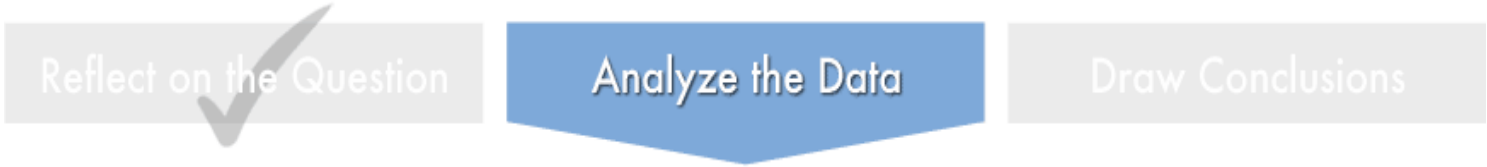

Primary Research Questions

1. Does a film's rating (PG, PG-13, or R) impact its cost to produce?
2. Does a film's rating (PG, PG-13, or R) influence its IMDB score?

Breakdown Your Analysis

Let's break this analysis into its required steps:

For each ANOVA:

1. Identify the number of films in each rating group (PG, PG-13, R).
2. Compute the mean and standard deviation of the variable of interest for each group.
3. Create boxplots to help visualize group differences and check test assumptions.
4. Run ANOVA.
5. If the F statistic is significant, run a Tukey HSD test to determine which groups are different.

Here is the code you will use:

```
# Show how many films are in each group
table(film$Rating)
```

Question 1

```
# Calculate avg film budget of each group
aggregate(Budget~Rating,film,mean)

# Calculate sd of film budget within each group
aggregate(Budget~Rating,film,sd)

# Visualize the group means and variability
boxplot(film$Budget~film$Rating, main= "Film Budgets by Rating",
ylab= "Budget", xlab= "MPAA Rating")

# Run ANOVA
```

Lab due May 03, 2016 at 17:00 UTC

Problem Set
Problem Set due May 03, 2016 at 17:00 UTC

```
modelbud <- aov(film$Budget~film$Rating)
summary(modelbud)

# Run post-hoc test if F statistic is significant
TukeyHSD(modelbud)
```

Question 2

```
# Calculate avg IMDB score of each group
aggregate(IMDB~Rating,film,mean)

#Calculate sd of IMDB scores within each group
aggregate(IMDB~Rating,film,sd)

# Visualize the group means and variability
boxplot(film$IMDB~film$Rating, main= "IMDB Scores by Rating",
        ylab= "IMDB Score", xlab= "MPAA Rating")

# Run ANOVA
modelscore <- aov(film$IMDB~film$Rating)
summary(modelscore)

# Run post-hod text if F statistic is significant
TukeyHSD(modelscore)
```

(1/1 point)

1. What does **aov** stand for?

- it doesn't mean anything; it is just an indicator that we want to create a vector of scores
- an open variable
- analysis of variance ✔

Click here for a video explanation of how to answer this question.

You have used 1 of 1 submissions

(1/1 point)

2. Which of the following comes closest to what it sounds like to "read aloud" this line of code?

```
aggregate(Budget~Rating,film,mean)
```

- Find the Budget **for each** Rating group, and then calculate the overall mean for the dataset film.

- Look across all the Budget cases **and then** find the mean Rating for each film.
- For all the cases in film, take the variable Budget and, **given the** Rating group, find the mean ✔

You have used 1 of 1 submissions

(1/1 point)

3. If group differences **are** present, what should be true about the output of this line of code?

```
aggregate(Budget~Rating,film,mean)
```

- The average budget for each group should be different. ✔
- The mean for the each rating group should be about twice as large as the next.
- The average rating should appear about the same for each group.

Click here for a video explanation of how to answer this question.

You have used 1 of 1 submissions

(1/1 point)

4. If we are to **satisfy** the assumptions of ANOVA, what should be true about the output of this line of code?

```
aggregate(Budget~Rating,film,sd)
```

- The standard deviation of each group's budget should vary.
- The largest standard deviation should be no more than twice the smallest standard deviation. ✔
- The standard deviation of each group's ratings must be identical.

Click here for a video explanation of how to answer this question.

You have used 1 of 1 submissions

(1/1 point)

5. Suppose we wanted to test if each type of Genre had the same level of Ratings. What has caused the error below? (You may want to refer to the dataset in R for help.)

```
film <- FilmData
modelRating <- aov(film$Rating~film$Genre)

Warning messages:
1: In model.response(mf, "numeric") : using type = "numeric" with a
factor response will be ignored
2: In Ops.factor(y, z$residuals) : - not meaningful for factors
```

- (●) We should have run a Chi Square Test of Independence. ✔
- ( ) We cannot store the aov() function information into "modelRating."
- ( ) One of the variable names is spelled differently in our dataset.
- ( ) We should have run a two-sample t-test.

Click here for a video explanation of how to answer this question.

You have used 1 of 1 submissions