Twitter Conversations Project
Descriptive Statistics and Preliminary Analyses

Spring 2022:

Sparse: *Are comments shorter than their parent tweets?*

- Average word count of a comment on a given tweet relative to the word count of the tweet (i.e. Say we have Tweet_i, which has n comments. What is the average word count of the n comments on Tweet_i, relative to the word count of Tweet_i?).
    - *Reason:* This would be to get at whether each comment is (on average) shorter than the tweet and by how much. If we repeat this measure for each Tweet, we could say something like: On average, comments on tweets are X words shorter than the tweet on which they were based.
    - *Deliver:*
        - Variable added to the dataset
        - Report of the average word count of a comment across entire dataset
        - Report of the average word count of comments in just the Jan 6 case study
        - Report of the average word count of *parent tweets* across entire dataset
        - Report of average word count of parent tweets in just the Jan 6 case study

- Difference in word count between the parent tweet and the comments. Say Tweet 1 has 20 words

Comment Number	N Words in Comment	Difference Between Comment and Parent Tweet
1	10	20-10=10
2	15	20-15=5
3	30	20-30=-10
4	5	20-5=15
5	20	20-20=0

Then, we'd calculate the average change in word count for Tweet 1 as: [10+5+(-10)+15+0]/5 = 4, so comments on Tweet 1 have 4 fewer words than the parent tweet.

- *Reason:* Tells us how much shorter comments are than their parent tweets on average.

- *Deliver:*
    - Variable added to the dataset
    - Report of the average difference in word count across the entire dataset
    - Report of the average difference in word count for just the Jan 6 case study

Biased: ***Do comments contain more partisan bias than their parent tweets?***

- Use Kaufman's measure of partisan bias applied to each parent tweet in the dataset and all subsequent comments. See here: https://arkaufman.shinyapps.io/textbias/ and email communication with him on resources to do this at scale.
- *Deliver:*
    - Variable added to dataset for partisan bias for each comment and partisan bias for each parent tweet.
    - Estimate the difference in bias estimate between parent tweet and each comment on that tweet (as with word count above). Add this variable to the dataset.
    - Report the average partisan bias for all comments in the dataset.
    - Report the average partisan bias for all parent tweets in the dataset.
    - Report the average difference in partisan bias between parent tweets and comments in the dataset.
    - Plot: Nice-looking plot showing the distribution of partisan bias in parent tweets (one line, solid), distribution of partisan bias in comments (one line, dashed), all in black and white and clearly labeled (e.g. score will range from -2 (extremely liberal) to +2 (extremely conservative) – label axes as such). Left panel should show distributions for the full dataset, right panel should show distributions for just January 6th data.

Mobilizing: ***Are comments more likely to contain calls to action than their parent tweets?***

- Create a dichotomous variable for whether parent tweets and comments contain a call to action.
    - Keyword Search Approach: search for the following terms/phrases: must, need* to, research, vot*, act*, try, you should, we should, they should, look at, go, think, call, stop
    - Machine Learning Approach: use labeled data from the experiments as the training set and apply to the Twitter data to estimate whether a parent tweet or comment is mobilizing or not.
- *Reason:* Tells us whether comments (social information) contain more mobilizing calls to action than tweets by news outlets.
- *Deliver:*
    - Variables added to dataset for each parent tweet and each comment for whether they contain mobilizing information (1) or not (0)
    - Report the average number of parent tweets that contained mobilizing calls to action in the entire dataset.

- Report the average number of comments that contained mobilizing calls to action in the entire dataset.
- Report the average number *and percentage* of comments that contained mobilizing calls to action, but their parent tweets *did not* (e.g. parent tweet == 0) for the entire dataset. Example: Across the entire dataset, on parent tweets by news outlets that did not contain any mobilizing calls to action, X% of the comments on that tweet contained a call to action.
- Report the average number of parent tweets that contained mobilizing calls to action in the Jan 6th case study.
- Report the average number of comments that contained mobilizing calls to action in the Jan 6th case study.
- Report the average number *and percentage* of comments that contained mobilizing calls to action, but their parent tweets *did not* (e.g. parent tweet == 0) for the Jan 6th case study. Example: On January 6th, on parent tweets by news outlets that did not contain any mobilizing calls to action, X% of the comments on that tweet contained a call to action.

Fall 2021:

Descriptive Statistics:

- Average word count of a comment on a given tweet relative to the word count of the tweet (i.e. Say we have Tweet_i, which has n comments. What is the average word count of the n comments on Tweet_i, relative to the word count of Tweet_i?). This would be to get at whether each comment is (on average) shorter than the tweet and by how much. If we repeat this measure for each Tweet, we could say something like: On average, comments on tweets are X words shorter than the tweet on which they were based.
    - If this turns out to be interesting, maybe creating a bar graph showing the average word count for comments vs. average word count for tweets would be useful, but only if the gap ends up interesting. Otherwise I'll just describe in text.

-Another way to look at this would be to look at the difference in word count between the parent tweet and the comments. So something like, let's say Tweet 1 has 20 words

Comment Number	N Words in Comment	Difference Between Comment and Parent Tweet
1	10	20-10=10
2	15	20-15=5
3	30	20-30=-10
4	5	20-5=15
5	20	20-20=0

Then, we'd calculate the average change in word count for Tweet 1 as: [10+5+(-10)+15+0]/5 = 4, so comments on Tweet 1 have 4 fewer words than the parent tweet.

- Sentiment of comments and original tweets. There's a Young and Soroka dictionary, there's a liwcalike package that is a bootlegged version of LIWC, etc.
- Similarity. Calculate the similarity score for each comment, comparing it to its parent tweet (i.e. how similar is each comment to the parent tweet). Can use cosine similarity, jaccard similarity, and other measures.
- Time. Amount of time between each comment and its parent tweet.

Preliminary Analyses:

- Are comments shorter than their parent tweets?

  On average **comments** are **6 words shorter** than their parent tweets **among Jan 6** tweets, currently.

  On average **comments** are **3 words shorter** than their parent tweets **among all** tweets available for analysis, currently.

- Are comments more emotional/negative/positive than their parent tweets?

ALL CURRENTLY AVAILABLE (EXCEPT JAN 6)	Original Tweet	Comments
Positive	41%	32%
Neutral	40%	44%
Negative	18%	23%

JAN 6 ONLY	Original Tweet	Comments
Positive	41%	27%
Neutral	39%	49%
Negative	19%	24%

- Do partisan words show up more in comments than in parent tweets (e.g. keyword search: Democrat*, Republican*, Trump, Clinton, Biden
- Do political action words show up more in comments than in parent tweets (e.g. keyword search: vot*, research, protest,

ALL CURRENTLY	Original Tweet	Comment

AVAILABLE (EXCEPT JAN 6)		
Democrat	2.3%	1.6%
Republican	1.8%	1.3%
Trump	16.4%	8.6%
Clinton	1.1%	0.7%%
Biden	4.8%	0.1%
Vot*	3.7%	2.1%
Research	.3%	0.1%
Protest	1%	0.4%
Riot	0.2%	0.3%
Antifa	0.03%	0.07%
Terrorist	0.1%	0.3%

JAN 6 ONLY	Original Tweet	Comment
Democrat	5.5%	1.8%
Republican	6%	2%
Trump	25.3%	12.4%
Clinton	0.1%	0.08%
Biden	8.7%	1.8%
Vot*	10.4%	2.5%
Research	0.13%	0.13%
Protest	5.3%	3.3%
Riot	5.1%	2.9%
Antifa	0.38%	2.9%
Terrorist	0.46%	2.8%

- Are comments on negative parent tweets more negative than comments on positive parent tweets?

Jan 6 ONLY	Positive Parent Tweet	Negative Parent Tweet -
Positive Comments	24.74%	24.81
Neutral Comments	49.21%	48.85%
Negative Comments	25.99%	26.28%

- Do negative parent tweets generate more comments than neutral parent tweets?

	Number of Original Tweets	Number of Comments
Positive Parent Tweet		
Negative Parent Tweet		
Neutral Parent Tweet		

- Do negative parent tweets have comments with lower similarity scores than neutral parent tweets?

- Do negative parent tweets have comments with lower similarity scores than positive parent tweets?

- Do parent tweets about the insurrection generate more comments than parent tweets that were not about the insurrection?

- Are comments on insurrection parent tweets more negative than comments on non-insurrection parent tweets?