Social Data Science and the Digital Economy
CARMA 2022
─Keynote─

Dr Fabian Braesemann (1,2)

(1) Oxford Internet Institute, University of Oxford
(2) Datenwissenschaftliche Gesellschaft Berlin

Agenda

(A) Relevant concepts of the digital economy

- Two-sided markets
- The long tail
- The winner takes all
- Datafication

(B) Social data science applications

- The social data science pipeline
- Social data science for nowcasting: predicting sustainable tourism in Europe
- Social data science to understand the digital economy: online labour markets

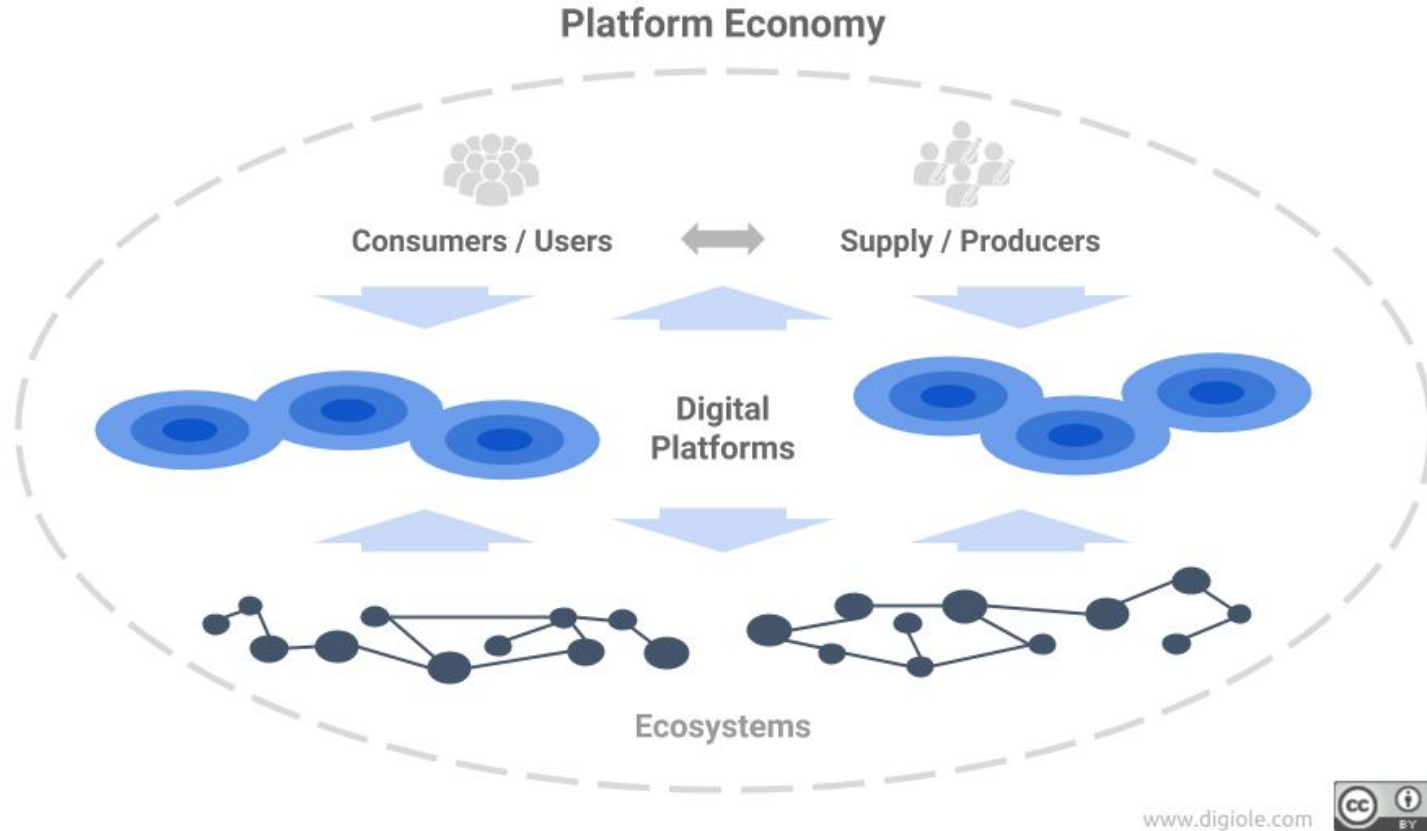# About myself

Study of economics
In Berlin and Vienna

Social Data Science at Oxford
since 2017

Understanding the *"digital economy embroidery"*.

What is the digital platform economy?

Platform Economy

Consumers / Users ⟷ Supply / Producers

Digital Platforms

Ecosystems

www.digiole.com

Braesemann | CARMA 2022 Keynote | Social Data Science and the Digital Economy

How did we get here? Two "trailblazers" of the digital economy

1994

amazon.com

Only selling books

2000

amazon.com
EARTH'S BIGGEST BOOKSTORE

Diversification after the survival of the ".com" bubble

2010

amazon Prime

50% of all US e-commerce retail

2022

More than three million jobs worth over $1 billion

1998 Elance work differently   **+**   **2003** oDesk   **=**   **2015** upwork

The two-sided market: Why are digital platforms successful

A physical bookshelf has limited space

Amazon's bookshelves are (in theory) endless

Advantages and disadvantages in the platform economy

A market with more than "just" a price tag

A market with space for variety
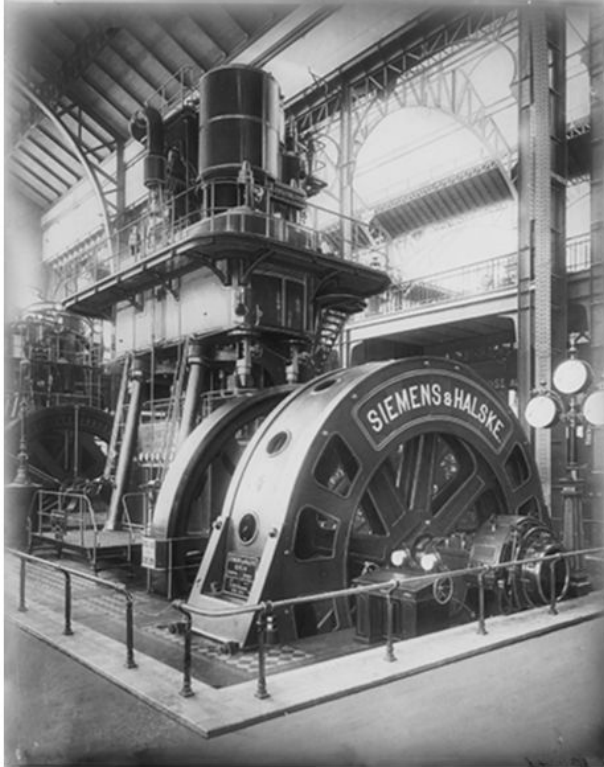
But often a winner-takes-all market

Why should we care about data?
Datafication is everywhere

Siemens 1914:
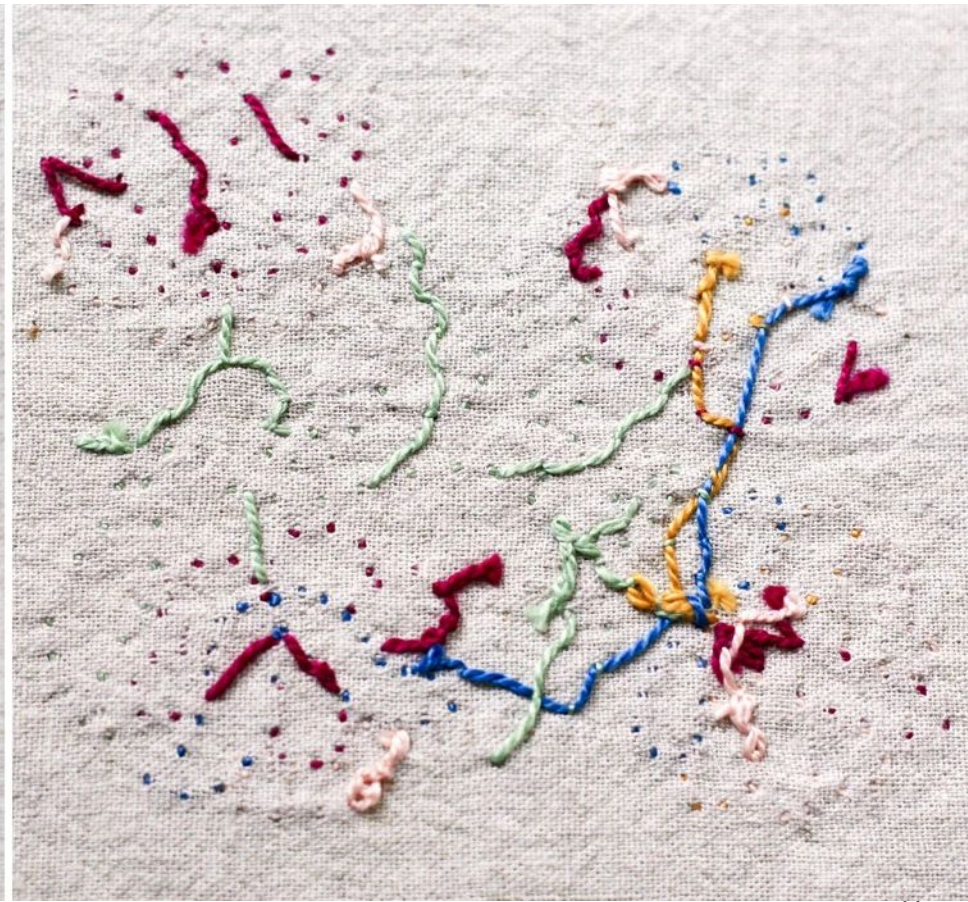

Siemens 2014:

We live in a universe of data.

We need to understand that data is everywhere.

We need to understand data as an asset and turn it into a value.

Braesemann | CARMA 2022 Keynote | Social Data Science and the Digital Economy

Summary - the digital economy and datafication

- Two-sided markets give rise to digital platforms

- Customer preferences, fix costs and variety

- The winner takes all

- Datafication

Looking at the backside of the *"digital economy embroidery"*

Braesemann | CARMA 2022 Keynote | Social Data Science and the Digital Economye: Kate and Rose Patterns)

Social Data Science and the Digital Economy

Novel social research

Effective policy action
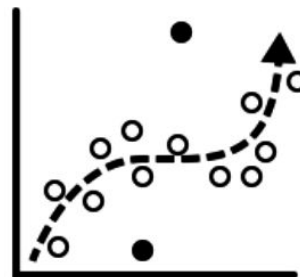
How to flip the *"digital economy embroidery"*

1. Web mining
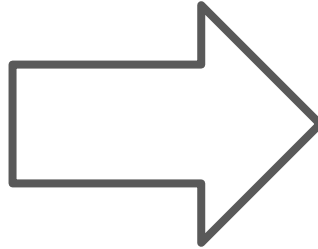2. Data pre-processing
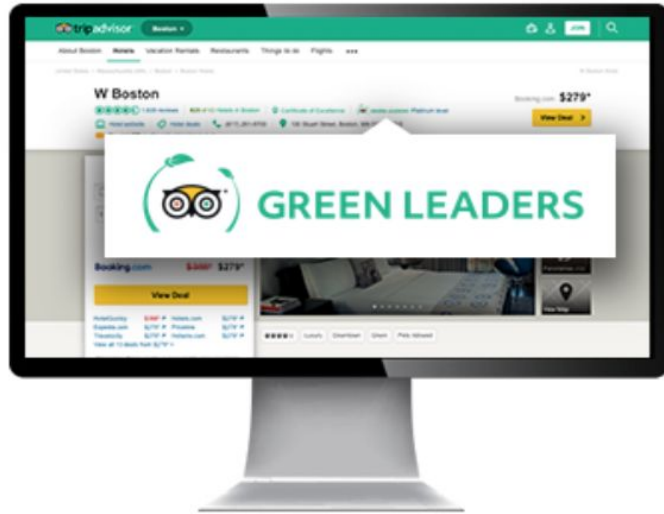3. Unsupervised learning
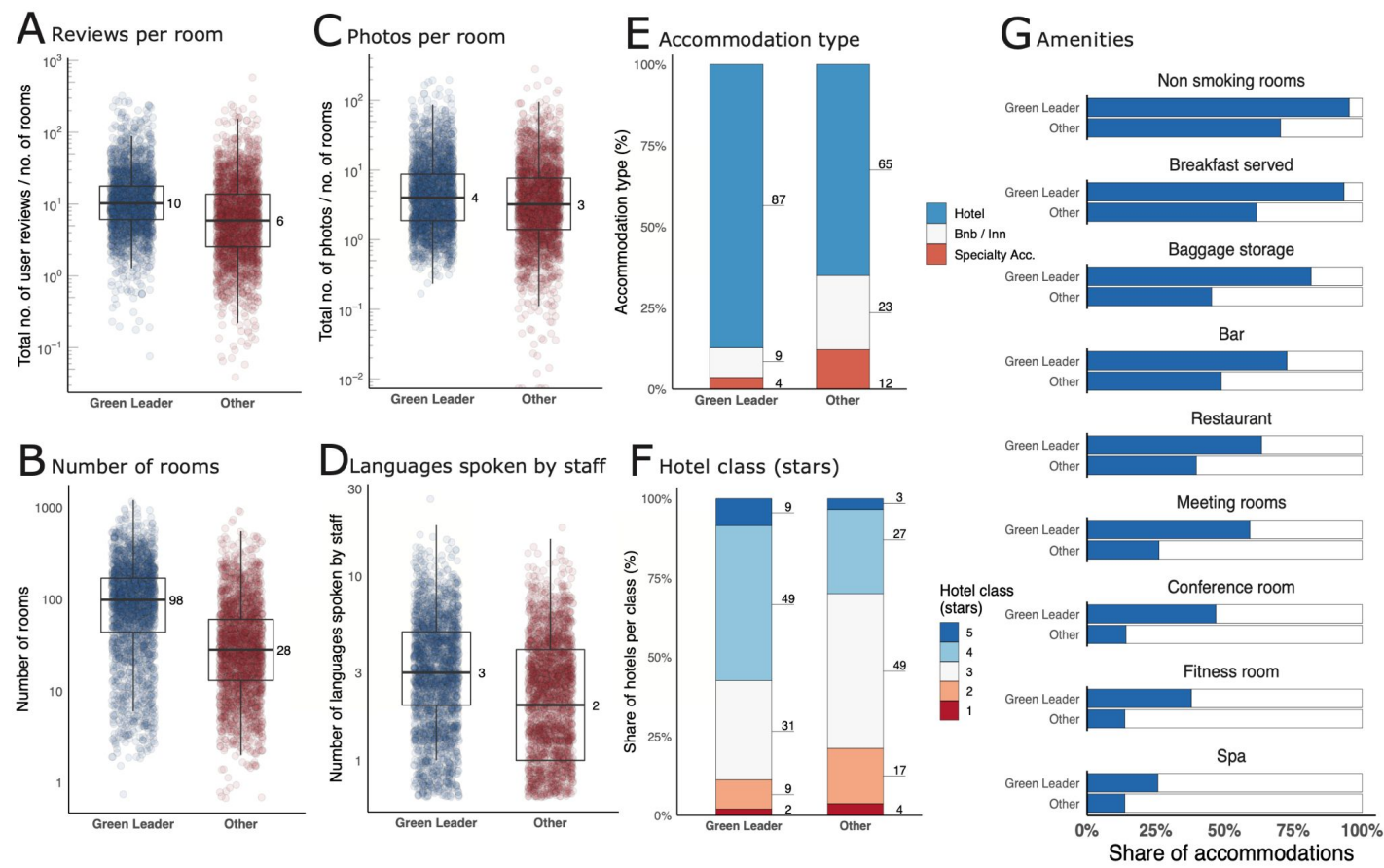4. Predictive Analytics
5. Data visualisation
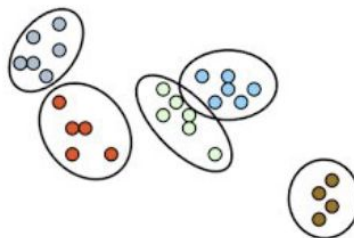
Flipping the *"digital economy embroidery"*
Using data science for nowcasting of Sustainable Tourism
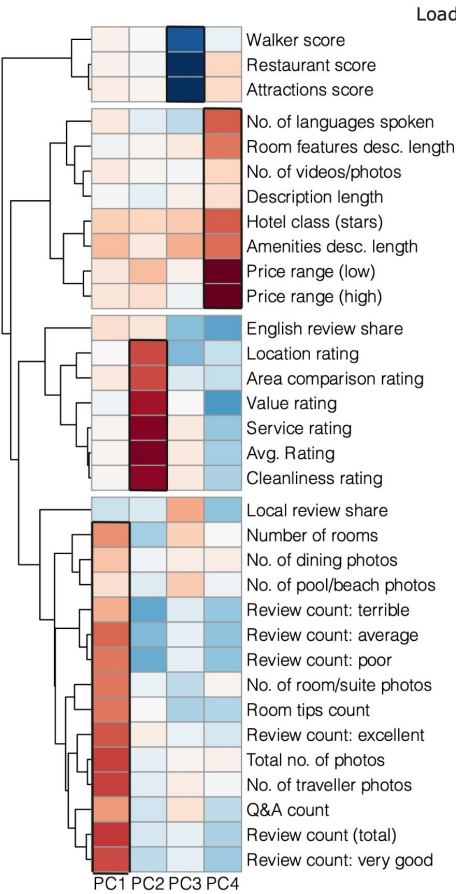
Flipping the *"digital economy embroidery"*: Sustainable Tourism

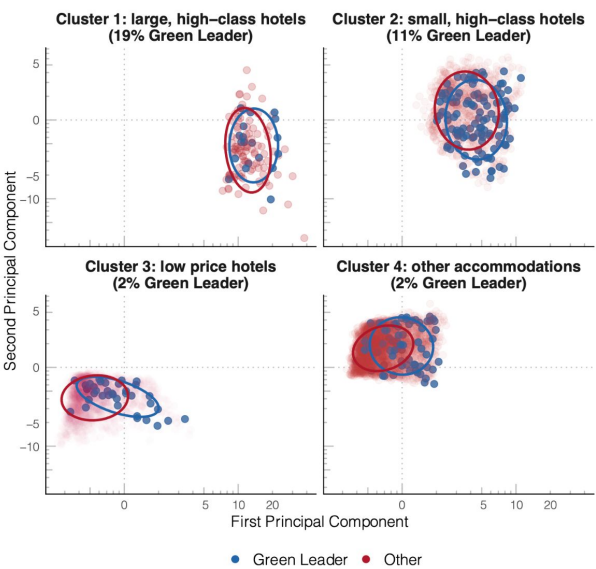# Flipping the *"digital economy embroidery"*: Sustainable Tourism

A Heatmap of principal component loadings

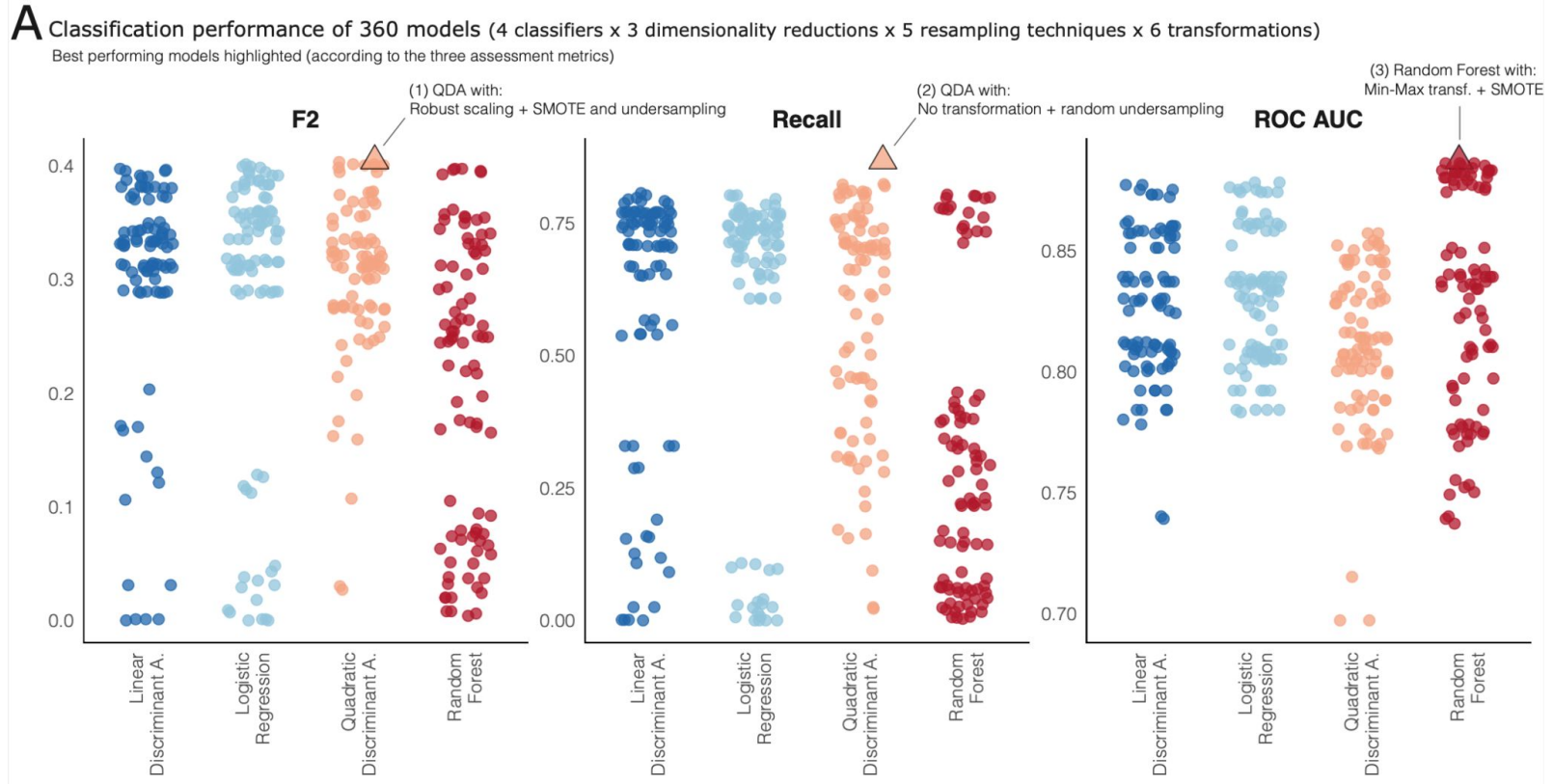**B** Summary statistics of quality and user interaction features per cluster

Cluster:	(1)	(2)	(3)	(4)
No. of accommodations	1,214	10,067	20,303	33,931
Share	2 %	15 %	31 %	52 %
Stars (mean)	3.9	3.7	2.8	3.1
Rooms (median)	279	86	33	20
Photos/rooms (med.)	6.4	7.2	1.7	4.0
Reviews/rooms (med.)	12.3	13.0	3.5	6.9
Price p. n. (EUR, med.)	147	132	80	99
Share of english reviews	71 %	59 %	21 %	32 %
Share of Green Leaders	**19 %**	**11 %**	**2 %**	**2 %**

C Clusters (panels) and Green Leader hotels (colour) in 2-dimensional space

Cluster 1: large, high–class hotels (19% Green Leader)
Cluster 2: small, high–class hotels (11% Green Leader)
Cluster 3: low price hotels (2% Green Leader)
Cluster 4: other accommodations (2% Green Leader)

Flipping the *"digital economy embroidery"*: Sustainable Tourism

A Classification performance of 360 models (4 classifiers x 3 dimensionality reductions x 5 resampling techniques x 6 transformations)
Best performing models highlighted (according to the three assessment metrics)

(1) QDA with:
Robust scaling + SMOTE and undersampling

(2) QDA with:
No transformation + random undersampling

(3) Random Forest with:
Min-Max transf. + SMOTE
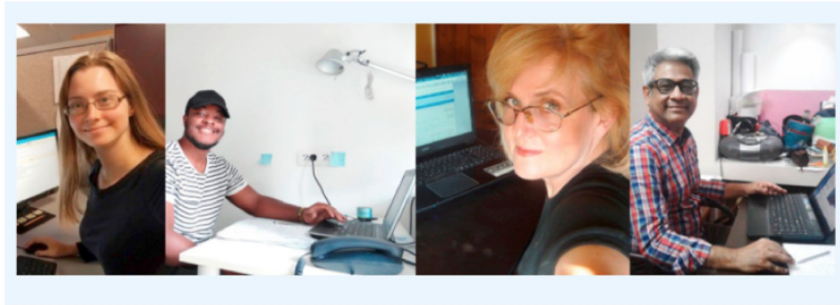
Flipping the *"digital economy embroidery"*: Sustainable Tourism

Predicted share of Green Leader accommodations in Europe (NUTS-2)

Share of Green Leader hotels (%)
Predicted from QDA model (1)

50 – 67
20 – 50
10 – 20
5 – 10
1 – 5
0 – 1
no data
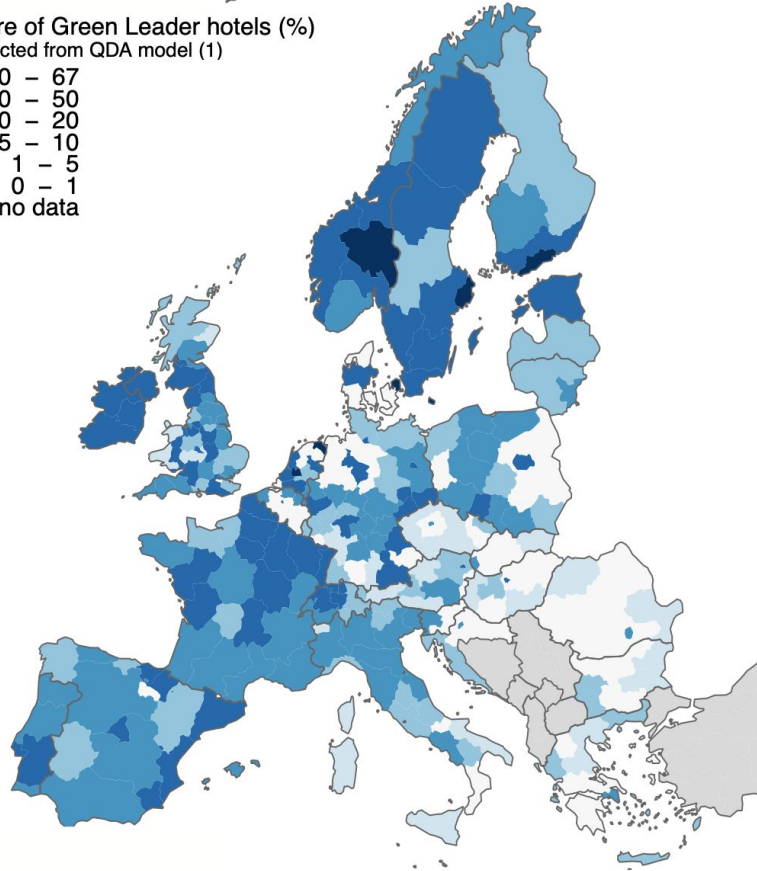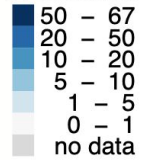
Understanding the Digital Economy through Data Science: The Global polarisation of remote work

The Global polarisation of remote work: our argument

Global polarisation:

Digital Taylorism and the Global Auction (Brown et al. 2011)

Urban-rural polarisation:

Paul Collier's *Future of Capitalism*
'Booming metropolis vs broken provincial cities'

Polarisation between occupations:

David Autor
1. *The skill content of recent technological change (2003)*
2. *The growth of low-skill service jobs and the polarization of the US labour market (2013)*
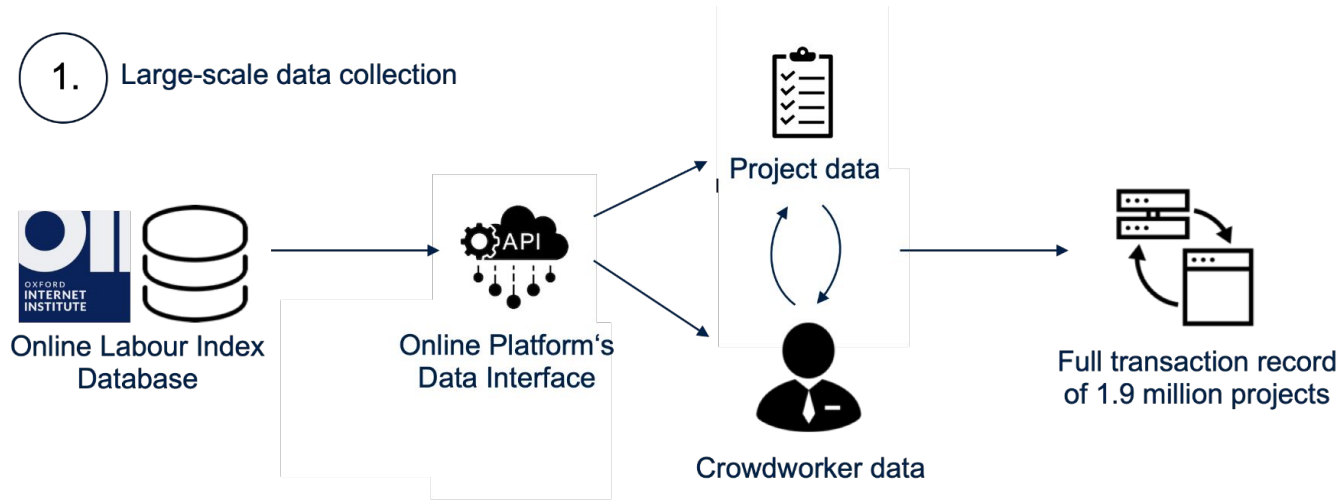

→ Polarisation has a direct impact in the global and largely unregulated online labour market
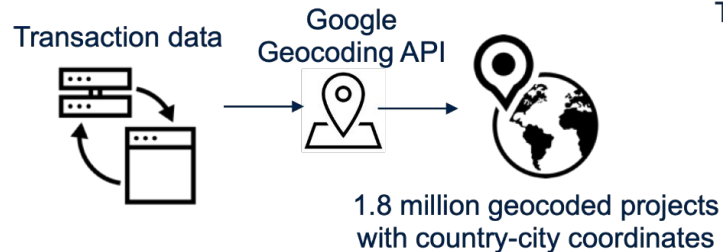
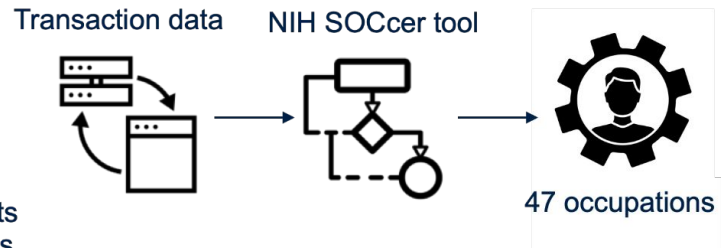# The Global polarisation of remote work: data science pipeline

1. Large-scale data collection

Online Labour Index Database → Online Platform's Data Interface → Project data / Crowdworker data → Full transaction record of 1.9 million projects

2. Geocoding

Transaction data → Google Geocoding API → 1.8 million geocoded projects with country-city coordinates

3. Algorithmic mapping of job types to official occupations

Transaction data → NIH SOCcer tool → 47 occupations

Online labour markets are polarised across space

City:
- ● Buyer
- ● Seller
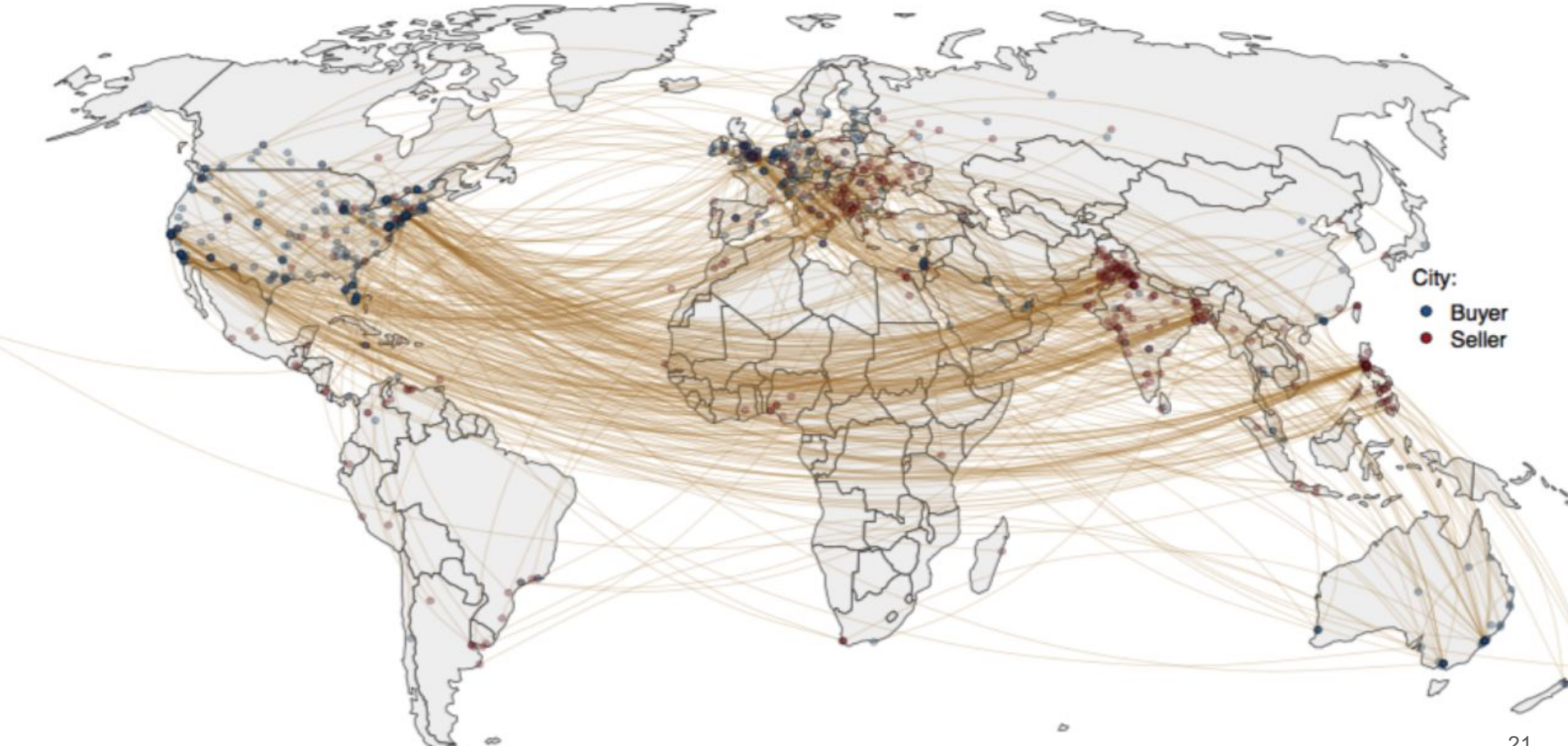
# Online labour markets are polarised across space
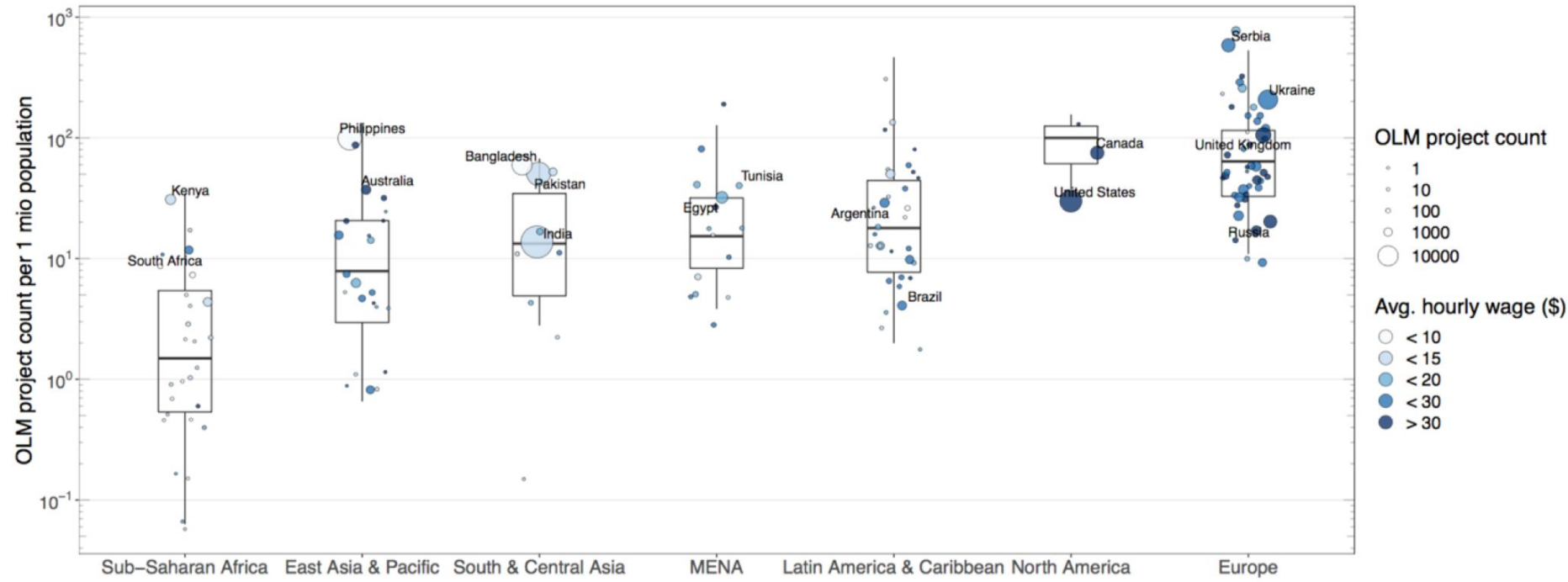
The polarisation can be explained by regional factors

Dependent variable:	Yearly online labour project count[a]			Online labour avg. wage per hour[a,b]		
Level:	Countries	Sub-national regions		Countries	Sub-national regions	
Geography:	Global	OECD+[c]	Global South[d]	Global	OECD+[c]	Global South[d]
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Population Population, total (log scale)	**0.96*** (0.02)	**0.85*** (0.09)	**0.54*** (0.04)	**0.04*** (0.01)	0.05 (0.04)	**0.06*** (0.02)
Education Model (1), (4): share of pop. with secondary education Model (2), (5): share of pop. with tertiary education Model (3), (6): avg. years of education	**0.06*** (0.003)	0.004 (0.01)	**0.08*** (0.03)	**0.01*** (0.001)	**0.01*** (0.002)	**0.06*** (0.01)
Income per capita Model (1), (4): GDP per capita (in 1,000 $) Model (2), (5): GDP per capita (2015 PPP $, log scale) Model (3), (6): Gross National Income p. c. (2011 PPP $)	**−0.01*** (0.003)	**−0.33*** (0.10)	**0.05*** (0.02)	**0.002*** (0.001)	**−0.12*** (0.05)	**0.02*** (0.01)
Internet connectivity Model (1), (4): fixed broadband subscriptions per 100 people Model (2), (5): share of HHs with internet broadband access Model (3), (6): share of HHs with internet access	**0.04*** (0.01)	**0.03*** (0.004)	**0.01*** (0.004)	**0.01*** (0.002)	**0.005*** (0.001)	**0.003*** (0.001)
IT specialisation of the economy Model (1), (4): ICT share of all service exports (log scale) Model (2), (5): Gross value added in ICT (2015 PPP $, log)	**0.22*** (0.04)	**0.56*** (0.05)		**−0.02*** (0.01)	**0.09*** (0.02)	
English language Indicator: English is official language	**0.68*** (0.11)			**−0.07**** (0.03)		
Price level PPP conversion factor (per 1,000 int. $)	**−0.32*** (0.08)			−0.04 (0.02)		
Capital region Indicator: region holds country capital		0.13 (0.10)	**1.74*** (0.11)		**−0.08**** (0.03)	**−0.12*** (0.05)
Constant					**3.21*** (0.31)	**2.53*** (0.11)
n (regional units)	139	292	299	112	253	52
Observations	1,136	2,384	2,074	763	1,536	244
Fixed / Random Effects	Yearly FE	Country-year Fixed Effects		Yearly FE	Country-year Random Effects	
R^2	0.72	0.70	0.42	0.44	0.79	0.62
Adjusted R^2	0.71	0.67	0.35	0.43	0.79	0.61

23

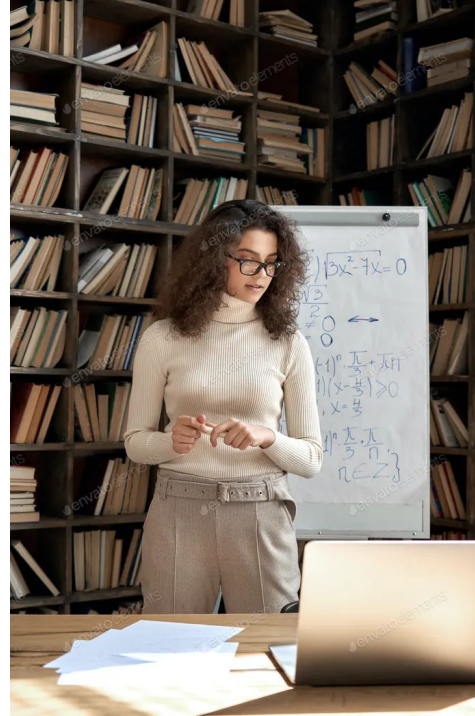# Skills are the main polarising force in the online labour market
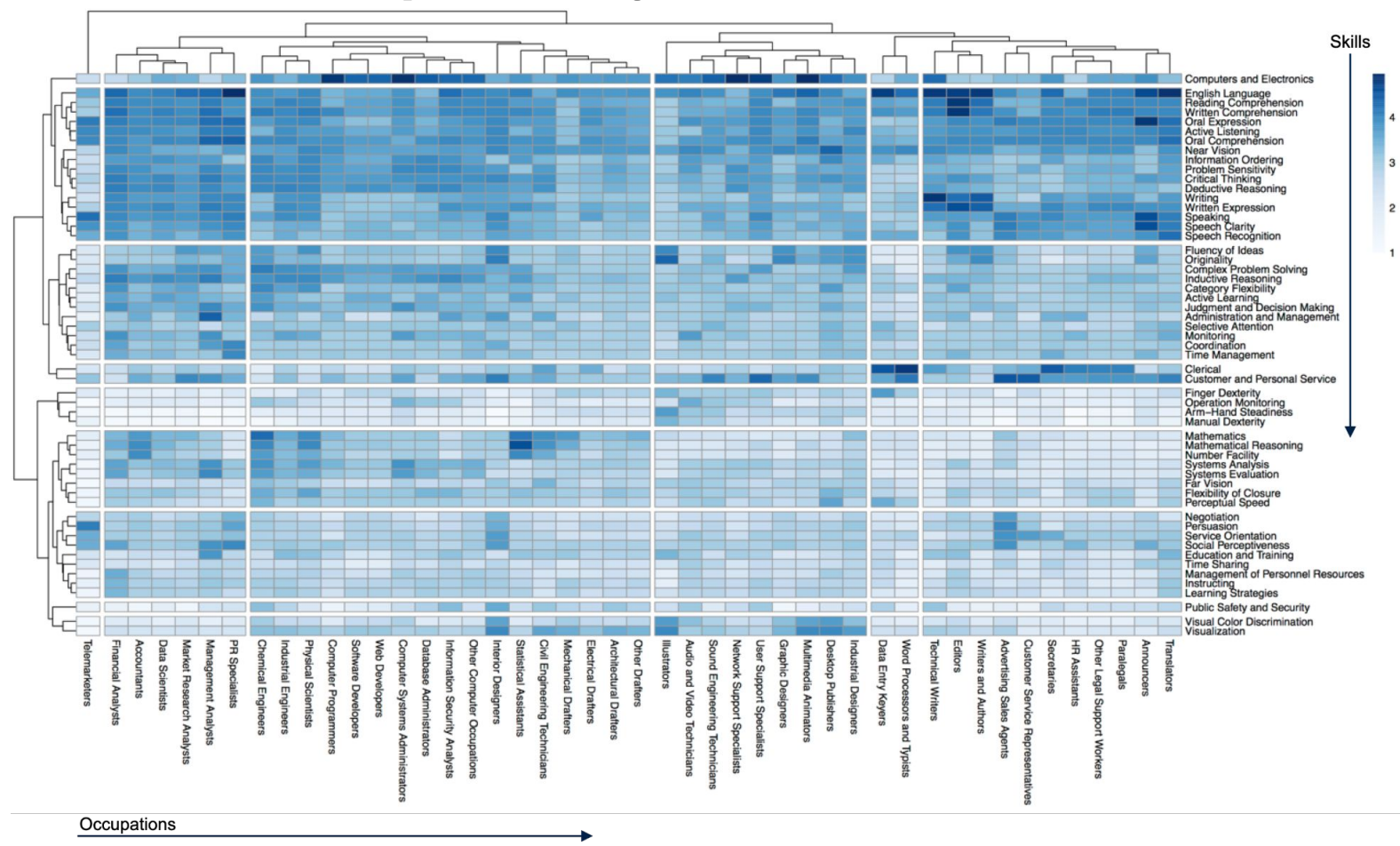
Steps for effective and sustainable digital economy participation

Braesemann et al. (2021). The polarisation of remote work.

Provide internet access   Teach marketable skills   Improve platform infrastructure

Summary - Social Data Science and the digital economy

1. Digital technologies enabled important digital markets to emerge
2. These digital markets have their own 'economic laws'
3. Data play a crucial role in these markets

4. Using social data science methods we can use this data for research & policy
5. Online data can inform novel measures and be used for 'nowcasting'
6. Social data science can be used to understand and manage the digital economy

References:

Hoffmann, Braesemann, Teubner (2022) Measuring Sustainable Tourism with Online Platform Data. EPJ Data Science. Forthcoming

Braesemann et al. (2022) The global geography of remote work. PLOS ONE. Revise & Resubmit (Preprint: The polarisation of remote work)