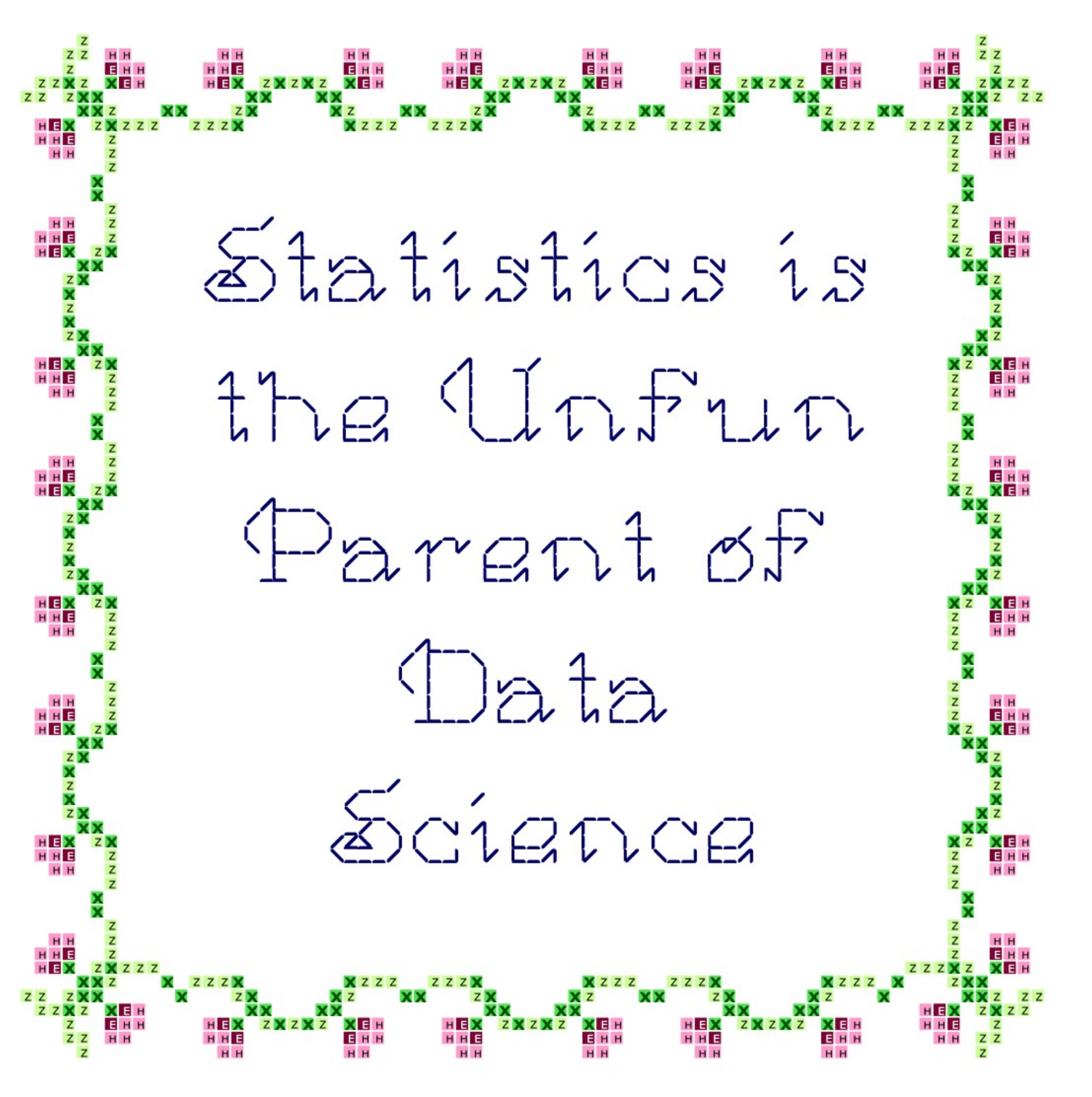
Is Statistics giving you fits? How to think about data and statistical methods for driver behavior and safety in a changing transportation world

Oct 20, 2022

Carol Flannagan, Ph.D.

University of Michigan Transportation Research Institute (UMTRI)

Some Background First


Statistics is
the Unfun
Parent of
Data
Science

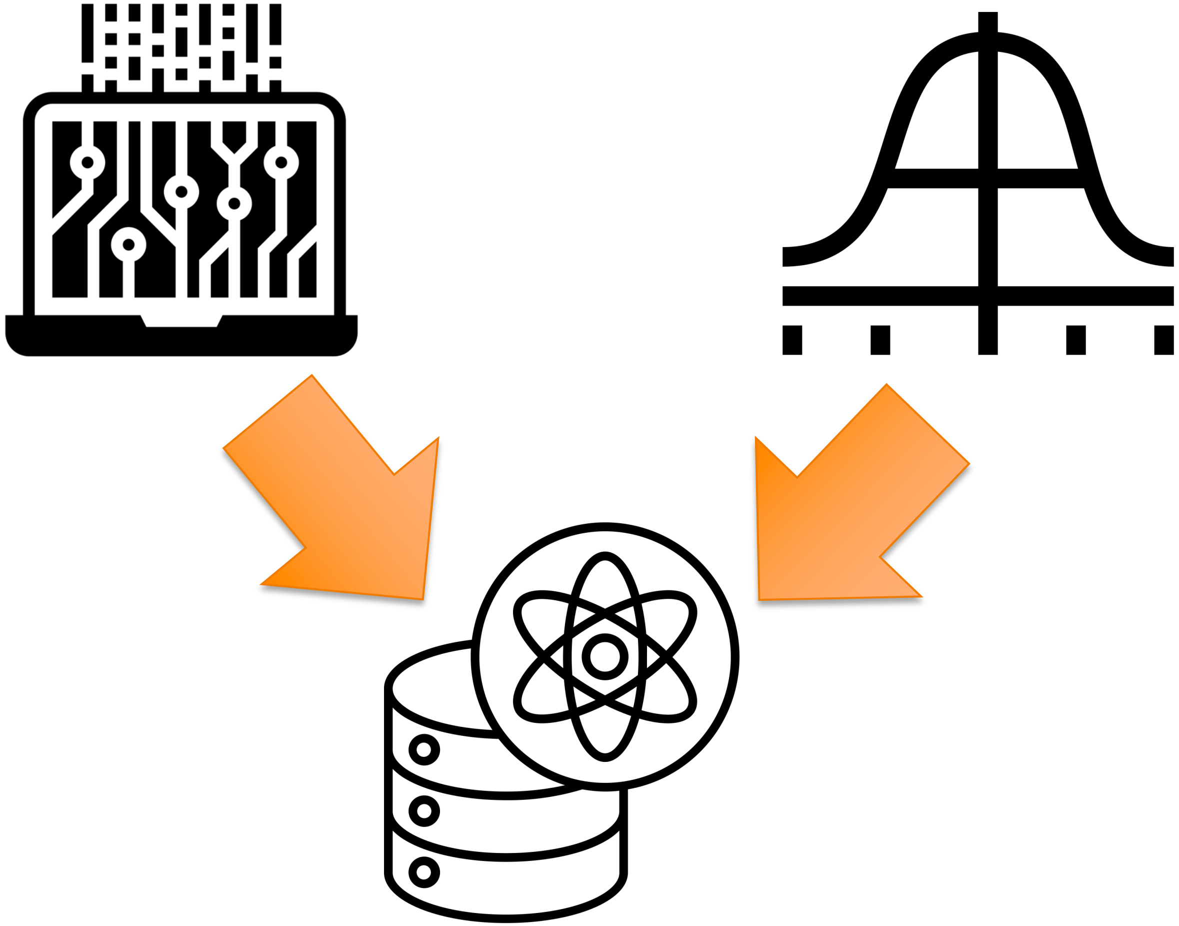
Statistics is the Unfun Parent of Data Science

“Let me fill you in on a little secret…When I met your Dad, I was fun too. But I had to give all that up, because you can’t have two fun parents. That’s a carnival.”

- Claire Dunphy

Statistics is the Unfun Parent of Data Science

Fun parent



Unfun parent

If Data Science = Computer Science + Statistics…

Good things that happen to other people are not bad things that happen to you (and vice versa)

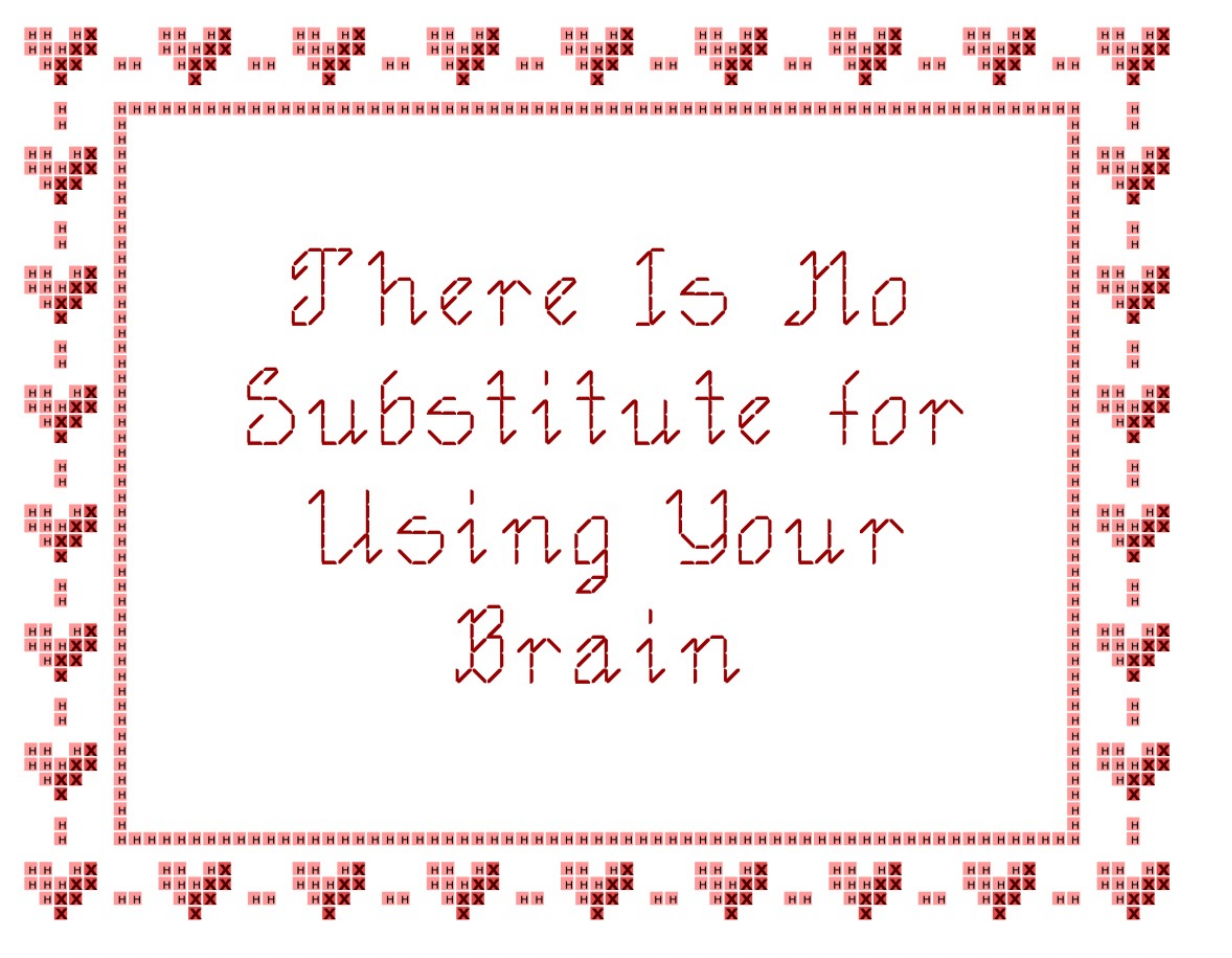
You need to have a plan and a goal

Long-delay feedback systems tend to oscillate

Separate the hazards

Fun Parent

Unfun Parent



There Is No Substitute for Using Your Brain

There is No Substitute for Using Your Brain

Know your analytical goals

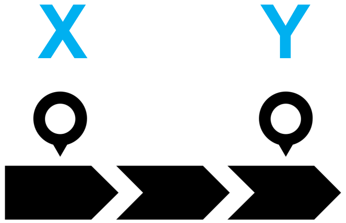
Description

Prediction

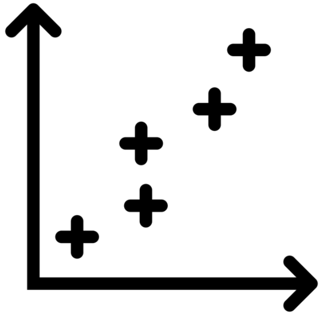
Inference

There is No Substitute for Using Your Brain

Causal inference requires:



Temporal precedence

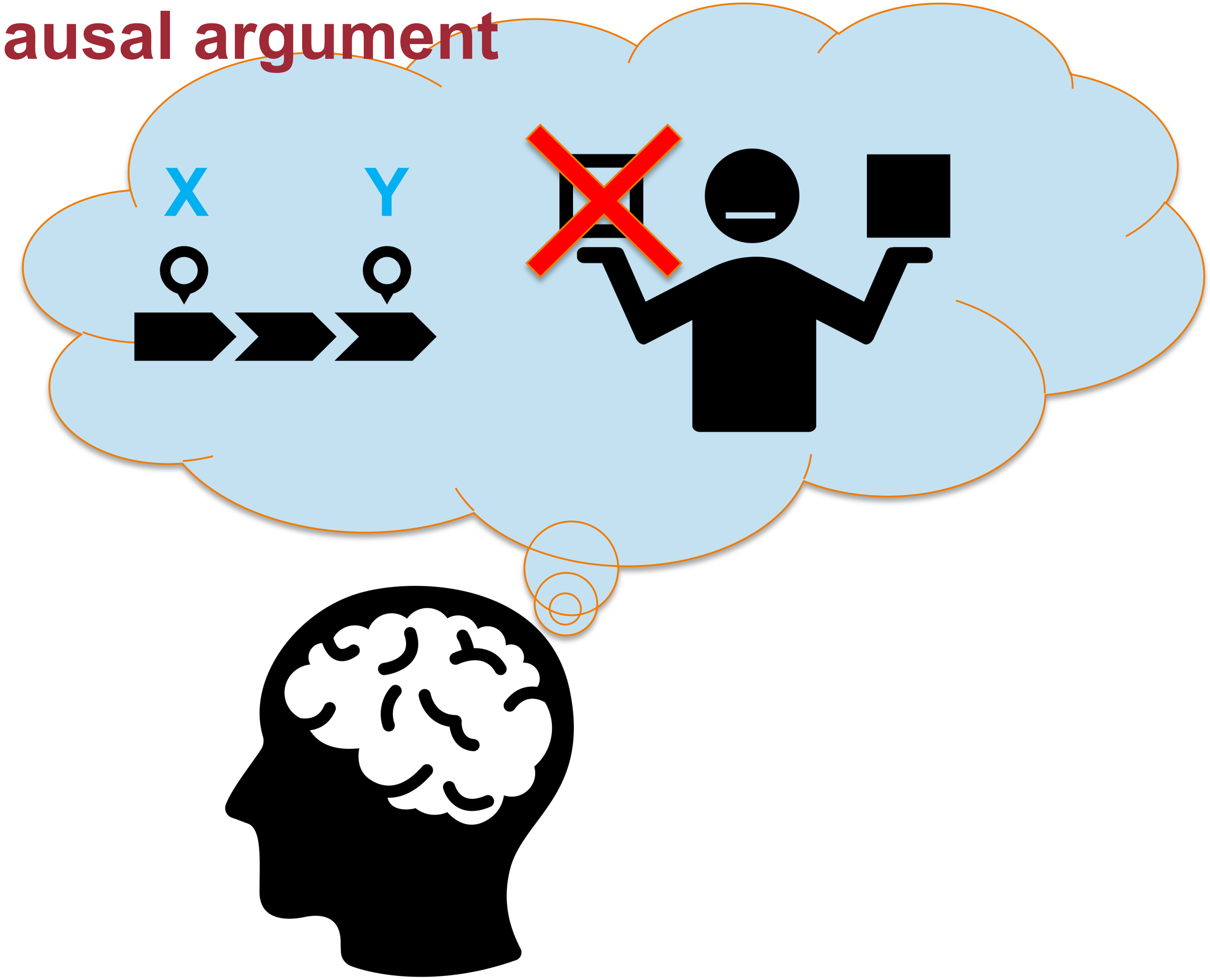


Covariation

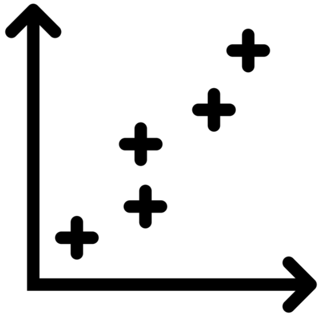
No Alternative Explanations

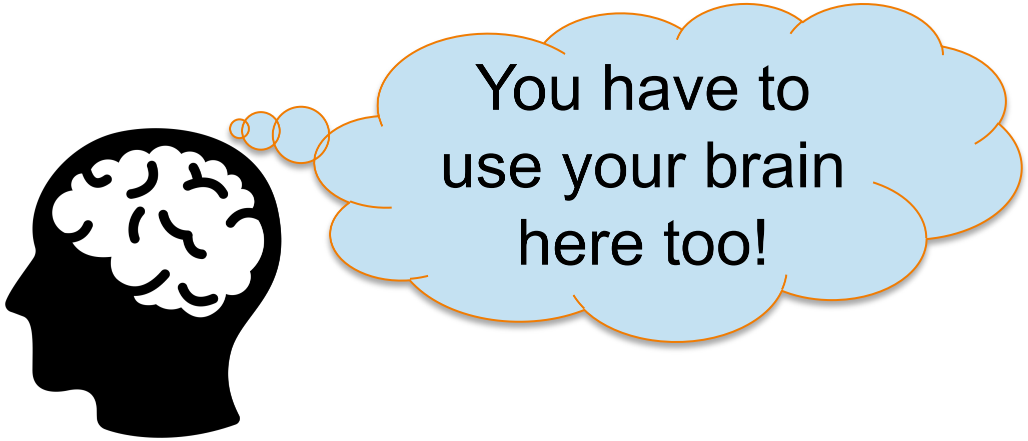
There is No Substitute for Using Your Brain

Causal argument

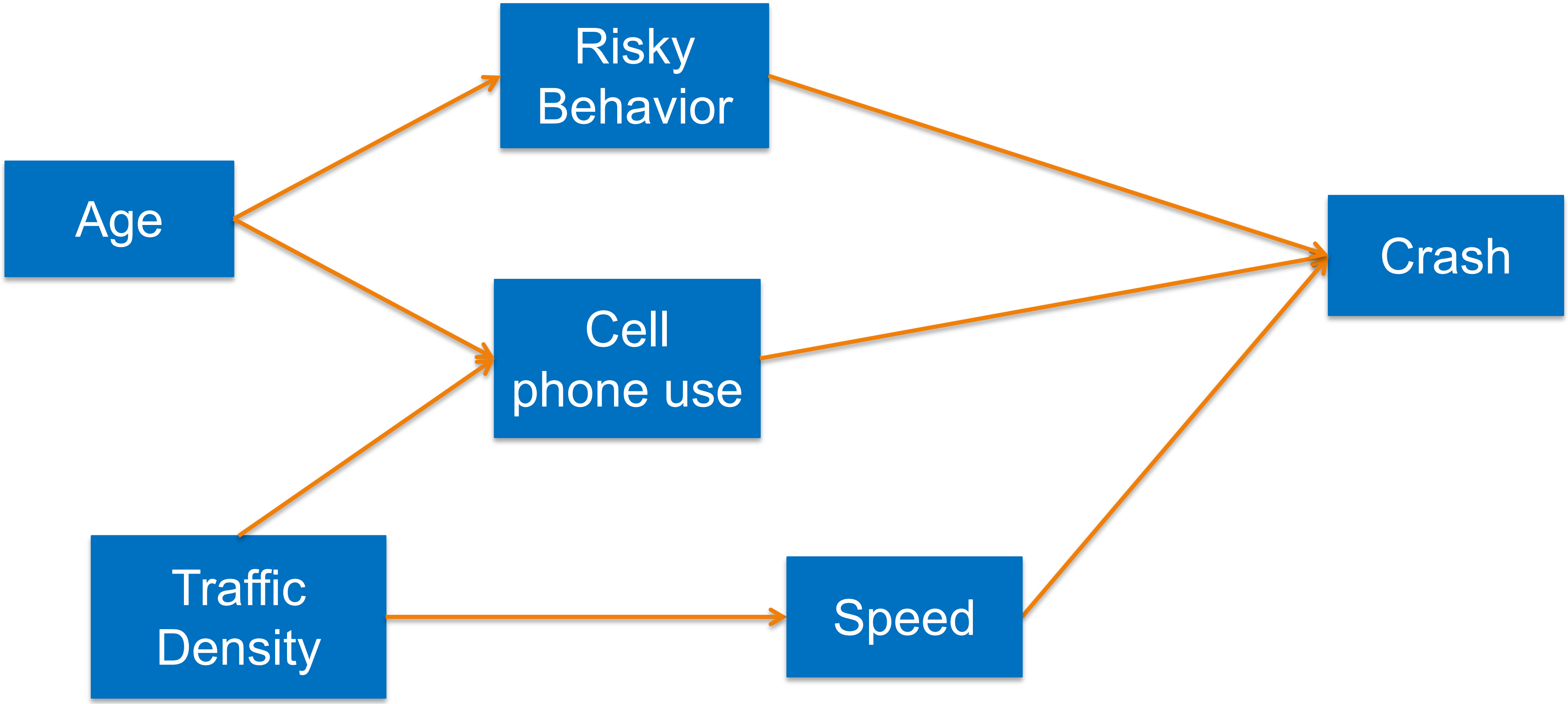


Statistics and data





There is No Substitute for Using Your Brain



Directed Acyclic Graph => DAG

There is No Substitute for Using Your Brain

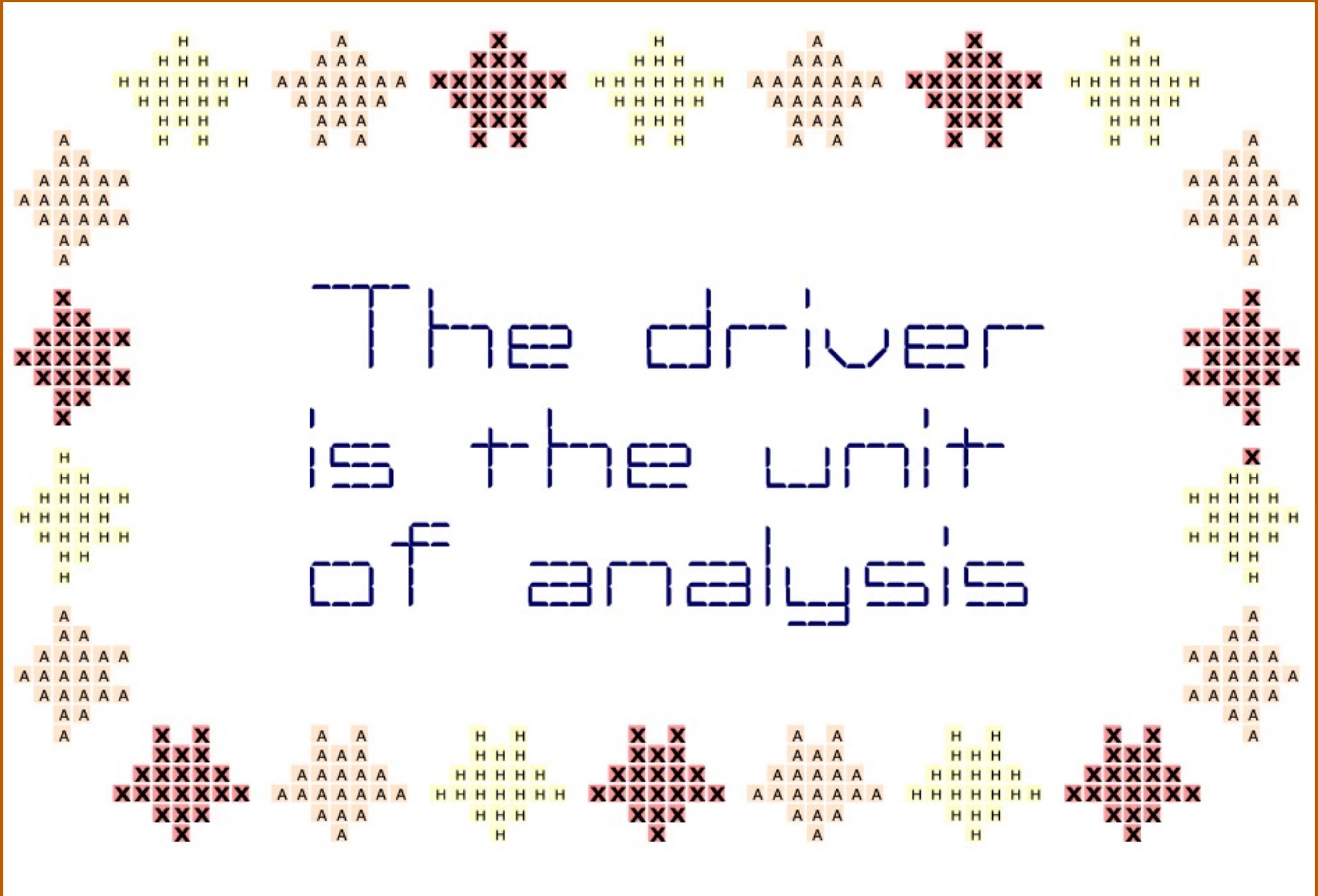
Requirements for estimation of causal effects:

1. *Consistency*: the cause is sufficiently well-defined
2. *Non-interference*: observations are independent
3. *Exchangeability*: “cause present” and “cause absent” groups are the same in all other ways that matter
4. *Positivity*: anyone in the dataset could have been in the “cause present” or “cause absent” group

There is No Substitute for Using Your Brain

Requirements for estimation of causal effects:

1. *Consistency*: the cause is sufficiently well-defined
2. *Non-interference*: observations are independent
3. *Exchangeability*: “cause present” and “cause absent” groups are the same in all other ways that matter
4. *Positivity*: anyone in the dataset could have been in the “cause present” or “cause absent” group

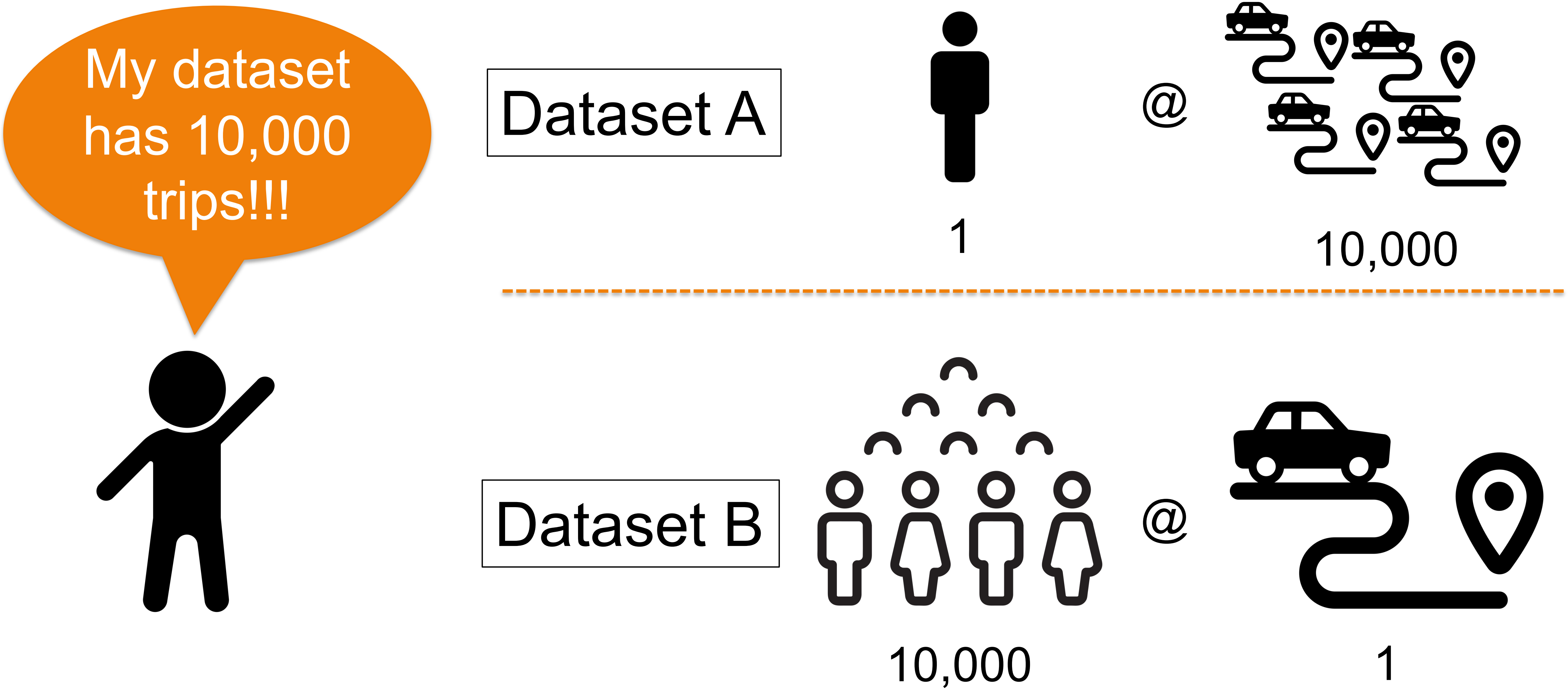


The driver is the unit of analysis

The Driver is the Unit of Analysis

Example: Do L3 ADS drivers respond to takeover requests more slowly when they are on their cell phone?

The Driver is the Unit of Analysis



My dataset has 10,000 trips!!!

Dataset A

1 @ 10,000

Dataset B

10,000 @ 1

The Driver is the Unit of Analysis

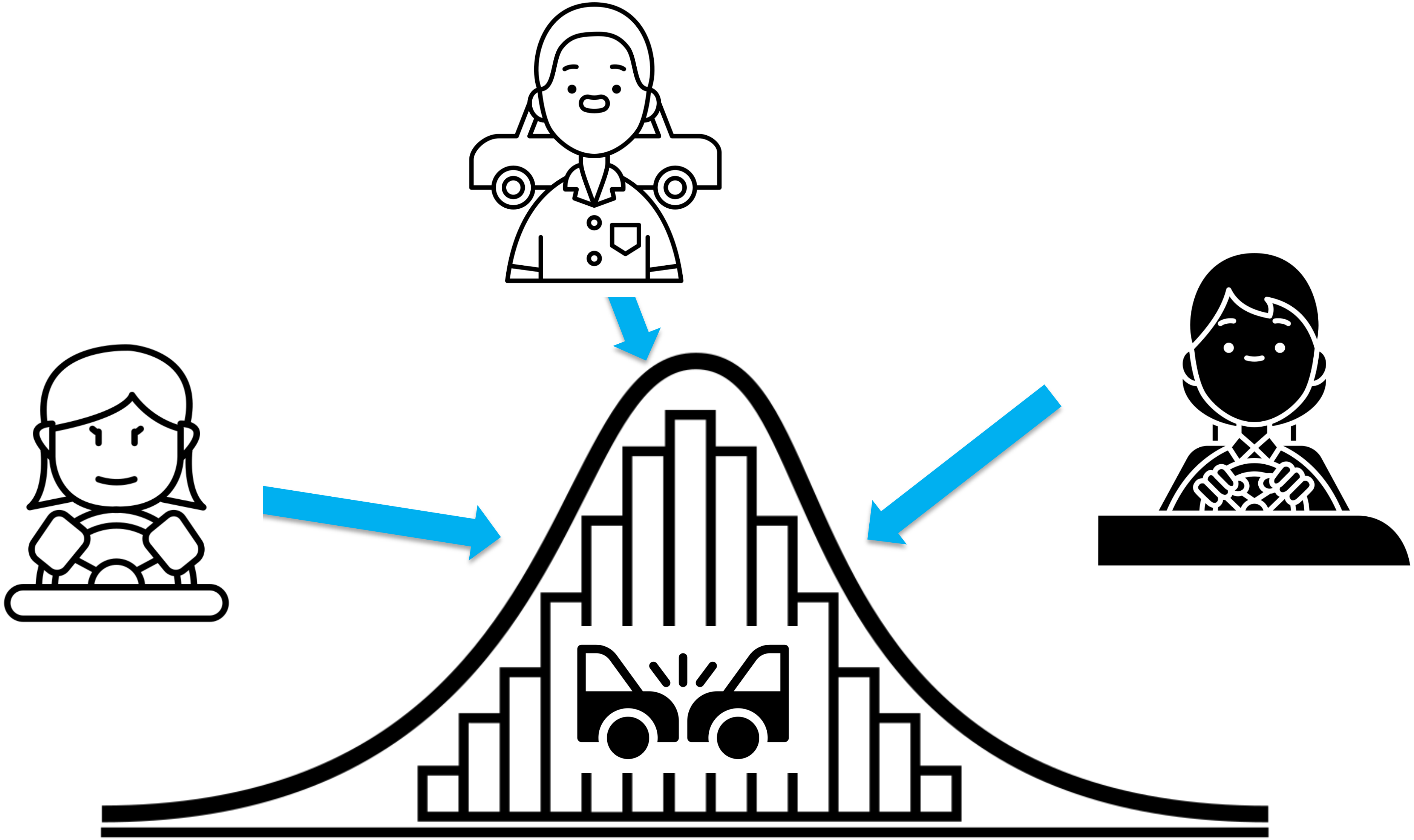
Why do we care about Big Data?

- Big sample size

Why do we care about sample size?

- Generalizability
- Statistical power

The Driver is the Unit of Analysis



The Driver is the Unit of Analysis

Statistical power depends on:

1. Sample size
2. Variability in the outcome
3. The size of the difference of interest
4. The selected significance, or Type I error level (typically 0.05)
5. The statistical test being used

The Driver is the Unit of Analysis

Statistical power depends on:

1. Sample size
2. Variability in the outcome
3. The size of the difference of interest
4. The selected significance, or Type I error level (typically 0.05)
5. The statistical test being used

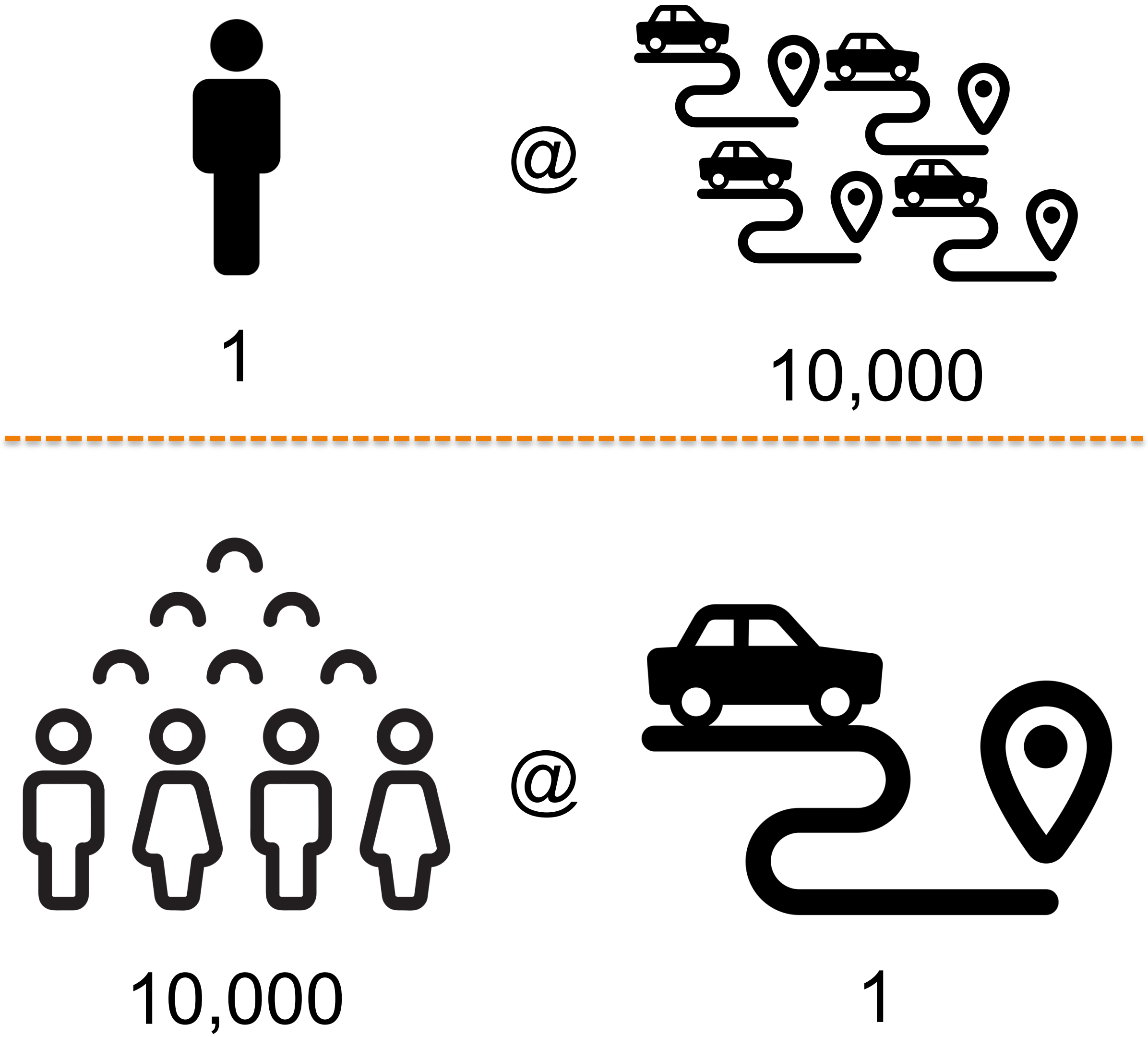
The Driver is the Unit of Analysis

Statistical power:

- More trips/person = reduces variability in the estimates of takeover time
- But sample size = # drivers

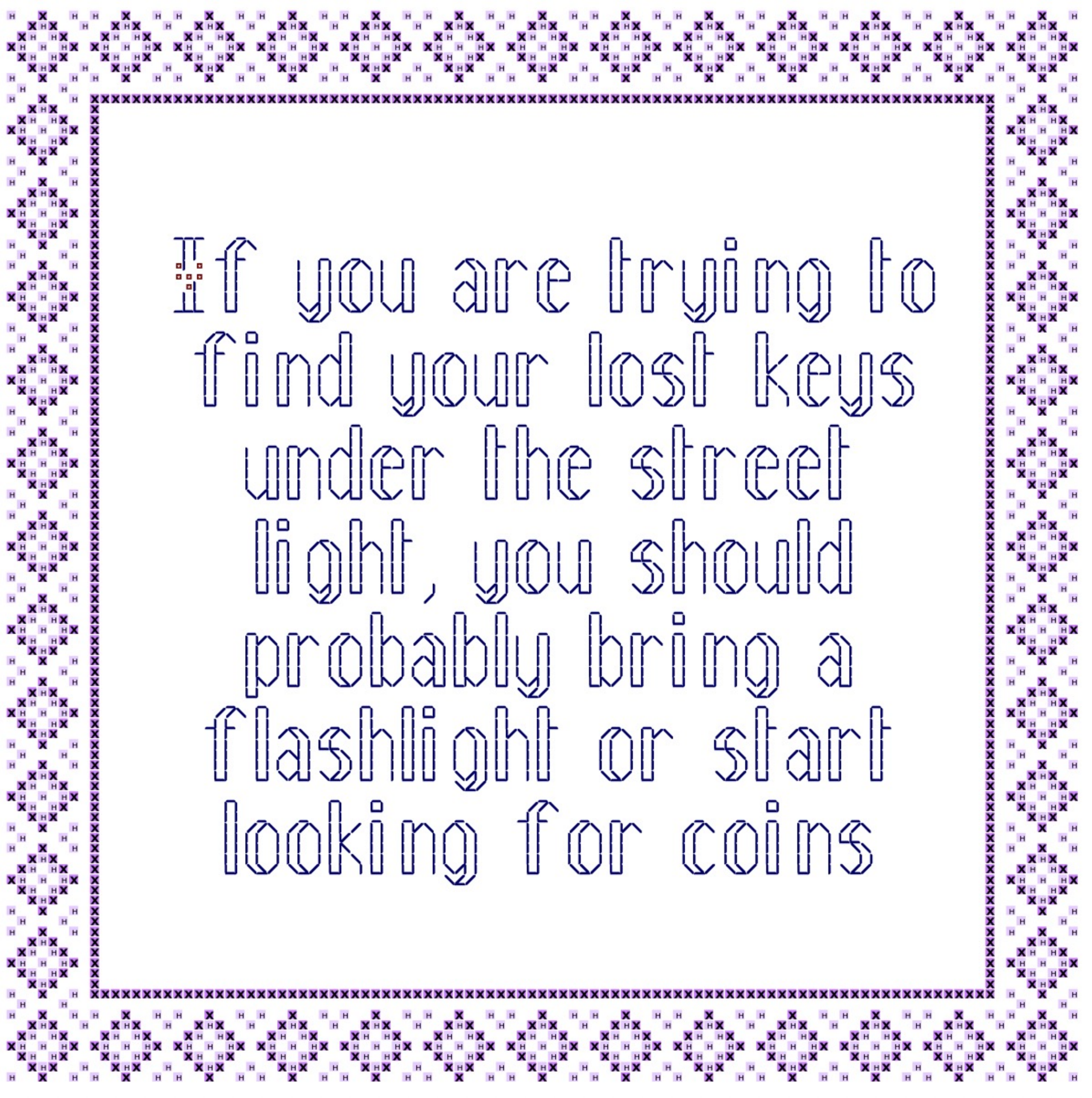
Generalizability:

- More observed variability sources = more generalizability



1 @ 10,000

10,000 @ 1

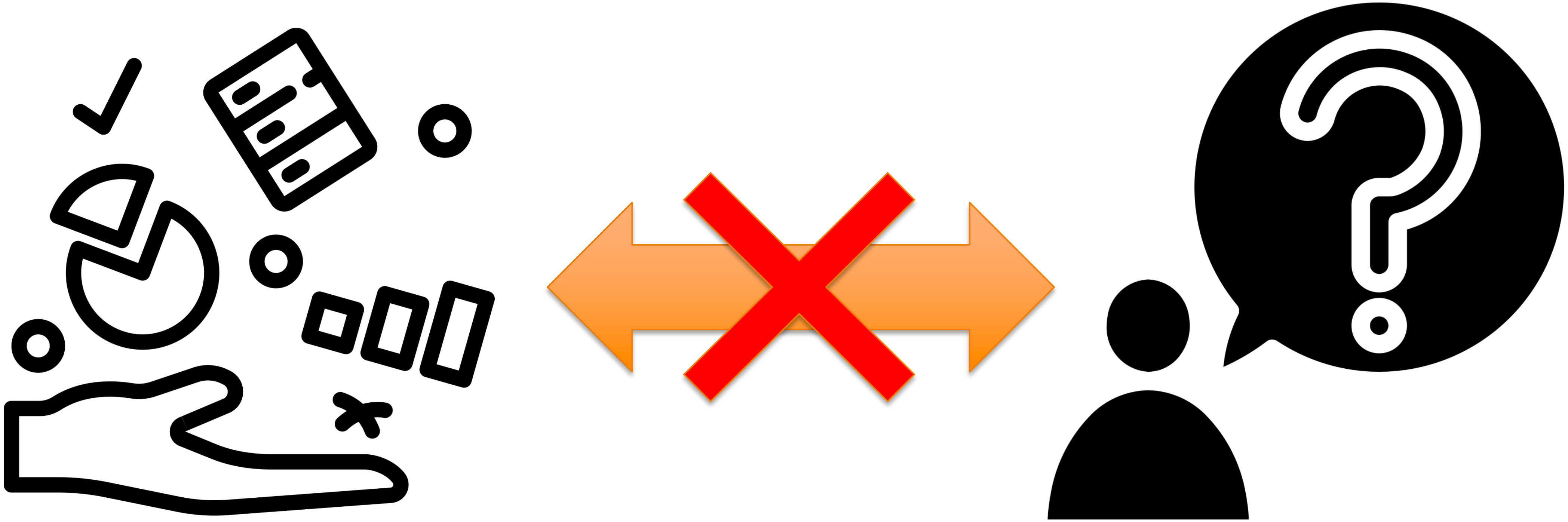


If you are trying to find your lost keys under the street light, you should probably bring a flashlight or start looking for coins

If you are trying to find your lost keys under the street light, you should probably bring a flashlight or start looking for coins

THIS IS WHERE YOU LOST YOUR WALLET?
NO, I LOST IT IN THE PARK. BUT THIS IS WHERE THE LIGHT IS.

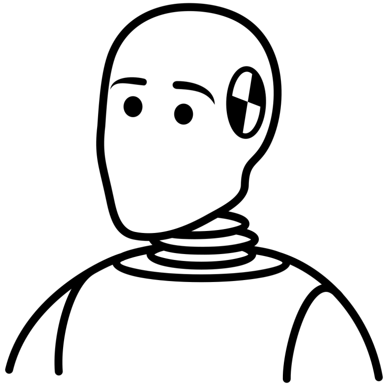
If you are trying to find your lost keys under the street light, you should probably bring a flashlight or start looking for coins



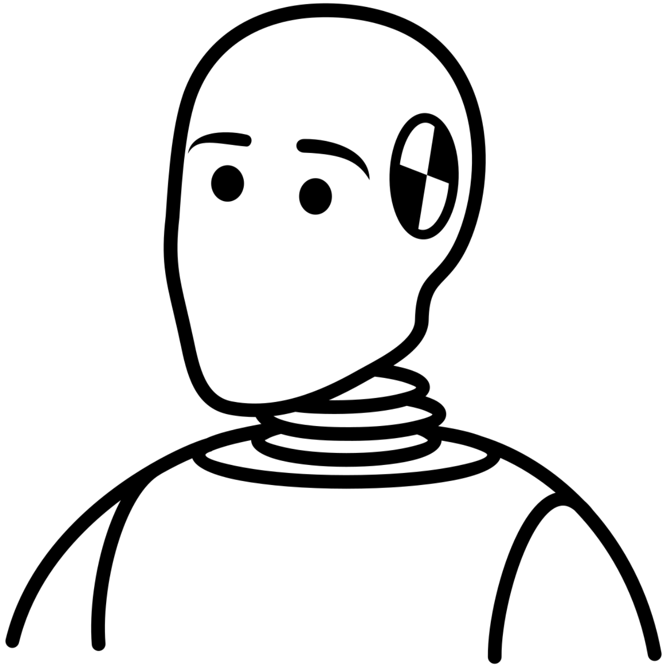
If you are trying to find your lost keys under the street light, you should probably bring a flashlight or start looking for coins

Adjust the research approach

1. Make inferences about the right things
2. Use surrogates

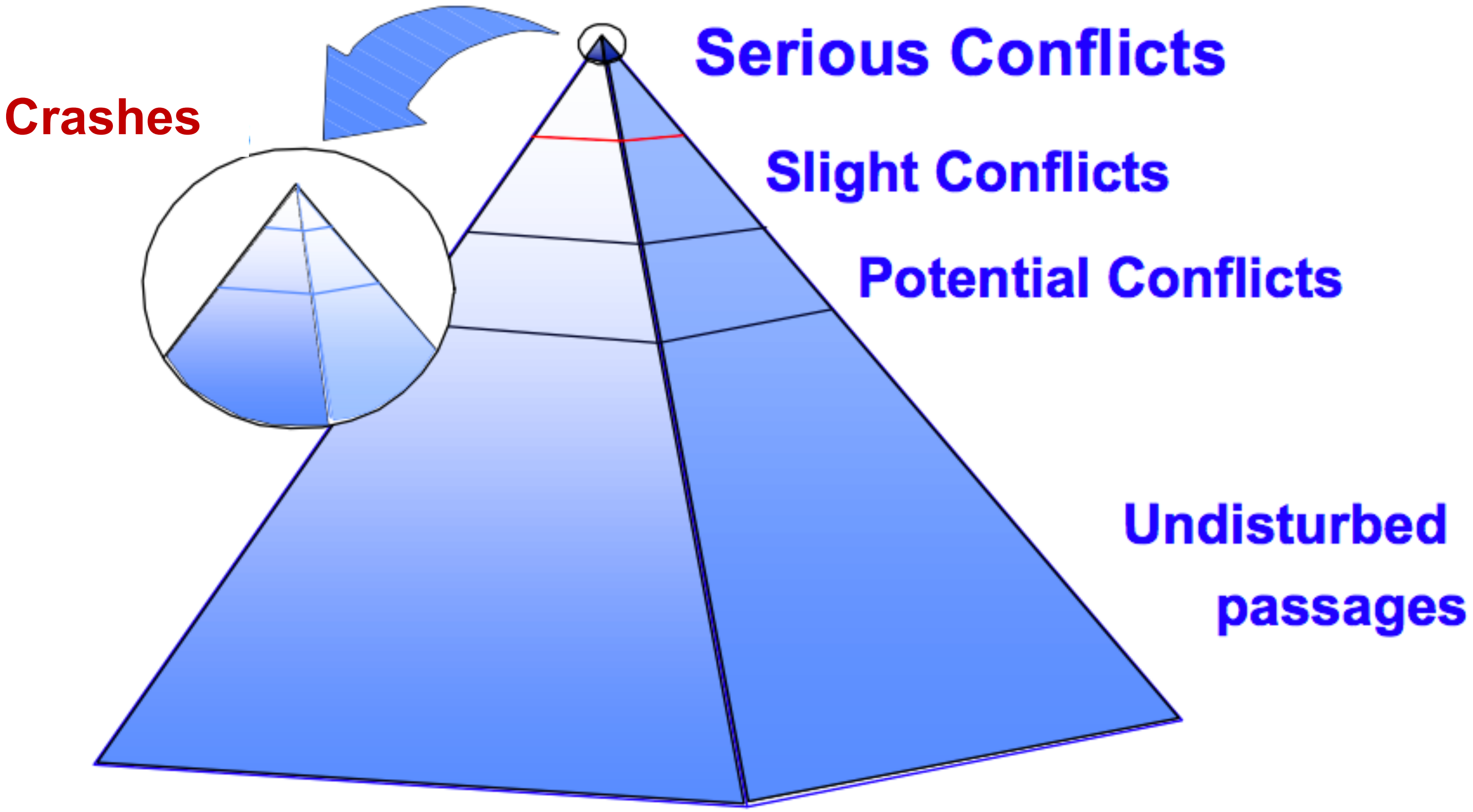


If you are trying to find your lost keys under the street light, you should probably bring a flashlight or start looking for coins



If you are trying to find your lost keys under the street light, you should probably bring a flashlight or start looking for coins

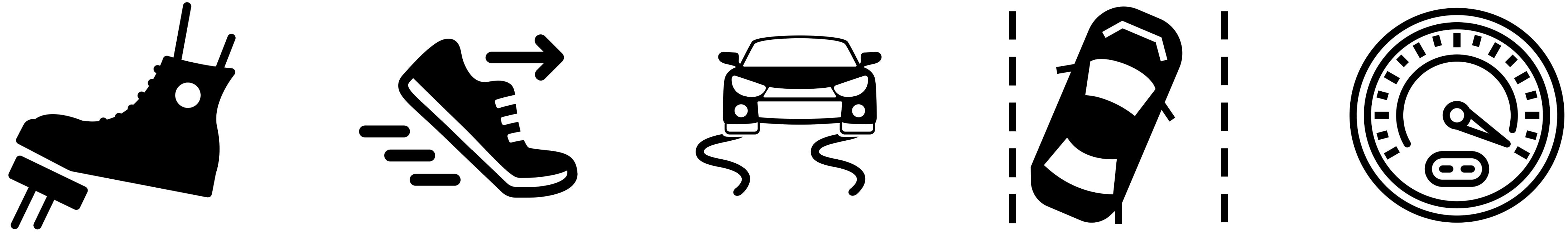
If you are trying to find your lost keys under the street light, you should probably bring a flashlight or start looking for coins



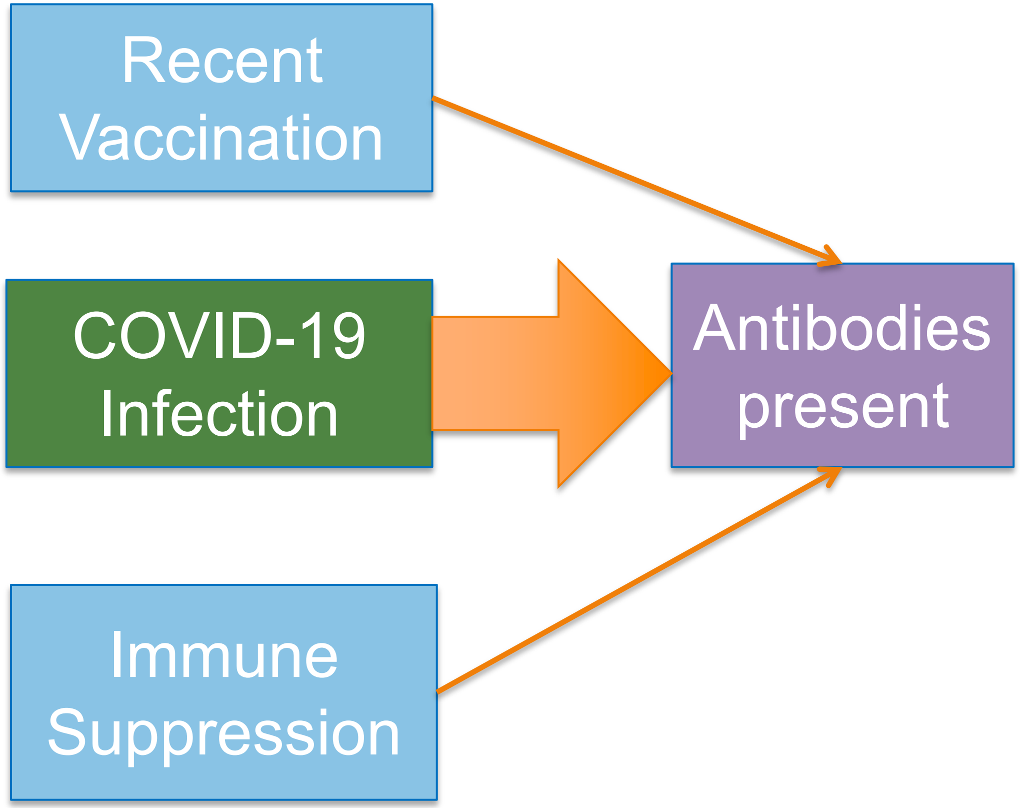
http://www.tft.lth.se/fileadmin/tft/video_in_traffic/Swedish_conflict_technique.pdf

If you are trying to find your lost keys under the street light, you should probably bring a flashlight or start looking for coins

For safety, common surrogates are *driver-controlled* kinematics



If you are trying to find your lost keys under the street light, you should probably bring a flashlight or start looking for coins



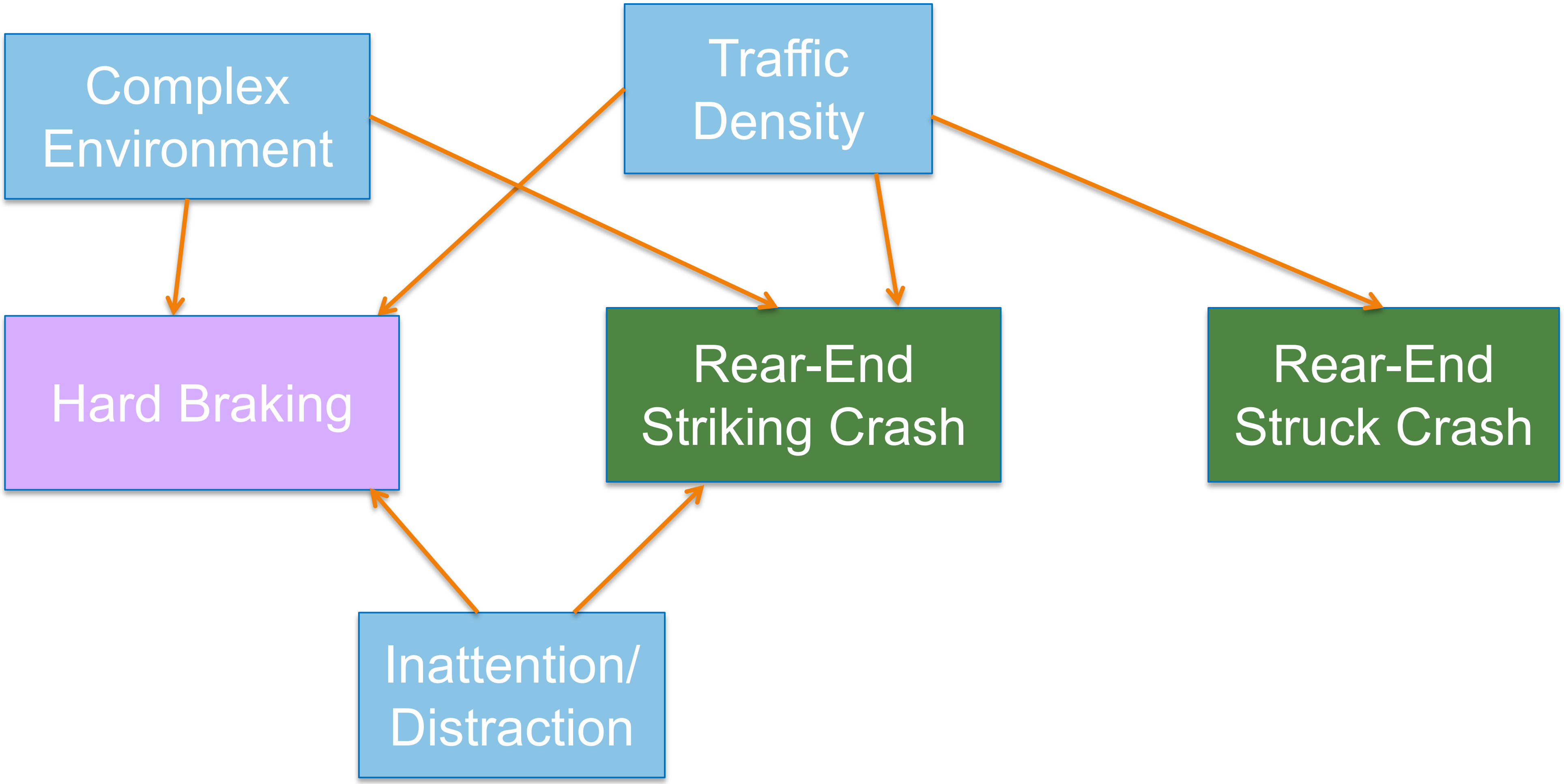
Recent Vaccination

COVID-19 Infection

Immune Suppression

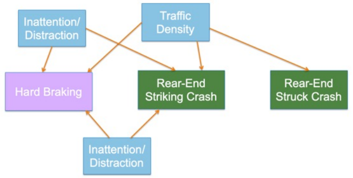
Antibodies present

If you are trying to find your lost keys under the street light, you should probably bring a flashlight or start looking for coins



If you are trying to find your lost keys under the street light, you should probably bring a flashlight or start looking for coins

What to do?



1. Know your surrogate causal model
2. Narrow your focus
3. Clarify interpretation

Context matters. A lot.

Context Matters. A Lot.

Environment, Vehicle, or Driver?

Drunk driving (predicts increased crash risk)

Unprotected left turns (predicts increased crash risk)

L3 ADS Mode and Driver takeovers (predicts increased crash risk)

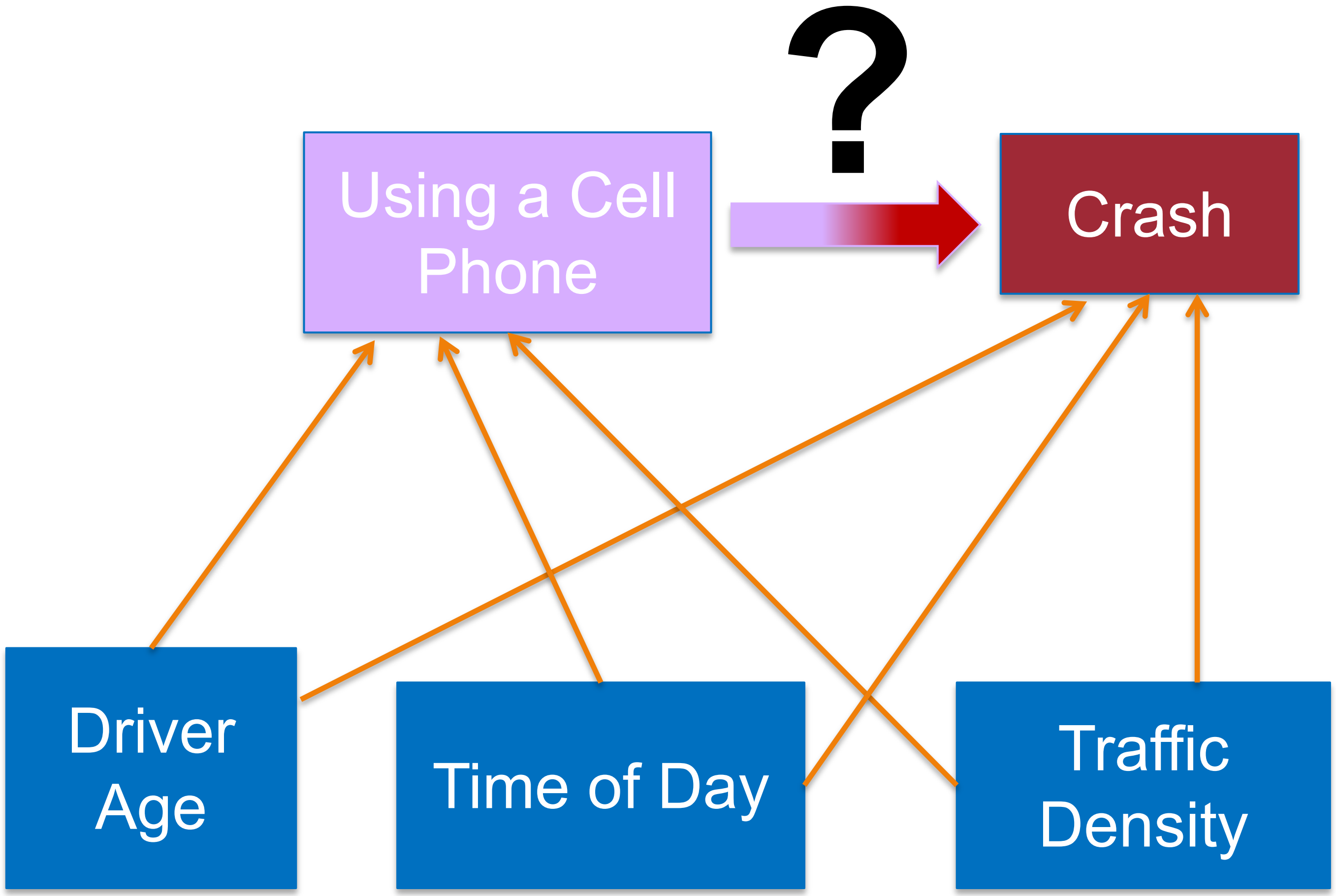
DSM [illegible] less [illegible] time (context-based)

Context Matters. A Lot.

Why do we care?

EXCHANGEABILITY

Context Matters. A Lot.

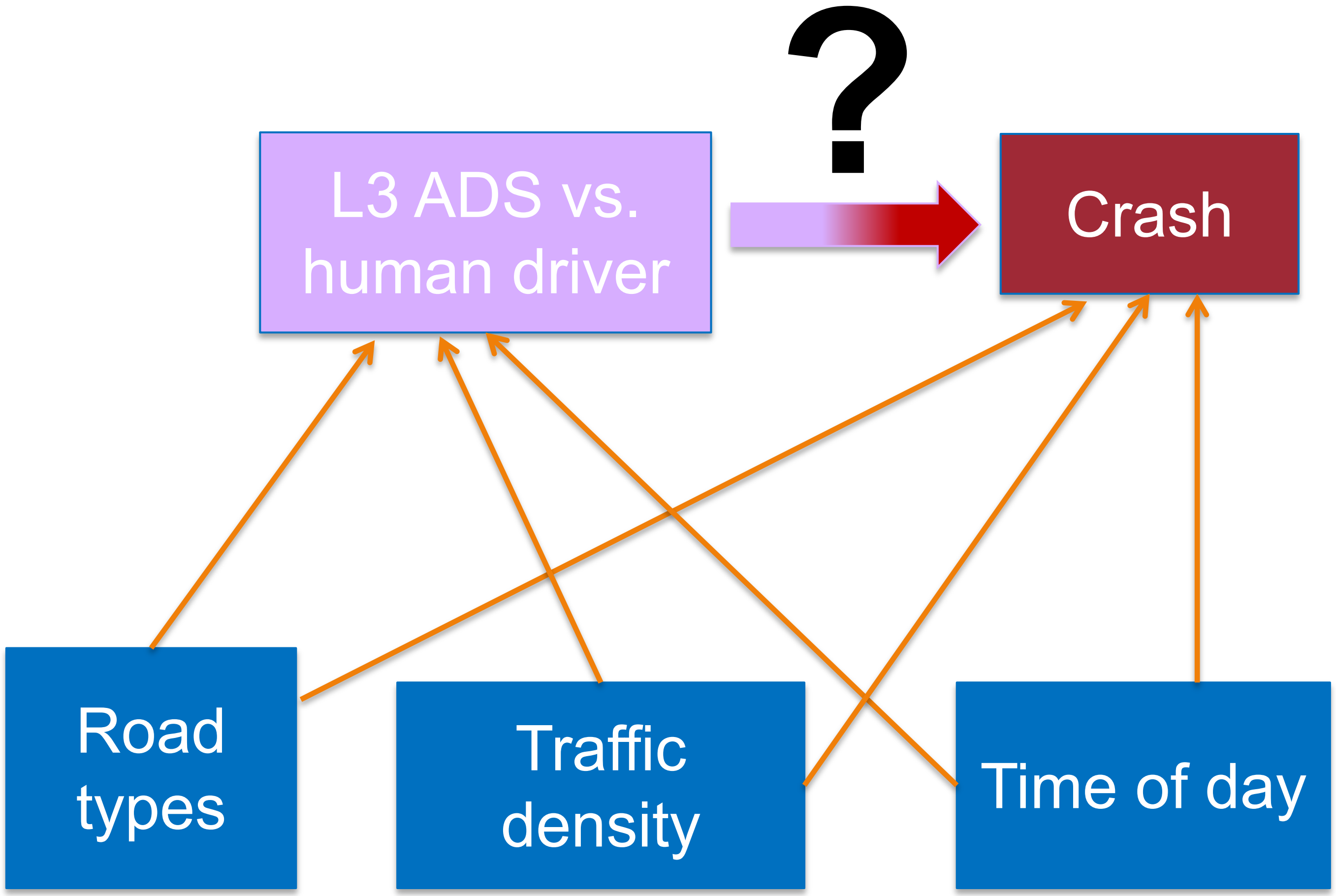


No adjustment:
OR = 2.38

With adjustment:
OR = 1.98

From: Flannagan, C., Bärgman, J., & Bálint, A. (2019). Replacement of distractions with other distractions: A propensity-based approach to estimating realistic crash odds ratios for driver engagement in secondary tasks. *Transportation research part F: traffic psychology and behaviour*, *63*, 186-192.

Context Matters. A Lot.



L3 ADS vs. human driver

?

Crash

Road types

Traffic density

Time of day

Statistics is the Unfun Parent of Data Science

Context matters. A lot.

There Is No Substitute for Using Your Brain

The driver is the unit of analysis

If you are trying to find your lost keys under the street light, you should probably bring a flashlight or start looking for coins

Questions?