Big data skills in geo-spatial

Richard Kingston

Senior Lecturer in Urban Planning & Smart Cities

Head of Department of Planning & Environmental Management

School of Environment, Education & Development

richard.kingston@manchester.ac.uk

@gisplanner

13th October 2016

MANCHESTER
1824
The University of Manchester

Big data skills in geo-spatial

1. What is geo-spatial data?

2. Why (if at all) is it different to other data?

3. Does geo big data need different skills?

4. Data sources – what's changed in the geo-world?

5. Which software?

6. Summary

   Questions / discussion

Oxford St. 1900

Oxford St. 2010

- The visual depictions of future cites have not materialised

- Instead we have 'cities of bits' (Mitchel, 1995) and a 'space of flows' (Castells, 1989)

1. So, what is geo-spatial data?

- Think of an Excel / SPPS / SAS spreadsheet linked to a map
  - each row is a place / location
  - each column an attribute of that place / location
  - there are nearly always multiple attributes of each place / location
  - places / locations can be lines, points or polygons
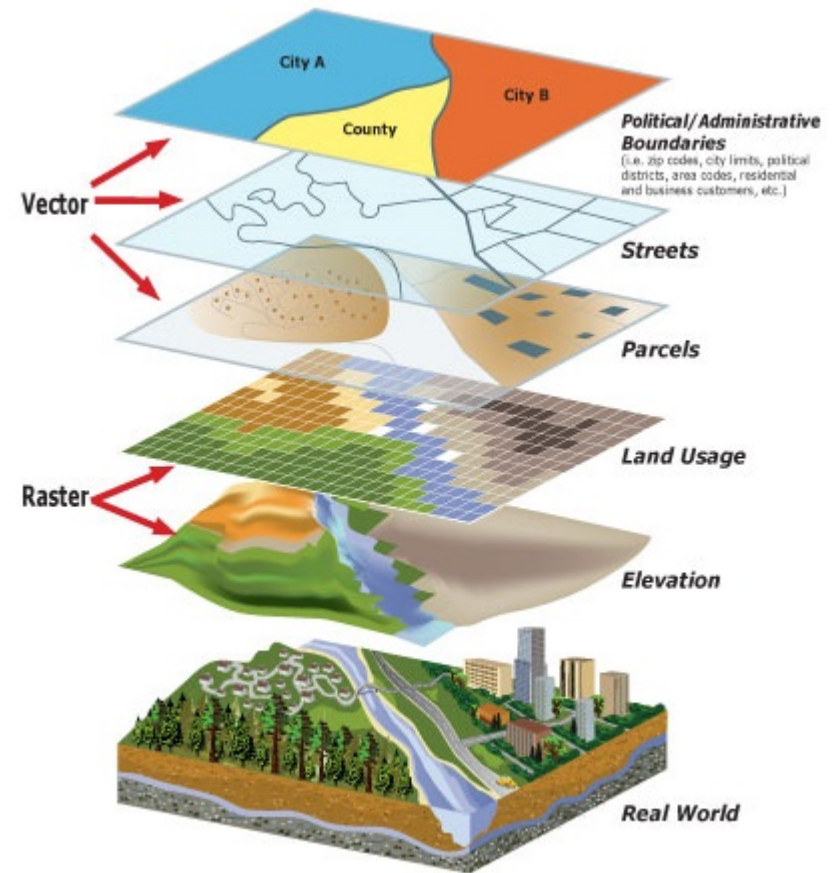
1. What is geo-spatial data?

- YOU are geo-spatial data and you have lots of attributes
  - where you live, work, play, shop
  - where to travel to from
  - where you spend your money, purchase goods and services
- You should all be in the 2011 & 2001 census (possibly the 1991, '81 and '71 ones too)

www.ppgis.manchester.ac.uk

2. Why (if at all) is it different to other data?

- Location, location, location
- Spatial resolution matters
    - England LSOA map ungeneralised 450mb
    - England LSOA map generalised 125mb
- Classic example
    - if coastline of GB is measured using 100km long units, then coastline is approx. 2,800km.
    - with 50km units, then coastline is approx. 3,400km. i.e. approx. 600km longer as so...

        see Mandelbrot 1967.

www.ppgis.manchester.ac.uk

2. Why (if at all) is it different to other data?

- Temporal resolution matters too
  - frequency of sampling massively increase data size
  - here we can see 9,678 planes in real-time
  - quickly multiplies into millions, billions, trillions of data points
  - same applies to movement of people, goods vehicles etc. in our cities

www.p

3. Does geo big data need different skills?

- In the geo-spatial word maybe not...?
- Tools and methods already exist to analyse big spatial data
  - see Openshaw's work in the 1990s on GAM
  - analysis of health data to detect cancer clustering
- The biggest problem is not analysis of the data but
  - methods to deal with increasingly distributed and disparate data
  - with such BIG data how do we deal with error propagation?
  - are we asking the right questions of the data?

According to Akoko et al (2015)

"Big Data has been a buzzword in the last decade. The term, coined by Roger Magoulas , refers to large data sets almost impossible to manage and process using traditional data management tools. It refers to various forms of large information sets requiring complex computational platforms in order to be analyzed."
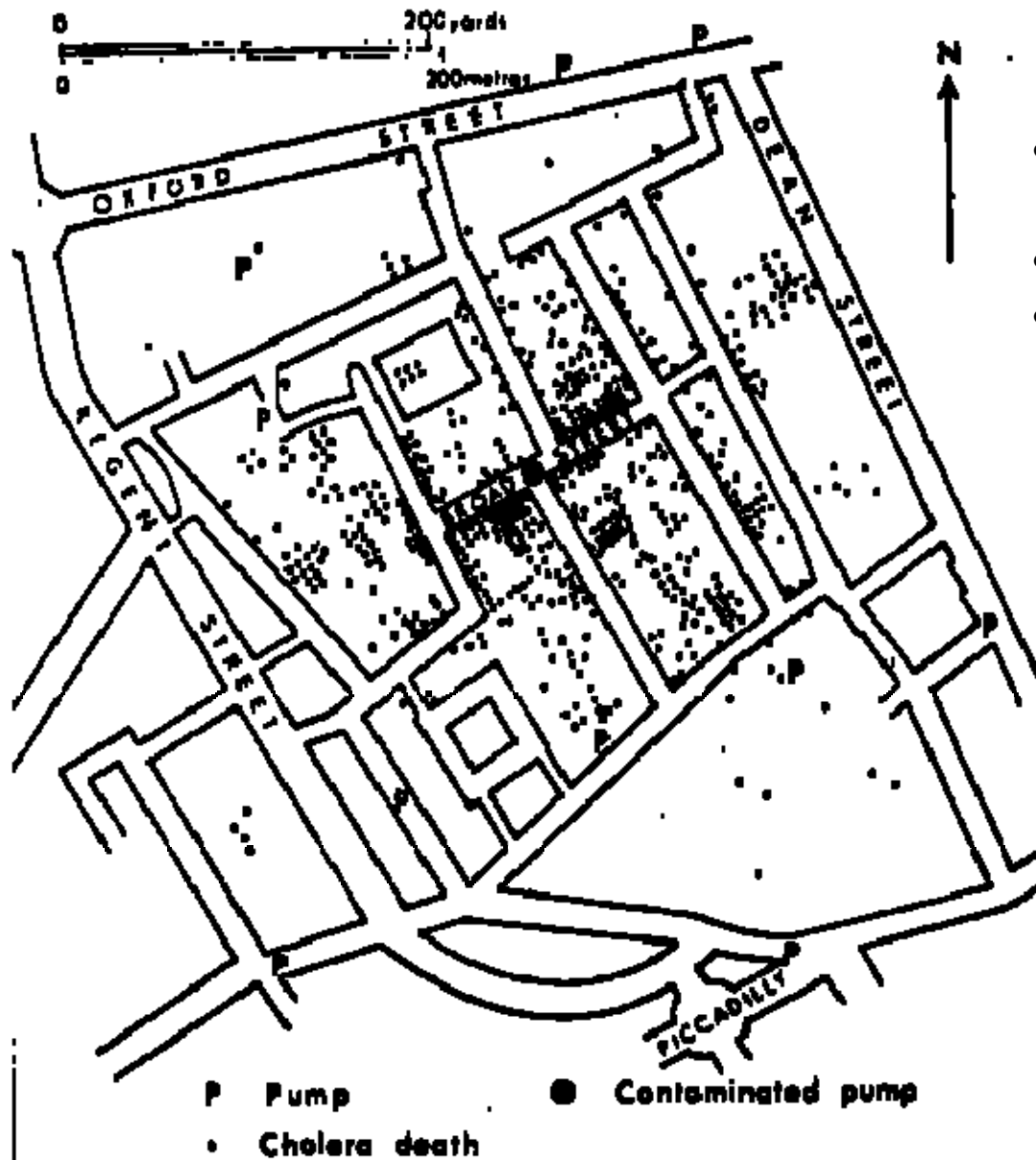
- We can use the same models and methods it is just a question of scaling up
- But, databases need to cope with ever increasing rows and columns – noSQL?
- Distributed data tables?
- Can current software cope? NO!

P **Pump**   ● **Contaminated pump**

• **Cholera death**

www.ppgis.manchest

4. Data sources – what's changed in the geo-world?

- From static to dynamic

- Local to cloud – Web Mapping Services

- In the past I grabbed my own copy of the data
  - it was out of date very quickly
  - my version was different to your version
  - who's version is correct?
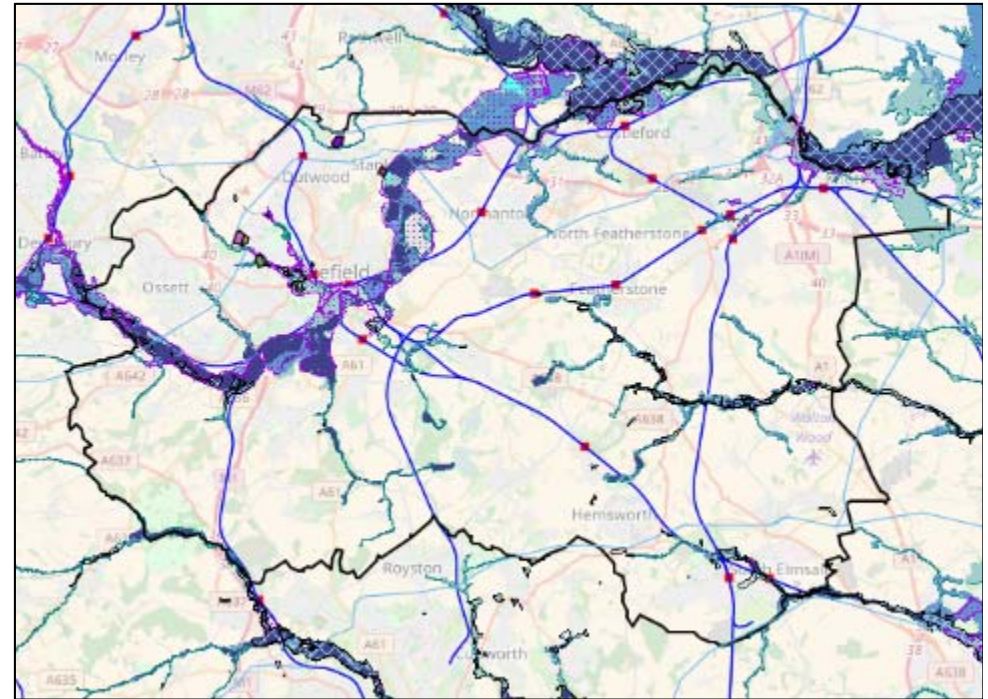
5. Which software?

- If this is new to you the software is not the starting point
  - most of it does the same thing anyway e.g. ESRI's proprietary GIS (www.esri.com) vs. open source free to download QGIS (www.qgis.org)
- Understanding the quantitative methods and analysis tools is the starting point
- Which tool for you data?
  - is your data vector or raster?
- We have a problem with current software falling over even with small data

6. Summary

- Geo-spatial is multi-faceted
  - great for bring together disparate data in to one place
  - Excellent at visualising patterns – ESDA

- Data is getting bigger with more opportunities for collecting and analysing in real-time

- We have the analysis tools

- **BUT,** we need distributed databases that can cope

- You need to ask the right question to frame your data mining/analysis

Questions and discussion

Seminar:

Commuteflow - a geodemographic classification of commuting for England and Wales: Methods and Applications

Richard Kingston & Stephen Hincks

Tuesday 25th October 5pm Studio 2.3, Humanities Bridgeford St.

Richard Kingston

Senior Lecturer in Urban Planning and Smart Cities

Head of Department of Planning & Environmental Management

School of Environment, Education & Development

richard.kingston@manchester.ac.uk

@gisplanner

www.ppgis.manchester.ac.uk

References

Akoka, J., I. Comyn-Wattiau, I. and Laoufi, N. (2015) Research on Big Data. <u>Advances in Conceptual Modeling: ER 2015 Workshops AHA, CMS, EMoV, MoBID, MORE-BI, MReBA, QMMQ, and SCME, Stockholm, Sweden, October 19-22, Proceedings.</u> M. A. Jeusfeld and K. Karlapalem. Cham, Springer International Publishing**:** 173-183.

Castells, M. (1989) The Informational City: Information Technology, Economic Restructuring, and the Urban Regional Process. Chichester: John Wiley & Sons.

Mandelbrot, B. (1967) How Long Is the Coast of Britain? Statistical Self-Similarity and Fractional Dimension, *Science,* New Series, 156(3775), pp. 636-638. [doi](doi):10.1126/science.156.3775.636

Mitchell, W. (1995) City of bits: Space, Place and the Infobahn. Cambridge: MIT Press.