GEOG5927M

Predictive Analytics

Roger Beecham
www.roger-beecham.com

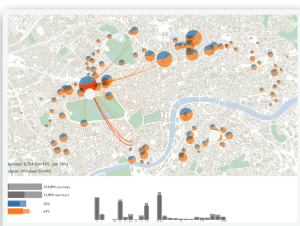
Roger Beecham

Lecturer in Geographic Data Science

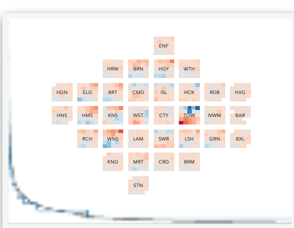
www.roger-beecham.com

Technical: Data Visualization, New and Computational Statistics
Applied: Transportation, Crime Science, Political Science

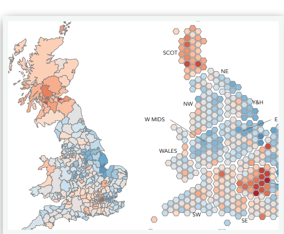
Beecham & Wood
Exploring gendered cycling behaviours
doi: 10.1080/03081060.2013.844903
Processing, R | Estimation-based stats



Beecham et al.
Studying (inferring) commuter workplaces
doi: https://doi.org/10.1016/j.compenvurbsys.2013.10.007

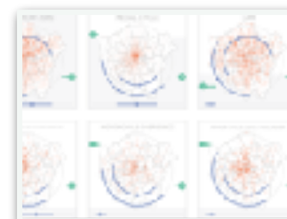


Beecham & Wood
Characterising (inferring) group-cycling
doi:10.1016/j.trc.2014.03.007
Java, Processing | Network stats

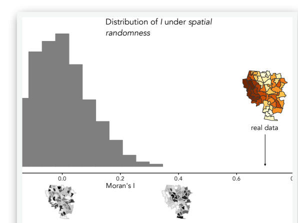


Beecham et al.
Locally varying explanations
doi:10.5311/JOSIS.2018.16.377
R | GW stats, penalised regression, permutation

Beecham et al.
An update to ecological analysis in Political Science
in press
R | penalised regression, permutation



Beecham et al.
Multivariate small multiples
doi10.1111/cgf.12900
Processing | Circular stats, regression



Beecham et al.
Map LineUps
doi:10.1109/TVCG.2016.2598862
R | Spatial stats, regression, permutation

Beecham et al.
Statistical Process Control for uncertainty analysis
in press
Javascript, d3 | Frequentist

Beecham & Slingsby
Characterising labour market self-containment
doi: 10.1177/0308518X19850580
R | Estimation-based stats

GEOG5927M

Predictive Analytics

Roger Beecham
www.roger-beecham.com

Module content and philosophy

Spatial modelling • Data mining • Response modelling • Microsimulation • Agent-based modelling	to	similuate and predict consumer behaviour	[content]
Research and industry case studies	to	evaluate modelling techniques in practice	[philosophy]

Outcomes

By the end of this module you should be able to

1. **explain** and **critically evaluate** the role of spatial analytics in simulating and predicting consumer behaviours

----

2. **apply** geocomputational modelling and simulation techniques on real data sets

----

3. **devise** and **employ** spatial modelling tools to address business problems, presenting and justifying recommendations in an appropriate context

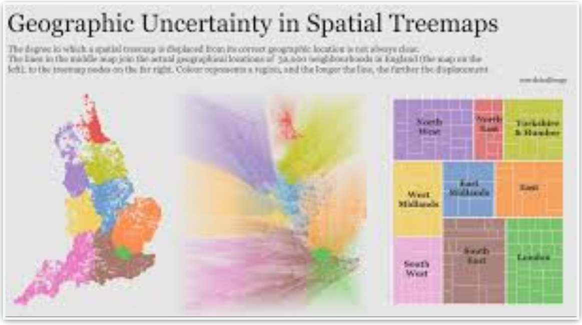
Who’s who

Roger Beecham
Assignment 1 | Convener

Jaiqi Ge | Nick Malleson
Assignment 2

Rob Radburn
Guest lecturer

Guest Lecture



Rob Radburn
Leicestershire County Council

Where & When

Lectures | Mondays

1400-1600

Roger Stevens LT 25

12.25

Labs | Wednesdays

0930-1230

Miall computing labs

10.19

1400-1700

Chemical and Process Engineering labs

GR.06

Outline : Lectures

Roger Beecham	Session 1 Predictive Analytics & Microsimulation
Roger Beecham	Session 2 Response Modelling & Targeted Marketing
Jiaqi Ge | Nick Malleson	Session 3 Behavioural and Agent-based Models
Rob Radburn | Roger Beecham	Session 4 Guest Lecture
Roger Beecham	Session 5 Wrap-up

Outline : Labs

Roger Beecham	Session 1 Simulating Behaviour
Roger Beecham	Session 2 Targeted Marketing
Jiaqi Ge | Nick Malleson	Session 3 Behavioural and Agent-based Models
Rob Radburn | Roger Beecham	Session 4 Coursework Surgery
Roger Beecham	Session 5 Assignment 2 — Presentations

Assessment

Assignment 1

Individual data analysis based on practicals 1 and 2
1,000 words, 4 figures
Thursday 16th January 2020 by 2pm

Assignment 2

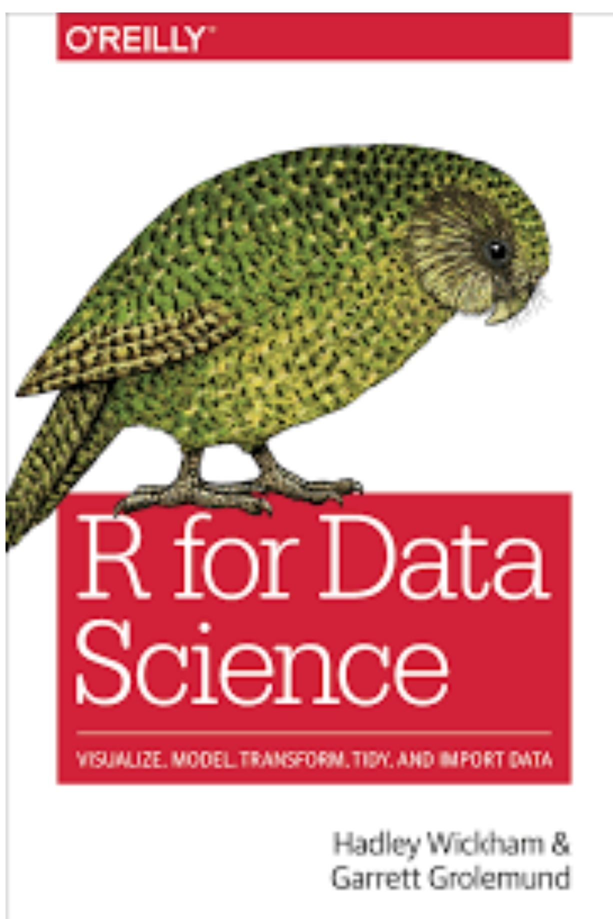
Group presentations based on practical 3
Presentations held Wednesday 11th December 2019

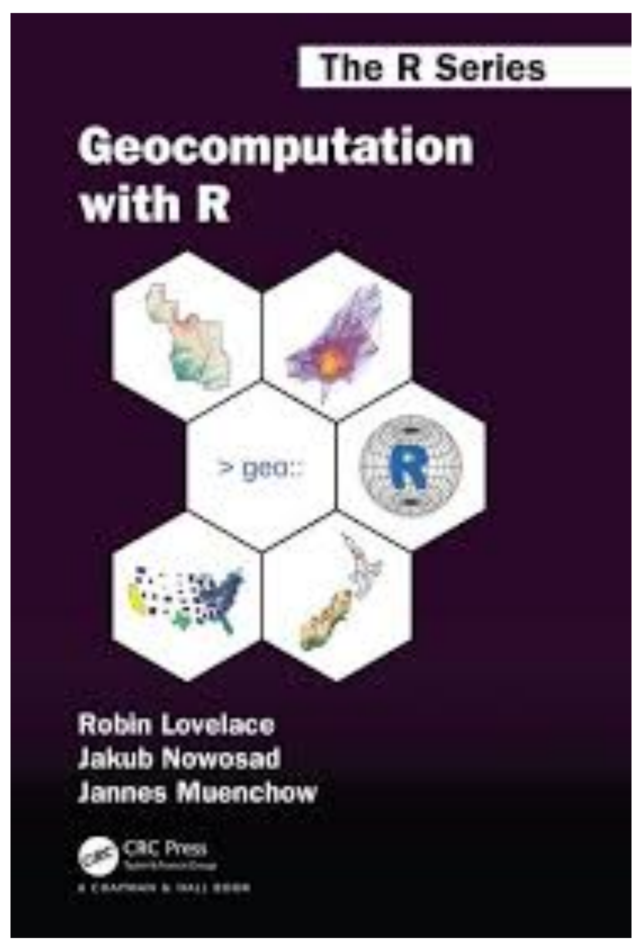
Technologies

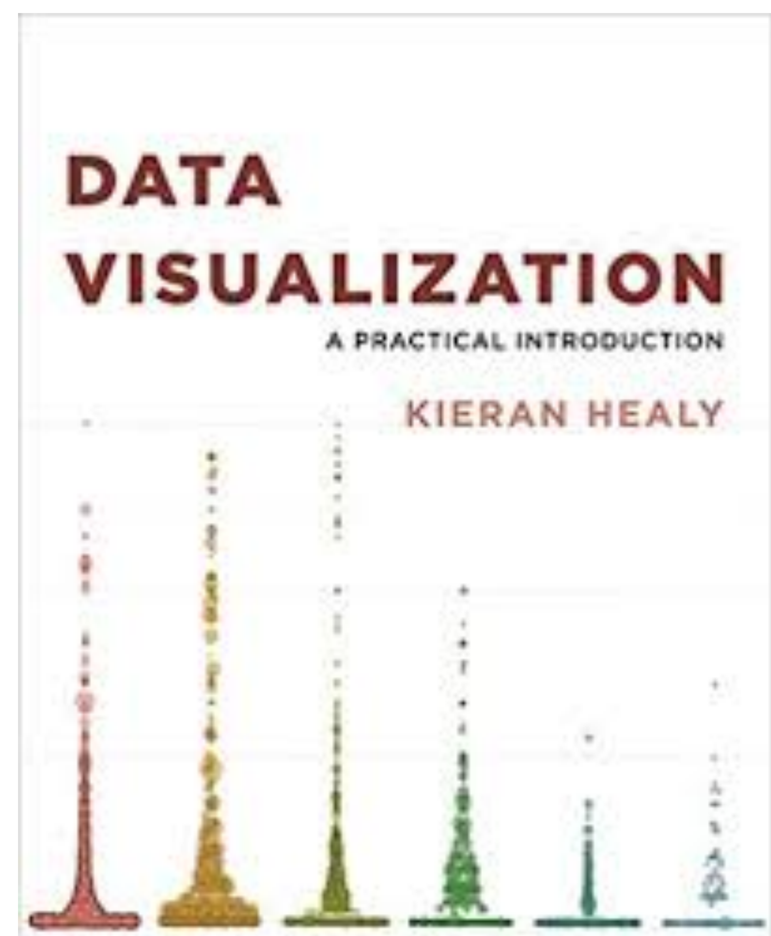
Textbook?

No fixed text book

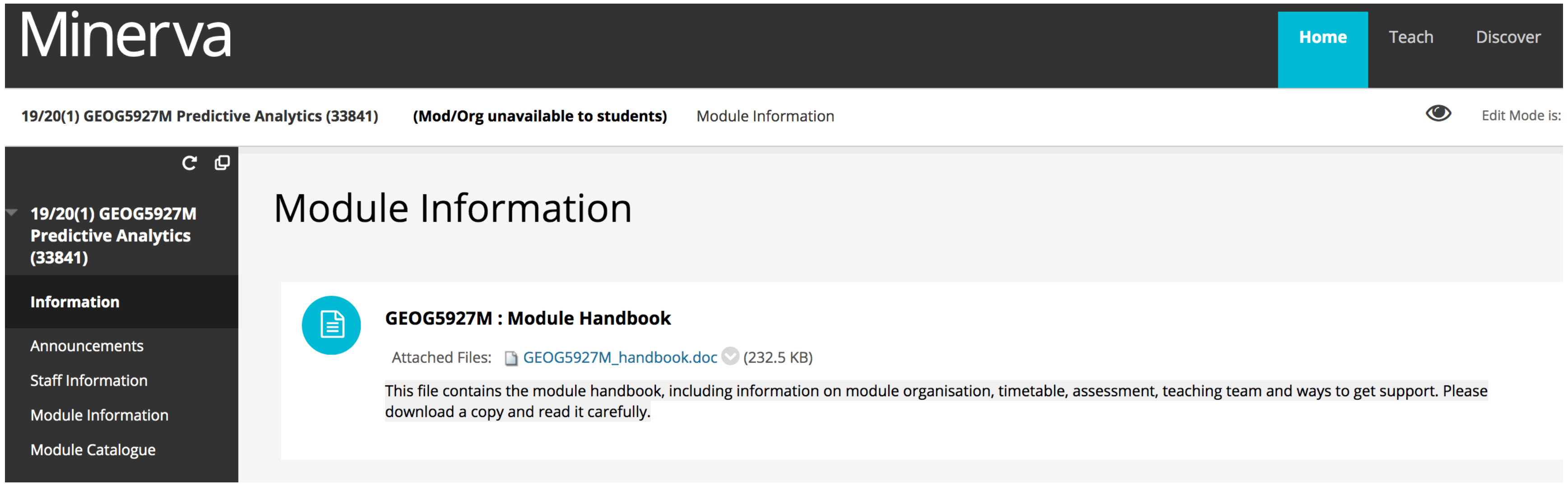
… But





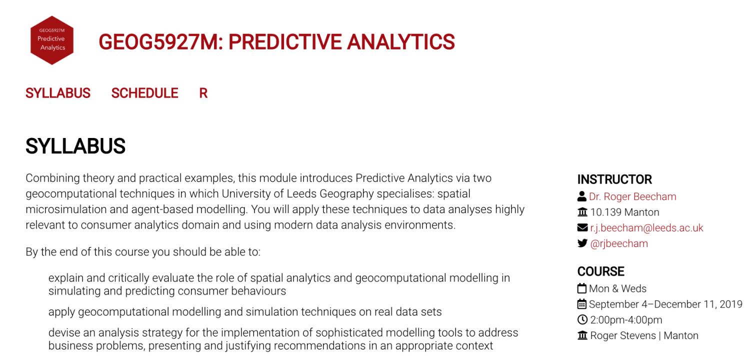


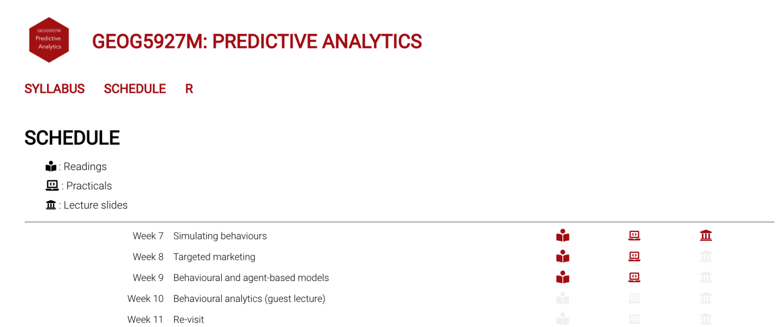
Module Resources



- Module handbook
- Assessment details
- Assessment submission page

Module Resources





- Schedule — week-by-week overview
- Lecture notes
- Lab exercises

How to learn

Come to lectures and labs

try stuff out, engage, ask questions

Independent learning

read and explore – be curious

Coursework

coursework throughout (labs and lectures)

?

Introduction to Predictive Analytics

GEOG5927M
Predictive
Analytics

Forbes Billionaires Innovation

EDITOR'S PICK | 9,075 views | Sep 22, 2019, 08:26pm

A Long View On How Big Data And AI Have Transformed Business Culture

McKinsey & Company

Search jobs Dating Sign in Search UK edition

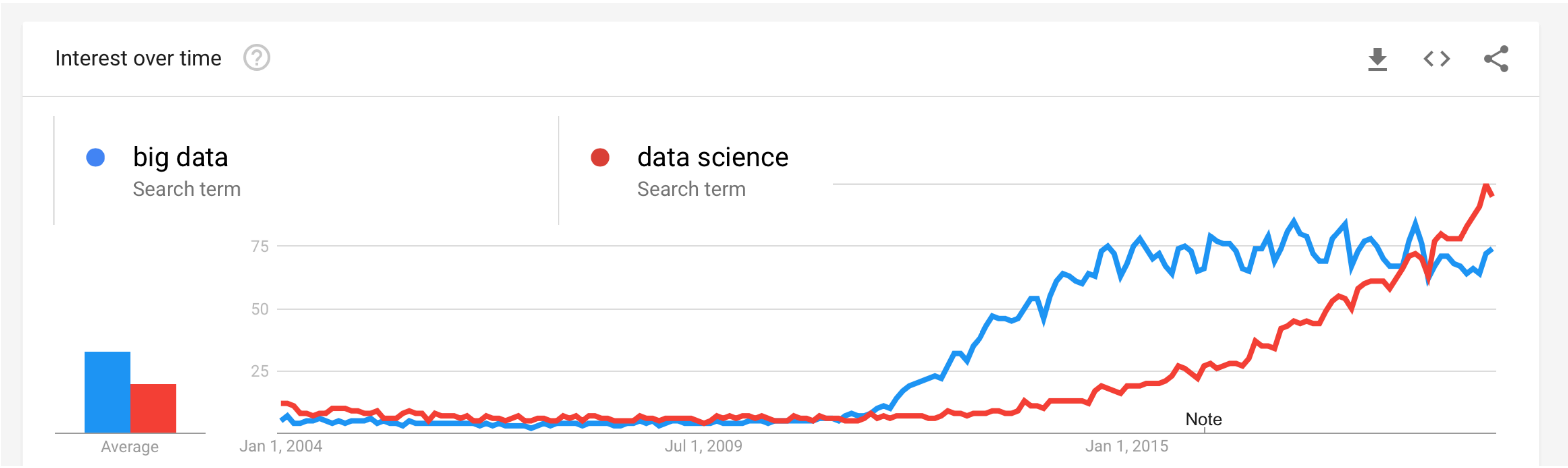
The Guardian

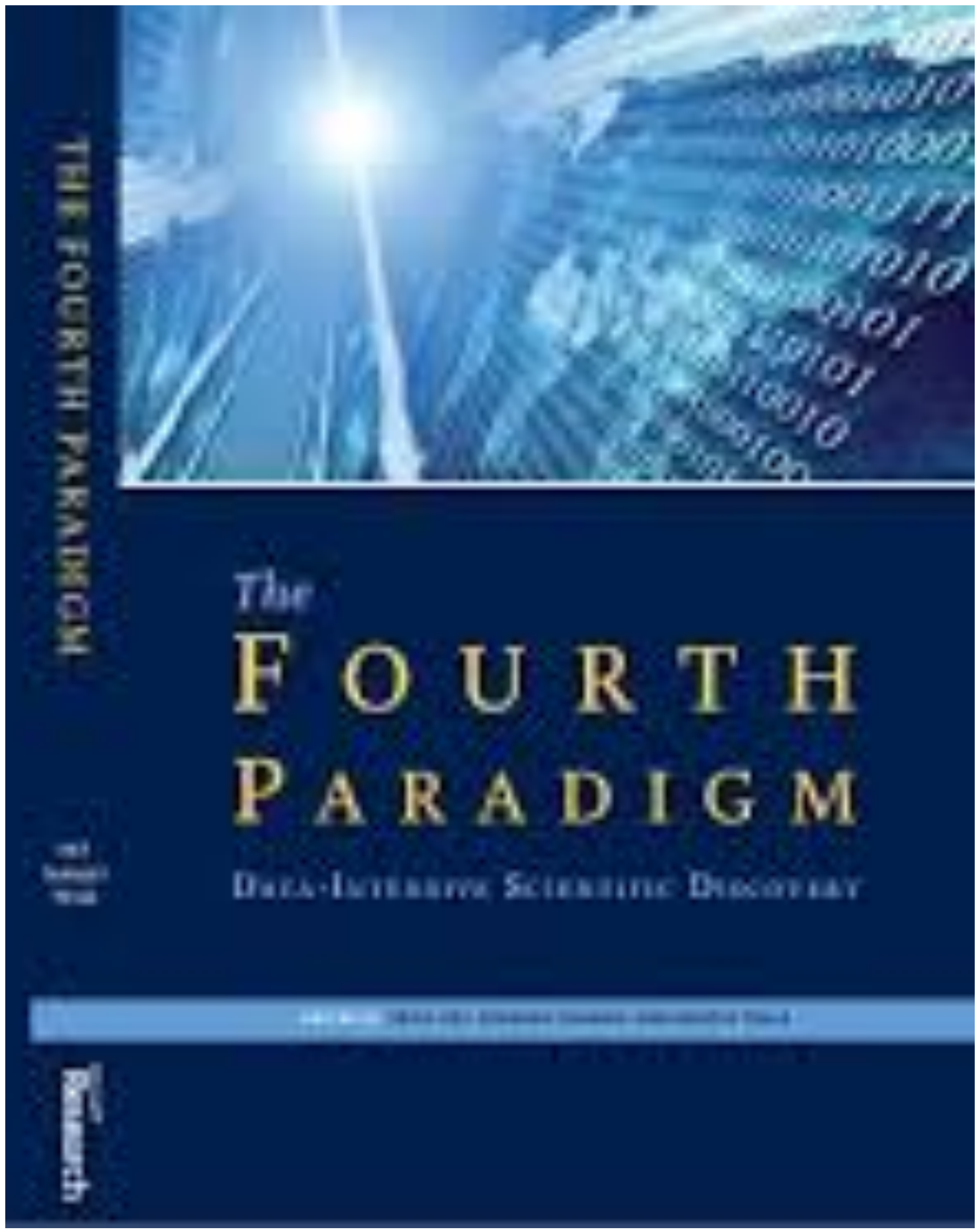
News Opinion Sport Culture Lifestyle More

Open data economy
Media & Tech Network

All the data you need to power your business is free online

Open data available online is improving and streamlining businesses across Europe. We show you how to make the most of it “**big data**” and “**data science”** on Google Trends, Oct 2019





Hey, Tony, Stewart Tansley, and Kristin M. Tolle. The fourth paradigm: data-intensive scientific discovery. Vol. 1. Redmond, WA: Microsoft research, 2009.

1000 years ago – **experimental science**
description of natural phenomena

100s years ago – **theoretical science**
Newton's laws, Maxwell's Equations

<50 years ago – **computational science**
Simulate complex phenomena

today – **data-intensive science**
Generate knowledge through observation (again)

CHRIS ANDERSON MAGAZINE 06.23.08 12:00 PM

THE END OF THEORY: THE DATA DELUGE MAKES THE SCIENTIFIC METHOD OBSOLETE

Illustration: Marian Bantjes

The next generation of scientific discovery will be data-driven as previously unrecognised patterns are discovered by analysing massive and mixed datasets.

David Willets MP, 2013,
Then Minister for Universities and Science

Google

nature

Letter | Published: 19 February 2009

Detecting influenza epidemics using search engine query data

Jeremy Ginsberg, Matthew H. Mohebbi, Rajan S. Patel, Lynnette Brammer, Mark S. Smolinski & Larry Brilliant

Nature **457**, 1012–1014 (2009) | Cite this article

Video: http://goo.gl/4ysAmw

Science Contents News

POLICY FORUM | BIG DATA

The Parable of Google Flu: Traps in Big Data Analysis

David Lazer[1,2,*], Ryan Kennedy[1,3,4], Gary King[3], Alessandro Vespignani[5,6,3]
+ See all authors and affiliations

Liu et al. *EPJ Data Science* (2019) 8:4
https://doi.org/10.1140/epjds/s13688-019-0182-z
EPJ.org
EPJ Data Science
a SpringerOpen Journal
REGULAR ARTICLE Open Access

Inside 50,000 living rooms: an assessment of global residential ornamentation using transfer learning

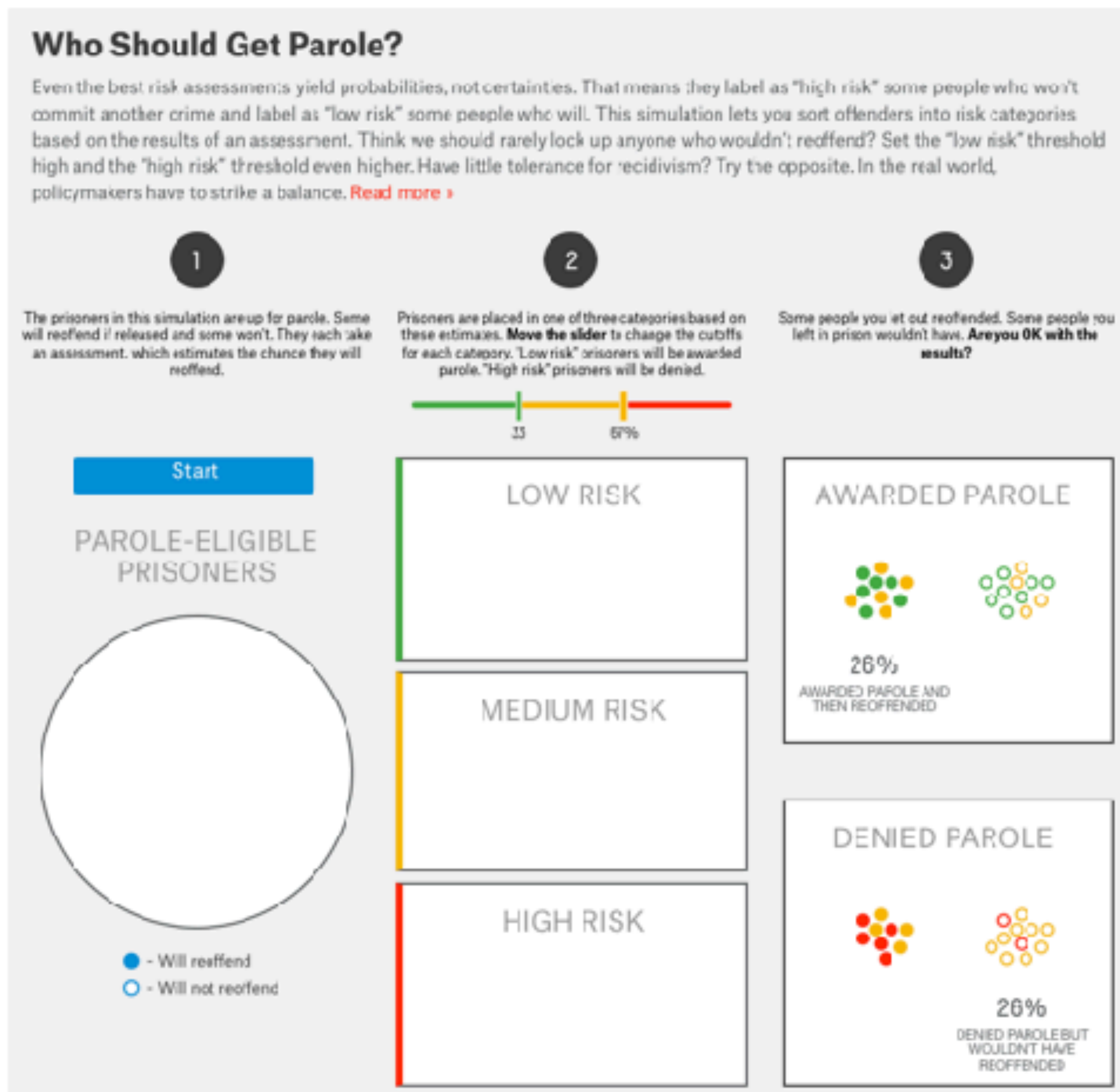
Xi Liu[1], Clio Andris[1*], Zixuan Huang[2] and Sohrab Rahimi[3]

Figure 2 Object detection. Object detection examples for living room images. Ivory and orange bounding boxes in (**a**) and (**b**) are the model's results for plant and book identification, respectively. Green and blue bounding boxes in (**c**) and (**d**) are the model results for wall art and decor identification, respectively

FiveThirtyEight

Should Prison Sentences Be Based On Crimes That Haven't Been Committed Yet?

By Anna Maria Barry-Jester, Ben Casselman and Dana Goldstein
Graphics by Matthew Conlen, Reuben Fischer-Baum and Andy Rossback
Filed under Criminal Justice
Published Aug. 4, 2015



Data mining and machine learning –
Detect hidden patterns in data

Information Visualization –
Explore complex structure and patterns in data that are difficult to expose using computation alone

Predictive analytics —
Use these patterns to predict, under uncertainty, what will happen in future

Assignments

In Assignment #1 we'll be generating a large synthetic dataset of customers and looking for behavioural and demographic associations between individuals to better *target* marketing activity.

GEOG5927M
Predictive Analytics

In Assignment #2 we'll be using data and heuristics to explore and predict how customers will behave and respond to different store formats.

?

break

?

PRACTICAL 1 : SIMULATING BEHAVIOUR

The aim of this session is to create a synthetic population of households in Leeds. You will use an individual-level survey dataset describing individuals' holiday-making behaviour and apply spatial microsimulation to estimate a population-level dataset of these holiday-making behaviours at the household-level in Leeds. In Practical 2, you will use this population to undertake a data analysis to support a targeted marketing strategy.

microdata.csv

15,189 records

Person_ID	OA_GRP	Sex	Ageband	NumberCh	CombinedH	OverSeasAi	UKAirport	OverallHoli	AgeSex	Supergroup
11603	8c	F	a35to49	2	26-30K	LEI	MAN	Excellent	F35to49	Hard-Presse
11285	8c	F	a25to34	0	0-10K	IBZ	MAN	Fair	F25to34	Hard-Presse
13938	8c	M	a50to64	1	16-20K	LCA	BHX	Fair	M50to64	Hard-Presse
10255	8c	F	a25to34	1	26-30K	ALC	LBA	Poor	F25to34	Hard-Presse
831	8c	M	a50to64	0	26-30K	AGA	MAN	Good	M50to64	Hard-Presse
1754	8c	M	a65over	0	Not Answei	DLM	MAN	Good	M65over	Hard-Presse
2330	8c	F	a65over	0	Not Answei	DLM	MAN	Excellent	F65over	Hard-Presse
10818	8c	M	a25to34	0	36-40K	KGS	MAN	Fair	M25to34	Hard-Presse
8237	8c	M	a65over	2	16-20K	FUE	MAN	Good	M65over	Hard-Presse
11508	8c	F	a35to49	2	71-80K	ZTH	LBA	Poor	F35to49	Hard-Presse

ageBand	demographics
incomeBand	demographics
oac	geodemographics
originAirport	preference
destinationAirport	preference/attitude
satisfactionScore	preference/attitude



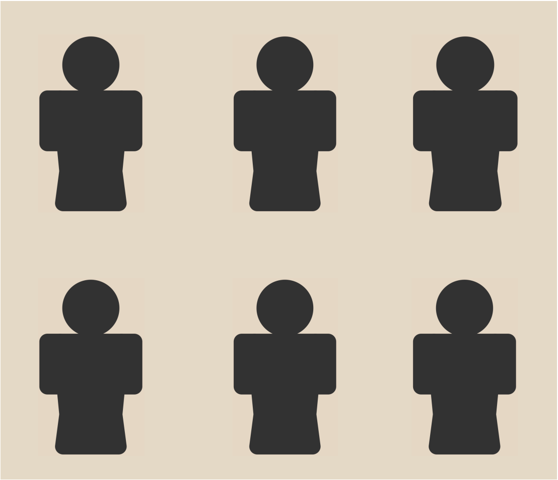


microdata.csv
15,189 records

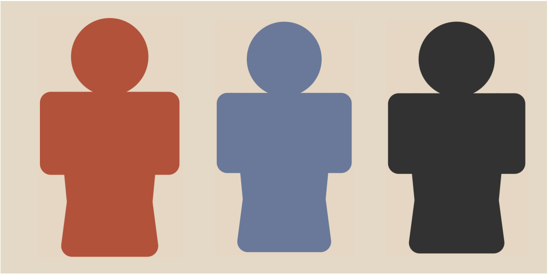
Person_ID	OA_GRP	Sex	Ageband	NumberCh	CombinedH	OverSeasAi	UKAirport	OverallHoli	AgeSex	Supergroup
11603	8c	F	a35to49	2	26-30K	LEI	MAN	Excellent	F35to49	Hard-Presse
11285	8c	F	a25to34	0	0-10K	IBZ	MAN	Fair	F25to34	Hard-Presse
13938	8c	M	a50to64	1	16-20K	LCA	BHX	Fair	M50to64	Hard-Presse
10255	8c	F	a25to34	1	26-30K	ALC	LBA	Poor	F25to34	Hard-Presse
831	8c	M	a50to64	0	26-30K	AGA	MAN	Good	M50to64	Hard-Presse
1754	8c	M	a65over	0	Not Answe	DLM	MAN	Good	M65over	Hard-Presse
2330	8c	F	a65over	0	Not Answe	DLM	MAN	Excellent	F65over	Hard-Presse
10818	8c	M	a25to34	0	36-40K	KGS	MAN	Fair	M25to34	Hard-Presse
8237	8c	M	a65over	2	16-20K	FUE	MAN	Good	M65over	Hard-Presse
11508	8c	F	a35to49	2	71-80K	ZTH	LBA	Poor	F35to49	Hard-Presse

Spatial Microsimulation

Census data
high-level | low in detail
population-level | complete



Survey data
individual-level | rich in detail
small-scale | unrepresentative



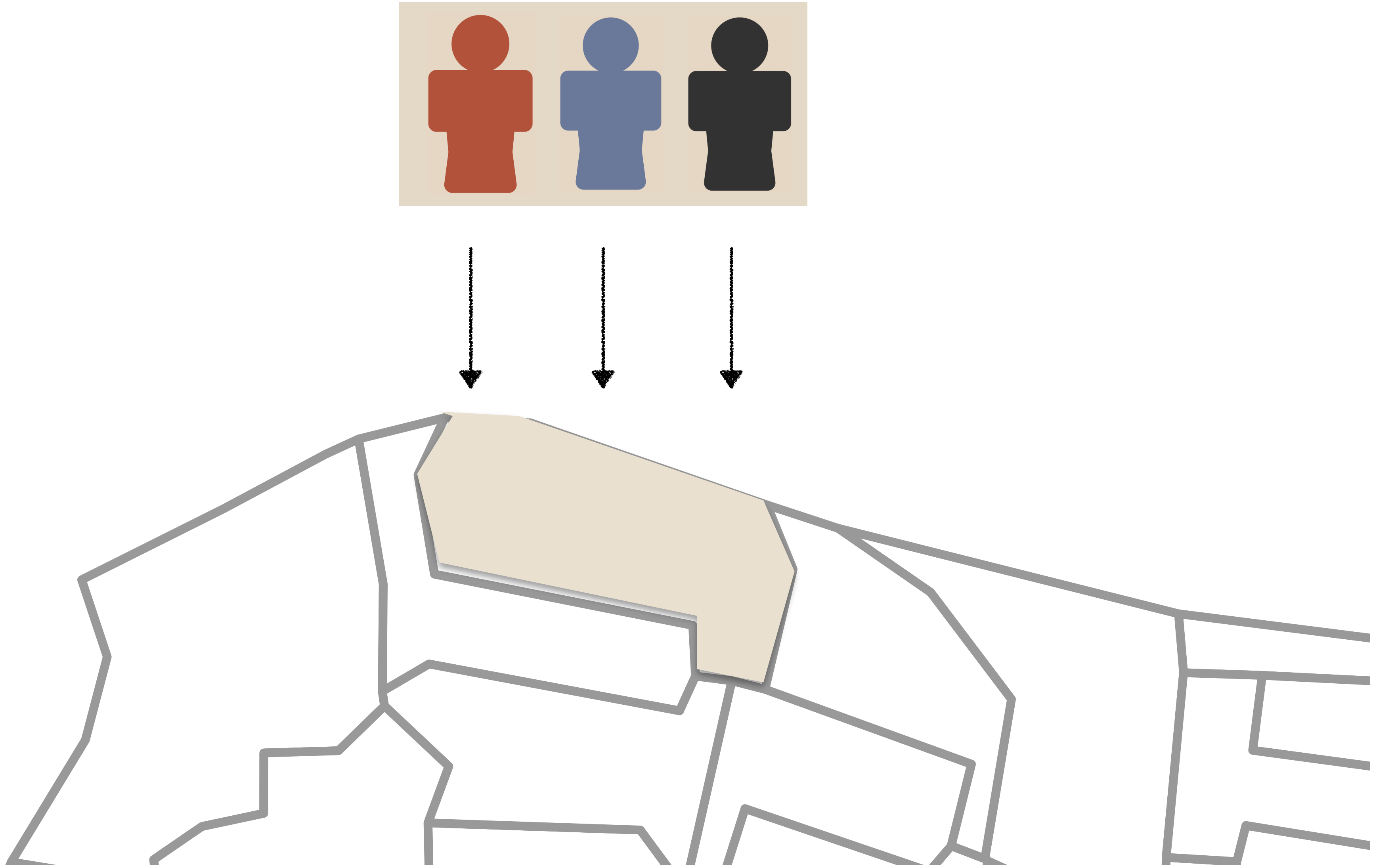
Spatial Microsimulation

The creation, analysis and modelling of individual-level data allocated to geographic zones.

Lovelace & Dumont 2016

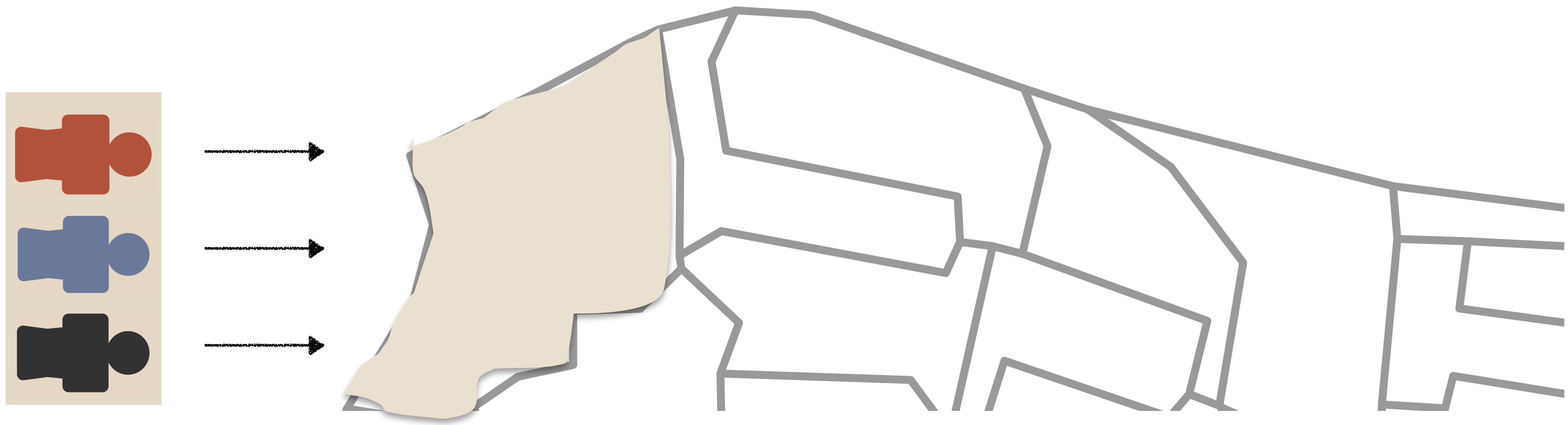
Spatial Microsimulation

survey data
individuals

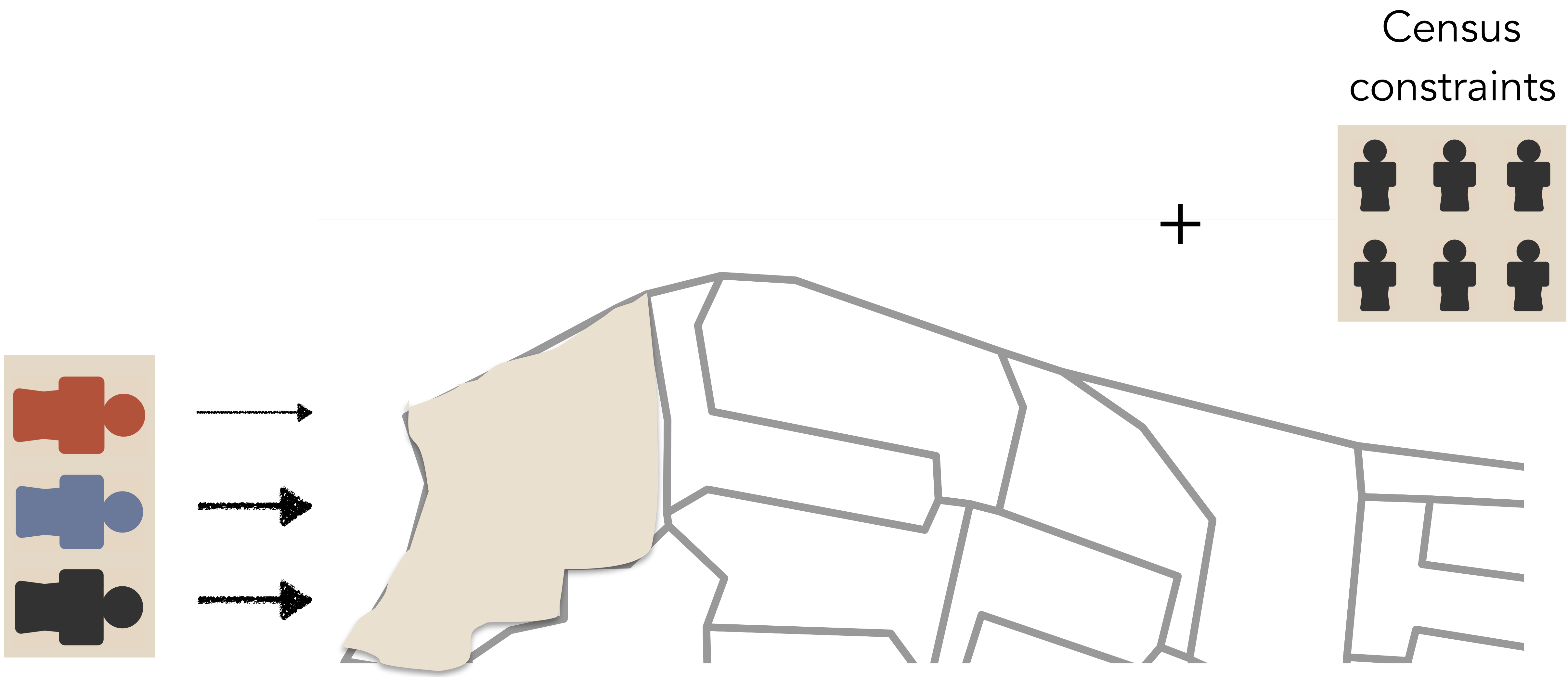


Spatial Microsimulation

Spatial Microsimulation



Spatial Microsimulation



Microsimulation does not generate ***new*** *data*

--

merely copies of existing data

Lovelace & Dumont 2016

Examples

Health : smoking [Tomintz et al. 2008]

- **Why?** Reported in individual surveys, but not population-level and not from place-to-place

- **Benefits** : could be used to target/locate smoking support clinics

- '**Benefits'** : could be used by a Tobacco company for targeting investment

Examples

Economic policy : evaluation [De Agostini et al. 2014}

- **Why?** Simulate / spread impacts inferred from individual-level data over an entire country

- **Benefits** : quantify (under uncertainty) the impacts of a regressive welfare reform at the country-level

- **Benefits** : evidence-based decision-making

Examples

Transport : simulating behaviour [Lovelace 2014]

- **Why?** When designing infrastructure, want to know about the distribution of individuals meeting a particular set of characteristics

- **Benefits** : provide evidence around likely winners and losers of a new infrastructure investment

Assumptions

- Individual-level microdata are representative of the study area

- Target variable is dependent on the constraint variables in a way that is relatively constant over space and time

- Input microdataset and constraints are sufficiently rich and detailed to reproduce the full diversity of individuals and areas in the study region

GEOG5927M: PREDICTIVE ANALYTICS

SYLLABUS SCHEDULE R

PRACTICAL 1 : SIMULATING BEHAVIOUR

The aim of this session is to create a synthetic population of households in Leeds. You will use an individual-level survey dataset describing individuals' holiday-making behaviour and apply spatial microsimulation to estimate a population-level dataset of these holiday-making behaviours at the household-level in Leeds. In Practical 2, you will use this population to undertake a data analysis to support a targeted marketing strategy.

R Studio

Dataset

Person_ID	OA_GRP	Sex	Ageband	NumberCh	CombinedH	OverSeasAi	UKAirport	OverallHoli	AgeSex	Supergroup
11603	8c	F	a35to49	2	26-30K	LEI	MAN	Excellent	F35to49	Hard-Presse
11285	8c	F	a25to34	0	0-10K	IBZ	MAN	Fair	F25to34	Hard-Presse
13938	8c	M	a50to64	1	16-20K	LCA	BHX	Fair	M50to64	Hard-Presse
10255	8c	F	a25to34	1	26-30K	ALC	LBA	Poor	F25to34	Hard-Presse
831	8c	M	a50to64	0	26-30K	AGA	MAN	Good	M50to64	Hard-Presse
1754	8c	M	a65over	0	Not Answe	DLM	MAN	Good	M65over	Hard-Presse
2330	8c	F	a65over	0	Not Answe	DLM	MAN	Excellent	F65over	Hard-Presse
10818	8c	M	a25to34	0	36-40K	KGS	MAN	Fair	M25to34	Hard-Presse
8237	8c	M	a65over	2	16-20K	FUE	MAN	Good	M65over	Hard-Presse
11508	8c	F	a35to49	2	71-80K	ZTH	LBA	Poor	F35to49	Hard-Presse

microdata.csv

15,189 records

ZoneID	Person_ID	OA_GRP	Sex	Ageband	NumberCh	CombinedH	OverSeasAi	UKAirport	OverallHoli	AgeSex	Supergroup
E00056750	11603	8c	F	a35to49	2	26-30K	LEI	MAN	Excellent	F35to49	Hard-Pressed Living
E00056750	11285	8c	F	a25to34	0	0-10K	IBZ	MAN	Fair	F25to34	Hard-Pressed Living
E00056750	13938	8c	M	a50to64	1	16-20K	LCA	BHX	Fair	M50to64	Hard-Pressed Living
E00056750	10255	8c	F	a25to34	1	26-30K	ALC	LBA	Poor	F25to34	Hard-Pressed Living
E00056750	831	8c	M	a50to64	0	26-30K	AGA	MAN	Good	M50to64	Hard-Pressed Living
E00056750	1754	8c	M	a65over	0	Not Answe	DLM	MAN	Good	M65over	Hard-Pressed Living
E00056750	2330	8c	F	a65over	0	Not Answe	DLM	MAN	Excellent	F65over	Hard-Pressed Living
E00056750	10818	8c	M	a25to34	0	36-40K	KGS	MAN	Fair	M25to34	Hard-Pressed Living
E00056750	8237	8c	M	a65over	2	16-20K	FUE	MAN	Good	M65over	Hard-Pressed Living
E00056750	11508	8c	F	a35to49	2	71-80K	ZTH	LBA	Poor	F35to49	Hard-Pressed Living

simulated_population.csv

320,596 records

?