IVfinalcode.R

shriyakumbhoje

2023-12-12

```
#installing packages
if(! require(ggplot2)) install.packages("ggplot2")
if(! require(tidyverse)) install.packages("tidyverse")
if(! require(tidygraph)) install.packages("tidygraph")
if(! require(visNetwork)) install.packages("visNetwork")
if(! require(igraph)) install.packages("igraph")
if(! require(wordcloud)) install.packages("wordcloud")

#Attaching packages
library(ggplot2)
library(tidyverse)
library(tidygraph)  #network analysis
library(visNetwork) #interactive network visualization
library(igraph)
#for 4
library('RColorBrewer')
library('wordcloud')
library('tm')

data <- read.csv("GibhliS.csv", header = TRUE, sep = ",", skip = 2, col.names = c("Year", "Title", "Directors",
"Screenwriters", "Producer", "Music", "Commercial premiere", "Running Time", "Rotten Tomatoes"))
getwd()
```

```
## [1] "/Users/shriyakumbhoje/Desktop/IV /report "
```

```
head(data)
```

```
##   Year                   Title      Directors   Screenwriters       Producer         Music
## 1 1986       Castle in the sky  Hayao Miyazaki Hayao Miyazaki  Isao Takahata  Joe Hisaishi
## 2 1988  Grave of the Fireflies  Isao Takahata   Isao Takahata      Toru Hara Michio Mamiya
## 3 1988     My Neighbour Totoro Hayao Miyazaki Hayao Miyazaki       Toru Hara Joe Hisaishi
## 4 1989 Kiki's Delivery Service Hayao Miyazaki Hayao Miyazaki  Hayao Miyazaki Joe Hisaishi
## 5 1991          Only Yesterday  Isao Takahata   Isao Takahata  Toshio Suzuki  Masaru Hoshi
## 6 1992             Porco Rosso Hayao Miyazaki Hayao Miyazaki   Toshio Suzuki  Joe Hisaishi
##   Commercial.premiere Running.Time Rotten.Tomatoes
## 1              2-8-86        2h 5m             96%
## 2             16-4-88       1h 28m            100%
## 3             16-4-88       1h 26m             93%
## 4             29-7-89       1h 43m             98%
## 5             20-7-91       1h 59m            100%
## 6             18-7-92       1h 33m             96%
```

```
#summarising the data for data transformation
summary(data)
```

```
##       Year         Title            Directors         Screenwriters        Producer
##  Min.   :1986   Length:21          Length:21          Length:21          Length:21
##  1st Qu.:1992   Class :character   Class :character   Class :character   Class :character
##  Median :2001   Mode  :character   Mode  :character   Mode  :character   Mode  :character
##  Mean   :2001
##  3rd Qu.:2010
##  Max.   :2023
##     Music           Commercial.premiere Running.Time       Rotten.Tomatoes
##  Length:21          Length:21           Length:21          Length:21
##  Class :character   Class :character    Class :character   Class :character
##  Mode  :character   Mode  :character    Mode  :character   Mode  :character
##
##
##
```

```
table(data$Directors) #this will show the frequency of each unique value in the specified variable
```

```
##
##       Goro Miyazaki        Goro Miyazaki        Hayao Miyazaki Hiromasa Yonebayashi
##                   1                    1                    10                    2
##     Hiroyuki Morita        Isao Takahata      Yoshifumi Kondo
##                   1                    5                    1
```

```
#PARSING
#converting Commercial.premiere column into class: Date
data$Commercial.premiere <- as.Date(data$Commercial.premiere, format = "%d-%m-%Y")
class(data$Commercial.premiere)
```

```
## [1] "Date"
```

```
#converting Running.time coulmn "Xh Ym" format to  minutes
print(data$Running.Time)
```

```
##  [1] "2h 5m"  "1h 28m" "1h 26m" "1h 43m" "1h 59m" "1h 33m" "1h 59m" "1h 51m" "2h 13m"
## [10] "1h 43m" "2h 5m"  "1h 15m" "1h 59m" "1h 55m" "1h 41m" "1h 34m" "1h 31m" "2h 17m"
## [19] "2h 6m"  "1h 43m" "2h 4m"
```

```
convert_to_minutes <- function(time_string) {
  # Extract hours and minutes using regular expression
  time_components <- as.numeric(regmatches(time_string, gregexpr("\\d+", time_string))[[1]])
  total_minutes <- time_components[1] * 60 + time_components[2]
  return(total_minutes)
}

data$Running.Time <- sapply(data$Running.Time,convert_to_minutes)  #applying the function to the entire column
class(data$Running.Time)
```

```
## [1] "numeric"
```

```
print(data$Running.Time)
```

```
##  [1] 125  88  86 103 119  93 119 111 133 103 125  75 119 115 101  94  91 137 126 103 124
```

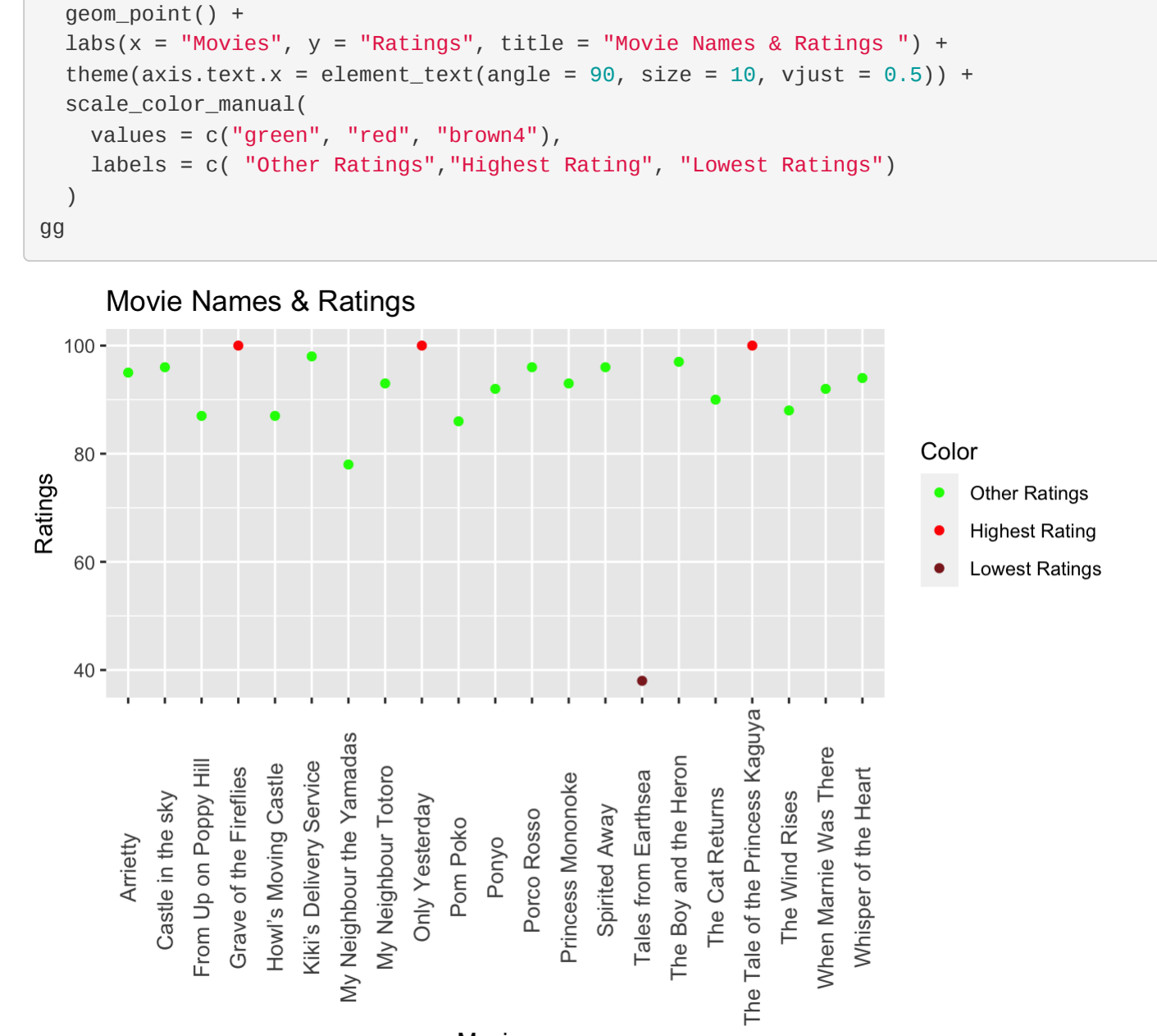
```
#converting column Rotten.Tomato ratings into numeric
data$Rotten.Tomatoes<- as.numeric(gsub("%", "", data$Rotten.Tomatoes))

#1 Calculate max and min (ggplot) ---------------------------------------------------------------
#movie ratings to check the movie with the highest and lowest rating
#scatterplot
#aes() function in ggplot binds two variables you want to plot, geom_point used to create scatterplots

max_rating <- max(data$Rotten.Tomatoes)
min_rating <- min(data$Rotten.Tomatoes)

#Create a new column for color based on ratings
data$Color <- ifelse(data$Rotten.Tomatoes == max_rating, "green",
                     ifelse(data$Rotten.Tomatoes == min_rating, "red", "brown4"))

# Plotting
gg <- ggplot(data, aes(Title, Rotten.Tomatoes, color = Color)) +
  geom_point() +
  labs(x = "Movies", y = "Ratings", title = "Movie Names & Ratings ") +
  theme(axis.text.x = element_text(angle = 90, size = 10, vjust = 0.5)) +
  scale_color_manual(
    values = c("green", "red", "brown4"),
    labels = c( "Other Ratings","Highest Rating", "Lowest Ratings")
  )
gg
```



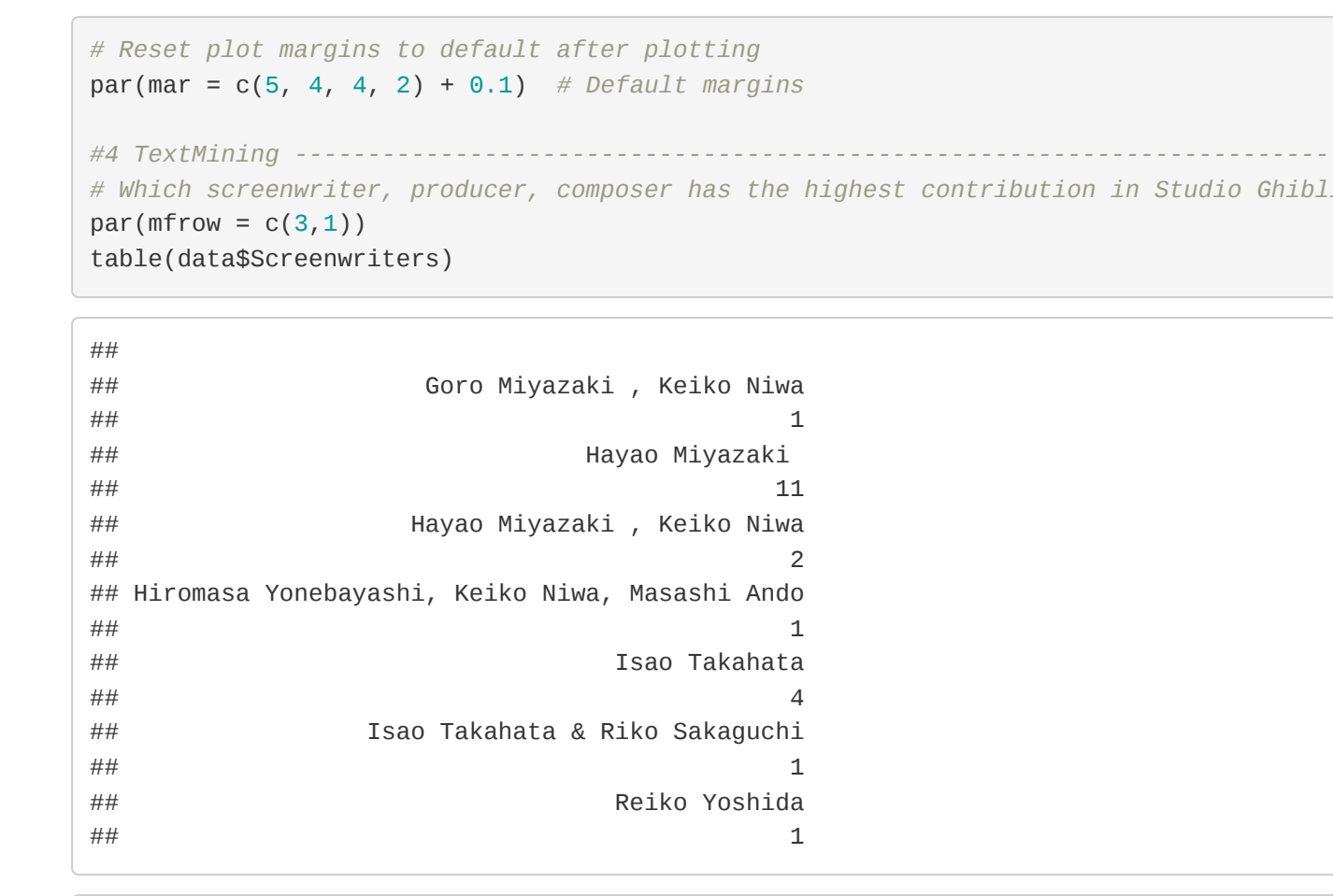

```
#2 Animation -------------------------------------------------------------------
#Is there a trend in rating over the years
library(timetk)
library(gganimate)
df2 <- data_frame(
  Release.Date= data$Year,
  Rating= data$Rotten.Tomatoes
)
df2 %>%
  ggplot(aes(x= Rating,y=Release.Date)) +
  geom_line() +
  labs(x='Date',y='Rotten tomatoes Rating', title='Rating') +
  transition_reveal(Rating)
```

```
## Warning: No renderer available. Please install the gifski, av, or magick package to create
## animated output
```

```
## NULL
```

```
#3 TimeSeries -------------------------------------------------------------------
#Is there any association between the run time of a movie and its rating over the years
par(mfrow = c(2,1))
par(mar = c(4, 4, 2, 2))

plot(data$Year, data$Running.Time, type = "l",lwd=2,ylim = c(1, 140), xlim = c(1988, 2023),
     xlab = "Year", ylab = "Running time", main = "Running time Trend",
     col = "blue")
plot(data$Year, data$Rotten.Tomatoes, type = "l",lwd=2,ylim = c(1, 120), xlim = c(1988, 2023),
     xlab = "Year", ylab = "Rotten Tomatoes", main = "Rotten Tomatoes Ratings Trend",
     col = "orange")
```




```
# Reset plot margins to default after plotting
par(mar = c(5, 4, 4, 2) + 0.1)  # Default margins

#4 TextMining -------------------------------------------------------------------
# Which screenwriter, producer, composer has the highest contribution in Studio Ghibli production?
par(mfrow = c(3,1))
table(data$Screenwriters)
```

```
##
##                     Goro Miyazaki , Keiko Niwa
##                                              1
##                                 Hayao Miyazaki
##                                             11
##                    Hayao Miyazaki , Keiko Niwa
##                                              2
## Hiromasa Yonebayashi, Keiko Niwa, Masashi Ando
##                                              1
##                                  Isao Takahata
##                                              4
##                   Isao Takahata & Riko Sakaguchi
##                                              1
##                                   Reiko Yoshida
##                                              1
```

```
table(data$Producer)
```

```
##
##                   Hayao Miyazaki                    Isao Takahata
##                                1                                1
## Nozomu Takahashi , Toshio Suzuki                        Toru Hara
##                                1                                2
##                    Toshio Suzuki               Yoshiaki Nishimura
##                               14                                2
```

```
table(data$Music)
```

```
##
##         Akiko Yano       Cecile Corbel        Joe Hisaishi        Masaru Hoshi
##                  1                   1                  11                   1
##      Michio Mamiya     Satoshi Takebe Shang Shang Typhoon Takatsugu Muramatsu
##                  1                   1                   1                   1
##   Tamiya Terashima            Yuji Nomi
##                  1                   2
```

```
data$Screenwriters <- trimws(data$Screenwriters)
# Create the frequency table and plot the word cloud
freq1 <- table(data$Screenwriters)
wordcloud(words = names(freq1), freq = freq1, min.freq = 0, scale=c(2,1), col = "#301934")
```

```
## Warning in wordcloud(words = names(freq1), freq = freq1, min.freq = 0, scale = c(2, : Hayao
## Miyazaki , Keiko Niwa could not be fit on page. It will not be plotted.
```

```
## Warning in wordcloud(words = names(freq1), freq = freq1, min.freq = 0, scale = c(2, :
## Hiromasa Yonebayashi, Keiko Niwa, Masashi Ando could not be fit on page. It will not be
## plotted.
```

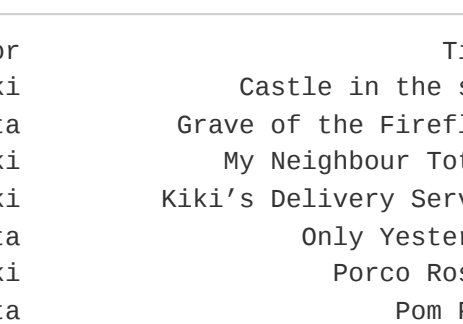
```
## Warning in wordcloud(words = names(freq1), freq = freq1, min.freq = 0, scale = c(2, : Isao
## Takahata & Riko Sakaguchi could not be fit on page. It will not be plotted.
```

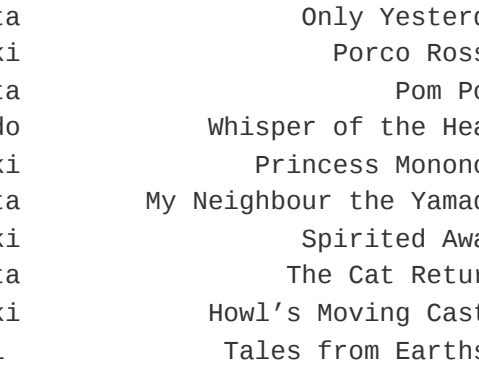
```
## Warning in wordcloud(words = names(freq1), freq = freq1, min.freq = 0, scale = c(2, : Hayao
## Miyazaki could not be fit on page. It will not be plotted.
```

```
data$Producer <- trimws(data$Producer)
freq2 <- table(data$Producer)
wordcloud(words = names(freq2), freq = freq2, min.freq = 0, scale=c(2,1), col = "#AA336A")
```

```
## Warning in wordcloud(words = names(freq2), freq = freq2, min.freq = 0, scale = c(2, :
## Nozomu Takahashi , Toshio Suzuki could not be fit on page. It will not be plotted.
```

```
data$Music <- trimws(data$Music)
freq3 <- table(data$Music)
wordcloud(words = names(freq3), freq = freq3, min.freq = 0, scale=c(2,1), col = "#66023C")
```




```
#5Interactive Network ---------------------------------------------------------------
#Which Director has directed the most movies for Studio Ghibli?
data_new <- data.frame(Director = data$Directors,Title = data$Title)
data_new
```

```
##                Director                             Title
## 1        Hayao Miyazaki                 Castle in the sky
## 2         Isao Takahata            Grave of the Fireflies
## 3        Hayao Miyazaki               My Neighbour Totoro
## 4        Hayao Miyazaki           Kiki's Delivery Service
## 5         Isao Takahata                    Only Yesterday
## 6        Hayao Miyazaki                       Porco Rosso
## 7         Isao Takahata                          Pom Poko
## 8       Yoshifumi Kondo               Whisper of the Heart
## 9        Hayao Miyazaki                 Princess Mononoke
## 10        Isao Takahata         My Neighbour the Yamadas
## 11       Hayao Miyazaki                     Spirited Away
## 12      Hiroyuki Morita                   The Cat Returns
## 13       Hayao Miyazaki               Howl's Moving Castle
## 14       Goro Miyazaki                Tales from Earthsea
## 15       Hayao Miyazaki                             Ponyo
## 16 Hiromasa Yonebayashi                          Arrietty
## 17       Goro Miyazaki               From Up on Poppy Hill
## 18        Isao Takahata The Tale of the Princess Kaguya
## 19       Hayao Miyazaki                    The Wind Rises
## 20 Hiromasa Yonebayashi           When Marnie Was There
## 21       Hayao Miyazaki              The Boy and the Heron
```

```
#create a "tbl_graph" object
network <- data_new%>%
  as_tbl_graph()


network %>%
  activate(nodes) %>%
  mutate(degree= centrality_degree()) %>%
  as_tibble() %>%
  arrange(desc(degree))
```

```
## # A tibble: 28 × 2
##    name                          degree
##    <chr>                          <dbl>
##  1 "Hayao Miyazaki"                  10
##  2 "Isao Takahata"                    5
##  3 "Hiromasa Yonebayashi"             2
##  4 "Yoshifumi Kondo"                  1
##  5 "Hiroyuki Morita"                  1
##  6 "Goro Miyazaki "                   1
##  7 " Goro Miyazaki "                  1
##  8 "Castle in the sky "               0
##  9 "Grave of the Fireflies"           0
## 10 "My Neighbour Totoro"              0
## # ℹ 18 more rows
```

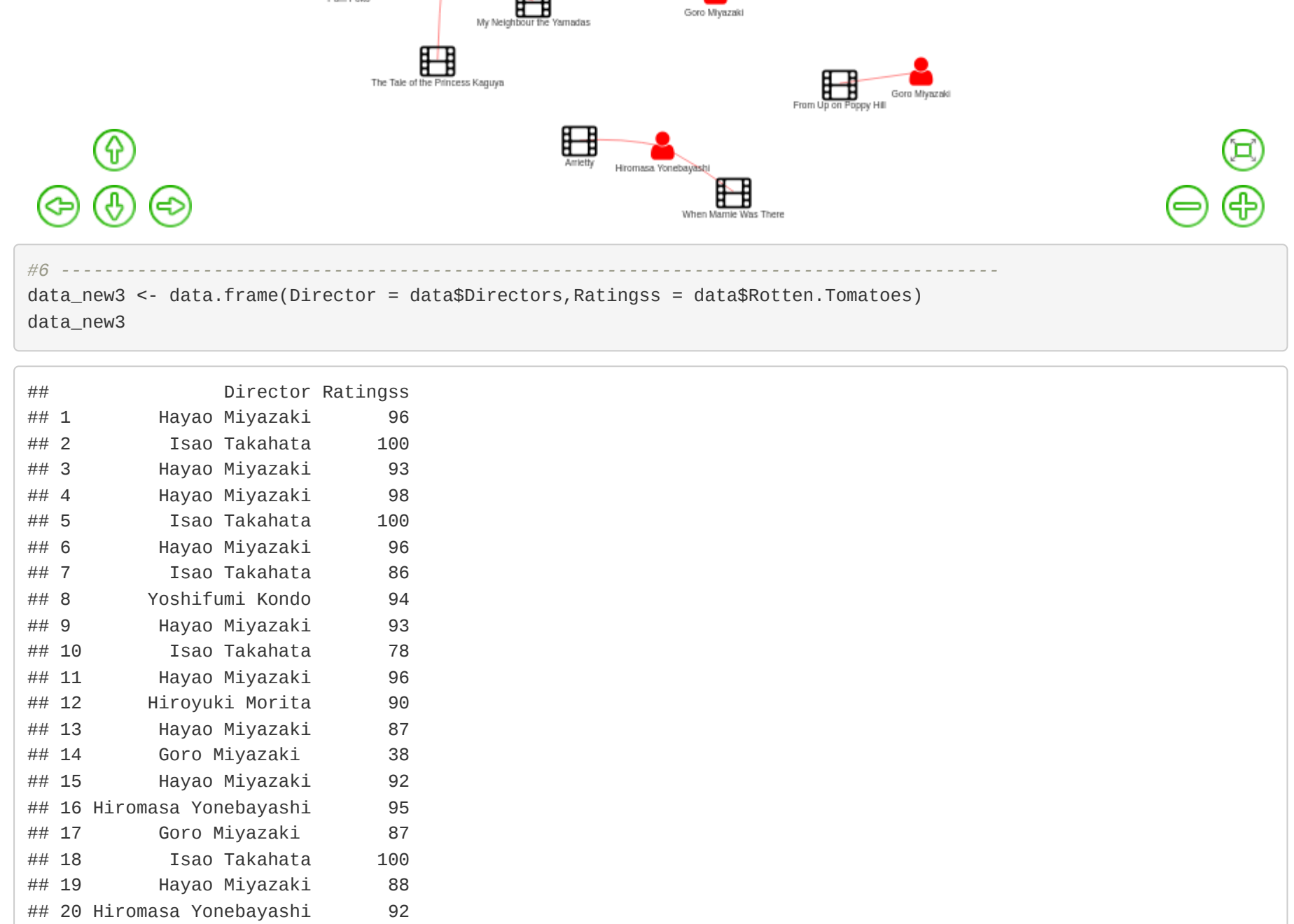
```
#Interactive Network

vis_network <- network %>%
  mutate(group = if_else(condition = name %in% unique(data_new$Director),
                         true="Director",
                         false="Title")) %>%
  toVisNetworkData()
#interactive part

visNetwork(nodes = vis_network$nodes, edges = vis_network$edges,
           width = "100%", height = "600px", main = "The Ghibli Movie Network") %>%
  visLayout(randomSeed = 1000) %>%
  addFontAwesome() %>%
  visGroups(groupname = "Title", shape = "icon",
            icon = list(code = "f008", color = "black")) %>%
  visGroups(groupname = "Director", shape = "icon",
            icon = list(code = "f007", color = "red")) %>%
  visOptions(highlightNearest = list(enabled = TRUE, hover = TRUE), nodesIdSelection = TRUE) %>%
  visInteraction(navigationButtons = TRUE)
```

The Ghibli Movie Network

Select by id



```
#6 ------------------------------------------------------------------------------------------
data_new3 <- data.frame(Director = data$Directors,Ratingss = data$Rotten.Tomatoes)
data_new3
```

```
##                Director Ratingss
## 1        Hayao Miyazaki       96
## 2         Isao Takahata      100
## 3        Hayao Miyazaki       93
## 4        Hayao Miyazaki       98
## 5         Isao Takahata      100
## 6        Hayao Miyazaki       96
## 7         Isao Takahata       86
## 8       Yoshifumi Kondo       94
## 9        Hayao Miyazaki       93
## 10        Isao Takahata       78
## 11       Hayao Miyazaki       96
## 12      Hiroyuki Morita       90
## 13       Hayao Miyazaki       87
## 14       Goro Miyazaki        38
## 15       Hayao Miyazaki       92
## 16 Hiromasa Yonebayashi       95
## 17       Goro Miyazaki        87
## 18        Isao Takahata      100
## 19       Hayao Miyazaki       88
## 20 Hiromasa Yonebayashi       92
## 21       Hayao Miyazaki       97
```

```
#create a "tbl_graph" object
network6 <- data_new3%>%
  as_tbl_graph()


network6 %>%
  activate(nodes) %>%
  mutate(degree= centrality_degree()) %>%
  as_tibble() %>%
  arrange(desc(degree))
```

```
## # A tibble: 21 × 2
##    name                          degree
##    <chr>                          <dbl>
##  1 "Hayao Miyazaki"                  10
##  2 "Isao Takahata"                    5
##  3 "Hiromasa Yonebayashi"             2
##  4 "Yoshifumi Kondo"                  1
##  5 "Hiroyuki Morita"                  1
##  6 "Goro Miyazaki "                   1
##  7 " Goro Miyazaki "                  1
##  8 "96"                               0
##  9 "100"                              0
## 10 "93"                               0
## # ℹ 11 more rows
```

```
#Interactive Network

vis_network1 <- network6 %>%
  mutate(group = if_else(condition = name %in% unique(data_new$Director),
                         true="Director",
                         false="Ratingss")) %>%
  toVisNetworkData()
#interactive part

visNetwork(nodes = vis_network1$nodes, edges = vis_network1$edges,
           width = "100%", height = "600px", main = "The Director-Rating Relation") %>%
  visLayout(randomSeed = 1000) %>%
  addFontAwesome() %>%
  visGroups(groupname = "Ratingss", shape = "circle", color="lightgray") %>%
  visGroups(groupname = "Director", shape = "icon",
            icon = list(code = "f007", color = "lightblue")) %>%
  visOptions(highlightNearest = list(enabled = TRUE, hover = TRUE), nodesIdSelection = TRUE) %>%
  visInteraction(navigationButtons = TRUE)
```

The Director-Rating Relation

Select by id