Challenge 1 Questions

* + Given the provided data, what are three conclusions that we can draw about crowdfunding campaigns?
  + 1. ‘US’ had the greatest number of campaigns overall per parent-category and ‘theater’ sub-category has found the most success with crowdfunding. ‘Journalism’ is the least pursued though it has usually found success when tried so far. Both ’film & video’ and ‘music’ have a moderate level of successful and failed campaigns second to ‘theater’.
  + 2. ‘Plays’ within ‘theater’ draws the largest funding successfully across all other subcategories, although its success rate is just about 50-50 against failed & canceled campaigns.
  + 3. Specific months throughout the ‘Years’ do not have any *drastic* effect on whether the associated campaigns were successful, failed, canceled, or currently live. Relatedly ‘Jul’ has the far most of successful campaigns.
  + What are some limitations of this dataset?
  + We don’t know the reason for cancelled campaigns, i.e. did the funds run out early, etc.?
  + Data set is too small: only 7 countries listed worldwide.
  + Depending on currency rate, is the backers\_count sufficient (needed or be higher or lower?) for funding a **successful** campaign in other countries?
  + What are some other possible tables and/or graphs that we could create, and what additional value would they provide?
  + Pie chart; which could show how much funding is generally allocated to each Category.
  + ‘Percent Funded’ vs ‘outcome’; shows the rate of being ‘successful’ is nearly double (over 100%) vs ‘failed’ campaigns because pledged funding surpasses the expected goal.
  + Table showing campaigns without outliers, would help minimize any unexpected trends.
* Use your data to determine whether the mean or the median better summarizes the data.
  + Median better summarizes the data here because the distribution is not quite symmetrical; majority of the data points are closer to the ~100-200 number of backers\_count. The presence of many outliers skews the data so Mean is not a good measure.
* Use your data to determine if there is more variability with successful or unsuccessful campaigns. Does this make sense? Why or why not?
  + There is more variability with successful campaigns and rightfully so because the total number of successful campaigns encompass larger data set compared to failed ones and thus derives a higher mean (average).