

A REPORT
FOR THE
SEGMENTATION OF STATISTICS CANADA'S PROXIMITY
MEASURES

June 2023

Ricky Heinrich
Noman Mohammad
Avishek Saha
Jonah Edmundson

Statistics Canada Liaison - Jérôme Blanchet, Bjenk Ellefsen
UBCO Project Supervisor - His Excellency Dr. Firas Moosvi

Contents

1	Introduction	1
2	Background	2
3	Data	3
3.1	Primary Dataset	3
3.2	Missing Values	3
3.3	Other Data	3
4	Methods	4
4.1	Data Exploration	4
4.2	Clustering Tendency	4
4.3	Quintiles	4
4.4	Minima Identification	4
4.5	Clustering	4
4.5.1	Number of Clusters	4
4.5.2	Comparison of Algorithms	4
4.5.3	Cluster Profiles	4
5	Results	5
5.1	Data Exploration	5
5.1.1	Summary Statistics	5
5.1.2	Distributions	5
5.2	Clustering Tendency	6
5.3	Quintiles	6
5.4	Minima Identification	6
5.5	Clustering	6
5.5.1	Cluster Profiles	6
6	Discussion	7
6.1	Comparison of Approaches	7
6.2	Interpretation of Cluster Profiles	7
7	Limitations	8
8	Conclusion	9
8.1	Successful Methods	9
8.2	Unsuccessful Methods	9
8.2.1	Univariate Clustering	9
8.2.2	Multivariate Clustering	9
8.3	Extra Plots and Tables	9
9	References	10
A	Appendix	11

List of Figures

1	Comparison of distributions	6
---	---------------------------------------	---

List of Tables

1 Introduction

The

2 Background

3 Data

text

3.1 Primary Dataset

3.2 Missing Values

Symbol	Meaning
.	not available for any reference period
..	not available for a specific reference period
...	not applicable
F	too unreliable to be published

	DBs with Data Available	Percentage
Employment	423,602	86.5
Pharmacy	178,521	36.5
Childcare	243,964	49.8
Healthcare	300,465	61.4
Grocery	141,063	28.8
Pri. Educ.	225,359	46.0
Sec. Educ.	141,213	28.8
Library	112,655	23.0
Parks	234,068	47.8
Transit	181,305	37.0
DB Pop.	487,526	99.6

3.3 Other Data

4 Methods

text

4.1 Data Exploration

4.2 Clustering Tendency

4.3 Quintiles

4.4 Minima Identification

4.5 Clustering

4.5.1 Number of Clusters

4.5.2 Comparison of Algorithms

4.5.3 Cluster Profiles

5 Results

text

5.1 Data Exploration

5.1.1 Summary Statistics

Variable	Counts
DBs Per Province	
<i>Alberta</i>	66,749
<i>British Columbia</i>	52,850
<i>Manitoba</i>	30,669
<i>New Brunswick</i>	14,345
<i>Newfoundland and Labrador</i>	8,756
<i>Northwest Territories</i>	1,495
<i>Nova Scotia</i>	15,279
<i>Nunavut</i>	792
<i>Ontario</i>	133,214
<i>Prince Edward Island</i>	3,639
<i>Quebec</i>	106,251
<i>Saskatchewan</i>	54,118
<i>Yukon</i>	1,519
CMA Type	
<i>CMA (B)</i>	206,709
<i>Untraced CA (D)</i>	53,061
<i>Traced CA (K)</i>	16,992
<i>Not a CMA or CA</i>	212,914
Amenity Dense	
<i>Low Density (0)</i>	442,179
<i>Medium Density (1)</i>	37,303
<i>High Density (2)</i>	4,827
<i>Too unreliable to publish (F)</i>	5,367
Suppressed	
<i>Not suppressed (0)</i>	484,309
<i>Info. Suppressed (1)</i>	5,367

5.1.2 Distributions

example text [Figure 1](#) ...

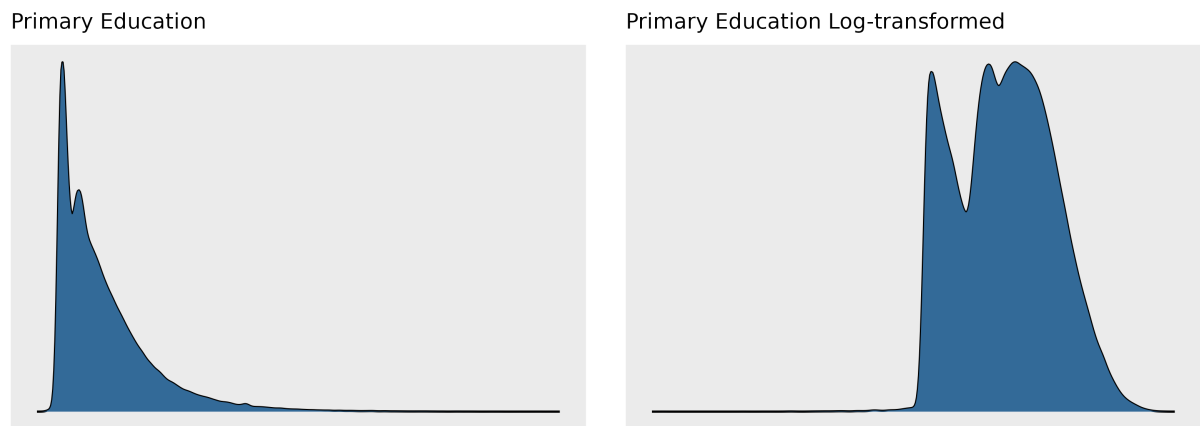


Figure 1: Distribution of the proximity measure to primary education services before and after log-transformation.

5.2 Clustering Tendency

5.3 Quintiles

5.4 Minima Identification

5.5 Clustering

5.5.1 Cluster Profiles

6 Discussion

text

6.1 Comparison of Approaches

6.2 Interpretation of Cluster Profiles

7 Limitations

8 Conclusion

8.1 Successful Methods

8.2 Unsuccessful Methods

8.2.1 Univariate Clustering

8.2.2 Multivariate Clustering

8.3 Extra Plots and Tables

9 References

A Appendix

text

	Employment	Pharmacy	Childcare	Healthcare	Grocery	Pri. Educ.	Sec. Educ.	Library	Parks	Transit	DB Pop.
1 Dec.	1e-04	0.0075	0.0079	2e-04	0.0144	0.0319	0.0374	0.0508	0.0127	0.0011	0
2 Dec.	4e-04	0.0098	0.0152	7e-04	0.0221	0.0416	0.0421	0.0558	0.0203	0.0026	0
3 Dec.	0.0013	0.0146	0.0241	0.0018	0.0289	0.0582	0.0485	0.0624	0.0278	0.0045	5
4 Dec.	0.003	0.0193	0.0348	0.0032	0.0348	0.072	0.0586	0.0707	0.0372	0.0067	16
5 Dec.	0.0065	0.0256	0.0476	0.005	0.0434	0.09	0.0745	0.0814	0.0481	0.0094	29
6 Dec.	0.0127	0.0341	0.0636	0.0074	0.0555	0.1105	0.091	0.096	0.0614	0.0131	45
7 Dec.	0.0217	0.0457	0.0846	0.0111	0.0719	0.1366	0.1141	0.1168	0.0793	0.0184	66
8 Dec.	0.0368	0.0641	0.1167	0.0184	0.0985	0.172	0.1492	0.1488	0.105	0.0272	100
9 Dec.	0.0726	0.0983	0.1751	0.0343	0.154	0.233	0.2128	0.2106	0.1494	0.0442	173
Min.	0	0	0	0	1e-04	4e-04	5e-04	1e-04	0	0	0
Median	0.0065	0.0256	0.0476	0.005	0.0434	0.09	0.0745	0.0814	0.0481	0.0094	29
Mean	0.02541	0.04438	0.07584	0.01372	0.06991	0.11617	0.104	0.11462	0.0692	0.01805	72
Max.	1	1	1	1	1	1	1	1	1	1	7607
Std. Dev.	0.0491	0.0579	0.0874	0.0279	0.0783	0.0917	0.0869	0.0978	0.0685	0.027	146
Skew	4.656	4.555	2.807	7.041	3.201	1.963	2.462	3.439	2.824	5.692	8
Kurtosis	38.08	37.81	14.82	95.45	17.83	8.72	11.84	18.48	17.2	72.96	152

	Counts	Percentages	Log Transformed Counts	Log Transformed Percentages
Employment	45,390	9.27	0	0.00
Pharmacy	13,416	2.74	478	0.10
Childcare	15,397	3.14	140	0.03
Healthcare	31,007	6.33	50	0.01
Grocery	11,904	2.43	794	0.16
Pri. Educ.	10,205	2.08	98	0.02
Sec. Educ.	8,683	1.77	215	0.04
Library	8,867	1.81	2,295	0.47
Parks	12,703	2.59	910	0.19
Transit	14,165	2.89	3,596	0.73