**Text as Data Final Project**

**Topic:** Sentiment analysis of the War in Israel and Gaza.

**Scope:**

* Media/news articles or official government statements?
* English only? According to Max, it’s possible to include non-English texts, however, it will require some sort of validation method to ensure comparability.
* Focus geographical split (western vs. non-western) or political split (left vs. right)?
* Time period? Limit to the first 30 days of the conflict?
* Topic modeling + sentiment analysis?

**Methodology:**

**Data sources:**

* Media/news outlets
  + BBC (UK)
  + The Guardian (UK)
  + CBC (Canada)
  + DW (Germany)
  + France24 (France)
  + CNN (USA)
  + Fox News (USA)
  + NBC (USA)
  + New York Times (USA)
  + Vox (USA)
  + Al Monitor (USA)
  + Al Jazeera (Qatar)
* Government statements
  + [The White House](https://www.whitehouse.gov/?s=israel+gaza) (USA)
  + [PM Justin Trudeau](https://www.canada.ca/en/privy-council/pm/search.html?q=israel+gaza&wb-serch-sub=Search&st=s&num=10&s5bm3ts21rch=x&st1rt=0&langs=eng&cdn=pm&default_text=) (Canada)
  + [UK Parliament](https://hansard.parliament.uk/search?startDate=2018-10-07&endDate=2023-11-13&searchTerm=israel%20gaza&partial=False&sortOrder=1) (UK)
  + [Australian Parliament](https://www.aph.gov.au/Parliamentary_Business/Hansard/Search?__VIEWSTATEGENERATOR=20B6B7A5&ind=0&st=1&sr=0&q=gaza&hto=1&expand=False&drvH=7&drt=2&pnu=47&pnuH=47&f=26%2F07%2F2022&to=14%2F11%2F2023&pi=0&pv=&chi=0&coi=0&ps=10) (Australia)
* UN Speeches
  + [UN Security Council speeches](https://digitallibrary.un.org/search?ln=en&cc=Speeches&p=&f=&rm=&sf=&so=d&rg=50&c=Speeches&c=&of=hb&fti=0&fct__2=Security%20Council&fct__8=PALESTINE%20QUESTION&fti=0)

**Examples from later year:**

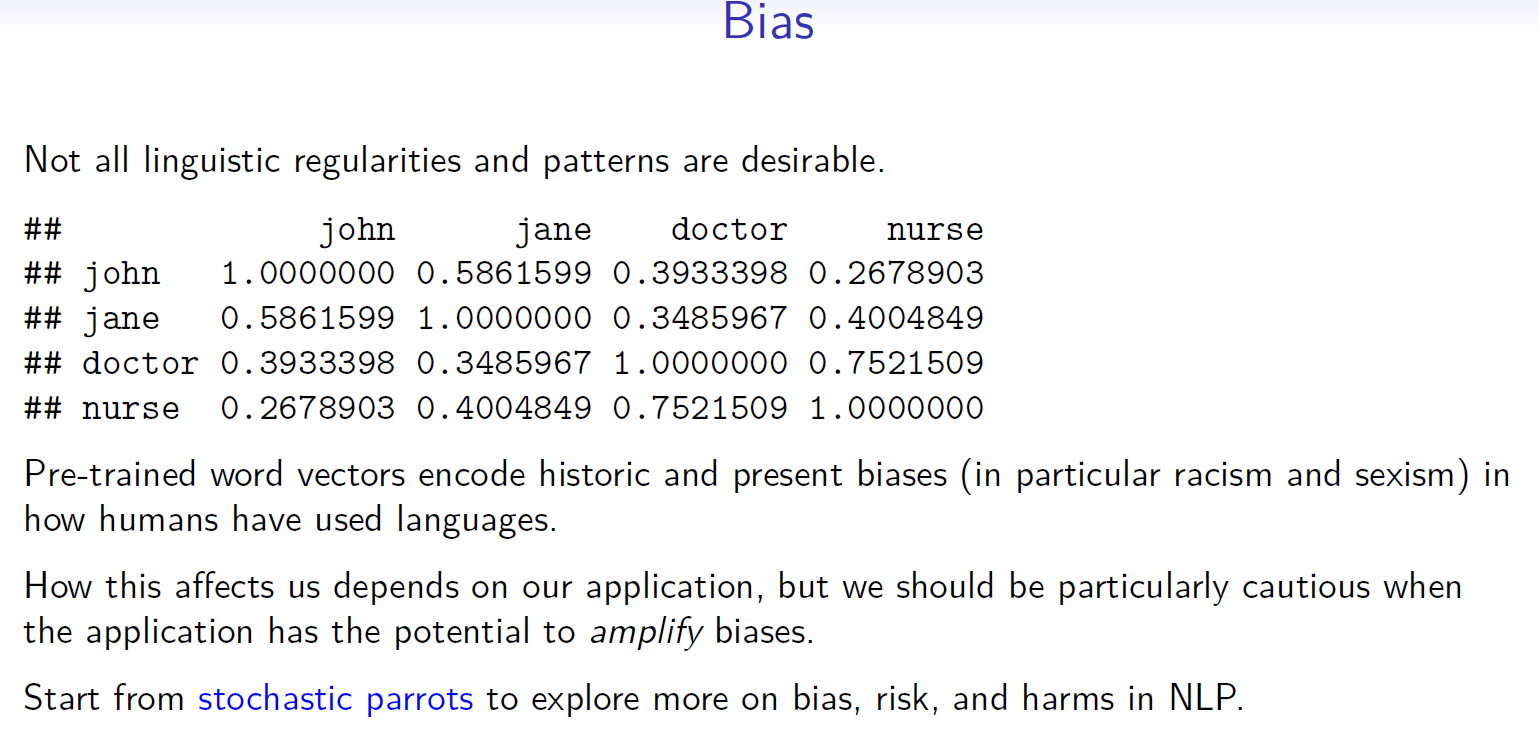
* <https://github.com/lwarode/text-as-data_political_scaling>
* <https://github.com/annaccd/assignment_3_brazil_elections>
* <https://github.com/benediktstroebl/Emotionality-in-Political-Speech-of-the-EU-Commission>

Possibl extension:

Database for measuring levels of violences/incidents:  
  
The Armed Conflict Location & Event Data Project (ACLED) logs and maps political violence, demonstrations, civilian deaths, and many more types of individual incidents at a granular level. (<https://acleddata.com/israel-palestine/>). This could be used as a measurement of what happens on the ground and how different types of incidents. I have made a quick download of the dataset, it includes: "Riots",  
"Protests",  
"Strategic developments" ,  
"Explosions/Remote violence"  
 "Violence against civilians"   
"Battles"   
  
We could use an algorithm to match articles with the individual incidents? I am not exactly sure yet how to link articles to the dataset. E.g. If we find articles that mention specific incidents, we can match it with the database, if it’s an article that is generally commenting on the situation, we could decide to correlate it to an overall measure of “violence” for that point in time?

**Minimal viable product**

Motivation: potential bias in Palestine and Israel by typing in words in word embedding space



word embeddings: cosine similarities dimension of suffering with further away from