MATH2270 Assignment 2

Code

Visualising Open Data

Student Details

- Casey-Ann Charlesworth (s3132392)

Hide

```
# Load your packages
library(readr)
library(dplyr)
library(magrittr)
library(ggplot2)
```

Data

Hide

```
# Load your data and prepare for visualisation
Movies <- read.csv("IMDB-Movie-Data_mod.csv")
Movies$Rev_male <- with(Movies, ifelse(Gender==0, Revenue_mill, NA))
Movies$Rev_female <- with(Movies, ifelse(Gender==1, Revenue_mill, NA))
No_Data <- nrow(Movies) - nrow(subset(Movies, Revenue_mill>-1))
```

Visualisation

Hide

```
# Visualise Your Data
gg <- ggplot(Movies, aes(Rev_female))
gg + geom_histogram(aes(x=Rev_female, y= ..density..), colour="white", fill="#fe6ba3",
                    binwidth=18) +
  geom_histogram(aes(x=Rev_male, y=-..density..), colour="white", fill="royalblue",
binwidth=18) +
  coord_flip(ylim=c(-0.013,0.013), xlim=c(0,937)) +
  theme_bw() +
  labs(title="Do popular films earn more revenue with a male or female lead?",
       subtitle="Gross revenue of the 1,000 most popular film titles from 2006 to 2016*",
       x="Revenue (millions)", y="Proportion of films",
       caption=paste("Source: https://www.kaggle.com/PromptCloudHQ/imdb-data \n
       *", No_Data, "films with no gross revenue available")) +
  theme(plot.title = element_text(size = 30, face = "bold"),
        plot.subtitle = element_text(size= 15),
        plot.caption = element_text(size=10, hjust=0)) +
  annotate("text", label="Star Wars: Episode VII - The Force Awakens",
           x=936.5, y=0.0006, size=4, hjust=0) +
  geom_segment(x=936.5, y=0.0005, xend=936.5, yend=0.0003,
               arrow = arrow(length=unit(0.1, "cm"))) +
  annotate("text", label=paste(nrow(subset(Movies, Rev_male>-1)), "movies with male leads"),
           x=-20, y=-0.0005, hjust=1, size=4) +
  annotate("text", label=paste(nrow(subset(Movies, Rev_female>-1)), "movies with female lead
s"),
           x=-20, y=0.0005, hjust=0, size=4)
ggsave("assign2.png", height = 8, width = 15)
```

Commentary (200 words)

Although the release and popularity of Star Wars: Episode VII has completely affected the shape and story of the back-to-back histograms, the outlier itself demonstrates an interesting point - that big box office films can succeed with female leads.

That said, of the 1,000 films, the lead actor was likely to be male approximately 71.6% of the time compared to 28.4% for females.

However, once the films with no revenue were removed, the male to female lead actor ratio changed to 3:1 respectively.

Due to the higher number of male led films, the scale up from low grossing films to the higher ones is quite naturally curved – thus meaning that more male led films have gross revenue at every step of the box office ladder up to about $750 million.

It differs for female led films, in that most gross revenues are on the lower end of the scale. With each step up the histogram, there are fewer female films at this level, and thus the curve has been replaced with a histogram that contains gaps and a more erratic design.

Granted there are some higher grossing films with female leads (such as Star Wars and even Passengers listed Jennifer Lawrence first, even though I would argue that Chris Pratt is the lead), however, for the most part female led films earn less at the box office.

Perhaps this is due to the types of films women lead – most blockbusters tend to have the male action star (with Star Wars ep 7 bucking that trend) while most female led popular films are more along the drama, romance or comedy genres.

In conclusion, we must never forget that women led blockbusters cannot only be popular but also successful at the box office.