Analyzing and Visualizing WeRateDogs

By Ryan Reardon

Introduction

Maybe you have Twitter account; or, perhaps read or heard about the 2016 interaction with another Twitter user, when Matt Nelson; founder of WeRateDogs purposefully misnamed him "Brent" as is common in Weird Twitter, spawned the catchphrase "They're good dogs, Brent", which became one of the biggest memes of 2016.   Or used phrases puppo, pupper, doggo, or floofer in an internet conversation about a dog.
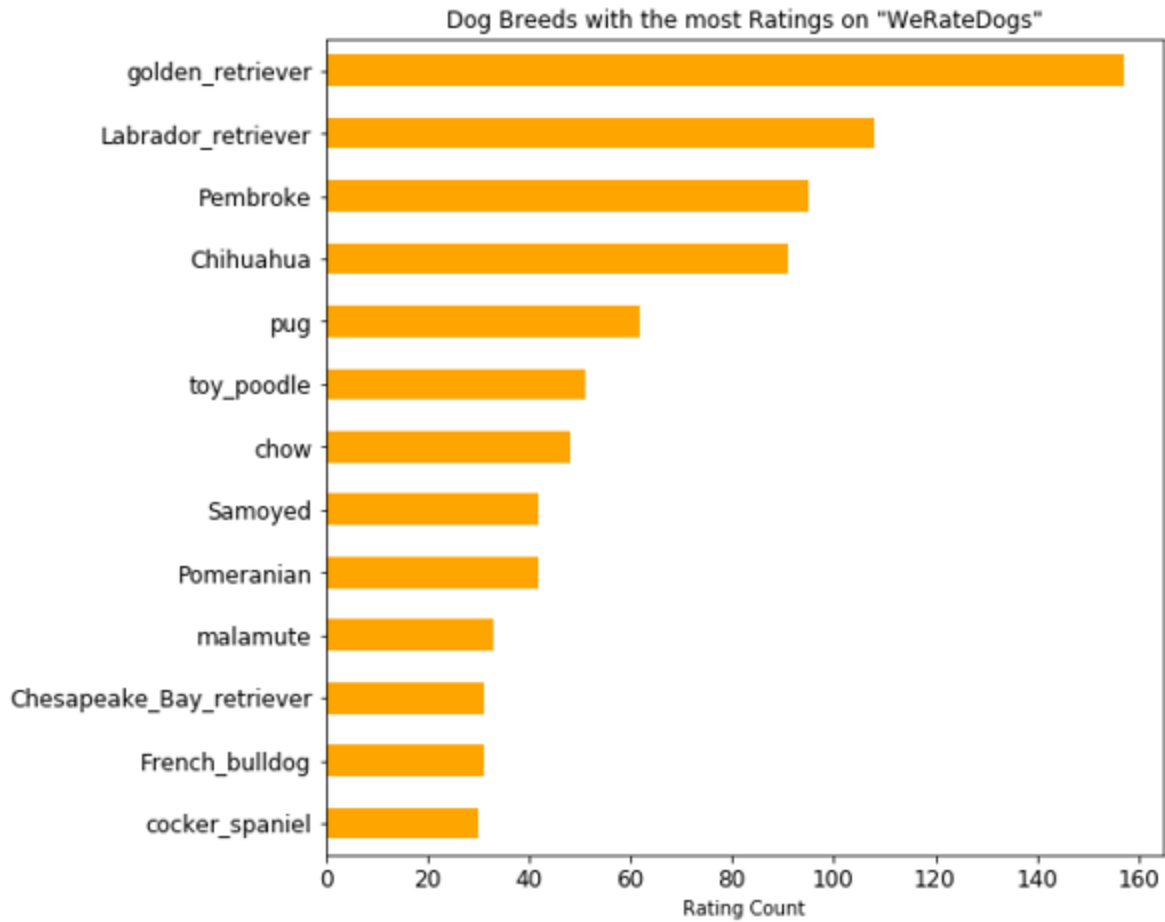
WeRateDogs is a Twitter account with over 4 million followers that rates people's dogs with a humorous comment about the dog. I was started 2015 by college student Matt Nelson and has received media coverage for its popularity and attention drawn to social media copyright law when it was suspended by Twitter.

WeRateDogs asks people to send photos of their dogs, then tweets selected photos rating and a humorous comment. Dogs are rated on a scale of 1 to 10, but almost always exceed the rating of 10, such as 13/10. After all, don't all dogs exceed a rating of 10.
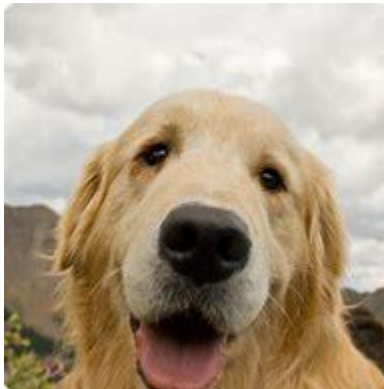
Dog Breeds Relationship to other variables:

Let's look at the Dog breeds with the most ratings!

```
golden_retriever          157
Labrador_retriever        108
Pembroke                   95
Chihuahua                  91
pug                        62
toy_poodle                 51
chow                       48
Samoyed                    42
Pomeranian                 42
malamute                   33
French_bulldog             31
Chesapeake_Bay_retriever   31
cocker_spaniel             30
```

Dog Breeds with the most Ratings on "WeRateDogs"

- These are dog breeds with the most ratings. Golden Retriever has the most at 157 followed by Labrador Retriever at 108.
- Is this due to the algorithm recognizing the breed easier or were these more popular dog breeds?

Let's look at the top 10 dog breed predictions!

	dog_breed	confidence_prediction
1845	komondor	0.999956
631	chow	0.999953
1129	Labrador_retriever	0.999885
1519	chow	0.999876
980	chow	0.999837
357	dalmatian	0.999828
555	chow	0.999823
260	Old_English_sheepdog	0.999715
1383	pug	0.999365
1131	Great_Dane	0.999223

- Top 10 dogs with the highest prediction of breed are dogs that are easily distinguishable; basically, dogs with looks that standout out.
- The Golden Retriever isn't on the list, but other dogs from the most rated are on the list and most have odd looks; the exception is the Labrador Retriever.
- The dog breeds with the most ratings is more so do to the breeds popularity and less with algorithm, since these photos are sent in by users and rated.

Let's look at dog breeds average rating compared to their count!

Earlier we looked at dog breeds with the most ratings and that seems related to popular dog breeds amongst ownership and the Golden Retriever is clearly the most popular among dog owners at least the most uploaded photo on WeRateDogs.

```
Bouvier_des_Flandres    13.000000
Saluki                  12.500000
briard                  12.333333
Tibetan_mastiff         12.250000
Border_terrier          12.142857
standard_schnauzer      12.000000
silky_terrier           12.000000
Gordon_setter           11.750000
Irish_setter            11.750000
Samoyed                 11.690476
```
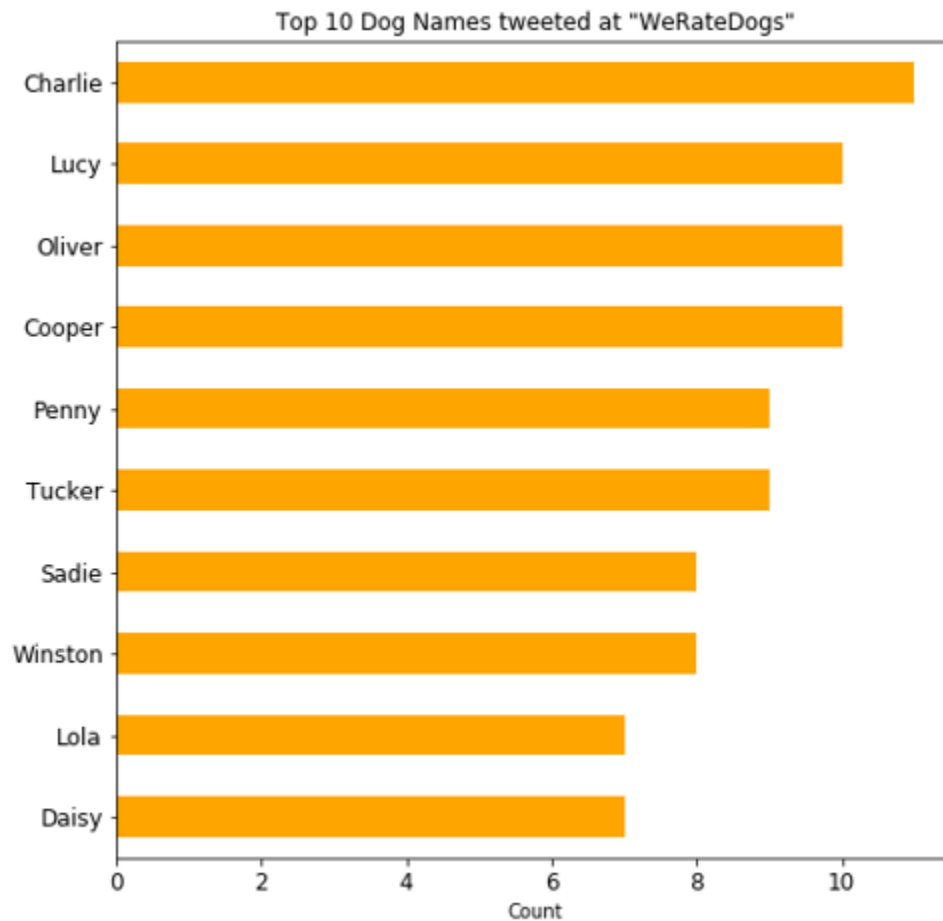

- Comparing Top dog counts or most rated to highest average rating is surprising. Again, dog breeds with the most ratings common dog breeds among owners that probably get more uploaded photos.
- Less common dog breeds have higher averages.
- Clearly WeRateDogs users' rate fewer common dogs higher, but they do have the motto "All dogs deserve a rating higher than a 10"
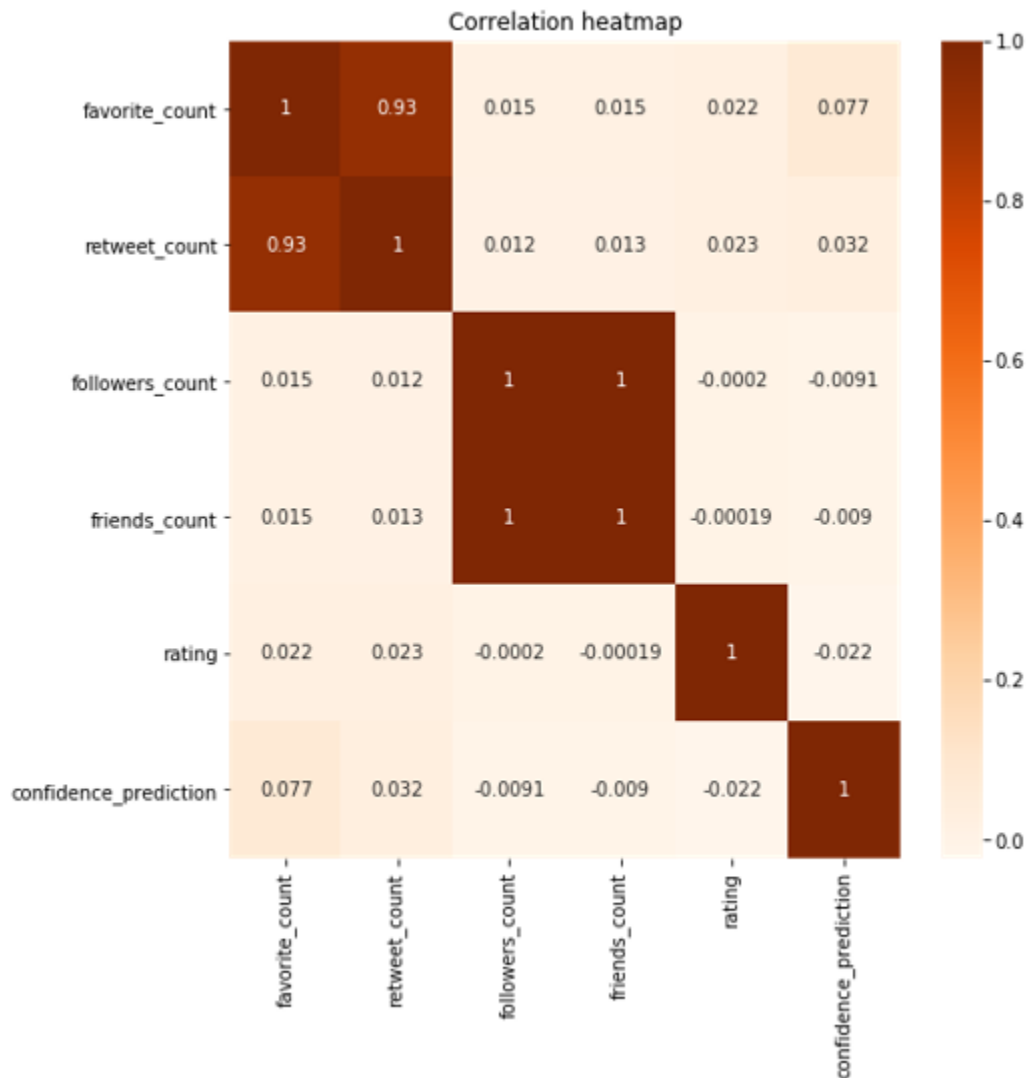
Let's look at the top 10 dog names mentioned in a tweet for fun!

```
Charlie      11
Lucy         10
Oliver       10
Cooper       10
Penny         9
Tucker        9
Winston       8
Sadie         8
Daisy         7
```


Top 10 Dog Names tweeted at "WeRateDogs"

- Plotted this for fun and "Charlie" is mentioned the most in tweets.
- Probably due to Brent the archnemesis of WeRateDogs, which both reconciled over "Charlie" the puppy of Brent.
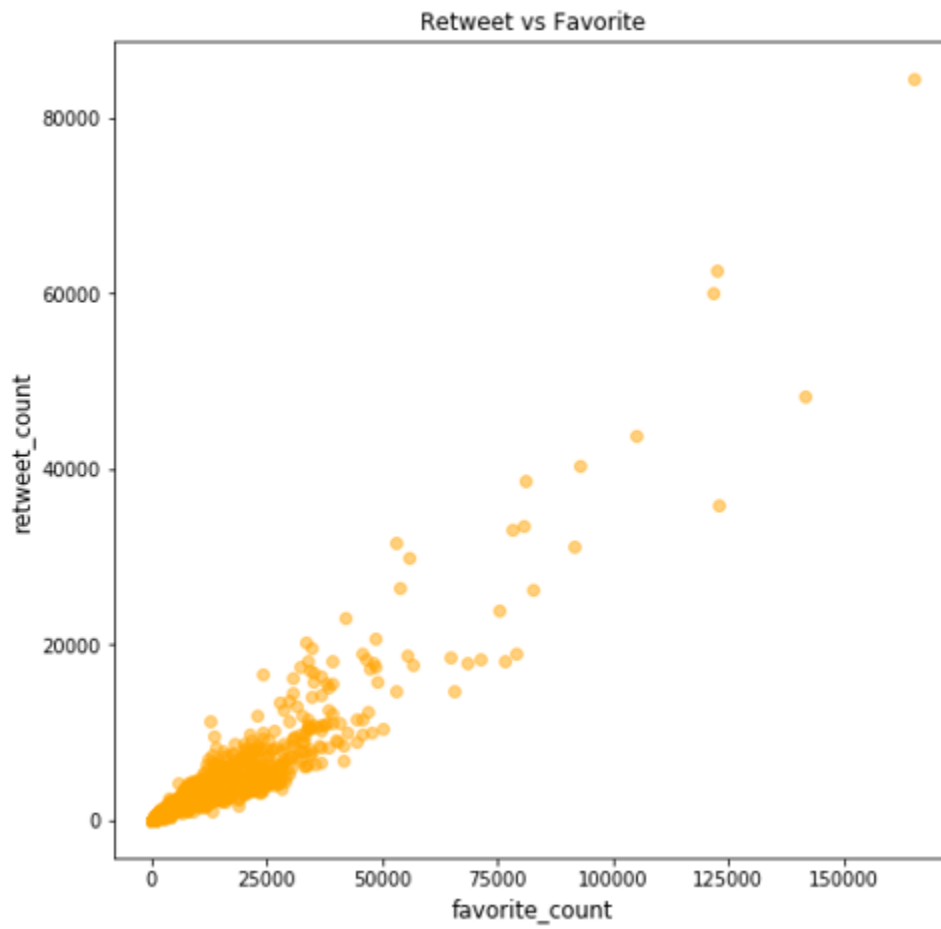- In addition, there are probably other Charlies too.

Relationship between variables:

Correlation heatmap

Reviewing the correlation between quantitative variables in our datasets:
- Most variables have very weak positive and very weak negative correlations.
- Friend count and follower count show strong positive correlation of 1 but looking at it further the friend counts are all 10. This wouldn't be something to investigate further.
- Retweet count and favorite count shows a strong positive correlation of .93; seems realistic, more favorites more retweets.
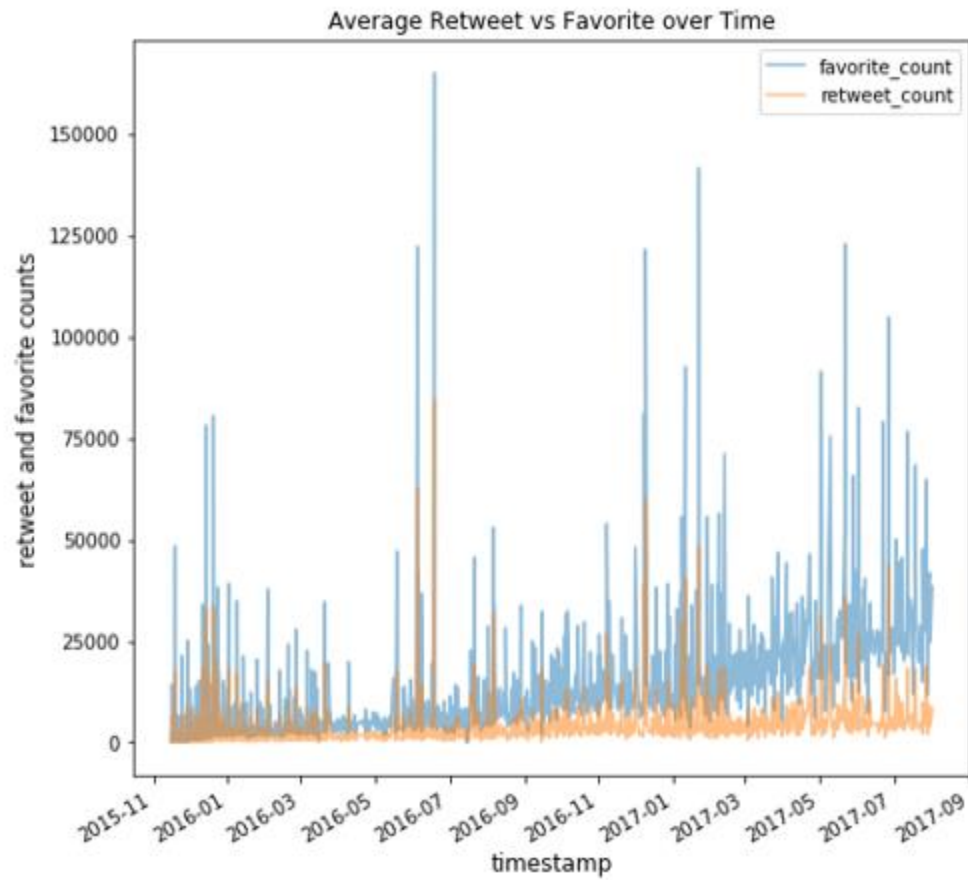
Let's investigate favorite count and retweet count some more!

- As the correlation map showed; the scatterplot visualizes, as retweets increase favorite count increase.
- Below is an image with the highest retweets and favorite counts.

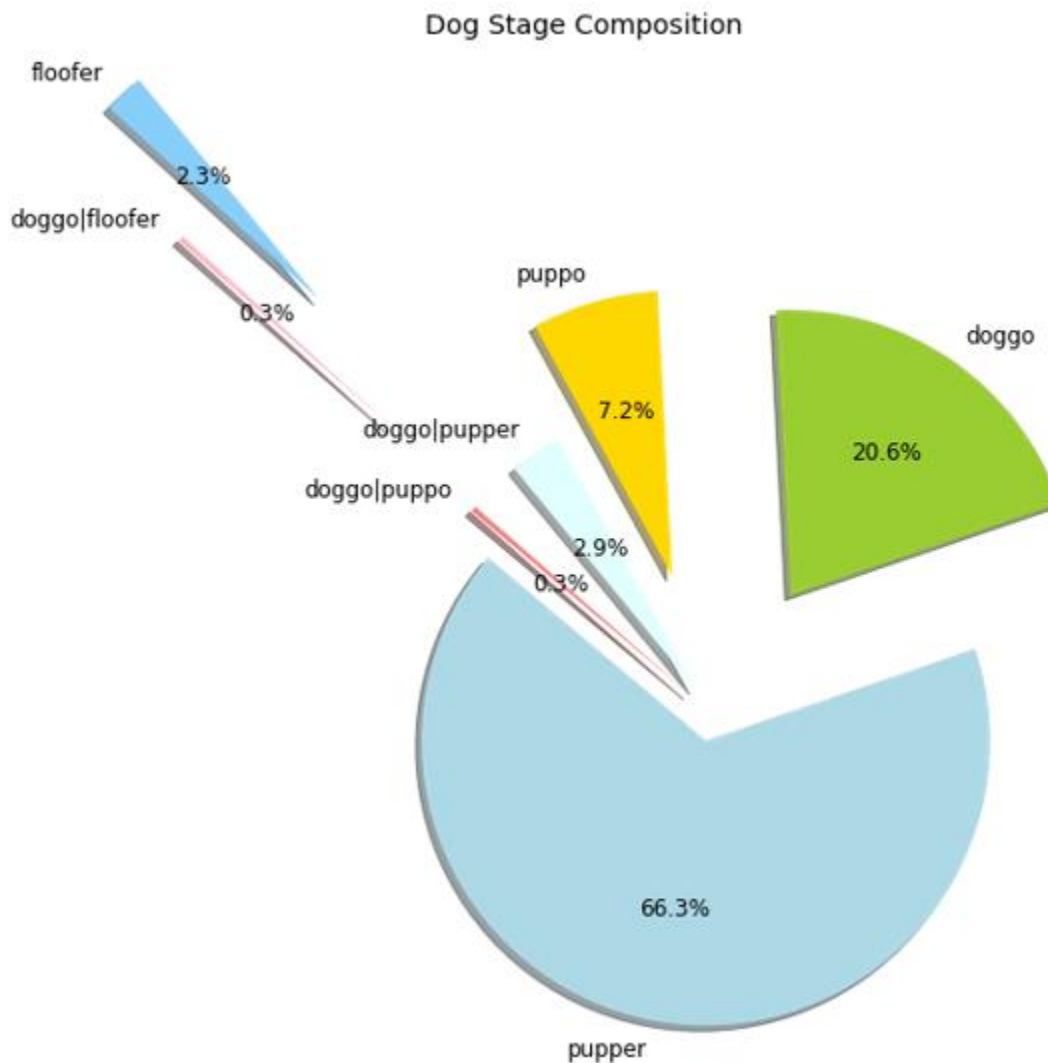Let's look at the average retweet and favorite over time!

Average Retweet vs Favorite over Time

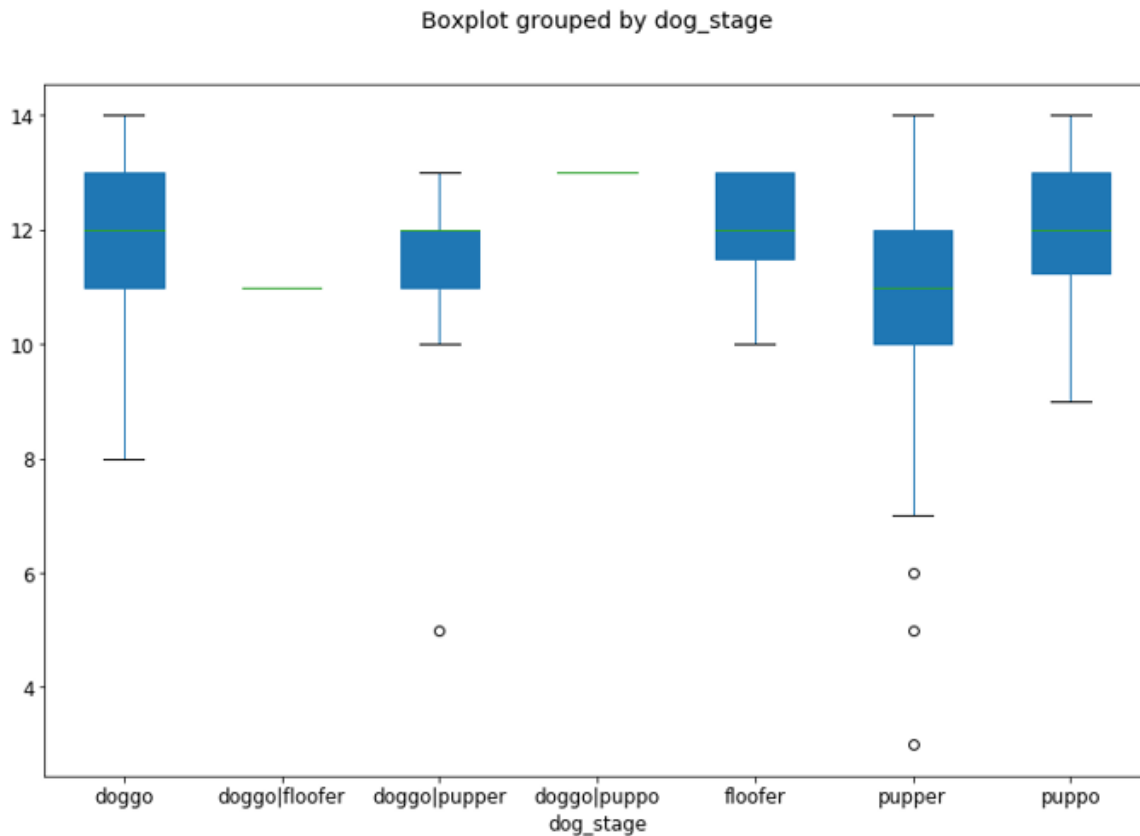| timestamp	favorite_count	retweet_count
2016-06-18 18:26:18	165056	84489
2017-01-21 18:26:02	141597	48245
2017-05-22 00:28:40	122911	35809
2016-06-04 23:31:25	122282	62651

- Retweet and favorite count increase around holidays and special occasions.
- WeRateDogs started 2015, so you can see the retweet and favorites increasing overtime, but the pattern is similar just more volume.

Dog Stage variable composition:

Let's look at the breakdown of the dog stage variable!

Dog Stage Composition

floofer 2.3%
doggo|floofer 0.3%
puppo 7.2%
doggo 20.6%
doggo|pupper
doggo|puppo 2.9%
0.3%
pupper 66.3%

- Excluding unidentified dog stages, pupper is 2/3 of the dog stage variable and the other 1/3 is made up of doggo, floofer, puppo, or a combination of both.
- Doggo: Dog, Pupper: puppies or puppy, Puppo: puppy, and floofer: fluffy
- Its's apparent users upload a lot of puppy photos.
- I'm sure there's a few users that consider their dog still a puppy regardless of age.

Boxplot grouped by dog_stage

	count	mean	std	min	25%	50%	75%	max
dog_stage								
doggo	63.0	11.888889	1.471351	8.0	11.00	12.0	13.0	14.0
doggo|floofer	1.0	11.000000	NaN	11.0	11.00	11.0	11.0	11.0
doggo|pupper	9.0	11.000000	2.397916	5.0	11.00	12.0	12.0	13.0
doggo|puppo	1.0	13.000000	NaN	13.0	13.00	13.0	13.0	13.0
floofer	7.0	12.000000	1.154701	10.0	11.50	12.0	13.0	13.0
pupper	203.0	10.636798	1.728179	3.0	10.00	11.0	12.0	14.0
puppo	22.0	12.000000	1.309307	9.0	11.25	12.0	13.0	14.0

dog_stage	
doggo	12.0
doggo|floofer	11.0
doggo|pupper	12.0
doggo|puppo	13.0
floofer	12.0
pupper	11.0
puppo	12.0

- Pupper makes up 2/3 of the dog stage variable; so you would except a lower mean with the higher count and few outliers, but mean is slightly lower than the median giving it a slight left skew.
- Standard deviations are low indicating distribution is clustered around the mean for all stages in the dog stage variable except the combination stages.

Conclusion:

WeRateDogs is a fun hilarious twitter account for dog lovers or any person that likes animals and looking to have fun.

From analysis the increased usage and popularity of the twitter account, the hilarious rating scale because every dog should be rated above 10, an accurate algorithm for recognition of breeds. The internet language has become main stream; Doggo: dogs, Pupper: puppies, Puppo: puppy, and Floofer: fluffy dog.

Here are a few of the photos tweeted that clearly were rated way above the scale and one isn't a dog; clearly a joke it's Snoop Dogg (rated 420) and Atticus got a bonus points for the celebrating the 4th (rated 1776), but you get the idea about the twitter account!