

Big Data for the smaller organisation

TFMA 2015

Terry Hogan
Golden Orb

What is “Big Data”?

45,000,000,000,000,000,000

- bytes of consumer internet traffic per month

80% of data created in the past 12 months

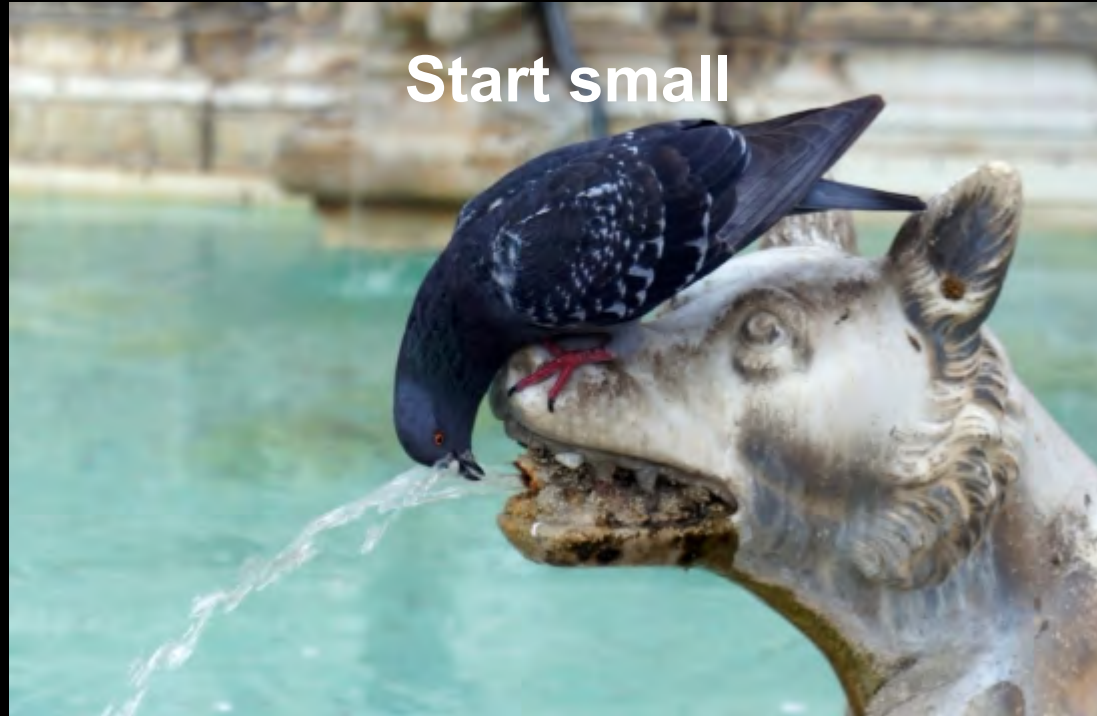
What is the relevance of **Big Data** to smaller businesses?

- New analytical capabilities
 - Much is open source
- New sources of data
 - Many free to use
- Real-time data/analysis
 - Adaptive websites
- Reduced cost of entry
 - Low-cost, cloud-based tools

Big Data - why bother?

- Because it's there!
 - Better insight
 - Better customer service
 - Improved profitability
 - New revenue streams
 - Competitive action

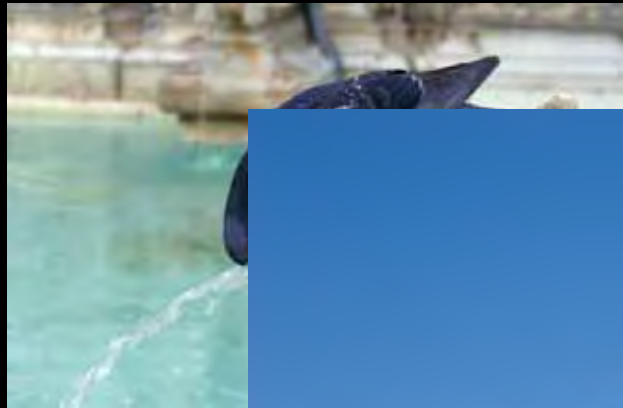
Getting started with Big Data



Getting started with Big Data



Getting started with Big Data

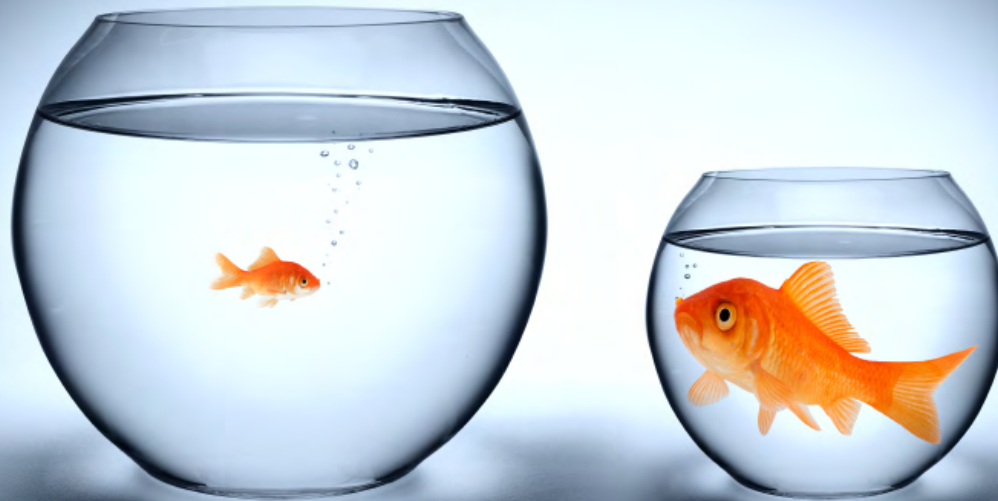


Visualise



Getting started with Big Data

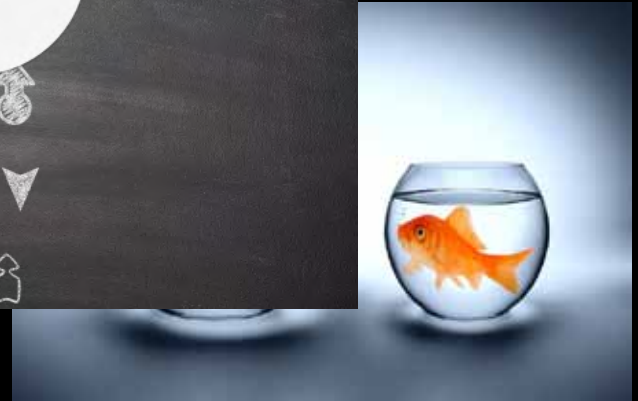
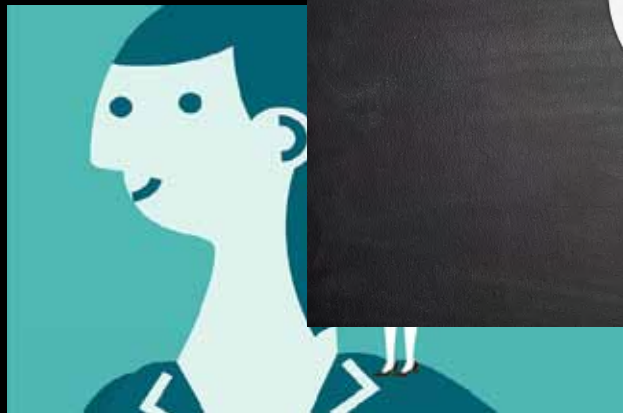
Join Big with Small



Getting started with Big Data



Think Big Data!

A central graphic on a dark grey background. At the top, the text "Think Big Data!" is written in white. Below it is a white cloud shape containing the words "BIG DATA" in large, bold, black capital letters. Underneath the cloud, there are several white arrows of various shapes and sizes, some pointing up and some pointing down, symbolizing data flow.

Tools and resources to get started with Big Data

- Platform
 - Google Cloud Platform
 - Amazon Web Services
- Language
 - Python, R
 - Google Apps Script
- Visualisation
 - Python/R built-in visualisation
 - Google Fusion Tables/Google Sheets
 - Tableau

Tools and resources to get started with Big Data

•Platform

```
import tweepy

auth=tweepy.OAuthHandler(consumer_key="xxxxx", consumer_secret="yyyyy")

api=tweepy.API(auth)

results=[]
for tweet in tweepy.Cursor(api.search, q="#tfma2015").items(500):
    results.append(tweet.text)
```

- Python/R built-in visualisation
- Google Fusion Tables/Google Sheets
- Tableau

Tools and resources to get started with Big Data

The screenshot shows a Tableau Public dashboard. At the top, there is a navigation bar with 'tableau public Beta', 'GALLERY', 'AUTHORS', 'BLOG', 'RESOURCES', and a 'SIGN IN' button. Below the navigation bar, there is a 'Gallery' header with navigation arrows and a 'Subscribe' button. The main content area features a map of the Soho neighborhood. The map is overlaid with data points: red circles of varying sizes representing the number of deaths at each address, and blue and green water pump icons representing clean and fouled pumps, respectively. A legend on the right side of the map provides details for the data points. The legend includes a 'No. Deaths at Address' section with four circle sizes corresponding to 1, 5, 10, and 15 deaths, and a color scale from 1 (light red) to 15 (dark red). The 'Water Pumps' section shows a blue water drop icon for 'Clean' and a green water drop icon for 'Fouled'.

Then, Snow mapped these data points onto a map of the Soho neighborhood.

The results were startling.

NB: The map reprinted here is Snow's original. I geocoded the addresses, introduced the number of deaths as a dimension in the dataset, and uploaded the map as a background image. Then, I conformed the dimensions of the map to the appropriate lat/long. Thus, the locations appear as they would have on Snow's map.

No. Deaths at Address

1
5
10
15

1 15

Water Pumps

- Clean
- Fouled

— Tableau

Tools and resources to get started with Big Data

- Public tools and data sources
 - Google
 - Correlate, BigQuery, Custom Search, Prediction
 - Public data explorer, shared fusion tables
 - Amazon
 - Elastic Map Reduce / Hive
 - Open Government Licence
- Learning resources
 - Coursera
 - Roshan data mining videos

Tools and resources to get started with Big Data

- Public

- Go

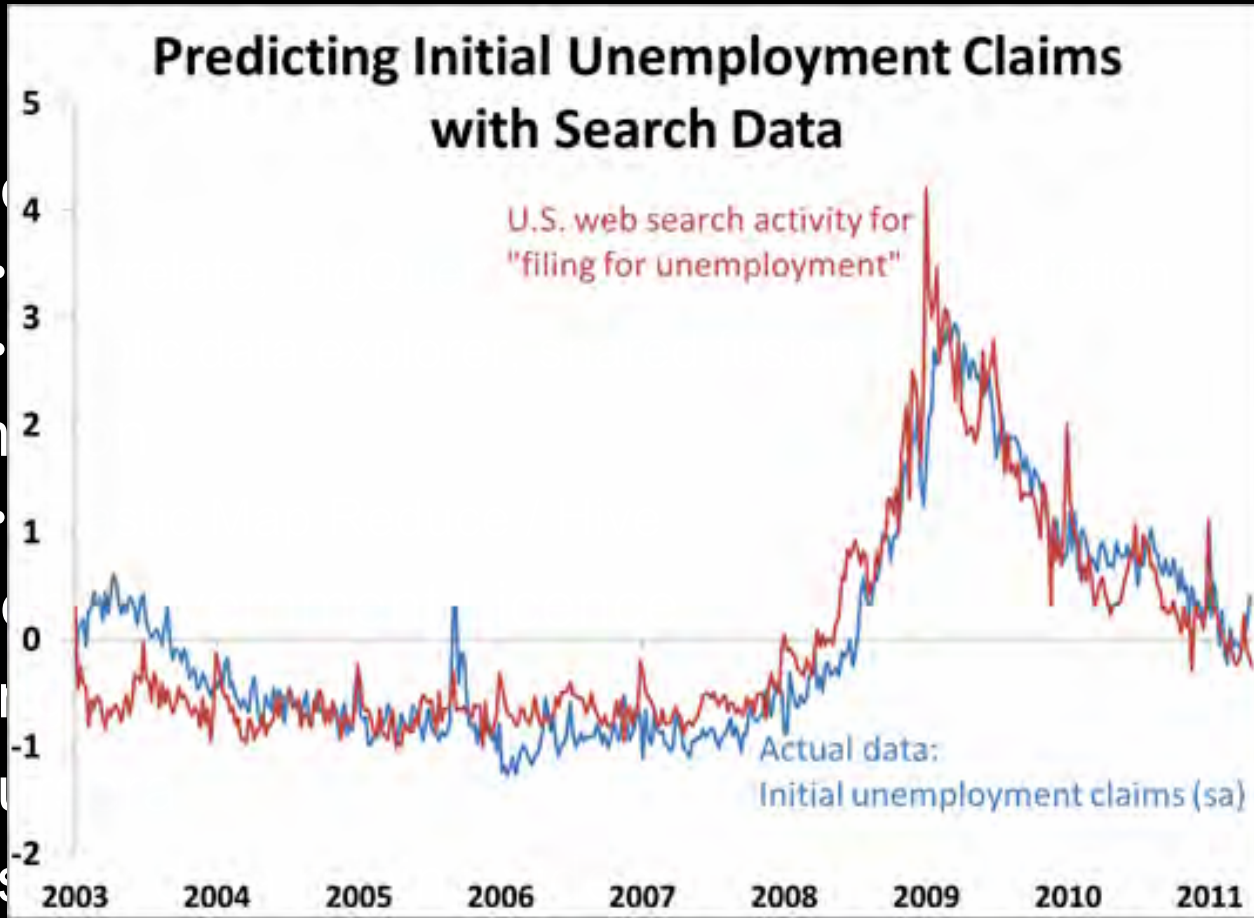
- Am

- Op

- Learni

- Co

- Ros

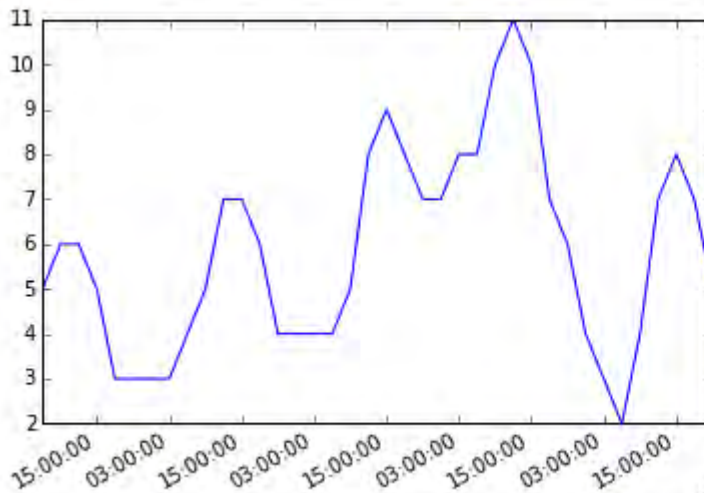


Tools and resources to get started with Big Data

• P

```
In [3]: import metoffer
M = metoffer.MetOffer(METOFFICE_KEY)
x = M.nearest_loc_forecast(51.4033, -0.3375, metoffer.THREE_HOURLY)
y = metoffer.parse_val(x)
```

```
In [4]: labels, temp = zip(*[(d['timestamp'][0], d['Temperature'][0]) for d in y.data])
p = plt.plot(labels, temp, 'b-')
p[0].figure.autofmt_xdate()
```



• L

Thank You

Terry Hogan

www.golden-orb.ltd.uk

[@GoldenOrbLtd](https://twitter.com/GoldenOrbLtd)