In order to effectively wrangle the data, I first started by looking for missing data values. Since there didn't appear to be any key missing values aside from animal names, I proceeded to look into the data's tidiness. I felt that the data could be presented in a cleaner manner by combining each of the data archive, image, and tweet like tables into a single data table. In order to merge these tables, I updated the "id," column from the tweet_likes datafield to match the other two dataframes. Once this was done, I was able to perform an inner join to merge each table togetherOnce the separate tables were merged, I moved on to condensing the doggo, pupper, puppo and floofer columns into a single column. This would allow for comparisons and visualizations to be performed off of this combined information. Lastly for tidiness, all unnecessary columns were dropped from the data field to allow for a cleaner appearance.

Once tidiness was addressed, I moved on to the data's quality. Although the numerators were generally above 10, I removed any values that were greater than 400 as I feared these values would skew my results too greatly. To ensure consistency across my dataset, I removed any rows that had denominator values different than 10. I converted the timestamp field to a datetime function in order to remove the unnecessary "+0000," endings from each row's data. To ensure consistency among results, I decided to convert all figures to lowercase. I wanted to ensure retweets didn't skew my insights, so I removed all retweets from my data set. Lastly, I wanted to ensure that my insights were based on one photo only, and believed that multiple photos under the same tweet id could skew my results. I decided to remove any rows that contained multiple photos to ensure consistent results.