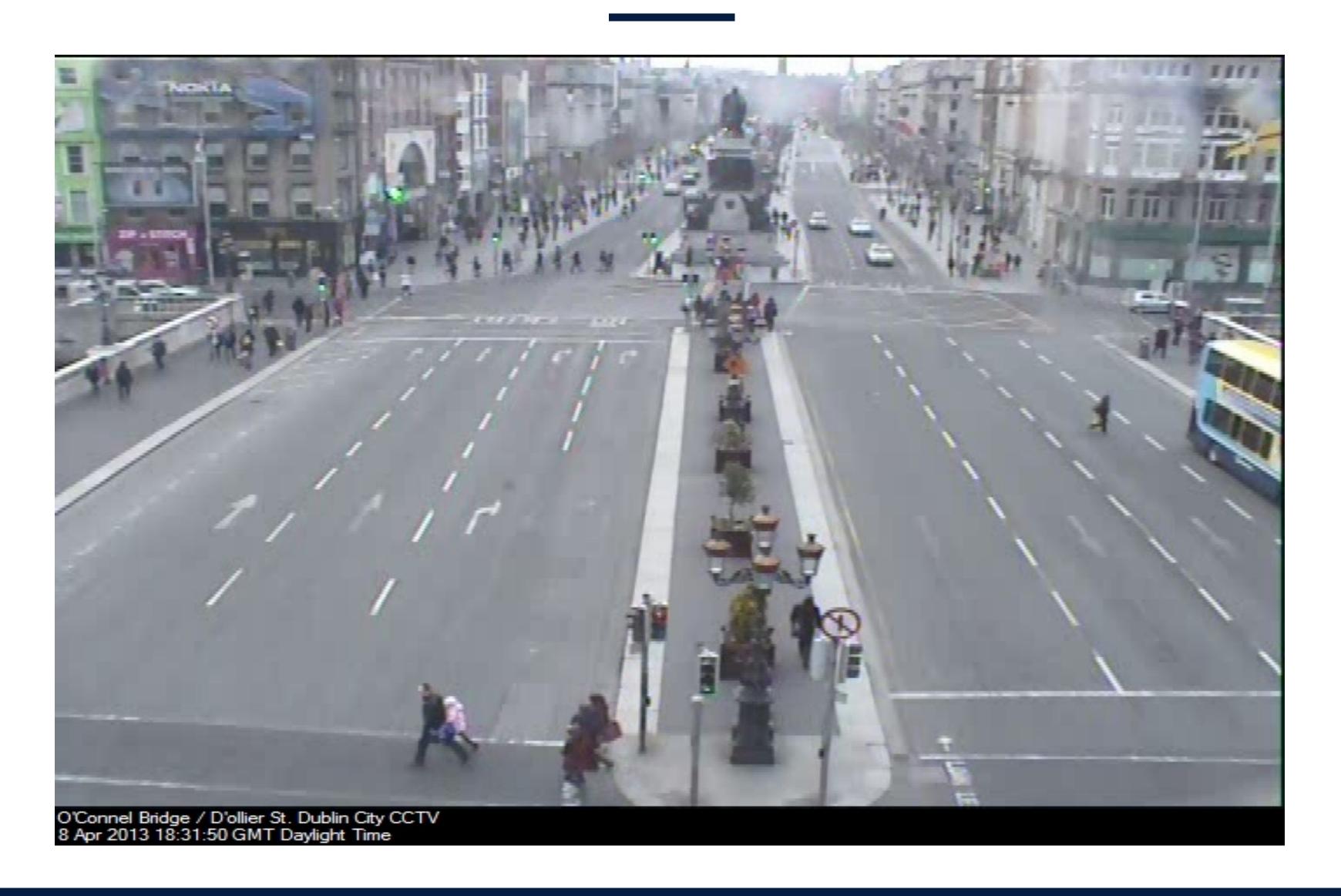




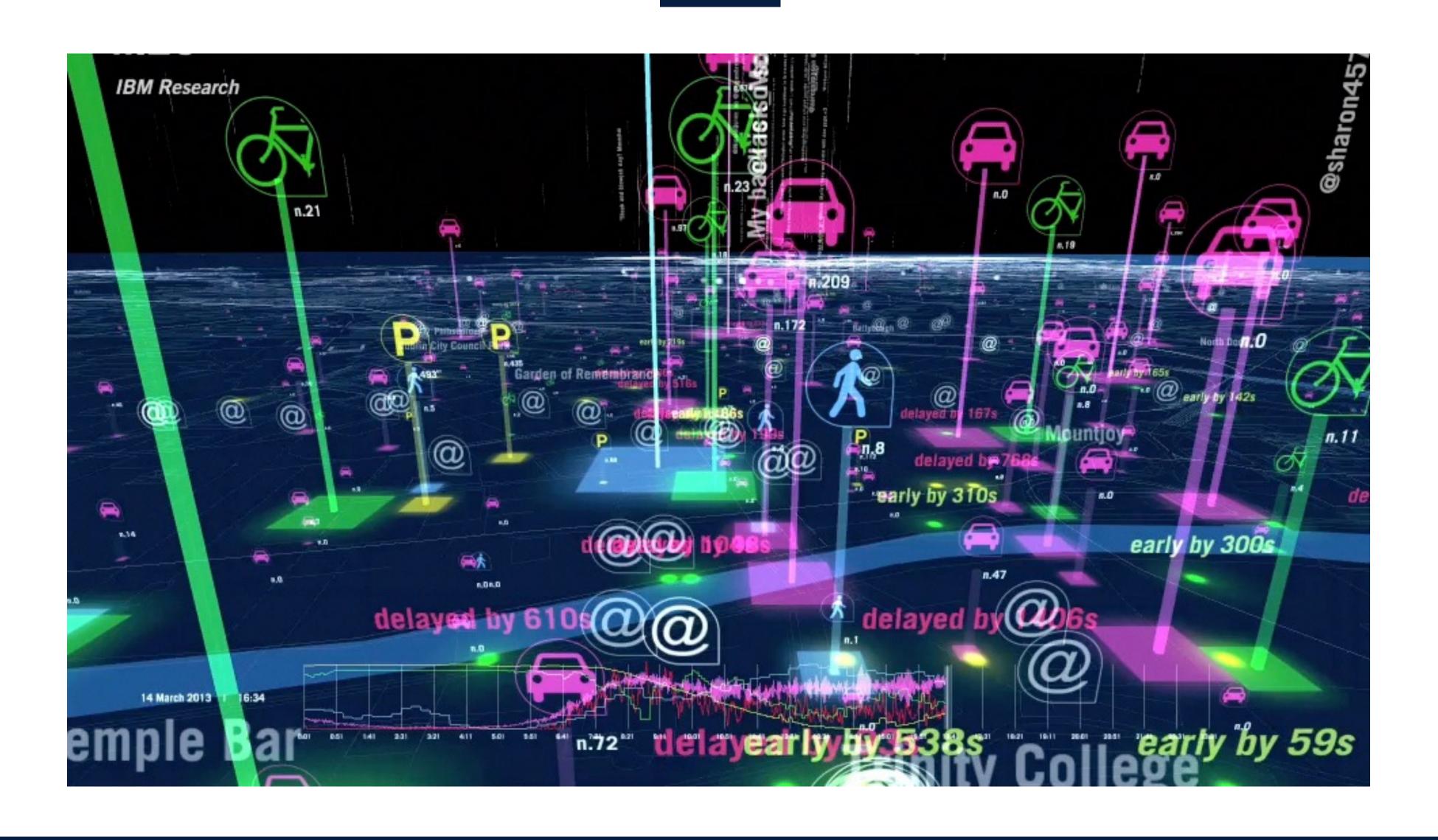
Data + Algorithms -> Knowledge -> Benefits

Data wrangling (Pandas, Numpy) Machine Learning (scikit-learn)

# What do you see?



### A fantastic source of data





### Data is the new oil



201 May Economist,

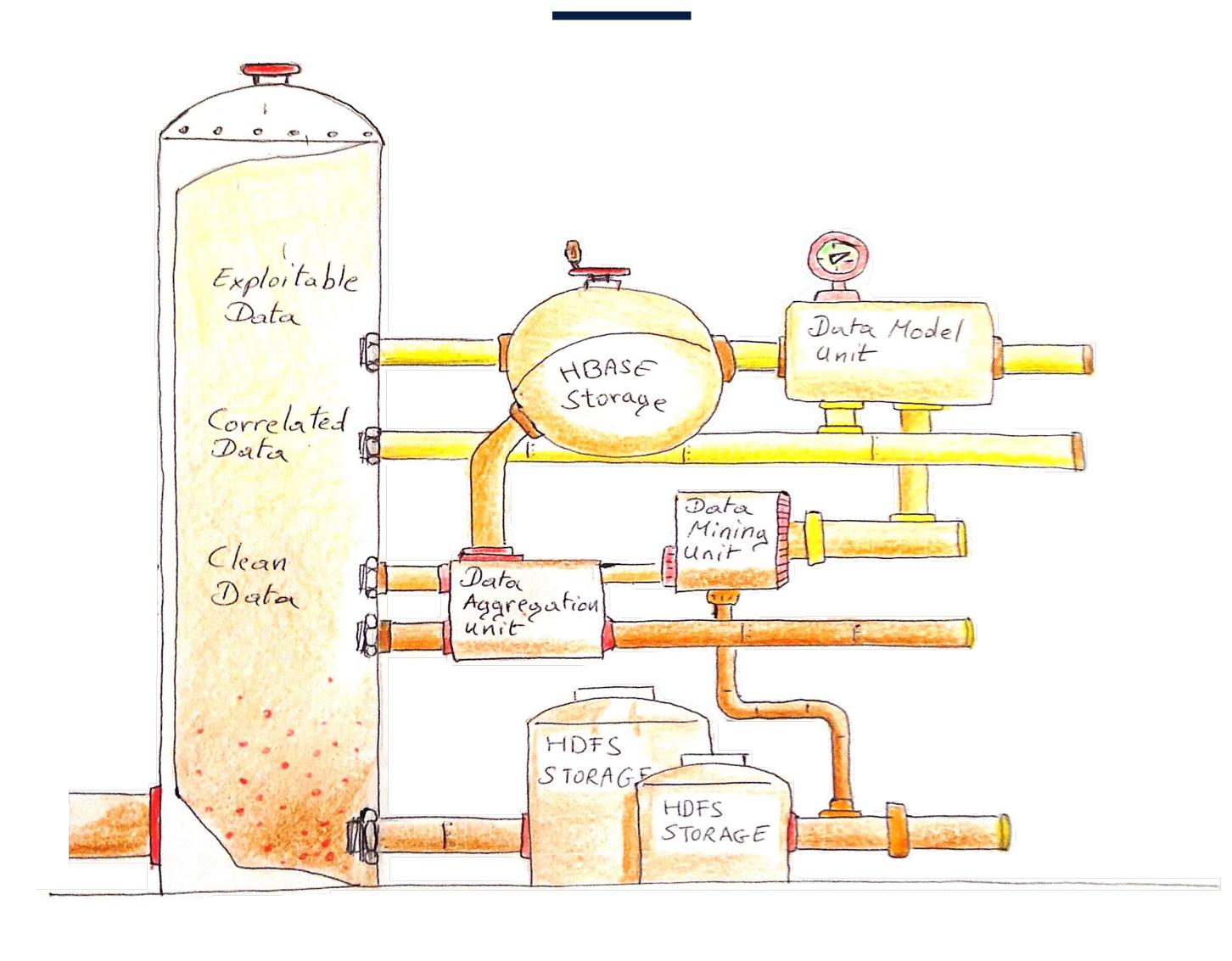
### Big data, Bad data



BY LOWE FOR THE SUN-SENTINEL, FLOR



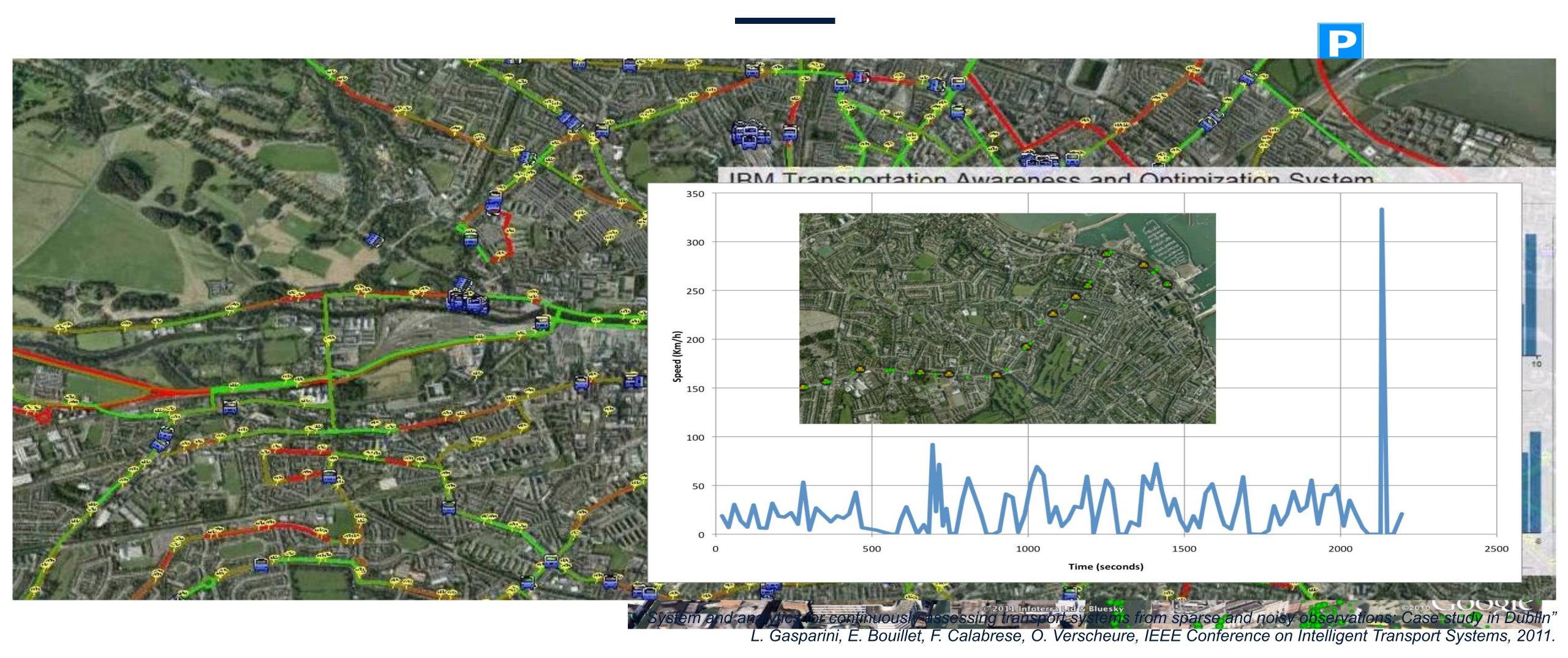
## Like oil, data must be refined





### From raw data to information

Real-time tool for situational awareness



Data + Algorithms -> Knowledge -> Benefits

Data wrangling (Pandas, Numpy) Machine Learning (scikit-learn)

# How can data science improve Urban Mobility?





#### 1. From raw data to information

- Assimilate complex sensor data
- Deal with volume, accuracy, diversity

#### 2. From information to models

- Understand how people move
- Model and predict demand

# How can data science improve Urban Mobility?





#### 1. From raw data to information

- Assimilate complex sensor data
- Deal with volume, accuracy, diversity

#### 2. From information to models

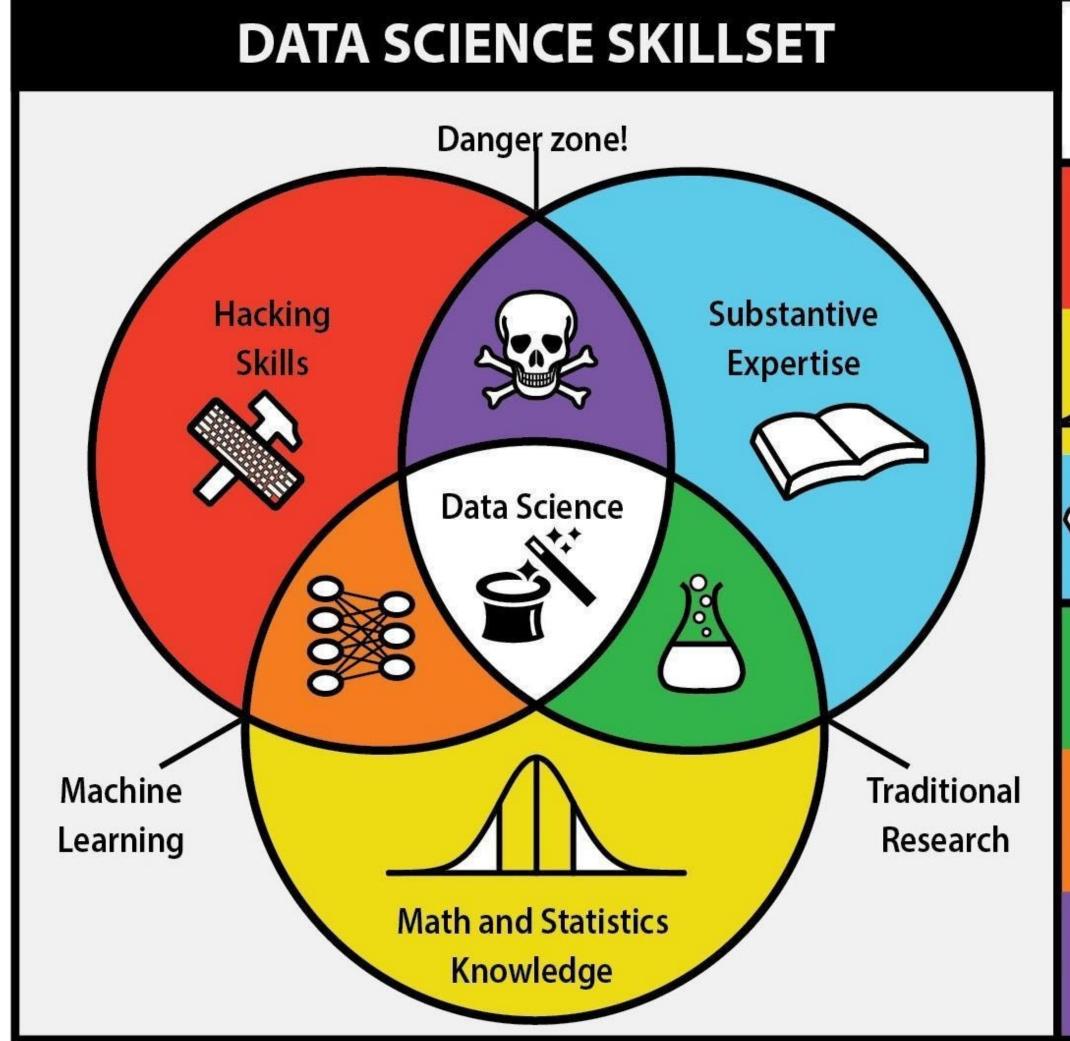
- Understand how people move
- Model and predict demand

#### 3. From models to actions

- Optimize planning & operations
- Deal with uncertainty



## From Machine Learning to Data Science





Data science, due to its interdisciplinary nature, requires an intersection of abilities: hacking skills, math and statistics knowledge, and substantive expertise in a field of science.



**Hacking skills** are necessary for working with massive amounts of electronic data that must be acquired, cleaned, and manipulated.



**Math and statistics knowledge** allows a data scientist to choose appropriate methods and tools in order to extract insight from data.



**Substantive expertise** in a scientific field is crucial for generating motivating questions and hypotheses and interpreting results.



**Traditional research** lies at the intersection of knowledge of math and statistics with substantive expertise in a scientific field.



**Machine learning** stems from combining hacking skills with math and statistics knowledge, but does not require scientific motivation.



**Danger zone!** Hacking skills combined with substantive scientific expertise without rigorous methods can beget incorrect analyses.

Domain expertise

Visualization (Matplotlib)

Data + Algorithms -> Knowledge -> Benefits

Data wrangling (Pandas, Numpy) Machine Learning (scikit-learn)

# QUESTIONS?



