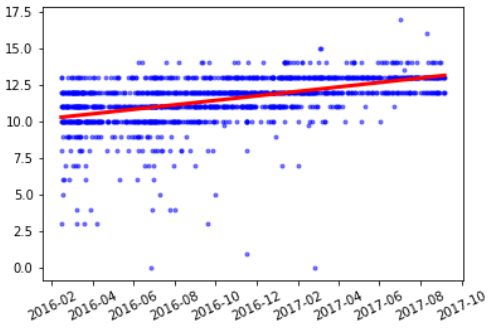
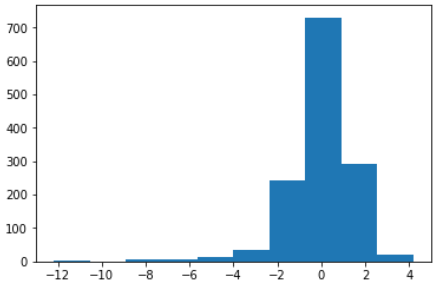
# Not All Dogs Are Created Equal.

Alex Macdonald
November 20, 2017, 4:20 PM

Contrary to popular belief, the internet is not made entirely of cats. There are also dogs (and lots of them). Among the popular internet activities of today (including memeing, and supporting various social and environmental causes, and more), looking at and rating pictures of dogs has become a common pastime for many. There is a Twitter account that goes by the name “WeRateDogs”, where other users can send in pictures of their own dogs to be rated by the WeRateDogs team. This afternoon, I quickly glanced at the most recent posts by WeRateDogs, and have found that a lot of the ratings are “13/10”. This begs the question: are these ratings for real? Of course the rating of a cute dog is subjective to taste and what the dog is doing (bonus points for being in costume), but have these high ratings always been the case for this Twitter account? Or perhaps were the ratings at one point in time more appropriate? Let’s use some data science to investigate.




Using the magic that is the internet, I was presented with the ratings from WeRateDogs dating back to February of 2016. I was interested in seeing if the ratings have always been around the ~130% level, or if these high ratings were a byproduct of showing only the cutest of the cute dogs. By plotting the scores onto a chart (as seen to the right), and calculating a trend line that shows the general trend that the data takes over time, it’s evident that there’s something fishy going on here. As time has progressed, the rating scores have shown an increase over the last year and a half. However, my mother always told me to be weary of statistical analysis (those pharmaceutical companies always try to fool you!) and so I need to verify that my instincts are correct.




How can I confirm my beliefs that the ratings are increasing over time? Linear regression (a.k.a that thing you might have learned in a statistics class long ago). Using the same data, I can perform an analysis, and based on the resulting p-value (or probability value that I’m wrong) I can confirm my intuition. And so, I created this graph on the left. It kind of looks like the data is flipping you off, but that’s exactly what we want to see: that the data follows a normal distribution. Because of this shape, the p-value can be recognized as statistically relevant, and because it was such a low number (7.79578475263e-77, otherwise known as 0. with 77 zeroes before you hit the number 7), I can conclude that the ratings are increasing over time. Whether or not this is due to a bias of cuter dogs being selected for the site or the WeRateDogs team going soft over time is another study in itself, but this is one example where what you’re seeing isn’t often the whole story.