**Homework 1 - Kickstarter Tom Callegari**

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

**Seasonality of campaign starts**: It appears that in the spring more campaigns are started, with a gradual decline throughout the summer until the beginning of fall where a slight increase in new campaigns occur with what seems to be the beginning of the school year. There are many potential reasons for this seasonality; could it be that the human psyche is more hopeful in the spring? Or that colleges and universities encourage this as a way for students to finance projects?

**The bigger the ask, the larger the chance of failure: I**llustrated with the graph Goal Success Percentages, this dataset shows a developing mirrored linear trend between the amount a goal is set as and how much is pledged. As the amount of a campaigns goal increases so does the amount pledged in turn decrease. A big takeaway from this is to set your goal for funding reasonable and within the bounds of what might be acceptable for the type of project that you are trying to complete. As with most things, the bigger the goal the harder it is.

**Performing arts projects represent two thirds of all campaigns (within the dataset):** Within this dataset most of the campaigns would fall under a meta category of performing arts. If you add up all the film & video, music and theater category labeled campaigns they represent almost 64% of total observations. If this dataset is representative of the real population of Kickstarter campaigns, then it would be fair to say that drama geeks have embraced crowd funding; or maybe it is just that the dataset was originally queried for someone looking to start a performing arts project.

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

**It requires further investigation**: Further analysis of the relationships between variables should be conducted to more accurately draw conclusions. Both the Goal Success Percentage line graph and the goal modeled to pledged amounts linear regression illustrate potential limitations in the data integrity and/or collection techniques. Without further information on the background of the dataset it would be difficult to come to any conclusion about the data.

**Performing arts bias?:** From the graphs and tables created it shows that a significant portion of campaigns present within the dataset are to support the performing arts. Does the two thirds proportion that exists represent a biased sample of the theoretical population of Kickstarter campaigns, or was it a result of a more specific database query?

**What is with the ‘live’ projects?**: Of the 4114 campaigns listed, only 50 of them were labeled as ‘live’. This represents 1.22% of the projects within the dataset and curiously they only occur from January to March. The remaining projects also occurred throughout the year, including the first quarter, so they must also have been ‘live’, at least in the sense that they were running and open for donations. It is not entirely clear to me what the live label represents.

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

* Boxplots showing the relationship between qualitative and quantitative variables. An interesting one might be campaign category and goal amount.
* Additional scatterplots and regressions, the relationship between goal and backers would be interesting to explore.