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

Theater, specifically plays, is the most common campaign category. We could then interpret that the theater community uses crowdfunding more than any other industry.

Successful campaigns have more backers. This is shown since mean, median, min, + max for successful campaigns are all larger that that of failed campaigns.

If you look at the “Outcomes Based on Goal” graph, I would only hold value in the data below $9,999, since there are a lot of data points in this range. For the other goal ranges, I don’t feel comfortable drawing any conclusions. There are too few data points from $10,000 to $49,999. For greater than or equal to $50,000, there are a lot of data points. However, every other “bin” graphed is no more than a range of $5,000. If you look at the campaigns that are greater than $50,000, it goes all the way up to $199,200! This is too large a range of goal values and shouldn’t be compared to the smaller “bins.” Therefore, the only conclusion I can draw from this graph is that campaigns are more likely to succeed vs fail if goal is below $10,000.

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

The data ranges across 10 years, but only has 1,000 data points. This seems to be very few considering this large window of time and how popular crowdfunding is.

The parent category, journalism, may be likely to skew the data since it has an impressive 100% success rate, although only 4 campaigns. So, if you were to take the success rates across all categories and then average them, I’d consider leaving the 100% value out.

The US campaigns contribute 3x more data than all other countries combined. So, if you do not filter by country and just look at all countries, then any conclusions you draw are most likely due to the US data.

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

I think creating a table that lists outcomes in terms of % would be a better way to compare results across categories. Comparing counts isn’t that useful, since categories that have more funding attempts (theater) will likely have a higher count of success vs those that have fewer attempts (journalism). Then you could do a stacked bar graph with %successful, %live, %failed, %cancelled across all categories and you’d get a better idea of which categories are most likely to exceed funding goals.

It might be more valuable if we had the raw data of donation values, not just the average donation value for each campaign. You could then create a box + whisker plot to see what value of donations you’re likely to receive. You’d then be able to start a campaign knowing what to expect and would be able to set backer perk levels accordingly.

Through the statistical analysis we did of backers\_count, we see that successful campaigns have more backers. I’d like to know if they were only more successful due to “word of mouth” or if they were also more successful because the average donation was greater for successful vs failed campaigns. I’d do the same statistical analysis, just use average donation data instead of backers\_count data.