HW Assignment #1

1. Given the provided data, what are three conclusions we can draw about Kickstarter campaigns?
   * Campaigns with lower goals succeeded more often than campaigns with higher goals. Campaigns with goals of less than $1,000 had a success rate of over 70 percent (71%). On the other end of the scale, the average the success rate of campaigns with goals of $5,000 to over $50,000 was 41%.
   * The success of a campaign was directly correlated with the amount of backers a specific campaign received. With the exception of a couple successful campaigns that had small goals and only a few backers, the more backers a campaign had, the more likely it was to succeed. For example, campaigns with 25 or more backers had a 86.95% success rate.
   * The Arts (theater and music) had the most success amongst most other categories. For instance, theater and music combine to have 1379 successful campaigns, equaling 63.1% of all the successful campaigns. With a success rate of 67.3%, not only did the categories of theater and music have the greatest number of successful campaigns, but also they had a success rate higher than average of the rest of the categories (40%).
2. What are some limitations of this dataset?
   * There is no data for individual backers. While we can find the average amount of money per backer, it does not tell the entire story. It is not entirely clear whether or not some of these campaigns raised most of their funds through individuals who gave a large sum of money or small donations from many people.
   * From the data provided, we cannot tell when each campaign produced their money during the course of fundraising. For instance, did a campaign raise most of its money during the first month, the middle or during the last month and is there any correlation between the different campaigns?
3. What are some other possible tables and/or graphs that we could create?
   * The success rate based on the country of origin of each campaign. It would tell us whether or not the country of origin plays a significant role in the success rate of the campaign, which could provide some insight to individuals who are not sure whether to launch in or out of the U.S. This could be done with a simple bar graph.
   * It would interesting to see if there is any correlation between the duration of a campaign and whether or not it was successful. It could be useful to see if longer or shorter campaigns have more success. This information could be pulled by using a pivot table and the result could be compiled in a separate table similar to how we plotted data point for the bonus assignment with the number of backers.

Bonus Part 2

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

In this particular instance, I believe that that the median (62) is a better indicator when looking at the successful campaigns due to two outliers. The campaign with the second most and greatest number of backers had 20,242 and 26,457 respectively. The campaign with the third most backers had under 10,000 backs (8,359). These outliers greatly skewed the mean (194.4), making the median a better representation of the “average” amount of campaign backers for successful campaigns.

For unsuccessful campaigns, the median (3) summarizes the data more meaningfully once again. Even though there are no outrageous outliers like for the successful campaigns, the vast majority of backers per campaign is under 15. For example, out of the 1,879 unsuccessful campaigns, 1,495 of them have 15 or less backers, totaling nearly 80 % of the data. The mean (19.5), does not capture the majority of the data, let alone the average of it, therefore it does not play a meaningful role in this instance.

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

In this particular dataset, the successful campaigns had a much greater variance than the unsuccessful campaigns. This is because the backers for successful campaign ranged from 1 to 26,457. These outliers drastically skewed the data and the variance as well.