WeathR: Twitter Perspectives on the Weather

**Abstract**

Twitter is an information product generated by its users. We can use tweets to examine the relationship between what users post – their perceptions of the world around them – and physical conditions in the world. By collecting both tweets containing weather words and the true temperatures in the Twitter users’ locations, we can better understand how people in different cities perceive seasons, weather, and temperature. Through visualization of both tweets and meteorological data we found that traditionally warmer cities tweet more about being cold than traditionally cold cities, even though the actual temperature in the warmer city is often much higher than the temperature in the colder city. We conclude that users are more likely to tweet about things that break from their normal environment.

**Background and significance**

Twitter has become a data source for data scientists working on projects that cover a wide range of applications, including several related to natural phenomena. Because Twitter provides individuals’ reactions to different events in real-time and often with location information attached, it has become a useful primary data source. This spatial-temporal information has been used to analyze the perceptions and reactions to natural disasters such as forest fires [1], floods [2], and earthquakes [3]. The analysis of tweets has also been used to augment weather forecasts [4] and disaster readiness programs [5].

Perception of temperature is to some extent relative. How do people in different geographic areas of the country react to weather and the change of the seasons? By combining the frequency of weather and seasonal terms on Twitter with meteorological measurements of the weather in the places where the tweets originated, we can compare how people in different parts of the United States perceive their weather. We want to answer questions such as: What does “cold” mean to people in Minnesota versus those in California? We would also like to explore how we could use tweets to follow and visualize the general flow of heat and cold waves across the United States.

**Methods**

 *Data Collection*

 We monitored tweets coming from nine cities with varying climates: Atlanta, GA; Chicago, IL; Denver, CO; Los Angeles, CA; Minneapolis, MN; New Orleans, LA; New York City, NY; San Antonio, TX; and Seattle, WA. We also collected tweets that include weather- and season-related search terms such as leaves, hot chocolate, cold, chilly, snuggie, et cetera. A tweet-cleaning R function threw away the geo-located tweets (from the nine cities) that do not mention any of our terms. Each member of our team was in charge of three cities. For example, every three days the team member in charge of Chicago, Minneapolis and New Orleans would set three consecutive loops to run: one to collect tweets from Chicago and save to a spreadsheet, one to do the same for Minneapolis, and one to do the same for New Orleans. Each of these loops takes about a day to run. Unfortunately, many Twitter users do not enable geo-location, so we collected many tweets that contained one of our weather-related search terms without knowing if they were in one of our nine cities.

In addition to Twitter data, we required weather data. A team from the California Soil Resource Lab wrote an R function to access Wunderground.com data. We modified this freely available code to suit our needs. For each city and day within our range of tweet- and weather-collecting dates, we used this modified function to return the day’s average temperature.

*Variables*

 Our final data are tweets containing at least one of our weather-related search terms, with some geo-located as being in one of our nine cities. In addition, for every day on which we collected tweets (November 1, 2013 – December 1, 2013), we have the day’s average temperature in each of our nine cities. We tallied each weather term in each city on each day to determine the most frequently tweeted term per day, per city.

*Analytic Methods*

 We created a Shiny app, a web application made in R, to visualize the data. We made all of work accessible via Github.There are four tabs on the application: 1) “Map,” to view the most frequently tweeted weather-related term and average temperature for three cities on a given day as a slideshow; 2) “Time Series-Frequency,” a time-series to view the changing tweet frequency of a given weather-related term (cumulative across all nine cities) with a drop down menu to choose a search term; 3) “Time Series-Frequency vs. Temperature,” a time-series to view both the changing tweet frequency of a given weather-related term and the changing daily average temperature for a given city with menus to choose the term and location; and 4) “Random Tweets,” an interactive map where you can hover your mouse above a data point to view a random tweet from that location.

**Results**

On the “Map” tab, as the application plays, thirty days' worth of information appears on the map. Popular terms in all cities on any given day are common terms like “cold”, “leaves” and “wind”, though as November comes to a close, several cities' most frequent term on a given day is “Thanksgiving”. On the “Time Series-Frequency” tab, words like “cold”, “hot chocolate”, “holiday” and “Thanksgiving” peak decisively at the cold, bitter end of November. Others fluctuate; it would be interesting to see what weather events occurred when “storm” mentions crested in mid-November. Strangely, “pumpkin”s highest-frequency day occurs on November 1, and mentions of “warm” do not seem to ebb, even as we approach December! On the “Time-Series Frequency vs. Temperature” tab the cities with the coldest weather do not also have the most mentions of “cold”. On the “Random Tweets” tab we can check the raw data for context to make sure the weather mentions are actually being made in reference to the weather and not as part of, or describing, something like a song lyric.

Let’s look at descriptive visuals not included on our Shiny. The left figure below shows the average temperature per day in LA (red) and Minneapolis (blue). The words represent the most frequently tweeted word on that day in that location. We can see that “cold” is tweeted more frequently in LA than in Minneapolis even though the temperature is about 20 degrees warmer. Minneapolis is more concerned with the rain. The right figure below shows the distribution of weather word tweets that we collected overall. We can see that “cold”, “rain”, and “snow” are the most frequently tweeted words while words like “snuggie” and “earmuffs” were not tweeted at all in our cities during our time frame.



**Discussion/Conclusions**

We found expected and unexpected relationships between words, weather, and location. When tweeting about the weather, those in the coldest cities do not necessarily tweet the most about “chilly weather” terms, which we intuitively expected. Instead we see that in cities that are less used to the cold weather (e.g. LA), users tweet more about cold weather terms even though they have higher temperatures than other more hardy cities (e.g. Minneapolis). The most frequently tweeted weather word per day in LA is “cold” on almost every day, but in Minneapolis where the temperature is about 20 degrees colder, “cozy” is a most frequently tweeted word in mid-November. The data we collected suggests that different cities have different relationships with the seasons and the passing of time.

Because social media provides users with a venue to express their thoughts, feelings and ideas in real time, we have access to data that would not be otherwise available. The National Oceanic and Atmospheric Administration (NOAA) has recognized the utility of Twitter as a data source and has recently awarded grants for research about how to better facilitate the sharing of weather updates on Twitter and how to better manage a weather emergency (e.g. how to get users to respond to warnings to seek shelter) by examining trends in weather word tweets [6]. Our work suggests that users in cities that experience more extreme weather are less likely to tweet about it, as they are accustomed to their local weather. Therefore, NOAA would benefit from targeting cities that frequently have weather emergencies and placing more emphasis on facilitating weather updates in these places as locations with less experience with weather emergencies will create Twitter chatter on their own. In the future, it would be nice to collect more geo-located tweets over a longer period of time and quantify our results.

**References**

1. Bertrand De Longueville, Robin S. Smith, and Gianluca Luraschi. 2009. "OMG, from here, I can see the flames!": a use case of mining location based social networks to acquire spatio-temporal data on forest fires. In *Proceedings of the 2009 International Workshop on Location Based Social Networks* (LBSN '09). ACM, New York, NY, USA, 73-80.<http://doi.acm.org/10.1145/1629890.1629907>

2. Sarah Vieweg, Amanda L. Hughes, Kate Starbird, and Leysia Palen. 2010. Microblogging during two natural hazards events: what twitter may contribute to situational awareness. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '10). ACM, New York, NY, USA, 1079-1088.<http://doi.acm.org/10.1145/1753326.1753486>

3. Takeshi Sakaki, Makoto Okazaki, and Yutaka Matsuo. 2010. Earthquake shakes Twitter users: real-time event detection by social sensors. In *Proceedings of the 19th international conference on World wide web* (WWW '10). ACM, New York, NY, USA, 851-860.<http://doi.acm.org/10.1145/1772690.1772777>

4. Jeff Cox, Beth Plale. 2011. Improving Automatic Weather Observations with the Public Twitter Stream. *Technical Report*. Indiana University, School of Informatics and Computing. <http://www.cs.indiana.edu/cgi-bin/techreports/TRNNN.cgi?trnum=TR691>

5. Amanda Lee Hughes, Leysia Palen. 2010. Twitter adoption and use in mass convergence and emergency events. *International Journal of Emergency Management.* Vol 6, Number 3-4/2009, 248-260.<http://www.inderscience.com/info/inarticle.php?artid=31564>

6. NOAA. 2012. New NOAA awards to fund studies of weather warnings, social media, Internet tools and public response.<http://www.noaanews.noaa.gov/stories2012/20120827_oarsocalscienceawards.html>

Appendix



This is the first tab of our Shiny application. The map changes like a slide show to go through each day, showing the average daily temperature and the most frequently tweeted weather word. We can see on the first of November, LA is talking about it being cold even when the average temperature on that day is about 70 degrees.



This is the second tab of the Shiny app. Here we see the frequency of “leaves” tweets across all locations during the month of November. The frequency fluctuates across the month of November.



Here is the third tab of our Shiny application. We can see the frequency of “rain” tweets in Seattle and compare it to the average daily temperature in that day. The temperature and tweet frequency seem to peak at about the same time on the 12th of November.



Here is an example of the Random Tweet tab. This Twitter user in Iowa wishes she was home with her electric blanket.



Here we can compare more cities in the same way that we compared Los Angeles and Minneapolis. Chicago seems to be less hearty than Minneapolis; users frequently tweet about being cold, and San Antonio seems to be heartier than LA, as users tweet mostly about leaves. It is interesting to see the temperatures in Atlanta and New York City overlap. Atlanta most frequently tweets about “leaves”, but this word could actually refer to the verb rather than the noun. Atlanta has the most air traffic in its major airport, so these mentions of “leaves” could be related to this.



Here we wanted to test the hypothesis that frequency of “warm” tweets and “cold” tweets move together as people are tweeting about being cold but also tweeting about wishing that it were warmer. Here we see evidence to support our hypothesis. However, between the 10th and 20th, the correlation breaks down.



Here we compare all of the cities to see where “cold” was most frequent. These visuals display the average temperature of each city, and the dots mark the days on which “cold” was the most frequently mentioned term in each city. We can see (and intuitively predict) that this does not occur on the same day for each city; however, the most “cold” tweets also did not occur on the coldest day.



We can see a similar relationship for “Thanksgiving”. The following visual depicts when “Thanksgiving” was the most frequently mentioned term for each city, which did not occur on the same day, and not always on November 28th, Thanksgiving.