**Text Mining with "R":**If you work in analytics or data science you are familiar with the fact that data is being generated all the time at ever-faster rates. Analysts are often trained to handle tabular or rectangular data that are mostly numeric, but much of the data proliferating today is unstructured and text-heavy. Many of us who work in analytical fields are not trained in even the simplest approaches to analyzing unstructured text data. This short course serves as an introduction to text mining with the R programming language.

Upon successfully completing this course, students will be able to use R to:

* Assess regular expressions within unstructured text
* Tidy unstructured text data
* Perform word frequency analysis
* Quantify the sentiment in text
* Assess the frequency and importance of terms across documents
* Understand the relationship between words
* Perform topic modeling analysis

## Outline

Day 1

0900-0930: **Introduction**

0930-1030: **Review of tidyverse**

1030-1045: Break

1045-1200: **Regular expressions**

1200-1300: Lunch

1300-1345: **Unnest & clean**

1345-1430: **Frequency analysis**

1430-1450: Break

1450-1600: **Case study / tutorial**

Day 2

0900-0930: **Warm-ups**

0930-1030: **Sentiment analysis**

1030-1045: Break

1045-1130: **tf-idf**

1130-1200: **mini-case study**

1200-1300: Lunch

1300-1400: **Topic modeling**

1400-1500: **Text for prediction**

1500-1520: Break

1520-1600: **Case study / tutorial**