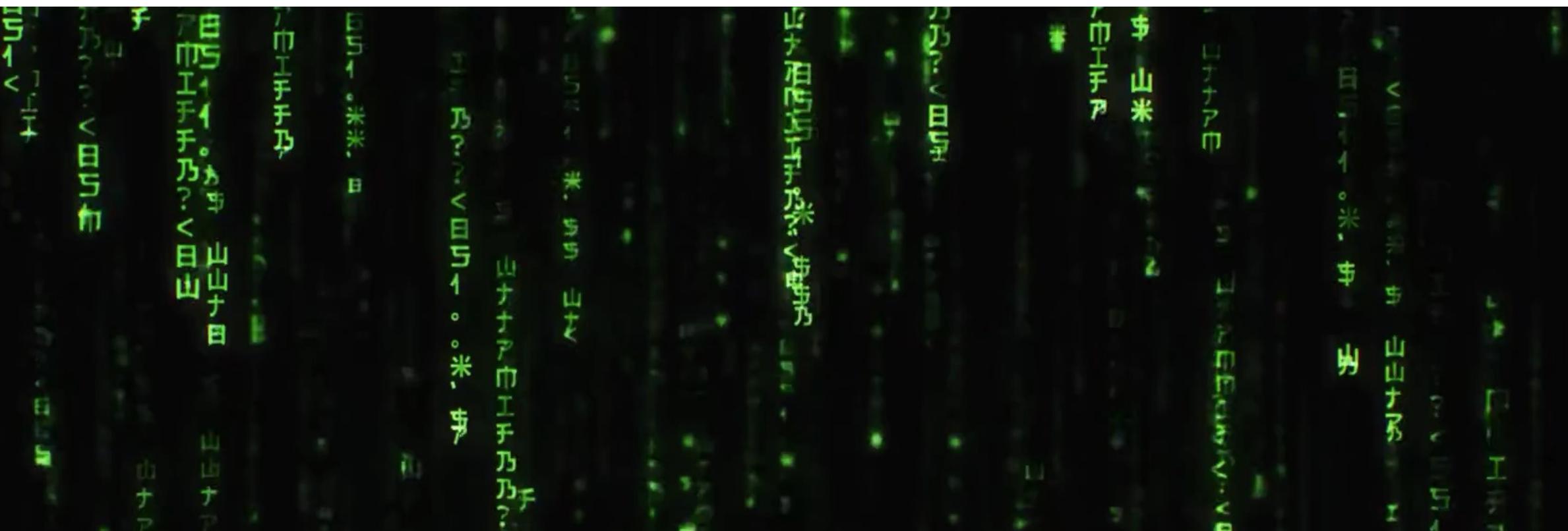


Data, Politics and Society

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 data 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 has its own reading list; available on the dedicated [Data, Politics, and Society](#) webpage (link also on Moodle).
- During the interactive seminars you are expected to come prepared and actively participate in discussions. The seminars are a core element of the module.

Content



Content

The screenshot shows a web browser window with the title 'Data, Politics and Society'. The URL is jtvdijk.github.io/GEOG0163/. The page content is as follows:

Jump to:

[Week 01](#) [Week 02](#) [Week 03](#) [Week 04](#) [Week 05](#) [Week 06](#) [Week 07](#) [Week 08](#) [Week 09](#)
[Week 10](#)

Week 01 - Data I: The Good

Reading list

This week's reading consists of:

1. Kitchen, R. 2014. Conceptualising data. In: Kitchen, R. *The data revolution: Big Data, open data, data infrastructures and their consequences*, pp.1-26. [\[Link\]](#)
2. Lansley, G., Li, W. and Longley, P. 2019. Creating a linked consumer register for granular demographic analysis. *Journal of the Royal Statistical Society Series A: Statistics in Society* 182:4, pp.1587-1605. [\[Link\]](#)
3. Schläpfer, M., Dong, L., O'Keeffe, K. et al. 2021. The universal visitation law of human mobility. *Nature* 593, pp.522-527. [\[Link\]](#)
4. Van Dijk, J. and Krygsman, S. 2018. Analyzing travel behavior by using GPS-based activity spaces and opportunity indicators. *Journal of Urban Technology* 25:2, pp.105-124. [\[Link\]](#)
5. Trasberg, T. and Cheshire, J. 2021. Spatial and social disparities in the decline of activities during the COVID-19 lockdown in Greater London. *Urban Studies*. [\[Link\]](#)

Seminar preparation

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

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

You do not have to submit this, however, do bring everything that you prepared with you to the seminar.

[Back to top](#)

Assessment

- The assessment for Data, Politics and Society consists of two parts: (1) a critique and essay outline and (2) a commentary article on the technical, ethical, political and/or social questions raised by data science and the analysis of large-scale human-generated datasets and their impacts on society.
- Document with detailed instructions available on Moodle.
- Submission Part 1 (30%): 16th of November 2023
- Submission Part 2 (70%): 8th of January 2024

Communication

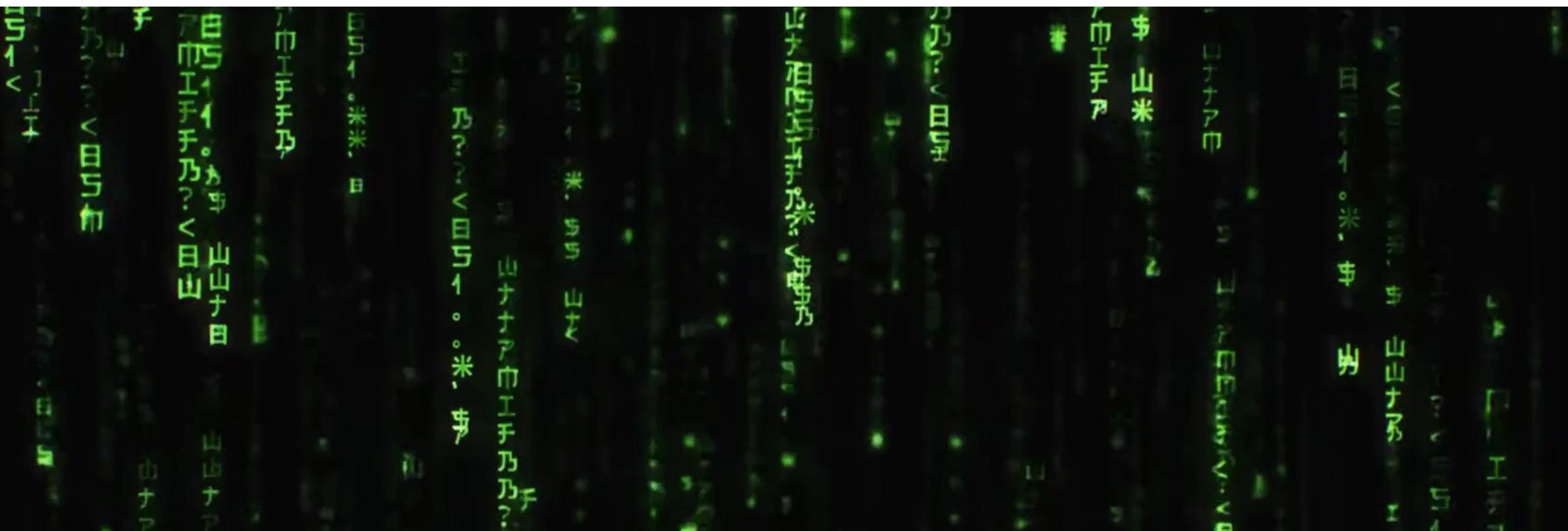
- All important information will be communicated through Moodle.
- Questions can be asked after each Lecture and during the Seminars.
- For specific questions or other support, my Academic Support and Feedback (ASF) Office Hours are scheduled on Monday mornings (11h00-13h00) and Thursday afternoons (13h00-14h00).

Illness

- In case you do feel unwell, even if it is 'just' a cold or sore throat, we strongly encourage you to stay at home if you can, to avoid passing on infection to others.-
- You are required to inform the Geography Office about your absence:
geog.office@ucl.ac.uk

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 (not per se qualitative)

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)

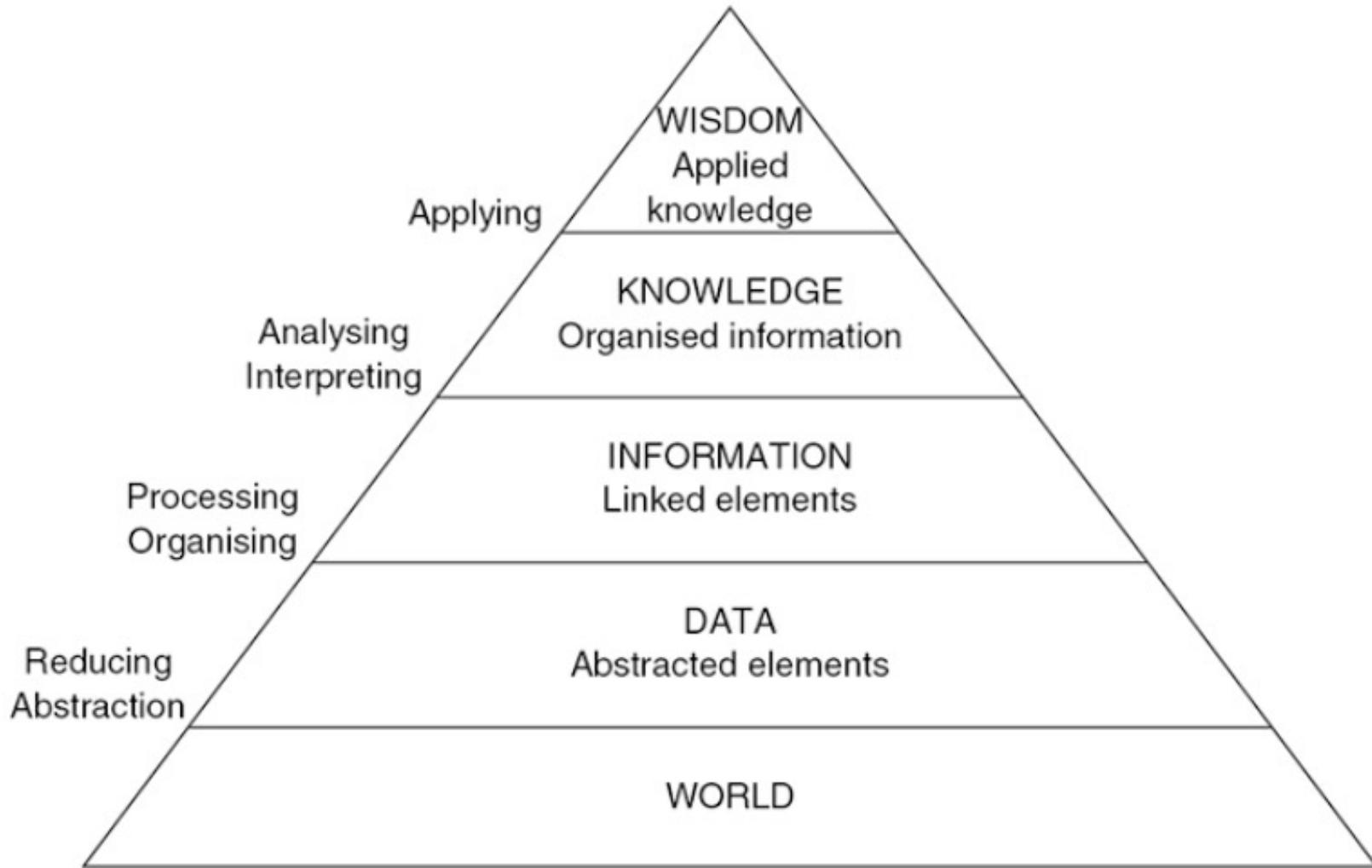
Data

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

Data

- Indexical data: data that enable identification (e.g. student 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

- Datafication: Our social and daily lives are also quantified in ways not seen before – our friendships, interests, conversations, information searches.
- Human-generated large-scale datasets come into existence from our direct use of technology which act as "sensors" to our actions and activities.
- How: actively collected data as well as passively collected data
- Why interesting to academics and businesses: 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. Revealed behaviour?

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 three examples of different human-generated datasets and research projects which make use of different types of data:

- 1:** App-based GPS data collected through smartphones (UCL)
- 2:** Linked Consumer Registers and Residential Mobility (UCL)
- 3:** 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 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

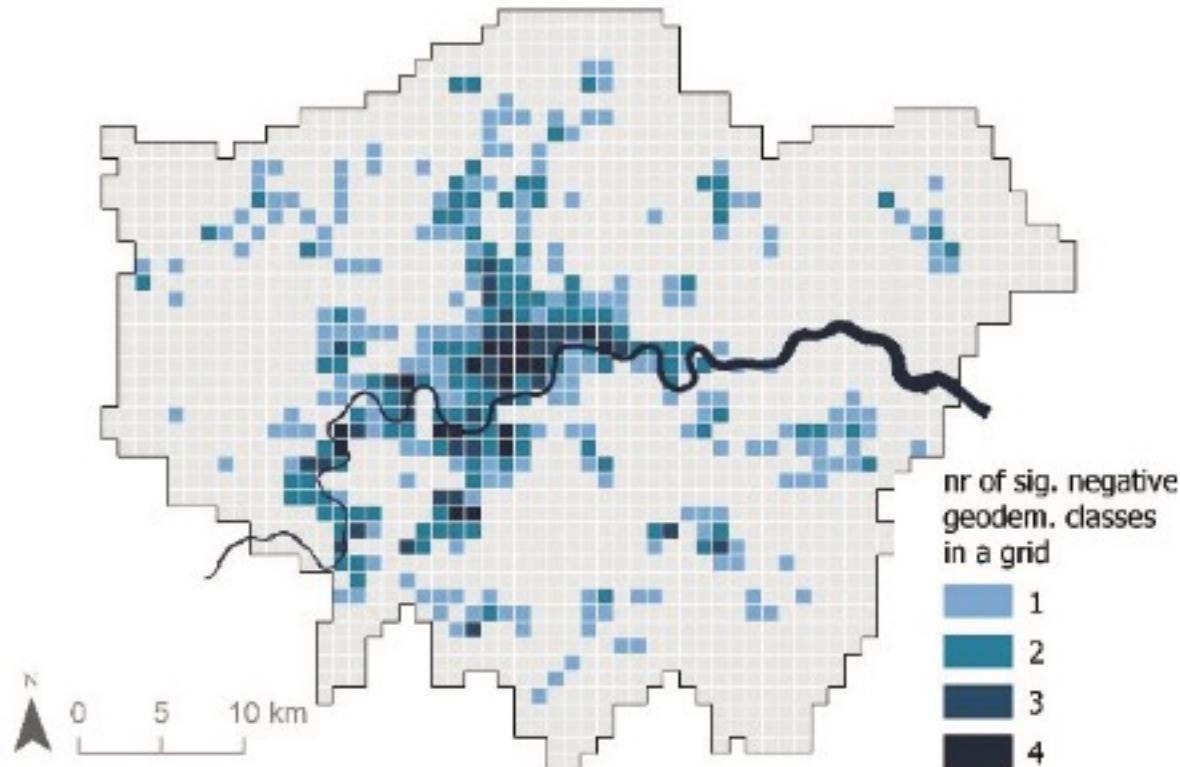
- Location data are collected and stored through a Software Development Kit (SDK) embedded into smartphone apps.
- The location data are further used by app developers for commercial purposes, such as location-based ad targeting, and are monetised by being sold to firms that mine the data for business 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

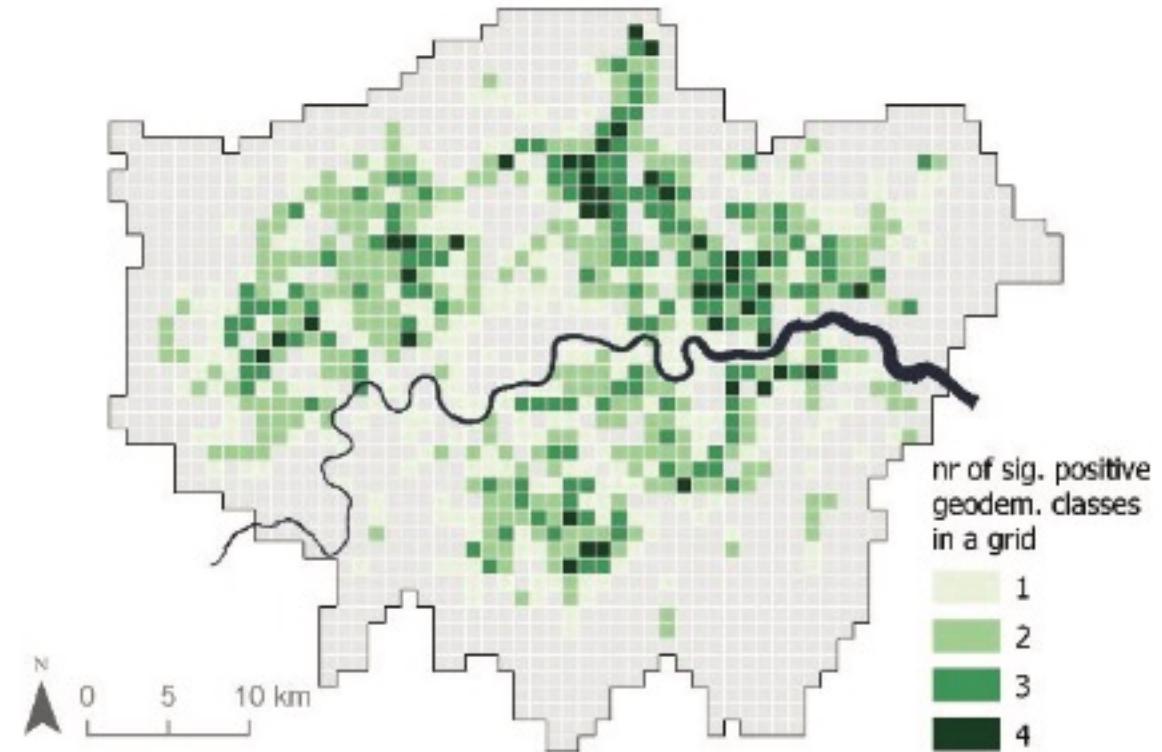
- The location data in this study are spatially aggregated to 1 km² grid cells so that we only know the number of unique devices per hour in each grid cell but do not have any information to construct digital traces of any of the devices.
- The location data has been collected from 308,311 unique devices 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

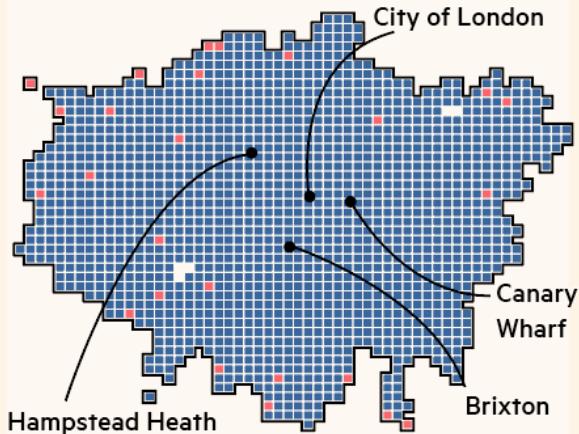
Change in the number of unique devices used in a km² area compared with a previous snapshot

Activity rose ■ Activity fell □

Peak lockdown:

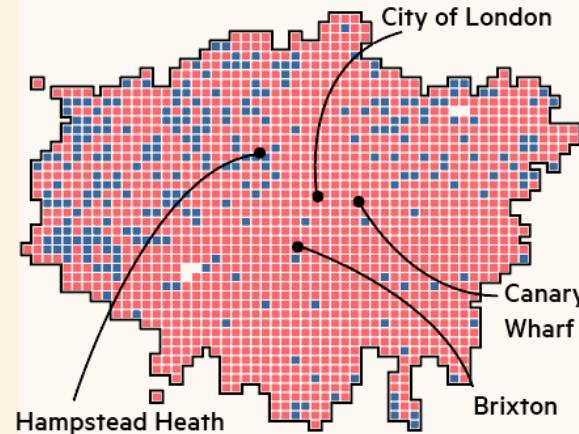
April 13-20 activity levels fell compared with pre-lockdown (March 16-22)

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 Dr Terje Trasberg.

Using consumer data for demographic analysis

- Lansley *et al.* 2019, Van Dijk *et al.* 2021
- Lack of annually updated administrative data sources. Census? Surveys?
- UK-wide population dataset 1997-2023 (LCR v3)
- UK Public Version of the Electoral Roll 1997 – 2023, supplemented with Consumer Data Sources (e.g. loyalty card data) for selected years

Using consumer data for demographic analysis

Forename	Surname	Address
Justin	van Dijk	Flat 18 Terry House SW22NT London

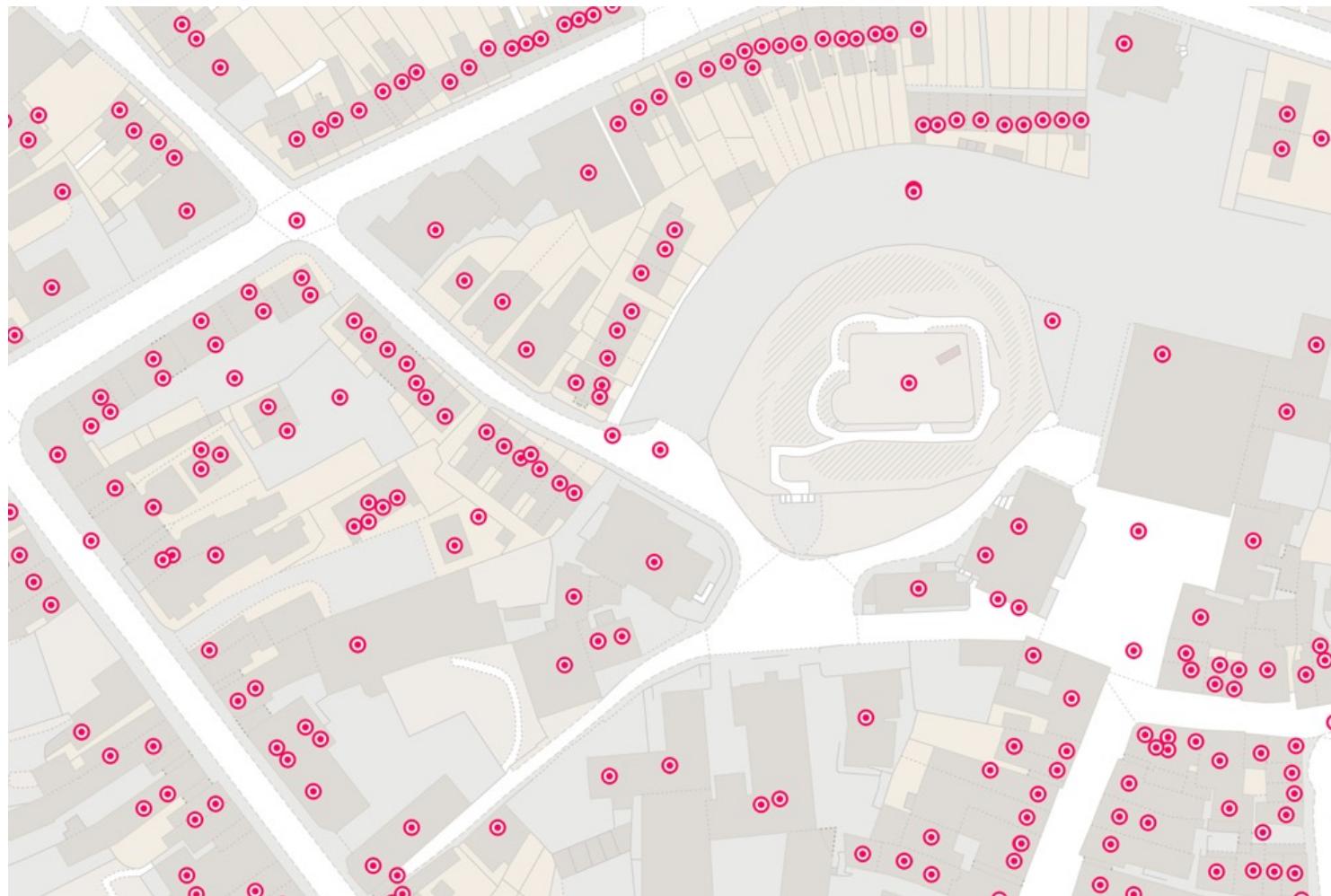
Using consumer data for demographic analysis

Forename	Surname	Address
Justin	van Dijk	Flat 18 Terry House SW22NT London

x

1,000,000,000

Using consumer data for demographic analysis



Using consumer data for demographic analysis

Address	Name	...	2019	2020	2021	2022	2023
a1000	Stephen Law	X	X	X			
a1000	S Law				X	X	X

Using consumer data for demographic analysis

Address	Name	...	2019	2020	2021	2022	2023
a1000	Stephen Law	x	x	x	x	x	x
a2000	S Swainr				x	x	x

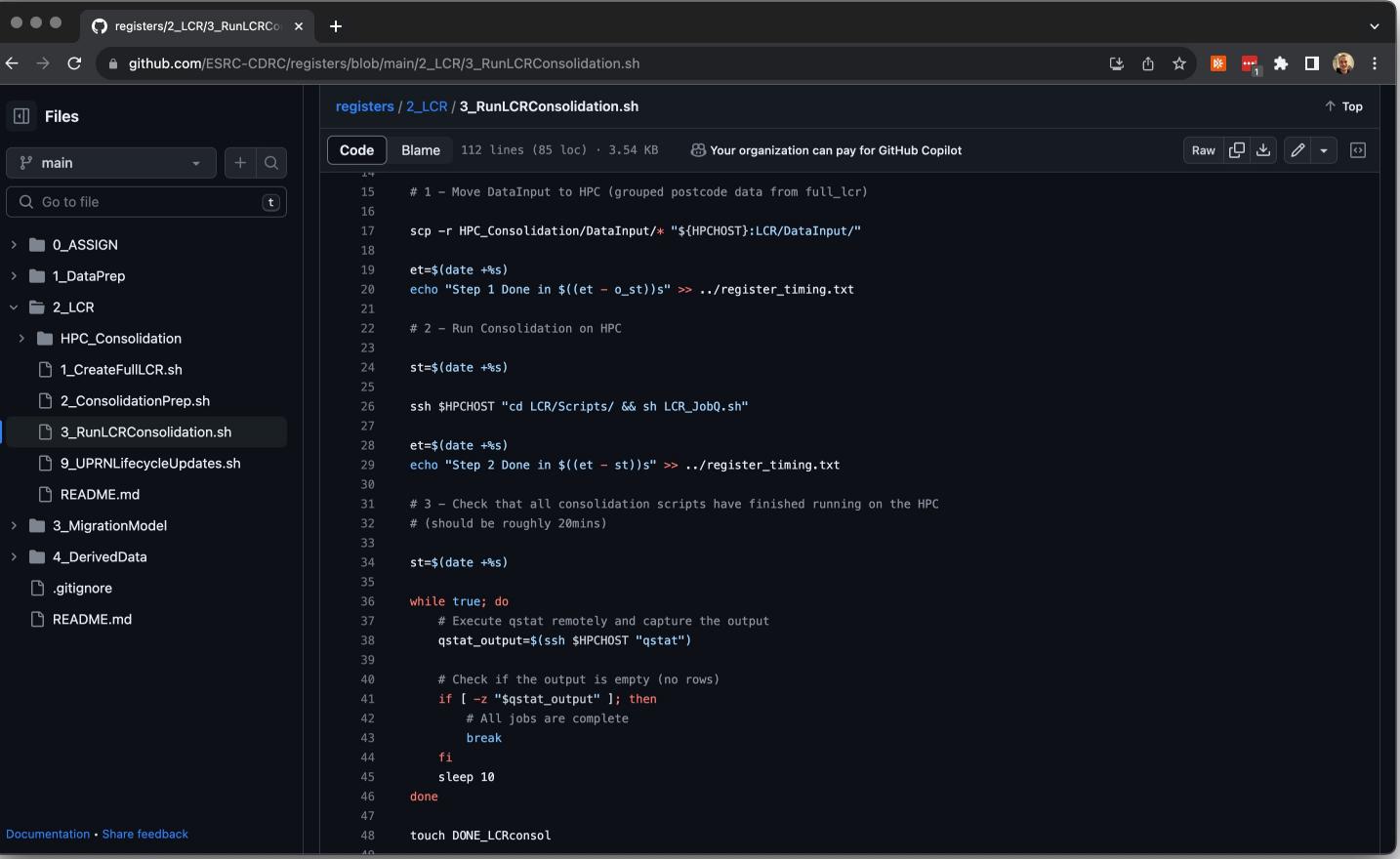
Using consumer data for demographic analysis

Address	Name	...	2019	2020	2021	2022	2023
a2000	Anwar Musah	x		x	x	x	
a2000	Rory Coulter		x		x	x	x

Using consumer data for demographic analysis

Address	Name	...	2019	2020	2021	2022	2023
a2000	Anwar Musah	x	x	x	x	x	
a2000	Rory Coulter		x	x	x	x	x

A lot of code



The screenshot shows a GitHub browser interface with the URL github.com/ESRC-CDRC/registers/blob/main/2_LCR/3_RunLCRConsolidation.sh. The left sidebar displays a file tree for the 'main' branch, showing directories like '0_ASSIGN', '1_DataPrep', '2_LCR' (which contains 'HPC_Consolidation', '3_RunLCRConsolidation.sh', '9_UPRNLifecycleUpdates.sh', 'README.md', and '3_MigrationModel'), '4_DerivedData', and files like '.gitignore' and 'README.md'. The right panel shows the content of the '3_RunLCRConsolidation.sh' script. The script is a shell script with the following content:

```
# 1 - Move DataInput to HPC (grouped postcode data from full_lcr)
scp -r HPC_Consolidation/DataInput/* "${HPCHOST}:LCR/DataInput/"

et=$(date +%s)
echo "Step 1 Done in $((et - o_st))s" >> ../register_timing.txt

# 2 - Run Consolidation on HPC

st=$(date +%s)

ssh $HPCHOST "cd LCR/Scripts/ && sh LCR_JobQ.sh"

et=$(date +%s)
echo "Step 2 Done in $((et - st))s" >> ../register_timing.txt

# 3 - Check that all consolidation scripts have finished running on the HPC
# (should be roughly 20mins)

st=$(date +%s)

while true; do
    # Execute qstat remotely and capture the output
    qstat_output=$(ssh $HPCHOST "qstat")

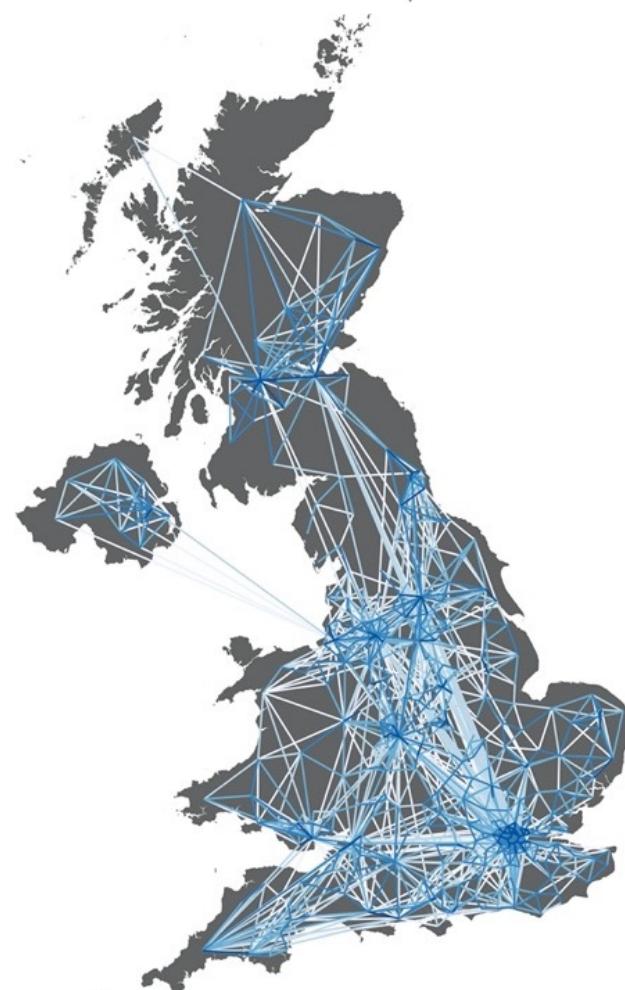
    # Check if the output is empty (no rows)
    if [ -z "$qstat_output" ]; then
        # All jobs are complete
        break
    fi
    sleep 10
done

touch DONE_LCRconsol
```

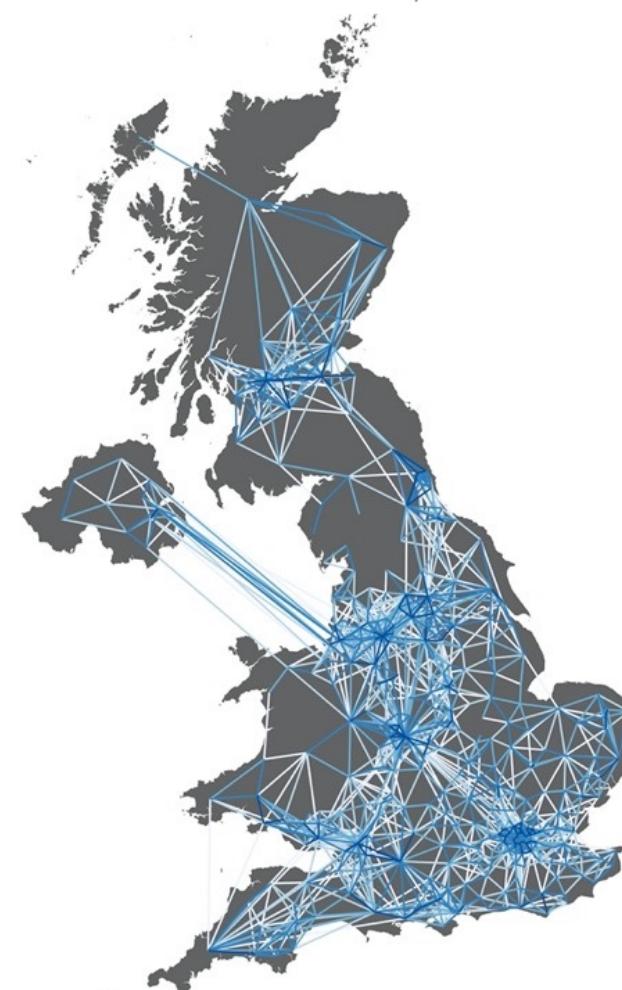
Using consumer data for demographic analysis

- 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 and model demographic change in inter-Census years.

Using consumer data for demographic analysis

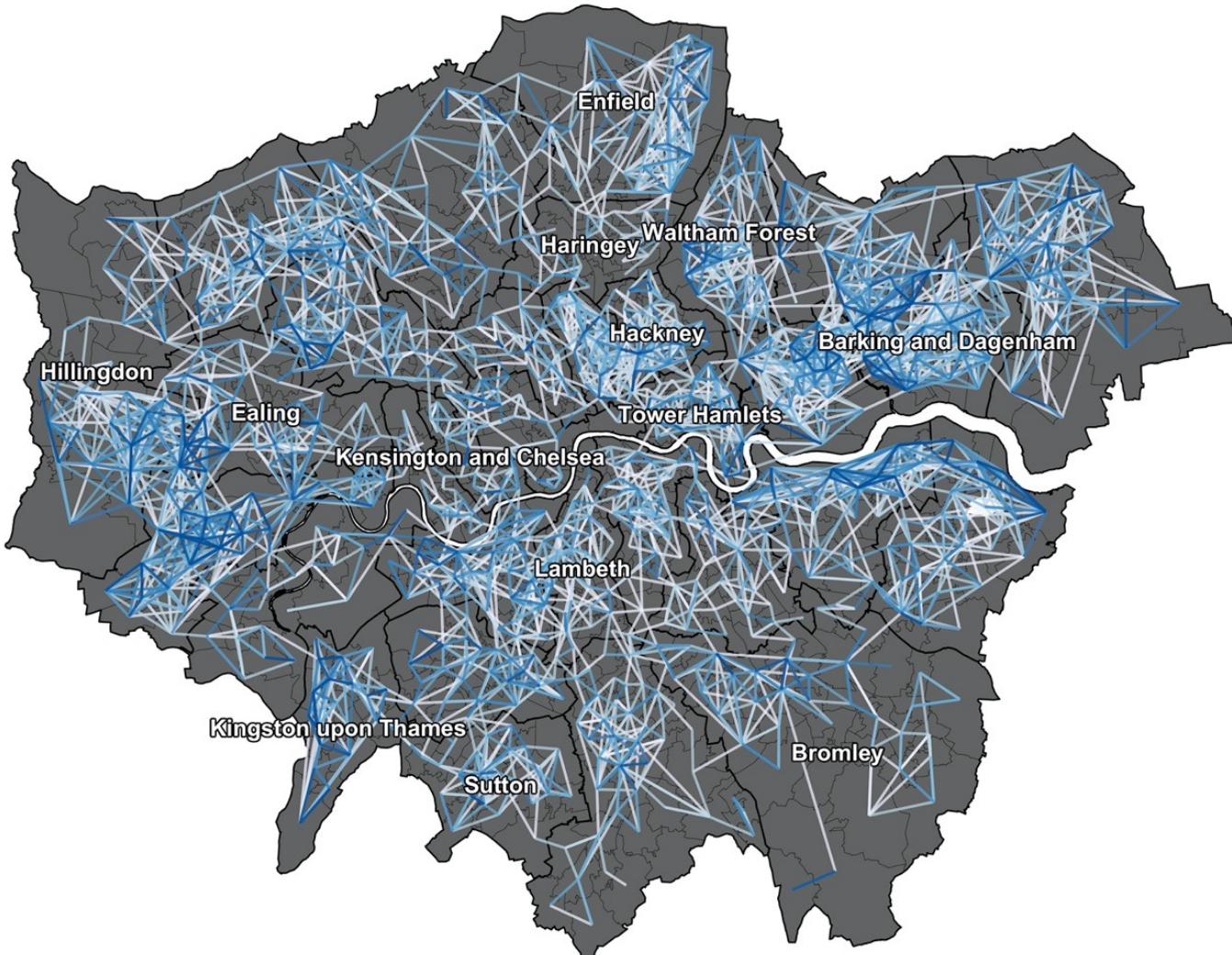


census



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 three examples of different human-generated datasets and research projects which make use of different types of human-generated data and with different applications. **Good!**
- Mobile phone application data > Understanding population movements
- Consumer data > Micro-demographic data infrastructure
- Mobile phone data (CDR) > Discovering new “Laws” 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 **three 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 areas of concern of the data used in the article.

Questions

Justin van Dijk

j.t.vandijk@ucl.ac.uk

