

URBAN BIG DATA ANALYTICS

Course Introduction

July 18, 2018

Instructor: [Andy Hong, PhD](#)
Postdoctoral Research and Teaching Fellow
School of Population and Public Health
University of British Columbia

A BIT ABOUT ME

- ▶ Postdoctoral fellow in the School of Population and Public Health
- ▶ Co-founder of [Healthy Cities Network](#)
- ▶ Studied informatics, geography, and public policy
- ▶ Bike commuter and hiker
- ▶ A father of two kids



COURSE OVERVIEW

- Updated course syllabus (link)
- Assignments
- Group sessions
- Group project

What this course is

- Introduction to *urban* data science
- Learn to think and talk like data scientists
- Learn how to wrangle, explore, and analyze *data*

What this course is NOT

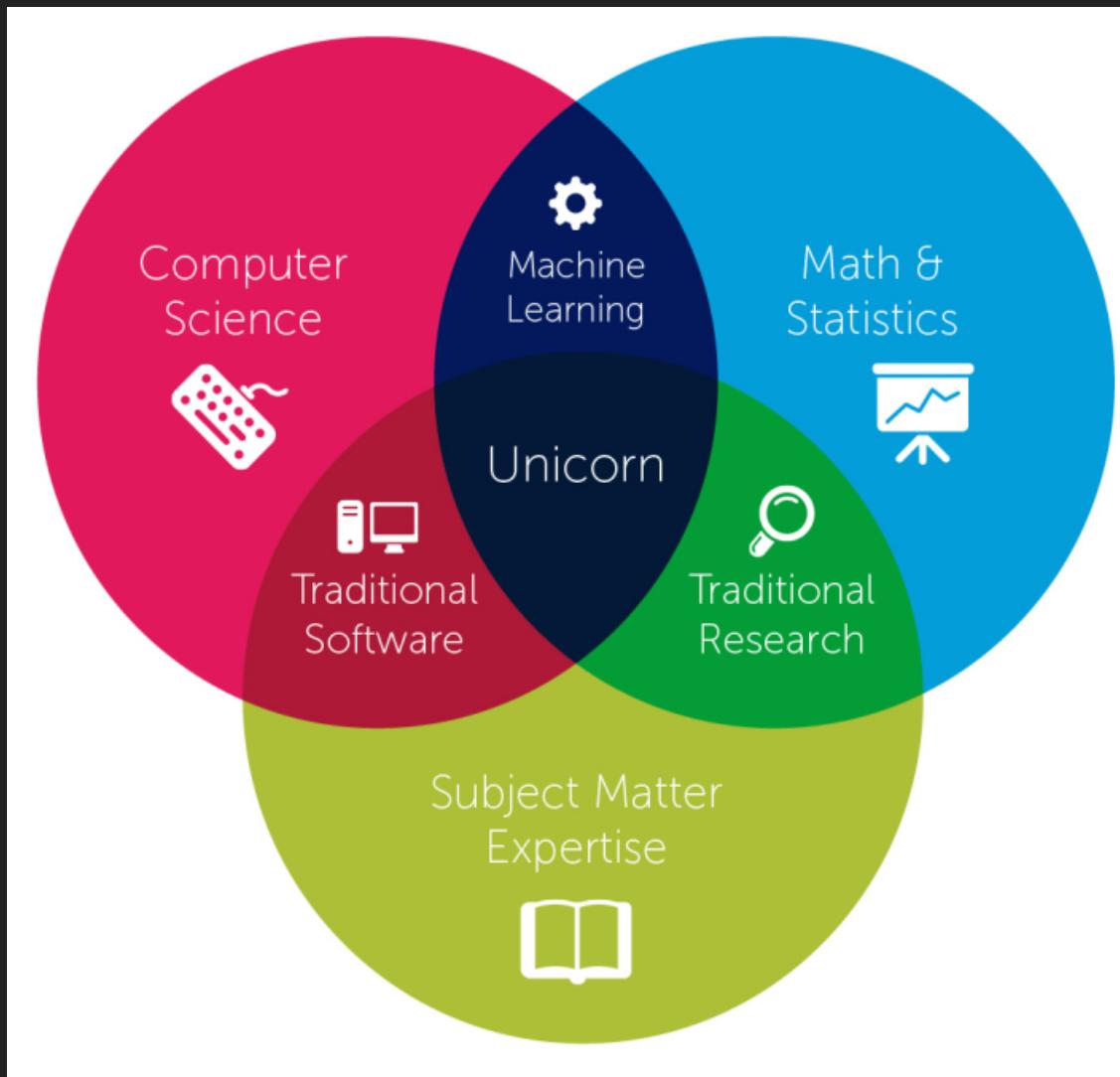
- Learn how to program
- Learn statistics and math
- Learn GIS and mapping skills

LET'S BREAK THE ICE

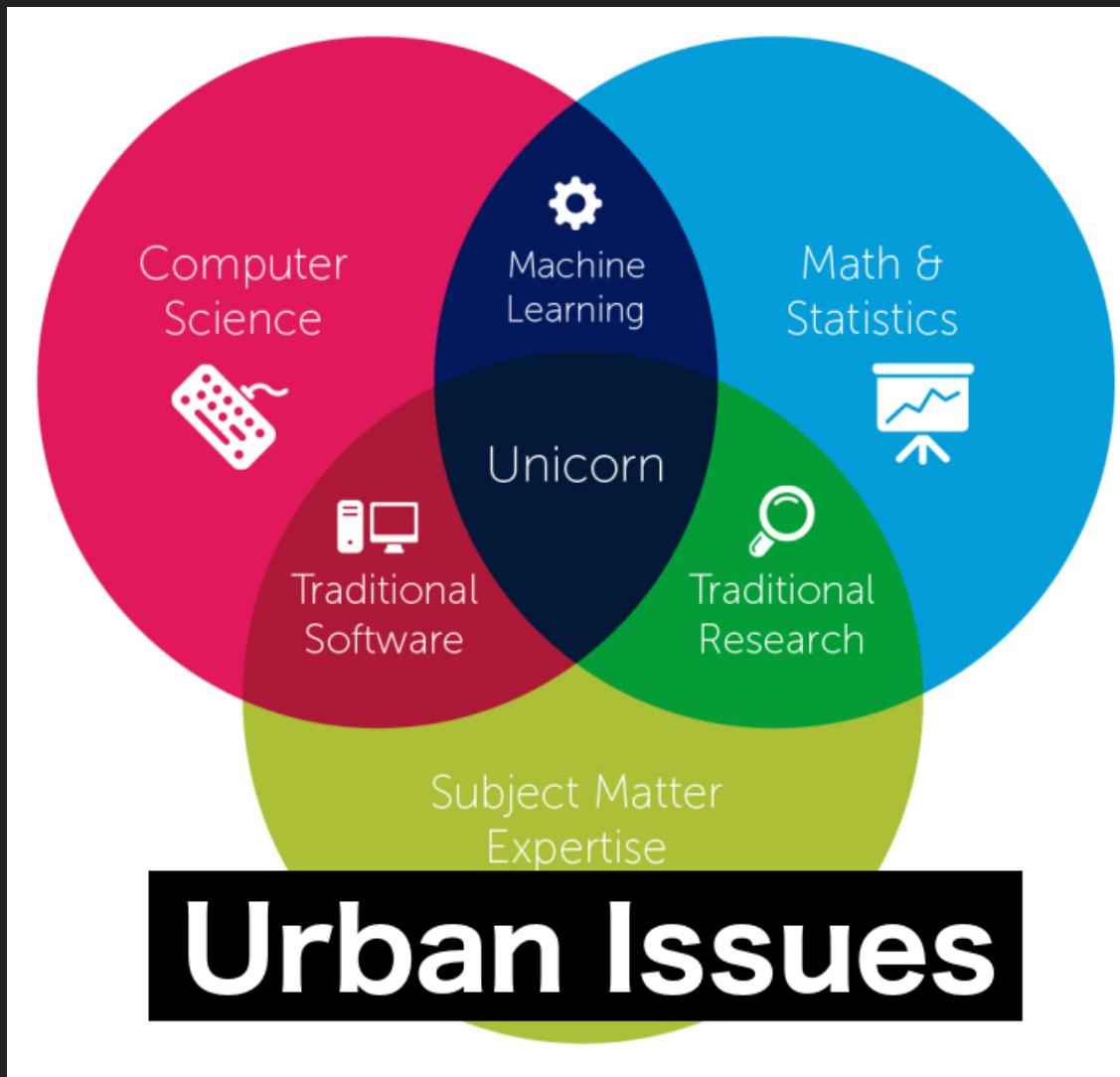
- Tell us your name and where you are from
- Your job or school major
- Why are you taking this course?
- How good is your coding skills? (from 1 to 5)



WHAT IS DATA SCIENCE?



WHAT IS URBAN DATA SCIENCE?



Question to
everyone



WHAT ARE URBAN ISSUES?

- Health issues: air pollution, noise pollution
- Environmental issues: climate change (coastal cities), extreme heat
- Traffic congestion, safety, crime
- Housing, income inequality, racial segregation

USING DATA SCIENCE TO SOLVE URBAN HEAT ISLAND ISSUES

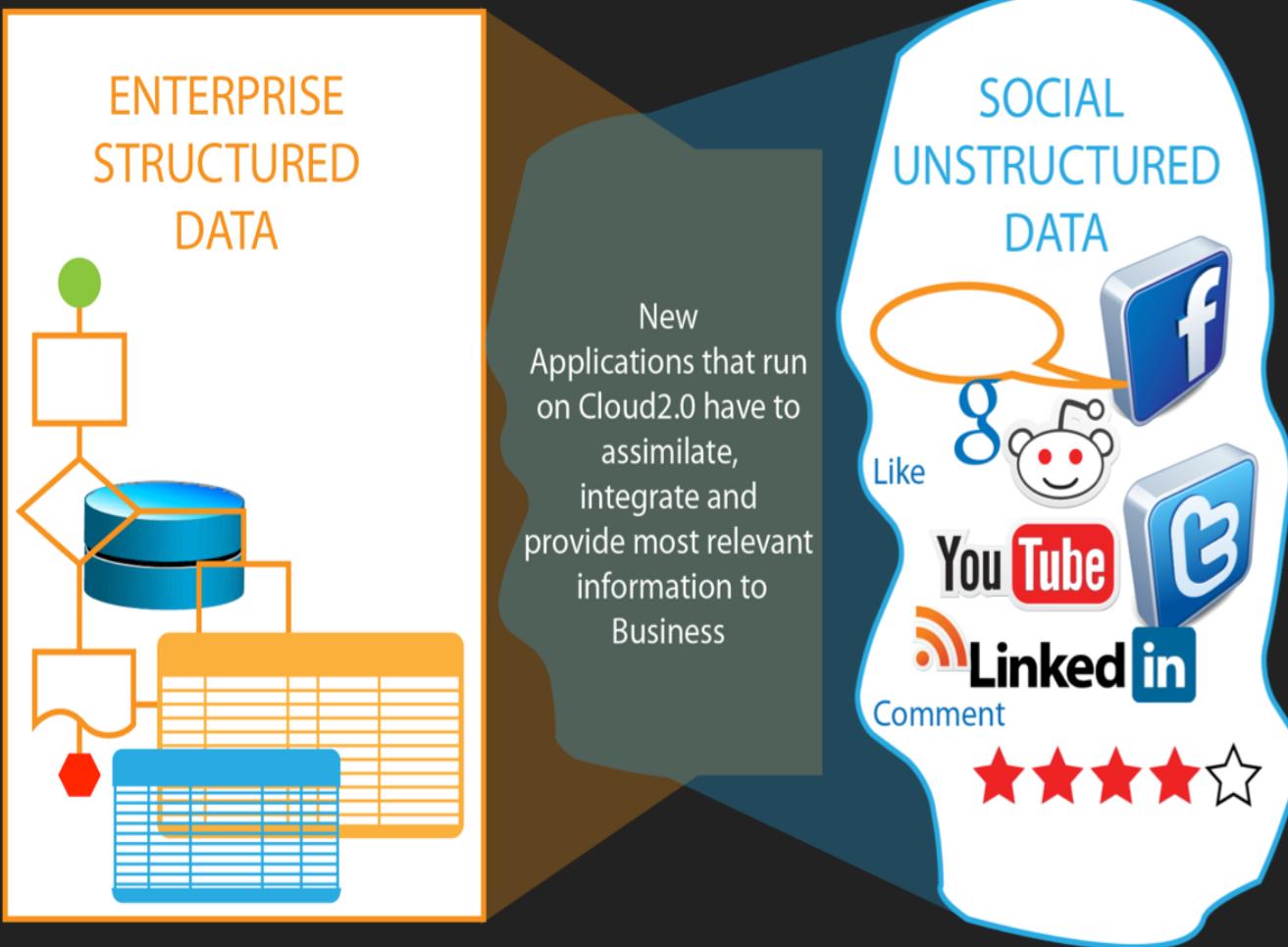
URBAN "BIG" DATA

- Every day, we create 2.5 quintillion bytes of data
- 90% of the data today has been created in the last 2 years alone.

WHAT IS BIG DATA?

- Three Vs: Volume, Velocity, Variety
- Very large data that cannot be possibly handled with typical softwares, like Excel
- Data volume continues to increase. ex) Sensor data (IoT), crowdsourced data
- Everything as data, not just text but also images, videos, tweets, etc

BIG DATA = STRUCTURED + UNSTRUCTURED DATA

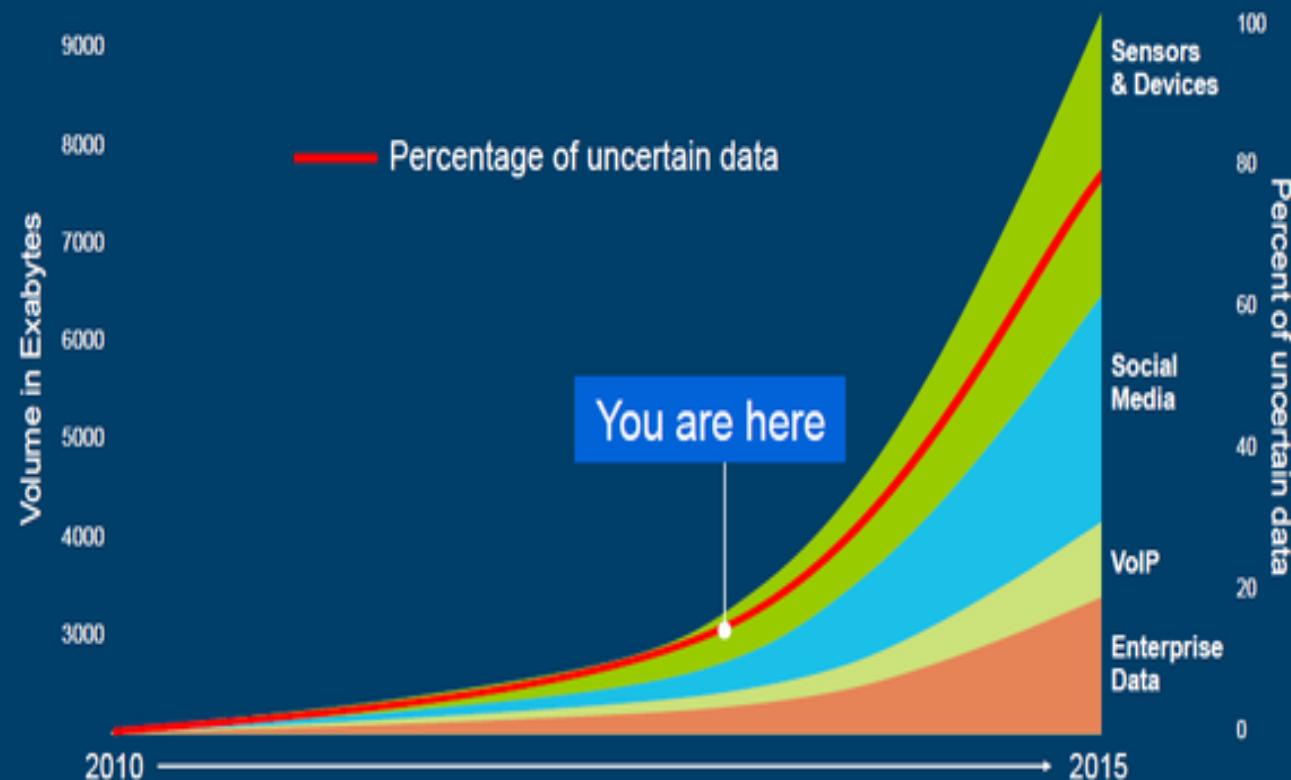


BIG DATA TREND

IBM Investor Briefing



Big Data: This is just the beginning



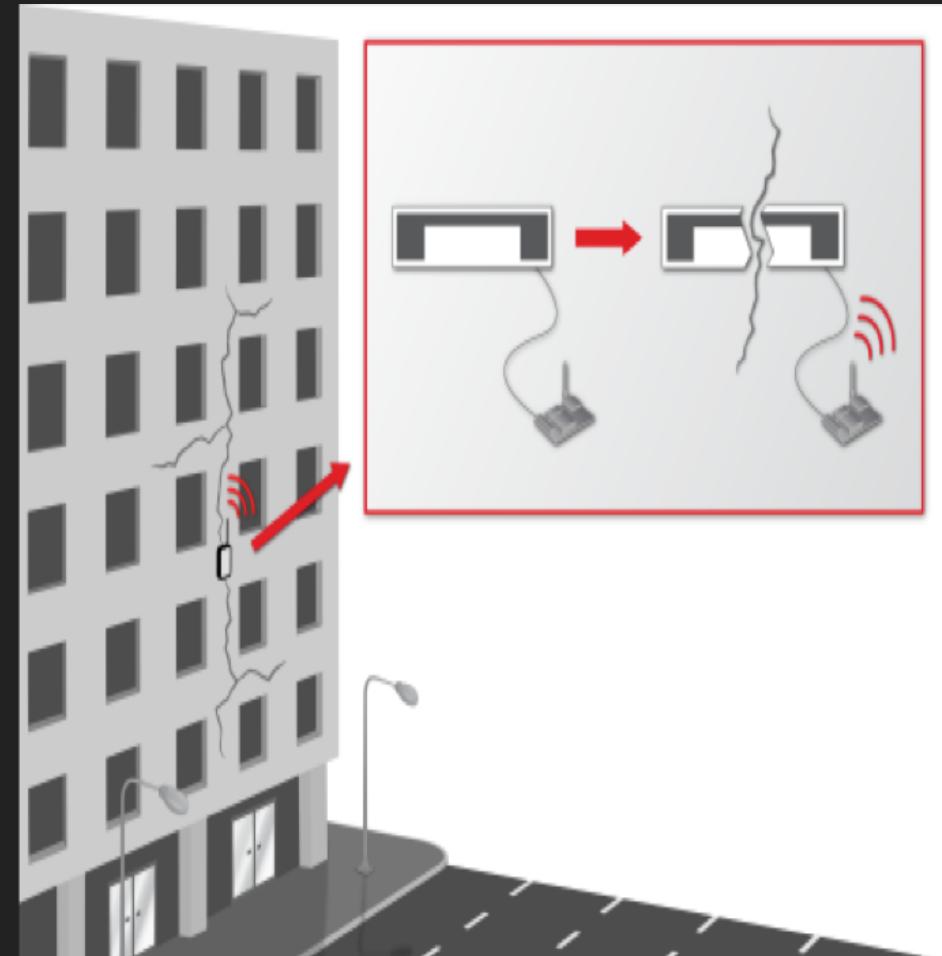
SMART CITIES AND INTERNET OF THINGS



REAL-TIME SENSORS = BIG DATA



Air Quality Sensors



Building Structure Sensors

CITY SENSOR

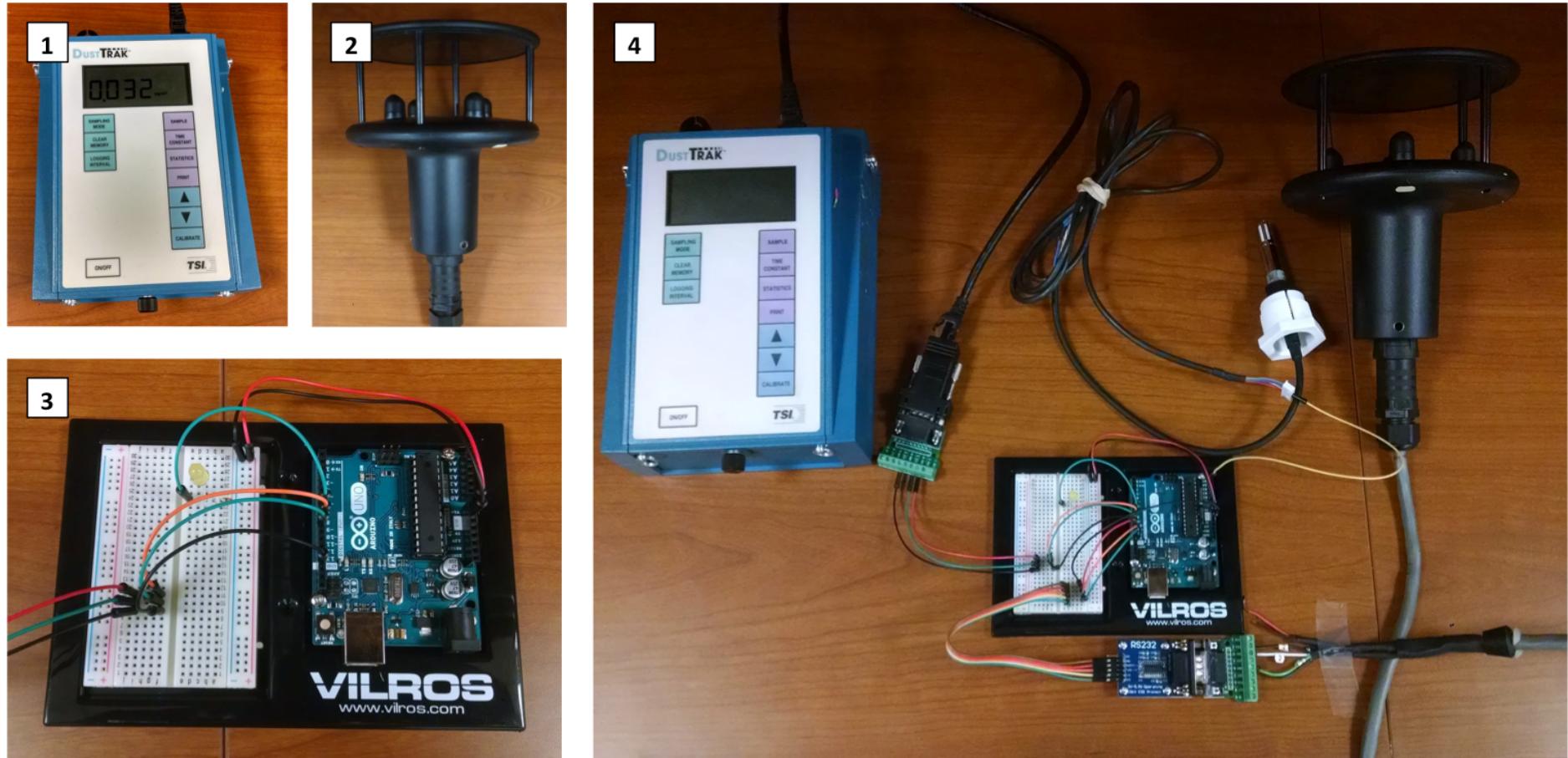


Figure 2. IoT Prototyping Environment; 1) DustTrak (PM2.5 monitor); 2) WindSonic (wind direction and speed); 3) Arduino board; 4) Complete prototype setup

RISE OF SPATIAL BIG DATA



Google Earth Engine



Landsat

4, 5, 7, and 8



MODIS

Daily, NBAR, etc.



Terrain

SRTM, NED, etc.



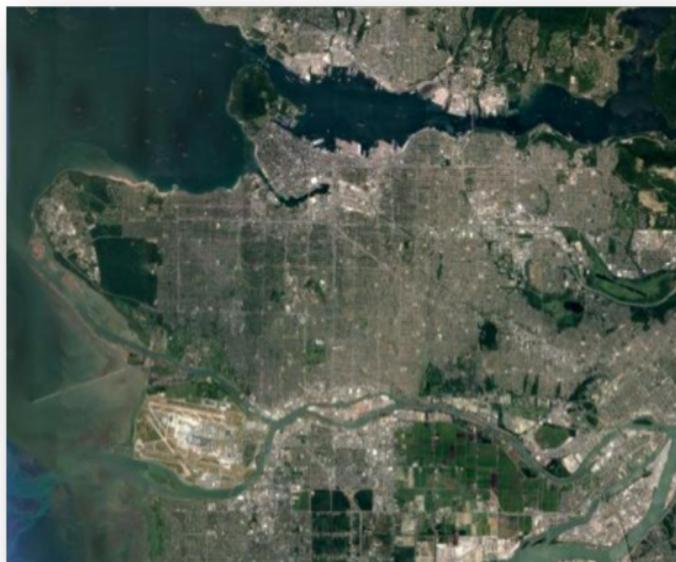
Land Cover

GlobCover, NLCD, etc.

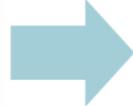


Atmospheric

NOAA NCEP, etc.



Satellite imagery



Green space measure

AGE OF CLOUD COMPUTING

Pre cloud days

- Downloading and patching together satellite images
- Long time to load gigabytes of satellite images
- Take several days to process large raster data

Post cloud days

- All the sattellite data stored on a public cloud
- No need to install software. Everything in a browser
- Google Earth Engine: data processing on the fly

GOOGLE EARTH ENGINE DEMO

Earth engine code editor:
<https://code.earthengine.google.com/>

