Challenge Module 1

Daniel Meyerowitz

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

1. While tech campaigns have the highest ratios of successes to failures/cancelations, most of the other categories are consistent in that successes are a slight majority compared to the sum of cancelations and failures.
2. Campaigns started in June and July have a significantly higher chance of success than all other months. Summer lovin’!
3. The overall success rate of crowdfunding campaigns has increased over the time of the study, from 54% in 2010 to 63% in 2020 with a similar sample size.

**What are some limitations of this dataset?**

I’m not really sure…

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

I think the most valuable additions would be additional graphs that include percentages.

**Statistical Analysis**

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

I think the median makes the most sense, as this is an extremely wide range of data with outliers at both ends that could confound a mean calculation. The sample size is large enough that the median would be a better representation of the “average” amount of backers.

**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 successful campaigns, because the range is much higher, and the sample doesn’t include lots of 0s like the failures. Some failed campaigns had lots of backers, but they are outliers. Many successful campaigns had double digit followers, but many had thousands of followers! This is indicated in the astronomical SD (1267) compared to the SD of the failures at 961.