Module 1 Challenge Written Portion

**Given the provided data, what are three conclusions that we can draw about crowdfunding campaigns?**

Crowdfunding campaigns based on theatre, specifically within the sub-category of plays, are the most successful at bringing in money. This is true across countries.

Journalism based crowdfunding initiatives were the least common, however, all 4 of those launched were successful.

Campaigns with mid-range goals of 15000 to 34999 had the greatest chance of success.

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

The data fails to capture the ratio of success to failure in a concise way. Although theatre was the most successful of the crowdfunding categories, it also had a significantly large chance of failure. If a campaign has a 50/50 shot of being successful or not, that should be highlighted better in the data.

The data does not tell us why these patterns exist, such as in the case of the outcomes based on the goals. It is hard to predict causation from viewing the data on its own, and not knowing the specifics of the initiatives.

**What are some other possible tables and/or graphs that we could create, and what additional value would they provide?**

Adding on to my point from the second question, it could be beneficial to add in another table or column highlighting the ratio of success to failure. When it comes to creating and putting together crowdfunding initiatives, it would be insightful to know what projects have the greatest chance of success. A graph would be created to visualize this.

Statistical Analysis

* Does the mean or the median better summarize the data?
  + The data is positively skewed, meaning the mean is greater than the median in both sets of data. As such, it is better to use the median to summarize the data as the mean takes outliers into account and is thus not a great representation. The mean is a better measure of central tendency when the data is normally distributed.
* Variance?
  + There is more variability within the data for unsuccessful campaigns as the variance and standard deviation numbers are larger. This aligns with the reality that unsuccessful campaigns do not have one phenotype that encompasses what all of them present as. An unsuccessful campaign could have had no backers at all, or many but of whom all pledged very little. There is more stability for what encompasses a successful campaign.