



New data sources to study airport competition

Riccardo Gallotti, Marc Fuster, Jose Javier Ramasco
IFISC (UIB-CSIC) Mallorca, Spain

Belgrade, November 29th, 2017



Exploratory research



BigData4ATM



Exploratory research



DATA DRIVEN

- explore the opportunity for new data-informed, modelling of passenger's behavior
- point out available data sources
- open the modelling route very roughly, so that more refined techniques can follow



Focus on new data

We integrate traditional data-sources...



Advantages:

- framed and targeted
- established and reliable

Disadvantages

- limited to a-posteriori analysis
- aggregated
- expensive

Focus on new data

... with new sources of data



Advantages:

- huge statistical population sizes (big)
- open access
- microscopic information
- opportunity for now-casting

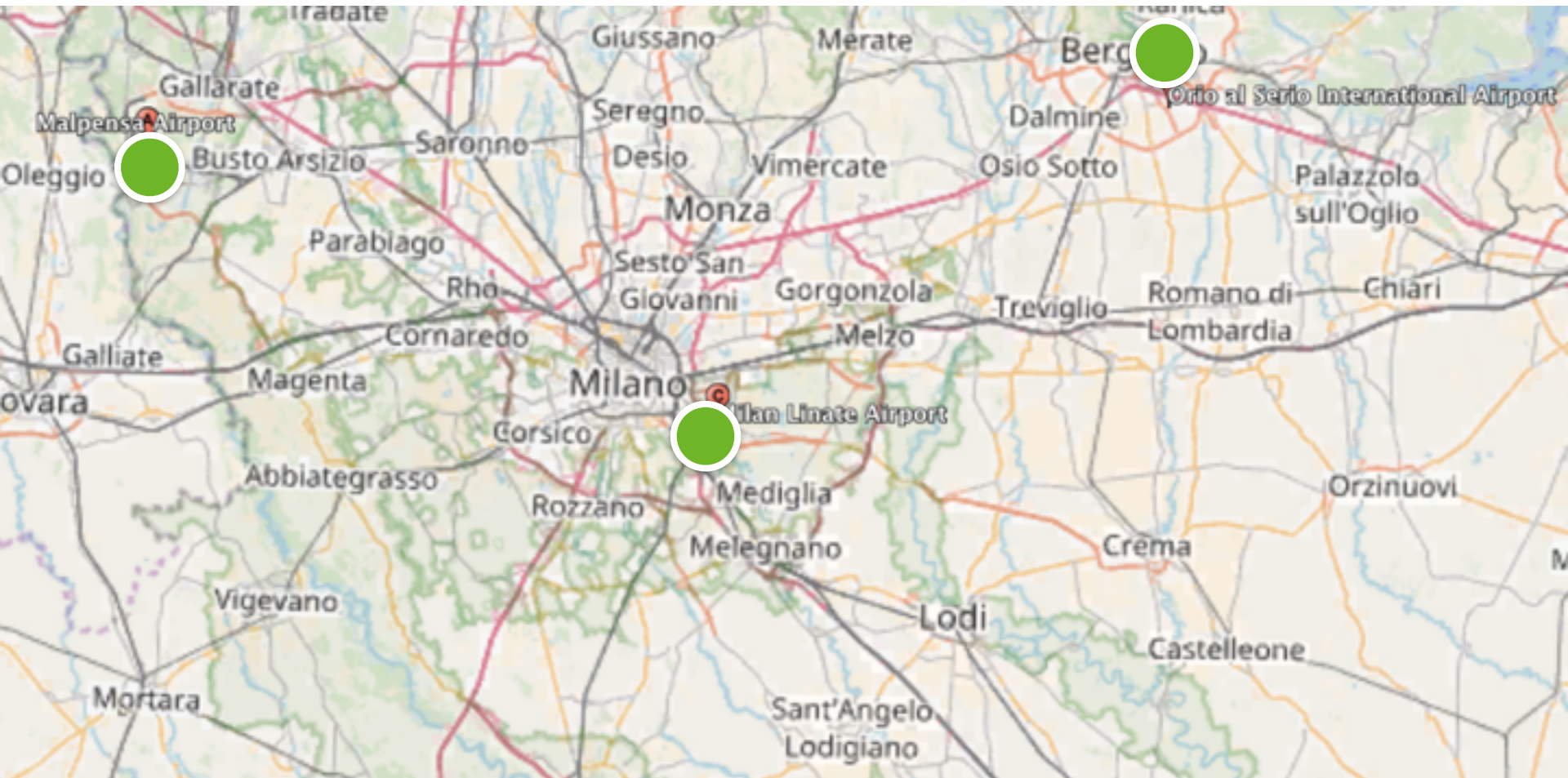
Disadvantages:

- non targeted to the system in analysis
- need for testing and new methods

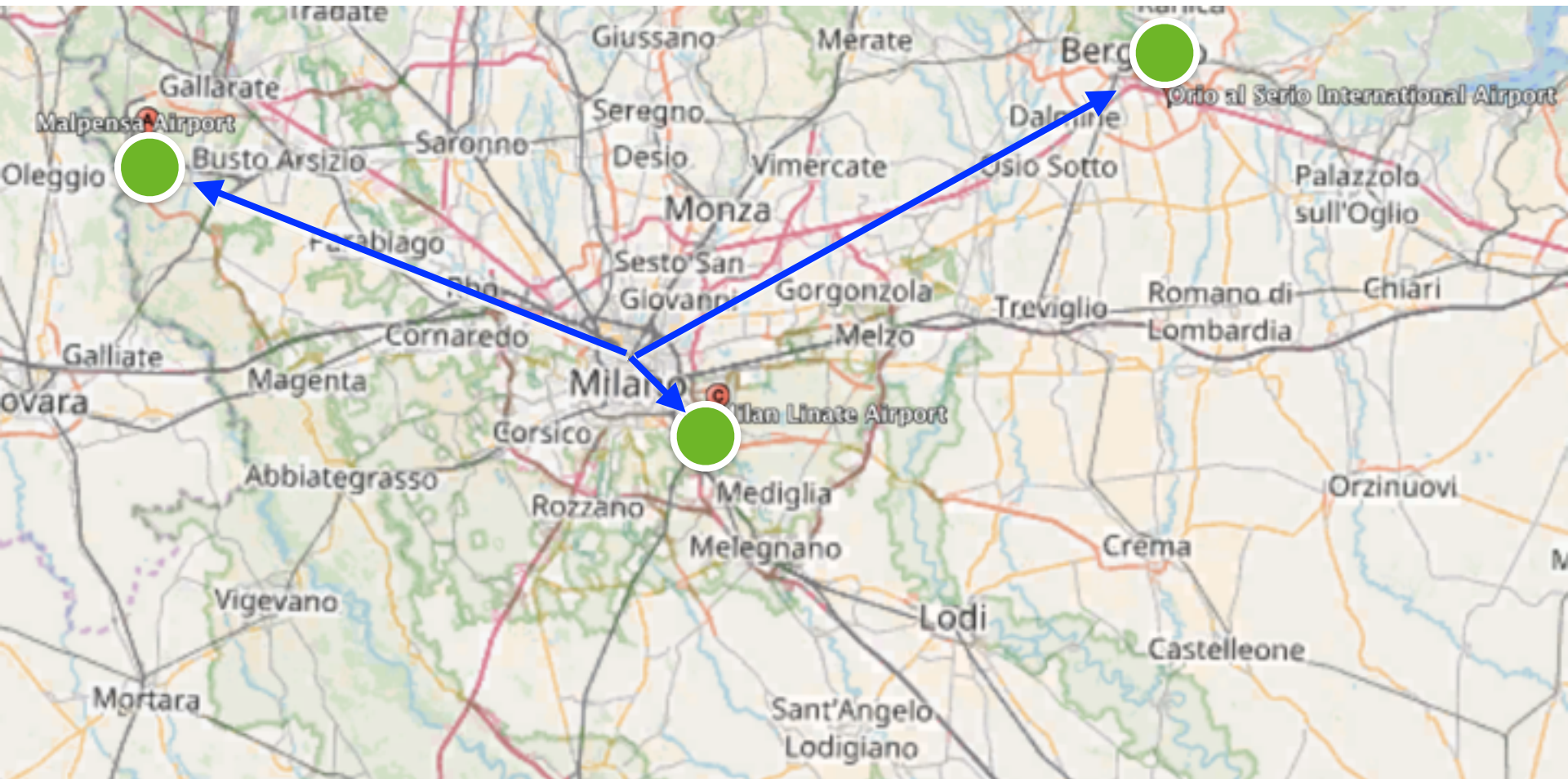
Competing airports



BigData4ATM



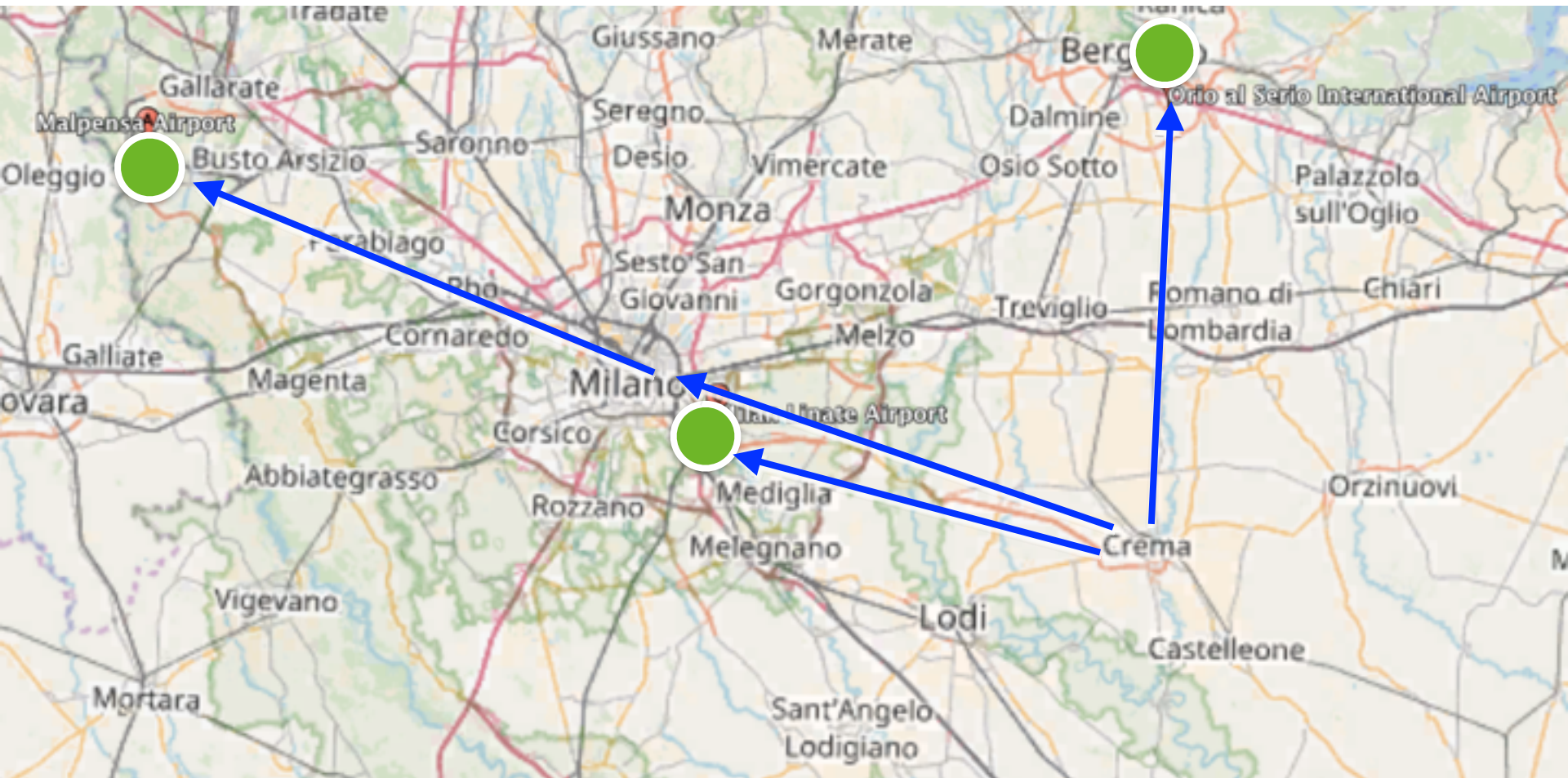
Competing airports



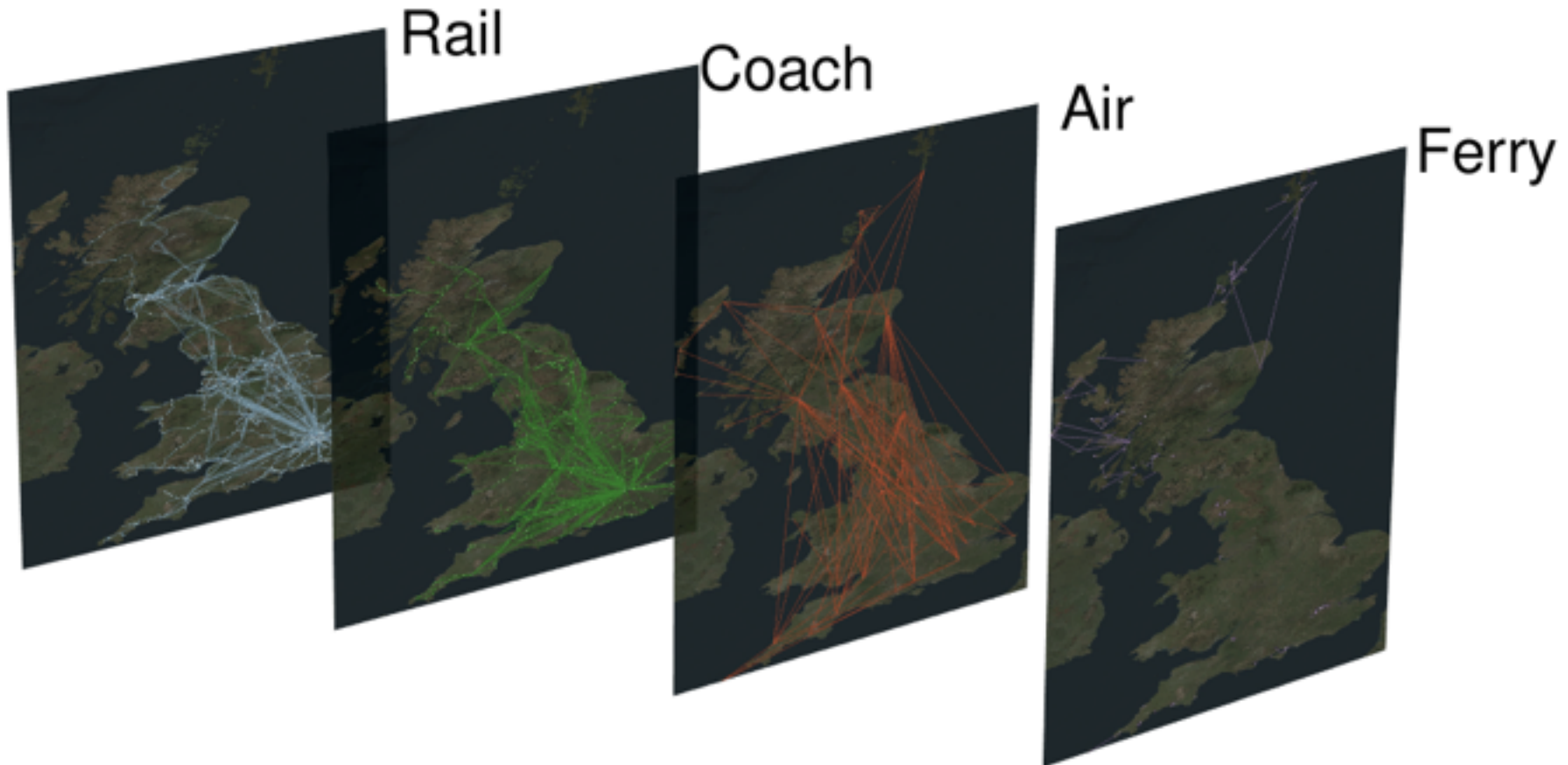
Competing airports



BigData4ATM



Study the interactions between different modes of transport



Gallotti and Barthelemy, 2014.

Three case studies with open-access data

Taxi Pickups in NYC



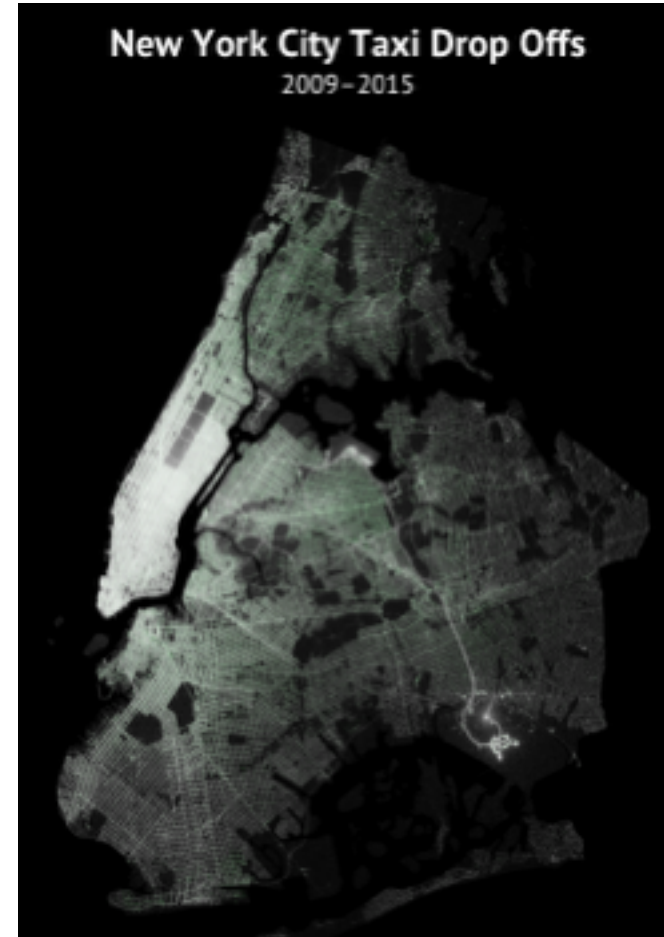
Geo-located tweets in London



Google Maps travel-times in London and Paris



Case study #1: Taxi Pickups



<http://toddwschneider.com>

Case study #1: Taxi Pickups



http://www.nyc.gov/html/tlc/html/about/trip_record_data.shtml
(just 1-click away from raw data)

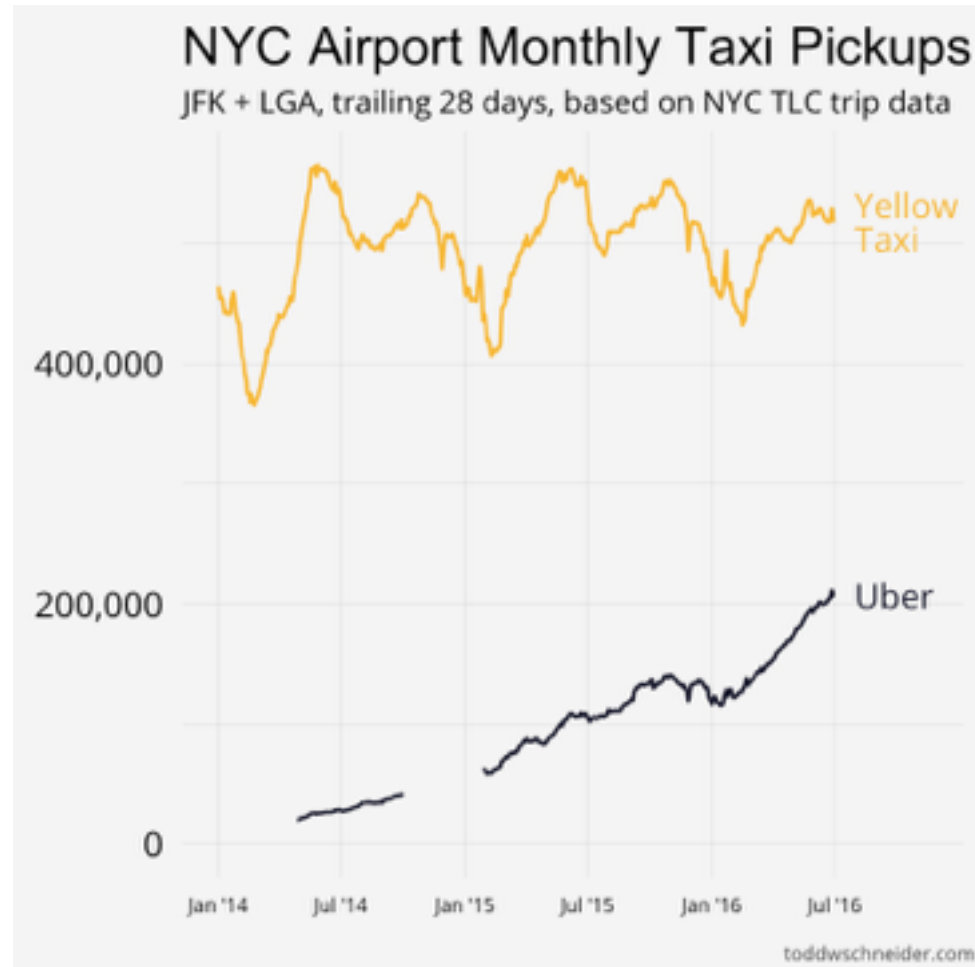
Trip Sheet Data (CSV Format)

▼ 2017

January	Yellow	Green	FHV
February	Yellow	Green	FHV
March	Yellow	Green	FHV
April	Yellow	Green	FHV
May	Yellow	Green	FHV
June	Yellow	Green	FHV
July			
August			
September			
October			
November			
December			

Case study #1: Taxi Pickups

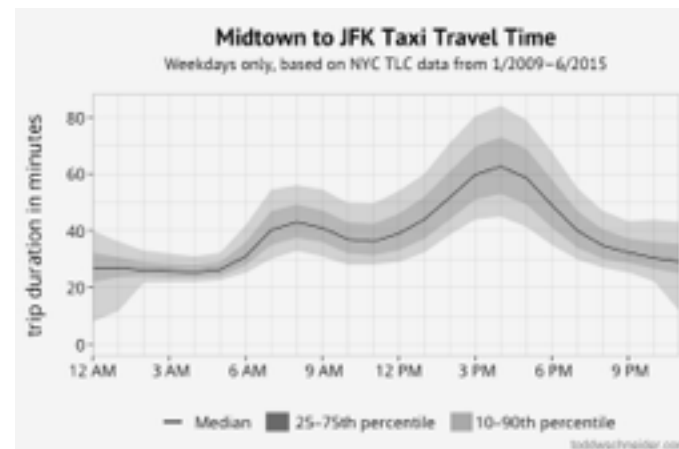
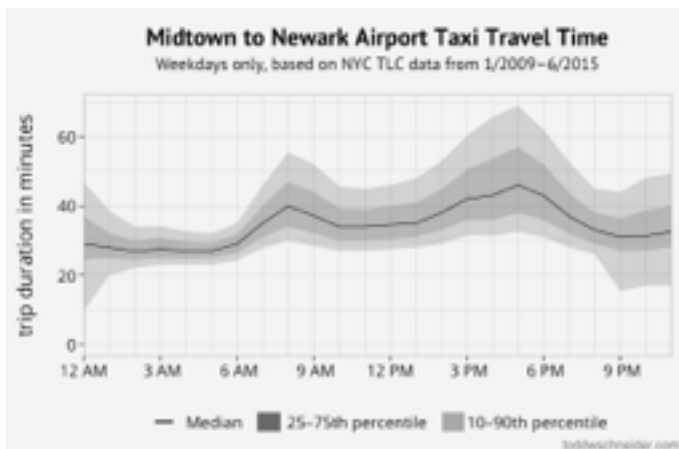
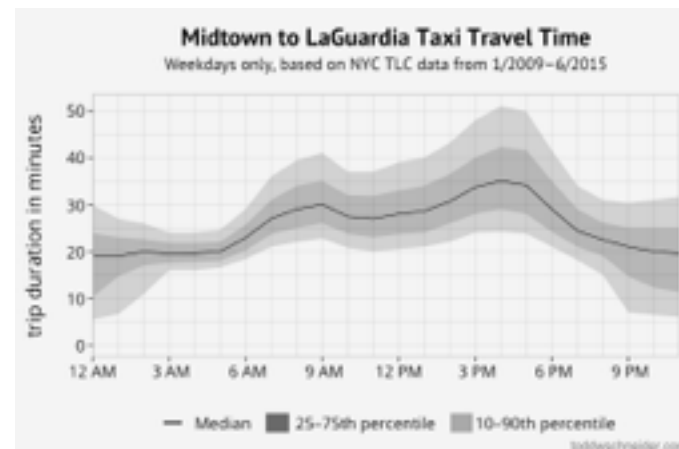
Example: the
rise of Uber



<http://toddwschneider.com>

Case study #1: Taxi Pickups

Example:
travel-times
from midtown

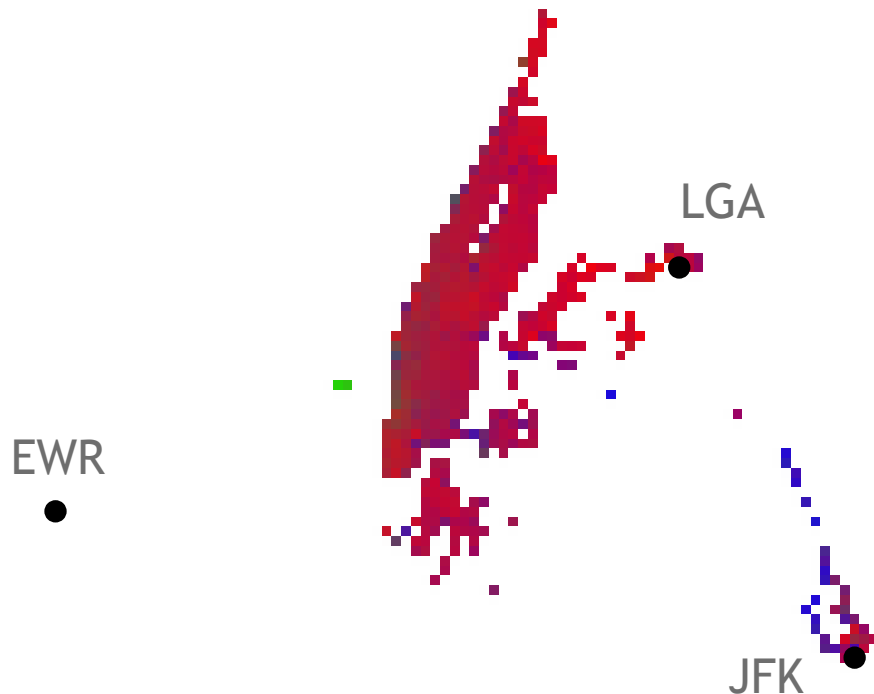


<http://toddwschneider.com>

Case study #1: Taxi Pickups

Real data on passengers choices

Red:LGA, Green:EWR, Blue:JFK



- we only have the pickups in within the NY state
- we map where occurred the pickups with an airport as destination
- we know travel duration and costs, including extra fares

LGA is the most frequent destination
(closer and not connected by train)

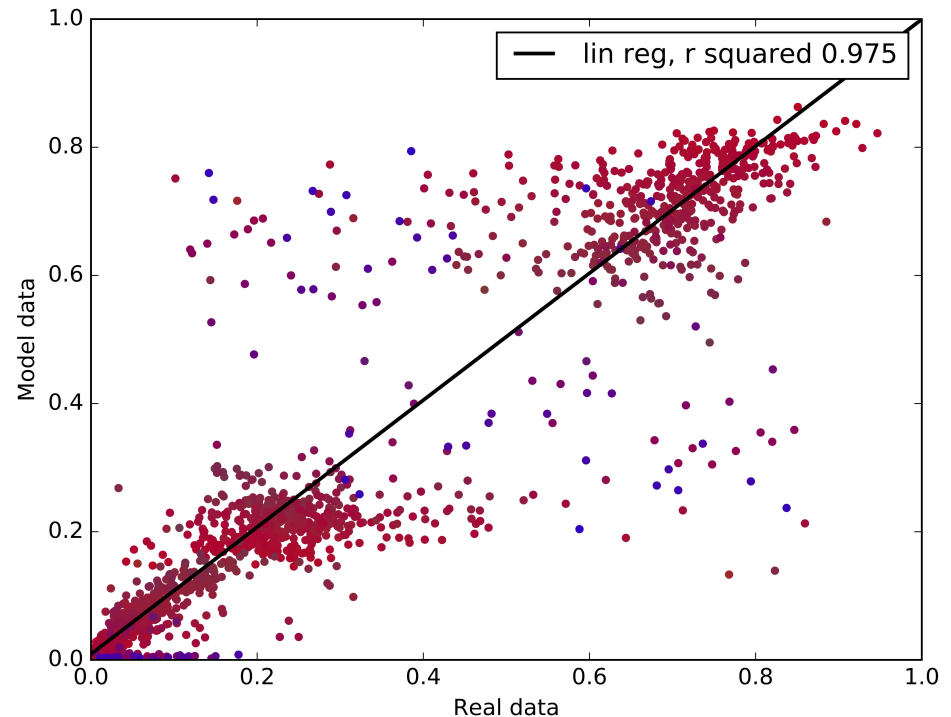
Case study #1: Taxi Pickups

For modelling, we only consider
time and cost of the travel to the
airport



$$C_i(a) = c_i(a) + V_T t_i(a)$$

$$P_i(a) = \frac{\exp(-C_i(a)/k)}{\sum_i \exp(-C_i(a)/k)}$$

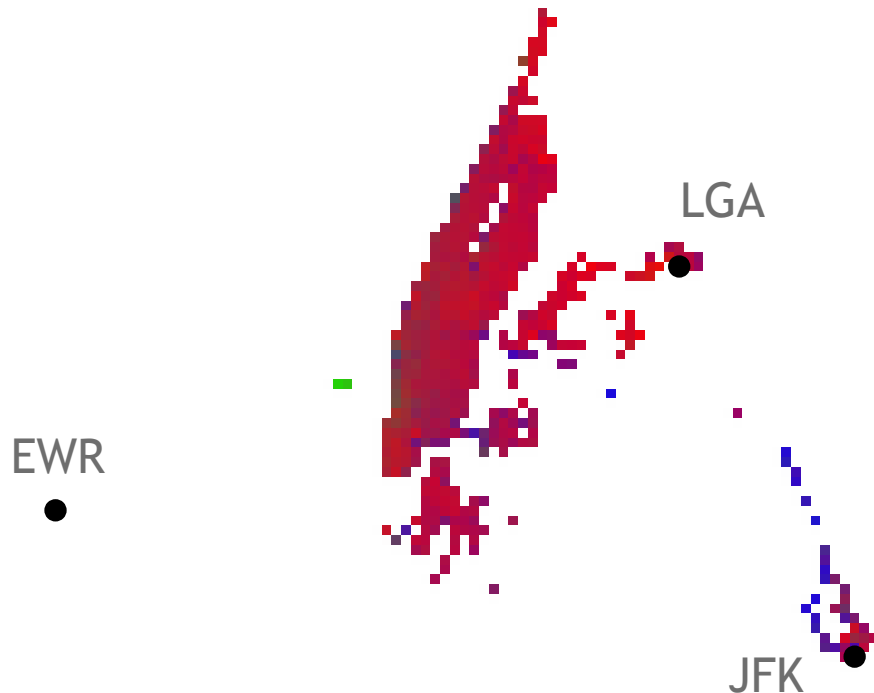


Machete model works well here!

Case study #1: Taxi Pickups

Real data on passengers choices

Red:LGA, Green:EWR, Blue:JFK

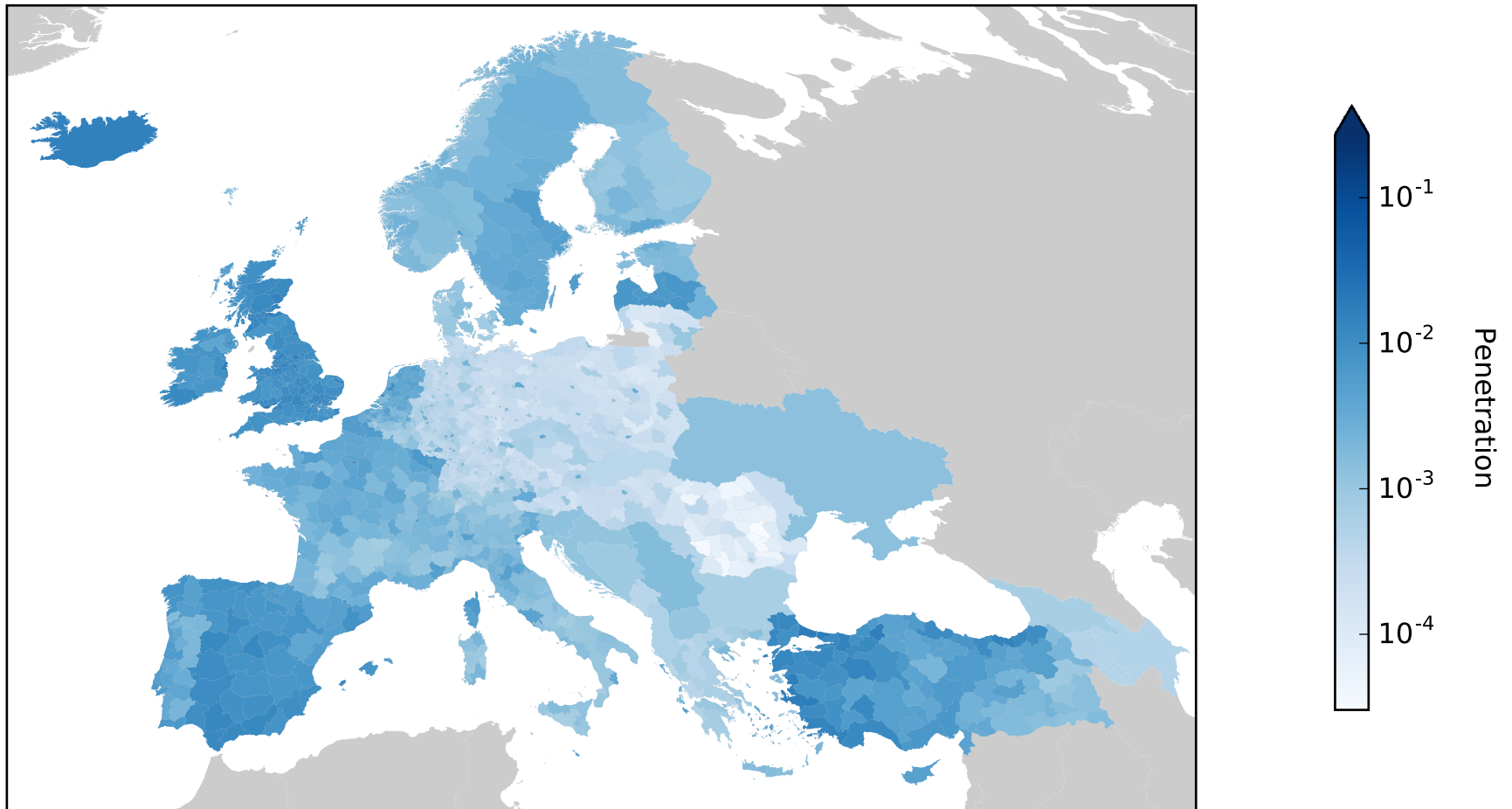


Airport decision with 0.35 USD per minute

Red:LGA, Green:EWR, Blue:JFK

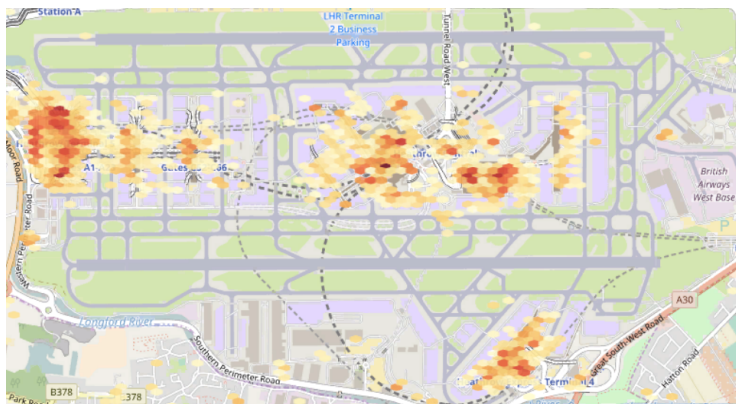
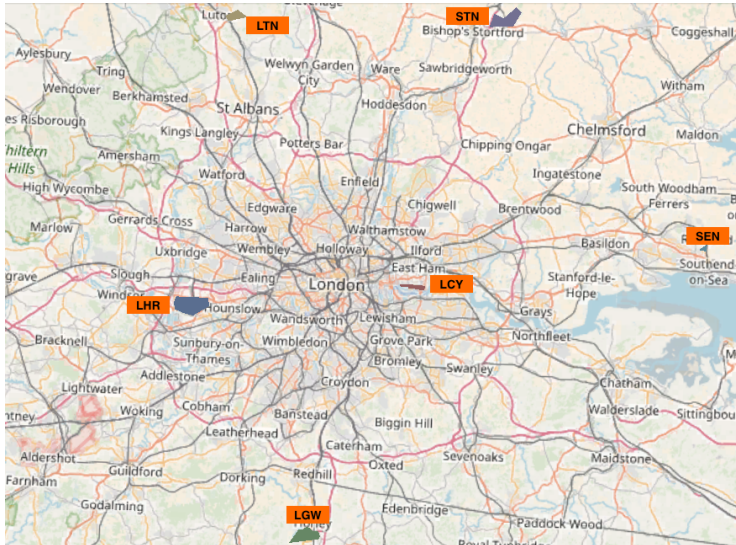


Case study #2: Geo-located tweets



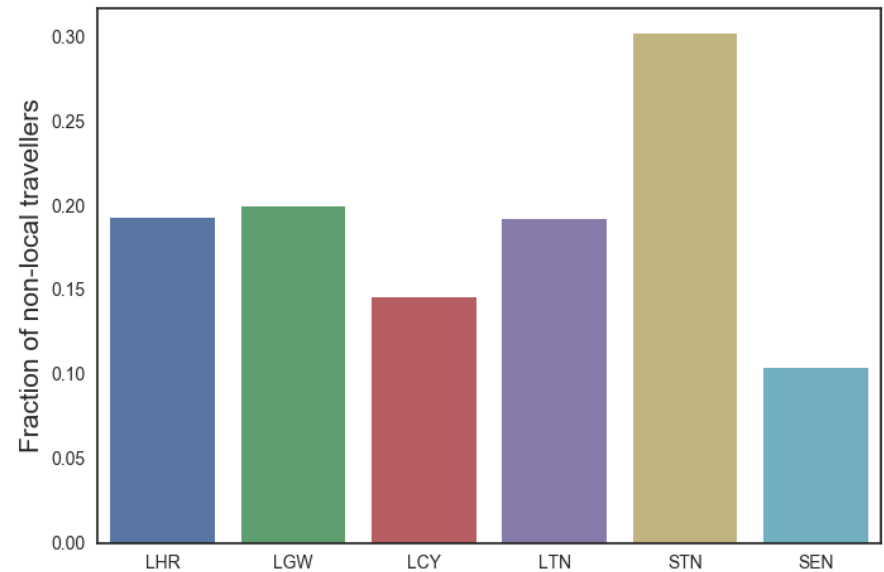
Case study #2: Geo-located tweets

Inside the airport



(mixing arrivals and departures)

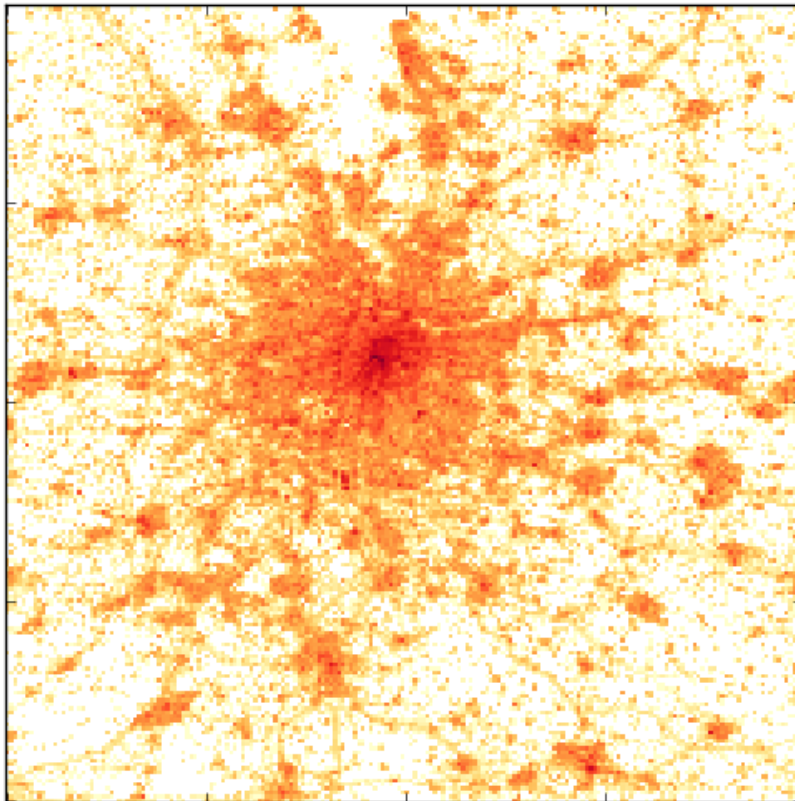
Fraction of tourists



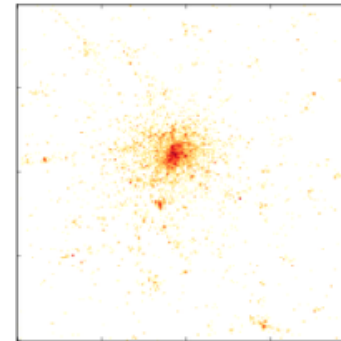
Case study #2: Geo-located tweets

Home/Main destination location for locals and tourists
(also not seen the airport)

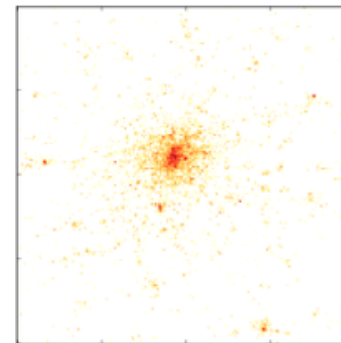
UK



FR



ES



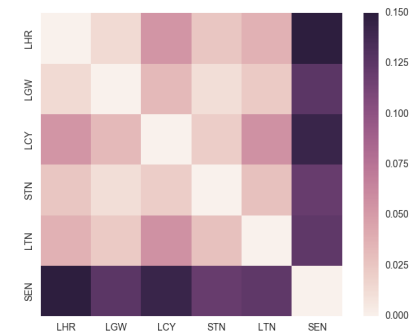
Case study #2: Geo-located tweets

Home location users seen in the different airports (UK)

LHR

LGW

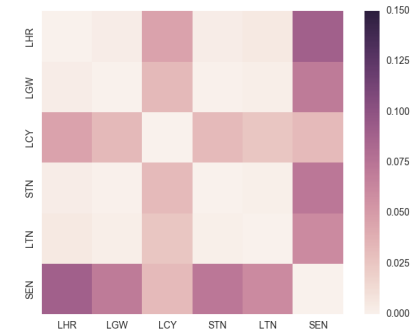
STN



LCY

LTN

SEN

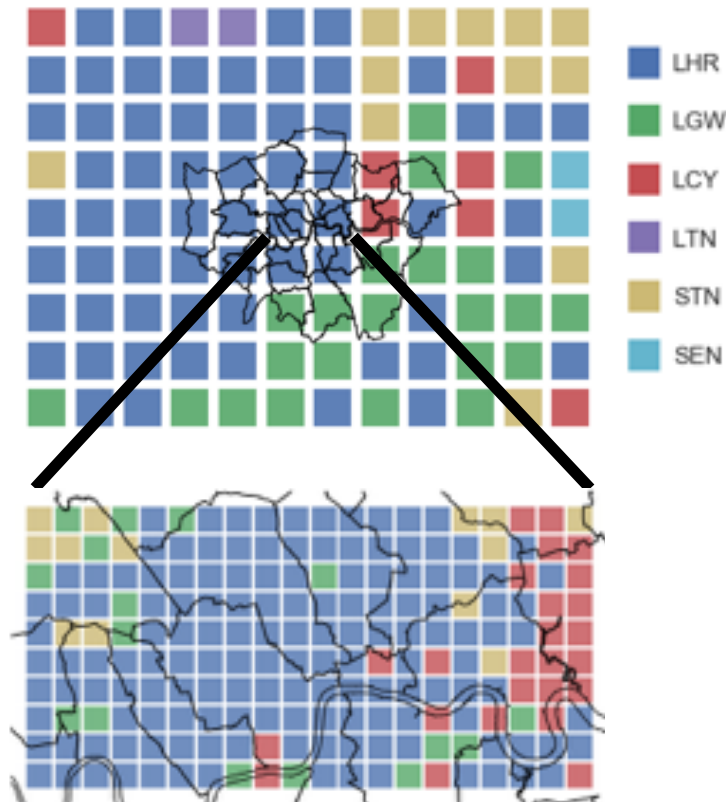


UK (up) show more
difference among airports

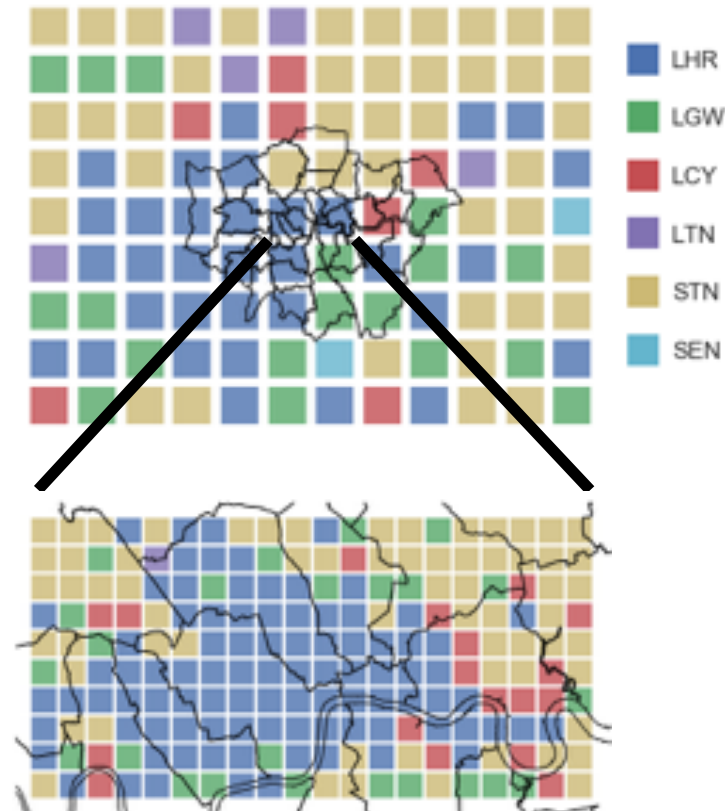
Case study #2: Geo-located tweets

Approximate the catchment areas: most frequent airport for each cell

Locals



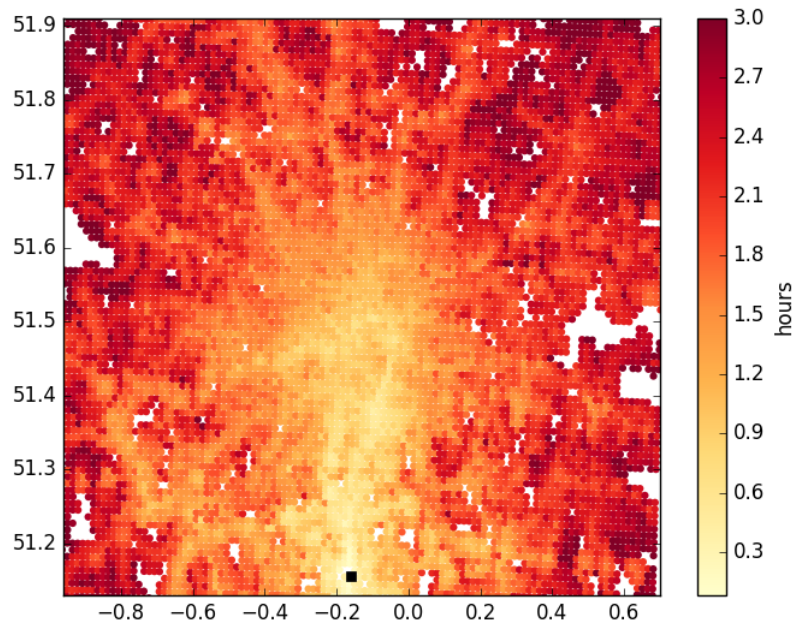
Tourists



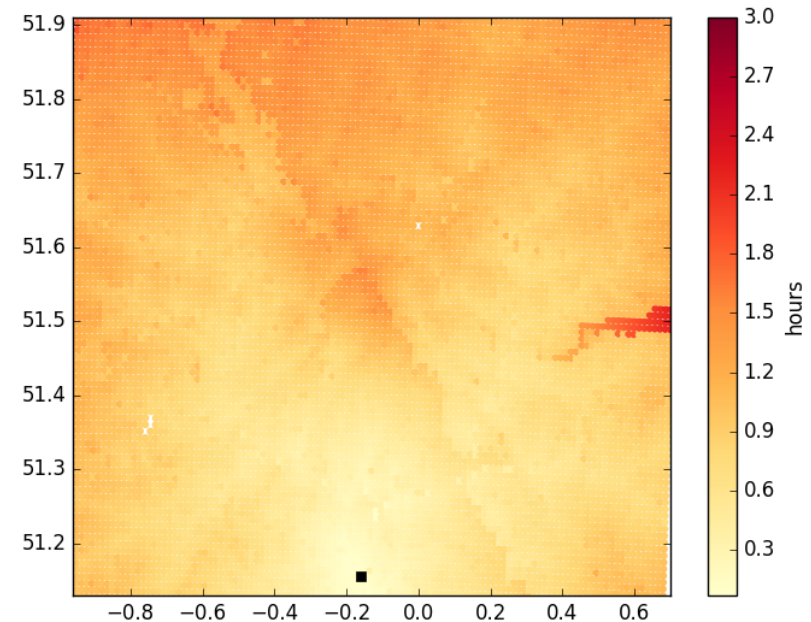
Outside the center, data is too scarce for higher resolution analysis

Case study #3: Google Maps travel-times

Transit to LGW



Driving to LGW

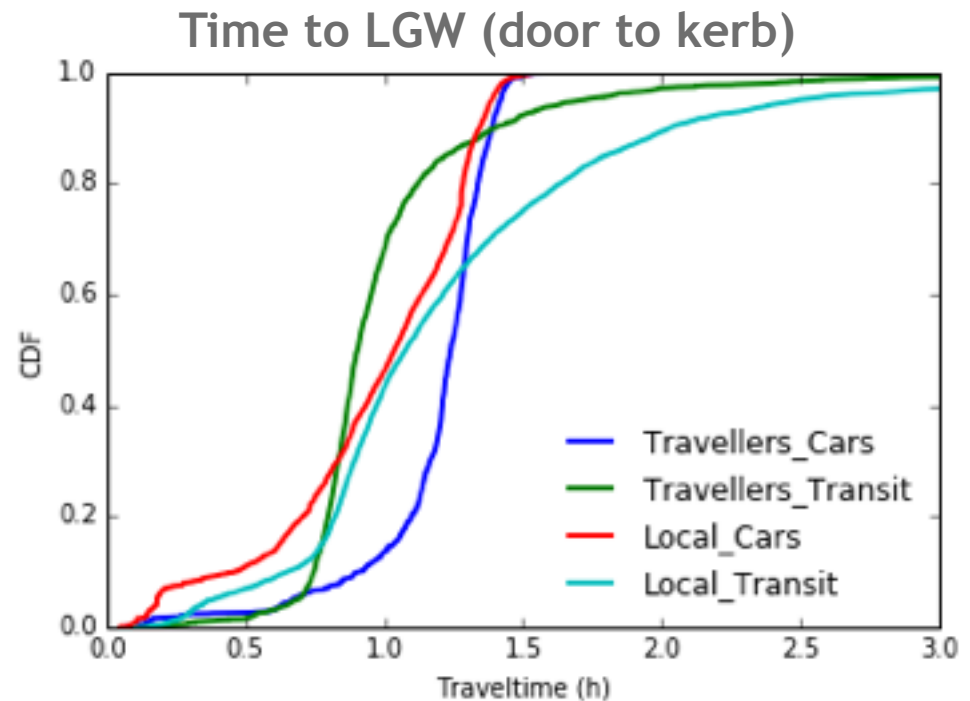
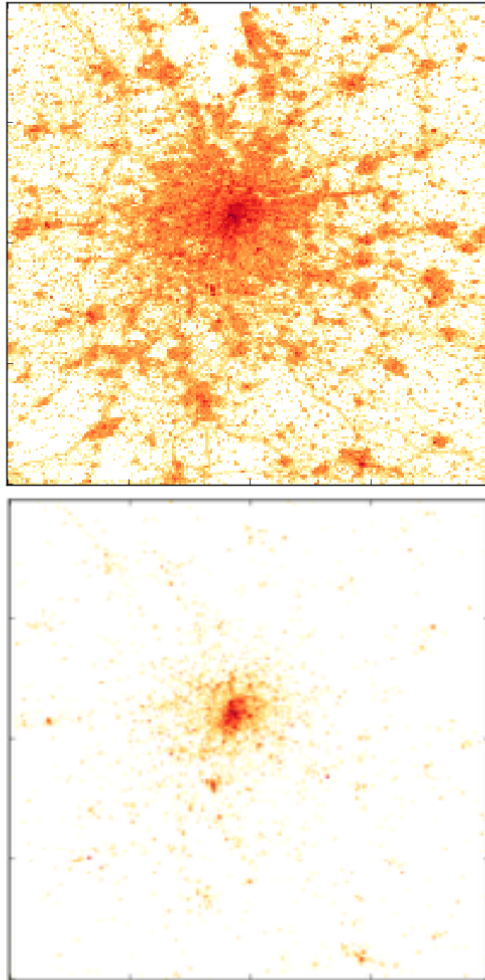


No hour dependence or congestion



Case study #3: Google Maps travel-times

And population distribution estimated from Twitter



Case study #3: Google Maps travel-times



BigData4ATM



Modelling approach for the choice of airport a for reaching destination b

- Fixed mode of transport m
- Ground costs = time t_{ij}
- Ticket prices $c(a, b)$ set as constant in time
- Choice of final destination independent on the residence area
- Value of time V_T constant across the population
- No alternative option for the final destinations



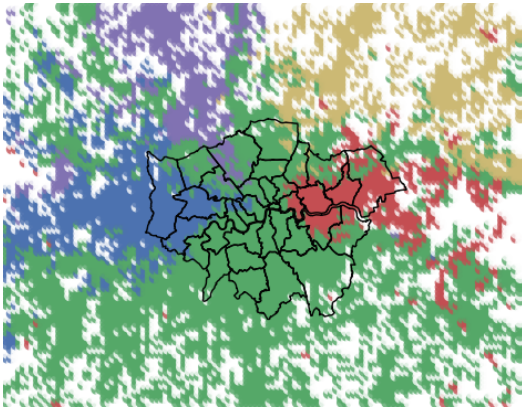
$$C_{ij}(a, b, m) = c(a, b) + V_T t_{ij}(a, m)$$

$$P_{ij}(a; b, m) = \frac{\exp(-C_{ij}(a, b, m)/k)}{\sum_i \exp(-C_{ij}(a, b, m)/k)}$$

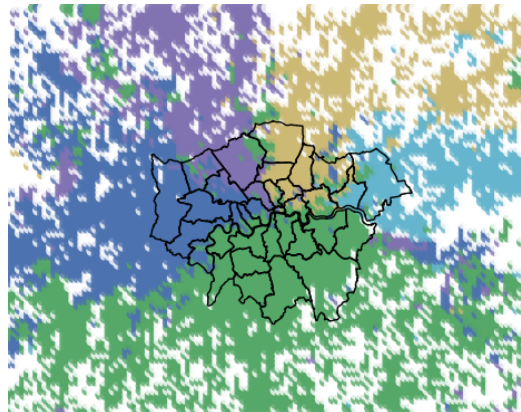
$$F(a, b, m) = \frac{\sum_{i,j} P_{ij}(a; b, m) Pop_{ij}}{\sum_{i,j} Pop_{ij}}$$

Case study #3: Google Maps travel-times

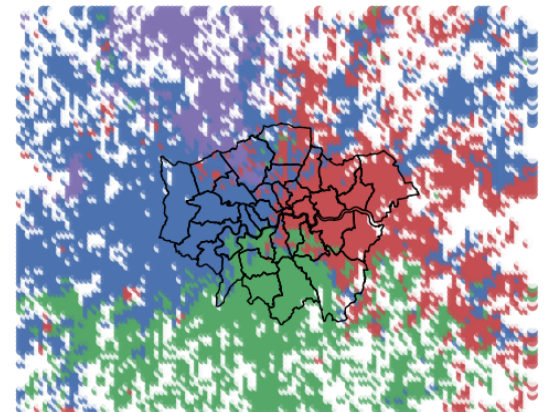
To MAD (transit)



To PMI (transit)



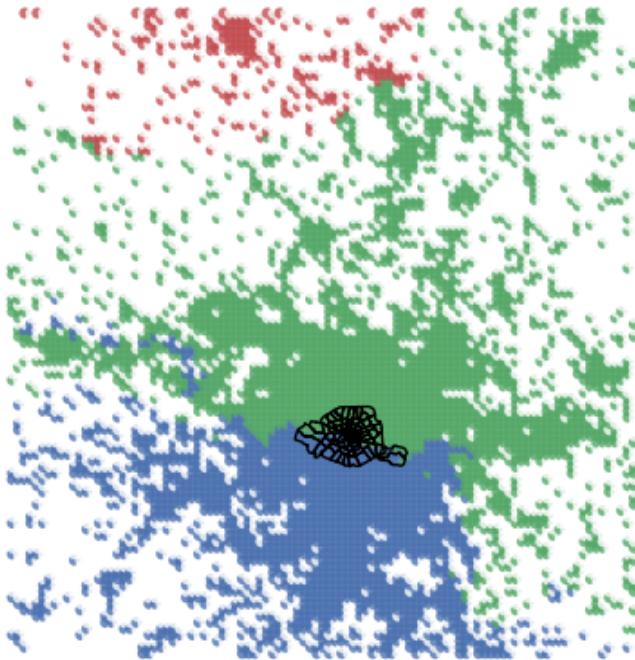
To ZRH (transit)



- $V_T = 150$ USD/h

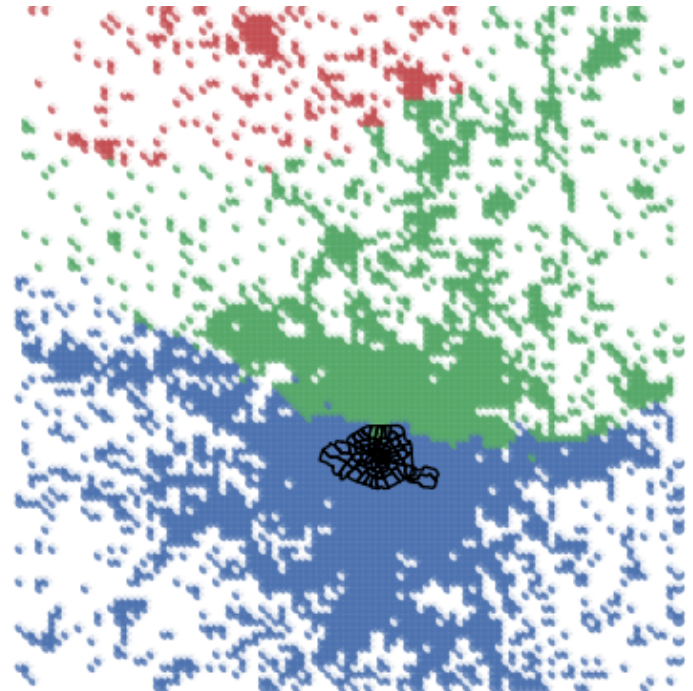
Case study #3: Google Maps travel-times

To MAD (driving)



ORY
CDG
BVA

To PMI (driving)



- $V_T = 190$ USD/h

Conclusions

- The spread of ICT technologies opens new research avenues by offering new types of data not traditionally used.
- Some of this data is openly available and can be easily accessed. Availability, accessibility, and granularity is destined to improve in the future.
- Even with some drastic simplification, we have been able to use Taxi, Twitter and Google Maps data to illustrate the effect of the interaction between air transport system and other transportation modes.
- This path is ready to be followed by more research, using more parsimonious methodologies to integrate these new data-sources in the modelling of passenger's choice behavior.



Thank you very much for your attention!



This project has received funding from the SESAR Joint Undertaking under the European Union's Horizon 2020 research and innovation programme under grant agreement No [699260]



Founding Members

