Crowdfunding Questions

1. Given the provided data what are three conclusions that we can draw about crowdfunding campaigns?
   1. Being a Staff pick or a Spotlight pick had no correlation with whether the outcome of money pledged would be successful.
   2. The 3 most successful parent categories are theater, film& video, and music. Each country has their topmost successful categories. Also, the most successful sub-category regardless of country is Plays.
   3. The most successful time to crowdfund seems to be in the summer months. Although, those months still had their failures.
2. What are some limitations of this data set?
   1. There is very little explanation as to what the crowd was actually funding. The categories and blurb only provided so much information. We are also assuming an average donation, when in reality one or two backers could have been a major contributor as to why certain funds actually had a successful outcome.
3. What are some other possible tables and/or graphs that we could create, and what additional value would they provide?
   1. We could probably create a table that shows correlation between whether the campaign was a success or failure based on if it was chosen as a staff pick or featured as a spotlight campaign. The whole purpose of the staff picks and spotlight feature for crowdfunding is to help the campaign get the attention of people browsing. It would be nice to see if this extra attention does in fact lead to a positive correlation of being able to reach the set fundraising goal.
4. Does the mean or median of the backers better summarize the data?
   1. The median is a better summarization of the data since it accounts for skewed distribution. The mean is better when the values have a symmetrical distribution because, one or two outliers can greatly influence the mean. Basically, one high outlier can shift the mean to the top of the distribution which in turn makes it less useful in trying to predict where the rest of the distribution points might fall.
5. Use your data to determine if there is more variability with successful or unsuccessful campaigns. Does this make sense? Why or why not?
   1. There is more variability with successful campaigns. There are roughly 200 more campaigns that were successful compared to failed. These higher numbers lead to a higher standard deviation and in turn a higher variance since they are exponentially related.