Problem Set 1

MGSC 310, Fall 2019, Professor Hersh

Elmer Camargo + Nick Trella

Libraries Needed

```
library("tidyverse")
library("ggplot2")
library("ggthemes")
library('ggridges')
```

Question 1: Getting and Setting Working Directories)

```
getwd()
## [1] "C:/Users/Elmer/Documents/R/Statistical Modeling/PSET1/imdb_dataset"
setwd("C:/Users/Elmer/Documents/R/Statistical Modeling/PSET1/imdb_dataset")
```

Question 2: Reading CSV File)

```
imdb = read.csv("movie_metadata.csv")
```

Question 3: Dimensions

```
dim(imdb)
## [1] 5043   28
```

Question 4: Variable/Column Names

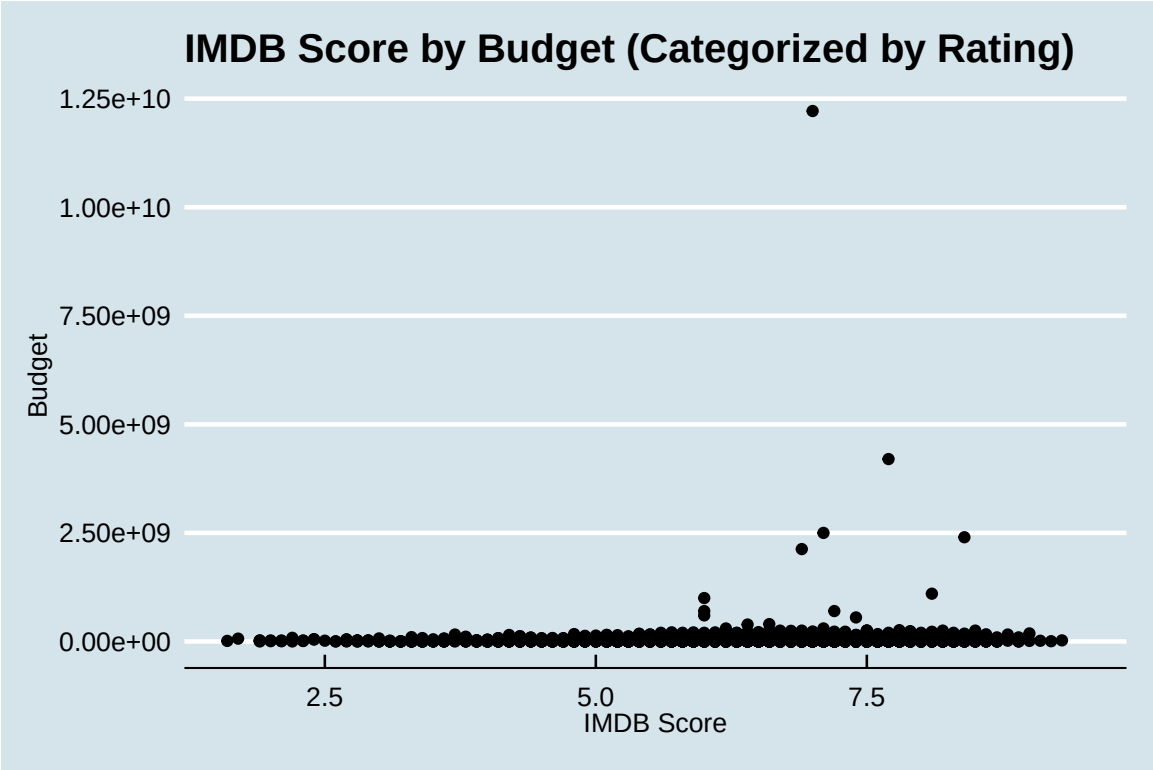
```
names(imdb)
##  [1] "color"                     "director_name"
##  [3] "num_critic_for_reviews"    "duration"
##  [5] "director_facebook_likes"   "actor_3_facebook_likes"
##  [7] "actor_2_name"              "actor_1_facebook_likes"
##  [9] "gross"                     "genres"
## [11] "actor_1_name"              "movie_title"
## [13] "num_voted_users"           "cast_total_facebook_likes"
## [15] "actor_3_name"              "facenumber_in_poster"
## [17] "plot_keywords"             "movie_imdb_link"
## [19] "num_user_for_reviews"      "language"
## [21] "country"                   "content_rating"
## [23] "budget"                    "title_year"
## [25] "actor_2_facebook_likes"    "imdb_score"
## [27] "aspect_ratio"              "movie_facebook_likes"
```

Question 5: Using ggplot()

```
sp1 <-
  ggplot(data = imdb)+
```

```
  geom_point( mapping = aes(x = imdb_score, y = budget)) +
  labs(x = "IMDB Score",
       y = "Budget",
       title = "IMDB Score by Budget (Categorized by Rating)")+
  ggthemes::theme_economist()

sp1
```



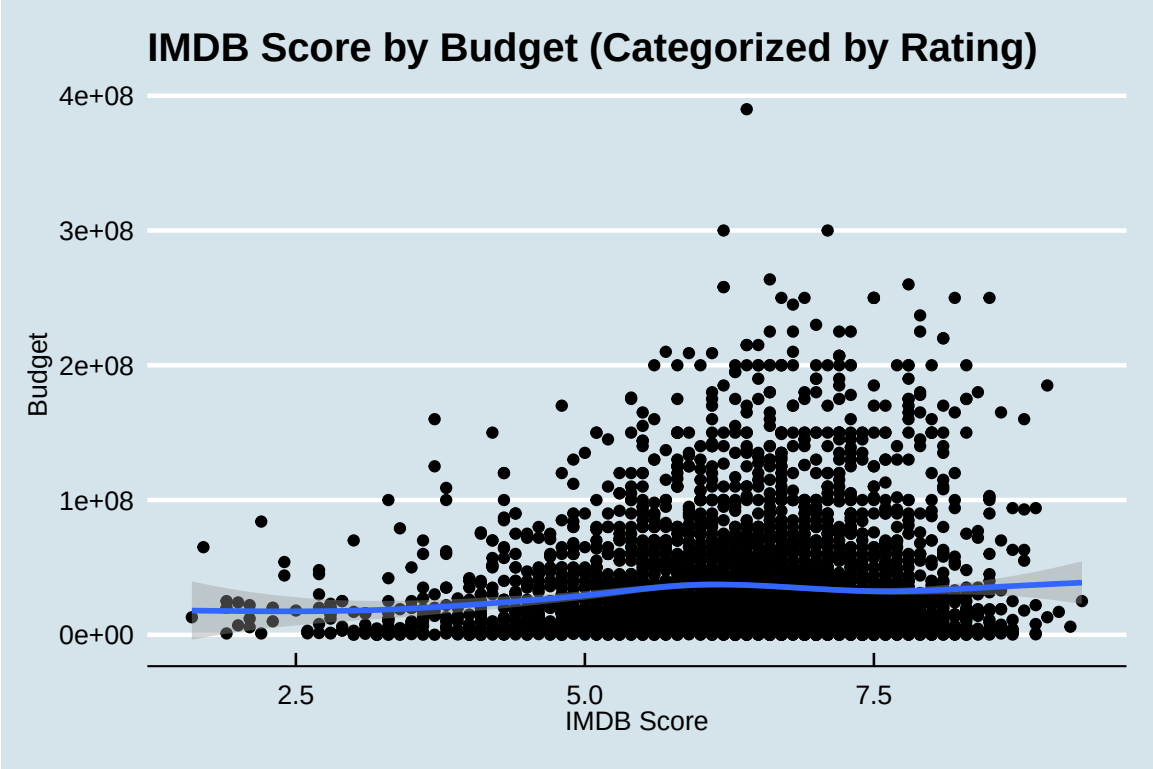
Question 6: Validating Data and Filtering

```
imdb <- imdb %>% filter(budget < 400000000)
dim(imdb)
## [1] 4539   28
```

Question 7: Using stat_smooth()

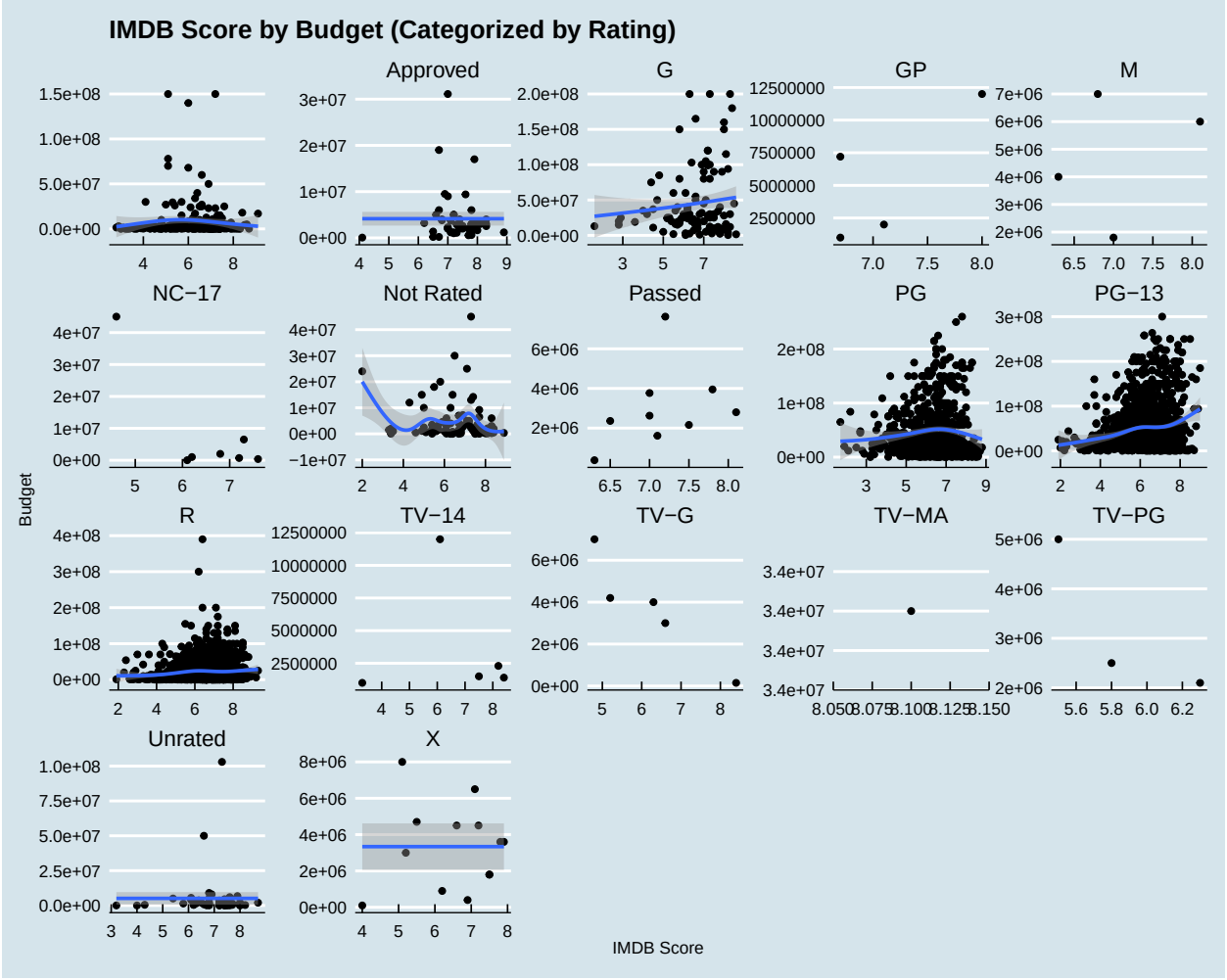
```
sp2 <-
  ggplot(data = imdb)+
  geom_point( mapping = aes(x = imdb_score, y = budget)) +
  stat_smooth( mapping = aes(x = imdb_score,y = budget) )+
  labs(x = "IMDB Score",
       y = "Budget",
       title = "IMDB Score by Budget (Categorized by Rating)")+
  ggthemes::theme_economist()

sp2
```

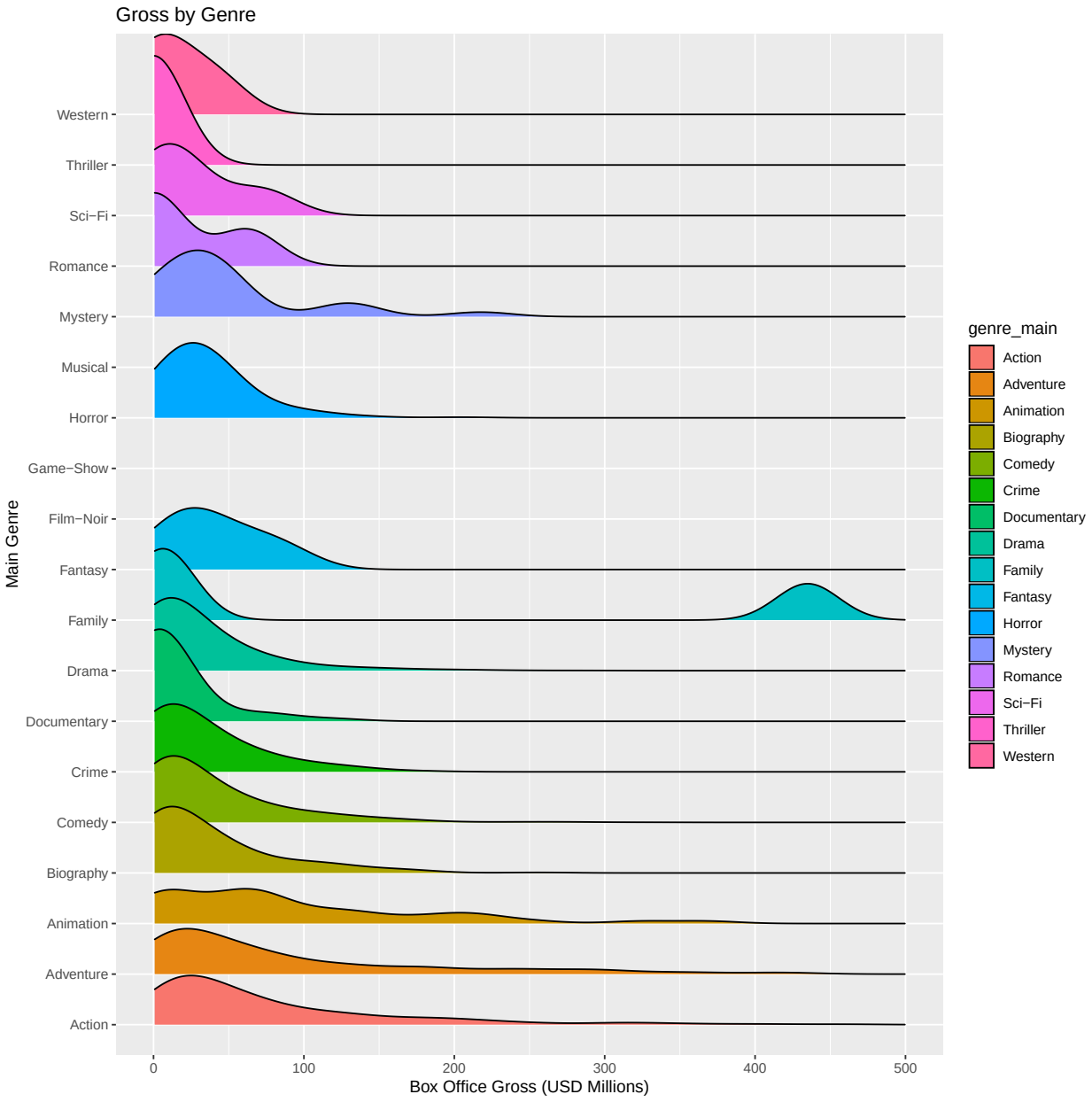


Question 8: Using facet_wrap()

```
sp3 <-
  ggplot(data = imdb)+
  geom_point( mapping = aes(x = imdb_score, y = budget)) +
  stat_smooth( mapping = aes(x = imdb_score,y = budget) )+
  facet_wrap(~ content_rating, scales = "free") +
  labs(x = "IMDB Score",
       y = "Budget",
       title = "IMDB Score by Budget (Categorized by Rating)")+
  ggthemes::theme_economist()
sp3
```


IMDB Score by Budget (Categorized by Rating)
Approved
G
GP
M
NC−17
Not Rated
Passed
PG
PG−13
R
TV−14
TV−G
TV−MA
TV−PG
Unrated
X
Budget
IMDB Score

Question 9: Using ggridges



Question 10 Exploring Data

We chose to compare budget and gross across rate PG -13 and R rated movies. Big Budget PG -13 Movies are made for more of a mass appeal in mind (e.g. Fast and Furious, Transformers) with the hopes of major grossing while movies rated R have a much wider range for budgets and grossing. Some rated R movies such The Shape of Water will have a relatively lower budget but still get a high gross because of the critical acclaim. They tend to be more specialized movies where budget isn't as important of a factor in the success of the movie

```
filterd <- imdb %>% filter(content_rating %in% c("PG","R"))

sp5 <-
```

```
  ggplot(data = filterd)+
  geom_point( mapping = aes(x = budgetM, y = grossM)) +
  stat_smooth( mapping = aes(x = budgetM,y = grossM) )+
  facet_wrap(~ content_rating, scales = "free") +
  labs(x = "Budget",
       y = "Gross",
       title = "Budget by Gross (Categorized by Rating)")
```

sp5