Topic: Election Day Tweets

Report (5-10 pages) https://docs.google.com/document/d/1mUyNOLmRr7wlAMfOn32TojwCkv2FikMKhkk4I_Z_-V8/edit?usp=sharing

- Summary
- Dataset (50 pts)
    - Description of data
    - Source of data
    - License restrictions, if any
    - Description of tables and significant attributes
    - https://www.kaggle.com/manchunhui/us-election-2020-tweets
- "Awesome cool thing" (100 pts)
    - Such as: charts and data visualizations, screen captures of software, interesting queries, and text descriptions
- Technical challenges with dataset
    - Less than 10 GB of data

Timeline

Date	Task	Notes
11/13 (Fri)	Make timeline, choose a dataset	
11/20 (Fri)	Choose what direction to take the project -- choose scope (everyone) Create paper outline (Karah) Start cleaning data, based on ideas	Everyone should: Come up with a scope prepared to discuss
11/28 (Sat)	Write intro Start creating visualizations Start sql queries	-- think about sql queries -- research how to put data into csci 403 database
12/1 (tue)	Samson & Karah go to office hours and upload cleaned datasets to CSCI403	This week: Jared can do word maps. Margaret will work on the report, will do research on bot prediction parameters. Karah can do geography maps and misc. Visualizations. Samson will finish off the technical challenges section and add on to the reports.
12/4 (Fri)	Completed draft, do final revisions	Meet for 2 hours. Format/edit report draft during meeting time. -- do queries

12/9 (Wed)	**Project 9 Due**	

Scope: word trends in tweets

- Similarities/trends within most popular tweets by likes, retweets, replies
- Location trends: do trends vary by state?
- Android vs apple word trends

As a stretch:

-

Friday, 11/13/2020 Team Meeting

- Topic idea: Trump vs. Biden Tweets
- Scope Ideas:
  - Weight tweets by likes and user followers
  - Device: Android vs. iPhone
  - State trends:
    - Tweets vs. election results
    - Election called vs. state respond
- Cleaning data:
  - Only English tweets
  - Find most popular words
  - Maybe don't scrap all null columns, maybe make a limit
  - Make limit on what to clean based on number of nulls (?)
- Visualization ideas:
  - Word maps
  - Charts
- To do:
  - Choose scope
  - Make write-up outline
  - Clean data
  - Research data visualization ideas

Friday, 11/20/2020 Team Meeting

- Determining scope:
  - Trends based on words
  - Candidate popularity:
    - Time
    - Location
    - Reach
  - Main Question: Is it possible to detect a bot?
- Problems:
  - How to take out non-English tweets?
    - Answer: only include data from America
- Keep time/location correlation maps as a backup plan.
- For Next Week:
  - Research SQL queries
  - Figure out how to put in database (maybe reach out to CPW)

    - ~~SELECT COUNT(state_code), state_code FROM biden WHERE country = 'United States of America' AND state_code IS NOT NULL GROUP BY state_code LIMIT 50;~~
      ~~SELECT COUNT(state_code), state_code FROM trump WHERE country = 'United States of America' AND state_code IS NOT NULL GROUP BY state_code LIMIT 50;~~