MATH2270 Assignment 2

Visualising Open Data

Student Details

- Casey-Ann Charlesworth (s3132392)

<div style="text-align: right;">Hide</div>

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

Data

<div style="text-align: right;">Hide</div>

```
# 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))
#View(Movies)
```

Visualisation

<div style="text-align: right;">Hide</div>

```
# 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)",
       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=9, hjust=0)) +
  annotate("text", label="Star Wars: Episode VII - The Force Awakens",
           x=936.5, y=0.0006, size=3, 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=3) +
  annotate("text", label=paste(nrow(subset(Movies, Rev_female>-1)), "movies with female lead
s"),
           x=-20, y=0.0005, hjust=0, size=3)
```


Do popular films earn more revenue with a male or female lead?
Gross revenue of the 1,000 most popular film titles from 2006 to 2016*

Source: https://www.kaggle.com/PromptCloudHQ/imdb-data

* 128 films with no gross revenue available
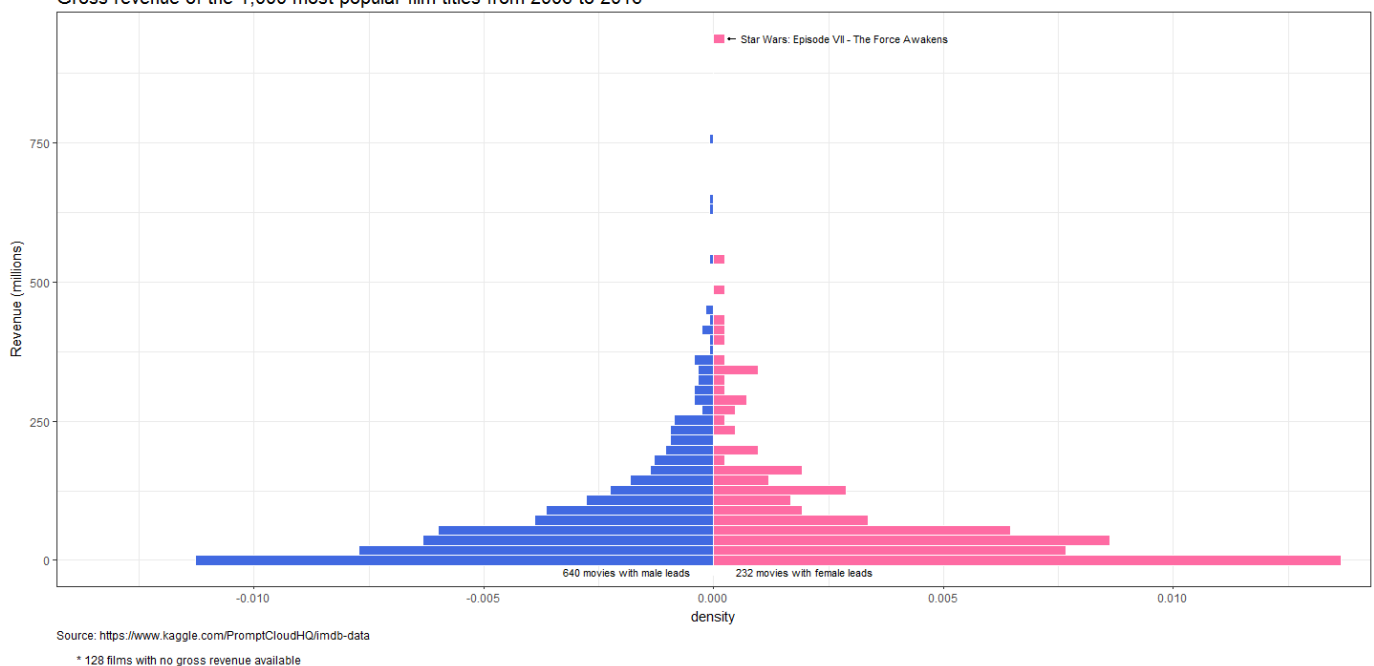
Hide

```
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 70% of the time compared to 30% for females. That would account for the graduated gross revenue scale that can be seen on the male histogram side. Compare that to the females, who had a larger amount of low grossing films but whose films also made a good showing further up the scale. However, as can be seen by the gaps and erratic pattern, higher grossing films with women leads were not as consistently occuring as can be seen with the men.