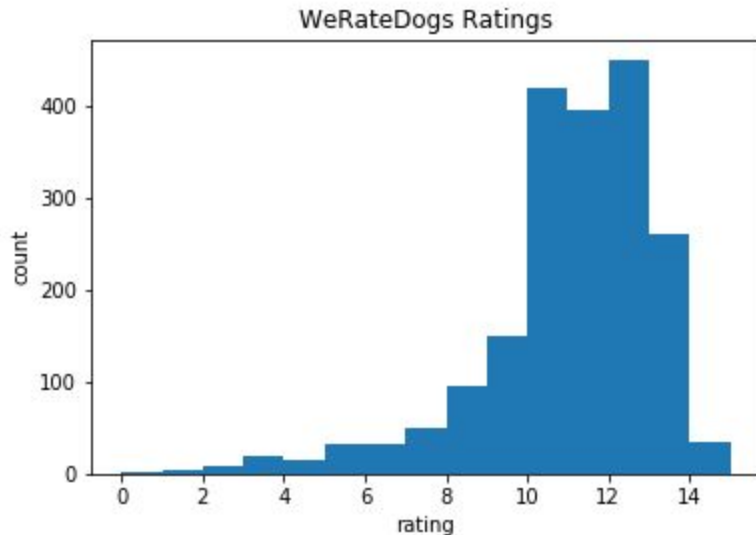# WeRateDogs Analysis

I focused my WeRateDog analysis on understanding how the site rates dogs. First, I created a histogram to understand the distribution of ratings. It is pictured below.

This chart actually removes a couple of outliers. There were ratings of 26, 27, and 75 that were actually decimal values. In the future, I would like to convert them to their proper values of 11.26, 11.27, and 9.75, but for now I removed them. There was also a 420 rating, which was actually a picture of snoop dogg. And there was a 1776, which was a dog in an Uncle Sam costume for July 4.

Besides those outliers, it appears the majority of ratings are between 10 and 13. I was curious why some pictures are rated higher than other, so I calculated the the percentage of ratings from 0 to 14 where the neural net predicted a dog.
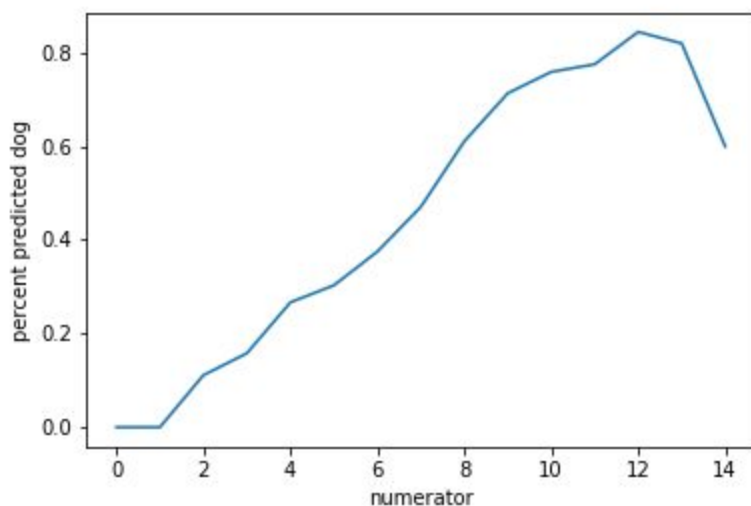
Rating increases with the probability an image is actually a dog. I double checked many images to be sure low ratings were for non dog pictures. For the most part, they were. There were a couple exceptions however. An example is below.

What kind of person sends in a picture without a dog in it? 1/10 just because that's a nice table

The site had a little fun with dog blending in pictures.

Next, I was unsure why 14s had a lower percentage of predicted dogs, so I manually explored it. It turns out the neural net struggled to label video or multimedia data. Lots of such data was tagged "website". From there, I wondered what else the neural net struggled with. A couple things stood out. First, owners took a lot of pictures of their dogs driving, which the neural net labeled "seatbelts". Second, a lot of fluffy dogs were labeled "teddy" Finally, dogs with tennis balls in their mouths were labeled "tennis balls".

In the future, I would like to predict rating numerator from tweet and dog data