



# **EXAMINING GENDERED PLACES & SPACES IN UK PHD STUDY USING DATA SCIENCE & QUANTITATIVE GEOGRAPHY**

**Laura Sheppard**

**Jon Reades**

**Richard Freeman**

Centre for Advanced Spatial Analysis, UCL

15/11/24

UKCGE conference

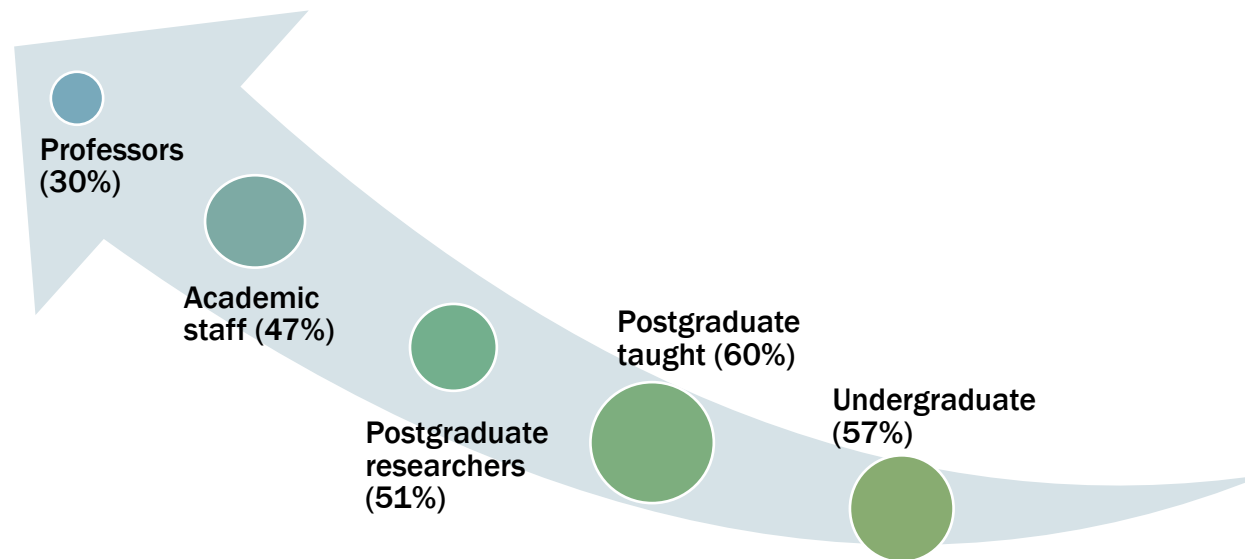
# THE PROBLEM

**Women and gender minorities are underrepresented in PhD study and academia, especially in positions of leadership and seniority and in particular disciplines and higher education institutions**

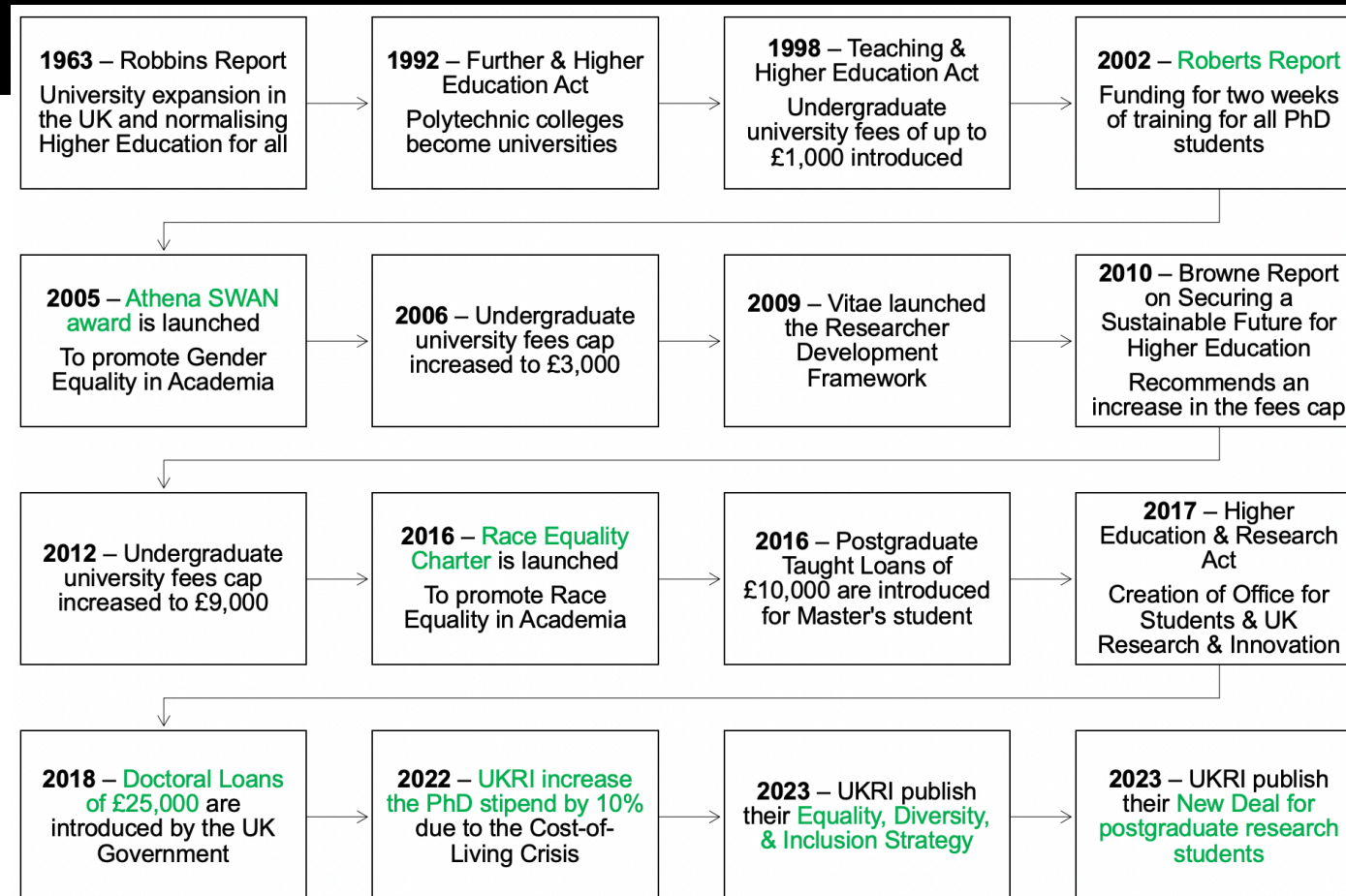
**There is a doctoral data gap to explore this:**

- UKRI – the data is only for 20-25% of PhD studentships**
- HESA – the data is not openly available at the individual level and does not have data before 1994/5**

# THE 'LEAKY' PIPELINE

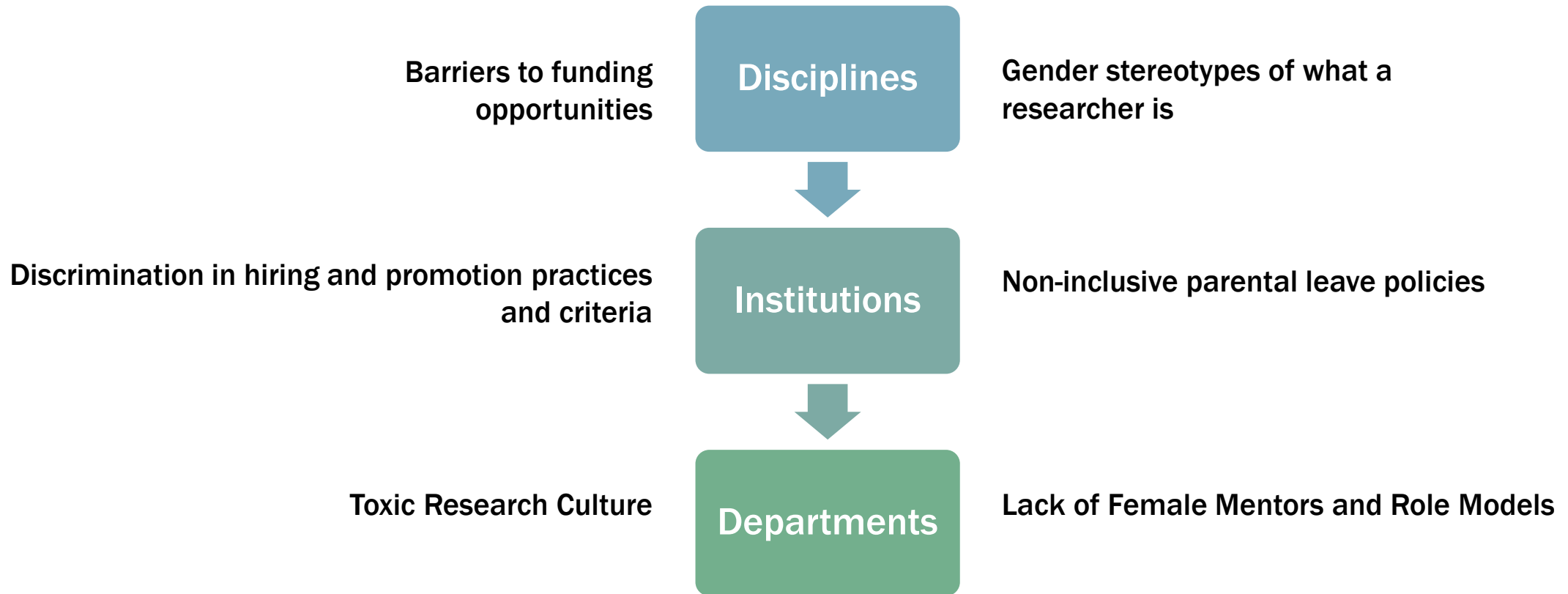


# HIGHER EDUCATION POLICY



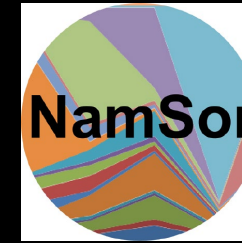
# AIM

To examine and understand the gendered patterns and dynamics of PhD students at departmental, institutional and disciplinary scales, across the UK and historically, using the British Library's EThOS metadata.



# ETHOS DATA

- As of October 2022, ~610,500 doctoral records
- 98% of UK doctoral students are included in EThOS – a population, rather than a sample
- Variables include:
  - Author name
  - Title
  - Year of award
  - Qualification type (e.g. PhD, DClInPys, EdD)
  - Institution
  - Discipline (based on assigned Dewey Decimal Classification)
  - Department (sometimes)
  - Supervisor/s (rarely)
  - Funder/s (rarely)



# GENDER INFERENCE ALGORITHMS

- No personal characteristics of PhD students included in EThOS
- Needed a tool to create a gender variable from data available in EThOS e.g. names
- Used gender inferencing services:
  - o Free: Gender R, Gender Guesser
  - o Paid: Genderize, Gender API, Namsor
- Issues of algorithmic biases
  - o Reinforcing the gender binary
  - o Misgendering people
  - o Inferring personal data from public data
  - o Biases of incorrect inferencing for non-Western names

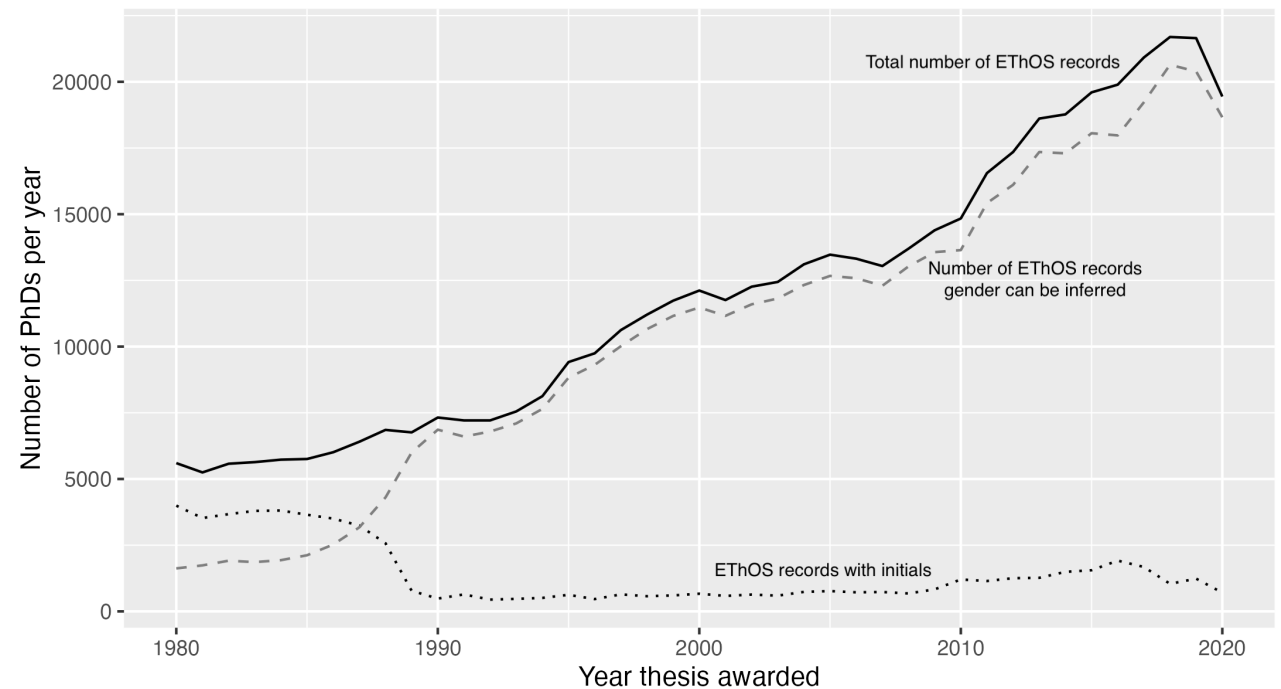
# MISSING NAME DATA

From 1980 onwards:

- 512,000 PhD records
- 60,000 PhD records with initials as first name
- Issues with initials in UCL, Salford, and Swansea

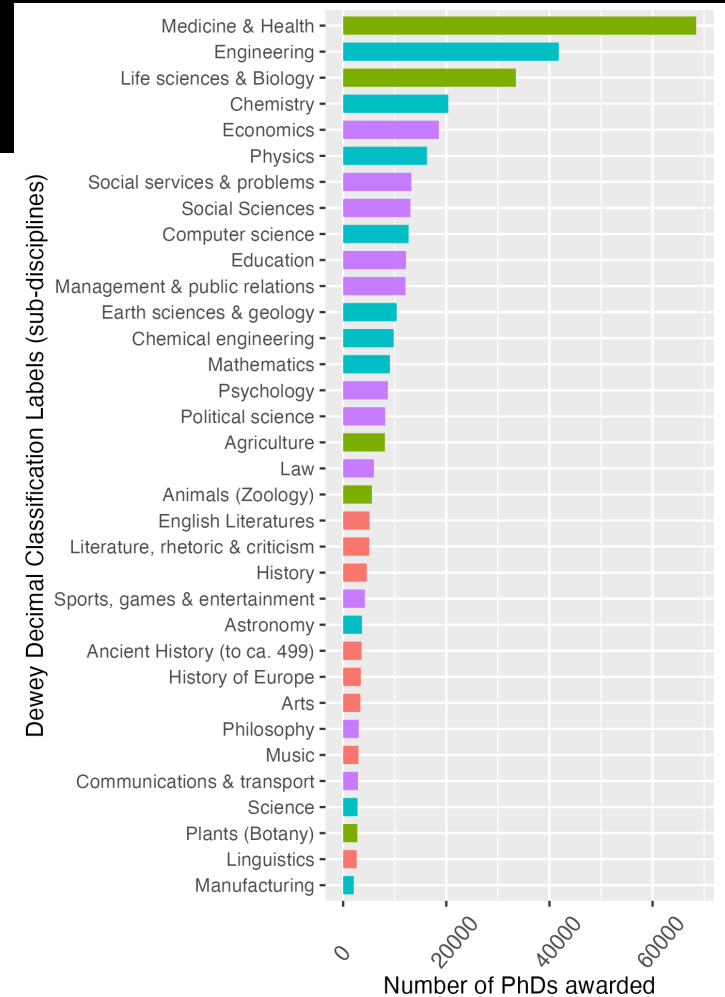
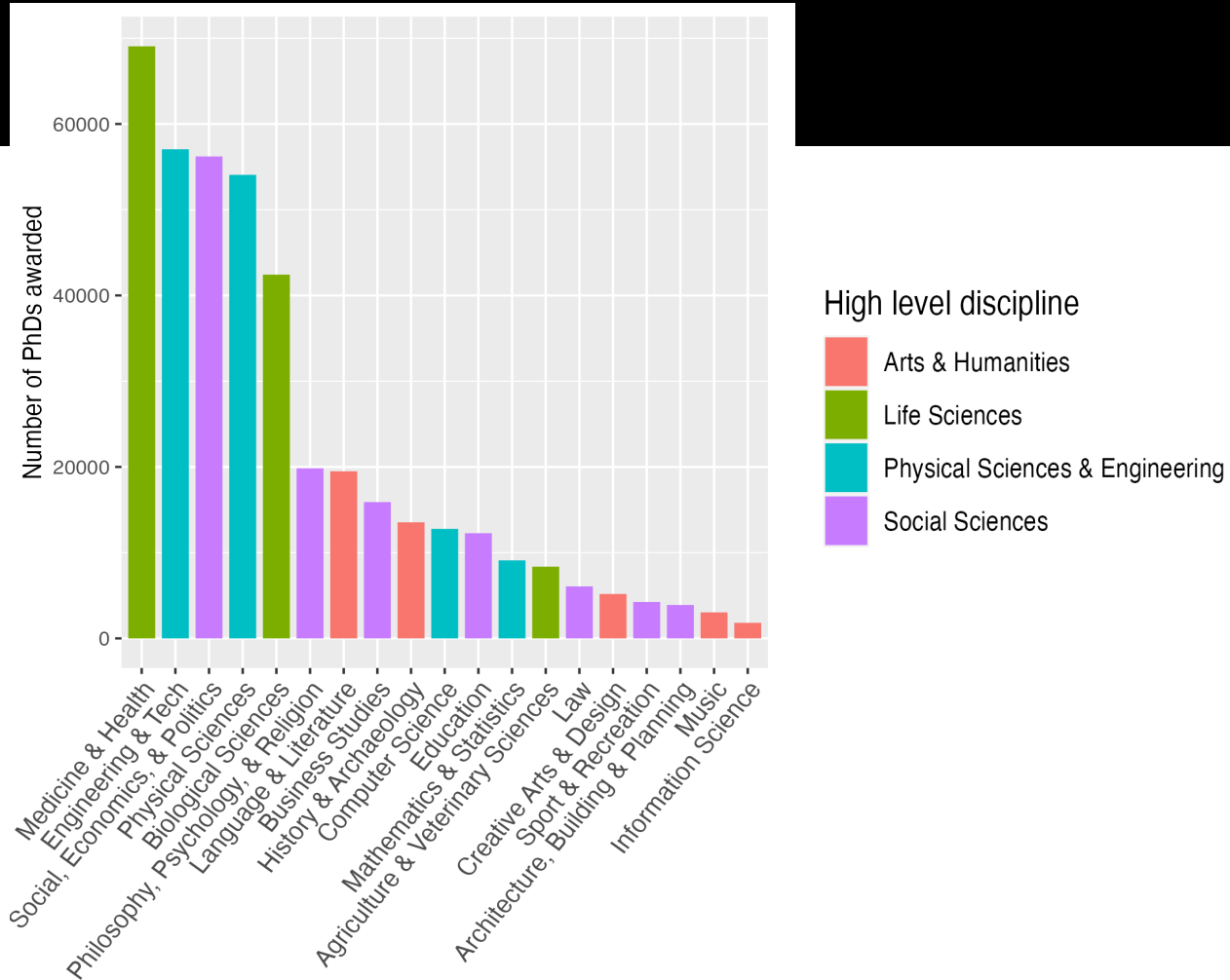
Therefore, we used data from 1990 onwards

Number of students and records with initials as names in EThOS  
126 Higher Education Institutions (1980-2020)



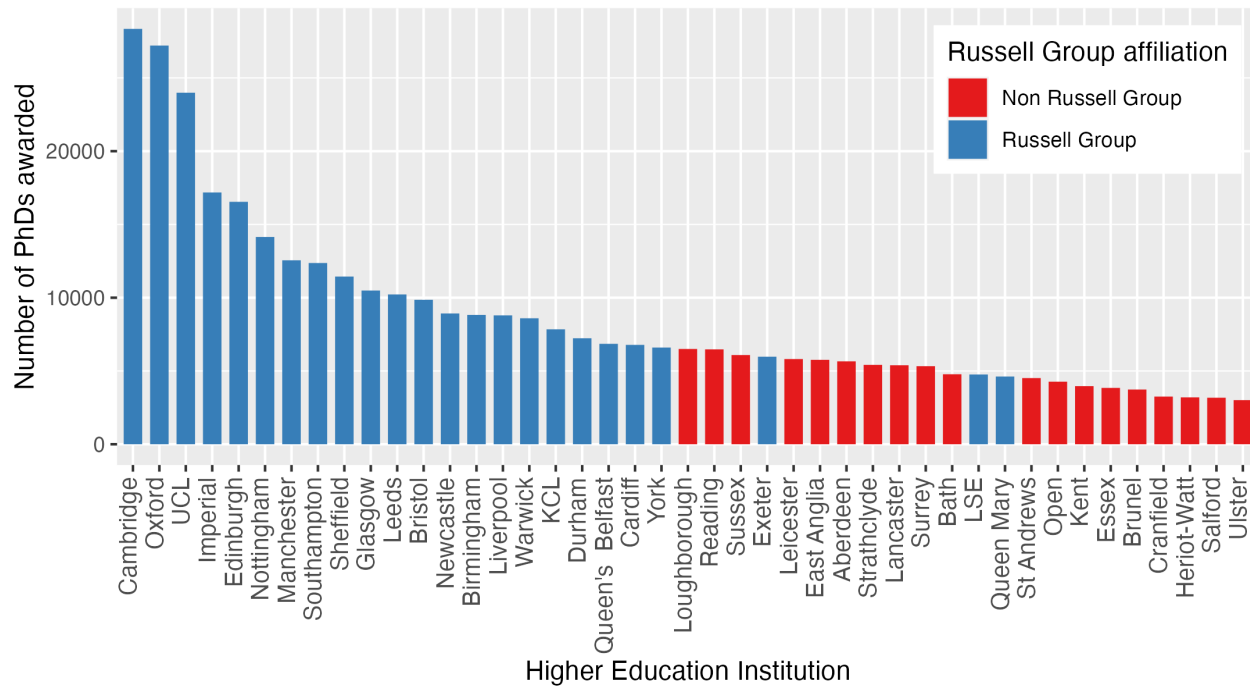


# DISCIPLINARY PATTERNS

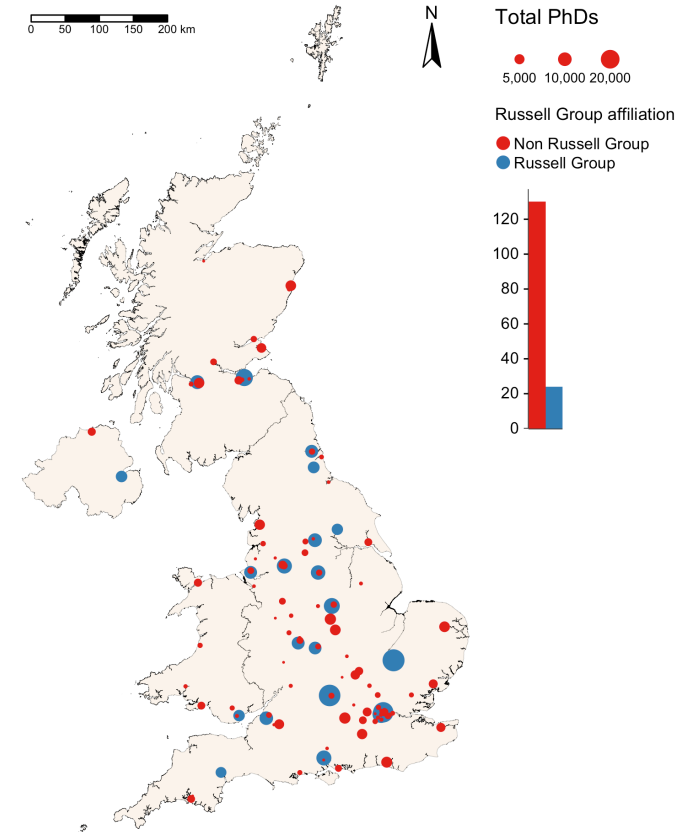


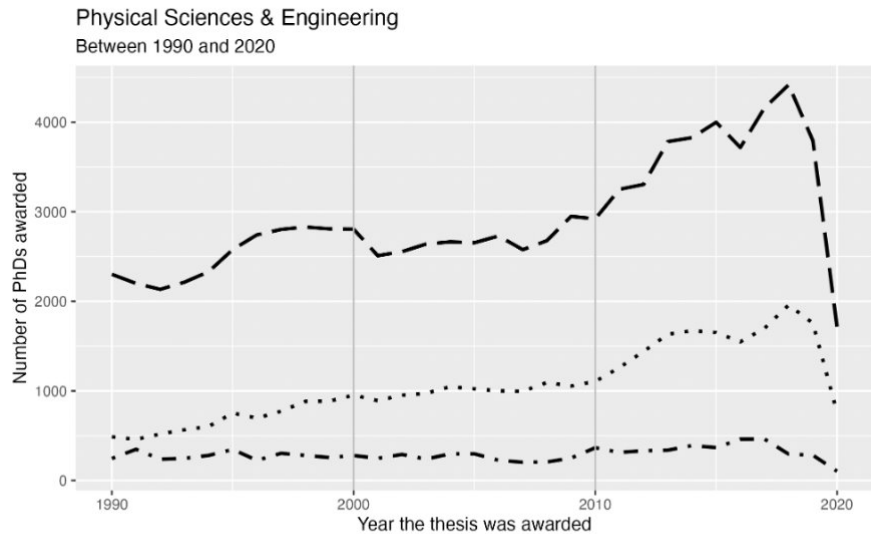
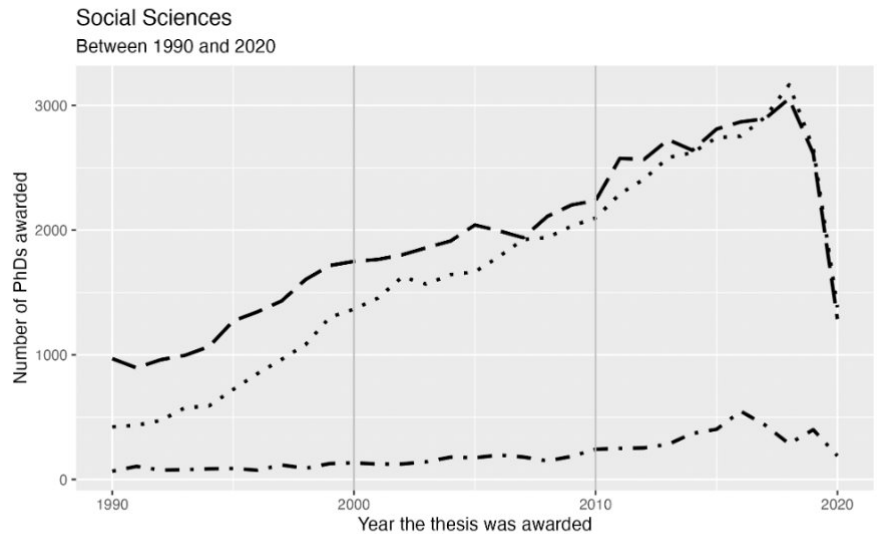
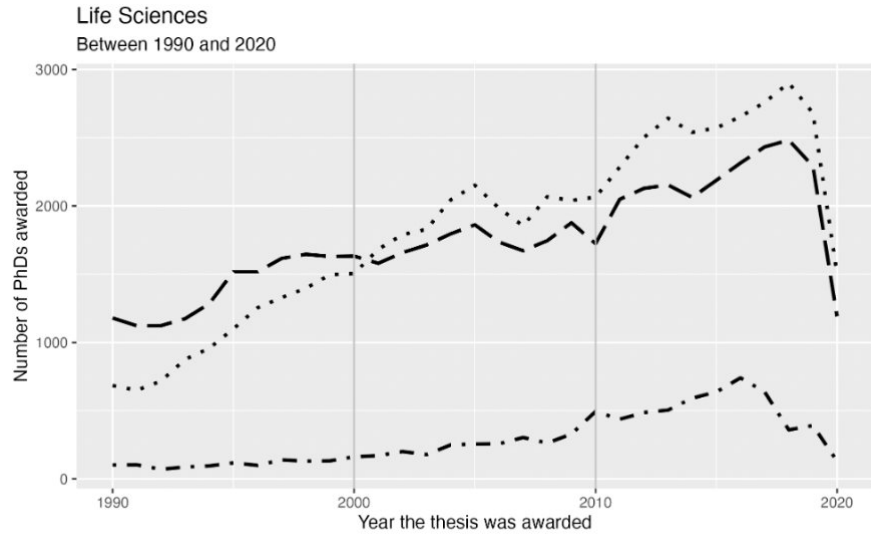
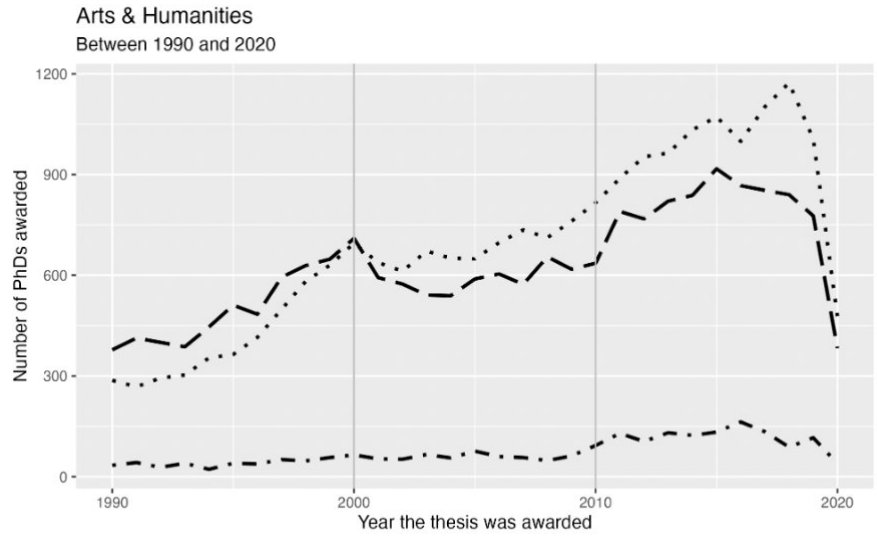
# INSTITUTIONAL PATTERNS

PhDs awarded by UK universities spilt by Russell Group affiliation  
Over 3,000 PhDs between 1990 and 2022



Total PhDs awarded by UK universities





••• female    — male    - - - unknown

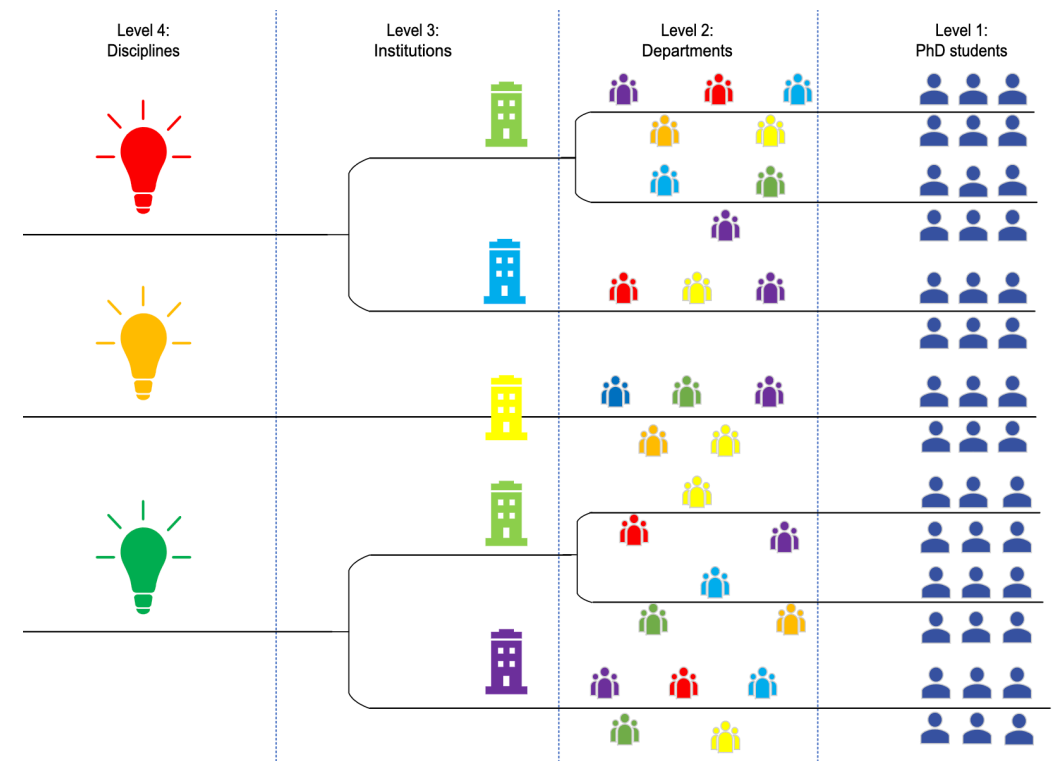
# MULTILEVEL MODELS

**Multilevel Models** are often used in health and education research and split the residual variance between different scales/levels within the hierarchical structure of the data.

In this example with EThOS, the hierarchical structure is:

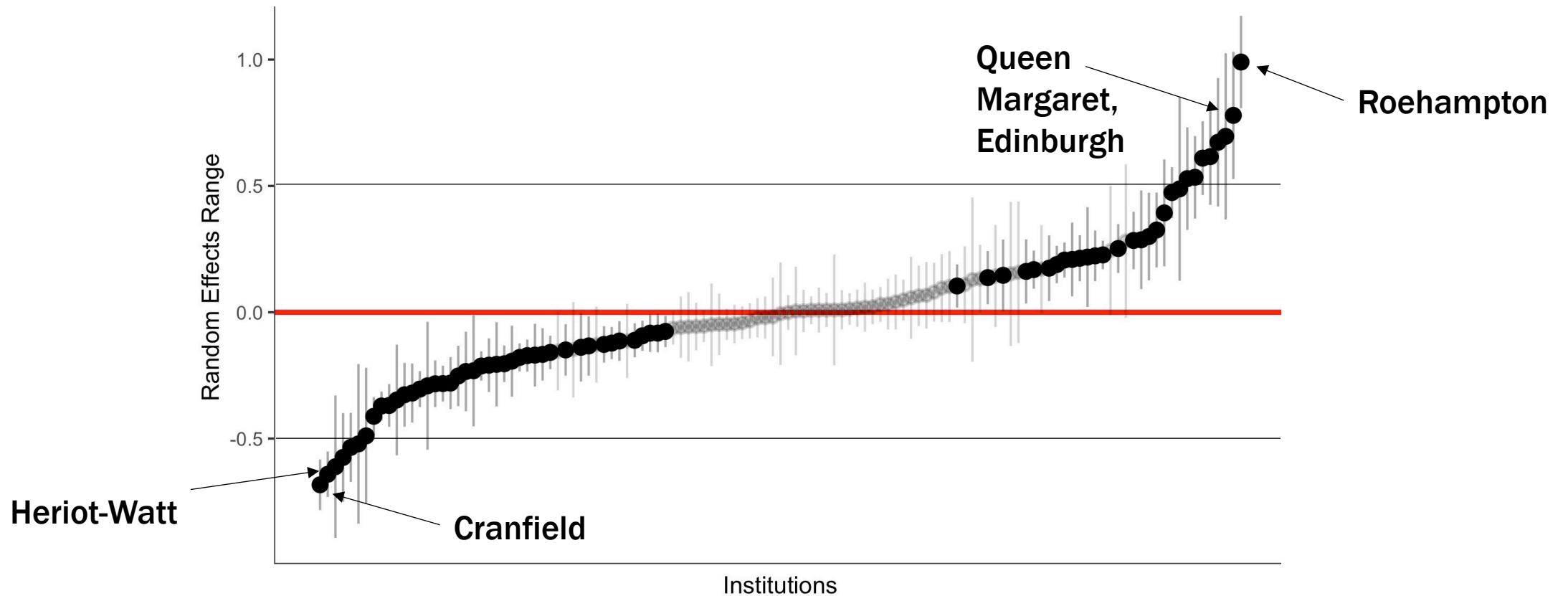
- **PhD students** *nested in*
- **Departments** *nested in*
- **Institutions** *nested in*
- **Disciplines.**

This allows us to understand **where** the variation in the probability a PhD student is female and which levels of the data account for what amount of variation.



# UK UNIVERSITIES

Number of observations: 420,246, Institutions: 121, % variance at level 2: 2.6%

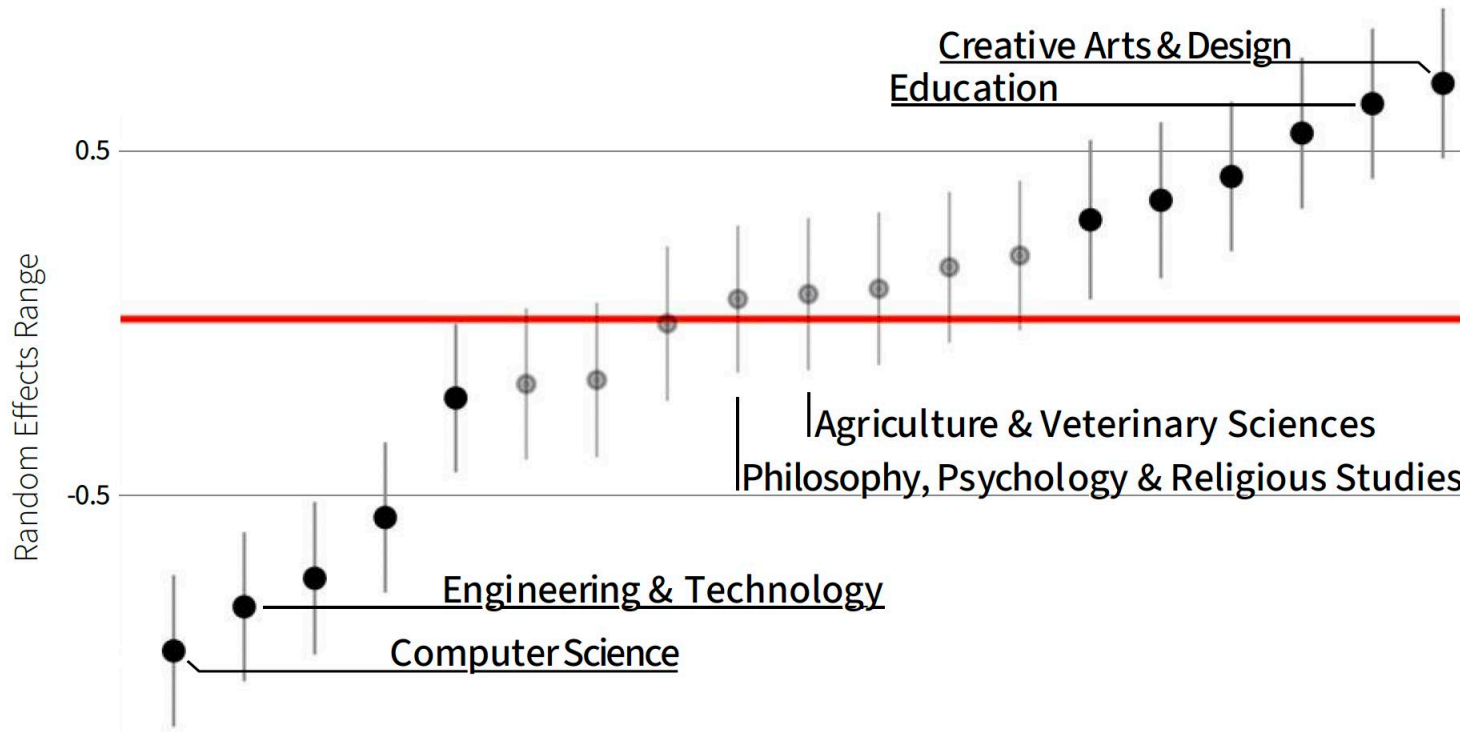


# DISCIPLINES

Number of observations: 384,199, Disciplines:, 19, % variance at level 2: 6.5%

## Male-skewed:

- Computer Science
- Engineering & Technology
- Mathematics & Statistics



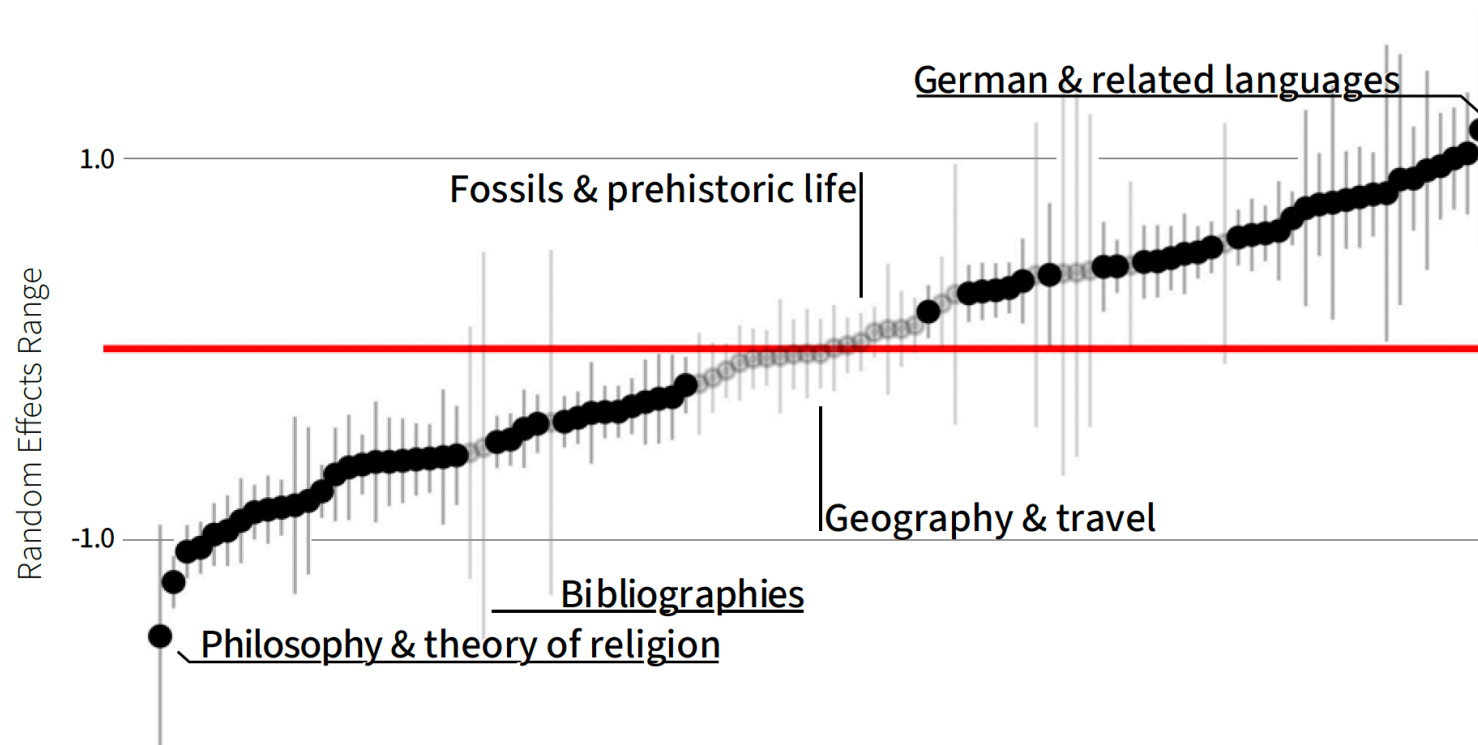
## Female-skewed:

- Creative Arts & Design
- Education
- Literature & Language

# SUB-DISCIPLINES

Number of observations: 380,697, Sub-disciplines:, 99, % variance at level 2: 10.7%

- Male-skewed:**
- Philosophy & Theory of Religion
  - Physics
  - Computer Science



- Female-skewed:**
- German language
  - Italian & Romanian literatures
  - Customs, etiquette, and folklore

# DEPARTMENTS

Number of observations: 377,386, Sub-disciplines:, 6,775, % variance at level 2: 9.9%

Missing department data in EThOS, therefore, have made proxy-departments from Institution + Sub-discipline.

## Female skewed

Sub discipline & university	n	Variance
Leicester: Arts	34	1.443
Oxford: Italian, Romanian & related literatures	32	1.388
Southampton: Psychology	319	1.341
City, University of London: Psychology	150	1.335
Goldsmith's: Education	76	1.314

## Male skewed

Sub discipline & university	n	Variance
Bath: Manufacturing	31	-1.416
Nottingham: Manufacturing	98	-1.389
Cambridge: Computer Science	487	-1.386
Birmingham: Physics	198	-1.360
Brighton: Engineering	99	-1.338



# FUTURE RESEARCH

- Applying for ESRC grant funding to extend the work into unequal outcomes by ethnicity and intersectionality (gender and ethnicity)
- Examine the potential impact of the Covid-19 pandemic on PhD completions (when more data is released from the BL, post cyberattack in October 2023)
- Write a paper with a focus on STEM PhDs and analyse gendered topics over space and time

# EXAMINING GENDERED PLACES & SPACES IN UK PHD STUDY USING DATA SCIENCE & QUANTITATIVE GEOGRAPHY

## Publications:

Sheppard, L., Reades, J. and Freeman, R. P. J. (2023). Gendering and Diversifying the Research Pipeline: A Quantitative Feminist Geographical Approach to Gender in Higher Education, *Annals of the American Association of Geographers*, 113(4), pp. 817–833. doi: 10.1080/24694452.2022.2160302.

Sheppard, L., Reades, J. and Freeman, R. P. J. (Forthcoming). Examining the 'gendered' places and spaces of UK doctoral education using multilevel modelling. *Transactions of the Institute of British Geographers*. (Minor revisions accepted)

Sheppard, L., and Reades, J. (Forthcoming). Finding the female geographers: The gendered dynamics of UK geography PhD study. *The Professional Geographer*. (Minor revisions accepted)

15/11/24

Dr Laura Sheppard

@laurahsheppard

UKCGE conference

