BIG DATA and Human Mobility

THE EUNOIA PROJECT

Maxi San Miguel

IFISC

Universitat de les Illes Balears

CSIC

BIG DATA

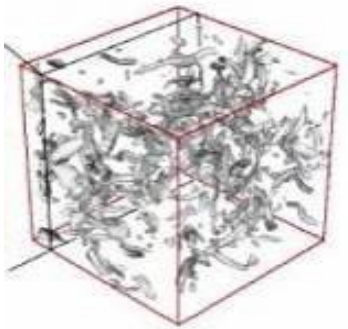
Big Data is like teenagers sex

Fourth Paradgim, J. Gray, 2007

- Thousand years ago:
  science was **empirical**
  describing natural phenomena

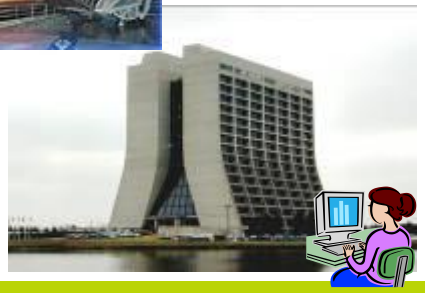
- Last few hundred years:
  theoretical branch
  using models, generalizations

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{4\pi G\rho}{3} - \mathrm{K}\frac{c^2}{a^2}$$



- Last few decades:
  a **computational** branch
  simulating complex phenomena

- Today:
  data exploration (eScience)
  unify theory, experiment, and simulation
  Information/Knowledge stored in computer
  Scientist analyzes database / files
  using data management and statistics



Data Science

-What do we understand when we know everything?:

On the face of this 'data deluge', it has been argued we are witnessing the end of theory and that the scientific method is becoming obsolete:
"The new availability of huge amounts of data, along with the statistical tools to crunch these numbers, offers a whole new way of understanding the world. Correlation supersedes causation, and science can advance even without coherent models, unified theories, or really any mechanistic explanation at all."

C. Anderson (2008) The end of theory: The data deluge makes the scientific method obsolete. Wired Magazine, see http://www.wired.com/science/discoveries/magazine/16-07/pb_theory

BIG DATA

Twitter

Cell Phone

Electronic Transactions

IFISC

Twitter Data

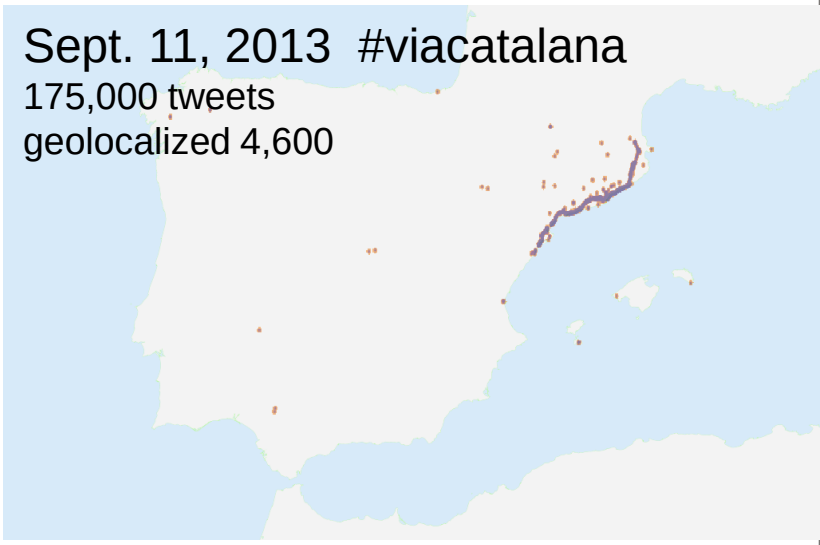
Geolocalized twitter

Mobility

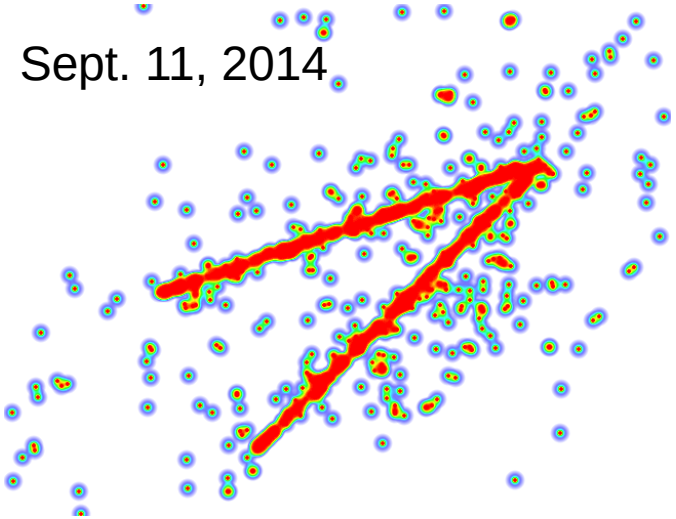
Sept. 11, 2013 #viacatalana

175,000 tweets

geolocalized 4,600



Sept. 11, 2014

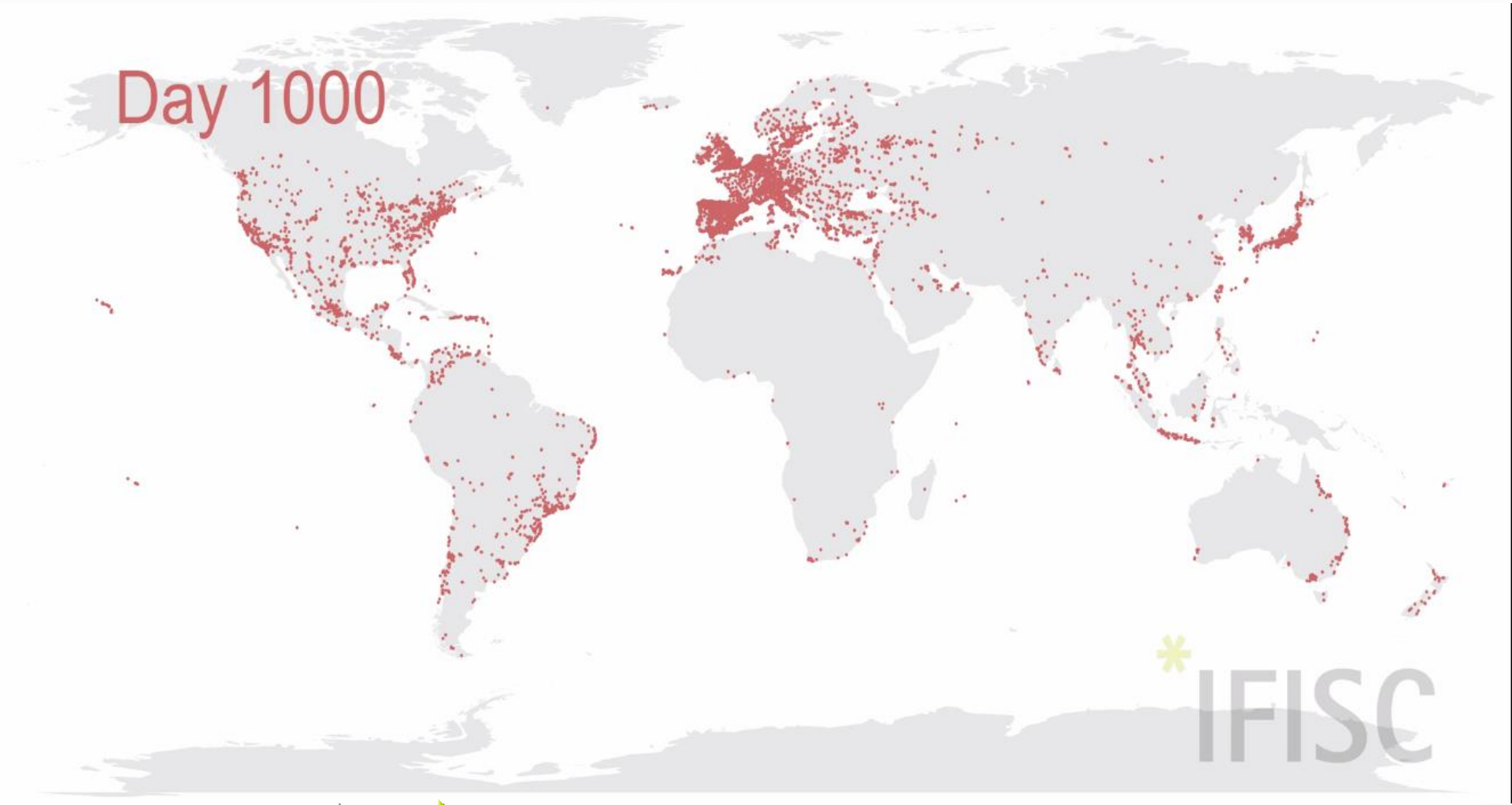


Data Mining: 15-25 Mill. Tweets/day

http://ifisc.uib-csic.es/humanmobility/

Twitter, Mobility and City Influence

J. Royal Society Interface 12, 20150473 (2015)



Day 1: Mallorca

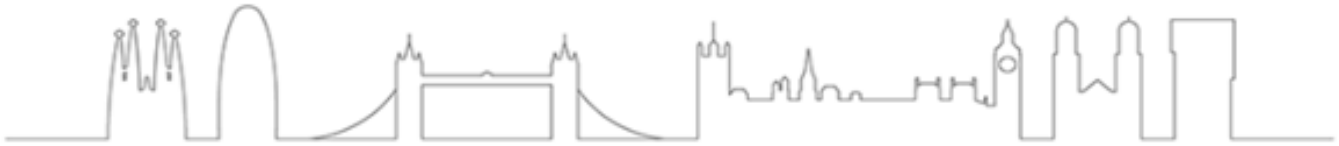
ELECTRONIC TRANSACTIONS

BBVA

http://www.youtube.com/watch?v=Zel6wych9p0

eunoia

EVOLUTIVE USER-CENTRIC NETWORKS FOR INTRAURBAN ACCESSIBILITY



SEVENTH FRAMEWORK PROGRAMME

European Commission DG.CONNECT

GSS Global Systems Science

http://eunoia-project.eu

CONSORTIUM

CSIC Instituto de Física Interdisciplinar y Sistemas Complejos — IFISC

NOMMON

ETH Eidgenössische Technische Hochschule Zürich Swiss Federal Institute of Technology Zurich

Centre for Advanced Spatial Analysis casa

cea energie atomique • energies alternatives

Antonio Lucio Independent Consultant

Ajuntament de Barcelona

Subcontractors:

iermB INSTITUT D'ESTUDIS REGIONALS I METROPOLITANS DE BARCELONA

DYM MARKET RESEARCH

Supporting Institutions:

GREATER LONDON AUTHORITY

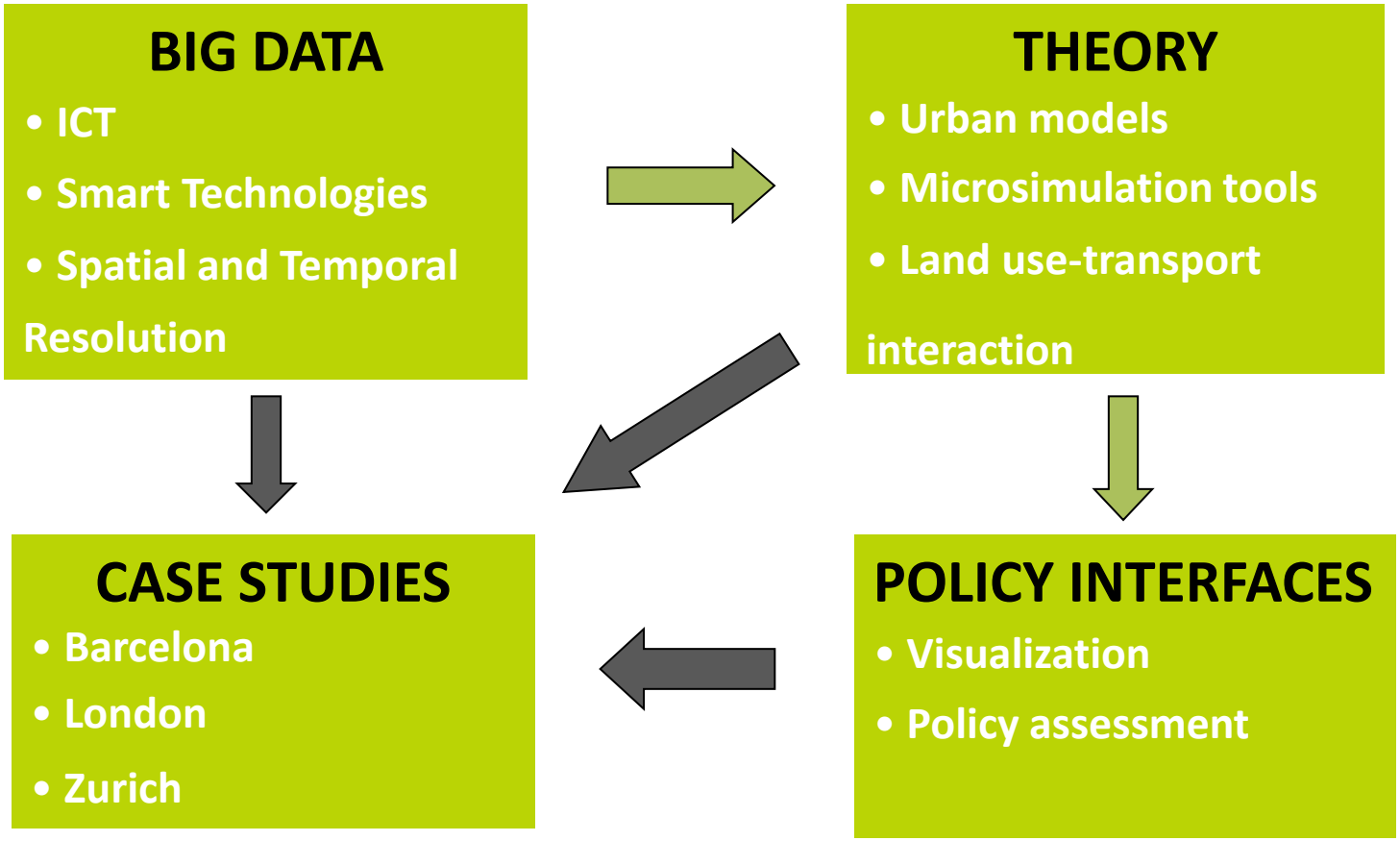
Stadt Zürich

BBVA INNOVATION CENTER

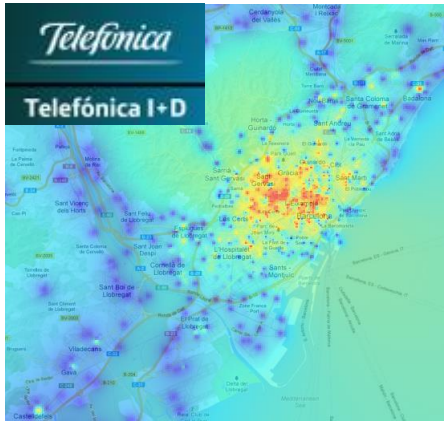
Telefónica Telefónica I+D

EUNOIA: Scope and Objectives

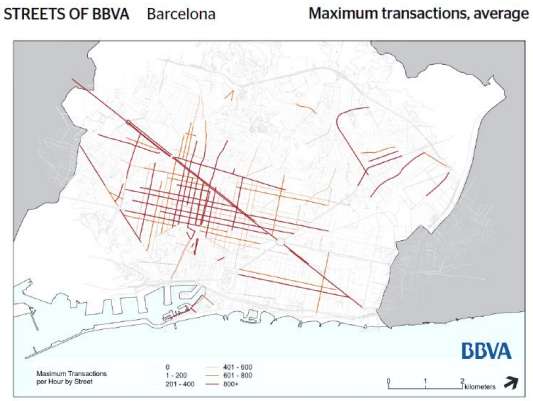
The overall goal of EUNOIA is to take advantage of the opportunities brought by smart city technologies and complex systems science to develop new urban models and ICT tools empowering city governments and their citizens to design better mobility policies



EUNOIA Data

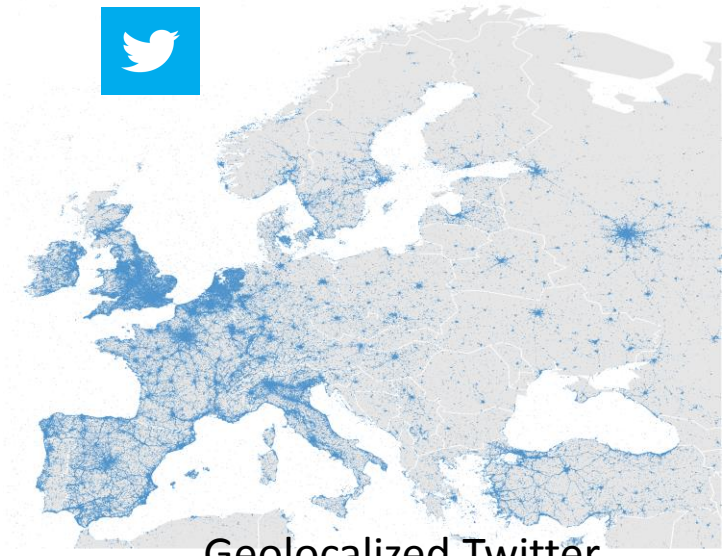


Mobile Phone



Electronic transactions

Transport Card London



Geolocalized Twitter

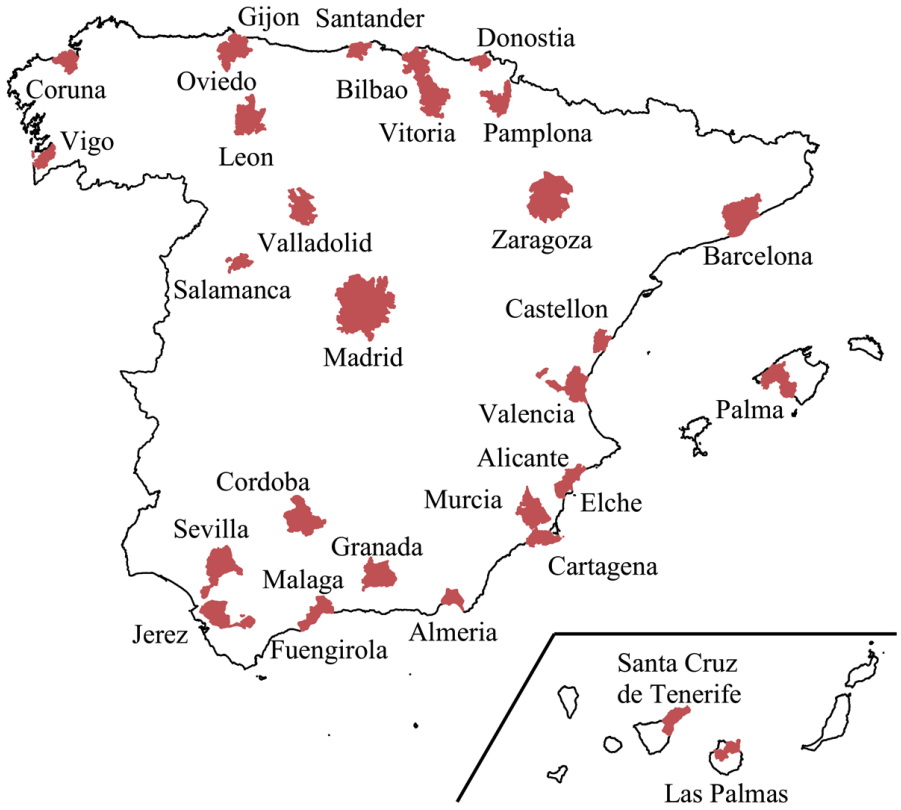
- GIS Data
- Census and conventional information from Local Authorities on mobility and transport.
- Surveys (Barcelona)

2.1 Geographical data (GIS files)
2.2 Land use
2.3 Surveys and census
2.4 Transport
2.5 Geolocated Twitter

IFISC data repository

Filename	City	Description
Barcelona_tweets_20130523.zip	Barcelona	Geolocated tweets in BCN.
Barcelona_network_20130523.zip	Barcelona	Follower network between users posting gelocated tweets in BCN.
London_tweets_20130523.zip	London	Geolocated tweets in London.
London_network_20130523.zip	London	Follower network between users posting gelocated tweets in London.

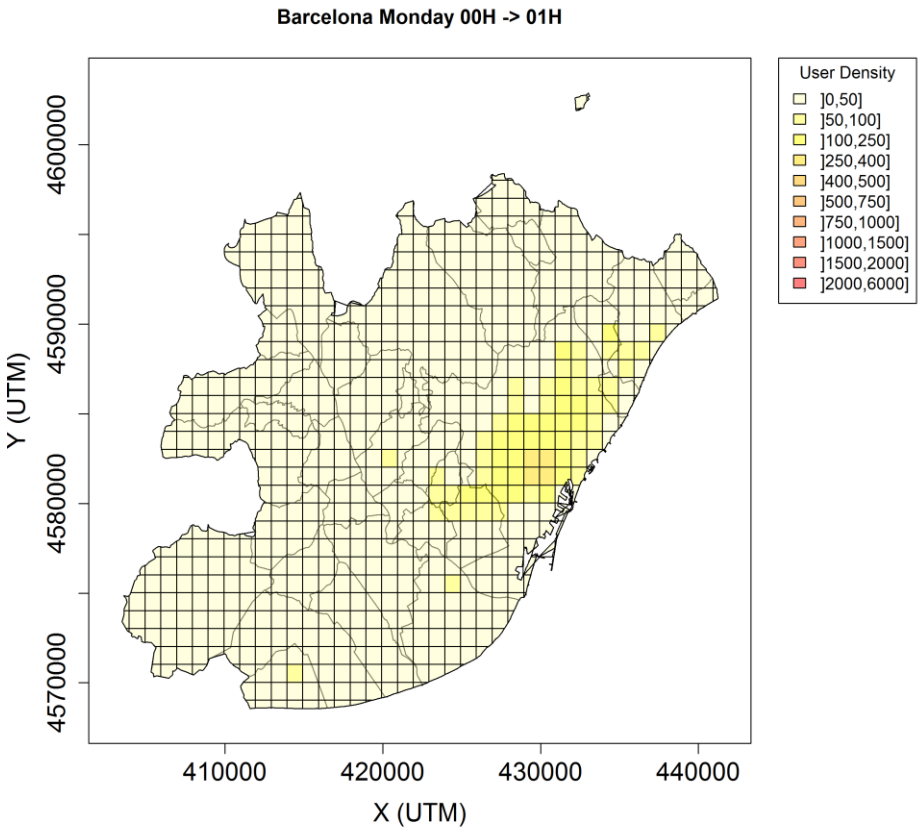
Cell phone data

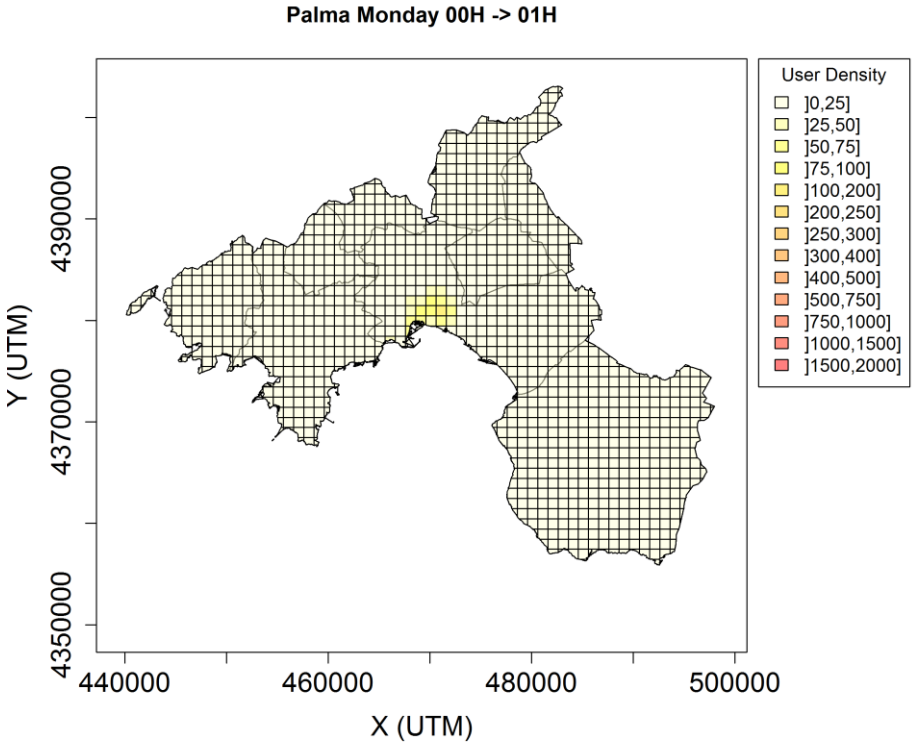


31 urban areas with more than 200,000 inhabitants
55 days in fall 2009
Number of users (per hour): on average 2% population (max 5%)

CELL PHONE DATA

Time distribution #users/km^2

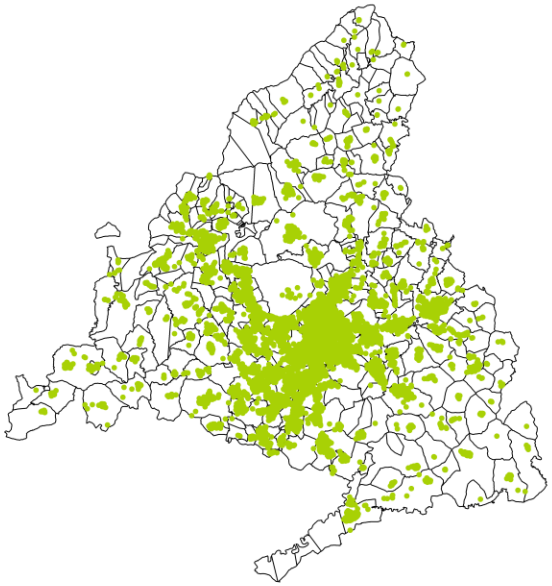


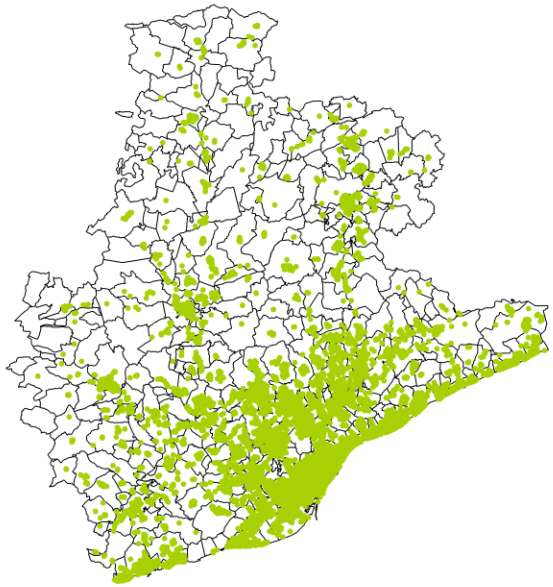


Provinces of Madrid and Barcelona in 2011-12

- 130 M of transactions
- 3.5 M of **BBVA** customers
- 320,000 businesses

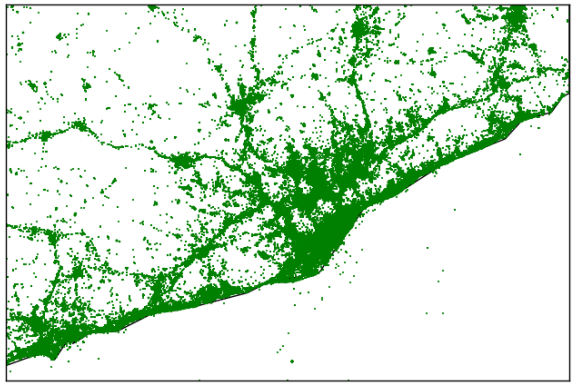
BBVA INNOVATION CENTER



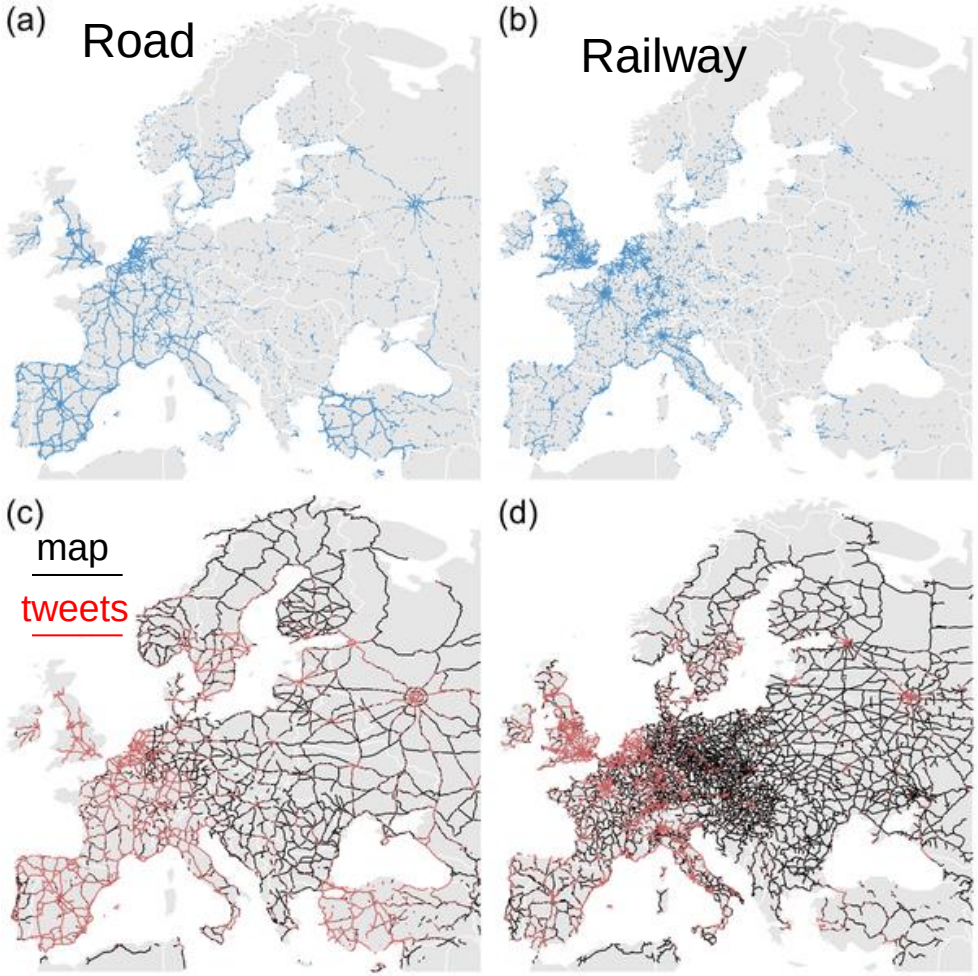


TWITTER DATA

15-25 Mill. tweets per day. 1 year
1,4 Mill. geoloc in BCN
25,000 users geolocalized in BCN



6 Mill. geoloc in LON
260,000 users geolocalized in LON



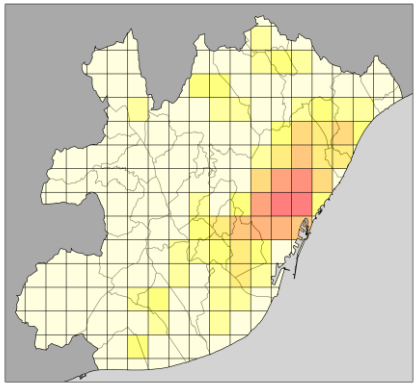
Tweets on the road, M. Lenormand et al PLoS ONE 9(8): e105407 (2014)

- Is this good data? Validation
- How can we use it to replace traditional sources of data?
- What can we learn from it?

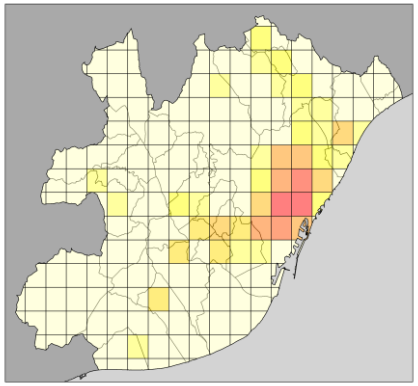
EUNOIA: Validation of ICT data

PLoS ONE ***9, e105184 (2014)***

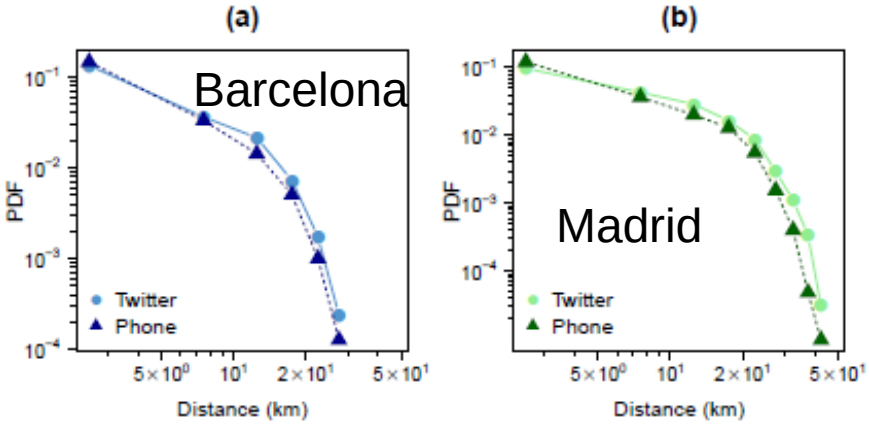
Phone



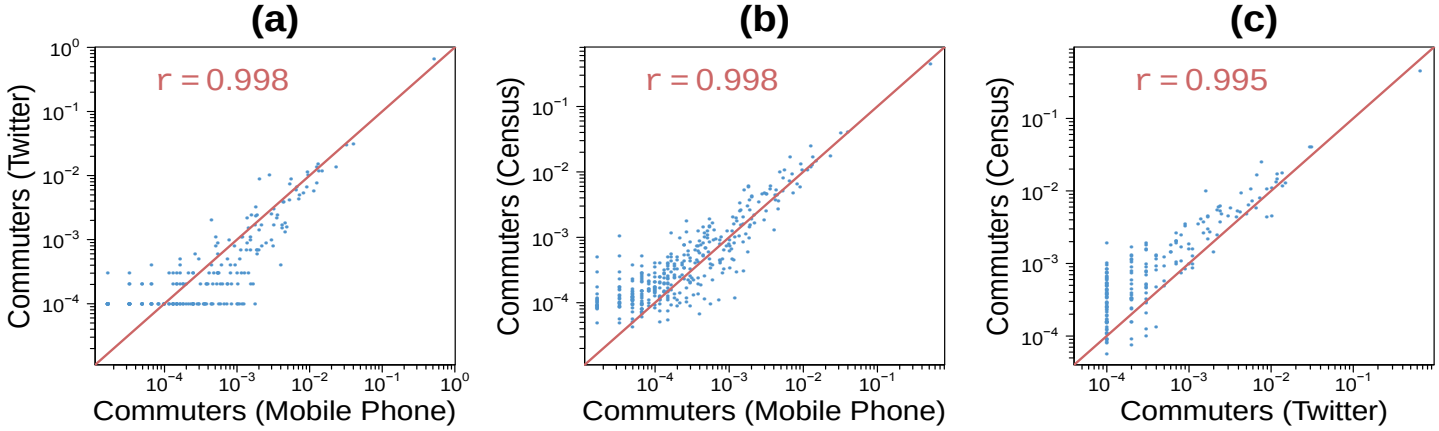
Twitter



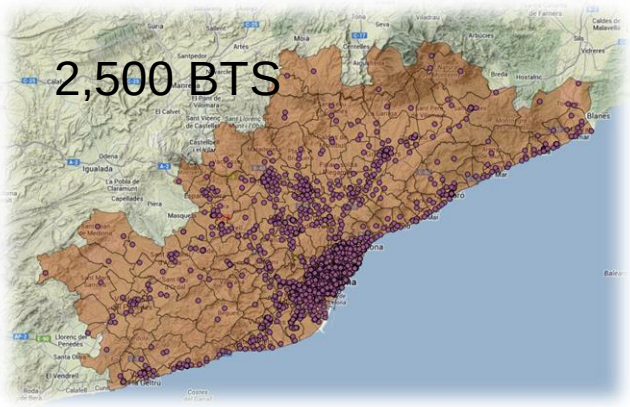
Distribution of commuting distances

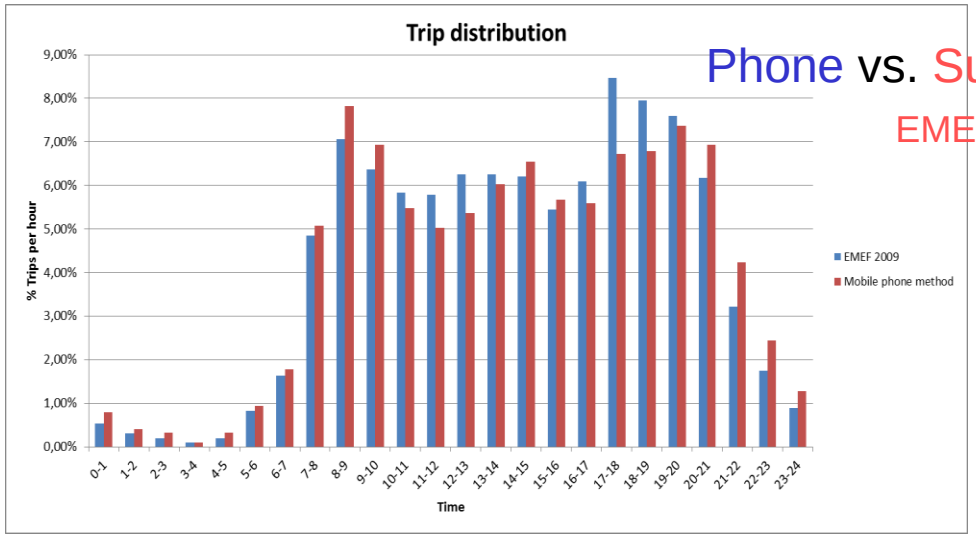


Twitter vs. Phone vs. Census: Ex. Commuting fluxes in Barcelona



ODs from mobile phone data





Phone vs. Survey

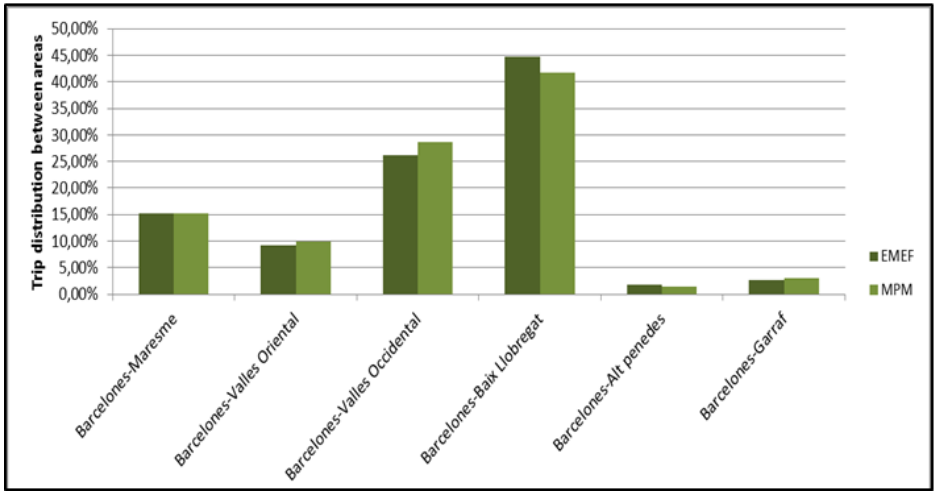
EMEF2009

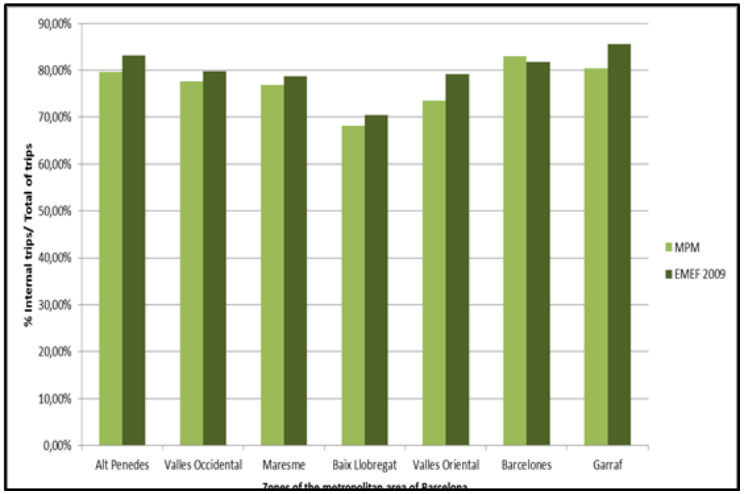
LOW COST ODs

Trip distribution (Barcelonès)

Phone vs. Survey

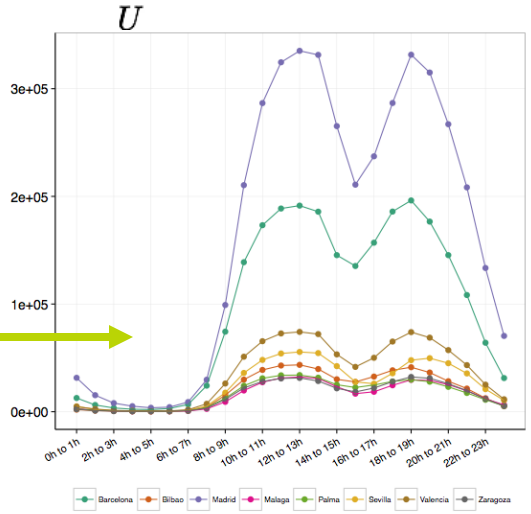
Intra-zone trips (Municipalities)

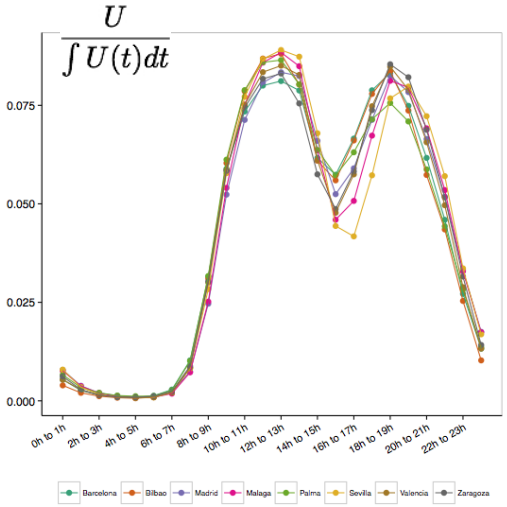




Urban rhythm

number phone users per hour during an average weekday



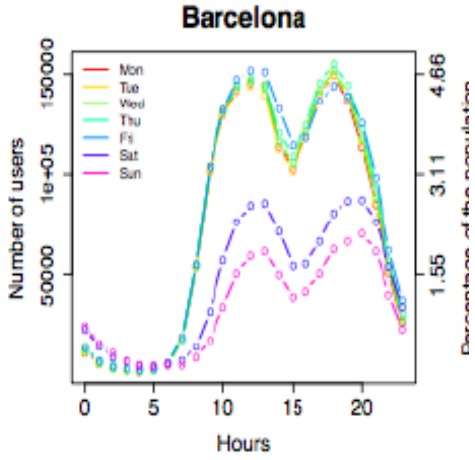


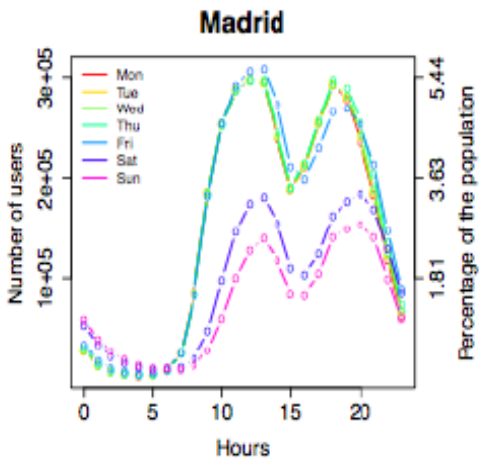
TELEPHONE

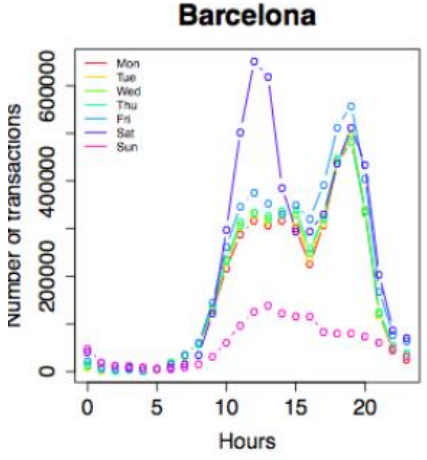
Telefónica I+D

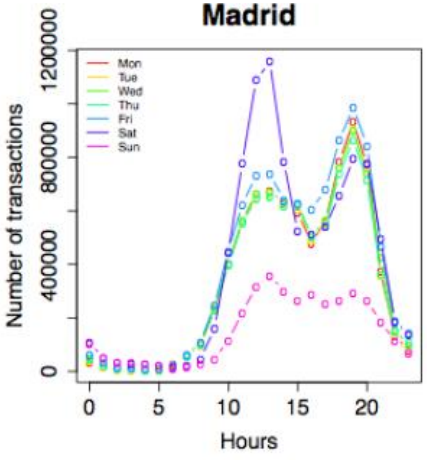
CREDIT CARD

BBVA INNOVATION CENTER



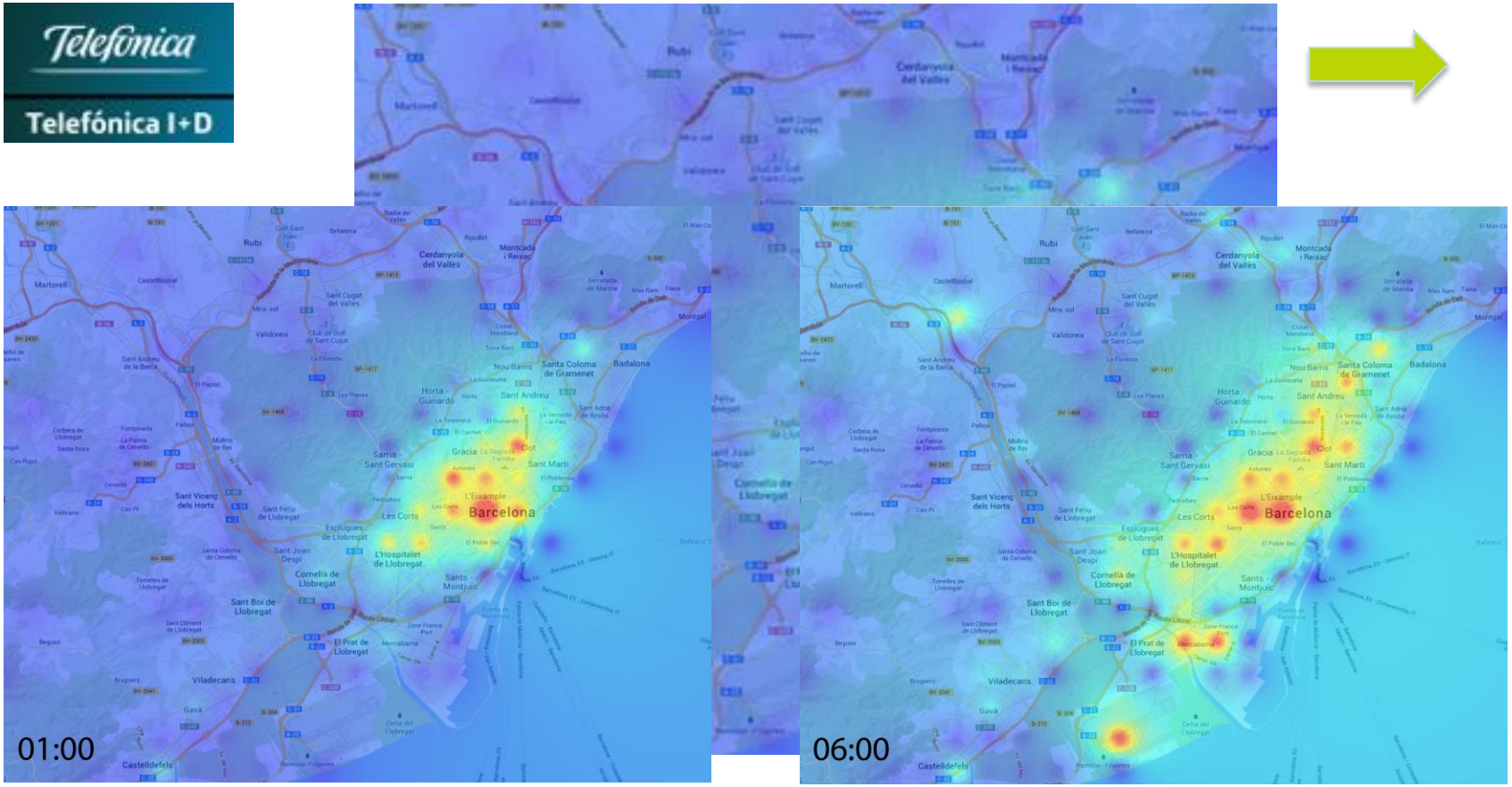






Hotspots:

Telefónica I+D



Scientific Reports 4, 5276 (2014)

Hotspots and Polycentric cities

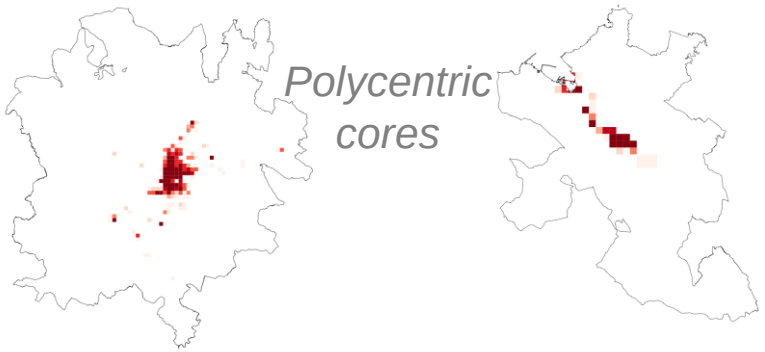
Sevilla (1.200.000)

Granada (470.000)

Monocentric cores

Madrid (6.100.000)

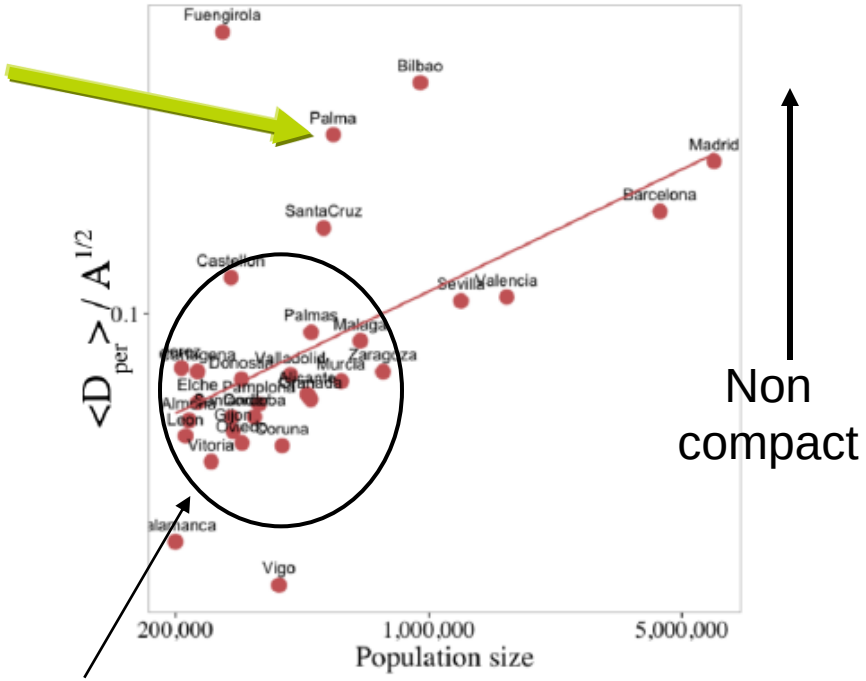
Bilbao (950.000)



Polycentric cores

31 cities:
Number of hot spots is not proportional to population *(sublineal)*

Compacity coefficient

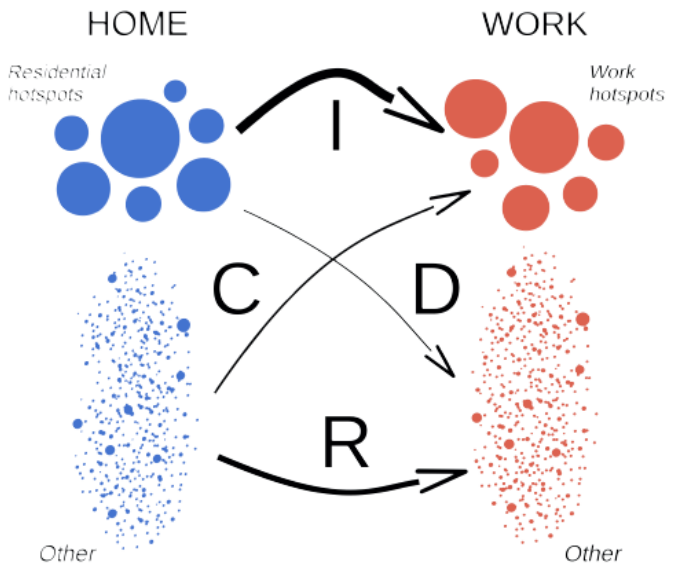


Non compact

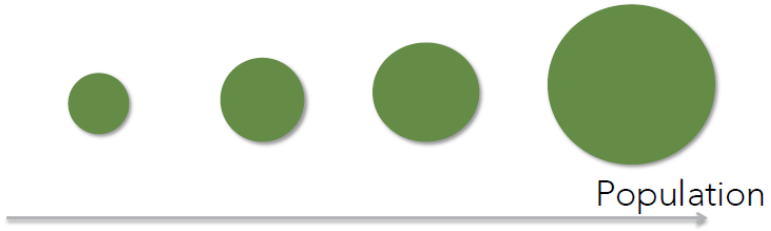
Compact cities

Nature Communications **6,** 6007 (2015)

Mesoscale description of urban mobility



Structure of flows in Spanish cities



	Cordoba Gijon Vitoria	Zaragosa Malaga	Valencia Sevilla	Madrid Barcelona
R	27%	36%	41%	46%
I	43%	37%	31%	25%

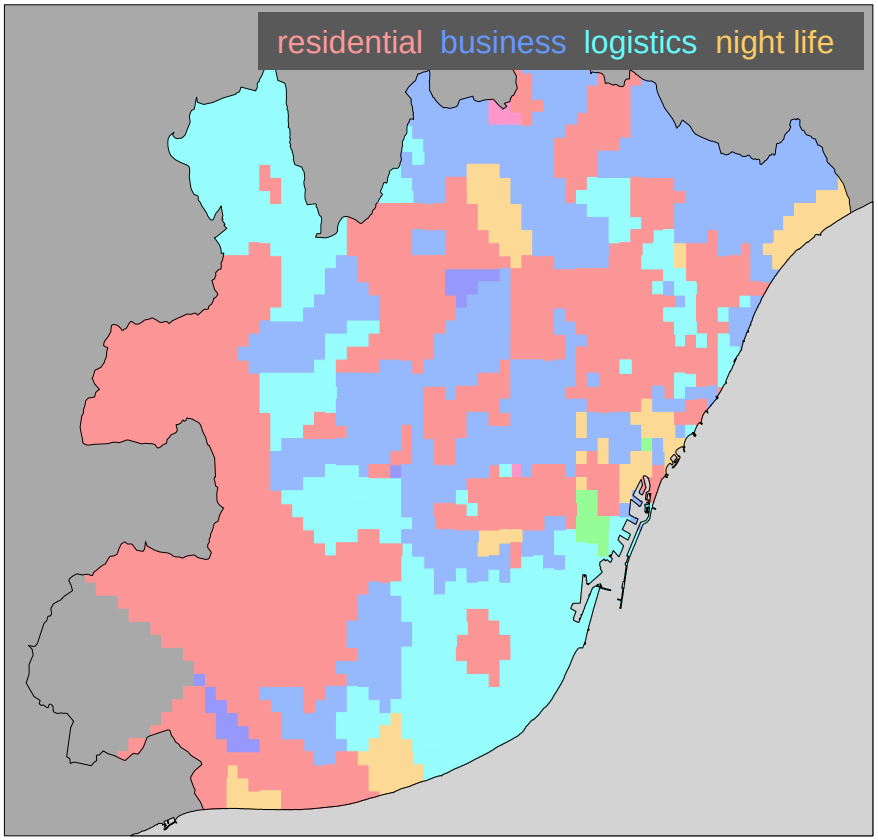
Increased possibility to commute
from anywhere to anywhere in bigger cities

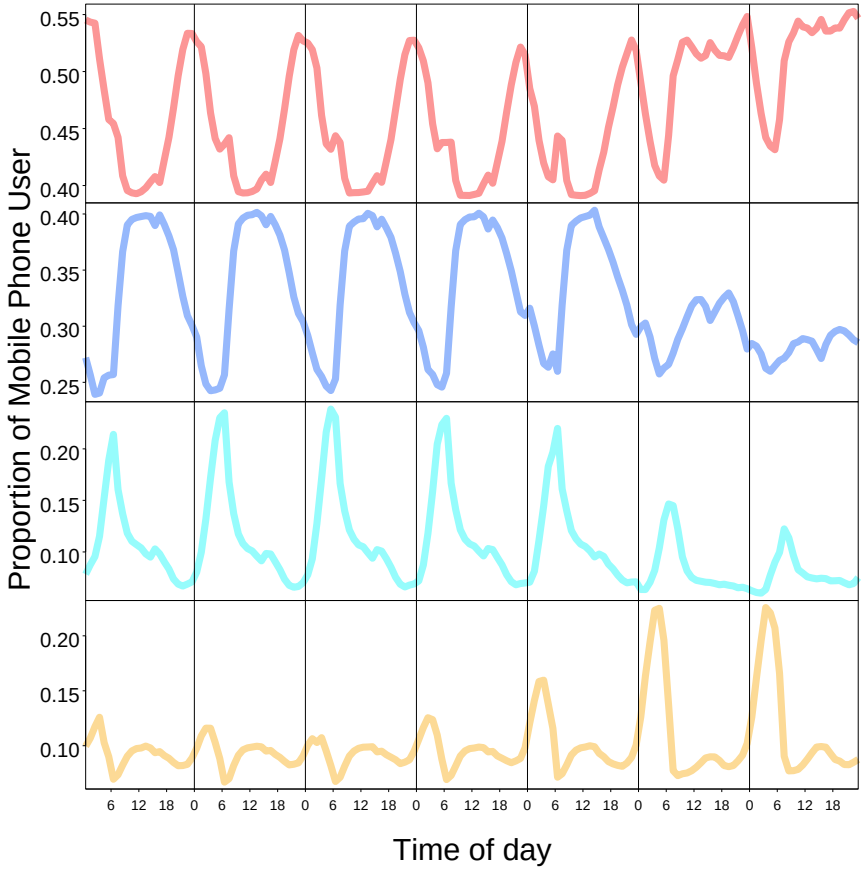
Dominance of R flows in large cities confirmed in Oyster card data for London

From Data to Urban Knowledge

M. Lenormand et al., Royal Soc. Open Science (2015) arXiv: 1503.06152

Landuse from the functional network of the city





eunoia EVOLUTIVE USER-CENTRIC NETWORKS FOR INTRAURBAN ACCESSIBILITY

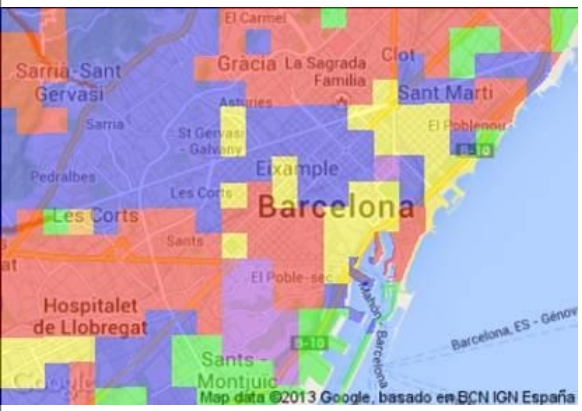
insight

M. Lenormand et al., Royal Soc. Open Science (2015) arXiv: 1503.06152

Landuse from the functional network of the city

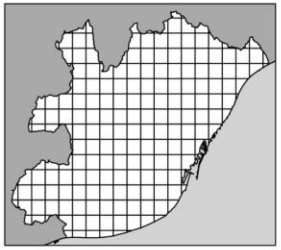
Network approach to determine land-uses from mobile phone data

Automatic detection of 4 main land use whose relative proportions are very close from one city to another

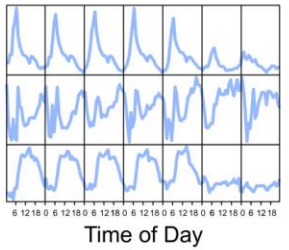


business
residential
logistics
night life

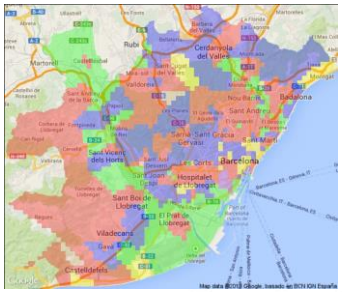
Metropolitan Area



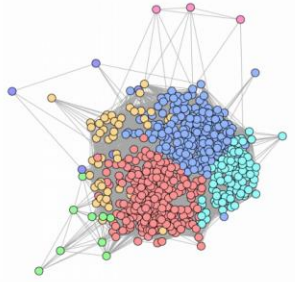
Recordings sites

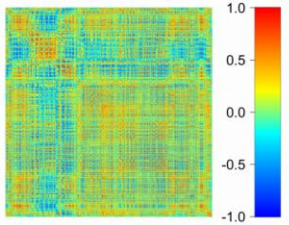


Signals



Functional Network

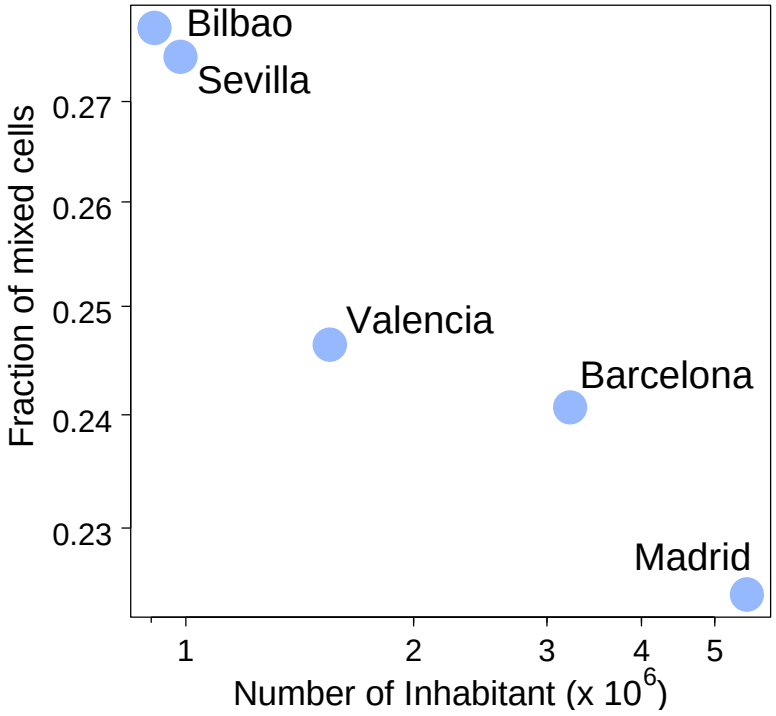


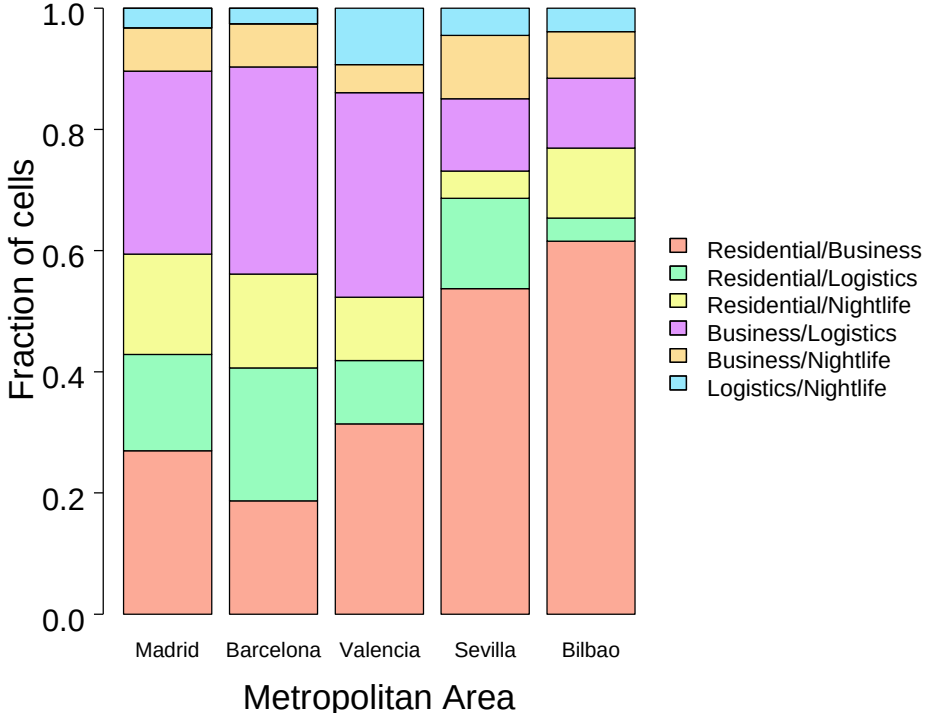


Correlation Matrix

M. Lenormand et al., Royal Soc. Open Science (2015)arXiv: 1503.06152

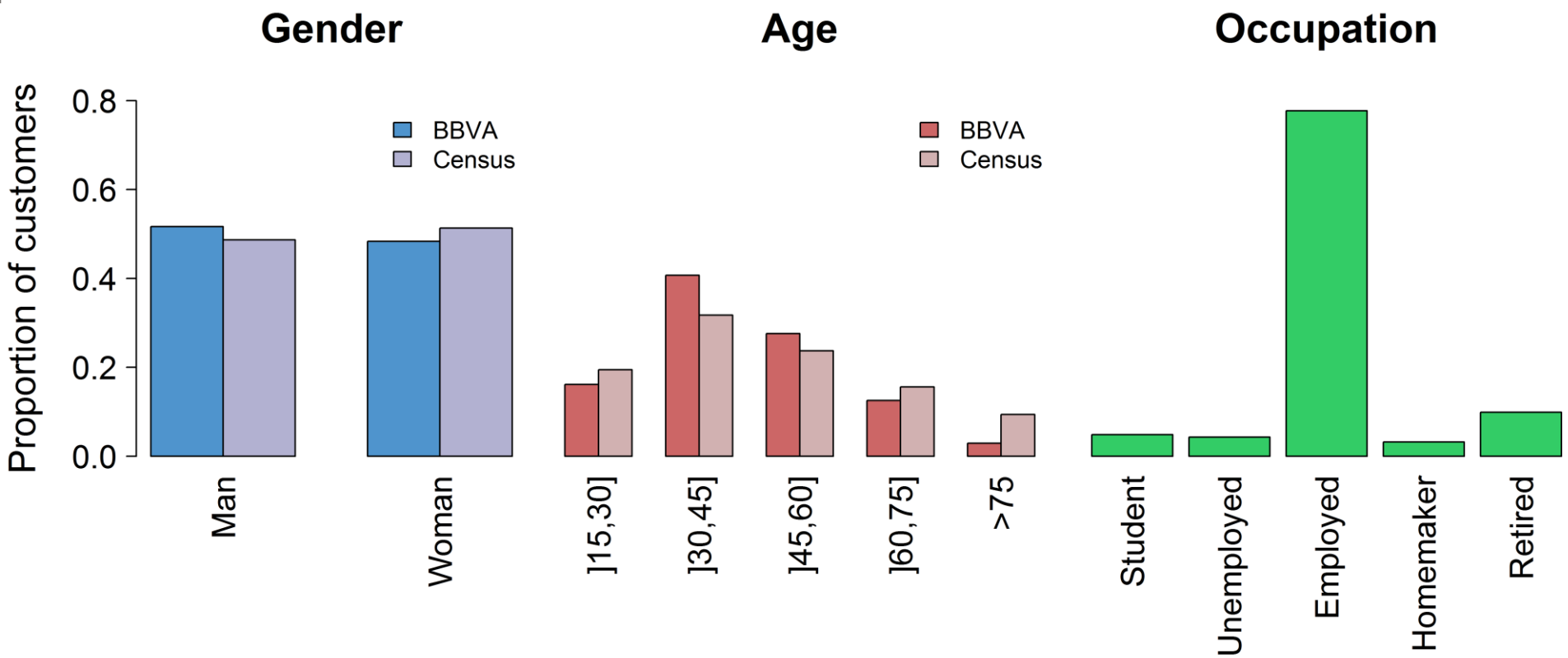
Landuse from the functional network of the city





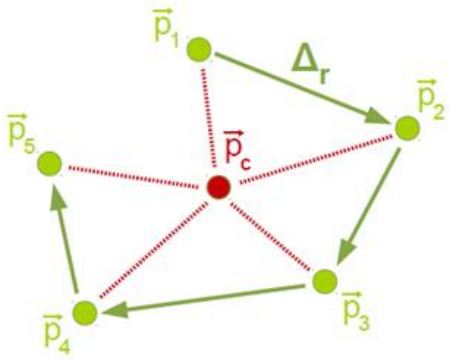
BBVA Data vs Census

M. Lenormand, et al. Sci. Rep. 20 May 2015



Sociodemographics of mobility: occupation

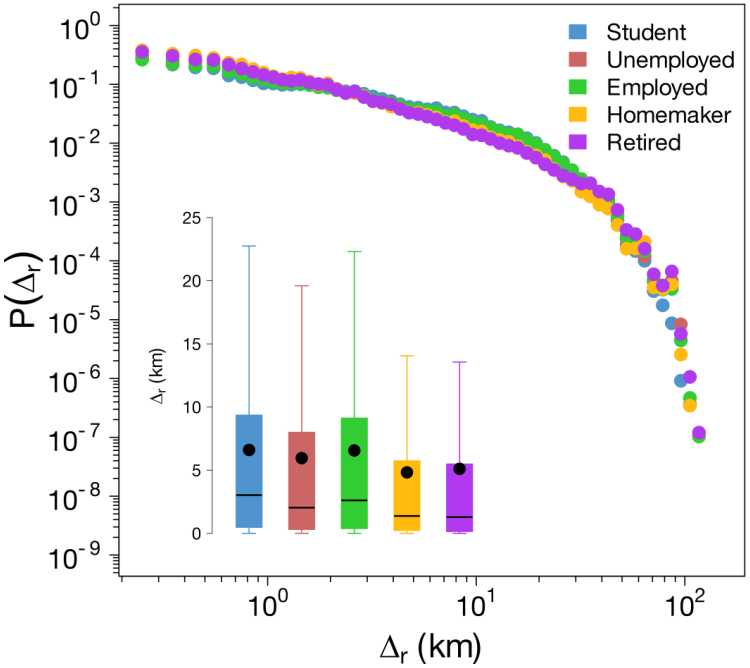
M. Lenormand, et al. Sci. Rep. 20 May 2015



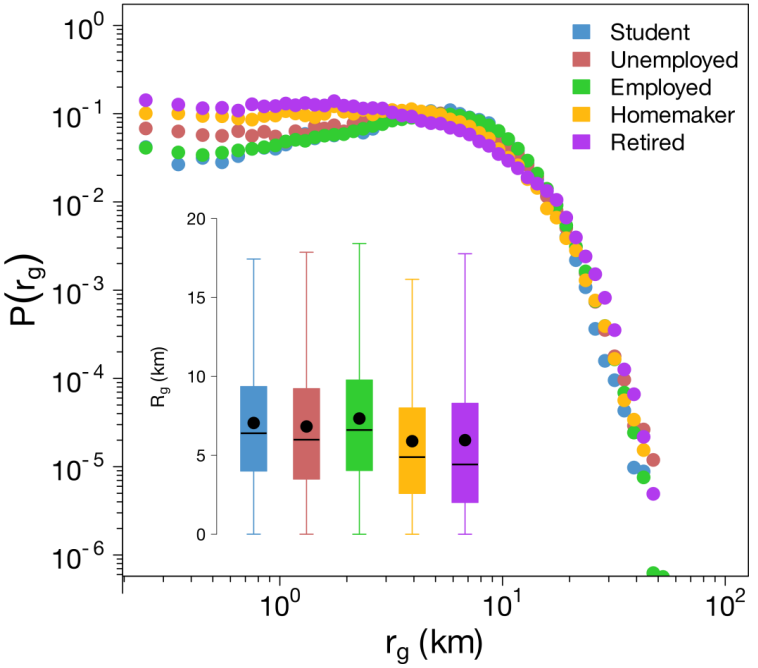
$$\Delta_r = \sqrt{(\vec{p}_i - \vec{p}_{i+1})^2}$$

$$r_g = \sqrt{\frac{1}{n}\sum_{k=1}^{n}(\vec{p}_k - \vec{p}_c)^2}$$

(a)

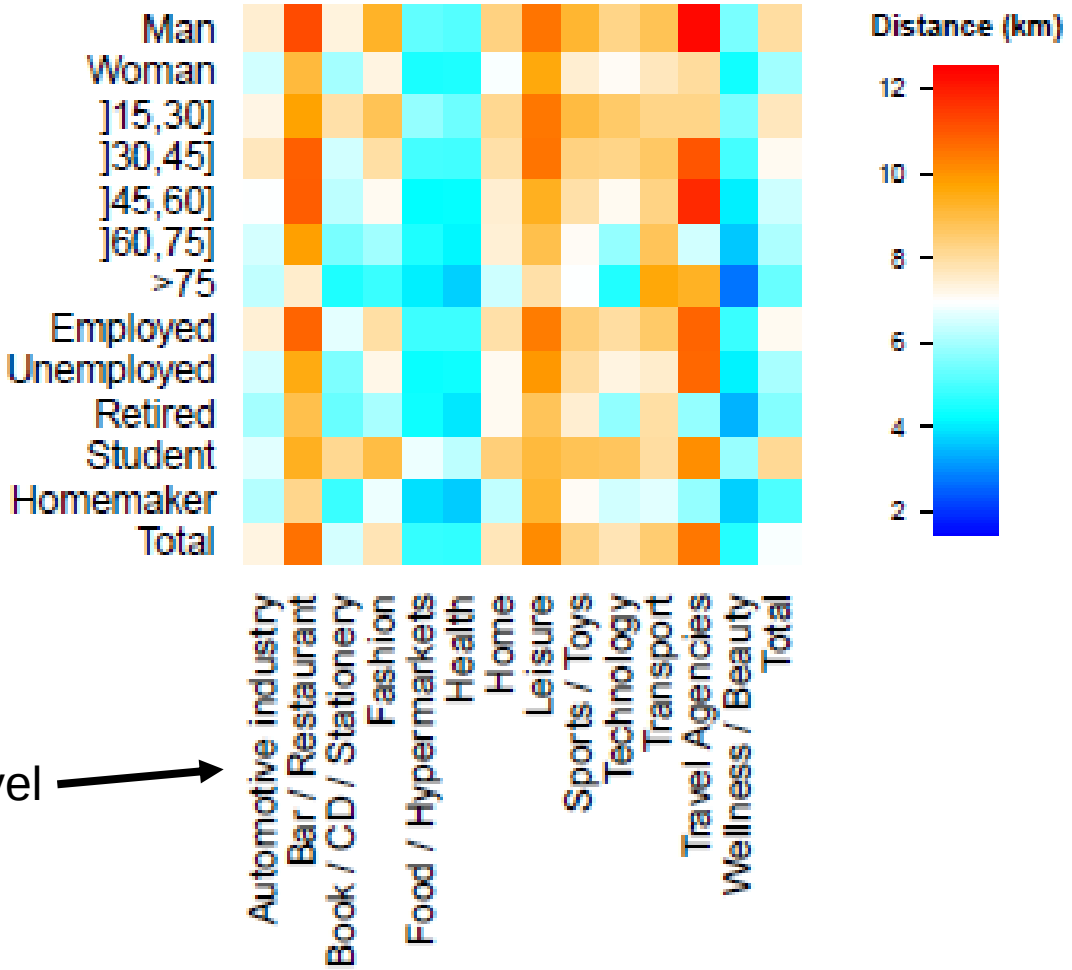


(b)



M. Lenormand, et al. Sci. Rep. 20 May 2015

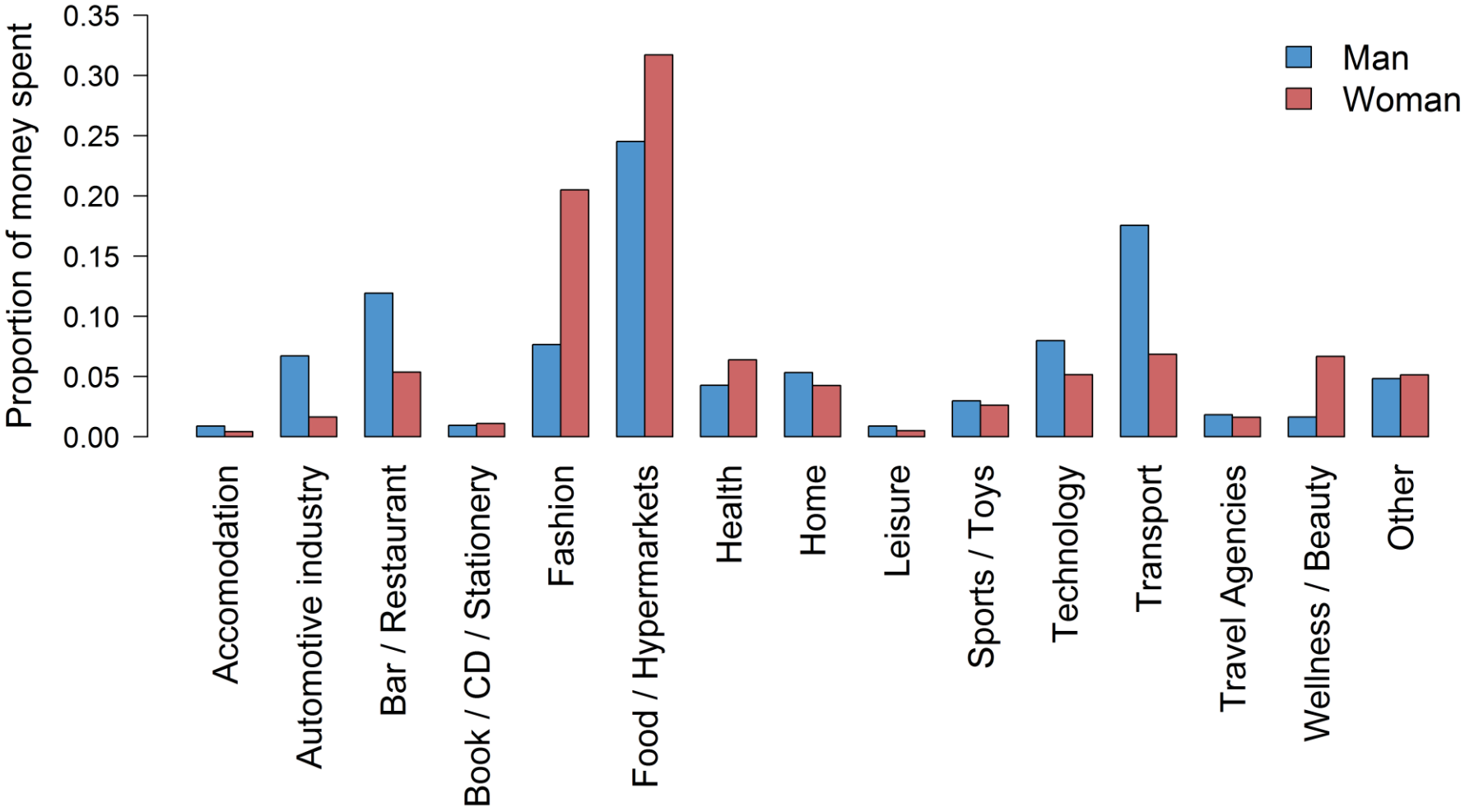
Influence of sociodemographic characteristics on human mobility



Purpose of travel →

Money spent

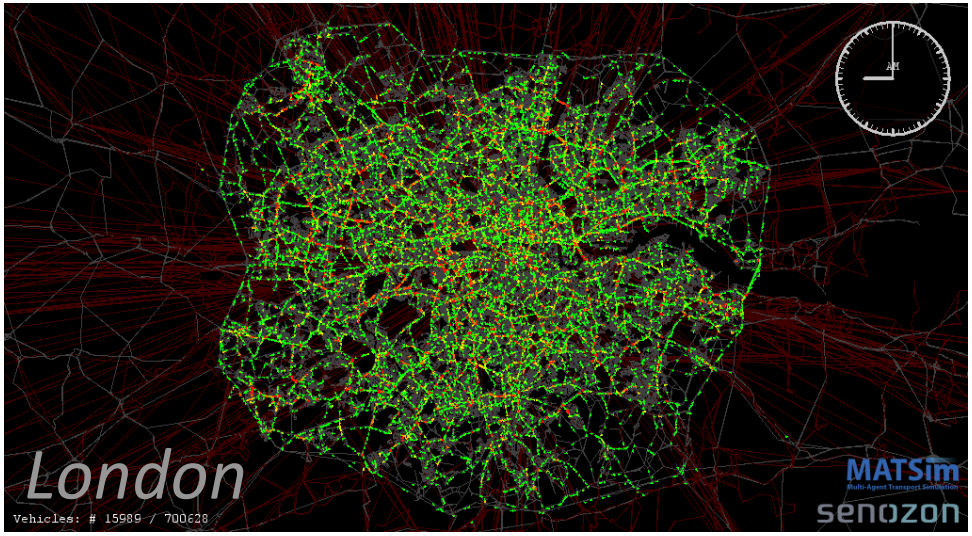
M. Lenormand, et al. Sci. Rep. 20 May 2015



IFISC

MATSim
Multi-Agent Transport Simulation

Agent Based Modelling

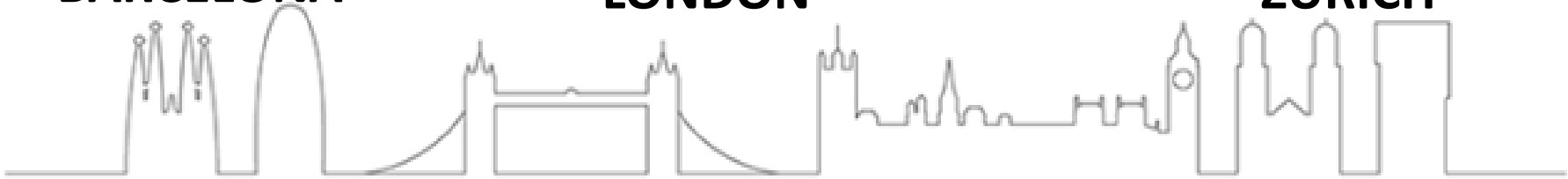


EUNOIA Developments:

- **New type of input data (mobile phone, oyster card…)**
- **Integrate social relationships: Social network** ***(findings from phone data)***
- **Joint decisions: interindividual interactions. Game theory for social coordination**
- **Joint resources: Bycicle module**

EUNOIA Case studies: A tale of three cities

BARCELONA **LONDON** **ZURICH**



Policy context:

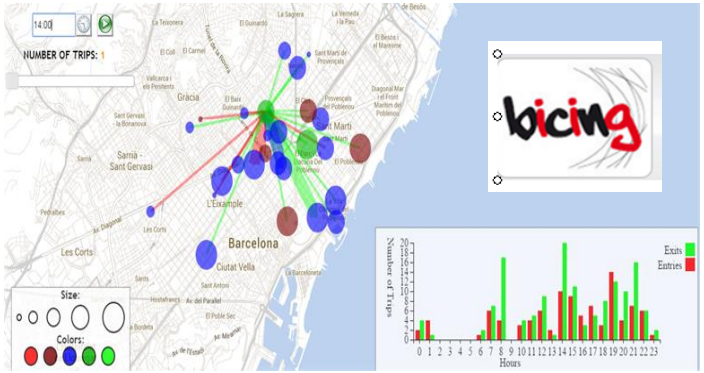
- Bike redistribution
- What if scenarios

Methodology development:

- Conventional + nonconventional data
- Implementation of bike MATSim module
- Implementation of visualization tool

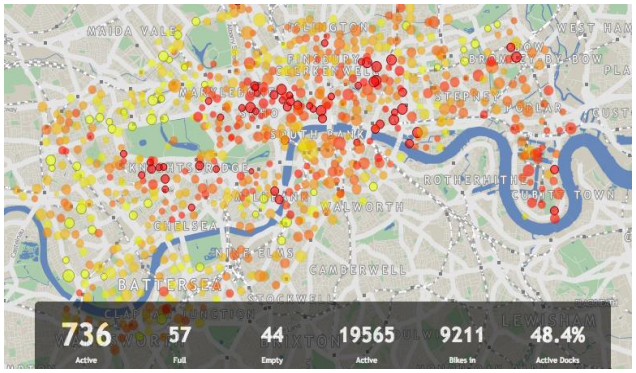
EUNOIA Case studies: A tale of three cities

BARCELONA



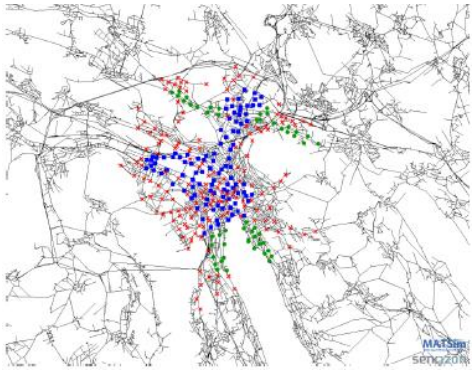
- Mobile Phone data input
- Uphill/downhill trips
- Visualization

LONDON



- Oyster card data input

ZURICH



- Joint choice:
  -escort to school
  -leisure travel
- Uphill-downhill

BEFORE AND AFTER EUNOIA

- Validation of ICT data for urban mobility
- New methods of data analysis for urban mobility
- New ICT Data driven approach to modeling and microsimulation

→ Results call for a radical change of urban policy praxis by the use of new, cheap, non-aggregated ICT data

Some practical challenges

- **Data:** -Multilevel and multiscale
  -Data mining vs Data analytics
  -Data driven vs. Question driven
  -Data privacy
  -Data sharing/property: Academics, Companies, Government

- **Limits of prediction/forecasting vs decision making**

- **Science in support of policy options:**

  -Policy makers are generally not interested in discussing challenges unless they come with solutions.

  -Policy makers are interested in a single answer, but a single answer often is not found

The end of politics?

Do not care to vote, an algorithm is in charge!

What do we understand when we know everything?

Data analysis: 2nd great age of mapping

Science as the art of abstraction:

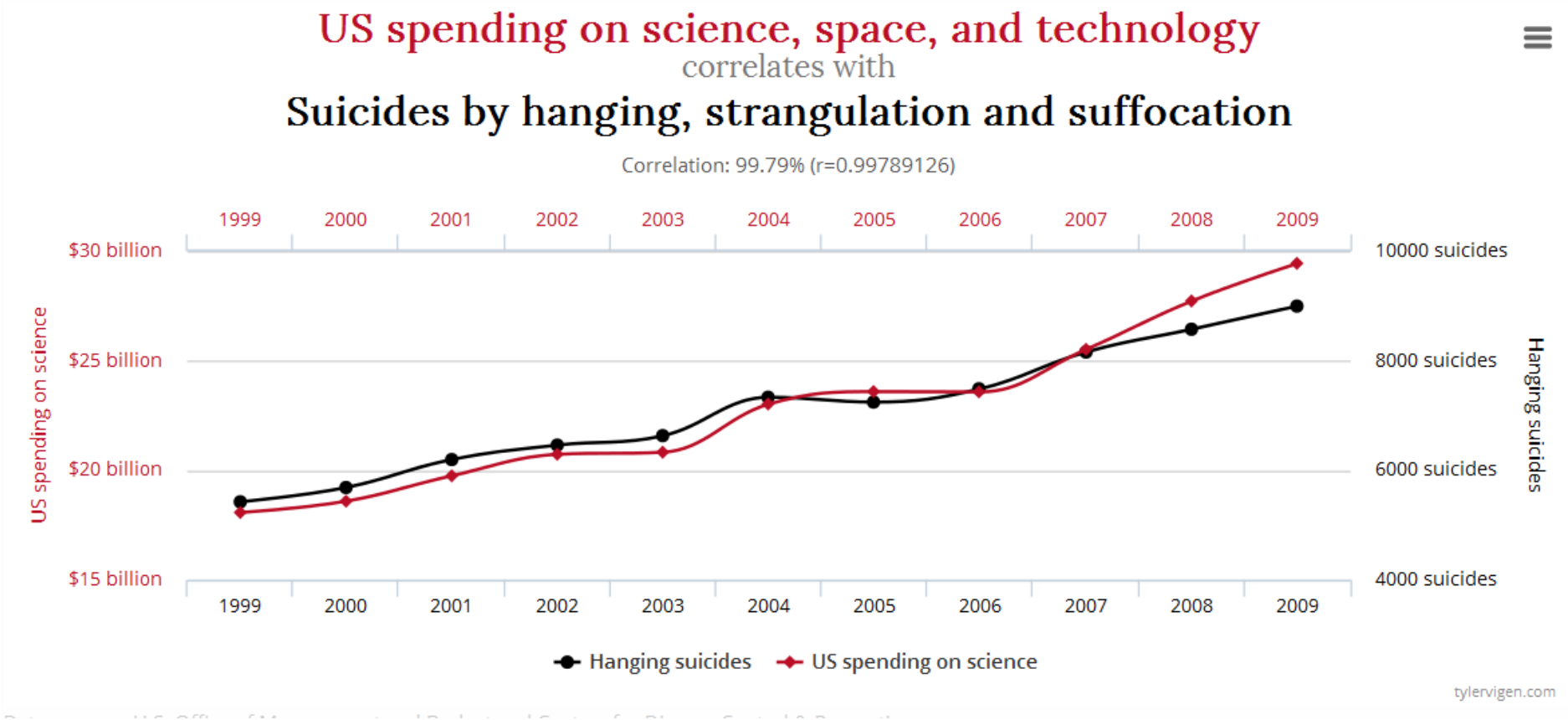
"What do you consider the largest map that would be really useful?" "About six inches to the mile." "Only six inches!" exclaimed Mein Herr. "We very soon got six yards to the mile. Then we tried a hundred yards to the mile. And then came the grandest idea of all! We actually made a map of the country, on the scale of a mile to the mile!" "Have you used it much?" I enquired. "It has never been spread out, yet," said Mein Herr: "The farmers objected: they said it would cover the whole country, and shut out the sunlight! So now we use the country itself, as its own map, and I assure you it does nearly as well (From Lewis Carroll)

Questions and answers:

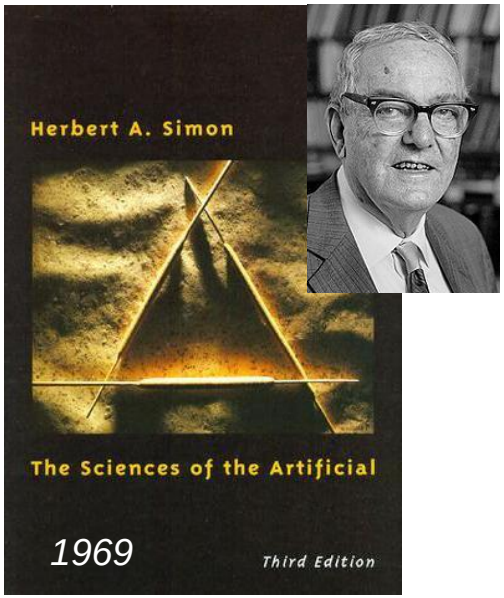
Computers are useless: They only provide answers! (Pablo Picasso)

http://www.tylervigen.com/

SPURIOUS CORRELATIONS



Participatory Challenge



Herbert A. Simon

-Nobel Prize in Economics decision-making process (1978)

-Founding father of

Artificial Intelligence
Bounded Rationality
Complex Systems
….

What we ordinarily call "style" may stem just as much from these decisions about the design process as from alternative emphases on the goals to be realized through the final design.

When we come to the design of systems as complex as cities, or buildings, or economies, we must give up the aim of creating systems that will optimize some hypothesized utility function, and we must consider whether differences in style of the sort I have just been describing do not represent highly desirable variants in the design process rather than alternatives to be evaluated as "better" or "worse."

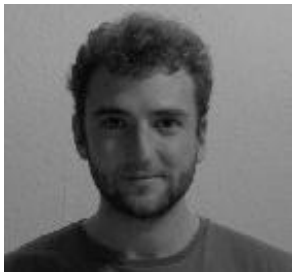
We have usually thought of city planning as a means whereby the planner's creative activity could build a system that would satisfy the needs of a populace. Perhaps we should think of city planning as a valuable creative activity in which many members of a community can have the opportunity of participating if we have wits to organize the process that way.

IFISC

THANKS!

José J. Ramasco

Pere Colet



Maxime Lenormand

Fabio Lamanna

Antònia Tugores

Aleix Bassolas