Local Patterns of National Household Survey Non-Response in Canadian Cities

Scott Bell, Kelsey Bates, Kyle Snarr and Jessica Alegria

June 22, 2018

Local Patterns of National Household Survey Non-Response in Canadian Cities

Scott Bell[1]
Kelsey Bates[1]
Kyle Snarr
Jessica Alegria

1 Geography and Planning, University of Saskatchewan, Saskatoon, SK

Abstract

Statistics Canada and the Canadian government invoked a dramatic change in the collection of detailed demographic and other data for the Census year 2011. Despite reverting in 2016 to the traditional "long form" census format, the National Household Survey (NHS) of 2011 represents an important and meaningful opportunity for study. Furthermore, with a 10-year gap between instances of the more reliable "long form" survey format, users of detailed census data products face challenges if interested in demographic, economic, social, and other changes that happened between 2006 and 2016 or trends in such data over a period that includes the 2011 NHS. Here we examine patterns of non-response, using the variable Global Non-Response (GNR) in several Canadian cities using dissemination areas (DA) as the unit of analysis. We will also show patterns of similarity and dissimilarity with GNR and other NHS variables (social, demographic, ethnic, housing, etc.).

Background and Relevance

For the Census year 2011, Statistics Canada (and the Canadian government) invoked substantial changes in the way they collect census data regarding the Canadian populace. Canada has shifted from a legally-enforced "long form" with a 20% sample, to a voluntary National Household Survey (NHS) with a 30+% sample (Canada, 2011a). This change raised concerns regarding the spatial variability of uncertainty in NHS data across Canada and reliability more generally. Understanding the distribution of data uncertainty is essential for researchers to consider, particularly if they are considering using such data for research. Understanding data uncertainty is also central to communicating patterns present in populations represented by that sampled data. For decision makers (planners, policy makers, and elected and non-elected government officials) the need to understand underlying demographic patterns is essential to making informed decisions. Understanding, or at least having access to, the range of possible outcomes of a sampled dataset plays an important role in the likelihood of invoking a decision, the voracity with which a spatial pattern is defended, or even the likelihood of it being used as part of the decision making process. Decision making in Canadian cities is normally based on emerging patterns of population change summarized by census data and mapped according to standard units of analysis. The above described change demands that we examine the availability as well as variability of uncertainty of the collected data (Kardos, Benwell, & Moore, 2005). It is important to note that this research is not about the uncertainty of spatial data, but the spatial nature of data uncertainty.

Methods and Data

The key variable used in this research is the Global Non-response Rate (GNR) of the National Household Survey 2011 using Dissemination Areas as the unit of analysis. GNR combines complete non-response (household) and partial non-response (question) into a single rate, and is used as an indicator of data quality (Canada, 2011b). Smaller values of GNR indicate a lower risk of inaccuracy. According to NHS user guide in 2011 (Canada, 2011a), products of any geographic areas with GNR greater or equal to 50% is not released due to the high level of error which exceed an acceptable threshold. As a result, DAs without GNR are excluded from the analysis and mapping processes.

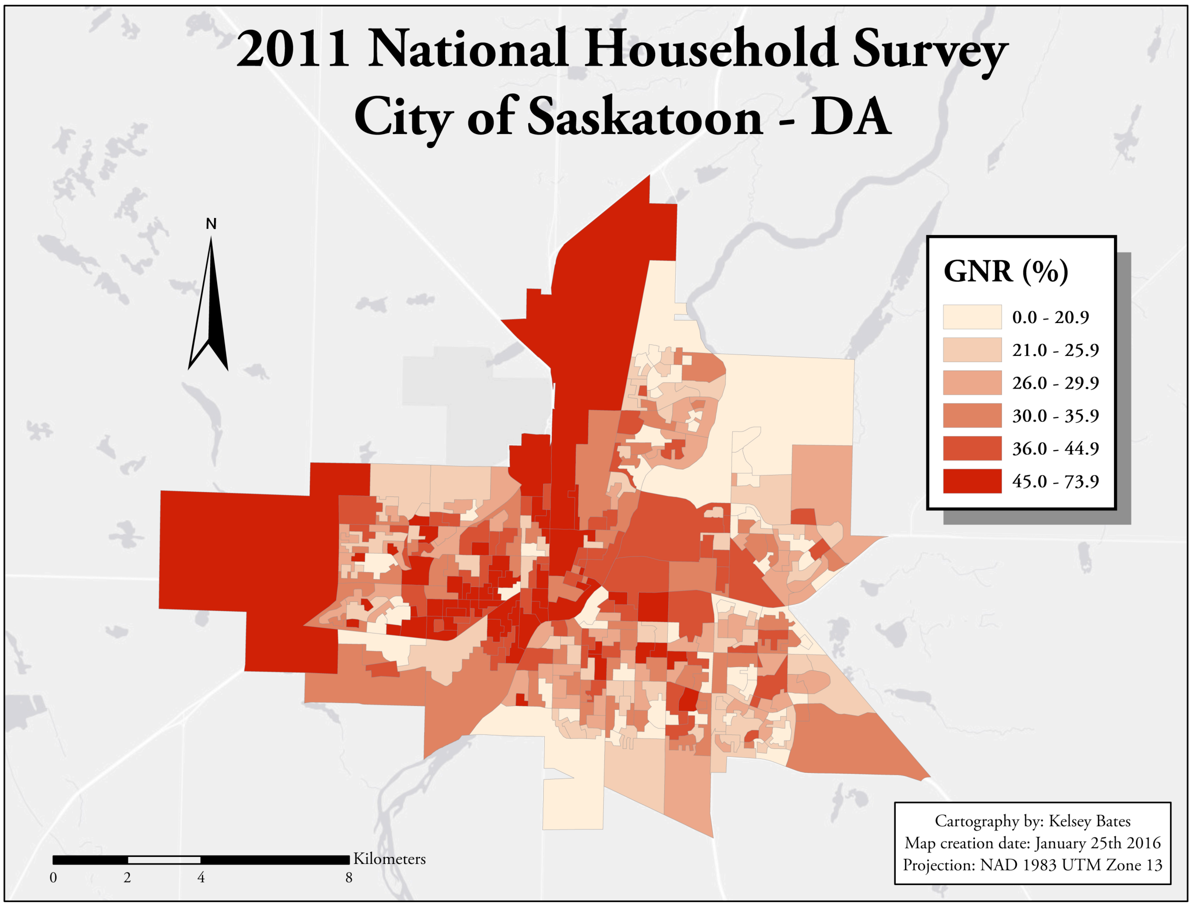
Our earlier work (2013 and 2014) was restricted to much larger units of analysis than the DA. This restriction was the result of substantial delays in the release of NHS data at all scales of analysis and for all non-spatial variables. It is our assumption that these delays resulted from the high rate of non-response and the resulting need to increase the sampling rate for many geographic areas. Furthermore, as mentioned above, data were not released for places (in this case defined by individual units of analysis) with a GNR greater than 50%. This policy results in a disappointing patchwork of spatial and non-spatial data. In this study we focus on urban areas, as they are more densely populated. This is substantiated in our earlier work:

> "In remote areas, all the households were invited to participate the NHS 2011 survey (Canada, 2011a). Consequently, it is not surprising that Northwest, Nunavut, and Yukon territories have relatively lower GNR. In terms of urbanization, Peri-Urban and Rural geographies have higher GNR than Census Metropolitan geographies; this pattern is true for most provinces (figure 1). In other words, densely populated metropolitan areas have higher response rates than more sparsely populated non-metropolitan areas." (2014)

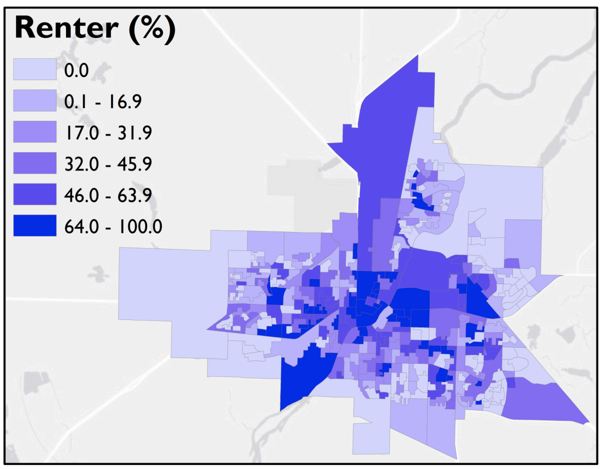
In order to summarize, describe, and explain GNR in Canadian cities we have mapped thirteen Canadian cities in nine provinces. Unfortunately, we did not feel comfortable including St. John's NL as the number of DAs did not provide enough statistical power to conduct the analysis used to relate non-GNR variables from the NHS to GNR. The maps presented here are a selection of all maps produced. Additionally, we include a selection of non-GNR variable maps that we think show interesting patterns related and not related to the distribution of GNR.

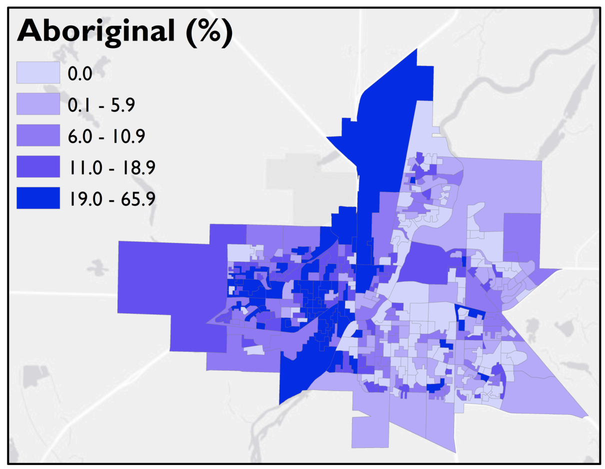
Results

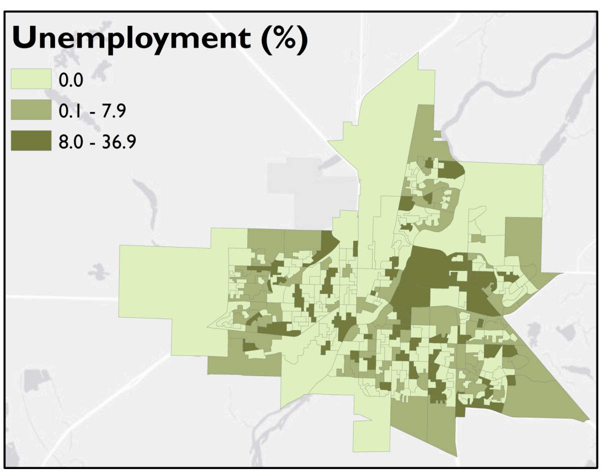
We present several sample maps from our complete atlas (unpublished) of GNR and significant predictor variables. Predictor variable maps represent variables that emerged as significant predictors from a spatial regression analysis of each city. We created both city specific models as well as a "national model" that used the same four predictors for all cities. These regression model results are not presented here as they are part of a larger publication and research project. Mapped patterns suggest that there is intra-urban clustering of GNR in all Canadian cities (see appendix, following references) and that these patterns match social, economic, housing status, and aboriginal variables. Interestingly, there are inconsistencies in the patterns when multiple cities are examined. In some cities (Calgary) areas with more aboriginal residents have higher GNR, the same pattern is not present (or as strong) in Saskatoon. Not surprising, population transience is an important predictor of non-response. For instance, in Vancouver's downtown and, more so, lower eastside, GNR is high, as would be expected for a population that is perhaps less likely to reside at the same address during the time of the survey, making compliance more difficult.

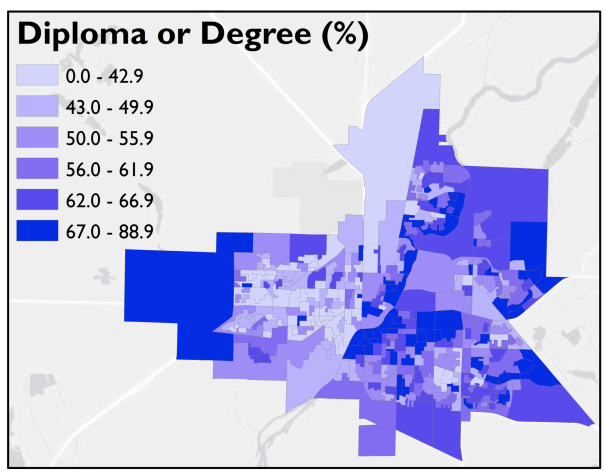


2011 National Household Survey City of Saskatoon - DA









Kilometers
0 4 8 16

Cartography by: Kelsey Bates
Map creation date: April 5, 2016
Projection: NAD 1983 UTM Zone 13N

Conclusions

Canada's National Household Survey was never promoted as an improvement on earlier (and subsequently reinstated) survey tools. It was initially promoted as a means to increase the freedom of Canadians to participate or not. The maps presented here suggest that the NHS should be used with caution. Even when oversampling resulted in the return of an adequate number of survey responses researchers should be concerned. If an underlying non-spatial pattern of non-response exists (something we believe is likely), then oversampling will simply ensure that more of the population in the biased sub-sample will return the survey. Our primary finding is that while sweeping generalizations are problematic, there are patterns that deserve additional attention. Additionally, an important consideration is the implication of using data from an unreliable tool to predict that tool's unreliability; such conclusions are troublingly circular and deserve further attention.

References

Bell, S., Jones, M., & Wei, T. (2013). Communicating with uncertain data: A search for spatial and geographic patterns Paper presented at the COSIT 2013 Workshop on Visually Supported Reasoning with Uncertainty, Scarborough, North Yorkshire, UK.

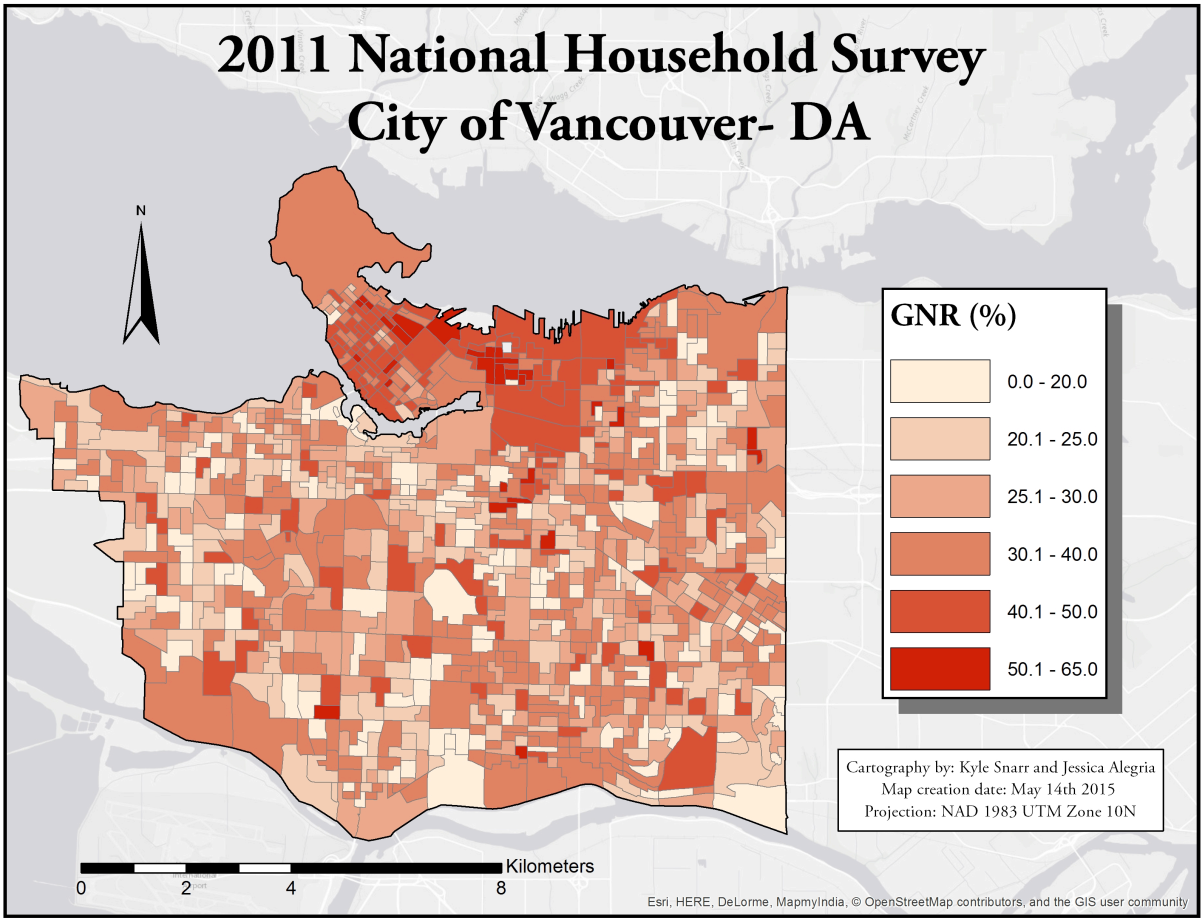
Bell, S., & Wei, T. (2014). Mapping the Spatial Pattern of the Uncertain Data: A comparison of Global Non-response Rate (GNR) between Metropolitan and Non-Metropolitan Area. *Spatial Knowledge and Information-Canada*.

Canada, S. (2011a). National Household Survey User Guide, 2011. Retrieved October 24th, 2013, from http://www12.statcan.gc.ca/nhs-enm/2011/ref/nhs-enm_guide/99-001- x2011001-eng.pdf.

Canada, S. (2011b). NHS Profile, 2011 – About the data. Retrieved October 24th, 2013, from http://www12.statcan.gc.ca/nhs-enm/2011/dp-pd/prof/help-aide/aboutdataaproposdonnees.cfm?Lang=E.

Kardos, J., Benwell, G., & Moore, A. (2005). The visualisation of uncertainty for spatially referenced census data using hierarchical tessellations. Transactions in GIS, 9(1), 19-34.

Appendix A: GNR for Vancouver (other cities available, file size constraints resulted in few maps included with submission)



Appendix B: Predictor Variables for Vancouver