Social Data Science

SOCIOL 114
Winter 2026

Lecture 1: Welcome

Learning goals for this course

By the end of this course, you will be able to

- connect theories about inequality to quantitative empirical evidence
- evaluate the effects of hypothetical interventions to reduce inequality
- conduct data analysis using the R programming language

How computing looked **in the 1950s**

Source: NASA

How computing looked **in the 1980s**

Source: Wikimedia

How computing looks **today**

Source: Apple

Social Data Science

Our class will combine

- new data science tools for estimation
- longstanding social science research designs

We will seek to tell stories with data.

Figure from England et al. (2020)

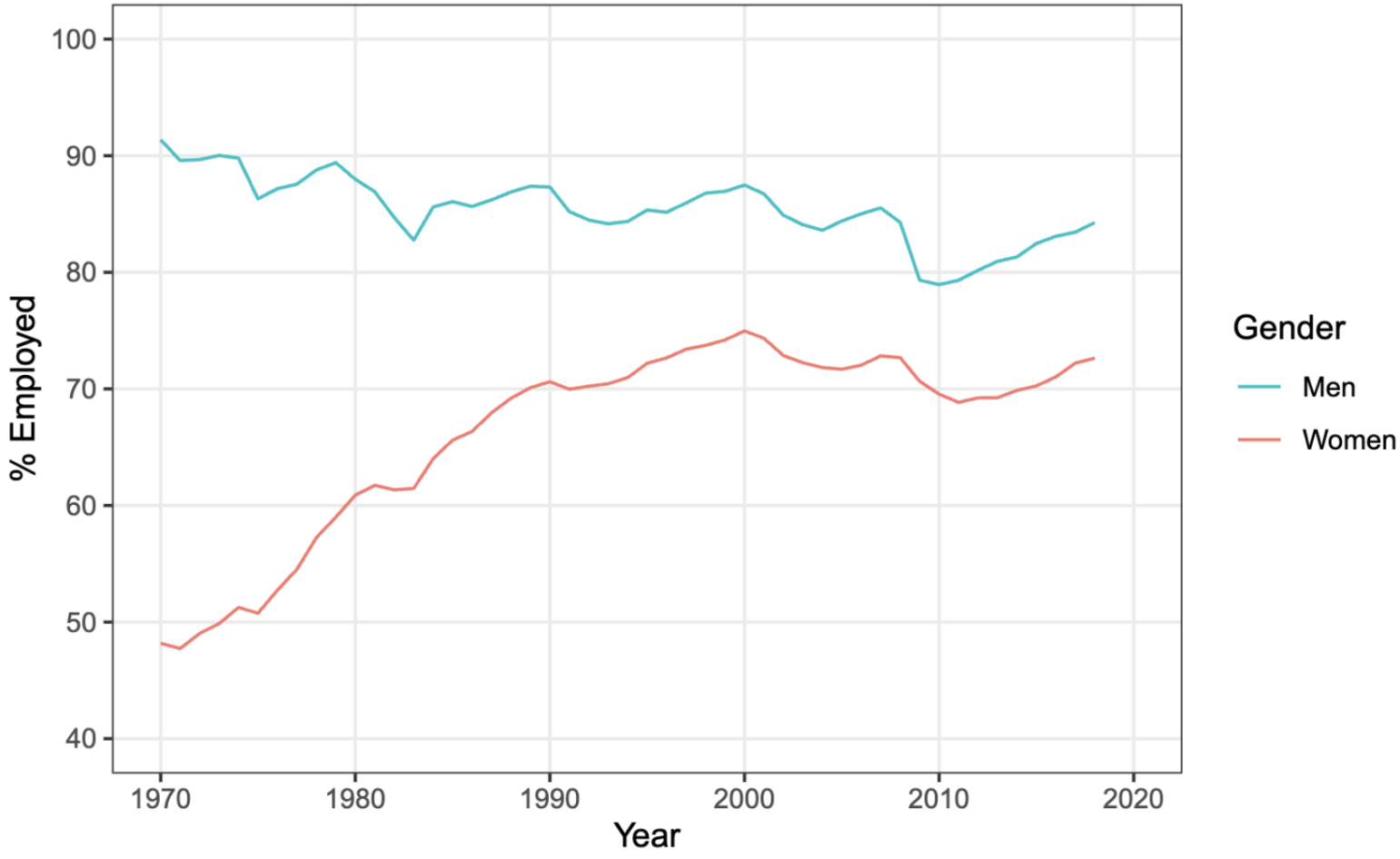

Fig. 1. Percentage of women and men, age 25 to 54, employed in the last week, 1970 to 2018. Source: Authors' computations from IPUMS CPS ASEC samples for 1970 to 2018.

Research questions in social data science

Which of the following can be answered by data science?

1. The gender employment gap closed from 1970–1990 but has remained roughly constant since then.
2. We should enact new policies to eliminate the remaining gap.

Research questions in social data science

Which of the following can be answered by data science?

1. The gender employment gap closed from 1970–1990 but has remained roughly constant since then.
2. We should enact new policies to eliminate the remaining gap.

(1) is an **empirical question** (more amenable to data science)
(2) is a **normative question** (less amenable to data science)

Elements of a data science question

Elements of a data science question

1. a unit of analysis
    - a row of your dataset

Elements of a data science question

1. a unit of analysis
   - a row of your dataset
2. an outcome
   - a variable with a value for each unit

Elements of a data science question

1. a unit of analysis
   - a row of your dataset
2. an outcome
   - a variable with a value for each unit
3. a target population
   - a set of units about whom to infer
   - clear who is included and who is not

Course logistics

Who should take this course?

The course is designed for upper-division undergraduate students.

Attendance

Public health matters—stay home if sick!

Otherwise, we hope to see you in class.

Course materials

All materials will be posted here:

soc114.github.io

Course support

- Post questions on Piazza
- Office hours

Software

As soon as possible, you should

- Install R (statistical software)
- Install RStudio (user interface)

There are also cloud-based options.

Method of assessing student achievement

Quizzes	50%
Problem sets	50%

Quizzes

- Due MW at 5pm
- Submit in BruinLearn after each lecture
- Automatically graded
- Attending class will make them easier
- Lowest 2 scores dropped at end of the quarter

Problem sets

- Due Friday at 5pm
- Material covered by Tuesday
- Graded by PhD student reader
- Content includes
  - Code to analyze data
  - Written summaries in English

Late work

- 0.5% penalty for each hour late
- 1 minute late = 0.5% penalty
- 23:01 late = 24 x 0.5 = 12% penalty
- Automatic in BruinLearn

No assignments will be accepted after Mar 20 at 5pm.

Collaboration

- encouraged to work together
- consulting help is great
- should never involve one student having possession of a copy of all or part of work done by someone else, in the form of an email, an email attachment file, or a hard copy

Academic integrity

Each student in this course is expected to abide by the UCLA Academic Integrity policies. Any work submitted by a student in this course for academic credit must be the student's own work.

Students with disabilities

You belong in this course. We are happy to work with you on appropriate accomodations—see the syllabus for details about working with CAE.

Mental health and wellbeing

Your health and wellbeing are important to us!

See syllabus for links to mental health resources. We hope our course helps you thrive at UCLA, and your thriving is far more important than anything in this course.

Honors section (Soc 189)

- Capped at 20 seats
- Open to any student in this course!
- W 1–1:50pm
- Leads to an independent research report analyzing an empirical question of your choosing

Questions about logistics?

Introduction to R

Using the course website, we will learn about

- Software Prerequisites
- Basics of R