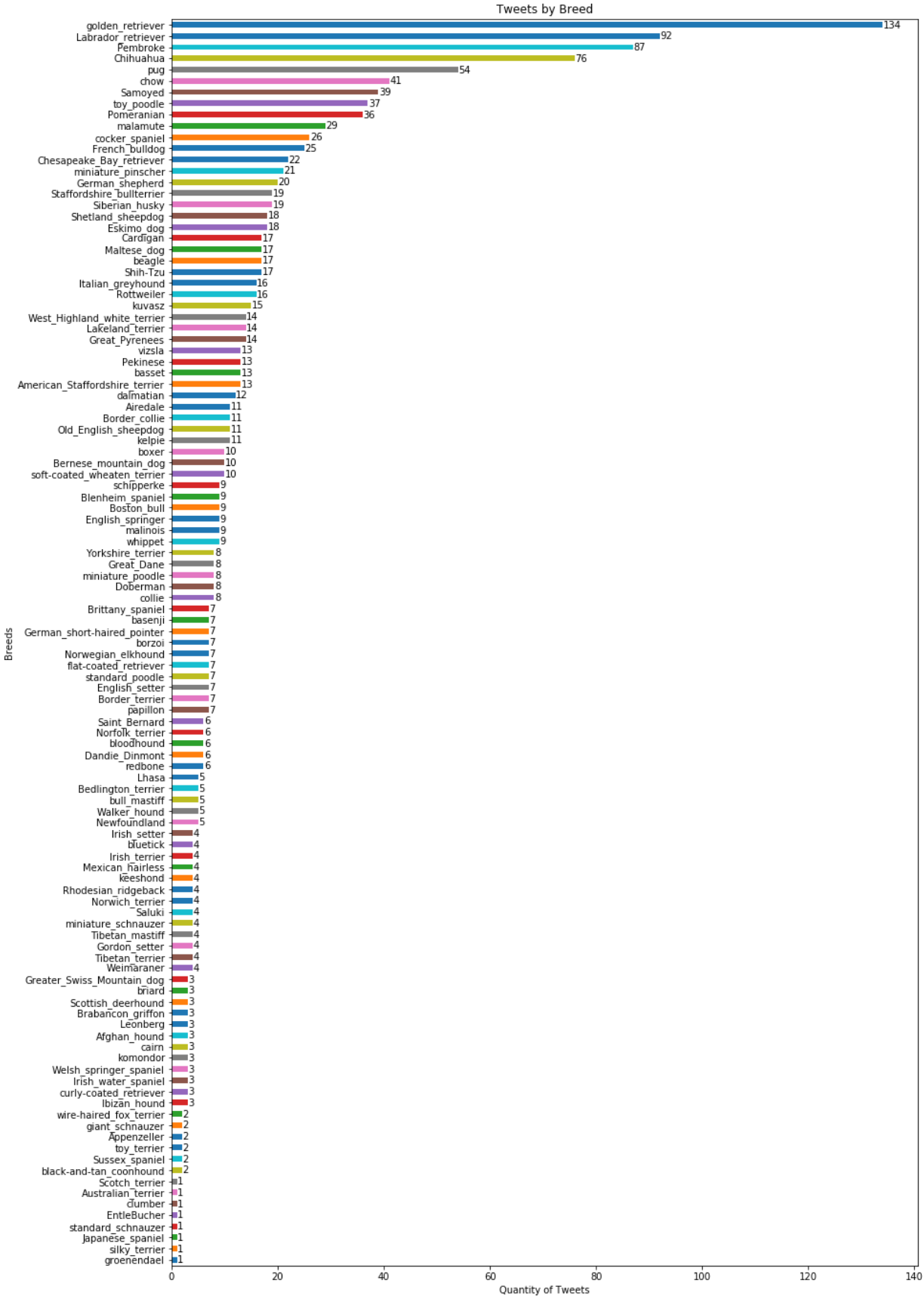
Exploring WeRateDogs

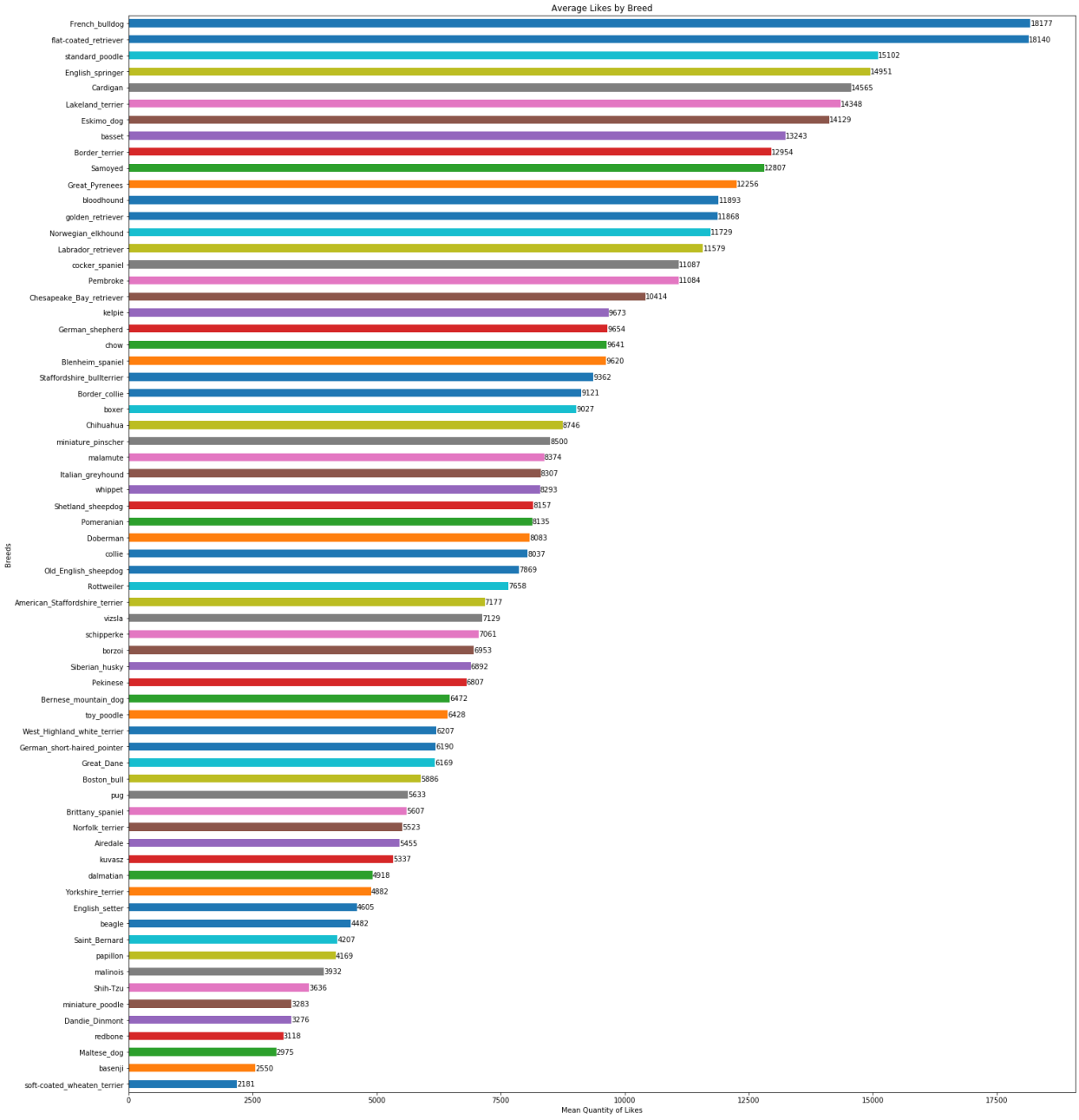
The exploration of the gathered WeRateDogs data mainly involved an analysis of the various breeds, including charting the number of tweets, likes, retweets, and ratings by breeds; as well as an analysis of correlations in dataset, including positive correlations with the ratings score.

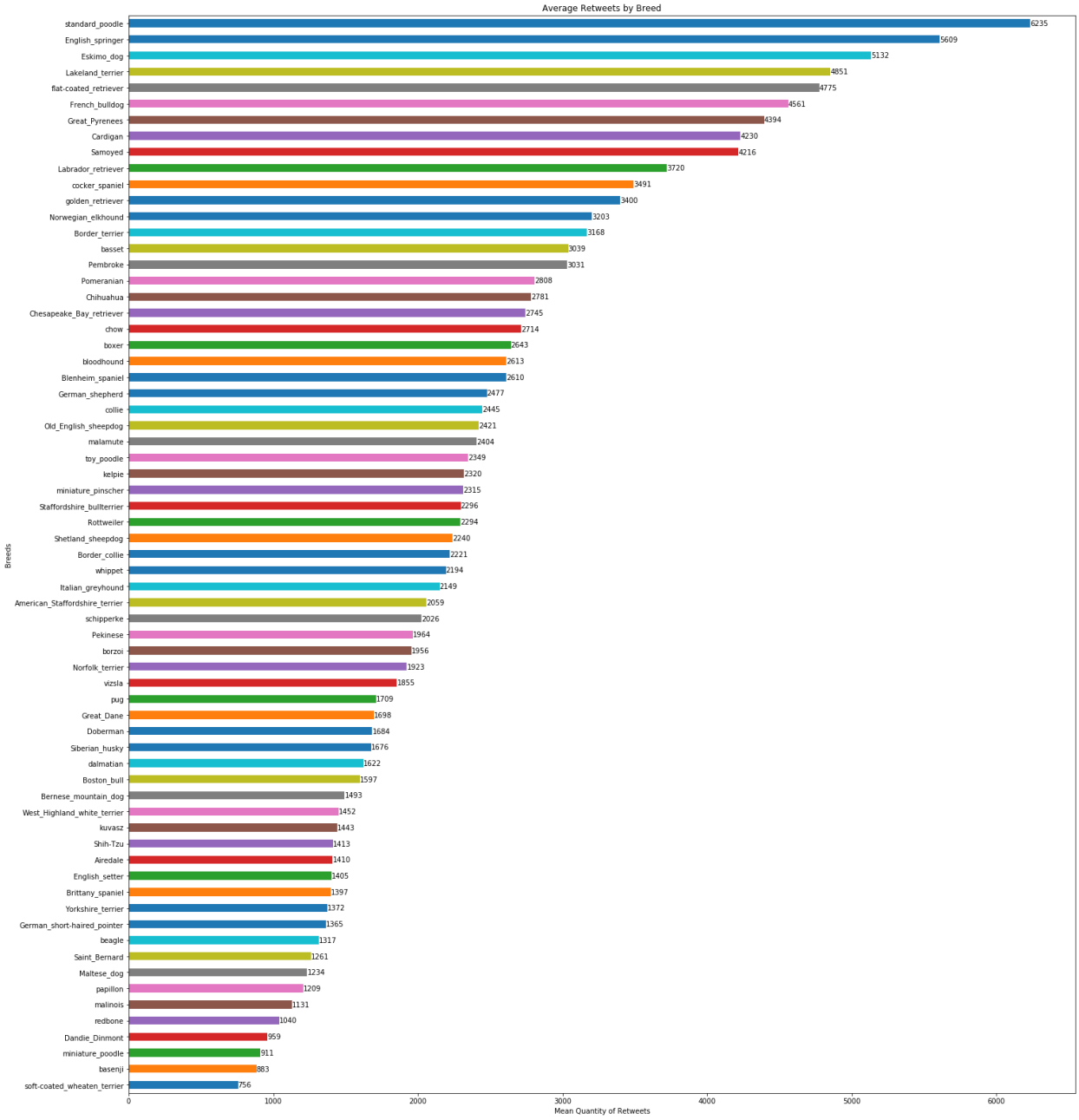
Breed Analysis

To begin, I plotted the quantity of tweets for each breed so I can which breeds were popular. As seen below, perhaps unsurprisingly, Golden Retrievers and Labrador Retrievers were the most popular. I'm guessing that the number of tweets just reflect how common the breeds are in general.

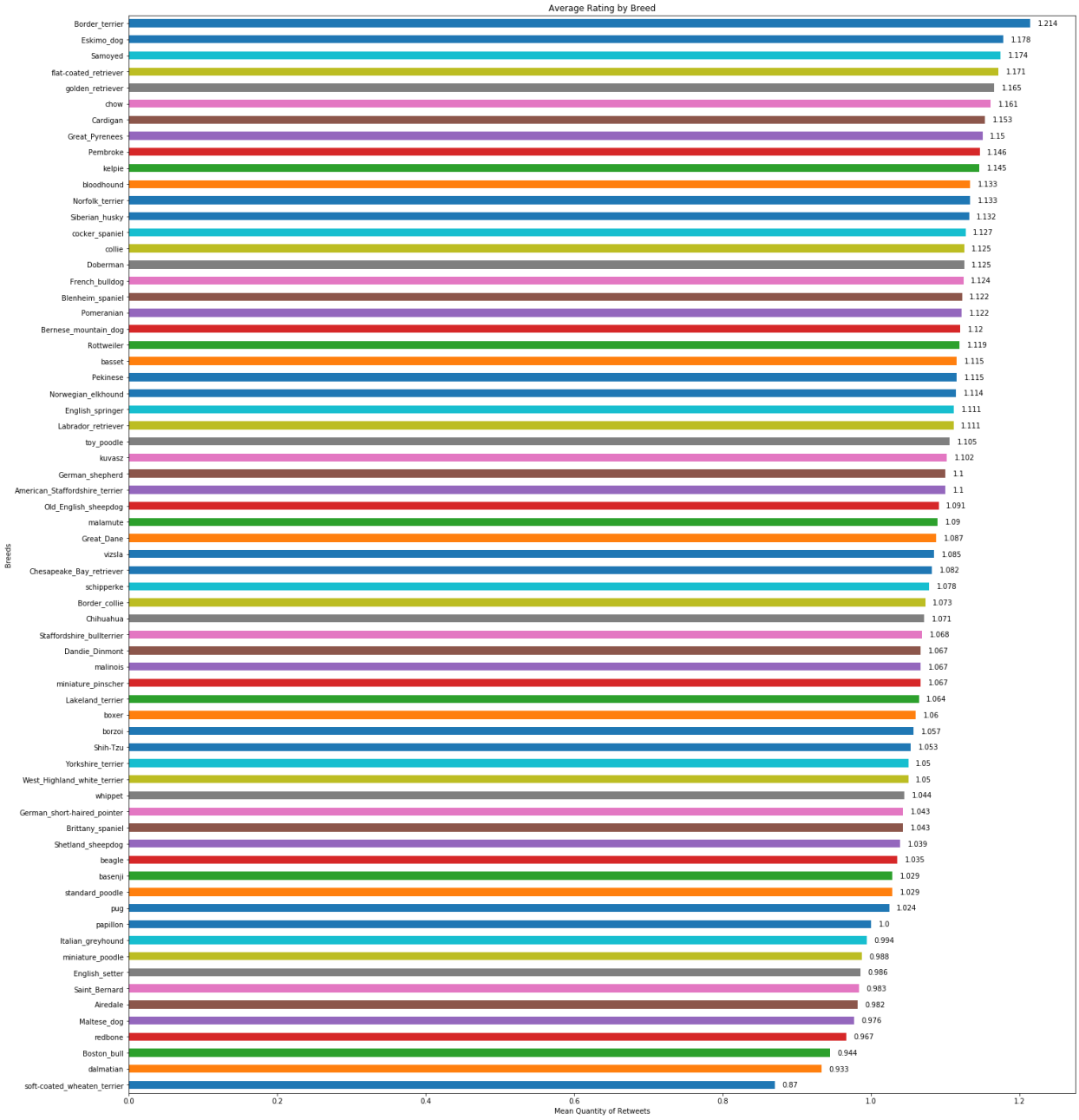

Tweets by Breed
golden_retriever 134
Labrador_retriever 92
Pembroke 87
Chihuahua 76
pug 54
chow 41
Samoyed 39
toy_poodle 37
Pomeranian 36
malamute 29
cocker_spaniel 26
French_bulldog 25
Chesapeake_Bay_retriever 22
miniature_pinscher 21
German_shepherd 20
Staffordshire_bullterrier 19
Siberian_husky 19
Shetland_sheepdog 18
Eskimo_dog 18
Cardigan 17
Maltese_dog 17
beagle 17
Shih-Tzu 17
Italian_greyhound 16
Rottweiler 16
kuvasz 15
West_Highland_white_terrier 14
Lakeland_terrier 14
Great_Pyrenees 14
vizsla 13
Pekinese 13
basset 13
American_Staffordshire_terrier 13
dalmatian 12
Airedale 11
Border_collie 11
Old_English_sheepdog 11
kelpie 11
boxer 10
Bernese_mountain_dog 10
soft-coated_wheaten_terrier 10
schipperke 9
Blenheim_spaniel 9
Boston_bull 9
English_springer 9
malinois 9
whippet 9
Yorkshire_terrier 8
Great_Dane 8
miniature_poodle 8
Doberman 8
collie 8
Brittany_spaniel 7
basenji 7
German_short-haired_pointer 7
borzoi 7
Norwegian_elkhound 7
flat-coated_retriever 7
standard_poodle 7
English_setter 7
Border_terrier 7
papillon 7
Saint_Bernard 6
Norfolk_terrier 6
bloodhound 6
Dandie_Dinmont 6
redbone 6
Lhasa 5
Bedlington_terrier 5
bull_mastiff 5
Walker_hound 5
Newfoundland 5
Irish_setter 4
bluetick 4
Irish_terrier 4
Mexican_hairless 4
keeshond 4
Rhodesian_ridgeback 4
Norwich_terrier 4
Saluki 4
miniature_schnauzer 4
Tibetan_mastiff 4
Gordon_setter 4
Tibetan_terrier 4
Weimaraner 4
Greater_Swiss_Mountain_dog 3
briard 3
Scottish_deerhound 3
Brabancon_griffon 3
Leonberg 3
Afghan_hound 3
cairn 3
komondor 3
Welsh_springer_spaniel 3
Irish_water_spaniel 3
curly-coated_retriever 3
Ibizan_hound 3
wire-haired_fox_terrier 2
giant_schnauzer 2
Appenzeller 2
toy_terrier 2
Sussex_spaniel 2
black-and-tan_coonhound 2
Scotch_terrier 1
Australian_terrier 1
clumber 1
EntleBucher 1
standard_schnauzer 1
Japanese_spaniel 1
silky_terrier 1
groenendael 1
Breeds
0
20
40
60
80
100
120
140
Quantity of Tweets

While the last result was unsurprising, perhaps there are other breeds that are more popular likes and retweets-wise. So, for breeds with more than 5 tweets, I plotted the average likes and retweets for each breed. Here, we see that Golden Retrievers are not at the top. French Bulldogs generate the most likes, and Standard Poodles generate the most retweets.





Just because the data was available, I also plotted the average ratings. Apparently Border Terriers are the best dogs among very good dogs.

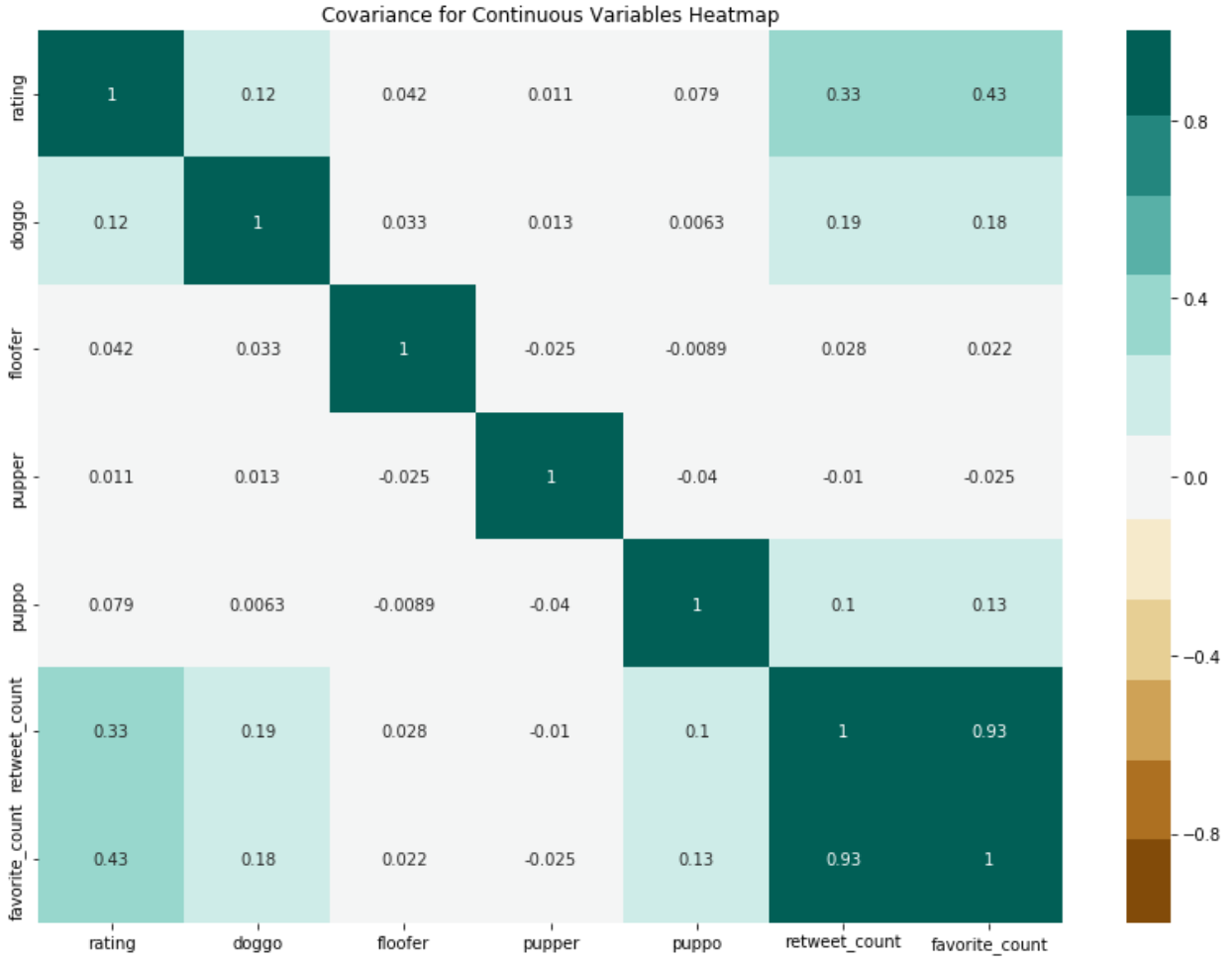


Exploring Correlations

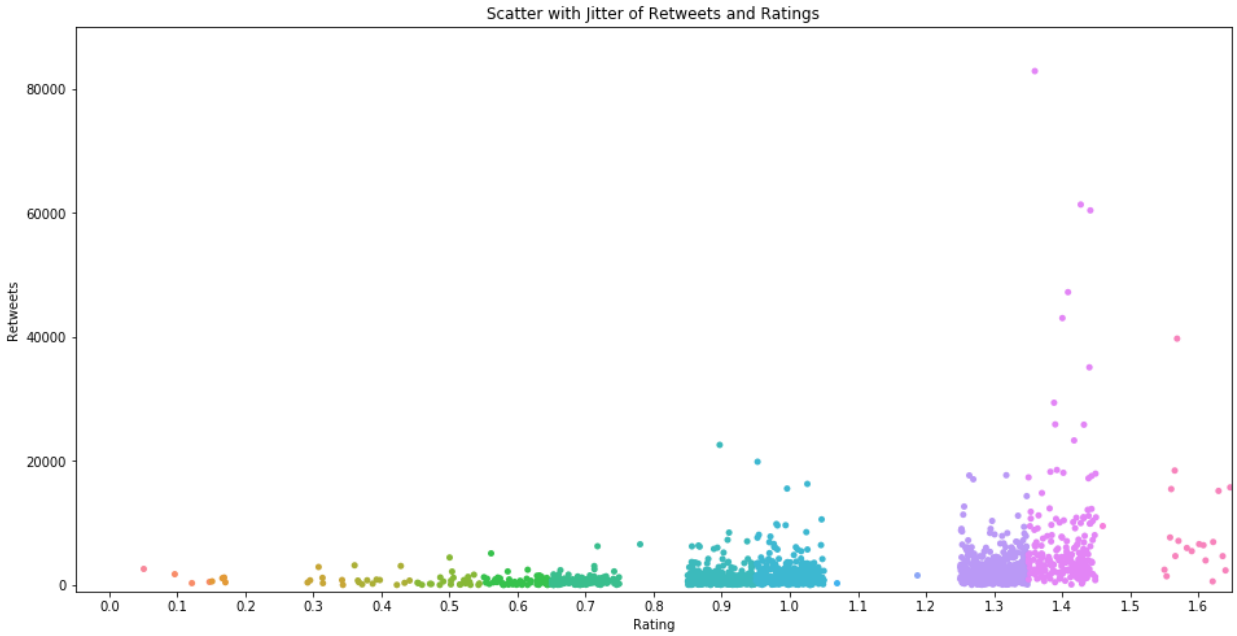
Given there were a number of continuous variables in the dataset, I generated a correlation-table heatmap to calculate and visualize significant correlations.

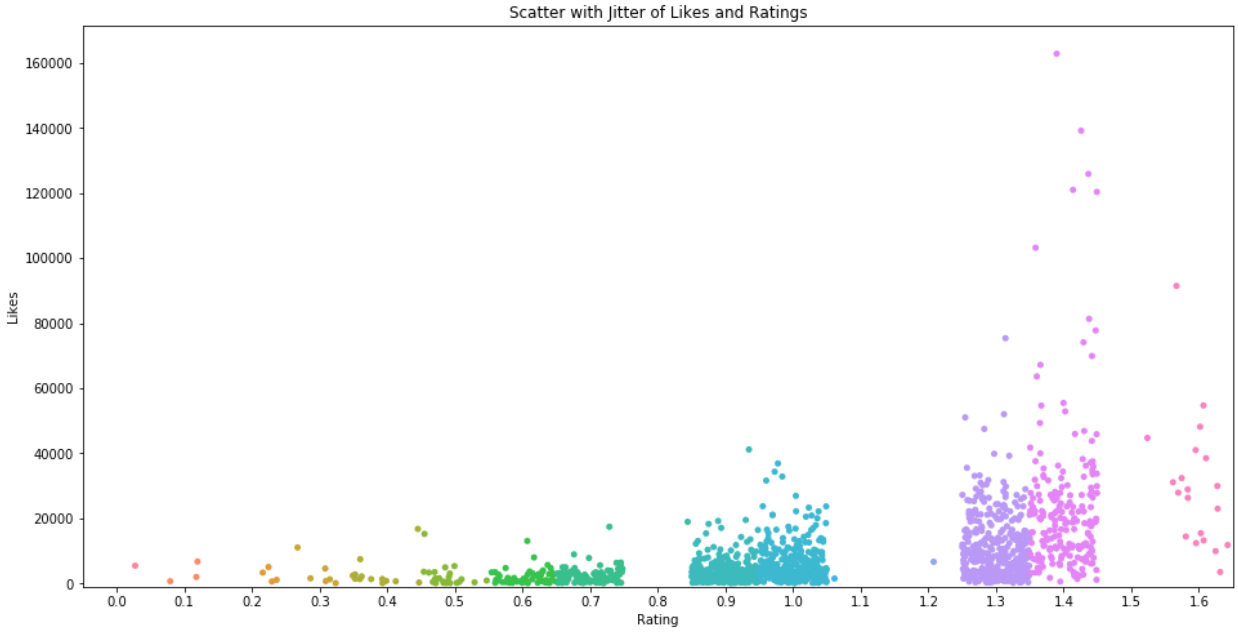
From reviewing this table, I found that (1) of the dog stages (“puppo”, “pupper”, “floofer”, and “doggo”), only “doggo” and “puppo” generated notable correlations, and that (2) ratings appear to have a significant correlation with likes and retweets.

For (1), “doggo” and “puppo” appear to slight positive correlations with likes and retweets. I guess the insight here would be that it wouldn’t hurt to tweet “doggo” and “puppo” more often (or post more pictures of “doggos” and “puppos”? I’m not sure what these words mean).



For (2), given that significant correlations were found for the ratings, I wanted to visualize the relationship to see if the correlations were visible. Even though the ratings were technically continuous values, they were mostly discrete and falling mostly about 0.1 increments. Therefore, I scattered the variables with jitter so the points didn't overlap and misrepresent the quantity of points at or near a particular coordinate. Funny enough, it does appear that the higher ratings are associated with more retweets and likes.





Findings

To recap, the exploration produced the following insights:

1. Although Golden Retrievers are by far the most common breed on WeRateDogs, there are other breeds that garner more likes and retweets.

2. Whether a tweet has a "floofer" and "pupper" doesn't seem impact likes and retweets, but there is a slight positive correlation for "doggos" and "puppos".

3. Higher ratings do appear to garner more likes and retweets.

So maybe for the next rating, WeRateDogs should seek a very good French Bulldog and award the "doggo" a 14/10.