Journal of Rural and Community Development

Triangulating Neighborhoods: A Research Note on Improving Links Between People and Places in Smaller Cities and Rural Areas

Authors: Brittany Barber, Rachel McLay, Daniel Rainham, & Howard Ramos

Triangulating Neighborhoods
A Research Note on Improving Links Between People and Places in Smaller Cities and Rural Areas

Brittany Barber
Dalhousie University
Halifax, Nova Scotia, Canada
Brittany.Barber@dal.ca

Rachel McLay
Dalhousie University
Halifax, Nova Scotia, Canada
Rachel.McLay@dal.ca

Daniel Rainham
Dalhousie University
Halifax, Nova Scotia, Canada
Daniel.Rainham@dal.ca

Howard Ramos*
Dalhousie University
Halifax, Nova Scotia, Canada
Howard.Ramos@dal.ca

*Corresponding author

Abstract

There is growing interest among social and behavioral scientists in exploring concepts of place through characteristics of social and physical space. In smaller regions, such research faces a number of obstacles due to limitations of the geo-spatial units available, and postal code linkages to these units can be particularly unreliable. This research note explores how triangulating postal code data with open field survey questions on the name of a neighbourhood and adjacent streets can help improve understandings of communities. The research note offers a practical overview of how triangulating missing information can help resolve misclassification errors in postal codes or other geocoding.

Keywords: methodology; socio-spatial; geocoding; postal code; neighborhood

---

1.0 Introduction

There is growing interest in exploring concepts of place through characteristics of social and physical space (Kwan, 2012; Longley, 2012). A wide array of research draws upon spatially integrated perspectives (Anselin, 1999; Goodchild, Anselin, Appelbaum, & Harthorn, 2000) to answer important research questions of socio-economic, health, and environmental significance (Bannister-Tyrrell et al., 2017; Busch & Ferretti-Gallon, 2017; Longhi, Nijkamp, & Poot, 2010; Nieuwenhuis & Hooimeijer, 2016). Within Canada, research increasingly focuses on the distinction

between urban and rural geographies for development strategies and policies (Alasia, 2004), for enhancements to service delivery and access to resources (Mantler, Jackson, & Ford-Gilboe, 2018; Marr, 2015), and for economic and environmental sustainability (Halseth, Markey, Ryser, Hanlon, & Skinner, 2017; Hvenegaard, Hallstrom, & Brand, 2019). However, socio-spatial research on smaller regions faces a number of obstacles due to limitations of the geo-spatial units available (Kaida, Ramos, Singh, & McLay, 2020). Moreover, postal code linkages to these units can be particularly unreliable in rural areas because community mailboxes and shared postal codes across large geographical regions often prevent specificity. This research note lays out the challenges encountered when using postal codes to link individuals to their communities' census geographies. It offers a practical overview of how survey questions can be used to triangulate missing information and resolve misclassification errors in postal codes.

Postal codes are one of the most commonly used geo-spatial units, along with civic addresses and place names, and they are often the only type of geographic identifier available to Canadian researchers (Khan, 2018). However, the use of postal codes for identifying location is subject to several challenges (Goldberg & Jacquez, 2012), including poor match rates or incomplete matches from minor data entry errors, the restricted use of partial postal codes (e.g., FSAs) for assigning locations (Chow, Dede-Bamfo, & Dahal, 2016), and misclassification errors, where a postal code is incorrectly assigned to geographic boundaries (Jacquez, 2012). Misclassification errors arise more commonly in rural areas because their postal code areas are much larger, crossing multiple census boundaries. The Postal Code Conversion File Plus (PCCF+) is a popular research tool for linking postal codes to Statistics Canada's standard geographic areas. It uses population-weighted probabilistic matching to assign postal codes to larger geographic units when more than one geographic assignment is possible (Statistics Canada, 2017). The tool is derived from earlier versions of the PCCF, which do not use proportional assignment of postal codes based on population counts and links them to standard geographic areas without accounting for postal codes that could link to more than one geographic area. The PCCF+ is thus a valuable and effective tool for understanding macro-level trends; however, in studies that require exact linkages between individuals and their census geographies, its weighting and probabilistic matching inevitably introduce some error (Terashima & Kephart, 2016) and this is compounded in smaller cities and rural regions.

Challenges and limitations associated with the use of postal codes for spatial analysis are further compounded by the fact that there are few technical resources available in the social science literature to assist in validating or improving the accuracy of locations identified through geocoding or other approaches. One approach, triangulation, involves the use of multiple methods or data sources to assist in improving location accuracy, offering more comprehensive or complete information (Bekhet & Zauszniewski, 2012; Risjord, Moloney & Dunbar, 2001). Across-method triangulation does this using both quantitative and qualitative data-collection techniques (Casey & Murphy, 2009). In surveys, or within methodology approaches, this can mean a combination of open questions and more traditional closed or numeric questions; it can also be done through using multiple surveys or quantitative sources. Given the challenges faced by those using postal codes within social science research, gathering different kinds of data to triangulate locations can be an important or necessary step to validate location information. In this research note, we draw upon our own experience from the Perceptions of Change Project to

illustrate how completeness and positional accuracy of postal code-derived locations can be triangulated using supplementary and open-field survey questions.

Our research set out to explore people's perceptions of change in their community and sought to assess how that compared to outside measures. We did this through surveys and compared results from them against neighborhood- and community-level census data. Our survey was conducted by (landline and mobile) telephone with 1,848 participants in four small and mid-sized cities: Halifax, Nova Scotia, Charlottetown, Prince Edward Island, St. John's, Newfoundland, and Moncton, New Brunswick. We surveyed within the Census Metropolitan Area (CMA) or Census Agglomeration (CA), which includes suburban areas as well as outlying areas that are considered rural. The Charlottetown CA, in particular, covers a large amount of rural area. Participants were asked 75 questions about changes they observed in their neighborhoods and city as a whole. Of these, three questions asked about participants' location and neighborhood. The questions were created with the knowledge that telephone surveys typically suffer from high non-response rates (Statistics Canada, 2008), in part due to the use of invasive or sensitive questions, including those inquiring about participants' location or their home address (Kennedy & Vargus, 2001). For these reasons, we asked participants for their postal code, and we also asked them to name their neighborhood and identify several neighboring streets. This allowed us to understand participants' location, sometimes to the neighborhood block level, without asking them to provide their civic address or street name. We then linked our survey data to census data using the PCCF+ 7A. Using supplementary and open-field location questions, we were able to triangulate missing postal codes and correct errors generated when using the probabilistic matching of the PCCF+. The remainder of this research note broadly outlines our approach of 'triangulating neighborhoods' and specifies its benefits.

2.0 Triangulating Neighborhoods

Our process of triangulation began with the research team identifying any data entry errors resulting in false postal codes. We did this by analyzing alpha-numeric strings that resemble real postal codes but which did not match any existing codes. This can be done using the PCCF or PCCF+ and examining the errors or non-matches. Common data entry errors included mistaking "S" for similar-sounding "F," the number "0" for the letter "O," and so on. When participants' neighborhood and street names corresponded to a postal code closely resembling the recorded code (e.g.,"A1L1H6" instead of "A!L1H6" or "B2V1A4" instead of "B2B1A4"), the code was replaced. Through this process, the team was able to recover 27 postal codes among the 63 non-matches that were not considered valid codes by Canada Post; these cases would have otherwise been lost. Recovering such cases, especially when working with samples across smaller regions, can substantially improve the number of cases included in analysis.

While data entry errors did not disproportionately affect rural areas, misclassification errors did, and this problem was aggravated by the limitations of geo-spatial units available in small cities like Charlottetown (see Kaida et al., 2020). The PCCF+ considers matches to be unique when a postal code can be linked to a single dissemination area (DA), dissemination block, or block face, and the majority of urban postal codes meet this criterion (Statistics Canada, 2017). When codes have multiple matches (covering multiple DAs), probabilistic weighting is used to assign these cases to specific geographies. Even when codes do span multiple matches,

these DAs are likely to fall within the same census tract (CT); in mid-sized and larger cities, CTs are commonly used as a proxy for neighborhoods. Random spot-checking of assignments made using probabilistic weighting in Halifax, Moncton, and St. John's found that CT-level assignments were validated by the contextual data on participants' locations, even when DA-assignments were not.

Smaller cities, including Charlottetown, are not assigned CTs, so other geo-spatial units must be used to study neighborhood- or community-level data (Kaida et al., 2020). We chose census subdivisions (CSDs) as the best option for rural and outlying areas within the Charlottetown CA, but we found that DA-level probabilistic assignments were often spread across multiple CSDs, which resulted in more arbitrary assignments. For this reason, we decided that Charlottetown postal codes with non-unique linkages required further investigation. This included Charlottetown postal codes with a link type code lower than 9 in the PCCF+ 7A, where 9 signals unique linkages (Statistics Canada, 2017). One-hundred-sixty-four postal codes were triangulated. Among these, 124 were in rural areas, out of 149 rural codes (83.2%), compared to 40 in "urban" Charlottetown, out of 281 urban codes (14.2%). In total, we manually reassigned 30 cases (27 rural, 3 urban) to different CSDs where there was evidence of a mismatch between the PCCF+ assignment and contextual information. We also removed 19 assignments (16 rural, 3 urban) made by the PCCF+ in cases where there was not enough information to support any assignment.

Triangulating geo-location through postal codes and neighboring streets can also be important for improving completeness when only partial postal codes or forward sortation areas (FSAs) are available. Cross-checking such matches against neighborhood names and nearby streets allows us to retain more accurate data and overcome some of the challenges associated with using the PCCF+ by validating the assignments made by probabilistic matching. More specific information about nearby street names is often most valuable in urban areas, especially when participants name side-streets or intersections. In many rural areas, however, residents live near a single long road or highway, which often crosses multiple geographic boundaries. In these cases, the name of their community is often sufficient to validate an assignment. Each of these practices improves the geocoding of responses and the accuracy of postal codes or other socio-spatial units used.

3.0 The Value of Triangulation

In this note, we have shared our experiences using supplemental survey questions inquiring about participants' neighborhood or location to investigate the completeness and accuracy of postal codes and improve subsequent data linkages and spatial analysis. Our example shows how the use of strategic and non-invasive survey questions to gather geographical information can help researchers triangulate participants' neighborhoods and improve the accuracy and quality of geographic identifiers. Across-method triangulation can be employed in surveys by gathering multiple kinds of locational information, including postal codes, neighborhood or community names, and the names of nearby streets.

Ultimately, the extra effort to derive multiple types of location information beyond residential postal code can provide valuable dividends, especially to improve coverage in rural areas and in studies where exact matches between participants and locations are necessary. Inaccurate location information compromises the validity of inferences made from location-based relationships and limits the usefulness of

spatial analysis (Yang, Bilaver, Hayes, & George, 2004). Given the limited examples of mixed-methods approaches to the determination of location across social and behavioral sciences, future research should consider employing multiple approaches to collecting spatially oriented data when possible (Pavlovskaya, 2006). As our experience suggests, gathering open-ended responses to supplement and contextualize postal codes or other geographic indicators helps ensure more accurate geographical linkages for individual data, which is needed for understanding the effects of social and physical space within socio-spatial analysis.

Acknowledgements

References

Alasia, A. (2004). *Mapping the socio-economic diversity of rural Canada: A multivariate analysis.* (Catalogue No. 21-601-MIE). Ottawa, Statistics Canada: Agriculture Division. Retrieved from https://www150.statcan.gc.ca/n1/pub/21-601-m/21-601-m2004067-eng.pdf

Anselin, L. (1999). The future of spatial analysis in the social sciences. *Annals of GIS*, *5*(2), 67–76. https://doi.org/10.1080/10824009909480516

Bannister-Tyrrell, M., Verdonck, K., Hausmann-Muela, S., Gryseels, C., Muela Ribera, J., & Peeters Grietens, K. (2017). Defining micro-epidemiology for malaria elimination: systematic review and meta-analysis. *Malaria Journal*, *16*(1). https://doi.org/10.1186/s12936-017-1792-1

Bekhet, A. K., & Zauszniewski, J. A. (2012). Methodological triangulation: An approach to understanding data. *Nurse Researcher*, *20*(2), 40–43. https://doi.org/10.7748/nr2012.11.20.2.40.c9442

Busch, J., & Ferretti-Gallon, K. (2017). What drives deforestation and what stops it? A meta-analysis. *Review of Environmental Economics and Policy*, *11*(1), 3–23. https://doi.org/10.1093/reep/rew013

Casey, D., & Murphy, K. (2009). Issues in using methodological triangulation in research: Dympna Casey and Kathy Murphy explore the advantages and disadvantages of using triangulation. *Nurse Researcher*, *16*(4), 40–55. https://doi.org/10.7748/nr2009.07.16.4.40.c7160

Chow, T. E., Dede-Bamfo, N., & Dahal, K. (2016). Geographic disparity of positional errors and matching rate of residential addresses among geocoding solutions. *Annals of GIS, 22*(1), 29–42. https://doi.org/10.1080/19475683.2015.1085437

Goldberg, D. W., & Jacquez, G. M. (2012). Advances in geocoding for the health sciences. *Spatial and Spatio-temporal Epidemiology, 3.* 1–5. doi:10.1016/j.sste.2012.02.001pmid:22469486

Goodchild, M. F., Anselin, L., Appelbaum, R. P., & Harthorn, B. H. (2000). Toward spatially integrated social science. *International Regional Science Review*, *23*(2), 139–159 https://doi.org/10.1177%2F016001760002300201

Halseth, G., Markey, S., Ryser, L., Hanlon, N., & Skinner, M. (2017). Exploring new development pathways in a remote mining town: The case of Tumbler Ridge, BC, Canada. *The Journal of Rural and Community Development, 12*(2/3), 1–22.

Hvenegaard, G. T., Hallstrom, L. K., & Brand, K. L. P. (2019). Implementation dynamics for sustainability planning in rural Canada. *The Journal of Rural and Community Development, 14*(1), 54–76.

Jacquez, M. G. (2012). A research agenda: Does geocoding positional error matter in health GIS studies? *Spatial and Spatio-temporal Epidemiology, 3*(1)*,* 7–12. https://doi.org/10.1016/j.sste.2012.02.002

Kaida, L., Ramos, H., Singh, D., & McLay, R. (2020). How to capture neighborhood change in small cities. *Canadian Studies in Population, 26*(8)*,* https://doi.org/10.1007/s42650-020-00026-8

Kennedy, J. M., & Vargus, B. (2001). Challenges in survey research and their implications for philanthropic studies research. *Nonprofit and Voluntary Sector Quarterly, 30*(3), 483–494.

Khan, S. (2018). Positional accuracy of geocoding from residential postal codes versus full street addresses. *Health Reports*, *29*(82), 9.

Kwan, M.-P. (2012). How GIS can help address the uncertain geographic context problem in social science research. *Annals of GIS*, *18*(4), 245–255. https://doi.org/10.1080/19475683.2012.727867

Longhi, S., Nijkamp, P., & Poot, J. (2010). Joint impacts of immigration on wages and employment: review and meta-analysis. *Journal of Geographical Systems*, *12*(4), 355–387. https://doi.org/10.1007/s10109-010-0111-y

Longley, P. A. (2012). Geodemographics and the practices of geographic information science. *International Journal of Geographical Information Science*, *26*(12), 2227–2237. https://doi.org/10.1080/13658816.2012.719623

Mantler, T., Jackson, K. T., & Ford-Gilboe, M. (2018). The CENTRAL hub model: Strategies and innovations used by rural women's shelters in Canada to strengthen service delivery and support women. *The Journal of Rural and Community Development, 13*(3), 115–132.

Marr, E. (2015). Assessing transportation disadvantage in rural Ontario, Canada: A case study of Huron County. *The Journal of Rural and Community Development, 10*(2), 100–120.

Nieuwenhuis, J., & Hooimeijer, P. (2016). The association between neighborhoods and educational achievement, a systematic review and meta-analysis. *Journal of Housing and the Built Environment*, *31*(2), 321–347. https://doi.org/10.1007/s10901-015-9460-7

Pavlovskaya, M. (2006). Theorizing with GIS: A tool for critical geographies? *Environment and Planning A: Economy and Space*, *38*(11), 2003–2020. https://doi.org/10.1068/a37326

Risjord, M., Moloney, M., & Dunbar, S. (2001). Methodological triangulation in nursing research. *Philosophy of the Social Sciences*, *31*(1), 40–59. https://doi.org/10.1177/004839310103100103

Statistics Canada. (2017). *Postal Code Conversion File Plus (PCCF+) Version 7A, Reference Guide.* (Catalogue no. 92-154-G). Ottawa, Statistics Canada. Retrieved from https://www150.statcan.gc.ca/n1/pub/92-154-g/92-154-g2017001-eng.htm

Terashima, M., & Kephart, G. (2016). Misclassification errors from postal code-based geocoding to assign census geography in Nova Scotia, Canada. *Can J Public Health*, *107*(4–5), 424. https://doi.org/10.17269/cjph.107.5459

Yang, D. H., Bilaver, L. M., Hayes, O., & Goerge, R. (2004). Improving geocoding practices: Evaluation of geocoding tools. *Journal of Medical Systems*, *28*(4), 361–370. https://doi.org/10.1023/B:JOMS.0000032851.76239.e3