# Analyzing and Visualizing WeRateDogs Twitter Archive after Data Wrangling

By: Gregory Gardner

Introduction

I have analyzed and visualized a twitter archive for a popular twitter account called WeRateDogs. WeRateDogs twitter page features pictures of dogs along with ratings. The following is an analysis, coupled with visualizations, of twitter data for the WeRateDogs twitter archive.

Analysis and Visualization
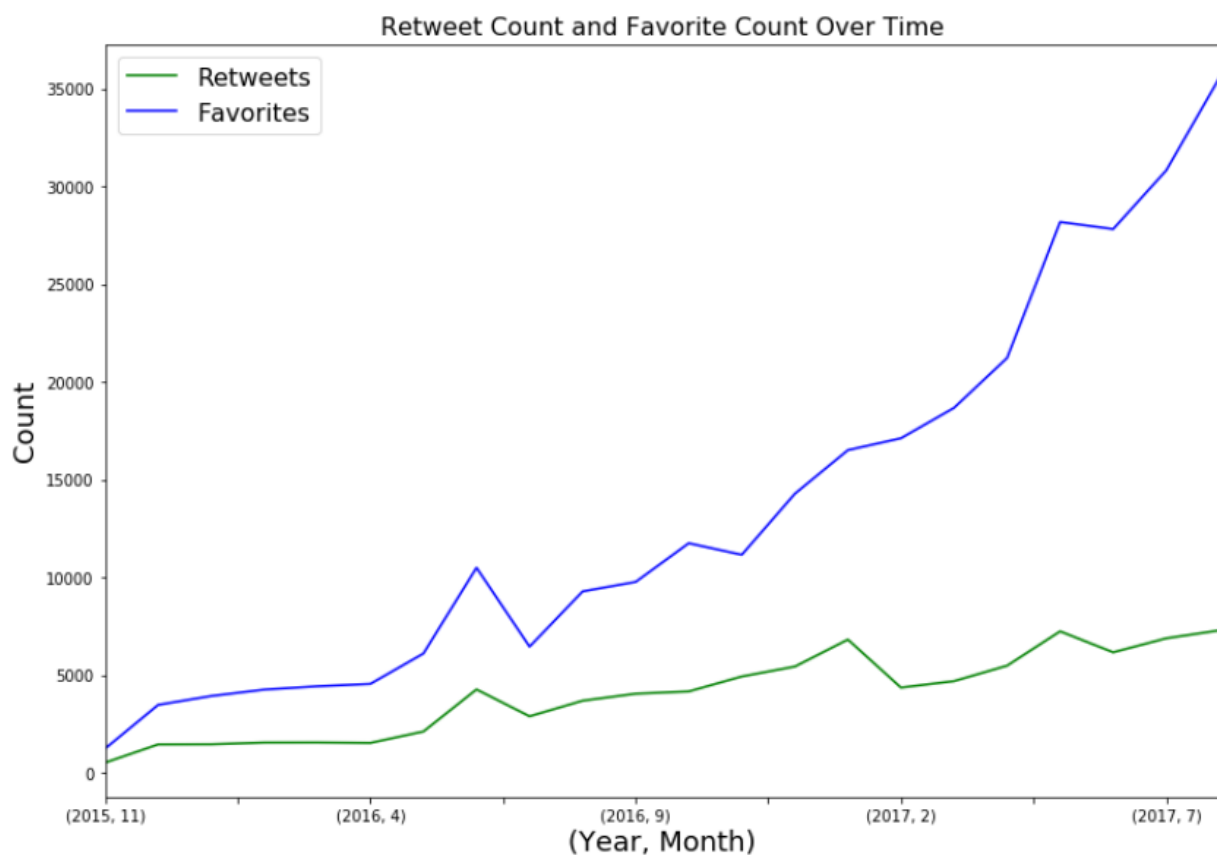
The first graph here shows the retweets and favorites over a year and a half time period. Both retweets and favorites increase over time, however, favorites increases at a much higher rate.

Most Common Dog Names
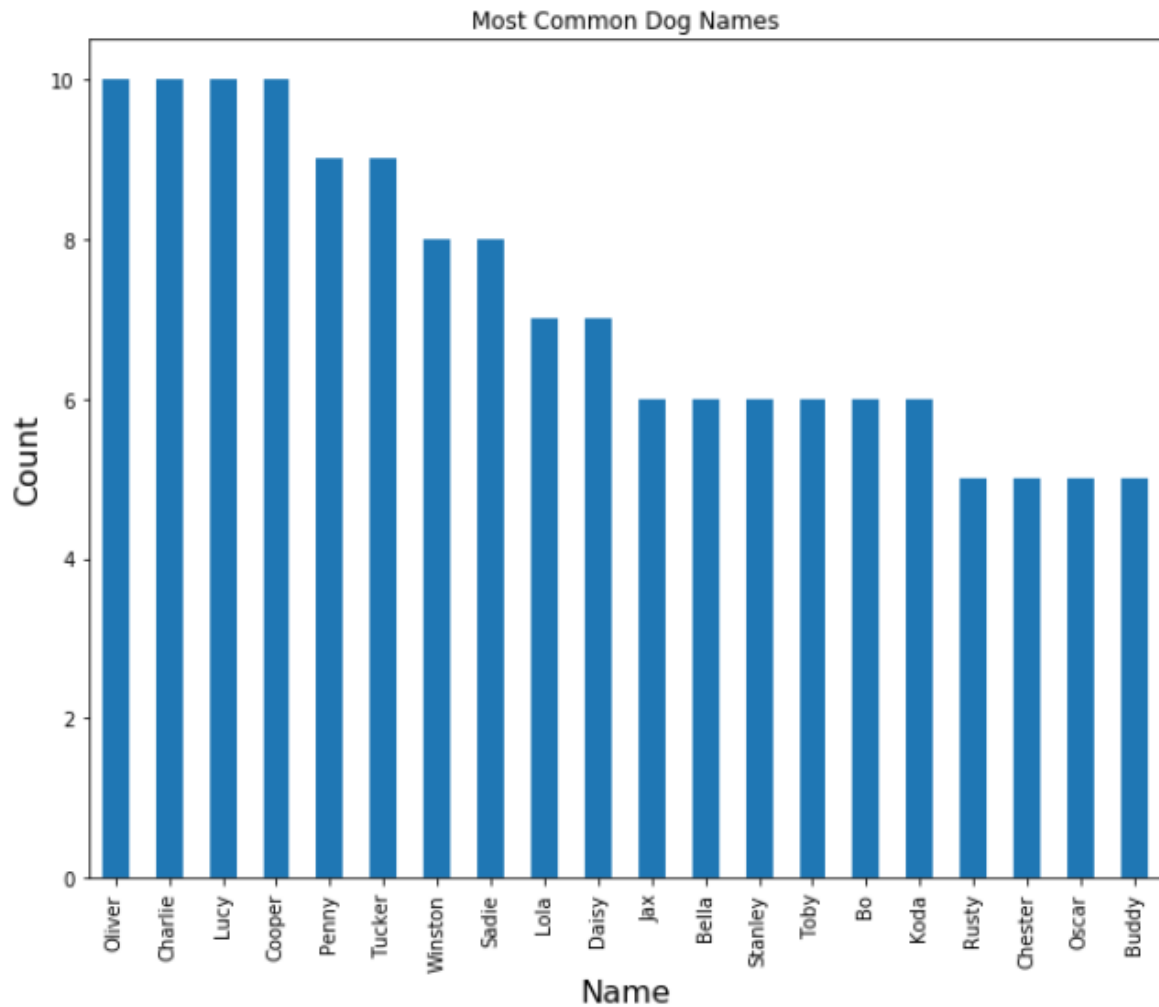
The following bar graph shows the twenty most common dog names to have been tweeted on WeRateDogs. Oliver, Charlie, Lucy and Cooper are the most common dog names to have been tweeted on WeRateDogs with ten each.

Boxplot grouped by dog_stage

rating_numerator
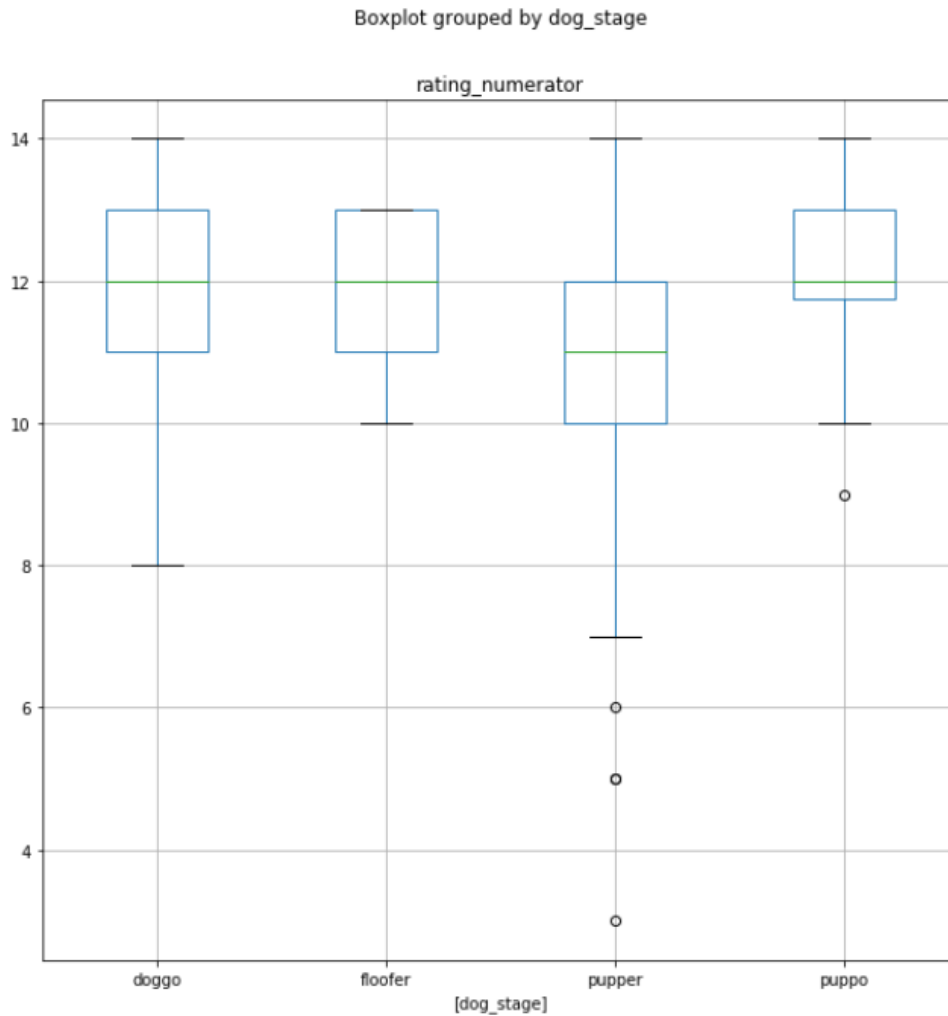
The following box plot plots the dog stage with the ratings they have received without any outliers that may have been present.

dog_stage	count	mean	std	min	25%	50%	75%	max
doggo	67.0	11.970149	1.476792	8.0	11.00	12.0	13.0	14.0
floofer	8.0	11.875000	1.125992	10.0	11.00	12.0	13.0	13.0
pupper	218.0	10.706422	1.766575	3.0	10.00	11.0	12.0	14.0
puppo	24.0	12.041667	1.267629	9.0	11.75	12.0	13.0	14.0

Pupper is the most common stage of dogs tweeted while floofers are not tweeted much at all. Puppers have the lowest average rating than any of the other stages. We cannot be sure if that is because they are more common on WeRateDogs or if that is because puppers are generally younger dogs while doggos are generally older in age.

Dog Rating Distribution
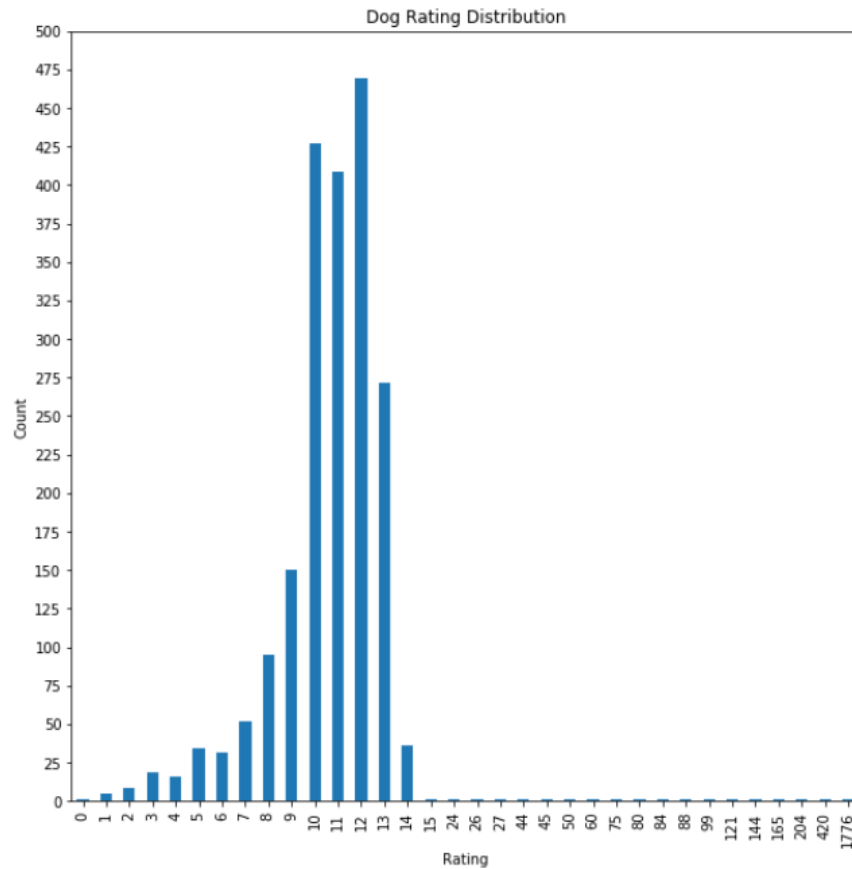
This graph shows the count for each rating in our data set.

```
0          1
1          5
2          9
3         19
4         16
5         34
6         32
7         52
8         95
9        150
10       427
11       409
12       469
13       272
14        36
15         1
24         1
26         1
27         1
44         1
45         1
50         1
60         1
75         1
80         1
84         1
88         1
99         1
121        1
144        1
165        1
204        1
420        1
1776       1
Name: rating_numerator, dtype: int64
```

As we can see, dogs rated 10 out of 10, 11 out of 10, and 12 out of 10 are the most common. The ratings above 15 could be a typo in a tweet or an error in our data set. We cannot be sure of that but it is highly unlikely that a dog received a 1776 rating.