Section 3: Data Filtering

We have one extra line of code in this script - line 4. Run the script and look at the output. What does this do?

Change the filter term to something else and re-run the script (be careful not to remove the quotation marks). Be warned, this might produce a lot of output!

Line 6 has also been edited. Rather than printing 10 tweets, we print all of the tweets in twitter_data. The function len() returns the number of items in the list passed to it - in this case, the number of tweets in twitter_data.

Try filtering one of the accounts by a random collection of letters (e.g. 'aofhafgas'). What happens? Why?

Coding Challenge: Calculate the mean number of likes for a filtered set of tweets. You will need to get the number of tweets in a different way here – how is this being determined in the for statement?

Questions to Investigate:

- Find some topics that get large amounts of likes and retweets for each person. Be scientific about this – make predictions before you run the code for a term. Compare terms between the two accounts.
- What topics are more popular when Barack Obama tweets about them compared to Taylor Swift. Are these easy to find? Why?
- Is there a way that we can find out how popular a tweet is in comparison to other tweets from the same account? (HINT: Ideally, we want to rescale the number of likes for each tweet such that the most popular tweet from an account scores 1 and all other values are between 0 and 1). This is a maths question!