Md Nakir Ahmed

3. Read

a. Foster et al. 1, 2

The Big Data and Social Science textbook introduces how modern data sources and computational tools can enhance social science research and policymaking. It begins by framing the role of big data and then moves into practical methods for acquiring and using web-based data.

Chapter 1 – Introduction

Chapter 1 sets the stage by arguing for the value of big data methods in social science and public policy. It emphasizes that while large-scale, unconventional datasets are becoming increasingly available, the real contribution comes when these are used to answer questions in familiar social science ways—from hypothesis through measurement to inference—rather than simply doing flashy analyses. The chapter introduces a "use case" around research & development investments and societal outcomes, illustrating how open, non-confidential data sources can serve typical research goals. It also highlights crucial issues of privacy, reproducibility, and how researchers should think through ethical concerns as part of their workflow.

Chapter 2 – Working with Web Data and APIs

Chapter 2 focuses on how external data from the web—whether embedded in web pages or delivered via APIs—can complement traditional datasets in social science research. It explains methods for gathering structured or unstructured data (web scraping, accessing APIs), discusses challenges like data quality, changing webpage structure, access restrictions, and the need for careful planning. The chapter also stresses that data collection should be driven by research questions: knowing why you collect what you collect is as important as getting the data itself. It introduces tools (using Python) for scraping web content and calling APIs, with examples from social media and government open data portals.

b. Breiman, Leo. 2001. "Statistical modeling: The two cultures (with comments and a rejoinder by the author)." Statistical science 16, no. 3 199-231.

Leo Breiman's (2001) essay "Statistical Modeling: The Two Cultures" contrasts two dominant approaches in statistics: the data modeling culture and the algorithmic modeling culture.

In the data modeling culture, statisticians assume data follow a specified stochastic model (e.g., linear or logistic regression), and the focus is on estimating parameters and testing hypotheses. In the algorithmic modeling culture, common in machine learning, the emphasis is on predictive accuracy using flexible methods (e.g., decision trees, random forests, neural networks) without assuming a fixed data-generating model.

Breiman argues that the data modeling culture dominated academic statistics but often produced weak predictions, while the algorithmic approach—though less interpretable—achieves higher predictive performance and better adapts to complex, real-world data. He calls for statisticians to embrace algorithmic methods and prioritize prediction alongside inference to stay relevant in modern data science.

4. Data

a. Identify a data method you want to develop expertise

I plan to develop expertise in object detection using deep learning, a data method within computer vision that enables the identification and classification of objects in image datasets. This method integrates feature extraction and predictive modeling, making it especially useful for analyzing large and complex visual data.

5. Discussion for next week:

a. What is data?

Data is information collected in raw form about facts, figures, or observations and that can be processed and analyzed to generate knowledge.

b. What is big data?

Big data refers to extremely large, complex, and high-velocity datasets that require advanced tools and methods to store, process, and analyze effectively.

c. Small data?

Small data is manageable in size, often structured, and can be processed with traditional statistical methods on a single computer.

d. Data generation process

The data generation process involves collecting observations, recording measurements, storing them systematically, and preparing them for analysis through cleaning and transformation.