

# Data, Politics and Society

## Module overview



# About this Module

Data, Politics and Society provides an interdisciplinary introduction to the politics and ethics of using large-scale, primarily human-generated, data, drawing together insights from data science, political geography, digital anthropology, legal studies and sociology.

This module will provide those interested in social and geographic data science as well as data ethics with a comprehensive background in the complexities of using large-scale human generated datasets.

# About this Module

We will be critically looking at data and their usage with the ultimate aim to become a more conscious practitioner.

# Outline

## *Part I: Data and its role in society*

W1

W2

W3



Data: The Good, The Bad, The Ugly

W4

W5



Societal and environmental impacts of data and technology

# Outline

## *Part II: Mitigating the risks of working with large-scale datasets*

W6

W7

Regulations and governance

W8

Crowdsourcing, VGI, and Geographic Citizen Science

W9

*with Dr Artemis Skarlatidou*

W10

Critical Data Studies

# Module structure

- This module consists of ten lectures (Wednesday mornings) and ten interactive seminars (Friday mornings). All sessions last about 50 minutes ('UCL Hour').
- Each week will have its own reading list; available on the dedicated [Data, Politics, and Society](#) webpage (link on Moodle).
- On the same webpage the Lecture slides will be made available prior to each Lecture.

# Assessment

- The assessment for Data, Politics and Society is a 3,000-word commentary piece on the technical, ethical, political and social questions raised by data science and the analysis of large-scale geographic and social datasets and their impacts on society.
- Detailed instructions in the GEOG0163 Coursework Instructions document on Moodle. Examples of last year's submissions with indicative grades also available.
- Submission: 9 January 23

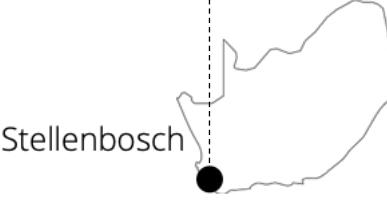
# Communication

- All important information will be communicated through Moodle.
- Questions can be asked after each Lecture and during the Seminars.
- For specific questions or other support, my Academic Support and Feedback (ASF) Office Hours are scheduled on Tuesday mornings (10h00-12h00) and Friday afternoons (13h00-14h00). Slots can be booked through Microsoft Bookings: [\[link\]](#)

# About me



B.Sc. Human Geography and Planning  
M.Sc.. Human Geography and Planning



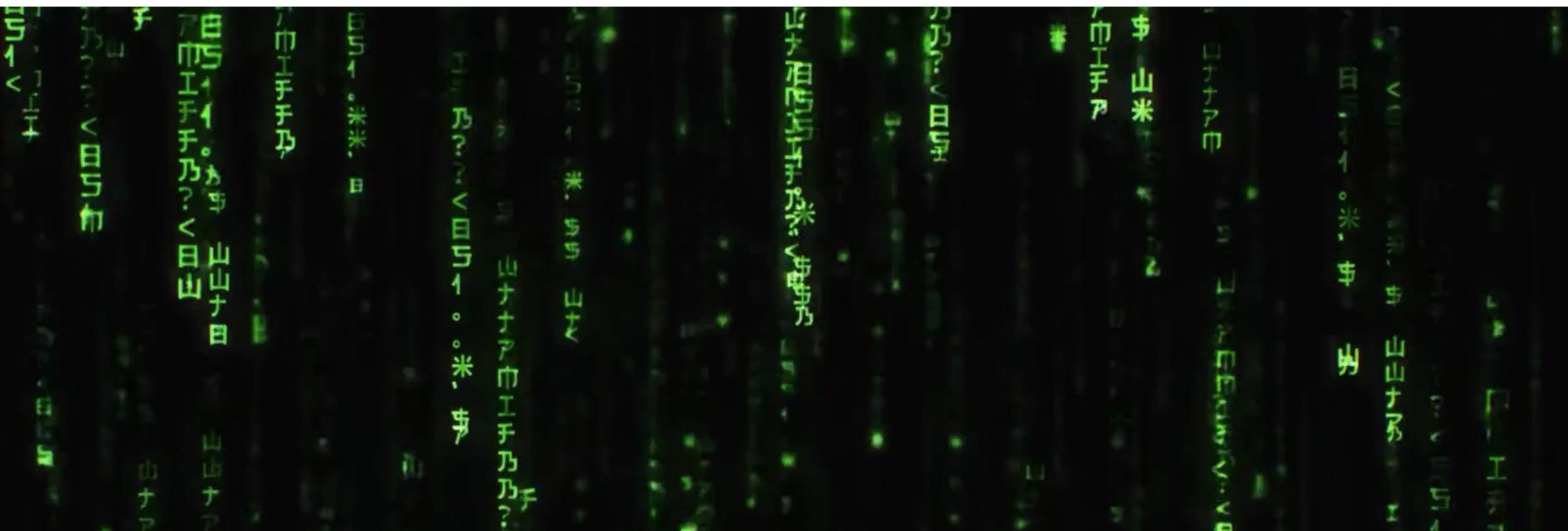
Ph.D. Transport Economics



Postdoc Geospatial Analytics and  
Computing Research Group,  
Lecturer in Social and Geographic  
Data Science

# Data, Politics and Society

# W1 - Data I: The Good



Data: The Good

# Today

- Data in the context of this module
- Datafication
- Examples of recent research with large datasets

# Data

What are data?

# Data

Kitchen 2014:

"Data are commonly understood to be the raw material produced by abstracting the world into categories, measures and other representational forms – numbers, characters, symbols, images, sounds, electromagnetic waves, bits – that constitute the building blocks from which information and knowledge are created." (p.2)

# Data

- Quantitative: numeric, properties of phenomena  
measurement scale (nominal, ordinal, interval ratio)
- Qualitative: non-numeric

# Data

- Structured: data that is easily organised, defined date model
- Semi-structured: loosely structured data, often hierarchical (XML)
- Unstructured: data without an identifiable structure

# Data

- Captured: data that is measured with some instrument (e.g. survey)  
deliberate
- Exhausted: data as a by-product (e.g. transaction data)
- Derived: data derived from other data (e.g. model output)

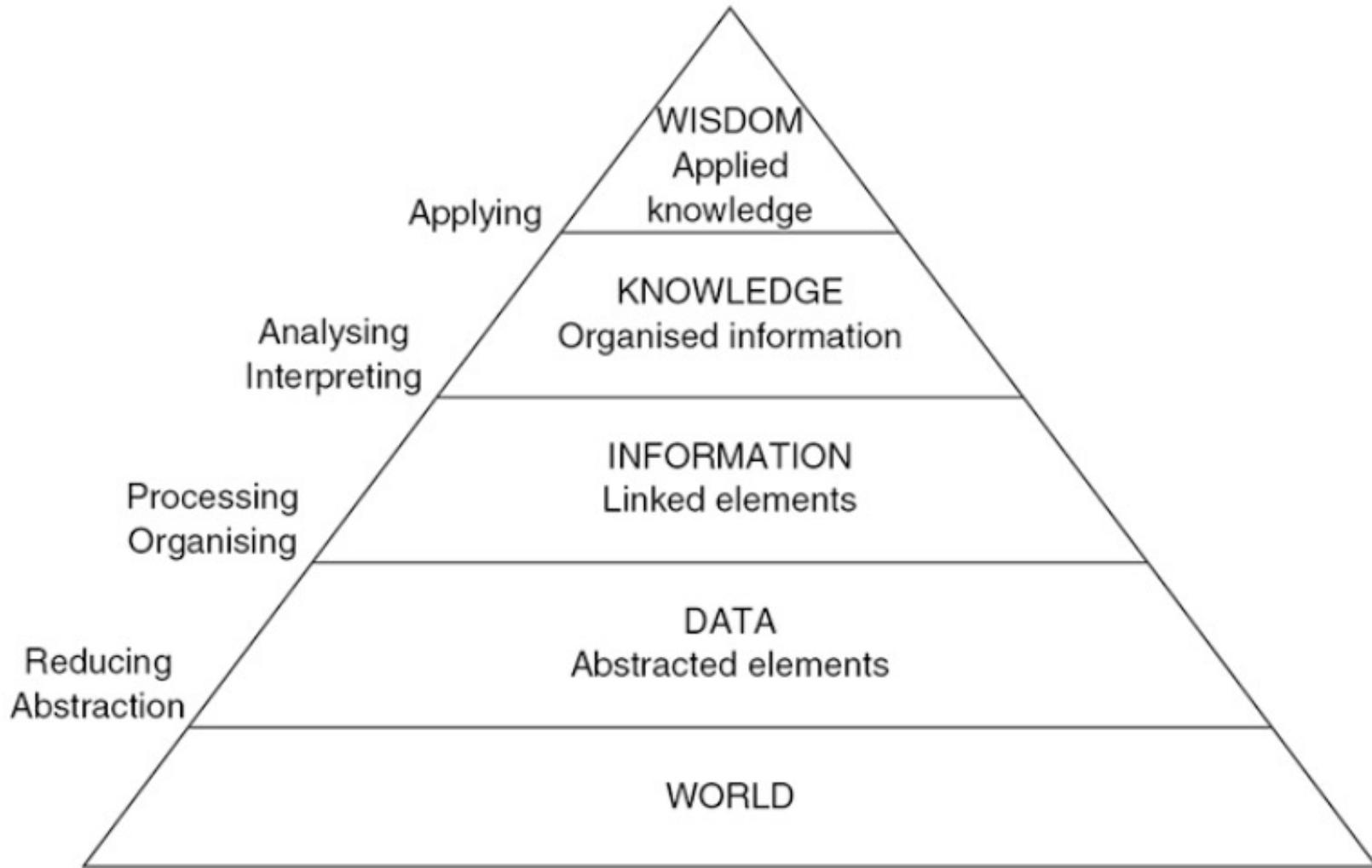
# Data

- Primary: data captured by a researcher, deliberate
- Secondary: "analysis-ready" dataset, pre-processed
- Tertiary: "amplified" data (e.g. geodemographics)

# Data

- Indexical data: data that enable identification (e.g. national insurance number)
- Attribute data: data representing aspects of a phenomenon
- Metadata: data about data (e.g. field descriptions, dataset description)

# Data



Kitchen. 2014. Knowledge pyramid adapted from Adler (1986) and McCandless (2010)

In the context of this module

What types of data will we be talking about?

# Datafication

- Human-generated large-scale datasets come into existence from our direct use of technology which act as "sensors" to our actions and activities.
- Datafication: Our social and daily lives are also quantified in ways not seen before – our friendships, interests, conversations, information searches.
- Why: To gain unprecedented knowledge and “critical insight” into human characteristics and behaviour, e.g. socio-economic statistics, movement and mobility, expenditure patterns, and derive value from these findings.

# Datafication

“Taking all aspects of our lives and turning them into data”

(Cukier and Mayer-Schoenberger, 2013).

# Datafication

Datafication has grown to become an accepted new paradigm  
for understanding sociality and behavior. Should we?

# Human-generated datasets

In the remainder of this lecture, we will explore four examples of different human-generated datasets and research projects which make use of different types of data:

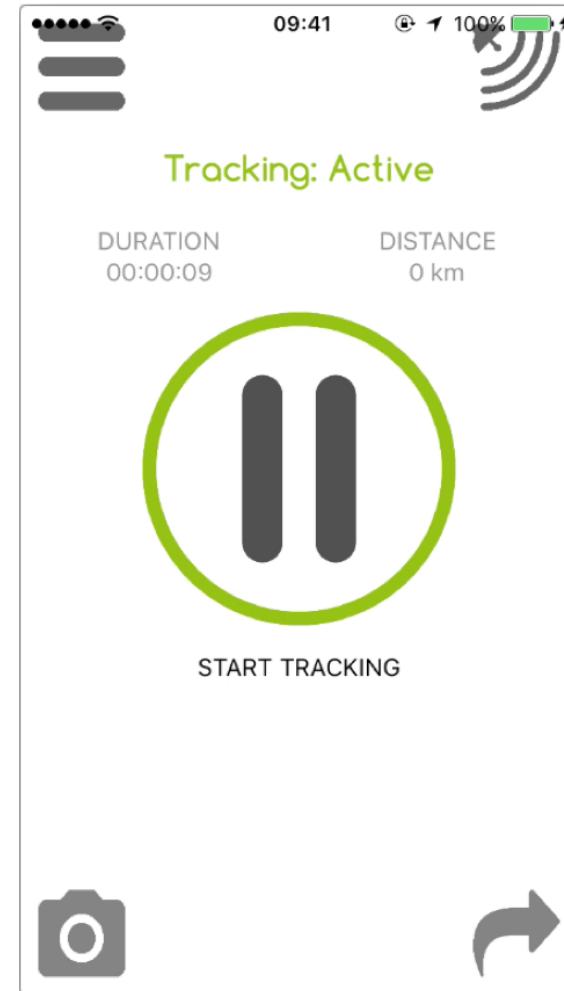
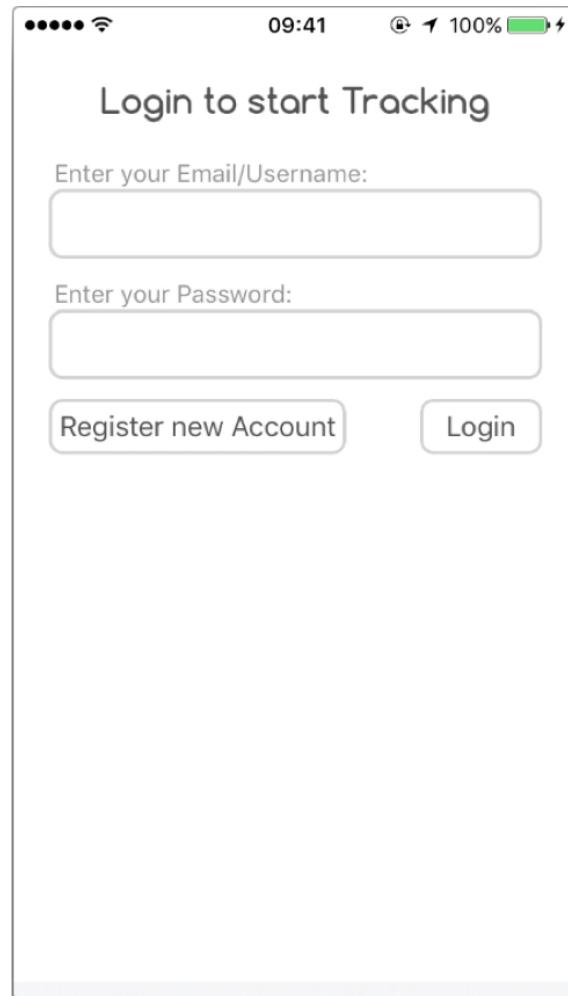
- 1:** GPS data collected through smartphones (SU)
- 2:** Linked Consumer Registers and Residential Mobility (UCL)
- 3:** Data on human mobility obtained through mobile applications (UCL)
- 4:** The Universal Visitation Law of Human Mobility (MIT)

*There is a corresponding article for each of these 'datasets' in your reading list. See link on Moodle to access your reading list.*

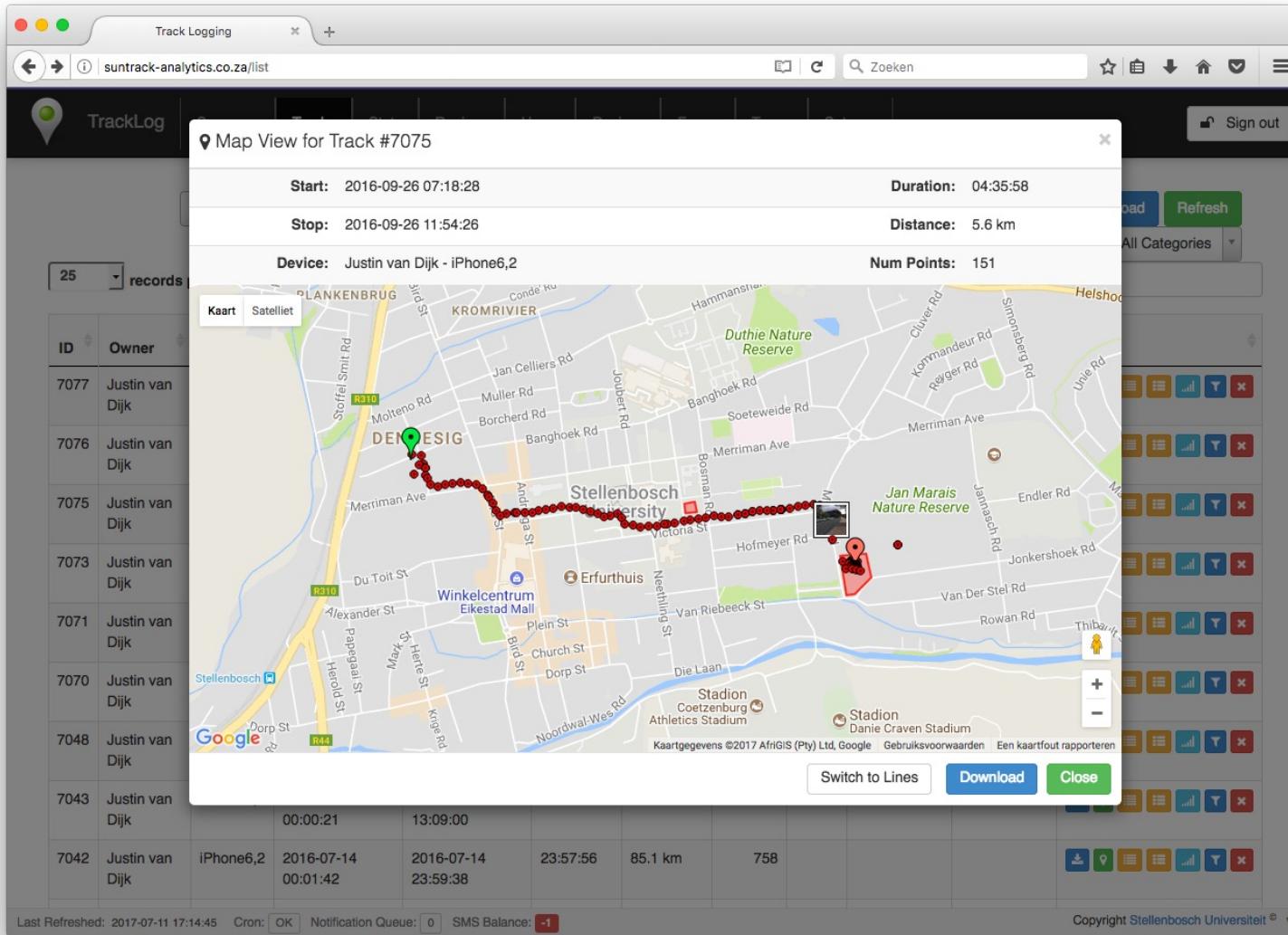
# Capturing daily mobility with GPS

- Van Dijk and Krygsman 2018
- Travel demand management to mitigate negative externalities of fossil-fuel based travel.
- Category of 'soft measures': Voluntary Travel Behaviour Change Programmes.
- "The objective of these programmes is to allow people to choose to change travel behaviour rather than to expect or force reactions in response to external stimuli or pressures."
- How to measure? Small expected changes, limitations of travel surveys and dairies.

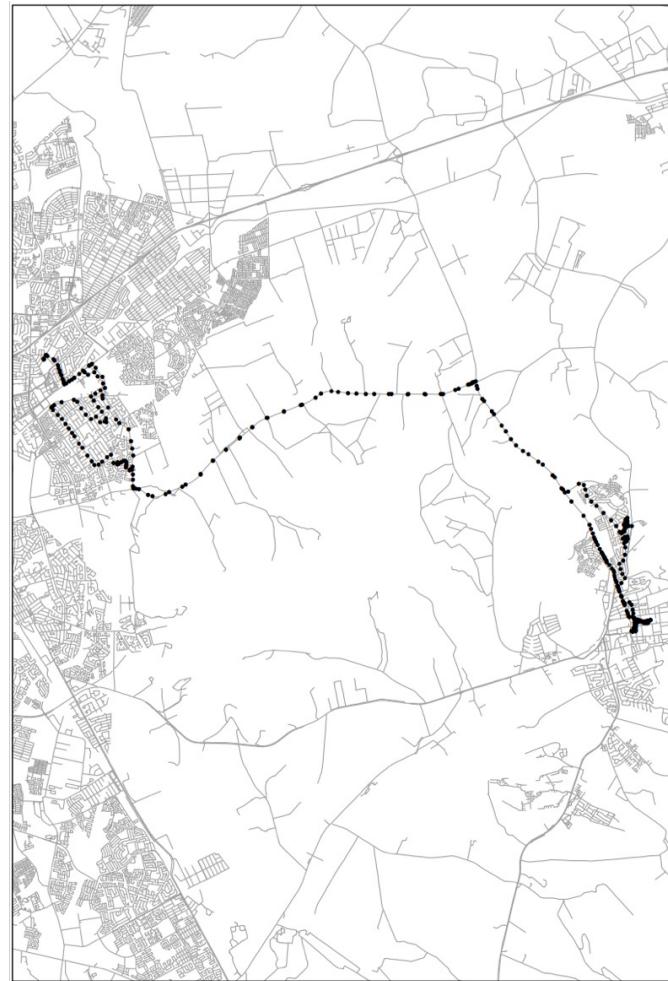
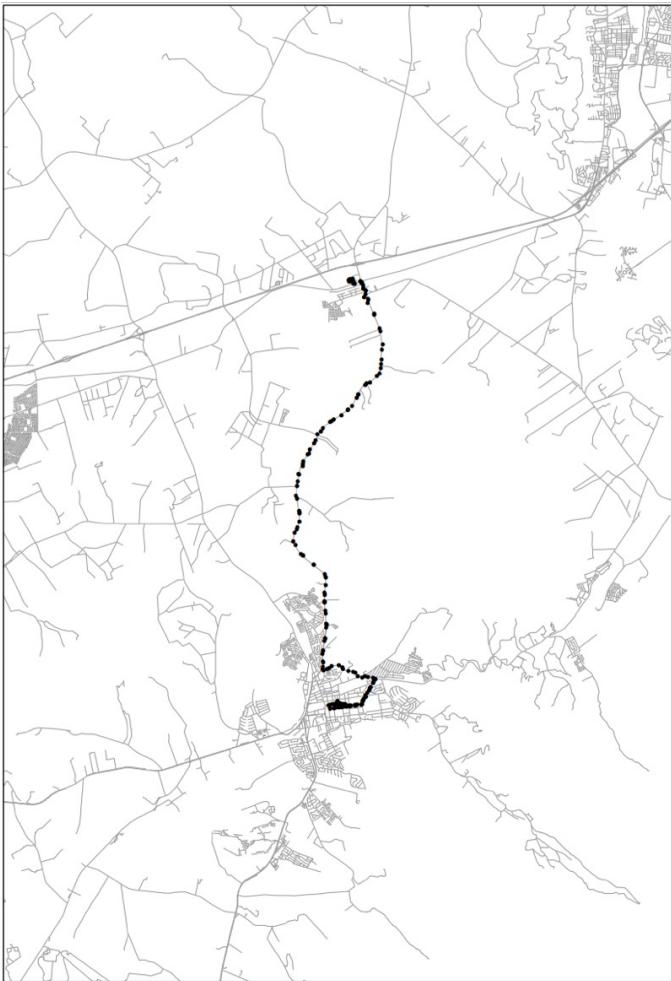
# Capturing daily mobility with GPS



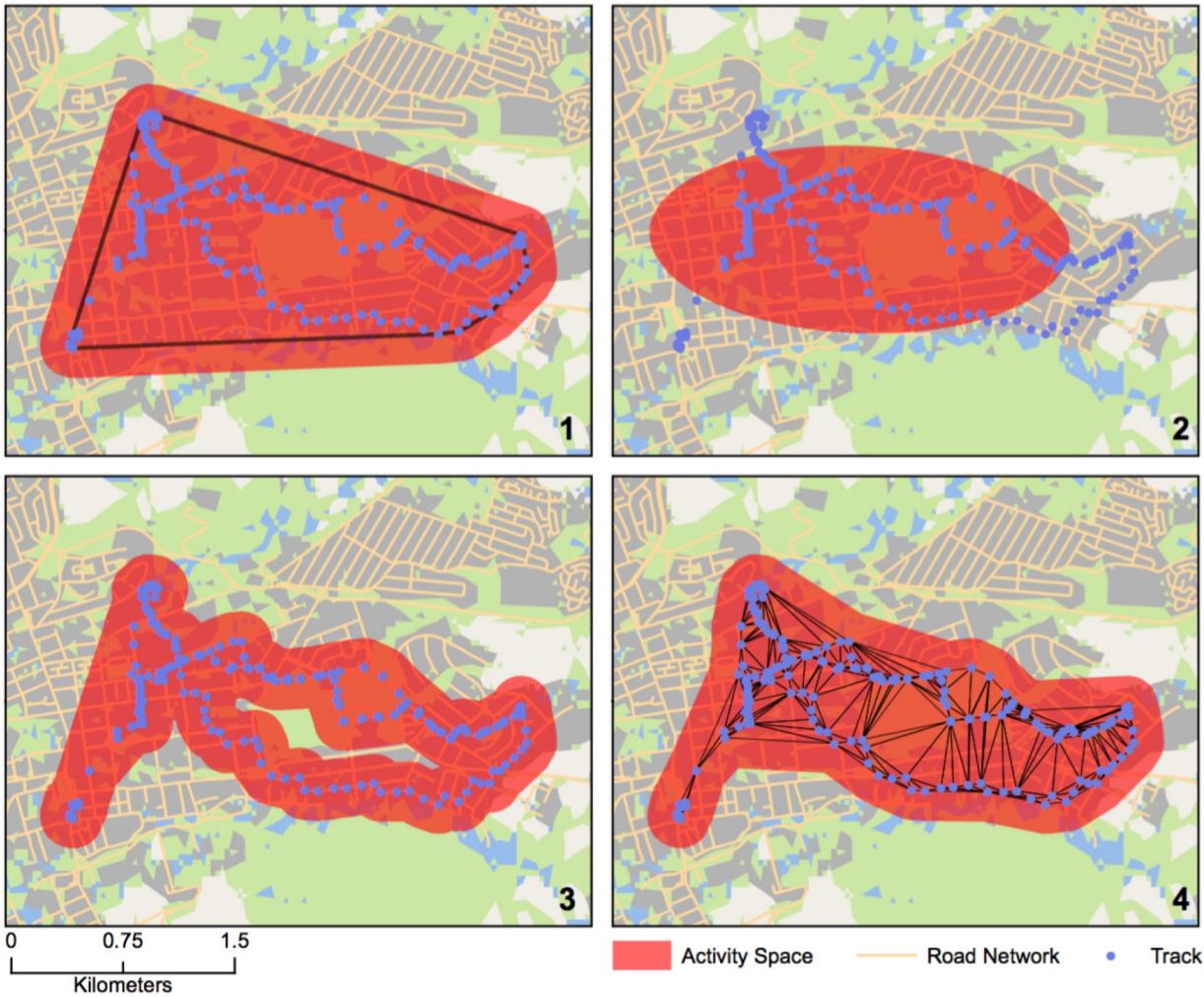
# Capturing daily mobility with GPS



# Capturing daily mobility with GPS



# Capturing daily mobility with GPS



# Capturing daily mobility with GPS

- Measuring an individual's spatial footprint ("revealed spatial behaviour").
- Using activity spaces as a way of defining an individual's neighbourhood.
- Relate these activity spaces as a way of conceptualising accessibility to opportunities ("travel as derived demand").

# Capturing human mobility with mobile applications

- Trasberg and Cheshire 2021
- Data on human mobility obtained from mobile applications to explore the activity patterns in London after the first wave of COVID-19 lockdown restrictions.
- The authors link spatially aggregated mobile locations data to the geodemographic classifications, with the aim of identifying socioeconomic characteristics that could explain the differing rates of decline in neighbourhood activity volumes.

# Capturing human mobility with mobile applications

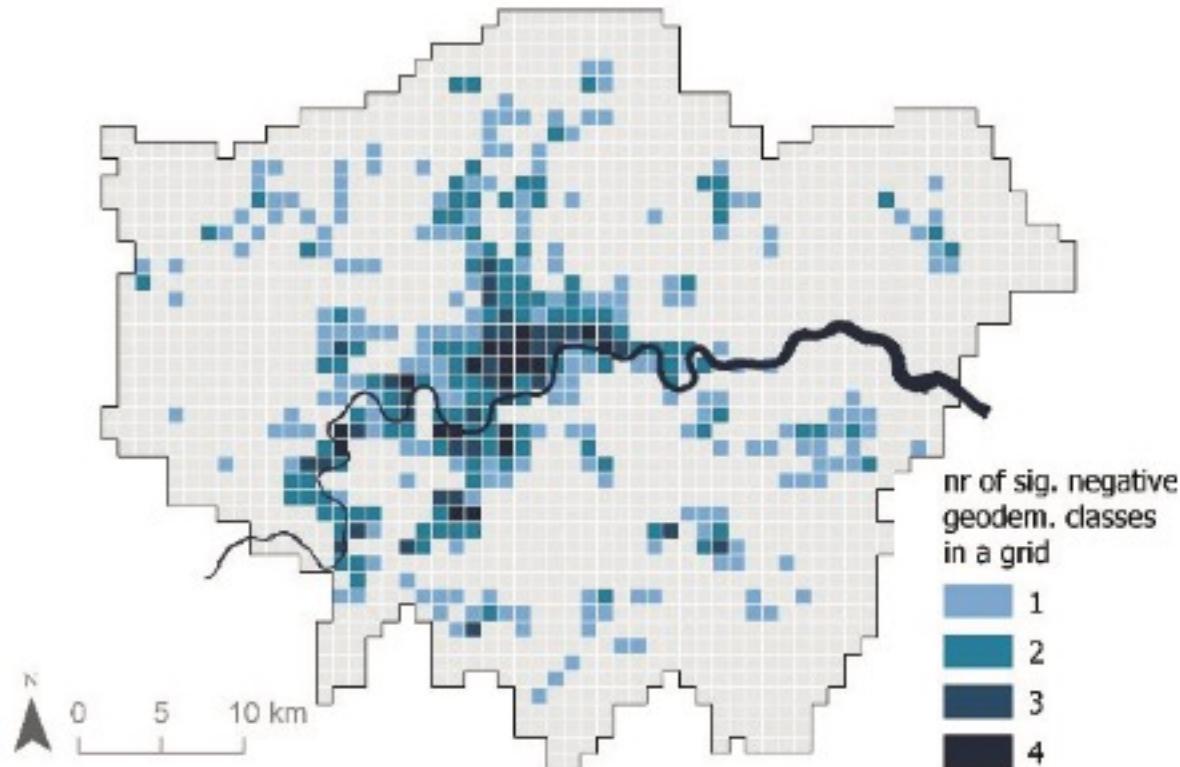
- Location data are collected and stored through a Software Development Kit (SDK) embedded into smartphone apps.
- The location data are further used by app developers for commercial purposes, such as location-based ad targeting, and are monetised by being sold to firms that mine the data for business insights.
- Although the users are given the choice to turn off the location tracking from their mobile devices, the consumers do not necessarily have an indication of when their data are being collected.

# Capturing human mobility with mobile applications

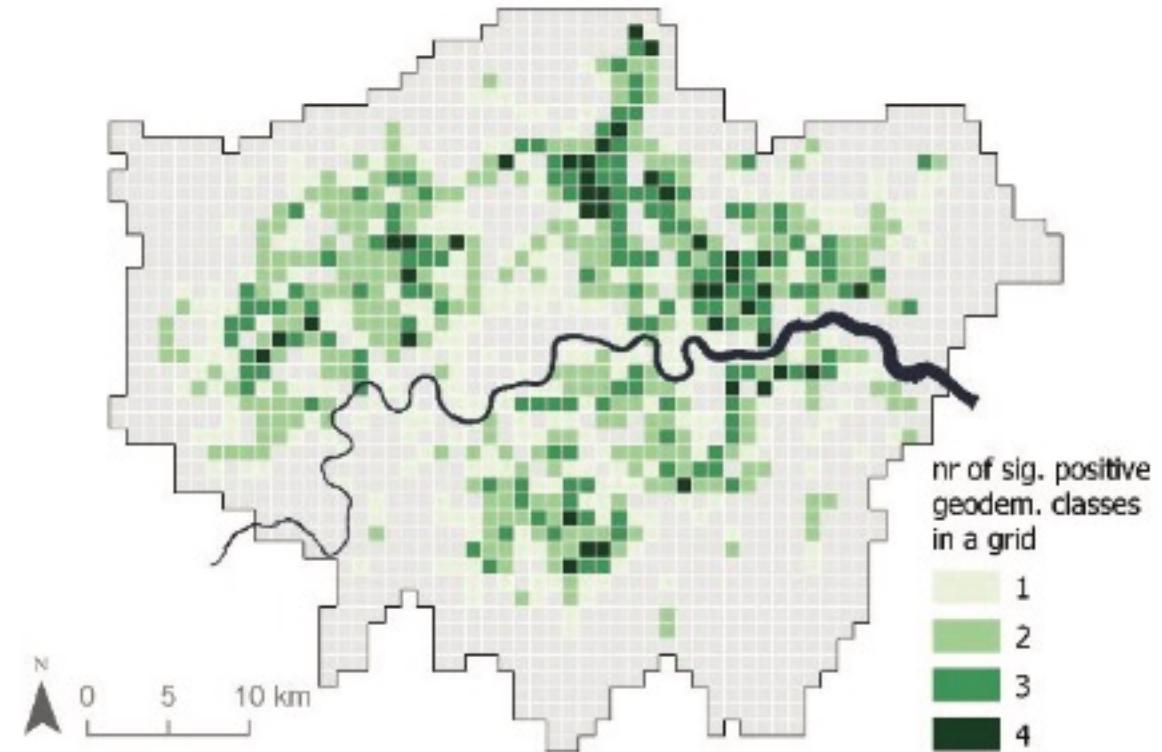
- The location data in this study are spatially aggregated to 1 km<sup>2</sup> grid cells so that we only know the number of unique devices per hour in each grid cell but do not have any information to construct digital traces of any of the devices.
- The location data has been collected from 308,311 unique devices during the study period.
- Results provide insights into the impacts of mobility restrictions in different demographic groups across Greater London.

# Capturing human mobility with mobile applications

Significantly **less decline** than average



Significantly **more decline** than average



# Capturing human mobility with mobile applications

Activity levels in many cities had begun rebounding even before changes to the lockdown had been announced

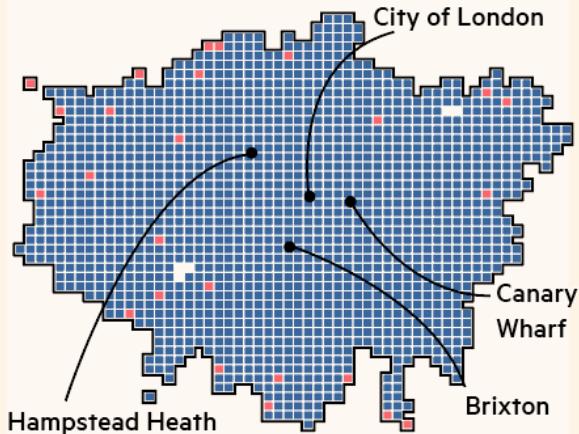
Change in the number of unique devices used in a km<sup>2</sup> area compared with a previous snapshot

Activity rose ■ Activity fell ■

## Peak lockdown:

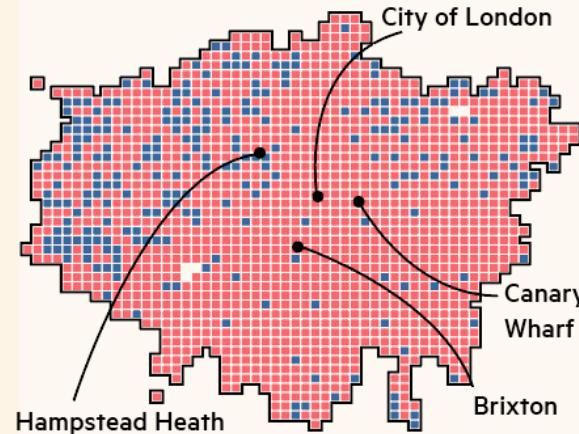
April 13-20 activity levels fell compared with pre-lockdown (March 16-22)

### London



## Anticipating a loosening:

April 30-May 6 activity levels picked up compared with peak lockdown (April 13-20)



The Financial Times. 2020. Britain on the move even before Johnson eased lockdown, data show. [Online] <https://www.ft.com/content/cc70d690-99a6-4056-9ebe-d0b39c40a359> Data visualisation by Terje Trasberg.

# Using consumer data for demographic analysis

- Lansley *et al.* 2019
- Lack of annually updated administrative data sources. Census? Surveys?
- UK-wide population dataset 1997-2020 (LCR v2)
- UK Public Version of the Electoral Roll 1997 – 2020, Consumer Data Sources 2002 - 2017

# Using consumer data for demographic analysis

<b>Forename</b>	<b>Surname</b>	<b>Address</b>
Justin	van Dijk	Flat 18 Terry House SW22NT London

# Using consumer data for demographic analysis

<b>Forename</b>	<b>Surname</b>	<b>Address</b>
Justin	van Dijk	Flat 18 Terry House SW22NT London

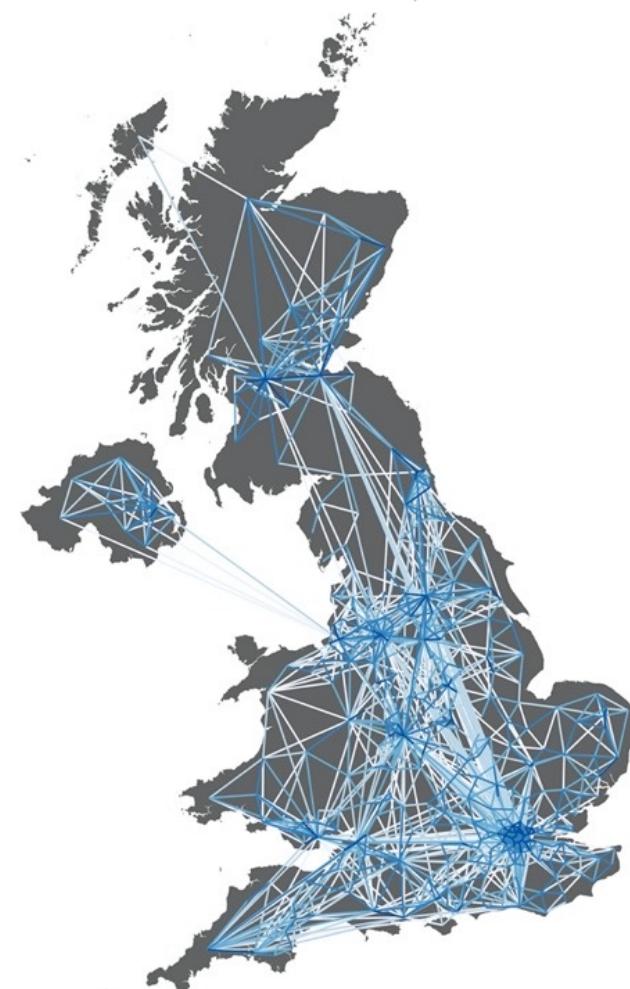
x

804,321,891

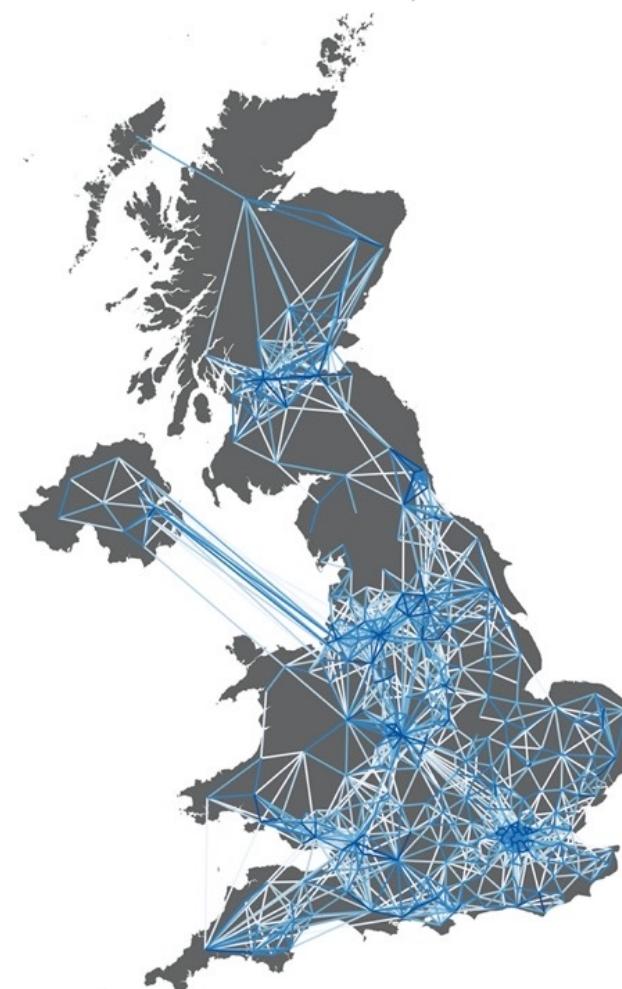
# Using consumer data for demographic analysis

- Linkage of all records to address dataset.
- Fill in the gaps of years without data.
- Reconciliation of likely duplicates – both within and between households.
- Usage: development of household structure indicators, analysing segregation along ethnic lines, intergenerational social mobility through linkage with historic data sources.
- Development of a model to estimate residential mobility.

# Using consumer data for demographic analysis

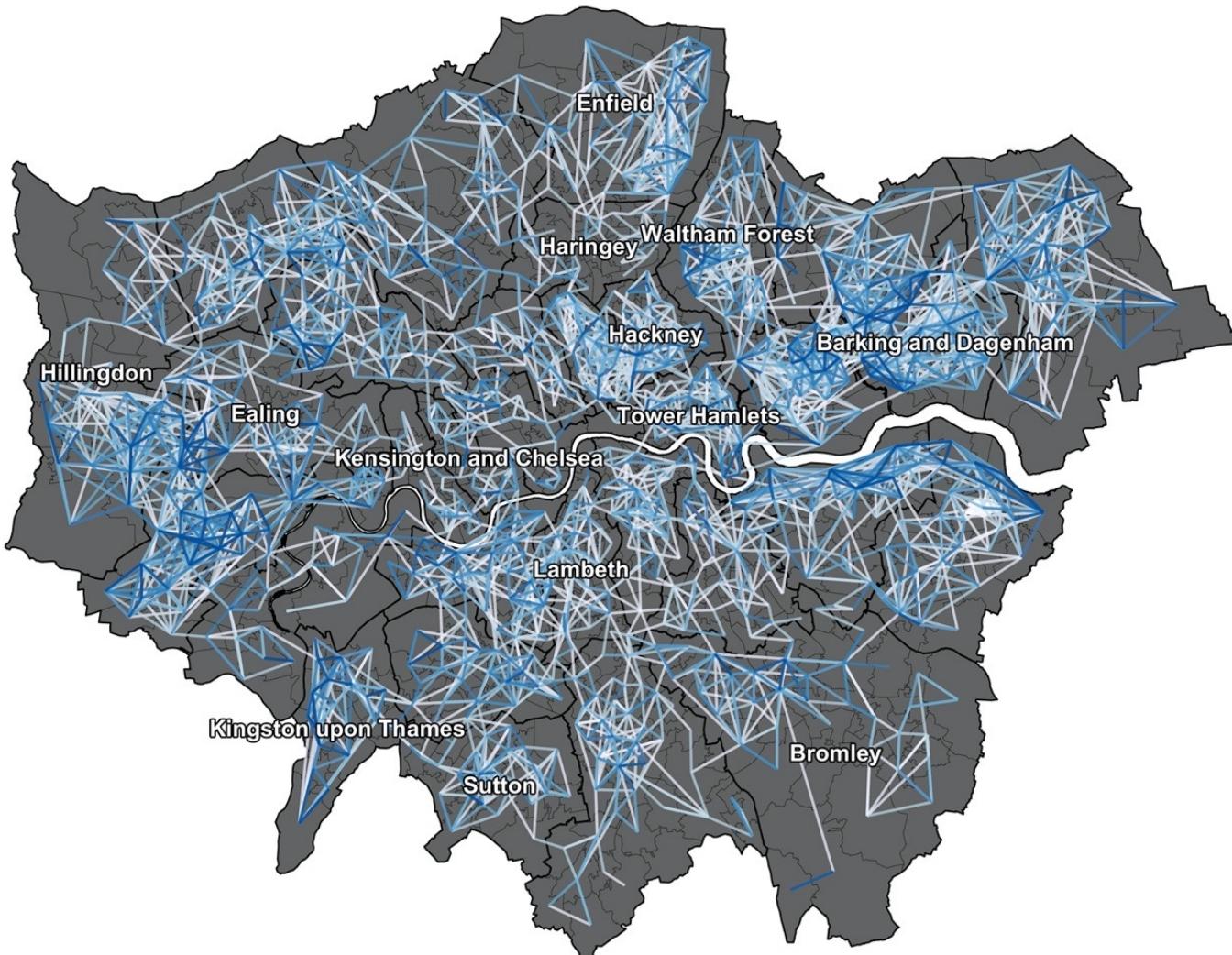


census



model

# Using consumer data for demographic analysis



# The Universal Visitation Law of Human Mobility

- Schläpfer *et al.* 2021
- “The universal visitation law of human mobility”
- Cell Detail Records. Billions of data point consisting of an anonymised ID of the corresponding user, latitude, longitude and a time stamp.
- Proposal of a scaling law that captures the temporal and spatial spectrum of population movement on the basis of large-scale mobility data from diverse cities around the globe: “the number of visitors to any location decreases as the inverse square of the product of their visiting frequency and travel distance”.

# The Universal Visitation Law of Human Mobility



# The Universal Visitation Law of Human Mobility

- Proposal of a “scaling law” that captures the temporal and spatial spectrum of population movement on the basis of large-scale mobility data from diverse cities around the globe.
- “The number of visitors to any location decreases as the inverse square of the product of their visiting frequency and travel distance”.
- Potential applications could be found in the predictions of recurrent flows, providing a basis for applications in urban planning, traffic engineering and the mitigation of epidemic diseases.

# Conclusion

- We explored four examples of different human-generated datasets and research projects which make use of different types of human-generated data and with different applications. **Good!**
- GPS data > Understanding individual travel behaviour
- Mobile phone application data > Understanding population movements
- Consumer data > Granular demographic data
- Mobile phone data (CDR) > “Law” of Geography?
- Human-generated datasets allow us to do things that are otherwise impossible.
- Clear **implications** for geographical analysis.

# Seminar preparation

In preparation for the next seminar, please first read the book chapter by Kitchen. After this, carefully read the remaining **four articles** on the reading list and for each article:

- Write a 100-words summary of what you think is the article's main contribution.
- Identify three strong points of the data used in the article.
- Identify three points of concern of the data used in the article.

# Questions

Justin van Dijk

j.t.vandijk@ucl.ac.uk

