



JOINT PROGRAM IN SURVEY METHODOLOGY

NATURAL LANGUAGE PROCESSING WITH R

An Online short course sponsored by the Joint Program in Survey Methodology

January 20 – January 31, 2025

ROBYN FERG

Westat

COURSE OBJECTIVES

In this two-week course, participants 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 pre-processing and exploratory methods. The topics that follow will include machine learning models used for topic modeling, clustering, classification, sentiment analysis, and word embeddings. Participants will also be introduced to web scraping. Considerations will be given to both long and short texts of various subject matter. Class examples will be demonstrated primarily in R.

WHO SHOULD ATTEND

Anyone interested in learning and applying methods for analyzing text data is encouraged to attend. This course is designed for persons with a bachelors-level background in Statistics or related field. Basic knowledge of R is expected.

THE INSTRUCTOR

ROBYN FERG (Ph.D., Statistics, University of Michigan) is a Senior Statistician at Westat and teaches graduate and undergraduate courses at the University of Maryland and the University of Michigan. She has given seminars on text analysis at the Census Bureau and University of Maryland and has presented text analysis research at several international and national conferences. Her doctoral and post-doctoral research focuses on developing methods for extracting insights from social media text data.



JOINT PROGRAM IN SURVEY METHODOLOGY

CLASS STRUCTURE

The course will be in an online format from January 20 to January 31, 2025.

Participants will have online access to the course packet (slides) and to the recorded lectures. The recorded lectures will be divided into eight sections (four per week) of

1-1.5 hours each. Participants can watch the videos at their own pace. **Live group Online discussions are scheduled for Friday 01/24/25 and 01/31/25, 10-11:30 AM.**

These live sessions will be used to discuss questions that came up over the week when watching the videos. Small assignments will be posed throughout the week to provide the participants hands-on experience. These assignments are not mandatory nor will they be collected, but participants are highly encouraged to complete the assignments before the live sessions. Solutions to these assignments will be discussed during the live sessions.