Challenge 1-Crowdfunding Data

1. Given the provided data we can draw the following conclusions:
   1. It seems that about an equal number of concluded campaigns end in success and failure, regardless of the parent category they are in.
   2. There is a large jump of successful campaigns in the middle months when looking at all years together, maybe meaning that campaigns may be more likely to succeed if they are in the summer.
   3. The most popular type of thing to crowdfund (within this dataset) is theater, specifically plays. The number of entries for plays is the largest by a large amount.
2. What are some limitations of the data set:
   1. There are very few countries represented by this dataset, meaning it’s only useful if you want to know the trends in these specific countries.
   2. We don’t know where this dataset is from, and it could be aimed at things like theater which would be why so many entries are theater related rather than other types of entertainment.
3. What are some other tables and/or graphs that we could create and what additional value would they provide?
   1. Comparison of the types of projects and the amount of money they are asking for would be interesting. Maybe the reason plays are so numerous is because they cost less than other types of projects?
   2. A comparison of the country of the project vs. the success and failure per year could be interesting, you could compare to other economic data and see if crowdfunding success rate changes if the economy is “up” or “down” in the country the data is from.

Statistical questions:

1. Does the mean or median describe the data better?
   1. In both cases, I think the Median describes the data better. Both the successful and failed data have a small number of very high outliers which skews the data so the median will be more central overall.
2. Is there more variability within the successful or unsuccessful campaigns? Does this make sense? Why or why not?
   1. The successful campaigns have a higher variance and standard deviation, which would indicate that the number of backers is more variable than the failed campaigns. This makes sense because campaigns with fewer donors are much more likely to fail, so there is less room for variability because the numbers overall are lower.