Assignment 1

**Part 1**

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

- for campaigns with goals between 15000-19999, 20000-24999, and 30000-34999 the percentage successful is 100%.

- the theater category has the highest total of campaigns while journalism has the lowest with only 4. Looking at the stacked-column pivot chart, we see a general trend of an approximate half and half distribution of successful and failed campaign across categories

- July has the highest successful cases across the years

What are some limitations of this dataset?

Limited Variables: The dataset may not include all relevant variables that are required to accurately analyze what is resulting in a higher success rate for example. The kind of fundraising approach is also a fundamental part of a successful campaign, however, this dataset only included timeline and category which may lead to an incomplete understanding of the issue being studied.

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

By calculating the duration of the campaign, from date started to date ended, we can use a double histogram to determine whether there is a relationship between the number of successful and failed cases to the duration of the campaign.

**Part 2 - Statistical Analysis**

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

- the median would better summarize the data, I've determined this by generating 2 histograms for successful and failed with all the data points and both are significantly right skewed, this means that the means are no longer the accurate representation of the data

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 with the successful campaigns, looking at the dataset, there are more successful campaigns than failed ones, which makes sense for there to be more variability given a larger number of data points. And looking at the difference in median number of backers, we see successful campaigns also tend to have more number of backers and larger campaigns generally have more variability in the number of backers due to their larger sample size. Additionally, successful campaigns may attract a wider range of backers with varying levels of interest and investment, leading to more variability in the number of backers. So yes, it makes sense.