Assignment 1

David

2024-08-30

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.3.3
```

```
## ── 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 1.9.3     ✔ 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
```

```
##Reading the data
read.csv("c://Users/akind/OneDrive - Florida State University/Floridal/Classes/Political Analysis/R_class/data/kdrama (1).csv")-> movie
##Number of K-drama
nrow(movie)
```

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

```
##Just to view the data to be sure there is a variable "number of episodes"

names(movie)
```

```
##  [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"
```

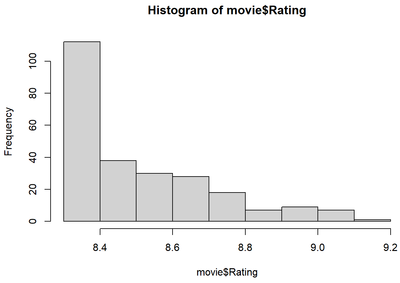
```
##OR
colnames(movie)
```

```
##  [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"
```

```
##To determent the mean of "Number of Episodes"
mean(movie$Number.of.Episodes)
```

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

```
##Histogram plot of Rating

hist(movie$Rating)
```




```
###Showing number of rating more than 9
colnames(movie)
```

```
##  [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"
```

```
##showing the rating variable with higher than 9 points
above9 <- movie%>%
  filter(movie$Rating >= 9)


##showing the number of shows with rating higher than 9
length(above9$Rating)
```

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

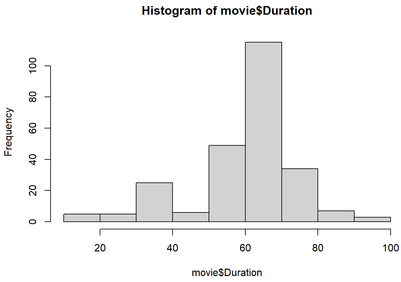
```
##renaming the Year of release viarable

movie <- movie%>%
  rename("Year" = Year.of.release )



##Number of shows between 2020-2022

released <- movie%>%
  filter(Year >= 2020, Year<= 2022)
length(released$Year)
```

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

```
##type of duration variable

class(movie$Duration)
```

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

```
##Changing the Duration variable to minutes

mins <- movie$Duration

##removing "hr"
mins <- str_remove(mins, "hr\\.")

##removing "mins"
mins <- str_remove(mins, "min\\.")

##substituting "1" with "60"
mins <- gsub("\\b1\\b", "60", mins)

##splitting the string and converting it to a numeric vector

sum_numbers <- function(x){
  numbers<- as.numeric(unlist(strsplit(x, "\\s+")))
  sum(numbers)
}
##summing the string
sums <- sapply(mins, sum_numbers)

##Bringing the new data into the old
movie <- movie %>%
  mutate(Duration = sums)

##viwing to be sure it worked
view(movie)

##histogram of the new viarable

hist(movie$Duration)
```




```
##shows with Original.Network being Netflix

netflix<- movie%>%
  filter(Original.Network == "Netflix")

##average rating score netflix

mean(netflix$Rating)
```

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

R Markdown