```r
library(readr)
```

```
## Warning: package 'readr' was built under R version 4.4.3
```

```r
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.4.3
```

```
##
## 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
```

```r
library(stringr)

#q1: download and read the number of rows
library(readr)
kdrama <- read_csv("C:/Users/barna/Downloads/kdrama.csv")
```

```
## Rows: 250 Columns: 17
```

```
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (14): Name, Aired Date, Original Network, Aired On, Duration, Content Ra...
## dbl  (3): Year of release, Number of Episodes, 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.
```

```r
View(kdrama)

nrow(kdrama)
```

```
## [1] 250
```

```r
#Total 250 rows are present


#q2: List of Variables
names(kdrama)
```

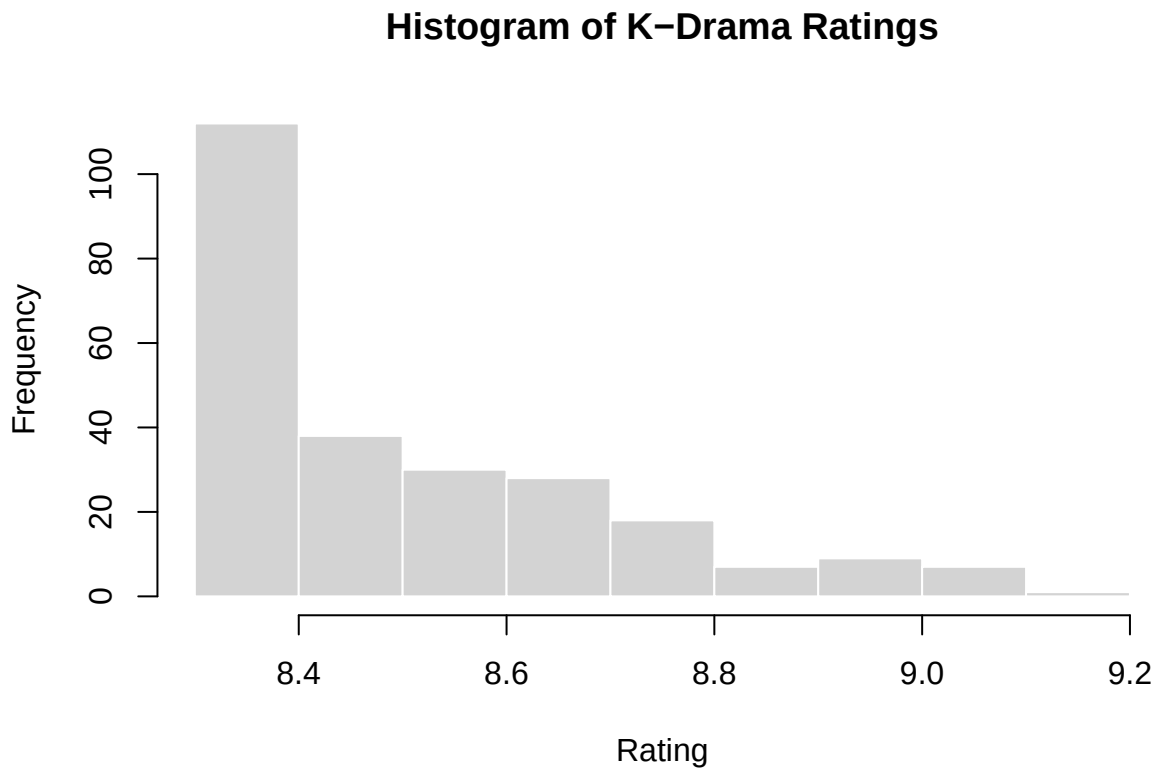
```
##  [1] "Name"                "Aired Date"          "Year of release"
##  [4] "Original Network"    "Aired On"            "Number of Episodes"
##  [7] "Duration"            "Content Rating"      "Rating"
## [10] "Synopsis"            "Genre"               "Tags"
## [13] "Director"            "Screenwriter"        "Cast"
## [16] "Production companies" "Rank"
```

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

```r
#q3. mean value of total number of episodes for all the kdramas
mean_episodes <- mean(na.omit(kdrama$`Number of Episodes`))
mean_episodes
```

```
## [1] 19.064
```

```r
#[1] 19.064


#q4. histogram of the shows rating

hist(kdrama$Rating,
     main = "Histogram of K-Drama Ratings",
     xlab = "Rating",
     border = "white")
```




```r
#q5:rating higher than 9 points
sum(na.omit(kdrama$`Rating`))
```

```
## [1] 2133.5
```

```r
#[1 2133.5
sum(na.omit(kdrama$`Rating` > 9))
```

```
## [1] 8
```

```r
#Ans: 8
```

```r
#q6. Rename variable Year.of.release to simply Year without creating a new variable

library(dplyr)

kdrama <- kdrama %>%
  rename(Year = `Year of release`)
#ANother way to create, other than tidyverse, as we know year is in third column,
colnames(kdrama)[3] <- "Year"

#q7. dataset were released in 2020-2022
kdrama %>%
  filter(between(Year, 2020, 2022)) %>%
  nrow()
```

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

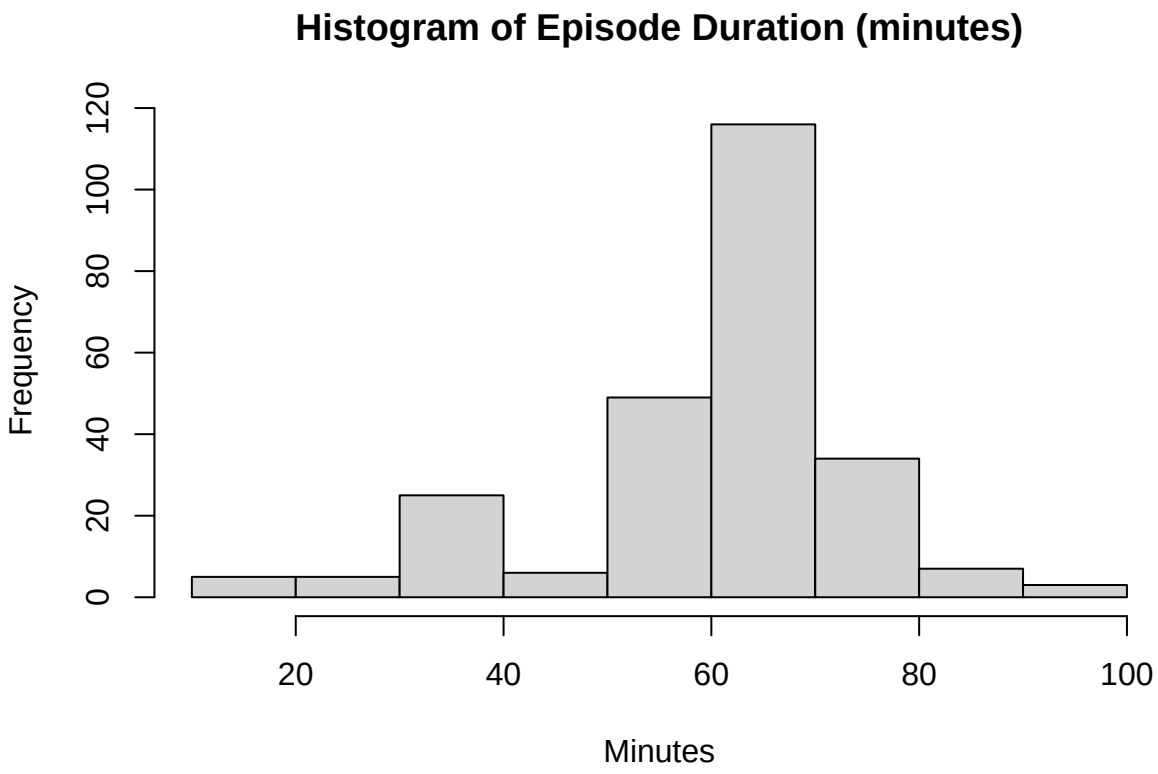
```r
#couldhave donw like sum(na.omit(kdrama$Year) >= 2020 and na.omit(kdrama$Year) <= 2022) but and symble
#Ans: 106


#q8. type of variable is Duration
class(kdrama$`Duration`)
```

```
## [1] "character"
```

```r
#Ans:  "character"


#q9. Recode variable Duration to a numerical variable measuring duration in minutes. plot histogram of

#tidyverse way:

library(dplyr)
library(stringr)
class(kdrama$`Duration`)
```

```
## [1] "character"
```

```r
kdrama <- kdrama %>%
  mutate(Duration_min =
           coalesce(as.numeric(str_extract(Duration, "\\d+(?=\\s*hr)")), 0) * 60 +
           coalesce(as.numeric(str_extract(Duration, "\\d+(?=\\s*min)")), 0))

hist(na.omit(kdrama$Duration_min),
     main = "Histogram of Episode Duration (minutes)",
     xlab = "Minutes")
```




```
#function way: though not sure, took online help..... :(
Duration_to_minutes <- function(x) {
  hrs <- ifelse(grepl("hr", x, ignore.case = TRUE),
                as.numeric(sub(".*?(\\d+)\\s*hr.*", "\\1", x)), 0)
  mins <- ifelse(grepl("min", x, ignore.case = TRUE),
                 as.numeric(sub(".*?(\\d+)\\s*min.*", "\\1", x)), 0)
  hrs[is.na(hrs)]  <- 0
  mins[is.na(mins)] <- 0
  60*hrs + mins
}



#q10. dataset that will include shows with Original.Network being Netflix.
library(dplyr)
library(stringr)

netflix_exact <- kdrama %>%
  filter(str_trim(`Original Network`) == "Netflix")

nrow(netflix_exact)
```

```
## [1] 12
```

```
#Ans: 12
```

```r
#q11. What is the average rating score for the shows that have Netflix as an Original Network.
mean(na.omit(netflix_exact$`Rating`))
```

```
## [1] 8.65
```

```r
#or,
mean(netflix_exact$Rating, na.rm = TRUE)
```

```
## [1] 8.65
```

```r
#Ans: 8.65
```