

DATA Information

You will find three .csv files: clinton_raw, trump_raw, and congress_raw on mycourse.

Basically, they contains all 2016 presidential and congress candidates' tweets and some other information.

The clinton_raw and trump_raw have similar structure.

Each line represents a tweet and is in the following format:

created_at, favorite_count, retweet_count, text, time, AffectCount, MoralCount, shared

You only need to care the first four columns:

created_at: the time of the tweet;

favorite_count: the number of favorites of that tweet;

retweet_count: the number of retweets of that tweet;

text: the text of the tweet.

The congress_raw is a little bit different.

For the congress_raw, each line represents a tweet and is in the following format:

text, elite, retweet_count, media, gender, dw_extr, dwextr_rs, dw_score, race, age, followers

text: the text of the tweet;

elite: the twitter account;

retweet_count: the number of retweets of that tweet;

media: if the tweet contains media information (e.g., pictures). -0.5: yes; 0.5: no;

gender: male (0.5) or female (-0.5);

dw_extr, dwextr_rs, dw_score: all refers political ideology, they are calculated based on each person's voting history. You should only use "dw_score", <0: libreal, >0 conservative;

Race: White (0.5) or Non-White (-0.5);

followers: Number of followers.