

Teaching / Research Guide for Archive #1

Background:

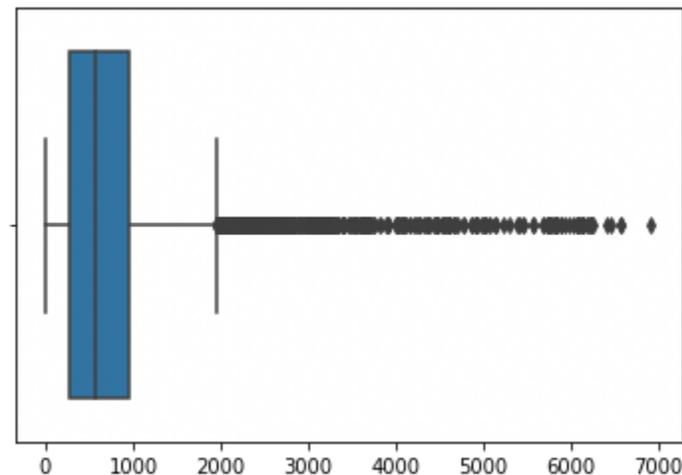
Beginning in 2017, I set up a google alert for “young people protest” and “student protest.” As a historian of student activism in higher education, the initial purpose of the alert was to follow contemporary forms of student activism. In revisiting the alerts in 2020, I realized that the google alerts provide a tidy temporal lens on student activism during the three years of the Trump administration. From a scholarly standpoint, I was curious if the google alerts (as a dataset/text corpus) can offer any insight into contemporary forms of student activism over the last three years, both under the Trump administration and, more generally, in the context of the rising of authoritarian tendencies in regions like India and Brazil. More recently, as I explored ways to integrate digital inquiry into my fall 2020 course, I also began to see the google alerts (as a dataset/text corpus) as a potential teaching resource. Could my students, using tools like Voyant and ArcGIS, identify trends in that data/text corpus? How would they interpret those trends as students? And what arguments might they make as young scholars about contemporary student activism

Corpora Creation Notes

Step 1: Extracted 20,327 news URLs from 2,326 emails

Step 2: Downloaded and parsed urls into title, body text, etc. 17,014 articles were successfully downloaded and parsed.

Step 3: Checked the word count of each article to remove potentially bad data. Limited articles to those between 500 and 2000 words, which is approximately the 50th percentile up to the outlier limit (75th percentile + (1.5 * IQR)). 8,875 articles were of this length.



Step 4: Sample the articles into 4 corpora:

- a) Top Publishers: All articles from the top 5 publishers represented were extracted.
- b) Sentiment: All articles were quantified. The 500 most negative and 500 most positively ranked articles were extracted.
- c) Politics: All articles were searched for four term groups. 4,095 articles had at least one term. The 250 articles with the most mentions in each term group were extracted. There can be overlap between these groups.
- d) Chronology: The articles were split into three years (4-2017 to 3-2018, 4-2018 to 3-2019, 4-2019 to 3-2020). A random sample of 30 articles from each month of those yearly groupings was taken.

Limitations:

Given that this database relied on google alerts (and its ever-shifting algorithms) and narrow set of search terms (“young people protest” and “student protest”), I am mindful of its range of limitations. The reliance on “protest” as the operable term and google alerts limited the database and may not capture the smaller moments of organizing and meetings among students. In this regard, I recognize that this corpora archive does not capture the breadth and depth of student protest over the last three years, especially those instances that might not capture national attention. I also recognize that the sources are news websites reporting on student activism. For teaching purposes, see references below to contextualize the limitations.

Research / Teaching Explorations:

While not providing direct primary source material on contemporary student activism, the limitations of the text corpora presents a unique avenue of inquiry for students: how national and international newspapers (mainstream) framed student activism with local contexts and across national borders (e.g. "narrative analysis"). Consider some of the following questions when experimenting with the corpora archive:

- Compare and contrast the sentiment analysis corpora. What do you notice in comparing “negative reporting” versus “positive reporting”? What does comparing the two reveal about student activism?
- Compare and contrast the top publishers corpora. How did newspapers report on student activism? Were there differences across country papers? (NOTE: You can run this experiment with the sentiment analysis).
- Compare and contrast the political groupings. What was the nature of student activism the past three years according to this database (That is, what were the political motivations, interests, and/or tactics)?
- Visualize each corpora using DreamScape and examine the links in context. Then, consider: To what extent and in what ways (how) was student activism linked over the past 3 years?
- Examine the corpora (chronology) using the trends tool. What trends emerge when exploring student activism from 2017-2020?

References

- Blevins, Cameron. 2019. "Intro to the Stacks: A Tour of the Virtual Stacks." *Modern American History*, 2 doi:10.1017/mah.2019.16.
- Brier, Stephen. 2017. "Confessions of a Premature Digital Humanist." *The Journal of Interactive Technology & Pedagogy*. 11.
- Drucker, Johanna. 2011. "Humanities Approaches to Graphical Display." *Digital Humanities Quarterly*. 5, 1.
- Noble, Safiya. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: New York University Press.
- Poirier, Lindsay. 2020. "Ethnographies of Datasets: Teaching Critical Data Analysis through R Notebooks." *The Journal of Interactive Technology & Pedagogy*. 18.
- Putnam, Lara. 2016. "The Transnational and the Text-Searchable: Digitized Sources and the Shadows They Cast." *The American Historical Review*. 121, 2.