

FML Assignment 1

2023-09-09

```
library(tidyverse)
```

```
## — Attaching core tidyverse packages — tidyverse 2.0.0 —
## ✓ dplyr      1.1.3      ✓ readr      2.1.4
## ✓ forcats    1.0.0      ✓ stringr    1.5.0
## ✓ ggplot2     3.4.3      ✓ tibble     3.2.1
## ✓ lubridate  1.9.2      ✓ tidyr      1.3.0
## ✓ purrr      1.0.2
## — Conflicts — tidyverse_conflicts() —
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
##(1)
#This dataset is downloaded from "Kaggle" site
#https://www.kaggle.com/datasets/rajugc/imdb-top-250-movies-dataset

##(2) Importing dataset
data <- read_csv("/Users/akhilchintu/Downloads/IMDB Top 250 Movies.csv")
```

```
## Rows: 250 Columns: 13
## — Column specification —
## Delimiter: ","
## chr (10): name, genre, certificate, run_time, tagline, budget, box_office, c...
## dbl (3): rank, year, rating
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
data
```

```
## # A tibble: 250 × 13
##   rank name    year rating genre certificate run_time tagline budget box_office
##   <dbl> <chr> <dbl> <dbl> <chr> <chr> <chr> <chr> <chr>
## 1     1 1 The ... 1994  9.3 Drama R          2h 22m Fear c... 25000... 28884504
## 2     2 2 The ... 1972  9.2 Crim... R          2h 55m An off... 60000... 250341816
## 3     3 3 The ... 2008  9 Acti... PG-13          2h 32m Why So... 18500... 1006234167
## 4     4 4 The ... 1974  9 Crim... R          3h 22m All th... 13000... 47961919
## 5     5 5 12 A... 1957  9 Crim... Approved 1h 36m Life I... 350000 955
## 6     6 6 Schi... 1993  9 Biog... R          3h 15m Whoever... 22000... 322161245
## 7     7 7 The ... 2003  9 Acti... PG-13          3h 21m The ey... 94000... 1146457748
## 8     8 8 Pulp... 1994  8.9 Crim... R          2h 34m Girls ... 80000... 213928762
## 9     9 9 The ... 2001  8.8 Acti... PG-13          2h 58m The Le... 93000... 898204420
## 10    10 10 The ... 1966  8.8 Adve... Approved 2h 58m They f... 12000... 25253887
## # i 240 more rows
## # i 3 more variables: casts <chr>, directors <chr>, writers <chr>
```

```
spec(data)
```

```
## cols(
##   rank = col_double(),
##   name = col_character(),
##   year = col_double(),
##   rating = col_double(),
##   genre = col_character(),
##   certificate = col_character(),
##   run_time = col_character(),
##   tagline = col_character(),
##   budget = col_character(),
##   box_office = col_character(),
##   casts = col_character(),
##   directors = col_character(),
##   writers = col_character()
## )
```

#(3) Descriptive statistics for quantitative variables:

```
summary(data[, -c(2,5:13)])
```

```
##           rank           year           rating
##  Min.      : 1.00    Min.    :1921    Min.      :8.000
## 1st Qu.: 63.25    1st Qu.:1966    1st Qu.:8.100
## Median :125.50    Median :1994    Median :8.200
## Mean    :125.50    Mean    :1986    Mean     :8.307
## 3rd Qu.:187.75    3rd Qu.:2006    3rd Qu.:8.400
## Max.    :250.00    Max.     :2022    Max.     :9.300
```

#(3) Descriptive statistics for qualitative variables:

```
summary(data[, -c(1,3,4)])
```

```
##      name          genre      certificate      run_time
## Length:250      Length:250      Length:250      Length:250
## Class :character Class :character Class :character Class :character
## Mode  :character Mode  :character Mode  :character Mode  :character
##      tagline      budget      box_office      casts
## Length:250      Length:250      Length:250      Length:250
## Class :character Class :character Class :character Class :character
## Mode  :character Mode  :character Mode  :character Mode  :character
##      directors      writers
## Length:250      Length:250
## Class :character Class :character
## Mode  :character Mode  :character
```

#(4) Transform one variable:

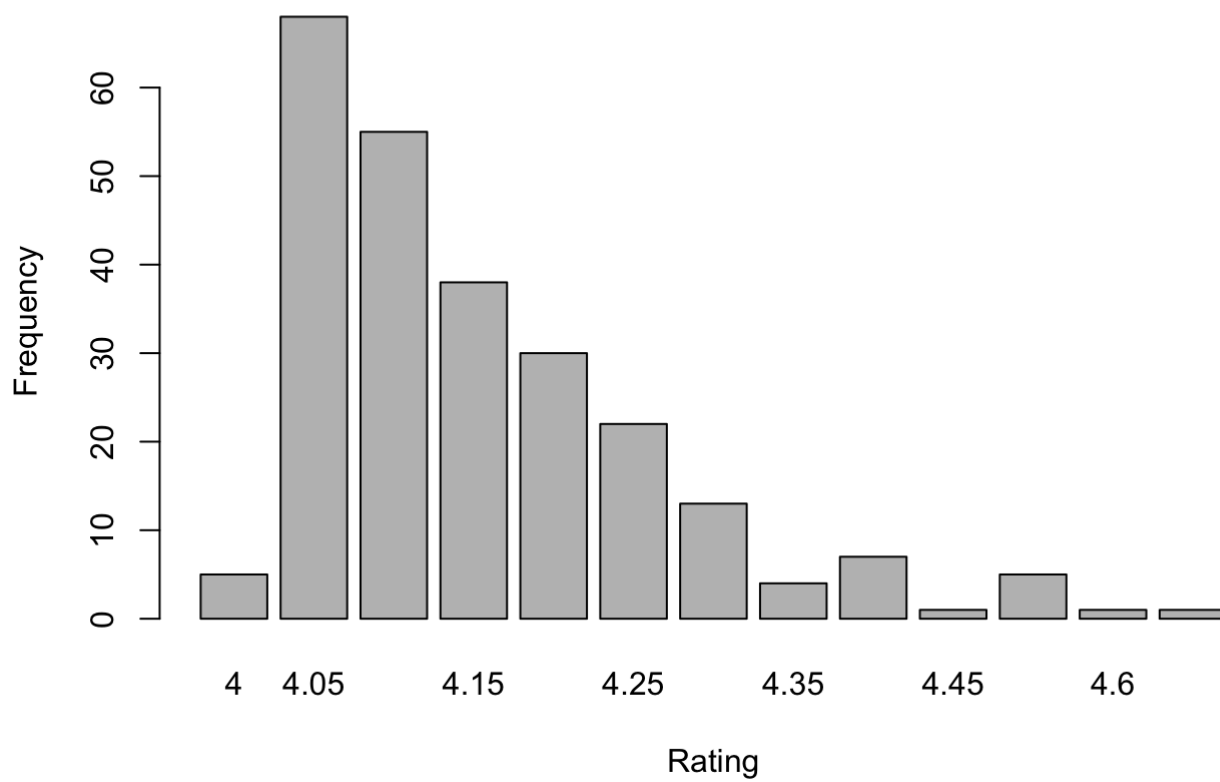
```
data <- data %>% mutate(rating = rating / 2)
print(data$rating)
```

```
##      [1] 4.65 4.60 4.50 4.50 4.50 4.50 4.50 4.50 4.45 4.40 4.40 4.40 4.40 4.40 4.40 4.35
##      [16] 4.35 4.35 4.35 4.30 4.30 4.30 4.30 4.30 4.30 4.30 4.30 4.30 4.30 4.30 4.30 4.25
##      [31] 4.30 4.25 4.25 4.25 4.25 4.25 4.25 4.25 4.25 4.25 4.25 4.25 4.25 4.25 4.25 4.30
##      [46] 4.25 4.25 4.25 4.25 4.25 4.25 4.25 4.25 4.25 4.20 4.20 4.20 4.20 4.20 4.20 4.20
##      [61] 4.20 4.20 4.20 4.20 4.20 4.20 4.20 4.20 4.20 4.20 4.20 4.15 4.20 4.20 4.15 4.20
##      [76] 4.20 4.20 4.20 4.20 4.15 4.15 4.20 4.20 4.15 4.15 4.15 4.20 4.20 4.15 4.15
##      [91] 4.15 4.15 4.20 4.15 4.15 4.15 4.15 4.15 4.15 4.15 4.15 4.15 4.15 4.15 4.15
##     [106] 4.15 4.15 4.20 4.15 4.15 4.15 4.15 4.15 4.15 4.15 4.15 4.10 4.10 4.10 4.10 4.10
##     [121] 4.15 4.15 4.10 4.10 4.15 4.10 4.10 4.15 4.10 4.10 4.10 4.10 4.10 4.10 4.15 4.10
##     [136] 4.10 4.10 4.10 4.10 4.10 4.10 4.10 4.10 4.10 4.10 4.10 4.10 4.10 4.10 4.10
##     [151] 4.10 4.10 4.10 4.10 4.10 4.10 4.10 4.10 4.10 4.10 4.10 4.10 4.10 4.10 4.10
##     [166] 4.10 4.05 4.05 4.10 4.05 4.05 4.05 4.05 4.05 4.10 4.05 4.05 4.05 4.05 4.10
##     [181] 4.05 4.05 4.05 4.05 4.05 4.05 4.05 4.05 4.05 4.05 4.05 4.05 4.10 4.05 4.10
##     [196] 4.05 4.05 4.05 4.05 4.05 4.05 4.05 4.05 4.05 4.05 4.05 4.10 4.10 4.05 4.05
##     [211] 4.05 4.05 4.05 4.05 4.05 4.05 4.05 4.05 4.05 4.05 4.05 4.05 4.15 4.05 4.05
##     [226] 4.05 4.00 4.10 4.05 4.05 4.05 4.05 4.05 4.40 4.10 4.05 4.05 4.05 4.05 4.05
##     [241] 4.05 4.05 4.05 4.00 4.05 4.05 4.10 4.00 4.00 4.00
```

#(5) Plotting of quantitative variable:

```
barplot(table(data$rating),
        main = "Distribution of Ratings",
        xlab = "Rating",
        ylab = "Frequency")
```

Distribution of Ratings



#(5) Scatter plot

```
plot(data$rating,data$year ,main = "Scatter Plot of Year vs Rating",  
      xlab = "Rating",  
      ylab = "Year")
```

