A Study of Area Wise Mapping of Assaults on Women in Delhi

Rajani Kant Awasthi[1], Dr. Atul Sangal[2]

[1]Research Scholar at Sharda University, Greater Noida U.P., India

[2]Associate Professor, Sharda University, Greater Noida U.P., India

Abstract: ***The national capital recorded a 17 per cent rise in crimes against women in the first six-and-half months of this year as compared to the corresponding period last year, with an average of six rapes being reported daily, according to Delhi Police data. Cases of assault on women increased by almost 19 per cent, while instances of cruelty by husbands and in-laws rose by 29 per cent during the period. According to the data, Delhi recorded 6,747 cases of crimes against women between January 1 and July 15 in 2021 and the number went up to 7,887 in 2022.Till July 15 this year, the city reported 1,100 rape cases as compared to 1,033 in the same period last year. Cases of assault on women with intent to outrage her modesty have also gone up and a total of 1,480 such cases have been reported. The number was 1,244 last year. Cases of insult to the modesty of women have marginally gone down from 229 in 2021 to 225 in 2022, the data showed. Cases of kidnapping of women increased by around 17 per cent, while cases of abduction of women went down by almost 43 per cent. A total of 2,197 cases of kidnapping of women have been reported this year, 317 more than last year, it stated. The number of cases of abduction of women was 105 in 2022 and 184 in 2021.A total of 2,704 cases of cruelty by husbands and in-laws were reported in the national capital in 2022 and the figure was 2,096 in 2021, it said, adding that 69 cases of dowry death and seven under the Dowry Prohibition Act have also been reported. A senior police officer said awareness among women has increased. Police are regularly sensitising women in colleges, schools and colonies. In most rape cases, the accused is known to the victim. Earlier, women were not much aware about laws, but now they are coming forward to report the incidents happening against them, the officer said.***

Keywords: Principal component analysis, Eigenvalue, Score plot, Scree plot, Bi-plot, type of crimes

1. Introduction

Of all the crimes, sex related crimes are the most barbarous and humiliating, women and children are the most vulnerable group for such type of crimes. The alarming rise in cases of sexual assaults in Delhi and worldwide represents a major public health problem.

Table 1: Statistics of various crimes in Delhi during last ten years

Year	Dacoity	Murder	Att.to murder	Robbery	Riot	Kid for Ransom	rape	Snatching	Hurt	Burglary	M.V.Theft	House Theft	Other Theft	M.O.Women	Other Kid/Abd.	Fatal Accident	Simple Accident
2011	33	543	386	562	50	25	572	1476	1946	1419	14668	1918	6313	657	3767	2047	5233
2012	28	521	439	608	79	21	706	1440	1747	1715	14391	1746	5895	727	3949	1822	5115
2013	33	517	585	1245	113	30	1636	3638	1768	2835	14916	3216	11992	3515	6294	1778	5788
2014	82	586	770	6464	160	38	2166	7350	2077	10309	23384	12735	42634	4322	7105	1629	6994
2015	75	570	770	7407	130	36	2199	9896	1898	12848	32729	15318	56385	5367	7694	1582	6503
2016	46	528	646	4761	79	23	2155	9571	1489	14307	38644	14721	77563	4165	6596	1548	5827
2017	36	487	645	3147	50	16	2146	8231	1352	9819	40972	10739	114054	3422	6079	1565	5105
2018	25	513	529	2444	23	19	2135	6932	1508	4117	46433	3727	138596	3314	6032	1657	4858
2019	15	521	487	1956	23	15	2168	6266	1312	3026	46215	2630	190874	2921	5886	1433	4177
2020	9	472	570	1963	689	11	1699	7965	1064	2199	35019	2036	132419	2186	4051	1163	3015
2021	26	459	761	2333	68	17	2076	9383	1360	2637	37910	2485	150203	2551	5510	1206	3514

The primary objective of this study to earmark the geographical areas of Delhi with respect to type of crimes so that a foolproof framework can be prepared and required measures placed depending upon the type of crime, area of crime and methods/tactics adopted by criminals in executing the crime successfully.

In this study, a particular methodology using Principal Component Analysis along with various algorithms has been followed.

Variance Decomposition

Consider a linear statistical system -- a random matrix (multidimensional set of random variables) X of size n×m where the first dimension denotes observations and the second variables. Moreover, if we recall that linear statistical systems are characterized by two inefficiencies: 1) noise and 2) redundancy. The former is commonly measured through the *signal(desirable information) to noisepublic health problem. (undesirable information) ratio* SNR=σ^2signal/ σ^2noise, and implies that systems with larger signal variances σ^2signal relative to their noise counterpart, are more informative. Assuming that noise is a nuisance equally present in observing each of the m variables of our system, it stands to reason that variables with larger variances have larger SNRs, therefore carry relatively richer signals, and are in this regard relatively more important, or *principal*.

Whereas relative importance reduces to relative variances across system variables, redundancy, or relative *uniqueness* of information, is captured by system covariances. Again we recall that covariances (or normalized covariances called correlations) are measures of variable dependency or co-movement (direction and magnitude of joint variability). In other words, variables with overlapping (redundant) information will typically move in the same direction with similar magnitudes, and will therefore have non-zero covariances. Conversely, when variables share little to no overlapping information, they exhibit small to zero linear dependency, although statistical dependence could still manifest nonlinearly.

Together, system variances and covariances quantify the amount of information afforded by each variable, and how much of that information is truly unique. In fact, the two are typically derived together using the familiar *variance-covariance* matrix formula:

$$\Sigma X = E(X \top X) = (\top)$$

where ΣX is an $m\times m$ *square symmetric* matrix with (off-) diagonal elements as (co)variances, and where we have *a priori* assumed that all variables in X have been demeaned. Thus, systems where all variables are unique will result in a diagonal ΣX, whereas those exhibiting redundancy will have non-zero off-diagonal elements. In this regard, systems with zero redundancy have a particularly convenient feature known as *variance decomposition*. Since covariance terms in these systems are zero, total system variation (and therefore information) is the sum of all variance terms, and the proportion of total system information contributed by a variable is the ratio of its variance to total system variation.

Although the variance-covariance matrix is typically not diagonal, suppose there exists a way to diagonalize ΣX, and by extension transform X, while simultaneously preserving information. If such transformation exists, one is guaranteed a new set of at most m variables (some variables may be perfectly correlated with others) which are uncorrelated, and therefore linearly independent. Accordingly, discarding any one of those new variables would have no linear statistical impact on the $m-1$ remaining variables, and would reduce dimensionality at the cost of losing information to the extent contained in the discarded variables. In this regard, if one could also quantify the amount of information captured by each of the new variables, order the latter in descending order of information quantity, one could discard variables from the back until sufficient dimensionality reduction is achieved, while maintaining the maximum amount of information within the preserved variables. We summarize these objectives below:

1) Diagonalize ΣX.
2) Preserve information.
3) Identify principal (important) information.
4) Reduce dimensionality.

So how does one realize these objectives? It is precisely this question which motivates the subject of this entry.

Principal Component Analysis

Recall that associated with every matrix X is a *basis* -- a set (matrix) of *linearly independent* vectors such that *every* row vector in X is a linear combination of the vectors in the basis. In other words, the row vectors are *projections* onto the column vectors in B. Since the covariance matrix contains all noise and redundancy information associated with a matrix, the idea driving *principal component analysis* is to re-express the original covariance matrix using a basis that results in a new, diagonal covariance matrix -- in other words, off-diagonal elements in the original covariance matrix are driven to zero and redundancy is eliminated.

Change of Basis

The starting point of PCA is the *change of basis* relationship. In particular, if B is an $m\times p$ matrix of geometric transformations with $p\leq m$, the $n\times p$ matrix $Q=XB$ is a projection of the $n\times m$ matrix $X=[X^{\top}1,\ldots, X^{\top}n]$ onto B. In other words, the rows of X are linear combinations of the column vectors in $B=[B_1,\ldots,B_p]$, Formally,

$$Q = \begin{vmatrix} X1 \\ X2 \end{vmatrix} \quad [B1 \ldots \ldots \ldots B2]$$

$$= \begin{vmatrix} X1B1 \ldots X2Bp \\ XnB1 \ldots \ldots XnBp \end{vmatrix}$$

More importantly, if the column vectors {B1,…, Bp} are also linearly independent, then B, by definition, characterizes a matrix of basis vectors for X. Furthermore, the covariance matrix of this transformation formalizes as:

$$\Sigma Q = E(Q^{\top}Q)=E(B^{\top}X^{\top}XB)=B^{\top}\Sigma_X B$$

It is important to reflect here on the dimensionality of ΣQ, which, unlike ΣX, is of dimension $p\times p$ where $p\leq m$. In other words, the covariance matrix under the transformation B is at most the size of the original covariance matrix, and possibly smaller. Since dimensionality reduction is clearly one of our objectives, the transformation above is certainly poised to do so. However, if *the objective is simply dimensionality reduction, then any matrix* B *of size* $m\times p$ *with* $p\leq m$ *will suffice; so question arises here why especially does* B *have to characterize basis.*

The answer is simple: dimensionality reduction is not the *only* objective, but one among *preservation of information* and *importance of information*. As to the former, we recall that what makes a set of basis vectors special is that they characterize *entirely* the space on which an associated matrix takes values and therefore *span* the multidimensional space on which that matrix resides. Accordingly, if B characterizes a basis, then information contained in X is never lost during the transformation to Q. Furthermore, that the channel for dimensionality reduction that motivated our discussion earlier was never intended to go through a sparser basis. Rather, the mechanism of interest was a diagonalization of the covariance matrix followed by variable exclusion. Accordingly, any dimension reduction that reflects basis sparsity via p≤m, is a consequence of perfect co-linearity (correlation) among some of the original system variables. In other words, p=rk(X), where rk(·) denotes the matrix *rank*, or the number of its linearly independent columns (or rows).

Diagonalization

We argued earlier that any transformation from X to Q that preserves information must operate through a basis transformation B. Suppose momentarily that we have in fact found such B. Our next objective would be to ensure that B also produces a diagonal ΣQ. In this regard, Two famous results in linear algebra are very important:

1) [Thm. 1:] A matrix is symmetric if and only if it is orthogonally diagonalizable.

- In other words, if a matrix A is symmetric, there exists a diagonal matrix D and a matrix E which diagonalizes A, such that A=EDE⊤⊤. The converse statement holds as well.
- [Thm. 2:] A symmetric matrix is diagonalized by a matrix of its orthonormal eigenvectors.
- Extending the result above, if a q×q matrix A is symmetric, the diagonalizing matrix E= [E1,…,Eq] = [1,…,], the diagonal matrix D=diag[λ1,…,λq], and Ei and λi are respectively the ith eigenvector and associated eigenvalue of A.
- Note that a set of vectors is orthonormal if each vector is of length unity and orthogonal to all other vectors in the set. Accordingly, if V=[V1,…,Vq] is orthonormal, then $V^{\top}_i V_i=1$ and $V^{\top}_i V_k=0$ for all j≠k. Furthermore, V⊤V=I_q⊤= where Iq is the identity matrix of size q, and therefore, $V^{\top}=V^{-1}$.
- Further that eigenvectors of a linear transformation are those vectors which only change magnitude but not direction when subject to said transformation. Since any matrix is effectively a linear transformation, if v is an eigenvector of some matrix A, it satisfies the relationship Av=λv. Here, associated with each eigenvector is the eigenvalue λ quantifying the resulting change in magnitude.
- Finally, observe that matrix rank determines the maximum number of eigenvectors (eigenvalues) one can extract for said matrix. In particular, if rk(A)=r≤q, there are in fact only r orthonormal eigenvectors associated with A. To see this, use a geometric interpretation to note that q−dimensional objects reside in spaces with q orthogonal directions. Since any n×q matrix is effectively a q−dimensional object of vectors, the maximum number of orthogonal directions that characterize these vectors is q. Nevertheless, if the (column) rank of this matrix is in fact r≤q, then q−r of the q orthogonal directions are never used. For instance, think of 2d drawings in 3d spaces. It makes no difference whether the drawing is characterized in the xy, the xz, or the yz plane -- the drawing still has 2 dimensions and in any of those configurations, the dimension left out is a linear combination of the others. In particular, if the xz plane is used, then the z−direction is a linear combination of the y−direction since the drawing can be equivalently characterized in the xy plane, and so on. In other words, one of the three dimensions is never used, although it exists and can be characterized if necessary. Along the same lines, if A indeed has rank r≤q, we can construct q−r additional orthogonal eigenvectors to ensure dimensional equality in the diagonalization A=EDET, although their associated eigenvalues will in fact be 0, essentially negating their presence.

By extension of the previous point, since A is a q−dimensional object of q−dimensional column vectors, it can afford at most q orthogonal directions to characterize its space. Since all q such vectors are collected in E, we are guaranteed that E is a spanning set and therefore constitutes an eigenbasis.

Since Cov(X)) is a symmetric matrix by construction, the 1st result above affords a re-express of equation (1) as follows:

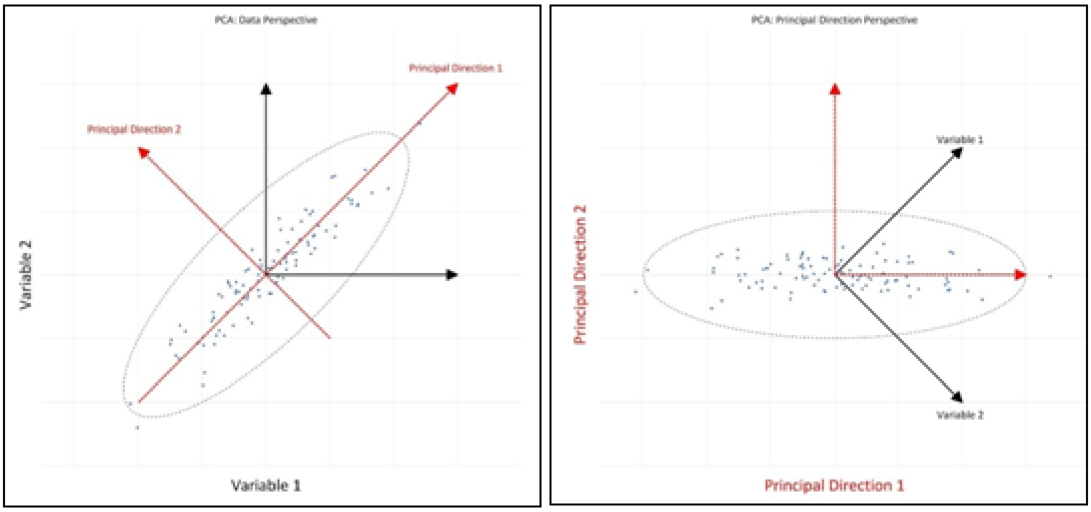
$$\Sigma Q = B^{\top}\Sigma_X B = B^{\top}E_X D_X E^{\top}{}_X B \qquad (2)$$

where E_X=[E1,…,Em]=[1,…,] is the orthonormal matrix of eigenvectors of ΣX and D_X=diag[λ1,…,λq] is the diagonal matrix of associated eigenvalues.

Now, since we require Σ_Q to be diagonal, we can set $B^{\top}=E^{1}$ in order to reduce Cov(Q) to the diagonal matrix D_X. Since the 2nd linear algebra result above guarantees that E_X is orthonormal, we know that $E^{-1}=E^{\top}$.

Accordingly ΣQ =D_X if and only if B=E_X

The entire idea is visualized below in Figures 1 and 2. In particular, Figure 1 demonstrates the ``data perspective" view of the system in relation to an alternate basis. That is, two alternate basis axes, labelled as ``Principal Direction 1" and ``Principal Direction 2" are superimposed on the familiar x and y axes. Since the vectors of a basis are mutually orthogonal, the principal direction axes are naturally drawn at 90^0 angles. Alternatively, Figure 2 demonstrates the view of the system when the perspective uses the principal directions as the reference axes.



Consistency

In practice, ΣX, and by extension ΣQ, E^X, and D^X, are typically not observed. Nevertheless, we can apply the analysis above using sample covariance matrices:

$S_Q=1/nQ^TQ \xrightarrow{p} \Sigma Q$

$n\to\infty$

and $S_X=1/nX^TX \xrightarrow{p} \infty\Sigma X$

$n\to\infty$

$$Where\ \frac{p}{n\to\infty}$$

indicates weak convergence to asymptotic counterparts. In this regard, the result analogous to equation (2) for estimated 2nd moment matrices states that $S_Q=\hat{}E^T{}_XS_X\hat{}E_X=\hat{}E^T{}_X=E^T{}_X(\hat{}E_X\hat{}D_X\hat{}E^T{}_X)\hat{}E_X=\hat{}D_{XW.}$
Where $\hat{}E_X$ and $\hat{}D_X$ now represent the eigen basis and respective eigen values associated with the square symmetric matrix S_X. It is important to understand here that while $\hat{}E_X\neq E_X\hat{}\neq$ and $\hat{}D_X\neq D_X\hat{}$, there is a long-standing literature far beyond the scope of this entry which guarantees that $\hat{}E_X\hat{}$ and $\hat{}D_X\hat{}$ are both consistent estimators of E_X and D_X, provided $m/n\to 0$ as $n\to\infty$. In other words, as in classical regression paradigms, consistency of PCA holds only under the usual ``large n and small m" framework. There are modern results which address cases for $m/n\to c>0$, however, they too are beyond the scope of this text. In proceeding however, in order to contain notational complexity, unless otherwise stated, we will maintain that E_X and D_X now represent the eigenbasis and respective eigenvalues associated with the square symmetric matrix S_X.

Preservation of Information

In addition to diagonalizing SQ, we also require preservation of information. For this we need to guarantee that B is a basis. Here, we recall the final remark under the 2nd linear algebra result above, which argues that S_Q affords at most m orthonormal eigenvectors and associated eigenvalues, with the former also forming an eigenbasis. Since all m eigenvectors are collected in $E_X=B$, we are guaranteed that B is indeed a basis. In this regard, we transform X into m statistically uncorrelated, but exhaustive directions. We are careful not to use the word variables (although technically they are), since the transformation $Q=XE_X$ does not preserve variable interpretation. That is, the jth column of Q no longer retains the interpretation of the jth variable (column) in X. In fact, the jth column of Q is a projection (linear combination) of all m variables in X, in the direction of the jth eigenvector Ej. Accordingly, we can interpret XE_X as m orthogonal weighted averages of the m variables in X. Furthermore, since E_X is an eigenbasis, the total variation (information) of the original system X, namely SX, is preserved in the transformation to Q. Unlike S_X however, $S_Q=D_X$ is diagonal, and total variation in X is now distributed across Q without redundancy.

Principal Directions

Since preservation of information is guaranteed under the transformation $Q=XE_X$, the proportion of information in S_X associated with the jth column of S_Q is in fact λj. By extension, each column in Q has standard deviation $\sqrt{\lambda j}$ or variance λj. Moreover, since S_Q is diagonal and information redundancy is not an issue, it stands to reason that the total amount of system variation is the sum of variations due to each column in Q. In other words, total system variation is $tr(S_Q)=\lambda 1+\ldots+\lambda m$, where $tr(\cdot)$ denotes the matrix trace operator, and the jth orthogonalized direction contributes to : $\Lambda j/(\lambda 1+\ldots+\lambda m)\times 100\%$ of total system variation (information).

If we now arrange the columns of Q, or equivalently those of E_X, according to the order $\lambda(1)\geq\lambda(2)\geq\ldots\geq\lambda(m)$, where $\lambda(j)$ are ordered versions of their counterparts λj, we are guaranteed to have the directions arranged from most principal to least, measured as the proportion of total system variation contributed by that direction.

Another useful feature of the vectors in E_X is that they quantify the proportion of directionality each original variable contributes toward the overall direction of that vector. In particular, let ei,j denote the ith element in $Ej=[e1,j,\ldots,em,j]=[1,\ldots,]$, where $i\in 1,\ldots,m$, and observe that since Ej are the eigenvectors of S_X, each element ei,j is in fact associated with the ith variable (column) of X. Furthermore, since the vectors Ej each have unit length due to (ortho)normality, we know that they must lie inside the unit circle and that $e^2_{i,j}\times 100\%$ of the direction Ej is due to variable i. In other words, we can quantify how principal each variable is in each direction.

Principal Components

Principal directions, the eigenvectors in EX, are often mistakenly called principal components. Nevertheless, correct literature reserves the term principal components for the projections of the original system variables onto the principal directions. That is, principal components refer to the column vectors in Q=[Q1,…,Qm]=XEX, and are sometimes also referred to as scores. Like their principal direction counterparts, principal components contain several important properties worth observing.

As a direct consequence of the diagonalization properties discussed earlier, the variance of each principal component is in fact the eigenvalue associated with the underlying principal direction, and principal components are mutually uncorrelated. To see this formally, let Cj=[0,…,0,1,0,…,0]⊤ denote the canonical basis vector in the jth dimension. Then, using the result in equation (4), the correlation between the jth and kth principal components Qj=QCj and Qk=QCk, respectively, is obviously:
sQj,Qk=1/nQ⊤jQk=C⊤j(1/nQ⊤Q)Ck=C⊤jSQCk=C⊤jDXCk which equals λj when j=k= and 0 otherwise.

Moreover, we can quantify how (co)related the original variables are with the principal directions. In particular, consider the covariance between the ith variable Xi=XCiand the jth principal component Qj, formalized as: SXiQj=1/nX⊤iQj=C⊤i(1nX⊤Q)Cj=C⊤i(1nX⊤XEX)Cj=C⊤iSXEXCj=C⊤iEXDXE⊤XEXCj=C⊤iEXDXCj=ei,jλj(5)where the antepenultimate line invokes Theorem 1 to SX, and the cancelation to identity in the penultimate line follows by Theorem 2 and orthonormality of EX, and the ultimate line is the product of the jth element of the principal direction Ej and the jth principal eigenvalue.

Dimension Reduction

At last, we arrive at the issue of dimensionality reduction. Assuming that the columns of Q are arranged in decreasing order of importance (more principal columns come first), we can discard the g<m least principal columns of Q until sufficient dimension reduction is achieved, and rest assured that the remaining (first) m−g columns are in fact most principal. In other words, the m−g directions which are retained, contribute to

$$\sum_{j=1}^{m-g} \lambda\,(j)/(\lambda 1+\cdots+\lambda m)\times 100\%$$

of the original variation in X. Since directions are ordered in decreasing order of importance, the first few directions will capture the majority of variation, leaving the less principal directions to contribute information only marginally. Accordingly, one can significantly reduce dimensionality whilst retaining the majority of information. This is particularly important when we want to measure the complexity of our data set. In particular, if the r most principal directions account for the majority of variance, it stands to reason that our underlying data set is in fact only r−dimensional, with the remaining m−r dimensions being noise. In other words, dimensionality reduction naturally leads to data denoising.

So how does one select how many principal directions to retain? There are several approaches, but we list only several below:

1) A very popular approach is to use a scree plot -- a plot of the ordered eigenvalues from most to least principal. The idea here is to look for a sharp drop in the function, and select the bend or elbow as the cutoff value, retaining all eigenvalues (and by extension principal directions) to the left of this value.
2) Another popular alternative is to use the cumulative proportion of variation explained by the first r principal directions. In other words, select the first r principal directions such that
$$\sum_{j=1}^{r} \lambda\,(j)(\lambda 1+\cdots+\lambda m)\geq 1-\ \alpha$$
where α ∈ [0, 1] ∈ [0, 1]. Typical uses set α = 0.1 in order to retain r most principal directions that capture at least 90% of the system variation.
3) A more data driven result is known as the Guttman-Kaiser (Guttman (1954), Kaiser (1960), Kaiser (1961)) criterion. This criterion advocates the retention of all eigenvalues, and by extension, the associated principal directions, that exceed the average of all eigenvalues. In other words, select the first **r** principal directions such that λ (1) +…+ λ (k) ≥ r¯ λ,
$$Where\ {}^{-}\lambda = 1/m\sum_{j=1}^{m}\lambda j$$
4) An entirely data-driven approach akin to classical information criteria selection methods borrows the Bai and Ng (2002) paper on factor models. In this regard, consider Xj=β1Q1+…+βrQr+U(j,r), as the regression of the jth variable in X on the first r principal components of S_X, and let ˆU(j,r) denote the corresponding residual vector. Furthermore, define SSR(j,r)=1/nˆU(j,r)⊤ˆU(j,r), as the sum of squared residuals from said regression, and define

$$SSR\ (r) = 1/\ m\sum_{j=1}^{m} SSR\ (j,r)$$

as the average of all SSR (j, r) (,) across all variables j for a given r. We can then select r as the one that minimizes a particular penalty function. In other words, the problem reduces to:

min {ln (SSR (r)) + rg (n, m)}
r

where g (n, m) is a penalty term which leads to one of several criteria proposed in Bai and Ng (2002). For instance when n>m, one such option is the ICp2(r) criterion, and the problem above formalizes as:

min{ln(SSR(r))+r(n+m/nm)ln(m)}
r

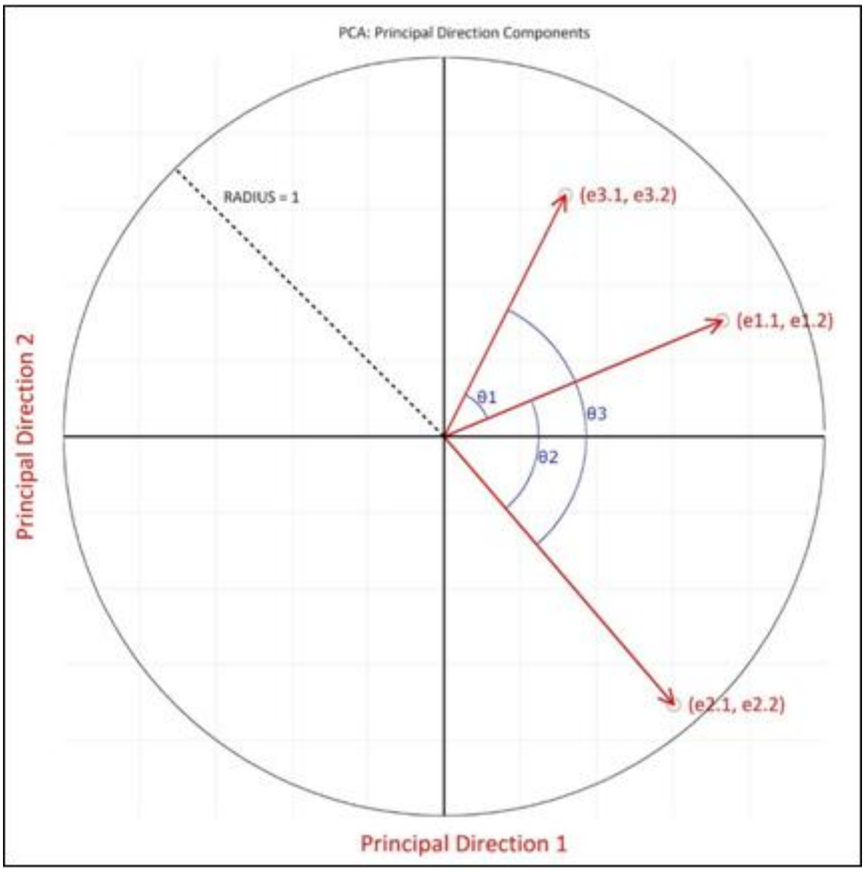
Of course, it goes without saying that discarding information comes at its own cost, although, if dimensionality reduction is desired, it may well be a price worth paying.

Loading Plots

A powerful inferential tool unique to PCA is element-wise comparison of two principal directions. In particular, consider two principal directions Ej=[e1, j,…, em, j] and Ek=[e1, k,…, em, k], and let {V1, j, k,…,Vm, j, k} denote the set of vectors from the origin (0,0) to (ei, j, ei, k) for i∈1, In other words, Vi, j, k= (ei, j, ei, k)ᵀ. Then, for any (j, k) principal direction pairs, a plot of all m vectors Vi, j, k, for i∈1,…, m, on a single plot, is called a loading plot.

There is an important connection between the vectors Vi, j, k and original variable covariances. In particular, consider SXi, Xs-- the finite sample covariance between Xi and Xs -- and, assuming we have ordered eigenvalues from most principal to least, note that:

SXi,Xs=$C^\top{}_i S_X Cs$=$C^\top iE_X D_X E^\top{}_X Cs$=λ(1)ei,es.1+λ(2)ei,2es.2+…+λ(m)ei,es.m=$V^\top i,_{1,2} L_{1,2} Vs,_{1,2}$+…+$V^\top i,_{m-1}$,m$L_{m,m-1} V_{s,m-1,m}$where Lj,k=diag[λ(j),λ(k)] denotes the appropriate scaling matrix. In other words, for any (j,k) principal direction pairs, $V^\top{}_{i,j}$,k$L_{j,k} V_{s,j,k}$ explains a proportion of the covariance SXi,Xs. Accordingly, when Xi and Xs are highly correlated, we can expect $V^\top{}_{i,j,k} L_{j,k} V_{s,j,k}$ to be larger values. In this regard, let $\theta_{i,s,j,k}$ denote the angle between any two vectors $V_{i,j,k}$and $V_{s,j,k}$ and recall that $\cos_{\theta i,s,j,k}=V^T{}_{i,j,k} V_{s,j,k} \| V_{i,j,k} \| \| V_{s,j,k} \|$.To accommodate the use of the scaling matrices Lj,k, observe that we can modify this result as follows:$V^\top{}_{i,j,k} L_{j,k} V_{s,j,k}=V^\top{}_{i,j,k} L_{j,k} (V_{i,j,k} V^\top{}_{i,j,k})^{-1} V_{i,j,k} \| V_{i,j,k} \| \| V_{s,j,k} \| \cos\theta_{i,s,j,k}$(6)Now, when $\theta_{i,s,j,k}$ is small, say between 0 and π/2, we can expect $V^\top{}_{i,j,k} L_{j,k} V_{s,j,k}$ to be large, and by extension, Xi and Xs to be more correlated. In other words, vectors that are close to one another in a loading plot indicate stronger correlations of their underlying variables. Figure 3 below gives a visual representation.



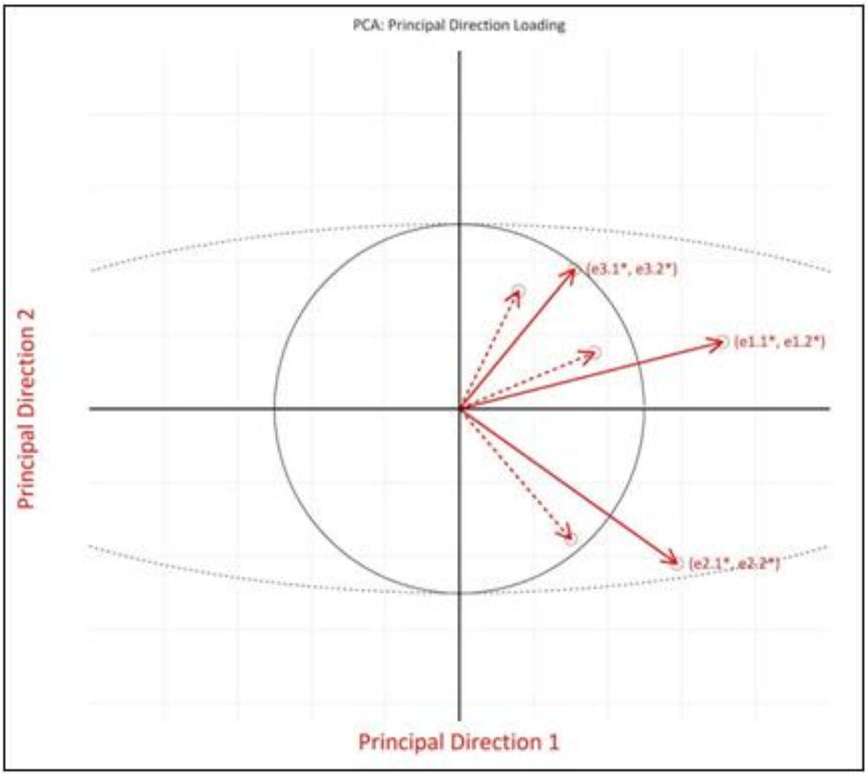
It is important to realize here that since θi, s, j, k is in fact the angle between Vi, j, k and Vs, j, k, the interpretation of how exhibitive θi,s,j,k is of the underlying correlation SXi, Xs is made more complicated by the presence of Lj, k in equation (6). Accordingly, to ease interpretation, the vectors Vi,j, k are sometimes scaled appropriately, or loaded with scaling information, leading to the term loadings. In this regard, consider the vectors $\tilde{V}_{i,j,k}=V_{i,j,k} L^{1/2}{}_{j,k}$. Here, loading is done via $L^{1/2}{}_{j,k}$, and we have:

SXi,Xs=˜V⊤i,1,2˜Vs,1,2+…+˜V⊤i,m−1,m˜Vs,m−1,mand˜V⊤i,j,k˜Vs,j,k=∥˜Vi,j,k∥∥˜Vs,j,k∥cos˜θi,s,j, kAs such, ˜θi,s,j,k more closely exhibits the true angle between Xi and Xs than θi, s, j, k, and loading plots using ˜Vi,j,k tend to be more exhibitive of the underlying correlations SXi, Xs than those based on Vi, j, k. Of course, one does not have to resort to the use of L1/2j,k as the loading matrix. In principle, one can use Lαj, k for some 0≤α≤1, although the underlying interpretation of what such a loading means ought to be understood first.

Of course, it is not difficult to see that $\tilde{V}$i,j, k=Vi,j, kL^αj, k is in fact the ith "XY"-pair between Ej$\lambda^{\alpha j}$ and Ekλ^αk. In other words, it is the ith "XY"-pair using the "loaded" jth and kth principal directions. Accordingly, the term *loading vector* is sometimes used to denote a loaded principal direction. In particular, the entire matrix of loading vectors ˜EX can be obtained as follows:

˜EX=EXD$^{\alpha}$X

Figure 4 below demonstrates the impact of using a loading weight. In particular, the vectors in Figure 3 are superimposed on the set of loaded vectors where the loading factor is $D^{1/2}_{X}$. Clearly, the loaded vectors are much more correlated with the general shape of the data as represented by the ellipse.



Scores Plots

A score plot across principal direction pairs (j, k) is essentially a scatter plot of the principal component vector Qi vs. Qj. In fact, it is the analogous version of the loading plot, but for observations as opposed to variables. In this regard, whereas the angle between two loading vectors is exhibitive of the underlying correlation between some variables, the distance between observations in a score plot exhibits homogeneity across observations. Accordingly, observations which tend to cluster together tend to move together, and one typically looks to identify important clusters when conducting inference.

Recall also the expression derived in the last line of 5, namely, SXiQj=ei,jλj=(ei,jλ$^{1/2}$j)λ$^{1/2}$j. Notice that the latter expression states that the correlation between the ith variable and the jth score vector is in fact a product of the jth element of a loaded ith principal direction (ith loading vector), and λ1/2j. Accordingly, in order to achieve a more natural interpretation, one can proceed in a manner analogous the creation of loading vectors, and either scale or entirely remove the remaining scaling factor. This leads to the idea of loaded score vectors. In particular, using the context above, if one wishes to interpret the correlation between the ith variable and the jth score vector as just a loaded principal direction without the additional factor λ1/2j, then doing so is as simple as computing SXiQjλ−1/2j=ei,jλ1/2j where we now interpret Qjλ−1/2j as a loaded score vector. Of course, an infinite array of such scaling options is achievable using Qjλ−αj, although, as before, their interpretation ought to be understood first.

Outlier Detection

An important application of PCA is to outlier detection. The general principle exploits the first few principal directions to explain the majority of variation in the original system, and uses data reconstruction to generate an approximation of the original system using the first few principal components.

Formally, if we start from the matrix of all principal components Q, it is trivial to reconstruct the original system X using the inverse: $QE^{\top}_{X}=XE_{X}E^{\top}_{X}=X$, On the other hand, if we restrict our principal components to the first r≪m most principal directions, then ˜Q˜E⊤X=˜X≈X, where ˜Q~ and ˜E_X~ are respectively the matrix Q and E_X with the last m−r− columns removed, and ≈≈ denotes an approximation. Then, the difference ξ=˜X−X is known as the reconstruction error, and if the first r principal directions explain the original variation well, we can expect $\|\xi\|^2$ where $\|\cdot\|_D$ denotes some measure of distance.

We would now like to define a statistic associated with outlier identification, and as in usual regression analysis, the reconstruction error (residuals) plays a key role. In particular, we follow the contributions of Jackson and Mudholkar (1979) and define SPE=ξξ⊤ as the squared prediction error most resembling the usual sum of squared residuals. Moreover, Jackson and Mudholkar (1979) show that if observations (row vectors) in X are independent and identically distributed, Gaussian random variables, SPE has the following distribution

$$SPE \sim \sum_{j+1}^{m} \lambda\,(j) Z^2 j = \psi\,(k)$$

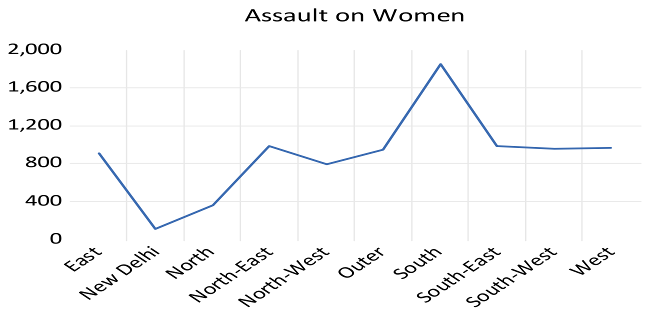
where $\chi 2p$ denotes the $\chi 2$distribution with p degrees of freedom, and Zj are independent $\chi 2_1$ variables. Noting that the ith diagonal element of SPE, namely SPEii=CTi(SPE)Ci=T is associated with the itervation, we can now derive a rule for outlier detection. In particular, should SPEii, for any i, fall into some critical region defined by the upper $(1-\alpha)$ percentile of $\Psi(k)$, that observation would be considered an outlier.

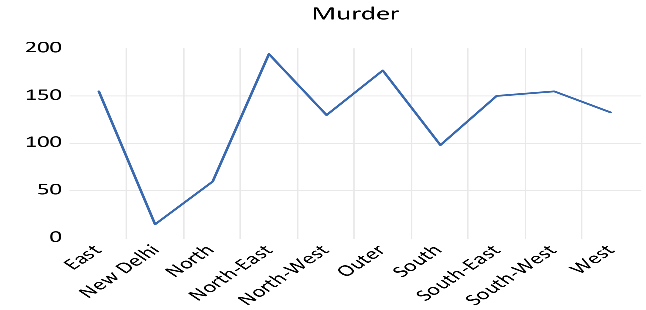
Data Source

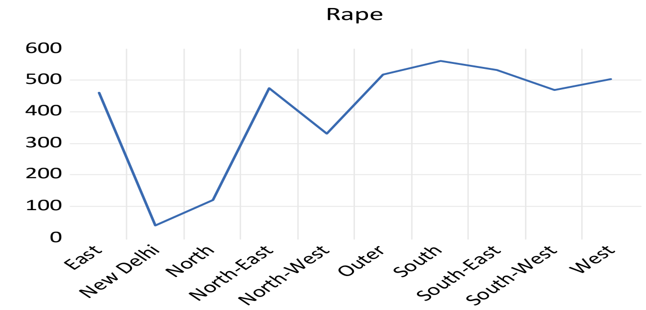
Secondary data from official website of Delhi Police, Govt. open data base.

The requirement of complete data to conduct PCA was available for only 2014, hence the base of study is data of year 2014. A descriptive statistics of data is shown below:

Date: 10/04/23 Time: 10:29			
Sample: 1 10			
	Assault_ on Women	Murder	Rape
Mean	887.4000	126.0000	400.4000
Median	949.5000	141.0000	471.0000
Maximum	1855.000	194.0000	560.0000
Minimum	115.0000	14.00000	40.00000
Std. Dev.	453.6749	54.90396	180.7240
Skewness	0.377959	-0.870593	-1.176003
Kurtosis	3.816601	2.816592	2.822309
Jarque-Bera	0.515937	1.277236	2.318126
Probability	0.772620	0.528022	0.313780
Sum	8874.000	1260.000	4004.000
Sum Sq. Dev.	1852388.	27130.00	293950.4
Observations	10	10	10







A graphical representation of assault, rape and murders on women is shown above

Covariance Analysis: Ordinary			
Date: 10/04/23 Time: 10:30			
Sample: 1 10			
Included observations: 10			
Correlation	Assault on Women	Murder	Rape
Assault on Women	1.000000		
Murder	0.486901	1.000000	
Rape	0.850994	0.813933	1.000000

A quick interpretation of the correlation structure indicates that assault on women is highly correlated with rape and moderately correlated with murder of women with intention to outrage her modesty.

Principle component analysis of crime data:

Principal Components Analysis					
Date: 10/04/23 Time: 10:31					
Sample: 1 10					
Included observations: 10					
Computed using: Ordinary correlations					
Extracting 3 of 3 possible components					
Eigenvalues: (Sum = 3, Average = 1)					
				Cumulative	Cumulative
Number	Value	Difference	Proportion	Value	Proportion
1	2.445782	1.931951	0.8153	2.445782	0.8153
2	0.513831	0.473443	0.1713	2.959613	0.9865
3	0.040387	---	0.0135	3.000000	1.0000
Eigenvectors (loadings):					

Paper ID: SR231010120234 DOI: 10.21275/SR231010120234

Variable	PC 1	PC 2	PC 3		
Assault on Women	0.554296	-0.682555	0.476313		
Murder	0.542187	0.730302	0.415563		
Rape	0.631497	-0.027906	-0.774876		
Ordinary correlations:					
	Assault on Women	Murder	Rape		
Assault on Women	1.000000				
Murder	0.486901	1.000000			
Rape	0.850994	0.813933	1.000000		

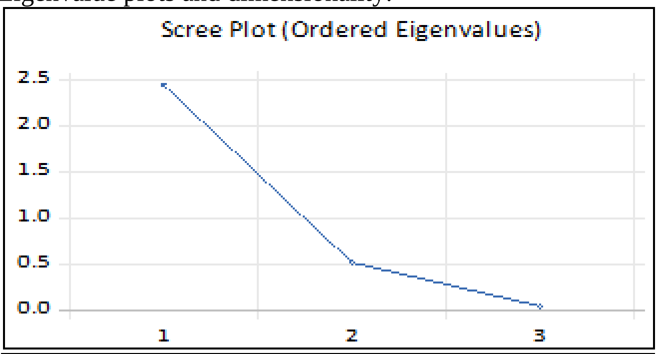
The first table summarizes the information on eigenvalues. The later are sorted in order of principality (importance), measured as the proportion of information explained by each principal direction. Here, we see that the first principal direction explains roughly 81% of the information contained in the underlying correlation matrix, the second roughly 9%. The cumulative proportion of information explained by first two is roughly 91%.

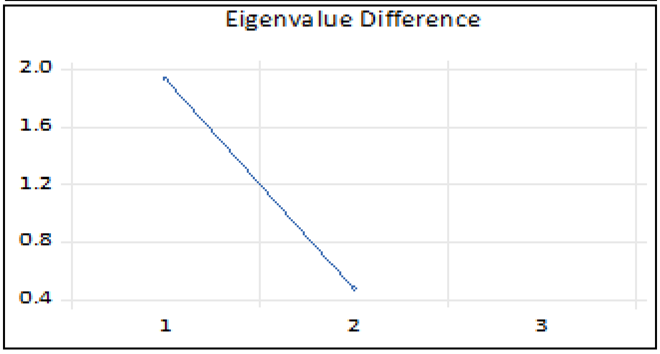
The second (middle) table summarizes the eigenvectors associated with each of the principal eigenvalues., the eigenvectors are also arranged in order of principality. Furthermore, whereas the eigenvalues highlight how much of the overall information is extracted in each principal direction, the eigenvectors reveal how much weight each variable has in each direction.

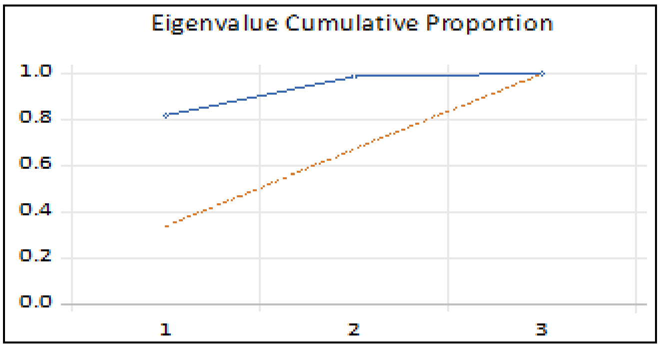
Accordingly, the relative importance of any variable in a given principal direction is effectively the proportion of the eigenvector length (unity) attributed to that variable. As shown here as under:

	PC1 account for		PC2 account for	
Assault on women	0.554296	0.307244056	-0.682555	
Murder	0.542187	0.293966743	0.730302	0.533341011
Rape	0.631497	0.398788461	-0.027906	

Eigenvalue plots and dimensionality:







The first is the scree plot-a line graph of eigenvalues arranged in order of principality. Superimposed on this a red dotted horizontal line with a value equal to the average of the eigenvalues. The idea here is to look for a kink point, or an elbow, and retain all eigenvalues, and by extension their associated eigenvectors, that from the first portion of the kink, and discard the rest. From the plot, it is evident that a kink occurs at the 2nd eigenvalue, indicating that we should retain the first two eigenvalues.

A slightly more numeric approach discards all eigenvalues significantly below the eigenvalue average. We see that the average of the eigenvalues is 1.00, and the second eigenvalue is just below cut off. Since the 2nd value is close to this average, it would be safe to conclude that the scree plot analysis indicates that only first two eigenvalues ought to be retained.

The second graph plots a line graph of the differences between successive eigenvalues. Superimposed on this graph is another horizontal line, this time with a value equal to the average of the differences of successive eigenvalues. Which is 0.80. The idea here is to retain all eigenvalues whose differences are above this threshold. Clearly, only the first two eigenvalues satisfy this criterion.

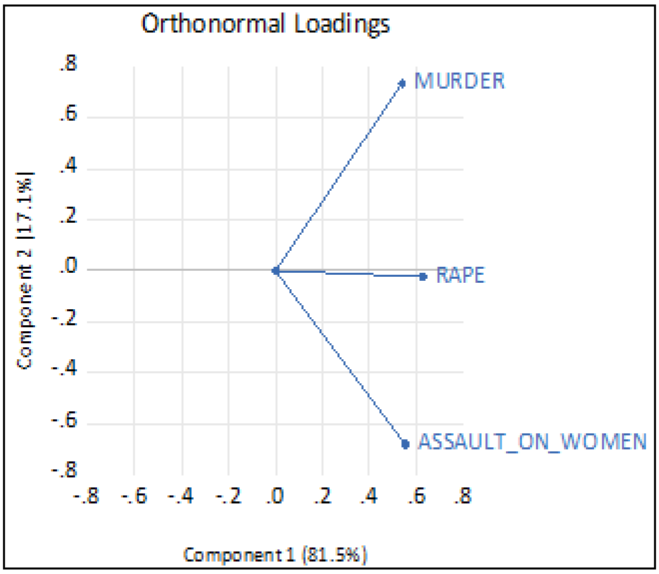
The final graph is a line graph of the cumulative proportion of information explained by successive principal eigenvalues. Superimposed on this graph is line with a slope to the average of the eigenvalues namely 1, the idea here is to retain those eigenvalues that form segments of the cumulative curve whose slopes are at least as steep as the line with slope 1.In our case, only two eigenvalues seem to form such a segment: eigenvalues1 and 2.

All three graphical approaches indicate that one ought to retain the first two eigenvalues and their associated eigenvectors.

Principal Direction Analysis:

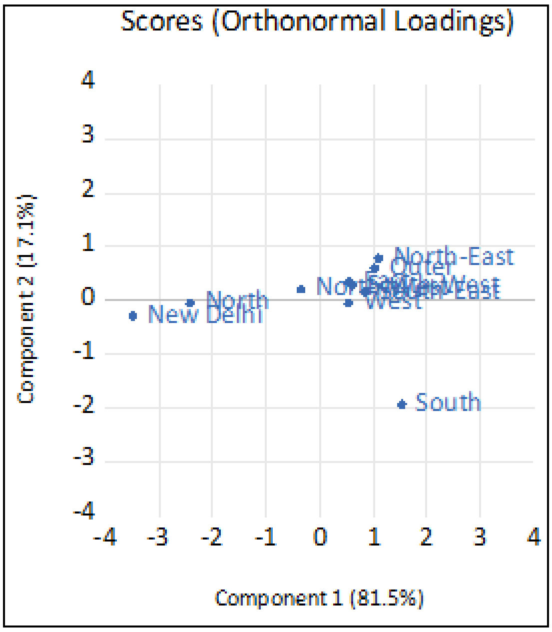
Variable Loading Plot:

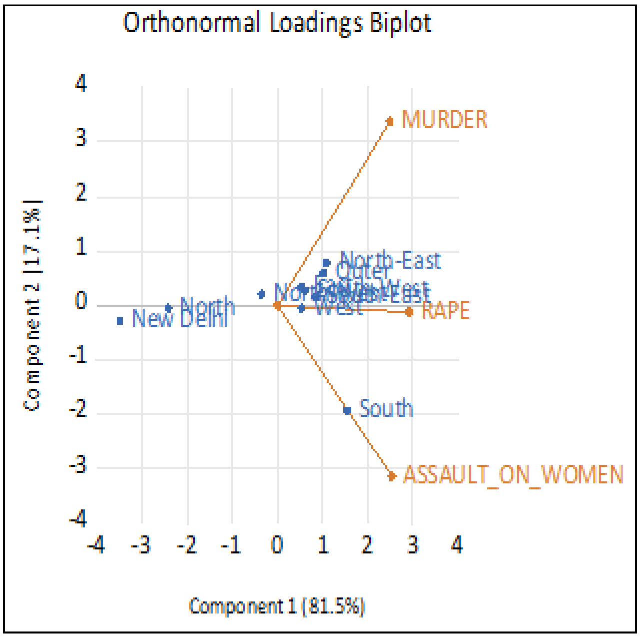
We plot variable loading graph that produces "X Y" pair plots of loading vectors. Here, we will plot the true eigenvectors since scaling is unity.



In this plot, the angle between the vectors in a loading plot is related to the correlation between the original variables to which the loading vectors are associated. Accordingly, we see that Assault on women with intent to outrage and modesty and rape of women, are strongly positively correlated with rape. Further, we also conclude that Assault on women with intent to outrage and modesty and murder of women is nearly uncorrelated since they form a near 90 degree angle.

Component Score Plot and Bi-plot:





Now, from Score plot as shown above, We conclude that if we refer east of the zero vertical axis, a state is located comprising NE, Outer, SE and SW part of Delhi which is positively dominated by the crimes of Assault on women with intent to outrage and modesty , rape of women and murder. Conversely, North, North Delhi part of Delhi are typically less associated with above said crimes.

Further, the north of the zero horizontal axis a state is located, the more positively correlated it is with the second principal direction. Since the latter is dominated positively (north of the zero horizontal axis) by the variable murder. We conclude that such states are positively correlated with said crimes in area of NW Delhi. Conversely, the area of South Delhi is typically less dominated by these crimes.

2. Conclusion

Concluded that area of NE, SE, SW and outer part of Delhi is highly dominated by the crimes of Assault, Rape and Murder. The north-east district is turning out to be the crime underbelly of Delhi. This area reported the largest number of heinous crimes among the 11 police districts—about 40% of 16,273 last year. These include cases of murder and attempt to murder, rape, molestation, riots and kidnappings. Outer district was the second most crime-infested area,,,,,,,,, with a 30% share. This has been highlighted in a white paper on “State of Policing and Law and order in Delhi”, released by NGO Praja and the data obtained through RTI.

The north-east district is an amalgamation of resettlement and unauthorised colonies and has easy access of Uttar Pradesh. The colonies include Seemapuri, Gokulpuri, Bhajanpura, Shahdara, Seelampur and Khajuri Khas besides some middle class areas like Mansarovar Park. On the other side is outer which has a mix of rural areas and unauthorised colonies, bordering Haryana, besides the middle class Rohini and other similar pockets. The areas include Narela, Alipur, Bawana, Khanjawla, Shahbad Dairy, Mangolpuri, Sultanpuri and SamaipurBadli.

According to the police, most of these areas have a huge migrant population from across the country. Poverty, lack of amenities, vulnerability to exploitation, minority and majority communities living cheek by jowl, out of school children lrft to fend for themselves at home as their parents toil to eke out their daily existance and social factors like alcoholism and propensity for violence make these areas ripe for crime. The rural areas are caught between an agricultural past and an urban present.

North-West Delhi has earned the dubious distinction of being the most crime prone district in the city. The latest official statistics released by the Delhi Police reveals that the North-West District tops the crime figures among all the nine districts in Delhi.

About 500 cases of heinous crimes in NorthWest Delhi last year have not been unexpected and all-together surprising for the police. Indeed, as police officials in the district testify, the staggering crime figures for North-West Delhi have been along the expected lines and in keeping with the recent trends.

For instance, the 135 cases of homicide last year in the region have been only slightly higher than the 128 such cases in the preceding year. Till the first week of November this year, the region had so far witnessed 118 cases of killing and the final count at the end of the current year is also expected to hover in the vicinity of last year"s figures. Similarly, the police had till the first week of November recorded 80 cases pertaining to attempted murders, which is again close to the heels of such 107 total cases reported for the whole of last year in the northwest district.

The North-West District remains the most susceptible region in terms of public safety and from the point of view of security perceptions. Police officials maintain that the notoriety stems largely as a result of the wanton mushrooming of new slum clusters in the area in addition to the already large proliferation of slums and unauthorised colonies spreadout in this region, second only to East Delhi in terms of size and population.

"The region has an unusually high percentage of migrants from other states besides the disturbing presence of a large number of illegal nationals from Bangladesh, Nepal and Afghanistan," "People hailing from outside Delhi but living within the city have almost no stake in keeping the city safe and secure, and are wont to opportunistic adventures that are prejudicial to law and order,"

While the police have been devising strategies and game plans to be one up on criminals, they face some serious drawbacks of their own in effectively tackling crime. Police stations in places like Maurice Nagar, Timarpur, Pratap Nagar and Mukherji Nagar have all through been operating without custody cells, while some police stations operate from ramshackle buildings. There is a shortage of sufficient number of police personnel and police stations in many of the big areas like Rohini, where people have been demanding setting up of more police stations for several years now.

Further, there is a scope for an early predicting the crimes in identified hot spot areas using advanced algoriths. Machine learning ensemble-based algorithms can handle well such problems. Various methods like Weighted Moving Average, Functional Coefficient Regression and Arithmetic-Geometric Progression based prediction of the crime in coming years

References

[1] DL Weisburd, T McEwen - Available at SSRN 2629850, 2015 - papers.ssrn.com

[2] S Khalid, F Shoaib, T Qian, Y Rui, AI Bari… - Applied Spatial Analysis …, 2018 – Springer

[3] B Van Schaack - 2016 - papers.ssrn.com

[4] J Ratcliffe - Handbook of quantitative criminology, 2010 - Springer

[5] N Levine - Geographical analysis, 2006 - Wiley Online Library

[6] JH Ratcliffe - European Journal on Criminal policy and research, 2004 – Springer

[7] J Eck, S Chainey, J Cameron, R Wilson - 2005 - discovery.ucl.ac.uk

[8] R Paynich, B Hill - 2013 - books.google.com

[9] T Lorenc, S Clayton, D Neary, M Whitehead… - Health & place, 2012 – Elsevier

[10] S Chainey, L Tompson, S Uhlig - Security journal, 2008 – Springer

[11] P Kedia - International Institute of Information Technology, 2016 - researchgate.net

[12] SD Johnson, DJ Birks, L McLaughlin… - London: Home …, 2007 - researchgate.net