HW1

```r
library(ggplot2)
library(tidyverse)
library(stringr)
library(tinytex)
```

```r
drama <- read.csv("data/kdrama.csv")
summary(drama)
```

```
##      Name            Aired.Date        Year.of.release
##  Length:250         Length:250         Min.   :2003   
##  Class :character   Class :character   1st Qu.:2017   
##  Mode  :character   Mode  :character   Median :2019   
##                                        Mean   :2018   
##                                        3rd Qu.:2021   
##                                        Max.   :2022   
##  Original.Network     Aired.On         Number.of.Episodes
##  Length:250         Length:250         Min.   :  1.00    
##  Class :character   Class :character   1st Qu.: 16.00    
##  Mode  :character   Mode  :character   Median : 16.00    
##                                        Mean   : 19.06    
##                                        3rd Qu.: 20.00    
##                                        Max.   :133.00    
##    Duration         Content.Rating         Rating     
##  Length:250         Length:250         Min.   :8.300  
##  Class :character   Class :character   1st Qu.:8.300  
##  Mode  :character   Mode  :character   Median :8.500  
##                                        Mean   :8.534  
##                                        3rd Qu.:8.700  
##                                        Max.   :9.200  
##    Synopsis            Genre               Tags          
##  Length:250         Length:250         Length:250        
##  Class :character   Class :character   Class :character  
##  Mode  :character   Mode  :character   Mode  :character  
##                                                          
##                                                          
##                                                          
##    Director         Screenwriter           Cast          
##  Length:250         Length:250         Length:250        
##  Class :character   Class :character   Class :character  
##  Mode  :character   Mode  :character   Mode  :character  
##                                                          
##                                                          
##                                                          
##  Production.companies     Rank          
##  Length:250           Length:250        
##   Class :character     Class :character
##   Mode  :character     Mode  :character
##
##
##
```

```r
unique(drama) # 250
```

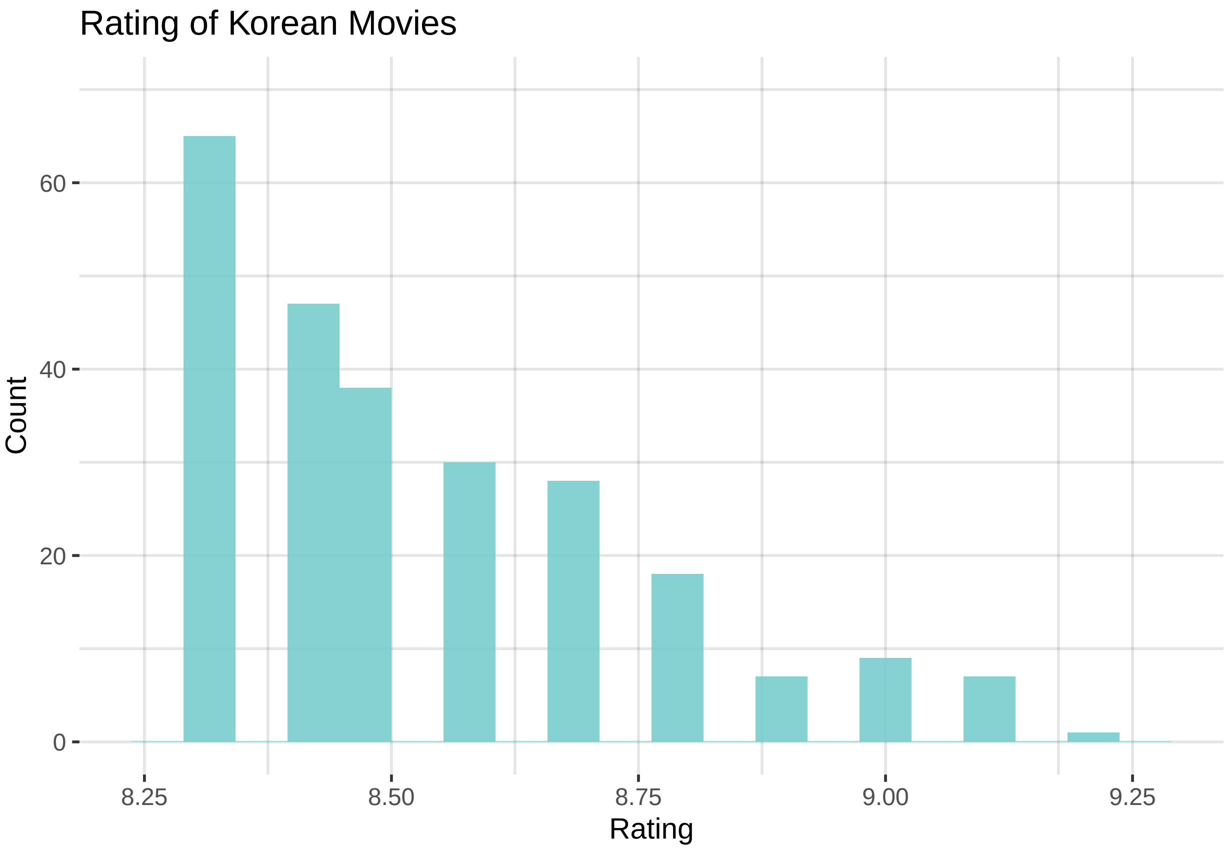
```r
colnames(drama) # 17
```

```
##  [1] "Name"                 "Aired.Date"
##  [3] "Year.of.release"      "Original.Network"
##  [5] "Aired.On"             "Number.of.Episodes"
##  [7] "Duration"             "Content.Rating"
##  [9] "Rating"               "Synopsis"
## [11] "Genre"                "Tags"
## [13] "Director"             "Screenwriter"
## [15] "Cast"                 "Production.companies"
## [17] "Rank"
```

```r
summary(drama$Number.of.Episodes) #19.06
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1.00   16.00   16.00   19.06   20.00  133.00
```

```r
ggplot(drama) +
  theme(
    panel.background = element_rect(fill = "white"),
    panel.grid.major = element_blank(),
    panel.grid.minor = element_blank()
  )+
  geom_hline(yintercept = c(0,10,20,30,40,50,60,70), alpha =.1)+
  geom_vline(xintercept = c(8.25, 8.375, 8.5, 8.625, 8.75, 8.875, 9.0,9.175, 9.25), alpha=.1)+
    geom_histogram(aes(x = Rating), bins = 20, fill = "darkslategray3", alpha =.9) +
  labs(y = "Count",
       title = "Rating of Korean Movies")
```



```
nrow(drama %>%
  filter(drama$Rating >=9))
```

```
## [1] 17
```

```
drama <- drama %>%
  rename(Year = Year.of.release)
```

```
nrow(drama %>%
     filter(drama$Year>=2020 & drama$Year<=2022))
```

```
## [1] 106
```

```
glimpse(drama$Duration) #chr
```

```
##  chr [1:250] "52 min." "1 hr. 10 min." "1 hr. 30 min." ...
```

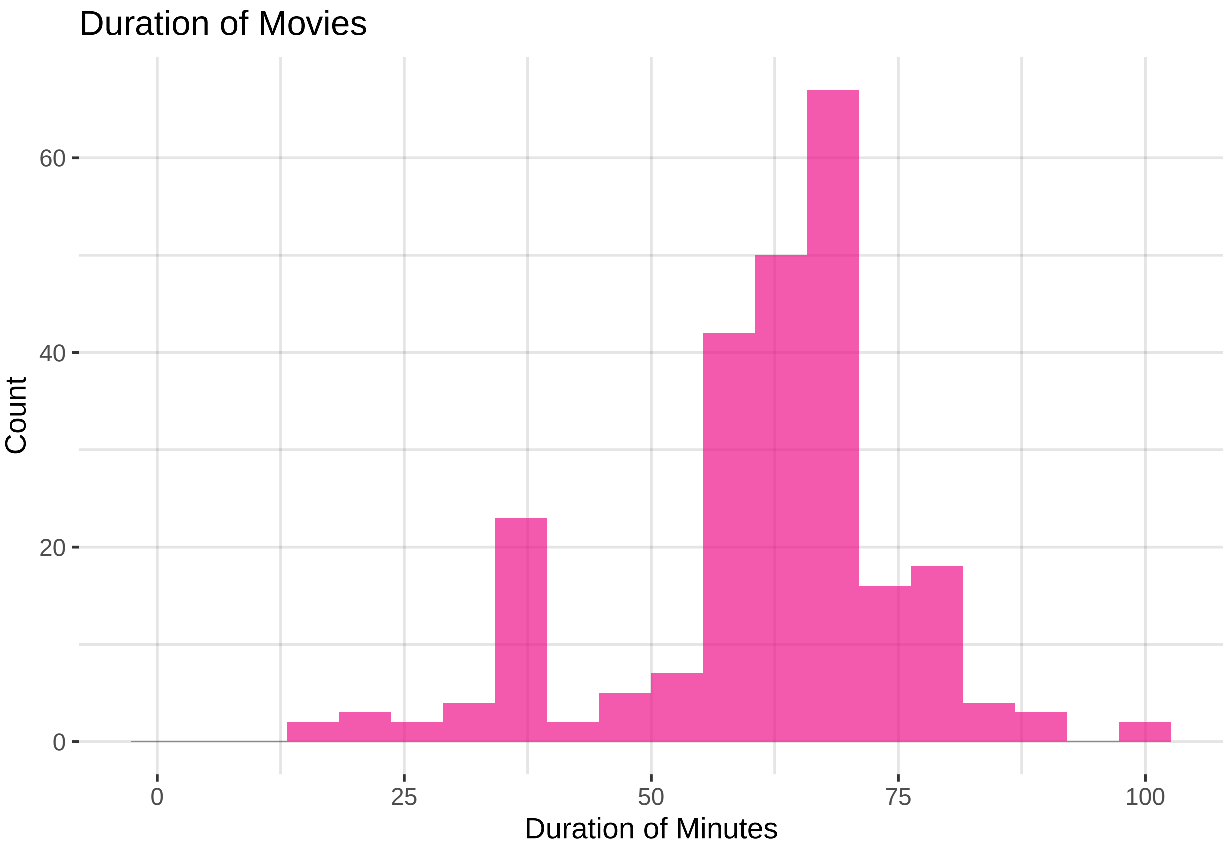
```
unique(drama$Duration)
```

```
##  [1] "52 min."        "1 hr. 10 min." "1 hr. 30 min." "1 hr. 40 min."
##  [5] "1 hr. 17 min." "40 min."        "1 hr. 32 min." "1 hr. 20 min."
##  [9] "1 hr. 3 min."  "1 hr. 15 min." "1 hr. 25 min." "1 hr. 5 min."
## [13] "45 min."        "60 min."        "1 hr. 22 min." "51 min."
## [17] "50 min."        "1 hr. 7 min."  "30 min."        "1 hr. 21 min."
## [21] "1 hr. 2 min."   "35 min."       "55 min."        "1 hr. 11 min."
## [25] "1 hr. 6 min."   "58 min."       "1 hr. 8 min."   "1 hr. 13 min."
## [29] "1 hr. 9 min."   "54 min."       "25 min."        "33 min."
## [33] "1 hr. 4 min."   "44 min."       " 1 hr. 3 min." "20 min."
## [37] "15 min."        "24 min."
```

```
min_func <- function(x){
  h <- as.numeric(str_extract(x,"\\d+(?=\\s*(?:hr)\\b)")); h[is.na(h)] <- 0
  m <- as.numeric(str_extract(x,"\\d+(?=\\s*(?:min)\\b)")); m[is.na(m)] <- 0
  h*60 +m
}

drama <- drama %>%
  mutate(duration_min = min_func(Duration))

ggplot(drama)+
  theme(
    panel.background = element_rect(fill = "white"),
    panel.grid.major = element_blank(),
    panel.grid.minor = element_blank()
  ) +
  geom_hline(yintercept = c(0,10,20,30,40,50,60), alpha =.1)+
  geom_vline(xintercept = c(0,12.5,25,37.5,50,62.5,75,87.5,100),alpha=.1)+
  geom_histogram(aes(x=duration_min), bins = 20, fill = "deeppink2", alpha =.7)+
  labs(y = "Count",
       x = "Duration of Minutes",
       title = "Duration of Movies")
```



```
netflix <- drama %>%
  filter(str_detect(Original.Network, "Netflix"))
```

```
summary(netflix$Rating) # 8.662
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   8.300   8.500   8.700   8.662   8.825   9.200
```