**Scraping NYT:**

**3/12/2024**

For the first part of the project, I used the NYT article API to scrape the URL of all the articles published by NYT under their “The Israel Hamas War” category. Because the API has a request limit, I was having some trouble fetching all the articles, so I ended up placing a timer for 15 seconds per page. It is a paginated API so I randomly decided to use 100 pages, it actually ended up giving me articles as far back as 2009. In order to not have to request again and deal with API request issues, I saved all the URLs in a csv as a dataframe.

I think my next step is to use beautifulsoup to scrape all the article urls. I also need to just scrape 2024 and 2023 articles so need to add a filter for that in my initial code.

**4/12/2024**

My initial approach was to scrape all the articles using beautifulsoup but I was blocked by the NYT. Darryl suggested that I scrape just the headline and that should suffice for the text analysis. At one point, I felt like I should give up this project.

**6/12/2024**

Was able to scrape the headlines. I am trying to replicate this analysis by [The Intercept](https://theintercept.com/2024/01/09/newspapers-israel-palestine-bias-new-york-times/) but they parsed every sentence so I am wondering how limited my approach of looking at headlines is and if I should retry scraping the main content of the article.

So far from the headlines, I have been able to count how many times the words Israel, Gaza, Hamas and Civilians appear in the title. Also used wordcloud to make a plot of the most commonly occuring words in the title, I feel that plot does tell the story a bit but still using headlines is very limiting.

Looking at the NYT API, I see that it also gives the “snippet”, “abstract” alongside with the headline so using that to get something that would be a little less limiting.

**7/12/2024**

After using the API to scrape the text from the headlines, I did some basic analysis and visualization. I looked for patterns in how NYT was reporting on the deaths of Palestinian versus those in Israel, in terms of whether the newspaper was providing specific numbers or using terms like “dozens”, “scores of” and so on. I also looked for attribution phrases when speaking about Israeli deaths and Palestinian deaths as well as the use of passive voice. Major newspapers like NYT have used the word “deadliest” for Oct 7 but when it comes to the more than 40,000 Palestinians who have been killed since then, there is no such strong emotive word used.

For visualisation, I plotted bar charts.

I got to know about word cloud generator in Python and used it to generate word clouds– I first separated all the texts that were specifically about Israel and then those about Palestine and generated word clouds.

My next step would have been to scrape Al Jazeera for comparative analysis but I am unable to find its API, I don’t know if it is public.

**15/12/2024**

Set up my story template but haven’t yet written the captions or put the analysis in word.

**Earthquake auto-update**

7/12/2024

Followed all the instructions. Initially filtered for the west coast but when it came to mapping it, the map looked too concentrated as I am using all of USA so I filtered for California.