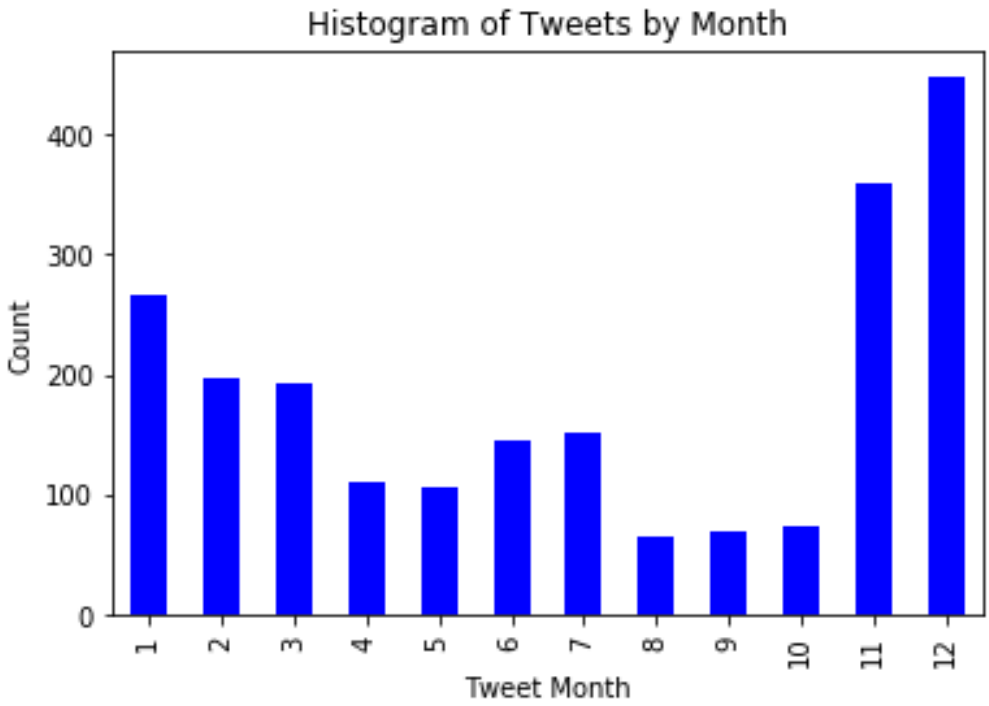
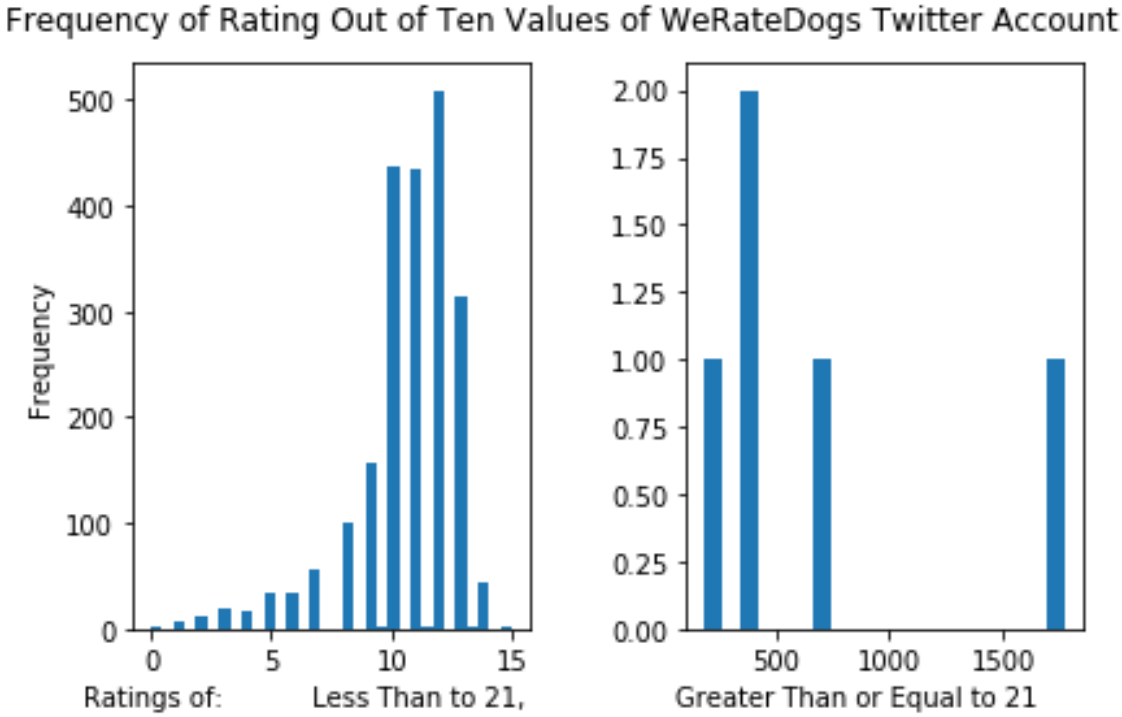
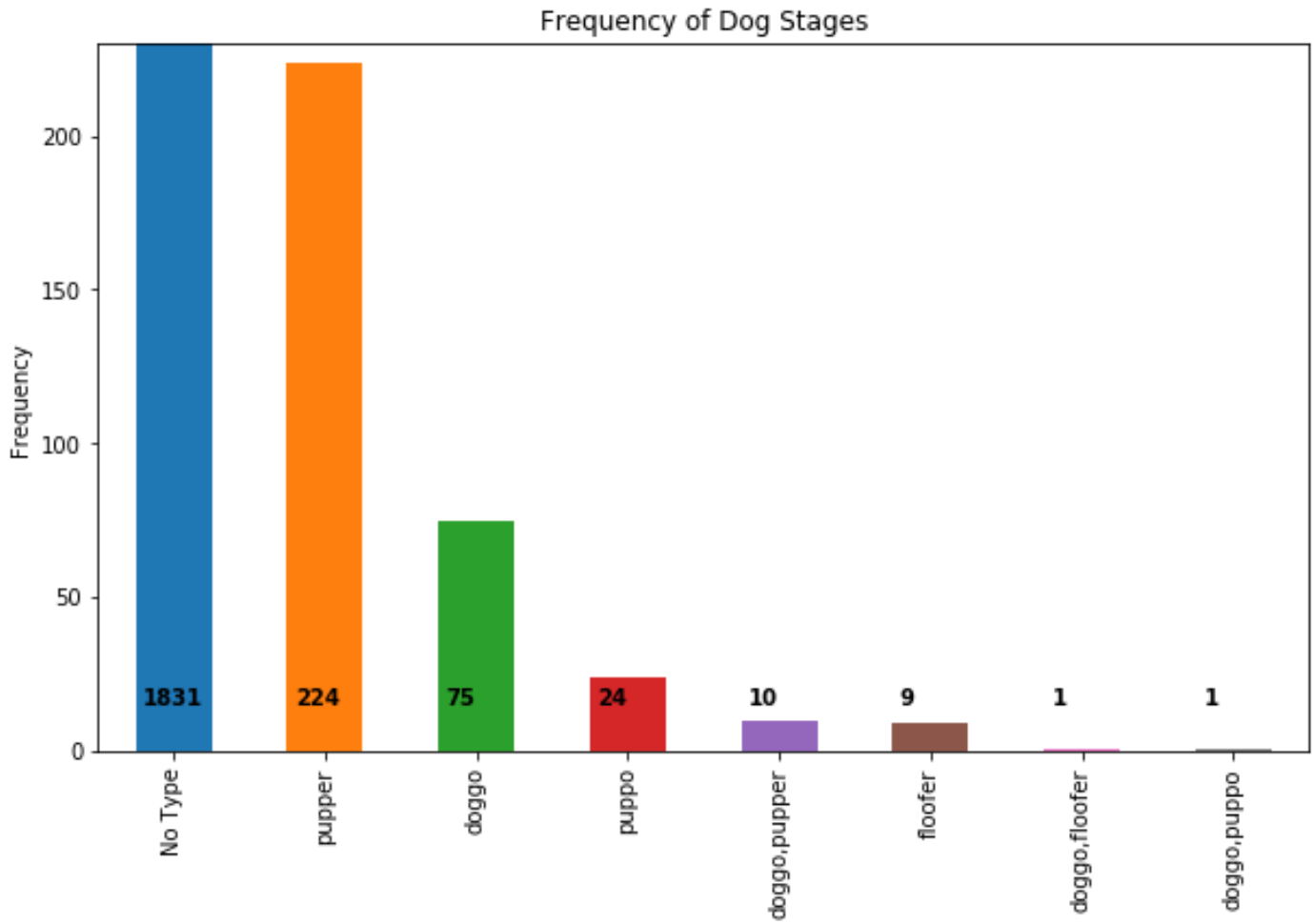
A Visualization of WeRateDogs' Tweets



I grouped each tweet by its published month. Whether because of work or perhaps seasonal activities, this person is more active in the late fall and early winter

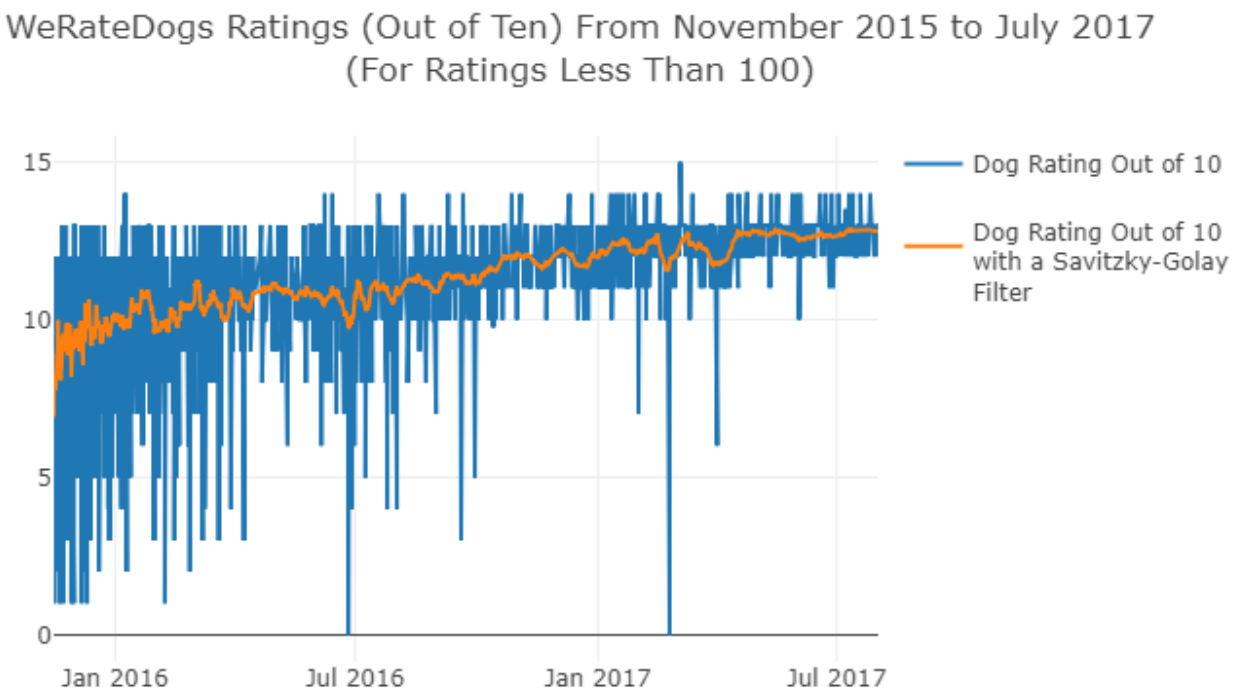


It looks as though most ratings are either a 10/10, an 11/10, a 12/10 or a 13/10. The top four ratings were so high that if shown together with the other ratings, it would be unreadable. So I separated them.

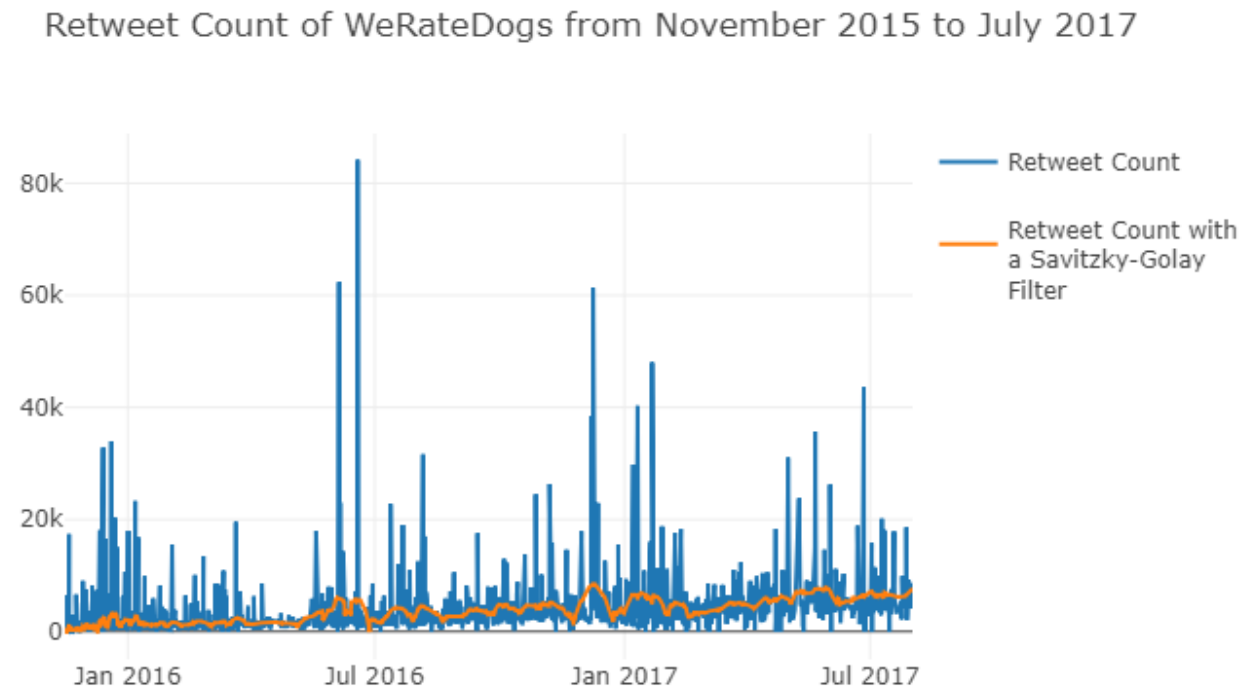


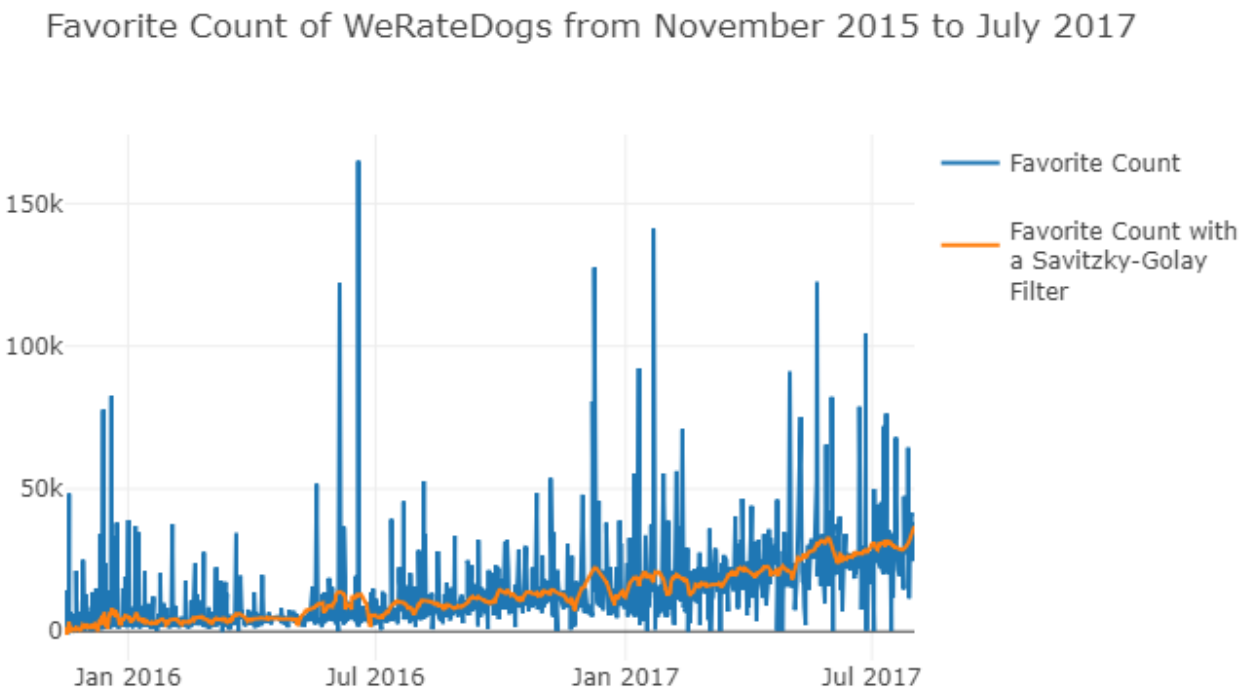
The number of dogs that had NaN, or no type, were so many that if allowed to be shown completely in this plot, the other types would be too small to see. So I cut off the plot at the 280 mark.

The following three plots were made using the Plotly library. I found it made nicer plots using a datetime value type for the x axis. It was also easier to put a filtered line plot overtop the original line plot.

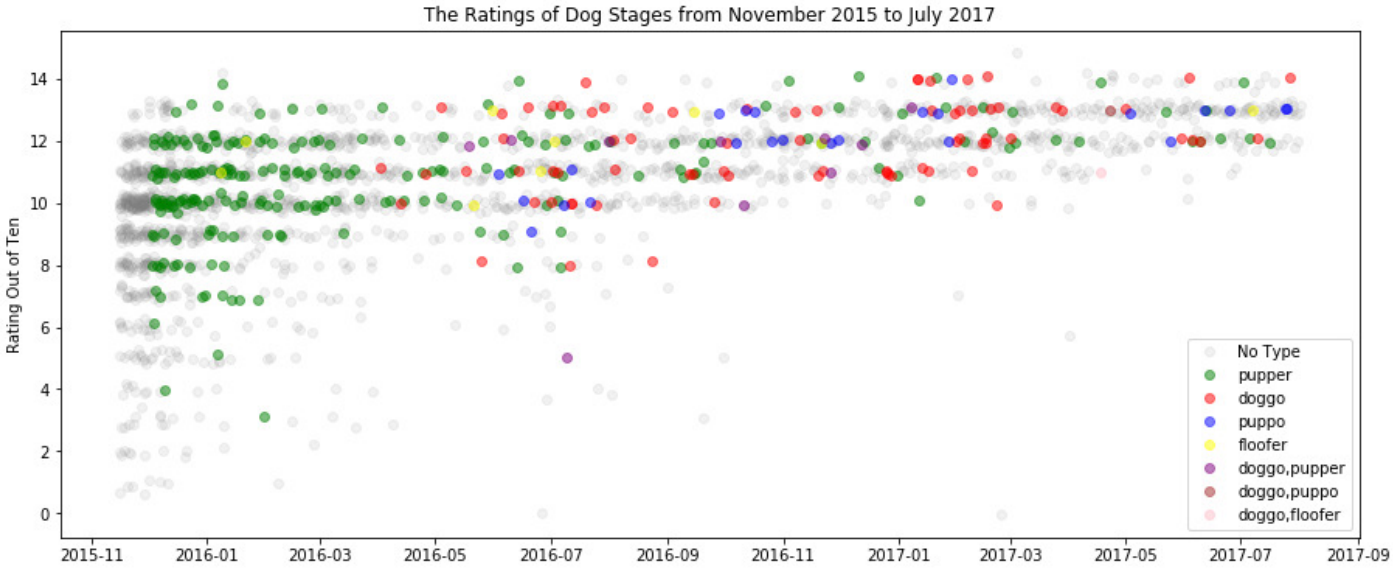


The above plot shows that this Twitter account has seen a gradual increase in the ratings it gives to dogs over the past two years. It also shows that the account was more varied in the ratings given early on.

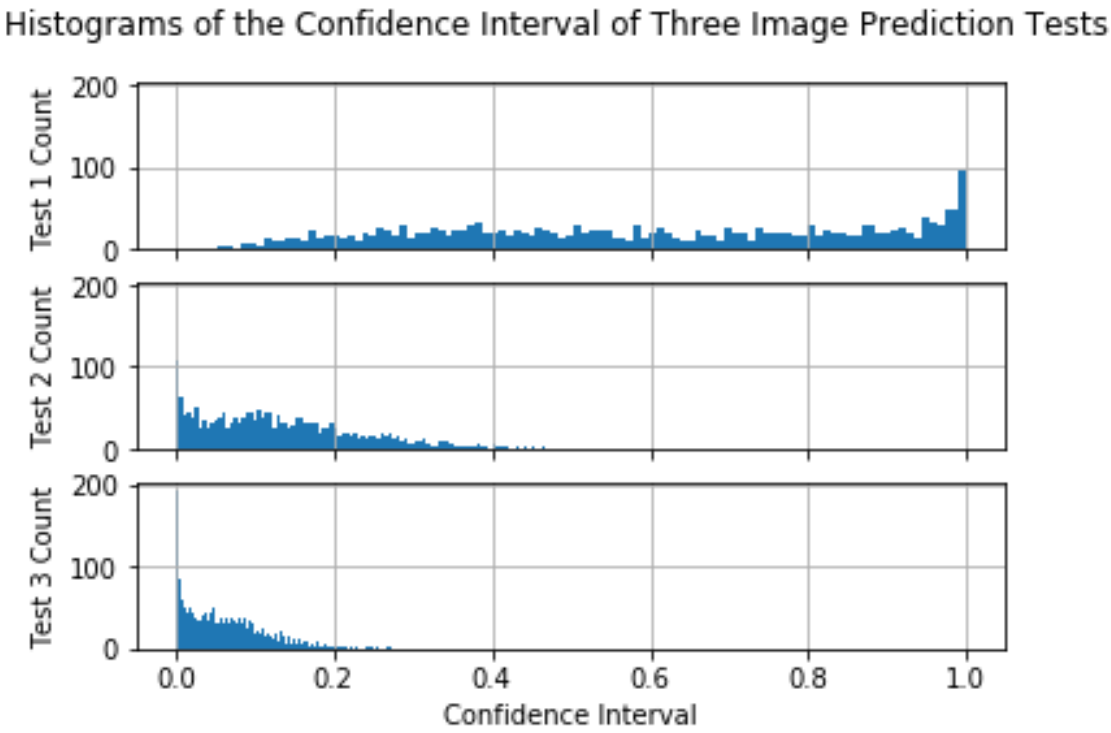




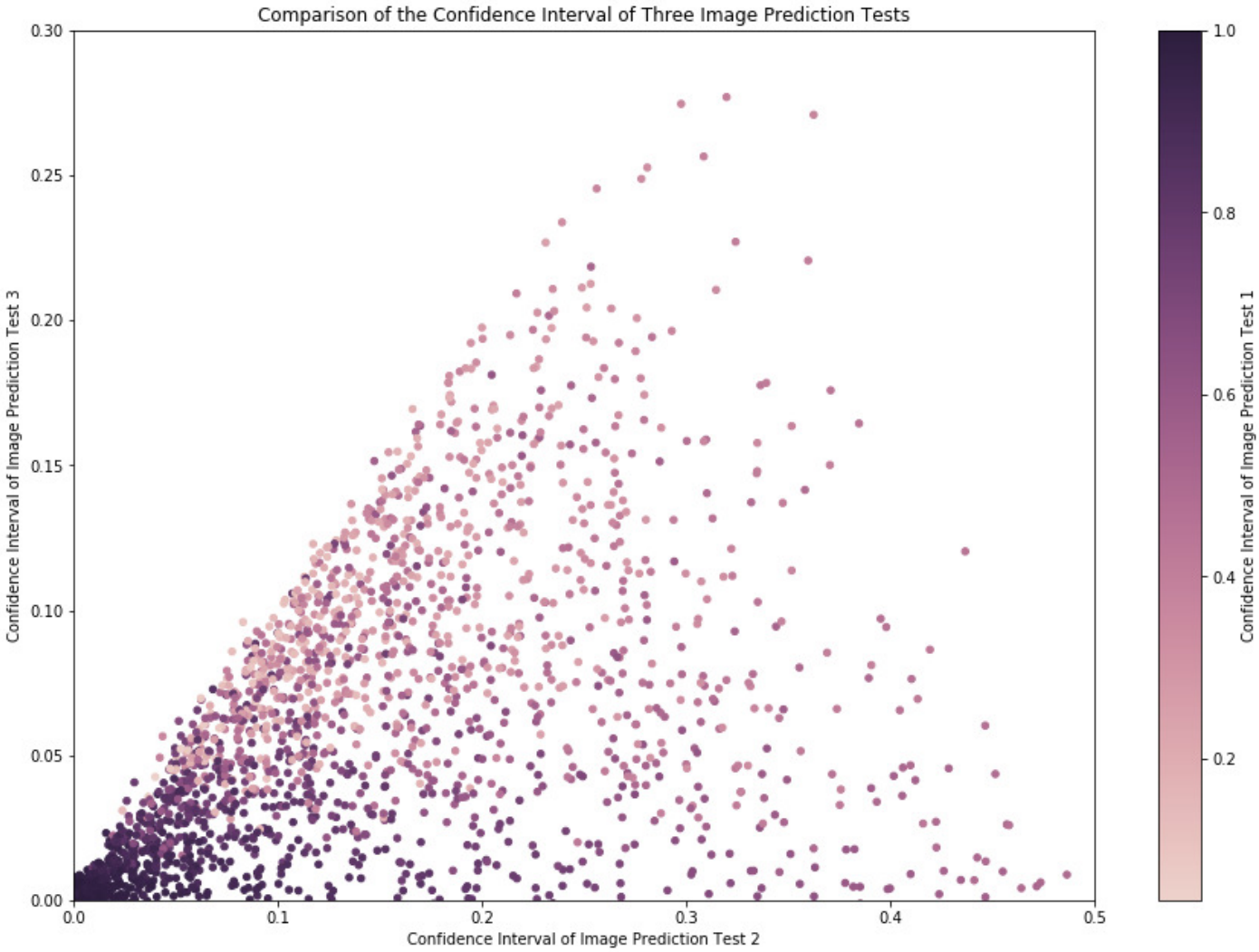
These two line plots show the increase of retweets and favorites over the past two years for WeRateDogs.



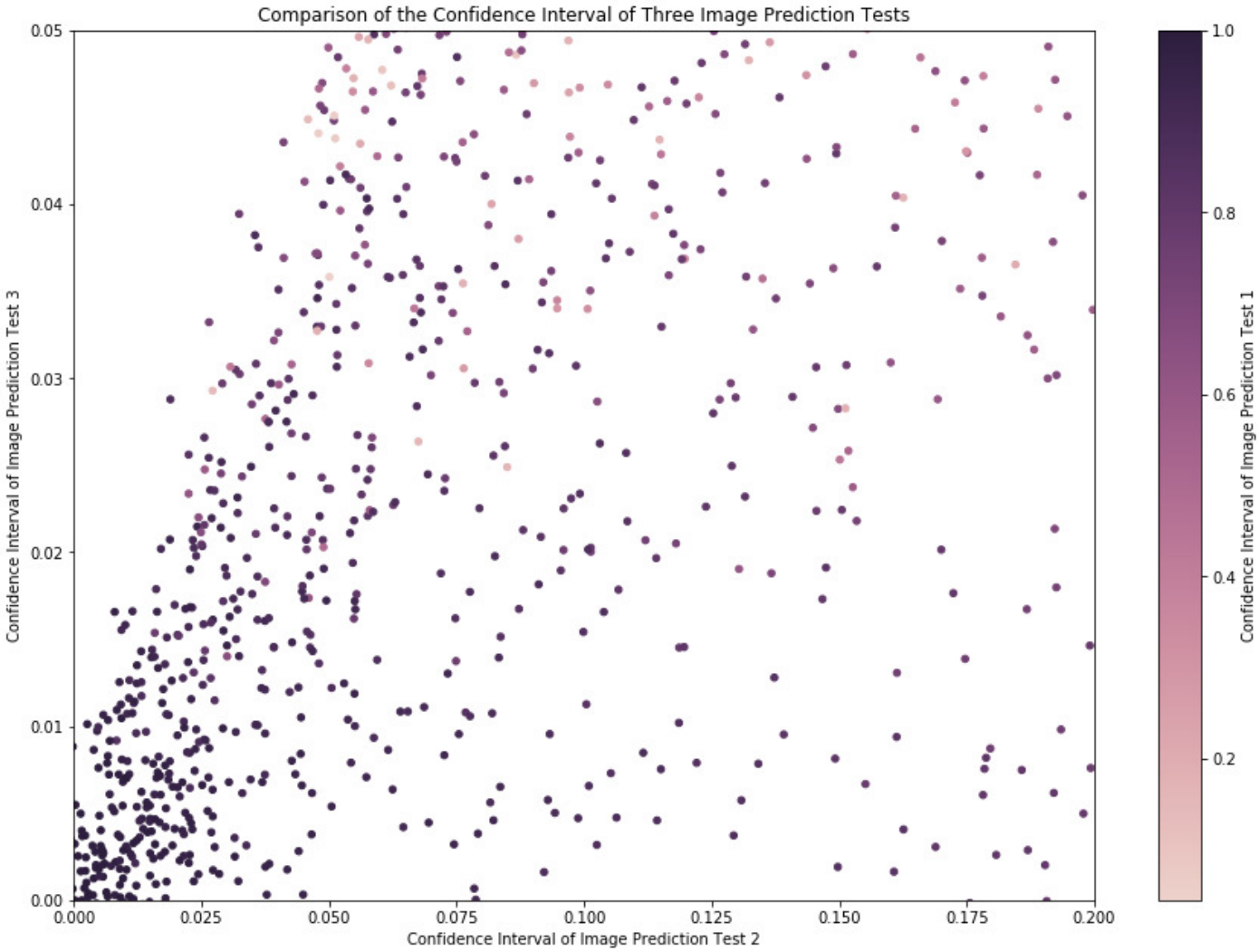
The above dot plot shows that as the giving of ratings below 10 decreased, the variety of dog stages given increased. ‘Pupper’ was solely used for the first five or six months.



Looking at these three histograms tells me that the overall confidence interval of the first test was the strongest and got subsequently lower with each following test. I want to look at these three tests in a different way.



All the data for the second and third prediction tests are in the lower confidence interval range. While the first test is mostly in the higher confidence interval range. It also seems that data in the first test that are a confidence interval of 1 or close to 1 are mostly below 0.2 for the second test and 0.05 for the third test. I want to take a closer look.



This zoomed in plot shows that a large portion of the data from test 1 that is or is close to a confidence interval of 1 is gathered in the lower bounds of the confidence interval for tests 2 and 3. Which implies to me, the higher the confidence interval for the previous test, the lower the confidence interval for all subsequent tests.