Introduction to Geographic Data Science for Urban Analytics

Alex Singleton

Professor of Geographic Information Science

Department of Geography and Planning

Northern Regional Data Facility

LIFE CHANGING | World Shaping

UNIVERSITY OF LIVERPOOL

Consumer Data Research Centre

www.cdrc.ac.uk

www.geographicdatascience.com

www.alex-singleton.com

@alexsingleton

Overview

- Geographic Data Science (inc **Systems)**
- *Open* “Big” **Data**
- Future Prospects
- The course

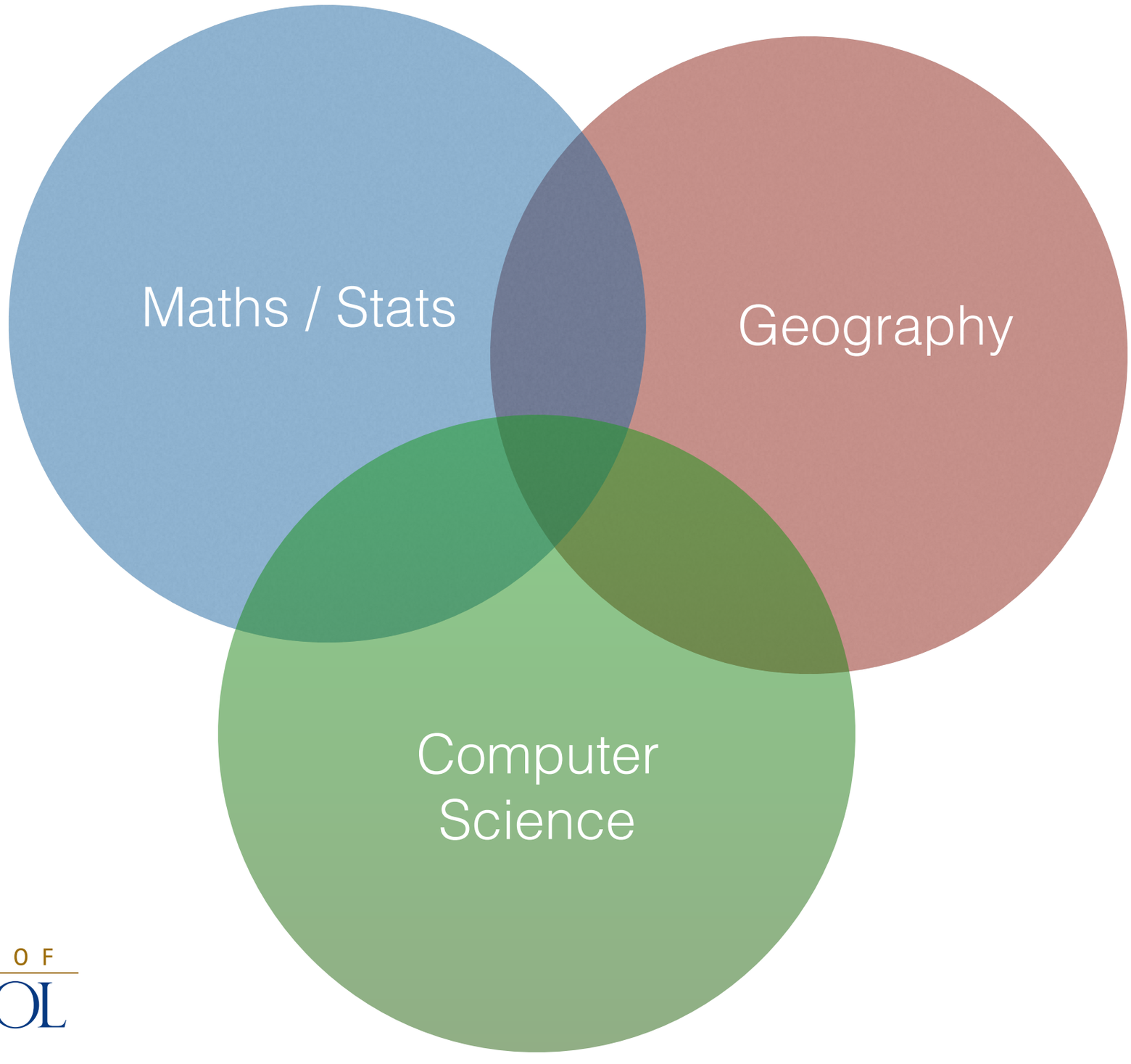
UNIVERSITY OF LIVERPOOL

Geographic Data Science

- Couples burgeoning new and dynamic data sources with advanced quantitative and computational methodology to advance debates around problems of global social and economic importance.

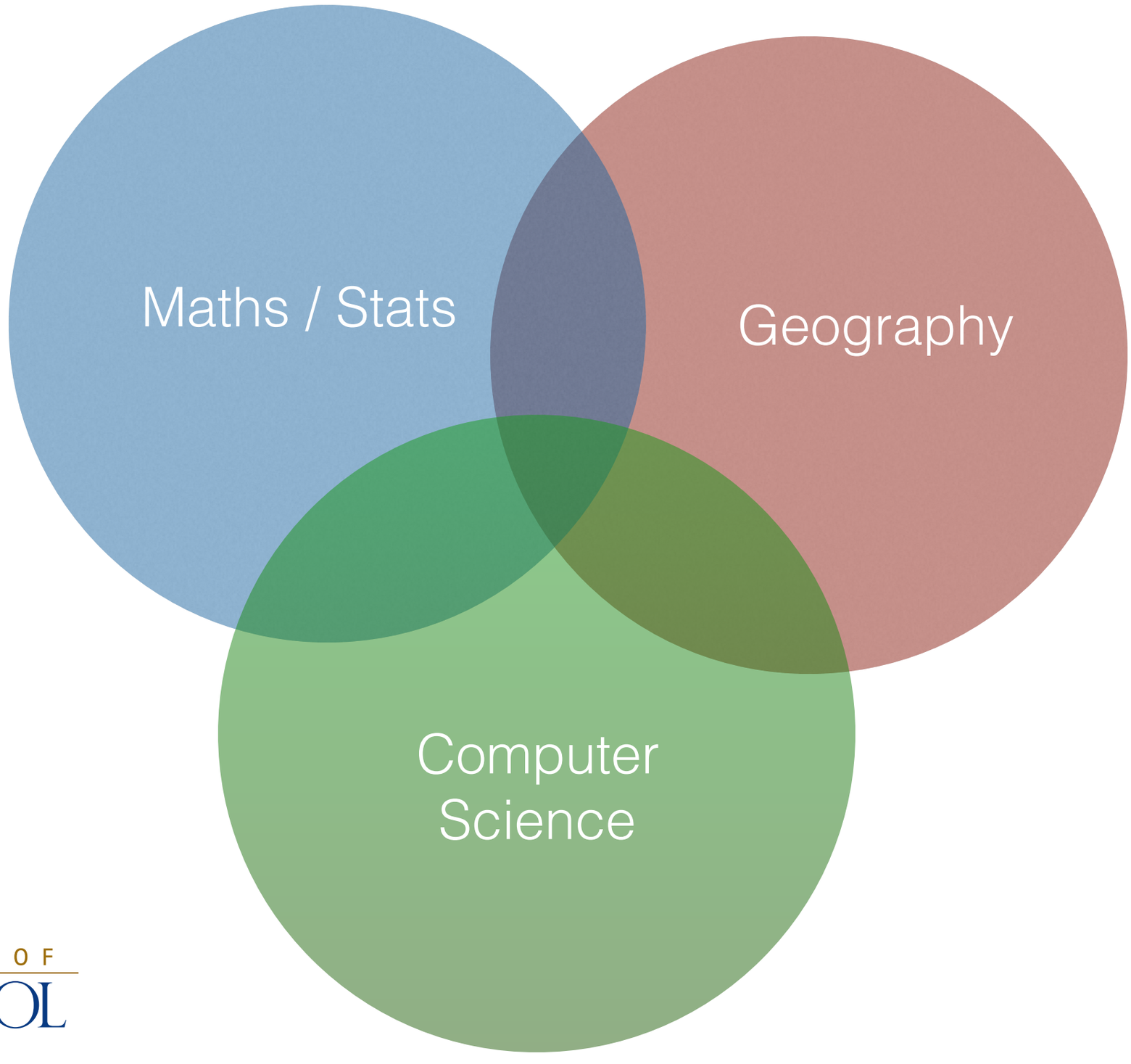
UNIVERSITY OF LIVERPOOL

Geographic Data Science



UNIVERSITY OF
LIVERPOOL

Geocomputation



UNIVERSITY OF
LIVERPOOL

WARWICK

Text only | Sign in

Statistics More...

Search Statistics

Department of Statistics

Undergraduate

Data Science » Admis

Undergrad

Computer science and

BSc Data Science is
sophisticated theory a
scale data analysis. It
Department of Compu
insight to operate at t

The role of data scient
This is because data s
created by the revolut
in the 20th Century: h
are generated on a da
shopping and social ne
detailed information o
interpretation, this dat
large scale data into in
in Statistics and Comp

Currently, global dema
statistical and computi
outstrips supply, with
predictions suggesting
this area for at least th
graduates in Data Scie
presents opportunities
one of the most crucia

coursera | Global Partners

Courses Specializations Institutions About | Sign In Sign Up

UNIVERSITY of WASHINGTON

Introduction to Data Science

Join the data rev
This specialized
conventional cur
and leave armed
data.

Preview Lectu

About the

Commerce and
prediction. Skills
management on
proficiency with
disciplines and a
techniques of da
data managemer
mining (e.g., clus
(e.g., linear and

IBM

Industries & solutions Services Products Support & downloads My IBM

Search

IBM Software › Information Management › InfoSphere Platform › Big Data Analytics ›

What is a data scientist

About data scientists

Rising alongside the relatively new technology of big data is the new job title data scientist. While not tied exclusively to big data projects, the data scientist role does complement them because of the increased breadth and depth of data being examined, as compared to traditional roles.

Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data

→ Download the ebook

So what does a data scientist do?

A data scientist represents an evolution from the business or data analyst role. The formal training is similar, with a solid foundation typically in computer science and applications, modeling, statistics, analytics and math. What sets the data scientist apart is strong business acumen, coupled with the ability to communicate findings to both business and IT leaders in a way that can influence how an organization approaches a business challenge. Good data scientists will not just address business problems, they will pick the right problems that have the most value to the organization.

The data scientist role has been described as "part analyst, part artist." Anjul Bhambhri, vice president of big data products at IBM, says, "A data scientist is somebody who is inquisitive, who can stare at data and spot trends. It's almost like a Renaissance individual who really wants to learn and bring change to an organization."

Whereas a traditional data analyst may look only at data from a single source – a CRM system,

Contact IBM

Considering a purchase?

- Chat now
- Email IBM
- Request a quote
- Or call us at: 1-877-426-3774
  Priority code: Info Mgmt

Forbes Article

Learn what a data scientist is from IBM's Anjul Bhambhri

→ Read the article

Blog posts by James Kobielus

- Data Scientists: Myths and mathemagical superpowers
- Data Scientist: Closing the Talent Gap
- Data Scientist: Master the Basics, Avoid The Most Common Mistakes
- Data Scientist: Exploration in the Age of the

UNIVERSITY O
LIVERPOOL

What is a Data Scientist?

Vice president of big data products at IBM, says:

"A data scientist is somebody who is inquisitive, who can stare at data and spot trends"

"It's almost like a Renaissance individual who really wants to learn and bring change to an organization."

UNIVERSITY OF LIVERPOOL

http://www-01.ibm.com/software/data/infosphere/data-scientist/

Google

Trends

Worldwide | 2004 - present | All categories | Web Search

Hot Searches

Top Charts

Explore

Subscriptions

NCAA March Madness

Compare

Search terms

Locations

Time ranges

Topics

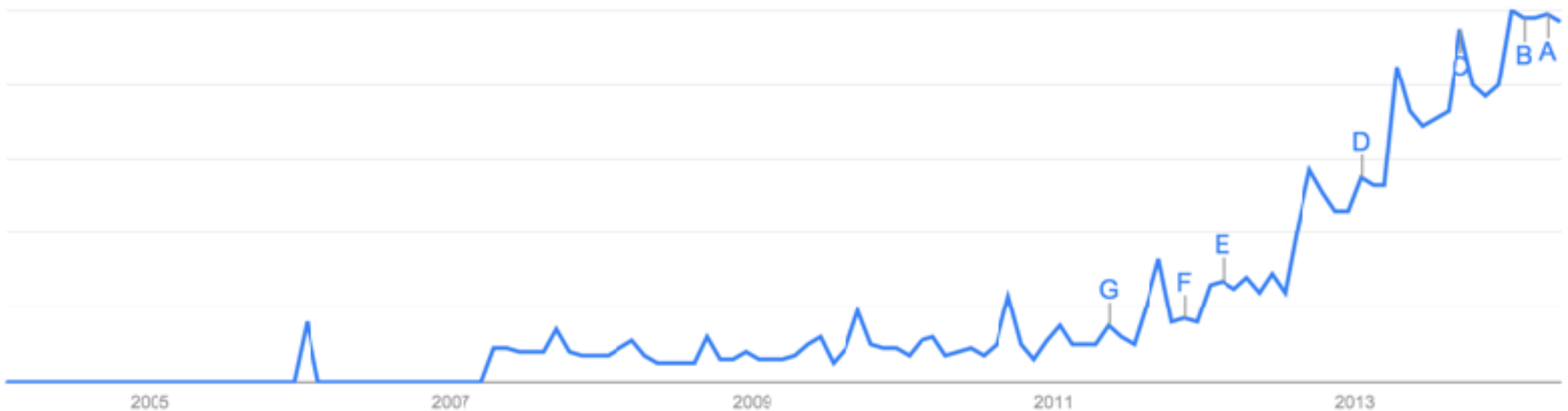
Subscribe

data scientist
Search term

+ Add term

Interest over time

News headlines Forecast



UNIVERSITY OF LIVERPOOL

Open Data

Open data is information that is available for anyone to use, for any purpose, at no cost.

http://theodi.org/guides/what-open-data
http://www.flickr.com/photos/mayhem/3899818862/

UNIVERSITY OF LIVERPOOL

DATA.GOV

DATA TOPICS APPLICATIONS DEVELOPERS CONTACT

The home of the U.S. Government's open data

Here you will find data, tools, and resources to conduct research, develop web and mobile applications, design data visualizations, and more.

GET STARTED

SEARCH OVER 85,000 DATASETS

Credit Card Complaints

BROWSE TOPICS

Agriculture Consumer Education Energy Finance Geospatial

DATA.GOV.UK Beta
Opening up Government

Home Data Apps Interact

Linked data, registries and talking about the weather

Of open data, sticky sweets and nights in Ibiza

Data.gov.uk to g
Integrated release of our
Drupal stack

What did open data ever do for us?

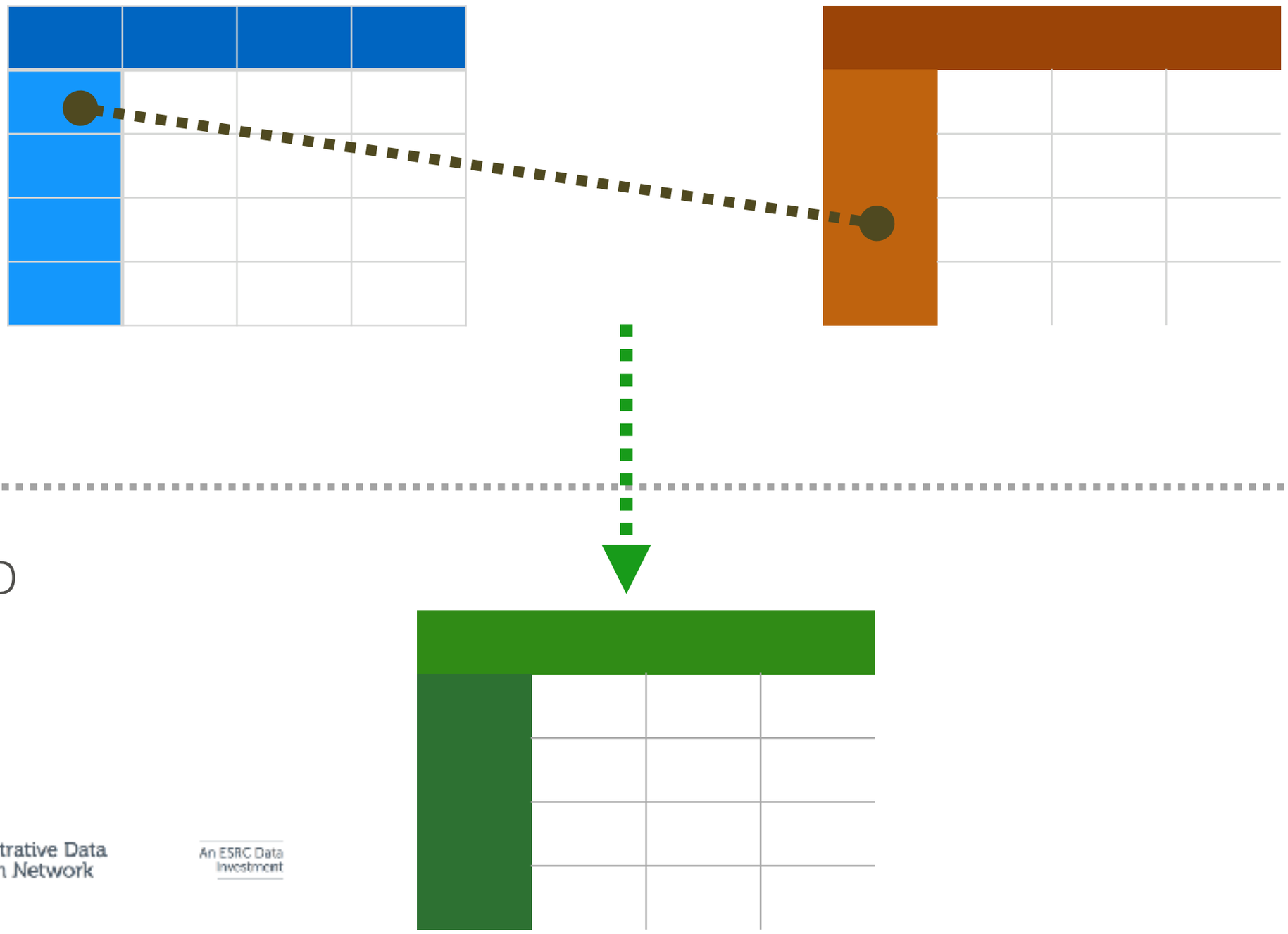
Improvements to data.gov.uk

A quick guid
working with
data

SECURE

School Data

University Data



SANITISED

Administrative Data Research Network

An ESRC Data Investment

UNIVERSITY OF LIVERPOOL

Singleton, A. (2010). The Geodemographics of Educational Progression and their Implications for Widening Participation in Higher Education. Environment and Planning A, 42(11):2560–2580.

BUT

Stan Openshaw

Open Data

- Open data has to have a **licence that says it is open** data. Without a licence, the data can't be reused. The licence might also say:
  - that people who use the data must credit whoever is publishing it (this is called **attribution**)
  - that people who mix the data with other data have to also release the results as open data (this is called **share-alike**)

http://theodi.org/guides/what-open-data

BUT

Stan Openshaw

Free Data is not Open Data

BIG DATA

Rio, Brazil

UNIVERSITY OF LIVERPOOL

http://www.npr.org/blogs/parallels/2013/05/31/187316703/Rio-Goes-High-Tech-With-An-Eye-Toward-Olympics-World-Cup

Reality

http://www.flickr.com/photos/x-ray_delta_one/8184264475/
http://www.flickr.com/photos/neoporcupine/1866929252/

Fantasy

YOUR PERSONAL "FLYING CARPET" Step into it, press a button, and off you go to market, to a friend's home, or to your job. Take off and land anywhere: no parking problems. Plug in to any electric outlet for recharging. They're working on it!

MORE POWER TO YOU!

America's independent light and power companies build for your new electric living

Tomorrow's higher standard of living will put electricity to work for you in ways still unheard of!

The time isn't too far off, the experts say, when you'll wash your dishes without soap or water—ultrasonic waves will do the job. Your beds will be made at the touch of a button. The kids' homework will be made interesting and even exciting when they are able to dial a library book, a lecture or a classroom demonstration right into your home—with sound. (Some of this is happening already.)

To enjoy all this, you'll want a lot more electric power, and the independent electric companies of America are already building new plants and facilities to provide it. Right now these companies are building at the rate of $5,000,000,000 a year, and planning to double the nation's supply of electricity in less than 10 years.

America has always had the best electric power service in the world. The electric companies are resolved to keep it that way.

AMERICA'S INDEPENDENT ELECTRIC LIGHT AND POWER COMPANIES

Company names on request through this magazine

XR

By Marc Waddington | 7 Comments | 3 Jun 2013 08:00

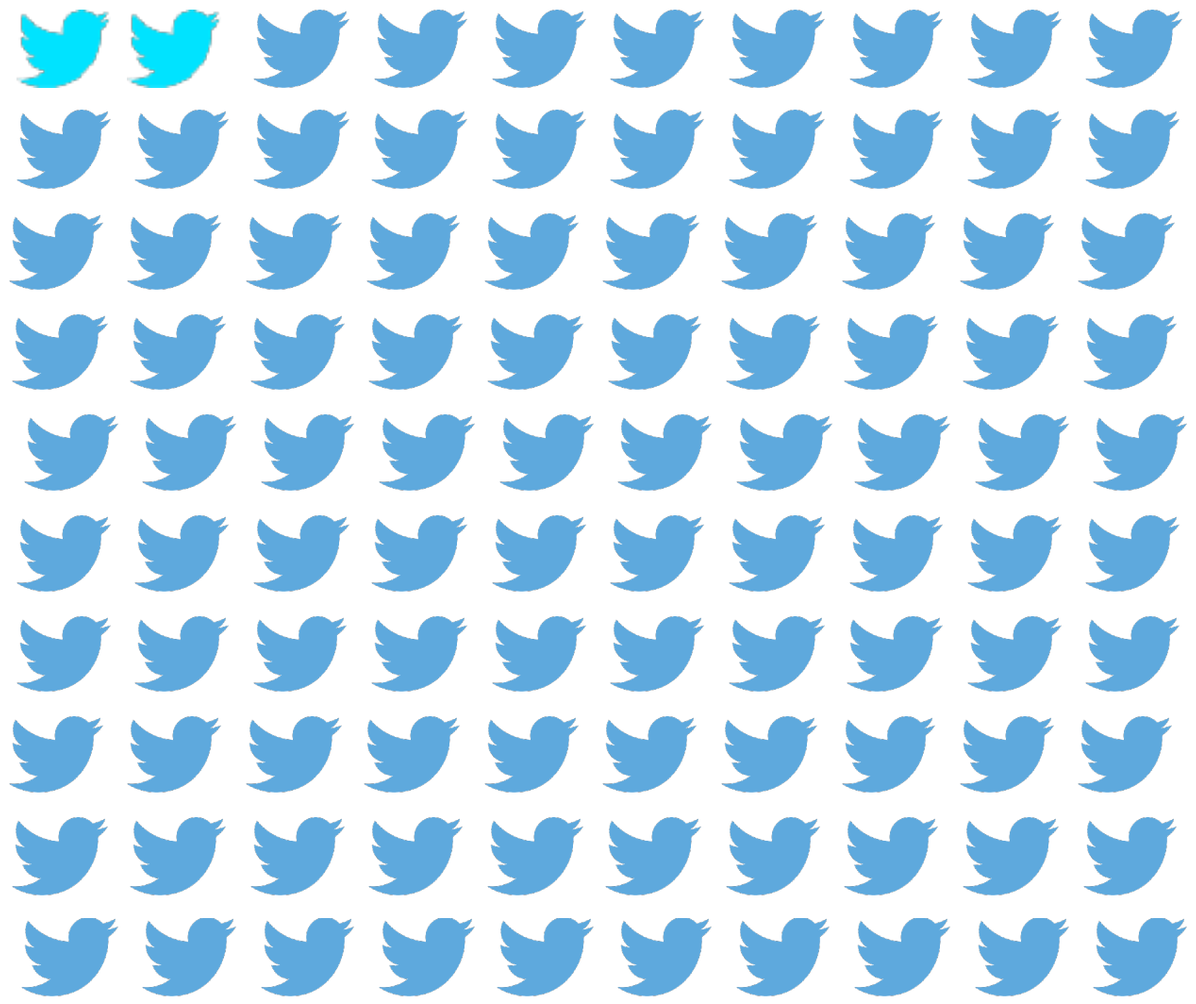
Liverpool council to borrow £200m to end pothole crisis

Login with Facebook

See more stories you'll love

You've turned off story recommendations. Turn

Sampling bias?



UNIVERSITY OF LIVERPOOL

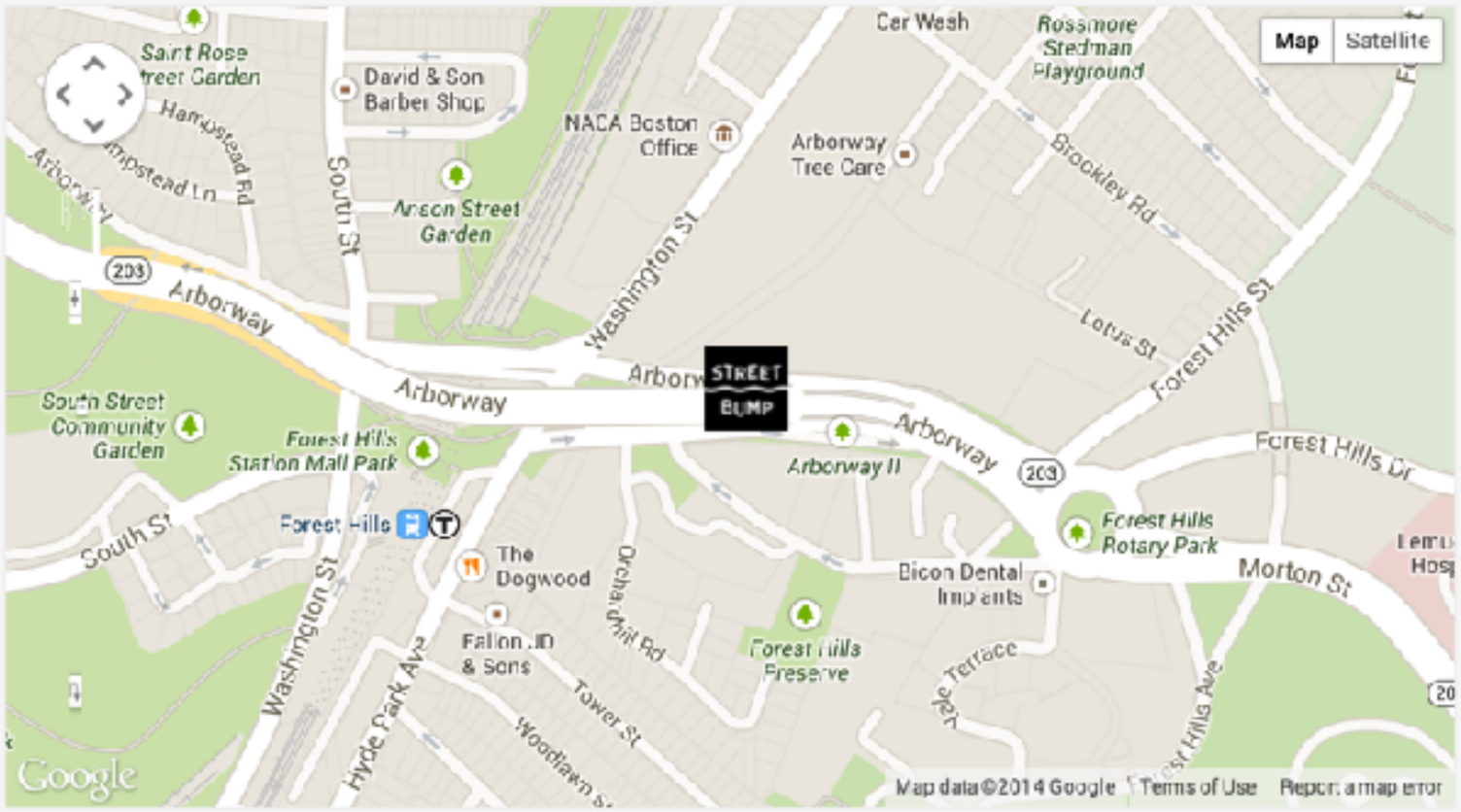
Sampling bias?

STREET BUMP About Sign In

Where's Street Bump being used?

430 trips, 32,577 bumps, 0 potholes filled, and 0 roadway problems identified

4 bumps reported in Monsignor William J Casey Hwy, Boston, MA 6 days ago



Want to use Street Bump to improve your community? Contact Us

What's Street Bump?

Street Bump is a crowd-sourcing project that helps residents improve their neighborhood streets. Volunteers use the Street Bump mobile app to collect road condition data while they drive. The data provides governments with real-time information to fix problems and plan long term investments.

Download on the App Store Learn More

UNIVERSITY OF LIVERPOOL

http://streetbump.org/

Sampling bias?

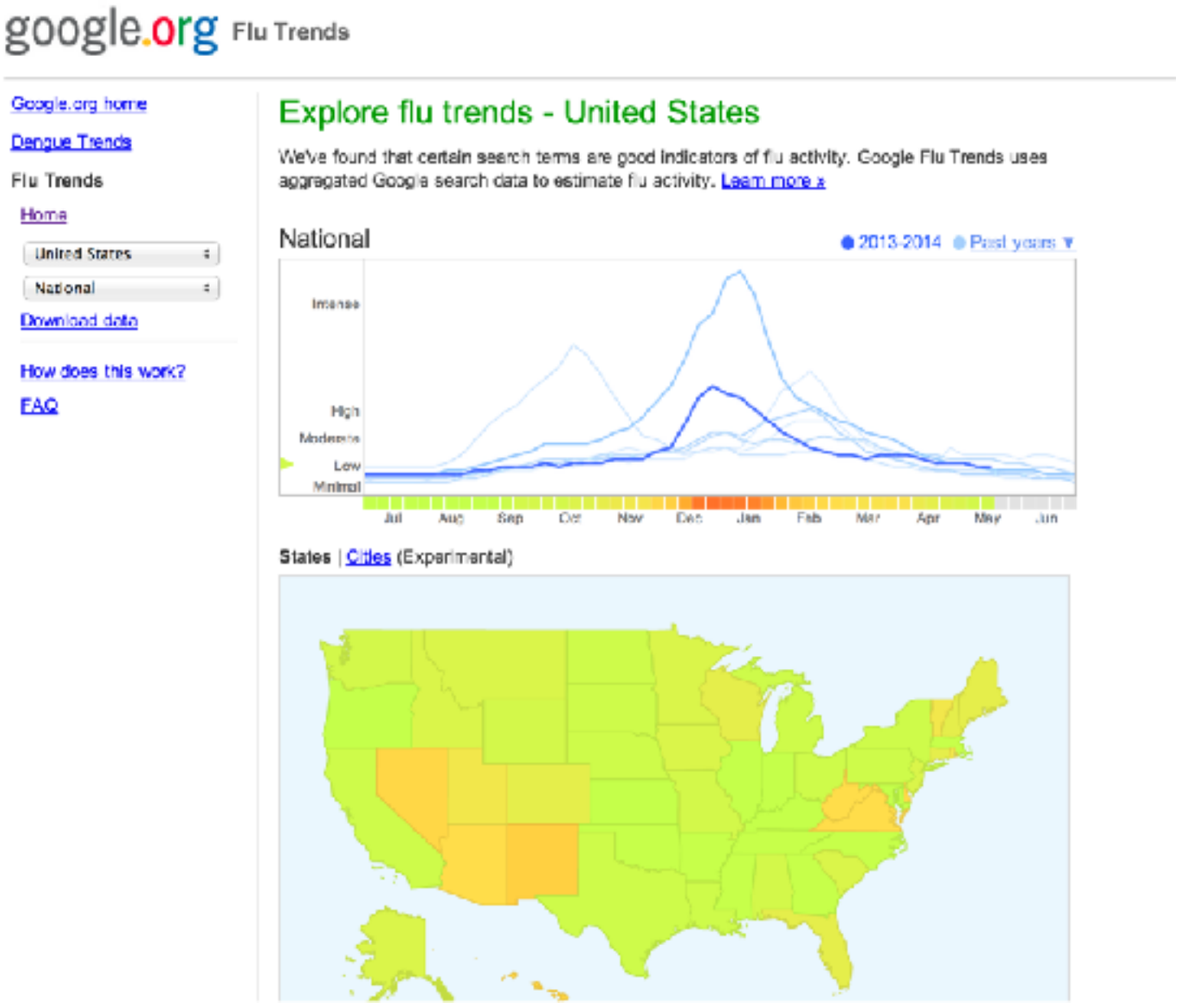
Mobile Devices + Twitter Use

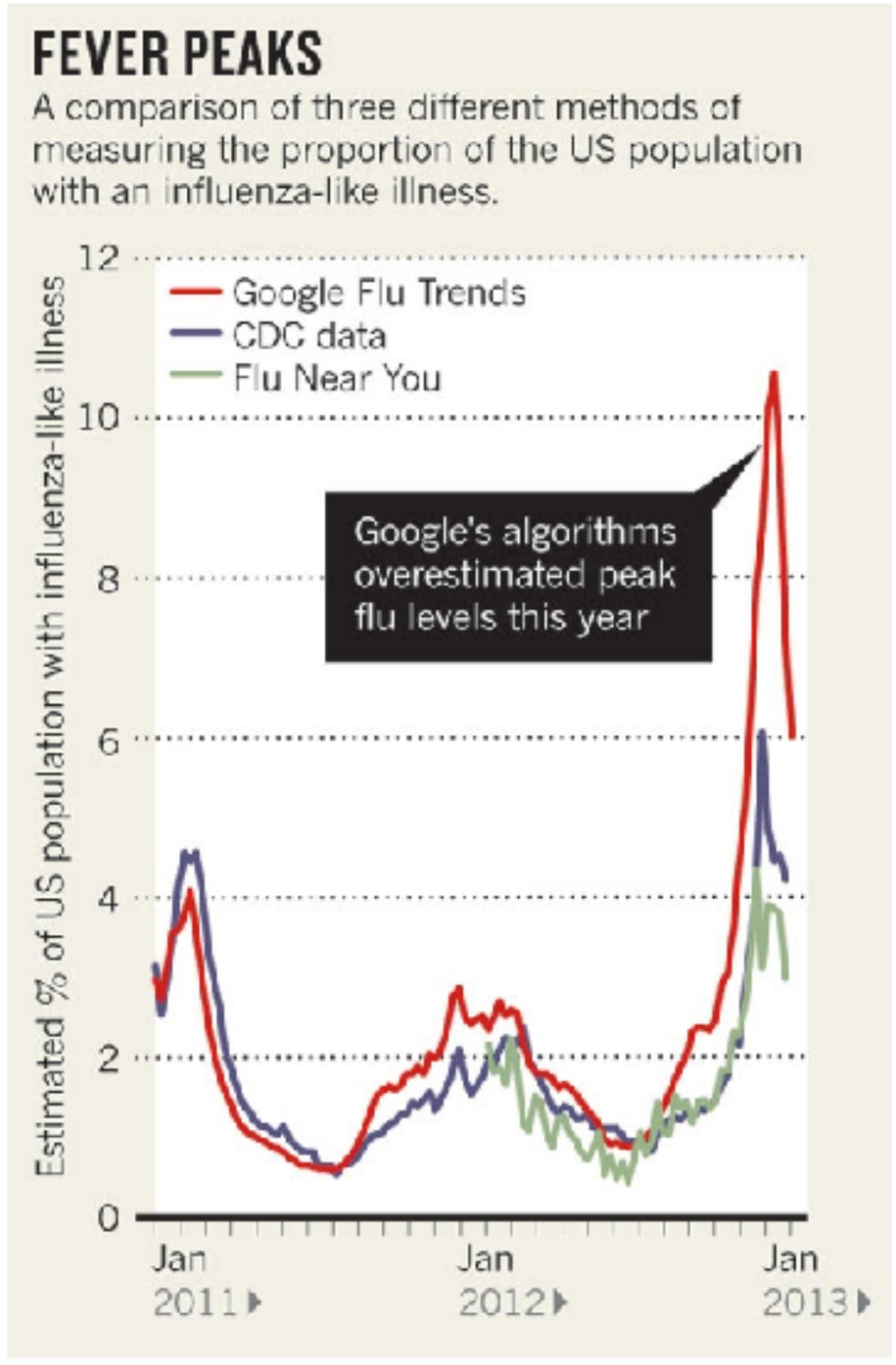
More than 280 million Tweets posted from mobile phones reveal geographic usage patterns in unprecedented detail.

Android | iPhone | Blackberry | Other

Eric Fischer (Twitter Map)

Complex problems, unstable environment





UNIVERSITY OF LIVERPOOL

http://www.nature.com/news/when-google-got-flu-wrong-1.12413

Software Systems

http://upload.wikimedia.org/wikipedia/commons/2/23/Noahs_Ark.jpg

Anything that won't fit on a spreadsheet!

Mike Batty, UCL

UNIVERSITY OF LIVERPOOL

PART 2 SOCIAL AREAS AND CLUSTERS

33 The city of Liverpool is shown divided into five types of area in figure 2

a high status area of owner occupied houses with stable families

a rooming house area of subdivided houses providing furnished privately rented accommodation chiefly for young people

the inner council estates, mainly though not exclusively the older blocks of flats

the outer council estates, mainly houses and newer blocks of flats

an area of older terraced housing, mostly unfurnished privately rented houses many lacking an inside wc.

34 The key characteristics of each area are shown in table 4, expressed as percentages of the average for the city as a whole. Thus the ratio to the city mean for owner occupied households in the high status area is 236, meaning that 71% of all households in that area are owner occupiers as the average in the city as a whole is 30%.

35 The relative size of each social area is shown in table 5. The second and third are smaller than the others and the fourth has the largest population, the largest number of clusters and also the largest number of basic data areas. Furthermore, the average population size of basic data areas is larger in the fourth area and particularly small in the second. This implies that the outer council estates contain large tracts of housing which are socially homogeneous whereas the rooming house area is more sharply split up into small areas with different social characteristics.

36 The social areas are most strongly related with the pattern of housing tenure in the city in two important respects. Each area as a whole is characterised by a single type of tenure not only in the composition of its own housing stock but also in the extent to which it contains within its boundaries a large proportion of all houses in that category in the city. Thus the high status area contains over half the city's owner occupied houses and virtually all of its newer houses in that category. The rooming house and older terraced areas split the rest of the private houses between them, the furnished rented in one and the unfurnished rented and most of the remaining (mainly older) owner occupied houses in the other. And the inner and outer council areas contain between them virtually all of the city's council housing.

37 Another important feature of the system is that to a very large extent each social area is comparatively homogeneous in the make up of its housing stock. That is, all clusters and basic data areas in a given area are pretty similar in their housing composition, as is shown in table 1. No other characteristic is so strongly related with the system as a whole, since for no other does the amount of variation retained stay as high as for housing tenure.

38 The social areas are comparatively poorly related to the distribution of social classes in the city in the sense that

Table 4 Social areas: general characteristics

	city mean (%)	social areas (city = 100) 1	2	3	4	5
housing						
owner occupied	*30.1*	236	87	7	34	107
council tenant	*40.0*	14	21	222	211	21
private unfurnished	*26.3*	79	158	30	19	218
private furnished	*3.5*	78	674	29	14	53
shared dwelling	*3.3*	76	508	57	35	75
lacking inside wc	*22.7*	17	44	29	46	273
7 or more rooms	*8.8*	233	180	27	38	66
1 or 2 room	*6.5*	44	485	161	60	44
over 1.5 persons/room	*2.7*	16	170	363	101	55
under 0.5 persons/room	*33.1*	138	83	57	83	110
rooms/person	*1.64**	122	101	71	88	106
rooms/dwelling	*5.01**	115	84	80	98	102
socio-economic status						
professional/managerial	*10.9*	253	124	31	53	49
non manual	*19.6*	169	131	48	81	75
skilled manual	*34.2*	79	89	66	115	114
semi skilled	*20.8*	46	99	126	114	118
unskilled	*14.4*	19	68	265	106	113
age/household structure						
aged 0-4	*7.8*	87	109	99	84	127
5-14	*17.4*	87	75	119	115	94
15-14	*16.2*	85	135	109	104	92
25-44	*21.6*	107	106	95	91	105
45-64	*24.5*	108	88	93	103	97
65+	*12.6*	117	96	86	94	99
new commonwealth born	*0.8*	70	495	133	18	80
over 15, married	*60.7*	106	90	85	99	106
2 adults, 5+ children	*2.1*	46	58	197	150	65
single non pensioner	*7.0*	82	298	109	60	93
5 year migrant	*27.8*	94	142	97	112	77
persons/household	*3.13**	92	82	111	111	95
education/employment						
students	*3.4*	148	183	51	90	62
HNC/degree	*5.6*	280	194	16	38	28
mining/manufacturing	*35.8*	74	85	92	115	110
services/government	*43.6*	131	118	96	86	88
mar female econ active	*44.3*	98	104	102	102	97
male, unemployed	*9.1*	36	126	201	108	99
male, sick	*1.9*	41	130	224	93	104
travel						
walk to work	*17.4*	65	103	193	82	118
bus to work	*48.9*	72	98	108	115	102
car to work	*23.7*	181	97	18	86	81
cars/person	*0.12**	189	105	20	81	77

*ratio

Table 5 Distribution of population, tenure and status by social area (%)

	social areas 1	2	3	4	5	city total
population	22	9	9	33	27	100
tenure						
owner occupied	51	8	1	11	29	100
council tenant	3	2	20	69	6	100
private unfurnished	17	14	3	6	60	100
private furnished	17	61	3	5	15	100
socio-economic status						
professional/managerial	55	11	3	18	14	100
non manual	37	12	4	27	21	100
skilled manual	17	8	6	38	31	100
semi skilled	10	9	11	37	32	100
unskilled	4	6	24	35	31	100

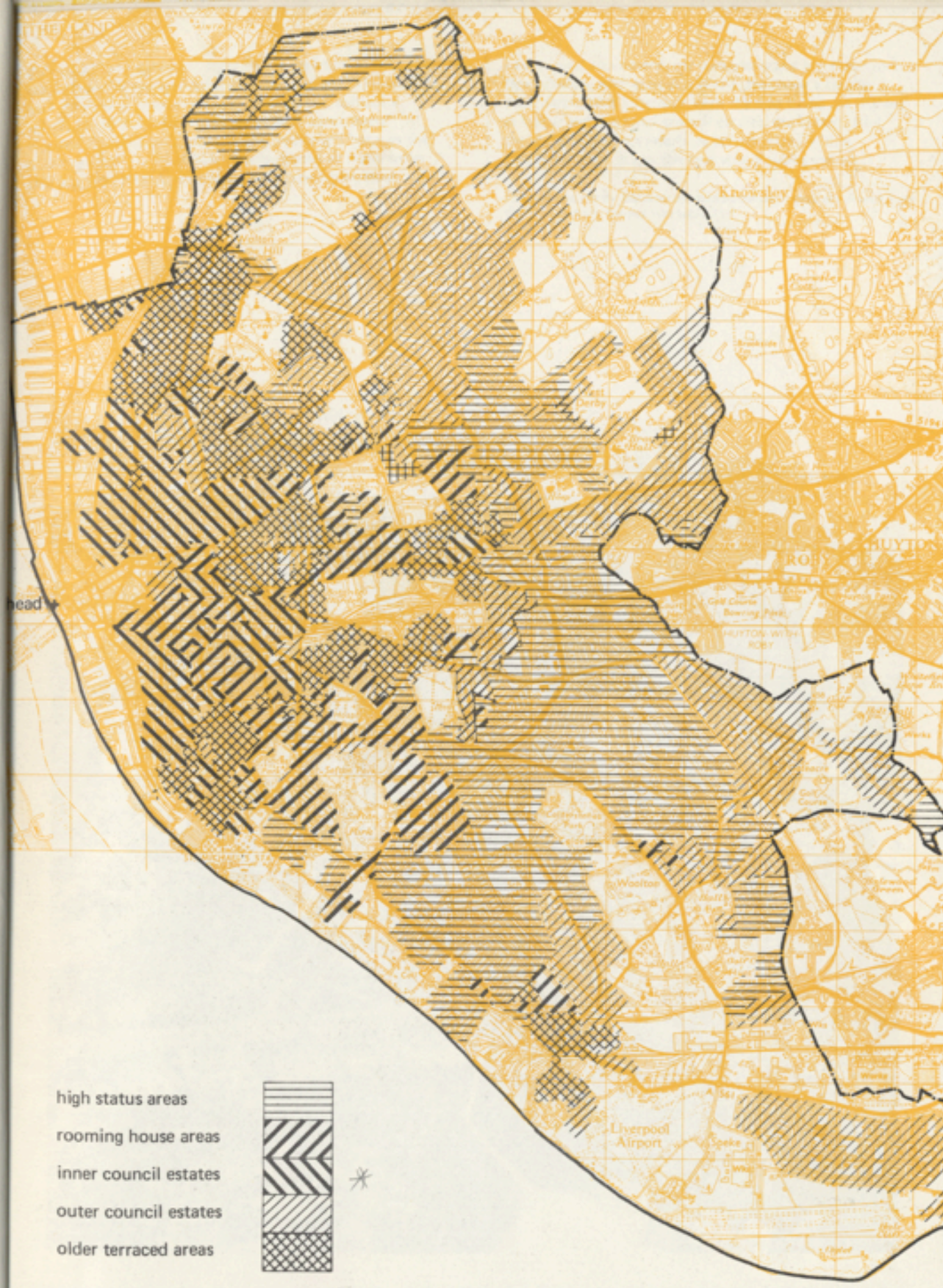


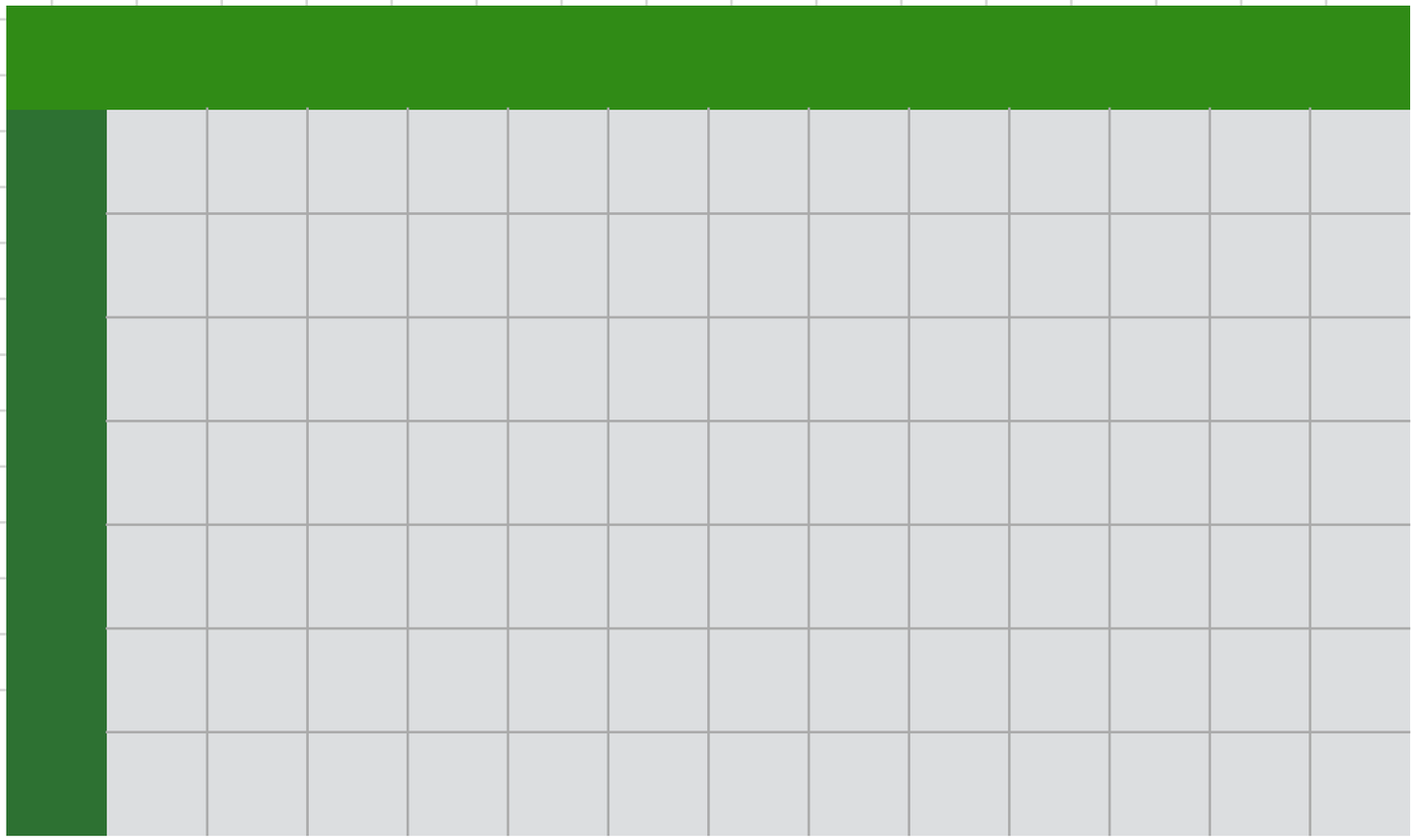
Figure 2 Social Areas

Attributes

Areas

PCA

"Pre" clustering - contiguity constraint



UNIVERSITY OF LIVERPOOL

Aggregate

Decades

Census 2001

Census 2011

Census 2022

Release 2003

Admin. Data

Open Data

Individual

Seconds

Social Media.Data

Closed Data

Business Data

? Data

UNIVERSITY OF LIVERPOOL

Consumer Data Research Centre

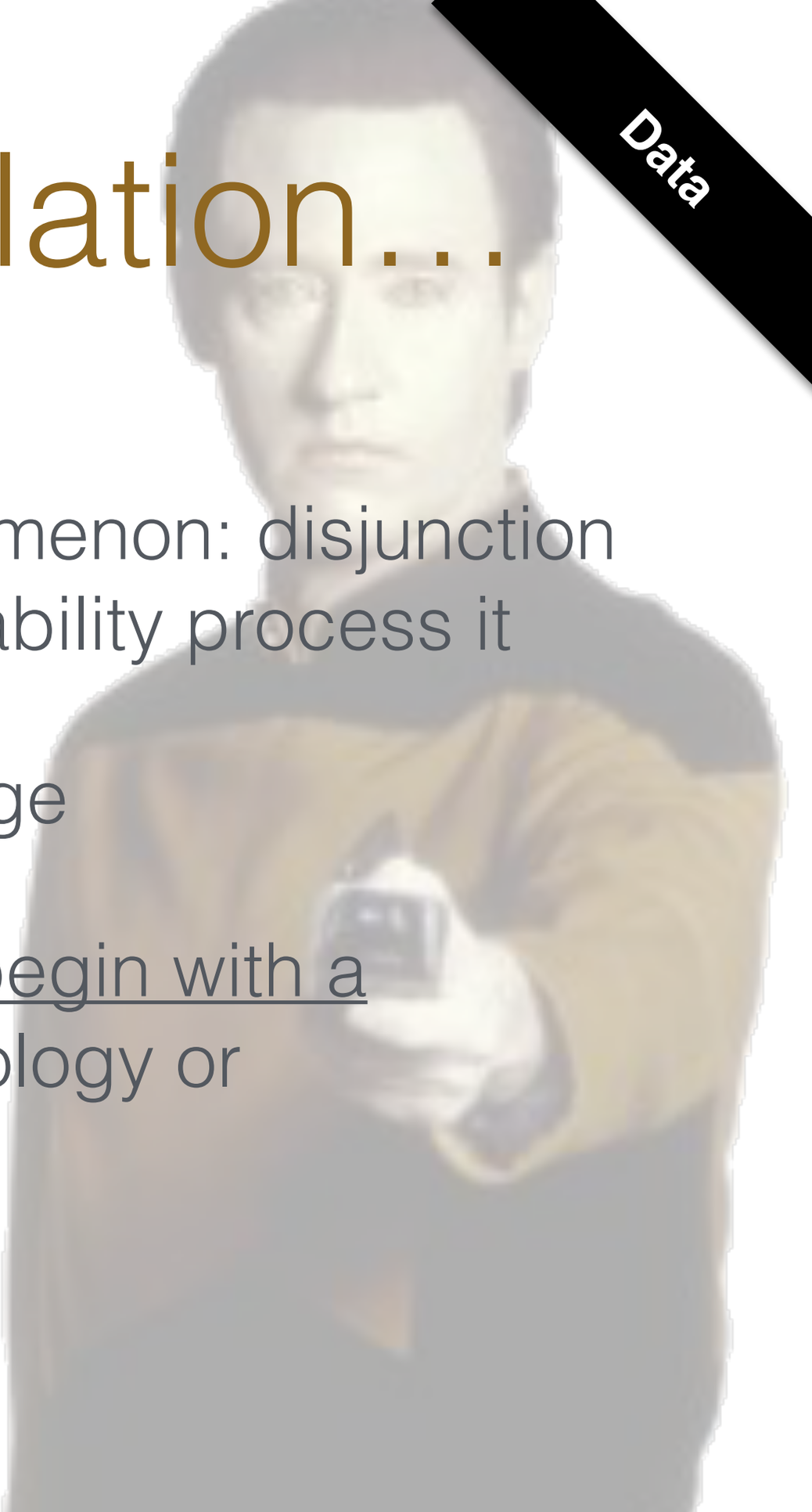
An ESRC Data Investment

Grand Challenges

- Of what are these new data **representative**?
- What should be **captured** (or not) and how?
- What are the **ethical / legal** considerations - privacy / surveillance?
- What new problems can be explored through **imaginative use** of data and software?

UNIVERSITY OF LIVERPOOL

Some Speculation…



- Big Data is **not** a **new** phenomenon: disjunction between available data and ability process it

- **New methodology** will emerge

- **Great Opportunity**: Should begin with a problem to solve not a technology or infrastructure

UNIVERSITY OF LIVERPOOL

Some speculation…

Market

- Opensource GIS
  - Reduce market share commercial desktop GIS
  - Commercial GIS, refocus on cloud - services

theguardian

News | Sport | Comment | Culture | Business | Money | Life & style | Travel | Environment

News › Technology › Microsoft

UK government plans switch from Microsoft Office to open source

Cabinet Office minister Francis Maude plans to standardise on open formats to cut costs on Office suite and break 'oligopoly' of IT suppliers

Share
Tweet this
+1 1.1k
Email

Press Association
theguardian.com, Wednesday 29 January 2014 09.46 GMT
Jump to comments (…)

Microsoft Office costs the UK government significant amounts every year, says

Technology
Microsoft · Open source · Google

Politics
Francis Maude

More news

More on this story

Francis Maude plays down universal credit IT row with DWP
Minister says it is normal for Cabinet Office to provide support to departments

UNIVERSITY OF LIVERPOOL

http://www.theguardian.com/technology/2014/jan/29/uk-government-plans-switch-to-open-source-from-microsoft-office-suite
https://www.whatdotheyknow.com/request/133909/response/323829/attach/3/RESPONSE%2028509371.pdf

Government pays Microsoft £5.5m to extend Windows XP support

The UK government is paying Microsoft £5.5m of taxpayers' money to extend support for the out-of-date Windows XP operating system

Support for Windows XP ends on 8 April 2014 Photo: PA

By Sophie Curtis
11:14AM BST 03 Apr 2014

Print this article

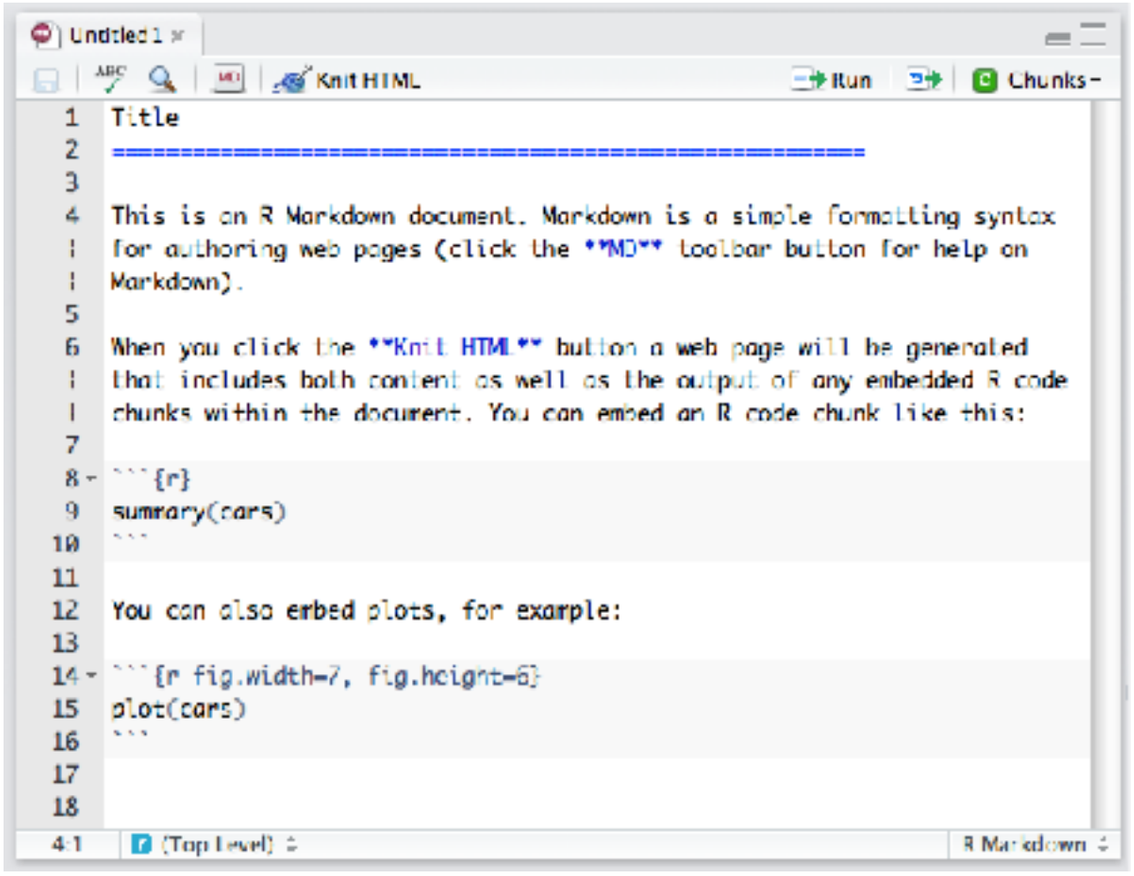
The Telegraph
Unlimited website access
START NOW

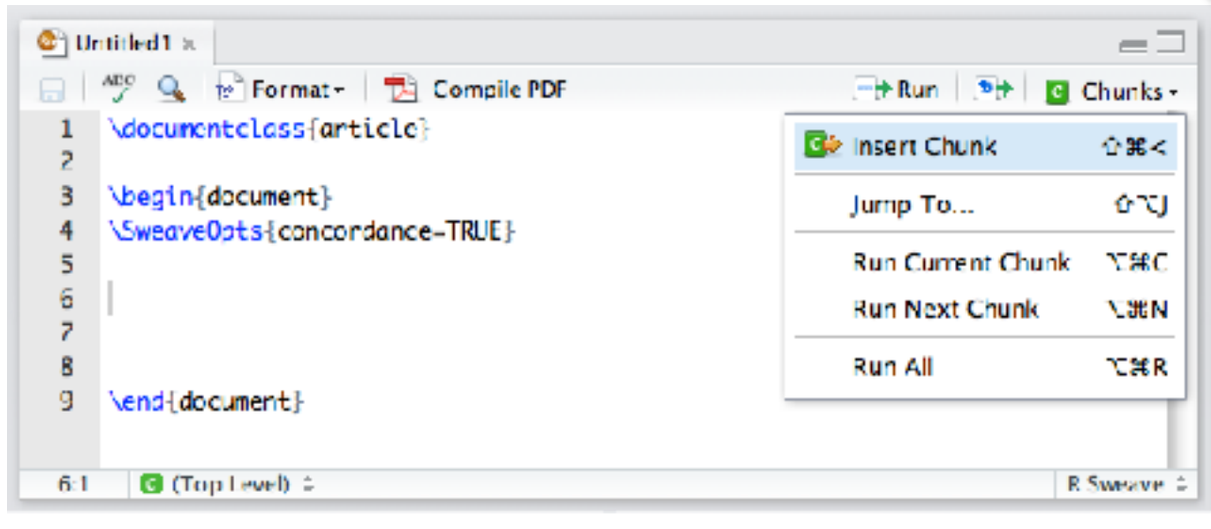
£43 tariff
£39.99 upfront
5GB data
Be more dog
Get yours now
Terms apply
iPhone 5 S
O2

Promotions »

Some Speculation…

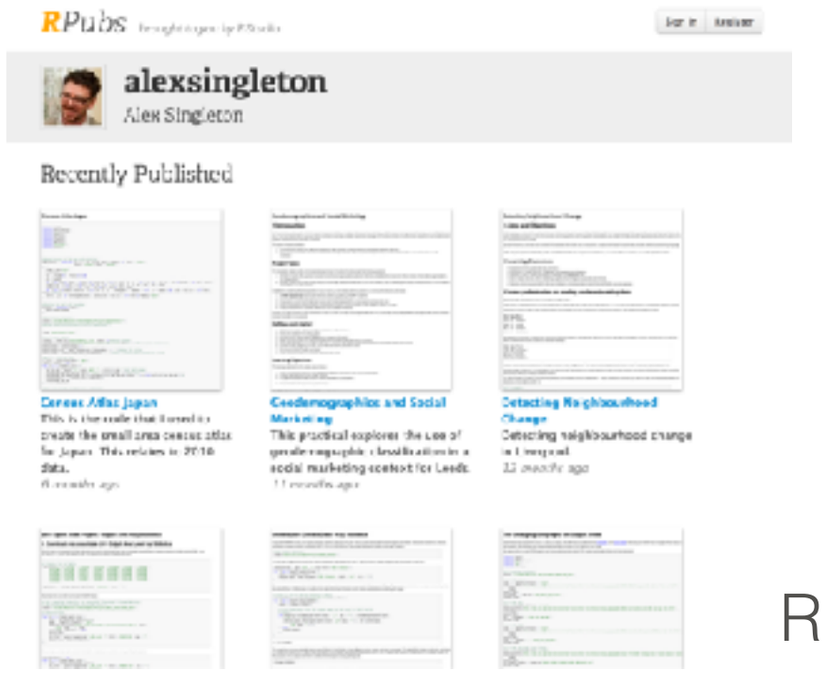
Reproducibility





Markdown (.Rmd)

Sweave (.Rnw)



Rpubs

UNIVERSITY OF LIVERPOOL

Some Speculation...

Education

Learning to code

ALEX SINGLETON

is a lecturer in geographic information science at the University of Liverpool

IN MY OPINION, a geography curriculum should require students to learn how to code, ensuring that they're equipped for a changed job market that's increasingly detached from geographic information systems (GIS) as they were originally conceived.

The ability to code relates to basic programming and database skills that enable students to manipulate large and small geographic data sets, and to analyse them in automated and transparent ways. Although it might seem odd for a geographer to want to learn programming languages, we only have to look at geography curriculums from the 1980s to realise that these skills used to be taught.

For example, it wouldn't have been unusual for an undergraduate geographer to learn how to programme a basic statistical model (for example, regression) from base principles in Fortran (a programming language popular at the time) as part of a methods course.

But during the 1990s, the popularisation of graphical user interfaces in software design enabled many statistical, spatial analysis and mapping operations to be wrapped up within visual and menu-driven interfaces, which were designed to lower the barriers of entry for users of these techniques. Gradually, much GIS teaching has transformed into learning how these software systems operate, albeit within a framework of geographic information science (GISc) concerned with the social and ethical considerations of building representations from geographic data. Some Masters degrees in GISc still require students to code, but few undergraduate courses do so.

The good news is that it's never been more exciting to be a geographer. Huge volumes of spatial data about how the world looks and functions are being collected and disseminated. However, translating such data safely into useful information is a complex task.

During the past ten years, there has been an explosion in new platforms through which geographic data can be processed and visualised. For example, the advent of services such as Google Maps has made it easier for people to create geographical representations online.

However, both the analysis of large volumes of data and the use of these new methods of representation or analysis do require some level of basic programming ability. Furthermore, many of these developments haven't been led by geographers, and there's a real danger that our skill set will be seen as superfluous to these activities in the future without some level of intervention.

Indeed, it's a sobering experience to look through the pages of job advertisements for GIS-type roles in the UK and internationally. Whereas these might once have required knowledge of a particular software package, they increasingly look like advertisements for computer scientists, with expected skills and experience that wouldn't traditionally be part of an undergraduate geography curriculum.

Many of the problems that GIS set out to address can now be addressed with mainstream software or shared online services that are, as such, much easier to use. If I want to determine the most efficient route between two locations, a simple website query can give a response within seconds, accounting for live traffic-volume data. If I want to view the distribution of a census attribute over a given area, there are multiple free services that offer street-level mapping. Such tasks used to be far more complex, involving specialist software and technical skills.

There are now far fewer job advertisements for GIS technicians than there were ten years ago. Much traditional GIS-type analysis is now sufficiently non-technical that it requires little specialist skill, or has been automated through software services, with a subscription replacing the employment of a technician. The market has moved on.

Geographers shouldn't become computer scientists; however, we need to reassert our role in the development and critique of existing and new GIS. For example, we need to ask questions such as which type of geographic representation might be most appropriate for a given dataset. Today's geographers may be able to talk in general terms about such a question, but they need to be able to provide a more effective answer that encapsulates the technologies that are used for display. Understanding what is and isn't possible in technical terms is as important as understanding the underlying cartographic principles. Such insights will be more available to a geographer who has learnt how to code.

Within the area of GIS, technological change has accelerated at an alarming rate in the past decade and geography curriculums need to ensure that they embrace these developments. This does, however, come with challenges. Academics must ensure that they are up to date with market developments and also that there's sufficient capacity within the system to make up-skilling possible.

Prospective geography undergraduates should also consider how the university curriculums have adapted to modern market conditions and whether they offer the opportunity to learn how to code.

UNIVERSITY OF LIVERPOOL

www.geographical.co.uk

The Course…

- Practicals
  - Introduction to R
  - Data Manipulation in R
  - Descriptive statistics
  - Charts and graphs
  - Mapping areas and contexts
  - Visualising point patterns
  - Mapping flows

- **For a city anywhere in the world; utilize those skills acquired during this course to create a data driven profile that describes:**
  - **Some aspects of population structure**
  - **An attribute that varies over both space and time**
  - **The spatial distribution of a social outcome**

UNIVERSITY OF LIVERPOOL

www.alex-singleton.com/GDS_UA_2017/

Many Thanks

UNIVERSITY OF LIVERPOOL