Combined Data EDA

Katie Schilling - 501130072

04/04/2022

```
library(readxl)
library(caTools)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
require(lmtest)
```

```
## Loading required package: lmtest

## Loading required package: zoo

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```
library(dlookr)
```

```
##
## Attaching package: 'dlookr'

## The following object is masked from 'package:base':
##
##     transform
```

Import bothe Clean datasets

```
Covid_monthly <- read.csv("C:/Users/Katie Schilling/Downloads/covid_monthly_clean.csv")
Vital_Events <- read.csv("C:/Users/Katie Schilling/Downloads/vital_events_clean.csv")
```

Combine the vital events data with the Covid Monthly data

```
Final_dataset <- merge(x=Vital_Events, y=Covid_monthly, all = TRUE)
```

Check the data and ensure data merged properly

```
summary(Final_dataset)
```

```
##      Date               Births        Marriages         Deaths
##  Length:336         Min.   :10020   Min.   :  597   Min.   : 5926
##  Class :character   1st Qu.:11260   1st Qu.: 2596   1st Qu.: 6706
##  Mode  :character   Median :11818   Median : 3559   Median : 7326
##                     Mean   :11763   Mean   : 5085   Mean   : 7500
##                     3rd Qu.:12288   3rd Qu.: 7627   3rd Qu.: 8094
##                     Max.   :13398   Max.   :11532   Max.   :11390
##                     NA's   :6       NA's   :6       NA's   :6
##   Stillbirths         Covid
##  Min.   :  0.00   Min.   :    11
##  1st Qu.: 73.00   1st Qu.: 22889
##  Median : 90.50   Median : 75935
##  Mean   : 90.74   Mean   : 95144
##  3rd Qu.:114.00   3rd Qu.:134128
##  Max.   :156.00   Max.   :395815
##  NA's   :6        NA's   :313
```

Change the N/A in the Covid Positive Cases to 0 so that the data is not omitted from the predictions

```
Final_dataset$Covid[is.na(Final_dataset$Covid)] = 0
```

See if there are anymore NA's in the dataset

```
Final_dataset %>% filter_all(any_vars(is.na(.)))
```

```
##        Date Births Marriages Deaths Stillbirths  Covid
## 1 2021-07-1     NA        NA     NA          NA  15968
## 2 2021-08-1     NA        NA     NA          NA  67913
## 3 2021-09-1     NA        NA     NA          NA 125560
## 4 2021-10-1     NA        NA     NA          NA  91834
## 5 2021-11-1     NA        NA     NA          NA  75935
## 6 2021-12-1     NA        NA     NA          NA 395815
```

Remove rows with NA as they will skew the results

```
Final_dataset <- na.omit(Final_dataset)
```

Check for NA's to confirm all have been removed

```
Final_dataset %>% filter_all(any_vars(is.na(.)))
```

```
## [1] Date        Births      Marriages   Deaths      Stillbirths Covid
## <0 rows> (or 0-length row.names)
```

```
summary(Final_dataset)
```

```
##      Date                Births        Marriages         Deaths
##  Length:330         Min.   :10020   Min.   :  597   Min.   : 5926
##  Class :character   1st Qu.:11260   1st Qu.: 2596   1st Qu.: 6706
##  Mode  :character   Median :11818   Median : 3559   Median : 7326
##                     Mean   :11763   Mean   : 5085   Mean   : 7500
##                     3rd Qu.:12288   3rd Qu.: 7627   3rd Qu.: 8094
##                     Max.   :13398   Max.   :11532   Max.   :11390
##   Stillbirths         Covid
##  Min.   :  0.00   Min.   :     0
##  1st Qu.: 73.00   1st Qu.:     0
##  Median : 90.50   Median :     0
##  Mean   : 90.74   Mean   :  4289
##  3rd Qu.:114.00   3rd Qu.:     0
##  Max.   :156.00   Max.   :237308
```

```
Final_dataset$Date <- as.Date(Final_dataset$Date,"%Y-%m-%d")
```

```
glimpse(Final_dataset)
```

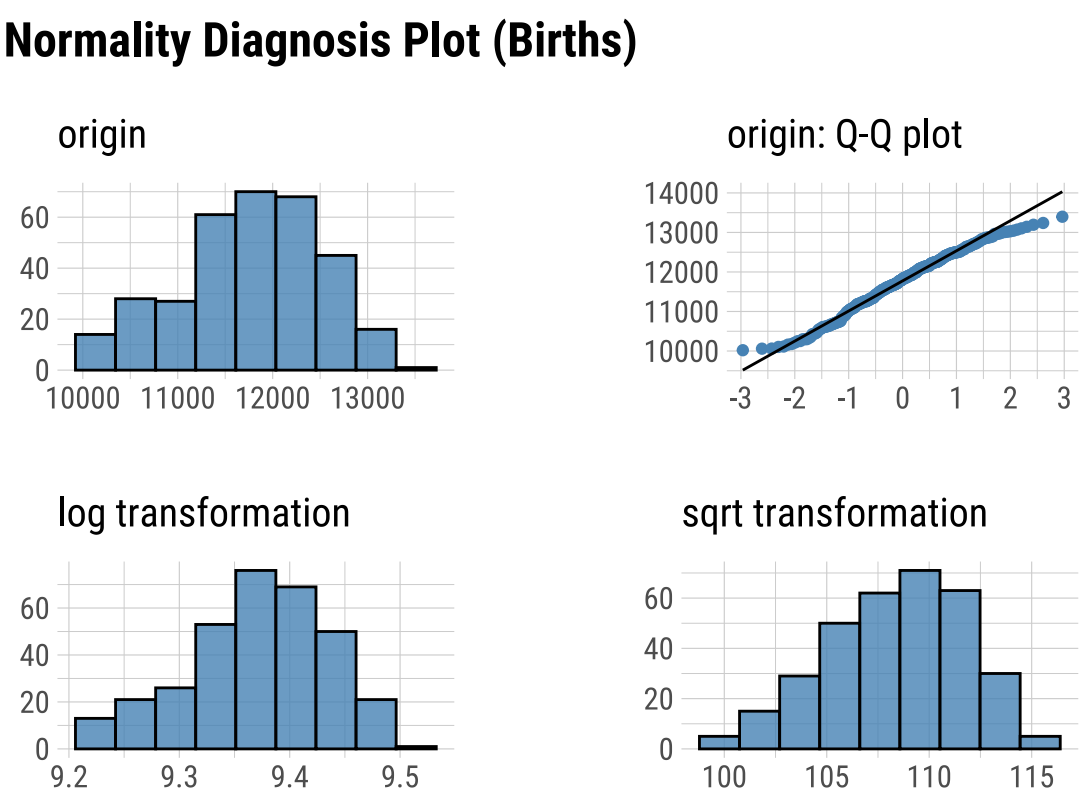
```
## Rows: 330
## Columns: 6
## $ Date        <date> 1994-01-01, 1994-02-01, 1994-03-01, 1994-04-01, 1994-05-0~
## $ Births      <int> 11631, 11254, 13003, 12576, 13240, 13072, 13045, 12982, 12~
## $ Marriages   <int> 2078, 2650, 2557, 3967, 6493, 7754, 9264, 9194, 8540, 7400~
## $ Deaths      <int> 8094, 6428, 6503, 6224, 6483, 6187, 6196, 5926, 6062, 6515~
## $ Stillbirths <int> 75, 62, 73, 74, 67, 66, 70, 79, 60, 59, 56, 43, 78, 84, 75~
## $ Covid       <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
```

Create function to view basic EDA

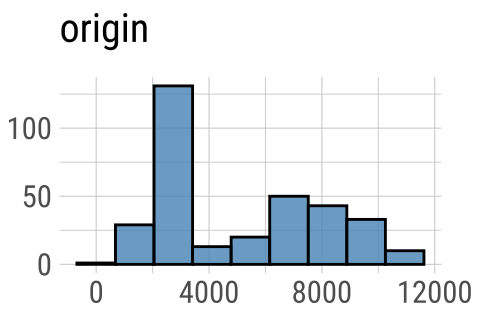
```
basic_eda <- function(data)
{
  glimpse(data)
  print(status(data))
  freq(data)
  print(profiling_num(data))
  plot_num(data)
  describe(data)
}
```

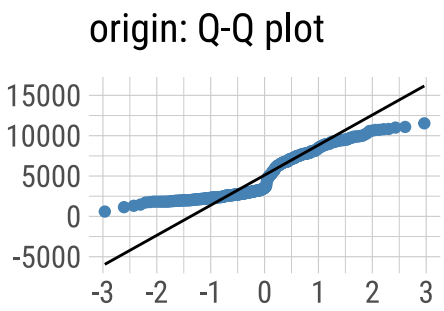
Visual analysis of the normality of the data.

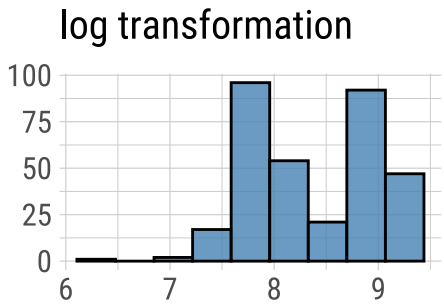
```
plot_normality(Final_dataset)
```

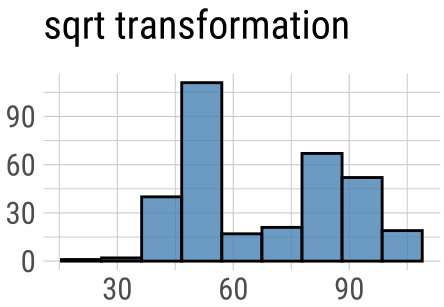


Normality Diagnosis Plot (Marriages)

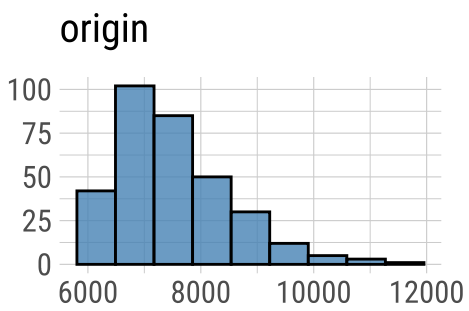


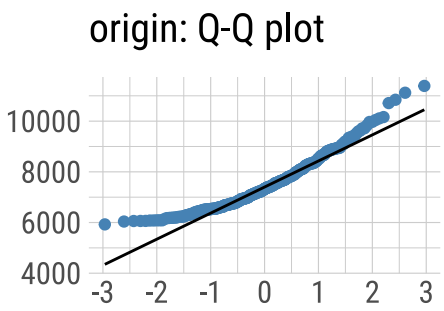


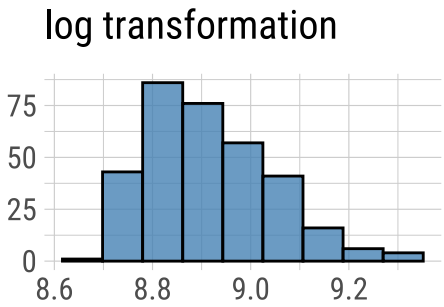


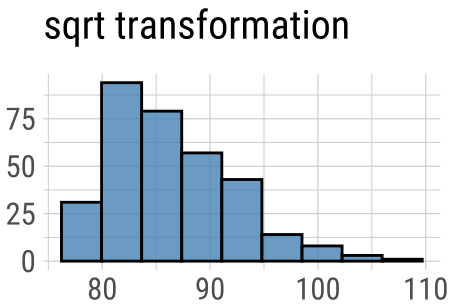


Normality Diagnosis Plot (Deaths)

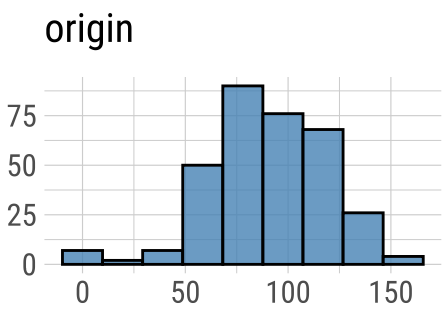


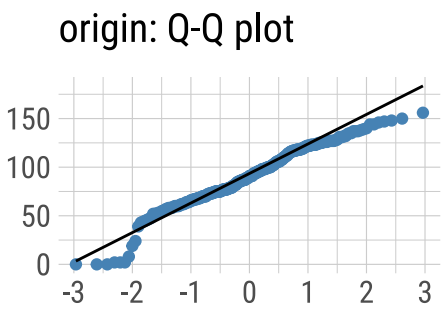


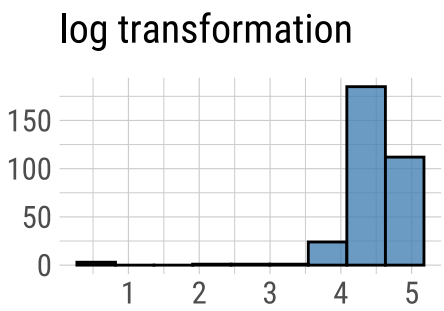


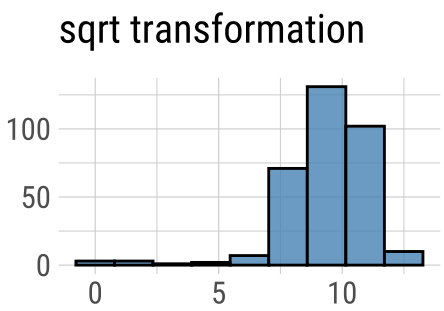


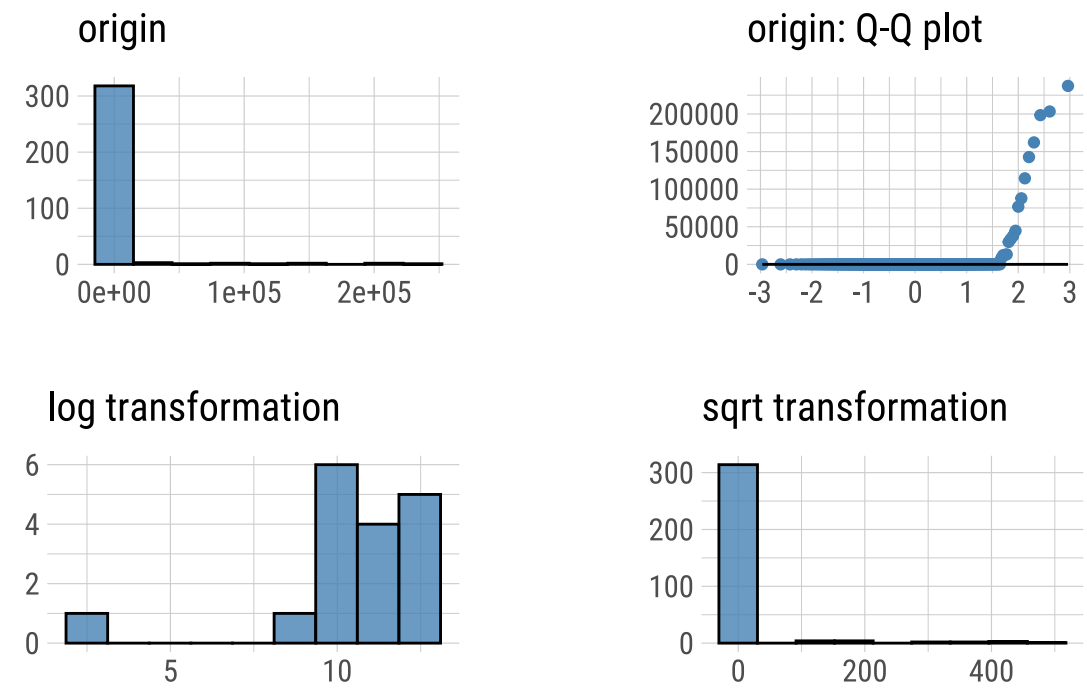
Normality Diagnosis Plot (Stillbirths)











Find the correlation, if any, between the variables in the data

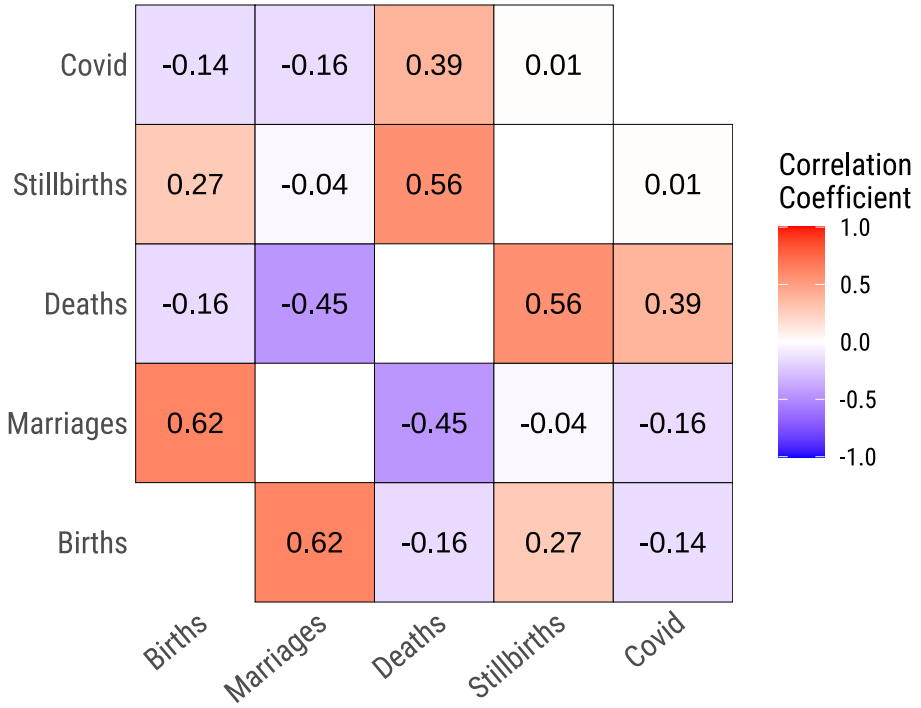
```
correlate(Final_dataset)
```

```
## # A tibble: 20 x 3
##    var1        var2        coef_corr
##    <fct>       <fct>           <dbl>
##  1 Marriages   Births          0.618
##  2 Deaths      Births         -0.158
##  3 Stillbirths Births          0.273
##  4 Covid       Births         -0.144
##  5 Births      Marriages       0.618
##  6 Deaths      Marriages      -0.451
##  7 Stillbirths Marriages      -0.0353
##  8 Covid       Marriages      -0.155
##  9 Births      Deaths         -0.158
## 10 Marriages   Deaths         -0.451
## 11 Stillbirths Deaths          0.564
## 12 Covid       Deaths          0.390
## 13 Births      Stillbirths     0.273
## 14 Marriages   Stillbirths    -0.0353
## 15 Deaths      Stillbirths     0.564
## 16 Covid       Stillbirths     0.0124
## 17 Births      Covid          -0.144
## 18 Marriages   Covid          -0.155
## 19 Deaths      Covid           0.390
```

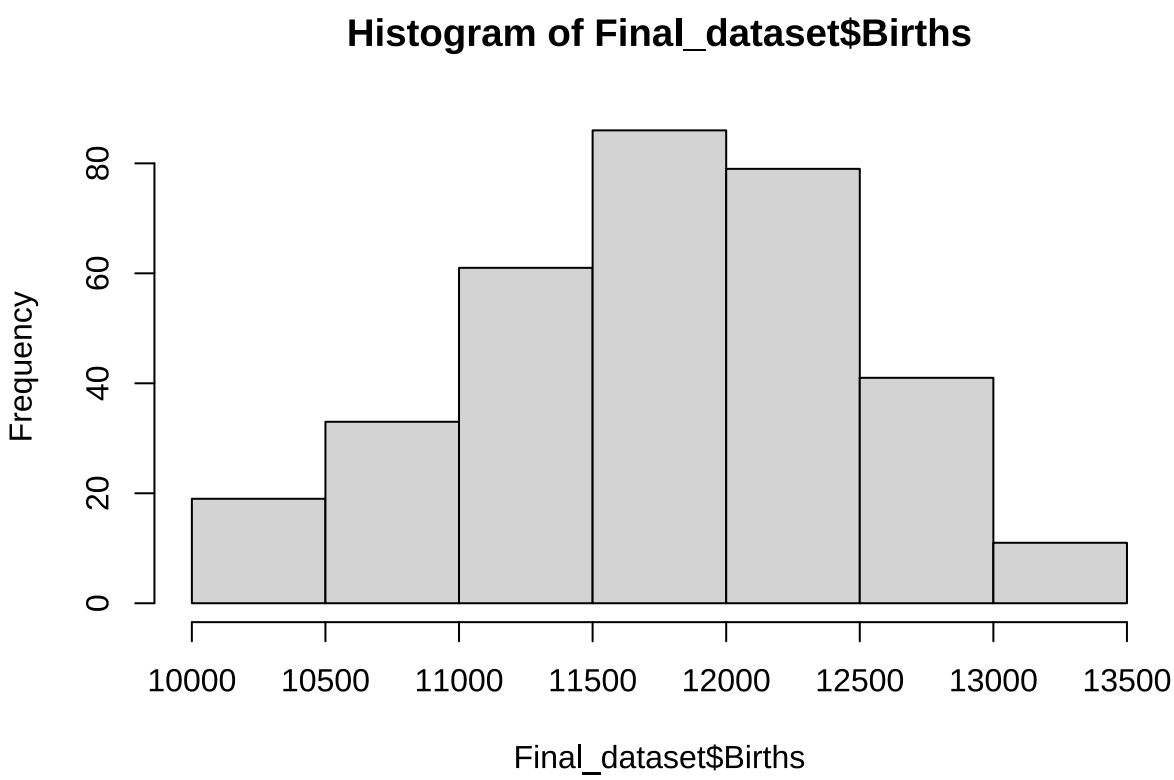
```
## 20 Stillbirths Covid          0.0124
```

visualize the correlation, if any, of the data

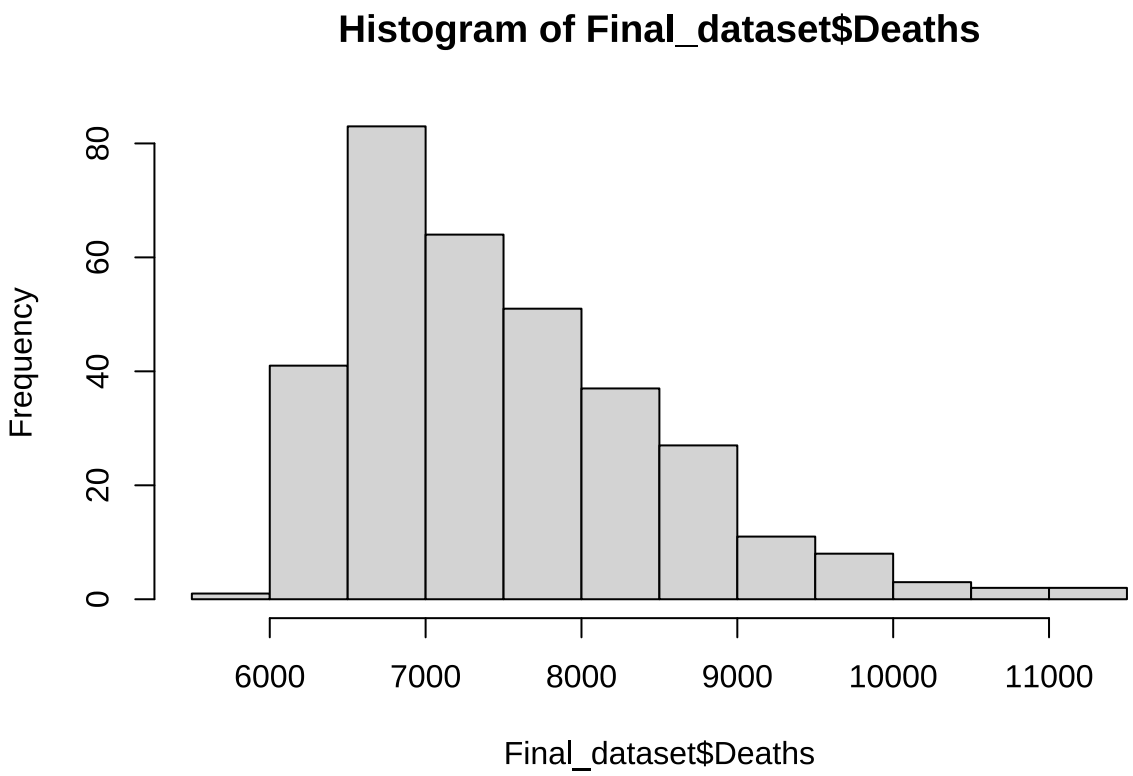
```
plot_correlate(Final_dataset)
```



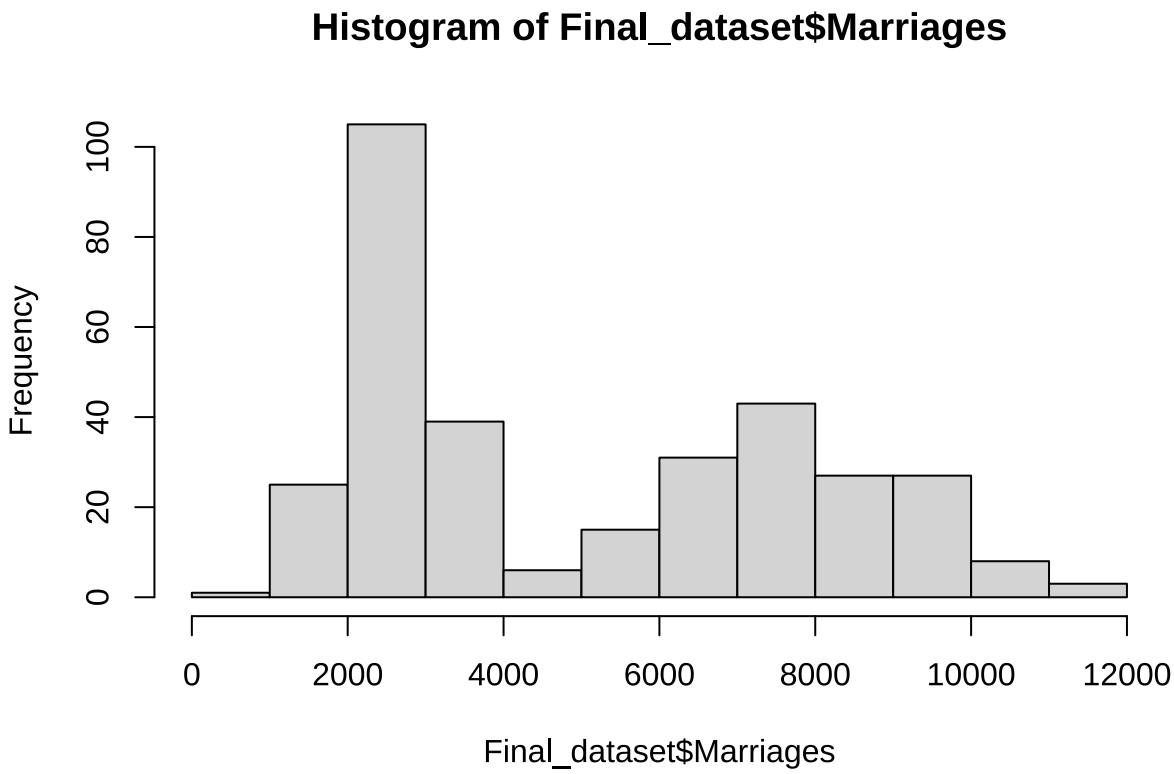

```
hist(Final_dataset$Births)
```



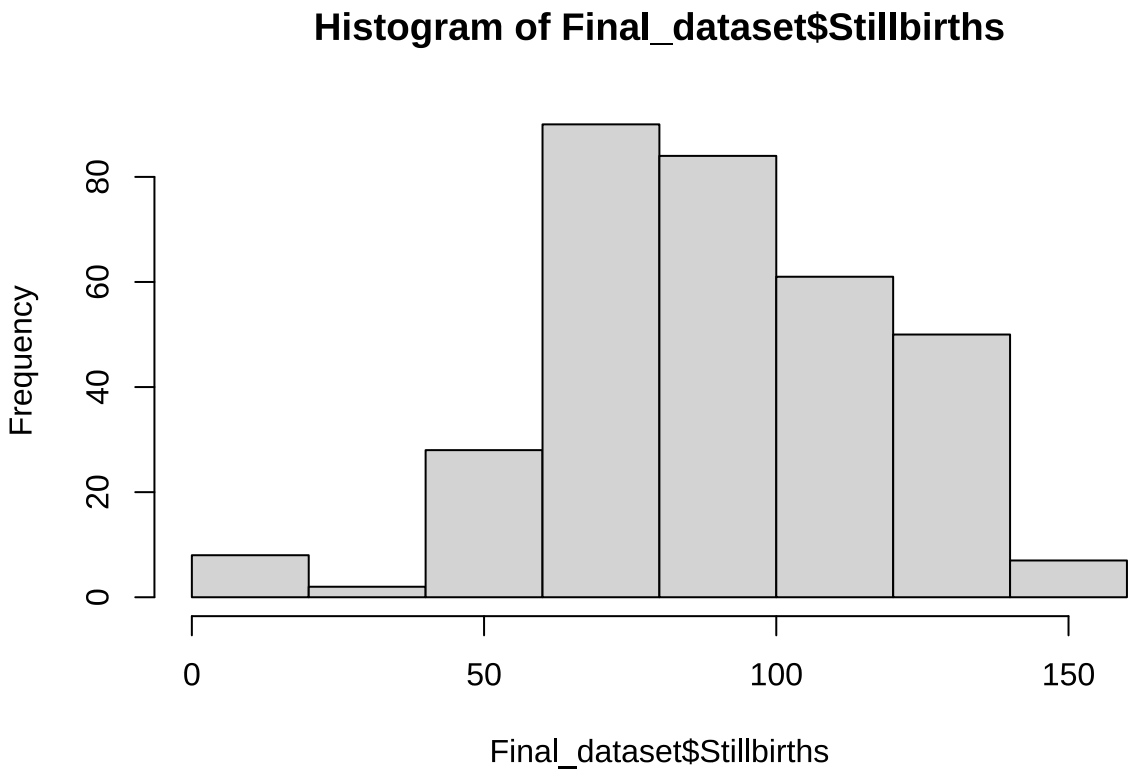

```
hist(Final_dataset$Deaths)
```



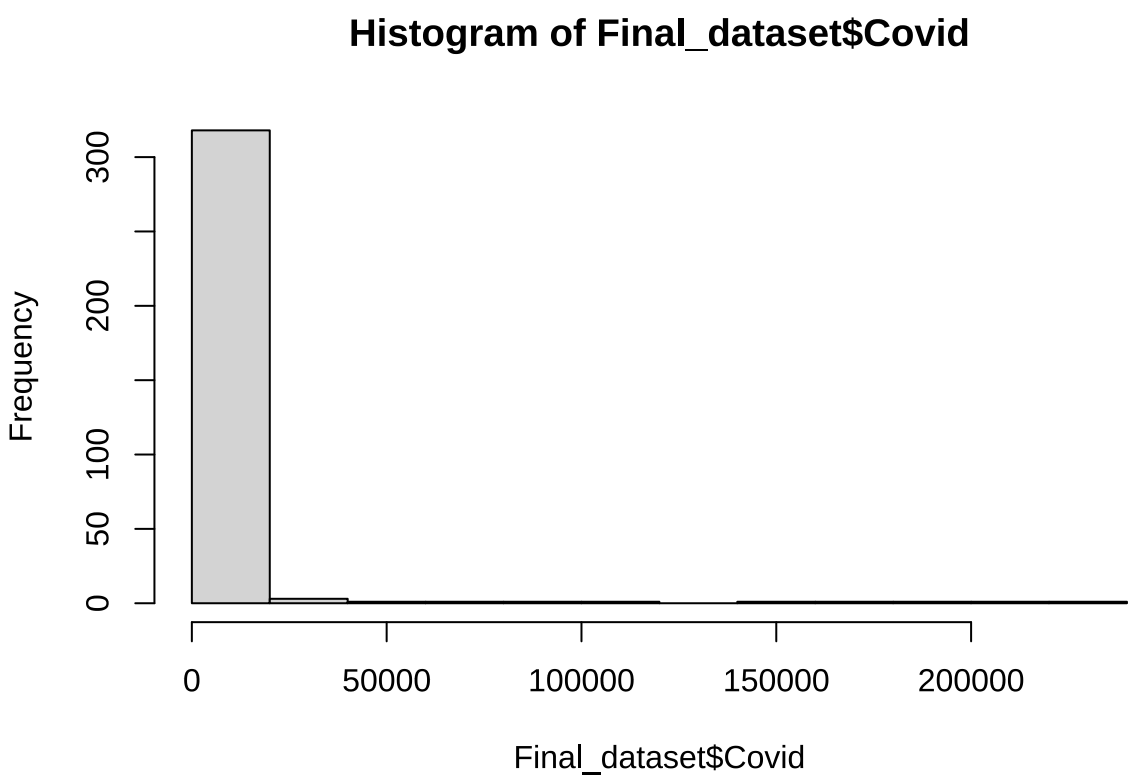

```
hist(Final_dataset$Marriages)
```



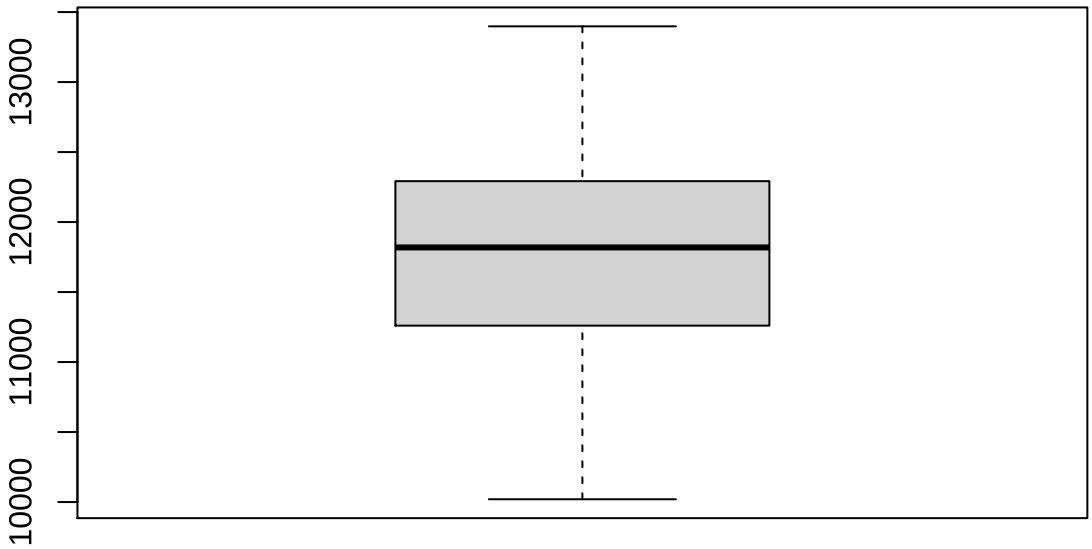

```
hist(Final_dataset$Stillbirths)
```



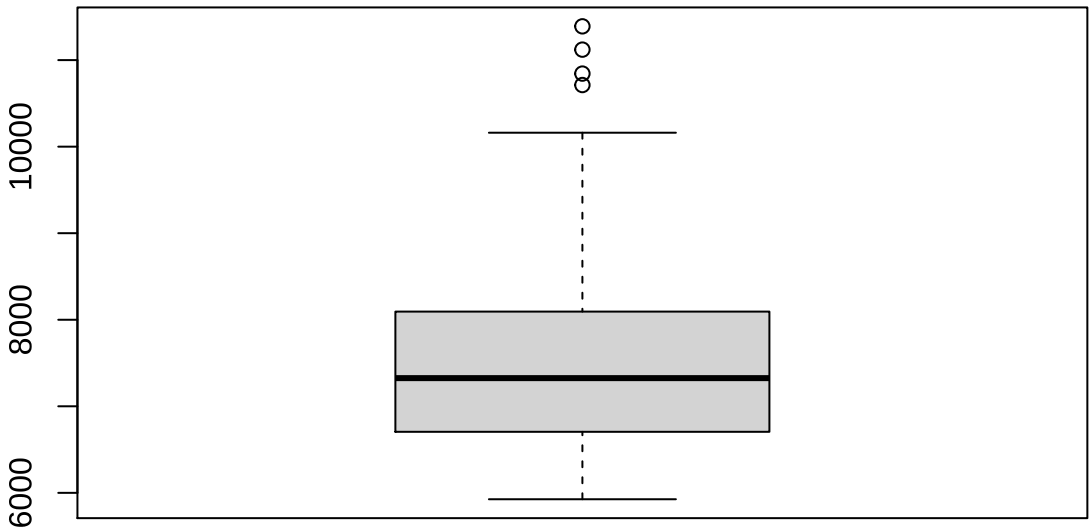

```
hist(Final_dataset$Covid)
```



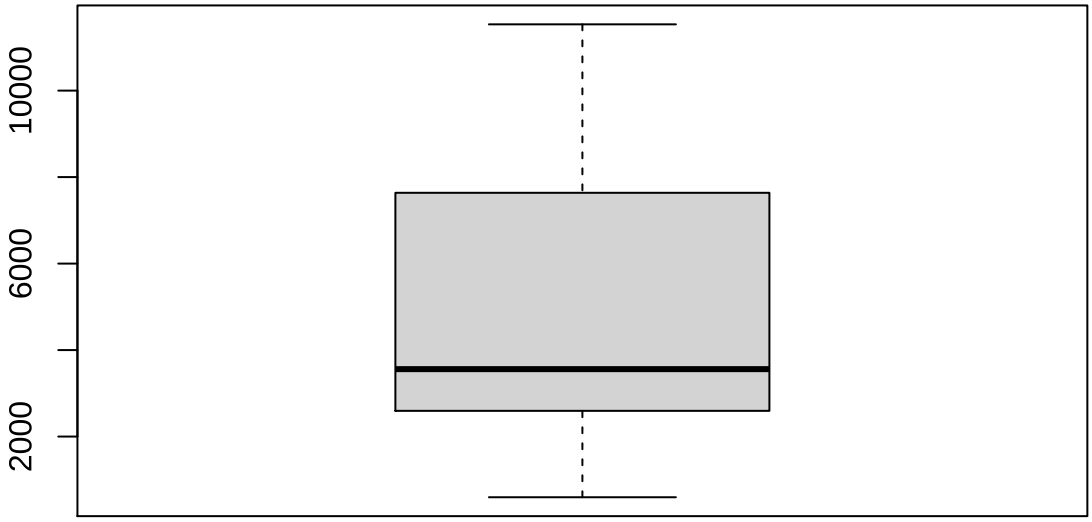

```
boxplot(Final_dataset$Births)
```



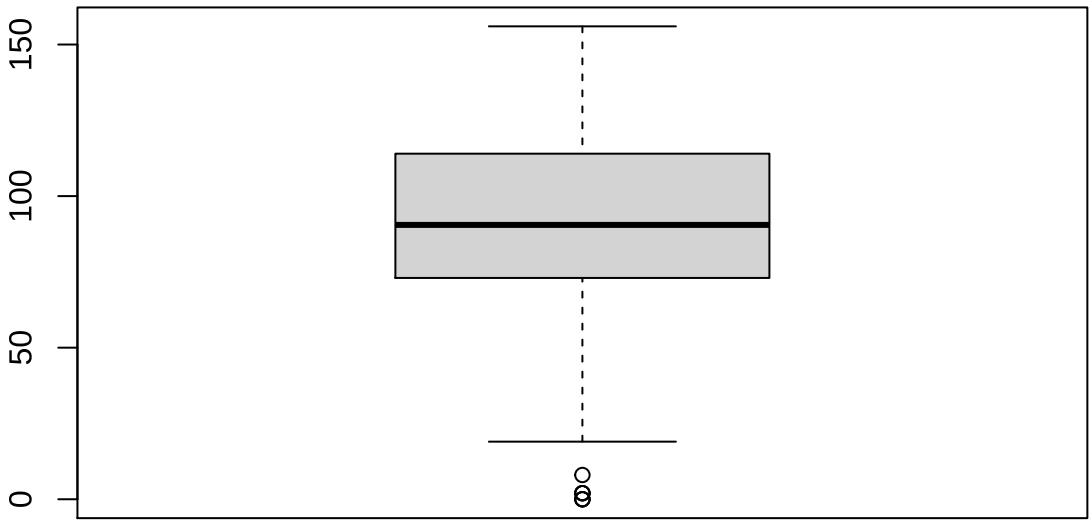

```
boxplot(Final_dataset$Deaths)
```



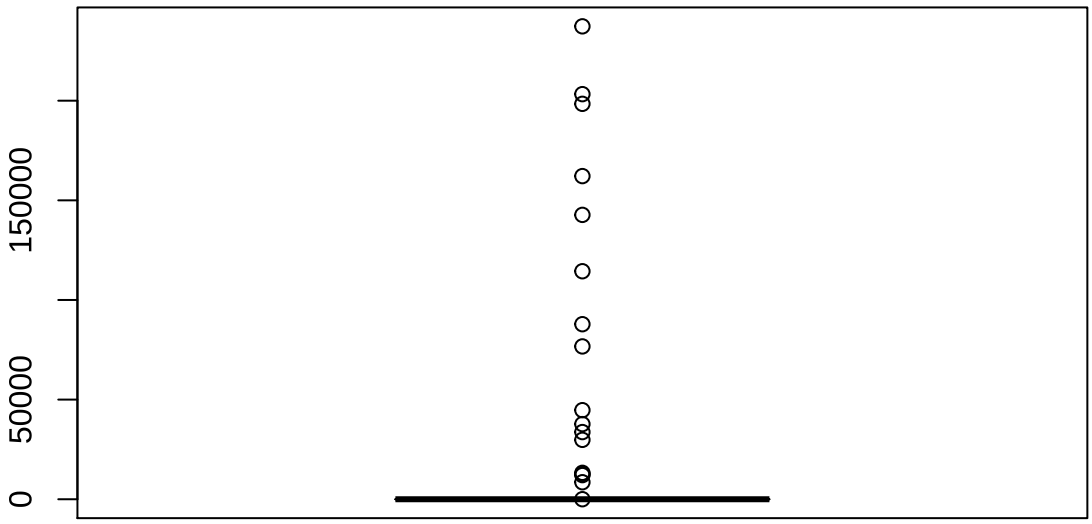

```
boxplot(Final_dataset$Marriages)
```




```
boxplot(Final_dataset$Stillbirths)
```




```
boxplot(Final_dataset$Covid)
```




```
eda_web_report(Final_dataset)
```

```
##
##
## processing file: eda_temp.Rmd

##   |                                                                      |
##   ordinary text without R code
##
##   |                                                                      |..
## label: setup (with options)
## List of 3
##  $ echo   : logi FALSE
##  $ warning: logi FALSE
##  $ message: logi FALSE
##
##   |                                                                      |...
##   ordinary text without R code
##
##   |                                                                      |....
## label: load_packages
##   |                                                                      |......
##   ordinary text without R code
##
##   |                                                                      |.......
```

```
## label: unnamed-chunk-19 (with options)
## List of 2
##  $ echo  : logi FALSE
##  $ engine: chr "css"
##
##   |                                                                      |........
##   ordinary text without R code
##
##   |                                                                      |.........
## label: udf (with options)
## List of 3
##  $ echo   : logi FALSE
##  $ warning: logi FALSE
##  $ message: logi FALSE
##
##   |                                                                      |..........
##   ordinary text without R code
##
##   |                                                                      |...........
## label: check_variables (with options)
## List of 4
##  $ echo   : logi FALSE
##  $ warning: logi FALSE
##  $ message: logi FALSE
##  $ comment: chr ""
##
##   |                                                                      |............
##   ordinary text without R code
##
##   |                                                                      |.............
## label: create-overview
##   |                                                                      |..............
##   ordinary text without R code
##
##   |                                                                      |................
## label: overview (with options)
## List of 1
##  $ results: chr "asis"
##
##   |                                                                      |.................
##   ordinary text without R code
##
##   |                                                                      |..................
## label: overview-pre (with options)
## List of 1
##  $ results: chr "asis"
##
##   |                                                                      |..................
##   ordinary text without R code
##
##   |                                                                      |..................
## label: unnamed-chunk-20 (with options)
## List of 1
##  $ results: chr "asis"
```

```
##
##   |                                                                      |.....................
##   ordinary text without R code
##
##   |                                                                      |.....................
## label: unnamed-chunk-21 (with options)
## List of 1
##  $ results: chr "asis"
##
##   |                                                                      |.....................
##   ordinary text without R code
##
##   |                                                                      |.....................
## label: variables (with options)
## List of 1
##  $ results: chr "asis"

##   |                                                                      |.....................
##   ordinary text without R code
##
##   |                                                                      |.....................
## label: normality (with options)
## List of 1
##  $ results: chr "asis"
##
##   |                                                                      |.....................
##   ordinary text without R code
##
##   |                                                                      |.....................
## label: normality-list (with options)
## List of 2
##  $ comment: chr ""
##  $ results: chr "asis"

##   |                                                                      |.....................
##   ordinary text without R code
##
##   |                                                                      |.....................
## label: unnamed-chunk-22 (with options)
## List of 1
##  $ results: chr "asis"
##
##   |                                                                      |.....................
##   ordinary text without R code
##
##   |                                                                      |.....................
## label: unnamed-chunk-23 (with options)
## List of 1
##  $ results: chr "asis"
##
##   |                                                                      |.....................
##   ordinary text without R code
##
##   |                                                                      |.....................
```

```
## label: compare_numerical (with options)
## List of 1
##  $ results: chr "asis"

##   |                                                                      |.................
##   ordinary text without R code
##
##   |                                                                      |.................
## label: unnamed-chunk-24 (with options)
## List of 1
##  $ results: chr "asis"
##
##   |                                                                      |.................
##   ordinary text without R code
##
##   |                                                                      |.................
## label: compare-category (with options)
## List of 1
##  $ results: chr "asis"
##
##   |                                                                      |.................
##   ordinary text without R code
##
##   |                                                                      |.................
## label: unnamed-chunk-25 (with options)
## List of 1
##  $ results: chr "asis"
##
##   |                                                                      |.................
##   ordinary text without R code
##
##   |                                                                      |.................
## label: unnamed-chunk-26 (with options)
## List of 1
##  $ results: chr "asis"
##
##   |                                                                      |.................
##   ordinary text without R code
##
##   |                                                                      |.................
## label: unnamed-chunk-27 (with options)
## List of 1
##  $ results: chr "asis"
##
##   |                                                                      |.................
##   ordinary text without R code
##
##   |                                                                      |.................
## label: correlation (with options)
## List of 1
##  $ results: chr "asis"
##
##   |                                                                      |.................
##   ordinary text without R code
```

```
##
##   |                                                                      |.................
## label: unnamed-chunk-28 (with options)
## List of 1
##  $ results: chr "asis"
##
##   |                                                                      |.................
##   ordinary text without R code
##
##   |                                                                      |.................
## label: plot-correlation (with options)
## List of 1
##  $ results: chr "asis"

##   |                                                                      |.................
##   ordinary text without R code
##
##   |                                                                      |.................
## label: unnamed-chunk-29 (with options)
## List of 1
##  $ results: chr "asis"
##
##   |                                                                      |.................
##   ordinary text without R code
##
##   |                                                                      |.................
## label: unnamed-chunk-30 (with options)
## List of 1
##  $ results: chr "asis"
##
##   |                                                                      |.................
##   ordinary text without R code
##
##   |                                                                      |.................
## label: group-numerical (with options)
## List of 1
##  $ results: chr "asis"
##
##   |                                                                      |.................
##   ordinary text without R code
##
##   |                                                                      |.................
## label: unnamed-chunk-31 (with options)
## List of 1
##  $ results: chr "asis"
##
##   |                                                                      |.................
##   ordinary text without R code
##
##   |                                                                      |.................
## label: group-categorical (with options)
## List of 1
##  $ results: chr "asis"
##
```

```
##   |                                                                      |.................
##   ordinary text without R code
##
##   |                                                                      |.................
## label: unnamed-chunk-32 (with options)
## List of 1
##  $ results: chr "asis"
##
##   |                                                                      |.................
##   ordinary text without R code
##
##   |                                                                      |.................
## label: group-correlation (with options)
## List of 1
##  $ results: chr "asis"
##
##   |                                                                      |.................
##   ordinary text without R code

## output file: eda_temp.knit.md

## "C:/Program Files/RStudio/bin/pandoc/pandoc" +RTS -K512m -RTS eda_temp.knit.md --to html4 --from mark

##
## Output created: C:\Users\KATIES~1\AppData\Local\Temp\Rtmp8WmGij/EDA_Report.html
```