Cuadernos de historia económica

The persistence of segregation in education: Evidence from historical elites and ethnic surnames in Colombia

Por:
Juliana Jaramillo-Echeverri
Andrés Álvarez

Núm. 58
Junio, 2023

BANCO DE LA REPÚBLICA
CENTRO DE ESTUDIOS ECONÓMICOS REGIONALES (CEER) - CARTAGENA

The persistence of segregation in education: Evidence from historical elites and ethnic surnames in Colombia

Juliana Jaramillo-Echeverri* Andrés Álvarez‡

The series **Cuadernos de Historia Económica** is a publication of Banco de la República in Cartagena. The opinions contained in this document are the authors' sole responsibility and do not commit Banco de la República or its Board of Directors.

Abstract

Inequality in access to high-quality education can hinder the ability of education to promote intergenerational mobility. Looking at the case of Colombia, one of the most unequal and least mobile countries in Latin America, we evaluate whether contemporary differences in access to high-quality education have deep roots in the past. We use several past and contemporary sources to define social status attributes for several historical groups. Assuming that sufficiently rare surnames are part of the same extended family, we trace dynasties of indigenous, *encomenderos* (Spaniard colonial officers), 19th century slave-owners, and members of different educational, social, and business elites of the 17th, late 19th and the beginning of the 20th centuries. Using microdata from administrative sources, we provide evidence of social segregation in education and test if the historical status of each social group is associated with access to disadvantageous or privileged educational institutions. The results show that the original social status of the historical groups is highly associated with their contemporary performance in educational outcomes. We explore assortative mating as a mechanism for perpetuating segregation in education. We find evidence of homogamy within the historical elites and ethnic surnames. We conclude that the educational system in Colombia reproduces patterns of social exclusion rooted in the past.

Keywords: Education, Segregation, Persistence, Assortative mating, Colombia
JEL Classification: O15, D63, I24, J15, J12, N36

*Junior Researcher, Center for Regional and Economic Studies. Contact info: jjaramec@banrep.gov.co
†Facultad de Economía, Universidad de Los Andes. Contact info: ca.alvarez967@uniandes.edu.co

‡We thank Catalina Bernal and Jorge Guerra for excellent research assistance. We are also in debt with Karina Acosta, Jhorland Ayala, Jaime Bonet, Gregory Clark, Mario Cuenda, Neil Cummins, Sara Horrell, Javier Mejia, Chris Minns, Aurelius Noble, Fabio Sanchez, Eric Schneider, and the seminar participants several the workshops and conferences for their insightful feedback. We thank Leopoldo Fergusson and Sergio Florez for their generosity in sharing the data from their research. All errors and omissions are ours.

Persistencia de la segregación en educación: Evidencia a través de apellidos de élites históricas y grupos étnicos en Colombia

Juliana Jaramillo-Echeverri[*] Andrés Álvarez[†‡]

Resumen

La desigualdad en el acceso a educación de alta calidad puede obstaculizar el papel de la educación como motor de movilidad social. Estudiando el caso de Colombia, uno de los países más desiguales y menos móviles del América Latina, nuestro objetivo es evaluar si las diferencias contemporáneas en el acceso a educación de alta calidad tienen sus raíces en el pasado. Con fuentes históricas y contemporáneas definimos atributos de estatus social de varios grupos históricos. Asumiendo que los apellidos suficientemente raros son parte de la misma familia extensa, seguimos dinastías de indígenas, encomenderos, dueños de esclavos miembros de diferentes élites educativas, sociales y empresariales de los siglos XVII, finales del siglo XIX y principios del siglo XX. Usando fuentes administrativas evaluamos si el estatus histórico de cada grupo social está asociado con el acceso a instituciones educativas privilegiadas. Los resultados muestran que el estatus social original de los grupos históricos predice el acceso a educación de alta calidad. Los grupos étnicos continúan siendo segregados de la educación de alta calidad contemporánea. Mientras que, entre más antigua es la élite más probabilidad hay de que converja a la media en estatus social. Además encontramos evidencia de homogamia contemporánea dentro de las élites históricas y los apellidos étnicos. Los resultados permiten concluir que el sistema educativo en Colombia reproduce patrones de exclusión social que están arraigados en el pasado.

Palabras clave: Educación, Segregación, Persistencia, Homogamia, Colombia
Clasificación JEL: O15, D63, I24, J15, J12, N36

[*]Investigadora Júnior, Centro de Estudios Económicos Regionales. Contacto: jjaramec@branrep.gov.co
[†]Facultad de Economía, Universidad de Los Andes. Email: ca.alvarez967@uniandes.edu.co
[‡]Agradecemos profundamente a Catalina Bernal y Jorge Guerra por su excelente asistencia en la investigación. También estamos en deuda con Karina Acosta, Jhorland Ayala, Jaime Bonet, Gregory Clark, Mario Cuenda, Neil Cummins, Sara Horrell, Javier Mejia, Chris Minns, Aurelius Noble, Fabio Sanchez, Eric Schneider y con los participantes de varios seminarios y conferencias por sus valiosos comentarios. Todos los errores y omisiones son nuestros.

1 Introduction

Colombia is among the most unequal and stratified countries in Latin America. According to several contemporary studies, Colombia has one of the highest intergenerational income elasticity and one of the highest Gini coefficients, and mobility in the country is lower in the upper and lower parts of the socioeconomic distribution.[1] Although the literature has pointed to education as a key mechanism to close social gaps (Torche, 2014; Behrman, Gaviria, & Szekely, 2001), several studies show that the educational system in the region reproduces patterns of social exclusion, which deepens the gap between social classes (Behrman, Gaviria, & Székely, 2003; Marshall & Calderon, 2006).

Recently, Cardenas, Fergusson, and Garcia-Villegas (2021) studied in dept what they call *educational apartheid* in Colombia. They provide evidence of contemporary social segregation in access to educational institutions and show that access to school is one of Colombia's most accurate markers of social status. According to Fergusson and Flórez (2021a) while children from households in the lowest socioeconomic level represent more than 40% of students who complete secondary education, this group barely represents 10% of those who access tertiary education. Conversely, children from households of the highest socioeconomic level represent less than 5% of students who complete secondary education but represent more than 50% of those who access tertiary education.[2] Similarly, Duarte, Bos, and Moreno (2012) show that, in Colombia, there are high inequalities in students' academic results associated with the socioeconomic level of their families and the type of school they attend (public vs. private). This gap is more important between schools than within them, which denotes high segregation of Colombian schools according to the socioeconomic level of the students.

This research aims to evaluate whether social segregation in education today has deep roots in the past. There are several reasons to believe social status has persisted in Colombia since colonial times. First-pass evidence suggests that the rigid social structure of the Spanish colonial regime is reflected in contemporary Colombia, where Indigenous and Afro-Colombians appear at the bottom of the social ladder, with the worse socioeconomic outcomes. Empirically, the literature has shown that the colonial regime set the type of institutions that we

[1] In 2019 the World Bank reported a Gini index for Colombia of 51.3 and the OECD report of 2018 shows that in Colombia it would take at least 300 years for offspring of low-income families to reach the mean income. See also Narayan et al. (2018); OECD (2018); Angulo, Azevedo, Gaviria, and Páez (2012); Alvaredo and Londoño Velez (2013)

[2] See Figure 9 in Cardenas et al. (2021). The socioeconomic level corresponds to a stratification system used to target cross-subsidies for the payment of utilities in Colombia.

expect to be persistent and that extractive institutions such as the *encomienda* and slavery had negative long-term effects at the national and sub-national levels (García-Jimeno, 2005; Acemoglu, Bautista, Querubin, & Robinson, 2007; Acemoglu, García-Jimeno, & Robinson, 2012; Maloney & Valencia Caicedo, 2016; Kalmanovitz, 2001; Alvarez, 2018). Still, the effects of these institutions on access to education have not been measured yet.

Due to the lack of reliable longitudinal data, it has been challenging to show if the contemporary educational apartheid has been persistent over time. However, using rare surnames and their relative representation in socioeconomic outcomes has shown to be a very productive way to overcome the limitations of scarce long–term longitudinal data.[3] We study social groups from different historical periods and ethnic backgrounds linked with a clear position on the social ladder in the past to evaluate their position in the educational system today.[4] Our analysis is based on using rare surnames to follow multiple generations and take advantage of the availability of rich contemporary administrative data–sets and historical information (Clark et al., 2015; Guell, Mora, & Telmer, 2013; Jaramillo-Echeverri, Álvarez, & Bro, 2021). We also exploit the official rule of registering paternal and maternal surnames inherited from the Spanish naming customs to describe the marriage patterns of our historical groups in the educational system. Following Chiappori, Dias, and Meghir (2020), we measure assortative mating in the marriage market as another measure of segregation in the educational system.

Our results sheds light on the origins of social segregation in the educational system. We find significant differences in access to contemporary education between historical groups, corroborating educational segregation's deep roots. In other words, group affiliation accurately predicts access to education. Historically excluded social groups (Afro-Colombians and Indigenous) show the lowest social status in all the educational categories we use to measure it. These results indicate that today, like in the colonial period, this segregation occurs at the top, from which ethnic groups are absent. However, we uncover different patterns of persistence for the elite groups. The encomenderos of the 17$^{\text{th}}$ century and the slave

[3] A whole branch of literature about social mobility and persistence has been built around using surnames. For an extensive review of name-based methods, see Santavirta and Stuhler (2020). Some relevant examples of fruitful applications of these methods in different historical and geographical contexts are summarised in Clark (2014) and Clark, Cummins, Hao, and Diaz-Vidal (2015).

[4] Our historical groups are: Indigenous pre–Hispanic groups, 17$^{\text{th}}$ century Spaniard colonial officers (*encomenderos*), African enslaved people, members of privileged families with access to higher education in the 17$^{\text{th}}$ and late 18$^{\text{th}}$ century, slaves owners in 1851, and members of different social and business elites of the late 19$^{\text{th}}$ and the beginning of the 20$^{\text{th}}$ centuries.

owners of the 19th century seem to have lost their high-status position. On the contrary, the elites related to the emerging bourgeoisie of the end of the 19th century present strong persistence of their privileged high status. The social status of those families who accessed high quality and high prestige education during the 17th to the early 19th centuries show weaker persistence in comparison to the post-1870s elite. We conclude that the educational system in Colombia reproduces patterns of social exclusion rooted in the past and partially reflects the persistence of colonial segregated social structures. At the sub-national level, we find that most regions of the country follow a similar pattern to that found at the national level. Historical elite groups have had privileged access to high-quality education, while ethnic groups have been systematically absent.

Many contributions point to the persistence of colonial institutions in Latin America and their effects on contemporary economic performance (Sokoloff & Engerman, 2000; Acemoglu, Johnson, & Robinson, 2005; Williamson, 2015; Dell, 2010; Valencia Caicedo, 2019; Bonet & Roca, 2007). For Colombia, in particular, there are several analyses at the sub-national level, showing the importance of certain forms of colonial institutions as determinants of long-term differences in development in different regions of the country (García-Jimeno, 2005; Acemoglu et al., 2007, 2012; Maloney & Valencia Caicedo, 2016; Kalmanovitz, 2001; Alvarez, 2018). Most of these studies conclude that the presence of extractive institutions based on forced labour was detrimental to the long-term development of the regions. For example, Acemoglu et al. (2012) find the presence of gold mines in the 17th and 18th centuries is associated with contemporary lower school enrolment and higher land inequality at the municipal level. We contribute to this literature by showing that the long-term effects of the colonial institutions were uneven across population groups. We find that Indigenous and Afro-Colombians at the bottom of the social ladder during the colonial period appear at the bottom in the contemporary educational system, while the elite status of the encomenderos and slave owners, on the contrary, has not persisted over time. Additionally, our results indicate that more attention should be place to understanding the persistence of the elites from the early republic, as this group was the most successful in maintaining its social status.

We provide a new picture of the marriage market in the country and in line with previous research, we show that assortative marriage reflects the educational segregation of the country. In particular, those at the bottom of the social ladder continued to be trapped with scarce possibilities to move upward. These results confirm that the segregated educational system reinforces assortative mating, as schools are essential for social interaction. At the

same time, homogamy contributes to the persistence of segregation, creating a feedback loop that is difficult to break.

Additionally, we contribute to the literature on the case study of Colombia, enlarging the perspective on the long–term evolution of relative social status rather than focusing on contemporary short–term intergenerational correlations. The findings contribute to the literature by showing that educational systems matter in the perpetuation of the elites (Cardenas et al., 2021; Clark & Cummins, 2013). Finally, we contribute to the literature by assembling several data–sets to track persistence and social mobility in the presence of fragmented information.

2 Education in Colombia

Today, the educational system in Colombia displays symptoms of social segregation, which has been recently shown by Cardenas et al. (2021). This segregation happens because the supply of quality education in Colombia is concentrated in private and expensive institutions that are out of reach for most. However, this phenomenon is not unique to Colombia, it has been observed in other countries like the US. Chetty, Friedman, Saez, Turner, and Yagan (2017) and Chetty et al. (2017) present compelling evidence of inequality of opportunity in access to high-quality education and its effects in the US schooling system. Both in the US and in Colombia these studies conclude that inequality in access to high-quality education limits the ability of education to promote intergenerational mobility.

Recently, Blanden, Doepke, and Stuhler (2023) suggested that educational inequality is at the root of low social mobility because if only the children of wealthy parents have access to high-quality education, inequality will be more persistent across generations compared to a society where education is less dependent on family background.

In Colombia, primary and secondary schooling access has enlarged since the last constitutional reform (1991). Coverage rates in primary education are above 90%, and in secondary education, they are close to 75%. Also, access to higher education has improved during the last decade, going from around 35% (2010) to 52% (in 2020). However, the quality of education is still poor (Camacho, Messina, & Uribe Barrera, 2017). Colombia has the lowest scores in the PISA standardised tests among the members of the OECD, and it is below the average level in South America. In 2018, the country reached an average score of 406 out of

600, while Mexico and Chile obtained 416 and 438, respectively.

Beyond improvements in coverage, the most salient characteristic of the educational system is the heterogeneity in opportunities to access high-quality institutions. In a country with an average low-quality level, a small number of high-quality secondary schools, located mainly in the urban areas, indicates segregation and status-driven access to high-quality schools. As shown in Fig. 1, high-quality schools are strongly associated with high-income families, while low-quality schools are primarily associated with low-income families.

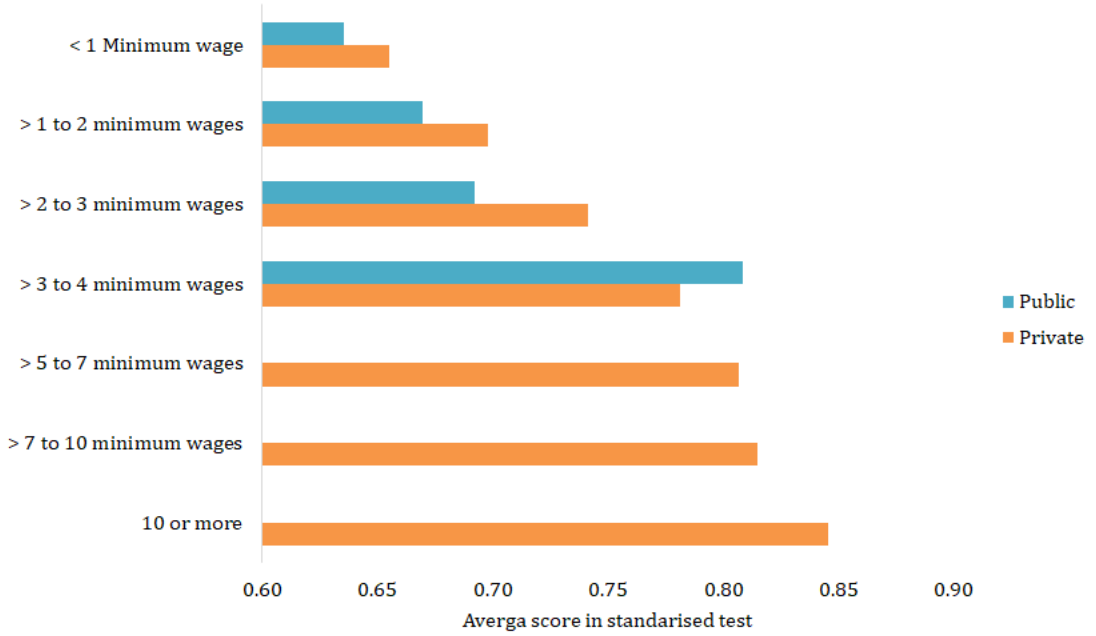
Additionally, school quality is strongly linked with the public/private character of the institution, as shown in Fig. 1. Public schools generally display worse quality than private schools, and public schools are also concentrated in low-income families. This implies that social status is highly correlated with school performance.[5]

All this generates a bottleneck in access to good quality higher education with a clear bias favouring high-income households and generating social apartheid as Cardenas et al. (2021) called it.[6] The probability of accessing higher education for a member of the lower 25% of the income distribution is around 32%, while it is 68% for a member of a household in the top 25% (See Cardenas et al. (2021) page 99). Furthermore, the probability of accessing higher education depends strongly on the level of education attained by the mother. The intergenerational correlation between mothers and children who access higher education is 92.4%. In comparison, this same correlation is 19.6% for mothers who only complete primary education and young people who only complete primary education (García Jaramillo, Rodríguez Orgales, Sánchez Torres, & Bedoya Ospina, 2015).

[5]Fig. 1 shows the monthly total income of households whose children attend a school in multiples of legal minimum wage. Colombia's legal monthly minimum wage was around 200 US dollars by January 2022. It is worth noticing that due to the high level of informality in the country, more than 40% of the workers earn less than the legal minimum wage.

[6]A similar pattern is observed in the US where college access varies greatly by parent income. See Chetty et al. (2017).

Figure 1: School quality and family income by type of school



Note: Figure 1 shows the average score in the national standardised test by family income and by the type of school. Source: Fergusson and Flórez (2021b).

But educational quality is not the only feature offered by private schools. Additionally, there is a clear elitist bias and a social status signalling mechanism related to attending a private instead of a public school. Although public schools perform worse on average than private schools on different quality measures, many private schools have lower quality performance than public schools with exceptionally good performance. Amid this heterogeneity, Cardenas et al. (2021) shows that there is a marked preference among the middle classes (especially the lower-middle income classes) to make an extraordinary economic effort for their children to attend private schools, even those of similar or lower quality than public schools where their children could attend for free. The work of Fergusson and Flórez (2021b) shows in detail evidence of an elitist bias in school preferences, which explains these decisions. More precisely, private schools seek to imitate the characteristics of very high-quality private schools, both in their names and educational proposals. The top of these schools is located in the capitals of the country's wealthiest regions and among high-quality private schools, the most elitist institutions are international private schools. Those institutions organise in exclusive associations as the group we focus on in Bogota: UNCOLI, an association of 27 international elite schools.

The causes behind this educational segregation are unknown. Cardenas et al. (2021) suggests that at least since the early-republic, political and ideological tensions prevented the advancement of public education. The limited public education created differences in social capital that can explain today's educational apartheid. In this research, we test if these historical roots also go back to colonial times.

3 Data

In this section we provide a comprehensive understanding of the data sets that were utilised for the quantitative analysis. The data collected for this paper are diverse, including surnames, administrative records, and historical and anthropological sources.

3.1 Surnames

The use of surnames is relevant as these are almost always passed from one generation to another carrying relevant characteristics strongly related to social status.[7] The assumption behind the rarity of the last names is that sufficiently rare surnames in a population represent observations of the same familiar dynasty.[8]

Since, at least the Bourbon reforms (18^{th} century), the modern use of the Spanish naming system was adopted in Colombia (*Viceroyalty of New Granada*) and during the 19^{th} century acquired legal enforcement. This system consists of the customary or legal use of two surnames after a given name. The first surname is the father's first surname, and the second is the mother's. This implies that both family lines are present in a given name. This naming system was conserved and reinforced in Colombia after independence from the Spanish Crown. For our purpose, this contributes to preserving the record of the mother's surnames for at least one additional generation compared with naming systems conserving only one surname (usually the father's). Instead of losing the mother's surname, we can identify both family links by observing contemporary cross-sectional data.[9]

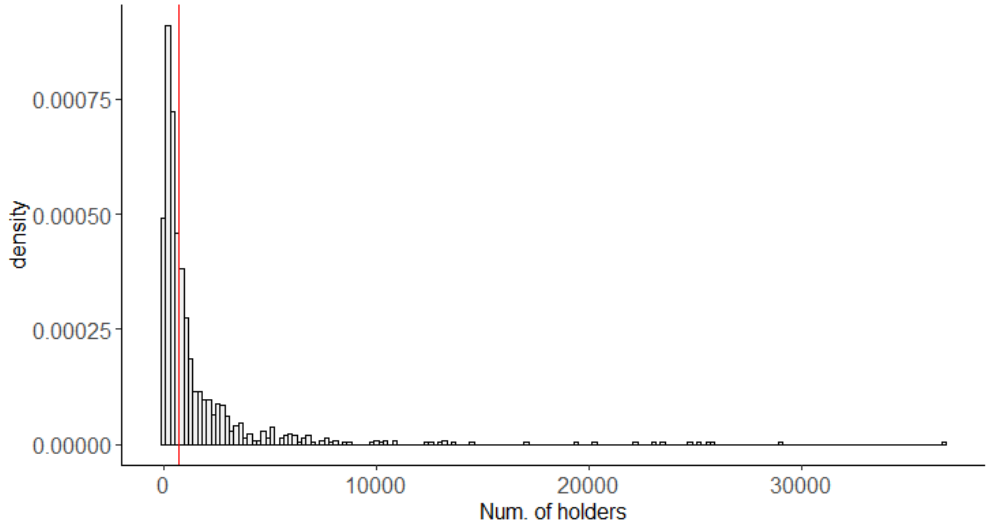
[7] For example, Guell et al. (2013) show that rare surnames correlated with observable metrics of social and economic outcomes for the Spanish case.

[8] See for example Clark and Cummins (2013) for a study of Oxbridge. See Guell, Rodríguez Mora, and Telmer (2007) and Clark (2014) for a more general application of surnames to measure social mobility in the long run and Barone and Mocetti (2021) for an application of surnames in the historical context of Florence.

[9] Usually, in Colombia, the first surname of all children is the paternal surname followed by the maternal surname. Therefore maternal surnames are usually the first name of their father. In the 1990s it was

To define a set of sufficiently rare surnames in the educational system in contemporary Colombia, we use a database that contains all the students enrolled in third grade of primary school in Colombia in 2016, corresponding to more than 90% of the families with children at this age (8-9 years old).[10] This sample has the advantage of including a good representation of low, high, and middle-income families. Using the Integrated System of Enrolment (SIMAT- Sistema Integrado de Matrícula), we have access to the individual data for 997,036 students, including their full names, the schools they attend, and the regions where they are enrolled. The frequency of appearance of every surname in the SIMAT measures the rarity and commonality. The distribution of these frequencies can be seen in Fig. 2.[11]

Figure 2: Distribution of surnames in Colombia, 2016



Note: Figure 2 shows the distribution of paternal and maternal surnames. The red line indicates the threshold of our preferred definition of rarity. Source: SIMAT, 8-9 years old enrolled in third grade of primary school, 2016.

We find 9,117 surnames held by only one student, such as *Savatoni* or *Abacuk*. The last point in the distribution is the surname *Rodriguez*, the most common surname in Colombia,

introduced the possibility of adopting as a first surname the mother's surname. However, if both parents are known and officially registered as such, both surnames must be included as part of a person's legal name.

[10] We focus on this age group as they provide the most complete sample of surnames. For primary school (1 to 5 grade) in 2016, Colombia's total net enrolment ratio was 93,58, and the gross enrolment was 113.56, according to the National Ministry of Education.

[11] The pattern of this distribution of surnames is well established in the literature (Santavirta & Stuhler, 2020) and it is worth noticing that this common feature of the distribution of surnames across *Western* societies presents an advantage for the method because the number of rare surnames is large.

which appears 36,678 times in the dataset. From this sample of Colombian surnames, we select those with less than 2% of the number of holders of the most common surname in the sample (Rodriguez), and more than three holders.[12] The latter limit is set to avoid possible remaining misspellings registered as a one–holder's surname. This sample defines the list that we consider as being part of the same extended family and, therefore, belonging to dynasties whose status may have persisted from the past.

We define a rarity as a proportion of the most common surname to define a reference standard for navigating different data sources. If our dataset of reference (SIMAT) is a good representation of the distribution of surnames in the whole population, keeping a reference to the frequency of Rodriguez holders as a measure of rarity or commonality of the surnames seems to be a reasonable strategy.[13] The total number of rare surnames obtained is 9,902, and we also include in our database the 20 most common surnames as a comparison group.

As reported in Table 1, from a total of 997,036 students, we obtained 23,144 unique paternal and maternal surnames. Some particular surnames needed a slight transformation to avoid confounders or typos. We transformed these surnames by removing accents or special characters (ñ, ü) and for the compound surnames like "De la Cruz", we merged them as a unique word.

Table 1: Surnames summary statistics

Number of distinct names	23,144
Mean frequency per name	86.16
% of people holding top-20 surnames	20%
% of people holding top-50 surnames	32%
% of people holding rare surnames	24%

Note: Rare surnames are defined as those held by more than 3 holders to avoid misspellings and less than 2% of the most common surname in the sample (Rodriguez). This is our preferred definition of rarity but we modifier this threshold as part of the robustness checks.

[12] We modify the definition of rarity using different thresholds and do not find significant differences in our results. See results in the appendix

[13] It is worth noticing that, according to the national official registry (*Registraduría Nacional del Estado Civil*), holders of Rodriguez represent around 1.5% of the entire population in Colombia. In our sample, this surname represents 1.8% of the total.

3.2 Historical legacy: elites and ethnic groups

To study the origins of educational segregation, we connect past and present by identifying groups of surnames that contain information about their historical social status. We are interested in those groups that occupy an extreme place at some end of the continuum of social status at different points in history. We select two ethnic groups, four groups that represent the pre-industrial elite and two groups that belong to the modern elite. For these groups, we have information on their historical social status linked to their surnames, and with this information, we follow them in the educational system today. We define a *group* as a set of families that share the same socioeconomic status measured, for example, by relevant ethnic social markers, access to elite education in the 17^{th}-18^{th} centuries, or owing African enslaved people at the end of the slavery in Colombia in 1851 and we collect the surnames of these families using several sources, including primary and secondary sources. In Section 7.A, we describe in detail the sources from which we collected the data.

After the first Spaniards settled in Colombia around 1500, the Colonial power established the institution of the *encomienda* in which the notable Spaniards living in the colony were allowed to collect tribute such as gold, fabrics, emeralds, and labour from the Indigenous communities in exchange for Christian education and protection (Colmenares, 2007). Soon after, the Indigenous were also forced to work in gold mines, *haciendas* and craft shops. By 1560 the *encomienda* was well-established in all the territory of Nueva Granada. Still, several social disagreements between the Crown and the encomenderos and the demographic collapse of the Indigenous population caused its abolition at the end of the 17^{th} century (Kalmanovitz, 2008). Despite the extractive aspect of this institution, several authors have suggested that the *encomienda* had persistent positive effects on the long-term development of the municipalities in which this institution was present and that its presence is associated with lower levels of poverty and higher secondary school enrolments. However, it appears to have no effect on inequality (García-Jimeno, 2005; Faguet, Matajira, & Sánchez, 2017; Mora Bustamante, 2016). This literature suggests that the causal channel of this persistence runs through the strengthened local presence of stronger public institutions since the *encomienda* was the precursor of the state. However, others have also pointed to the persistence of pre-colonial prosperity that is reflected in contemporary higher incomes (Maloney & Valencia Caicedo, 2016).[14] Our first list of surnames comes from an official document written in 1674 by Juan Florez Ocariz from which we obtained the list of names of Encomenderos.

[14]The *encomienda* settled in more dense areas where the presence of the pre-colonial population was higher.

Then, we collect a list of 131 Indigenous surnames from a detailed anthropological study (Instituto Colombiano de Cultura Hispanica, 1992).

The striking demographic collapse of the Indigenous population during the 16th and 17th centuries resulted in the importation of enslaved people from Africa to work in the gold mines. By the 18th century, the economy of Colombia was based on mining, agriculture, and commerce; in turn, these activities were established around slave labour. The regions historically associated with gold mining and the production of sugarcane and other forms of plantations were the preeminent destination of African enslaved people, mainly in the departments of Cauca, Antioquia, Chocó and Bolivar, while most of the African enslaved people entered through the port of Cartagena (Jaramillo Uribe, 1963). As with the *encomienda*, researchers have measured the long-term impact of slavery on economic development (García-Jimeno, 2005; Acemoglu et al., 2012) concluding that the presence of gold mines in the 17th and 18th centuries is associated with contemporary lower school enrolment and higher land inequality at the municipal level. Slavery was finally abolished in 1851 and the Afro-descendant populations founded several communities near the Caribbean and Pacific coasts. We use a list of slave owners complied by Tovar (2007) and Tovar and Tovar (2009) and we use two lists of surnames compiled by afrocolombianists, built upon a set of past studies (Zapata Olivella & Mina Aragon, 2014; Mosquera, 2014).

Beyond slavery, one of the salient traits of the segregation patterns during the colonial period was the use of blood purity. This royal legislation that originated in Spain during the 15th century to exclude Jews and Arabs from critical public positions was established in the colonies to avoid racial mixing and maintain control of the region's political and religious institutions (Villalobos, 2005). The legislation required proof of the purity of blood, legitimate birth, and honorability for appointment to several positions, including ecclesiastic and military, and admission to secondary and higher education (Helg, 2014).

Schools and higher-education institutions appeared during the second half of the 16th century to educate the heirs of Spaniard families. These institutions were the principal educational centre of the ruling elites during colonial times but also of the leaders of the bourgeoisie pro-republican groups during the Independence process that started after 1810. We collected a list of 224 students of the Colegio San Bartolome during the colonial period until the Independence, using the book *Real Colegio Mayor y Seminario de San Bartolome, 1605 to 1820*. We built a list of 5,482 male students of the Universidad del Rosario using the work by

Guillén de Iriarte (2006) and Guillén de Iriarte (2008) who reports a list of all the graduates from 1773 until 1842.

Independence movements against the Spanish Crown took place in the first decade of the 19th century, and the consolidation of an independent republic was achieved in the 1820s. The long-term effects of the Independence process in Latin American countries during the first half of the 19tth century are still a matter of open debate in the literature (Grafe & Irigoin, 2006). For example, it is argued that the fiscal system was destroyed while internal conflicts (wars) increased in the regions (Coatsworth, 1993; Bulmer-Thomas, 2003). A good part of the Colombian historiography coincides with pointing out that the break with the Spanish monarchy did not constitute a radical change in the political, economic, and cultural elites of these countries, as regional elites defended their political and economic interests. Ripoll Echeverría (2006) calls it a "revolution without renovation", while Olarte (1993) points out that the Creole elites (whites from Spanish families born in the colonial territory), who led the Independence movements once Napoleon overthrew the King of Spain in 1808, consolidated a highly hierarchical system, maintaining the social class segregation of the old regime.

In the long-run, Latin American countries also benefited from institutional and economic modernisation (Coatsworth, 1993; Prados de la Escosura, 2009). The importance of the slave economy started decreasing as new industries emerged while mining lost importance compared to new agricultural products (Bértola & Ocampo, 2012). The expansion of agricultural exports was accompanied by rapid colonisation of new lands in the lower slopes of the mountain ranges and in the country's interior that continued until the beginning of the 20th century. The expansion of the tropical export crops, and especially the production of coffee, laid the foundation of the modern economic era with the emergence of early forms of industrial development in Antioquia and Cundinamarca and the banking sector in 1870.

However, the persistence and the economic effects of the new institutions and elites that emerged after the independence has been less studied in the literature. But in any case, by the beginning of the 20th century, an economic and cultural elite was settled in cities like Bogota, Medellin, and Cartagena. These elite groups were linked to productive activities such as the financial and commercial system and also promoted the creation of cultural (but exclusive) spaces such as social clubs for gentlemen. As discussed in the next section, these groups also established high-quality *private* schools and universities that contributed to the

persistence of a highly segregated school system. To study the elite of the early 20th century, we collected a list of individuals from the founding acts of the Jockey Club and a list of the founders of the first 12 banks established in the country, between 1870 and 1883, from the foundation letters of each firm. Table 2 summarises these groups, and we describe in detail the historical groups, their economic and social position in history, and the sources from which we collected the data in Section 7.A.

Table 2: Historical groups

Social condition	Historical group	Period	Source
Ethnic groups	Indigenous	Forced work since 15th century	Instituto Colombiano de Cultura Hispanica (1992)
	Afrocolombian	Forced work since 16th century	Zapata Olivella and Mina Aragon (2014) and Mosquera (2014)
Pre-industrial elite	Encomenderos	16th – 17th century	Carrizosa (1990)
	Slave owners	Received compensation from the state in 1851	Tovar (2007) and Tovar and Tovar (2009)
	Colegio San Bartolome	1605 – 1820	ICCH, 1996
	Colegio del Rosario	1773 – 1842	Guillén de Iriarte (2006) and Guillén de Iriarte (2008)
Modern Elite	Bank founders	1870 – 1885	Founding acts
	Jockey Club	1874 – 1902	Foundation letters

Note: Detailed information about the historical groups, their economic and social position in history, and the sources from which we collected the data can be found in Section 7.A.

It is essential to recognise that the contemporary Colombian population, as in much of Latin America, is primarily a mixture of pre-Hispanic and Spanish populations, although some groups have mixed less. For example, a genetic study comparing self-reported racial belonging with evidence of mtDNA found that populations that recognise themselves as "mulattoes", "mestizos," or "Afro-descendants" have a dominant component of African mtDNA (Salas et al., 2008), suggesting little miscegenation. The study also shows that, contrary to the massive migratory waves in the southern part of the continent, Colombia did not have evidence of genetic drift after the Colonial period, as is the case in Venezuela and Peru. However, the country's distribution patterns of genetic origins differ considerably between regions. The coastal populations near the Pacific Ocean and the Caribbean sea show an essential component of the Afro-descendant population with little evidence of admixtures for more than a century and a half. Therefore, despite the importance of miscegenation, the regions with a more significant presence of Afro-descendants, both self-reported and by mtDNA studies, present minimal genetic admixture.

On the other hand, the regions with higher income levels and the large cities of the interior (located on the Andes) present a greater dominance of mestizos, with a strong presence of pre-Hispanic origins. But, even in large cities, there is no evidence of significant changes in the gene pool during the last 150 years (Salas et al., 2008). These findings allow us to conclude that although Colombia is a country with substantial *mestizaje*, Afro-Colombians

populations tended to remain relatively isolated from people of indigenous or European origin, and there is no evidence of significant changes introduced by migrations since the 19^{th} century in the mass of the population.

Then, the social groups analysed in this work can be considered extreme representatives of genetic distribution. On the one hand, people who retain Pre-Hispanic and Afro-Colombians surnames tend to be populations with little miscegenation. On the other hand, although elites with more remote origins in the early colonial period may have had greater miscegenation, our analysis shows considerable segregation reinforced by patterns of assortative mating. In this sense, our research does not allow us to capture the behaviour of the bulk of the population where miscegenation is more pervasive.

3.3 Measures of access to education

To measure contemporary access to education, we use different administrative datasets collected at the individual level that contain both surnames and relevant information on the quality of education, depending on the school or university the individual attended. To define contemporary access to education, we construct eight categories (e.g., high and low-quality education, public low-quality education, and high-prestige high schools) described in Table 3. Appendix Section 7.B describes each dataset in detail, the sources we obtained, and provides additional information about the educational system.

Table 3: Description of contemporary access to education

Access to education	Outcome	Definition
High status	Los Andes University	Most prestigious private university in Bogota
	International Schools	Most prestigious and expensive schools located in Bogota
	High Quality school	school being on top 5% of the score in the national standardised test
	HQ Public school	Public school being on top 5% of the score in the national standardised test
	HQ Private school	Private school being on top 5% of the score in the national standardised test
Low status	Low Quality school	school being on bottom 5% of the score in the national standardised test
	LQ Public school	Public school being on bottom 5% of the score in the national standardised test
	LQ Private school	Private school being on bottom 5% of the score in the national standardised test

Note: The variables come from the registration forms filled out by students who finish high school and take a standardised test known as SABER 11, similar to the SAT in the USA or the A-levels in Europe. The variable of Los Andes comes from the graduation act of the University. See Appendix Section 7.B for more details.

4 Methodology

4.1 Relative Representation in the educational system

With this information, the relative representation of any surname j is calculated as described in equation 1.[15] Then, we go from individual-level data to surname-level data by taking the average relative representation of each surname.[16]

$$RR_j = \frac{\text{Share of j in access to education}}{\text{Share of j in SIMAT}} \tag{1}$$

This measure captures how much a surname is over-represented or under-represented in access to high and low quality education given its frequency in the total population. For the common surnames, which tend to behave similarly to the average population, the relative representation in any period and in any social outcome will not be statistically different from one. For high-status surnames in high-status outcomes, the relative representation of the surname will exceed one, but in low-status outcomes, it will fall below one. On the contrary, for low-status surnames, the relative representation will exceed one in low-status outcomes, and be lower than one in high-status outcomes.[17] With this measure it is possible to assess the persistence is historical status because, in the presence of social mobility, historically elite/marginalised groups should not appear today significantly over-represented/under-represented in high-quality schools.

4.2 Persistence of historical groups

To evaluate whether the contemporary differences in the educational system have deep roots in the past, we regress the relative representation at the surname level (RR_j) on a set of dummy variables for each social group. Following equation 2, we test wheter the mean level of the RR of rare surnames belonging to historical groups is statistically different from the mean value of this metric for the most common surnames.

[15] It is important to note here that equation 1 refers to the relative representation of an individual surname. With the RR_j we can estimate the average level of the relative representation of a historical group.

[16] We adapt this measure to different subnational levels in Section 5.2.

[17] The distribution of the relative representation at the surname level is clustered around zero and it has a long right-hand tail as shown in the Appendix 11

$$Z[RR_j] = c + \sum_{g=1}^{n} \delta D_g + \varepsilon_j \tag{2}$$

Where g is the social group to which surname j belongs, D_g is the matrix of dummies for each group g, $Z[RR_j]$ is estimated at the surname level for each educational outcome, and c is the constant term. The standard errors are robust to account for potential heteroscedasticity given that it is likely that, like other social status measures, the variance of the Relative Representation is different across the historical groups. To allow for an easier interpretation of the results, we standardise the RR_j values using a *Z-score* method, and we will refer to this transformation as $Z[RR]$. This transformation allows us to interpret the coefficients of our regressions in terms of standard deviations with respect to the mean value. As previously discussed, a common surname is expected to be equally present in any social condition in a similar proportion as it is within the entire population and it has been largely documented that the average social status of common surnames, in Western cultures, is similar to the average social status of the society studied (Clark et al., 2015; Chetty, Hendren, Kline, & Saez, 2014). Given that the most common surnames have an average social status, the value of the standardised relative representation (RR) of those surnames (i.e. Rodriguez) is close to 0, or at least not statistically different from 0.

4.3 Assortative mating

For our analysis, it is worth studying the patterns of homogamy within elites and ethnic groups and to what extent those patterns are present in the educational system. Assortative mating refers to the tendency of individuals to choose partners who belong to the same social group, even when other options are available. These patterns can be observed and measured through the prevalence of couples who share common social characteristics, in our case, surnames that belong to the same historical group.[18] More importantly, if the attendance to different schools and higher education institutions is segregated between elites and historically marginalised groups, the possibility of encounters between those groups is reduced.

Assortative mating can contribute to perpetuate segregation in education through homogamy.

[18]For example, there is abundant evidence of a strong correlation among spouses by religion, ethnic group, and social status; whether measured by occupation or education level (Blossfeld, 2009; Schwartz, 2013; Ermisch, Francesconi, & Siedler, 2006).

For example, individuals who attend high-quality schools tend to marry others with similar educational backgrounds. Their children have a higher probability to attend high-quality schools and later complete higher education. In that case, this can lead to the concentration of highly educated individuals in specific social environments. These groups may have better access to resources and opportunities, which can further reinforce the educational advantage of these individuals and their children. In contrast, individuals with lower levels of education may be more likely to marry partners with similar educational backgrounds, leading to the concentration of less educated individuals in other neighbourhoods or communities. These groups may have fewer resources and opportunities, which can contribute to the perpetuation of educational disadvantage.

We can consider that a segregated educational system can reinforce assortative mating because schools are an important place for social interaction. At the same time, homogamy can reinforce the barriers to social mobility, contributing to the persistence of segregation.[19] With these assumptions in mind, we document marriage patterns in different types of educational institutions by looking at the proportion of paternal-maternal surname pairs belonging to each historical group. Exploiting the Spanish naming custom allows us to make use of both surnames (maternal and paternal) to observe marriage patterns as we consider each pair of surnames of any individual student as a marriage. First, we present a description of the marriage market in the whole population using the data from SIMAT and then we focus our analysis on three types of access to education: low-quality public schooling, high-quality private schooling, and Los Andes University. Looking at this, we test whether assortative mating is more pronounced among elite groups attending high-quality private schools compared to those in low-quality public schools. Additionally, we explore whether there is a higher degree of assortativeness in the marriage market within ethnic groups in various types of educational institutions.

Cultural preferences, the protection of social status, or a social norm can explain the preference for a partner of the same group. If there is a clear preference for a couple of the same group we can assume it is motivated by complementarity between members of the same social group. Furthermore, the preference for a partner of a different social group may be decreasing on the social distance from that group.

[19]The literature argues that homogamy allows elite social groups to maintain their social status for future generations and can contribute to a segregating pattern reproducing the exclusion of underclass groups.

According to this logic, we follow Chiappori et al. (2020) by computing the Separable Extreme Value (SEV) index they propose. This index captures the distance between a hypothetical pure random matching, with the actual distribution of couples. The index is derived from a model of marriage matching with no search friction to isolate the individual incentives to prefer to form a couple with someone with similar characteristics. The model focuses on the stability of a couple resulting from a comparison of the "surplus" obtained by each individual from forming a couple with someone of the same group compared with forming a mixed couple. Under the assumption of transferable utility, the model defines assortativeness as a measure of the deterministic component of the surplus derived from a marriage that depends only on a particular characteristic of the partner: i. e. income, education level, or belonging to a particular social group.

The advantage of the approach proposed in Chiappori et al. (2020) is that we can calculate the deterministic part of the surplus using observable patterns of marriage between social groups and this requires only 3 parameters. The first two are the proportion of paternal and maternal surnames belonging to a particular group: noted m and n respectively. The third is the proportion of marriages where both surnames are from the same group: noted r in what follows.[20] If there is perfect assortative matching, then r would equal the minimum of m and n and the index will be higher than zero. If the matching was random, then r would equal the product of m and n, in this case, the SEV index equals zero. Before giving the actual expression of the index, we show in Table 4 the different possible scenarios in the marriage market comparing two generic groups as an illustrative example: elite and ethnic group.

[20]To explain the construction of the index we will assume that we only have two categories: elite and ethnic group surnames, but in our analysis, we will use five categories: Indigenous, Afro-Colombians, pre-industrial elite, modern elite, and commons.

Table 4: Assortative matching scenarios between groups

a. Observed matching

paternal/maternal	elite	ethnic group
elite	r	m-r
ethnic group	n-r	1+r-m-n

b. Random matching

paternal/maternal	elite	ethnic group
elite	mn	m(1-n)
ethnic group	n(1-m)	(1-m)(1-n)

c. Perfectly Assortative

paternal/maternal	elite	ethnic group
elite	n	0
ethnic group	0	1-n

Note: The Table 4 represents the different scenarios of matching with only two categories: elite and underclass surnames. Where r denotes the share of marriages where both surnames are from an elite group, and m and n denote the number of elite paternal and paternal surnames, respectively. In Panel C, we assume the surnames distributions for paternal and maternal surnames are identical. The table is based on Chiappori et al. (2020).

The index is defined in Eq. (3). To summarise, mn is the proportion of couples if randomly matched and there is positive assortative mating if $r \geq mn$. The I_{SEV} is the log of the ratio between the proportion of observed matches between couples of the same group and the observed proportion of couples formed by surnames belonging to different social groups. If this ratio equals one, then the index will be zero, and we can conclude that there is random matching and, thus, that there is no positive surplus of assortativeness.

$$I_{SEV} = ln\left(\frac{r(1+r-m-n}{(n-r)(m-r)}\right) \tag{3}$$

We calculate the SEV index for every possible couple of historical groups and also for the couples formed with one common surname. If the ratio is higher than one, this means there are more couples in the same group than couples formed by different social groups, which indicates high assortative complementarity.

Another interesting property of the SEV index is that it allows assortativeness to be measured

locally for a particular social environment. When considering the probability for two groups to form a couple the degree of assortativeness can vary depending on the possibilities of interaction between them.[21] In sum, the SEV index allows us to estimate a measurement of assortativeness that is comparable across different types of education and by looking at the index for each pair of historical groups we can determine if there are different levels of assortativeness between pairs of groups.

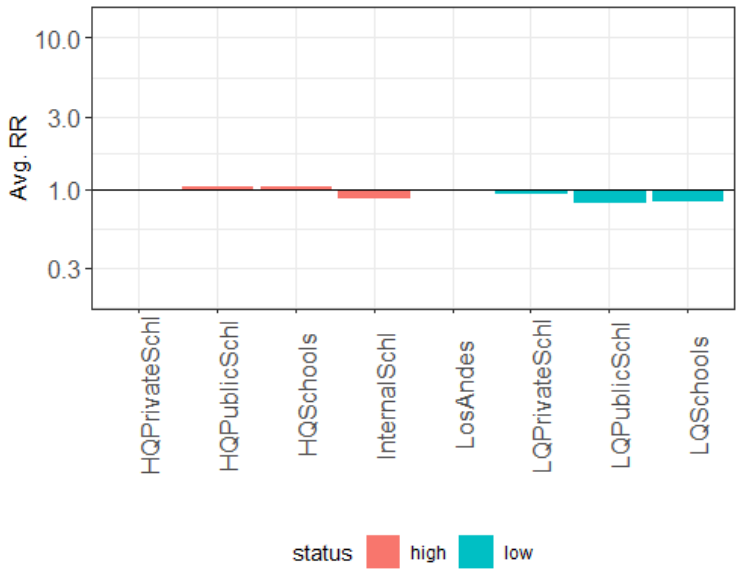
5 Results

5.1 Persistence of historical groups

Fig. 3 and Fig. 4 show the distribution of the average relative representation within each social group in educational categories in 2016. Fig. 3 shows this information for the 20 most common surnames in Colombia, and the figure can be used as the standard to compare our metrics of over or under-representation in a particular contemporary social status marker. A RR smaller than one implies that the group of surnames is underrepresented while a RR higher than 1 indicates that the group of surnames is over-represented in any given educational category. As shown in the graph, common surnames do not appear over or under-represented in any category, with an average RR close to 1.

[21] Chiappori et al. (2020) highlight assortativeness as a local property, which means that it can also capture assortativeness within couples of elite social groups (e.g. Jockey members and Bank shareholders) in a particular social environment, namely an elite school. While elite groups can exhibit different patterns of assortativeness in a different social environment, like a low-quality public school. Furthermore, the SEV index can show low values within other pairs of groups in the same social context (e.g. Bank shareholders and Indigenous).

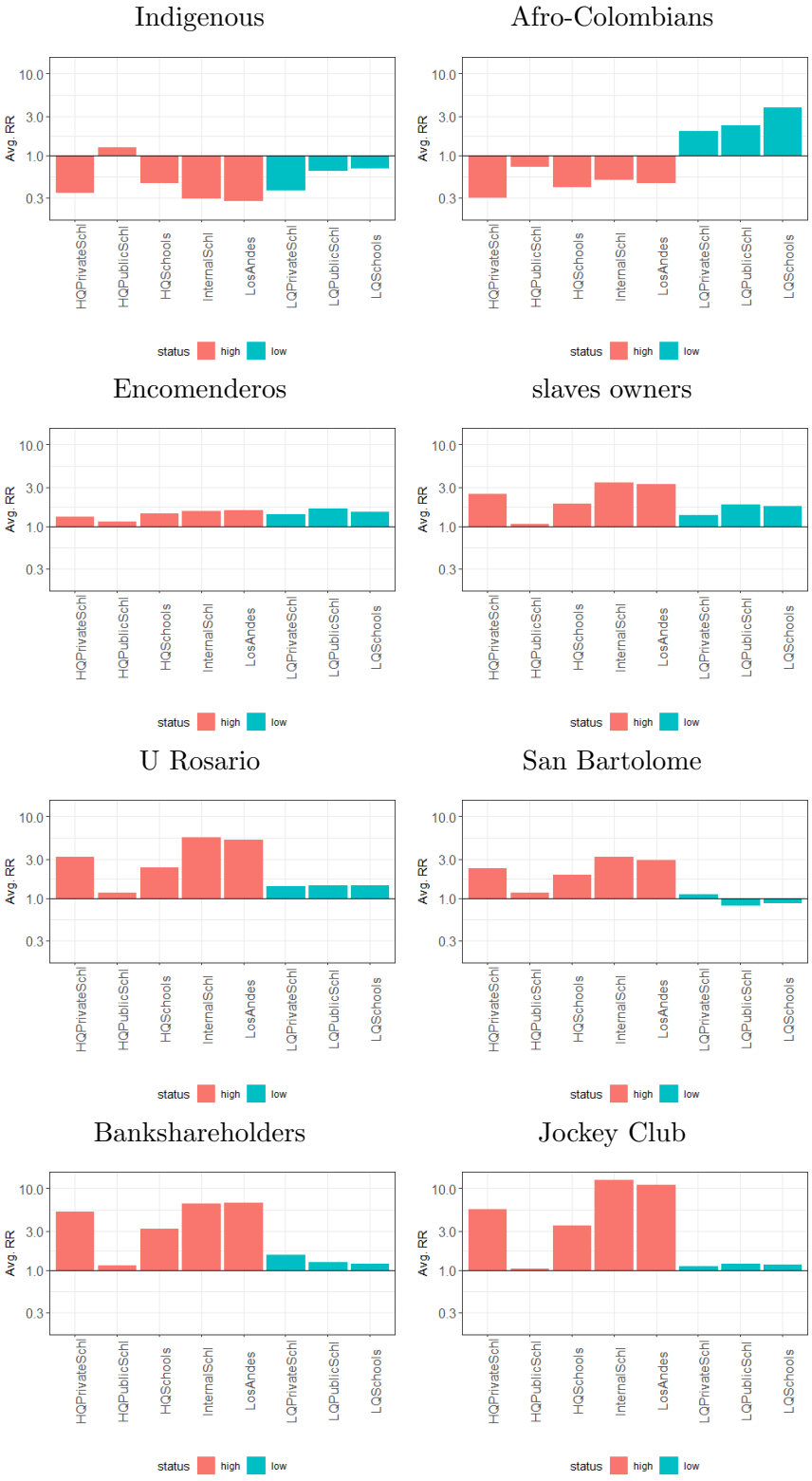
Figure 3: Relative Representation in access to education - Common surnames



Note: Fig. 3 presents the average relative representation of the 20 most common surnames in Colombia. The horizontal black line is set at $RR_j = 1$, corresponding to the neutral level of relative representation. The y-axis is log-scaled to ease interpretation.

The most striking result from Fig. 4 is the long persistence of the ethnic groups. The Indigenous and the Afro-Colombians remain underrepresented and overall absent from high-quality education and elite institutions. On the opposite side of the social ladder the Jockey Club, the banks–shareholders and, to a lesser degree, the graduates of the colonial schools, San Bartolome and Universidad del Rosario, show systematic over-representation in the high-quality institutions. Of special attention is the overrepresentation of the member of the Jockey Club in the graduates from Los Andes, as this group appears at a rate nearly 12 times higher than in the general population. On the opposite side, Indigenous appear at an extremely low rate in Los Andes, about 0.3 times their frequency in the population. In comparison, Clark (2014) estimates that Ashkenazi Jews appeared 6 times more among physicians in the US in 2012 than in the general population, while Native Americans occur at about 6% of the expected rate.

Figure 4: Relative Representation in access to education



Note: The horizontal black line is set at $RR_j = 1$. The y-axis is log-scaled to ease interpretation.

It is interesting to notice that the group of encomenderos seems to practically behave as the commons, as it does not appear systematically over or underrepresented in any category. These results suggest a process of convergence to the mean.[22] Similarly, the group of slave owners appears overrepresented in several high-quality and low-quality outcomes, suggesting that the group lost its elite-type characteristic, also indicating a regression to the mean process. This result is surprising given that Ager, Boustan, and Eriksson (2021) found that in the case of the US, the sons and grandsons of former slaveholders surpassed their counterparts in educational and occupational attainment. However, it should be noticed that these two social groups are not only defined far back in time, but they are also related to the pre-modern, and colonial society elites, while the other elite categories are not only recent but, as is the case of the graduates from the colonial schools, cover professional categories as doctors and leaders of the bourgeoisie pro-republican groups during the Independence process. In particular, we observe the slave owners right at the collapse of slavery, and it is likely that several of these families could not enrol in the modern era's new productive activities.

As for the surnames linked to schools of higher education in the Colonial period, these surnames seem to have preserved better their status. Notably, surnames associated with the University of Rosario are more prevalent in high-status positions than those associated with the Colegio San Bartolome. One possible explanation for this discrepancy is that the University of Rosario was established to train civil authorities under the direct protection of the King, while the mission of Colegio San Bartolome was to educate religious individuals. Thus, the social status associated with the University of Rosario may have persisted more strongly over time due to its roots in the ruling class.

These results indicate that although status has persisted for certain groups, especially for Indigenous, Afro-Colombians, and elite groups of the late 19^{th} century, some pre-modern groups have lost their original high status. These findings suggest that the long-term effects of colonial institutions, such as the *encomienda* and slavery, were uneven across population groups and more lasting in the indigenous and Afro-Colombians.

[22] In Jaramillo-Echeverri et al. (2021), we find a similar pattern for the Chilean case, where the Spaniard colonial officers appear as if they were slowly losing their elite status. We confirm that this group shows the lowest persistence of all elite groups in both Chile and Colombia when measured by the relative representation in high-status categories. In Chile, they behave almost as the commons, only slightly over-represented in high-income groups. This regression to the mean is more evident in Colombia as they appear slightly under-represented in high-quality schooling.

The results from table 5 confirm the patterns shown in figure 4. For example, looking at all high-quality education measures, Indigenous and Afro-Colombian surnames appear underrepresented by around 0.5 standard deviations compared to the common surnames. On the contrary, members of the Jockey Club appear overrepresented by more than 0.7 standard deviations. In line with the findings of the previous section, Encomenderos are significantly underrepresented in elite-type schools by around 0.3 standard deviations while slave owners and alumni from Colegio San Bartolome do not appear to be to behave differently from the common surnames.

In general, the late 19^{th} century elite groups are over-represented in high-quality education. In contrast, the pre-industrial groups behave in all categories as common surnames, showing small and non-significant coefficients in almost all outcomes. On the contrary, Indigenous and Afro-Colombian surnames are absent from all privileged education outcomes.

Regarding the low-quality institutions, most groups do not behave significantly differently from the commons, the exceptions being the students from the San Bartolome, that are underrepresented by almost 0.2 standard deviations, the Afro-Colombians that are over-represented by around 0.6 standard deviations in low-quality schools and the indigenous surnames that appear underrepresented in these schools by 0.2 standard deviations. One fact should be highlighted: even though SIMAT covers around 90% of school-age children, according to the 2005 Census Report, the Indigenous population shows the most lagging indicators in all the education variables, including enrolment in primary school. This implies that within the SIMAT the poorest Indigenous children are not even registered; therefore, we may be underestimating their participation in the lower strata. This could be a potential concern because this can bias the estimations upwards. However, as the results indicate high persistence, Indigenous could face even higher persistence.

Overall, the results show that contemporary educational segregation in Colombia occurs mainly at the top, from which ethnic groups are absent. At the bottom, in low-quality schools, there is no significant differentiation between elite surnames and common surnames. However, ethnic groups do show a distinct pattern: Afro-Colombians are over-represented in low-quality institutions, while indigenous are once again absent.

Finally, the rare surnames not identified in any historical group are less represented in both ends of the educational outcomes than in the commons. They appear underrepresented by around 0.1 to 0.3 standard deviations compared to the common surnames, which confirms

that having a rare surname is insufficient to predict access to particular education.

The results in Appendix Table 12 show that when we include all surnames, the general findings hold, but, as expected, the significance and explanatory power of the regressions fall considerably as the predictive power of frequent surnames is lower than the predictive power of rare surnames (Santavirta & Stuhler, 2020). On the contrary, the patterns remain when we impose a more restrictive definition of rarity (1% of the most common surname). Still, the significance of every coefficient and the overall fit of the regression improves, as shown in Appendix Table 13.

Table 5: Differences in educational categories by historical groups

	Los Andes	International schools	High Quality Schools	HQPrivateSchool	Low Quality Schools	LQPublicSchool
Ethnic groups						
Indigenous	-0.519^{***}	-0.359^{**}	-0.485^{***}	-0.426^{***}	-0.220^{***}	-0.200^{**}
	(0.145)	(0.118)	(0.079)	(0.092)	(0.070)	(0.067)
Afro-Colombians	-0.485^{***}	-0.336^{**}	-0.511^{***}	-0.444^{***}	0.568^{***}	0.179
	(0.137)	(0.109)	(0.072)	(0.083)	(0.166)	(0.099)
Pre-industrial elites						
Encomendero	-0.390^{***}	-0.298^{**}	-0.191	-0.286^{***}	-0.008	0.030
	(0.113)	(0.090)	(0.102)	(0.075)	(0.137)	(0.145)
slaves owners	-0.109	-0.069	-0.014	0.025	0.085	0.098
	(0.146)	(0.120)	(0.083)	(0.094)	(0.085)	(0.062)
URosario	0.212	0.171	0.137	0.132	-0.008	-0.011
	(0.178)	(0.154)	(0.119)	(0.125)	(0.07)	(0.064)
SanBartolome	-0.306^{*}	-0.217^{*}	-0.063	-0.127	-0.194^{***}	-0.188^{***}
	(0.127)	(0.110)	(0.080)	(0.091)	(0.054)	(0.061)
modern elites (late 19^{th} century)						
Bankshareholders	0.047^{***}	-0.110	0.286	0.432	-0.067	-0.052
	(0.259)	(0.214)	(0.163)	(0.239)	(0.094)	(0.088)
JockeyClub	1.448^{***}	1.320^{***}	0.580^{***}	0.756^{***}	-0.063	-0.049
	(0.366)	(0.299)	(0.161)	(0.215)	(0.065)	(0.059)
Rare No Groups	-0.385^{***}	-0.250^{***}	-0.296^{***}	-0.263^{***}	-0.098	-0.078
	(0.144)	(0.117)	(0.075)	(0.090)	(0.064)	(0.058)
Constant	0.342^{*}	0.219	0.260^{***}	0.227^{*}	0.085	0.068
	(0.144)	(0.117)	(0.074)	(0.089)	(0.063)	(0.057)
Observations	9,922	9,922	9,922	9,922	9,922	9,922
Adjusted R^2	0.044	0.029	0.020	0.024	0.005	0.002

Note: The equations are estimated using OLS. The 20 most commons surnames is the latent group of comparison.The RR_j is normalised. Robust standard errors in parentheses and *** p<0.01, ** p<0.05, * p<0.1 indicate statistical significance.

It should be noticed that it is possible that while Indigenous and Afro-Colombians are exclusive categories, some elite surnames appear in more than one elite group. For example, there is one surname that appears in all the elite groups, while the surnames 7 surnames appear in 5 of the 6 elite groups. Of particular attention are those surnames that appear both in historical groups from the pre-modern era as well as from the late 19th century. These surnames represent the long-lasting elite that has persisted over time and that was able to reinvent itself, despite the failure of the *encomienda* or the independence process that sought to break the colonial institutions. To evaluate whether these long-surviving surnames have been more successful to maintain their social status in comparison to those that appear in only one period, we classify elite surnames into a new category *multiperiod elite* that includes all surnames that belong to at least one pre-modern group and one modern group.

Table 6 shows the results by educational categories. The coefficients of the *multiperiod elite* group show that in all high-quality categories, this group appears around one standard deviation more represented than the common surnames and reflects the systematic over-representation of elite surnames in access to high-quality schooling. Similar to the elite of the 19th century this group is overrepresented at the top but not underrepresented at the bottom. Once again, the size of the coefficients of the elite surnames that appear only in the pre-industrial period suggests that this group has been slowly converging to the mean. These findings corroborate that within the elite there are two different patterns of persistence. The first one corresponds to the elite surnames from the pre-industrial groups that were unable to enrol in the modern elite groups and show that this group has slowly moved towards the mean status. The other pattern is one of the long-surviving elites' surnames and the modern elite groups which indicates that these two groups have been more successful to maintain their social status and accessing high-quality education.

Table 6: The long-lasting elite

	Los Andes	International schools	High Quality Schools	HQPrivateSchool	Low Quality Schools	LQPublicSchool
Indigenous	-0.258^{***}	-0.124^{**}	-0.275^{***}	-0.222^{***}	-0.207^{**}	-0.131^{*}
	(0.056)	(0.040)	(0.054)	(0.044)	(0.075)	(0.064)
Afro-Colombians	-0.259^{***}	-0.132^{**}	-0.333^{***}	-0.271^{***}	0.578^{***}	0.238^{**}
	(0.066)	(0.048)	(0.050)	(0.047)	(0.147)	(0.089)
Multiperiod elite	1.208^{***}	1.053^{***}	0.811^{***}	0.998^{***}	-0.082	-0.010
	(0.247)	(0.206)	(0.141)	(0.189)	(0.090)	(0.074)
Only pre-industrial elite	0.142^{*}	0.162^{**}	0.211^{***}	0.206^{***}	0.030	0.100
	(0.069)	(0.049)	(0.057)	(0.050)	(0.080)	(0.067)
Only modern elite	1.634^{*}	1.375^{**}	0.846^{***}	1.139^{***}	-0.090	0.026
	(0.646)	(0.524)	(0.221)	(0.313)	(0.104)	(0.105)
Rare No Groups	-0.124^{*}	-0.015	-0.086	-0.058	-0.085	-0.009
	(0.054)	(0.035)	(0.047)	(0.039)	(0.068)	(0.054)
Constant	0.081	-0.016	0.050	0.022	0.072	-0.001
	(0.054)	(0.034)	(0.045)	(0.038)	(0.068)	(0.053)
Observations	9,922	9,922	9,922	9,922	9,922	9,922
Adjusted R^2	0.038	0.023	0.020	0.024	0.004	0.001

Note: Persistent elite is defined if a surname belongs to at least one pre-modern and one modern group. The equations are estimated using OLS.The 20 most commons surnames is the latent group of comparison.The RR_j is normalised. Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1.

The institutional design of the Colonial system created segregated access to education, and as a consequence, elite–exclusive access. The Spanish Crown defined rules of access to higher education that forced the candidates to demonstrate their connection with family traditions linked to the traditional Spanish nobility. Towards the end of the colonial period, there was an elite of families of several generations born in the vice-royalty. These *Creole* elites were, in large part, protagonists of the process of independence and creation of the Republic, at the beginning of the 19^{th} century. Despite this political transformation, which sought to break with colonial forms, the rules of access to education did not transform substantially. Although the requirements for the purity of blood and nobility were relaxed, we show that *de jure* segregation in access to higher education continues.

In conclusion, we find significant and robust differences in the relative representation between historical groups in access to education. These differences confirm the persistence of social segregation between ethnic and elite groups. Furthermore, a suggestive pattern of heterogeneity among the elite groups emerges. First, the existence of convergence to the mean status of the pre-colonial elites. Second, the elites related to modern productive activities or the emerging bourgeoisie of the end of the 19^{th} century and the long-lasting elite present strong persistence of their privileged high status.

5.2 Regional view

The previous findings revealed that disparities in access to education have deep historical roots, with historically marginalised groups being absent from high-quality education at the national level. But could these results be explained by the association between surnames and regions? Or could it be that Indigenous and Afro-Colombian surnames come disproportionately from isolated or unusual populations? If this was the case, then the groups we are observing may be different from the rest of the population. Additionally, the concentration of high-quality schools in a limited number of cities may be driving these patterns.

To address these concerns, we calculated the relative representation of each historical group in low and high-quality schools across the six regions of the country. We defined each region by including departments that share both geographical and historical contexts. The country could be divided into 6 macroregions: Caribe, Pacifica, Cafetera, Centro, Oriente and Amazonas based on their proximity and geographical characteristics. The departments that form each of these macroregions share similar historical and cultural characteristics.

Topographic factors have been tied to the location of the population since pre-colonial times and were determinant in the foundation of settlements during the colonisation process. Also, these geographical features such as Los Andes, kept the regions isolated before the end of the 19th century when investments in roads and railroads started to grow.[23]

Given the significant regional inequality in education quality present in Colombia, we defined high-quality schools at the regional level as those that were in the top 20% of the standardised test scores in each region and low-quality schools as those in the bottom 20%. We defined rare and common surnames at the regional level using the same criteria as at the national level. Specifically, common surnames are the 20 most popular surnames in the region, while rare surnames are those held by less than 2% of the population compared to the most popular surname in the region.

Fig. 12 presents the map of Colombia divided by regions and Table 14 summarises the number of surnames by each historical group found in each region, as well as the 10 most popular surnames in each region.

The findings are presented in Fig. 5 and Fig. 6. The patterns observed in most regions are consistent with the national-level results shown in Fig. 4. Specifically, in the Pacifica, Cafetera, Centro, and Amazonas regions, Indigenous and Afro-Colombian groups are underrepresented in high-quality private schools, while elites are significantly overrepresented. This finding is interesting because it shows that even in places where elites had historically little presence the elite groups appear overrepresented in elite-type schools. The results of low-quality schools show that elite surnames do not behave differently from common surnames, that is, elite surnames do not appear significantly underrepresented in low-quality schools. Indigenous surnames are consistently underrepresented in most regions, indicating absence from the entire educational system rather than just high-quality education, while Afro-Colombians appear overrepresented in low-quality education in most regions.

When we compare the behaviour of elites and their persistence in status, a regional pattern of interest appears within modern elites. In the coffee region, the most visible persistence

[23]The Caribe region includes La Guajira, Magdalena, Cesar, Atlantico, Bolivar, Sucre, Cordoba, and San Andres y Providencia. The Pacifica region includes Choco, Valle del Cauca, Cauca, and Narino. The Cafetera region includes Antioquia, Caldas, Quindio, and Risaralda. The Centro region includes Huila, Tolima, Cundinamarca, Boyaca, and Bogota. The Oriente region includes Santander and Norte de Santander. Finally, the Amazonas and Orinoquia region includes Amazonas, Meta, Putumayo, Caqueta, Guaviare, Casanare, Arauca, Vichada, Guainia and Vaupes.

in their high social position is the bankers. While in the central area, the preeminence of the members of the Jockey is evident. This difference is explained by the evolution of these different elite groups in the regions. On the one hand, it was precisely in Antioquia that the modern development of the manufacturing industry began, and the banking system achieved its most significant growth at the end of the 19^{th}century. On the other hand, the central region strongly influenced Bogota's elites, where the Jockey Club was founded and where this institution was closer to the economic and political power at the national level.

Perhaps the most thought-provoking finding is that of the Caribe region, which historically had an important presence of Indigenous and Afro-Colombians. In this region, Indigenous and Afro-Colombians do not appear underrepresented in high-quality private schools, suggesting that these groups are not absent from elite-type schools. It is possible to speculate that in this region ethnic surnames have been more integrated into the elites. Interestingly, this is not the case in the Pacifica or Amazonas regions, even though both of them also had an important presence of ethnic groups.

Figure 5: Relative Representation in regions

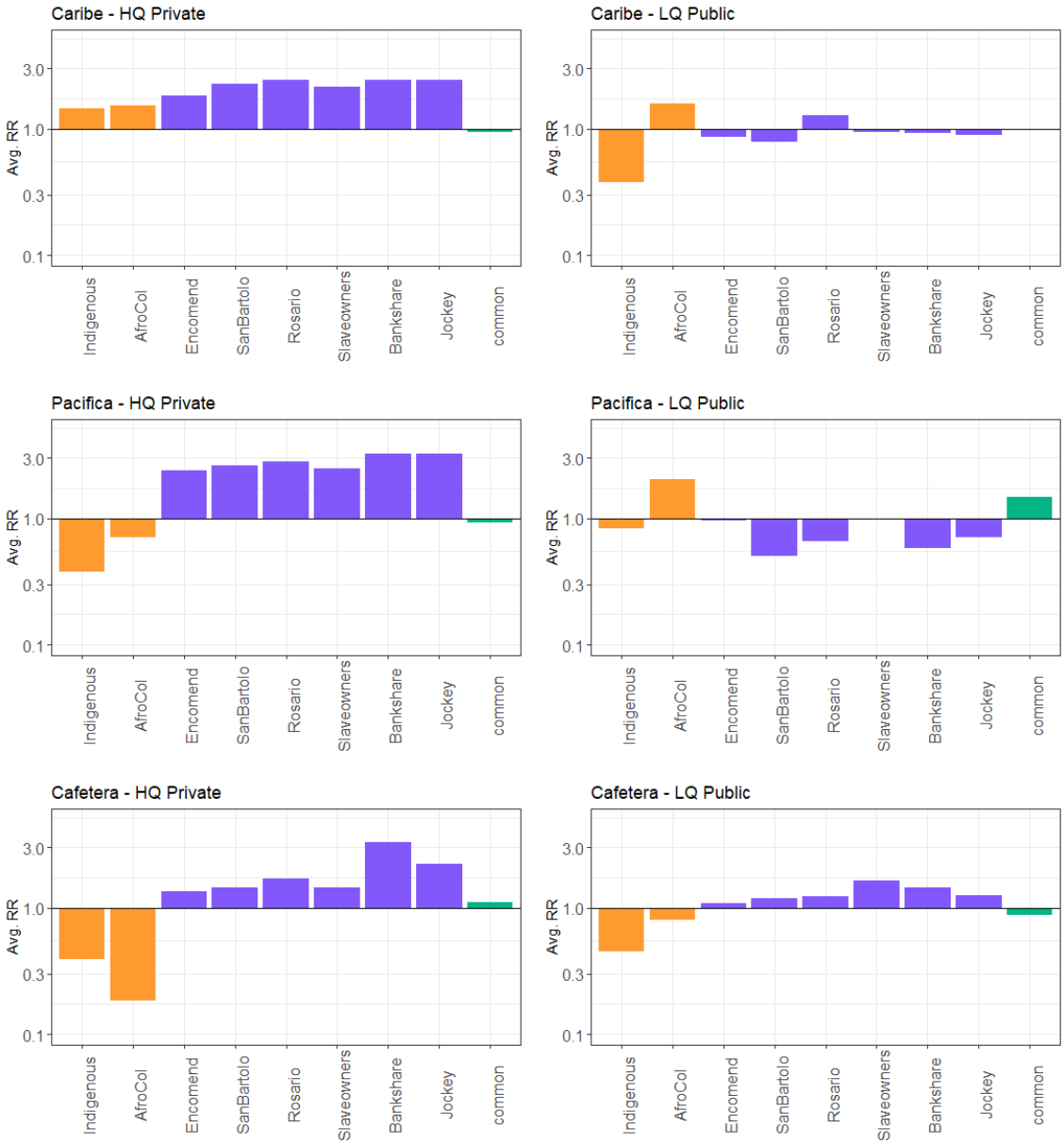
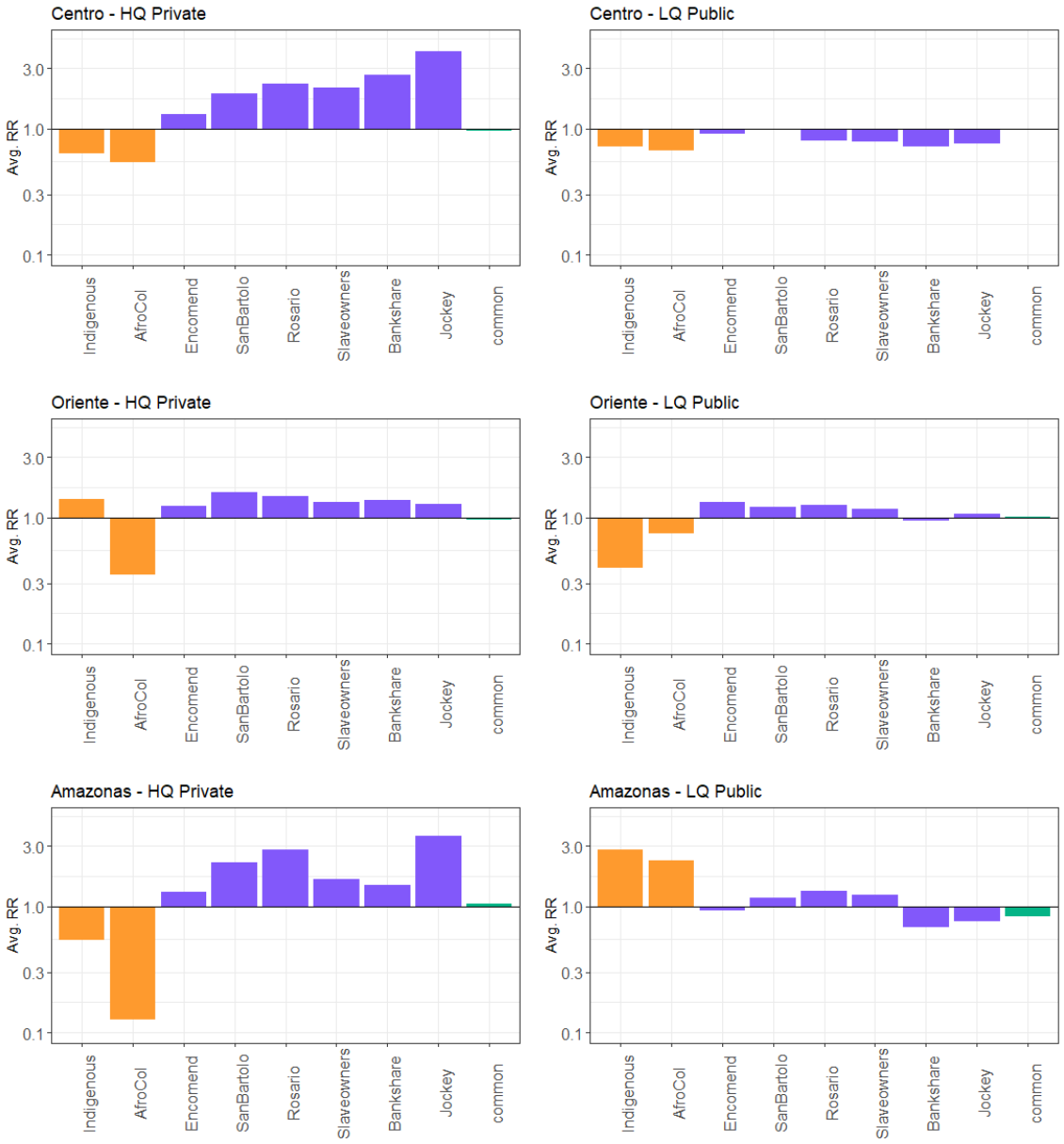


Figure 6: Relative Representation in regions



Notes: The figure shows the distribution of the relative representation by historical group in the access to schools in Bogota. Source: Authors' calculations with Saber 11 and SIMAT.

Taken together, these results indicate that most regions of the country follow a similar pattern to that found at the national level. Historical elite groups have had privileged access to high-quality education while ethnic groups have been systematically absent. However, the patterns are less clear when it comes to low-quality schools, which indicates that educational

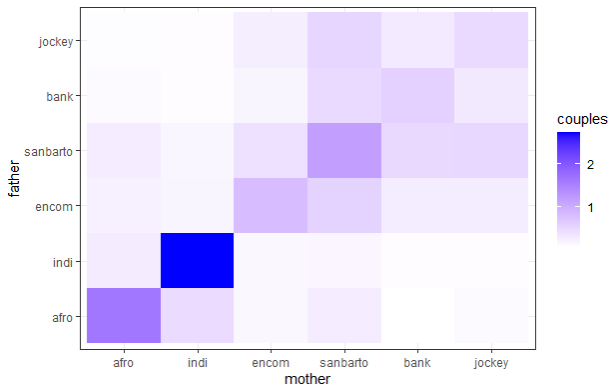
segregation occurs mainly at the top.

5.3 Assortative mating

We first show direct evidence of inter-group marriage patterns by calculating the proportion of paternal-maternal rare surname pairs belonging to each historical group. Fig. 7 shows the results looking at the SIMAT sample. The diagonal of the matrix shows the proportion of intra-group marriage, while the cells outside the matrix show inter-group matches.

As expected, intra-group marriages are more prevalent and this is especially noticeable in the case of the ethnic groups that rarely match with other groups. In particular, ethnic groups seldom match elite groups of the late 19^{th} century. On the contrary, elite groups from any period appear to match frequently with other elite groups. The results show that in the general population, there is high assortative mating within ethnic groups, little matching between indigenous and Afro-Colombians, and high levels of homogamy within modern elite groups.

Figure 7: Who is marrying who: Population (SIMAT)



Notes: The results come from individuals that had two rare surnames identified in our historical groups. The total number of pairs found in each combination is normalised using the number of surnames in each list. Source: Authors' calculations with SIMAT.

To observe if homogamy reflects similar patterns of segregation as the ones shown in Fig. 4, we focus on three different types of access to education: low-quality public schooling, high-quality private schooling and Los Andes University. When we compare the two most extreme types of schools, namely low-quality public and high-quality private schools, the results in Fig. 8a and Fig. 8b show that the marriage market is very different at these two points of

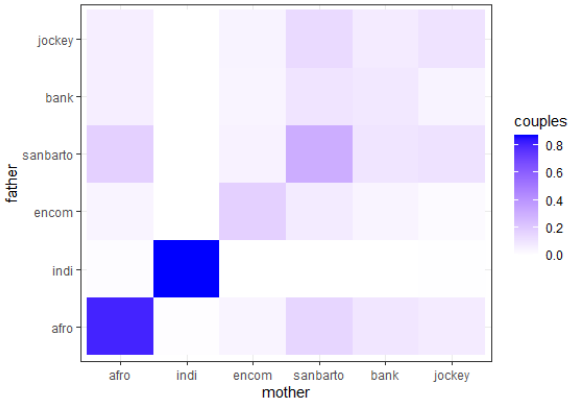
the educational system. The first thing worth noticing is that these results are in line with what we observed in Fig. 4, related to the relative presence of different historical groups. Afro-Colombians and Indigenous appear more in low-quality public institutions, while they seldom appear in high-quality private institutions (the opposite is true for the elite groups).

In low-quality public schools, Afro-Colombians and elite surnames appear frequently matched to surnames from a distinct historical group, although in most cases it is more common to see a match between paternal and maternal surnames from the same group (as shown in the diagonal of the figure). On the contrary, Indigenous surnames show the highest incidence of homogamy as they only appear matched to other Indigenous surnames. The results are somehow similar to the marriage market in the general population, but there is more matching between Afro-Colombians and elite groups in low-quality public schools. Overall, we can conclude that low-quality public schools exhibit some degree of inter-group marriage between elite and Afro-Colombians, but none with Indigenous.

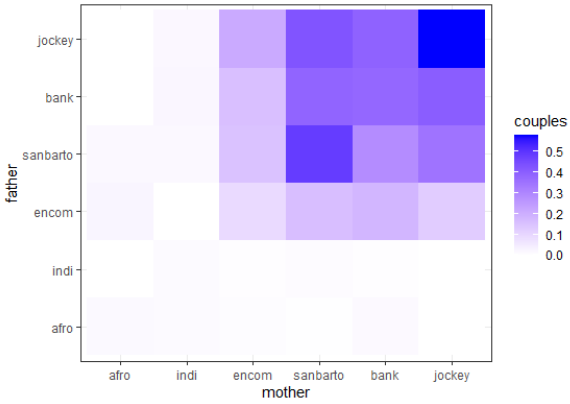
The marriage patterns clearly change when we look at the results in high-quality private schools. Maternal surnames from any elite group are rarely linked to Afro Colombian or Indigenous paternal surnames, and this is mainly because these groups are virtually absent in high-quality private schools. We observe a high frequency of marriage within elite groups which implies that the elite marriage market is very dynamic between elite groups. These figures reveal a striking difference between these two educational conditions. In low-quality public schools, there is more association between Afro-Colombians and elite groups, while in high-quality private schools, there is clear segregation.

Figure 8: Who is marrying who: Public and private schools

(a) Low quality public schools



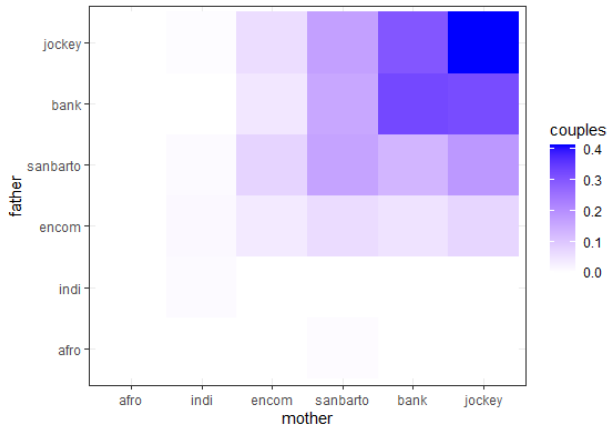
(b) High quality private schools



Notes: The results come from individuals that had two rare surnames identified in our historical groups. The total number of pairs found in each combination is normalised using the number of surnames in each list. Fig. 8a shows the marriage market in low-quality public schools while Fig. 8b represents the marriage market in high-quality private schools. Source: Authors' calculations with Saber 11.

When we repeat the exercise using graduates of Los Andes the results from Fig. 9 confirm that the same patterns are found in elite high schools and elite higher education. Afro-Colombians and Indigenous surnames are practically absent in Universidad de Los Andes, while maternal surnames from the elite are mostly found together with paternal surnames from the elite groups of the late 19^{th} century. The results confirm that segregation in the educational system is also reflected in segregated marriage markets.

Figure 9: Who is marrying who: High quality higher education



Notes: The results come from individuals that had two rare surnames identified in our historical groups. The total number of pairs found in each combination is normalised using the number of surnames in each list. Source: Authors' calculations with Los Andes graduates.

The previous results show that there is a marked trend toward homogamy among Indigenous and Afro-Colombians. Also, the different elite groups appear highly matched to members of the Jockey Club. This is particularly clear when we observe the patterns in high-quality private schools and universities. By contrast, the most interesting result is the way in which most intergroup marriages appear only for children in low-quality public schools.

Table 7 reports the results of the Separable Extreme Value (SEV) index comparing different pairings of groups. Given that assortativeness is a local property the top part of the table reports the index between historical groups that are less likely to appear together, that is elite and ethnic. The second part of the table reports matching our historical groups and the 20 most common surnames. Since the theoretical model underlying this index allows us to estimate the degree of assortativeness of a group in the presence of pairing with a different group, the value of the index is higher when we consider two groups between which there is little incentive to pair.

It is worth keeping in mind that when the SEV index is equal (or close to zero), there is random matching between the groups. Similarly, when the index is greater than zero, there is positive assortative matching or homogamy, that is that paternal surnames are more frequently associated with maternal surnames of their own historical group.

The results confirm the pattern of high assortativeness of the historical groups. Above all the

results show that, even in social situations where elite groups can be paired with one of the ethnic groups, there is little probability of finding this pairing. Also, the results show that in high-quality private schools and in elite higher education there are no matches between Afro-Colombians or indigenous surnames with elite surnames (including pre-modern elites). For this reason, the SEV results in a value that tends to infinity. This is partly explained by the absence of Afro-Colombians and indigenous from these educational institutions. But, even when there is some presence of these groups, as in the case of the indigenous surnames in Los Andes, we found only a few couples with indigenous surnames and common surnames.

There is homogamy in the three types of educational institutions studied. But, there are three findings of great interest when comparing the degree of assortativeness between the institutions. First, as we mentioned above, and confirming the literature on genetic analysis for Colombia, we find that indigenous and Afro-Colombians groups tend not to mix, we could even say that they tend to avoid each other, as shown by the SEV index with the highest value in low-quality public schools (SEV is greater than 9).

Second, modern elites tend to have greater assortativeness than pre-modern elites and it is higher in elite-type institutions than in low-quality institutions. This shows that the pattern of status persistence, that was evidenced in the analysis of relative representation in the previous sections, goes hand in hand with greater homogamy among the groups that manage to maintain their high status. In this way, we find evidence that the barriers that prevent upward mobility of historically excluded groups, especially to achieve high levels of educational social status, is related to the homogamy of these elite socialisation spaces.

Third, the SEV index reports that when matched to common surnames both pre-industrial and modern elite groups have an index close to zero, suggesting randomness in the matching between elite surnames and common across the different educational categories. On the contrary, the Afro-Colombians and the Indigenous appear always with high levels of assortativeness even when compared with the common surnames and the SEV index is always higher than 2.

Table 7: Separable Extreme Value index

	SIMAT	Low quality public schools	High quality private schools	Los Andes
pre-modern elite, Indigenous	3.95	8.05	3.64	∞
pre-modern elite, afro	3.05	1.98	3.28	∞
modern elite, Indigenous	4.75	∞	5.02	∞
modern elite, afro	4.17	3.23	∞	∞
Indigenous, afro	3.25	9.08	∞	∞
pre-modern elite, common	0.27	0.55	0.23	0.44
modern elite, common	0.52	0.75	0.58	1.04
afro, common	3.75	2.63	3.52	∞
Indigenous, common	3.99	7.38	2.28	3.53

Note: The table shows the results from Eq. (3). When matching is random the index will be equal to zero and if there is positive assortative matching then the index will be greater than zero.

Overall, we can conclude that there is virtually no mixing between elite and ethnic groups in Colombia in the educational system, not even in low-quality public schools. These results present novel empirical evidence on the presence of assortative matching in the contemporary marriage market. In particular, these findings suggest that the high levels of homogamy reinforce the persistent segregation of the educational system in contemporary Colombia, especially for ethnic surnames.

5.4 Caveats

There are several cautions to consider in our estimations. First, our sample of rare surnames is defined using a contemporary source of surnames. It could be the case that certain historical groups were more likely to become extinct and therefore our estimations would be biased given that we are selecting a sample of families that behave differently from the population. To minimise this concern we report the percentage of extinguished surnames, or those that appear in the historical record but not in SIMAT in the last column of Table 8.

Table 8: Linking historical groups with contemporary surnames

Historical Group	Number of surnames	Number surnames in SIMAT	Percentage of non matched
Indigenous	131	113	11%
Afro-Colombian	271	65	66%
Encomenderos	363	55	46%
slaves owners	1,607	578	33%
Bank-shareholders	241	66	12%
Jockey Club	309	97	12%
San Bartolome	224	224	34%
U Rosario	711	238	24%

Notes: The first column shows the number of surnames in each list of historical groups. The second column is the number of surnames we define as rare using the number of holders in SIMAT. The last column shows the percentage of extinguished surnames, or those that appear in the historical record but not in SIMAT.

From the table, we observe that encomenderos and the Afro-Colombians seem to have the highest extinction rates, while the rest of the groups are more stable. Following genealogical trees from different sources, we have seen that in the case of the encomenderos the transmission of surnames between parents and children has not always been through the paternal and maternal first surnames, as it is the case today. Up until the end of the 18$^{\text{th}}$ century, surnames could vary among members of the same family, weakening the link between the list of encomenderos and contemporary status measures. Additionally, compound surnames have been transformed and simplified over time. An example of this are the surnames Arias de Monroy or Enciso y Cardenas. The addition of the particle "y" used to be customary to avoid confusion between given names and surnames.[24] Therefore, it is likely that surnames before the late 18$^{\text{th}}$ century are less informative and that the links between generations are weaker.

In the case of the Afro-Colombians, the sources that we use to compile this list do not provide detailed information about the number of holders of the surnames at the time of compilation. Furthermore, part of the surnames come from registers of slave importation, or African enslaved people that entered the port of Cartagena, which means that it is possible that some surnames from the list did not stay in Colombia.[25]

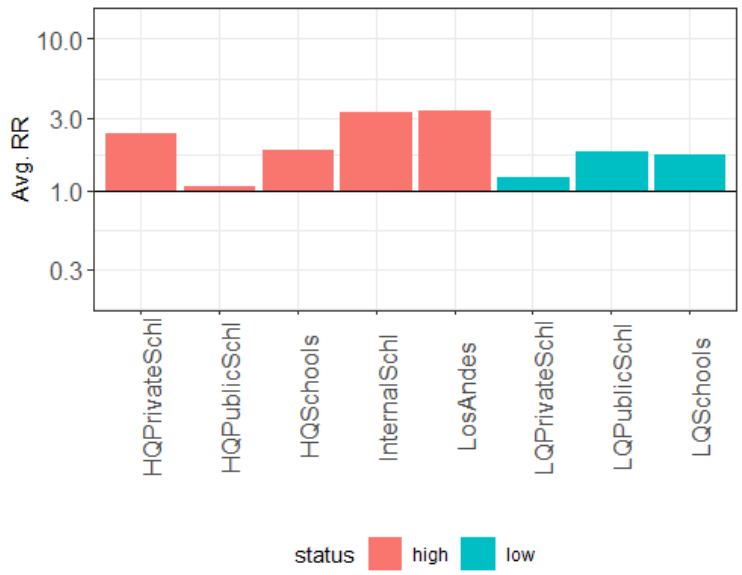
[24] In Spanish, it is common to find surnames, like Ramon, that are also first names. In which case, surnames are separated by the particle as in Rodrigo Ramon y Casales. This rule was also adopted if both surnames have the particle "de", to avoid confusion and to clearly differentiate both surnames, as in Francisco de Goya y Luciente.

[25] We also checked with Census data and found 152 surnames in the 2018 census. Checking the origin of some of the surnames that do not match, we found that several of these surnames appear in other Latin American countries but not in Colombia.

A similar concern is that by looking only at contemporary surnames, we select surnames that were common in the past and only recently became rare. To minimise this concern, we consider several definitions of rarity and report the results in the Appendix. Additionally, we should highlight that the common surnames that we use as our comparison group corresponds to surnames that have been traditionally common even in Spain. The 20 surnames that belong to our commons group end in *-ez* which meant *"son of"*, similar to the ending *-son* in English surnames.

Regarding the results of the slave owners, it is possible that the presence of surnames that also appear in Afro-Colombians can bias the results downwards. To alleviate this concern, we calculate the average relative representation of all surnames from the slave owners list, excluding those that appear with at least one incidence in the 2010 and 2013 waves of the longitudinal survey, ELCA.[26] The results in Fig. 10 show very similar patterns that the original results with all the slave owner's surnames. Overall this group appears overrepresented in both low and high-quality education, indicating that the regression to the mean of this group is not driven by the presence of Afro-Colombian surnames.

Figure 10: Relative Representation in access to education - slaves owners



Note: Fig. 3 presents the average relative representation of the 20 most common surnames in Colombia. The horizontal black line is set at $RR_j = 1$, corresponding to the neutral level of relative representation.

[26]This survey follows approximately 10,000 Colombian households in urban and rural areas of Colombia every three years and reports among other characteristics, ethnicity.

Finally, we have previously discussed that there is a potential measurement error in the status of certain surnames as we are assuming that all groups are homogeneous in status. This seems to be the case in particular with encomenderos and slave owners. However, as shown in Table 9, the number of elite families from the encomenderos and slaves owners that are found both in the pre-industrial and in the industrial elites is lower in comparison to the surnames of the students from the colonial schools. This suggests that both encomenderos and an important part of the slave owners' families were less likely to belong to the modern elite and were unable to reinvent themselves, which could explain the results of regression to the mean.

Table 9: Number of families by period

	Only pre-industrial elite	Only modern elite	Multi-period elite
Encomendero	58	-	8
slaves owners	585	-	87
Urosario	206	-	68
SanBartolome	181	-	49
Bankshareholders	-	17	62
JockeyClub	-	29	88

Notes: The table shows the number of surnames by group that appear in each period. The first column shows the number of surnames in only pre-industrial elites. The second column is the number of surnames that appear only in modern elites. The last column shows the number of surnames that appear in both periods.

6 Conclusions

This research brings novel evidence that shows that social segregation is present in the contemporary educational system in Colombia. We confirm that this segregation reproduces patterns of social exclusion that go as far as the Colonial period. For example, the Spanish Crown defined rules of access to higher education that forced the candidates to demonstrate their connection with family traditions linked to the Spanish nobility. Although the requirements for the purity of blood and nobility were relaxed after the Independence in the early-19$^{\text{th}}$ century, we show that *de jure* segregation in access to higher education continues.

In this research, we assume that rare surnames can be treated as belonging to the same extended family, which allows us to establish links between multiple–generations. Using historical primary sources and ethnographic studies, we identified surnames belonging to

elites and ethnic groups throughout history. We find very sticky floors for the ethnic groups, while the elite of the past has been slowly moving down the social ladder. In particular, the persistent under-representation of historically excluded social groups (Afro-Colombians and Indigenous) in high–quality education is noticeable. The results are similar at the regional level. The main exception was the Caribe region, where both Afro-Colombians and Indigenous appear over-represented in high-quality institutions, suggesting that in this region, ethnic groups moved upwardly.

We find three different mobility patterns in the history of Colombia. First, pre-industrial elite surnames have slowly moved towards the mean status, as these surnames do not appear over or are underrepresented in almost any educational category. Second, elite groups from the late 19^{th} century and surnames that appear in both periods show considerable persistence in their social status, appearing at the top of the social ladder in several markers of status. The elite could move down the social ladder, but this is a very slow process. Third, ethnic groups have the highest persistence of social status among the groups we studied. While we observe some reinvention of the elite, the bottom part of the social status has continued, with very few possibilities to move upward.

Additionally, we present novel empirical evidence on the presence of assortative matching in the contemporary marriage market. In particular, our findings suggest that the high levels of homogamy reinforce the persistent segregation of the educational system in contemporary Colombia, especially for ethnic surnames. We show that there is high assortative matching in the marriage market, particularly in the case of ethnic groups, and this reflects social segregation at the school and university levels. We conclude there is no matching in the marriage market between elite and ethnic groups in Colombia.

While previous literature has shown that extractive institutions such as the *encomienda* and slavery had different long-term effects at the municipal level, we find that the effects were uneven for different social groups. Although the *encomienda* had long-term positive effects on development, it negatively affected the Indigenous population. On the contrary, the social status of the encomenderos has converged to the mean, especially for those families that were unsuccessful in enrolling in new modern activities. Similarly, the literature has shown that slavery had detrimental effects on several socioeconomic outcomes at the sub-national level. Our results uncover that these effects are especially persistent for Afro-Colombians, although the social status of slave owners has regressed to the mean, as in the case of the

encomenderos. These results show that the Indigenous and Afro-Colombians, which were at the bottom of the social ladder during the colonial period, appear at the bottom of the contemporary educational system. The elite status of the encomenderos and slave owners, on the contrary, has not persisted over time. However, the social status of elite groups of the late 19^{th} century and the status of elite families that persisted from the colonial period to the modern one predict contemporary social status today.

References

Acemoglu, D., Bautista, M. A., Querubin, P., & Robinson, J. (2007). *Economic and Political Inequality in Development: The Case of Cundinamarca, Colombia* (Working Paper No. 13208). NBER: National Bureau of Economic Research.

Acemoglu, D., García-Jimeno, C., & Robinson, J. A. (2012). Finding El Dorado: slavery and long-run development in Colombia. *Journal of Comparative Economics*, *40*(4), 534–564.

Acemoglu, D., Johnson, S., & Robinson, J. A. (2005). Institutions as a fundamental cause of long-run growth. In P. Aghion & S. Durlauf (Eds.), *Handbook of economic growth* (Vol. 1, pp. 385–472).

Ager, P., Boustan, L., & Eriksson, K. (2021). The intergenerational effects of a large wealth shock: white southerners after the civil war. *American Economic Review*, *111*(11), 3767–94.

Alvaredo, F., & Londoño Velez, J. (2013). *High incomes and personal taxation in a developing economy: Colombia* (Working Paper No. 12). CEQ: Commitment to Equity.

Alvarez, D. (2018). *Explaining informality: Extractive states and the persistent incentives for being lawless.* (Mimeo)

Angulo, R., Azevedo, J. P., Gaviria, A., & Páez, G. N. (2012). *Movilidad social en Colombia* (Working Paper No. 30). Documento CEDE: Universidad de Los Andes.

Arrazola, R. (1970). *Palenque: primer pueblo libre de América.* Cartagena de Indias: Ediciones Hernández.

Barone, G., & Mocetti, S. (2021). Intergenerational mobility in the very long run: Florence 1427–2011. *The Review of Economic Studies*, *88*(4), 1863–1891.

Behrman, J., Gaviria, A., & Szekely, M. (2001). Intergenerational Mobility in Latin America. *Economia*, *2*(1), 1–31.

Behrman, J., Gaviria, A., & Székely, M. (2003). *Who's in and who's out: social exclusion in Latin America.* Inter-American Development Bank.

Bértola, L., & Ocampo, J. A. (2012). *The economic development of Latin America since independence.* OUP Oxford.

Blanden, J., Doepke, M., & Stuhler, J. (2023). Educational inequality. In E. A. Hanushek, L. Wößmann, & S. J. Machin (Eds.), *Handbook of the economics of education.* Elsevier.

Blossfeld, H.-P. (2009). Educational assortative marriage in comparative perspective. *Annual Review of Sociology*, *35*, 513–530.

Bonet, J., & Roca, A. M. (2007). El legado colonial y el desarrollo regional en Colom-

bia. *Revista de Historia Economica-Journal of Iberian and Latin American Economic History*, *25*(3), 367–394.

Bulmer-Thomas, V. (2003). *The economic history of Latin America since independence*. Cambridge University Press.

Camacho, A., Messina, J., & Uribe Barrera, J. (2017). *The expansion of higher education in Colombia: Bad students or bad programs?* (Working Paper No. 13). Documento CEDE: Universidad de Los Andes.

Cardenas, J.-C., Fergusson, L., & Garcia-Villegas, M. (2021). *La quinta puerta: de como la educacion en Colombia agudiza las desigualdades en lugar de remediarlas*. Bogota: Ariel.

Carrizosa, E. (1990). *Indices de los árboles de las genealogías del Nuevo Reino de Granada de Juan Florez de Ocariz* (Instituto Caro y Cuervo, Ed.). Bogota.

Chetty, R., Friedman, J. N., Saez, E., Turner, N., & Yagan, D. (2017). *Mobility report cards: The role of colleges in intergenerational mobility* (Working Paper No. 23618). NBER: National Bureau of Economic Research.

Chetty, R., Hendren, N., Kline, P., & Saez, E. (2014). Where is the land of opportunity? The geography of intergenerational mobility in the United States. *The Quarterly Journal of Economics*, *129*(4), 1553–1623.

Chiappori, P.-A., Dias, M. C., & Meghir, C. (2020). *Changes in assortative matching: Theory and evidence for the US* (Working Paper No. 26932). NBER: National Bureau of Economic Research.

Clark, G. (2014). *The son also rises: surnames and the history of social mobility*. Princeton University Press.

Clark, G., & Cummins, N. (2013). *Surnames and social mobility: England 1230-2012* (Working Paper No. 181). Department of Economic History: The London School of Economics.

Clark, G., Cummins, N., Hao, Y., & Diaz-Vidal, D. (2015). Surnames: A new source for the history of social mobility. *Explorations in Economic History*, *55*(1), 3–24.

Coatsworth, J. H. (1993). La independencia Latinoamericana: hipótesis sobre sus costes y beneficios. In L. Prados de la Escosura & S. Amaral (Eds.), *La independencia Americana: consecuencias económicas* (pp. 17–30). Alianza.

Colmenares, G. (2007). La Formación de la Economía Colonial (1500 – 1740). In *Historia económica de Colombia* (pp. 13–47). Editorial Planeta.

Correa, J. S. (2009). Banca y región en Colombia, 1850-1880. *Revista de Economía Institu-*

cional, *11*(20), 15–44.

Davila-Ortiz, A. (2010). *El jockey club que yo conoci*. Bogota: Villegas Editores.

Dell, M. (2010). The persistent effects of Peru's mining mita. *Econometrica*, *78*(6), 1863–1903.

Duarte, J., Bos, M. S., & Moreno, M. (2012). *Calidad, igualdad y equidad en la educación colombiana (análisis de la prueba saber 2009)* (Tech. Rep. No. 396). Notas Técnicas: Banco Interamericano de Desarrollo.

Ermisch, J., Francesconi, M., & Siedler, T. (2006). Intergenerational mobility and marital sorting. *The Economic Journal*, *116*(513), 659–679.

Faguet, J.-P., Matajira, C., & Sánchez, F. (2017). *Is extraction bad? encomienda and development in Colombia since 1560* (Working Paper No. 48). Documento CEDE: Universidad de Los Andes.

Fergusson, L., & Flórez, S. (2021a). Desigualdad Educativa en Colombia. In J. C. Cárdenas, L. Fergusson, & M. García-Villegas (Eds.), *La Quinta Puerta: De cómo la educación en Colombia agudiza las desigualdades en lugar de remediarlas* (pp. 81–114). Bogota: Ariel.

Fergusson, L., & Flórez, S. (2021b). Distinción escolar. In J. C. Cárdenas, L. Fergusson, & M. García-Villegas (Eds.), *La Quinta Puerta: De cómo la educación en Colombia agudiza las desigualdades en lugar de remediarlas* (pp. 81–114). Bogota: Ariel.

García Jaramillo, S., Rodríguez Orgales, C., Sánchez Torres, F. J., & Bedoya Ospina, J. G. (2015). *La lotería de la cuna: la movilidad social a través de la educación en los municipios de Colombia* (Working Paper No. 31). Documento CEDE: Universidad de los Andes.

García-Jimeno, C. (2005). *Colonial institutions and long-run economic performance in Colombia: is there evidence of persistence?* (Working Paper No. 7967). Documento CEDE: Universidad de los Andes.

Grafe, R., & Irigoin, M. A. (2006). The Spanish Empire and its legacy: fiscal redistribution and political conflict in colonial and post-colonial Spanish America. *Journal of Global History*, *1*(2), 241–267.

Guell, M., Mora, J. V., & Telmer, C. I. (2013). The informational content of surnames, the evolution of intergenerational mobility, and assortative mating. *Review of Economic Studies*, *82*(2), 693–735.

Guell, M., Rodríguez Mora, J. V., & Telmer, C. I. (2007). *Intergenerational Mobility and the Informative Content of Surnames* (Working Paper No. DP6316). CEPR: Centre

for Economic Policy Research.

Guillén de Iriarte, M. C. (2006). *Los estudiantes del Colegio Mayor de Nuestra Señora del Rosario, 1773-1826*. Bogota: Universidad del Rosario.

Guillén de Iriarte, M. C. (2008). *Los estudiantes del Colegio Mayor de Nuestra Señora del Rosario, 1826-1842*. Bogota: Universidad del Rosario.

Helg, A. (2014). La limpieza de sangre bajo las reformas borbónicas y su impacto en el Caribe neogranadino. *Boletín de historia y antigüedades*, *101*(858), 143–180.

Instituto Colombiano de Cultura Hispanica. (1992). *Geografia humana de Colombia. Region del Pacifico - Tomo IX*. Bogota: Instituto Colombiano de Antropología e Historia.

Jaramillo-Echeverri, J., Álvarez, A., & Bro, N. (2021). *Surnames and social rank: Long-term traits of social mobility in colombia and chile* (Working Paper No. 01). CAF: Development Bank of Latin America.

Jaramillo Uribe, J. (1963). Esclavos y señores de la sociedad colombiana del siglo XVIII. *Anuario Colombiano de Historia Social y de la Cultura*, *1*, 3–62.

Kalmanovitz, S. (2001). *Las instituciones y el desarrollo económico en Colombia*. Bogota: Norma.

Kalmanovitz, S. (2008). *La economía de la Nueva Granada*. Universidad de Bogotá Jorge Tadeo Lozano, Ciencias Económico-Administrativas.

Maloney, W. F., & Valencia Caicedo, F. (2016). The persistence of (subnational) fortune. *The Economic Journal*, *126*(598), 2363–2401.

Marshall, J., & Calderon, V. (2006). *Social exclusion in education in Latin America and the Caribbean* (Working Paper). Sustainable Development Department: Inter-American Development Bank.

Mejía, W. J. (1996). *Real Colegio Mayor y Seminario de San Bartolomé: nobleza e hidalguía; colegiales de 1605 a 1820*. Instituto Colombiano de Cultura Hispánica.

Melo, J. O. (2017). *Historia mínima de Colombia*. El Colegio de Mexico AC.

Mora Bustamante, A. (2016). *Consecuencias a largo plazo de la encomienda en Colombia: persistencia institucional.* (Mimeo)

Mosquera, S. A. (2014). *Afrochochoanos: orígenes y troncos familiares.* Quibdó: Muntú - Bantú: Fundación Social Afrocolombiana.

Narayan, A., Van der Weide, R., Cojocaru, A., Lakner, C., Redaelli, S., Gerszon Mahler, D., ... Thewissen, S. (2018). *Fair Progress?: Economic Mobility across Generations around the World*. The World Bank.

OECD. (2018). *A Broken Social Elevator? How to Promote Social Mobility*. Paris: OECD

Publishing.

Olarte, R. S. (1993). Estructuración y disolución de un sistema de reclutamiento de elites: El caso de Santafé de Bogotá, en Nueva Granada. *Boletín Cultural y Bibliográfico*, *30*(32), 3-23.

Prados de la Escosura, L. (2009). Lost decades? Economic performance in post-independence Latin America. *Journal of Latin American Studies*, *41*(2), 279–307.

Ripoll Echeverría, M. T. (2006). *La élite en Cartagena y su tránsito a la República-revolución política sin renovación social*. Universidad de los Andes, Facultad de Ciencias Sociales.

Salas, A., Acosta, A., Alvarez-Iglesias, V., Cerezo, M., Phillips, C., Lareu, M., & Carracedo, A. (2008). The mtdna ancestry of admixed colombian populations. *American Journal of Human Biology*, *20*(5), 584–591.

Santavirta, T., & Stuhler, J. (2020). *Name-Based estimators of intergenerational mobility: Evidence from Finnish veterans.* (Mimeo)

Schwartz, C. R. (2013). Trends and variation in assortative mating: Causes and consequences. *Annual Review of Sociology*, *39*, 451–470.

Sokoloff, K. L., & Engerman, S. L. (2000). Institutions, factor endowments, and paths of development in the New World. *Journal of Economic perspectives*, *14*(3), 217–232.

Torche, F. (2014). Intergenerational mobility and inequality: The Latin American case. *Annual Review of Sociology*, *40*, 619–642.

Tovar, J. (2007). *La manumisión en Colombia: 1821-1851. Un análisis cuantitativo* (Working Paper No. 28). Documento CEDE: Universidad de Los Andes.

Tovar, J. A., & Tovar, H. (2009). *El oscuro camino de la libertad. Los esclavos en Colombia, (1821-1851)*. Bogota: Ediciones Uniandes.

Valencia Caicedo, F. (2019). The mission: Human capital transmission, economic persistence, and culture in South America. *The Quarterly Journal of Economics*, *134*(1), 507–556.

Velasquez Murillo, R. (1962). Gentilicios africanos del occidente de Colombia. *Revista colombiana de folclor*, *II*(7), 107–149.

Villalobos, Á. R. A. (2005). La limpieza de sangre en el Colegio de Abogados de Caracas a finales del siglo XVIII. *Fronteras de la Historia*(10), 305–328.

Williamson, J. G. (2015). Latin american inequality: Colonial origins, commodity booms or a missed twentieth-century leveling? *Journal of Human Development and Capabilities*, *16*(3), 324-341.

Zapata Olivella, M., & Mina Aragon, W. (2014). *El árbol brujo de la libertad: África*

en Colombia, orígenes-transculturación-persistencia, ensayo histórico mítico. Bogota: Ediciones Desde abajo.

7 Appendices

7.A Description of Historical Groups

In this Appendix we present the historical groups we find in different moments of Colombian history, from the beginning of the colony to the start of the republic. We describe the groups, their economic and social position in history and the sources from which we collected the data.

7.A.1 The Colonial legacy: 16th - 19th centuries

Encomenderos

The *encomienda* was a formal institution that Spain established in different Latin American countries. It was the right of notable Spaniards living in the colony to collect tribute such as gold, fabrics, emeralds, and labour from the Indigenous communities while offering them religious education. In Colombia, the size of one *encomienda* ranged from 100 to 3,900 Indigenous per *encomienda*, and in the regions where there was a higher presence of agricultural societies, there was a higher number of *encomiendas*, creating significant variation in the wealth of an encomendero. The *encomienda* lasted from 1550 to the end of the 17th century when the demographic collapse of the Indigenous population made it unsustainable (Kalmanovitz, 2008). The encomenderos were an heterogeneous group and their social status changed significantly from 1560 to the moment when the *encomienda* was abolished. But these first "recognised" settlers in the colony had access to the best resource available in the colonies and they created the first economic elite of the 17th century. After the collapse of the *encomienda*, forced labour continued to be part of the economic elite of the territory, but other activities like commerce and trade emerged in the colonies, and a political elite was established by the 1750s.

Our first list of surnames comes from an official document written in 1674 by Juan Florez Ocariz. It is an extensive report about the people living in that period in the New Kingdom of Granada. This text, entitled *Genealogia del Nuevo Reyno de Granada*, contains information about various personalities that were related to the conquest and founding of the country.

The book contains detailed lists of members of the military campaigns of occupation of contemporary Colombia and Venezuela, including also some parts of Ecuador and Peru. From Carrizosa (1990) we obtain the names and surnames of the people that appear in the original book with their original spelling.[27] Particularly, we focus on those that appear under the title of *encomendero*.

Indigenous

It is well-known that Indigenous populations in Latin America have been historically marginalised. As discussed above, soon after the arrival and establishment of the Spaniards the *encomienda* was established as a way of assigning Indigenous labour and goods to the Spaniards. For more than one century Indigenous communities had the obligation of sending workers to the mines (mita), the haciendas or the craft shops. After 1590 a series of changes took place in order to guarantee the safety of the communities. Among those changes, the Crown established the prohibition of using Indigenous for personal services or to provide food and crops for the encomenderos. As the reduction of the Indigenous population continued, the Crown decided to import African enslaved people to replace Indigenous labour, particularly in the mines, while organising the Indigenous population in small shelters near the cultivated lands. This division in the territory laid the foundations of a remarkably hierarchical society in which Indigenous groups kept losing part of their original territories. To reverse this process between 1988 and 1990, approximately 20% of the national territory was given to the control of communities that represented less than 2% of the population (Melo, 2017).

In Colombia surnames such as Cárdenas, Tobar o Cuero which are Spanish in their origin, are common in some Indigenous communities. However, other communities have retained their pre-hispanic surnames as Sastoque or Botina. The official use of surnames among Indigenous communities appeared during the 16$^{\text{th}}$ century thanks to the adoption of the institution of *resguardos* in the Viceroyalty of New Granada. Similar to the congregaciones in New Spain, the *resguardos* allowed for a certain autonomy on the exploitation of land for an Indigenous family or group of families outside the territory of an *encomienda*. The main goal of this institution was to maintain control over Indigenous groups. They were forced to be baptised and to adopt a Christian name to which the original given name of the heads of the family, usually men, was added as a surname. On other occasions, surnames were assigned to the

[27] This means that some of the surnames like Ximenez that have been transformed in time to become Jimenez appears in our list of genealogies, but we cannot link them with their contemporary spelling.

whole group based on a topographic characteristic of the region. We can use surnames to track families over time because in most Indigenous groups the patrilineal inheritance in which the children inherit the surnames from the parents is preserved (Instituto Colombiano de Cultura Hispanica, 1992). In this research, we use Indigenous names only, since in this way we can link the surname with its historical social status. We have collected a list of 131 Indigenous surnames from a detailed anthropological study (Instituto Colombiano de Cultura Hispanica, 1992).

Slaves owners

Between the 16th and the 17th century, the Crown started to import African enslaved people to substitute for the declining Indigenous population. By the 18th century the economy of Nueva Granada was based on mining, agriculture and commerce, and in turn, these activities were established around slave labour. However, after independence, the importance of the slave economy started decreasing as new industries emerged, and mining lost importance in comparison with new agricultural products. Likewise, international pressure made the manumission process irreversible. In 1851 Congress issued the law of May 21st, 1851, by which the African enslaved people would be free from January 1st, 1852, and the slaves owners would be compensated with bonds of the state.

Tovar (2007) and Tovar and Tovar (2009) compile the list of the slaves owners who were covered by this law and received compensation from the state. We use this list of 1,667 surnames of slaves owners that received this compensation between 1851 and 1852. This list represents the elite of the Colonial legacy that by 1851 was in decline. Several caveats should be made with this list, and with the social status that these surnames can capture. This list does not give us a consistent social status, as we do not have complete information about the number of African enslaved people for each individual. Tovar (2007) present a list of 30 people who had more than 50 African enslaved people for those who he could assign the number of African enslaved people. For example, in this list, Napoleon Lozano appears with 196 African enslaved people, while Ramon Cabezas with 52. The variance in the number of African enslaved people may indicate variance in social status, as well as variance in the capital received by the state. It is not clear then if this elite continued to be an elite group after they freed the African enslaved people, or if on the contrary, this elite was already in decline. Similarly, an important part of the Colombian historiography suggests that slavery was only profitable in very exceptional cases (Kalmanovitz, 2008).

An additional caveat regarding the social status captured by the surnames is that after the process of manumission many of the African enslaved people took the surname of their owner, and therefore surnames like Mosquera may be related to the black population, a population that has been systematically marginalised in the country.

Afro-Colombians

The Afro-Colombians population in Colombia has its origins, in its immense majority, in the African enslaved people that arrived during the 16^{th} to the beginning of the 19^{th} centuries. The regions historically associated with gold mining and the production of sugarcane and other forms of plantations were the preeminent destination of black African enslaved people, mainly in the departments of Cauca, Antioquia, Chocó and Bolivar (Jaramillo Uribe, 1963). During the colony, the African enslaved people were at the bottom of the social ladder, and flights and rebellions were common. After the Independence process and the definitive abolition of slavery in 1851, Afro-descendant populations founded several communities in the vicinity of the Caribbean and Pacific coasts. A considerable part of the African enslaved people adopted the surnames of their owners at the moment of gaining freedom, but the population of these regions sought to reconstruct their African ancestral identities, or at least avoid identification with the former slaves owners. These populations thus adopted surnames that corresponded to topographical surnames, ethnic names or designated physical characteristics of their ancestors before being Christianised or officially being given a Spaniard or Mestizo surname.

Therefore, in Colombia, a set of surnames that are properly Afro-descendant and that have almost zero frequency of occurrence among non-Afro-Colombian populations remain to this day. Using the main anthropological studies that have sought to define inventories of these typically Afro-descendant surnames, we compiled a list of 276 of them. We use two lists of surnames compiled by afrocolombianists, built upon a set of past studies (Zapata Olivella & Mina Aragon, 2014; Mosquera, 2014). They discuss and group different ethnographic scholarly works, in particular the seminal contributions on this subject by Velásquez-Murillo (1962) and Arrazola (1970). We then compared these lists of surnames with our other groups to avoid coincidences with surnames of slave–owners (as the very frequent case of the surname Mosquera).

7.A.2 Elite schools and universities: 17th century

During the Spanish colonial period (mid-16th to mid-19th centuries) and until the republican reforms of the mid-19th century, the access to high school and higher education was ruled by a selection process based on noble heritage and "purity of blood". Only the heirs of Spanish families, and in particular those with noble privileges were admitted to the "colegios mayores" (higher education). The selection process was based on the examination of candidates' files, and recommendation letters from prestigious acquaintances of the candidate. An important part of those archival materials was conserved. Two of the most important, and the most ancient, institutions of the Colonial era exists today: *Colegio Mayor de San Bartolomé* and *Colegio Mayor de Nuestra Señora del Rosario*.

Colegio San Bartolomé

The *Colegio San Bartolome* is a private primary and secondary school founded in Bogota in 1604, and since these times it has been a flagship in high-quality education. It was founded by archbishop Bartolome Lobo Guerrero to educate Spaniards, born out of a legitimate marriage, who already knew how to read and write. In particular, the descendants from Conquistadores were preferred (Mejía, 1996). It is the oldest school in Colombia and one of the most prestigious institutions from which 28 presidents of Colombia have graduated.

We collected a list of 224 students of the Colegio San Bartolome during the colonial period until the Independence, using the book *Real Colegio Mayor y Seminario de San Bartolome, 1605 to 1820* published by the Instituto Colombiano de Cultura Hispanica in 1996. This book accounts for the history of the school from its foundation until the Colombian Independence.

Universidad del Rosario

The *Colegio Mayor de Nuestra Señora del Rosario*, known today simply as Universidad del Rosario, is a higher education institution founded in Bogota in 1653. It is one of the oldest universities in Latin America and the oldest with uninterrupted functioning in Colombia since Colonial times. From the 17th to the 18th century Universidad del Rosario was the most prestigious higher education institution in the country. It was ruled under the Dominican Order and proof of blood purity or nobility (*hidalguia*) was necessary for gaining access to it. As a result, this institution was the main educational centre of the ruling elites during colonial times. Philosophers, lawyers, theologians and physicians were formed at this university.

We established a list of 5,482 male students of the Universidad del Rosario using a secondary source: a two volumes work by Maria Clara Guillen de Iriarte (Guillén de Iriarte, 2006, 2008) who reports a list of all the graduates from 1773 until 1842. This work contains information about the students, their parents and eventually their patrons. The books offer a reliable transcription of the full set of manuscript archival sources of curricula of graduates.

7.A.3 The beginning of modernity: late 19th and early 20th century

Jockey Club

By the end of the 19th century, aiming at emulating the British elite's practices, the Colombian elite, in particular in Bogota, founded social and sports clubs. One of the most exclusive social clubs is, still today, the *Jockey Club*.[28] The purpose of founding the Jockey Club was to create a private space for social gatherings for its members. When the Jockey Club started, "the members of the Club were businessmen, renters and heirs, partially dedicated to the agriculture of grains (wheat and barley), coffee growers, milkmen, promoters or entrepreneurs of public services, senior politicians, some humanists, literati, lawyers, doctors and a few other professionals" (Davila-Ortiz, 2010). This club has been one of the most exclusive institutions in Colombia since its foundation. Currently, the Jockey Club continues to be a flagship institution of the Colombian elite. There are no memberships available, and those who want to be part of the group must wait for a member to retire or die without leaving an heir to take their place, which is indicative of the exclusiveness of the club. Even more, the number of members who can make decisions and be elected to management positions is limited to only 600 people.

We collected a list of 811 individuals who were associated with or were founders of this elitist (and all-male) institution. From the founding acts of the Jockey Club, we obtained a list of the 123 men that participated in its foundation on September 20th, 1874. Also, we have the list of the associates in 1875, and the following "re-foundations" acts of 1888, 1890 and 1902.

Bank founders

[28] This social club and its members, had such influence on the country's politics and economic development during the 19th and 20th centuries that some called it "the other Casa de Nariño", meaning the other presidential house. This wordplay comes from the fact that the President's residence is named after one main figure of the independence process: Antonio Nariño and the original building where the Jockey Club was established was Nariño's house.

A period of free banking took place in Colombia between 1870-1885. During this period private banks could issue banknotes, backed by their reserves of silver and gold. In total 36 banks were established in different regions of the country. These banks were opened mainly by merchants from some regions in search of more liquidity. They also intended to link the banking system with local and international trade. Likewise, "the creation of banks in the most prosperous regions of the country allowed the regional and local elites to control the monetary and credit system, consolidating their power over the financial market" (Correa, 2009).

We have a list of the founders of the first 12 banks established in the country, between 1870 and 1883, from the foundation letters of each firm. This list includes the 3 biggest banks of the period: *de Colombia*, *de Bogota* and *de Antoquia*. The former two are still among the biggest private banks today. So we can observe a business elite of the period, as these firms were the first forms of stock-holding corporations in the country and signal the first steps of economic modernity well before industrial development (towards the 1920s).

7.B Description of contemporary educational data

To define contemporary access to education we construct several educational categories (e.g. high and low-quality education, high prestige high schools and higher education institutions). We provide a short description of each dataset and the variables we use to define access to education under consideration in our analysis of the persistence of social status.

The first dataset is an administrative register of the SABER 11. It contains information at the individual level compiled from the registration forms filled out by students who finish high school and take a standardised test, known as SABER 11, similar to the SAT in the USA or the A-levels in Europe. This test is a control of the quality of the educational attainments of students and serves also as a college admission test. It is administered and applied by the ICFES (Instituto Colombiano para el fomento de la educación superior). We use information from these records for the years 1996 to 2016 on an annual basis. The database contains a total of 12'456,376 individuals. Fig. 11 shows the number of students enrolled in the last year of secondary school by the historical group from 1996 to 201. There are no considerable differences in the trends of the historical groups, but the differences in levels are significant. The ratio between the number of students in the third grade of primary school and the number of students enrolled on the last year of secondary schooling in 2016

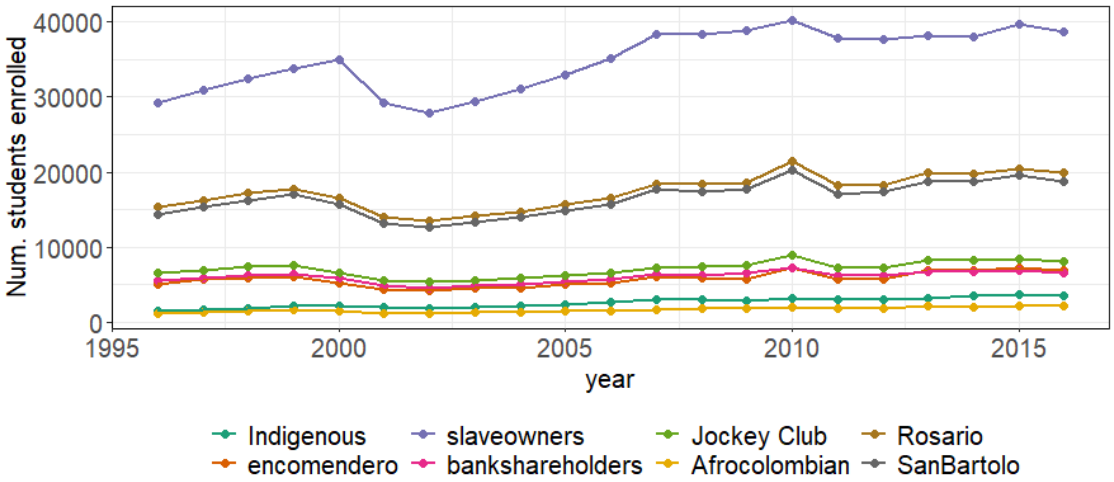
is around 35%, as shown in Table 10.

Table 10: Completion rates by historical groups.

Historical group	Num students enrolled SIMAT, 2016	Num students enrolled SABER 11, 2016	Saber11/SIMAT (%)
Indigenous	10,685	3,577	33.48
Afro-Colombian	6,816	2,097	30.77
Encomenderos	17,487	6,943	39.70
slaves owners	108,386	38,606	35.62
San Bartolome	50,471	18,700	37.05
Rosario	53,641	19,877	37.06
Bankshareholders	16,711	6,618	39.60
Jockey Club	21,426	8,086	37.74
Common surnames	395,654	134,203	33.92

Note: The second column shows the number of students enrolled in third grade of primary schooling in 2016 by historical groups. The third column is the number of the number of students enrolled in the last year of high school in 2016 by historical groups. The last column is the proportion between column 2 and 3 as a way to calculating the share of students in high school in comparison to primary schooling.

Figure 11: Students enrolled in high school, 1996-2016



Note: The figure shows the number of students enrolled in the last year of high school by historical group using rare surnames, from 1996 to 2006. Source: SABER 11.

We use the school-level scores in the standardised tests to define High or Low-quality schools. We label these types of schools as High or Low-Quality Schools in general (HQRanking, LQRanking) or High-Quality Private or Public Schools (HQPrivateSchool, HQPublicSchool).[29] The distinction between private and public is crucial in a country like Colombia where private schooling covers a significant proportion of the schooling supply.

[29]It is important to note that a significant part of the population does not belong to any of those extremes. We choose to create such an exclusive set of schools because we are aiming at understanding the persistence and the mobility of elites and ethnic groups along the social ladder.

Besides the high-school institutions' quality defined above, we also use the dataset SABER 11 to identify a marker of an elite education: the UNCOLI schools, for Union de Colegios Internacionales (International High–Schools Union). This union groups the most prestigious, selective and expensive schools in the country, located in Bogota.

Finally, Los Andes is a private institution founded in Bogotá in 1948 and was established to be independent of any political party or religious institution. Since then it has been considered a particularly elite institution. Since 1950, less than 0.03% of people between 20 and 30 years have graduated from this university. Of the 363,039 students who registered in a university in 2016, 2,940 enrolled in Los Andes, that is, only 0.8%. Nevertheless, the last government was composed of 33% of graduates from Los Andes. Graduates from this university are the most represented in high-level policy-making. Also, alumni of Uniandes (as known in Colombia) are over-represented among the private managers of the biggest companies in the country.

7.C Description of relative representation by educational outcomes

Table 11: Summary statistics of the relative representation by historical groups

Educational category	Min.	1stQu.	Median	Mean	3rdQu.	Max.
Uniandes	0	0	0	1.12	0.65	121.09
Uncoli - International schools	0	0	0	1.16	0	211.28
HQ Universities	0	0.18	0.61	0.83	1.14	29.57
HQ Ranking	0	0	0.36	0.99	1.16	46.46
HQ Public school	0	0.09	0.52	0.91	1.10	40.47
HQ Private school	0	0	0	1.03	0.95	64.06
LQ Ranking	0	0	0.15	1.21	1.37	164.61
LQ Public school	0	0	0	1.23	1.19	191.20
LQ Private school	0	0	0	0.86	0.97	54.54

Source: Authors' calculations with Los Andes graduates, SIMAT, and SABER 11.

7.D Robustness: Different thresholds of rarity

To address potential selection in our rare surnames we consider other definitions of rarity. First, we include all the surnames. The results show that the general findings hold, but, as expected, the significance and explanatory power of the regressions fall as the predictive

power of frequent surnames is lower than the predictive power of rare surnames. Then, we impose a more restrictive definition of rarity (1 % of the most common surname). The results reveal that the patterns remain but the significance of every coefficient improves as well as the overall fit of the regression.

Table 12: All surnames. Differences in educational social categories - OLS z-scores

	Los Andes	International schools	High Quality Schools	HQPrivateSchool	Low Quality Schools	LQPublicSchool
Ethnic groups						
Indigenous	−0.632***	−0.511***	−0.618***	−0.578***	−0.232**	−0.121
	(0.105)	(0.106)	(0.106)	(0.106)	(0.106)	(0.106)
AfroColombians	−0.558***	−0.452***	−0.584***	−0.519***	0.102	0.095
	(0.112)	(0.112)	(0.112)	(0.112)	(0.113)	(0.113)
Pre-industrial elites						
Encomendero	−0.383***	−0.297***	−0.301***	−0.344***	−0.039	−0.027
	(0.080)	(0.080)	(0.080)	(0.080)	(0.081)	(0.081)
Slaveowners	−0.390***	−0.332***	−0.258***	−0.257***	0.036	0.078
	(0.055)	(0.055)	(0.055)	(0.055)	(0.055)	(0.056)
URosario	0.008	−0.099*	−0.029	−0.115**	0.002	−0.017
	(0.056)	(0.056)	(0.056)	(0.056)	(0.056)	(0.056)
SanBartolome	−0.401***	−0.259***	−0.212***	−0.257***	−0.038	−0.116**
	(0.059)	(0.059)	(0.059)	(0.059)	(0.059)	(0.059)
modern elites (late 19th century)						
Bankshareholders	0.097	0.004	0.121	0.234***	0.068	−0.063
	(0.081)	(0.082)	(0.082)	(0.081)	(0.082)	(0.082)
JockeyClub	0.831***	0.780***	0.503***	0.660***	−0.103	−0.057
	(0.074)	(0.074)	(0.074)	(0.074)	(0.074)	(0.074)
No Group	−0.538***	−0.424***	−0.492***	−0.449***	−0.157***	−0.087
	(0.054)	(0.054)	(0.054)	(0.054)	(0.055)	(0.055)
Constant	0.521***	0.412***	0.473***	0.433***	0.145***	0.081
	(0.054)	(0.054)	(0.054)	(0.054)	(0.054)	(0.054)
Observations	23,144	23,144	23,144	23,144	23,144	23,144
Adjusted R^2	0.014	0.009	0.010	0.010	0.002	0.001

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Note: The equations are estimated using OLS.
The 20 most commons surnames is the latent group of comparison.
The RR_j is normalised.

Table 13: Rare surnames : 1% of Rodriguez. Differences in educational social categories - OLS z-scores

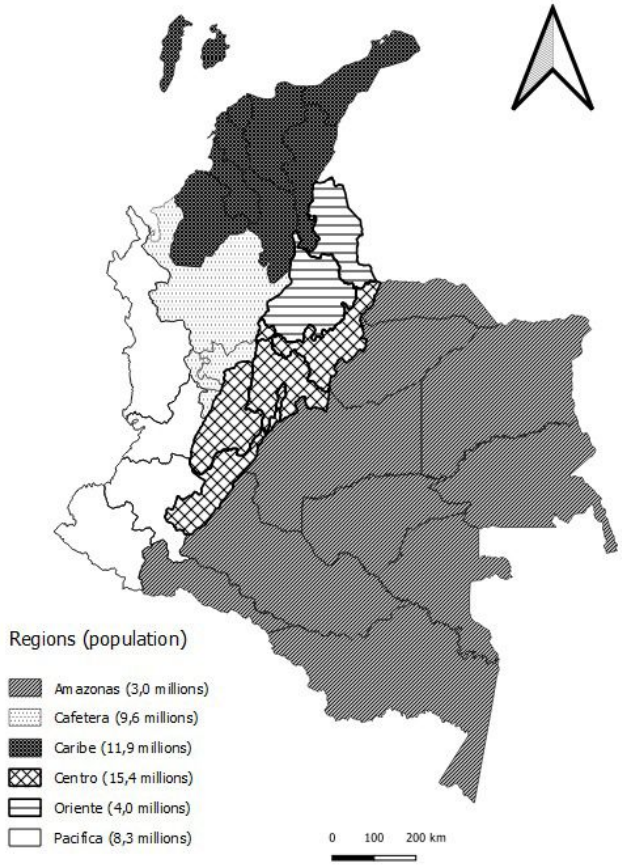
	Los Andes	International schools	High Quality Schools	HQPrivateSchool	Low Quality Schools	LQPublicSchool
Ethnic groups						
Indigenous	−0.484***	−0.325***	−0.456***	−0.384***	−0.293**	−0.178
	(0.120)	(0.121)	(0.122)	(0.121)	(0.123)	(0.123)
AfroColombians	−0.468***	−0.317**	−0.496***	−0.417***	0.207	0.173
	(0.135)	(0.137)	(0.137)	(0.137)	(0.139)	(0.139)
Pre-industrial elites						
Encomendero	−0.313**	−0.224	−0.122	−0.223	0.116	0.143
	(0.159)	(0.160)	(0.161)	(0.160)	(0.163)	(0.163)
Slaveowners	−0.045	−0.014	0.039	0.095	0.014	0.125
	(0.076)	(0.077)	(0.078)	(0.077)	(0.078)	(0.078)
URosario	0.325***	0.263***	0.208***	0.218***	0.078	0.017
	(0.079)	(0.080)	(0.080)	(0.080)	(0.081)	(0.081)
SanBartolome	−0.293***	−0.202**	−0.010	−0.073	−0.086	−0.200**
	(0.087)	(0.088)	(0.088)	(0.088)	(0.089)	(0.089)
modern elites (late 19^{th} century)						
Bankshareholders	0.117	−0.101	0.381***	0.557***	0.094	−0.046
	(0.137)	(0.138)	(0.139)	(0.138)	(0.140)	(0.140)
JockeyClub	1.679***	1.536***	0.663***	0.873***	−0.111	−0.066
	(0.112)	(0.113)	(0.113)	(0.113)	(0.114)	(0.114)
No Group	−0.353***	−0.219***	−0.276***	−0.226***	−0.158**	−0.062
	(0.077)	(0.077)	(0.078)	(0.078)	(0.079)	(0.079)
Constant	0.309***	0.186**	0.240***	0.189**	0.142*	0.053
	(0.076)	(0.077)	(0.077)	(0.077)	(0.078)	(0.078)
Observations	9,614	9,614	9,614	9,614	9,614	9,614
Adjusted R^2	0.053	0.035	0.023	0.029	0.003	0.002

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Note: The equations are estimated using OLS.
The 20 most commons surnames is the latent group of comparison.
The RR_j is normalised.

7.E Regions, surnames and historical groups

Fig. 12 shows the map of Colombia divided by 6 regions based on both geographical and historical proximity. The Caribe region includes La Guajira, Magdalena, Cesar, Atlantico, Bolivar, Sucre, Cordoba, and San Andres y Providencia. The Pacifica region includes Choco, Valle del Cauca, Cauca, and Narino. The Cafetera region includes Antioquia, Caldas, Quindio, and Risaralda. The Centro region includes Huila, Tolima, Cundinamarca, Boyaca, and Bogota. The Oriente region includes Santander and Norte de Santander. Finally, the Amazonas and Orinoquia region includes Amazonas, Meta, Putumayo, Caqueta, Guaviare, Casanare, Arauca, Vichada, Guainia and Vaupes. The map also includes the population of each region based on the population census of 2018.

Figure 12: Colombia regions by population



Notes: The map shows Colombia divided in six regions and included an estimation of the population of each region. Source: Authors' calculations with 2018 population census from DANE.

Table 14 summarises the number of surnames by each historical group found in SIMAT in each region, as well as the 10 most popular surnames in each region.

Table 14: Historical and common surnames by regions

	Amazonas	**Cafetera**	**Caribe**	**Centro**	**Oriente**	**Pacifica**
Top 10 common	Rodriguez, Gomez, Garcia, Lopez, Gonzalez, Hernandez, Sanchez, Perez, Martinez, Ramirez	Lopez, Gomez, Garcia, Ramirez, Giraldo, Gonzalez, **Restrepo**, **Zapata**, Hernandez, Perez	Martinez, Perez, Rodriguez, Hernandez, Gonzalez, Diaz, Lopez, Garcia, **Torres**, Gomez	Rodriguez, Gonzalez, Sanchez, Martinez, Gomez, Garcia, Hernandez, Ramirez, Lopez, **Rojas**	Rodriguez, Gomez, Garcia, Perez, Hernandez, Martinez, Sanchez, Diaz, Ramirez, **Quintero**	**Valencia**, **Mosquera**, Lopez, Garcia, Gomez, Rodriguez, Sanchez, Martinez, Gonzalez, **Caicedo**
Indigenous	89	57	32	98	27	73
AfroColombians	55	38	41	60	26	69
Encomenderos	65	83	89	75	64	67
Slaves owners	595	591	678	719	522	541
U Rosario	272	289	312	307	229	246
San Bartolome	219	221	245	256	199	195
Bankshareholders	76	78	110	101	76	76
Jockey Club	101	118	139	135	103	105
Rare no group	6500	4455	8432	8557	3402	6414

Note: The text in bold highlights common surnames that are common only in that region. Source: Source: Authors' calculations with SIMAT.