R Notebook

Description of data

This is the raw data behind the story "Be Suspicious Of Online Movie Ratings, Especially Fandango's. (http://fivethirtyeight.com/features/fandango-movies-ratings/). The dataframe contains every film that has a Rotten Tomatoes rating, a RT User rating, a Metacritic score, a Metacritic User score, and IMDb score, and at least 30 fan reviews on Fandango.

Load libraries

```
rm(list=ls(all=TRUE))

setwd("C:/Users/Neil/Desktop/Class folder/Fall 2017/RM/Analyses/MyClassActivities")
library(tidyverse)
```

Import data

```
fandango <- read_csv("fandango.csv")
```

Wrangle the data

In order to create the plots from the story we must first flip the data from wide to long.

```
fandango_long <-  fandango %>%
  mutate(fandango_actual_stars = round((fandango_ratingvalue/.5))*.5) %>%
  select(film, rt_norm_round, rt_user_norm_round, metacritic_norm_round,
         metacritic_user_norm_round, imdb_norm_round,
         fandango_stars, fandango_actual_stars) %>%
  gather(rt_norm_round, rt_user_norm_round, metacritic_norm_round,
         metacritic_user_norm_round, imdb_norm_round,
         fandango_stars, fandango_actual_stars,
         key = rating, value = stars) %>%
  mutate(website = factor(rating,
                          levels = c("rt_norm_round", "rt_user_norm_round",
                                     "metacritic_norm_round", "metacritic_user_norm_round",
                                     "imdb_norm_round",
                                     "fandango_stars", "fandango_actual_stars"),
                          labels = c("Rotten Tomatoes", "Rotten Tomatoes users",
                                     "Metacritic", "Metacritic users",
                                     "IMDb users",
                                     "Fandango", "Fandango Corrected"))) %>%
  mutate(stars.f = factor(stars, levels = c(.5,1,1.5,2,2.5,3,3.5,4,4.5,5)))

fandango_long_grp <- fandango_long %>%
  group_by(website, stars.f) %>%
```

```
  summarize(count = n()) %>%
  complete(stars.f) %>%
  mutate(count = ifelse(is.na(count), 0, count), percent_stars = (count/146 * 100)) %>%
  ungroup()

fandango_long_grp
```

```
## # A tibble: 70 x 4
##            website stars.f count percent_stars
##             <fctr>  <fctr> <dbl>         <dbl>
##  1 Rotten Tomatoes     0.5    14      9.589041
##  2 Rotten Tomatoes       1    10      6.849315
##  3 Rotten Tomatoes     1.5    16     10.958904
##  4 Rotten Tomatoes       2     4      2.739726
##  5 Rotten Tomatoes     2.5    17     11.643836
##  6 Rotten Tomatoes       3    13      8.904110
##  7 Rotten Tomatoes     3.5    11      7.534247
##  8 Rotten Tomatoes       4    15     10.273973
##  9 Rotten Tomatoes     4.5    23     15.753425
## 10 Rotten Tomatoes       5    23     15.753425
## # ... with 60 more rows
```

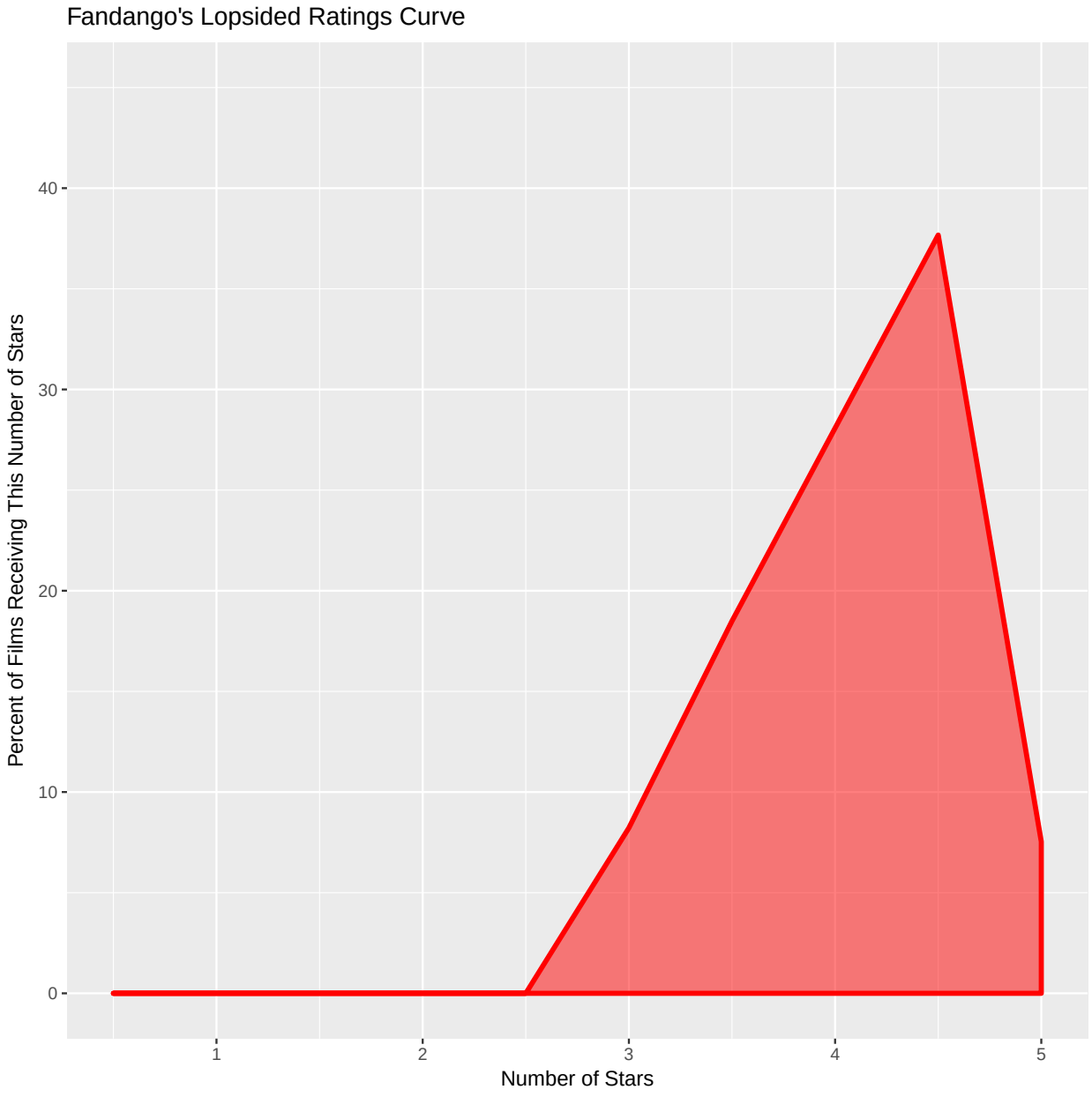
Create the three plots from the FiveThirtyEight Story

Create plot 1

This plot demonstrates that Fandango has a very lopsided distribution to their ratings.
According to Fandango, none of the movies are lower than 3 stars.

```
plot1 <- filter(fandango_long_grp, website == "Fandango") #" == Fandago" is telling us to only look at

ggplot(plot1, aes(y = percent_stars, x = as.numeric(as.character(stars.f)))) +
  geom_area(size = 1.25, fill = "red", color = "red", alpha = .5) +
  labs(title = "Fandango's Lopsided Ratings Curve",
       y = "Percent of Films Receiving This Number of Stars", x = "Number of Stars") +
  ylim(0,45)
```

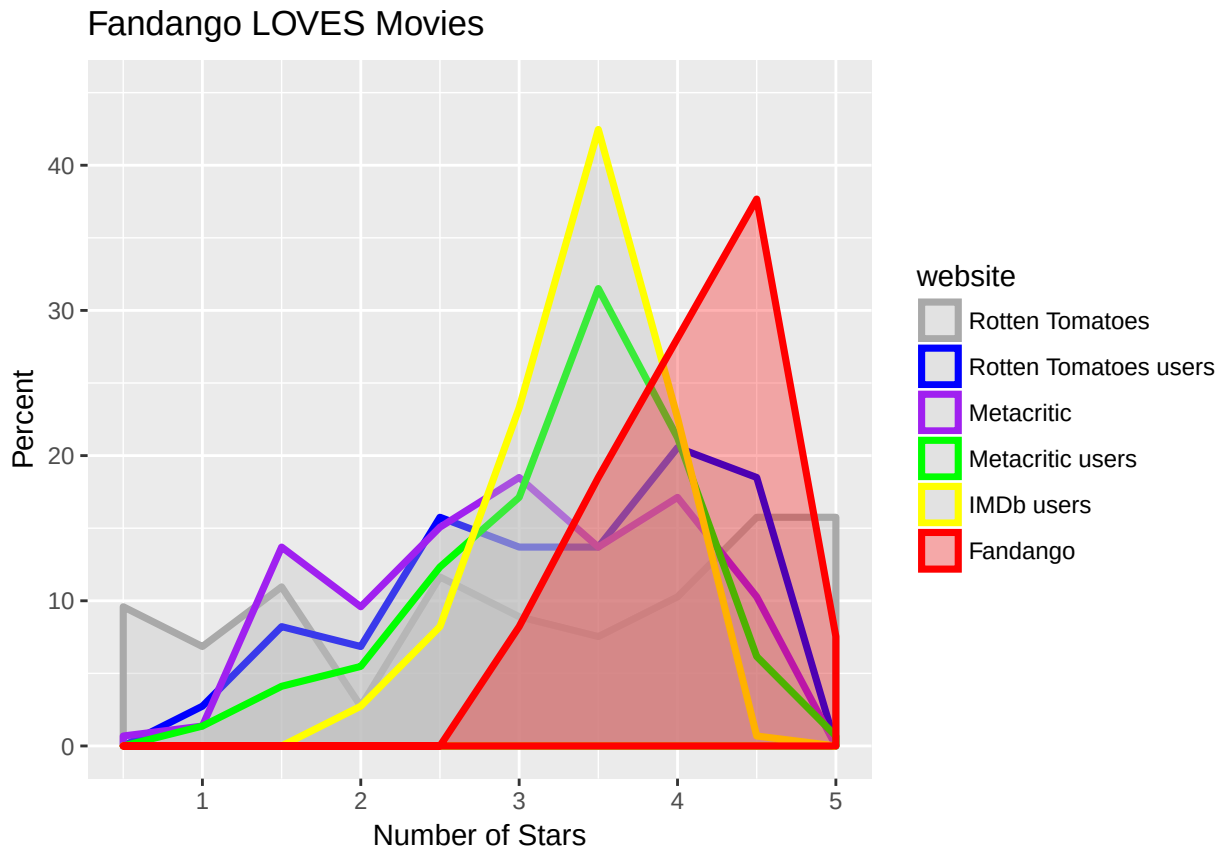


Create plot 2

This plot demonstrates that other websites use the full rating scale.

```
plot2 <- filter(fandango_long_grp, website != "Fandango Corrected")

ggplot(plot2, aes(y = percent_stars, x = as.numeric(as.character(stars.f)),
                  colour = website, fill = website)) +
  geom_area(size = 1.25,
            position = "identity") +
  labs(title = "Fandango LOVES Movies",
       y = "Percent", x = "Number of Stars") +
  scale_fill_manual(values=alpha(c("grey", "grey", "grey", "grey", "grey", "red"),
                                 alpha = .3)) +
  scale_color_manual(values=c("dark grey", "blue", "purple", "green", "yellow", "red")) +
  ylim(0,45)
```



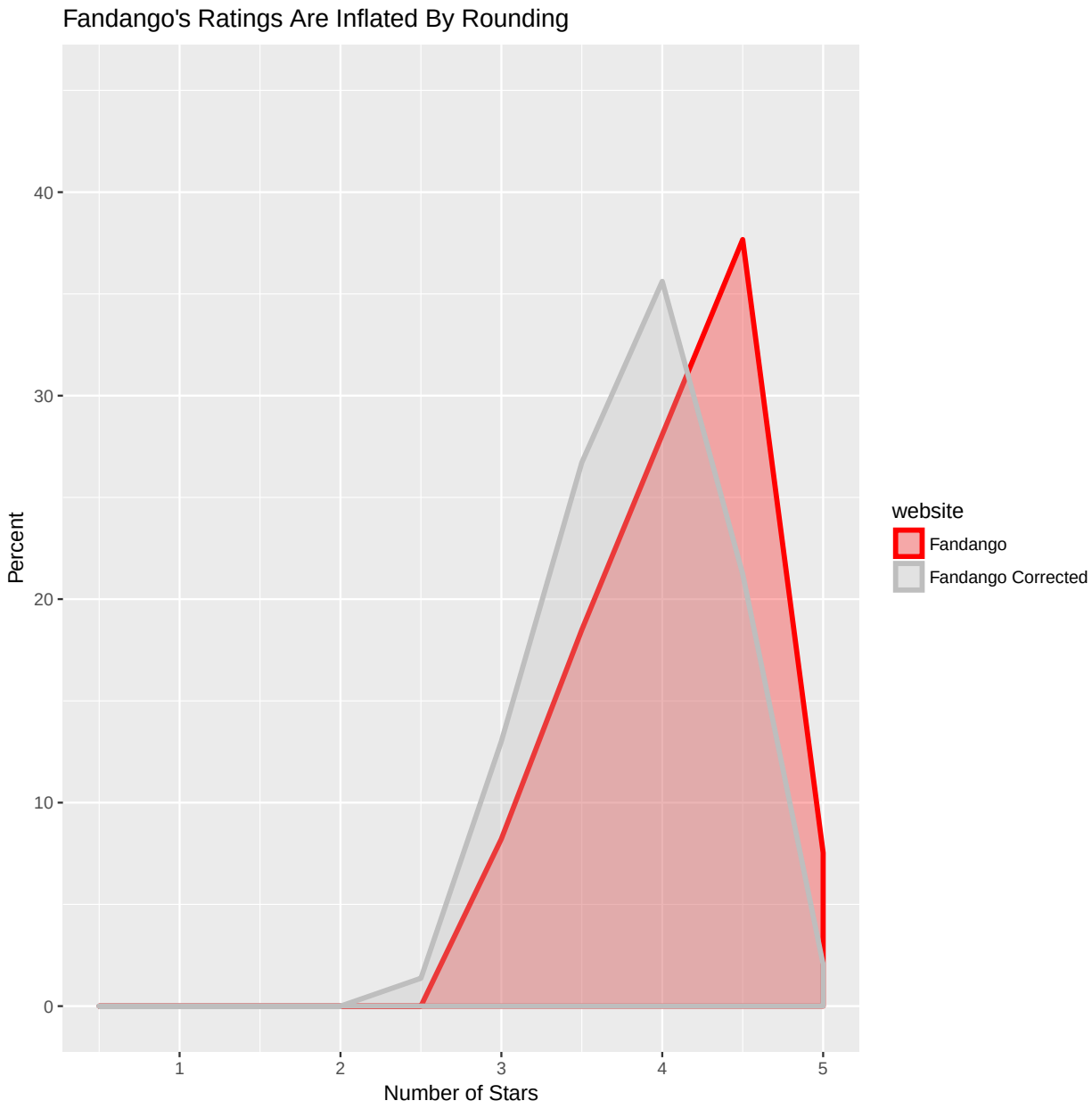
Create plot 3

By looking at Fandango's rounded rating, it becomes apparent that Fandango always rounded up!

```
compare <- fandango %>%
  select(film, fandango_stars, fandango_ratingvalue)

plot3 <- filter(fandango_long_grp,
                website == "Fandango" | website == "Fandango Corrected")

ggplot(plot3, aes(y = percent_stars, x = as.numeric(as.character(stars.f)),
                  colour = website, fill = website)) +
  geom_area(size = 1.25,
            position = "identity") +
  labs(title = "Fandango's Ratings Are Inflated By Rounding",
       y = "Percent", x = "Number of Stars") +
  scale_fill_manual(values=alpha(c("red", "grey"), alpha = .3)) +
  scale_color_manual(values=c("red", "grey")) +
  ylim(0,45)
```


Fandango's Ratings Are Inflated By Rounding
Percent
40
30
20
10
0
1
2
3
4
5
Number of Stars
website
Fandango
Fandango Corrected