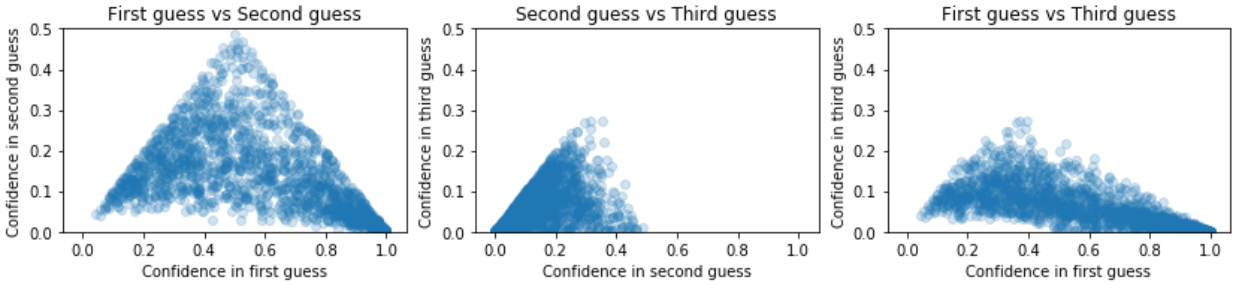
I conducted a short analysis of the Twitter data after completing my wrangling. The first thing I did was create a correlation matrix using Pandas to see if any obvious correlations exist. I noted a weak correlation between the number of retweets and the number of images in a tweet and decided to investigate. A histogram of the number of images in each tweet revealed that the vast majority of tweets had only one image in them. This means there's a lot less data on tweets with a high number of images (eg more than 1) in them, but I went ahead and created a boxplot comparing the number of images to the number of retweets. This boxplot revealed a weak linear correlation between the mean number of retweets and the number of images.

I also investigate the relationships between the image prediction model's confidence in its first, second, and third guesses for what was featured in each tweet image. I created this visualization to help me understand what was happening here:



The leftmost plot has sharp edges that could likely be modeled by a bilinear model. These edges represent points where the confidence of the first guess plus the confidence of the second guess equal one.

The second plot shows how the second and third guesses relate. In general, both the second and third guesses are low-confidence guesses – neither one appears to ever have confidence above 0.5. Again, this visualization has a sharp edge that could be modeled with a linear model.

The third plot show the first guess' confidence compared to the third guess' confidence. It's heavily skewed to the right, which indicates that the model is much more confident in its first guess than its third guess.