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

Given the provided data, we can draw many conclusions about crowdfunding campaigns. First, we can see that the categories of film & video, music, and theater have more crowdfunding campaigns compared to other categories. As a sub-category, plays specifically has more crowdfunding campaigns compared to other sub-categories. We can also tell that different categories and sub-categories of crowdfunding campaigns differ by country with some countries having more crowdfunding campaigns focused on one category/sub-category and other countries focusing on others. For example, while most countries crowdfunding campaigns are predominantly based on the film & video, music, and theater categories, Denmark also has a high percentage of their overall crowdfunding campaigns focused on publishing compared to other countries. Lastly, we can see that outcomes for crowdfunding campaigns varied substantially based on the year of the crowdfunding campaigns. For example, some years had higher numbers of successful campaigns than others and the success of the campaigns fluctuated drastically based on the time of year.

What are some limitations of this dataset?

There are a few limitations to the dataset. Namely, the variation in types of currency is one limitation. This dataset would pose a challenge in adequately comparing currency amounts unless these amounts were converted, and the dataset also does not have a note on whether the goal and pledged amounts are in USD or in another form of currency. Also, this dataset lacks any indication of what a spotlight or staff pick means so analyzing this data would be challenging without any context. There is also no context or distinction between the differences in canceled and failed crowdfunding campaigns. Lastly, the blurbs provided on the campaigns are filled with jargon and it is challenging to understand the connections between some of them and their associated categories/sub-categories.

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

Given that we are looking at percentages for the outcomes, pie charts may be better for data visualization. For example, on some of the stacked bar charts, it is very challenging to see the “live” or “canceled” outcomes and pie charts may help with digesting the information more easily. Also, using pie charts to look at outcome with a legend of categories/sub-categories may be more informative. It may have also been useful to create individual bar graphs or pie charts by month in order to analyze which crowdfunding campaigns are more successful during different times of the year and see if there is a pattern across different years. Additionally, adding a couple of scatterplots to see if there is a trend between number of backers and a crowdfunding campaign being successful or failing could be informative. This could let us know that there is or is not a relationship between having more backers and being more successful, less backers and failing, etc.

Use your data to determine whether the mean or the median better summarizes the data.

The mean number of backers for both the “failed” and “successful” outcomes better summarizes the data. The mean is a greater representation of the middle or center of the datasets as both distributions are heavily skewed to the right.

Use your data to determine if there is more variability with successful or unsuccessful campaigns. Does this make sense? Why or why not?

There is more variability in successful campaigns. This makes sense because some campaigns will likely have more backers that donate less money while others have less backers that donate more money.