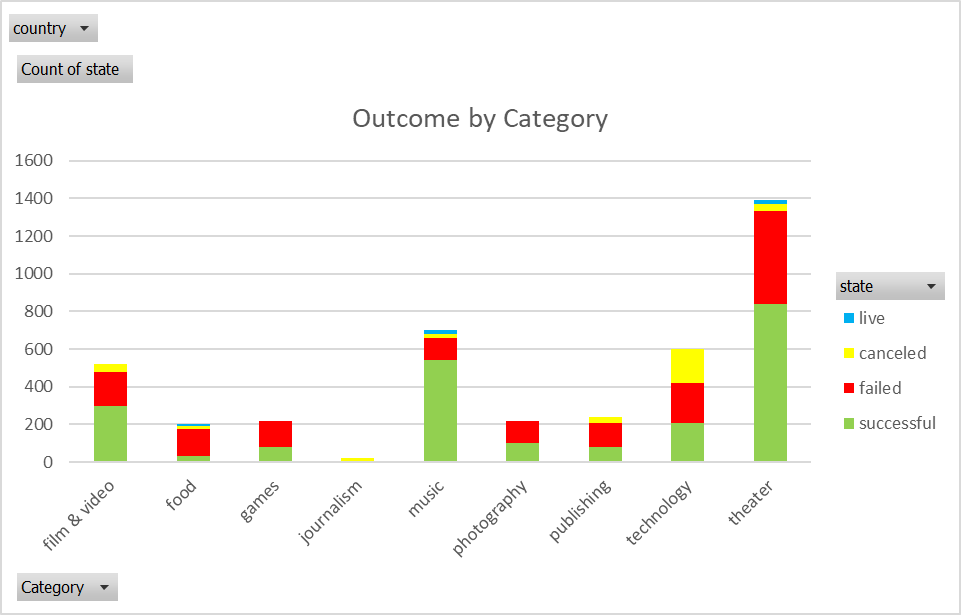
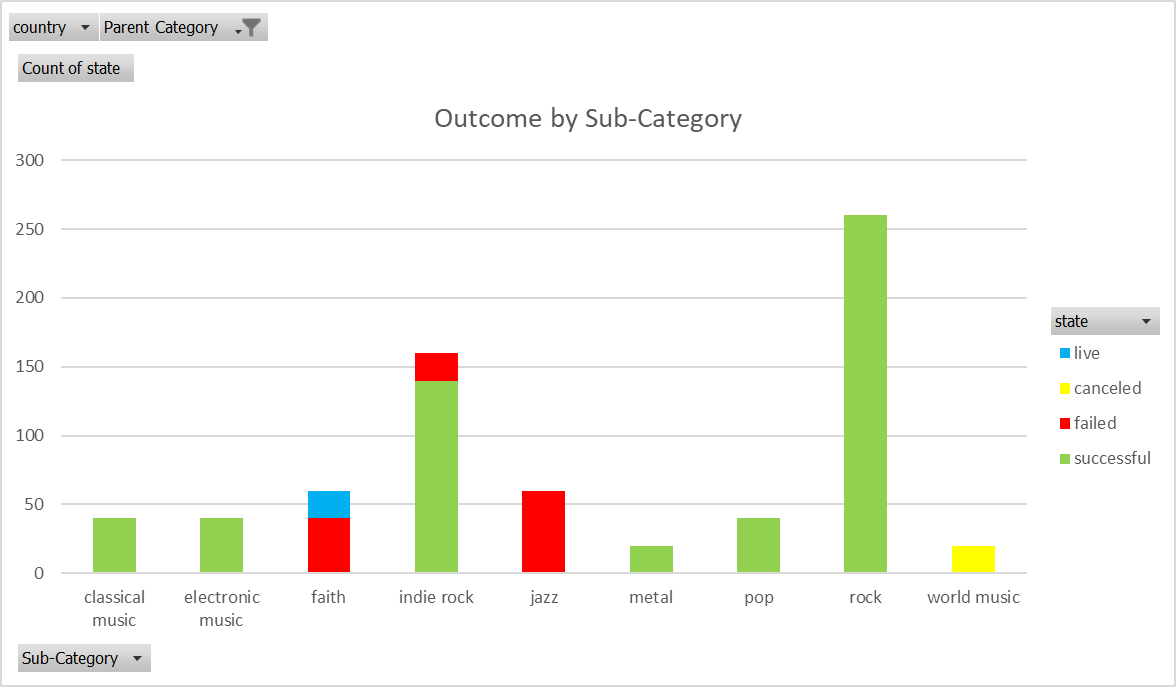
**Unit 1 Homework: Kickstart My Chart**

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

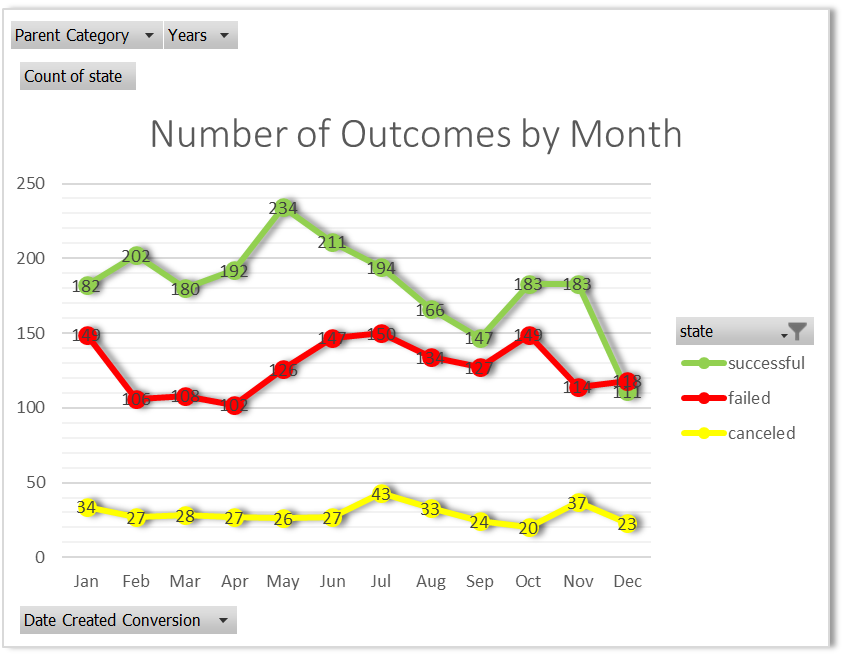
* Based on the sample of funding data for Kickstarter projects, it is apparent that there are more startups in the theater category than any other. The categories that historically have the highest proportion of successes are music, theater, and film/video.



* Drilling down into sub-categories, it is noted that within the category of music, most sub-categories have overall been successful, with the exception of jazz, faith, and world music.



* Kickstarter businesses started in December have the lowest success rate, while the highest number of successes resulted from businesses started in May.



1. **What are some limitations of this dataset?**

This data set contains only information about projects that utilize Kickstarter to collect funding to launch their business. Conclusions drawn from this study can only be applied to startups using Kickstarter.

Conclusions drawn from this study are based on a sample of the population. Given that there are over 300,000 projects in Kickstarter which have raised over $2 billion, the sample is only a small percentage of the population. We have data on 4114 startups that raised around $46 million. So, for a sample that is around 1.4% of all Kickstarter projects, it contains around 2.3% of the total funds raised. It appears that this sample contains a disproportionate amount of higher funded projects than is present in the population. Perhaps changing the sampling technique could ensure random selection of projects.

1. **What are some other possible tables and/or graphs that we could create?**

A moving average of percent funded across time could help to see how the trends in funding change over the years.

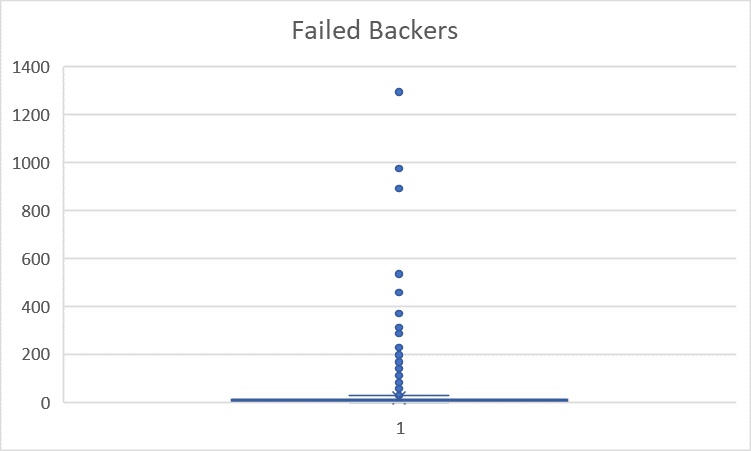
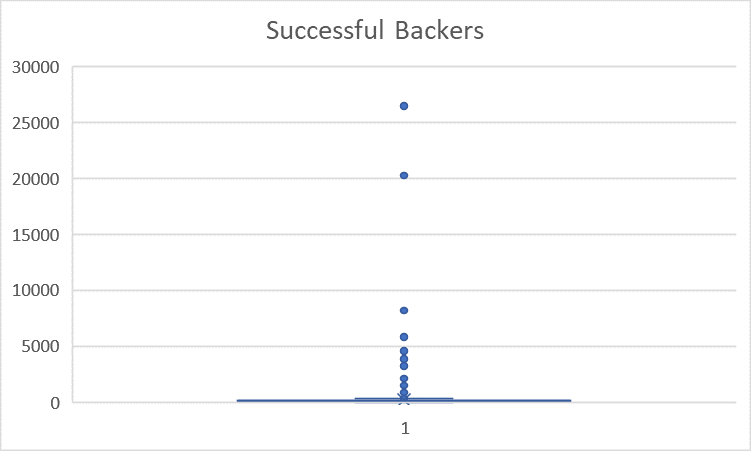
A new variable could be created by subtracting the date created from the date ended to get the amount of time donations were accepted from a project. A scatterplot and a regression analysis could indicate whether this variable is predictive of percent funded.

**Statistical Analysis**

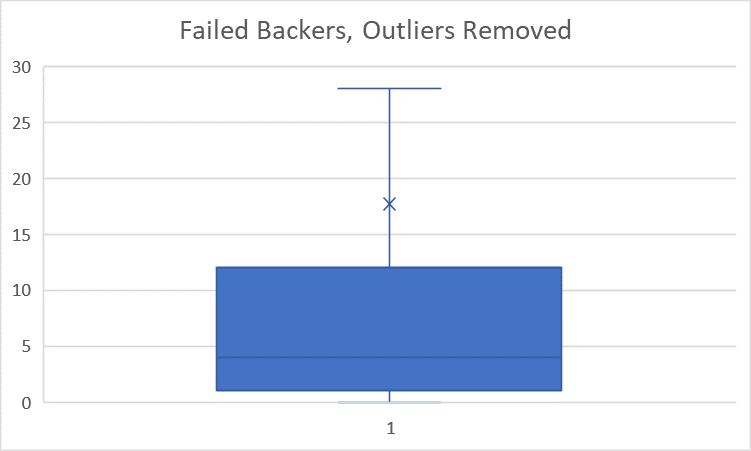
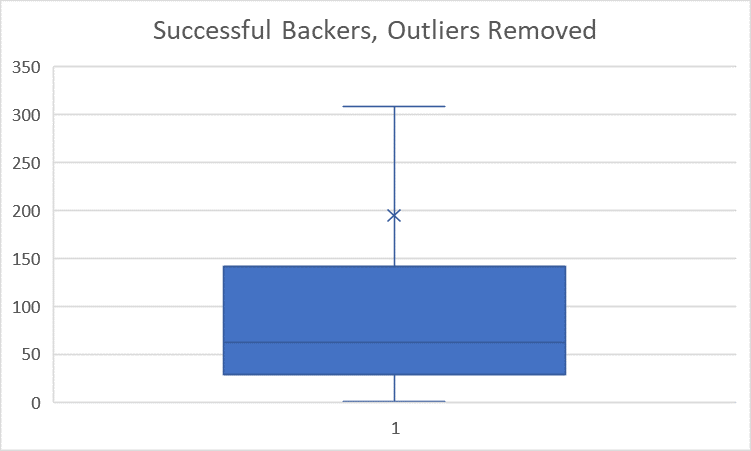
Based on the descriptive statistics, the distribution of the number of backers for successful and failed startups appears to be very skewed.



For both the successful and failed startups, the mean number of backers is much higher (3 times) than the median number of backers. That means that half of the startups were backed by fewer contributors than 1/3 of the mean number of backers. Looking at a graphical representation of the distribution, for both successful and failed startups, a box and whisker plot shows there are outliers pulling the mean away from the center of the data.



When the outliers are removed from the plot, the shape of the distribution is visible. Notice that the mean, marked with an “X” is above the 75th quartile of the data.



Because the mean is so far from the bulk of the data, the median is a better statistic to summarize how many backers contributed to a startup.