Data Science for Smart Cities

CE88

Prof: Alexei Pozdnukhov
GSI: Madeleine Sheehan

115 McLaughlin Hall

alexeip@berkeley.edu
m.sheehan@berkeley.edu
CE88 in title

My background

MSc Mathematical Physics
Lomonosov Moscow State University, Russia

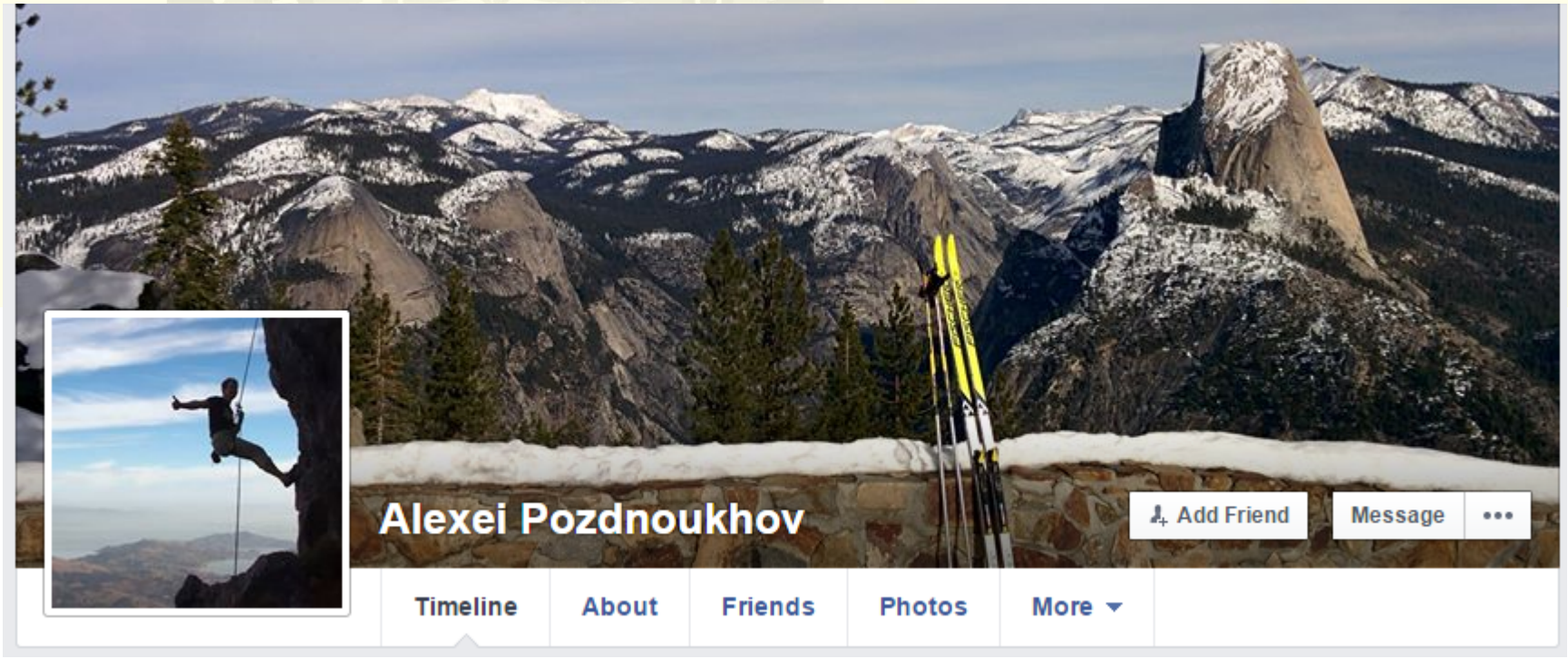
Research Asisstant
IDIAP Research Institute, Switzerland

PhD Computer Science
Ecole Polytechnique Federale de Lausanne, Switzerland

Premier Asisstant
University of Lausanne, Switzerland

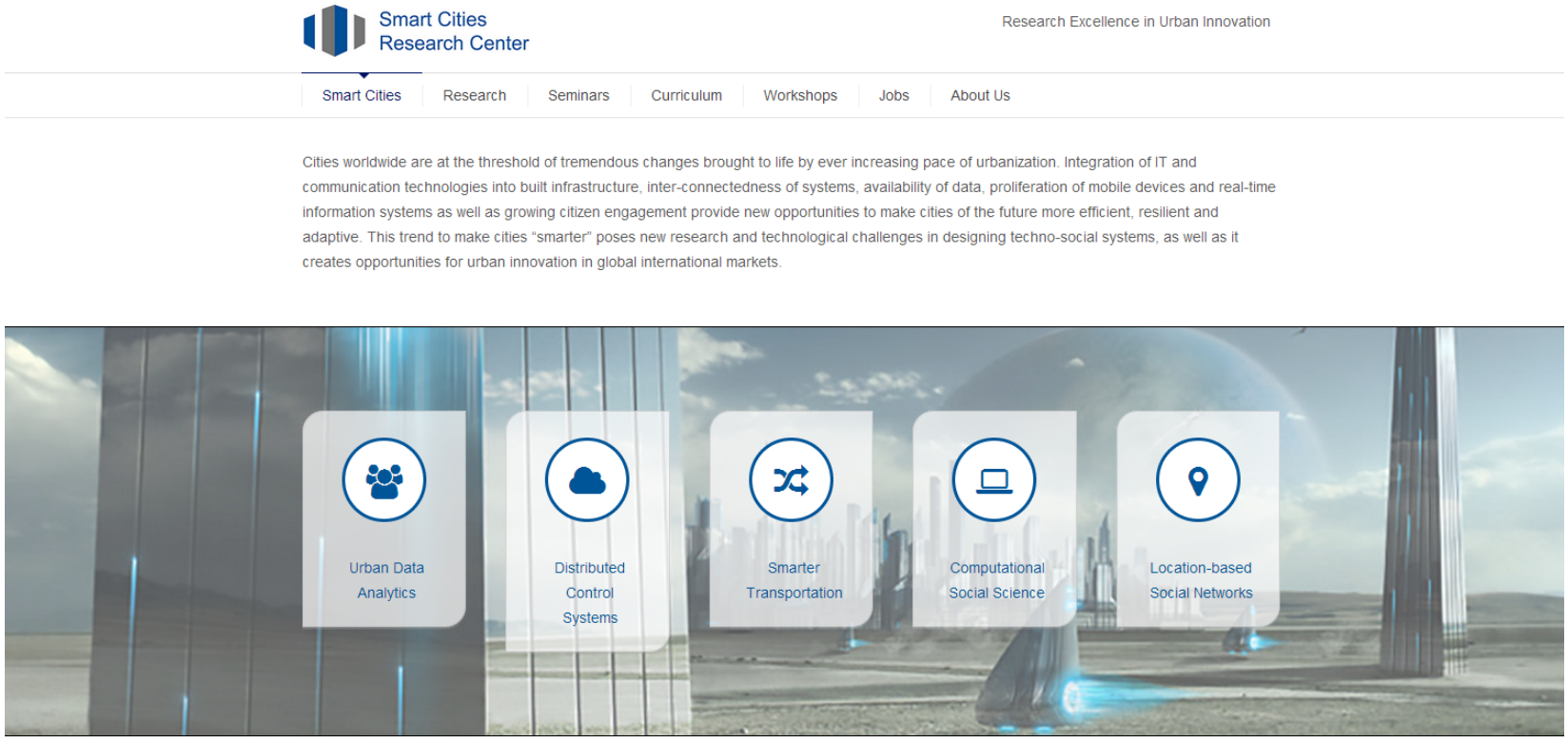
Stokes Lecturer
National Centre for Geocomputation, Ireland

More about me



So, if I am not answering your email during a weekend, I am most likely…

~~…off the grid, climbing or skiing in the Sierras.~~



…working on our research projects!

http://smartcities.berkeley.edu/smartbay/

http://vcresearch.berkeley.edu/signatures/2015-2016-fellows

Maddie Sheehan – GSI

Email: m.sheehan@berkeley.edu
Office Hours: TBD – will send out a Piazza poll!

About me:

- I am a PhD student in Transportation Engineering
- I am working with the Smart Cities Research Group
- I grew up in Maine ☺
- I studied EE at Brown University
- In my free time I like to play ultimate frisbee, ski, explore new areas, and hang with friends

Course logistics

A typical class: lecture + seminar + lab

Q&A: Piazza

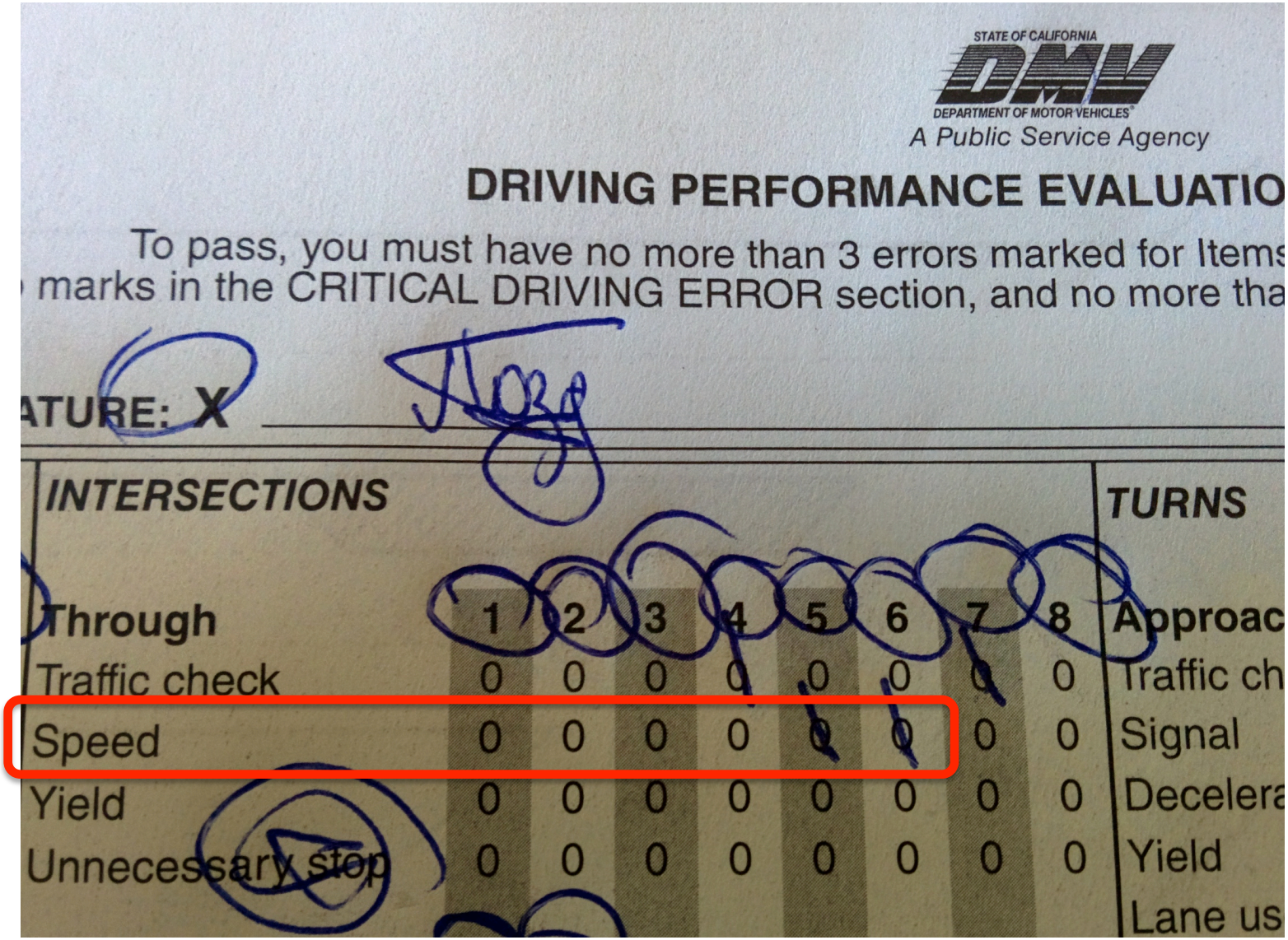
Homework: bCourses

Coding: Python

Grading:
Homework (~9 problems) 50%
Midterm (take home) 20%
Final Project 30%

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	Approac
Traffic check	0	0	0	0	0	0	0	0	Traffic ch
Speed	0	0	0	0	0	0	0	0	Signal
Yield	0	0	0	0	0	0	0	0	Decelera
Unnecessary stop	0	0	0	0	0	0	0	0	Yield
									Lane us

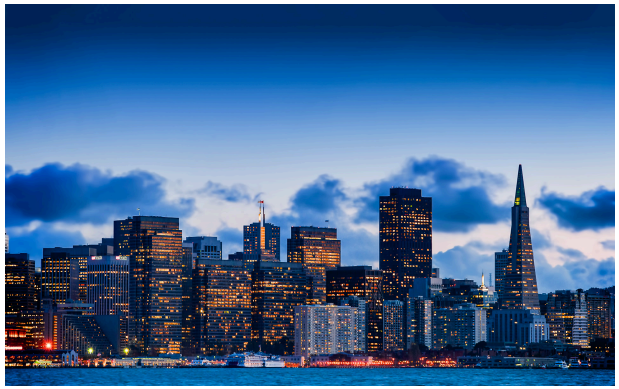
Today's agenda

- Introductions

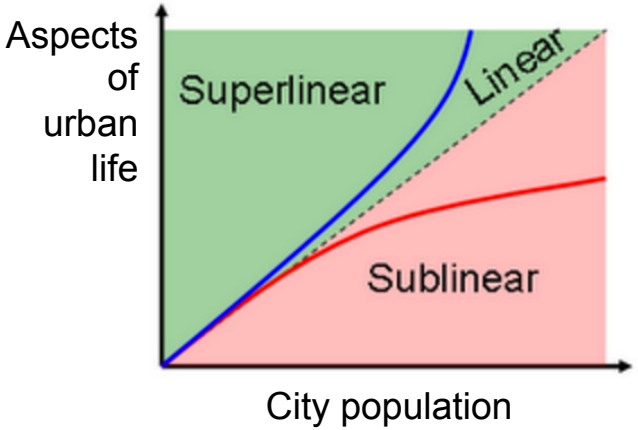
- Cities
- What makes a city smart
- The role of data and data science

- 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'

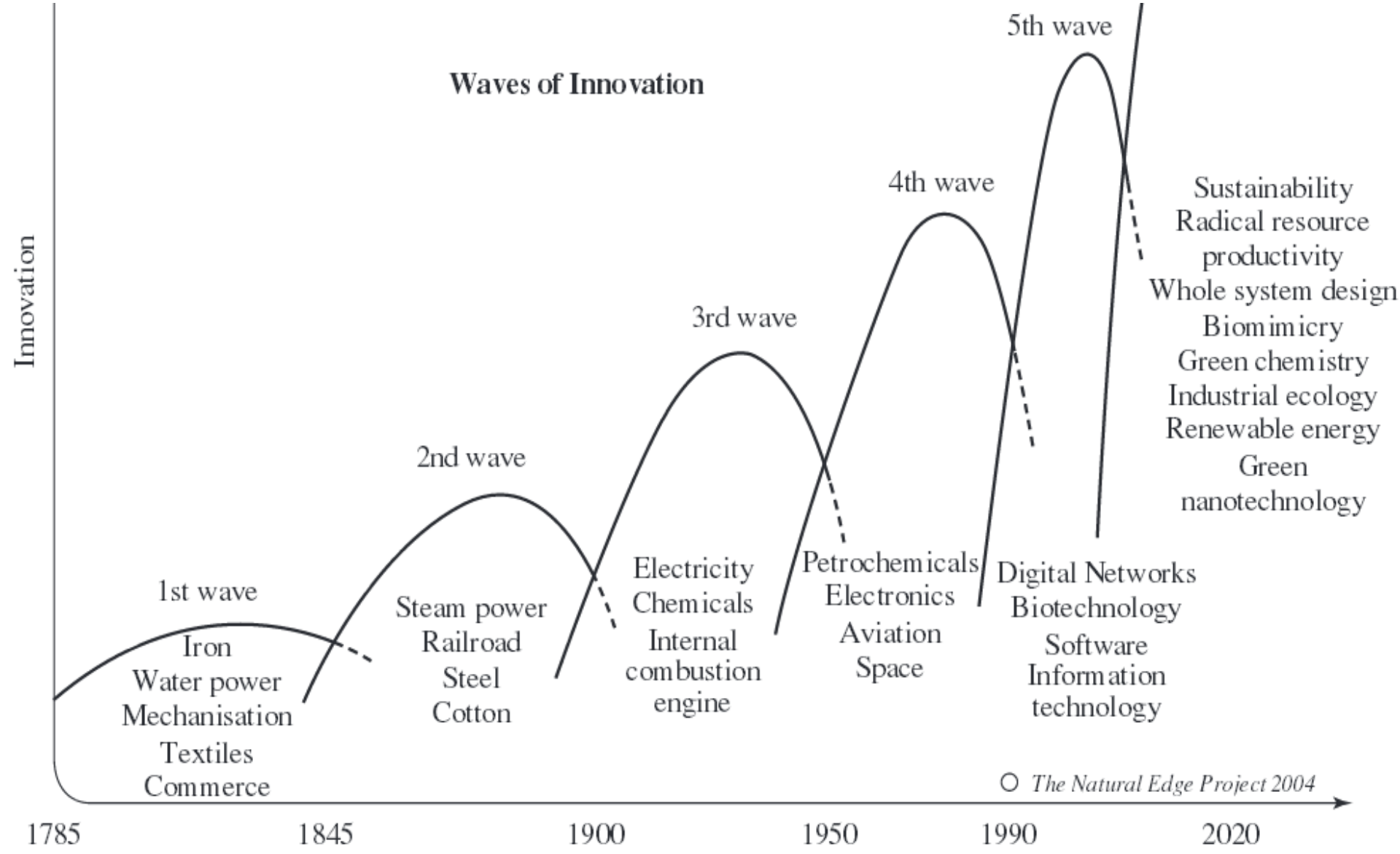
- Total wages
- Inventions/New Patents
- R&D employment
- 'Supercreative' professionals

} super-linear ☺☺☺ 'knowledge production'

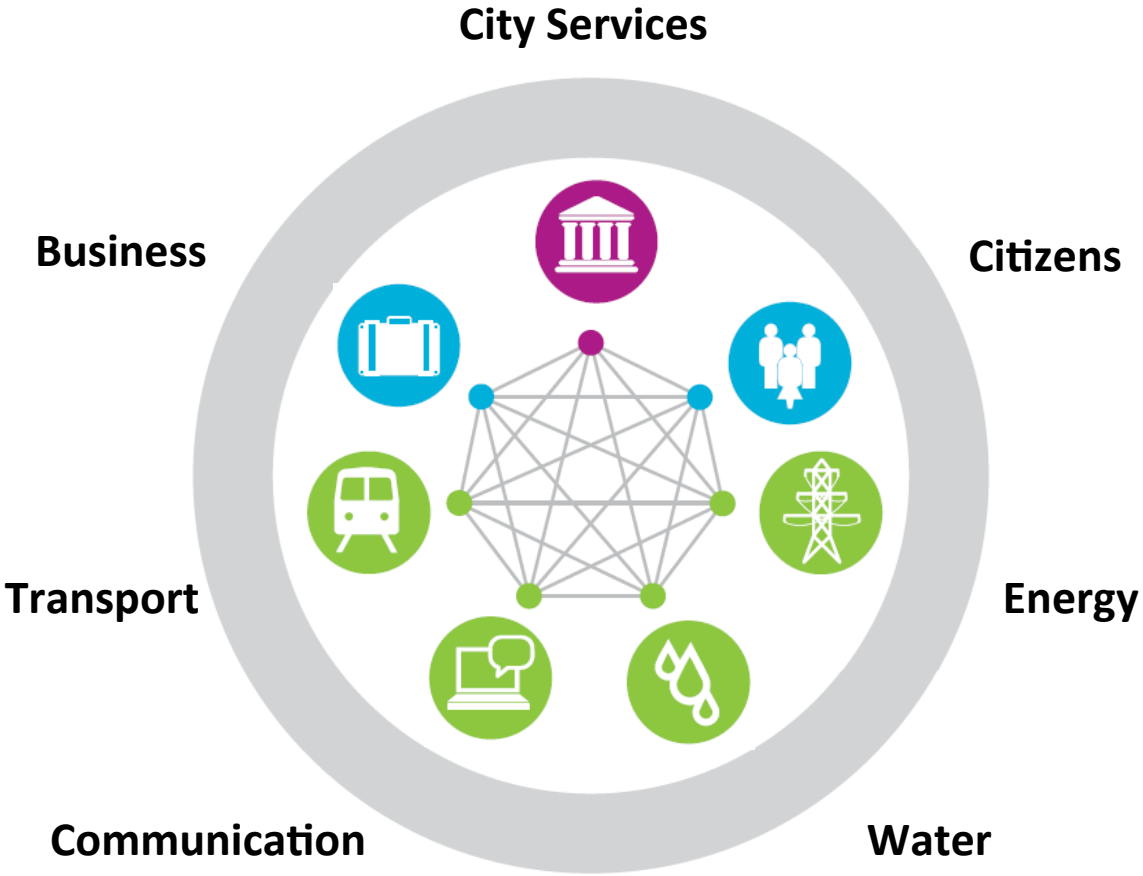
- Crime
- AIDS

} super-linear ☹☹

Cities: incubators of innovation

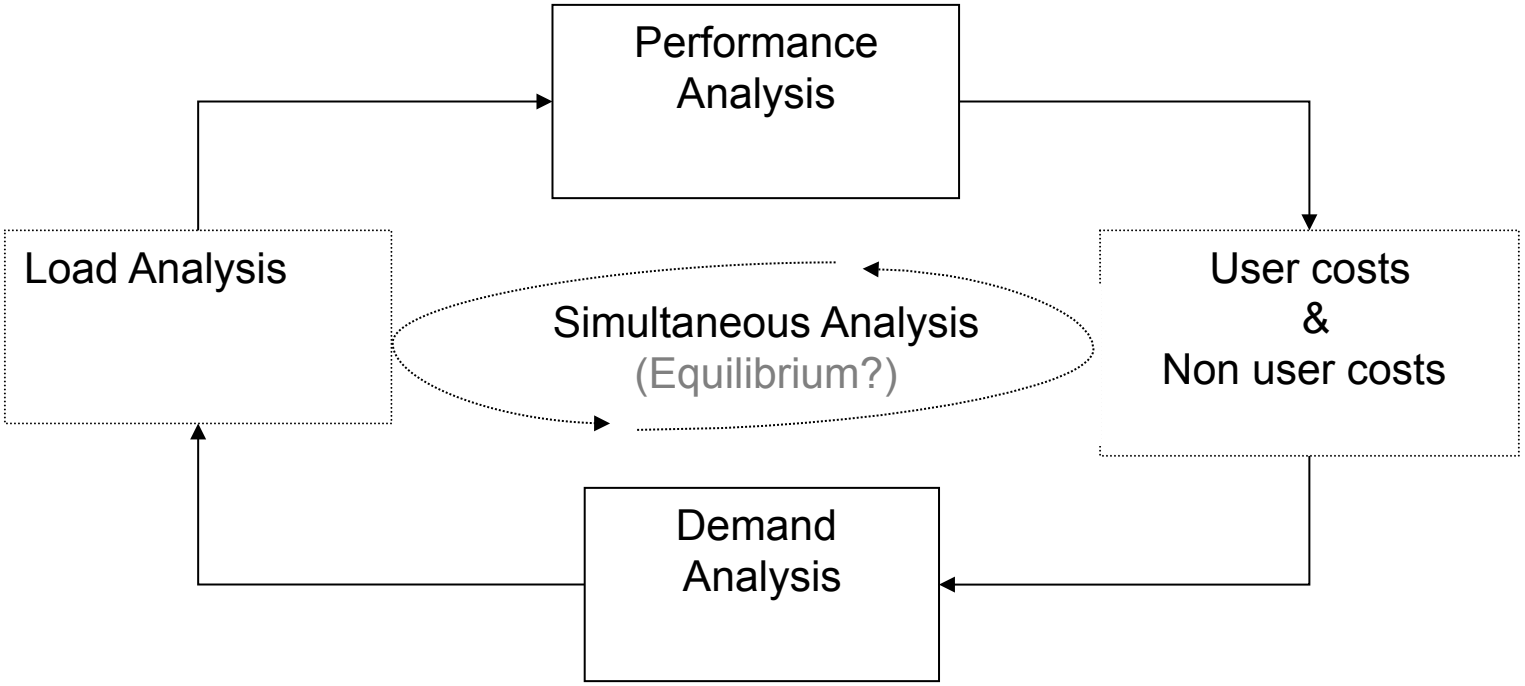


Interconnected infrastructures



A framework to study infrastructures

Introduction to urban systems: how we will study infrastructures



Demand…

Supply…

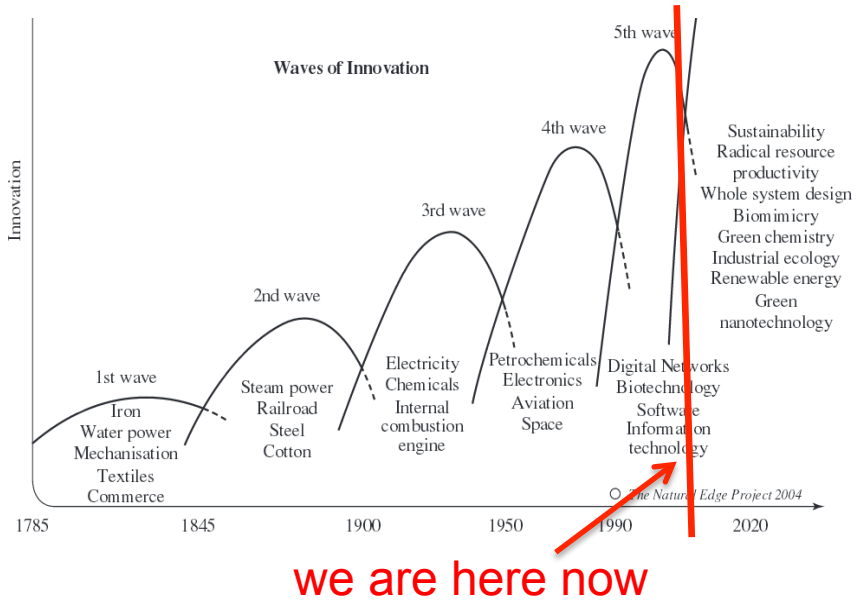
Performance analysis

The role of data

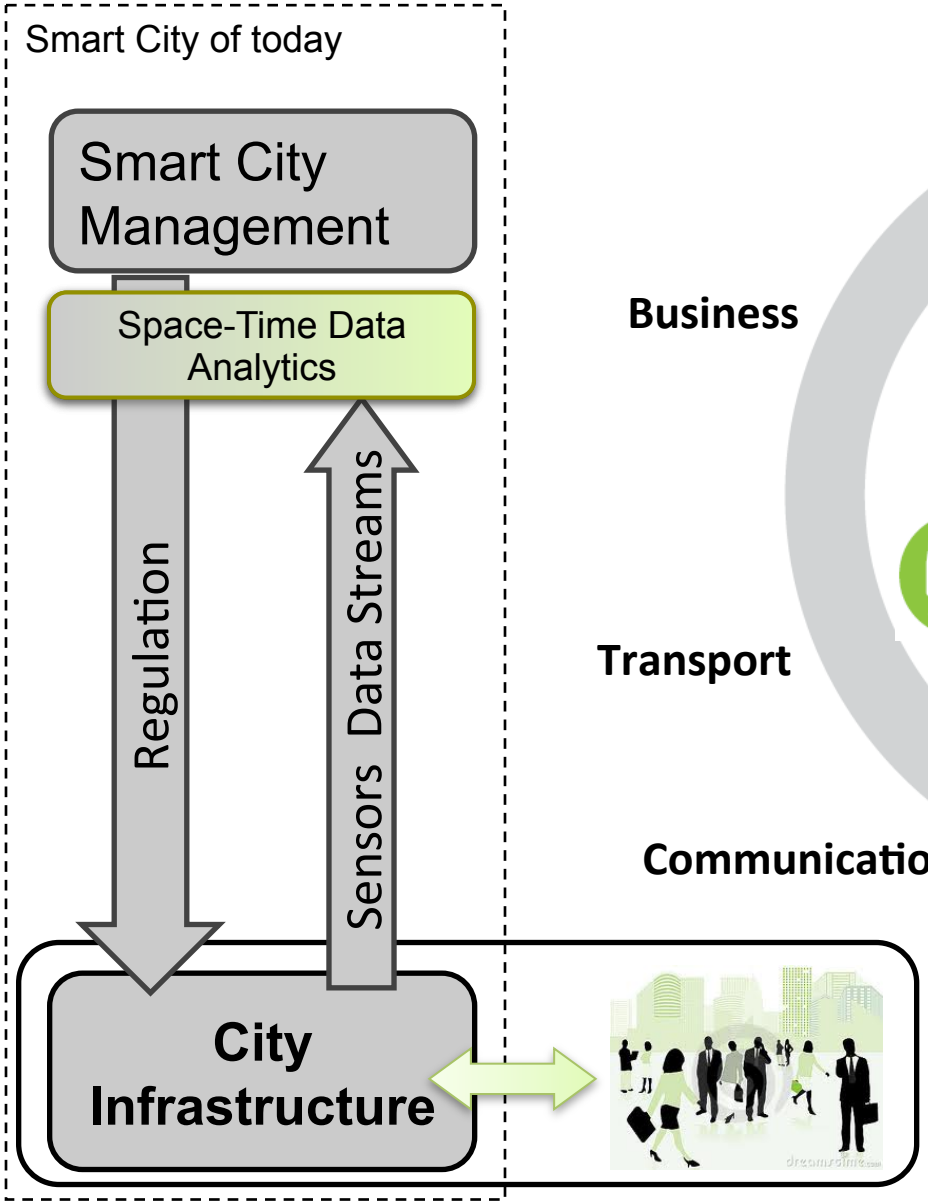
"If you are looking for a career where your services will be in high demand, you should find something where you provide a scarce, complementary service to something that is getting ubiquitous and cheap.

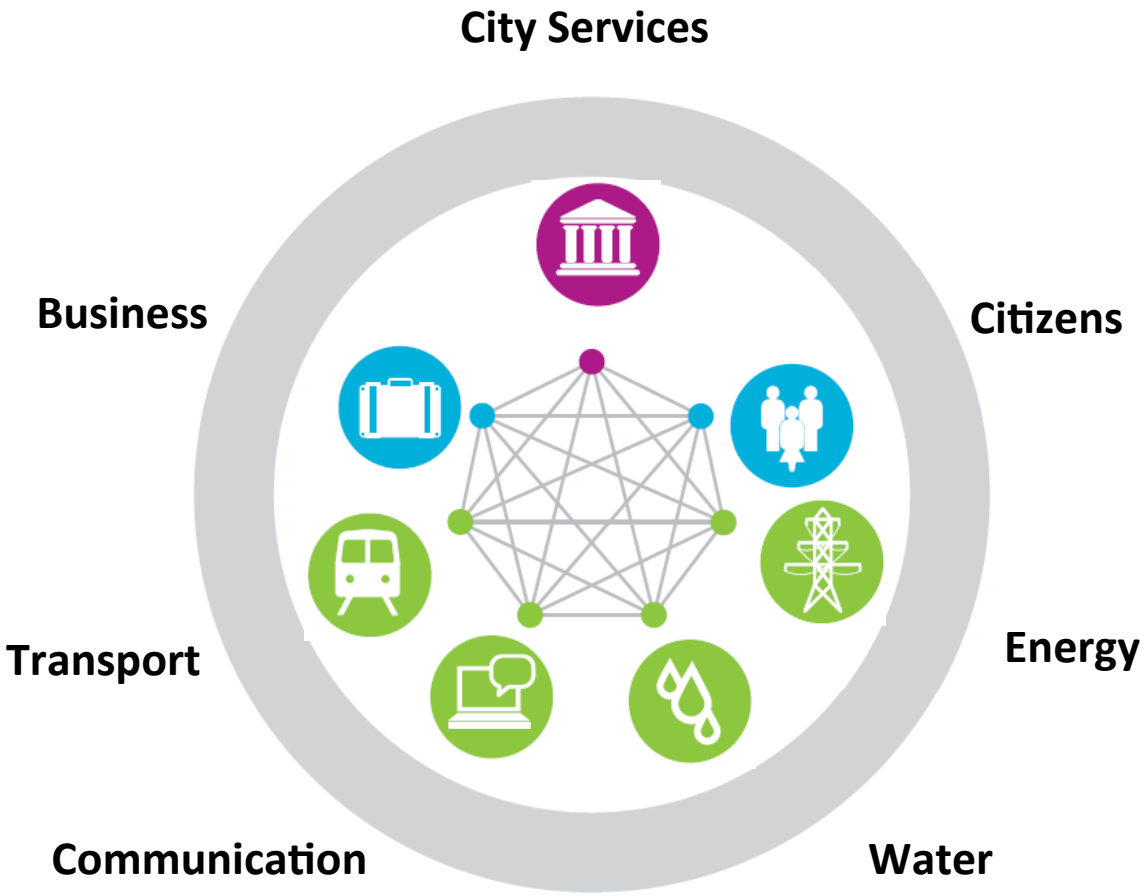
So what's getting ubiquitous and cheap? Data. And what is complementary to data? Analysis."

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

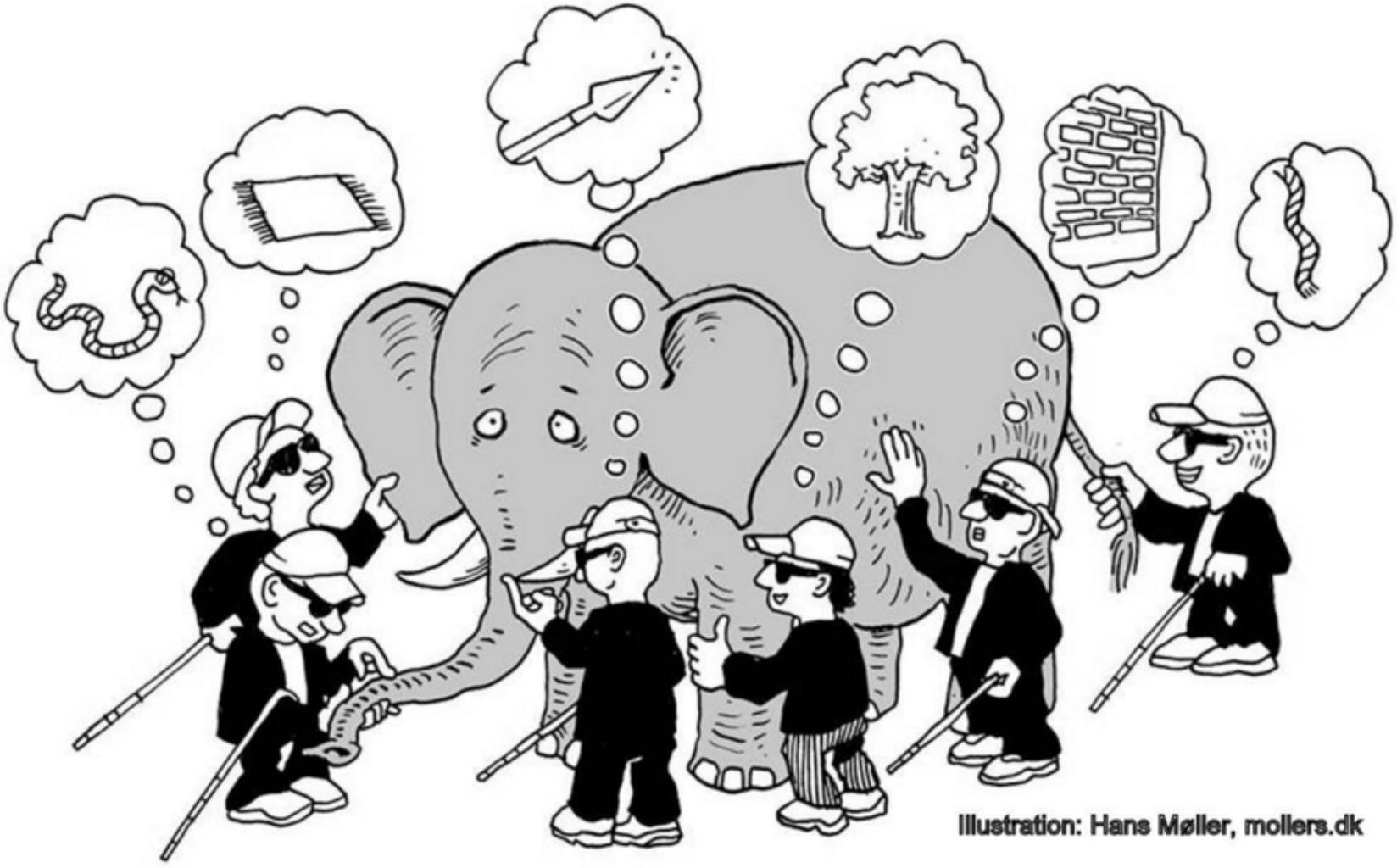


Smart City of today

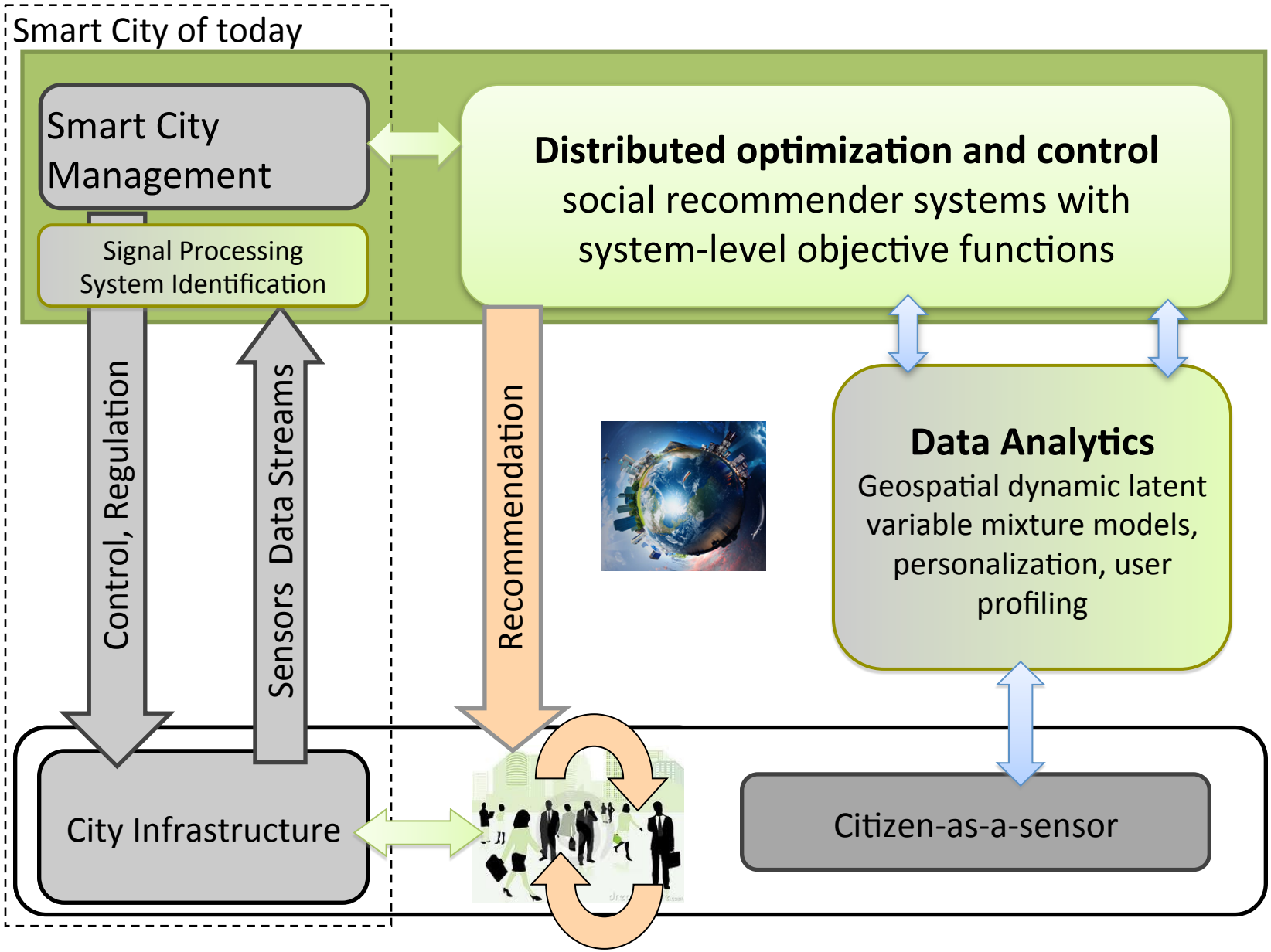




Why study systems as inter-dependent / inter-connected ?



Smart City vision: Human in the loop



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.

Course contents

Weeks 3-5. Urban data collection, handling and processing.

Lecture 5. Data acquisition: measurement and crowd-sensing.
Lecture 6. Community surveys, population census, open government data, APIs.

Lecture 7. Demand data exploration. Measuring human decisions and behaviors.
Lecture 8. Supply data exploration. Embedded sensors for infrastructure performance monitoring.

Lecture 9. Impact data exploration: environmental sensor networks, smart meters, crowd-sensing. Spatial visualizations: air quality and noise.
Lecture 10. Remote sensing technologies overview.

Course contents

Weeks 6-9. Data exploration and analysis. Demand, supply and impact.

Lecture 11. Visualizations and exploratory analysis of mobility data: bar charts and histograms.
Lecture 12. Multivariate data visualization. Understanding and modeling total number of trips between different spatial areas.
Lecture 13. Travel mode choice data, comparing two groups: drivers and transit users.

Lecture 14. Signal processing and sensor data integration. Traffic volumes and energy consumption data.
Lecture 15. Understanding spread: central limit theorem, normal distribution.
Lecture 16. Exploratory analysis, novelty detection, interactive dashboards.

Lecture 17. Concepts of environmental data modeling. Examples of air and noise pollution.
Lecture 18. Uncertainty analysis. Modeling population exposure levels: confidence intervals and decision making.

Course contents

Weeks 10-13. **Modeling and forecasting. Regression and classification.**

Lecture 19. Understanding travel data: are there differences in space?
Lecture 20. Understanding travel data: are there differences in time?

Lecture 21. Modeling travel volumes: the tale of two cities.
Lecture 22. Correlation. Direct demand models: regressing total number of trips.

Lecture 23. Causality and decision making in urban planning.
Lecture 24. Travel mode choice data revisited: mode choice as classification problem.

Lecture 25. Mode choice: nearest neighbor and linear classification.
Lecture 26. Feature selection: socio-economic and level-of-service variables.

Course contents

Weeks 14-15. **Decision making, planning and governance.**

Lecture 27. Decision support systems, conditional probability.
Lecture 28. Data fusion for scenario evaluation, Bayes theorem applications.

Lecture 29. System performance evaluation, principles of control systems.
Lecture 30. Multi-criteria decision making. Policy and equity.

Thank you!