Mapping Twitter Sentiment
An Exploration of Open Source Web GIS

Josiah Parry
GE 4280 GIS Applications, Dr. Hyun Joong Kim

OVERVIEW

- The advent of social media has enabled social discourse to transcend physical interactions. As a globalizing society, we experience and partake in discourse online. Our fervent Facebook posts and our biting quips on Twitter provide a snapshot into spatiotemporal variation of human emotion and response. I have come to view social media as the largest corpus of data on human behavior and interaction. I seek to explore, uncover, and interpret sentiment, be it about public policy, sport, television, or what have you.
- Following, I shall describe an exploration that has become a bite size application that explores the sentiment of Twitter users in the largest cities in the united states. The application enables the user to enter a search phrase, select a number of cities, and visualize the aggregate sentiment of tweets about the entered phrase in the presented cities.

DATA

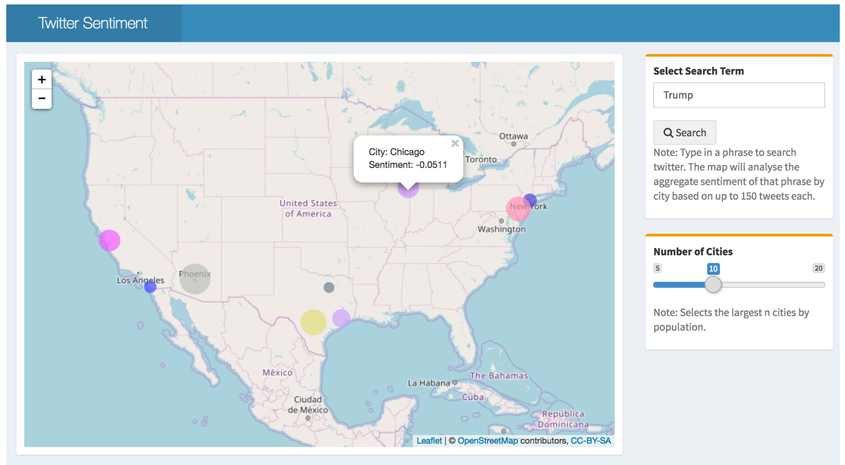
- The US cities shapefile used was first downloaded from ESRI's ArcGIS online *USA Major Cities* group layer (ESRI, U.S. Census Bureau). The layer displays "the locations of cities within the United States with populations of approximately 10,000 or greater, all state capitals, and the national capital," (ESRI). This file was then imported into ArcMap and then converted into a shapefile to be used within R (statistical computing language).
- Tweets are gathered using the Twitter API via *rtweet*

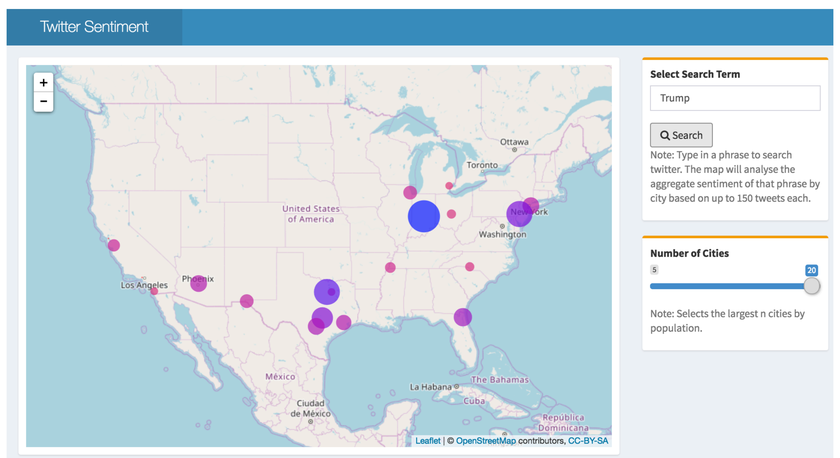
RESULTS

	name	tweet	avg_sent
1	Los Angeles	@_Deplorable_Me @JohnFugelsang @POTUS Trump has done MORE golfing, more tweeting more attacking US Citizens, more attacking our allies, more kissing Putin's ass, more treason that Benedict Arnold, more lining his own pockets w taxpayer $$$, more lying	-0.0734871426676328
2	San Diego	@Bumperstik @NancyPelosi @realDonaldTrump hahahahhaha! Byrd renounced his ties to KKK! HAIL TRUMP! https://t.co/n7HAtmH4s4	-0.0586252531988911
3	San Francisco	@quanta_chaos @ColdQuo @jsheppard18 @KayaJones I am not paid to defend anyone. I am not even saying that sin is right in any way. I support Trump because he is worthy of my support. I do not agree with every of his policies, but I agree with most of them and I love that he is fixing the damage caused by the previous presidnts	-0.0785060775336489
4	San Jose	Sen. Gillibrand Responds To Donald Trump's Vulgar Tweet: "You Cannot Silence Me" https://t.co/l6UNHQ1ap4 via @deadline	-0.0499569556852556
5	Phoenix	Perhaps they were discussing the polls that said the word most associated with Trump is "idiot." https://t.co/2P9IZe1qij	-0.0439892311661609
6	El Paso	Gowdy mode activated. He brought up the anti-trump text messages. #Rosenstein https://t.co/ToX4Vt5rxj	-0.048959744451893
7	Austin	@kazmc39 @RobinTweets247 @itchstdnts @realDonaldTrump @daveweigel @theresa_may Oh, I'm doing great! Jones won....It's only a matter of time before your beloved Trump is out of the oval office! Lol. Who cares who started a tweet, you've been absent. Start a Trump supporting tweet then leave? Oh yeah, look at all the other republicans fleeing! Lol. I get it	-0.0328661335880481
8	Dallas	@konoboe Trump is a racist. That is why he is terrified.	-0.0535136965925677
9	Fort Worth	@ananavarro ED MARTIN IS ONE OF THE RUDEST MOST UNPROFESSIONAL CONDESCENDONG SMUG HOLIER-THAN-THOU DISGUSTING INDIVIDUALS AROUND. Thus, the PERFECT Trump et al surrogate! ... though sorry Ana you even have to associate with such human-underbelly.	-0.0292750447394807
10	Houston	President Trump: America 'Cannot Afford' for Roy Moore to Lose Alabama Senate Election https://t.co/q4Ti28U3O8	-0.0434455009951939

METHODS

- Project was generated using R and open source packages entirely
- Made interactive using *shiny* and *shinydashboard*
- Shapefile collected by US Census and filtered to 20 largest cities in the US by population
- Twitter API is searched for a provided term within 10 miles of the largest cities
  - Up to 150 tweets are collected (retweets are excluded)
  - Utilized *rtweet*
- Each tweet has a sentiment score assigned to it
  - Scores are generated using the *sentimentr*
- Sentiment is aggregated by each city and averaged
- Each city circle size is the absolute sentiment score * 10^5.2
- Circle color is denoted on a continuous scale from minimum average sentiment observed to maximum average sentiment.
  - Spatial objects are manipulated using *sf* and visualized with *leaflet*





METHODS VISUALIZED

```
cities <- st_read("cities.shp") %>%
 select(OBJECTID, NAME, geometry, POPULATION) %>%
 rename(name = NAME, CID = OBJECTID, pop = POPULATION)


function(input, output, session) {

 update_cities <- eventReactive(input$search, {
  search_term <- input$search_terms
  input$n_cities
  temp <- cities %>%
   top_n(input$n_cities, pop) %>%
   mutate(query = paste(geometry[]))

  # Create a character vector will house each lat long pair
  # Initialize vector
  search_xy <- character(5)
  for (i in 1:nrow(temp)) {
   search_xy[i] <- paste0(temp$geometry[[i]][2],",", temp$geometry[[i]][1],",10mi")
  }
  # Bind the query vector to the cities internal dataframe
  temp$query <- search_xy

  city <- temp
  tweets <- map_df(city$query, function (x) {search_tweets(search_term, n = 150, geocode = x, lang = "en", include_rts =
FALSE) %>% mutate(query = x)})

  city_tweets <- left_join(tweets, city, by = "query") %>%
   select(screen_name, text, retweet_count,
          favorite_count, CID, geometry, name) %>%
   rename(city = name) %>%
   mutate(id = row_number()) # Create unique identifier
```

CONCLUSIONS

- The development of interactive and open source geospatial research requires knowledge of not only GIS methods, but about the underlying aspects of each tool used, user design, and experience. This research project was a first foray into developing geospatial products using the open source statistical programming language R.
- The nature of these tools seem to be anarchic and unruly. However, these tools are open to public scrutiny and enable reproducible content. This project runs using two R scripts that are currently publicly available on Github (https://github.com/JosiahParry/twitter_sentiment_map). Unlike with standard ESRI products, analyses can be followed through logically with a single reading of the script (assuming one understands the language). I believe this is where the strength of the project lays.

RESOURCES

- Use the applcication at josiahparry.shinyapps.io/twitter_sentiment
- Find the project on Github: github.com/josiahparry/twitter_sentiment_map
- Learn R:
  - DataCamp.com
- Text Analysis:
  - Tidy Text Mining : Julia Silge & David Robinson
  - tidytextmining.com
  - *sentimentr: github.com/trinker/sentimentr*