Nathaniel Diamond

Data Visualization Boot Camp

Excel Homework

1. Given the provided data, what are three conclusions we can draw about Kickstarter campaigns?
   1. In this data set, for almost the entire year there are more successful campaigns than there are failed campaigns; and for several months there are more successful campaigns than there even are failed and canceled campaigns combines. The number of successful campaigns appears to be seasonal to some extent. It peaks in May at 236 successful campaigns, and the lowest point is in December at 111 successful campaigns. December is the only month where there are more failed campaigns than successful campaigns in this data set; there are 118 failed campaigns in December.
   2. The most common category is “theatre” at 1,393 total campaigns, and the second most common category is music at 700 total campaigns. By an extreme margin, the most common sub-category is “plays” at 1,066 total campaigns, which the second most common sub-category, “rock” only has 260. The least common category is “journalism” at only 24 campaigns in this data set.
   3. The most successful campaigns are in the “music” category at a success rate of 77%, despite being only the second most common campaign. Interestingly, the most common category is the second most successful; “theatre” has a success rate of 60%. The least successful category is also the least common, “journalism” has a 0% success rate – every campaign was cancelled.
2. What are some limitations of this dataset?
   1. I would like to see some rationale for why these 4,000 campaigns were chosen to be included in the data set. According to Statista (<https://tinyurl.com/y8xvaml4>), only about 37% of Kickstarter campaigns are successful, while in this data set the success rate is about 53%; by that metric alone it lends some doubt in my mind to this data set being an accurate representation of Kickstarter campaigns.
3. What are some other possible tables and/or graphs that we could create?
   1. A pie chart for each “state” (live, canceled, successful, failed” and what the proportion are per category, and sub-category. For example in the “journalism” category, even though it has a relatively small amount of entries, all of which are failed, while “music” is the second most popular category in the dataset and has a significant majority of successful campaigns. I would like to see each “state” and see more easily be able to see which categories are the most and least successful with percentages.
   2. A histogram depicting how long each campaign state lasts for. It would be interesting to see what the mean and variance length of time for each state is, and perhaps a box & whisker plot to easily see where each timeframe falls into the quartiles of what is to be expected of a “successful”, “failed” or “cancelled” campaign.
   3. A table to depict the average donation for each campaign and the relationship with the category and sub-category to determine which type of campaign is most likely to have the highest or lowest average donation.