International Journal of Innovation, Creativity and Change. www.ijicc.net
Volume 8, Issue 2, 2019

An Exploratory Analysis of the Development of Philippine Regions

Starr Clyde L. Sebial[a], [a]JHCerilles State College Mati, San Miguel, Zamboanga del Sur, 7029, Email: [a]clyde21g@gmail.com

The Philippines is one of the fast-growing economies in the South-East Asian and the Pacific region. This study considered eight factors: HEI PRC rate, crime rate, education, employment, health, poverty, income, and basic family amenities of the 17 regions of the country, all taken from the year 2012 databases of the Philippine Statistics Authority (PSA), the Philippine Institute for Development Studies (PIDS) and Open Data Philippines. Principal Component Analysis (PCA) generated the indices of the six factors and Cluster Analysis (CA) to group regional similarities. Clustering 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 five and 10. The study revealed 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

The Philippines has been labeled as one of the fast-growing economies in South-East Asia and the Pacific. Several measures of economic development such as gross national product (GNP), social structures, labor force and others have shown solid growth in the Philippine economy. However, sustaining it will require policy continuity among all local government units (LGUs), from municipality, provincial and regional levels, to support the development of infrastructure and human capital, improvements to the investment climate, and governance reforms.

Various researches, local and international, have been using a number of indicators and measurement methodologies to quantify the fiscal and social status of states, countries and local governments. One factor 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 include income and expenditures, labor and employment, prices, finance, agriculture, health, education, communication, tourism, transportation and public order, 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 inequality and disparity across regions in the country and the slow pace of development. As defined by the UNDP, human development is a process of enlarging people's choices. For people to lead better lives, they must be able to enjoy a healthy and long-lasting existence measured by access to knowledge in its different expressions, as measured by basic enrolment ratios and literacy rates, have the material resources for a decent standard of living measured by income, and 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) providing analytically and empirically grounded discussions of major development issues, trends and policies. The HDI is said to be the summary of total development, an averaged achievement in three basic dimensions namely; a long and healthy life, knowledge, and a decent standard of living. The 2012/2013 PHDR has revealed rankings of provinces in terms of HDI. The highest HDI levels were all in Luzon. Nine out of 10 provinces with the lowest HDI levels were in Mindanao.

Alipour (2014) mentioned that the United Nations has defined four new groups of development indicators:

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).

Problems like unemployment, poverty and crime threaten not only rural areas but even highly urbanised areas, throughout a developing country like the Philippines. To give room for growth and improvement of social and economic life, it is important to analyse the rates and attend to these economic development indicators.

Leones (2006) studied the crimes associated with urbanisation in the Philippines. A comprehensive study focused on several crime indices of the Philippines' 17 regions from January to November 2004. Among the regions, the NCR has registered the highest index crime volume. She further claimed that developing and maintaining safe communities, be they urban or rural, is central to good governance. She conjectured that a holistic approach with emphasis on empowering active citizenry, where society gives importance to its capacity to create justice and human growth, should be implemented vigorously to reduce crime rates.

Krishnamurthy, et al., (2012) used clustering techniques to mine crime surveys. He focused on crime reporting systems and classifications. He processed the data into partition clustering methods, which primarily classified into K-means, AK-mode and Expectation-Maximisation algorithms. The study revealed that in a sensitive domain of crime data, the efficient clustering techniques are vital for crime analysts and law-enforcers, before investigating a case, to help solve crimes faster. The similarity measures in his study are important in finding unresolved crimes.

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

The above-mentioned studies are evidence that the evolution of exploratory data analysis (EDA) provides an outlook on what the data can tell us, beyond formal modeling and hypothesis testing. The Cluster Data Analysis (CDA), an EDA tool in statistics, was formulated to cluster a set of similar individuals or objects. In this analysis, we do not formulate any assumptions about the underlying data set and its distribution. This analysis does not identify a particular statistical model as do the other EDA tools. But the results of this analysis could better target appropriate programs and treatment, to resolve certain social, economic, environmental factors and indicators of community development. The researcher believes that the competitive clusters of the 17 regions of the Philippines could provide

opportunities, from researchers to policy makers, to meet the socio-economic challenges of globalisation.

Overall, this study analysed clusters, with respect to identified economic and social factors, of the 17 regions in the Philippines for the academic year 2012. Those factors were taken from the Philippine Statistics Office (PSA) and Philippine Institute for Development Studies (PIDS). The analysis focused on the *Higher Education Institution's (HEI) PRC board exam performance, employment, crime rate, education, health, income and expenditure and basic family amenities*.

All factors and variables were classified by region. The employment factor contains indices of the labor force, and the employment rate in the year 2012. The education index comprised completion rate and functional literacy rates from the 2008 census. The health index contains infant mortality rates and the prevalence of underweight children, both from Census 2011. 'Income and expenditure' includes recent average family income and expenditure. The figure for basic family amenities indexes access to electricity, potable water, and sanitary toilet facilities, and the proportion of families living either in makeshift housing or informal settlements respectively.

Conceptual Framework

Data mining is the computer-assisted process of digging through and analysing enormous data sets, extracting its meaning, analysing data from different perspectives, and summarising it usefully (Bharathi, 2014). In this research, data mining is important for predicting and analysing performance indicators of the 17 regions of the Philippine archipelago.

Cortright (2006) cited different regions as having distinct sets of economic development opportunities, and stated that clustering them can help diagnose regional economic strengths and challenges, and identify realistic ways to shape the region's economic future. Grouping regions into clusters could facilitate a wide range of analysis, from people such as researchers to policy-makers, in targeting programs and treatments to resolve certain social, economic, environmental factors and indicators of community development.

Research Design and Methods

This research is descriptive, exploratory and analytical. It used Principal Component Analysis (PCA) and Cluster Analysis. The Philippines Statistics Authority (PSA) provided the researcher a 2012 statistical dataset of average family income and expenditures, and employment factors. The percentages of crime, education, health, poverty, and housing basic amenities were all taken from the 2012 Economic and Social Database, of the Philippine

Institute for Development Studies (PIDS). Percentage performance statistics of Higher Education Institutions (HEIs) were from Open Data Philippines. All of these factors and variables were categorised into regions.

Before the cluster is analysed, each factor must be indexed using PCA. Each factor has several variables. Accordingly, 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 incidence 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, access to potable water, access to sanitary toilet facility, proportion of families living in makeshift housing and proportion of families living in informal settlement.

Together with the *crime index* and *HEI performance index*, the PCA indices were used in the cluster analysis to group the Philippine regions into competitive clusters. That will provide insights on the internal similarities and external differences of each cluster's performing regions.

Results and Discussions

This section reports the interpretations and analysis of the results. The 17 variables excluding crime rate and HEI PRC percentage performance were treated with PCA. The resulting six indices, together with crime and HEI rates, were analysed using CA. Quantitative measures of the analysis: the MINITAB results of the Principal Component Analysis (PCA), together with the bar graphs of each factor, and the Cluster Analysis of the Philippine Regions, were attached in the appendices.

The Appendix 3, Chart 1 showed that the NCR, CAR, Western Visayas and Central Visayas regions performed in a high percentage in the PRC board examinations in the year 2012; the Autonomous Region in Muslim Mindanao (ARMM) in the lowest percentage. Also, the Central Mindanao or 'SOCCSKSARGEN' region has the lowest rate while the Caraga region and ARMM have the highest crime rate in the academic year 2012.

Principal Component Analysis (PCA)

The indices of education, employment, poverty, health, income and expenditure and basic housing amenities were generated using MINITAB and EXCEL. MINITAB generated the components of the PCA index equation, and EXCEL was used to compute the regional

indices on each factor. Appendix 4, Chart 2 showed that low employment indices were associated with NCR, Cagayan Valley, Central Luzon and ARMM. It also showed that in 2012, the Bicol, Zamboanga Peninsula, and Northern Mindanao regions rated highest on employment index. Appendix 4, Chart 3 illustrates that the ARMM region ranked the lowest on the education index, followed by Zamboanga Peninsula. All the other 15 regions scored above 80 on the education index.

In Appendix 4, Chart 4, the NCR and Central Luzon regions clearly ranked lowest in the health index, while MMAROPA and the Eastern Visayas regions ranked highest. Among the regions with the lowest poverty scores were the ARMM, NCR, Central Luzon and Davao. This is clear in Appendix 4, Chart 5. The chart also showed that the MMAROPA region rated highest on the poverty index in 2012. The NCR region ranked first both in the income index and basic housing amenities index, while the ARMM ranked last in both income and basic housing amenities. Appendix 4, Chart 6, also showed that CALABARZON and Central Luzon regions have high incomes. Appendix 4, Chart 7, showed that the 16 regions have enough access to basic housing amenities like potable water, electricity and sanitary toilet facilities, while the ARMMs has an extremely low index compared to the other regions.

Cluster Analysis (CA)

The researcher used standard cluster analysis to group similar regions, and so distinguish them from different regions. The clustering factors were: *HEI PRC rate, crime rate, employment index, education index, health index, income index and basic family amenities index*. Multivariate statistics tools, Cluster Analysis (CA) in particular, was used to analyse the dataset with the MINITAB software.

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

The Dendrogram in the Appendix 5 showed four competitive clusters, formulated in the standard cluster analysis regarding the eight factors. Table 1 shows clustering in the Philippines Regions. Cluster 1, the biggest group containing 10 regions, is composed of NCR, CAR, Cagayan Valley, Central Luzon, CALABARZON, Central Visayas, Northern Mindanao, Davao, Davao, SOCCSKSARGEN and Caraga regions. Cluster 2 has five elements, namely the Ilocos, Bicol, Western Visayas, Eastern Visayas and Zamboanga Peninsula regions. The MIMAROPA and ARMM regions clustered independently as Clusters 3 and 4 respectively.

The Principal Component Analysis computed the factors and percentages of crime and HEI PRC. Based on the indices computed thus, most Cluster 1 regions have similarly low indices of health mortality, poverty and in crime, and high indices in HEI PRC performance, employment, education, income and in the basic housing amenities. This can be shown from 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 when a community has established good governance monitoring education, employment, income and basic household needs, eventually this will lower crime rates, mortality rates as well as the poverty settings of the constituents.

The Cluster 2 regions revealed similarly low rates in the HEI PRC performance, employment, income and basic housing indices. However, this cluster displayed a high index in crime and poverty. The Cluster 2 regions have demonstrated the same indirect relationship of the factors only in a different manner. Their performances in board exams, employment status, income and housing amenities has been deteriorating due to high rates of crime and poverty. Government officials and policy makers of Eastern and Western Visayas, Bicol, Ilocos and Zamboanga Peninsula regions must carefully plan deterrent programs, to lessen their crime rates and poverty index scores. This will improve income and basic housing amenities, as well as their performances in the PRC board exams.

MIMAROPA and ARMM regions have clustered independently and displayed different characteristics from the first two clusters. The lone region in the third cluster, MIMAROPA, has relatively high index scores of HEI PRC performance and employment, and a very low

crime rate. However, the region has a very high poverty score and high mortality rate. It also shows very low indices in education, income and basic housing amenities. It clearly shows that income, housing amenities and education indirectly relate to mortality and poverty. As the mortality and poverty rates increase, education, income and housing will probably decrease. However, the crime rate of MIMAROPA inversely affects the HEI PRC performance and employment score, already displayed by Clusters 1 and 2.

The ARMM region has grouped itself solely as Cluster 4 in the cluster analysis. The region performed poorly in all of the factors. ARMM ranked lowest in HEI PRC performance, employment, education, poverty, income and housing amenities. The region also ranked first in health mortality and second in crime. Its constituents performed the lowest in the entire country, regarding education and PRC board exams. Their families also have the least access to basic housing amenities like potable water, electricity and sanitary toilet facilities. These ratings basically explained why ARMM has a very high crime rate and poverty index. Local ARMM officials must conduct intensive study, planning and supervision, to improve the region.

The analysis of this paper is insufficient to comprehensively evaluate the performance, economic status and social factors of the 17 regions of the country. Conversely, the research may serve a purpose; to provide government officials and policy-makers the objective of creating programs, and give researchers ideas as to opening new perceptions about analysing related concerns.

Conclusion

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/

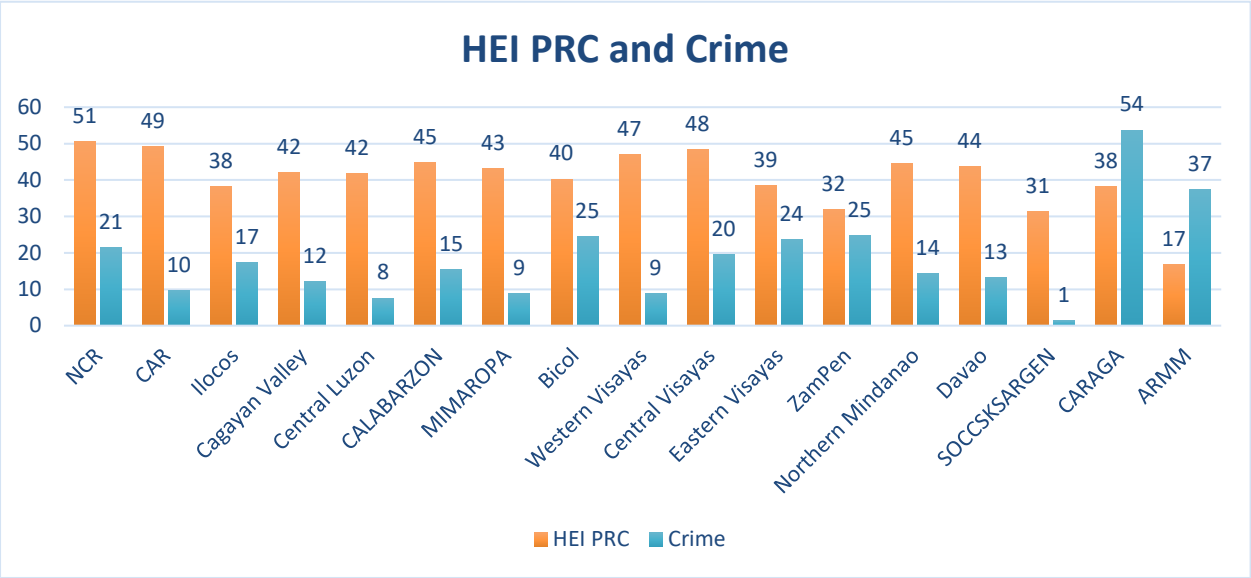
Appendix I. Identified Variables

Variable
HEI PRC Performance
Labor 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

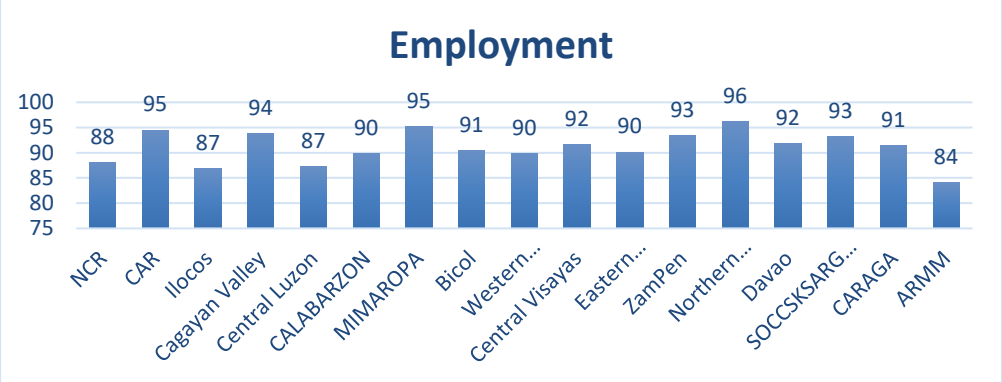
Principal Component Analysis: C1, C2, C3, C4

Eigenanalysis of the Covariance Matrix

Eigenvalue	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

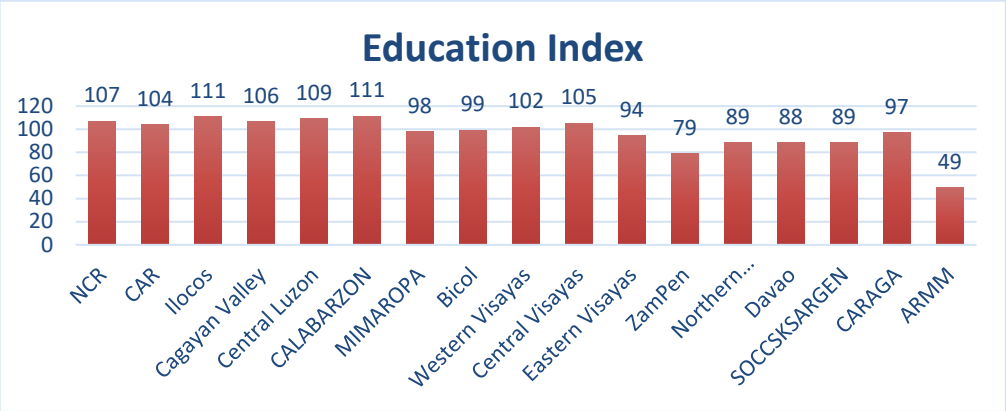
Principal Component Analysis: C5, C6

Eigenanalysis of the Covariance Matrix

Eigenvalue	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

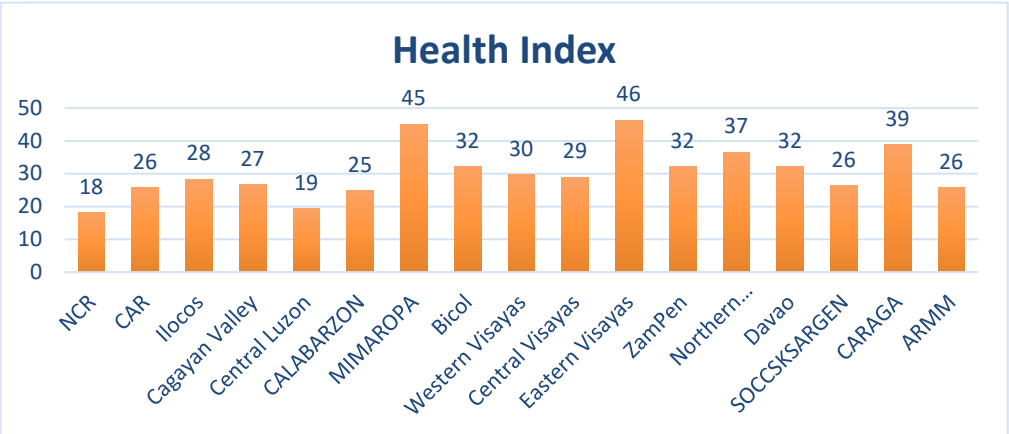
Principal Component Analysis: C7, C8

Eigenanalysis of the Covariance Matrix

Eigenvalue	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

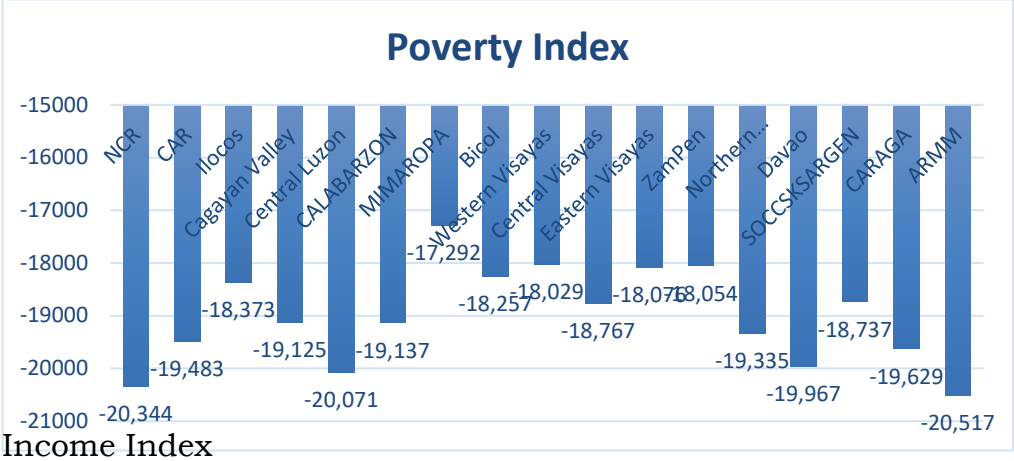
Principal Component Analysis: C9, C10

```
Eigenanalysis of the Covariance Matrix

Eigenvalue  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

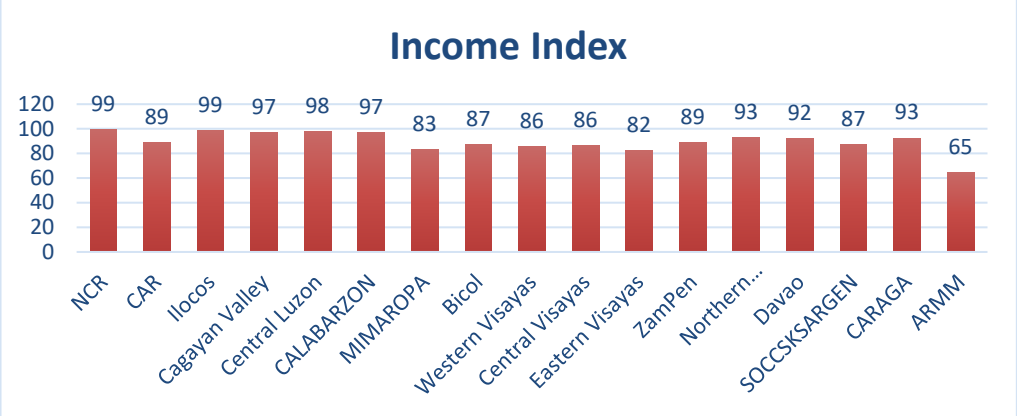
Principal Component Analysis: C11, C12

```
Eigenanalysis of the Covariance Matrix

Eigenvalue  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

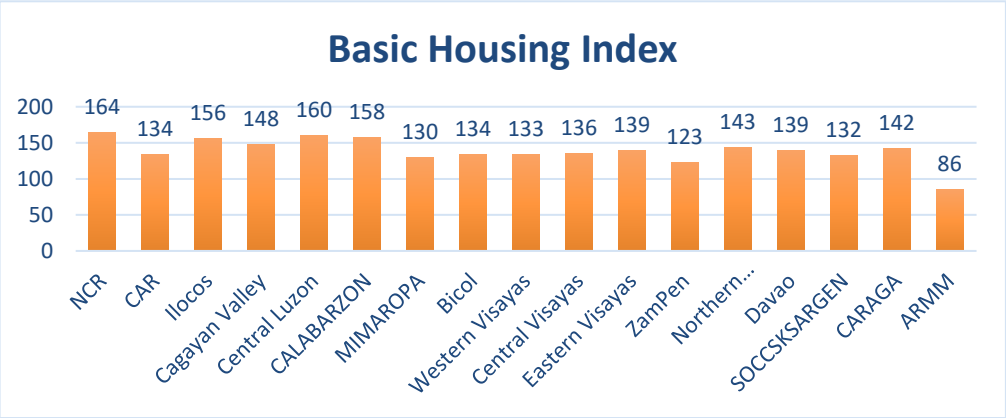
Principal Component Analysis: C13, C14, C15, C16, C17

Eigenanalysis of the Covariance Matrix

Eigenvalue	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

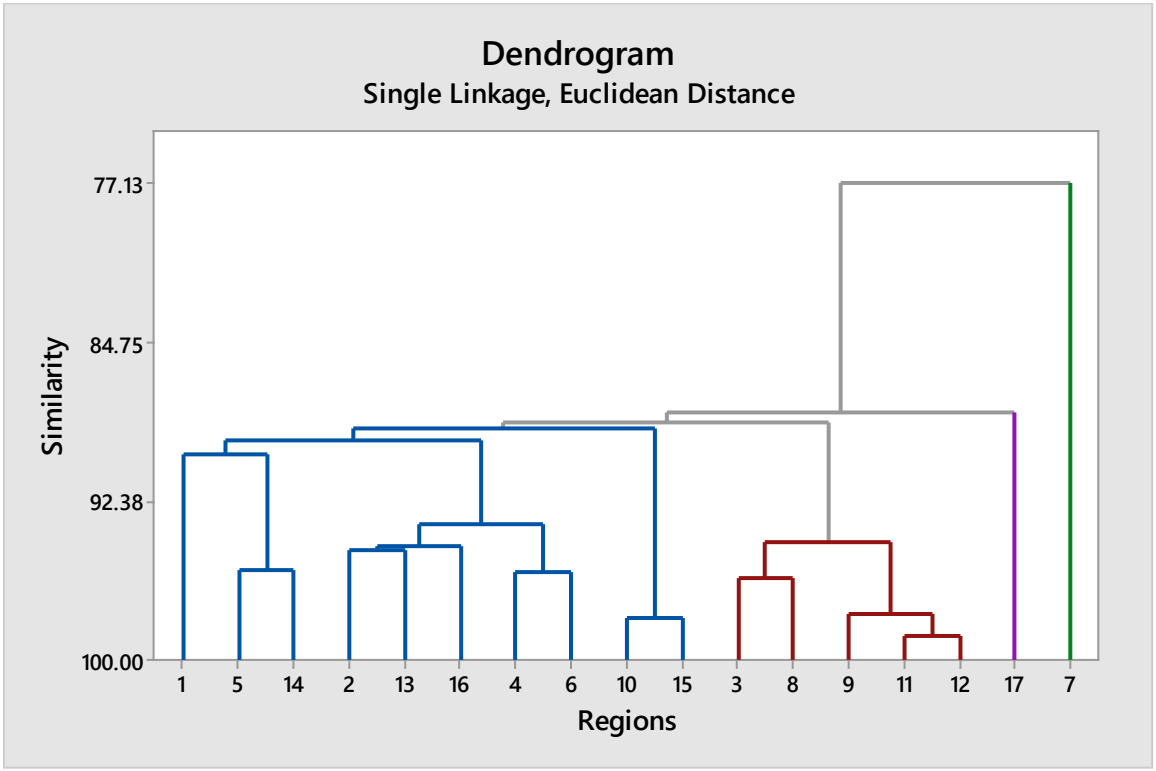
	Number of observations	Within cluster sum of squares	Average distance from centroid	Maximum distance from centroid
Cluster1	10	2751091	448.397	907.958
Cluster2	5	94960	129.463	218.609
Cluster3	1	0	0.000	0.000
Cluster4	1	0	0.000	0.000

Cluster Centroids

Variable	Cluster1	Cluster2	Cluster3	Cluster4	Grand centroid
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


Dendrogram
Single Linkage, Euclidean Distance
Similarity
77.13
84.75
92.38
100.00
1 5 14 2 13 16 4 6 10 15 3 8 9 11 12 17 7
Regions