Multivariate Statistical Techniques to Analyze Crime and its Relationship with Unemployment and Poverty: A Case Study

Anthony Crespo, Juan Brito, Santiago Ajala, Isidro R. Amaro, and Zenaida Castillo

Imbabura, Ecuador

April 2023

1. Problem statement

Problem statement

Ecuador is exposed to crime every day.

In 2021:
- Ecuador experimented one of the biggest prison crisis.
- The number of homicides doubled compared to 2020.

In addition to crime, unemployment, and poverty also affect Ecuadorians.

SCHOOL OF MATHEMATICAL AND COMPUTATIONAL SCIENCES

UNIVERSIDAD YACHAY TECH

Problem statement

In December 2021, 27.7% of Ecuadorians were immersed in poverty, living with less than USD 2.85 per day.

In April 2022, the unemployment rate reached 4.7%.

This work focuses on using HJ-Biplot and Clustering to:
- Analyze crime in Ecuador in 2021 and 2022.
- Determine a possible correlation between crime, unemployment, and poverty in Ecuador in 2019 and 2021.

SCHOOL OF MATHEMATICAL AND COMPUTATIONAL SCIENCES

UNIVERSIDAD YACHAY TECH

2. Materials and Methods

Geographically, Ecuador is divided into
4 natural regions:

- Coast
- Interandean region
- Amazon
- The Galapagos Islands

Data Description: Territorial organization of Ecuador

The political-administrative division of Ecuador is made up of 24 provinces:

- 7 in the Coast
- 10 in the Interandean region
- 6 in the Amazon
- 1 in the Galapagos Islands

SCHOOL OF MATHEMATICAL AND COMPUTATIONAL SCIENCES

UNIVERSIDAD YACHAY TECH

Crime Data

■ Cover the period from January 2021 to May 2022.

■ Data classified by month and by province.

■ Nine variables were taken:

1. Homicides
2. Personal robberies
3. Auto parts theft
4. Motorcycle theft
5. Car theft

6. House theft
7. Business robbery
8. Rapes
9. Number of inhabitants per province.

SCHOOL OF
MATHEMATICAL AND
COMPUTATIONAL SCIENCES

UNIVERSIDAD
YACHAY
TECH

Unemployment and Poverty Data

■ Cover the years 2019, and 2021. Data from 2020 is not available.

■ Data classified by month and by province.

■ Nine variables were taken:

1. Unemployment
2. Adequate employment
3. Unpaid employment
4. NEET people
5. Poor by income

6. Extreme poverty by income
7. Basic needs poverty
8. Multidimensional poverty
9. Extreme multidimensional poverty

Data on crime, unemployment, and poverty were provided by the National Institute of Statistics and Census of Ecuador (INEC).

Codes for provinces

Table 1: Codes for provinces

Province	Code	Province	Code	Province	Code
Azuay	A	Guayas	G	Pichincha	P
Bolívar	B	Imbabura	I	Tungurahua	T
Cañar	U	Loja	L	Zamora Chinchipe	Z
Carchi	C	Los Ríos	R	Galapagos	W
Cotopaxi	X	Manabí	M	Sucumbíos	K
Chimborazo	H	Morona Santiago	V	Orellana	Q
El Oro	O	Napo	N	Santo Domingo	J
Esmeraldas	E	Pastaza	S	Santa Elena	Y

Codes for crime, unemployment and poverty variables

Table 2: Codes for crime variables

Variable	Code	Variable	Code
Homicides	$X1$	House theft	$X6$
Personal robberies	$X2$	Business robbery	$X7$
Auto parts theft	$X3$	Rapes	$X8$
Motorcycle theft	$X4$	Number of inhabitants per province	$X9$
Car theft	$X5$		

To refer to the crime rates per 100,000 inhabitants, we added 100K as a suffix to its corresponding code, e.g. X1_100K is the homicide rate per 100 thousand inhabitants.

Table 3: Codes for unemployment and poverty

Variable	Code	Variable	Code
Unemployment	$Y1$	Extreme poverty by income	$Y6$
Adequate employment	$Y2$	Basic needs poverty	$Y7$
Unpaid employment	$Y3$	Multidimensional poverty	$Y8$
NEET people	$Y4$	Extreme multidimensional poverty	$Y9$
Poor by income	$Y5$		

SCHOOL OF MATHEMATICAL AND COMPUTATIONAL SCIENCES

UNIVERSIDAD YACHAY TECH

Multivariate Statistical Techniques

Two multivariate statistical techniques were used:

- HJ-Biplot

- Clustering

SCHOOL OF
MATHEMATICAL AND
COMPUTATIONAL SCIENCES

UNIVERSIDAD
YACHAY
TECH

HJ-Biplot

- Allows representing the data in a low-dimensional space.
- Individuals are represented as points, and variables as vectors.
- HJ-Biplot seeks to obtain maximum quality of representation in individuals and variables.

Interpretation

The orthogonal projection onto a vector indicates the order of individuals in that variable. For instance, $a_{i_3} > a_{i_1} > a_{i_2}$ with respect to b_{j_2}.

The variance of a variable is approximated by the length of the vector. To illustrate, b_{j_1} has the largest variance.

The distance between points indicates similarity. For example, a_{i_1} is more similar to a_{i_2} than to a_{i_3}.

The correlation between the variables is approximated by the cosine of the angle between the vectors. For example, b_{j_3} is highly correlated with b_{j_1} but uncorrelated with b_{j_2}.

Source: [11]

SCHOOL OF
MATHEMATICAL AND
COMPUTATIONAL SCIENCES

UNIVERSIDAD
YACHAY
TECH

Clustering

- Clustering takes the data and generates categories or clusters.

- K-means partitions a data set into **k** disjoint groups.

- The value of **k** can be obtained using the elbow method.

SCHOOL OF
MATHEMATICAL AND
COMPUTATIONAL SCIENCES

UNIVERSIDAD
YACHAY
TECH
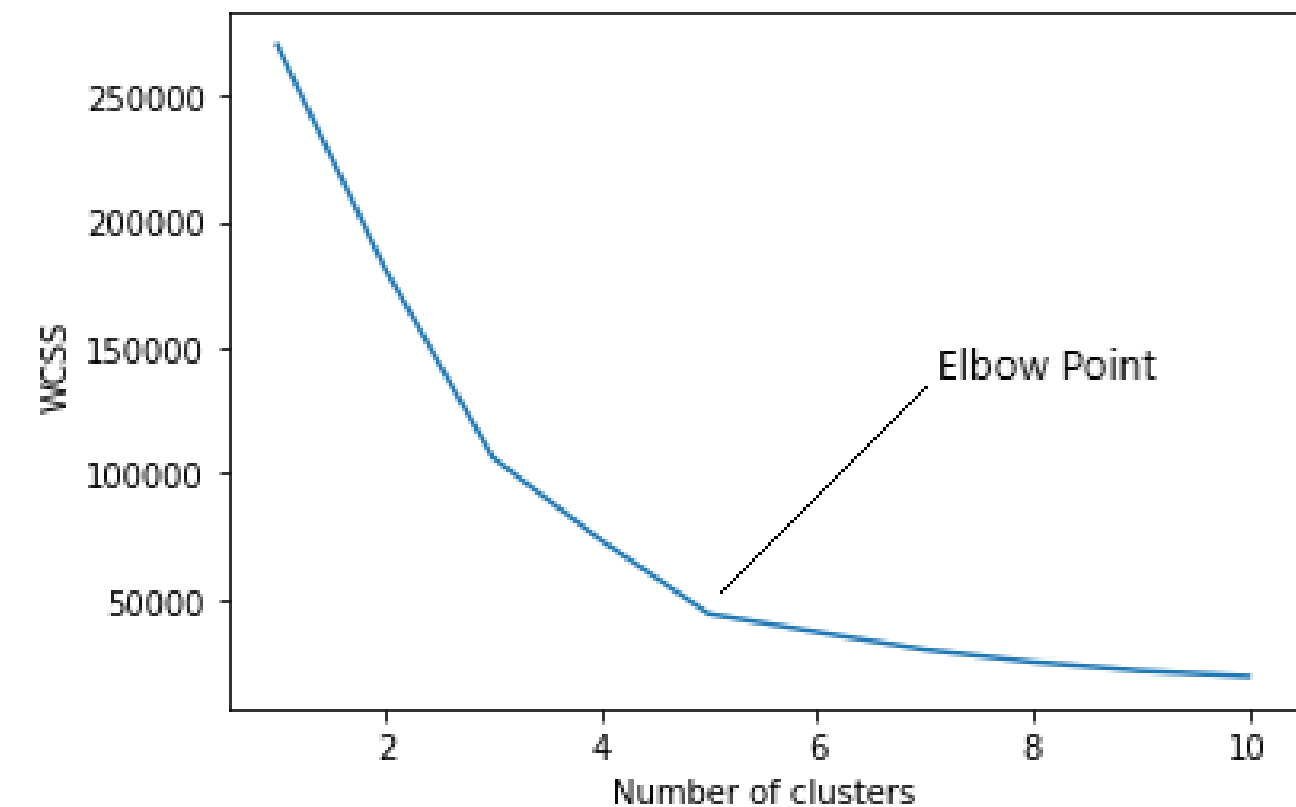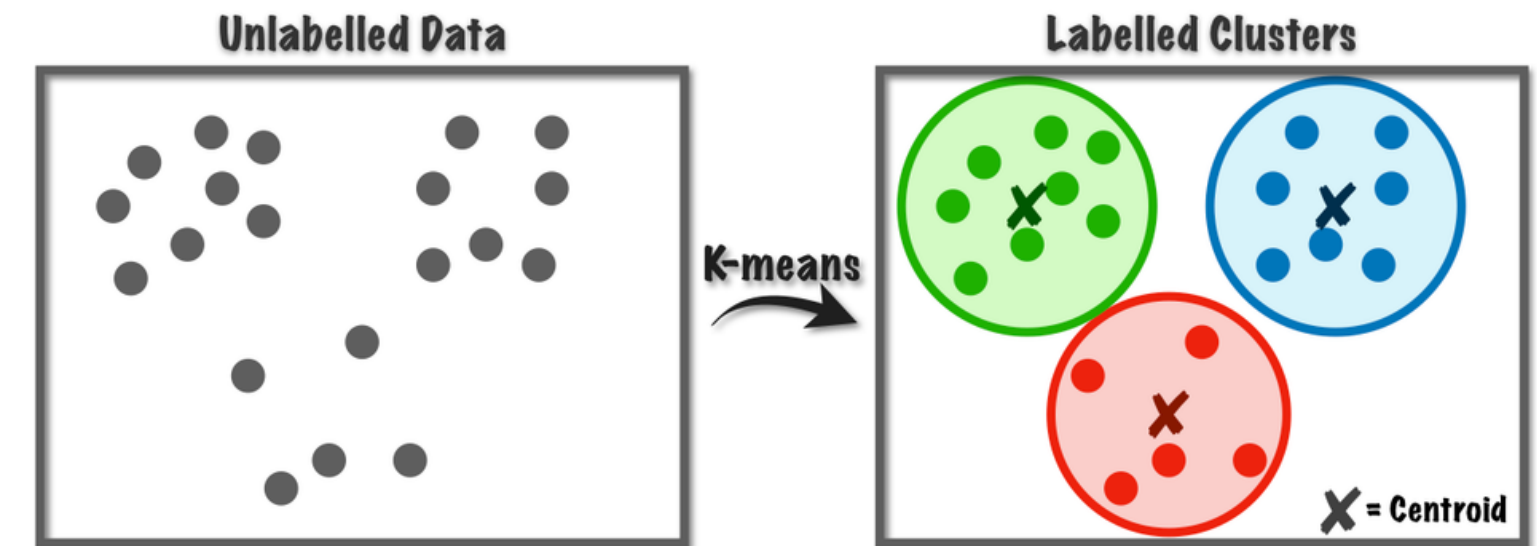
Methodology

Two analyses were performed using clustering and HJ-Biplot, over normalized data.

First Analysis

- Includes three subanalyses:
  - Crime per month
  - Crime per province
  - Crime rate per 100,000 inhabitants.

Second Analysis

- Combines data about crime, poverty, and unemployment to search for a possible correlation.

3. Results

Analysis of Crime from January 2021 – May 2022

- Crime per Month

- Crime per Province

- Crime Rate per 100,000 inhabitants

SCHOOL OF MATHEMATICAL AND COMPUTATIONAL SCIENCES

UNIVERSIDAD YACHAY TECH

Fig. 1: HJ-Biplot for months and number of crimes.
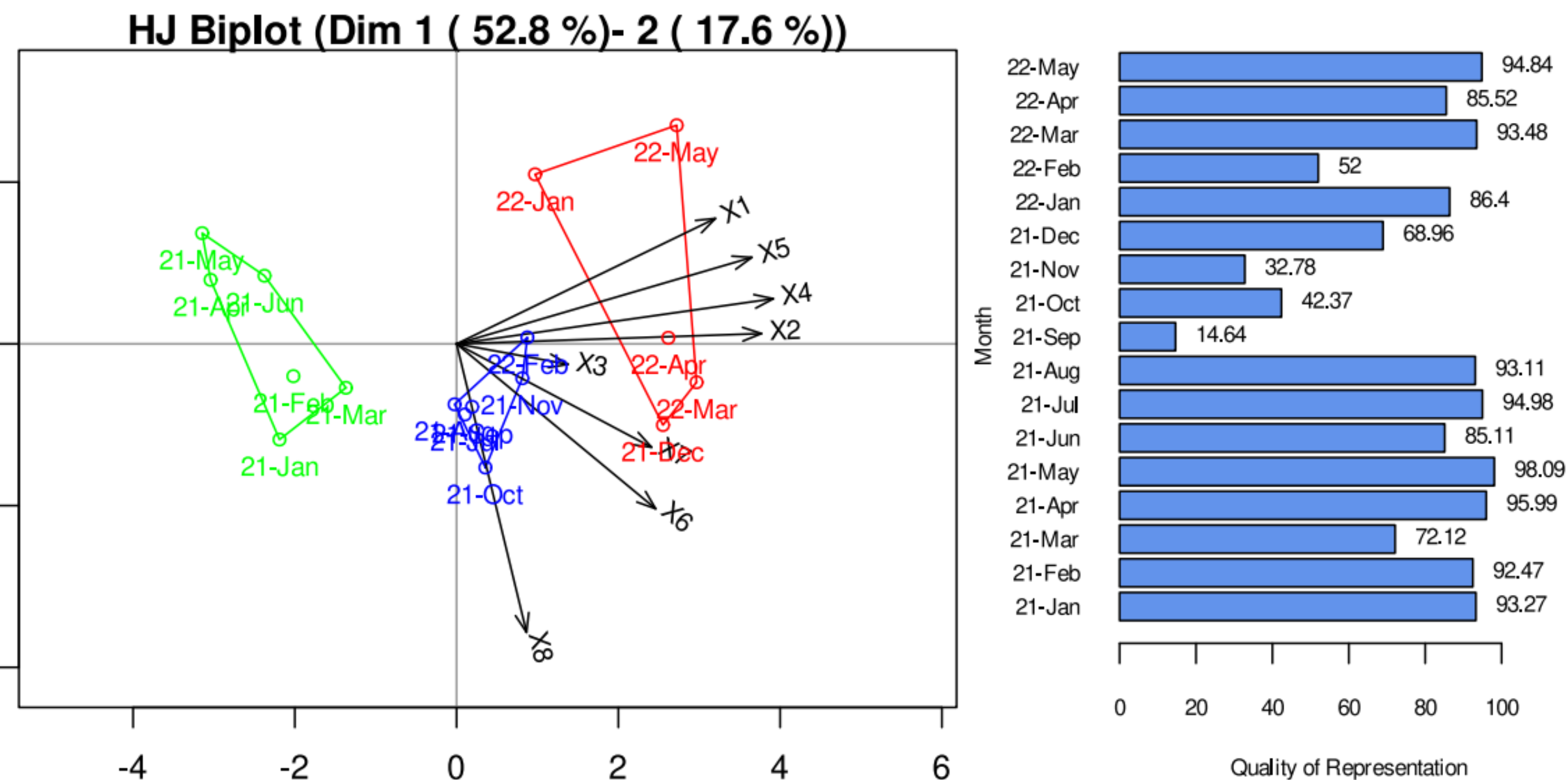
- The clusters correspond almost perfectly with the first and second semesters of 2021 and the first semester of 2022.

- Clusters classified as the months with the lowest, medium, and highest crime numbers.

- There is an increase in almost all variables of crime from one semester to another.

Crime per Province

Fig. 2: HJ-Biplot for provinces and number of crimes.
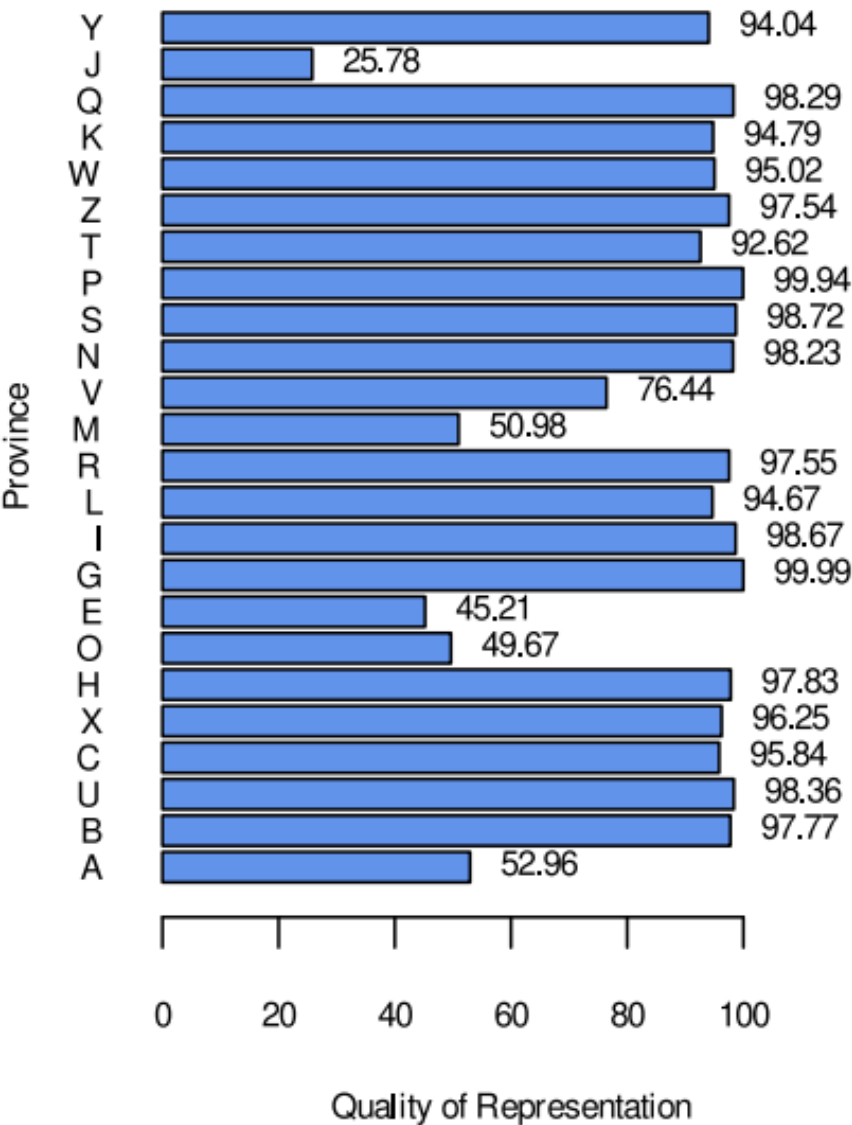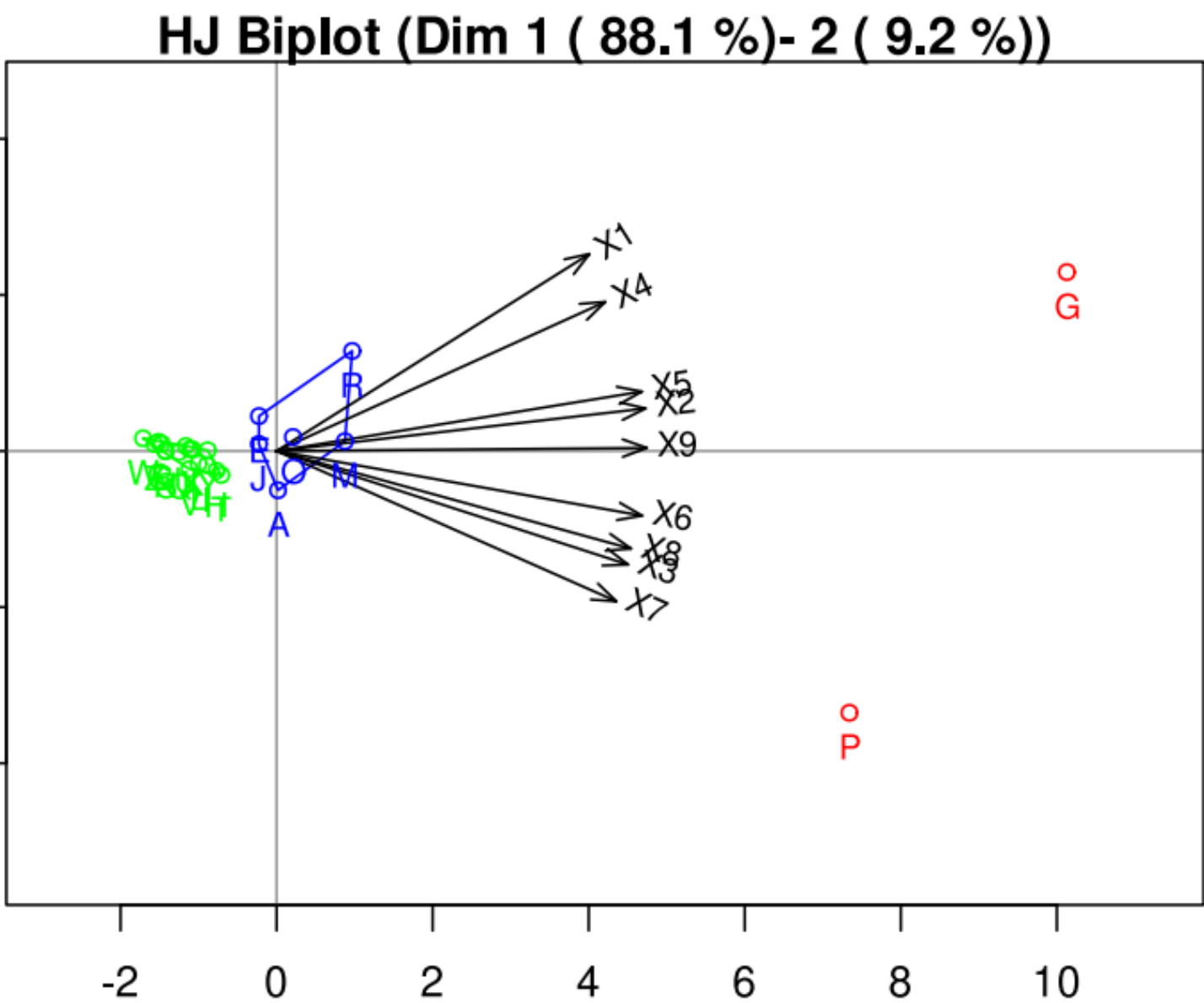
- High correlation of almost all the variables with respect to X9 (Population):

- E.g. Provinces with a high number of inhabitants tend to have a high number of crimes, and vice versa.

- The number of crimes in Guayas and Pichincha significantly exceeds the rest of the provinces.

Crime Rate per 100,000 inhabitants
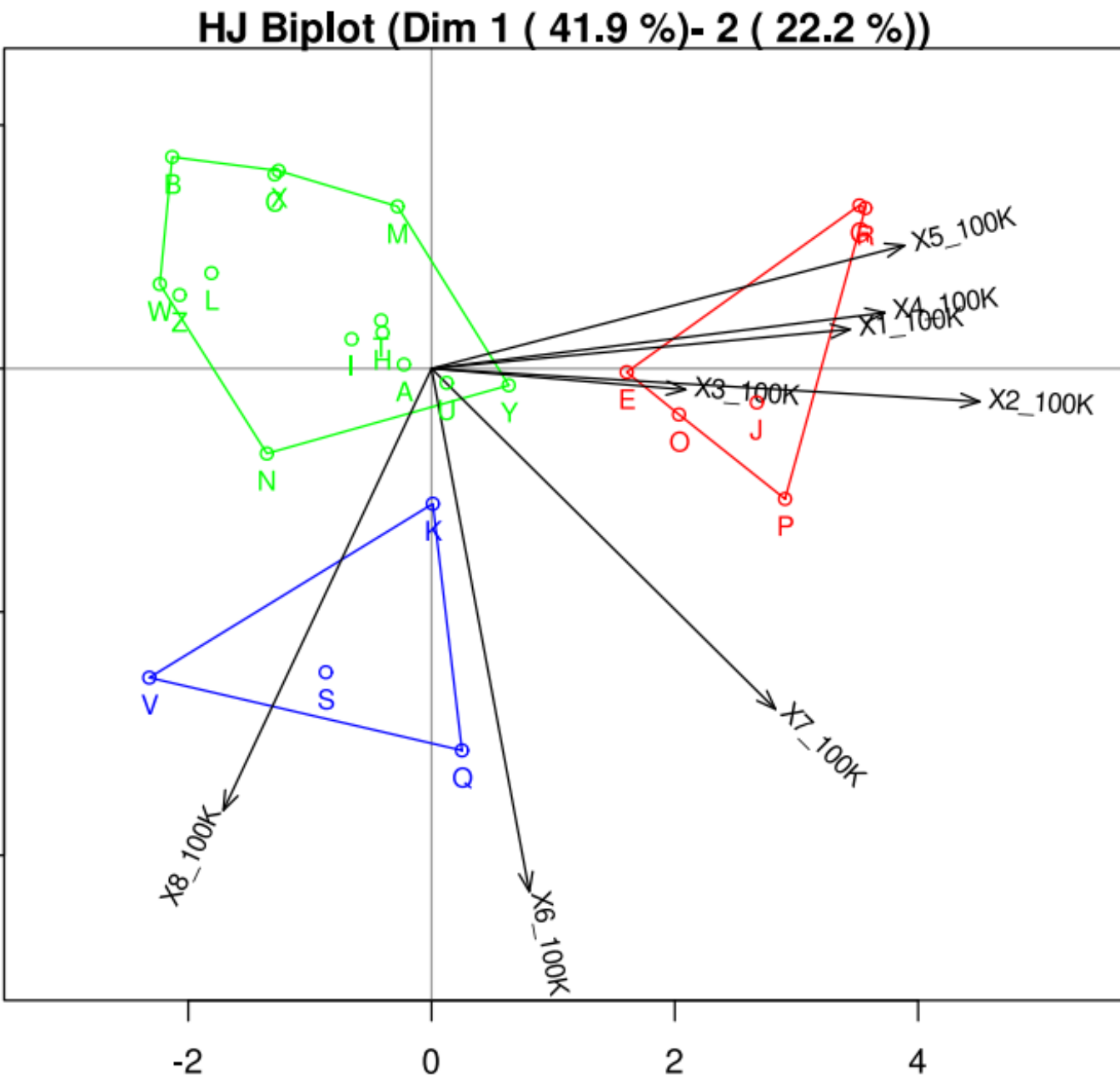
Fig. 3: HJ-Biplot for provinces and crime per 100,000 inhabitants.

- First group (highest rates in X1, X2, X3, X4, and X5): 4 of 7 provinces of the Coast region: El Oro, Esmeraldas, Guayas, Los Ríos, and 1 of the Interandean region: Pichincha.

- Second group (lowest crime rates in most variables): Manabí, Santo Domingo, Napo, Zamora Chinchipe, Galápagos, and all the provinces of the Interandean region, except Pichincha.

- Third group (Highest rates in X6 and X8): Morona Santiago, Pastaza, Sucumbíos, and Orellana, 4 of 6 provinces in the Amazon region.

- Crime, Unemployment, and Poverty in 2019

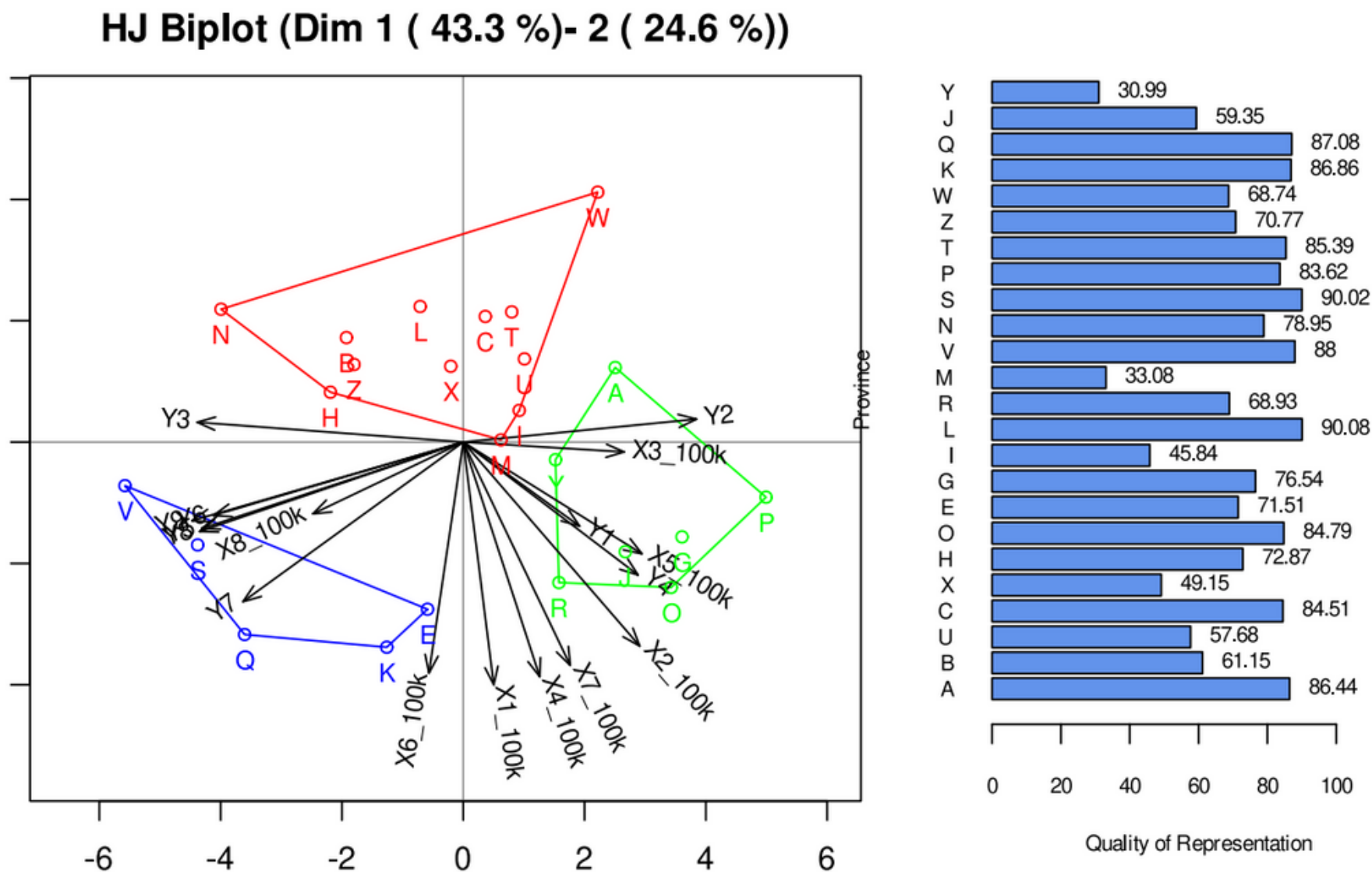- Crime, Unemployment, and Poverty in 2021

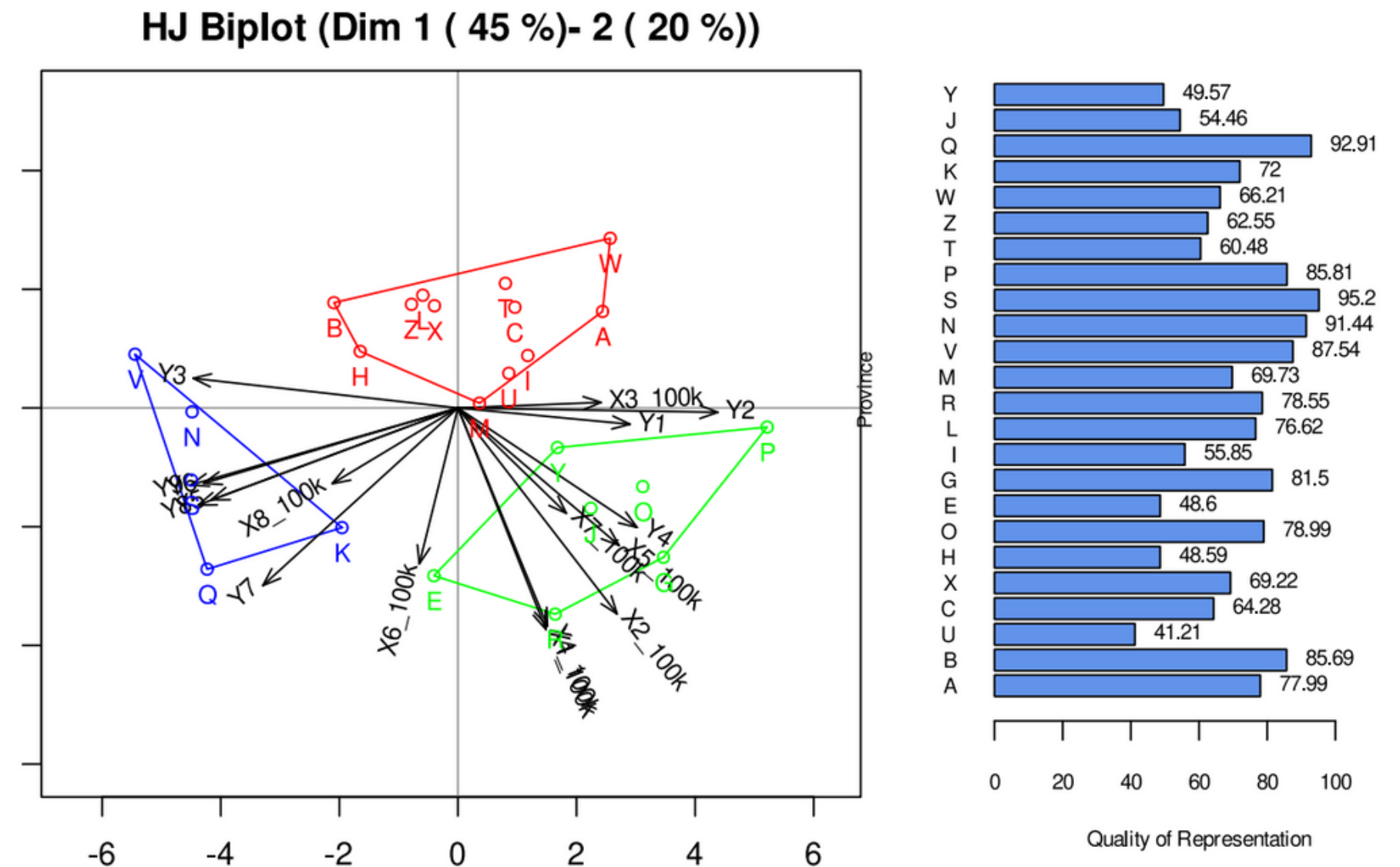Fig. 4: HJ-Biplot for Crime, unemployment, and Poverty of 2019.

Fig. 5: HJ-Biplot for crime, unemployment, and poverty of 2021.

- In both years, there is a strong correlation between poverty variables Y5, Y6, Y7, Y8, Y9, and X8_100K (rape rate) which suggests a statistical correlation between these variables.

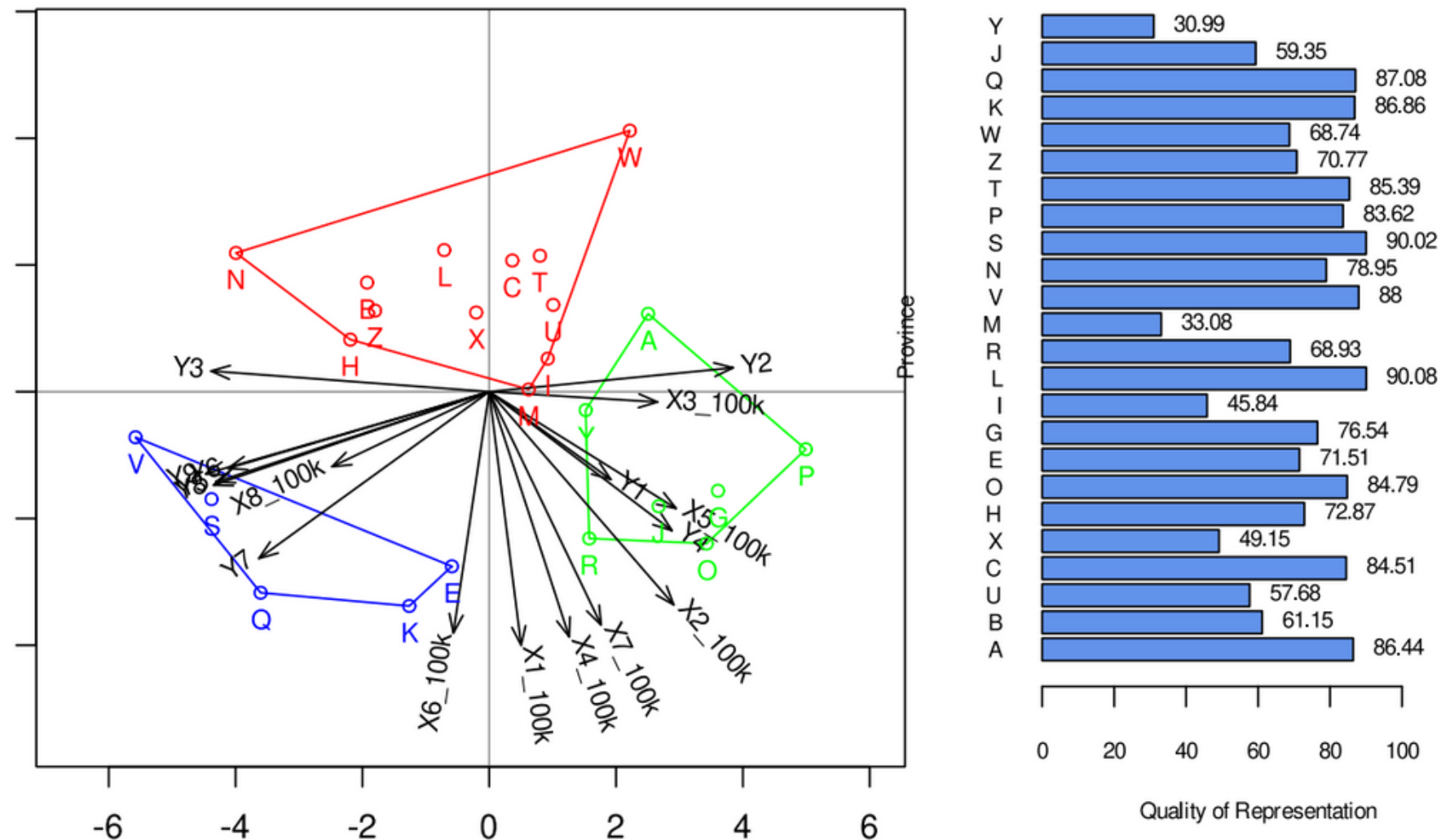- The cluster in blue presents high poverty and rape rates. Furthermore, it contains 4 of 6 provinces of the Amazon in 2019, and 5 of 6 in 2021.

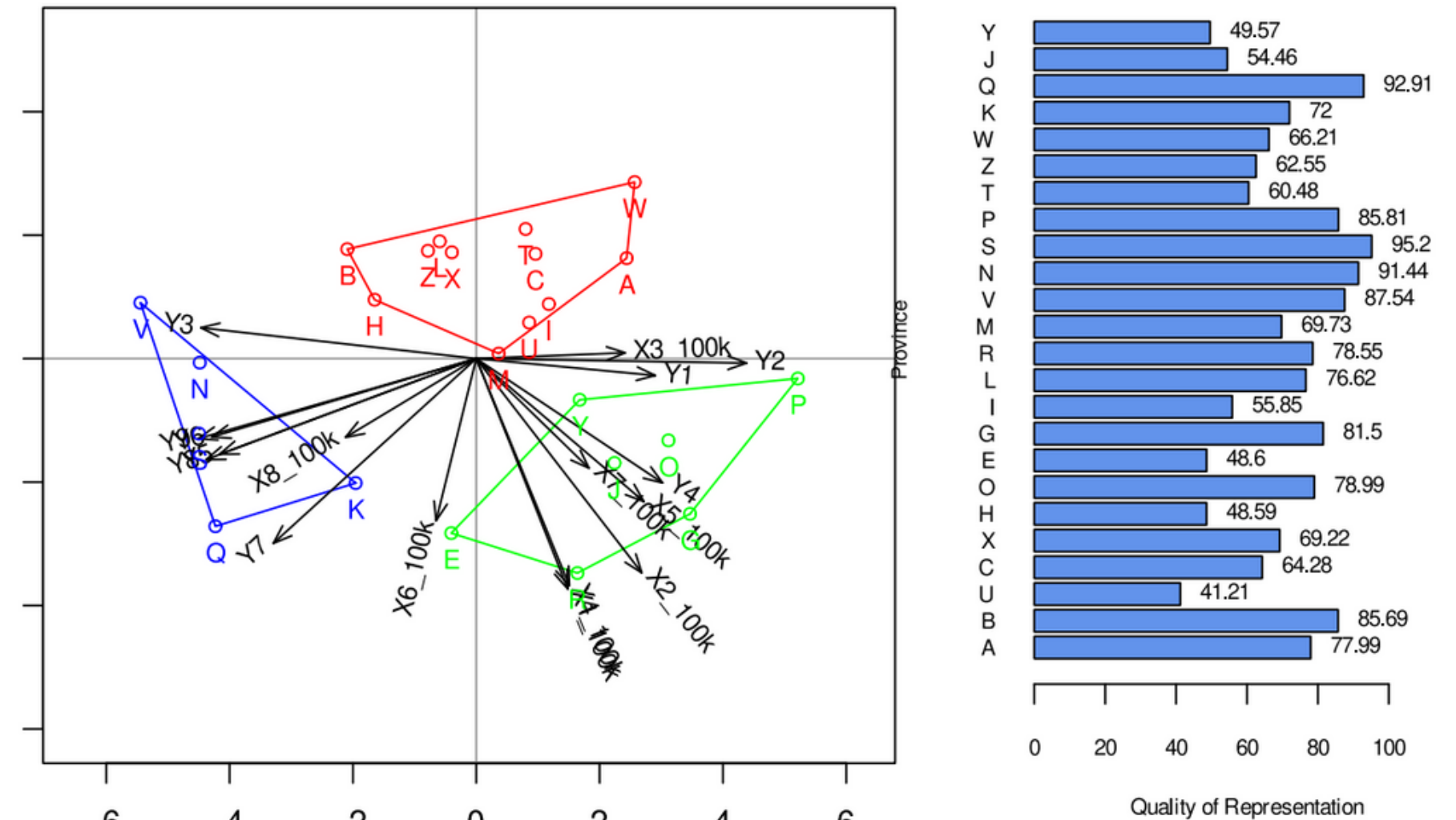Fig. 4: HJ-Biplot for Crime, unemployment, and Poverty of 2019.

Fig. 5: HJ-Biplot for crime, unemployment, and poverty of 2021.

- The 3 resulting clusters match almost perfectly with the 3 primary regions: Coast, Interadean region, and Amazon. This suggests that crime, unemployment, and poverty in Ecuador follow regional patterns.

- The relationship between the variables of poverty, unemployment, ad crime is very small or almost null.

4. Conclusion

Conclusion

In the study, we:

- Identified an increase in crime in Ecuador from January 2021 to May 2022.

- Found a high correlation between the number of crimes and the population.

- Identified a regional pattern in the variables of crime, poverty, and unemployment.

- Identified a statistical correlation between the poverty variables and rape rate.

- Concluded that **there is not a strong statistical correlation between the variables of unemployment, poverty, and crime** in Ecuador.

5. References

References

1. Instituto Nacional de Estadística y Censos, https://www.ecuadorencifras.gob.ec/institucional/home/

2. Cubilla-Montilla, M., Nieto-Librero, A.B., Galindo-Villardón, M.P., TorresCubilla, C.A.: Sparse hj biplot: A new methodology via elastic net. Mathematics 9(11), 1298 (2021). https://doi.org/10.3390/math9111298

3. España, S.: La inseguridad en Ecuador encierra en casa a los ciudadanos y saca a los militares a las calles. El País (Feb 2022)

4. Fu, L., Lin, P., Vasilakos, A.V., Wang, S.: An overview of recent multi-view clustering. Neurocomputing 402, 148–161 (2020)

5. Gabriel, K.R.: The biplot graphic display of matrices with application to principal component analysis. Biometrika 58(3), 453–467 (1971)

6. Galindo, M.: Una alternativa de representación simultánea: HJ-Biplot. Qüestiió 10(1), 13 – 23 (1986)

7. Golalipour, K., Akbari, E., Hamidi, S.S., Lee, M., Enayatifar, R.: From clustering to clustering ensemble selection: A review. Engineering Applications of Artificial Intelligence 104, 104388 (2021)

8. González, J.M., Fidalgo, M.R., Martín, E.J., Vicente, S.: Study of the evolution of air pollution in Salamanca (Spain) along a five-year period (1994–1998) using HJBiplot simultaneous representation analysis. Environmental Modelling & Software 21(1), 61–68 (2006)

Thanks for your attention!