Report on analyzing the WeRateDogs data

The twitter account *WeRateDogs* comments on dogs' images and gives rating scores. We briefly review our insights into its preprocessed dataset that includes about 1600 tweets combined as well as the dog breed predictions. Most tweets (>98%) were submited via iPhone users whereas 1.3% were sent via online version and only 0.5% from Twitter Deck.

1. Statistics of Dog Stages

The pie chart on the right shows that over 84% of rows had no information on the dog stage.
The most dominant dog stage is the pupper (10.4%) whereas other dog stages, such as doggo, puppo, and floofer, occupied 3.2%, 1.3%, and 0.5%, respectively.

2. Statistics of Dog Breeds

The relations of dog breeds with tweet counts, retweets, and favorites are shown on two bar charts on the right. Both **Golden and Labrador retriever**s won the 1st and 2nd ranks in terms of tweet counts. They took the same ranks in other rankings of retweets and favorites in the second bar chart.

In the rankings of retweets and favorites, we found almost the same kinds of breeds as in the previous case. A slight difference is French bull dog ranked in the 5th positition, for example. The overall tendency of the number of (re)tweets and favorites is that the number of tweets seems positively correlated with the number of retweets and favorites.

3. Statistics of Rating Scores

According to the bar plot below, the mean value of the rating scores is about **1.1** and the **25%** percentile gets already 1.00. In other words, most dogs are given high scores and generously evaluated.
The histogram, with the bin width equal to 0.1, shows that about **68%** of dogs were evaluated in the range between **0.9** and **1.2**.

4. Remarks

In the wrangling process, we restricted the results of dog predictions only to the most likely ones. It might be also interesting to observe how the statistics of predicted dog breeds will behave for the second and third likely predictions. It might be interesting if we could fit each bar plot onto a certain continuous function in mathematics.

Dog breeds ranked in top 10 of tweet counts

Dog breeds ranked in top 10 of retweets and favorites

Distribution of rating scores (bin width=0.1)