School of Environmental Sciences: Annual Progress Proforma for Year 1 PGR students

Name: Peter Prescott
Project Title: The Geodemographics of British Streets
Supervisors: Alex Singleton

What will your project achieve?

This PhD will develop a method that creates a public geodemographic classification; that in addition to utilising the best of open data; couple these with other data sources and generate more timely and accurate measures of populations and their contexts. The project will develop innovative techniques that utilise street geometry in their specification, estimation and testing; as the site for social interaction and the construction of neighbourhoods.

State why this PhD project is important.

Historically, humankind is in the early stages of an unprecedented *data revolution* (Kitchin, 2014) – and geographic/spatial data in particular pose unique challenges (Lee and Kang, 2015) *Economically*, a recent Cabinet briefing suggested use of geospatial data could unleash £10 billion (Cabinet Office, 2018). But *critically*, advances in data analytics can have negative consequences too (O'Neil, 2016), and so it is imperative that research be open, and reproducible (Peng, 2011).

State how the knowledge gained from the project will be put to use by other researchers or the non-academic community (industry, charities etc).

Geodemographic classifications are of use in every sector which deals with real people in particular places. In business it is useful for marketing; in politics, campaigning; in government, for evaluating policy; in churches and charities, for understanding how to see and serve the needs of people in particular places.

The underlying methodology that I develop will be applicable to other areas of spatial analysis than merely geodemographics itself, and will be used by statisticians, social scientists, and population geographers in all the various fields of work which consider human life in real world places.

The code that I write and the research that I do will be shared openly, and will be used and improved by those addressing similar problems.

Give three or four aims for the proposed research. Keep each of these research aims to one sentence.

- To develop a new methodology for zone design in small area statistics, based on the natural unit of *the street*, rather than ad hoc administrative areal divisions.
- To bring together a comprehensive array of available open data sources to generate new insight.
- To enable a better understanding of the nature and needs of local communities.

Matching each of the research aims (listed in 4) outline the objectives to this research project: what they will do and what they will achieve.

1. Use Street Network Graph (as available through Ordnance Survey's Master Map or Open Street Map) to partition the UK on the basis of street layout.
2. Create a new open geodemographic data resource for the whole United Kingdom, using machine learning to cluster street-level neighbourhoods into statistically similar groupings.
3. Use web technology to facilitate individualized interaction with current data.

State the three or four most important published manuscripts you have read this year and why you felt they were important to your PhD topic.

- Webber and Burrows (2018) give a historical overview of the development of geodemographics, while engaging with various critiques of geodemographics, and demonstrating its application to understanding contemporary social patterns.
- O'Sullivan and Unwin (2014) offer a foundational account of first principles that is both accessible and rigorous.
- Law (2017) includes a thorough account of the need for a well-defined Street-Based Local Area.
- Boeing (2017) is an excellent example of how to use Python to create software to do street network analysis.

Outline the progress you have made, with reference to the aims and objectives above, over the past year.

- Agreed thesis outline and progress plan with supervisor
- Taken seven MSc modules (receiving average mark of 74/100 for four completed modules)
- Produced multiple pieces of relevant work:
  - Gave a talk on the founding fathers of *Geographic Information Science*: Longley, Maguire, Goodchild & Rhind. (Prescott, 2019h)
  - Gave a presentation on the development and application of Agent-Based Modelling in the social sciences. (Prescott, 2019f)
  - Wrote software to create simulation of agents interacting in a spatial field. (Prescott, 2019d)
  - Created app to visualize research-paper citations as a network graph. (Prescott, 2019e)
  - Wrote a paper examining the QLFS 2012 and considering the question of secularization in Britain, and the relationship between religiosity, age, and sex. (Prescott, 2019b)
  - Wrote a paper analyzing the 2011 Census data to assess which features of small areas were the best predictors of home ownership. (Prescott, 2019a)
  - Gave presentation reviewing and critiquing Kim Knotts' spatial analysis of religious-secular relations. (Prescott, 2019g)
  - Wrote paper reviewing the qualitative method of urban walking, as practised within the framework of *psychogeography*, in critical dialogue with my own Evangelical epistemology, ethics, and experience. (Prescott, 2019c)
  - Combined multiple open datasets to create a geodemographic classification of the political constituencies of Britain. (Prescott, 2020a)
  - Created implementations in Python of the Perceptron Algorithm (Prescott, 2020c) and the K-Means Clustering algorithm (Prescott, 2020b).
- Learnt how to use the necessary tools to facilitate efficient and reproducible research writing and computation:
  - Markdown for undistracted plain-text writing,
  - Vim for efficient text-editing,
  - Zotero for citation tracking,
  - Git for version control (backup management),
  - Pandoc and Latex for publication-quality documents,
  - Docker for reproducible computational environments

Outline your plans for the next year, with reference to the aims and objectives above. Do you foresee any challenges?

- Complete draft literature review, focussing particularly on:
  1. The Genealogy of Geodemographics, Social Area Analysis and Urban Data Science.
  2. Building a Geodemographic Classification: Machine-Learning Clustering Methods.

3. The Possiblities and Problems of Zone Design in Small Area Statistics
4. The Theory of Neighbourhood Morphology and Community Ontology.

- Become familiar with available sources of Open Data.
- Partner in project developing a geodemographic classification of London.
- More MSc. Modules:
  - Social Analytics & Visualization
  - Database and Information Systems
  - Big Data Analysis
- Submit a first paper to a journal for review.

Please list the development workshops you have completed over the past year or so, delivered by LDC or by other means, e.g. through the DTP/CDT or external providers. What plans do you have for development workshops over the next year?

I took part in the PGR Induction, and the Library Tour.

I have no specific plans for future workshops – will consider opportunities as they come.

References

Boeing, G. (2017). *OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks*. Computers, Environment and Urban Systems. 65:pp. 126–139.

Cabinet Office (2018). *An Initial Analysis of the Potential Geospatial Economic Opportunity*. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/733864/Initial_Analysis_of_the_Potential_Geospatial_Economic_Opportunity.pdf.

Kitchin, R. (2014). *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. Sage.

Law, S. (2017). *Defining Street-based Local Area and measuring its effect on house price using a hedonic price approach: The case study of Metropolitan London*. Cities. 60:pp. 166–179.

Lee, J.G. and Kang, M. (2015). *Geospatial Big Data: Challenges and Opportunities*. Big Data Research. 2(2):pp. 74–81. ISSN 2214-5796. doi:10.1016/j.bdr.2015.01.003.

O'Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Penguin Random House USA Ex. ISBN 978-0-553-41881-1.

O'Sullivan, D. and Unwin, D. (2014). *Geographic Information Analysis*. John Wiley & Sons.

Peng, R.D. (2011). *Reproducible research in computational science*. Science. 334(6060):pp. 1226–1227.

Prescott, P. (2019a). *Examining Predictors for Home Owner-Occupancy*. https://geodemographics.co.uk/projects/regression-modelling/.

——— (2019b). *Examining the Association between Age and Religiosity in 21st Century Britain*. https://geodemographics.co.uk/projects/analytics-in-r/.

——— (2019c). *'Pedestrian Activist Exploration': Walking as Qualitative Method*. https://geodemographics.co.uk/projects/qualitative-methods/.

——— (2019d). *Project: Agent-Based Modelling*. https://geodemographics.co.uk/projects/agent-based-modelling/.

——— (2019e). *Project: Bibliographic Citation Graph*. https://geodemographics.co.uk/projects/citations/.

——— (2019f). *Talk: Agent-Based Modelling*. https://geodemographics.co.uk/projects/gds-abm/.

——— (2019g). *Talk: Critiquing a Qualitative Research Paper*. https://geodemographics.co.uk/projects/qualitative-research-paper/.

——— (2019h). *Talk: Longley, Maguire, Goodchild & Rhind*. https://geodemographics.co.uk/projects/gds-longleyetal/.

——— (2020a). *The Geodemographics of British Politics*. https://geodemographics.co.uk/projects/envs615_analysis-of-human-dynamics_201442927/.

——— (2020b). *Implementation of the K-Means Clustering Algorithm*. https://github.com/peterprescott/kmeans.

——— (2020c). *Implementation of the Perceptron Algorithm*. https://github.com/peterprescott/perceptron.

Webber, R. and Burrows, R. (2018). *The Predictive Postcode: The Geodemographic Classification of British Society*. Sage.