OPEN ACCESS

Journal of Advanced Research in Social Sciences
ISSN 2538-919X

Mathematical Relationships Between Legal and Genetic Data in Argentina

Gabriel A. Ruiz*, and Carmelo J. Felice

Ph.D., Laboratorio de Medios e Interfases, Departamento de Bioingeniería, FACET-UNT and INSIBIO-CONICET, Argentina

ARTICLE INFO

Keywords:
political instability,
contempt,
toxic social behavior,
genetic markers,
non-compliance with laws

ABSTRACT

It is known that the contempt for authority by both indigenous peoples and Spaniards during the conquest and colonisation of America is a cause of political instability in Argentina. Contempt for authority can be quantified by measuring the level of non-compliance with laws and regulations, and indigenous or Spanish origins can be determined through genetic quantification. This study investigates the underlying mathematical relationships among legal and genetic variables without taking causality into account. Using genetic, legal, and geographical data from 24 Argentine provinces, significant correlations are identified between vertical fiscal imbalance, property tax evasion, and fatal traffic accident rates with Amerindian and African genetic content, excluding CABA. Unregistered work shows a high correlation with the latitudes of capital cities, probably due to the geographical and temporal sequence of Argentina's colonisation. Multidimensional cluster analysis reveals three regions in Argentina with similar social behaviours based on genetic composition, law compliance, and geographical proximity. These findings highlight the complex interplay between genetic heritage and socio-legal dynamics, offering new insights into Argentina's political instability.

1. Introduction

Political instability in Argentina has been a persistent issue for over four centuries, as demonstrated in previous studies by our research team (Felice and Ruiz, 2023a, 2023b). These studies identified contempt for authority as a primary cause of this instability, manifesting either directly by altering the tenure of rulers or indirectly through the violation of laws. Chaotic patterns of social behaviour from both mathematical and human perspectives were discovered by analysing temporal series documenting leadership periods in Latin America from the time of conquest and colonisation to the present day ethical and human perspectives.

Contempt for authority significantly impacts a country's political stability. Our previous publications analysed these time series descriptively, proposing two primary hypotheses:

*Corresponding author's E-mail address: gruiz@herrera.unt.edu.ar

Cite this article as:
Ruiz, G. A., & Felice, C. J. (2025). Mathematical Relationships Between Legal and Genetic Data in Argentina. *Journal of Advanced Research in Social Sciences, 8*(2): 1-35. https://doi.org/10.33422/jarss.v8i2.1309

1) Contempt of the Spanish in America for the authority of the Spanish crown is a primary causal factor of political instability in Argentina.

2) Indigenous contempt for the authority of the white conqueror and coloniser is a primary causal factor of political instability in Argentina.

In subsequent publications (Felice & Ruiz, 2024; Felice & Ruiz, 2025), these hypotheses were validated by assessing historical, psychological, and sociological aspects. Historical aspects included historical data, archaeology, bioarchaeology, and statistical information. Psychological aspects considered suicides, acculturation, post-traumatic stress disorder, aggressive behaviors, addictions, loss of self-esteem, and other negative effects of the conquests of indigenous peoples. Sociological analysis evaluated the intergenerational transmission of social behaviours, both among indigenous peoples and different cultures around the world.

To further support these hypotheses, it was necessary to collect and analyse legal and genetic data quantitatively.

The present study examines violations of a large set of laws and regulations as an indirect indication of contempt for authority. Additionally, the Amerindian and extracontinental genetic content of inhabitants throughout Argentina was analysed to distinguish between indigenous and Spanish contempt as expressed in the two hypotheses.

2. Theoretical framework and literature review

To understand this work, it must be taken into account that political stability is defined by the indicator "time of a ruler's tenure in office" (Felice and Ruiz, 2023a, 2023b). This definition replaces the construct of political stability with the quantitative indicator of politicians' stability, focusing the problem on individuals rather than the abstract concept of politics. Politicians' stability is affected by the behaviour of the surrounding people, and such behaviour depends on their level of contempt for authority. The relationship between these concepts and the justification for using this definition is explored in depth in other publications (Felice and Ruiz, 2025).

To attribute toxic social behaviour to either Hispanic or Indigenous heritage in a genetically highly admixed society such as Argentinean society, this work analytically relates two groups of data: non-compliance with laws and norms, and the Amerindian, Eurasian, and African genetic content of the inhabitants of the 24 Argentine provinces.

The analysis carried out in this work is solely intended to find mathematical relationships between the variables of both data groups without indicating causality. The causality of the relationships is demonstrated in two previous works through propositionally logical statements and a logical analysis of them (Felice and Ruiz, 2024; Felice and Ruiz, 2025).

On the other hand, previous research has explored the relationships between genetics, geography, and social behaviours, providing a foundation for our study. For example, Cavalli-Sforza et al. (1994) demonstrated the role of genetics in shaping human societies, while studies by Comas et al. (2004) and Wang et al. (2013) highlighted the impact of genetic diversity on social behaviours and structures. Furthermore, research by Diamond (1997) and Olson et al. (2019) examined how geographical factors influence the development of political and social systems. These studies collectively underscore the complex interplay between genetic heritage and socio-political dynamics.

3. Research objectives and approach

Our results reveal correlations between certain legal and genetic data sets, while others do not show statistically significant correlations. To uncover underlying connections, geographical data was supplemented. For instance, a high correlation between unregistered work and the latitude of the capitals was found. Geographical coordinates of the capitals provide a different perspective, allowing provinces to be grouped into clusters with similar genetic, legal, and geographic characteristics through multidimensional cluster analysis. This approach overcomes the limitations of using only genetic and legal data to establish possible patterns of social behaviour.

4. Materials and methods

4.1. Legal data

This section describes the legal violations and regulations analysed, methods for obtaining each dataset, and their relevance to measuring contempt for authority. Numeric data is selected from sets of law and norm violations representing toxic social behaviours across all social classes and provinces in Argentina. The selection focuses on variables reflecting non-compliance with laws, ensuring they are verifiable and reasonably certain. Emphasis is placed on addressing limitations, including bribery, evasion, theft, forgery, and the absence of enforcement. Inclusion and exclusion criteria are applied to define the most representative dataset.

4.2. Legal violations dataset

Datasets representing toxic social behaviours across all social classes and provinces were selected. The selection of each indicator, along with the inclusion/exclusion criteria and quantification methods, are detailed in the supplementary materials "S4 Non-Technical Losses" and "S5 Legal Selection." The numerical values for selected violations are provided in “S3 Broken Laws”.

The following sets of laws and rules are included: (1) Electricity theft, (2) Traffic violations, (3) Fatal traffic accidents, (4) Evasion of gross taxes, (5) Evasion of property tax, (6) Prison population, (7) Detainees in police stations, (8) Crimes against economic and financial order, (9) Crimes against public administration, (10) Money laundering and terrorism financing crimes, (11) Unregistered labour, (12) Illegal street sale of products, (13) Claims for health services, (14) Workplace accidents and (15) Lawsuits.

4.3. Genetic data

Genetic markers are used to quantify the genetic composition of Argentinians, creating maps of provincial genetic composition. The process of selecting genetic markers (genetic data) to define "indigenous" and "Spanish" within a broader context of "Amerindian" and "Eurasian" is also described, along with the collection of quantitative values of the representative genetic percentage content of the inhabitants of all Argentine provinces. For each selected marker, its interpretation in terms of social behaviours is justified. Three sets of markers are analysed:

- **Y-DNA**: Indicates patrilineal descent.
- **mtDNA**: Reveals maternal lineage.
- **Autosomal DNA**: Identifies SNP markers across chromosomes.

The genetic markers used to quantify the genetic composition of Argentinians are chosen independently of their formal affiliation with a specific social group. This approach frees the analysis from any social classification, group membership, or ideology, aiming to validate our initial hypotheses. The analysis justifying genetic information inclusion and marker selection is in “S6 Genes Selection”. Data is available for the 23 provinces and Buenos Aires in “S2 Genetic Data.”

4.4. Cluster analysis

To identify similar social behaviour patterns in the provinces, clustering was performed using two algorithms:

- **Fuzzy C-Means Algorithm**: Minimises an objective function to obtain an optimal fuzzy partition.
- **K-Means Algorithm**: Groups objects into clusters by minimising distances between objects and cluster centroids.

4.5. Data used

- Latitude and longitude of provinces.
- Total Amerindian+African genetic content.
- 13 legal non-compliance indicators.

The clustering process aimed to define homogeneous province groups based on the selected features. The optimal cluster number was determined using the average silhouette coefficient. Further details can be found in the supplementary material "S9 Cluster Analysis".

4.6. Statistical analyses

The Pearson correlation coefficient was used to measure the linear relationship between legal non-compliance indicators and genetic content. The choice of Pearson correlation is justified due to its ability to quantify the strength and direction of the linear association between two continuous variables. However, given the complex and multifactorial nature of social behaviours, additional analyses such as Spearman's rank correlation and regression analysis could be employed to account for non-linear relationships and potential confounding variables. Cluster analysis was performed using Fuzzy C-Means and K-Means algorithms to identify patterns of social behaviour in different provinces. The use of these clustering algorithms is justified by their ability to group data points based on similarity, which is particularly useful in identifying homogeneous groups of provinces with similar legal non-compliance and genetic profiles. The determination of the optimal number of clusters using the silhouette coefficient ensures the reliability and validity of the clustering results.

4.7. Limitations and confounding variables

The primary limitation of the data sources is the potential for underreporting or misreporting of legal violations, which could affect the accuracy of the legal data. Additionally, genetic data may be limited by the availability and representativeness of the samples used in the analysed publications. Potential confounding variables include socio-economic factors, cultural influences, and historical events that were not directly measured but could influence both legal compliance and genetic composition. Future studies should aim to incorporate these factors to provide a more comprehensive understanding of the observed correlations.

All the information and datasets used for analysis, geographical visualisation, and mathematical processing are available in the supplementary material.

5. Regarding the geographical maps

The maps are created using an ad-hoc software developed in Python. In the graphics, 32 levels of grayscale are used, except in 2 cases: the map of non-compliance peaks and the map of province groupings in clusters. In the first case, provinces that have at least one non-compliance peak compared to the rest of the provinces are shown in red. In the second case, clusters are distinguished from each other using different hatchings.

6. Provinces analysed

For the various analyses and statistics in this work, the 23 Argentine provinces and one federal district, the Autonomous City of Buenos Aires -CABA- are considered, which can practically be regarded as another province with the same rights and obligations as the rest. The provinces are Buenos Aires, Catamarca, Chaco, Chubut, Córdoba, Corrientes, Entre Ríos, Formosa, Jujuy, La Pampa, La Rioja, Mendoza, Misiones, Neuquén, Río Negro, Salta, San Juan, San Luis, Santa Cruz, Santa Fe, Santiago del Estero, Tierra del Fuego, Antarctica and the South Atlantic Islands, and Tucumán.

7. Results

The results in this work are grouped into 8 sets: 1) Geographic maps of non-compliance, 2) Geographic maps of genetic content, 3) Genographic profile maps, 4) Ternary plots, 5) Genetic content-law violation relationship, 6) Latitude-law violation relationship, 7) Cluster analysis, and 8) Radial diagrams. All the legal and genetic data used are found in the supplementary material *"S7 Genetic and legal*".

7.1. Map of non-compliance peaks

The first overall indicator takes into account that all available data and non-compliance estimates have different degrees of uncertainty. This source of errors comes from incomplete statistics, variation in sampling sites over time, contradictory data without additional information for contrast, or distorted or seemingly incorrect data due to crimes such as bribery or forgery of public documents, which are very difficult or impossible to determine. To reduce uncertainty, the first indicator only shows in which province the maximum non-compliance values of a specific law were obtained. Peak values are defined as any value ≥ 0.99.

Figure 1 shows in red the provinces that exhibit peak values, clarifying which laws are violated in each case.

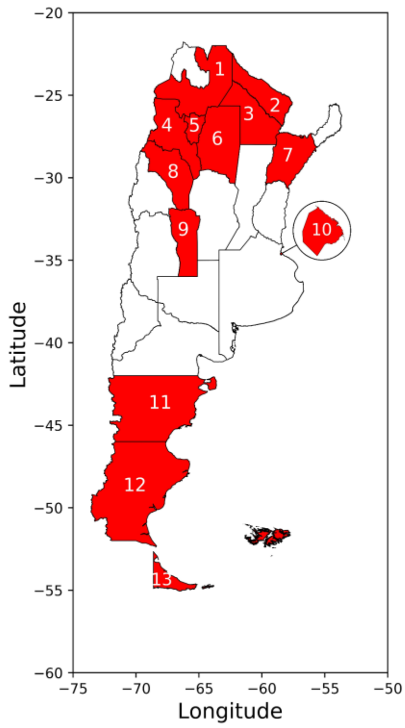

Figure 1. Province distribution of broken laws peaks. In the figure numbers correspond to the following broken laws: 1) Property tax, Incarcerated population; 2) Property Tax, Vertical fiscal imbalance; 3)Property Tax; 4)Vertical Fiscal Imbalance; 5) Unregistered labour; 6) Detainees in police stations, Tax Ineff., Fatality rate 7) Electricity theft; 8) Property tax, Vertical fiscal imbalance; 9) Accidents at work; 10) Money laundering, Administrative crimes, Traffic infractions, Illegal trade, Health claims; 11) Property tax, Tax inefficiency; 12) Property tax and 13) Economic crimes, Property tax, Illegal trade. High resolution Figure 1

Figure 1 displays the provinces where all non-compliance peaks occur for at least one of the mentioned laws. This means that the marked provinces have the highest percentage of non-compliance for one or more laws in the analysed set of law violations. In numerical terms, the provinces that have between 99% and 100% of the maximum value are shown. Tables with all values are attached in the supplementary material.

To rule out the possibility that the plot shown in Figure 1 is simply random, the probability of such a pattern of spikes occurring is calculated. For this, it is considered that the Argentine provinces are usually grouped into 5 regions (INDEC, 2023): Northwest (NOA) with 6 provinces, Northeast (NEA) with 4, Cuyo with 3, Pampeana with 6 and finally Patagonia with 5 provinces. Table 1 shows which provinces are associated with each region in Argentina.

Table 1.
Distribution of the Argentine provinces in regions

Region	**Number of provinces**	**Provinces**
NOA	6	Jujuy, Salta, Tucumán, Catamarca, La Rioja y Santiago del Estero
NEA	4	Chaco, Formosa, Corrientes y Misiones
CUYO	3	Mendoza, San Juan y San Luis
PAMPEANA	6	Santa Fé, Entre Ríos, La Pampa, Córdoba, CABA y Buenos Aires
PATAGÓNICA	5	Neuquén, Río Negro, Chubut, Santa Cruz y Tierra del Fuego e Islas del Atlántico Sur

Now, to calculate the probability of selecting 5 provinces from NOA Region, 3 provinces from NEA Region, 1 provinces from CUYO Region, 1 province from PAMPEANA Region and 3 provinces from PATAGÓNICA Region when selecting 13 provinces at random, it should be used the concept of combinations, specifically the multinomial coefficient, whose formula is given by the equation 1:

$$C(n; k_1, k_2, \dots k_m) = \frac{n!}{k_1! \cdot k_2! \cdot \dots k_m!} \qquad \text{Equation 1}$$

Where n is the total number of items (13 provinces in this case) and k_1,k_2, ..., k_m are the individual group sizes (5, 3, 1, 1 and 3 in this case). Using this formula, the number of possible combinations that satisfy the given conditions can be calculated. This number is then divided by the total number of possible outcomes (total combinations when selecting 13 provinces).

- Number of favourable outcomes = C(6; 5) x C(4; 3) x C(3; 1) x C(6; 1) x C(5; 3) = 6 x 4 x 3 x 6 x 10 = 4320
- Total number of possible outcomes = C(24; 13) = 24! / (13! x (24-13)!) = 2,496,144

Therefore, the probability that 5 provinces belong to NOA Region, 3 provinces belong to NEA Region, 1 province belongs to CUYO Region, 1 province belongs to PAMPEANA Region and 3 provinces belong to PATAGÓNICA Region when selecting 13 provinces at random is approximately 0.00173

7.2. Map of overall averages

In Figure 2, the average distribution per province is shown for all offences, including Money Laundering, Administrative Crimes, Economic Crimes, Electricity Theft, Accidents at Work, Property Tax Evasion, Vertical Fiscal Imbalance, Unregistered Labour, Incarcerated Population, Detainees in Police Stations, Tax Inefficiencies, Traffic Infractions, Traffic Accidents, and Illegal Mall and Stalls.

It should be noted that all non-compliance graphs presented are referenced to the range, meaning that when a province is blank (value 0), it does not imply that it has not violated that law but rather that its non-compliance is minimal compared to other provinces.

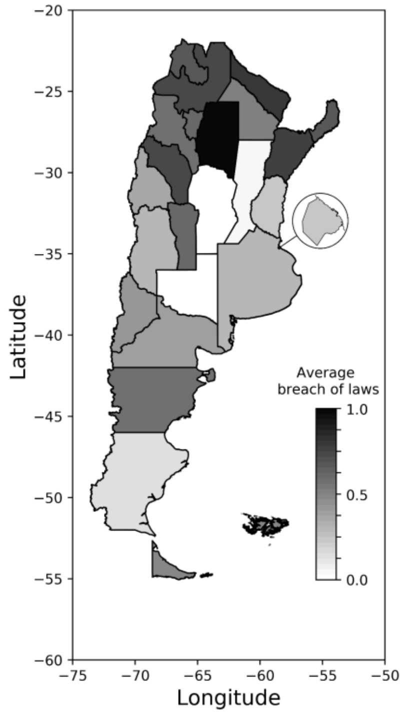

Figure 2. Map of Average Law Non-Compliance Values per Province. [High resolution Figure 2]

The distribution reveals a higher number of law non-compliance instances in the northern part of the country. It's important to clarify that the distribution does not imply economic impact; values in this regard must be defined case by case and were not analysed in this study.

7.3. Crimes against economic-financial order

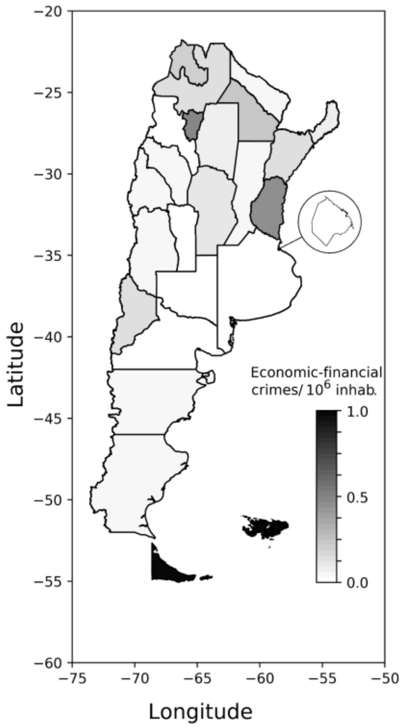

Figure 3. Crimes against the economic and financial order are depicted by province and per million inhabitants. [High resolution Figure 3]

The northern region of the country exhibits provinces with the highest number of crimes against the economic and financial order (Figure 3). The maximum value is found in Tierra del Fuego.

The distribution in 2019 is similar to the distribution in 2018. However, the 2017 distribution has an anomaly in Río Negro with a data point 10 times higher than the rest of the provinces, as shown in Figure 4.

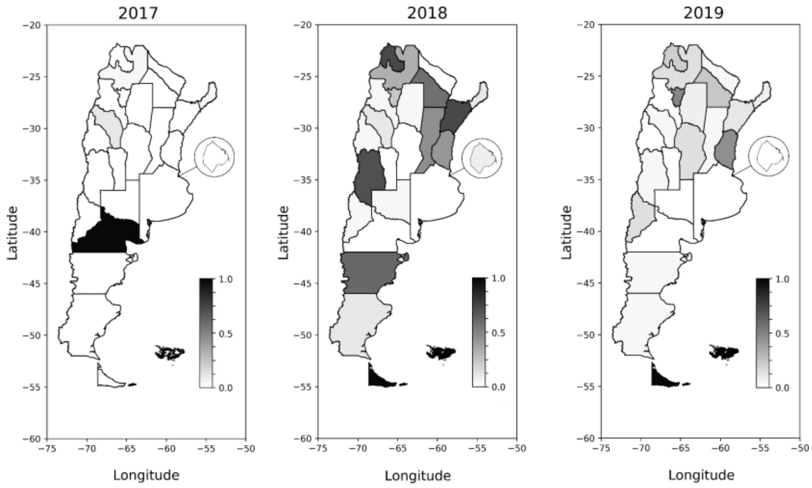

Figure 4. Crimes against the economic and financial order by province and per million inhabitants between 2017 and 2019. High resolution Figure 4

The values for Río Negro in the years 2017, 2018, and 2019 are 468, 0, and 1, respectively. The distribution of 2019 is adopted as representative.

7.4. Crimes against public administration (Administrative Crimes)

Crimes against public administration per million inhabitants in 2018 are illustrated in Figure 5A. The southern and central regions of the country show the highest values of the crime rate, with a peak in the province of Mendoza.

7.5. Money laundering and terrorism financing crimes

The map in Figure 5B shows that the maximum value occurs in CABA (Autonomous City of Buenos Aires), while several provinces exhibit minimum values.

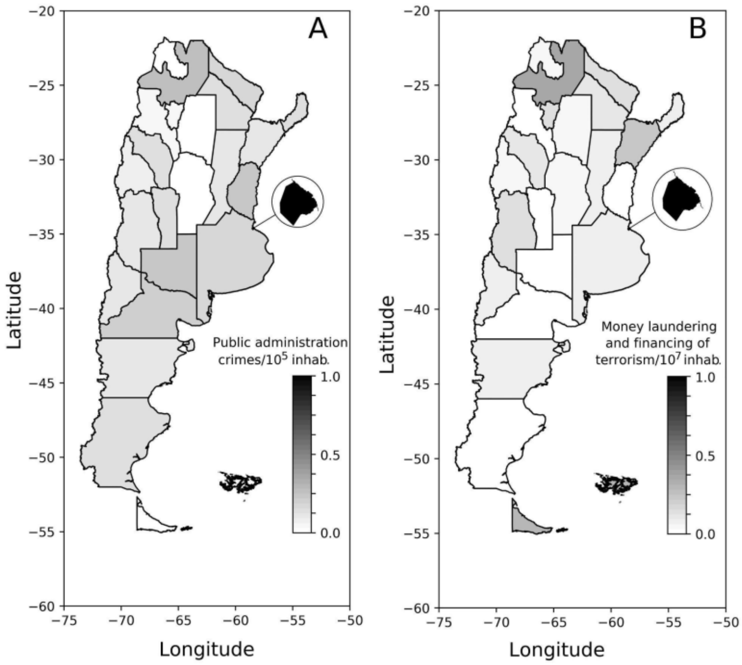

Figure 5. (A) Crimes against Public Administration per million inhabitants 2018. (B) Crimes of money laundering and terrorism financing are shown by province and per 10 million inhabitants. High resolution Figure 5

7.6. Tax evasion of gross income

In Figure 6A, a preliminary estimation of ***Tax ineff. A*** between 1960 and 2002 is shown, indicating the percentage increase needed in each province to reach the efficiency frontier (Di Gresia, 2003).

In Figure 6B, another estimation of the same tax using the same method for the period 1991-2012 (Rossignolo, 2017) is presented, expressing the level of inefficiency ***Tax ineff. B*** as a percentage of effective collection.

In both estimations, the northern provinces of the country exhibit the highest percentages of inefficiency in tax collection.

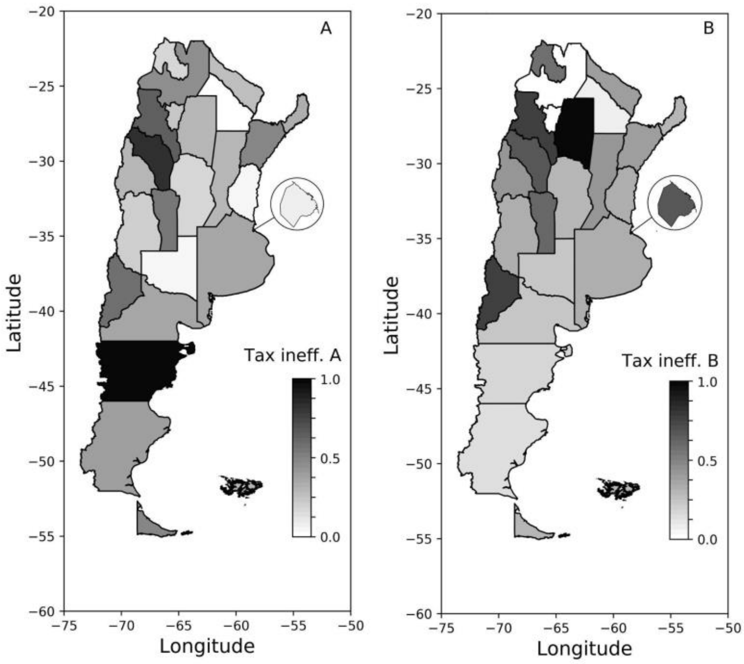

Figure 6. (A) Tax inefficiency A: The level of inefficiency is defined as the percentage increase in gross income collection leading to the efficiency frontier. Source: Di Gresia, 2003. (B) Tax inefficiency B: The level of inefficiency is represented as a percentage of effective collection Source: Rossignolo, 2017. Underutilization of the Potential Gross Income Tax between 1991-2012. Model I - Effective Tax Rate as an Explanatory Variable. High resolution Figure 6

7.7. Electricity theft

Figure 7A presents an estimation of theft, burglary, or fraud in the distribution of electrical energy in Argentina during the year 2019.

The northern region of Argentina shows the highest percentages of non-technical losses of electrical energy. The map includes losses between 2% and 33% with an average of 8%. To place in context, the average rate of electricity T&D losses -ETD- in Argentina is much higher than the OECD average (15% compared to 7% in OECD). ETD is the ratio between transport and distribution losses and the electricity consumption.

7.8. Fatalities in traffic accidents

The distribution of fatalities in traffic accidents per province per ten thousand registered vehicles is depicted in Figure 7B for the year 2019.

Provinces in the northern part of the country show the highest number of fatalities in traffic accidents, with the highest value in Santiago del Estero.

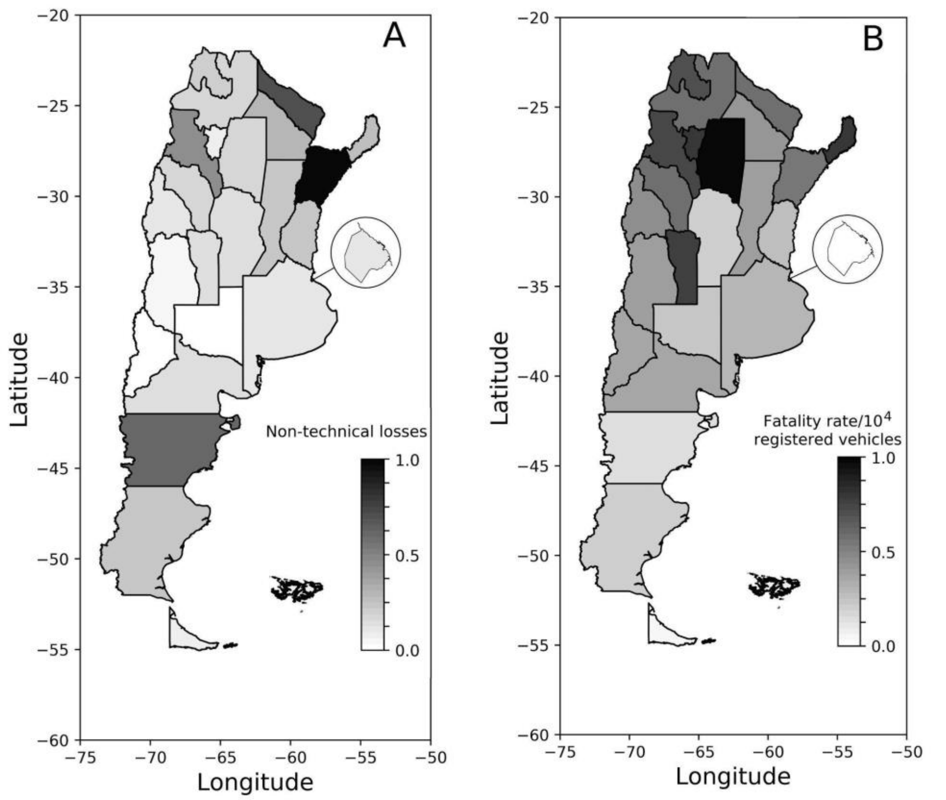

Figure 7. (A) Non-technical losses in percent by province. Source: *Informe Estadístico Anual 2019 from ADEERA.* (B) Fatalities in traffic accidents per ten thousand registered vehicles. Year 2019. Source: Tabla 35 in Anuario estadístico de seguridad vial. Dirección Nacional de Observatorio Vial. Dirección de Estadística Vial. High resolution Figure 7

7.9. Illegal street trading (IMS)

Figure 8A shows the peak of non-compliance in CABA, followed by Tierra del Fuego, Río Negro, Misiones, and Jujuy among the five highest values of the 24 provinces.

The case of CABA is very particular due to *La Salada*, the largest wholesale market in Argentina and the source of the figures attributed to CABA, is primarily composed of Bolivian, Paraguayan, Peruvian, and even Senegalese immigrants (Parra García, 2021; Benencia, 2017).

7.10. Claims for health services

Figure 8B reveals that the central region of Argentina has the highest number of complaints per inhabitants to the Superintendence of Health Services.

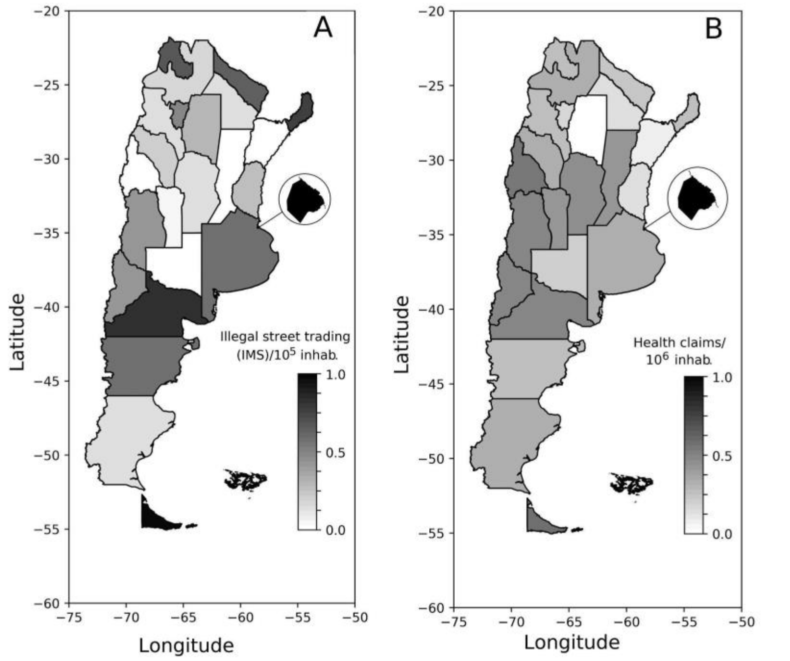

Figure 8. (A) Average per province of illegal vendors per 100,000 inhabitants, including stalls in street markets and illegal street vendors in November 2015. Data from Catamarca, La Pampa and Tierra del Fuego, CAME report from 2014. (B) Claims to the Superintendency of Health Services for different reasons, including medical surcharges. High resolution Figure 8

7.11. Unregistered labour

Values per province for the first quarter of 2019 are presented in Figure 9A (BCRA, 2019).

Figure 9A shows that the northern part of the country has the highest rates of unregistered work, with Tucumán showing the maximum value of unregistered work in Argentina at 48%.

7.12. Work accidents

The following map displays the rate of reported accidents per 1000 workers per province, according to the ***Annual Report on Occupational Accidents 2019*** from the Superintendence of Occupational Risks in Argentina (SRT, 2019).

In Figure 9B, the central region of Argentina exhibits the highest values of work accidents. Salta shows the maximum value of the work accident rate in the Northwest (NOA) region.

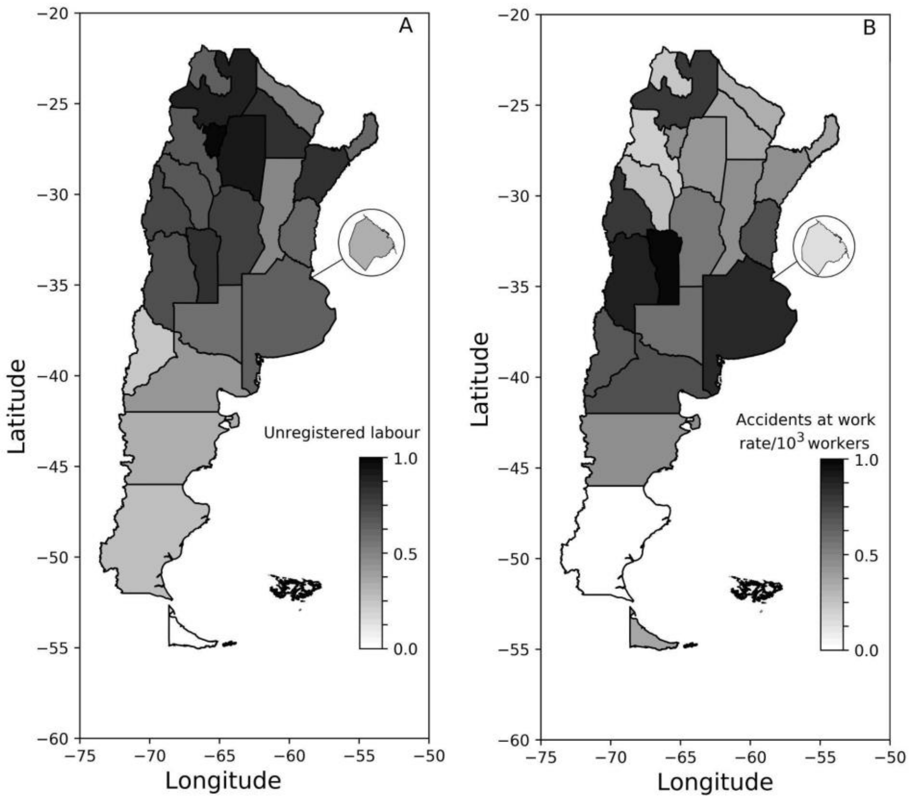

Figure 9. (A) Unregistered labour 2019. Salaried workers who do not have pension deductions. (B) Rate of work accidents per 1000 workers in 2019. High resolution Figure 9

7.13. Lawsuits

The map in Figure 10A illustrates the number of lawsuits for labour issues in the 4th quarter of 2019, distributed by province.

In this case, the provinces in the central region of Argentina show the highest number of lawsuits for labour issues.

7.14. Property tax collection

Indicator showing the relative revenue to the Gross Geographic Product. Figure 10B indicates that the highest relative revenue to the Gross Geographic Product is found in the central region of Argentina.

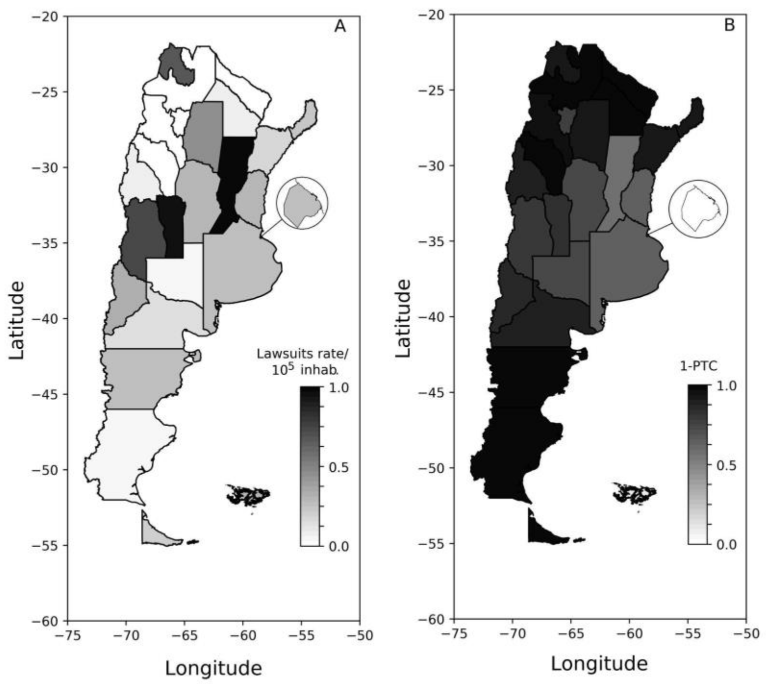

Figure 10. (A) Rate of lawsuits for labour issues per one hundred thousand workers. Source: *SRT (2019).* (B) PTC: Property Tax Collection by province as percentage of Total tax revenue. Figure 16 uses the indicator 1-PTC to better reflect the non-compliance with a law. Darker colours represent provinces with lower revenue collection. Source: Castro *et al.*, 2014. [High resolution Figure 10]

7.15. Vertical fiscal imbalance map

In Figure 11A, the northern and southern regions of Argentina show the highest dependencies of provinces on federal funds. The indicator reveals a geographical coincidence between provinces with lower revenue and those most dependent on federal funds.

7.16. Detained in police stations

In Figure 11B, it can be observed that the northern and southern regions of the country present the highest detention rates.

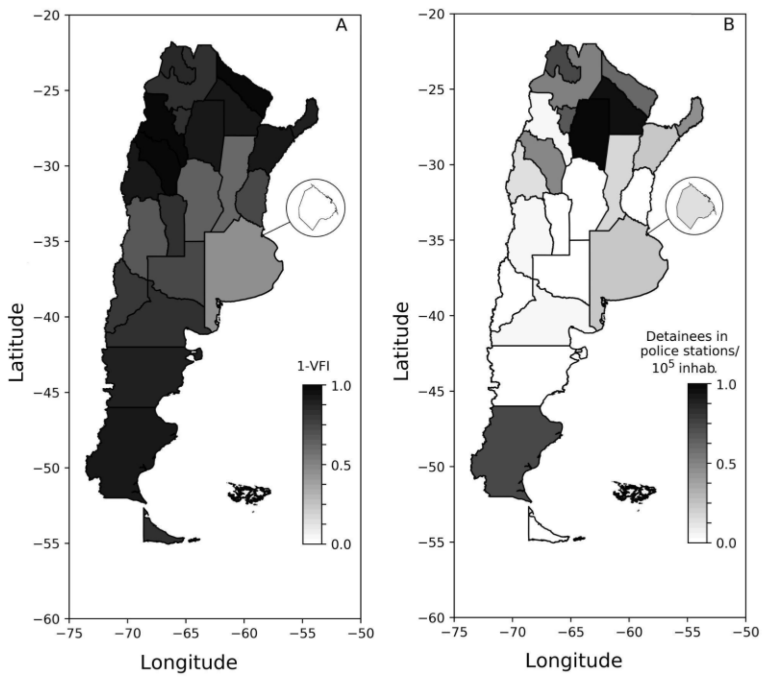

Figure 11. (A) Vertical Fiscal Imbalance. Figure uses the indicator 1-VFI to better reflect the non-compliance with a law. Dark colours indicate a higher dependence on federal funds. Source: Saiegh, 1999. (B) Detainees in police stations by province on Dec 31 2021. High resolution Figure 11.

7.17. Geographic maps of genetic content

All measurements of the genetic markers are found in the "S2 Genetic data" in the supplementary material. When there are multiple measurements in the same geographical location, the measurement with the highest number of samples is adopted to minimise measurement errors. Averages are not used because differences in measurement methods and experimental conditions are typical. For instance, different haplogroups or measurement methods are often employed.

In this publication, to minimise errors in statements regarding the ethnic composition of Argentine provinces, a diverse set of genetic methods from the literature is employed. The methods used, along with their respective references, can be found in the "S1 Genetic papers" in the supplementary material.

For calculations and graphics in this work, provincial capitals or the city with the highest population within a province are adopted as representative of the various law violations analysed. For example, Rosario (986,248 inhabitants) is adopted in Santa Fe instead of the capital Santa Fe de la Veracruz (404,602 inhabitants).

In the genetic distribution maps presented below, the genetic content data is assigned as a certain percentage to each province of Argentina. This method of visualisation allows us to compare against the different levels of law violations that, in the original publications, are also presented by provinces but in data tables.

To gain an overall view of the genetic content of Argentinians, two types of genetic marker distribution per province are analysed. One includes only Amerindian and African genes, and the other includes only European and Asian genes. In both cases, autosomal, mitochondrial, and sex-determining region Y markers are included.

Amerindian and African contents have been combined since they culturally have the same historical implications from the perspective of social behaviours, as analysed in this work. The same applies to European and Asian genes.

The first type of indicator is TAA or *Total Amerindian and African genetic load* defined in equation 1, which evaluates the average genetic content of the inhabitants of a province. The maximum value of this indicator is 600%, in the case of a total absence of European and Asian genes.

$$TAA = AA + AAF + MTA + MTAF + YA + YAF \quad (1)$$

where,

AA: Autosomal amerindian
AAF: Autosomal african
MTA: mtDNA amerindian
MTAF: mtDNA african
YA: Y-STR amerindian
YAF: Y-STR african.

The second indicator is TEA or *Total European and Asian genetic load* defined in equation 2, which evaluates the average genetic content of the inhabitants of a province. The maximum value of this indicator is 300%, in the case of a total absence of Amerindian and African genes.

$$TEA = AE + MTE + YE \quad (2)$$

where,

AE: Autosomal eurasian
MTE: mtDNA eurasian
YE: Y-STR eurasian

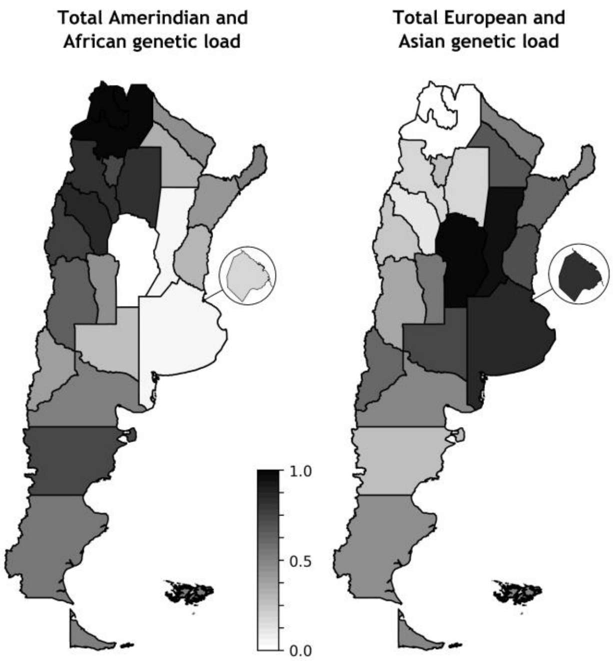

Figure 12. Genetic distribution maps: On the left, Total Amerindian and African genetic load, and on the right, Total European and Asian genetic load. High resolution Figure 12

For the maps (Figure 12), the values of TAA and TEA on a unit scale will be used, and to highlight the difference between colour tones by province, normalisation with respect to the data range, as indicated in equation 5 will be performed:

$$TAAu = TAA/600 \tag{3}$$

$$TEAu = TEA/300 \tag{4}$$

Normalising with respect to the data range:

$$TAAd = \frac{TAAu - MIN(TAAu)}{Max(TAAu) - Min(TAAu)} \tag{5}$$

7.18. Genographic profile maps

One way to understand the baseline values of Amerindian, Euro-Asian, and African components is to plot the genetic profile across Argentina. For the creation of genetic profile maps, the meridian M = -66.2214087 is adopted as the reference axial axis (see Figure 13). This axis traverses Argentina in its entirety, taking x=0 as the most northern point, parallel P1=21.7811538 south in the province of Jujuy, to the point where it intersects parallel P2=55.0566 south, which is tangent to the most southern continental point, Cape San Pío, located on the Grande Island of Tierra del Fuego. All distances from x=0 for the analysed locations are calculated by subtracting the value of the northern reference parallel from each parallel.

Figure 13. Location of the reference parallel and meridian. Map source: wikipedia. Creative Commons Atribución-CompartirIgual 4.0 Internacional. [High resolution Figure 13]

Latitude data must be corrected for flattening of the earth near the poles. In this work, a method based on the distance to the centre of the earth is used (Snyder 1987) to correct the data. Under the hypothesis that longitudes are always equal, each longitude degree is taken as constant in km. Then:

$$1 \text{ degree} = 111.12 \text{ km} \quad (6)$$

The correction coefficient (cc_{Lat}) is taken as:

$$cc_{Lat} = \left| cos\left(\underline{lat}\ \frac{\pi}{180}\right) \right| *111.12 \quad (7)$$

Where $\underline{lat}$ is the average of the latitudes in degrees of the Argentine capitals. Therefore, the product between latitudes in degrees and cc gives the corrected latitudes in km. Table 2 shows the latitudes of the Argentine capitals (https://www.geodatos.net/). The first two columns show the provinces and capitals respectively. In the third column the latitudes are expressed in degrees, in the fourth are corrected and expressed in km and in the last column the latitudes are corrected, expressed in km and referred to parallel P1 (see Figure 13).

Table 2.
Corrected latitude of the Argentine capitals and referred to the parallel P1

Argentine Province	**Capital**	**Lat [degree] S**	**Corrected Lat [km] S**	**Corrected Lat and referred to P1 [km] S**
Buenos Aires	AMBA*[1]	34,6132	3263,7705	1209,9677
CABA	CABA*[2]	34,6131	3263,7611	1209,9582
Catamarca	San Fernando del Valle de Catamarca	28.4667	630,3979	630,3979
Chaco	Resistencia	27.45	2588,3334	534,5305
Chubut	Comodoro Rivadavia*[3]	45.8641	4324,6478	2270,8450
Córdoba	Córdoba	31.4135	2962,0623	908,2595
Corrientes	Corrientes	27.4667	2589,9081	536,1053
Entre Ríos	Gualeguaychú*[4]	31.7327	2992,1606	938,3577
Formosa	Formosa	26.1833	2468,8929	415,0901
Jujuy	San Salvador de Jujuy	24.1833	2280,3076	226,5047
La Pampa	Santa Rosa	36.6167	3452,6859	1398,8831
La Rioja	La Rioja	29.4333	2775,3440	721,5412
Mendoza	Mendoza	32.8908	3101,3609	1047,5580
Misiones	Posadas	27.3833	2582,0441	528,2412
Neuquén	Neuquén	38.9516	3672,8498	1619,0470
Río Negro	Viedma	40.8135	3848,4133	1794,6105
Salta	Salta	24.7833	2336,8832	283,0803
San Juan	San Juan	31.5375	2973,7546	919,9518
San Luis	San Luis	33.295	3139,4740	1085,6711
Santa Cruz	Río Gallegos	51.6226	4867,6321	2813,8292
Santa Fé	Rosario*[5]	32,9468	3106,6413	1052,8384
Santiago del Estero	Santiago del Estero	27.7833	2619,7611	565,9583
Tierra del Fuego	Ushuaia	54.8108	5168,2559	3114,4531
Tucumán	San Miguel de Tucumán	26.8167	2528,6179	474,8150

*[1]: Although La Plata is the capital of Buenos Aires, AMBA is used instead La Plata because is the place where the highest population density of that province is concentrated (La Plata, 740k inhabitants; AMBA, ≈11000k inhabitants). In the AMBA population we do not include CABA inhabitants.
*[2]: CABA is the Autonomous City of Buenos Aires and is included as another province.
*[3]: The values from Comodoro Rivadavia were used as our provincial benchmark, as it is the most populous city in Chubut.
*[4]: The values from Gualeguaychú were used as our provincial benchmark, as there is only one data available. We assume that this data is historically representative of the population of Entre Ríos.
*[5]: The values from Rosario were used as our provincial benchmark, as it is the most populous city in Santa Fé.

Similarly to how latitudes were corrected and referenced to parallel P1, the same can be done with longitudes, which are referenced to meridian M (see Figure 13). In this case, the correction coefficient is given by equation 8.

$$cc_{Long} = \left| cos\left(\underline{long}\ \frac{\pi}{180}\right)\right| * 111.12 \tag{8}$$

Where $\underline{long}$ is the average of the longitudes in degrees of the Argentine capitals. Table 3 is the equivalent of Table 1 but for Longitudes.

Table 3.
Corrected longitudes of the Argentine capitals and referred to the meridian M

Argentine Province	Capital	Long [degree]	Corrected Long [km]	Corrected Long and ref to M [km]
Buenos Aires	AMBA*[1]	-58,377	-2873,576	386,124
CABA	CABA*[2]	-58,377	-2873,576	386,124
Catamarca	San Fernando del Valle de Catamarca	-65,783	-3238,136	21,564
Chaco	Resistencia	-58,983	-2903,411	356,289
Chubut	Comodoro Rivadavia*[3]	-67,497	-3322,469	-62,768
Córdoba	Córdoba	-64,181	-3159,266	100,435
Corrientes	Corrientes	-58,833	-2896,028	363,673
Entre Ríos	Gualeguaychú*[4]	-58,514	-2880,317	379,383
Formosa	Formosa	-58,183	-2864,032	395,669
Jujuy	San Salvador de Jujuy	-65,300	-3214,345	45,356
La Pampa	Santa Rosa	-64,283	-3164,300	95,401
La Rioja	La Rioja	-66,850	-3290,642	-30,942
Mendoza	Mendoza	-68,827	-3387,967	-128,267
Misiones	Posadas	-55,883	-2750,816	508,884
Neuquén	Neuquén	-68,059	-3350,160	-90,460
Río Negro	Viedma	-62,997	-3100,966	158,734
Salta	Salta	-65,417	-3220,088	39,613
San Juan	San Juan	-68,536	-3373,654	-113,953
San Luis	San Luis	-66,336	-3265,323	-5,623
Santa Cruz	Río Gallegos	-69,218	-3407,212	-147,512
Santa Fé	Rosario*[5]	-60,700	-2987,913	271,787
Santiago del Estero	Santiago del Estero	-64,267	-3163,480	96,221
Tierra del Fuego	Ushuaia	-68,316	-3362,801	-103,101
Tucumán	San Miguel de Tucumán	-65,217	-3210,243	49,458

In the following Figures, the genographic profile maps are shown. In Figure 14A, the data points represent the total Amerindian and African content; in Figure 14B, the autosomal Amerindian and African content; and in Figure 15A, the total Eurasian content. The data points have been connected using a Stepped Graph and interval-defined functions, composed of simpler function segments, which are defined in smaller intervals and joined at their endpoints with an adequate degree of smoothness known as splines (Data splines).

In Figure 14A, the Total Amerindian and African content is notably high in Argentina in the 21st century, despite the massive European immigration between 1860 and 1930. This asymmetry is explained because the proportion of male immigrants from Spain and Italy almost doubled the number of female immigrants (Tercer Censo Nacional, 1914; Avena, 2013; Resano, 2018), leaving the possibility of a large number of couples or marriages between European men and Amerindian or Amerindian descendants women. According to the 1914 census, the total population of Argentina was 7,905,502 inhabitants, of which 2,357,952 were foreigners, predominantly Italians and Spaniards, whose gender composition is shown in Table 4.

Table 4.
Number of women and men immigrants according to the 1914 census

Origin	Men	Women
España	512.742 (62%)	316.959 (38%)
Italia	587.497 (63%)	342.366 (37%)

Figure 14B indicates that the entire sampled population of the Argentine Republic presents Amerindian+African content fluctuating between 18% and 50%.

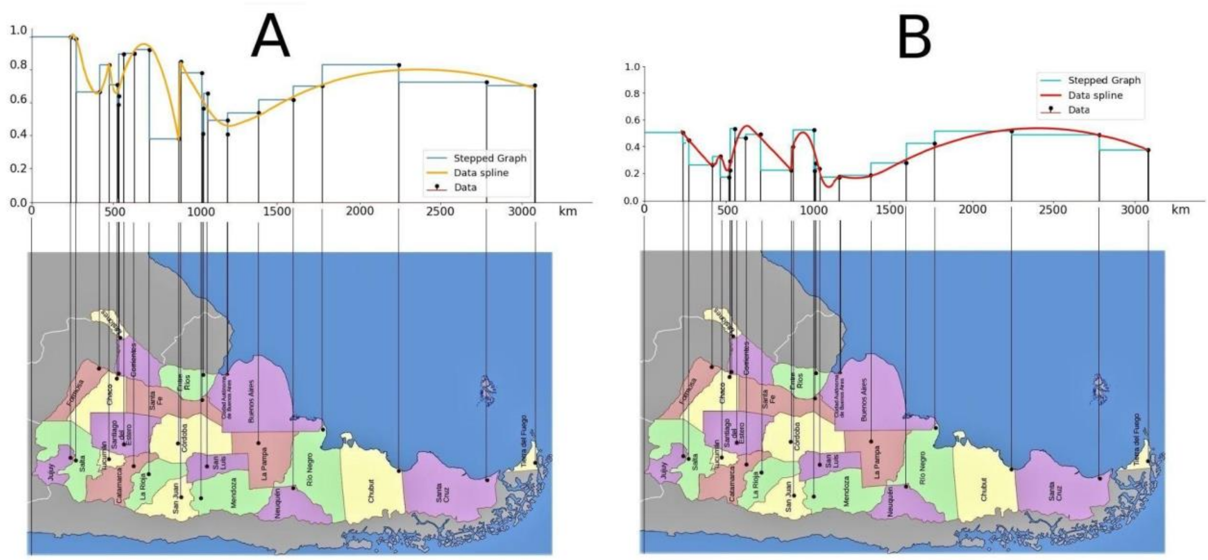

Figure 14. (A) Total Amerindian and African genographic profile, including mtDNA, autosomal and Y-STR markers. Orange curve is a second order spline. (B) Amerindian and African Autosomal genographic profile. Red curve is a second order spline. High resolution Figure 14

The elevated baseline level of Total European and Asian content, shown in Figure 15A, is an indicator that mestizaje occurred predominantly between Amerindian women and European men.

7.19. Ternary plots of genetic data

Ternary graphs of genetic data allow us to visualise three variables in a single graph without resorting to 3D diagrams. In our case, we use all markers but grouped by continental origin: Amerindian, Euro-Asian, and African.

Ternary graphs display all types of genetic tests from 10,795 individuals of both sexes in the Argentine Republic. Figure 15B shows data from autosomal, exosomal, mtDNA, and Y-STR/SNP genetic tests. The points are all clustered around the European-Amerindian axis, which could be interpreted as a highly admixed Argentina, where the mixture of Amerindian, European, Asian, and African genes has occurred at perhaps a unique level in the world. In the case of African genes, there is a dilution effect manifested in the autosomal marker, showing a concentration range between 3% and 24%, except for Córdoba, which shows 0%, and Jujuy, which shows 0.1%.

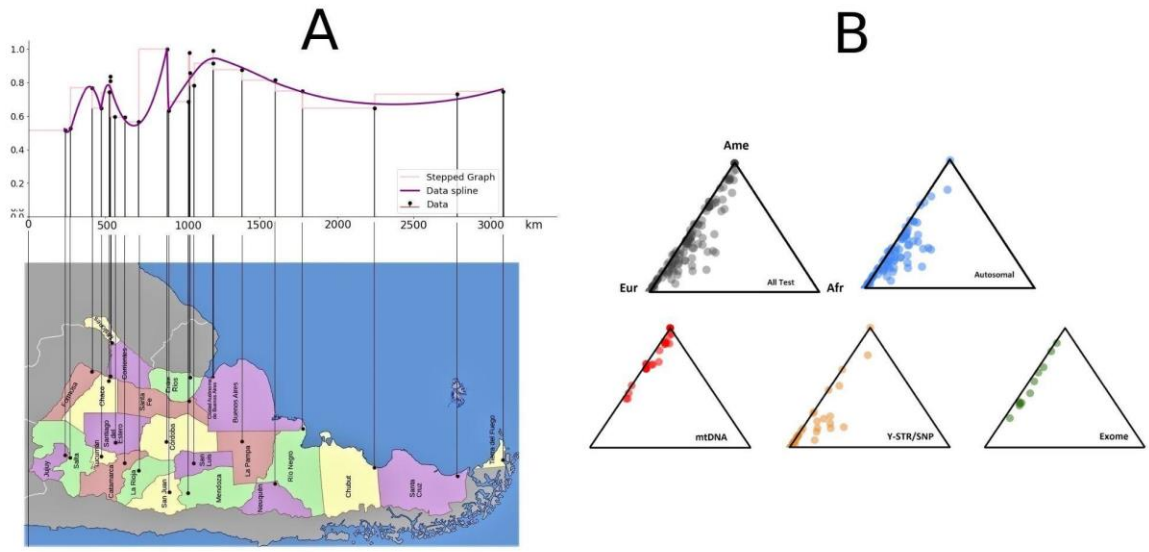

Figure 15. (A) Total european and asian genographic profile, include mtDNA, autosomal and Y-STR markers. Purple curve is a second order spline. (B) Ternary plots. High resolution Figure 15

7.20. Genetic content-law non-compliance relationship. linear correlation analysis

The linear correlation analysis of total Amerindian+African genetic content (TAA) versus each of the laws violated did not produce significant values except in three cases. Figure 16 shows a scatter plot of Property Tax Evasion versus TAA. The estimated linear regression line has an R^2=0.326 (CC=0.57), which, when excluding CABA with a much lower fiscal evasion value than the rest of the provinces, produces R^2=0.432 (CC=0.66).

A similar case occurs when comparing the Vertical Fiscal Imbalance against the total Amerindian+African genetic content with R^2=0.358 (CC=0.60), which, when excluding CABA with a much lower federal fund dependency value, produces R^2=0.453 (CC=0.67).

The last exception occurs when comparing the number of fatalities per ten thousand registered vehicles against the total Amerindian+African genetic content with R^2=0.305 (CC=0.55), which is a very poor correlation value for the hypotheses we want to demonstrate.

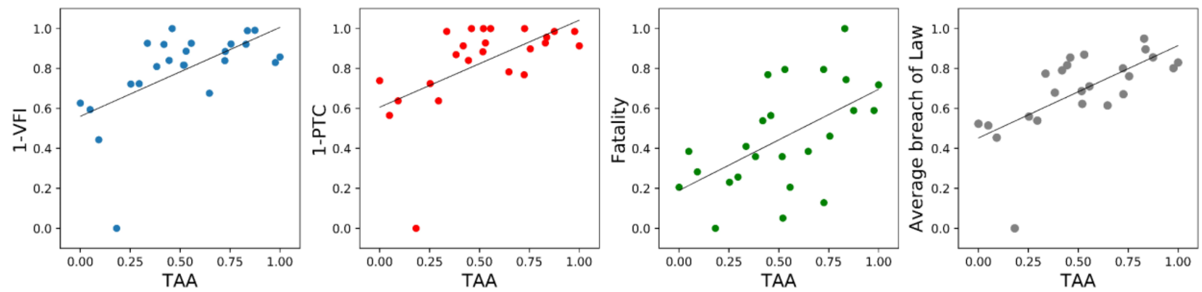

Figure 16. Scatter Diagram of Variables (1-VFI), (1-PTC), Fatality and Average breach of law versus TAA taking into account all provinces. Dot: data; Solid black line: estimated regression line. High resolution Figure 16

Table 5.
Correlation test to evaluate the association between TAA and Vertical Fiscal Imbalance, Property Tax Evasión, Fatalities rate /10^4 registered vehicles and the average of three items

All provinces		Vertical Fiscal Imbalance	Property Tax Evasión	Fatalities rate /10^4 registered vehicles	Average breach of law with three items
TAA	Pearson Correlation	0.592*	0.572*	0.555*	0.665*
	Sig. (2 tailed)	0.002	0.004	0.005	0.000
	N	24	24	24	24

*. Correlation is significant at the 0.01 level (2-tailed)

As observed in Table 5, the relationship between TAA and the three breaches of the law and the average of these three items is highly significant. In the first two cases, CABA appears as anomalous in a positive sense (much less evasion).

Figure 17 and Table 6 show the behaviour without CABA.

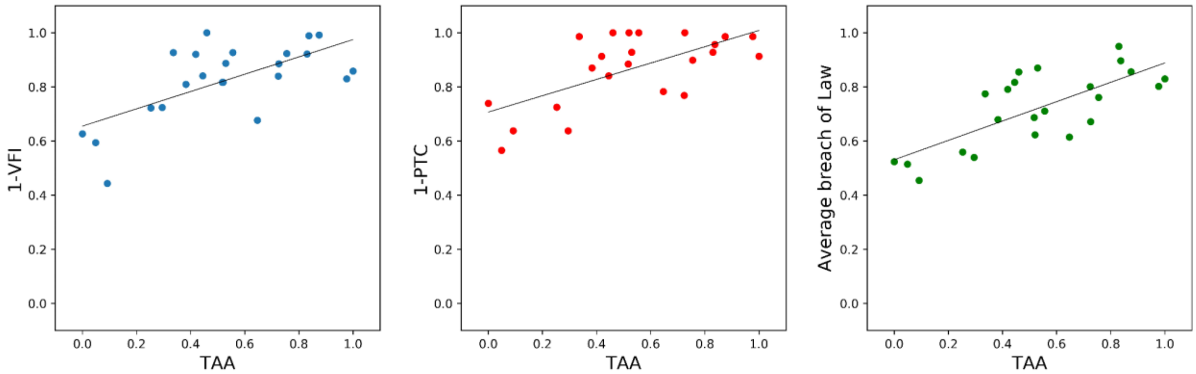

Figure 17. Scatter Diagram of Variables (1-VFI), (1-PTC), Fatality and Average breach of law versus TAA without considering CABA. Dot: Data, solid black line: Estimated regression line. High resolution Figure 17

Table 6.
Correlation test to evaluate the association between TAA and Vertical Fiscal Imbalance, Property Tax Evasión, Fatalities rate /10^4 registered vehicles and the average of three items without considering CABA

Without CABA		Vertical Fiscal Imbalance	Property Tax Evasión	Average breach of law with three items
TAA	Pearson Correlation	0.655*	0.650*	0.735*
	Sig. (2 tailed)	0.001	0.001	0.000
	N	23	23	23

*. Correlation is significant at the 0.01 level (2-tailed)

From the table above, it can be seen that both Pearson correlation coefficients and significance levels improve significantly.

The rest of the non-compliances show no statistically significant correlation with genetic data.

7.21. Latitude-law non-compliance relationship

7.21.1 Unregistered labour

As mentioned earlier, there is no statistically significant correlation between genetic content (TAA) and unregistered labour. However, when we analyse this non-compliance versus latitude (expressed in km and referred to parallel P1, see Table 2), it can be observed that the data are distributed around the estimated straight line, as shown in Figure 18.

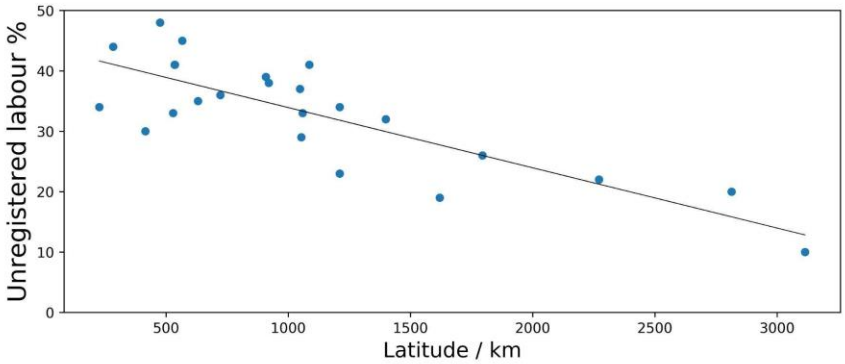

Figure 18. Scatter Diagram of Unregistered Labour versus Latitude. Dot: Raw data, solid black line: Estimated regression line. High resolution Figure 18

Table 7.
Correlation test to evaluate the association between latitude Unregistered labour

		Unregistered labour
Latitude	Pearson Correlation	-0.818**
	Sig. (2 tailed)	0.000
	N	24

**. Correlation is significant at the 0.01 level (2-tailed)

Table 7 shows a high correlation between latitude and unregistered labour. Figure 18 demonstrates that the level of non-compliance clearly increases toward the northern part of the country, where 19 out of the 24 provinces show a percentage higher than 25%, with a maximum of 48% in the province of Tucumán. To understand these values in a global context, it is helpful to know the percentages in the 27 countries of the European Union. In them, undeclared work in the private sector ranges between 7.1% in Germany and 27.3% in Poland. It is defined as the difference/discrepancy between the reported labour supply by workers and the reported use of labour by employers (Williams, 2017).

In the rest of the cases, the correlation between genetic content and non-compliance with this law does not show a significant value. However, when we analyse non-compliance versus geographical position, an apparent relationship is observed. This can be observed in detail in the supplementary material "*S8 Data without relationships*".

7.22. Cluster analysis

Variations in S (silhouette coefficient) with the number of clusters were analysed using the 24x20 matrix that characterises the provinces. Figure 19 shows how the mean silhouette coefficient varies with the number of clusters analysed. In this case, the optimal cluster number is three. This means that there are 3 (three) clusters, within which the provinces exhibit similar behaviours regarding the analysed characteristics.

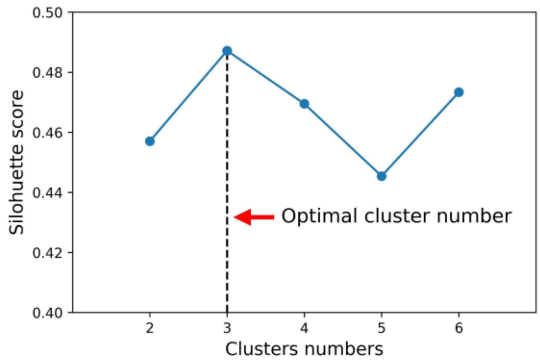

Figure 19. Variation of S with the number of clusters. [High resolution Figure 19]

Figure 20 illustrates how the provinces are grouped into the three clusters.

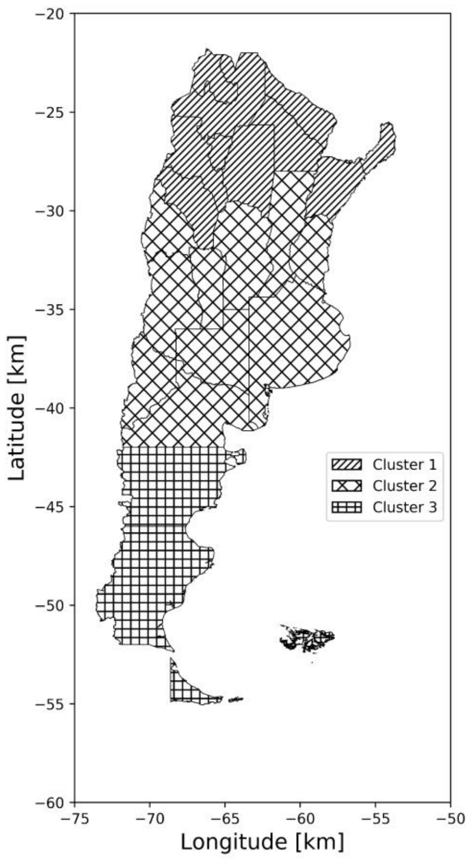

Figure 20. Cluster analysis with fuzzy c-means and k-means algorithms. [High resolution Figure 20]

From Figure 20, it can be observed that, through another approach, the same result as mentioned earlier was reached: the northern provinces exhibit similar geographic, genetic, and law non-compliance characteristics, as do the central or southern provinces of Argentina.

7.23. Radial diagrams

The radial diagrams presented in Figure 21 allow for a quick recognition of each province's behaviour regarding the analysed laws.

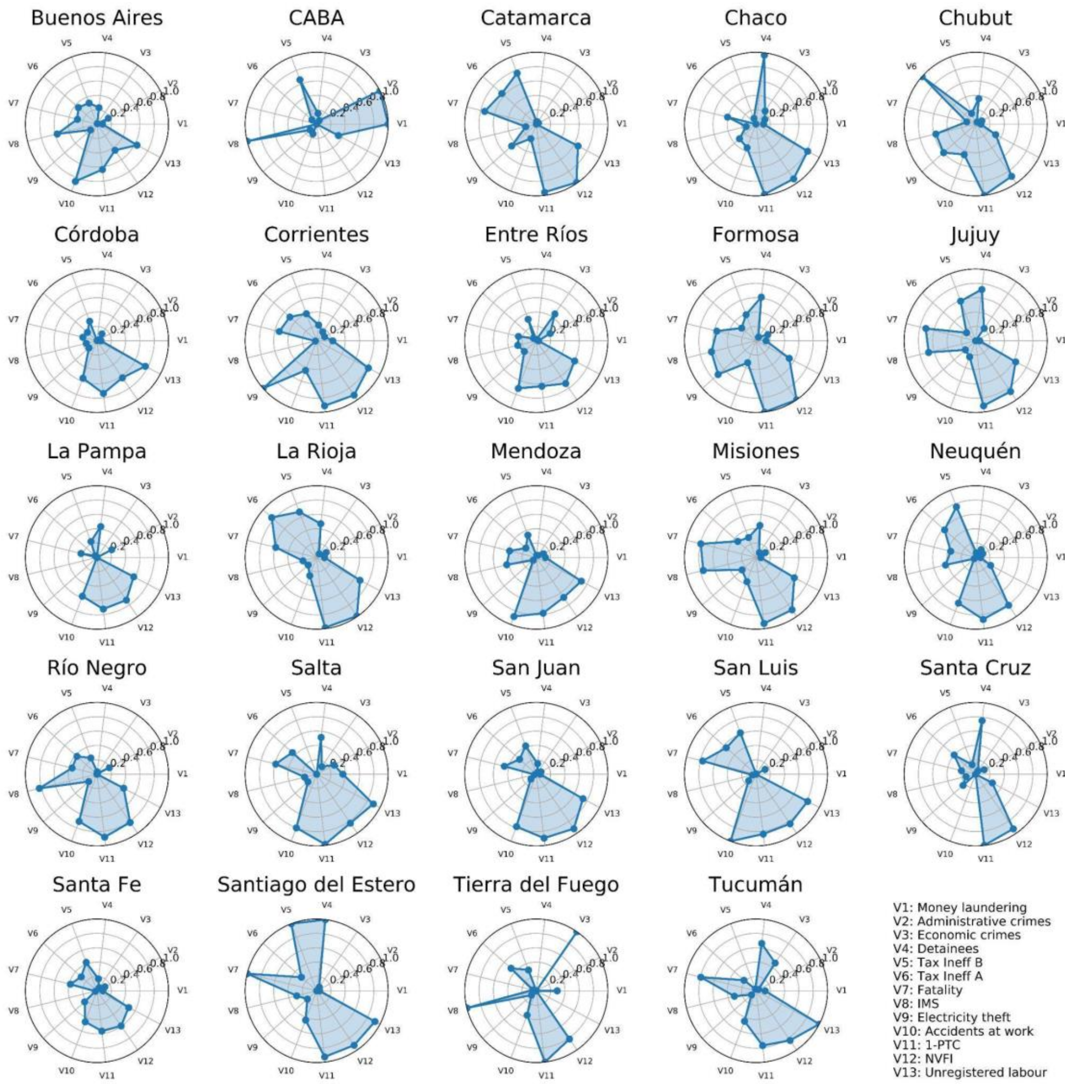

Figure 21. Radial graphs of non-compliance with laws by province. High resolution Figure 21

A radar or spider chart is a way to quickly and graphically visualise in two dimensions a set of three or more quantitative variables plotted on axes originating from the same point. In our case, the 13 law non-compliances for each province were represented on the radar chart, excluding lawsuits and health claims as they are indirect indicators. The radar chart allows for a single glance at all non-compliances per province, providing an idea of which ones are the most important or significant. It should be noted that, since the non-compliances are normalised, small values are located at the centre of the radar, while high values are on the periphery. For example, Santiago del Estero shows high non-compliance values for variables V4 (Detainees in police stations), V5 (Tax Inefficiency B - gross income-), V7 (Fatal victims of traffic accidents), and V11 (1-PTC), V12 (1-VFI), and V13 (Unregistered labour), although the highest non-compliance occurs in the first three.

8. Discussion

For the analysis of the obtained results, the underlying relationship between concepts must be considered. Figure 22 graphically shows this relationship.

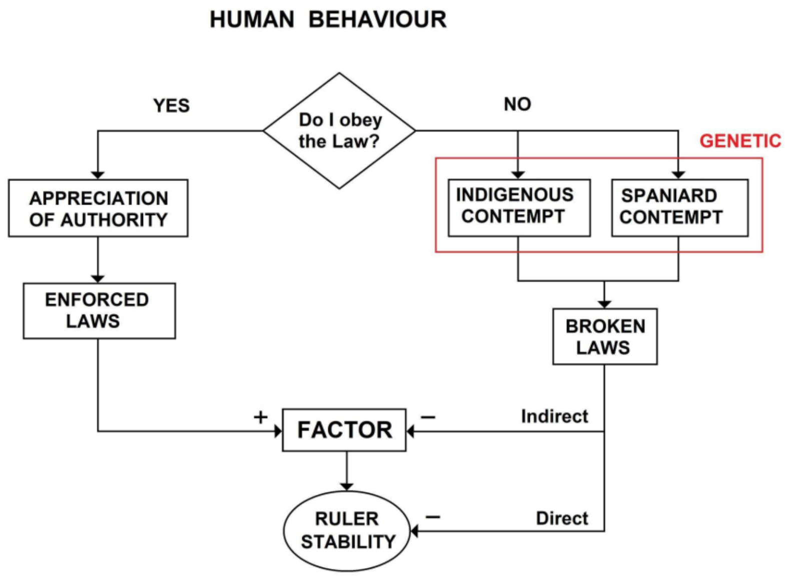

Figure 22. Relationship between contempts, genetic and ruler stability. High resolution Figure 22

Figure 22 shows that behind each violation of a law or regulation, there is behaviour related to either Hispanic or indigenous origins. Due to the extensive Amerindian, African, and Eurasian genetic mix, there are no clear boundaries between an indigenous or Spanish citizen. Instead, there is a genetic gradient that is analysed in this work. The figure also shows how these behaviours influence factors that could affect the stability of a ruler's position. Examples of such factors include Gross Domestic Product, the number of coups d'état, the number of political parties, the amount of unregistered work, the number of deaths in traffic accidents, or the number of military personnel in government positions.

This approach provides a comprehensive view of the problem, facilitating a better understanding of how the relationships between variable groups are analysed in this work and their connection to political stability. Throughout the analysis, statistically significant mathematical relationships emerged between certain groups of legal and genetic data, a correlation between unregistered work and geographic location, and similar social behaviour patterns in three groups of provinces.

Out of the 15 sets of laws analysed, a clear mathematical relationship was found in three cases: the Vertical Fiscal Imbalance, Property Tax Evasion, and Fatality rate per 10^4 registered vehicles, versus the total Amerindian + African genetic content. These relationships could indicate an apparent indigenous origin of contempt for authority, indirectly manifested through the violation of respective laws in each analysed group.

When analysing unregistered work versus genetic content, no relationship was found between these variables. However, replacing genetic content with geographic location revealed a high linear correlation with a Pearson coefficient of -0.818. The lack of correlation with genetic content and the high correlation with distance suggests two main causes. The first is that it is a widespread custom independent of Hispanic or indigenous origin, as no correlation was found between the level of unregistered work per province and the Amerindian, Eurasian, and African genetic contents. The second is a notable decrease in unregistered work as the distance from north to south of the country increases. A possible explanation for this behaviour is that unregistered work, historically derived from indigenous slave labour (Felice and Ruiz, 2024), is a custom more deeply rooted in the north due to a higher indigenous population density at the beginning of the conquest and colonisation of America in the 16th and 17th centuries. In the south, this custom is much more recent, dating back to the late 19th and early 20th centuries.

In the last group of data, where no acceptable correlation was found between genetics and law violations, cluster analysis identified three zones in Argentina that might share common social behaviours. While the highest proportion of traffic accidents with fatalities and detainees in police stations is found in the northern provinces, workplace accidents occur in the central provinces. However, some behaviours, such as unregistered work or property tax evasion, are widespread throughout the country, indicating that these toxic behaviours could be independent of genetics or proximity to northern Argentina.

The information analysed in this work can be condensed into conclusions that are based on the results of mathematical analysis of statistical data collected from official pages of the Argentine Republic and from works published in journals of significant impact in the scientific community.

8.1. Key findings

The conclusions obtained from different mathematical and statistical tools are presented as statements that have been numbered for clarity.

For instance, based on Figure 1, it can be concluded that:

1. *The provinces showing maximum non-compliance values are in the northern and southern regions of the country and in CABA.*

Regarding this statement, it should be noted that while Jujuy and Misiones do not show a maximum non-compliance in at least one of the analysed laws, they do present non-compliance indices above 90%.

This distribution is not random, as shown by the probability calculation, the probability of the occurrence of the non-compliance configuration shown in Figure 1 is extremely low (0.00173).

The map of overall non-compliance averages provides further detail of what Figure 1 illustrates.

2. *On average, the highest non-compliances mainly occur in the northern and southern provinces of the country and to a lesser extent in the central provinces. CABA is the population showing the highest number of non-compliances.*

Based on the ternary plots shown in Figure 15B, it can be concluded that:

3. *Argentina has people with a range of skin tones, where a unique blend of Amerindian, European, Asian, and African genes has come together.*

This statement is supported by Figures 14A, 14B, and 15A, which show that: (a) the total Amerindian + African genetic content, including mtDNA and autosomal, is remarkably high in Argentina from south to north and from east to west in the 21st century, (b) the total Y-STR European + Asian genetic content is also remarkably high throughout Argentina, and (c) the autosomal Amerindian + African content shows an average value above 25%.

Figures 16 and 17 show that:

4. *There is a high linear correlation between the total Amerindian and African genetic content and the violation of three laws (1-PTC, 1-VFI and Fatality).*

There is no coincidence or chance when establishing a linear relationship between two variables that exhibit a high correlation coefficient. The underlying causality in this strong mathematical relationship has a cultural origin and is substantiated with statements from the analysis in other papers (Felice and Ruiz 2024; Felice and Ruiz, 2025).

Figure 18 and those shown in the supplementary material "S8 Data without relationships" show that:

5. *There is a relationship between the breach of laws and the geographic location of provincial capitals, and this relationship is particularly linear in the case of 'unregistered workers'.*
6. *There is a significant baseline level of non-compliances in all provinces. This implies that, from a social perspective, all inhabitants of the Argentine Republic violate laws with a significant baseline value.*

In the cases of non-compliance with Money laundering, Economic-financial crimes, Electricity theft, Detainees in police stations, and Traffic infractions, there is an increase in the number of non-compliances in the northern and southern regions of the country, while in Accident at work, Lawsuits, and Health claims, a significant increase in these non-compliances is observed in the central region of the country, compared to the northern and southern regions. CABA is an exceptional case that shows the highest number of non-compliance peaks compared to all provinces.

Cluster analysis does not provide information on the optimal number of clusters to be used to group provinces with similar characteristics. For this reason, the silhouette coefficient is calculated, which is a measure of how similar provinces are within a cluster and how different clusters are from each other. For cluster analysis, only non-compliances that do not show explicit linear correlation with either genetic aspects or geographic coordinates are considered. This excludes the possibility that the results of cluster analysis are biassed by the shown linear relationships.

Based on the results obtained through cluster analysis (see Figure 20), it can be concluded that:

7. *The northern provinces (NOA and NEA) forming cluster 1, the southern provinces (Chubut, Santa Cruz, and Tierra del Fuego and South Atlantic Islands) forming cluster 3, and the rest of the central provinces forming cluster 2, exhibit homogeneity within their respective clusters and heterogeneity with respect to the other clusters regarding the studied characteristics (genetic material, geographical location, and levels of non-compliance with laws that did not show explicit linear correlation).*

The results of cluster analysis are consistent with those shown in non-compliance scatter plots.

Finally, radial diagrams allow for a quick visual representation of the non-compliance level, per province, of all studied laws, through simple inspection.

8.2. Ethical considerations and societal implications

Given the sensitive nature of this research, it is important to consider the ethical implications and potential societal impacts:

1. **Data Privacy**: We have always used public domain data, including public databases and scientific publications also in the public domain. In each analysed work, the respective authors adhered to proper handling of information in terms of ethics and confidentiality. As examples, in studies where samples for genetic analysis are obtained, it is stated that informed consent was given by all participants. The protocol and procedures employed were reviewed and approved by the review committee of various organisations such as universities, hospitals, and governmental institutions. Informed consent was given by all persons prior to inclusion. In other cases, the samples were obtained as requirements for Civil and Criminal cases, to establish biological

relationships or identification in forensic cases, and consequently, the sample collection was carried out by court order. For the analyses, the samples were coded to ensure anonymity.

2. **Stigmatization and discrimination:** To avoid stigmatisation and discrimination, all genetic and legal information is classified only by province in Argentina and by genetic content, without making any individual or ethnic group distinctions. However, it is recognized that the results can be misinterpreted or used inappropriately. To mitigate this risk, several measures have been implemented:

    a. **Contextualization of Results:** It has been emphasised in our analyses and discussions that the results must be interpreted within a broader socio-cultural context. This means that the results of the study should not be interpreted to suggest a direct and immutable causal relationship between the genetic makeup of individuals and their likelihood of breaking laws. It is cautioned that the results should not be oversimplified to imply that genetic makeup alone determines behaviour. It should be clear that our results do not refer to individuals; they refer to population groups differentiated only by province.

    b. **Objectivity of Results:** All presented results come from mathematical and statistical analyses, constituting an objective description of reality. That is, the findings in the study are based on rigorous mathematical and statistical methods. These methods ensure that the results are unbiased and not influenced by personal opinions or subjective interpretations. This statement emphasises that the conclusions drawn are grounded in empirical data and quantitative analysis, providing a factual and impartial representation of the relationships observed in the study.

    In other words, although the results are objectively accurate, it is crucial to interpret them within a broader sociocultural context. This means that although the observed correlations are real and statistically significant, they must be understood as part of a more complex network of social, cultural and historical factors. This avoids simplistic or deterministic interpretations that could lead to erroneous or stigmatising conclusions.

3. **Policy and Public Perception**: The conclusions of this paper cover highly sensitive social aspects for Argentine society. Our results can be seen as a set of data and suggestions for political decision-making, which must be planned with great caution and based on consensus. Participants should come from the government, non-governmental organisations, sociologists, statisticians, scientists, psychologists, historians, economists, and all the necessary specialists to design public policies. It is essential to consider ethnicity in the planning of such public policies. An extreme example is the measures to be taken to reduce suicide rates in our indigenous communities. It is also important to consider ethnicity in social plans and resource allocation, implement cultural measures that strengthen Argentine multiculturalism and eliminate pigmentocracy, integrate non-European contributions into Argentine history, socially recognize indigenous peoples alongside the history of our European and African immigrants, and integrate the American and African worldviews with the Eurasian to create a culture that includes all Argentine citizens.

4. **Further Research and Ethical Standards**: Future studies should continue to explore these complex relationships with a strong ethical framework. Multidisciplinary approaches involving sociologists, geneticists, and ethicists can provide more comprehensive insights and mitigate potential negative impacts.

9. Conclusions

The mathematical relationships found between legal and genetic data sets of the inhabitants of Argentina are presented in this work, without taking into account the underlying causality in each relationship. Profound social implications could be attributed to the findings, such as the assignment of stereotypes or toxic behaviours to certain ethnicities or social groups. However, the results must be interpreted in a broader context that includes, in addition to the mathematical relationships, the historical origins of contempt for authority, which are measured by quantifying non-compliance with laws and regulations. Stigmatisation or discrimination against specific population groups can be avoided through this type of contextual analysis.

This work presents a rigorous mathematical and statistical analysis of collected data, following the guidelines of inclusion and exclusion defined in supplementary materials "S5 Legal selection" and "S6 Genes selection." The analysis reveals that some sets of legal and genetic data exhibit linear relationships with high statistically tested correlation coefficients, while others do not show statistically significant relationships. For example, Vertical Fiscal Imbalance, Property Tax Evasion, and Fatality are linearly linked to the total Amerindian+African genetic content (TAA) with correlation coefficients ranging between 0.650 and 0.735 when excluding CABA from the analysis due to its anomalous behaviour.

Since the visual analysis of non-compliance maps shows an apparent relationship between genetic and legal data that cannot be validated through Pearson analysis due to the lack of a linear pattern, additional geographical data was introduced to reveal underlying connections. The geographical coordinates of the capitals provide a different perspective, as the representation of genetic and legal data alone is not sufficient to identify solid mathematical relationships. For instance, a linear relationship is observed between unregistered labour and the latitudes of the capitals, with a Pearson coefficient of -0.818, indicating that the number of unregistered workers increases linearly from south to north, ranging from 12.8% in Tierra del Fuego to 41.6% in Jujuy.

Geographical maps alone have limitations, as visual information alone does not provide a firm basis for establishing clear patterns of social behaviour. Despite the inability to determine specific mathematical relationships, scatter plots show that certain provinces exhibit higher levels of specific non-compliances. For example, in the northern provinces of Argentina, levels of non-compliance with Money Laundering, Economic-Financial Crimes, Electricity Theft, Detainees in Police Stations, and Traffic Infractions are higher than in the central or southern provinces, while levels of non-compliance with Workplace Accidents, Lawsuits, and Health Claims are more significant in the central provinces.

Multidimensional cluster analysis helps to overcome the limitations of the current and previous studies. From the cluster analysis with legal data that do not exhibit a mathematical relationship with genetic or geographical data, consistent results are found with those observed from scatter plots. It is found that the northern provinces of Argentina (NOA and NEA) exhibit similar behaviours regarding genetic material, the level of non-compliance with certain laws, and geographical proximity, forming cluster 1. The provinces of Chubut, Santa Cruz, Tierra del Fuego, and the South Atlantic Islands form cluster 3 due to their shared characteristics. Finally, the central provinces are grouped in cluster 2. Despite these groupings, all provinces exhibit toxic social behaviours regarding law compliance with high baseline levels, regardless of their geographical position or genetic content.

10. Limitations and future research

While this study provides valuable insights, several limitations must be acknowledged. The potential for underreporting or misreporting of legal violations could affect the accuracy of the legal data. Additionally, genetic data may be limited by the availability and representativeness of the samples used in the analysed publications. Potential confounding variables, such as socio-economic factors, cultural influences, and historical events, were not directly measured but could significantly influence both legal compliance and genetic composition.

Future research should aim to incorporate these factors to provide a more comprehensive understanding of the observed correlations. Expanding the dataset to include more detailed socio-economic variables and historical contexts could help clarify the relationships between genetic composition, legal compliance, and geographical factors. Moreover, employing other statistical methods, such as multivariate regression analysis and machine learning techniques, could offer deeper insights into the complex interactions between these variables.

In summary, while this study identifies important patterns and correlations, it also highlights the need for more nuanced and multifaceted approaches to fully understand the dynamics of legal compliance and genetic composition across Argentine provinces.

Acknowledgments

This work was supported by institutional funds from Facultad de Ciencias Exactas y Tecnología de la Universidad Nacional de Tucumán, Laboratorio de Medios e Interfases and Instituto Superior de Investigaciones Biológicas (UNT-CONICET).

Supplementary material

All material is either here or in the QR code below

- *S1 Genetic Papers:* A document containing detailed descriptions and comments on all bibliographic references from which genetic analysis data were extracted.
- *S2 Genetic Data:* A spreadsheet containing percentage values of genetic content differentiated by their American, Eurasian, and African continental origins. Each percentage set in the spreadsheet tabs includes bibliographic references, sampling location, geographic position, and marker type. It also includes how haplogroups were grouped continentally for certain scientific works and how data were grouped for ternary plots.
- *S3 Broken Laws:* A spreadsheet containing numerical percentage data by province for claims or violations related to non-compliance with different sets of laws or norms.
- *S4 Non-Technical Losses:* A spreadsheet containing how non-technical losses of electricity were estimated for each province, indicating how the electricity distribution companies in Argentina were grouped.
- *S5 Legal selection:* contains a detailed analysis of how each indicator is selected, the inclusion and exclusion criteria, and the methods used for their quantification.
    - *S5-1 Agro tax evasion*
    - *S5-2 Newspapers*
- *S6 Genes selection:* contains an analysis that justifies the inclusion of genetic information as a means to demonstrate the hypotheses, as well as the process of selecting the specific genetic markers.
- *S7 Genetic and legal:* contains the legal and genetic datasets used for geographical visualisation and mathematical analyses.

- *S8 Data without relationships:* contains the legal and genetic datasets that do not show statistically significant relationships with the geographical location of the capitals' latitudes.
- *S9 Cluster Analysis:* contains a complete and detailed explanation of the algorithms and data used in cluster analysis, including the silhouette coefficient, which is a method for determining the optimal number of clusters for the study.

References

ADEERA (2019). *Informe Estadístico Anual*. Retrieved from https://www.adeera.org.ar/

Avena, S., Di Fabio Rocca, F., Postillone, M. B., & Dejean, C. (2013). ¿Existió el crisol de razas en Argentina? Una respuesta desde la antropogenética. In *La complejidad después de Babel* (pp. 281-314).

Banco Central de la República Argentina (BCRA). (2019). *Informe de inclusión financiera. Primer Semestre de 2019*. Retrieved from: https://www.bcra.gob.ar/Pdfs/PublicacionesEstadisticas/iif0119.xlsx

Benencia, R., & Canevaro, S. (2017). Migración boliviana y negocios. De la discriminación a la aceptación. La salada como fenómeno social. *REMHU: Revista Interdisciplinar da Mobilidade Humana*, 25, 175-196. https://doi.org/10.1590/1980-85852503880004910

Castro, L., Díaz Frers, L., Alfieri, A. C., & Bovino, A. (2014). *Desafíos y potencial del impuesto inmobiliario en la Argentina. Documento de Trabajo*, (120).

Cavalli-Sforza, L. L., Menozzi, P., & Piazza, A. (1994). *The History and Geography of Human Genes*. Princeton University Press.

Comas, D., Calafell, F., Mateu, E., Pérez-Lezaun, A., Bosch, E., & Bertranpetit, J. (2004). Trading genes along the silk road: mtDNA sequences and the origin of Central Asian populations. *American Journal of Human Genetics*, *63*(6), 1824-1838. https://doi.org/10.1086/302133

Diamond, J. (1997). *Guns, Germs, and Steel: The Fates of Human Societies*. W.W. Norton & Company.

Di Gresia, L. (2003). *Impuesto sobre los Ingresos Brutos: Análisis Comparativo de su Evolución y Perspectivas*. Documentos de Federalismo Fiscal.

Dirección Nacional del Observatorio Vial, Dirección de Estadística Vial. (2021, July). *Road Safety Statistical Yearbook: 2019*. Retrieved from https://www.argentina.gob.ar/sites/default/files/2018/12/ansv_ov_anuario_estadistico_2019_final.pdf

Efficiency of power plants. (2024, February 14). Rate of electricity T&D losses, *DATA MAPPER, Power sector*. Retrieved from https://biee-cepal.enerdata.net/en/datamapper.

Felice, C. J., & Ruiz, G. A. (2023a). Argentine Chaotic Term Length Series in an American Historical Context: Part 1: Time Domain Analysis. *Journal of Advanced Research in Social Sciences*, *6*(2), 1–21. https://doi.org/10.33422/jarss.v6i2.932

Felice, C., & Ruiz, G. A. (2023b). Argentine Chaotic Term Length Series in an American Historical Context: Part II: Complex System Analysis. *Journal of Advanced Research in Social Sciences*, *6*(3), 1–34. https://doi.org/10.33422/jarss.v6i3.937

Felice, C. J., & Ruiz, G. A. (2024). From Indigenous Resistance to the Contemporary Argentine Contempt for Authority. *Journal of Advanced Research in Social Sciences*, 7(4), 1–32. https://doi.org/10.33422/jarss.v7i4.1301

Felice, C. J., & Ruiz, G. A. (2025). The Behavior of Spanish Conquerors and Colonizers in America Is a Root Cause of Political Instability in Argentina. *Journal of Advanced Research in Social Sciences*, *8*(1), 1–25. https://doi.org/10.33422/jarss.v8i1.1300

Guy, D. J. (1978). The Rural Working Class in Nineteenth-Century Argentina: Forced Plantation Labor in Tucuman. *Latin American Research Review*, *13*(1), 135-145. https://doi.org/10.1017/S0023879100030715

INDEC (Instituto Nacional de Estadísticas y Censos de Argentina). (2023).

Leveau, C. M. (2013). Spatial variations in motorcycle registrations and the mortality of motorcycle users due to traffic injuries in Argentina. *Salud colectiva*, *9*(3), 353-362. https://doi.org/10.1017/S0023879100030715

Nassif, S. (2015). Condiciones de vida y subjetividad de los obreros de la agro-industria azucarera tucumana hacia la década de 1960. *Revista Nuestro NOA*, (6).

Olson, S. H., Gangnon, R., Silveira, G. A., & Patz, J. A. (2019). Deforestation and malaria in Mâncio Lima County, Brazil. *Emerging Infectious Diseases*, *16*(7), 1108-1115. https://doi.org/10.3201/eid1607.091785

Parra García, H. (2021). Ensamblajes transnacionales de la colectividad boliviana en Buenos Aires. *Diálogo Andino*, *(66)*, 337-349. https://doi.org/10.4067/S0719-26812021000300337

Resano, M., & Moral, P. (2018). Mestizaje genético en las poblaciones humanas actuales de Argentina: Revisión. *Antropo*, *39*, 77-96.

Rossignolo, D. (2017). La ineficiencia tributaria en las provincias argentinas: el impuesto sobre los ingresos brutos. *Cuadernos de Economía*, *36*(70), 95-135. https://doi.org/10.15446/cuad.econ.v36n70.55025

Rosso L. J. y Cía. (1916). *Tercer Censo Nacional: Año 1914, Ordenado por Ley 9108 bajo la presidencia del Dr. Roque Saenz Peña*. Retrieved from http://www.estadistica.ec.gba.gov.ar/dpe/Estadistica/censos/C1914-T1.pdf

Saiegh, S. M., & Tommasi, M. (1999). Why is Argentina's fiscal federalism so inefficient? Entering the labyrinth. *Journal of Applied Economics*, *2*(1), 169-209. https://doi.org/10.1080/15140326.1999.12040535

Snyder, J. P. (1987). *Map projections—A working manual*. US Government Printing Office. https://doi.org/10.1080/15140326.1999.12040535

SRT (Superintendencia de Riesgos del Trabajo). (2019). *Informe anual de accidentabilidad laboral 2019*. Departamento de Estudios y Estadísticas.

Tercer Censo Nacional (1916). Realizado el 1ro de Junio de 1914, Ley Nro 9108 bajo Presidencia Roque Sáenz Peña, ejecutado bajo presidencia de Victorino de la Plaza. Buenos Aires, Talleres Gráficos de L.J. Rosso y Cía.

Wang, S., Lewis, C. M., Jakobsson, M., Ramachandran, S., Ray, N., & Bedoya, G. (2013). Genetic variation and population structure in native Americans. *PLoS Genetics, 3*(11), e185. https://doi.org/10.1371/journal.pgen.0030185

Weinberg, M., & Mercolli, P. H. (2017). Sweet Death: Indigenous Labour Exploitation in the San Martín de Tabacal Sugar Mill, Salta, Argentina. *Cultural Studies*, *31*(1), 70-92. https://doi.org/10.1080/09502386.2016.1167929

Williams, C., Bejaković, P., Mikulić, D., Franić, J., Kedir, A., & Horodnic, I. A. (2017). An evaluation of the scale of undeclared work in the European Union and its structural determinants: Estimates using the labour input method. *Publications Office of the European Union*. https://doi.org/10.2139/ssrn.3092080