DRAFT SYLLABUS: March 28, 2026

A DATA-SCIENTIFIC INTRODUCTION TO JUSTICE AND DEMOCRACY

RUTGERS UNIVERSITY | DEPARTMENT OF STATISTICS
01:960:492 | FALL 2026
PROFESSOR RUOBIN GONG
SEC-209 (BUSCH CAMPUS) | WEDNESDAYS 8:30AM – 11:20AM

COURSE DESCRIPTION

We live in a data-centric, evidence-driven world. While AI changes the world order by reimagining work, study, and the ways we think, some values are invariable and we hold them dear, including our commitment to civil and criminal justice, a democracy society, and the rules of law. This course examines fundamental questions of justice and democracy from a data scientist's perspective. Through reading judicial opinions, including landmark Supreme Court cases and academic papers, we will come to see how data and statistical reasoning provide support and clarity for important debates, and how data science intersects with law, social policy, and our civic life.

LEARNING OBJECTIVES

At the end of this class, you will be able to:

- Understand statistical reasoning in legal contexts;
- Apply data analysis to justice and policy questions;
- Critically evaluate empirical arguments.

PREREQUISITES

This is an undergraduate level topics course. There are no prerequisites for this course.

TENTATIVE LIST OF TOPICS

MODULE 0: LET THE DATA SPEAK

This introductory module motivates the use of quantitative evidence to support the making of prudent judicial decisions and sound social policy. As a historical case study, we will learn about the Brandeis Brief, named after its inventor Louis Brandeis. Before Brandeis became a Supreme

Court Justice, he represented the State of Oregon to defend its laws that protected women's rights in the workplace. He created, as no lawyer has ever done before, a massive compilation of data and reports and documentation tending to show the detrimental effect of unrestricted working hours on the health of female employees. Brandeis' win in the landmark case of *Muller v. Oregon* affirmed the value of protective legislation, a rare recognition during the Supreme Court's *Lochner* era jurisprudence.

Data science elements: As a preparation for the semester, the introductory module familiarizes us with foundational concepts in probability and statistics.

MODULE 1: EQUAL PROTECTION

The Fourteenth Amendment of the U.S. Constitution provides that no state shall "deprive any person of life, liberty, or property, without due process of law; nor deny to any person within its jurisdiction the equal protection of the laws." This Module explores the evolving interpretation of "equal protection" in the contexts of access to education and criminal justice. We will read the landmark Supreme Court cases of *Brown v. Board of Education* ("separate is not equal"), *Grutter v. Bollinger* (affirmative action in education), and most recently *Students for Fair Admissions v. Harvard* (the end of affirmative action). We will then turn to examine, via a series of academic papers, the apparent impossibility to resolve certain observed biases.

Data science elements: As part of this module we will study conditional probabilities, the Bayes Rule, and the infamous Simpson's Paradox.

MODULE 2: TRUTH-SEEKING IN THE COURTROOM

A plaintiff who has been hurt and seeks justice in court must demonstrate that the defendant did something wrong that caused her injury. She must present enough proof to convince a jury that the defendant was to blame. This Module examines how causation is proven in a court of law, and how that process relates to – and differs from – how scientists approach the causal question.

Data science elements: We will study correlation, regression, and causal inference as part of this module.

MODULE 3: DEMOCRACY'S DATA INFRASTRUCTURE

Article I of the U.S. Constitution mandates Congress to carry out a census every ten years. The census is an "actual enumeration" of the population, and forms the basis of apportionment of taxes and of House seats among the states. The accuracy and reliability of the census is crucial to the well-functioning of this democracy.

The authority to conduct the Decennial Census is delegated to the U.S. Census Bureau. To do its job, the Census Bureau must abide by the laws and regulations that grant it the authority in the first place, one of them being the Constitutional mandate that the Census be an "actual enumeration." Statistical and legal questions arise when the Census misses some people – most

trickily, when the Bureau knows that it is missing some people but cannot get them to respond. What can the Census Bureau do, and what must it not do?

Data science elements: In this module, we will study the statistical concepts of sampling and sampling bias, missing data, and imputation.

MODULE 4: THE RIGHT TO INFORMATIONAL PRIVACY

The right to informational privacy and confidentiality is said to be one of the implicit rights conferred to individuals by the U.S. Constitution. The meaning of informational privacy, however, is elusive to the courts as it is to data scientists. This Module walks us through the following formulations of the notion of "informational privacy": the judicial; the legal-academic; the philosophical; the social-scientific; the statistical; the cryptographic; and finally, the intuitive. How are these formulations relate to, and differ from, one another?

Data science elements: As part of this module we will delve into the subjects of Bayesian inference and differential privacy.

MODULE 5: ONE PERSON, ONE VOTE

Gerrymandering refers to the manipulation of electoral maps in ways that systematically advantages one constituent interest over another. Gerrymandering, when used to suppress the voting power of one racial group over another, is unconstitutional because it is a violation of the Equal Protection guarantee of "one person, one vote." We will discuss the use of statistical analysis to detect and prove when gerrymandering is suspect.

Data science elements: In this module, we will study statistical hypothesis testing and the concept of Markov chain Monte Carlo.

ASSIGNMENTS AND GRADING

The course evaluation scheme consists of the following components:

- Active participation in class discussions (20%);
- Weekly problem sets (40%);
- (Mandatory but non-credit) midterm (0%);
- Final exam (40%).