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

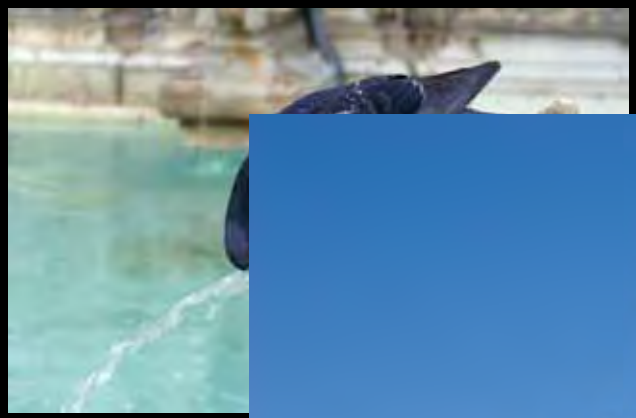
Big Data - why bother?

- 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



Getting started with Big Data

Getting started with Big Data

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

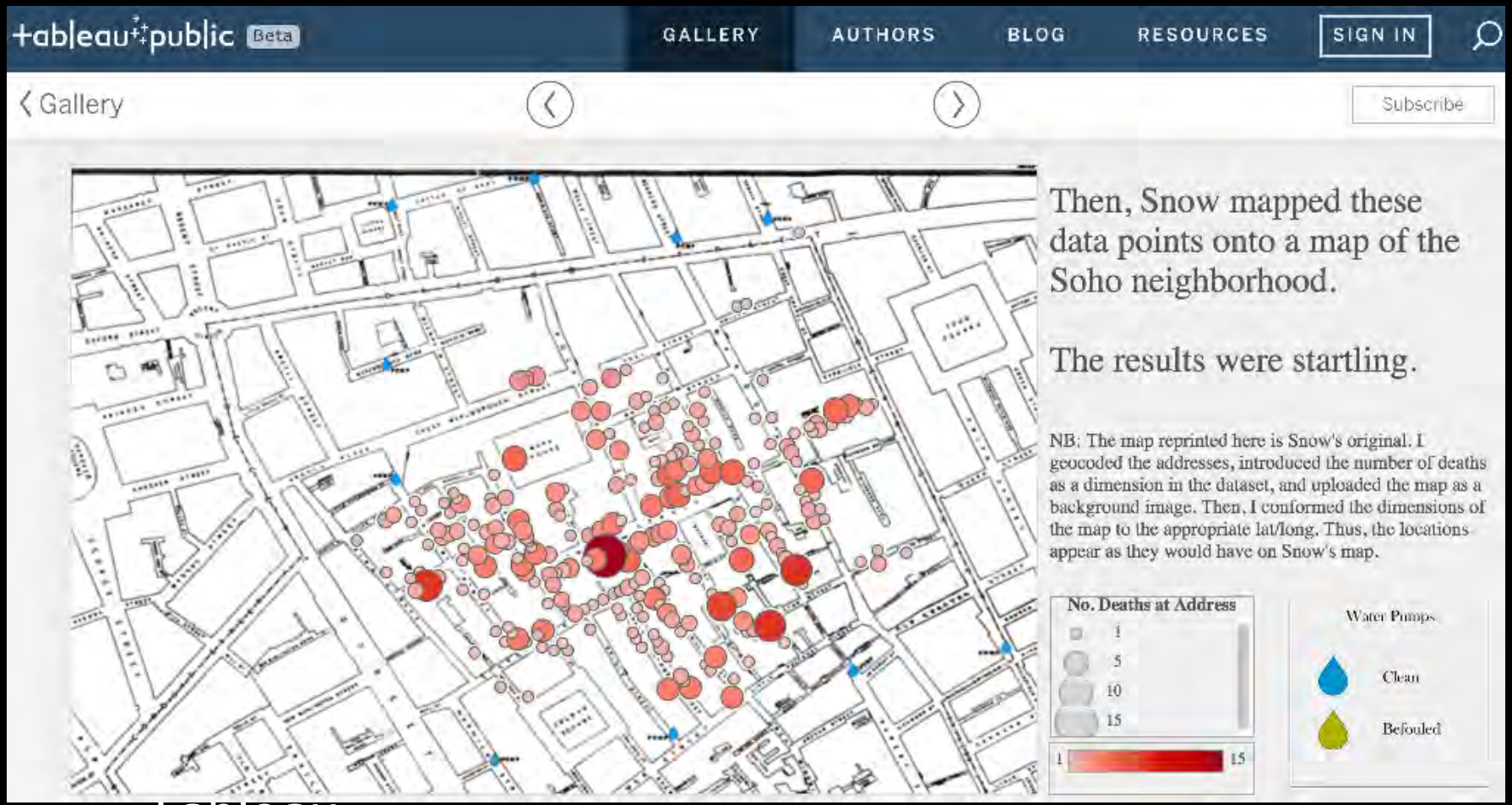
Tools and resources to get started with Big Data

- 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



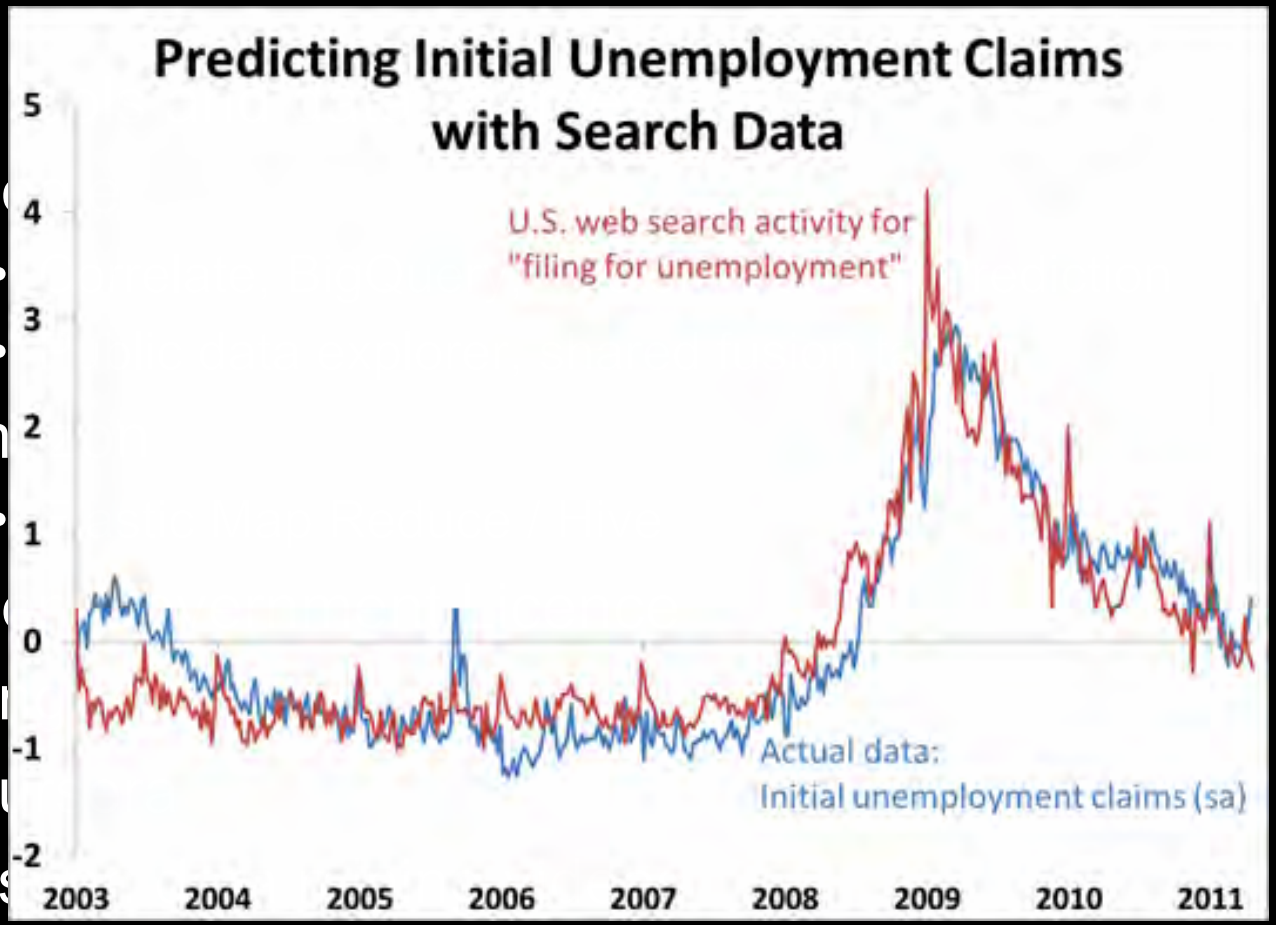
– Tableau

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

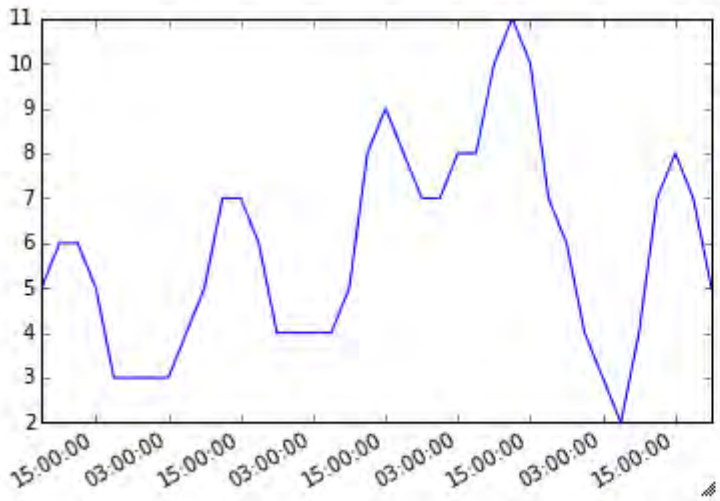
- Public
  - –Go
    -
    -
  - –Am
    -
  - –Op
- Learni
  - –Co
  - –Ros



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