

Regionalizing and Understanding Commuter Flows: An Open Source Geospatial Approach

Lorraine Barry

School of Natural and Built Environment, Queen's University Belfast
l.barry@qub.ac.uk
January 2017

Summary

This contribution describes an open source geospatial approach to modelling, visualising and understanding the complex patterns of aggregate origin destination commuter flows within Northern Ireland. It explores the existing problems with modelling such flows and proposes methods to analyse fully the population demographics of movement patterns. Such methods develop the use and potential of recently available programming libraries such as PySAL* and D3*.js which have the capacity to integrate, regionalise, visualise and analyse spatially at an advanced level.

*PySAL Python Spatial Analysis Library © Copyright 2014-, PySAL Developers; 2009-13 Sergio Rey.

*D3 Copyright 2015 Mike Bostock.

KEYWORDS: Clustering, Geovisualisation, SIM, PySAL, Open Source GIS

1. Current situation and history of commuter modelling

This research outlines the main limitations of the use of current aggregate data for the analysis of commuting flows using the example of Northern Ireland. Current aggregations available from census data at ward level vary greatly in size and internal commuting flows and are therefore a highly unsuitable geography for the basis of spatial interaction modelling and the understanding of commuting patterns based on socio-demographic population characteristics. This unsuitability was demonstrated by counter-intuitive results for rural parts of the province from research conducted by Lloyd, Shuttleworth & Catney in 2008 (Lloyd et al, 2008) and importantly demonstrate the susceptibility of flow data analysis to the effects of aggregation and zone structure. Figure 1 and 2 show the variation in ward sizes and proportions of internal commuters for the Belfast area.

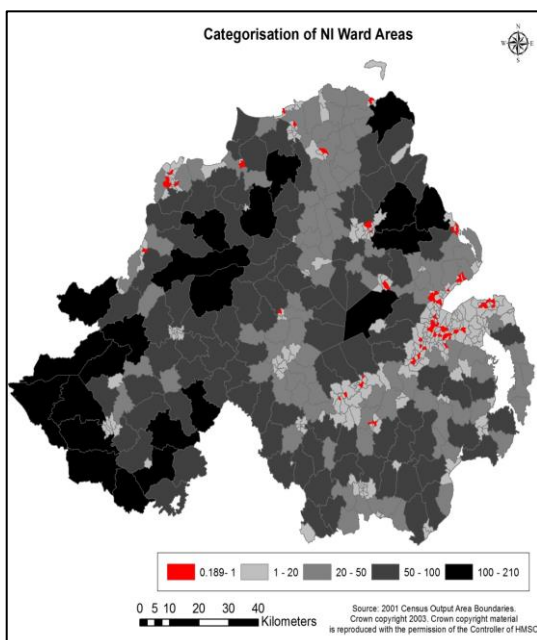


Figure 1 Variation in Ward Sizes NI

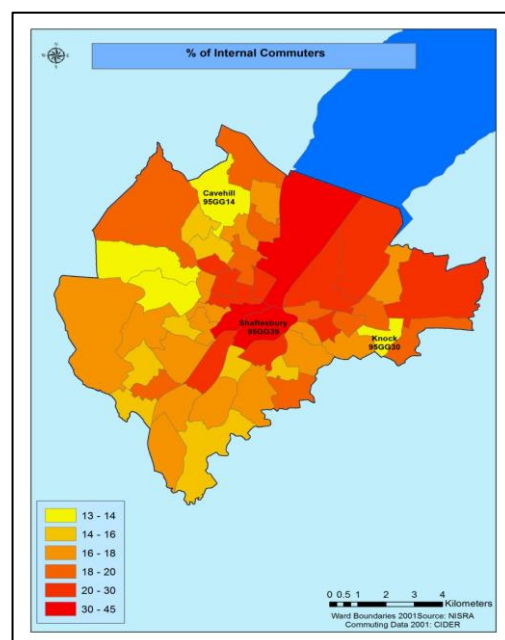


Figure 2 Internal Commuters for Belfast Wards

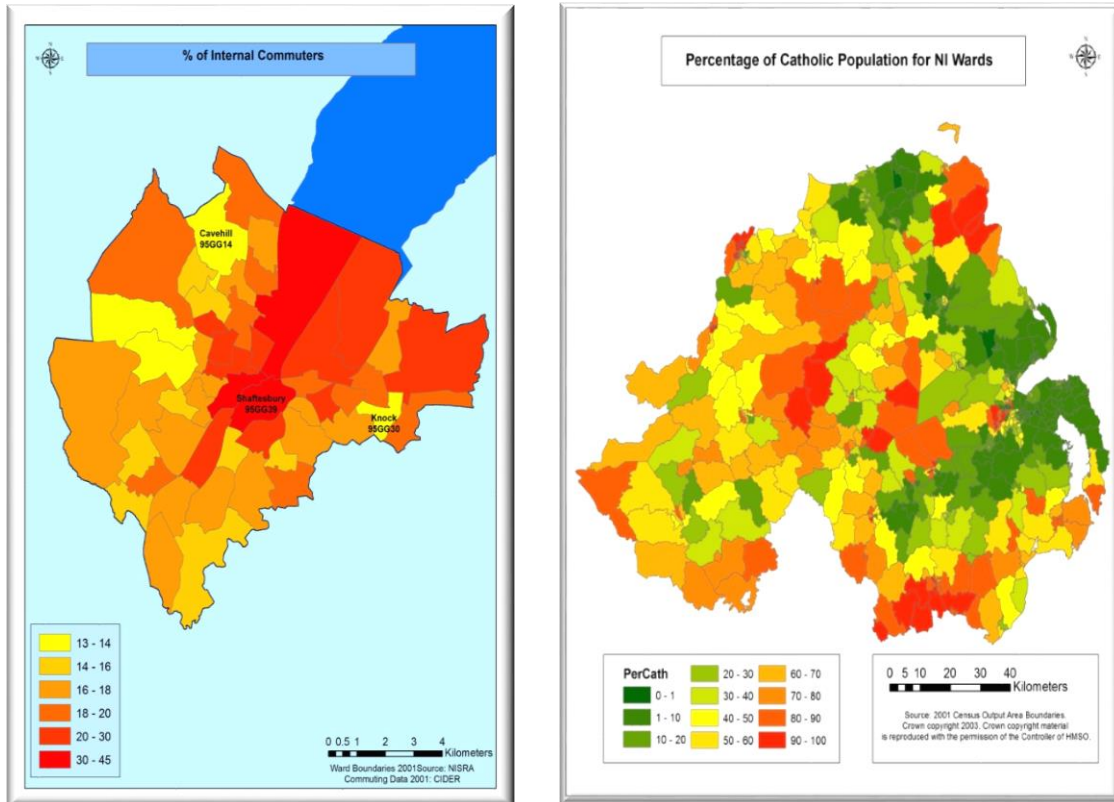


Figure 5 Test Created Regions Run based on regionalisation

Figure 6 Illustration of the levels of religious segregation within Northern Ireland

4.2. Spatial Interaction Modelling

OResidPopA	ONoWorkers	OCath	OPerCath	DestID	DResidPopA	DNoWorkers	DCath	DPerCath	DPerCath	DPerCath	Residual	DevianceResidual	StdDevianceResidual
5764	1591	5402	93.719639	95GG01	5764	1591	5402	93.719639	0.00000	0.00000	261.475	28.438	28.443
5764	1591	5402	93.719639	95GG02	6602	1256	5720	86.640412	7.079227	7.079227	-29.820	-7.723	-7.724
5764	1591	5402	93.719639	95GG03	5693	2366	491	8.624627	85.095012	85.095012	-33.270	-8.157	-8.164
5764	1591	5402	93.719639	95GG04	4942	1087	2200	44.516390	49.203249	49.203249	-28.636	-6.568	-6.570
5764	1591	5402	93.719639	95GG05	5251	2289	2632	50.123786	43.595853	43.595853	-28.388	-6.534	-6.536
5764	1591	5402	93.719639	95GG06	6020	2008	113	1.877076	91.842563	91.842563	-33.586	-8.196	-8.204
5764	1591	5402	93.719639	95GG07	5504	1637	4773	86.718750	7.000889	7.000889	37.183	5.842	5.843
5764	1591	5402	93.719639	95GG08	4942	1618	2736	55.362202	38.357437	38.357437	-31.158	-7.894	-7.896
5764	1591	5402	93.719639	95GG09	6025	2241	167	2.771784	90.947855	90.947855	-30.544	-6.827	-6.833
5764	1591	5402	93.719639	95GG10	3964	1344	66	1.664985	92.054654	92.054654	24.404	3.812	3.816

Figure 7 – Poisson Regression for a test set of created regions

Spatial interaction modelling is utilised to evaluate the patterns between volume of flows and the underlying socio-economic tendencies of the origin and destination zones (Lloyd et al, 2011). The level of segregation is significant for a large proportion of wards within the province, as illustrated in figure 6. Figure 7 illustrates the initial findings for a local Poisson regression for a set of created regions from the regionalisation process.

4.3. Geovisualization

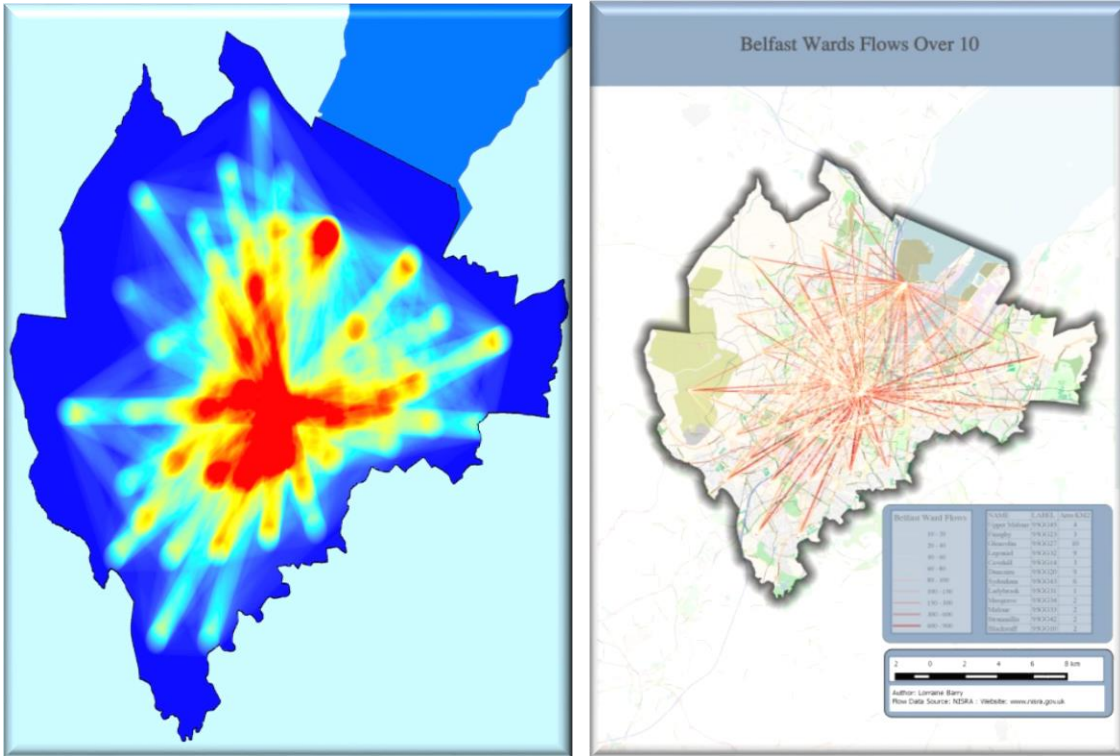


Figure 8 - Raster density of Belfast area flows **Figure 9** – Open source QGIS visualisation

Geovisualization of flow data is an important aspect which compliments the regionalisation and spatial interaction process. It is important for the effective display of results and to simplify the understanding and analysis of inherently voluminous data. Geovisualization is facilitated through the adoption of open source libraries such as D3 and also open source software such as QGIS. Figure 8 and 9 represent the geovisualizations of voluminous flow data using QGIS.

5. Summary

The presentation will show the fundamentals of how open source core geospatial libraries can evaluate and explore the spatial patterns of complex interaction data. This understanding impacts on city and local planning, social cohesion and community wellbeing.

The three strands of this research – Regionalization, Geovisualization and Spatial Interaction Modelling – are vital to understanding population flows and only possible through an adoption and development of open source methodologies. The methods presented are largely applicable to any aggregate analysis of interaction data - commuting, migration, and travel and consumer patterns. The use of both python and new or improved regionalization and visualization methods are being actively applied to the study of lifestyle data (Folch & Spielman 2014).

6. Acknowledgements

Census commuter data and population demographic data provided by the Northern Ireland Statistics and Research Agency (NISRA). Background geospatial data provided by Land and Property Services, Department of Finance, Northern Ireland under the NIMA agreement MOU203. Acknowledgement for use and development of PySAL, the Python Spatial Analysis Library © Copyright 2014-, PySAL Developers.

7. Biography

I have been a GI Scientist for over 18 years and am currently undertaking a part-time PhD in open GIScience to model and better understand commuter flows. I have a keen interest in open source GIS and have presented at FOSS4G-NA in 2015 and FOSS4G 2016.

References

Alvanides, S., Openshaw, S. and Duke-Williams, O. (2000) Designing zoning systems for flow data. In P. M. Atkinson and D. J. Martin (Eds.) *Innovations in GIS 7: GIS and Geocomputation*. London: Taylor and Francis, pp. 115–134.

Duque, J., L. Anselin and S. Rey. 2012. The Max-p-Regions Problem. *Journal of Regional Science* 52 (3): 397-419.

Folch, D. and Spielman, S.E. (2014) Identifying Regions based on Flexible User Defined Constraints. *International Journal of Geographic Information Science*.

Lloyd, C. D. and Congdon, P. (2011) A spatial Random Effects Model for Interzone flows: Commuting in Northern Ireland. *Journal of Applied Statistics*.

Lloyd, C. D. Shuttleworth, I. G. and Catney, G. (2008) Origin-specific models for analysing commuting flows in Northern Ireland: scale effects and other problems. In D. Lambrick (Ed.) *GISRUK 2008. Proceedings of the GIS Research UK 16th Annual Conference. Manchester Metropolitan University, 2nd–4th April 2008*. Manchester: Manchester Metropolitan University, pp. 145–149.

Openshaw, S. & Taylor, P. 1981. The modifiable unit problem. In N. Wrigley, ed. *Quantitative geography, a British view*. London, Pion. pp. 127-144

Rey, S. and L. Anselin, 2007. PySAL: A Python Library of Spatial Analytical Methods. *The Review of Regional Studies* 37(1), 2007: 5–27.