

# BIOF 021: R for Analysis of Text Data

This workshop will provide an introduction to working with text data in R and explore various approaches to analyzing text data. The first session will cover principles for wrangling text data as well as some basic text mining applications. The subsequent two sessions will delve into specific techniques to enable automated analysis of text data.

## Learning Objectives

- Read text data into R and prepare it for analysis
- Understand and select from various options in preparing text, such as stemming, lemmatization, term frequency weighting, term frequency-inverse document frequency weighting (tf-idf), and tokenization
- Conduct simple text mining to explore content of a text corpus
- Describe how unsupervised approaches can be used to identify clusters of related documents
- Process text data to prepare for unsupervised analysis; o Build, train, and evaluate models for text clustering
- Interpret outputs of clustering algorithms
- Describe how supervised approaches can be used to develop text-based models for multi-class classification
- Process text data to prepare for supervised analysis; o Build, train, and test models for text classification

**Class Type:** Workshop

**Prerequisites:**

BIOF 017

BIOF 018

BIOF 019

BIOF 043

BIOF 339

BIOF 501

Any of the above courses and workshops or basic familiarity with R.

**Program:** Bioinformatics and Data Science