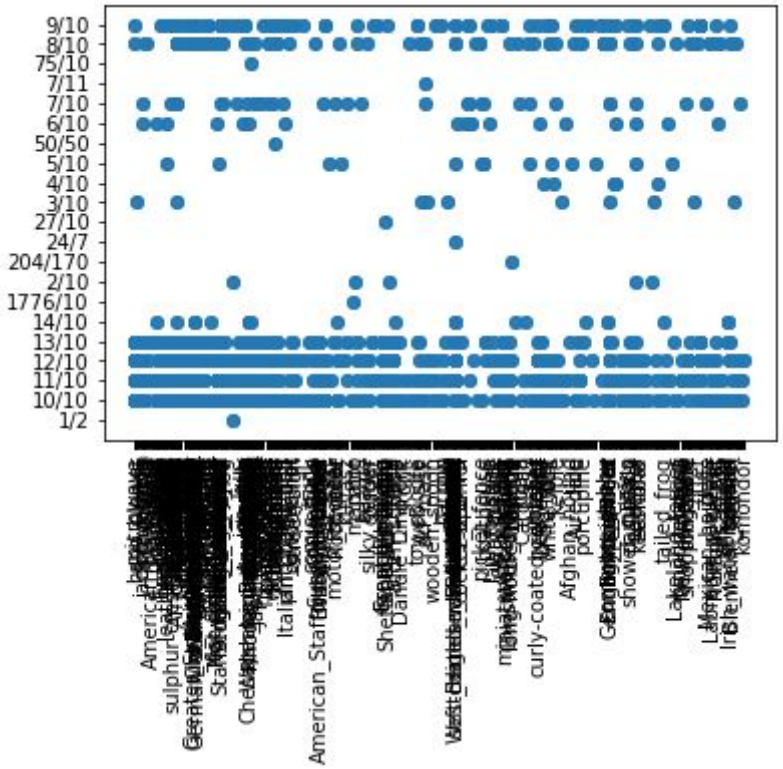
The dataset initially seemed difficult to plot because the ratings were so weird, however, there are certainly aspects that are trackable- for instance, the category of dog type. In addition, despite roughly 200 ‘weird’ ratings, the ratings are generally a number over 10. In further examining this data, it would be possible to remove the ratings that are ‘outliers’, with denominators not equal to 10, and try to judge how ratings is correlated with the other variables.

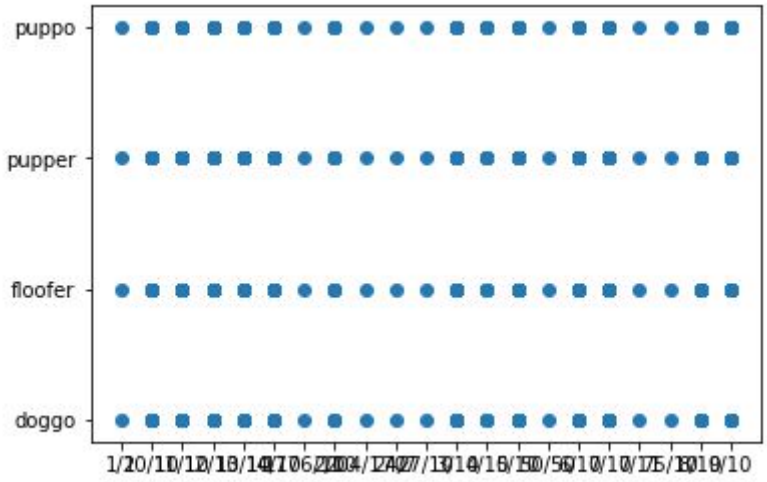
Variables that require further investigation are dog_type relating to many other categories- eg, the predictions category (are golden retrievers mostly classified as ‘puppos’, for instance?), the ‘likes’ categories, the ‘retweets’ category, and the rating.

Finally, to continue on with analysis of this dataset, I think it would be interesting to look into the text of the tweets to see different correlations between recurring words and ratings. Are there certain words that relate to different dog types? Do higher numerators correspond to certain words?

The first plot I made was an attempt to see the relationship between the prediction of dog breed, and the rating given. The ratings are now strings, so the order here doesn’t make sense, and there were too many breeds of dogs, so the x-axis is impossible to read.



The second plot I made was a comparison between dog type and rating.



Visualizations Citations
- Used for help: https://matplotlib.org/gallery/shapes_and_collections/scatter.html
- Used for help: https://matplotlib.org/gallery/ticks_and_spines/ticklabels_rotation.html#sphx-glr-gallery-ticks-and-spines-ticklabels-rotation-py