

Catalogue no. 17-20-0002  
ISBN 978-0-660-43023-2

# Canadian Social Environment Typology User Guide

by Rajendra Subedi, Nicole Aitken, and Lawson Greenberg

Release date: May 9, 2022



Statistics  
Canada

Statistique  
Canada

Canada

---

## How to obtain more information

For information about this product or the wide range of services and data available from Statistics Canada, visit our website, [www.statcan.gc.ca](http://www.statcan.gc.ca).

You can also contact us by

**Email at** [infostats@statcan.gc.ca](mailto:infostats@statcan.gc.ca)

**Telephone**, from Monday to Friday, 8:30 a.m. to 4:30 p.m., at the following numbers:

- Statistical Information Service 1-800-263-1136
- National telecommunications device for the hearing impaired 1-800-363-7629
- Fax line 1-514-283-9350

## Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner. To this end, Statistics Canada has developed standards of service that its employees observe. To obtain a copy of these service standards, please contact Statistics Canada toll-free at 1-800-263-1136. The service standards are also published on [www.statcan.gc.ca](http://www.statcan.gc.ca) under "Contact us" > "[Standards of service to the public](#)."

## Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued co-operation and goodwill.

Published by authority of the Minister responsible for Statistics Canada

© Her Majesty the Queen in Right of Canada as represented by the Minister of Industry, 2022

All rights reserved. Use of this publication is governed by the Statistics Canada [Open Licence Agreement](#).

An [HTML version](#) is also available.

*Cette publication est aussi disponible en français.*

---

---

## Table of contents

<b>1. Purpose</b> .....	<b>4</b>
<b>2. Background</b> .....	<b>4</b>
<b>3. Uses of the Canadian Social Environment Typology (CanSET)</b> .....	<b>4</b>
<b>4. Data used to create the CanSET</b> .....	<b>5</b>
<b>5. Methodology</b> .....	<b>5</b>
5.1 Variable Selection .....	6
5.2 Number of Clusters .....	6
5.3 Cluster Names, Characteristics and Uses.....	7
5.3.1 Six clusters solution .....	7
5.3.2 Ten clusters solution.....	8
5.3.3 Twenty clusters solution.....	9
<b>6. How to link the CanSET with your data</b> .....	<b>11</b>
<b>7. Summary and Conclusion</b> .....	<b>11</b>
<b>8. References</b> .....	<b>12</b>
<b>Appendix A</b> .....	<b>13</b>

# Canadian Social Environment Typology User Guide

by **Rajendra Subedi, Nicole Aitken, and Lawson Greenberg**

## 1. Purpose

The purpose of this user guide is to define the concept of identical Dissemination Area (DA) clusters used in the Canadian Social Environment Typology (CanSET) and to give an overview of how the clusters can be used to explore DA level health and social inequalities. More detailed technical information on the formation of the social environment clusters can be found in the *Canadian Social Environment Typology: Methodology Guide* developed by the health inequality unit of the Centre for Population Health Data (CPHD).

## 2. Background

In recent years there has been a growing demand for relevant health information at neighbourhood level. However, due to the lack of an available classification approach, it has been difficult to compare and contrast small areas within a city or among the cities in Canada. To overcome this shortcoming, the health inequality unit of the Centre for Population Health Data (CPHD) at Statistics Canada collaborated with the Urban Public Health Network (UPHN) to develop a Dissemination Area (DA) classification approach called the Canadian Social Environment Typology (CanSET). A methodology guide has been already in place, which describes the technical and methodological process involved in the development of the CanSET (available upon request). The main objectives of this document is to give basic information on how the CanSET was developed and also to provide detail information on how to use the CanSET classification to understand area level inequalities.

The CanSET is a hierarchical clustering of DAs within Canadian Census Metropolitan Areas (CMA)<sup>1</sup> and Census Agglomerations (CA)<sup>2</sup>. The typology generated through this cluster analysis can be taken as a social unit of analysis that indicates the geographic distribution of different kinds of combinations of population characteristics throughout CMAs and CAs. In other word, each social environment cluster includes similar DAs scattered across Canada. Therefore, the CanSET is not defining neighbourhoods but defining areas of common social composition which may be entirely different than the neighbourhoods defined by different local authorities in Canada. The new CanSET data can be used to better understand health inequalities in relation to the composition of sub-city geographies in Canada.

The main objective of this user guide is to facilitate the reader on how to use the new social environment clusters to understand social and health inequalities in more populous areas in Canada. The social environment clusters are developed by using 30 different socioeconomic, demographic, and ethno-cultural variables from the 2016 Census aggregated at the DA level. All the variables for this product were taken directly from the 2016 Census of Population microdata except the population density variable, which was derived from Census of Population 2016 profile table.

## 3. Uses of the Canadian Social Environment Typology (CanSET)

The individual level socioeconomic, demographic and ethnocultural status are useful to understand health and social outcomes but are often not available in the most commonly used dataset like surveys or administrative data. Therefore, area based socioeconomic and demographic measures have been widely used to measure inequalities where the individual level information is unavailable or hard to reach. Area level analysis helps to understand whether living in a socioeconomically disadvantaged area gives additional health risk beyond the individual level socioeconomic status.

The new CanSET portrays the complex social composition of Canadian CMA and CA. Unlike the previous studies that focused only on marginalization or health inequalities, the CanSET is more comprehensive as it includes a variety of dimensions of urban social environment and covers all populous areas of Canada. This is the first

---

1. A CMA is a geographic unit with a total population of at least 100,000 of which 50,000 or more must live in the core based on adjusted data from the previous Census of Population Program.  
2. A CA is a geographic unit with a core population of at least 10,000 based on adjusted data from the previous Census of Population Program.

analysis to include all DAs from the CMAs and CAs simultaneously in a single typology rather than analyzing them separately for each CMA or CA. In addition, it includes the microdata from the Census of Population 2016 which covers all population of the study area. Therefore, the CanSET is very important for urban research as it allows cities to make comparison within themselves or with other cities and facilitates them to set benchmarks and track progress in combating health and social inequalities.

For instance, in light of the recent COVID-19 pandemic, the CanSET data can be used to understand which social environment cluster is performing better to minimize the rate of infection, hospitalization and mortality. It helps to understand which factors play important role in determining health behaviour and response strategies during a pandemic like the COVID-19. This understanding could help in the planning and allocating of resources that are required for specific types of communities.

## 4. Data used to create the CanSET

Data selected and analyzed for the CanSET was sourced from the Census of Population 2016 microdata for all the DAs within 35 CMAs and 117 CAs in Canada. The DA was used as a geographic unit of analysis. Statistics Canada defines a DA as a small, relatively stable geographic unit composed of one or more adjacent dissemination blocks with an average population of 400 to 700 persons (Statistics Canada, 2016). The DAs are the smallest geographic units for which Statistics Canada disseminates its Census data for public use. The DAs are easy to aggregate small geographic units as they can be combined to form other larger geographic units. Although there are 56,590 DAs in Canada (2016 Census Geography), only the 43,144 DAs that belong to the CMAs or CAs were included in the CanSET, mainly because of their high population densities, and also due to the availability of the variables of interest.

The DAs for which Census data (i.e., either short form or long form) were not released due to confidentiality or data quality issues were not included in the creation of the CanSET. The DAs associated with Indian reserves were excluded from the analysis because some Census questions were either not asked or the concepts were not applicable or they were incompletely enumerated.

**Table 1**  
**Dissemination Area (DA), average population and households by provinces and territories**

Province/Territory	Total number of DAs	Number of DAs in CanSET	Average population per DA	Average number of households per DA
Newfoundland and Labrador	1,073	451	613	259
Prince Edward Island	295	152	563	244
Nova Scotia	1,658	986	610	269
New Brunswick	1,454	817	568	250
Quebec	13,658	10,544	627	281
Ontario	20,160	17,325	695	271
Manitoba	2,183	1,416	642	259
Saskatchewan	2,474	1,120	630	258
Alberta	5,803	4,282	774	300
British Columbia	7,617	5,984	678	278
Yukon	67	33	855	357
Northwest Territories	98	34	575	211
Nunavut	50	0	NA	NA

Source: Canadian Census of Population 2016.

## 5. Methodology

The CanSET was created using hierarchical cluster analysis of DAs into three levels of nested social environment clusters. Cluster analysis attempts to assign observations to groups (clusters) based on their similarity or differences using a measure of statistical (i.e., Euclidian) distance from each other. Observations within each group are found to be similar to one another with respect to variables or attributes of interest. In other words, the goal is to group the observations into homogeneous and distinct clusters. A hierarchical cluster analysis method was applied for cluster analysis with a 'Fast Ward' option in JMP® 13 (SAS Institute Inc., 2016) analytical software. The Fast Ward method applies an algorithm that computes Ward's method more quickly for large size data, therefore this method was used considering the data size for this analysis. Details of the methodology can be found in the "Canadian Social Environment Typology: Methodology Guide", available upon request.

## 5.1 Variable Selection

The selection of variables was the result of a consultation with UPHN members. An extensive review of the past literature associated with neighbourhood typology was also performed before selecting the variables. Variables describing the demographic, socioeconomic, and ethnocultural determinants of health within the DAs across Canada were used in the clustering algorithm to group similar DAs. Numeric variables were selected that were reliable and readily available from the Census of population 2016. The variables chosen for creating the CanSET cover a wide range of subjects including: demographic structure (age structure, family size, etc.), socioeconomic status (income, education, labour market status, housing condition, etc.) and ethno-cultural background status (Aboriginal status, immigration status, visible minority status, etc.). Health-related variables were deliberately not used in the creation of the typology to make the typology equally applicable in other fields. General socioeconomic and demographic variables were derived directly from the short form Census, whereas specific variable that were not available in short form Census (e.g., level of education, immigration status, occupation etc.) were derived from the long form Census and weights were applied. Some variables were available at person level whereas others were available at census family, economic family or household level. A preliminary list of 93 diverse variables were selected from the microdata of the Census of Population 2016. However, after an extensive review of the variables and after a round of preliminary analysis, some variables that were strongly correlated with each other were excluded. In addition, some similar variables were grouped together that resulted into 30 final variables as shown in Appendix A. All the variables were aggregated at DA level geography and median values were used.

The variables selected for cluster analysis were measured on different scales, or on a common scale with differing variances. Therefore, they were standardized in order to mitigate the effect of these differences among the variables. All the 30 variables were standardized with mean 0 and variance 1 prior to performing the cluster analysis. Some variables with highly skewed distributions were still much skewed even after normalization. So, all variables were capped at their 99<sup>th</sup> percentile.

## 5.2 Number of Clusters

The optimal number of clusters were six, ten and twenty for the first, second and third level of hierarch respectively. The clusters are defined as the set of DAs that are similar in terms of the selected characteristics (variables). Davies-Bouldin Index (DBI) for hierarchical clustering was used to determine the optimal number of clusters nested in three hierarchical levels. The smaller value of the DBI means a better clustering solution. Using the DBI, three different optimal 'k' values (number of clusters) were determined.

The optimal 'k' values were 6 in the range of 2 to 9 clusters, 10 in the range of 10 to 19 clusters and 20 in the range of 20 to 30 clusters as the Davies-Bouldin index was the smallest for these solutions. Therefore, a nested hierarchy of six, ten and twenty clusters was created as outlined in Table 2.

**Table 2**  
**Number of optimal clusters in the CanSET by hierarchy**

Cluster levels	Total number of Dissemination Areas
<b>A</b>	<b>19,127</b>
A1	6,650
A11	6,650
A2	7,170
A21	2,600
A22	3,349
A23	1,221
A3	5,307
A31	3,295
A32	2,012
<b>B</b>	<b>10,914</b>
B1	7,787
B11	4,553
B12	1,183
B13	2,051

**Table 2**  
**Number of optimal clusters in the CanSET by hierarchy**

Cluster levels	Total number of Dissemination Areas
B2	3,127
B21	1,215
B22	1,912
<b>C</b>	<b>1,328</b>
C1	1,328
C11	1,328
<b>D</b>	<b>4,418</b>
D1	4,418
D11	2,578
D12	887
D13	953
<b>E</b>	<b>4,709</b>
E1	4,709
E11	2,334
E12	847
E13	1,528
<b>F</b>	<b>2,648</b>
F1	1,388
F11	1,388
F2	<b>1,260</b>
F21	1,260

**Note:** The alphabets A to F represent six clusters of the first level of the hierarchy; the alphanumeric values of A1 to F2 represent 10 clusters of the second level of the hierarchy; and the alphanumeric values of A11 to F21 represent 20 clusters of the third level of the hierarchy.

**Source:** Canadian Social Environment Typology Data.

### 5.3 Cluster Names, Characteristics and Uses

Three levels of nested clusters were created for the CanSET using hierarchical clustering method (Table 2). The optimal number of clusters were determined at six, ten and twenty. The median values were calculated for each variables for each cluster solution and specific characteristics were developed based on the position of the median value of each cluster in a quintile distribution of all the DAs in the analysis.

The CanSET data table comes with Dissemination Area Unique ID (DAUID) and associated cluster membership for each of the three cluster solutions. For example, one DA may fall into cluster A when we use six cluster classification but the same DA may fall into cluster A3 and A32 respectively for ten and twenty cluster classification. Users can select the cluster solution that distinguishes their area of interest and can use the respective value to classify the neighbourhoods. Users can select the cluster solution to use depending on the socioeconomic, demographic and ethno-cultural composition of a CMA or CA of interest. For instance, if users wants to use CanSET data to understand health inequalities in large CMAs like Toronto, Montreal or Vancouver, they may use all six, ten or twenty cluster solutions as all type of DAs are present in large CMAs having large population. However, only a few clusters types may be present in small CMAs or CAs. In that case, they may end up using only six or ten clusters to compare health outcomes.

#### 5.3.1 Six clusters solution

Level one of the hierarchy creates six clusters as the optimum cluster solution. The number of DAs in each cluster ranges from 1,328 to 19,127 whereas the median population density ranges from 2,176 to 8,160 people per square kilometre. In this solution, each DA is given a value of 1 to 6 based on their cluster membership. Following are some of the major characteristics of each cluster for this solution.

**Table 3**  
**Name and description of first level of clusters**

Cluster Number	Cluster Name	Cluster Description
Cluster 1	A	DAs in this cluster have medium population density but higher than average number of people per household; lower than average proportion of single parent families; high proportion of households with a university degree at bachelor's level or above; low unemployment rate and higher than average household income; higher than average proportion of people in managerial or professional occupations; high dwelling ownership rate and low proportion of households in need of major repair.
Cluster 2	B	DAs in this cluster have relatively low population density; lower than average number of people per household but higher than average proportion of single parent families; very low proportion of households with a university degree at bachelor's level or above; very low proportion of recent immigrant population but higher than average proportion of Aboriginal population; relatively high proportion of labour force in manufacturing, and sales and service occupation; relatively low median dwelling value and low adjusted family income.
Cluster 3	C	DAs in this cluster have very small household size; very low proportion of population under the age of 14 years but very high proportion of elderly population aged 65 years and above; very high proportion of institutionalized population; very high proportion of low income households; very high proportion of government transfer of payment recipients; low dwelling ownership rate; and very low adjusted family income.
Cluster 4	D	DAs in this cluster have very high population density and very low proportion of children 14 years of age and under; very small household size; very low proportion of labour force in manufacturing occupations but high proportion of population in professional occupations; higher than average proportion of households with a university degree; very low dwelling ownership rate and very high proportion of population spending more than 30% of income on housing costs; and higher than average dwelling value. Most of these DAs are located in the provinces of Quebec, Ontario, Alberta and British Columbia.
Cluster 5	E	DAs in this cluster have very high population density; relatively high proportion of population 14 years of age and under; very high proportion of lone parent families and very high proportion of government transfer of payment recipients; high unemployment rate; very high proportion of immigrants and recent immigrant population; high proportion of labour force working in sales and service related occupations; very low dwelling ownership rate; and very low adjusted family income. DAs in this cluster are mostly from the provinces of Quebec, Ontario and Alberta.
Cluster 6	F	DAs in this cluster have high population density; very large household size; very high proportion of immigrant population and very high proportion of visible minorities of South and East Asian origin; very high proportion of the population not speaking either of the official languages of Canada; and very high dwelling value. DAs in this cluster are mostly from the Montreal, Toronto, Calgary and Vancouver CMAs.

Source: Canadian Social Environment Typology Data.

### 5.3.2 Ten clusters solution

Level two of the hierarchy creates ten clusters as an optimum cluster solution (Table 2). The number of DAs varies from a minimum of 1,260 to the maximum of 7,787. The median population density ranges from 1,834 to the maximum of 8,160 per square kilometre. The ten clusters developed are nested in the six clusters described above (see Table 2). Therefore, the clusters in each level share some characteristics with the higher level of cluster to which they belong. The ten clusters have the characteristics as outlined below.

**Table 4**  
**Name and description of second level of clusters**

Cluster Number	Cluster Name	Cluster Description
Cluster 1	A1	DAs in this cluster are from the Canadian coast to coast with low population density; relatively low proportion of low income households; low unemployment rate; low proportion of immigrant population and very low proportion of visible minority population; high dwelling ownership rate; and high adjusted family income.
Cluster 2	A2	DAs in this cluster have relatively large family size; very low proportion of population receiving government transfer of payment; relatively low unemployment rate; high proportion of population working in managerial and professional occupations; and a very high proportion of households with a university degree. This cluster has mixed DAs from coast to coast with the highest adjusted family income.
Cluster 3	A3	DAs in this cluster have relatively high young population aged 14 years and under but low proportion of population aged 65 years and above; relatively large household size; high proportion of households with a university degree; high proportion of immigrant and visible minority population but low proportion of Aboriginal population; relatively high dwelling value and high adjusted family income.
Cluster 4	B1	DAs in this cluster have low population density; small household size; relatively high proportion of lone parent families; low proportion of immigrant population but high proportion of Aboriginal population; low dwelling value. Although the DAs in this cluster are found all across the country, most of the DAs from the territories and Northern areas of the provinces belong to this cluster.



**Table 4**  
**Name and description of second level of clusters**

Cluster Number	Cluster Name	Cluster Description
Cluster 5	B2	DAs in this cluster have very small household size; very high proportion of lone parent families, low income families and government transfer of payment recipients; high unemployment rate; low proportion of immigrant population but high proportion of Aboriginal population; very low proportion of household with a university degree; high proportion of population working in sales and service occupation but very low in managerial occupation; relatively high proportion of dwellings in need of major repair; very low dwelling value; and very low adjusted family income.
Cluster 6	C1	DAs in this cluster have very small household size; very low proportion of population under the age of 14 but very high proportion of elderly population aged 65 and above; very high proportion of institutionalized population; very high proportion of low income households; very high proportion of government transfer of payment recipients; low dwelling ownership rate; and very low adjusted family income.
Cluster 7	D1	DAs in this cluster have very high population density; very small household size; very low proportion of labour force in manufacturing occupations but very high proportion in professional occupations; higher than average proportion of household having a member with a university degree; high proportion of immigrants; very low dwelling ownership rate and very high proportion of population spending more than 30% of income on housing costs; and higher than average dwelling values. DAs in this cluster are mostly from the provinces of Quebec, Ontario, Alberta and British Columbia.
Cluster 8	E1	DAs in this cluster have very high population density; relatively high proportion of population under the age of 15; very high proportion of lone parent families and very high proportion of government transfer of payment recipients; high unemployment rate; very high proportion of immigrant and recent immigrant population; high proportion of labour force working in sales and service related occupations; very low dwelling ownership rate; and very low adjusted family income. DAs in this cluster are mostly from the provinces of Quebec, Ontario and Alberta.
Cluster 9	F1	DAs in this cluster have relatively high population density; very large household size; high proportion of households having a member with a university degree at bachelors level or above; very high proportion of immigrant population; high proportion of visible minorities of East Asian origin; very low proportion of Aboriginal population; relatively very high proportion of population not speaking either of the official languages of Canada; and very high median dwelling value. DAs in this cluster are mostly from Montreal, Toronto, Calgary and Vancouver CMA.
Cluster 10	F2	DAs in this cluster have high population density but relatively low proportion of elderly population 65 years of age and above; very large household size; very high proportion of immigrant population; very high proportion of South Asian and Black visible minorities; very low proportion of population working in professional occupations; and relatively high dwelling value. DAs in this cluster are mostly from Ontario and British Columbia.

Source: Canadian Social Environment Typology Data.

### 5.3.3 Twenty clusters solution

Level three of the hierarchy creates twenty clusters as an optimum cluster solution. The twenty clusters developed are nested within the ten clusters which in turn are nested within six clusters described above (see Table 2). Therefore, the clusters in each level share some characteristics with the previous level of cluster to which they belong. In the twenty cluster solution, the number of DAs in each cluster ranges from 847 to 6,650 whereas the median population density ranges from 32 to 19,750 per square kilometre. The twenty clusters have the following characteristics.

**Table 5**  
**Name and description of third level of clusters**

Cluster Number	Cluster Name	Cluster Description
Cluster 1	A11	DAs in this cluster are from the Canadian coast to coast with low population density; relatively low proportion of low income households; low proportion of immigrant population and very low proportion of visible minority population; low unemployment rate; high dwelling ownership rate; and high adjusted family income.
Cluster 2	A21	DAs in this cluster have low population density; low proportion of lone parent families; low proportion of government transfer of payment recipients; relatively low proportion of low income households; high proportions in managerial and professional occupations; and high median dwelling value and adjusted family income.
Cluster 3	A22	DAs in this cluster have large household size; very low proportion of lone parent family; the lowest proportion of government transfer of payment recipients; very low proportion of low income families; low unemployment rate; low proportion of Aboriginal and visible minority populations; highest dwelling ownership rate; high dwelling value; very high adjusted family income.
Cluster 4	A23	DAs in this cluster have relatively high proportion of children 14 years of age and under; very high proportion of households having a member with a university degree; very low proportion of government transfer of payment recipients; low average unemployment rate; very high proportion of population working in managerial and professional occupations; very low proportion of Aboriginal population; very high dwelling value; very high adjusted family income.

**Table 5**  
**Name and description of third level of clusters**

Cluster Number	Cluster Name	Cluster Description
Cluster 5	A31	DAs in this cluster have relatively large household size; low proportion of low income families; high proportion of immigrant population; relatively low proportion of Aboriginal population; relatively high dwelling ownership proportion and high adjusted family income.
Cluster 6	A32	DAs in this cluster have high population density; very large household size; relatively low proportion of lone parent families; very high proportion of immigrant population but low proportion of Aboriginal population; very high proportion of visible minority groups from South Asian and East Asian origin; high proportion of dwellings not suitable for accommodation; high dwelling value; and relatively high adjusted family income.
Cluster 7	B11	DAs in this cluster have relatively small household size and high proportion of lone parent families; very low proportion of households having a member with a university degree; relatively high proportion of population working in sales, service and manufacturing occupations; low proportion of immigrant population but high proportion of Aboriginal population; low dwelling value and high proportion of dwellings in need of major repair; low adjusted family income.
Cluster 8	B12	DAs in this cluster have low population density; high proportion of children 14 years of age and under but low proportion of seniors aged 65 years and over; low proportion of households having a member with a university degree; very high proportion of Aboriginal population but low proportion of immigrants and visible minority population; relatively high proportion of dwellings in need of major repair; relatively low dwelling values.
Cluster 9	B13	DAs in this cluster have the lowest population density but high proportion of seniors aged 65 years and above; low proportion of lone parent families; lowest proportion of households having a member with a university degree; low proportion of government transfer of payment recipients; very low proportion of immigrants and visible minority population but high proportion of Aboriginal population; relatively low dwelling value but high variability of the dwelling values.
Cluster 10	B21	DAs in this cluster have very small household size but very high proportion of lone parent families; high proportion of low income families and government transfer of payment recipients; very high unemployment rate; very high proportion of Aboriginal population; very low proportion of the families own their dwellings; very low dwelling value and very low adjusted family income.
Cluster 11	B22	DAs in this cluster have very high proportion of seniors but very low proportion of children 14 years of age and under; very small household size; very high proportion of population with below high school education; very high proportion of low income families and very high proportion of government transfer of payment recipients; high unemployment rate; low proportion of visible minority population; very low dwelling value and very low adjusted family income.
Cluster 12	C11	DAs in this cluster have very small household size; very low proportion of population 14 years of age and under but very high proportion of elderly population aged 65 and above; very high proportion of institutionalized population; very high proportion of low income households; very high proportion of government transfer of payment recipients; low dwelling ownership rate; and very low adjusted family income.
Cluster 13	D11	DAs with very high population density but very low proportion of children 14 years of age and under; very small household size; very high proportion of the population working in professional occupations; high proportion of households having a member with a university degree; very low proportion of the population own their dwellings; high average dwelling value.
Cluster 14	D12	DAs in this cluster have very low proportion of children but very high proportion of seniors aged 65 years and above; very small household size; low proportion of lone parent families; low proportion of population receiving government transfer of payment; mostly working in managerial and professional occupations; high variability of dwelling value.
Cluster 15	D13	DAs in this cluster have very high population density but very low proportion of children 14 years of age and under; high proportion of government transfer recipients; high proportion of low income families; very high proportion of immigrant population and visible minorities of East Asians and West Asian origin; very high proportion of families spending more than 30% on housing cost; high median dwelling value but very low adjusted family income.
Cluster 16	E11	DAs in this cluster have high population density; low proportion of population aged 65 years and above; very high proportion of lone parent families; relatively high unemployment rate; very high proportion of recent immigrant population; very high proportion of Latino and Black visible minorities; and very high proportion of dwellings not suitable for occupancy.
Cluster 17	E12	DAs in this cluster have relatively high population density; high proportion of lone parent families and relatively large household size; high proportion of government transfer of payment recipients; very high proportion of immigrant population; very high proportion of population not speaking either of the official languages of Canada; very high proportion of Latino and Black visible minorities groups; mostly working in sales and service occupation; and relatively high dwelling value.
Cluster 18	E13	DAs in this cluster have very high population density; very high proportion of lone parent families and high proportion of low income households; very high unemployment rate; very high proportion of immigrant and visible minority population but low proportion of Aboriginal population; low proportion of households having a member with a university degree; very high proportion of the households spending more than 30% on housing cost; very high proportion of dwellings not suitable for occupancy; very low adjusted family income.

**Table 5**  
**Name and description of third level of clusters**

Cluster Number	Cluster Name	Cluster Description
Cluster 19	F11	DAs in this cluster have low proportion of children 14 years of age and under; very large household size; high proportion of government transfer recipients; very high proportion of immigrants; very low proportion of Aboriginal population; very high proportion of visible minority of East Asian origin; very high dwelling value; high proportion of population without the knowledge of either of the official languages of Canada.
Cluster 20	F21	DAs in this cluster have high population density; relatively low proportion of seniors aged 65 years and above; very large household size; high proportion of households having a member with a university degree; very high proportion of immigrant population; relatively high unemployment rate; very high proportion of South Asian and Black visible minorities; relatively high proportion of population working in sales and service, and manufacturing occupations; relatively high dwelling value.

Source: Canadian Social Environment Typology Data.

## 6. How to link the CanSET with your data

The CanSET is designed to use with any dataset for which the DA or postal code information is available. The CanSET data can be easily linked with other data sources using the dissemination area unique identification (DAUID) number from the Census of population 2016. DAUID from any other Census of population may need a correspondence file to match them with the 2016 Census DAUID. The data set that do not have DA information but have the postal code information can also be linked with CanSET data using the Postal Code<sup>OM</sup> Conversion File plus (PCCF+) produced by Statistics Canada.

The CanSET data comes with a table that includes DAUID and hierarchal cluster membership for each of the three cluster solutions. The table contains four columns of DAUID, Cluster 6, Cluster 10 and Cluster 20 with numeric value of 1 to 6, 1 to 10 and 1 to 20 respectively as shown in a sample table (Table 6) below.

**Table 6**  
**Example rows of CanSET data**

DAUID	CanSET6	CanSET10	CanSET20
10010213	4	7	14
10010214	5	8	16
10010215	2	5	11
10010216	4	7	13
10010217	2	4	7
10010218	2	5	10
10010219	2	5	11
10010220	2	4	7

Source: Canadian Census of Population 2016 and Canadian Social Environment Typology Data.

## 7. Summary and Conclusion

This user guide provides an overview of the methodology used to develop CanSET and provides information on how to use these social environment clusters with the users own health or social data. To create the Can SET Canadian DAs were grouped together based on their demographic structure, socioeconomic status, and ethno-cultural background. The Can SET used 30 different variables from the Census of Population 2016 to classify Canadian DAs into three nested levels of clusters using the Fast Ward hierarchical clustering method. The optimal number of clusters were determined at six, ten and twenty using the Davies-Bouldin index. The median values were calculated for each variables for each cluster solution and specific characteristics and names were given to each cluster based on the value of the variables in each clusters.

The new CanSET portrays the complex social composition of Canadian urban geography. The CanSET is more comprehensive as it includes a variety of dimensions of urban social environment and covers all populous areas of Canada. This study is unique as it include all DAs from the CMAs and CAs simultaneously in a single typology rather than analyzing them separately for each CMA or CA. The analysis is based on the microdata from the Census of Population 2016 which covers all population of the study area. Therefore, the CanSET data allows cities to make comparison within themselves or with other cities and facilitates them to set benchmarks and track progress in combating health and social inequalities.

## 8. References

Davies, D. L., & Bouldin, D. W. (1987). A Cluster Separation Measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1(2), 224-227.

Johnson, R., & Wicheren, D. (2007). *Applied Multivariate Statistical Analysis* (6th ed.). Prentice Hall.

SAS Institute Inc. (2016). JMP 13 Multivariate Methods: Hierarchical Cluster. Cary, North Carolina, USA: SAS Institute Inc.

Statistics Canada. (2016). [Dictionary, Census of Population](http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/geo021-eng.cfm). Ottawa: Statistics Canada. Retrieved from <http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/geo021-eng.cfm>

Statistics Canada. (2018). *Postal Code Conversion File Plus (PCCF+) Version 7B, Reference Guide*. Ottawa: Statistics Canada.

## Appendix A

### Appendix A

#### Description of the variables included in the CanSET cluster analysis

Variable name	Description	Universe and Remarks
PopDen	Population density; total number of people per square kilometre of area.	Includes all population
InstHlthShelt	Percent of population that are institutionalized and residing in medical or long-term care facilities or shelters	Includes all population
Pop14	Percent of population aged 0 to 14 years	Includes all population
Pop65	Percent of population aged 65 and above	Includes all population
HhldSize	Average number of people in a household	Includes occupied private dwellings and units in senior residence but excludes other collective dwellings
LnePrnt	Percent of lone parent census families	Includes occupied private dwellings and units in senior residence but excludes other collective dwellings
NoHghSch	Percent of private households with the highest level of education of all its members "no certificate, degree or diploma"	Includes number of people aged 15 and over in occupied private households
UnivDgr	Percent of private households with at least one member of the household having "university certificate, diploma or degree at bachelor level or above"	Includes number of people aged 15 and over in private households
Inc	Household size adjusted median household income after tax	Calculated as the median household income*number of households in a DA/total number of households adjusted by the number of people in the households
GovTrfs	Percent of population receiving specific government transfers.	Includes population aged 15 and over in an occupied private dwellings and units in senior residence but excludes other collective dwellings
UErate	Unemployment rate for population aged 15 years and above	Includes population aged 15 and above who were available for work but unemployed in the census reference week
OccMang	Percent of employed labour force in management and administration occupation	Includes population aged 15 and above who were available for work in the census reference week
OccManuf	Percent of employed labour force in manufacturing, construction and trade related occupation	Includes population aged 15 and above who were available for work in the census reference week
OccuProf	Percent of employed labour force in professional occupation	Includes population aged 15 and above who were available for work in the census reference week
Imm	Percent of immigrant population	Includes persons in occupied private households. This question is not asked in Indian Reserves.
Reclmm	Percent of recent immigrant population (landed in Canada between 2011 and 2016)	Includes persons in occupied private households. This question is not asked in Indian Reserves.
AboRate	Percent of people who identified themselves with Aboriginal peoples of Canada	Includes respondents to this question in occupied private households.
VisMin	Percent of people who identified themselves as belonging to a visible minority group	Includes respondents to this question in occupied private households.
SthAsn	Percent of people belonging to South Asian visible minority group	Includes respondents to the visible minority question in occupied private households.
EastAsn	Percent of people belonging to Chinese, Filipino, Southeast Asian, Korean or Japanese visible minority groups	Includes respondents to the visible minority question in occupied private households.
Black	Percent of people belonging to Black visible minority group	Includes respondents to the visible minority question in occupied private households.
Latino	Percent of people belonging to Latin American visible minority group	Includes respondents to the visible minority question in occupied private households.
ArbWstAsn	Percent of people belonging to Arab or West Asian visible minority groups	Includes respondents to the visible minority question in occupied private households.
NoOffLang	Percent with no knowledge of either of official languages	Includes all population

**Appendix A****Description of the variables included in the CanSET cluster analysis**

<b>Variable name</b>	<b>Description</b>	<b>Universe and Remarks</b>
OwnDwl	Percent of private dwellings occupied by owner	Includes households that own or rent their private dwelling. Excluded is shelter occupancy on Indian reserves or settlements.
HouNONAff	Percent of households spending more than 30% of its average total income shelter costs	This variable is calculated for private households living in owned or rented dwellings who reported a total household income greater than zero. Excluded are band housing, farm dwellings.
DwNotSutab	Percent of private households living in not suitable accommodations (according to the National Occupancy Standard)	A dwelling is considered suitable if it has enough bedrooms for the size and composition of the household. Includes only private households.
DwRpair	Percent of occupied private dwellings in need of major repairs	Includes only occupied private dwellings
DwValue	Average value of privately owned dwellings	Includes only occupied private dwellings
Relative_iqrDV	Measure of variability in dwelling value within a DA	Includes only occupied private dwellings

**Source:** Canadian Census of Population 2016.