ANA 515 Assignment 2 - Sean Anggani

Sean Anggani

2023-06-18

R Markdown

Dataset: https://github.com/fivethirtyeight/data/blob/master/bechdel/movies.csv

Section 1 This dataset measures the representation of women in film and other fiction, which is known as the Bechdel test. A movie passes the Bechdel test if they satisfy 3 criterias: 1. At least 2 named women in the movie 2. These 2 women have a conversation with each other 3. The conversation is not about a male character

The data analyzes 1795 movies from 1990 to 2013 and aims to answer what is the ROI when movies have passed the Bechdel test, and also captures other meaningful data such as budget. The file is stored in a CSV that are delimited by commas.

Section 2

```
library(dplyr)
```

```
##
## 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
```

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v forcats   1.0.0     v readr     2.1.4
## v ggplot2   3.4.2     v stringr   1.5.0
## v lubridate 1.9.2     v tibble    3.2.1
## v purrr     1.0.1     v tidyr     1.3.0
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(knitr)
library(bslib)
```

```
##
## Attaching package: 'bslib'
##
## The following object is masked from 'package:utils':
```

```
##
##     page
```

```
download.file(url = "https://raw.githubusercontent.com/fivethirtyeight/data/master/bechdel/movies.csv",
movies<-read.csv("movies.csv", stringsAsFactors = FALSE)
movies_df <- as.data.frame(movies)
```

Section 3

```
clean_movies<-rename(movies, behavior=test)
new_movies<-filter(movies, year>2010)
```

Section 4 This dataframe has

```
nrow(movies_df)
```

```
## [1] 1794
```

rows and

```
ncol(movies_df)
```

```
## [1] 15
```

columns. The names of the columns and a brief description of each are in the table below:

```
library(knitr)
movies_finances <- select(movies_df,budget, domgross, intgross)
columns_summary <- data.frame(
Columns = c(colnames(movies_finances)),
Description = c(
"Budget of making the movie", "Gross domestic revenue", "Gross international revenue")
)

kable(columns_summary, caption = "Movies Finances")
```

Table 1: Movies Finances

Columns	Description
budget	Budget of making the movie
domgross	Gross domestic revenue
intgross	Gross international revenue

Section 5

```
movies_df$domgross <- as.integer(movies_df$domgross)
```

```
## Warning: NAs introduced by coercion
```

```
movies_df$intgross <- as.integer(movies_df$intgross)
```

```
## Warning: NAs introduced by coercion
```

```
## Warning: NAs introduced by coercion to integer range
```

```
movies_budget <- select(movies_df,budget)
movies_domgross <- select(movies_df,domgross)
movies_intgross <- select(movies_df,intgross)

sum_budget<-summary(na.omit(movies_budget))
```

```r
sum_domgross<-summary(na.omit(movies_domgross))
sum_intgross<-summary(na.omit(movies_intgross))

print(sum_budget)
```

```
##      budget
##  Min.   :     7000
##  1st Qu.: 12000000
##  Median : 28000000
##  Mean   : 44826463
##  3rd Qu.: 60000000
##  Max.   :425000000
```

```r
print(sum_domgross)
```

```
##     domgross
##  Min.   :        0
##  1st Qu.: 16311571
##  Median : 42194060
##  Mean   : 69132048
##  3rd Qu.: 93354918
##  Max.   :760507625
```

```r
print(sum_intgross)
```

```
##     intgross
##  Min.   :8.280e+02
##  1st Qu.:2.610e+07
##  Median :7.642e+07
##  Mean   :1.478e+08
##  3rd Qu.:1.896e+08
##  Max.   :1.328e+09
```