Comparison between Keywords from Social Media and Actual Weather Data

Zani Dembure

Earth and Environmental Science Department, Furman University, Greenville, SC 29613

Abstract

Since the birth of social media in 1997, billions of active users around the world uses some sort of social network on a daily basis to either connect with one another, engage in news content, or share information that can provide real time data for everyday use. Some of this data can be information on real-time happenings with nearby traffic, time and location of local sports, or even reviews on the best eateries in the area. This project focuses on the use of targeted keywords based on the temperature of weather on Twitter and compares the results with actual weather data to test whether social media can be reliable for this type of information. The tweets were gathered on March 14 and 15, primarily in South Carolina, in order to provide enough data for analysis. The keywords that were used the most were 'chilly', 'cold', 'snow', 'cool', 'cooler', and 'freezing'. Since there wasn't enough tweets from each county of South Carolina, the ones with the most tweets on local weather were selected. Weather station data was used to create a map of the actual weather data from the two days. The results show that the keywords were used appropriately for the selected counties compared to the actual weather data. The findings show that temperatures of the weather station map match to the tweets description of the area. Although there were positive results on comparing tweet data and weather data, tweets can still be proven to be unreliable at times in the form of misinformation or false analysis.

Lit Review

An article from the gisuser website talks about the benefits of using the Twitter feed along with a map to provide real-time water, weather and flood forecasting for the state of Texas. The map is named the USGS Texas Water Dashboard which can be accessed anywhere from desktop, smartphone or other mobile devices. The article addresses the benefits of having social media in the events of an emergency by relating to the 2015 flooding in Texas when many people didn't have power and had to rely on social media to get the latest information on the current water conditions. Feeds like USGS TX Floodwatch and USGS TX Rainwatch gave water level and precipitation data within minutes of any drastic change in weather. The site also claims that this information can prove to be useful for water managers in making crucial decisions about the local resources.

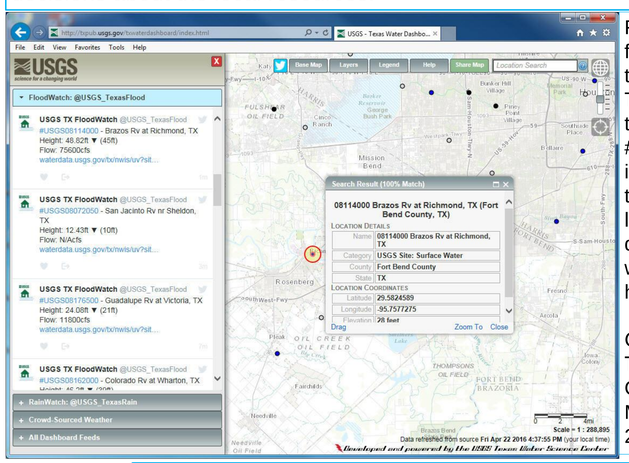
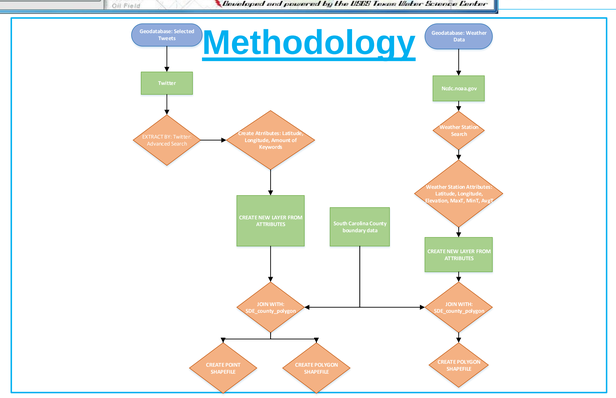
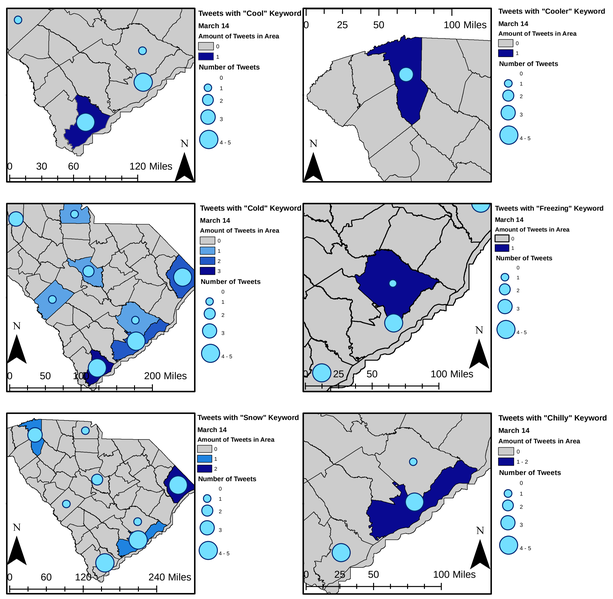


Figure 1: Twitter feed showing tweets from USGS TX Floodwatch and the hashtag, #nmwx. To the right is a map of Texas that has selectable layers and point data regarding weather and natural hazards.

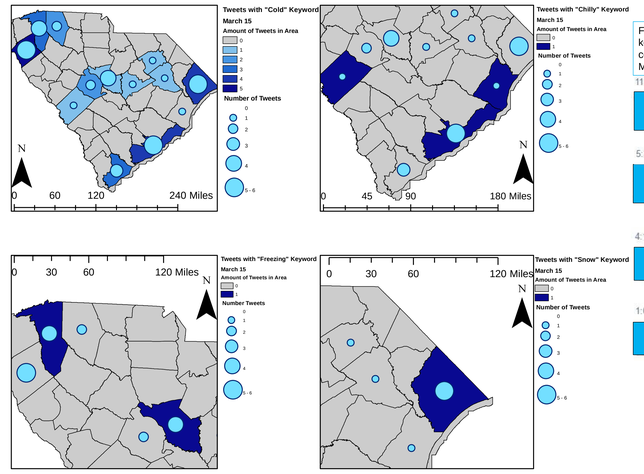
GISUSER. "USGS TX FloodWatch." GISUSER, Spatial Media LLC, 2003-2015

Methodology





Tweets Collected from the Following Counties: Greenville, York, Richland, Aiken, Berkeley, Charleston, Beaufort, and Horry



Tweets Collected from the Following Counties: Spartanburg, Greenville, Anderson, Aiken, Lexington, Richland, Sumter, Darlington, Florence, Horry, Georgetown, Charleston, and Beaufort

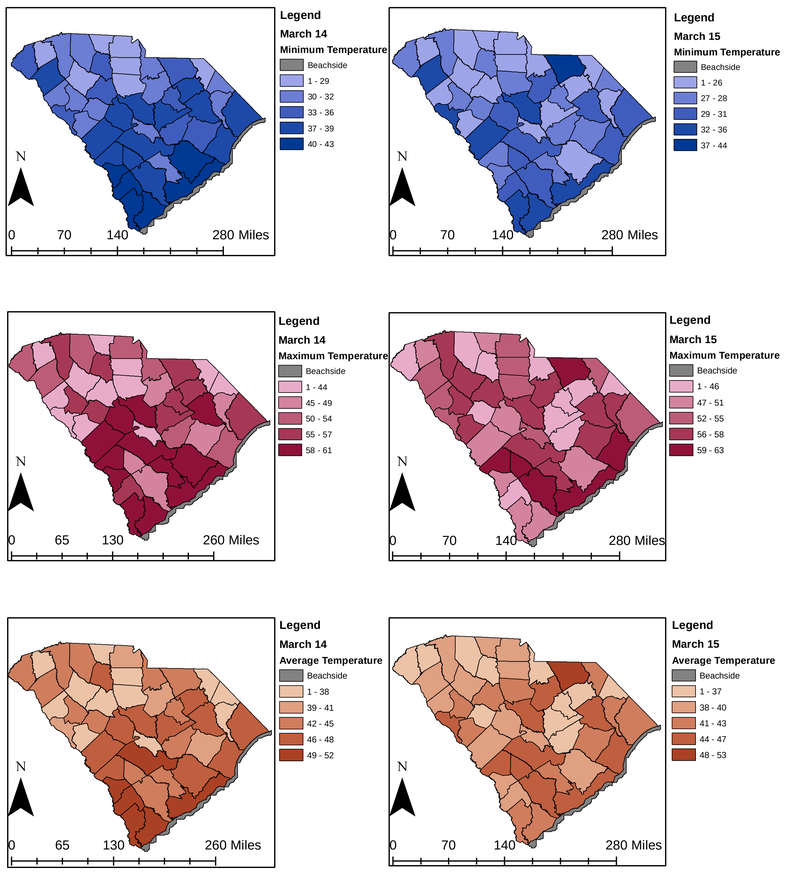
Figures 2-5: Examples of selected tweets based by keywords like cold and sun. These tweets were collected from Myrtle Beach and Hilton Head SC on March 14 and 15.

11:35 AM - 14 Mar 2017 from Hilton Head Island, SC
Mar 14
it was **cold** @ Hilton Head Island, South Carolina
instagram.com/p/BRoPiGcBtl6/

5:35 AM - 15 Mar 2017 from Hilton Head Island, SC
Mar 15
☀Good morning ☀ it may be **cold**, the **sun** is still shining on Hilton Head 🌴 #happywednesday... instagram.com/p/BRqLIqzjBdv/

4:11 PM - 14 Mar 2017 from Myrtle Beach, SC
Mar 14
Cold but pretty. @ Myrtle Beach, South Carolina
instagram.com/p/BRovJHWD2Dig...

1:07 PM - 15 Mar 2017 from Myrtle Beach, SC
Mar 15
It's **cold** but you can't beat the view 😎 @ Palace Oceanfront Resort
instagram.com/p/BRq-2qrjcXk/



Minimum, Maximum, and Average Temperatures for South Carolina on March 14 and 15

Conclusion

The results of this project shows how social media like Twitter can play a role by giving somewhat reliable data to locals like the current weather status. Due to the weather being colder than usual in South Carolina during these two days in March, considerable feedback was made from the local Twitter users on the unusual change in temperature. One of the disadvantages of relying on social media data is that there might not be any tweets that day regarding weather or there are few tweets in the area with conflicting interests. Although the data from the weather stations themselves are more reliable and accurate, it is still useful to have backup sources like the social media to count on during times of need.

References and Data Sources

Acknowledgements: I would like to thank Dr. Suresh and Mike Winiski for helping me with my project.

"Advanced search." *Twitter*, 2017, www.twitter.com/search-advanced/.

"Data Tools: Find a Station." *NOAA*, 2017, www.ncdc.noaa.gov/cdo-web/datatools/findstation.

"DLG County Boundary 1:24,000 scale." DNR, 2015, www.dnr.sc.gov/GIS/gisdnrdata.html.

GISUSER. "Twitter feed and map provide real-time water, weather and flood forecasting for Texas." *GISUSER*, 27 May 2016, www.gisuser.com/2016/05/twitter-feed-and-map-provide-real-time-water-weather-and-flood-forecasting-for-texas/.

FURMAN