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

Using the provided data we can draw some conclusions. Theatre is the category with the most amount of campaigns at around one third of the campaigns, and is the most popular crowd funded parent category. Breaking down by subcategory, the most popular subcategory of crowdfunding is plays. This makes sense as it is a subcategory of the most popular category, theatre. A third conclusion we could draw about the data is the fundraising goal. Campaigns are most successful when they have a campaign goal of crowdfunding that is over $10,000 but less than $25,000. All campaigns were successful in this range which makes it an ideal goal for crowdfunding campaigns.

1. What are some limitations of this data set?

There are some limitations to this data set. Firstly, the age of the data makes conclusions drawn from it not as directly applicable to current crowdfunding campaigns as there has been so much advancement in the way that campaigns are constructed, marketed, shared, and methods of donation due to the advance of technology.

Another limitation is that when breaking down by sub category, plays is quite a broad category compared to other categories that broke into multiple categories, such as breaking down music into genres like jazz and rock. This could contribute to why the data is skewed towards plays being the highest subcategory.

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

Another table and graph that could be created to provide some additional insight is comparing the number of backers, broken down by category and subcategory, and not just the number of campaigns. This could help determine if there are categories with less backers but more money, and determine if there are some backers that are skewing the data. There also could be some conclusions drawn if the data is broken up by country, to see if there is a country that is particularly influencing certain campaigns.

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

The median is more useful to summarize the data because the data is not normally distributed. There are many campaigns with smaller number of backers, and fewer campaigns with large number of backers, so the data is skewed. There is also a high variance and thus a high standard deviation of the data, which suggests the median is more of a reliable summary of the data.

1. 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, as they are distributed among the entire amount of backers: low backers and high backers. The failed campaigns were more clustered around the lower amount of backers. This makes sense because campaigns with high amounts of backers seem to be more likely to be successful, because if there are many backers then they are able to reach their goal and become successful. More statistical analysis may be required to see whether campaigns are more likely to be successful or not when there are low backers, but the option of having higher amounts of backers makes the variability higher.