Homework 1 – Kickstart My Chart

Loretta Cortez

1. Given the provided data, what are three conclusions we can draw about Kickstarter campaigns?
   * The Categories that ran the most campaigns: Theater, Music and Technology
   * Categories that produce most success: Film & Video, Music and Theater
     1. The most successful subcategories were:
        1. Film & Video: Documentary
        2. Music: Rock and Indie Rock
        3. Theater: Plays
   * Categories that produces the most failed campaigns: Theater, Technology, Film & Video
     1. Technology produced a significate number of failed campaigns, given the total number of campaigns run in the category. Additionally, this category had the highest number of Cancelled campaigns. This leads to the conclusion that this category is not ideally suited to Kickstart campaigns.
   * Countries that had the most successful campaigns were CA, GB and US. However, these same countries also had the most campaigns overall, so this is not necessarily significant.
2. What are some limitations of this dataset?

* Calculation of the average contribution is assumed to be even, given the lack of additional information. It seems unlikely that all contributions would be even, so this information may be misleading.
* The number of campaigns per country is not evenly distributed, so comparison by country is not necessarily significant.

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

* Rather than just looking at numbers, calculating the % of successful, failed, etc. campaigns may be more instructive, since the total campaigns is variable. Creating a line graph based on these percentages may be instructive.
* Viewing the top three countries (with respect to quantity of campaigns) would normalize the data set, rather than retaining the countries with very small numbers of campaigns. A line graph for these three countries would then give valuable insight.

Bonus Statistical Analysis

In comparing the number of backers for the Successful vs the Failed campaigns, I believe the Mean has more meaning. The Mean and the Median are far apart, indicating this dataset is not evenly distributed. This makes sense, given the differences in quantity of campaigns by type. The Median only gives the middle value of the data set, which is not particularly useful here.

There is more variability in the Successful campaigns, which makes sense given the variety in types of campaigns.