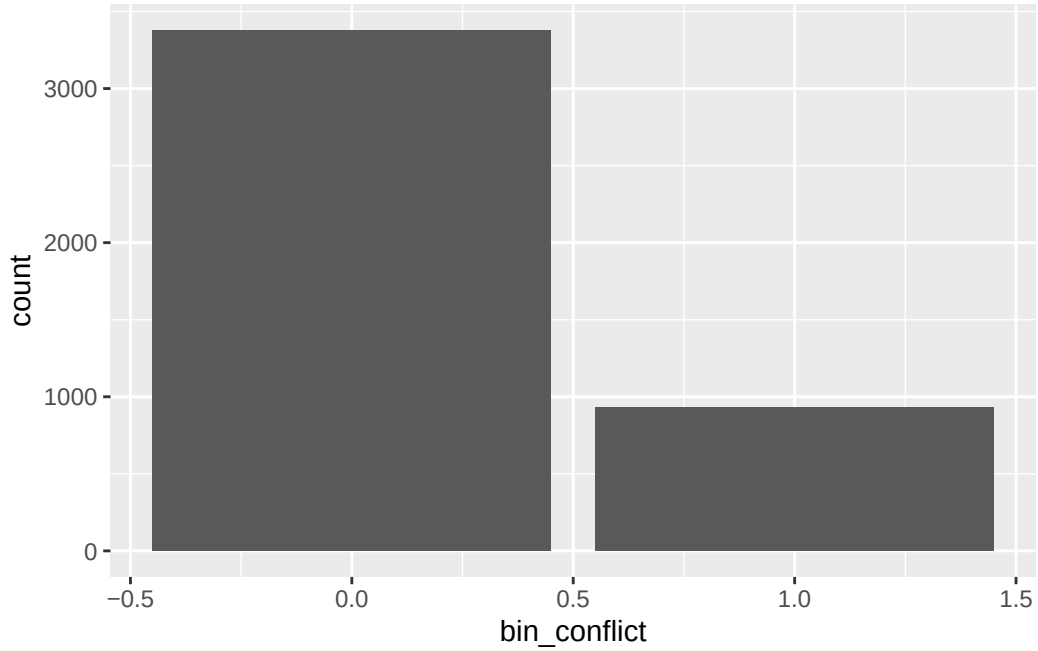



```
`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



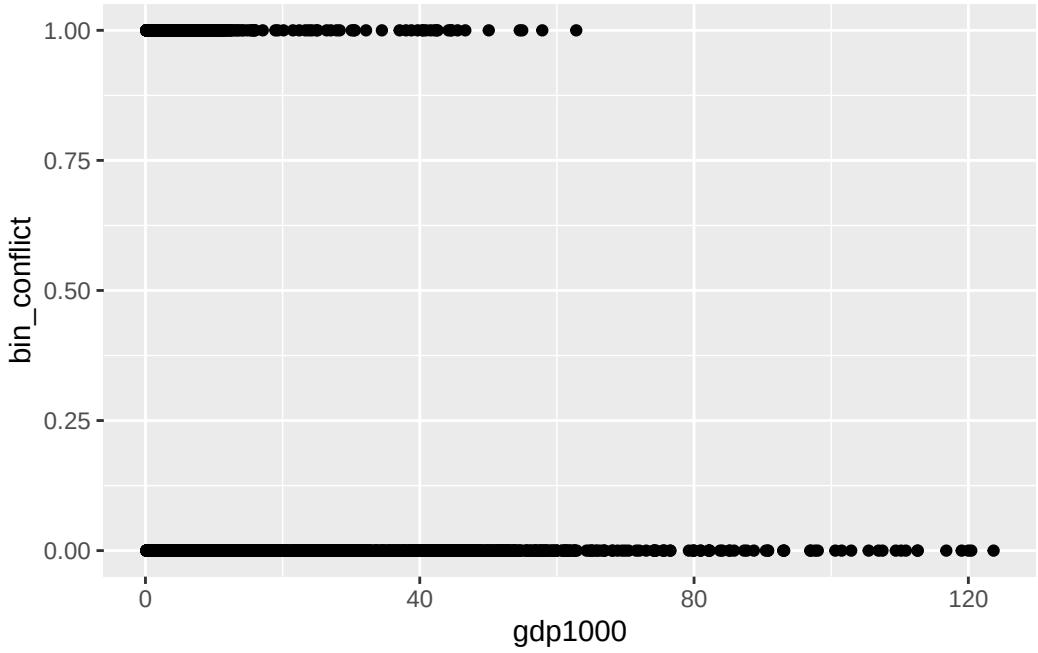
Lets use a bar graft to examine the binary variable of conflict for all countries and years included

```
ggplot(data = combo_df) +
  geom_bar(mapping = aes(x = bin_conflict))
```



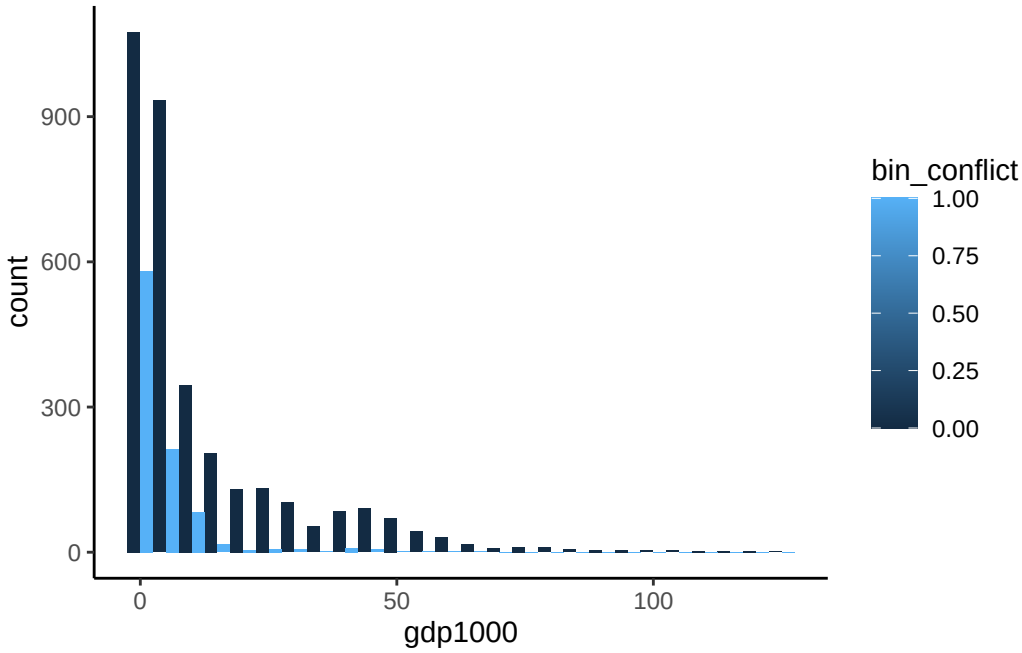
Lets plot gdp100 by conflict using ggplot

```
ggplot(data = mice_est) +
  geom_point(mapping = aes(x = gdp1000, y = bin_conflict))
```



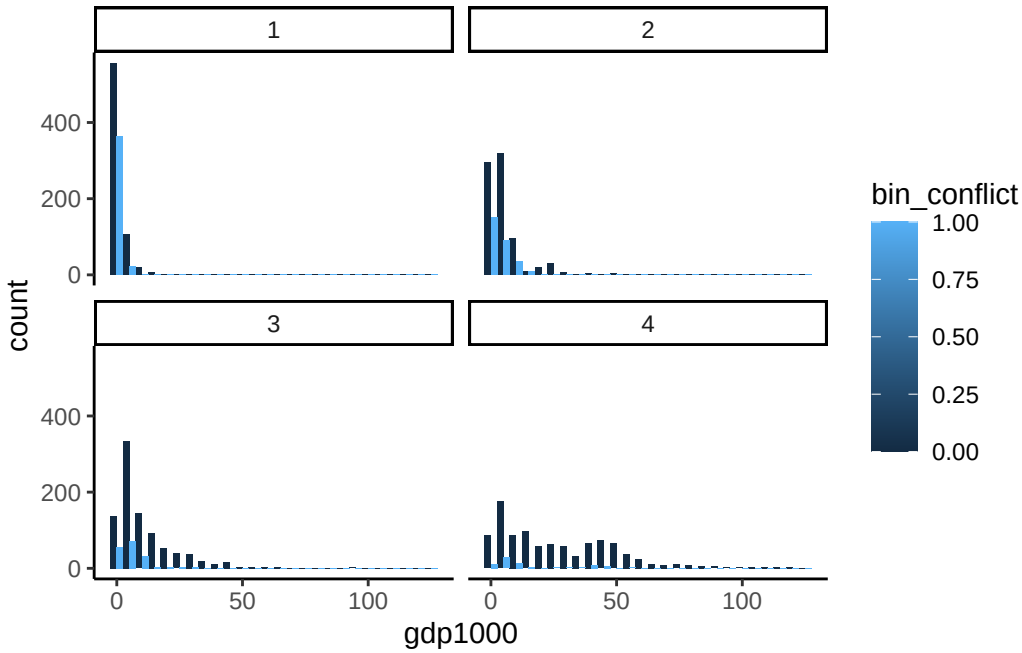
Lets make a comparative histogram to see how bins of gdp compare in those with conflict and those without.

```
ggplot(data = mice_est, aes(x = gdp1000, group = bin_conflict, fill = bin_conflict )) +
 geom_histogram(position='dodge', binwidth = 5) +
 theme_classic()
```



Lets further examine this relationship by creating a new variable which divides data into quartiles based on male education

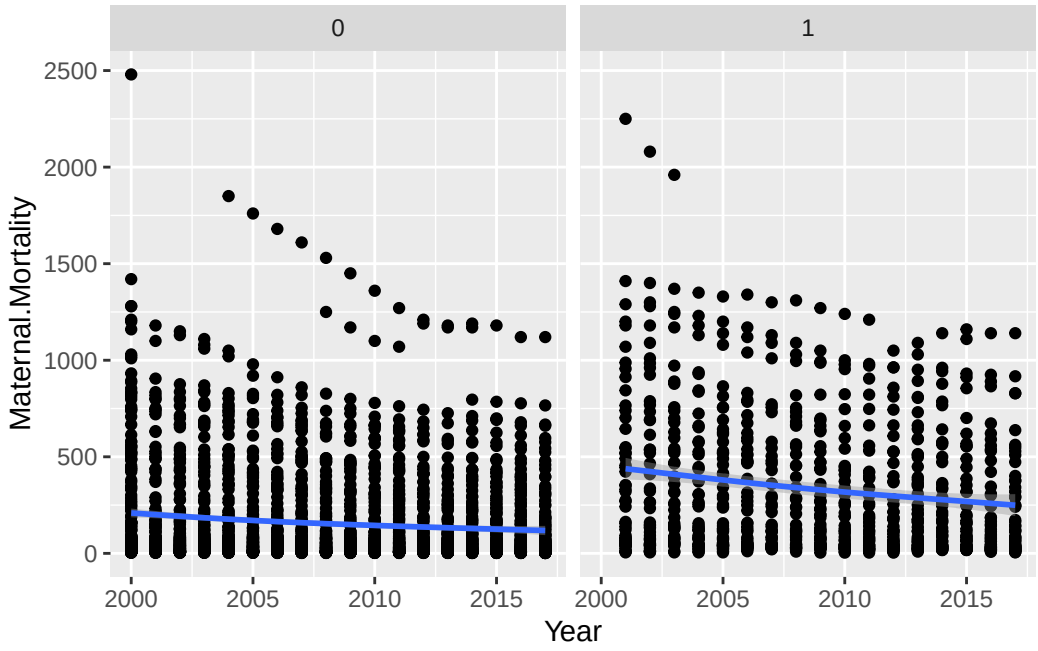
```
new_combo_df <- mice_est
new_combo_df$male_edu_qt <- ntile(new_combo_df$male_edu, 4)
ggplot(data = new_combo_df, aes(x = gdp1000, group = bin_conflict, fill = bin_conflict )) +
  geom_histogram(position='dodge', binwidth = 5) +
  theme_classic() +
  facet_wrap(vars(male_edu_qt))
```



Let us examine the trend in maternal mortality over the years and stratify into separate plots based on conflict

```
ggplot(data = remove_na, aes(x = Year, y = Maternal.Mortality)) +
  geom_point() +
  geom_smooth() +
  facet_wrap(vars(bin_conflict))
```

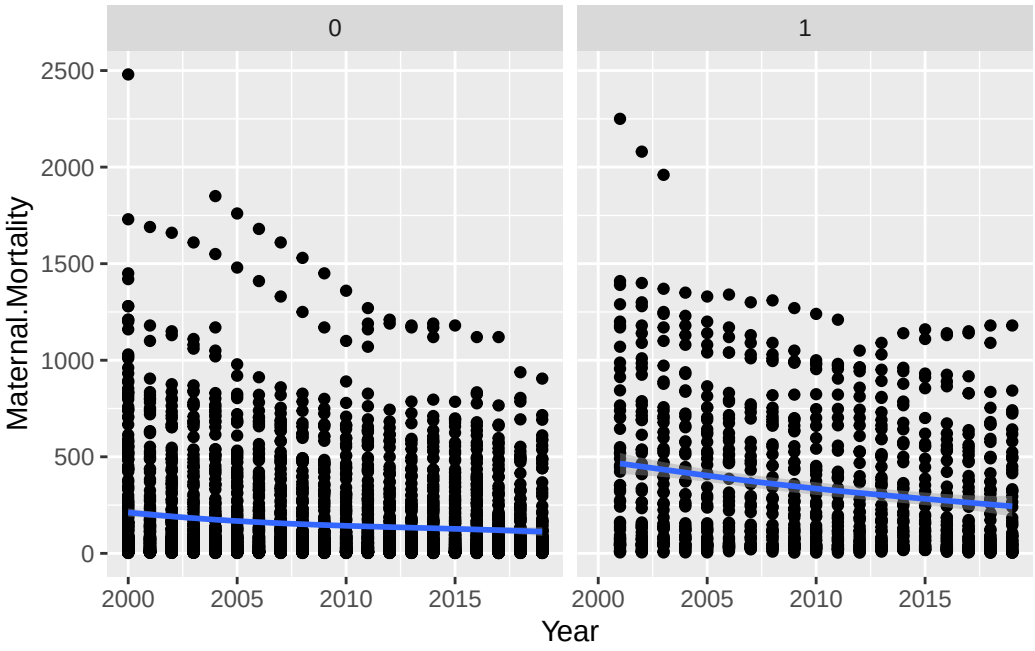
```
`geom_smooth()` using method = 'gam' and formula = 'y ~ s(x, bs = "cs")'
```



Lets do the same plot, but use our dataset with multiply imputed missing variables

```
ggplot(data = mice_est, aes(x = Year, y = Maternal.Mortality)) +
  geom_point() +
  geom_smooth() +
  facet_wrap(vars(bin_conflict))
```

```
`geom_smooth()` using method = 'gam' and formula = 'y ~ s(x, bs = "cs")'
```


0
1
2500
2000
1500
1000
500
0
Maternal.Mortality
2000
2005
2010
2015
Year