# Social Data Science

Course Summary

Goals of the course

- ▶ visualize economic inequality
- ▶ connect theories to quantitative evidence
- ▶ evaluate the effects interventions
- ▶ conduct data analysis using R

We learned to ask quantitative questions

- ► Define a population of units
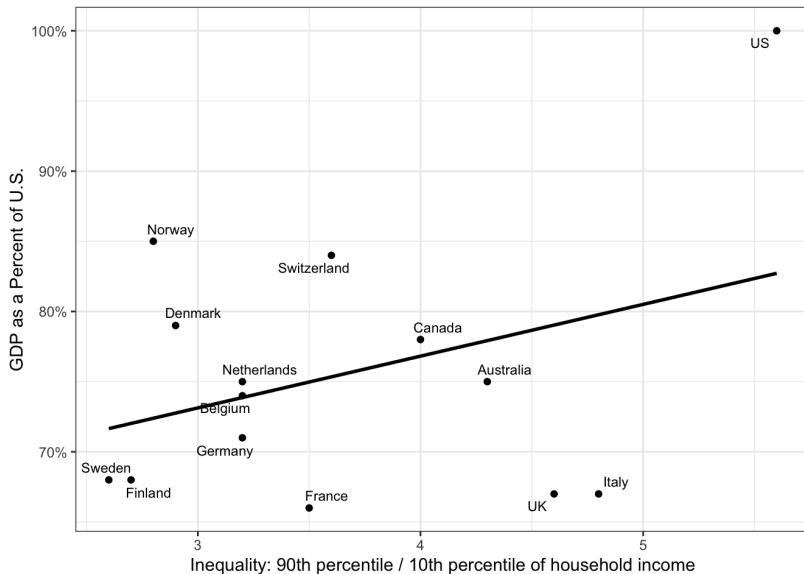- ► Outcome for every unit
- ► Summary statistic

We learned to ask quantitative questions

90th percentile ($212,110)

gets 14x as much as the

10th percentile ($15,640)
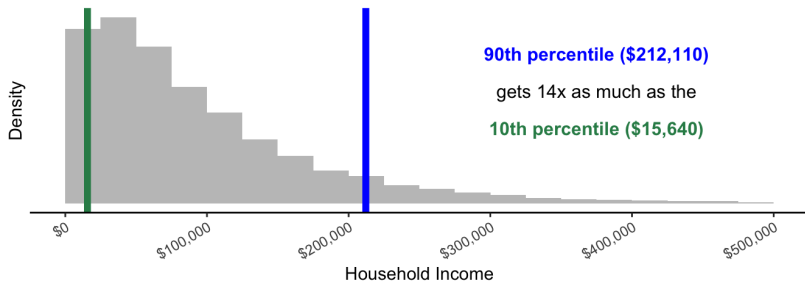
We learned how to make a visualization in R

- ▶ Call to ggplot()
- ▶ Adding layers
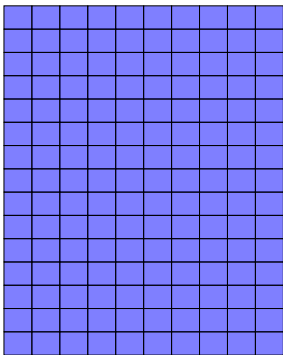- ▶ Customizing axis labels

We learned how to make a visualization in R

We learned about population sampling
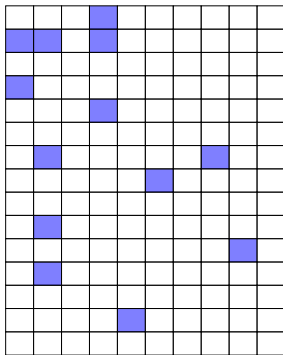
Full Count Enumeration

Back of Room

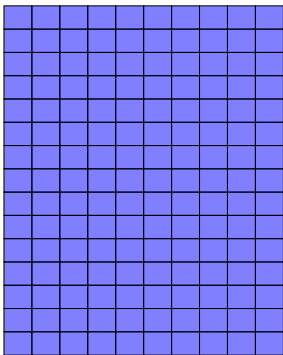Front of Room

Probability Sample

Back of Room

Front of Room

We learned about population sampling

Full Count Enumeration
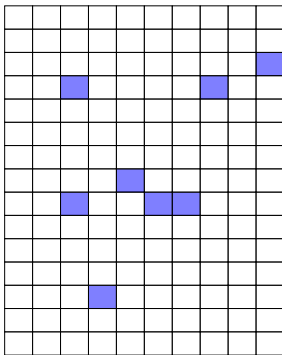
Back of Room

Probability Sample

Back of Room

Front of Room
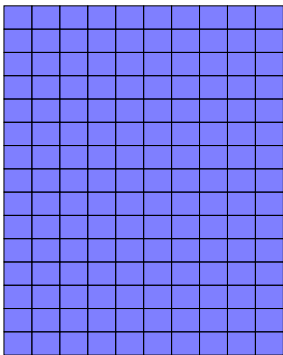
Front of Room

We learned about population sampling

Full Count Enumeration

Back of Room

Probability Sample

Back of Room

Front of Room

Front of Room

We learned about population sampling

Sample Design

Back of Room

P(Sampled) = 0.1

P(Sampled) = 0.5

Front of Room

Sample

Back of Room

Front of Room

We learned about population sampling

Sample Design

Back of Room

P(Sampled) = 0.1

P(Sampled) = 0.5
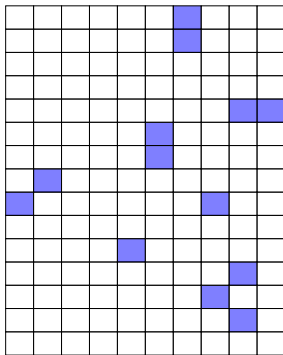
Front of Room

Sample

Back of Room

Front of Room

We learned about population sampling

Sample Design

Back of Room

P(Sampled) = 0.1

P(Sampled) = 0.5
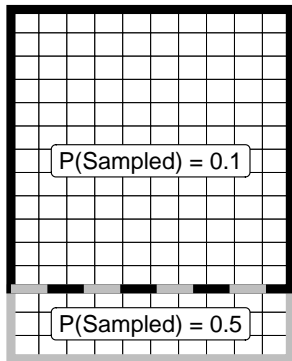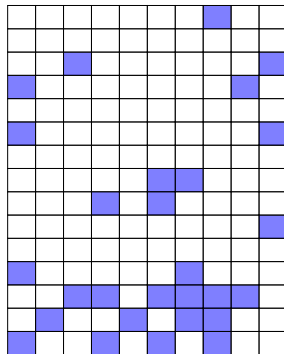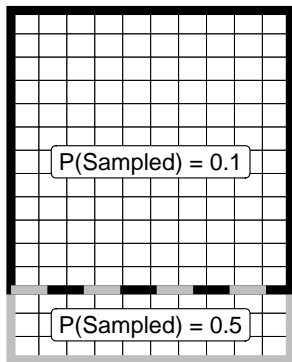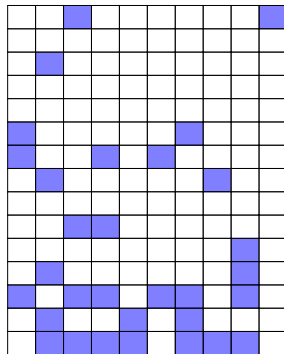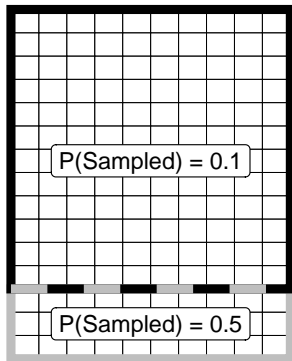
Front of Room

Sample

Back of Room
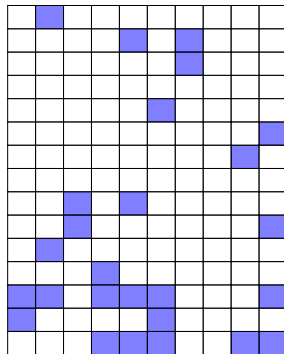
Front of Room

We learned to define and identify causal effects

Population Potential Outcomes		Randomized Treatment	Observed Outcomes	
Y^1_{Maria}	Y^0_{Maria}	$A_{Maria} = 1$	Y^1_{Maria}	
$Y^1_{William}$	$Y^0_{William}$	$A_{William} = 0$		$Y^0_{William}$
Y^1_{Rich}	Y^0_{Rich}	$A_{Rich} = 0$		Y^0_{Rich}
Y^1_{Sarah}	Y^0_{Sarah}	$A_{Sarah} = 1$	Y^1_{Sarah}	
$Y^1_{Alondra}$	$Y^0_{Alondra}$	$A_{Alondra} = 0$		$Y^0_{Alondra}$
$Y^1_{Jesús}$	$Y^0_{Jesús}$	$A_{Jesús} = 1$	$Y^1_{Jesús}$	

We learned to define and identify causal effects

We learned about DAGs

- ▶ Nodes and paths
- ▶ Structures

Causal path	$\bullet \rightarrow \bullet \rightarrow \bullet$
Fork	$\bullet \leftarrow \bullet \rightarrow \bullet$
Collider	$\bullet \rightarrow \bullet \leftarrow \bullet$

- ▶ Selecting a sufficient adjustment set

We learned models for description

Estimator 1: Subgroup sample mean

Player Salary

Dodgers

Past Win−Loss Record
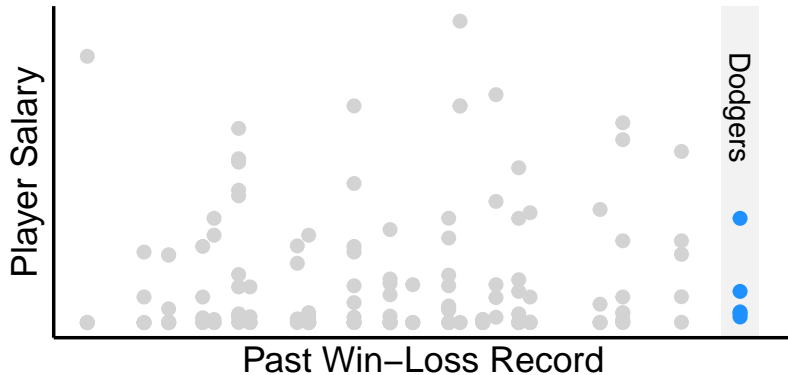
We learned models for description

Estimator 2: Full sample mean

Player Salary

Past Win−Loss Record

Dodgers
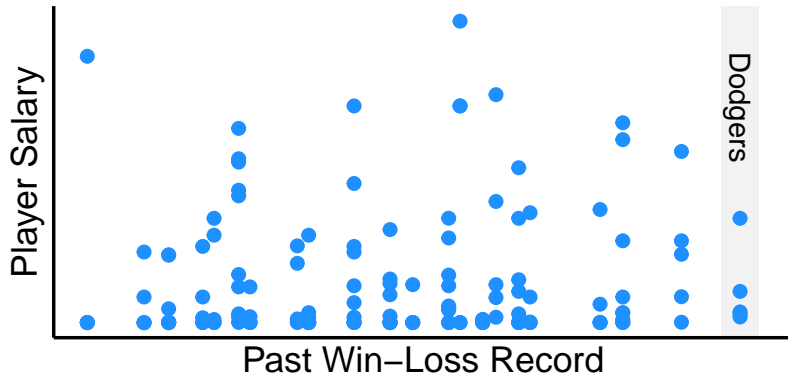
We learned models for description

Estimator 3: Regression prediction
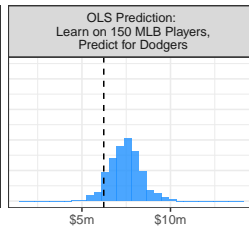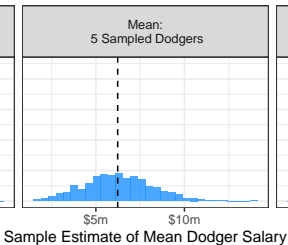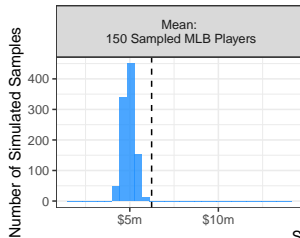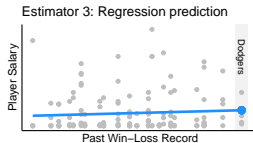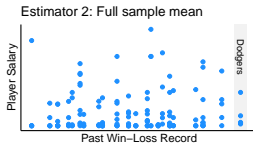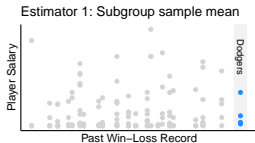
Player Salary

Dodgers

Past Win–Loss Record

We learned models for description

We learned models for causal inference

$Y_1^{\text{Takes 114}}$	?	
$Y_2^{\text{Takes 114}}$	?	
$Y_3^{\text{Takes 114}}$	?	
$Y_4^{\text{Takes 114}}$	?	
$Y_5^{\text{Takes 114}}$	?	
$Y_6^{\text{Takes 114}}$	?	
Outcome under 114	Outcome under no 114	

Each Row is a Student in This Class

We learned models for causal inference

Each Row is a Student in This Class

Outcome under 114	Outcome under no 114
$Y_1^{\text{Takes 114}}$	$\hat{Y}_1^{\text{No 114}}$
$Y_2^{\text{Takes 114}}$	$\hat{Y}_2^{\text{No 114}}$
$Y_3^{\text{Takes 114}}$	$\hat{Y}_3^{\text{No 114}}$
$Y_4^{\text{Takes 114}}$	$\hat{Y}_4^{\text{No 114}}$
$Y_5^{\text{Takes 114}}$	$\hat{Y}_5^{\text{No 114}}$
$Y_6^{\text{Takes 114}}$	$\hat{Y}_6^{\text{No 114}}$

We learned models for causal inference

Outcome under 114	Outcome under no 114
$Y_1^{\text{Takes 114}}$	$\hat{Y}_1^{\text{No 114}}$
$Y_2^{\text{Takes 114}}$	$\hat{Y}_2^{\text{No 114}}$
$Y_3^{\text{Takes 114}}$	$\hat{Y}_3^{\text{No 114}}$
$Y_4^{\text{Takes 114}}$	$\hat{Y}_4^{\text{No 114}}$
$Y_5^{\text{Takes 114}}$	$\hat{Y}_5^{\text{No 114}}$
$Y_6^{\text{Takes 114}}$	$\hat{Y}_6^{\text{No 114}}$

Each Row is a Student in This Class

General approach

We learned models for causal inference

Outcome under 114	Outcome under no 114
$Y_1^{\text{Takes 114}}$	$\hat{Y}_1^{\text{No 114}}$
$Y_2^{\text{Takes 114}}$	$\hat{Y}_2^{\text{No 114}}$
$Y_3^{\text{Takes 114}}$	$\hat{Y}_3^{\text{No 114}}$
$Y_4^{\text{Takes 114}}$	$\hat{Y}_4^{\text{No 114}}$
$Y_5^{\text{Takes 114}}$	$\hat{Y}_5^{\text{No 114}}$
$Y_6^{\text{Takes 114}}$	$\hat{Y}_6^{\text{No 114}}$

Each Row is a Student in This Class

General approach

1) Define potential outcomes

We learned models for causal inference

	Each Row is a Student in This Class	$Y_1^{\text{Takes 114}}$	$\hat{Y}_1^{\text{No 114}}$
		$Y_2^{\text{Takes 114}}$	$\hat{Y}_2^{\text{No 114}}$
		$Y_3^{\text{Takes 114}}$	$\hat{Y}_3^{\text{No 114}}$
		$Y_4^{\text{Takes 114}}$	$\hat{Y}_4^{\text{No 114}}$
		$Y_5^{\text{Takes 114}}$	$\hat{Y}_5^{\text{No 114}}$
		$Y_6^{\text{Takes 114}}$	$\hat{Y}_6^{\text{No 114}}$
		Outcome under 114	Outcome under no 114

General approach

1) Define potential outcomes
2) Define target population

We learned models for causal inference

Each Row is a Student in This Class

$Y_1^{\text{Takes 114}}$	$\hat{Y}_1^{\text{No 114}}$
$Y_2^{\text{Takes 114}}$	$\hat{Y}_2^{\text{No 114}}$
$Y_3^{\text{Takes 114}}$	$\hat{Y}_3^{\text{No 114}}$
$Y_4^{\text{Takes 114}}$	$\hat{Y}_4^{\text{No 114}}$
$Y_5^{\text{Takes 114}}$	$\hat{Y}_5^{\text{No 114}}$
$Y_6^{\text{Takes 114}}$	$\hat{Y}_6^{\text{No 114}}$

Outcome under 114

Outcome under no 114

General approach

1) Define potential outcomes

2) Define target population

3) Make causal assumptions

We learned models for causal inference

General approach

1) Define potential outcomes
2) Define target population
3) Make causal assumptions
4) Model unobserved outcomes

We learned models for causal inference

Each Row is a Student in This Class

$Y_1^{\text{Takes 114}}$	$\hat{Y}_1^{\text{No 114}}$
$Y_2^{\text{Takes 114}}$	$\hat{Y}_2^{\text{No 114}}$
$Y_3^{\text{Takes 114}}$	$\hat{Y}_3^{\text{No 114}}$
$Y_4^{\text{Takes 114}}$	$\hat{Y}_4^{\text{No 114}}$
$Y_5^{\text{Takes 114}}$	$\hat{Y}_5^{\text{No 114}}$
$Y_6^{\text{Takes 114}}$	$\hat{Y}_6^{\text{No 114}}$
Outcome under 114	Outcome under no 114

General approach

1) Define potential outcomes

2) Define target population

3) Make causal assumptions

4) Model unobserved outcomes

5) Predict them

We learned models for causal inference

$Y_1^{\text{Takes 114}}$	$\hat{Y}_1^{\text{No 114}}$
$Y_2^{\text{Takes 114}}$	$\hat{Y}_2^{\text{No 114}}$
$Y_3^{\text{Takes 114}}$	$\hat{Y}_3^{\text{No 114}}$
$Y_4^{\text{Takes 114}}$	$\hat{Y}_4^{\text{No 114}}$
$Y_5^{\text{Takes 114}}$	$\hat{Y}_5^{\text{No 114}}$
$Y_6^{\text{Takes 114}}$	$\hat{Y}_6^{\text{No 114}}$
Outcome under 114	Outcome under no 114

Each Row is a Student in This Class

General approach

1) Define potential outcomes

2) Define target population

3) Make causal assumptions

4) Model unobserved outcomes

5) Predict them

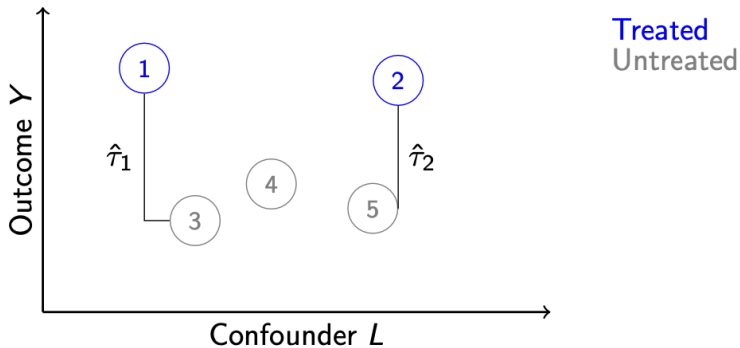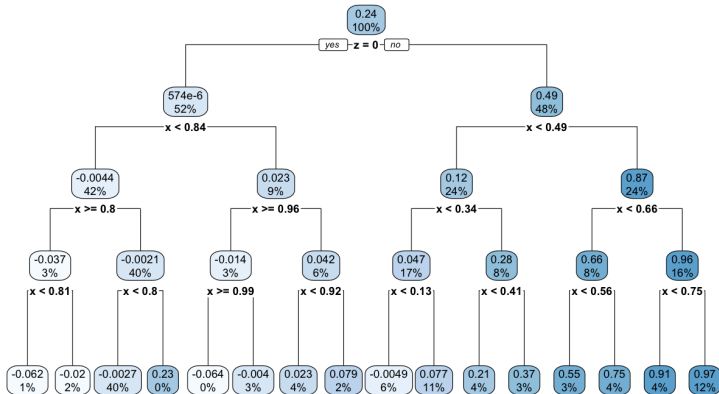6) Report an average

We learned about matching

We learned about trees

We learned lots of ways to predict Y
given a set of predictors $\vec{X}$

(supervised machine learning)

You put it all together in a project

You put it all together in a project

You got data

You put it all together in a project

You got data

- ▶ You visualized your findings
- ▶ You answered a descriptive and causal question
- ▶ You presented what you found!

Feedback

1. What is your favorite thing you learned this quarter?

Feedback

1. What is your favorite thing you learned this quarter?
2. What worked especially well in the course structure / topics?

Feedback

1. What is your favorite thing you learned this quarter?
2. What worked especially well in the course structure / topics?
3. How could this course be even better?

Course evaluations

Course evaluations

- Specific examples are especially helpful
- Especially specific experiences with your TA

Thank you for a great quarter!