Boxplots and Chance

Unit 2 - Lab 5

Directions: Follow along with the slides and answer the questions in **BOLDED** font in your journal.

Some background...

- For this lab, we will be using the `movie` data set to "make the call" for differences between boxplots.
- The `movie` data set was compiled using the Rotten Tomatoes website, which is an online movie and TV review database, and Box Office Mojo, which is an online database of movie revenues.

Let's investigate

- Perhaps we want to know if there is a difference in the runtime between rated G and rated R movies.
- **Without looking at any data, do you have an opinion about this? Do you think rated R movies run longer than G rated movies? Why?**

Creating one boxplot

- We can create a boxplot of just the runtimes of the movies first.

```
bwplot(~runtime, data = movie)
```

- **Record the values for Q1, Q3, and the median.**

Expanding further

- We can separate this boxplot into multiple boxplots simply by sorting runtime by the MPAA rating.

```
bwplot(~runtime | mpaa_rating,
       data = movie)
```

- But this is not very easy to read or compare between the ratings. We could change the layout by using the `layout` option in the `bwplot()` function to correct this.

```
bwplot(~runtime | mpaa_rating,
       data = movie, layout = c(1,6))
```

It's still pretty confusing

- Right now, there are 6 different MPAA ratings to compare between. We're really interested in only 2 though: rated R and rated G.
- Let's further simplify our boxplot function by using the `subset` argument.

```
bwplot(~runtime | mpaa_rating,
       data = movie,
       subset = mpaa_rating == "G" |
          mpaa_rating == "R",
       layout = c(1,2))
```

- **Write down how to use the `subset` argument.**

The boxplot

- **Include sketches of the boxplots in your DS Journal.**
- **What are the Q1, Q3, and median values for rated R movies?**
- **What are the Q1, Q3, and median values for rated G movies?**
- **Do the data overlap? Do the boxes overlap?**

Make the call!

- **How do the median values compare for each of the 2 ratings?**
- **How do the box sizes compare?**
- **Make the call about whether or not rated R movies and rated G movies have differing runtimes. Explain your reasoning.**

On your own:

- Make boxplots of the `critics_rating` variable for movies rated *PG-13* and _R+. Then do the following:
  - **Write down the code you used to make these boxplots.**
  - **Compute the exact *median*, *Q1*, *Q3* and *IQR* for each boxplot. Look back at lab 2.2 if you need help computing these values.**
  - **Make the call: Are movies that are rated *R* rated more or less favorably than films rated *PG-13*. Justify your answer by referencing your numerical summaries from above.**