* Given the provided data, what are three conclusions that we can draw about crowdfunding campaigns?
* What are some limitations of this dataset?
* What are some other possible tables and/or graphs that we could create, and what additional value would they provide?

Given the provided data, one conclusion that can be drawn is that more crowdfunding campaigns succeed than fail. Over fifty percent (565 out of 1,000) of all crowdfunding campaigns, across all categories, succeeded. However, a considerable number of campaigns also failed (364 out of 1,000). We can also conclude that more campaigns belong to the category of theater, specifically plays, than any other category. They also have the largest number of successful campaigns of any category or sub-category. One final conclusion that we can draw is that there is a rise in the number of successful campaigns created in June that continues into July, but this number then quickly drops around August. There is an increase in cancelled projects in August as well.

One limitation of this dataset is that it doesn’t tell us how well any of these campaigns were publicized. This could affect how many backers a campaign is able to gain, and therefore affect the outcome of the campaign regardless of date created, goal, or category. Another limitation of this data set is that we cannot always determine why a campaign is canceled. We can guess that it might have something to do with how long the campaign lasted compared to how much of their goal was funded, but we cannot know for sure. We also cannot know if there were reasons not included in our data that a project failed.

Another possible table we could add is how many backers pledged money to different campaigns. This would tell us how the number of backers varied by category of campaign. We could then create a chart that would tell us how the number of backers impacted the success or failure of certain campaigns. We could also create a table that analyzes the outcome of campaigns in regards to their initial goal amount. This would tell us if the amount of money set as a goal affected whether the campaign was successful or not. We could also create a chart that shows which category/sub-category of campaigns had more percent funded than others. This would show if any particular category had more successful campaigns that exceeded their goal. We could also create a table that compares average donation to the outcome of the campaign, telling us whether the size of the donations impacted success or failure.

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

The median better summarizes the data because it is the exact midpoint of the data. It tells us that 50% of backers for the campaigns were below that point or above that point. The mean tells us the average number, but, because our numbers vary so much, there are a few very large numbers affecting the mean. This makes our average a lot higher than the actual middle point of our data.

There is more variability in the successful campaigns. This makes sense because the difference between the median of the data for successful campaigns and the mean is larger than it is for unsuccessful campaigns.