PS1

Taylor Hill

2024-08-29

1. Reading in the data

```
library(tidyverse)
library(pander)
library(reshape2)
kdrama <- read_csv("kdrama.csv")
```

2. Returning a 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"
```

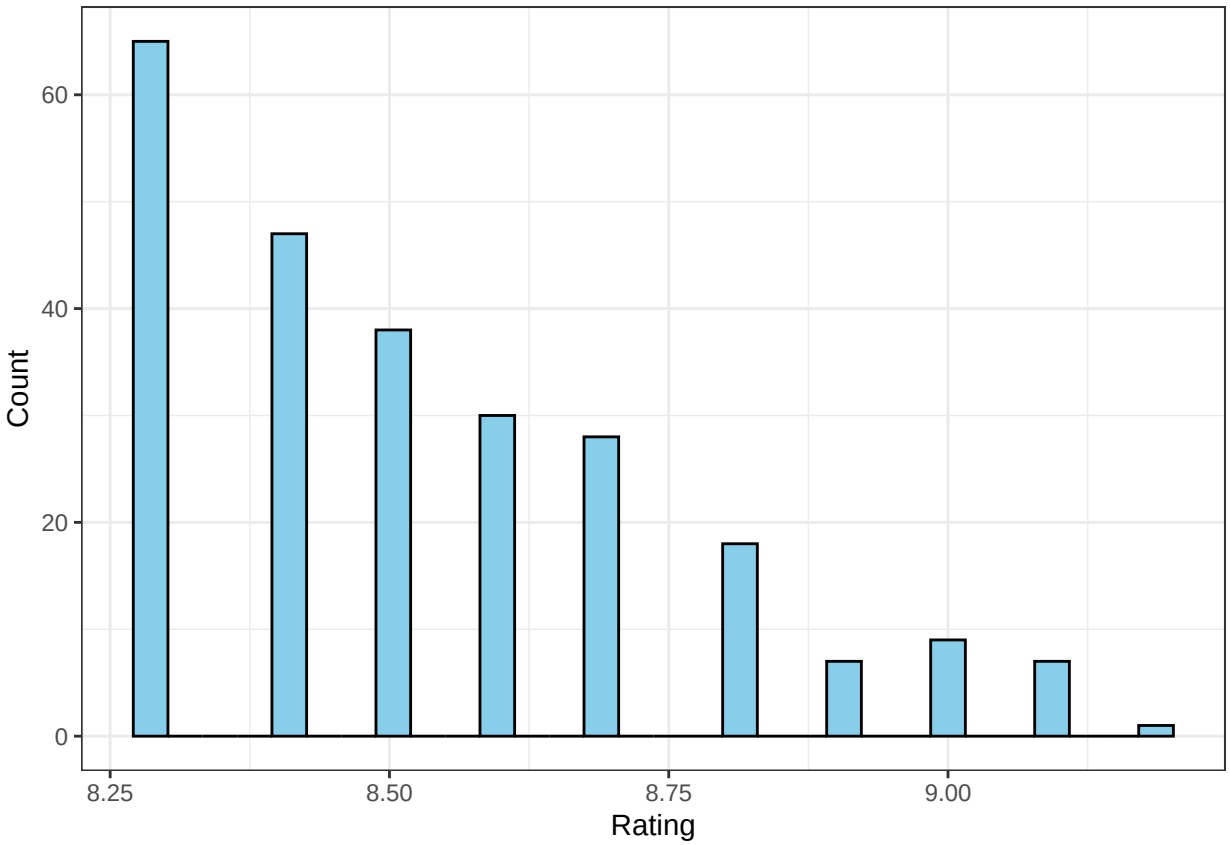
3. Mean number of episodes

```
mean(kdrama$"Number of Episodes") |> pander()
```

19.06

4. Histogram that shows rating

```
ggplot(kdrama, aes(x=Rating))+
  geom_histogram(color="black", fill="skyblue")+
  theme_bw()+
  labs(y="Count")
```



5. Rating higher than a 9

```
table(kdrama$Rating) |>
  pander()
```

8.3	8.4	8.5	8.6	8.7	8.8	8.9	9	9.1	9.2
65	47	38	30	28	18	7	9	7	1

6. Rename variable

```
kdrama<-kdrama |>
  rename(Year="Year of release")
```

7. Numher of shows from 2020-2022

```
number_shows<-kdrama |>
  count(Year) |>
  filter(Year>=2020,
         Year<=2022)
sum(number_shows$n)
```

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

8. Type of variable

```
class(kdrama$Duration)
```

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

9. Recode duration to numerical

```
kdrama<-kdrama |>
  mutate(colsplit(kdrama$Duration, "hr.", names=c("hour", "minute")))

kdrama<-kdrama |>
  mutate(hour=parse_number(hour), minute=parse_number(minute),
         minute=case_when(
          is.na(minute)~ 0,
          TRUE~minute),
         hour=case_when(
           hour==1~60,
           TRUE~hour),
         Duration=hour+minute)
```

10. Network shows dataset

```
net_kdrama<-kdrama |>
  filter(`Original Network`=="Netflix")
```

11. Average rating for Netflix shows

```
mean(net_kdrama$Rating) |> pander()
```

8.65