

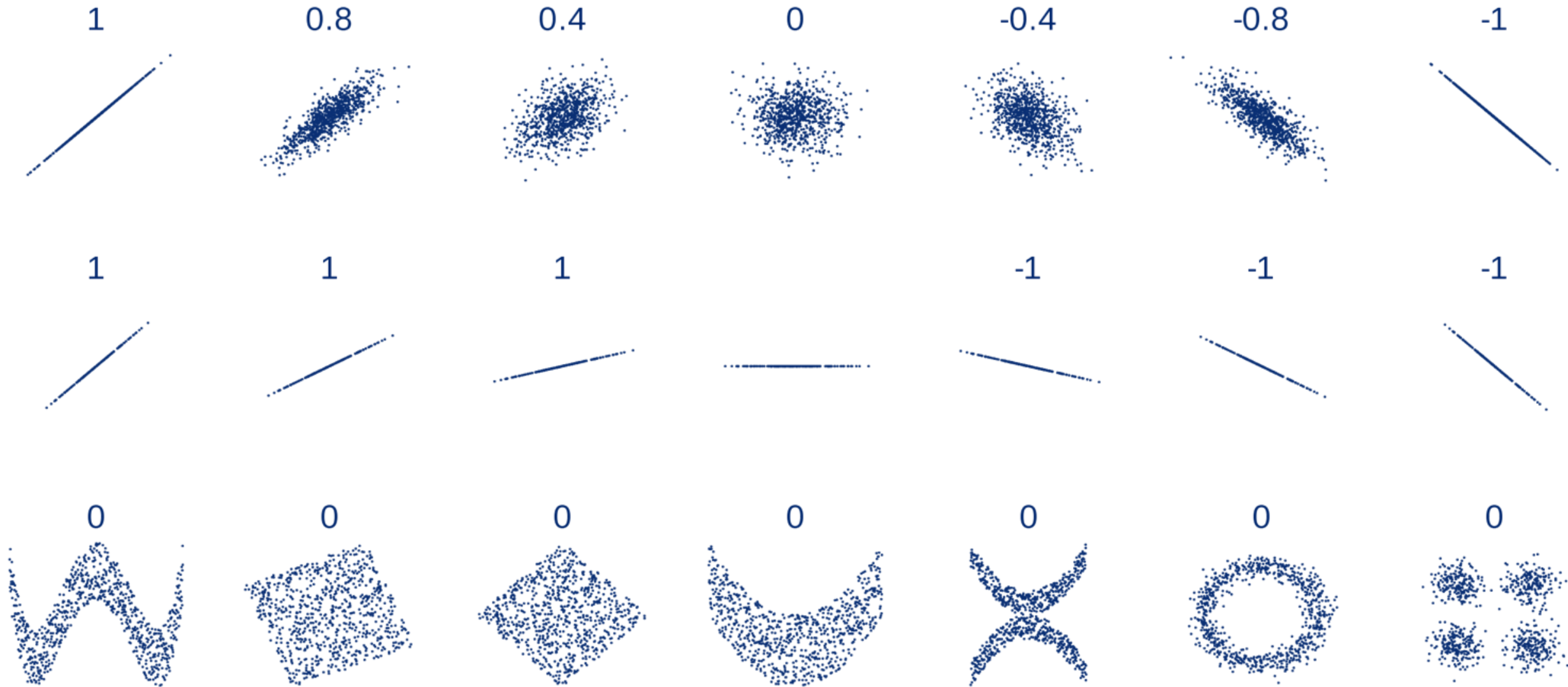


# CAUSALITY VERSUS CORRELATION IN THE READY-MIX INDUSTRY THROUGH USE OF IOT SYSTEMS

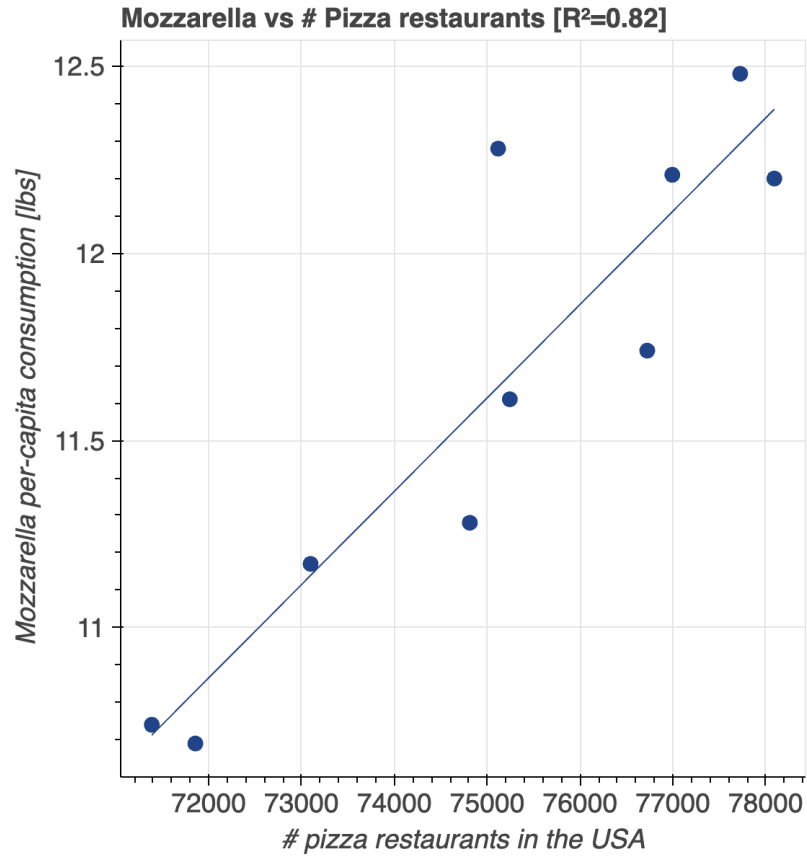
NATHAN TREGGER  
VERIFI, LLC



# CORRELATION – STATS RELATIONSHIP BETWEEN TWO RANDOM VARIABLES



# CORRELATION DOES NOT IMPLY CAUSATION...



foursquare.com/



ottopizza.blog

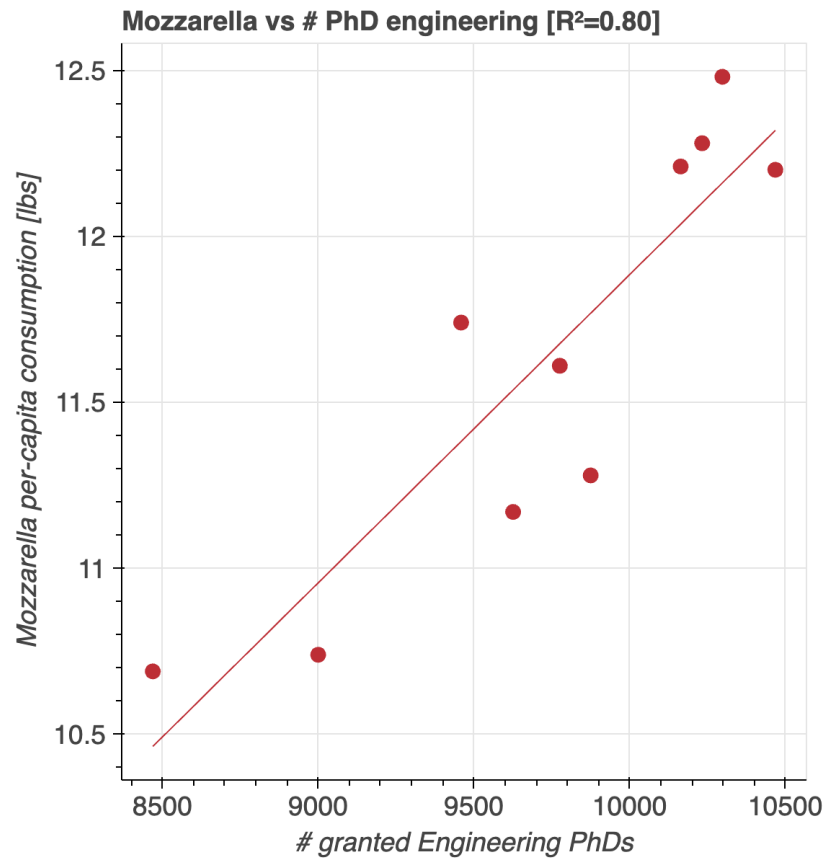
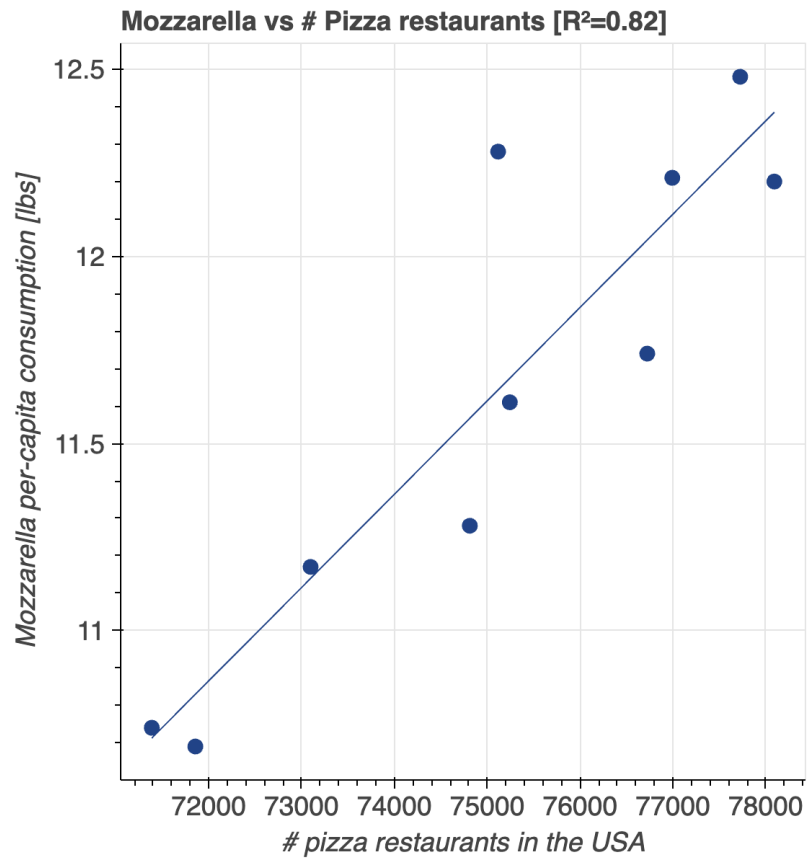


DALL-E (17 iterations)



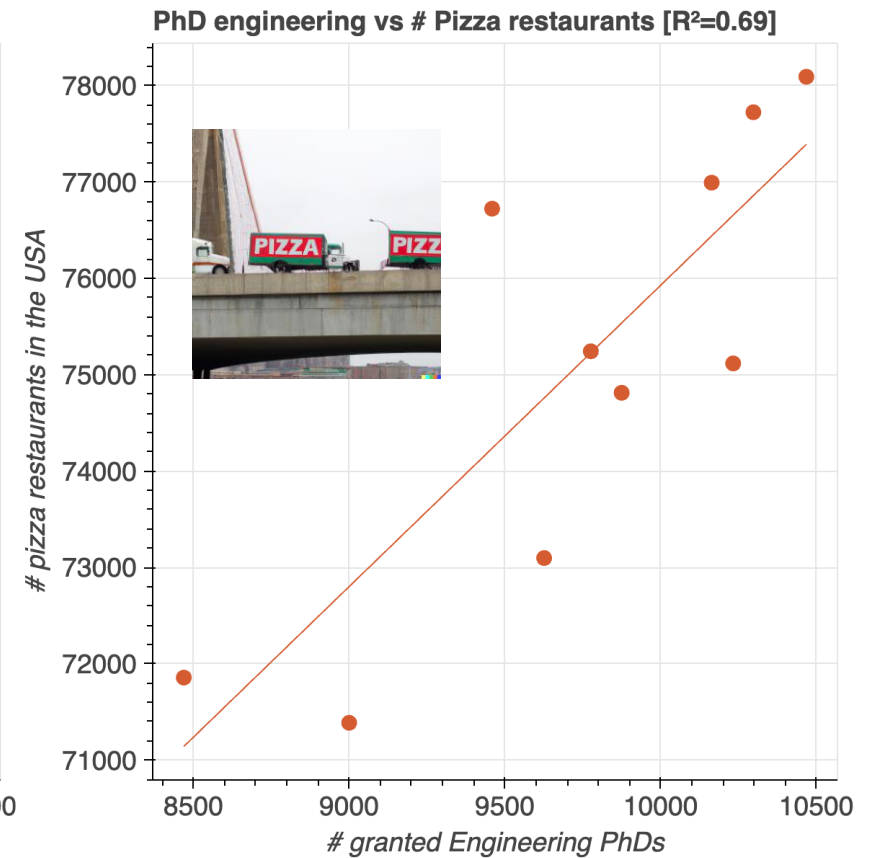
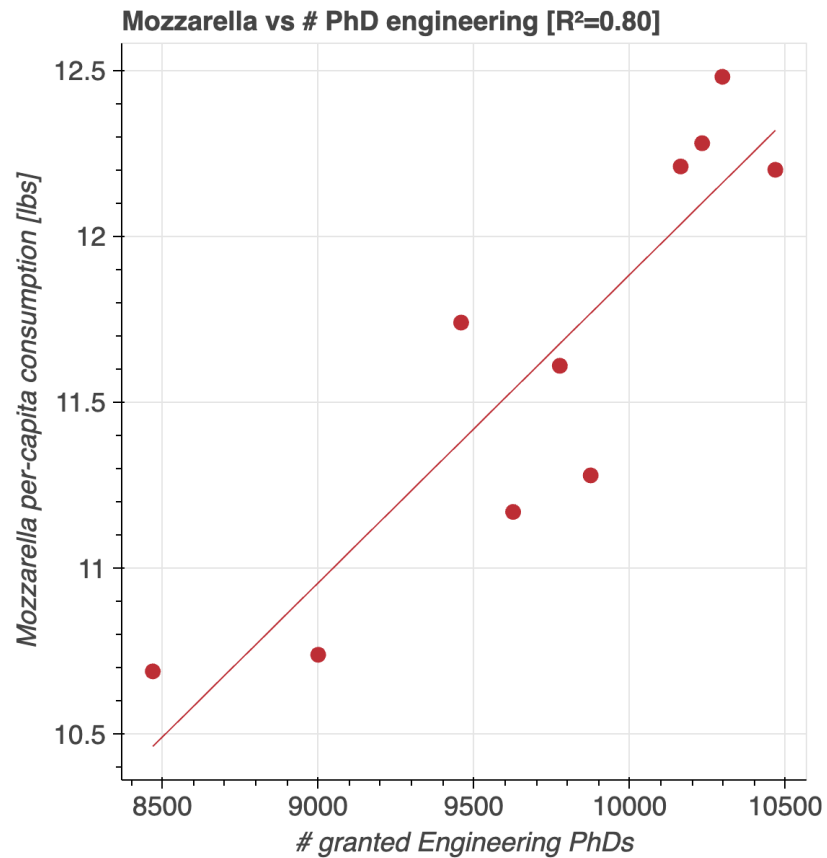
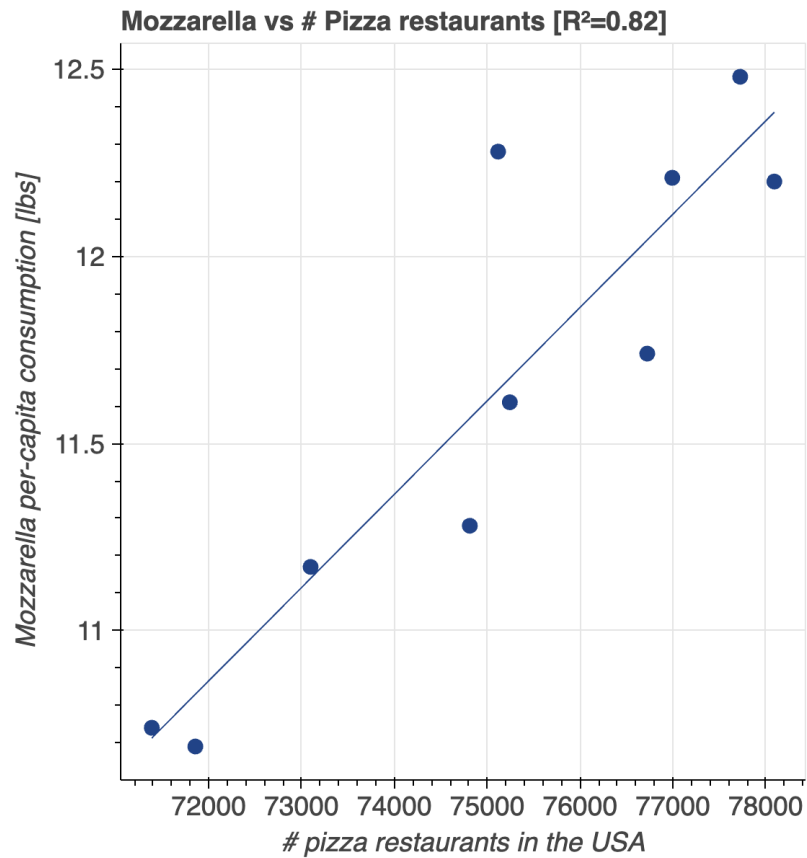
restaurantguru.com

# CORRELATION DOES NOT IMPLY CAUSATION...



DALL-E (30 iterations)

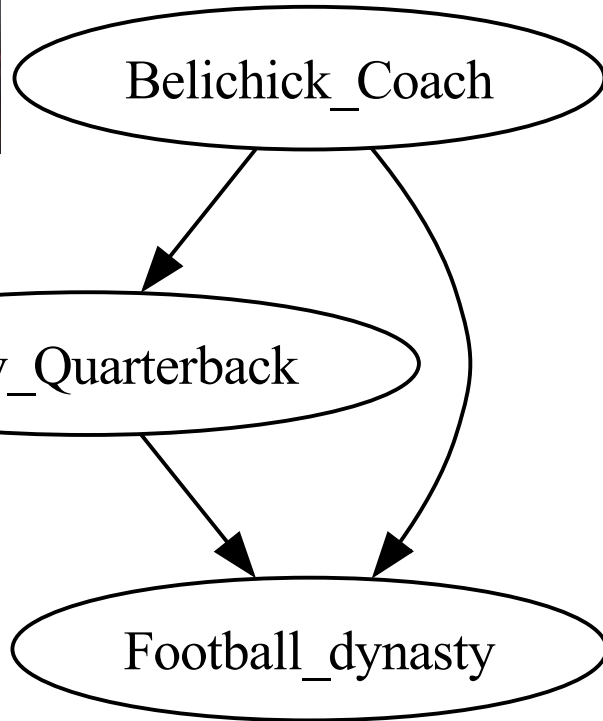
# CORRELATION DOES NOT IMPLY CAUSATION...



[statista.com/statistics/183948/per-capita-consumption-of-mozzarella-in-the-us-since-2000/](https://statista.com/statistics/183948/per-capita-consumption-of-mozzarella-in-the-us-since-2000/)  
[statista.com/statistics/377597/number-of-pizza-restaurants-us/](https://statista.com/statistics/377597/number-of-pizza-restaurants-us/)  
[nces.nsf.gov/pubs/nsf24300/data-tables#group3](https://nces.nsf.gov/pubs/nsf24300/data-tables#group3)

# CAUSAL MAPS – A WAY TO INTRODUCE THE CONCEPT OF CAUSATION

twincities.com



Arrows denote *cause*

Two immediate points:

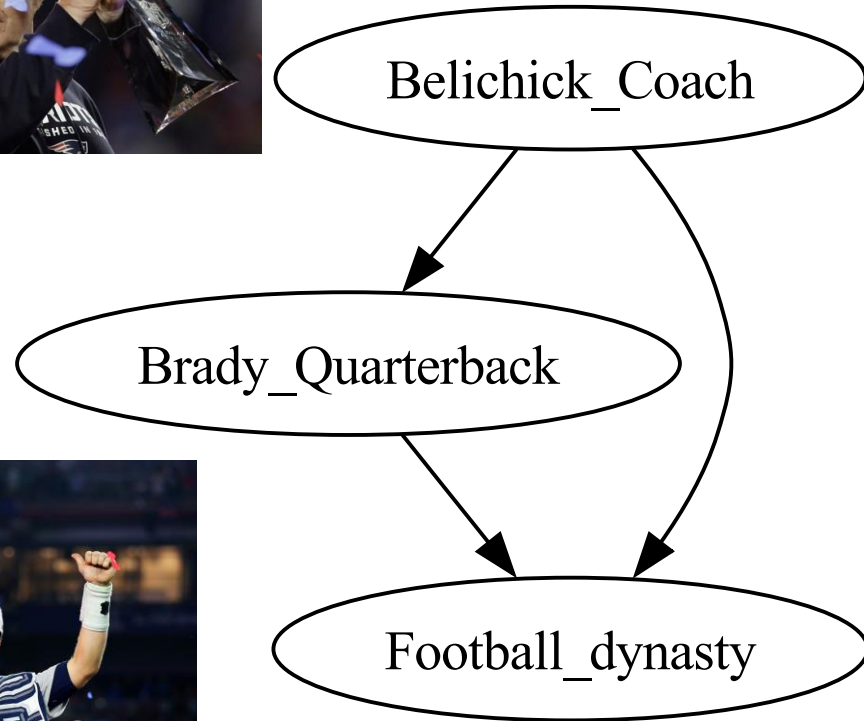
1. Depends on your expertise (i.e. domain knowledge)
2. Highlights confounding variables

sportscasting.com



# CAUSAL MAPS – A WAY TO INTRODUCE THE CONCEPT OF CAUSATION

twincities.com



sportscasting.com



Arrows denote *cause*

Two immediate points:

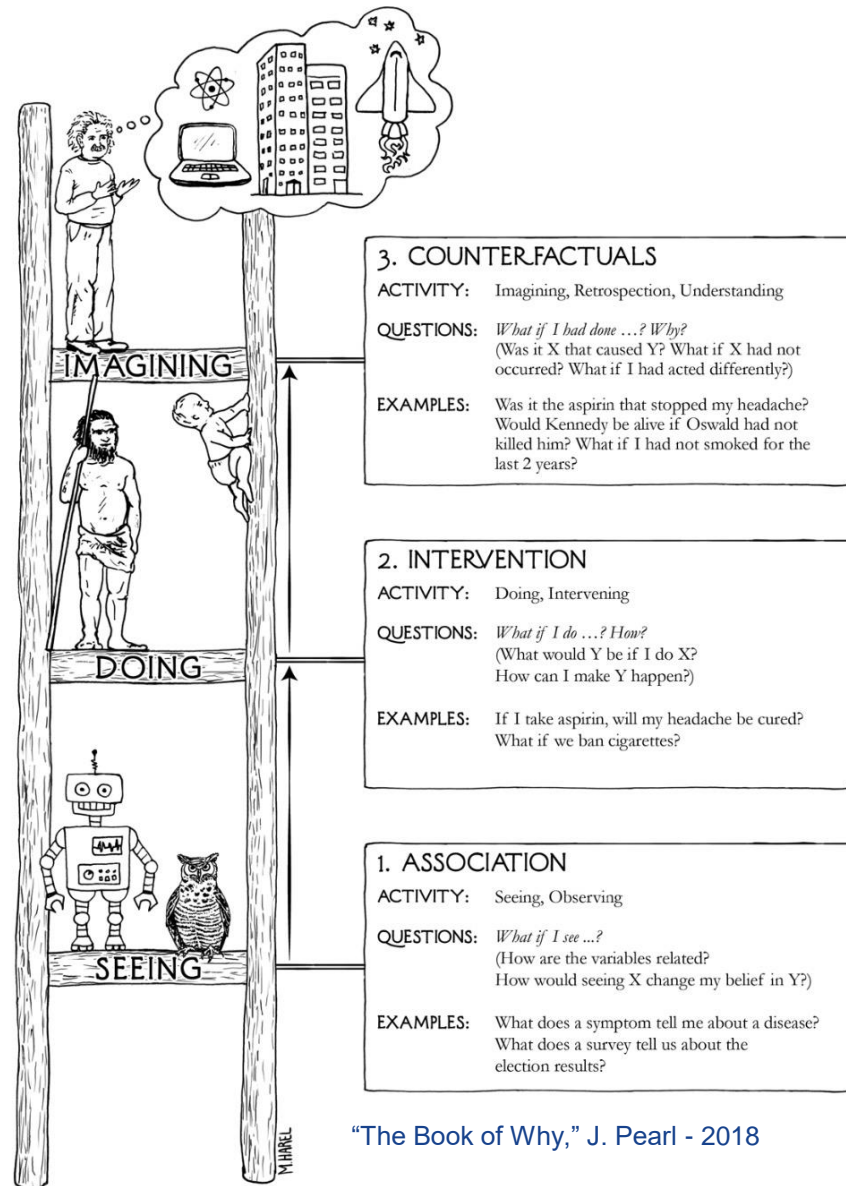
1. Depends on your expertise (i.e. domain knowledge)
2. Highlights confounding variables



sportskeeda.com



# BENEFITS OF CAUSAL MAPS AND MODELS



“The Book of Why,” J. Pearl - 2018

Briefly how it works:

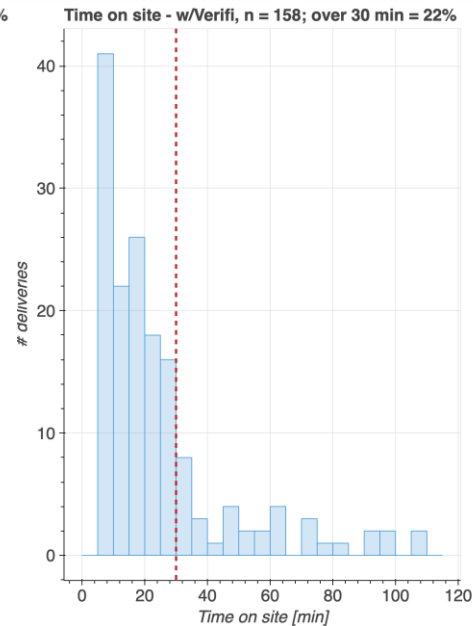
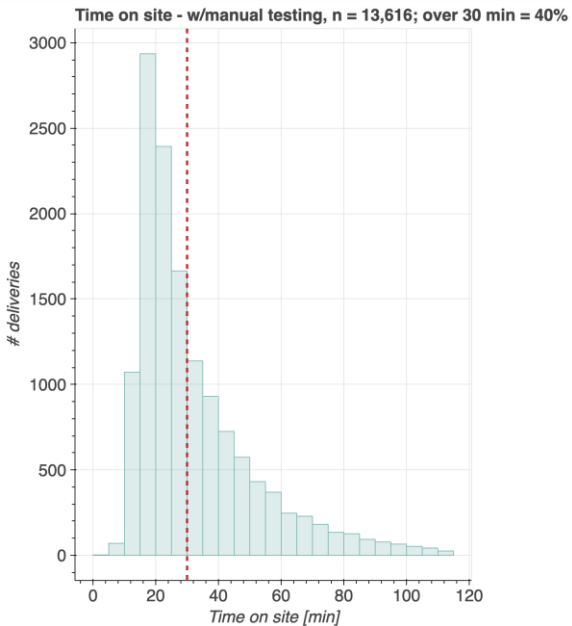
- Domain experts (YOU!) map out
- Identify the “estimand” i.e. what we’re interested in estimating
- Estimate the causal effect
  - Via control of the flow of data
  - Using “Do-Calculus” – i.e. set of rules for manipulating the flow
- Refute the estimate
  - Essentially validate your model

See, for example:

[https://www.pywhy.org/dowhy/v0.8/getting\\_started/intro.html](https://www.pywhy.org/dowhy/v0.8/getting_started/intro.html)

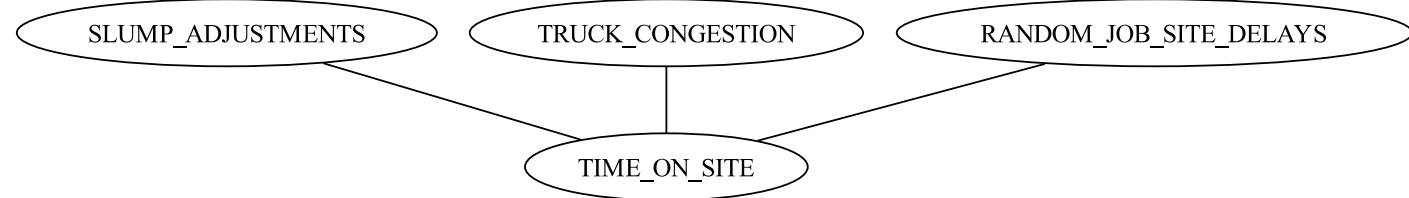


# EXAMPLE PROBLEM STATEMENT: WHAT CAUSES “IDLE” TIME AT THE SITE?



Reduction in billable idle time by 18%

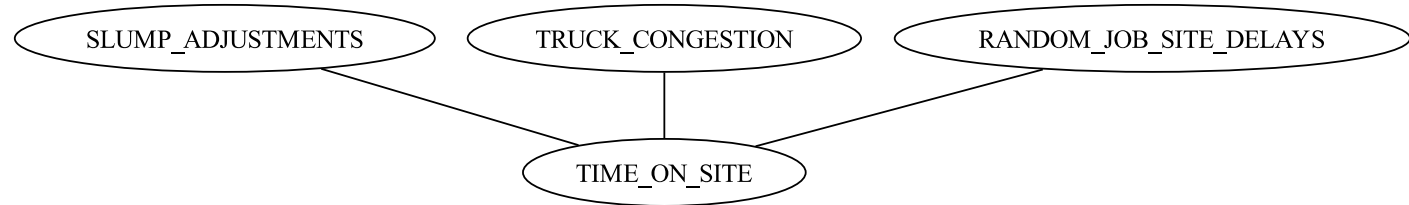
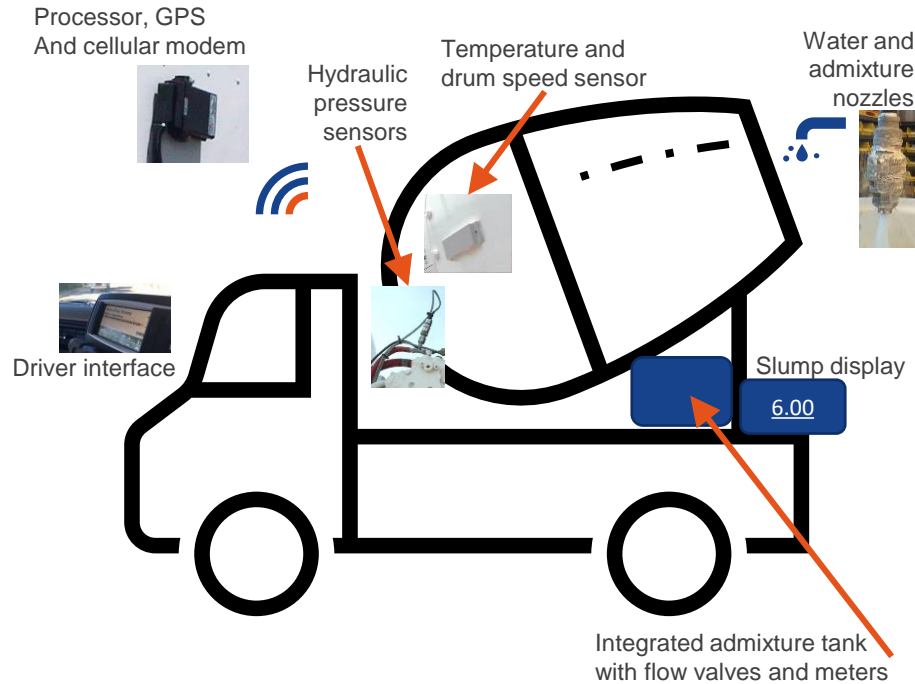
Ok, so how might we do this in a traditional sense...?



Essentially:

- Is the truck ready to pour?
- Is the contractor ready to pour?
- Is Murphy's law in effect?

# EXAMPLE PROBLEM STATEMENT: WHAT CAUSES “IDLE” TIME AT THE SITE?

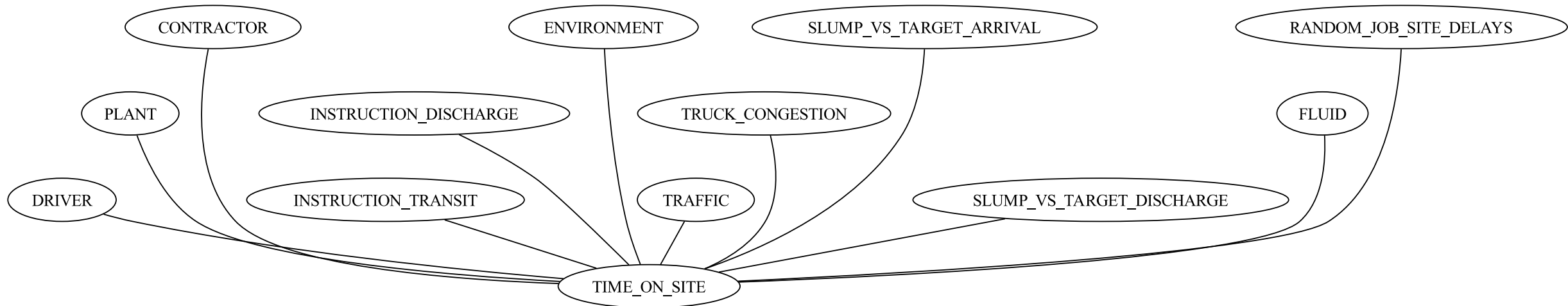


- A commercial IoT measurement/management system is used to feed the models
- Over 180 MM cubic yards have been generated through these systems globally
- Billions of data points... slump, temperature, GPS, water adds, admixture adds...

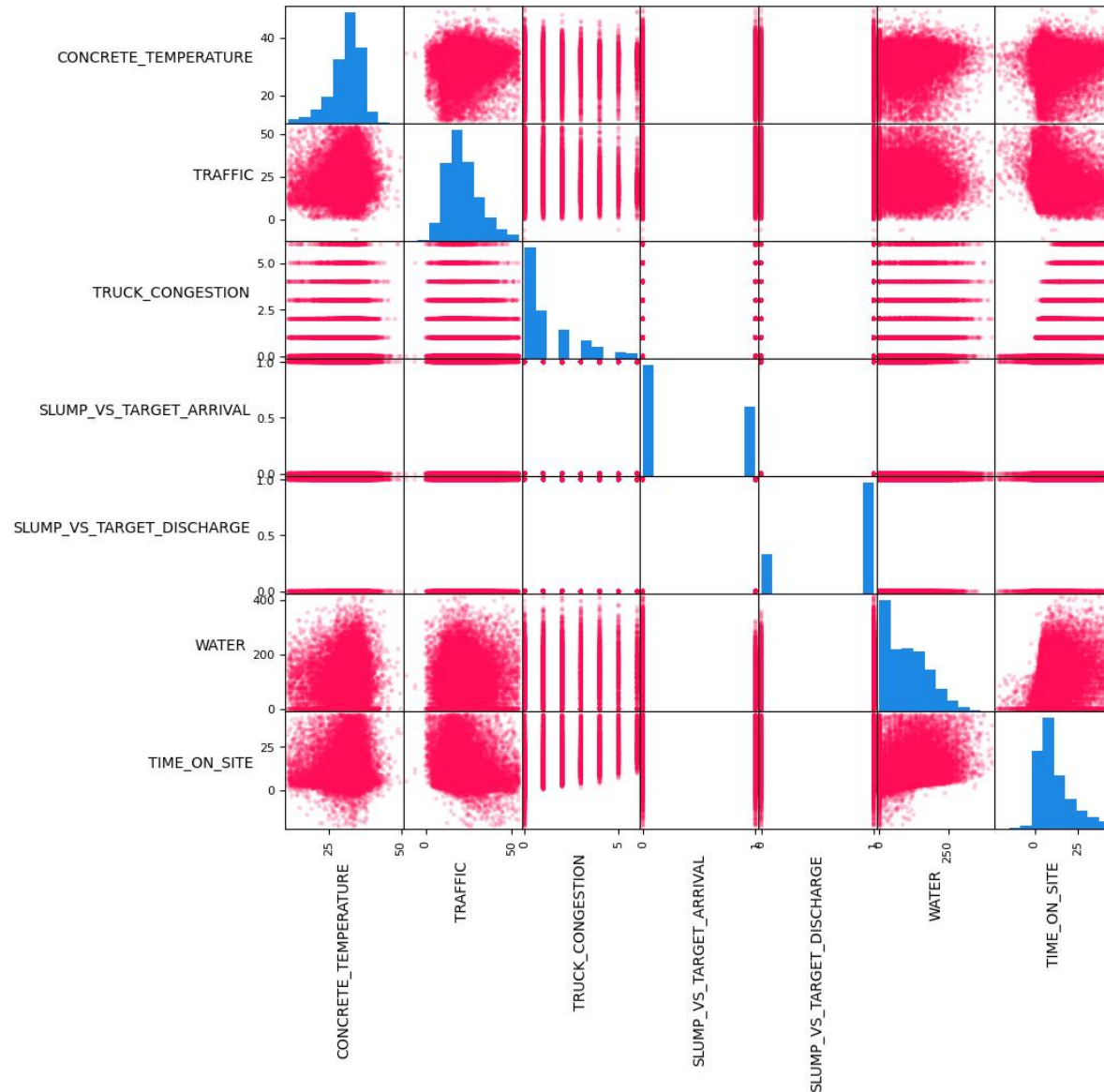
# EXAMPLE PROBLEM STATEMENT: WHAT CAUSES “IDLE” TIME AT THE SITE?



Let's be a little more detailed...



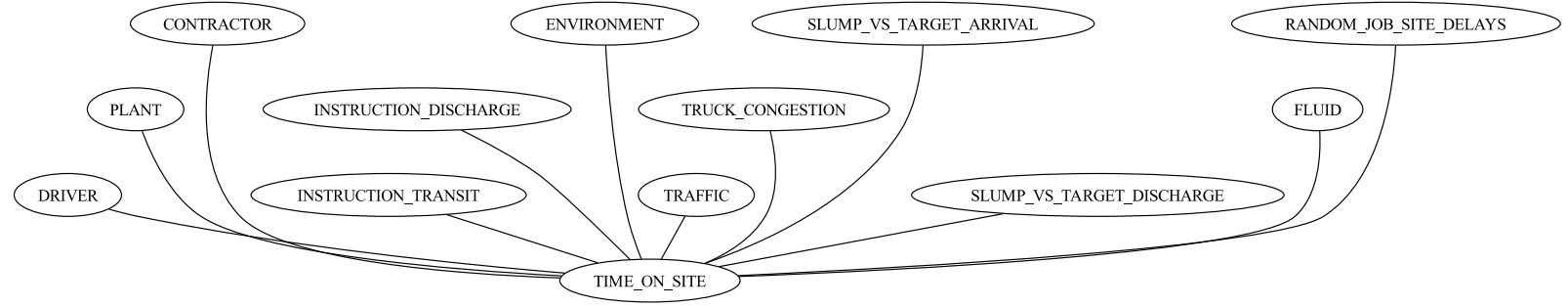
# A QUICK LOOK AT THE DATA



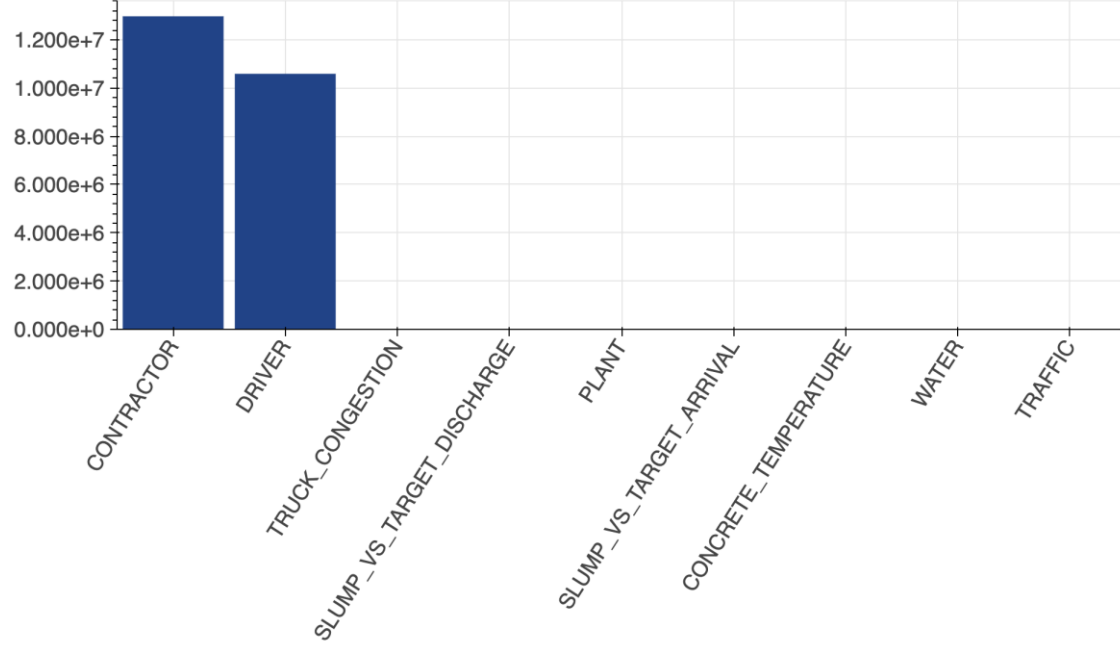
## One producer

- 19,067 loads over three months
- SLUMP\_VS\_TARGET\_ARRIVAL, SLUMP\_VS\_TARGET\_DISCHARGE converted to TRUE/FALSE
- Manage instructions constant
- Outliers removed
- Looking to see what causes TIME\_ON\_SITE

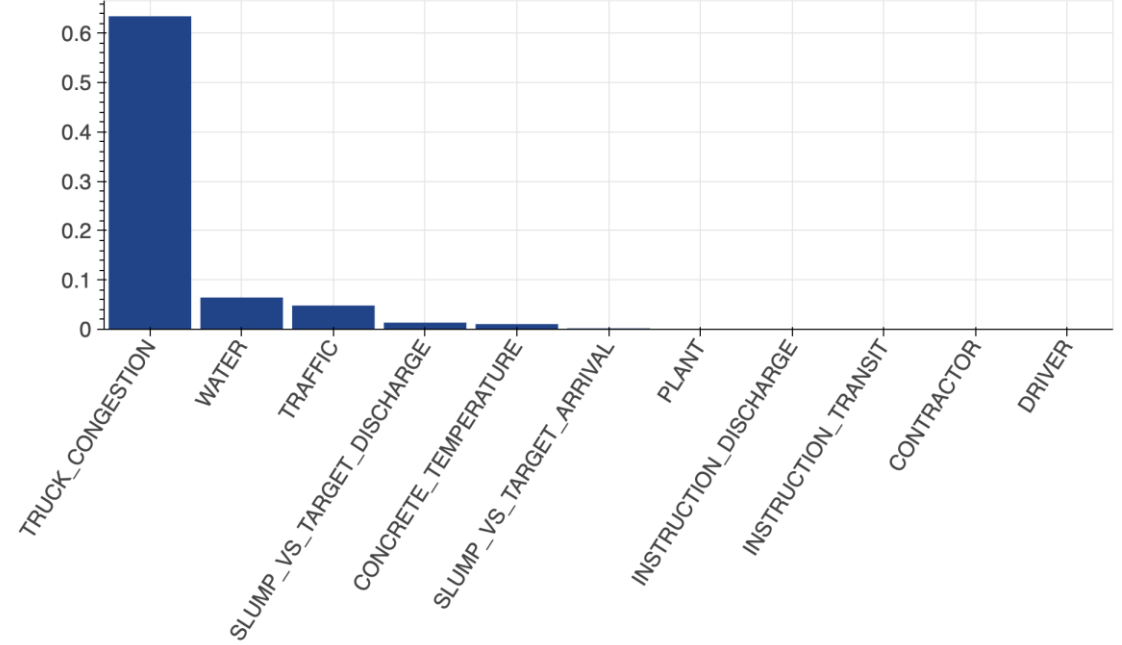
# TRADITIONAL METHODS



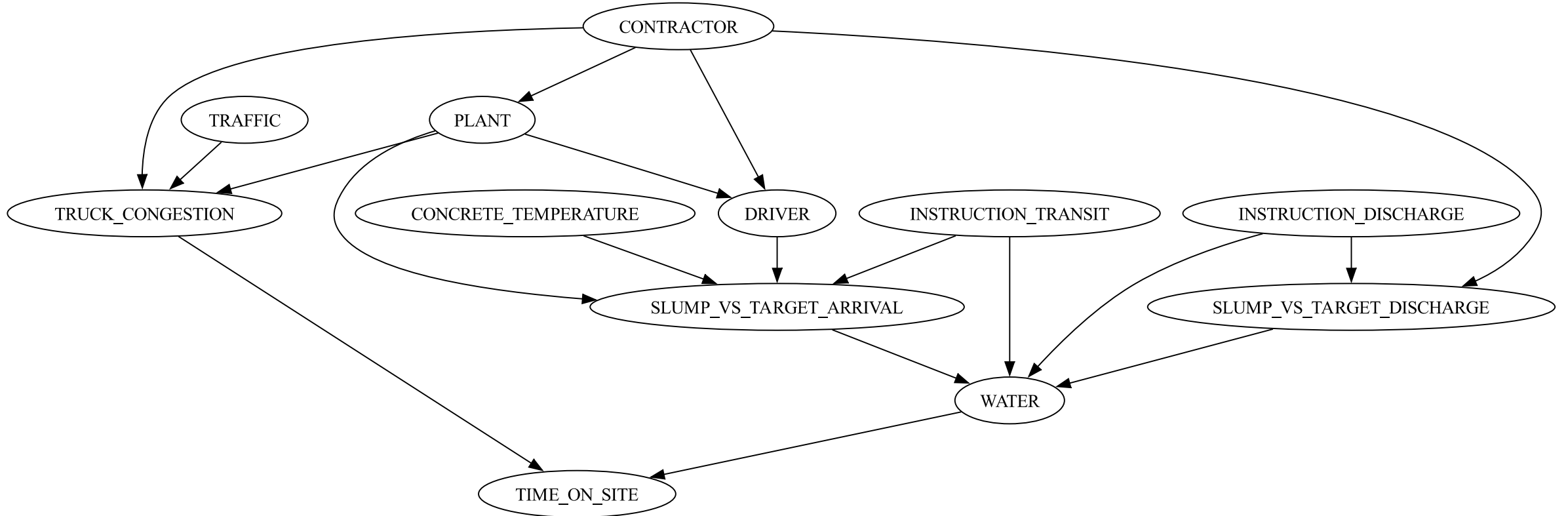
Feature importance: Linear regression; [avg PWT = 58.9%]



Feature importance: Random forest; [avg PWT = 61.9%]

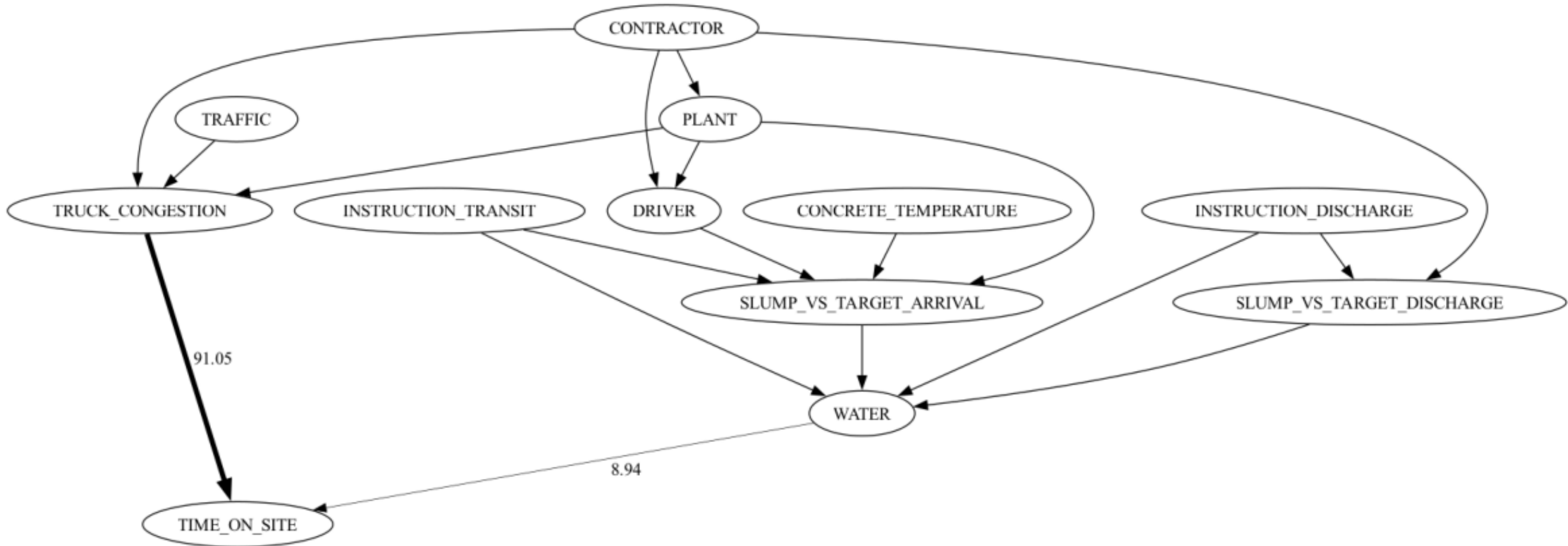


# THE CAUSAL MAP



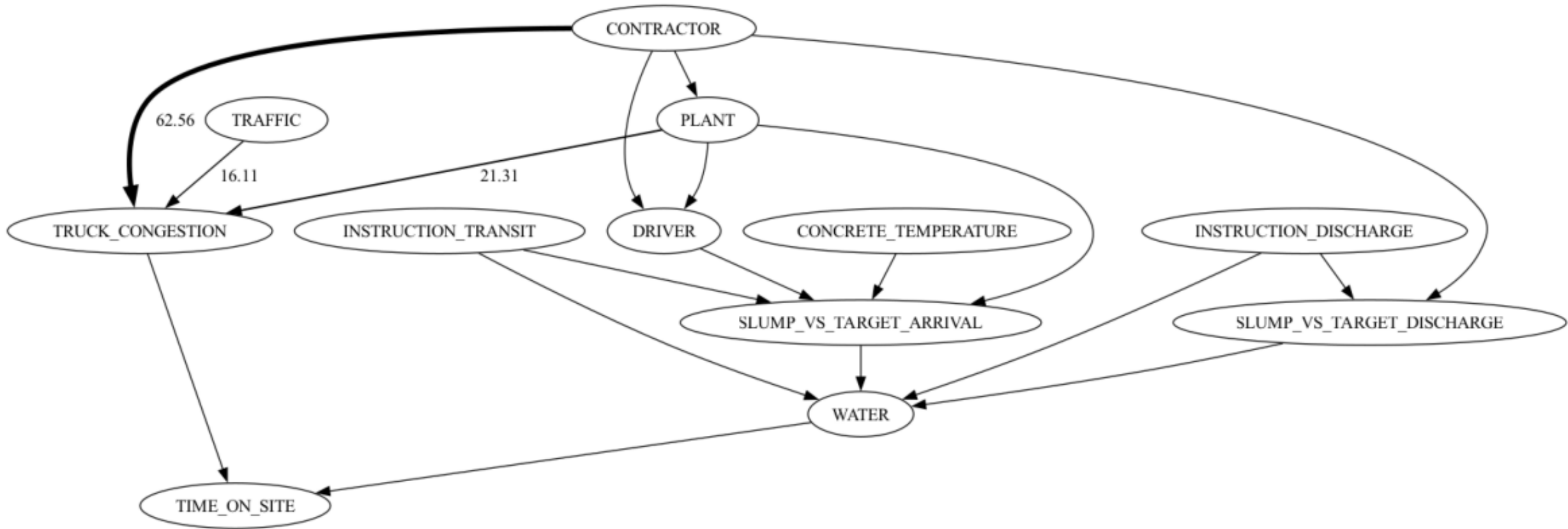
# APPLYING THE “DO-CALCULUS” AND FINDING “ARROW STRENGTH”

In other words, how much does each node contribute to the variance in a child node?



# APPLYING THE “DO-CALCULUS” AND FINDING “ARROW STRENGTH”

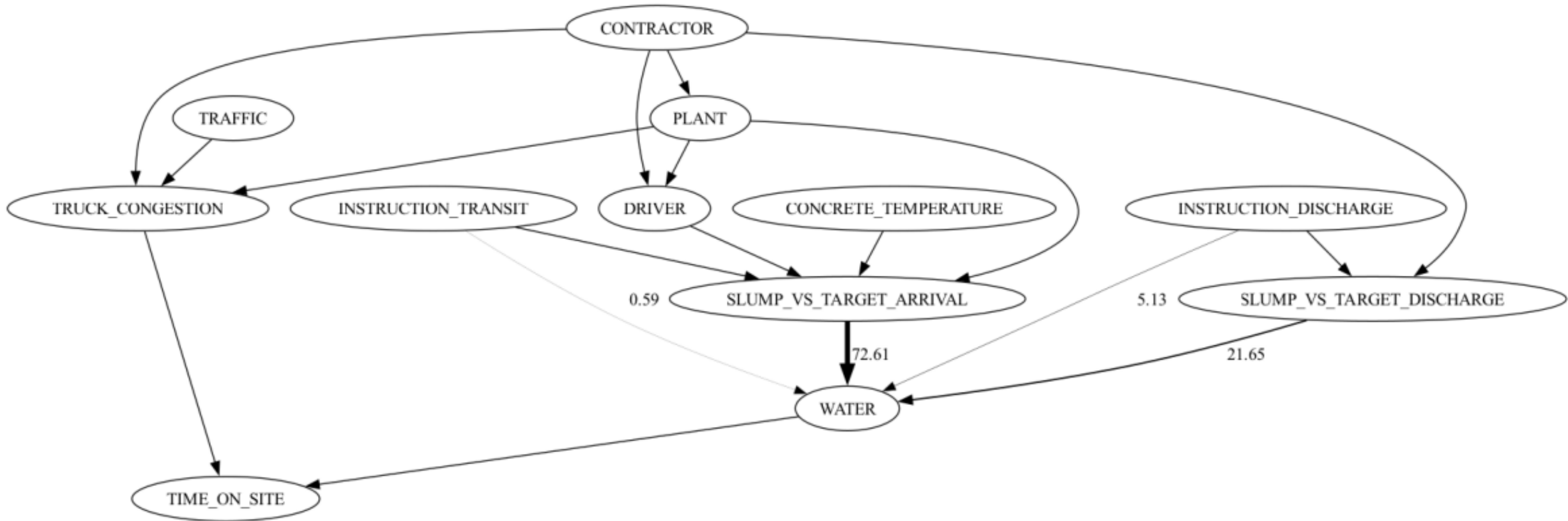
In other words, how much does each node contribute to the variance in a child node?





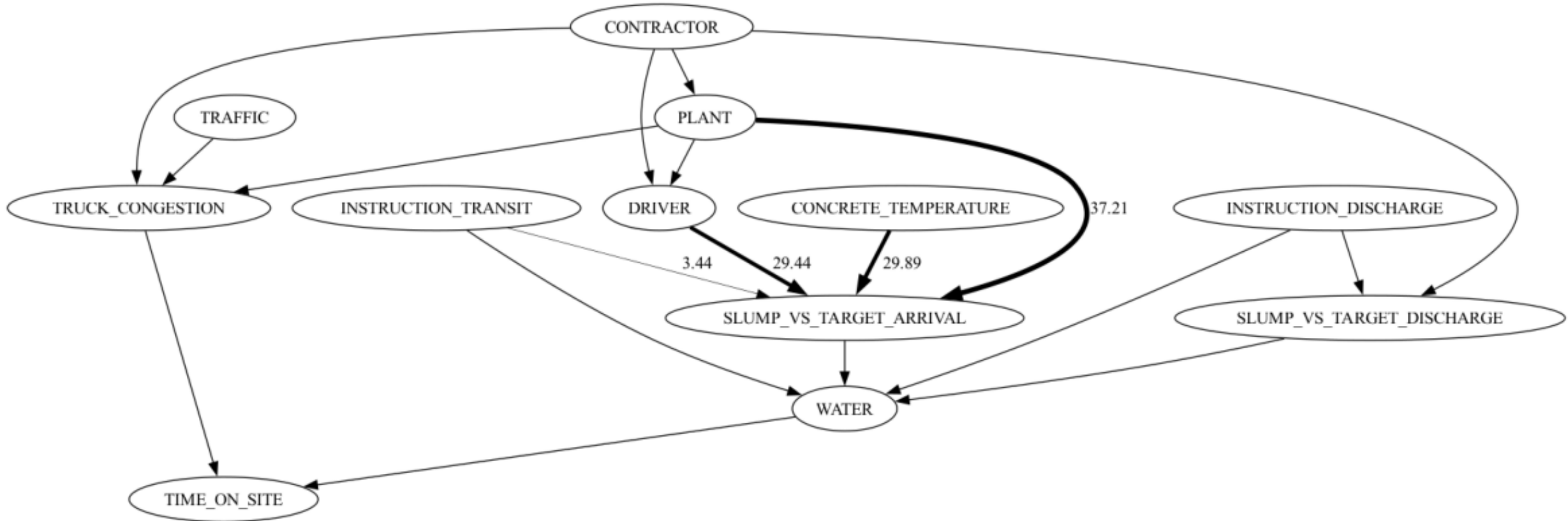
# APPLYING THE “DO-CALCULUS” AND FINDING “ARROW STRENGTH”

In other words, how much does each node contribute to the variance in a child node?



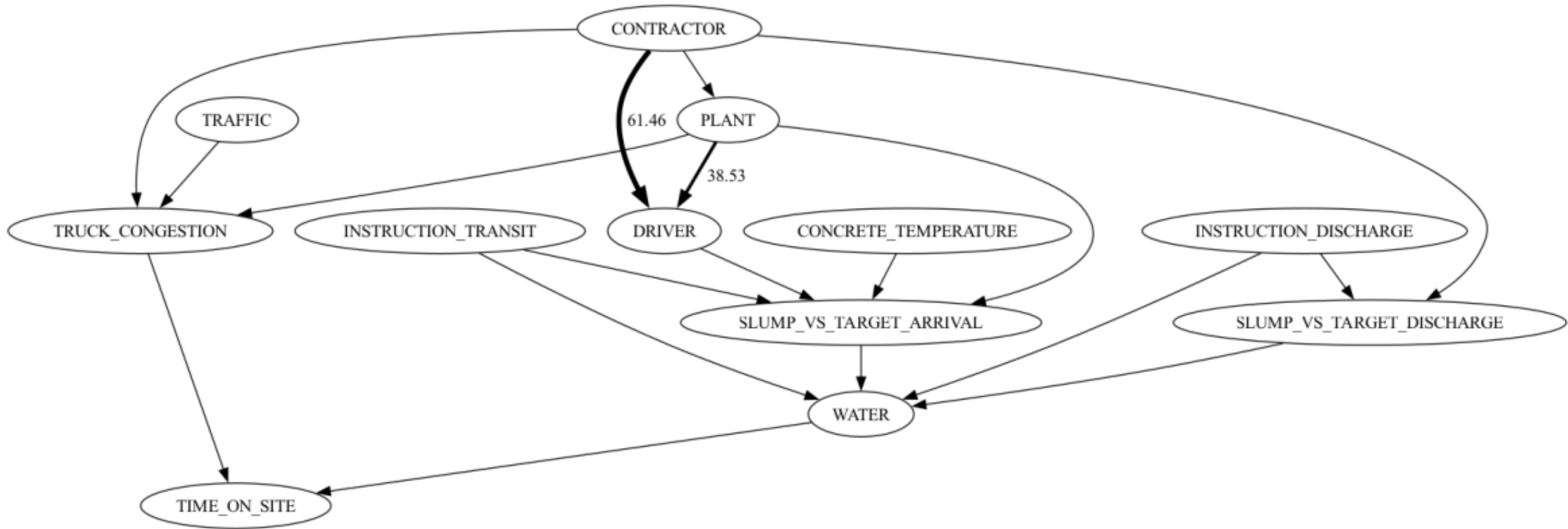
# APPLYING THE “DO-CALCULUS” AND FINDING “ARROW STRENGTH”

In other words, how much does each node contribute to the variance in a child node?

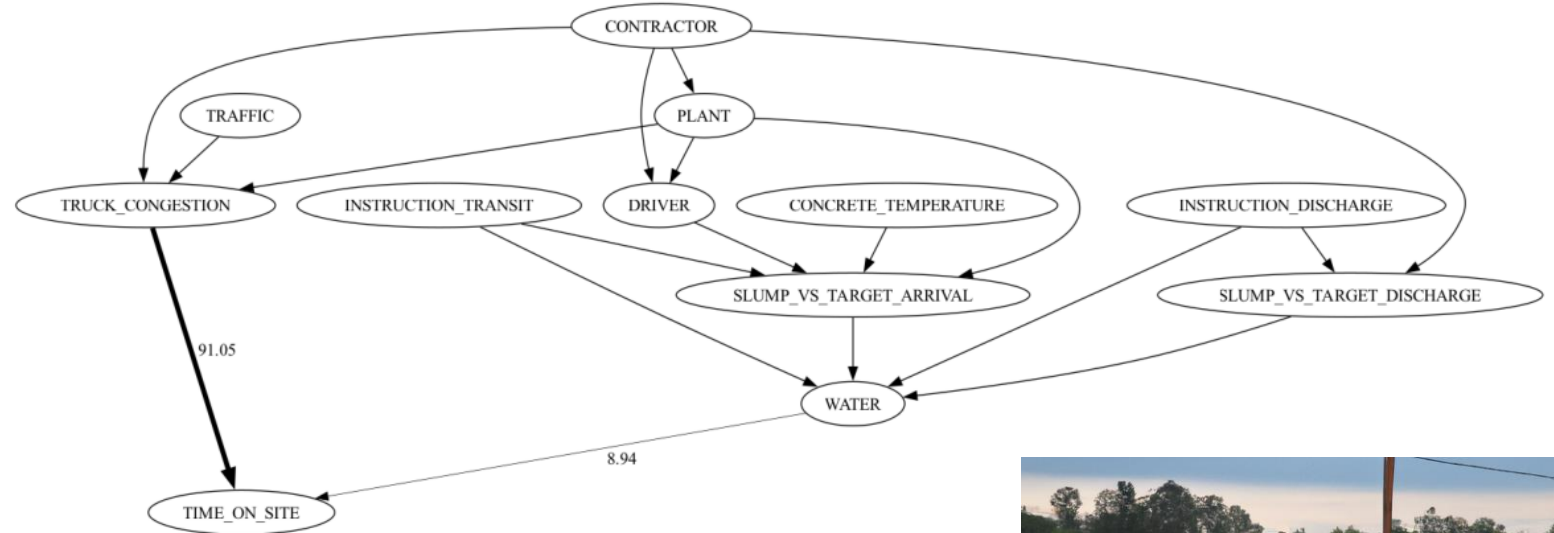
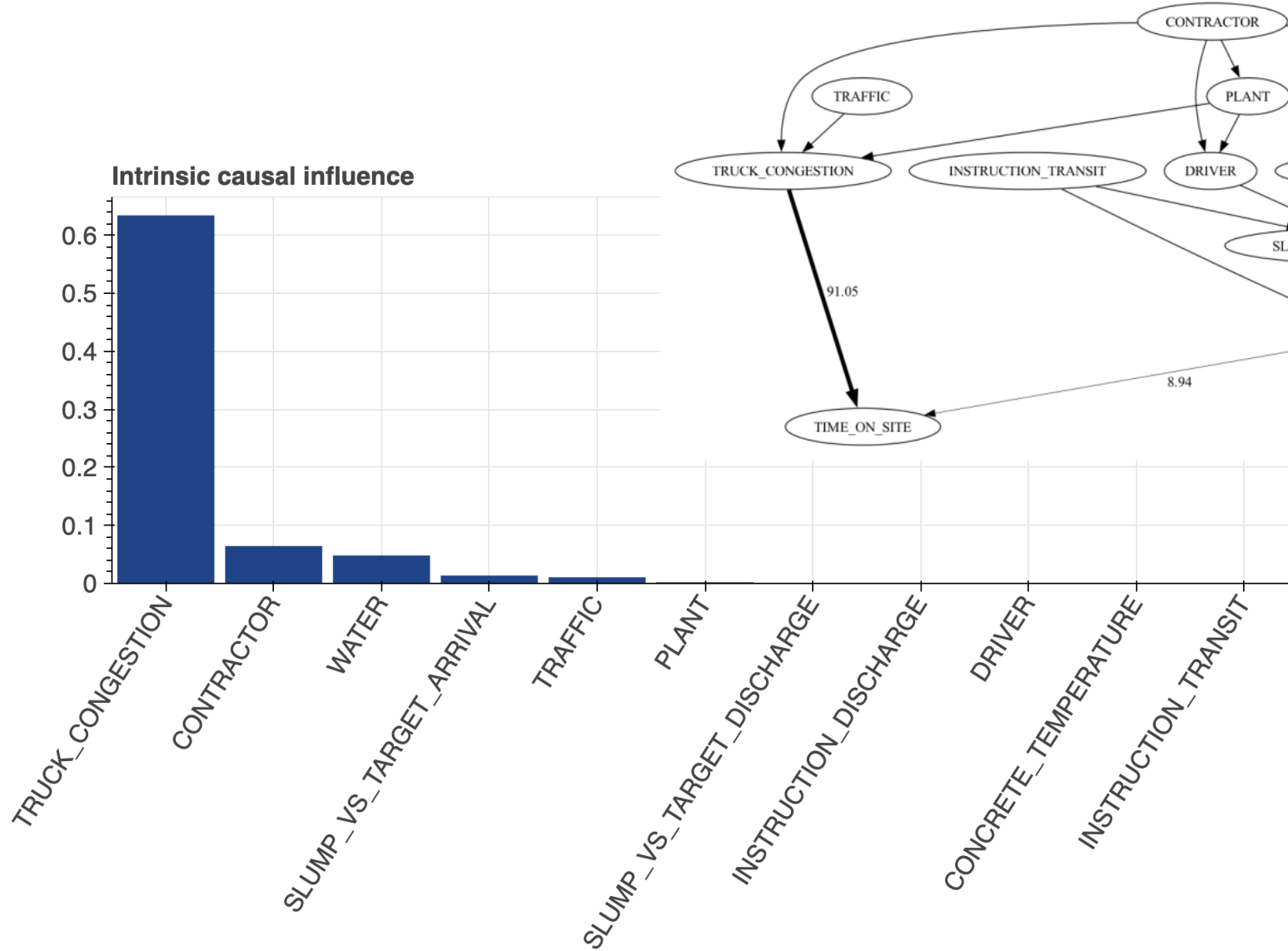


# APPLYING THE “DO-CALCULUS” AND FINDING “ARROW STRENGTH”

In other words, how much does each node contribute to the variance in a child node?

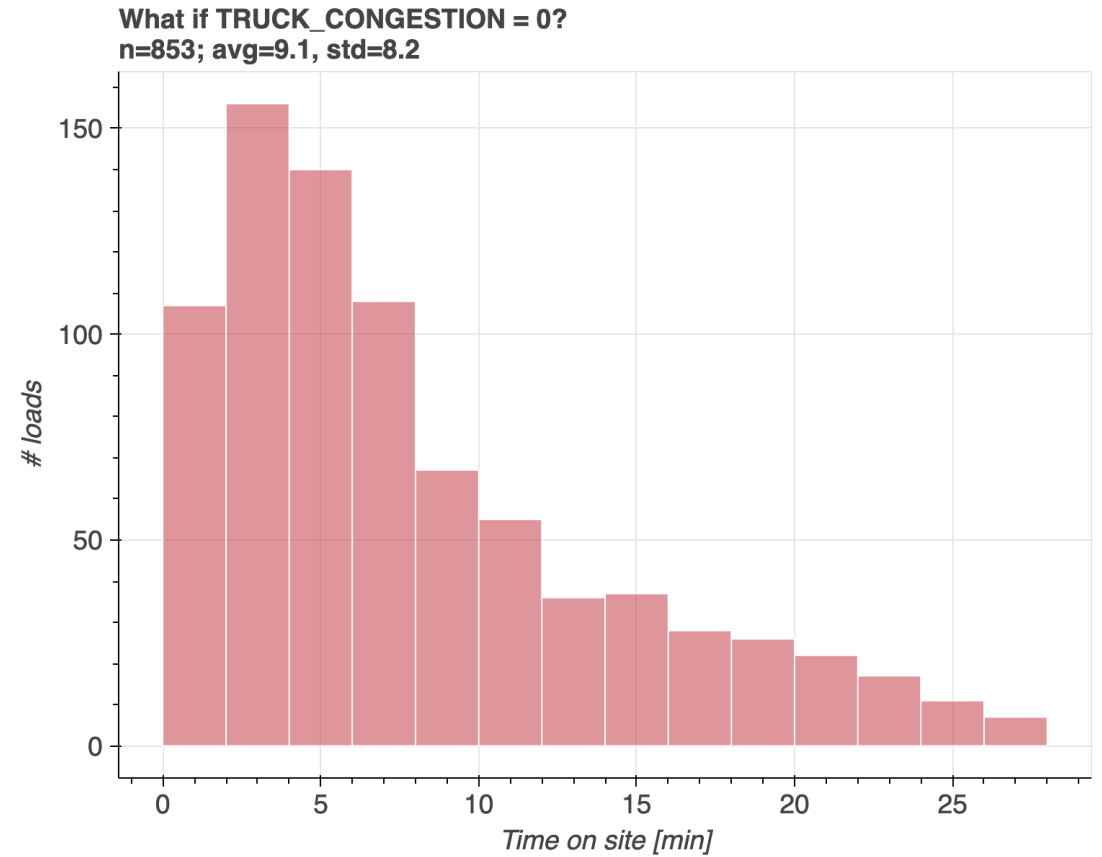
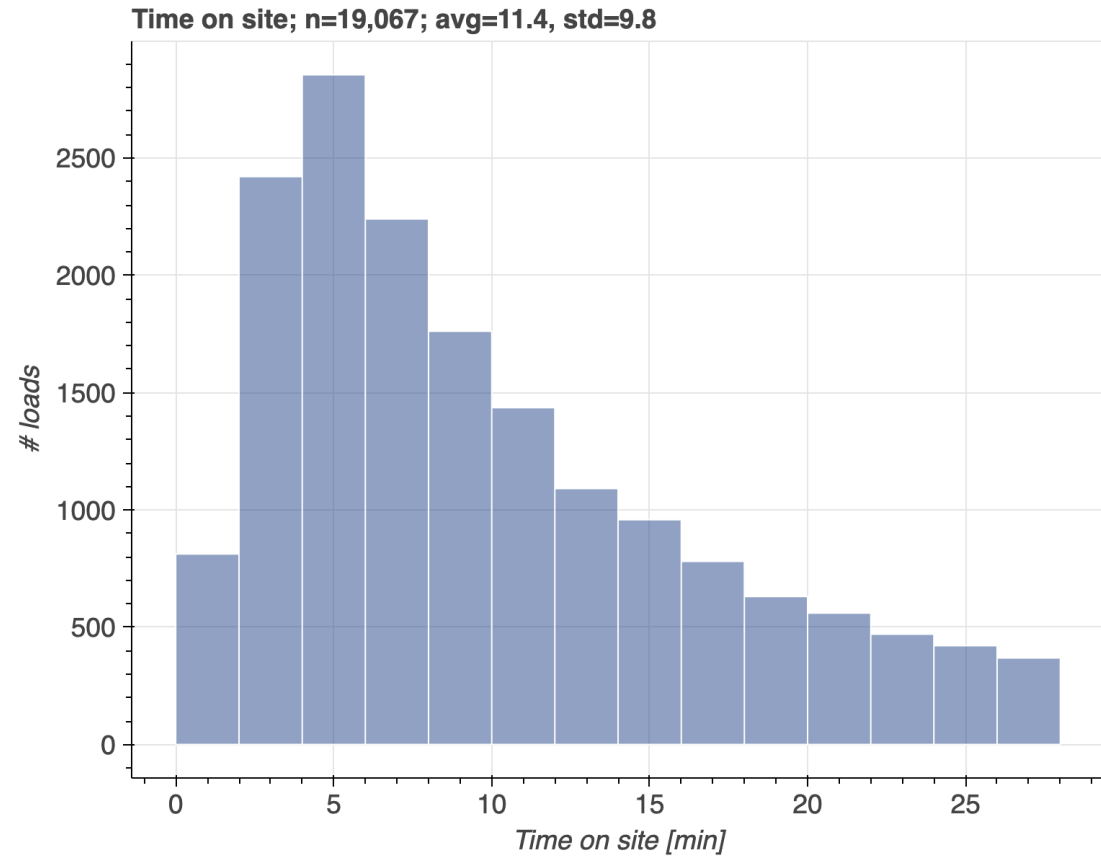


# INTRINSIC CAUSAL INFLUENCE (ISOLATING THE EFFECTS)

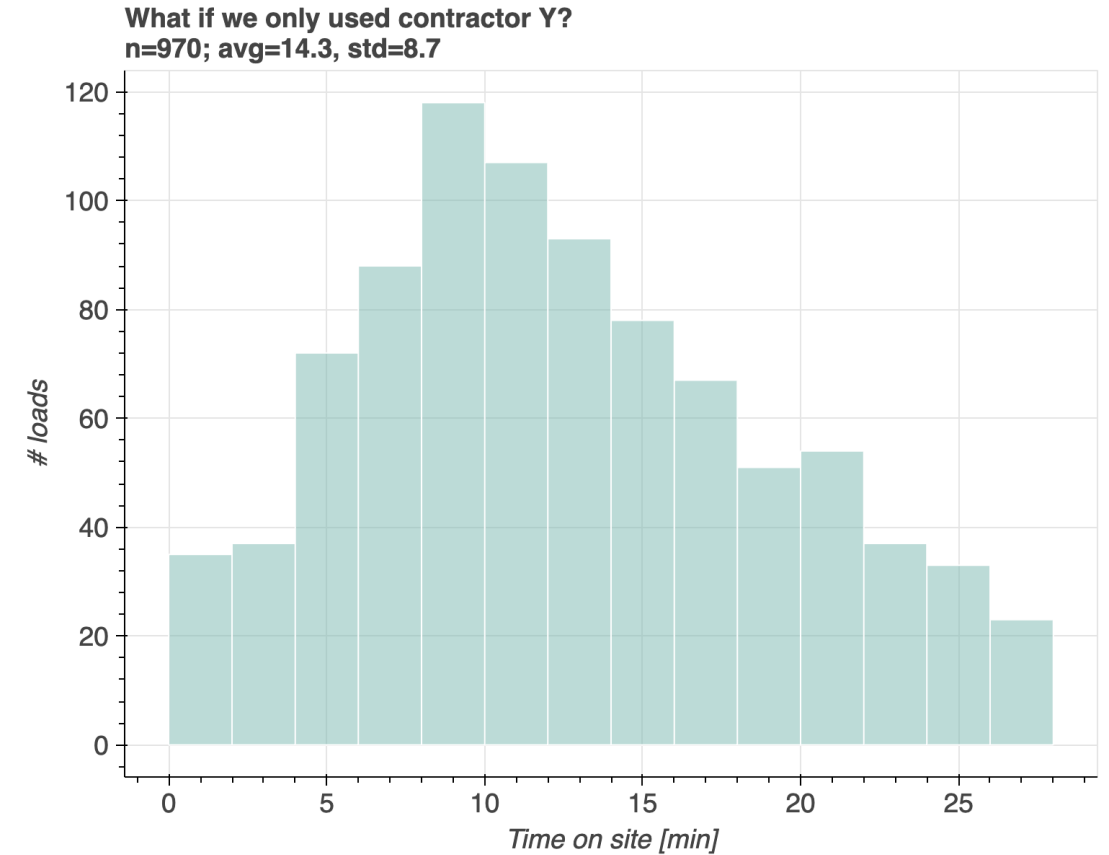
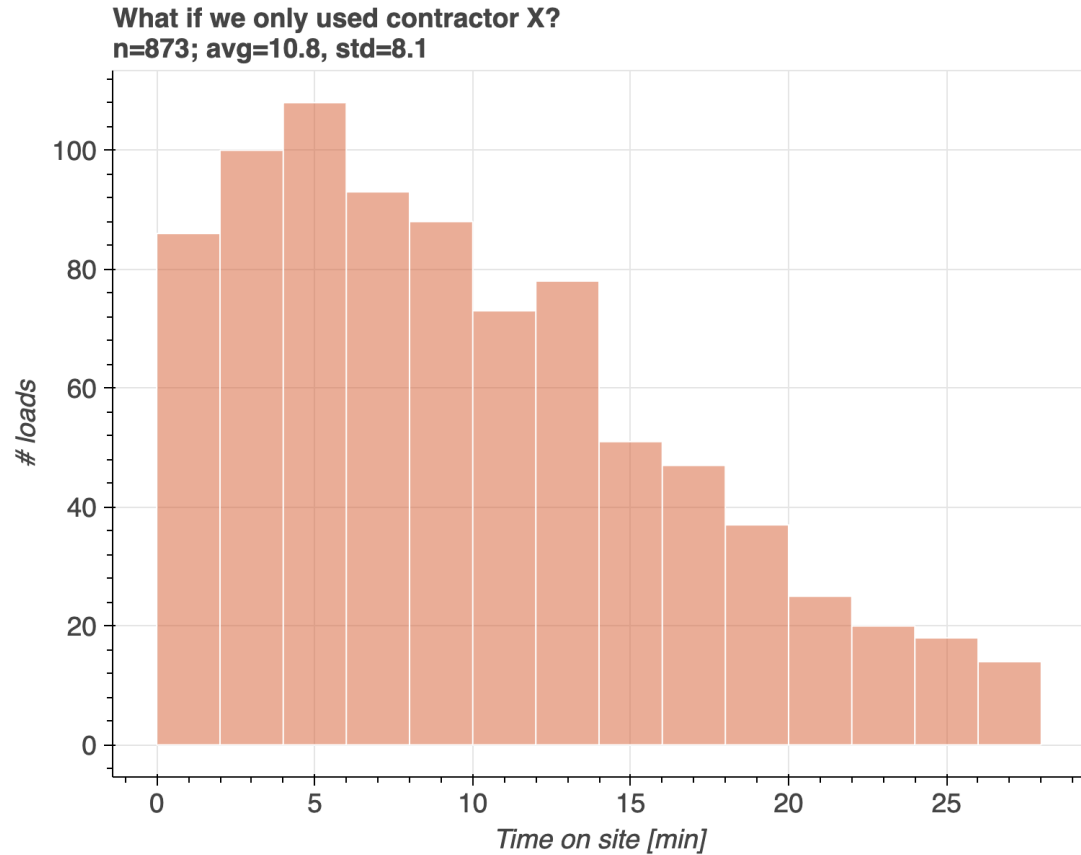


DALL-E (12 iterations)

# DIGGING DEEPER - INTERVENTIONS



# DIGGING DEEPER - INTERVENTIONS





**QUESTIONS?**