Geographic Data Science

Data Science Lecture I:

(New) Spatial Data

Dani Arribas-Bel

t e r d a y "

a revolution

ience to make sense of it all

d a y

quick) refresher
of (spatial) data

les

d) spatial data

ld) spatial data

stics (+ and -)

rces of (spatial)
ata

rces of (spatial) ata

olution

ame categories (polygons,

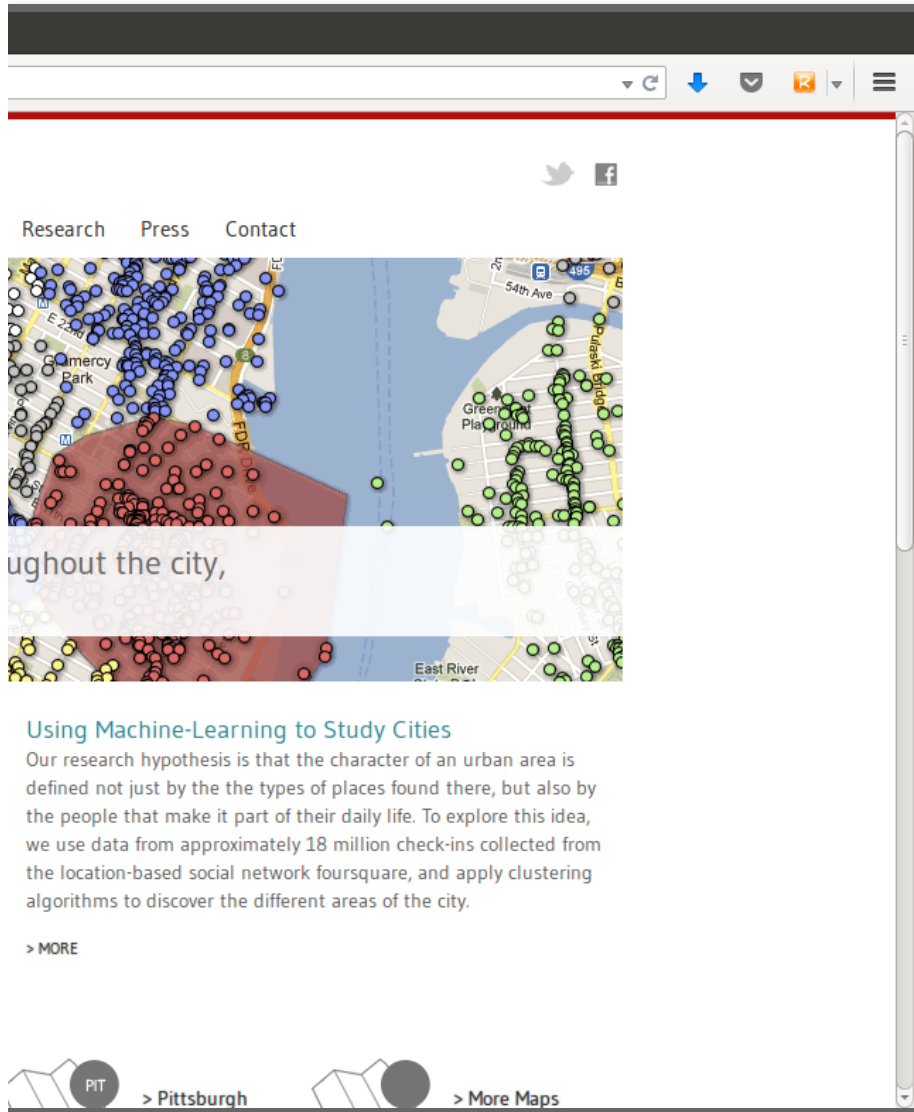t nature

ginated:

sensors"

sinesses/businesses going

ment Data

s as sensors

despread adoption of sensors
s...)
man life is subject to leave a
ected, stored and analyzed
/ data creators (sensors,

' --> Most of it (80%?) has
sion

Livehoods

ughout the city,

Using Machine-Learning to Study Cities

Our research hypothesis is that the character of an urban area is defined not just by the the types of places found there, but also by the people that make it part of their daily life. To explore this idea, we use data from approximately 18 million check-ins collected from the location-based social network foursquare, and apply clustering algorithms to discover the different areas of the city.

> MORE

PIT   > Pittsburgh   > More Maps

moving online

arts of bussiness activities
the last decades
hange in the final product or
gital data is "available" about

siness activities have been
chnologies ("internet natives")
researchers better

Walkscore

Sutro Baths

Presidio San Francisco

San Francisco is the 2nd most walkable large city in the US with 805,235 residents.

San Francisco has excellent public transportation and is very bikeable.

Find apartments in San Francisco's most walkable neighborhoods: Chinatown, Financial District and Downtown.

San Francisco Homes for Sale

$3,750 and the average home price is $1,099,999. [?]

ata for open
rnments

se (part of) their internal data

Shadbolt, 2010):

d accountability
cial value
provement
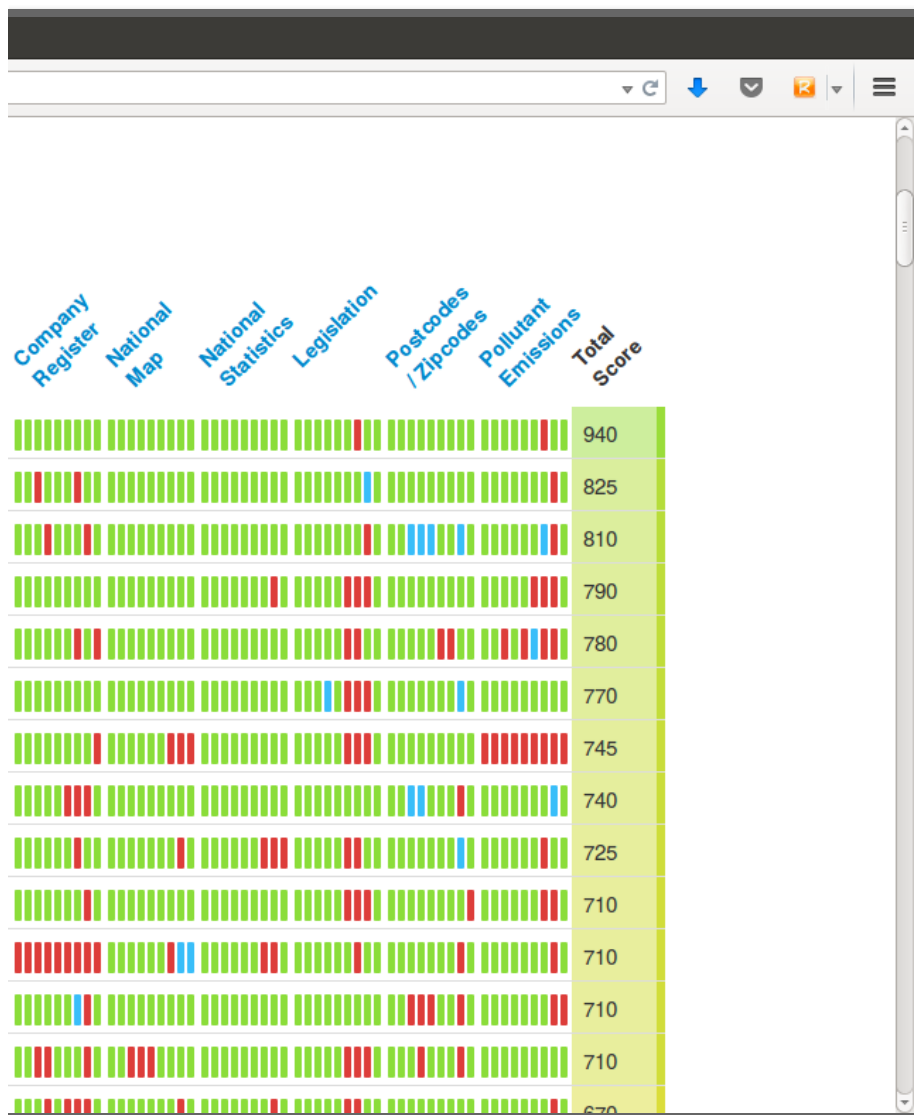industries and jobs

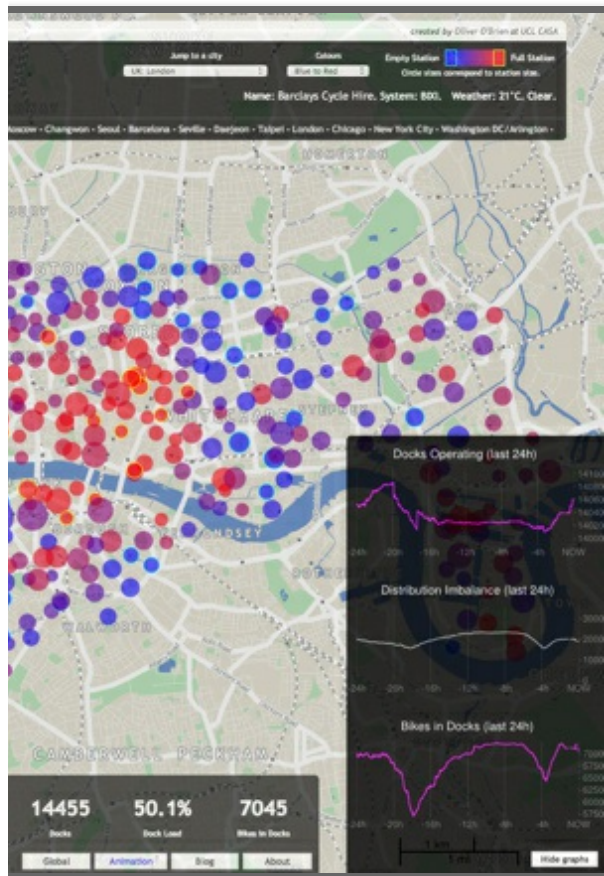pen Data Index
)14

Company Register	National Map	National Statistics	Legislation	Postcodes / Zipcodes	Pollutant Emissions	Total Score
						940
						825
						810
						790
						780
						770
						745
						740
						725
						710
						710
						710
						710
						670

keShare Map

ture

s Quiz

he origin of the following

ets
e transaction values
rant listing
dices
e service station status

uiz - Answers

ottom-up
tion values --> Open

ing --> Digital businesses
> Traditional (not accidental!)
station status --> Open

llenges

i a s

urban researchers meets some
tativity, accuracy…)
new data sources will not

ed to have extra care and put
clusions they can reach from
of data
run in favour of researchers,
en for granted

barriers to access

lable
ature makes them not be

kills is required to tap into

erstanding of the internet,

a. "go beyond Word and

Methods

exactly the same as that of

example:

ns Vs. Points

uency): Decadal Vs. Real-time

ell" with techniques employed

Geography.

nsight as possible from these

techniques from other

ones

ine learning (but also others

tance)

- Visualization

aries

aphy, but more emphasis

lore the data) and

ories with data)

ston

ls - Machine
rning

nce, blended with statistics

tern recognition

omputer some true
model, then use the model to
diction is available (e.g.

ta speak"... and the machine

Livehoods)

d？ NewOld！

both worlds
entary