# Introduction to Text Analysis
## Summer Institute
## July 2023

**Time:**    July 10-20, 2023
            Live classes Monday, Wednesday, and Thursday 1:00-2:30 (Eastern Time)

**Location:**    Zoom

**Instructor:**    Robyn Ferg, Ph.D.
            Westat
            fergr@umich.edu

## Overview:

In this two-week course, students will learn a variety of natural language processing methods for analyzing and extracting meaning from text data. The course will start with an introduction to text data, including text preprocessing and exploratory methods. The topics that follow will include machine learning models used for topic modeling, clustering, classification, sentiment analysis, and word embeddings. Students will also be introduced to web scraping. Considerations to both long and short texts of various subject matter. Class examples will be demonstrated primarily in R. This course assumes a bachelors-level background in Statistics or related field and knowledge of R or Python; no prior knowledge of text analysis is assumed.

## Course Materials:

All course materials will be available online on the class Canvas website. Course materials on the website include video lectures, recommended readings, assignments, and code.

## Evaluation:

This is a one credit course. All participants, whether taking the class for credit or not for credit, are expected to fully participate in the class. All students will receive a grade determined by class participation and completion of three assignments. If a non-credit student does not want to receive a letter grade, they must complete the audit form (available by emailing Patsy Gregory, pagregor@umich.edu) and submit it to the Summer Institute office. Assignments are optional, but recommended, for students auditing the course.

## Accommodations for Students with Disabilities:

If you believe you need an accommodation for a disability, please contact the Services for Students with Disabilities (SSD) office to help us determine appropriate academic accommodations. SSD (734-763-3000; http://ssd.umich.edu) typically recommends

accommodations through a Verified Individualized Services and Accommodations (VISA) form. Any information you provide is private and will remain confidential.

**Course Schedule**

**Monday, July 10**
**1:00-2:30**      Course introduction, student introductions, course expectations,
                Lecture: Introduction to Text Data

Assignment 1 released

**Tuesday, July 11**
Video Lecture:
- Remaining introduction to text data videos
- Topic Modeling

**Wednesday, July 12**
**1:00-2:30**      Q+A on topic modeling
                Introduction to text data code, topic modeling code

**Thursday, July 13**
**1:00-2:30**      Lecture: Sentiment Analysis
                Sentiment analysis code
                Lecture: Text Summarization
                Text summarization code

Assignment 2 released

**Friday, July 14**
Assignment 1 due (midnight)

**Monday, July 17**
**1:00-2:30**      Lecture: Word Embeddings
                Code for word embeddings

Assignment 2 due (midnight)

**Tuesday, July 18**
Video Lecture:
- Classification
- Clustering

Assignment 3 released

**Wednesday, July 19**
**1:00-2:30**     Q+A on classification and clustering
                Code for classification, code for clustering

**Thursday, July 20**
**1:00-2:30**     Lecture: Web Scraping
                Web scraping code
                Course wrap-up

Assignment 3 due (midnight)