Week 4- in class activity_EDA

```
library(here)
```

```
Warning: package 'here' was built under R version 4.2.3
```

```
here() starts at C:/Users/rebec/OneDrive/Old Documents/Version Control/Armed-conflict
```

```
library(tidyverse)
```

```
Warning: package 'tidyverse' was built under R version 4.2.3
```

```
Warning: package 'ggplot2' was built under R version 4.2.3
```

```
Warning: package 'tibble' was built under R version 4.2.3
```

```
Warning: package 'tidyr' was built under R version 4.2.3
```

```
Warning: package 'readr' was built under R version 4.2.3
```

```
Warning: package 'purrr' was built under R version 4.2.3
```

```
Warning: package 'dplyr' was built under R version 4.2.3
```

```
Warning: package 'stringr' was built under R version 4.2.3
```

```
Warning: package 'forcats' was built under R version 4.2.3
```

```
Warning: package 'lubridate' was built under R version 4.2.3
```

```
-- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
v dplyr     1.1.3     v readr     2.1.4
v forcats   1.0.0     v stringr   1.5.1
v ggplot2   3.4.3     v tibble    3.2.1
v lubridate 1.9.3     v tidyr     1.3.0
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
```

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

names(finaldata)
```

```
 [1] "country_name"       "ISO"                "region"
 [4] "Year"               "gdp1000"            "OECD"
 [7] "OECD2023"           "popdens"            "urban"
[10] "agedep"             "male_edu"           "temp"
[13] "rainfall1000"       "totdeath"           "armconf1"
[16] "Maternal_Mortality" "Infant_Mortality"   "Neonatal_Mortality"
[19] "Under5_Mortality"   "drought"            "earthquake"
```

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

```
   country_name ISO           region Year  gdp1000 OECD OECD2023  popdens
1        Canada CAN Northern America 2000 24.27100    1        1 66.19704
2        Canada CAN Northern America 2001 23.82206    1        1 66.45361
3        Canada CAN Northern America 2002 24.25534    1        1 66.71112
4        Canada CAN Northern America 2003 28.30046    1        1 66.96384
5        Canada CAN Northern America 2004 32.14368    1        1 67.21715
6        Canada CAN Northern America 2005 36.38251    1        1 67.47283
7        Canada CAN Northern America 2006 40.50406    1        1 67.73674
8        Canada CAN Northern America 2007 44.65990    1        1 67.99444
9        Canada CAN Northern America 2008 46.71051    1        1 68.25765
10       Canada CAN Northern America 2009 40.87631    1        1 68.53354
11       Canada CAN Northern America 2010 47.56208    1        1 68.80739
12       Canada CAN Northern America 2011 52.22370    1        1 69.04842
13       Canada CAN Northern America 2012 52.66909    1        1 69.27604
```

```
14       Canada CAN Northern America 2013 52.63517    1        1 69.50772
15       Canada CAN Northern America 2014 50.95600    1        1 69.76876
16       Canada CAN Northern America 2015 43.59614    1        1 69.98853
17       Canada CAN Northern America 2016 42.31560    1        1 70.21484
18       Canada CAN Northern America 2017 45.12943    1        1 70.40863
19       Canada CAN Northern America 2018 46.54864    1        1 70.63614
20       Canada CAN Northern America 2019 46.32867    1        1 70.83794
      urban   agedep male_edu     temp rainfall1000 totdeath armconf1
1  56.14335 46.34463 12.30281 5.486244    0.9971559       11        0
2  56.40270 45.89632 12.35258 6.469105    0.8644873       23        0
3  56.67093 45.46660 12.40182 5.979147    0.9460938        1        0
4  56.94365 45.07468 12.45053 5.416964    1.0189234        0        0
5  57.20020 44.67374 12.49870 5.556961    1.0008237        0        0
6  57.41671 44.26641 12.54635 6.187472    1.0367199        0        0
7  57.59143 43.96370 12.59349 6.895084    1.0917386        0        0
8  57.75691 43.83612 12.64015 5.900051    1.0134091        0        0
9  57.97905 43.85426 12.68634 5.650118    1.0693435        0        0
10 58.24228 43.94937 12.73207 5.398867    0.9928497        0        0
11 58.52809 44.13587 12.77735 6.781766    1.0379754        0        0
12 58.81437 44.53578 12.82218 6.269133    1.1343442        0        0
13 59.05573 45.18393 12.86660 7.249497    0.9747708        0        0
14 59.19713 45.95404 12.91059 5.954381    1.0282075        0        0
15 59.30361 46.75493 12.95414 5.584650    1.0377695        0        0
16 59.42627 47.59164 12.99723 6.436884    0.9632446        0        0
17 59.50521 48.41410 13.03988 7.184514    0.9677826        0        0
18 59.59325 49.14806 13.08210 6.539669    1.0995322        0        0
19 59.68433 49.80166 13.12388 6.539677    1.0991469        0        0
20 59.75984 50.47739 13.16522 6.539633    1.0987523        0        0
   Maternal_Mortality Infant_Mortality Neonatal_Mortality Under5_Mortality
1                   9              5.3                3.8              6.2
2                  10              5.3                3.8              6.2
3                  10              5.3                3.9              6.2
4                  10              5.3                3.9              6.2
5                  10              5.3                3.9              6.1
6                  11              5.2                3.9              6.1
7                  11              5.2                3.9              6.0
8                  11              5.1                3.8              6.0
9                  12              5.1                3.8              5.9
10                 12              5.0                3.8              5.8
11                 11              5.0                3.8              5.7
12                 11              4.9                3.7              5.7
13                 11              4.9                3.7              5.6
14                 11              4.8                3.6              5.5
```

```
15                11              4.7               3.6              5.4
16                11              4.7               3.6              5.4
17                10              4.6               3.5              5.3
18                10              4.6               3.4              5.2
19                NA              4.5               3.3              5.1
20                NA              4.4               3.3              5.1
   drought earthquake
1        0          0
2        0          0
3        0          0
4        0          0
5        0          0
6        0          0
7        0          0
8        0          0
9        0          0
10       0          0
11       0          0
12       0          0
13       0          0
14       0          0
15       0          0
16       0          0
17       0          0
18       0          0
19       0          0
20       0          0
```

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

```
   country_name ISO                          region Year  gdp1000 OECD OECD2023
1       Ecuador ECU Latin America and the Caribbean 2000 1.451531    0        0
2       Ecuador ECU Latin America and the Caribbean 2001 1.904814    0        0
3       Ecuador ECU Latin America and the Caribbean 2002 2.184209    0        0
4       Ecuador ECU Latin America and the Caribbean 2003 2.438344    0        0
5       Ecuador ECU Latin America and the Caribbean 2004 2.703566    0        0
6       Ecuador ECU Latin America and the Caribbean 2005 3.014310    0        0
7       Ecuador ECU Latin America and the Caribbean 2006 3.340841    0        0
8       Ecuador ECU Latin America and the Caribbean 2007 3.579032    0        0
9       Ecuador ECU Latin America and the Caribbean 2008 4.260433    0        0
10      Ecuador ECU Latin America and the Caribbean 2009 4.240703    0        0
```

11	Ecuador	ECU	Latin America and the Caribbean	2010	4.640246	0	0	
12	Ecuador	ECU	Latin America and the Caribbean	2011	5.202656	0	0	
13	Ecuador	ECU	Latin America and the Caribbean	2012	5.678456	0	0	
14	Ecuador	ECU	Latin America and the Caribbean	2013	6.050355	0	0	
15	Ecuador	ECU	Latin America and the Caribbean	2014	6.374631	0	0	
16	Ecuador	ECU	Latin America and the Caribbean	2015	6.130587	0	0	
17	Ecuador	ECU	Latin America and the Caribbean	2016	6.079089	0	0	
18	Ecuador	ECU	Latin America and the Caribbean	2017	6.246404	0	0	
19	Ecuador	ECU	Latin America and the Caribbean	2018	6.321349	0	0	
20	Ecuador	ECU	Latin America and the Caribbean	2019	6.233258	0	0	

	popdens	urban	agedep	male_edu	temp	rainfall1000	totdeath	armconf1
1	23.27432	36.19963	67.44216	7.738627	19.54855	1.4201653	0	0
2	23.39372	36.67994	66.57356	7.843942	19.66622	1.1667746	0	0
3	23.52087	37.08903	65.65488	7.949449	20.24695	1.4577981	2	0
4	23.58358	37.23792	64.71472	8.055240	20.05016	1.5781807	0	0
5	38.43743	37.39268	63.78049	8.161433	20.10136	1.0683450	26	1
6	38.55361	37.36968	62.86530	8.268176	19.88163	0.8555447	0	0
7	38.65018	37.47567	61.97042	8.375587	20.07087	1.1114502	0	0
8	38.76505	37.68172	61.11422	8.483729	19.49536	1.0899082	0	0
9	38.83977	37.67445	60.31015	8.592603	19.85711	1.6184816	0	0
10	38.92613	37.39437	59.55262	8.702180	20.39298	1.0870796	25	1
11	39.03066	37.26838	58.83793	8.812409	20.11160	1.7045703	0	0
12	39.09586	37.61553	58.16553	8.923172	19.86633	1.4518388	0	0
13	39.13343	38.00733	57.51051	9.034284	20.19000	1.7520003	0	0
14	39.18619	38.22511	56.84804	9.145523	19.85177	1.3735605	0	0
15	39.27871	38.12421	56.17001	9.256679	20.42252	1.2572257	0	0
16	39.38824	38.15633	55.46511	9.367582	20.95595	1.7284273	0	0
17	39.46201	38.45745	54.73369	9.478071	20.77476	1.3168761	0	0
18	39.53609	38.65993	53.99096	9.587993	20.53262	1.9544485	0	0
19	39.58380	38.87253	53.12249	9.697221	20.53714	1.9573265	0	0
20	39.75109	39.05144	52.29278	9.805670	20.54169	1.9602443	0	0

	Maternal_Mortality	Infant_Mortality	Neonatal_Mortality	Under5_Mortality
1	122	24.7	14.1	29.5
2	117	23.4	13.4	28.0
3	110	22.4	12.7	26.6
4	100	21.5	12.1	25.4
5	94	20.7	11.6	24.4
6	94	19.9	11.1	23.5
7	90	19.2	10.6	22.6
8	85	18.5	10.2	21.7
9	82	17.7	9.7	20.8
10	80	17.0	9.3	19.9
11	78	16.3	8.9	19.0

```
12                76              15.6                8.5              18.1
13                71              14.9                8.1              17.3
14                67              14.3                7.8              16.6
15                65              13.7                7.5              15.9
16                63              13.2                7.3              15.4
17                61              12.8                7.1              14.8
18                59              12.4                6.9              14.4
19                NA              12.0                6.9              13.9
20                NA              11.6                6.8              13.4
   drought earthquake
1        0          0
2        0          0
3        0          0
4        0          0
5        0          0
6        0          0
7        0          0
8        0          0
9        0          0
10       1          0
11       0          0
12       0          0
13       0          0
14       1          0
15       0          1
16       0          0
17       0          1
18       0          0
19       0          0
20       0          1
```

```
####### EXPLORATORY DATA ANALYSIS #######


finaldata |>
  head()
```

```
  country_name ISO        region Year   gdp1000 OECD OECD2023  popdens    urban
1  Afghanistan AFG Southern Asia 2000        NA    0        0 14.13654 16.25324
2  Afghanistan AFG Southern Asia 2001        NA    0        0 14.23156 16.25661
3  Afghanistan AFG Southern Asia 2002 0.1835328    0        0 14.32270 16.42654
4  Afghanistan AFG Southern Asia 2003 0.2004626    0        0 14.40691 16.60701
```

```
5  Afghanistan AFG Southern Asia 2004 0.2216576    0        0 15.21947 16.71367
6  Afghanistan AFG Southern Asia 2005 0.2550551    0        0 15.33619 16.85096
    agedep male_edu     temp rainfall1000 totdeath armconf1 Maternal_Mortality
1 108.3466 2.762086 12.69959    0.2763704     5065        1               1450
2 108.9899 2.856936 12.85570    0.2793079     5394        1               1390
3 109.3472 2.954241 12.71081    0.3805710     5553        1               1300
4 109.4475 3.054121 12.16592    0.4288939     1157        1               1240
5 109.2868 3.156706 13.04643    0.3754336      944        1               1180
6 107.9646 3.262133 12.23141    0.4415680      817        1               1140
  Infant_Mortality Neonatal_Mortality Under5_Mortality drought earthquake
1             90.5               60.9            129.2       1          0
2             87.9               59.7            125.2       0          1
3             85.3               58.5            121.1       0          1
4             82.7               57.2            116.9       0          1
5             80.0               55.9            112.6       0          1
6             77.3               54.6            108.4       0          1
```

```
finaldata |>
  head()
```

```
  country_name ISO        region Year   gdp1000 OECD OECD2023  popdens    urban
1  Afghanistan AFG Southern Asia 2000        NA    0        0 14.13654 16.25324
2  Afghanistan AFG Southern Asia 2001        NA    0        0 14.23156 16.25661
3  Afghanistan AFG Southern Asia 2002 0.1835328    0        0 14.32270 16.42654
4  Afghanistan AFG Southern Asia 2003 0.2004626    0        0 14.40691 16.60701
5  Afghanistan AFG Southern Asia 2004 0.2216576    0        0 15.21947 16.71367
6  Afghanistan AFG Southern Asia 2005 0.2550551    0        0 15.33619 16.85096
    agedep male_edu     temp rainfall1000 totdeath armconf1 Maternal_Mortality
1 108.3466 2.762086 12.69959    0.2763704     5065        1               1450
2 108.9899 2.856936 12.85570    0.2793079     5394        1               1390
3 109.3472 2.954241 12.71081    0.3805710     5553        1               1300
4 109.4475 3.054121 12.16592    0.4288939     1157        1               1240
5 109.2868 3.156706 13.04643    0.3754336      944        1               1180
6 107.9646 3.262133 12.23141    0.4415680      817        1               1140
  Infant_Mortality Neonatal_Mortality Under5_Mortality drought earthquake
1             90.5               60.9            129.2       1          0
2             87.9               59.7            125.2       0          1
3             85.3               58.5            121.1       0          1
4             82.7               57.2            116.9       0          1
5             80.0               55.9            112.6       0          1
6             77.3               54.6            108.4       0          1
```

```r
finaldata |>
  slice_sample(n = 10)
```

```
     country_name ISO                          region Year    gdp1000 OECD
1           China CHN                    Eastern Asia 2002  1.1485083    0
2      Uzbekistan UZB                    Central Asia 2001  0.4567063    0
3      Kazakhstan KAZ                    Central Asia 2018  9.8126264    0
4         Ukraine UKR                  Eastern Europe 2015  2.1246624    0
5            Fiji FJI                       Melanesia 2000  2.0158812    0
6  Czech Republic CZE                  Eastern Europe 2007 18.4665479    1
7      Costa Rica CRI Latin America and the Caribbean 2010  8.1472440    0
8         Liberia LBR              Sub-Saharan Africa 2005  0.2905412    0
9         Algeria DZA                 Northern Africa 2008  4.9465640    0
10      Argentina ARG Latin America and the Caribbean 2006  5.8909780    0
   OECD2023   popdens     urban   agedep  male_edu      temp rainfall1000
1         0 38.164130 38.520615 43.28138  7.329050 15.171960    1.0523652
2         0  8.552784 16.280515 68.62683 10.714807 13.826433    0.2616566
3         0 13.883015 27.138575 56.23014 10.906643  7.962870    0.3163183
4         0 44.611764 45.405937 44.74021 12.788181 10.433425    0.5011876
5         0 34.881073  4.082326 60.95684  8.285971 24.651346    4.7108078
6         1 27.105318 31.569711 40.36846 11.901235  9.549768    0.7374430
7         0 44.682131 54.005884 47.41980  8.675282 23.848313    3.8399109
8         0 27.228367 31.247208 86.00699  5.497026 26.176176    2.5750404
9         0 29.978571 38.610737 48.40744  5.663582 17.325458    0.4228914
10        0 36.119863 56.054092 58.37503  9.266019 17.449441    0.9319829
   totdeath armconf1 Maternal_Mortality Infant_Mortality Neonatal_Mortality
1         0        0                 51             25.6               18.4
2       297        1                 41             48.4               26.9
3         0        0                 NA              9.2                4.7
4      4528        1                 21              8.1                5.6
5         0        0                 51             19.1               10.1
6         0        0                  5              3.2                2.0
7         0        0                 32              9.2                6.6
8         0        0                816             90.7               38.4
9       495        1                117             25.3               18.9
10        0        0                 57             14.4                8.7
   Under5_Mortality drought earthquake
1              31.5       1          1
2              57.4       0          0
3              10.3       0          0
4               9.5       0          0
5              22.4       0          0
```

```
6               3.9       0          0
7              10.5       0          0
8             129.6       0          0
9              29.5       0          0
10             16.1       0          0
```

```
# Check the structure of the data #
str(finaldata)
```

```
'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 ...
 $ totdeath          : int  5065 5394 5553 1157 944 817 1711 4982 7020 5660 ...
 $ armconf1          : int  1 1 1 1 1 1 1 1 1 1 ...
 $ Maternal_Mortality: int  1450 1390 1300 1240 1180 1140 1120 1090 1030 993 ...
 $ Infant_Mortality  : num  90.5 87.9 85.3 82.7 80 77.3 74.6 71.9 69.2 66.7 ...
 $ Neonatal_Mortality: num  60.9 59.7 58.5 57.2 55.9 54.6 53.2 51.7 50.3 48.9 ...
 $ Under5_Mortality  : num  129 125 121 117 113 ...
 $ drought           : int  1 0 0 0 0 0 1 0 1 0 ...
 $ earthquake        : int  0 1 1 1 1 1 1 0 0 1 ...
```

```
# Get summary statistics for the data #
summary(finaldata)
```

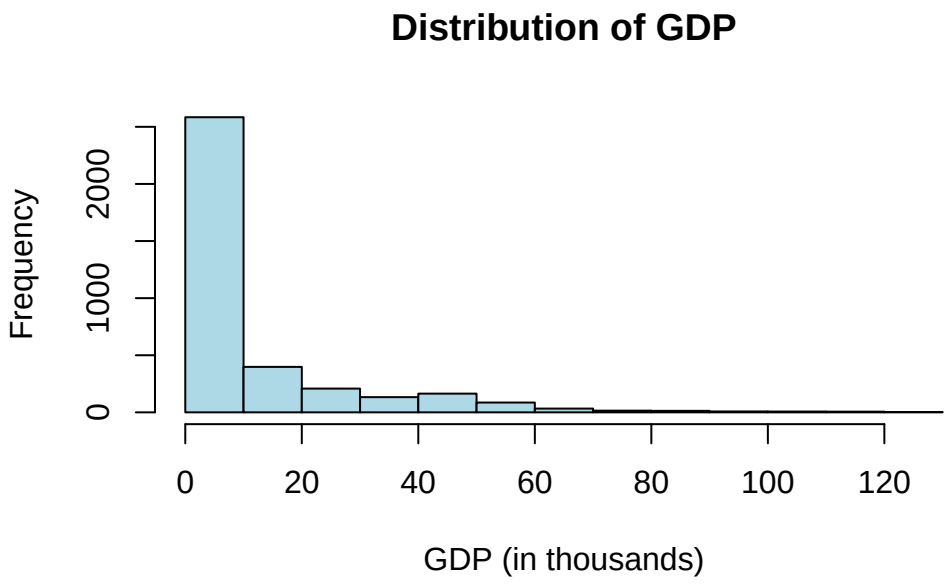
```
 country_name           ISO               region               Year
 Length:3720        Length:3720        Length:3720        Min.   :2000
 Class :character   Class :character   Class :character   1st Qu.:2005
 Mode  :character   Mode  :character   Mode  :character   Median :2010
                                                          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        totdeath          armconf1      Maternal_Mortality
Min.   :0.01993   Min.   :    0.0   Min.   :0.0000   Min.   :   2.0
1st Qu.:0.59146   1st Qu.:    0.0   1st Qu.:0.0000   1st Qu.:  17.0
Median :1.01288   Median :    0.0   Median :0.0000   Median :  66.0
Mean   :1.20216   Mean   :  361.1   Mean   :0.1892   Mean   : 210.6
3rd Qu.:1.68706   3rd Qu.:    2.0   3rd Qu.:0.0000   3rd Qu.: 299.8
Max.   :4.71081   Max.   :78644.0   Max.   :1.0000   Max.   :2480.0
NA's   :20                                           NA's   :426
Infant_Mortality Neonatal_Mortality Under5_Mortality    drought
Min.   :  1.60   Min.   : 0.80      Min.   :  2.00   Min.   :0.00000
1st Qu.:  7.60   1st Qu.: 4.90      1st Qu.:  9.00   1st Qu.:0.00000
Median : 18.90   Median :12.10      Median : 22.20   Median :0.00000
Mean   : 28.90   Mean   :16.18      Mean   : 40.50   Mean   :0.08737
3rd Qu.: 44.52   3rd Qu.:25.32      3rd Qu.: 61.33   3rd Qu.:0.00000
Max.   :138.10   Max.   :60.90      Max.   :224.90   Max.   :1.00000
NA's   :20       NA's   :20         NA's   :20
  earthquake
Min.   :0.00000
1st Qu.:0.00000
Median :0.00000
Mean   :0.08333
3rd Qu.:0.00000
Max.   :1.00000
```

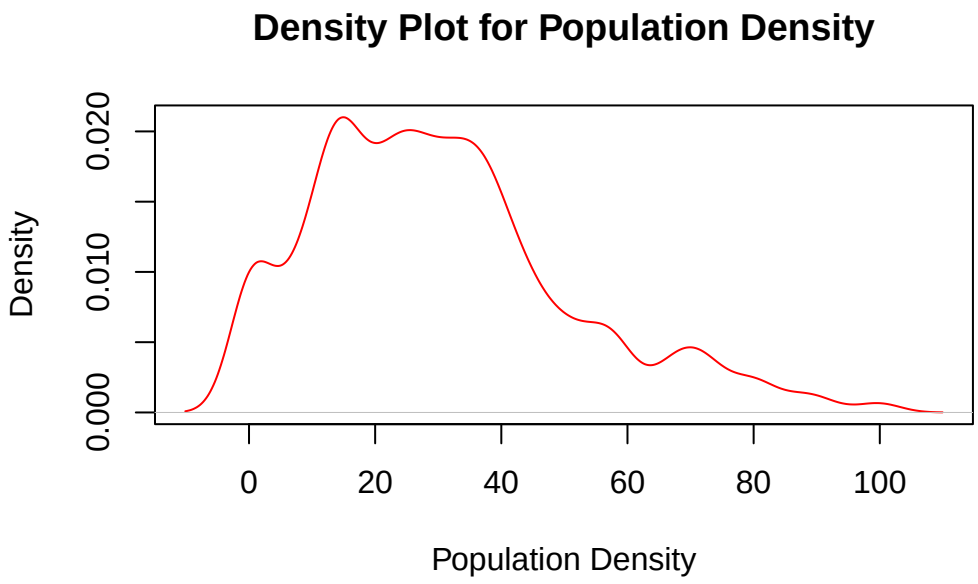
```r
# 20 missing obs for popdens, urban, male_edu, temp, rainfall1000, infant_mortality
# neonatal_mortality, under5_mortatlity

# 426 missing obs for maternal_mortality and 62 missing obs for gdp1000

# So, we have data that is MNAR

missing_data <- finaldata[!complete.cases(finaldata), ]


# View countries with the 20 missing obs

missing_countries <- finaldata[is.na(finaldata$rainfall) |
                           is.na(finaldata$male_edu) |
                           is.na(finaldata$popdens) |
                           is.na(finaldata$urban) |
                           is.na(finaldata$temp) |
                           is.na(finaldata$infant_mortality) |
                           is.na(finaldata$neonatal_mortality) |
                           is.na(finaldata$under5_mortality), ]

# View the countries with missing data
missing_countries$country_name
```

```
character(0)
```

```r
# All 20 missings obs for popdens, urban, male_edu, temp and rainfall1000 are from
# Cote d'Ivoire.

# All 20 missing obs for infant, neonatal and under5 mortality are from Puerto Rico

# No obvious pattern for the missing obs for maternal mortality

# All obs for gdp1000 are missing for South Sudan, North Korea, Eritrea, Venezuela. For Soma


### EXPLORE DISTRIBUTION AND PROPERTIES OF VARIABLES #####


# Histogram for GDP
hist(finaldata$gdp1000, main = "Distribution of GDP", xlab = "GDP (in thousands)", col = "lig
```




```
## Most values for GDP1000 fall between 0-10 (GDP in thousands) ##


# Density plot for Population Density
plot(density(finaldata$popdens, na.rm = TRUE), main = "Density Plot for Population Density",
```




```
## Popdens values between 5 and 45 per 100,000 are more likely to occur or occur more freque


# Create a contingency table of OECD membership by country
table(finaldata$country_name, finaldata$OECD)
```

```
                          0  1
  Afghanistan            20  0
  Albania                20  0
  Algeria                20  0
  Andorra                20  0
  Angola                 20  0
  Antigua and Barbuda    20  0
  Argentina              20  0
  Armenia                20  0
  Australia               0 20
  Austria                 0 20
  Azerbaijan             20  0
  Bahrain                20  0
  Bangladesh             20  0
  Barbados               20  0
```

Belarus	20	0
Belgium	0	20
Belize	20	0
Benin	20	0
Bhutan	20	0
Bolivia	20	0
Bosnia and Herzegovina	20	0
Botswana	20	0
Brazil	20	0
Brunei	20	0
Bulgaria	20	0
Burkina Faso	20	0
Burundi	20	0
Cambodia	20	0
Cameroon	20	0
Canada	0	20
Cape Verde	20	0
Central African Republic	20	0
Chad	20	0
Chile	10	10
China	20	0
Colombia	20	0
Comoros	20	0
Congo	20	0
Costa Rica	20	0
Cote d'Ivoire	20	0
Croatia	20	0
Cuba	20	0
Cyprus	20	0
Czech Republic	0	20
Democratic Republic of the Congo	20	0
Denmark	0	20
Djibouti	20	0
Dominica	20	0
Dominican Republic	20	0
Ecuador	20	0
Egypt	20	0
El Salvador	20	0
Equatorial Guinea	20	0
Eritrea	20	0
Estonia	10	10
Ethiopia	20	0
Federated States of Micronesia	20	0

Fiji	20	0
Finland	0	20
France	0	20
Gabon	20	0
Georgia	20	0
Germany	0	20
Ghana	20	0
Greece	0	20
Grenada	20	0
Guatemala	20	0
Guinea	20	0
Guinea-Bissau	20	0
Guyana	20	0
Haiti	20	0
Honduras	20	0
Hungary	0	20
Iceland	0	20
India	20	0
Indonesia	20	0
Iran	20	0
Iraq	20	0
Ireland	0	20
Italy	0	20
Jamaica	20	0
Japan	0	20
Jordan	20	0
Kazakhstan	20	0
Kenya	20	0
Kiribati	20	0
Kuwait	20	0
Kyrgyzstan	20	0
Laos	20	0
Latvia	16	4
Lebanon	20	0
Lesotho	20	0
Liberia	20	0
Libya	20	0
Lithuania	18	2
Luxembourg	0	20
Macedonia	20	0
Madagascar	20	0
Malawi	20	0
Malaysia	20	0

Maldives	20	0
Mali	20	0
Malta	20	0
Marshall Islands	20	0
Mauritania	20	0
Mauritius	20	0
Mexico	0	20
Moldova	20	0
Mongolia	20	0
Montenegro	20	0
Morocco	20	0
Mozambique	20	0
Myanmar	20	0
Namibia	20	0
Nepal	20	0
Netherlands	0	20
New Zealand	0	20
Nicaragua	20	0
Niger	20	0
Nigeria	20	0
North Korea	20	0
Norway	0	20
Oman	20	0
Pakistan	20	0
Panama	20	0
Papua New Guinea	20	0
Paraguay	20	0
Peru	20	0
Philippines	20	0
Poland	0	20
Portugal	0	20
Puerto Rico	20	0
Qatar	20	0
Romania	20	0
Russian Federation	20	0
Rwanda	20	0
Saint Lucia	20	0
Saint Vincent and the Grenadines	20	0
Samoa	20	0
Sao Tome and Principe	20	0
Saudi Arabia	20	0
Senegal	20	0
Serbia	20	0

Seychelles	20	0
Sierra Leone	20	0
Singapore	20	0
Slovakia	0	20
Slovenia	10	10
Solomon Islands	20	0
Somalia	20	0
South Africa	20	0
South Korea	0	20
South Sudan	20	0
Spain	0	20
Sri Lanka	20	0
Sudan	20	0
Suriname	20	0
Swaziland	20	0
Sweden	0	20
Switzerland	0	20
Syria	20	0
Tajikistan	20	0
Tanzania	20	0
Thailand	20	0
The Bahamas	20	0
The Gambia	20	0
Timor-Leste	20	0
Togo	20	0
Tonga	20	0
Trinidad and Tobago	20	0
Tunisia	20	0
Turkey	0	20
Turkmenistan	20	0
Uganda	20	0
Ukraine	20	0
United Arab Emirates	20	0
United Kingdom	0	20
United States	0	20
Uruguay	20	0
Uzbekistan	20	0
Vanuatu	20	0
Venezuela	20	0
Vietnam	20	0
Yemen	20	0
Zambia	20	0
Zimbabwe	20	0

```r
# Only Lithuania, Latvia, Estonia and Chile changed OECD membership status through the obser

# Filter the data for Latvia, Lithuania, and Chile
subset_country <- subset(finaldata, country_name %in% c("Latvia", "Lithuania", "Chile"))

table(subset_country$country_name, subset_country$OECD, subset_country$Year)
```

```
, ,  = 2000


           0 1
  Chile    1 0
  Latvia   1 0
  Lithuania 1 0

, ,  = 2001


           0 1
  Chile    1 0
  Latvia   1 0
  Lithuania 1 0

, ,  = 2002


           0 1
  Chile    1 0
  Latvia   1 0
  Lithuania 1 0

, ,  = 2003


           0 1
  Chile    1 0
  Latvia   1 0
  Lithuania 1 0

, ,  = 2004
```

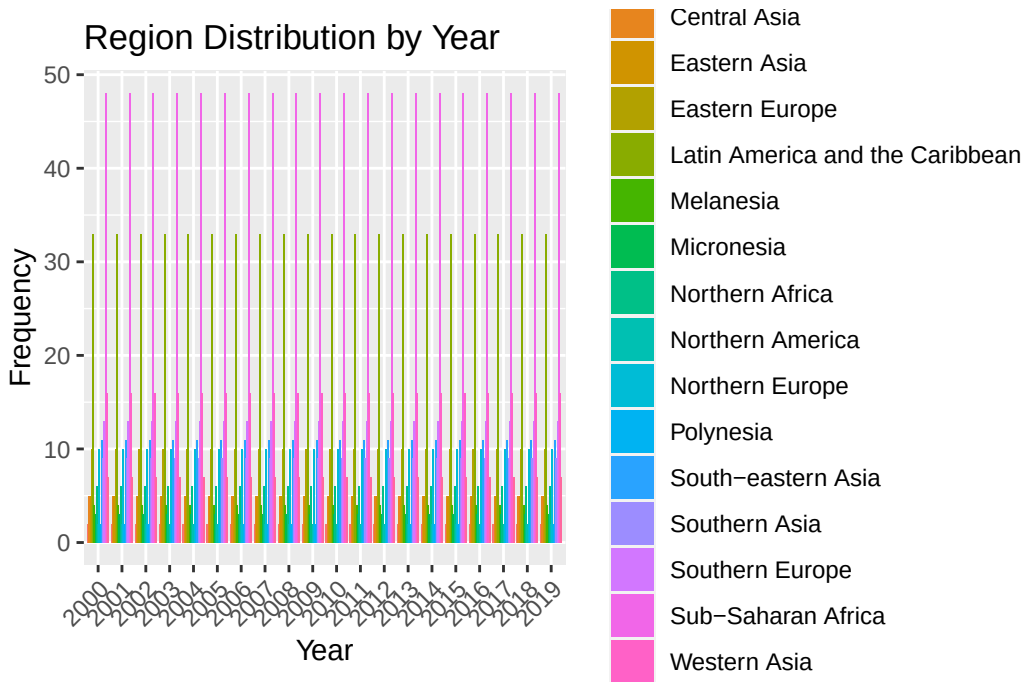
```
           0 1
  Chile    1 0
  Latvia   1 0
  Lithuania 1 0

, ,  = 2005


           0 1
  Chile    1 0
  Latvia   1 0
  Lithuania 1 0

, ,  = 2006


           0 1
  Chile    1 0
  Latvia   1 0
  Lithuania 1 0

, ,  = 2007


           0 1
  Chile    1 0
  Latvia   1 0
  Lithuania 1 0

, ,  = 2008


           0 1
  Chile    1 0
  Latvia   1 0
  Lithuania 1 0

, ,  = 2009


           0 1
  Chile    1 0
  Latvia   1 0
```

```
  Lithuania 1 0

, ,  = 2010


           0 1
  Chile    0 1
  Latvia   1 0
  Lithuania 1 0

, ,  = 2011


           0 1
  Chile    0 1
  Latvia   1 0
  Lithuania 1 0

, ,  = 2012


           0 1
  Chile    0 1
  Latvia   1 0
  Lithuania 1 0

, ,  = 2013


           0 1
  Chile    0 1
  Latvia   1 0
  Lithuania 1 0

, ,  = 2014


           0 1
  Chile    0 1
  Latvia   1 0
  Lithuania 1 0

, ,  = 2015
```

```
           0 1
  Chile    0 1
  Latvia   1 0
  Lithuania 1 0

, ,  = 2016


           0 1
  Chile    0 1
  Latvia   0 1
  Lithuania 1 0

, ,  = 2017


           0 1
  Chile    0 1
  Latvia   0 1
  Lithuania 1 0

, ,  = 2018


           0 1
  Chile    0 1
  Latvia   0 1
  Lithuania 0 1

, ,  = 2019


           0 1
  Chile    0 1
  Latvia   0 1
  Lithuania 0 1
```

```
#Chile became an OECD country in 2011, Latvia in 2017 and Lithuania in 2019.
```

```
### DENSITY PLOTS FOR CATEGORICAL VARIABLES ###

# Plot for region by year

library(ggplot2)

ggplot(finaldata, aes(x = as.factor(Year), fill = region)) +
  geom_bar(position = "dodge") +
  labs(title = "Region Distribution by Year",
       x = "Year",
       y = "Frequency",
       fill = "Region") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



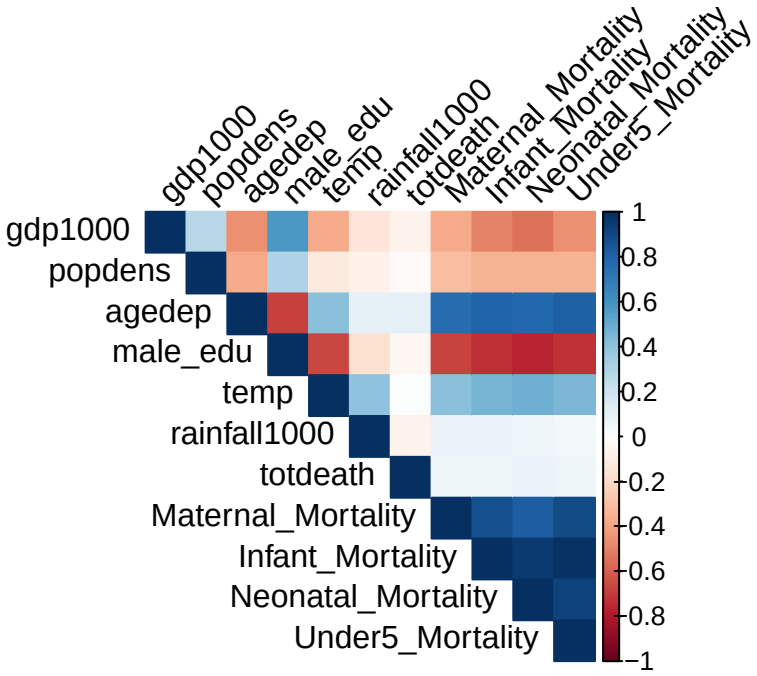

```
## Sub-Saharan Africa and Latin American & the Caribbean were the regions that had the highe

### CORRELATION MATRIX FOR CONTINUOUS VARIABLES ###

# Compute correlation matrix
continuous_vars <- finaldata[, c("gdp1000", "popdens", "agedep", "male_edu", "temp", "rainfa
cor_matrix <- cor(continuous_vars, use = "complete.obs")
```

```
# Visualize the correlation matrix
library(corrplot)
```

```
Warning: package 'corrplot' was built under R version 4.2.3
```

```
corrplot 0.92 loaded
```

```
corrplot(cor_matrix, method = "color", type = "upper", tl.col = "black", tl.srt = 45)
```

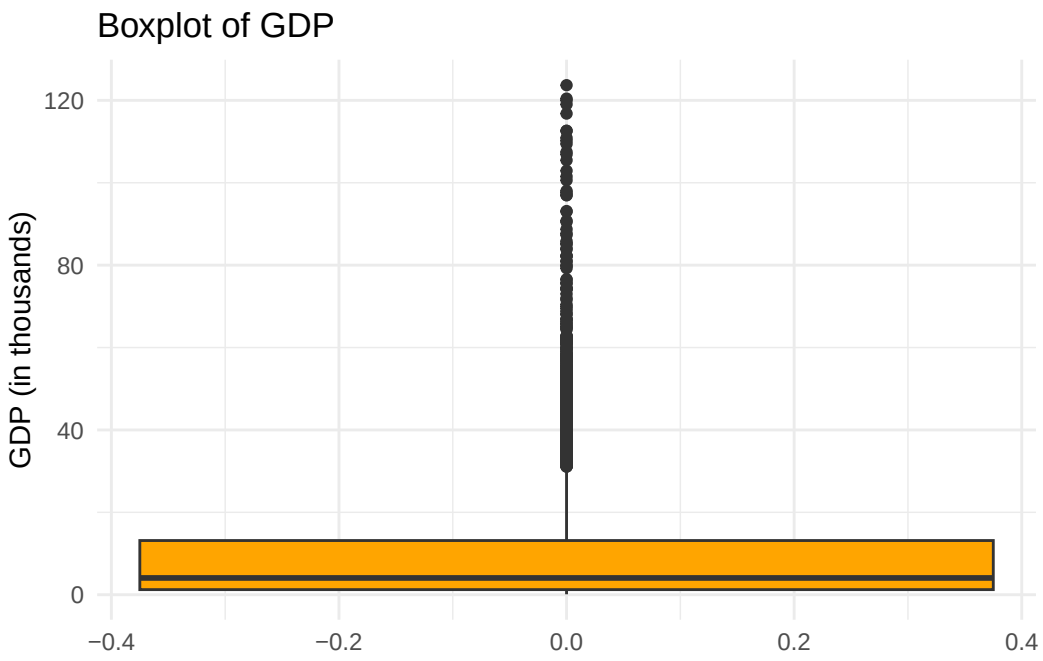



```
## All the mortality variables are correlated with each other, suggesting that countries wit

## GDP seems to be negatively correlated with maternal, infant and neonatal mortality, sugge


### SUMMARIZE DATA BY GROUP ###
# Summarize data by region #

library(dplyr)
finaldata %>%
  group_by(region) %>%
  summarise(mean_gdp = mean(gdp1000, na.rm = TRUE),
            mean_popdens = mean(popdens, na.rm = TRUE))
```

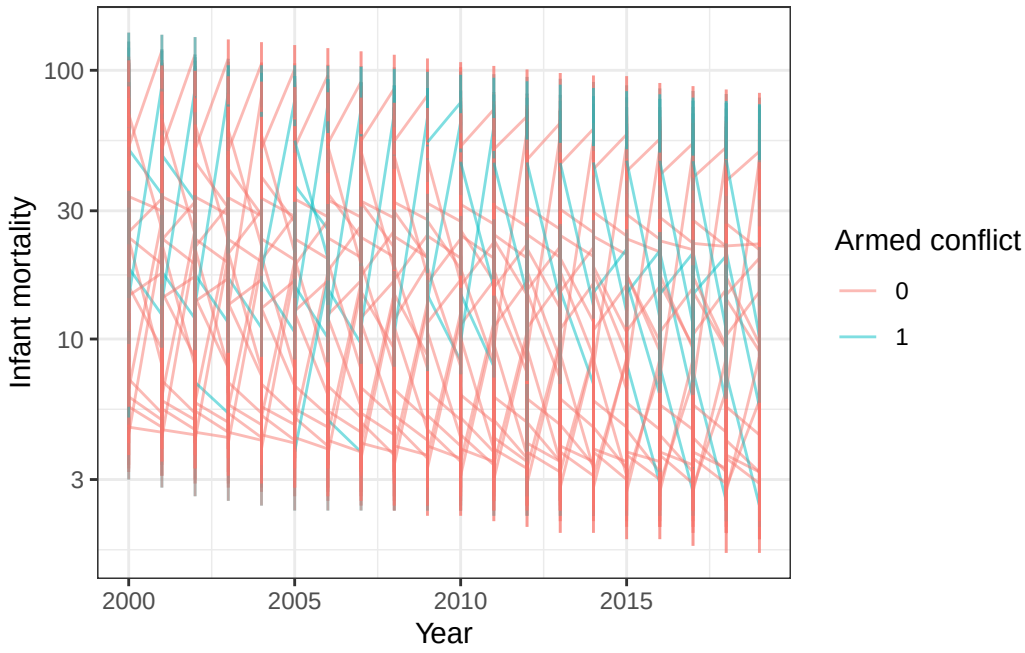
```
# A tibble: 17 x 3
   region                           mean_gdp mean_popdens
   <chr>                               <dbl>        <dbl>
 1 Australia and New Zealand          38.8          67.5
 2 Central Asia                        2.89         13.9
 3 Eastern Asia                       17.4          41.0
 4 Eastern Europe                      8.66         30.0
 5 Latin America and the Caribbean     7.93         33.2
 6 Melanesia                           2.43         18.7
 7 Micronesia                          2.40         34.1
 8 Northern Africa                     3.81         37.6
 9 Northern America                   45.3          61.4
10 Northern Europe                    40.3          35.1
11 Polynesia                           3.22          0
12 South-eastern Asia                  8.78         32.5
13 Southern Asia                       2.24         34.3
14 Southern Europe                    16.7          32.8
15 Sub-Saharan Africa                  2.01         21.9
16 Western Asia                       16.0          37.1
17 Western Europe                     53.1          36.7
```

```r
### CHECK OUTLIERS ###

# Boxplot for GDP

library(ggplot2)

ggplot (finaldata, aes(y = gdp1000)) +
  geom_boxplot(fill = "orange") +
  labs(title = "Boxplot of GDP", y = "GDP (in thousands)") +
  theme_minimal()
```

```
Warning: Removed 62 rows containing non-finite values (`stat_boxplot()`).
```

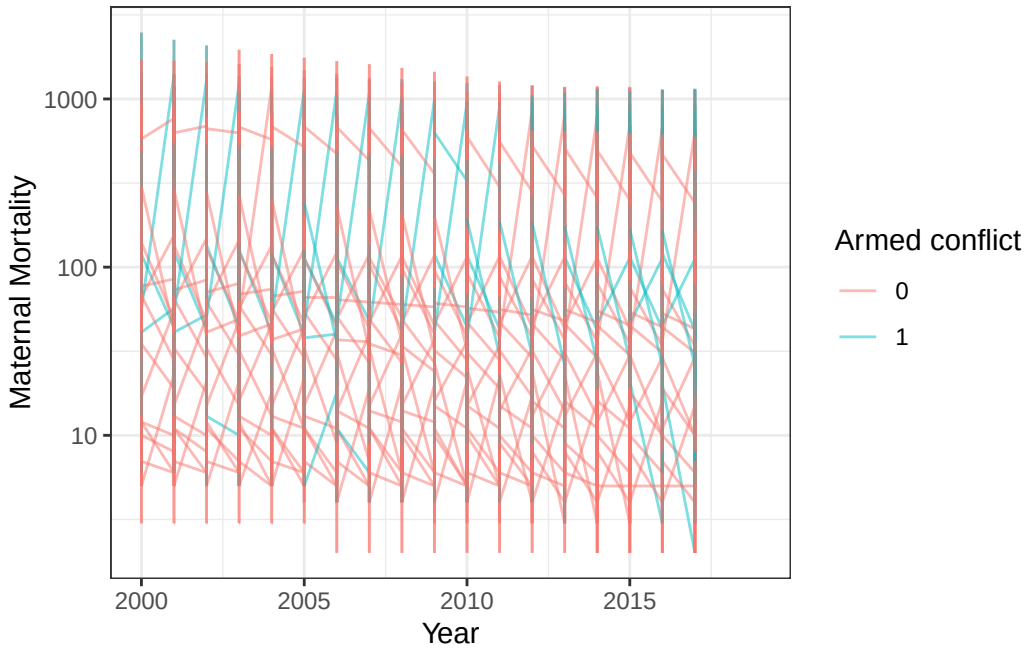



```
 # the 62  missing values for gdp1000 were removed from this plot



# Plot and look at infant mortality by armed conflict status #

finaldata |>
  ggplot(aes(x = Year, y = Infant_Mortality, group = region)) +
  geom_line(aes(color = as.factor(armconf1)), alpha = 0.5) +
  xlim(c(2000,2019)) +
  scale_y_continuous(trans='log10') +
  labs(y = "Infant mortality", x = "Year", color = "Armed conflict") +
  theme_bw()
```

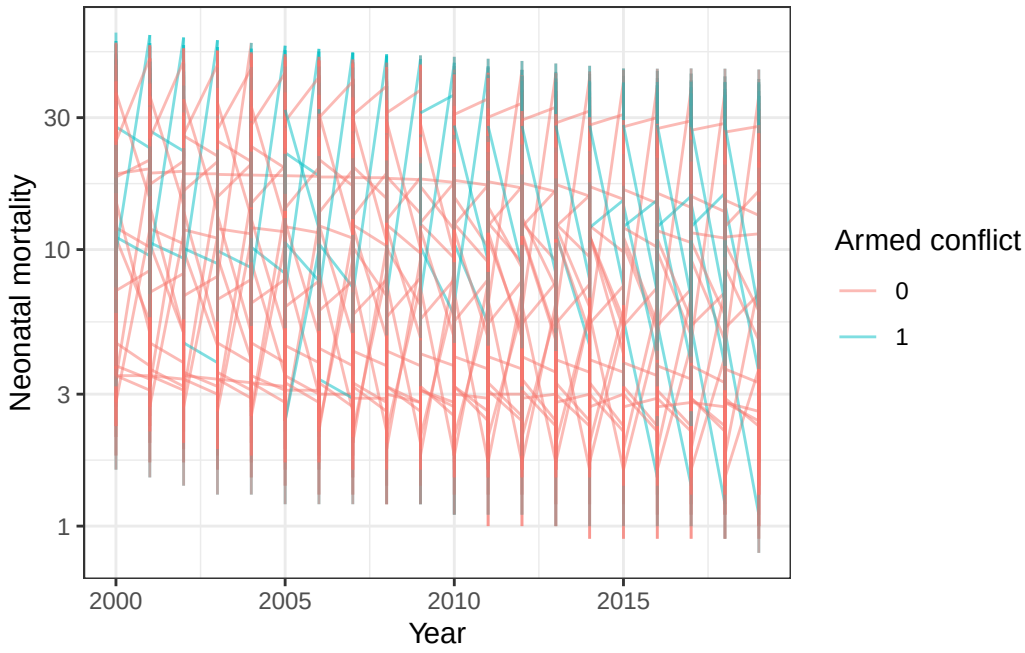



```
# Plot and look at maternal mortality by armed conflict status #

finaldata |>
  ggplot(aes(x = Year, y = Maternal_Mortality, group = region)) +
  geom_line(aes(color = as.factor(armconf1)), alpha = 0.5) +
  xlim(c(2000,2019)) +
  scale_y_continuous(trans='log10') +
  labs(y = "Maternal Mortality", x = "Year", color = "Armed conflict") +
  theme_bw()
```

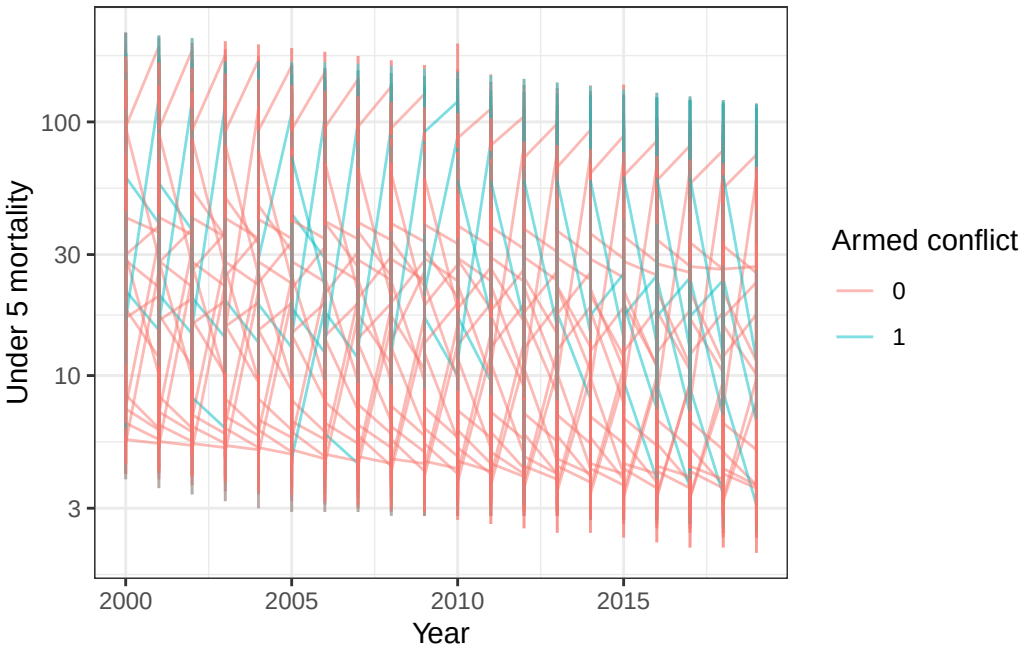
```
Warning: Removed 373 rows containing missing values (`geom_line()`).
```




```
# Plot and look at neonetal mortality by armed conflict status #

finaldata |>
  ggplot(aes(x = Year, y = Neonatal_Mortality, group = region)) +
  geom_line(aes(color = as.factor(armconf1)), alpha = 0.5) +
  xlim(c(2000,2019)) +
  scale_y_continuous(trans='log10') +
  labs(y = "Neonatal mortality", x = "Year", color = "Armed conflict") +
  theme_bw()
```




```
# Plot and look at under 5 mortality by armed conflict status #

finaldata |>
  ggplot(aes(x = Year, y = Under5_Mortality, group = region)) +
  geom_line(aes(color = as.factor(armconf1)), alpha = 0.5) +
  xlim(c(2000,2019)) +
  scale_y_continuous(trans='log10') +
  labs(y = "Under 5 mortality", x = "Year", color = "Armed conflict") +
  theme_bw()
```


Under 5 mortality
100
30
10
3
2000
2005
2010
2015
Year
Armed conflict
0
1