**Homework 1**

**Rebecca Pham**

**Due Date: Sat. Sept. 26, 2020**

**1. Given the provided data, what are three conclusions we can draw about Kickstarter campaigns?**

The first conclusion I can draw from the 1st graph is that majority of the campaigns are in "theatre” whereas there are not that many campaigns for “journalism” and the ones that are were all cancelled. Although there are more campaigns in “theatre”, “music” seems to have a better “successful” to “failed” ratio.

The second conclusion I would draw from 2nd graph is the Kickstarter campaigns with sub-category “play” has the highest number of campaigns so it seems a lot of people on Kickstarter create campaigns under “play” but more than a third of them have “failed”

The third conclusion I see from the data from the 3rd graph is that although more of the campaigns succeed than fail, a very small amount of them go live and less than 10% get cancelled.

The last conclusion I observe from the 3rd graph is that during the number of campaigns with state “failed” exceeds “successful”. For majority of the year, campaigns leaned towards “successful” than “failed”.

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

One of the limitations of this datasets is that the distribution for categories is wide. For example, there was only 24 campaigns for “journalism” and 1393 campaigns for “theater”. Although all “journalism” campaigns were cancelled it’s possible that if a higher number of campaigns were held under this category we may see more successful campaigns and the data collected may not be an accurate representation of “journalism” campaigns since the data collected under this category is low compared to the others.

Another limitation of the datasets is that the “state” of the campaign in Kickstarter could be dependant on the type of demographic that go on this platform. For example, if the demographic on Kickstarter are generally older adults with disposable incomes then we may see that “film & video” are more successful than “games”.

**3. What are some other possible tables and/or graphs that we could create?**

One additional graph that we could create is a pie graph. A pie graph would allow us to see the ratio of success to failure for certain categories or subcategory campaigns in Kickstarter. Another graph we could use is the cluster bar graph because instead of a stacked bar graph it would be easier to tell whether a category had more state “successful” or “failed”. One last graph we could look at is a stacked graph examining spotlight and state to determine if there is a relationship between if a campaign was more likely to be “successful” if there was a spotlight on them on Kickstarter.

**Bonus Statistical Analysis**

**1. Use your data to determine whether the mean or the median summarizes the data more meaningfully.**

For this data the mean would best summarize the data more meaningfully. Since we are calculating the mean and median from the # of backers in “successful” or “failed” campaigns, the mean would be more accurate to use as it includes all the data for backers. If we plot out the data, the data is almost evenly distributed except for a couple of outliers, in that case mean would be a more accurate method of summarizing the data.

**2. Use your data to determine if there is more variability with successful or unsuccessful campaigns. Does this make sense? Why or why not?**

From this data set, successful campaigns had more variability. This makes more sense because the range for successful campaigns is the number of backers needed to reach the goal amount and infinitely up whereas unsuccessful campaigns the range is from 0 to the number of backers required to reach the goal amount. In this case where the number of backers set can be infinite, this means the range for unsuccessful campaigns would be less than the range for successful campaigns.