Data Science for Smart Cities

CE88

Prof: Alexei Pozdnukhov

OH: Monday 2-4pm
115 McLaughlin Hall

My background

MSc Mathematical Physics
Lomonosov Moscow State University, Russia

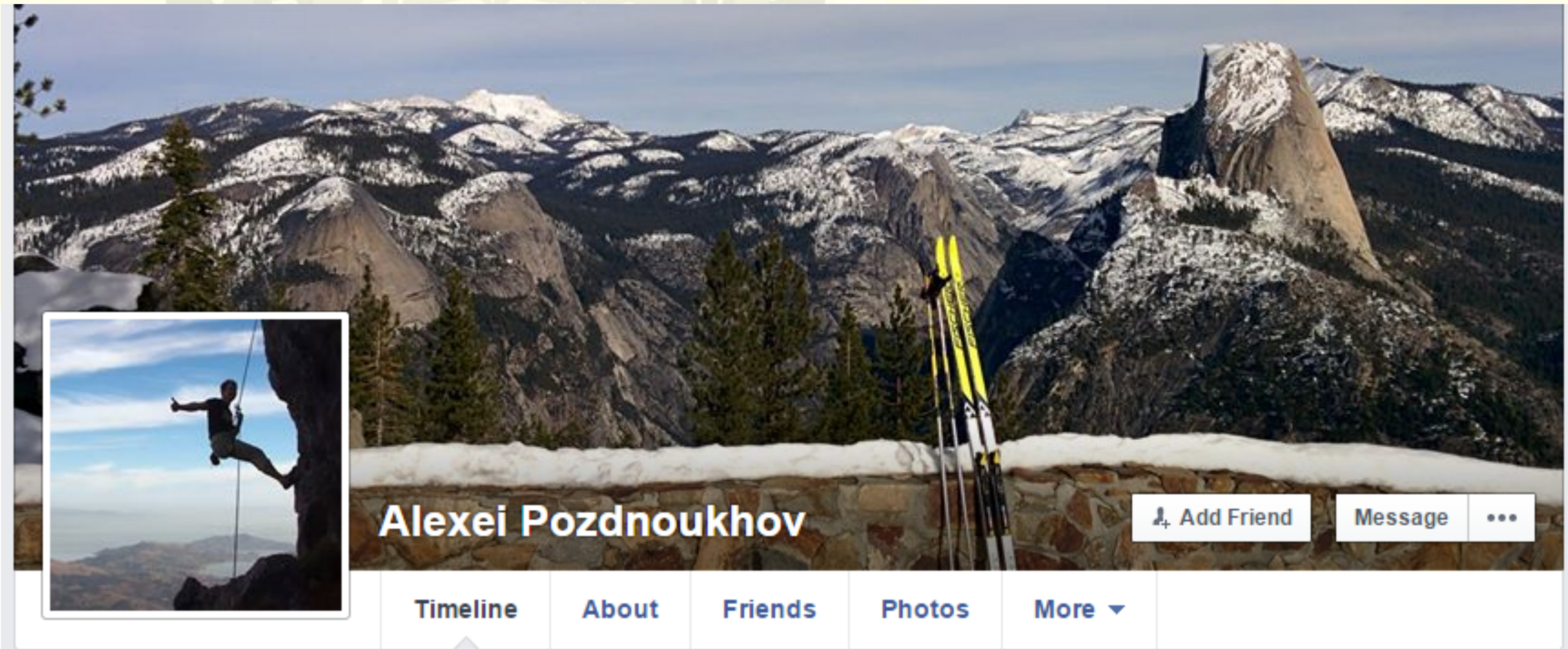
Research Asisstant
IDIAP Research Institute, Switzerland

PhD Computer Science
Ecole Polytechnique Federale de Lausanne, Switzerland

Premier Asisstant
University of Lausanne, Switzerland

Stokes Lecturer
National Centre for Geocomputation, Ireland

More about me



Course logistics

A typical class: lecture + discussion + lab

Q&A: Piazza

Homework: bCourses

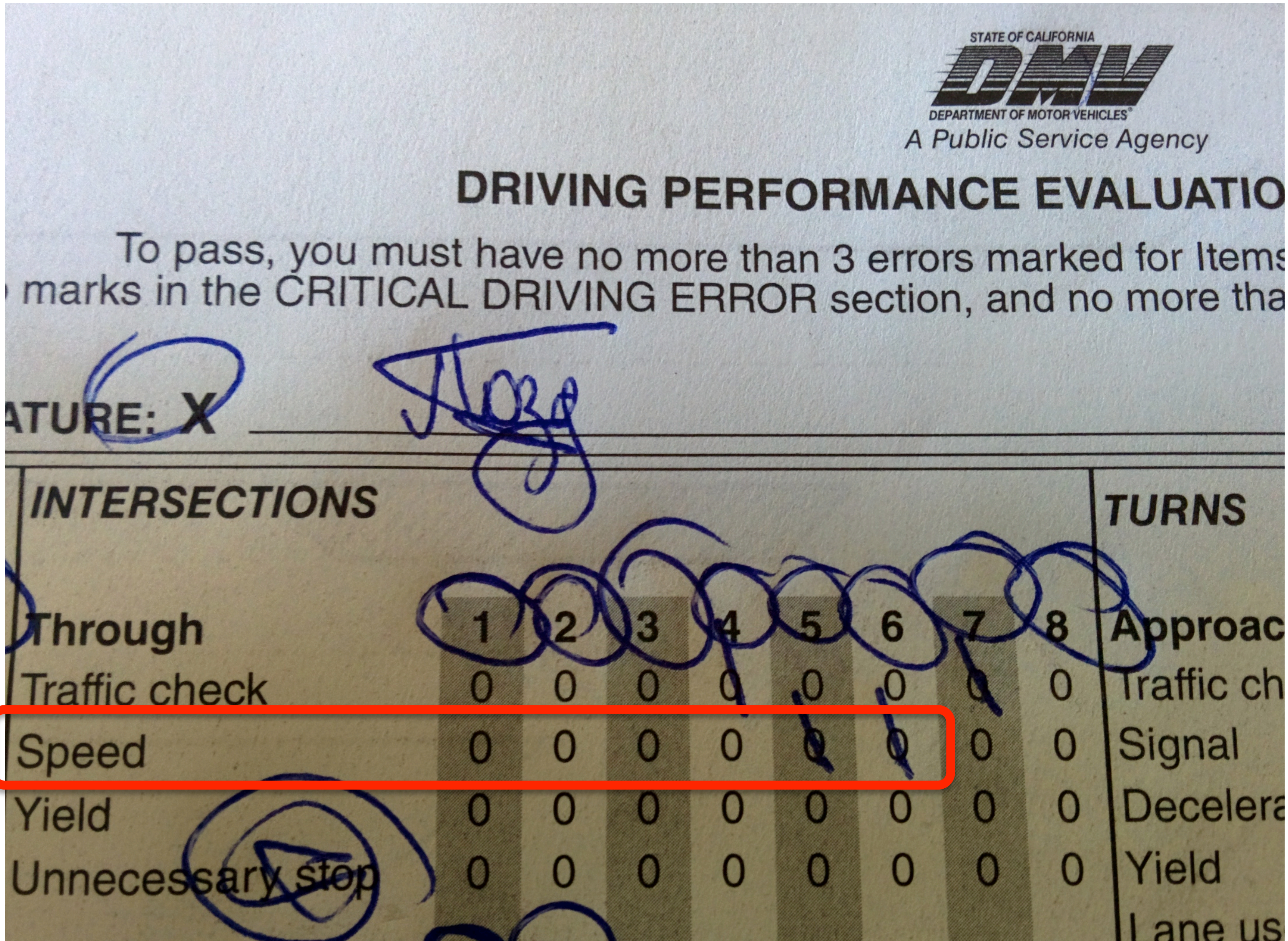
Coding: Python, datahub notebooks

Grading: Homework (~8 problems) 50%

Midterm and/or Final Project 50%

This is a 2 Units course, so required workload is lower – but tasks are open-ended!

Course Pace



STATE OF CALIFORNIA
DMV
DEPARTMENT OF MOTOR VEHICLES
A Public Service Agency
DRIVING PERFORMANCE EVALUATIO
To pass, you must have no more than 3 errors marked for Items
marks in the CRITICAL DRIVING ERROR section, and no more tha
ATURE: X
INTERSECTIONS
TURNS
Through 1 2 3 4 5 6 7 8
Traffic check 0 0 0 0 0 0 0 0
Speed 0 0 0 0 0 0 0 0
Yield 0 0 0 0 0 0 0 0
Unnecessary stop 0 0 0 0 0 0 0 0
Approac
Traffic ch
Signal
Decelera
Yield
Lane us

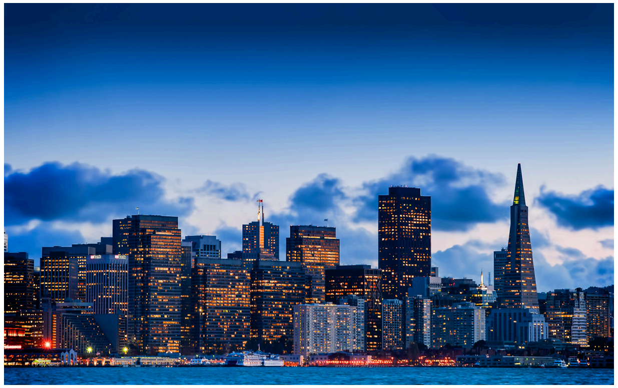
Today's agenda

- Introductions

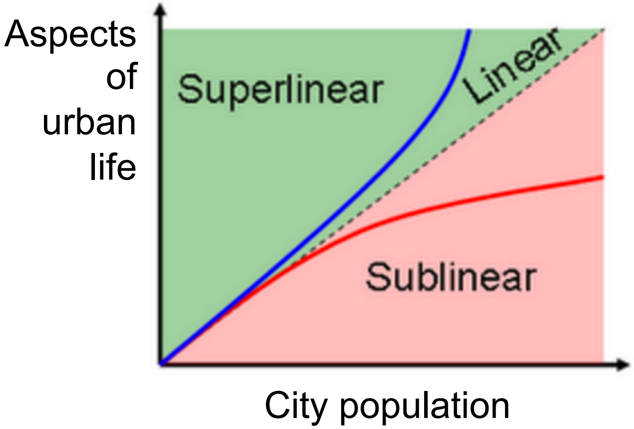
- Cities
- What makes a city smart
- The role of data and data science

- Course overview

Cities



"Cities could persist - as they have for thousands of years - only if their advantages offset the disadvantages"



Scaling of urban life with city population

- Electricity consumption
- Water consumption

} linear

- Length of roads
- Length of electric cables
- Number of gasoline stations

} sub-linear ☺ 'economies of scale'

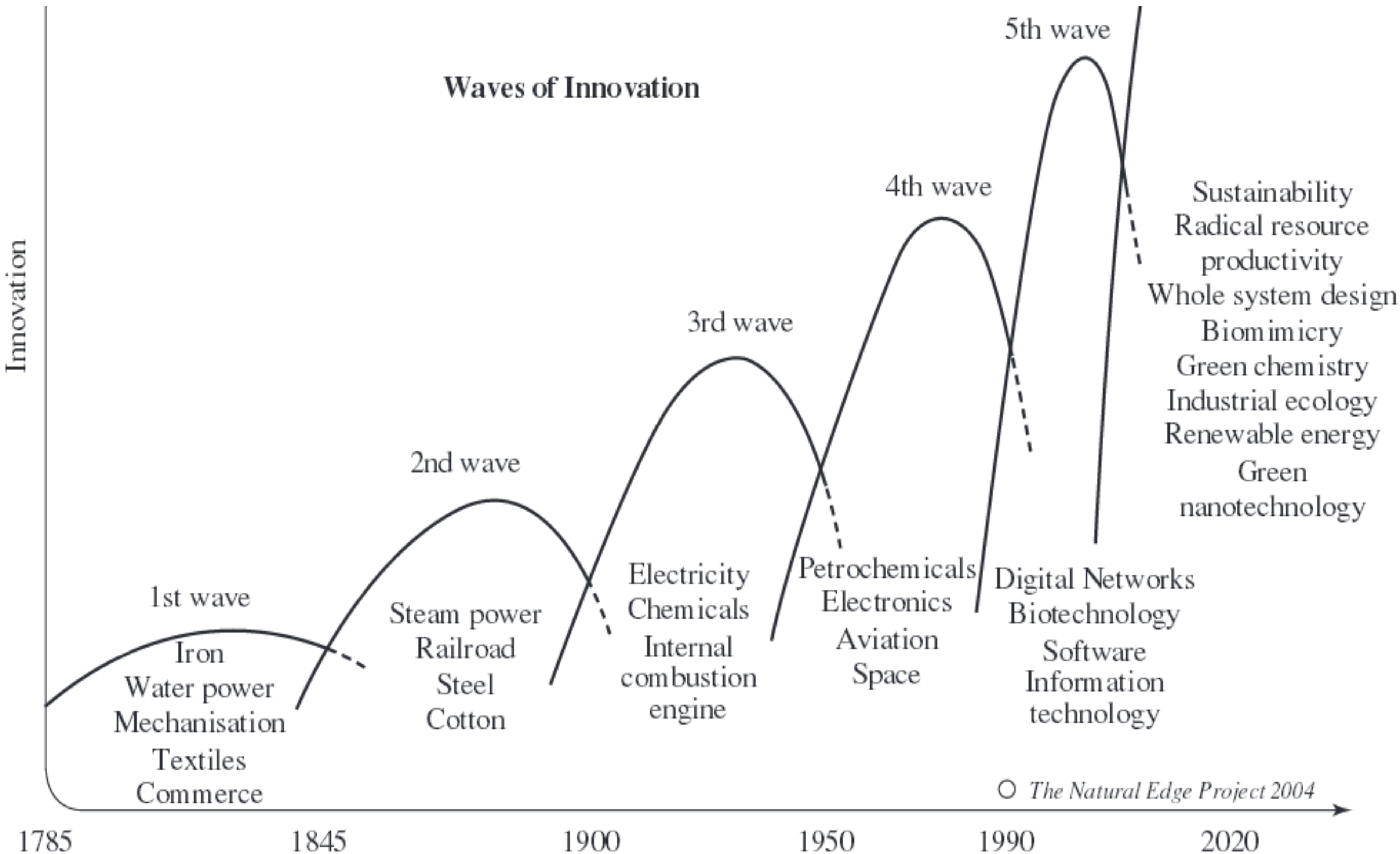
- Total wages
- Inventions/New Patents
- R&D employment
- 'Supercreative' professionals

} super-linear ☺☺☺ 'knowledge production'

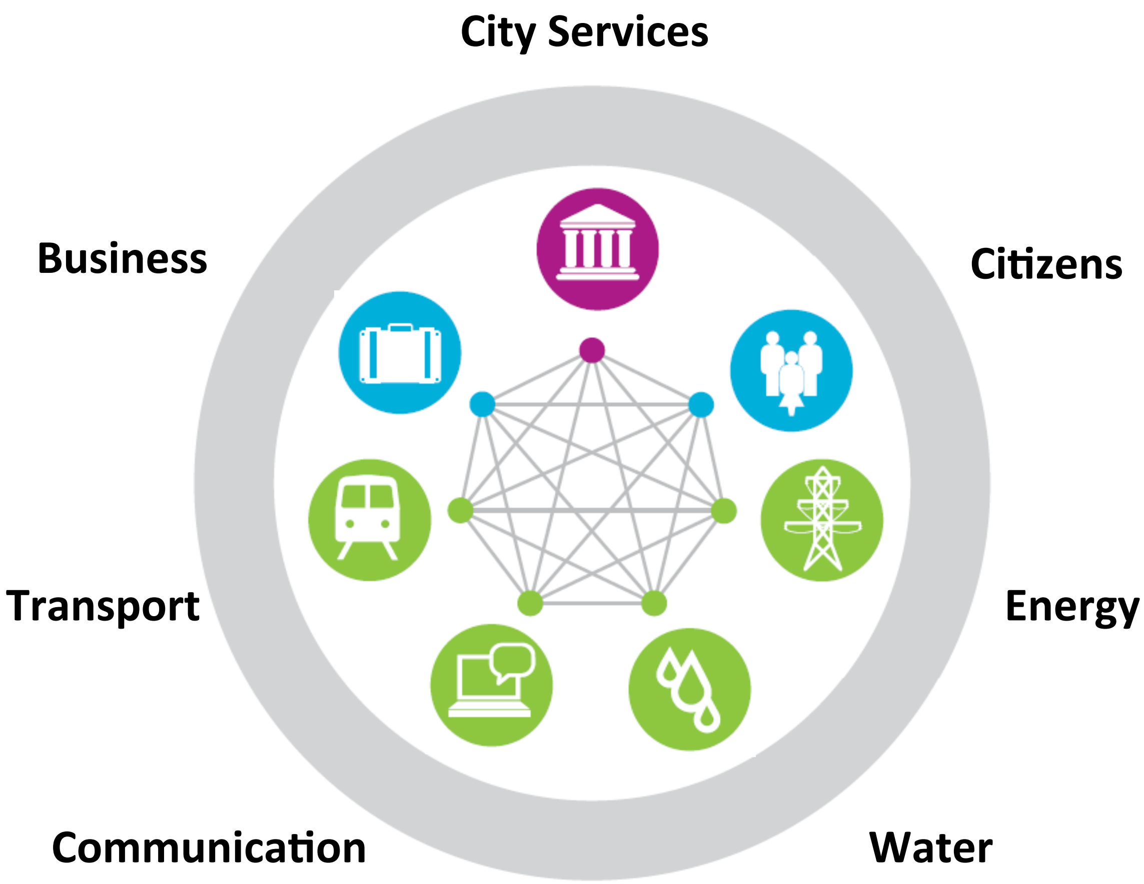
- Crime
- AIDS

} super-linear ☹☹

Cities: incubators of innovation



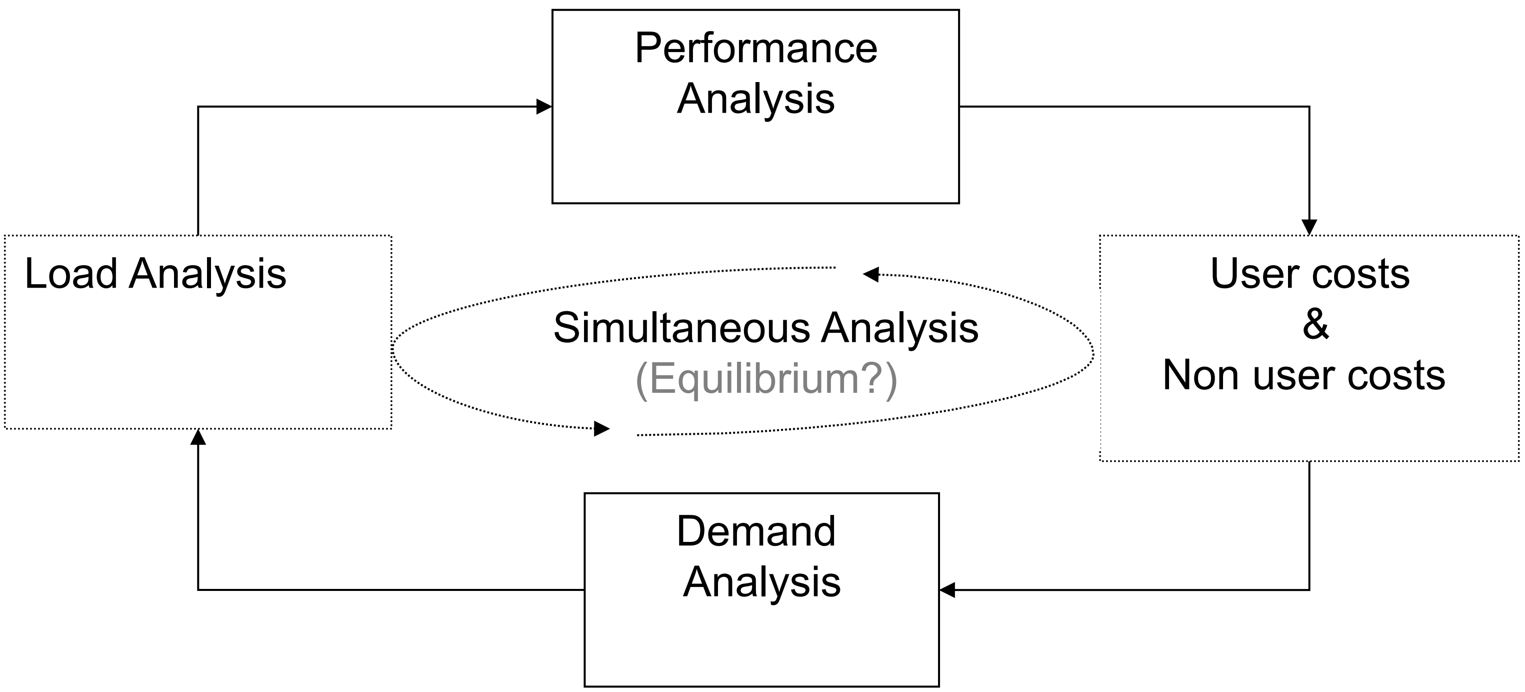
Interconnected infrastructures



(c) IBM

A framework to study infrastructures

Introduction to urban systems: how we will study infrastructures

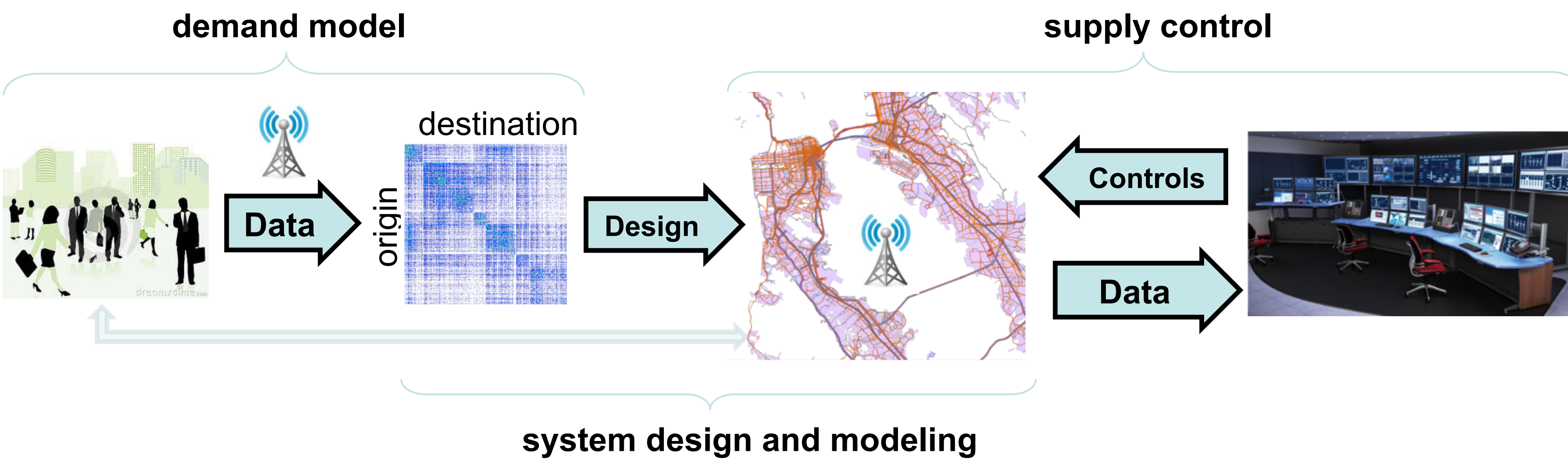


Demand…

Supply…

Performance analysis

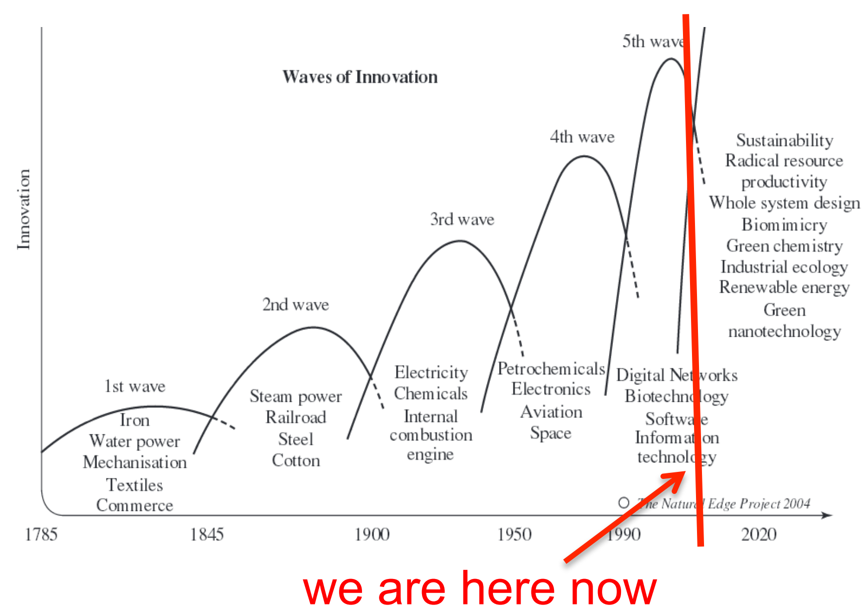
The role of data



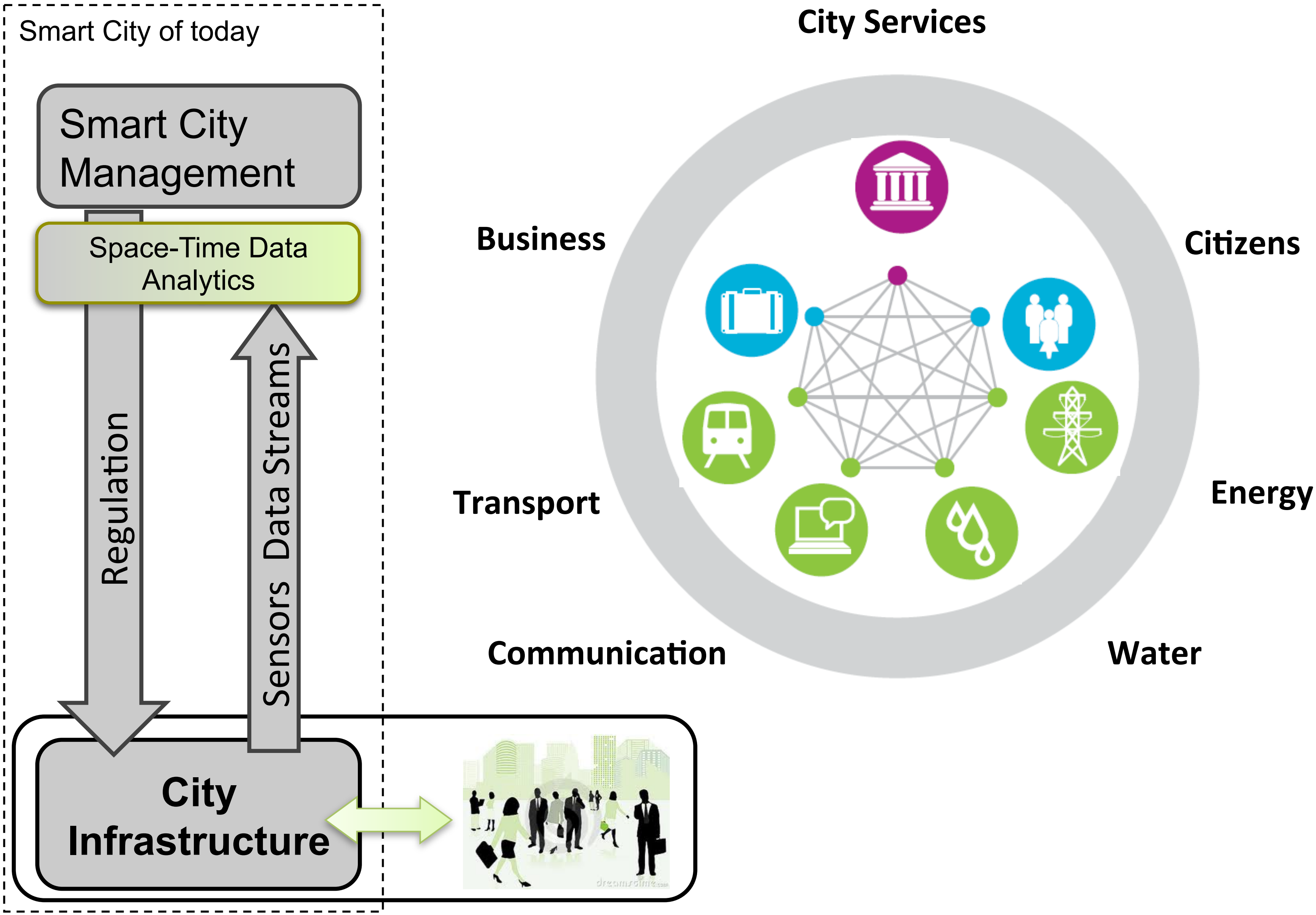
"If you are looking for a career where your services will be in high demand, you should find something where you provide a scarce, complementary service to something that is getting ubiquitous and cheap.

So what's getting ubiquitous and cheap? Data. And what is complementary to data? Analysis."

– Prof. Hal Varian, UC Berkeley, Chief Economist at Google



Smart City of today



Why study systems as inter-dependent / inter-connected ?

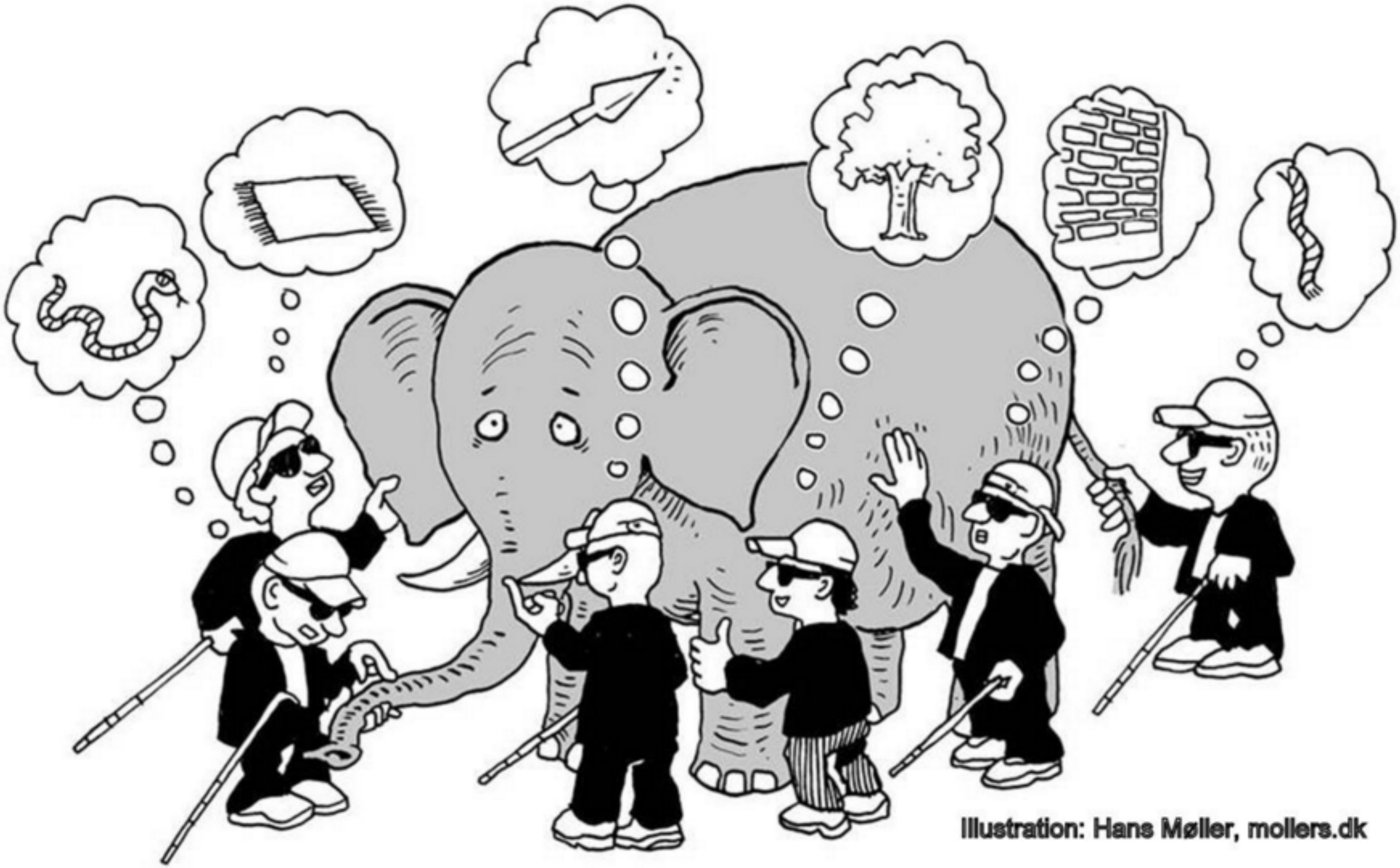
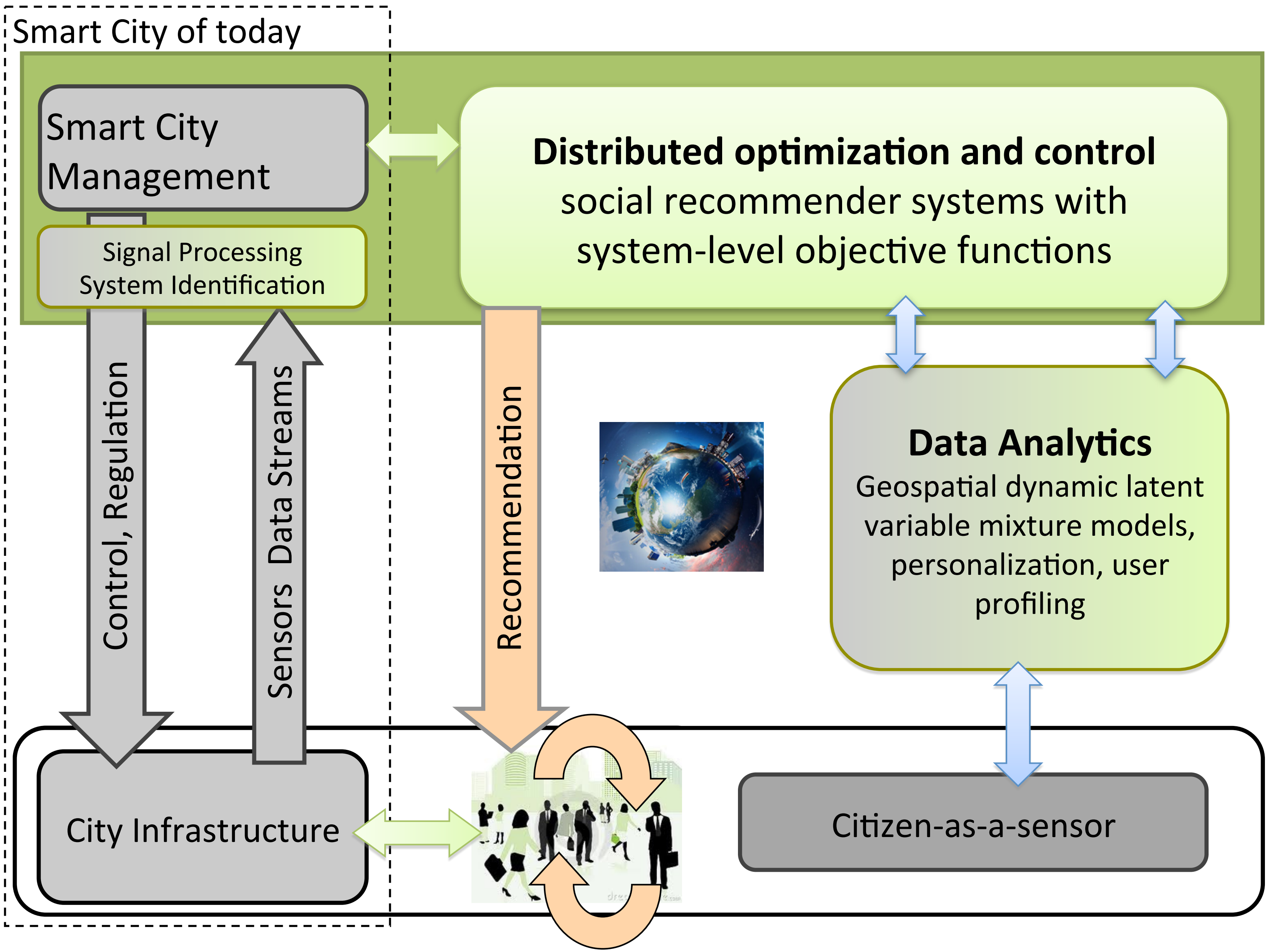
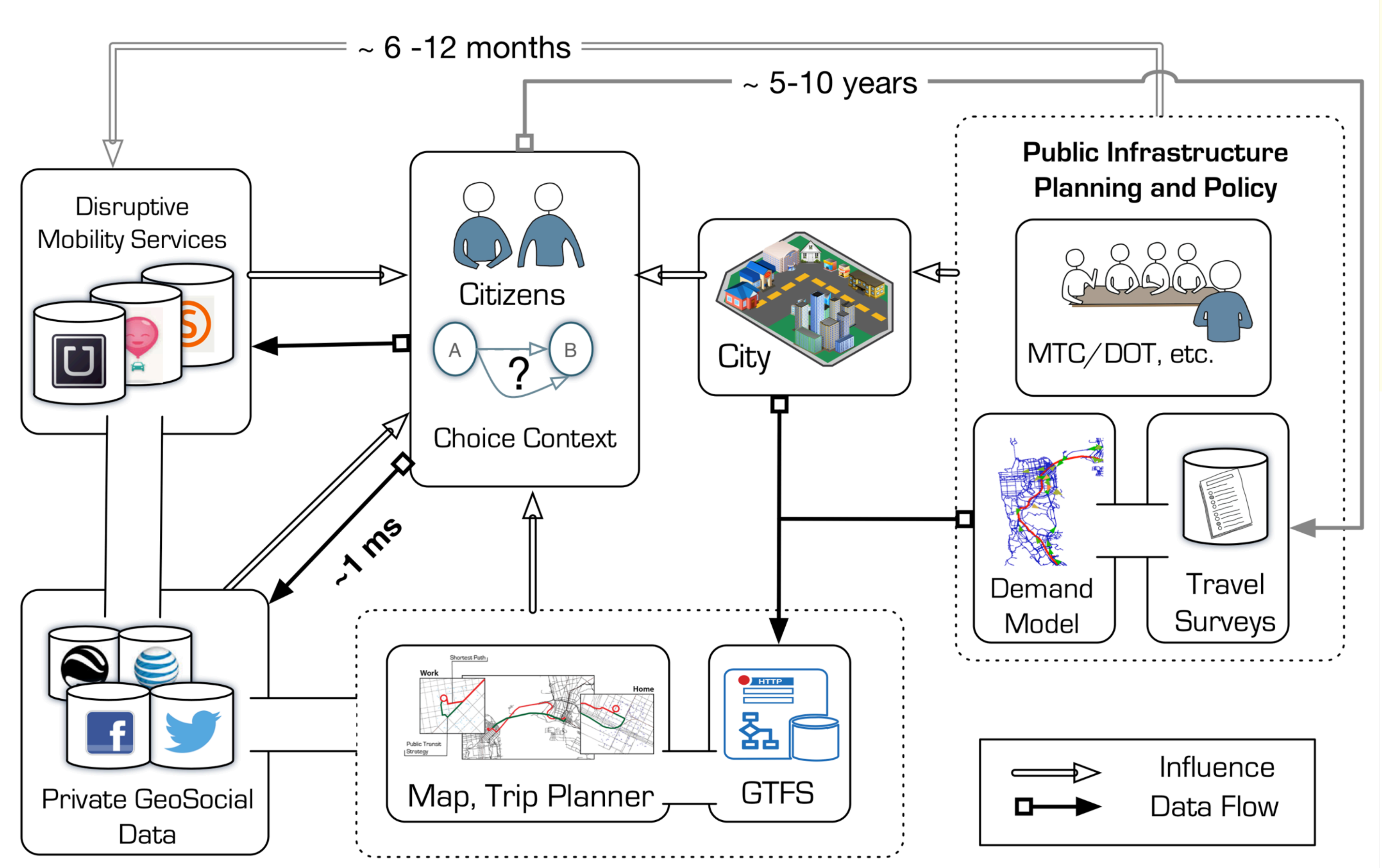


Illustration: Hans Møller, mollers.dk

Smart City vision: Human in the loop



Data ecosystem in cities



Course structure

Introduction and motivation: cities as complex systems.

↓

Urban data collection, handling and processing.

↓

Data exploration and analysis. Demand, supply and impact.

↓

Modeling and forecasting. Regression and classification.

↓

Decision making, planning and governance.

Course contents

Weeks 1-2. Introduction and motivation: cities as complex systems.

Lecture 1. Introduction to urban systems. Inter-dependent infrastructures with human in the loop.
Lecture 2. Modeling principles. Causality and experiments in demand- and supply-side data analysis.

Lecture 3. Spatio-temporal nature of urban data.
Lecture 4. Data flows in cities. Purpose of data analysis: decision making feedback loops.

Introduction and motivation: cities as complex systems.

Lecture 1. Introduction to urban systems. Inter-dependent infrastructures with human in the loop.
Lecture 2. Modeling principles. Causality and experiments in demand- and supply-side data analysis.

On evolution of Science

- **Observational Science**
  – Scientist gathers data by direct observation
  – Scientist describes data

- **Analytical Science**
  – Scientist builds analytical model
  – Makes predictions

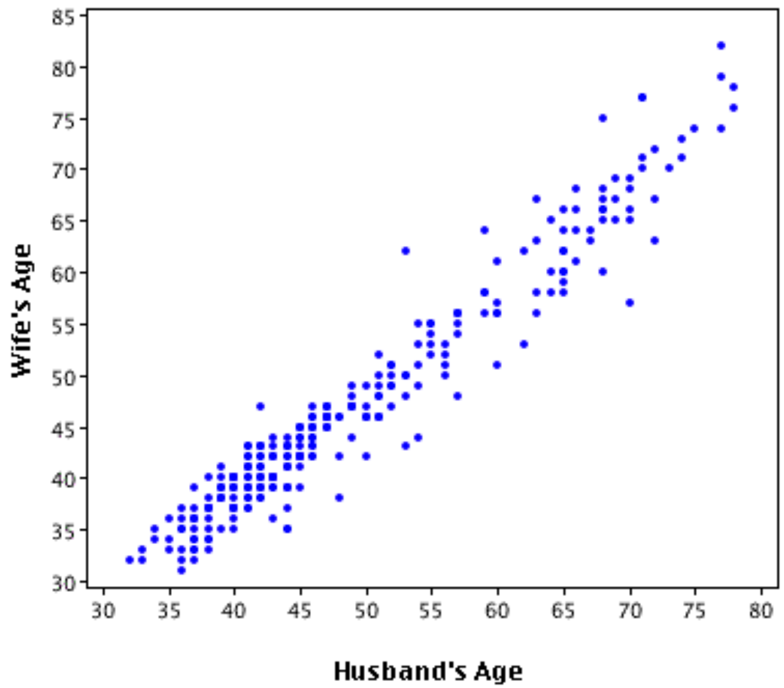
- **Computational Science**
  – Simulate analytical model
  – Validate model and makes predictions

- **Data Exploration Science**
  – Data-driven science: data are measured and processed by PC
  – Scientists analyze databases / files

Observational models

- Scientist gathers data by direct observation

- Scientist describes observed relationships

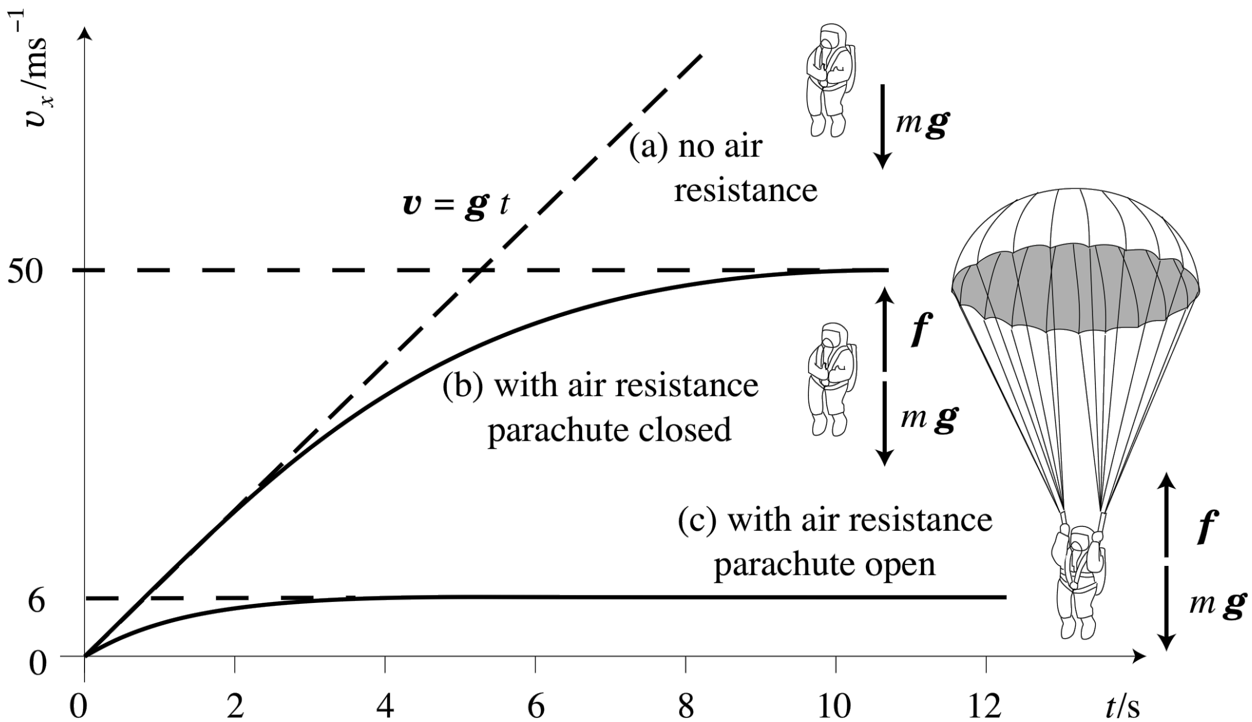


Observational models are very universal and straightforward to apply. While they can hardly describe situations that were not observed previously, they can help formulate hypothesis about the underlying first principles.

Interestngly, they can be applied when the process at hand is complicated to the extent that the 'first principles' are hard or even impossible to define (which is often true for many aspects involving human/ social behaviors – hence most urban phenomena).

Analytical models

- Scientist builds an analytical model from the stated first principles, introducing more simplifying assumptions if necessary
- It is used to make predictions and control system behavior in the range of parameters whithin which the model assumptions hold true

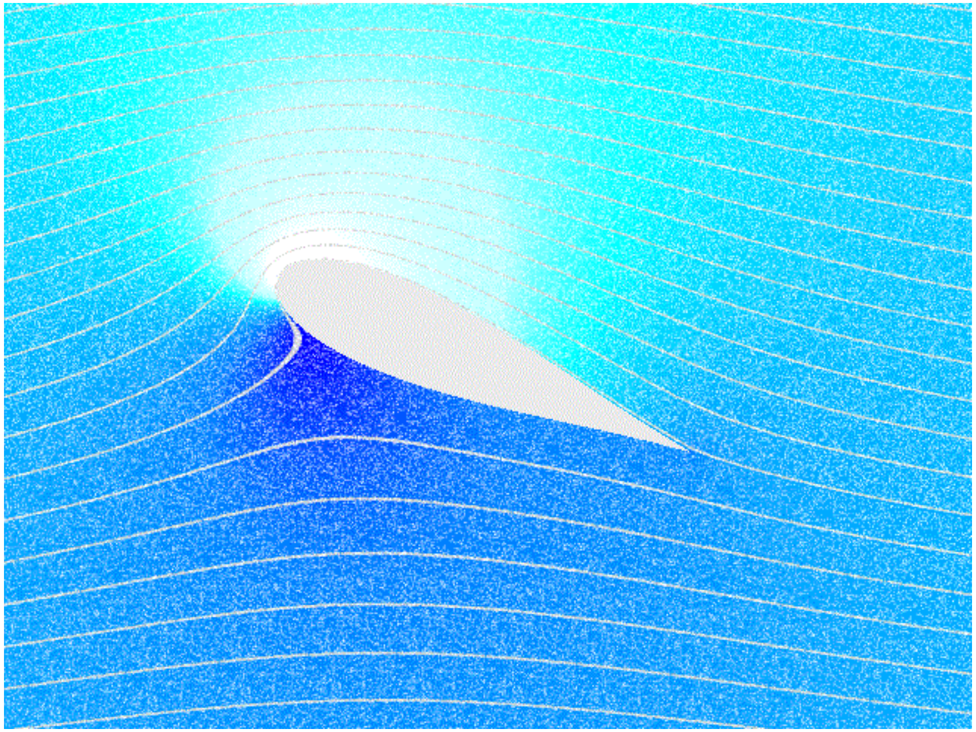


Computational models

- If the system gets complicated, one can simulate its main processes with a computer, using numerical approximations to analytical models
- Computational models can be validated against empirical data, and used to make predictions for unseen situations and engineer better systems

While analytical models of air resistance can be derived for bodies of simple geometries, air flow around complicated bodies can only be derived numerically using computational models. Such models can be used to build supersonic airplanes!

Such models are mainly used in engineering, where laws of physics provide solid first principles for analytical models.



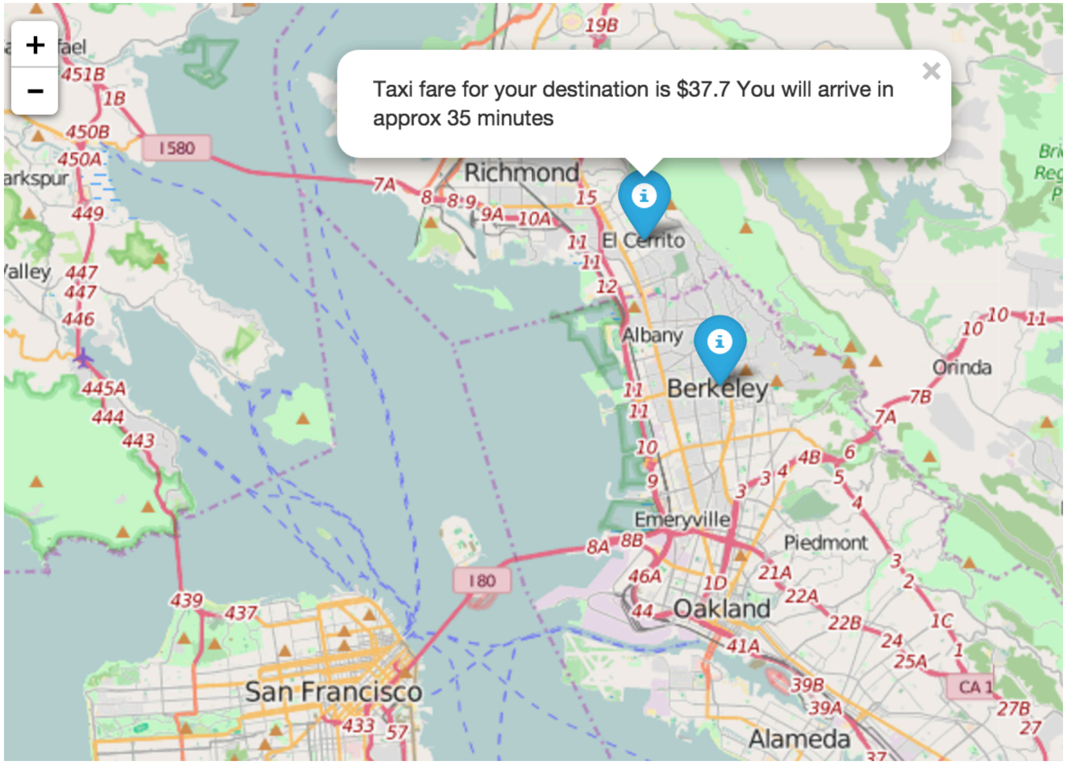
Data-driven models

How does this landscape of modeling approaches change with the deluge of data, especially in cases where first principles are hard or impossible to apply?

Option 1. Develop an **algorithmic approach** to building observational models, i.e. build a computer system able to capture the essence of dependencies from empirical data in a way that can be used for making predictions, as well as helpful for studying the cause-effect relationships in the observed phenomena.

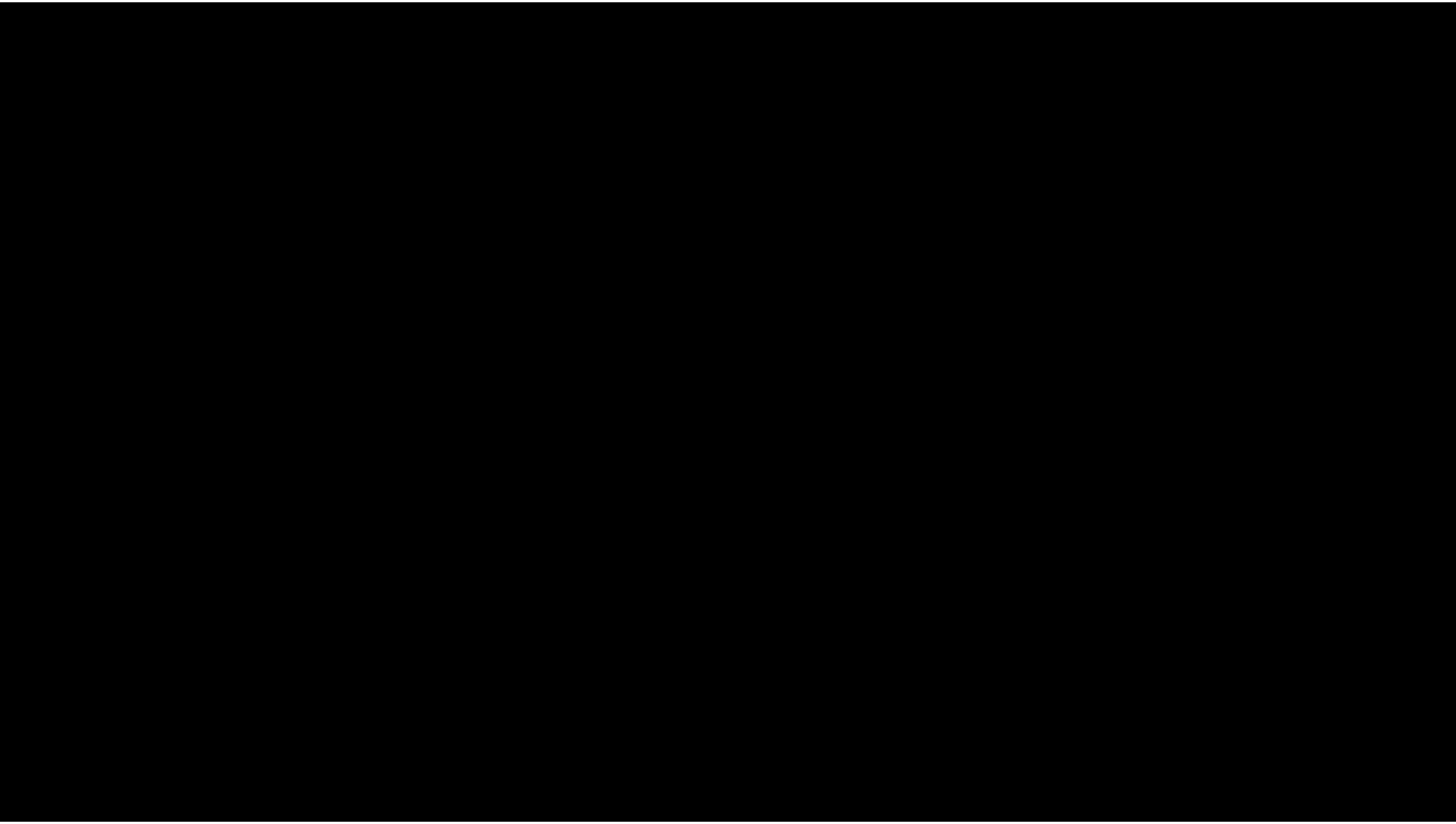
Examples

- Predict travel time and fare of a taxi trip in a city
- Predict electricity consumption profile of a planned office building
- Predict the queue length in the Qualcomm café in Sutardja Dai Hall at a given time and day.



Data-driven models

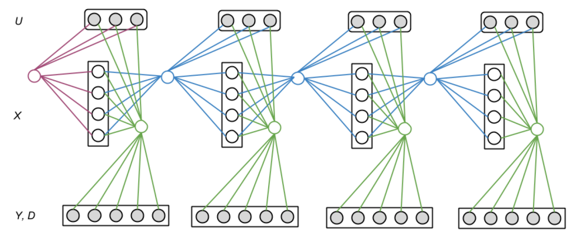
Option 2. Inform a computer micro-simulation with rich data. Video: https://goo.gl/JULmBl



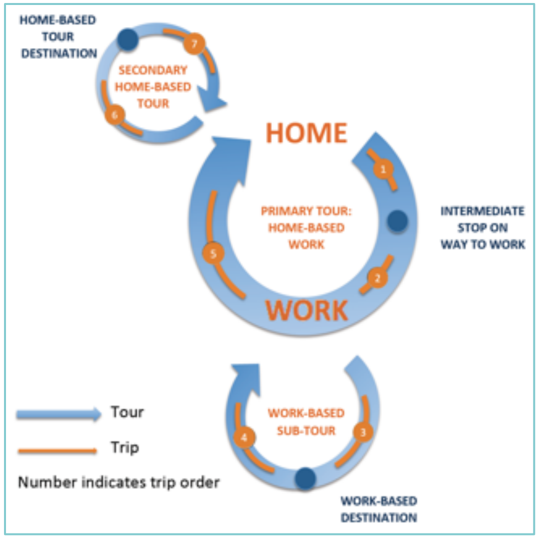
Example: the SmartBay project

•

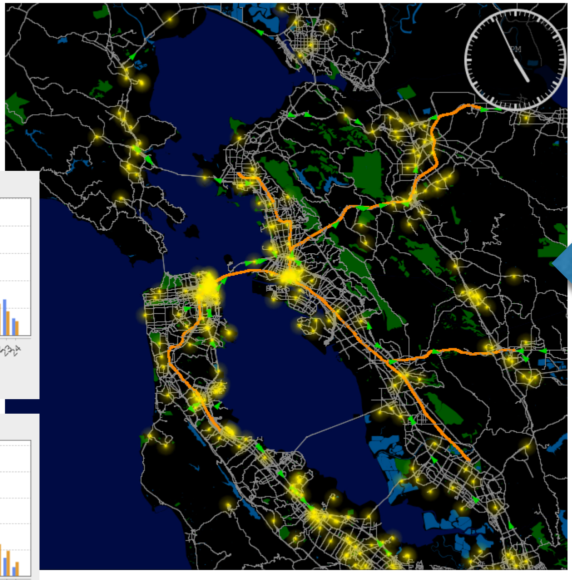
Micro-simulation of SF Bay Area mobility from cell phone data.

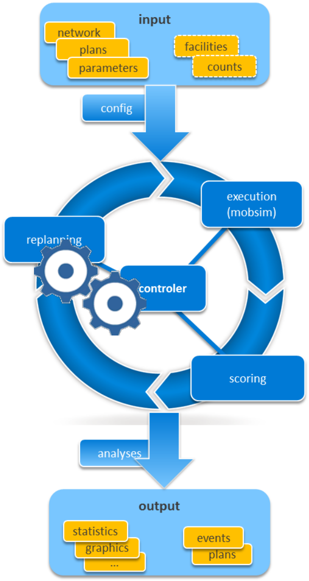


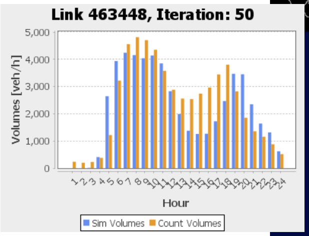
NN-HMM Architecture

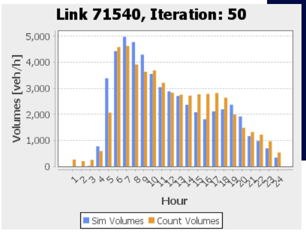


1M agents, 500 shown in visualization

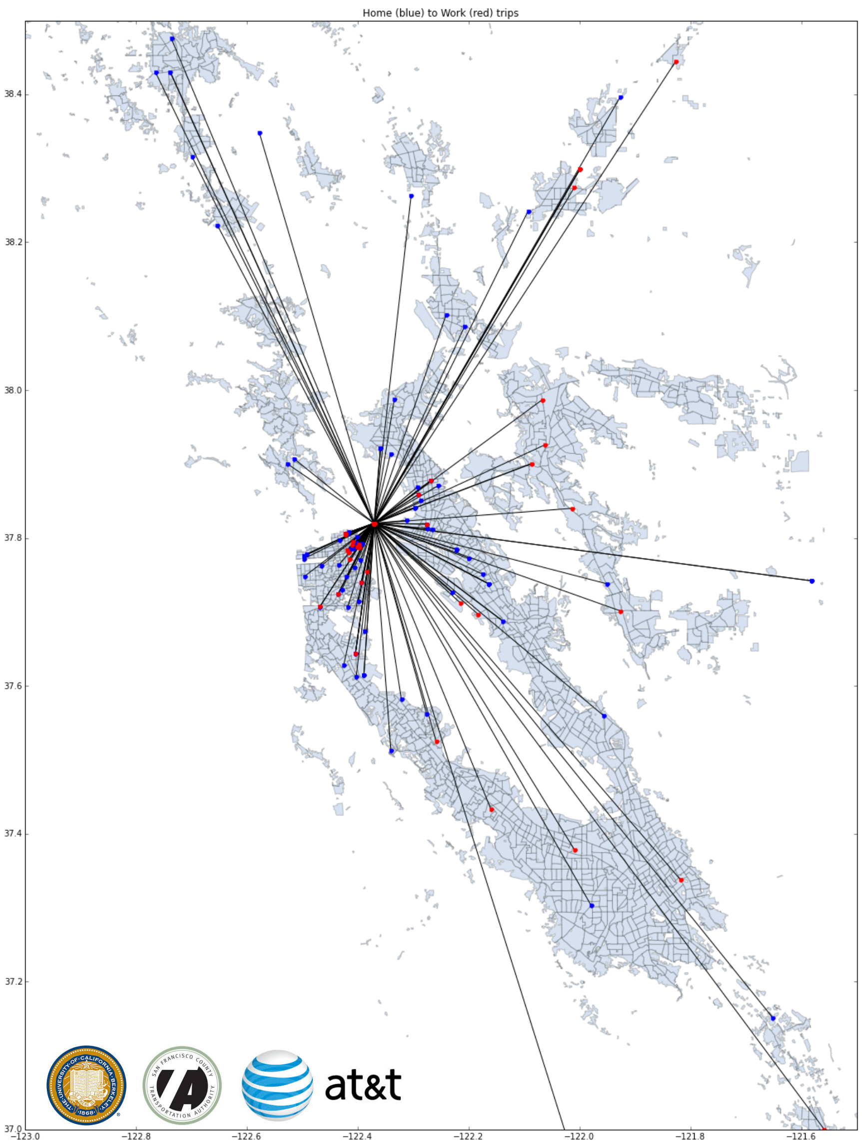








Discussion point: Treasure Island development



SAN FRANCISCO COUNTY TRANSPORTATION AUTHORITY

Fact Sheet

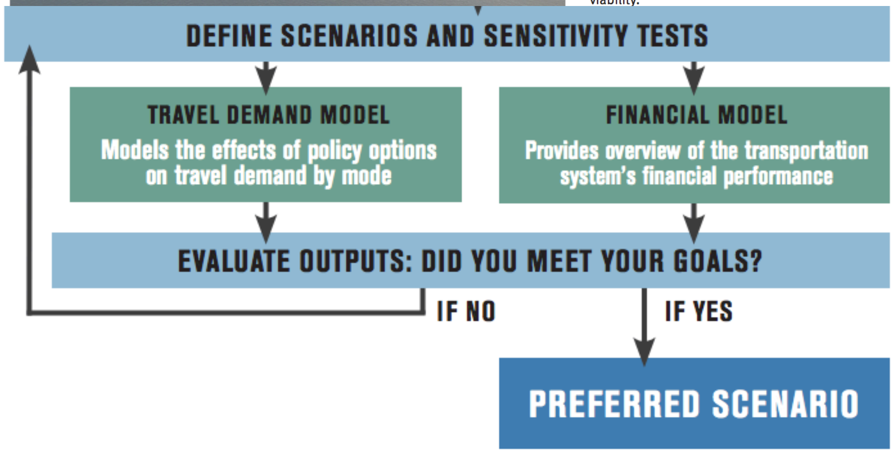
LAST UPDATED July 2014

CONGESTION PRICING TO ENABLE NEW NEIGHBORHOOD DEVELOPMENT

Treasure Island Mobility Management

Current Activities:
Pricing Program Policy Analysis

The Treasure Island Mobility Management Study, currently underway, will analyze and recommend pricing program policies, and establish financial viability.

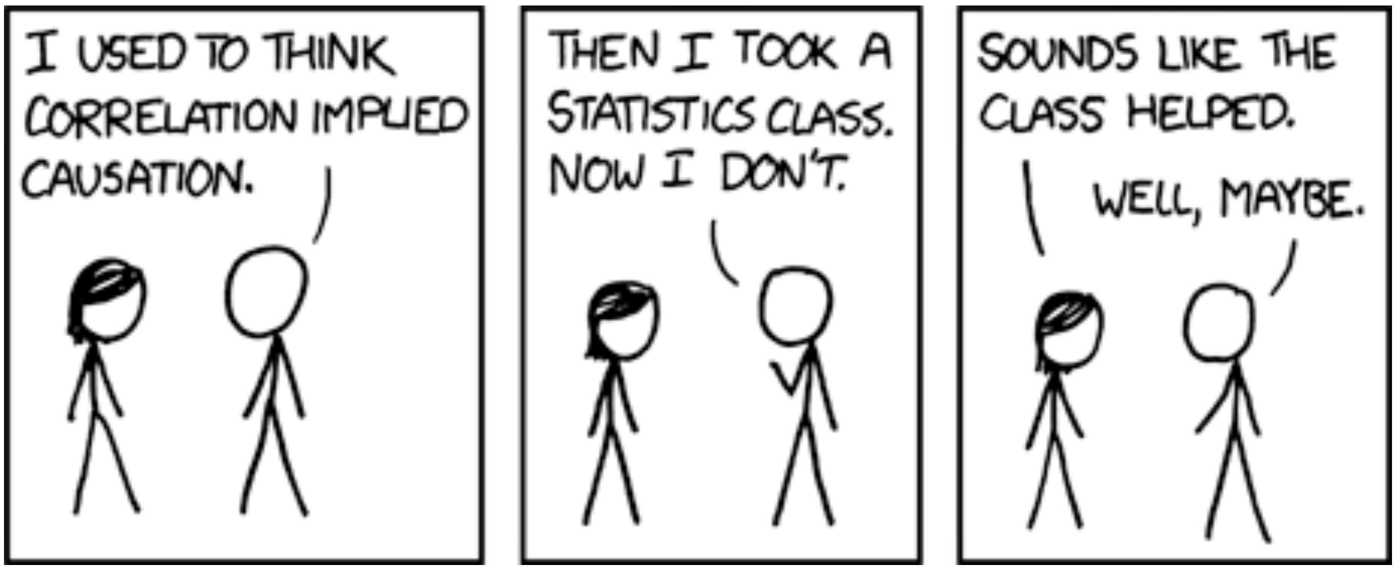


What kinds of models would one need to support planning of a new urban development? What data are needed?

Course topic

Introduction and motivation: cities as complex systems.

Lecture 1. Introduction to urban systems. Inter-dependent infrastructures with human in the loop.
Lecture 2. Modeling principles. **Causality and experiments in demand- and supply-side data analysis.**



Course contents

https://github.com/alexeipberkeleyedu/dssc2018

Thank you!