Text Mining Demo in R

Previous slides discussed some of the theory behind text mining and NLP.
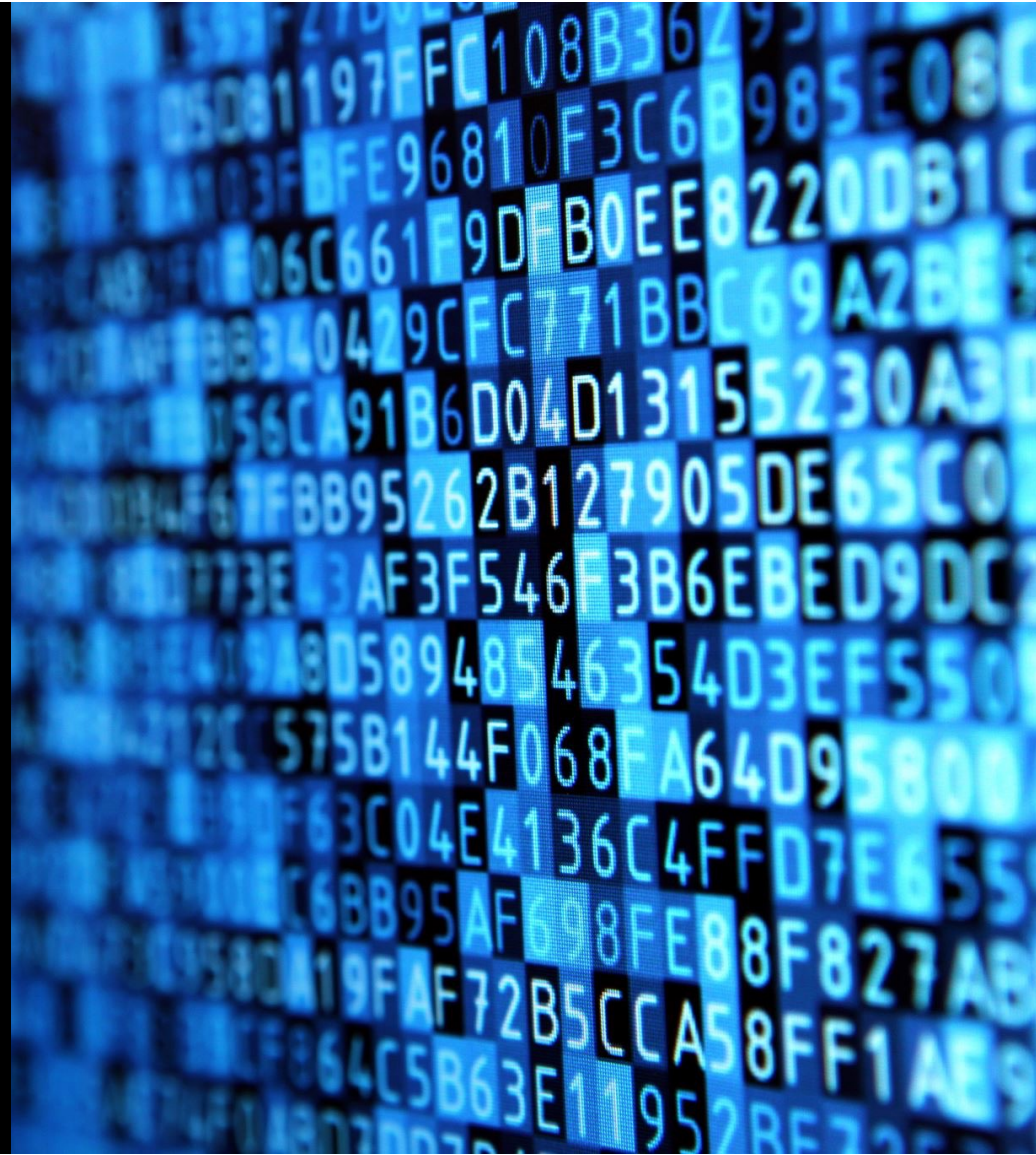
What does this look like in practice?

In this deck, we'll take a look at R code that:

Creates a word cloud out of text data stored in a csv file

Extracts key words from text data using an R NLP lib

WordClouds

- Great for generating a quick visual summary of large amounts of text

- Can be very evocative

- The nature of your text will have a strong impact on the exact effect. Example: Word cloud of restaurant reviews

WordCloud Code Demo

See supplemental text file for R code.

Keyword Extraction

Keyword extraction is still quite a difficult problem. It often relies on NLP, which involves identifying parts of words, rather than text mining, which focuses on statistical patterns.

To carry out keyword extraction you must first match words to their language parts – e.g. noun, verb.

You can play around with the algorithm used to extract keyword by, e.g. placing more emphasis on nouns, verbs, combinations, etc.

Data exploration is useful here.