Summary of Data

Goal Analysis

Given the provided data, it appears that projects categorized as film/video, games, journalism, music, and theater received greater rates of success in comparison to the other categories. Looking at a more detailed scope of the subcategories, audio and world music had a 100% success rate. Animation, documentaries, photography books, world music, non-fiction, and television were amongst the subcategories that had high success rates as well. There is also a correlation of outcome to time of year. Success peaked from May to August, while failures peaked from June to September, and cancelations peaked between the months of August and November. In conclusion, campaigns that were amongst the highest rates of success were film/video, games, journalism, and music oriented, and occurred between the months of May to August.

The sample size of this data set is a major limitation. This data set includes 1,000 individuals represented disproportionately in categories. Such a small data set contains an immense amount of bias, which leaves room for error. Due to the lack of representation for categories such as audio and world music, an outlier is created. Audio had a 100% success rate; however, it represents only 4 individuals, while plays had a 54% success rate, but represents 344 individuals.

Another table and graph that could have been created would be a scatterplot for percent funded vs. parent category because it would create a visual for exactly how successful a particular campaign was at obtaining the pledged amount. Which is important because it would show not only how successful it was, but a proportion of how much money was actually raised. Another useful visual would be one for “Date Created” and “Date Ended”, so that we could see exactly how long it took for a particular campaign to reach its goal. Some of the categories may have shown less success because they didn’t run the campaign as long as they should have to reach their goal.

Statistical Analysis

Use of the median would be best to use to summarize both Failed and Successful outcomes because they both exhibit skewed data. Both graphs are skewed left, which means that the median is a better way to summarize the data because it is not as influenced by outliers in the data set. There is also more variability in the Successful Outcome data set, which makes sense because it has a larger range of values, 98-1425 while Failed had a range of 0-452.