CORRELATION BETWEEN WEATHER AND TWITTER SENTIMENT ACROSS THE U.S.

Yiting Wang
ywang803@ucsc.edu

ABSTRACT

In this project, I tried to find out if there is a relationship between the weather conditions and people's moods, and I visualized the data of weather conditions and Twitter sentiment within the U.S. to help to find the answer.

1   INTRODUCTION

Weather affect our daily lives, from what we eat, what we wear, what activities we do, what transportation we use.Daggumati et al. (2016) In common sense, we knew the weather made us moody because we were gloomy on rainy days and cheery when the sun shone.Livni To confirm this idea, for this project, I'm planning to measure the connection between moods and meteorology on a large scale with the help of Twitter, which is an online social network that allows users to upload short text messages–tweets–of up to 140 characters. This restriction encourages users to construct focused and timely updates. Accordingly, analysis of Twitter users to understand their sentiments and predict human behaviors has gained popularity recently, **e.g.** applications to predicting stock prices, forecasting U.S. presidential elections, etc.

However, studies of the relationship between sentiment and weather is relatively rare, in this study, I use a huge amount of Twitter data to get the U.S. residents' sentiments, and then present the relationship between the sentiment data and weather variables such as temperature, humidity, and sea level pressure based on nationwide meteorological records.

2   MOTIVATION

There are many factors that can affect a person's mood and sometimes, the same factors will have the opposite effect on people's emotions. Take myself as an example, before coming to California, the city I lived in had the very similar weather condition as Seattle. So when the sun came out, everyone will go out to enjoy the sunbathe happily. After living in California for a year, if it rains today, my mood must be super good because I really miss the humid and fresh air on rainy days, and rain can reduce the chance of wildfire, which is also cheerful because wildfire is very common in California.

When it comes to "good weather", we often assume that it is a sunny day. But after I have this experience, I want to know the relationship between the weather conditions and people's moods. And visualization of various data involved takes an important role in my project to more directly find the answer.

3   RELATED WORK

- Twitter Sentiment AnalysisRipamonti: A Kaggle notebook to analyze the sentiment of Tweets.
- TweerAnalyzerDoshi et al. (2017): Twitter Trend Detection and Visualization: a system to extract real-time Twitter data and to represent the trending Twitter hashtags and active users on a bar graph.
- Twitter sentiment classification using stanford NLPPhand & Phand (2017)

Figure 1: weather and mood

- Enhanced Sentiment Classification Using Geo Location TweetsAnil Phand & Chakkarwar (2018)
- Visually comparing weather features in forecastsQuinan & Meyer (2016): visualize the weather forecasts data to offer support for Meteorologists when exploring the uncertainty in the simulations.
- EarthBeccario: A project to visualize global weather conditions.

Visualizations about sentiment and weather are rare; however, there are a lot of studies showing the separated two parts of work. Visualization of weather has been very mature now, and there are also many visualizations of sentiment online as shown in the figure 2 and figure 3.
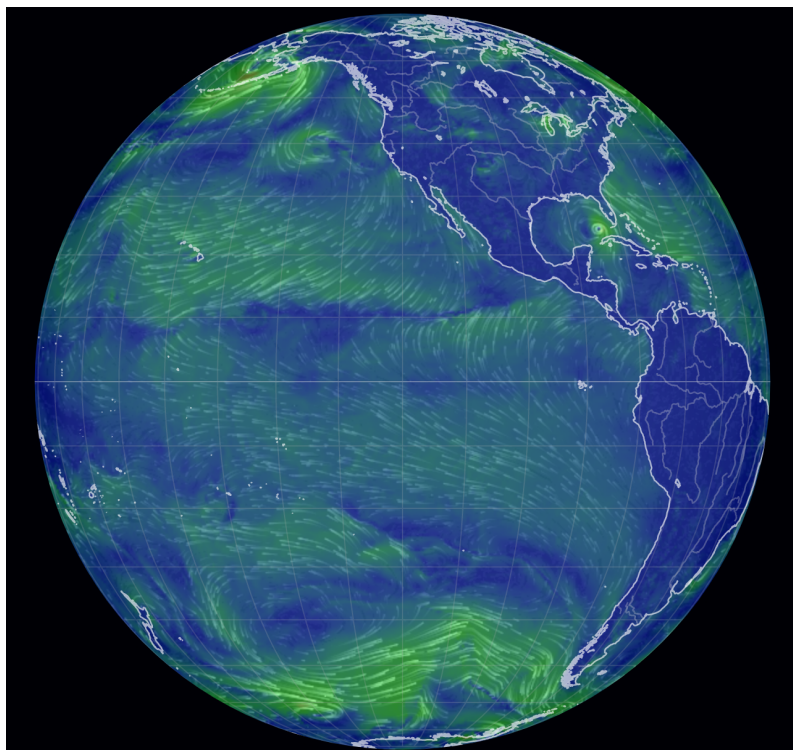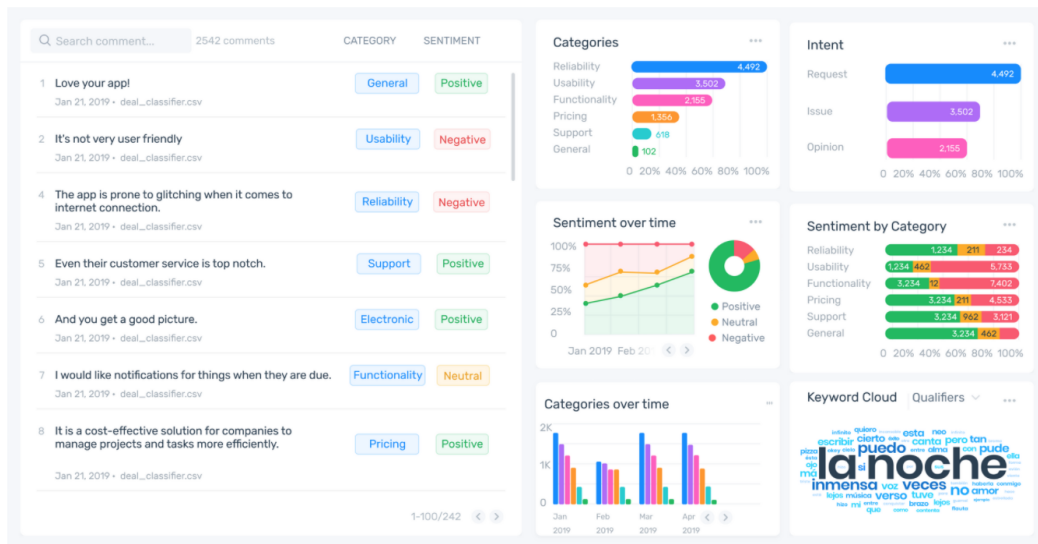

Figure 2: earth: visualization of weather

Figure 3: visualization of sentiment

4 PROPOSED RESEARCH DIRECTIONS

- Gather the Tweet data that is needed.
- Analyze Tweets' sentiment in the United States in April 2019.
- Collect weather data **e.g. temperature, humidity, sea level pressure** in the United States in April 2019.
- Visualize the relationship of above data using statistics.

5 DATA

Twitter data

I have crawled tweet data for half a month(April 2019) in the United States, and I choose the biggest city of each state to represent each state. Firstly, I run geocode query on twitter to find the place ID of each city **e.g.** place ID of Los Angeles is 3b77caf94bfc81fe, and use snscrape which is a scraper for social networking services (SNS) to gather URL/Tweet_id. JustAnotherArchivist Tweet_id is a unique identifier provided to each tweet. Then I use Twitter API (via tweepy) to extract historical tweets. I have gathered 1,695,175 tweets in total for analyzing.

Sentiment data

The sentiment data made by the posters were analyzed using Linguistic Inquiry and Word Count(LIWC), a tool used as a measure of emotion expression. It offers powerful sentiment analysis, and we can get the positive score and negative score of each tweet.

Weather data

The weather data were collected from national weather service and weather underground, which provides historical as well as real-time weather data within the United States. (http://wunderground.com). The averaged weather variables from multiple weather stations are used to represent the weather condition for the state. Park et al. (2013) The weather variables include:

- Temperature: the mean temperature during daytime
- Sea level pressure: the atmospheric pressure at sea level at a given location
- Precipitation: the amount of rain over one day
- Relative humidity: the amount of water vapor present in the air
- Cloud cover: the fraction of the sky obscured by clouds when observed from a particular location
- Dew point: the temperature the air needs to be cooled to (at constant pressure) in order to achieve a relative humidity (RH) of 100%

6 RESULTS

Firstly, I present the average positive sentiment of each state during the half of a month(April, 2019) on a choropleth map as shown in Figure 4, the color depth reflects the level of positive affect. Figure 5 shows there is a clear relationship between latitude and positive sentiment. As the latitude increases from the south to the north, the positive sentiment from Twitter decreases. Since the biggest difference in weather between the north and the south is temperature(Usually the temperature in the north is higher than in the south), I predict that people live in the states where the temperature are higher have more positive sentiment than people live in colder states. To confirm my prediction, I further analyze the relationship between weather variations and Twitter sentiments in more detail. I try to analyze the relationship between the sentiment and weather using linear
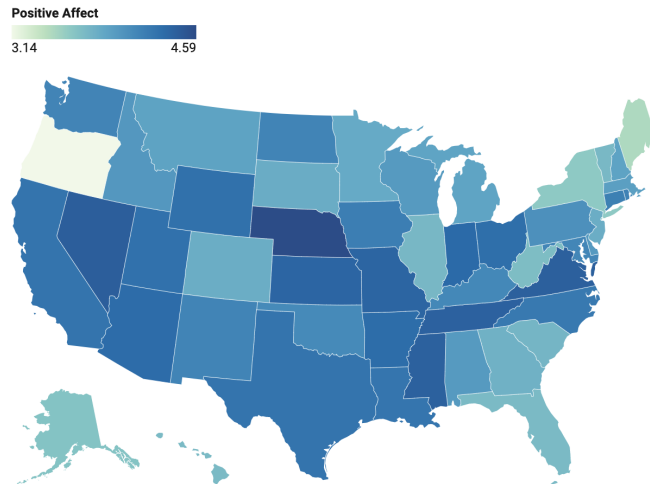

Figure 4: positive sentiment across states during Apr.1-Apr.14

regression, and calculate Pearson correlation r for each relationship. Among the various weather variables, I found three significant factors that shows correlation with mood. Except for them, all other weather variables I have analyzed shows negligible correlation.

Figure 6 shows the relationship between the mean temperature, relative humidity, sea level pressure and positive mood across all 50 states. The x-axis represents the mean weather variable of each state for the half of a month, and the y-axis represents the positive mood for each state.Park et al. (2013) The best linear fit is the solid line and the shaded area represents the 95% confidence interval. Figure 6 clearly shows a positive relationship between the mean temperature and positive mood, and negative relationship between the mean relative humidity and positive mood, between mean sea level pressure and positive mood, with r=0.248, r=-0.241, r=-0.185 for each relationship. Suggesting that people living in a state having higher mean temperature, or lower mean relative humidity, or lower mean pressure do express more positive affect on Twitter than those living in states where the mean temperature is colder, or the mean relative humidity is higher, or the mean pressure is higher.
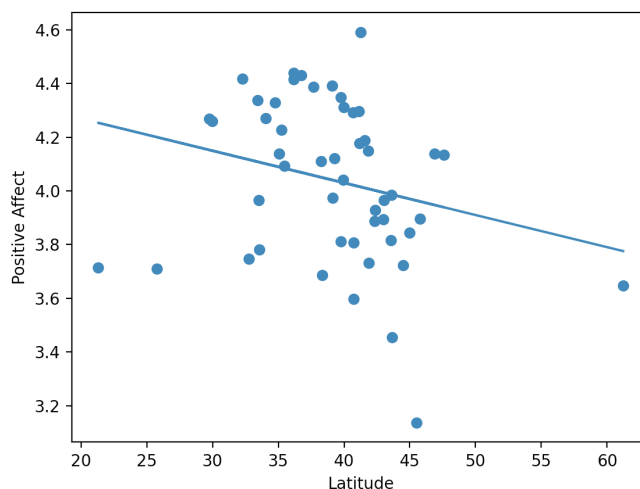
Figure 5: relationship between latitude and positive sentiment across 50 states
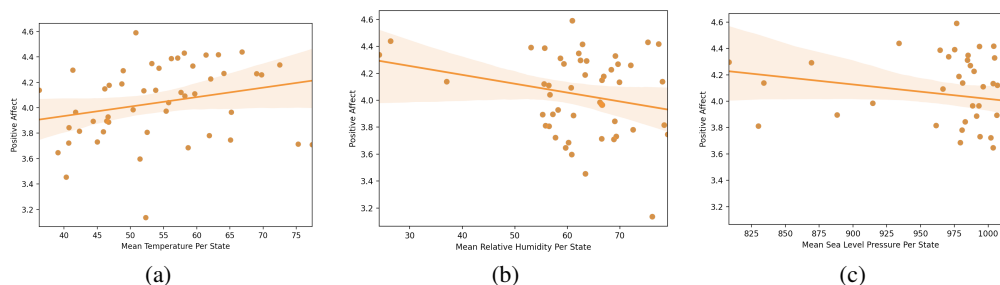

| (a) | (b) | (c) |

Figure 6: relationship between three variables and positive sentiment across 50 states

To confirm above results, I analyze the relationship of weather variables and positive sentiments within each state, trying to find if the same pattern exits within each state during April 1, 2019 to April 15, 2019.

Temperature

The relationship between temperature and positive sentiment within each state is in consistent with the relationship between the mean temperature and positive sentiment across 50 states. Figure 7 shows Pearson's coefficient value in each state, the deeper blue color represents the greater degree of positive correlation, and the deeper red color represents the greater degree of negative correlation. As shown in Figure 8, in most of states, people feel positive emotions when the temperature is high, but there is still some states showing that people feel positive emotions when the temperature is low.

Relative Humidity

The relationship between humidity and positive sentiment within each states is surprisingly inconsistent with the relationship between the mean Relative Humidity and positive sentiment across 50 states. Figure 9 shows Pearson's coefficient value in each state. As shown in Figure 10, there are more positive relative relationships between relative humidity and positive sentiment, which contradicts the previous conclusion. It implies that many people feel more positive emotions when the relative humidity is high, while some people feel less positive emotions when the relative
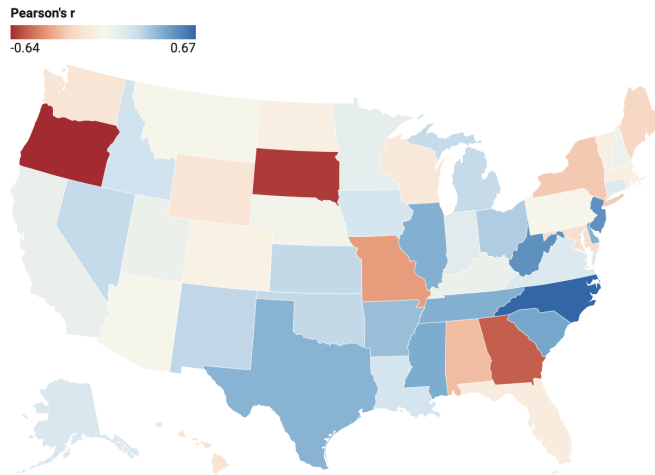
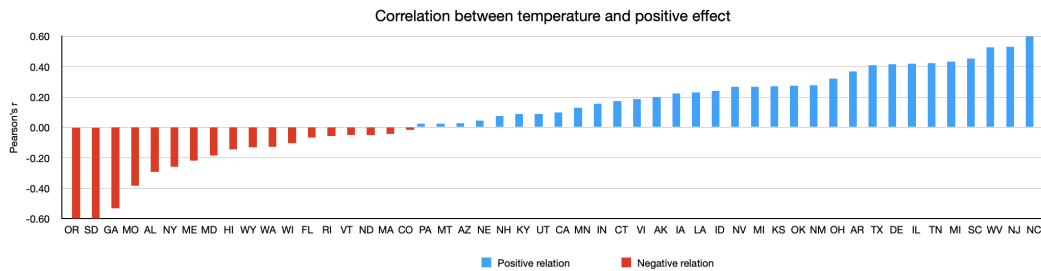Figure 7: correlation between temperature and positive sentiment

Figure 8: correlation between temperature and positive sentiment

humidity is low.

While this might be an issue to explored in future studies, if we take a closer look at some specific area, we can find some interesting results. Some states like Seattle, shows a negative correlation in that an increase in humidity led to a decrease in positive sentiments-perhaps humidity is a very common factor in a "rain city". And some states like Arizona, Nevada, however, shows a positive correlation between relative humidity and positive sentiment, it's still reasonable given that humidity is a welcome factor on a desert landscape.Park et al. (2013)

Sea Level Pressure

Finally, I also found negative relationship between the pressure and positive sentiments. Figure 11 and Figure 12 shows the relationship between atmospheric pressure and positive affect across all states. The majority of states exhibited positive affect on days with low atmospheric pressure, yet a few states showed the opposite trend. This is in consistent with the general trend across all 50 states.

It is also worth noting that people living in the west of the U.S. show positive correlation and eastern people show negative correlation between atmospheric pressure and positive affect.
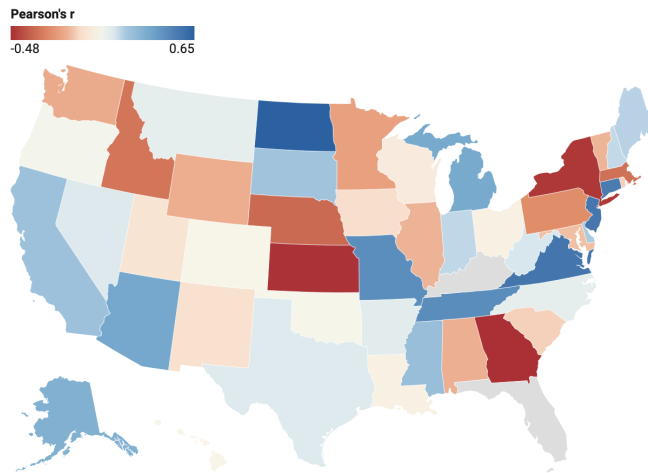
6

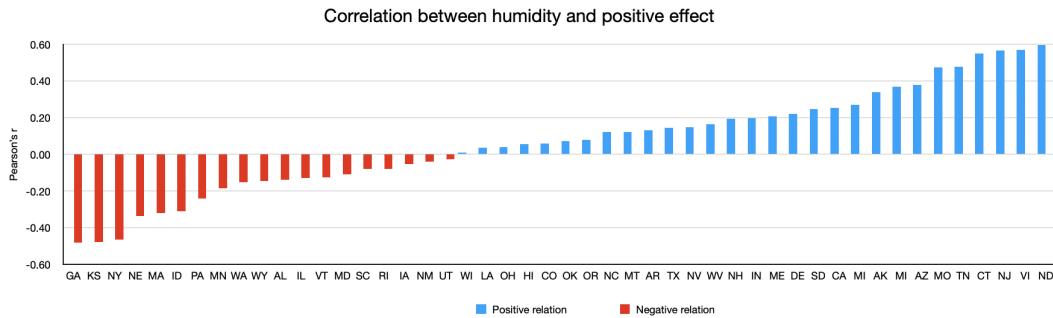Figure 9: correlation between relative humidity and positive sentiment

Figure 10: correlation between relative humidity and positive sentiment

7 CONCLUSION

For this project, I tried to find out if there is linear relationship between weather variables and positive mood of Twitter users and I do find out there is interesting relationship between them.

I found that the temperature is positively correlated with positive sentiment, while humidity is negatively correlated with positive sentiment across biggest cities of each state. However, when analyzing the relationship between the mean temperature, relative humidity, sea level pressure and positive mood over the 14 days in April, 2019, across all states, the relationship between humidity and positive sentiment shows a different trend, which requires more research to find out the reason. Furthermore, not all states show the same trends when analyzing the relationship between temperature and sea level pressure, which suggests further investigation into how meteorological factors may combine with others to influence moodPark et al. (2013). **e.g.** People live in Nevada do have better mood when relative humidity is high.

While my work do find out a correlation between some weather variables and sentiment of Twitter users, this work is just a small step toward the final conclusion, and has shortcomings: Firstly, the observation period was too short(only half a month on April, 2019), Secondly, the sentiment data is only generated from Twitter data, which does not include sentiment data from other social platforms, Thirdly, the quantitative analysis was limited to evaluating linear relationships between weather variables and sentiment data, while more sophisticated multivariate analyses would be a better choice to find out the correlation between weather and mood.
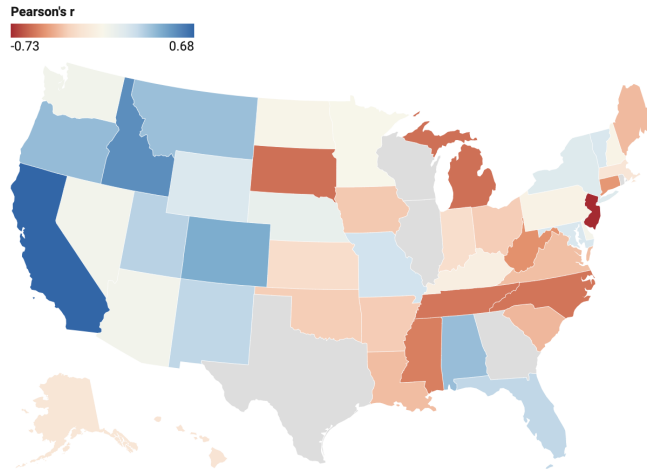
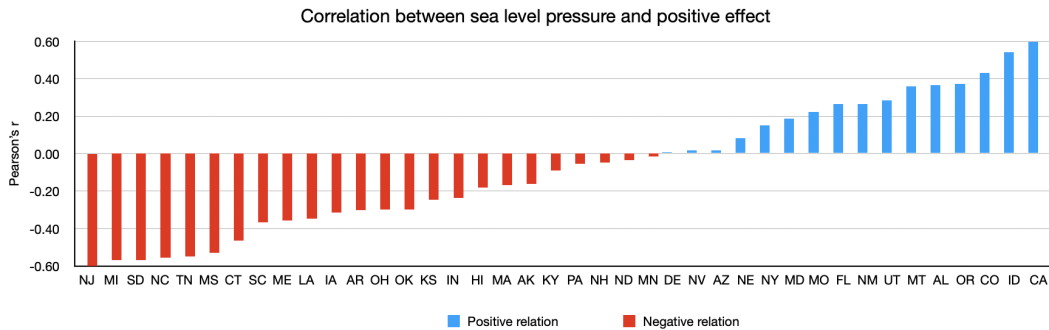Figure 11: correlation between sea level pressure and positive sentiment

Figure 12: correlation between sea level pressure and positive sentiment

REFERENCES

S. Anil Phand and V. A. Chakkarwar. Enhanced sentiment classification using geo location tweets. In *2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT)*, pp. 881–886, 2018. doi: 10.1109/ICICCT.2018.8473048.

Cameron Beccario. Earth. URL `https://earth.nullschool.net/#current/wind/surface/level/orthographic`.

Shruti Daggumati, Igor Soares, Jieting Wu, David Cao, Hongfeng Yu, and Jun Wang. Tweether: A visualization tool displaying correlation of weather to tweets. In *Visualization and Data Analysis*, 2016.

Z. Doshi, S. Nadkarni, K. Ajmera, and N. Shah. Tweeranalyzer: Twitter trend detection and visualization. In *2017 International Conference on Computing, Communication, Control and Automation (ICCUBEA)*, pp. 1–6, 2017. doi: 10.1109/ICCUBEA.2017.8463951.

JustAnotherArchivist. A social networking service scraper in python. URL `https://github.com/JustAnotherArchivist/snscrape`.

Ephrat Livni. The data that prove bad weather alters your mood. URL `https://qz.com/1088037/`.

Kyung-Soon Park, Soogeun Lee, Eunjeo Kim, Jeeyeon Park, and M. Cha. Mood and weather: Feeling the heat? *Proceedings of the 7th International Conference on Weblogs and Social Media, ICWSM 2013*, pp. 709–712, 01 2013.

S. A. Phand and J. A. Phand. Twitter sentiment classification using stanford nlp. In *2017 1st International Conference on Intelligent Systems and Information Management (ICISIM)*, pp. 1–5, 2017. doi: 10.1109/ICISIM.2017.8122138.

P. S. Quinan and M. Meyer. Visually comparing weather features in forecasts. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):389–398, 2016. doi: 10.1109/TVCG.2015.2467754.

Paolo Ripamonti. Twitter sentiment analysis. URL https://www.kaggle.com/paoloripamonti/twitter-sentiment-analysis/comments.