

# Homework 7 Part 1

Amit Arora

October 29, 2017

## Wordcloud from Twitter data

The tweets.py program was run to pull data from twitter for a user provider #hashtag and number of tweets and then generate a wordcloud after cleaning and tokenizing the tweet data. The program takes command line arguments, but if no arguments are provided it downloads 30 tweets for #ggplot2. A maximum of 30 tweets are allowed, if the user provides any number greater than 30 then the program automatically clamps it at 30.

Following is a sample invocation of the program.

```
python tweets.py --hashtag=datascience --count=30
```

Note that the leading `#` for the hashtag argument is optional, the program automatically adds it if the user does not provide it.

The following sections show wordcloud for tweets with #datascience. Note that a 100 tweets were downloaded by changing the code temporarily just to get a good wordcloud visualization. Also, retweets were excluded from the twitter query.

## Wordclouds made in Python

Two wordclouds were created in Python. The screenshots are provided below.



Figure 1: Data science wordcloud1 in Python



1 tweets.py: this is the python code for the entire part 1 homework

5. \*.png: wordcloud png files.