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

From the graphs created, it appears that crowd funding campaigns created over the midsummer months a higher number of successes. That is 64% and 62% success rates during June and July, respectively.

We can see that the average success rate of crowdfunding ventures is 57%, with ventures in animation, audio, nonfiction, photography books, television, translations, wearables, and web categories having a success rate of over 60%.

Over recent years, the success rate of crown funding campaigns has increased, with funding campaigns started between 2017 and 2019 having a 63% success rate, significantly higher than previous averages. This could be a temporary blip or due to other factor during those years.

* What are some limitations of this dataset?

This data set is by no means all-encompassing. While I may try to draw information regarding the success of campaigns per category, for some categories there are only 3 listed campaigns (an a 100% related success rate). N=1000 is satisfactory when analyzing by factors such as year wherein each year has an average N of 100, but less so for factors with more categories.

We also don’t immediately have any information on which campaigns may have had outside marketing to skew the success rate. Additionally, a campaigns success in this case is defined as having met its funding goal, as opposed to having actually delivered on campaign promises.

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

I personally made graphs that included rates of successful campaigns per total number of campaigns while adjusting my categories. Additionally I tried to filter outcome by parent category, year, country, and number of backers. I also looked at values for average goal amount per category and success rate. This provided much of the information I directly discussed previously.

NOTE: I answered these questions before reading ahead to the next section of the challenge…

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

The median values appear to be the best descriptors in this case. Neither successful nor failed distributions are standard normal curves, and so standard deviation and mean can fail to express the data’s skew.

* 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 a larger variance among the successful campaigns. This does make sense, as the average and median number of backers for successful campaigns are higher. In general, the distribution of successful campaigns extends towards more backers (more backers more likely to be successful!), and it stands to reason that both failed and successful campaigns can be made from a solo backer.