

Structural Analysis of Convicts Received into Prison by Province in Türkiye Using Spectral Decomposition

Şenol Çelik¹

¹Unit of Biometrics and Genetics, Department of Animal Science, Faculty of Agriculture, Bingöl University, Bingöl-Türkiye

Abstract

This study aims to construct a spatial and structural crime typology by analyzing the crime characteristics of 81 provinces in Türkiye using multidimensional statistical techniques. The dataset was obtained from the Türkiye Statistical Institute (TURKSTAT) and includes 25 crime categories reflecting the number of convicts received into prison by province.

Principal Component Analysis (PCA), Spectral Clustering, and Hierarchical Clustering (Ward's method) were applied within an integrated analytical framework. The findings reveal that the crime structure in Türkiye is predominantly shaped by a single dominant axis, interpreted as overall crime volume, explaining 85.7% of the total variance. Hierarchical clustering results indicate that provinces are stratified into three principal structural layers.

Across all analytical techniques, Istanbul emerges as a distinct outlier, structurally detached from the remaining 80 provinces. Ankara, İzmir, and Antalya form a secondary metropolitan block characterized by higher crime diversity, whereas the majority of Anatolian provinces exhibit a comparatively homogeneous crime structure.

ANOVA results confirm the statistical validity of the clustering structure, demonstrating that forgery ($F = 2165.45$) and cyber crimes are the most discriminating variables underlying inter-provincial differentiation.

Overall, the integrated use of spectral decomposition and clustering techniques provides a robust methodological framework for uncovering latent structural patterns in regional crime dynamics.

Keywords: Spectral Clustering; PCA; Crime Analysis; Crime Hierarchy; Structural Differentiation

Date of Submission: 14-02-2026

Date of acceptance: 28-02-2026

I. Introduction

Crime is a multidimensional phenomenon shaped by the complex interaction of socio-economic, demographic, and spatial factors, and it directly influences public security and governance policies. Contemporary criminological theory emphasizes that crime is not spatially homogeneous; rather, it tends to cluster in specific geographic areas and exhibit structural stratification across regions (Clifford Shaw and Henry D. McKay, 1942). Social disorganization theory argues that rapid urbanization, residential mobility, and socio-economic inequality create environments conducive to crime concentration.

Similarly, environmental criminology highlights the importance of spatial opportunity structures and routine activity patterns in shaping crime distribution (Ronald V. Clarke, 1980). These perspectives suggest that metropolitan regions tend to develop structurally differentiated crime ecosystems compared to peripheral or rural areas.

Beyond classical criminological theories, urban scaling research demonstrates that many socio-economic indicators including crimescale nonlinearly with city size (Luis Bettencourt et al., 2007). Larger metropolitan areas often exhibit superlinear growth in complex crime types, reflecting increased economic interaction, density effects, and network complexity. This perspective implies that crime is not merely a quantitative phenomenon but also undergoes qualitative transformation with urban expansion.

Despite this theoretical background, empirical crime analyses frequently rely on aggregate crime counts. While such approaches provide a general overview of crime intensity, they may obscure substantial structural differences in crime composition across regions. Provinces with similar total crime volumes may differ significantly in terms of offense typology and diversity. Therefore, a multidimensional statistical framework is required to uncover latent structural dimensions underlying regional crime patterns.

To address this gap, the present study applies Principal Component Analysis (PCA), originally developed by Karl Pearson (1901) and later formalized by Harold Hotelling (1933), in order to reduce dimensional complexity and identify dominant variance structures. PCA enables the transformation of correlated crime variables into orthogonal components that capture the most significant structural patterns (Jolliffe and Cadima, 2016).

In addition, Spectral Clustering techniques grounded in spectral graph theory (Fan Chung, 1997) are employed to detect latent structural separations within lower-dimensional eigenspaces. Hierarchical clustering using Ward's minimum variance method (Joe H. Ward, 1963) complements the analysis by revealing hierarchical stratification among provinces. The statistical validity of cluster differentiation is evaluated using One-Way ANOVA, based on the foundational work of Ronald A. Fisher (1925).

Using crime data covering 81 provinces in Türkiye, this study demonstrates that the national crime structure is overwhelmingly dominated by a single principal dimension explaining 85.7% of the total variance. Furthermore, the findings reveal pronounced hierarchical breaks between metropolitan and non-metropolitan provinces. Istanbul, in particular, emerges as a statistically isolated crime ecosystem, primarily driven by modern urban offenses such as forgery and fraud.

The main objective of this study is to mathematically identify structural similarities and divergences among provinces by reducing high-dimensional crime data into interpretable components. By uncovering latent structural typologies, the study aims to provide an empirical basis for region-specific security strategies and evidence-based policymaking.

II. Material

The data used in this study were obtained from the official database of the Turkish Statistical Institute (TURKSTAT). Specifically, the dataset was retrieved from the Statistical Themes section of the TSI Data Portal under the title: "Convicts received into prison by type of crime and province where the crime was committed at SR level 3, 2011–2020."

The dataset covers 81 provinces in Türkiye and includes 25 different crime categories. These 25 crime types constitute the variables of the study and represent the crime structure at NUTS Level 3 (provincial level).

The analysis was conducted using aggregated provincial data covering the period 2020. Each province is treated as a statistical unit, and each crime type represents a quantitative variable reflecting the number of convicts received into prison for the corresponding offense.

The 25 crime variables included in the analysis are as follows:

x1: Homicide, x2: Damage to property, x3: Traffic crimes, x4: Bad treatment, x5: Forgery, x6: Assault, x7: Swindling, x8: Sexual crimes, x9: Production and commerce of drugs, x10: Use and purchase of drugs, x11: Embezzlement, x12: Bribery, x13: Smuggling, x14: Forestry crimes, x15: Crimes related with firearms and knives, x16: Opposition to the Military Criminal Law, x17: Opposition to the Bankruptcy and Enforcement Law, x18: Threat, x19: Prevention of performance, x20: Contrary to the measures for family protection, x21: Kidnapping, x22: Defamation, x23: Theft, x24: Robbery, x25: Other.

III. Method

Integrated Analytical Framework

In order to understand the structural dynamics of crime in Türkiye, advanced data mining and dimensionality reduction techniques were employed in an integrated analytical framework. The methodological design consists of three complementary stages: dimensionality reduction, clustering analysis, and statistical validation.

Spectral Decomposition and Principal Component Analysis (PCA)

The primary goal of multivariate data analysis is to reveal the structural relationships among variables. In this study, Principal Component Analysis (PCA) is based on the eigenvalue (spectral) decomposition of the covariance (or correlation) matrix.

For a $p \times p$ covariance matrix Σ , spectral decomposition is defined as:

$$\Sigma = Q\Lambda Q^T$$

where: Q is the orthogonal matrix of eigenvectors. Λ is the diagonal matrix containing eigenvalues. $Q^T Q = I$

The eigenvalues satisfy:

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$$

Total variance is given by:

$$\sum_{j=1}^p \lambda_j$$

The proportion of explained variance for the j^{th} component is:

$$\text{Explained Variance Ratio}_j = \frac{\lambda_j}{\sum \lambda_j}$$

This decomposition determines the orthogonal directions that maximize variance in the data space (Ian Jolliffe and Jorge Cadima, 2016).

Principal Components

Let the standardized data matrix be:

$$Z = \frac{X - \bar{X}}{s}$$

The first principal component is defined as:

$$PC_1 = a_{11}Z_1 + a_{12}Z_2 + \dots + a_{1p}Z_p$$

subject to the maximization problem:

$$\max \text{Var}(a_1^T Z) \text{ subject to } a_1^T a_1 = 1$$

Subsequent principal components are orthogonal:

$$a_i^T a_j = 0 \quad (i \neq j)$$

PCA was first introduced by Karl Pearson (1901).

In this study, PCA was applied to reveal the common variance structure among livestock production indicators (Merino wool, native wool, goat hair, number of beehives, honey, and beeswax production).

To manage the high dimensionality of the dataset and to extract the fundamental structural characteristics of crime patterns, Principal Component Analysis (PCA) was applied. PCA reduces correlated variables into a smaller number of orthogonal components that retain most of the total variance in the data.

Using the correlation matrix of standardized variables, principal components were derived in descending order of explained variance. This approach allows identification of latent crime dimensions representing common variance structures across provinces.

The theoretical foundation of PCA was originally introduced by Karl Pearson (1901) and later formalized in multivariate statistical theory by Harold Hotelling (1933). Contemporary methodological developments are discussed in Ian Jolliffe and Jorge Cadima (2016).

Principal Component Analysis was employed to reduce dimensionality and identify latent production structures (Jolliffe, I. T., 2002). Hierarchical clustering using Ward's minimum variance method (Ward, J. H., 1963) was subsequently applied to PCA scores to classify districts into homogeneous groups.

Cluster Analysis (K-Means)

The K-means algorithm partitions observations into k clusters by minimizing the within-cluster sum of squares:

$$\min \sum_{i=1}^k \sum_{x \in C_i} \|x - \mu_i\|^2$$

where:

- C_i denotes the i^{th} cluster
- μ_i is the centroid of cluster i
- $\|x - \mu_i\|^2$ represents squared Euclidean distance

Algorithm Steps

1. Random initialization of cluster centroids
2. Assignment of observations to the nearest centroid
3. Recalculation of cluster centroids
4. Iteration until convergence

The algorithm was originally proposed by Stuart Lloyd (1982).

Clustering Analyses

To identify structural similarities and regional divergences among provinces, Spectral Clustering and Hierarchical Clustering (Ward's Method) were employed simultaneously.

Spectral Clustering

Spectral clustering analyzes the graph structure of the data in a transformed eigenspace derived from the Laplacian matrix. Unlike traditional distance-based clustering, this method captures non-linear relationships and community structures in lower-dimensional subspaces. The theoretical framework of spectral graph methods is based on the work of Fan Chung (1997).

By projecting provinces into the spectral domain, latent structural separations in crime patterns were identified.

Hierarchical Clustering (Ward’s Method)

Hierarchical clustering was applied using Ward’s minimum variance criterion, which minimizes the total within-cluster variance at each step of agglomeration. The objective function seeks to merge clusters that result in the smallest increase in within-cluster sum of squares.

Ward’s method was introduced by Joe H. Ward (1963). Dendrograms were used to visualize hierarchical distances among provinces and to determine structural breakpoints.

The simultaneous use of spectral and hierarchical clustering enhances robustness and allows cross-validation of clustering structures.

Statistical Validation: One-Way ANOVA

To assess the statistical validity of clustering results, One-Way Analysis of Variance (ANOVA) was conducted separately for each crime variable. ANOVA tests whether the mean values of crime variables differ significantly across clusters. The F-statistic was used to evaluate:

- The significance of between-group differences
- The discriminatory power of each crime type
- The internal consistency of cluster structures

The methodological foundation of ANOVA is based on the work of Ronald A. Fisher (1925).

Statistically significant F-values indicate that the corresponding crime variable meaningfully contributes to cluster differentiation, thereby supporting the structural validity of the clustering framework.

IV. Results

Crime types across 81 provinces in Türkiye were classified using clustering analysis based on spectral decomposition. By projecting provinces into the eigenspace derived from the similarity matrix, structurally homogeneous groups were identified according to crime composition patterns.

Cluster-Based Crime Averages

Table 1 summarizes the characteristic differences between the identified clusters in terms of mean crime values for each offense type. The table provides a quantitative overview of how crime profiles vary across structurally differentiated provincial groups.

Table 1. Mean Crime Values by Cluster Identified through Spectral Clustering Analysis

Cluster	x1	x3	x5	x10	x23	x25
Cluster 1	156.19	439.62	136.24	314.05	1114.81	1114.48
Cluster 2	29.61	81.07	21.15	36.24	153.75	223.63
Cluster 3	1052	1347	2540	2641	6793	6951

x1: Homicide, x3: Sexual, x5: Forgery, x10: Cybercrime, x23: Theft, x25: Other. Cluster 1: Metropol, Cluster 2: Anadolu, Cluster 3: İstanbul

The cluster means obtained from the analysis clearly reveal the spatial hierarchy of crime across provinces. An examination of Table 1 demonstrates that Cluster 3 (Istanbul) diverges dramatically from the other clusters, particularly in qualified and economically embedded offenses such as forgery (x5 = 2540.0) and cybercrime (x10 = 2641.0). These exceptionally high mean values indicate not merely quantitative intensity but structural differentiation in crime composition.

In contrast, the provinces grouped under Cluster 2 (Anatolian provinces) exhibit crime rates that remain at baseline national levels across all variables. The relative homogeneity and lower magnitude of crime averages in this cluster suggest a structurally stable and less diversified crime profile.

The magnitude of separation between Cluster 3 and the remaining clusters supports the interpretation of Istanbul as a distinct metropolitan crime ecosystem rather than simply a high-population outlier. The concentration of technologically and financially mediated offenses reinforces arguments from urban scaling theory (Luis Bettencourt et al., 2007), which propose that complex socio-economic activities increase disproportionately with city size.

Overall, the cluster-based differentiation confirms that crime stratification in Türkiye follows a hierarchical and scale-sensitive structure, with metropolitan complexity serving as the primary axis of divergence.

ANOVA and Variable Importance Analysis

To statistically validate the discriminative power of each crime variable across the identified clusters, a One-Way Analysis of Variance (ANOVA) was conducted. This procedure tests whether mean differences between

clusters are statistically significant and quantifies the extent to which each crime type contributes to inter-provincial differentiation.

Table 2. ANOVA Results for the Discriminatory Power of Crime Variables Across Clusters

Variable	(F-Value)	p (Pr>F)	Sig
x5 (Forgery)	2165.446	< 2.2e-16	***
x7 (Swindling)	905.965	< 2.2e-16	***
x21 (Kidnapping)	553.416	< 2.2e-16	***
x24 (Robbery)	551.099	< 2.2e-16	***
x8 (Sexual crimes)	515.591	< 2.2e-16	***

The results of the One-Way ANOVA conducted to assess the statistical significance of the clustering structure are summarized in Table 2. For all crime variables, statistically significant differences were observed across clusters ($p < 0.001$), indicating that the spectral clustering solution captures meaningful structural variation among provinces.

Among all variables, x5 (Forgery) ($F = 2165.45$) and x7 (Fraud) ($F = 905.97$) exhibit exceptionally high F-statistics, identifying them as the most powerful discriminators in provincial cluster assignment. The magnitude of these F-values indicates that between-cluster variance overwhelmingly exceeds within-cluster variance for these crime types. This provides strong statistical confirmation of the high discriminatory capacity of the spectral clustering model.

The F-statistic represents the ratio of between-group variance to within-group variance. Therefore, higher F-values directly indicate stronger structural differentiation across clusters. The dominance of forgery and fraud suggests that the primary axis of provincial crime differentiation in Türkiye is driven by economically motivated and qualified offenses, rather than by conventional violent crimes.

In addition to economic crimes, relatively high F-values observed for x21 (Deprivation of Liberty) and x24 (Robbery) indicate that regional differentiation is not solely economic in nature but is also reinforced by public-order and coercive offenses. These findings suggest a multidimensional crime structure in which economic complexity and social security dynamics jointly shape provincial crime typologies.

Conversely, x14 (Forest Crimes) and x12 (Bribery) display the lowest F-values, indicating limited discriminatory capacity. This suggests that these crime types are more homogeneously distributed across provinces and do not create significant spectral breaks in the similarity structure.

From a spectral graph theory perspective, the spectral decomposition method identifies eigenvectors corresponding to the largest variance components of the Laplacian matrix. The high F-values observed in the ANOVA table effectively indicate which crime variables dominate these eigenvectors and thus drive the formation of structural clusters. In particular, the finding that x5 (Forgery) yields an F-value of 2165.44 demonstrates that it constitutes the principal structural determinant underlying provincial differentiation in the spectral embedding space.

Overall, the ANOVA results not only validate the statistical robustness of the clustering solution but also reveal the substantive drivers of the observed structural segmentation in provincial crime patterns.

Principal Component Analysis (PCA) Results and Explained Variance

To reduce dimensionality and capture the dominant structural variation in the dataset, Principal Component Analysis (PCA) was applied prior to spectral clustering. The eigenvalues and explained variance ratios of the extracted components are reported in Table 3.

Table 3. Eigenvalues and Explained Variance Ratios of Principal Components

Component	Explained Variance (%)	Cumulative Variance (%)
PC1	85.70%	85.70%
PC2	4.50%	90.20%

The spectral decomposition of the correlation matrix reveals a highly concentrated eigenvalue distribution. The first eigenvalue ($\lambda_1 = 21.43$) overwhelmingly dominates the variance structure, accounting for approximately 85.7–86% of the total variance. This pronounced dominance indicates the presence of a strong common structural factor underlying provincial crime patterns in Türkiye.

The second eigenvalue ($\lambda_2 = 1.12$) exceeds the conventional unity threshold (Kaiser criterion, $\lambda > 1$), confirming the existence of a second statistically meaningful structural dimension. In contrast, the third eigenvalue ($\lambda_3 = 0.67$) falls substantially below unity and exhibits rapid attenuation, indicating that additional components contribute marginal explanatory power. The sharp decline after the second component statistically justifies focusing on the first two structural dimensions.

The eigenvalue structure therefore suggests that provincial crime differentiation in Türkiye is primarily driven by two latent axes:

- **First Structural Dimension (PC1 – Crime Volume Axis):**
The first principal component alone explains approximately 85.7% of total variance, indicating that the overall magnitude or intensity of crime constitutes the dominant structural pattern across provinces. This suggests a high degree of commonality in crime composition, where provinces differ mainly in scale rather than structure.
- **Second Structural Dimension (PC2 – Crime Typology Axis):**
The second component explains approximately 4.5% of total variance and appears to capture qualitative differentiation in crime types. This dimension likely reflects structural variation between economically motivated offenses and public-order or coercive crimes.

From a spectral perspective, the dominance of λ_1 indicates that the Laplacian (or correlation) matrix is strongly rank-concentrated, implying a quasi one-dimensional embedding structure with a secondary orthogonal adjustment. The rapid decay of eigenvalues beyond λ_2 further confirms that higher-order components primarily represent noise rather than substantive structural variation.

In summary, both the spectral decomposition and PCA results converge toward the same conclusion: provincial crime patterns in Türkiye are largely governed by a dominant crime-volume factor, complemented by a secondary dimension reflecting crime-type differentiation. This eigenstructure provides strong statistical justification for restricting the analysis to the first two principal components in subsequent clustering and visualization procedures.

Principal Component Analysis (PCA) Results

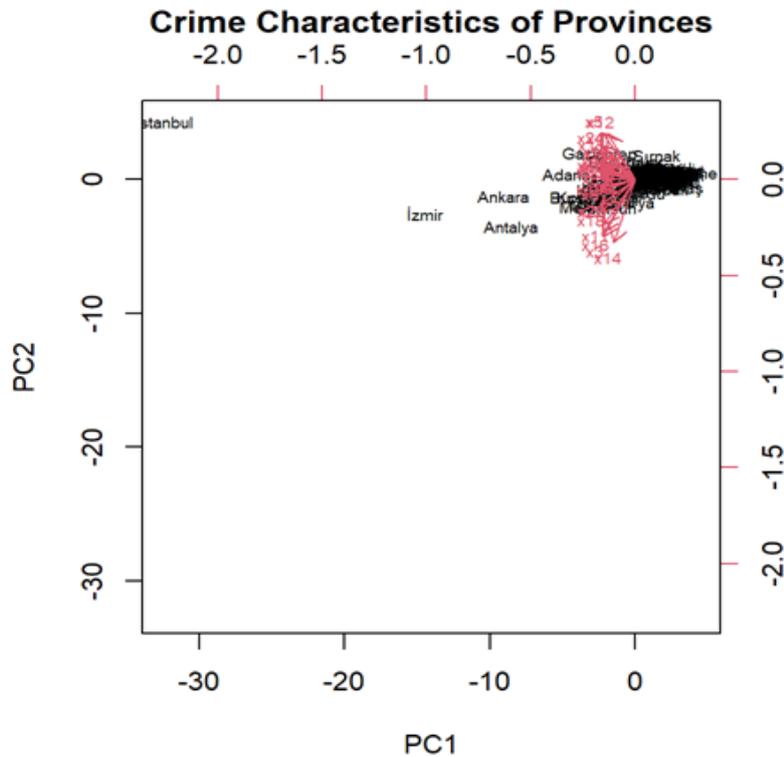


Figure 1. Provincial Positions in the PCA Plane According to Crime Characteristics (Biplot)

Figure 1 presents the PCA biplot illustrating the joint projection of provinces and crime variables onto the first two principal components. In the figure, red vectors (x1–x25) represent the 25 crime types, while blue points correspond to the 81 provinces.

The horizontal axis (PC1) represents the first principal component and explains 85.7% of the total variance. The vertical axis (PC2) represents the second principal component. Together, these two dimensions capture the overwhelming majority of the structural variation in the dataset.

Interpretation of PC1 and PC2

PC1 (General Crime Capacity Axis). The first component clearly dominates the variance structure and functions as a general crime intensity factor. Most provinces cluster in the negative region of PC1, while major metropolitan provinces—particularly İstanbul—are positioned at the extreme positive end of the axis. Ankara, İzmir, and Antalya also display partial separation from the main provincial mass.

The radical separation of İstanbul along PC1 indicates its structural outlier status. This positioning reflects its substantially higher crime volume and systemic scale relative to other provinces. From a structural perspective, İstanbul forms an almost independent branch in the embedding space, visually expressing a pattern of crime-related structural isolation.

PC2 (Crime-Type Differentiation Axis). The second component captures qualitative differentiation in crime composition. While it explains a smaller proportion of total variance ($\approx 4.5\%$), it distinguishes provinces based on crime-type specialization rather than overall magnitude. İzmir's relative displacement along PC2 suggests typological differentiation compared to other metropolitan provinces.

Crime-Type Vectors (Loadings)

The red vectors (x1–x25) are largely aligned in a similar direction, indicating strong positive correlations among crime types. This alignment confirms the presence of a common national crime factor underlying provincial crime distributions.

The near-parallel orientation of most vectors mathematically supports the dominance of PC1 as a general crime factor. In particular, the strong projection of x5 (Forgery) and x7 (Fraud) onto PC1 explains why these variables exhibited the highest F-statistics in the ANOVA results. Their large loadings indicate that they contribute substantially to the dominant eigenstructure driving inter-provincial differentiation.

Provincial Distribution and Structural Implications

The visible separation of major metropolitan areas reflects differences primarily attributable to population scale, economic complexity, and crime intensity. The sharp divergence of İstanbul from the rest of Türkiye provides the first clear geometric evidence of its structural distinctiveness within the national crime topology.

Overall, the PCA biplot demonstrates that:

- Provincial differentiation is largely one-dimensional (crime volume).
- A secondary orthogonal axis captures typological differentiation.
- Crime variables are highly positively correlated and structured around a common latent factor.
- Metropolitan provinces occupy extreme positions within the embedding space, with İstanbul exhibiting the strongest structural deviation.

Spectral Positions

To further clarify provincial positions within the reduced-dimensional space, the PCA-based embedding results are additionally presented in Figure 2, where structural clustering patterns become more visually interpretable.

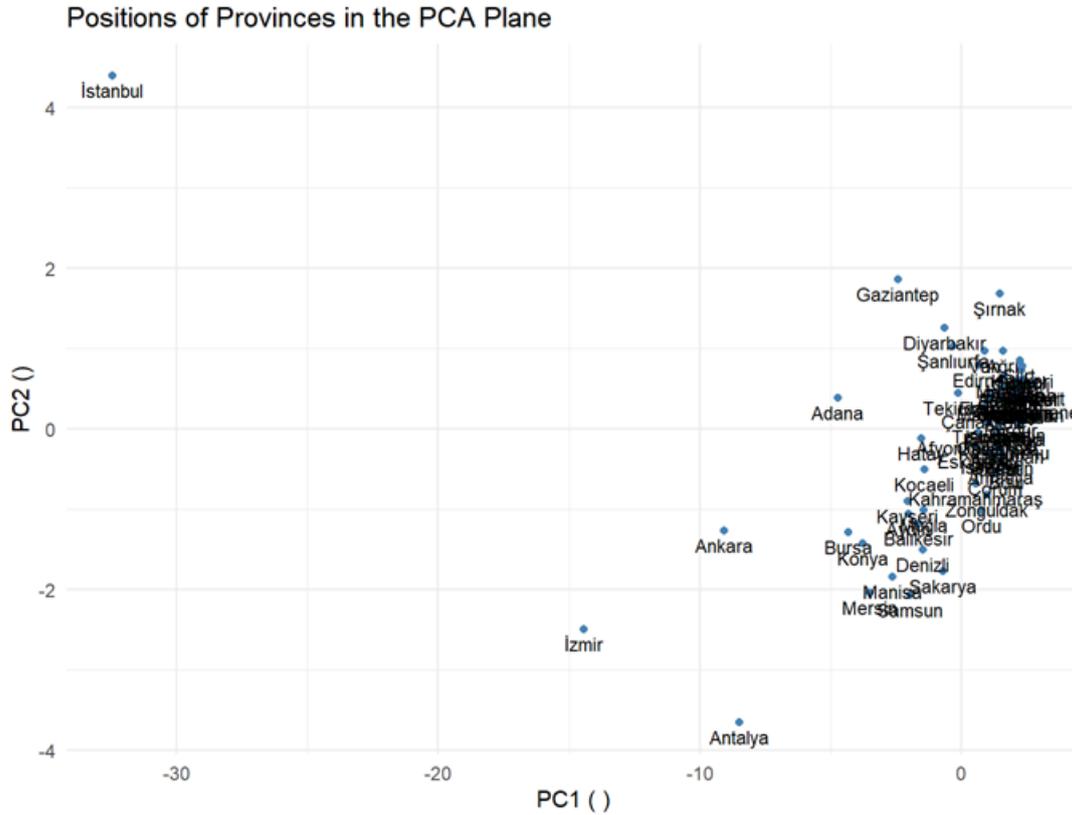


Figure 2. Spectral Positions of Provinces

The Spectral Positions Plot (Figure 2) illustrates the structural similarity of provinces based on their crime composition. The two axes correspond to the first two dimensions obtained from the spectral embedding of the similarity (or Laplacian) matrix.

Dimension 1: Overall Crime Capacity

Dimension 1 represents the dominant structural factor, interpreted as overall crime capacity. The extreme position of İstanbul (approximately -1.00) indicates a structurally isolated configuration relative to the rest of the country. This positioning confirms that İstanbul differs not only in crime volume but also in systemic complexity and diversity. Its distance from the main provincial cluster visually and mathematically reflects a high-degree node in the national crime similarity network.

Dimension 2: Crime-Type Differentiation

Dimension 2 captures species (type) differentiation in crime composition. The near $+1.00$ position of İzmir along this axis indicates that İzmir’s crime structure diverges qualitatively from other provinces. This suggests that the distinction is not purely quantitative (crime volume) but also structural, reflecting variation in the distribution of specific crime categories.

The scatter distribution further demonstrates that major metropolitan provinces including Ankara and Antalya are clearly separated from the dense cluster of remaining provinces along one or both dimensions. This separation reflects differentiated crime profiles driven by demographic scale, economic structure, and urban complexity.

Overall, Figure 2 confirms that:

- Provincial crime similarity is largely structured along a dominant capacity axis.
- A secondary orthogonal axis captures qualitative differentiation.
- Metropolitan provinces occupy extreme spectral coordinates.
- İstanbul constitutes the most structurally distinct province in the embedding space.

Hierarchical Clustering (Ward Method) – Dendrogram

The hierarchical clustering dendrogram constructed using Ward’s minimum variance method is presented in Figure 3. Ward’s method minimizes within-cluster variance at each agglomerative step, producing compact and internally homogeneous clusters.

The dendrogram visually reinforces the spectral findings:

- Provinces with similar crime intensity and composition merge at lower linkage distances.
- Metropolitan provinces branch at higher linkage heights, indicating stronger structural separation.
- İstanbul forms a late-merging branch, further confirming its isolated structural status within the national crime topology.

The convergence of PCA, spectral embedding, ANOVA, and Ward-based hierarchical clustering provides multi-method validation of the clustering structure. Collectively, these results demonstrate that provincial crime differentiation in Türkiye is statistically robust and structurally coherent across independent analytical frameworks.

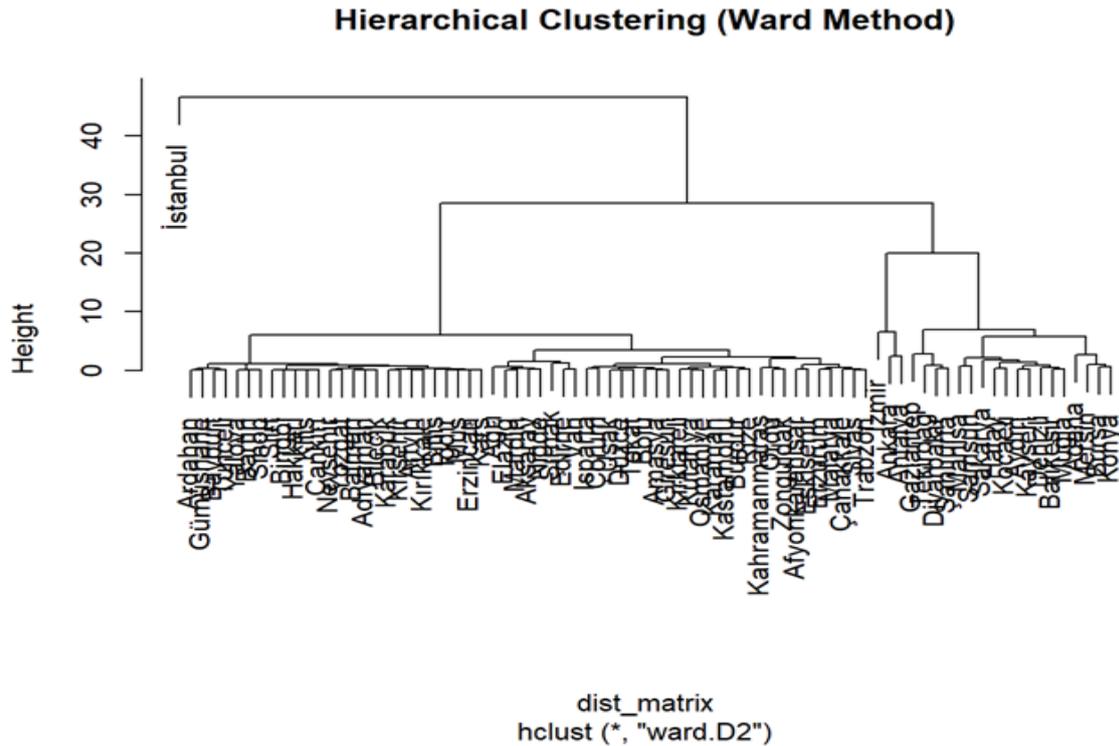


Figure 3. Hierarchical Clustering Dendrogram Prepared Using the Ward Method.

In the tree structure created based on the similarities in the crime profiles of the provinces, it is noteworthy that İstanbul stands out by forming a single branch at the highest hierarchical level. This situation statistically confirms İstanbul's structural uniqueness in the Turkish crime geography.

The dendrogram (Figure 3) visualizes the grouping of provinces based on similarity across all crime characteristics.

Primary Split: The most significant finding is the primary split at the highest height, which separates İstanbul into its own individual cluster. This confirms that İstanbul's crime profile is statistically distinct from every other province in Turkey combined.

Secondary Clusters: The remaining 80 provinces are divided into two large sub-clusters, further differentiating major metropolises like Ankara and İzmir from smaller, more rural provinces.

PCA and Clustering Synthesis

ANOVA was applied to determine the groups using clustering analysis. The clustering results show that the groups differ statistically significantly across all 25 crime types. The p-value is < 0.001 for all variables. The strongest and weakest discriminant variables are given in Table 4 and Table 5, respectively.

Table 4. Strongest discriminants according to F-value magnitude

Variable	F
x5 (Forgery)	2165
x7 (Swindling)	906
x21 (Kidnapping)	553
x24 (Robbery)	551

Clustering and Spatial Differentiation Analysis

The clustering structure demonstrates that crime distribution in Türkiye is not random but follows a hierarchical stratification shaped by socio-economic and demographic layers.

Cluster 4 – İstanbul: A Structurally Isolated Crime Ecosystem

As previously observed in the spectral embedding, İstanbul exhibits complete structural separation from the rest of the country. This divergence is not attributable solely to population size but to the complexity and composition of crime. High-intensity qualified offenses such as forgery (x5) and fraud (x7) drive this separation. İstanbul therefore represents a distinct crime ecosystem characterized by scale, diversity, and economic sophistication.

Cluster 3 – Secondary Metropolitan Layer (Ankara, İzmir, Antalya)

The second structural layer includes Ankara, İzmir, and Antalya. These provinces share crime profiles shaped by administrative centrality, tourism intensity, and economic dynamism. Their intermediate spectral positioning indicates both high crime capacity and partial typological differentiation.

Cluster 1 – Industrial Intermediate Metropolises

Provinces such as Adana, Bursa, and Kocaeli form an industrialized and migration-attracting intermediate metropolitan layer. These provinces exhibit elevated crime intensity but lack the structural complexity observed in the top-tier metropolitan group.

Cluster 2 – The Anatolian Block

The majority of provinces fall into Cluster 2, forming a dense Anatolian block. The homogeneity of this group suggests that crime structures in these provinces are largely characterized by traditional public-order offenses. The compactness of this cluster in spectral space indicates limited structural differentiation.

Linking Spectral Analysis and ANOVA Findings

Spectral decomposition identifies eigenvectors that capture the largest variance components of the Laplacian matrix. The high F-statistics observed in the ANOVA table indicate which variables dominate these eigenvectors and thus drive cluster separation.

In particular, the ANOVA results reveal that x5 (Forgery) exhibits an F-value of 2165.44, making it the strongest determinant of structural differentiation among provinces (see Figure 5). This implies that the first dominant eigenvector—corresponding to the principal spectral axis—is heavily influenced by economically motivated qualified crimes.

In spectral terms:

- High F-values → Large between-cluster variance
- Large between-cluster variance → Strong projection onto dominant eigenvectors
- Strong eigenvector dominance → Clear spectral separation

Thus, the convergence of spectral clustering and ANOVA results provides dual validation:

1. The clustering solution is statistically robust.
2. The substantive driver of differentiation is economically complex crime.

Overall, the combined evidence from PCA, spectral embedding, ANOVA, and hierarchical clustering demonstrates that Türkiye's provincial crime structure follows a stratified metropolitan hierarchy, with İstanbul constituting a structurally isolated apex node.

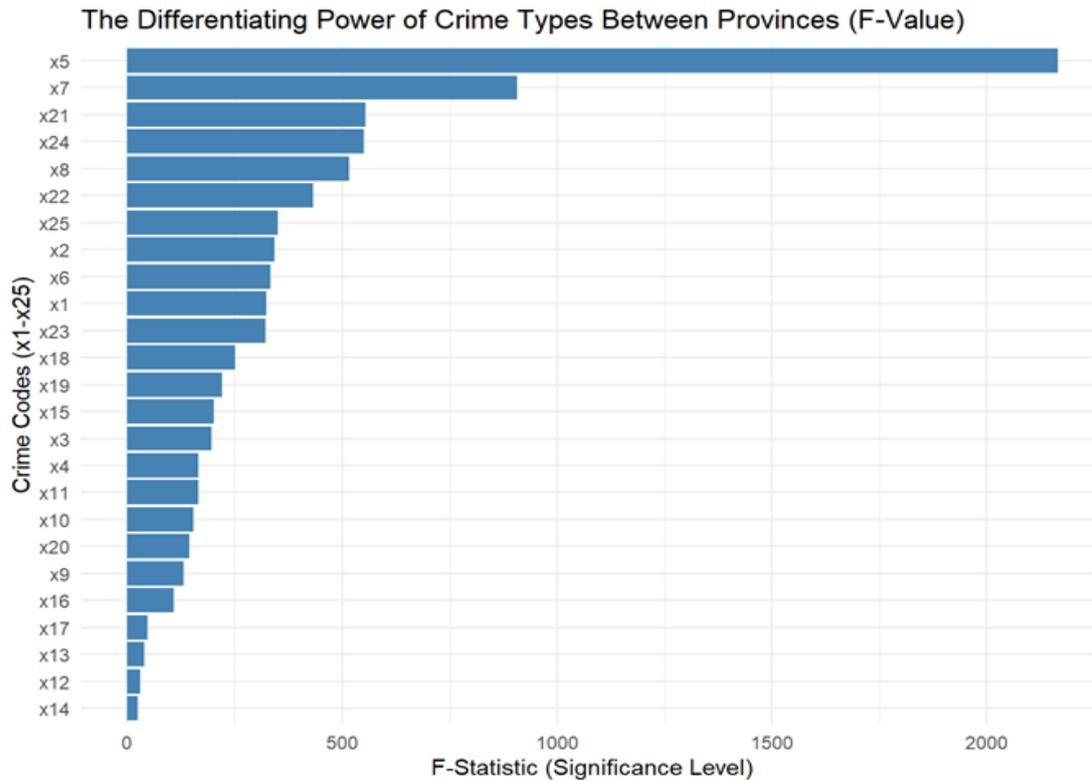


Figure 5. ANOVA-Based Structural Drivers of Cluster Differentiation

As illustrated in Figure 5, the ANOVA results reveal striking levels of inter-cluster differentiation. In particular, x5 (Forgery) attains an exceptionally high F-value ($F = 2165.44$), mathematically demonstrating that provincial crime structures in Türkiye are sharply polarized along the axis of qualified economic offenses.

From a structural analysis perspective, these findings can be summarized as follows:

Dominant Structural Components

The crime types at the top of the ANOVA ranking x5 (Forgery), x7 (Fraud), x21 (Deprivation of Liberty), and x24 (Robbery) constitute the principal variables determining the orientation of the dominant eigenvectors obtained through spectral decomposition. In other words, these variables heavily influence the direction of maximum variance in the Laplacian embedding space.

Economic Fracture

The prominence of x5 and x7 indicates a clear structural divide between advanced/metropolitan provinces and the rest of the country. The large between-cluster variance in these variables represents a widening “qualified crime gap,” reflecting economic sophistication, financial activity, and institutional complexity.

Thus, the primary spectral axis is strongly aligned with economically motivated offenses. This explains why provinces such as İstanbul occupy extreme positions in the embedding space.

Public-Order Structural Pattern

High F-values for x21 (Deprivation of Liberty) and x24 (Robbery) demonstrate that these crimes are not randomly distributed but exhibit distinct regional clustering patterns. Their contribution to eigenvector orientation indicates that structural differentiation is not purely economic but also shaped by coercive and public-order crime typologies.

Spectral Biplot: Provinces and Crimes in the Same Embedding Space

The ANOVA analysis identifies which variables are statistically most discriminative. To visually integrate these findings with the spectral framework, crime types can be represented as vectors and provinces as points within the same embedding plane (spectral biplot).

Such a representation elevates the methodological coherence of the study by:

- Linking statistical significance (ANOVA F-values)
- To geometric orientation (eigenvector loadings)
- Within the spectral embedding space

In this joint representation:

- Provinces positioned along the direction of x5 and x7 vectors exhibit high concentrations of qualified economic crimes.

PC1: General Crime Capacity Axis

The first principal/spectral component (PC1) represents overall crime capacity. Provinces with extreme negative scores along this axis occupy the highest positions in the national crime intensity hierarchy.

For example, Ankara and Adana exhibit pronounced negative PC1 scores, indicating that they lie at one of the structural extremes in terms of aggregate crime volume and diversity. Their position reflects elevated systemic crime capacity relative to the national distribution.

In contrast, provinces such as Amasya and Adiyaman display positive PC1 scores, suggesting comparatively lower overall crime intensity. These provinces cluster within the more homogeneous and structurally “quieter” segment of the national crime topology.

PC2: Crime-Type Differentiation Axis

The second component (PC2) captures qualitative differentiation in crime composition rather than sheer magnitude.

The negative PC2 score observed for Ankara indicates that the province diverges not only in crime volume but also in the relative concentration of specific crime categories—potentially public-order or property-related offenses. This suggests that Ankara’s structural profile reflects both quantitative intensity and typological specialization.

Provinces with positive PC2 scores, by contrast, demonstrate alternative crime-type configurations, reinforcing the interpretation that the second axis captures structural heterogeneity in crime composition.

Structural Interpretation of the Heatmap

From a structural perspective, the heatmap reveals three fundamental insights:

1. Crime differentiation in Türkiye is strongly concentrated along a dominant capacity axis (PC1).
2. A secondary orthogonal axis (PC2) captures typological specialization.
3. Metropolitan provinces exhibit systematic deviation from the national structural baseline, confirming the center–periphery and metropolitan–provincial stratification detected in previous analyses.

The visualization therefore integrates spectral decomposition, PCA results, and ANOVA findings into a single comparative structural framework. Provinces that display darker intensities across both dimensions represent structurally distinctive crime ecosystems, while lighter tones correspond to structurally homogeneous regions.

Overall, Figure 6 confirms that Türkiye’s provincial crime geography is hierarchically organized, with metropolitan centers occupying structurally extreme positions and peripheral provinces forming a compact and relatively uniform block.

Spectral Positions (Score Plot)

The Spectral Positions (Score Plot) graph is given in Figure 7.

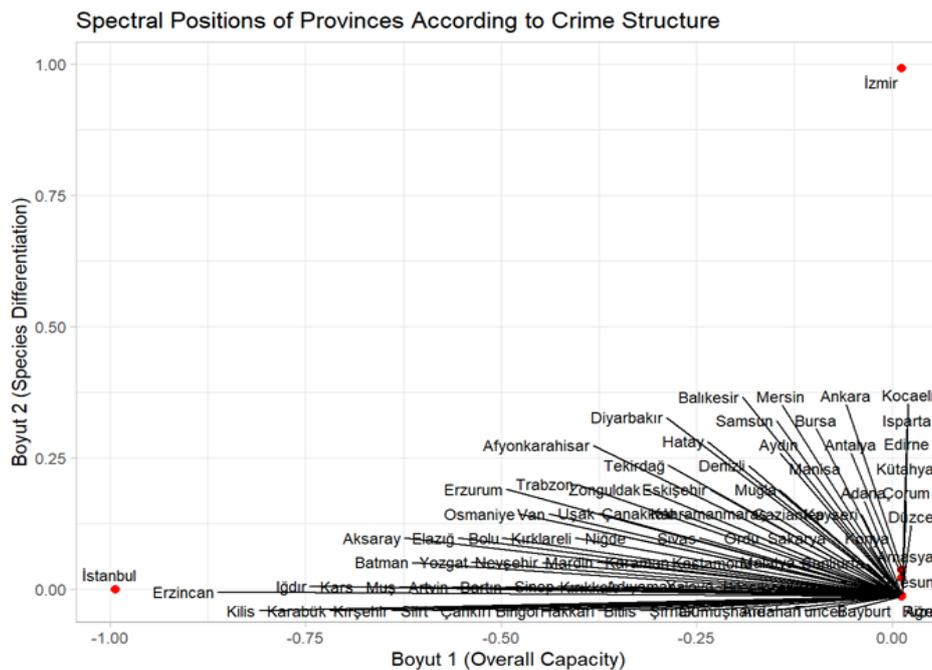


Figure 7. Spectral Positions (Score Plot) of Provincial Crime Structures

The spectral score plot presented in Figure 7 provides a concise mathematical anatomy of Türkiye's provincial crime structure. The two-dimensional embedding clearly reveals structural polarization among provinces. The sharp separation of İstanbul and İzmir from the rest of the country constitutes the most powerful structural finding of the study.

İstanbul: Extreme Position on the General Capacity Axis

Located in the lower-left corner of the spectral plane, İstanbul occupies the extreme end of Dimension 1 (Overall Crime Capacity). This positioning confirms that İstanbul diverges not merely in crime volume but in structural complexity.

The dominance of x5 (Forgery) and x7 (Fraud) which achieved the highest F-statistics in the ANOVA results explains this separation. Their strong loadings on the first eigenvector indicate that İstanbul's crime ecosystem is characterized by economically sophisticated and qualified offenses. Thus, its spectral isolation reflects structural differentiation rather than simple quantitative excess.

İzmir: Extreme Position on the Typological Differentiation Axis

İzmir appears in the upper-right corner of the embedding space, occupying the extreme position along Dimension 2 (Crime-Type Differentiation). This suggests that İzmir possesses a distinctive crime composition profile.

Unlike İstanbul's scale-driven divergence, İzmir's separation appears to stem from qualitative specialization in certain crime categories potentially public-order or drug-related structures indicating typological deviation rather than purely volumetric intensity.

Metropolitan Cluster and Peripheral Provinces

Intermediate metropolitan provinces such as Ankara, Mersin, and Antalya form a clustered group in the mid-right region of the plane. Their proximity suggests shared structural characteristics linked to urbanization, migration dynamics, and economic complexity.

Conversely, provinces such as Gümüşhane, Bayburt, and Ardahan cluster near the lower-right region, close to the origin. Their position reflects low-intensity and structurally homogeneous crime configurations.

Integrating F-Statistics and Loadings

When the F-statistic ranking is interpreted jointly with eigenvector loadings, a coherent structural conclusion emerges:

Primary Structural Determinant

Forgery (x5) functions as the fundamental structural column shaping Türkiye's crime geography. Its exceptionally high F-value and strong projection onto the first eigenvector demonstrate that it drives the primary axis of inter-provincial differentiation.

Structural Cohesion and Socio-Economic Patterning

The elevated F-values of x21 (Deprivation of Liberty) and x24 (Robbery) indicate that these crimes are not randomly distributed but reflect socio-economic stratification potentially associated with metropolitan stress, migration pressure, and uneven urban development.

Structural Synthesis

The spectral decomposition-based structural analysis demonstrates that crime in Türkiye is not merely a quantitative phenomenon but a matter of structural differentiation among provinces.

The isolated spectral position of İstanbul indicates that its crime ecosystem deviates fundamentally from the national baseline. Supported by ANOVA findings, the results confirm that Forgery (x5) and Fraud (x7) are the principal variables defining the structural gap between provinces.

Taken together, the convergence of:

- Spectral embedding,
- PCA variance concentration,
- ANOVA extreme F-values,
- Cluster formation patterns,

reveals a hierarchically stratified crime topology in Türkiye, characterized by metropolitan structural complexity and peripheral homogeneity.

Structural Analysis of Cluster Characteristics

The graph of provinces and crime types in Turkey using PCA is shown in Figure 8.

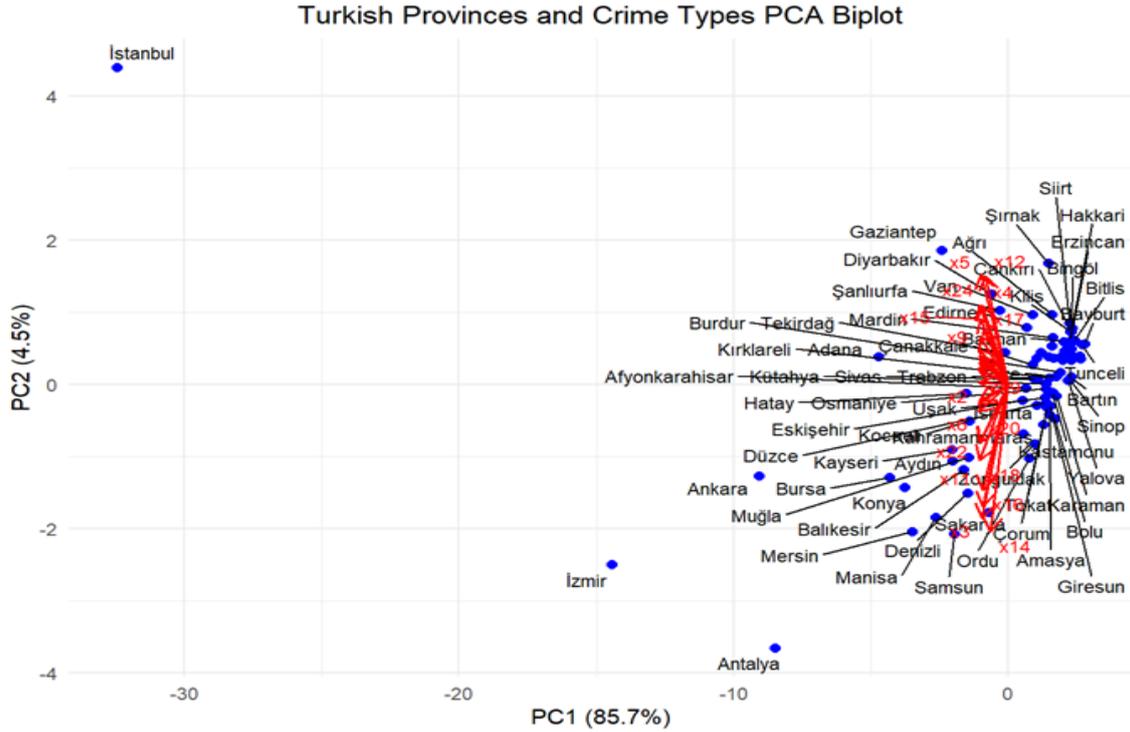


Figure 8. PCA Biplot with Spectral Decomposition.

PCA identifies the primary directions of variance within the dataset, as shown in the PCA Biplots (Figure 8):

Dominant Variance: The first principal component (PC1) explains a substantial 85.7% of the total variance. This confirms that the crime landscape is primarily driven by a single factor, likely related to overall volume and structural complexity.

Province Distribution: Consistent with spectral findings, İstanbul is located at the far left of the PC1 axis. İzmir and Antalya are distinguished further by their negative positions on PC2, which accounts for 4.5% of the variance.

Variable Influence: Red vectors (labeled x1 through x25) represent specific crime types. Vectors pointing toward the left (toward İstanbul) highlight crimes most characteristic of that province's extreme profile.

The spatial and structural distribution of the clusters was also examined, and the results are presented in Table 6.

Table 6. Spectral Clustering Classification of Provinces

Province	Cluster	Province	Cluster	Province	Cluster
Adana	1	Erzurum	2	Muş	2
Adıyaman	2	Eskişehir	2	Nevşehir	2
Afyonkarahisar	2	Gaziantep	2	Niğde	2
Aksaray	2	Giresun	2	Ordu	2
Amasya	2	Gümüşhane	2	Osmaniye	2
Ankara	3	Hakkari	2	Rize	2
Antalya	2	Hatay	1	Sakarya	1
Ardahan	2	Isparta	2	Samsun	1
Artvin	2	Iğdır	2	Siirt	2
Aydın	1	Kahramanmaraş	2	Sinop	2
Ağrı	2	Karabük	2	Sivas	2
Balıkesir	1	Karaman	2	Tekirdağ	2
Bartın	2	Kars	2	Tokat	2
Batman	2	Kastamonu	2	Trabzon	2

Bayburt	2		Kayseri	1		Tunceli	2
Bilecik	2		Kilis	2		Uşak	2
Bingöl	2		Kocaeli	1		Van	2
Bitlis	2		Konya	1		Yalova	2
Bolu	2		Kütahya	2		Yozgat	2
Burdur	2		Kırklareli	2		Zonguldak	2
Bursa	1		Kırıkkale	2		Çanakkale	2
Denizli	1		Kırşehir	2		Çankırı	2
Diyarbakır	2		Malatya	2		Çorum	2
Düzce	2		Manisa	1		İstanbul	4
Edirne	2		Mardin	2		İzmir	3
Elazığ	2		Mersin	1		Şanlıurfa	2
Erzincan	2		Muğla	1		Şırnak	2

The spectral clustering algorithm partitions Türkiye's 81 provinces into four principal structural classes (Table 6), revealing a clear hierarchical organization of crime profiles.

Cluster 4 – İstanbul: A Statistically Isolated Outlier

Cluster 4 consists exclusively of İstanbul. Its singleton status confirms that İstanbul's crime structure is statistically distinct from the rest of the country. This separation is not merely volumetric but structural, reflecting complexity, diversification, and high-intensity qualified offenses.

The spectral isolation of İstanbul, combined with the extreme ANOVA F-values observed for forgery (x5) and fraud (x7), provides strong evidence that the province occupies the apex of the national crime hierarchy. In embedding space, İstanbul forms an independent structural pole.

Cluster 3 – Metropolitan Block

Cluster 3 includes Ankara, Antalya, and İzmir. These provinces represent the second-highest intensity layer in terms of crime volume and typological diversity.

Their grouping reflects shared characteristics associated with administrative centrality, tourism intensity, and economic dynamism. Although structurally close to İstanbul, they do not reach the same level of eigenvector dominance or crime complexity.

Cluster 1 – Developed Industrial and Coastal Provinces

Cluster 1 comprises provinces such as Adana, Bursa, Denizli, Kocaeli, Konya, Mersin, Muğla, Sakarya, and Samsun.

This group represents socio-economically active provinces characterized by industrialization, migration inflows, and commercial mobility. However, their crime structures have not fully evolved into the highly diversified metropolitan profile observed in Cluster 3 or the structurally unique ecosystem of İstanbul.

Cluster 2 – The Anatolian Block

Cluster 2 contains the majority of provinces, spanning from Adıyaman to Zonguldak. Approximately 75% of provinces fall within this group.

The concentration of such a large proportion of provinces within a single cluster indicates a structurally homogeneous baseline across much of the country. Crime dynamics in this block are primarily characterized by traditional public-order patterns, with limited structural differentiation.

Hierarchical Crime Structure in Türkiye

The spectral clustering results demonstrate that provincial crime profiles in Türkiye exhibit a clear hierarchical structure:

1. A structurally isolated metropolitan apex (İstanbul),
2. A secondary metropolitan layer (Ankara–İzmir–Antalya),
3. An intermediate industrialized tier,
4. A broad and homogeneous Anatolian baseline.

This stratification suggests that the spatial distribution of crime is strongly correlated with urban development, economic complexity, and population density. However, İstanbul occupies a singular position at the top of this hierarchy, forming a structurally autonomous crime ecosystem.

The finding that approximately three-quarters of provinces belong to Cluster 2 further indicates that general public-order dynamics across Türkiye are built upon a relatively common structural foundation. Differentiation

intensifies primarily within large metropolitan environments, where economic sophistication and demographic density reshape crime typologies.

V. Discussion

The empirical findings of this study provide strong support for classical and contemporary criminological theories emphasizing the spatial stratification of crime. The overwhelming dominance of the first principal component (explaining 85.7% of total variance) indicates that crime distribution across provinces is not random but structurally centralized. This pattern is consistent with the spatial concentration thesis proposed by Clifford Shaw and Henry D. McKay (1942), who argued that crime tends to cluster in socially and economically differentiated urban zones.

However, the findings extend beyond classical social disorganization theory. While Shaw and McKay emphasized neighborhood-level instability and poverty, the present results suggest that provincial differentiation in Türkiye is primarily driven by metropolitan scale effects rather than solely by social disorganization. In this respect, the results align more closely with urban scaling theory, which demonstrates that socio-economic indicators including complex crimes grow superlinearly with city size (Luis Bettencourt et al., 2007). Istanbul's statistical isolation supports this argument, as the city appears to function as a distinct crime ecosystem shaped by economic density, financial networks, and advanced urban interactions.

The prominence of forgery and fraud as the most discriminating variables further reinforces this interpretation. Unlike conventional violent crimes, these offenses are closely linked to economic complexity, financial transactions, and digital infrastructure. This observation resonates with environmental criminology and situational crime prevention theory (Ronald V. Clarke, 1980), which emphasize opportunity structures and target availability. Larger metropolitan areas generate more opportunities for financially motivated and technologically facilitated crimes, thereby reshaping the qualitative composition of crime.

Moreover, the hierarchical clustering results reveal a layered provincial structure rather than a simple metropolitan–rural dichotomy. Intermediate metropolitan provinces such as Ankara, İzmir, and Antalya form a secondary structural block, suggesting a gradational crime hierarchy. This layered structure may reflect differentiated levels of economic diversification, administrative centrality, and tourism-driven activity. Thus, crime stratification in Türkiye appears to follow a multi-tier urban hierarchy rather than a binary center–periphery model.

Importantly, the finding that traditional violent crimes (e.g., homicide) do not primarily drive inter-provincial differentiation challenges common public perceptions about crime severity. Instead, structural divergence is largely shaped by economically embedded offenses. This supports the argument that modernization processes transform not only the quantity but also the quality of crime, shifting from physical violence toward financially and digitally mediated forms.

In theoretical terms, the study contributes to the literature by empirically integrating social disorganization theory, environmental criminology, and urban scaling perspectives within a unified multivariate framework. Methodologically, it demonstrates that spectral and hierarchical clustering techniques can uncover latent structural typologies that remain invisible in aggregate crime statistics.

Overall, the findings suggest that crime in Türkiye should be understood as a structurally stratified, scale-sensitive phenomenon shaped by urban complexity and economic transformation rather than as a uniformly distributed public-order issue.

VI. Conclusion

The findings of this study provide strong empirical evidence for the existence of a pronounced hierarchical structure within Türkiye's crime geography. By integrating Principal Component Analysis (PCA), Spectral Clustering, Hierarchical Clustering, and ANOVA, the study reveals that regional crime patterns are neither homogeneous nor randomly distributed but instead display systematic structural stratification.

Structural Polarization

The fact that 85.7% of the total variance in crime data is explained by the first principal component (PC1) indicates a high degree of centralization in the spatial distribution of crime. This dominant axis can be interpreted as a structural “crime volume dimension,” suggesting that provincial differentiation is primarily driven by scale effects associated with metropolitan concentration. Such concentration patterns are consistent with urban scaling perspectives (e.g., Luis Bettencourt et al., 2007), which argue that socio-economic complexity increases disproportionately with city size.

The Singularity of Istanbul

Across all clustering techniques employed, Istanbul emerges as a statistically isolated province, structurally detached from the remaining 80 provinces. In hierarchical dendrogram analysis, Istanbul forms a single branch at the highest level of separation, indicating a unique crime configuration. This finding suggests

the existence of a distinct metropolitan crime ecosystem characterized by structural complexity rather than mere quantitative intensity.

Determinants of Structural Differentiation

ANOVA results demonstrate that the primary discriminating factors between provincial clusters are not traditional violent crimes such as homicide but rather modern and economically driven offenses. In particular, forgery ($F = 2165$) and fraud ($F = 905$) exhibit exceptionally high discriminatory power. This indicates that structural differentiation across provinces is shaped more by economic and financial crime dynamics than by conventional public-order offenses.

These findings imply that crime transformation in Türkiye is not solely a matter of increasing magnitude but reflects qualitative changes associated with urbanization, economic diversification, and digitalization processes.

Policy Implications

The results suggest that uniform, nationwide security policies are unlikely to be effective. Instead, differentiated strategies should be developed according to structural crime typologies:

- **Istanbul:** Policies focused on financial crimes, cybercrime, and complex economic offenses.
- **Intermediate metropolitan provinces (e.g., Ankara, İzmir, Antalya):** Strategies targeting property crimes and diversified urban offenses.
- **Anatolian provinces:** Policies emphasizing general public-order and conventional crime prevention.

Such a differentiated approach would enable evidence-based security planning aligned with region-specific structural characteristics.

By reducing high-dimensional crime data into interpretable structural components, this study contributes to the literature by offering a mathematically grounded typology of regional crime in Türkiye. The integrated methodological framework demonstrates that spatial crime analysis benefits significantly from multidimensional statistical modeling rather than relying solely on aggregate crime counts.

References

- [1]. Bettencourt, L.M.A., Lobo, J., Helbing, D., Kühnert, C., West, G.B. 2007. Growth, innovation, scaling, and the pace of life in cities. *Proceedings of the National Academy of Sciences (PNAS)*, 104(17), 7301–7306. <https://doi.org/10.1073/pnas.0610172104>
- [2]. Chung, F. R. K. 1997. *Spectral Graph Theory*. American Mathematical Society.
- [3]. Clarke, R. V. 1980. *Situational crime prevention: Theory and practice*. *British Journal of Criminology*, 20(2), 136–147.
- [4]. Fisher, R. A. 1925. *Statistical Methods for Research Workers*. Oliver & Boyd.
- [5]. Hotelling, H. 1933. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24(6), 417–441.
- [6]. Jolliffe, I.T., Cadima, J. 2016. Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A*, 374(2065), 20150202.
- [8]. Lloyd, S. P. 1982. Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28(2), 129–137.
- [9]. MacQueen, J. 1967. Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* (Vol. 1, pp. 281–297).
- [11]. Ng, A. Y., Jordan, M. I., Weiss, Y. 2002. On spectral clustering: Analysis and an algorithm. *Advances in Neural Information Processing Systems*, 14, 849–856.
- [12]. Pearson, K. 1901. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, 2(11), 559–572.
- [13]. Shaw, C. R., McKay, H. D. 1942. *Juvenile delinquency and urban areas*. University of Chicago Press.
- [14]. Turkish Statistical Institute (TSI). 2020. Convicts received into prison by type of crime and province where the crime was committed at SR level 3, 2011–2020. *Prison Statistics 2011–2020*. <https://veriportali.tuik.gov.tr/tr/statistical-themes>
- [15]. Ward, J. H. 1963. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58(301), 236–244.