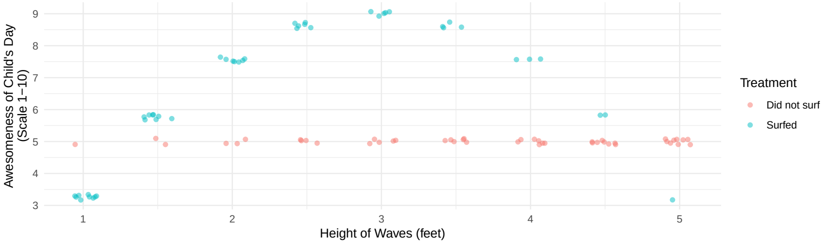
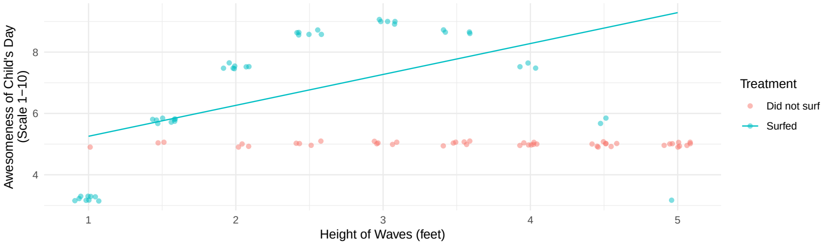
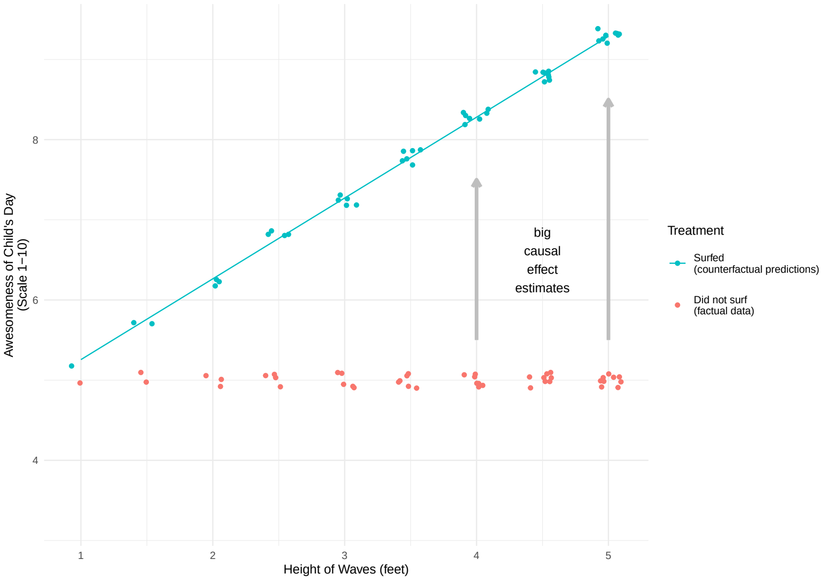
# Doubly Robust Slides

# A motivating example



$$\tau = \frac{1}{n_{\textsf{Not Surfed}}} \sum_{i:A_i=\textsf{Not Surfed}} \left(Y_i^{\textsf{Surfed}} - Y_i^{\textsf{Not Surfed}}\right)$$

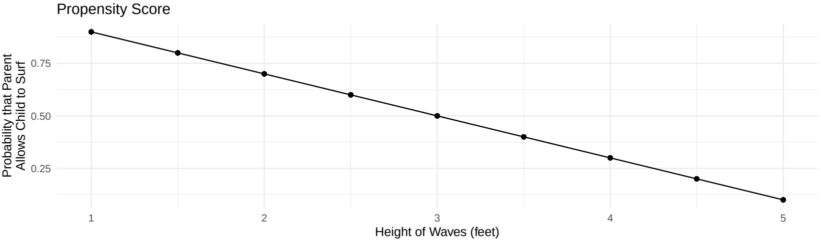

Awesomeness of Child's Day
(Scale 1–10)
8
6
4
1
2
3
4
5
Height of Waves (feet)
Treatment
Did not surf
Surfed



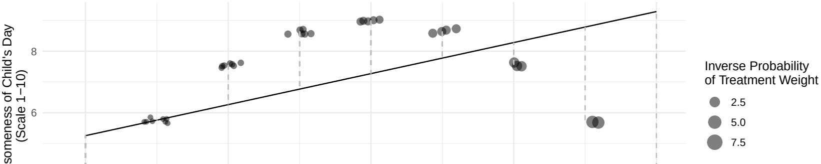
Awesomeness of Child's Day
(Scale 1–10)
8
6
4
1
2
3
4
5
Height of Waves (feet)
big
causal
effect
estimates
Treatment
Surfed
(counterfactual predictions)
Did not surf
(factual data)

We are always the child

- model is too simple for the world
- model is always wrong




To be even more concrete, consider 5-foot wave days. The child knows that they surfed only 10% of these days. Every day surfed really stands in for 9 days not surfed. If we were estimating the average treatment effect on the untreated by weighting, we would weight the 5-foot day by $\mathsf{P}(\text{Not Surfed} \mid X = 5)/\mathsf{P}(\text{Surfed} \mid X = 5) = .9/.1 = 9$. We therefore might estimate the weighted error of the linear model, weighted by these weights.



## Augmented inverse probability weighting

If the child corrected for the error, they would be using an augmented inverse probability weighting estimator.

$$\hat{\tau}_{\text{AIPW}} = \frac{1}{n_{\text{Not Surfed}}} \sum_{i:A_i=\text{Not Surfed}} \left(\hat{Y}_i^1 - Y_i\right) - \frac{\sum_{i:A_i=\text{Surfed}} w_i Y_i}{\sum_{i:A_i=\text{Surfed}} w_i}$$

where $w_i$ is the inverse probability weight for estimating the average treatment effect on the untreated.

$$w_i = \frac{\text{P}(A = \text{Not Surfed} \mid X = x_i)}{\text{P}(A = \text{Surfed} \mid X = x_i)}$$

In this example, the outcome model is misspecified (a line for a parabola) but the weights are correct. Below, we see that the weights allow us to correct the wrong outcome model. First, we calculate the truth because in the simulation we know the potential outcomes.

```
# A tibble: 1 x 1
```

# Targeted learning

AIPW is not the only way to update a model. Another option is called targeted learning (Van der Laan and Rose 2011). We first introduce targeted learning through one concrete example, then generalize the procedure in the sections that follows.

## Modeling $Y^1$ for the ATC

In the surfing example, our goal is to estimate the mean outcome under surfing, for the observations on days when there was surfing. We begin with the child's initial fit: linear regression. Following notation that is common in targeted learning,[1] we will refer to this regression line as $\hat{Q}^0$,

$$\underbrace{\hat{Q}^0(\vec{x})}_{\substack{\text{The 0 superscript}\\ \text{indicates an untargeted}\\ \text{initial estimate}}} \qquad = \hat{\mathsf{E}}(Y \mid A = 1, \vec{X}) = \hat{\alpha} + \hat{\beta}\vec{x}$$

where $\hat{\alpha}$ and $\hat{\beta}$ are the OLS coefficients when modeling $Y$ given $X$ among the treated observations.

# Sample splitting for machine learning

Under classical statistical approaches to inference, one often worries about model misspecification. The problem of model misspecification is that if one approximates $\text{E}(Y \mid A, \vec{X})$ by an additive regression and the true response function is nonlinear, then the model will be an inconsistent estimator of $\text{E}(Y \mid \vec{X} = \vec{x})$ for at least some $\vec{x}$. Double robustness solves this problem: as long as $\text{E}(Y \mid \vec{X})$ or $\text{P}(A = 1 \mid \vec{X})$ is estimated by a consistent estimator, then the causal effect estimate is consistent.

Machine learning approaches seem to upend this logic: flexible models can be consistent estimators by construction. Without any assumption of a statistical model, a random forest can yield a consistent estimator of $\text{E}(Y \mid A, \vec{X})$ and $\text{P}(A = 1 \mid \vec{X})$. By consistent, we mean that the forest would come to equal these estimands in an infinite sample. Does double robustness then have any use?

The problem with flexible machine learning estimators is that they converge to the true response surface at slower rates than

# Exercises

Carry out the child's analysis by

- an outcome model.
- inverse probability weighting, with weights estimated by logistic regression
- augmented inverse probability weighting