

Gmail Package by Asha Elizabeth Gupta

STAT 290 Final Project, 15 March 2011

The purpose of this package is to retrieve emails between two people (the 'user', who is running the analysis, and a 'friend' of their choosing), and analyze the patterns of communication between the two people.

This package analyzes reply times, volume and frequency of emails, and words used in subject lines in emails between the user and the friend.

Outputs include:

- a word cloud made using words in the subject line. Common words (e.g. 'the', 'that', 'of', etc. are excluded)
- histograms that compare the number of emails sent vs. received by the user with the friend by day of the week, hour of day, and over time (from the start of the date range of emails, to the end of the date range)
- a jitter-plot of reply times in hours, by day of the week and hour of the day, along with the mean reply time for the user and friend
- a data.frame of the all emails fetched including the following parameters: MsgID, Date, To, From, Subject, ReplyID, Reply.Time, Refs

This package depends on the ggplot2, tm, png and snippets packages.

Note: This package requires python, and fetches emails using python's imaplib module. This also only works for gmail users. Gmail users will first have to enable IMAP. One can do this from their gmail account -> Email Settings --> Forwarding and POP/IMAP --> IMAP Access --> Enable

I hope to continue working on this package. Future versions will analyze the plain text body of the emails, provide statistics on emails to and from the user independent of a specific friend, and will explore some of the more interesting visualizations I saw in the ggplot2 documentation.