Homework #1 Questions

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

* There are more successful & live (57.9%) than failed & canceled (42.1%) crowdfunding campaigns, which should encourage more crowdfunding campaigns to start.
* There is little correlation between the amount of a set goal for a campaign and the likelihood of success of the campaign. Campaigns should “high-ball” their goal between 15,000 and 45,000 since the percentage of success is relatively unchanged in this interval in order to maximize contributions.
* In primary categories with at least 40 total projects (all but journalism, which has only 4), technology has the greatest chance of success (69%). Technology based campaigns should be the most confident in their likelihood of success going into the crowdfunding process.

1. What are some limitations of this dataset?
   * There is no measure for the viability/appeal of individual ideas. For example, a campaign for desalination technology to secure a clean, sustainable water supply should receive more backers and money than a campaign to change the colors of all traffic lights to blue, pink, orange.
   * A larger point is there is no specific information about why a certain project was a success or failure. Bad ideas, poor management, human error, etc.
   * The currency information is reported in several different currencies, each with differing exchange rates between them. Meaning any conclusions drawn from the set goal or pledged amount will be incorrect until all of those numbers are converted into one currency.
2. What are some other possible tables and/or graphs that we could create, and what additional value would they provide?
   * One such comparison not done in the homework activity is between percent funded and outcome. Analyzing his would show how the likelihood of success is affected by the how much of the set goal was funded. One would think as percent funded goes up, so does the likelihood of success, but to what degree is unknown. Campaigns can use this information to judge their chances for success based on their current funding numbers.
   * Another, is a comparison between the length of time a campaign is active and the percent of their goal that is funded. If a correlation exists, then new campaigns could use this insight to judge whether they should keep the project active or cancel it based on their current levels of funding “x” number of days into the campaign.

Bonus Homework #1 Questions

1. Does the mean or median fit these two data sets more meaningfully?
   * Median, as both data sets have a large variance (contain many outliers), which makes the data set right skew far from a normal distribution.
2. Do successful or failed campaigns have a larger variance? Does this make sense? Why or why not?
   * Successful campaigns have a larger variance than failed ones. One would assume for a given campaign the likelihood of success increases with the number of backers, thus there are more high-end outliers in the successful campaign data set. We should expect a larger variance from successful campaigns because of the positive correlation between number of outliers and variance.