**Unit 1 Homework: Kickstart My Chart**

1. Given the provided data, what are three conclusions we can draw about Kickstarter campaigns?
2. The dataset contained more successful projects (53.1%) in comparison to failed projects (37.2%) and “Music” has the highest success rate (77.1%) while food has the highest fail rate (70.0%) as comparing between parent categories.
3. There were subcategories where all projects got either cancelled or success such as “classical music” or “animation”
4. May has the highest success rate (60.6%) while February has the lowest failure rate (31.6%) and December is the only month that failure rate is higher than success rate.
5. It looks like the higher backers count, the higher probability of getting spotlight as well as higher change of getting success state.
6. What are some limitations of this dataset?

Sample size between each category (or sub-category) is very different, so it is difficult to compare between each parent category as well as sub-category. This also applies even we converted to the percentage because the impact of one-unit change is different if each category (or sub-categories) has different total number of data sets. For example, category, “journalism” is very sensitive to the percentage change because it has small sample size while we will not see any significant percentage change from changing one unit of “Music” category.

Therefore, it would be good to scale variables, so we can compare different variables (categories or subcategories) on equal footing.

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

Multi-level category chart and table that has both main category and subcategory labels because there are sub-categories that belong to different categories.