INTRODUCTION TO DATA

Observational studies and experiments

Types of studies

- **Observational study:**
  - Collect data in a way that does not directly interfere with how the data arise
  - Only correlation can be inferred
- **Experiment:**
  - Randomly assign subjects to various treatments
  - Causation can be inferred

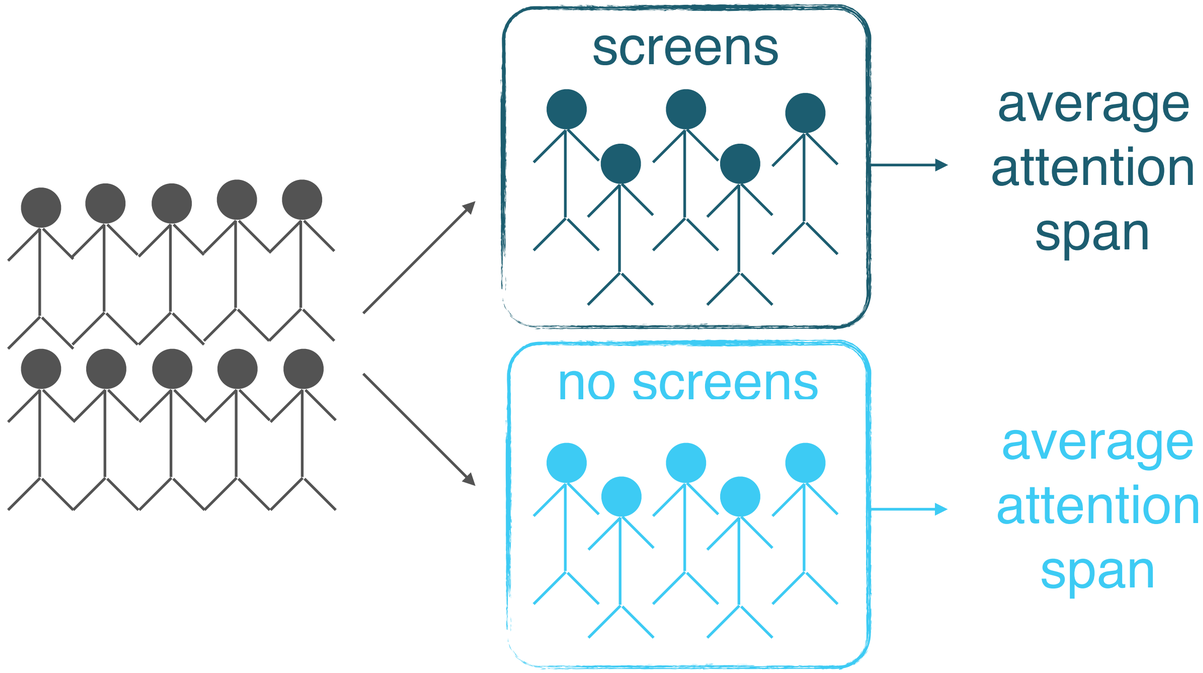
Design a study

Screens at bedtime and attention span

observational study

Association?

experiment



Causation?

INTRODUCTION TO DATA

Let’s practice!

INTRODUCTION TO DATA

Random sampling and random assignment

Random...

- **Random sampling:**
  - At selection of subjects from population
  - Helps generalizability of results
- **Random assignment:**
  - At selection of subjects from population
  - Helps infer causation from results

Scope of inference

	Random assignment	No random assignment	
Random sampling	Causal and generalizable	Not causal, but generalizable	Generalizable
No random sampling	Causal, but not generalizable	Neither causal nor generalizable	Not generalizable
	Causal	Not causal	

INTRODUCTION TO DATA

Let’s practice!

INTRODUCTION TO DATA

Simpson’s paradox

Explanatory and response

x (explanatory)	y (response)

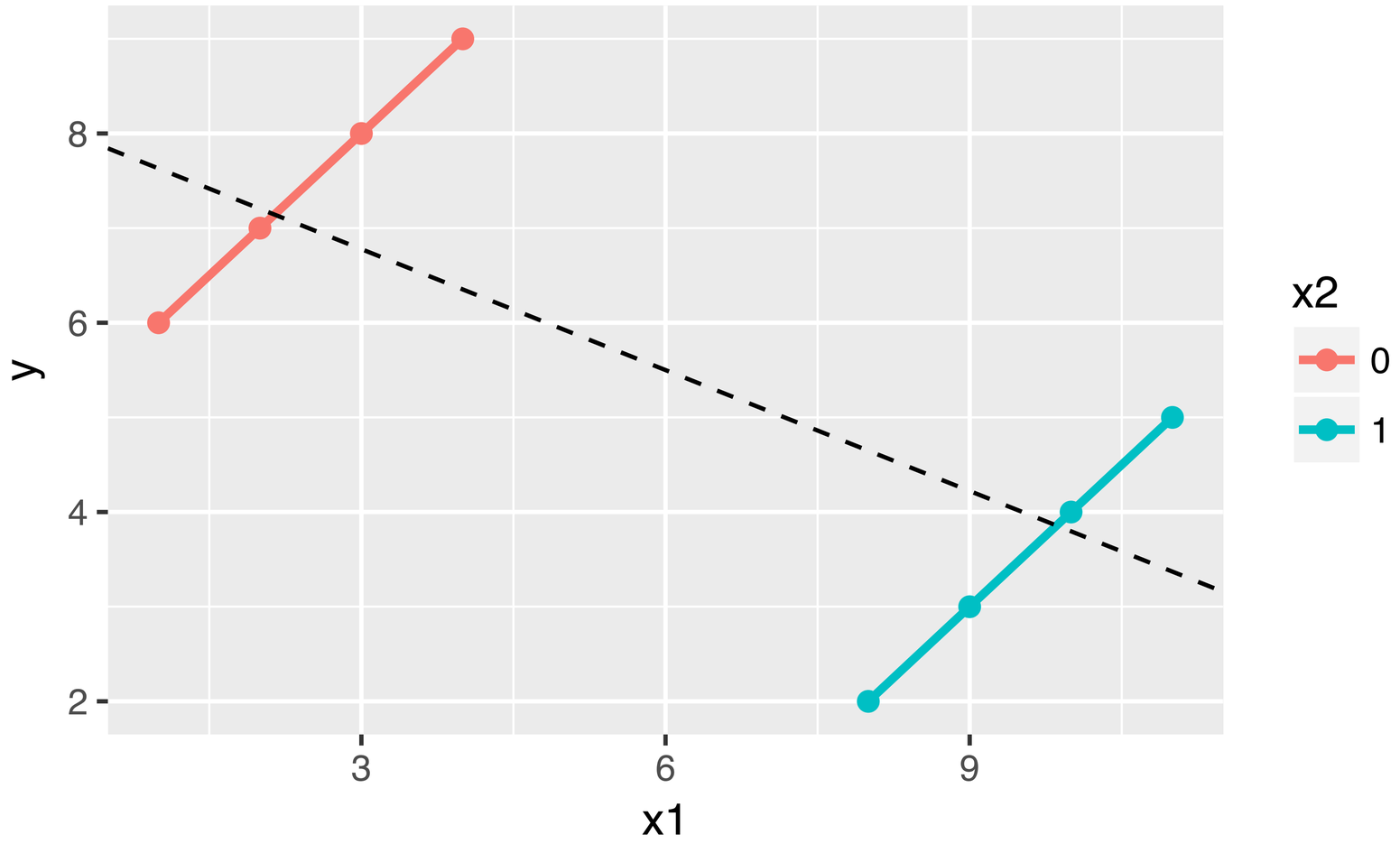
Multivariate relationships

x_1 (explanatory)	x_2 (explanatory)	x_3 (explanatory)	y (response)

Multivariate relationships

calories (explanatory)	age (explanatory)	fitness (explanatory)	heart health (response)

Simpson's paradox



Berkeley admission data

	Admitted	Rejected
Male	1198	1493
Female	557	1278

INTRODUCTION TO DATA

Let’s practice!

INTRODUCTION TO DATA

Recap: Simpson's paradox

Simpson's paradox

- Overall: males more likely to be admitted
- Within most departments: females more likely
- When controlling for department, relationship between gender and admission status is reversed
- Potential reason:
  - Women tended to apply to competitive departments with low admission rates
  - Men tended to apply to less competitive departments with high admission rates

INTRODUCTION TO DATA

Let’s practice!