

Catalogue no. 18-001-X
ISBN 978-0-660-47705-3

Reports on Special Business Projects

Toward a Classification of Communities by Remoteness: A Proposal

by Fatemeh Hosseininasabnajar, Anne Munro, and Haaris Jafri

Release date: June 30, 2023



 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.

You can also contact us by

Email at 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 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

© His Majesty the King in Right of Canada as represented by the Minister of Industry, 2023

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

Acknowledgments	5
Summary	6
1 Introduction	7
2 Concepts and Data Sources	8
2.1 Remoteness concepts	8
2.2 Geographical concepts.....	9
2.3 Data sources.....	10
3 Methodology	11
3.1 Overview	11
3.2 Application	12
4 Results	16
4.1 Geographic Distribution of CSDs by Non-remote and Remote Class.....	16
4.2 Non-remote and Remote CSDs by Indigenous and Non-Indigenous Communities	16
4.3 Non-remote and Remote CSDs by Mode of Transportation	18
4.4 CSD Characteristics by Remoteness Class for Populated CSDs Connected to the Main Road/Ferry Network	18
5 Conclusion	20
6 References	21
7 Appendix	22

Table of tables

Table 1	Number of CSDs by k=2 and k=3 cluster	12
Table 2	Descriptive statistics of $AggPop_{SurrCSD}$ and $AggPop_{SurrPOPCTR}$ for populated CSDs within intermediate-remote class by IR range.....	13
Table 3	Descriptive statistics of $AggPop_{SurrCSD}$ and $AggPop_{SurrPOPCTR}$ for populated and connected CSDs; IR in [0.3793, 0.5010]	13
Table 4	Descriptive statistics for IR for groups of CSD that have factor values for $Agg_{PopSurrCSD}$ and $AggPop_{SurrPOPCTR}$ greater than the respective means of each factor; populated and connected CSDs; IR in [0.3793, 0.4500]	15
Table 5	IR descriptive statistics by mode of transportation; populated CSDs.....	18
Table 6	Descriptive statistics of CSDs by remoteness class (IR cut-off: 0.40); populated and connected CSDs.....	18
Table 7	IR descriptive statistics by remoteness class (IR cut-off: 0.40); populated and non-connected CSDs	19
Table A-1	Number of CSDs by mode of transportation, by two- and three-class natural clusters on the IR continuum	22

Table of figures

Figure 1	Frequency distribution of the IR of CSDs in Canada.....	9
Figure 2	Scatter plot of IR vs. the aggregate population of surrounding CSDs (i.e., $AggPop_{SurrCSD}$); IR in [0.3793, 0.5010]	14
Figure 3	Geographic distribution of non-remote and remote CSDs, IR cut-off point 0.40.....	16
Figure 4	Distribution of Indigenous and non-Indigenous populated CSDs by remoteness class	17
Figure 5	Population Distribution of Indigenous and non-Indigenous CSDs by remoteness class	17
Figure A-1	Summary at each step of narrowing the search interval to identify the IR cut-off.....	23
Figure A-2	Optimal number of clusters k-means clustering- Elbow method	24
Figure A-3	Optimal number of clusters k-means clustering- Average Silhouette method	25
Figure A-4	Proportion of CSDs with aggregate population of surrounding CSDs (i.e., $AggPop_{SurrCSD}$) greater than 139,000, by IR score ranges; IR in [0.3793, 0.5010].....	26
Figure A-5	Geographic distribution of remote, intermediate-remote, and non-remote CSDs obtained using k-means method (k=3).....	27

Acknowledgments

The authors of this report would like to acknowledge the following at Statistics Canada for their support and inputs: Alessandro Alasia, Pulkit Aggarwal, Eric Baxter, Mahamat Hamit-Haggar, Peter Murphy, and Jason Wong. The following colleagues at Indigenous Services Canada are also acknowledged for their valuable feedback: Eric Guimond, Bruno Powo Fosso, and Jennie Thompson.

Toward a Classification of Communities by Remoteness: A Proposal

by **Fatemeh Hosseinasabnajar**, **Anne Munro**, and **Haaris Jafri**

Summary

This study proposes a categorical classification of geographic communities (i.e., census subdivisions) into remoteness classes using a continuous index of remoteness. The methodology and the results of its application are discussed herein.

Alasia et al. (2017) used the travel cost from a community to population centres (POPCTRs) and the size of those POPCTRs to develop a continuous Index of Remoteness (IR) at the census subdivision (CSD) level in Canada. This index provides a measure of geographical proximity to what are typical points of service availability, taken in that work to be POPCTRs. The index has values ranging from 0 for the least remote community to 1 for the most remote community. This IR provides policymakers and researchers with a practical measure on a continuous scale for gauging the relative remoteness of communities.

However, in some applications, it is desirable to have discrete levels of remoteness defined by thresholds on the continuous IR. These levels would group communities into distinct categories of remoteness. This categorical classification would identify communities that are relatively less remote (i.e., starting with a “non-remote” class), and other classes with relatively more remote communities (i.e., ending with a “remote” class). When a two-level or binary classification is called for, a single threshold or cut-off is needed on the continuous IR. This would divide communities into a “remote” class and a “non-remote” class.

In this study, the IR was subdivided into classes based on natural clusters found on the IR continuum. To narrow the search interval to identify a single IR cut-off point on the [0, 1] range, the index values were initially grouped into three classes, with cut-off values of 0.2721 and 0.5010. In a following step, the class in the middle parts of the range, i.e., [0.2721, 0.5010], was split further, resulting in two classes over this narrowed range. Finally, the binary cut-off was identified at the IR value of 0.40.

In order to carry out the above steps to narrow the search range for a binary cut-off, in addition to looking at the natural breaks in the IR, factors were selected and used that relate to a more generalized concept of remoteness than what the IR represents. Such a remoteness concept encompasses the availability of services in the environs of a community, as proxied by the surrounding population. Therefore, the aggregate population of surrounding POPCTRs and the aggregate population of surrounding CSDs (i.e., communities) were selected to bring in this extended element of remoteness. These factors were evaluated in concert with the IR to achieve progressive narrowing of the search range. As a result of this procedure, the [0.4000, 0.4500] interval was identified as being of most interest. Applying a criterion whereby priority was given to putting borderline CSDs (i.e., IRs within [0.4000, 0.4500]) in the remote class, the lower end of this range, 0.40, was selected as the binary IR cut-off.

Using the 0.40 IR cut-off, about one-third (31.7%) of populated CSDs overall were flagged as remote. Among Indigenous CSDs (see section 2.2 for definition used), almost three-fifths (60.2%) of these communities were remote, while this proportion dropped to about one-fourth (25.3%) for non-Indigenous communities. In terms of population, about one in twenty-five (4.1%) Canadians lived in a remote area (based on the 2016 Census of Population), while almost three-fifths (60.3%) of the population of Indigenous CSDs also lived in remote communities.

This study outlines a methodology that is generic and is not tied to addressing any specific problem or application. It is acknowledged that specific applications might have their own unique needs for groupings by remoteness and might also follow some other valid approach than the one used here to achieve them. Users might also adapt the methodology used in this study to address their needs.

1 Introduction

Two of the main factors in defining the remoteness of a community are population density and proximity to relatively more densely populated and larger centres. By combining these factors, a continuous remoteness index can be computed with values that would range from a minimum value that would be assigned for the least remote communities, to a maximum value, for the most remote communities. An example would be values on a continuum from 0 to 1.

In some applications, it is desired or required to have two or more discrete levels of remoteness as opposed to a continuous index. Starting from the continuous Index of Remoteness (IR) for Canadian communities created by Alasia et al. (2017), the objective of this study is to develop a binary IR cut-off point to group CSDs into non-remote and remote classes (alternatively, less-remote, and more-remote, etc.).

The proposed binary IR cut-off point is identified by progressive narrowing of the range of IR values over which to search for such a point. The methodology finds natural breaks on the IR scores and exploring the cut-offs by examining the IR and the aggregate population within a certain degree of proximity to a given community.

The result of this analysis is a proposed classification that can be further tested in actual applications. A qualified proposal is put forward for the choice of a single cut-off point from a progressively narrowed search range on the full IR continuum. A rationale is provided for each successive narrowing of the search range and for the choice of a single value as the proposed binary IR cut-off. While the methodology used here is generic and is not tied to addressing any specific problem or application, it is acknowledged that different applications might require their own specialized groupings by remoteness and that the cut-off or threshold for other binary groupings could vary from what is proposed in this work.

The paper is organized as follows: section 2 presents an overview of the remoteness and geographic concepts and the data sources. Section 3 outlines the methodology used for finding the binary IR cut-off point. Section 4 presents the results of applying the identified binary cut-off. The conclusion in section 5 provides an overview of key findings and the position of the approach used here relative to other possible approaches and proposals for a binary IR cut-off.

2 Concepts and Data Sources

This section provides brief information on remoteness and the geographic concepts and data sources used in this study.

2.1 Remoteness concepts

Remoteness generally refers to the isolation of a community in physical terms. Population density and proximity to relatively more densely populated and larger centres are the two main attributes mostly used to capture the notion of the remoteness of a community. These two attributes are considered to compute the continuous Index of Remoteness (IR) of Canadian communities which is used in this study. A detailed methodological description of IR is presented in Alasia et al. (2017). The index is based on a gravity model and is computed for each census subdivision (CSD)¹ of Canada, reporting some population or a road connection. The model accounts for distance between a CSD and any population centres within a 2.5-hour travel time, as well as the population size of these population centres.² The continuous index that results from these computations ranges from 0 to 1, with values near 0 representing the least remoteness and closest proximity to population centers.³ This continuous index provides policymakers, researchers, and other stakeholders with an important tool for identifying communities within a specific range of remoteness.

Figure 1 displays the frequency distribution of the 2016 IR. The distribution is skewed to the right, implying that the number of communities with high remoteness are relatively few compared with those that have low or intermediate remoteness. The largest number of CSDs fall in the two bins with ranges from 0.3 to 0.35 and from 0.35 to 0.40.

Examining Figure 1 could arguably lead to three groupings being visualized: a first grouping including CSDs at the lower end of the range with IR values less than 0.15; a second grouping containing CSDs with IR values falling in the $[0.1500, 0.4500)$ ⁴ interval; and a third grouping with IR scores equal to or greater than 0.45. Over two-thirds (61.6%; 3,158 out of 5,125) CSDs have IR scores between $[0.1500, 0.4500)$. Subsequent sections of this paper will describe the methodology and the results obtained, but this initial observation leads to a preliminary expectation that the binary IR cut-off point falls in this middle interval.

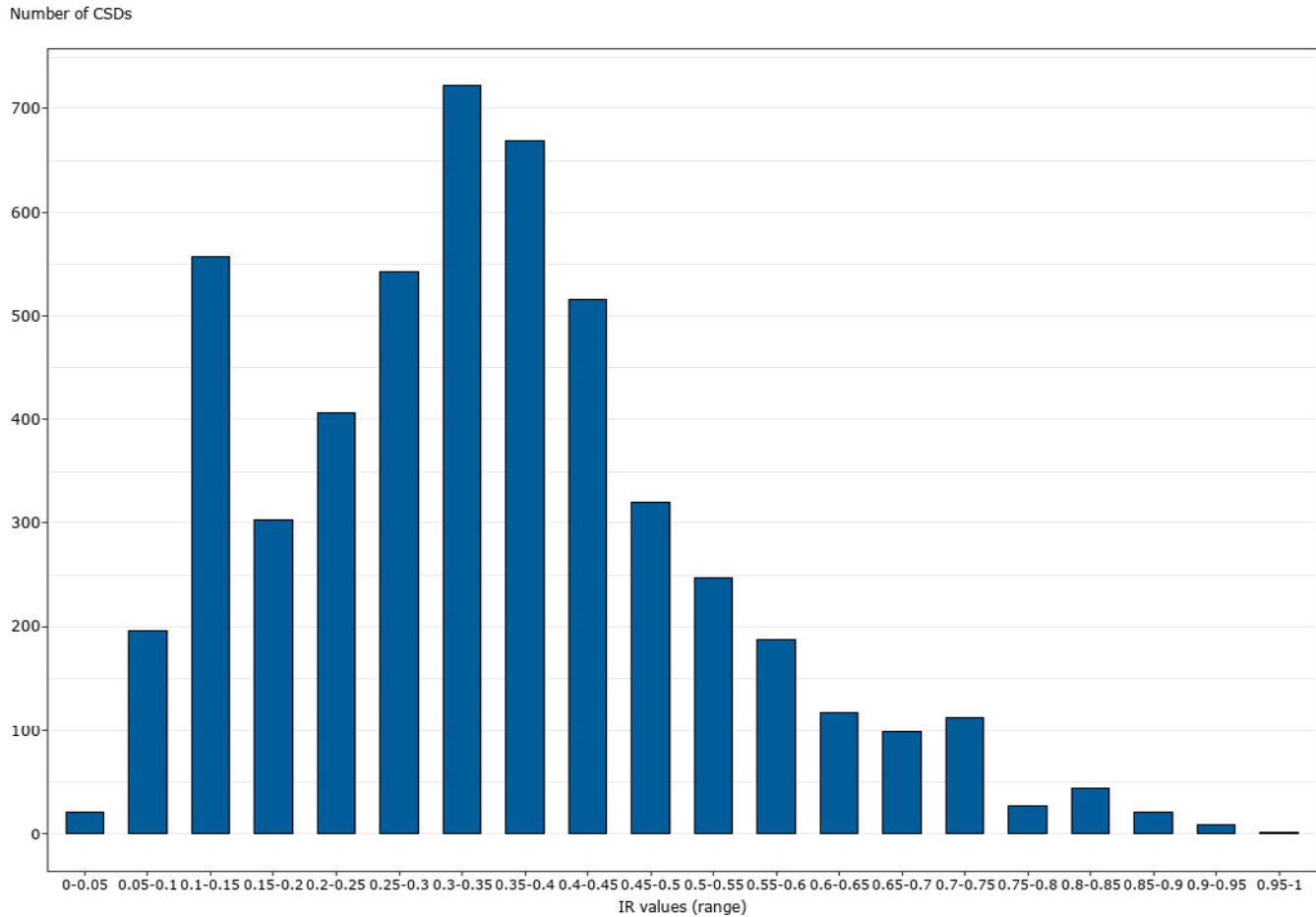
1. The detailed information on CSDs is available at [Dictionary, Census of Population, 2016 - Census subdivision \(CSD\) \(statcan.gc.ca\)](https://www150.statcan.gc.ca/n1/pub/92-627-x/2016001/article/00001-eng.htm).

2. The detailed information on POPCTR is available at [Dictionary, Census of Population, 2016 - Population centre \(POPCTR\) \(statcan.gc.ca\)](https://www150.statcan.gc.ca/n1/pub/92-627-x/2016001/article/00001-eng.htm).

3. An update to the IR based on the 2016 Census of Population and the 2016 SGC was released in 2020. This study is based on this update. See [Index of Remoteness \(statcan.gc.ca\)](https://www150.statcan.gc.ca/n1/pub/92-627-x/2020001/article/00001-eng.htm).

4. In the notation used in this study for indicating intervals, a square bracket (i.e., “[” or “]”) implies that the end value nearest the symbol occurrence is included in the interval, while an open or close parenthesis symbol (i.e., “(” or “)”) implies that the end value nearest the symbol occurrence is excluded.

Figure 1
Frequency distribution of the IR of CSDs in Canada



Source: Authors' computation based on Index of Remoteness (IR) values calculated using the 2016 Census of Population.

One of the ways of setting about defining discrete levels of remoteness is to identify natural clusters within the continuous IR. However, identifying the natural clusters does not suffice to flag non-remote and remote CSDs. The IR was calculated based on the proximity of a community to POPCTRs and the population of these POPCTRs. However, in the generalized sense, there are also other factors on which more comprehensive conceptualizations of remoteness may be based. For example, the IR does not adequately capture the condition of a community that is far from a POPCTR but is located in an area where it is surrounded by small communities that together contain a number of services. Therefore, other related factors that bring in a larger or more generalized conception of remoteness could be considered to advantage for the task at hand. In this analysis, to identify communities that could be classified as remote, factors such as the population in the surrounding areas of a community were also explored to identify a cut-off that can be used to classify CSDs.

2.2 Geographical concepts

This study is using Statistics Canada's 2016 Standard Geographical Classification (SGC).⁵ The 2016 SGC has 5,162 census subdivisions (CSDs, i.e., communities) and 1,005 population centres (POPCTRs). A CSD is an administrative region defined along municipal and reserve boundaries by the province or territory to which each

5. For details information on the 2016 SGC, see [Standard Geographical Classification \(SGC\) 2016 - Volume 1, The Classification \(statcan.gc.ca\)](https://www150.statcan.gc.ca/n1/pub/92-626-x/2016001/article/00001-eng.htm).

municipality or reserve belongs.⁶ In this study, CSDs are also referred as communities, since they are almost analogous to municipalities which can be thought of as making up a type of community. The term community may be more easily understood by some stakeholders than CSD. Whereas POPCTRs are defined as areas with a population of at least 1,000 and a population density of 400 persons or more per square kilometer.⁷

In this study, CSD are classified as Indigenous or non-Indigenous. Here, Indigenous CSDs are defined in terms of the 2016 SGC as any of the six CSD types legally affiliated with First Nations or Indian bands: Indian reserve (IRI), Indian settlement (S-É), Indian government district (IGD), Terres réservées aux Cris (TC), Terres réservées aux Naskapis (TK), Nisga'a land (NL).⁸ In addition, ISC provided a list including Inuit CSDs and a list of CSDs in Northwest Territories and Yukon which were affiliated with First Nations bands.⁹ CSDs included in these two lists were also classified as Indigenous in the current study.

Based on consultations with subject matter experts at ISC, there were only eight self-governing Métis settlements in Alberta with land-base agreements. These settlements were defined as Designated Places (DPLs) and created by provinces and territories, in cooperation with Statistics Canada.¹⁰ These DPLs were smaller than a CSD, and therefore the CSDs in which they located could not be classified as Indigenous CSDs.

All other CSDs are classified as non-Indigenous. Note that the classification of a CSD as non-Indigenous does not imply that the entire population of that CSD was non-Indigenous. The population of a CSD classified here as non-Indigenous could comprise of both non-Indigenous and Indigenous people.

Overall, 1,070 Indigenous CSDs (as defined above) with IR values were included in the analysis.

2.3 Data sources

This study relied mainly on the 2016 version of the continuous IR for Canadian communities whose original version was published in 2017 (Alasia et al., 2017). Whereas the original version of the IR is based on the 2011 Census of Population and the 2011 SGC, the 2016 version¹¹ is based on the 2016 Census of Population and the 2016 SGC.¹²

The 2016 IR contains values for 5,125 CSDs, including for the 4,882 populated CSDs, as well as for 243 non-populated CSDs that were connected to the main road/ferry network.

In the current study, the population of CSDs and POPCTRs was obtained in nearly all cases from the 2016 Census of Population.¹³

The travel time matrices¹⁴ used in this study were a copy of those used for the 2016 IR version. The travel-time matrices contain the travel time from each CSD to POPCTRs within 300 kilometres straight-line distance and from each CSD to other CSDs within this distance.

6. The detailed information is available at [Dictionary, Census of Population, 2016 - Census subdivision \(CSD\) \(statcan.gc.ca\)](https://www150.statcan.gc.ca/n1/pub/92-626-x/2016001/article/00001-eng.htm).

7. The detailed information is available at [Dictionary, Census of Population, 2016 - Population centre \(POPCTR\) \(statcan.gc.ca\)](https://www150.statcan.gc.ca/n1/pub/92-626-x/2016001/article/00001-eng.htm).

8. For detailed information on CSD types and CSD types associated with 'on reserve' population, see [Dictionary, Census of Population, 2016 - Census subdivision \(CSD\) \(statcan.gc.ca\)](https://www150.statcan.gc.ca/n1/pub/92-626-x/2016001/article/00001-eng.htm).

9. ISC provided a list of CSDs with these eight CSD types in Northwest Territories and Yukon which were affiliated with First Nations bands: Chartered Community (CC) in Northwest Territories, Community Government (CG) in the Northwest Territories, Hamlet (HAM) in the Northwest Territories, Settlement (SÉ) in the Yukon, Settlement (SET) in the Northwest Territories, Self-Governing (SG) in the Yukon, Teslin Land (TL) in the Yukon, Village (VL) in the Yukon and Northwest Territories.

10. For detailed information on Designated places (DPLs), see [Dictionary, Census of Population, 2016 - Designated place \(DPL\) \(statcan.gc.ca\)](https://www150.statcan.gc.ca/n1/pub/92-626-x/2016001/article/00001-eng.htm).

11. The 2016 IR scores are available at [Index of Remoteness \(statcan.gc.ca\)](https://www150.statcan.gc.ca/n1/pub/92-626-x/2016001/article/00001-eng.htm).

12. For details on the 2016 SGC, see [Standard Geographical Classification \(SGC\) 2016 - Volume I, The Classification \(statcan.gc.ca\)](https://www150.statcan.gc.ca/n1/pub/92-626-x/2016001/article/00001-eng.htm).

13. Information on data products of the 2016 Census of Population is available at [2016 Census of Population - Data products \(statcan.gc.ca\)](https://www150.statcan.gc.ca/n1/pub/92-626-x/2016001/article/00001-eng.htm).

14. The travel time matrices were created in 2017 and consistent with the 2016 Census of Population and 2016 boundary files.

3 Methodology

In order to classify CSDs into two groups based on their IR scores, several steps were taken. An overview of these steps is presented in the following sub-section, while the details of their application are discussed in the sub-section 3.2.

3.1 Overview

The initial step narrowed the range of the IR values over which a discrete cut-off for classifying communities into non-remote and remote classes could be found. This was done by finding natural breaks that occur in the distribution of the IR scores of all the CSDs using the k-means clustering method.¹⁵ This allowed for the identification of discrete groups within the continuous IR.

The k-means method generated clusters (or classes) in the IR score distribution by minimizing the “within class” variation and maximizing the “between classes” variation. Given the distribution observed in Figure 1, CSDs were grouped into three clusters based on their IR scores using a k-means algorithm with $k = 3$.¹⁶ The results of three clusters were combined with the two clusters (i.e., $k = 2$) to narrow the search interval for the desired binary IR cut-off.

In a second step, in addition to looking at natural breaks in IR scores, other factors were explored in order to distinguish groups of communities with relatively similar IR scores. It can be expected that residents of a community that have access to a set of other communities within a reasonable proximity, i.e., in their “surroundings”, that together have a relatively large aggregate population will have more services available to them. By contrast, residents of a community with a smaller aggregate population in their surrounding areas can be expected to have access to fewer services. This presumption is both intuitive and corroborated by evidence in the literature supporting the assumption that population size can be used as a proxy for service availability (see Alasia et al., 2017 and Department of Health and Aged Care in Australia, 2001).

Therefore, the aggregate population in CSDs and the aggregate population in POPCTRS around a community within a 2.5-hour travel time were selected for use in this study along with the natural breaks in IR scores. The 2.5-hour duration was used in order to limit the travel time to a plausible range of commuting over a single day. These factors describe the situation in the surrounding areas of a given community with a specific IR score and provide a lens for making relevant differentiations between it and other communities with relatively similar IR scores. For example, a community that is far from a POPCTR but is surrounded by several small communities, with a particular aggregate population, would likely be considered less isolated. Therefore, in a more generalized sense, it would be less remote compared to one with fewer near neighbors or a smaller aggregate population, even though both might have fairly close IR values.

The first factor, the “aggregate population in surrounding CSDs” or $\text{AggPop}_{\text{SurrCSD}}$, was calculated by finding the sum of the populations of all CSDs within a travel time of 2.5 hours and within a 300 km straight-line distance from a reference CSD. The population of the reference CSD itself was not included in the calculation of this indicator. The second factor, the “aggregate population in surrounding POPCTRS” or $\text{AggPop}_{\text{SurrPOPCTR}}$, was calculated by finding the sum of the populations of all POPCTRS within 2.5 hours travel time and within 300 km geodesic distance from a reference CSD, including the population of the POPCTR located in the reference CSD itself.¹⁷

The $\text{AggPop}_{\text{SurrCSD}}$ and $\text{AggPop}_{\text{SurrPOPCTR}}$ were utilized to search for a cut-off value by searching for patterns in their relationship with the IR within the intermediate interval, which was obtained based on natural breaks in the IR. This was done by exploring the measures of central tendency (mean and median) of these two factors and relating these with the distribution of IR scores. Communities with IR scores falling within the intermediate interval identified earlier were included in this analysis. The objective was to identify the IR score below which most or

15. Different clustering methods including hierarchical methods were tested and k-means was determined to be the optimal method. Using the Jenks natural breaks classification, similar results were obtained. The information on k-means clustering method is widely available. For example, see Hastie, T., Tibshirani, R., & Friedman, J. (2008). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd ed.). Springer.

16. The optimal number of clusters also turned out to be three. In order to determine that, CSDs were grouped into a number of clusters ranging from two to ten. These clusters were then evaluated to find the optimal number of clusters by the Elbow method, average silhouette, etc. (see Figure A- 2 and Figure A- 3 in Appendix). The Elbow method measures the compactness of the clustering, and average silhouette measure the quality of clustering, see Hastie, T., Tibshirani, R., & Friedman, J. (2008).

17. Other factors related to surrounding CSDs and POPCTRS were calculated and examined to select the factors which were associated with the remoteness of an area. Some examples of these quantities are the minimum travel time to (populated) CSDs, maximum/mean population of surrounding CSDs within a limited travel time, and minimum/mean travel time to small/medium and large POPCTRS. These factors were not selected because there was no clear-cut relationship between these factors and the IR which would enable their use to categorize communities into two remoteness classes, i.e., their correlation with the IR was quite low and statistically insignificant.

nearly all CSDs are surrounded by a greater aggregate population ($AggPop_{SurrCSD}$) than the mean of $AggPop_{SurrCSD}$ for all CSDs. The same process was repeated using the aggregate population of surrounding POPCTRs, i.e., using $AggPop_{SurrPOPCTR}$.

In the final step, after identifying some candidates for IR cut-off, a single IR cut-off was selected. This was done based primarily on a qualitative argument that prioritizes classifying borderline CSDs as remote. The decision was also supported by examining IR characteristics by mode of transportation.

The next section explains the application of these steps, see the summary of these steps in Figure A-1 in Appendix.

3.2 Application

Natural breaks in the IR scores were identified using the k-means method as the first step in categorizing CSDs into natural clusters. This allowed for the identification of discrete groups within the continuous IR. Applying this method, CSDs were grouped into two and three clusters (i.e., $k = 2$ and $k = 3$). The classification of CSDs into three clusters yielded an intermediate (or middle) range within which the search for the binary IR cut-off would proceed. Note that intervals were displayed mathematically in this study such that a square bracket (i.e., $]$) was used when including the end value and a round bracket (i.e., $)$ to exclude the end value.

Table 1 shows the number of CSDs by class obtained by running k-means algorithm for $k = 2$ and $k = 3$.

Table 1
Number of CSDs by k=2 and k=3 cluster

Two-class	Three-class		
	Non-remote; IR range [0, 0.2717]	Intermediate-remote; IR range [0.2721, 0.5010]	Remote; IR range [0.5014, 1]
	CSD counts		
Non- remote; IR range [0, 0.3791]	1,717		1,415
Remote; IR range [0.3793, 1]	...		1,127
			866

... not applicable

Note: CSDs are classified into two and three classes based on their IR scores using k-means algorithm.

Source: Authors' computation based on Index of Remoteness (IR) scores calculated using the 2016 Census of Population.

As seen in Table 1, CSDs with IR scores at either end of the IR distribution (1,717 CSDs with IRs in $[0, 0.2717]$ and 866 CSDs with IRs in $[0.5014, 1]$) were classified respectively as being non-remote (close to 0) or remote (close to 1) based on being sorted into either two and three clusters ($k = 2$ or $k = 3$). However, the classification of CSDs with score in the range of $[0.2721, 0.5010]$, called "Intermediate-Remote", would change depending on whether two or three clusters were created. This interval is defined a transitional working category between non-remote and remote when two natural breaks were being assessed.

Given the overlap of classes based on two and three clusters, the IR score of 0.3791 (i.e., the natural break/cut-off in IRs obtained based on k-means algorithm for $k = 2$) divides the intermediate-remote class into two groups: (i) CSDs with IR scores within $[0.2721, 0.3791]$ interval and (ii) CSDs with IR scores within $(0.3791, 0.5010]$ interval. With a view to explore further this intermediate-remote class of CSDs, Table 2 displays the descriptive statistics of $AggPop_{SurrCSD}$ and $AggPop_{SurrPOPCTR}$ in the IR ranges obtained by regarding the single natural break obtained based on $k = 2$ means algorithm for populated CSDs with IR scores that fall between 0.2721 and 0.5010.

The table shows that CSDs with IR in the range of $[0.2721, 0.3791]$ were in areas that had a minimum aggregate population of surrounding population centres ($AggPop_{SurrPOPCTR}$) of 30,707. This is close to the lower population threshold of a medium POPCTR, which is 30,000.¹⁸ Looking at the median value of $AggPop_{SurrCSD}$ for the CSDs with IRs in $[0.2721, 0.3791]$ in Table 2, it is seen that more than half of these communities also have access to aggregate surrounding population in the range of the size of a large POPCTR (i.e., 100,000 or more).¹⁹ From these estimates of the equivalent level of access in their surrounding areas when comparing to POPCTRs with relatively large populations, it follows that the binary cut-off point should be considered to fall outside this interval (i.e., outside $[0.2721, 0.3791]$). Furthermore, it should be located within an interval that represents greater remoteness, being the $[0.3793, 0.5010]$ interval. This latter range has 1,127 CSDs, with the majority of them (94%, or 1,058 CSDs) being populated CSDs connected to the main road/ferry networks (Table A- 1 in Appendix).

18. For detailed information see [Dictionary, Census of Population, 2016 - Population centre \(POPCTR\) \(statcan.gc.ca\)](https://www150.statcan.gc.ca/n1/pub/92-627-x/2016001/article/00001-eng.htm).

19. The detailed information is available at [Dictionary, Census of Population, 2016 - Population centre \(POPCTR\) \(statcan.gc.ca\)](https://www150.statcan.gc.ca/n1/pub/92-627-x/2016001/article/00001-eng.htm).

Table 2
Descriptive statistics of $AggPop_{SurrCSD}$ and $AggPop_{SurrPOPCTR}$ for populated CSDs within intermediate-remote class by IR range

Factor	IR Range	Aggregate Population				
		Min	Max	Mean	Standard deviation	Median
$AggPop_{SurrCSD}$	[0.2721, 0.3791]	1,477	1,830,417	460,103	229,372	408,938
	[0.3793, 0.5010]	0	1,100,780	138,352	81,584	129,413
$AggPop_{SurrPOPCTR}$	[0.2721, 0.3791]	30,707	1,165,823	325,483	170,437	295,917
	[0.3793, 0.5010]	0	218,666	72,524	39,124	68,820

Note: Given the overlap of classes based on two and three clusters by using k-means method, the IR score of 0.3791 (i.e., the cut-off obtained for k=2) divides CSDs falling within the intermediate-remote class (i.e., IRs within [0.2721, 0.5010] obtained for k=3) into two groups within specific IR ranges: [0.2721, 0.3791] and [0.3793, 0.5010]. There are 2,431 populated CSDs in intermediate-remote class of which 1,066 CSDs have IR scores within [0.3793, 0.5010] interval. Note that there is no CSD with IR scores in the (0.3791, 0.3793) interval.

Source: Authors' computation based on Index of Remoteness (IR) scores calculated using the 2016 Census of Population and the 2016 Census of Population.

The next step continues to narrow the search interval for the IR cut-off further by examining the characteristics of CSDs with IR scores falling inside the [0.3793, 0.5010] interval. For this purpose, the following factors which represent the aggregate population of surrounding areas are used (i) $AggPop_{SurrCSD}$ (the aggregate population of surrounding CSDs); and (ii) $AggPop_{SurrPOPCTR}$ (the aggregate population of surrounding POPCTRs).

Table 3 displays the descriptive statistics for $AggPop_{SurrCSD}$ and $AggPop_{SurrPOPCTR}$ for populated and connected CSDs with IR scores between 0.3793 and 0.5010, the narrowed IR range identified earlier.

Table 3
Descriptive statistics of $AggPop_{SurrCSD}$ and $AggPop_{SurrPOPCTR}$ for populated and connected CSDs; IR in [0.3793, 0.5010]

Factor	Min	Max	Mean	Standard deviation	Median	95th percentile
$AggPop_{SurrCSD}$	325	1,100,780	139,044	81,173	130,036	270,281
$AggPop_{SurrPOPCTR}$	0	218,666	72,606	39,078	68,820	136,154

Note: Only CSDs with non-zero population in 2016 which were connected to the main road/ferry network were included in the calculation.

Source: Authors' computation based on Index of Remoteness (IR) scores calculated using the 2016 Census of Population and the 2016 Census of Population.

With the mean values of $AggPop_{SurrCSD}$ and $AggPop_{SurrPOPCTR}$ noted from Table 3, the distribution of IR scores versus these factors is examined to identify any insight-giving patterns that might be present in IR scores with respect to the mean values of these factors.

Figure 2 displays the scatter plot of the IR scores (y-axis) versus $AggPop_{SurrCSD}$. As expected, there is a negative correlation between the IR and $AggPop_{SurrCSD}$. The vertical line in this figure corresponds to the mean of $AggPop_{SurrCSD}$ (see Table 3). This line divides the CSDs into two groups based on whether or not their $AggPop_{SurrCSD}$ is above the average for all populated and connected CSDs.

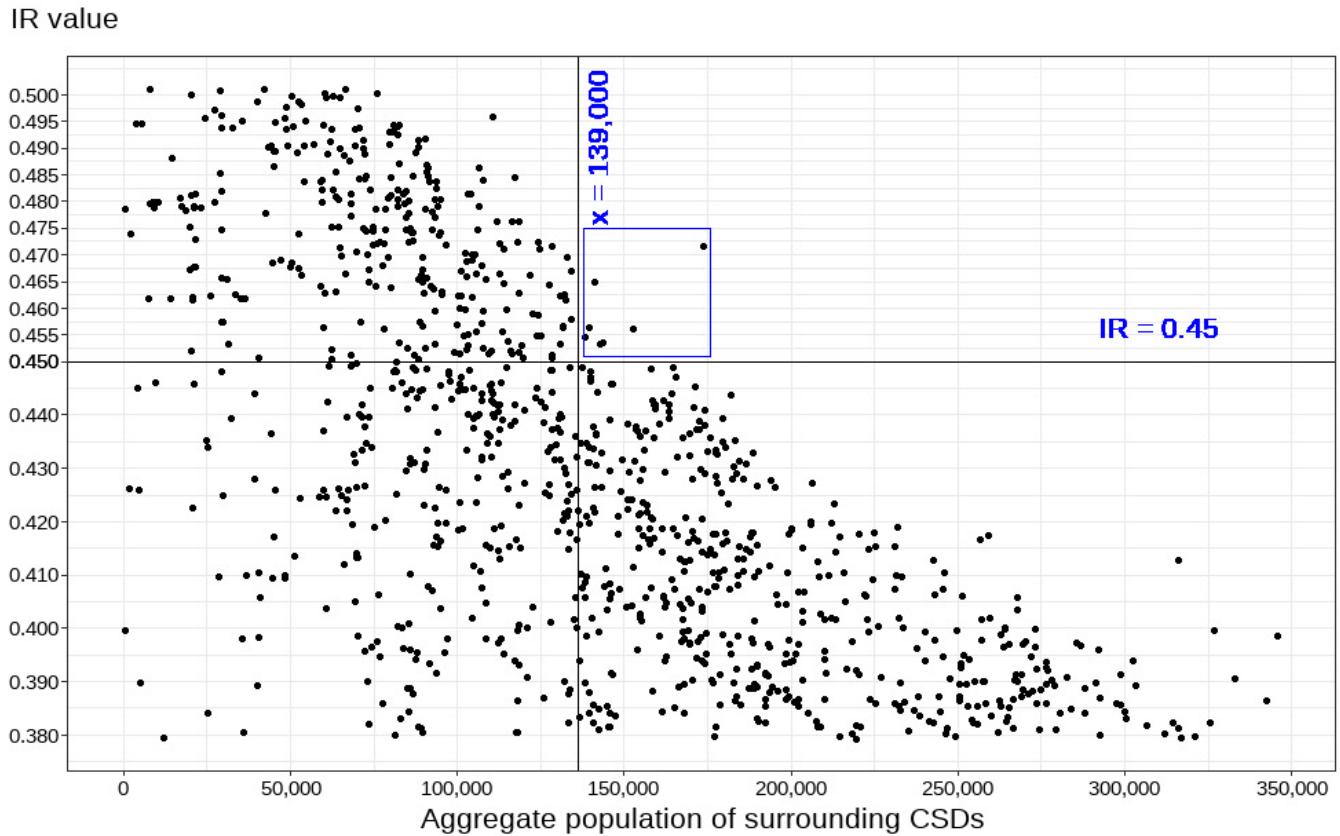
Next, moving up the IR axis in Figure 2, it is examined whether there is an IR score below which most or nearly all CSDs have $AggPop_{SurrCSD}$ that is greater than the mean for this factor for all CSDs. This IR value can be a potential cut-off point as it separates CSDs with relatively greater aggregate surrounding populations from those with smaller aggregate populations in surrounding areas.

As seen in Figure 2, nearly all CSDs (except for 6 CSDs; see the top right of the Figure, above the IR = 0.45 horizontal line) that have an above-average $AggPop_{SurrCSD}$ have IR scores of greater than 0.45 (i.e., to the right of the vertical line). Also, dividing the interval of [0.3793,0.5010] into smaller intervals with a step size of 0.0050, the proportion of CSDs with $AggPop_{SurrCSD}$ greater than 139,000 was calculated for CSDs with IR scores within each small interval (e.g., [0.4400,0.4450]), see Figure A- 4 in the Appendix. As may be seen in Figure A- 4, the proportion of CSDs with $AggPop_{SurrCSD}$ greater than 139,000 is less than 10.0% (i.e., from no CSDs to two CSDs) in every small interval for IR values greater than or equal to 0.45.

Note that the mean and median IR scores of CSDs falling below the mean of $AggPop_{SurrCSD}$ (139,000) (i.e., to the left of the vertical line) are equal to 0.45 (rounded to two decimal points). This result suggests that 0.45 can be a potential upper limit for the IR cut-off, since CSDs with IR scores above this value have access to areas with smaller aggregate populations.

Similarly, examining the distribution of IR scores versus $\text{AggPop}_{\text{SurrPOPCTR}}$, 98.4% of CSDs with access to above average mass of POPCTR aggregate population (see Table 3) have IR scores less than 0.45. This result also supports 0.45 as a potential limit value for a cut-off point, as it separates CSDs with relatively large aggregate population in their surrounding population centres from CSDs with smaller values for this factor.

Figure 2
Scatter plot of IR vs. the aggregate population of surrounding CSDs (i.e., $\text{AggPop}_{\text{SurrCSD}}$); IR in [0.3793, 0.5010]



Notes: Each point represents a populated CSDs which is connected to the main road/ferry network with IR value in [0.3793, 0.5010].
 The vertical line ($x = 139,000$) represents the mean of the aggregate population of surrounding CSDs within 2.5 hours.
Source: Authors' computation based on Index of Remoteness (IR) values calculated using the 2016 Census of Population.

This analysis further narrows the search interval of a possible IR cut-off for a binary classification, from [0.3793, 0.5010] to [0.3793, 0.4500). The value of 0.4500 is excluded from the search interval when proceeding to the next step, because the CSDs that have this IR score are more remote compared to those with smaller IR scores. There is only one CSD that happens to have an IR score of exactly 0.4500 while having an $\text{AggPop}_{\text{SurrPOPCTR}}$ of less than 139,000.

The next step consisted of examining IR characteristics of CSDs within this IR range by partitioning them around the mean values of the two population factors. CSDs whose factor values in each case are above the respective means of the factor values have a relatively large aggregate population in their surrounding areas compared to CSDs whose factor values fall below the mean. This provides a suitable way for distinguishing less remote from more remote communities. Communities with relatively small aggregate surrounding populations, on average, have greater IR scores and are likely to be more remote in the generalized sense of remoteness compared with those with greater aggregate populations in their environs, and vice versa.²⁰

20. For CSDs with IR scores falling in [0.3793, 0.4500) that have factor means below average, the mean and median IR scores are 0.42 (rounding to decimal points).

Table 4 shows that these mean values of $AggPop_{SurrCSD}$ and $AggPop_{SurrPOPCTR}$ are 164,000 and 87,000, respectively, and it also depicts the descriptive statistics of the CSD groups for which the values of these two factors are greater than their respective mean values (within the IR range of interest, i.e., [0.3793, 0.4500]).

Table 4

Descriptive statistics for IR for groups of CSD that have factor values for $AggPop_{SurrCSD}$ and $AggPop_{SurrPOPCTR}$ greater than the respective means of each factor; populated and connected CSDs; IR in [0.3793, 0.4500]

Factor	Factor mean	CSD counts	IR for Group of CSDs having factor values > factor mean				
			Min	Max	Mean	Standard deviation	Median
$AggPop_{SurrCSD}$	164,000	360	0.3793	0.4489	0.4015	0.0164	0.3974
$AggPop_{SurrPOPCTR}$	87,000	357	0.3793	0.4448	0.4018	0.0160	0.3986

Source: Authors' computation based on Index of Remoteness (IR) scores calculated using the 2016 Census of Population and the 2016 Census of Population.

Table 4 shows that the mean IR for the group created using either factor is approximately 0.40 (rounding to two decimal places). The median IR for these groups is also 0.40, implying that the middle point or the pivot point of the IRs is 0.4. This analysis introduces 0.40 as another potential cut-off point/lower limit of the search range.

Examining the reduced interval defined by 0.40 as its upper point, i.e., [0.3793, 0.4000],²¹ it is first noted that there are 277 CSDs with an IR score within the interval which have mean and median IR scores of 0.3900. These CSDs have a minimum value for $AggPop_{SurrPOPCTR}$ of 18,884, and nearly all of them (98%) have $AggPop_{SurrCSD}$ and $AggPop_{SurrPOPCTR}$ values over 30,000 (close to the lower population threshold of a medium POPCTR). Looking at a closer level of proximity, over 80% of these communities also have access to aggregate population of 30,000 and more within 1.5 hours travel time.

Based on this analysis, remote CSDs would be expected to have access to aggregate population in their surroundings within 2.5 hours that is smaller than 30,000 when compared to non-remote CSDs. This analysis provide support to a cut-off value of 0.40 as a lower limit of the IR value for the search range.

The cut-off point was selected such that, to the greatest extent possible, it would classify borderline CSDs as remote. To achieve this, the lower limit of the [0.4000, 0.4500] range, i.e., 0.4000,²² was selected as the preferred cut-off point. In specific field applications, practitioners might consider tailoring the approach used here to their requirements. For example, they could investigate the CSDs with IR scores relatively close to the 0.40 cut-off somewhat more comprehensively in terms of focusing on an aspect which is important for their application.

One of the strengths of the methodology used in this study lies in exploring the natural breaks on the IR continuum to yield clusters. These clusters indicate conceptual groupings of communities in terms of similarities in their degree of remoteness as expressed by their IR score. These clusters imply a degree of similarity between communities within them. However, the methodology is still dependent upon the use of aggregate populations in surrounding areas, which may not capture the full picture for individual communities. Another limitation of the methodology lies in the relative uncertainty as to the extent to which a distributed population can serve as a proxy for service availability. Although much research has been done on population size as a measure of what services are available, it is less certain that a collection of smaller communities would offer the same services as would be found in an urban hub of a similar total population size.

21. The single CSD with an IR score of 0.40 is more remote compared to those with smaller IR scores, therefore it is excluded from the search interval in the next step. Only one CSD happened to have an IR score of 0.4000 with 135,618 aggregate population of CSDs within 2.5 hours.

22. For brevity, the IR value of "0.4000" is also expressed as "0.40" in certain parts of the text as well as in some figures and charts.

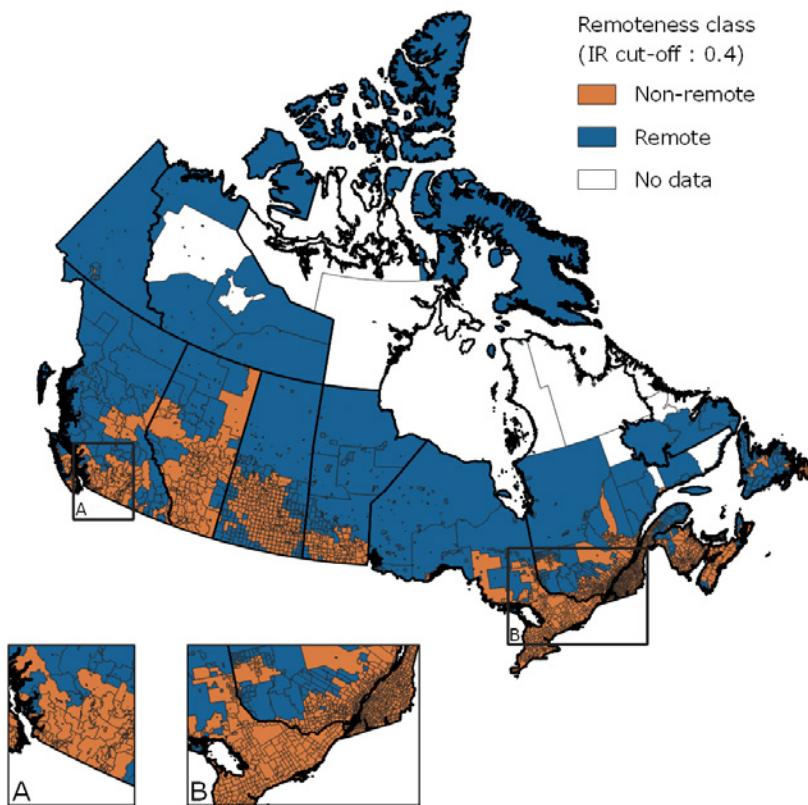
4 Results

The results of applying the proposed IR binary cut-off of 0.40 to all CSDs in Canada is discussed in this section. While discussing these results, it should be recalled that communities classified using the binary cut-off into the two groups (non-remote and remote) do still have different levels of remoteness on the IR continuum compared to others in the same group.

4.1 Geographic Distribution of CSDs by Non-remote and Remote Class

Figure 3 shows the geographic distribution of non-remote and remote CSDs based on the 0.40 cut-off.²³ In general, vast regions of the northern parts of several provinces are predominantly made up of remote CSDs. Most parts of the territories are also made up of remote CSDs. However, the distribution is more mixed in most of the Atlantic provinces except for Newfoundland and Labrador, as well as in Alberta and parts of British Columbia.

Figure 3
Geographic distribution of non-remote and remote CSDs, IR cut-off point 0.40



Note: The white parts in the map represent CSDs with no reported population in 2016 and not connected to the main road/ferry network.
Source: Authors' computation based on Index of Remoteness (IR) scores calculated using the 2016 Census of Population.

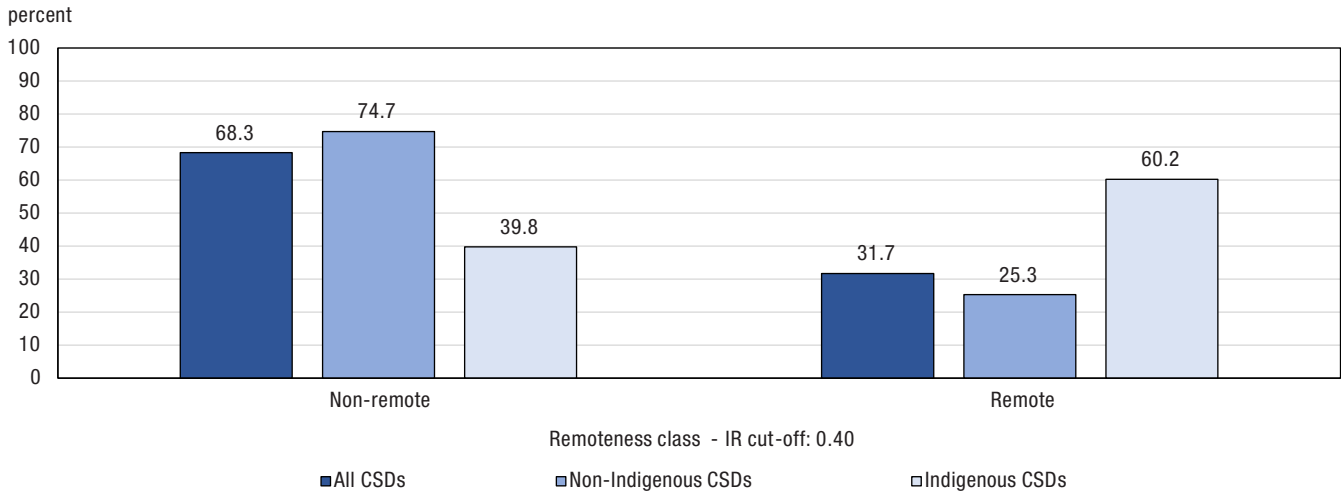
4.2 Non-remote and Remote CSDs by Indigenous and Non-Indigenous Communities

Figure 4 displays the distribution of populated CSDs by remoteness class based on the 0.40 IR cut-off, including breakdowns by Indigenous and non-Indigenous communities.²⁴

23. The geographic distribution of CSDs based on the three-class of remoteness obtained using the k-means method for k=3 was displayed in Figure A-5 Appendix.
 24. For details on the classification of CSDs as Indigenous and non-Indigenous in this study, see section 2.2.

Among all populated CSDs, about one-third (31.7%) of them were remote based on this IR cut-off. Focusing on Indigenous CSDs, almost three-fifths (60.2%) of these CSDs were classified as remote. This proportion dropped to one-fourth (25.3%) for non-Indigenous communities.

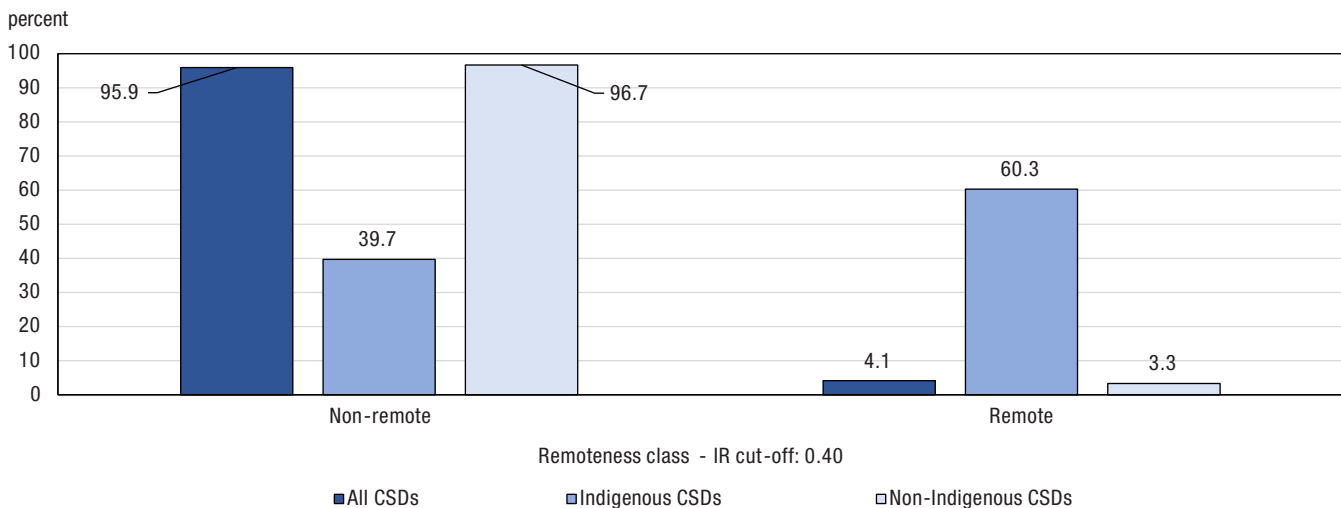
Figure 4
Distribution of Indigenous and non-Indigenous populated CSDs by remoteness class



Note: Only CSDs with non-zero population according to the 2016 Census of Population are included. The CSDs were classified as remote if the IR score was 0.40 or greater.
Source: Authors' computation based on Index of Remoteness (IR) scores calculated using the 2016 Census of Population.

In terms of population distribution, Figure 5 shows that almost three-fifths (60.3%) of the residents of Indigenous CSDs lived in remote communities, compared to only about 3.3% of the population of non-Indigenous CSDs.²⁵ Overall, about one in twenty-five (4.1%) of all Canadians lived in remote areas (based on the 2016 Census of population).

Figure 5
Population Distribution of Indigenous and non-Indigenous CSDs by remoteness class



Note: Only CSDs with non-zero population according to the 2016 Census of Population are included. The CSDs were classified as remote if the IR score was 0.40 or greater.
Source: Authors' computation based on Index of Remoteness (IR) scores calculated using the 2016 Census of Population.

25. For details on the classification of CSDs as Indigenous and non-Indigenous in this study, see section 2.2.

4.3 Non-remote and Remote CSDs by Mode of Transportation

Table 5 depicts the IR descriptive statistics for all populated CSDs by mode of transportation. All 47 CSDs which were connected to other CSDs via air only have IR scores greater than 0.40 and therefore were considered to be remote based on the 0.40 IR cut-off. However, CSDs which were connected to other communities via the main road/ferry network and by a combination of air, train, winter road, charter boat and/or seasonal ferry could be classified as either non-remote or remote, since the minimum IR score of these CSDs is less than 0.40. The next two sub-sections look at these CSDs in more detail.

Table 5
IR descriptive statistics by mode of transportation; populated CSDs

Mode of transportation	CSD		IR				
	Counts	Percent	Min	Max	Mean	Standard deviation	Median
Air	47	1.0	0.4671	1	0.7862	0.1472	0.8523
Combination of air, train, winter road, charter boat, and/or seasonal ferry	91	1.9	0.2532	0.9173	0.7302	0.1392	0.7811
Main road/ferry network	4,744	97.2	0	0.8571	0.3301	0.1533	0.3297
Total	4,882	100.0	0	1	0.3419	0.1680	0.3344

Note: Only CSDs with non-zero population according to the 2016 Census of Population are included.

Source: Authors' computation based on Index of Remoteness (IR) scores calculated using the 2016 Census of Population and the 2016 Census of Population.

4.4 CSD Characteristics by Remoteness Class for Populated CSDs Connected to the Main Road/Ferry Network

Table 6 displays the descriptive statistics of population and IR scores of populated CSDs which were connected to the main road/ferry network by the remoteness class obtained based on the 0.40 IR cut-off. It shows less than one-third (29.7%) of populated and connected CSDs were flagged as remote communities, and about one in twenty-five (3.8%) of Canadians lived in these remote communities.

Table 6
Descriptive statistics of CSDs by remoteness class (IR cut-off: 0.40); populated and connected CSDs

Remoteness Class	CSD		Population		IR		
	Counts	Percent	Sum	Percent	Mean	Standard deviation	Median
Non-remote	3,334	70.3	33,745,749	96.2	0.2528	0.0978	0.2695
Remote	1,410	29.7	1,323,913	3.8	0.5127	0.0947	0.4847

Note: Only CSDs with non-zero population according to the 2016 Census of Population and that are connected to the main road/ferry network are included.

Source: Authors' computation based on Index of Remoteness (IR) scores calculated using the 2016 Census of Population and the 2016 Census of Population.

4.5 CSD Characteristics by Remoteness Class for CSDs Connected to Other CSDs via a Combination of Modes of Transportation

Table 7 shows descriptive statistics of population and IR scores of CSDs which are connected to other communities via some combination of air, train, winter road, charter boat and/or seasonal ferry (e.g., CSDs that were non-connected to the main road/ferry network), by remoteness class based on the 0.40 IR cut-off. Among all such CSDs, only two were non-remote with one located in Quebec and the other in Ontario. Overall, the vast majority (97.8%; 89 out of 91) of these non-connected CSDs were flagged as remote.

Table 7
IR descriptive statistics by remoteness class (IR cut-off: 0.40); populated and non-connected CSDs

Remoteness Class	CSD Population				IR		
	Count	Sum	Min	Max	Mean	Standard deviation	Median
Non- remote	2	379	0.2532	0.3461	0.2997	0.0657	0.2997
Remote	89	59,790	0.4084	0.9173	0.7398	0.1243	0.7850

Note: Only CSDs with non-zero population according to the 2016 Census of Population and that are connected to other communities via some combination of air, train, winter road, charter boat and/or seasonal ferry are included.

Source: Authors' computation based on Index of Remoteness (IR) scores calculated using the 2016 Census of Population and the 2016 Census of Population.

Given the number of remote CSDs connected to other communities via some combination of mode of transportation (i.e., 89; see Table 8) and 47 remote CSD connected via air (see Table 5), nearly all (136 out of 138; 98.6%) communities which were not connected to other CSDs via the main road/ferry network had an IR score greater than 0.40. These results support the selection of 0.40 as a binary cut-off point, since this cut-off separates communities which were not connected to the main road/ferry network. Therefore, these communities would naturally be expected to be more geographic isolated or remote.

5 Conclusion

This study describes the methodology and results of proposing a single cut-off point that can be applied to a continuous Index of Remoteness (IR), previously developed at Statistics Canada, to classify all CSDs in Canada into two classes of remoteness. Of these two classes, one would be non-remote (or less-remote, an IR value of less than 0.40) and the other, remote (or more-remote, an IR value of greater than or equal to 0.40). This was done by first identifying natural clusters on the IR continuum and using the resulting classes to select a single cut-off value using supplementary factors that relate to a broader concept of remoteness.

The classification of CSDs into three natural clusters yielded an intermediate range of IR values which resulted in narrowing the search range into a range of IR values of [0.3793, 0.5010].

The aggregate surrounding populations of a CSD (e.g., the population within a certain proximity) was used to describe the remoteness of an area in a more generalized sense than would be captured by IR. This surrounding population would include both the surrounding CSDs and the surrounding POPCTRs. Therefore, in addition to the natural clusters in the IR distribution, the aggregate population of CSDs and the aggregate population of POPCTRs within 2.5 hours of travel time of a reference CSD were used to identify the binary cut-off point.

The narrowed range of [0.3793, 0.5010] that was obtained based on the natural breaks in IR scores was explored using the two factors relating to the aggregate surrounding population to narrow the range further in a progressive manner. This process yielded the range to be [0.4000, 0.4500]. Applying a criterion whereby priority is given to putting borderline CSDs (i.e., IRs within [0.4000, 0.4500]) into the remote class, the lower end of this range, 0.40, was selected as the binary IR cut-off.

Overall, about one-third (31.7%) of populated CSDs were flagged as remote, and only one in twenty-five (4.1%) of Canadians lived in these remote communities (based on the 2016 Census of Population). Based on the 0.40 IR cut-off, almost three-fifths (60.2%) of all Indigenous CSDs were classified as being remote while this proportion drops to one-fourth (25.3%) for non-Indigenous communities.

Nearly all (136 out of 138, or 98.6%) communities which were not connected to other CSDs via the main road/ferry network had an IR score greater than 0.40. This observation supports the selection of 0.40 as the desired cut-off, since the chance of geographic isolation of communities (or being remote) which are not connected to the main road/ferry network would naturally be expected to be higher compared to connected ones.

While the methodology developed is generic and is not tied to addressing one problem or application, it is acknowledged that different applications might require their own specialized groupings by remoteness.

6 References

Alessandro Alasia, Frédéric Bédard, Julie Bélanger, Eric Guimond and Christopher Penney (2017). [Measuring remoteness and accessibility: A set of indices for Canadian communities](#). Statistics Canada- Centre for Special Business Project.

Department of Health and Aged Care (2001), [Measuring Remoteness: Accessibility/Remoteness Index of Australia \(ARIA\)](#). Occasional Papers: New Series No. 14. Australian Government.

Hastie, T., Tibshirani, R., & Friedman, J. (2008). The Elements of Statistical Learning: Data Mining, Inference, and Prediction (2nd ed.). Springer

Rajendra Subedi, Shirin Roshanafshar and T. Lawson Greenberg (2020). [Developing Meaningful Categories for Distinguishing Levels of Remoteness in Canada](#). Statistics Canada- Centre for Population Health Data (CPHD). Analytical Studies: Methods and References No. 026.

7 Appendix

Table A-1
Number of CSDs by mode of transportation, by two- and three-class natural clusters on the IR continuum

Two-class	Three-class		
	Non-remote; IR range [0, 0.2717]	Intermediate-remote; IR range [0.2721, 0.5010]	Remote; IR range [0.5014, 1]
CSD counts			
Air			
Non-remote; IR range [0, 0.3791]
Remote; IR range [0.3793, 1]	...	2	45
Combination of air, train, winter road, charter boat and/or seasonal ferry			
Non- remote; IR range [0, 0.3791]	1	1	...
Remote; IR range [0.3793, 1]	...	6	83
Main road/ferry network			
Non- remote; IR range [0, 0.3791]	1,693	1,364	...
Remote; IR range [0.3793, 1]	...	1,058	629
No population			
Non- remote; IR range [0,0.3791]	23	50	...
Remote; IR range [0.3793, 1]	...	61	109
All CSDs			
Non- remote; IR range [0, 0.3791]	1,717	1,415	...
Remote; IR range [0.3793, 1]	...	1,127	866

... not applicable

Note: CSDs are classified into two and three classes based on their IR scores using k-means algorithm.

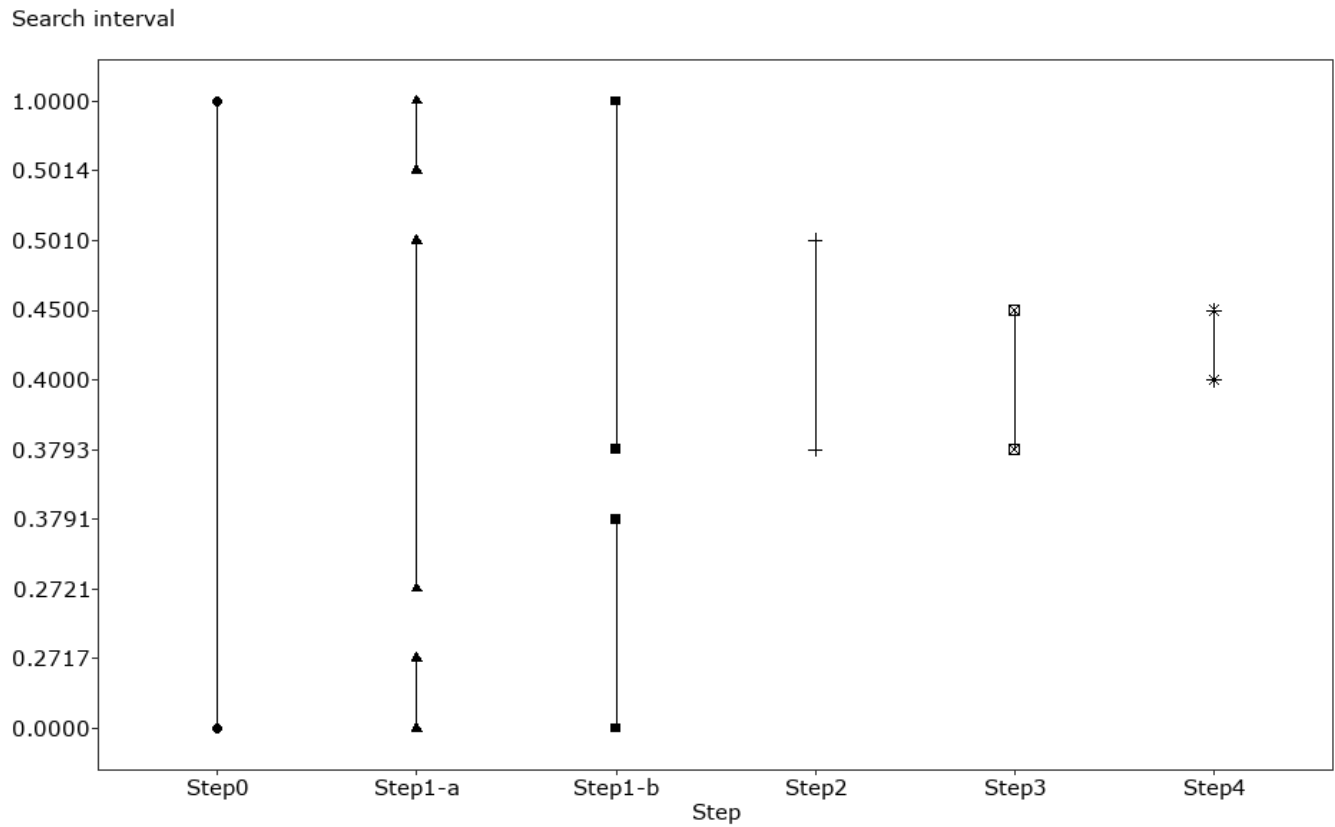
Source: Authors' computation based on Index of Remoteness (IR) scores calculated using the 2016 Census of Population.

The progressively narrowed search intervals within which to explore for the IR cut-off at are displayed in Figure A-1. The following provides an overview of the steps for determining these narrowed intervals:

- Step1-a: Applying k-means algorithm for $k = 3$ to find natural breaks in IR scores and to identify the transitional interval.
- Step1-b: Applying k-means algorithm for $k = 2$ to reduce the search interval based on the overlap between intervals obtained in this step and Step1-a. The search interval is narrowed to [0.3793, 0.5010].
- Step2: Given the information obtained in Step1-a and Step1-b, the search interval in this step is [0.3793, 0.5010]. Focusing on the populated and connected CSDs to the main road/ferry network with IR scores in this interval, this step explores IR scores versus $AggPop_{SurrCSD}$ and $AggPop_{SurrPOPCTR}$ to identify a pattern to narrow the search interval. It results in reducing the search interval from [0.3793, 0.5010] to [0.3793, 0.4500]. The 0.45 is also identified as a potential cut-off.
- Step3: Focusing on populated and connected CSDs with IR scores in [0.3793, 0.4500] and with $AggPop_{SurrCSD}$ and $AggPop_{SurrPOPCTR}$ greater than 164,000 and 87, 000, respectively, the search interval is narrowed to [0.3793, 0.4000]. The reduced search interval is obtained based on the mean and median of IR scores of these CSDs. The 0.40 is also identified as another potential cut-off.
- Step4: The smallest interval which is likely contain the IR cut-off is determined to be [0.4000, 0. 45000], and the IR cut-off is selected to be 0.40.

Note that there is no CSD with IR score within the (0.2717, 0.2721), (0.3791, 0.3793) and (0.5010, 0.5014) intervals.

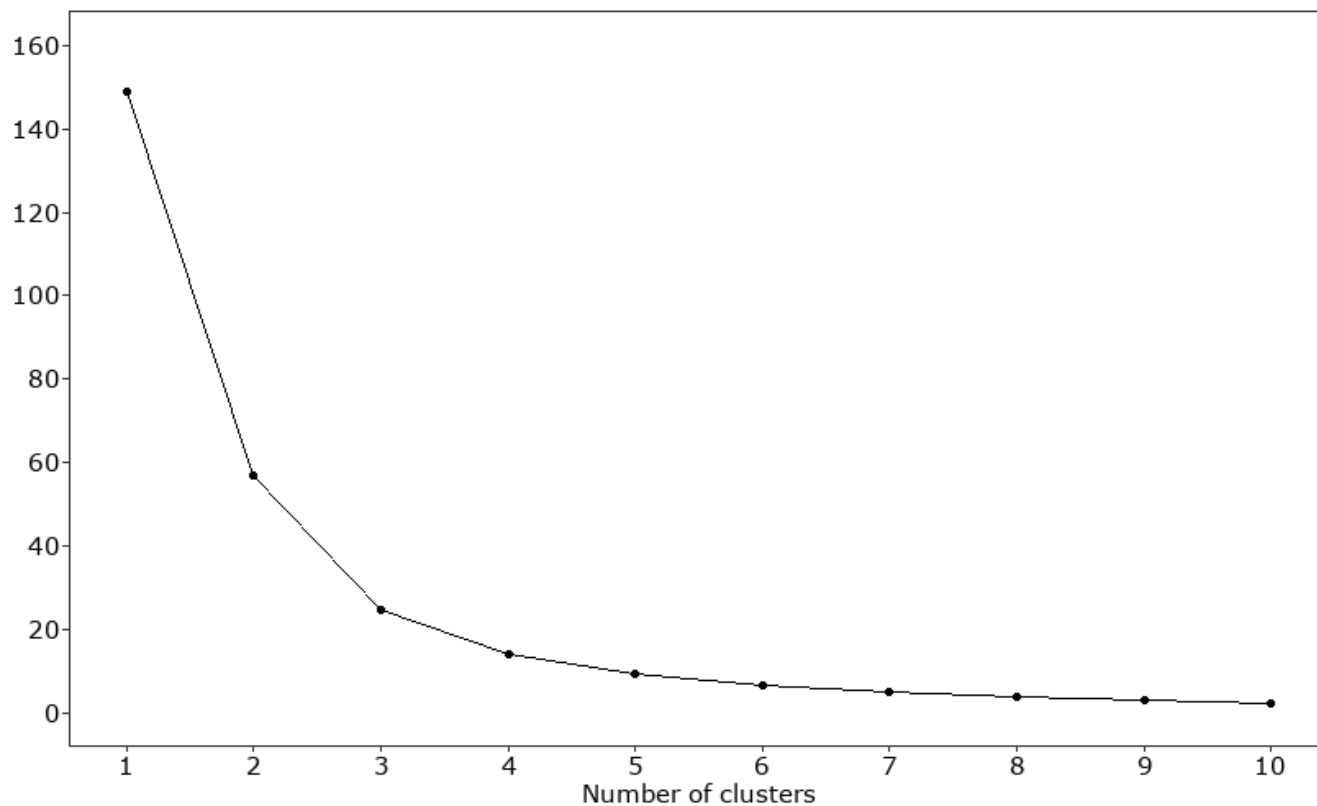
Figure A-1
Summary at each step of narrowing the search interval to identify the IR cut-off



Note: Each interval represents the search interval to identify the IR cut-off. Note that the interval in Step0 displays the range of the IR which varies between 0 and 1.
Source: Authors' computation based on Index of Remoteness (IR) scores calculated using the 2016 Census of Population.

Figure A-2
Optimal number of clusters k-means clustering- Elbow method

Total within sum of squares

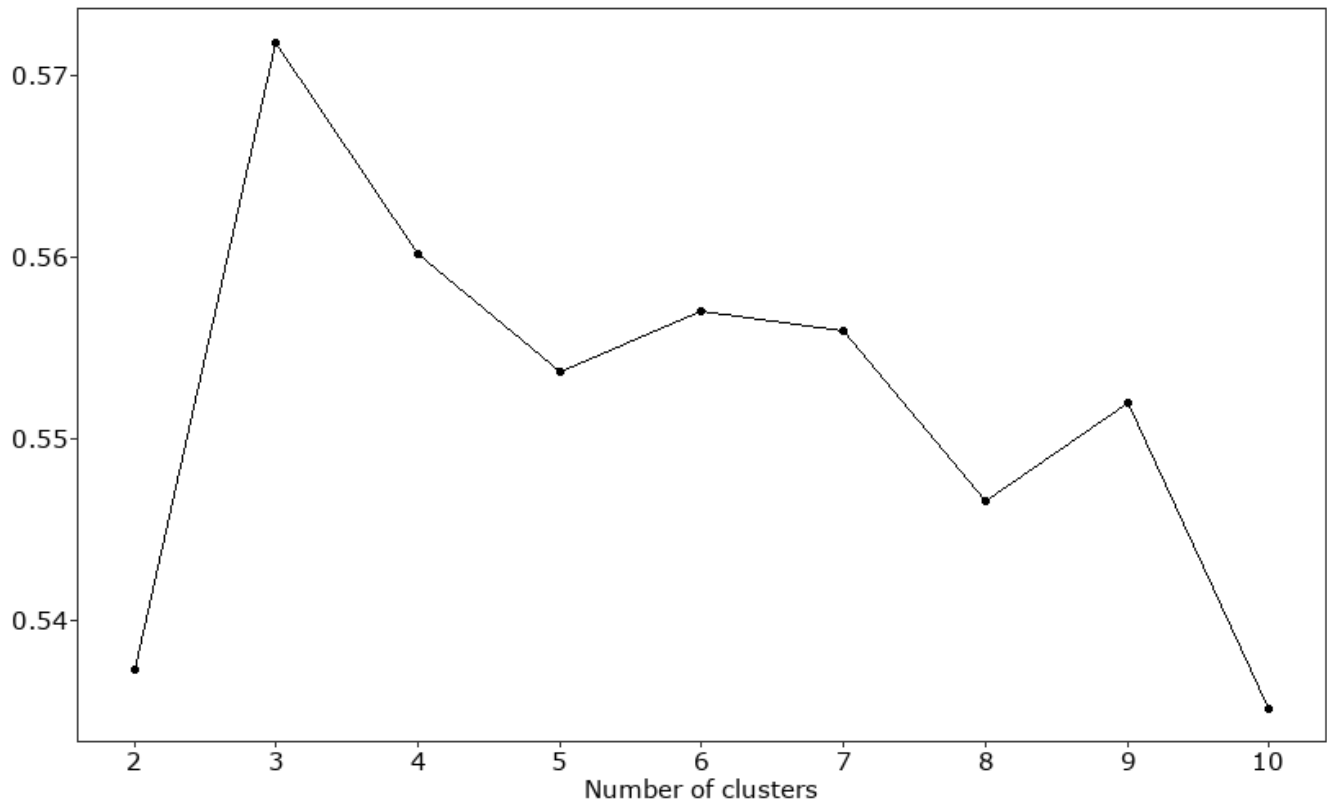


Notes: All CSDs were classified into different number of groups based on their IR scores using the k-means clustering algorithm for $k = 1, \dots, 10$. This Figure displays the total within groups variation for each number of clusters. As seen, the optimal number of clusters is three.

Source: Authors' computation based on Index of Remoteness (IR) scores calculated using the 2016 Census of Population.

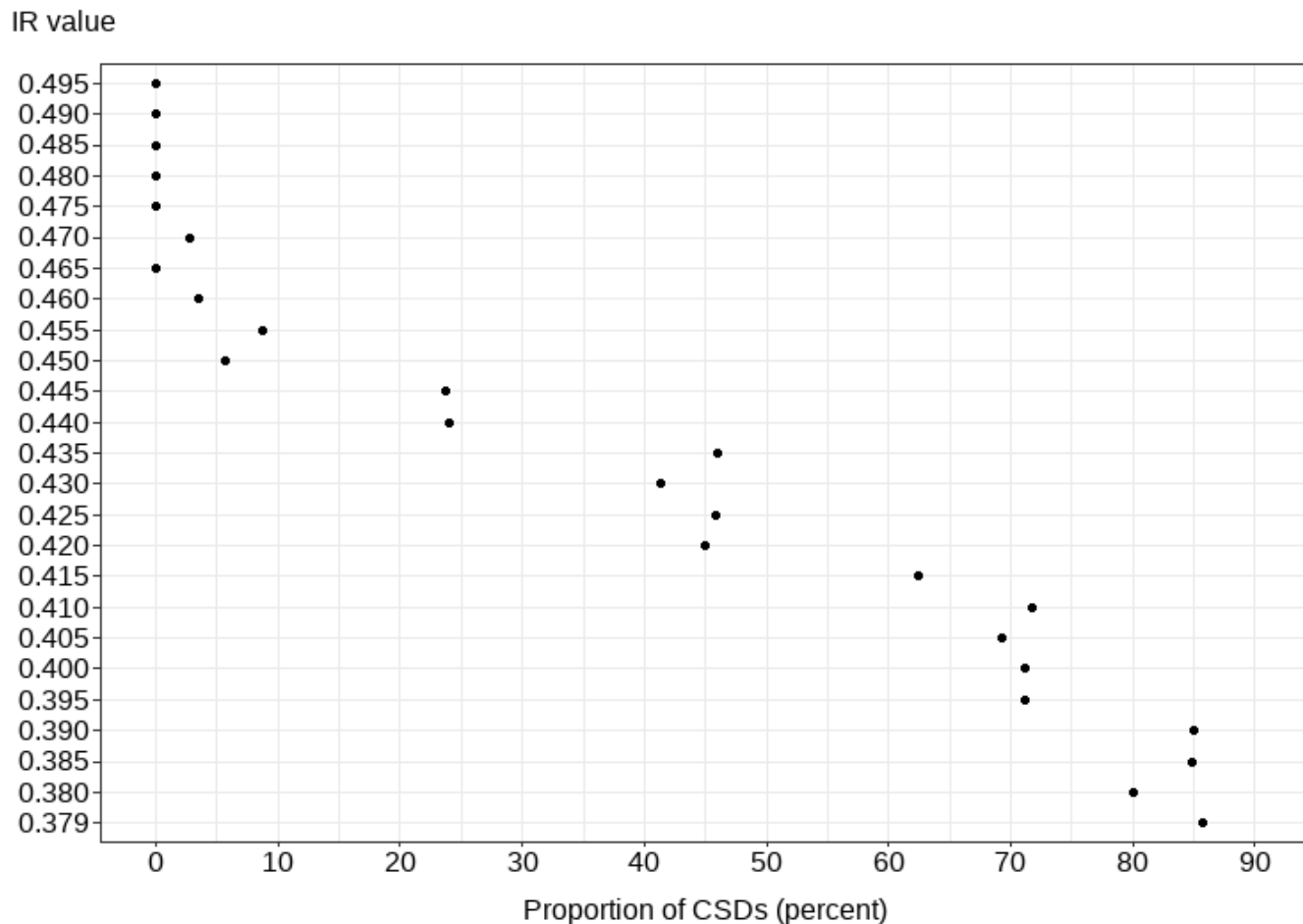
Figure A-3
Optimal number of clusters k-means clustering- Average Silhouette method

Average silhouettes



Notes: All CSDs were classified into different number of groups based on their IR scores using the k-means clustering algorithm for $k = 1, \dots, 10$. This Figure displays the average Silhouette for each number of clusters. As seen, the optimal number of clusters is three.
Source: Authors' computation based on Index of Remoteness (IR) scores calculated using the 2016 Census of Population.

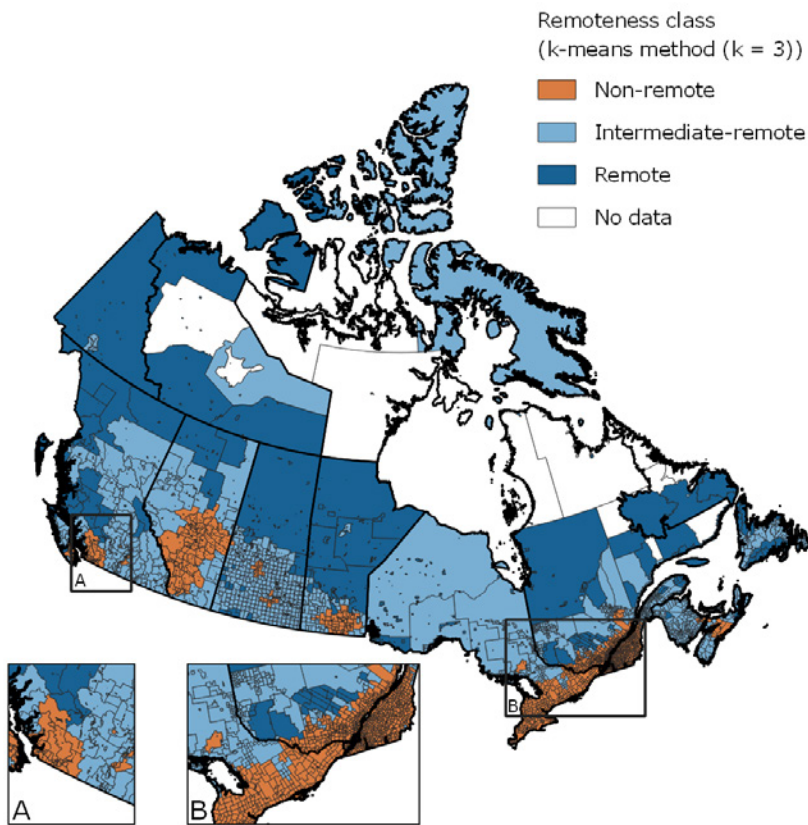
Figure A-4
Proportion of CSDs with aggregate population of surrounding CSDs (i.e., $AggPop_{SurrCSD}$) greater than 139,000, by IR score ranges; IR in [0.3793, 0.5010]



Notes: Only populated and connected CSDs to the main road/ferry network with IR scores in the [0.3793, 0.5010] interval were included to create this Figure. Each value on Y-axis represents an interval with the lower bound equals to the value displayed on the axis, and the upper bound equals to that next value displayed. For instance, 0.405 represents the interval of [0.4050, 0.4100). Each point represents the proportion of populated and connected CSDs with $AggPop_{SurrCSD}$ greater than 139,000 with IR score less than or equal to the IR value displayed in this Figure. For example, given the IR value of 0.455 on the Y-axis, 2 out of 23 (8.6%) CSDs with IR scores within [0.4550, 0.4600) have $AggPop_{SurrCSD}$ greater than 139,000 in their surrounding CSDs within 2.5 hours of travel time.

Source: Authors' computation based on Index of Remoteness (IR) scores calculated using the 2016 Census of Population and the 2016 Census of Population.

Figure A-5
Geographic distribution of remote, intermediate-remote, and non-remote CSDs obtained using k-means method (k=3)



Notes: The white parts in the map represent CSDs with no reported population in 2016 and not connected to the main road/ferry network. CSDs were classified into three classes based on their IR scores using k-means algorithm for k=3.
Source: Authors' computation based on Index of Remoteness (IR) scores calculated using the 2016 Census of Population.