

Project Lunacy, Mental Health, and Well-Being:

Visualizing the change in online media's discussion of mental health



A thesis submitted in part fulfilment of the degree of
BSc (Hons) in Computer Science with Data Science

Patrick Löwe 16725829
Patrick.lowe@ucdconnect.ie

Mark Matthews
Mark.Matthews@ucd.ie

Table of Contents

Section Number	Title	Page Number
1	Abstract	4
2	Project Specification	5
3	Introduction	6
4	Related Work	7
5	Data Considerations	10
6	An Outline of my Approach	11
7	Project Work Plan	13
7.1	▶ Outline	13
7.2	▶ Risk Mitigation Plan	15
	→ Data Feasibility	
	→ Technical Feasibility	
	→ Evaluation Feasibility	
8	Data & Context	16
8.1	▶ Data Pre-processing	16
8.2	▶ Sentiment Analysis	18
8.3	▶ Topic Modelling	20
9	Core Contributions	23
10	Evaluation	30
11	Summary & Conclusions	31
12	Acknowledgements	32
13	References	32

List of Figures

Figure Number	Title	Page Number
1	Biopsychosocial Venn Diagram	1
2	Presidential Election, 2016	9
3	NYT API Limits	10
4	FYP Gantt Chart	14
5	Data-Process Flow Chart	17
6	Excel Data Structure	18
7	Lemmatizing Examples	19
8	Average Sentiment Score	20
9	Threshold Establishment	21
10	Topics with their top 5 associated keywords	23
11	Articles with their Topic Distribution	24
12	Count of mental health articles VS all other articles	24
13	Total mental health articles, monthly	25
14	Sentiment score, per topic	26
15	Gun Control debate article	27
16	Mass shootings infographic	27
17	Total compound score for guns	27
18	Average compound score for guns	28
19	Article title on prison suicide attempt	28
20	Average compound score for Prison	28
21	Mean compound score for Healthcare, Research, Guns, Prison	29
22	Average compound score overall	29
23	Sentiment using different keywords	30
24	Sentiment by month, since COVID-19	31

Abbreviations

LDA	<i>Latent Dirichlet Allocation</i>
CTM	<i>Correlated Topic Model</i>
PLSI	<i>Probabilistic Latent Semantic Indexing</i>
NYT	<i>The New York Times</i>
<i>IT</i>	<i>The Irish Times</i>
API	<i>Application Programming Interface</i>
AWS	<i>Amazon Web Services</i>

1. Abstract

Mental health has been shown to be a core element of wellbeing. However, the 21st century is placing unprecedented demands on our mental health. It's believed that a significant proportion of the population will require some form of mental health care (Aoun, 2004). Unfortunately, negative social discourse can generate stigma which can then reduce help-seeking behaviours (Mak, 2007). This research aims to explore the role that media plays in shaping this stigmatised social discourse.

This project therefore aims to examine any changes in sentiment towards mental health within online media over the past decade. It achieves this through three key objectives:

1. To retrieve a large and credible data source of online media relating to mental health.
2. To classify the articles into their top 'n' topics using an appropriate topic modelling algorithm
3. To identify any shift or trend in sentiment bias, both positive and negative, towards mental health using appropriate sentiment analysis tools, then graph the results using appropriate visualization methods.

Specifically, this project focuses on investigating how the sentiment related to the term 'mental health' have changed over a period of time in one online media website (e.g. The New York Times). More precisely, this work focuses on analysing how the associations related to the term 'Mental Health' have changed over a period of time (e.g. 10 years) and what related insights can be derived about the current societal discourse related to mental health and its future trajectory as it relates to positive or negative sentiment.

It achieves this by analysing the sentiment associated with mental health in the core topics identified within the corpus. Initial data was collected using the New York Times Article Search API covering a minimum of a 10-year period. Topic modelling was then used to categorize each article as a topic depending on recurring words within each article. Topic modelling is a method of representing a corpus as a mixture of themes based on a probability distribution over the corpus (Blei et al., 2003). After a series of topics were identified, sentiment analysis was used to examine in depth articles that portray mental health in a positive, negative, or neutral (mixture of positive and negative without either being prevalent) outlook.

The core challenge here was identifying an appropriate number of topics, determining the number of passes since it is a manual process, and then how to accurately analyse the sentiment attached to mental health. To overcome these challenges, Latent Dirichlet Allocation (LDA) was used as there were a concise number of topics expected to be associated with mental health, and a large set of topics in combination with LDA could return incoherent/uninterpretable topics since it is based off the popular words across multiple articles.

Based on previous work (Rentier, 2018), the ideal number of topics are between 10 to 15 as anything larger would create incoherent topics too niche/subset of core topics (e.g. instead of Arts it might suggest books, or films, or home design etc), and a smaller set of topics might not create enough popular words to differentiate the topics (e.g. core words surrounding one business could be interpreted as automotive, instead of business trading).

This research illustrates any changes towards the sentiment of mental health using an interactive visualisation within Jupyter Notebooks. This tool will allow users to select a topic and view Positive/Negative/Compound scores across the time period. It also can highlight areas of interest which are then discussed.

2. Project Specification

This project researches and summarizes the way media constructs mental health, how it continues to change over time, as well as considering its future trajectory. It uses the most relevant applications for topic modelling of newspaper articles, the different hyperparameters of the algorithms that can be used, appropriate sentiment analysis measures and tries to overcome the common challenges faced with sentiment analysis (also known as Opinion Mining). Finally, this project demonstrates how we can effectively visualize how sentiment in relation to mental health changes over time using informative visualization.

The end product is a report on the change in sentiment associated with mental health, on a topic level that visualizes how the key term 'Mental Health' is received by target audiences, as communicated in one large news website (The New York Times) and how it has changed over the course of a 10 year period. This gives insight into which sectors are quicker to adopt a positive view, such as arts and film compared with sports. Since the sentiment analysis is being scored on a large amount of text, manual reading is not viable. Therefore, programming is used. The definition of positive and negative scoring must be clarified and therefore manually reading a number of articles to clarify what is positive/negative.

3. Introduction

Social discourse on mental health over the past century has changed dramatically. As a society we have moved from seeing mental health issues as 'lunacy', requiring removal from society, to an experience which in many cases needs medical intervention. The past few decades have however seen this discourse within professional arenas further refined with mental health being understood as a biopsychosocial phenomenon which requires a biopsychosocial response. (Papadimitriou, 2018)

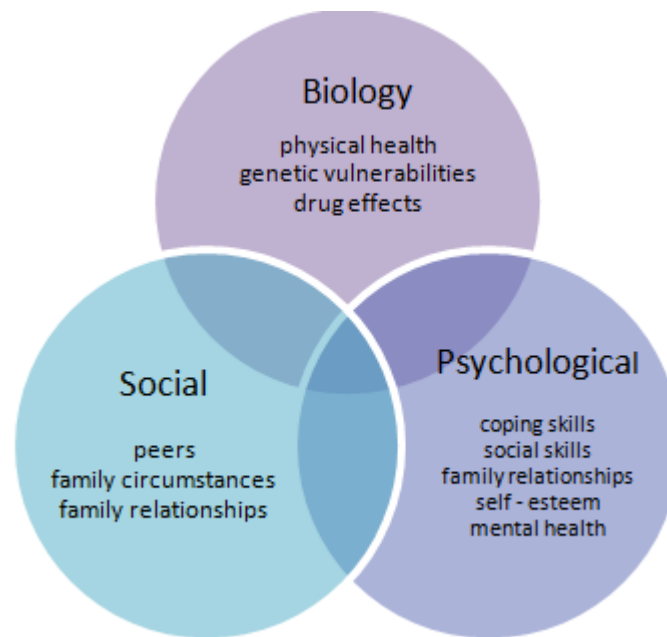


Figure 1: Biopsychosocial Venn diagram

Introducing this more systemic view of mental health into social discourse has been slow (Ohlsson, 2018) and stigma still is a challenge which reduces help-seeking behaviours. Examining how the media plays a role in shaping and reflecting the social discourse, and therefore sentiment on mental health, can play an important role. An examination such as this also contributes to the discourse directly, by making the shift towards this way of conceptualising mental health more explicit.

This research investigates how online media discusses mental health, whether positively, negatively, or neutrally (a mixture). This will help make explicit whether stigma (negative bias) is still being sustained as part of the social discourse in this context.

It has been shown that internalised stigma due to social discourse impacts negatively on help-seeking behaviour (Henderson, 2013). For example, when people feel that they can openly express themselves without fear of judgment then they are more likely to seek help with their mental wellbeing.

In Ireland, a significant number of LGBT adolescents (over 55% of 14 to 18-year old's) disclosed self-harming (BeLonGTo, 2018). The common reasoning they offered was because they felt like they could not talk to their friends or family (with fears including that of being judged, bullying, being isolated, etc.).

This is just one small snapshot of mental health challenges. As a global society, with many contexts affecting mental health, stigma is seen as the most harmful block to wellbeing. The converse is also true whereby if we can take responsibility for a more positive discourse on mental health, we promote resilience.

A significant contributor to this social discourse lies with the media's representation of mental health. Traditionally, the business of media relied heavily on two key elements of human experience – curiosity and negative bias

(stigma). In addition, their power to influence social discourse is well documented. For example, the theory that the media does not just tell us the news but tells us what to think about the news. (McCombs & Reynolds, 2002) (Gansinger, 2019)

Therefore, understanding media sentiment towards mental health, both past and present, and to showcase its future potential trajectory is important. This project aims to use topic modelling techniques and sentiment analysis tools to explore how the discourse on mental health has changed in one popular online media source in North America.

4. Related Work and Ideas

The challenges with this project were in collecting valuable and interpretable data from an online media source. Not all articles returned using a keyword will have it as the focus of the article. Accurately interpreting the top keywords associated with each topic which allows a clear and coherent topic name, and finally how to accurately assess the sentiment analysis of the corpus as different methods of pre-processing article text can alter the sentiment scoring.

The first challenge to this project was resolved by The New York Times API search engine. Previous research papers, (Jonathan Chang, 2009) (Smith, 2016) , have used this as a source of data as it allowed them to gather insight attributes in which to build their project (author name, news desk source, time/date, keyword searches, etc.). It has been commonly used by many projects. The limitation to this is the requirement for a paid license to access the full article, without this only the headline and lead paragraph is accessible. This does not satisfy the needs for sentiment analysis as the story can take a turn throughout the article, performing a quick sentiment analysis on headlines and lead paragraphs lead to very different results when the main article is then interpreted.

Then the appropriate topic modelling technique in relation to the research aim was considered. From a paper looking into humans' ability to interpret topic modelling (Jonathan Chang, 2009), the report's findings are that the LDA performed closer to humans perception of associated topics when using word intrusion, that is the ability for a model to apply semantically cohesive topics to an article.

The CTM did however give better predictive results when using held-out likelihood, the ability for the training set results and the unseen results (for testing) was closely linked. It outperformed the results from human testing, which hints at it being overly fit. Then pLSI became less efficient as topic numbers increased, implying that overfitting was affecting interpretability. Since LDA is an approximation algorithm it can produce slightly different results each initialization, meaning each topic should return the same top N words when run. (Baumer, 2017)

For sentiment analysis there have been mixed opinions on what tools to use and why., (Jongeling et al., 2015) found that certain tools work better for different domains of data. Due to their results on NLTK it suffices as an appropriate tool for

newspaper domain-based analysis. Their research was carried out on its ability to achieve accurate sentiment on software engineering domain-based data. The pre-processed data for topic modelling cannot be used since there was too much removal of punctuation, lemmatization, stop words etc. As such, the original data was used and reprocessed towards sentiment analysis. The goal was to group newly labelled data by publishing month, break down the sentiment of each article by sentence, then gather a total and store it for each article. The result was to visualize sentiment on a monthly basis across the 10 years, and also on a topic level basis.

The challenges with working on a large amount of data to apply topic modelling is that it was repeatedly reported at taking days on high-powered machines to classify (Rita Hamad, 2015). For this reason, this research initially limited the analysis to no lower than 10 topics. Rentier (2018) proposes that finding the number of topics is a manual process, the best approach is to create a model that looks at a large number of topics, run it across all the data multiple times so that it can perform a comparative against other articles. This in turn will result in a large amount of associated words per topic. The idea is to reduce the number of topics and rerun until a solid list of descriptive, associated words with each topic is found. When the number of categories, or in this case topics, are limited then the algorithm has a limited ability to detect variation (Rita Hamad, 2015). Then these topics must be applied against the data to see how well they fit the articles.

Finally, to be able to quickly construe the results the correct types of visualizations are to be considered. Almost every blog post, research paper, or study shows a timeline of sentiment analysis, some with a secondary attribute (stock prices, presidential candidates, tweet topic) also displayed. For this project, a line graph monitoring the sentiment versus month, year by year, is presented. There are also multiple lines, each for the 10-20 topics, depending on the manual findings later in the project. An example of this presentation is below.

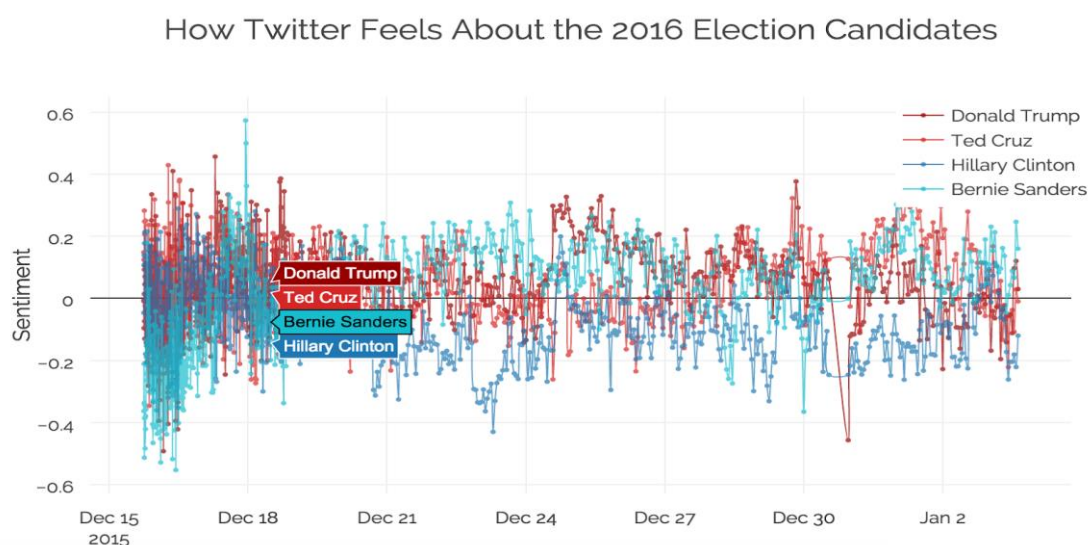


Figure 2. (Bohan, 2016) Sentiment Analysis of presidential candidates

An interactive graph showing the sentiment change throughout time would have easily shown any changes. However, given that the final report is printed hardcopy this interactivity will be a secondary consideration. Other graphs include a breakdown of months comparing each month with itself to previous years. This demonstrates the changes from the initial period to the current period, but also highlights that mental health appears to become a topic of discussion during certain months, for example during mental health awareness week in October, seasonal depression, etc. The best tool identified so far has been Gapminder, an opensource tool used in largely recognised TedTalk presentation by Hans Rosling (Gapminder, n.d.) but has been tricky to use.

5. Data Considerations

The data collected for this project relates to articles from The New York Times. A more advanced approach to this research was considered but deemed unfeasible. This was to examine a second news source, such as The Irish Times, and perform a comparative between the two sources. After contacting The Irish Times (IT) and UCD Library it was not feasible. The IT does not have an API for collecting the data, and the Nexus (through UCD Library) does have the data but downloading en masse is not possible, only the metadata. This meant that topic modelling and sentiment analysis was not possible. This could have led to interesting insights as the newspapers are operating in different countries and could have shown cultural differences or perhaps highlighted different journalistic conventions between countries. Another approach considered but rejected due to lack of processing power for the researcher would have been to include other key terms that relate to mental health i.e. depression, circadian rhythms, schizophrenia, depression, etc. These articles could also be analysed separately, sentiment could then be compared but the topics encountered will vary depending on the available data with each set of results for each keyword. It would be possible to combine all datasets and perform topic modelling thus providing greater depth to our understanding of the societal discourse around mental health.

A few possibilities were considered for determining if an article focused on mental health or was pulled for mentioning the keyword once. First, analysing the article and focusing on the sentence that contained 'mental health'. However, an unknown amount of context (therefore additional sentiment) could be lost. Next, was to analyse the paragraph it occurred in. This could be done using the paragraph tags (<p>) in the HTML data parsed, but again sentiment could be lost. Instead, the number of times the keyword appears in the article was counted and a threshold found to cut off articles. This allows appropriate sentiment to be gathered with little skewness in data due to unrelated articles.

The articles being returned by the API are those containing the keyword 'Mental Health' therefore based on previous work (Rentier, 2018) 800 to 1,000 articles per year are expected. Performing a quick check on retrieving article ID and publish date for 2019 returned 836 articles. These articles are all historical, but as reported by Laban, (2016) they may have duplicate entries if they must make

minor corrections. With this in mind, the data pre-processing involved a search for duplicate articles and only the most up-to-date version retained.

There are limitations to the data being pulled, only 4,000 articles can be requested a day, or 10 a minute. There is also a results limit of 1,000 per request. In order to bypass these limits, the data requests were as follows:

Goal: ~10,000	Requests / min	Daily Limit	Result (for expected articles)
NYT	10	10,000	1 Day
The Guardian	5	5,000	2 Days

Figure 3. NYT API limits

There will be 10 requests made, one for each in the time period. Based on the previous work by (Rentier, n.d.) It is expected to return roughly 700 articles per year. Due to inconsistent HTML structure within some articles (blogs, opinion, etc) these articles were removed, resulting in 369 articles in 2010 to 716 to 2019.

In terms of data licensing, all the data cannot be used for commercial use. Educational use is permitted, but since this research uses a paid license to access the full article story, that section is not permitted to be shared. Therefore, the findings and analysis are appropriate to share online but not the data itself.

The core challenge here is identifying an appropriate number of topics, determining the number of passes since it is a manual process, and then how to accurately analyse the sentiment attached to mental health. To overcome these challenges, Latent Dirichlet Allocation (LDA) was used as there were a concise number of topics expected to be associated with mental health, and a large set of topics in combination with LDA could return incoherent/uninterpretable topics since it is based off the popular words across multiple articles. Based on previous work (Rentier, 2018), the ideal number of topics are between 10 to 15 as anything larger would create incoherent topics too niche/subset of core topics (e.g. instead of Arts it might suggest books, or films, or home design etc), and a smaller set of topics might not create enough popular words to differentiate the topics (e.g. core words surrounding one business could be interpreted as automotive, instead of business trading).

In terms of sentiment analysis challenges, pre-processing was considered. This involved removing named entities, such as cities or people. Often, this has little impact on the overall sentiment attached to an article and removing it helped narrow down the associated words to create a topic. For topic modelling, tokenization involved breaking down articles into a 'bag-of-words', this helped speed up applying a topic.

The data also had punctuation removed as it bears no impact on topic modelling (but may impact sentiment analysis), the data was converted to lowercase, and a

list of stop words was also removed (then, are, we, you) as it would only affect sentiment but not topic modelling. (Eric Baumer, 2015).

Lemmatization within the NLTK package in python was also used. This involves changing a word to its core, for example 'Studied' or 'studies' will become 'studying'. Several papers have reported that stemming, another method similar to lemmatization, can work effectively on their data but also report of words being shortened by mistake/misinterpretation, for example 'bushfires' gets shortened to 'bushfir' removing the -es. While this can be workable in most cases if it shortens words incorrectly and combines them with other shortened words the stem of that word may be identical, but the unprocessed words can have 2 distinct meanings. Stanford NLP was tested as a comparison to NLTK. However, NLTK is more suited towards beginners and is python based, whereas Stanford NLP is java based and more complex to implement within the constraints of this research.

6 An Outline of My Approach

The python package 'YanyTapi' was used to parse the JSON format of the returned data from The New York Times API into an easily readable and usable format within python. This enabled easily transformable data into a data frame for the next steps. It is specifically built for the NYT API. The data was then exported and stored in a CSV file for further use. With considerations for request limits, the data was appended on publishing month for all 10 years, then duplicates removed, based on their article ID and matching count. Any matching articles with a different word count were removed if it is not the most recent publication of that article.

For data pre-processing the 'NLTK' python package will be used. This will allow for quick, and easy processing for tokenization, lemmatization, breaking articles into sentences for further analysis, punctuation, etc. Although Gensim can also be used for this step, it has been reported as running slowly on larger dataset for pre-processing, whereas NLTK runs faster.

The python package 'Gensim' has been frequently quoted for applying Topic Modelling. (Jonathan Chang, 2009) showed that an LDA approach worked the best in terms of human interpretation vs machine interpretation. Since the ideal number of topics is manual, the original model will start with 20 topics and work down towards 10, any lower has shown a problem with variation (Rita Hamad, 2015). The model parameters for passes will be set to 20 initially and increasing as the number of topics reach 10. This allows for the model to run quicker at the start as it isn't evaluating every iteration. Alpha will be set to 0, allowing it to learn asymmetric priori from the data, this has been reported as fitting the data more naturally when topics are lower (below 50) (Rehurek, 2013).

The results from the topic model at each run will be a list of:

- a) topic number
- b) the top words associated with that topic, and
- c) the weight of that word within that topic.

This means that to have a name or title on the topic it will need to be manually/humanly interpreted. Each run of the model may yield different associated words and weightings. For this reason, the number of passes increases as the number of topics decreases. This allows for more meaningful words to be perceived from the topic.

Gensim also comes with a built-in function for model evaluation. This allows easy calculations on the models Perplexity and Coherence. Perplexity is an evaluation of the model's ability to predict a sample, this is usually only beneficial when comparing models as a lower perplexity score is better. This measure will be considered when lowering the number of topics. Coherence is a more accurate score as it measures how semantically meaningful it is based off of the other high scoring words, the higher coherence scores the better.

For graphing these results, there are several options. Considering interactive graphs would be easier to gain an understanding, Tableau is considered. However, this is a paid service and while there is a free account for students in approved academic centres, the license may expire mid-year 2020 as it was used for a previous module. SAS Visual Analytics was also considered but does not seem to be a part of the academic bundle. The main tool falls to Gapminder, an open-source software used for interactive visualizations. The backup, if unseen circumstances prevent previously mentioned tools from being used, is plotting graphs within python's matplotlib package, while not aesthetically pleasing it appears that a lot of visualization tools are paid services, with the trial/free level accounts being too limited to achieve what is necessary.

7. Project Work Plan

7.1 Outline

For semester 2 the project will entail a) gathering the data, b) pre-processing the data, c) apply topic modelling, d) run sentiment analysis and finally e) create a visual report and discussion on the findings. These plans follow on after another and rely heavily on each being completed. With the limitations to data being pulled through NYT API, some preprocessing can begin in a gated process, but topic modelling requires all data be collected and pre-processed.

Sentiment analysis requires that the data be mildly pre-processing in comparison to the level of pre-processing needed for topic modelling. Therefore, sentiment analysis and topic modelling could be run in parallel, depending on computing demands and availability of software. Research papers have used online resources such as Amazon Warehouse to process large volumes of data but that is a paid service. The goal is to use a local machine for sentiment analysis, as this process can be batched for monthly articles. However, topic modelling will require a much larger processing power as all the data must be considered. (Rentier, 2018) mentions that the run time on a high-powered AWS took 8 hours. This was due to running 12-16 topics on 200 passes. The failsafe here is that this work has already been done, so choosing 14 topics but on a lower set of passes (50) could greatly increase the processing ability to run on a local machine. If this is unachievable, there is the student version of AWS that comes with \$75 credit on the account.

Using Amazon Comprehend as a backup, they have 2 tools of use. Topic Modelling and Sentiment Analysis. The pricing for this is \$1 per job and first 100MB and then \$0.004 per MB after that. Pulling 1,000 test articles on mental health for 2019, I then duplicated this to bring up to the estimated 30,000 mark. This created a CSV ~ 7MB in size. However, this does not include the full article text, so the real file is expected to be 4-5 times larger. If the file is expected to be ~35MB this still falls under the flat rate of \$1. This appears very feasible using the \$75 credit on the account. This is however dependent on what is used for the remainder of our Data Mining module, the reason the student account was set up.

FYP GANTT

TASK NAME	START DATE	DAY OF MONTH*	END DATE	DURATION* (WORK DAYS)	DAYS COMPLETE*	DAYS REMAINING*
New York Times API						
Pull 2007 -> 2010	2/2	2	2/4	2	0	2
Pull 2011 -> 2014	2/3	3	2/5	2	0	2
Pull 2015 -> 2018	2/4	4	2/6	2	0	2
Amalgamate Results	2/5	5	2/7	2	0	2
Pre-Processing						
Sentiment Analysis 2007 -> 2012	2/10	10	2/14	4	0	4
Sentiment Analysis 2013 -> 2018	2/12	12	2/16	4	0	4
Topic Modelling 2007 -> 2012	2/14	14	2/18	4	0	4
Topic Modelling 2013 -> 2018	2/16	16	2/20	4	0	4
Sentiment Analysis						
Article Level 2007 -> 2009	2/24	24	2/27	3	0	3
Article Level 2010 -> 2012	2/27	27	3/1	3	0	3
Article Level 2013 -> 2016	3/1	1	3/4	3	0	3
Article Level 2017 -> 2018	3/4	4	3/7	3	0	3
Combine to Monthly Level	3/7	7	3/10	3	0	3
Topic Modelling						
10 Topics	3/16	16	3/19	3	0	3
12 Topics	3/19	19	3/23	4	0	4
14 Topics	3/22	22	3/27	5	0	5
16 Topics	3/25	25	3/31	6	0	6
Comparison of above	4/1	1	4/3	2	0	2
Graphing Results						
Topic Level Sentiment	4/3	3	4/8	5	0	5
Monthly Sentiment	4/8	8	4/13	5	0	5
Trajectory	4/13	13	4/16	3	0	3
Report						
Draft	4/17	17	4/18	1	0	1
Finalise	4/18	18	4/20	2	0	2

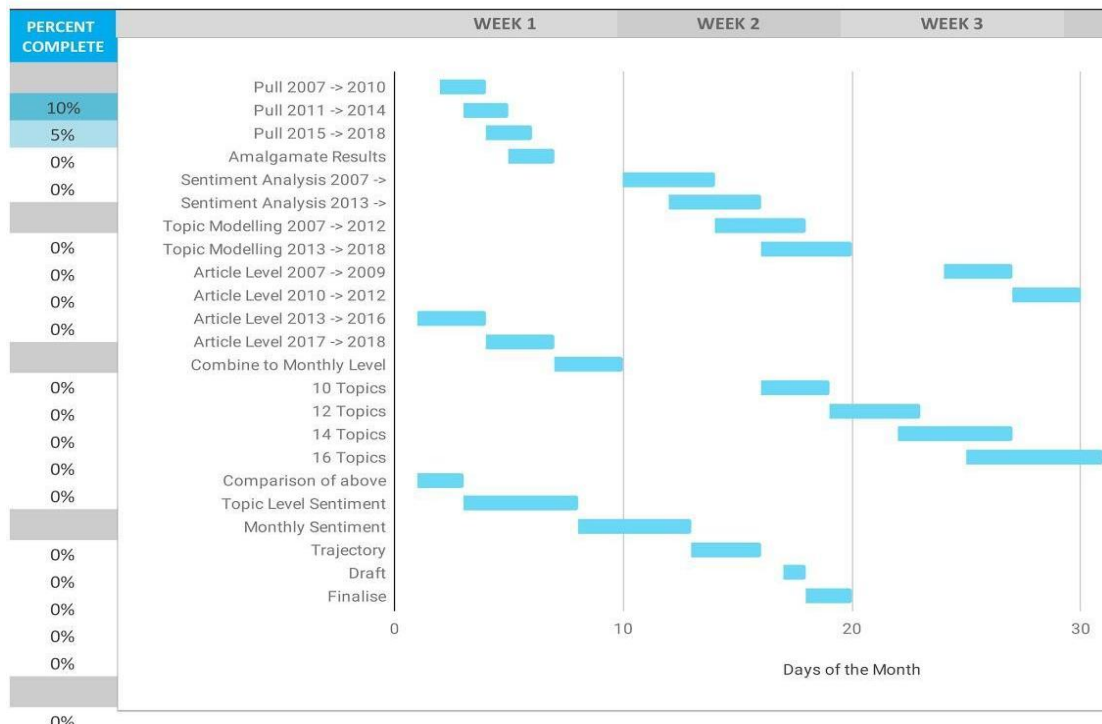


Figure 4. Gantt of project breakdown

Once the results are obtained the best comparison would be with the previous work carried out by (Rentier, 2018). Sentiment analysis can then be broken down to an article level, and furthermore to a sentence level building up to a total for the article. This allows for sentiment to be scaled up and analysed on a monthly level or topic level, both over the time period.

7.2 Risk Mitigation Plan

Data Feasibility - The data from the NYT API is all openly sourced and usable for non-commercial purposes. The downside is that the article content is not allowed to be publicly shared, and as such I will only be permitted to share the results but not the source data. For the advanced approach, there appears to be a limited number of newspapers with API systems implemented. For the data comparison the Irish Times unfortunately does not appear to have an API gateway, as such I have gained access to Die Zeit, a German newspaper but their articles will be in Deutsch, this is still feasible for topic modelling (although special characters may need to be converted, e.g. ß can be converted to 'ss' or ö can become 'o', the trouble occurs with sentiment analysis. This can be resolved using the Google Cloud Natural Language API as it supports German. While the tools identified for English (NLTK, Stanford NLP) have some ability to process German it does not have all the same abilities and therefore Google would be used instead. The alternative to Die Zeit is the Guardian, an English newspaper. A license costs £7.49 a month and would allow access to the same level of detail as the NYT.

Technical Feasibility - Almost every tool being used for this project is open-source/free. The limitations may occur if the back-up options are to be used (i.e. AWS Comprehend). Based on previous calculations, the amount of data to be processed seems feasible on the student credit attached with a new account (\$75). The data is expected to be ~10,000 articles and will fit within the CSV file type limitations. The risk is with the 1,000 article limits per request, this means that the results will either have to be appended to the file or saved separately and all files then appended after. Both options are feasible but pose risk of crashing if transferring large amounts at once. For this I propose a safe approach by appending one file/1,000 request at a time.

Evaluation Feasibility - I believe the best method for evaluation is in comparison to the article by (Rentier, 2018). This is a great comparison for evaluation for topic modelling only as she has not yet performed sentiment analysis. For that, it has been identified as tricky to evaluate and I therefore believe that choosing N articles at random to evaluate by hand and compare to the sentiment applied by the tool is the best measure of evaluation for sentiment analysis.

Ethical Feasibility - There is a need for this analysis to be performed on online media which is freely available to all, with some limitations to a paid service for additional pay-wall articles, but the article text will not be shared publicly after analysing. There are also no privacy concerns in using the data as any names/places cannot be used for topic modelling or sentiment analysis, and as such entity removal will be used to remove these before processing.

8. Data and Context

8.1 Data Pre-processing

This chapter will explain how the data was collected, refined, and processed, and give an understanding of the context, limitations, and points of improvement of the data.

Starting with python code to retrieve the articles through the New York Times API, Yanytapi package was used. All metadata was able to be pulled but the full article text was not. The NYT was contacted about retrieving this but required a specific licensing I was not able to get. Instead, I subscribed to a student edition and used the 'URLLIB' package to retrieve full text from each article. All articles containing the keyword 'mental health' were retrieved one year at a time to prevent any problems on access limitations; from 1st Jan 2010 to 31st Dec 2019 inclusive.

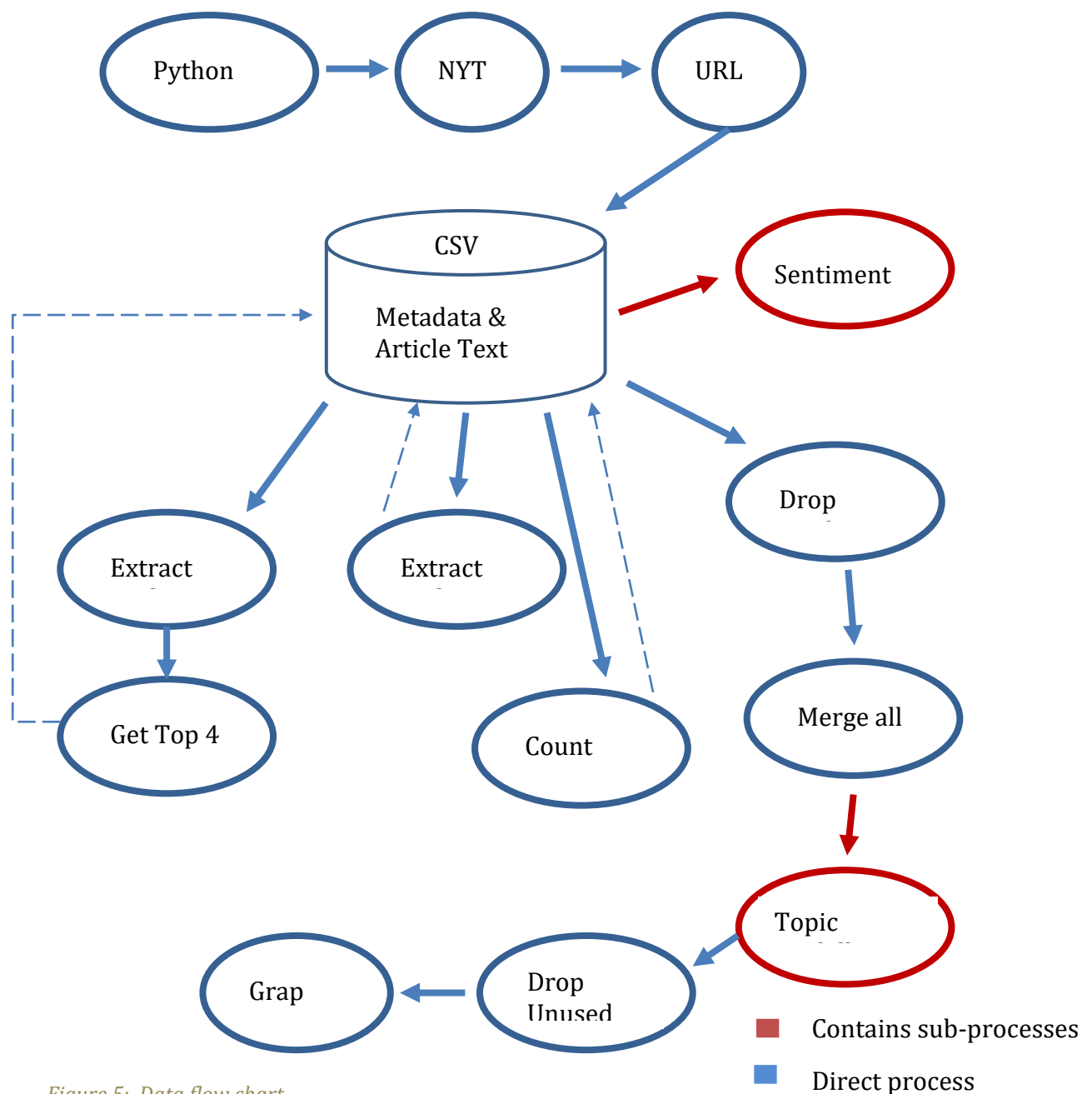


Figure 5: Data flow chart

Then each data frame was used to gather the article text using URLLib requests, this package pulls all the HTML writeup and using python the core article text was extracted. This section of code however skipped articles that contained interactive (no text, but interactive graphics), blog posts (HTML structure varied too much to get consistent text), deal-book (same issue as blog posts), finally any article defined as 'opinion' or 'article' was used but each had a different HTML structure and an if statement was used to run the appropriate code. Saving the data frames with their article text hit a CSV character limit, and so 4 columns of 'body_text' were created to store the overflow of characters. These are later combined into a single column when read back in from their CSV to python.

Using the by-line from the metadata, the authors with each article are extracted. Initially this was to be used to view sentiment towards mental health of each author and find if their sentiment has changed over time, expanded into a range of topics or key topics, and identify which authors are the leading cause of change in sentiment, if any. However not all articles have an author and articles with multiple authors is hard to decipher who is the lead author. While this data has been kept it was not used in graphing for these reasons.

The 'headline' was used to extract the kicker, an additional line of text to give quick insight into an article but these lacked consistent structure as some used keywords, some kickers identified as 'The Bay Citizen' while others used the opening paragraph, and some had none at all. The plan was to use these for later analysis, sort of like topics but identified by the author/NYT themselves. The last section of this also investigates the article text and counts every occurrence of 'mental health' as 'cases'. This is later used to filter articles that may have been pulled for containing the keyword but not relate to mental health as a whole, by using a count it is assumed that a higher count (later determined in analysis of >2) means an article is strongly likely to focus on mental health, while a low count (<2) simply mentions mental health without being a key focus.

Once all the necessary alterations were performed on the metadata and article text was gathered, the 10 CSV files are used to get sentiment for each article. At this stage they remain as 10 separate files as sentiment is performed at an article level, but later combined into one source file for topic modelling since all articles must be used in the calculations.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	id	author_1	author_2	author_3	author_4	blog	body_text	body_text	body_text	body_text	cases	compund	document	keywords	kicker	lead	para	meta
2	nyt://artic	Benedict Carey					Young children in military families are abc				4	-0.6444	article	[('name': 'None	Young chil	["hits": 43]	
3	nyt://artic	Randal C. Archibold					MEXICO C ãœThe director said that the r				4	0.9821	article	[('name': 'None	MEXICO C	["hits": 43]	
4	nyt://artic	Charles Zanol					Narcissists nicians are accustomed to think				3	0.9874	article	[('name': 'Mind	Narcissists	["hits": 43]	
5	nyt://artic	Sharon La Dan Levin					LOUHE, Ch Mental He00 renminbi, or abou				25	-0.9793	article	[('name': 'Neglect at	LOUHE, Ch	["hits": 43]	
6	nyt://artic	Sharon LaFraniere					XIZHEN, Cl, in the Soio the Hepu County P				23	-0.9985	article	[('name': 'None	XIZHEN, Cl	["hits": 43]	
7	nyt://artic	James Warren					James Warren writes a column for the Chi				3	0.5242	article	[('name': 'Chicago N	James Wa	["hits": 43]	
8	nyt://artic	Ronald D. Castille					PhiladelphiaEVERY month Americans com				2	-0.9446	article	[('name': 'OpEd Cor	Philadelph	["hits": 43]	
9	nyt://artic	Nicholas C. Elizabeth A. Harris					It was, some might argue, an apt dry run f				3	-0.567	article	[('name': 'None	It was, sor	["hits": 43]	
10	nyt://artic	Benedict Carey					For decades, researchers have ransacked				3	0.91	article	[('name': 'None	For decad	["hits": 43]	
11	nyt://artic	Elizabeth Bumiller					WASHINGTON cause he needs that adrenaline				2	-0.9952	article	[('name': 'Military W	WASHINGTON	["hits": 43]	
12	nyt://artic	Douglas Martin					Elizabeth l r work in the Bronx for the sim				5	0.3818	article	[('name': 'None	Elizabeth l	["hits": 12]	
13	nyt://artic	Neil Genzlinger					Dario Dã€"Ambrosi certainly makes ãœœl				3	-0.9118	article	[('name': 'Theater R	Dario Dã€	["hits": 12]	
14	nyt://artic	Debra Weiner					As Cpl. Ricago and the Illinois National Gu				1	-0.7783	article	[('name': 'Chicago N	As Cpl. Ric	["hits": 12]	
15	nyt://artic	Matthew Brunwasser					SOFIA ãœ" human rights researcher who t				12	-0.9932	article	[('name': 'None	SOFIA ãœ"	["hits": 12]	
16	nyt://artic	John Eligon					The trial of a 42-year-old man charged wit				4	-0.9782	article	[('name': 'None	The trial o	["hits": 12]	
17	nyt://artic	John Eligon					The trial of David Tarloff, who is accused c				7	-0.8338	article	[('name': 'None	The trial o	["hits": 12]	
18	nyt://artic	Shoshana Walter					On the mc complexities that officers face				22	-0.9994	article	[('name': 'The Bay C	On the mc	["hits": 12]	
19	nyt://artic	Benedict Carey					The ordinary miners are long gone, left be				1	-0.631	article	[('name': 'None	The ordin	["hits": 12]	
20	nyt://artic	Nicholas Bakalar					A new study suggests that a 50-year-old d				3	-0.9871	article	[('name': 'None	A new stu	["hits": 43]	

Figure 6. Excel data structure

8.2 Sentiment Analysis

The next step is to gather sentiment of the articles. This used a combination of NLTK to manipulate the text and Vader for sentiment analysis as it had better scoring metrics. Firstly, the function will combine all 4 body_text columns from the CSV file into 1 column of the data frame. Looking at scoring the plain text, for example using an article

(nyt://article/1b332228-8cdd-55a3-8b3c-25432146b382) we have the text (first sentence only in report) of:

Young children in military families are about 10 percent more likely to see a doctor for a mental difficulty when a parent is deployed than when the parent is home, researchers are reporting Monday in the most comprehensive study to date of such families' use of health insurance during wartime

Then we get a scoring of:

*Text score: {'neg': 0.086, 'neu': 0.865, 'pos': 0.049, 'compound': -0.952}
Adjusted Score: {'neg': 0.114, 'neu': 0.784, 'pos': 0.102, 'compound': 0.0847}
Stemmed Score: {'neg': 0.104, 'neu': 0.826, 'pos': 0.07, 'compound': -0.7624}
Lemma: {'neg': 0.111, 'neu': 0.786, 'pos': 0.103, 'compound': 0.3246}*

Looking at these we can see the changes that occur, after reading a number of articles (listed below) the scoring after lemmatization felt the most appropriate. What needs to be clarified here is what is defined as positive and negative sentiment. From investigating articles that rated highly negative or highly positive (viewing compound score), it seems that the negative sentiment is focused towards problem-focused viewpoints on mental health, whereas positive sentiment was solution-focused. For example, using a few problem-focused sentences (listed in the code repository) we find the problem-focused scored:

Lemma: {'neg': 0.252, 'neu': 0.672, 'pos': 0.076, 'compound': -0.9721}

While the solution-focused sentences gave a scoring of:

Lemma: {'neg': 0.14, 'neu': 0.657, 'pos': 0.202, 'compound': 0.7783}

Article Title	Article ID	Link	Scoring
Mental Health Visits Rise as Parent Deploys	nyt://article/1b332228-8cdd-55a3-8b3c-25432146b382	https://www.nytimes.com/2010/11/08/us/08child.html	Text: - 0.95 Lemma: 0.33
Harriet Shetler, Who Helped to Found Mental Illness Group, Dies at 92	nyt://article/acdc861a-77da-5152-8e52-b7da00ff5bc1	https://www.nytimes.com/2010/04/04/us/04shetler.html	Text: 0.87 Lemma: 0.73

Ricky Wyatt, 57, Dies; Plaintiff in Landmark Mental Care Suit	nyt://article/d3cd1c76-0591-500a-bbfd-c1d98c4618f9	https://www.nytimes.com/2011/11/04/health/ricky-wyatt-57-dies-plaintiff-in-landmark-mental-care-suit.html	Text: 0.13 Lemma: 0.67
Cuomo Adviser Takes Pay From Health Industry	nyt://article/97471c2f-6836-5478-a573-9541b29b05d4	https://www.nytimes.com/2011/02/23/nyregion/23cuomo.html	Text: 0.99 Lemma: 0.99
Oregon Congressman, Named in Sex Case, to Resign	nyt://article/3c870be7-1d3f-5747-96f5-e7e5495b6f62	https://www.nytimes.com/2011/07/27/us/politics/27wu.html	Text: -0.84 Lemma: -0.72

Figure 7: Example of lemmatizing affecting score

With this, if we look at the average compound score across the years we can see that the sentiment declines towards 2014 but takes a turn and becomes positive trending, with scoring even crossing the x-axis to now be in the positive. The dip in 2018 (marked in red) is likely due to the increase in Gun topic, and the more negative scoring. Whereas the drop in 2014 is a combination of less articles pertaining to mental health, and an increase of articles relating to Prison which consistently has highly negative scoring.

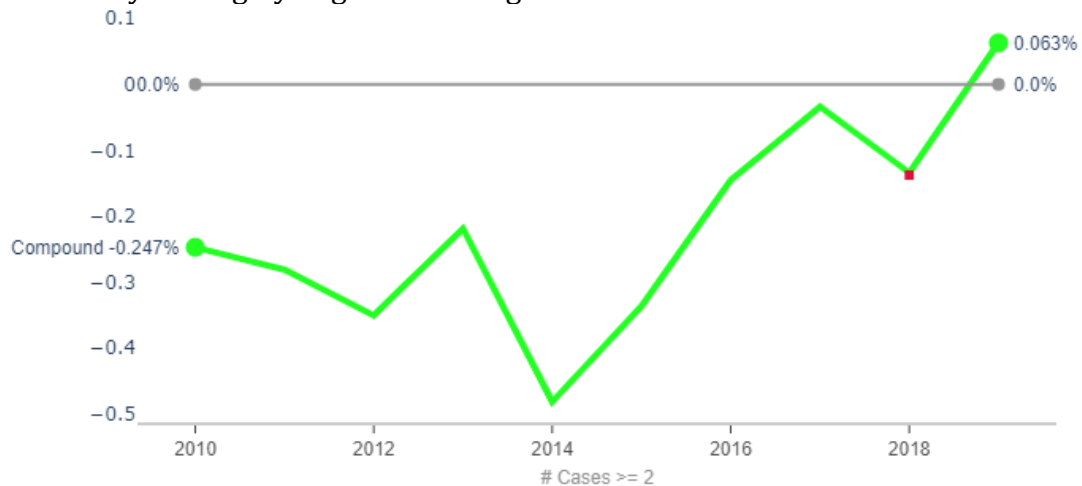


Figure 8: Average compound score over 10 years, grey line is 0%

After the sentiment is added to our data frame the code looks into a count of “mental health” occurrences within each article. The idea is to look into the average counts per month and find a reasonable threshold so that we can discard articles with a low count, i.e. articles pulled for containing the keyword but are very unlikely to focus around mental health. Using this, we find an

average cut-off point of ≥ 1.7 (shown below), meaning we can look at articles that mention mental health at least twice in their articles.

Threshold Establishment

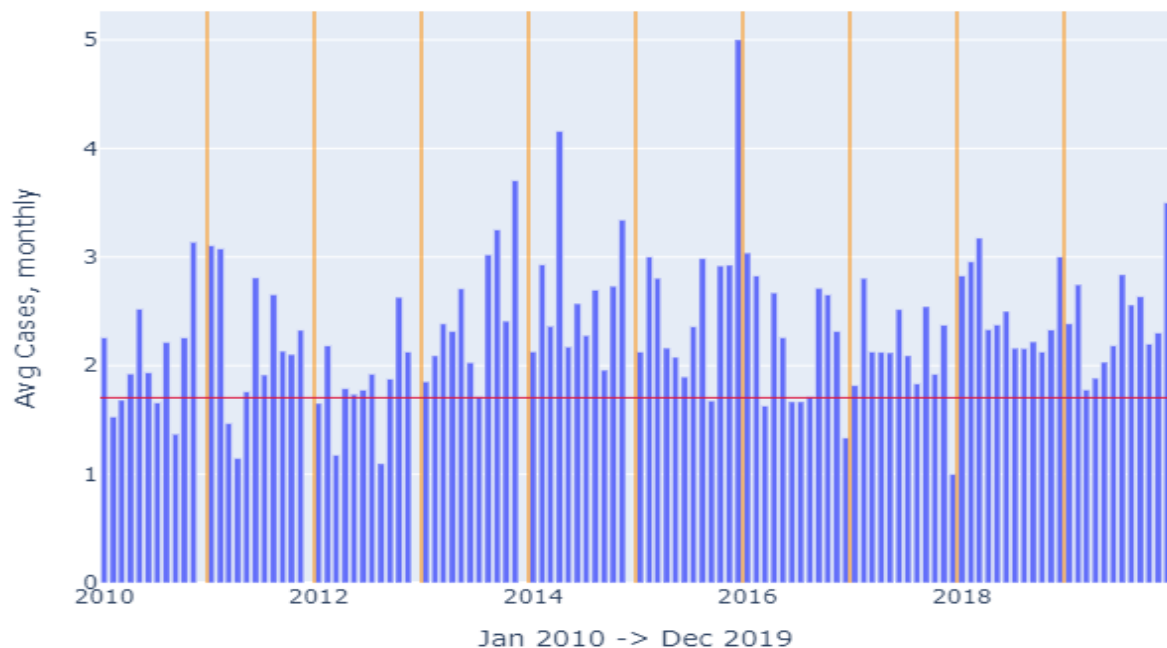


Figure 9 Threshold Establishment, red is threshold, yellow separates years

8.3 Topic Modelling

From here, I used Topic Modelling across all $\approx 5,000$ articles. The first step combines the 4 body_text columns into one, as the data frames do not have the same character limit CSV has. Then, it looks into entity removal. This is a method in which strings of text are firstly tokenized, each word is separated into its own token instead of the entire article text as one token. Then, it uses Part of Speech (POS or POST) tagging to identify it with their appropriate tags (noun, verb, adjective, adverbs, etc). Once these entities have been identified the next step is to remove them or in this case replace them with an empty string. In terms of topic modelling this helps reduce computation time as there are fewer words to cross validate. These entities tend to add little meaning to topic interpretation. For example, without removing the name of a person topic modeling could return their names as a core contributor to a certain topic. Instead of getting descriptive words such as 'politician', 'government', 'judicial' etc topics would be clouded with names. While some names may have a clear association, we can't tell if the article may relate to healthcare reform, advances in research, incidents in prison, by removing these entities we are left with a clearer description of topics, allowing a more accurate and concise topic name.

After Tokenization and Entity Removal, the program will lemmatize the words (Heidenreich,2018). This is a method in which text is transformed to its core meaning. Both Stemming and Lemmatization changes the word to its root meaning but stemming will do so by removing parts of a word, cutting it down. For example "studies" becomes "study" with lemmatize but "studi" with

stemming. For this reason, we use only lemmatization. Afterwards the program will also remove some additional stop-words that were not picked up through previous data pre-processing, this includes words like [may, among, is, with, etc] that don't provide any context to topics. Additionally, some "words" had to be added here that showed up commonly within topic results but had no meaning such as: ['boyd','rev','wu'] this is likely due to a pre-processing error on large scale data.

Vectorization then creates a Count Vector to find similarities between words and reduce the dictionary, thus further reducing the text space we need to analyse to find our topics. For example, "Man" and "woman" would be close together as would "king" and "queen", for our texts we could use this to limit "studies" and "research" to one word, giving us a more accurate set of words for each topic.

The penultimate step is performing passes over the texts to find topics. It starts on 10 topics and works up to 20 expected topics. This range was chosen as anything less than 10 topics would yield little insight expected from the corpus, and more than 20 would dilute the meanings of topics too drastically (for example politics could change to 3 new topics; "state-level politics", "National-level politics", and "international politics". The program uses 50 passes, this allows it to look over the articles and find the best set of words, since this was run on my local machine it was not set higher. Alpha was kept at "auto" as this would learn the distribution of words as it progressed. The program evaluated, or updated, the topics after every 2,000 articles. Too low and this slowed down the process and too high it would not get an accurate reading of the topics, for this reason it remained at 2,000. This resulted in a result of [Topic Number, word, frequency in corpus] for example:

```
0,[ ('study',0.015833242),
    ('child',0.015114548),
```

Limiting the results for each topic to a threshold of first 5 topics (using a number such as 0.015 resulted in too few words for some topics), we can interpret a title for each topic number. The resulting words are attached to the Jupyter Notebooks. **14 Topics appeared to be the best result:**

Research	[('study',0.015833242), (('child',0.015114548), (('health',0.01393376), (('patient',0.013755058), (('mental',0.010072122))]
Guns	[('gun',0.071704715), (('blog',0.037593443), (('shooting',0.031465273), (('school',0.023948966), (('law',0.019910177))]
Healthcare	[('health',0.030876134), (('care',0.023253199), (('state',0.01545426), (('service',0.01030877), (('program',0.010270733))]

Politics	('trump',0.023648681), ('president',0.01861281), ('republican',0.015171797), ('state',0.009938171), ('house',0.009177081)]
Family	[('family',0.016192341), ('child',0.013620538), ('home',0.010397714), ('mother',0.009100331), ('life',0.00843489)]
Veterans	('veteran',0.03443446), ('military',0.020306224), ('war',0.0134340525), ('army',0.010332687), ('soldier',0.009944349)]
Entertainment/Arts	('art',0.011690392), ('film',0.011043259), ('music',0.010293929), ('play',0.008379868), ('artist',0.008312408)]
Gender/Wellbeing/Sexual Healthcare	('woman',0.07482121), ('abortion',0.018366568), ('sexual',0.017263388), ('gay',0.013557949), ('marijuana',0.013556056)]
Government	[('government',0.010643251), ('united',0.010468426), ('country',0.010430458), ('american',0.0075890063), ('state',0.0074350797)]
Judicial	[('court',0.030144297), ('case',0.024635823), ('lawyer',0.024123223), ('judge',0.020129142), ('trial',0.01251834)]
Prison	[('police',0.029628336), ('city',0.024066722), ('officer',0.021117607), ('prison',0.0129525075), ('department',0.01160563)]
Property/Business	[('city',0.013329343), ('company',0.012195402), ('food',0.006127273), ('los',0.0052945227), ('angeles',0.005032553)]
Education	[('student',0.057598058), ('school',0.056078043), ('university',0.023558477), ('college',0.022262461), ('teacher',0.009268459)]
Disability	[('disability',0.06879156), ('psychiatric',0.055413824), ('disorder',0.041535333), ('diagnosis',0.04098582),

	['autism',0.032482862]]
--	-------------------------

Figure 10; The 14 topics and their 5 top keywords

While there is still overlap in the topics, Judicial and Prison, their top words throughout the 10 to 20 topics appeared to clearly define them as Judicial versus Prison and as such were created as 2 separate topics.

Finally, the last step is to get the topic weighting for each article, this involves comparing the article to each topic and finding the distribution within. This results in [topic name, distribution].

	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	
1	negative	neutral	positive	s	compound	Research	Guns	Healthcar	Politics	Family	Veterans	Entertainr	Gender/H	Governme	Judicial	Prison	Property/I	Education	Disability
2	0.111	0.78	0.11	0.5617	0.639282		0	0.045892	0	0.107979	0.17148	0	0	0	0	0	0.026562	0	0
3	0.144	0.725	0.131	-0.8983	0.234845		0	0.219971	0.046993	0.13255	0	0	0.017761	0.153266	0.015615	0.126026	0	0	0.031861
4	0.11	0.756	0.133	0.9003	0.164422		0	0	0.11588	0	0.021118	0	0	0	0	0	0	0.017941	0.294149

Figure 11 Topic Modelling Distribution of each topic per article

Since these results are difficult to graph, an excel formula is used to get the “core topic”. This calculates which topic is the most prominent for each article, which is the topic used in graphing results. We now have 10 years of article data, their sentiment, and applied topic modelling to get their topic distribution as well as core topic. Next, is a dive into analysis on mental health articles and their sentiment, with a look into their trends.

9. Core Contributions

This project examines whether there have been changes in sentiment towards mental health within one online media published (i.e. NYTimes) over the past decade, achieving this through three key objectives:

1. Retrieve a large and credible data source of online media relating to mental health.
2. Classify the articles into their top ‘n’ topics using an appropriate topic modelling algorithm
3. To score each article, then identify any shift or trend in sentiment bias, both positive and negative, towards mental health using appropriate sentiment analysis tools, then graph the results using appropriate visualization methods.

Article Count

Firstly, if we look at the total number of articles containing mental health versus all other articles by The New York Times (note the dual axes) we find that there has been a decline in overall articles, which may be worth future work, but an increase in articles containing mental health.

Non-Mental Health Articles VS Mental Health Articles

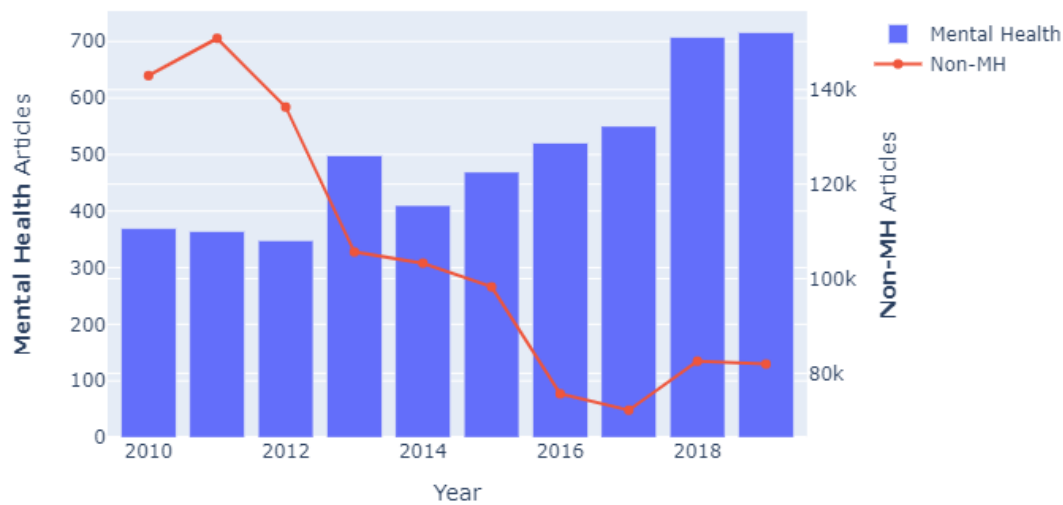


Figure 12: Count of Mental Health (MH) articles vs all other NYT articles

Now let us look at the number of times articles contain mental health, month by month. With this graph we can see multiple spikes; Jan 2011 (Tucson Shooting), April 2014 (Police Brutality, Prison Suicides), Aug 2015 (Prison reform), and Feb 2018 (Parkland Shooting) which all mention the key 'mental health' far more than usual. Also note how December of each year has little to no articles on mental health.

Count of "Mental Health" occurrences, Monthly

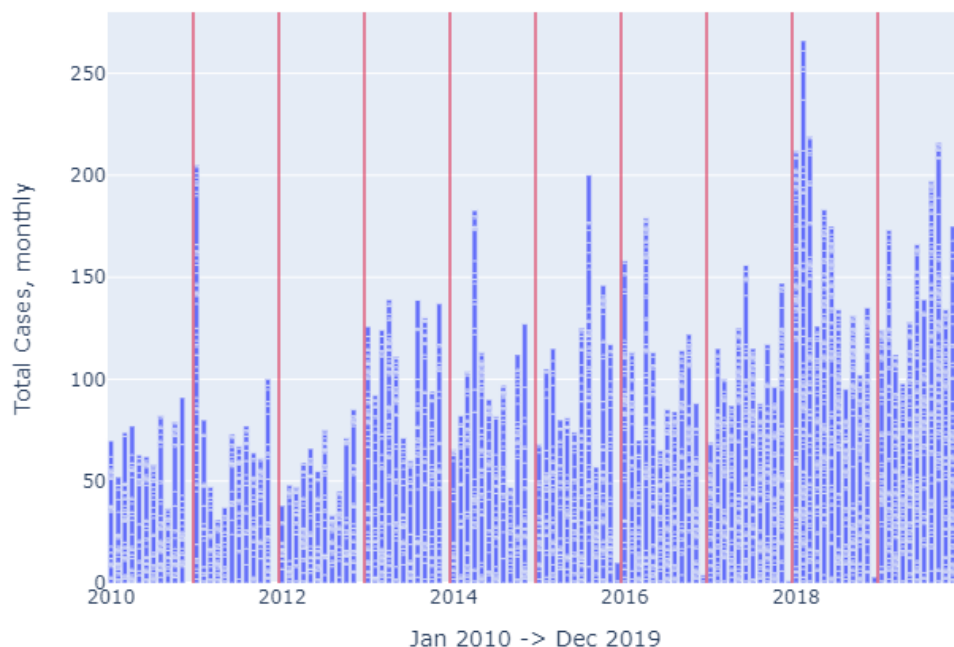


Figure 13: Total counts of 'Mental Health' in an article per month (red line divides each year)

Each vertical blue bar consists of stacked bars (or ticks). Each tick on a bar is an article for that month. The red line is to help segment each year. To limit the results to articles which are focused on mental health instead of being pulled for

containing our keyword only once, we will look at the mean number of cases per month to find an appropriate threshold. Setting it to 1.7 yields the below graph. Going above 1.7 cuts off too many articles, so we'll look into articles that contain at least 2 mentions of "mental health"

Threshold Establishment

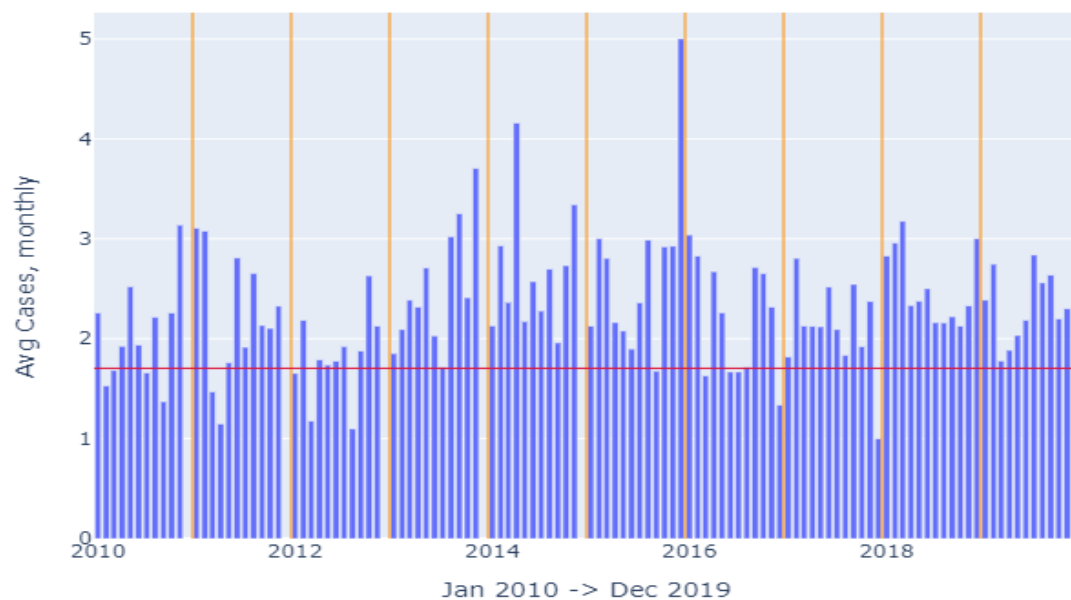


Figure 9, repeated: Threshold Establishment, red is threshold, yellow separates years

To analyse the scoring, we need to understand what compound scoring is. It is computed by summing the valence scores of each word in the lexicon, adjusted according to the rules, and then normalized to be between $[-1, +1]$, a normalized, or weighted composite score.

The first topic, Prison, is negatively rated in almost every case. **This implies that views on mental health for articles relating to prison are problem-focused.** Healthcare has an almost even distribution between positively and negatively rated articles. This could be due to articles discussing the current situation of healthcare (problems) in regards to mental health vs the changes due to be implemented (solutions). Politics shows another even distribution but a total number of articles smaller than Healthcare. Likely due to mixed political responses towards mental health.

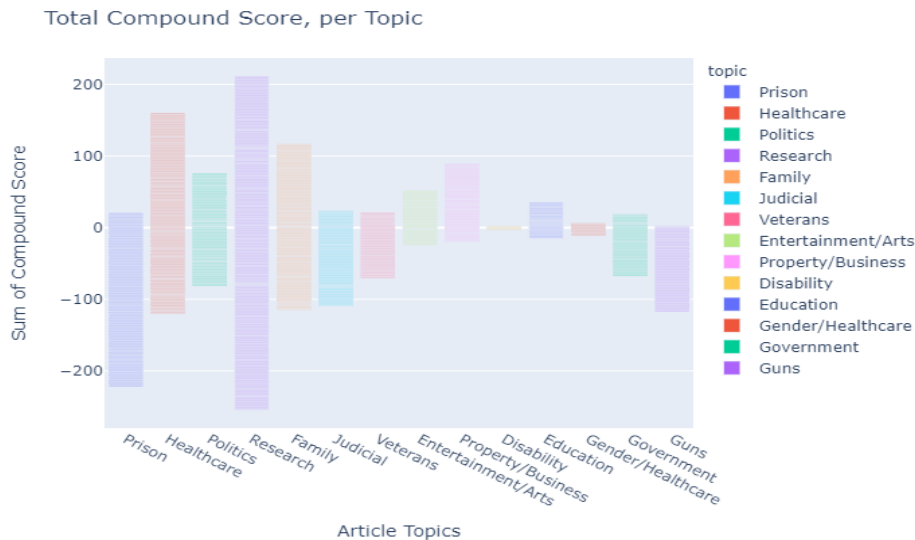


Figure 14:

Total compound score per topic

Research contains the most amount of articles and again is evenly distributed. Due to the definition of sentiment within this project, this result is expected. Research will tend to focus on the problems and then discuss the solutions. This is a good result, if it was skewed towards one rated there would be cause for alarm there.

GUNS

The last topic in purple is Guns. This is nearly always associated with negative sentiment. This is likely due to the spark of gun control articles after mass shootings (a word showing up highly within this topic modelling results).

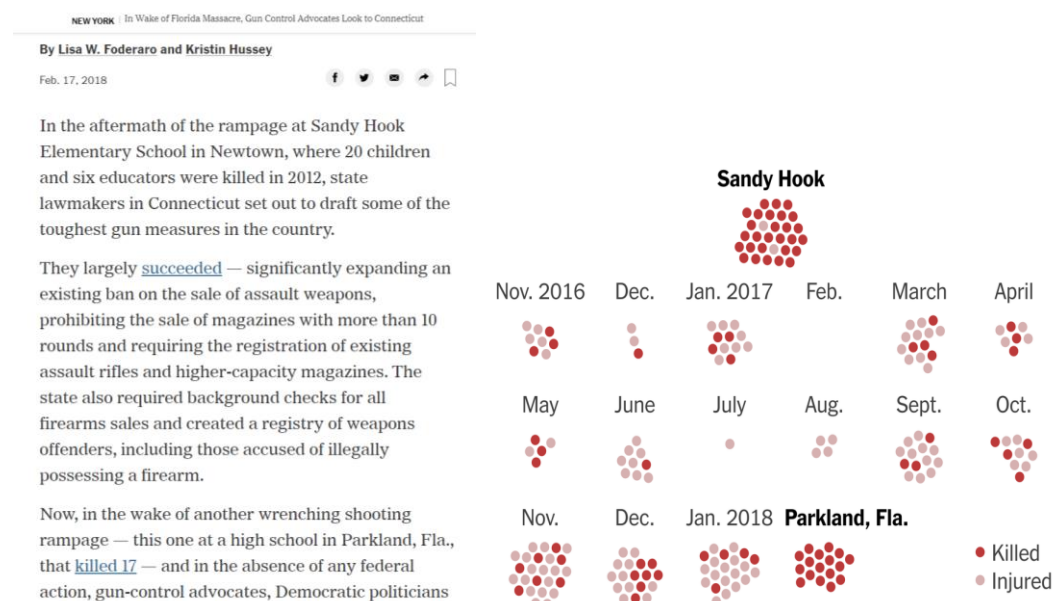


Figure 15/16; Article discussing Gun Control, Infographic on Mass shootings

If we investigate just the sentiment on Guns topic, we find one of the spikes is in Feb 2018, identified in the earlier graph on article counts. Looking at these articles within the results shows a spark in gun control conversation after the mass shooting in Parkland, Florida. This came only months after America's largest mass shooting in Las Vegas, October 2017 and Pulse Nightclub in June 2016. Before, 'Gun' articles focused on deaths or altercations involving guns but began changing to gun control, its current limitations in comparison to other countries. These articles refer to the mental health of the shooter and lack of treatment / healthcare. The spikes in negative sentiment within 'Guns' correlate to mass shootings.

"Mass shootings by people with serious mental illness represent less than 1% of all yearly gun-related homicides. In contrast, deaths by suicide using firearms account for the majority of yearly gun-related deaths." - (Mass shooting and Mental Illness)

It would be interesting to see the changes in sentiment on gun-related articles from non-biased data. Since these articles are gathered for containing 'mental health', some future work might be to perform a comparison on gun-related articles that contain/do not contain mental health.

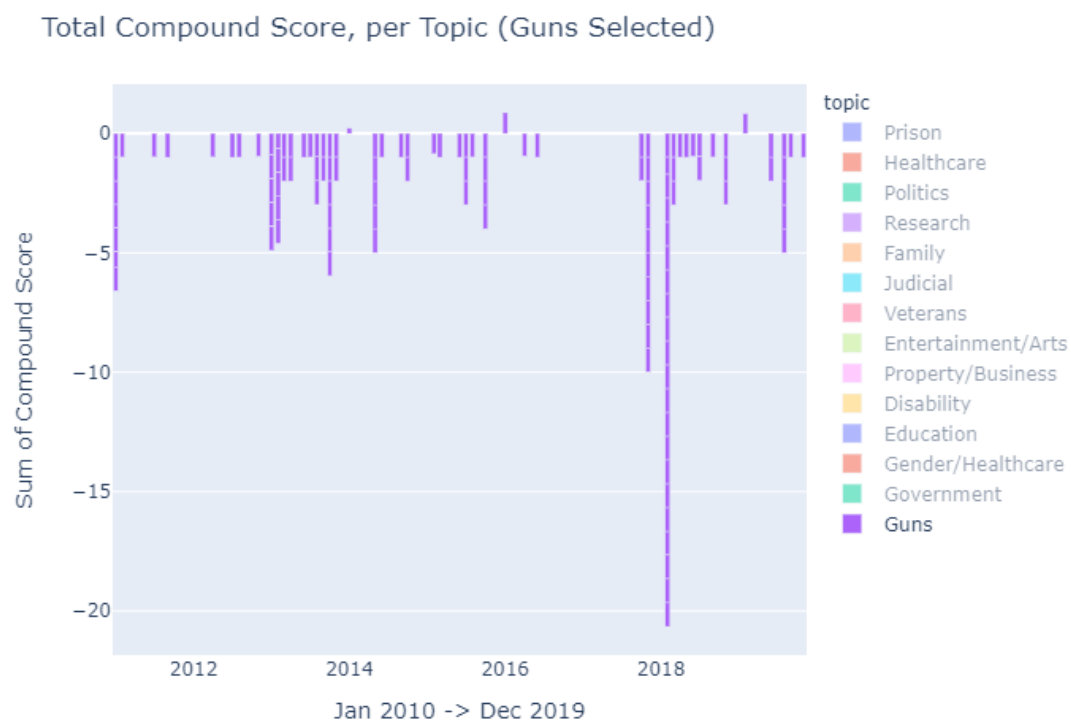


Figure 17; Total compound score for 'Guns', 2010 to 2019

Now, if we average these ratings and look at the sentiment again, for one topic at a time, we find that from 2018 onwards the sentiment towards Guns has almost doubled in negativity.

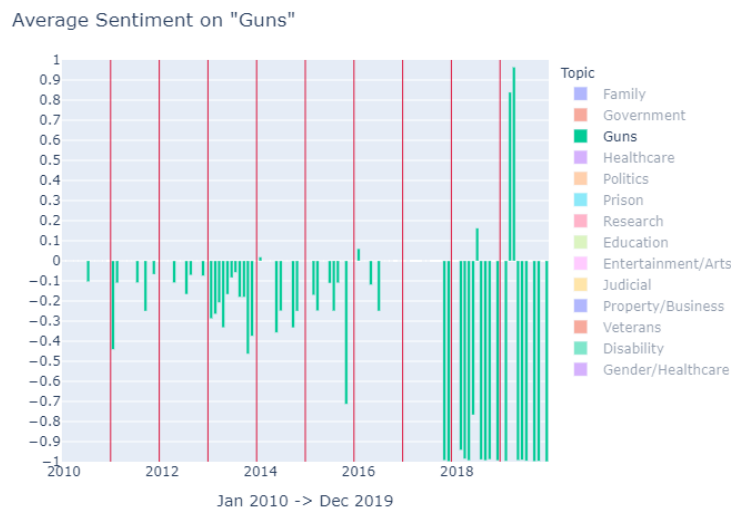


Figure 18; Average compound score for 'Guns'

Prisons subsection

Looking at prison we can see that there are a few spikes in articles throughout the time frame, but a majority of these articles have a consistent and strongly negative sentiment, this is likely due to inmates attempting suicide. Such articles focus on the lack of mental health services available to inmates, the lack of responsibility of the prison staff, and the overall tone being problem-focused.

'They Saw Him Hanging and Did Nothing': A Teen's Agony in Rikers

Nicholas Feliciano, 18, had a history of suicide attempts, but guards stood by for seven minutes as he tried to kill himself.

Figure 19; Article title on prison suicide attempt

Average Sentiment on "Prison"

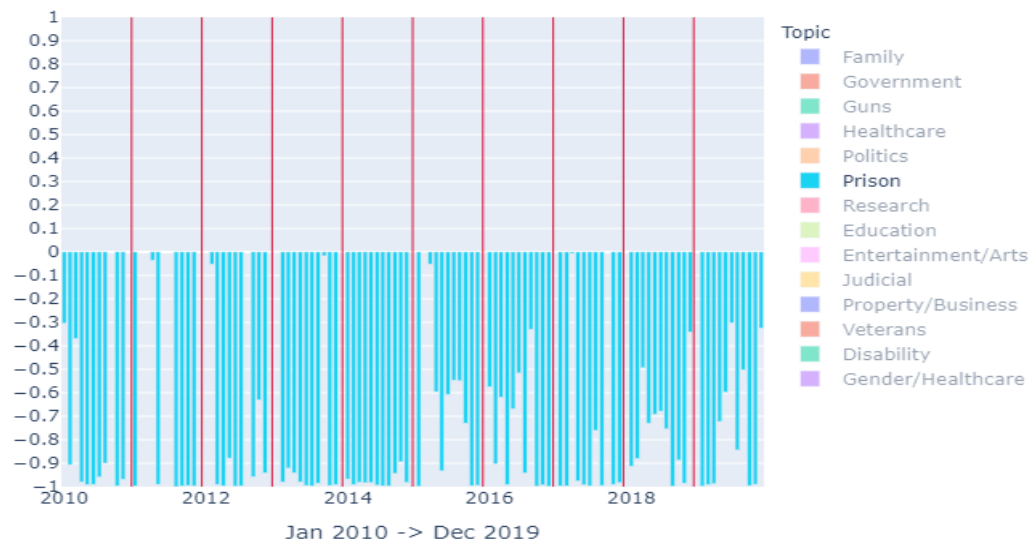


Figure 20; Average compound score on 'Prison'

Since Prison and Guns have a large number of articles, but both are strongly negative, looking at another 2 topics with large coverage and which are more positive we can get a better idea how overall sentiment is improving, and if perhaps some topics will continue to remain negative (i.e problem-focused).

Strongly Negative Topics in Comparison to More Positive Topics

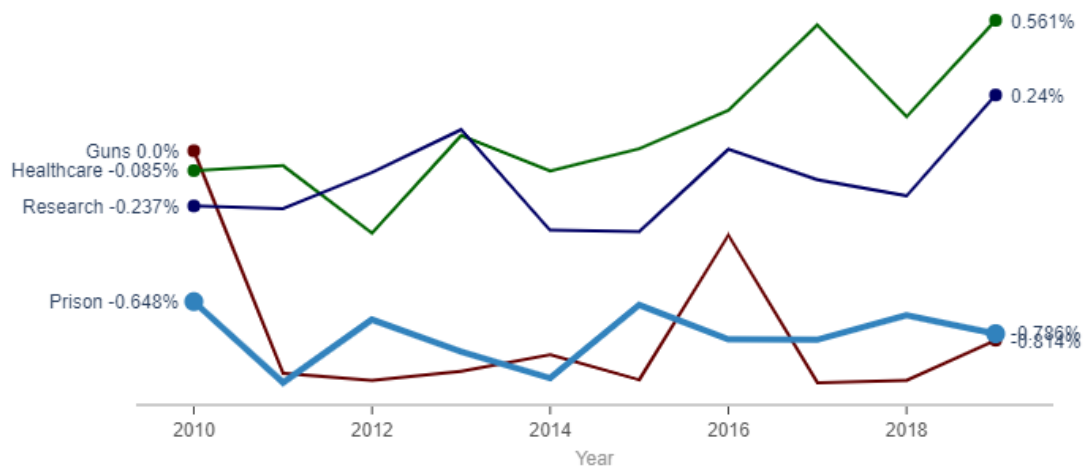


Figure 21; Mean compound score of Guns, Healthcare, Research, Prison

Looking at the mean compound score for some of the topics mentioned above we can spot some trends. Research and Healthcare both take a slight dip until 2014, likely due to less articles pertaining to the respective topic, but begin to increase in positive sentiment by 2019 due to a combination of more articles covering said topic and a solution-focused outlook. This can be seen using the interactive Total Compound Score graph and highlighted respective topics. Guns and Prison both decline and remain at a low, negative sentiment by 2019 with little reassurance of improving. This could be due to more frequent gun-violence

matched with little improvement on regulations within the U.S. With more gun-control debates circulating within the NYT, the outlook is likely problem-focused and therefore will have negative sentiment attached.

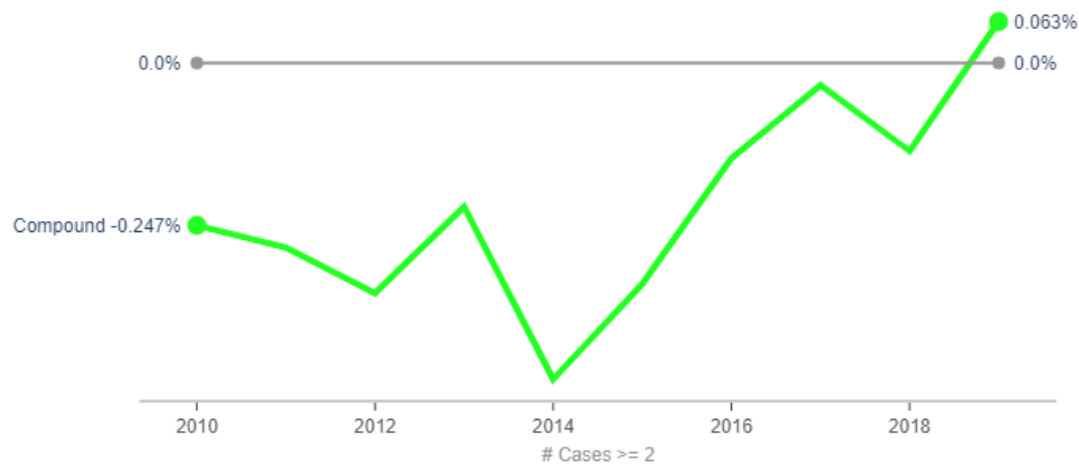


Figure 22; Average compound score over all articles, no topics, 2010 to 2019

If we look at the overall mean compound score the trend appears to decline towards 2014 but makes a quicker recovery in the preceding years, even crossing slightly into the positive. This shows that sentiment towards mental health in online media appears to be improving overall, and at a faster rate than it declined between 2010 to 2014. Both sentiment and count of articles are on the rise. The overall compound score has barely crossed the 0 mark, at 0.063, but is enough to be out of the neutral range $[-0.05, +0.05]$ and can therefore be considered positive.

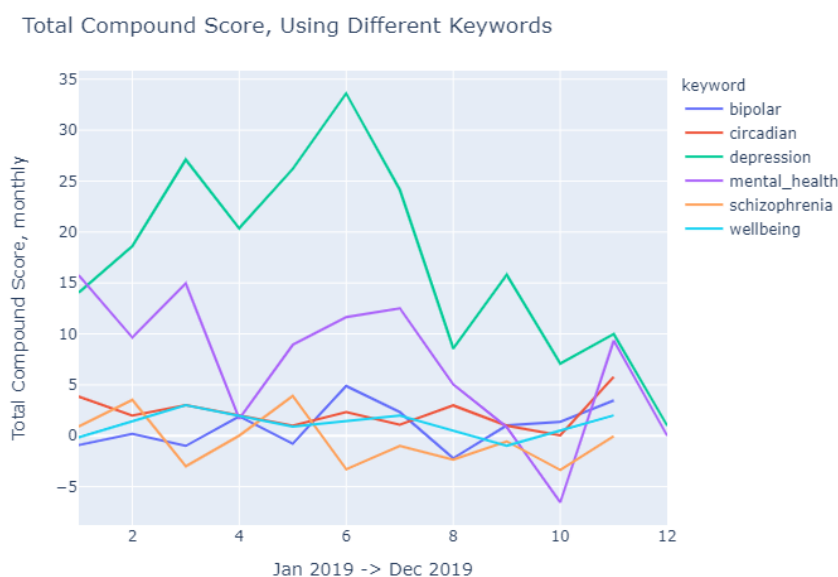


Figure 23; sentiment using different keywords

In comparison, looking at the same sentiment for articles using different

keywords (figure 23) throughout 2019, we find that articles on depression have a higher score over mental health, but also a higher number of articles. Looking

at the mean scoring instead shows a score around 0.25, lower than Circadian Rhythms due to there being such few articles on Circadian Rhythms it is likely an inaccurate score. For example, here mental health has an average scoring that lies relatively closely to the 0 line over all other keywords which also is the case in the in-depth analysis above. Overall, the keywords have too few articles to gather better insight into their attached sentiment, future work here may look into a larger time frame (30 years or more), or a larger data source. Depression would be another very interesting area to explore further, with enough articles being pulled it could be worth looking into how the topics changed as fewer articles containing depression may be focused on mental health, and could provide an interesting breakdown between problem focused and solution focused sentiment.

A Glance at COVID-19 and Mental Health

With the recent outbreak of COVID-19 and the ensuing global pandemic, I felt it was worthwhile to investigate any possible changes towards the sentiment attached to mental health. This meant performing a new analysis on separate data, topics relating to Mental Health appear to have spiked in March most likely due to the COVID-19 lockdown/limitations, since articles containing either “corona” or “covid” had increased from 9 in February to 61 in March. This work is only a prospective insight, highlighting the media’s ability to change their sentiment when necessary.

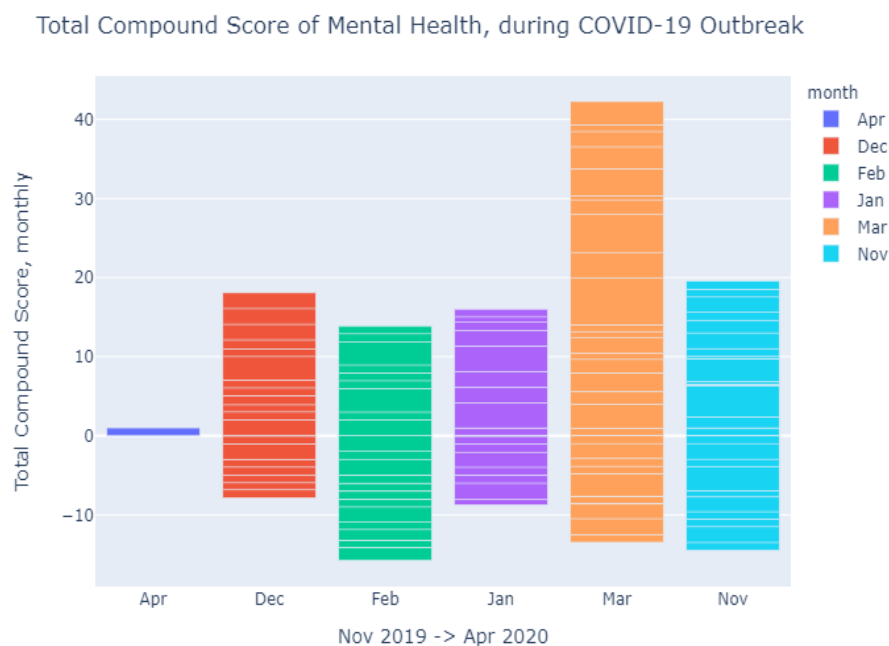


Figure 24;
sentiment by
month since
Corona

10. Evaluation

Findings

Investigating at a topic level has yielded insights into the distribution of articles pertaining to mental health across a range of identified topics. For example, there is far greater coverage (a larger number of articles) within Research, Healthcare, and Prison than within Education or Government yet the change in sentiment of each topic varied greater within Guns (an average coverage of mental health compared to other topics), with more spikes throughout the period that other topics and even showed a solid change, doubling in negative sentiment from 2018 onwards (see Fig 18).

This research has also found that the average sentiment over all topics is shifting from moderately negative in 2010(-0.25%), hitting its lowest in 2014 (-0.5%) , to a smoother path towards positivity in 2019 (0.06%. The compound score breakdown defined within VADER (VADER Sentiment) defines a neutral range of [-0.05,+0.05] therefore sentiment has just slightly crossed the threshold to be deemed positive heading into 2020. It will be interesting to see how 2020 performs in terms of the global pandemic, and America's response.

Strengths

This project aimed to evaluate the change in sentiment towards mental health over a 10 year period, from an online media source, and to perform a comparison on 2 data sources, then create interactive graphs that users can filter to gain a clearer understanding of findings. In terms of topic modelling, meeting the criteria is evaluated in comparison to a blogpost by Charissa Rentier (Rentier, 2018) who performed very similar topic modelling on NYT articles, but over a larger time frame, 1980 to 2017. Sentiment was not performed here, and as such the suggested evaluation is by hand reading of N articles and comparing to the compound score of the lemmatized articles.

Once sentiment within this project has been clearly defined as solution focused being positive compound score and problem focused being negative compound score, I feel that this research has provided solid insight into how online media can perpetuate different bias towards a subject. This project improves upon previous work by Charissa Rentier, where the focus is on sentiment.

The full article text was limited to a unique paid-subscription, while I did have a student subscription to use the API which allowed for both subscription-based and free articles metadata to be gathered, I fear that articles may have been missed, and access to said articles blocked when using URLLib to, more manually, access to the text. For example, in the blogpost by Charissa Rentier the count of 'Mental Health' articles is average around 700-1,000 while this project saw 500-700. However, since this project avoided links to Deal books, Interactives, and blogs due to their inconsistent web-structure for extracting data and their nature of being user-based sourced I feel that focusing on articles by NYT authors gave a better insight into an online media's sentiment towards mental health.

Overall, it shows that some topics are already on a path to improvement (politics has less articles that are negatively rated), and others remain on a low rating (the topic of 'guns' has a consistent and highly negative sentiment). However, as a whole the sentiment and therefore contribution towards social discourse on mental health is improving and appears to be remaining on that trajectory.

Weaknesses

From the data above, figure 22, the ***sentiment towards mental health appears to be improving***. However, this project still had some limitations that are worth investigating for future work. Originally planned for this research was a comparison between 2 media sources, ideally the Irish Times. However, there is no API for The Irish Times, and using the Nexus to gather articles also failed. While there is a repository of articles that can be pulled, each article can be retrieved, in full, by a single word file only. To gather data en-masse gives only the metadata which cannot be used for analysis within this project.

The limitations due to processing power (online resources such as AWS, Google, etc were behind a paywall) may also have hindered the results. Allowing more passes, a different alpha and evaluation more/less often may produce more insightful results. Topic Modelling had to be performed on a corpus level, while sentiment analysis only required processing at an article level. For sentiment analysis, different tools and even pre-processing methods could yield a different angle on the results. Instead of including articles that mentioned our keyword X amount of times, it could analyse only the paragraph of the articles that contained mental health. Other sentiment analysis methods involved using machine learning to label articles as either positive, neutral, or negative. This required a larger amount of data as test/training data was required to set up the machine learning and would also require more human interaction/time than available for this project.

Unfortunately, due to inconsistent data structuring within the data sources (NYT API), metadata could not be graphed for in-depth analysis. Being able to identify the lead author of articles and perform an analysis on sentiment for their articles to identify whether authors are predisposed to a positive/negative sentiment would have been interesting. Also, to perform the test on specific News Desks and see if, for example, National has a different view than Foreign. The tools used to graph, although a minor addition, would be ideal to provide more interaction and a better understanding, perhaps unlocking new insights, through a timeline animation, on-the-go filtering etc.

11. Summary and Conclusions

This project examined the sentiment towards mental health over a 10-year period, in order to assess changes throughout different topics. It investigated

4. A large and credible data source of online media relating to mental health.
5. Classified the articles into their top 'n' topics using an appropriate topic modelling algorithm
6. It identified a shift in sentiment bias, both positive and negative, towards mental health using appropriate sentiment analysis tools, and found underlying trends within topics such as Guns/Prison. The results were graphed with some interactive visualization methods using Jupyter Notebooks.

The key-findings are a positive shift in overall sentiment while some articles became more strongly negative (Guns), and some remained unchanged (Prison, still strongly and consistently negative). The dataset allowed for a clear set of topics to be modeled with enough variety in topics to allow for underlying trends to be identified. For example, Family, Prison, Research are more diverse than Police, Judicial, Politics, which were identified using different thresholds of topic numbers. Overall, this project has found a stigma towards mental health within certain topics. For example, the misconception that the relationship between people with mental illness and gun violence is represented significantly by mass shootings is perpetuated by online media when in fact less than 1% of yearly gun-related homicides is related to mass shootings by people with mental illness. While the shift in sentiment is becoming positive some topics continue to have their stigma towards mental health.

It would be worth investigating, over a larger time period, if it is appropriate to say that mental health is becoming more positive only due to a larger range of topics with slight positive sentiment outweigh topics which are more consistently discussed. For example, if Guns have a track record of strongly negative sentiment and a frequent appearance in media, but several other topics mentioning mental health have positive sentiment yet appear less frequently, is it safe to say that online media is becoming less stigmatized towards mental health, and therefore deconstructing any stigma? Perhaps article scoring could be weighted by the number of views it got, potentially showing the impact it had on readers.

Farrelly, (2015) proposes that discourse within a social setting has the ability to construct an object (say the topic of mental health) in such a way that it is normal for it to be talked about or engaged in the understanding it was constructed under, rather than how it 'really' exists. Discourse analysis is then the ability to deconstruct this object and hopefully allow us to reconstruct it in a more meaningful way. It is hoped this research contributes to this process in a positive way as it demonstrates an emerging shift to a more understanding, respectful and useful way of conceptualising mental health while identifying the media's ability to assist in this construction.

The project is published on GitHub (code and results but not source data, due to licensing), be posted online (such as Medium) to showcase the technical abilities and findings of the project, and hopefully allow other researchers to look further into the the changing discourses around mental health in other online media

(social media for example) and perhaps perform some of the recommended future work.

12. Acknowledgements

I wish to thank Mark Matthews for his help, insights, and direction on this project. Special thanks to Joanna Sweeney for her insights into the mental health field.

13. References

- Anon., n.d. *Christian Views on Suicide*. [Online]
Available at: https://en.wikipedia.org/wiki/Christian_views_on_suicide
- Aoun, S. P. D. a. W., 2004. Assessing population need for mental health care: a review of approaches and predictors.. *Mental Health Services Research*, 6(1), pp. 33-46.
- Baumer, E., 2015. *Testing and Comparing Computational Approaches for Identifying the language of framing in political news..* [Online].
- BBC, 2018. *Why is it hard to talk about problems?*. [Online]
Available at: <https://www.bbc.co.uk/newsround/46425575>
- BeLonGTo, 2018. *LGBT-Ireland*. [Online]
Available at: <http://belongto.org/wp-content/uploads/2018/05/LGBT-Ireland-Full-Reportpdf.pdf>
- Bohan, M., 2016. [Online]
Available at: <https://plot.ly/~mattybohan>
- David M. Blei, A. Y. N. M. I. J., n.d. *Latent Dirichlet Allocation*. [Online]
Available at: <http://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf>
- Farrelly, M., 2015. *The Discursive Construction of Mental Health Problems in Irish Print News Media..* [Online].
- Gansinger, M., 2019. *Social impact of media discourse in the age of iDeology. A perspective from the global periphery..* s.l.:Anchor Academic Publishing..
- Heidenreich, H., 2018. *Stemming and Lemmatization*. [Online]
Available at: <https://towardsdatascience.com/stemming-lemmatization-what-ba782b7c0bd8>
- Heid, M., 2018. *Is it bad for you to read the news constantly*. [Online]
Available at: <https://time.com/5125894/is-reading-news-bad-for-you/>
- Henderson, C. E.-L. S. a. T. G., 2013. Mental illness stigma, help seeking, and public health programs.. *American journal of Public Health*, 103(5), pp. 777-780.
- Jonathan Chang, J. B.-G., n.d. *Reading Tea Leaves: How Humans Interpret Topic Models..* [Online].
- Jongeling, R., n.d. *Choosing Your Weapons: On Sentiment Analysis*. [Online]
Available at: <https://www.win.tue.nl/~aserebre/ICSME2015ERARobbert.pdf>
- Mak, W. P. C. P. L. a. C. S., 2007. Meta-analysis of stigma and mental health.. *Social science & medicine*, 65(2), pp. 245-261.
- McCombs, M. & Reynolds, A., 2002. *News influence on our pictures of the world. Media Effects: Advances in Theory and Research..* s.l.:s.n.
- Ohlsson, R., 2018. Public discourse on mental health and psychiatry: Representations in Swedish newspapers.. *Health*, 22(3), pp. 298-314.
- Papadimitriou, G., 2018. The" Biopsychosocial Model": 40 years of application in Psychiatry.. *Psychiatriki*, 28(2), pp. 107-110.

Philippe Laban, M. H., 2016. *Building and Visualizing long-ranging news stories*. [Online].

Rehurek, R., 2013. *LDA in Gensim*. [Online]
 Available at: <https://rare-technologies.com/python-lda-in-gensim-christmas-edition/>

Rentier, C., n.d. *Using LDA Topic Modelling*. [Online]
 Available at: <https://towardsdatascience.com/thats-mental-using-lda-topic-modeling-to-investigate-the-discourse-on-mental-health-over-time-11da252259c3>

Rita Hamad, J. P., n.d. *A Novel Computational Method for Obesity Policy Research..* [Online].

Samaritans, n.d. *Media Guidelines*. [Online]
 Available at: <https://www.samaritans.org/ireland/about-samaritans/media-guidelines/media-guidelines-ireland/>

The New York Times, n.d. *NYT API*. [Online]
 Available at:
<https://www.datameer.com/documentation/current/New+York+Times+API>
https://en.wikipedia.org/wiki/2011_Tucson_shooting (Tucson Shooting)
<https://www.nytimes.com/2014/04/12/opinion/mentally-ill-in-prison-and-outside.html> (Prison suicide)
<https://www.nytimes.com/2014/04/02/us/police-shootings-of-mentally-ill-suspects-are-on-the-upswing.html> (Police Brutality)
<https://www.nytimes.com/2015/08/06/us/los-angeles-agrees-to-overhaul-its-jail-system.html> (Prison Reform)
<https://www.nytimes.com/2018/02/16/us/politics/fact-check-parkland-gun-violence-mental-illness.html> (Parkland Shooting)
<https://psychiatryonline.org/doi/pdf/10.5555/appi.books.9781615371099>
 (Mass Shootings and Mental Illness), James L. Knoll IV, M.D. George D. Annas, M.D., M.P.H
<https://github.com/cjhutto/vaderSentiment> (VADER Sentiment)