# Udacity data wrangling project – data analyst nanodegree

Elad Shahar

Sept 2018

## Intro

The act report provides a closer look at some of the visualizations and analysis from the udacity weratedogs data wrangling project in the data analyst nanodegree.
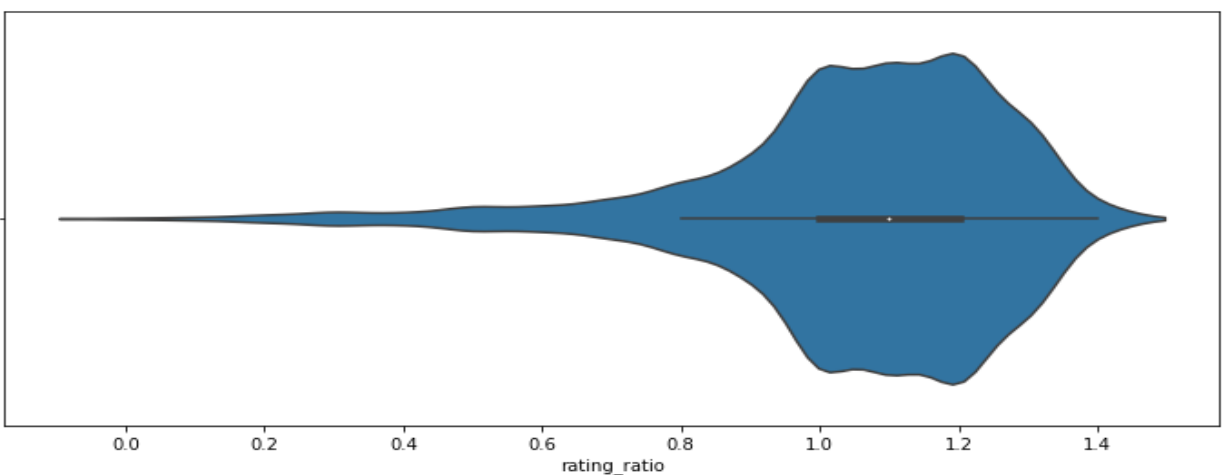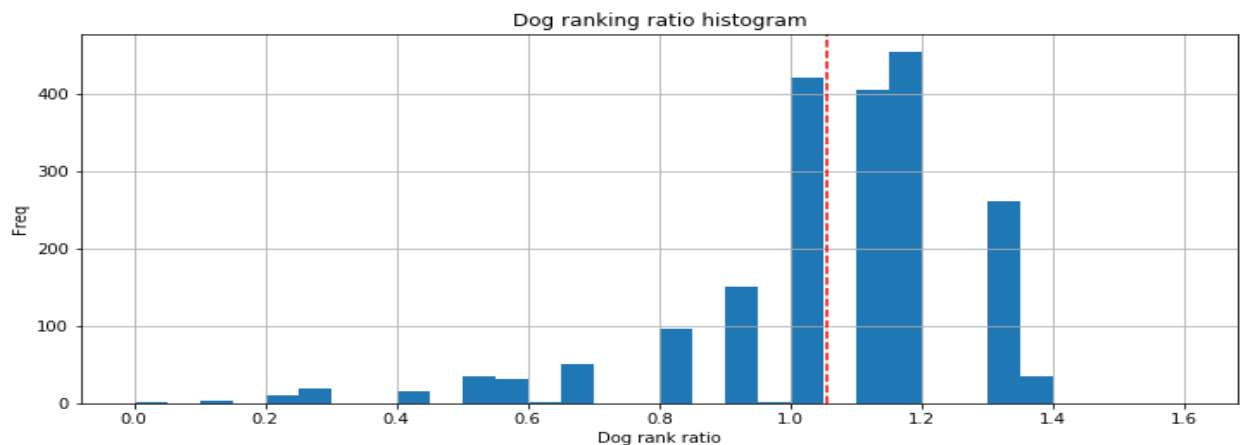
## Dog rating ratio distribution

When analyzing the distribution of rating ratios I found that the mean result is 1.05 with the median at 1.1.

This means that the average ratio between numerator and denominator for dogs on _weratedogs_ is higher than 1, part of the appeal of this crowd pleasing tweeter account with many of the tweets with a higher numerator than the denominator.

Frequent ranking ratios are between 1 and 1.2 (interquartile range) with the 95 confidence interval between 0.8 and 1.4

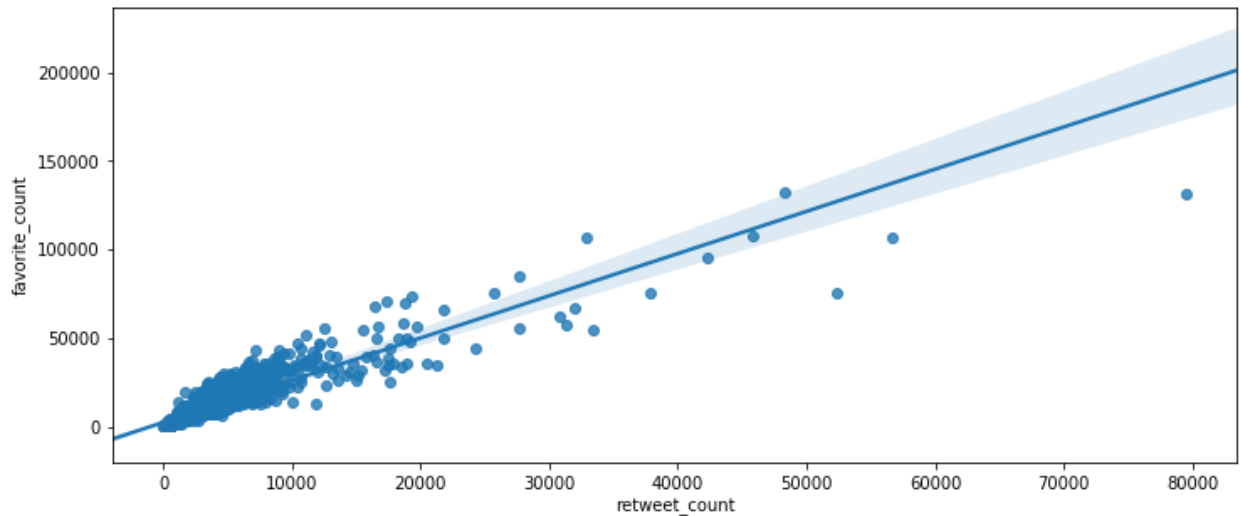This is shown in both plots I chose to use, the first an histogram and the second a violin plot

## Retweet count versus favorite count

The scatterplot below shows the correlation between number of retweets a tweet receives on weratedogs and the received number of favorite (likes) for the same tweet

The correlation coefficient is 0.91 and shows an extremely strong positive correlation
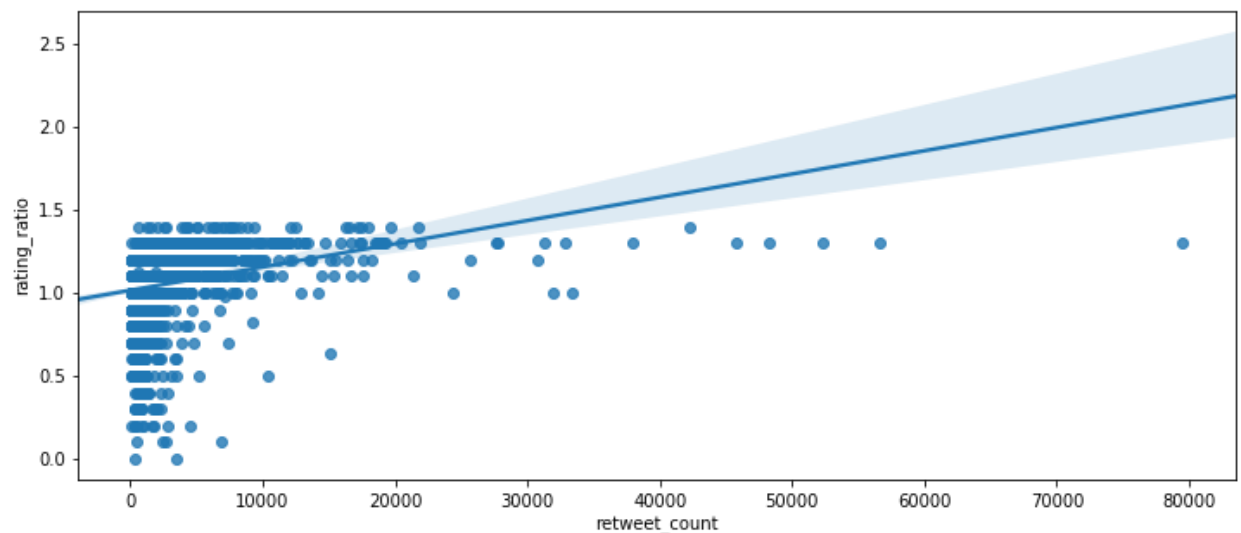
The more retweets a tweet receives, the more likely it is to have also received high number of favorite votes



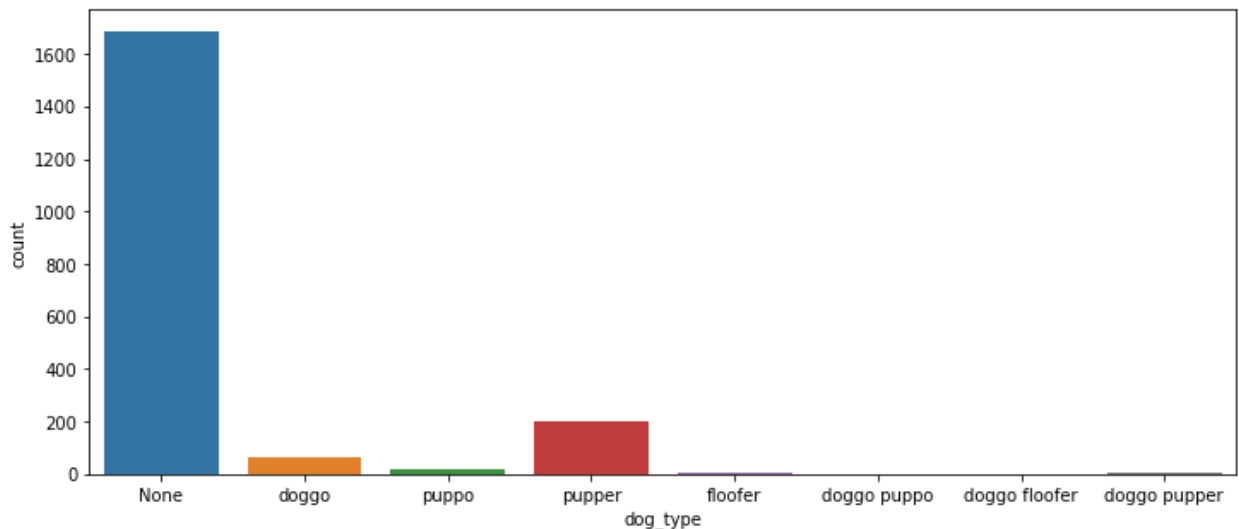## Retweet count to dog rating ratio

The plot below shows the relation between a dogs rating ratio and the retweet count.

We can see a positive correlation but not a strong one (+0.3). an interesting insight from the plot is that the tweets with the highest number of retweets are all for dogs rated 1.3 and above. No tweet passed 35000 retweets if it did not receive a high rating ratio.
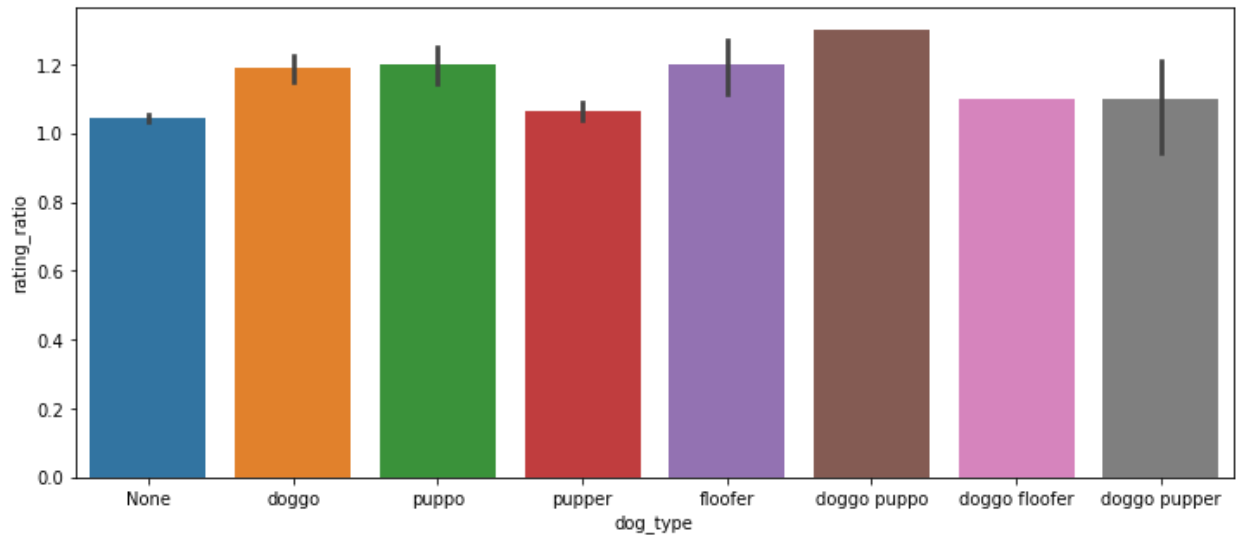
## Distribution by dog type

The plot below shows the distribution of dog types in the dataset, we can see that most dogs are unassigned to either 'doggo' , 'puppo', 'pupper or floofer. From those that are assigned the most common assignment is 'pupper' followed by 'doggo'.
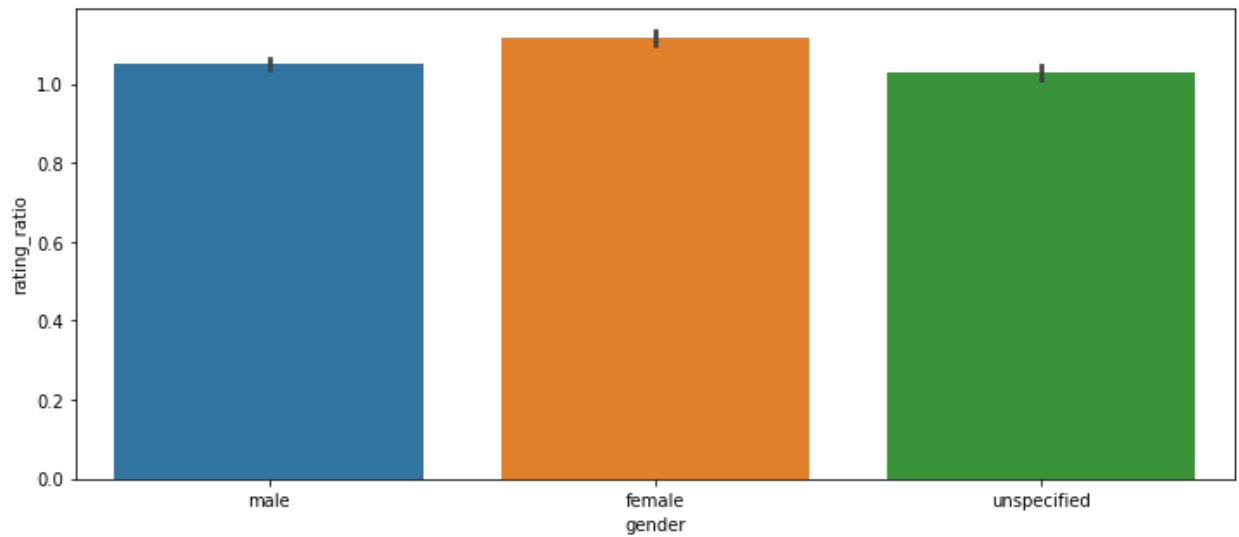


## Dog rating ratio by dog type

The distribution of dog type did not provide sufficient insight so I decided to further explore dog types in relation to the rating ratio. Interestingly here we can see that dogs that are unassigned have a lower average rating to unassigned dogs. From the assigned dogs the types with the highest average rating ratio are 'doggo', 'puppo' and 'floofer'. The most numerous assigned dog type 'pupper' surprisingly has the lowest average score from the dogs assigned a dog type, very close to the average score of the non assigned dogs.

**Rating ratio by gender**

Further exploration of rating ratio in relation to gender can be seen in the plot below.

Females receive higher rating on average, with males second. The lowest average rating ratio belongs to dogs with unspecified gender.



Thank you for reading!