Gathering the data:

The instructions gave how the data should be gathered: twitter_archive_enhanced.csv was as given image_predictions.tsv was to be downloaded programmatically tweet_json.txt was to be constructed via API

The initial data was given, downloaded, saved, then placed in the same folder as wrangle_act.ipynb so that it could be easily accessed. Named: dftweet

The image recognition link was given. It was downloaded, saved as a .tsv, Named: dfimg

The Twitter API data needed tweet ids, made an array of ids based on appending the ones from dftweet and dfimg, and then grabbing the unique values. Then using the given code, which outputs as a .txt file, before being transformed into a list, then a dataframe. Named: dfapi

Assessing:

Just going through the DataFrames to get more familiar with the files using both programmatic and visual assessment.

Cleaning:

Tidiness: Remove retweets from dftweet
Tidiness: Remove the unnecessary columns from dftweet
Tidiness: Remove retweets from dfapi
Tidiness: Remove retweets from dfapi (those beginning with 'RT')
Tidiness: Remove the unnecessary columns from dfapi

Tidiness: Add dog_stage columns for dfapi
Tidiness: Add dog_stage column for dftweet
Quality: Drop other doggo-lingo columns for dftweet

Tidiness: Extract base of rating_numerator and rating_denominator from full_text
Quality: Remove punctuation (except periods/decimal points) and letters from before
Quality: Remove punctuation (except periods/decimal points) and letters from after

Tidiness: Get name of dog following format of, 'This is..." and getting the next word
Tidiness: Remove working columns ('pre' and 'post)
Tidiness: Rename columns in dfapi to match those of dftweet

Tidiness: Filter out dfimg with most common type of dog
Tidiness: Remove unnecessary columns from dfimg

Make copies before merging into each other

Tidiness: Find columns in dftweetimg that are not in dfapiimg
Tidiness: Get the list of rows in dftweetimg that are not in dfapiimg
Tidiness: Find out what columns to add to tweets_to_add so that it can be appended to dfapi
Tidiness: Append tweet_to_add to dfapiimg
Tidiness: Reorder the columns to make it easier to understand
Tidiness: Change the columns to the correct types