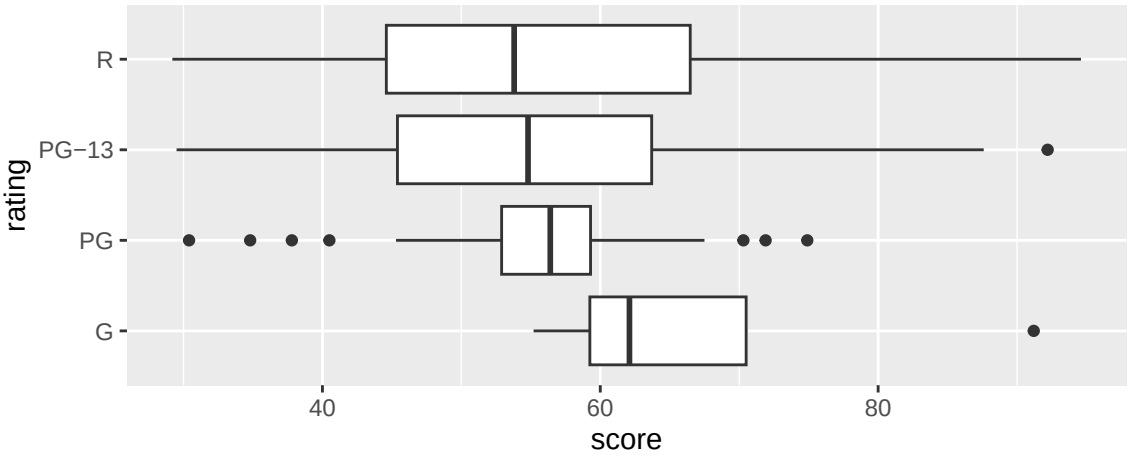
Team members:

In this lab we will use a dataset of movies. For each movie we have information on their genre, critic score (from 0-100), rating (G, PG, PG-13, or R), and box office earnings (in millions of dollars).

Question 1



List 3 factual observations about the plot shown above.

1.

2.

3.

Question 2

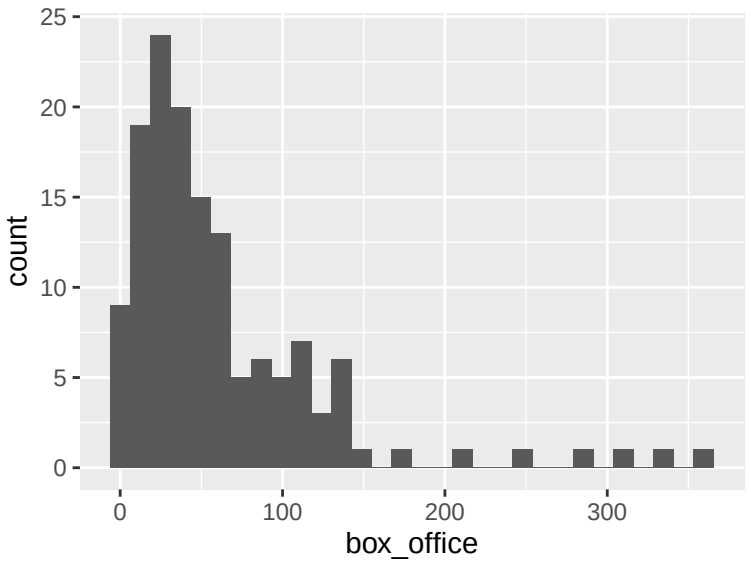
Which ratings categories appear to be similar in critic score? Which appear to be the least similar? Does this make sense?

Question 3

Identify each plotting step from the code (i.e., circle and number the step):

1. Call `ggplot()` with the data.
2. Specify variables to be mapped to aesthetics (e.g., axes, colors, etc)
3. Add a layer to specify the plot type.

```
ggplot(movies, aes(x = box_office)) +
  geom_histogram()
```



Question 4

How would you describe the distribution visualized in question 3 (right-skewed, left-skewed, symmetric)?

Question 5a

If we doubled the dataset by duplicating every row, how would the plot from question 1 change, if at all?

Question 5b

If we doubled the dataset by duplicating every row, how would the plot from question 3 change, if at all?

Question 6

I want to produce a bar plot of movie ratings. Identify which code I should use.

Option 1:

```
ggplot(movies, x = rating) +
  geom_bar()
```

Option 2:

```
ggplot(aes(movies), x = rating) +
  geom_bar()
```

Option 3:

```
ggplot(movies, aes(x = rating))
  geom_bar()
```

Option 4:

```
ggplot(movies, aes(x = rating)) +
  geom_bar()
```

Option 5:

```
ggplot() +
  geom_bar(movies, x = aes(rating))
```

Question 7

Work with your team to sketch a box plot of box office earnings using information provided by the plot in question 3.

Question 8

I want to investigate which genre (action, comedy, or horror) had more rated R movies and which genre had the fewest PG movies. What type of plot would you recommend?

Create a sketch of what this plot might look like. (I haven't shown you any data for genre, so this exercise is for you to think about what would be on the x-axis, y-axis, legend, and what the data could look like).