Introduction to Quantitative Methods

Quantitative Methods in Political Economy
Pre-sessional Lecture 1

Wouter Leenders
wouter.leenders@kcl.ac.uk

Purpose of pre-sessional lectures

- Set foundation for module 7SSPN107
- Help you decide whether to take 7SSPN107
- Guidance on using quantitative methods in your projects
- Guidance on using R and RStudio for data analysis (tomorrow's lecture)

What are qualitative methods?

- Methods for studying social and political phenomena in depth and in context
- Often work with non-numerical evidence: interviews, documents, archives, observation, field notes
- Common approaches include interviews, ethnography, case studies, process tracing, and archival analysis
- Particularly useful for studying mechanisms, institutions, and historical processes

What are quantitative methods?

- Methods that use data expressed or represented in quantitative form
- Used to describe patterns, make predictions, estimate relationships, and assess causal claims
- Usually involve quantifying uncertainty
- Data can include conventional numerical data *and* text as data

Text as data

Text can be transformed into quantities that we can analyse systematically.

- Parliamentary speeches
- Party manifestos
- Newspaper articles
- Social-media posts
- Historical documents

Examples include word frequencies, dictionaries, topic models, and numerical representations of text.

Mixed methods

- Combine quantitative and qualitative methods in the same research project
- The methods can answer different parts of the same question
- Example: establish a broad statistical pattern, then use interviews or archival evidence to understand the mechanism
- Or begin with qualitative evidence and use it to develop hypotheses for quantitative analysis

What questions can quantitative methods answer?

Descriptive

What is happening? How much? How does it vary across people, places, or time?

Predictive

Given what we know, what is likely to happen or which observations are likely to have a particular outcome?

Causal

What would happen to an outcome if we changed one factor while holding everything else relevant fixed?

What is causality?

A causal question compares two states of the world:

- Outcome if a unit receives treatment
- Outcome if the same unit does not receive treatment

The fundamental difficulty is that we never observe both outcomes for the same unit at the same time.

Research design is therefore central to causal inference.

Descriptive analysis

Examples of descriptive questions:

- How high is inequality?
  Atkinson, Piketty & Saez (2011)
- How has support for democracy evolved over time?
  Pew Research Center (2024)

Answering apparently simple descriptive questions still requires choices about measurement, data, and comparison.

Causal analysis

Examples of causal questions:

- Why are some countries rich and some countries poor?
  Acemoglu, Johnson & Robinson (2001)
- What is the effect of minimum wages on employment?
  Card & Krueger (1994)
- What is the effect of education on earnings?
  Card (1995)

Causal analysis

More examples:

- Does democracy cause economic growth?
  Acemoglu, Naidu, Restrepo & Robinson (2019)
- Does the media affect political behaviour?
  Martin & Yurukoglu (2017)
- Does campaigning increase turnout?
  Gerber & Green (2000)

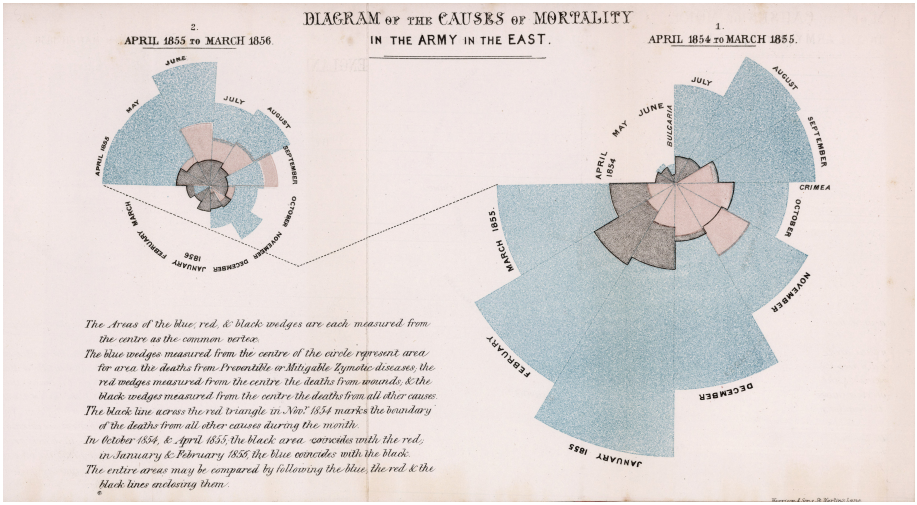
Florence Nightingale

- Nurse, social reformer, and statistician
- Analysed mortality among British soldiers during the Crimean War
- Used statistical evidence to argue for sanitary reform
- Founded the first professional school of nursing at St Thomas' Hospital, now the King's Florence Nightingale Faculty of Nursing, Midwifery & Palliative Care

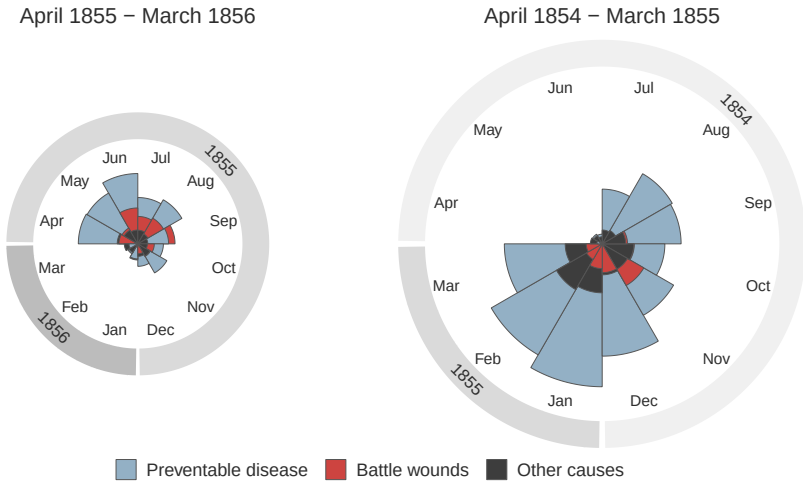
Nightingale's data analysis

- Collected and organised administrative records on deaths among British soldiers
- Classified deaths by cause: disease, wounds, and other causes
- Compared mortality across time, before and after sanitary reforms
- Used data visualisation to communicate the magnitude of and changes in mortality

Nightingale's polar-area diagram



Nightingale's polar-area diagram recreated in R



What did the analysis show?

- Disease killed many more soldiers than battlefield wounds
- Mortality fell substantially after sanitary reforms
- The evidence helped strengthen the case for public-health reforms, including improved sanitation and hygiene

Quantitative methods in research

A common workflow in quantitative social science is:

1. Research question
2. Theory
3. Hypothesis
4. Research design
5. Data
6. Analysis
7. Interpretation and conclusions

Research question

- The question the research is designed to answer
- It should determine which evidence and methods are appropriate
- Quantitative questions can be descriptive, predictive, or causal

Example: Does door-to-door canvassing increase voter turnout?

Theory

- A theory provides an explanation for why we expect a relationship or pattern
- It identifies mechanisms linking explanatory factors to outcomes
- Theory helps determine what evidence would support or challenge an argument

Example: An individual's decision whether or not to vote may depend on the salience of an election, or on whether voters believe it is the "right" thing to do. Campaigns may increase that salience and may appeal to the moral duty to vote.

Hypothesis

- A hypothesis is an observable implication of a theory
- It should state a pattern or relationship that can be confronted with data
- A single theory can generate several hypotheses

Example: Voters contacted by canvassers are more likely to vote than otherwise similar voters who are not contacted.

Research design

- Specifies how the research question will be answered
- Determines which comparisons are informative
- Particularly important when the goal is to make a causal claim

Example: Randomly assign households to a canvassing group and a control group.

Data

Data consist of observations and variables.

- Surveys
- Experiments
- Administrative records
- Historical and archival sources
- Text and digital trace data

Example: Treatment assignment, turnout, and relevant characteristics for each voter or household.

Analysis

- Summarise and visualise the data
- Estimate quantities or relationships of interest
- Quantify uncertainty
- Assess whether the evidence is consistent with the hypothesis

Example: Compare turnout rates in the canvassing and control groups; later in the module, express this comparison using regression.

Interpretation and conclusions

- What do the results imply for the research question?
- How large is the estimated relationship or effect?
- How uncertain is the estimate?
- What assumptions are required for the interpretation?
- What are the limitations and implications?

The data: `AER::CPSSW8`

- AER (*Applied Econometrics with R*) is an R package providing datasets and tools for econometric analysis
- 61,395 US full-time workers, aged 21–64, from the Current Population Survey
- Earnings: average pre-tax hourly earnings, in US dollars
- Other variables: years of education, age, gender, and region

Data frame

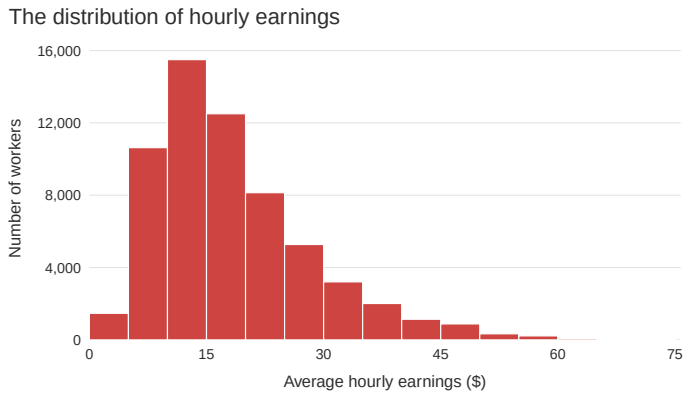
A data frame

Each row is a worker; each column is a variable.

Row	Earnings ($ per hour)	Education (years)	Age (years)	Gender
1	20.67	14	31	Male
2	24.28	12	50	Male
3	10.15	12	36	Male
		⋮		
61,393	12.02	12	44	Female
61,394	13.94	12	39	Male
61,395	11.92	10	24	Male

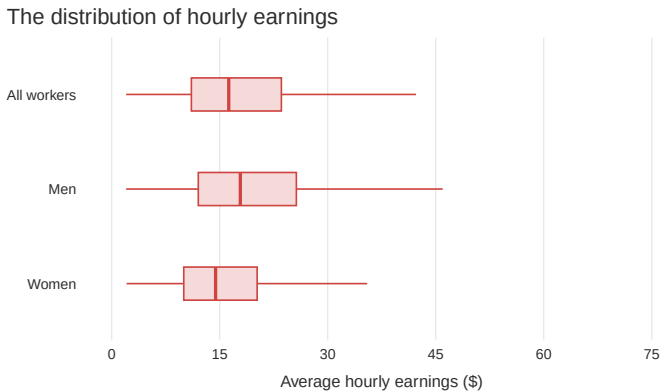
Rows are workers, columns are variables, and cells contain their values.

Histogram



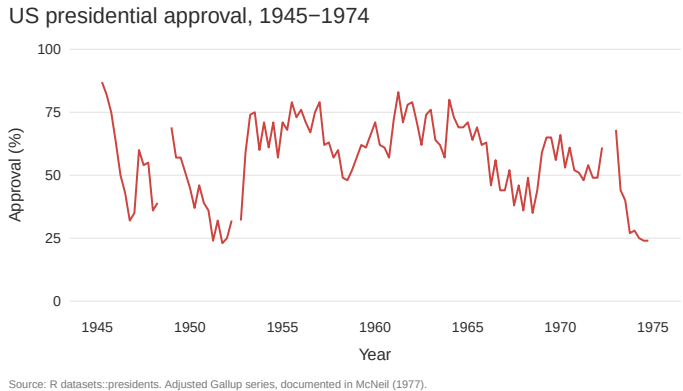
Each bar spans $5 per hour. Its height shows the number of workers in that earnings range.

Box plot



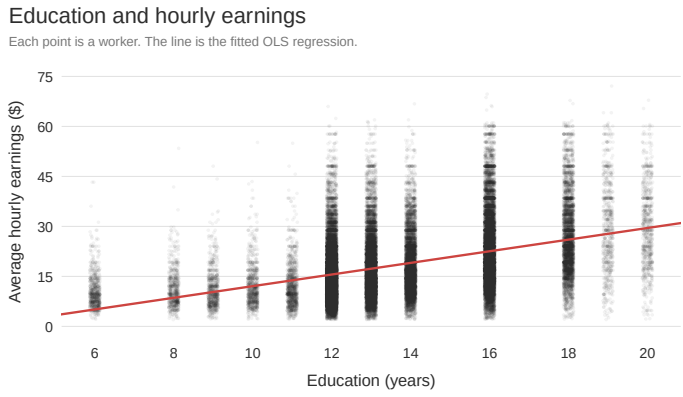
Box plots compare earnings distributions using medians and quartiles.

Line graphs



Line graphs show change over time, here in US presidential approval.

Scatter plots



The scatter plot shows the relationship between years of education and hourly earnings.

Regression table

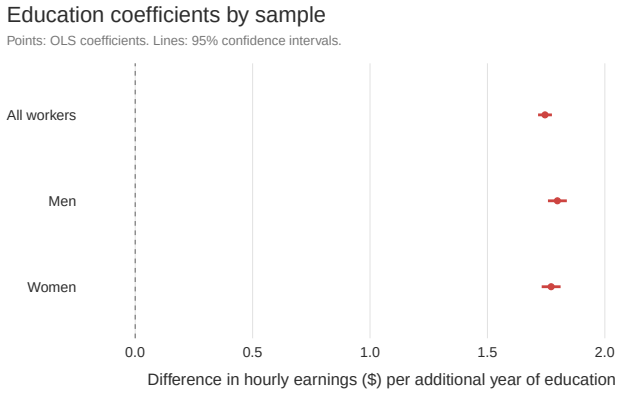
Hourly earnings and education

Outcome: average hourly earnings ($). Ordinary least squares (OLS).

	All workers	Men	Women
Education (years)	1.75 (0.02)	1.80 (0.02)	1.77 (0.02)
Constant	−5.38 (0.21)	−4.25 (0.28)	−8.07 (0.29)
Observations	61,395	34,348	27,047
R−squared	0.18	0.18	0.22

Coefficients summarise estimated relationships. Standard errors in parentheses quantify uncertainty.

Coefficient plot



Points show the education coefficients for each group. Lines show 95% confidence intervals.

Module outline: lectures

1. Descriptive statistics
2. Normal distribution
3. Confidence intervals
4. Hypothesis testing
5. Regression I
6. Regression II
7. Regression III
8. Diagnostics
9. Panel data I
10. Panel data II

Theoretical material is based primarily on Stock & Watson, *Introduction to Econometrics* (4th ed.).

Module outline: R seminars

- R fundamentals and workflow
- Exploring and visualising data
- Transforming and combining data
- Regression in R
- Regression and panel data

Practical material is based primarily on Wickham, Çetinkaya-Rundel & Grolemund, *R for Data Science* (2nd ed.).

Module outline: assessment and prerequisites

- Problem set: 60%
- Two-hour in-person examination: 40%
- No prior knowledge of statistics or R is expected
- Familiarity with basic high-school algebra is assumed
- Mathematical notation and formulas will be used extensively

Should you take this module?

The module is likely to be a good fit if you want to:

- Understand quantitative evidence used in political economy and the social sciences
- Learn how to summarise, visualise, and analyse data
- Understand statistical inference and regression
- Conduct your own quantitative analysis in R