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

The parent categories with the highest number of successful campaigns are theater, film & video, and music. Furthermore, the sub-categories with the highest number of successful campaigns are plays and rock. Overall, there’s an increasing number of successful campaigns from May to July.

* What are some limitations of this dataset?

Theater, music, and film & video campaigns make up 697 of the 1000 observations. So, these types of projects may appear to be more successful overall because of the large number of observations that they represent. In contrast, journalism campaigns make up only 4 of the 1000 observations. All 4 of these campaigns were successful. Thus, one might assume that all journalism campaigns will be successful. However, this success rate may be due to amount of data we have for that type of campaign. Another issue is that US campaigns make up 763 of the observations. So, the conclusions we draw for the overall dataset may not be accurate for campaigns in other countries.

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

I think it would help to create a trendline graph of average donation by date created conversion. This graph would have to be filtered by currency to avoid confusion due to the different values of each currency. It may also be helpful to filter it by parent category. With this graph we may be able to draw conclusions about what months consumers tend to make high donations, which would then contribute to the success of a campaign. I think it would also be helpful to look at a graph of average percent funded by parent category or date created conversion. This may give insight into the conditions under which campaigns overperform donations wise. Another graph to look at would be average goal by parent category. Perhaps certain campaign parent categories tend to be successes because they have a low average goal.

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

We know the successful campaigns data is right skewed because the mean is greater than the median. Thus, the median best summarizes the data. Similarly, the median best describes the failed campaigns as well.

* 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 in successful campaigns because their variance is higher than that of failed campaigns. This makes sense because the difference between the mean and median of the successful campaigns is greater than that of the failed campaigns.