

COVID-19 AND MENTAL HEALTH IN THE NEWS

Natasha Carpio Castellanos,
Jayda Hart, Yue Yin, Kexin Zhang



BACKGROUND

- Mental health is historically a taboo; discussion is uncommon due to associated stigmas.
- In recent years, normalization of mental health discussion, even in mainstream media.
- The COVID-19 outbreak increased focus on mental health due to isolation; time for reflection.
- Analyze the online presence of mental health discussions, specifically anxiety and depression, two common emotional issues.



RESEARCH QUESTIONS

How are anxiety and depression framed in newspaper articles?

How has the frequency of articles related to anxiety and depression changed from the pre-COVID-19 era through the pandemic and into the present day, specifically in the time frames of lockdown?

DATA

The New York Times &
The Guardian



The New York Times

New York Times'
“Article Search API”.

```
begindate = "{}0101".format(year)
enddate = "{}1231".format(year)
nyt_params = {"q": word, "fq": f'headline:{word}',
              "begin_date": begindate, "end_date": enddate,
              "page": page, "api-key": api_key}

response = requests.get(base_url, params= nyt_params)
page_data = response.json()
time.sleep(14)
```

- Year 2004 - Year 2024 (Feb 7)
- 1115 articles for “anxiety”
- 716 articles for the “depression”
- Pro: articles’ authors as a proxy of the voice of society
- Con: filtering by section was not straightforward --> dirty data

	year	specific_date	section	subsection	headlines	abstracts	Score
0	2004	2004-12-10T05:00:00+0000	U.S.	None	Study Pursues a Genetic Link to Depression	Scientists at Duke University find genetic var...	8.0
1	2004	2004-12-17T05:00:00+0000	U.S.	None	Study Suggests Way to Predict Whom Antidepressants Work Best	Harvard and UCLA researchers find that common ...	8.0
2	2004	2004-12-07T05:00:00+0000	Health	None	Tracking Stress and Depression Back to the Womb	Studies indicate that stress and depression in...	8.0
3	2004	2004-09-12T05:00:00+0000	Business Day	None	Depression, a Frequent Visitor to Wall St.	Rate of chronic depression seems to be high am...	8.0
4	2004	2004-08-25T05:00:00+0000	U.S.	None	New Therapy On Depression Finds Phone Is Effective	Study reported in Journal of the American Medi...	8.0

The Guardian

```
: search_guardian_api(api_key, query, from_date, to_date):
base_url = "https://content.guardianapis.com/search"

from_date = '2004-01-01' #20 years ago
to_date = '2024-02-07'
query = 'anxiety'

params = {
    'api-key': '93eb884c-6abd-444e-aba9-95bba2d84dad',
    'q': query,
    'order-by': 'newest',
    'from-date': from_date,
    'to-date': to_date,
    'show-fields': 'headline,bodyText,webPublicationDate',
    'section': 'society' or 'wellness' #section is setted as society or
}
```

Total Sample Size:

3092 articles for depression and 2949 for anxiety

Time Period:

January 1, 2004 - Feburary 7,2024

Query parameters:

search query ('anxiety', 'depression'), ('from_date' and 'to_date'), section of the newspaper ('society' or 'wellness').

Con:

Dirty data

```
all_articles = []
page = 1
total_articles = float('inf') # Initialize to infinity

while len(all_articles) < total_articles:
    params['page'] = page
    response = requests.get(base_url, params=params)

    if response.status_code == 200:
        data = response.json() Loading...
        articles = data.get('response', {}).get('results', [])
        total_articles = data.get('response', {}).get('total', 0)

        for article in articles:
            headline = article.get('fields', {}).get('headline', '')
            publication_date = article.get('webPublicationDate', '')
            body_text = article.get('fields', {}).get('bodyText', '')
            all_articles.append({'Headline': headline, 'Publication Date': publication_date, 'Body Text': body_text})

        page += 1
    else:
        print(f"Error: {response.status_code}, {response.text}")
        break

return all_articles

api_key = '93eb884c-6abd-444e-aba9-95bba2d84dad'
query = 'anxiety'
from_date = '2004-01-01'
to_date = '2024-02-07'

anxiety_articles = search_guardian_api(api_key, query, from_date, to_date)
df = pd.DataFrame(anxiety_articles)
df.to_csv('anxiety_articles.csv', index=False)
```

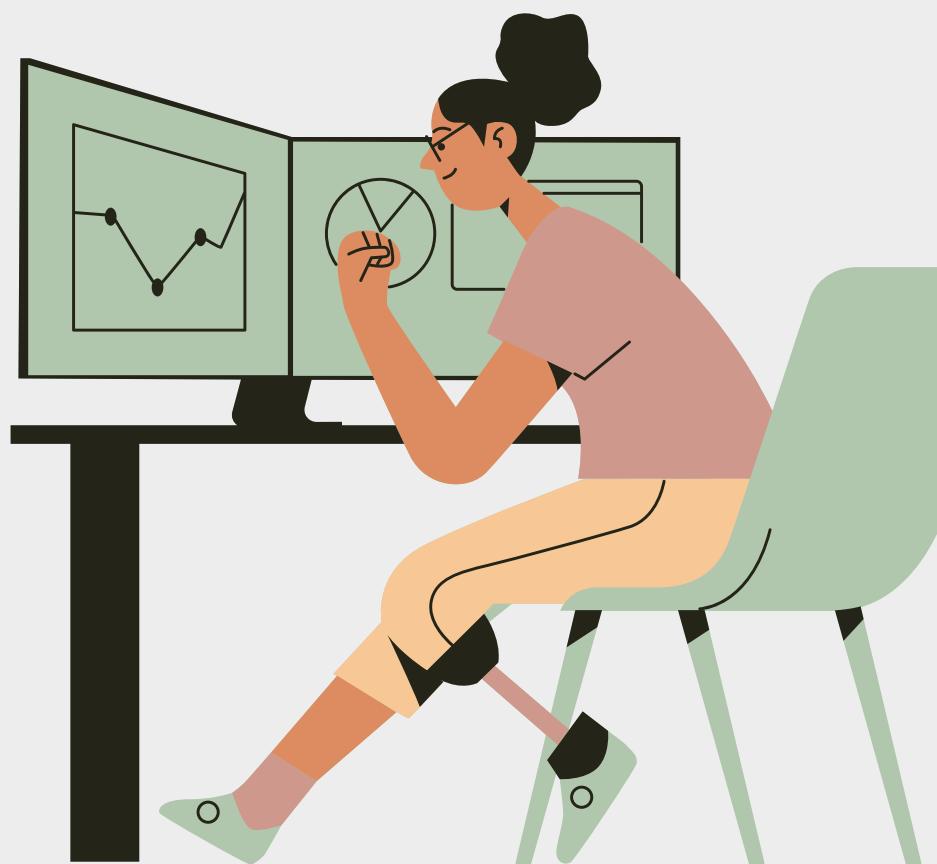
We manage pagination and rate limits by fetching articles **page by page** to avoid overwhelming the API with too many requests at once.

```
"text": [
    "Total number of articles with 'anxiety' in the headline: 2949\n",
    "\n",
    "
                    Headline          Publication Date
"0 Children's emergency mental health referrals i... 2024-02-07T00:01:04Z
"1 Women with polycystic ovary syndrome may have ... 2024-02-05T17:00:37Z
"2 'Rage would bubble out of nowhere': my battle ... 2024-02-04T10:00:42Z
"3 Nearly a third of UK secondary pupils avoid sc... 2024-02-02T17:24:46Z
"4 'I plan fewer outings': Britons on the scarcit... 2024-02-02T16:00:06Z
"5 Trans people in England more likely to have lo... 2024-02-01T07:00:43Z
"6 Early release of mental health patients in Eng... 2024-02-01T00:01:36Z
"7 Pliers, abscesses and agonising pain: Britain'... 2024-01-30T10:00:26Z
"8 UK ministers vow to close loopholes in disposa... 2024-01-29T08:25:14Z
"9 Disposable vapes to be banned in UK as part of... 2024-01-28T22:53:27Z
"\n",
"
                    Body Text
"0 The number of children referred to emergency m... \n",
"1 Women diagnosed with polycystic ovary syndrome... \n",
"2 The first time I learned anything useful about... \n",
"3 Almost three in 10 secondary pupils are avoidi... \n",
"4 Urinating in public made the headlines this we... \n",
"5 Transgender people in England are much more li... \n",
"6 Mental health patients in England are being pu... \n",
"7 When Lucy finally found a dentist that would a... \n",
"8 UK ministers will aim to stop firms from skirt... \n",
"9 Ministers are to ban disposable vapes as part ... \n"
]
```

DATA CLEANING AND WRANGLING

NEW YORK TIMES

- Filter unrelated articles (ex: “Great Depression”, “economy”)
- Create variable for different time periods, topic (anxiety or depression), newspaper source
 - March 11, 2020 — WHO Declares COVID-19 a Pandemic
 - May 11, 2023 — End of public health emergency
- Tokenize words for sentiment analysis and word counts



THE GUARDIAN

- Use `pd.concat` to combine anxiety and depression articles into a single DataFrame.
- Use ``duplicated`` function to filter our duplicated Titles.
- Create mask to find and remove duplicates articles.

	NYT	Guardian	Final DF
Total	1323	4745	6068
Anxiety	833	1688	2521
Depression	490	1804	2294
Both		1253	1253
Pre Covid	1036	3776	4812
Covid	248	775	1023
Post Covid	39	194	233

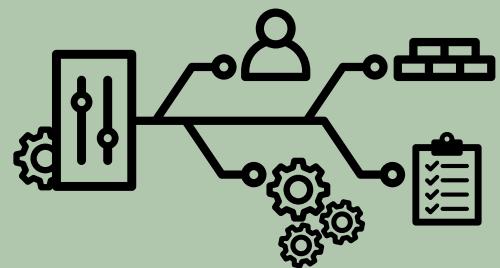
DATA ANALYSIS & VISUALIZATION

Insights from the Proportion and Frequency of Published Mental Health-Related Articles

- We calculated and visualized the proportion of anxiety and depression articles published each year from 2004 to 2023.

Why Proportion?

Method

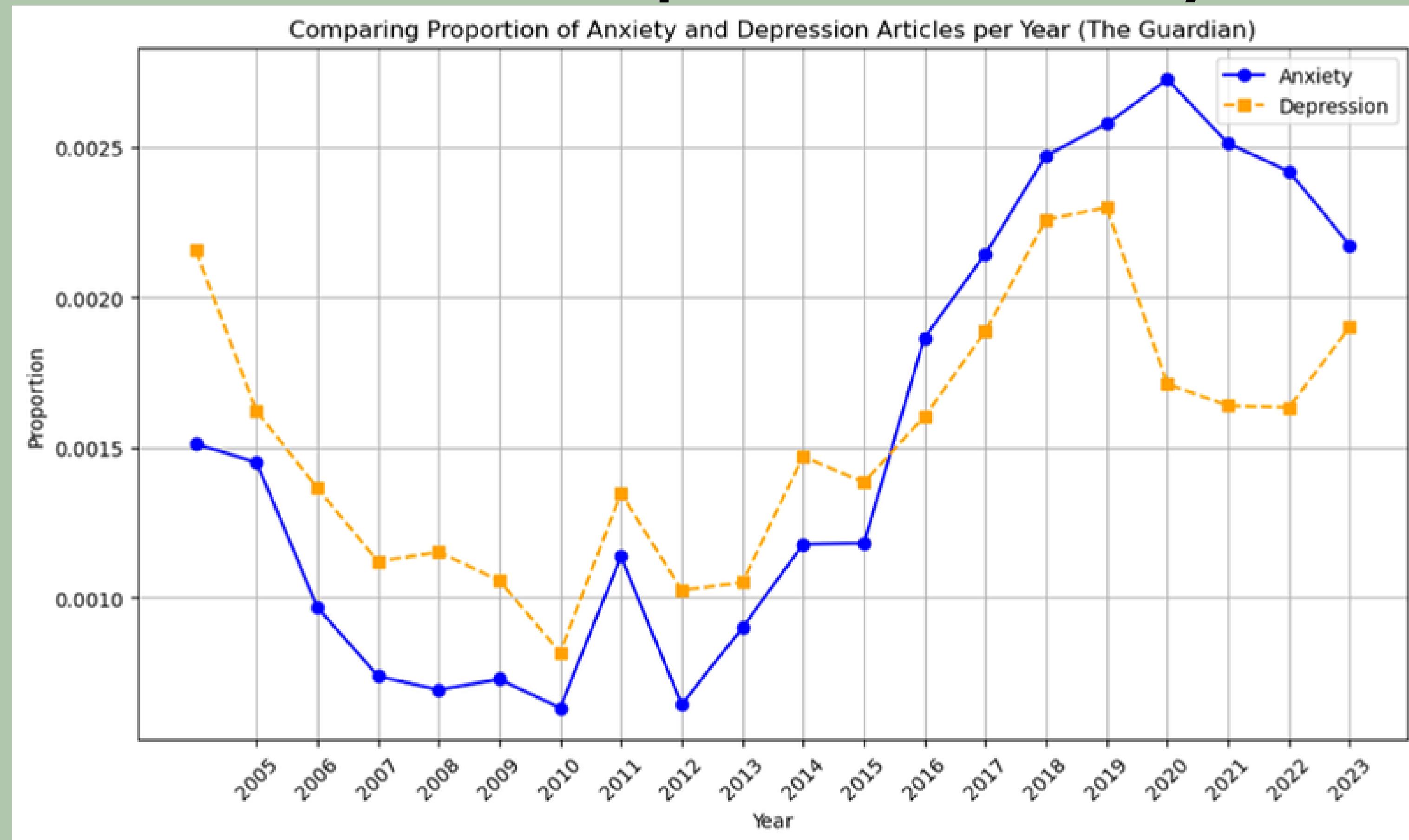


- We first extracted data on the total number of articles and the number of anxiety or depression articles for each year.
- We then calculated the proportion and plot line graphs serve any trends or changes

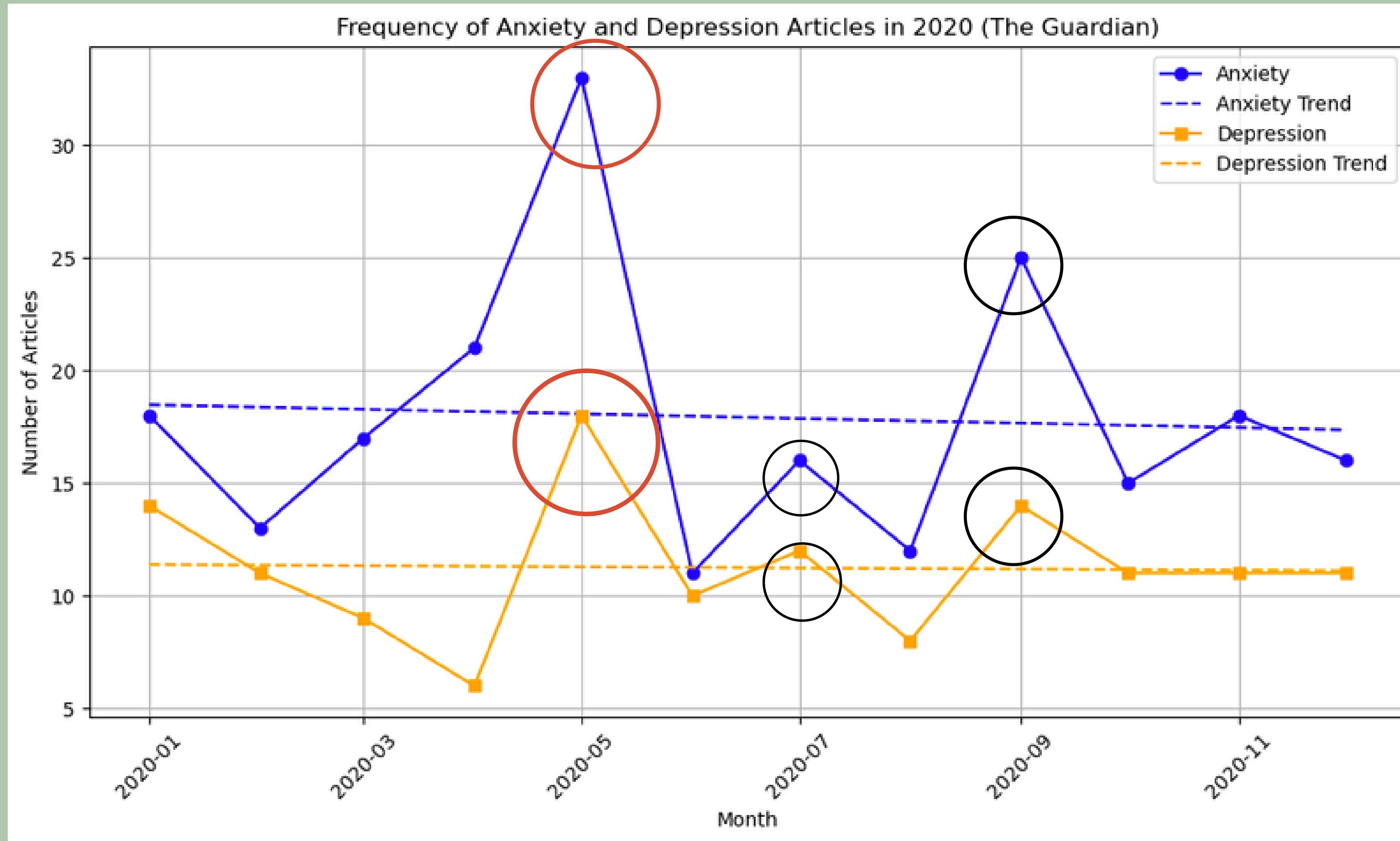


Year	Total Articles
2004	97873
2005	99939
2006	108381
2007	125860
2008	144090
2009	122035
2010	108920
2011	118663
2012	120930
2013	125551
2014	123105
2015	118486
2016	110396
2017	85363
2018	80131
2019	78753
2020	78924
2021	76842
2022	78965
2023	77315
2024	11255

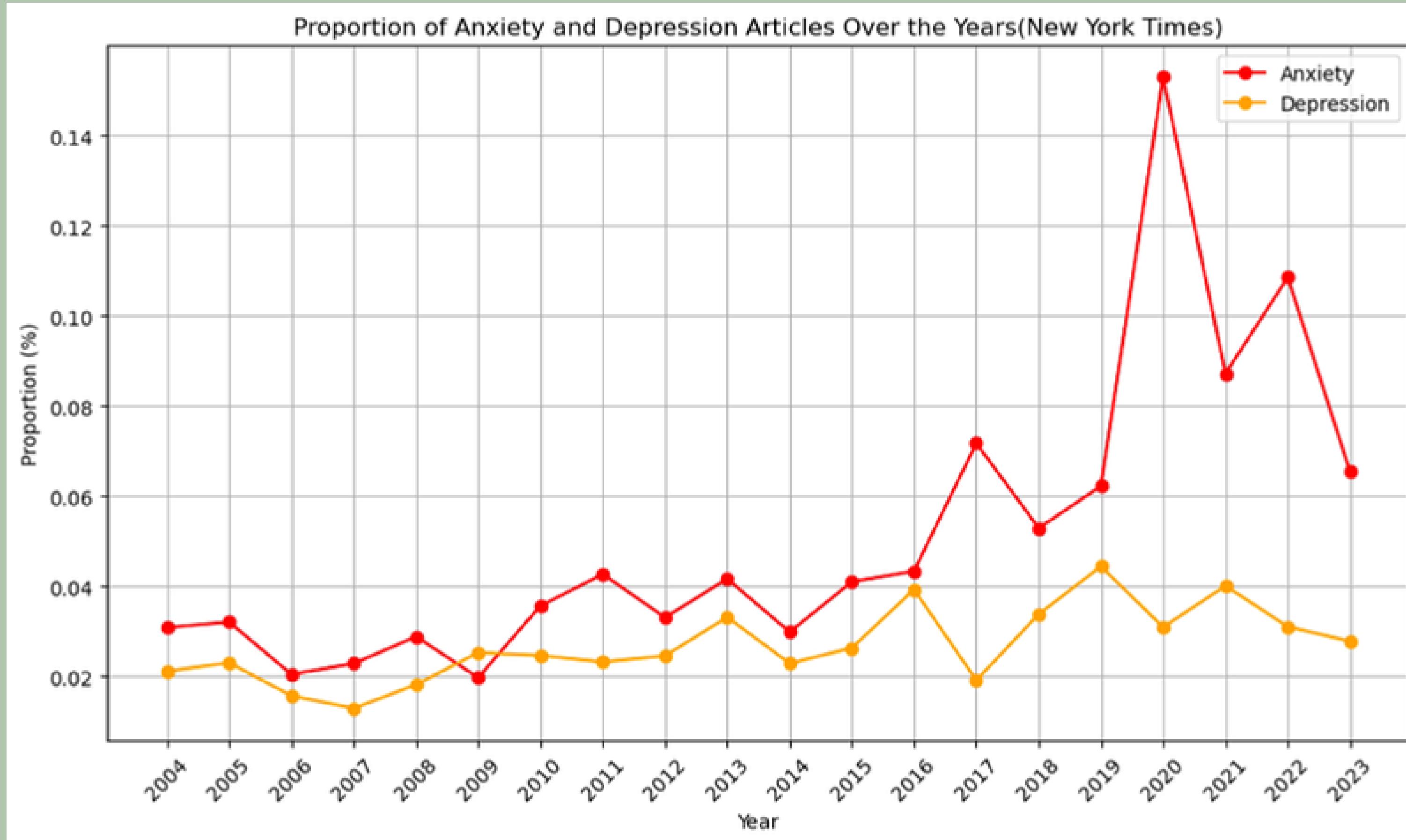
The Guardian - Proportion over 20 years



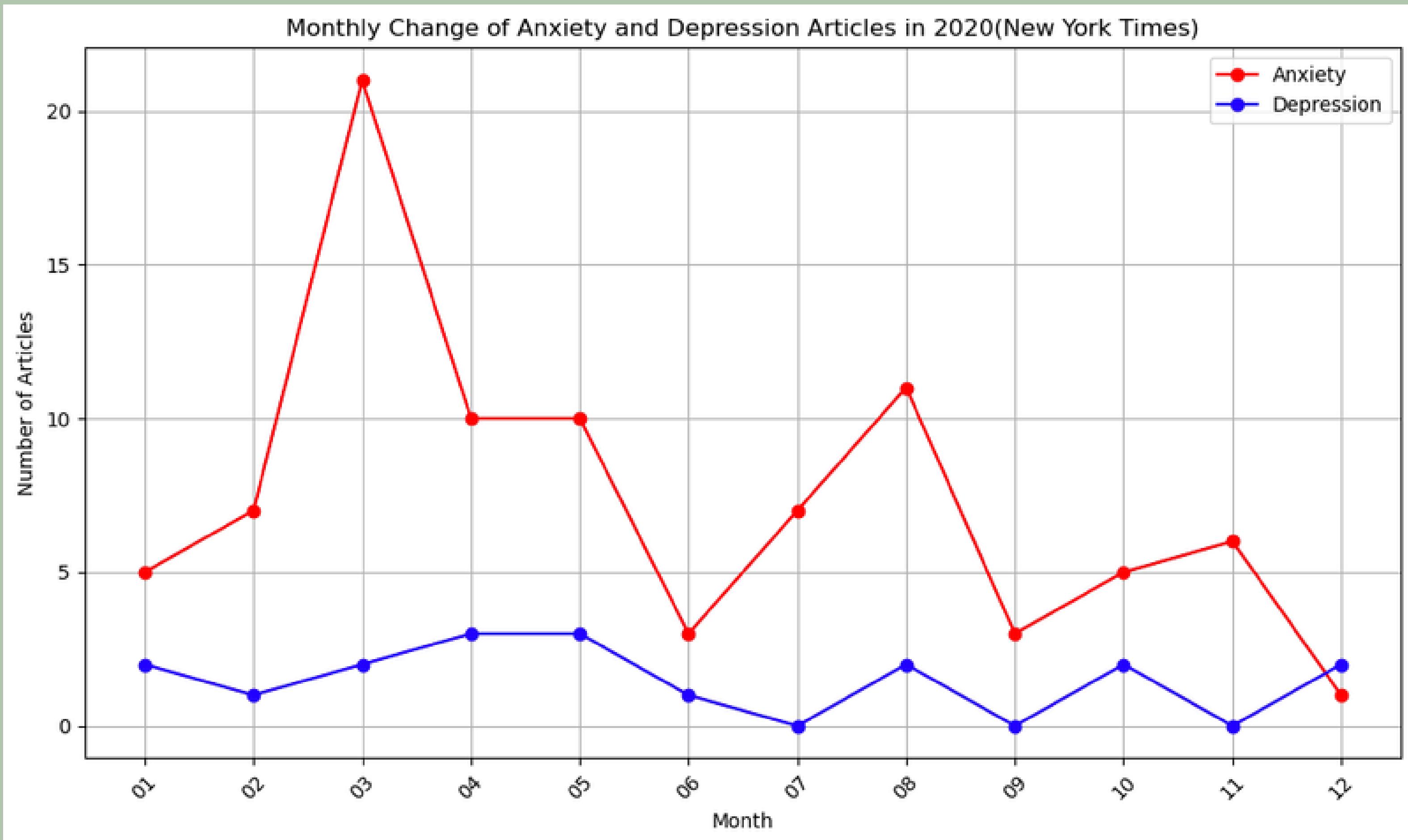
The Guardian - Frequency in 2020



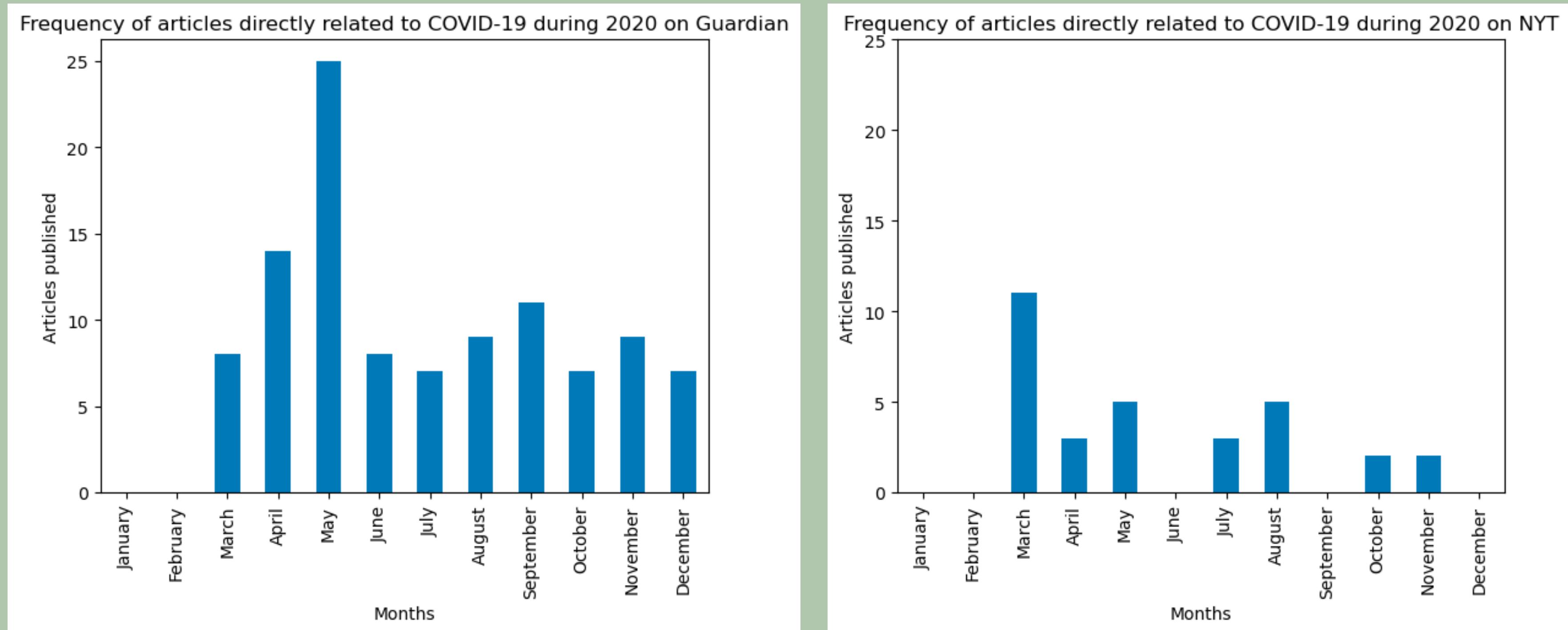
The New York Times - Proportion over 20 years



The New York Times – Frequency in 2020



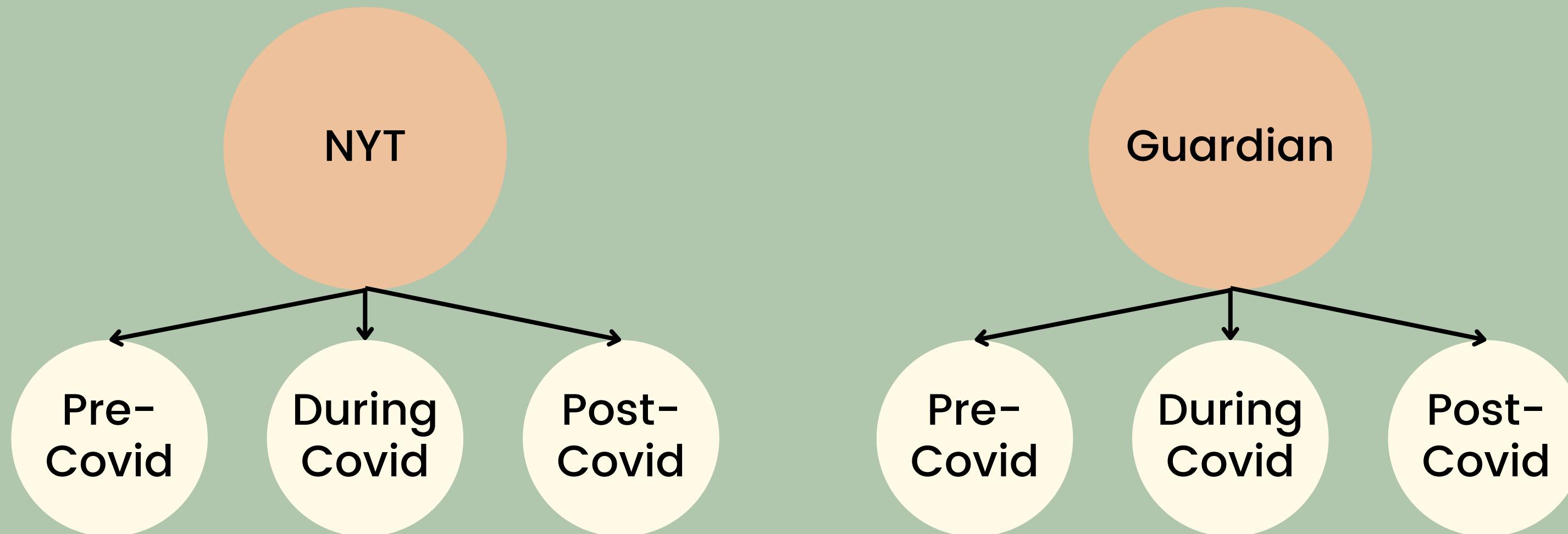
Frequency of COVID-19 articles during 2020



Surprisingly, not many articles were strictly related to covid-19 even if we did see a peak in 2020. Increase of mental health discussion could be also influenced by other factors such as different generations, normal evolution, etc.

WORDS' FREQUENCIES

- We separated the cleaned data into six groups, separating by newspaper and time frame

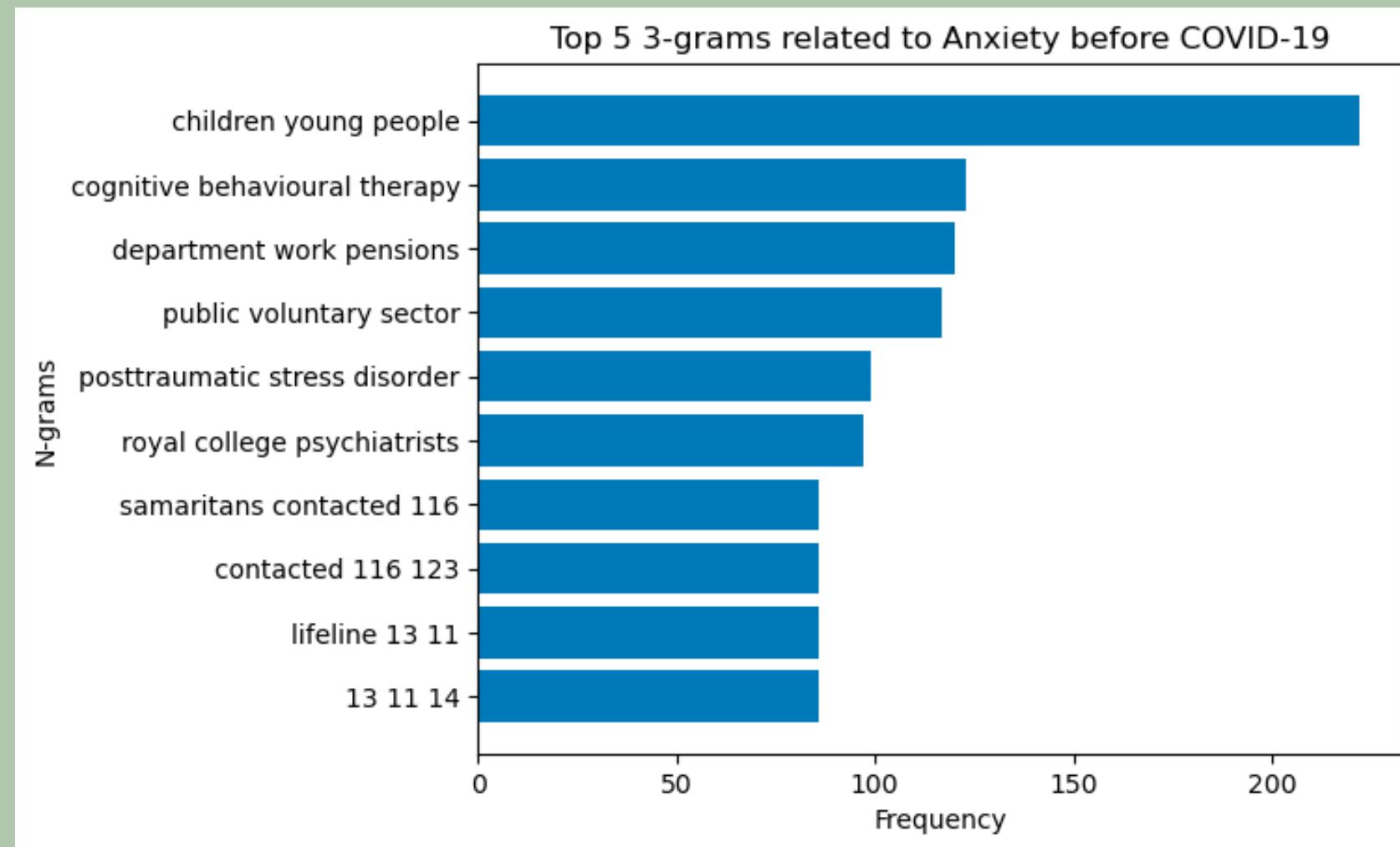


Most used words in NYT articles during COVID-19

Most used words in The Guardian articles during COVID-19

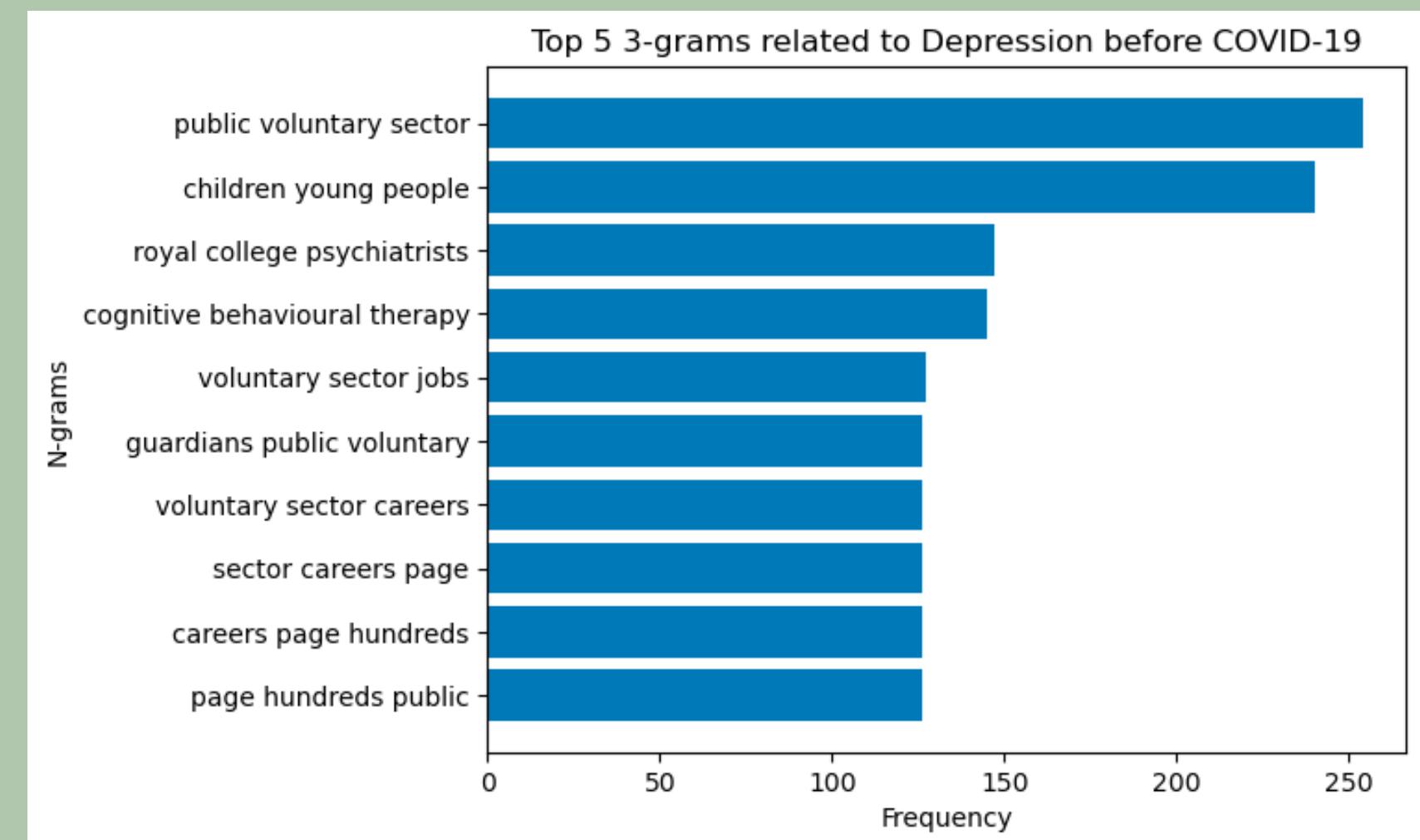


N GRAMS

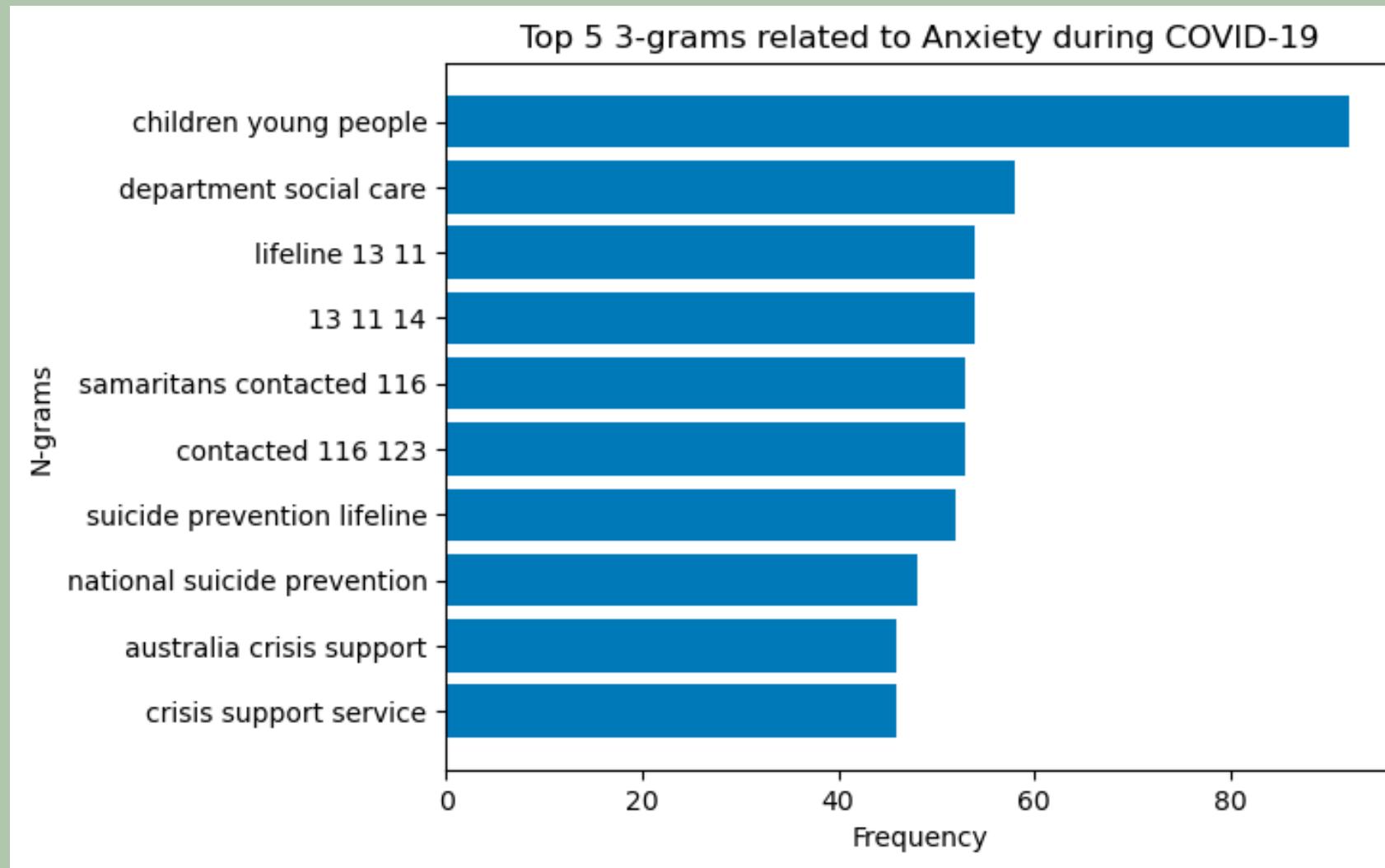


Help numbers come up

Before COVID-19 Anxiety vs Depression

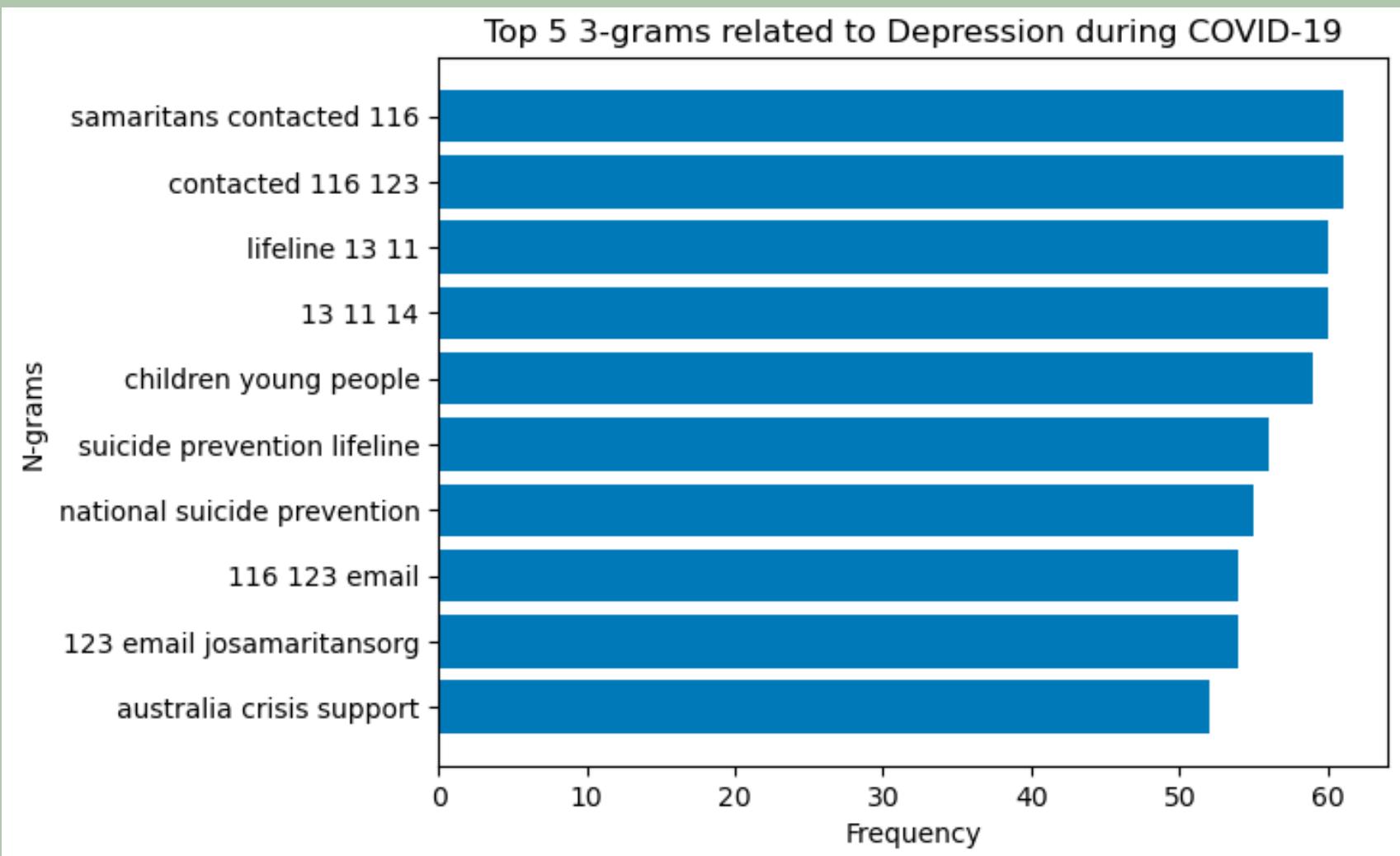


During COVID-19 Anxiety vs Depression

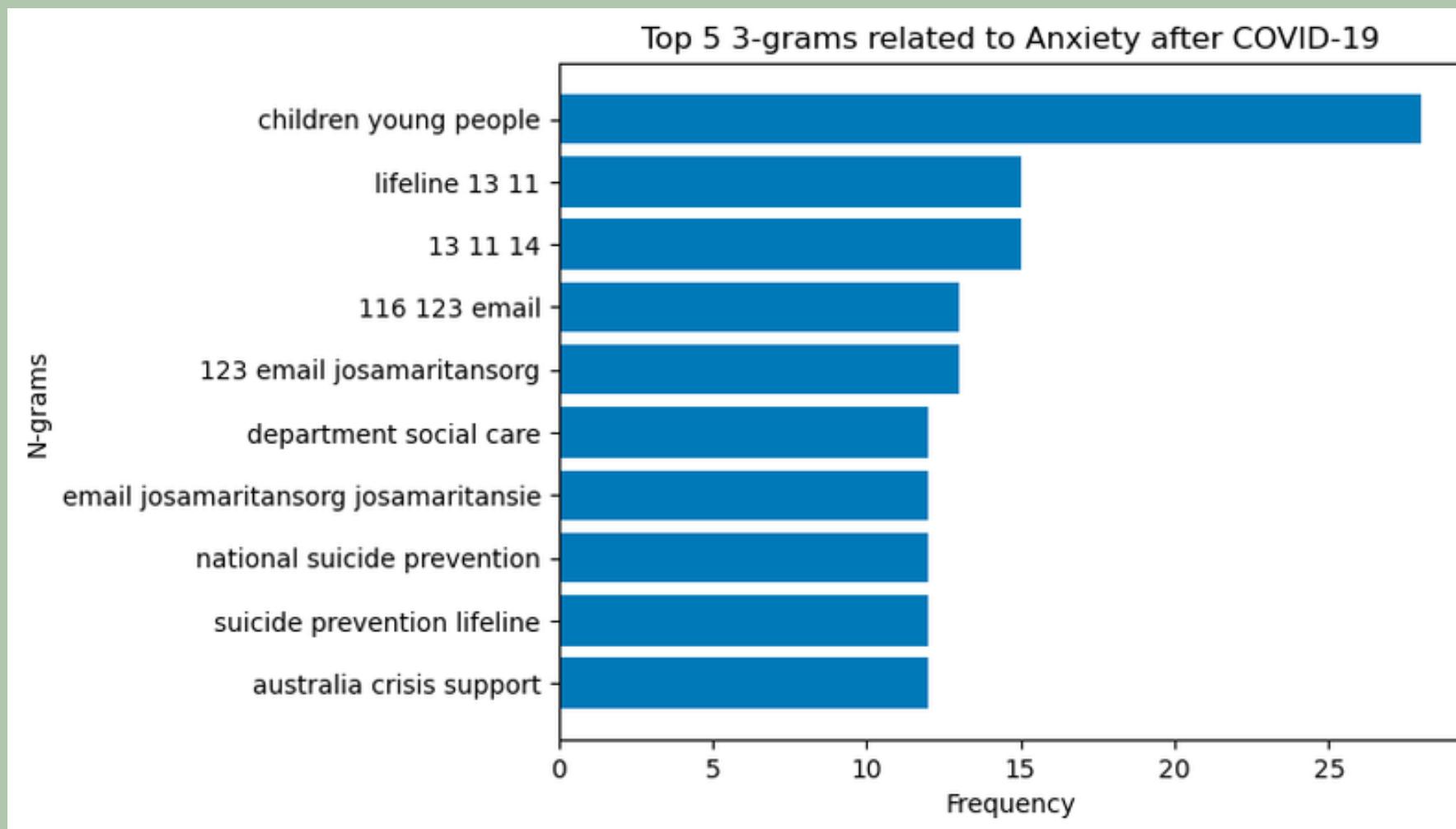


Help numbers still come up, including
suicide prevention

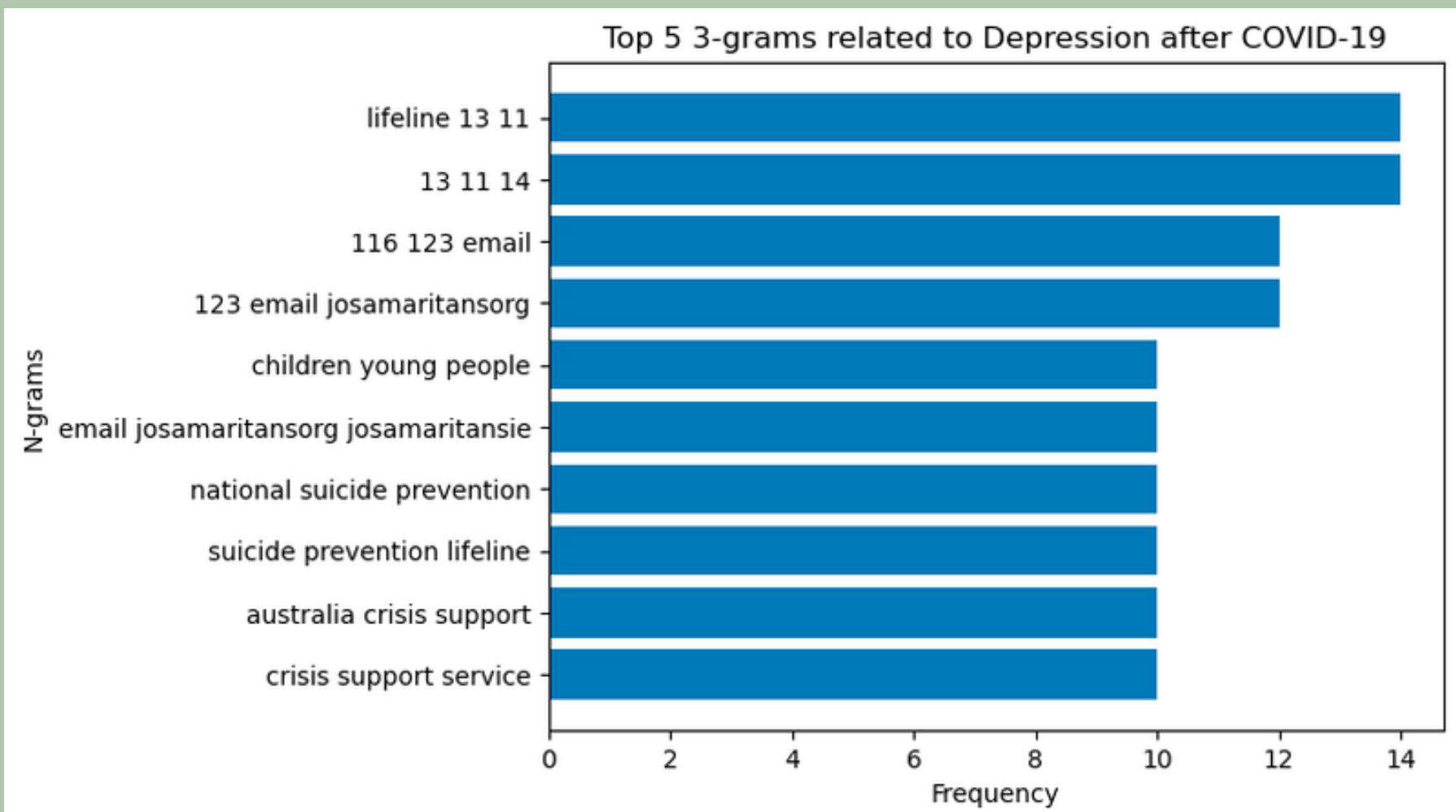
Emails also come up



After COVID-19 Anxiety vs Depression

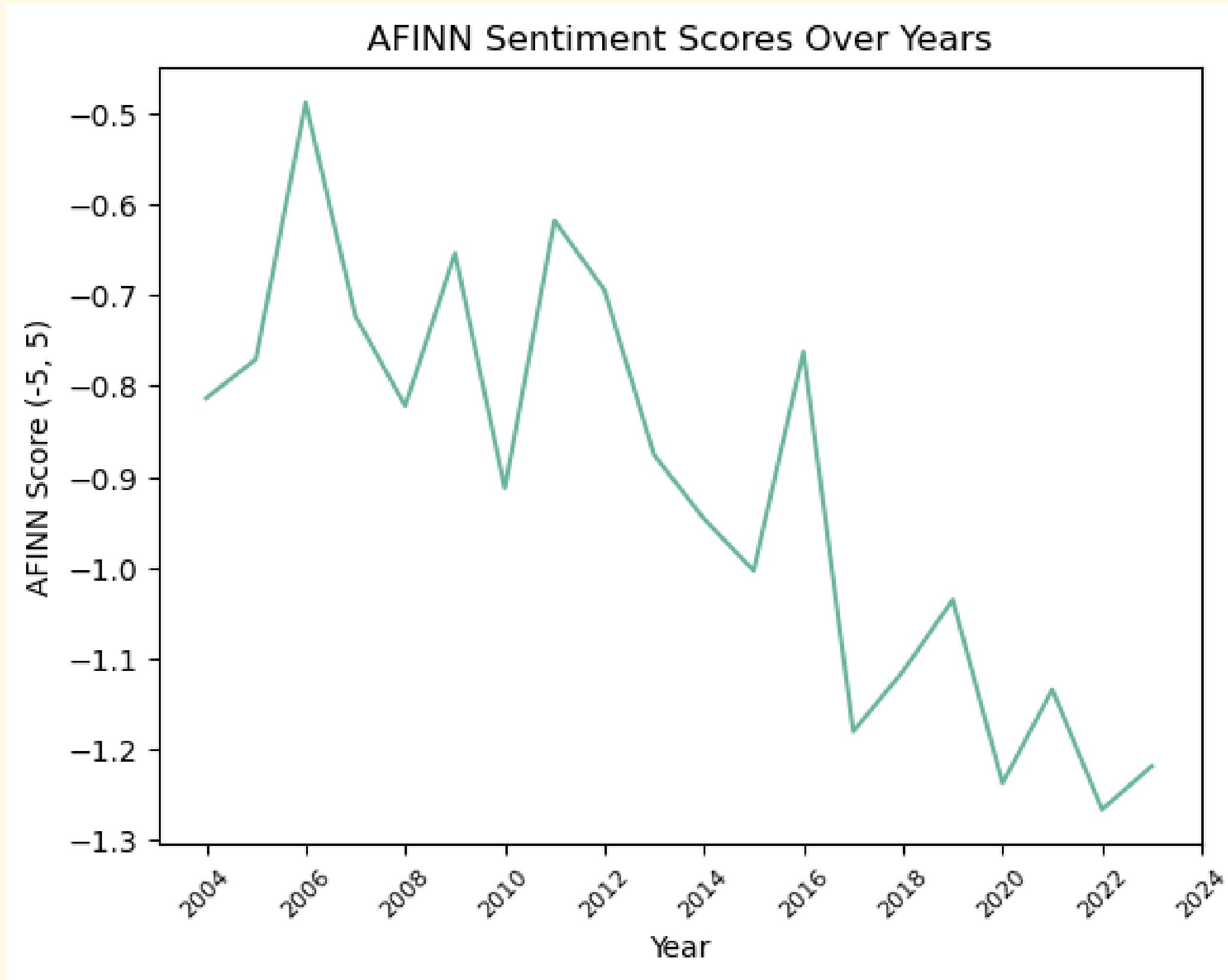


Numbers and emails still appear,
but the proportion is reduced

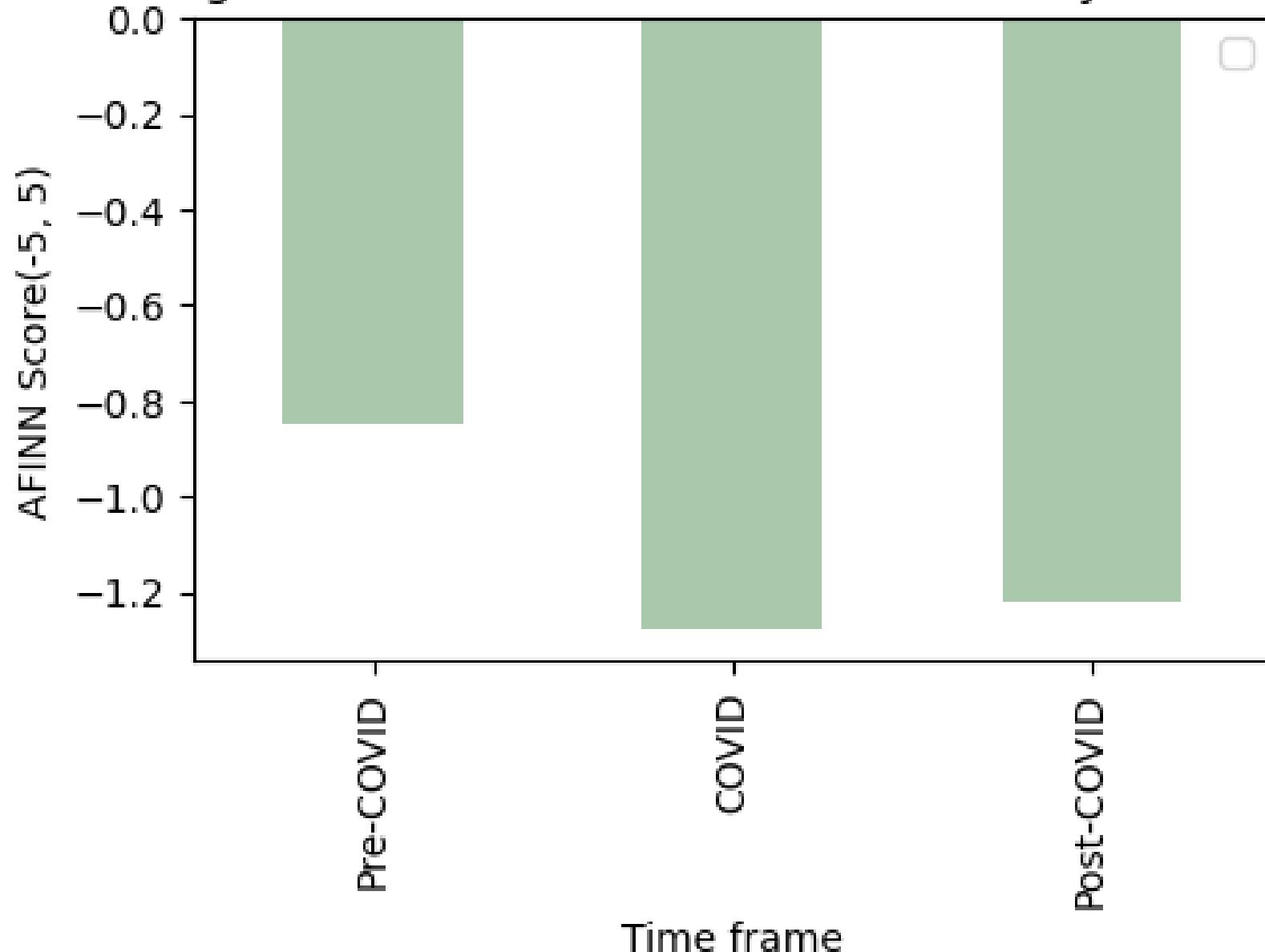


AFINN Sentiment Lexicon

A list of English words rated for
valence by Finn Årup Nielsen
from negative to positive, on a
scale from -5 to 5



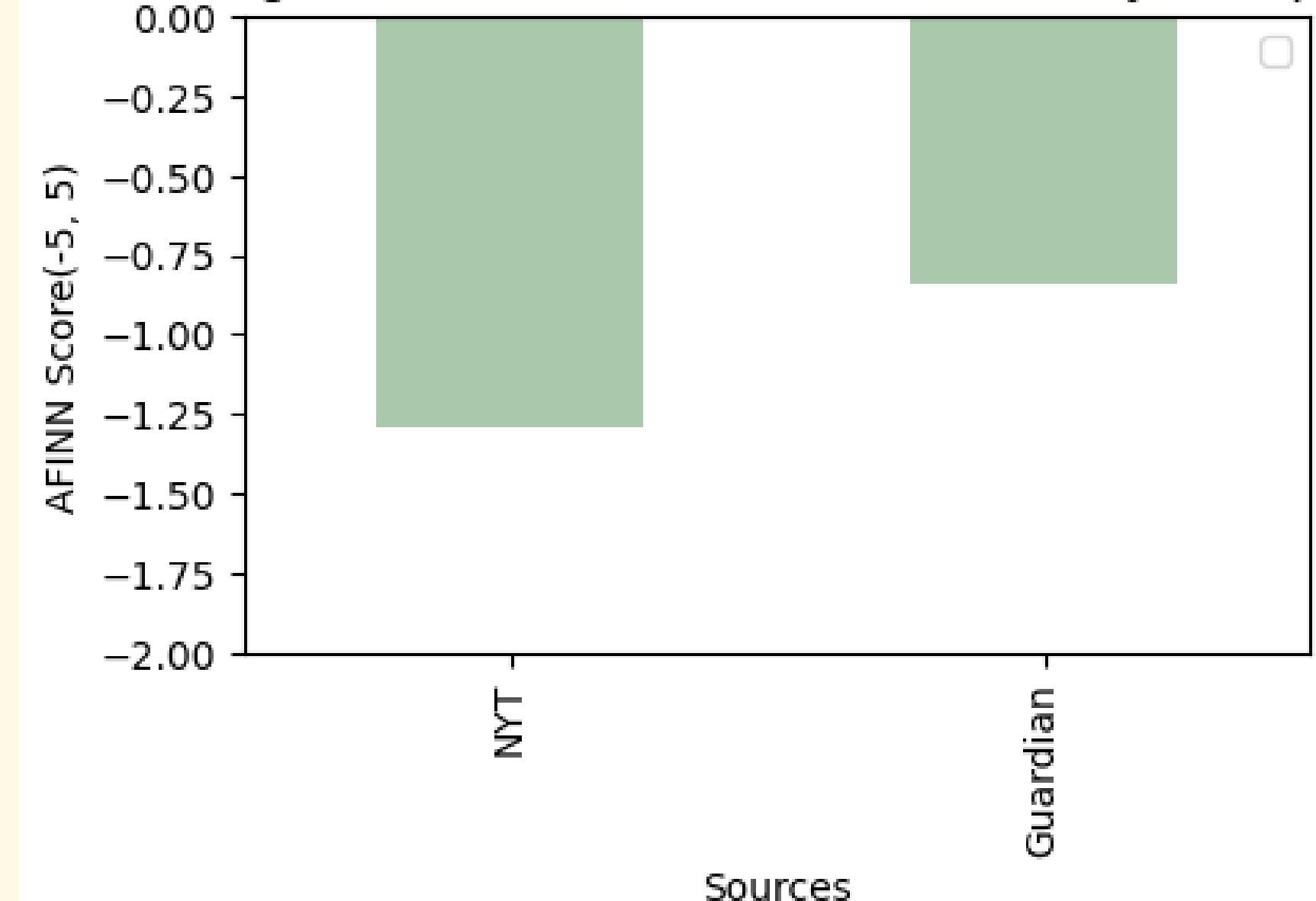
Average AFINN Sentiment Score for Titles by Time Frame



AFINN Score is most negative during Covid and remains low afterwards

NYT has more negative language overall

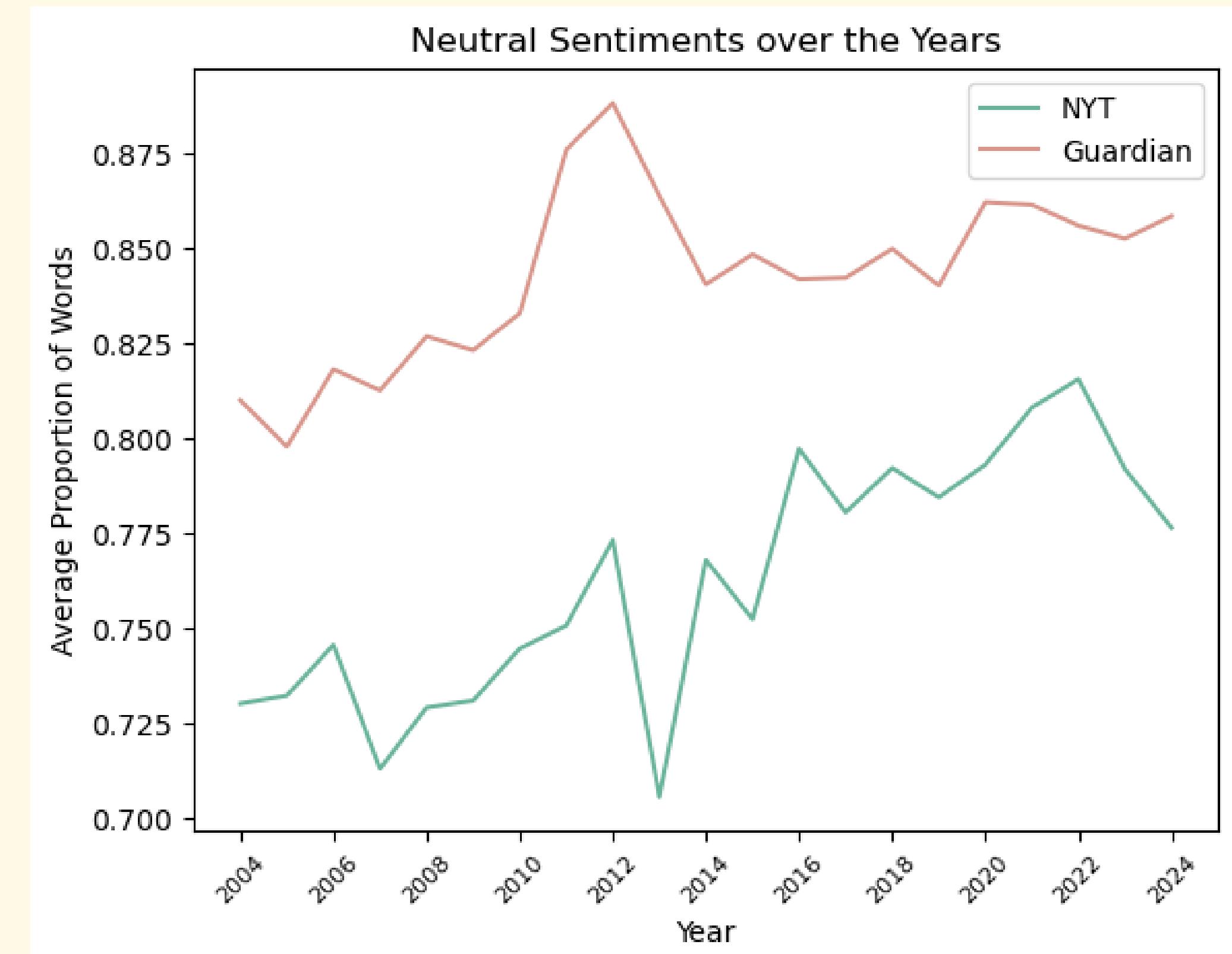
Average AFINN Sentiment Score for Titles by Newspaper



NRC Emotion Lexicon

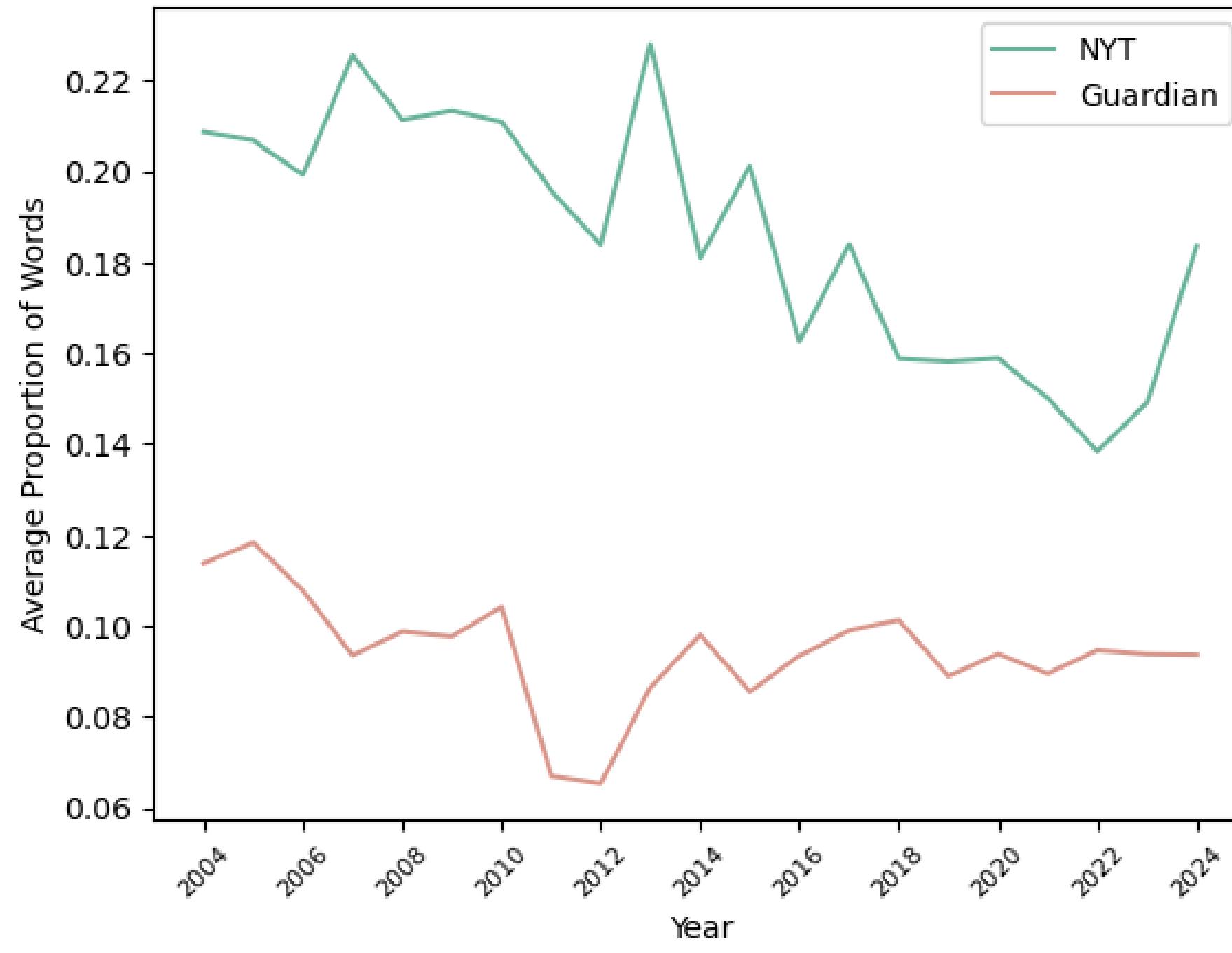
A list of English words and associations with eight emotions (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust) and the sentiments (negative and positive), rated either 1 or 0. This data

source was created by crowd-sourcing through Mechanical Turk



Increase in proportion of words that are neither negative nor positive

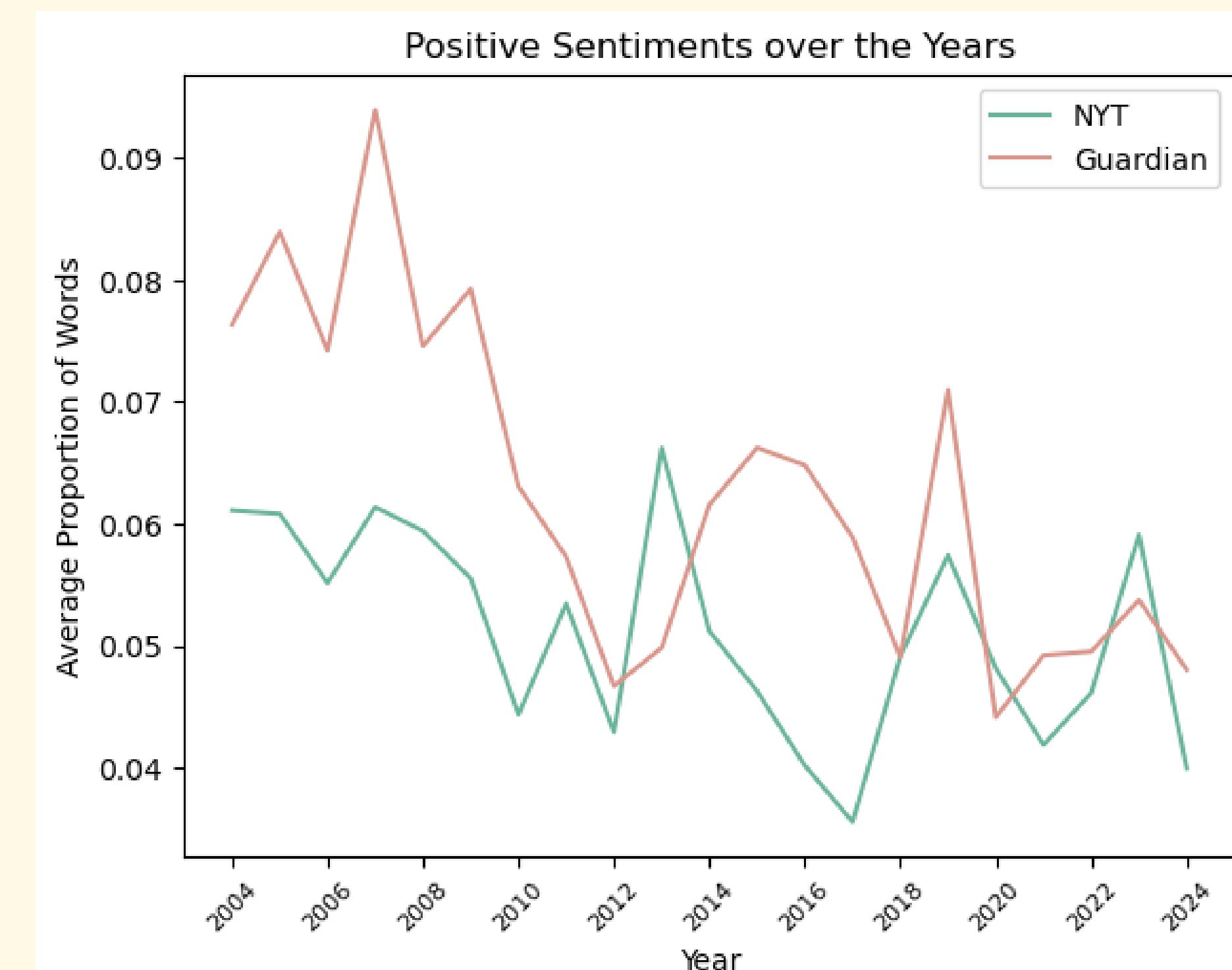
Negative Sentiments over the Years



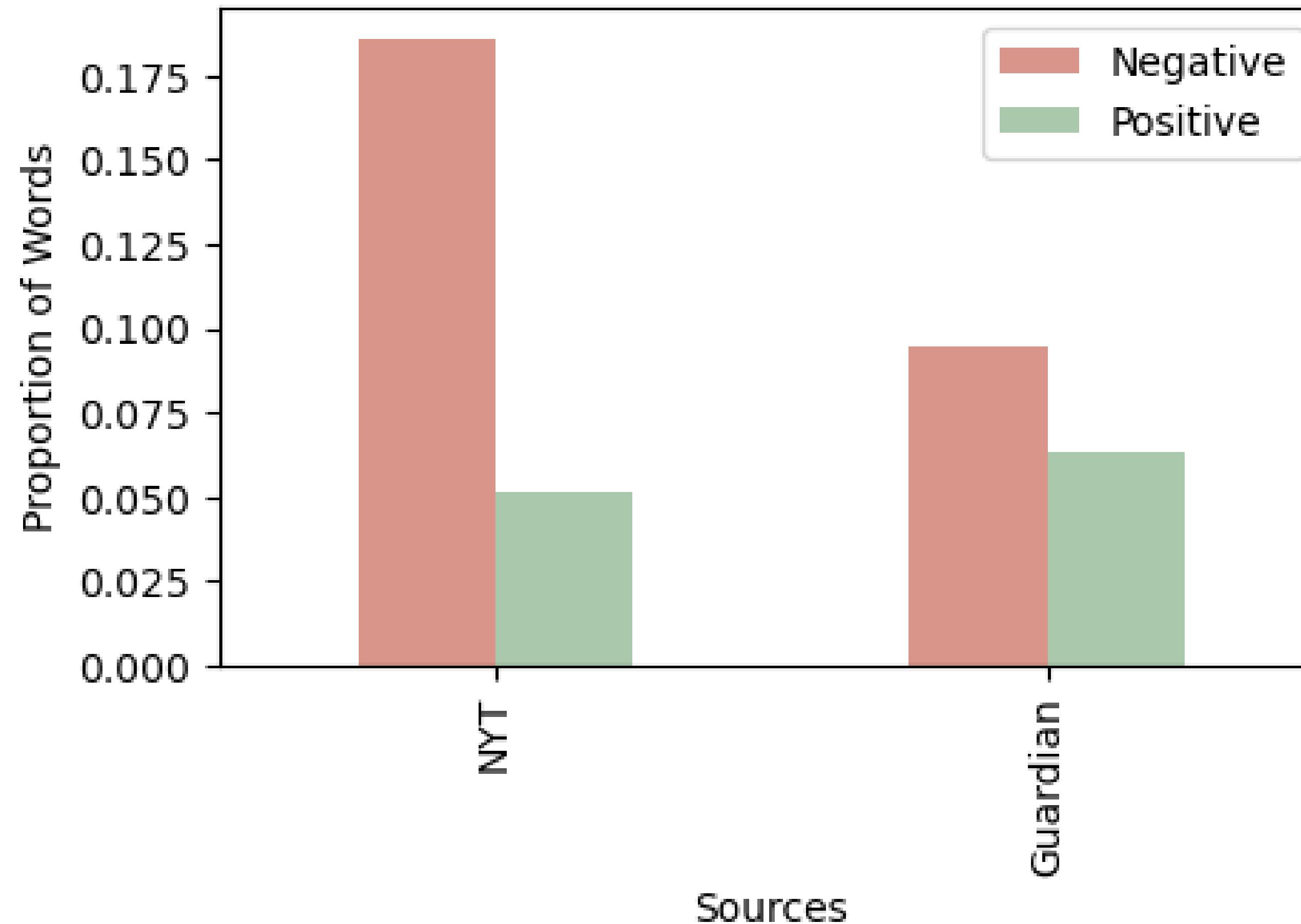
NYT is more negative and decreases slowly and The Guardian remains generally consistently low.

Guardian starts more positive and decreases at a more rapid rate, while the NYT starts out low and remains consistently low

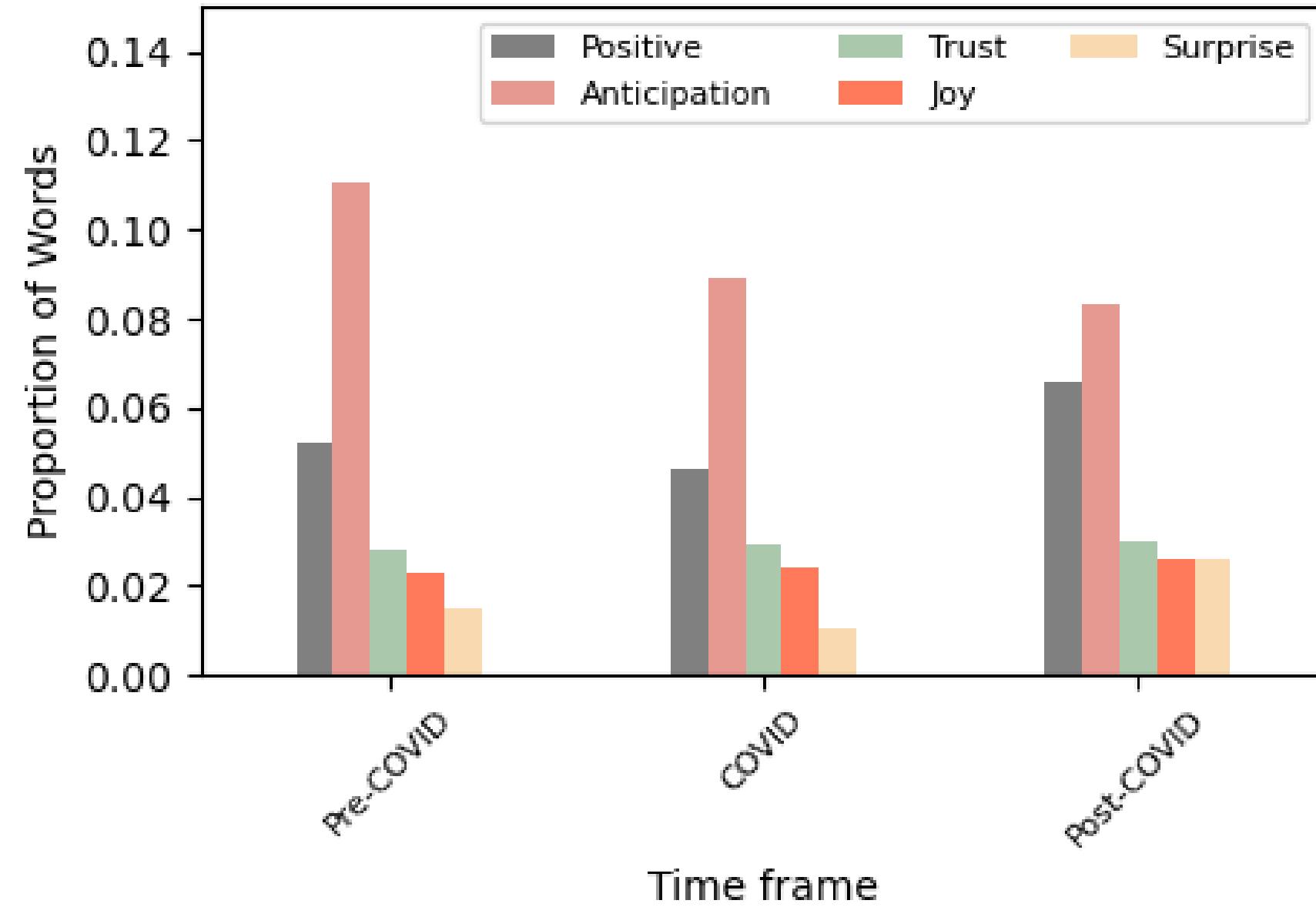
Positive Sentiments over the Years



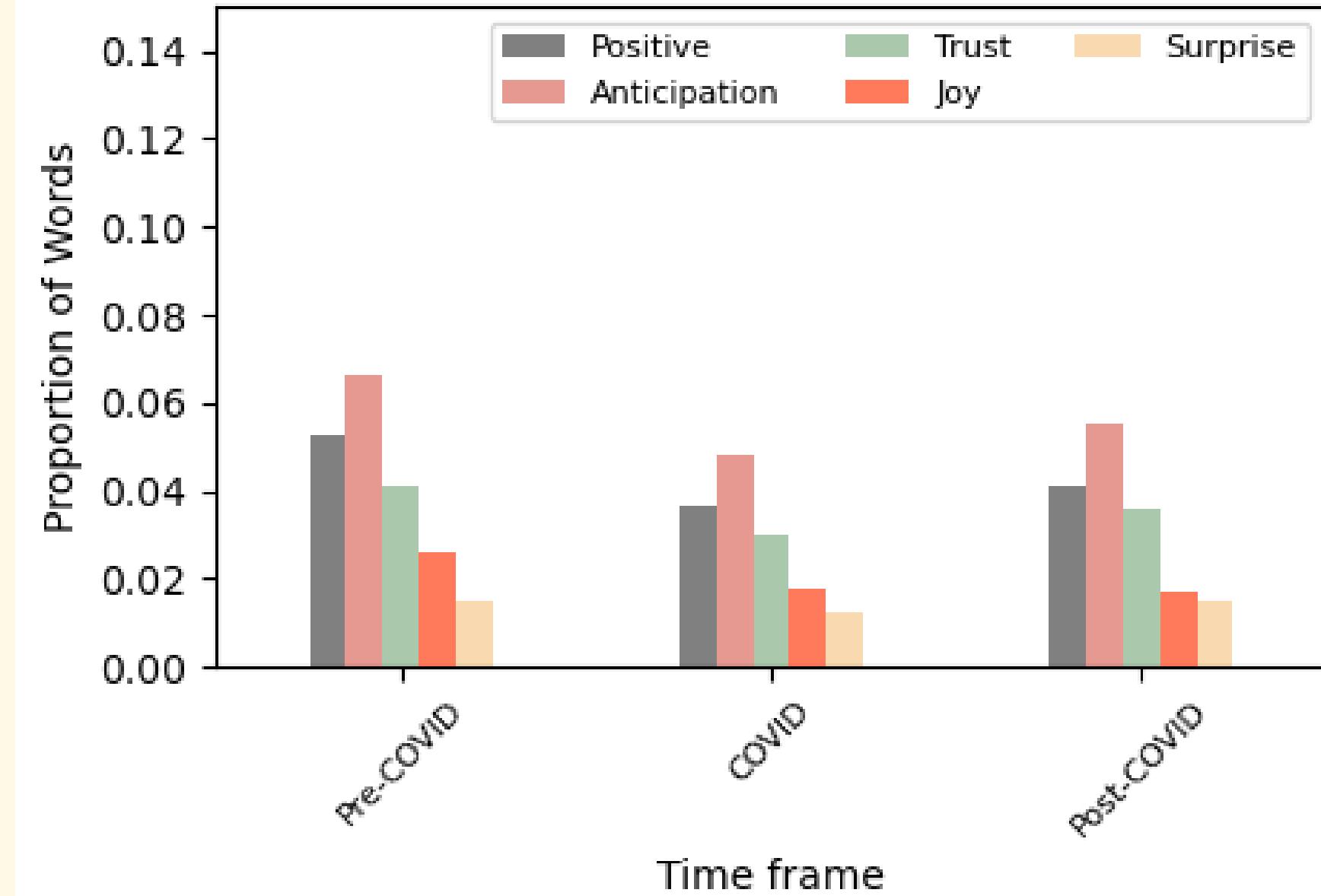
Proportion of Negative vs Positive Words in Titles by Source



NYT Positive Sentiments by Time Frame

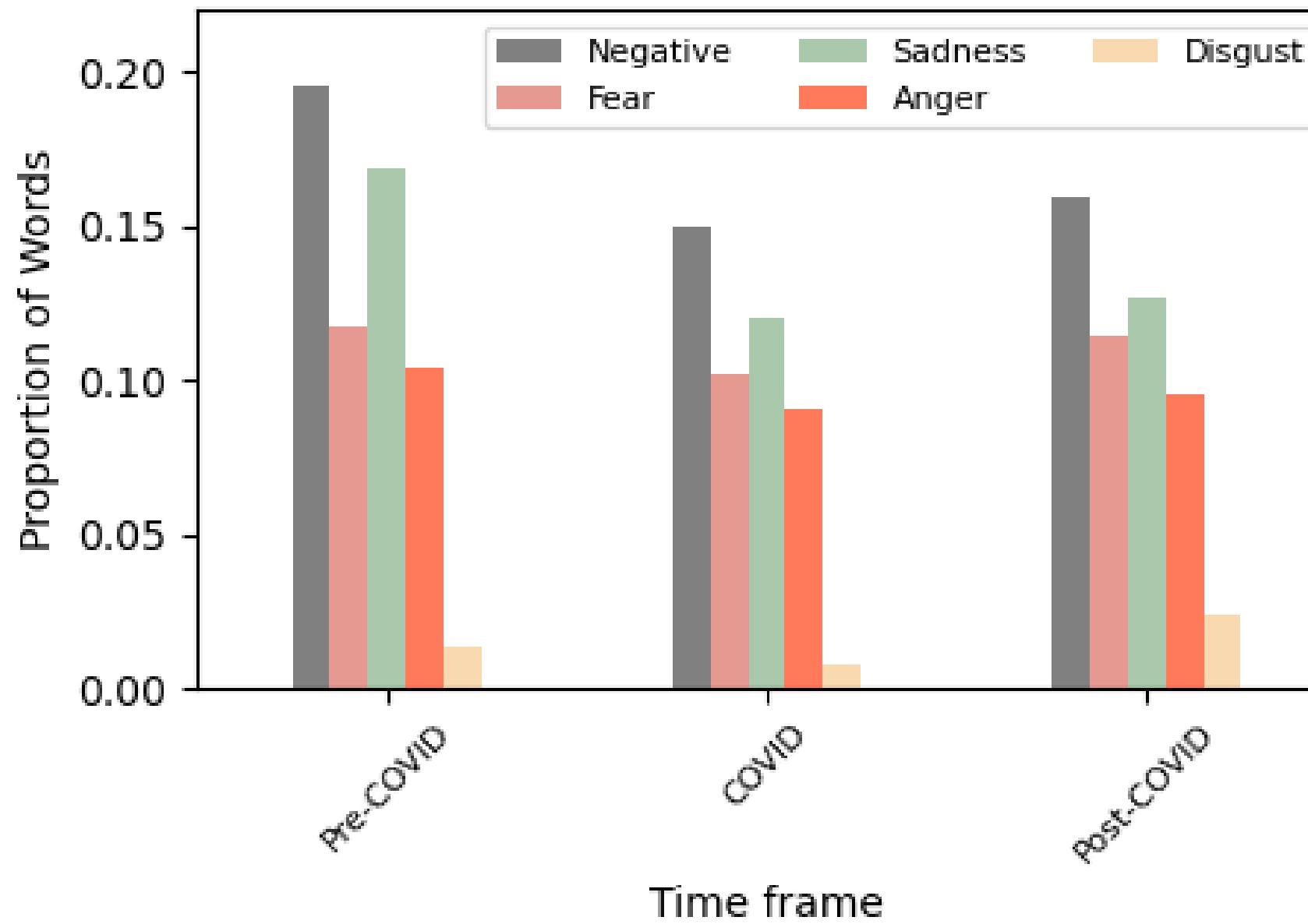


Guardian Positive Sentiments by Time Frame

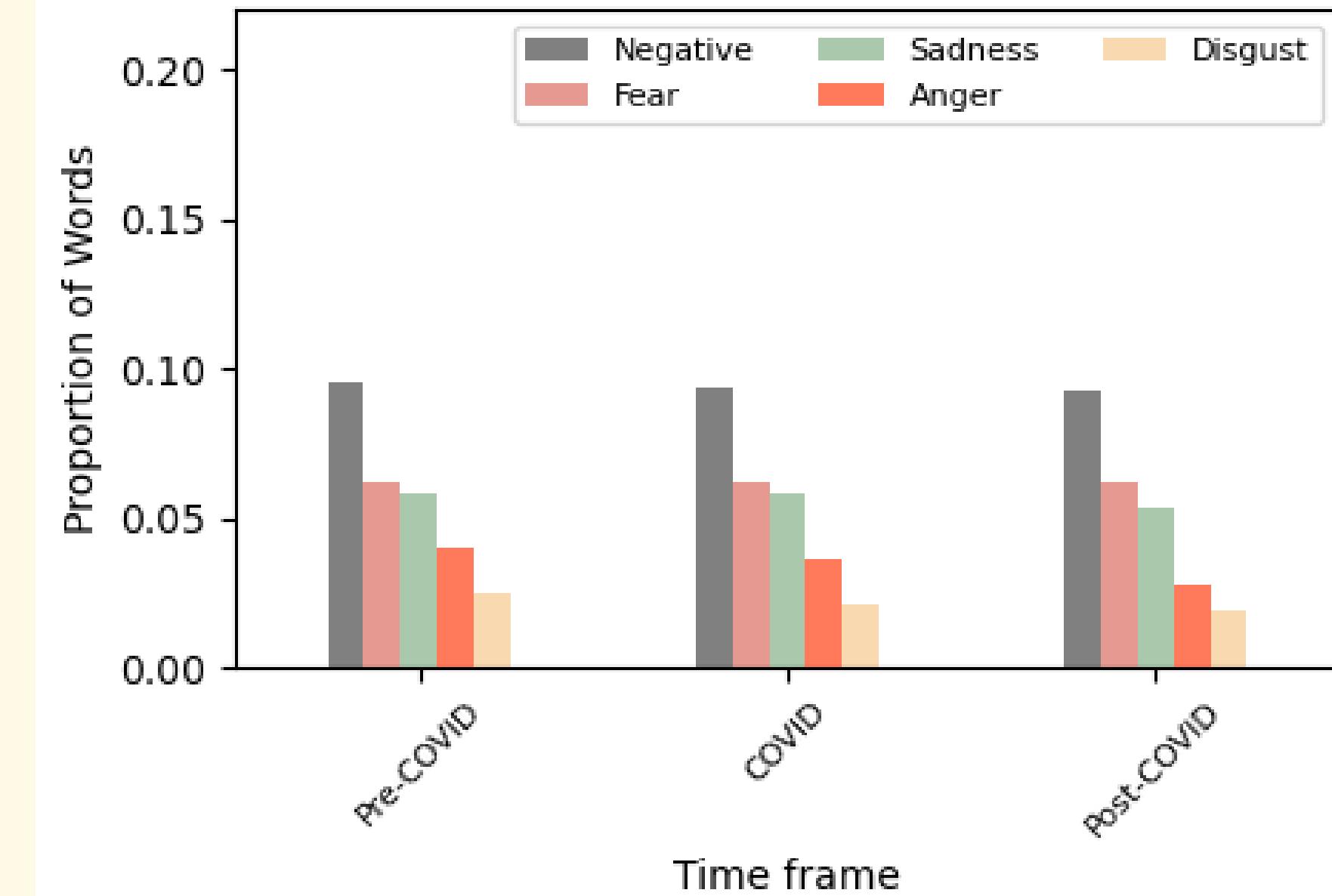


Generally consistent throughout different time periods.
Anticipation is the most popular emotion across both newspapers.

NYT Negative Sentiments by Time Frame

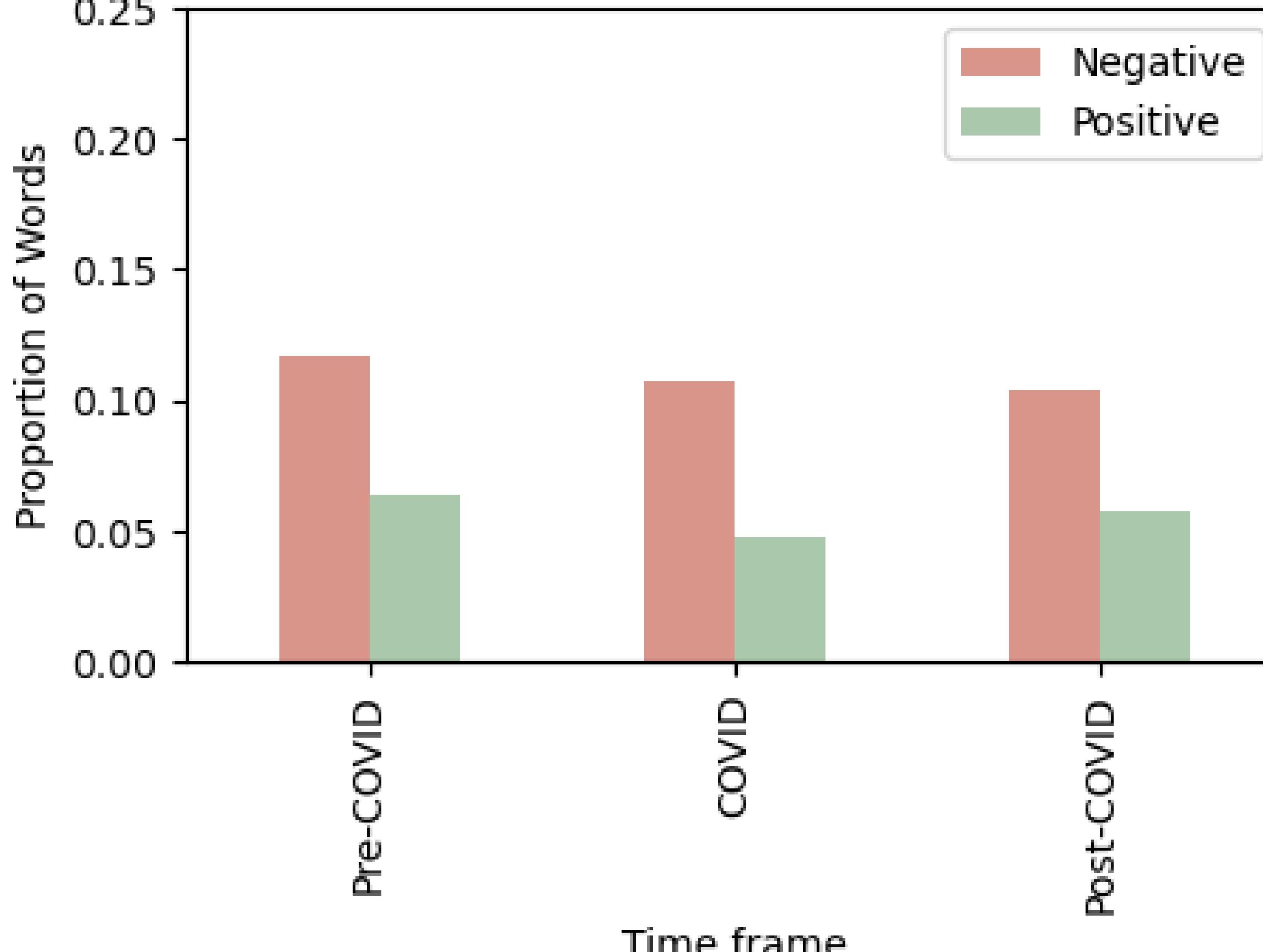


Guardian Negative Sentiments by Time Frame



Increases disgust and decrease in sadness in NYT. Sadness is most used in NYT and fear is most used in Guardian.

Proportion of Negative vs Positive Words by Time Frame

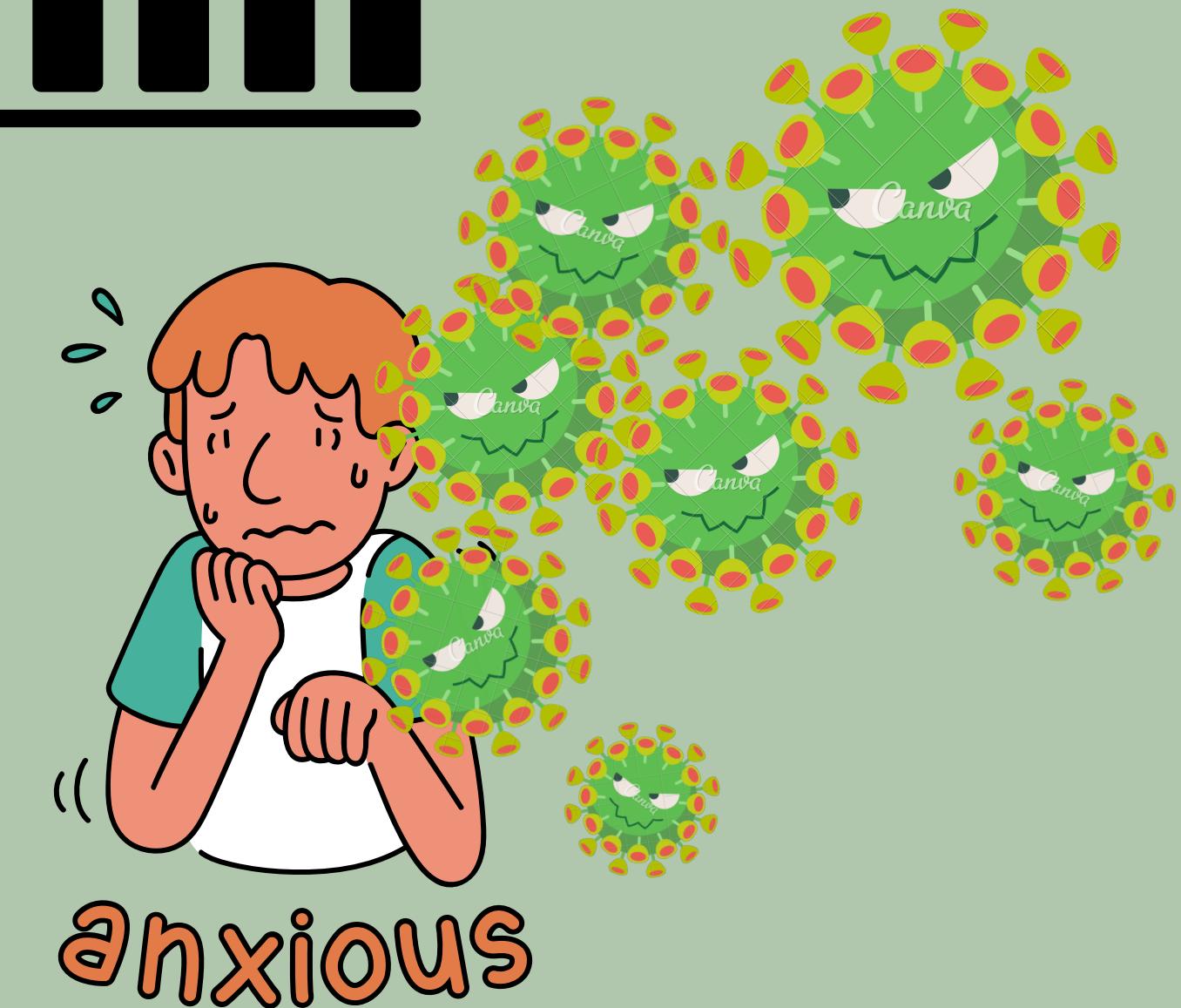


TAKE AWAYS



SUMMARY OF TRENDS

- Overall an increasing media trend, particularly during the COVID-19 pandemic.
- Our search keywords effectively captured the evolving discourse on mental health, with anxiety receiving more attention than depression during the initial phases of the pandemic, potentially because of Covid triggered anxiety more.



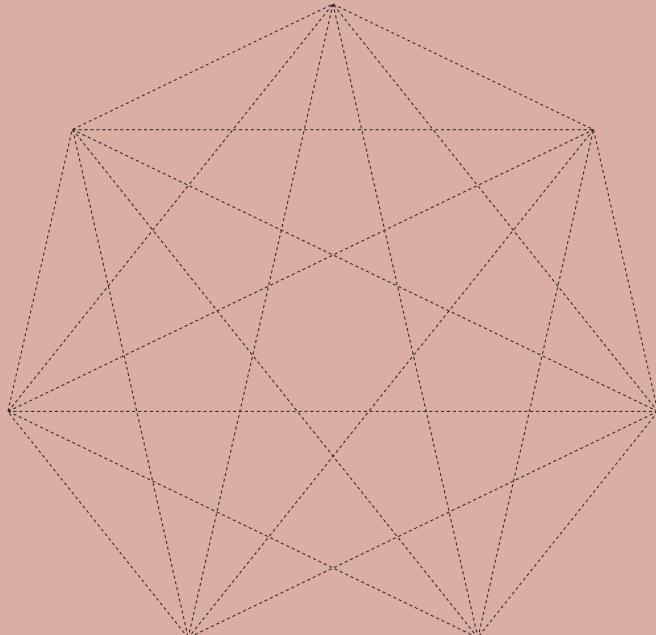
SENTIMENTS

- Mixed results on changes in negative or positive sentiment over time
 - AFINN implies increase in negativity
 - NCR implies decrease in both negativity and positivity
 - Possible increase in neutrality
- Differences between sources
 - NYT uses more negative language (both with AFINN and NCR)
 - NYT also uses more emotive language overall (positive and negative)



IMPLICATIONS

- Different resources --> different approaches/perspectives on mental health
- Mental health discussion is becoming more relevant and less stigmatized (more neutral but not more positive)



FUTURE DIRECTIONS

- Other platforms (Reddit, etc.)
 - Social media vs the news?
- Other keywords (loneliness, etc.)
- Other sentiment dictionaries

LESSONS

- Data collection & cleaning --> 80% of the time
- Great need for attention to detail (cleaning)
- Coding projects in group:
 - Importance of communication



THANK
you