Erasmus University Rotterdam

Erasmus School of Economics

MSc Economics and Business: specialization in Behavioral Economics (Marketing track)

Thesis title: Do longer lockdowns have an impact on depression?

Insights of what Twitter can reveal for us in California & Florida states before and during lockdown periods.

Evangelina Settecase (student number: 576252)

Supervisor: Dr. Aurélien Baillon
Second assessor: Dr. Sophie van der Zee

Date: 6th of March 2023

Abstract

The outbreak of the COVID-19 pandemic has led to a clear deterioration in people's mental health. Since then, tools that allow monitoring the evolution of individuals' feelings have been increasingly gaining ground. Twitter, like many other social media platforms, played a significant role as it is widely used to voice opinions and thoughts with others. This study contributes to the growing corpora of research that uses social media platforms to analyze people´s mental health across the pandemic phases. The main aim is to shed some light on the effects of government responses on citizens' well-being. This study uses a sample of more than 4,000 tweets from seventeen depressed individuals based in the states of Florida and California. To spot differences across pandemic phases and states MANOVA and post-hoc tests are performed. Results are presented with descriptive statistics as the methods used shows the existence of differences (MANOVA test) and which variables are responsible for the differences (post-hoc test). This research shows that people become more active on Twitter at night after the lockdown restrictions were put in place (Stay at Home orders). Moreover, negative emotions were stronger before the lockdown's implementation, which can be associated with higher levels of uncertainty and fear in the early stages of the spread of the virus.

Table of Contents

1. Introduction

The irruption of the COVID-19 pandemic has deteriorated the global scenario of mental health, leading to a 25% increase in the prevalence of both depression and anxiety (World Health Organization, 2022). Considering that depression already affected 5% of adults around the globe, one of the consequences of the pandemic is the increase of this figure up to 6,25% of adults worldwide (World Health Organization, 2021). This virus formally declared as a pandemic on March 11th, 2020, includes diverse consequences that help the exacerbation of poor mental health. Among the main drawbacks of the pandemic are lockdowns imposed by governments to minimize or delay the spread of the virus, job losses, reduction in household income, loneliness, the death of beloved ones, and fear of infection, which are associated with an increased in poor population mental health (Ettman et al., 2022; World Health Organization, 2022).

Having relevant and updated information on the current mental health status of the population has become an imperative condition for health institutions to respond accordingly (Santomauro, 2021). However, traditional methods to diagnose depression (questionnaires and self-reporting) depend on the willingness to collaborate with the patients, which in many cases do not want to be assisted (Varghese Babu and Kanaga, 2021). As people express their feelings and thoughts with friends and relatives through various social media, many researchers have focused on alternative ways of detecting depression that does not involve traditional psychologist/psychiatrist appointments. Many studies use people's posts on social media as a way of analyzing their behavior or even anticipating a possible future depression diagnosis through how people write on the media, using techniques such as sentiment analysis (De Choudhury et al., 2013, Chandra Guntuku et al., 2017, Varghese Babu et al., 2021, among others). The consequences of the pandemic on the population's mental health -as with any other traumatic situation- can be seen in the language used on social media (Monzani, Vergani, Pizzoli, Marton, and Pravettoni, 2021).

Until now, many publications had emerged to support the idea that the irruption of the pandemic had increased levels of depression worldwide (Wang, Fan, Palacios, Chai, Guetta-Jeanrenaud, Obradovich, Zhou, and Zheng, 2022; Zhou, Zogan, Yang, Jameel, Xu, and Chen, 2021; Safa, Bayat, and Moghtader, 2022). This is mainly supported by the idea that at the beginning, when the vaccine was not available, both governments and citizens had to deal with uncertainty and the fear

of the unknown it entails. There is a growing amount of research that compares country performance on depression level during lowdown periods. However, to the best of my knowledge, little has been done at state (province) levels.

This thesis aims to analyze depression-related patterns in tweets over time of people who have been diagnosed with depression or Post-Traumatic Stress Disorder (PTSD) after the outbreak of COVID-19 in two states of The United States: California and Florida, where the duration of lockdown has been different. The main hypothesis is that individuals who live in a state which had implemented a longer quarantine showed more intense (negative) emotions compared with those who lived in a state with lighter restrictions. This hypothesis is sustained by the idea that longer lockdown periods trigger many depression symptoms, such as looniness feelings, job losses -with the economics and psychological implications it entails-, marital problems -which also increases the possibility of changing from a family home to a mono-parental home-, and the like. Therefore, people who live in a state with a longer quarantine might be more likely to experience depression symptoms sooner due to the mentioned situations that might appear because of the imposed restrictions.

To address the mentioned hypothesis, the research question of this thesis is the following: *Do people who live in California (a state with a longer quarantine period) show more depressed signals than those who live in Florida (a state with a shorter lockdown period) during the lockdown?* Additionally, this thesis will also analyze if there are any differences in the way people expressed themselves in social media before and during the lockdown in the mentioned states.

To answer the proposed questions, the content of the tweets of people who have been diagnosed with depression or PTSD after the outbreak of the COVID-19 pandemic (April – July 2020) are going to be analyzed. The sample used only considers people from California and Florida. The comparison of the tweets content is going to be done with the help of the Linguistic Inquiry and Word Count (LIWC) software and it will include both comparisons across states, and before (since January 2020) and during quarantine periods (Stay at Home orders in the US). The differences among those groups are going to be evaluated using a MANOVA test and two post-hoc tests that allow making more precise conclusions.

California and Florida are not governed by the same party, being the first one a democrat state while the second one is republican, and are far from each other. However, they have been selected as treatment and control groups respectively for this study as they share many similarities. In terms of population density, California is the most populated state of the country, and Florida is in third place, with 39,6 and 21,9 million habitants in 2022.[1] They also have similar conditions, being both coastal locations and having a wide variety of recreational and cultural activities, conditions that - in case of differences- might contribute (or not) to creating more depressive settings.

In Florida, the Stay-At-Home (SAH) restriction lasted for 27 days, from April 3rd to April 30th of 2020 while in California it encompassed 88 days, from March 19th until June 15th of the same year.[2] The most populated state of the country was the first one to establish a 'Stay-At-Home' order, demanding all residents to remain home when they are not part of the sixteen critical sectors identified by the national government (that entails health, nutrition, supply chain workers and the like). Citizens were allowed to do groceries and buy medicines while keeping social distance among individuals (Executive Department State of California, 2022). On the other hand, Florida was the 34th state to declare such an act when Governor Ron DeSantis communicates the measure on the 1st of April stating that individuals with a 'significant underlying medical condition' must stay at home and only senior citizens involved in obtaining or providing essential services and activities are allowed to leave their places (Klas and Contorno, 2020).

2. Literature review

2.1. Advantages of user-generated content

The use of social media platforms such as Twitter, Reddit, or Facebook to analyze people's behavior is gaining traction, especially on psychological grounds. In this sense, De Choudhury, Counts, and Horvitz (2013) in the pioneering publication `Social Media as a Measurement Tool of Depression in Populations´ propose the use of social media for analyzing depression. The benefits of using social media posts over traditional questionaries for depressed individuals are broad.

[1] According to World Population Review: https://worldpopulationreview.com/states
[2] For more information visit: https://www-nytimes-com.eur.idm.oclc.org/interactive/2020/us/coronavirus-stay-at-home-order.html?searchResultPosition=1 and https://www-nytimes-com.eur.idm.oclc.org/interactive/2020/us/states-reopen-map-coronavirus.html

Social media platforms allow live monitoring of people's feelings, behaviors, and thoughts while identifying and following users across time (Leis et al., 2019; Monzani et al., 2021). They also avoid incurring in very expensive survey costs (Wang, et al., 2022). Tyler McCormick from the Center for Statistics and the Social Sciences of Washington University states that "…*Twitter is the largest observational study of human behavior we've ever known…*" (Frizell, 2014).

Moreover, the use of mobile devices enables accessing social media platforms at every moment and place (Leis et al., 2019). User-generated content also provides large amounts of up-to-date information about people's activity and language usage in a natural environment, which is not biased by an experiment setting or lack of memory of the responder (De Choudhury, Counts, and Horvitz, 2013; Zhou; Hamad; Shuiqiao, Shoaib, Guandong, and Fang, 2021).

2.2. Use of social media to monitor depression

A common feature found in papers that measure depression levels using social media is to build and train classification models (logistic regression, Naïve Bayes, decision trees) that can predict the probability of suffering from a mental disorder based on the way people write. Some of those papers do so while analyzing post features using Bag-of-Words (which counts the number of times certain words appear in a block of text) or the widely known LIWC psycholinguistic dictionary (De Choudhury, Counts, and Horvitz, 2013; De Choudhury, Gamon, Counts, and Horvitz, 2013; Coppersmith, Dredze, and Harman, 2014; Nadeem, Horn, Coppersmith, and Sen, 2016).

In the beginning, pioneering papers in the field use both psychological questionaries that determine whether a person has depression -or any other mental disease- to classify the post on social media of the same users (when they consent to do so). Therefore, researchers were using mainly two sources of information: medical questionaries and social media posts. Shortly after, a newer generation of papers emerged using only social media posts to analyze people´s mental health. They mainly classify depressed individuals by making use of self-reported statements (Coppersmith, Dredze, and Harman, 2014; Safa, Bayat, and Moghtader, 2022)

Among the pioneering papers, De Choudhury, Counts, and Horvitz (2013) were able to develop a classification model that predicts depression with 73% accuracy. The authors use a crowdsourcing

methodology to classify individuals in their sample. This methodology consists of using the mentioned questionaries and asking people if they consent revealing their Twitter username to analyze their post. They made an index that considers different dimensions of their tweets (language, emotion, style, and user engagement) and build each of them using -among others- the post text analyzed using LIWC, number of followers, likes, and time of the post. They also test the index using geolocated tweets in 50 states of the United States and the 20 unhappiest cities of the country.[3] In that sense, it is worth mentioning that both California and Florida obtained the same level of depression based on their model. In the same year (2013) De Choudhury, Gamon, Counts, and Horvitz, using the same database[4], found that depress users manifest a decrease in social media usage, an increase in negative emotional posts, talk more amount religion and medical concerns while related to a smaller number of users.

Coppersmith, Dredze, and Harman (2014) build on top of the previously mentioned research findings widening the scope of mental illness to consider 441 depressed users, 244 users with post-traumatic stress disorder (PTSD), 394 individuals with bipolar disorder, and 389 with seasonal affective disorder (SAD). These authors formed the database by considering self-reported mental illness diagnosed by users who post 'I was diagnosed with X', being X any of the mentioned mental disorders. From this corpus of users, tweets made between 2008 and 2013 were extracted and compared their way of writing with a randomly selected control group of 5728 users. Coppersmith et al., (2014) also conduct LIWC analysis using the categories: swear, posemo (positive emotions), negemo (negative emotions), anx (anxiety), and some pronoun expressions. The authors performed a correlation analysis founding that swears, anger, and negemo were highly correlated and triggered by the same words. They show significant differences in ways of writing among the treatment categories and the control group based on the selected LIWC dimensions.

Safa, Bayat, and Moghtader (2022) are also part of the second group of papers that only use user-generated content to predict depressive symptoms. These authors also analyzed LIWC positive and negative emotions as well as the usage of first personal pronouns, which are known to be more used by depressed individuals (Safa, Bayat, and Moghtader, 2022; Coppersmith, Dredze, and

[3] Based on article: http://images.businessweek.com/ss/09/02/0226_miserable_cities/index.htm
[4] Crowdsourcing to identify 171 Twitter users who had been diagnosed with depression and collect one year of tweets from them before the diagnosis.

Harman, 2014). In this sense, Leis et al. (2019) also proved that depressed people used significantly more first-person singular pronouns when comparing tweets. However, Tackman, et al., (2019) proved that the use of self-referential language should be seen as an indicator of negative emotions or general distress rather than depression.

Many papers also use other information social media provides. For instance, the time in which people post and the level of engagement they manifest with others. Some authors study the time of the post, considering a night-window from 9 PM and 5:59 AM (De Choudhury, Counts, and Horvitz, 2013; De Choudhury, Gamon, Counts, and Horvitz, 2013) or between midnight and 4 AM (Coppersmith, Dredze, and Harman, 2014) in which depress users proved to be more active. Time-analysis are relevant because in 80% of depressed individuals, symptoms tend to worsen during the night (Lustberg and Reynolds, 2000). The level of engagement of users with others can also be measured. In this case by considering the number of posts made, the proportion of replies to other's posts, the number of mentions (@USERNAME), the size of their social network considering the number of unique mentions, (De Choudhury, Counts, and Horvitz, 2013; Coppersmith, Dredze, and Harman, 2014; Zhang, Lyu, Liu, Zhang, Yu, Luo, 2021).

2.3. LIWC for analyzing blocks of text

Until now, many papers' approaches had been discussed using dictionary-based techniques to analyze blocks of text. In this section, this kind of technique is going to be discussed more in depth as it is going to be the one used in this study.

Sentiment analysis is an emerging trend used to understand people's sentiments in multiple situations in their everyday life (Varghese Babu and Kanaga, 2021). It consists of the classification of a block of text as either positive, neutral, or negative. The two most common sentiment analysis tasks are subjectivity and polarity detection (Thelwall, Buckley, and Paltoglou, 2012). The former predicts whether a given text is subjective or not getting values in the range [0, +1], being 0 very objective and +1 very subjective. And the latter whether a text is positive or negative, gets values in the range [-1, +1], being -1 extremely negative, +1 extremely positive, and 0 neutral. Both variables can be obtained from blocks of text using Python's library TextBlob (Loria, 2020).

There are two big approaches for analyzing the strength of sentiments: lexical algorithms and machine-learning algorithms. According to Thelwall, Buckley, and Paltoglou (2012), the first one involves assigning a sentiment orientation to a set of terms. Therefore, based on the occurrences of the defining words, the algorithm predicts the sentiment of the whole block of text. This approach includes the use of a psychological dictionary, -such as *the Linguistic Inquiry and Word Count (LIWC)*-, the General Inquirer Lexicon, or the SentiWorldNet- that will the help of a lexical algorithm, can be used to predict the sentiment of a text-based upon the occurrences of certain words.

On the other hand, the second approach involves the usage of machine-learning algorithms after defining a method for feature extraction on text. This method can be just words, stemmed words (words roots), or part-of-speech tagged words. Those selected features are going to be used to train the classification modes, which can be categorized into traditional machine learning-based approaches and deep learning-based approaches. Machine learning-based methods include Support Vector Machine (SVM), Naïve Bayes (NB), Maximum Entropy (ME), Decision tree learning, and Random Forests. Those methods are further categorized into supervised and unsupervised learning methods (Yadav and Vishwakarma, 2020).

The usage of dictionary-based approaches dates back to the 20th century when a group of psychologists identified clusters of words in specific dictionaries that shed light on people's desires, achievements, and affiliations (Boyd and Pennebaker, 2016). Those types of methods give insights into individual behavior and thoughts by analyzing the usage of their words.

Pennebaker, Boyd, Jordan, and Blackburn (2015) mentioned that the LIWC 2015 Dictionary is equipped with more than 6400 words and can generate approximately 90 variables for each block of text the program analyzed. Those variables are classified into different categories:

- Summary language variables (including the variables: analytical thinking, emotional tone, and authenticity). [5]
- Descriptor categories (words per sentence, proportion of the analyzed words that are captured by the dictionary).

[5] These variables were introduced in the 2015 edition of the software (Pennebaker, Boyd, Jordan, and Blackburn, 2015).

- Linguistic dimensions (percentage of articles, pronouns, negations).
- Psychological process (affections, cognition, health, family).
- Personal concerns (words related to work, religion, death, leisure activities).
- Informal language (netspeak, swear words).
- Punctuation categories (commas).

Every time a word is matched with a dictionary category, the corresponding variable will be incremented. The authors give the example of the word *cried*, which is linked to five LIWC categories: sadness, negative emotion, overall affect, verbs, and past focus.

2.4. Application of sentiment analysis after the outbreak of COVID-19

Many papers had recently emerged to shed light on the pandemic consequences on mental health around the world. Zhang, Lyu, Liu, Zhang, Yu, and Luo (2021) claimed to have constructed the biggest depression-related database in English identifying 2575 depressed Twitter users in The United States (based on their biography information and post where they manifest being diagnosed with depression). The authors consider the same number of control base individuals, and their past tweets during three months starting on April 18th, 2020. They trained a model to identify depressed and non-depressed posts based on some LIWC features (tone, clout, analytical thinking, anger, and anxiety, among others). They used their model to construct depression scores for the states of California, Florida, and New York and compared their behavior with the national level. They spot that Florida had a lower depression score -compared to the other states and the national level- both before and during the COVID-19 outbreak. Additionally, both California and Florida paid relatively more attention to the government policy regarding the pandemic (also compared to New York and the national level).

In the same line but using a different approach, Wang, Fan, Palacios, Chai, Guetta-Jeanrenaud, Obradovich, Zhou, and Zheng (2022) found that the pandemic generated a sharp decline in sentiments scores around the globe that were followed by slower and asymmetric recoveries when considering data from 1st of January to 31st of May of 2020. These authors retrieved more than 650 million geotagged posts around the world from Twitter and Weibo (the Chinese counterpart). Unlike other authors, in this case, posts directly associated with the pandemic or the COVID-19

virus were excluded. The authors justify this choice by saying that pandemic-related posts might not be a good sample representation of the emotional state of the general population and can be contaminated with political discussions or campaigns.

In Italy, Monzani, Vergani, Francesca, Pizzoli, Marton, and Pravettoni (2021) run a descriptive study. They used Italian COVID-19-related tweets to evaluate the change of three variables they construct using LIWC[6] (emotional tone, analytical thinking, and somatosensory processes). The first quarantine phase in Italy (24/2/2020 – 14/6/2020) was considered. The authors found that at the beginning of the quarantine period, there were lower records of emotional tone and analytical thinking. However, when daily cases and death raised, the use of negative emotions and somatosensory words also increase.

In Australia, Zhou, Zogan, Yang, Jameel, Xu, and Chen (2021) retrieve tweets in the New South Wales (NSW) state during the period January 1st, 2020 – May 22, 2020, with the help of Tweepy Python library and Twitter API.[7] The authors divide their sample into two. One part for depressed tweets and, another one for non-depressed (control) tweets. Depressed tweets were defined as those published by persons who identify themselves as being diagnosed with depression. From each tweet, they extract three features: emotions (based on emojis and slang usage), topic-level, and domain-specific features. Using the following classification methods: logistic regression (LR), linear discriminant analysis (LDA), and Gaussian Naïve Bayes (GNB) they found that people were more depressed after the government implementation of the quarantine restrictions in the state.

Also in Australia, Wang, Huang, Hu, Zhang, Li, Ning, Corcoran, Khan, Liu, Zhang, and Li (2022) retrieved 244.406 tweets that contains searching terms related to the pandemic such as *pandemic*, *COVID-19**, *coronavirus*, *vaccin** to track the changes in mental health across eights Australian capital cities during the period: 1st January 2020 – 31st May 2021. The period was chosen to capture the evolution of the polarity of the tweets and eight selected emotions during three phases of the pandemic in the country. They found that people move from being pessimistic and having negative emotions in the first phase (1/1/2020 – 10/3/2020), to a more optimistic outlook of the pandemic

[6] These authors use the Italian LIWC2015 version of the software.
[7] API stand for Application Programming Interface.

in the second phase (11/3/2020 – 25/3/2021). And return to a more negative and pessimistic final phase (26/3/2021 – 31/5/2021).

In Spanish-speaking countries, Leis, Ronzano, Mayer, Furlong, and Sanz (2019) also use Twitter API to identify depression patterns in the Spanish language. The authors create three databases: one for depressive users (those who mention in their profile suffering from depression), another for depressive tweets (manual selection of tweets from the depressive users), and a control database. They mention that depressive users are less active on Twitter, but they use it more during the night, being this a symbol of insomnia. They also found a higher frequency of tweets related to sadness in the depressive tweets database, and a predominance of anger in both depressive users and tweets samples.

3. Data collection and methodology

In this section the way data has been collected and cleaned is presented, as well as the methodology used to answer the research questions. The first subsection introduces the sample selection and data cleaning criteria. The second one consists of the methodology used: MANOVA test and consequent post-hoc tests. It also presents assumptions of the test, and advantages of running it.

3.1. Data Collection and Preparation

For this research, Twitter data was collected using Twitter API with a research account that allows gathering Twitter information without any time constraints and with a maximum of ten million tweets per month. This includes the contents of the tweets, date and time of each of them, username, location, number of followers, and retweets of the users. The connection to the API and the preparation of the data was done using Python and R software.

The data extraction was structured as follows: firstly, individuals who declared to be diagnosed with depression or PTSD were identified. This also encompassed those who communicated to had started taking antidepressants. These statements were filtered considering those done in the states of Florida and California after the outbreak of COVID-19 (during the period April 2020 – July 2020). Secondly, previous tweets of the selected individuals were extracted. Other variables such as date, time, and location of the tweet were also obtained.

In the US, antidepressants must be prescribed by a physician, psychiatrist, or nurse practitioner.[8] Therefore, we are assuming that those who declared to start taking antidepressants had been diagnosed with depression. The inclusion of PTSD people is based on papers that deal with depressed and PTSD individuals together, especially as many of the symptoms are shared among the diseases (Coppersmith, Dredze, and Harman, 2014; Nadeem, Horn, Coppersmith, and Sen, 2016). Given that the lockdown has been roughly three times longer in California than in Florida, people from those states are going to be considered as treatment and control groups respectively.

For the identification of users Twitter API and Python libraries 'Tweepy' and 'Pandas' were used. Once the users were identified, the platform 'Export Comments'[9] was used to extract the tweets of the first half of 2020 (from the 1st of January until the 30th of June of 2020). In this case, retweets, replies, and posts that only included links were omitted since insights can only be extracted when the sender of the message generates the information themselves from the use of their own language (Van Der Zee, Poppe, Havrileck, and Baillon, 2021).

Following other researchers, for this analysis individuals whose tweets contain the words: "*I * diagnosed * with depression*", "*I * diagnosed * with PTSD*" or "*I * start * antidepressants*" are going to be considered (Ettman, et al., 2021; Chandra Guntuku, et al., 2017). Note that the aesthetic symbol implies that the order of words can be modified, other words can be added in the middle, and that the words can be alternated. For instance, the word 'started' is accepted as it contains 'start' in it. To clarify, some fragments of the obtained tweets are presented with the search terms marked in bold: "*Today, my therapist* ***diagnosed*** *me* ***with*** *Clinical* ***Depression***. ***I*** *tried to tell my primary care doctor TWICE that* ***I*** *was depressed and both times* ***I*** *was ignored...*"; *"...**I** was recently **diagnosed with PTSD**...*"; "***I*** *was* ***diagnosed with*** *anxiety/****depression*** *because the last two weeks* ***I****'ve been having attacks where my body shakes and...*", "***Start****ed* ***antidepressants*** *today.* ***Start*** *of actually getting mental health care that* ***I****'ve been needing for so long...*"

Twenty-seven tweets were identified using the mentioned filtering expressions in users with locations either in Florida or California in the time frame April-July 2020. One tweet was removed

[8] https://khealth.com/learn/antidepressants/how-to-get-antidepressants/#who-can-prescribe-antidepressants
[9] https://exportcomments.com

from the original list as it was from the same person. Of those distinct twenty-six individuals, eight were omitted (six from California and two from Florida) because of not reaching a certain threshold of a minimum number of Tweets for the study (which I set on a minimum of 50 days with a post in Twitter for individuals in the state of California and 20 days for those from Florida). The threshold was arbitrarily set and differs per state as California SAH order was longer and, because from the original twenty-six individuals, a disproportionated majority come from this state. Hence, this reduction of individuals aimed to balance the database. Additionally, another user has been removed from the database after analyzing the Tweets content as the person manifest moving to another state (Atlanta) in the considered period. Therefore, the final database has **seventeen individuals**, nine from California and eight from Florida.

The primary database consists of 5.101 tweets made by the selected users in the first six months of 2020. However, this database has been shortened to **4.131 tweets**, as the purpose of this study is to compare those tweets in the two selected states before and during the Stay-at-Home order (SAH). This number of posts ensures that in each state there is at least one tweet (observation) per day. Table Nº1 summarized the period in which tweets of the selected users were collected. The mentioned 4.131 tweets are obtained considering 165 days of tweets in California, and 129 days of tweets in Florida, in both cases starting on the 1st of January 2020.

Table Nº 1: Time frame considered in the analysis

	Before SAH Period	During SAH Period
California	01/01/2020 - 18/03/2020 **(77 days)**	19/03/2020 - 15/06/2020 **(88 days)**
Florida	01/01/2020 - 12/04/2020 **(102 days)**	13/04/2020 - 30/04/2020 **(27 days)**

Given the current limitations in the number of characters in a tweet (280 characters), the use of abbreviations or short text forms is widely used in posts (Chua, Storey, Li, and Kaul, 2019). Those abbreviations together with spelling mistakes can jeopardize the estimation of the variables done in this study from the block of texts (Newman et al., 2003; Van Der Zee et al., 2021). Therefore, each tweet was manually corrected to ensure better estimations of the mentioned variables. Consult Appendix A for an extensive list of the short forms that have been manually corrected in every tweet.

3.2. Methodology

This study aims to dig into the differences in the way people express themselves on Twitter in two different locations (California and Florida) and during two different moments in time (before and during the SAH order). Therefore, a Two-Way Multivariate Analysis of Variance (MANOVA) is going to be conducted by taking as factors (independent or group variables) both the state and the phase of each tweet (before or during the SAH order). The explanatory variables are formed by relevant variables created with both LIWC and Python software. This section will introduce the MANOVA test, its advantages, and assumptions. After that, two post-hoc tests are going to be presented, the traditional Two-Way Analysis of Variance (ANOVA) conducted per each response variable and the Discriminant Descriptive Analysis (DDA). The results of the presented test are going to be explained in the following section.

3.2.1. Two-Way MANOVA Test: in theory

The Multivariate Analysis of Variance (MANOVA), as the Analysis of Variance (ANOVA), is part of the family of General Linear Models (GLM) that assess correlations and effect-sizes equivalent to the coefficient of determination R^2, which states the participation of the variation of the independent variable in the dependent one (Reichwein Zientek and Thompson, 2009). Following Loewen and Plonsky (2016) MANOVA can be defined as a kind of analysis of variance that contains more than one dependent or response variable together with one or more independent or factor variables.[10] The purpose of the MANOVA is to assess whether multiple levels of the group variables -on their own or combined- have an impact on the dependent variables (Bray and Maxwell, 1985).

According to Stevens (2009), $k - group$ MANOVA compares each group simultaneously on p dependent variables. Bray and Maxwell, (1985) present the MANOVA as a test that similarly to the ANOVA, analyses the amount of variation within every independent variable and states whether the variation within those variables is smaller than the one calculated between them. Therefore, a larger between-subjects variance, than the within-subject variance, in the independent

[10] Note that there is no consensus in the literature on how to address these variables, in the reviewed literature they are mentioned as: factor, group, categorical or independent variables indistinguishable.

variables implies that this variable has a significant effect on the dependent variable. That means that the independent variables do affect the dependent ones.

As in the univariate case of ANOVA (that considers only one dependent variable) the null hypothesis is presented as follows:

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \cdots = \mu_k$$

In this case, the null hypothesis represents the same population means (μ) of all dependent variables for every level of the independent or group variable. Note that in the multivariate case, the last equation represents population means vectors (instead of just population means) that are equal for all dependent variables on all the categories of the independent variable (Stevens, 2009). At this point, it is worth anticipating that as only one mean being different leads to the rejection of the null hypothesis, further analysis (post-hoc tests) is needed to establish which independent variables differ.

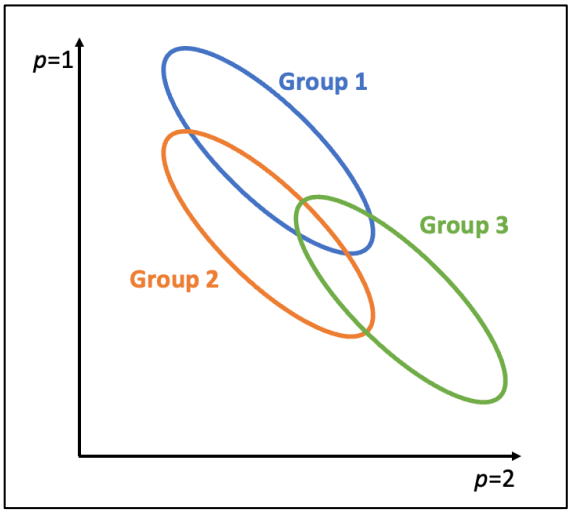
3.2.1.1. Intuition behind the MANOVA test

The idea behind the MANOVA is to check if different groups (usually one or more treatment groups versus one control group) show a difference in the behavior in different responses or dependent variables. The null hypothesis assumes that all groups are equal in all the selected dimensions, or in other words, that there is no treatment effect (Stahle and Wold, 1990). Following these authors, the intuition of the test graphically is presented in Figure Nº1. This graph considers three groups (J=3) and two dependent variables (p=2). Note that considering only two dependent variables is useful for the graphical representation as they determined the number of axes. Warne (2014) mentioned that in case of perfect correlation between two dependent variables and considering three independent variables (p=3), a three-dimension graph will collapse into a two-dimensional one.

Stahle and Wold (1990) mentioned that every observation represents a dot in the plot that is grouped to form one ellipse per group level. Hence, under the null hypothesis, a perfect overlap of the three ellipses is expected. The authors specify that the null hypothesis is rejected when the

distances *between* the centroids of the ellipses (mean points) are larger compared to the *within* variation of the groups.

Figure Nº 1: Representation of three bivariate ellipses (scatters)



Source: made by the author based on Stahle and Wold (1990)

3.2.1.2. MANOVA advantages

Given that researchers should either conduct a MANOVA or the univariate equivalent ANOVA for every dependent variable (Warne, 2014) it is worth mentioning the two main benefits of conducting this multivariate test over the ANOVA variant:

1. As MANOVA does not run a test for each dependent variable (while this is the case of the ANOVA), it decreases the experiment-wise Type I error. This is because there is less chance of committing a Type I error when a smaller number of statistical tests are needed. Being the Type I error rejecting the null hypothesis while it is true, or in other words: wrongly assuming that the groups differ (Loewen et al., 2016).
2. MANOVA is set to be more realistic as only this type of analysis (multivariate) considers simultaneously all the possible interactions among dependent variables (Reichwein Zientek and Thompson, 2009). This is particularly important in social disciplines, that are also interested in the combined effects of dependent variables on independent ones (Warne, 2014).

3.2.1.3. MANOVA assumptions: in theory

The MANOVA test, like any other mathematical model, tries to approximate reality. Therefore, violations of model assumptions are innate to the usage of the model. However, what matters is the extent to which assumptions are violated. In this respect, Stevens (2009) explains in detail that some violations of assumption are more serious than others. Following this author, the three assumptions that dependent variables on the MANOVA test must meet are going to be presented together with their impact on the Type I error and power of the test.[11]

1. Independence assumption observations should be independent, being the violation of this assumption a very serious problem (Stevens, 2009).

2. Multivariate normality assumption in each group, observations of the dependent variables should have a multivariate normal distribution. However, due to the complexity of testing this assumption, the author recommends testing normality for each dependent variable in each group using the Shapiro-Wilk test, in other words, checking marginal normality for each variable would be sufficient (Stevens, 2009).

   The Shapiro-Wilk test is a *goodness-of-fit* check that assesses to what extent the sample resembles a normally distributed data set. The test constructs a statistic (W) that takes values between 0 and 1, representing 1 a perfect match. For the construction of the statistic, the sample is ordered and standardized.[12] If the sample data follows a normal distribution the quantile values of the observations would be equally spaced. The intuition of this test suggests that values of the statistic W close to 1 represent a sample data that fits a normal distribution (King and Eckersley, 2019). As in this case, the null hypothesis states that the data is normally distributed, with a α level of 0.05, the $p - value$ associated with this statistic needs to be greater than 0.05 to not reject the null hypothesis (Stevens, 2009).

[11] Type I error is the probability of rejecting a null hypothesis when it is actually true in the population, in this sense 'the probability of committing a Type I error', 'α' or 'level of statistical significance of the test' are all used as synonyms (Banerjee, Chitnis, Jadhav, Bhawalkar, & Chaudhury, 2009). Sherry (2006) defined the Type I error as the chance of spotting significant results when they shouldn't exist. On the other hand, the probability of failing to reject the null hypothesis when it is false (committing a type II error) is known as β and the value 1 - β is defined as the power of the test (Banerjee et. al., 2009). The power of the test is the probability of rejecting the null hypothesis while it is false. Intuitive we can easily elaborate on the relation between a and the power of the test as follows: as a increases, so does the power of the test, *ceteris paribus* because a larger significance level (α) implies a larger area of rejection for the test and therefore a greater chance of rejecting the null hypothesis (and a more powerful test), in other words, the price you pay for increasing the power of the test is working with a greater a (AP Central, 2022).

[12] Standardizing a sample refers to converting it into a sample with a distribution that has a mean equal to zero and a standard deviation equal to one (μ=0 and σ=1).

Considering that the $p - value$ shows the smallest significance level at which H_0 can be rejected (Wooldridge, 2020). However, it is worth mentioning that Stevens (2009) also states that not fulfilling this assumption has a small effect on Type I error and power, unlike the previous one.[13] Another way to test for normality, widely used by the literature, is by the performance of quantile-quantile plots (Q-Q plots). This methodology is explained in Appendix C.

3. Homogeneity of variance assumption indicates having the same population covariance matrix on each dependent variable (homogeneity of covariance matrices). The covariance matrix is formed by the covariance of each variable, which is a measure of the relation between the variation of two variables. In this sense, a positive covariance means that the variables tend to move together (increase or decrease). A negative covariance shows an inverse relationship and a covariance equal to zero shows that the variables are not related at all. However, to measure the strength of the (linear) relation, the correlation needs to be used as it shows both if the variables move together -direction- and the strength of the relation -how close to one in absolute value the coefficient is- (King and Eckersley, 2019). According to Stevens (2009), this assumption is very restrictive as for two matrices to be equal, all the corresponding elements of the matrix need to be the same.

   This assumption is usually tested using Box's M test (Stevens, 2009; Bray and Maxwell, 1985; Sarma and Vishnu Vardhan, 2018). In this test, the null hypothesis states that each group has an equal covariance matrix, and it is rejected when the $p - value$ of the test is smaller than the significance level α. However, two remarks should be mentioned when the null hypothesis is rejected.

   First, the homogeneity of variance condition is robust if the groups have a similar number of observations. Or in other words, if the ratio between the observations of the largest and the smallest group is less than 1,5 (Stevens, 2009). Secondly, the Box's M test is very sensitive to non-normality. Stevens (2009) illustrates this by mentioning that the null

[13] Stevens (2009) also provides a selection of studies that proves small effect on the Type I error and power of the test by deviations from multivariate normality.

hypothesis might be rejected not because of unequal covariance matrices, but because of a lack of nonnormality.

The three presented assumptions can be easily reduced to two. Lars (1990) mentioned that MANOVA rests mainly on two assumptions: the independence of the observations (first assumption) and the equal covariance matrix for the residuals of all groups or independent variables (third assumption). This is mainly because the second assumption of multivariate normality in the dependent variables acts as a precondition to meet the third assumption of homogeneity in the covariance matrices.

3.2.1.4. MANOVA statistics

The most common multivariate measures used to calculate MANOVA are Wilk's lambda, Pillai's trace, Hotelling-Lawley trace, and Roy's largest root. These tests differ in the way they combine the dependent variables to assess the amount of variation in the data (Bray and Maxwell, 1985).

Wilk's lambda (L) shows the amount of variance of the dependent variable that is not explained by the (different levels of) the factor or independent variable (Bray and Maxwell, 1985). Therefore, the smaller the value of the statistic, the larger the difference between the analyzed groups. As Wilk's lambda can only get values between zero and one, the result of subtracting the statistic from one ($1 - L = \eta_p^2$) shows the amount of variance of the dependent variable that is explained by the independent variable, which is also known as effect size or generalize eta-square (Patel and Bhavsar, 2013; Steyn Jr and Ellis, 2009). In the case of a statistic equal to zero (ideal scenario), there is no variance not being explained by the independent variable leading to an effect size equal to 1 (Bray and Maxwell, 1985).

Pillai's trace measures the amount of variance of the dependent variable that is explained by a larger separation of the factor variables (Bray and Maxwell, 1985). It can also get values from 0 to 1. Contrary to Wilk's lambda, this is a positive value statistic, and therefore higher values indicate that the effects contribute more to the model, in other words, with values close to one, the null hypothesis should be rejected (IBM, 2022). This statistic is considered to be the most reliable as it accounts for Type I errors when the sample size is small (Bray and Maxwell, 1985).

The Hotelling-Lawley trace calculates the most significant linear combination of dependent variables (Bray and Maxwell, 1985). This statistic is larger than Pillai's trace, however, in presence of small eigenvalues of the test matrix,[14] both statistics are similar (IBM, 2022). Finally, Roy's largest root is obtained in a similar way as Pillai's trace with the exception that in this case, only the largest eigenvalues are considered (Bray and Maxwell, 1985). This statistic is smaller or equal to Hotelling's trace. When the two statistics have the same value, the effect is mainly associated with only one dependent variable (IBM, 2022). With larger sample sizes, the level of significance of the four presented tests tends to converge (Bray and Maxwell, 1985). However, Wilk's lambda remains the most widespread test used, due to its simplicity (Bray and Maxwell, 1985).

3.2.2. Post-hoc test

After performing a MANOVA test and finding a statistically significant result among the defined groups, it is worth understanding from which variables the statistical differences come from. As MANOVA only tests the hypothesis that at least one mean is significantly different among the groups a *post-hoc*[15] test should be used (Foster, et al., 2018). In this section, the commonly used ANOVA test is going to be presented as a first alternative. Secondly, the Discriminant Descriptive Analysis (DDA) is introduced as a more refined alternative.

3.2.2.1. ANOVA as a post-hoc test: in theory

The Analysis of Variance (ANOVA), as was explained before, aims to detect if all the levels of the factor variables (in this case state and phase) have the same mean regarding one specific variable or if at least one differs significantly from the rest (King and Eckersley, 2019). This test is the most used post-hoc test after performing a MANOVA, as it can be run per response variable. Therefore, it allows disentangle those response variables that (individually) show differences among the group variables and those that (also individually) show no differences among groups.

In this study, nine Two-Way ANOVA tests are going to be run (one per response variable). Two-way because, as in the case of MANOVA, both the state and the phase variables, together with the interaction effect among them, are going to be the independent or group variables.

[14] This matrix is obtained by calculating the error-term matrix which is obtained by taking the inverse of the within-groups sum of squares and cross-products matrix, and then multiplying this matrix by the between-groups sum of squares and cross-products matrix.
[15] Latin expression that means 'after the event'.

Even though the test aims to analyze differences in means, it is called Analysis of Variance, as it calculates two different estimations of the population variance (σ^2), one that is sensitive to differences in the mean between groups and another one that it is not (King and Eckersley, 2019). In this way, in the case of all the group means being the same, the two estimations made by the test will also match. Therefore, the test will not reject the null hypothesis of all means being the same. However, when the group means differ, the estimations will differ. Consequently, the null hypothesis is going to be rejected, being unlikely that the differences in means between the groups occur by chance (King and Eckersley, 2019).

Following Tabachnick and Fidell (2007), the simplest form of ANOVA is going to be explained. This model has an independent variable with two group levels (or factor variable) and a dependent variable, which is expressed on a continuous scale. Each level of the independent variable is formed by many observations per level. Putting these variables into an equation and considering that each observation has a score that is represented by Y_{ij} where observations are denoted by $i = 1, 2, \dots, n$ and every group level is denoted by j. The grand mean is denoted as GM, which is calculated for all the observations in all the group levels. Therefore, the difference between every score and the grand mean ($Y_{ij} - GM$) can be divided into two. One part represents the difference between each score and its group level mean ($Y_{ij} - \bar{Y}_j$) that also represents the error term.[16] Another part shows the difference between the group means and the grand mean ($\bar{Y}_j - GM$). This second part represents the effects of the independent variable and an error term (Tabachnick and Fidell, 2007).

$$(1) \quad \left(Y_{ij} - GM\right) = \left(Y_{ij} - \bar{Y}_j\right) + \left(\bar{Y}_j - GM\right)$$

Every term in equation (1) is then summed and squared. By doing so it is assured that negative and positive terms do not cancel among each other (Tabachnick and Fidell, 2007). The first term of the summed and squared equation on the right of the equal sign shows the total sum of the squares (SS) within the group (that represents the error). The last term shows the sum of the squares between groups (treatments), equation (1) can be rewritten as follows:

[16] Considering that all the observations received the same treatment (are part of the same group) differences between each score and the group mean are unexplained.

$$(2)\ SS_{total} = SS_{wg\ or\ error} + SS_{bg}$$

At this point, it is clear why the test is called Analysis of Variance, as the variance should be interpreted as a deviation or difference between two elements. This deviation is the one present in the three terms of equation (1). However, for the sum of squares (SS) to become variance, they should be averaged (Tabachnick and Fidell, 2007). The three terms of equation (2) are averaged by the corresponding degrees of freedom (df).[17] When dividing each term in equation (2) by the corresponding df term, the three variances of the test calculated as the *averaged* sum of squared are obtained. This is shown in equation (3), which can be rewritten as equation (4), given that each term is called mean squared (MS) (Tabachnick and Fidell, 2007).

$$(3)\ \frac{SS_{total}}{df_{total}} = \frac{SS_{wg\ or\ error}}{df_{wg\ or\ error}} + \frac{SS_{bg}}{df_{bw}}$$

$$(4)\ MS_{total} = MS_{wg\ or\ error} + MS_{bg}$$

To test the null hypothesis that the population means of every group level are the same ($\mu_1 = \mu_2 = \cdots = \mu_a$), the F statistic needs to be calculated. The F is a ratio of two variances, in this case, the mean square between groups and the mean square of the error term as it is shown in the following equation (Tabachnick and Fidell, 2007).

$$(5)\ F = \frac{MS_{bg}}{MS_{wg\ or\ error}}$$

If the null hypothesis is true and there are no main differences between the group levels that make the numerator bigger, the F ratio reduces to be a ratio of two estimates of the same error. Therefore, F ratio will be close to 1 (Tabachnick and Fidell, 2007).

Note that the assumptions needed for performing ANOVA are omitted as they are the same ones already presented in the MANOVA case.

[17] Where the df_{total} (total degrees of freedom) are calculated as the number of scores or observations N minus 1. The $df_{wg\ or\ error}$ (within-group degrees of freedom) are N minus the number of group levels a. And the last term, df_{bw} (between-groups degrees of freedom) are represented as the a group means minus 1 (Tabachnick & Fidell, 2007).

3.2.2.2. Discriminant Descriptive Analysis (DDA) as a post-hoc test: in theory

Among the many different *post-hoc* test that exist Warne (2014) and, Huberty and Olejnik (2006) provide reasons to use a Discriminant Descriptive Analysis (DDA) instead of the more widespread ANOVA method. Reichwein Zientek and Thomson (2009) mentioned that ANOVA and MANOVA are answering different research questions, as the former address questions related to observed variables while the latter, those related to unobserved latent variables made from observed variables. Those latent constructions, such as attitudes or beliefs are in the scope of social sciences and therefore, worth being analyzed (Warne, 2014). Furthermore, as mentioned before, performing one ANOVA per every dependent variable will increase the probability of having a Type I error (Sherry, 2006).

As MANOVA and ANOVA, DDA is also part of the GLM, and its conceptually and mathematically equivalent to a multiple regression. The coefficients obtained in DDA functions are equivalent to the betas calculated in regressions. The main difference is that DDA linearly combines dependent variables creating synthetic dependent variables that maximized the differences among groups (Sherry, 2006). The main aim of this method is to point out the dependent variables that are linked to group differences (Sherry, 2006).

According to Sherry (2006) before analyzing each DDA function (done per group variable) it is important to first compute the (Canonical) Discriminant Functions. These functions are calculated per each group variable as they represent a synthetic indicator of the degree of group separation given the chosen response variables (polarity, subjectivity, etc.). This author also mentioned that in presence of non-significant Discriminant Functions, the DDA analysis should be stopped.

From the outputs of the discriminant functions, three variables are pivotal: *squared canonical correlation (R^2), eigenvalues, and $p-values$*. The first one accounts for the proportion of the variance that is explained by the correlation between the grouping variable and the dependent variables of the model. Eigenvalues are the ratios of between-groups to within-groups sum of squares (Sherry, 2006). Therefore, large values of the R^2 and eigenvalues indicate that the function succeeds in separating groups. Finally, $p-value$ is related to the null hypothesis of the function where all canonical correlations in the model are all equal to zero. $P-values$ smaller than the level α are needed and desirable.

The number of DDA functions generated in the model can vary but the minimum corresponds to one less than the number of groups or response variables, whichever is the smallest (Enders, 2003). To spot the differences among groups some coefficients should be considered: *standardized coefficients, structure coefficients* (r_s), and *parallel discriminant ratio coefficients* of the discriminant functions and the group centroids (Sherry, 2006; Warne, 2014).

According to Sherry (2006), *the standardized coefficients* describe the relative importance of the variables in the function as they help to build the synthetic or discriminant variable (also known as the DDA score). However, they cannot rank response variables in terms of importance as those coefficients are made to simultaneously consider the contributions of the other variables. Therefore, if two or more variables are correlated, they share their contribution to the discriminant score and the individual contribution of each of the variables is not possible to be disentangled (Sherry, 2006).

The *structure coefficients* (r_s) are simply the Person's correlation coefficient that ranges from +1 to -1 between the observed variable and the synthetic one, created from all the predictor variables in the equation (Sherry, 2006). Considering the *squared of these values,* (r_s^2) show the variance participation in the synthetic or composite score of the function (Sherry, 2006).

Standardized coefficients cannot be used to elaborate a ranking of variables in the equation. However, following Thomas and Zumbo (1996) and Warne (2014), *the parallel discriminant ratio coefficients* are introduced as a measure of the relative importance of response variables in the discriminant function. Nevertheless, parallel discriminant ratio coefficients cannot be considered as an indicator of variable importance in presence of suppressor variables (Thomas and Zumbo, 1996). These last variables influence the synthetic variable through relations with other response variables (Sherry, 2006). These coefficients are calculated as the multiplication of the standardized coefficients and the structure coefficient of each response variable.

4. Results

In this section the results of the MANOVA, ANOVA, and DDA test are going to be presented. However, first other descriptive information will be shown to contextualize the insights provided by the run tests.

4.1. Preliminary descriptive information on the collected data

Daily number of infections can have an impact on population's feelings. This section will present the evolution of both infections and deaths per state in the analyzed period distinguishing the SAH order phases. After that, following De Choudhury et al., (2013) this subsection will also briefly present the diurnal and night activity on Twitter of the selected users in both states and phases. Both characteristics are going to be presented in a descriptive way.

4.1.1. Daily cases and deaths evolution

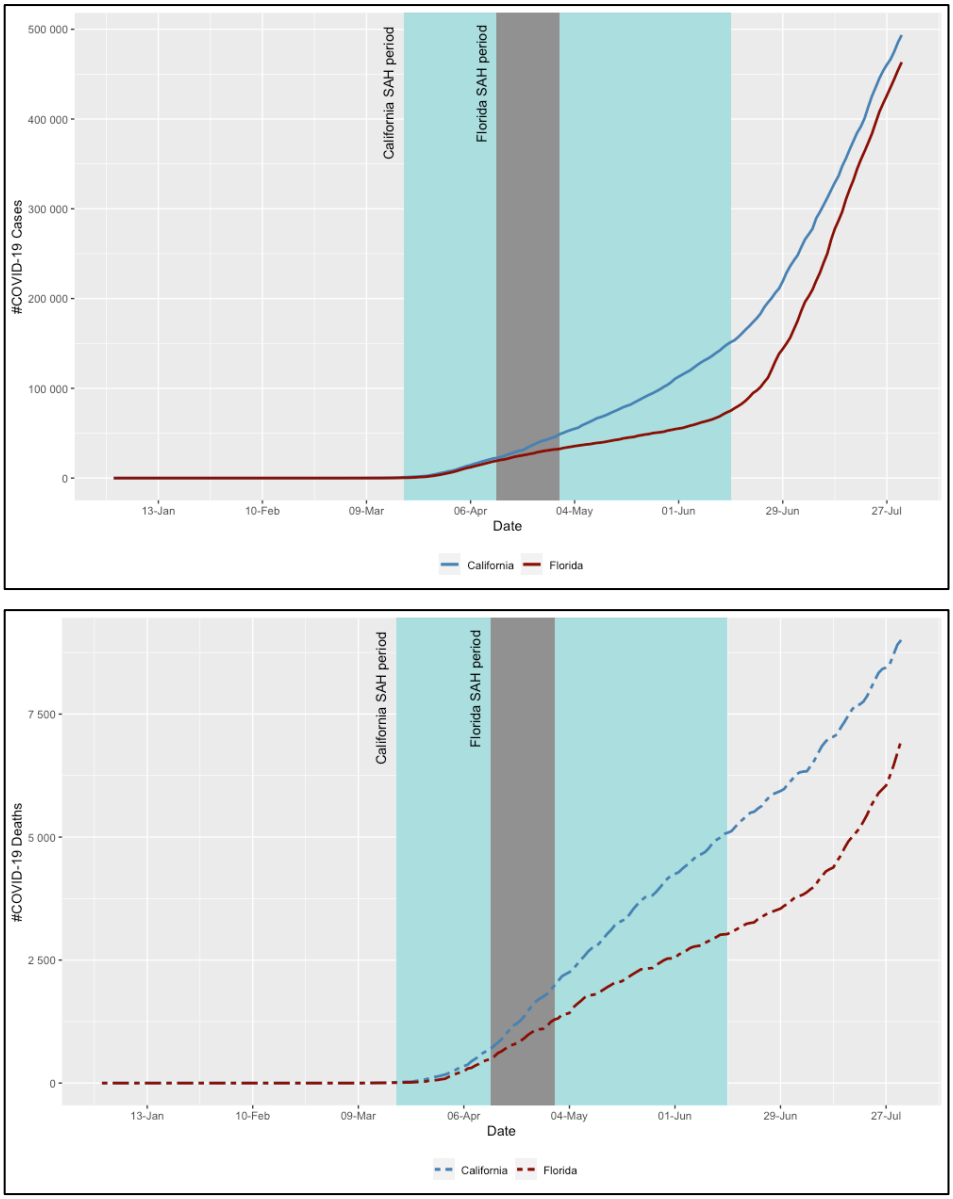
Figure Nº2 show the evolution of both COVID-19 daily cases and deaths in the studied geographies (California and Florida). When comparing the curves, the slopes of the death's figures are steeper, probably explained by the lack of vaccines and treatments in the early stages of the pandemic. California takes the lead in both the number of infections and deaths. However, those are absolute values that shouldn't be compare as the population of California roughly doubles the one in Florida. The figures also help to relativize the SAH duration in each state. In both states, the curves of cases rise sharply once the lockdown restrictions (SAH) were removed. Which shows the contribution of these measures to delay the spread of the virus.

4.1.1. Hourly analysis

De Choudhury, Counts, and Horvitz (2013) present evidence that depressed people tend to be more active on Twitter during the night and in the early morning. This is also backed by the idea that depression symptoms tend to worsen during the night (Lustberg and Reynolds, 2000). On the other hand, non-depressed individuals proved to be more active during the day.

Figure Nº3 presents per state and phase the number of tweets the selected number of users generate per hour (in the y-axis). Additionally, and to make more accurate comparisons (considering that we have less observation in Florida during the SAH order as it entails fewer days) each bar is labeled with the proportion of tweets per hour in each panel (state and phase).

Figure Nº 2: Evolution of the number of COVID-19 daily cases (first panel) and daily deaths (second panel) in California and Florida in the period 01-01-2020 – 31-07-2020

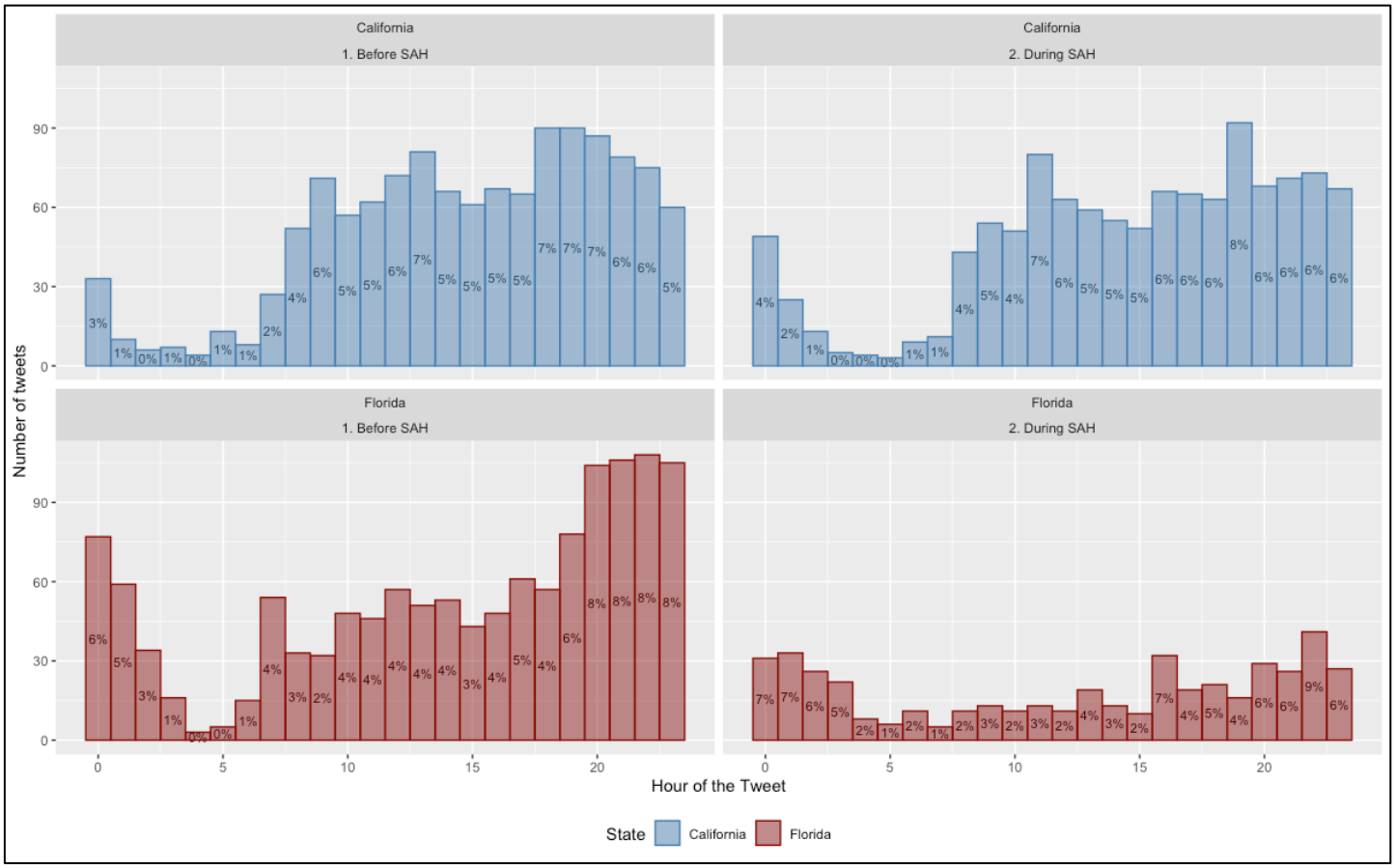


Note: graphs were made by the author using daily cases and deaths figures from: 'The COVID Tracking Project' available at: https://covidtracking.com/data/download

De Choudhury, Counts, and Horvitz (2013) defined a 'night window' between 9 PM and 5:59 AM where depressed people tend to be more active. Following these authors, I found that in California, before the SAH order period, 23,1% of the tweets were done in the night window. However, during the SAH order 27,2% of the tweets were done at night in the same state. In Florida, 39,7% of the tweets before the SAH order were done during the night window, while during the SAH order the

proportion was 48,5% of the post. Therefore, in both states, a slight increase in night activity was registered during the SAH order, although Florida users present a higher level of night activity in both phases when compared with California's users.

Figure Nº 3: Hourly distribution of tweets per state and phase



Note: Percentage labels are calculated per panel.

To make this analysis, the time of the tweets was modified to match the real-time of the sender. In the exported dataset, the date and time were expressed based on UTC (Universal Time Coordinated or Coordinated Universal Time). Therefore, the time of the reported tweets has been adjusted to express the local time of each state. In the case of Florida, the state has two time zones with 1 hour difference between them: UTC -5 and UTC -6 (Time and Date, 2022). From them, the first time zone was chosen as it is the one present in the most populated cities of the state such as Miami, Orlando, Jacksonville, and the state's capital Tallahassee.[18] On the other hand, California has only

[18] For more information visit: https://www.timeanddate.com/worldclock/usa/florida

one-time zone across the whole state (UTC -8)[19]. Hence, eight hours need to be subtracted from the reported date and time in California, while in Florida, five hours have been subtracted.

4.2. Two-Way MANOVA Test: in practice

At this point, it is clear that a Two-Way MANOVA with 2 independent factors (the phase of the SAH order and the state) is going to be conducted. However, one of the challenging parts of this analysis is choosing the more accurate dependent variables to include in the model. This subsection will present the chosen variables used in the MANOVA analysis, together with the test of the assumptions and the results of the test. The following subsection will introduce post-hoc test results.

4.2.1. Variables selection

The chosen dependent variables must not correlate with each other, firstly because it does not have any scientific value. Secondly, as has been explained above, because it will alter the graphical intuition of the MANOVA collapsing one of its dimensions. Bray and Maxwell (1985) highlight that correlation between dependent variables reduces the power of the test, which is what the researcher wants to prevent by performing a MANOVA instead of several ANOVA.

Table Nº 2 presents a selection of the twenty-six prospective dependent variables I had preselected to include in the MANOVA with their specifications. They have been chosen to capture both the way in which people communicate and the emotions they convey while posting on Twitter based on the literature review. Polarity and subjectivity were included as they are the most common sentiment analysis indicators (Thelwall, Buckley, and Paltoglou, 2012). Pronouns were also included as, especially the higher usage of first-person pronouns is an indicator of depression behavior (Safa, Bayat, and Moghtader, 2022; Coppersmith, Dredze, and Harman, 2014). Following other authors indicators of analytical thinking (analytic, authentic, clout) and emotional tone (tone, affect, positive/negative emotions, anxiety, sadness, anger, swear, informal) were also included (Coppersmith, Dredze, and Harman, 2014; Leis, Ronzano, Mayer, Furlong, and Sanz, 2019; Zhang, et al., 2021; Monzani, et al., 2021). Additionally, as depressed people tend to isolate

[19] For more information visit: https://www.timeanddate.com/time/zone/usa/california

themselves and interact less with others, I have included indicators of that behavior such as: family, social, friends.

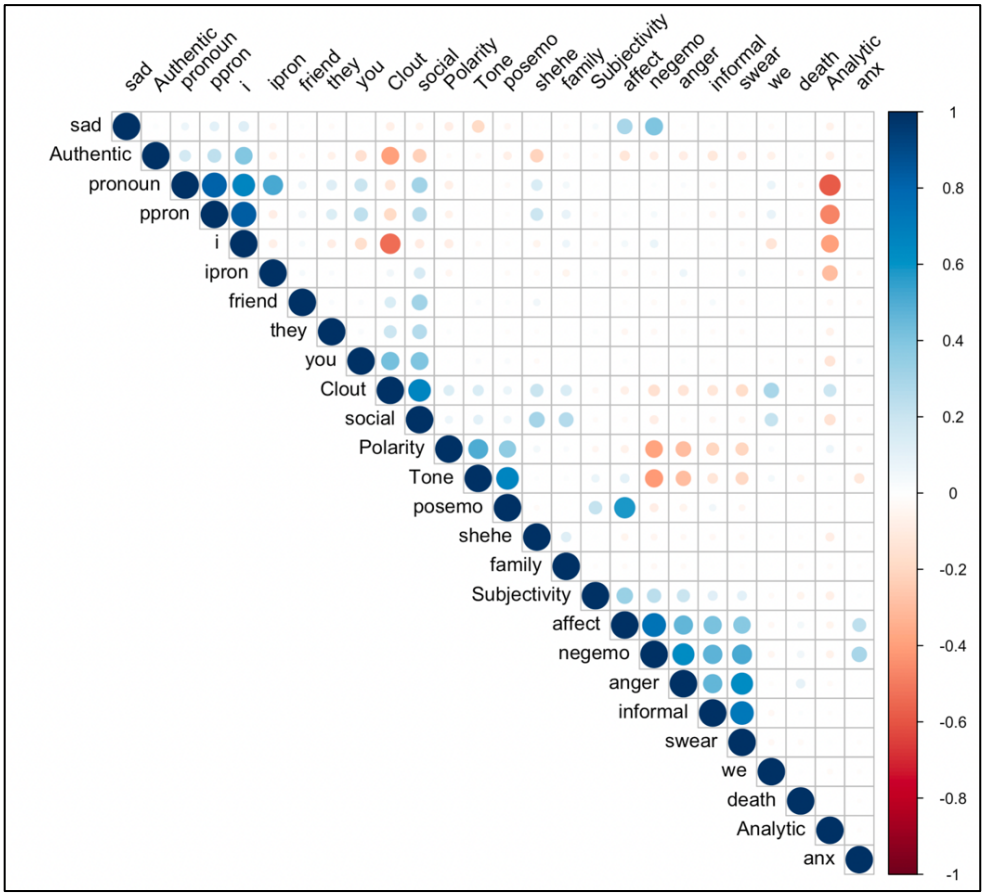
Table Nº 2: Prospective dependent variables to include in the analysis

Abbreviation	Source	Function	Explanation
Polarity	Generated by the researcher using TextBlob (Python library)	Summary Language Variable	Returns the polarity score, float within the range [-1.0, 1.0] (Loria, 2020).
Subjectivity			Returns the subjectivity score, float within the range [0.0, 1.0] where 0.0 is very objective and 1.0 is very subjective (Loria, 2020).
Analytic	LIWC		The analytical thinking variable is a factor-analytically derived dimension based on several categories of function words. Originally published as the Categorical-Dynamic Index, or CDI, Analytic Thinking captures the degree to which people use words that suggest formal, logical, and hierarchical thinking patterns. People low in Analytical Thinking tend to write and think using language that is more intuitive and personal. Language scoring high in Analytic Thinking tends to be rewarded in academic settings and is correlated with things like grades and reasoning skills. Language scoring low in Analytic Thinking tends to be viewed as less cold and rigid, and more friendly and personable (LIWC, 2022).
Clout			Clout refers to the relative social status, confidence, or leadership that people display through their writing or talking. The Clout algorithm was developed based on the results from a series of studies where people were interacting with one another (e.g., Kacewiczet al., 2013). Note that Clout is different from the concept of "Power" (including the LIWC-22 "power" variable). Power or, more accurately, the need for power, reflects people's attention to or awareness of relative status in a social setting. You can have a confident leader who has no interest in other people's standing in the social hierarchy (LIWC, 2022).
Authentic			Authentic reflects to which degree a person is self-monitoring. Therefore, low scores of this variable include prepared texts (speeches) and texts where the person has been socially cautious. On the other hand, high scores are associated to spontaneous conversations among friends with little social inhibitions are present (LIWC, 2022).
Tone			Emotional Tone: includes both positive and negative tone dimensions. In this sense, higher numbers are associated to a positive tone and numbers below 50 suggest a more negative tone (LIWC, 2022).
pronoun		Function words	Total pronouns (I, them, itself).
ppron			Personal pronouns (I, them, her).
i			1st pers singular (I, me, mine).
we			1st pers plural (we, us, our).
you			2nd person (you, your, thou).
shehe			3rd pers singular (she, her, him).
they			3rd pers plural (they, their, they'd).
ipron			Impersonal pronouns (it, it's, those).
affect		Psychological Processes	Affective or emotional processes (happy, ugly, bitter,cried).
posemo			Positive emotion (love, nice, sweet).
negemo			Negative emotion (hurt, ugly, nasty).
anx			Anxiety (worried, fearful).
anger			Anger (hate, kill, annoyed).
sad			Sadness (crying, grief, sad).
social			Social processes (mate, talk, they).
family			Family (daughter, dad, aunt).
friend			Friend (buddy, neighbor).
death		Personal concerns	Death (bury, coffin, kill).
informal			Informal language.
swear			Swear words (fuck, damn, shit).

Note: this table has been constructed for the selected variables following Pennebaker et al. (2015) classification of variables obtained from the LIWC software. The variables obtained using TextBlob Python library have been categorized following the guidelines set by Pennebaker et al. (2015).

To prevent correlations between them and to also reduce the number of variables to include in the model, a correlation matrix has been performed. [20] This matrix (Figure Nº 4) presents in every entry the Person product-correlation coefficients (r_{XY}) calculated for all the variables. The diagonal entries always show the number 1 as they represent the correlation of the variable with itself (Reichwein Zientek and Thompson, 2009). In this graph, the size of each bubble indicates the strength of the correlation (number in absolute value) and the color, the direction (sign). In this way, the correlation of the diagonal is depicted with a big dark blue circle.

Figure Nº 4: Correlation matrix for the pre-selected twenty-six dependent variables



Source: done by the author with variables obtained using LIWC and TextBlob Python's library.

As it is expected, many of the variables show strong correlations, making their presence redundant in the analysis. Some of them are very logical, for instance, the variable *pronoun* which indicates the participation of the total number of pronouns in the tweet, is strongly correlated to the one

[20] Note that only the top right part of the matrix is presented to avoid repletion of variables.

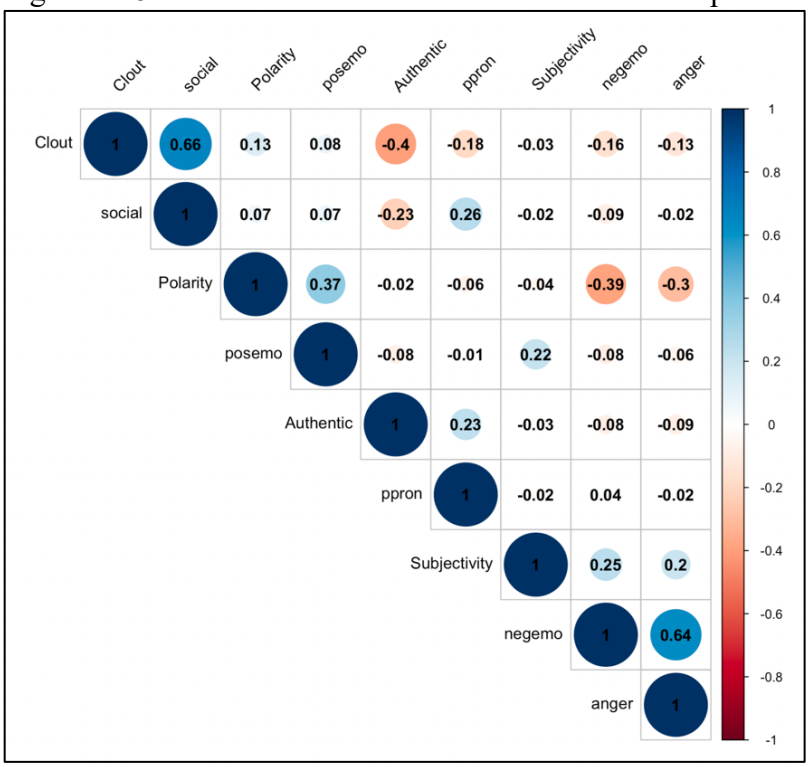
indicated personal pronouns (*ppron*), first-person singular pronouns (*I*), and impersonal pronouns (*ipron*) separately. Also, the variable is negatively associated with 'Analytic' indicating that block of texts with low analytical thinking tend to use more pronouns. The variable that depicts negative emotions (*negemo*) is positively correlated -as expected- with the variables *anger*, *informal*, and *swear*. This indicates that a less formal (nasty) and rude register is used in the presence of irritation. The variable *polarity* is positively correlated to positive emotions (*posemo*) and negatively correlated to the negative ones (*negemo*).

Out of the initial twenty-six prospective response variables, nine have been selected. This selection accounts for the correlation issue and ensures that many aspects of depressed behavior are identified (use of personal pronouns, interactions with others, negative emotions, etc.). In the case of the group of pronouns variables, only personal pronouns remain as they are the type of pronouns most used by depressed individuals. Analytical thinking was excluded despite being use by the literature (Zhang, et al., 2021; Monzani, et al., 2021) because of the high correlation with personal pronouns. Polarity and subjectivity were selected regardless their high correlation with other selected variables (positive/negative emotions) as they are the most used indicators in text analysis (Thelwall, Buckley, and Paltoglou, 2012). Positive and negative emotions were also chosen because the analysis of feelings is at the core of this thesis.

However, variables such as tone, sad, affect, informal, swear were excluded because of their correlation with the selected ones. Clout, associated with confidence and leadership, was selected -apart from being mentioned in the cited literature- as an indicator of the users' self-esteem. It will clarify if depression can, in this case, be associated with low self-esteem. Social was picked to represent interactions with others in the tweets while family was excluded. Finally, authentic was chosen -regardless its correlation with clout- to account for impulsive reactions, also because it is supported by the literature (Zhang, et al., 2021; Monzani, et al., 2021).

The corresponding correlation matrix with the correlation coefficients for the selected dependent variables is presented in Figure Nº5.

Figure Nº 5: Correlation matrix for the selected nine dependent variables



Source: done by the author with variables obtained using LIWC and TextBlob Python's library.

As mentioned before, the mild strong correlation of the polarity with positive emotions (positive correlation) and negative emotions and anger (negative correlation) is still present, as well as the correlation between clout and social (positive) and clout and authentic (negative). This implies that those who refer to their social status, confidence, or leadership abilities (clout), are more likely to tweet about colleagues and other peers (social). They also use more careful speech and pay attention to the selection of words (being less authentic).

4.2.2. Descriptive statistics

As the MANOVA test uses the variation of the variables to estimate differences among groups, some indicators of variation per variable are going to be presented before conducting the test. To illustrate the level of variation of each variable across the groups, Table Nº 3 was created. Appendix B also shows per dependent and group variable the violin plots -density plots combined with Box plots- to have a comprehensible idea of the distribution of the variables across groups.

In the four groups, the variables Polarity, Positive Emotions, Negative emotions, and Anger show a median value of zero. After observing the violin plots shown in the Appendix B, it is clear that

not only the middle observation is zero, but also that there is a big concentration of observations around zero in those variables. This means that both TextBlob Python library (in the case of polarity) and LIWC software (for anger, positive and negative emotions) were not able to identify those emotions in most of the tweets. Therefore, on average both before and during the SAH order, the selected people in California and Florida tend to tweet non-emotional related tweets. While reading the tweets, I noticed that when some degree of irony is present, both softwares fail to identify the emotions the user intended to transmit. This is because words with both positive and negative connotations are involved. Unfortunately, the fact that the four variables directly related to either positive or negative emotions of the sample show medians=0 should not be associated with non-emotional tweets during that period, but to a limitation of this study.

In contrast, the variable that reflects the usage of personal pronouns is not affected by ironic tweets and should be an accurate representation of the writer's intentions after the removal of short form and grammar mistakes by the author. Moreover, it is a variable that many researchers used to identify people with depression as it tends to increase when they experiment mental health disorders (Coppersmith, Dredze, and Harman, 2014; Safa, Bayat, and Moghtader, 2022; Leis et al., 2019; Tackman, et al., 2019). In this case, the state of Florida depicts the expected behavior of the median of *personal pronouns* variable as this one increases by 7% between the phases 'Before SAH' and 'During SAH'. However, the same variable decreases by 6% in California in the same time frame. This might be explained by the fact that the SAH act was almost three times longer in California than in Florida. The median of the variable social, which refers to people talking about others, increases in both states between both phases, 15% in California and from zero to 3,6 in Florida, which contrary to our hypothesis, demonstrates that the selected Twitter users talked more about other people during the SAH order than before it.

Subjectivity, clout, and authentic are variables that show observations all over the distribution (from 0 to 1 in the case of Subjectivity and from 0 to 100 in the other two cases). They also have a greater spread, with standard deviation values which are big enough to represent on average two-thirds of the median values. When comparing the medians in each state, subjectivity in California shows a very small decrease in the median (-4%) while it decreases by 11% in Florida between the two analyzed stages. Authentic shows a small decrease (-3%) in Florida while it drops in

California (between the 'Before SAH' and ´During SAH´ phases) by 12%. Clout´s medians, on the other hand, show different directions in both states in the comparison, it increases in California by 16% while it drops in Florida by 19%.

Overall, the analyzed texts indicate that people were less subjective and authentic during the SAH. However, people tended to be more aware of their social status, leadership abilities, and confidence (clout) during the quarantine in California, while they tend to be less aware of that in Florida during the same period. Which might indicate that people had lower self-esteem in Florida during the SAH compared to the previous period. This is not surprisingly, especially considering that in Florida people used more personal pronouns (classical indicator of depression) during the SAH order than before.

Table Nº 3: Descriptive statistics

	California - Before SAH				California - During SAH			
	Mean	Median	SD	Variance	Mean	Median	SD	Variance
Polarity	0,018	0,000	0,342	0,117	0,030	0,000	0,302	0,091
Subjectivity	0,421	0,467	0,328	0,108	0,411	0,447	0,316	0,100
Clout	33,720	21,180	33,136	1097,986	36,650	24,510	34,103	1163,030
Authentic	65,800	85,210	37,205	1384,182	62,530	74,760	37,599	1413,702
Personal pron.	13,170	13,330	8,711	75,889	12,860	12,500	8,291	68,747
Positive emotions	4,017	0,000	6,777	45,925	3,644	0,000	5,920	35,044
Negative emotions	4,959	0,000	8,907	79,341	4,097	0,000	6,493	42,159
Anger	1,963	0,000	5,264	27,715	1,865	0,000	4,379	19,178
Social	7,894	5,560	9,587	91,914	8,189	6,380	9,250	85,557

	Florida - Before SAH				Florida - During SAH			
	Mean	Median	SD	Variance	Mean	Median	SD	Variance
Polarity	0,054	0,000	0,287	0,082	0,032	0,000	0,268	0,072
Subjectivity	0,384	0,437	0,313	0,098	0,357	0,387	0,316	0,100
Clout	33,090	21,180	32,255	1040,353	32,030	17,230	33,513	1123,153
Authentic	68,160	89,630	37,049	1372,597	64,920	86,720	38,983	1519,652
Personal pron.	12,490	12,500	9,286	86,234	12,990	13,330	9,916	98,328
Positive emotions	3,639	0,000	7,158	51,244	3,971	0,000	7,109	50,542
Negative emotions	3,842	0,000	8,990	80,814	3,586	0,000	7,483	55,991
Anger	1,848	0,000	5,771	33,306	1,521	0,000	4,105	16,852
Social	7,121	0,000	9,387	88,121	7,453	3,570	9,763	95,316

4.2.3. MANOVA assumptions: in practice

In this subsection, the MANOVA assumptions are going to be tested with the selected variables of this study.

1. Independence assumption as every observation represents a tweet that comes from one of the seventeen chosen individuals, this assumption cannot be assured. Although I have analyzed individual autocorrelations and partial correlations over time in each dependent variable. No strong correlation has been spotted which has led to discard a time series analysis.
2. Multivariate normality assumption will be tested with a Shapiro-Wilk test performed on each variable for each group. The results are shown in Table Nº 4.

Table Nº 4: Shapiro-Wilk test results for each variable in each of the selected groups

Dependent Variable	California	
	Before SAH	During SAH
Polarity	W = 0.95143, p-value < 2.2e-16	W = 0.95033, p-value < 2.2e-16
Subjectivity	W = 0.9107, p-value < 2.2e-16	W = 0.9194, p-value < 2.2e-16
Clout	W = 0.84702, p-value < 2.2e-16	W = 0.86045, p-value < 2.2e-16
Authentic	W = 0.79451, p-value < 2.2e-16	W = 0.818, p-value < 2.2e-16
Personal pron.	W = 0.96011, p-value < 2.2e-16	W = 0.96172, p-value < 2.2e-16
Positive emotions	W = 0.65257, p-value < 2.2e-16	W = 0.65379, p-value < 2.2e-16
Negative emotions	W = 0.59594, p-value < 2.2e-16	W = 0.67455, p-value < 2.2e-16
Anger	W = 0.42667, p-value < 2.2e-16	W = 0.49365, p-value < 2.2e-16
Social	W = 0.7889, p-value < 2.2e-16	W = 0.81477, p-value < 2.2e-16

Dependent Variable	Florida	
	Before SAH	During SAH
Polarity	W = 0.92537, p-value < 2.2e-16	W = 0.90372, p-value = 2.416e-16
Subjectivity	W = 0.90414, p-value < 2.2e-16	W = 0.8881, p-value < 2.2e-16
Clout	W = 0.84801, p-value < 2.2e-16	W = 0.82646, p-value < 2.2e-16
Authentic	W = 0.76887, p-value < 2.2e-16	W = 0.77288, p-value < 2.2e-16
Personal pron.	W = 0.94013, p-value < 2.2e-16	W = 0.93912, p-value = 1.118e-12
Positive emotions	W = 0.56295, p-value < 2.2e-16	W = 0.62675, p-value < 2.2e-16
Negative emotions	W = 0.47444, p-value < 2.2e-16	W = 0.54369, p-value < 2.2e-16
Anger	W = 0.35726, p-value < 2.2e-16	W = 0.4323, p-value < 2.2e-16
Social	W = 0.77845, p-value < 2.2e-16	W = 0.78048, p-value < 2.2e-16

After exanimating the output, it is clear that none of the variables per group shows a normal distribution, invalidating the normality assumption of the MANOVA test (as in all cases, the $p - value < 0.05$). By visually exploring the Q-Q plots conducted in this case for all the dependent variables it is also clear that even without dividing observation into the four groups we want to analyze (California and Florida before and during the SAH order), none of the dependent variables follow a normal distribution. Although some of the variables

are closer to a normal distribution than others, such as personal pronouns, clout, and polarity. Q-Q plots of each dependent variable can be found in Appendix C.

3. Homogeneity of variance assumption is tested with Box's M test, where in the H_0 each group has the same covariance matrix. Note that both for this test and the MANOVA itself the group variables included are the state, phase of SAH order, and the interaction between both variables.

Table Nº 5: summary of Box's M test results

Summary for Box's M-test of Equality of Covariance Matrices	
Chi-Sq:	7.540.167
df:	135
p-value:	< 2.2e-16

As it is depicted in Table Nº 5, the null hypothesis is rejected, as the $p - value$ of the test is lower than the significance level α set as 0.05. The rejection of H_0 is serious as groups do not have the same size. The group with more observations is 'Florida before the SAH order' (1293 tweets) and the one with fewer observations is in the same state during the SAH order (454 observations). On the other hand, California has 1243 and 1141 tweets respectively. Therefore, the ratio between the largest and smallest group is 2.8. This implies the lack of robustness for this condition as more variation might come from the group with a smaller sample size. As Box's M test is very sensitive to nonnormality, it will produce biased statistics to be concerned about. Explained by both nonnormality of observations and unequal group sizes.

None of the assumptions that support the findings of the MANOVA test are met. Having this a critical impact on the probability of rejecting the null hypothesis when it is actually true in the population (Type I error). Therefore, the presented results of the test, as well as the consecutive post-hoc test performed, should be considered very carefully.

4.2.4. Two-way MANOVA results

Table Nº 6 shows the results of the MANOVA test with the described statistics. They consist of the F value, $p - value$, degrees of freedom (df) used to calculate the F statistic, and the $p - value$.

Additionally, a measure of the effect size or partial eta squared (η_p^2) with its 95% confidence interval is presented. Note that df is in all the cases equal to one as each group variable has two levels and in the case of the interaction, with only two independent variables, only one interaction is possible. All independent variables have statistically significant effects on at least one dependent variable, with the same significance level (α) across the different tests. In all the cases, the variable state proves to be significant at the 0.1% level while the phase and interaction are also significant but at the 5% level.[21]

On the same line, the effect sizes are very small, although the ones of the state are slightly bigger (η_p^2=0.0097) than the interaction (η_p^2=0.0048) and phase (η_p^2=0.0041). From these results, I can infer that based on the analyzed variables, the state has a stronger impact than the phase or the interaction of both on the way people express themselves on social media. However, as the 95% confidence interval cover all the range of possible variables for the effect size [0 – 1], this variable should be taken as an insight rather than a serious indicator of the effect of the factor variables on the dependent variables.

Table Nº 6: Two-Way MANOVA Results

Independent variable	Test Statistic	df	Value	F	Sig. (*p-value*)	η_p^2
State	Wilk's Lambda	1	0.99028	4.4929	6.78e-06 ***	9.72e-03 [0.00, 1.00]
Phase		1	0.99589	1.8879	0.04918 *	4.11e-03 [0.00, 1.00]
State*Phase		1	0.99512	2.2464	0.01680 *	4.88e-03 [0.00, 1.00]
State	Pillai's Trace	1	0.0097216	4.4929	6.78e-06 ***	9.72e-03 [0.00, 1.00]
Phase		1	0.0041081	1.8879	0.04918 *	4.11e-03 [0.00, 1.00]
State*Phase		1	0.0048845	2.2464	0.01680 *	4.88e-03 [0.00, 1.00]
State	Hotelling's Trace	1	0.0098170	4.4929	6.78e-06 ***	9.72e-03 [0.00, 1.00]
Phase		1	0.0041251	1.8879	0.04918 *	4.11e-03 [0.00, 1.00]
State*Phase		1	0.0049085	2.2464	0.01680 *	4.88e-03 [0.00, 1.00]
State	Roy's Largest Root	1	0.0098170	4.4929	6.78e-06 ***	9.72e-03 [0.00, 1.00]
Phase		1	0.0041251	1.8879	0.04918 *	4.11e-03 [0.00, 1.00]
State*Phase		1	0.0049085	2.2464	0.01680 *	4.88e-03 [0.00, 1.00]

Note: significance codes: '***' 0.001 '**' 0.01 '*' 0.05

4.3. Post-hoc test

Although MANOVA assumptions are not met, given the fact that the results were significant it is worth presenting the performing post-hoc test results. In this case, as a first exploration nine Two-

[21] Note that the interaction effect refers to the fact that the effect of state (or phase) on the dependent variables: polarity, subjectivity, clout, and the like, depends on the level of the other independent variable phase (or state).

Way ANOVA tests are conducted to spot significant results per variable. After that, the results of the second post-hoc test: DDA are going to be presented.

4.3.1 ANOVA as a post-hoc test: in practice

In this section, the output of the calculation of the nine two-way ANOVA run for every response variable is going to be presented. As in the case of MANOVA, the interaction effect between the two independent variables (state and phase) is considered. The output for every test run includes degrees of freedom, total sum of squares, total mean of squares, F statistics, and $p - values$. Note that as in all cases $df = 1$, and therefore $SS_{total} = MS_{total}$.

On the one hand, it is interesting to highlight those cases in which the F statistic is close to 1, as with small mean differences, there are no significant differences among the groups. Those are found in the variable social (across phases), and the variable anger (across states); both with F values greater than 0.9. To a lower extent, we found the variable subjectivity across both states and phases (interaction effect). Therefore, we can say that when comparing the way people express themselves on Twitter before and during the SAH order, there were no differences found in the way people refer to others (social), although significant differences were found in this aspect across states ($p - value < \alpha$). There were no differences in the way people express irritation (anger) among states. Finally, there was no evidence to support differences in the level of subjectivity in the speeches across states and phases taken together. Although, there are significant differences in terms of subjectivity across states ($p - value < \alpha$).

On the other hand, it is relevant to analyze Table Nº 7 focusing on the significance level. Apart from the already mentioned significant variables (social, and subjectivity across states), the following variables should be added to the group. Polarity (the degree of positive/negative tone), and clout (the degree of social status/leadership/confidence) that has only been significantly different across states while authentic (speech preparation), and negative emotions are significantly different across states and phases (considered separately).

Table Nº 7: Two-Way ANOVA Results

Response variable	Independent variable	df	SS_{total}	MS_{total}	F	Sig. (*p-value*)
Polarity	State	1	0.6	0.6029	6406	0.0114 *
	Phase	1	0.0	0.0002	0.002	0.9637
	State*Phase	1	0.3	0.2522	2679	0.1017
Subjectivity	State	1	1.5	1.5342	15117	0.000103 ***
	Phase	1	0.2	0.2378	2343	0.125939
	State*Phase	1	0.1	0.0622	0.613	0.433677
Clout	State	1	5367	5367	4876	0.0273 *
	Phase	1	2047	2047	1860	0.1728
	State*Phase	1	3421	3421	3108	0.0780 .
Authentic	State	1	9571	9571	6819	0.00905 **
	Phase	1	9902	9902	7055	0.00794 **
	State*Phase	1	0	0	0.000	0.98763
Personal pron.	State	1	166	165.60	2080	0.149
	Phase	1	0	0.35	0.004	0.947
	State*Phase	1	139	138.71	1742	0.187
Positive emotions	State	1	13	12.85	0.285	0.593
	Phase	1	13	13.13	0.291	0.589
	State*Phase	1	107	106.88	2370	0.124
Negative emotions	State	1	599	598.8	8942	0.0028 **
	Phase	1	385	385.3	5754	0.0165 *
	State*Phase	1	79	78.7	1176	0.2783
Anger	State	1	24	23.73	0.916	0.339
	Phase	1	30	30.43	1174	0.279
	State*Phase	1	11	11.20	0.432	0.511
Social	State	1	691	690.9	7733	0.00545 **
	Phase	1	89	88.6	0.992	0.31930
	State*Phase	1	0	0.3	0.003	0.95576

Note: significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

4.3.2. Discriminant Descriptive Analysis (DDA) as a post-hoc test: in practice

As in the presented model, there are four groups formed in total (given the existence of two states and two phases), three DDA functions are going to be generated. One was calculated to spot the differences among the states (California and Florida), another for the phase (before or during the SAH order), and a third one for the interaction of the two. In each function, the most relevant response variables (those who contribute the most to the group's differences) are going to be identified.

In this section, the three DDA functions and the (Canonical) Discriminant Function are running using R software following the guidelines provided by Smith et al. (2020) and the code they publish on the Open Science Framework for conducting a DDA as a post-hoc analysis after running a MANOVA.[22] As in the case of the ANOVA, the assumptions needed for performing a DDA analysis are omitted as they are the same used for MANOVA. Although it must be highlighted that

[22] https://osf.io/vyxgt/

given that none of the assumptions are met, the results of the DDA should also be taken as guidelines and put into question.

Before disclosing the results of the three DDA analyses, (Canonical) Discriminant Functions are going to be analyzed. This will ensure that there are significant interactions among the response and group variables. Table Nº 8 provides the outputs of the three-run Discriminant Functions.

Table Nº 8: Main outputs of the canonical discriminant functions

Function	R^2	Eigenvalue	p-value
Canonical Discriminant Function for State	0.0095716	0.0096641	8.726e-06 ***
Canonical Discriminant Function for Phase	0.0041081	0.0041251	0.04905 *
Canonical Discriminant Function for State*Phase	0.0048845	0.0049085	0.01674 *

Note: Significance codes: '***' 0.001 '**' 0.01 '*' 0.05

According to the table, the squared canonical correlation (R^2) that shows the degree of correlation among the group variable and the selected response variables, and the eigenvalues (ratios of between-groups to within-groups sum of squares) are very similar. Group differences are slightly bigger among states (R^2= 0.96%) than among phases of the SAH order (R^2=0.41%), and when we consider the interaction of the two as grouping variable (R^2=0.49%).

Finally, every $p - value$ associated with the Discriminant Functions is significant at the standard level of significance $\alpha = 0.05$. This is paramount to continue with the DDA analysis as it depicts that at least one correlation between any of the response variables and the group variables is significant. Although all functions show significant results, the big majority of the variation among groups is not explained by this model.

Th following three tables will show the results of the DDA analysis per group (state, phase, and the interaction of both). These includes standardized coefficients (analog to the betas in regressions), structure coefficients (r_s), squared structure coefficients (r_s^2), and the parallel discriminant ratio coefficients. First, the results of the DDA function run for the state group are going to be presented, then the corresponding one for the phase, and finally, the one that deals with the interaction of the two.

Table Nº 9: Standardized Discriminant Function Coefficients, Structure Coefficients, and Parallel discriminant ratio coefficients **for State Effect** on nine response variables

Response variable	Standardized coefficient	r_s	r_s^2	Parallel discriminant ratio coefficient
Subjectivity	-0.57544501	-0.6554275	0.42958515	0.377162459
Clout	-0.19537296	-0.2938884	0.08637039	0.057417843
Personal pron.	-0.21918839	-0.2262929	0.05120849	0.049600782
Positive emotions	-0.07944663	-0.1021302	0.01043057	0.008113896
Negative emotions	-0.53240081	-0.5474847	0.29973950	0.291481296
Social	-0.23330711	-0.3960071	0.15682165	0.092391279
Polarity	0.32728730	0.3910213	0.15289766	0.127976305
Anger	0.34552044	-0.1863429	0.03472367	-0.064385275
Authentic	0.20377036	0.3112230	0.09685977	0.063418027

Considering that standardized coefficients are analog to beta coefficients in regular regression, they can also be presented in the following form[23]:

$$DDA\ Score_{state} = -0.58Subjectivity - 0.20Clout - 0.22Personal\ pron. - 0.08Positive\ emotions - 0.53Negative\ emotions - 0.23Social + 0.32Polarity + 0.34Anger + 0.20Authentic$$

Note that the dependent variable of the equation is called *score* as it is used to calculate the score of every observation in this discriminant function (that has state as the dependent variable). Following Warne (2014) the first coefficient (-0.58) of the equation can be interpreted as follows: for every 1 standard deviation increase in the subjectivity value of the tweet, the DDA score is predicted to decrease by 0.58 standard deviations, keeping all the other variables constant. In the same way, we can say that for every 1 standard deviation increase in the polarity of the tweet, the DDA score is set to increase by 0.32 standard deviations, ceteris paribus.

Regarding the **structure coefficients** (r_s), subjectivity (-0.66) and negative emotions (-0.57) are the variables with the strongest correlations with the state grouping variable, followed by social (-0.40) and polarity (0.39). The **squared of these values (r_s^2)** the results show that subjectivity accounts for 37,7% of the variance in the score of this function, followed by negative emotions (29,1%) and polarity (12,8%).

[23] Note that the coefficients of the regression are the standardized coefficients presented in the table rounding to the nearest two decimals.

The **parallel discriminant ratio coefficients** give a measure of the relative importance of the coefficients in the equation. In this case, subjectivity (0.38) is the variable that contributes the most to group's separations (differences among California and Florida), followed by negative emotions (0.29).

Considering now the phase of the quarantine as the grouping variable. Table Nº 10 show that authentic is the variable that accounts for most of the variation of the score of these functions (r_s^2=41.4%). Followed by negative emotions (r_s^2=33.8%) and subjectivity (r_s^2=13.8%). By looking at the parallel discriminant ratio coefficient, the variables that contribute most to group separations are also authentic (0.49) and negative emotions (0.46).

Table Nº 10: Standardized Discriminant Function Coefficients, Structure Coefficients, and Parallel discriminant ratio coefficients **for Phase Effect** on nine response variables

Response variable	Standardized coefficient	r_s	r_s^2	Parallel discriminant ratio coefficient
Subjectivity	-0.206780185	-0.37112575	0.1377343212	0.0767414511
Clout	0.001960056	0.33076206	0.1094035371	0.0006483122
Personal pron.	0.177249953	-0.01634198	0.0002670602	-0.0028966144
Positive emotions	-0.105929485	-0.13117035	0.0172056618	0.0138948081
Negative emotions	-0.794025881	-0.58172537	0.3384044111	0.4619050030
Social	-0.025967309	0.24143490	0.0582908095	-0.0062694147
Polarity	-0.249237563	-0.01072377	0.0001149993	0.0026727675
Anger	0.139438890	-0.26309896	0.0692210608	-0.0366862263
Authentic	-0.761449941	-0.64377956	0.4144521245	0.4902059095

Table Nº 11 presents the results of the interaction effect between the state and phase variables as grouping factor. There, the variables that contribute the most to the function's variance are clout (r_s^2=15.3%), followed by polarity (r_s^2=13.3%) and positive emotions (r_s^2=11.8%). In the same fashion, clout, positive emotion, and polarity contribute the most to group separations. They have respectively, parallel discriminant ratio coefficients equal to 0.32, 0.22, and 0.20.

Table Nº 11: Standardized Discriminant Function Coefficients, Structure Coefficients, and Parallel discriminant ratio coefficients **for State by Phase Interaction Effect** on nine response variables

Response variable	Standardized coefficient	r_s	r_s^2	Parallel discriminant ratio coefficient
Subjectivity	-0.346326084	-0.172623555	0.029798891612	0.0597840397
Clout	-0.813745446	-0.391525966	0.153292582184	0.3186024719
Personal pron.	-0.007001249	0.294246015	0.086580717608	-0.0020600895
Positive emotions	0.638474753	0.342923272	0.117596370354	0.2189478513
Negative emotions	0.454369150	0.242097864	0.058611375623	0.1100018005
Social	0.525059339	0.013155345	0.000173063115	0.0069073370
Polarity	-0.552166036	-0.364890083	0.133144772840	0.2014799107
Anger	-0.606285675	-0.145958079	0.021303760721	0.0884922922
Authentic	-0.190998616	0.002807826	0.000007883887	-0.0005362909

5. Discussion

Results of both the MANOVA test and post-hoc tests emphasize the existence of greater differences across states (than phases or even the interaction of both). Nevertheless, these results should be taken very carefully because of two main reasons. Firstly, none of the assumptions of the conducted test were met. Secondly, despite having a sample of more than 4000 observations (tweets), they are generated by seventeen individuals (nine in California and eight in Florida), who have a great influence on the results. Unfortunately, only information about the mentioned individuals was gathered because of the restriction imposed on the data: the scope was limited to individuals with public Twitter accounts with georeferenced profiles (from where the state location was extracted). They should also need to manifest in Twitter being diagnosed with depression or PTSD in the period April – June of 2020.

However, differences across the pandemic phases (before and during the SAH acts) are more interesting to look at. They involve changes in the way the same group of people expresses themselves on social media. In this sense, *authenticity* (degree of speech preparation) and *negative emotions* tend to differ across phases, considering the parallel discriminant ratio coefficients. However, to understand what happened to those variables we need to consider the descriptive statistics presented. They show that the degree of authenticity decreased in both states by 5% when considering the mean differences. Nevertheless, as those variables can be impacted by extreme values it is worth mentioning that the median values also decrease, 12% in California and 3% in Florida. Therefore, people were less authentic and natural during the SAH period, implying that they were more cautious and prepared their messages better during the imposed quarantine.

When it comes to differences between negative emotions across phases, in all cases the median levels were cero. Nevertheless, the mean values show a 17% decrease in California and a 7% drop in Florida. Those decreasing negative emotions values contradict the main hypothesis of this thesis that people in California express more negative emotions during the SAH order. However, the literature review supports those findings. This is the case of Wang, Fan, Palacios, Chai, Guetta-Jeanrenaud, Obradovich, Zhou, and Zhen (2022) who report that globally, considering the first wave of the pandemic (from 1st of January to 31st of May, 2020) a sharp decline in sentiments scores was followed by a slower and asymmetric recovery, Additionally, Wang, Huang, Hu, Zhang, Li, Ning, Corcoran, Khan, Liu, Zhang, and Li (2022) also found in Australia a pessimistic phase (between 1st of January and 10th of March, 2020) followed by a more optimistic one (in the following three months). Therefore, as the pessimistic phases of these studies match the ´Before SAH´ phase of this one, it can be said that the results found are aligned with the literature.

The interaction group (state*phase) shows differences in the variables: clout, positive emotions, and polarity when considering parallel discriminant ratio coefficients. When taking a look at the descriptive statistics, positive emotions and polarity shows contradictory results on both states and are not going to be analyzed. However, the variable clout, shows an increase in California (during the SAH) and a decrease in Florida. This result is important as this variable can be considered as an indicator of self-esteem.

The main hypothesis of this thesis was to prove if California experienced more negative emotions during the SAH order than Florida. However, what I missed was that the pandemic exacerbates negative emotions, but it happened at the very beginning of it, even before the pandemic declaration or the lockdown restrictions. Therefore, the hypnotized increased levels of fear happened before the declaration of the SAH orders and might even decrease or deaccelerating during lockdown periods. This might be supported by feelings of security associated with governmental measures, along with fewer levels of uncertainty time brings. Moreover, it the coming month of the start of the pandemic, medical information on treatments and vaccines also helps people to feel safer.

Another interesting insight is that people were more active at night during the SAH order as the participation of night tweets increases in both states considerably. However, this can be associated with getting familiar with home office work schemas and the reduction of recreational options lockdowns entails. Therefore, it is difficult to disentangle if this is an indication of earlier depress-related symptoms. Which are in line with the reviewed (Lustberg and Reynolds, 2000; De Choudhury, Counts, and Horvitz, 2013; Coppersmith, Dredze, and Harman, 2014; Zhang, Lyu, Liu, Zhang, Yu, Luo, 2021).

As it was mentioned before, differences among states should not be considered given the small number of individuals studied. However, it is still worth analyzing the obtained results and comparing them with the available literature. In this sense, both post-hoc tests pointed out that the variables social, subjectivity, and polarity account for the greatest differences among states. Those variables might differ just because very few different individuals are considered, who might have singular ways of expressing their emotions. However, it is still relevant to look at the descriptive stats. The comparison of both states across the two phases shows that in California people are slightly more subjective and social (talk more about others) while in Florida people are slightly more optimistic (slightly higher polarity scores). This last finding is aligned with Zhang, Lyu, Liu, Zhang, Yu, and Luo (2021) publication which states that people in Florida have shown lower depression scores (when compared to the national level and the states of California and New York).

Although the biggest limitations of this study were already presented, there are other worth mentioning. They are going to be enumerated as follows:

- The use of a dictionary-based approach is not able to account for irony or sarcasm, which is present in many of the analysed tweets. This was spotted by the researcher while manually correcting for short forms and spelling mistakes and comparing the first round of polarity, subjectivity, positive, and negative emotions scores. To clarify, a sentence such as *´What a wonderful day, my girlfriend is done with me´* might be receiving some positive emotions points because of the inclusion of the word *wonderful*.
- Demographic variables in the analysis such as age, gender, and marital status are not available. Therefore, they cannot be used as control variables.

- The sample is only formed by Twitter users, which might not be an accurate representation of the total community of each state. In this respect, De Choudhury, Counts, and Horvitz (2013) manifest that one of the limitations in their study is an inherent population bias as from the people who use internet, only a small portion use Twitter. A recent study shows that only 23% of American adults use the bird social platform (Meltem, 2022).

6. Conclusion

This study aims to contribute to the growing plethora of research in the sentiment analysis arena using social media posts. This technique proves to be very efficient as significantly reduces the cost of gathering information and avoids questionary and memory bias. It also allows reaching a huge amount of people in real-time. Needless to say, this research method has a promising future ahead. However, it will also require researchers to master and develop more advanced techniques that properly deal with user-generated content. Among the limitations of those observations are, the use of ironical expression by the sender of the message, which can bias the results obtained especially by dictionary-based approaches. Moreover, the scope of this analysis is restricted by only considering social media users, which might not be an accurate representation of the overall population. However, when it comes to analyzing depression, its power relies on being able to collect data where people are less reluctant to openly manifest those insights in a survey or medical environment.

This is, to the best of the author's knowledge, the first attempt to detailed compare the performance of the states of California and Florida before and during the State at Home (SAH) act on depressed individuals. However, the limited number of individuals that use Twitter to share their depression or PTSD diagnosis after the outbreak of the COVID-19 pandemic is the main limitation of this analysis. Which jeopardizes comparisons done across states and the test of the main hypothesis of this research.

Nevertheless, interesting insights aligned with the literature were found when comparing the post done before and during the SAH order in each state. Considering a timespan between the 1st of January 2020 and the start of the lockdown (SAH order) in California and Florida, both states registered more negative emotions than in the consecutive phase of lockdown. This finding is

aligned with the literature review and supported by the idea that higher uncertainty levels and fear of the unknown consequences of the virus were present before governments took action. Additionally, in both states, authenticity levels decrease during the SAH order, showing that people were more careful when posting in Twitter.

On the other hand, in both states, an increase in night activity (considering the participation of night tweets) was registered during the SAH order. However, this finding, associated with depression, should not be taken as a contradiction to the decrease in negative emotions during this phase. This is because, new routines associated with lockdown (home office, less free-time activities) might have alter people's night window.

References

AP Central. (2022, December 18). Power in Tests of Significance. Retrieved from AP Central College Board: https://apcentral.collegeboard.org/courses/ap-statistics/classroom-resources/power-in-tests-of-significance

Banerjee, A., Chitnis, U., Jadhav, S., Bhawalkar, J., & Chaudhury, S. (2009). Hypothesis testing, type I and type II errors. Industrial Psychiatry Journal, 127-31.

Boyd, R. L., & Pennebaker, J. W. (2016). A Way With Words: Using Language for Psychological Science in the Modern Era. In C. V. Dimofte, C. P. Haugtvedt, & R. F. Yalch, Consumer Psychology in a Social Media World (pp. 222-236). New York: Routledge.

Bray, J., & Maxwell, S. (1985). Multivariate Analysis of Variance. SAGE Publications, Inc, 1-80.

Chandra Guntuku, S., Yaden, D., Kern, M., Ungar, L., & Eichstaedt, J. (2017). Detecting depression and mental illness on social media: an integrative review. Current Opinion in Behavioral Sciences, 43-49.

Chen, X., Sykora, M., Jackson, T., & Elayan, S. (2018). What about mood swings? Identifying depression on Twitter with temporal measures of emotions. Companion Proceedings of The Web Conference 2018, (pp. 1653–1660). Lyon.

Chua, C. E., Storey, V., Li, X., & Kaul, M. (2019). Developing insights from social media using semantic lexical chains to mine short text structures. Decision Support Systems, 1-10.

Coppersmith, G., Dredze, M., & Harman, C. (2014). Quantifying Mental Health Signals in Twitter. Proceedings of the workshop on computational linguistics and clinical psychology: From linguistic signal to clinical reality, 51-60.

De Choudhury, M. (2013). Role of Social Media in Tackling Challenges in Mental Health. Microsoft Research, 49-52.

De Choudhury, M., Counts, S., & Horvitz, E. (2013). Social Media as a Measurement Tool of Depression in Populations. In Proceedings of the 5th ACM International Conference on Web Science, 1-10.

De Choudhury, M., Gamon, M., Counts, S., & Horvitz, E. (2013). Predicting Depression via Social Media. Microsoft Research, 1-10.

Enders, C. K. (2003). Performing Multivariate Group Comparisons Following a Statistically Significant MANOVA. Measurement and Evaluation in Counseling and Development, 40-56.

Ettman, C., Cohen, G., Abdalla, S., Sampson, L., Trinquart, L., Castrucci, B., . . . Galea, S. (2022). Persistent depressive symptoms during COVID-19: a national, population-representative, longitudinal study of U.S. adults. Lancet, 10091-10096.

Executive Department State of California. (2022, March 19). Retrieved from Executive Order N-30-20: https://covid19.ca.gov/img/Executive-Order-N-33-20.pdf

Foster, G., Lane, D., Scott, D., Hebl, M., Guerra, R., Orsherson, D., & Zimmer, H. (2018). An Introduction to Psychological Statistics. Missouri: Open Educational Resources Collection.

Frizell, S. (2014, January 14). Time. Retrieved from How Twitter Knows When You're Depressed: https://time.com/1915/how-twitter-knows-when-youre-depressed/

Huberty, C. J., & Olejnik, S. (2006). Applied MANOVA and Discriminant Analysis. Georgia: Wiley.

IBM. (2022, September 13). Multivariate Tests of Within-Subjects Effects. Retrieved from SPSS Statistics Documentation: https://www.ibm.com/docs/en/spss-statistics/29.0.0?topic=variance-multivariate-tests-within-subjects-effects

Kaur, H., Ahsaan, S., Alankar, B., & Chang, V. (2021). A Proposed Sentiment Analysis Deep Learning Algorithm for Analyzing COVID-19 Tweets. Information Systems Frontiers, 1417–1429.

King, A., & Eckersley, R. (2019). Numerical Methods to Investigate Whether a Sample Fits a Normal Distribution. In A. King, & R. Eckersley, Statistics for Biomedical Engineers and Scientists How to Visualize and Analyze Data (pp. 156-158). London: Academic Press.

Klas, M. E., & Contorno, S. (2020, April 1). Tampa Bay Times. Retrieved from Florida Gov. Ron DeSantis issues statewide stay-at-home order: https://www.tampabay.com/news/health/2020/04/01/florida-gov-ron-desantis-issues-statewide-stay-at-home-order/

Leis, A., Ronzano, F., Mayer, M., Furlong, L., & Sanz, F. (2019). Detecting Signs of Depression in Tweets in Spanish: Behavioral and Linguistic Analysis. Journal of Medical Internet Research.

LIWC. (2022, December 18). LIWC Help Page. Retrieved from LIWC ANALYSIS: https://www.liwc.app/help/liwc

Loewen, S., & Plonsky, L. (2016). An A – Z of Applied linguistics Research Methods. London: Macmillan Education.

López-Chau, A., Valle-Cruz, D., & Sandoval-Almazán, R. (2020). Sentiment Analysis of Twitter Data Through Machine Learning Techniques. In M. Ramachandran, & Z. Mahmood, Software Engineering in the Era of Cloud Computing (pp. 185-210). Mexico City: Springer.

Loria, S. (2020, April 26). Textblob Documentation. Release 0.16.0. Retrieved from https://buildmedia.readthedocs.org/media/pdf/textblob/latest/textblob.pdf

Lustberg, L., & Reynolds, C. F. (2000). Depression and insomnia: questions of cause and effect. Sleep Medicine Reviews, 253–262.

Meltem, O. (2022, May 5). Pew Research Center. Retrieved from 10 facts about Americans and Twitter: https://www.pewresearch.org/fact-tank/2022/05/05/10-facts-about-americans-and-twitter/

Mike, T., Buckley, K., & Georgios, P. (2012). Sentiment strength detection for the social web. Journal of the American Society for Information Science and Technology, 163-173.

Monzani, D., Vergani, L., Francesca, S., Pizzoli, S., Marton, G., & Pravettoni, G. (2021). Emotional Tone, Analytical Thinking, and Somatosensory Processes of a Sample of Italian Tweets During the First Phases of the COVID-19 Pandemic: Observational Study. Journal of Medical Internet Research, 1-11.

Nadeem, M., Horn, M., Coppersmith, G., & Sen, S. (2016). Identifying Depression on Twitter. ArXiv, 1-9.

Newman, M., Pennebaker, J., Berry, D., & Richards, J. (2003). Lying Words: Predicting Deception From Linguistic Styles. Personality and Social Psychology Bulletin, 665-675.

Patel, S., & Bhavsar, C. D. (2013). Analysis of pharmacokinetic data by Wilk's lambda. International Journal of Pharmaceutical Science Invention, 36-44.

Pennebaker, J., Boyd, R., Jordan, K., & Blackburn, K. (2015). The Development and Psychometric Properties of LIWC2015. Austin, TX: University of Texas at Austin.

Reichwein Zientek, L., & Thompson, B. (2009). Matrix Summaries Improve Research Reports: Secondary Analyses Using Published Literature. Educational Researcher, Vol. 38, Nº 5, 343-352.

Safa, R., Bayat, P., & Moghtader, L. (2022). Automatic detection of depression symptoms in Twitter using multimodal analysis. The Journal of Supercomputing, 4709-4744.

Santomauro, D. (2021). Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. Lancet, 1700-1712.

Sarma, K., & Vishnu Vardhan, R. (2018). Multivariate Statistics made simple. A practical approach. New York: CRC Press.

Shepherd, J. (2022, August 4). Social Shepherd. Retrieved from 22 Essential Twitter Statistics You Need to Know in 2022: https://thesocialshepherd.com/blog/twitter-statistics

Sherry, A. (2006). Discriminant Analysis in Counseling Psychology Research. The Counseling Psychologist, 661-683.
Smith, K. N., Lamb, K. N., & Henson, R. K. (2020). Making Meaning out of MANOVA: The Need for Multivariate Post Hoc Testing in Gifted Education Research. SAGE, 41-55.
Stahle, L., & Wold, S. (1990). Multivariate analysis of variance (MANOVA). Chemometrics and Intelligent Laboratory Systems, 127–141.
Stevens, J. P. (2009). Applied Multivariate Statistics for the Social Sciences, Fifth Edition. Routledge Academic.
Steyn Jr, H. S., & Ellis, S. M. (2009). Estimating an Effect Size in One-Way Multivariate Analysis of Variance (MANOVA). Multivariate Behavioral Research, 106-129.
Tabachnick, B., & Fidell, L. (2007). Experimental Designs using ANOVA. Belmont: Duxbury.
Tackman, A., Sbarra, D., Carey, A., Donnellan, B., Horn, A., Holtzman, N., . . . Mehl, M. (2019). Depression, negative emotionality, and self-referential language: A multi-lab, multi-measure, and multi-language-task research synthesis. Journal of Personality and Social Psychology, 817-834.
Thelwall, M., Buckley, K., & Paltoglou, G. (2012). Sentiment strength detection for the social web. Journal of the American Society for Information Science and Technology, 163-173.
Thomas, R., & Zumbo, B. D. (1996). Using a Measure of Variable Importance to Investigate the Standardization of Discriminant Coefficients. Journal of Educational and Behavioral Statistics, 110-130.
Time and Date. (2022, November 27). Retrieved from Time Zones in Florida, United States: https://www.timeanddate.com/time/zone/usa/florida
University of Virginia Library. (2022, December 11). Research Data Services + Sciences. Retrieved from Understanding Q-Q Plots: https://data.library.virginia.edu/understanding-q-q-plots/
Van Der Zee, S., Poppe, R., Havrileck, A., & Baillon, A. (2021). A Personal Model of Trumpery: Linguistic Deception Detection in a Real-Wold High-Stakes Setting. Association for Psychological Science, 3-17.
Varghese Babu, N., & Kanaga, E. G. (2021). Sentiment Analysis in Social Media Data for Depression Detection Using Artificial Intelligence: A Review. SN Computer Science, 1-20.
Wang, J., Fan, Y., Palacios, J., Chai, Y., Guetta-Jeanrenaud, N., Obradovich, N., Zheng, S. (2022). Global evidence of expressed sentiment alterations during the COVID-19 pandemic. Nature Human Behaviour, 349-358.
Wang, S., Huang, X., Hu, T., Zhang, M., Li, Z., Ning, H., . . . Li, X. (2022). The times, they are changing: tracking shifts in mental health signals from early phase to later phase of the COVID-19 pandemic in Australia. BMJ Global Health, 1-9.
Warne, R. (2014). A Primer on Multivariate Analysis of Variance (MANOVA) for Behavioral Scientists. Practical Assessment, Research & Evaluation. Vol 19, Nº 17, 1-10.
Wooldridge, J. M. (2020). Introductory Econometrics. A Modern Approach. Boston: Cengage.
World Health Organization. (2021, September 13). Retrieved from Depression: https://www.who.int/news-room/fact-sheets/detail/depression
World Health Organization. (2022, March 2). Retrieved from COVID-19 pandemic triggers 25% increase in prevalence of anxiety and depression worldwide: https://www.who.int/news/item/02-03-2022-covid-19-pandemic-triggers-25-increase-in-prevalence-of-anxiety-and-depression-worldwide
Yadav, A., & Vishwakarma, D. K. (2020). Sentiment analysis using deep learning architectures: a review. Artificial Intelligence Review, 4335–4385.

Zhang, Y., Lyu, H., Liu, Y., Zhang, X., Wang, Y., & Luo, J. (2021). Monitoring Depression Trends on Twitter During the COVID-19 Pandemic: Observational Study. JMIR Publications: Public Health Emergency Collection, 1-12.

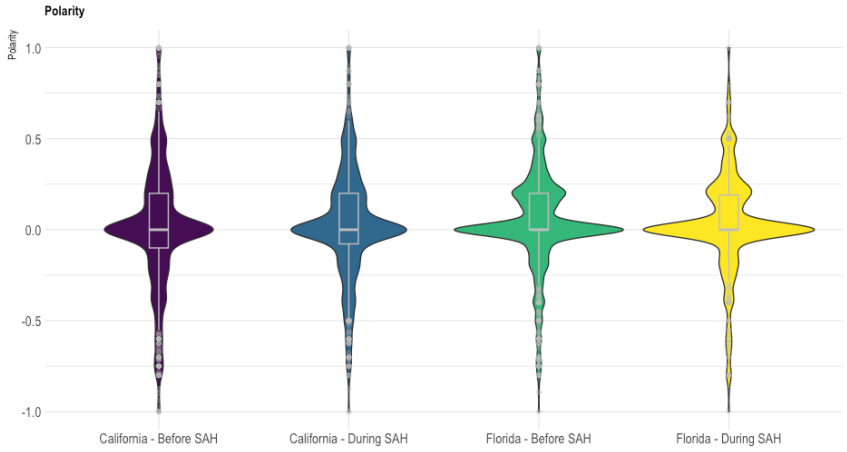
Zhou, J., Zogan, H., Yang, S., Jameel, S., Xu, G., & Chen, F. (2021). Detecting Community Depression Dynamics Due to COVID-19 Pandemic in Australia. IEEE TRANSACTIONS ON COMPUTATIONAL SOCIAL SYSTEMS.

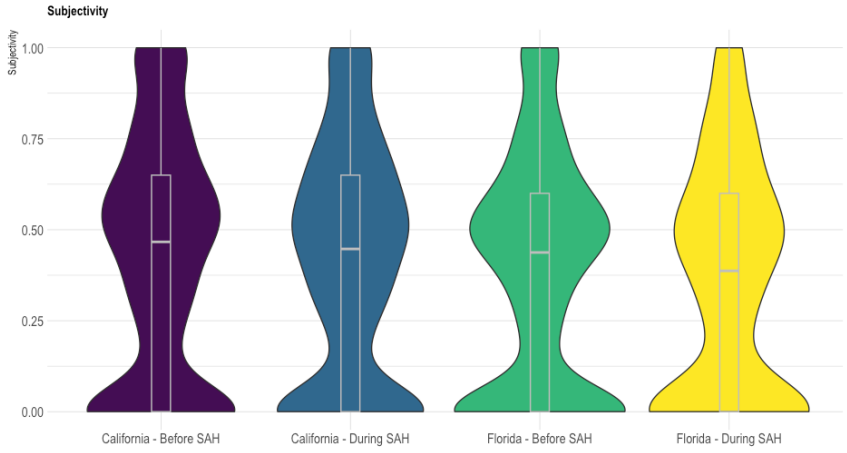
Appendix A

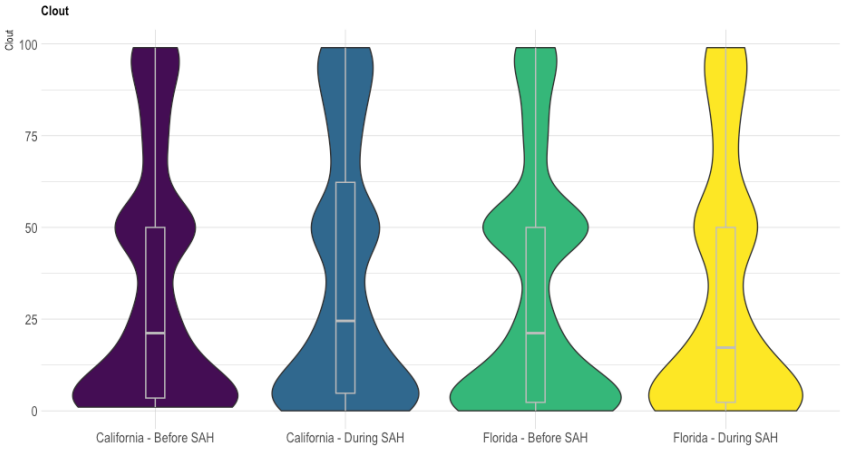
Short Form	Expression
ABT	About
ACAB	All Cops Are Bastards
AF	As fuck
ASF	As fuck
ATL	Above the line
ATP	At this point
BBW	big beautiful woman
BC	Because
BFF	Best friend
BLM	Black Lives Matter
BM	Black Man
DM	Direct message
DMV	Department of Motor Vehicles
DND	do not disturb
DTF	down to fuck
EDM	Electronic dance music
FKN	Fucking
FR	For real
FYE	excellent or in a state of excitement.
IBS	Irritable bowel syndrome
IDC	I don't care
IDGAF	I don't give a fuck
ILY	I love you
IMMA	I'm going to
IRL	in real life
ISH	Near or about; approximately
IWNY	Internet Week New York
LDR	long-distance relationship
LIL	Little
LMAO	Laughing my ass off
LMK	Let me know
LOL	Laugh out loud
MFS	motherfucker
NFL	National Football League
NGL	Not gonna lie
NVM	Never Mind
OMFG	Oh my fucking God
OMG	Oh my God
OTW	On the way
PLZ	Please
SZN	season
TF	The fuck
TMI	Too much information
UTI	Urinary tract infection
V-Day	Valentine's Day
vacay	vacation
WBY?	What about you?
WRT	With reference to
WTH	What the hell

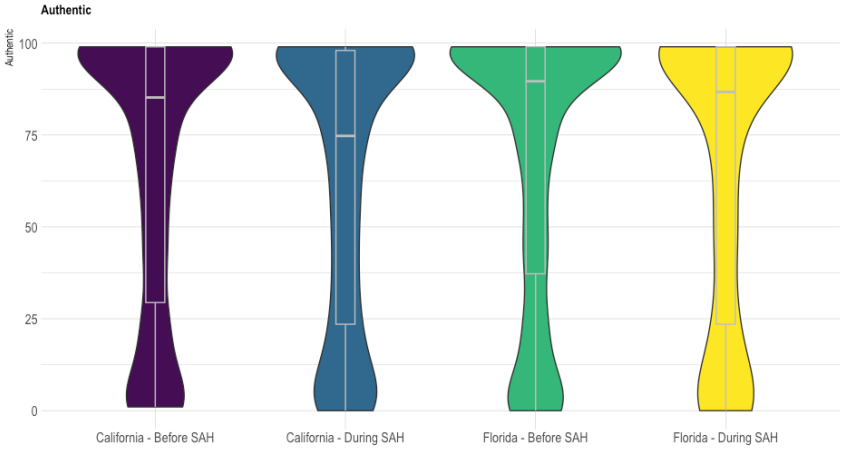
Appendix B

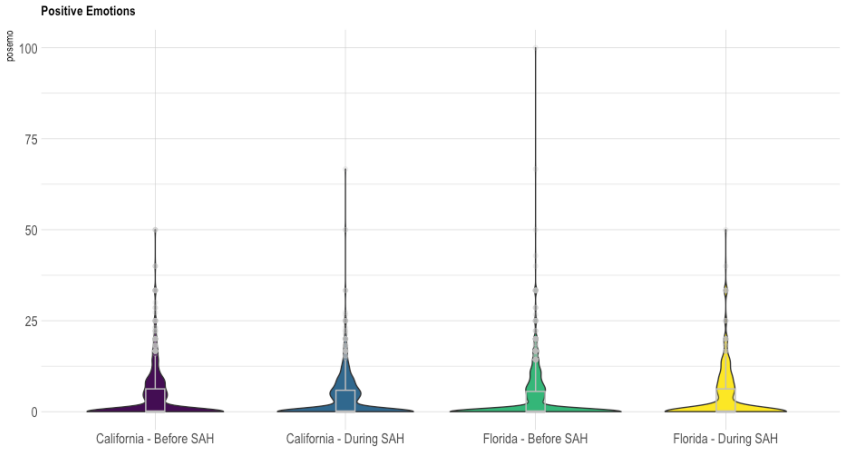
Violin Plots

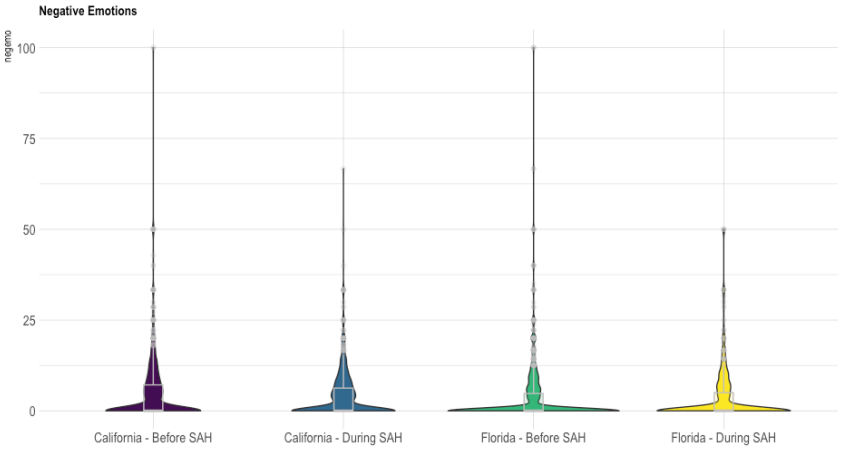


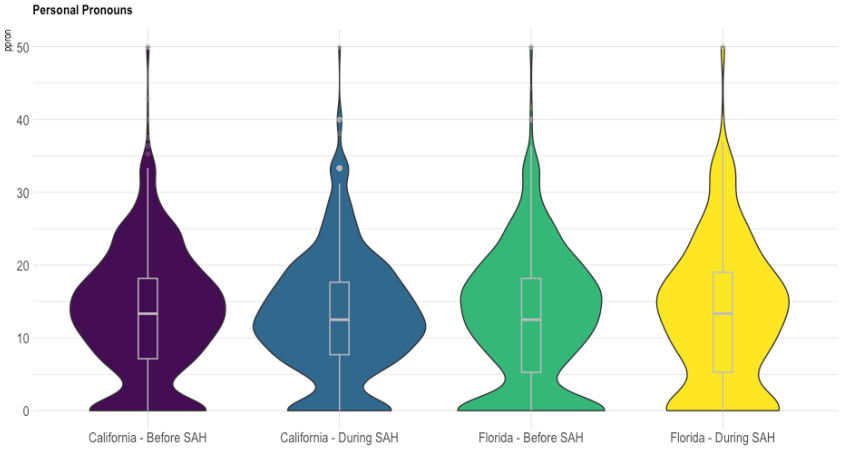


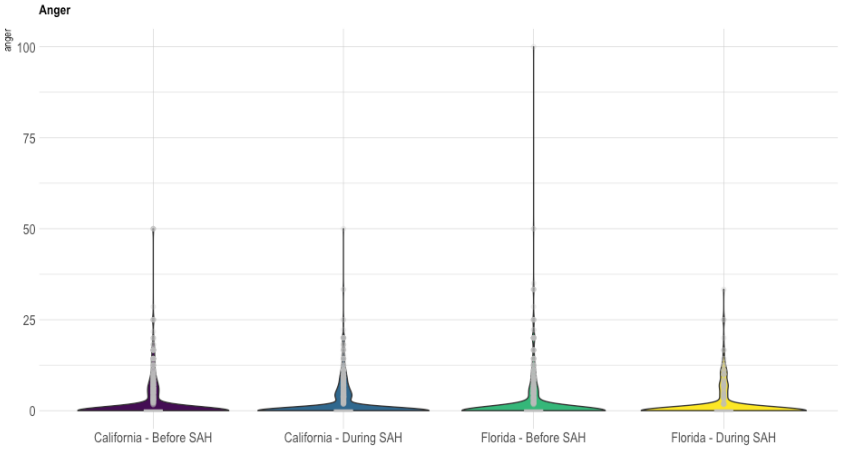

Clout
Clout
100
75
50
25
0
California - Before SAH
California - During SAH
Florida - Before SAH
Florida - During SAH

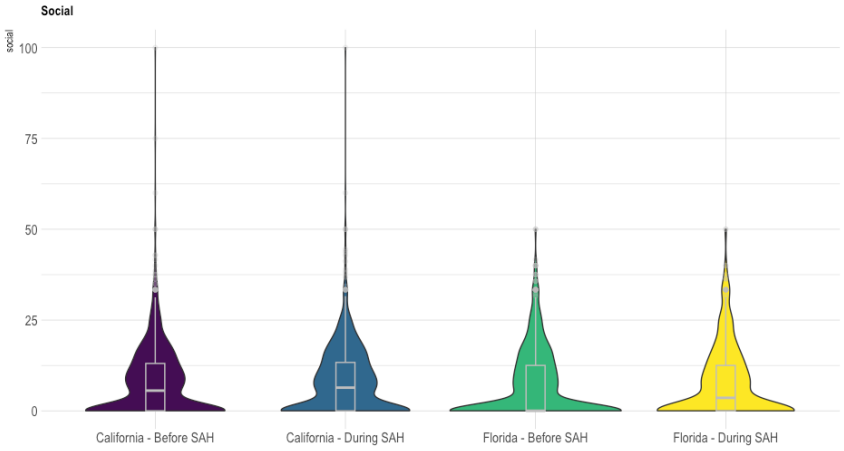

Authentic
Authentic
100
75
50
25
0
California - Before SAH
California - During SAH
Florida - Before SAH
Florida - During SAH


Positive Emotions
posemo
100
75
50
25
0
California - Before SAH
California - During SAH
Florida - Before SAH
Florida - During SAH


Negative Emotions
negemo
100
75
50
25
0
California - Before SAH
California - During SAH
Florida - Before SAH
Florida - During SAH


Personal Pronouns
ppron
50
40
30
20
10
0
California - Before SAH
California - During SAH
Florida - Before SAH
Florida - During SAH


Anger
anger
100
75
50
25
0
California - Before SAH
California - During SAH
Florida - Before SAH
Florida - During SAH


Social
social
100
75
50
25
0
California - Before SAH
California - During SAH
Florida - Before SAH
Florida - During SAH

Appendix C

Q-Q Plot consist of a scatter plot that is performed by dividing the dataset into quantiles (or percentages) and plotting them against the ones that come from the distribution we want to test, which in this case is the normal distribution (University of Virginia Library, 2022). Therefore, if the resulting plot forms a line of 45º, the two samples follow the same distribution. The link between Shapiro-Wilk test and Q-Q plot relies in being goodness-of-fit measures that test the extent to which a sample data fit a normal distribution (King and Eckersley, 2019).

Figure C shows Q-Q Plots performed per variable for the total database. This was done to have an idea of the distribution of each variable, without distinguishing by state or phase in time (before or during SAH). Note that the x-axis refers to the theoretical quantiles of the normal distribution and the y-axis to the ones of the analyzed variable, that is why they might differ among variables as they have different scales.

Figure C: Q-Q Plots per each dependent variable