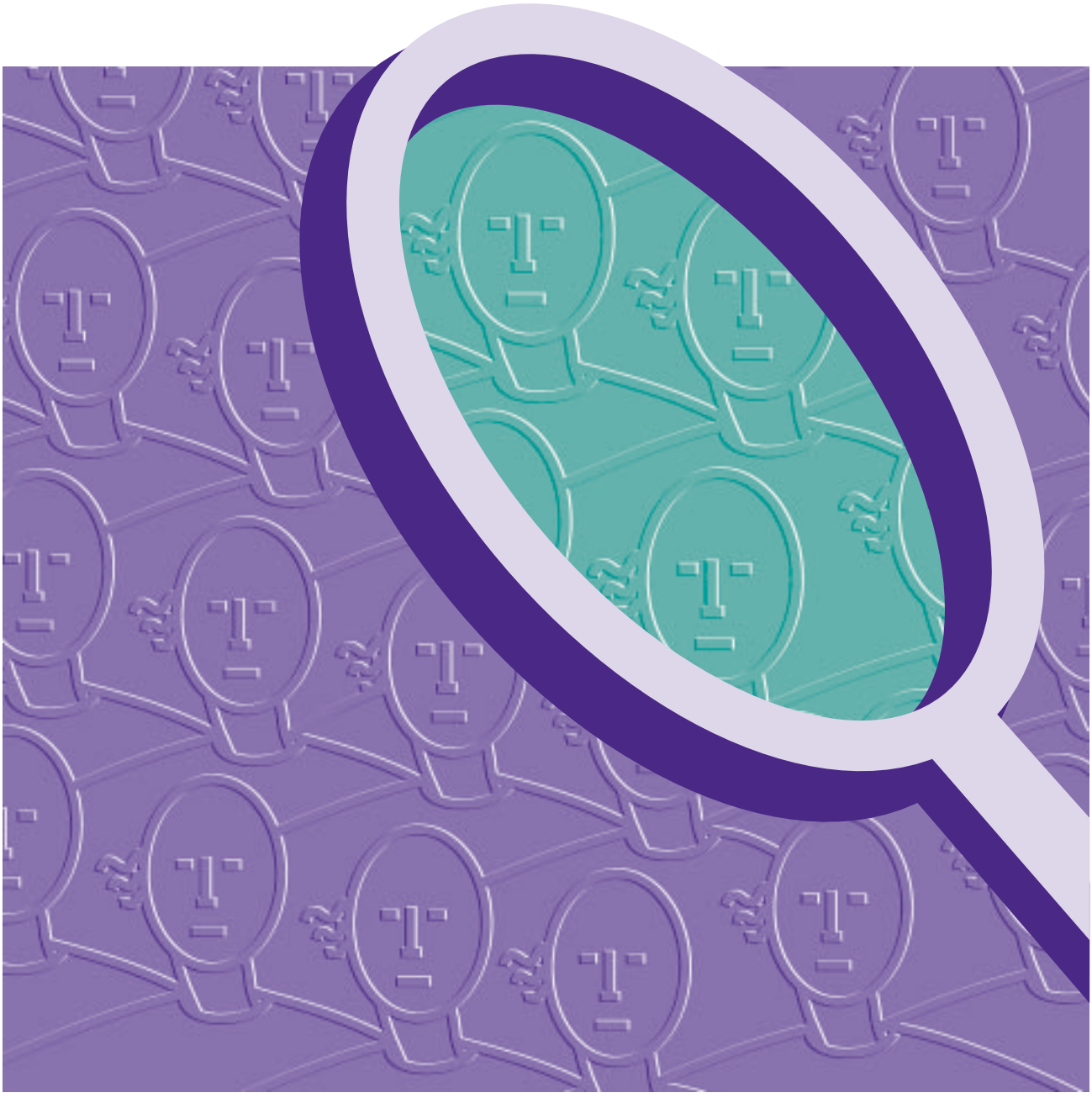
Health Region Peer Groups



Statistics Canada
Statistique Canada

Canada

Health Region Peer Groups

Larry MacNabb
Health Statistics Division
Statistics Canada

Phone: (613) 951-4269
Email: larry.macnabb@statcan.ca

Preface

This paper was developed with considerable input and guidance from the Peer Group Project Working Group comprised of members from Statistics Canada, provincial ministries of health, health regions, universities and the Canadian Institute for Health Information.

- Lorna Bailie – Statistics Canada, Assistant Director Health Statistics Division, Working Group Chair
- Yves Beland – Statistics Canada, CCHS Senior Methodologist
- Ray Bollman – Statistics Canada, expert researcher in rural Canadian issues
- Dr. David Elliott – Nova Scotia Ministry of Health, past member of the NPHS/CCHS Advisory Committee
- Jason Gilmore – Statistics Canada, Analyst, Health Indicators Project, Health Statistics Division
- Larry MacNabb – Statistics Canada, CCHS Dissemination Manager
- Jean Marie-Berthelot – Statistics Canada, Manager - Health and Analysis Modelling Group
- Dr. Cory Neudorf – Director of Strategic Health Information and Planning, Saskatoon District Health, member of the NPHS/CCHS Advisory Committee
- Roger Pitblado – Laurentian University, Department of Geography and Centre for Rural and Northern Health Research.
- Indra Pulcins – Canadian Institute for Health Information
- Nancy Ross – McGill University, Department of Geography
- Donald Schopflocher – Alberta Health and Wellness, biostatistician and member of the NPHS/CCHS Advisory Committee
- Sylvain Thivierge – Statistics Canada, CCHS Methodologist
- Brenda Wannell - Statistics Canada, Manager, Health Indicators, Health statistics Division
- Jennifer Zelmer – Canadian Institute for Health Information

Table of Contents

List of Tables

List of Figures

1. Introduction

The inception of the Canadian Community Health Survey in conjunction with the expansion of existing data products for the provision of health region[1] level information has necessitated the need to develop a method of comparing regions with similar socio-economic determinants of health. This paper will present one possible method for placing health regions into groupings with similar socio-economic characteristics or "peer groups". After the effects of the various social and economic characteristics known to influence health status have been removed it is then possible to compare regions by measures of health status and compare the relative effectiveness of health promotion and prevention activities across regions.

Development of the criteria used to define peer groups required careful consideration of their intended use. The requirement that peer groups be used as a method for comparing health related issues ultimately eliminated all variables directly describing health as potential candidates for analysis. Further, it was required that all variables used must be reliable and available for all health regions.

The need for objectivity required that peer groups be developed using empirical techniques. Consideration of the need for simplified and relevant comparison also required that peer groups have approximately 5 to 10 health regions per grouping and that there be representation across the country within each peer group.

In the actual application of the above parameters several limiting factors resulted which required some minor modifications. All criteria were followed to the extent possible and any deviations are explained in detail.

2. Data

In order to achieve maximum statistical differentiation between health regions, variables were chosen to cover as many of the social and economic determinants of health as possible. Areas covered include:

Population Change
Demographic Structure
Social Status
Economic Status
Ethnicity
Aboriginal Status
Housing
Urbanisation/Metropolitan Influence
Income Inequality
Labour Market Conditions

When exploring potential data sources the Census of Canada was a logical starting point. Census information is readily available at various levels of geography and covers a broad range of topics. Thus it is possible to represent the majority of the broad topic areas, for all regions, meeting the main requirement that all health regions receive equal variable coverage.

A complete list of variables and their definitions are included as Appendix A.

[1] *Health regions are defined by provincial governments as the areas of responsibility for regional health boards (i.e., legislated) or as regions of interest to health care authorities.*

3. Methodology

3.1. Clustering Methods

Generally speaking cluster analysis attempts to organise variables or observations into distinct groups based on a measure of their distance from each other or a distinct point in p-dimensional space, where p represents the number of variables used to describe each observation. There are two basic types of cluster analysis: hierarchical and non-hierarchical.

Hierarchical methods organise a set of N observations into a series of m clusters, where m can range from 1 (a single cluster containing all observations) to N (each observation occupying its own cluster). The ability to split a set of observations into a series of groups having higher degrees of similarity is one of the strengths of this class of algorithms. However, the ability to adapt to relationships uncovered in the data at later stages is one of their fundamental weaknesses (Andberg, 1973).

Non-hierarchical algorithms attempt to partition a set of observations into a pre-defined set of groups using a specified optimisation criterion. This approach appeared best suited to meeting the original objectives of the peer group project. Consider the task of splitting N observations into *k* clusters, where each cluster k has mean $\bar{x}_k$ and varying size n_k. Many of the most commonly used clustering algorithms attempt to minimise the within cluster sum-squared error denoted by:

Equation 1

$$E = \sum_{i=1}^{k} \sum_{j=1}^{j=n_i} (x_{ij} - \bar{x}_i)^2$$

(Everitt, 1993).

3.2. K-Means Clustering

The SAS procedure FASTCLUS uses a k-means algorithm to assign observations to a pre-defined set of k clusters. A description of k-means clustering and several variants of the method can be found in Andberg, 1973. The basic steps for placing observations into k clusters are as follows:

I. Select k observations as cluster seeds.
II. Observations are assigned to the nearest cluster seed. After all observations are assigned, cluster seeds are replaced by their respective cluster means. This step is repeated until the change in cluster seeds becomes or approaches zero.
III. Form final clusters by assigning each observation to its nearest cluster seed.

Complete details of the FASTCLUS procedure can be found in the SAS OnlineDoc®, Version 8.

3.3. Number of Clusters

One of the major problems with cluster analysis is selecting the appropriate number of clusters. Several criteria have been suggested (Everitt, 1993) which generally involve the optimisation of one or more test statistics. From a practical perspective it is generally left up to the analyst to determine the number that best suits a given need. For the purpose of this analysis a maximum number of 20 clusters[2] was deemed a practical starting for comparing health regions. This would give an average number of 7 health regions to a cluster in line with the original study objectives.

[2] *To remain consistent with the statistical exercise of defining peer groups through cluster analysis the term cluster will be used instead of peer group.*

4. Results

4.1. Standardisation of Variables

In order to mitigate the effect of differing variance among variables all were standardised to mean 0, unit variance. The two variables measuring the proportion of low income amongst persons 15 + and low income children could not be calculated for some of the more remote health regions. One advantage of the fastclus procedure is that it has the ability of imputing values based on characteristics of other cluster members. For the purpose of the principal component analysis missing values were replaced with the mean value of 0 for these variables.

4.2. Initial Clusters

In order to determine a starting point for analysis the clustering algorithm was instructed to group the 139 health regions into a set of 20 clusters. In total, 6 health regions were placed into a cluster with only themselves as a member (Table 4.2-1). This indicated that 20 clusters would be too many, given the spread of data used in the analysis and considering the objective of comparing regions within a cluster. Overall clusters appear to be equally spaced with approximately equal within cluster variation.

In order to simplify the exercise of visualising 24 dimensions, health region clusters were plotted against the first two principal components of the 24 by 24 correlation matrix derived from the 24 variables used in the analysis (Figure 4.2-1). Overall, the first two principal components account for approximately 53% of the total variability. The first principal component appears to be an overall measure of urbanicity, while the second is highly weighted on variables associated with poverty. A summary of the results from the principal component analysis including all eigenvalues and the first six eigenvectors for the correlation matrix can be found in Appendix B.

Health regions have been identified by their cluster membership from Table 4.2-1 and by colour. Health regions assigned to their own cluster have been identified using larger symbols. Visually the clusters appear to be evenly spread with minimal overlap. This is and indication that the first two principal components explain most of the variation in the data and serve as a useful tool for cluster assessment. Any overlap is the result of other factors influencing the assignment to a specific cluster and is explored further in the general discussion.

Ideally given a set number of groupings all clusters should be separated by approximately the same distance. A plot of cluster frequency versus cluster radius and distance between clusters provides a good indication of the appropriate cluster spacing given the number of clusters required (Figure 4.2-2). It appears that setting a cluster radius of 5 would ensure adequate spacing between clusters and ensure that most health regions can be assigned to a cluster while minimising the effect of outlying health regions.

Table 4.2-1: Initial classification of health regions into 20 clusters.

Cluster	Frequency	RMS Std	Radius	Nearest Cluster	Distance Between Cluster Centers
A	2	0.67	2.32	G	7.03
B (4711 – Northern Health Services Branch (K) Service Area)	1	..	-	D	6.74
C	7	0.56	3.03	R	4.06
D	2	..	2.29	S	5.18
E	3	0.59	2.79	M	4.50
F (1006 – Health Labrador Corporation)	1	..	-	I	4.46
G (5916 – Vancouver)	1	..	-	O	5.78
H	13	0.58	3.26	J	2.85
I	8	0.52	3.82	R	3.21
J	21	0.46	3.03	H	2.85
K	22	0.54	3.81	Q	2.40
L (5918 – North Shore)	1	..	-	C	5.60
M	6	0.49	3.08	E	4.50
N	2	..	2.96	D	6.34
O	2	0.51	1.76	G	5.78
P (4817 – Northwestern Regional Health Authority)	1	..	-	I	4.83
Q	34	0.49	3.35	K	2.40
R	8	0.53	3.27	Q	3.11
S	3	0.62	2.85	I	4.39
T (4690 – Churchill)	1	..	-	S	4.93

Figure 4.2-1: Plot of initial set of 20 health region clusters by first two principal components.

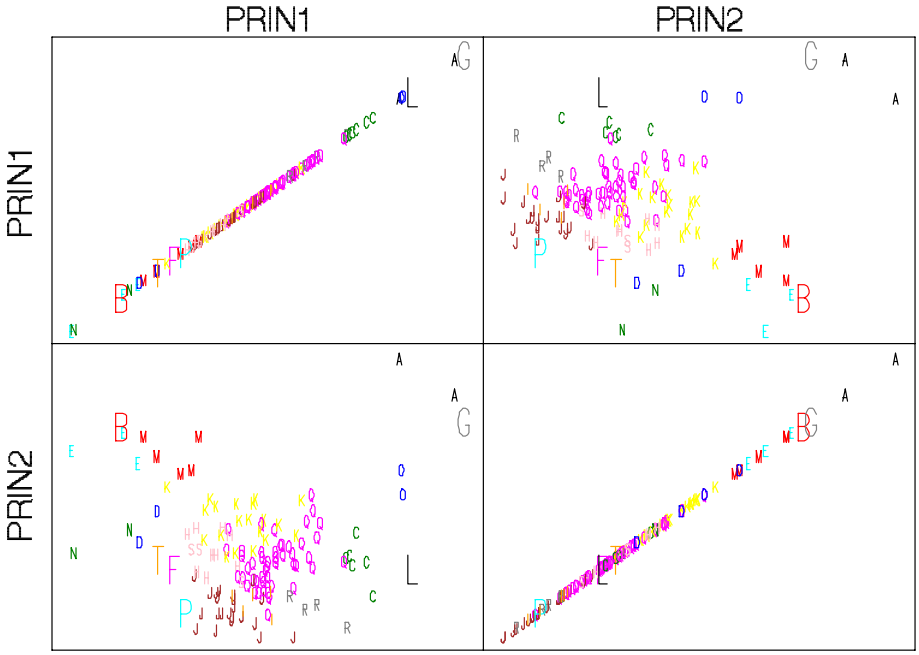
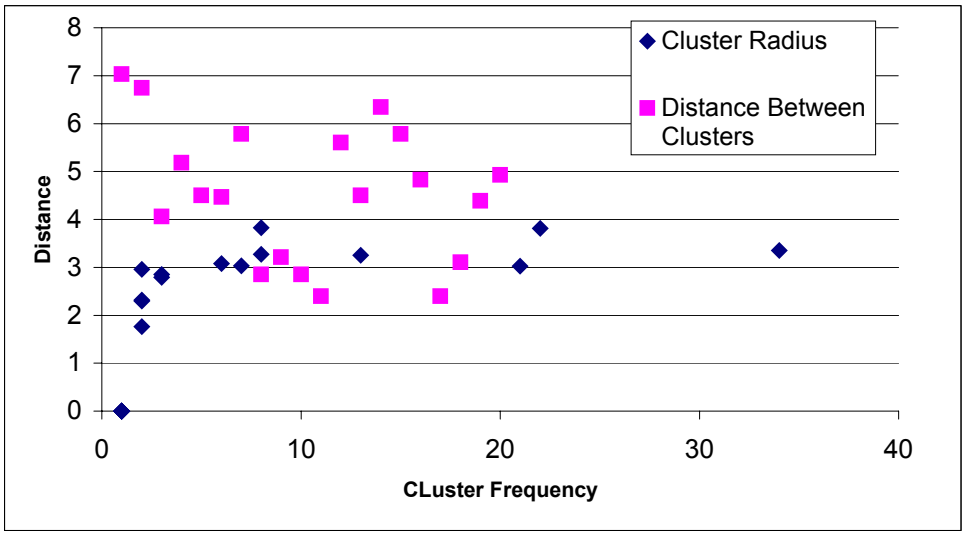


Figure 4.2-2: Cluster frequency versus radius and distance between clusters for the initial 20 health region peer groupings.



4.3. Exclusion of Outliers

The FASTCLUS procedure was rerun using the cluster means from Table 4.2-1 for all clusters with a frequency greater than one. Further, any health region with a distance of five or greater to its nearest cluster center was excluded from the analysis and assigned to the nearest cluster after the analysis was completed.

Table 4.3-1 provides a result summary from the analysis. The results are similar to those presented in the first attempt with generally equal cluster separation and within cluster variance. Only three health regions were excluded from the analysis due to their exceeding the distance criteria to the nearest cluster center.

Figure 4.3-1 shows the cluster results for each health region against the first two principal components. Health regions that did not meet the distance criteria with nearest cluster for these health regions are identified by a larger symbol. The largest clusters are grouped in the center and there appears to be minimal overlap between clusters. Looking at the first two principal components, it is obvious that for some health regions there are other factors influencing assignment of health regions to a cluster than what is captured in the first two principal components. During the final step excluded health regions were combined with their nearest cluster. Overall this had minimal effect on the final cluster statistics (Table 4.3-2).

Table 4.3-1: Maximum cluster radius of 5 using cluster means for groupings from **Table 4.2-1** with more than one health region as seeds.

Cluster	Frequency	RMS Std	Radius	Nearest Cluster	Distance Between Cluster Centers
A	2	0.67	2.32	K	9.05
B	7	0.56	3.03	M	4.06
C	2	.	2.29	N	4.75
D	3	0.59	2.79	I	4.50
E	13	0.58	3.53	G	2.85
F	10	0.62	4.83	M	3.61
G	21	0.46	3.05	E	2.85
H	22	0.54	3.91	L	2.40
I	6	0.49	3.08	D	4.50
J	2	.	2.96	C	6.34
K	2	0.51	1.76	B	5.81
L	34	0.49	3.39	H	2.40
M	8	0.53	3.20	L	3.11
N	4	0.71	3.69	F	4.28

Unassigned

-K (5916 – Vancouver)

-C (4711 – Northern Health Services Branch (K) Service Area)

-B (5918 – North Shore)

Figure 4.3-1: First two principal components versus cluster, with cluster radius of five identifying excluded health regions.

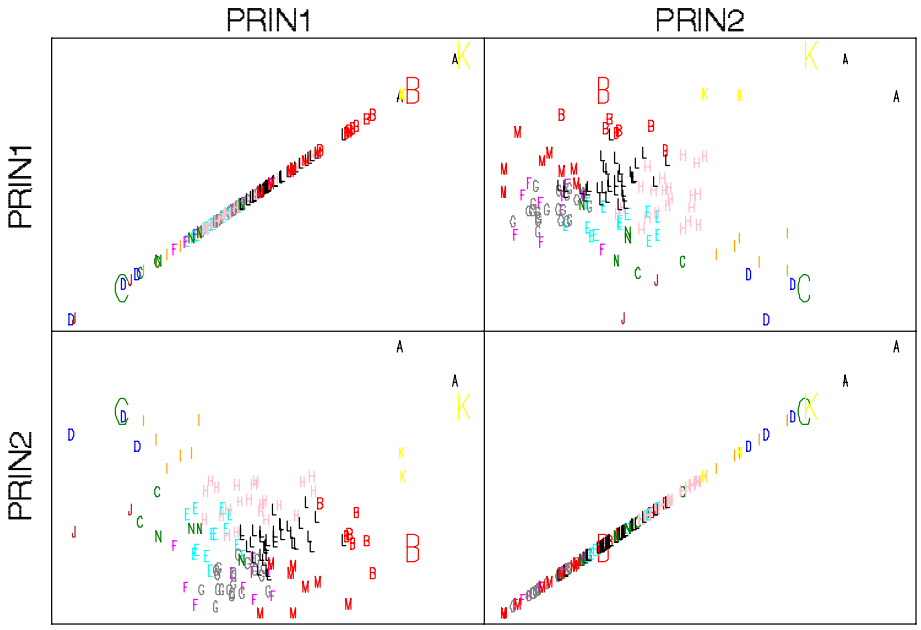


Table 4.3-2: Cluster statistics after combining excluded health regions with their closest cluster.

Cluster	Frequency	RMS Std	Radius	Nearest Cluster	Distance Between Cluster Centers
A	2	0.67	2.32	K	7.98
B	8	0.66	5.60	M	4.12
C	3	0.99	6.74	N	5.57
D	3	0.59	2.79	I	4.50
E	13	0.58	3.26	G	2.85
F	10	0.62	4.35	M	3.61
G	21	0.46	3.03	E	2.85
H	22	0.54	3.81	L	2.40
I	6	0.49	3.08	D	4.50
J	2	.	2.96	C	6.96
K	3	0.77	5.78	B	6.82
L	34	0.49	3.35	H	2.40
M	8	0.53	3.27	L	3.11
N	4	0.71	3.69	F	4.28

5. Discussion

5.1. Lack of Cluster Homogeneity on First Two Principal Components

Scrutiny of Figure 4.3-1 indicates that some clusters may be further defined by information other than that, which is explained in the first two principal components. Specifically, clusters that graphically appear to be problematic include clusters C, F, K and N. In order to ensure that the cluster analysis is assigning health regions to appropriate clusters, the final clusters were plotted against the third, fourth, fifth and sixth principal components.

Initial observation of Figure 4.3-1 would lead towards the conclusion that health region 4711 (Northern Health Services Branch (K) Service Area) denoted by the large C would be closer to cluster D. From Figure 5.1-1 we see that cluster C is distinctly separated from D on both the third and fourth principal components. Loosely interpreted, the third principal component is highly weighted on variables associated with age while the fourth principal component is highly weighted on employment and wealth. These two principal components also distinctly separate cluster F from G and N from E.

The only other glaring irregularity is observed for health region 5916 (Vancouver) denoted by the large letter K. From Figure 4.3-1 it would appear that this health region is closer to cluster A which is represented by Toronto and Montreal (Appendix D). It is only when plotted against the fifth and sixth principal components (Figure 5.1-2) that the true picture arises and we see that the Vancouver health region does belong in cluster K, which contains the Burnaby and Richmond health regions. It appears that the fifth principal component, which is highly weighted on variables associated with immigrants and visible minorities, is the difference.

Figure 5.1-1: Final health region clusters plotted against their third and fourth principal components.

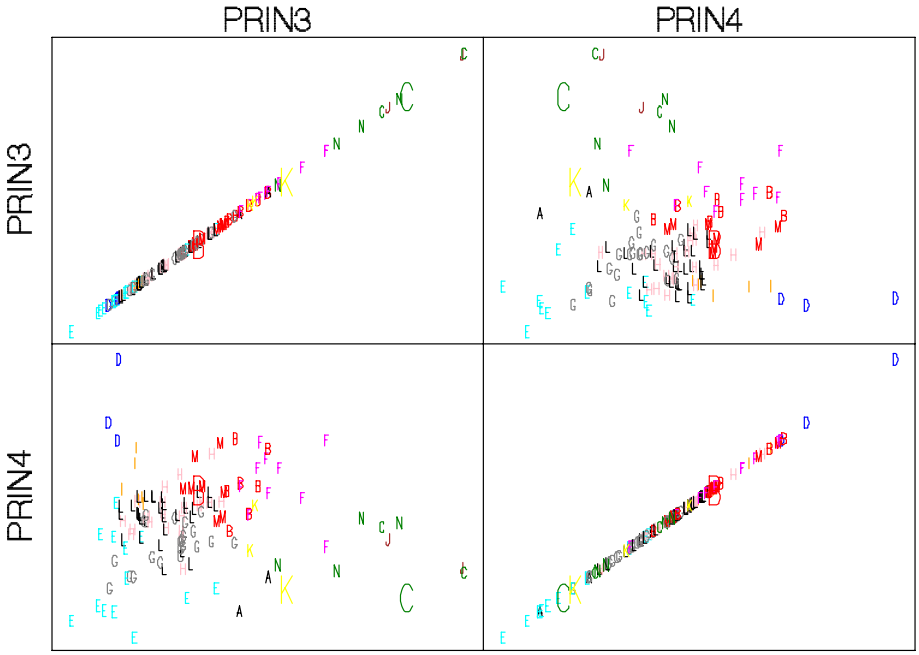
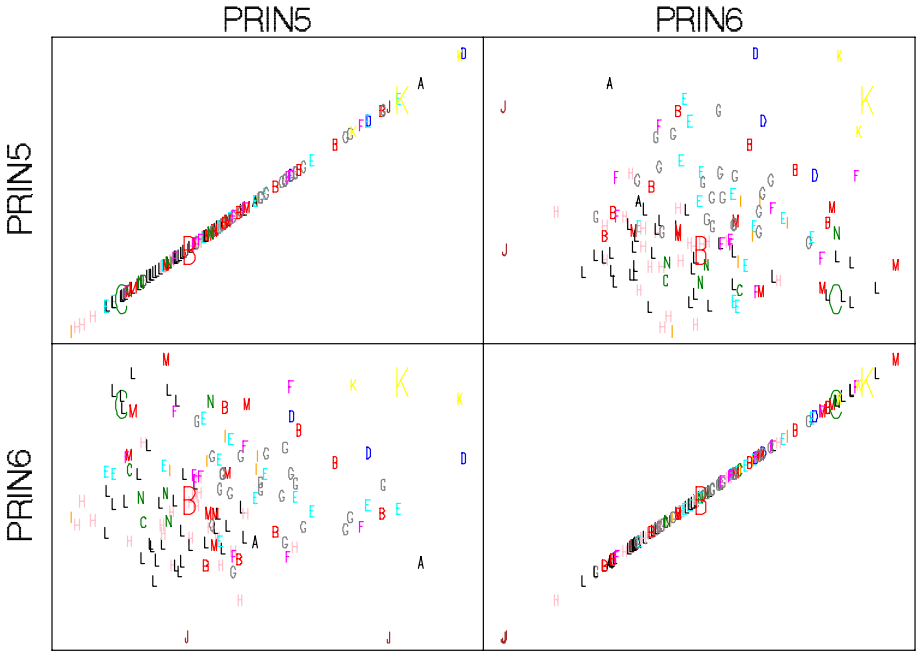


Figure 5.1-2: Final health region clusters plotted against their fifth and sixth principal components.



5.2. Strongest Predictors

In order to determine which variables played a key role in defining the health region peer groups, the final clusters were run against all 24 variables in a stepwise descriminant analysis. Partial R-SQ statistics for entry and removal were set at 0.15. Any variable which had an R-SQ of 0.5 or higher when regressed against a variable already in the model was removed from the analysis. A summary of the results can be found in Table 5.2-1.

The strongest predictors of the final peer groupings are Aboriginal Percentage and Percent Visible Minority. Looking at correlations with other variables, Aboriginal percentage is highly correlated with youthfulness and Visible Minority appears to be highly correlated with factors defining urban centers. The remaining factors in descending order that seem to define the clusters include unemployment, size, age, income inequality and growth.

Table 5.2-1: Stepwise descriminant analysis of final health region groupings on all 24 variables.

Step	Variable	Partial R-SQ	R-SQ Variables in Model
1	**AboPer**	**0.9075**	..
	Removed		
(Aboriginal)	Pop15	0.8168	0.6471
2	**% VisMin**	**0.8952**	..
	Removed		
	AvgDwl	0.6969	0.6093
	PopDen	0.8425	0.5726
(Urban)	HouAff	0.6393	0.5338
	OwnDwl	0.5852	0.5070
	AvgSchl	0.6634	0.5728
3	**Unepm**	**0.8127**	..
	Removed		
	Emp	0.7971	0.8847
(Unemployment)	LTUnemp	0.7827	0.7903
	GovTran	0.8046	0.5957
4 (Size)	**Pop96**	**0.7369**	..
5	**Pop65**	**0.6093**	..
	Removed		
	MFRat	0.5520	0.6709
	AvgInc	0.5595	0.5888
(Age)	LowKids	0.4910	0.5729
	Growth	0.4643	0.5277
	Low15	0.4818	0.5212
	LnePrnt	0.4570	0.5085
6	**IncInq**	**0.5121**	..
(Income	Removed		
Inequality)	MIZ	0.4216	0.5034
7	**MigMob**	**0.2803**	..
	Removed		
(Growth)	ImmPer	0.2707	0.5025

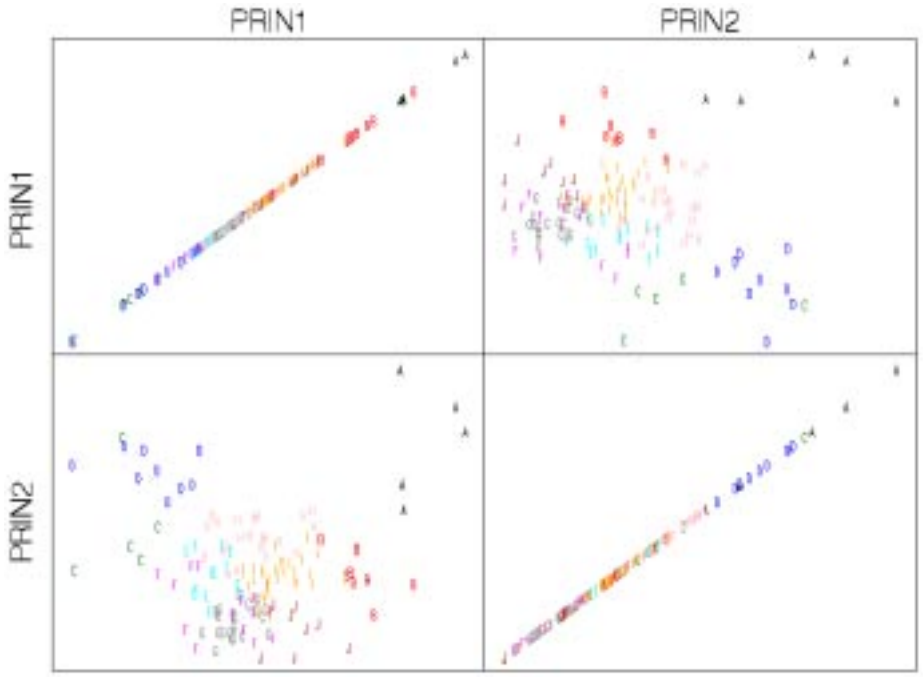
5.3. Collapsing Small Clusters

The results from section 4.3 represent clusters that are evenly spaced and have minimal within cluster variance given the parameters used by the clustering algorithm. From a practical perspective having a cluster with less than five health regions does not provide many options for comparison. In order to provide more peers for comparison, clusters with less than five members were combined with their nearest neighbour.

Specifically, cluster A, which is comprised of regions 2406 (Région de Montréal-Centre) and 3595 (City of Toronto Public Health Unit) was combined with cluster K consisting of regions 5916, 5917 and 5919 or Vancouver, Burnaby and Richmond respectively. Cluster J, consisting of regions 2417 (Région du Nunavik) and 2418 (Région des Terres-Cries-de-la-Baie-James) was combined with cluster C which is made up of regions 4680, 4711 and 6201 or Burntwood, Northern Health Services Branch (K) Service Area and Nunavut respectively.

Finally, clusters N and F, consisting of mostly Northern health regions in western Canada, were combined as were clusters D and I made up of health regions on the east coast. The resulting clusters are plotted against the first two principal components (Figure 5.3-1). Summary statistics for each peer grouping can be found in Appendix C and a list of health regions in each peer group can be found in Appendix D along with their respective groupings through each step of the analysis.

Figure 5.3-1: Final 10 peer groupings plotted against first two principal components.



5.4. Peer Group Descriptions

The seven final variables from the stepwise descriminant analysis were used to represent each of the factors, which appear to differentiate among health regions. The mean values for each peer grouping can be found in Appendix C. For each of the seven variables, the median, 1st and 3rd quartiles of the mean peer group estimates were calculated. Values were classified based on the following ranges.

High: X > Median + 1.5*Interquartile Range
Medium: Median + 1.5*Interquartile Range >= X > Median
Low: Median >= X > Median – 1.5*Interquartile Range
Very Low: Median – 1.5*Interquartile Range >= X

The results from the classification can be found in Table 5.4-1. Though not perfect, this appears to simplify the task of describing each of the final 10 peer groups. For each peer group, only the variables required to distinguish them from another peer group are reported. For example, peer group J is the only peer group with medium Aboriginal and medium urban concentration.

Table 5.4-1: Final peer grouping descriptions based on seven factors resulting from the stepwise descriminant analysis.

	Aboriginal	Urban	Unemploy-ment	Size	Age	Income Inequality
A	Low	High	Medium			
B	Low	High	Low			
C	High	Low				
D	Low	Low	High			
E	Medium	Low	Low	Low	Medium	Low
F	High	Medium				
G	Medium	Low	Low	Low	Medium	Medium
H	Low	Low	Medium			
I	Low	Medium	Low			
J	Medium	Medium				

5.5. Geography Limitations

Each province and territory defines the geographic boundaries for a health region based on administrative preference and in some instances are composed of several smaller administrative areas to ensure that sample survey estimates will attain a sufficient coefficient of variation to be reportable. This is one of the major limiting factors affecting the peer grouping exercise.

Health regions can be strictly urban or rural or some combination of the two. This lack of homogeneity in defining health region boundaries makes the exercise of assigning health regions to peer groups much more difficult as it can have a large impact on the degree to which a variable represents a specific region and in some cases important defining factors may be missed.

5.6. Values Bordering Clusters

Ideally, when searching for the closest statistical comparators, an analyst in a specific health region would measure the statistical distance from all health regions to their own and use a selection of the closest for comparison purposes. In order to validate the final set of peer groupings, the five closest neighbours for each health region, based on euclidian statistical distance, were identified. For P variables, the euclidian distance between points x_i^p and x_j^p is defined by:

Equation 2

$$D = \sqrt{\sum_{k=1}^{p} (x_i^k - x_j^k)^2}$$

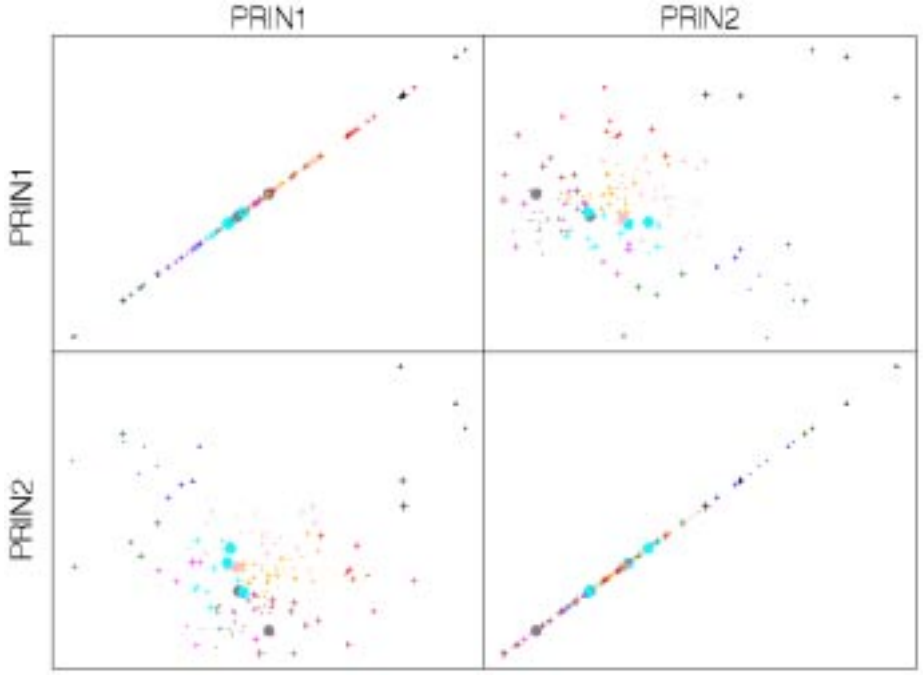
For each health region a count of the five closest health regions that were included in the same peer group was recorded. The results are summarised in Table 5.6-1. In total over 86% of the 139 health regions are assigned to a peer group containing at least 2 of their closest statistical neighbours.

A plot of the final peer groups by the first two principal components and the number of nearest statistical neighbours is shown in Figure 5.6-1. Health regions in clusters without any of their closest neighbours are identified by a solid circle, while the remaining health regions are identified with +'s of varying size. The larger the + the fewer closest neighbours in the same peer group. All of the health regions having 0 neighbours in the same peer group are located on the boundaries of the largest peer groups.

Table 5.6-1: Number of closest neighbours included in the same peer group.

Number of Closest Neighbours	Freq.	Percent	Cumulative Percent
5	44	31.65	31.65
4	23	16.55	48.20
3	33	23.74	71.94
2	20	14.39	86.33
1	13	9.35	95.68
0	6	4.32	100.00

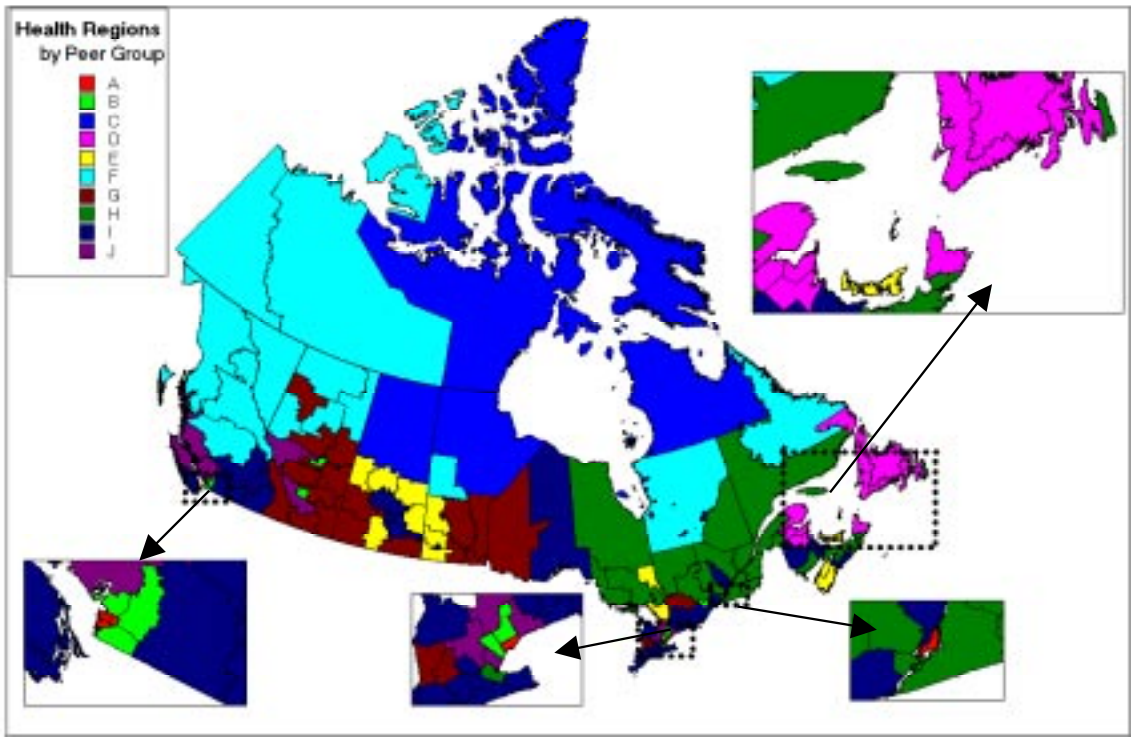
Figure 5.6-1: Plot of peer groups by first two principal components and number of closest neighbours.



5.7. Geographic Representation of Final Peer Groups

When maps of the final peer groups are plotted, some interesting trends become apparent. Large regional centers appear to split out into peer groups based on their size. It then appears that health regions surrounding these regional centers tend to group together. There are some obvious relationships based on both latitude and longitude with health regions in the West, Prairies, Central and Eastern Canada clustering together. Further, there are definite groupings of northern health regions based on their location in western or eastern parts of the country. This is consistent with the Aboriginal make-up of these territories. All peer groups have representation across provincial and territorial boundaries.

Figure 5.7-1: Map of final peer groupings.



6. Summary

Expansion of existing data products to the health region level of geography and collection of health status indicators at the health region level through the Canadian Community Health Survey has led to the need for a method of comparing health regions with similar socio-economic determinants of health. Using data collected at the health region level, mostly through the 1996 Census of Canada, this paper outlines one method of grouping regions into "Peer Groups" or clusters with similar social and economic health determinants.

Health Regions were grouped using a non-hierarchical clustering algorithm which minimised the within cluster sum of squared errors for a predefined number of clusters. Starting with an initial set of 20 clusters and ensuring that each cluster contained at least two health regions, the results indicated that regions naturally grouped themselves into 14 distinct peer groups. The number of regions in each grouping ranged from 2 to 34 health regions. Visual analysis of the resulting clusters show that clusters were evenly spaced with approximately equal radius, indicating that attempts to split clusters with higher membership was not appropriate given the distribution of the data used in the analysis.

Stepwise descriminant analysis was used to determine which variables had the most influence on determining the final peer groupings. Overall, the three most important defining variables were the proportion of Aboriginals and visible minorities in a region closely followed by unemployment. Population size, age and income inequality also has a relatively strong influence on the groupings. A health region's rate of growth appeared to have a minor influence on the groupings.

From a practical perspective peer groupings with fewer than five health regions were combined with their nearest neighbours to provide a sufficient number of health regions for comparison purposes. The final result is a set of 10 peer groups ranging in size from 5 to 34 health regions with membership crossing provincial boundaries. When mapped, peer groups appear to fall out based on their geography in relationship to the east and west coasts as well as southern and northern Canada. Further they appear to be defined by their distance from large urban centers. Tests using a strict measure of statistical distance demonstrate that over 86% of health regions are in a peer group with at least two of their closest neighbours.

7. References

Andberg MR., Cluster Analysis for Applications. New York: Academic Press, 1973.

Everitt BS., Cluster Analysis, 3rd Edition. Toronto: Halsted Press, 1993.

SAS Institute Inc., SAS OnlineDoc®, Version 8, Cary, NC: SAS Institute Inc., 1999.

Appendix A . Variable Definitions

1. **1996 Population (Pop96)**

   Definition: Estimate of the total number of individuals living in a region.

   Source: Statistics Canada, Demography Division (special tabulation)

2. **Aboriginal Percentage (AboPer)**

   Definition: Proportion of a regions' total population self-identifying with an Aboriginal group.

   Sources: Statistics Canada, 1996 Census, 1996 Census Coverage Studies, and Demography Division (population estimates)

3. **Average Dwelling Value (AvgDwl)**

   Definition: Average expected value of an owner-occupied, non-farm, non-reserve dwelling (including the value of the land the dwelling is on) at the time of the Census.

   Source: Statistics Canada, 1996 Census

4. **Average Income (AvgInc)**

   Definition: Average personal income for persons aged 15 and over, from all sources.

   Source: Statistics Canada, 1996 Census

5. **Average Number of Years of Schooling (AvgSchl)**

   Definition: Average number of years of schooling (elementary, secondary, university and non-university) for the population aged 25 to 54.

   Source: Statistics Canada, 1996 Census (special tabulations)

6. **Employment Rate (25 to 54) (Emp)**

   Definition: Number of employed persons aged 25 to 54 divided by the total number of individuals between the ages of 25 and 54 in a given region.

   Source: Statistics Canada, 1996 Census

7. **Growth Rate (Growth)**

   Definition: Percent change in a regions population estimate from 1995 to 1997.

   Source: Statistics Canada, Demography Division (special tabulation)

8. **Government Transfer Income (GovTran)**

   Definition: Proportion of all income that came from government transfers (eg., GIS/OAS, C/QPP, EI, etc.) for the population 15 years of age and older.

   Source: Statistics Canada, 1996 Census

9. Housing Affordability (HouAff)

Definition: Proportion of total households spending 30% or more of total household income on shelter.

Source: Statistics Canada, 1996 Census

10. Immigrant Percentage (ImmPer)

Definition: The proportion of immigrants in a geographic area, and the proportion of those immigrants who came to Canada from 1981 to 1996.

Source: Statistics Canada, 1996 Census

11. Income Inequality (IncInq)

Definition: Proportion of income (from all sources) held by the bottom half of all households, based on the median household income for that specific community.

Source: Statistics Canada, 1996 Census (special tabulations)

12. Internal Migrant Mobility (MigMob)

Definition: Proportion of people that lived in a different Canadian municipality at the time of the previous Census (5-year internal migrants). Excludes Canadians in households outside Canada (military and government personnel).

Source: Statistics Canada, 1996 Census

13. Lone-Parent Families (LnePrnt)

Definition: Proportion of lone-parent families among all census families living in private households. A census family refers to a married or common-law couple or lone parent with at least one never-married son or daughter living in the same household.

Source: Statistics Canada, 1996 Census

14. Long Term Unemployment Rate (LtUnemp)

Definition: Proportion of the labour force aged 15 and over who did not have a job any time during the current or previous year.

Source: Statistics Canada, 1996 Census

15. Low Income 15+ (Low15)

Definition: Proportion of persons in economic families and unattached individuals with 1995 incomes below the Statistics Canada low-income cut-off (LICO). The cut-offs represent levels of income where people spend disproportionate amounts of money for food, shelter, and clothing. LICOs are based on family size and degree of urbanization; cut-offs are updated to account for changes in the consumer price index. Data were not derived for economic families or unattached individuals in the Territories or on Indian Reserves

Source: Statistics Canada, 1996 Census

16. Low Income Children (LowKids)

Definition: Proportion of children under age 18 living in economic families with 1995 incomes below Statistics Canada's low-income cut-offs (LICO). Data were not derived for economic families or unattached individuals in the Territories or on Indian Reserves.

Source: Statistics Canada, 1996 Census

17. Male-Female Ratio (MFRat)

Definition: Total number of males in a given region in 1996 divided by the total number of females.

Source: Statistics Canada, Demography Division (population estimates)

18. Owner-Occupied Dwellings (OwnDwl)

Definition: Proportion of dwellings in which the owner also lives. Band housing and collective dwellings (i.e. rooming houses, nursing homes, military camps etc .) are excluded from both numerator and denominator.

Source: Statistics Canada, 1996 Census

19. Population Density (PopDen)

Definition: Number of people per square kilometer.

Source: Statistics Canada, 1996 Census and Geography Division (special tabulations)

20. Population Under 15 (Pop15)

Definition: Proportion of the population in a given region under the age of 15 (1996 population).

Source: Statistics Canada, Demography Division (population estimates)

21. Population 65 Years and Older (Pop65)

Definition: Proportion of the population in a given region aged 65 years and older (1996 population).

Source: Statistics Canada, Demography Division (population estimates)

22. Strong MIZ (MIZ)

Definition: Strong MIZ (Census Metropolitan and Census Agglomeration **I**nfluenced **Z**ones represents the proportion of the population living in Census Metropolitan Areas (CMAs), Census Agglomerations (CAs) and communities that fall outside CMAs/CAs that have at least 30% of the employed labour force commuting to CMAs/CAs. The larger the proportion, the stronger the relationship between the specific community and a nearby CMA/CA.

Source: Statistics Canada, Geography Division , 1996 Census and Health Statistics Division (special tabulations)

23. Unemployment Rate (Unemp)

Definition: Total number of unemployed individuals 15 and older divided by the total number of individuals 15 and older participating in the labour force.

Source: Statistics Canada, 1996 Census

24. Visible Minority (VisMin)

Definition: Proportion of the population belonging to a visible minority group. As defined by the Employment Equity Act (1986), visible minorities are persons (other than Aboriginal people) who are non-Caucasian in race or non-white in colour.

Source: Statistics Canada, 1996 Census

Appendix B . Principal Component Analysis

	Eigenvalue	Difference	Proportion Explained	Cumulative Proportion
1	6.704	0.781	0.279	0.279
2	5.924	1.782	0.247	0.526
3	4.141	2.755	0.173	0.699
4	1.387	0.264	0.058	0.757
5	1.123	0.072	0.047	0.803
6	**1.051**	**0.396**	**0.044**	**0.847**
7	0.655	0.035	0.027	0.874
8	0.620	0.082	0.026	0.900
9	0.538	0.148	0.022	0.923
10	0.390	0.065	0.016	0.939
11	0.325	0.090	0.014	0.952
12	0.234	0.037	0.010	0.962
13	0.197	0.026	0.008	0.970
14	0.171	0.070	0.007	0.978
15	0.101	0.008	0.004	0.982
16	0.094	0.018	0.004	0.986
17	0.075	0.010	0.003	0.989
18	0.065	0.008	0.003	0.991
19	0.057	0.007	0.002	0.994
20	0.051	0.014	0.002	0.996
21	0.037	0.010	0.002	0.998
22	0.027	0.008	0.001	0.999
23	0.019	0.004	0.001	0.999
24	0.015		0.001	1.000

Eigenvectors for the first six principal components

	Prin1	Prin2	Prin3	Prin4	Prin5	Prin6
IncInq	-0.119	-0.236	0.030	0.298	0.240	-0.196
MFRat	-0.202	-0.150	0.306	0.044	0.141	0.177
Pop96	0.237	0.170	0.048	-0.017	0.162	-0.240
Pop15	-0.187	-0.123	0.355	-0.098	-0.037	-0.010
AvgDwl	0.309	0.012	0.105	0.109	0.041	0.229
OwnDwl	-0.034	-0.127	-0.359	0.185	0.149	0.338
MigMob	0.083	-0.236	0.137	-0.044	-0.232	0.514
GovTran	-0.247	0.219	-0.223	-0.020	0.059	0.100
VisMin	0.289	0.120	0.131	0.017	0.375	0.136
LnePrnt	0.015	0.253	0.247	-0.052	-0.467	-0.173
Emp	0.190	-0.301	-0.024	-0.321	0.050	-0.166
PopDen	0.227	0.206	0.089	-0.174	0.334	0.017
MIZ	0.297	0.042	-0.027	0.243	-0.270	-0.232
LTUnemp	-0.117	0.327	0.024	0.343	-0.077	0.063
Unemp	-0.215	0.263	-0.003	0.397	0.005	0.131
Low15	0.038	0.363	0.087	-0.227	-0.005	0.133
LowKids	0.019	0.361	0.070	-0.240	-0.022	0.175
AvgInc	0.259	-0.163	0.166	0.280	-0.097	-0.157
HouAff	0.307	0.139	-0.074	0.121	-0.241	0.204
Pop65	0.071	0.019	-0.413	-0.346	-0.032	0.037
AboPer	-0.191	0.012	0.373	-0.203	-0.071	-0.076
ImmPer	0.124	0.178	0.204	0.111	0.419	-0.100
Growth	0.157	-0.158	0.278	0.013	0.000	0.385
AvgSchl	0.353	-0.054	-0.086	0.073	-0.112	-0.027

Appendix C . Cluster Descriptive Statistics

Cluster		Percent Aboriginal	Percent Visible Minority	Unempl. Rate	1996 Population	Percent 65+	Income Inequality	Internal Migrant Mobility
A	*N*	5	5	5	5	5	5	5
	Mean	1.0	37.9	10.1	1,031,800	12.8	19.8	14.4
	Min	0.3	18.7	7.8	155,005	10.7	18.0	10.0
	Max	2.1	49.3	13.2	2,462,510	14.7	22.6	20.5
B	*N*	8	8	8	8	8	8	8
	Mean	1.5	19.8	7.6	610,864	9.8	22.9	19.0
	Min	0.3	15.0	5.6	176,772	7.1	19.9	14.1
	Max	3.6	31.2	8.8	881,794	13.6	25.4	28.9
C	*N*	5	5	5	5	5	5	5
	Mean	80.6	0.7	16.6	24,716	3.1	23.6	15.3
	Min	65.0	0.3	13.7	8,905	2.1	20.5	11.4
	Max	92.1	1.6	20.2	45,167	4.2	29.2	19.5
D	*N*	9	9	9	9	9	9	9
	Mean	2.8	0.4	28.3	85,546	11.8	22.5	9.2
	Min	0.3	0.2	21.7	17,899	8.7	21.1	5.5
	Max	9.6	1.1	42.8	142,625	14.2	24.8	11.9
E	*N*	13	13	13	13	13	13	13
	Mean	8.1	1.0	9.6	63,829	16.9	22.6	16.2
	Min	0.9	0.3	3.7	35,345	12.6	20.6	9.3
	Max	24.9	2.2	15.9	127,547	22.2	24.9	20.7
F	*N*	14	14	14	14	14	14	14
	Mean	23.5	2.9	12.1	47,468	5.1	23.7	22.8
	Min	1.2	0.4	7.1	1,111	1.8	21.9	15.7
	Max	45.5	7.4	20.3	129,112	8.4	26.3	29.7
G	*N*	21	21	21	21	21	21	21
	Mean	6.9	1.6	6.8	78,193	13.0	23.5	20.1
	Min	0.3	0.7	3.5	21,194	7.0	22.0	15.3
	Max	22.2	4.7	11.6	181,472	18.3	25.1	25.5
H	*N*	22	22	22	22	22	22	22
	Mean	2.7	2.0	12.7	312,847	12.0	22.1	14.6
	Min	0.1	0.3	8.1	54,706	7.4	21.3	8.1
	Max	10.2	11.7	17.4	1,287,115	14.7	23.7	25.5
I	*N*	34	34	34	34	34	34	34
	Mean	2.8	3.6	9.6	205,086	13.9	23.0	18.9
	Min	0.2	0.7	6.6	47,252	9.3	21.7	10.1
	Max	10.0	9.2	13.0	418,334	19.2	24.7	29.9
J	*N*	8	8	8	8	8	8	8
	Mean	3.3	4.5	7.5	217,391	10.5	24.0	25.4
	Min	0.4	1.4	4.6	71,177	8.9	22.3	20.3
	Max	6.8	9.9	11.3	472,754	12.8	25.7	31.5

Appendix D . Summary of Cluster Assignments

Final Peer Group A	Step 4.2	Step 4.3	Step 5.3
2406 - Région de Montréal-Centre	A	A	A
3595 - City of Toronto Public Health Unit	A	A	A
5916 - Vancouver	G	K	A
5917 - Burnaby	O	K	A
5919 - Richmond	O	K	A

Final Peer Group B	Step 4.2	Step 4.3	Step 5.3
3551 - Ottawa Carleton Public Health Unit	C	B	B
3553 - Peel Public Health Unit	C	B	B
3570 - York Public Health Unit	C	B	B
4804 - Calgary Regional Health Authority	C	B	B
4810 - Capital Health Authority	Q	B	B
5907 - South Fraser Valley	C	B	B
5908 - Simon Fraser	C	B	B
5918 - North Shore	L	B	B

Final Peer Group C	Step 4.2	Step 4.3	Step 5.3
2417 - Région du Nunavik	N	J	C
2418 - Région des Terres-Cries-de-la-Baie-James	N	J	C
4680 - Burntwood	D	C	C
4711 - Northern Health Services Branch (K) Service Area	B	C	C
6201 - Nunavut	D	C	C

Final Peer Group D	Step 4.2	Step 4.3	Step 5.3
1004 - Health and Community Services Western Region	M	I	D
1002 - Health and Community Services Eastern Region	E	D	D
1003 - Health and Community Services Central Region	E	D	D
1005 - Grenfell Regional Health Services Board	E	D	D
1205 - Zone 5	M	I	D
1305 - Region 5	M	I	D
1306 - Region 6	M	I	D
1307 - Region 7	K	I	D
2411 - Région de la Gaspésie-Îles-de-la-Madeleine	M	I	D

Final Peer Group E	Step 4.2	Step 4.3	Step 5.3
1102 - Rural Health Region	J	E	E
1201 - Zone 1	H	E	E
1202 - Zone 2	Q	E	E
3545 - Muskoka-Parry Sound Public Health Unit	H	E	E
3563 - Timiskaming Public Health Unit	H	E	E
4650 - Marquette	H	E	E
4655 - South Westman	J	E	E
4660 - Parkland	H	E	E
4702 - Moose Jaw (B) Service Area	Q	E	E
4705 - Yorkton (E) Service Area	H	E	E
4708 - Melfort (H) Service Area	H	E	E
4709 - Prince Albert (I) Service Area	H	E	E
4710 - North Battleford (J) Service Area	H	E	E

Final Peer Group F	Step 4.2	Step 4.3	Step 5.3
1006 - Health Labrador Corporation	F	F	F
2410 - Région du Nord-du-Québec	J	F	F
4670 - Norman	S	N	F
4690 - Churchill	T	N	F
4813 - Mistahia Regional Health Authority	J	F	F
4815 - Keeweetinok Lakes Regional Health Authority	S	N	F
4816 - Northern Lights Regional Health Authority	I	F	F
4817 - Northwestern Regional Health Authority	P	F	F
5912 - Cariboo	J	F	F
5913 - North West	I	F	F
5914 - Peace Liard	I	F	F
5915 - Northern Interior	I	F	F
6001 - Yukon Territory	I	F	F
6101 - Northwest Territories	S	N	F

Final Peer Group G	Step 4.2	Step 4.3	Step 5.3
3539 - Huron Public Health Unit	J	G	G
3549 - Northwestern Public Health Unit	J	G	G
3554 - Perth Public Health Unit	Q	G	G
3557 - Renfrew Public Health Unit	Q	G	G
4620 - North Eastman	J	G	G
4625 - South Eastman	J	G	G
4630 - Interlake	J	G	G
4640 - Central	J	G	G
4701 - Weyburn (A) Service Area	J	G	G
4703 - Swift Current (C) Service Area	J	G	G
4707 - Rosetown (G) Service Area	J	G	G
4801 - Chinook Regional Health Authority	Q	G	G
4802 - Palliser Health Authority	J	G	G
4805 - Health Authority #5	J	G	G
4806 - David Thompson Regional Health Authority	Q	G	G
4807 - East Central Health Authority	J	G	G
4809 - Crossroads Regional Health Authority	Q	G	G
4811 - Aspen Regional Health Authority	J	G	G
4812 - Lakeland Regional Health Authority	J	G	G
4814 - Peace Regional Health Authority	J	G	G
5901 - East Kootenay	Q	G	G

Final Peer Group H	Step 4.2	Step 4.3	Step 5.3
1001 - Health and Community Services St. John's Region	K	H	H
1203 - Zone 3	Q	H	H
1204 - Zone 4	K	H	H
1302 - Region 2	K	H	H
1304 - Region 4	K	H	H
2401 - Région du Bas-Saint-Laurent	K	H	H
2402 - Région du Saguenay - Lac-Saint-Jean	K	H	H
2403 - Région de Québec	K	H	H
2404 - Région de la Mauricie et Centre-du-Québec	K	H	H
2405 - Région de l'Estrie	K	H	H
2407 - Région de l'Outaouais	K	H	H
2408 - Région de l'Abitibi-Témiscamingue	K	H	H
2409 - Région de la Côte-Nord	K	H	H
2412 - Région de la Chaudière-Appalaches	K	H	H
2415 - Région des Laurentides	K	H	H
2416 - Région de la Montérégie	K	H	H
3526 - Algoma Public Health Unit	K	H	H
3537 - Hamilton-Wentworth Public Health Unit	Q	H	H
3547 - North Bay Public Health Unit	Q	H	H
3556 - Porcupine Public Health Unit	K	H	H
3561 - Sudbury Public Health Unit	K	H	H
4610 - Winnipeg	Q	H	H

Final Peer Group I	Step 4.2	Step 4.3	Step 5.3
1101 - Urban Health Region	Q	L	I
1206 - Zone 6	Q	L	I
1301 - Region 1	Q	L	I
1303 - Region 3	Q	L	I
2413 - Région de Laval	Q	L	I
2414 - Région de Lanaudière	K	L	I
3527 - Brant Public Health Unit	Q	L	I
3531 - Elgin-St Thomas Public Health Unit	Q	L	I
3533 - Bruce-Grey-Owen Sound Public Health Unit	Q	L	I
3534 - Haldimand-Norfolk Public Health Unit	Q	L	I
3535 - Haliburton-Kawartha-Pine Ridge Public Health Unit	Q	L	I
3538 - Hastings and Prince Edward Public Health Unit	Q	L	I
3540 - Kent-Chatham Public Health Unit	Q	L	I
3541 - Kingston-Frontenac-Lennox and Addington Public Health Unit	Q	L	I
3542 - Lambton Public Health Unit	Q	L	I
3543 - Leeds-Grenville-Lanark Public Health Unit	Q	L	I
3544 - Middlesex-London Public Health Unit	Q	L	I
3546 - Niagara Public Health Unit	Q	L	I
3552 - Oxford Public Health Unit	Q	L	I
3555 - Peterborough Public Health Unit	Q	L	I
3558 - Eastern Ontario Public Health Unit	Q	L	I
3562 - Thunder Bay Public Health Unit	Q	L	I
3565 - Waterloo Public Health Unit	Q	L	I
3568 - Windsor-Essex Public Health Unit	Q	L	I
4615 - Brandon	Q	L	I
4704 - Regina (D) Service Area	Q	L	I
4706 - Saskatoon (F) Service Area	Q	L	I
5902 - West Kootenay-Boundary	Q	L	I
5903 - North Okanagan	Q	L	I
5904 - South Okanagan Similkameen	Q	L	I
5905 - Thompson	Q	L	I
5906 - Fraser Valley	Q	L	I
5910 - Central Vancouver Island	Q	L	I
5920 - Capital	Q	L	I

Final Peer Group J	Step 4.2	Step 4.3	Step 5.3
3530 - Durham Public Health Unit	R	M	J
3536 - Halton Public Health Unit	R	M	J
3560 - Simcoe Public Health Unit	Q	M	J
3566 - Wellington-Dufferin-Guelph Public Health Unit	R	M	J
4803 - Headwaters Health Authority	J	M	J
4808 - WestView Regional Health Authority	J	M	J
5909 - Coast Garibaldi	R	M	J
5911 - Upper Island/Central Coast	Q	M	J