The wrangle report analysis WeRateDogs data from twitter downloaded file, twitter API and an image file. The twitter API was extracted from Tweepy and then converted to a data frame. The data from each of these files were merged on the common column of twitter_ID. The combined data was assessed and several quality and tidiness issues were identified.

Cleaning the data consisted of several steps. First the retweet_account data column was drop as it was empty . Tweet_id was converted to a numeric data type. The columns in_reply_to_user_id and in_reply_to_status_id had all null values and so was dropped.Also 'retweeted_status_id', 'retweeted_status_user_id', 'retweeted_status_timestamp were dropped as it was retweet data. The rating numbers were extracted from the text column, to ensure the decimal values were also extracted. Timestamps were converted to date time data types. The dog stages of Doggo, Puppo, Pupper, Blep, Snoot, and Floof were extracted from the text column. They were all converted to lowercase for uniformity. The old columns of dog stages were dropped. The p1,p2 and p3 dog data was evaluated and if the the dog which had a true probability was selected and the False was ignored to find the possible type of dog for each twitter_id and then converted to name with proper case. Also, the source of the data was extracted from the source name.

Now the data was used for analysis and answering questions. First, we wanted to find which sources the twitter feeds were coming from and which was the source with the max tweets. Next we looked at the dogs with the 10 largest rating. We were also able to find the dogs and accounts with 5 largest favorite counts. They were then displayed in a bar chart. And lastly we grouped the data by type of dog to evaluate which dog had 10 largest ratings. They were then displayed in a bar chart.