The Weather Blues: Wrangling Weather Data U

# Abstract

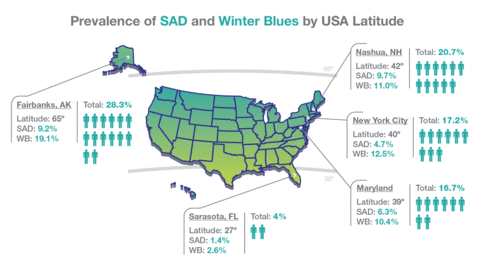
The digital evolution has had profound changes on the study of Depression and mood related disorders. With computers infiltrating every aspect of our lives, non-intrusive methods of psychological research have started to evolve. While many researchers focus on Social Media as means to analyze human behavior, large open source datasets from consumer products like the Yelp Review Dataset provide rich, clean sources of data to explore and analyze human behavior. My goal was to use natural language processing (NLP) on the Yelp review dataset to explore variations in sentiment, mood, and rating with respect to various weather-related data. Along these lines, I wanted to explore the physiological implications of weather: does weather cause depression? Does weather impact business? I used an open source Python text-analysis library to perform basic sentiment analysis and assist in parts-of-speech tagging. I then created a simple “term frequency, inverse document frequency” (TF-IDF) script to analyze pronoun usage using the methods popularized by Pennebaker (2011). Combining the weather data with each review, I found no evidence that weather has an impact on sentiment or mood. There was however an expected relationship between usage of “I words” (indicating depression or worse mood) and lower review scores (i.e a higher star rating). From a personal growth prospective, working on this project provided me with a wealth of useful information on project management, data collection, data visualization, API’s, and much more.

# Seasonal Affective Disorder (SAD)

Seasonal affective disorder is a mood disorder that exhibits similar characteristics of depression by virtue of the change in seasons. According to Cedars Sinai, SAD is caused by a chemical change in the brain that is associated with less daylight and shorter days (“Seasonal Affective Disorder”, n.d). Colloquially, most people associate SAD with Winter, where the days are shorter, and the temperature is cooler. However, the onset of SAD is subjective to an individual’s respective background. People already diagnosed with mood disorders are often diagnosed with SAD when their symptoms worsen in the winter or summer. They are uniquely classified as Fall-onset or Spring-onset, and they are diagnosed through a professional evaluation.

SAD shares many traits similar to clinical depression, and as a result, it is under researched and the causes of SAD are not entirely understood. That being said, there are some interesting avenues for exploration. A 2007 study found that SAD is more prevalent and severe amongst college aged students (Sullivan & Payne, 2007). A 2001 study found levels of melatonin in patients diagnosed with SAD were higher than a control group (Wehr & Duncan, 2001). These findings went against the findings of a study performed by Checkley and Murphy in 1993 that found no link between melatonin levels and SAD (Checkley & Murphy, 1993). Possibly contributing to the maligned “winter only” definition of SAD is that geographic location also plays a role in SAD. Prevalence of winter SAD is much higher in northern latitudes then it is in southern latitudes (Rosen, 1990). In the United States, a 2017 study found that daylight savings time contributed to a 11% increase in SAD episodes (Hansen, 2017).

Figure



# Personal Anecdotes on Seasonal Affective Disorder

I’ve been lucky enough to grow up in Los Angeles most of my life. Unlike many other places in the world, when I woke up for school as a child, I never had to think twice about wearing a t-shirt and flips flops to school. This changed for me when I decided to play basketball at a liberal arts college in Massachusetts. Moving across the country was what initially opened my eyes to the impact weather could have on my mood. I distinctly remember early on my Freshman year in Massachusetts, after five straight days of lighting and thunder, I felt significantly less happy. Even in LA, there are weeks where it rains a lot and while it may seem like the world is ending, I had never experienced mood changes because of it. But in Massachusetts, these changes in weather had a profound effect on me. As the semester progressed and the days became shorter, I realized my mood was almost constantly tied to the weather. Surprisingly, overcast days were more impactful on me then cold days. It wasn’t until a conversation with another student from California that I learned of seasonal affective disorder. After some research, I realized I had been experiencing many of the SAD symptoms.

After one year in Massachusetts, the warm mild climate of Los Angeles was calling my name again, so I decided to transfer to USC. As expected, the familiar climate immediately solved any resemblance of SAD, but my time in Massachusetts left with a defining experience of the affects weather can have on a “thin skinned” Angeleno like myself. After settling in at USC, I didn’t think much of my experience with weather in Massachusetts. Until, that is, my sister announced she’d be moving to a considerably more sever climate than Massachusetts.

My sister also grew up in Los Angeles, but unlike me, she attended USC straight out of high school, making her first 22 years of life all in Los Angeles. Once she was accepted to a medical school in Milwaukee, Wisconsin, my family and I all worried that she would also experience SAD once she moved there. Surprisingly, her first year in Wisconsin went well; she studied so much that she was barely ever outside.. But in her second year, the winter weather worsened considerably, and she too felt the consequences of her disposition as a life-long Angeleno. I empathized with her situation, as she began to describe many of the same mood swings that I experienced from the weather.

Around the same time my sister was sharing her experiences with SAD was when I began to explore options for a final project in PSYCH499. My experience combined with my sister’s led me to start asking questions that are essential to any research project. I wondered if I was predisposed to getting SAD because of where I grew up. I also wondered how long it takes someone to acclimate to a new climate. Given the difficulty of answering these questions, there wasn’t much research on SAD in the specific context of migration. Most research surrounding SAD focus on the biochemical reactions that cause and alleviate SAD, not the very specific subpopulation of people who are moving to new places. Given the unique possibilities that digital technologies provide psychological research, I decided I would attempt to answer these questions.

# Methods

What are the physiological and behavioral impacts of weather on different demographics of people? This was my initial research question when I first started this project. Initially, I brainstormed a few ways to answer this question. Option one was to use geographically filtered Tweets that responded to various weather patterns. For example, there’s a heat wave in San Francisco; I collect tweets from before, during, and after that the heat weave to see how sentiment and the contents of tweets changed. I ditched the Twitter method due to the restrictions of the API and the lack of time I had to complete the project. My second option was scraping travel blogger post entries. This method would successfully capture the effect that moving from one place to another would have on mood and sentiment, so it seemed like a good option. However, after further exploration it became apparent that most travel bloggers tag their posts by country. As a result, the weather data associated with each post would’ve been an average of the entire country’s weather— making the weather data too broad and effectively useless. The final option I laid out was to use a Kaggle data set containing reviews from a large time period. This option gave me a clean, large dataset, that allowed me to explore multiple relationships between weather and the contents of the reviews.

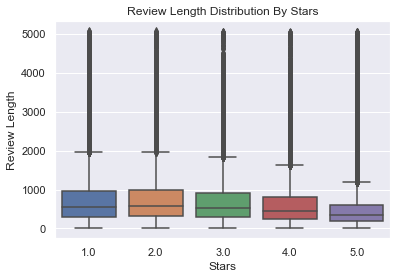
## Yelp Dataset

Yelp is a massively popular crowd-sourced review forum that was founded in 2004 (Yelp, 2019). The founders of Yelp hoped to hedge the internet to fuel the democratization of public recommendations for businesses of all types. They’ve grown tremendously since 2004 and have overthrown the previous standard of using Yellow Pages for business information. Given that they are a digital platform, they possess an incredible amount of valuable data. With this trove of data, they created the Yelp Dataset Challenge: a competition for students to find meaningful insight from their open sourced dataset. The dataset and competition is updated every year, and currently the competition is in it’s thirteenth iteration (Yelp Dataset, 2019).

The dataset consists of over 6.7 million reviews spanning over 10 metropolitan areas. One of the many issues with physiological research is the necessity of performing lab experiments with a limited sample sizes. With over 1 million unique users, the dataset is large, nonreactive, complete, and clean. One issue with the dataset is its representativeness. Some assumptions I’ve made on the dataset is that it represents a tech-savvy population of users and might not represent older populations. Usage drift and system drift are other issues with large datasets like the Yelp dataset. Salganik defines behavioral drift as the change in how people use a data collection system, and system drift as the change in the data collection system itself (Salganik, 2018). Given the growth and changes of the platform over time, the Yelp datasets longitudinal research is unclear.

The dataset consists of six unique files for businesses, users, reviews, check-ins, tips, and photos. For this project I used the businesses, the users, and the reviews files. In order the load the dataset without crashing my computer, I sampled 500,000 reviews from the 6.7 million reviews. Using Pandas, I merged the two datasets so that each review contained a geographic location. Finally, I performed exploratory data analysis to develop an understanding of the data I was working with.

|  |  |
| --- | --- |
| C:\Users\rasap\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\EF31050B.tmp | C:\Users\rasap\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\9CADAE51.tmp |
|  |  |



## Weather Data Collection

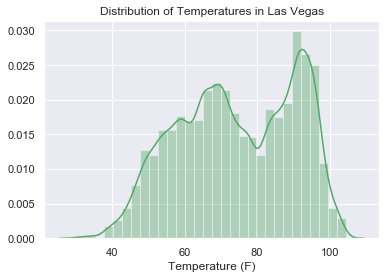
##### Attempt One:

Since 1970, the National Oceanic and Atmospheric Administration (NOAA) has sought to accurate weather information to the general public. NOAA uses a national gird of certified weather stations to collect and send weather data to a national database (*About NCDC*, n.d.). Every single news channel or weather application in the United States creates their predictions from the open source data that NOAA provides. The National Center for Environmental Information (NCDC), a subsidiary of NOAA, uses an API to distribute satellite, radar, and temperature data to the entities that need it. Given my needs for this project I sought to use this API to provide comprehensive historical weather data for all 500,000 reviews that I was using.

While planning this project, I assumed that collecting the weather data would be one of the easier tasks. I was wrong. The first problem I encountered was the NCDC API daily request limit of 10,000. With the assumption that each request returns data for one review, I would’ve needed 50 days to collect weather data for all 500,000 reviews that I was using. To work around this, I found the oldest and newest review for every zip code in the dataset. This effectively established a range of dates for every zip code that I needed weather data for. By doing requesting data only within this range, I was able to reduce the time I needed to collect the weather data from 50 days to a couple hours.[[1]](#footnote-1)

After I finished running the collection script, I ran into my second problem: messy, incomplete data. For starters, some zip codes had no historical data, making them completely useless. For the zip codes that returned historical data, the data returned was incomplete. Some zip codes only returned mean temperature. Some zip codes only returned snow and rain counts. What I had was almost completely useless.

##### Attempt Two:

After some brainstorming with my TA, I changed my approach. In EDA, I found that nearly a quarter of the dataset consisted of reviews from Las Vegas. By limiting the dataset to one zip code, I narrowed the focus of finding the historical data considerably. I Googled “Las Vegas Historical Weather” and found a few free non-government website containing *clean* daily historical data. I chose usclimatedata.com and proceeded to build a script that scraped the entire website in a few minutes. This time, the data was perfect as it included average, high, and low temperature and rain and precipitation for every review in Las Vegas.

## Natural Language Processing

##### Sentiment Analysis

Since my goal was to understand if the reviewers exhibit depressive traits in response to climatic changes, I considered building a domain specific sentiment analysis model. Ideally, the model would classify the review from 0 to 1 if it exhibited traits of depression. The difficulty with building this model however was that there was no way to reliably train it. Instead, I used TextBlob, a Python text processing library. Built in to the library was a sentiment analysis tool that I used to classify the sentiment of every review.

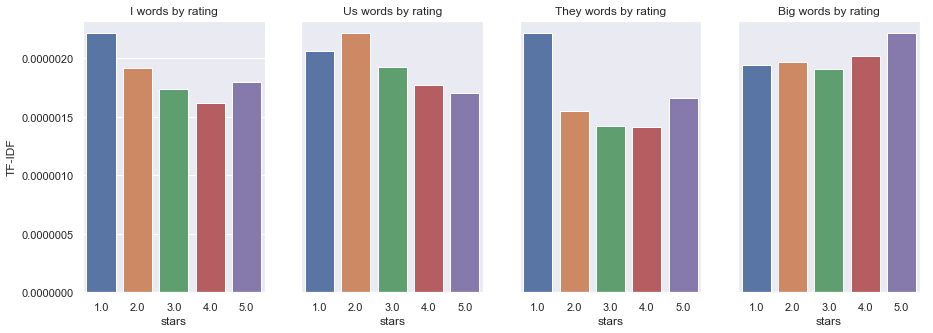
##### Pennebaker Pronoun Usage

Early work in social sciences using natural language processing was done by word count. By comparing the relative word frequencies of a list of words from two documents, we can unlock information about the author. Pennebaker has shown that people who are exhibiting depression like symptoms use pronouns differently (Pennebaker, 2011). More specifically, the usage of “I words” tends to increase while “we words” decrease simultaneously in depressed individuals.

For every review, I counted the number of “I words”, “we words”, “they words”, and “big words.” I classified a big word as any word longer than 9 characters. After, I found the TF-IDF rating of each of these categories of words for each review. The TF-IDF provides the relative frequency of these words compared to the usage in all the reviews.

# Results and Findings

## Study 1:



Study one attempted to show the variations of the different pronoun group usage with respect to stars. I did not attempt to use weather in this study to see if variations in the data exist exclusive of weather. By calculating the mean TF-IDF for each group and rating, I was able to produce bar charts that showed some expected results. “I word” and “they word” usage increased when users gave a bad reviews. Intuitively this makes sense. A person that is unsatisfied with their experience is more likely to speak on the specific occurrences of their experience, using more pronouns to describe the interactions between themselves and the business. Satisfied users on the other hand used fewer “I and we words” and instead used big words at a higher frequency. This also makes sense as satisfied users might be more inclined to talk about their feelings they had with respect to their experience.

What didn’t show here was an inverse relationship between “I word” and “we word” frequency. Had the usage of “us words” increased with rating, their might’ve been evidence that reviewers that give one star reviews exhibit traits of depressed writers.

## C:\Users\rasap\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\2F48CA33.tmpStudy 2:

For Study 2, I compared the sentiment rating from each review with the mean temperature from that day. As shown, there is no significant relationship between the two variables.

My expectation was that there would’ve been a negative correlation between sentiment and mean temperature. This lack of correlation is not necessarily an indication that weather doesn’t affect someone’s mood. Instead, it shows that a review contains a far too narrow context to identify the implications of an external variable like weather. A review is a short and narrow account of an experience a user has had with a business. The likelihood of an extraneous factor like weather impacting reviews over the totality of a dataset is very unlikely.

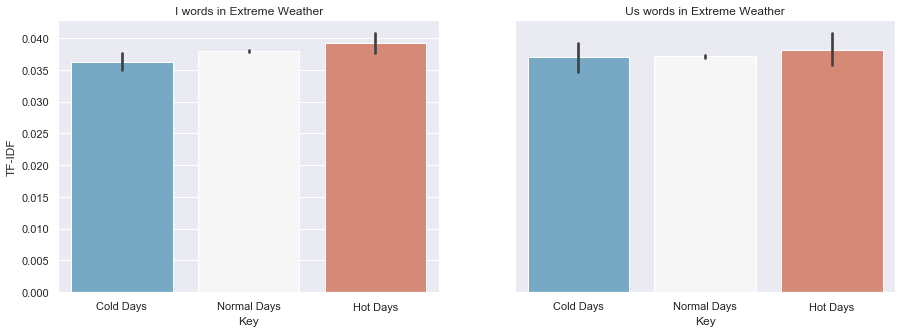
## Study 3:

Study three attempts to capitalize on the shortcomings of study two by narrowing the focus to a subset of users. Here, I selected the four users with the most number of reviews, and ran a regression on each users usage of “I words” and “us words.” Again, the data is messy and doesn’t appear to show much of a meaningful relationship between temperature and depressive pronoun usage. However, “Us words” appear to be less frequently used as temperature increases. While this might be significant, the error padding on all the regression lines are very wide, indicating the variability of the data.

|  |  |
| --- | --- |
| C:\Users\rasap\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\3F8B2867.tmp | C:\Users\rasap\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\D04160D.tmp |

## Study 4:

During my time in Massachusetts, extreme weather had significant impacts on my mood. I decided to analyze some of our findings when the weather was particularly cold or particularly cold. Since weather histogram is not normal, I decided to set arbitrary upper and lower temperature limits to create the subgroups. For the hot days, I filtered out all the days where the high surpassed 112 degrees Fahrenheit. For the cold days, anything below 35 degrees Fahrenheit would be considered cold. I then compared the two to the rest of the population to produce some interesting results. “I words” were used more frequently when the weather was very hot. Contrastingly, on days where it was cold, I word usage was considerably lower. The results for “us words” followed a similar pattern, but the error was considerably higher.



## Takeaways

Natural language proccesing has already had a massive impact on many different industries and fields of study. However, in order to see this impact, some basic requirements are neccsary. This project in particular has proven to me that a combination of high quality data and good research design practices are essential for a data science project of this size.

My intial research question was to find the physiological and behavioral impacts of weather on different demographics of people. While there were some interesting patterns, the findings from this project show no evidence that weather influences our physiological status and our behavior. Substantial research into seasonal affective disorder has shown that many different climatic variables impact our phycological state. However, as shown by this project using natural language processing on business reviews is unlikely to provide any meaningful results.

##### Limitations

Although my project did not provide many findings, I want to at least point out some of the limatations of the proccess I used.

###### Dataset

The Yelp dataset has already proven to be a fantastic resoruce for research and natural language proccessing. However, attempting to use review data to detect psyccoolcigical impacts of weather is a very difficult task. While context is very important in any sort of natural languge proccessing task, prupose is also very important. The purpose of a Yelp review is to describe the experience you had in the minutes or hours that you were at a business. It’s a very narrow and direct task, and very different from a Tweet where there are no bounds on what is appropratie to write about. For this reason, I think the Yelp dataset shouldn’t be used for projects similar to this one. Instead, using Tweets or blog posts might provide more represenative data.

###### Weather Data

Another limatiation of this study was the limited weather data. While I was able to gain access to temperatutre and rain data, SAD has been shown to be caused by light. Humidity and cloud cover are two other variables that would’ve made this analysis comprehensive.

###### Only one city

Las Vegas has a very unique climate, so it’s possible that the residents of Las Vegas are immune to chagnes in weather. Ideally, we would compare these results with another city or municplaity with a completely different weather profile. Due to time constratints I was unable to accomplish this.

1. I will also note that the servers went down for 48 hours for maintenance while I was testing my script. It took a couple of hours before I realized that it wasn’t my code, but the downtime of the server. This too, was a considerable setback. [↑](#footnote-ref-1)