Jacob Zahn

Homework 1

**PART 1**

Conclusion 1:

From the PCategory tab, the data suggests that most crowdfunding campaigns occur in the United States (76.3%). Furthermore, the most common crowdfunding category is theater in every country in this dataset except Switzerland (Music) and the United Kingdom (Film & Video).

Conclusion 2:

From the SCategory tab, the data suggests that within the parent category of games, the number of crowdfunding campaigns for video games is more than double that of mobile games, while also exhibiting a higher success rate.

Conclusion 3:

From the Outcome tab, the data suggests that the best time of year for crowdfunding in general is in July. This is based on the fact that July is the time of year when there is a spike in the number of successful outcomes and the difference in the number of successful and unsuccessful outcomes is greatest.

Limitations:

The greatest limitation to any data set is the sample size. While this data set has 1,000 crowdfunds, it is unclear whether new trends would emerge with more data. This is especially true when considering countries outside the US, for which there are only 237 crowdfunds to consider.

Another limitation of this data set is that I do not know how this data was collected, which impacts the reliability of the data set. If the data was collected from financial documents, it is likely to be robust. However, if the data was collected through surveys, it might not be very reliable.

A third limitation of this data set is that there is no information on the specific amount each individual contributed to a crowdfund campaign. Since we only have the total pledged amount and the backer count, we are limited to knowing only the average donation per backer per campaign.

Further Analysis:

Currently, the PCategory pivot table can be filtered by country, which made it convenient to look at each country or a combination of countries, but it would be useful to have a chart that compares the campaigns in each country. For example, I could make a pivot table with countries in the rows and campaign outcomes in the columns, with outcome counts filling in the table.

Another useful chart would be one that looks at percent funded by category. We created the designations “successful”, “failed”, and “canceled” to evaluate the success of a campaign, but in doing so lost some of the information on how successful a campaign was. A campaign that raised 1000% of its goal was more successful than a campaign that raised 100% of its goal. Perhaps some interesting trends would emerge on which types of campaigns tended to raise far more than their goal. This chart would have the campaign categories on the x-axis and percent funded on the y-axis.

**STATISTICAL ANALYSIS**

The median summarizes the data better than the mean. For both successful and unsuccessful campaigns, the data is skewed to the left, so the mean is going to “pulled” toward the right where the higher backer counts are. The median is not influenced by those higher values. If the data sets were more symmetrical, however, the mean would be better.

There is more variability in successful campaigns based on the variability/standard deviation. This makes sense to me because it is more likely to have successful campaigns that have large backer counts, while still having successful campaigns with lower backer counts. The weight of the campaigns with large backer counts increases the standard deviation.