Social Media data analysis with R

Data Inquiry Lab Winter 2017. Instructor: Whitt Kilburn, kilburnw@mail.gvsu.edu

Hi! Welcome to Social Media Data Analysis with R. We'll start with accessing Twitter data. We'll briefly review a few other capabilities for social media in R. To do so, you'll need to take care of a few preliminaries. First, setup a Twitter account. Then:

1. Register to access Twitter's API. After setting up your account, log in and go to: `https://apps.twitter.com` It will say "You don't currently have any Twitter Apps."

2. So click 'Create new App'. Fill out the form on the following page — name, website (use your GVSU personal page or something similar), and provide a brief description of it. Remember these fields could be for someone trying to develop, for example, iphone Twitter apps, so Twitter asks about a 'Callback URL'. Just leave that field blank. For my name, I just used my Twitter handle, and for description I typed "use of twitter for investigations".

3. Next agree to the terms and conditions, click "Yes, I agree..." and then click "Create your Twitter Application".

4. On the next page, click the tab marked "Keys and Access Tokens". You need to save a record of four keys and tokens.

5. I recommend saving each piece of info in a plain text file, organized as follows:

    (a) consumer_key $<-$ "YOUR CONSUMER KEY HERE"
    (b) consumer_secret $<-$ "YOUR CONSUMER SECRET HERE"

6. Next, you need an "Access Token". At the bottom of the screen, click "Create my access token". Then record the following:

    (a) access_token $<-$ "YOUR ACCESS TOKEN HERE"
    (b) access_secret $<-$ "YOUR ACCESS SECRET HERE"

7. Make sure you copy-paste the entire key/token. Once you have these four pieces of info recorded you can close your Twitter apps page.

What will we do in the workshop?

We'll spend most of our time on Twitter. We'll use R packages for interacting with Twitter's API and pulling data. We'll collect tweets as a corpus and visualize word elements of it. We'll do a sentiment analysis of tweets.

Why use R packages for analyzing social media data?

Pulling data off of social media networks typically involves the use of the network's API, which requires both coding time and skill. The social media oriented packages in R are intended to do this heavy lifting for us, which is one of the upsides of using R packages for social media data analysis. The packages provide a relatively easy interface with social media data, without requiring a lot of coding. And we can

structure the data we pull into formats amenable to other analyses in R, such as a sentiment analysis of Tweets or a visualization using popular R graphics packages. The main downside, however, is that we are limited to the functionality of the R package. Fortunately, however, most of the packages provide great functionality. I'll provide an R script file to get us started and point you toward additional resources.

A few things to know about Twitter API limits

There's a time limit of working through the Twitter API. Finding tweets on a specific hashtag or key word can't be collected in their entirety — beginning with the first time someone ever used it. (At least not without paying Twitter for the data, I guess.) But looking to the future, however, it is possible to record tweets over a period of time and save these tweets to analyze later.

Twitter limits the number of calls (requests for data) to Twitter over a 15 minute window. There are different limits for different types of requests. See `https://dev.twitter.com/rest/public/rate-limits`. There's lots to read over here, and I've just skimmed it. From what I have read, it appears the use of targeted searches results in all recent tweets (within search parameters) that match, given your API limit. How these are selected when more tweets exist that can fit within the API limit, I don't know.

Resources for further study

The library has e-books directly relevant to social media data analysis. First, for an introduction to R, try *R for Data Science* by Hadley Wickham. `http://r4ds.had.co.nz/` And then for applying text mining methods to social media, *Text Mining with R* by Julia Silge and David Robinson, to be published in May. Finally, there's *Social Media Mining with R* by Nathan Danneman and Richard Heimann. This last text discusses theories of data analysis in the context of social media and presents some examples.

Other social media R packages

instaR `https://cran.r-project.org/web/packages/instaR/instaR.pdf` allows access to Instagram profile data, and automatic downloading of images given various criteria such as hashtag or location. Google "Instagram R package" for a few examples.

SocialMediaLab This tool is perhaps a bit overhyped as the-only-tool-you'll-ever-need for social media. Very useful for tracking website sharing over social media. Lots of other functionality on different social media sites. See `https://www.hastac.org/blogs/herrcafe/2014/07/01/introducing-socialmediaminer-social-media-tool-r` And see `https://cran.r-project.org/web/packages/SocialMediaLab/index.html` A beginner's guide to it: `https://goo.gl/bGCjBh`

RFacebook "This package provides a series of functions that allow R users to access Facebooks API to get information about users and posts, and collect public status updates that mention specific keywords." `https://cran.r-project.org/web/packages/Rfacebook/Rfacebook.pdf`.

New packages are introduced continually and existing packages are updated to adjust for changes in social media. A user-friendly guide to R packages is `https://rdrr.io/`.