Milestone_2.0

Daniel Smith

3/27/2019

```
knitr::opts_chunk$set(echo = TRUE)
#load needed packages
library(tidyverse)
```

```
## -- Attaching packages -------------------------------------------------------------------------------


## v ggplot2 3.1.0        v purrr   0.3.0
## v tibble  2.0.1        v dplyr   0.8.0.1
## v tidyr   0.8.2        v stringr 1.4.0
## v readr   1.3.1        v forcats 0.4.0


## -- Conflicts ----------------------------------------------------------------------------------------
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(twitteR)
```

```
##
## Attaching package: 'twitteR'


## The following objects are masked from 'package:dplyr':
##
##     id, location
```

Project Changes

Change of Project Direction and Justification

Originally, this project was meant to serve as an explanation of the twitter conversation occuring around the global epidemic of chronic kidney disease of unknown etiology (CKDu). In order to determine the feasibility of analyzing tweets related to this topic, an initial search on Twitter showed that there were people ad organizations tweeting about the epidemic and these tweets served as the basis for identifying keywords and hashtags to use for the colleciton of tweets. From the beggining, there was worry about having access to a sufficient number of these tweets to proivde a robust analysis. Given the fewer number of twitter users tweeting about CKDu and the inherent limitations of Twitter's API (i.e. only allowing free access to tweets in the last 7 days), the focus of this project has shifted to analyzing tweets related to climate change and health.

Change in Project Question 3.

Due to the general lack of geocoding of tweets, question 3 "Where are people from who are tweeting about the CKDu epidemic?" has been dropped. Instead, an analysis of the word network in the collected tweets will be done using the libraries `igraph` and `ggraph`.

Background

Data Acquisition

Data for this project was acquired from http://twitter.com and were acquired through Daniel Smith's API account, which has been connected to R below. All tweets were acquired on March 4, 2019.

```r
#Connect R to twitter.com with API information
#First, API information is entered into R as vectors.
api_key <- "sbUlW5AkRQtN9j9qj9BuW13tn"
api_secret <- "Oql2n57GRzfDmvUWYdOy9BJxq5kwPdK77KCOEKIxMOpOUO61gl"
access_token <- "906672440-UZYZCHBALNuo5Qt0Fg15bWH3AMq6WxJnaxqkTOrp"
access_token_secret <- "UfYImM6LN1PeXuJtPN6RIlK8Zz9mRQa1pHsDAf9bOuUbw"
#Second, connect R to twitter
setup_twitter_oauth(api_key, api_secret, access_token, access_token_secret)
```

```
## [1] "Using direct authentication"
```

A mixture of key words/phraces and hashtags were used to query tweets about the CKDu epidemic. Keywords used in this search were "chronic kidney disease of unknown etiology", "enfermedad renal crónica no tradicional", "mesoamerican nephropathy", and "chronic interstitial nephritis in agricultural communities". Hashtags used in this search were #CKDu #CINAC #MeN and #ERCnt.

```r
#Search for tweets that contain the phrace "chronic kidney disease of unknown etilogy"
CKDu_keyword <- c(searchTwitter("chronic kidney disease of unknown etiology", n=5))
```

```
## Warning in doRppAPICall("search/tweets", n, params = params,
## retryOnRateLimit = retryOnRateLimit, : 5 tweets were requested but the API
## can only return 0
```

```r
ERCnt_keyword <- c(searchTwitter("enfermedad renal crónica no tradicional", n=5))
```

```
## Warning in doRppAPICall("search/tweets", n, params = params,
## retryOnRateLimit = retryOnRateLimit, : 5 tweets were requested but the API
## can only return 0
```

```r
MeN_keyword <- c(searchTwitter("mesoamerican nephropathy", n=5))
```

```
## Warning in doRppAPICall("search/tweets", n, params = params,
## retryOnRateLimit = retryOnRateLimit, : 5 tweets were requested but the API
## can only return 2
```

```r
CINAC_keyword <- c(searchTwitter("chronic interstitial nephritis in agricultural communities", n=5))
```

```
## Warning in doRppAPICall("search/tweets", n, params = params,
## retryOnRateLimit = retryOnRateLimit, : 5 tweets were requested but the API
## can only return 0
```

```r
Climate_Change <- c(searchTwitter("climate change+health", n=5))
#Search for tweets that contain the hashtags of interest for this project.
CKDu_hashtag <- c(searchTwitter("#CKDu", n=5))
ERCnt_hashtag <- c(searchTwitter("#ERCnt", n=5))
```

```
## Warning in doRppAPICall("search/tweets", n, params = params,
## retryOnRateLimit = retryOnRateLimit, : 5 tweets were requested but the API
## can only return 0
```

```r
CINAC_hashtag <- c(searchTwitter("#CINAC", n=5))
```

```
## Warning in doRppAPICall("search/tweets", n, params = params,
## retryOnRateLimit = retryOnRateLimit, : 5 tweets were requested but the API
## can only return 4
```

```r
MeN_hashtag <- c(searchTwitter("#MeN", n=5))
#make tweets a data frame.
Climate_Change.df <-twListToDF(Climate_Change)
#print data frame
Climate_Change.df
```

```
##
## 1 RT @sandrogalea: Time for a #HealthNewDeal. The #GreenNewDeal has animated the #climatechange deba
## 2 RT @MBNowlin: This makes me so mad I could spit nails! Such hypocrisy with climate change and publ
## 3  Researchers analyzed 2 billion social media posts; When record weather events occur consistently
## 4 @Puglaas @janephilpott @AndrewScheer From where I sit armed with nothing but my vote...health care
## 5          At UNDP in New York. Discussing HIV/AIDs, Health and Climate Change and mitigation.
##   favorited favoriteCount replyToSN            created truncated
## 1     FALSE             0     <NA> 2019-03-04 22:13:00     FALSE
## 2     FALSE             0     <NA> 2019-03-04 22:11:10     FALSE
## 3     FALSE             0     <NA> 2019-03-04 22:10:35      TRUE
## 4     FALSE             7 ltmcdies 2019-03-04 22:04:39      TRUE
## 5     FALSE             0     <NA> 2019-03-04 22:02:54     FALSE
##            replyToSID                  id replyToUID
## 1                <NA> 1102693449621483521       <NA>
## 2                <NA> 1102692990315827201       <NA>
## 3                <NA> 1102692844723044352       <NA>
## 4 1102689287986831360 1102691348115685376  104037382
## 5                <NA> 1102690908787609600       <NA>
##                                                                    statusSource
## 1   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
## 2 <a href="http://twitter.com/download/android" rel="nofollow">Twitter for Android</a>
## 3                <a href="http://twitter.com" rel="nofollow">Twitter Web Client</a>
## 4            <a href="https://mobile.twitter.com" rel="nofollow">Twitter Web App</a>
## 5   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
##        screenName retweetCount isRetweet retweeted longitude latitude
## 1      madorairene            2      TRUE     FALSE        NA       NA
## 2 ChrisLa83115764            1      TRUE     FALSE        NA       NA
## 3 _TheLastBastion            0     FALSE     FALSE        NA       NA
## 4        ltmcdies            0     FALSE     FALSE        NA       NA
## 5        DrIfereimi           0     FALSE     FALSE        NA       NA
```