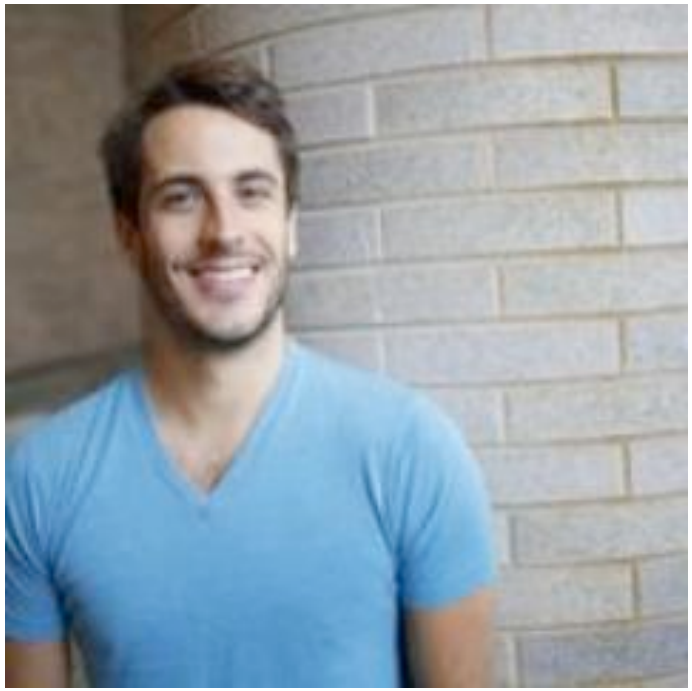
Change the World In a Day

Brian d'Alessandro
VP Data Science – m6d
Data Ambassador - DataKind

“The Best Minds of My Generation are Thinking About How to Make People Click Ads…That Sucks”



- Jeff Hammerbacher,
  Chief Scientist Cloudera

Ad Supported Companies Have Tools, Data and Resources.

Google

Y!

ebay

foursquare

Cassandra

HIVE

Linux

mahout

It just so happens that…

The Same Tools and Methods that can improve CTR, Search Relevancy and Product Recommendations can also…

Save Lives
Improve Cities
Empower Rural Villages

The Same Tools and Methods that can improve CTR, Search Relevancy and Product Recommendations can also…

Data Science is Highly Transferrable: From Business as Usual to Technical Charity

Improve Cities
Save Lives
Bring Fresh Water to People that Need It
Empower Rural Villages

Google | Giving

"…we help Googlers **match their skills** to **specific nonprofit needs**, and allow them to use up to **20 hours of work time** across the year to **volunteer**"

Is 20 hours enough?

Is 20 hours enough?

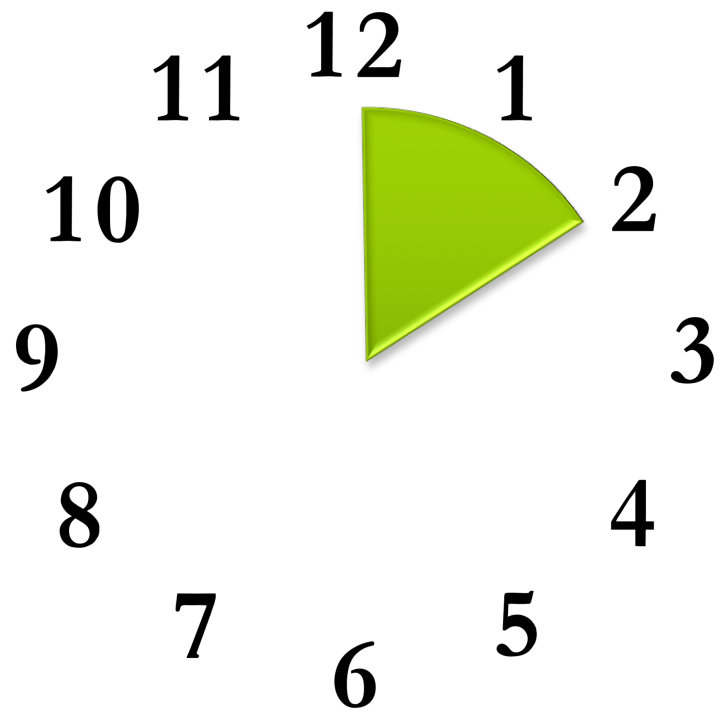
YES!!!

(Enter Stage Left)

The Wee Hours

Defining The Problem

Scoping the Project

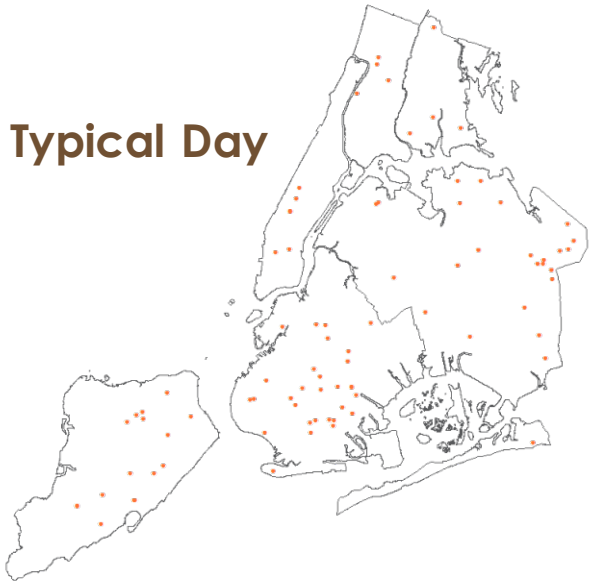


When Good Trees Turn Bad

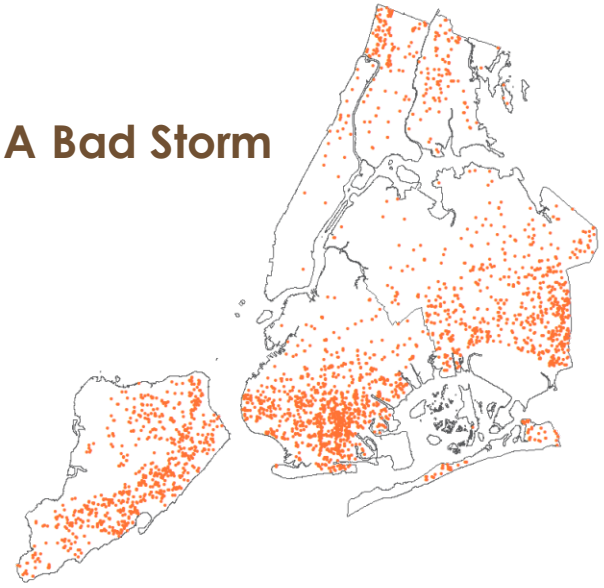
Can block pruning reduce future tree hazards?

Tree Hazards By Neighborhood

Typical Day



A Bad Storm



Data Source: 311 Requests, NYC Parks Forestry Management System

DS Translation

Quantify the effects that NYC Parks' pruning program has on reducing future tree hazards.

DS Translation

Quantify the effects that NYC Parks' pruning program has on reducing future tree hazards.

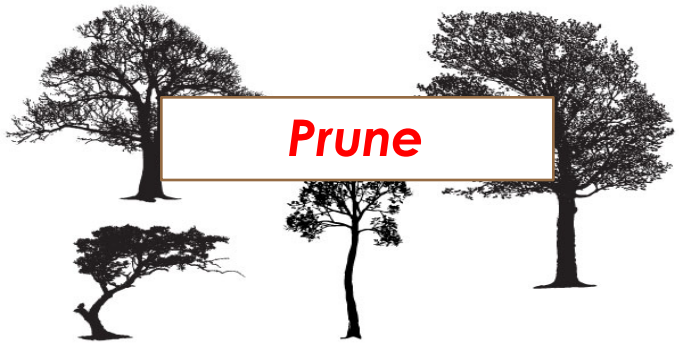
This is a question of CAUSALITY

What do ad conversions and fallen trees have in common?

Controlled Experimentation

Today

Future



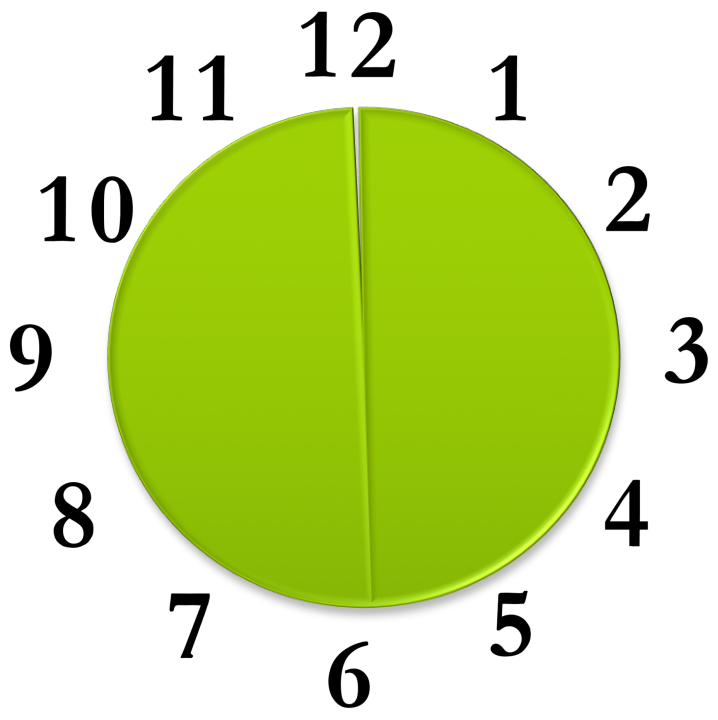
You can't A/B test in the past!

(But luckily we have data)

Observational methods allow us to statistically recreate an A/B test

The Data Wrangling Hours

Munging, Joining, Exploring



Block Level Data

Treatment

Year block was pruned

Features

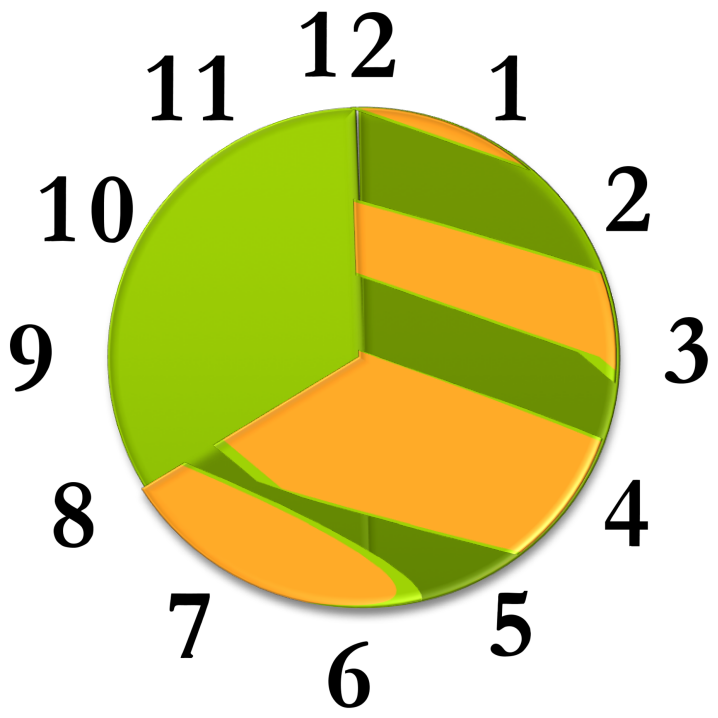
Number of trees on block
Number of trees by tree size
Number of trees by species type
Work Orders by type in Prior Years

Outcome

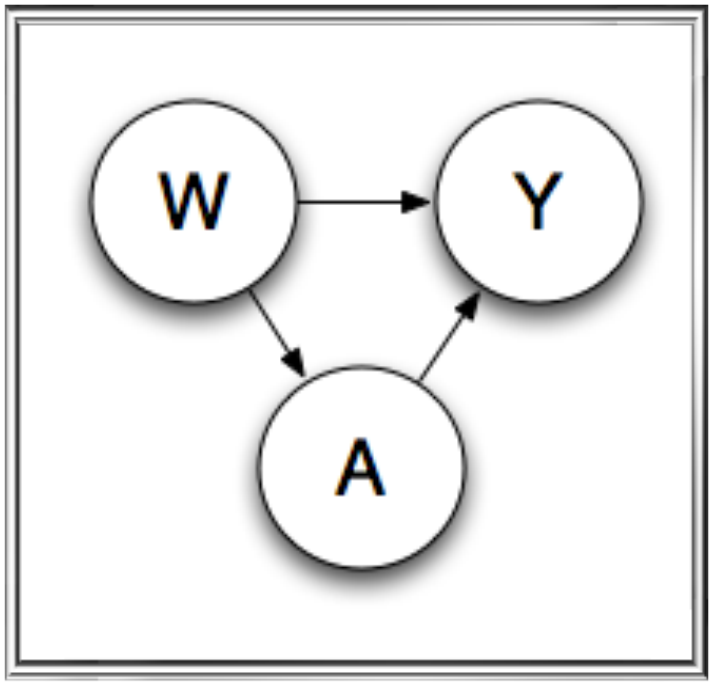
Hazardous Work Orders in Next Year

The Insight Hours

Statistical Programming & Analysis



Confounding



Y = Future Hazards
A = Block Pruning
W = All relevant info about Blocks

Certain characteristics that cause hazard conditions also make the blocks more likely to receive block pruning.

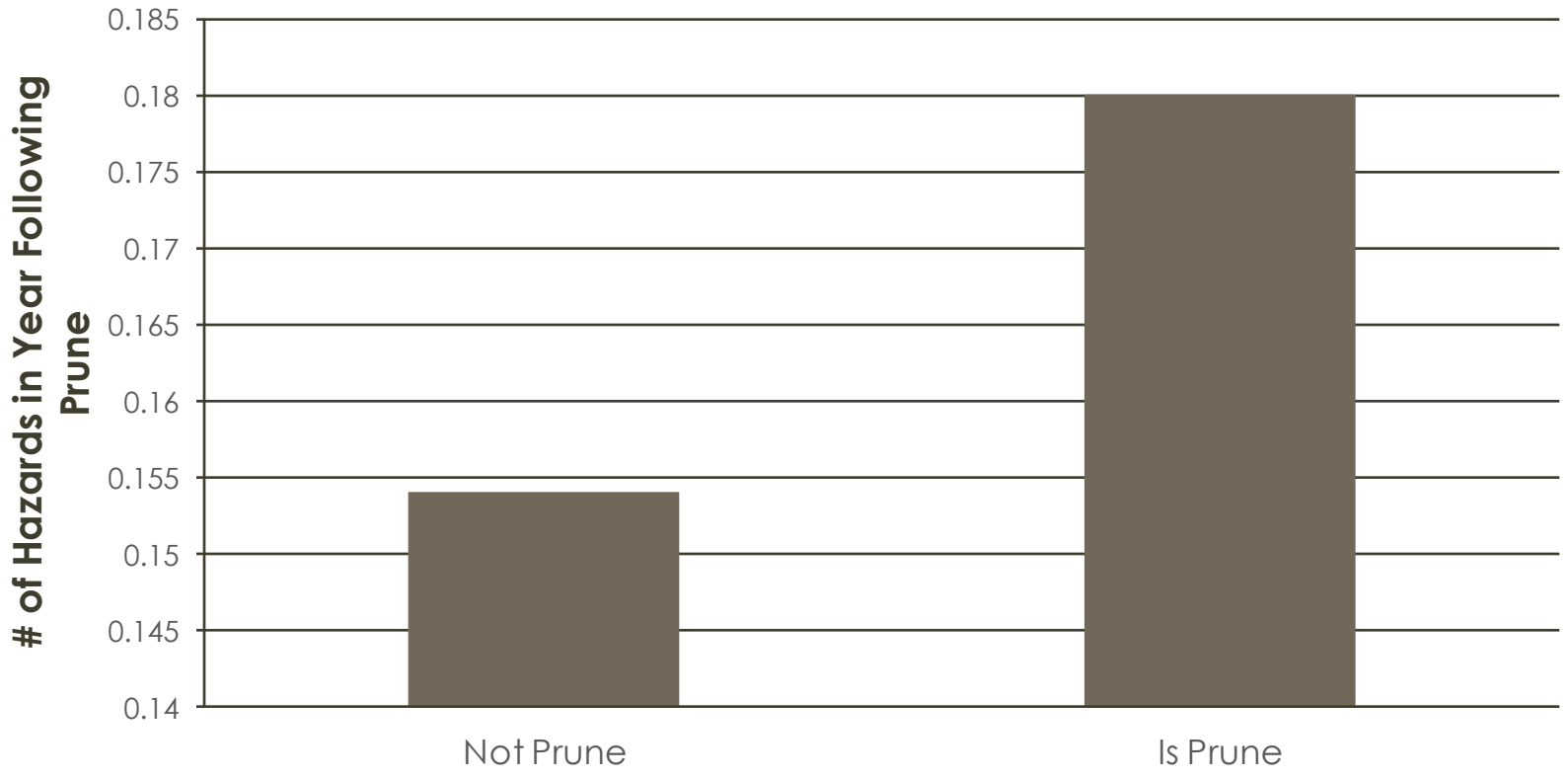
The key to observational methods is controlling for these confounders using statistical models and effectively recreating a control group that "looks" like the treated group.

A (Doubly) Robust Method for Observational Causal Effect Estimation

$$\psi^a_{n,A-IPTW} = \frac{1}{n}\sum_{i=1}^{n}\frac{I(A_i = a)}{g_{A_n}(A_i, W_i)}(Y_i - Q_{Y_n}(a, W_i)) + \frac{1}{n}\sum_{i=1}^{n}(Q_{Y_n}(a, W_i)).$$

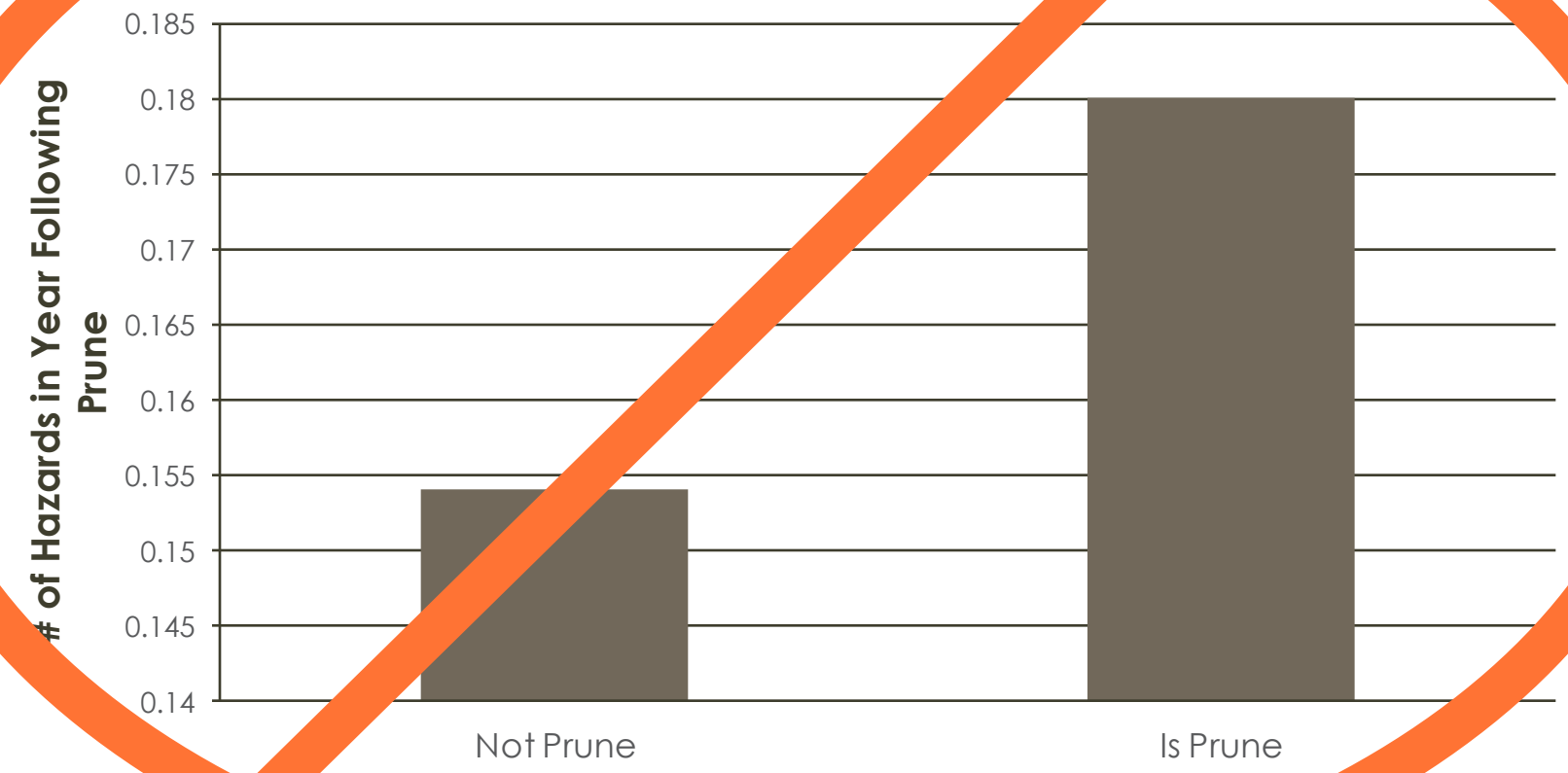
Naïve Results

This is the "causal" effect when you don't adjust for confounding?



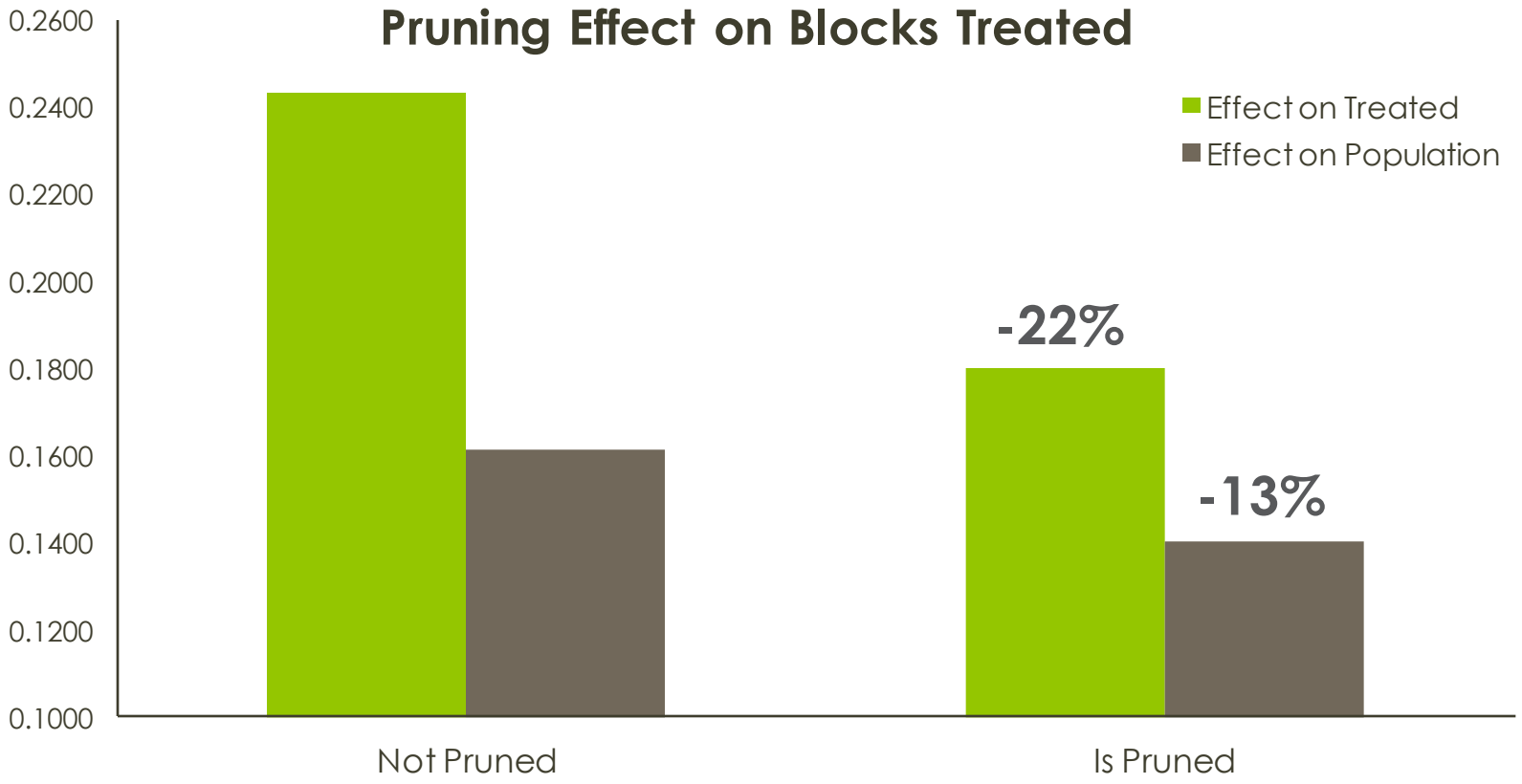
Naïve Results

This is the "causal" effect when you don't adjust for confounding?



Doubly Robust Results

Pruning on average reduced following year hazardous work orders on the blocks pruned by 22%.



**These are estimates made from 1000 bootstrapped samples of the data. Both estimates have p-values <0.001 in a standard two-way t-test.*

In ~ 20 hours of work time…

- Established baseline results for how an often endangered program improves living conditions of NYC
- Discovered that NYC Parks can optimize pruning ROI by targeting “at-risk” pruning sites

Beyond Trees…

What will it take to scalably enable and motivate large corporations to donate more Human Capital?

- Matching services (DataKind, Code for America, etc.)
- Motivating Data Scientists
- Motivating Corporations

With a small donation of your employee's time…

Have your Click and Eat it Too!