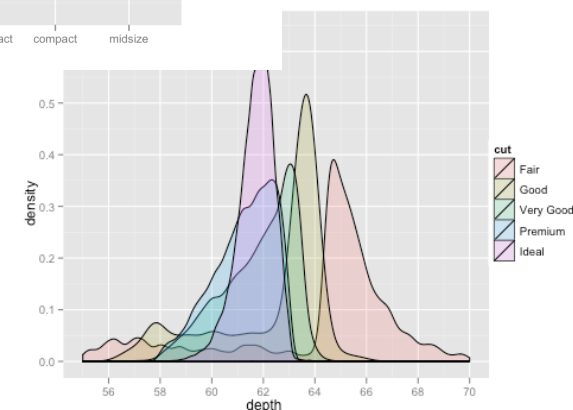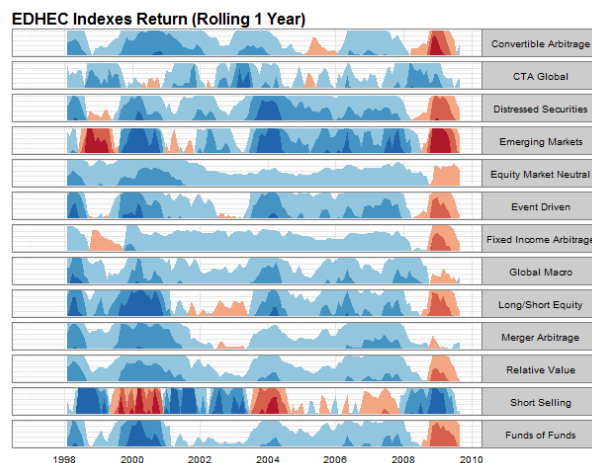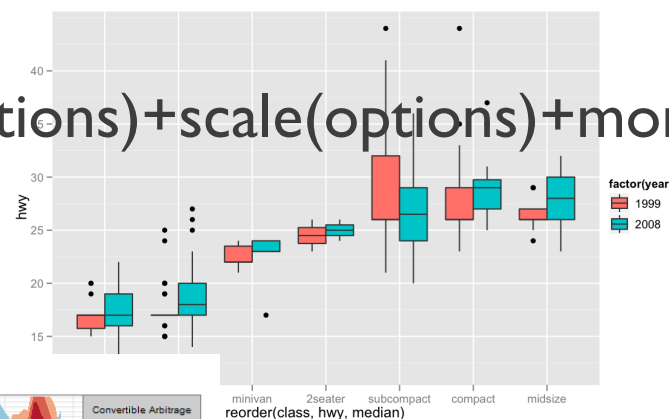# TEXT MINING AND VISUALIZATIONS IN R

ALLEN ROSS

WHAT IS TEXT MINING AND WHY DOES IT MATTER?

- Set of tools to extract information from groups of text data
  - Sources can be books, journals, newspapers, facebook, twitter, etc
  - Health related: clinic notes, doctor notes, patient free text, etc
    - https://www.ncbi.nlm.nih.gov/pubmed/?term=shah+n+lyalina
    - https://www.ncbi.nlm.nih.gov/pubmed/27626235
- Word count, trends, sentiment analysis, NLP

GGPLOT2

- Style of plotting popular in R

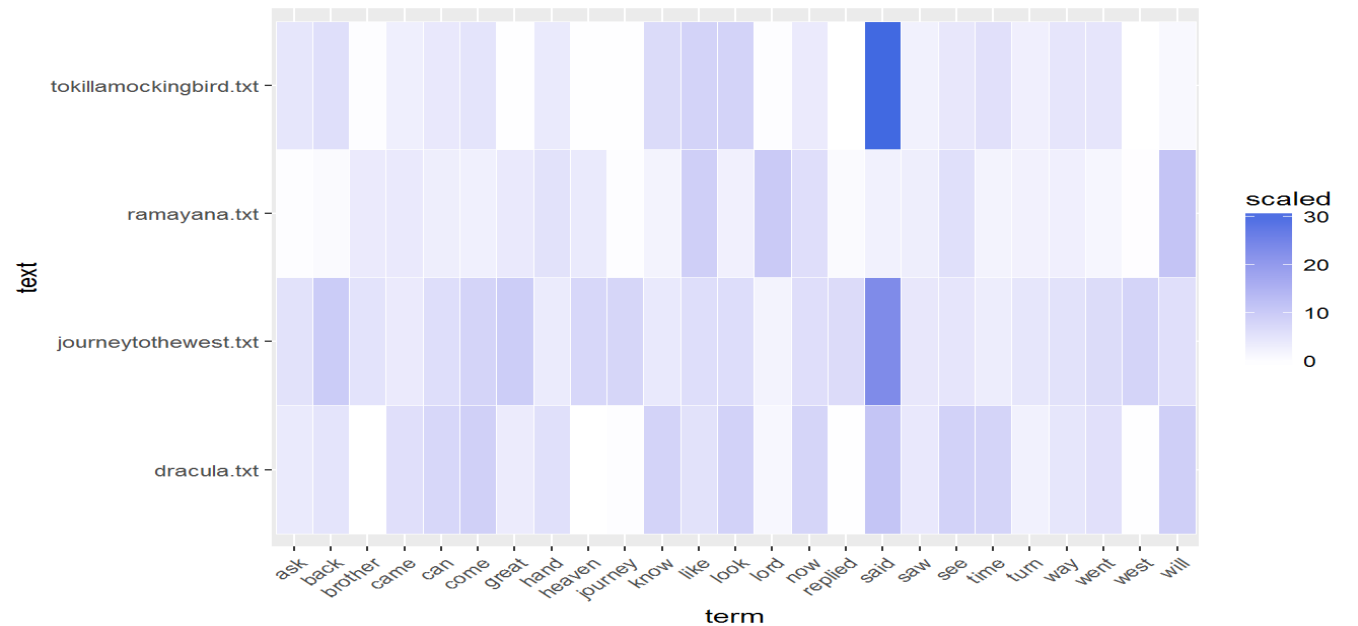- Format: ggplot(data)+geom_<plot type>(options)+scale(options)+more…

- Useful links:

  - http://www.cookbook-r.com/Graphs/

  - http://docs.ggplot2.org/current/

EXAMPLE 1

https://www.gutenberg.org/ebooks/

- Book Analysis
  - Journey to the West
  - Ramayana
  - Dracula
  - To Kill a Mockingbird
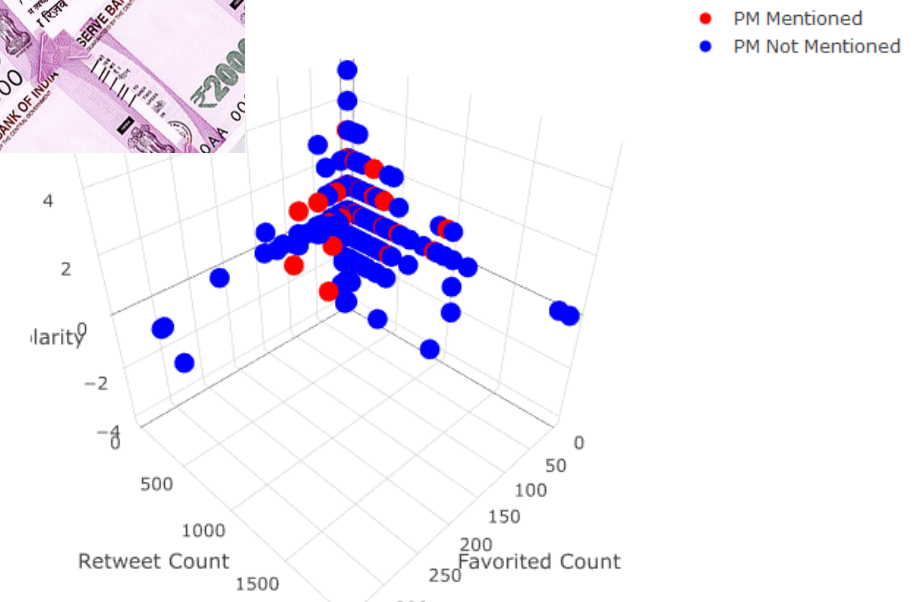
EXAMPLE 2

- Twitter Analysis
  - Demonetization in India
  - 8000 tweets directly following announcement
  - Link: https://www.kaggle.com/arathee2/demonetization-in-india-twitter-data

SOME RESOURCES

- ftp://cran.r-project.org/pub/R/web/packages/tm/vignettes/tm.pdf

- http://onepager.togaware.com/TextMiningO.pdf

- https://rstudio-pubs-static.s3.amazonaws.com/31867_8236987cf0a8444e962ccd2aec46d9c3.html