An Exploratory Analysis of the Development of Philippine Regions

Starr Clyde L. Sebial[a], [a]JH Cerilles State College Mati, San Miguel, Zamboanga del Sur, Philippines, Email: [a]clyde21g@gmail.com

The Philippines is one of the fastest growing economies in Southeast Asia and the Pacific. This study aims to explore and analyse the development of 17 Philippine regions according to eight key economic and social factors, including HEI PRC rate; crime rate; education; employment; health; poverty; income, and basic family amenities. Data is taken from the 2012 databases of the Philippine Statistics Authority (PSA), the Philippine Institute for Development Studies (PIDS) and Open Data Philippines. The Principal Component Analysis (PCA) method was used to generate the indexes of six factors and Cluster Analysis (CA) was used to group the similarities of the regions. The clustering process revealed that ARMM and MIMAROPA regions have clustered independently in two individual clusters while the rest of the 15 regions have clustered in two groups of 5 and 10. The study reveals that crime, poverty and health mortality rates have generally inverse relationships with HEI PRC percentage performance, employment, education, income and basic housing amenities.

Key words: *Data mining, social structures, economic, regions, cluster analysis.*

Introduction

Within Southeast Asia and the Pacific, the Philippines has a significant and rapidly-growing economy. Several measures of economic development, such as gross national product (GNP), social structures and labour force, have largely contributed to the country's solid economic growth. In order to sustain this growth, policy continuity is required in all local government units (LGUs), including municipality, provincial and regional levels. This policy stability will support developments in infrastructure and human capital, improvements to the investment climate, and governance reforms.

Various local and international studies have used a number of indicators and measurement methodologies to quantify the fiscal and social status of states, countries and local governments. One such factor of particular importance is good governance. As mentioned by Virola et.al. (2004), the National Statistical Coordination Board (NSCB) has defined governance as the manner in which power is exercised in the management of the country's economic and social resources for development. This refers to the carrying out of economic, political and administrative authority to administer the constituents' affairs at all levels. These includes income and expenditures; labour and employment; prices and finance; agriculture; health; education; communications; tourism; transportation and public order, and safety and justice.

The Human Development Report Office (HDRO) at the United Nations Development Programme (UNDP) has been using the Human Development Index (HDI) to assess the inequality, disparity and development progress across regions in the country. As defined by the UNDP, human development is a process of expanding and improving people's choices. For people to lead better lives, they must be able to enjoy a healthy and long-lasting existence through a number of key humanitarian factors. These include having access to knowledge in its different expressions and as measured by basic enrollment ratios and literacy rates; having the material resources for a decent standard of living as measured by income, and being able to freely participate in community life and collective affairs (2015 Human Development Report).

Since 1990, the UNDP has been publishing a series of Human Development Reports (HDRs) which provide analytical and empirically grounded discussions of major development issues, trends and policies. The HDI is said to be the summary of total development with an average achievement in three basic dimensions: long and healthy lifespan, knowledge, and decent standard of living. The 2012/2013 PHDR has revealed rankings of provinces in terms of the HDI, showing that the highest HDI levels were all in Luzon and 9 out of 10 provinces with the lowest HDI levels were in Mindanao.

Alipour (2014) outlines the four groups of development indicators as defined by the United Nations:

a) Social indicators (such as employment, health, education, housing, poverty, welfare and social justice, ethical values, the role of women, access to land and other natural resources, and social structure);
b) Economic indicators (such as independence or economic dependence, energy, production and consumption patterns, transportation, management, infrastructure, trade and productivity);

    c) Environmental indicators (such as groundwater, freshwater, agriculture and food, urbanisation, natural resources, air pollution and ozone depletion and global climate change); and

    d) Institutional factors (such as integrated decision making, capacity building, science and technology, public awareness, government and the role of civil society and legal framework).

Social issues like unemployment, poverty and crime have also been shown to threaten both highly urbanised and rural areas of developing countries like the Philippines. To allow for growth and improvement in social and economic areas, it is important to appropriately attend to these economic development indicators.

Leones (2006) devised a comprehensive study on the crimes associated with urbanisation in the Philippines. This study focused on several crime indexes of the 17 regions of the Philippines from January to November 2004. Among these regions, the NCR has registered the highest index crime volume. Leones further claimed that developing and maintaining safe communities, whether urban or rural, was central to the issue of good governance. She conjectured that a holistic approach focused on active citizenry empowerment, wherein society purposefully creates justice and human growth of high social importance, should be vigorously implemented to reduce the crime rate.

Krishnamurthy et. al., (2012) implemented data mining on crime survey analysis using clustering techniques involving crime reporting systems and crime classifications. The researchers processed the data into partition clustering methods which were primarily classified into K-means, AK-mode and Expectation-Maximisation algorithms. The study revealed that within the sensitive domain of crime data, efficient clustering techniques played vital roles for crime analysts and law enforcers to precede investigations and to assist with solving crimes quickly. The similarity measures in this study are important factors which help to identify patterns in unresolved crimes.

In examining Canada's 139 health regions, Shields et.al. (2002) used cluster and discriminant analyses in formulating peer groups with similar health characteristics and socio-demographic profiles. The study focused on life expectancy, disability-free life expectancy and risk factor estimates that were based on the data of 1996 Census of Canada, the Canadian Vital Statistics Database and Canadian Community Health Survey 2000. The study revealed that socio-demographic factors and risk factors, such as smoking and obesity, play critical roles in differentiating between communities in health outcomes.

The aforementioned studies support the evolution of exploratory data analysis (EDA) which has provided an outlook on what data can reveal beyond formal modeling and hypothesis

testing. The Cluster Data Analysis (CDA), an EDA statistical tool, was formulated to appropriately group similar individuals or objects. In this analysis, no assumptions are formulated about the underlying data set or its distribution. Unlike other EDA tools, this analysis also does not identify a particular statistical model. Results can rather assist in targeting appropriate programs and treatments to resolve certain social, economic or environmental factors that affect community development. The present study posits that competitive clusters of the 17 Philippine regions may provide opportunities for researchers and policy makers to meet the socio-economic challenges of globalisation.

This study aims to provide an analysis of clusters regarding identified economic and social factors of 17 regions in the Philippines. Data is taken from the 2012 academic year and is sourced from the databases of the Philippine Statistics Office (PSA) and the Philippine Institute for Development Studies (PIDS). The analysis focuses on the Higher Education Institution's (HEI) PRC board exam performance; employment rate; crime rate; education; health; income and expenditure, and basic family amenities.

All factors and variables are classified by region. The employment factor contains indexes in the labour force,and employment rates are taken from the year 2012. The education index contains the completion functional literacy rates as obtained from the 2008 Census. The health index contains infant mortality rates and the prevalence of underweight children, obtained from the 2011 Census. Income and expenditure include average family income and expenditure for the year 2012, as well as basic family amenities comprised of access to electricity, access to potable water, access to sanitary toilet facilities, the proportion of families living in makeshift housing and the proportion of families living in informal settlements.

Conceptual Framework

Data mining refers to the computer-assisted process of finding and analysing large sets of data and extracting meaning from that data. Analysis is then performed from various perspectives to summarise the data into useful and accessible information (Bharathi, 2014). In this research, data mining plays an important role in the prediction and analysis of performance indicators of the 17 Philippine regions.

Cortright (2006) suggests that different regions have distinct sets of economic development opportunities; clustering can assist in diagnosing a region's economic strengths and challenges and can identify realistic ways to shape the region's economic future. Grouping regions into these clusters according to identifiable factors could provide a wide range of analyses for researchers and policy makers, targeting appropriate programs and treatment to

resolve certain social, economic, environmental factors and indicators of community development.

Research Design and Methods

This research employs descriptive, exploratory and analytical methods which are grounded in the Principal Component Analysis (PCA) and Cluster Analysis (CA) techniques. The Philippines Statistics Authority (PSA) databases provide a 2012 statistic dataset of the average family income and expenditures as well as relevant employment factors. The percentages of crime, education, health, poverty, and housing basic amenities are sourced from the 2012 Economic and Social Database of the Philippine Institute for Development Studies (PIDS), while the percentage performance of Higher Education Institutions (HEIs) are taken from the Open Data Philippines database. All of these factors and variables were appropriately categorised into regions.

Prior to conducting the cluster analysis, indexes of each factor must be generated using the Principal Component Analysis (PCA). Given several variables for each factor, the index scores of each region were individually computed using the MINITAB and EXCEL software: *education factor index* from completion rate and functional literacy rate; *employment factor index* from employment rate and labor force rate; *health factor index* from infant mortality rate and prevalence of underweight children; *poverty factor index* from per capita poverty threshold and poverty circumstances among families; *income and expenditure factor index* from average family income and average family expenditure, and *housing and basic amenities factor index* from access to electricity, water, sanitary toilet facilities, proportion of families living in makeshift housing and proportion of families living in informal settlements.

Together with the *crime index* and *HEI performance index*, the generated indexes from the principal component analysis were used in the cluster analysis to group the Philippine regions into competitive clusters. This provides insights into how the performing regions are similar to each group member and how they differ from the members of the other clusters.

Results and Discussions

This section reports the interpretations and analysis of the study's results. The 17 variables excluding crime rate and HEI PRC percentage performance were treated with Principal Component Analysis (PCA). The resulting 6 indexes, together with crime and HEI rate, were analysed using Cluster Analysis (CA). Quantitative measures of the analyses included the MINITAB results of the Principal Component Analysis (PCA), bar graphs representative of each factor, and a Cluster Analysis of the Philippine regions. These quantitative components can be found in the attached appendices.

Chart 1 in Appendix 3 shows that the NCR, CAR, Western Visayas and Central Visayas regions received high percentage performance in the 2012 PRC board examinations while the ARMM received the lowest percentage performance. The statistics also indicate that the region SOCCSKSARGEN received the lowest crime rate while the CARAGA and ARMM regions reported the highest crime rates for 2012.

Principal Component Analysis (PCA)

The indexes of education, employment, poverty, health, income and expenditure and basic housing amenities were generated using the MINITAB and EXCEL software. The components of the PCA index equation were generated by MINITAB and the indexes of the regions on each factor were computed using the EXCEL software. Chart 2 in Appendix 4 shows that regions containing low employment indexes were NCR, Cagayan Valley, Central Luzon and ARMM. It also showed that Bicol, Zamboanga Peninsula, and Northern Mindanao regions produced the highest employment indexes in the year 2012. Chart 3 in Appendix 4 illustrates that the ARMM region has the lowest education index followed by Zamboanga Peninsula. All other 15 regions were above 80 in the education index.

Chart 4 in Appendix 4 clearly shows that the NCR and Central Luzon regions recorded the lowest rankings in the health index while MMAROPA and Eastern Visayas regions recorded the highest. Among the regions with the lowest poverty indexes were the ARMM, NCR, Central Luzon and Davao. This has been clearly illustrated in Chart 5 of Appendix 4. The chart also showed that the MMAROPA region got the highest poverty index in the year 2012. The NCR region ranked first both in the income index and basic housing amenities index while the ARMM ranked last also both in the income index and basic housing amenities index. Chart 6 of Appendix 4 also shows that CALABARZON and Central Luzon regions have a high-income index. Chart 7 of Appendix 4 shows that the 16 regions have enough access to basic housing amenities like access to potable water, electricity and sanitary toilet facilities while the ARMMs has an extremely low index compared to the other regions.

Cluster Analysis (CA)

The standard Cluster Analysis (CA) method was used to appropriately group the regions into clusters of similar objects, and which were distinct from other identifiable clusters. Each region was clustered based on eight factors: *HEI PRC rate, crime rate, employment index, education index, health index, income index and basic family amenities index*. Multivariate statistics tools, particularly the Cluster Analysis (CA), were used to analyse the resultant dataset using the MINITAB software.

Figure 1. Dendrogram of 17 Philippine regions
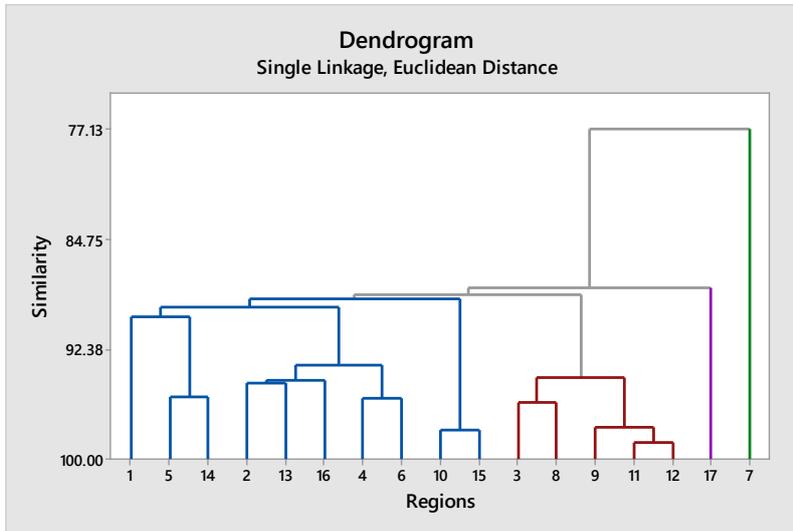
The dendrogram in Figure 1 above indicates that four competitive clusters are formulated in the standard Cluster Analysis regarding the 8 factors. Table 1 below represents this clustering of the Philippines regions. Cluster 1 is the biggest group containing 10 regions composed of NCR; CAR; Cagayan Valley; Central Luzon; CALABARZON; Central Visayas; Northern Mindanao; Davao; SOCCSKSARGEN, and CARAGA regions. Cluster 2 includes the 5 regions of Ilocos; Bicol; Western Visayas; Eastern Visayas, and Zamboanga Peninsula. Lastly, the MIMAROPA and ARMM regions have clustered independently as Clusters 3 and 4 respectively.

Table 1: Clustering of regions of the Philippines

Regions	Region Name	Cluster
1	NCR - National Capital Region	1
2	CAR - Cordillera Administrative Region	1
3	Region I - Ilocos	2
4	Region II - Cagayan Valley	1
5	Region III - Central Luzon	1
6	Region IVA - CALABARZON	1
7	Region IVB - MIMAROPA	3
8	Region V - Bicol	2
9	Region VI - Western Visayas	2
10	Region VII - Central Visayas	1
11	Region VIII - Eastern Visayas	2
12	Region IX - Zamboanga Peninsula	2
13	Region X - Northern Mindanao	1

14	Region XI - Davao	1
15	Region XII - SOCCSKSARGEN	1
16	Region XIII - CARAGA	1
17	ARMM - Autonomous Region in Muslim Mindanao	4

Based on the factor and percentage indexes computed in the Principal Component Analysis (PCA) for crime and HEI PRC, the majority of Cluster 1 regions have similarly low indexes in health mortality, poverty and crime. These same regions produced high indexes in the HEI PRC performance, employment, education, income and in basic housing amenities. This data is further detailed in the charts presented in the PCA results in Appendix 4. Cluster 1 generally exhibited an indirect relationship of the factors in this study. This simply suggests that communities with well-established governances that effectively monitor education, employment, income and basic household needs consequently have lower crime rates, mortality rates and poverty settings of the constituents.

The regions in Cluster 2 reveal similarly low rates in the HEI PRC performance, employment, income and basic housing indexes, though produced high crime and poverty indexes. These regions have therefore demonstrated the same indirect relationship as Cluster 1, though with a reversal of index factors. Cluster 2 regions' performances in board exams, employment status, income and housing amenities have been deteriorating due to its high rates of crime and poverty. The government officials and policy makers of Eastern and Western Visayas, Bicol, Ilocos and Zamboanga Peninsula regions must therefore provide well-planned deterrent programs to lessen their crime rates and lower their poverty indexes. Such implementations will improve income and basic housing amenities in the regions, as well as heighten performances in PRC board exams.

The MIMAROPA and ARMM regions were clustered independently with displays of different characteristics from the first two clusters. The lone region in the third cluster MIMAROPA shows relatively high indexes in the HEI PRC performance and employment with a very low crime rate, though also indicates a high poverty index and high mortality rate. This region also showed very low indexes in education, income and basic housing amenities. These results indicate that the region's income, housing amenities and education have an indirect relationship with its mortality rate and poverty index, meaning that higher mortality and poverty rates subsequently lower rates in education, income and housing. However, the crime rate of MIMAROPA was found to inversely affect HEI PRC performance and employment, a scenario displayed in Clusters 1 and 2.

The ARMM region has also grouped itself solely as Cluster 4 in the CA process due to its distinctly poor performance in all factors. ARMM ranked lowest in HEI PRC performance, employment, education, poverty, income and housing amenities. The region also ranked first

and second in the health mortality and crime rates respectively. Its constituents have performed the lowest in the entire country regarding education and PRC board exams. Families living in the region also have the least access to basic housing amenities like potable water, electricity and sanitary toilet facilities. These ratings thus explain the region's very high crime rate and poverty index. Intensive study, planning and supervision must be appropriately conducted by local officials of ARMM to resolve these pressing issues.

The analysis of this paper is not sufficient to comprehensively evaluate the performances, economic status and social factors of the 17 examined Philippine regions. Results may, however, provide government officials and policy makers with the grounds to create and implement programs for reform, particularly in the identified problematic regions. Further research on these issues may also be conducted to generate new perceptions and methods on conducting analysis for related concerns.

Conclusion

From this study, it is evident that crime, poverty and health mortality rates have generally inverse relationships with HEI PRC percentage performance, employment, education, income and basic housing amenities.

REFERENCES

Alipour, H. & Gholamipour, L. *Assessment of the Degree of Development of the Cities of Gilan Province, Iran (Economic Approach).* Trends in Life Sciences: An International Peer-reviewed Journal

Bharathi, A & Shilpa, R.(2014). *A Survey on Crime Data Analysis of Data Mining using Clustering Techiniques.* International Journal of Advance Research in Computer Science and Management Studies

Cortright, J. (2006). *Making Sense of Clusters: Regional Competitiveness and Economic Deelopment.* A Discussio Paper Prepared for the The Brookings Institution Metropolitan Policy Program

Krishnamurthy, R & Kumar, J. (2012). *Survey of Data Mining Techniques on Crime Data Analysis.* International Journal of Data Mining Techniques and Applications.

Leones, C. (2006). *The Current Situation of Crime Associated with Urbanization: Problems Experienced and Countermeasures Initiated in the Philippines.* 129th International Senior Seminar Visiting Experts' Papers

United Nations Development Programme (UNDP). *2015 Human Development Report.*

Virola, Al., de Costo, S., Addawe, J., & Reyes, L., (2004). *The Best and Worst   Provinces in the Philippines: What Happened to Their Leaders in the 2004 Elections?* 9th National Convention on Statistics (NCS)

Economic and Social Database. Philippine Institute for Development Studies (PIDS), http://econdb.pids.gov.ph/

Philippine Statistics Authority (PSA) Database., https://psa.gov.ph/

Open Data Philippines Database., http://data.gov.ph/

Appendix I. Identified Variables

| Variable |
| --- |
| HEI PRC Performance |
| Labour Force |
| Employment Rate |
| Unemployment Rate |
| Underemployment Rate |
| Crime |
| Completion Rate |
| Functional Literacy Rate |
| Infant Mortality Rate |
| Prevalence of Underweight Children |
| Per Capita Poverty Threshold |
| Poverty Incidence Among Families |
| Average Income |
| Average Expenditure |
| Access to Electricity |
| Access to Potable Water |
| Access to Sanitary Toilet Facility |
| Proportion of Families Living in Makeshift Housing |
| Proportion of Families Living in Informal Settlements |

Appendix 2. Map of the Philippines by Region

Appendix 3. HEI PRC Percentage Performance and Crime Rate

Chart 1. HEI and Crime Rates
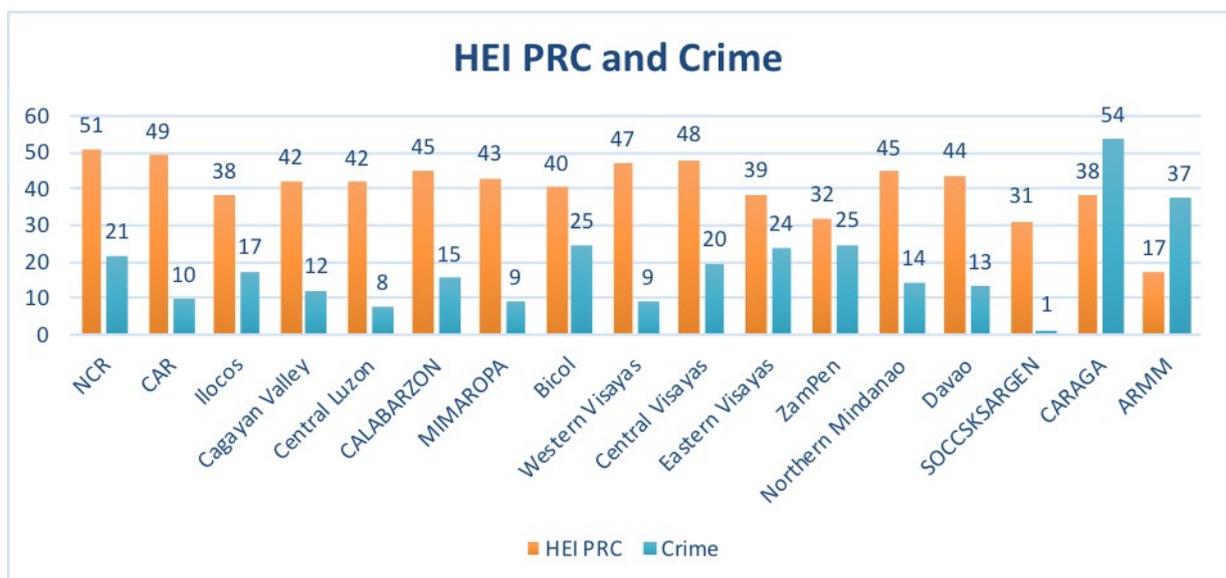
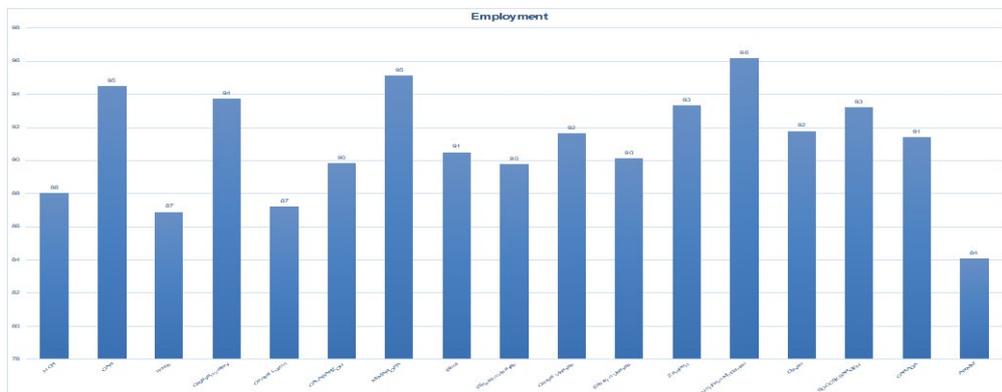**Appendix 4. Principal Component Analysis (PCA) Results**

Employment Index

Principal Component Analysis: C1, C2, C3, C4

Eigen analysis of the Covariance Matrix

Eigen value 40.915 9.912 6.774 -0.000
Proportion 0.710 0.172 0.118 -0.000
Cumulative 0.710 0.882 1.000 1.000

Variable	PC1	PC2
C1	0.242	0.564
C2	0.109	0.553
C3	-0.109	-0.553
C4	0.958	-0.268

Chart 2. Employment Index

Education Index

Principal Component Analysis: C5, C6

Eigen analysis of the Covariance Matrix

Eigen value 231.56 19.72
Proportion 0.922 0.078
Cumulative 0.922 1.000

Variable	PC1	PC2
C5	0.941	-0.339
C6	0.339	0.941

Chart 3. Education Index

Health Index

Principal Component Analysis: C7, C8
Eigen analysis of the Covariance Matrix

Eigen value 60.689 14.898
Proportion   0.803   0.197
Cumulative   0.803   1.000

Variable	PC1	PC2
C7	0.940	-0.341
C8	0.341	0.940

Chart 4. Health Index
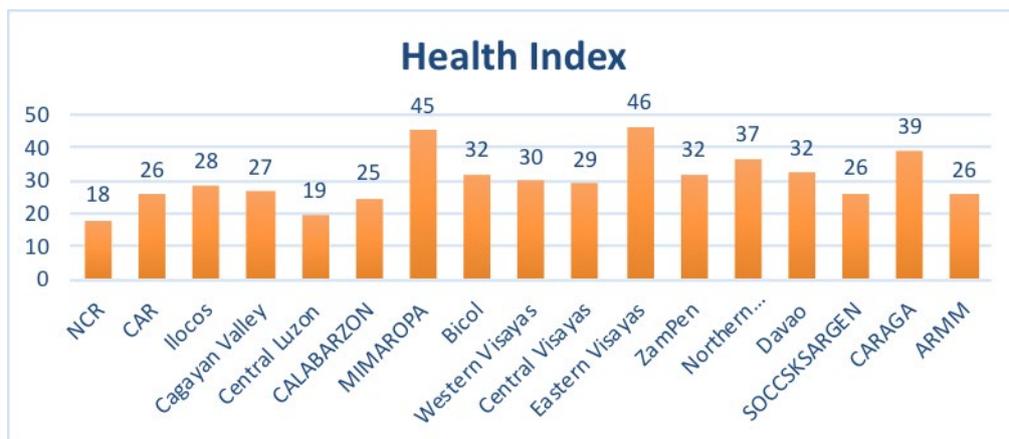
Poverty Index

Principal Component Analysis: C9, C10

Eigen analysis of the Covariance Matrix

Eigen value 852537   143
Proportion   1.000 0.000
Cumulative   1.000 1.000

Variable   PC1   PC2
C9   -1.000 -0.002
C10   0.002 -1.000
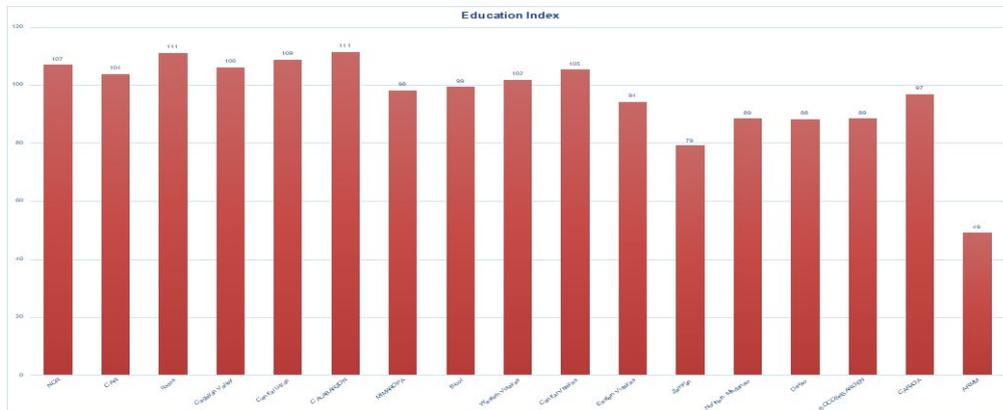
Chart 5. Poverty Index
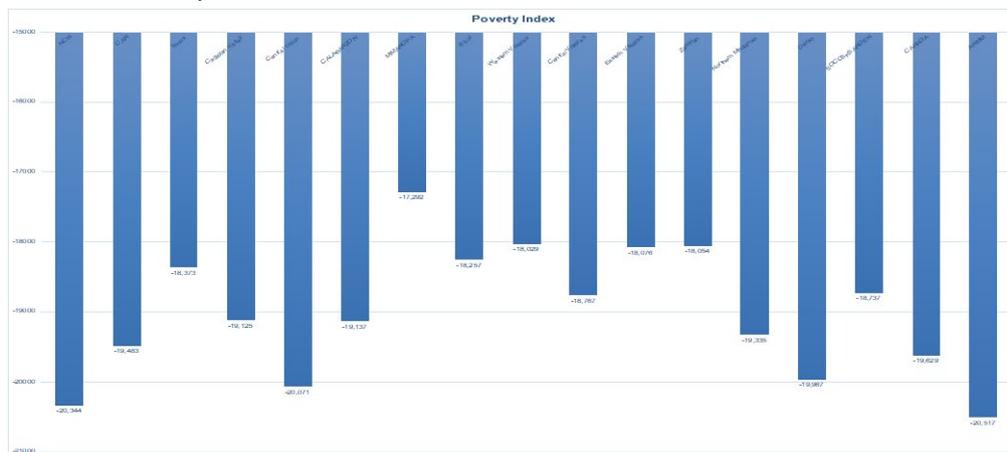
Income Index

Principal Component Analysis: C11, C12

Eigen analysis of the Covariance Matrix

Eigen value 6181.1   60.6
Proportion   0.990 0.010
Cumulative   0.990 1.000

Variable   PC1   PC2
C11   0.753 -0.658
C12   0.658   0.753

Chart 6. Income Index

Basic Housing Index

Principal Component Analysis: C13, C14, C15, C16, C17
Eigen analysis of the Covariance Matrix

Eigen value 320.40 11.09 10.97 3.94 0.34
Proportion 0.924 0.032 0.032 0.011 0.001
Cumulative 0.924 0.956 0.988 0.999 1.000

Variable	PC1	PC2	PC3
C13	0.524	-0.073	-0.848
C14	0.730	0.507	0.397
C15	0.438	-0.777	0.351
C16	0.006	0.076	0.014
C17	0.024	0.359	0.025

Chart 7. Basic Housing Index

Appendix 5. Cluster Analysis (CA) Result

Cluster Analysis of Observations: HEI PRC rate, crime rate, employment in, education in...

Euclidean Distance, Single Linkage
Amalgamation Steps

Step	Number of clusters	Similarity level	Distance level	Clusters joined		New cluster	Number of obs. in new cluster
1	16	98.8780	36.199	11	12	11	2
2	15	97.9513	66.098	10	15	10	2
3	14	97.7982	71.036	9	11	9	3
4	13	96.0614	127.073	3	8	3	2
5	12	95.7867	135.937	4	6	4	2
6	11	95.7095	138.426	5	14	5	2
7	10	94.7255	170.175	2	13	2	2
8	9	94.5004	177.436	2	16	2	3
9	8	94.3589	182.002	3	9	3	5
10	7	93.4410	211.618	2	4	2	5
11	6	90.0920	319.669	1	5	1	3
12	5	89.4219	341.289	1	2	1	8
13	4	88.8609	359.386	1	10	1	10
14	3	88.5798	368.458	1	3	1	15
15	2	88.1006	383.918	1	17	1	16
16	1	77.1270	737.967	1	7	1	17

Final Partition
Number of clusters: 4

	Number of observations	Within cluster sum of squares	Average distance from centroid	Maximum distance from centroid
Cluster1	10	2751091	448.397	907.958
Cluster2	5	94960	129.463	218.609
Cluster3	1	0	0.000	0.000
Cluster4	1	0	0.000	0.000

Cluster Centroids

	Cluster1	Cluster2	Cluster3	Cluster4	Grand centroid
Variable					
HEI PRC rate	43.5	39.2	43.1	16.9	40.6
crime rate	16.8	19.8	8.8	37.4	18.5
employment index	43.4	49.7	48.2	36.0	45.1
education index	100.5	97.2	98.2	49.2	96.4
health index	27.8	33.7	45.2	25.9	30.5
poverty index	-19459.4	-18157.7	-17291.9	-20516.9	-19011.3
income index	295.8	229.7	225.6	172.9	265.0
basic housing index	145.4	137.0	129.7	85.7	138.5

Distances Between Cluster Centroids

	Cluster1	Cluster2	Cluster3	Cluster4
Cluster1	0.00	1303.46	2168.79	1068.03
Cluster2	1303.46	0.00	865.99	2361.11
Cluster3	2168.79	865.99	0.00	3226.36
Cluster4	1068.03	2361.11	3226.36	0.00