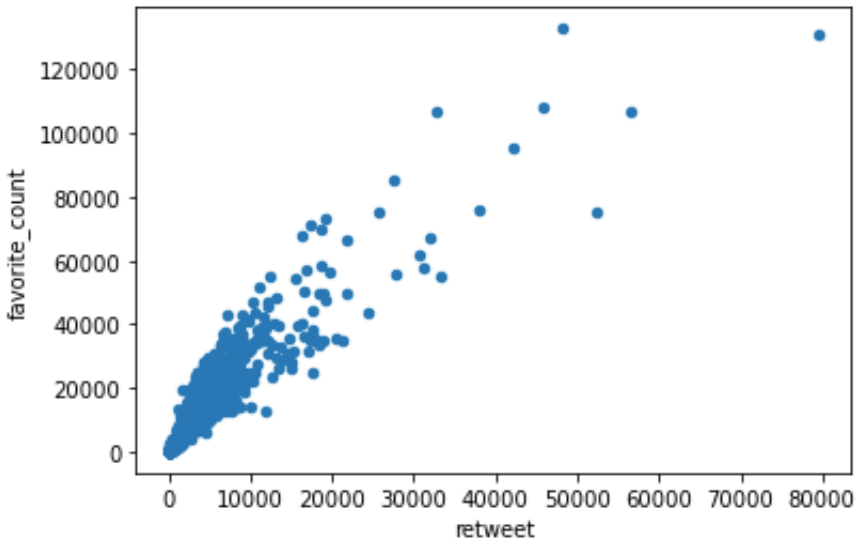
Whose a good dog?

All of them actually :) But is there a relationship between retweets and favorites? What category of dog is most favorited? And what breed image is more correctly predicted from tweets analyzed by a neural network? These are a few of the analyses possible based on the meld of three datasets related to WeRateDogs: a twitter account that allows readers to do exactly that – rate funny or eye-catching pictures of dogs. The meld integrated data from three sources into a single repository – a twitter archive of the tweets and accompanying images, additional data (retweets and favorites) distilled from the Twitter API, and the results of breed image predictions distilled from the passing of the Twitter archive through a neural network.

Is there a relationship between the retweets and favorites of posts?

Yes there is. This scatterplot illustrates a broadly positive relationship between the number of retweets a tweet gets, and its favorite count, as the upwards slant to the right of this plot shows: higher retweets seem to be associated with more favorites.



A basic linear regression model to check the strength of the relationship between retweets and favorites confirms this. We can see from the regression output below that there is a good fit of this model: R-squared is positive and high close to 1.0 - variability in favorite_count can be well-explained by retweets. And the coefficient for retweet shows us that for every retweet, favorite_count increases by just over 2. So retweet those floofers sheeple!

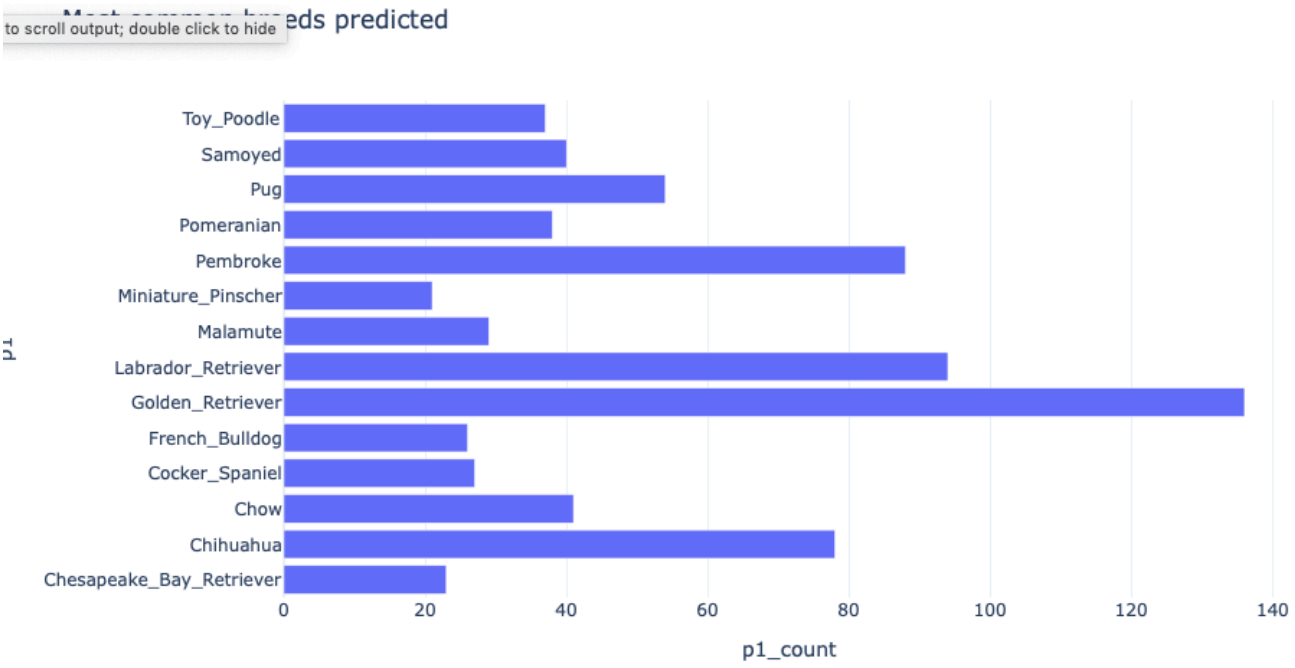
OLS Regression Results

Dep. Variable:	favorite_count	**R-squared:**	0.834
utput; double click to hide	OLS	**Adj. R-squared:**	0.834
Method:	Least Squares	**F-statistic:**	9858.
Date:	Sat, 22 Aug 2020	**Prob (F-statistic):**	0.00
Time:	08:34:28	**Log-Likelihood:**	-19576.
No. Observations:	1970	**AIC:**	3.916e+04
Df Residuals:	1968	**BIC:**	3.917e+04
Df Model:	1		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
intercept	2310.3110	131.142	17.617	0.000	2053.119	2567.503
retweet	2.3843	0.024	99.289	0.000	2.337	2.431

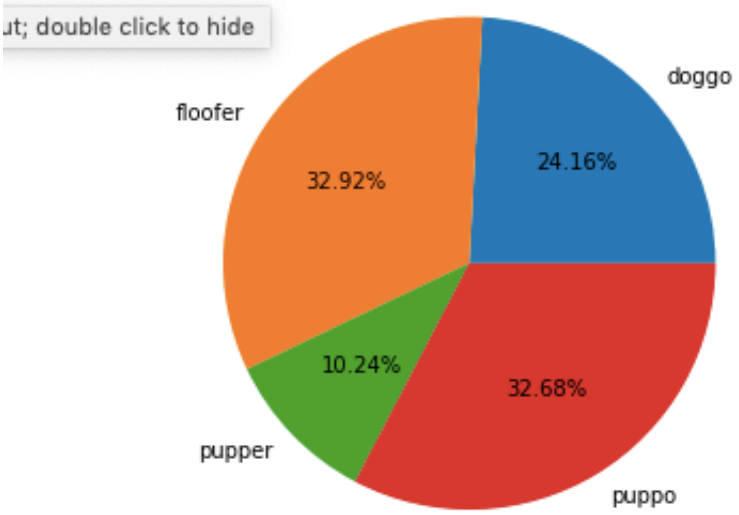
Omnibus:	533.171	**Durbin-Watson:**	0.772
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	38146.051
Skew:	-0.261	**Prob(JB):**	0.00
Kurtosis:	24.551	**Cond. No.**	6.35e+03

What breeds is most frequently predict based on the twitter images?



Retrievers for the win! Golden Retrievers and Labrador Retrievers are most frequently predicted by the neural network based on tweet data, followed by Corgis (Pembroke) and Chihuahuas. It could be that Golden Retrievers and Labs are the most common breeds to be tweeted about on WeRateDogs but more analysis here would be required. (It also doesn't mean that they are necessarily the most popular – looking at the retweets and favorites would give a sense of this.)

Rock, paper, scissors – what category of dog is highly favorited – Doggos, floofers, puppers, or puppos?



In very unscientific dog-lover speak these terms are used to assign a categorisation to dogs – floofers can be big and hairy, doggos are dogs in general, and puppers and puppos are puppies. (There's a world dogtionary out there – here is one source: https://www.rover.com/blog/uk/do-you-speak-doggolingo-dog-internet-jargon-explained/ (https://www.rover.com/blog/uk/do-you-speak-doggolingo-dog-internet-jargon-explained/)

Based on the categorisations in the data, floofers are. But if one collasped puppers and puppos into the same category, then puppers do. Either way you look at it, size seems to matter This is not conclusive however, as less than a third of the database was assigned to a categorisation.