Module 1 Challenge

Written Analysis:

* Given the provided data, what are three conclusions that we can draw about crowdfunding campaigns?
  + Technology campaigns tend to be a pretty successful category of campaign to run. Out of the 96 technology campaigns that were run, 64 of them were successful (67%).
  + Theater campaigns tend to be the most popular category of campaign to run. Out of the 1,000 campaigns that were run, 344 of them were theater campaigns (34.4%) this is followed by Film & Video (17.8%) and Music (17.5%). In terms of sub-categories the most popular campaigns that were run were plays at 34.4%. No other sub-category had the sample size that plays did.
  + June and July tend to be the best months to run a successful campaign. The number of failed campaigns in June and July remains relatively similar to the rest of the year at about 30. Canceled campaigns are also comparable to the rest of the year with 3 in June and 4 in July. The number of successful campaigns, however, increases to 55 in June and 58 in July, which are increases of 17% and 23% respectively when compared to the average.
* What are some limitations of this dataset?
  + Some categories and sub-categories for the campaigns are not evenly comparable given that the sample sizes are so small. For example, journalism campaigns had a 100% success rate, but only 4 journalism campaigns were run, while theater only had a success rate of about 54% with 344 total campaigns ran. Same would apply for World Music and Audio under sub-categories. Each have 100% success rates but have only 3 and 4 campaigns that were ran respectively. It would be difficult to say those categories or sub-categories were the most successful given that the number of campaigns ran for each category and sub-category were not evenly distributed.
  + Total campaign goal varies largely between the campaigns, which makes it difficult to compare the popularity of each one fairly. Some campaigns have a goal as small as $100 while others have goals as large as $199,200. It would be much easier to reach a smaller funding goal, but that doesn’t necessarily mean it the most popular campaign. For example, the Hampton, Lewis and Ray campaign has a goal of $600 and were able to pledge a total of $4,022. While it was a successful campaign, it would be difficult to compare that to a campaign like Romero-Hoffman that had a goal of $199,200 and were able to pledge a total of $184,750.
* What are some other possible tables and/or graphs that we could create, and what additional value would they provide?
  + One graph that I think would be helpful would be a line graph that compared backer count to the average donation that could be filtered based on category and sub-category. This would allow us to see a) what campaign type people are most interested in and b) how much money they are contributing
  + Another graph that I think would be helpful would be a bar or pie graph that allows you to see what percentage of funds are coming from each country that could be filtered by category and sub-category.. By doing so, we would be able to see what campaign types could be ran in each individual country in order to have the most success.

Statistical Analysis:

* For this situation, the median would better summarize the data. The reason that would be the case is because a vast majority of both the successful and unsuccessful campaigns have a backer count that is lower than the mean. Out of the 565 successful campaigns 403 have a backer count less than the mean of 851.15. As far as the unsuccessful campaigns, 253 of 364 campaigns have a backer count less than the mean of 585.62. Essentially, the means are being heavily skewed by a few campaigns with a significantly large backer count, so the median would be a much more accurate representation of the data as a whole.
* Based on the variance and standard deviation for both successful and unsuccessful campaigns, its clear that there is more variability amongst successful campaigns. This does make sense because if you look at the data campaigns with higher backer counts tended to be more successful so you’ll have larger outliers. With that there are still some campaigns with small funding goals that were successful even with smaller backer count. Which increases the variability. On the other hand, unsuccessful campaigns were more consistent in that the backing within these campaigns was a lot smaller thus having less variability.