How can communities be spatially defined to better reveal community-level happiness?

Hugh Shiplett, John F. Helliwell & Christopher P. Barrington-Leigh

Extended Abstract

Local, national, and global interest in subjective well-being has been growing rapidly over the past twenty-five years. This is now starting to be matched by official collection of relevant happiness data. Of the three general types of measure --- life evaluations, positive affect and negative affect --- the former has been found to best capture the overall quality of life in a community or country. Thus, while the OECD has recommended a substantial slate of measures of subjective well-being (OECD 2013), the slate is anchored by a core question asking people how satisfied they currently are with their lives as a whole, on a scale running from 0 to 10. Although most OECD countries and some elsewhere are now collecting measures of the satisfaction with life (SWL) on a national basis, only a few have samples sufficiently large to deliver robust measures at the community level (Brezzi & Ramirez 2016, Lu et al. 2015, OECD 2015).

Since it will be some time before these data become less scarce, it is important to consider in the meantime how they can most efficiently be used to measure community-level quality of life, to explain these differences, and to suggest ways to make community-level improvements in the quality of life. To provide good measures requires sufficient sample size, and with small communities aggregated so as to reflect natural boundaries between distinct communities. To explain the resulting differences requires that the chosen boundaries must match census boundaries, so that potential explanatory variables can be drawn as much as possible from administrative data. Explanation also requires a sufficient number of resulting communities to provide statistically robust explanations. Finally, to provide the basis for considering how policies have and could affect the quality of life, it is best if the community boundaries also match as closely as possible with the various local and regional administrative boundaries.

There are a number of empirical and practical reasons for wanting high geographic resolution in accounts of well-being. These have mostly to do with the general finding that the drivers and

supports of well-being have strong local components. Because dimensions of the social context have proven such strong predictors of life satisfaction, variation across local communities is likely to exist for both sides of the equation. Trust in neighbours and sense of belonging to one's local community, for instance, exhibit variation and predict life satisfaction beyond their influence on other measured community and individual characteristics (Helliwell and Wang, 2011).

Social norm- and reference-setting, along with other spatial spillovers, are also likely to occur at small geographic scales. Analyses of income reference effects, which are generally found to be strong compared with the more direct, positive effect of income, have been carried out at census tract scales and smaller.

Inequality of well-being, which we measure by the standard deviation of life evaluations within each community, has been argued to provide a broader and more relevant measure of inequality than does inequality in the distribution of income (Goff et al 2016). We find evidence in our new data as well that inequality of life satisfaction varies a lot among communities, and quite differently from income inequality.

Numerous other social, economic, and demographic determinants including ethnicity, housing type and housing costs, access to services, and so on all vary locally and have natural implications for life satisfaction. This is particularly important given the high but still growing urbanization rates in developed and developing countries, because aspects of the built and social environment in cities vary greatly over small distances.

Studies which average spatially over all these sources of variation will tend to underestimate their importance. This lack of variation, combined with the resulting drop in the number of communities under study, renders it difficult or impossible to identify the underlying relationships. Our procedures are especially designed to avoid these difficulties by respecting natural boundaries and while achieving sufficiently large and equal sample sizes.

The usefulness of high-resolution SWL datasets is also ultimately linked to the availability of other data. The smallest geographic scales at which census data are compiled represent natural targets for analysing SWL, and there is now a growing wealth of spatial analytic data from government and other sources, which can be brought to bear on the task of understanding the determinants of life satisfaction.

On the other hand, life satisfaction is particularly challenging to measure at small geographic scales. It has a large idiosyncratic component which is manifested as unexplained variance in most modeling efforts. As a result, for reasons of cost, there are relatively few datasets available with local sampling. National surveys tend anyway to stratify at larger spatial scales, and very large samples must be accumulated in order to have both full coverage and the ability to statistically discriminate at fine spatial scales. Indeed, until recently there were few examples even of sub-national SWL datasets, with the United Kingdom and Canada having perhaps the largest samples of SWL data collected as part of the official system of integrated surveys. This situation is starting to improve as more countries initiate special surveys designed to be integrated with available samples of administrative data. Full-country coverage with survey samples large enough to provide a fine-grained geographic breakdown is still rare, but is very likely to become more common in the next few years. This helps to sharpen our interest is trying to find the most efficient ways to choose geographic units so as to make best use of the available data.

This paper makes two contributions. One is a methodology for choosing and validating community regions for aggregating spatially identified data, using both data-driven and exogenous criteria. The other is a preliminary analysis of the features and drivers of life satisfaction and its distribution among geographic communities in Canada. Our dataset will be available for public use as Supplementary Material to this article.

Two major Canadian surveys – the Canadian Community Health Survey and the General Social Survey – have now been asking the same consistently worded and scaled life satisfaction question since 2010, providing a national sample now exceeding 400,000 respondents. In our first analysis we shall use all of these data to consider various ways of defining community

boundaries to get the sharpest delineation of the average satisfaction with life in each region. After some experimentation, we have found that by using a target sample size of 300 observations, with a lower limit of 250, by respecting census tract boundaries, and aggregating nearby communities using both built and natural community structure to group together neighbouring populations, provides us with a set of 1215 community-level geographic entities together covering all of Canada's geography. These conditions ensure some reasonable attainment of our objectives, and deliver a remarkably high degree of differentiation.

In this paper we also present the first estimates of life satisfaction in these communities, along with a substantial range of census data covering exactly the same communities. We also perform a number of diagnostic tests to ensure that we have found a near-optimal trade-off between sample size and number of communities, and to test whether our deliberate use of natural, built and administrative structures to define borders gives us a significantly better trade-off than we could have achieved using more arbitrary methods to define community boundaries.

A good data set, given our objectives, will have community-level SWL averages with standard errors of estimate that are small relative to the differences in means across communities. The data should also have high coherence, in the sense that correlations with census-level variables that explain a significant part of the total inter-community variance.

The community-level means are quite tightly estimated in our data, with standard errors only about 1% of the mean. The differences among the community level averages are large, ranging from 7.0 to 8.9, more than twenty standard errors. Comparing the average values of a number of variables in the top and bottom quintiles, we show sharp differences in some variables, e.g. community belonging, population density, inequality of well-being, and time in residence, and either slight or no differences in other variables often found important at the individual level, e.g. income, unemployment, and education. Together, these variables account for 50% of the large variation in life satisfaction among communities.

Our experiments with alternative, less theory-driven selection of boundaries and sample sizes indicate that we have found a combination of sample size, number of communities and reliance

on important political and geographic boundaries that gives us a high degree of explainable variance and good power to determine some of the key correlates of community-level happiness. These data are useful on their own to study community-level differences. They can also be used to provide contextual variables for multi-level modelling using individual life satisfaction data set in a community context. We also hope that it will be possible to apply and test our spatial analysis in other countries.

References

Brezzi, M., & Diaz Ramirez, M. (2016), Building subjective well-being indicators at the subnational level: A preliminary assessment in OECD regions, *OECD Regional Development Working Papers*, No. 2016/03, OECD Publishing, Paris.
DOI: http://dx.doi.org/10.1787/5jm2hhcjftvh-en

Helliwell, J. F., & Wang, S. (2011). Trust and Wellbeing. *International Journal of Wellbeing*, *1*(1).

Lu, C., Schellenberg, G., Hou, F., & Helliwell, J. F. (2015). How's Life in the City? Life Satisfaction across Census Metropolitan Areas and Economic Regions in Canada. *Economic Insights.* Statistics Canada: Ottawa, ON.

OECD (2013), *OECD Guidelines on Measuring Subjective Well-being*, OECD Publishing, Paris.
DOI: http://dx.doi.org/10.1787/9789264191655-en

OECD (2015), *Measuring Well-being in Mexican States*, OECD Publishing, Paris.DOI: