Analysis:

Conclusions:

From the stacked bar chart of the parent category vs number of outcomes we may conclude that three major categories 'theater', 'film & video', 'music' has significantly higher number of outcomes compared to other categories. On average we may also conclude that the number of successful campaigns is higher than other campaigns.

Also, from the stacked bar chart for subcategories, we observe that 'plays' in theater, 'Rock' in music, and 'documentary' in film & video categories have higher number of outcomes as well as number of successful campaigns than other categories and campaigns.

In addition, according to the line chart for 'count of outcome' vs 'created month' we observe that on average the number of successful campaigns throughout a year is significantly higher than failed, canceled, or live campaigns.

Finally, from the summary of the number of backers for successful and unsuccessful campaigns we observe that since the distributions are not symmetric, and both are skewed to the right, median is a better measure of central tendency for both successful and unsuccessful campaigns. Also, for both types of campaigns, interquartile range might be a good measure of dispersion, but standard deviation and variance all describe the variability in the subsets very well. We can observe that unsuccessful campaigns have smaller variance and standard deviation than successful campaigns.

Limitations of the data set:

Category ‘theater’, which has the highest number of campaigns, has only ‘plays’ as its subcategory in this data set. Is this really the only subcategory that theater has? The variety in this category is a lot less than the other categories such as ‘music’ and ‘film & video’.

The way the data is categorized in this data set is probably not the best. For example, it might be better if we consider ‘theater’, ‘photography’, ‘publishing’, and ‘journalism’ in one major category, let’s say ‘visual art’ and then make the comparison between the major categories.

Potential plots:

Scatter or box plot of the subcategories to find out their approximate distribution. If we know the distribution of the subsets, we can have a better comparison among different type of campaigns.

A line plot of the date ended variable to figure out how long each campaign last compared to the goal or pledged amounts.