**Kendal Bergman**

**Challenge 1**

**Given the provided data, what are three conclusions that we can draw about crowdfunding campaigns?**

1. The most common parent category that occurred was theater. The least common parent category that occurred was journalism.
2. Theater is the most common parent category occurring for campaigns in the U.S., but film & video is the most common parent category occurring for campaigns in Great Britain.
3. More campaigns were launched in the U.S. than in any of the other countries.

**What are some limitations of this dataset?**

This dataset includes multiple countries, but there are far more American campaigns represented in the dataset vs. other countries. Drawing conclusions about campaigns in China, for example, is going to be based on a limited sample size (31 total campaigns). Since we only have the total pledged amount and the number of backers, we can only determine the average donation per person for each campaign. We don’t have any way to look at particularly large or small donations and whether there is any link between campaign success and outlier amounts. We can also see when campaigns are created and when they ended, but we don’t know when in the lifecycle of the campaigns that the most funds were donated.

**What are some other possible tables and/or graphs that we could create, and what additional value would they provide?**

We could create a pivot table that looks at outcomes by average donation. That might provide insight on the most successful level of giving (do campaigns with lower average donations have more success?). We could also create a graph showing outcomes based on the goal amount per campaign. That could give us insight into whether certain goals are more likely to be successful. We could also create a new column showing the length of each campaign, and look at whether longer or shorter campaigns seem to have more successful outcomes. That could help target ideal lengths of campaigns in the future.

**Statistical Analysis:**

**Use your data to determine whether the mean or the median better summarizes the data.**

In this instance the median better summarizes the data. That is because the distribution of the data is skewed in both cases (as opposed to normal distribution). There are some campaigns with very large numbers of backers, but then there are many more campaigns with small amounts of backers.

**Use your data to determine if there is more variability with successful or unsuccessful campaigns. Does this make sense? Why or why not?**

There is more variability with successful campaigns – the standard deviation for successful campaigns is greater. This makes sense because, in general, you would expect unsuccessful campaigns to have fewer backers overall. If a campaign is successful, it means it met its goal, and that probably means more people supporting the campaign. However, there is no rule saying that successful campaigns have to have lots of backers. So, a campaign could ultimately be successful if there were a small number of passionate supporters.