École Polytechnique Fédérale de Lausanne

BodyTweets: Foundations for Social Media Analysis of Body-Based Language

by Daniel Smarda

Master Thesis

Approved by the Examining Committee:

Robert West
Thesis Advisor

Navid Rekab-saz
Johannes Kepler University Linz, Austria
External Expert

EPFL IC IINFCOM DLAB
INN 310 (Bâtiment INN)
Station 14
CH-1015 Lausanne

August 20, 2021

For Grandpa John, his wife, his five daughters, their children, and all who have found their way into the perpetual love of this family.

Acknowledgments

The completion of this project and the accumulation of research skills required to complete it would not have been possible without the tactful guidance of my supervisor, Navid Rekab-saz. It takes one kind of teacher to "give someone a fish" when they encounter an obstacle and another to guide them on "how to fish", to speak. For the gentle, consistent encouragement of the latter, I am extremely grateful.

This work is an accumulation of skills, opinions, and perspectives that extend well outside the technical experiences of this Masters program. I thank the London mindfulness community, who taught me how to take a break when needed, and Ms. Cathryn McNamara, whose teachings remind me that the context is always as important as the content.

Thank you to my new friends in Lausanne that welcomed a foreigner with open arms. In particular, Harsh and Matthieu: during late nights and difficult exams, thank you for bringing joy to writing even more lines of code and decomposing increasingly abstract renewal theory during even the most stressful of projects. And Jonathan: thank you for always encouraging me to look further and aim higher for the potential impact of our work.

For opening up their home to me as a warm, loving place for me to study, to rest, to enjoy fondue, and to laugh at some kittens, I cannot express enough appreciation to Nicolas, Severine, Louis, and Mathilde Jourdain. Thank you making studying a Masters degree an experience in life as much as in career.

To my parents and brother, for whom these words are not new, thank you for riding the waves that comes with supporting a son's and sibling's move abroad.

And finally, to Nicole: thank you for being a friend, partner, and companion in every aspect of life that I could not possibly begin to name.

August 20, 2021 Daniel Smarda

Abstract

The language we use to talk about our bodies has direct links to our self esteem, which, in turn, has dramatic effects on measurable health outcomes. Psychology and medicine-grounded studies studying this effect are case-based, and computational social science studies either focus on visual media (Instagram, Facebook), or study online harassment as a whole. We offer BodyTweets, a methodology to create and analyze a geolocated Twitter dataset specific to the body-based language domain using Twitter's free Full Academic Search API. Following the methodology, we show that the resulting dataset is representative of both the numerous Twitter epochs over the last decade and the geography of the United States at the state level. We then analyze a dataset from this methodology at the nation and state granularity levels using the relatively nascent DeepMoji emotion analysis tool. Before discussing the limitations of the study, we find that between joy, surprise, fear, anger, sadness, and guilt, the most prevalent emotion is joy, followed by sadness and disgust, and that the distributions of emotions are similar across the states.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Understanding the way we talk about bodies is crucial. In the news, from government recommendations, we hear stories of insidious messages that, somehow or another, find their way into the brains of ourselves, our peers, our friends. These consequences of body shaming are devastating and, increasingly, documented. Beyond personal experience alone, it is clear that misdirected words can and do cause real harm to the psyche of ourselves and society.

The problem has been approached from a variety of angles. Psychologists have studied collegiate populations to see what messages are received, how, and how quickly. Philosophers have ideated on the nature of how we form the concept of the "self" (are we our thoughts? Our actions? Our decisions?). Linguists (and, increasingly, computer scientists) have attempted to empirically investigate how these messages find their way into the field of our attentions' consumption.

With so many avenues with which to learn about the nature of this body-shaming (individual and interpersonal) phenomena, it is worth considering which route can most effectively, genuinely analyze the most crucial aspects of it. Considering all of the above, under the umbrella of quantitative analysis, the following are missing from the empirical literature on this topic. First, a dataset with which we can consider the natural-language phenomena of everyday conversations. Second, an understanding at the nation level (and global, though that is outside the scope of a single Masters student) of the dynamics of body shaming. And third, an understanding of the causes, or at least contributing factors, of the propagation of negative body-based language.

While the third goal is alluded to, rigorous analysis (e.g., causational studies) of particular aspects of this issue ultimately warrants individual attention. However, a quality dataset and preliminary geographic analysis is well within the scope of this project. The structure of this thesis is therefore as follows. In Chapter 2 we discuss in more detail the relevant background research in this project. In Chapter 3 we take great care to translate the goals of this study to a tractable scale, detailing a specific methodology to collect a representative dataset and analyze the tweets for emotional content. In Chapter 4 we characterize the dataset using descriptive statistics by geography and body-related word. In Chapter 5, we use the DeemMoji tool to analyze the emotional content of the collected data at the nation and state granularity levels. Finally, we discuss key results, limitations, and directions for future study in Chapter 6.

Chapter 2

Background and Related Work

2.1 Psychology and Medical Research

Empirical evidence of the devastating nature of negative body image is abundant. In 2006, Jennifer Dyl and colleagues [11] studied medical symptoms in 208 adolescents and found clear correlations between body dysmphoric disorder (BDD) and measures of anxiety and depression. Similarly, in 2008, Yazdani and colleagues found that “body image defects caused by obesity could lie in negative psychological well-being in all aspects” in a sample of 124 obese patients [39]. These and other medical studies are small-scale and often limited to patients with other previous diagnoses, but give us numerical evidence of severe detrimental health effects of having a negative body image.

In the most recent two decades, social and visual media have contributed to these phenomena. Ross Krawczyk and Kevin Thompson studied both women and men (contrary to the common belief that issues of body image seldom have significant impact on men) and found that participants of both genders experienced a higher rate of body dissatisfaction after viewing images of sexually objectified women [18]. Similarly, in studies of adolescent girls, Tiggemann and colleagues have found quantitative links between internet exposure and body shame (via self-reported indicators of anxiety, body dissatisfaction, and self-objectification) [33]

In particular, *visual* social media sites and their impact on the mental health of users is comprehensively documented. In 2016, Lewallen and colleagues studied 118 women in the

United States and found that individuals who follow more fitness-related boards on Pinterest "are more likely to report intentions to engage in extreme weight-loss behaviors" [20]. Using a similar methodology to Tiggeman (2015), Meier and colleagues studied 103 adolescent women and found an increase in "weight dissatisfaction, drive for thinness, thin ideal internalization, and self-objectification" as the overall girls' appearance exposure increased. Significantly, these increases in unhealthy mental behaviors correlated specifically with increases in the use of Facebook's photo features – not with Facebook use overall.

Despite this last finding, a limited but emerging body of research is beginning to demonstrate the importance of linguistic text features, both spoken and written, on our perceptions of self. In the most directly-relevant experiment to this current study, Kaaren Watts and colleagues "primed" participants (undergraduate psychology students) by exposing them to 8 positive and 8 negative body-related nouns (*hips*, *stomach*, etc.) before exposing them to a target adjective (*beautiful*, *magnificent*), and measuring the response times to the target words [38]. Using a 2 x 2 x 2 factorial design to simultaneously measure the relationship between response times and self-reported body image metrics, the researchers found a clear link (lower response time between congruent pairs vs. incongruent pairs) between mental schematics of body parts and emotionally-charged adjective words. Other experiments, particularly in neuroscience, are reaching the same conclusions. The volume of medical content is limited here, but more details can be found papers by Shirao [29], which uses functional magnetic resonance imaging (fMRIs) to document changes in brain biology in response to exposure to negative body-image-related words, and Unterhalter and colleagues [35], in which 20 undergraduate women showed "a general memory bias for weight information *independent* of their personal preference of being thinner".

2.2 Technological Advancements

So, it is clear that both men and women show negative mental, physical, and neurophysiological responses to particular types of body-related words. While this is certainly alarming, it is not wholly counterintuitive, and in the last two decades, informatics-grounded research began to analyze cyberbullying as a health threat. From a theoretical perspective, Vandebosch and colleagues [36] interviewed adolescents to extract the common qualities of cyberbullying, finding that the most detrimental cases are characterized by a) a power differential between a "perpetrator" and "victim", and b) an element of repetition, either of an online nature, or an offline nature that compliments the online aggression. In 2012, public policy professor Colette Langos [19] recognized the evolving nature of online cyberbullying, differentiating between "direct" cyberbullying (as described by Vandebosch) and "indirect" cyberbullying, which occurs more on public online platforms (Facebook news feed, Twitter feed) than private, message-based spaces.

To attempt to counter violent language on these sites, several studies have leveraged machine learning tools to detect online hate speech. Framing the task as a classification task, Al-garadi and colleagues combined network features, activity features (e.g. likes and comments), user features (age, gender), and content features (vulgarity, point of view), and fed them into classical machine learning algorithms: Naïve Bayes, Support Vector machines, Random Forest, and k-Nearest-Neighbor [15]. Using 4 different feature selection experiments, the authors obtained F-scores well above 90%. Research into feature engineering and method exploration has continued – in 2020, Muneer and colleagues expanded Al-garadi's work to include Logistic Regression, Stochastic Gradient Descent, AdaBoost, and Light Gradient Boosting Machines [22]. Regression methods have also been explored [9], though they are significantly less common.

In addition to classical machine learning representations, the abundance of attention lent to this area in recent years has led to creative feature engineering and model combination methods. In 2015, Mangaonkar and colleagues [21] introduced a collaborative approach comprised of a network of autonomous detection nodes, each of which use different machine learning algorithms and each of which may be trained on a different dataset. In a test on an unbalanced dataset of 177 and 1,163 bullying and non-bullying tweets respectively, they demonstrated a time speedup in computation with comparable results compared to individual classifiers alone. Similarly, an example of using network-based features can be found in the XBully tool, which learns network activity features such as network hotspots and neighborhood relationships [7]. Then, it applies additional feature embedding techniques to account for a diversity of social network characteristics (e.g., sparsity), resulting in state of the art performance in almost all Instagram and Vine experiments.

Even more creatively, cyberbullying detection studies in the most recent few years have started to leverage emotional and sentiment-based features in their attempts to improve the accuracy of these systems. In 2017, Balakrishnan and colleagues extracted the personality traits of users by passing Twitter user analytics data and tweet linguistic features to the (now deprecated) IBM Watson Personality Insights API [1]. The result was a consistent F-measure increase of at least 0.1 when compared to baseline models that did not incorporate any of the personality traits. Dani and colleagues [8] combined content (one-hot encoded vocabulary) methods with sentiment scores and similarly found an increase in F-measure of 0.5 when compared to content classification alone.

When considering these bodies of research independently and together, we notice a niche in this domain that needs to be filled. Research from the medical community demonstrates the potential for relationships between body-based language and health outcomes, but only does so in case settings with relatively small (N<1000) sample populations. The technological community provides practical ideas for stopping the spread of cyberbullying or hate speech in general, and demonstrative insights from the visual social media domains, but similar studies in the linguistic domain are currently missing. In the next section, we detail a methodology to curate a dataset of tweets that is as representative of the total Twitter population as possible using only the free level of Twitter's API services. This methodology and the accompanying dataset can be used as foundations for the study of body-related language at the social network scale.

Chapter 3

Methodologies

This chapter is divided into three content-distinct subsections: sample construction, emotion analysis methodologies, and the geolocation algorithm.

3.1 Constructing a Representative Sample

Due to the unique domain-specific nature of this project, a novel dataset was necessary which needed to satisfy the following criteria:

- As representative as possible of the United States Twitter user population
- Geographically diverse, to allow for meaningful analysis of demographic attributes
- Robust to the time-fluctuating nature of Twitter users and trending topics

This problem was particularly difficult when introducing a low-resource constraint, as the Twitter API only offers limited functionality. The Full Academic Search track [34] (at time of writing) offers the ability to search Twitter's entire database by keyword, date, and language, but only starting from a single end date and returning results chronologically backwards.

Discussion of geographic diversity is delegated to Section 3.3, but temporal diversity is discussed here. The Twitter API's Full Academic Search Track does not offer time-interval granularity. Instead, only single times and a maximum desired n results are specified. All relevant results are returned chronologically backward in paginated form until the desired number of results is reached. For this reason, it is vital to select a diverse set of times from which the desired tweets can be obtained.

In order to be representative of the Twitter population as a whole, the temporal queries should satisfy the following conditions:

Robust to year-scale time fluctuations: This mans it is robust to the cultural changes that occur over the course of a decade, both online (considering changes in Twitter word choice) and offline (for example, the presidential election of Donald Trump).

Robust to seasonal changes Consideration of fluctations in seasonal uses of langugage is important, as summer discussions of "beach bod" may be different than winter discussions of more heavily-clothed bodies.

Robust to time-of-day and day-of-week time scale fluctuations Empirical research has demonstrated shifts in mood based on the day of week in the United States [32]. It is therefore important that the dataset contain data from multiple days throughout the week.

Holiday Considerations Depending on the scope of any particular desired study in the area, the trending disposition of Twitter and should be considered especially in the context of holidays. In the dataset collected for this project, for example, we take great care to avoid major United States holidays including Christmas, Thanksgiving, and Valentine's day, with all queries at least 2 weeks away from any major holiday.

3.2 Emotion Analysis Methodologies

A huge number of tools are available for sentiment analysis [26], [40], broken down into lexicon-based approaches and machine-learning based approaches. Of the tools available, only a small number of them are:

- Free or low cost
- Well-evaluated
- Demonstrably effective *across domain areas* and *without any training data*

Of the tools available satisfying the above criteria, the most widely-used and verified approaches were Vader Sentiment Analysis [17], the Linguistic Inquiry and Word Count tool (LIWC) [25], and Deepmoji [14], as analyzed in [26]. Compared to Vader, DeepMoji carries the distinct advantage that it classifies texts into a variety of categories. Compared to LIWC, DeepMoji provides the best performance and domain-specificity to Twitter. DeepMoji is particularly adept at recognizing slang and sarcasm [14]. The package includes a tokenizer, so that raw feeds can be fed in and tokenized before being processed.

DeepMoji provides a PyTorch implementation which assigns percentage confidence scores to each of 64 categories in the researchers' selected target set. Figure 3.1 shows an example of this output.

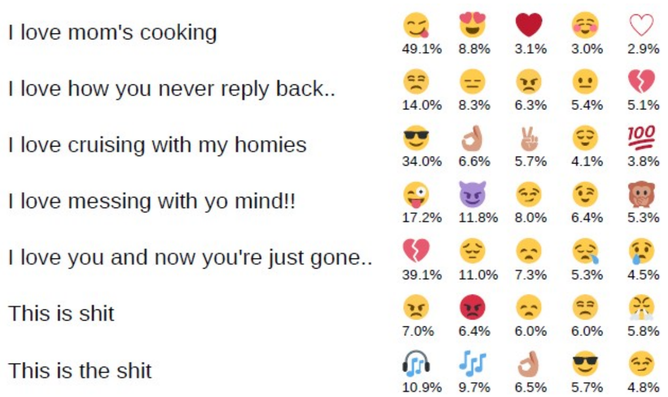

Figure 3.1: Example output for tweets from DeepMoji.

To translate these emojis into digestible emotions aligned with psychology also requires an intentional choice of emotion model. While several models exist, the most widely used (e.g., [37]) and, arguably, simplest, is that of Ekman's 6 "Basic" emotions [12]. This model also carries the interpretive advantage over other emotion models of having comparatively distinct, non-overlapping emotion categories [1].

3.3 Geolocation

Tweet geolocation is yet another area of active research relevant to the present study. The easiest method of geolocating tweets is using the Twitter API's built-in 'geo' objects. The main issue with this method is that only approximately 1% of tweets are geotagged [31]. While a multitude of other algorithms have been proposed and assessed (see for example, [28]), many are quite involved to implement (e.g., [16]), and very few have been consolidated or tested into reliable software tools.

To find a middle ground between complexity and availability, we choose to implement an algorithm proposed and validated by [3] and used again in [4]. The original algorithm in [3] was used to identify tweets at the city-level. We modify this algorithm to focus on state-level geolocation and detail the algorithm here.

We start with the user's 'location' string, a free-text optional entry on Twitter profiles. Starting with the raw string, we first match unigrams and bigrams extracted from the location string [2] to exact names and two-letter United States Postal Code abbreviations of all 50 US states. 980 Tweets matched to Puerto Rico were excluded from the final dataset because the United States American Community Survey, the gold-standard for most demographic analysis, does not contain information about Puerto Rico. Washington, D.C. was matched to "District of Columbia".

Once all tweets with state strings were matched to states, the user strings were parsed for city matches based on city strings in the United States American Community Survey [5]. This was a much more difficult challenge due to the generic and sometimes confusing names of cities (e.g., Kansas City in the state of Missouri, not Kansas). To overcome this, we created a dataset of potentially-confounding names of global geographical entities (again borrowing in methodology from [4]). This dataset consisted of:

- All Canadian Metropolitan Areas and Census Agglomerations [6]

[1] Ekman later added a 7th emotion, "contempt", but this is so far less widely-used in the empirical emotion analysis literature

[2] Tokenization here is completed using the NLTK TweetTokenizer() tool: https://www.nltk.org/api/nltk.tokenize.html.

- All United Kingdom Cities, Regions, Authorities, Wards, and Electoral Division [23]
- All names of nations recognized by the United Nations [10]
- All global metropolitan areas with more 100,000 people [24]

When cities were matched to multiple locations (e.g. "London" being matched to the locations *London, United Kingdom* and *London, Ohio, United States*), the location of the tweet was assumed to be in the most populous area.

Figure 3.2 illustrates this algorithm visually. A random sample of 100 tweets returned an accuracy of 93%.

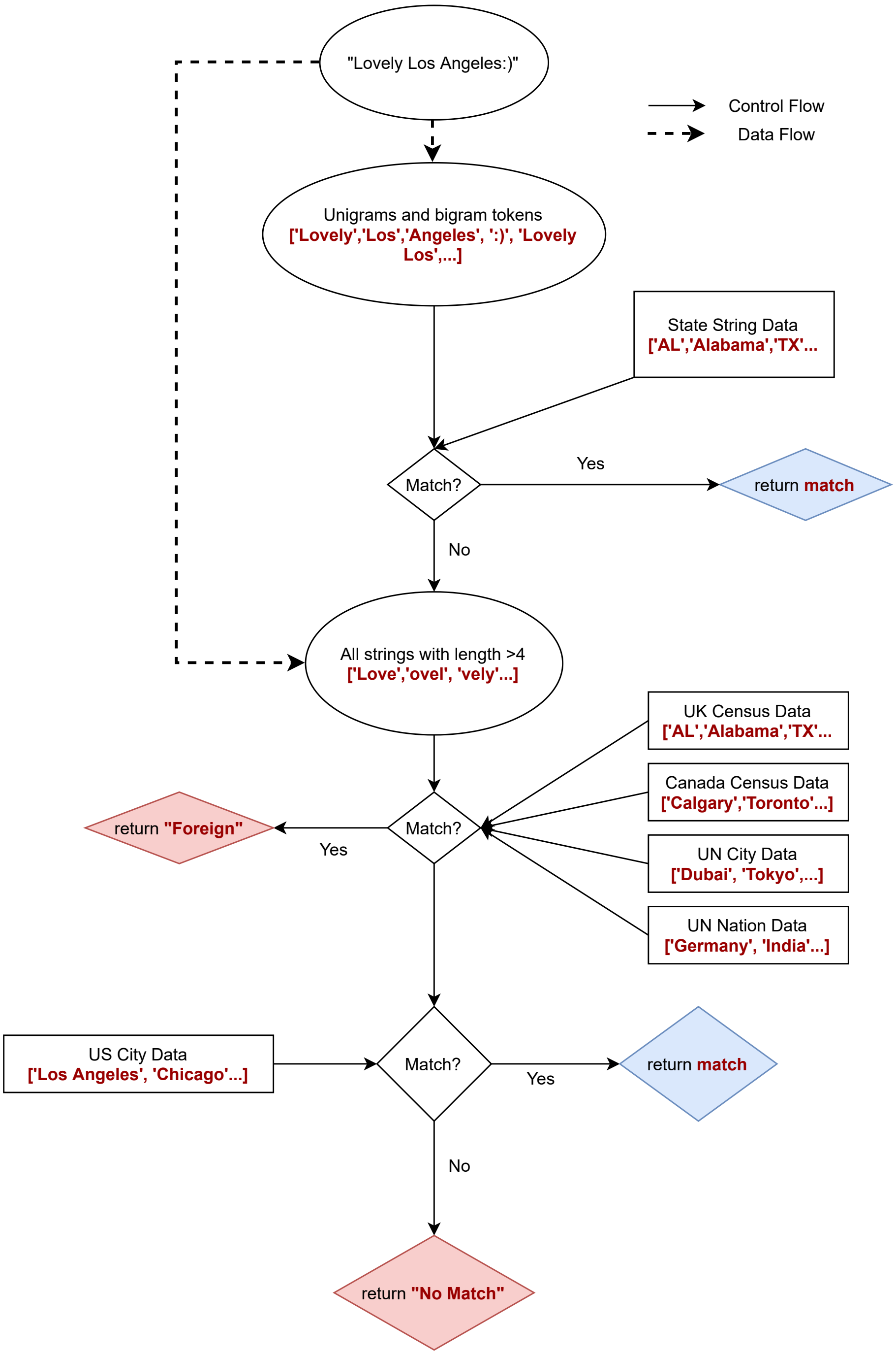

Figure 3.2: State Geolocation Algorithm.

Chapter 4

National and State-Disagreggated Dataset

Using the methodologies outlined in Chapter 3, we collected a representative dataset of the United States Twitter population using body-based language keyword and time queries. In the following sections we detail the choices used for this specific dataset and characterize it using standard frequency-based descriptive statistics.[1].

4.1 Time Intervals

Due to the configuration of the Twitter API, tweets were collected in 10 batches of 200,000 tweets per batch. Because the limit for the queries was placed on the number of tweets rather than the time range, the time interval of each batch varied from 99 minutes to 8 hours, 40 minutes. The queries spanned late autumn (early December), winter (late January and early March), and spring (early June). Unfortunately, summer was omitted by error.

Data from all days of the week except for Wednesday were collected. This includes Monday morning, Friday evening, and weekends as studied in [32] and [13]. As explained in Chapter 3, data were collected from every other year starting in 2013 until 2021, ensuring equal coverage of the entire decade.

[1]All code used for this project can be found at https://github.com/danieljsmarda/BodyTweets

Table 4.1 shows the details of the time aspects of the queries. All times are in Central Standard Time (CST), which is the time zone of Texas and Illinois. The corresponding times for the West and East Coast can be found by subtracting 2 hours and adding 1 hour, respectively.

Table 4.1: Query Time Information; all times in CST (Chicago).

Day of Week	Time	Date	Year
Monday	10:21 - 14:00	01.28	2013
Saturday-Sunday	22:40 - 07:00	12.05-06	2015
Thursday	14:21 - 21:00	06.01	2017
Tuesday	16:54 - 21:00	10.15	2019
Sunday-Monday	02:20 - 07:00	02.28	2021
Saturday	10:11 - 14:00	01.26	2013
Thursday-Friday	22:20 - 07:00	12.03-04	2015
Tuesday	14:28 - 21:00	05.30	2017
Sunday	16:29 - 21:00	10.13	2019
Thursday-Friday	03:31 - 07:00	02.26	2021

In total, 2,000,000 tweets were collected based on keyword and time interval parameters. After discarding tweets for which the user had no location string, 414,560 unique tweets remained. In the case of tweets from users that were geolocated to more than one state, one row was added to the dataset for each geolocated state, resulting in additional 7,727 rows for a total dataset of 422,287 tweets. The tweets were published by 317,566 unique authors with an average of 1.33 tweets per user.

4.2 Tweet Counts

Figure 4.1 shows the number of tweets allotted to each state. California, Texas, and New York, the top 3 most populous states in the United States, also contribute the most heavily to the dataset, with the least populous 3 states (North Dakota, Vermont, and Wyoming) contributing the least. In general, the distribution very closely follows the general population distribution of the United States.

This population-heavy skew has both advantages and disadvantages for data analysis. National statistics, such as those discussed in Chapter 5, may be biased towards more populated areas. However, policy-based analyses (e.g. combination with demographic data as demonstrated briefly in Section 5.2) would benefit from equal-weight assignment.

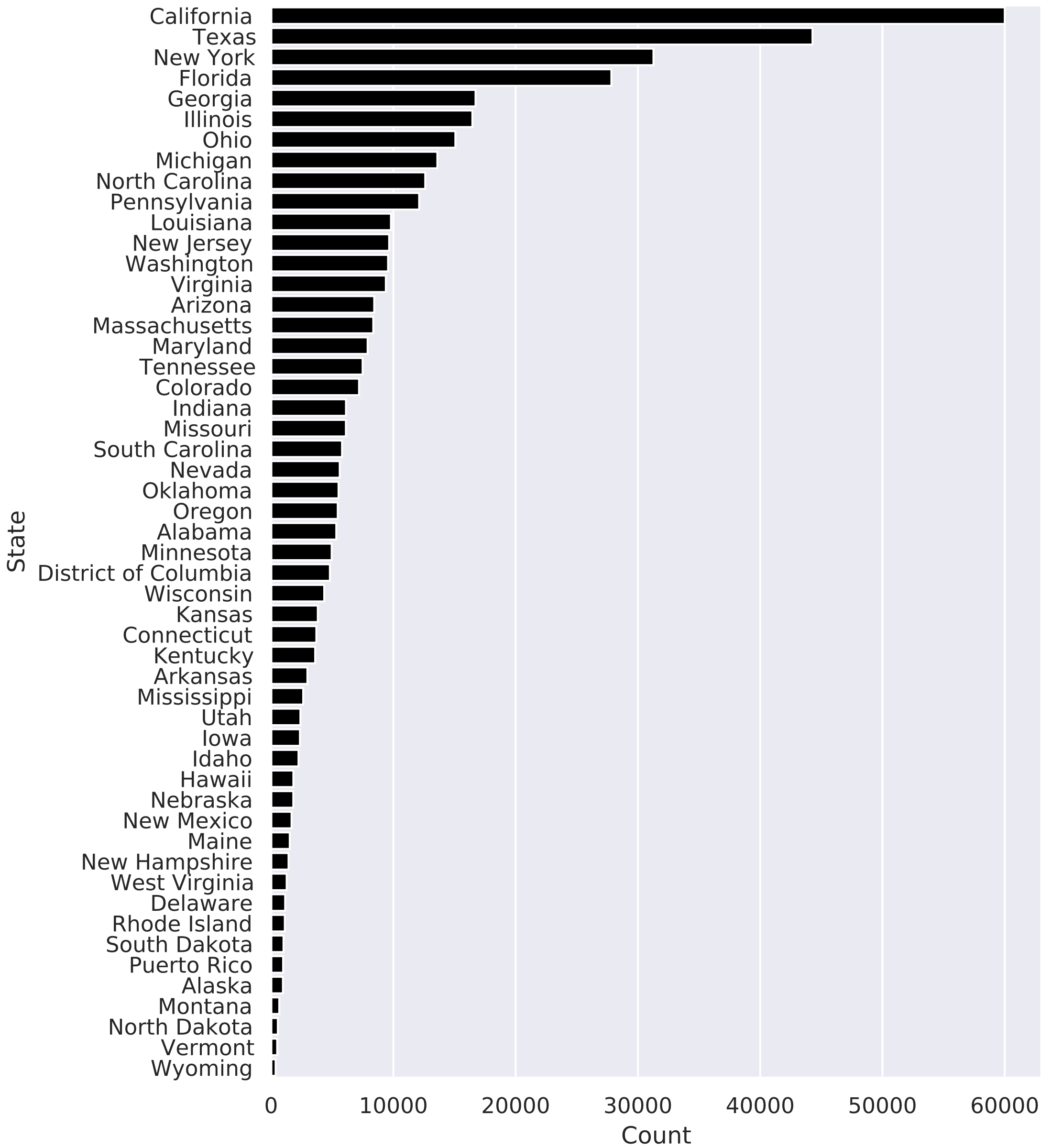

Figure 4.1: Distribution of Counts of Tweets across the US States.

Figure 4.2 shows the cumulative distributions more clearly. The top 40% of the data is covered by the most tweet-dense states, while the 25 least-tweet-dense states contribute only 20% of the tweets to the overall dataset. Numerical information for these distributions can be found in Table 4.2.

Table 4.2: State Tweet Count Distribution Data

	Tweet Count Distribution
count	52.00
mean	8120.90
std	11044.49
min	369.00
25%	1778.50
50%	5142.00
75%	9428.75
max	59999.00

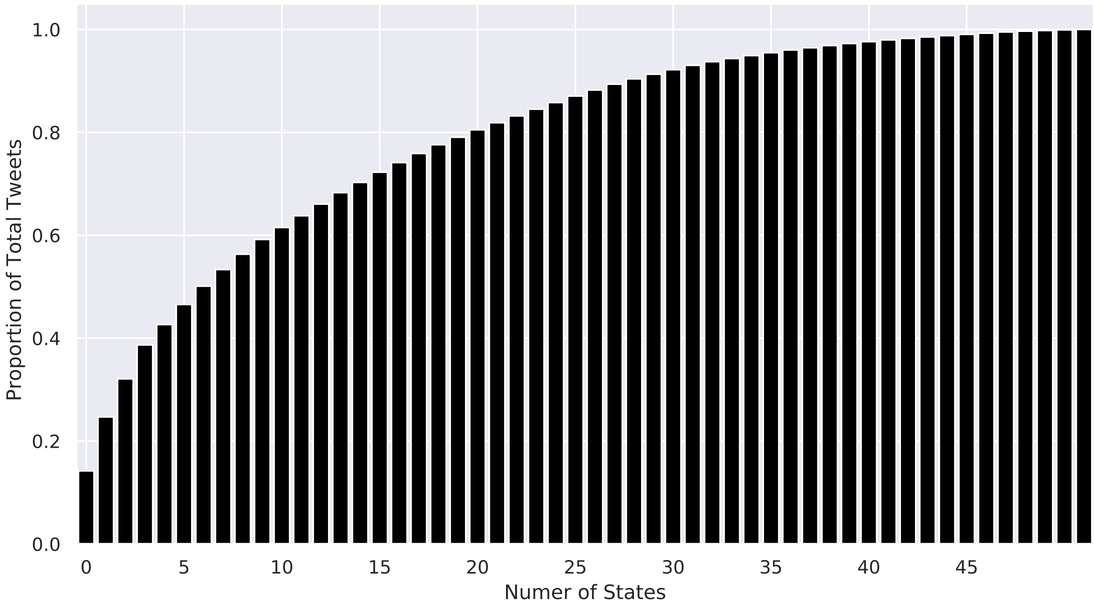

Figure 4.2: Cumulative Coverage of Total Tweets by State

4.3 Token Matches

4.3.1 Token Choice

The data was collected based on 24 body-related nouns and phrases. The choice of nouns is not new [38], but a unique list was generated for this project. The primary list consisted of 19 body-related nouns that, as the discretion of the researcher, satisfied a number of properties relevant for this study. The words were commonly-used anatomical body nouns. Intentionally, words belonging to any of the following groups were omitted:

1. Words and slang-words that had a high probability of being emotionally charged ("ass", "tits"). This ensured a higher probability that the word token was being used to refer to an actual body part and not another unrelated topic (e.g., "That's kick-ass!").

2. Body part names used primarily in the medical domain. This increased the likelihood of returning tweets representative of a greater proportion of the public rather than solely members of the scientific community.

3. Words related to sexual anatomy ("penis", "vagina", "butt", "dick", etc.). Determining when these tokens were being used to refer to body parts and not for an unrelated purpose was deemed to be too difficult a task, similar to Item 1.

The final list was comprised of intentionally-chosen words that, in the context of language in general (not only Twitter) have a high likelihood of being used to identify with the self (see, for example, [2]), as assigned either autonomously (by the speaker to themself, e.g. "my eyes") or heteronomously (by the speaker to someone else, e.g. "your eyes" or "his eyes"). This meant that where appropriate the plural version of a token was chosen but the singular excluded ("teeth" included, but not "tooth", "lip" but not "lips"). [2] Finally, the goal was to create a list of commonly-used nouns that were context-independently neutral, such that the emotion analysis would reflect the emotion values of the context of the words rather than of the words themselves, and so words with multiple possible contexts (e.g., "body") were favored over words with singular contexts (e.g. "eyebrows"). Put alternatively, the list favored words for which the hypothesized outcome would be less clear.

Tokens were all searched for individually, with the exception of "face", for which pronouns were appended due to too many ambiguous uses (e.g., "Time to face my fear."). The final list of tokens is shown in Table 4.3.

Table 4.3: List of words matched for finding body-related tweets

Body Area	Words
Head/Face	eyes, ears, mouth, nose, lips, teeth, neck
Torso	chest, breast, breasts, belly, torso, stomach, waist, hips
Limbs	legs, arms
General	body, skin
Phrases	"my face", "your face", "his face", "her face", "their face"

[2]Studies in each of these areas, while important, would require linguistic and philosophical research related to sense of self that is outside the scope of this general overview descriptive study.

4.3.2 Token Analysis

While the emphasis for the project was creating a geographically and culturally-representative dataset, I also analyzed the dataset by token. Figure 4.3 shows the distribution of counts of Tweets that contained each token. The words "body" and "eyes" were the most commonly-occurring tokens, each occurring in nearly 15% of all tweets, followed by "mouth", "skin", "my face", "legs", and "arms". In total, 13 out of the 24 tokens occurred in more than 10,000 tweets, and 18 out of the 24 tokens occurred in more than 5,000 tweets.

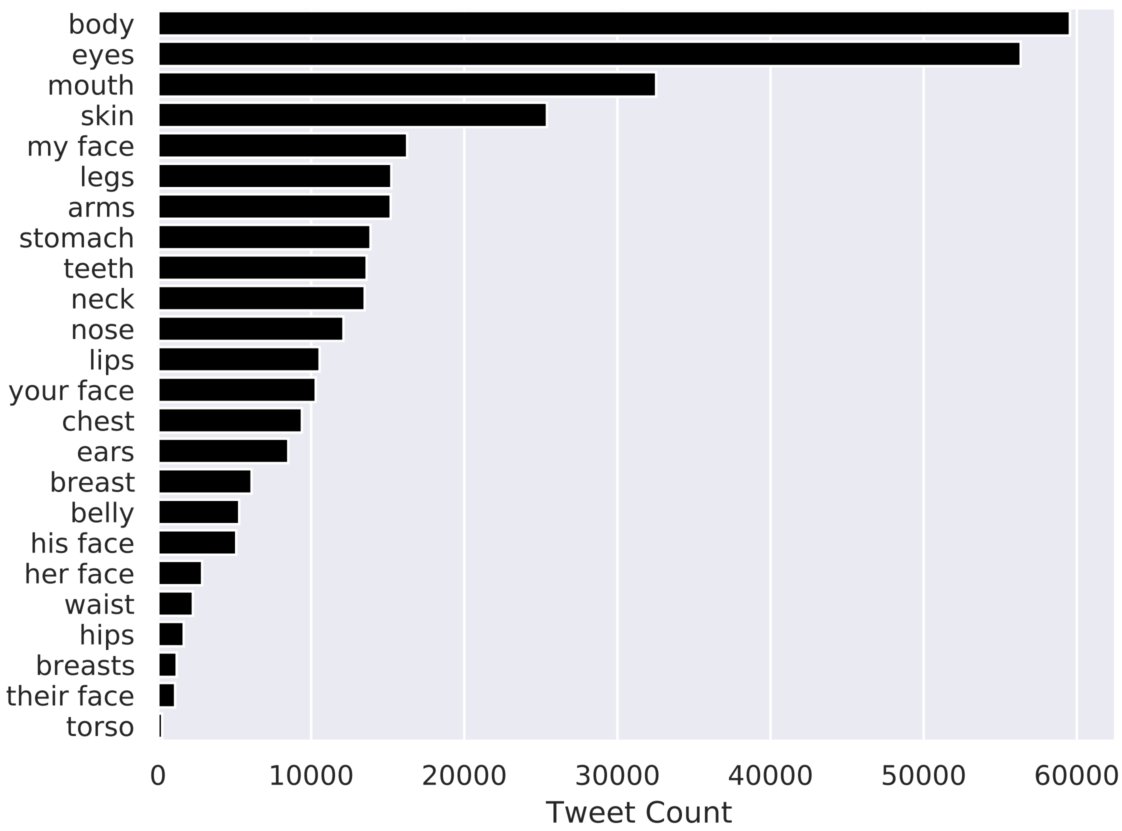

Figure 4.3: Distribution of Number of Tweets Containing Each Token

Chapter 5

Emotion Analysis

Finally, using the dataset accrued in Chapter 4, we now present an introductory study on the emotional content of Twitter's body-based language corpus. We analyze the data at first the nation scale, then at the state level.

As discussed in Section 3.2, the Deepmoji algorithm returns percentage scores for 64 emojis. To convert these scores into commonly-used terms, each emoji was manually mapped to one of Ekman's "Basic Emotions". For each tweet, the 64 scores were grouped by emotion, resulting in a 6-category emotion distribution of each tweet.

While Ekman asserts that these emotions are distinct, the way that emotions are categorized in other studies vary by study and tool. LIWC, for example, has two higher-level categories: "Positive Emotion" and "Negative Emotion", with the "Negative Emotion" category further divided into "Anxiety", "Anger", and "Sadness", while VADER sentiment ranks its scores on a scale from -1 (negative sentiment) to +1 (positive sentiment).

As discussed in Subsection 4.3.1, the query words for this dataset were intentionally chosen to be neutral nouns with multiple contexts. As such, it was unclear prior to the study which contexts of each word would be dominant. At the state granularity level, we hypothesized that variation between states would exist, though it was unclear what the nature of that variation would be.

5.1 Nation-Level Analysis

Figure 5.2 shows the scores of all 422,287 tweets aggregated to the national level, and Table 5.1 shows these numerical values in tabular form. In this figure, emotions are demonstrated as unique from one another. By far, joy is the most represented emotion at 44.8%. Sadness and disgust are the next-most represented, at 17.8% and 15.5% respectively, and "surprise", "fear", and "anger" are all represented at less than 10% each. By far, "joy" is the most represented emotion.

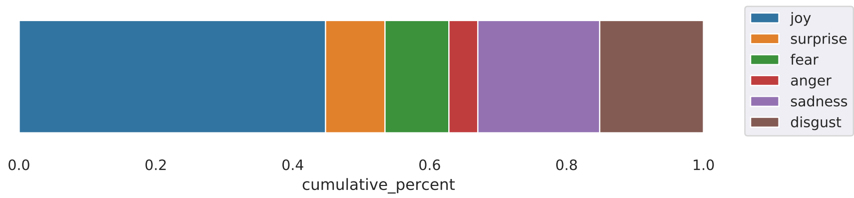

Figure 5.1: National Emotion Distribution - Distinct Model

Table 5.1: Numerical National Emotion Values

Emotion	National Value
joy	0.448
surprise	0.087
fear	0.093
anger	0.042
sadness	0.178
disgust	0.151

To more closely align with other sentiment analysis tools and LIWC, Figure 5.2 shows the same information, though color-coded in roughly positive/negative form. "Joy" and "surprise" were considered as positive emotions – though we recognize that surprise could be a negative emotion – and "fear", "anger", "sadness", and "disgust" as negative emotions. Overall the tweets are scored slightly more negative (53.5%) than positive (46.5% positive).

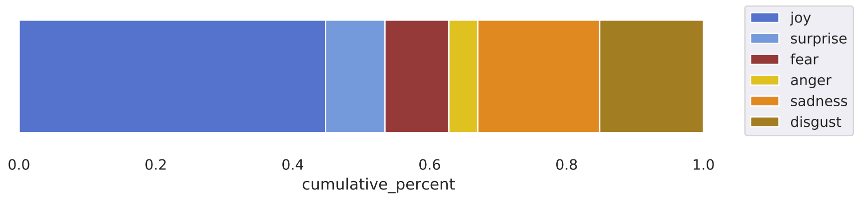

Figure 5.2: National Emotion Distribution - Positive/Negative Model

Measuring Spread Measuring the spread of the emotion scores becomes less defined than $\mathbb{R}$-spanning datasets primarily because in this case, the proportion distributions are dependent on one another. Nonetheless, understanding the nature of the spread is important to understanding the semantic meaning of the emotion scores. In particular, outlier spikes in the data could point to linguistic properties indicative of the potential for further study. Figure 5.3 shows the distribution of tweet scores for each emotion.

Understandably, most of the non-joy emotions are skewed towards the negative direction. Importantly, though, we notice that the scores are reasonably equally-distributed (i.e., smooth). In other words, the mean "joy" score of 44.8% is comprised of scores distributed smoothly across $[0, 1]$ and not, for example, due to two clusters of scores around 0 and 0.8.

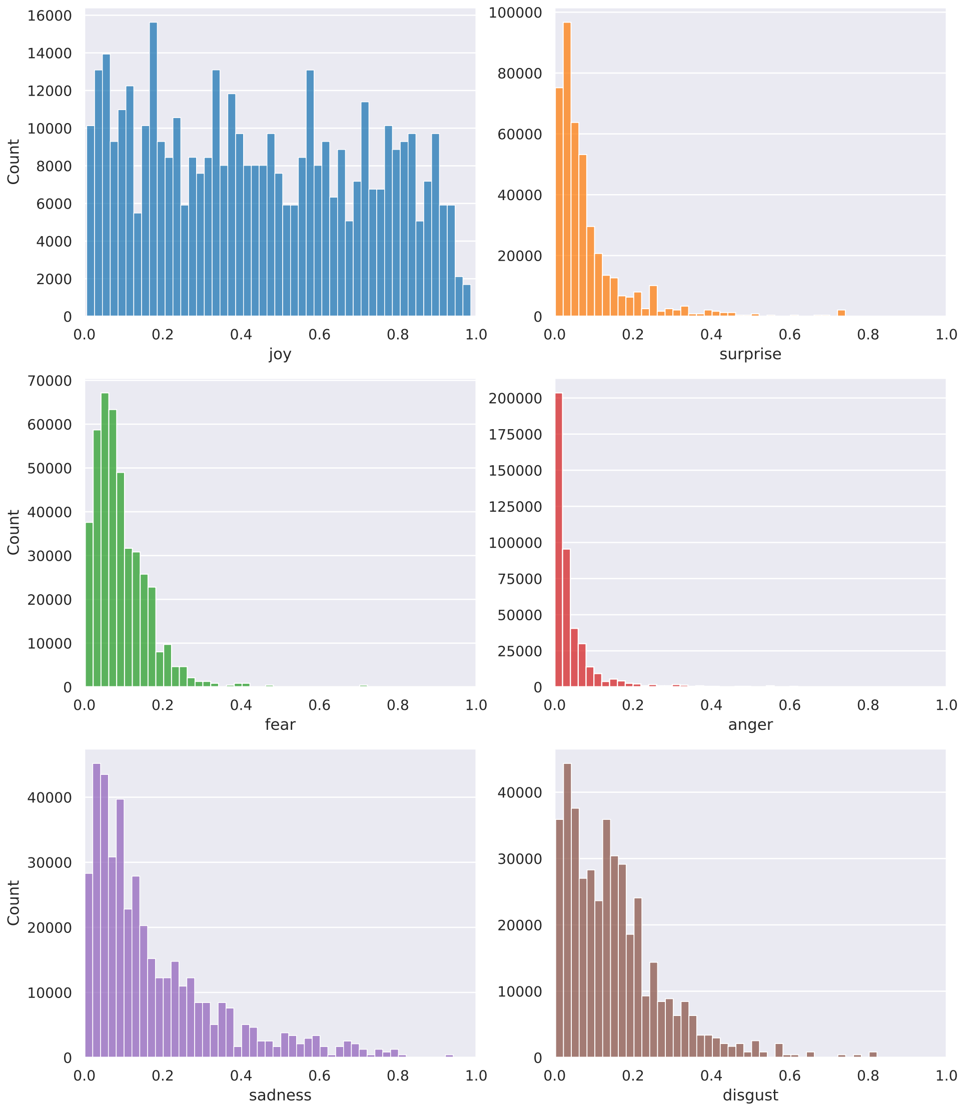

Figure 5.3: Distribution of Counts of Emotion Scores

5.2 State-Level Analysis

Finally, we discuss the distribution of scores across the states. For each emotion, the distribution of the 51 states (48 mainland, Hawaii, Alaska, and District of Columbia) is calculated. Surprisingly, we see that the domain covered by the individual distributions is extremely small, even though the mean scores are distributed across the percentage domain. Specifically, the interquartile-range (IQR) for each of the emotions is less than 0.005.

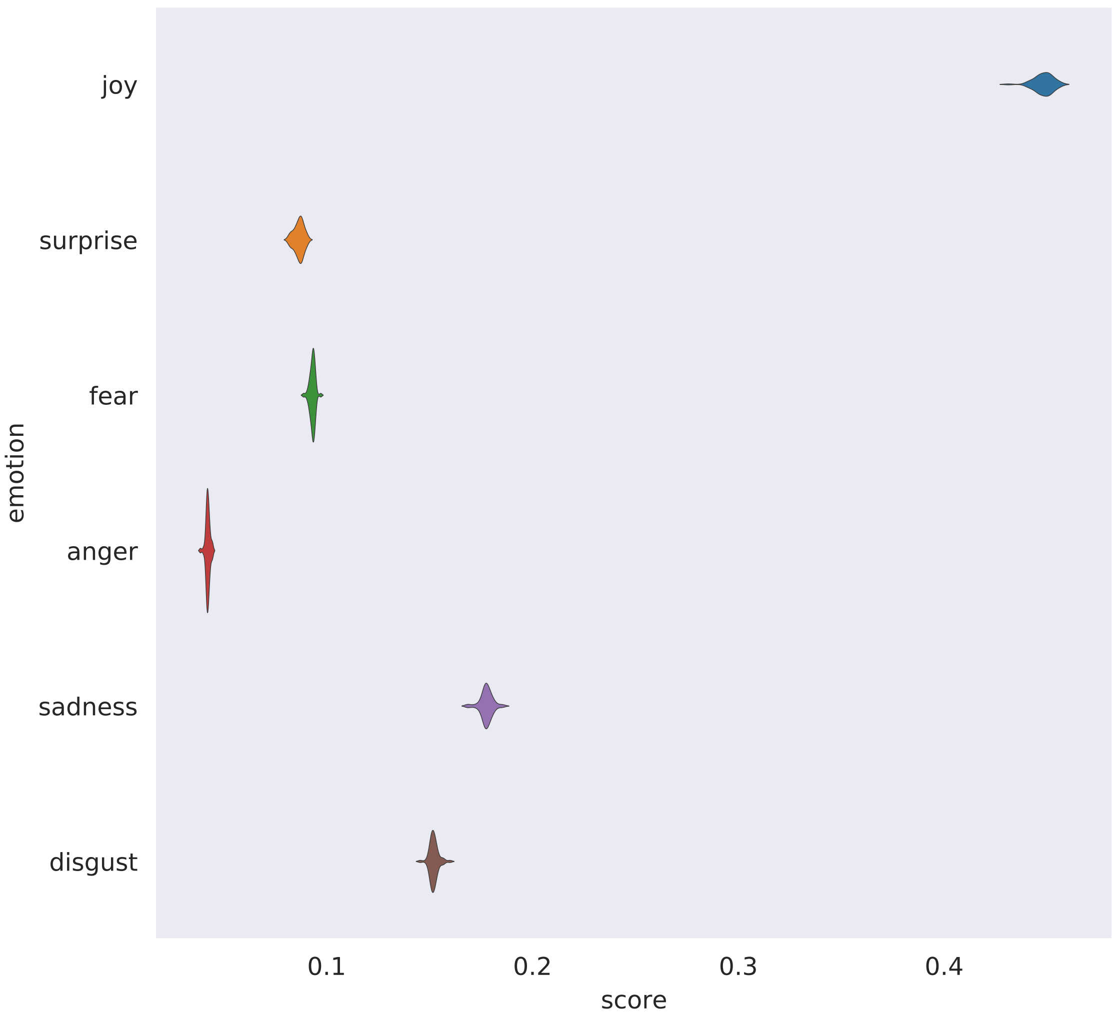

Figure 5.4: Distribution of Emotion Scores over States.

.

Table 5.2: State distribution Numerical Data

	joy	surprise	fear	anger	sadness	disgust
mean	0.448	0.087	0.093	0.042	0.178	0.152
std	0.005	0.002	0.001	0.001	0.003	0.002
min	0.431	0.081	0.089	0.039	0.168	0.146
25%	0.446	0.085	0.092	0.042	0.177	0.151
50%	0.449	0.087	0.093	0.042	0.177	0.152
75%	0.451	0.088	0.094	0.043	0.179	0.153
max	0.456	0.091	0.097	0.045	0.186	0.160

Upon seeing these scores, it becomes less evident whether the variance among scores is sufficient to conduct meaningful demographic analysis. Despite this, we plot five demographic attributes against these scores and see if any hidden trends emerge. The five demographic attributes are taken from the American Community Survey 1-year and 5-year estimates ending in 2019 (for the specific table, see [5]. The variables for each state are:

- The median age of the the state population in years
- The proportion of people in the state that are married
- The proportion of people in the state that have a bachelor's degree or higher
- The proportion of people in the state below the government-determined poverty line
- The population of the state born outside of the country.

Correlation with a single emotion was simpler and more easily-interpretable than correlating with multiple variables. So, we correlate each demographic variable against joy, which we believe gives the most comprehensive characterization of each state in terms of emotional content compared to the other emotions. Via both visual (Figure 5.5) and numerical (Table 5.3) analysis, it is evident that little correlation exists in this dataset.

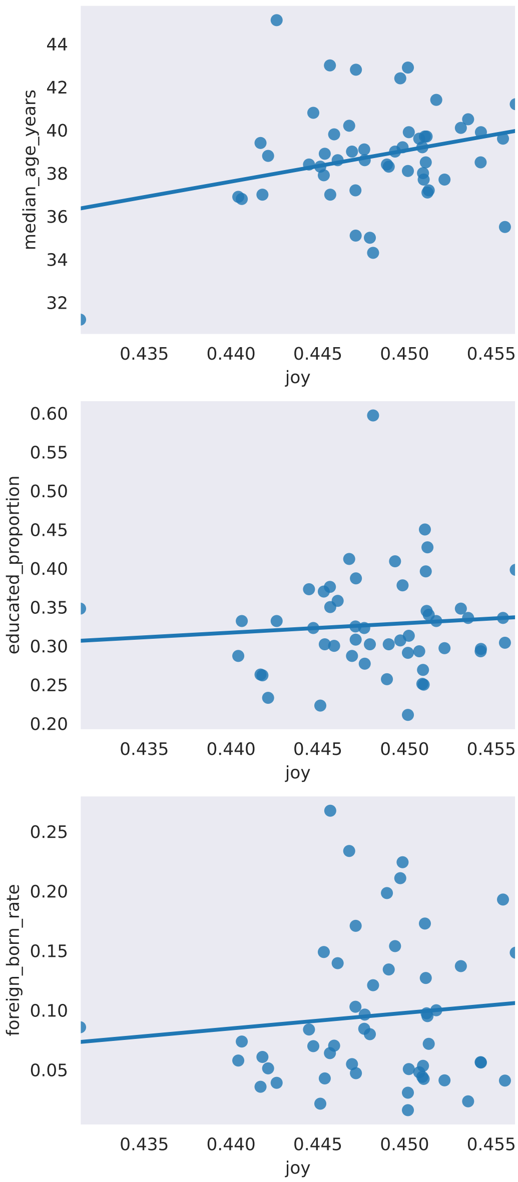

Figure 5.5: Correlations Between Joy Score and Selected Demographic Variables

Table 5.3: Correlation Coefficient values between selected demographic attributes and state emotion scores

	Pearson R	p-value
median_age_years	0.282	0.045
married_proportion	-0.181	0.203
educated_proportion	0.086	0.551
in_poverty_proportion	-0.068	0.634
foreign_born_rate	0.098	0.493

Chapter 6

Conclusion

We conclude by recapitulating the main contributions of this work, clarifying limitations of the study, and outlining the (numerous) directions for future work.

6.1 Contributions

In this project, we identified a discipline-specific gap in Twitter analysis methodologies: that of body-related language. Our main contributions are as follows:

- We provide a methodology for accumulating domain-specific, representative data on this topic. We detail necessary query parameters, offer a state-level geolocation algorithm, and provide a methodology to analyze emotions.
- We collect a dataset of approximately 400,000 tweets distributed across the United States. These tweets are based on carefully-selected keywords. The dataset is representative of all 50 US states and 23 body tokens.
- We analyze the emotion characteristics of the dataset. We show that joy accounts for 44.8% of tweet sentiment nationally, and that all states show similar emotion characteristics.

6.2 Limitations

Despite our best attempts to create a methodology that was representative of the United STates Twitter population as a whole, there are a number of limitations that should be stated explicitly.

Most notably, the geolocation algorithm was only minimally validated. While the self-evaluated 93% is quite high, this was on small sample of only 100 tweets. The validity of the conclusions drawn in Section 5.2 would be greatly improved by an accuracy analysis of a larger batch of tweets.

Also related to rigorous evaluation is the determination of the mappings of emojis to emotions discussed in Section 3.2. This mapping was also determined by the individual researcher. A crowd-sourced validation of which emojis should be assigned which emotion would potentially alter the national and state distributions. Furthermore, the analyses in Chapter 5 rely on a single tool: DeepMoji. Reliance on a single tool for any computational sociological or psychological analysis carries the need for intentional interpretation of the numerical findings, and this is no exception.

The results in Section 3.2 describe the population of *Twitter users* in a given state, *not* necessarily the population of the state as a whole. In particular, this means that the data is likely overrepresentative of urban areas and underrepresentative of rural areas. Linking the demographics of Twitter data to the demographics of the general population is an area of active research in and of itself (see [31] and [30]).

6.3 Future Work

Each of the limitations discussed previously provide room for improvement of the reliability of the dataset in question. Beyond these potential corrections, several other avenues for possible extensions exist.

The ultimate goal of this research is to influence policies that may contribute to negative online body sentiment. Applying the statistical methods used in studies such as [4] to the demographic variables in Section 5.2 would hopefully yield greater insights into these factors.

Because efficacious longitudinal analysis requires advanced understanding of the nature of the way the Twitter population changes over time, time-based variation analysis was excluded from this study, but could yield insights as to what situational factors (seasonal, holidays) contribute to more positive or negative body sentiment.

Finally, the DeepMoji system was chosen because it was the system with the highest demonstrated performance that also had an easy-to-use public API available. However, since it's publication it has been outperformed on certain emotional analysis tasks (see, for example, [27]). Complementing or replacing the results found here with an implementation of one of the more recent algorithms would additionally yield more reliable emotion analysis results.

Bibliography

[1] Vimala Balakrishnan, Shahzaib Khan, Terence Fernandez, and Hamid R. Arabnia. "Cyberbullying detection on twitter using Big Five and Dark Triad features". In: *Personality and Individual Differences* 141 (2019), pp. 252–257. ISSN: 0191-8869. DOI: https://doi.org/10.1016/j.paid.2019.01.024. URL: https://www.sciencedirect.com/science/article/pii/S0191886919300364.

[2] Roy F. Baumeister. "Self-concept, self-esteem, and identity." In: Nelson-Hall series in psychology. Chicago, IL, US: Nelson-Hall Publishers, 1999, pp. 339–375. ISBN: 0-8304-1422-3 (Paperback).

[3] Khandis R. Blake, Brock Bastian, Thomas F. Denson, Pauline Grosjean, and Robert C. Brooks. "Income inequality not gender inequality positively covaries with female sexualization on social media". In: *Proceedings of the National Academy of Sciences* 115.35 (2018), pp. 8722–8727. ISSN: 0027-8424. DOI: 10.1073/pnas.1717959115. eprint: https://www.pnas.org/content/115/35/8722.full.pdf. URL: https://www.pnas.org/content/115/35/8722.

[4] Khandis R. Blake, Siobhan M. O'Dean, James Lian, and Thomas F. Denson. "Misogynistic Tweets Correlate With Violence Against Women". In: *Psychological Science* 32.3 (2021). PMID: 33593204, pp. 315–325. DOI: 10.1177/0956797620968529. eprint: https://doi.org/10.1177/0956797620968529. URL: https://doi.org/10.1177/0956797620968529.

[5] United States Census Bureau. *American Community Survey 2019 Table B01003: Total Population.* 2021. URL: https://data.census.gov/cedsci/table?t=Populations%5C%20and%5C%20People&g=0100000US.160000&y=2019&d=ACS%5C%205-Year%5C%20Estimates%5C%20Detailed%5C%20Tables&tid=ACSDT5Y2019.B01003 (visited on 07/15/2021).

[6] Statistics Canada. *Census Profile, 2016 Census.* 2017. URL: https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/details/page.cfm?Lang=E&Geo1=PR&Code1=01&

Geo2=&Code2=&SearchText=Canada&SearchType=Begins&SearchPR=01&B1=All&TABID=1&type=0 (visited on 07/15/2021).

[7] Lu Cheng, Jundong Li, Yasin N. Silva, Deborah L. Hall, and Huan Liu. "XBully: Cyberbullying Detection within a Multi-Modal Context". In: *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*. WSDM '19. Melbourne VIC, Australia: Association for Computing Machinery, 2019, pp. 339–347. ISBN: 9781450359405. DOI: 10.1145/3289600.3291037. URL: https://doi.org/10.1145/3289600.3291037.

[8] Harsh Dani, Jundong Li, and Huan Liu. "Sentiment Informed Cyberbullying Detection in Social Media". In: *Machine Learning and Knowledge Discovery in Databases*. Ed. by Michelangelo Ceci, Jaakko Hollmén, Ljupčo Todorovski, Celine Vens, and Sašo Džeroski. Cham: Springer International Publishing, 2017, pp. 52–67.

[9] Laura P. Del Bosque and Sara Elena Garza. "Aggressive Text Detection for Cyberbullying". In: *Human-Inspired Computing and Its Applications*. Ed. by Alexander Gelbukh, Félix Castro Espinoza, and Sofía N. Galicia-Haro. Cham: Springer International Publishing, 2014, pp. 221–232. ISBN: 978-3-319-13647-9.

[10] United Nations Statistics Division. *Standard Country or Area Codes for Statistical Use (M49)*. 2021. URL: https://unstats.un.org/unsd/methodology/m49/ (visited on 07/15/2021).

[11] J. Dyl, J. Kittler, K. A. Phillips, and J. I. Hunt. "Body dysmorphic disorder and other clinically significant body image concerns in adolescent psychiatric inpatients: prevalence and clinical characteristics". In: *Child Psychiatry Hum Dev* 36.4 (2006), pp. 369–382.

[12] Paul Ekman. "Basic emotions". In: *Handbook of cognition and emotion* 98.45-60 (1999), p. 16.

[13] D. A. Ellis, R. Wiseman, and R. Jenkins. "Mental Representations of Weekdays". In: *PLoS One* 10.8 (2015), e0134555.

[14] Bjarke Felbo, Alan Mislove, Anders Søgaard, Iyad Rahwan, and Sune Lehmann. "Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm". In: *Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 2017.

[15] Mohammed Ali Al-garadi, Kasturi Dewi Varathan, and Sri Devi Ravana. "Cybercrime detection in online communications: The experimental case of cyberbullying detection in the Twitter network". In: *Computers in Human Behavior* 63 (2016), pp. 433–443. ISSN:

0747-5632. DOI: https://doi.org/10.1016/j.chb.2016.05.051. URL: https://www.sciencedirect.com/science/article/pii/S0747563216303788.

[16] Bo Han, Paul Cook, and Timothy Baldwin. "Text-Based Twitter User Geolocation Prediction". In: *J. Artif. Int. Res.* 49.1 (Jan. 2014), pp. 451–500. ISSN: 1076-9757.

[17] C.J. Hutto and Eric Gilbert. "VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text". In: Jan. 2015.

[18] Ross Krawczyk and J. Kevin Thompson. "The effects of advertisements that sexually objectify women on state body dissatisfaction and judgments of women: The moderating roles of gender and internalization". In: *Body Image* 15 (2015), pp. 109–119. ISSN: 1740-1445. DOI: https://doi.org/10.1016/j.bodyim.2015.08.001. URL: https://www.sciencedirect.com/science/article/pii/S1740144515000960.

[19] Colette Langos. "Cyberbullying: The Challenge to Define". In: *Cyberpsychology, Behavior, and Social Networking* 15.6 (2012). PMID: 22703033, pp. 285–289. DOI: 10.1089/cyber.2011.0588. eprint: https://doi.org/10.1089/cyber.2011.0588. URL: https://doi.org/10.1089/cyber.2011.0588.

[20] Jennifer Lewallen and Elizabeth Behm-Morawitz. "Pinterest or Thinterest?: Social Comparison and Body Image on Social Media". In: *Social Media + Society* 2.1 (2016), p. 2056305116640559. DOI: 10.1177/2056305116640559.

[21] Amrita Mangaonkar, Allenoush Hayrapetian, and Rajeev Raje. "Collaborative detection of cyberbullying behavior in Twitter data". In: *2015 IEEE International Conference on Electro/Information Technology (EIT)*. 2015, pp. 611–616. DOI: 10.1109/EIT.2015.7293405.

[22] Amgad Muneer and Suliman Mohamed Fati. "A Comparative Analysis of Machine Learning Techniques for Cyberbullying Detection on Twitter". In: *Future Internet* 12.11 (2020). ISSN: 1999-5903. DOI: 10.3390/fi12110187. URL: https://www.mdpi.com/1999-5903/12/11/187.

[23] United Kingdom Office of National Statistics. *Nomis Census 2011 Tables*. 2021. URL: https://www.nomisweb.co.uk/census/2011/data_finder (visited on 07/15/2021).

[24] United Nations. *Demographic and Social Statistics*. 2019. URL: https://unstats.un.org/unsd/demographic-social/products/dyb/dyb_2019/ (visited on 07/15/2021).

[25] James Pennebaker, Ryan Boyd, Kayla Jordan, and Kate Blackburn. *The Development and Psychometric Properties of LIWC2015*. Sept. 2015. DOI: 10.15781/T29G6Z.

[26] Filipe N. Ribeiro, Matheus Araújo, Pollyanna Gonçalves, Marcos André Gonçalves, and Fabrício Benevenuto. "SentiBench - a benchmark comparison of state-of-the-practice sentiment analysis methods". In: *EPJ Data Science* 5.1 (July 2016), p. 23. ISSN: 2193-1127. DOI: 10.1140/epjds/s13688-016-0085-1. URL: https://doi.org/10.1140/epjds/s13688-016-0085-1.

[27] Elvis Saravia, Hsien-Chi Liu, Yen-Hao Huang, Junlin Wu, and Yi-Shin Chen. "CARER: Contextualized Affect Representations for Emotion Recognition". In: Jan. 2018, pp. 3687–3697. DOI: 10.18653/v1/D18-1404.

[28] Stephan Schlosser, Daniele Toninelli, and Michela Cameletti. "Comparing Methods to Collect and Geolocate Tweets in Great Britain". In: *Journal of Open Innovation: Technology, Market, and Complexity* 7.1 (2021). ISSN: 2199-8531. DOI: 10.3390/joitmc7010044. URL: https://www.mdpi.com/2199-8531/7/1/44.

[29] N. Shirao, Y. Okamoto, G. Okada, and S. Yamawaki. "Temporomesial Activation in Young Females Associated with Unpleasant Words concerning Body Image". In: *Neuropsychobiology* 48.3 (2003), pp. 136–142. ISSN: 0302-282X. DOI: 10.1159/000073630. URL: https://doi.org/10.1159/000073630.

[30] L. Sloan and J. Morgan. "Who Tweets with Their Location? Understanding the Relationship between Demographic Characteristics and the Use of Geoservices and Geotagging on Twitter". In: *PLoS One* 10.11 (2015), e0142209.

[31] Luke Sloan, Jeffrey Morgan, William Housley, Matthew Williams, Adam Edwards, Pete Burnap, and Omer Rana. "Knowing the Tweeters: Deriving Sociologically Relevant Demographics from Twitter". In: *Sociological Research Online* 18.3 (2013), pp. 74–84. DOI: 10.5153/sro.3001. eprint: https://doi.org/10.5153/sro.3001. URL: https://doi.org/10.5153/sro.3001.

[32] Arthur Stone, Stefan Schneider, and James Harter. "Day-of-week mood patterns in the United States: On the existence of 'Blue Monday', 'Thank God it's Friday' and weekend effects". In: *The Journal of Positive Psychology* 7 (July 2012), pp. 1–9. DOI: 10.1080/17439760.2012.691980.

[33] Marika Tiggemann and Amy Slater. "The Role of Self-Objectification in the Mental Health of Early Adolescent Girls: Predictors and Consequences". In: *Journal of Pediatric Psychology* 40.7 (Mar. 2015), pp. 704–711. ISSN: 0146-8693. DOI: 10.1093/jpepsy/jsv021. eprint:

https://academic.oup.com/jpepsy/article-pdf/40/7/704/5107584/jsv021.pdf. URL: https://doi.org/10.1093/jpepsy/jsv021.

[34] Inc. Twitter. *Twitter Academeic Research Track description*. 2021. URL: https://developer.twitter.com/en/products/twitter-api/academic-research (visited on 06/15/2021).

[35] Gina Unterhalter, Simon Farrell, and Christine Mohr. "Selective memory biases for words reflecting sex-specific body image concerns". In: *Eating Behaviors* 8.3 (2007), pp. 382–389. ISSN: 1471-0153. DOI: https://doi.org/10.1016/j.eatbeh.2006.11.015. URL: https://www.sciencedirect.com/science/article/pii/S1471015306000936.

[36] Heidi Vandebosch and Katrien Van Cleemput. "Defining Cyberbullying: A Qualitative Research into the Perceptions of Youngsters". In: *CyberPsychology & Behavior* 11.4 (2008). PMID: 18721100, pp. 499–503. DOI: 10.1089/cpb.2007.0042. eprint: https://doi.org/10.1089/cpb.2007.0042. URL: https://doi.org/10.1089/cpb.2007.0042.

[37] Zhaoxia Wang, Chee Seng Chong, Landy Lan, Yinping Yang, Seng Beng Ho, and Joo Chuan Tong. "Fine-grained sentiment analysis of social media with emotion sensing". In: *2016 Future Technologies Conference (FTC)*. 2016, pp. 1361–1364. DOI: 10.1109/FTC.2016.7821783.

[38] K. J. Watts and J. Cranney. "Automatic evaluation of body-related words among young women: an experimental study". In: *BMC Public Health* 10 (June 2010), p. 308.

[39] N. Yazdani, S. V. Hosseini, M. Amini, Z. Sobhani, F. Sharif, and H. Khazraei. "Relationship between Body Image and Psychological Well-being in Patients with Morbid Obesity". In: *Int J Community Based Nurs Midwifery* 6.2 (Apr. 2018), pp. 175–184.

[40] David Zimbra, Ahmed Abbasi, Daniel Zeng, and Hsinchun Chen. "The State-of-the-Art in Twitter Sentiment Analysis: A Review and Benchmark Evaluation". In: 9.2 (Aug. 2018). ISSN: 2158-656X. DOI: 10.1145/3185045. URL: https://doi.org/10.1145/3185045.