MODULE ONE: INTRODUCTION AND OVERVIEW TO DATA SCIENCE

This module will provide a social science perspective on data science, introducing you to the objectives of the course via a visual overview of each module. It will discuss how data science is changing social science and statistics, and will cover reliability, generalizability, and reproducibility.

MODULE TWO: ETHICS IN DATA SCIENCE

This module will teach you about the shortcomings and problems of data science in respect to the groups of people it affects, who it’s representing, and how to responsibly acknowledge these issues in research. Beyond problems in data collection and data sources are issues of privacy, sampling, population size, interpretation, and application. This module importantly emphasizes issues of deidentification and reidentification, and data security.

MODULE THREE: SURVEYS AND CROWDSOURCING DATA

This module will give an overview of how to construct a survey and crowdsource responses. Specifically, you will be introduced to Qualtrics and learn about other freely available tools for building surveys. We will discuss Amazon’s Mechanical Turk, which is becoming the new norm for online data collection in the social sciences.

MODULE FOUR: DATA SCIENCE TOOLS

This module will introduce you to the data science tools commonly used in social science research. We will discuss the value of open-source programming languages, specifically R and Python, for research of this nature and weigh the advantages and disadvantages of each.

This module introduces Jupyter notebooks, and concludes with a brief overview of Git and GitHub as they have become essential for collaborative research programming projects.