**Proliferation of Twitter Bots During 2016 U.S. Presidential Election Cycle**

**Team Members**

* Chris Cruz
* Rufel Estrada
* Juan Flores
* Arafath Khan
* Raul Martinez

**Dataset Summary**

Source: Twitter Tweets from November 2016 - focus on U.S. Presidential Election

Observations: ~800k

Features: 13

File Size: 2.3 GB

**Problem Definition**

The rampant use of Twitter bots is suspected to have been used during the 2016 U.S. presidential election cycle. Our analysis of this dataset will consist of two main components:

* Overt Bot Identification: Identifying those tweets that are likely to have been generated from a bot source.
* Bot Likelihood: Using text analysis to determine what tweets bear a semantic similarity to those previously classified as bot-generated.

**Projected Output**

The final output of this analysis will be the quantification of bot involvement among Twitter traffic related to the presidential cycle from September 2016 to November 2016 (e.g. was it as prevalent as suspected?).

We are targeting usage of R, Python and Jupyter Notebook for this project.