Scott Barnett

Data Analysis Bootcamp

Homework #1

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

Based on the 4,114-count sample provided in the dataset, there are several conclusions that can be drawn about Kickstarter campaigns, of which three will be detailed in this report. Firstly, it would appear that plays are overwhelmingly the most popular category of project featured on the website, accounting for 26% of the sample. Secondly, the data from the goal amount-to-success rate dataset suggests that there is a negative correlation between the success of a project and the initial goal amount, with the percentage of projects canceled outnumbering the percentage of projects successfully funded when the goal amount is $50,000 or greater. Thirdly, there is evidence that certain project categories may lend themselves more to success in the Kickstarter realm. The subcategories “hardware” and “rock” featured in this dataset 140 and 260 times, respectively, and both boasted a 100% funding success rate.

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

Projects with smaller goal amounts will naturally have a higher success rate than those with loftier ambitions, skewing the data a bit and potentially drawing misleading conclusions about the kinds of campaigns that are likely to succeed on Kickstarter. Additionally, using average donation as a metric in lieu of median donation to estimate the contribution of the average backer to each project is not necessarily the most statistically sound way to determine this information, as one or two particularly generous benefactors will distort this figure.

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

Since this data can be pivoted by country, a filled map could be useful for quickly determining where certain project categories are most successful, and which projects are popular in a given country. A treemap can help visualize the categories that comprise the largest portion of the sample. Lastly, a scatterplot can display filtered data points plotted by, for instance, total amount pledged as a function of goal amount, with the added dimensions of dot size (number of backers) and conditional formatting for the color (whether the project succeeded, failed, was canceled, or is still live).