LESSON HANDOUT

The Data Analytics Environment

The context:

Data is everywhere and is the byproduct of nearly every action we take. Often we think of data as binary numbers stored under lock and key but your text messages, videos online and a post-it reminder on your desk are all data.

Every process, every sensor and every machine **produces** data but most importantly our processes, systems and government are increasingly being **driven** by data.

How did we get here?

Four factors have contributed to the unprecedented availability and opportunity that is presented by the data environment today.

1. **Data collection**. The growth of connected, electronic devices in the last few decades has exponentially increased the amount of data we collect.
    - Example. It is estimated that from the **invention of writing until 2013** we collected **13 exabytes of data** (in total) but now, we currently collect that amount of information **every day**.
2. **Data storage**. This data collection has been enabled by the growth of inexpensive and electronic data storage options.
    - Example. A 2mb Hard Drive disk cost the equivalent of US 1 million in 1967 vs two cents now.
3. **Processing power**. This proliferation of data is useless without the computing power required to process this data at scale.
    - Example. The Apple iPhone 5 has 2.7 times more processing power than the fastest supercomputer in 1985 - the Cray 2 Supercomputer.
4. **Statistical and programmatic methods.** Developments in programming such as machine learning allow us to handle and interpret this large scale of data. Many of the statistical methods used were invented by academics decades or even centuries ago.