/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\* README

\* Twitter Clustering project

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

Names: Aldrin Barraquio, Michael Christiansen,

Tanner Stirrat, Thao Nguyen

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\* List the names of your source files along with a brief description

\* of each. Your filenames may vary, but here is a list of my files:

\* func.R, clean.R, do.R,

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

1. combine.R: script to combine 2 data set - August and November

into 1 single data set.

* File inputs: tweets.Rdata, tweetsInd.Rdata
* File outputs: tweetsAll.rda

2. textMining.R: script for natural language processing,

creating tweets-vector data frame and frequency analysis

* File inputs: tweetsAll.rda
* File outputs: tweetsDF.rda
* Source code: func.R

3. distance.R: script to calculate distance and cluster tweets

* File inputs: tweetsDF.rda
* File outputs: tweetsClus.rda
* Source code: func.R

4. func.R: includes functions for textMining.R, distance.R and stackPlot.R

5. analysis.R: creates visualizations for the analysis

* File inputs: tweetsClus.rda

6. treemapanalysis.R : script to sort tweets into time zones and to construct a treemap

* File inputs: tweetsClus.rda

7. Split time.R : script to parse created\_at and create variables for stackPlot.R

* File inputs: tweetsClus.rda

8. stackPlot.R : script to create stackplots

* File inputs: tweetsClus.rda, Split time.R
* Source code: func.R

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Dependent R packages for the code:

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

treemap

stringR

tm

SnowballC

devtools

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\* List whatever help (if any) that you received. Include any

\* external resources and the names of any students that helped you.

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

Bags-of-word method: <https://www.edx.org/course/analytics-edge-mitx-15-071x-0#.VNGRx_7AKR8>

Stacked plot:

<http://stackoverflow.com/questions/13084998/streamgraphs-in-r>