

# Insights and visualizations produced from WeRateDogs™ wrangled data

## About the data...

WeRateDogs is a Twitter account that rates people's dogs with a humorous comment about the dog. It was started in 2015 by college student Matt Nelson, and has received international media coverage both for its popularity and for the attention drawn to social media copyright law when it was suspended by Twitter. Dogs are rated on a scale of one to ten, but are invariably given ratings in excess of the maximum, such as "13/10". Popular posts are re-posted on Instagram and Facebook.

## About the data wrangling...

In order to analyse the data, I collected the information from 3 sources, merged the sources cleaned and tidied up the information collected.

## Questions asked & Findings on the data...

Q1:What is the most popular dog breed?

Looks like the most popular dog breed in the data selection provided is the Golden retriever, followed by the Labrador retriever and the Pembroke. This appears to be in accord with the world pet population of dogs as well according to the 2017 study of American Kennel Club cited by CBS.

Q2:What is the most popular dog name?

Looks like there are many missing names, but other than that Cooper is the most popular name, followed by Tucker. This is also in accord with a 2017 study made by <https://www.rover.com/blog/2017-top-100-dog-names/>, where Cooper is among the top 3 male dog names.

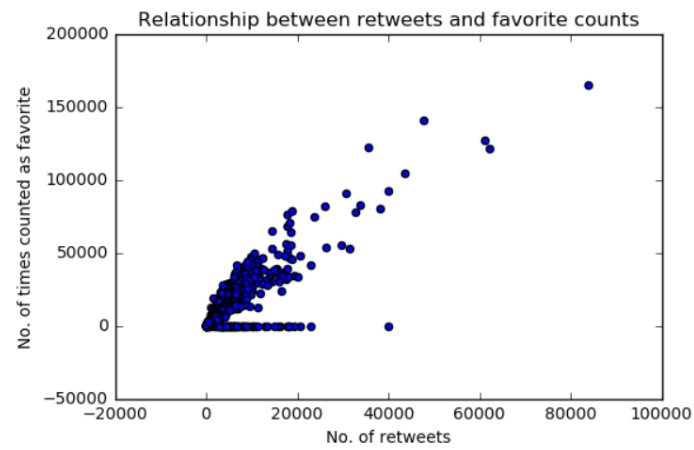
Q3:What are some descriptive statistics of the rating?

From the descriptive analytics table below we notice that the average numerator is 12.2 while the average denominator is 10.5 with standard deviations of 40.8 and 6.7 for the two measures. The maximum favorite count seems to have reached 164,525, while the maximum retweet count reached only 83,799. Given that there are ratings in excess of the maximum of "10/10" it shouldn't be a surprise that there are such large variations within the rating numerator and denominator information provided.

	rating_numerator	rating_denominator	img_num	p1_conf	p2_conf	p3_conf	retweet_count	favorite_count
count	2044.000000	2044.000000	2044.000000	2044.000000	2.044000e+03	2.044000e+03	2044.000000	2044.000000
mean	12.212329	10.459393	1.202544	0.595354	1.347722e-01	6.023668e-02	2825.574853	8474.721135
std	40.884170	6.730299	0.560639	0.271279	1.010962e-01	5.099345e-02	4907.587944	12749.427511
min	0.000000	2.000000	1.000000	0.044333	1.011300e-08	1.740170e-10	12.000000	0.000000
25%	10.000000	10.000000	1.000000	0.365868	5.300115e-02	1.609288e-02	600.000000	1598.750000
50%	11.000000	10.000000	1.000000	0.588620	1.178485e-01	4.922740e-02	1329.500000	3720.000000
75%	12.000000	10.000000	1.000000	0.847062	1.963338e-01	9.174633e-02	3259.750000	10618.500000
max	1776.000000	170.000000	4.000000	1.000000	4.880140e-01	2.734190e-01	83799.000000	164525.000000

Q4:Is there any apparent relationship between no. of times a tweet was counted as favorite and its no. of retweets?

Judging by the scatter plot below the number of of times a tweet was counted as favorite does seem to increase with the number of times it was retweeted.



## Resources:

<https://en.wikipedia.org/wiki/WeRateDogs>

<https://www.esquire.com/lifestyle/a54940/we-rate-dogs-matt-nelson-interview/>

<https://www.cbsnews.com/pictures/most-popular-dog-breeds-in-the-u-s/>

<https://www.rover.com/blog/2017-top-100-dog-names/>