Visualising Origin-Destination Data for Geographical Analysis: An Evaluation of Techniques

Kim Butterfield[*], Roger Beecham[†], and Alison Heppenstall[‡]

School of Geography, University of Leeds

January 2019

Summary

This paper evaluates a selection of innovative visualisation techniques used to analyse origin-destination flow data for geographical analysis. Firstly, the study identifies the most popular presentation techniques and tasks used within geographical studies, highlighting common issues associated with each method. Secondly, to demonstrate the effectiveness of these techniques, commuting between London boroughs using the 2011 Census Travel-to-Work data is briefly analysed. Commuting data has been chosen as it is an understandable and habitual flow for which robust and complete data are available. From this study, each visualisation technique is evaluated suggesting which method is best suited for various analytical tasks.

KEYWORDS: origin-destination, visualisation, Travel-to-Work, flow-mapping, 2011 Census

1. Introduction

Over recent years, data concerning the movement of people, goods, and services have become widely available; these movements between two locations are known as origin-destination data (OD-data) (Horner, 2004). Large, information-rich OD-datasets are important to analyse, notably by geographers, to uncover hidden spatial structures, and understand the movement of people. To analyse OD-data, statistical methods are often used, along with flow-line maps to visualise spatial patterns (Yang *et al.*, 2018). Whilst both methods are suitable for the data type, they are not the most useful depending on the study research task. It is integral that researchers consider the use of sophisticated visualisation techniques to gain geospatial insight into the dataset, thus uncovering underlying spatial patterns and flow relationships, and is an effective method of communicating these flows to decision-makers. However, many visualisation techniques are subject to various presentation issues (Boyandin, 2014).

This paper evaluates three innovative OD-visualisation techniques developed within the Information Visualisation (InfoVis) discipline. A short case study on commuting between London boroughs is provided using the 2011 Census. The visualisation techniques have been categorised into two types: flow maps as lines, and flow maps as matrices. Each technique is assessed in its ability to visualise data patterns, suitability for various tasks, and how effectively it presents spatial structures. The research tasks have been carefully chosen to reflect applied user needs when working with commuting data specifically, and to enable researchers to assess these visualisation techniques across a range of tasks commonly associated with flow data.

2. Background

Since 1837, straight-line flow mapping has become a popular technique for visualising OD-data and was computerised by 1959 (Tobler, 1987). The automation of OD flow maps began with simply using lines to connect OD-points (Jenny *et al.* 2016), and now includes the ability to vary lines by thickness,

[*] k.m.butterfield1@leeds.ac.uk
[†] r.j.beecham@leeds.ac.uk
[‡] a.j.heppenstall@leeds.ac.uk

transparency, and colour to present flow magnitude (Kadmon, 1971). Though this method is effective, geographers should consider applying innovative visualisation techniques as new underlying spatial patterns are yet to be discovered across many datasets.

3. Data and Methods

The Office for National Statistics (ONS) 2011 Census Travel-to-Work Table WU03UK containing flows between locations of usual residence and places of work at district-level by mode of transport was used. The dataset includes the following transport modes: underground, train, bus/minivan/coach, taxi, motorcycle/scooter/moped, car/van, passenger of car/van, bicycle, foot, or other. This dataset has not been heavily analysed within geographic journals (Rae, 2009; Andrienko *et al*., 2010), therefore there are potential new movement patterns to be discovered when applying novel or innovative methods.

To create the visualisations, ggplot2, Simple Features, and rmapshaper packages were used in RStudio. To produce flow maps as matrices, a gridded layout of London was used, created by ‘After the flood’ (https://aftertheflood.com/projects/future-cities-catapult/) (**Figure 1**). The grid repositions each London borough relative to its geographies, transforming each borough into the same size square for graphical analysis i.e. embedding matrices.

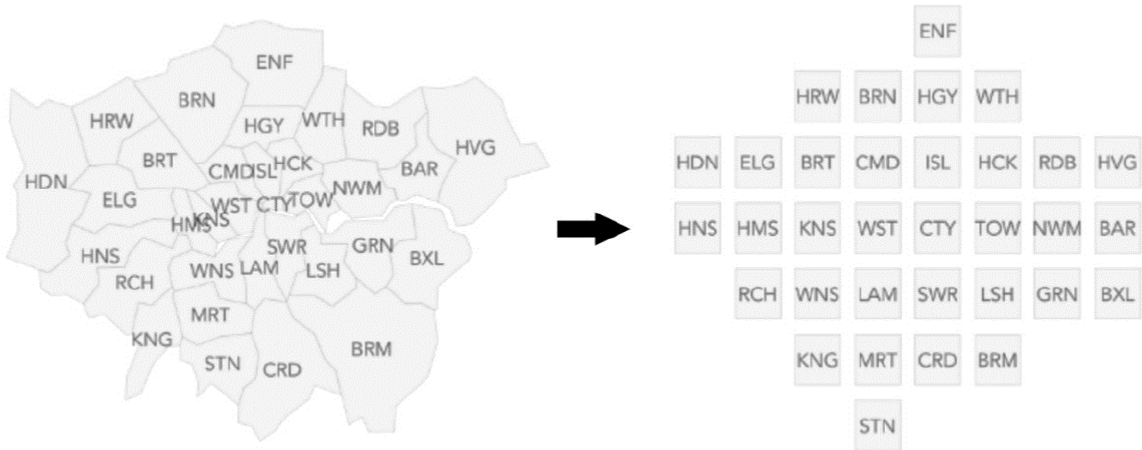

Figure 1 London boroughs transformed into a spatially-arranged grid map

4. Tasks and Common Issues

During visual OD-data analysis, researchers must consider the characteristics they are presenting, identify the most important factors of the data, and decide which technique is most appropriate. This section presents a list of visualisation tasks often associated with flow data types, and a list of commonly found issues in visualisations.

4.1 Visualisation Tasks

Visualisations should be easily interpreted, and researchers may want to emphasize specific aspects of their data to reveal hidden structures. There are numerous ‘tasks’ that the researcher will want to achieve through the visualisation.

The following analytical tasks were identified for commuting by mode of transport in London:
Task 1: To easily observe OD-flows to and from each borough in one visualisation.
Task 2: To preserve the spatial layout of the study area.
Task 3: To reveal hidden structures within the dataset that other techniques cannot.

Task 4: To visualise flow direction or bi-directionality.
Task 5: To visualise flow frequency/magnitude.

4.2 Visualisation Issues

All visualisation techniques have presentation problems, examples are presented in **Table 1**.

Table 1 Summary of Common Visualisation Issues

Issue	Summary
Salience bias	Extremely long or incidental flow lines are more prominent than others. However according to Tobler's First Law of Geography, the closer flow lines are, the more related and meaningful they are (Tobler, 1970; Wood *et al.* 2011, Beecham, 2012).
Visual clutter	Visualisations that contain large amounts of flows become unintelligible and difficult for the user to identify patterns within the data (Schneiderman, 1996; Card and Mackinlay, 1997; Ellis and Dix, 2007).
Occlusion	Overly thick lines that overlap one another will hide flows, leading to misrepresented results (Wood *et al.* 2010; Andrienko and Andrienko, 2013).
Flow direction	Traditional straight-line flow maps cannot show direction; thus, users cannot differentiate between in-flows and out-flows (Guo, 2009; Yang *et al.* 2018).

5. Visualisation Techniques

For this abstract, bike or foot travel is presented as these modes are often visualised for London-based geography/transport studies (Beecham, 2012; Wood *et al.* 2011; Wood *et al.* 2014). Each technique is evaluated on its ability of achieving the tasks in **Section 4.1**.

Flow Maps as Straight Lines and Bézier Curves

The most common technique is straight-line flow mapping (Cheshire and Uberti, 2014; Singleton, 2015; Lovelace *et al*., 2016; Rae, 2017). **Figure 2** presents a straight-line map of commuting between OD-pairs, using line encodings to present the data. Line thickness and saturation are weighted on journey frequencies; the darker/thicker the line, the more popular that journey. The most common journey pairs are between Westminster and Wandsworth, Lambeth, and Southwark. Commuting is also common from Wandsworth and the City of London, Camden and Hackney, suggesting that commuting by bike/foot is popular for those living/working in the city centre and between neighbouring districts.

Whilst this visualisation effectively identifies commuting pairs, it suffers from visualisation issues. Flow-direction is not visualised; thus, it is unknown whether more people are travelling into certain boroughs or out. Salience bias and occlusion are prevalent as flows are hidden 'underneath' thicker, saturated flow lines. Thus, the visualisation does not achieve Task 1, 3, or 4. Tasks 2 and 5 are achieved as the geography of London is preserved, and the frequency of flows are visualised.

Bézier curves are an alternative line-encoded technique (Wood *et al*., 2011; Beecham and Wood, 2014; Guo and Zhu, 2014; Jenny *et al*., 2017). In **Figure 3**, lines curve towards the destination of the OD-pair. Compared to **Figure 2**, it is now identifiable that popular commutes to Westminster and the City of London are in-flows as opposed to out-flows. This minor adjustment now visualises bi-directionality, thus partly achieving Tasks 2, 4, and 5. However, Bézier curves suffer from occlusion, as less frequent/shorter flows are hidden by longer/more salient lines therefore not achieving Tasks 1 and 3.

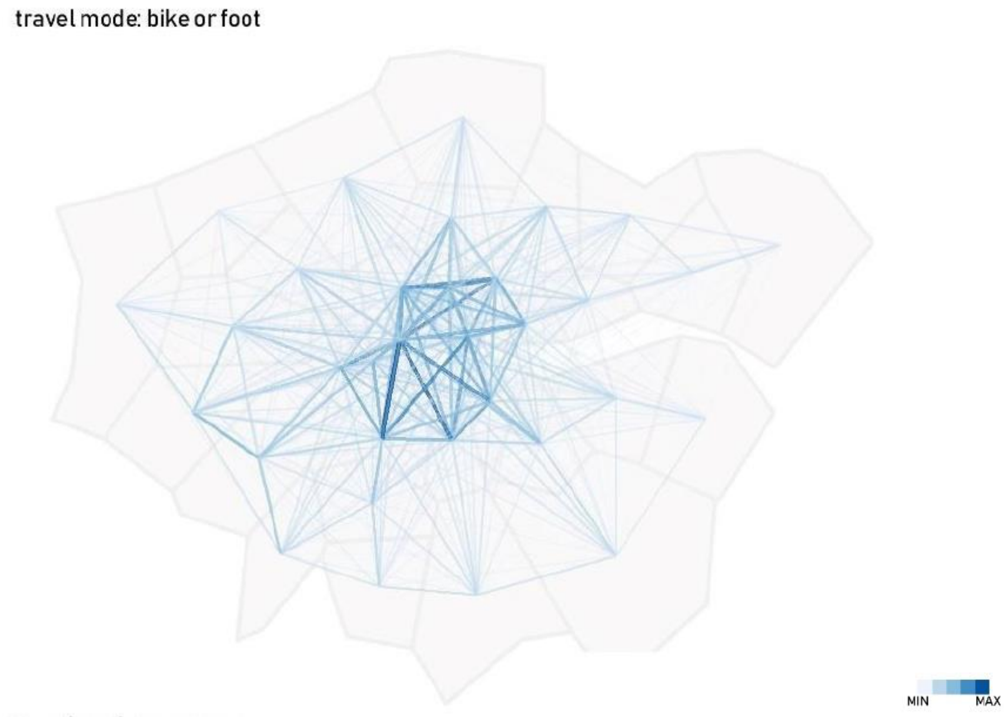

Figure 2 Straight-line flow map – commuting between London boroughs by bike or foot (Source: author).

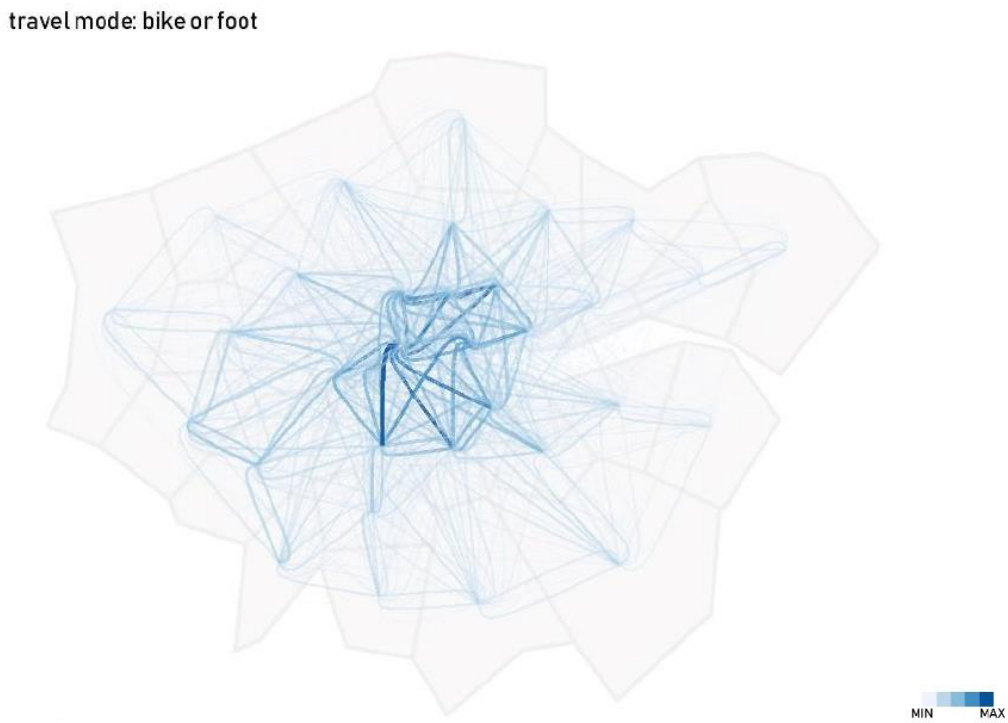

Figure 3 Bézier curve flow map – commuting between London boroughs by bike or foot (curve towards destination) (Source: author)

Flow Maps as Spatially-arranged OD-Matrices and Choropleth Maps

Spatially-arranged OD-matrices is an innovative technique proposed by Jo Wood. This entails converting geographies into spatially-arranged semi-abstract grids. This slight preservation of geography visualises in-flows and out-flows of each district without overlapping lines thus reducing occlusion and salience bias. This method is ideal for areas with grid-like geographies, such as Ireland (Kelly *et al*., 2013), and the USA (Wood *et al.*, 2010).

Figure 4 presents spatially-arranged OD-matrices with smaller spatially-arranged matrices embedded within. The matrices are encoded by saturation (darker colours represent more frequent commutes), and destination districts are outlined in black. Commutes by bike/foot are primarily performed between neighbouring districts. However, Camden, Islington, Westminster, City of London and Southwark have larger in-flows from districts further away; due to these small central districts being clustered together with cycle paths. The City of London experienced low numbers of out-commuters to almost all boroughs; this was overlooked in **Figure 3**. This visualisation achieves Tasks 1, 3, 4, and 5 - Task 2 to a lesser extent as the spatial layout has been distorted. This technique is useful for census data as the flow points are aggregated to district-totals, thus the visualisation does not suggest commutes occur between actual geocoded points as line-encoded maps suggest.

Figure 5 contains embedded choropleths. This allows the achievement of Task 2; however, some boroughs are physically small thus reducing readability. This technique is useful when boundary shape is important. This technique achieves Tasks 1, 2, 3, 4 and 5, though the data is subject to ecological fallacy, and commuters may live and work on edges of boroughs; however, this is a data issue, not a technique flaw.

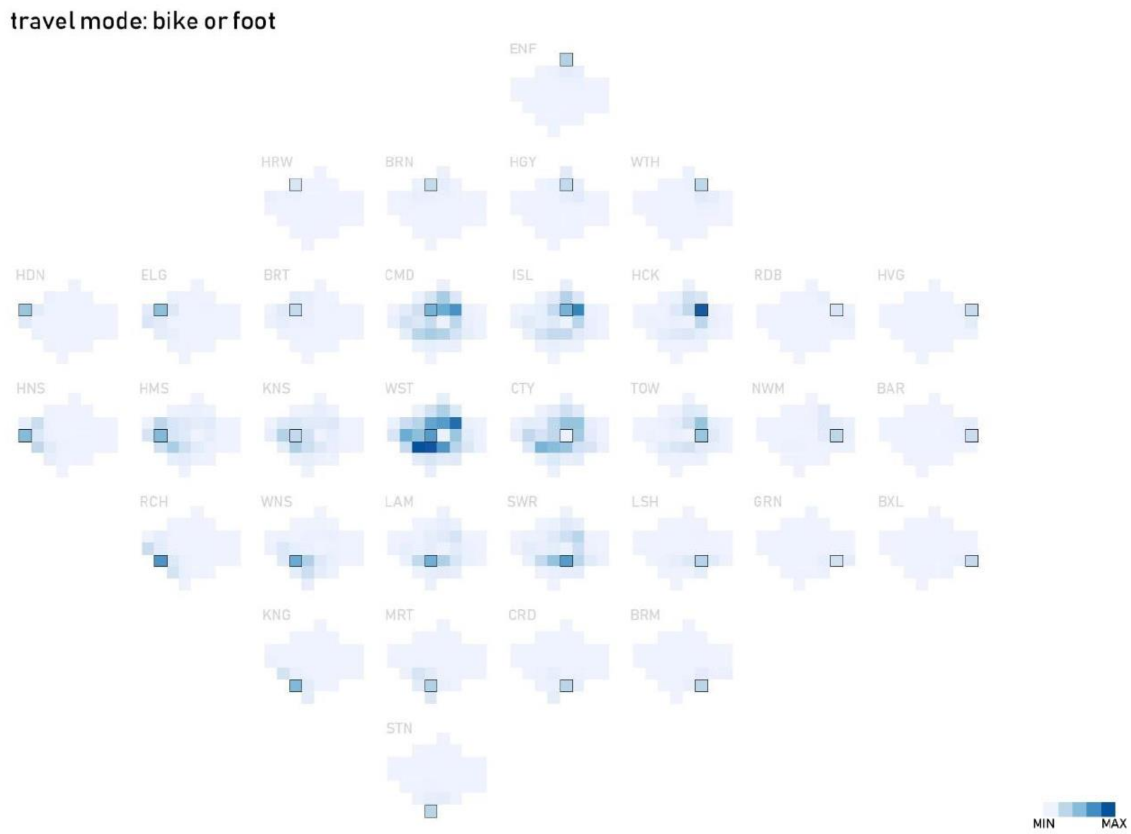

Figure 4 Spatially-arranged OD-Matrices – commuting between London boroughs by bike or foot (Source: author; Layout: aftertheflood.com)

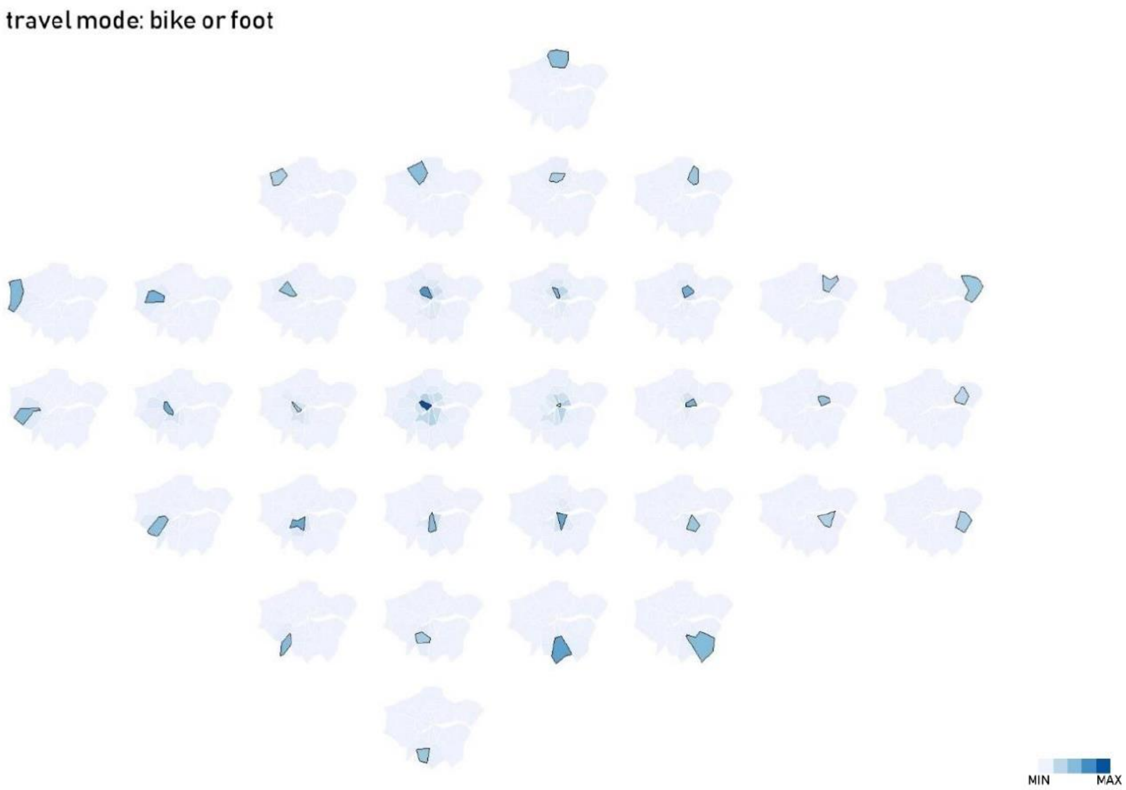

Figure 5 Spatially-arranged Choropleths – commuting between London boroughs by bike or foot (Source: author; Layout: aftertheflood.com)

6. Conclusions

This research demonstrates the importance of considering the use of alternative visualisation techniques for data analysis as not all methods are suited for all tasks. When choosing a visualisation technique, researchers should consider the geographic precision of their data points, what phenomena is to be identified (patterns, self-containment, etc.), and whether the visualisation produced meets all task criteria. Geographers are encouraged to apply visualisation techniques from the InfoVis discipline when analysing OD-data, and to consider learning to code visualisations using high-level processing frameworks such as ggplot2. This offers the ability to present OD-data using a range of techniques, and is a popular language amongst data scientists. Each visualisation technique has issues, and not all techniques can achieve all user tasks. To produce the best visual analysis of OD-data, it is essential to be mindful in visualisation decision-making. It is acknowledged that to fully evaluate the usefulness of these techniques, different forms of flow data need to be visualised such as non-complete datasets.

7. Acknowledgements

We would like to thank the ESRC White Rose Doctoral Training Partnership for their support; this piece of research was undertaken during Kim's +1 Masters year of her Collaborative Award.

8. References

Andrienko, G., Andrienko, N., Demsar, U., Dransch, D., Dykes, J., Fabrikant, S.I., Jern, M., Kraak, M.-J., Schumann, H. and Tominski, C. 2010. Space, time and visual analytics. *International Journal of Geographical Information Science*. **24**(10), pp.1577–1600.

Beecham, R. 2012. Using visual analytics to explore and explain cycling behaviours within a large-scale, transactional dataset. London, UK: City University London.

Beecham, R. and Wood, J. 2014. Exploring gendered cycling behaviours within a large-scale behavioural data-set. *Transportation Planning and Technology*. **37**(1), pp.83–97.

Boyandin, I. 2014. *Visualization of Temport Origin-Destination Data Thesis*.

Card, S.K. and Mackinlay, J. 1997. The structure of the information visualization design space. In *Information Visualization, 1997. Proceedings., IEEE Symposium on* (pp. 92-99). IEEE.

Cheshire, J. and Uberti, O. 2014. *London-The Information Capital*. Particular Books.

Ellis, G. and Dix, A. 2007. A taxonomy of clutter reduction for information visualisation. *IEEE transactions on visualization and computer graphics*. **13**(6), pp.1216-1223.

Guo, D. 2009. Flow mapping and multivariate visualization of large spatial interaction data. *IEEE Transactions on Visualization and Computer Graphics*, **15**(6).

Guo, D. and Zhu, X. 2014. Origin-Destination Flow Data Smoothing and Mapping. *IEEE Transactions on Visualization and Computer Graphics*. **20**(12), pp.2043–2052.

Horner, M.W. 2004. Spatial dimensions of urban commuting: a review of major issues and their implications for future geographic research. *The Professional Geographer*, **56**(2), pp.160-173.

Jenny, B., Stephen, D.M., Muehlenhaus, I., Marston, B.E., Sharma, R., Zhang, E. and Jenny, H. 2016. Design principles for origin-destination flow maps. *Cartography and Geographic Information Science*. **45**(1), pp.62–75.

Jenny, B., Stephen, D.M., Muehlenhaus, I., Marston, B.E., Sharma, R., Zhang, E. and Jenny, H. 2017. Force-directed layout of origin-destination flow maps. *International Journal of Geographical Information Science*. **31**(8), pp.1521–1540.

Kadmon, N. 1971. KOMPLOT “Do-it-yourself” computer cartography. *The Cartographic Journal.* **8**(2), 139–144.

Kelly, M., Slingsby, A., Dykes, J. and Wood, J. 2013. Historical Internal Migration in Ireland.

Lovelace, R., Goodman, A., Aldred, R., Berkoff, N., Abbas, A. and Woodcock, J. 2017. The Propensity to Cycle Tool: An open source online system for sustainable transport planning. *Journal of Transport and Land Use*, **10**(1), pp.505-528.

Rae, A. 2009. From spatial interaction data to spatial interaction information? Geovisualisation and spatial structures of migration from the 2001 UK census. *Computers, Environment and Urban Systems*. **33**(3), pp.161–178.

Rae, A. 2017. The Geography of Travel to Work in England and Wales: Extracts from the 2011 Census. *Applied Spatial Analysis and Policy*. **10**(4), pp.457–473.

Shneiderman, B. 1996, September. The eyes have it: A task by data type taxonomy for information visualizations. In *Visual Languages, 1996. Proceedings., IEEE Symposium on* (pp. 336-343). IEEE.

Singleton, A.D. 2015. Featured graphic. Transport mode choice and the commute to work. *Environment and Planning A: Economy and Space*. **47**(7), pp.1401–1403.

Tobler, W.R. 1987. Experiments In Migration Mapping By Computer. *The American Cartographer*. **14**(2), pp.155–163.

Wood, J., Dykes, J. and Slingsby, A. 2010. Visualisation of Origins, Destinations and Flows with OD Maps. *The Cartographic Journal*. **47**(2), pp.117–129.

Wood, J., Slingsby, A. and Dykes, J. 2011. Visualizing the Dynamics of London's Bicycle-Hire Scheme. *Cartographica: The International Journal for Geographic Information and Geovisualization*. **46**(4), pp.239–251.

Wood, J., Beecham, R. and Dykes, J. 2014. Moving beyond sequential design: Reflections on a rich multi-channel approach to data visualization. *IEEE transactions on visualization and computer graphics*, **20**(12), pp.2171-2180.

Yang, Y., Dwyer, T., Jenny, B., Marriott, K., Cordeil, M. and Chen, H., 2018. Origin-Destination Flow Maps in Immersive Environments. *IEEE transactions on visualization and computer graphics*, **25**(1), pp.693-703.

9. Biographies

Dr **Roger Beecham** is a lecturer at the University of Leeds working in Social and Geographic Data Science. His expertise includes data visualisation, usable statistics, and open source computing.

Professor **Alison Heppenstall** is a professor in Geocomputation at the University of Leeds and holds an ESRC-Alan Turing Institute Fellowship. Her research interests are focused on the development of artificial intelligent and machine learning methods.

Ms. **Kim Butterfield** is currently studying (first year) for a Ph.D. at School of Geography, University of Leeds. Her interests are in agent-based modelling, data visualisation, and retail location planning.