Analysing climate change related misinformation discourse from May 2019 - July 2022 on Twitter

School of Computer Science and Informatics, Cardiff University

Brandon Davies
MSc Computing and IT Management

September 2022

Abstract

Twitter has proven instrumental in the spread of misinformation, with fake news posts propagating faster on the platform than factual posts, with no policies within Twitter for flagging or removing misinformation or disinformation, Twitter will only remove posts if it breaks their policies. This report aims to analyse social media data from Twitter gathered from May 2019 to July 2022, tweets were only collected if they contained predefined misinformation terms. Out of the original misinformation tweets, I filtered them by climate change related terms which were "climate", "climate change" and "warming". The first objective was to investigate the trends of climate change related tweets in the last 4 years and to monitor how world events have affected it. The results clearly displayed a drop in climate change related tweets in 2020 due to the Covid-19 pandemic and a continuous upwards trend from 2021 to the present day. The second objective was to use Natural Language Processing models from Python on the tweets to derive insight from the tweets and represent a story of themes from each month. Exploratory data analysis, pre-processing text and NLP methods such as Ngrams, WordClouds, Hashtag co-occurrence, Top hashtags, Concordance and TF-IDF were used.

Acknowledgements

I would like to thank my supervisor Professor Alun Preece alongside David Rogers and David Tuxworth for all their advice and continued support during this project. I would also like to thank my fiancé Ly for her continued support.

Contents

List of Figures

Listings

Chapter 1

Introduction

The plan for this project is to conduct a thorough analysis on misinformation related tweets collected from Twitter in the years 2019-2022. My investigation will focus on climate relating tweets inside the callout data which was collected if a tweet contained a misinformation related term inside the the tweet text, this does not mean it is always misinformation, it could be satire or a users calling out misinformation. This data set was collected by Cardiff's Crime & Security Research Institute (CSRI). First, I will cover background literature around conspiracies and techniques used to analyse social media data. Further analysis will compare the volume and diversity of climate related tweets over the past 4 years, which can tell us how world events have affected climate discussions. Finally, I will investigate the outlier month of July 2022 where multiple climate events occurred which encouraged heightened climate related discourse.

1.1 Why Conspiracy Theories are significant

Conspiracy theories, as defined by Douglas et al (2019, p. 4), are "attempts to explain the ultimate causes of significant social and political events with claims of secret plots by two or more powerful actors". Ripp and Roer (2022) explain that these theories are often misleading in nature, void of scientific theoretical background. These narratives have become a huge threat to public health and society through attacks on social workers and disobeying government guidelines which in the case of COVID-19, can increase the spread of the disease, therefore harming the health of the population (Alam et al. 2020). Conspiracy theories typically fall into common characteristics:

the world or an event is held to be not as it seems or there is a believed cover up by powerful others (Freeman et al,. 2022).

In recent years, partly due to the pandemic there has been misinformation, conspiracy theories, and fear messages circling about the COVID-19 pandemic on various social media platforms (Gao et al., 2020). This was also witnessed in the 2014 West African Ebola epidemic (Maffioli, 2020). The COVID-19 pandemic has facilitated the spread of misinformation and conspiracy theories at a scale and pace that is unprecedented (Kearney et al., 2020). "Misinformation, disinformation and conspiratorial thinking have been a problem throughout and of course, not limited to the coronavirus pandemic" (Copping, 2022). In February 2020, Dr. Tedros Adhanom Ghebreyesus, the Director General of the World Health Organization (WHO), warned that the world is "not just fighting an epidemic, we're fighting an infodemic. Fake news spreads faster and more easily than this virus, and is just as dangerous" (Imhoff and Lamberty 2020). Such widespread of misinformation on social media has had great impact on various aspects of our society, including public elections, financial markets, environment protection, violent uprising etc (Wang et al., 2021). These irrational beliefs make people hesitant to engage in vaccinations and preventative health measures resulting in a loss of trust in public health, therefore conspiracy theories are creating damaging real-world effects (Copping, 2022). "According to the Pew Research Center, the influence of social media has outpaced traditional news outlets with 68% of US adults using social media as their primary sources of news" (Wang et al., 2021). Therefore, it becomes urgent for us to understand the dynamics of misinformation on social media, so that we can better promote accurate information, deter the spread of misinformation, and mitigate its negative effects on our society (Wang et al, 2022).

The influences of conspiratorial thinking are often severe and far-reaching through the rise of social media like QAnon and online forums of 4Chan, 8Chan and 8Kun (Funk, 2022). Funk (2022) states that with support and encouragement from major political figures including former US President Donald Trump, QAnon became more vocal and bolder. This in turn has increased more conspiratorial thinking and distrust in authoritarian figures between the population as research suggests that people who believe in one conspiracy theory often believe in multiple, even if they are contradictory theories (Freeman et al., 2020; Wood et al., 2012). This is based more on the increasing distrust of authoritarian figures compared to the details of the conspiracy theory itself. In recent years, during the pandemic, not only are infections spreading,

but also conspiracy narratives (Ripp and Roer, 2022), this was witnessed during the Ebola pandemic also (Alam and Shome, 2020).

According to Sen & Zadrozny (2020) the largest Facebook groups dedicated to QAnon had more than three million members in August 2020. The group discussed many conspiracy theories, from Donald Trump waging a secret war on an underground cabal of liberal paedophiles to Pizzagate, which became a popular hashtag on social media regarding a paedophile ring being hosted in a pizzeria (Funk, 2022). These wild accusations resulted in a shootout at Comet Ping Pong and became a threat to society. In 2021, this escalated to the insurrection of the capitol where conspiracy theorists and Trump supporters raided the capitol with claims of a false election resulting in more violence, shootouts and 5 deaths (Healy, 2021).

Social media platforms have been major contributors to the COVID-19 infodemic and beyond (Islam et al., 2020), overloading users with misinformation (Zarocostas, 2020). A number of conspiracy theories have arisen in social media; from fake and dangerous treatments to schemes that the virus is a part of a plan of the global elite to take over the world (Antypas et al., 2021). This has led us to today, where conspiracy theories threaten real information, they encourage xenophobic behaviours to divide our communities and often result in attacks on individuals. Early in the pandemic, COVID-19 led to negative attitudes against Asian people (Sorokowski et al., 2020) as well as an increased support for xenophobic public policies (Oleksy et al., 2021). It is extremely important that we conduct research to examine how conspiracy theories form, spread and convince people to believe them. Through research we can track the timeline of a conspiracy and analyse how threatening it can be to true knowledge, society and the innocent people involved.

Conspiratorial thinking is likely to bring short-term benefits to an individual such as a reduction in uncertainty and increase in control, they usually gain access to echo chambers with like minded individuals and gain a sense of privilege by accessing what they may perceive as secret misinformation (Freeman et al,. 2022). Conspiracy theories have elevated themselves from the fringes to the mainstream due to social medias, TV, documentaries and other digestible news, resulting in nearly everyone having heard of at least one conspiracy such as Princess Diana's murder or the JFK assassination and this leads a large population to believe at least one of them which could act as a gateway into more serious conspiracies (Freeman et al,. 2022). Freeman et al. (2022) believes that healthy mistrust may have tipped over into a breakdown

of trust, this would be understandable as conspiracy believes are commonly found in marginalised communities who believe they cant rely or trust authoritarian figures.

1.2 Motivations behind conspiracy theories

Copping (2022) believes that fear was a major factor contributing to the spread of misinformation during Covid-19 and not everyone was deliberately spreading misinformation. Motivations for conspiracy theories include epistemic, existential and social. Epistemic motivations focus on gathering an explanation in the face of uncertainty, people seek for patterns in randomness and coincidence when impactful events lack official explanations (Copping, 2022). Existential motivations emerge under conditions of threat, anxiety and a desire to alleviate these feelings whereas social motivations can be gained from creating a sense of community and boosting self image amongst these people (Copping, 2022). These motivations can all be witnessed from the uncertainty and anxiety that the world experienced during the COVID-19 pandemic. Mulukom supports this, declaring that higher levels of uncertainty and intolerance or avoidance of uncertainty has been related to higher levels of COVID-19 conspiracy beliefs and conspiracy mentality (Mulukom et al. 2022.). In addition to this, belief in Covid-19 conspiracies has been linked to other problematic attitudes such as prejudice (He et al,. 2020; Roberto et al., 2020), discrimination, decreased well-being, lack of personal control and xenophobia (Mulukom et al. 2022).

Belief in conspiracy theories usually goes against scientific evidence, which decreases the overall education and awareness levels of individuals involved or near the believers (Mulukom et al. 2022). Covid-19 conspiracy beliefs were also generally associated with lower psychological well-being and mental health issues, all these factors being enforced by the conspiracy community can be a massive recipe for disaster and a threat to individuals near them or society as a whole (Mulukom et al. 2022). Uniquely Freeman el al. 2022 found that those who believed in conspiracies were associated with being more likely to share opinions which could result in the vocal minority showing often on social medias. Individuals are commonly observed seeking out misinformation or information and conspiracies that already match their world views, this has happened to prominent anti vaccination protesters, they now use COVID-19 related vaccine conspiracies as a bolster to their misinformation and opinions using that shock value as a promoter (Ball and Maxmen, 2020).

1.3 Why Twitter?

The phrase 'social media' has become omnipresent in everyday life ranging from every age group, it has become crucial to today's society and understanding it. In 2021, over 4.26 billion people were using social media worldwide, a number projected to increase to almost six billion in 2027 (Dixon, 2022). More users than ever are sharing their thoughts, opinions, beliefs online through the accessibility and openness of social media platforms.

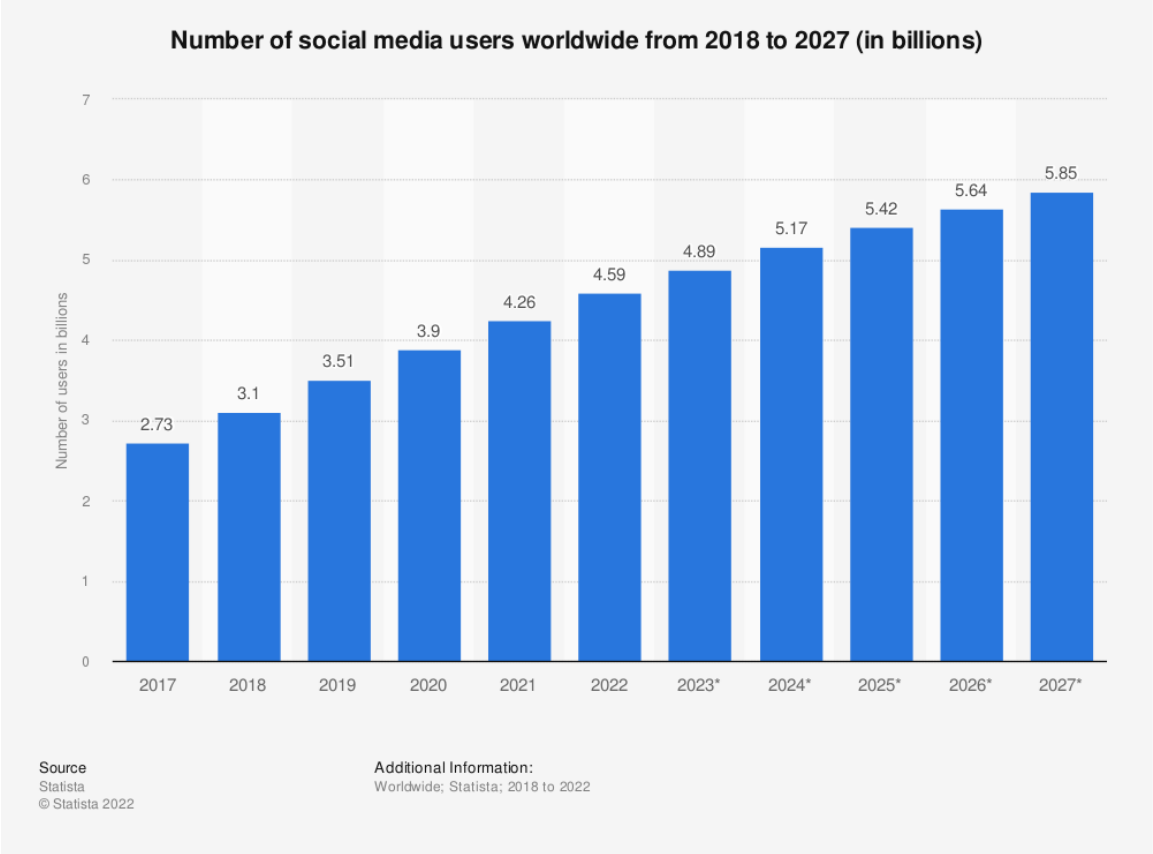


Figure 1.1: Number of social media users worldwide from 2018 to 2027 (in billions)

Twitter has also been the origin of misinformation that has affected the lives of the population, in March 2020, technology entrepreneurs and investors shared a study prematurely explaining the benefits of the drug chloroquine, a drug used previously for malaria, as an antiviral against COVID-19 (Ball and Maxmen, 2020). The study which claimed to benefit users the negative affects of COVID-19 was shared around social media before any kind of medical proof or trials were conducted. Shortly after, a small, non-randomised French trial was posted regarding a related drug named

hydroxychloroquine. The following day, Fox News had already aired a segment with one of the authors of the original document, following that, Trump called the drugs "very powerful" at a press briefing (Ball and Maxmen, 2020). All this escalation started on Twitter with a lack of evidence. This resulted in disruption and worry for patients with conditions such as lupus, who require these drugs to treat themselves. Hospitals also reported poisonings in people who experienced toxic side affects from pills containing chloroquine and also derailing clinical trials of other treatments due to the demand (Ball and Maxmen, 2020).

In this study, I will focus on information posted on Twitter's social media platform due to its open nature and relevance in society. Twitter is an important resource for researching and understanding society at large (Weller et al., 2013) and has been valuable for practical analysis such as Natural Language Processing. Twitter is a unique social platform which incorporates the concept of 'micro-blogging'. Users must abide by a strict 280 character limit to create short posts which contain often opinionated posts that can be directed at other users without the need of accepting friend requests or awaiting permissions to join certain groups. Twitter also invented the popular 'hashtag' categorisation technique which can help users find like minded individuals to engage discussion with or alternatively encounter opposition using the same 'hashtag'. Through these viral keywords we can often find relevant categorised data with a large amount of entries to study. Twitter is also very accessible as they provide an easy to use API with supported documentation to collect and analyse data. Through Twitters API we can access metadata such as the amount of retweets, replies, likes and user information. This data can be accessed through Twitter's developer portal and requires approval from Twitter staff to collect and manipulate public tweets.

There are many cases where people, either unintentionally or deliberately (Fetzer 2004), share unreliable information which causes confusion and suspicion amongst the general population. In addition to this, some users may engage with popular hashtags or keywords with satire which can result in decreasing trust online and ambiguity regarding the theory.

1.4 Aims and Objectives

To be able to achieve my aims, the following objectives will need to be fulfilled:

1. Manage and data wrangle the callout datasets.
2. Extract the climate related tweets from May and June in years 2019-2022, and use exploratory data analysis and quantitative methods to derive insight.
3. Implement further analysis Natural Language Processing methods such as n-grams, top hashtags, hashtag co-occurence and TF-IDF.
4. Analyse the recent history breaking heatwave month, July 2022 and compare these results to the past years.
5. Visualise my results using bar plots, heatmaps, word clouds etc
6. Investigate the results using qualitative analysis.

Chapter 2

Background

This project was carried out over 11 weeks from July 11th to 23rd September 2022. This included biweekly one hour meetings with members of the CSRI team, where I presented my analysis and progress. I was provided feedback from the researchers which guided me and assisted me with the scope of the dissertations aims and objectives.

2.1 Covid-19 conspiracies

Mulukom et al. (2022) state that conspiracy theories have severe consequences and therefore its crucial to understand those theories, why they form and continue to exist in the modern world. Tuxworth et al. (2021) revealed that a popular conspiracy theory from early 2020 was that people believed Covid-19 originated in a laboratory, various other adaptations of this were also rumoured, such as, Covid-19 was used as a biological weapon to control the population. Antypas et al. (2021) further investigated false treatments that were rumoured to treat Covid-19, such as using chlorine to treat Covid-19 without any medical testing or scientific evidence existing at that time. The subtopics identified in this research included "Covid/Weapon", "5G" and "Politics" (Antypas et al., 2021). Other misinformation campaigns have sometimes originated from governments, The Soviet Committee for State Security claimed HIV to be a biological weapon developed by the United States (Geissler & Sprinkle, 2013).

Ripp and Roer (2022) continued this investigating by looking at the correlation be-

tween Covid-19 related conspiracy narratives and vaccination willingness and infection-preventative behaviours, which resulted in a negative association which agrees with other research. In the SEM model of Freeman et al. (2020), conspiracy belief and vaccine hesitancy were positively associated: $\beta = 0.38, p < 0.001$. In addition to this, belief in Covid-19 conspiracies has been linked to other problematic attitudes such as prejudice (He et al,. 2020; Roberto et al,. 2020), discrimination decreased well being, lack of personal control and xenophobia (Mulukom et al,. 2020). Belief in conspiracy theories usually go against scientific evidence, which decreases the overall education and awareness of individuals involved or other parties (Mulukom et al., 2022).

Covid-19 conspiracy beliefs were also generally associated with lower psychological well being and mental health issues, all these characteristics combined can create a recipe for disaster and a threat to individuals (Mulukom et al., 2022). A study from Oost et al. (2022) collected two independent samples at two different times for a total of N = 8264 non vaccinated participants, 26.4% of respondents agreed "moderately", "a lot" or "completely" to the statement "the spread of the virus is a deliberate attempt to reduce the size of the global population". One quarter of the surveyed respondents seem to agree that there is some conspiracy existing beneath society and held some conspiratorial thinking. Many conspiracy theories revolve around historical events such as pandemics, as well as terrorist attacks which can be explained by the insecurity they create and the feeling of a lack of control among the population which further supports the research on conspiracy motivations (Ripp and Roer, 2022). Providing explanations is psychologically advantageous for several reasons, with one common reason that rises in the previous literature: granting an illusion of control. Considering this reasoning, it has been observed that a lack of control has been identified as one of the key drivers of conspiracy beliefs (Imhoff and Lamberty, 2020). For example, the ongoing Covid-19 pandemic was nearly an ideal breeding ground for conspiracies to flourish, as there is no easily comprehensible mechanistic explanation of the disease, it is an event of massive scale, it affects people's lives globally and leaves them with a lot of uncertainty (Imhoff and Lamberty, 2020). A complete change of lifestyle and new imposed restrictions to offset the routine of humans globally will always be met with resistance and agitation.

2.2 Conspiratorial thinking

Oliver and Wood (2014) reported that half of the American public endorsed at least one conspiracy theory. This is supported by a poll by Zogby International showing 49% of the sampled New York Residents beieved that officials in the US government were aware of the 9/11 attacks in advance (Sunstein and Vermeule 2009). This is an alarming number, given how close to the majority of the polled population believed that some conspiracy was alive during and after the 9/11 attack. 9/11 was an inside job, climate change is a hoax, JFK was assassinated by the CIA, the earth is flat, the pharmaceutical industry is suppressing a cure for cancer, vaccines cause autism, Princess Diana was murdered by the royal family, Barack Obama was born in Kenya, the world is ruled by lizards (Lawton, 2022), these are examples of the scale and variety of conspiracies that exist today from completely irrational to maybe plausible. They all follow similar conspiratorial thinking, Lawton (2022) explains that our brains have cognitive biases that make us susceptible to conspiracy theories. They are, proportionality bias, a belief that major events have major causes, intentionality bias, which makes us assume that events are planned by somebody or something and confirmation bias, which means we seek out evidence that supports our beliefs. Personality types also play a part, people who are naturally suspicious of received wisdom and authority are more likely to believe (Lawton, 2022). The conspiratorial mindset may have been an asset in the past, but is now a liability. When it comes to dealing with important issues such as climate change or Covid-19, conspiracy theories are a major obstacle to reasoned debate and evidence-based action (Lawton, 2022).

Freeman et al. (2022) conducted a study involving 2501 adults residing in England and was managed online through Lucid, a survey promoting website. The results found that 50% of the sample population showed little evidence of conspiratorial thinking, 25% showed a degree of endorsement, 15% showed a pattern of consistent pattern of endorsement and 10% had a very high level of endorsement. This study also found that higher endorsement resulted in less adherence to all government guidelines and protective healthcare behaviours. These results are similar across other countries also, in a study conducted in France in 2014 (N = 1500), 20% of respondents believed that the Illuminati were responsive for controlling all international economic activity (Longuet, 2014), for such an absurd conspiracy theory, 1/5 of the sampled population is a surprising number. Uniquely Freeman et al. (2022) found that those who believed in Covid-19 conspiracies were also associated with being more likely to share opin-

ions with others, they also associated with paranoia, climate change denial, general vaccination conspiracy beliefs and a shared distrust in intuitions and professionals. This reveals how intertwined multiple conspiracies and a conspiratorial mentality can be. Perhaps most disturbingly, conspiracy thinking has been shown to be associated with being more accepting to violence (Uscinski & Parent, 2014). Another attribute common to conspiracists who reject science is their reliance on the internet (Diethelm & McKee, 2009; Lewandowsky et al., 2013). Swami et al. (2012) also found that a conspiratorial mindset associated with having less egalitarian human rights attitudes.

Conspiracy theories have elevated themselves from the fringes to the mainstream due to social medias and the abundance of accessible digestible news, nearly everyone has heard of at least one conspiracy and this leads a large amount of people to believe some of them, or at least think they could be possible, which acts as a gateway into more serious conspiracies (Freeman et al,. 2022). Freeman et al. (2022) believes that healthy mistrust may have tipped over into a breakdown of trust, this is evident in marginalised communities who often believe they cant rely on or trust authoritarian figures. Conspiratorial thinking is likely to bring short-term benefits to an individual such as a reduction in uncertainty and increase in an illusion of control, they usually gain access to echo chambers which contain perceived secret information which supports their beliefs (Freeman et al., 2022).

Wood et al. (2012) conducted a study involving 137 participants, they found that the more participants that believed Princess Diana faked her own death, the more they believed that she was murdered. A similar result happened in their second study, the more participants believed that Osama Bin Laden was already dead when U.S. special forces raided his compound in Pakistan, the more they believed he is still alive. This is evidence that individuals can believe in 2 contradictory theories at the same time, with the large amount of theories out there, people associate more with distrust than the actual context and details of a conspiracy theory, spurred in part by the growth of new media, conspiracies have become a major sub-cultural phenomenon and a hobby for some (Wood et al., 2012). If these theories and the associated mindset keep spreading among the common population, over time, the view of the world as a place ruled my conspiracies is a definite threat. Wood et al. (2012) explains this as a threat of conspiracies becoming the default explanation for any given event which generates closed off worldviews "where beliefs come together in a mutually supportive network known as a monological belief system". Wood et al. (2012) suggests that some-

one who believes in conspiracy theories would naturally begin to see authorities as fundamentally deceptive and new future conspiracies would become more believable, leading people to believe multiple conspiracies at once. It is therefore unsurprising to predict that exposure to conspiracies increases uncertainty and then that uncertainty can lead to endorsing a wider range of conspiracies, even if they are contradictory (Wood et al., 2012) Individuals are commonly observed seeking out misinformation or information and conspiracies that match their world views, this has happened to prominent anti vaccination protesters, they now use Covid-19 vaccine conspiracies and outrage to bolster their own campaigns and interests. Not all conspiracy theories fall under the 'deceptive officialdom' umbrella, antisemitic conspiracy theories are an important historically important exception, instead of accusations against the elite for abuse of power, there existed theories around Jewish communities, usually of attempts to seize power for themselves (Wood et al., 2012).

2.3 QAnon

QAnon is a recent example of fringe conspiracies becoming mainstream and threatening the balance of society, QAnon emerged from the fringe anonymous forums such as 4Chan, 8Chan and 8Kun but was occasionally supported by American politicians and even the current president at that time, Donald Trump which raised them to notoriety (Funk and Speakerman, 2022). QAnon influenced the spreading of conspiracies with serious accusations such as #Pizzagate which involved Comet Ping Pong, a pizzeria and family friendly restaurant. The venue was accused of holding a child sex ring and linking it to satanism (Funk and Speakerman, 2022). Usually these conspiracies would be seen as nothing but ridiculous by the average person, but QAnon had gained a huge following on social media, reportedly 3 million members across multiple Facebook groups according to Sen & Zadrozny (2020) in August 2020 so their influence was noticed. This conspiracy was debunked multiple times, this included an official response from the Metropolitan Police Department of the District of Columbia. Ultimately when Donald Trump lost the election in 2020, QAnon was involved in influencing another conspiracy regarding the election where people believed the election was fraudulent and rigged for Trump to lose (Funk and Speakerman, 2022). This resulted in QAnon, alongside Trump himself, influencing the masses to stand up for their country resulting in the population storming the capitol, this caused disruption, carelessness and harm to the population involved (Funk and

Speakerman, 2022). We need to understand how these conspiracies theories emerge and counteract the infodemic of misinformation otherwise more harm will come to societies, chaos will ensue, trust will be lost along with knowledge and education.

2.4 Climate change conspiracies

In the past years, as climate disasters have become more common, climate denial has also increased, for example, a 2013 poll conducted in the United States indicated that almost 40% of the sampled population believed that climate change was a hoax (Uscinski et al., 2017). In a study conducted by Bolsen and Druckman (2018), candidates were chosen randomly on Amazons Mechnical Turk platform, 484 responded, respondents were asked to measure how strongly they agree or disagree to the statements. When faced with the statement "To what extent do you agree with the following statement: the idea that climate change is primarily due to human activities is a hoax or a conspiracy?", 52% of the sample responded with 1 (strongly disagree), 15% responded with 2 (mostly disagree) (Bolsen and Druckman 2018). From this small sample, the study portrayed a much more positive picture of how many believe in a climate change hoax, but this sample is only directed at the demographic looking for work on mTurk, which would be quite a restrictive sample. Climate skeptics believe that the well publicised consensus of climate change and the actions it has on the planet is either manufactured or illusory with a hidden agenda to serve the interests of a nefarious force, some examples that are often named are, United Nations, liberals, communists, or authoritarians (Uscinski et al., 2017). Climate deniers often believe it is just another hoax or cover story to exert control over the population and take away their freedom (Uscinski et al., 2017), others call it the "biggest scam in history" (Sussman, 2010, p.215; Lewandowsky et al., 2013). As previously seen for Covid-19 conspiracies, climate deniers are less likely to participate politically or take actions to lessen their carbon footprint or take part in activities to counter the effects of climate change (Uscinski et al., 2017).

97% of climate scientists have concluded that human-caused global warming is getting progressively worse (Cook et al., 2013). Climate change conspiracies therefore represent a unique case in that scientific agreement has already been solidified, but the public opinion at the same time is split (Uscinski et al., 2017). Theories claim that climate scientists purposely fake data to receive research funding or that cli-

mate change is a hoax to undermine local sovereignty are examples of climate change conspiracies (Douglas & Sutton, 2015). Lewandowsky et al. (2013) found that 20% of respondents believe climate change is a hoax perpetrated by corrupt scientists to spend more of the tax payers money on climate research. Other U.S polls find that 37% of respondents believe global warming is a hoax and 41% say that its definitely possible that global warming is a myth concocted by scientists (Jensen, 2013; Cassino, 2016). This is a similar phenomena witnessed with moon landing conspiracies, there is a clear rejection of scientifically proven facts, alongside scepticism against large organisations such as NASA in regards of accusing scientists of faking data and keeping the hoax secret over decades of research.

Climate scepticism gained enough attention from the public and academics that there were nine independent investigations in the United States and United Kingdom in connection to the "Climategate" incident, also known as Climatic Research Unit email controversy in 2009, which involved the hacking of thousands of emails from a University of East Anglia Climatic Research Unit, this resulted in exoneration of the climate scientists involved of any falsifications or wrongdoing (Lewandowsky, 2014). From the numerous amounts of research, literature and vast amount of people, scientists, agencies and governments involved in climate science, this likelihood of a conspiracy escaping exposure over the course of decades would be incredibly low (Keeley, 1999). Due to the bad reputation of conspiracy theories, they are muted in political discourse, because of this, conspiracy theories have cemented themselves online, particularly in anonymous forums (Uscinski et al., 2017) and on forums such as Reddit using 'throwaway accounts' with no repercussions.

Jolley and Douglas (2014) show that exposure to climate conspiracies can reduce people's intentions to reduce their personal carbon footprint, this can affect thousands of people through spreading narratives on social medias or people in their social circles. Conspiracy theories have also had influence during political campaigns, while Obama was president, the Republican controlled Congress was unwilling to address climate change, therefore President Obama acted alone to limit carbon emissions, this was also met with scepticism and conspiracy in itself (Uscinski et al., 2017). Science challenging conspiracies usually fall into two categories, ones that accuse industry and corporations, or accuse government. Conspiracies that accuse large industries such as the pharmaceutical industry tend to be accused of reaping huge profits at the expense of the common population or in this example, the ones who are in need of

pharmaceuticals (Uscinski et al., 2017). A case study for this would be the HIV and AIDS epidemic where groups in the United States and South Africa claim that the link between them is a fraudulent lie to sell phoney drugs (Nattrass, 2013). Climate change conspiracies can fall into both categories as some conspiracy thinkers believe the private organisations such as National Geographic are as much to blame for spreading false narratives as the governmental bodies such as the Intergovernmental Panel on Climate Change (IPCC) and United Nations (UN).

While the political right is generally more likely to be responsible for conspiracy theories that call into question the legitimacy of climate science, the political left is not immune to believing in conspiracy theories generally (Uscinski et al., 2017). A t-test conducted by Van der Linden (2015) revealed that, conservatives were significantly more likely than liberals to endorse the statement that "global warming is a hoax" (M = 3.87, SE = 0.17) vs. (M = 2.18, SE = 0.13), t(228) = 8.24, p ¡ 0.001. This could also be linked back to the fact that more than 90% of books endorsing scepticism towards environmentalism that were published since 1972 were sponsored by conservative think tanks (Jacques et al., 2008). Another study conducted by Lewandowsky et al. (2013) found that in their main SEM model, it showed a negative association between conspiracy theorising and conservatism, suggesting that conspiratorial thinking is more prevalent on the political left. This could be due to certain conspiracies being favoured by one side of the political spectrum, such as 9/11 being accused of being "an inside job" is favoured more by the political left whereas the denial of climate science is usually favoured more by the political right (Lewandowsky et al. 2013). Besides these findings, people on both sides of the political spectrum are capable of rejecting scientific finding that do not confirm their ideologies (Lewandowsky and Oberauer, 2016). In 1986 only a minority of Americans had heard of climate change but by 1988 heightened media coverage had made a majority of the public aware. By 2006 more than 90% of the U.S. public had heard of the threat of climate change (Nisbet & Myers, 2007). This number will only increase as Covid-19 threat and coverage decreases, climate change is seen as the new threat and the new hoax to keep the populations under control.

2.5 How to counter conspiracies and the spread of misinformation

Considerations must be made on what countermeasures might be available to reduce the scope, influence and spread of conspiracy theories along with misinformation. Conspiracy thinking is, by definition, difficult to correct because any evidence which doesn't align with their beliefs is itself considered part of the conspiracy or evidence that a conspiracy exists (Lewandowsky et al., 2013; Bale, 2007; Sunstein & Vermeule, 2009). A central feature of conspiracy theories is that they are extremely resistant to correction, direct denials or counter speech by government officials, or any contrary evidence can usually be processed as a product of the conspiracy itself (Sunstein and Vermeule 2009). Examples of this have been during the ever increasing heatwaves across the world, in the UK, the most recent heatwave in July 2022 was the target for a lot of controversy, social media users were accusing weather channels of creating a false emergency disaster and scaremongering viewers with increasing the darkness of colours on their maps or changing from green temperatures to red. After fact checking, these viral posts were discussing maps with different purposes out of context and comparing them to the current heat maps to show temperatures, so this was classed as another piece of misinformation to spread a false narrative (Evon, 2022). When faced with clear evidence that temperatures have increased over the years, conspiracists still find a way to turn the narrative in their favour by falsifying evidence. Banas and Miller (2013) found that fact-based anti conspiracy arguments were effective in reducing belief in conspiracy theories about 9/11. These techniques and strategies could be tested to reduce belief in climate change conspiracy theories and reduce misinformation spreading (Lewandowsky et al., 2012). Sunstein and Vermeule (2009) suggest that instead of debunking a single conspiracy theory, scientists and policymakers should try to debunk many at the same time. This is based on prior research that conspiracist beliefs tends to be wide in scope. Alternatively a consequence of multiple rebuttals at the same time could raise the complexity of possible conspiracist responses, with conspiracist ideology it might make other conspiracies mentioned more believable and drive them deeper down the rabbit hole but with the hopes of making the conspiracy seem increasingly ludicrous and unbelievable. Alternatively, providing additional scientific information may only amplify the rejection of such evidence, rather than the conspiracist accepting the evidence (Lewandowsky et al. 2013).

Another method that research has found to decrease belief in conspiracy theories is promoting an analytical thinking style, Swami et al. (2014) found that an analytical mindset is negatively associated with belief in conspiracy theories and prompts more careful information processing, therefore increasing attention on fact based evidence and where the content originated from. This was confirmed in the results of the study by Swami et al. (2014), greater belief in conspiracy theories was significantly predicted by lower analytic thinking, greater intuitive thinking and lower open minded thinking. Belief in conspiracy theories was associated with a lesser tendency to rely on analytic processing of information and a greater tendency to rely on intuitive information processing (Swami et al., 2014). Study 2 provided evidence that an experimental manipulation designed to activate analytical thinking was effective at reducing belief in conspiracies (Swami et al., 2014), this is a preemptive measure to combatting conspiracies and people believing or spreading misinformation but a crucial part of slowing down the growth of conspiracies and false narratives.

To prevent conspiracy narratives, we must learn how they spread, some theories bubble up spontaneously and appear over many different social networks, others are spread quite intentionally by conspiracy entrepreneurs who profit directly or indirectly from pushing these narratives and doubts to a wide audience (Sunstein and Vermeule 2009). The most common case of conspiracies being pushed for personal gain are political campaigns, very commonly conspiracies and slander will be spread before voting to sway audiences to vote for a perceived preferable candidate, this was famously seen for the 2016 presidential election between Donald Trump and Hillary Clinton. Bovet and Makse (2019) found that out of all the tweets linking to new articles in the 6 months before the election, fake news represents 10% and extremely biased news 15%. The #Pizzagate conspiracy also formed around this time attacking Hillary's character. Conspiracies are also used to create outrage, by exaggerating reality and turning it into a conspiracy it causes people to take action and creates a sense of community to work together, even though conspiracies are held by the minority currently, if only a small fraction of believers act on their beliefs it can still cause harm to innocent parties (Sunstein and Vermeule 2009).

2.6 Data collection and analysis methodologies

Abd-Alrazaq et al., (2020) collected tweets between February 2nd, 2020 and March 15th, 2020, using the standard Twitter API searching for predefined search terms ("corona", "2019-nCov" and "COVID-19"). This method is a very common way to collect large amounts of social media data as Twitter is open for all to view unless the account tweeting is a private account, for example, Facebook is locked by multiple levels of privacy due to the friending feature with multiple tiers in privacy settings or private groups. This method must be used carefully, if the specified terms are case sensitive, the data collected will only contain one variation of the search time, this would require making the tweets all lower case and setting search terms to lower case.

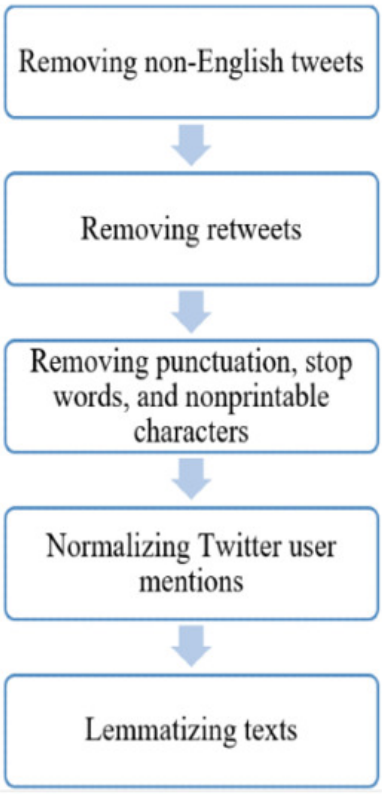


Figure 2.1: Pre-processing workflow (Abd-Alrazaq et al., 2020)

The tweets were then stored in a PostgreSQL database recording the tweet text, metadata such as, number of likes and retweets, profile information, followers and the time the tweet was posted. The Python library Tweepy was used to access the Twitter API and search for the specified terms (Abd-Alrazaq et al., 2020). The stored tweets were then pre-processed by first removing non-English tweets, this can be done via Twitter API metadata, although not always 100% accurate. Then Abd-Alrazaq et al. (2020) removed retweets, removed punctuation, stop words and non-printable characters using the Natural Language Toolkit Python library. This is a very effective and common library used in Natural language processing for cleaning, pre-processing and analysing. Proceeding that, they normalized the Twitter mentions and finally used WordNetLemmatizer from NLTK to lemmatize each token in the tweets (Abd-Alrazaq et al., 2020). After pre-processing, the tweets were analysed

and visualised into word clouds using single word and double word combinations. Abd-Alrazaq et al. (2020) also used Latent Dirichlet Allocation (LDA) topic modelling algorithm from the Python package sklearn, an unsupervised machine learning generative statistical model which identified a set of topics, then described as China, outbreak, wearing masks etc. A sentiment analysis was also performed on the tweets using the Textblob library, this was conducted by extracting the mean number of retweets, likes, followers for each tweet and then each topic to calculate perceived interaction and reach of the tweets (Abd-Alrazaq et al., 2020). Ultimately, the results show that the highest mean amount of likes were among the 'economic loss' topic and on the other side, 'travel ban' and 'warning' related topics had the lowest amount of mean likes. For the mean retweets, 'panic buying' topic averaged 0.89 and 'eating meat' averaged 7.11 retweets (Abd-Alrazaq et al., 2020). I believe these methods were very effective at achieving the desired goal. For my project, I would need to remove all usernames for privacy reasons, also lemmatizing could make word clouds harder to interpret as words like media changing into medium could change the meaning. LDA topic modelling is a great tool to check for common themes within conspiracy tweets and provided effective results in the study research by Abd-Alrazaq et al. (2020).

Gruzd and Mai (2020) conducted a study regarding the viral hashtag #FilmYourHospital, which was a popular conspiracy branch during the Covid-19 pandemic. This study used Netlytic to collect and analyse data with Gephi to visualise the resulting communication network over time by the means of social network analysis. The Python library Twarc was used to check if accounts had been deleted or suspended by Twitter and the Botometer API was used to analyse if an account was acting like a bot with automated behaviours (Gruzd and Mai, 2020). These methods were very effective at displaying a social network analysis visualisation which can clearly track how the hashtag originated and which actors were most responsible for its virality. As I am not familiar with these programs and with a limited time frame, I would look for a more Pythonic related method to visualise my data. In my analysis, I will not differentiate between bot like behaviour and the average user as the data is not scraped immediately after posting, which means Twitter should hopefully ban the junk tweets.

As Gruzd and Mai (2020) discovered, aside from 15,699 tweets in Portuguese and 73,010 tweets in English, there were tweets in 33 other languages, suggesting that the hashtag gained some international reach. My analysis will be focused primarily on

the English language tweets using Twitters language metadata. The social network analysis shows one of the biggest influences to the virality was @DeAnna4Congress, a verified account owned by DeAnna Lorraine, a former Republican Congressional candidate who recently ran against Nancy Pelosi for the U.S. House California District 12, Ms. Lorraine added legitimacy to this campaign and directly asked her 150k+ followers to continue this trend whereas the original poster only received 30 retweets Gruzd and Mai (2020). Other methods of data collection include using the "twitter2stata" package in Stata and downloading tweets mentioning certain keywords or hashtags as used by Kearney et al. (2020). As I am more familiar with Python, collecting my own subset of data would be easily accessible through the Twitter API and Tweepy Python library.

2.7 Natural Language Processing and Machine Learning

Tuxworth et al. (2021) uses the Python libraries SpaCy, Gensim and NLTK to preprocess tweets and model them. Then they are further categorised into European and non-European tweets and categorised into 3 languages: English, French and Spanish. The data was collected from Twitter metadata over 2 months for 2019 and 2020, specifically geolocatalisation data, unfortunately this was only available for 0.34% of the tweets.

The BERT model which is a transformer-based machine learning technique for natural language processing (NLP) was used to break down the most common topics using a set of terms which included "conspiracy", "misinformation" etc, while for Spanish and French BETO and FlauBERT were used. The BERT model was used to classify tweets into a specific geolocation using 3166 country codes as only 0.34% of the tweets had geolocation metadata (Tuxworth et al., 2021).

The BERT models all achieved over 85% accuracy, whereas the English models performed better than the other languages at 92% accuracy. There were also attempts to train the BERT model using the 2019 subset but this only made the accuracy of the English dataset drop by 1% (Tuxworth et al., 2021). The mBERT model was also trained and tested, a multilingual BERT model, unfortunately this only had inferior results. Lexical specificity and word embeddings are used to explore the classified tweets and reveal insights into disinformation. For example, it is shown that the

conspiracies surrounding the origin of COVID-19 are revealed through comparing the most similar words to a relevant keyword (Tuxworth et al., 2021). This shows that these methods are highly effective at identifying current and emerging trends in social media data, early detection can prevent misinformation and shut down conspiracy theories. Word2Vec was used from the library Gensim to find words that are similar to a specific query, for example "vaccine" (Tuxworth et al., 2021). Future work could research more into accurately predicting which tweets were misinformation as this can be a very difficult task with some users using sarcasm and the existing NLP models not dealing very effectively with that. NLP could be crucial to stopping misinformation through fact checking algorithms and knowledge-based detections to evaluate the reliability of content, such as evaluating whether the knowledge from text content is false via manual fact-checking (e.g., expert-based, or crowd-sourced fact-checking) (Grinberg et al., 2019) or automatic fact-checking through Natural Language Processing. NLP can also be used to extract relevant information, Alam and Shome (2020) used Artificial Neural Network (ANN), Long Short Term Memory (LSTM), and Gated Recurrent Unit (GRU), which are branches of deep learning techniques to detect news which involved attacks on health workers (AoHW-news) and GRU resulted in an accuracy of 94%.

2.8 Summary of literature

Research has shown that conspiracy theories are becoming more popular as the years progress but surprisingly, interest in some conspiracy theories sometimes increase as the events become older (Goertzel, 1994). A survey in 1963 found that 29% of respondents believed the official account that Lee Harvey Oswald acted alone in assassinating President Kennedy, but in 2001 that dropped to only 13% of respondents believing that he acted alone (Carlson, 2001). Academic research on conspiracies and countering misinformation will need to be increased as conspiratorial mindsets spread. Covid-19 has been covered extensively in academic research involving the individual hashtags, related conspiracy theories and the conspiratorial mindset which affects their daily actions and beliefs. There has been clear proof from multiple academic studies that strongly correlates people who endorse one conspiracy theory tend to endorse others, illustrating a so-called "conspiracy mentality" (Oost et al., 2022).

Lesser-known conspiracies haven't been researched in as much detail, there are gaps

in academic literature for flat earth, creationist and climate change conspiracies etc. These conspiracies also bring a break down in trust while also brainwashing vulnerable people and decreases the effectiveness of scientific facts and education. Recently there has been history breaking heatwaves in the UK around July 19th, hitting 40 degrees celsius, in conjunction with heatwaves across the world becoming more frequent, commonly occurring wildfires in California and devastating droughts in Africa. All these in conjunction are predicted effects of climate change. As we see these world events happen more often, discussion and media reporting has increased rapidly on this subject. The more discussion and media on these topics generates more conspiracies and climate change denial, I believe there needs to be more research into climate change conspiracies to tackle the waves of misinformation and conspiracy spread.

Successful studies using Twitter as a way to retrieve and analyse data such as Antypas et al. (2021) and Tuxworth et al. (2021) have confirmed my original method of using Twitter as the main source of information, this will involve less ethical and legal barriers as I can anonymise the usernames and not worry about breaching privacy ethics and Twitter developer API agreements. In my study, I will analyse the volume of climate change discussion over the last 4 years, taking into account the expected drop over the peak Covid-19 years 2020, 2021. I will then use data pre-processing from NLTK and analyse the content to derive insight from the tweets over each year. I believe BERT modelling would be more effective with short form text such as tweets, it has provided better results in studies such as Tuxworth et al. (2020) compared to the other popular models such as mBERT or LDA topic modelling.

As a collective, we can see that conspiracy theories do have real world effects, such as attacks on health workers during pandemics (Alan and Shome, 2020) or QAnon escalating violence to storm the capitol (Funk and Speakerman, 2022). With this in mind, research into ways to counter misinformation could also be used in the future to combat real world effects such as the World Health Organisation using adverts on Facebook to remind people to wear masks or Facebook itself auto flagging Covid-19 fake news posts with a link to a fact checker. Facebook started its fact-checking programme in December 2016 to rate and review the accuracy of content on Facebook for false news and misleading information (Facebook, 2019). These systems will be improved over the years and studies will look into them to see their effectiveness.

Future research may also examine the potential positive consequences of conspiracy theories and ways to counter misinformation. For example, conspiracy theories and

doubt along society may allow people to challenge existing social knowledge and encourage government transparency as a countermeasure to conspiracy doubt and misinformation spreading (Clarke, 2002; Jolley and Douglas, 2014).

2.9 Tools used for research and analysis

To conduct my analysis on the tweets, I chose Python version 3.9.7 as the programming language as its a very popular and versatile language with a large amount of supporting libraries and developer documentation which is perfect for data science. In combination with this, I used the web based Jupyter Notebooks through Anaconda to conduct, process and visualise the analysis. This is commonly used in data analysis due to the ability of running independent cells of code and showing immediate visualisation to compare and debug code. This can also save time and space as you are not required to run the whole script every time. Another perk of Jupyter notebooks is formatting, you can see a clear documentation of your code and use Jupyter Markdown Notation language to decorate the code with clear headings to segment code. This facilitated data modelling, dataframe wrangling, visualisation and results of models in an easily accessible and visual display.

2.10 Data Collection for this project

The data was provided to me from Cardiff's Crime and Security Research Institute which was extracted from a platform called Sentinel to scrape social media data from Twitter's API (Preece et al., 2017). This data was supplied in a .gz file and ultimately a JSON file was extracted from that. The data was collected over 4 years from 2019-2022, collecting April, May, June each year with the exception of July 2022 as a major heatwave and discussion of climate change occurred in the UK and various other locations in the world. The tweets were collected on a query using a set of search terms focused around misinformation, propaganda and fake news, see Appendix 1 for full list, if the tweet text contained any of these misinformation related terms, they would qualify. This does not mean they are classified as misinformation but the tweet does involve misinformation related terms or have 'called out' misinformation.

The tweets were also queried on the language metadata, attempting to filter out non English tweets. The data used in this project consisted of 248 days of tweets in

individual JSON files, this involved missing data, 8 days from May 2019, 5 days June 2019, 14 days from June 2020. In total, this included 1,068,916,461 'misinformation' tweets that were used in this project.

2.11 Libraries used

Library	Use case
Pandas	Pandas is a fast, powerful, and flexible open-source data analysis and processing tools, it is centred around creating dataframes and being able to manipulate data easily. This makes it very visual and easy to use in the Jupyter Notebook. This was used to manipulate large datasets and save as pickles.
NumPy	NumPy provides support for large, multi-dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on arrays. In my project, it was used to create a custom-built matrix.
Matplotlib, Seaborn	Matplotlib is a comprehensive library for creating static, animated, and interactive visualizations in Python, Seaborn is a library which is built on top of Matplotlib, which provides additional features and customisations. This was used to visualise trends and statistics.
JSON	JavaScript Object Notation is an open standard file format and data interchange format that uses human-readable text to store and transmit data objects, this was used to store the collected data from the Twitter API and then read it in and extract into a dataframe.
Scikit-Learn	Scikit-learn is a machine learning library, It features various classification, regression and clustering algorithms including support-vector machines, random forests, gradient boosting, k-means and DBSCAN. This was used to conduct the TF-IDF model through the TfidfVectorizer and TfidfTransformer.
Natural Language Tool Kit	The Natural Language Toolkit, or more commonly NLTK, is a suite of libraries for natural language processing for primarily the English language. This was used for pre-processing text, stop words, lemmatizing and ngrams.
Itertools	Itertools is a module in Python, it is used to iterate over data structures that can be stepped over using a for-loop. This was used for the combinations method to see the top occurrences for the Twitter hashtags
WordCloud	WordCloud is a module which creates customisation and visually digestible wordclouds in Python. This was used to create wordclouds of the most popular words used in the Tweets.
Regex (Re)	Regular expressions library provides pattern matching operations on strings. This was used to find hashtags and user mentions in the tweets and is commonly used in Natural Language Processing.
Collections	This module implements specialized container datatypes providing alternatives to Python's general purpose built-in containers, dict, list, set, and tuple. This library was only used for the counter feature to create Bag of words and an alternative to ngrams.
Glob	The Glob module is used to search for patterns in file names, this was used to combine all the JSON files and loop through them.
OS	This module provides a portable way of using operating system dependent functionality. This was used to read files.

Figure 2.2: Libraries used

The three most important libraries used in this project were Pandas, Natural Language Tool Kit and Scikit-Learn.

Pandas was crucial for storing data in their personal data structure, the dataframe. Dataframes are a unique way to store, filter and manipulate data easily, Pandas offers a wide range of useful and efficient methods such as storing tables as pickles and reading them then being able to filter using string methods on every object in the dataframe using str.contains(). I also used Pandas functions to iterate through the dataframe using df.iterrows(). Pandas is a fundamental tool for Python users when analysing any data due to it having the ability to efficiently handle large data and create an accessible way to manage data.

Natural Language Toolkit (NLTK) is an essential library when applying Natural Language Processing to a corpus or analysing text with a wide range of methods and functions available, this helped me process Ngrams, lemmatize text, stem words and use an existing list of stop words from their library. This saved me a lot of time and helped me filter the non important tokens and adding my own stop words as I went along. NLTK is the most popular NLP library and get regular updates which makes it reliable and effective at providing suitable functions to analyse at any corpus.

Scikit-Learn (Sklearn) was only used for TfidfVectorizer and TfidTransformer in this project, but this was fundamental during the calculations of the TF-IDF values. To create a TF-IDF matrix in less than 5 lines of code saved me a lot of time and produces effective results quickly. Sklearn is the most useful and robust library for machine learning in Python. It provides a selection of efficient tools for machine learning and statistical modelling including classification, regression, clustering via a consistence interface in Python. In future research, I hope to develop my skills in all these libraries and use machine learning in more complex studies analysing data and creating models.

Chapter 3

Analysis of climate related misinformation tweets from May 2019-June 2022

3.1 Aims of this chapter

In this chapter, I will try to investigate climate related tweets from the misinformation related tweets and compare the statistics, alongside the context of these tweets from May to June 2022 looking at specifically May and June in each year. As these tweets are already filtered by misinformation related terms, I further filtered based on climate terms, these were selected only if the tweet contained any of these terms, "climate", "climate change" or "warming".

I decided to focus on May and June in 2019, 2020 and 2022 in this section and include the 2021 analysis in the appendix, I choose these years to compare before covid (2019), during covid (2020) and after the height of the covid pandemic in 2022.

To achieve these results, I will use exploratory data analysis over all datasets then use various NLP methods to pre-process the tweets. To conclude, I will use NLP models to derive insight and context from the tweets, the methods used will be Ngrams, Word Clouds, popular hashtags, hashtag co-occurrences and TF-IDF.

3.2 Extracting climate data and creating clean dataframes

The misinformation 'callout' datasets were provided via Microsoft Teams from the CSRI, I proceeded to extract the JSON files from the .gz files, this included 217 JSON files split into individual days from the months May and June ranging from 2019-2022, there were 27 missing or unreadable days. These were extracted using WinRAR into a separate folder based on each month, I then created a Jupyter notebook for the purpose of combining the JSON files from each month into into monthly dataframes. This was done using the Python library Glob to use pattern matching and select all files in a specific folder to loop through.

```
files = glob.glob(r'C:\Users\Dissertation work\Data\June_2022\/*')
```

```
# read each line of the files, extract relevant data
final = []
rows = []
for name in files:
    with open(name, "r") as f:
        for line in f:
            try:
                data = json.loads(line)
            except:
                pass

            tweet_id = data["id_str"]

            is_reply = data["in_reply_to_screen_name"]

            is_retweet = data["retweeted"]

            user_name = data["user"]["screen_name"]

            hashtags = []
            for hashtag in data["entities"]["hashtags"]:
                hashtags.append(hashtag["text"])

            if "extended_tweet" in data:
                text = data["extended_tweet"]["full_text"].lower()
            elif data["text"]:
                text = data["text"].lower()

            created = data["created_at_src"]
```

```
30
31              rows.append((tweet_id, hashtags, text, is_reply,
    is_retweet, user_name, created))
32
33          final.extend(rows)
34
35 df = pd.DataFrame(data=final, columns=["tweet_id", "hashtags", "
    tweet_text", "is_reply", "is_retweet", "user_name", "created"])
```

Listing 3.1: Loop for extracting data from JSON

Then I created a loop to extract all the necessary data I needed for my project (see Listing 3.1).

	tweet_id	hashtags	tweet_text	is_reply	is_retweet	user_name	created
613	15209154623762	[]		None	False		Sun May 01 23:57:58 +0000 2022

Figure 3.1: Example of dataframe consisting of monthly misinformation tweets

See Figure 3.2 for the definition and use case for each field extracted. After the fields were extracted into a list of tuples, I created a dataframe to store and access the data, see Figure 3.1 for an example, I have hidden any personal details of the users.

As the monthly tweets ranged from 55,011,182 to 172,276,436, this process took a significant amount of time, when the dataframe was formed, I filtered it by using the dataframe method str.contains() on the tweet text column.

```
climate_df = df[df['tweet_text'].str.contains("climate|climate change|warming")]
```

This cut down the amount of tweets by at least 99% of the original dataset, now ranging from 54,078 to 636,175 now. I then proceeded to save this as a pickle (.pkl) to save space and be able to access the pickle easily for analysis.

```
climate_df.to_pickle(r'C:\Users\Dissertation
↪  work\Data\June_2022\climate_df_June22.pkl')
```

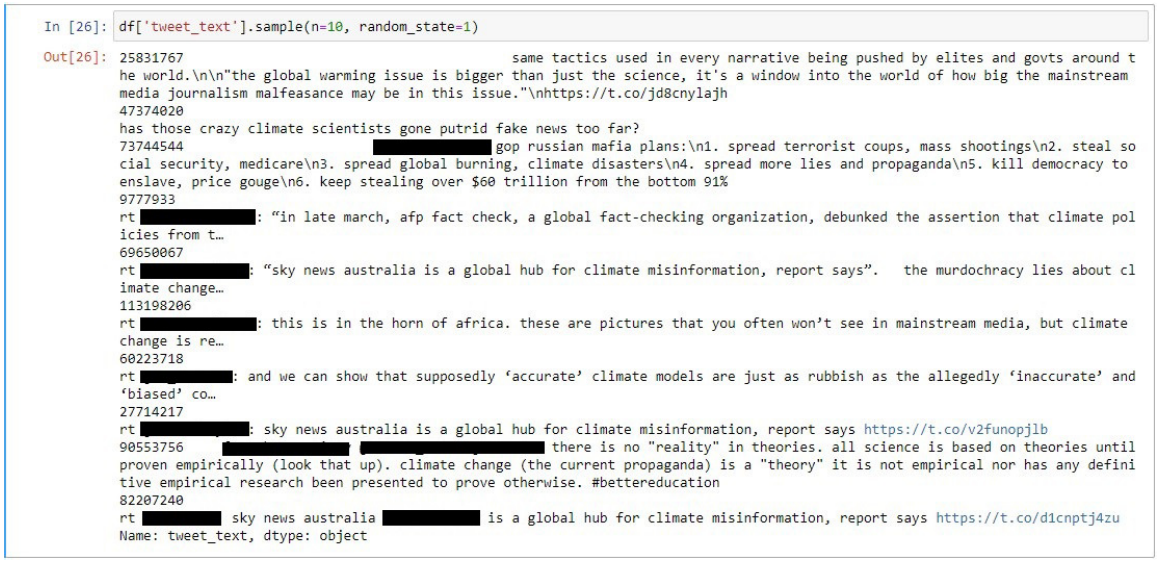
Fields extracted	Use case
tweet_id = data["id_str"]	This is the unique tweet ID and is used as an identifier, this will help me clarify that there are no duplicate tweets.
is_reply = data["in_reply_to_screen_name"]	This field shows if the tweet is replying to any other users, if this is not a reply it will show None, if it is a reply, it shows the username of that person. This was used to identify if users are creating unique posts or replying to other users.
is_retweet = data["retweeted"]	This field identifies if the tweet is a retweet, if it is a retweet, it will show True, if not, it will be False. This was used to identify virality of tweets and if they are being posted uniquely by individual users.
user_name = data["user"]["screen_name"]	This field identifies a user's Twitter username which is associated with the tweet, this was only used to gain context and identify who the most popular tweets came from, for privacy and ethics reasons, usernames won't be shown unless they are a mainstream figure.
hashtags = [] for hashtag in data["entities"]["hashtags"]: hashtags.append(hashtag["text"])	The hashtags field shows all hashtags present in the tweet. This came in the format of a list of lists, so we extracted the list into our dataframe. This was initially extracted to identify top hashtags and hashtag co-occurrence.

if "extended_tweet" in data:
text = data["extended_tweet"]["full_text"].lower()
elif data["text"]:
text = data["text"].lower()
	This field is the tweet text, the most important field for us to study in Natural Language Processing and to identify the context of the time period selected. This was used in the NLP analysis and language models.

created = data["created_at_src"]	This field indicated when the tweet was posted onto social media, this was used to identify exactly what day each tweet was posted and to run statistical tests on how many tweets were posts each day.

Figure 3.2: Fields extracting from JSON files

3.3 Sample tweets before pre-processing



```
In [26]: df['tweet_text'].sample(n=10, random_state=1)

Out[26]: 25831767                                          same tactics used in every narrative being pushed by elites and govts around t
         he world.\n\n"the global warming issue is bigger than just the science, it's a window into the world of how big the mainstream
         media journalism malfeasance may be in this issue."\nhttps://t.co/jd8cnylajh
         47374020
         has those crazy climate scientists gone putrid fake news too far?
         73744544                          [redacted] gop russian mafia plans:\n1. spread terrorist coups, mass shootings\n2. steal so
         cial security, medicare\n3. spread global burning, climate disasters\n4. spread more lies and propaganda\n5. kill democracy to
         enslave, price gouge\n6. keep stealing over $60 trillion from the bottom 91%
         9777933
         rt [redacted]: “in late march, afp fact check, a global fact-checking organization, debunked the assertion that climate pol
         icies from t…
         69650067
         rt [redacted]: “sky news australia is a global hub for climate misinformation, report says”.   the murdochracy lies about cl
         imate change…
         113198206
         rt [redacted]: this is in the horn of africa. these are pictures that you often won’t see in mainstream media, but climate
         change is re…
         60223718
         rt [redacted]: and we can show that supposedly ‘accurate’ climate models are just as rubbish as the allegedly ‘inaccurate’ and
         ‘biased’ co…
         27714217
         rt [redacted]: sky news australia is a global hub for climate misinformation, report says https://t.co/v2funopjlb
         90553756    [redacted] [redacted] there is no "reality" in theories. all science is based on theories until
         proven empirically (look that up). climate change (the current propaganda) is a "theory" it is not empirical nor has any defini
         tive empirical research been presented to prove otherwise. #bettereducation
         82207240
         rt [redacted] sky news australia [redacted] is a global hub for climate misinformation, report says https://t.co/d1cnptj4zu
         Name: tweet_text, dtype: object
```

Figure 3.3: Sample tweets from climate dataframe with usernames hidden

Figure 3.3 is an example of how the tweets were formatting in the climate dataframe before pre-processing using the Pandas library sample method to select 10 random tweets.

3.4 Exploratory data analysis

Initially I wanted to visualise the average amount of climate tweets per day to compare the sheer amount of volume over the last 4 years of May and June, July 2022 was added as an extra, this was because a history breaking temperature record was achieved in July 2022 in the UK and this was a recent event that created discussion. To create this bar plot, I loaded the individual monthly pickles and used the "created" column to slice the day off to create its own column called "day".

```
df1["day"] = df1["created"].str[0:10]
```

This turned "Sun May 01 23:57:58 +0000 2022" into "Sun May 01" as these were already sorted into individual months. Then I used this to count how many tweets were on each day and calculate the average as there was missing data from May 2019, June 2019 and June 2020, this was the best way to visualise a fair comparison. This plot was created using the Python library Seaborn.

```
avg_daily = sns.barplot(x=months, y=avg_tweets)
```

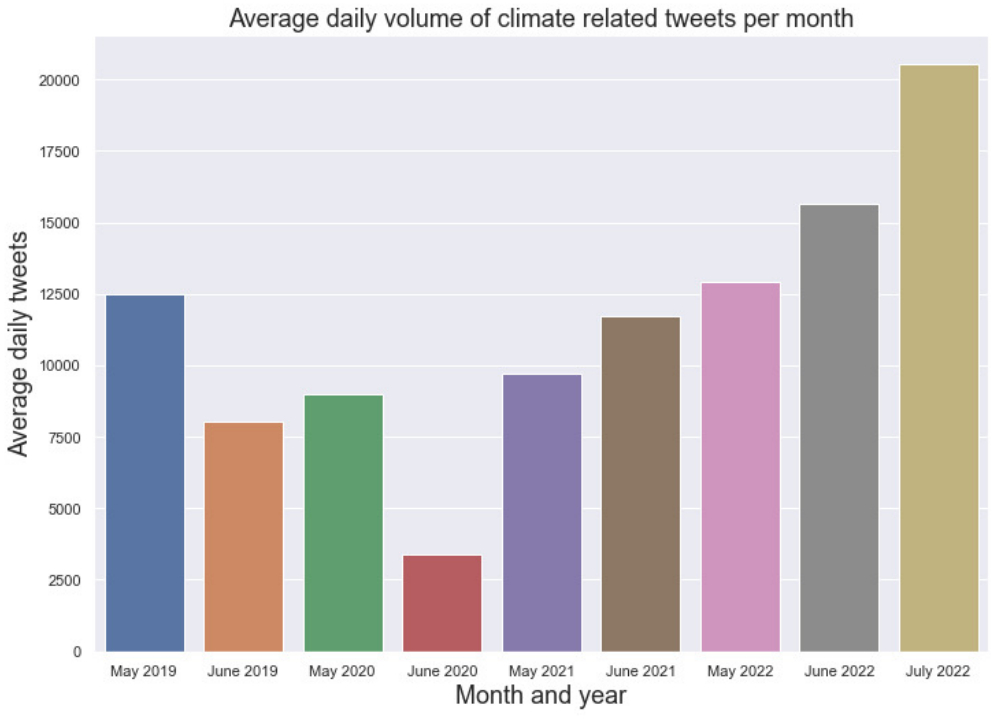


Figure 3.4: Average amount of climate related tweets daily per month 2019-2022

As we can see from Figure 3.3, climate misinformation related tweets dropped in June 2019 to an average 8,015 tweets per day, recovered a little in May 2020 at 8,975 average tweets per day but took a huge hit in June 2020 at 3,380 average tweets. May 2020 was the first month after Covid-19 lockdowns and the rise in climate tweets could be due to media discussing ecosystems regenerating due to lack of human pollution or it could be due to the news reporting that May 2020 was the hottest month in history at that time (Newburger, 2020). Emma Newburger reporting from CNBC wrote that in the last 12 month period, from June 2019 to May 2020, temperatures were nearly 0.7 degrees Celsius warmer than average. Globally, May was 0.63 degrees Celsius warmer than the average May recorded from 1981 to 2010. This could also be due to the increased amount of people using social media at this time due to users being indoors with the Covid-19 pandemic lockdowns.

After the drop in June 2020, primarily due to assumed Covid-19 outrage and misinformation over lockdowns and vaccinations, it was a consistent upwards climb peaking at June 2022 with 15,630 tweets per day from the original May/June comparison or alternatively 20,522 in the July 2022 comparison. As Covid-19 slowed down in discussion, the next big issue was climate change which was always discussed but overshadowed by more immediate threats. Climate disasters have occurred nearly every year but media did not cover these events as frequently during the height of the pandemic. Floods, wildfires and record breaking heats all around the world have been reported in the last 4 years, one of the biggest events happened in early September 2019 until March 2020, Australia had one of the worst bush fire seasons in its recorded history. The bushfires burned more than 46 million acres (72,000 square miles). At least 3,500 homes and thousands of other buildings were lost and 34 people died in the thousands of fires (CDP, 2020).

I also wanted to compare the percentage of the misinformation callout dataset that is climate related to give a more fair and proportional comparison to each of the months, this would help eliminate the bias if there were generally a lot more tweets collected that month, the average amount of climate tweets would also go up.

Figure 3.4 shows us a more representative comparison and a different narrative to investigate, we can identify a clear picture that climate related tweets were a higher proportion of the misinformation related tweets in 2019 but drop significantly as Covid-19 starts. I believe this supports my general prediction that Covid-19 misinformation overshadows climate discussion for the year of 2020, which was the height

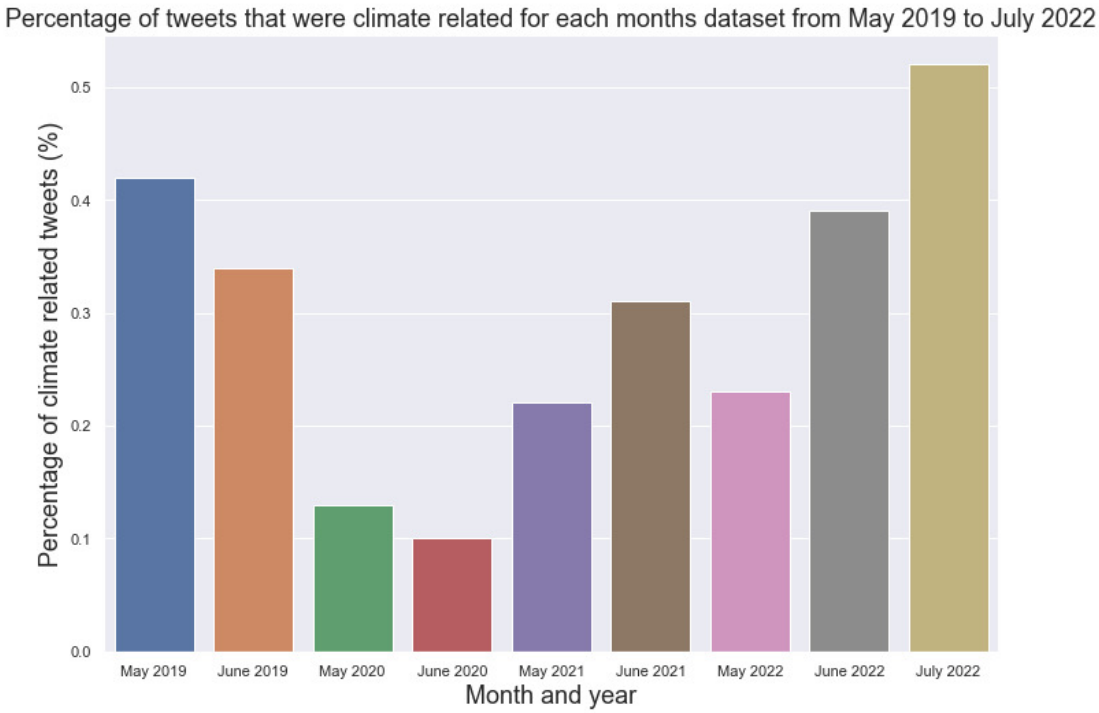


Figure 3.5: Percentage of climate related tweets from the callout dataset

of the pandemic and instability across the world. In May 2019 climate related tweets were 0.42% of the misinformation dataset and in June 2019 it fell to 0.34%, these were before the first Covid-19 case. After the first Covid-19 case which happened in December 2019, the climate discussion fell to a mere 0.13% in May 2020 and just 0.10% in June 2020. As covid was slowly becoming less severe to the healthy population and vaccines were released, climate discussion started to rise, it wasn't until June 2022 that it hit 0.39%, near the same levels as May 2019 and proceeded to hit 0.52%, the highest recorded during the heatwave month.

Figure 3.5 shows general exploratory data and statistics that I retrieved using the Twitter API, Regex and Pandas. When investigating the retweets, I noticed that my dataframe column "is_retweet" was completely full with "False" values, this may be due to an error in my extraction or an error with the Twitter API. Regardless of this set back, I decided to use Regex to find retweets instead, the Twitter API adds "RT" at the start of every retweet, so I created a regex expression and counted the amount of matches it found.

To make sure the regex match did not pick up any tweets that were not retweets, I configured the regex to match the start of every retweet with a specific pattern.

```
retweet = re.match(r"^rt @+", row.tweet_text)
```

The replies column extracted from the Twitter API did provide insightful results, I calculated the amount of replies by finding the sum of the "None" values and subtracting that away from the total amount of climate tweets.

To calculate the unique tweets, I used a set, a data structure which only stores unique values. All the tweets were collected in lower case so I compiled a set of all the tweets and counted them.

For the original tweets, I simply subtracted retweets from the overall amount of climate tweets, this was a metric I created to indicate all tweets that were posted by a user manually and wasn't just a retweet. This could indicate that it was original content created by the user or copy pasted from another source.

Retweets stayed fairly consistent with the exception of May 2019 with 71% of the tweets being retweets, without this month, retweets ranged from 44%-58%. This makes the original tweets, by far the lowest at only 29% of climate tweets being

2019-2022	May 2019	June 2019	May 2020	June 2020	May 2021	June 2021	May 2022	June 2022	July 2022
Total number of climate tweets	287,660	200,368	278,239	54,078	301,631	351,511	400,491	468,904	636,175
Original tweets (not retweets)	83,254	83,313	143,452	25,664	168,648	157,216	210,016	242,244	335,374
Unique tweets	7,935	7,258	9,345	3,723	11,168	11,478	13,985	16,256	25,708
Retweets	204,406	117,055	134,787	28,414	132,983	194,295	190,475	226,660	300,801
Replies	247,470	159,904	180,787	36,536	192,962	245,721	261,395	321,259	404,981
Total Tweets in callout dataset	68,421,942	59,149,395	219,123,503	55,011,182	136,265,329	115,067,784	172,276,436	120,946,752	122,654,138
Percentage of climate tweets in callout	0.42%	0.34%	0.13%	0.10%	0.22%	0.31%	0.23%	0.39%	0.52%
2019-2022 (%)	**May 2019**	**June 2019**	**May 2020**	**June 2020**	**May 2021**	**June 2021**	**May 2022**	**June 2022**	**July 2022**
Original tweets (not retweets)	29%	42%	52%	47%	56%	45%	52%	52%	53%
Unique tweets	3%	4%	3%	7%	4%	3%	3%	3%	4%
Retweets	71%	58%	48%	53%	44%	55%	48%	48%	47%
Replies	86%	80%	65%	68%	64%	70%	65%	69%	64%
Highest volume of tweets on one day	May 4th	June 6th	May 6th	June 7th	May 18th	June 2nd	May 1st	June 14th	July 18th

Figure 3.6: General stats showing the volume and % of retweets, replies etc

original tweets in May 2019. This might indicate a high virality in the tweets which encourages users to retweet and reply to certain popular figures. May 2019 consistently stood out as a record breaking month compared to the other 8 months, with the highest amount of replies being 86% of the tweets in the dataset, this concludes that only 14% of the tweets were not in response to another member. May 2019 also has the lowest amount of unique tweets at 3%, on par with May 2020, June 2021, May 2022 and June 2022. June 2020 stands out as the largest amount of unique tweets that month at 7% of the dataset being unique, this indicates more diverse narratives and users tweeting their opinions.

May 2020 was the first month in my data that was after Covid-19 lockdowns started, we can clearly see a huge increase in misinformation related tweets but a low amount of climate change tweets. June 2020 was the lowest lowest percentage of climate tweets in relation to misinformation, it was also by far the lowest amount of volume, this was because there were 2 weeks of missing data in this month. This should not majorly effect percentages and other results but is something to note. When compared to May 2020, the results are similar.

The 3 months in 2022 also have very similar results which would be expected, the most significant change is the percentage of climate tweets in the misinformation dataset going from 0.23% to 0.52%, with a minor 1% decrease in retweets.

3.5 Tweet pre-processing

While extracting tweet text, I used the .lower() method on all the strings, this was used to filter the dataframes and catch both upper and lower states of the words, this was also the first step of pre-processing the tweets for Natural Language Processing analysis.

After reading the pickle through Pandas, I split all the tweet text into tokens and lemmatized them using functions from the NLTK library inside a function that was inspired by a stack overflow user (Foz, 2021). Initially I was planning to use Parts of Speech (POS) to analyse tweet structures and grammar but after planning my objectives, I decided to focus on other methods. I also removed all usernames before applying these functions, I once again used Regex to search for any words that came after the "@" sign as this is the way to mention users on Twitter, all usernames were removed for privacy and ethic reasons.

```
climate_strings = re.sub(r'@([A-Za-z0-9_]+)','', climate_strings)
```

I then proceeded to load up stop words using NLTK corpus and then added my personal stop words that I noticed were commonly appearing in a bag of words model I used for testing on other tweets.

```
from nltk.corpus import stopwords.
stop_words = stopwords.words('english')
newStopWords = ['RT','I','T','S','U','http','co','s','n','u','p','amp','rt']
stop_words.extend(newStopWords)
```

For example, 'amp' came from the ampersand sign & when it gets converted into the JSON format.

I used a list comprehension to filter out any stop words in the tweets and applied the preprocess_text function over the list of tokens (see Listing 3.2).

```
from nltk.stem import WordNetLemmatizer
from nltk.tokenize import word_tokenize
from nltk.corpus import wordnet

normalizer = WordNetLemmatizer()

def preprocess_text(text):
    cleaned = re.sub(r'\W+', ' ', text).lower()
    tokenized = word_tokenize(cleaned)
    normalized = " ".join([normalizer.lemmatize(token,
   get_part_of_speech(token)) for token in tokenized])
    return normalized

def get_part_of_speech(word):
    probable_part_of_speech = wordnet.synsets(word)
    pos_counts = Counter()
    pos_counts["n"] = len(  [ item for item in
   probable_part_of_speech if item.pos()=="n"]  )
    pos_counts["v"] = len(  [ item for item in
   probable_part_of_speech if item.pos()=="v"]  )
    pos_counts["a"] = len(  [ item for item in
   probable_part_of_speech if item.pos()=="a"]  )
    pos_counts["r"] = len(  [ item for item in
   probable_part_of_speech if item.pos()=="r"]  )
    most_likely_part_of_speech = pos_counts.most_common(1)[0][0]
    return most_likely_part_of_speech
```

Listing 3.2: Pre-processing functions

Initially I also used the PorterStemmer from nltk.stem.porter, but after using word clouds and Ngrams on these tokens, it removed some of the context and meaning to the words so I reverted this. After applying preprocess_text on the tokens, the tweets were a list of cleaned tokens, ready to be analysed using NLP methods.

3.6 Ngrams

In this section, I will discuss the results from applying unigrams, bigrams and trigrams on each of the 8 months from May 2019-June 2022.

Ngrams are simply the most common occurrence of the phrase consisting of any

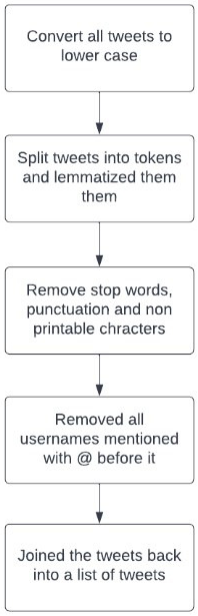


Figure 3.7: Process of cleaning tokens

number of words, for example, unigrams are the most common phrase consisting of 1 word, bigrams are the most common 2 consecutive words and trigrams are 3. From investigating these phrases and comparing each set of ngrams, we can begin derive context and patterns from what was commonly being discussed at that current time.

Initially I used the method Counter from the python library Collections to calculate unigrams or Bag of Words model (BOW) but as I progressed onto bigrams, I found the ngrams function from nltk.util a lot more useful and versatile as it can be scaled up easily by changing the value of N. This was used in conjunction with the Counter method to create lists of ngrams.

```
n = 3

trigrams = ngrams(processed_climate, n)

ngrams_climate = Counter(unigrams)
ngrams_climate.most_common(10)
```

I will be discussing the ngrams of the most important months to my narrative, which are 2019 (before covid), 2020 (after and during the height of covid) and finally 2022 (as

covid slows down and becomes less discussed). The ngrams for 2021 will be included in Appendix 2.

Unigrams (BOW)	May 2019	June 2019
1st	climate,247807	climate, 166406
2nd	change, 107120	change, 93384
3rd	propaganda, 94792	fake, 72300
4th	fake, 86441	news, 70649
5th	news, 67264	propaganda, 51588
6th	hoax, 55498	warming, 31692
7th	science, 41705	global, 31400
8th	crisis, 37776	trump, 24430
9th	lie, 35599	terrorism, 22121
10th	fraud, 33019	medium, 20693

Bigrams	May 2019	June 2019
1st	('climate', 'change'), 104150	('climate', 'change'), 91129
2nd	('fake', 'news'), 58087	('fake', 'news'), 66660
3rd	('climate', 'crisis'), 33208	('global', 'warming'), 28872
4th	('hoax', 'fraud'), 26702	('mainstream', 'medium'), 17091
5th	('crisis', 'lie'), 26662	('great', 'threat'), 11943
6th	('science', 'logic'), 26648	('see', 'great'), 11860
7th	('lie', 'hoax'), 26641	('news', 'mainstream'), 11805
8th	('fraud', 'affront'), 26634	('medium', 'see'), 11759
9th	('affront', 'science'), 26634	('threat', 'global'), 11713
10th	('logic', 'travesty'), 26634	('warming', 'terrorism'), 11671

	May 2019	June 2019
1st	('climate', 'crisis', 'lie'), 26656	('see', 'great', 'threat'), 11860
2nd	('lie', 'hoax', 'fraud'), 26639	('fake', 'news', 'mainstream'), 11805
3rd	('crisis', 'lie', 'hoax'), 26636	('news', 'mainstream', 'medium'), 11805
4th	('hoax', 'fraud', 'affront'), 26634	('mainstream', 'medium', 'see'), 11738
5th	('fraud', 'affront', 'science'), 26634	('medium', 'see', 'great'), 11738
6th	('affront', 'science', 'logic'), 26634	('great', 'threat', 'global'), 11713
7th	('science', 'logic', 'travesty'), 26634	('threat', 'global', 'warming'), 11713
8th	('logic', 'travesty', 'economic'), 26585	('global', 'warming', 'terrorism'), 11671
9th	('travesty', 'economic', 'social'), 26561	('sean', 'hannity', 'destroy'), 11654
10th	('economic', 'social', 'sinkho'), 26266	('hannity', 'destroy', 'fake'), 11654

Figure 3.8: Ngrams for 2019

As I expected, common words and phrases that were used to filter the dataset reside at the top of unigrams and bigrams such as "climate", "change", "propaganda", "fake", "news, "hoax". We can then see other stories emerge such as "science", "crisis",

"global", "lies", "terrorism", "trump" and "medium". Medium would have been the lemmatized version of media as mainstream media is commonly discussed around misinformation and was a selected term for the misinformation callout dataset.

May 2019 looks mostly generic with the resulting ngrams but focusing on a climate crisis and the defrauding of science, possibly accusing users of creating a climate change hoax and denying the science involved, calling it a lie. There is also another theme appearing in the trigrams involving an economic aspect in combination with the social issues. The tweet that gathered a lot of attention mentioned that the 'climate crisis' was a lie, a hoax, a fraud and an affront to science and logic, he then proceeded to call it a travesty, an economic and social sinkhole. This tweet gathered over 26,000 retweets and dominated the narrative for May 2019.

Looking into June 2019, we can see a lot less unigrams mentioning "climate" but in bigrams, "climate change" is still a popular ngram, this is due to people using climate paired with other words in the same environmental space, such as "climate propaganda", "climate deniers" etc. There were also some cases which were not talking about the environmental climate, but discussing a "climate of fear" or the economic climate, but these were the minority.

Unique stories emerge in June 2019 which involve themes that are not usually directly related to climate change such as "terrorism" and "trump" alongside highly correlated associations of climate change like "global warming". Trump appearing in the unigrams was caused by his controversial take where he quoted a tweet from an apparent co-founder of Greenpeace. Patrick Moore was an ex president of Greenpeace Canada who was voted out in 1986 but appeared on Fox and friends, a popular morning news show in the USA, as a 'Co-founder of Greenpeace' (BBC, 2019). On the show, Patrick declared that the climate crisis was fake news and later Trump tweeted, quoting Patrick, as Trump was the current president, it reached a huge audience and created discussion over misinformation.

Greenpeace later corrected this by tweeting that Patrick Moore was not a co-founder of Greenpeace and that he does not represent Greenpeace, they stated he is a paid lobbyist spreading misinformation (BBC, 2019).

The final theme that completes the picture in June 2019 is again associated with Fox News, this time it is in regards to Sean Hannity who hosted a commentary program, Hannity, on Fox News. The tweet which gathered attention discussed Hannity de-

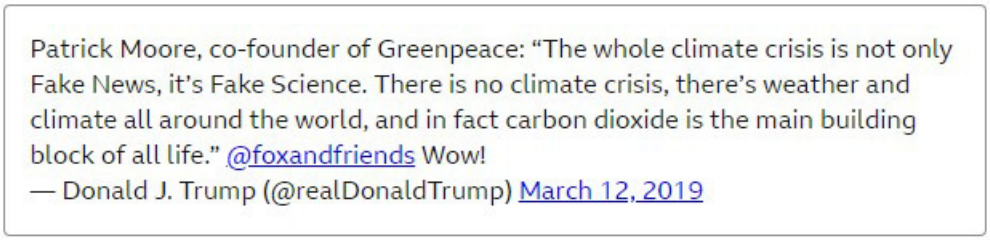


Figure 3.9: Trump tweet in 2019

stroying fake news media and declaring that fake news had become more of a threat than terrorism and global warming. As research has shown, misinformation has become a major problem, although these themes are contradictory to each other as they are both hosted by the Fox network which also spread false news during the same month.

In May 2020 (see figure 3.9), the ngrams display a familiar picture with the top unigrams being all the search terms used for the misinformation dataset or the filtered climate terms. Bigrams and trigrams are the biggest difference to other months, we can identify that the dominating discussion in May 2020 was Michael Moore's documentary called "Planet of the humans", this was a documentary focused on climate change and the reality of green energy, with an emphasis on exposing misconceptions (Sky News AU, 2020). The documentary was released on April 21st, 2020 and received a lot of controversy and discussion on Twitter.

Climate scientists branded this documentary as dangerous and misleading, Michael Mann, a leading climate scientist tweeted "Michael Moore is now promoting the very same agenda of climate inaction that is being pursued by the fossil fuel beholden Trump administration and Vladimir Putin." (Sky News, 2020)

The only exception to this narrative are the 2 trigrams referring to "climate", "propaganda", "like" and "look", "like", "5xejcsnc2y". These are part of a tweet which just included a link to what they accused of being climate propaganda, as this tweet was from 2020, the link has now been deleted off the platform.

Unigrams (BOW)	May 2020	June 2020
1st	climate, 198097	climate, 41242
2nd	change, 82940	propaganda, 28564
3rd	propaganda, 79116	call, 12189
4th	conspiracy, 62869	year, 12004
5th	news, 44745	political, 11565
6th	like, 36287	child, 11078
7th	warming, 35459	old, 10921
8th	fake, 34477	brainwash, 10823
9th	global, 33019	plan, 10811
10th	misinformation, 32340	leftist, 10772

Bigrams	May 2020	June 2020
1st	('climate', 'change'), 78582	('call', 'climate'), 11107
2nd	('fake', 'news'), 29595	('political', 'propaganda'), 10666
3rd	('global', 'warming'), 29168	('year', 'old'), 10616
4th	('climate', 'propaganda'), 19709	('propaganda', 'call'), 10563
5th	('conspiracy', 'theory'), 19593	('old', 'child'), 10558
6th	('look', 'like'), 18815	('five', 'year'), 10553
7th	('propaganda', 'look'), 18047	('abuse', 'nj'), 10549
8th	('conspiracy', 'theorist'), 13285	('nj', 'leftist'), 10549
9th	('michael', 'moore'), 10645	('leftist', 'plan'), 10549
10th	('mainstream', 'medium'), 9667	('plan', 'brainwash'), 10549

Trigrams	May 2020	June 2020
1st	('propaganda', 'look', 'like'), 17969	('propaganda', 'call', 'climate'), 10563
2nd	('climate', 'propaganda', 'look'), 17905	('abuse', 'nj', 'leftist'), 10549
3rd	('look', 'like', '5xejcsnc2y'), 10047	('nj', 'leftist', 'plan'), 10549
4th	('climate', 'change', 'denier'), 8784	('leftist', 'plan', 'brainwash'), 10549
5th	('michael', 'moore', 'documentary'), 8020	('plan', 'brainwash', 'traumatize'), 10549
6th	('documentary', 'planet', 'human'), 7979	('brainwash', 'traumatize', 'five'), 10549
7th	('expose', 'swindler', 'peddling'), 7961	('traumatize', 'five', 'year'), 10549
8th	('swindler', 'peddling', 'misinformation'), 7961	('five', 'year', 'old'), 10549
9th	('moore', 'documentary', 'planet'), 7955	('year', 'old', 'child'), 10549
10th	('planet', 'human', 'expose'), 7955	('old', 'child', 'political'), 10549

Figure 3.10: Ngrams for 2020

In June 2020, many unique themes emerge such as "political", "brainwash", "leftist". In bigrams and trigrams, we consistently find a variation of "five year old child" appear, which is a very strange narrative without context. After some investigating, I traced this tweet back to an American conservative public figure named Tom Fitton that dominated the narrative in June 2020. His outage lead him to tweeting about a news story from CNN (see figure 3.10).

Tom Fitton is the president of Judicial watch, an activist group that files Freedom of Information Act lawsuits to investigate misconduct by governmental officials with a specific bias to target Democratic presidents such as Bill Clinton and Barack Obama. At the time of writing this, he had 1.5 million followers and as we can see from the

ngrams being covered in his narrative in June 2020.

Figure 3.10 suggests only 537 retweets and 79 quote tweets but as the trigrams show, this reached a much larger audience. Analysing the words Tom uses in the tweet such as "brainwash", "political propaganda" and "so called climate change", we can clearly see he believes that climate change is a hoax, so he is spreading misinformation to thousands, potentially 100s of thousands would have seen this tweet.

Figure 3.11: Tweet from Tom Fitton

Unigrams (BOW)	May 2022	June 2022
1st	climate, 297465	climate, 433439
2nd	propaganda, 133891	change, 138860
3rd	change, 118528	propaganda, 119237
4th	disinformation, 76734	misinformation, 109583
5th	misinformation, 55083	news, 87146
6th	conspiracy, 54499	disinformation, 80246
7th	global, 47099	global, 78127
8th	spread, 46929	conspiracy, 68762
9th	warming, 42631	report, 56424
10th	medium, 42564	say, 56314

Bigrams	May 2022	June 2022
1st	('climate', 'change'), 109776	('climate', 'change'), 131822
2nd	('global', 'warming'), 38095	('climate', 'misinformation'), 51516
3rd	('propaganda', 'abc'), 30843	('sky', 'news'), 51025
4th	('happy', 'spread'), 30692	('news', 'australia'), 47279
5th	('correct', 'propaganda'), 30686	('fact', 'check'), 32391
6th	('abc', 'happy'), 30662	('hub', 'climate'), 29182
7th	('climate', 'disinformation'), 24877	('global', 'hub'), 28673
8th	('conspiracy', 'theory'), 23228	('fake', 'news'), 27784
9th	('extreme', 'weather'), 20264	('australia', 'global'), 27221
10th	('false', 'claim'), 20230	('conspiracy', 'theory'), 26230

Trigrams	May 2022	June 2022
1st	('correct', 'propaganda', 'abc'), 30662	('sky', 'news', 'australia'), 47212
2nd	('propaganda', 'abc', 'happy'), 30662	('hub', 'climate', 'misinformation'), 29074
3rd	('abc', 'happy', 'spread'), 30662	('global', 'hub', 'climate'), 28468
4th	('global', 'warming', 'make'), 19980	('australia', 'global', 'hub'), 27207
5th	('bbc', 'panorama', 'documentary'), 19868	('news', 'australia', 'global'), 27203
6th	('panorama', 'documentary', 'global'), 19868	('climate', 'misinformation', 'report'), 25025
7th	('documentary', 'global', 'warming'), 19868	('misinformation', 'report', 'say'), 24973
8th	('warming', 'make', 'number'), 19868	('central', 'source', 'climate'), 18760
9th	('make', 'number', 'false'), 19868	('misinformation', 'around', 'world'), 18413
10th	('number', 'false', 'claim'), 19868	('murdoch', 'sky', 'news'), 16967

Figure 3.12: Ngrams for 2022

In May 2022, we can see the common defined search terms with the exception of one unigram "spread", this would be assumed as the spread of misinformation or fake news. After further inspection into bigrams and trigrams, we can identify ABC and BBC, two popular news channel along with "extreme weather", "BBC panorama documentary". BBC Panorama global warming documentary was released in November 2021 but resurfaced in May 2022 as media claimed that there were numerous false claims in the documentary. Panorama declared "the death toll is rising around the world and the forecast is that worse is to come" but this has been proven wrong in a recent report from the World Meteorological Organisation that while the number of weather-related disasters, such as floods, storms and droughts have risen in the past 50 years, the number of deaths caused by them has fallen because of improved early warnings and disaster management (Fernandez 2022).

Another theme targeted ABC, Australian Broadcasting Company for spreading propaganda, these tweets weren't directly associated with climate change but did involve a climate communal crowd funding group called Climate200 who support political candidates to create better climate policies. The tweet that was met with backlash involves Climate200's convenor Simon Holmes à Court and a rather immature tweet from a co-presenter at ABC calling Simon a "strange cat".

From June 2022's unigrams, we can see a similar story to the previous months, with two unique phrases "report" and "say". Another news story dictates the narrative for June 2022, Sky News Australia and the owner Rupert Murdoch have been under scrutiny for continuously broadcasting misinformation ranging from climate, Covid-19 and hate speech against multiple groups. In an article by the Guardian, analysts found that Murdoch-owned channel creates and distributes content promoting climate scepticism across the world and was consistently ranked highly for traction, pushing the partisan views of its hosts and guests through social media. (Readfearn, 2022).

The report conducted by UK thinktank the Institute for Strategic Dialogue said "A failure to stem "mis- and disinformation online had allowed junk science, climate delayism and attacks on high-profile individuals working on the climate crisis to become mainstreamed" (Readfearn, 2022). While investigating this theme, I also noticed the tweets were usually accompanied by a hashtag named #boycottMurdoch as users try to call upon the owner of Sky News Australia to direct better content being broadcasted to the national audience.

3.7 Word clouds

Word clouds are a visual way to present the most common phrases that appear in a corpus, I created the word clouds using the Wordcloud module in Python. This was an extremely accessible library to use, within 2 lines of code, it was created and saved as a .png.

```
from wordcloud import WordCloud
wordcloud = WordCloud(width = 1000, height = 500,).generate(climate_corpus)
wordcloud.to_file("word_cloud.png")
```

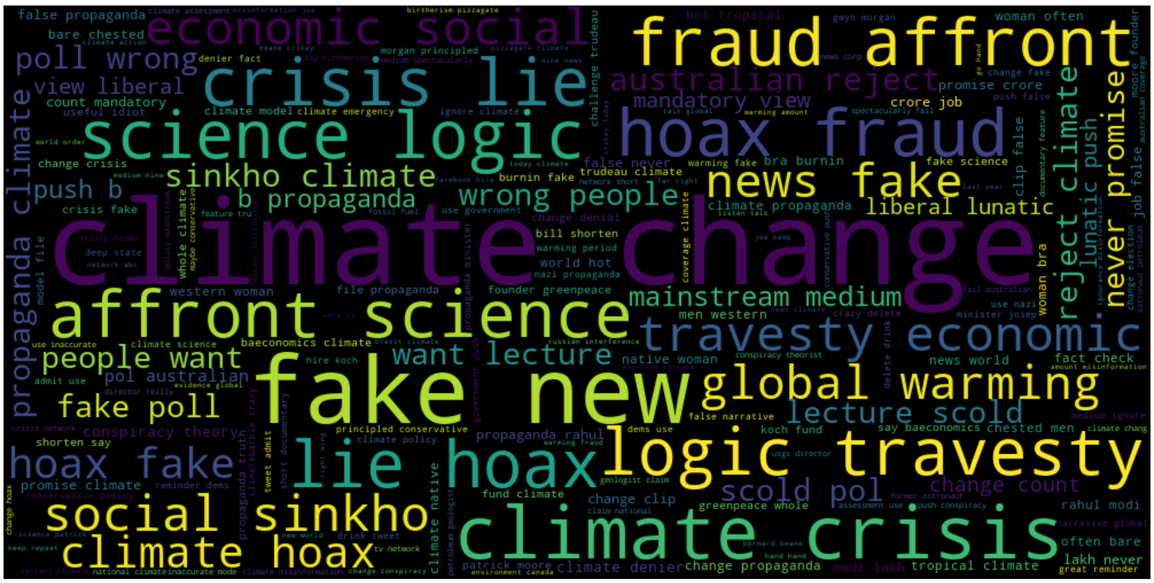


Figure 3.13: Word cloud for May 2019

The word clouds will inherently be very similar to the ngrams findings, with some themes that did not make the top 10's such as "Australian reject" in May 2019 and "President Trump" and "Bill Nye" in June 2020. As the ngrams most popular themes have already been investigated, I will display 2019, 2020 and 2022 word clouds in the main body and include 2021 word clouds in Appendix 3.

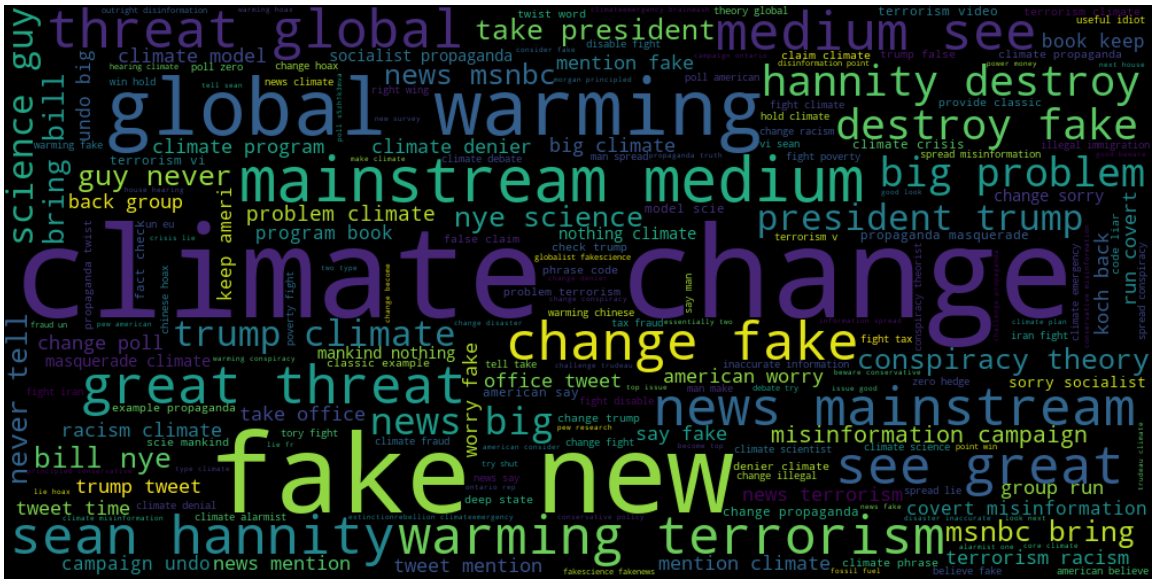


Figure 3.14: Word cloud for June 2019

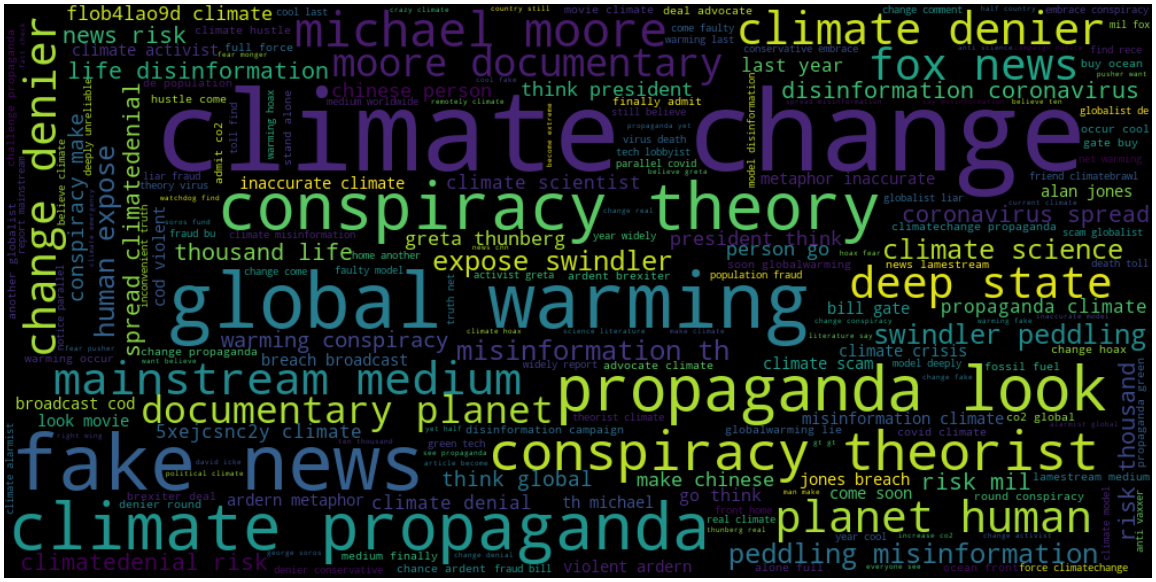


Figure 3.15: Word cloud for May 2020

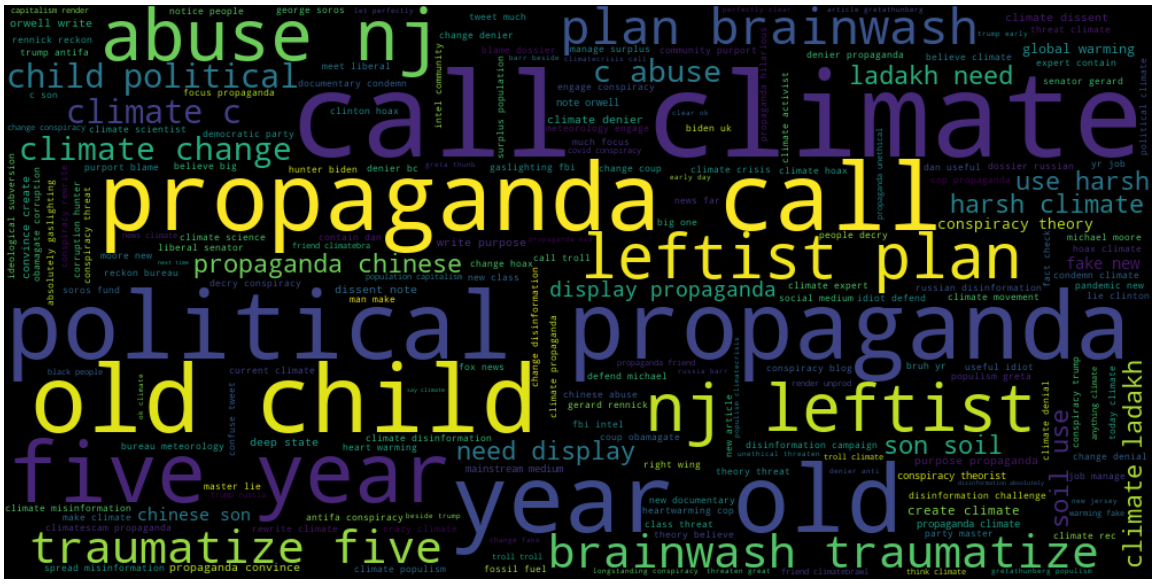


Figure 3.16: Word cloud for June 2020

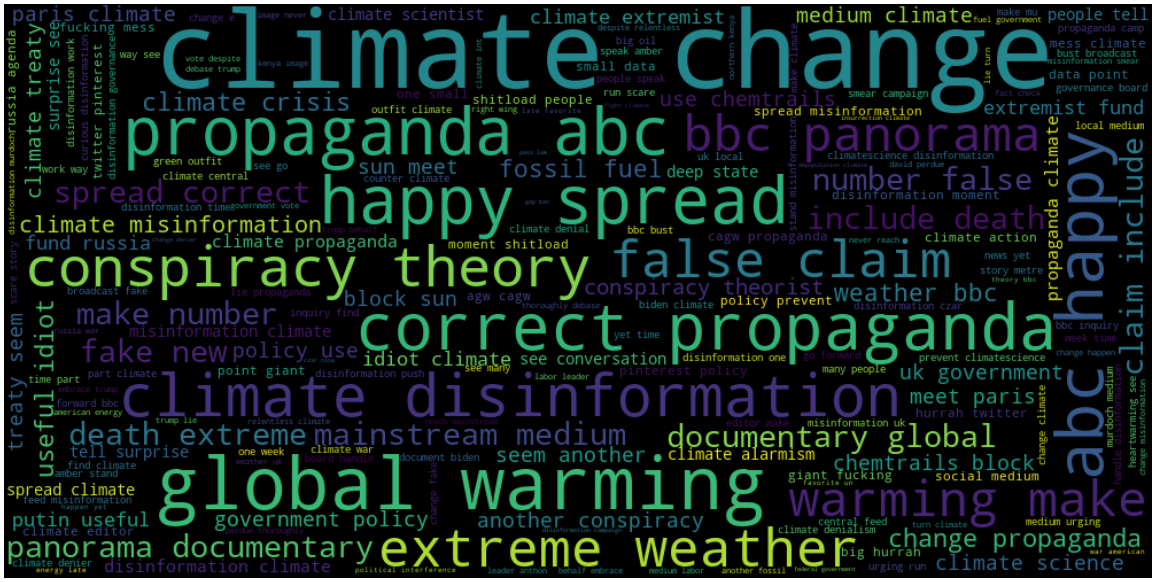


Figure 3.17: Word cloud for May 2022

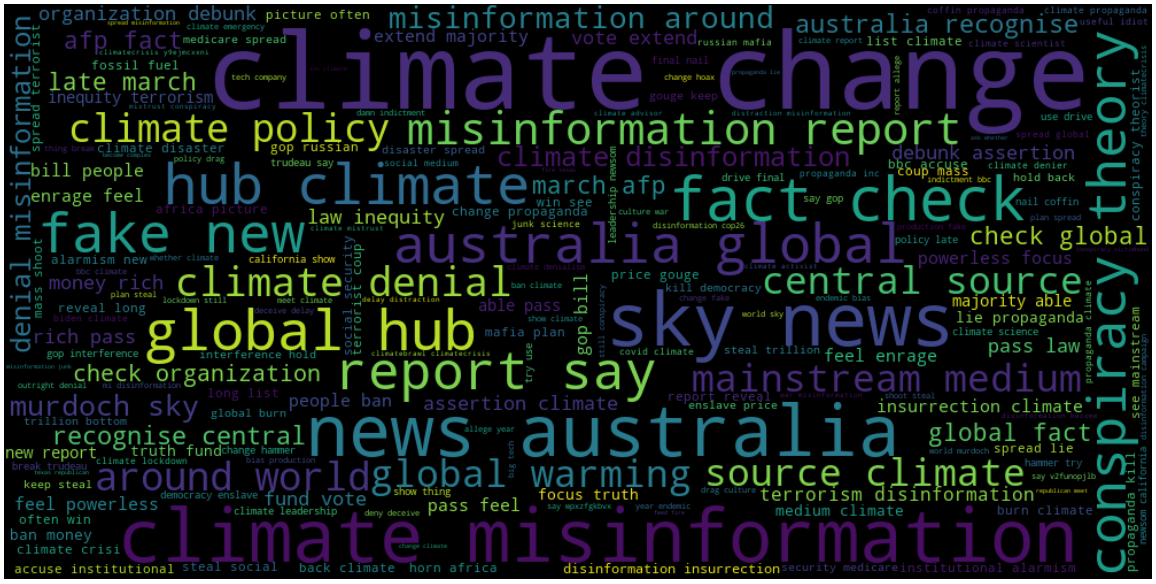


Figure 3.18: Word cloud for June 2022

3.8 Most popular hashtags

To find all hashtags in tweets, I used Regex to find any patterns that came after the # symbol, I used this method as it seemed more reliable than Twitters API hashtag field, I commonly saw words not being picked up by the hashtag field in the dataframe and I wanted to scrape every hashtag regardless of position in the tweet.

```
regex = r"(?<!RT\s)#\w+"

hashtag_list = []
for index, row in df.iterrows():
        htags = re.findall(regex, row.tweet_text.lower())
        hashtag_list.append(htags)

hashtags_refined = []
for _ in hashtag_list:
    if _ != []:
        hashtags_refined.extend(_)

count_htags = Counter(hashtags_refined)
count_htags.most_common(10)
```

Listing 3.3: Process of creating and counting common hashtags

I used hashtags_refined to remove all the empty lists and clean up hashtag_list, then

used the Counter object and most_common method to find the top 10 hashtags for each month.

Top hashtags	May 2019	June 2019
1st	('#climatechange', 10360),	('#climatechange', 7212),
2nd	('#climatebarbie', 3192),	('#climate', 4254),
3rd	('#climateemergency', 2460),	('#fakenews', 2572),
4th	('#auspol', 1540),	('#climateemergency', 2447),
5th	('#propaganda', 1495),	('#propaganda', 1879),
6th	('#climate', 1489),	('#eu', 1841),
7th	('#populism', 1423),	('#extinctionrebellion', 1756),
8th	('#globalwarming', 1329),	('#un', 1603),
9th	('#climatebrawl', 1261),	('#fakescience', 1582),
10th	('#cdnpoli', 1171)	('#globalist', 1567)

Figure 3.19: Top hashtags for 2019

From the top hashtags of May 2019, they show a very different narrative compared to the ngrams, #climatebarbie is a unique theme which refers to Catherine Mary McKenna, former politician who served as a Cabinet minister as a member of the Liberal Party and minister of environment and climate change from 2015 to 2019. These hashtags were majority verbally malicious attacks at McKenna and Bill Nye accusing them of liberal propaganda surrounding climate science. The hashtag itself is dehumanising and promotes gender discrimination which results in distasteful tweets spreading misinformation and attacking McKenna's character.

There are more general themes that could have links such as "#cdnpoli" which is short for Canadian politics and "#auspol" which stands for Australian politics, there hashtags are commonly used over the years and have no exclusivity to 2019. "#populism" is unique, it is usually used in the context of political strategies that strive to appeal to ordinary people who feel that their concerns are disregarded by established elite groups. There are many variations of these beliefs but commonly there is a divide between the 'elites' and the 'common people'.

In June 2019, the top hashtags again share similar climate hashtags "climateemergency", "climatechange" and "climate". The unique tags are "extinctionrebellion" which is a decentralised, international and politically non-partisan movement using non-violent direct action and civil disobedience to persuade governments to act justly on the climate. The tweet which contained the majority of these hashtags contained no context other than calling all these hashtags part of the brainwashing, they tagged #globalist #fakescience #fakenews #climate fraud #un-#eu-#4threich

#extinctionrebellion #climateemergency, this became a popular retweet to accuse the people involved with these hashtags as promoting a hoax or fake news. The tweet could have originated from Piers Corbyn, a well known climate denier and brother of British politician Jeremy Corbyn, this was suspected as the tweets were retweets mentioning his name, I could not verify this at this current time as the account was suspended.

Top hashtags	May 2020	June 2020
1st	('#climatechange', 11296),	('#climatechange', 1334),
2nd	('#coronavirus', 9827),	('#climatecrisis', 549),
3rd	('#climatedenial', 7897),	('#populism', 366),
4th	('#globalwarming', 4330),	('#climatebrawl', 338),
5th	('#lies', 3615),	('#covid19', 315),
6th	('#covid19', 3502),	('#climatescam', 313),
7th	('#climatebrawl', 3358),	('#climate', 251),
8th	('#climatecrisis', 3131),	('#gretathunberg', 235),
9th	('#green', 2828),	('#misinformation', 211),
10th	('#climate', 2102)	('#climatebra', 203)

Figure 3.20: Top hashtags for 2020

2020 hashtags are a lot more environmental centered than other years with "green", "climatecrisis", "climatebrawl", 'globalwarming", "climatedenial". We could assume that the tweets attached to the hashtag "climatedenial" are trying to combat misinformation and climate denial as conspiracy mindsets don't often refer to themselves as deniers, they prefer to see them in a more positive tone, using truth seekers or realists etc. As this is the first hashtag set after Covid-19, we can see that #coronavirus and #covid19 are very popular even after being filtered into climate related tweets. From the previous exploratory analysis and the findings here, we could assume that Covid-19 was more the focus in May and June 2020 but climate was still discussed and more than likely combined with covid to push their conspiracy narratives.

June 2020 was missing 2 weeks of data so the numbers are a lot less, but consistently we see #climatebrawl appear nearly every month, this was a hashtag used to combat climate denial and misinformation while promoting real climate science. Covid-19 appears again as this was a time which included lockdowns and covid was in every piece of media with soaring number of cases. Unique themes include #gretathunberg, who is a young Swedish environmental activist who is known for challenging world leaders to take immediate action for climate change mitigation and reached international news for her speeches. One clear hashtag which would be promoting climate

denial and climate misinformation would be #climatescam, discussing the climate with this negative connotation indicates that they believe it is a hoax and political propaganda.

Top hashtags	May 2022	June 2022
1st	('#climatecrisis', 8375),	('#climatecrisis', 14887),
2nd	('#climatebrawl', 5537),	('#vote', 14675),
3rd	('#climatechange', 5032),	('#climate', 7757),
4th	('#climate', 4229),	('#climatebrawl', 6903),
5th	('#climatescience', 3430),	('#climatechange', 6067),
6th	('#propaganda', 3325),	('#climateemergency', 4174),
7th	('#climateemergency', 2713),	('#cop26', 3848),
8th	('#misinformation', 2588),	('#misinformation', 3756),
9th	('#disinformation', 2422),	('#disinformation', 3233),
10th	('#vote', 2202)	('#globalwarming', 3149)

Figure 3.21: Top hashtags for 2022

2022 saw a very similar result as previous years, with one exception at number 10, "vote" appears. We would assume that political candidates were talking about climate related policies and this was used in a political context. In May 2022, there were the local elections in the UK and the 2022 Australian federal election occurring, as prominent as Australian politics has been in the past tweets and with sampling the tweets, I would believe that Australian politics were dominating the narrative around the hashtag #vote.

As we move into June 2022, most of the hashtags are the same as the previous month, #vote actually increases in volume by over 6 times, when sampling the tweets from this month, they were clearly dominated by a US narrative now. The tweets mentioned GOP which is a reference to the Republican Party, the most common retweet mentioned that voting for the GOP would make positive change by passing laws against inequity, terrorism, disinformation, insurrection, climate and more. As witnessed in the past months, this is a user promoting a positive narrative to counter misinformation and false media. "#cop26" is a strange hashtag to appear in June 2022 as it took place in November 2021, Cop26 is the 2021 United Nations Climate Change Conference. After some investigation, the tweets revealed that The Institute for Strategic Dialogue (ISD) released a report on COP26 called "Documenting and responding to climate disinformation at COP26 and beyond", and aimed to correct all the misinformation involved in the conference and viral posts that were referencing COP26 during that time period.

3.9 Hashtag co-occurrence

Hashtag co-occurrence is created by compiling a list of the most common hashtags to occur together in the same tweet, to achieve this I used the hashtag list referenced above, sorted them and then used the "combinations" function from the library Itertools alongside the Counter object to count the all the possible combinations of two hashtags on the same tweet.

```
from collections import Counter
from itertools import combinations

counter = Counter()

for tag in hashtag_list:
    tag.sort()
    combos = list(combinations(tag, r=2))
    counter.update(combos)


counter.most_common(50)
```

Listing 3.4: Creating list of combination hashtags

I will continue to focus on the 3 most important years to my narrative and include the co-occurence hashtags for 2021 in Appendix 4 for comparison.

Co-occurence hashtags	May 2019	June 2019
1st	('#migration', '#populism'), 1135	('#climateemergency', '#extinctionrebellion'), 1614
2nd	('#climateemergency', '#migration'), 1119	('#climateemergency', '#fakenews'), 1604
3rd	('#climateemergency', '#populism'), 1119	('#eu', '#un'), 1590
4th	('#globalistsgonewild', '#globalwarminghoax'), 1069	('#climate', '#climateemergency'), 1571
5th	('#climatechange', '#propaganda'), 721	('#climate', '#eu'), 1570
6th	('#climatechange', '#united4climate'), 640	('#climate', '#un'), 1570
7th	('#agwhoax', '#cchoax'), 572	('#climateemergency', '#eu'), 1570
8th	('#agwhoax', '#globalistsgonewild'), 572	('#climateemergency', '#un'), 1570
9th	('#agwhoax', '#globalwarminghoax'), 572	('#climate', '#fakenews'), 1567
10th	('#cchoax', '#globalistsgonewild'), 572	('#climate', '#globalist'), 1567

Figure 3.22: Top co-occurring hashtags for 2019

When comparing themes with ngrams and top hashtags, again the narrative here is different, with the top combination hashtag referring to brexit and the challenges that the EU face. The tweet that gained traction discussed the problems that the EU face, the user believes that brexit, migration and populism are not the biggest threats but climate change, ageing population and digital revolutions are the real challenge

ahead. This would seem like a more coherent discussion and less of a rambling of a conspiracy mindset. Generally the popular hashtags will be repeated in the co-occurrence list as they would be the most commonly occurring hashtags and most tweets, if they contain hashtags, commonly attach more than one.

There are also new hashtags appears which are supporting climate denial such as #agwhoax and #cchoax which we can assume to mean anthropogenic global warming and climate change hoax. There are a lot more obvious hoax encouraging hashtags appearing in the co-occurrence table such as 3rd highest with 1069 tweets, #globalistsgonewild and #globalwarminghoax.

In June 2019, there is a complex mix with #climateemergency and #extinctionrebellion being first, these are usually used in supporting the belief in climate change and actively raising awareness to change lifestyles to help mitigate the effects of climate change. As we look at the 2nd most common co-occurrence, it appears that climate deniers have used a common hashtag and diluted it with #fakenews to spread their misinformation through the means of an already popular hashtag. The other hashtags are mostly generic combinations of #climate with #eu and #un, with #fakenews appearing again with #climate.

Co-occurence hashtags	May 2020	June 2020
1st	('#climatedenial', '#coronavirus'), 7697	('#gretathunberg', '#populism'), 222
2nd	('#globalwarming', '#lies'), 3517	('#climatecrisis', '#gretathunberg'), 215
3rd	('#climatechange', '#green'), 2791	('#climatecrisis', '#populism'), 215
4th	('#climatechange', '#covid19'), 1488	('#climatechange', '#misinformation'), 155
5th	('#climatechange', '#climatecrisis'), 1127	('#climatebrawl', '#climatechange'), 144
6th	('#climatecrisis', '#covid19'), 712	('#climatechange', '#covid19'), 140
7th	('#climatecrisis', '#disinformation'), 659	('#climatedenial', '#coronavirus'), 137
8th	('#auspol', '#climateemergency'), 642	('#climatechange', '#climatecrisis'), 124
9th	('#climatebrawl', '#climatecrisis'), 632	('#climatechange', '#populism'), 116
10th	('#climatechange', '#propaganda'), 559	('#climatechange', '#disinformation'), 83

Figure 3.23: Top co-occurring hashtags for 2020

In May 2020, we can see the first clear combination of Covid-19 and climate related hashtags, number 1 by a clear amount at 7697 tweets contains #coronavirus paired with #climatedenial. Once again, we can assume that these tweets would be looking to counter the misinformation as conspiracy believers prefer not to use #climatedenial and would prefer to use the word hoax or truth seeking. At 2nd, this would be a more preferable hashtag combination for climate deniers as the people who attach #lies to a tweet, are usually focusing on reacting negatively to media around climate or global warming.

In the mix of climate misinformation and climate science media, #auspol appears once again consistently every year. This could be in reaction to climate disasters happening in Australia such as the bushfires or it could be due to the reaction of climate events from Australian politicians, there was also mentions of Australian protests during this time regarding Covid-19 and environmental issues.

June 2020 continues to have similar combinations hashtags to May 2020, #gretathunberg appears as a crossover with the top hashtags and #populism to pair with it. This could indicate that Greta Thunberg had a speech referring to populism and the climate crisis. Covid-19 and climate related tweets are continuously used in conjunction with each other to further enable conspiracy mindsets. In 2021, the BBC wrote a report investigating how Covid-19 denial has enabled more people to believe climate denial, the conspiracy believers will often start to believe that other conspiracies are all part of the main plot and use one conspiracy as a gateway (Spring, 2021).

Co-occurence hashtags	May 2022	June 2022
1st	('#climatebrawl', '#climatecrisis'), 2986	('#climatebrawl', '#climatecrisis'), 3861
2nd	('#climatecrisis', '#climateemergency'), 1223	('#climatecrisis', '#climateemergency'), 2870
3rd	('#climatebrawl', '#climateemergency'), 722	('#bonnclimateconference', '#cop26'), 2796
4th	('#climate', '#disinformation'), 561	('#climatebrawl', '#climateemergency'), 2319
5th	('#climatechange', '#misinformation'), 558	('#covid19', '#globalwarming'), 2135
6th	('#climatecrisis', '#disinformation'), 543	('#co2', '#covid19'), 2120
7th	('#climatechange', '#propaganda'), 542	('#co2', '#geoingenierie'), 2120
8th	('#climatechange', '#disinformation'), 538	('#co2', '#globalwarming'), 2120
9th	('#climate', '#climatecrisis'), 458	('#co2', '#vivi'), 2120
10th	('#auspol', '#ausvotes'), 452	('#covid19', '#geoingenierie'), 2120

Figure 3.24: Top co-occurring hashtags for 2022

2022 continues a very similar trend as 2020, #climatebrawl, #propaganda, #disinformation etc. As we are 2 years past the start of covid, it has disappeared from the combination hashtags but it appears again in June 2022. As covid media coverage declines, it is still being used alongside #globalwarming, this supports the findings from the BBC's report on new conspiracy believers using it as a gateway conspiracy onto climate denial (Spring, 2021).

In June 2022, unique hashtags appear such as #geoingenierie, #vivi, #co2 and #bonnclimateconference. #cop26 and #bonnclimateconference appear in June 2022 most likely due to the The Institute for Strategic Dialogue (ISD) releasing a report on documenting and responding to climate disinformation at COP26 and beyond. This would cause users to doubt what was being discussed at the Bonn Climate Conference which was hosted in June 2022. #vivi and #geoingenierie came from Italian tweets,

this would have been an error in the Twitter API or the collection tool in misdefining these tweets as English.

3.10 Hashtag co-occurrence heatmaps

To create the heatmaps, I took the 50 most common hashtag combinations and extracted every individual hashtag that appeared in the top 50 combinations, I choose only the top hashtags because a heatmap with too many hashtags would decrease readability significantly.

After I extracted the hashtags, I created a matrix using numpy to count each combination ready to be mapped into the visualisation. I then gave each hashtag an id number to code the data, then looped through the hashtag list to increment each combination. After the loop had finished, I converted the matrix into a dataframe and used Seaborn to visualise the data as a heatmap.

```
import numpy as np

# creating matrix of zeros for co-occurence
matrix1 = np.zeros((len(uni_htags_1), len(uni_htags_1)))
htag_to_id1 = {uni_htags_1[i]:i for i in range(len(uni_htags_1))}

# filling matrix with data

for _ in hashtag_ref:
        hashtags_id1 = [htag_to_id1[x] for x in _ if x in
    uni_htags_1]
        for h in hashtags_id1:
            for o in hashtags_id1:
                if h != o:
                    matrix1[h, o] += 1

# convert matrix into dataframe
heatmap_df1 = pd.DataFrame(data=matrix1, columns=uni_htags_1, index=
    uni_htags_1)

# create heatmap
sns.heatmap(heatmap_df1, square=True, cmap='viridis')
```

Listing 3.5: Creating a matrix to prepare data for the heatmap

I will display heatmaps for 2019, 2020 and 2022 in the main body, 2021 will be included in Appendix 5.

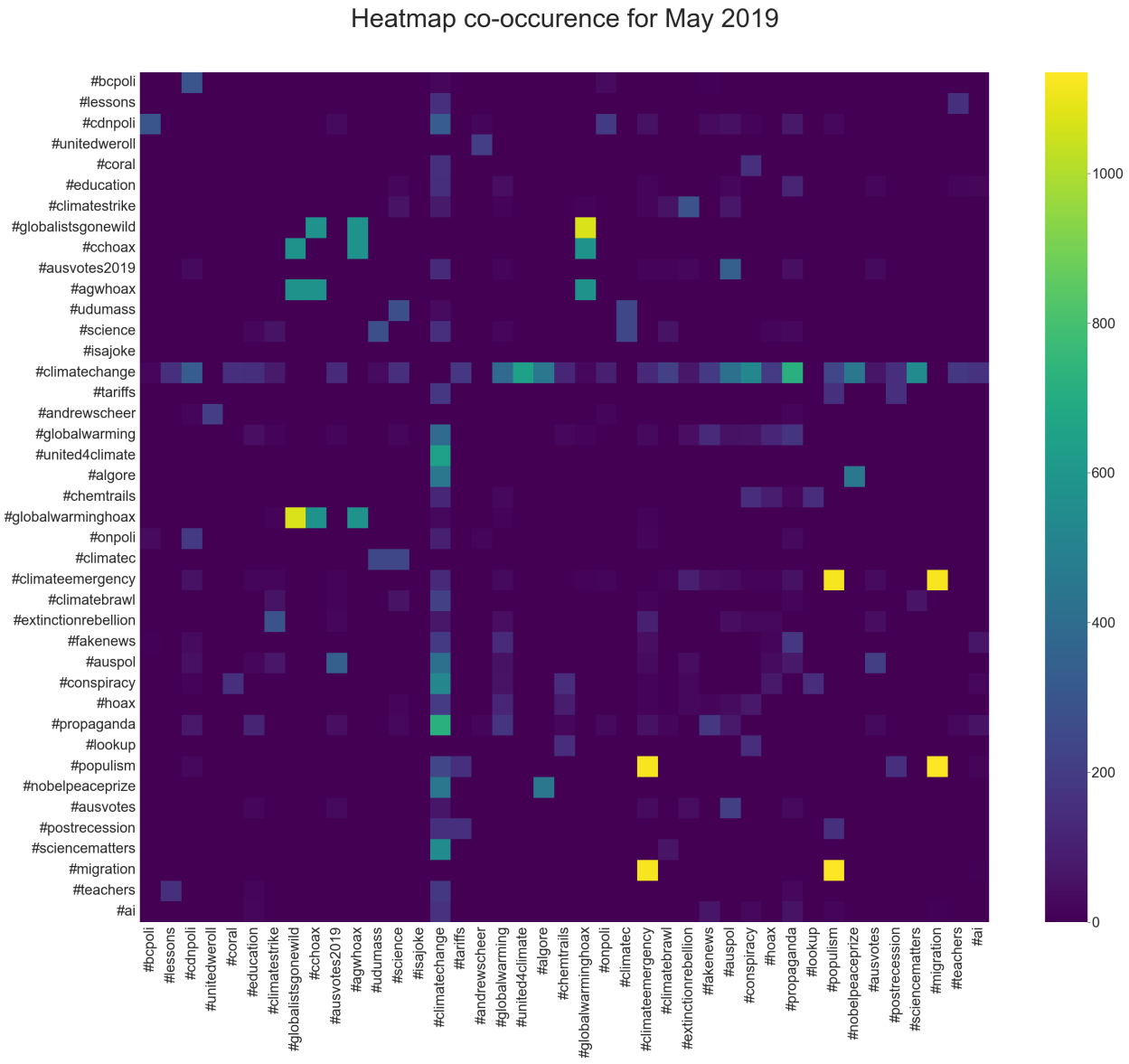


Figure 3.25: Hashtag heatmap for May 2019

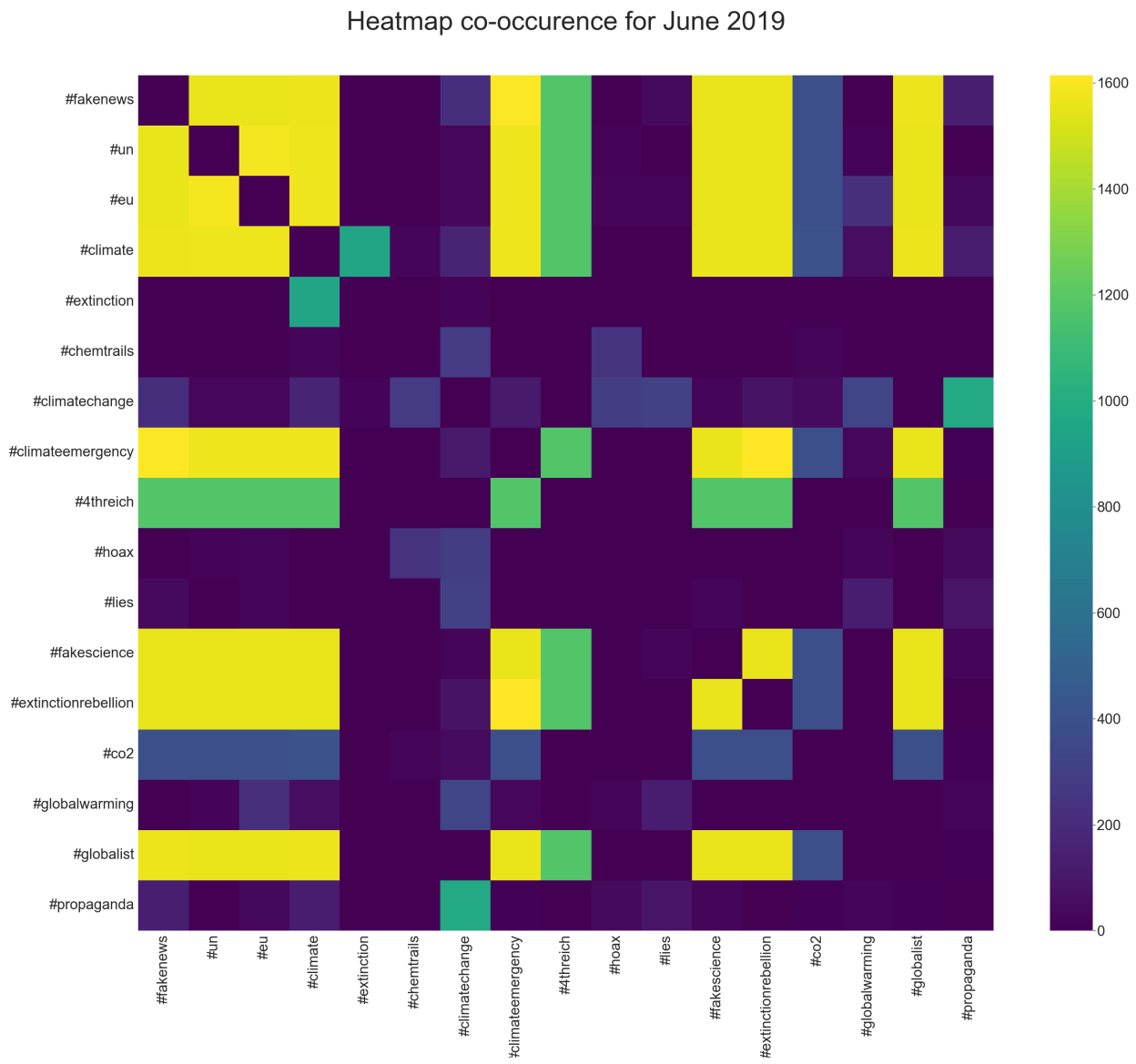


Figure 3.26: Hashtag heatmap for June 2019

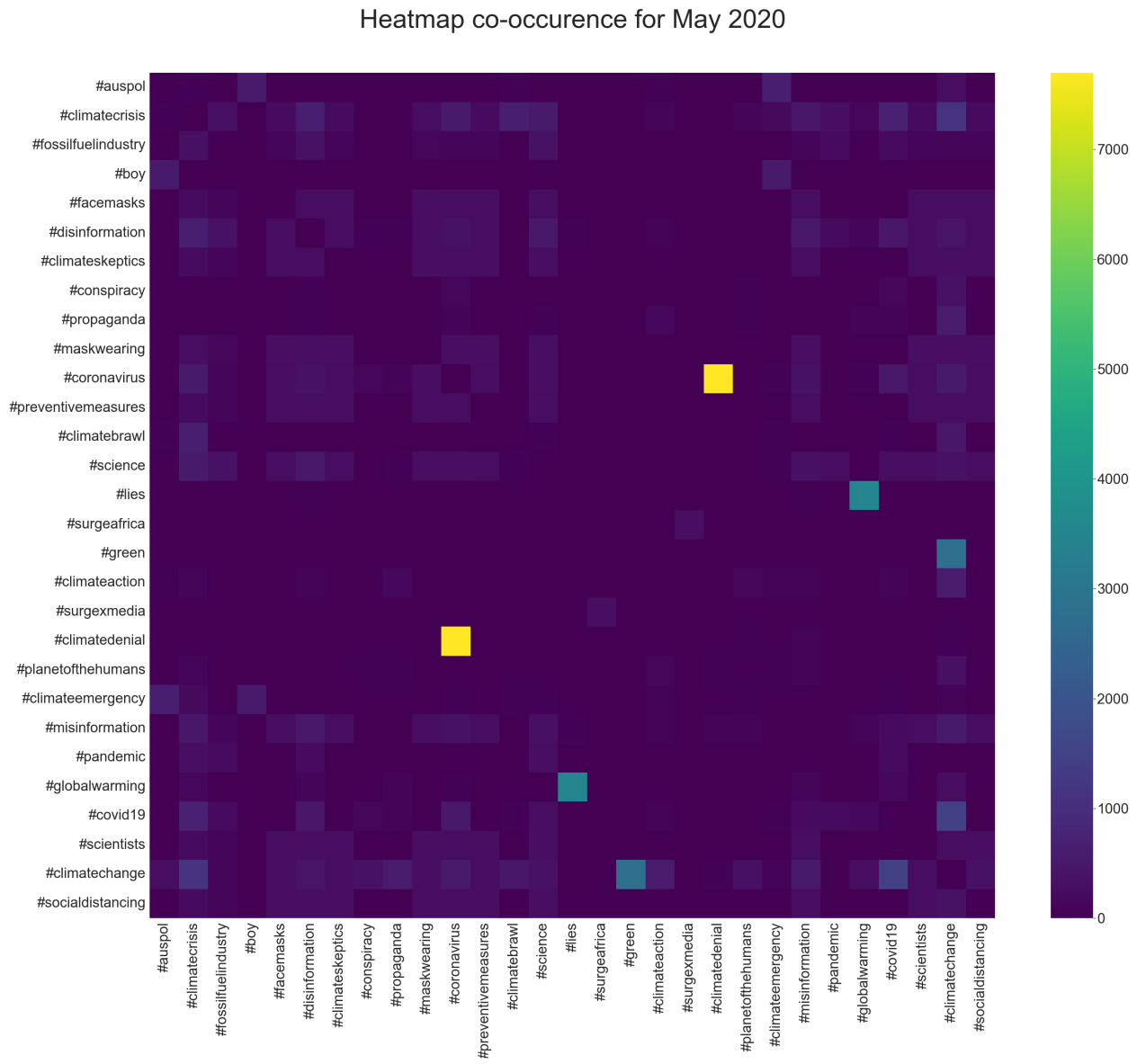


Figure 3.27: Hashtag heatmap for May 2020

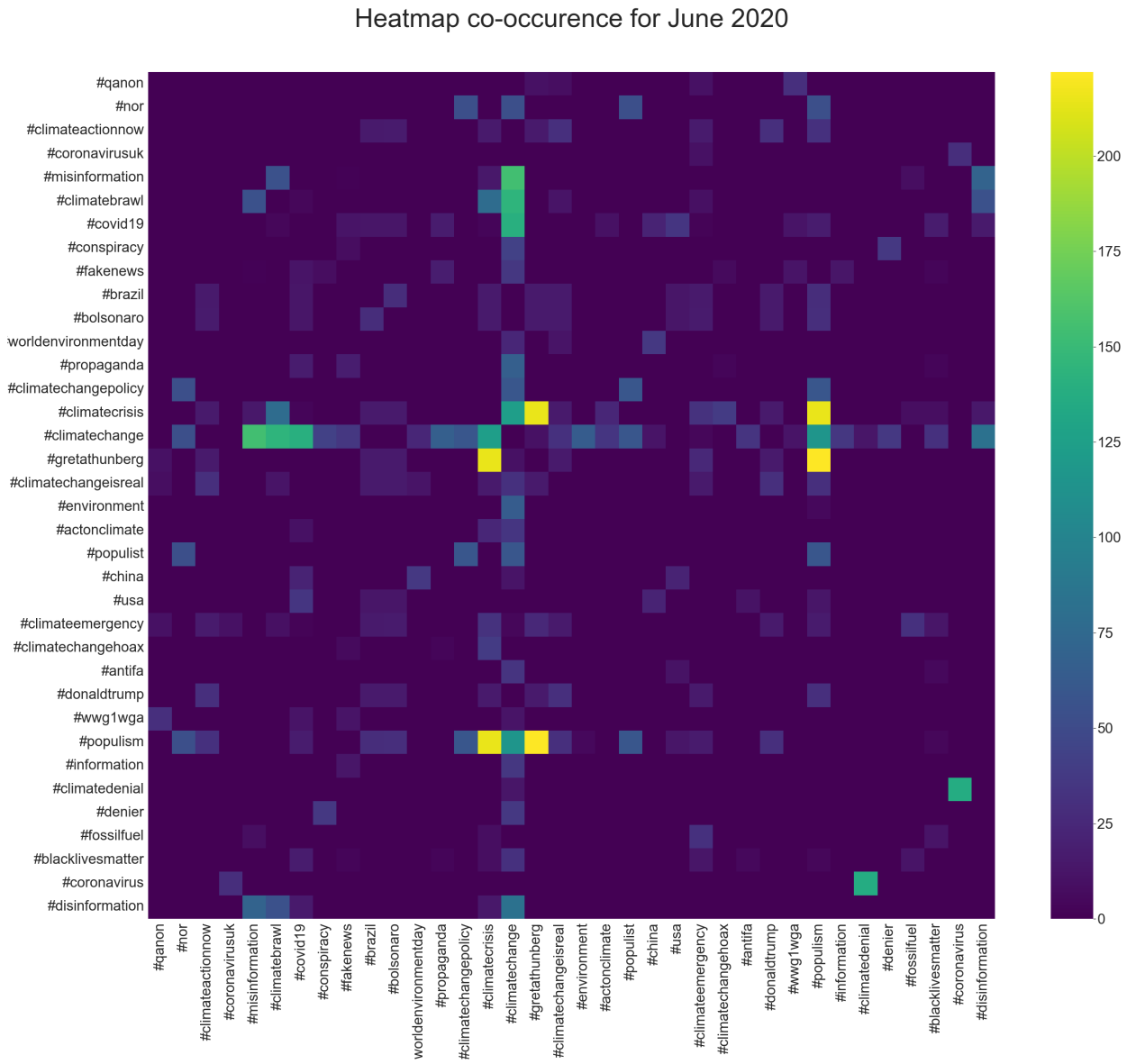


Figure 3.28: Hashtag heatmap for June 2020

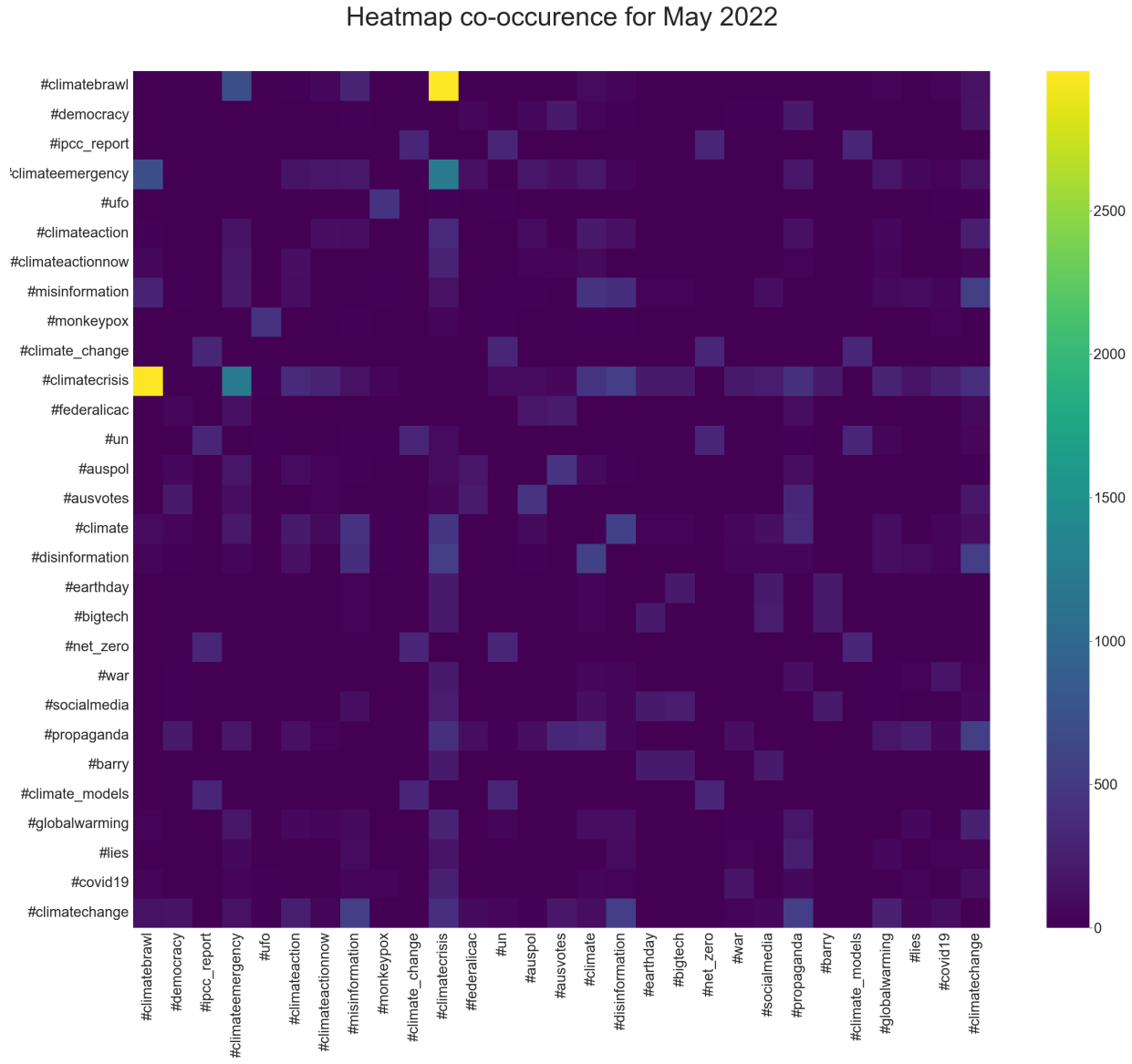


Figure 3.29: Hashtag heatmap for May 2022

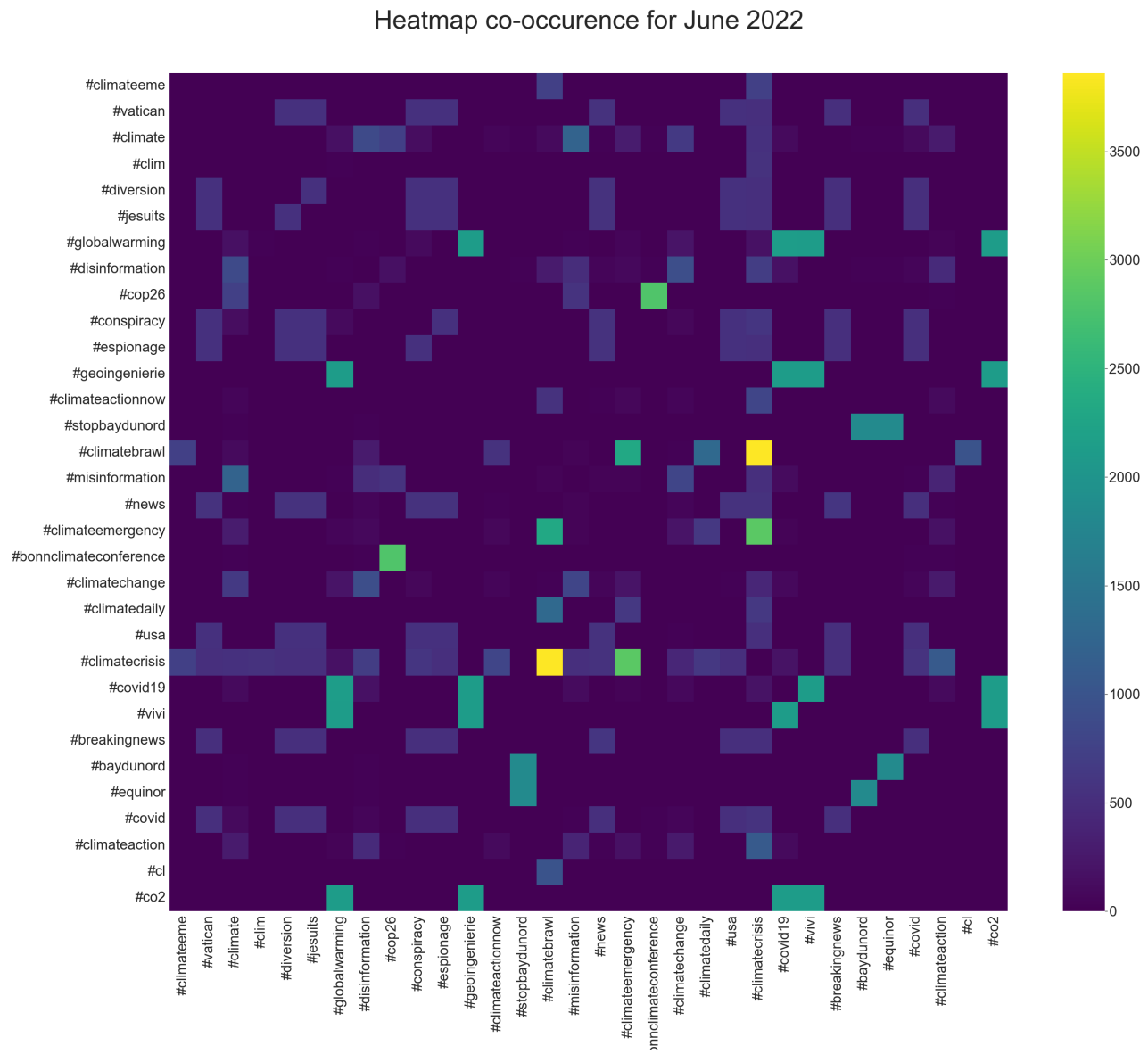


Figure 3.30: Hashtag heatmap for June 2022

3.11 Concordance

Concordance was used throughout the project to gain context and derive insights from popular hashtags, ngrams and co-occurrence hashtags. The method that was used came from the NLTK library and the Text object.

```
from nltk.text import Text

text = Text(climate_strings.split(" "))
concord_climate = text.concordance("traumatize", width=150, lines
    =50)
```

Listing 3.6: Concordance conducted from NLTK.text

For example, the trigrams in June 2020 revealed a theme around "traumatize", "five", "year", without context, this could be interpreted into many different stories. After concordance we can clearly see where this narrative came from and can trace the tweet back to a public figure and explain how and why this tweet gained popularity.

```
Displaying 50 of 10549 matches:
.. did i miss anything? rt : abuse: nj leftists plan to brainwash and traumatize five year old children with political propag
anda on so-called "climat
some sort of “dealing”. rt : abuse: nj leftists plan to brainwash and traumatize five year old children with political propag
anda on so-called "climat
n so-called "climate c… rt : abuse: nj leftists plan to brainwash and traumatize five year old children with political propag
anda on so-called "climat
n so-called "climate c… rt : abuse: nj leftists plan to brainwash and traumatize five year old children with political propag
anda on so-called "climat
n so-called "climate c… rt : abuse: nj leftists plan to brainwash and traumatize five year old children with political propag
anda on so-called "climat
n so-called "climate c… rt : abuse: nj leftists plan to brainwash and traumatize five year old children with political propag
anda on so-called "climat
n so-called "climate c… rt : abuse: nj leftists plan to brainwash and traumatize five year old children with political propag
anda on so-called "climat
n so-called "climate c… rt : abuse: nj leftists plan to brainwash and traumatize five year old children with political propag
anda on so-called "climat
n so-called "climate c… rt : abuse: nj leftists plan to brainwash and traumatize five year old children with political propag
anda on so-called "climat
```

Figure 3.31: Concordance for June 2020 trigrams

This was used for hashtags that were not self explanatory such as #vivi, #geoingenierie, after concordance, we found out they are non English hashtags. It was also used for stories such as "Sean Hannity" and "Michael Moore", therefore concordance was crucial to having a deeper understanding of what the social media climate looked like in the past months.

```
Displaying 50 of 2054 matches:
x il #covid19, ecco una nuova propaganda per imprigionarci in norme assurde.
#vivi #co2 #globalwarming #geoingenierie .

https://t.co/poejyjufps
x il #covid19, ecco una nuova propaganda per imprigionarci in norme assurde.
#vivi #co2 #globalwarming #geoingenierie .

https://t.co/9oteaiuzgr rt :
x il #covid19, ecco una nuova propaganda per imprigionarci in norme assurde.
#vivi #co2 #globalwarming #geoingenierie .

https://t.co/9oteaiuzgr  emer
x il #covid19, ecco una nuova propaganda per imprigionarci in norme assurde.
#vivi #co2 #globalwarming #geoingenierie .

https://t.co/9oteaiuzgr rt :
x il #covid19, ecco una nuova propaganda per imprigionarci in norme assurde.
#vivi #co2 #globalwarming #geoingenierie .

https://t.co/9oteaiuzgr rt :
x il #covid19, ecco una nuova propaganda per imprigionarci in norme assurde.
#vivi #co2 #globalwarming #geoingenierie .
```

Figure 3.32: Concordance for June 2022 trigrams

3.12 TF-IDF

The final method used was Term Frequency - Inverse Document Frequency or TF-IDF for short, this is a numerical statistic that is intended to reflect how important a word is to a document in a collection or corpus.

Term Frequency: TF is the frequency of a term or word is the number of times the term appears in a document compared to the total number of words in the document.

Inverse Document Frequency: IDF of a term reflects the proportion of documents in the corpus that contain the term. Words unique to a small percentage of documents (e.g., technical jargon terms) receive higher importance values than words common across all documents (e.g., a, the, and) (Karabiber, 2022).

To learn how to conduct this in python, I used various sources to read or watch and understand the model. The main resource used to assist me in conducting the TF-IDF vectors was an article named "TF-IDF Vectorizer scikit-learn" written by Chaudhary in 2020 on Medium.

The library that was used was Skikit-learn (Sklearn) and I imported the TdidfVectorizer object to use the fit_transform and get_feature_names functions.

```
from sklearn.feature_extraction.text import TfidfVectorizer
```

As the datasets were so large, the TF-IDF vector would not run on a whole months worth of tweets. To solve these memory issues, I removed all the retweets and at-

tempted to run the TF-IDF vectorizer again, this still didn't work so finally I removed all the replies.

The remaining tweets left were only tweets that were posted by the user manually, no retweets or replies, I believed these were the most valuable tweets to be looking at and it solved the memory issues so the vectorizer runs correctly.

The data was also cleaned with my personal stop words but I additionally used the argument for the TfidfVectorizer stop words.

```
no_retweets = []

for index, row in df.iterrows():
    retweet = re.match(r"^rt @+", row.tweet_text)
    if retweet:
        continue
    elif row.is_reply:
        continue
    else:
        no_retweets.append(row.tweet_text)

tfidf_data = [preprocess_text(tweet) for tweet in no_retweets]
```

Listing 3.7: Removing all retweets and replies then cleaning stop words

```
tfidfvectorizer = TfidfVectorizer(analyzer='word',stop_words= '
    english')

tfidf_wm = tfidfvectorizer.fit_transform(tfidf_data)

tfidf_tokens = tfidfvectorizer.get_feature_names()

df_tfidfvect = pd.DataFrame(data = tfidf_wm.todense(),index = range
    (0,len(tfidf_data)),columns = tfidf_tokens)
```

Listing 3.8: TF-IDF vectorizer and creating the dataframe

After the vectorized dataframe was complete, I extracted the highest scoring words using Pandas.

```
tokens_above_threshold = df_tfidfvect.max()[df_tfidfvect.max() >
    0.7].sort_values(ascending=False)
```

Listing 3.9: Extracting tokens with a high score

```
In [13]: tokens_above_threshold

Out[13]: fabrication          0.987049
         hoax                 0.972511
         cult                 0.963388
         chapter              0.961467
         genetics             0.953981
         nihilism             0.951844
         fmz32oobbz           0.941718
         contradict           0.925842
         emergency            0.923577
         propaganda           0.914346
         fiovlyv0zs           0.907134
         bullshit             0.906276
         dan                  0.902062
         sensible             0.898644
         vgiippgffx           0.894490
         blah                 0.894342
         fraud                0.890905
         climatechange        0.889784
         jab                  0.888430
```

Figure 3.33: TF-IDF high scoring tokens filtered from June 2022

When reviewing the top scoring tokens from the TF-IDF model, I also manually removed any nonsensical tokens such as "vgiippgffx", these are usually links to external sources from tweets or potentially throwaway usernames which wouldn't give any context or insight. This could have been prevented by cleaning all the links from the tweets beforehand.

Green = Top 10 scoring tokens
Yellow = 11th - 20th scoring tokens
Red = 21st - 30th scoring tokens

May 2019 TF-IDF consisted of 43,064 tweets.
June 2019 TF-IDF consisted of 42,849 tweets.

The top words from May 2019 that are related to climate misinformation are "disinformation", "hoax", "fakenewsmedia", "deforestation" and "conspiracy". The majority of the words are not related but might have been used as part of the climate tweet such as "foreign", "emergency", "junk", "leftist" and "right", but there are also some tokens that seem to have no relevance to the climate dataset or the misinformation dataset such as "mum", "squid", "barbie" etc. There is a possibility that these tokens were included in random tweets trying to get their tweets seen by using popular hashtags or phrases to appear in the trending results.

For June 2019, the top scoring token has no context or relevance to the climate dataset but could have been included in misinformation tweets. There are a lot less

May 2019	TF-IDF Score		June 2019	TF-IDF Score
disinformation	0.956855		blah	0.966488
fascist	0.950828		hoax	0.947942
hoax	0.950274		savetheenvironmentin5words	0.937521
coverup	0.882162		fortunately	0.914983
right	0.87802		smh	0.913716
progressive	0.875446		total	0.905111
foreign	0.864549		ffs	0.892256
fakenewsmedia	0.862163		globalwarming	0.891796
mudi	0.859461		emergency	0.861674
chineese	0.858069		reality	0.859639
squid	0.837177		indiana	0.853576
broadcaster	0.836756		epitome	0.852971
leftist	0.836382		thing	0.836032
4chong	0.835116		complete	0.833166
asteroid	0.829771		everyplace	0.82959
mum	0.823595		diversion	0.828982
label	0.801432		conspiracy	0.828131
modern	0.798368		snowpiercer	0.822107
deforestation	0.798141		newswirenow	0.819536
emergency	0.797954		short	0.812565
im	0.793991		delusional	0.811216
care	0.792029		climatechange	0.805648
conspiracy	0.783373		guess	0.805148
junk	0.778726		donnie	0.791136
pro	0.76856		real	0.784495
rich	0.761598		right	0.784402
pick	0.754632		climatealarmists	0.765503
newswirenow	0.751517		machine	0.758236
homework	0.749238		infront	0.757354
barbie	0.747476		buffoon	0.754028

Figure 3.34: TF-IDF high scoring tokens filtered from May and June 2019

uniquely named places or concepts in the top 10 of June 2019, we have "ffs", "smh", "fortunately", "total", these tokens alone have no meaning other than potential signs of frustration or disappointment from "smh" or "ffs" but no definite meaning from such tokens.

May 2020	TF-IDF Score		June 2020	TF-IDF Score
disinformation	0.935318		war	0.937245
beep	0.904349		blah	0.890321
whing	0.898014		card	0.843919
feminism	0.874024		struggle	0.839314
giggs	0.86298		climategate	0.827422
anti	0.857641		genderpaygap	0.822633
fake	0.84138		dead	0.802554
lie	0.808265		yep	0.794469
agenda	0.798952		economy	0.790859
office	0.789654		climatehoax	0.784042
dumb	0.774806		enemy	0.782108
model	0.772237		tho	0.771158
excruciate	0.771956		amazon	0.769521
potus	0.770559		awww	0.750203
fight	0.76436		smh	0.747614
climatehoax	0.757253		fake	0.740621
anthropogenic	0.75541		pm	0.736572
possess	0.717868		hoax	0.735746
frame	0.713966		mail	0.727003
thing	0.710932		record	0.722569
indoors	0.71		troll	0.713855
suit	0.709989		mad	0.708313
season	0.707181		countryfile	0.702985
assumption	0.70072		climatechangehoax	0.699327
mega	0.698834		proof	0.699093
favourite	0.696523		immunize	0.68412
ridiculous	0.695302		real	0.677006
theeppn	0.691082		beat	0.671901
mekong	0.690204		hearing	0.670741
covid19	0.680013		fear	0.65889

Figure 3.35: TF-IDF high scoring tokens filtered from May and June 2020

May 2020 TF-IDF consisted of 46,000 tweets.
June 2020 TF-IDF consisted of 8,122 tweets.

In 2020, we see a similar trend of "disinformation" and "blah" being very high scoring tokens, along with common ngrams such as "climatehox", "climategate", "fake". There are also tokens that appear that could be related to climate misinformation but were never seen in the ngrams or hashtags such as "war", "feminism", these themes can easily be introduced alongside misinformation and be discussed in the political themes that were witnessed previously.

May 2022	TF-IDF Score		June 2022	TF-IDF Score
fuck	0.970685		fabrication	0.987049
climatecult	0.957677		hoax	0.972511
thinker	0.953778		cult	0.963388
climatescam	0.952308		chapter	0.961467
hoax	0.902629		genetics	0.953981
laugh	0.899221		nihilism	0.951844
planter	0.899019		contradict	0.925842
ridicule	0.897461		emergency	0.923577
shove	0.890635		propaganda	0.914346
propaganda	0.886757		bullshit	0.906276
fake	0.880541		dan	0.902062
terribly	0.870305		sensible	0.898644
climatecrisis	0.865842		blah	0.894342
kill	0.860999		fraud	0.890905
czar	0.85479		climatechange	0.889784
nazi	0.851183		jab	0.88843
midwit	0.849596		fear	0.875792
heartwarmingly	0.848064		climatechangeisreal	0.874825
rep	0.845485		control	0.873562
suck	0.840595		toxic	0.866645
realscience	0.835447		ain	0.855994
fraud	0.834893		fake	0.849307
fabrication	0.834087		oops	0.833899
bulshit	0.83322		gpvm41de3p	0.832589
proud	0.828174		crisis	0.827947
defundthebbc	0.81527		warming	0.82624
passport	0.812058		gt	0.826008
lifetime	0.810422		climatechangehoax	0.729449
kcranews	0.872326		hoax	0.721
heartwarmingly	0.870795		permanently	0.715349

Figure 3.36: TF-IDF high scoring tokens filtered from May and June 2022

May 2022 TF-IDF consisted of 59,979 tweets.
June 2022 TF-IDF consisted of 51,426 tweets.

In 2022, we see the tokens seem to turn more aggressive, "fuck", "climatecult", "climatescam", "ridicule", "propaganda" in May and it continues into June 2022 with "cult", "bullshit", "fabrication", "contradict". There are even mentions of very serious themes on the extreme spectrum such as "nazi" and "kill", fortunately these words did not appear in the ngrams or hashtags which means this extremism comes from a small minority but should not be ignored as extremism and misinformation spread can still be very dangerous to the people exposed to these narratives.

The TF-IDF scores for May/June 2021 will be included in Appendix 6.

3.13 Summary of Chapter 3

In this chapter, I have established that there are clear indications that climate related discourse dropped during the start of the Covid-19 pandemic and that climate tweets were used in conjunction with covid tweets to push personal or political agendas. I have also been able to get a clear picture of the themes that dominated the narrative for each month in May/June for the years 2019, 2020, 2021 and 2022 by using Natural Language techniques. In the years that I have analysed above, the interest and participation in climate related discourse only increases after the year 2020.

In Chapter 4, I will further investigate the upwards trend into July 2022 which also involves the historical record breaking temperatures in the UK by reaching over 40 degrees Celsius in parts of England. There were worldwide heatwave related events happening in July 2022 so this month should be an extremely interesting month to investigate and compare to the other years. I expect July 2022 to involve a lot more focused climate related tweets as a new 'crisis' arrives.

Chapter 4

Investigation into how the recent heatwave in July 2022 affected climate related tweets

4.1 Daily volume of tweets in July 2022

After combining the JSON files into one, I used Pandas to sort all tweets into their specific days and counting them in preparation to create a bar plot using seaborn.

From Figure 4.1, we can clearly see an elevated amount of climate tweets starting around the 17th and falling drastically at the 21st. This tracks perfectly with the history breaking heatwave in multiple locations around the world. Referring back to Figure 3.5, we can see the massive increase in the amount of climate related tweets to all the other months explored, going from May 2022 with 172,276,436 tweets in the callout dataset with only 400,491 climate tweets to July 2022 with 122,654,138 callout tweets and 636,175 climate tweets. From these stats, we can clearly see an increase in 2022 of climate related media and narratives and it seems to only be increasing.

The UK experienced a brief but unprecedented extreme heatwave from 16 to 19 July 2022, with extreme temperatures recorded on both 18th and 19th. On 19th, 40.3°C was recorded at Coningsby (Lincolnshire), setting a new UK record by a margin of 1.6°C, and multiple stations across England also exceeded 40°C. This heatwave marked a milestone in UK climate history, with 40°C being recorded for the first time

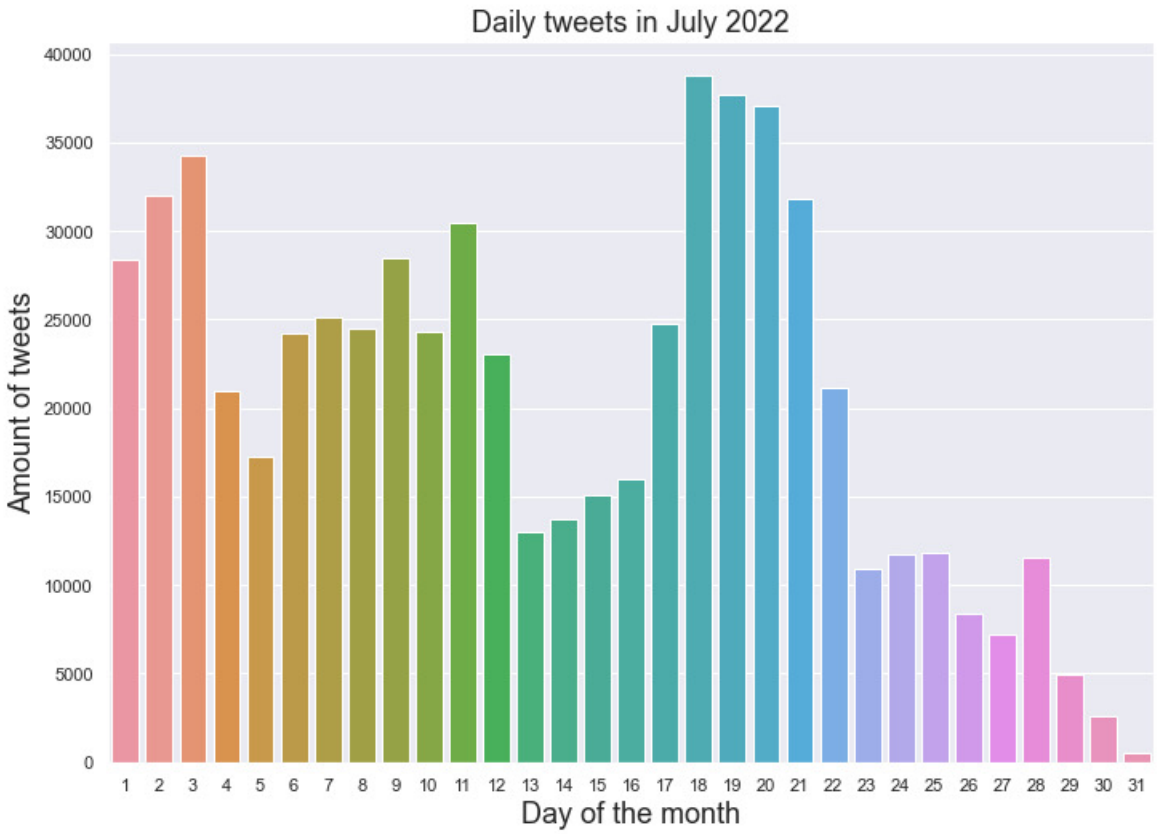


Figure 4.1: Amount of tweets per day in July 2022

in the UK (Met office, 2022b). The heat did not stop on the 19th, it continues until the 21st and the climate related tweets were also inflated along with the heat.

The heightened volume of tweets around the 8th July could have been due to the mainstream media projecting the Met Offices first warnings for extreme heat on the 8th of July. Conspiracists would often criticise these institutions for using such extreme language such as "national emergency" and claim that this is scaremongering. On July 15th, Met Office tweeted warning UK residents about the red extreme heat warning issued (Met Office, 2022a), from this point, the climate tweets ascend until the end of the heatwave.

Besides the UK, there were also heatwaves across Europe, From June to August 2022, heatwaves have affected parts of Europe, causing evacuations and heat-related deaths. The height of the temperatures was recorded in Pinhão, Portugal, on 14 July at 47 degrees celsius (Lusa, 2022)

Other parts of Europe that exceeded 40 degrees celsius were Brittany, Biscarosse (Landes), Cazaux (Gironde), Nantes (Loire-Atlantique), La Roche-sur-Yon (Vendée) and Lanmeur (Finistère) (Meteo France, 2022). This also included wildfires in Gironde, causing a total of nearly 37,000 people to be evacuated (Gironde, 2022). Other European countries included Germany, Spain, Hungary, Norway.

In the U.S., the heat was also inflated and placed more than 150 million people under heat warnings and advisories. Nearly every region of the U.S. experienced above average temperatures. Several states saw record-breaking triple digit highs in fahrenheit. With the added impact of high humidity in many regions, the extreme heat threatened the life and health of the residents (Pratt, 2022). This is enough evidence to back up the sheer increase of volume and outrage appearing on Twitter during July 2022 and more specifically July 16th to the 21st.

For comparison, I will include May/June 2019-2022 bar plots showing the individual tweets per day in Appendix 7.

4.2 July 2022 Ngrams

July 2022	Unigrams (BOW)	Bigrams	Trigrams
1st	climate, 558988	('climate', 'change'), 232367	('write', 'hit', 'piece'), 53390
2nd	change, 244331	('climate', 'misinformation'), 75769	('ban', 'climate', 'misinformation'), 36374
3rd	propaganda, 208090	('hit', 'piece'), 64351	('hit', 'piece', 'want'), 33045
4th	misinformation, 119458	('think', 'tank'), 62388	('hit', 'piece', 'try'), 31262
5th	conspiracy, 111664	('write', 'hit'), 53390	('recently', 'write', 'hit'), 30614
6th	think, 92027	('global', 'warming'), 40952	('think', 'tank', 'recently'), 30594
7th	disinformation, 79855	('bill', 'gate'), 38900	('tank', 'recently', 'write'), 30594
8th	fund, 69607	('ban', 'climate'), 36403	('climate', 'misinformation', 'guess'), 30499
9th	write, 67668	('conspiracy', 'theory'), 34954	('misinformation', 'guess', 'fund'), 27476
10th	piece, 67323	('piece', 'want'), 33045	('fund', 'think', 'tank'), 23408

Figure 4.2: Ngrams for July 2022

The discourse in July 2022 was very mixed but extremely focused on climate chatter, there are no other themes in this month compared to 2020 and 2021 where Covid-19 and other politics interfered. The top ngrams discuss the usual climate change discussion between climate deniers and climate believers, we can see this by the popular hashtag "#climatebrawl", which has been frequently examined as a hashtag used when debating climate deniers. Ngrams such as "conspiracy", "disinformation", "misinformation" and "propaganda" are also examples of the climate denial continuing to accelerate into July 2022, it should also be noted that people trying to combat this misinformation might use these words or hashtags.

The more popular unique theme came from a Swedish journalist tweeting about Bill Gates funding a think tank to deplatform the journalist for his views on climate misinformation. In the conspiracy mindset world, this would be an admittance of guilt if the climate activists tried to silence their views. These tweets didn't seem to have any clear evidence or reputable sources, it could have been used as a way to gain popularity for his off platform chat rooms using Telegram where his views could not be silenced.

Snippets from the popular tweets which assembled the Ngrams.

"organization wrote a defamatory hit piece trying to deplatform me for"

"a think tank funded by bill gates wrote a defamatory hit piece trying to deplatform me for"

4.3 July 2022 Word cloud

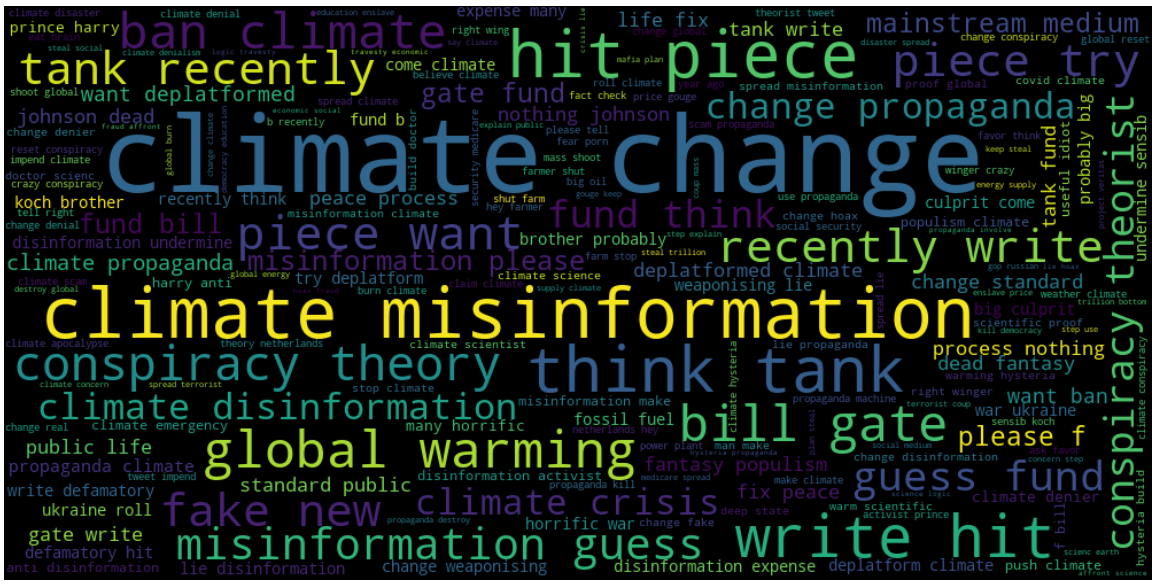


Figure 4.3: Word cloud for July 2022

The word cloud displays a visual aid to the combined ngrams, it can also be used to see smaller themes and observe them such as "prince harry", "ukraine roll" and "weaponising lie" included.

4.4 July 2022 Top hashtags

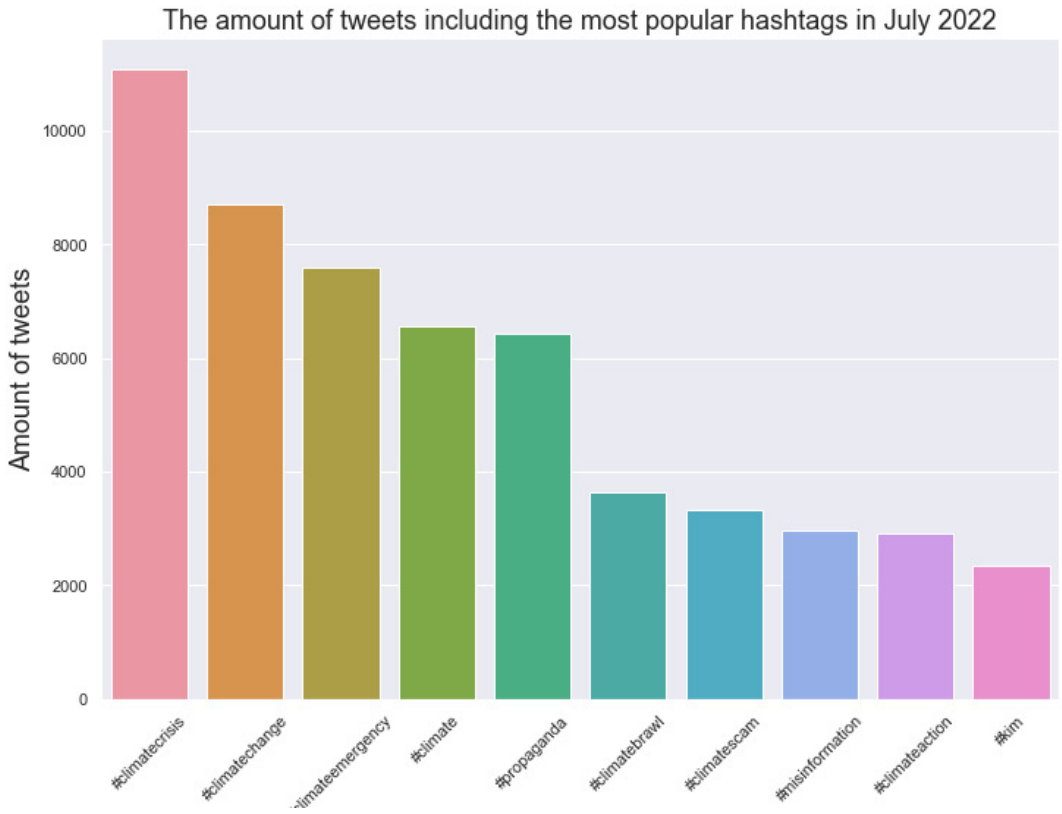


Figure 4.4: Top hashtag bar plot for July 2022

Top hashtags	July 2022
1st	('#climatecrisis', 11074),
2nd	('#climatechange', 8718),
3rd	('#climateemergency', 7601),
4th	('#climate', 6562),
5th	('#propaganda', 6441),
6th	('#climatebrawl', 3623),
7th	('#climatescam', 3319),
8th	('#misinformation', 2955),
9th	('#climateaction', 2902),
10th	('#kim', 2347)

Figure 4.5: Top hashtags table for July 2022

When we identify the most common hashtags for July 2022, we see a very similar picture as the other months examined but see a consistent focus on only climate, the tweets did not commonly include other topics to combine narratives. The only hashtag that is new and unique to July is "#kim" and very surprisingly "#climatescam", this

might indicate that there has been more of a negative sentiment to climate change media in recent months.

Taking a closer look at tweets containing the #kim, the tweet contents are referring to a skit mocking the conspiracy that global warming is a Chinese hoax, the sitcom is called "Unbreakable Kimmy Schmidt" and jokes about how bizarre and resourceful a climate change hoax would need to be.

Overall the hashtags were all climate related and only "#climatecrisis" is the most popular hashtag by a small amount, this could be used as a neutral hashtag as climate deniers and climate believers would both use this hashtag to refer to the climate crisis occurring. The only clear negative hashtag is down at 7th with 3319 tweets and refers to the climate change media as a climate scam.

4.5 July 2022 Hashtag co-occurrence

Co-occurence hashtags	July 2022
1st	('#climatecrisis', '#climateemergency'), 2377
2nd	('#covid19', '#misinformation'), 1139
3rd	('#climate', '#climatecrisis'), 1106
4th	('#climate', '#climateemergency'), 1049
5th	('#covid19', '#vaccination'), 1012
6th	('#misinformation', '#vaccination'), 1012
7th	('#climatecrisis', '#fossilfuel'), 942
8th	('#climatechange', '#climatecrisis'), 908
9th	('#crisiidrica', '#vivi'), 860
10th	('#climateemergency', '#democrats'), 843

Figure 4.6: Hashtag co-occurrence table July 2022

The co-occurrence hashtags show the combined narratives that appear, the majority are climate related as expected from the focused top hashtags, but the second most common combination was again referring to Covid-19 misinformation which shows that although media around covid has dwindled drastically, people were still tweeting about Covid-19 misinformation while combining the climate theme. To reinforce this theme, the 5th theme introduces "#covid19" and "#vaccination", in July 2022 most countries were already past the active vaccination stages with 3 covid vaccinations. This could be proof that Covid-19 denial and vaccination misinformation is introducing conspiracy mindsets to climate denial and a lack of trust in authority as literature proved.

Again a non English combination appeared in the top 10, using a past hashtag "#vivi"

and a new hashtag "#crisiidrica". There is also one hashtag mentioning politics, referring to "#democrats" in combination with "#climateemergency". This was used to blame the democrats for misinformation related to climate change calling out false claims and unrealistic policies.

Throughout the top hashtags and co-occurrence, there was no mention of the heat-waves happening, as this only happened for part of the month and the wider context of climate change was discussed throughout the month, the #heatwave hashtag fell below the top 10 but hit 18th in top hashtags with 1,173 tweets regarding this topic and didn't make it to the top 50 hashtag co-occurrences, in comparison #covid19 was the 13th top hashtag with 1,545 tweets.

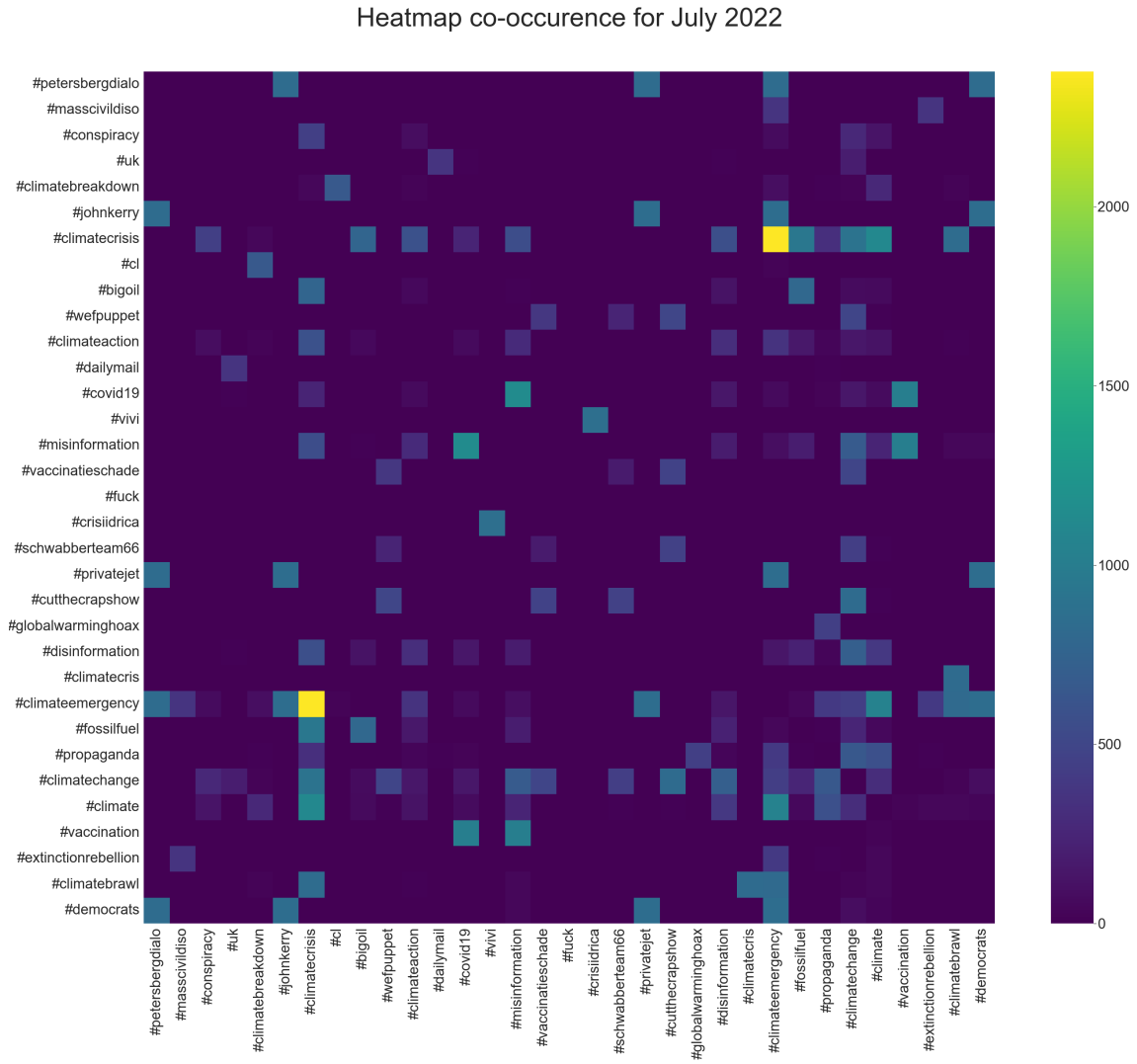


Figure 4.7: Hashtag co-occurrence heatmap July 2022

4.6 July 2022 TF-IDF scores

July 2022	TF-IDF Score
blah	0.977032
eveywhere	0.970219
drip	0.962022
insert	0.961816
overwhelm	0.957587
alright	0.956155
hoax	0.955242
climatehoax	0.952316
pathetic	0.949297
communism	0.93888
funny	0.936451
incoming	0.936316
best	0.935884
butt	0.933228
fuck	0.929842
business	0.929698
fool	0.929664
scare	0.928778
blast	0.928602
hysteria	0.928263
rich	0.927849
cbc	0.927571
disinformation	0.927554
surprise	0.920579
emergency	0.919885
sell	0.919743
record	0.91939
cult	0.919056
save	0.918751
lunaticleftists	0.915453

Figure 4.8: July 2022 TF-IDF scores

July 2022 TF-IDF scores reveal "blah" which has been consistently showing up with a high score, the top 10 tokens lack immediate climate relevance with the exception of "climatehoax". Other tokens involved in the top 10 are neutral terms, "hoax", "communism", "overwhelm" etc, these can be used in many different conspiracies and does not tell us if they are from climate denial or climate believers. Other unique tokens include "lunaticleftists" and "disinformation", these are commonly used by climate deniers as the 'left' are usually suspected to be believing in climate change and environmental issues.

There are no tokens referencing the heatwave exclusively but would have spurred climate discourse along with the other climate events happening worldwide. July 2022 did bring a heightened awareness of climate change media and climate disasters which did increase the amount of climate related tweets posted in July 2022 and I would assume keep increasing into August while dropping in volume going further into winter for western continents.

4.7 Summary of Chapter 4

In this chapter, I have established that there was a clear rise of climate related tweets during the heatwave event primarily focused around Europe, the days that contained heightened volume of tweets were around the 17th-22nd with extremely heightened tweets on the 18th, 19th, 20th and the 21st. The record breaking temperature was recorded on the 19th of July 2022 but the heatwave started on the 18th and the volume dropped significantly after the event and returned back to normal levels.

I have also established a clear theme from different perspectives of what topics were discussed in July 2022 and what topics were the most controversial or engaging for Twitter users during this time. From investigating certain hashtags and ngrams, we can predict which words were used by climate deniers or sceptics and what words were used by climate believers. Even though there was much less Covid-19 crossover in July, we can also derive that Covid-19 sceptics are now furthering their science disbelief into climate change by continuing to use co-occurrence hashtags such as "#covid19" with "#misinformation" while discussing climate topics.

As we do not have any data currently on the next months, it would be useful to see if the downwards trend at the end of July continues or if the upwards trend of the overall months keeps continuing. I would predict that August 2022 would be less than July but still higher than June 2022 and it would decline into Winter as temperatures drop and climate disasters see less coverage by western medias.

Chapter 5

Conclusion

Key findings:

- Covid-19 pandemic had a clear impact on climate related discourse on Twitter, decreasing significantly in the height of the pandemic in 2020.

- Climate related tweets during the Covid-19 pandemic were less focused on climate issues, focusing more on misinformation, disinformation, fake news, mainstream media and political debates.

- Covid-19 sceptics and generic science deniers commonly carried their beliefs into climate science denial as media shifted from covid to climate change, their mindset consistently uses one conspiracy to act as a gateway into other science denying conspiracies.

- Outrage to an event was the most common theme in the majority of months, as emotions become more heated, more people spread the narrative and increase the amount of users participating in sharing their opinions, one example of this was the famous Twitter user Tom Fitton sharing his outrage about adding climate sciences into the schools curriculum.

Originally my plan was to compare flat earth and climate change conspiracies, to find how much of the overall volume was flat earth related compared to climate change and then investigate the crossover between these two conspiracies as they are both environmental and science based conspiracies which have been scientifically proven. After initial exploratory data analysis, flat earth conspiracies seemed too small of

a subset to evaluate against a much more discussed real world and controversial topic such as climate change and global warming. These conspiracies have real world consequences which affect political discourse and human lives worldwide, so I choose climate change because of its impact and relevance.

The goal was then to evaluate climate change discussion over the past 4 years and investigating how the volume was affecting by world events such as the Covid-19 pandemic. This goal was met and my hypothesis was correct, the percentage of climate related tweets in May and June 2019 were 0.42% and 0.34% respectively which dropped significantly in May/June 2020 to 0.13% and 0.10#. While acknowledging the missing data for 2 weeks in June 2020 which still resulted in 55,011,182 misinformation tweets, when compared to May 2020 with 219,123,503 misinformation tweets, the percentages were still similar. It would have been beneficial to collect data closer to the start of the Covid-19 pandemic, around the time of the first covid cases in Europe and US for example, this would be insightful to compare media reactions around climate at this time as I would assume it would be completely overshadowed by covid conspiracies and hysteria.

My next goal was to investigate the discussion involved around climate related tweets in May/June for 4 years between 2019-2022. Using ngrams, top hashtags, hashtag co-occurrences and TF-IDF, I accomplished the most viral narratives each year with each model showing a unique side of Twitter and the real life scenarios happening in that time frame. I believe the most significant achievements from this project was that the methods used were successful in deriving context and applying Natural Language Processing techniques to tell stories of past events from large datasets full of tweets.

We can clearly see when controversial documentaries were released, for example May 2020 when Michael Moore's documentary "Planet of the Humans" was released or when governmental states release life changing policies such as changing the education curriculum in New Jersey to include climate change education starting at age 5, a lot of parents would be happy that their children would learn relevant environmental knowledge but the outrage of a famous figure "Tom Fitton" made this the topic of the month.

I am satisfied with the amount of data collected and analysed even with the 2 weeks missing from June and a couple of days missing from the other months, there was a

huge amount of data to analyse and the results were interesting and fair. If there was more time, I would have like to compare how popular the themes were compared to other months or other themes and also look at more than the top 10, for example, lower, less mainstream discussion.

I also evaluated consistent themes that kept dominating the discourse over the last 4 years, even with the pandemic, words and hashtags containing "climate crisis" was continuously discussed alongside support to help fight the climate crisis or on the other side declare that its a "hoax", "propaganda" or a "climate scam".

If there were no time constraints, I believe doing sentiment analysis on the tweets could have been very insightful, particularly filtering by certain words or hashtags to see if they were being posted in a positive or negative sentiment, broad hashtags like "climatechange" would be a good start and then using it to derive insight from hashtags such as "climatebrawl", "climateemergency" and "climateaction" which were used in a combative manner but sentiment might be able to confirm or deny if that was the case. I would have also liked to compare other conspiracies in the dataset such as Covid-19, the more recent Ukraine/Russia conspiracies and more obscure conspiracies regarding religion or fake moon landings.

I believe my visualisations were a successful way of creating readable and interpretable data but I would have liked to follow certain hashtags or phrases throughout the 4 years to see a more accurate time line of how popular certain themes such as "climatescam" or "climatecrisis" were, this would give information of when the negative or climate denial discourse was heightened. In future studies, location could also be a valuable factor as most of the tweets seemed to come from an Australian or American background and discussing the politics relevant for that location.

Regarding the visualisations used, I didn't want to use complex plots to obscure data but I also believed using bar plots for every table would make it extremely frustrating to extract relevant data and make the report tedious, in the end I used a mix of both, I would have liked the opportunity to create more complex graphics such as a bubble plot which could potentially show sentiments and volume of a certain hashtag, phrase or token.

In future research, there is space for other languages to be analysed around climate discourse, alongside using other forms of social media such as Reddit, 4Chan, Facebook etc. To analyse patterns more fairly, data could also be scraped throughout the

year and not exclusively May/June/July. I would have also liked to use the BERT topic modelling if I had more previous experience implementing it or more time to learn how to conduct it in this research paper.

Chapter 6

Reflection

Throughout my year as a student of MSc Computing and IT Management, I have developed a passion for Python and data science, ultimately setting my sights on starting my career in Data analysis. This project immediately peaked my interest as it was a suitable use of my passions to conduct a report that I would be proud of and would enjoy investigating, as this was my first time using Natural Language Processing and conducting a data analysis report using Python, I believe it has been extremely rewarding and successful.

I have learnt how to use Python libraries to scrape data from Twitter using Tweepy which was ultimately disregarded as the data collected by CSRI was much more extensive and suitable for a large scale analysis. Through extracting the data, I learnt how to manage large datasets, extracted from JSON and combined them with Glob, to be displayed into a Pandas dataframe. These are extremely useful skills for a job in a data focused career and the following cleaning of the data would help me become experienced and resourceful in data management and data cleaning.

During the data cleaning, I was faced with missing data and corrupted data, when I had combined every accessible piece of data, I took the averages of the metrics I needed to fairly compare these metrics.

After cleaning the data, filtering the dataframe and then using exploratory data analysis, I was immediately met with multiple paths to take my investigation, the narrative I choose was the one that intrigued me the most and continued to ensure that I enjoyed the process of finding the answers to tell the narrative.

I did have experience with visualisation libraries such as seaborn and matplotlib from a past MSc module but I have never needed to use custom matrices to map to a dataframe and then ultimately display it as a heatmap, this was extremely difficult for me at first and very rewarding to learn how to handle this problem and solve it.

I have continuously improved my time management skills since the start of this project, with the help of Dr Alun Preece and the other staff members which were assisting me during our bi-weekly meetings I was able to configure my scope of the project to fit the remaining time available. This was very important to me as originally I felt extremely lost and was only thinking of the bigger picture when in reality we only had 10 weeks to complete the project. I also presented my research to the members of the Cardiff Security and Crime Institute and receiving feedback to help me improve my report. Overall I am very satisfied with the amount of work I have completed and the analysis I conducted. To prevent time wasting, I alternated between writing the report and conducting analysis, using this method, if I got stuck or burnt out from one task, I would move on to the other while not leaving all the writing until the end.

During the project, I realised that analysing all 9 months of data would quickly become too extensive for this scale of project and the analysis would be very similar so I decided to cut out 2021 from the overall analysis to include just the most important months, I already conducted the NLP models on this data so I included it in the Appendix.

When I was conducting the main NLP analysis, I had to learn all the NLP techniques and models from the beginning, I used various resources such as CodeCademy, FreeCodeCamp, Youtube, StackOverFlow and Medium to learn the fundamentals of data science and Natural Language Processing. I had a lot of trouble with the Python library Sci-kit learn as the dataset was so large, it would keep crashing the Jupyter Notebook when trying to prepare a TF-IDF matrix. After consultation from my supervisors, they suggested that I remove retweets and replies to just analyse the original tweets. After this advice and all the debugging I experienced, I became very confident using TF-IDF. Alternative ways to fix this would have been to clean a lot of the nonsensical words such as links, usernames, hashtags, numbers etc, this would have also lowered the amount of memory the TF-IDF matrix was occupying.

Throughout the project, I have become confident with NLTK and I would be excited

to continue using these methods involved to clean text data and present it within my career. I also became very familiar with Anaconda and Jupyter Notebooks which are both essential tools for data analysts, I found efficient ways to create a notebook and load different datasets into it instead of repeating the same code, this was used for combining the JSON files and for analysing the individual months.

I believe my weakest asset throughout this project has been my writing skills and literature researching ability, I attempted to conduct a well structured and complete literature background, although I believe I have improved and I am proud of the work, I identify that this part of my skill set is lacking and my written structure still needs improvement.

Overall I also think the presentation of the project could be improved, as a first time user of LaTeX, I found it quite difficult and frustrating to do simple tasks like resizing and positioning images to remove white space and make the report look more aesthetically pleasing. On the other side, I found LaTeX an extremely useful tool for displaying functions, code snippets and formatting the whole document at once, I believe with more experience and learning, LaTeX would be the best tool for writing reports.

I also struggled with creating an appendix and bibliography. When I attempted to use .bib for the bibliography, I encountered many errors, so I resorted back to using Zotero to create a bibliography in word and attaching it as another PDF. Similarly for the appendix, I created that in word and attached it to the bottom. In the end, all goals were met in a timely manner and the process was enjoyable.

References

Abd-Alrazaq, A., Alhuwail, D., Househ, M., Hamdi, M. and Shah, Z. 2020. Top Concerns of Tweeters During the COVID-19 Pandemic: Infoveillance Study. *J Med Internet Res* 22(4), p. e19016. doi: 10.2196/19016.

Ahmed, W., Vidal-Alaball, J., Downing, J. and López Seguí, F. 2020. COVID-19 and the 5G Conspiracy Theory: Social Network Analysis of Twitter Data. *J Med Internet Res* 22(5), p. e19458. doi: 10.2196/19458.

Alam, M.M. and Shome, A. 2020. Attacks on Health Workers during COVID-19 Pandemic - Data Exploration and News Article Detection Using NLP and GRU Model. In: *7th International Conference on Networking, Systems and Security*. 7th NSysS 2020. New York, NY, USA: Association for Computing Machinery, pp. 1–11. Available at: https://doi.org/10.1145/3428363.3428366.

Ansah, E. 2020. Fear Experiences of Social Media Users in Ghana During the COVID-19 Pandemic-Lockdown: An Online Survey. *International Journal of Media and Information Literacy* 5, pp. 199–204. doi: 10.13187/ijmil.2020.2.199.

Antypas, D., Camacho-Collados, J., Preece, A. and Rogers, D. 2021. COVID-19 and Misinformation: A Large-Scale Lexical Analysis on Twitter. In: *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing: Student Research Workshop*. Online: Association for Computational Linguistics, pp. 119–126. Available at: https://aclanthology.org/2021.acl-srw.13.

Ari Sen and Zadrozny, B. 2020. QAnon groups have millions of members on Facebook, documents show. Available at: https://www.nbcnews.com/tech/tech-news/qanon-groups-have-millions-members-facebook-documents-show-n1236317 [Accessed: 21 April 2022].

Bale, J.M. 2007. Political paranoia v. political realism: On distinguishing between bogus conspiracy theories and genuine conspiratorial politics. *Patterns of prejudice* 41(1), pp. 45–60.

Ball, P. and Maxmen, A. 2020. The epic battle against coronavirus misinformation and conspiracy theories. Available at: https://www.nature.com/articles/d41586-020-01452-z [Accessed: 15 July 2022].

BBC 2019. Greenpeace hits back at Trump tweet on climate change denial. *BBC News* 12 March. Available at: https://www.bbc.com/news/world-us-canada-47543905 [Accessed: 7 September 2022].

Bolsen, T. and Druckman, J.N. 2018. Validating Conspiracy Beliefs and Effectively Communicating Scientific Consensus. *Weather, Climate, and Society* 10(3), pp. 453–458. doi: 10.1175/WCAS-D-17-0096.1.

Bovet, A. and Makse, H.A. 2019. Influence of fake news in Twitter during the 2016 US presidential election. *Nature Communications* 10(1), p. 7. doi: 10.1038/s41467-018-07761-2.

Carlson, D.K. 2001. Most Americans believe Oswald conspired with others to kill JFK. *Gallup News Service (April 11)*

Cassino, D. 2016. Trump Supporters More Conspiracy Minded than Other Republicans. Available at: https://portal.fdu.edu/fdupoll-archive/160504/ [Accessed: 27 August 2022].

CDP 2020. 2019-2020 Australian Bushfires. Available at: https://disasterphilanthropy.org/disasters/2019-australian-wildfires/ [Accessed: 6 September 2022].

Chaudhary, M. 2020. TF-IDF Vectorizer scikit-learn. Deep understanding TfidfVectorizer by… | by Mukesh Chaudhary | Medium. Available at: https://medium.com/@cmukesh8688/tf-idf-vectorizer-scikit-learn-dbc0244a911a [Accessed: 10 September 2022].

Clarke, S. 2019. Conspiracy theories and conspiracy theorizing. In: *Conspiracy Theories*. Routledge, pp. 77–92.

Cook, J. et al. 2013. Quantifying the consensus on anthropogenic global warming in the scientific literature. *Environmental Research Letters* 8(2), p. 024024. doi: 10.1088/1748-9326/8/2/024024.

Cook, J. et al. 2016. Consensus on consensus: a synthesis of consensus estimates on human-caused global warming. *Environmental Research Letters* 11(4), p. 048002. doi: 10.1088/1748-9326/11/4/048002.

Copping, L.T. 2022. Anxiety and covid-19 compliance behaviors in the UK: The moderating role of conspiratorial thinking. *Personality and Individual Differences* 192, p. 111604. doi: https://doi.org/10.1016/j.paid.2022.111604.

Daily Mail 2022. BBC climate editor made false claims on global warming for Panorama broadcast, inquiry finds | Daily Mail Online. Available at: https://www.dailymail.co.uk/news/article-10799153/BBC-climate-editor-false-claims-global-warming-Panorama-broadcast-inquiry-finds.html [Accessed: 8 September 2022].

Diethelm, P. and McKee, M. 2009. Denialism: what is it and how should scientists respond? *The European Journal of Public Health* 19(1), pp. 2–4.

Dixon, S. 2022. • Number of social media users 2025. Available at: https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/ [Accessed: 15 July 2022].

Douglas, K.M. and Sutton, R.M. 2015. Climate change: Why the conspiracy theories are dangerous. *Bulletin of the Atomic Scientists* 71(2), pp. 98–106. doi: 10.1177/0096340215571908.

Douglas, K.M., Uscinski, J.E., Sutton, R.M., Cichocka, A., Nefes, T., Ang, C.S. and Deravi, F. 2019. Understanding Conspiracy Theories. *Political Psychology* 40(S1), pp. 3–35. doi: https://doi.org/10.1111/pops.12568.

Evon, D. 2022. No, Weather Maps Aren't Scare-Mongering About Climate Change. Available at: https://www.snopes.com/news/2022/07/29/weather-maps-climate-change/ [Accessed: 29 August 2022].

Facebook 2019. Understanding Facebook's fact-checking programme. Available at: https://en-gb.facebook.com/gpa/blog/misinformation-resources [Accessed: 26 August 2022].

Fernandez, C. 2022. BBC climate editor made false claims on global warming. Available at: https://www.dailymail.co.uk/news/article-10799153/BBC-climate-editor-false-claims-global-warming-Panorama-broadcast-inquiry-finds.html [Accessed: 8 September 2022].

Ferrara, E. 2020. What types of COVID-19 conspiracies are populated by Twitter bots? *arXiv preprint arXiv:2004.09531*

Foz, G.A. 2021. CountVectorizer returning only zeros. Available at: https://stackoverflow.com/q/70537446 [Accessed: 7 September 2022].

Freeman, D. et al. 2020. COVID-19 vaccine hesitancy in the UK: the Oxford coronavirus explanations, attitudes, and narratives survey (Oceans) II. *Psychological Medicine* , pp. 1–15. doi: 10.1017/S0033291720005188.

Freeman, D. et al. 2022. Coronavirus conspiracy beliefs, mistrust, and compliance with government guidelines in England. *Psychological Medicine* 52(2), pp. 251–263. doi: 10.1017/S0033291720001890.

Funk, M. and Speakman, B. 2022. Setting a Q-uestionable attribute agenda: QAnon, far-right congressional candidates and irrational domains. *The Agenda Setting Journal* . doi: 10.1075/asj.21004.fun.

Gao, J. et al. 2020. Mental health problems and social media exposure during COVID-19 outbreak. *PLOS ONE* 15(4), p. e0231924. doi: 10.1371/journal.pone.0231924.

Geißler, E. and Sprinkle, R. 2014. Disinformation squared Was the HIV-from-Fort-Detrick myth a Stasi success? *Politics and the life sciences : the journal of the Association for Politics and the Life Sciences* 32, pp. 2–99. doi: 10.2990/32_2_2.

Gironde 2022. Fires in Landiras and La Teste-de-Buch | Gironde.FR. Available at: https://www.gironde.fr/actualites/incendies-landiras-et-la-teste-de-buch [Accessed: 10 September 2022].

Goertzel, T. 1994. Belief in conspiracy theories. *Political psychology* , pp. 731–742.

Gruzd, A. and Mai, P. 2020. Going viral: How a single tweet spawned a COVID-19 conspiracy theory on Twitter. *Big Data & Society* 7, p. 205395172093840. doi: 10.1177/2053951720938405.

He, J., He, L., Zhou, W., Nie, X. and He, M. 2020. Discrimination and Social Exclusion in the Outbreak of COVID-19. *International Journal of Environmental Research and Public Health* 17(8). Available at: https://www.mdpi.com/1660-4601/17/8/2933.

Healy, J. 2021. These Are the 5 People Who Died in the Capitol Riot. *The New York Times* 11 January. Available at: https://www.nytimes.com/2021/01/11/us/who-died-in-capitol-building-attack.html [Accessed: 1 May 2022].

Imhoff, R. and Lamberty, P. 2020. A Bioweapon or a Hoax? The Link Between Distinct Conspiracy Beliefs About the Coronavirus Disease (COVID-19) Outbreak and Pandemic Behavior. *Social Psychological and Personality Science* 11, p. 194855062093469. doi: 10.1177/1948550620934692.

Islam, M.S. et al. 2020. COVID-19–related infodemic and its impact on public health: A global social media analysis. *The American journal of tropical medicine and hygiene* 103(4), p. 1621.

Jacques, P.J. 2012. A general theory of climate denial. *Global Environmental Politics* 12(2), pp. 9–17.

Jacques, P.J., Dunlap, R.E. and Freeman, M. 2008. The organisation of denial: Conservative think tanks and environmental scepticism. *Environmental politics* 17(3), pp. 349–385.

Jensen, T. 2013. Democrats and Republicans differ on conspiracy theory beliefs. Available at: https://www.publicpolicypolling.com/polls/democrats-and-republicans-differ-on-conspiracy-theory-beliefs/ [Accessed: 27 August 2022].

Jolley, D. and Douglas, K.M. 2014. The social consequences of conspiracism: Exposure to conspiracy theories decreases intentions to engage in politics and to reduce one's carbon footprint. *British Journal of Psychology* 105(1), pp. 35–56. doi: 10.1111/bjop.12018.

Karabiber, F. [no date]. TF-IDF — Term Frequency-Inverse Document Frequency – LearnDataSci. Available at: https://www.learndatasci.com/glossary/tf-idf-term-frequency-inverse-document-frequency/ [Accessed: 10 September 2022].

Kearney, M., Chiang, S. and Massey, P. 2020. The Twitter origins and evolution of the COVID-19 "plandemic" conspiracy theory. *Harvard Kennedy School Misinformation Review* 1. doi: 10.37016/mr-2020-42.

Keeley, B.L. 1999. Of Conspiracy Theories. *The Journal of Philosophy* 96(3), pp. 109–126. doi: 10.2307/2564659.

Lawton, G. [no date]. Conspiracy theories. Available at: https://www.newscientist.com/definition/conspiracy-theories/ [Accessed: 25 August 2022].

Lewandowsky, S. 2014. Conspiratory fascination versus public interest: the case of 'climategate.' *Environmental Research Letters* 9(11), p. 111004. doi: 10.1088/1748-9326/9/11/111004.

Lewandowsky, S., Ecker, U.K.H., Seifert, C.M., Schwarz, N. and Cook, J. 2012. Misinformation and Its Correction: Continued Influence and Successful Debiasing. *Psychological Science in the Public Interest* 13(3), pp. 106–131. doi: 10.1177/1529100612451018.

Lewandowsky, S., Gignac, G.E. and Oberauer, K. 2015. Correction: The Role of Conspiracist Ideation and Worldviews in Predicting Rejection of Science. *PLOS ONE* 10(8), p. e0134773. doi: 10.1371/journal.pone.0134773.

Lewandowsky, S. and Oberauer, K. 2016. Motivated Rejection of Science. *Current Directions in Psychological Science* 25(4), pp. 217–222. doi: 10.1177/0963721416654436.

Lewandowsky, S., Oberauer, K. and Gignac, G.E. 2013. NASA Faked the Moon Landing—Therefore, (Climate) Science Is a Hoax: An Anatomy of the Motivated Rejection of Science. *Psychological Science* 24(5), pp. 622–633. doi: 10.1177/0956797612457686.

van der Linden, S. 2015. The conspiracy-effect: Exposure to conspiracy theories (about global warming) decreases pro-social behavior and science acceptance. *Personality and Individual Differences* 87, pp. 171–173. doi: 10.1016/j.paid.2015.07.045.

Lusa 2022. Portugal hits 47°C - The Portugal News. Available at: https://www.theportugalnews.com/news/2022-07-15/portugal-hits-47c/68702 [Accessed: 10 September 2022].

M. Alassad, M. N. Hussain, and N. Agarwal 2020. How to Control Coronavirus Conspiracy Theories in Twitter? A Systems Thinking and Social Networks Modeling Approach. In: *2020 IEEE International Conference on Big Data (Big Data)*., pp. 4293–4299. doi: 10.1109/BigData50022.2020.9378400.

McKee, M. and Diethelm, P. 2010. How the growth of denialism undermines public health. *Bmj* 341

Met Office 2022a. UK prepares for historic hot spell - Met Office. Available at: https://www.metoffice.gov.uk/about-us/press-office/news/weather-and-climate/2022/red-extreme-heat-warning [Accessed: 18 September 2022].

Met Office 2022b. Unprecedented extreme heatwave, July 2022 - Met Office. Available at: https://www.metoffice.gov.uk/binaries/content/assets/metofficegovuk/pdf/weather/learn-about/uk-past-events/interesting/2022/2022_03_july_heatwave.pdf [Accessed: 10 September 2022].

Meteo France 2022. Canicule intense et durable de juillet 2022 : que faut-il retenir ? | Météo-France. Available at: https://meteofrance.com/actualites-et-dossiers/actualites/canicule-intense-et-durable-de-juillet-2022-que-faut-il-retenir [Accessed: 10 September 2022].

Mulukom, V. van et al. 2022. Antecedents and consequences of COVID-19 conspiracy beliefs: A systematic review. *Social Science & Medicine* 301, p. 114912. doi: https://doi.org/10.1016/j.socscimed.2022.114912.

Nattrass, N. 2013. Understanding the origins and prevalence of AIDS conspiracy beliefs in the United States and South Africa. *Sociology of Health & Illness* 35(1), pp. 113–129. doi: 10.1111/j.1467-9566.2012.01480.x.

Newburger, E. 2020. Earth has hottest May on record, with 2020 on track to be one of the top 10 warmest years. Available at: https://www.cnbc.com/2020/06/05/climate-change-may-2020-is-hottest-month-on-record.html [Accessed: 6 September 2022].

Nisbet, M.C. and Teresa Myers 2007. Trends: Twenty Years of Public Opinion about Global Warming. *The Public Opinion Quarterly* 71(3), pp. 444–470.

Oleksy, T., Wnuk, A., Maison, D. and Łyś, A. 2021. Content matters. Different predictors and social consequences of general and government-related conspiracy theories on COVID-19. *Personality and Individual Differences* 168, p. 110289. doi: 10.1016/j.paid.2020.110289.

Oliver, J.E. and Wood, T. 2014. Medical Conspiracy Theories and Health Behaviors in the United States. *JAMA Internal Medicine* 174(5), pp. 817–818. doi: 10.1001/jamainternmed.2014.190.

Oost, P.V. et al. 2022. The relation between conspiracism, government trust, and COVID-19 vaccination intentions: The key role of motivation. *Social Science & Medicine* 301, p. 114926. doi: https://doi.org/10.1016/j.socscimed.2022.114926.

Pratt, S. 2022. A July of Extremes. Available at: https://earthobservatory.nasa.gov/images/150152/a-july-of-extremes [Accessed: 10 September 2022].

Preece, A., Spasic, I., Evans, K., Rogers, D., Webberley, W., Roberts, C. and Innes, M. 2017. Sentinel: A Codesigned Platform for Semantic Enrichment of Social Media Streams. *IEEE Transactions on Computational Social Systems* PP, pp. 1–14. doi: 10.1109/TCSS.2017.2763684.

Readfearn, G. 2022. Sky News Australia is a global hub for climate misinformation, report says | Sky News Australia | The Guardian. Available at: https://www.theguardian.com/media/2022/jun/14/sky-news-australia-is-a-global-hub-for-climate-misinformation-report-says [Accessed: 8 September 2022].

Ripp, T. and Röer, J.P. 2022. Systematic review on the association of COVID-19-related conspiracy belief with infection-preventive behavior and vaccination willingness. *BMC Psychology* 10(1), p. 66. doi: 10.1186/s40359-022-00771-2.

Roberto, K.J., Johnson, A.F. and Rauhaus, B.M. 2020. Stigmatization and prejudice during the COVID-19 pandemic. *Administrative Theory & Praxis* 42(3), pp. 364–378. doi: 10.1080/10841806.2020.1782128.

Simelela, N., Venter, W.D.F., Pillay, Y. and Barron, P. 2015. A Political and Social History of HIV in South Africa. *Current HIV/AIDS Reports* 12(2), pp. 256–261. doi: 10.1007/s11904-015-0259-7.

Sky News 2020. Planet of the Humans: Michael Moore green energy documentary branded "dangerous" by climate scientists | Climate News | Sky News. Available at: https://news.sky.com/story/planet-of-the-humans-michael-moore-green-energy-documentary-branded-dangerous-by-climate-scientists-11980420 [Accessed: 7 September 2022].

Sky News AU 2020. Michael Moore's documentary 'has exposed green energy as a fraud' | Sky News Australia. Available at: https://www.skynews.com.au/australia-news/michael-moores-documentary-has-exposed-green-energy-as-a-fraud/video/c3f9174f99a51412a4164f0978641308 [Accessed: 7 September 2022].

Spring, M. 2021. Covid denial to climate denial: How conspiracists are shifting focus - BBC News. Available at: https://www.bbc.co.uk/news/blogs-trending-59255165.amp [Accessed: 8 September 2022].

Sunstein, C.R. and Vermeule, A. 2009. Conspiracy Theories: Causes and Cures*. *Journal of Political Philosophy* 17(2), pp. 202–227. doi: 10.1111/j.1467-9760.2008.00325.x.

Sussman, A.B. 2010. *A veteran meteorologist exposes the global warming scam*. New York: WND Books.

Swami, V. 2012. Social Psychological Origins of Conspiracy Theories: The Case of the Jewish Conspiracy Theory in Malaysia. *Frontiers in Psychology* 3. Available at: https://www.frontiersin.org/article/10.3389/fpsyg.2012.00280.

Swami, V., Nader, I.W., Pietschnig, J., Stieger, S., Tran, U.S. and Voracek, M. 2012. Personality and individual difference correlates of attitudes toward human rights and civil liberties. *Personality and Individual Differences* 53(4), pp. 443–447.

Swami, V., Voracek, M., Stieger, S., Tran, U.S. and Furnham, A. 2014. Analytic thinking reduces belief in conspiracy theories. *Cognition* 133(3), pp. 572–585. doi: 10.1016/j.cognition.2014.08.006.

Tuxworth, D., Antypas, D., Anke, L.E., Camacho-Collados, J., Preece, A.D. and Rogers, D. 2021. Deriving Disinformation Insights from Geolocalized Twitter Callouts. *ArXiv* abs/2108.03067

Uscinski, J.E., Douglas, K. and Lewandowsky, S. 2017. Climate Change Conspiracy Theories. Available at: https://oxfordre.com/climatescience/view/10.1093/acrefore/9780190228620.001.0001/acrefore-9780190228620-e-328.

Uscinski, J.E. and Parent, J.M. 2014. *American Conspiracy Theories*. Oxford University Press. Available at: https://doi.org/10.1093/acprof:oso/9780199351800.001.0001 [Accessed: 27 August 2022].

Vermeule, C.A. and Sunstein, C.R. 2009. Conspiracy theories: causes and cures. *Journal of Political Philosophy*

Wang, X., Zhang, M., Fan, W. and Zhao, K. 2022. Understanding the spread of COVID-19 misinformation on social media: The effects of topics and a political leader's nudge. *Journal of the Association for Information Science and Technology* 73(5), pp. 726–737. doi: https://doi.org/10.1002/asi.24576.

Weller, K., Bruns, A., Burgess, J., Mahrt, M. and Puschmann, C. 2013. *Twitter and Society*. New York, United States of America: Peter Lang Verlag. Available at: https://www.peterlang.com/document/1109452.

Wood, M.J., Douglas, K.M. and Sutton, R.M. 2012. Dead and alive: Beliefs in contradictory conspiracy theories. *Social Psychological and Personality Science* 3(6), pp. 767–773. doi: 10.1177/1948550611434786.

Zarocostas, J. 2020. How to fight an infodemic. *The Lancet* 395(10225), p. 676. doi: https://doi.org/10.1016/S0140-6736(20)30461-X.

Appendix

Appendix 1:

Misinformation search terms for callout dataset:

Fake news, Propaganda, Disinformation, Active measures, Subversion, Interference, Influence, Conspiracy, Deep state, Misinformation, Fabrication, Manipulate, Deceive, Useful idiots, Mainstream media, Populism, Untrustworthy, Hoax, Made-up, Bogus, Inaccurate, Doctored, Fact Checking, eu False, eu Fraud, eu Hoax, eu Lies, eu Rumours, eu Troll, europe False, europe Fraud, europe Hoax, europe Lies, europe Rumours, europe Troll, european False, european Fraud, european Hoax, european Lies, european Rumours, european Troll

Appendix 2

2021 Ngrams:

Unigrams (BOW)	May 2021	June 2021
1st	climate, 246396	climate, 298587
2nd	propaganda,109726	change, 179273
3rd	change, 82195	propaganda, 133473
4th	misinformation, 44380	conspiracy, 102769
5th	conspiracy, 43796	think, 54092
6th	country, 32511	time, 52178
7th	false, 32310	next, 48127
8th	free, 31861	story, 47652
9th	racial, 31739	prove, 47286
10th	social, 31712	get, 47200

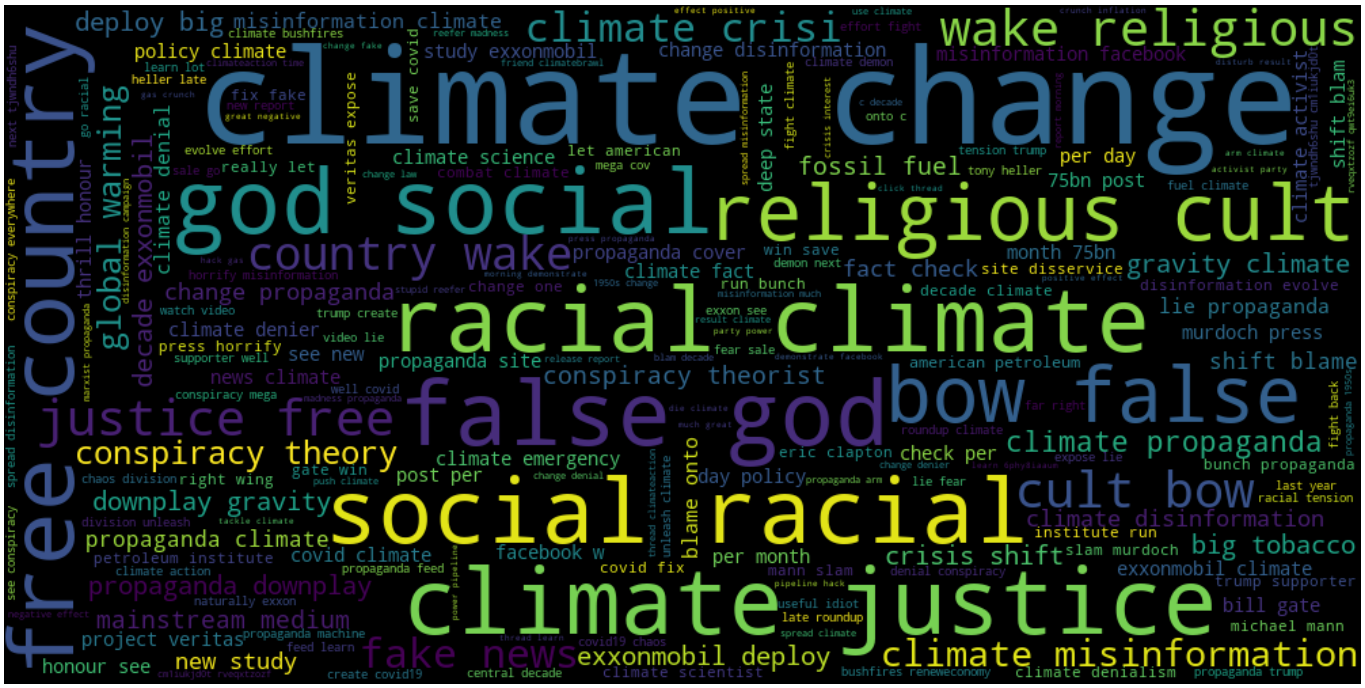
Bigrams	May 2021	June 2021
1st	('climate', 'change'), 75632	('climate', 'change'), 174152
2nd	('climate', 'justice'), 29507	('time', 'climate'), 44429
3rd	('racial', 'climate'), 29136	('next', 'time'), 44355
4th	('false', 'god'), 29084	('change', 'remember'), 44303
5th	('social', 'racial'), 29084	('conspiracy', 'story'), 44300
6th	('god', 'social'), 29070	('remember', 'conspiracy'), 44285
7th	('free', 'country'), 29044	('lock', 'next'), 44264
8th	('religious', 'cult'), 29036	('story', 'censor'), 44250
9th	('bow', 'false'), 29020	('censor', 'prove'), 44250
10th	('wake', 'religious'), 29006	('prove', 'lock'), 34195

Trigrams	May 2021	June 2021
1st	('social', 'racial', 'climate'), 29084	('time', 'climate', 'change'), 44398
2nd	('racial', 'climate', 'justice'), 29084	('climate', 'change', 'remember'), 44282
3rd	('false', 'god', 'social'), 29070	('change', 'remember', 'conspiracy'), 44266
4th	('god', 'social', 'racial'), 29070	('lock', 'next', 'time'), 44250
5th	('bow', 'false', 'god'), 29020	('next', 'time', 'climate'), 44250
6th	('wake', 'religious', 'cult'), 29006	('remember', 'conspiracy', 'story'), 44250
7th	('religious', 'cult', 'bow'), 29006	('conspiracy', 'story', 'censor'), 44250
8th	('cult', 'bow', 'false'), 29006	('story', 'censor', 'prove'), 44250
9th	('climate', 'justice', 'free'), 29006	('censor', 'prove', 'lock'), 34195
10th	('justice', 'free', 'country'), 29006	('prove', 'lock', 'next'), 34195

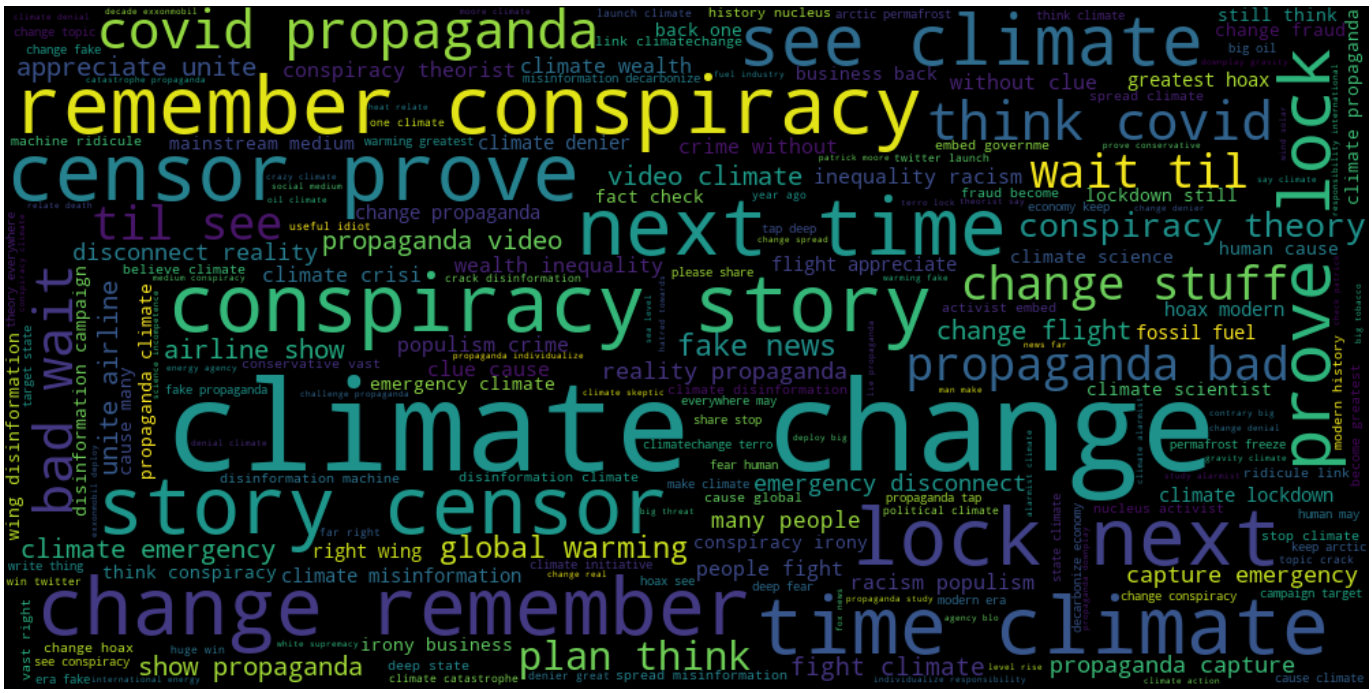
Appendix 3

2021 Word clouds:

May 2021



June 2021



Appendix 4

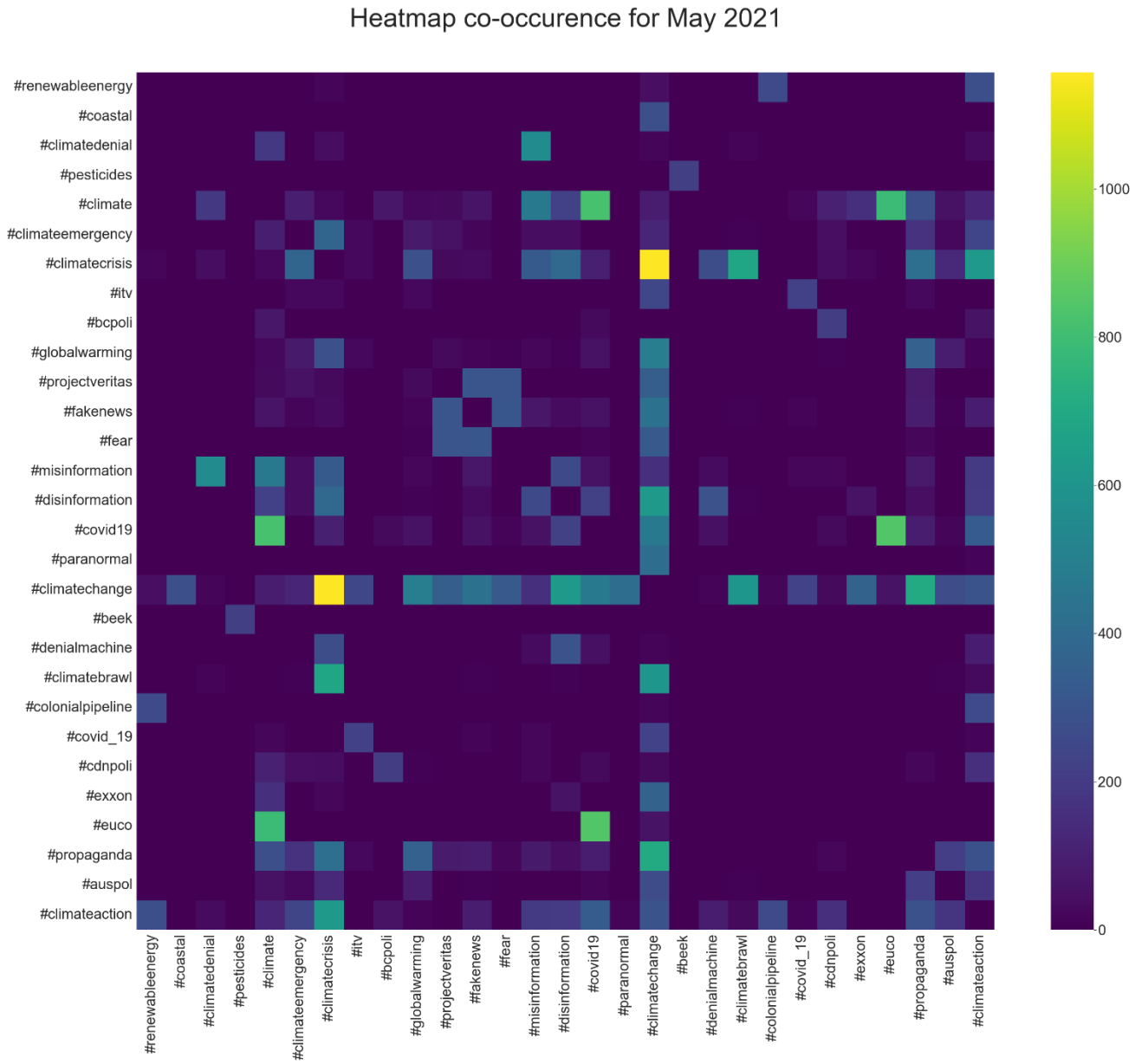
2021 hashtag co-occurrence:

Co-occurence hashtags	May 2021	June 2021
1st	('#climatechange', '#climatecrisis'), 1157	('#climatechange', '#climatecrisis'), 1277
2nd	('#covid19', '#euco'), 844	('#climatebrawl', '#climatecrisis'), 1243
3rd	('#climate', '#covid19'), 822	('#g7cor', '#g7summit'), 961
4th	('#climate', '#euco'), 788	('#climatebrawl', '#climatechange'), 753
5th	('#climatechange', '#propaganda'), 701	('#climateaction', '#climatecrisis'), 639
6th	('#climatebrawl', '#climatecrisis'), 684	('#climatebrawl', '#sciencematters'), 495
7th	('#climateaction', '#climatecrisis'), 628	('#climate', '#misinformation'), 487
8th	('#climatechange', '#disinformation'), 626	('#clim', '#climatebrawl'), 456
9th	('#climatebrawl', '#climatechange'), 609	('#climatechange', '#misinformation'), 452
10th	('#climatedenial', '#misinformation'), 552	('#climatechange', '#paranormal'), 441

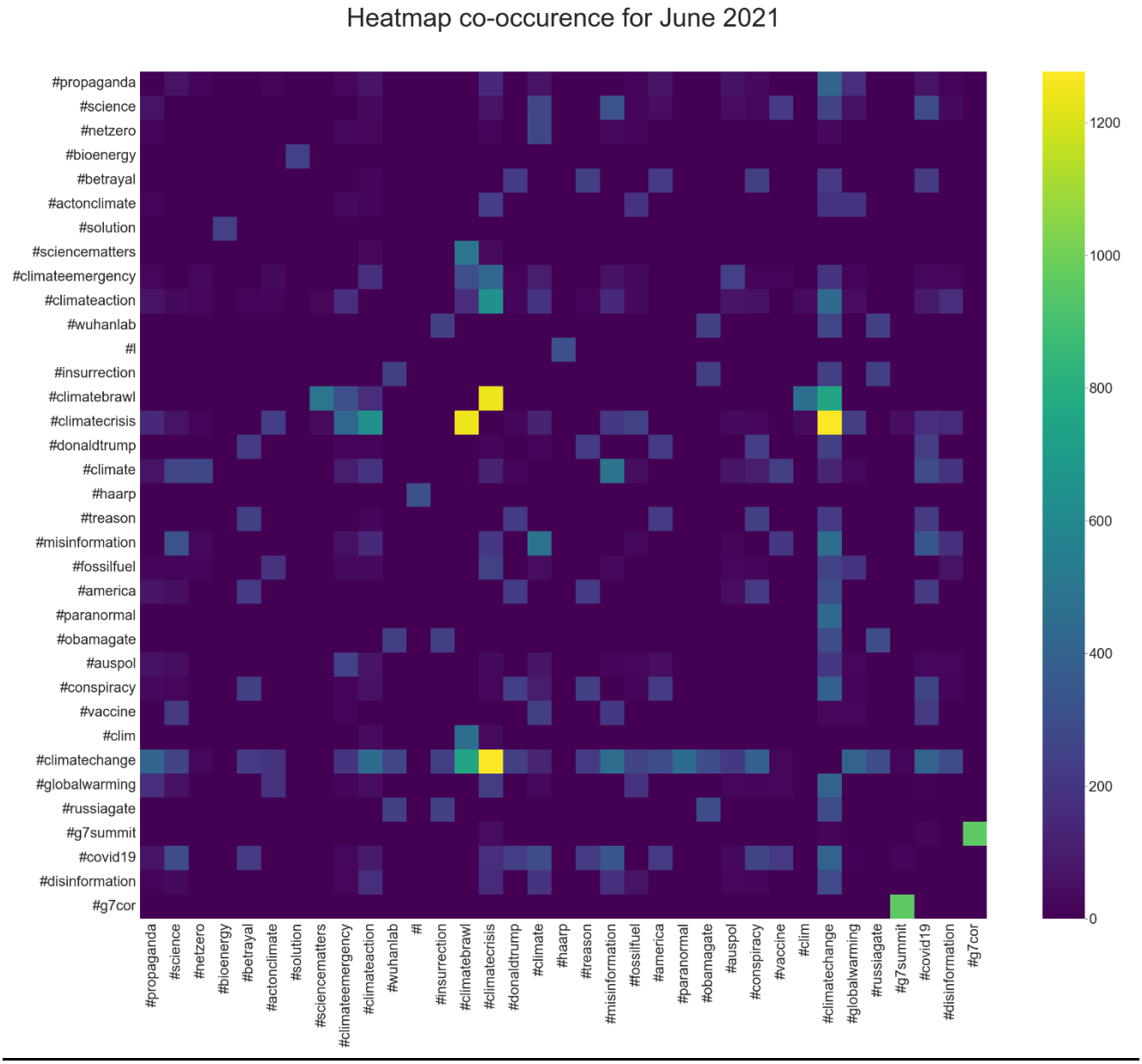
Appendix 5

2021 Heatmaps:

May 2021



June 2021


Heatmap co-occurence for June 2021
#propaganda
#science
#netzero
#bioenergy
#betrayal
#actonclimate
#solution
#sciencematters
#climateemergency
#climateaction
#wuhanlab
#l
#insurrection
#climatebrawl
#climatecrisis
#donaldtrump
#climate
#haarp
#treason
#misinformation
#fossilfuel
#america
#paranormal
#obamagate
#auspol
#conspiracy
#vaccine
#clim
#climatechange
#globalwarming
#russiagate
#g7summit
#covid19
#disinformation
#g7cor
1200
1000
800
600
400
200
0

Appendix 6

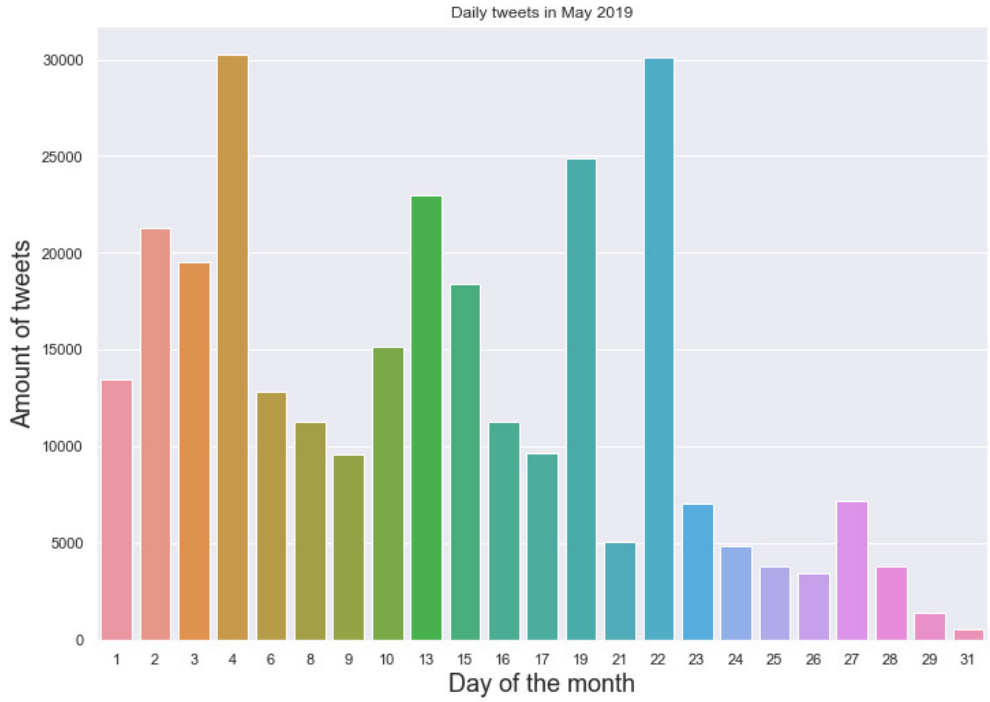
2021 TF-IDF scores:

May 2021	TF-IDF Score
climatescam	0.956016
heartwarming	0.940708
defundthebbc	0.923312
conspiracy	0.917775
expect	0.892475
surely	0.886787
climatecrisis	0.882737
ban	0.842527
criticise	0.820446
ponder	0.81728
refugee	0.816819
apos	0.812597
clickbait	0.798811
care	0.78921
scam	0.78758
catastrophe	0.780989
hoax	0.778766
cnn	0.766948
alarmism	0.761115
plan	0.742456
weather	0.740905
wake	0.739916
chiller	0.739287
dose	0.738425
relentless	0.733515
unhealthy	0.73302
commercial	0.730387
threat	0.728458
big	0.727851
fake	0.725554

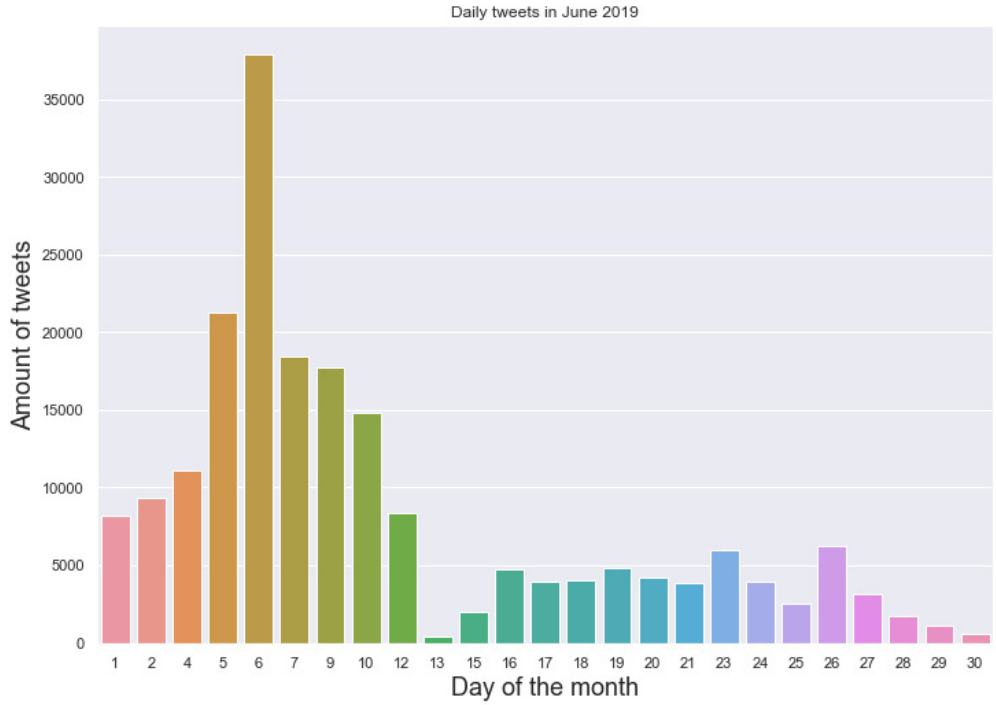
June 2021	TF-IDF Score
blah	0.977474
climatecult	0.963469
misinformation	0.92758
cult	0.91015
nut	0.908037
subject	0.904019
hoax	0.870551
drill	0.866933
anti	0.86544
factcheck	0.864172
love	0.864113
micro	0.849886
silence	0.845268
gop	0.843308
question	0.837773
climaterealists	0.833036
utc	0.832908
climatelockdown	0.826717
baseless	0.821707
amiright	0.803754
spread	0.803731
forever	0.798172
anthropogenic	0.792445
loki	0.789583
incoming	0.788839
real	0.78364
isnt	0.782913
shift	0.779504
drivel	0.778827
smh	0.778249

Appendix 7

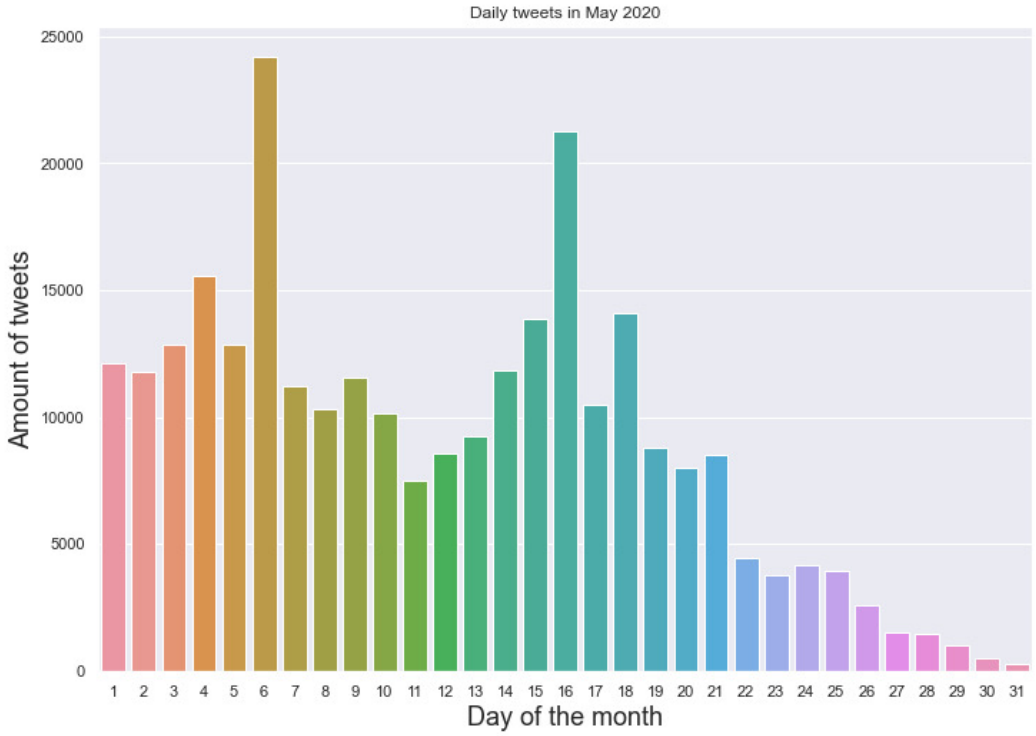
May 2019 Daily climate tweets:



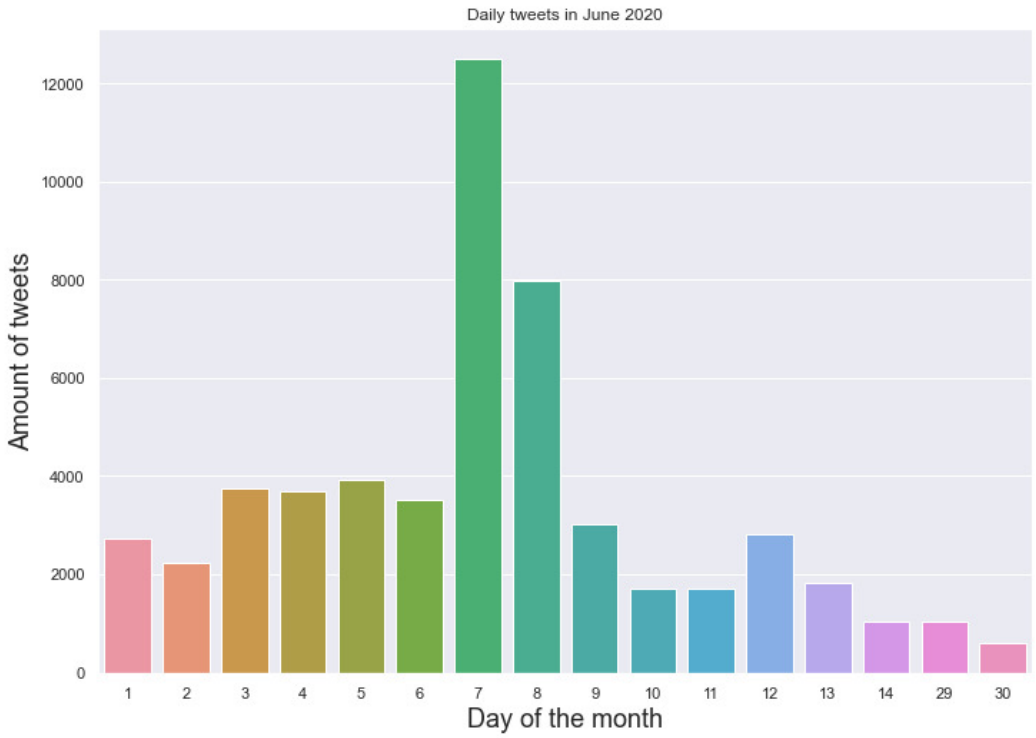
June 2019



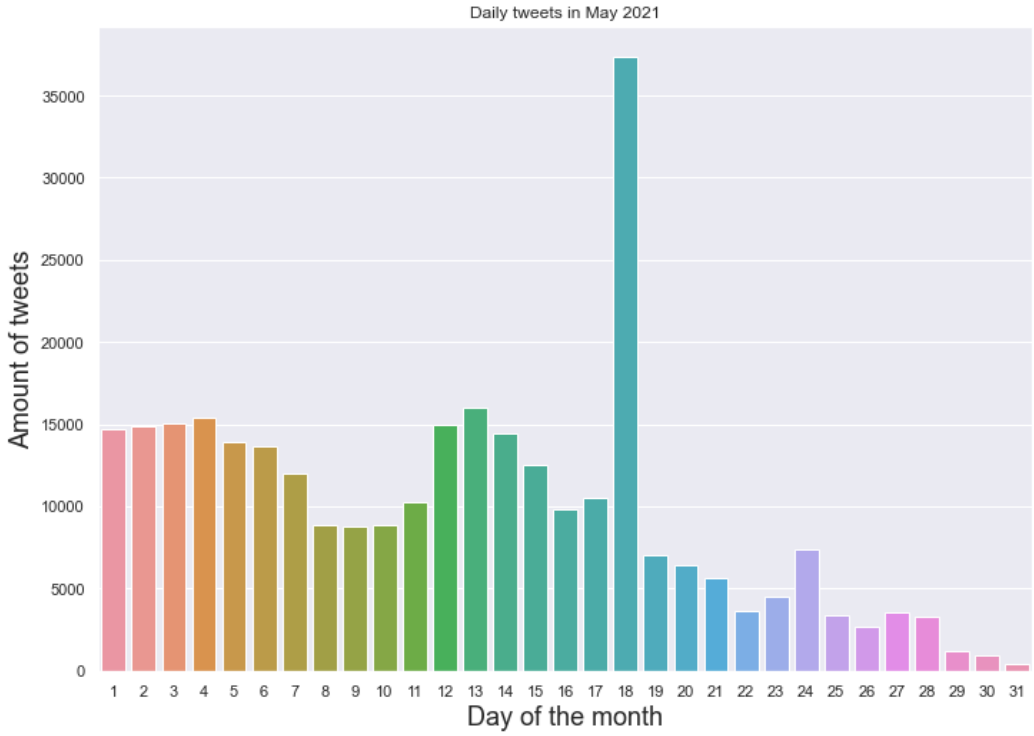
May 2020



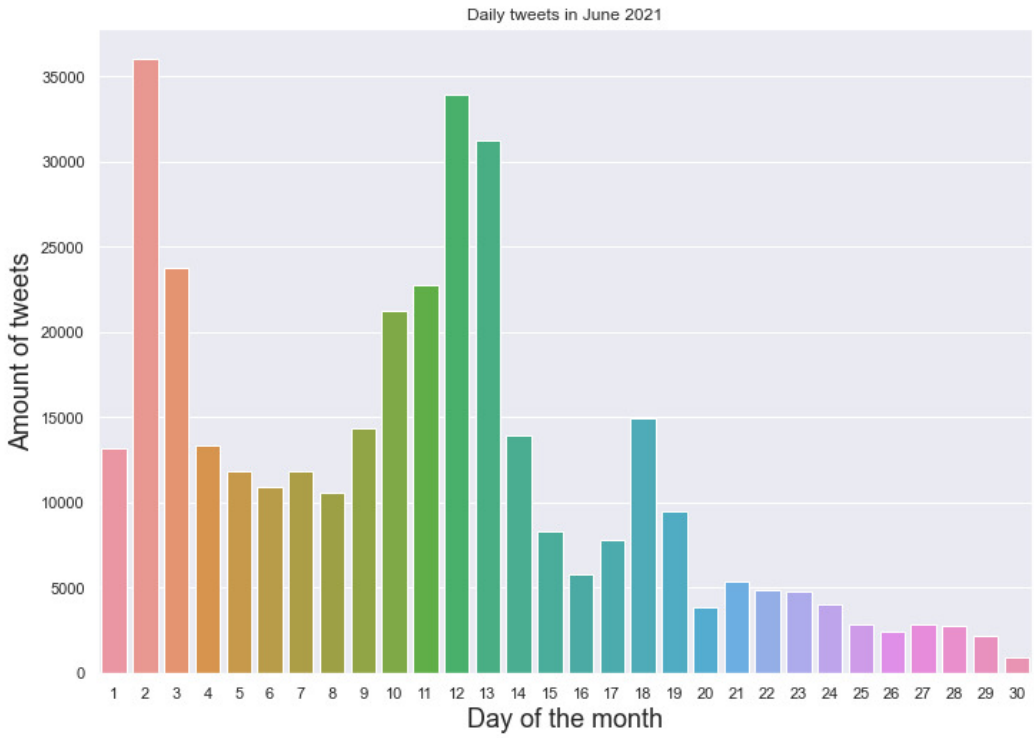
June 2020



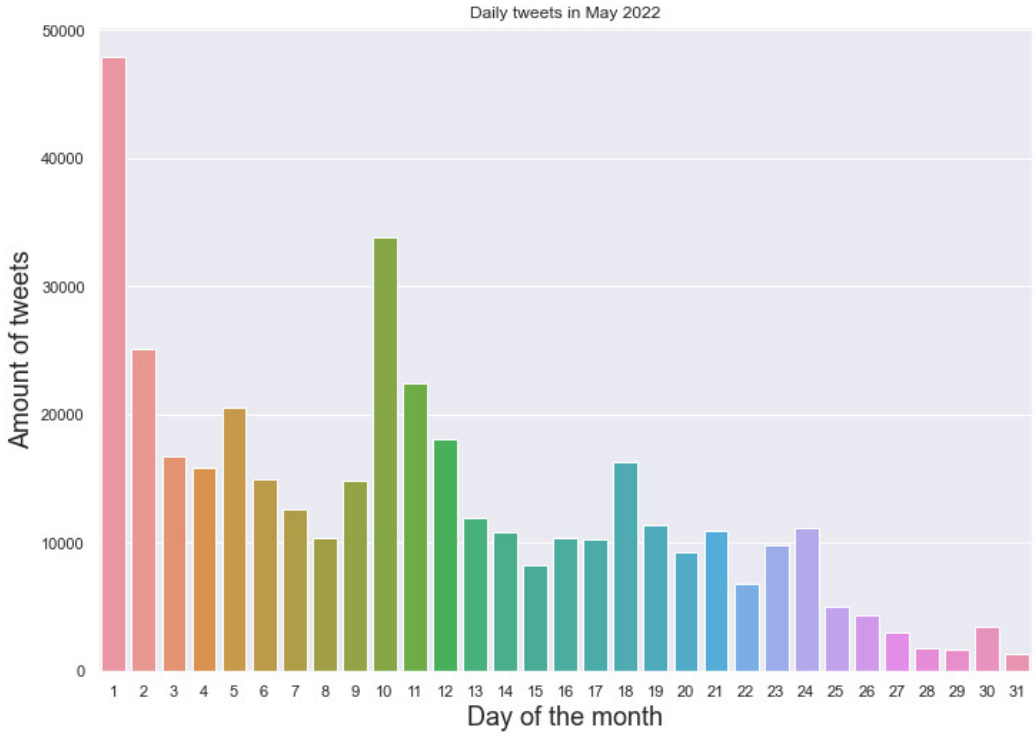
May 2021



June 2021



May 2022



June 2022