An analysis on the Location and Type of Index Crimes in the Philippines

Rena Sandy H. Baculinao[1] and **Roel F. Ceballos**[2]

Department of Mathematics and Statistics
University of Southeastern Philippines
Bo. Obrero, Davao City, Philippines 8000
[1]renasandybaculinao12@gmail.com
[2]roel.ceballos@usep.edu.ph

ABSTRACT

Crime statistics is a report that is used to monitor the occurrences, rates, and trend of crimes in the Philippines. It has been noted in recent reports that crime rates are decreasing and crime solution efficiency have improved. However,despite the improvement in the crime solution efficiency, crime prevention needs to be strengthened because the occurrence of index crimes is still relatively high. This paper aims to determine the hotspots or locations where a specific type of index crime is most likely to occur using the correspondence analysis and signed Chi-square statistic. Results show that there is a significant relationship between the location and type of index crimes in the Philippines. It is recommended that appropriate programs and policies shall be created to prevent or minimize the occurrence of specific types of index crimes in locations(regions) where they are found to be highly associated. CALABARZON, Caraga Region, and ARMM should focus on policies to address the high incidence of murder, while Cagayan Valley may explore plans and programs to addressed homicide occurrence. Bicol Region's top priority problem is rape, while for NCR is robbery. Lastly, theft is a common problem in Central Visayas, Northern Mindanao, Davao Region, and CAR. These regions may implement joint programs on theft prevention and control.

Keywords: Crime Statistics, Correspondence Analysis

1 INTRODUCTION

Crime is as an act committed or omitted in the violation of a law forbidding or commanding it and for which a punishment is imposed upon conviction. In the Philippines, the Revised Penal Code (Republic Act. No. 3815) serves as the fundamental law that defines criminal offenses and provides the penalties for the commission of such.Crime is divided into index and non-index crimes for statistical purposes and to create a standardized definition of crime classification. Index crimes refer to crimes against person and crimes against property such as murder, homicide, physical injury and rape, carnapping/carjacking, and cattle rustling. On the other hand, non-index crimes are violations of special laws such as illegal logging or local ordinances (Senate Economic Planning Office, 2013)

Based on the reports of the Philippine National Police(PNP), crime rates in the Philippines has a steady drop. Also, there was a significant decline in index crimes such as murder, homicide, physical injury, rape, robbery, theft, and carnapping in all regions of the country. The reduction index crime was recorded to be 20.56 percent reduction in 2017 compared to the same period in 2016 from January to October (Felipe, 2017).

In terms of Crime solution efficiency, which refers to the percentage of solved cases by the law enforcement agencies out of the total number of crime incidents, there was a steady increase from the year 2009 to 2012. Despite the improvement, crime solution efficiency remains low compared to other ASEAN countries.Managbanag (2017), found out that there is an association between the place of crime and the type of crimes committed by inmates of Davao Prison and Penal Farm. This paper aims to determine the crime hot spots or the location where a specific type of index crimeis most likely to

occur.Results of this study would provide necessary information that is useful in creating plans, policy, and programs for the prevention, management, and control of the different types of index crime for each region. Furthermore, the results and findings of the study could be used by future researchers as a basis for further research on crimes per location and related topics.

2 METHODS

Correspondence analysis is used to determine whether there is an association between the type and location of index crimes in the Philippines. The Signed -Chi-Square statistic was also used to aid in the interpretation of the bi-plot.

2.1 Procedure for Correspondence Analysis

The following are the step by step computational method in performing the analysis, according to Hintze (2007):

1. Established an *n*×*m* data matrix$\boldsymbol{K}$, where *n* is the number of rows and *m* is the number of columns.
2. Compute$\boldsymbol{P}$, the proportion matrix, by dividing the elements of $\boldsymbol{K}$ by the total of all numbers in $\boldsymbol{K}$**.** Mathematically, proportion matrix($\boldsymbol{P}$) is equal to
$$\boldsymbol{P} = \{p_{ij}\} = \left\{\frac{k_{ij}}{k}\right\}.$$
3. Compute the vectors$\boldsymbol{r}$ and $\boldsymbol{c}$. Using standard matrix notation, we have,
$$\boldsymbol{r} = \boldsymbol{P}\mathbf{1}$$
$$\boldsymbol{c} = \boldsymbol{P}'\mathbf{1}$$
where $\mathbf{1}$ is an appropriately dimensioned vector of ones.
4. Transform the vectors $\boldsymbol{r}$ and $\boldsymbol{c}$into diagonal matrices, then, get the square roots, and take the inverse of the resulting square matrices.
$$\boldsymbol{D}_{\boldsymbol{r}} = \left[\sqrt{diag(\boldsymbol{r})}\right]^{-1}$$
$$\boldsymbol{D}_{\boldsymbol{c}} = \left[\sqrt{diag(\boldsymbol{c})}\right]^{-1}$$
5. Compute the scaled matrix, $\boldsymbol{A}$ by taking the product of the diagonal matrices and matrix $\boldsymbol{P}$.
$$\mathbf{A} = \mathbf{D}_{\mathrm{r}}\mathbf{P}\mathbf{D}_{\mathrm{c}}$$
6. Calculate the Singular Value Decomposition (SVD) of **A**.
$$\langle \boldsymbol{B}, \boldsymbol{W}, \boldsymbol{C} \rangle = \boldsymbol{SVD}(\boldsymbol{A})$$
7. Compute the coordinate matrices,$\boldsymbol{F}$, and $\boldsymbol{G}$.
$$\boldsymbol{F} = \boldsymbol{D}_r\boldsymbol{B}\boldsymbol{W}$$
$$\boldsymbol{G} = \boldsymbol{D}_c\boldsymbol{C}\boldsymbol{W}'$$
8. Compute the eigenvalues$\boldsymbol{V}$.
$$\boldsymbol{V} = \boldsymbol{W}\boldsymbol{W}'$$
9. Compute the row distances (d_i) column distances (d_j).
$$d_i = \sum_j \left(\frac{1}{p_{.j}}\right)\left(\frac{p_{ij}}{p_{i.}} - p_{.j}\right)^2$$
$$d_j = \sum_i \left(\frac{1}{p_{i.}}\right)\left(\frac{p_{ij}}{p_{.j}} - p_{i.}\right)^2$$
10. The weighs ($w_i$andw_j)come from the vectors $\boldsymbol{r}$and $\boldsymbol{c}$ were formed in step 3.
$$w_i = \{r_i\} \text{ and } w_j = \{c_j\} \text{where } i = 1, \dots, n \text{ and} j = 1, \dots, m.$$

2.2 Statistical Treatment

1. **Proportion**was used to measure relative frequency or determines the percentage of the total individual in a specific attribute. The formula to get the proportion p is given by

$$p = \frac{n_{ij}}{n}$$

where

n_{ij} is the number of individual in the sample with the specific attribute and *n* is the total number of individuals

2. **Marginal Frequencies** are the row and column total of each cell frequencies. It was used to determine the expected frequency values of row (location of crimes) and column (types of index crimes) profiles and masses.

To get the marginal frequency of the ith row according to Rencher (2002), we used

$$n_{i+} = \sum_{j=1}^{b} n_{ij}$$

Likewise, to get the marginal frequency of the jth column, we used

$$n_{+j} = \sum_{i=1}^{a} n_{ij}$$

where

n_{ij} is the frequency for each combination of two or more variables in the contingency table with i rows ($i = 1,2, \dots, a$) and *j* columns ($j = 1,2, \dots, b$).

3. **Row and Column Profile** was used to determine the quotient between the frequency of each element in the contingency table and the marginal frequency of that row or column and is given respectively by (Rencher, 2002)

$$r_{ij} = \frac{n_{ij}}{n_{i+}}$$

$$c_{ij} = \frac{n_{ij}}{n_{j+}}$$

where

$i = 1, 2, \dots, a$ and $j = 1, 2, \dots, b$.

4. **Chi-Square Statistic (X^2)** was used to measure the association between variables and to examine the relationship between categories of the same variables. Moreover, correspondence analysis used this to measure the distance between the points on the biplot. The formula for chi-square statistic for testing independence between the two variables is given by (Rencher, 2002)

$$X^2 = \sum_{i=1}^{a} \sum_{j=1}^{b} \frac{\left(n_{ij} - \frac{n_{i+}n_{+j}}{n}\right)^2}{\frac{n_{i+}n_{+j}}{n}}$$

5. **Signed Chi-Square Value** refers to the individual chi-square value of every element in the contingency table. It was used to determine the similarity between the location of crimes and the type of index crimes, where positive values represent greater similarity while negative values represent lower similarity (Hair, 2010). To compute, we have:

$$Signed\ Chi - Square\ Value = \frac{\left(n_{ij} - \frac{n_{i+}n_{+j}}{n}\right)^2}{\frac{n_{i+}n_{+j}}{n}}$$

The computation for the Signed Chi-Square Value is similar to the computation for Chi-Square Statistic (X^2). The difference is that the Chi-Square Statistic computes for the overall chi-square value while Signed Chi-Square Value computes for the individual chi-square value. Also, for the Chi-Square Statistic, every chi-square value is always positive. While for Signed Chi-Square Value, chi-square value is positive when the observed value is higher than the expected value and negative when the expected value is higher than the observed value.

6. **Total Inertia** is defined as the product of the row total and the square of its distance to the centroid. It was used to assess the quality of its graphical representation in correspondence analysis (Yelland, 2010). It is given by

$$Total\ Inertia = \frac{X^2}{n}$$

where X^2 is the computed chi-square value, and *n* is the total table count.

7. **Biplot** is an enhanced scatterplot that uses both points and vectors to represent structure. It uses points to represent the scores of the observations, and it uses vectors to represent the coefficients of the variables on the principal components (Young, 1999). Rows with comparable patterns of counts will have points that are close together on the biplot and columns with comparable patterns of counts will also have points that are close together on the biplot (SAS Institute Inc., 2010).

   Also, in terms of angular distances, the closer the arrows to the axes, the greater contribution of the row category on that axis relative to the other axes. If the arrow is halfway between two axes, its row category contributes to both axes. The size of the angle between pairs of crimes measures the degree of relationship or association of crimes and location, the smaller the angle, the stronger the level of relationship or association (Dibal and Usman, 2018).

3 Results and Discussion

The distribution of the type of index crimes to the seventeen regions of the Philippines shows that, of the seventeen different regions in the Philippines, National Capital Region (NCR) has the highest total of incidence of index crimes with 98,020 crimes or 18.81%. It is followed by Western Visayas (Region 6) with a total of 67,517 index crimes or 12.95%. On the other hand, the Autonomous Region in Muslim Mindanao (ARMM) has the lowest number of index crimes with a total of 3,878 or 0.74%. However, crimes in ARMM are mostly understated because most people do not report crime occurrences to proper authorities. ARMM has the highest percentage for murder with 12.33%, while NCR has the lowest percentage with 0.73%. Also, Cagayan Valley has the highest percentage for homicide with 5.30%, and NCR has the lowest with 0.49%. Physical injury is rampant in Ilocos Region with a percentage of 63.23% while NCR has only 33.82%. For rape, the Cagayan Valley Region got the highest percentage with 3.65% while Northern Mindanao got the lowest with 1.25%. Robbery is also common in NCR with 20.34% while Ilocos Region has only 5.57%. Theft, carnapping, and cattle rustling are common occurrences in the Cordillera Administrative Region (CAR), ARMM, and SOCCSKSARGEN with 39.92%, 8.20%, and 1.23% respectively. Moreover, theft, carnapping, and cattle rustling arerare in Cagayan Valley, Eastern Visayas, and NCR with 16.30%, 0.57%, and 0% respectively. (See Appendix A. Table 1)

The result of the Correspondence Analysis is presented in Table 2. The chi-square statistic value is equal to 43,159.745 with a sufficiently very small p-value. Therefore, there is a significant association between the location and the type of index crime as analyzed in seven dimensions with a level of significance equal to 0.05. There are seven dimensions generated that can be interpreted to explain the model. Since it does not include all the dimensions, inertia does not always add up to 100%. The Inertia column gives the total variance explained by each dimension in the model.Based on our results, knowing something

about the location of crime explains around 8.3% of something about the occurrence of the type of index crime and vice versa. The association is weak, but it is highly significant, as indicated by the chi-square statistic. Each dimension has an indicated amount of variance explained in the model. As in most cases, Dimension 1 will always explain the most variance, which is 5.2% in this case.

Table 2. Summary of the Main Statistical Results

Dimension	Singular Value	Inertia	Proportion of Inertia	
			Accounted for	Cumulative
1	0.229	0.052	0.631	0.631
2	0.134	0.018	0.216	0.847
3	0.078	0.006	0.073	0.919
4	0.053	0.003	0.034	0.953
5	0.048	0.002	0.028	0.981
6	0.029	0.001	0.010	0.991
7	0.027	0.001	0.009	1
Total		0.083	1	1

Chi Square = 43159.745; pvalue<0.01

The Singular Value column consists of the square roots of the eigenvalues, which is the measure of the amount of variance of the original data. Also, the eigenvalues correspond to the total amount of information retained by each axis. In Correspondence Analysis, inertia and eigenvalues are synonymous in a way that, “each axis has an eigenvalue whose sum equals the inertia” (Benzecri, 1992).

The Proportion of Inertia columns consist of the percentage of variance for every dimension that explains the total variance explained by the model. Since there are eight columns, the average axis should account for 100/(8-1) = 14.3%.Now, in terms of rows, the average axis should account for 100/(17-1) = 6.3%. The maximum of the two percentages is 14.3%. Thus, any axis or dimension that contributes more than 14.3% is regarded as significant and included in the solution (Bendixen, 2003). Approximately, Dimension 1 explains 63.1% of the 8.3% total variance explained in the model. Dimension 2 explains 21.6% of the 8.3% total variance explained in the model. While Dimensions 3, 4, 5, 6, and 7 explain too little of the total variance with 7.3%, 3.4%, 2.8%, 1.0%, and 0.09% respectively and is lesser than 14.3%. Hence, there are only two dimensions that could explain the most variance (84.7%) of the model between the location and the type of index crime.

The biplot (See Appendix. Figure 1) is the graphical representation of the result in the correspondence analysis for the location and the type of index crime in the Philippines.The distance between any row points or column points gives a measure of their similarity or dissimilarity. Points that are closer to each other have similar profiles, and points that are far from each other have dissimilar or different profiles. Table 3 shows the calculated signed chi-square value as a similarity measure aside from the bi-plot. Signed chi-square value is an easier and more objective way to interpret the similarity and dissimilarity of each location and types of index crimes. ARMM, CALABARZON, and Caraga Region are associated with murder. Cagayan Valley isassociated with homicide, whileIlocos Region, Central Luzon, MIMAROPA, Western Visayas, Eastern Visayas, Zamboanga Peninsula are associated with physical injury. Bicol Region has the most occurrence of rape, while in NCR, the most common occurrence of crime is robbery. In Central Visayas, Northern Mindanao, Davao Region, and CAR theft cases. Furthermore, cattle rustling

and SOCCSKSARGEN are close in terms of angular distances; therefore, cattle rustling is a common occurrence in SOCCSKSARGEN.

Table 3. Calculated Signed Chi-Square Value

Region	Murder	Homicide	Physical Injury	Rape	Robbery	Theft	Carnapping	Cattle Rustling
1	-15.96	79.62	1127.54	13.16	-699.52	-763.97	-18.41	267.38
2	53.05	1030.61	282.53	189.87	-107.15	-766.81	42.80	0.02
3	-46.51	70.18	456.49	26.33	-190.95	-467.63	143.44	-64.39
4A	1250.11	455.25	-0.76	434.58	-15.79	-366.86	451.17	-51.59
4B	13.31	12.30	250.87	101.58	-101.68	-268.64	-9.51	-0.04
5	-1.80	-0.94	85.36	123.96	-256.23	-2.20	-107.41	14.76
6	-328.88	-44.51	1472.92	-79.19	-1892.57	-28.28	-815.62	-8.75
7	-33.26	-17.89	-18.50	-177.52	63.02	67.09	-41.22	-25.74
8	129.22	8.22	256.14	-13.46	-484.90	-16.81	-253.14	-19.02
9	148.65	-1.15	464.64	-4.90	-346.38	-340.72	3.29	27.04
10	23.87	39.45	-1429.57	-49.35	67.68	1268.02	80.81	445.22
11	-6.64	-88.36	-6.47	-0.28	-10.72	163.48	-309.95	-6.78
12	76.02	-2.17	5.29	-7.59	18.50	-112.89	44.09	496.67
13	333.77	41.64	-1.33	56.08	-0.91	-19.64	-6.17	-5.68
ARMM	2423.45	41.33	-35.94	43.85	0.15	-238.12	518.53	311.19
CAR	-74.56	-6.97	-63.28	4.50	-14.28	205.88	0.00	-13.28
NCR	-611.12	-567.42	-4262.22	-80.57	7212.37	1525.52	377.50	-368.99

Legend: Highly Associated Less Associated

Table 3 can also be used to determine dissimilarity between the location and type of index crimes. Looking at the most negative value in each row, we determine the location where each type of index crime is less likely to occur.There is a less occurrence of murder in CAR (-74.56), while Region 10 (-1429.57) and NCR (-4262.22) are less associated to physical injury. Also, Region 7 (-177.52) has less similarity to rape, while Region 5 (-256.23), Region 6 (-1892.57), Region 8 (-484.90), and Region 9 (-346.38) are less associated to robbery. Finally, Region 1 (-763.97), Region 2 (-766.81), Region 3 (-467.63), Region 4A (-366.86), Region 4B (-268.64), Region 12 (-112.89), Region 13 (-19.64), and ARMM (-238.12) have less similarities to the occurrence of theft.

.

4 Conclusion and Recommendation

Based on the results, there is a significant association between the location and type of index crimes in the Philippines. In general, this implies that there are specific crimes which are most likely to occur in certain regions. Thus, it is recommended that concerned government offices and agencies shall create programs and policies for crime prevention and management on a regional level. Specifically, ARMM, CALABARZON, and Caraga Region should focus on policies that will help prevent cases of Murder and improve crime solution efficiency. Ilocos region, Central Luzon, MIMAROPA, Western Visayas, Zamboanga Peninsula should address problems on physical injury. Central Visayas, Northern Mindanao, Davao Region, and CAR should create further research on prevention and control of theft cases.

5 References

Bendixen, M. (2003). *A Practical Guide to the Use of Correspondence Analysis in Marketing Research.* Marketing Bulletin, 2003, 14, Technical Note 2.

Benzecri, J. (1992). *Correspondence Analysis Handbook.* New York: Marcel Decker.

Dibal, N. P., & Usman, I. A. (2018). Correspondence Analysis as a Strategy to Explore the Association between Different Categories of Crime in Yobe State, Nigeria. *American Journal of Theoretical and Applied Statistics*, 118-125.

Felipe, C. S. (2017, November 19). *PNP: Index crimes dropped nationwide in 2017*. Retrieved from Philstar: https://www.philstar.com/metro/2017/11/19/1760546/pnp-index-crimes-drop-nationwide-2017

Hintze, J. L. (2007). Chapter 430. Correspondence Analysis. In *NCSS User's Guide I* (pp. 430-445). Kaysville, Utah.

Managbanag, D. A. (2017). *Correspondence Analysis between Place and the Type of Crimes by Inmates of Eastern Mindanao.*

Philippine National Police. (2017, June 22). *2013 National Crime Statistics*. Retrieved from Open Data Philippines: https://data.gov.ph/about

Rencher, A. C. (2002). *Methods of Multivariate Analysis 2nd Ed.* John Wiley and Sons, Inc. Publication.

SAS Institute Inc. (2010). *Knowledge Base/Focus Areas: Statistics*. Retrieved from http://support.sas.com/rnd/app/da/market/stat.html

Senate Economic Planning Office. (2013, June). *Crime Statistics At A Glance*. Senate of the Philippines. Retrieved from https://www.senate.gov.ph/publications/AAG%202013-05%20-%20Crime%20Statistics.pdf

Yelland, P. M. (2010). An Introduction to Correspondence Analysis. *Mathematica Journal 12*, 1-8.

6 APPENDIX

Table 1. Distribution of Index Crimes per Location or Regions in the Philippines

Legend: Lowest / Highest

Location of Crime	Type of Index Crime								TOTAL (%)
	Murder (%)	Homicide (%)	Physical Injury (%)	Rape (%)	Robbery (%)	Theft (%)	Carnapping (%)	Cattle Rustling (%)	
1	352(1.4%)	504(2.1%)	15468(63.2%)	506(2.1%)	1363(5.6%)	5523(22.6%)	497(2.0%)	249(1.0%)	24462(4.7%)
2	260(2.8%)	495(5.3%)	5638(60.4%)	341(3.7%)	714(7.7%)	1521(16.3%)	329(3.5%)	36(0.4%)	9334(1.8%)
3	578(1.3%)	803(1.9%)	23898(55.5%)	900(2.1%)	3882(9.0%)	11515(26.7%)	1451(3.4%)	60(0.1%)	43087(8.3%)
4A	1305(4.6%)	821(2.9%)	13688(48.0%)	970(3.4%)	2984(10.5%)	7479(26.2%)	1266 (4.4%)	33(0.1%)	28546(5.5%)
4B	262(2.2%)	208(1.8%)	6873(58.5%)	352(3.0%)	955(8.1%)	2826(24.0%)	237(2.0%)	43(0.4%)	11756(2.3%)
5	469(1.7%)	370(1.3%)	14619(52.2%)	741(2.6%)	2253(8.0%)	9020(32.2%)	418(1.5%)	145(0.5%)	28035(5.4%)
6	573(0.9%)	733(1.1%)	39549(58.6%)	882(1.3%)	3800(5.6%)	21275(31.5%)	49 (0.7%)	207(0.3%)	67517(13.0%)
7	746(1.4%)	605(1.2%)	24320(47.0%)	509(1.0%)	6423(12.4%)	17977(34.7%)	1045(2.0%)	124(.2%)	51749(9.9%)
8	508(2.9%)	285(1.6%)	9835(56.8%)	241(1.4%)	976(5.6%)	5354(30.9%)	98 (0.6%)	30(0.2%)	17327(3.3%)
9	661(2.9%)	303(1.3%)	13498(58.1%)	364(1.6%)	1659(7.2%)	5979(25.8%)	615(2.7%)	136(0.6%)	23215(4.5%)
10	710(2.1%)	596(1.8%)	11258(33.9%)	415(1.3%)	4237(12.8%)	14566(43.9%)	1075(3.2%)	361(1.1%)	33218(6.4%)
11	539(1.6%)	266(0.8%)	16022(47.4%)	583(1.7%)	3602(10.6%)	12403(36.7%)	325(1.0%)	98(0.3%)	33838(6.5%)
12	644(2.5%)	329(1.3%)	12675(49.3%)	394(1.5%)	3121(12.1%)	7427(28.9%)	800 (3.1%)	316(1.2%)	25706(4.9%)
13	381(4.4%)	191(2.2%)	4110(47.5%)	245(2.8%)	944(10.9%)	2595(30.0%)	177(2.0%)	19(0.2%)	8662(1.7%)
ARMM	478(12.3%)	101(2.6%)	1614(41.6%)	123(3.2%)	444(11.5 %)	718(18.5%)	318(8.2%)	82(2.1%)	3878(0.7%)
CAR	98(0.8%)	143(1.1%)	5577(43.4%)	258(2.0%)	1300(10.1%)	5126(39.9%)	316(2.5%)	23(0.2%)	12841(2.5%)
NCR	712(0.7%)	482(0.5%)	33147(33.8%)	1353(1.4%)	19934(20.3%)	39024(39.8%)	3368(3.4%)	0(0.0%)	98020(18.8%)
TOTAL	9276(1.8%)	7235(1.4%)	251789(48.3%)	9177(1.8%)	58591(11.2%)	170328(32.7%)	12833(2.5%)	1962(0.4%)	521191(100%)

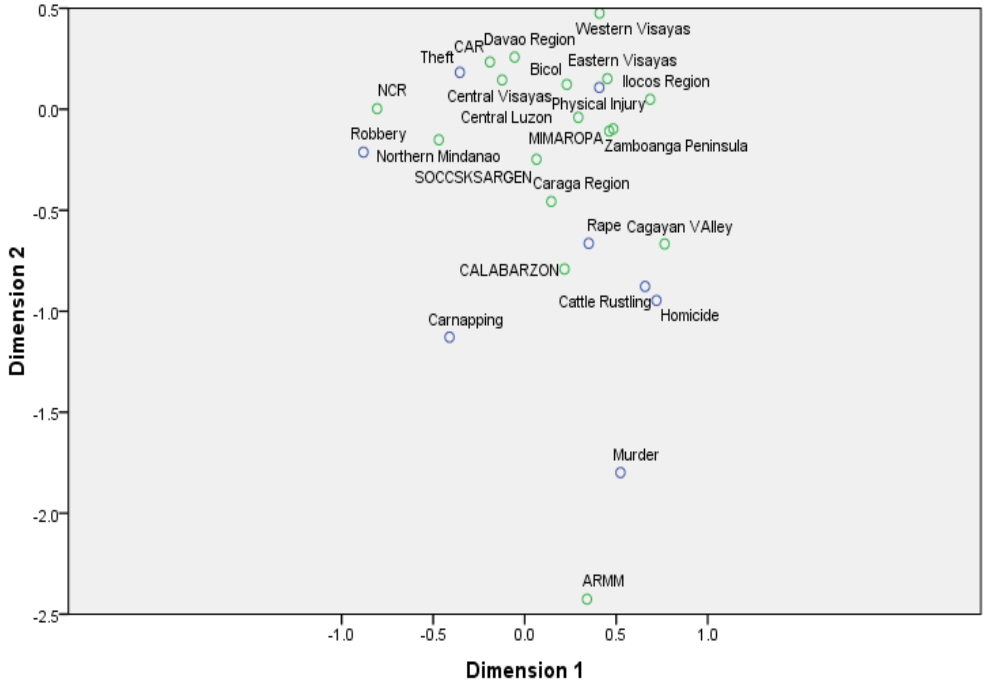

Figure 1. Association between the Location and the Type of Index Crime in Two Dimension