Data-driven methods
Machine learning

INFO/STSCI/ILRST 3900: Causal Inference

16 Nov 2023

Learning goals for today

At the end of class, you will have intuition for how sample splitting makes it easier to

1. choose among many estimands
2. choose among many estimators
3. develop new data science approaches

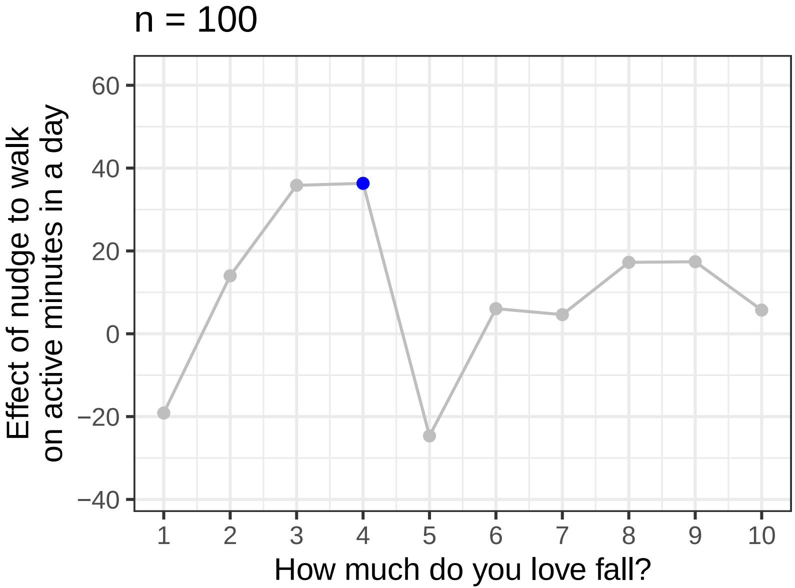
Targeted treatments

Want to find the subgroup
with the **biggest** causal effect

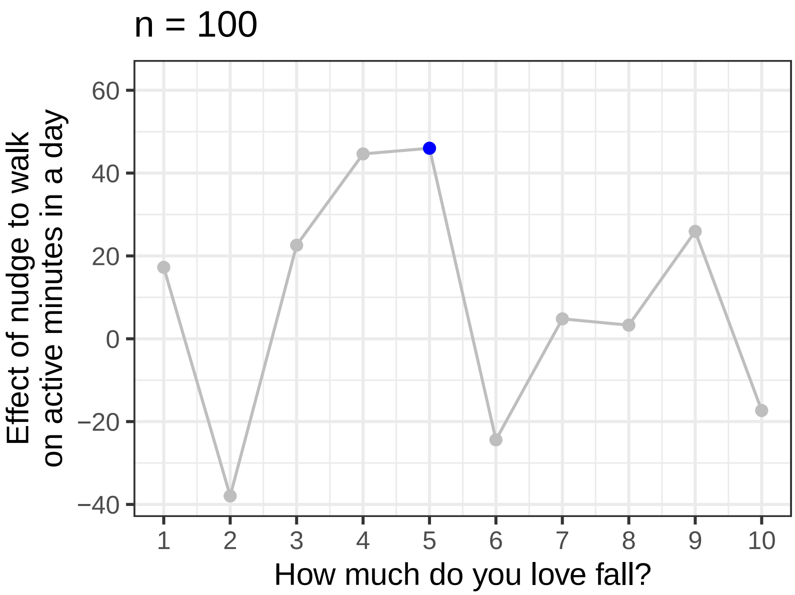
Targeted treatments

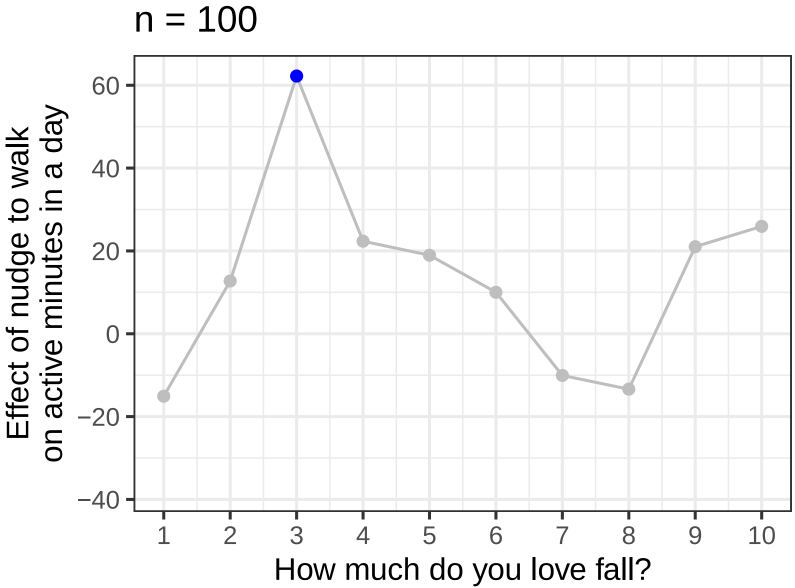
Want to find the subgroup
with the **biggest** causal effect

WATCH
fitbit
GARMIN



n = 100
Effect of nudge to walk
on active minutes in a day
60
40
20
0
−20
−40
1
2
3
4
5
6
7
8
9
10
How much do you love fall?


n = 100
Effect of nudge to walk
on active minutes in a day
60
40
20
0
−20
−40
1
2
3
4
5
6
7
8
9
10
How much do you love fall?


n = 100
Effect of nudge to walk
on active minutes in a day
60
40
20
0
−20
−40
1
2
3
4
5
6
7
8
9
10
How much do you love fall?

When you pick the biggest effect,
you select high positive **noise**

$$\hat{\tau}_x = \underbrace{\tau_x}_{\substack{\textbf{signal} \\ \text{true} \\ \text{effect}}} + \underbrace{\epsilon_x}_{\substack{\textbf{noise} \\ \text{sampling} \\ \text{variability}}}$$

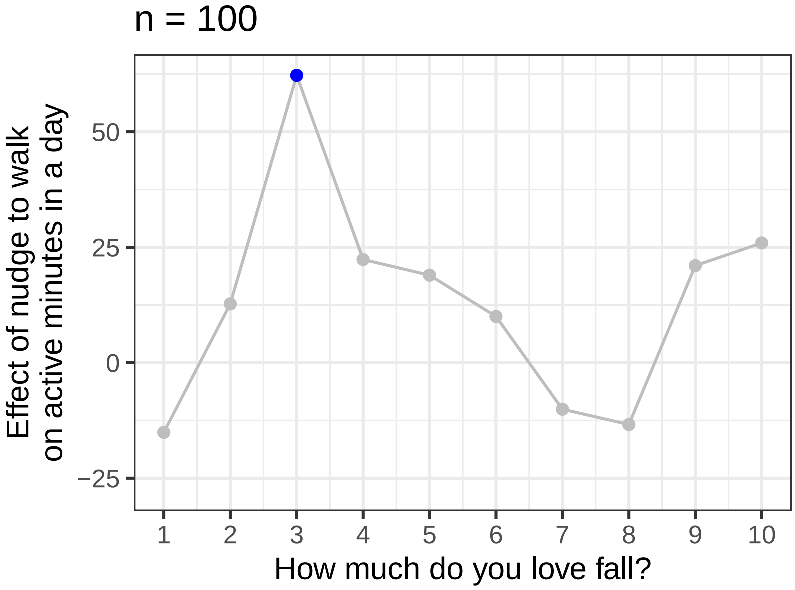
When you pick the biggest effect,
you select high positive **noise**

$$\hat{\tau}_x = \underbrace{\tau_x}_{\substack{\textbf{signal}\\ \text{true}\\ \text{effect}}} + \underbrace{\epsilon_x}_{\substack{\textbf{noise}\\ \text{sampling}\\ \text{variability}}}$$

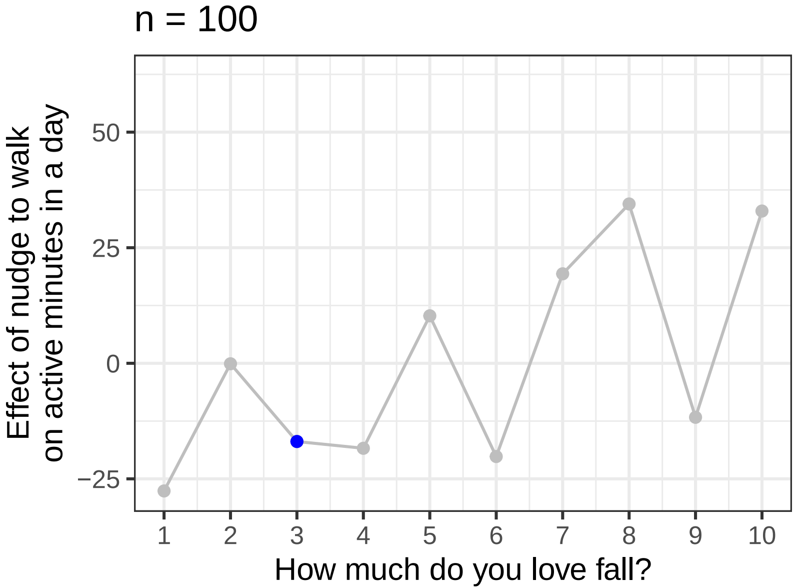
Solution: Two samples

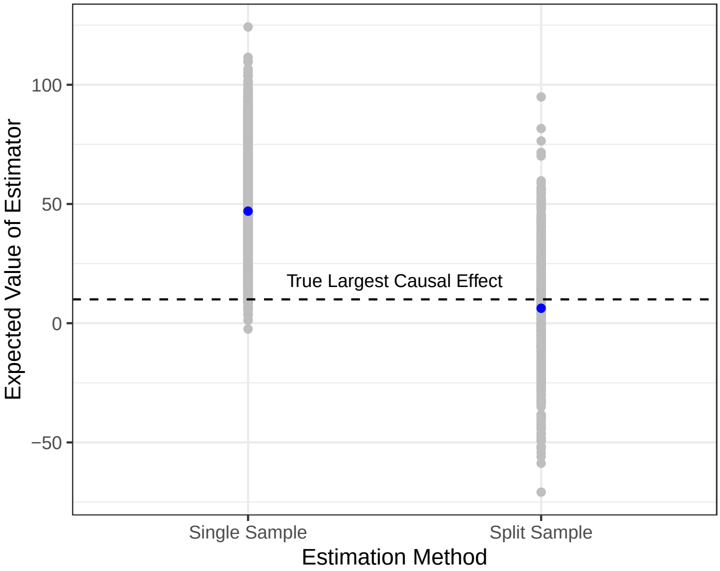
- select the X-value with the biggest effect
- estimate the effect in that subgroup

Selection sample

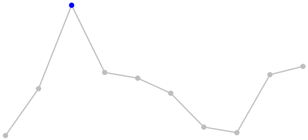


Estimation sample

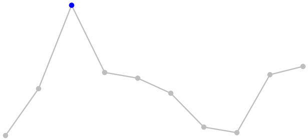




Expected Value of Estimator
100
50
0
−50
True Largest Causal Effect
Single Sample
Split Sample
Estimation Method

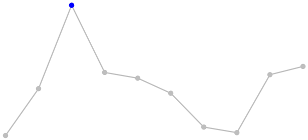


Where does the pick-the-largest estimator occur in practice?



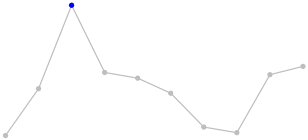
Where does the pick-the-largest estimator occur in practice?

- a data scientist searchers for effect heterogeneity



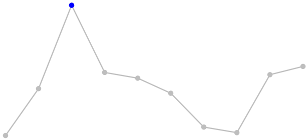
Where does the pick-the-largest estimator occur in practice?

- a data scientist searchers for effect heterogeneity
- they are excited about the biggest effect

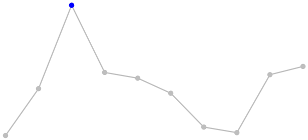


Where does the pick-the-largest estimator occur in practice?

- a data scientist searchers for effect heterogeneity
- they are excited about the biggest effect
- but it is really just noise

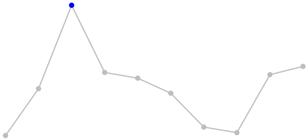


Where does the pick-the-largest estimator occur in practice?



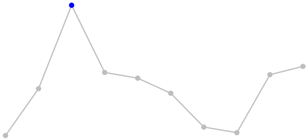
Where does the pick-the-largest estimator occur in practice?

- many psychology labs run experiments



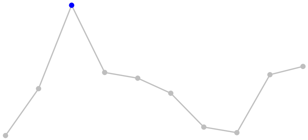
Where does the pick-the-largest estimator occur in practice?

- many psychology labs run experiments
- the biggest effect estimate gets published

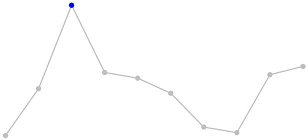


Where does the pick-the-largest estimator occur in practice?

- many psychology labs run experiments
- the biggest effect estimate gets published
- but it is really just noise

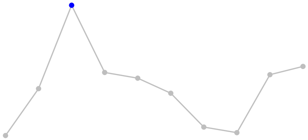


Where does the pick-the-largest estimator occur in practice?



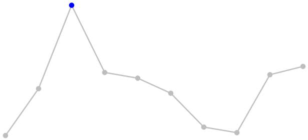
Where does the pick-the-largest estimator occur in practice?

- one researcher considers many estimators of one effect



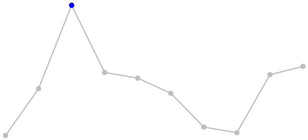
Where does the pick-the-largest estimator occur in practice?

- one researcher considers many estimators of one effect
- they like the biggest effect estimate

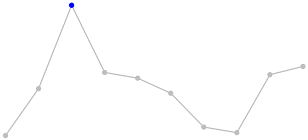


Where does the pick-the-largest estimator occur in practice?

- one researcher considers many estimators of one effect
- they like the biggest effect estimate
- but it is really just noise



whenever you make many estimates
but report only one
you are at risk of this problem



whenever you make many estimates
but report only one
you are at risk of this problem

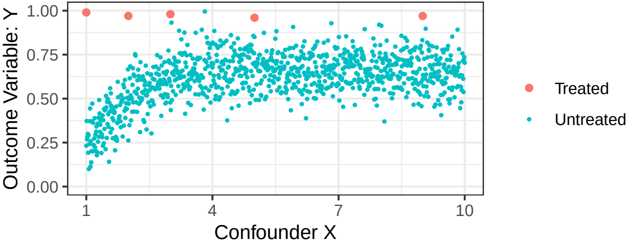
Sample splitting is an answer!

Learning goals for today

At the end of class, you will have intuition for how sample splitting makes it easier to

1. choose among many estimands
2. choose among many estimators
3. develop new data science approaches

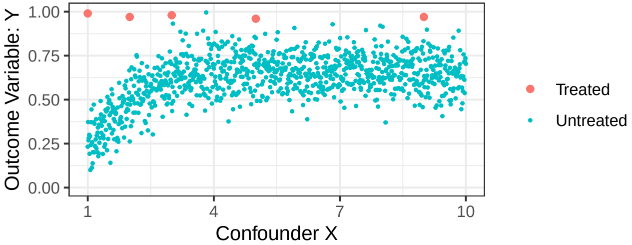
Sample splitting makes it easier to **choose among many estimators**



Estimand: Average Treatment Effect on the Treated

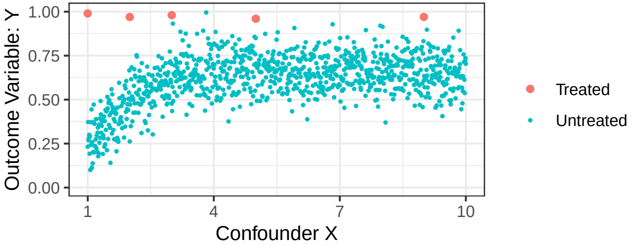
$$\mathrm{E}(Y^1 - Y^0 \mid A = 1) = \frac{1}{n_{A=1}} \sum_{i:A_i=1} \left(Y_i^1 - Y_i^0 \right)$$

Sample splitting makes it easier to **choose among many estimators**



$$\begin{tikzcd} X \arrow[r] \arrow[rr, bend right] & A \arrow[r] & Y \end{tikzcd}$$

Sample splitting makes it easier to **choose among many estimators**

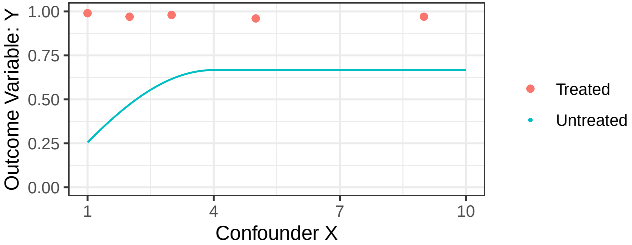


$$X \longrightarrow A \longrightarrow Y, \qquad X \longrightarrow Y$$

$$\mathrm{E}(Y^1 \mid A = 1) = \frac{1}{n_{A=1}} \sum_{i:A_i=1} Y_i$$

$$\mathrm{E}(Y^0 \mid A = 1) = \frac{1}{n_{A=1}} \sum_{i:A_i=1} \underbrace{\mathrm{E}(Y \mid A = 0, X = x_i)}_{\text{Need to estimate}}$$

Sample splitting makes it easier to **choose among many estimators**

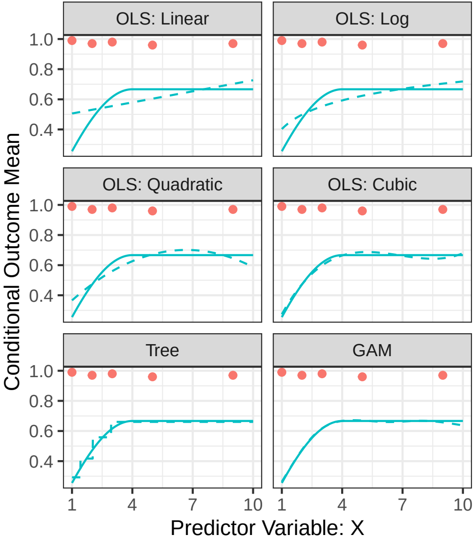


$$
\begin{tikzcd}
X \arrow[r] \arrow[rr, bend right] & A \arrow[r] & Y
\end{tikzcd}
$$

$$\mathrm{E}(Y^1 \mid A = 1) = \frac{1}{n_{A=1}} \sum_{i:A_i=1} Y_i$$

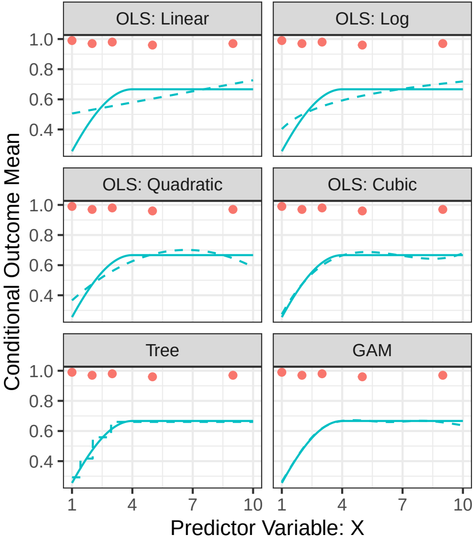
$$\mathrm{E}(Y^0 \mid A = 1) = \frac{1}{n_{A=1}} \sum_{i:A_i=1} \underbrace{\mathrm{E}(Y \mid A = 0, X = x_i)}_{\text{Need to estimate}}$$

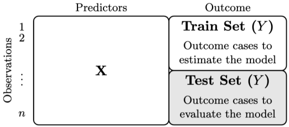
Sample splitting makes it easier to **choose among many estimators**



Example based on
[Lundberg, Brand, Jeon 2022](#)
Figure 5

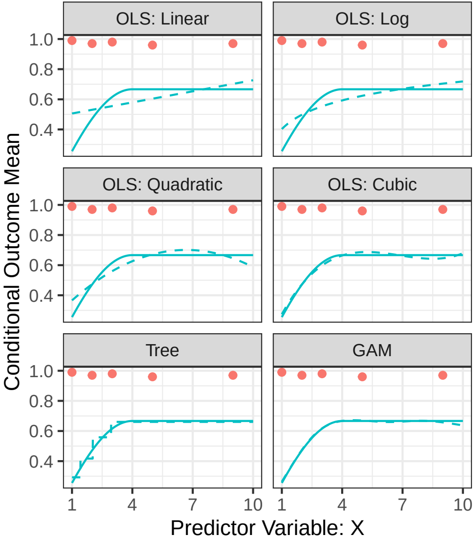
Sample splitting makes it easier to **choose among many estimators**

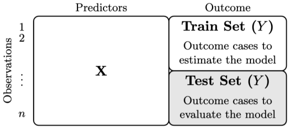


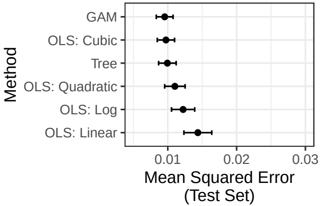


Example based on
Lundberg, Brand, Jeon 2022
Figure 5

Sample splitting makes it easier to **choose among many estimators**







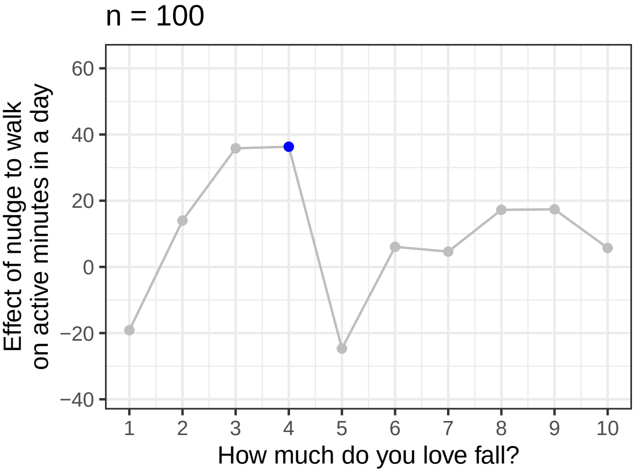
Example based on
Lundberg, Brand, Jeon 2022
Figure 5

Learning goals for today

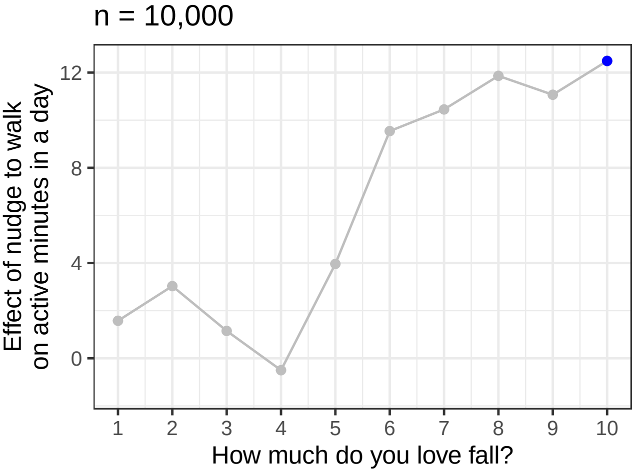
At the end of class, you will have intuition for how sample splitting makes it easier to

1. choose among many estimands
2. choose among many estimators
3. develop new data science approaches

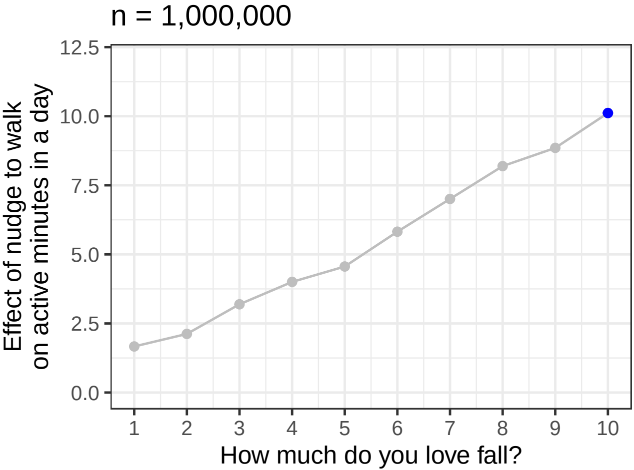
Sample splitting makes it easier to
develop new data science approaches



Sample splitting makes it easier to
develop new data science approaches



Sample splitting makes it easier to **develop new data science approaches**



Sample splitting makes it easier to **develop new data science approaches**

How much do you love fall?

1 10

Sample splitting makes it easier to **develop new data science approaches**

How much do
you love fall?

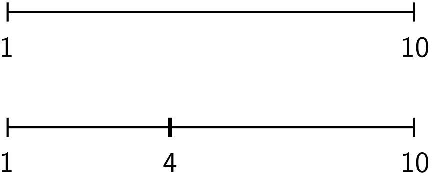
repeatedly split into two halves
with highly heterogeneous
treatment effects

1 10

Sample splitting makes it easier to **develop new data science approaches**

How much do you love fall?

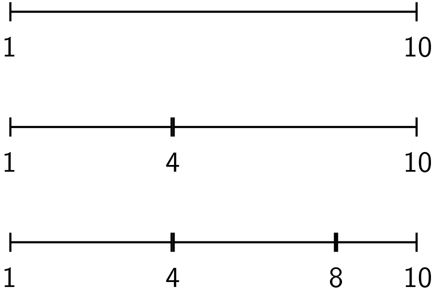
repeatedly split into two halves with highly heterogeneous treatment effects



Sample splitting makes it easier to **develop new data science approaches**

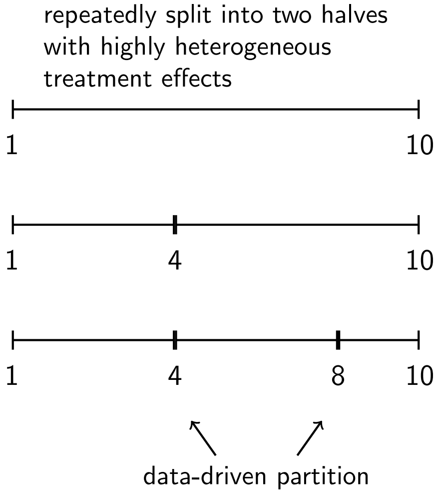
How much do you love fall?

repeatedly split into two halves with highly heterogeneous treatment effects



Sample splitting makes it easier to **develop new data science approaches**

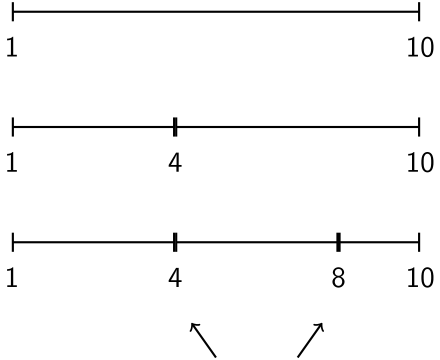
How much do you love fall?



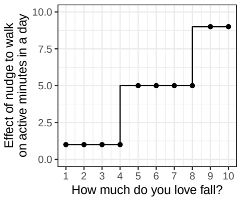
Sample splitting makes it easier to **develop new data science approaches**

How much do you love fall?

repeatedly split into two halves with highly heterogeneous treatment effects



data-driven partition

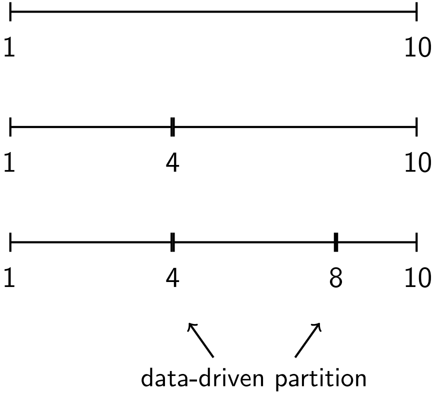


Sample splitting makes it easier to **develop new data science approaches**

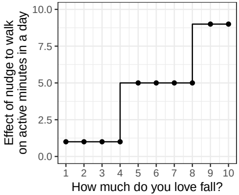
How much do you love fall?

In **Sample A**

repeatedly split into two halves with highly heterogeneous treatment effects



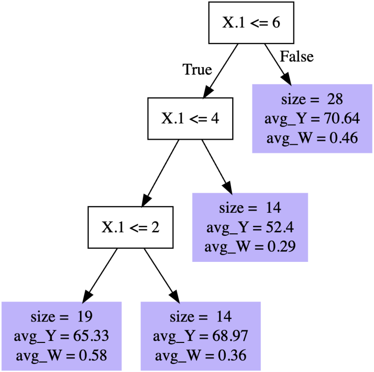
In **Sample B**
learn the response



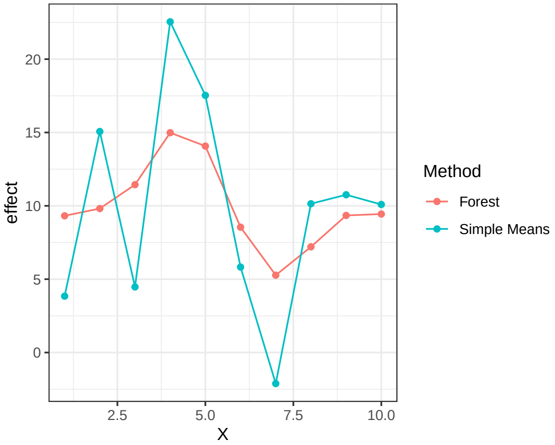
Sample splitting makes it easier to **develop new data science approaches**

Athey, S., & Imbens, G. (2016).
Recursive partitioning for heterogeneous causal effects.
Proceedings of the National Academy of Sciences
113(27), 7353-7360.

Sample splitting makes it easier to **develop new data science approaches**



Sample splitting makes it easier to **develop new data science approaches**



Learning goals for today

At the end of class, you will have intuition for how sample splitting makes it easier to

1. choose among many estimands
2. choose among many estimators
3. develop new data science approaches