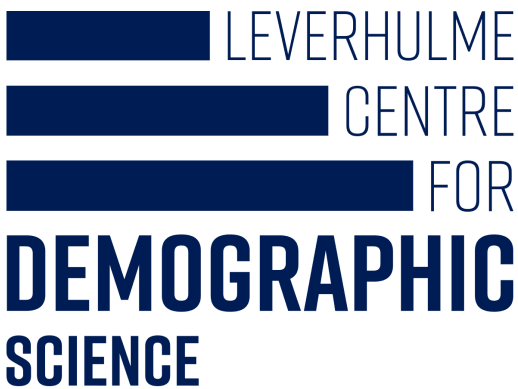
Adventures in Open & Reproducible Social Science

Charlie Rahal
LCDS, Department of Sociology and Nuffield College

University of Oxford, 18th April, 2023



In the spirit of 'Open Science':

crahal.github.io/files/adventures_in_social_science.pdf

and *all* of the research you see here today is at:

github.com/crahal

Lots of examples: hopefully something for everyone!

Everything here is made with free, and 'Open' tools:

- Linux (operating system)
- Python (general purpose programming)
- LaTeX (typesetting)

Everything here is made with free, and 'Open' tools:

- Linux (operating system)
- Python (general purpose programming)
- LaTeX (typesetting)

Why mention this?

Everything here is made with free, and 'Open' tools:

- Linux (operating system)
- Python (general purpose programming)
- LaTeX (typesetting)

Why mention this? We're talking about 'Open' Science!

What is 'Open Science'?

'The movement to make scientific research (including publications, data, physical samples, and software) and its dissemination accessible to all levels of an inquiring society, amateur or professional'

Pragmatically:

1. Allows creation of value and advancement.
2. Allows verification of the scientific record.
3. Allows public disclosure and engagement.
4. Allows science to become accessible and transparent.

What motivates a talk about 'Open Science'?

- Three anecdotes from my time as a doctoral student.

What motivates a talk about 'Open Science'?

- Three anecdotes from my time as a doctoral student.

  1. Make sure your coauthors have code.

What motivates a talk about 'Open Science'?

- Three anecdotes from my time as a doctoral student.
  1. Make sure your coauthors have code.
  2. Critically analyze everything.

What motivates a talk about 'Open Science'?

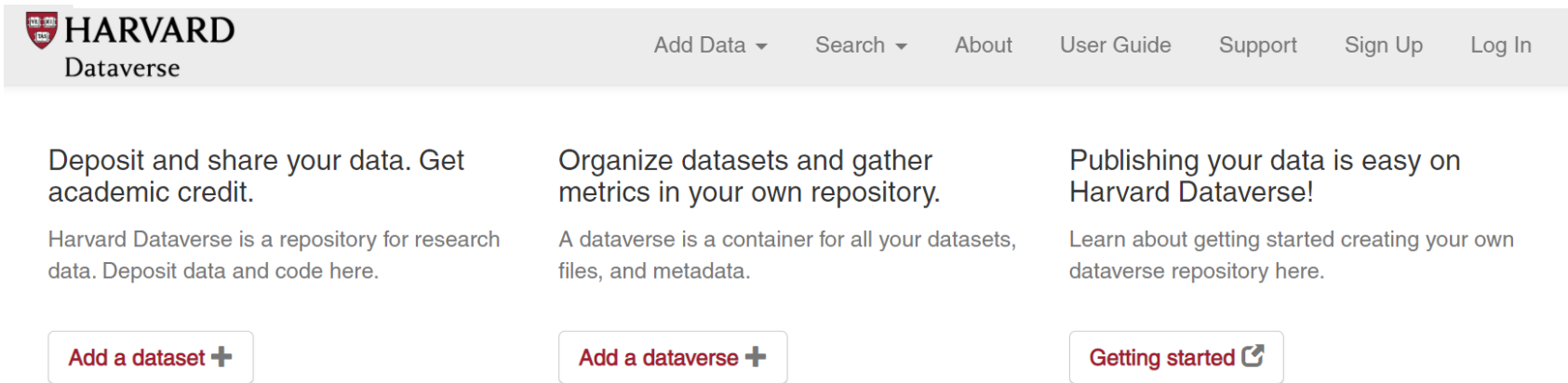
- Three anecdotes from my time as a doctoral student.
  1. Make sure your coauthors have code.
  2. Critically analyze everything.
  3. Consider using open-source software.

What does 'Open Science' look like?

1. Preprints and 'Open Access'.
2. Publicly available data.
3. Open source code

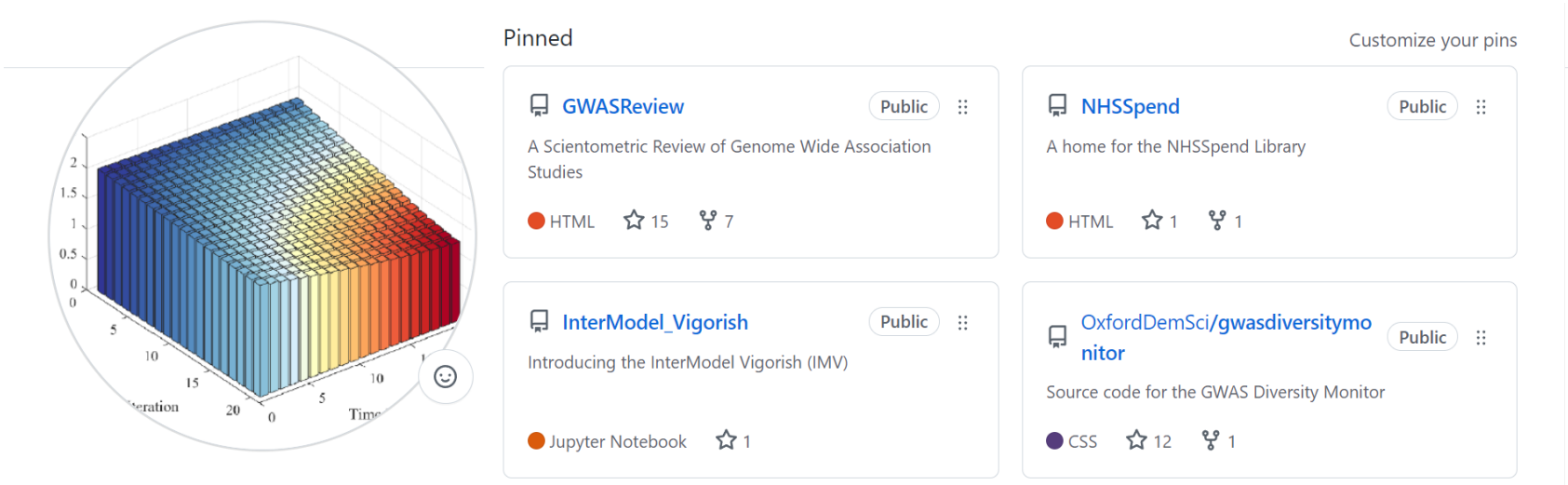
What does 'Open Science' look like?

1. Preprints and 'Open Access'.
2. Publicly available data.
3. Open source code



What does 'Open Science' look like?

1. Preprints and 'Open Access'.
2. Publicly available data.
3. Open source code



These slides are made with an open source typesetting language!

I_make_this_slide.tex

```
\begin{frame}[]
  These slides are made with an open source typesetting language!
  \begin{center}
    \begin{scriptsize}
      \begin{minipage}{.95\linewidth}
        \begin{myenv}{guess$\_$what$\_$this$\_$does.py}
          \begin{minted}{latex}
            # Hello friends :)
          \end{minted}
        \end{myenv}
      \end{minipage}
    \end{scriptsize}
  \end{center}
\end{frame}
```

Open source statistical code comes in the form of Python (or R):

guess_what_this_does.py

```
import json, io, imageio, requests
from PIL import Image

def get_cat_bounded(url):
    api_ret = requests.get(url)
    if api_ret.status_code==200:
          text = json.loads(api_ret.text)
        while (text[0]['width']<=400 or text[0]['width']>=600) or \
              (text[0]['height']<=400 or text[0]['height']>=600):
            api_ret = requests.get(url)
            if api_ret.status_code==200:
                 text = json.loads(api_ret.text)
        r = requests.get(api_text[0]['url'])
        return Image.open(io.BytesIO(r.content))

url = 'https://api.thecatapi.com/v1/images/search'
images = []
for _ in range(0, 50):
    images.append(get_cat_bounded(url))
imageio.mimsave('CATS.gif', images, fps=1)
```

Thats right: Cats!

Why show you this? Well, Python is great at lots of useful things:

1. Scientific Computing (SciPy, numpy)
2. Data visualisation (Matplotlib, seaborn)
3. Web scraping (requests)
4. Data wrangling (Pandas)

If you remember one thing from my talk today: try to learn a computer language as early as possible!

Why show you this? Well, Python is great at lots of useful things:

1. Scientific Computing (SciPy, numpy)
2. Data visualisation (Matplotlib, seaborn)
3. Web scraping (requests)
4. Data wrangling (Pandas)

If you remember one thing from my talk today: try to learn a computer language as early as possible!

Perhaps controversial, but most classes here now taught in R.

Why show you this? Well, Python is great at lots of useful things:

1. Scientific Computing (SciPy, numpy)
2. Data visualisation (Matplotlib, seaborn)
3. Web scraping (requests)
4. Data wrangling (Pandas)

If you remember one thing from my talk today: try to learn a computer language as early as possible!

Perhaps controversial, but most classes here now taught in R.

Python is great at building new datasets. But why bother?

> *'We have been using the same data sets over and over again, and since we wanted new answers, we have been developing new econometric techniques to try to transform the data, and get more meaningful information out of them. Just having a fresh, new data set brings a whole new perspective, and I think people are starting to realize that, and gradually people are becoming more interested in data collection itself'*
>
> *A. Cavallo, 15 November 2012 in Taylor et al. (2014)*

New datasets allow fresh insight and are typically open!

A lot of my work involves semi-structured data!

> *"A form of structured data that does not obey the tabular structure of data models associated with relational databases or other forms of data tables, but nonetheless contains tags or other markers to separate semantic elements"*

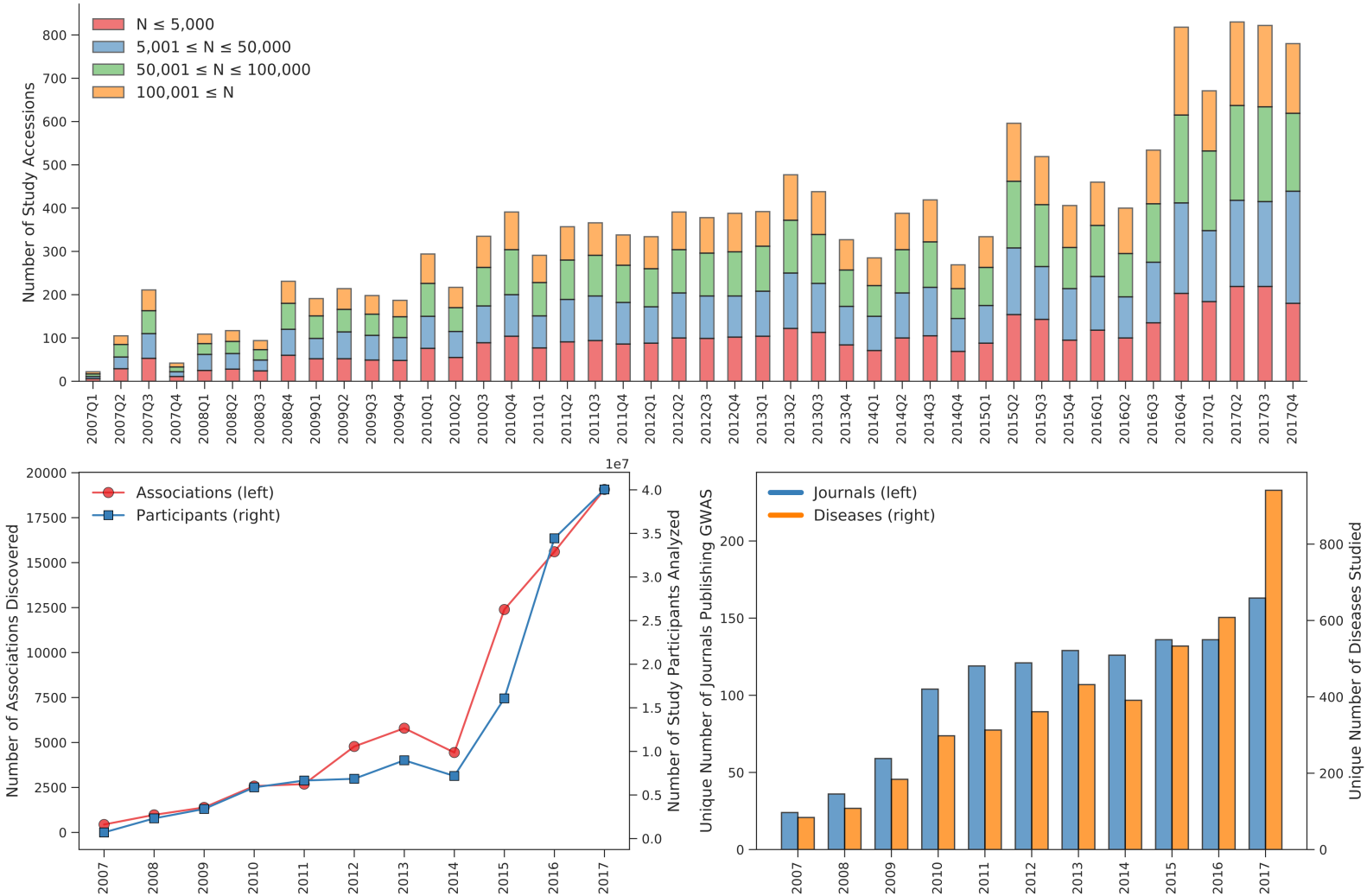
Differs from **tabular** surveys classical social scientists used to.

Often, our goal is to parse it and make it tabular all the same.

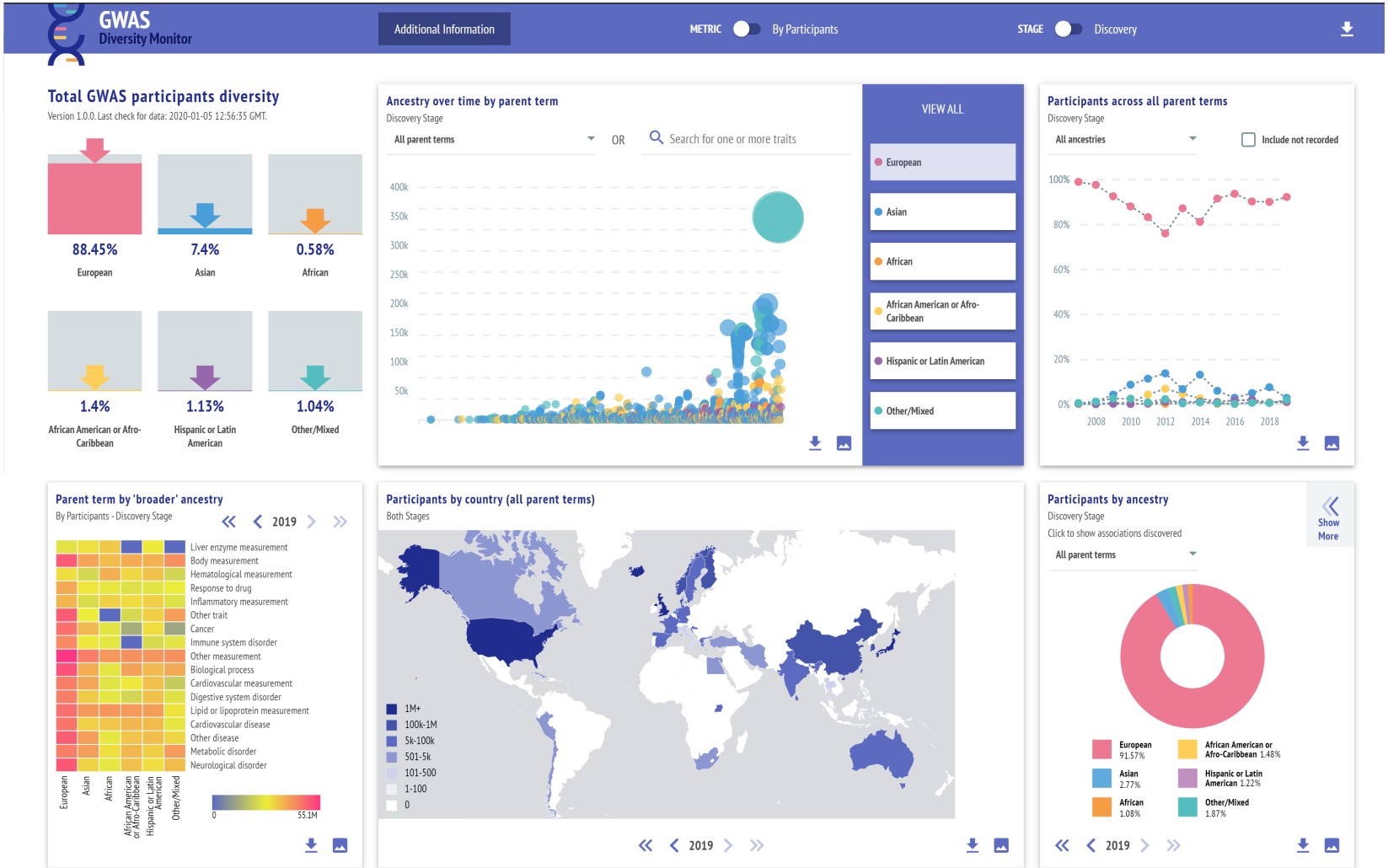
A lot of my semi-structured work is 'scientometric' and 'bibliometric' in nature:

- Scientometrics is the field of study which concerns itself with measuring and analysing science.
- Bibliometrics is the use of statistical methods to analyze quantitative information pertaining to books, articles and other publications.

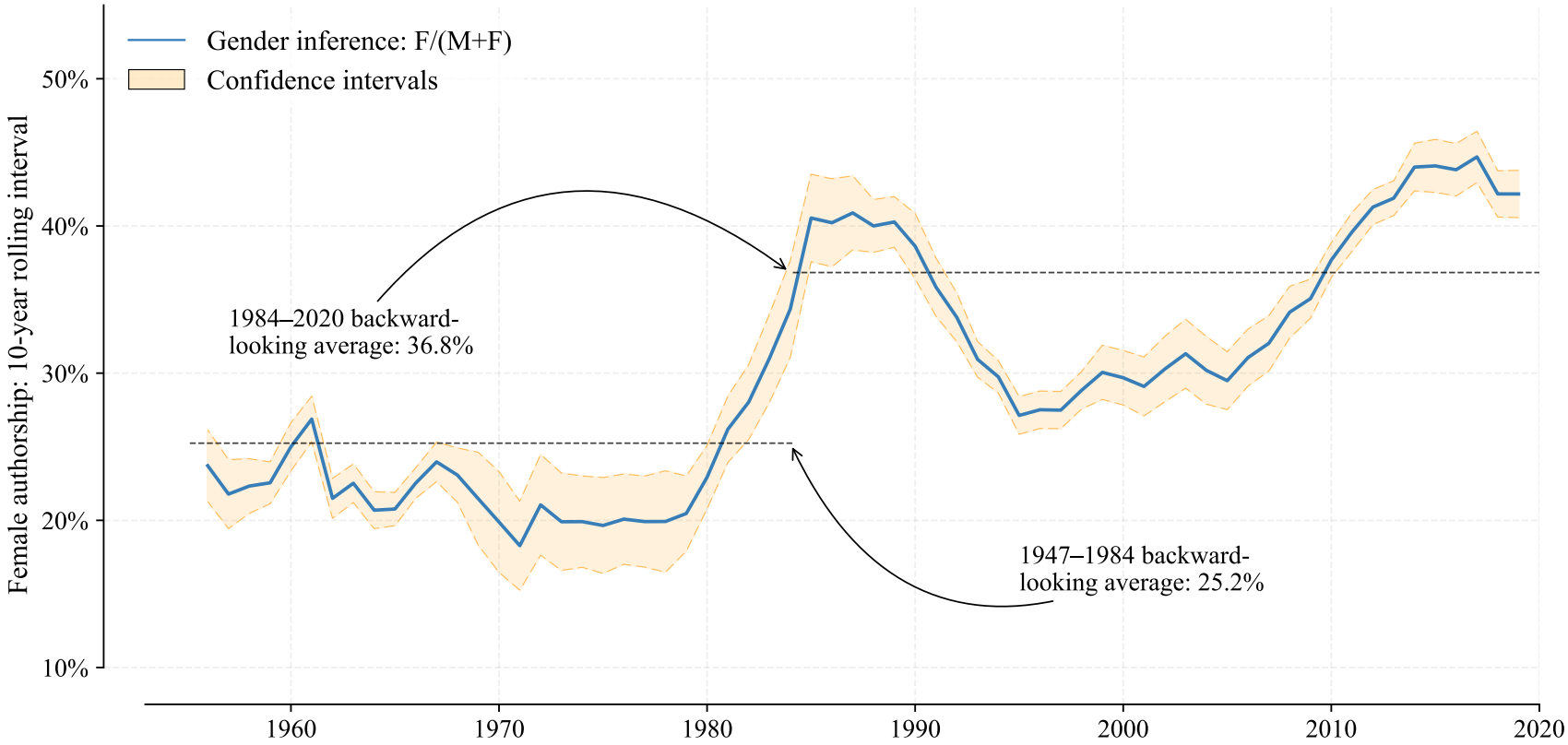
These fields are naturally related to 'Open Science'.



'The Scientometrics of Genome Wide Association Studies'
with Melinda C. Mills

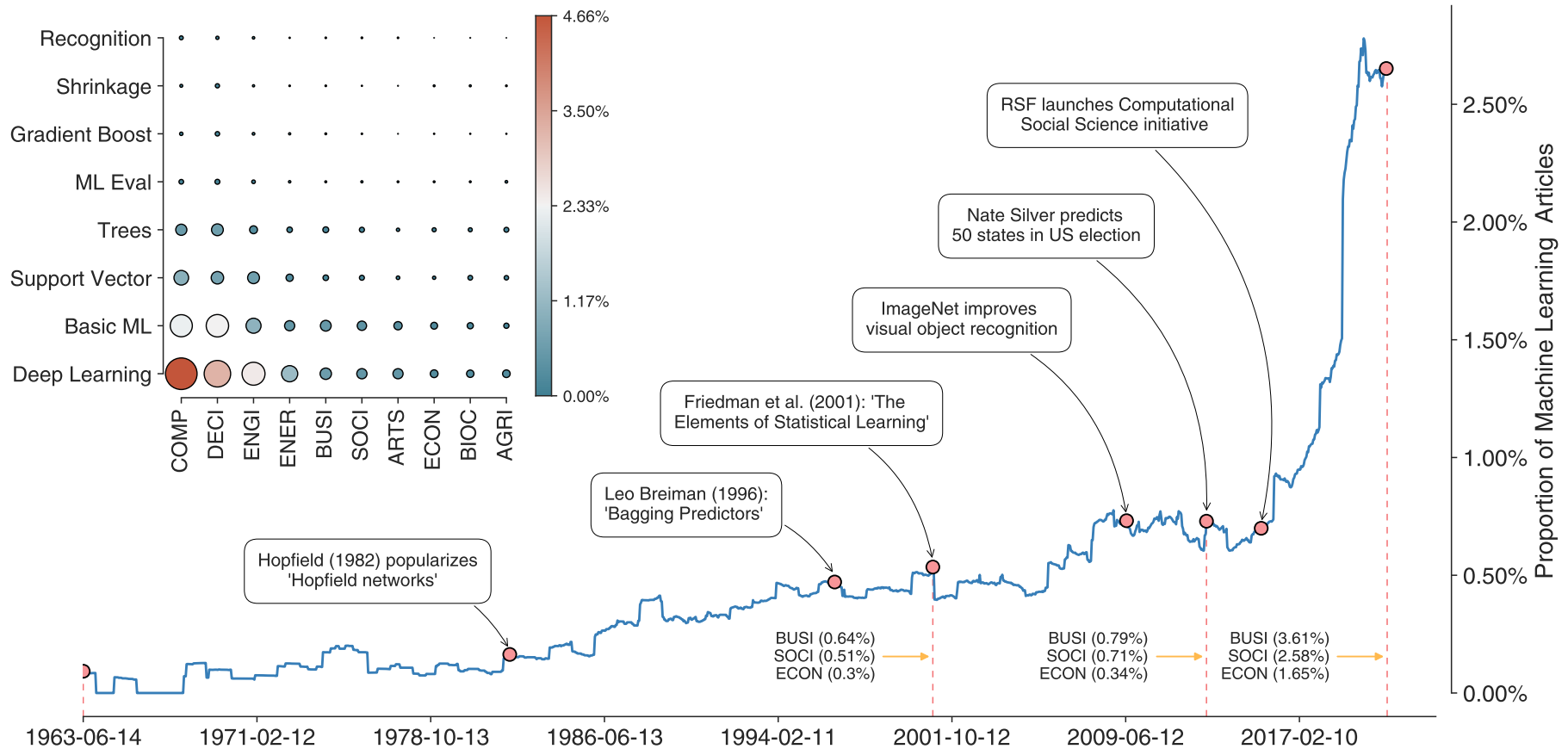


‘The GWAS Diversity Monitor Tracks Diversity in Real Time’
with Melinda C. Mills



‘Population Studies at 75 Years: An empirical review’

With Melinda C. Mills



'The Rise of Machine Learning in the Academic Social Sciences'

With Dave Kirk and Mark Verhagen

Administrative Data is an invalueable resource for Good.

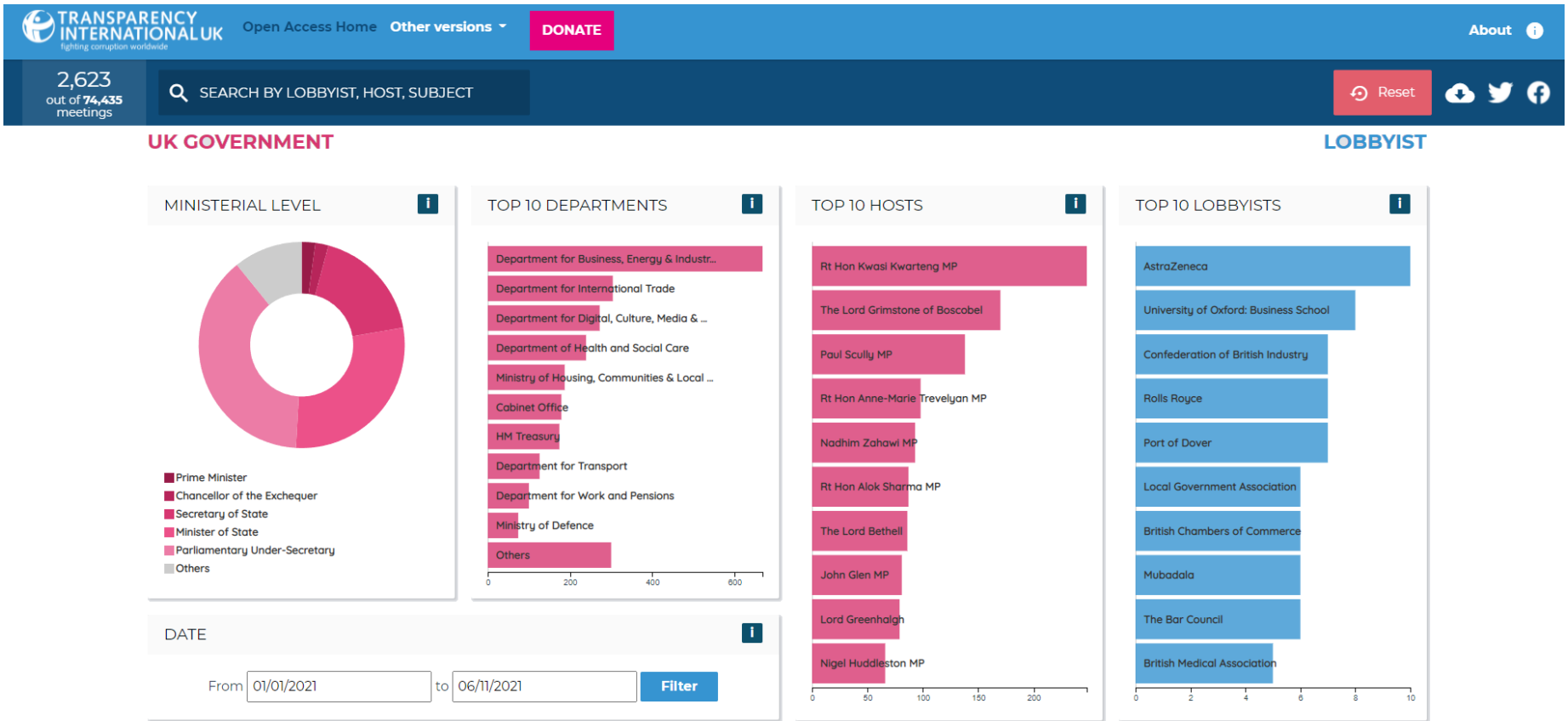
ADR UK (2018)

- UK leads world in terms of 'Openness' of government data.
- Value of public government data is increasingly recognised.
  - Increasingly recognised by the ESRC (e.g. ADRNs).
- For better or worse:
  - I'm interested in 'hard to use', semi-structured data.

Administrative Data is an invalueable resource for Good.

ADR UK (2018)

- UK leads world in terms of 'Openness' of government data.
- Value of public government data is increasingly recognised.
  - Increasingly recognised by the ESRC (e.g. ADRNs).
- For better or worse:
  - I'm interested in 'hard to use', semi-structured data.
  - Using curated data would be too easy.



openaccess.transparency.org.uk

With Ian Knowles, Paulo Serôdio, David Stuckler and TI UK

Procurement Data

- David Cameron opened data on government contracts (2010).
- This provided a totally new type of semi-structured data.
- From from local authorities, to the NHS, to central government.
- This sounds like it might be useful for academic research...
- However, the raw data is **unbelievably** messy.

Procurement Data

- David Cameron opened data on government contracts (2010).
- This provided a totally new type of semi-structured data.
- From from local authorities, to the NHS, to central government.
- This sounds like it might be useful for academic research...
- However, the raw data is **unbelievably** messy.

If only there was a **Government Data Service** to curate it...

Lets list a few of the problems with procurement data. Some are common to semi-structured data, some are not:

1. Some organisations don't believe in their legal obligation.
2. Every file is in a heterogeneous format in a different location (**thousands** of locations across the internet.)
3. Data commonly converted from a tabular file...

Lets list a few of the problems with procurement data. Some are common to semi-structured data, some are not:

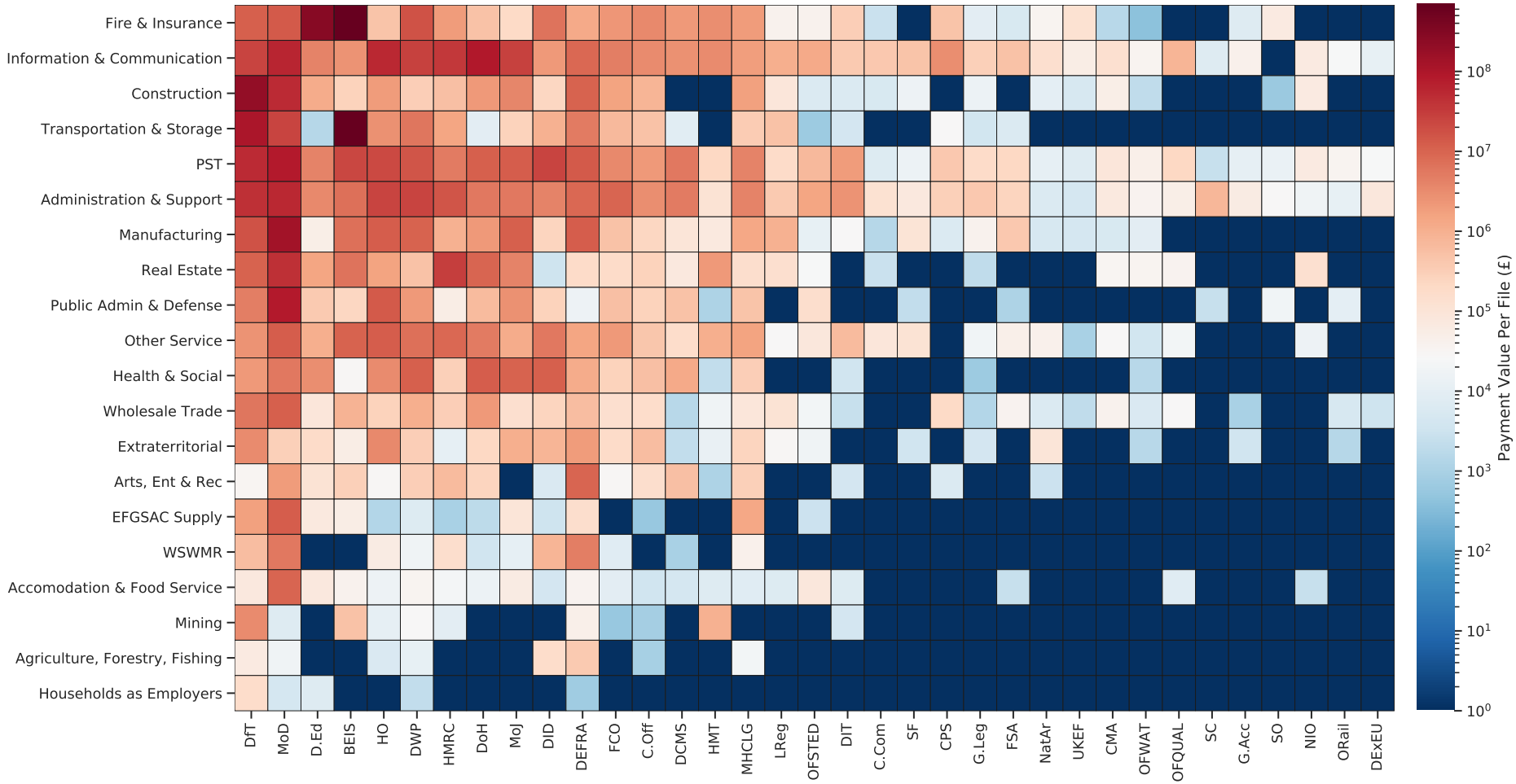
1. Some organisations don't believe in their legal obligation.
2. Every file is in a heterogeneous format in a different location (**thousands** of locations across the internet.)
3. Data commonly converted from a tabular file... **into** a .pdf! (This is a **direct** downgrade in the Berners-Lee 5*schema.)
4. Naturally, of course, there are no supplier UIDs.

Just one quick example, followed by what we can do...

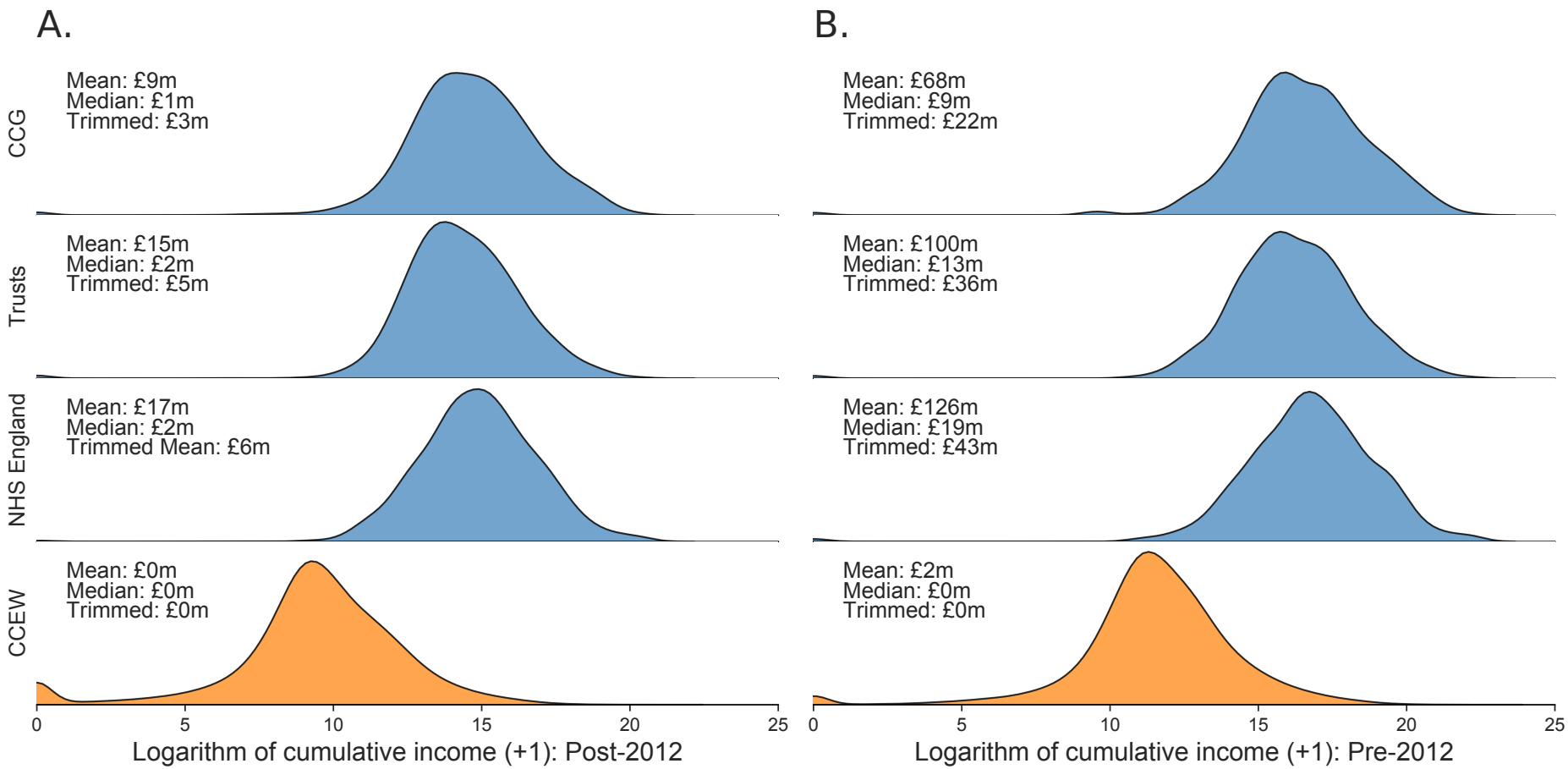
A3131

Expenditure Over Threshold

Department family	Entity	Date	Expense Type	Expense area	Supplier	Transaction number	AP Amount (£)	VAT registration number	Purchase invoice number
Department of Health	NHS Bexley CCG	30/04/201	Clinical&Medical-Othe Public Sector	COMMUNITY SERVICES	BEXLEY COUNCIL	11595633	658,000.00		6100071658
Department of Health	NHS Bexley CCG	30/04/201	Clinical&Medical-Othe Public Sector	COMMUNITY SERVICES	BEXLEY COUNCIL	11595652	100,000.00		6100071661
Department of Health	NHS Bexley CCG	30/04/201	Other Gen Supplies & Srv	COMMUNITY SERVICES	BEXLEY COUNCIL	11595677	90,000.00		6100071674
Department of Health	NHS Bexley CCG	30/04/201	Clinical&Medical-Othe Public Sector	Winter Resilience	BEXLEY COUNCIL	11595692	64,000.00		6100071687
Department of Health	NHS Bexley CCG	30/04/201	Hcare Srv Rec Fdtn Trust-Contract Baseline	ACUTE COMMISSIONING	UNIVERSITY COLLEGE LONDON HOSPITALS NHS FOUNDATION TRUST	11607291	68,256.00	654948687	5010086544
Department of Health	NHS Bexley CCG	30/04/201	Cont Care- Adult 100% Fully Funded	CHC ADULT FULLY FUNDED	CRAEGMOOR OAKS SIDCUP PRIORY	11609225	76,409.99		9510007000
Department of Health	NHS Bexley CCG	30/04/201	Clinical&Medical-Voluntary Sector	COMMISSIONING - NON ACUTE	BEXLEY COUNCIL	11615604	30,000.00		6100070989
Department of Health	NHS Bexley CCG	30/04/201	Clinical&Medical-Independent Sector	ACUTE COMMISSIONING	INHEALTH LTD	11615704	38,528.26	787424567	IHL00056882
Department of Health	NHS Bexley CCG	30/04/201	Hcare Srv Rec Fdtn Trust-Contract Baseline	ACUTE COMMISSIONING	MEDWAY NHS FOUNDATION TRUST	11623774	58,284.00		2383
Department of Health	NHS Bexley CCG	30/04/201	Clinical&Medical-Voluntary Sector	HOSPICES	GREENWICH & BEXLEY COMMUNITY HOSPICE	11628377	218,000.00		877
Department of Health	NHS Bexley CCG	30/04/201	Hcare Srv Rec NHS Trust-Contract Baseline	ACUTE COMMISSIONING	LEWISHAM & GREENWICH NHS TRUST	11635401	4,440,144.00		8466267
Department of Health	NHS Bexley CCG	30/04/201	Rent	PROGRAMME PROJECTS	NHS PROPERTY SERVICES LTD	11641657	218,724.14		48409294
Department of Health	NHS Bexley CCG	30/04/201	Rent	PROGRAMME PROJECTS	NHS PROPERTY SERVICES LTD	11642864	25,672.53		48413715
Department of Health	NHS Bexley CCG	30/04/201	Hcare Srv Rec NHS Trust-Contract Baseline	ACUTE COMMISSIONING	DARTFORD AND GRAVESHAM NHS TRUST	11642923	-32,306.00		2000004184
Department of Health	NHS Bexley CCG	30/04/201	Hcare Srv Rec NHS Trust-Contract Baseline	ACUTE COMMISSIONING	DARTFORD AND GRAVESHAM NHS TRUST	11642927	250,000.00		1000044139
Department of Health	NHS Bexley CCG	30/04/201	Hcare Srv Rec Fdtn Trust-Contract Baseline	PLANNED CARE	KING`S COLLEGE HOSPITAL NHS FOUNDATION TRUST	11654529	-3,042,119.26		CR00018970
Department of Health	NHS Bexley CCG	30/04/201	Clinical&Medical-Commercial Sector	COMMUNITY SERVICES	BEXLEY HEALTH LTD	11655039	26,930.40		BHL0662
Department of Health	NHS Bexley CCG	30/04/201	Hcare Srv Rec NHS Trust-Contract Baseline	AMBULANCE SERVICES	LONDON AMBULANCE SERVICE NHS TRUST	11663378	587,757.29	654967192	L0011214
Department of Health	NHS Bexley CCG	30/04/201	Hcare Srv Rec Fdtn Trust-Contract Baseline	PLANNED CARE	KING`S COLLEGE HOSPITAL NHS FOUNDATION TRUST	11669889	768,836.00		227489
Department of Health	NHS Bexley CCG	30/04/201	Hcare Srv Rec Fdtn Trust-Contract Baseline	PLANNED CARE	KING`S COLLEGE HOSPITAL NHS FOUNDATION TRUST	11669905	760,530.26		233541
Department of Health	NHS Bexley CCG	30/04/201	Cont Care- Adult 100% Fully Funded	FUNDED NURSING CARE	BUPA CARE HOMES	11669992	26,912.00	239731641	ALM038930TOALM038988
Department of Health	NHS Bexley CCG	30/04/201	Charges from CSU	RECHARGES	NHS SOUTH EAST CSU	11669996	85,259.00	GB654442045	7311608994
Department of Health	NHS Bexley CCG	30/04/201	Hcare Srv Rec Fdtn Trust-Contract Baseline	PLANNED CARE	KING`S COLLEGE HOSPITAL NHS FOUNDATION TRUST	11670000	279,925.00		228086
Department of Health	NHS Bexley CCG	30/04/201	Hcare Srv Rec Fdtn Trust-Contract Baseline	ACUTE COMMISSIONING	MOORFIELDS EYE HOSPITAL NHS FOUNDATION TRUST	11680115	82,568.00		IC0026903
Department of Health	NHS Bexley CCG	30/04/201	Hithcre-Foundation Trsts	COMMUNITY SERVICES	OXLEAS NHS FOUNDATION TRUST	11701702	228,989.00		37856
Department of Health	NHS Bexley CCG	30/04/201	Hithcre-Foundation Trsts	COMMUNITY SERVICES	OXLEAS NHS FOUNDATION TRUST	11701707	141,917.00		37832
Department of Health	NHS Bexley CCG	30/04/201	Hithcre-Foundation Trsts	COMMUNITY SERVICES	OXLEAS NHS FOUNDATION TRUST	11701710	807,726.00		37822
Department of Health	NHS Bexley CCG	30/04/201	Income tax <1Yr	BALANCE SHEET	HMRC	11720066	47,523.65	654462331	120PH00565770 MAR 16
Department of Health	NHS Bexley CCG	30/04/201	Income tax <1Yr-Student Loans	BALANCE SHEET	HMRC	11720066	338.00	654462331	120PH00565770 MAR 16
Department of Health	NHS Bexley CCG	30/04/201	National Insurance < 1 yr-NI- ERS	BALANCE SHEET	HMRC	11720066	29,114.84	654462331	120PH00565770 MAR 16
Department of Health	NHS Bexley CCG	30/04/201	National Insurance < 1 yr-NI-EES	BALANCE SHEET	HMRC	11720066	19,284.15	654462331	120PH00565770 MAR 16
Department of Health	NHS Bexley CCG	30/04/201	Hcare Srv Rec Fdtn Trust-Contract Baseline	PLANNED CARE	KING`S COLLEGE HOSPITAL NHS FOUNDATION TRUST	11722081	213,455.00		229026
Department of Health	NHS Bexley CCG	30/04/201	Hcare Srv Rec Fdtn Trust-Contract Baseline	ACUTE COMMISSIONING	KING`S COLLEGE HOSPITAL NHS FOUNDATION TRUST	11722088	265,141.00		229984
Department of Health	NHS Bexley CCG	30/04/201	Hcare Srv Rec Fdtn Trust-Contract Baseline	PLANNED CARE	KING`S COLLEGE HOSPITAL NHS FOUNDATION TRUST	11722091	273,009.00		230641
Department of Health	NHS Bexley CCG	30/04/201	Hcare Srv Rec Fdtn Trust-Contract Baseline	PLANNED CARE	KING`S COLLEGE HOSPITAL NHS FOUNDATION TRUST	11722099	256,424.00		231364
Department of Health	NHS Bexley CCG	30/04/201	Hcare Srv Rec Fdtn Trust-Contract Baseline	PLANNED CARE	KING`S COLLEGE HOSPITAL NHS FOUNDATION TRUST	11722114	224,799.00		232502
Department of Health	NHS Bexley CCG	30/04/201	Clinical&Medical-Othe Public Sector	REABLEMENT	BEXLEY COUNCIL	11722554	468,334.92		6890019276
Department of Health	NHS Bexley CCG	30/04/201	Hcare Srv Rec Fdtn Trust-Over/ Under Performance	ACUTE COMMISSIONING	QUEEN VICTORIA HOSPITAL NHS FOUNDATION TRUST	11738262	-26,066.24		1365
Department of Health	NHS Bexley CCG	30/04/201	Hcare Srv Rec Fdtn Trust-Over/ Under Performance	ACUTE COMMISSIONING	MOORFIELDS EYE HOSPITAL NHS FOUNDATION TRUST	11738298	30,540.00		IC0026446
Department of Health	NHS Bexley CCG	30/04/201	Hithcre-Foundation Trsts	COMMUNITY SERVICES	OXLEAS NHS FOUNDATION TRUST	11738402	15,605.00		37841
Department of Health	NHS Bexley CCG	30/04/201	Hcare Srv Rec Fdtn Trust-Contract Baseline	ACUTE COMMISSIONING	QUEEN VICTORIA HOSPITAL NHS FOUNDATION TRUST	11748386	47,900.00		2010008962
Department of Health	NHS Bexley CCG	30/04/201	Hcare Srv Rec NHS Trust-Contract Baseline	ACUTE COMMISSIONING	BARTS HEALTH NHS TRUST	11771729	84,072.08		254522
Department of Health	NHS Bexley CCG	30/04/201	Hcare Srv Rec NHS Trust-Contract Baseline	ACUTE COMMISSIONING	DARTFORD AND GRAVESHAM NHS TRUST	11771733	2,691,952.00		1000044308
Department of Health	NHS Bexley CCG	30/04/201	Hithcre-Foundation Trsts	MENTAL HEALTH CONTRACTS	SOUTH LONDON AND MAUDSLEY NHS FOUNDATION TRUST	11777527	25,202.00		B0007357
Department of Health	NHS Bexley CCG	30/04/201	Clinical&Medical-Voluntary Sector	COMMISSIONING - NON ACUTE	BEXLEY COUNCIL	11777582	34,361.83		6890018060
Department of Health	NHS Bexley CCG	30/04/201	Hithcre-Foundation Trsts	MENTAL HEALTH CONTRACTS	OXLEAS NHS FOUNDATION TRUST	11790994	1,749,240.00		37857
Department of Health	NHS Bexley CCG	30/04/201	Hithcre-Foundation Trsts	MENTAL HEALTH CONTRACTS	OXLEAS NHS FOUNDATION TRUST	11801176	50,000.00		37806
Department of Health	NHS Bexley CCG	30/04/201	Rent	ESTATES AND FACILITIES	NHS PROPERTY SERVICES LTD	11801410	51,009.75		48414351
Department of Health	NHS Bexley CCG	30/04/201	Hithcre-Foundation Trsts	CHILD AND ADOLESCENT MENTAL HEALTH	OXLEAS NHS FOUNDATION TRUST	11818020	255,781.00		37805
Department of Health	NHS Bexley CCG	30/04/201	Hithcre-Foundation Trsts	CHILD AND ADOLESCENT MENTAL HEALTH	OXLEAS NHS FOUNDATION TRUST	11818040	59,906.00		37804
Department of Health	NHS Bexley CCG	30/04/201	Clinical&Medical-Othe Public Sector	LEARNING DIFFICULTIES	BEXLEY COUNCIL	11818100	51,416.67		6890019263
Department of Health	NHS Bexley CCG	30/04/201	Clinical&Medical-Voluntary Sector	HOSPICES	GREENWICH & BEXLEY COMMUNITY HOSPICE	11818190	276,741.00		868
Department of Health	NHS Bexley CCG	30/04/201	Clinical&Medical-Commercial Sector	MATERNITY SERVICES	MARIE STOPES INTERNATIONAL	11835658	28,917.00	882426016	MS160477
Department of Health	NHS Bexley CCG	30/04/201	Hcare Srv Rec Fdtn Trust-Contract Baseline	ACUTE COMMISSIONING	KING`S COLLEGE HOSPITAL NHS FOUNDATION TRUST	11845602	1,869,631.00		234549
Department of Health	NHS Bexley CCG	30/04/201	Clinical&Medical-Othe Public Sector	LEARNING DIFFICULTIES	BEXLEY COUNCIL	11854667	250,000.00		6100071247
Department of Health	NHS Bexley CCG	30/04/201	Hcare Srv Rec Fdtn Trust-Contract Baseline	ACUTE COMMISSIONING	MOORFIELDS EYE HOSPITAL NHS FOUNDATION TRUST	11878126	82,568.00		IC0026968
Department of Health	NHS Bexley CCG	30/04/201	Clinical&Medical-Independent Sector	COMMUNITY SERVICES	BELLEGROVE ROAD SURGERY	11882991	65,392.39		BAC032016
Department of Health	NHS Bexley CCG	30/04/201	Healthcare Srv Rec CCG	NHS 111	NHS BROMLEY CCG	11923329	71,747.88		7022701149
Department of Health	NHS Bexley CCG	30/04/201	Clinical&Medical-Independent Sector	COMMUNITY SERVICES	CRAYFORD TOWN SURGERY	11923342	26,000.00		PHL01
Department of Health	NHS Bexley CCG	30/04/201	Hcare Srv Rec Fdtn Trust-Contract Baseline	PLANNED CARE	KING`S COLLEGE HOSPITAL NHS FOUNDATION TRUST	11923379	1,137,946.58		234497
Department of Health	NHS Bexley CCG	30/04/201	Hcare Srv Rec Fdtn Trust-Contract Baseline	ACUTE COMMISSIONING	GUYS & ST THOMAS HOSPITAL NHS FOUNDATION TRUST	11931340	1,455,983.00	654923417	3190559
Department of Health	NHS Bexley CCG	30/04/201	Hcare Srv Rec NHS Trust-Non Contract	PLANNED CARE	GUYS & ST THOMAS HOSPITAL NHS FOUNDATION TRUST	11935269	-3,129,822.00	654923417	9009585
Department of Health	NHS Bexley CCG	30/04/201	Hcare Srv Rec NHS Trust-Non Contract	PLANNED CARE	LEWISHAM & GREENWICH NHS TRUST	11935361	-128,548.00		8466655
Department of Health	NHS Bexley CCG	30/04/201	Hcare Srv Rec Fdtn Trust-Contract Baseline	PLANNED CARE	KING`S COLLEGE HOSPITAL NHS FOUNDATION TRUST	11951010	1,157,569.00		232736
Department of Health	NHS Bexley CCG	30/04/201	Hcare Srv Rec Fdtn Trust-Contract Baseline	PLANNED CARE	KING`S COLLEGE HOSPITAL NHS FOUNDATION TRUST	11951061	-388,405.89		CR00018929
Department of Health	NHS Bexley CCG	30/04/201	Hcare Srv Rec Fdtn Trust-Contract Baseline	PLANNED CARE	KING`S COLLEGE HOSPITAL NHS FOUNDATION TRUST	11951071	229,254.29		233540
Department of Health	NHS Bexley CCG	30/04/201	Hcare Srv Rec Fdtn Trust-Contract Baseline	PLANNED CARE	KING`S COLLEGE HOSPITAL NHS FOUNDATION TRUST	11951090	-473,000.00		CR00019017
Department of Health	NHS Bexley CCG	30/04/201	Cont Care- Adult 100% Fully Funded	CHC AD FULL FUND PERS HLTH BUD	BEXLEY COUNCIL	11951211	299,534.77		6100071865
Department of Health	NHS Bexley CCG	6	Hcare Srv Rec Fdtn Trust-Contract Baseline	ACUTE COMMISSIONING	UNIVERSITY COLLEGE LONDON HOSPITALS NHS FOUNDATION TRUST	11969298	68,256.00	654948687	5010089625

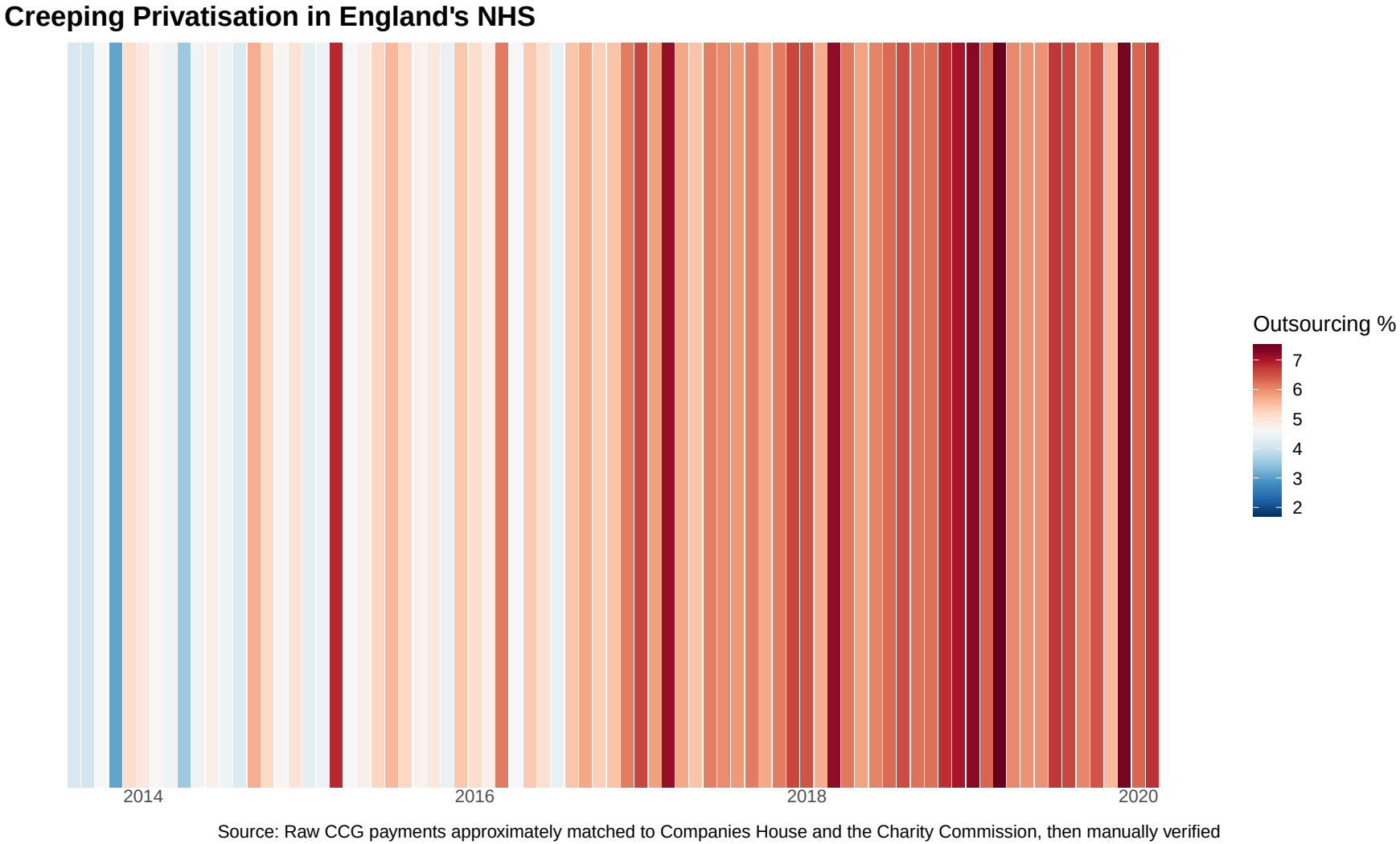


'Tools for Transparency in Central Government Spending'



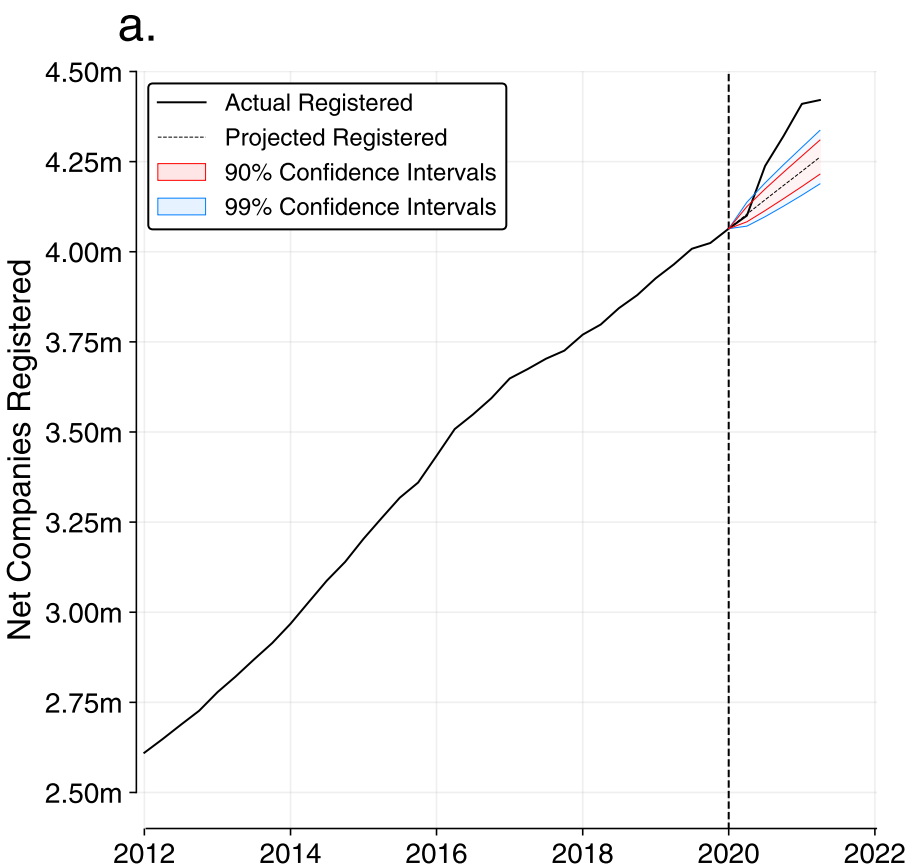
'The Role of Non-Profits in Public Health Service Provision'

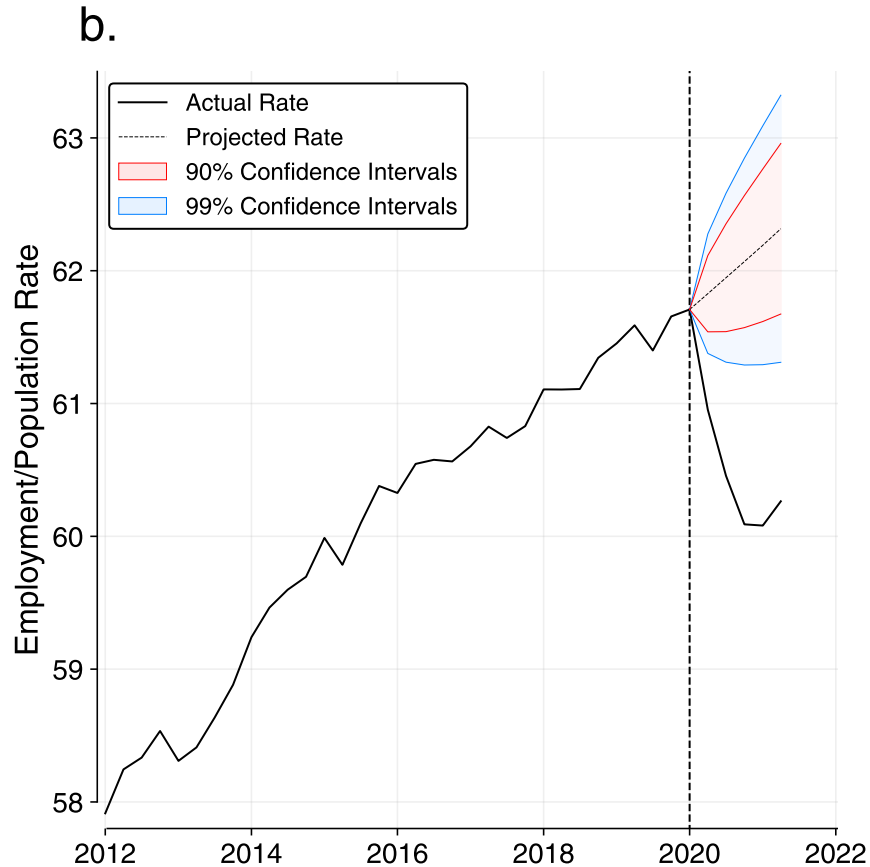
With John Mohan



'Creeping Commercialization in the NHS'

With Ben Goodair and Aaron Reeves



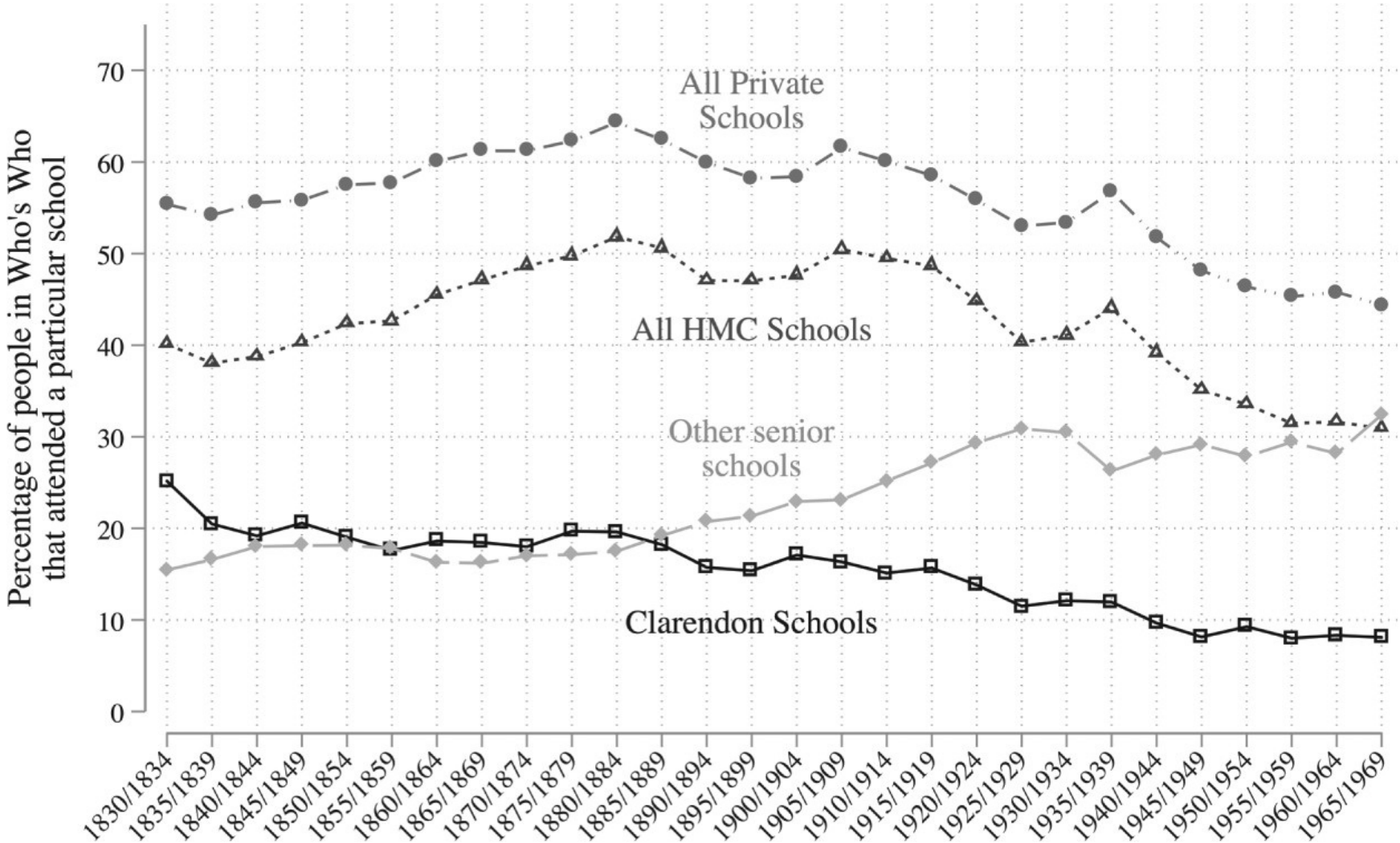


'Modelling Labour Market Inequalities and the Covid-19 Pandemic'

With Aaron Reeves and Naomi Muggleton

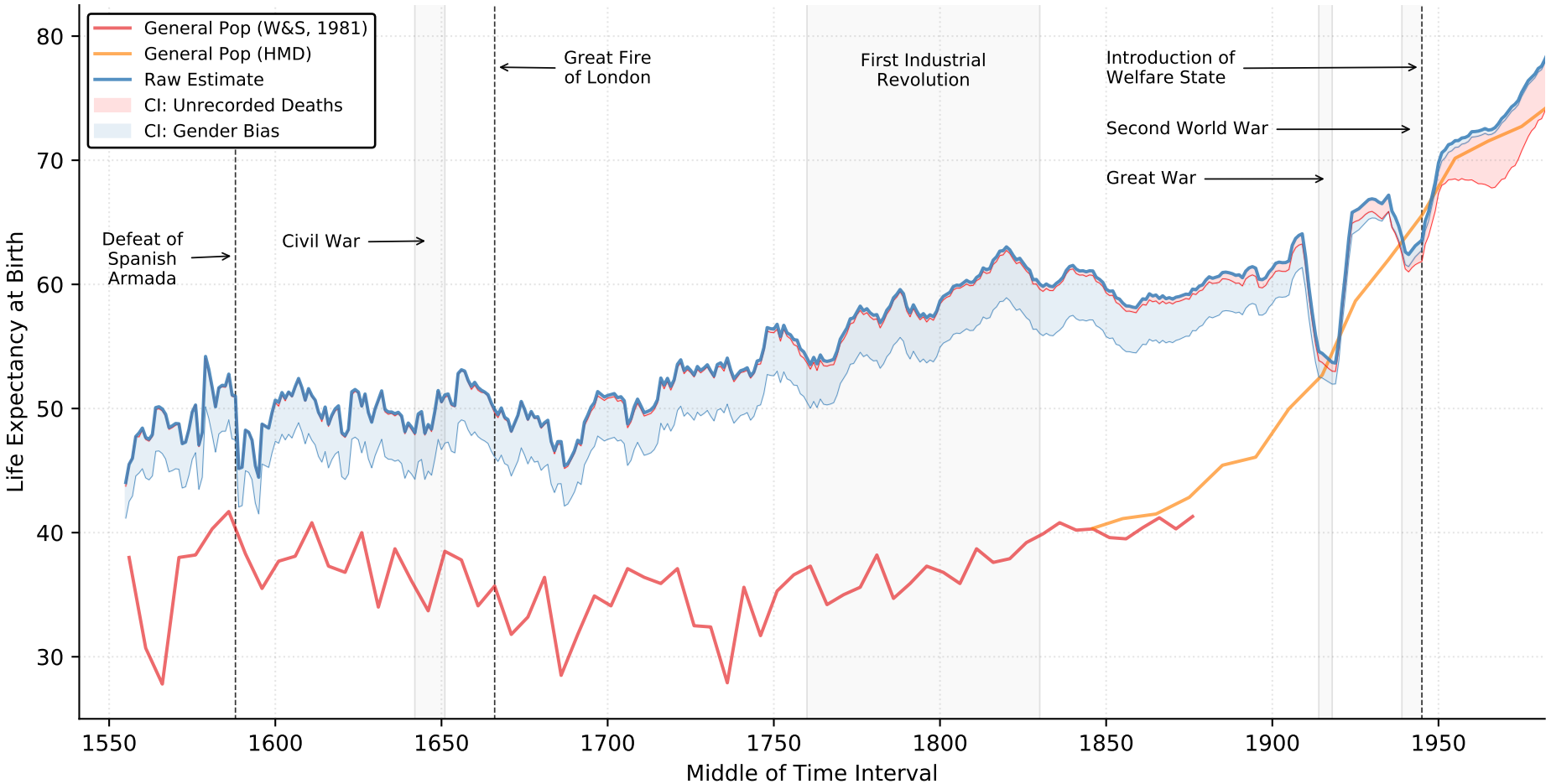
Semi-Structured Data for Inequality Research

- Data doesn't always come from APIs.
- Sometimes, it can come from raw `.html` itself.
- One example of this is when we scraped Who's Who to get information on education background of 'elites'.
- Another is when we scraped `thepeerage.com` to get data on dates of birth and death of British Peers.
- When doing this, we might choose to (there are competing views):
  - Check ToS and `robots.txt`.
  - Contact the webmaster first.

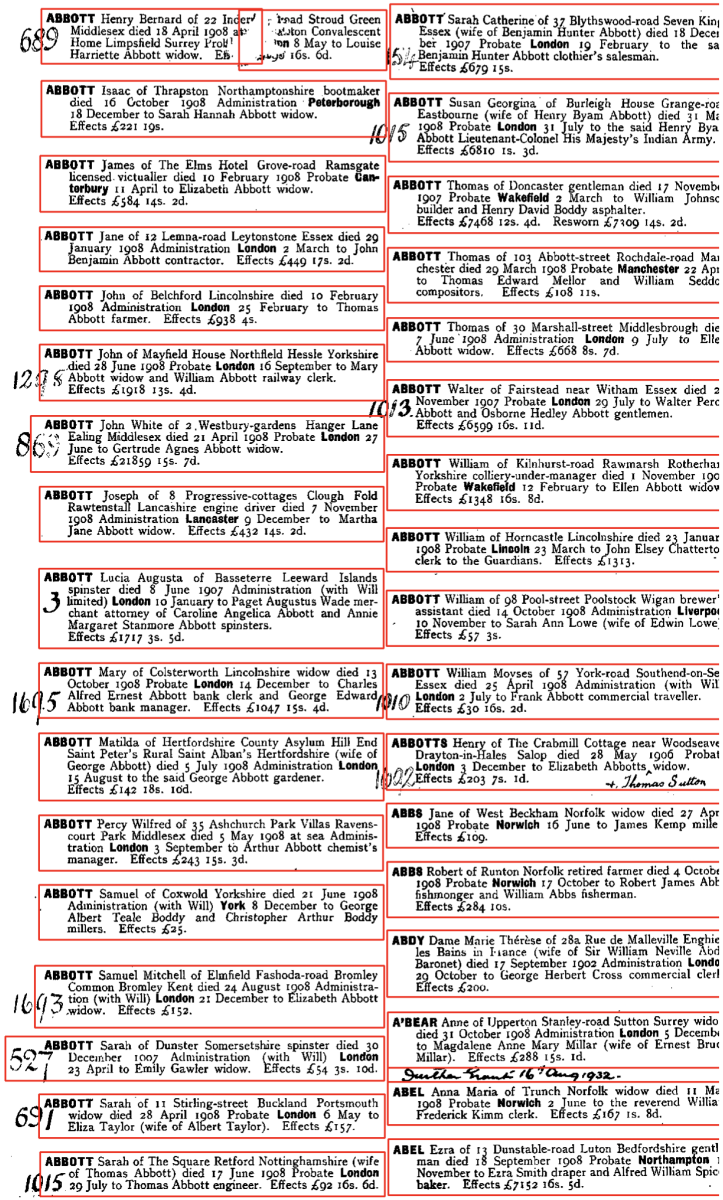


'The decline and persistence of the old boy'

With Aaron Reeves, Sam Friedman. and Magne Flemmen



'The Legacy of Longevity: Inequalities in UK Life Expectancy 1500-2016'
with Aaron Reeves, Felix Tropf and Darryl Lundy



ABBOTT Henry Bernard of 22 Inder[illegible] Middlesex died 18 April 1908 at[illegible] Home Limpsfield Surrey Prob[illegible] Harriette Abbott widow. Eff[illegible]

ABBOTT Isaac of Thrapston Northamptonshire bootmaker died 16 October 1908 Administration Peterborough 18 December to Sarah Hannah Abbott widow. Effects £221 19s.

ABBOTT James of The Elms Hotel Grove-road Ramsgate licensed victualler died 10 February 1908 Probate Canterbury 11 April to Elizabeth Abbott widow. Effects £584 14s. 2d.

ABBOTT Jane of 12 Lemna-road Leytonstone Essex died 29 January 1908 Administration London 2 March to John Benjamin Abbott contractor. Effects £449 17s. 2d.

ABBOTT John of Belchford Lincolnshire died 10 February 1908 Administration London 25 February to Thomas Abbott farmer. Effects £938 4s.

ABBOTT John of Mayfield House Northfield Hessle Yorkshire died 28 June 1908 Probate London 16 September to Mary Abbott widow and William Abbott railway clerk. Effects £1918 13s. 4d.

ABBOTT John White of 2 Westbury-gardens Hanger Lane Ealing Middlesex died 21 April 1908 Probate London 27 June to Gertrude Agnes Abbott widow. Effects £21859 15s. 7d.

ABBOTT Joseph of 8 Progressive-cottages Clough Fold Rawtenstall Lancashire engine driver died 7 November 1908 Administration Lancaster 9 December to Martha Jane Abbott widow. Effects £432 14s. 2d.

ABBOTT Lucia Augusta of Basseterre Leeward Islands spinster died 8 June 1907 Administration (with Will limited) London 10 January to Paget Augustus Wade merchant attorney of Caroline Angelica Abbott and Annie Margaret Stanmore Abbott spinsters. Effects £1717 3s. 5d.

ABBOTT Mary of Colsterworth Lincolnshire widow died 13 October 1908 Probate London 14 December to Charles Alfred Ernest Abbott bank clerk and George Edward Abbott bank manager. Effects £1047 15s. 4d.

ABBOTT Matilda of Hertfordshire County Asylum Hill End Saint Peter's Rural Saint Alban's Hertfordshire (wife of George Abbott) died 5 July 1908 Administration London 15 August to the said George Abbott gardener. Effects £142 18s. 10d.

ABBOTT Percy Wilfred of 35 Ashchurch Park Villas Ravenscourt Park Middlesex died 5 May 1908 at sea Administration London 3 September to Arthur Abbott chemist's manager. Effects £243 15s. 3d.

ABBOTT Samuel of Coxwold Yorkshire died 21 June 1908 Administration (with Will) York 8 December to George Albert Teale Boddy and Christopher Arthur Boddy millers. Effects £25.

ABBOTT Samuel Mitchell of Elmfield Fashoda-road Bromley Common Bromley Kent died 24 August 1908 Administration (with Will) London 21 December to Elizabeth Abbott widow. Effects £152.

ABBOTT Sarah of Dunster Somersetshire spinster died 30 December 1907 Administration (with Will) London 23 April to Emily Gawler widow. Effects £54 3s. 10d.

ABBOTT Sarah of 11 Stirling-street Buckland Portsmouth widow died 28 April 1908 Probate London 6 May to Eliza Taylor (wife of Albert Taylor). Effects £157.

ABBOTT Sarah of The Square Retford Nottinghamshire (wife of Thomas Abbott) died 17 June 1908 Probate London 29 July to Thomas Abbott engineer. Effects £92 16s. 6d.

ABBOTT Sarah Catherine of 37 Blythswood-road Seven King Essex (wife of Benjamin Hunter Abbott) died 18 December 1907 Probate London 19 February to the said Benjamin Hunter Abbott clothier's salesman. Effects £679 15s.

ABBOTT Susan Georgina of Burleigh House Grange-road Eastbourne (wife of Henry Byam Abbott) died 31 May 1908 Probate London 31 July to the said Henry Byam Abbott Lieutenant-Colonel His Majesty's Indian Army. Effects £6810 1s. 3d.

ABBOTT Thomas of Doncaster gentleman died 17 November 1907 Probate Wakefield 2 March to William Johnson builder and Henry David Boddy asphalter. Effects £7468 12s. 4d. Resworn £7109 14s. 2d.

ABBOTT Thomas of 103 Abbott-street Rochdale-road Manchester died 29 March 1908 Probate Manchester 22 April to Thomas Edward Mellor and William Seddon compositors. Effects £108 11s.

ABBOTT Thomas of 30 Marshall-street Middlesbrough died 7 June 1908 Administration London 9 July to Ellen Abbott widow. Effects £668 8s. 7d.

ABBOTT Walter of Fairstead near Witham Essex died 2 November 1907 Probate London 29 July to Walter Percy Abbott and Osborne Hedley Abbott gentlemen. Effects £6599 16s. 11d.

ABBOTT William of Kilnhurst-road Rawmarsh Rotherham Yorkshire colliery-under-manager died 1 November 1907 Probate Wakefield 12 February to Ellen Abbott widow Effects £1348 16s. 8d.

ABBOTT William of Horncastle Lincolnshire died 23 January 1908 Probate Lincoln 23 March to John Elsey Chatterton clerk to the Guardians. Effects £1313.

ABBOTT William of 98 Pool-street Poolstock Wigan brewer's assistant died 14 October 1908 Administration Liverpool 10 November to Sarah Ann Lowe (wife of Edwin Lowe) Effects £57 3s.

ABBOTT William Movses of 57 York-road Southend-on-Sea Essex died 25 April 1908 Administration (with Will) London 2 July to Frank Abbott commercial traveller. Effects £30 16s. 2d.

ABBOTTS Henry of The Crabmill Cottage near Woodseaves Drayton-in-Hales Salop died 28 May 1906 Probate London 3 December to Elizabeth Abbotts widow. Effects £203 7s. 1d.

ABBS Jane of West Beckham Norfolk widow died 27 April 1908 Probate Norwich 16 June to James Kemp miller Effects £109.

ABBS Robert of Runton Norfolk retired farmer died 4 October 1908 Probate Norwich 17 October to Robert James Abbs fishmonger and William Abbs fisherman. Effects £284 10s.

ABDY Dame Marie Thérèse of 28a Rue de Malleville Enghien les Bains in France (wife of Sir William Neville Abdy Baronet) died 17 September 1902 Administration London 29 October to George Herbert Cross commercial clerk Effects £200.

A'BEAR Anne of Upperton Stanley-road Sutton Surrey widow died 31 October 1908 Administration London 5 December to Magdalene Anne Mary Millar (wife of Ernest Bruce Millar). Effects £288 15s. 1d.

ABEL Anna Maria of Trunch Norfolk widow died 11 May 1908 Probate Norwich 2 June to the reverend William Frederick Kimm clerk. Effects £167 1s. 8d.

ABEL Ezra of 13 Dunstable-road Luton Bedfordshire gentleman died 18 September 1908 Probate Northampton 1 November to Ezra Smith draper and Alfred William Spicer baker. Effects £7152 16s. 5d.

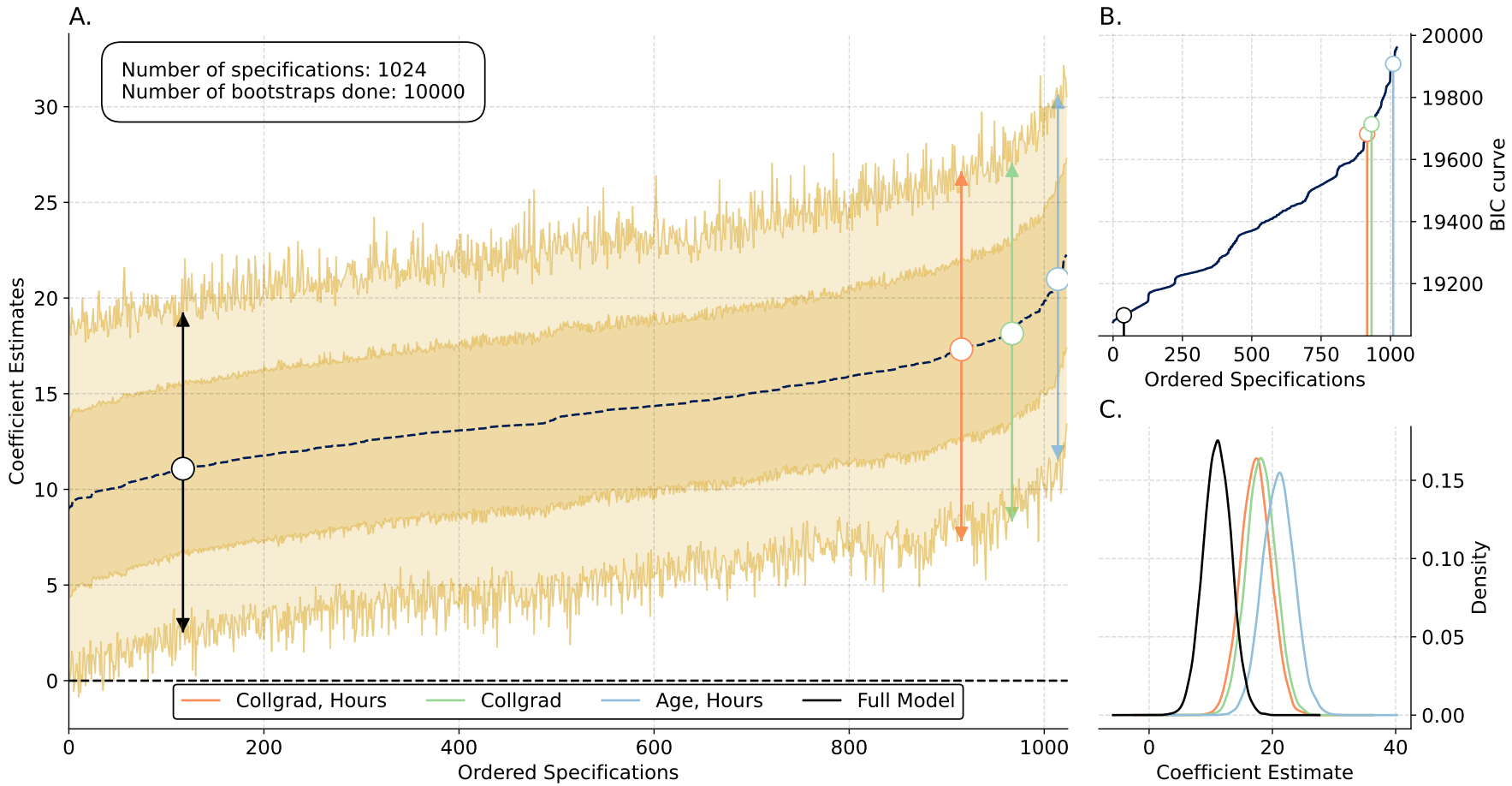
Ongoing research (With Aaron Reeves and Naomi Muggleton) uses and probate records from 1858-1996.

The idea is to:

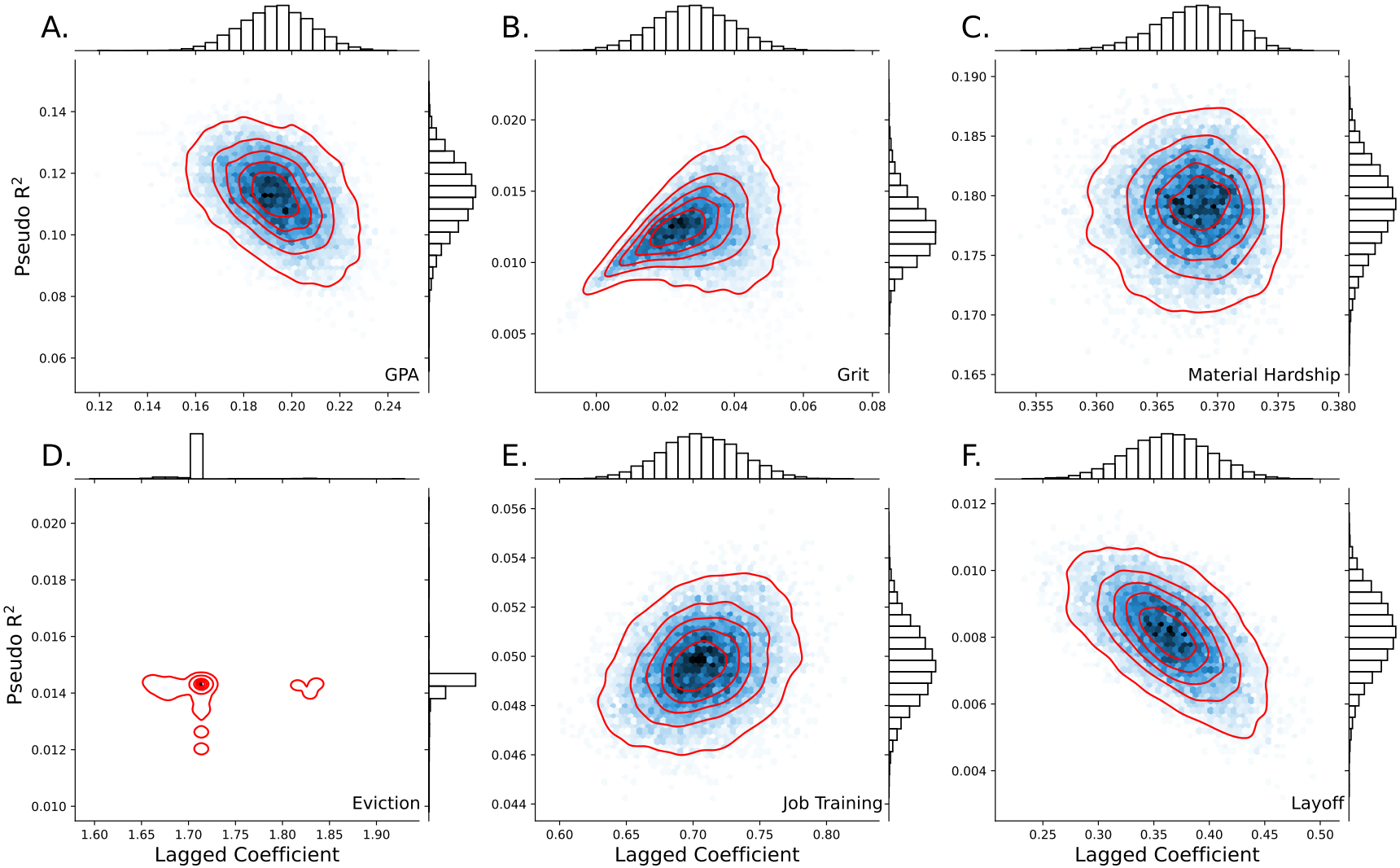
1. Scrape 800k .pdfs containing 24m probates: Done!
2. Preprocess images ready for OCR: Done!
3. Run OCR on HPC in parallel: Done!
4. Build lifetables to calculate e_0 by wealth at death...: Pending!

- There are close links between 'Open Science' and traditional survey based modeling paradigms!
- By opening research, we can examine the certainty of statistical results.
- Does β_1 really equal 7?
- How can we "take the 'con' out of econometrics"?
- Transparency and slightly more computational procedures!
- Current research culture around robustness only mandates examination of a couple of choices.
    - With current computational power; why not all combinations?

Our work to tackle specific types of uncertainty still in progress!

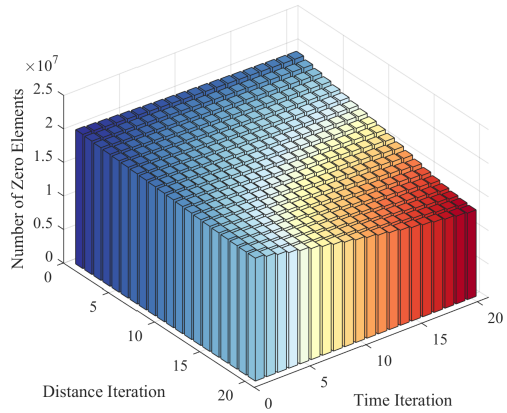


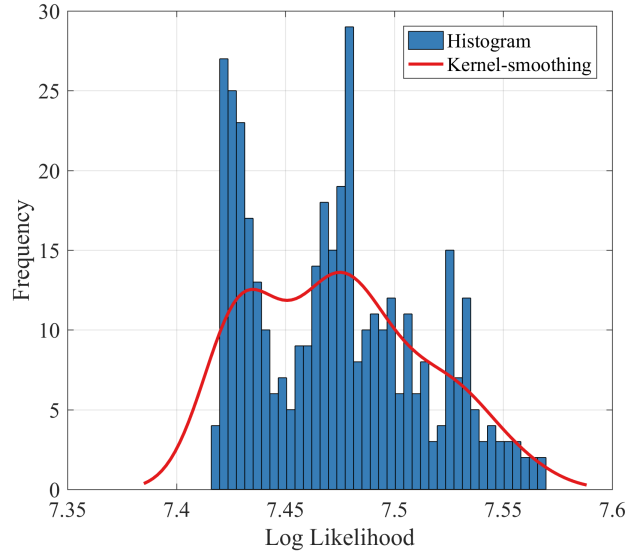
'Introducing NRobust: A multiversal library with bootstraps, selection, and model averaging' with Daniel Valdenegro-Ibarra

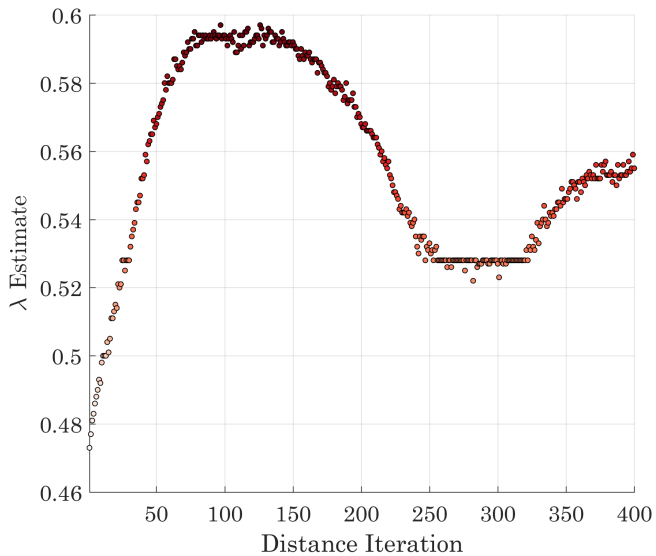


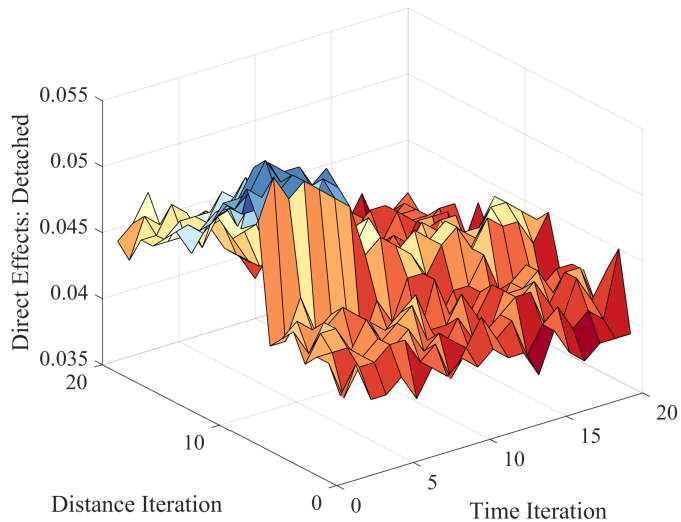
'From a Seed of Doubt Grows a Forest of Uncertainty'

With Arun Frey, Jiani Yan and Mark Verhagen



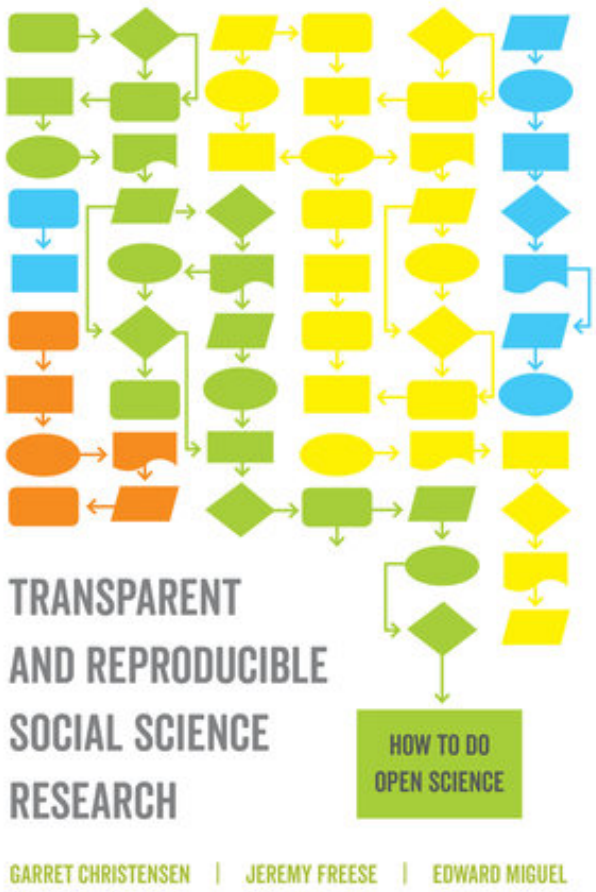






'A Grid Based Approach to Analysing Spatial Weighting Matrix Specification'

Just in case you're looking for something further:



This book is most definitely worth reading!

And not just because one of the authors is a coauthor of mine!

It does an **excellent** job of motivating Open Social Science.

Thank you!

Don't hesitate to get in touch at:

charles.rahal@sociology.ox.ac.uk

and

github.com/crahal!