

Gendering the research pipeline: A quantitative feminist geographical approach

Laura Sheppard* and Jon Reades

Centre for Advanced Spatial Analysis, University College London

GISRUK 2023

Summary

Women and gender minorities are underrepresented in positions of seniority in higher education (HE). Research on gender in HE varies in scale and methodology from large global surveys to smaller projects, with a noticeable gap of the experiences of ECRs and doctoral students, and how they vary from discipline-to-discipline, institution-to-institution, and department-to-department. We propose a new agenda by drawing together quantitative and feminist geography to focus on different 'platial' scales within HE, such as expanding the definition of gender and considering intersectionality, using open-source data, and using quantitative geography methods to develop a multi-scalar understanding of these dynamics.

KEYWORDS: Gender, Higher Education, Space & Place, Quantitative Geography, Feminist Geography

1. Introduction

We know that it is unsurprising that women, gender minorities, people of colour, people with long-term health conditions, and people from working class backgrounds are underrepresented throughout UK academia, especially in positions of seniority, such as as professors and Vice Chancellors. These axes of oppression and power are important to consider, and very much exist in UK HE, however, in this piece of research, we will solely focus on gender within academia, particularly in terms of PhD students.

Despite 57% of undergraduate students and 60% of postgraduate taught students being female in the UK, it has not translated to higher up the academic hierarchy with data from the Higher Education Statistics Agency (HESA) for 2019/20 showing that 46.6% of academic staff are female, and 0.16% identified as other (HESA, 2021). Of the 22,745 full professors, just 6,340 (27.9%) identified as female (HESA, 2021). In 2020/21, slightly more first-year postgraduate research students identified as female (50.6%) than those who identified as male (48.9%), and the number of enrolled females has been gradually increasing over the past five years (HESA, 2022).

The HESA data paints a picture of people of different genders working and studying in UK academia and this research aims to bring together quantitative and feminist geography to establish the differences in the 'platial' scales within academia and how the probability that someone of a certain gender varies across these scales. Using British Library data, we can build a detailed, and broad, picture of PhD students by gender across the UK and over time using a series of multilevel models.

2. Data

This research uses the British Library's (BL) E-Theses Online Service (EThOS), which is an electronic bibliographic meta database (data that describes and gives information about other data) for UK theses and is openly available to download from the BL's website (EThOS, 2021). EThOS contains approximately 98% of the doctoral theses awarded by UK HEIs and gives a near full record of

* Laura.sheppard.20@ucl.ac.uk

knowledge created through doctoral research (Gould 2016). It is published every six months, with the most recent version being October 2022.

In the most recent version of EThOS, there are around 610,00 UK PhD theses, from 1768 to 2022, and contains information on the author's name, the thesis title, the university, year of award, discipline, sometimes the department, and occasionally the supervisor. There are issues of missing data within the author's name and department columns as some names are recorded as initials, instead of full first names, and many EThOS records do not contain any information on the department that the thesis was studied in.

Prior to the modelling, the EThOS data was cleaned, and due to issues of missing data, EThOS was subsetted for PhDs from 1990 onwards to examine the gendered patterns of different scales and places within recent UK PhD study.

3. Methods

As gender is not included in EThOS, gender needed to be inferred using the student's names, an issue that can create ethical problems. A range of Gender Inferencing Algorithms were tested and validated, and it was concluded that 'Namsor' was the most suitable for the data, although it only infers a binary gender of male or female. Namsor provides a cultural context using surnames to inferring gender, unlike the other algorithms tested (Santamaría & Mihaljević, 2018). Namsor was used to infer the gender of PhD students in the post-1990 EThOS sample.

Multilevel models (MLMs) were used in this research and are used widely in quantitative geography (Goldstein, 2011; Snijders & Bosker, 2011). They are used when the data is hierarchical in nature, which is true for the EThOS data with PhD students nested in departments nested in universities and disciplines. MLMs can help explain if there is something about particular places that impacts the gender of the PhD students in HE that cannot be explained through simply assessing the individuals in isolation of the structures they exist within.

As Namsor only gives a binary response, MLMs for discrete dependent variables were used (Guo & Zhao, 2000). The dependent variable is coded as either 1 for a female PhD student and 0 for a male PhD student to show the presence of the variable (Powers, 2012). For this research a logit model was used due to the interpretation of the model output (how different variables and scales raise or lower the log odds of $Y = 1$) (Powers, 2012). Log likelihood tests, AIC scores, and variance partition coefficients will be used to assess the goodness of fit of the models and how much variation can be attributed to each level of the data, which helps us to examine the importance of particular places.

4. Results

4.1 Exploratory Data Analysis

From 1990 to 2022, the highest numbers of PhDs were awarded in the disciplines of: Medicine & Health, Engineering & Technology, and Social, Economic, & Political Studies (**Figure 1a**). The HE institutions who have awarded the highest numbers of PhDs were the universities of; Cambridge, Oxford, University College London, Imperial College London, and Edinburgh (**Figure 1b**).

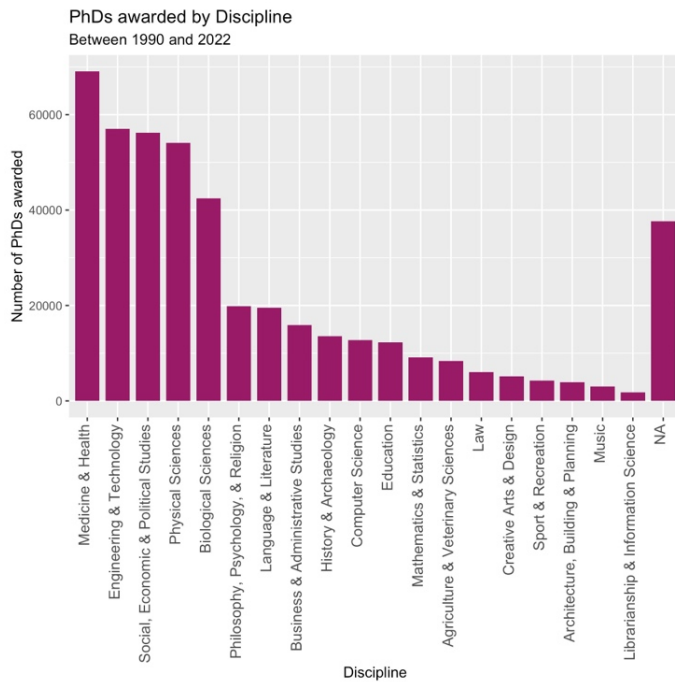


Figure 1a PhDs awarded by disciplines

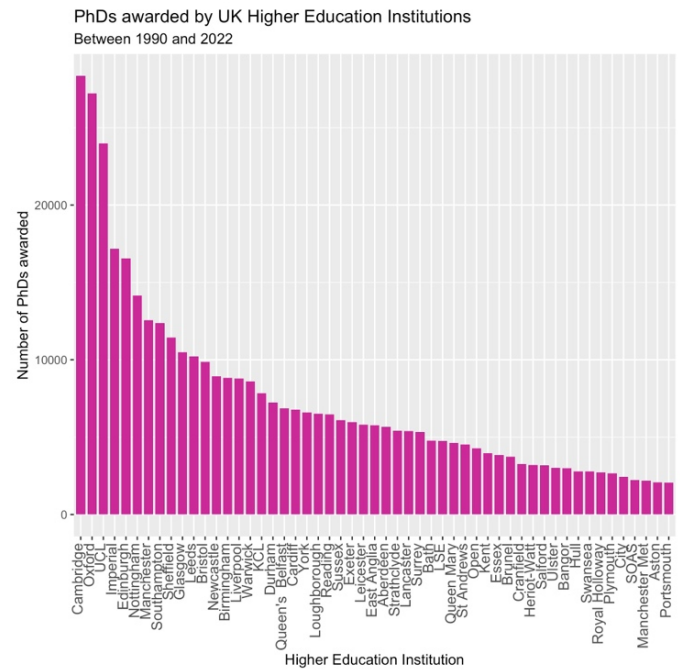


Figure 1b PhDs awarded by universities

In no year between 1990 and 2020 have the number of female PhD students overtaken the number male PhD students (**Figure 2a**). In terms of gender and disciplines, for the four main disciplines of Arts & Humanities, Life Sciences, Physical Sciences & Engineering, and Social Sciences, they vary in their numbers of female PhD students across time, as shown in **Figure 2b**. The Physical Sciences & Engineering remain well below gender parity and reached 29% female of their PhDs by 2020, compared to the Life Sciences, Arts, and Humanities which are over the 50% line for gender parity in 2020.

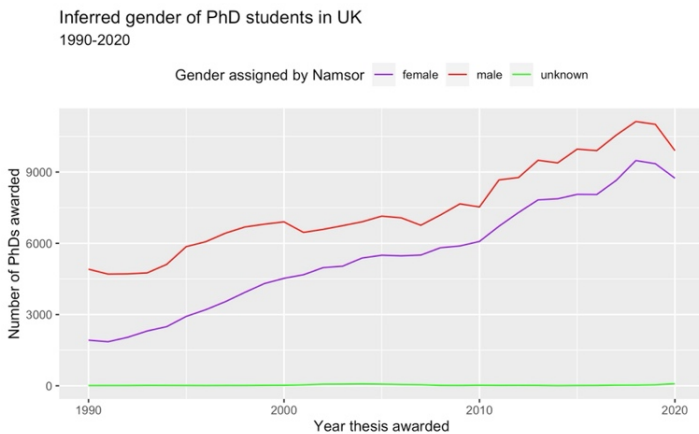


Figure 2a Gender of PhD students

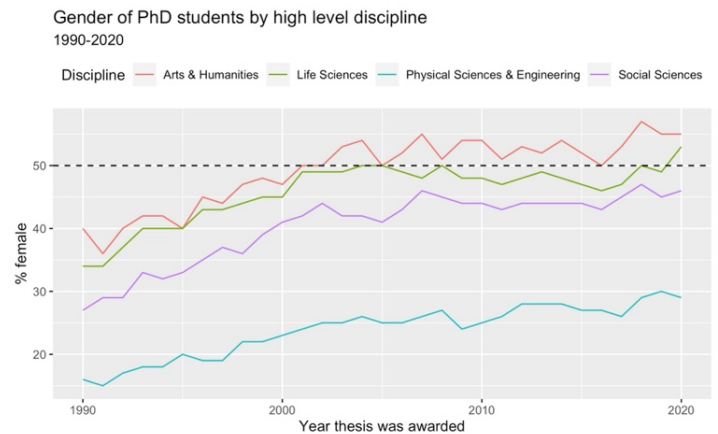


Figure 2b Gender of PhD students by disciplines

4.2 Multilevel Modelling

A series of models have been run including a single level logistic regression model, two level variance components models with no predictor variables, and random intercepts models with predictor variables. A few of the key models will be highlighted and discussed during the talk.

For some of the two-level variance components models run, the level two scales accounted for different amounts of variation, such as 6.5% for disciplines (n=19), 2.6% for universities (n=121), 1.8% for years (n=33), and 10.7% for sub-disciplines (n=99). All models had PhD students at level one. The model with sub-disciplines accounted for more variation in the gender variable than any other level two variables.

The disciplines of Computer Science, Engineering & Technology, and Mathematics & Statistics are skewed towards male PhDs. Whereas the disciplines of Arts & Design, Education, and Literature & Language are skewed towards female PhDs. These disciplines remain consistently skewed towards either female or male PhDs across the three decades within the sample.

In a three-level random intercepts model with institutions at level two and sub-disciplines at level one and various predictor variables, 13.8% of the variation in the gender variable was accounted for (0.12% by institutions, 8.1% by sub-disciplines). The log % female PhDs had a positive coefficient, whereas the Physical Sciences and Wales dummies both had negative coefficients. This indicates that as the log % female PhDs increases, as does the probability of the student being female, whereas if the PhD is in the Physical Sciences/Engineering and/or was completed at a Welsh university, then the probability of the PhD student being female decreases.

5. Conclusions

Even though this research is ongoing, some initial conclusions can be drawn. MLMs are suitable models to examine the importance of place in gender and academia as in all the MLMs run the log likelihood ratio test result that suggests that there are group differences in the likelihood that a PhD student is female. When explanatory variables are added into the MLMs and the intercepts are allowed to vary, more variation in the gender variable is accounted for and the explanatory power of the model increases.

6. Acknowledgements

This research is part of Laura's PhD on gender in UK HE. The work is funded by ESRC (part of UKRI) through the UBEL DTP, and by the British Library as industry partner.

References

HESA (2021). All staff (excluding atypical) by equality characteristics 2014/15 to 2019/20. URL: <https://www.hesa.ac.uk/data-and-analysis/staff/table-27>

HESA (2022). Higher Education Student Statistics: UK, 2020/21 - Student numbers and characteristics. URL: <https://www.hesa.ac.uk/news/25-01-2022/sb262-higher-education-student-statistics/numbers>

Harvey Goldstein (2011). *Multilevel Statistical Models*. Wiley and Sons Ltd, 4th ed.

Sara Gould (2016). UK theses and the British Library EThOS service: from supply on demand to repository linking. *Interlending and Document Supply*, 44(1)

G. Guo and H. Zhao (2000). Multilevel modeling for binary data. *Annual Review of Sociology*, 26: 441–462.

D. A. Powers (2012). Multilevel models for binary data. *New Directions for Institutional Research*, p. 57–75.

Santamaría L and Mihaljević H (2018) Comparison and benchmark of name-to- gender inference services. *PeerJ Computer Science*, doi:10.7717/peerj-cs.156.

T. A. B. Snijders and R. J. Bosker (2011). *Multilevel Analysis: An Introduction to Basic and Advanced*

Multilevel Modeling. SAGE Publications Ltd., 2nd ed.

British Library, 2021. EThOS: E-thesis online service. URL: <https://ethos.bl.uk/>

Biographies

Laura Sheppard is a PhD student at the Centre for Advanced Spatial Analysis at University College London. Her research interests include using quantitative human geography methods and data science to examine social and spatial inequalities.

Dr Jon Reades is an Associate Professor of Spatial Data Science at the Centre for Advanced Spatial Analysis at University College London. His interests include: Economic Geography; Housing and Neighbourhoods; Open Data & Teaching, and Reproducible Research; and Emerging Quantitative Methods.