

USING STATISTICS CANADA DATA TO MAP CANADA'S POPULATION

Julia Siemer, University of Regina, Saskatchewan, Canada

ABSTRACT

Statistics Canada, Canada's national statistics agency, offers a suite of spatial files for mapping and analysis of its various population data products. The following article showcases possibilities and shortfalls of the existing spatial files for mapping population data, and provides an overview of the structure of the available boundary files from the regional to the dissemination block level. Due to Canada's highly dispersed population, mapping its distribution and density can be challenging. Common mapping techniques such as the choropleth method are suitable only for mapping spatially high resolution data such as data at the dissemination area level. To allow for mapping of population data at less detailed levels such as census divisions or provinces, Statistics Canada has created a so-called ecumene boundary file which outlines the inhabited area of Canada and can be used to more accurately visualize Canada's population distribution and density.

ZUSAMMENFASSUNG

Statistics Canada, Kanadas nationale Statistikbehörde, bietet eine Reihe räumlicher Basisdaten zur Kartenerstellung und räumlichen Analyse ihrer Bevölkerungsdaten an. Der folgende Bericht zeigt die Möglichkeiten und Einschränkung der angebotenen räumlichen Daten für die Bevölkerungskartographie auf und stellt die vorhandenen Verwaltungsgrenzdateien bis zur kleinsten Einheit, dem so genannten ‚dissemination block‘, in ihrer Struktur vor. Die räumlich sehr ungleiche Verteilung der Bevölkerung Kanadas stellt eine besondere Herausforderung bei der kartographischen Visualisierung dar. Üblicherweise angewendete Methoden wie die Choroplethen-Methode sind im Falle Kanadas streng genommen nur zur Visualisierung kleinräumiger Bevölkerungsdichte geeignet. Um die Darstellung der Bevölkerungsdichte und -verteilung auch für größere Raumeinheiten, z. B. den so genannten ‚census divisions‘ oder die Provinzen, zu ermöglichen, veröffentlicht Statistics Canada einen Datensatz, der die Abgrenzung der Ökumene, also der bewohnten Fläche, des Landes veranschaulicht. Diese Datei ermöglicht es dem Nutzer die Bevölkerungsdichte und -verteilung des Landes angemessen und detailliert zu visualisieren.

1 INTRODUCTION

For the last few years, Statistics Canada, Canada's national statistical agency, has been offering many of their thematic data sets and other geographic products online, free of charge. This approach, adopted by the government in 2006 (Statistics Canada 2006) allows researchers (and the general public alike) relatively easy, unbureaucratic and user-friendly access to selected census data and digital boundary files for mapping and analysis.

The collection of statistical data is a federal responsibility in Canada and Statistics Canada produces statistics on the federal level as well as for the provinces and territories. The agency conducts a nation-wide census every 5 years¹ and receives (due to a legal requirement to participate in the survey) typically high response rates from its citizens. For the most recent census, held in May 2016, the overall national response rate was 98.4%, whereas the response rate for the 25% of the population who received the long-form of the census survey, was 97.8% (Statistics Canada 2016)². These rates were particularly high compared to the previous census in 2011 due to the re-introduction of the so-called long-form survey in 2016. In the 2011 census, the voluntary National Household Survey (NHS) replaced the previously mandatory long-form questionnaire of the survey and received a relatively low response rate of 69%. The overall response rate for the mandatory short-form of the 2011 census was at a considerably higher rate of 97% (Statistics Canada 2011a).

Population census data, in addition to other survey data collected by Statistics Canada (e.g., Agricultural Census, Labour Market Survey) and various indices (e.g., economic and health indicators) are considered the only source of reliable and detailed nationwide thematic data for planning purposes in Canada.

¹ The United Nations recommends that a census be held at least every ten years.

² Over the years, most census surveys have consisted of a short-form questionnaire, with six to ten basic demographic questions, which has been distributed to about 70% to 75% of the population and a long-form questionnaire, which has varied in length and has typically been distributed to 25% to 30% of the population. The 2016 long-form questionnaire included 70 questions, and was distributed to 30% of the population.

2 CANADA'S STANDARD GEOGRAPHICAL CLASSIFICATION (SGC)

All Statistics Canada data are based on the official Standard Geographical Classification (SGC) for geographic and statistical units and their names. Established in the early 1960s, the SGC has been updated over the years to accommodate changes in the population distribution at each census.

The SGC is based on a spatial classification of a set of discrete units that are mutually exclusive and, in total, cover the entire country. The SGC defines four levels of geographical units:

- » *Geographical region of Canada*
- » *Province or territory*
- » *Census division*
- » *Census subdivision*

Table 1 and 2 show the spatial and code structure of these units.

Region	Code (single digit)	Province or Territory	SGC Code (two digits)	Alpha Code (two digits)	Abbreviations
Atlantic	1	Newfoundland and Labrador	10	NL	N.L.
		Prince Edward Island	11	PE	P.E.I.
		Nova Scotia	12	NS	N.S.
		New Brunswick	13	NB	N.B.
Quebec	2	Quebec	24	QC	Que.
Ontario	3	Ontario	35	ON	Ont.
Prairies	4	Manitoba	46	MB	Man.
		Saskatchewan	47	SK	Sask.
		Alberta	48	AB	Alta.
British Columbia	5	British Columbia	59	BC	B.C.
Territories	6	Yukon	60	YT	Y.T.
		Northwest Territories	61	NT	N.W.T.
		Nunavut	62	NU	Nvt.

Table 1: Codes for geographic regions, and provinces and territories

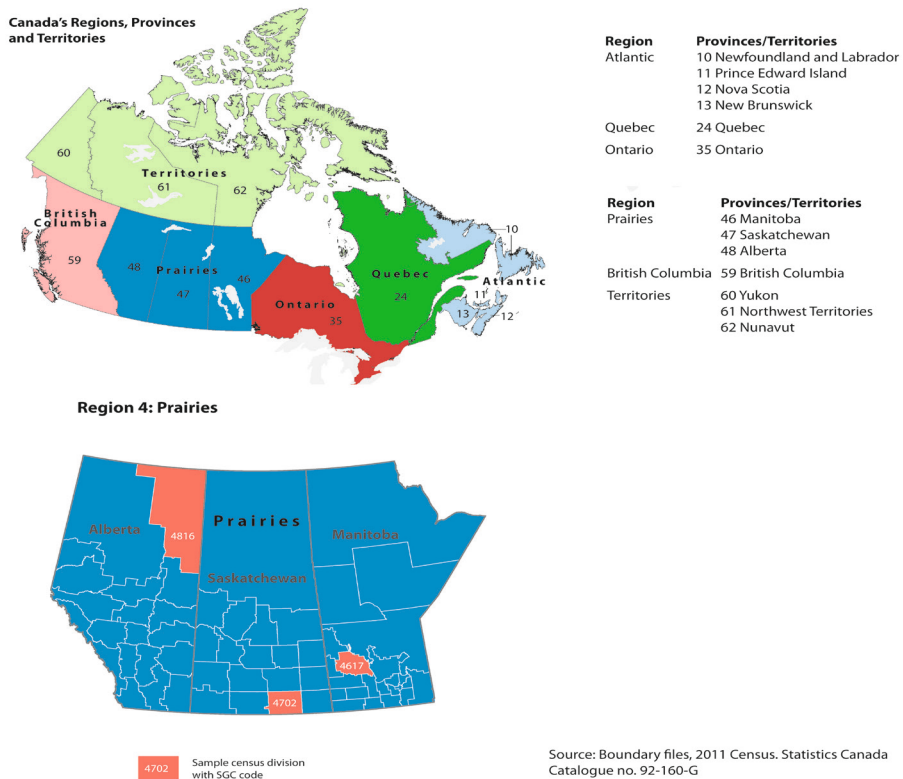
SGC Code (four digits)	Census Division (CD)	SGC Code (ten digits)	Census Subdivision (CSD)
10 01–62 08	Generic name (e.g., Division No. 1) or local geographic name	10 01 101 999–62 08 047 000	Local geographic names

Table 2: Sample codes for census divisions (CD) and census subdivisions (CSD)

In addition to these four geographic units, the SGC identifies several other geographic and statistical units, which follow the same general coding system. These units are: census metropolitan areas, census agglomerations and census influenced zones; consolidated census subdivisions, economic regions, agricultural regions, and smaller statistical areas such as census tracts, dissemination areas and dissemination blocks.

Map 1 illustrates the geographic distribution of the regions, provinces and territories of Canada and the next lower level of statistical areas (census divisions) of the Prairie region.

Based on the nature of the classification, each defined area is assigned a unique identifier (e.g., 47006: Census Division 6 in Saskatchewan). The design of the classification allows for any future changes of spatial units to be accommodated. The system supports the amalgamation of two or more areal units as well as possible splits of areas into two or more newly defined areas.



Map 1: Canada's statistical regions, provinces/territories, and Region 4: Prairies with census divisions 4617, 4702 and 4816 highlighted

3 GEOGRAPHIC BOUNDARY FILES

Statistics Canada makes a selection of boundary files available on their website (www.statcan.gc.ca). These files can be used with geographic information systems (GIS) and other mapping software. All files are available in the two official languages of Canada (English/French) and in three different file formats (ArcGIS (.shp), Geographic Markup Language (.gml), and MapInfo (.tab)). Boundary files are updated for each census, typically necessary due to changes in enumeration areas or the creation of new spatial units. In addition, updated spatial files for intercensal years are made available if necessary. For the 2016 census a first edition of boundary and other spatial files will be released in November 2016. Statistics Canada offers two types of spatial files of its areal units: cartographic boundary files, which are optimized for cartographic visualization (i.e., geometrically generalized) and digital boundary files, which describe the legal definition of the areal units. Particularly for coastal areas, the latter file type has special relevance as it includes areas beyond the Canadian landmass.

In addition to standard boundary files, Statistics Canada also creates files delineating the country's population and agricultural ecumene (i.e., the inhabited areas of the country). Ecumene boundary files can be used to map Canada's population distribution more accurately rather than generalizing distribution and density over entire statistical units. This is particularly useful as Canada exhibits a very uneven distribution of its population.

The ecumene files support both dot distribution and choropleth mapping techniques. Dot distribution mapping can be utilized when raw totals of a specific phenomenon (e.g., population counts) are collected for enumeration units and the cartographer wishes to show that the underlying phenomenon's nature is not uniform throughout the spatial unit (Slocum et al. 2009). Dot distribution maps rely on ancillary data which support the decision of placement of dots in locations where that particular phenomenon is most likely to occur. For dot distribution maps of population, created semi-automatically by GIS software from Statistic Canada data, the ecumene boundary file functions as a simple mask that restricts the placement of dots to populated areas and prohibits the placement of dots in unpopulated areas of the country.

Choropleth maps on the other hand visualize relative or standardized phenomena, such as population density for statistical or administrative areas assuming an even distribution throughout the area. Due to the availability of statistical data, the choropleth technique is probably the most commonly used thematic mapping technique. This technique is particularly appropriate when characteristics of a phenomenon change abruptly at boundaries and are uniform throughout the area.

(See MacEachren 1992, 1994 and Slocum et al. 2009 for more information on the relationship of the nature of phenomena and suitable mapping techniques.) Choropleth maps can also be an appropriate choice to display ‘typical’ values for areal units, even when the phenomenon does not change abruptly at the boundaries (Slocum et al. 2009). For population maps the choropleth technique is appropriate only if the average ratio (i. e., population density) can be considered a representative value for the entire area. Due to the dispersed distribution of Canada’s population, average density values for statistical units can only be considered typical for smaller units such as dissemination blocks or areas. Population density of larger areas such as census divisions or consolidated census subdivisions is not represented well by this technique. However, the Canadian population ecumene boundary file can be used to minimize this shortcoming by visualizing population density only for those parts of a statistical unit that have been defined as populated. Ideally, this results in a more accurate representation of the spatial distribution of population and its density.

Due to the nature of Canada’s population ecumene, which follows natural boundaries rather than administrative or statistical boundaries, it does not conform with SGC codes. The 2011 population ecumene boundary file defines the population ecumene as an area with a population density of 0.4 or more persons per square kilometre (see Map 2). For the purpose of creating the ecumene boundary file, population density was calculated per dissemination block. Each block then was classified as either being part of the ecumene (i. e., 0.4 persons/km² or higher) or being a non-ecumene block (i. e., less than 0.4 persons/km²) (Statistics Canada, 2012b).

As a result of the recommended use of this file for small scale mapping (1: 20,000,000 to 1: 25,000,000), the file displays a higher degree of geometric generalization than other boundary files and is not necessarily consistent with the positional accuracy of more detailed spatial files by Statistics Canada.

4 POPULATION DATA

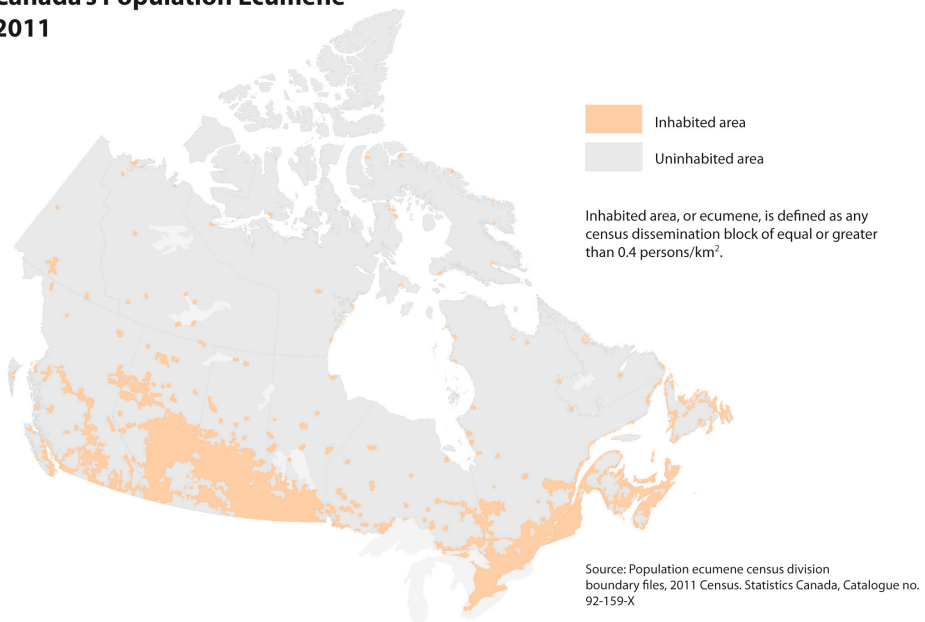
A first release of census data typically takes place in late Fall of a census year, followed by the release of more thematic data throughout the subsequent year. For the 2016 Census a first release of data (population and dwelling count) is announced for November 2016. In May 2017 a second data set (age and sex of the population, and the type of dwelling) will be released. The remaining data sets will be released until November 2017.

Summarized tables and analyses regarding the different characteristics and parameters of the Canadian population will be published on Statistics Canada's website. Complete data sets of basic demographic data, from national level down to dissemination areas (DA) are available through 'GeoSuite', a tool to query, save and download tabular census data. For the population maps of this study, census data from 2011 was used to illustrate mapping options and limitations of the Statistics Canada files. All maps are based on absolute population counts or population densities (persons/km²) per statistical unit. Other population characteristics, ratios or densities are possible to map in a similar fashion using the same general approach.

5 MAPPING THE POPULATION OF THE CANADIAN PRAIRIES

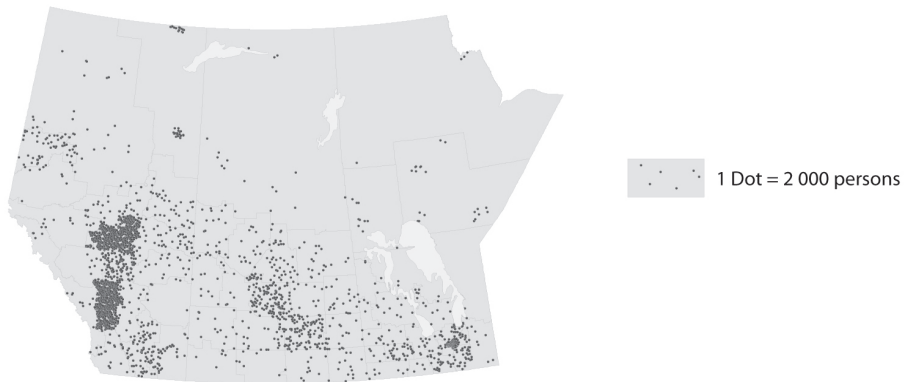
The following maps demonstrate the use of Statistics Canada spatial file for mapping population distribution and density. Map 3 (*Prairies: Population Distribution 2011*) illustrates how the population ecumene file is used as ancillary data or a mask to place dots – each dot represents 2000 persons – inside the inhabited areas of Canada. When deciding the placement of dots not only the location inside or outside certain areas is critical but also the dot value (i.e., how many people are represented per dot), the pattern of placement (e.g., uniform, random, or geographically weighted) and dot size (diameter of a single dot) are important parameters that affect the quality of the dot map (see Slocum et al. 2009 for more information of dot mapping). For this population dot map of the Prairie region, default settings for random placement in ESRI’s ArcMap were applied.

Canada’s Population Ecumene 2011



Map 2: Canada’s population ecumene in 2011

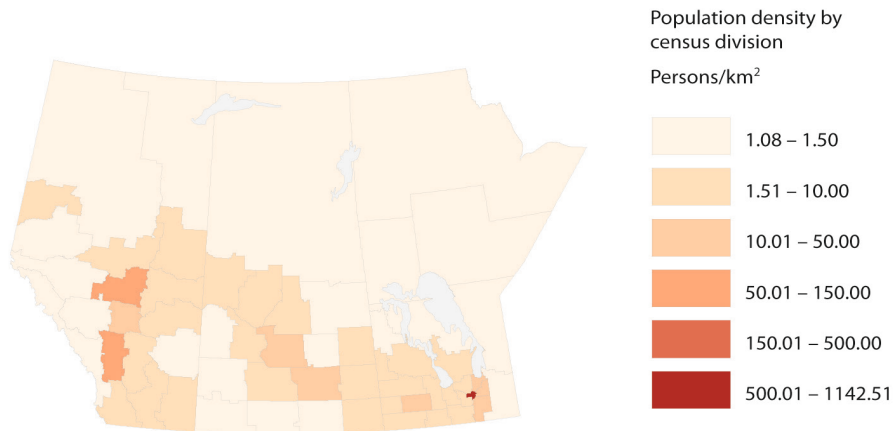
Prairies: Population Distribution 2011



Source: Population ecumene census division boundary files, 2011 Census. Statistics Canada, Catalogue no. 92-159-X; GeoSuite, 2011 Census. Statistics Canada, Catalogue no. 92-150-X.

Map 3: Dot distribution map of the population of the Prairies in 2011

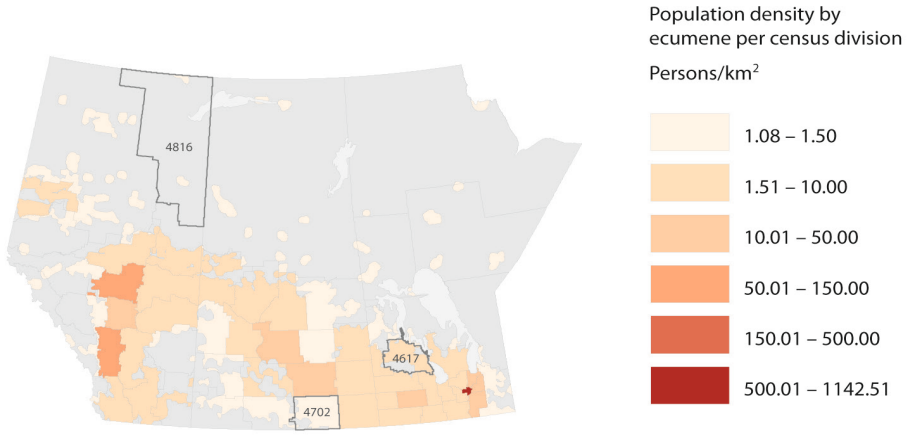
Prairies: Population Density 2011



Source: Population ecumene census division boundary files, 2011 Census. Statistics Canada, Catalogue no. 92-159-X; GeoSuite, 2011 Census. Statistics Canada, Catalogue no. 92-150-X.

Map 4: Choropleth map of the Prairie's population density in 2011

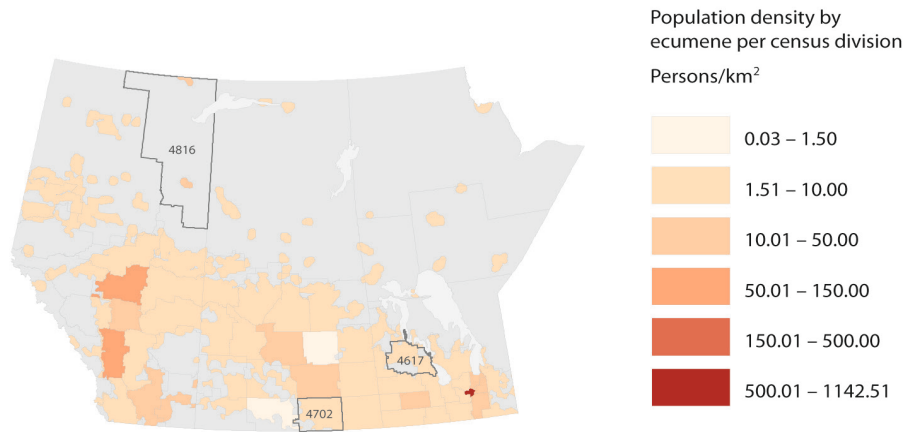
Prairies: Population Density of the Ecumene 2011



Source: Population ecumene census division boundary files, 2011 Census. Statistics Canada, Catalogue no. 92-159-X; GeoSuite, 2011 Census. Statistics Canada, Catalogue no. 92-150-X.

Map 5: Incorrect choropleth map with ecumene mask of the Prairie's population density in 2011

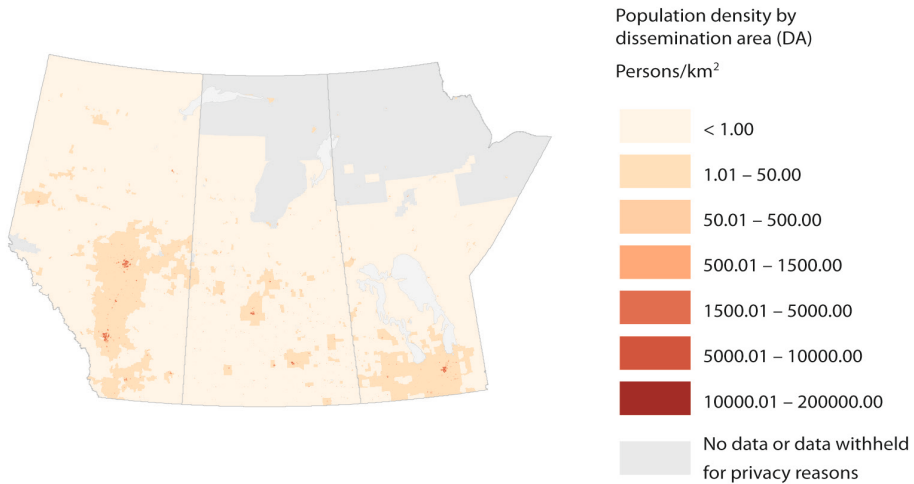
Prairies: Population Density of the Ecumene 2011



Source: Population ecumene census division boundary files, 2011 Census. Statistics Canada, Catalogue no. 92-159-X; GeoSuite, 2011 Census. Statistics Canada, Catalogue no. 92-150-X.

Map 6: Correct choropleth map with ecumene mask of the Prairie's population density in 2011

Prairies: Population Density by DA 2011



Source: Population ecumene census division boundary files, 2011 Census. Statistics Canada, Catalogue no. 92-159-X; GeoSuite, 2011 Census. Statistics Canada, Catalogue no. 92-150-X.

Map 7: Choropleth map of the population density of the Prairies by dissemination block in 2011

Maps 4 to 7 apply the choropleth method to visualize population density. Using this method reveals the limitations of the ecumene boundary file. Map 4 displays highly generalized population density (persons/km²) for the census division level of the Prairie region. Any internal variation of population distribution inside a census division is lost in this map. In an attempt to limit the display of population density to the ecumene of the Prairies (similar to the dot distribution mapping approach of map 3) the boundary mask exposes the restrictions of the file. By simply overlaying the choropleth map with the boundary mask an incorrect map is created with misleading average population density values (map 5). Naturally, populated areas of a spatial unit (e.g., census divisions) are smaller in size than the overall unit and need to be considered using their correct size (km²) when density is calculated. While this can be ignored in many cases due to minimal discrepancies, it particularly affects the population densities for the northern census divisions (e.g., Division 16 in Alberta, which includes the population centre Fort McMurray with a considerably higher population density than the surrounding area). Table 3 illustrates the differences in population density, sizes of census divisions and their populated areas for three selected divisions in the Prairie region.

Census Division	CD Code	Area (km ²)	Population	Population Density (persons/km ²)	Ecumene (km ²)	Population Density of Ecumene (persons/km ²)
Division 16	4816	104,990.43	67,516	0.64	1,974.78	34.19
Division 2	4702	17,309.07	22,266	1.29	14,470.60	1.54
Division 17	4617	14,499.98	22,208	1.53	10,012.69	2.22

Table 3: Population density 2011 for several Census Divisions of the Prairie region

Rather than simply masking unpopulated areas of census divisions with the ecumene boundary file, densities for populated areas require recalculation of population density. Map 6 shows a correct application of the ecumene boundary file in combination with population density, calculated using actual inhabited area sizes.

While the level of generalization of the ecumene boundary file does not match the level of detail of other boundary files (e.g., census tracts) differences in area sizes have only minimal impact on density rates for these smaller areal units. The use of the ecumene file as a mask over an existing population density map at the census tract level is acceptable but creates visually unpleasant maps.

To illustrate the actual population distribution of Canada more realistically, a smaller areal unit (e.g., the lowest level of publically available boundary and population data at the dissemination area (DA) level) should be used to map population distribution if the use of an ecumene boundary file is considered to be misleading. A sample of such a map is shown in map 7. However, due to the number of records in the data set (10,357 DAs for the Prairie region alone) processing and creating of such a map can be challenging and time consuming. Furthermore, calculated densities can be unrealistically high due to the very small size of some dissemination areas. Compared to other mapping techniques such as dasymetric mapping (a more demanding method which relies on additional ancillary data sets and often uses remotely sensed data) these additional steps can be considered as minimal. (See Slocum et al. 2009 for more information on dasymetric mapping.)

6 CONCLUSION

The maps created for this study demonstrate the value Statistics Canada's boundary files have for mapping and analysis of population data, particularly for dot distribution mapping. However, the study also reveals a serious shortfall of the ecumene boundary file which can lead to incorrect data visualization of choropleth maps at lower spatial resolution. This shortfall can be corrected by manually reassessing population density based on updated area calculations. This boundary file, on the other hand, allows for excellent visualization of Canada's dispersed population and can create a memorable impression of the phenomenon for the reader. Relatively few adjustments of the existing file structure are required to enable proper use of the boundary file which, in return, results in more accurate population density maps.

REFERENCES

- MACEACHREN, A. M. (1992): Visualizing Uncertain Information. In: Cartographic Perspectives. No. 13/Fall 1992, pp. 10–19.
- MACEACHREN, A. M. (1994): Some Truth with Maps. A Primer on Design and Symbolization. Washington, DC.
- SLOCUM, T. A.; McMaster, R. B.; Kessler, F. C.; Howard, H. H. (2009): Thematic Cartography and Geovisualization. 3rd Edition. Upper Saddle River: Pearson-Prentice Hall.
- STATISTICS CANADA (2006): Access to census information (abrufbar unter URL: <http://www12.statcan.gc.ca/census-recensement/2006/ref/acc-info-eng.cfm#dli>, Stand: 26.10.2016).
- STATISTICS CANADA (2011a): RDC proposals requesting Census or National Household Survey (NHS) (abrufbar unter URL: <http://www.statcan.gc.ca/eng/rdc/cennhs>, Stand: 26.10.2016).
- STATISTICS CANADA (2011b): Standard Geographical Classification (SGC) 2011. Volume 1. The Classification. Catalogue No. 12–571-X. Ottawa.
- STATISTICS CANADA (2011c): GeoSuite, Reference Guide, 2011 Census. Catalogue No. 92–150-G. Ottawa.
- STATISTICS CANADA (2012a): Boundary Files, Reference Guide, Second edition, 2011 Census. Catalogue No. 92–160-G. Ottawa.
- STATISTICS CANADA (2012b): Population Ecumene Census Division Cartographic Boundary File, Reference Guide, 2011 Census. Catalogue No. 92–159-G. Ottawa.
- STATISTICS CANADA (2016): 2016 Census of Population response rates (abrufbar unter URL: <http://www12.statcan.gc.ca/census-recensement/2016/ref/response-rates-eng.cfm>, Stand: 26.10.2016).