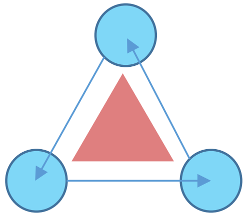


Indicator Frameworks

Joshua Tan (University of Oxford), Christine Kendrick (City of Portland), Abhishek Dubey (Vanderbilt University), and Sokwoo Rhee (NIST)

GCTC Expo

August 29, 2017

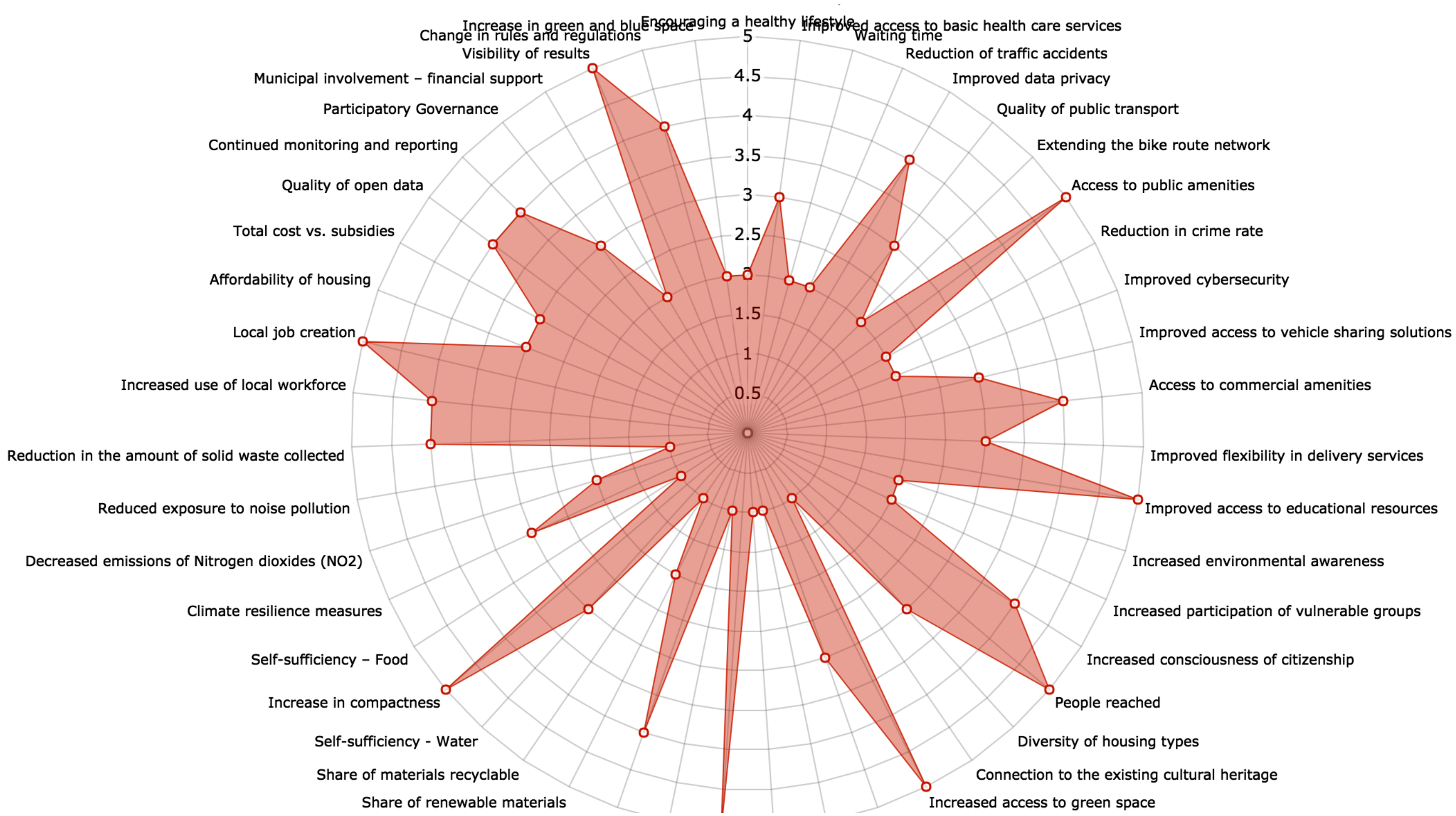
Start with the data you have,
not the data you want.

Why indicator frameworks?

- Sets the strategic **priorities** of your project, department, or city
- Communicate **progress** to stakeholders
- Enable **policy reactions** to data, especially in the optimization of processes
- **Simplify** your situation

How do we usually build indicators?

Indicator: “Access to public amenities”
Description: It is presumed that nearby availability of amenities leads to a lively neighbourhood and less car use. Amenities in the urban environment make an area more enjoyable and contribute to its desirability. It is assumed that these factors contribute to the success of smart city projects.
Definition: The extent to which public amenities are available within 500m
Calculation: Likert scale (1-5) 1. No amenities: no public amenities whatsoever are available (e.g. no basic nor additional). 2. Relatively few amenities: only few basic public amenities are available (e.g. a small park). 3. A reasonable number of amenities: basic public amenities are available including a few important amenities such as a park and a community center. 4. A sufficient number of amenities: basic public amenities are widely available (e.g. open green spaces, public recreation) as well as many important public amenities (theatres). 5. Relatively many amenities.
Expected data source: Google maps; project documentation and/or interviews with project leader, planning documents
Expected availability: High (everyone can access google maps); other relevant information should be available at the city planning office
Collection interval: After the project, but can also be used ex-ante to evaluate plans
Expected reliability: Because of the subjectivity that cannot be excluded, this indicator is not 100% reliable.
Expected accessibility: As a component of a successful project and selling point in a marketing sense, it is expected that this information will be accessible. No sensitivities expected.



Encouraging a healthy lifestyle
Improved access to basic health care services
Waiting time
Reduction of traffic accidents
Improved data privacy
Quality of public transport
Extending the bike route network
Access to public amenities
Reduction in crime rate
Improved cybersecurity
Improved access to vehicle sharing solutions
Access to commercial amenities
Improved flexibility in delivery services
Improved access to educational resources
Increased environmental awareness
Increased participation of vulnerable groups
Increased consciousness of citizenship
People reached
Diversity of housing types
Connection to the existing cultural heritage
Increased access to green space
Share of renewable materials
Share of materials recyclable
Self-sufficiency - Water
Increase in compactness
Self-sufficiency – Food
Climate resilience measures
Decreased emissions of Nitrogen dioxides (NO2)
Reduced exposure to noise pollution
Reduction in the amount of solid waste collected
Increased use of local workforce
Local job creation
Affordability of housing
Total cost vs. subsidies
Quality of open data
Continued monitoring and reporting
Participatory Governance
Municipal involvement – financial support
Visibility of results
Change in rules and regulations
Increase in green and blue space
5
4.5
4
3.5
3
2.5
2
1.5
1
0.5

Problems with the one-by-one approach

- It’s expensive.
- It’s subjective.
- It’s ad hoc.
- ~~Is it worth it?~~
- There’s just not enough data.*

* data that you **want** vs data that you **have**

Towards a science of measuring systems

- Cities are **cyber-physical systems**
- Cities are **systems of systems**
- These models are *mathematical descriptions*. They do not measure anything, per se.
- Other mathematical descriptions:
  - Network approaches
  - Economic models
  - Game theory

Problems with the systems approach

- People end up building toy models
  - See 90% of academic studies
- Or they build highly specific, technical models… thus not “systems”
- Or they build giant, unwieldy models
  - 175 indicators in CITYKeys, 212 in the Boston Indicators Project, etc.
  - More than 43 (!) indicator frameworks built for “smart and sustainable cities”
- The point is: models are useful locally, but they’re hard to sustain on bigger systems

Towards a science of measuring cities

- Problem: the **models** fail to describe the world perfectly. (Duh)
- Problem: the **data** doesn't either.*

* cities are **complex**: there isn't enough raw data
to describe all the interactions

The hypothesis

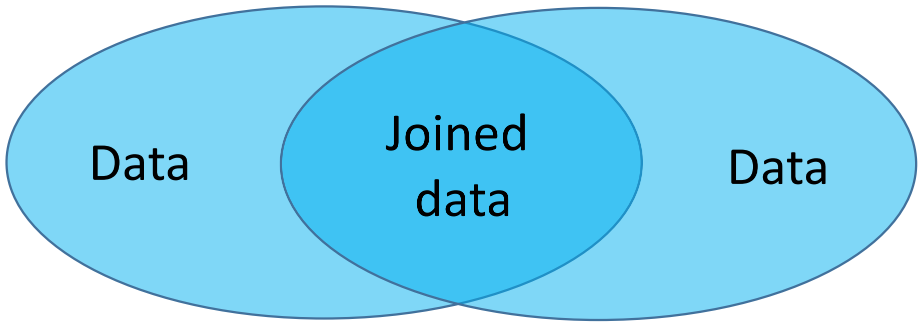
Instead of constructing indicator frameworks expensively and internally, meaning indicator-by-indicator, we can specify them abstractly and externally, by means of their causal and statistical relationships to other, already-extant sets of indicators. Our approach is especially suited to situations where heterogeneous data is distributed across many projects and many localities.

- Step 1: give a **mathematical semantics** for indicator frameworks.
- Step 2: test whether indicator frameworks can be upgraded to synthesize "**models** over **data**".

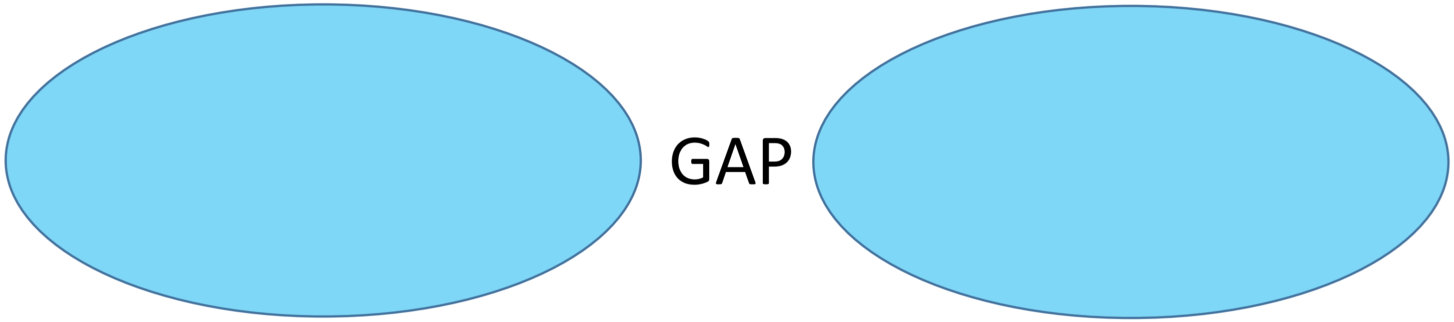
Data informs models,
while models constrain data.

How do we construct operational indicator frameworks?

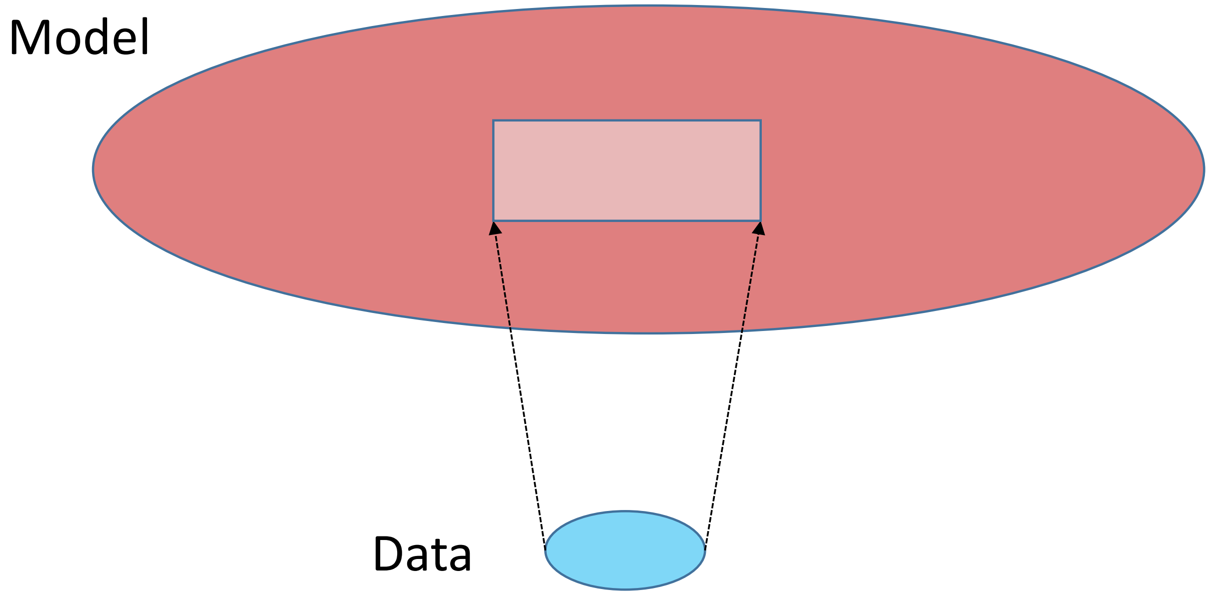
- Expert input, Likert scales, but also…
- We can refine old indicators by “cleaning” the data.
- We can combine two indicators A, B into a new indicator by “joining them”, i.e. forming A x B.
- We can compute or infer correlations between indicators, then describe proxies.



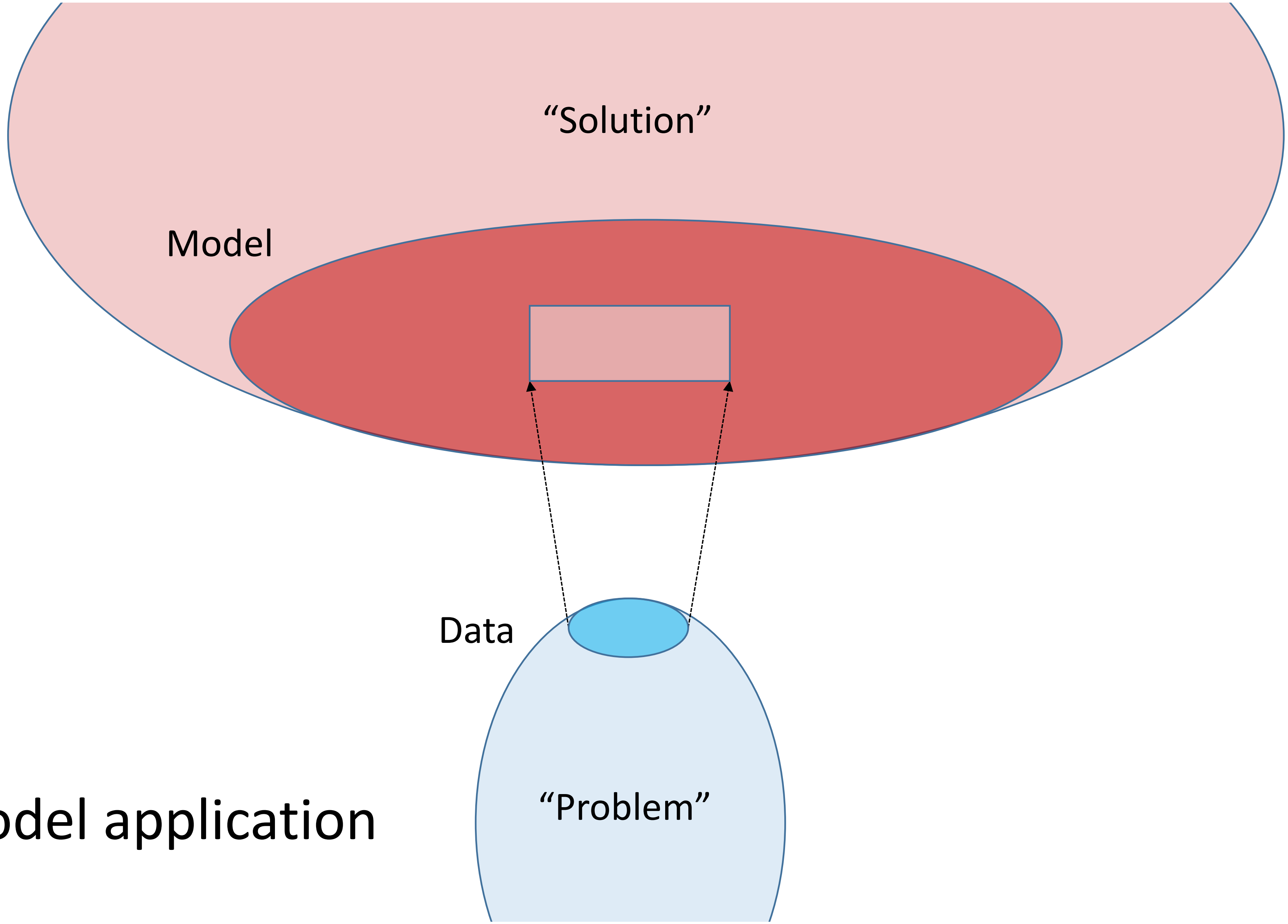
Joined data or “tensored” data



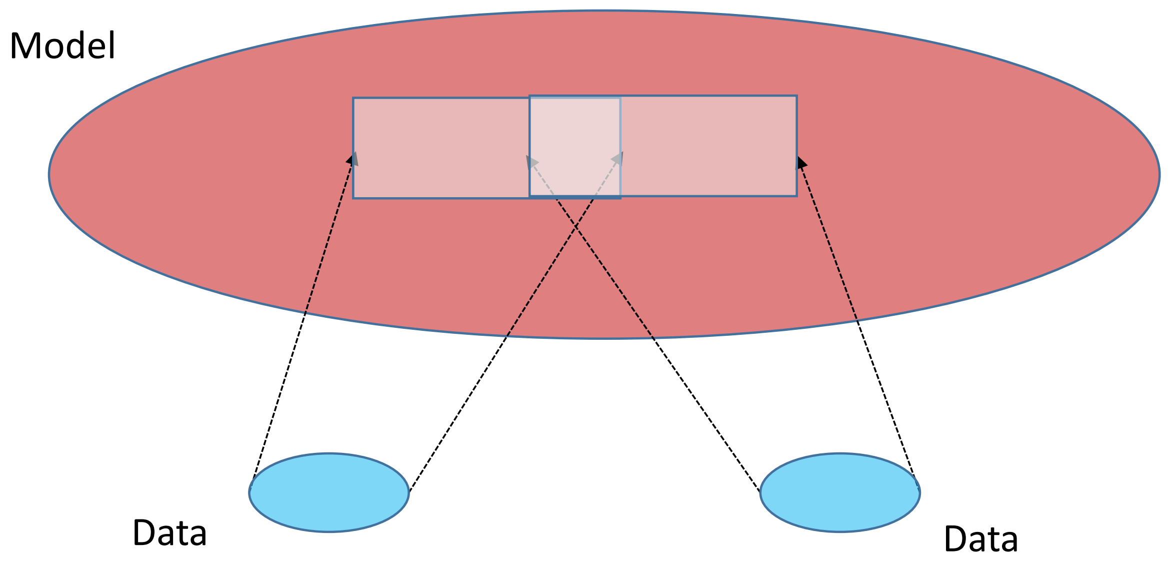
How do we join this data?



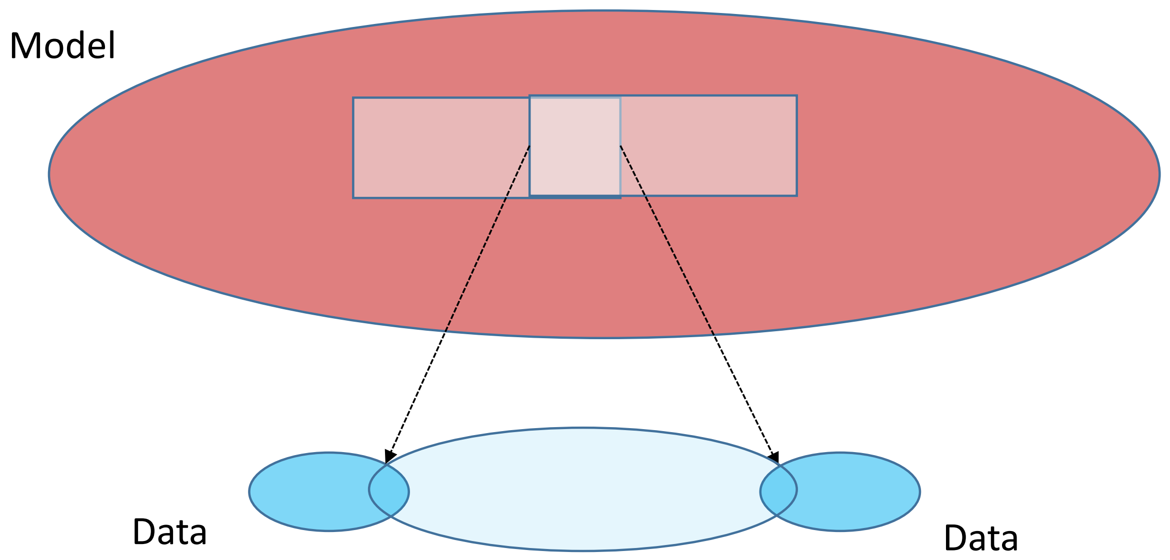
Model construction



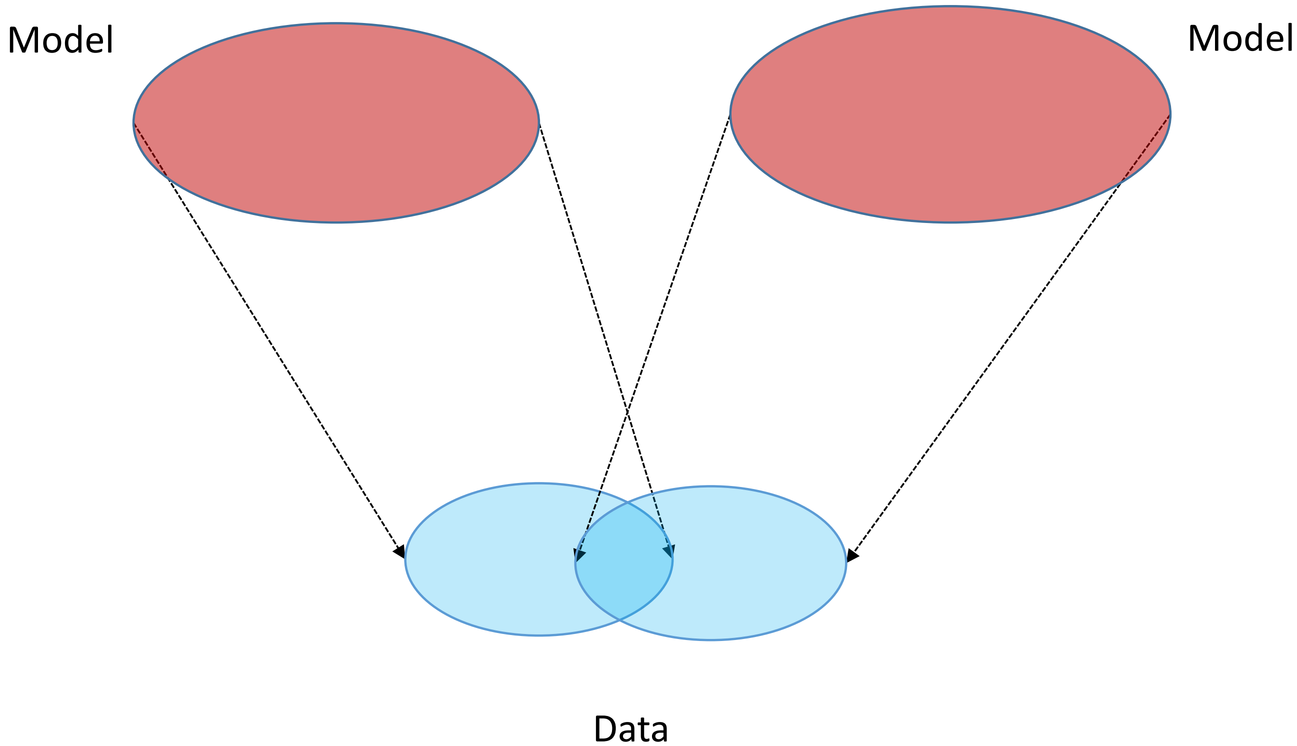
Model application



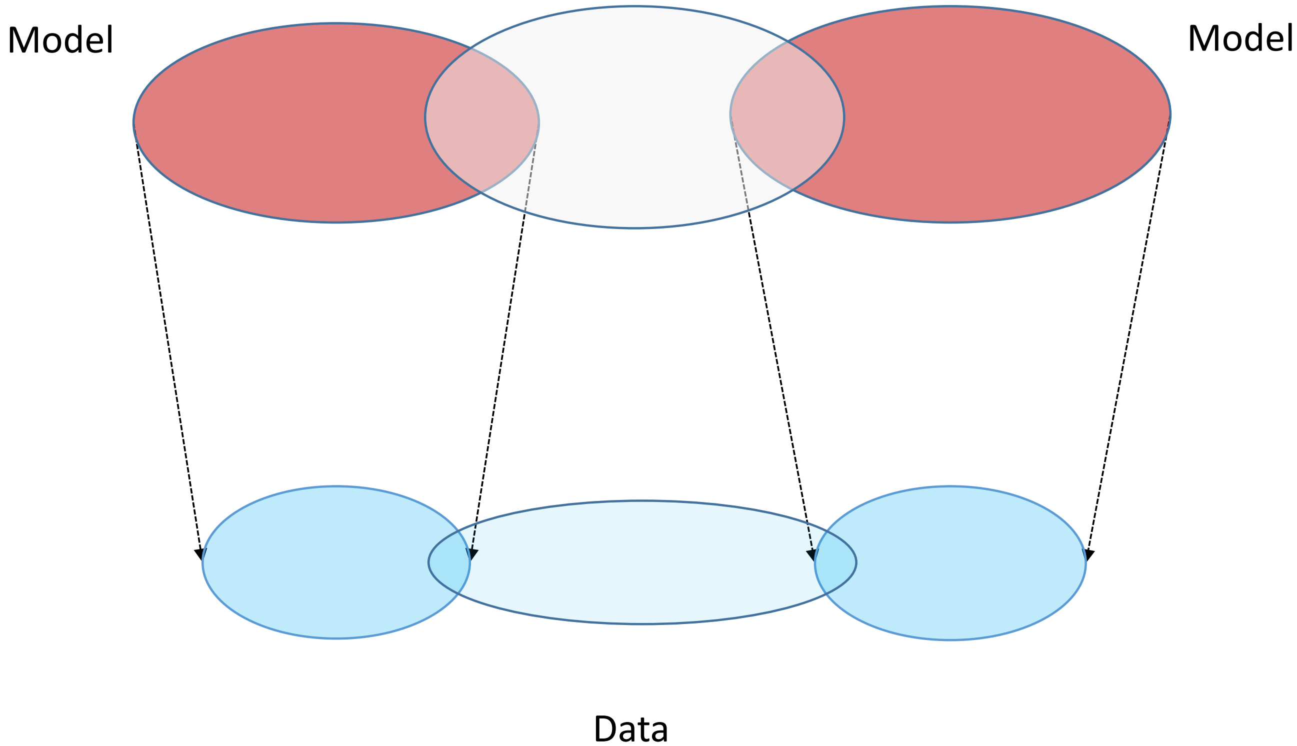
Model validation



“Integrating under a model”



Data simulation & integration



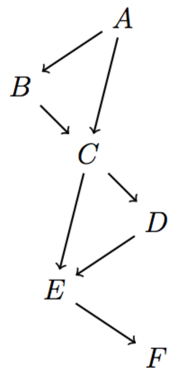
The general case?

Abstract indicator frameworks

- An abstract indicator framework is composed of:
  1. A **R**-valued matrix whose columns represent indicators and rows represent **data**
  2. An inner product operation between indicators, understood as their sample **correlation**
- The set of all abstract indicator frameworks forms something called a category

A simple model: causal diagrams

- Causal diagrams are directed acyclic graphs, e.g.

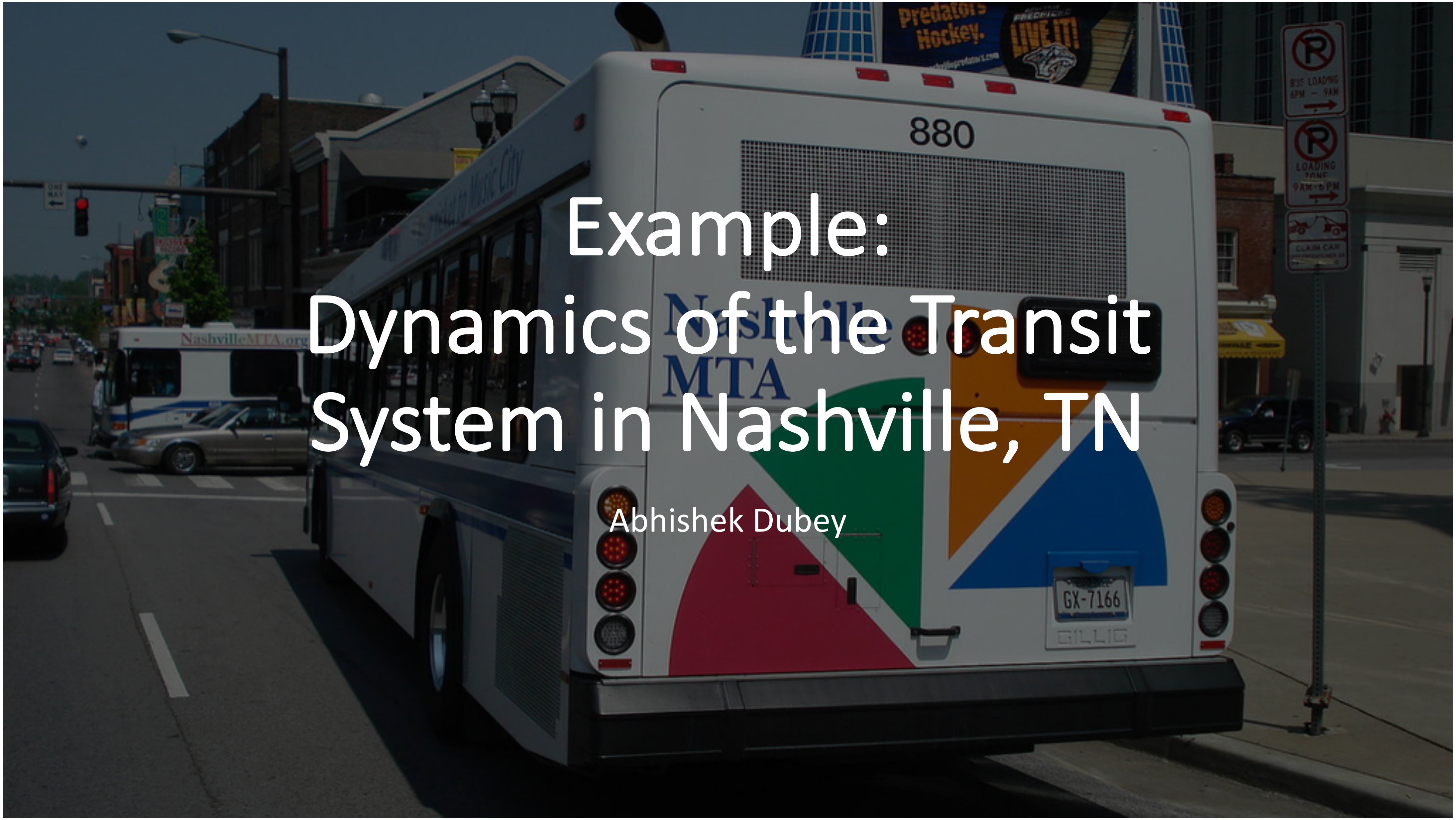


- Each causal diagram can be used to construct a monoidal category, called a *causal theory* [Fong '13].
- Each graph, essentially, defines a procedure for computing testing a **causal relationship** over a set of **correlations**.

Activity: build your own diagram!

Example: Dynamics of the Transit System in Nashville, TN

Abhishek Dubey



Date	Weekday (Mon=0)	Segment	Direction	Route	Actual Travel (Seconds)	Scheduled Travel (Seconds)	Actual Arrivel Time (24hr)	Trip_ID	Driver_ID	Jam Factor (0~10)	Actual Traffic Speed (mph)	Free Flow Speed Limit (mph)
20161011	1	MCC5_11 - HFOGG	FROM DOWNTOWN	1	826.99998	360	15:35:39	126346	1683	3.487592973	13.24569763	19.82003285
20161011	1	HFOGG - MTWD	FROM DOWNTOWN	1	580.99998	600	15:48:05	126346	1683	3.937953674	11.96260066	18.78503114
20161011	1	MTWD - 100O	FROM DOWNTOWN	1	805.99998	660	16:01:31	126346	1683	2.793803409	17.05275554	23.36003872
20161011	1	MCC5_11 - HFOGG	FROM DOWNTOWN	1	802.99998	420	16:38:47	126347	1683	5.28474093	10.79350626	19.82003285
20161011	1	HFOGG - MTWD	FROM DOWNTOWN	1	648	600	16:49:35	126347	1683	4.822438899	10.89503857	18.78503114
20161011	1	MTWD - 100O	FROM DOWNTOWN	1	507.99996	600	16:58:03	126347	1683	0.024484444	26.24559907	23.36003872
20161011	1	MCC5_11 - HFOGG	FROM DOWNTOWN	1	900.99996	420	17:37:25	126348	1683	4.263709706	12.1785496	19.82003285
20161011	1	HFOGG - MTWD	FROM DOWNTOWN	1	609.99996	540	17:47:35	126348	1683	4.862932428	10.63845745	18.78503114

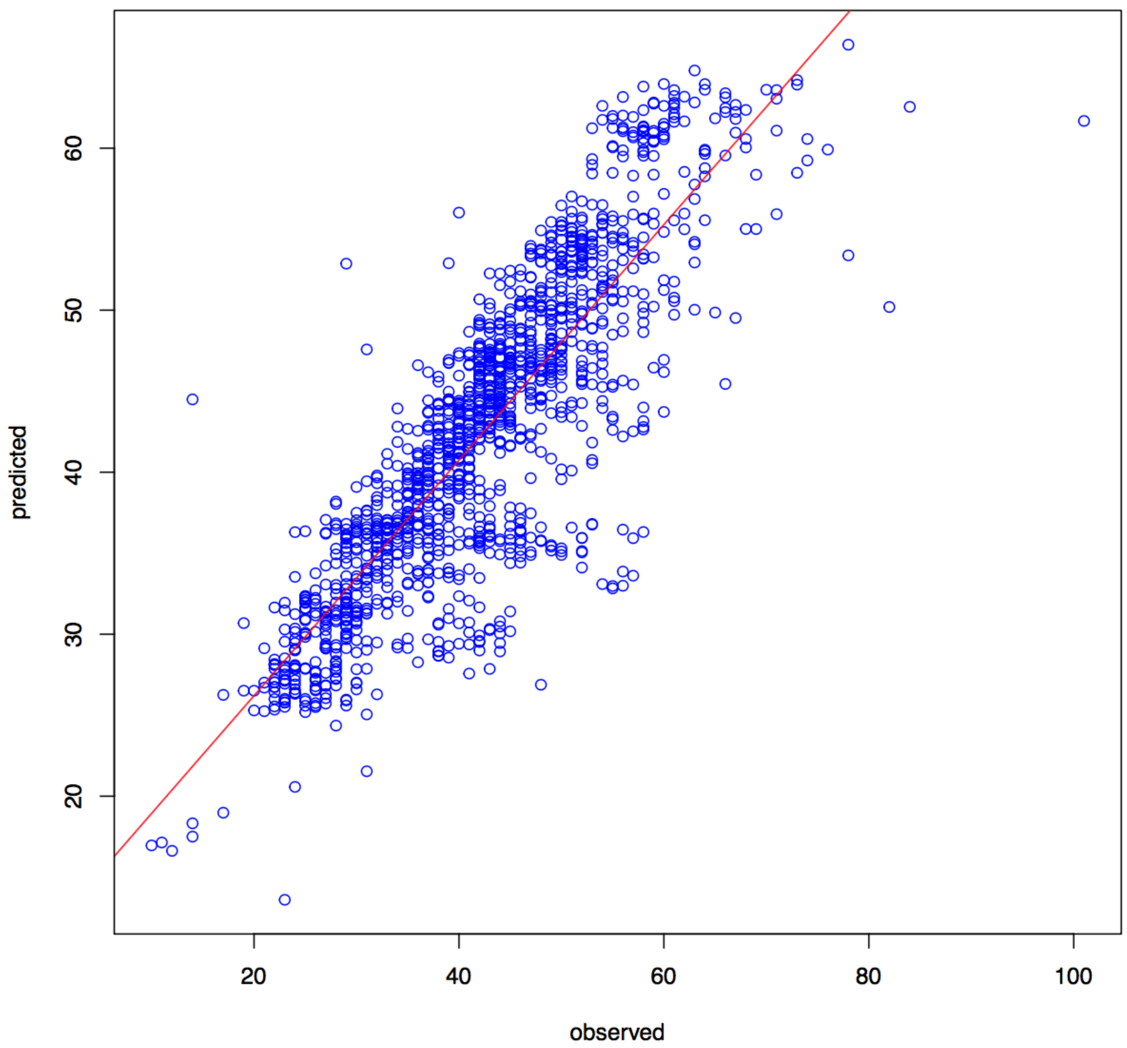
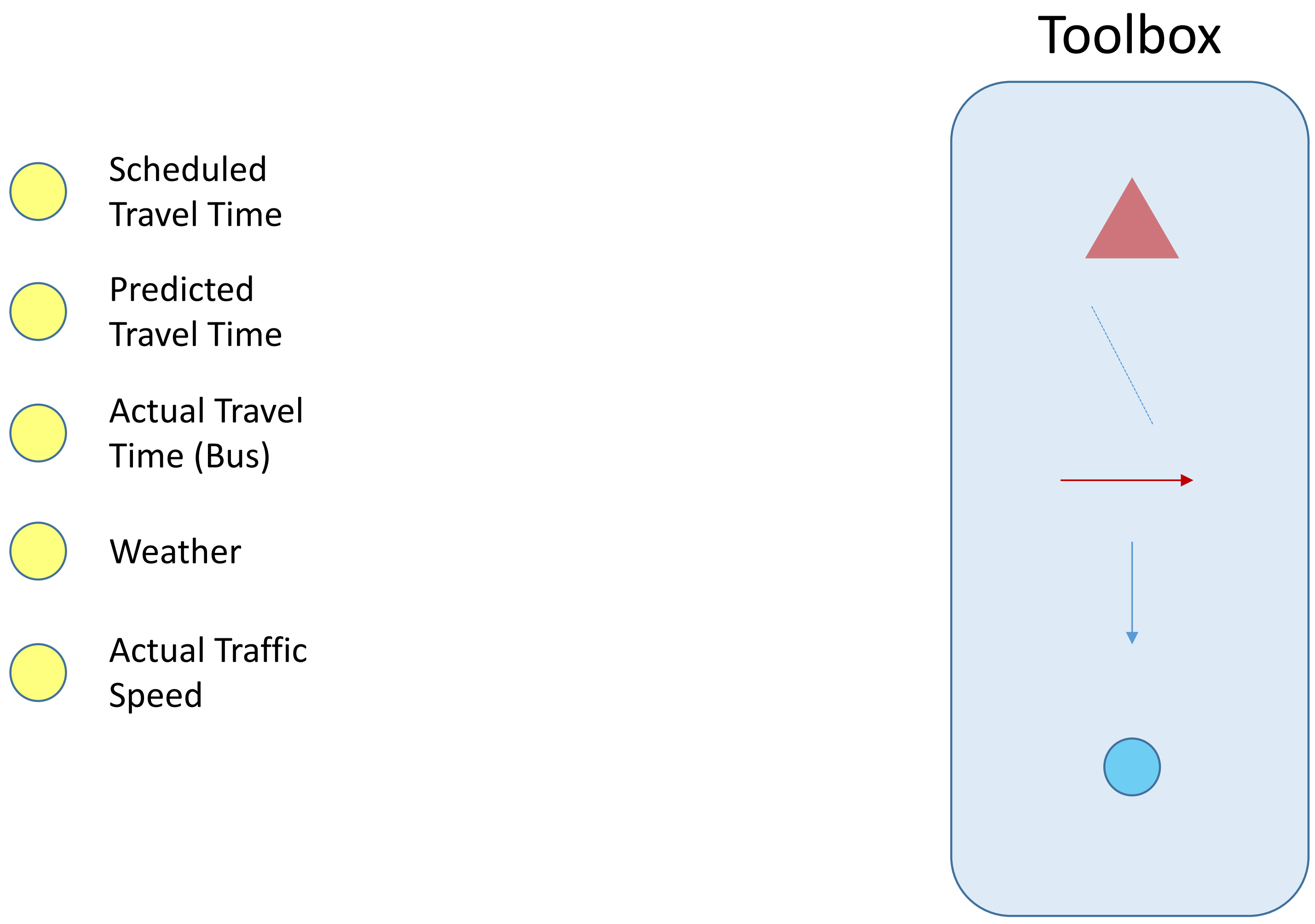
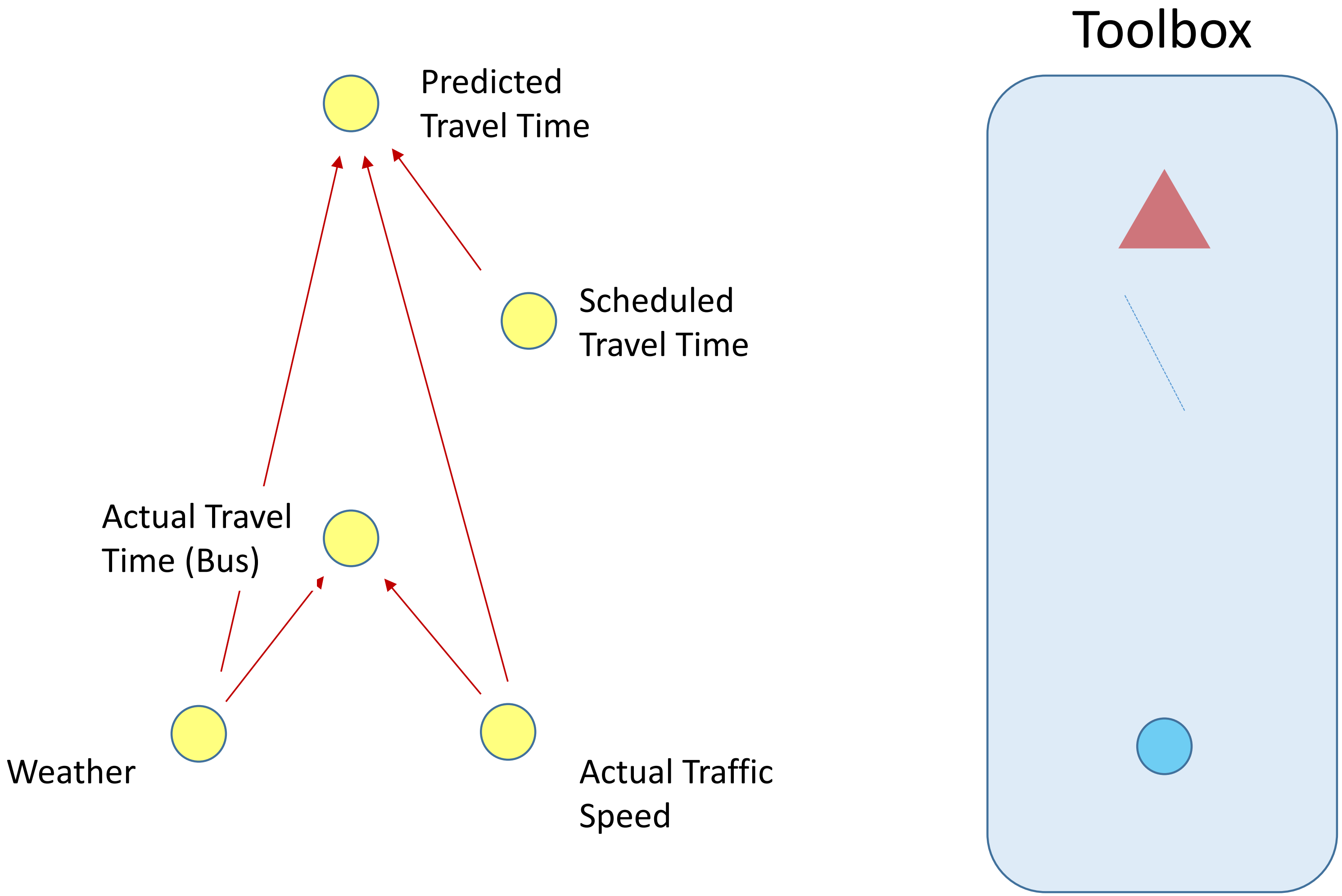
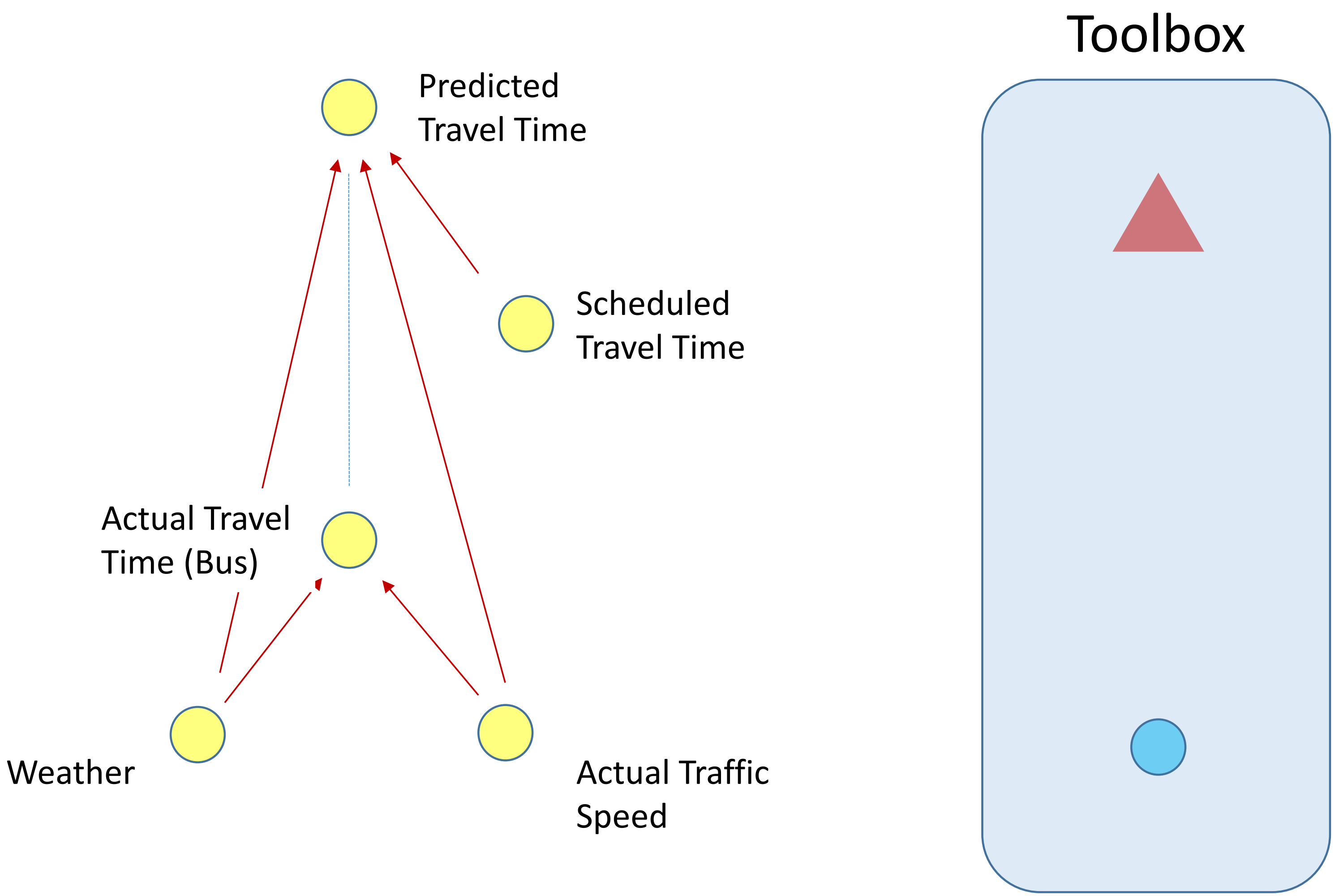


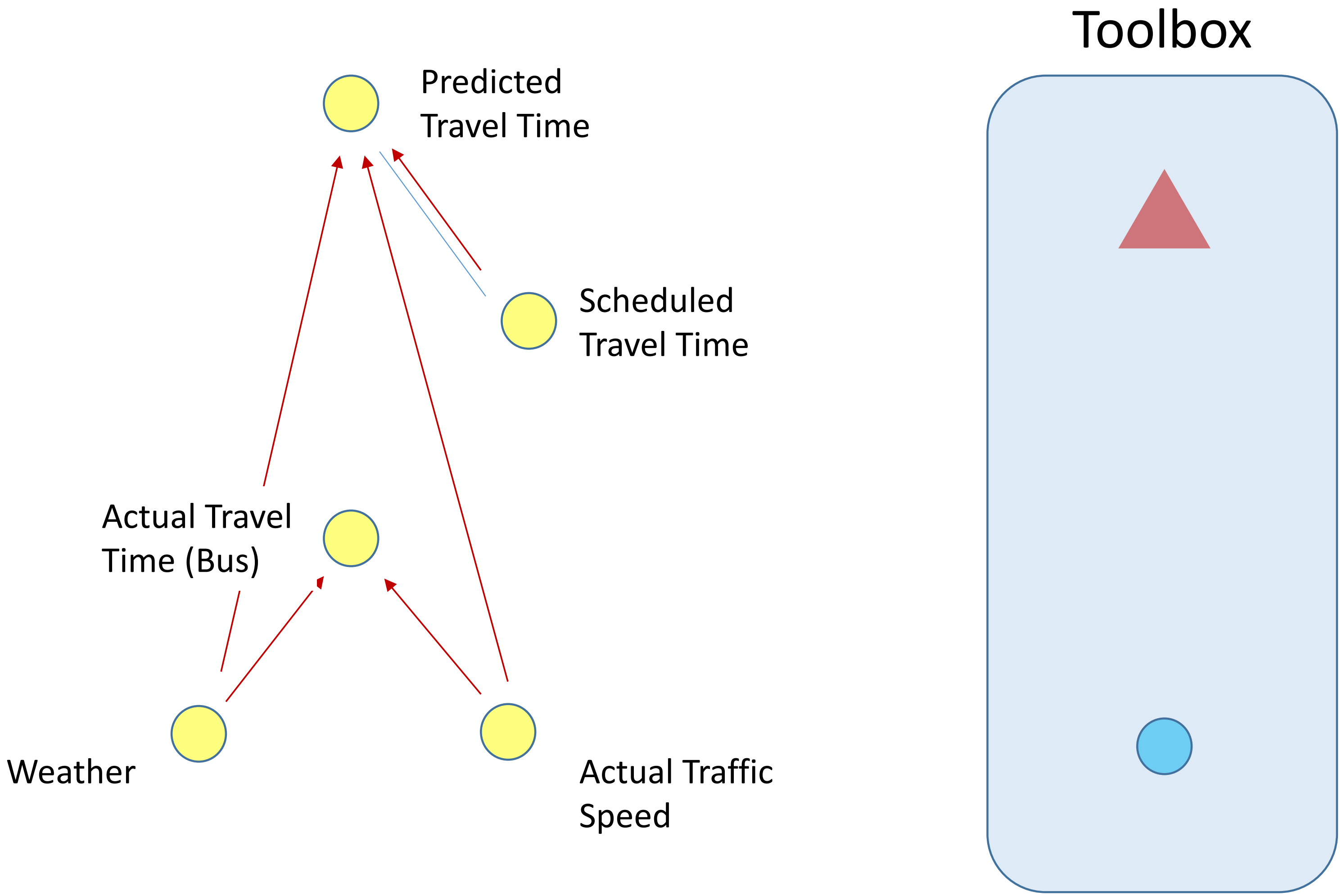
Figure 2. Observed Versus Predicted Travel Time.

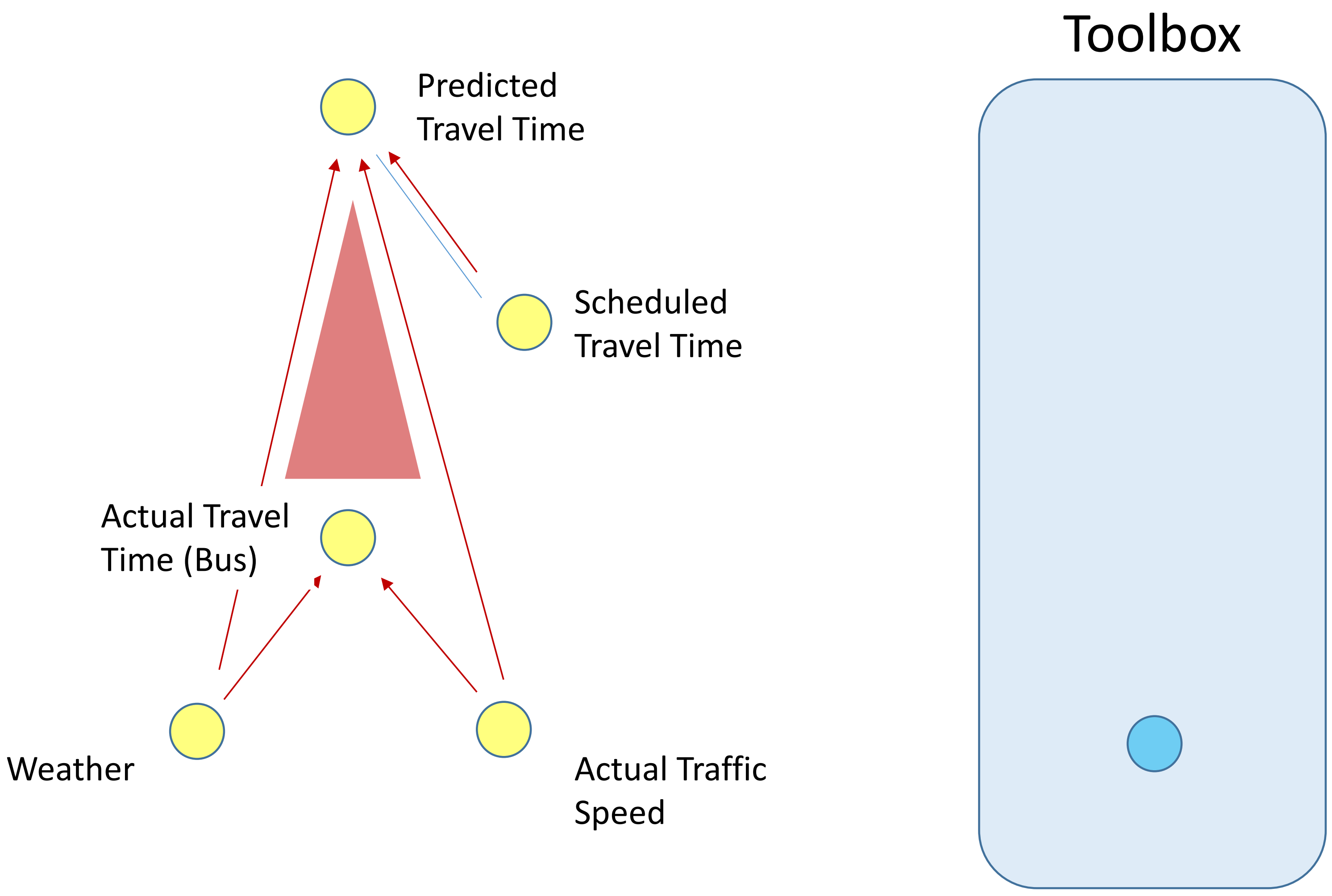

Scheduled Travel Time
Predicted Travel Time
Actual Travel Time (Bus)
Weather
Actual Traffic Speed
Toolbox



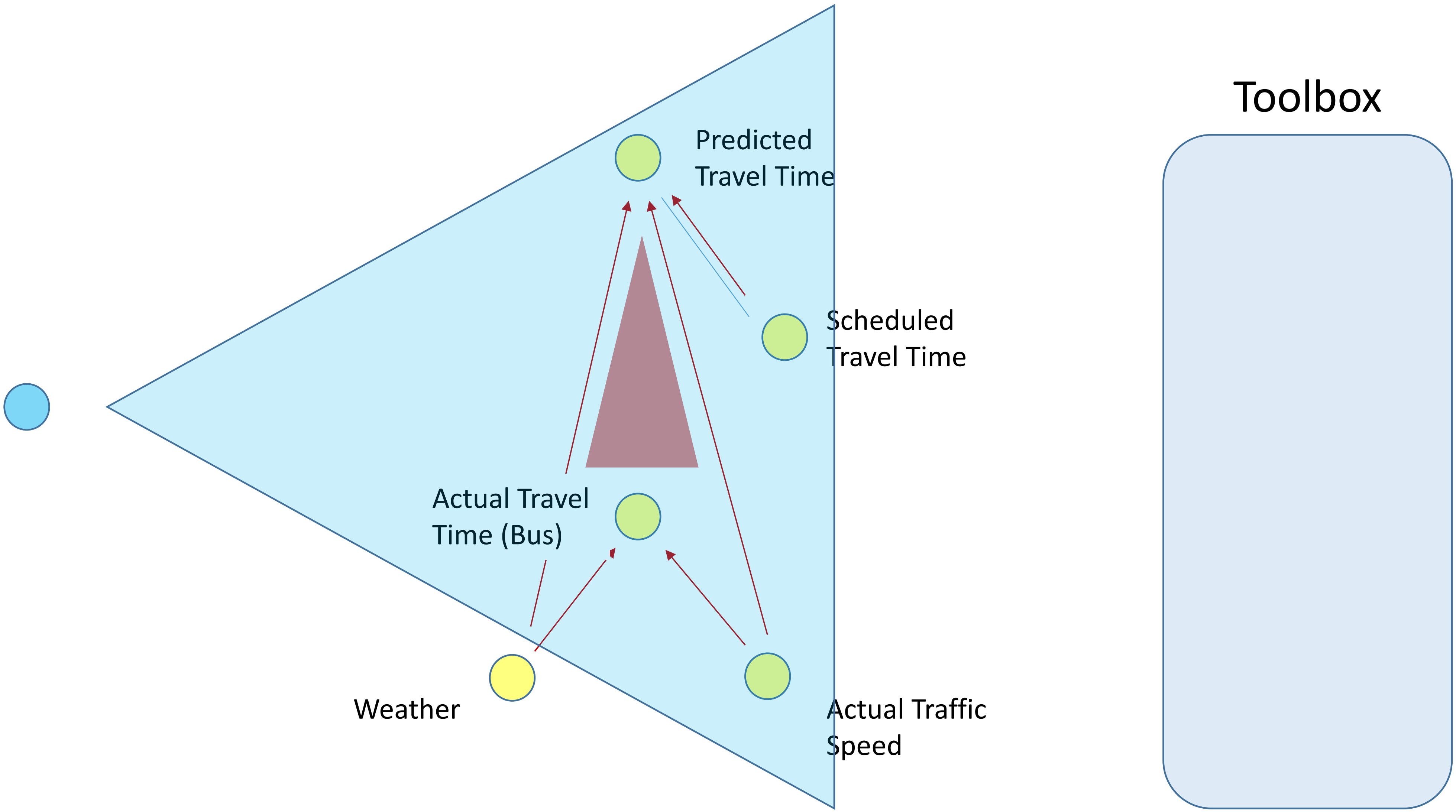
Predicted
Travel Time
Scheduled
Travel Time
Actual Travel
Time (Bus)
Weather
Actual Traffic
Speed
Toolbox


Predicted
Travel Time
Scheduled
Travel Time
Actual Travel
Time (Bus)
Weather
Actual Traffic
Speed
Toolbox


Predicted
Travel Time
Scheduled
Travel Time
Actual Travel
Time (Bus)
Weather
Actual Traffic
Speed
Toolbox



Predicted
Travel Time
Scheduled
Travel Time
Actual Travel
Time (Bus)
Weather
Actual Traffic
Speed
Toolbox



Predicted
Travel Time
Scheduled
Travel Time
Actual Travel
Time (Bus)
Weather
Actual Traffic
Speed
Toolbox

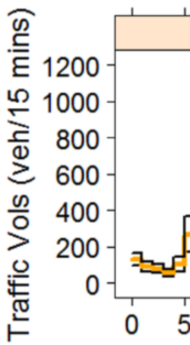
Understanding the Roadside Environment through Integrating Air Quality & Traffic-Related Data

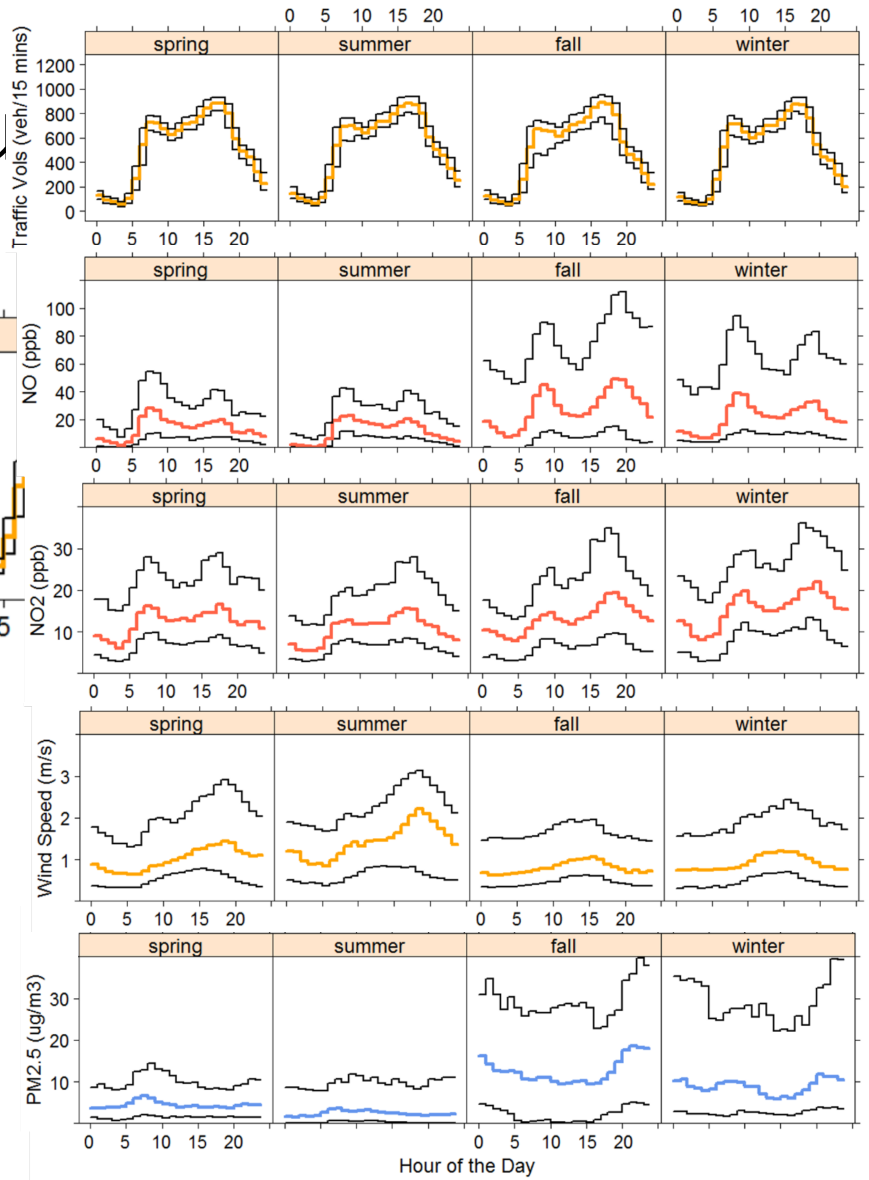
Christine Kendrick

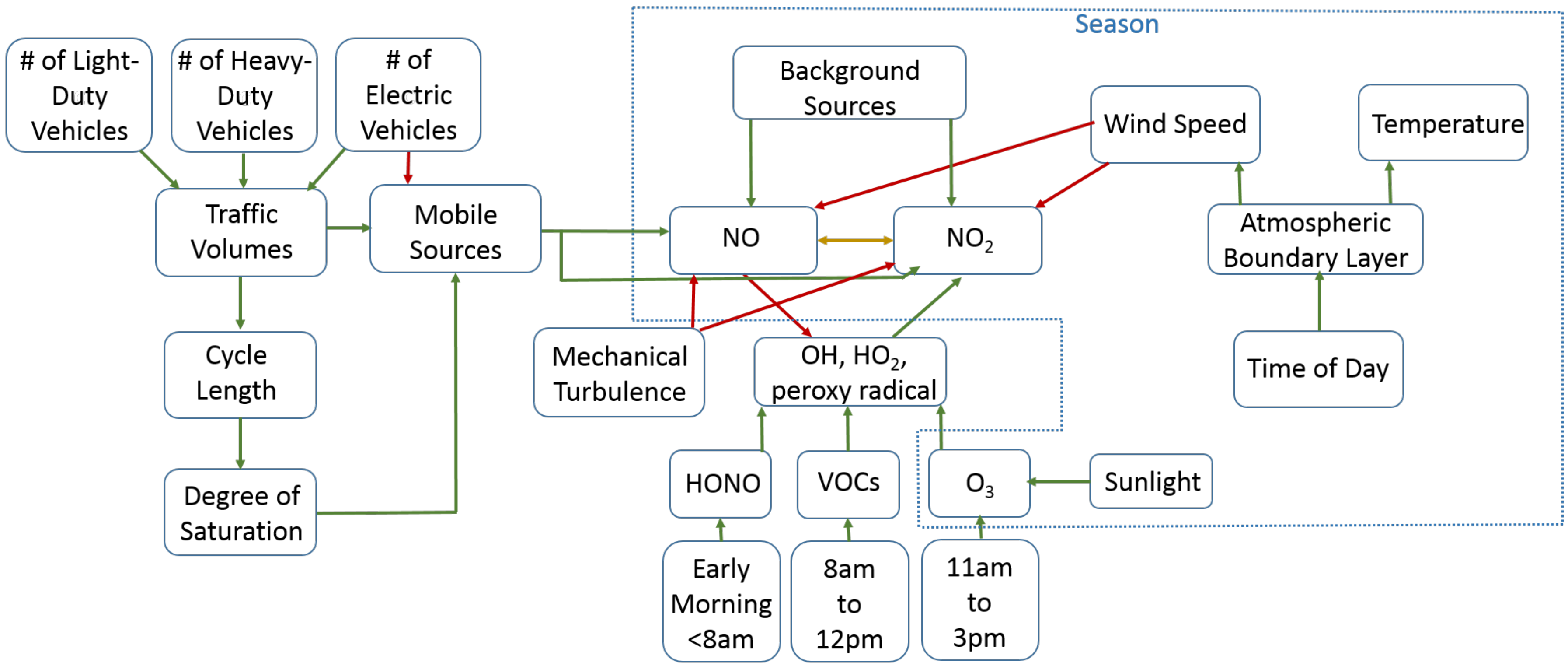
Findings

Season	Time Period	Coefficient NO per 100 vehicles per 15 mins	Standard Error of Coefficient NO per 100 vehicles per 15 mins	Adjusted r^2	Coefficient NO_2 per 100 vehicles per 15 mins	Standard Error of Coefficient NO_2 per 100 vehicle per 15 mins	Adjusted r^2
Fall	Morning	6.3 ** (7.9)**	1.4 (0.4)	0.1 (0.16)	1.2** (1.6)**	2.3 (0.1)	0.14 (0.24)
Winter	Morning	9.4** (11.2)**	1.4 (0.7)	0.14 (0.24)	1.5** (1.9)**	0.4 (0.2)	0.17 (0.26)
Spring	Morning	6.3** (6.7)**	0.8 (0.3)	0.41 (0.43)	2.5** (2.3)**	0.3 (0.2)	0.27 (0.28)
Summer	Morning	4.6** (4.4)**	(0.4) (0.2)	(0.45) (0.37)	1.3** (1.6)**	0.3 (0.1)	0.25 (0.23)
Fall	Evening	-1.3 (-1.6)	2.5 (1.1)	<0.001 (0.005)	1.2* (0.9)	0.6 (0.2)	0.05 (0.03)
Winter	Evening	-0.04 (0.09)	2.6 (1.3)	0.002 (0.001)	-0.4 (-0.08)	0.8 (0.4)	0.05 (0.03)
Spring	Evening	2.1 (2.2)**	1.1 (0.5)	0.03 (0.04)	0.9 (0.9)**	(0.8) (0.3)	0.007 (0.02)
Summer	Evening	2.1* (2.9)**	0.8 (0.3)	0.02 (0.07)	1.6* (1.9)**	0.7 (0.2)	0.03 (0.05)

Seasonal Diru







Season
# of Light-Duty Vehicles
# of Heavy-Duty Vehicles
# of Electric Vehicles
Traffic Volumes
Mobile Sources
Cycle Length
Degree of Saturation
Background Sources
Wind Speed
Temperature
NO
NO_2
Atmospheric Boundary Layer
Mechanical Turbulence
OH, HO_2, peroxy radical
Time of Day
HONO
VOCs
O_3
Sunlight
Early Morning <8am
8am to 12pm
11am to 3pm

Activity: build your own diagram!

- Step 1: pick one of the subject areas
- Step 2: write down, on yellow sticky notes, all the data you think would be relevant. Stick them on the paper
- Step 3: arrange them causally, from bottom (antecedents) to top (descendants)
- Step 4: draw **red arrows** for the causal model; make sure to join the variables if you need to
- Step 5: pick out a few **correlations** that you think you can *feasibly* verify
- Step 6: once you have a causal diagram… wait for further instructions!

Help us do science!

- Do you need help with selecting indicators?
- Contribute to a stable of examples.
- Get published.

Thank you.

https://indicator-frameworks.github.io

or **joshua.tan@cs.ox.ac.uk**