Caminante, sí hay Camino:

Networks in the First Transatlantic Migration, 1492-1540*

Leticia Arroyo Abad
CUNY & CEPR

José-Antonio Espín-Sánchez
Yale University & CEPR

Yannay Spitzer
Hebrew University of Jerusalem & CEPR

Ariell Zimran
Vanderbilt University & NBER

[PRELIMINARY AND INCOMPLETE: PLEASE DO NOT CITE]

Abstract

Between 1492 and 1540, over 21,000 Spaniards moved to the Americas. We study the patterns of the first transatlantic migration (1492–1540) using a novel individual-level dataset from archival and secondary sources. The records identify each migrant's origin locality, destination region, migration year, and occupation.We link each migrant to a new municipality dataset that measures population, taxes, social structure, geography, and travel costs to Seville. We document sharp shifts in destinations as the imperial frontier moved from the Caribbean to *Tierra Firme*, Mexico, and Peru, and we show that aggregate migration rose in surges rather than smoothly. Emigration concentrated in Castile and around Seville, the sole legal port of embarkation, and it concentrated in urban places. Networks shaped both the scale and the direction of migration. Sending regions "locked in" very early. Places that sent migrants in the 1490s and early 1500s kept sending far more migrants later. Migrants also followed origin-specific pathways: they disproportionately chose the same destinations, often sailed on the same ships, and relied on the same notaries to contract passage. We interpret this early lock-in as the interaction of contingency in the first expeditions, durable social ties within Spain, and a strong institutional bottleneck that routed migration through Seville.

JEL: D85, F22, F54, J15, J61, N34, N36, O15, Z13
Keywords: Migration; Networks; Colonization; Spain; Latin America

*Thanks to Andre Costa, Vidhi Maheshwari, Miguel Angel Lafuente Navarro, Fernanda Serna Godoy, and Shneyor Shprinstin for their excellent research assistance. We would like to thank Guillermo Esteban, Alejandra Irigoin, and Noel Maurer for their valuable input and to the participants to the Economic History seminars at George Mason University, University of Bocconi, Sapienza University, Northwestern University, Manchester University, Texas Tech University, Brown University, University of Colorado - Boulder, New York University, Universidad Pablo de Olavide, and the University of Zaragoza. We are also grateful for the comments from participants at the the CEPR & Bank of Spain conference in Madrid, the Economic History Association meetings in Sacramento, the LACEA - RIDGE workshop in Santiago de Chile, the CLADHE conference in Montevideo, the Groningen Growth and Development Centre conference, Economic History Association of Israel, and the World Economic History Association meeting in Lund. All views and errors are our own. We acknowledge the financial support of the Yale Economic Growth Center, Yale MacMillan Center, CUNY PSC Grants, and the National Science Foundation.

"To serve God and His Majesty and become rich"
Bernal Díaz del Castillo
The True History of the Conquest of New Spain, 1568

1 Introduction

In 1492, a single expedition left Spain headed by Columbus; within five decades, over 21,000 Spaniards followed. These migrants—expedition leaders and followers, sailors and artisans, clerics and merchants, servants and families—supplied the human capital that sustained early imperial expansion and shaped the initial composition of colonial societies. Yet we still know little about the basic contours of this first transatlantic migration: who moved, where they came from, where they went in the Americas, and how early social ties and institutions structured those choices. We address that gap by bringing the theory and tools of migration economics to Spanish migration to Spanish America from 1492 to 1540. We use newly assembled individual records linked to characteristics of municipalities and districts of origin. We construct a comprehensive quantitative account of who left Spain, where they came from, and how flows took shape across this first transatlantic migration. This setting offers an unusual opportunity to study network formation at its point of origin: we observe the first pioneers and the full early migrant network under extreme informational frictions, when communication across the Atlantic moved slowly and explorers' networks—rather than well-known destination opportunities—often anchored origin–destination links. By tracing how these early social ties interacted with the Crown's system for legal emigration, we explain why migration clustered so sharply in a subset of Spanish regions and why those clusters persisted even as the imperial frontier expanded.

We build the microdata by digitizing Boyd-Bowman's geographical index on migrants to the Americas, and expand it with hand-collected migration records from the *Archivo General de Indias* (AGI) in Seville and contracts of carriage from Seville's notarial archive. Crucially, we also transcribed all available passenger registers (*Asientos*) through 1540 and added the migrants from the four Columbian voyages (1492, 1493, 1498, 1502) and the Ovando colonizing expedition (1502), all absent from prior compilations. This expanded coverage is essential: it fills the very earliest years of exploration and gives us the first pioneers—the expedition leaders and crew members whose recruitment decisions seeded the networks that followed. Without these records, the moment of network formation would remain invisible. We code each migrant's place of origin, destination, occupation, year, and social status, and pair the individual records with a new dataset on Spanish municipalities covering population, tax assessments, social structure, and geographic characteristics.

A first set of results describes the temporal and spatial organization of these flows. On the destination side, migration tracked the moving imperial frontier. The Caribbean absorbed nearly all departures in the first two decades. Tierra Firme—roughly present-day Central America, Venezuela, and Colombia—entered as a destination in the 1510s. Mexico became central in the 1520s, and Peru, alongside short-lived settlement in the Southern Cone, rose sharply in the 1530s. Aggregate outflows were modest through this early period: only 3,435 migrants left in the first two decades. After that, outflows surged in waves that followed the opening of each new colonial foothold, culminating in a sustained takeoff during the 1530s when nearly 11,000 migrants—over half our sample—left Spain in a single decade. The destination mix shifted dramatically over these fifty years, but the geography of sending regions within Spain hardened early and barely changed thereafter. Over 90% of all migrants came from the same handful of Spanish regions in the 1530s as in the 1500s, even as total outflows tripled.

The migrants who made this journey were not drawn from any single rung of society. We find that migration cut across the full social hierarchy—*hidalgos* and nobles traveled alongside artisans, merchants, clergy, and servants. This first transatlantic flow was not, as a simple push story might suggest, a movement of the desperate poor seeking any escape. It was a broad social phenomenon in which the decision to migrate reflected access to information and networks as much as economic necessity.

The core contribution of the paper concerns why origins were so persistent. We show that social networks strongly shaped the flow even under severe informational and logistical constraints, and even when migrants moved toward destinations that few Spaniards had previously seen. The geography of sending regions locked in extraordinarily early: locations that had already sent migrants in the 1490s or early 1500s continued to send far more migrants thereafter. By 1500–1504, already-exposed districts sent migrants at rates about 15 times higher than non-exposed districts; by 1525–1529, that multiple approached 30. Origins also predicted destinations. Conditional on nationwide shifts in destination attractiveness, migrants disproportionately chose destinations that prior migrants from their own origin had chosen, consistent with destination-specific chain migration. The role of pioneers is central to this story. Expedition leaders recruited actively from their home towns, and our expanded dataset—which includes the earliest voyages and the contracts of carriage—lets us trace this process from its very first departures. We find that the arrival of a first explorer from a district predicts a sharp and persistent rise in subsequent emigration from that district, a pattern that would be invisible without coverage of these early years. The networks operated at a remarkably fine geographic scale: migrants from the same town were significantly more likely to travel on the same boat and to sign their passage contracts with the same notary, consistent with coordination through personal ties rather than

impersonal market intermediaries.

This persistence does not reflect a rich set of observable push fundamentals. Within Castile, once we account for proximity to Seville, observable local characteristics—income, taxes, soil quality, elevation—explain relatively little of the initial spatial pattern, which makes the subsequent persistence even harder to attribute to fixed local advantages. Instead, we interpret the early lock-in as the outcome of an interaction between contingency in the earliest expeditions, pre-existing social connectivity within late-medieval Castile, and a powerful institutional bottleneck. During our period, the Crown routed legal emigration through a single port and a centralized licensing system centered on Seville. Access to the transatlantic stream depended heavily on connections to the Seville-centered administrative and commercial apparatus. In this environment, networks could deepen within connected regions without spreading widely to regions that started outside the system.

We provide suggestive evidence on the historical roots of those pre-existing connections. The early sending belt aligns with the geography of participation in the final campaigns of the *Reconquista* and with modern markers of long-run social connectivity (including genetic and linguistic boundaries), which we use to proxy for durable interpersonal and community ties that predate overseas migration. This perspective reframes a classic question in the networks literature. Work on later mass migrations often emphasizes diffusion—networks spread outward over time and gradually draw new regions into participation (Spitzer and Zimran 2023). Our evidence instead highlights a setting where institutions constrained entry so strongly that networks formed quickly and then remained geographically frozen.

This paper makes three contributions. First, we bring systematic quantitative analysis to one of the most consequential migration episodes in human history—the founding movement of European settlement in the Americas—whose basic contours have remained poorly understood. Second, we construct a new individual-level dataset that substantially expands prior compilations, adding the earliest voyages, notarial contracts, and expedition records that make it possible to observe network formation from its very first moments. Third, our results show that the networks channeling Spaniards to the Americas were not built by the migration itself. They were inherited from a pre-existing social infrastructure rooted in the historical Kingdom of León and the Crown of Castile—visible today in the genetic and linguistic geography of the Iberian Peninsula—and frozen in place by the institutional bottleneck of the Seville monopoly.

1.1 Literature review

The quantitative study of early Spanish migration has relied on a small set of foundational compilations. Chaunu and Chaunu (1955) and Mörner (1975) inferred migration flows indirectly from shipping capacity. Boyd-Bowman (1964a, 1968) compiled the most widely used individual-level dataset, drawing on the *Catálogo de Pasajeros a Indias* (Archivo General de Indias 1940), but with significant gaps in the earliest years and no coverage of notarial contracts. Subsequent scholarship largely returned to the same archival bases (Martínez Shaw 1994). Notable exceptions include Márquez Macías (1995) and Macías Domínguez (1999), who used new sources but focused on the 18^{th} century, and a set of regional studies surveyed in García Hidalgo (2019). The qualitative literature on networks and chain migration during this period—particularly the role of expedition leaders in recruiting from their home towns, and the role of letters and personal ties in sustaining transatlantic links—is well developed (Almorza Hidalgo 2018; Altman 1989, 1991, 2000, 2021; Otte 1993; Stangl 2010). We build on this work by providing the first systematic quantitative test of these mechanisms across the full population of documented migrants. Our paper also connects to the literature on migrant networks in the Age of Mass Migration (1850–1914). Spitzer and Zimran (2023), studying Italian emigration between 1876 and 1920, show that migration spread gradually outward from a small number of epicenters in a spatial diffusion process, with emigration rates eventually converging across regions as networks reached previously unconnected communities. Spitzer (2021) documents a similar process for Jewish emigration from the Russian Empire's Pale of Settlement. Munshi (2003), working with 20^{th}-century data on Mexican migration to the United States, shows that a larger network at the destination raised both the likelihood of finding employment and wages. Munshi and Rosenzweig (2016), studying rural-to-urban migration in India, show that community networks shape not only the scale of migration but also its direction, as network ties bind individuals to specific destinations even when alternative labor markets might appear to offer higher returns.[1] Our results show that both mechanisms—larger flows from connected origins and persistent origin-destination specialization—were already operating in the very first years of European overseas expansion. Wegge (1998), using data on over 50,000 emigrants from the German principality of Hesse-Cassel in the mid-nineteenth century, shows that chain migration had a strong effect on subsequent migration from the same villages. Our results show that the same mechanisms—self-reinforcing chains, geographic persistence, network effects on destination choice—were already operating in the very first years

[1] That is, individuals might value the presence of networks, either for their own sake or because of their role in facilitating job placement, beyond simple wage differences across space. Moreover, an existing network at the destination might reflect unobserved origin-destination complementarities—an issue which generally poses a challenge to the identification of network effects, but which our setting of limited information about destinations helps to address.

of European overseas expansion, three centuries before the period that dominates the existing literature. Crucially, however, we find no spatial diffusion: the institutional bottleneck of the Seville monopoly prevented networks from spreading to new regions, in sharp contrast to the Italian and Russian cases. Our findings also speak to the literature on the long-run consequences of colonial settlement. Acemoglu et al. (2001) and subsequent work emphasize the role of colonial institutions in shaping long-run development. Our results suggest that the composition of founding settler populations—itself shaped by pre-existing networks rather than by systematic selection on individual characteristics—is an earlier link in this chain that has received less attention. Finally, we contribute to the growing literature connecting historical population genetics to economic and social history. Bycroft et al. (2019) show that the genetic structure of modern Spain closely reflects the boundaries of medieval kingdoms, tracing the population movements of the *Reconquista*. We show that this same geography predicts the origins of the first transatlantic migrants, connecting centuries of Iberian social history to the founding of colonial societies in the Americas.

2 Background

Columbus' expedition to the New World is well-known. With a few exceptions—perhaps Balboa, Cortés, and Pizarro—rather less famous are the hundreds of men who fanned out to explore (and sometimes conquer) the rest of the Caribbean and the American mainland. Moreover, we know even less about the fates of the thousands of voluntary migrants—men and women—who followed the first expeditions. The Crown had finished the *Reconquista* in 1492 when the Atlantic expansion began, but it did not finance this expansion as a fully funded state enterprise. The Spanish empire was explored, conquered, and settled by neither taxpayer-funded professional armies nor organized joint-stock companies. Rather, the first migrations consisted of private enterprises regulated by the Crown. It authorized voyages through *capitulaciones* that drew on legal experience from the *Reconquista* and earlier Atlantic expansion (e.g., Canary Islands). Castile's recent frontier mobilization, municipal fiscal coordination, and Atlantic access through Seville made this imperial model feasible.

2.1 Spain on the eve of overseas expansion

In 1492 the Iberian Peninsula was a composite monarchy rather than a unified state. The Crowns of Castile and Aragon remained legally distinct, yet under Ferdinand and Isabella they operated as a dynastic union that contemporaries increasingly referred to as "the Spains." Castile was larger, more populous, and fiscally

more centralized than the Crown of Aragon, whose territories remained oriented toward the Mediterranean.[2] By the late 15^{th} century Castile had emerged as the dominant partner within the union (Hillgarth 1978).

Castile's ability to sustain overseas expansion rested on institutional developments that predated Columbus. Throughout the later Middle Ages, the Crown had negotiated extraordinary fiscal services with the Cortes, creating a recurrent channel for mobilizing revenue. Under the Catholic Monarchs, the *Santa Hermandad* strengthened this arrangement by embedding royal taxation and enforcement within municipal structures. The *Hermandad* was not merely a policing instrument; it functioned as a fiscal-administrative mechanism that linked urban centers to royal projects (Triano Milán 2018a,b).[3]

This administrative consolidation was inseparable from the long Reconquista. For centuries, Christian kingdoms had expanded southward, repopulating conquered territories with settlers drawn from northern regions. The final campaign against Granada (1482–1492) mobilized nobles, municipal militias, and retainers across all of Castile (Hillgarth 1978). The *Hermandad* only recruited from a wide north–south corridor of Leon and Extremadura (Quesada 2001). These mobilization networks did not disappear with Granada's fall. They connected towns across western Castile, León, and Extremadura that had shared military service, settlement, and administrative integration for generations.

The demographic and cultural traces of this settlement process are visible in both linguistic and genetic evidence. Medieval political frontiers correspond closely to later dialect domains, including the Astur-Leonese corridor extending southward into Extremadura (Menéndez Pidal 1906; Penny 2002). Figure 14, shows modern population-genetic analysis likewise finds pronounced east–west differentiation within Spain and comparatively greater continuity along the western north–south axis (Bycroft et al. 2019). Galicia and the territories of the Crown of Aragon appear more segmented, while the Castilian–Leonese corridor shows evidence of sustained mixing over centuries.

Geography reinforced these institutional and social patterns. Spain is one of the most mountainous countries in Europe. Overland travel followed medieval roads and Mesta drove routes; navigable rivers were limited. Seville, though nearly ninety kilometers inland on the Guadalquivir, occupied a strategic position linking Atlantic access to the administrative heart of Castile. It lay at the southern terminus of the *Reconquista* settlement corridor and within the zone of greatest demographic connectivity. Proximity to Seville therefore reflected not merely distance but integration into Castile's fiscal and social core.

On the eve of overseas expansion, three structural features were thus in place: a centralized yet municipality-

[2]Granada was conquered in 1492; Navarra was subjugated in 1515; and Portugal would not join the union until a century later.

[3]The *Hermandad* was active during 1476-1498, and during this period the Castilian Parliament (*Cortes*) was not convened a single time.

embedded fiscal apparatus, dense north–south social connectivity within western Castile and Extremadura, and pronounced regional segmentation elsewhere on the peninsula. These conditions preceded Columbus's voyages. They did not determine that transatlantic migration would occur, but they structured which communities were institutionally and socially positioned to participate once overseas expansion began.

2.2 From frontier to empire: The making of migration

The institutional architecture that structured migration to the Americas did not emerge *de novo* after 1492. It drew directly on late medieval Iberian precedents. Overseas expansion was framed legally through *capitulaciones*, contractual grants by which the Crown authorized privately financed expeditions in exchange for a share of returns and recognition of royal sovereignty (Fernández and Sotelo 1987).[4] This contractual form had roots in frontier governance during the *Reconquista* and in earlier Atlantic experiments in the Canary Islands, where conquest, settlement, and revenue extraction were organized through delegated authority under royal oversight (Verlinden 1970).

Labor and tribute institutions were similarly transferred. Early colonial governance in the Caribbean adapted Iberian precedents for allocating rights over labor and tribute collection. Under Governor Nicolás de Ovando, indigenous populations were formally recognized as free vassals of the Crown but were subjected to compulsory labor and tribute obligations. These arrangements evolved into the *encomienda*, which granted Spaniards rights to indigenous tribute and labor while preserving ultimate royal sovereignty. The *encomienda* translated frontier practices of delegated authority and tribute extraction into an overseas setting (Arroyo Abad and Espín-Sánchez 2024; Zavala 1935).

If the legal and labor framework reflected *Reconquista* precedents, the administrative funnel reflected late 15^{th}-century Castilian centralization. Migration was tightly regulated through the *Casa de Contratación* in Seville. Prospective migrants were required to obtain licenses (licencias), document lineage—three-generations Christian—and good behavior—never convicted of a crime, no outstanding debts, and signed letters of good character from the local priest and mayor—and formalize passage contracts before notaries (Konetzke 1945). The choice of Seville was not accidental. Seville was already integrated into Castile's fiscal-administrative apparatus and lay within the north–south corridor of demographic connectivity created by the *Reconquista*. The *Casa de Contratación* did more than issue licenses. It coordinated navigation,

[4]The Santa Fe *capitulación* of 1492 signed between Columbus and the Crown kick-started Spain's transatlantic expansion a few months after the Crown had reconquered the last Muslim kingdom in the peninsula, effectively ending the *Reconquista*. Known as the "Magna Carta of discovery" (Ballesteros Beretta 1945), this three-page document became the cornerstone of the political and judicial institutions under which Madrid conquered and settled its overseas empire.

maintained records, certified pilots, and monitored trade flows.[5] In institutional terms, it functioned as a node that linked private enterprise to royal authority. The requirement that migrants physically present themselves in Seville before departure created a spatial bottleneck: access to overseas migration depended on integration into the Sevillian administrative orbit.[6] The contractual dimension of migration reinforced this centralization. Passage agreements were negotiated with ship masters and notarized in Seville's notarial offices (see Figure 1). Notaries were not passive recorders. They were intermediaries in credit, contracting, and information exchange (Oropeza Chávez 2016). The concentration of notarial activity in Seville embedded migration within a dense urban legal infrastructure, further tying access to overseas mobility to specific administrative channels.

The costs and risks of the voyage compounded these institutional filters. Transatlantic passage required provisioning for journeys that lasted months, including staged departures through Sanlúcar and the Canary Islands.[7] Travel expenses were significant relative to annual wages. For example, one-way fares to the Americas in our period had a median price of 12 ducats and a Mean of 23.3. 12 ducats came to about of six months unskilled wages (Hamilton 1934). This figure is consistent with Don Quixote's sidekick, Sancho Panza's, livelihood as a servant: "When I worked for Tom Carrasco, the father of the bachelor Samson Carrasco that your worship knows, I used to earn two ducats a month besides my food." (Cervantes Saavedra 1615, chapter XXVIII). If that were the only cost, then crossing the Atlantic in 1500 would have been comparable to making the same trip in 1900, when fares from Italy to the United States were equivalent to about 4 months' unskilled wages in southern Italy (Gomellini and Ó Gráda 2013, p. 276).[9]

[5]The *Casa de Contratación* played the role of NASA by bringing together the best minds of the time—cartographers such as Alonso de Santa Cruz, who produced the master map of the Spanish empire and pioneered the use of isogonic lines to correct compass readings, mathematician-astronomers such as Jerónimo de Chaves, who translated the standard astronomical treatise of the day into Spanish and taught navigation at the Casa, and master pilots such as Amerigo Vespucci, who served as its first chief pilot and gave his name to the new continent—to teach captains and pilots how to reach and explore the new continent and report back on their findings.

[6]In 16^{th} century Spain, making this journey to Seville was no small feat, requiring up to 80 days from the northern coast. Adding to the cost was the need to stay in Seville for the license processing.

[7]The physical commitment required of early migrants is difficult to overstate. A voyage from Seville to the Caribbean took three to four months, and the full journey to Peru could approach a year (Pérez-Mallaína 1998)—comparable in duration to a contemporary spacecraft mission to Mars, which requires six to nine months using standard trajectories (Kingdon 2025). At sea, the first stop was a week to ten days later in the Canary Islands to secure the two most important provisions for the trip: firewood and water. A month later, the ships would make a second stop in the Lesser Antilles to procure more water and firewood. The conditions on the ship were described akin to a narrow unstable prison that reeked like hell and was infested with lice, rats, and cockroaches. For those heading to New Spain, the trip would make another pit stop in Puerto Rico to arrive a month later to San Juan de Ulúa in Veracruz (Mexico). For Tierra Firme, the trip to the port of Cartagena (Colombia) would take two weeks. To make it all the way to Peru, the ships would head to Nombre de Dios (Panama). The insalubrious conditions of this port made it unlivable for two-thirds of the year, limiting the number of available trips to Peru. Peru-bound migrants would also have to cross the Isthmus of Panamá on foot. We know that travelers taking the Camino Real took four days to transit, but that road was not completed until 1540, meaning that the trip for the migrants in our sample took even longer.[8] To make it to the port of El Callao (near Lima), the traveler would add at least three weeks to the journey (Haring 1918; Martínez 1983; Pérez-Mallaína Bueno 1998; Schafer 1946).

[9]The difference was that a steerage fare in 1890 included food and a place to sleep—whereas a contract in 1500 included none of those things. It only included the provision of wood, salt, and water. Water was scarce and specified in the contract, varying from 1 to 2 liters per person per day. Passengers had to pack and pay for their luggage, including bedding and enough

Taken together, these features produced a distinctive configuration. The legal framework of overseas expansion was adapted from frontier contracts and delegated authority; labor and tribute institutions were transferred from Iberian precedents into colonial form; and migration was routed through a single administrative center embedded in Castile's existing fiscal and social networks. Rather than creating new channels of connectivity across the peninsula, the early transatlantic system activated and intensified pre-existing ones. Communities already integrated into Castile's administrative and social core could more readily access the Sevillian funnel. Regions more weakly connected—whether because of linguistic, political, or demographic segmentation—faced higher effective barriers to participation. In this sense, early Spanish migration to the Americas combined institutional transfer with institutional centralization. Frontier governance provided the template; Seville provided the gate.

2.3 The moving American frontier

If Spain's institutional and social infrastructure shaped who could migrate, the American frontier shaped where they went. Between 1492 and 1540, opportunity moved, and migration moved with it—though not in a straight line. Migration spread outward from the islands to the mainland as new conquests opened, while the older Caribbean corridor remained active. The first two decades after Columbus were dominated by the Caribbean. Early expeditions focused on Hispaniola, Cuba, and Puerto Rico. Settlement in this phase was fragile, often violent, and institutionally experimental. Administrative practices such as delegated authority through *capitulaciones* and the allocation of indigenous labor evolved in this environment (Altman 2021). Yet despite institutional instability, the Caribbean functioned as the initial anchor of migration.

Beginning in the 1510s, the frontier shifted toward Tierra Firme—roughly the mainland coasts of present-day Panama, Colombia, and Venezuela. Permanent footholds such as Nombre de Dios and Darién expanded the range of destinations (MacLeod 2010). This expansion was not gradual diffusion but punctuated movement. New destinations opened following exploratory expeditions, and migration surged in response. The conquest of Mexico in the early 1520s marked a decisive turning point. Hernán Cortés's campaign transformed the expected returns to migration. New Spain offered dense populations and the prospect of precious metals (Arroyo-Abad and Palma 2021; Hassig 2006). In the 1530s, the conquest of Peru by Francisco Pizarro generated a similar shift. The Andean region combined dense population, established political hierarchies, and access to precious metals (Lockhart 1968, 2013). Many of the initial settlements failed, most notably in

provisions and utensils for a trip that would last at least ten weeks (Slicher van Bath 1986)—compared to less than a week for a transatlantic migrant on a steamship. The full cost of passage then was the result of the fare plus provisions and luggage (Durán López 1991; Rodríguez Lorenzo 2017).

Rio de la Plata (modern-day Buenos Aires) and Florida (modern-day Florida).[10] Migration responded not smoothly but in waves tied to exploration, conquest, and settlement (Figure 4.a). These episodes illustrate the uncertainty migrants faced. The attractiveness of a destination could change abruptly depending on military outcomes, indigenous resistance, or administrative reorganization.

This moving frontier did not produce anonymous flows. It produced chains. Conquistadors and expedition leaders did not merely open destinations; they recruited from home, using banners, public announcements after mass, and town criers, while news of prior expeditions—especially stories of gold—circulated locally and made overseas ventures legible to people who had never seen the Indies (Altman 2021; Hennessy 1993). Information then kept the links alive. Migrants relied on letters and regular correspondence with settlers overseas, which conveyed not only opportunities but also bureaucratic requirements and practical guidance for the journey (Otte 1993; Stangl 2010). Over time, these channels turned pioneer movement into repeatable routes: migrants increasingly followed relatives and fellow townspeople, financed passages through loans or remittances, and drew on kin and paisanaje networks that structured both departure and settlement (Almorza Hidalgo 2018; Altman 1991, 2021; Núñez 2002; Vázquez de Prada Vallejo 1991).

This shifting frontier had two implications for migration patterns. First, the composition of destinations changed rapidly over time. The Caribbean dominated early; Mexico and Peru later absorbed growing shares of migrants. Second, information about these destinations traveled primarily through expedition participants, returning settlers, and correspondence. Whereas the epic stories of conquest and gold circulated rapidly all over Europe, the details on how to get there and how to take advantage of the opportunities that the New World offered did not. Under conditions of slow communication and high uncertainty, knowledge about new opportunities was localized. Early participants in a given expedition became conduits of information and recruitment for subsequent waves. Thus, while the American frontier was dynamic, the geography of Spanish origins proved far more stable. Migration responded elastically to new continental opportunities, yet participation within Spain remained concentrated in a relatively narrow set of regions. The contrast between fluid destinations and persistent origins lies at the center of the empirical analysis that follows.

[10] Of the 126 expedition leaders identified in our data, over half of them died prematurely—and a bit over a third of those died at the hands of their compatriots (see Richman (1921), Montana (1943), Esparza (2015), González Ochoa (2003), Kirkpatrick (1934), and a variety of biographical dictionaries such as the Real Academia de Historia's *Diccionario Biográfico electrónico*). Of our sample, 6% of expedition leaders died of unknown causes, 7% perished from disease, 12% suffered fatal accidents, 14% lost their lives at the hands of indigenous peoples, and 21% were killed by other Spanish explorers. The rest survived their expeditions and settled in América. Less than 1% returned to Spain.

3 Data and sources

Our novel database is the result of detailed data collection from an array of archival and secondary sources. Our database combines individual data on migrants, Spanish municipality-level population and economic data, and granular geographic characteristics.[11]

3.1 Migrant data

The key component of our database is individual-level migrant information. Gathering this information is possible because the Spanish imperial government regulated migration flows. These migration records survived in two main archives: the General Archives of the Indies, Spain (*Archivo General de Indias*, AGI) and at the Archive of Notary Records of Seville (*Archivo de Protocolos Notariales de Seville*, APNS). Our starting point was the digitization of Boyd-Bowman's work Boyd-Bowman (1964b, 1968) (henceforth, BB), based on the *Catálogo de Pasajeros a Indias* (Archivo General de Indias 1940) and added Spaniards living in the Americas from other secondary sources.[12] The *Catálogo* itself is a transcription of the applications to migrate to the Americas (*Licencias*)—akin to modern-day visas—and ship's passenger lists (*Asientos*), both housed in AGI. Thanks to the Duke of Alba's funding, the *Catálogo* preserved a vital body of evidence and directly inspired the pioneering work of Peter Boyd-Bowman.[13] These initial efforts revealed significant gaps. To fill these gaps, we hand-collected and transcribed original archival records from the notarized contracts of carriage between passengers and captains (*maestres*) housed at the APNS—see Figure 1.[14] The APNS holds over 20 million pages of documents, just for the 16^{th} century. We follow the collection of 16 volumes of (Instituto Hispano-Cubano de Historia de América 1930–2024), which contain summaries of the contracts related to the Americas. Within that collection, we look for any contract that refers to passengers crossing to the Americas, and extract the information on the date, the names, occupation, origin, and destination of the passengers. We also collect information on the name of the captain (*maestre*) and the notary (*escribano*). Finally, we also fully transcribed all the (*Asientos*) available until 1540, and added other information—missing in BB—such as the name of the ship's captain and the name of the migrants parents.[15] Finally, the

[11] For more details, see Appendix A.

[12] Boyd-Bowman's (1964a, 1968) magnum opus, *Indice Geobiográfico de 56.000 Mil Pobladores Españoles de América en el Siglo XVI* (originally refers to the number of migrants until 1600. These volumes offer aggregate statistics for the entire century and individual information until 1540. This compilation is based on information from Archivo General de Indias (1940), which itself uses two primary sources (*Asientos* and Informacion y Licencias) stored at the *Archivo General de Indias* (AGI).

[13] A fire in 1924, damaged the AGI's archives destroying an unknown number of travel records. This prompted the Duke to fund the remodelling of the archive and the creation of the *Catálogo*.

[14] For details on linking Spanish names see Appendix B.1 and Espín-Sánchez et al. (2022).

[15] 1542 marks a major institutional break (New Laws), so our window captures the pre–New Laws regime. These laws, promulgated by Charles I in November 1542, banned the enslavement of indigenous peoples, prohibited new *encomienda* grants, ended their heritability, and imposed stricter regulation on expeditions of conquest (Elliott 1963).

earliest voyages to the Americas—including Columbus's four expeditions (1492, 1493, 1498, and 1502) and Nicolás de Ovando's colonizing expedition (1502)—remain absent in BB. We added information on these migrants from secondary sources (Gil 1985; Gould 1984; León Guerrero et al. 2007; Mira Caballos 2000, 2014).

To match each migrant to a destination, we consulted historical atlases, dictionaries, traveler's accounts, and anthropological and historical secondary sources. We divide the destinations of migrants into five main destination groups—the Caribbean, Mexico, Tierra Firme (roughly modern Colombia, Venezuela, and Central America), Peru (Ecuador, Peru, Bolivia, and Chile), and La Plata (roughly modern Argentina, Paraguay, and Uruguay). While most of the destinations have retained their pre-colonial or colonial names, others have changed over time. This classification catches, roughly the region where the migration settled. The end result of these efforts was a comprehensive dataset covering more than 21,000 individuals who migrated from Spain to the Americas, before 1540. For each person, we systematically recorded a range of demographic and social attributes: the year of departure, place of origin, destination, occupational category, and social status or rank (including classifications such as noble, clergy, or servant).[16] In addition to individual travelers, we reconstructed the composition of collective expeditions, distinguishing between leaders, identified conquistadors, and ship crews. This dual structure—individual migrants and organized expeditions—allows us to analyze both the micro-level characteristics of migration and the broader organization of colonial ventures.[17]

3.2 Spanish population data

We created a dataset of Spanish population by municipality in the early 16^{th} century (see Figure 2). The majority of migrants listed his or her place of origin. We used a variety of censuses for the different Spanish kingdoms. These censuses were primarily fiscal in nature, designed to assess and collect revenue. Our main population source is the *Censo de Pecheros of 1528* (Instituto Nacional de Estadística 2008). This extraordinary source provides information on the number of households for each population center for the Crown of Castile. The *Censo* classifies into four categories: *pobres* (poor), *pecheros* (taxpayers), *nobleza* (nobility), and *clero* (clergy). Only the *pecheros* were liable for taxation, but the enumeration covered all four groups. For other regions, the sources available generally report number of households, often referred as *fuegos* ("hearths" in Spanish). See Appendix A.4 for details.

[16]For details on assigning occupational categories and social class to pre-modern Spanish sources see Appendix A and Espín-Sánchez et al. (2019).
[17]See Appendix A.1.

3.3 Geographic characteristics

Spain is a topographically diverse country, one of the most mountainous countries in Europe. To analyze migration, we must take these characteristics into account. Because Seville served as the epicenter of transatlantic migration, we must evaluate how easily people from different regions of the peninsula could reach the city in the 16^{th} century. This exercise implies taking into account ruggedness, navigable rivers, ports, and roads. It is based on GIS shapefiles of Spanish municipalities from Huwise (2023).

To assess ruggedness and elevation, we used GIS shapefiles from the Shuttle Radar Topographic Mission (Jarvis et al. 2008). Based on the work by Esteban-Oliver et al. (2024), we selected operational ports and navigable rivers for the period to account for water travel. The main forms of overland transportation in the period in question were the *Mesta* drove roads[18] and other main medieval roads based on Carreras Monfort and de Soto Cañamares (2010)—see Figure A.2.

With all this base information, we compute, for each municipality, a distance (in days) from the emigration port of Seville. Using algorithms for the computation of travel cost described by Donaldson and Hornbeck (2016) and Zimran (2020), we compute travel cost between each place of origin and Seville.[19] Figure 3 presents the results of this calculation.[20] We refer to this as a place's "travel cost" from Seville, allowing for water travel (panel (a)) and overland travel only (panel (b)).[21]

In most cases, our analysis is at the municipality level. In some cases, however, we find it helpful to aggregate the analysis to higher geographic levels. An obvious level of aggregation is the province, which we do use. As an intermediate level, we also aggregate municipalities into modern judicial districts (*partidos judiciales*). Though these districts did not exist historically, they provide a way to group municipalities that respects local connections and enables analysis of geographic units smaller than the province but larger than the municipality.

[18] The *Mesta* roads are the *cañadas reales*.

[19] We treat off-road travel as 4.2-times as costly as travel on roads. The rationale for this is that $4.2 = 3 \times 1.4$, where 1.4 is the correction often used to adjust straight-line distances for the fact that it is typically not possible to walk in a straight line (Donaldson and Hornbeck 2016; Zimran 2020).

[20] See Appendix A.5 for a detailed explanation on these estimations.

[21] For districts, we use the travel distance from the district's capital.

4 Empirical results

4.1 General trends

From 1492 to 1540, our lower bound estimates of the flow of voluntary migrants is 21,021. A first look at the migration flows (Figure 4.(a)) shows that the flow of migrants from Spain to the Americas migration was quite small in magnitude until about 1510, after which there was a brief but short-lived surge, which repeated again in the mid 1520s. Only in the 1530s was there finally a sustained surge in migration.Each surge followed the establishment of new colonial footholds in the Americas. 95% of migrants come from Castile. Over the entire period, the average annual rate of migration from a municipality in Castile was 4.35 per 100,000 (Table 1).[22] This rate was more than three times greater than in Granada, six times greater than in Navarra, and ten times greater than in Aragon. This difference emerged early and persisted throughout the period. The fivefold increase in migration rates among Castilian municipalities between 1492–1512 and 1513–1540 (from 1.27 to 6.51) mirrored the overall Spanish increase across the same periods.[23] 15% of migrants came from Seville, making it a unique player in the migration. This city exhibited a very high migration rate relative to its population (Figure B.3.(b)). This is unsurprising given the monopoly of the port of Seville in the trade with the Americas, but it also implies that Seville must receive special treatment in the analysis.[24]

The most popular destination was the Caribbean islands (23.2%)—mainly Cuba, *La Hispaniola*, and Puerto Rico—followed by Mexico (22.5%). Tierra Firme (Central America, Colombia, and Venezuela) attracted 17.2% of the migrants. Both Peru (Peru, Bolivia, Ecuador, and Chile) and La Plata (Argentina and Paraguay) received nearly 5.2% and 4.4% respectively, in part due to their late discoveries. The remaining went to other destinations such as the Philippines (see Figure 5). As the explorers opened up new destinations over time, migrants followed. Figure B.1 gives a sense of the importance of each major destination group over time, plotting the number of migrants to each destination from all sources over time. It shows that in the early years of our study period, the Caribbean was the main destination of migrants. Tierra Firme's initial exploration dates back to 1501 led by Rodrigo de Bastidas. It took almost a decade to establish a settlement in Nombre de Dios, Panamá), in 1510. Mexico became an important destination in the 1520s after the expedition of Hernán Cortés. Francisco Pizarro's expedition to Peru in 1531 motivated migrants to try their luck there. After very deadly expeditions to Buenos Aires (La Plata), Pedro de Mendoza headed

[22] Henceforth, all emigration rates are annual per 100,000 unless stated otherwise.

[23] We adopt 1512 as a breakpoint with the issuance of the Laws of Burgos that marked a shift toward a more formalized colonial regulatory framework.

[24] For details on the composition of migrants see Appendix A.2.

an expedition in 1535 that couldn't survive indigenous resistance.[25]

Figure 4 presents an overview of the time trends of the emigration data. The average annual number of migrants in the data is 437, with an overall rising trend throughout the period, and substantial year-to-year volatility. Panel (b) shows the number of migrants by source. Although BB is still the main source, our augmented dataset includes many new migrants, specially during the early years, where networks were forming. Throughout the period, the most important regions in terms of passenger origins were Andalusia, followed by Castile and Extremadura (panel c). Despite its location in the far north, far from Seville, the Basque Country was also an important source of passengers (though to a lesser extent than other regions), due to its shipbuilding industry, i.e., all Spanish Atlantic ships in this period where built there. Finally, the destinations of passengers, generally followed the trends of conquest—first the Caribbean, followed by Tierra Firme, and finally Mexico, and Peru, which forged ahead during their conquests (panel d). La Plata, as explained above, was a late and marginal destination.

4.2 Geographic origins of migrants

The origins of expedition crews and explorers mirror this uneven geography (Figure 6, panels (a) and (b)). Crew origins cluster around Seville and the lower Guadalquivir. Seville contribute the largest number of crew members followed by the neighboring provinces in western Andalusia and Extremadura (e.g., Badajoz and Huelva). Explorer origins concentrate even more sharply in inland Extremadura: Badajoz falls in the top category and adjacent Cáceres also ranks highly, with smaller counts in nearby western Andalusian provinces. Outside this southwest corridor, only a handful of provinces in north-central Castile register moderate counts, while most provinces in eastern and Mediterranean Spain contribute few crew members or explorers.

Castile contributes heavily to this migration wave, but departures cluster in a few areas within Castile. Figure 7 maps the geographic origins of migration in detail. Three clusters stand out. First, municipalities in the province of Seville—the primary gateway to transatlantic travel—form a major hub, with an average municipal emigration rate of 12.82.[26] Second, emigration rates are even higher in Extremadura, particularly in the province of Badajoz (14.33).[27] Several towns had astonishing emigration rates exceeding 50. This rural, rugged inland region contributed many participants to early conquest and colonization, including leading figures such as Ovando, Cortés, and Pizarro. Third, a smaller but distinct cluster appears in the Basque

[25]Two years later, they abandoned the settlement. Buenos Aires would not be re-founded until 1580.

[26]Provinces are according to modern boundaries.

[27]Five of the 13 Spanish towns with a population above 2,500 and emigration rates over 50 were in Extremadura: Albalá, Cafra, Cáceres, Mérida, and Trujillo. For comparison, only 16 out of all 299 Spanish towns above 2,500 were Extremaduran.

Country, with an average emigration rate of 13.73, consistent with the region's early role in shipbuilding and maritime labor (Odriozola Oyarbide 1998; Ortiz Arza 2019). Alongside these regional clusters, the map also shows that many urban centers served as local emigration hubs, a pattern we examine further below.

The spatial distribution is also defined by areas of low or no emigration. This is true for much of the eastern half of the peninsula (including the Kingdom of Aragon and eastern Castile) and for the northwestern region of Galicia. Distance from Seville does not fully account for this pattern, as we show below. Particularly striking is the absence of major clusters around large Mediterranean port cities such as Barcelona and Valencia.[28] These ports were active in imperial commerce but contributed relatively little to migrant cohorts in this period.

We next relate these patterns to observable municipal characteristics. Figure 8(a) bins municipalities by travel distance from Seville accounting for land and water routes.[29] Emigration rates are substantially higher for the closest 5 percent of municipalities—and, to a lesser extent, the next closest 5 percent. Beyond this range there is little additional decline with distance. There is also a strong negative relationship between elevation and emigration: average rates in the lowest altitude decile are four times those in the highest decile (4.14 vs. 1.02) (Figure 8(b)). Municipal size also predicts emigration. Emigration rates stay nearly flat among smaller localities, at roughly 2 per 100,000, but they rise sharply once municipalities pass 1,000 inhabitants (Figure 9, panel (a)). Municipalities with more than 5,000 inhabitants average 13.9 per 100,000, compared with 2.00 in municipalities under 1,000. Among the largest cities (with population above 10,000), city size does not explain much additional variation. Excluding Seville for clarity, larger cities do not systematically send more migrants, and the relationship is moderately negative and statistically insignificant (Figure 9, panel (b)). Instead, kingdom-level differences dominate. With the notable exception of Granada, large cities outside Castile—including Barcelona and Valencia—typically exhibit emigration rates between 1 and 5, whereas Castile's large cities average 30.46.

4.3 Local push factors

The descriptive patterns above suggest that geography and local economic conditions shaped who left Spain. We now quantify these relationships by estimating municipality-level regressions that relate emigration to a set of local characteristics. We measure emigration as a rate and, in our baseline specifications, we model it in logs so determinants shift emigration proportionally rather than by a constant amount, consistent with

[28] Emigration rates from the provinces of Barcelona and Valencia were only 0.42 and 0.29.
[29] Seville is included in the nearest bin.

standard models in which migration arises from individual binary choices.[30] Many municipalities record zero migrants, so we estimate log-rate specifications only after aggregating over sufficiently large units and time periods.[31] For tractability and comparability (and for a somewhat simpler interpretation), we therefore also report linear specifications, while noting that they impose constant marginal effects regardless of baseline emigration levels. Our preferred unit of analysis is the municipality; when aggregation becomes necessary, we aggregate to the modern-day district level and verify that the results remain consistent. Unless stated otherwise, we cluster standard errors by province.

Our baseline specification is a cross-sectional binomial maximum-likelihood regression of the form

$$\log\left(\frac{m_i}{1-m_i}\right) = x_i'\beta, \tag{4.3.1}$$

where m_i is the migration rate for municipality i over the entire period and x_i is a vector of geographic and other local characteristics of municipality i. Table 2 confirms that access to the Seville and urban scale explain a large share of cross-municipality variation in emigration. Travel costs to Seville correlate negatively with emigration in specifications without fixed effects and with province fixed effects: a one-standard-deviation increase in travel cost predicts roughly 33 percent lower emigration (columns (1)–(2)). Once we compare municipalities within the same district, the travel-cost coefficient shrinks and loses precision (column (3)), consistent with the idea that travel costs mainly differentiate Seville's broader hinterland from the rest of Spain rather than explaining within-district differences. By contrast, log population strongly and robustly predicts emigration in every specification (elasticity 0.46–0.67). Other geographic variables—elevation, soil quality, and proximity to rivers or the coast—show no systematic association with emigration. Finally, even after conditioning on local characteristics, large geographic gaps remain in column (1): Aragon's emigration rate is only about 9.5 percent that of the Castilian reference category, Navarra's is about 55 percent that of Castile, while Seville and especially Badajoz exhibit substantially higher emigration.[32] Comparing the results

[30] cf. Spitzer 2021; Spitzer et al. 2025; Spitzer and Zimran 2023.

[31] To retain the full municipality-level sample, we also estimate binomial maximum likelihood (BML) models with the emigration count as the dependent variable and municipal population as the exposure term. Appendix B.2 shows that BML corresponds to an individual-level logit model with municipality–level regressors (Spitzer and Zimran 2023). When emigration rates are very small—as in our setting—BML closely tracks Poisson regression at the municipality level, and both yield nearly identical semi-elasticity interpretations. When migration rates are very low, the coefficients in these regressions, which are equivalent to logit regression coefficients, can be interpreted as semi-elasticities for the migration probability.

Nonlinear binomial and Poisson estimators become computationally demanding with many fixed effects and, in panel settings, can suffer from incidental-parameter bias. The incidental-parameter problem implies finite-T bias of unknown sign in nonlinear fixed-effects models. In linear panel specifications with lagged outcomes, the analogous bias takes the form of Nickell bias, of order $\frac{1}{T}$ and with a known direction (Nickell 1981).

[32] District-level regressions (Table B.1) show similar results and reinforce the importance of travel costs, at least across small regions. In particular, even after controlling for province fixed effects, average district-level travel costs remain strongly predictive of emigration, with a much larger coefficient than in the corresponding municipality-level specification.

for early and later migrants (before and after 1513, columns (4) and (5)) shows that the over-representation of Seville and Badajoz appeared only in the latter period. Beyond that, there were no material differences between the two periods.[33]

Finally, to assess whether poorer municipalities sent more migrants, we augment the baseline specifications with a proxy for local income: the log of assigned taxes per capita (Table B.4). We restrict this exercise to Castile, the only kingdom for which the tax data are available. Because some Castilian municipalities paid no assigned taxes, we add an indicator for zero taxes per capita and interact it with the log tax measure.[34] Across specifications, the coefficient on log assigned taxes is negative, but the implied elasticity remains small and never reaches statistical significance.[35]

4.4 Dynamics

The geography of emigration changes little over time. The decade-by-decade origin maps show the same hotspots throughout the period, with only modest fluctuations (Figure B.2). Simple tests for convergence in migration rates across provinces over time shows this persistence. Figure 10.(a) tests for σ-convergence, plotting the coefficient of variation of emigration rates across provinces for each year in our study period. The figure shows an initial decline in cross-province dispersion, but from roughly 1510 onward the coefficient of variation stabilizes around 1.5–2, indicating little further σ-convergence. That is, there was no reduction over time in the variation in migration rates over provinces as would occur if provinces that were not participants in the early migration began to participate later.

Figure 10.(b) tests for β-convergence—whether laggard provinces experienced a greater increase in emigration rates, thus catching up to the emigration rate of leading provinces. The x-axis plots a province's average annual emigration rate for the per-1513 period, while the y-axis plots the ratio of a province's post-1513 migration rate to its pre-1513 migration rate.[36] The figure provides only weak evidence of β-convergence: the analogous univariate regression has an estimated coefficient of -0.155. This does indicate some degree of catching up by laggard provinces, but for two reasons does not reveal any *meaningful* evidence of catch-up. The degree of convergence exhibited here is substantially smaller than in other contexts of convergence,[37] Taken together, these results indicate that emigration quickly became spatially concentrated

[33]See Tables B.2 and B.3) in Appendix B.3.1 for more details.

[34]Many municipalities with zero assigned taxes received exemptions for institutional reasons (e.g., prior ties to the Crown) rather than because they were especially poor, so we treat zero taxes as a distinct category.

[35]Restricting the sample to municipalities with non-missing assigned taxes and adding the tax variables does not materially change the coefficients on the other covariates.

[36]To be precise, if m_{p0} is province p's pre-1513 migration rate, and m_{p1} is province p's post-1513 migration rate, Figure 10.(b) plots $\frac{m_{p1}}{m_{p0}}$ on the y-axis and m_{p0} on the x-axis.

[37]For instance, in Italy in the Age of Mass Migration, a similar regression yields a coefficient of -0.5 (Spitzer and Zimran

and remained so, with little evidence that late-entry regions gradually joined the movement.

In summary, the map of migration in Spain was set very early, and did not move afterwards in our period, and probably for centuries. The few towns and cities that happen to have an early pioneer among their network were the only ones that participated in the migration to the Americas. In Subsection 5.2 we discuss the implications of this lack of spreading of this network.

Table 3 quantifies this persistence by regressing district-level emigration after 1513 on emigration in the formative years before 1513.[38] We estimate a binomial maximum likelihood model in which a municipality's migration rate after 1513, m_{i1} is expressed in the logit form

$$\log\left(\frac{m_{i1}}{1-m_{i1}}\right) = \alpha \mathbb{1}\{m_{i0}=0\} + \beta \mathbb{1}\{m_{i0}>0\} \times \log(m_{i0}) + x_i'\gamma + \varepsilon_i, \tag{4.4.1}$$

where m_{i0} is municipality i's pre-1513 emigration rate and x_i is a vector of various controls. This approach separates persistence on the extensive margin (whether a district sent anyone before 1513), captured by the coefficient α from persistence on the intensive margin (how much it sent, conditional on sending), captured by the coefficient β.[39]

The extensive margin alone already implies large persistence. In column (1), districts with any pre-1513 emigration send about 8.6 times more migrants per capita after 1513 than districts with none. Column (2) adds the intensive margin by interacting log pre-1513 emigration with an indicator for positive pre-1513 emigration; among districts that sent anyone before 1513, a 10% higher pre-1513 emigration rate predicts about 8.2% higher post-1513 emigration. Adding controls attenuates these estimates but leaves them economically large (columns (3)–(4)). Province fixed effects substantially reduce the extensive-margin contrast, but persistence remains strong within provinces: districts with any pre-1513 emigration experience about 76% higher post-1513 emigration (column (5)), and the intensive-margin elasticity remains sizable at 0.632 (column (6)).

The same pattern appears when we track exposure over time. Figure 11.(a) reports, by lustrum, the over-representation of emigration among places that had already sent at least one migrant by the previous lustrum—shown for both districts and municipalities. Even in the early 1500s, already-exposed places

2023, p. 107). There is also strong reason to believe that measurement error and idiosyncratic shocks account for a substantial part of this relationship, and that the true coefficient is closer to -0.05. This arises from implementing the "falsification test" of Spitzer (2021) and Spitzer and Zimran (2023) for β-convergence by regressing $\frac{m_{p1}}{m_{p0}}$ on m_{p1} (instead of m_{p0} as in the figure).

[38] We aggregate to districts to reduce the prevalence of zero pre-1513 emigration (so more places contribute to variation among senders) and to allow for local spillovers through migration networks. We code zero migration as zero, as in the tax regressions.

[39] This has the additional benefit of enabling us to include municipalities with no pre-1513 migration in the analysis while still using log migration as the regressor of interest.

dominate subsequent departures: municipalities (districts) with any migrant by 1499 exhibit emigration rates 15.1 (8.4) times those of unexposed places in 1500–1504. These gaps remain large and, if anything, grow over time; by 1525–1529, over-representation reaches 16.3 for municipalities and 29.4 for districts.

Figure 11(b) uses a complementary measure: for each exposure cutoff, it compares emigration in all remaining years of the sample. Municipalities (districts) exposed by 1499 are 6.4 (4.2) times over-represented in emigration over 1500–1540; those exposed by 1514 are 14.2 (18.6) times over-represented in 1515–1540. A similar picture emerges from the decade maps in Appendix Figure B.2: by around 1510, the spatial distribution of origins appears largely settled, even though large areas of the peninsula remained largely unaffected by emigration.[40] One of the most peculiar features of early Spanish migration to the Americas was that the map of migration coalesced very quickly. In fact, it largely stabilized within the first voyages, making it difficult to identify systematic changes later in the geographic origins of migration. This pattern is particularly striking given the greater challenges to information transmission before the advent of railroads, steamships, and the telegraph.[41]

4.5 Networks

The strong spatial persistence documented above points to migrant networks: early movers lowered the costs of migration for later cohorts and, in doing so, helped lock in the initial geography of departures. We test two implications. First, recent emigration from a place should predict higher emigration in the next period, holding fixed time-invariant local characteristics. Second, networks should shape not only the decision to migrate, but also the choice of destination.

Short-run origin networks Table 4 uses a district×lustrum panel and measures network size by emigration in the previous lustrum.[42] Columns (1)–(4) estimate BML models for the emigration count (with population as the exposure term), allowing for both an extensive margin (any emigration in $t-1$) and an intensive margin (how much emigration in $t-1$, conditional on being active) by adapting equation (4.4.1)

[40] After our period of study, the Basque country lost its preeminence as one of the main sending regions. As we discuss in Subsection 5.1, this region is genetically and culturally distinct from the rest of Spain, and the rest of the region the sent migrants. In Subsection 5.2, we discuss how migration for more than two centuries after our period of study ended, migration was concentrated on the areas of western Castile, and Seville. Only after the liberalization of ports of origin in the 1760s do we see other regions of Spain sending migrants in large numbers.

[41] The dynamics of almost every well-documented case of modern mass migration reflect both local persistence and gradual but meaningful spatial shifts. One could expect that access to migration was initially restricted to areas around the port of embarkation and to places with close ties to the seafaring industry. While the advantage of early starters would persist for some time, it should gradually diminish as information and access to migration networks spread across the country, thereby bringing additional regions with strong migration potential into the broader migration flow. Yet, as we show using several empirical approaches, this did not occur.

[42] The appendix reports analogous results using cumulative migrant stocks rather than short-run flows.

to a panel setting with district-level data,

$$\log\left(\frac{m_{jt}}{1-m_{jt}}\right) = \alpha \mathbb{1}\{m_{jt-1}=0\} + \beta \mathbb{1}\{m_{jt-1}>0\} \times \log(m_{jt-1}) + z'_{jt}\gamma + \varepsilon_{jt}, \tag{4.5.1}$$

where m_{jt} is the emigration rate of district j in lustrum t, m_{jt-1} is its emigration rate in lustrum $t-1$, and z_{jt} is a vector of controls, including period fixed effects, local geographic characteristics, and fixed effects at various levels of aggregation.

Without district fixed effects, lagged emigration predicts higher current emigration among previously active districts; for example, column (3) implies an elasticity of roughly 0.4.[43]

Columns (9)–(12) report linear panel estimates as a tractable benchmark:

$$m_{jt} = \beta m_{jt-1} + z'_{jt}\gamma + \varepsilon_{jt}. \tag{4.5.2}$$

These estimates remain positive even with district fixed effects (column 12). The coefficient implies that one additional migrant per 10,000 population in the previous lustrum predicts about 0.3 additional migrants per 10,000 in the current lustrum, relative to a mean of roughly 3 per 10,000 per lustrum (Table 1). Because the corresponding fixed-effects bias in this linear setting (Nickell bias) runs downward,[44] this estimate provides a conservative lower bound on short-run network persistence.

Explorers as local network nodes Figure 12 isolates a historically salient channel: recruitment driven by prominent expedition leaders and explorers. Motivated by accounts of hometown recruiting (e.g., for Pizarro's Peruvian campaign), we estimate an event-study design around the first explorer from each district using Callaway–Sant'Anna (which requires a linear specification). The estimates show a flat pre-trend and a sharp, persistent rise in emigration beginning in the lustrum of first exposure. The magnitude corresponds to a sizable increase in subsequent emigration per 10,000 inhabitants.[45]

4.5.1 Destination-specific networks

Network models predict path dependence in destinations: migrants should disproportionately follow earlier migrants from the same origin to the same destination, conditional on period-specific destination attractive-

[43] Including district fixed effects may further absorb time-invariant heterogeneity (column (4)). In this case, the coefficient shrinks toward zero and loses precision probably due to the incidental-parameter bias, or Nickel bias in this case, which can distort nonlinear fixed-effects estimates in panels. This downward bias is not present in the linear specification.

[44] Nickell bias scales as $\frac{1}{T}$ and has a known sign (Nickell 1981).

[45] The longest-horizon estimate relies on very few early-treated districts and has large standard errors; we omit it from the figure for clarity.

ness. We capture this channel using an origin-destination over-representation measure:

$$OR_{odt} = \frac{S_{odt}}{S_{dt}} = \frac{\frac{M_{odt}}{M_{ot}}}{\frac{M_{dt}}{M_t}}, \tag{4.5.3}$$

where S_{odt} is the share of migrants from origin o in period t who choose destination d, and S_{dt} is the corresponding national share in period t. We work with $\log OR_{odt}$ (the distribution is approximately log-normal and centered on zero, the point of equal representation), and we aggregate to the district×lustrum level to reduce small-number noise and to match the spatial scope of plausible networks.[46]

Figure 13.(a) shows a strong raw correlation between current and lagged destination over-representation.[47] The implied elasticity is 0.268: if a district was twice as over-represented in destination d in $t-1$, it remains meaningfully over-represented in t. Figure 13(b) shows that this relationship persists after (i) incorporating zero prior networks via an indicator and (ii) absorbing destination×lustrum shocks with fixed effects; the elasticity remains positive and economically meaningful (0.141).

An alternative version of the same analysis is given in Figure 13.(b). Whereas panel (a) depicts the relationship captured by the coefficient β in the equation

$$\log OR_{odt} = \alpha + \beta \log OR_{odt-1} + \varepsilon_{odt}, \tag{4.5.4}$$

panel (b) depicts the relationship captured by the coefficient β in the equation

$$\log OR_{odt} = \beta \log OR_{odt-1} \times \mathbb{1}\{M_{odt-1} > 0\} + \mathbb{1}\{M_{odt-1} = 0\} + \delta_{dt} + \varepsilon_{odt}. \tag{4.5.5}$$

That is, panel (b) is a binned residual scatter plot where the treatment variable is an interaction between the over-representation variable $\log OR_{odt-1}$ and an indicator for non-zero network $M_{odt-1} \neq 0$, which is equal to zero if the size of the network is zero and one otherwise. We also control for an indicator for zero-sized network. This enables bringing in the zero-sized networks into the regression. Put another way, this figure shows the correlation conditional on having a non-zero network, while controlling for an indicator for having no network. Finally, we add destination-lustrum fixed effects δ_{dt}. The estimated elasticity is somewhat smaller, 0.141, but is still both economically and statistically significant. Put briefly, the

[46] To avoid running into cases where the network treatment is not defined (i.e. if either $M_{ot} = 0$—there was no migration at all from origin o in period t—or $M_{dt} = 0$—there were no migrants from anywhere in Spain to destination d in period t—we aggregate the treatment at the district-lustrum level.

[47] The binned scatter plot shows the network treatment $\log OR_{odt-1}$ on the horizontal axis and $\log OR_{odt}$ is on the vertical axis. On both axes, equal representation is signified by a value of zero.

relationship documented in panel (a) is robust to accounting for cases in which there is no prior origin-destination network and to capturing changes in the attractiveness of destinations over time.

An alternative approach to measure the effect of networks on destination choices is to zoom in to the individual level. This provides both fine tuning and a micro-foundation for the migration decision. For each individual migrant, we ask whether his destination choice, conditional on his choice to migrate to the Americas, was affected by the choices of migrants from the same place of origin in the previous lustrum. The treatment variable is defined in the same manner, except that the lustrum period is defined with respect to each migrant's year of migration as a rolling window. We estimate this using the following baseline conditional-logit specification:

$$\Pr(d_i = d) = \frac{\exp[\beta_1 \log OR_{od,t-1} \times \mathbb{1}(OR_{od,t-1} \neq 0) + \beta_2 \mathbb{1}(OR_{od,t-1} = 0) + \delta_{dt}]}{\sum_{k=1}^{J} \exp[\beta_1 \log OR_{ok,t-1} \times \mathbb{1}(OR_{od,t-1} \neq 0) + \beta_2 \mathbb{1}(OR_{ok,t-1} = 0) + \delta_{kt}]} \tag{4.5.6}$$

where $OR_{od,t-1}$ is the over-representation in years $[t-5, t-1]$ and δ_{dt} captures destination-specific time trends or period fixed effects.[48] The estimates of β_1 are presented in Table 5. In Column (1), we consider all migrants from 1493 and control for destination-specific time trends. The estimated coefficient of the log network treatment is 0.242, which represents the effect on the log odds ratio. For example, if in year t 20 percent of all migrants choose destination d, and migrant i's district o was twice over-represented in this destination in the previous lustrum ($OR_{od,t-1} = 2$), then the predicted probability that he will choose this destination is approximately 23 percent (compared to the baseline 20 percent).[49] The coefficient remains similar in column (2) when the sample is restricted to migrants in the 1530s, which is the period where all destinations were available. In Columns (3)–(6), we reduce the sample to migrants going to the four primary destinations. Even when the destination indicators are interacted with year indicators or a time trend, or the period was or was not restricted to the 1530s, the estimates are statistically significant and hover around 0.15. Suggesting that a double over-representation in the previous period would increase the share of migrants in the destination from 20 to 22 percent.

[48] In some specifications, the estimation did not converge and we resorted to a smaller set of destination-specific time trends. We also experimented with specifications that were restricted to the 4 main destinations (Indies/Caribbeans, Mexico, Tierra Firme, and Peru)—in which the omitted cases are La Plata and Florida—or to the period from 1530, in which all destinations were already known and available. Notice that migrants with zero migration from their district of origin in the previous lustrum, or whose destination had no migrants at all in the previous lustrum, are omitted.

[49] The calculation follows from the conditional logit specification in Equation 4.5.6. Let $\Pr(d_i = d \mid OR = 1) = 0.20$ be the baseline probability when the district has equal representation ($OR_{od,t-1} = 1$). With $OR_{od,t-1} = 2$ and $\hat{\beta} = 0.242$, the predicted probability is:

$$\Pr(d_i = d \mid OR = 2) = \frac{\exp(\hat{\beta} \log 2) \cdot \Pr(d \mid OR = 1)}{1 - \Pr(d \mid OR = 1) + \exp(\hat{\beta} \log 2) \cdot \Pr(d \mid OR = 1)} = \frac{2^{0.242} \times 0.20}{0.80 + 2^{0.242} \times 0.20} \approx 0.228$$

This represents an increase of approximately 18% in the odds of choosing destination d (from odds of 0.25 to 0.30).

Evidence of origin-destination specialization is a hallmark of a strong role of migrant networks in shaping migration (Munshi 2015, 2020). In general, however, it is difficult to distinguish between the role of networks in driving this specialization and alternative explanations in which it is the product of observable or unobservable factors that make migration to a particular destination more attractive to individuals from a particular origin, such as alignment in climate (Obolensky et al. 2024) or an advantage in the destination labor market. In our setting, limited knowledge about destinations, their rapidly changing composition, and the lack of either much climate variability in Spain that would lead to differential specialization makes such alternative explanations far less likely to explain our findings.

4.5.2 Evidence from contracts

We now turn to the notarial contracts (protocols) to test whether migrants who shared the same notary, or notarial office, were more likely to travel together. The notarial records offer a unique window into the administrative infrastructure of colonial migration, because every prospective migrant had to formalize their passage before a notary (*escribano*) working in a given notarial office (*oficio*) in Seville. If notaries served as nodes in migration networks—connecting passengers to ship masters, channeling groups of migrants onto the same vessels—we should observe that pairs who shared a notary are disproportionately likely to have traveled with the same ship master (*maestre*). The sample consists of 1,711 migrants from the notarial protocols who have non-missing information on year, ship master, surname, notary, and notarial office, covering the period 1500–1540. From these migrants, we construct all 1,462,905 ($\frac{N(N-1)}{2}$) unique pairs and estimate linear probability models with multi-way clustered standard errors following (Cameron et al. 2011) (see Appendix B.3 for details).[50]

Table 6.(a) reports the results for the probability that two migrants traveled with the same ship master. In column (1), sharing a surname increases this probability by 2.9 percentage points, which is very large relative to the baseline probability of 0.5 percent (580%). This effect is stable across all specifications. Notice that the analysis is done at the household level. This means these numbers are underestimating the true correlation since does not take into account migrants with the same last name that travel within the same household. Whereas households are typically make of spouses and their children, it is not unusual for brothers or cousins to travel together. The analysis here is undercounting them, and the true number is larger. In column (2), we add an indicator for whether the two migrants used the same notary. Sharing an *escribano* raises the probability of sharing a master by 2.6 percentage points, an effect that is both statistically and economically

[50]Fafchamps and Gubert (2007) apply dyadic regression to study network formation in developing economies.

large. Column (3) shows that sharing a notarial office also predicts traveling with the same master, with a coefficient of 1.7 percentage points. However, as shown in column (4)—which includes both the notary and the office indicators—the coefficient on same *escribano* remains large and significant at 2.6 percentage points, while the coefficient on same *oficio* falls to essentially zero and is no longer statistically significant. This pattern is consistent with the notary—not the office—being the operative unit of the migration network. The office-level correlation observed in column (3) is entirely driven by the fact that migrants who share an office often share a specific notary within that office. Once we condition on the notary, the residual office effect disappears.

Table 6.(b) presents the same specifications for the probability that two migrants shared the same voyage, defined as traveling with the same ship master within a one-year window. The results follow a similar pattern. Sharing a surname increases the probability of co-traveling by 1.3 percentage points relative to a baseline of 0.3 percent. Sharing a notary raises it by 2.2 percentage points in column (2) and 2.1 percentage points in the full specification of column (4). The main difference relative to Table 6.(a) is that in the full specification (column 4), the coefficient on same office, while greatly reduced from column (3), remains statistically significant at 0.1 percentage points.

Three findings emerge from this analysis. First, sharing a surname is a strong predictor of co-migration, consistent with the role of kinship networks documented in the literature (Almorza Hidalgo 2018; Altman 1991, 2021; Vázquez de Prada Vallejo 1991). Second, the notary emerges as a central figure in the migration process. The notary effect is at least as large as the surname effect, and in the case of same-voyage regressions, it is the single strongest predictor. Third, the notarial office matters only insofar as it houses specific notaries. The sharp drop in the office coefficient when we control for the notary shows that network effects operate at the level of individual notaries rather than at the level of the institutional office. These results are consistent with a model in which notaries served as brokers in the migration market, connecting prospective migrants with ship masters and facilitating group departures. The concentration of the effect at the notary level, rather than the office level, suggests that these brokerage relationships were personal rather than institutional.

4.5.3 Evidence from ships registers

Table 7 report results from the dyadic regressions, similar to above but using data from the ships registers (*asientos*) hand collected from the *Archivo General de Indias*. Each observation is a pair of migrants and the outcome indicates whether the pair traveled with the same ship master (panel (a)) or on the same voyage (panel (b)). We estimate linear probability models with multi-way clustered standard errors following

Cameron et al. (2011), and present several specifications that progressively add geographic controls and fixed effects. The main finding is that town-level ties are the strongest and most robust predictor of shared migration outcomes. In column (1), sharing a town of origin raises the probability of traveling with the same ship master by 0.43 percentage points ($p < 0.01$). Given the low baseline probability of any two migrants sharing a master (0.75 percent), this is a large effect. The coefficient on same town is remarkably stable across specifications: it remains positive and highly significant whether we control for province, municipality, or include year, decade, and province fixed effects. In the most demanding specification that includes all three geographic levels and province fixed effects (column (10)), the same-town coefficient is 0.53 percentage points for the same-master outcome and 0.58 percentage points for same voyage, both significant at the 1% level.

The results for the same-voyage outcome (Table 7.(b)) closely mirror those for same master, which is expected since same voyage is a stricter version of the same-master variable that additionally requires temporal overlap. The slightly smaller coefficients in the baseline specifications reflect this additional restriction, but the overall pattern of results is nearly identical. When we enter each geographic level separately (columns (1),(3), and (4)), the coefficients decline monotonically with the breadth of the geographic unit: 0.38 percentage points for same town, 0.30 for same municipality, and 0.16 for same province in the same-master regression. This gradient confirms that finer geographic ties carry more predictive power. However, much of the municipality and province effects reflects composition: when all three levels are included simultaneously (columns (2), (6), and (10)), the municipality coefficient becomes small and often insignificant, and in some cases turns negative.[51] This pattern suggests that the municipality variable adds little information once we already know whether two migrants come from the same town. In other words, networks are very local.

Taken together, these results provide strong evidence that migration to the Americas was organized through local, town-level networks. Migrants from the same town were significantly more likely to share a ship master and to travel on the same voyage, even after accounting for provincial geography and temporal trends. The persistence of this pattern across a wide range of specifications suggests that the coordination of transatlantic migration operated primarily through fine-grained social ties rooted in specific communities of origin.

[51] The city of Seville requires especial consideration here. The original documents records the neighborhood (*colación*) of residency of the migrant. We consider each Sevillian neighborhood as a separate "town" within the municipality of Seville.

5 Discussion

We find clear evidence that social networks played an important role in shaping the first waves of migration to the Americas. They determined both the aggregate size of a migration flow from each origin and the distribution of this flow across origins. Yet the nature of this influence is somewhat unusual compared with other episodes of network-driven migration: the "map" of emigration origins coalesced early and remained remarkably stable. Spain developed a migration belt running from Andalusia in the south to the Basque Country in the north, while large swaths of the country—particularly Galicia, Aragon, Navarra, and much of eastern Castile—consistently sent few migrants. Observable geographic, economic, or legal characteristics do not account for the stark contrast between these two migration regimes, nor does the pattern conform to the expectation that, over time, networks diffuse migration more broadly across origins.

5.1 Where do networks come from?

Our results raise a puzzle. Migration networks formed almost immediately after Columbus' maiden voyage and remained geographically fixed for the remainder of our period. They did not spread to new regions, even as the volume of migration grew and new American destinations opened up. The standard explanation for persistence in migration—that time-invariant local characteristics favor emigration from certain places—does not fit our findings well. Observable geographic and economic characteristics explain little of the variation in emigration once we control for distance to Seville. The persistence patterns remain present even after we control for local fixed effects. Something else kept migration anchored in a specific set of regions running from Andalusia through Extremadura and into Old Castile, but not into Galicia, Eastern Castile, or Eastern Andalucia, let alone into the Kingdom of Aragon.[52] The migration networks channeling Spaniards to the Americas were not created by the migration itself. They were inherited from a much older structure of social connections rooted in the *Reconquista* and visible today in both the genetic makeup and the linguistic geography of the Iberian Peninsula.

Recent work in population genetics provides the most striking piece of evidence. Bycroft et al. (2019) used haplotype-based methods to analyze the DNA of modern-day Spanish individuals and mapped the resulting genetic clusters onto geography. Their central finding is that the boundaries of these clusters—inferred purely from shared DNA, with no geographic information used in the inference—correspond closely to the boundaries of medieval kingdoms and linguistic regions that existed over 500 years ago. The main

[52] Martínez Shaw (1980) shows that there were no legal restrictions for Aragonians to travel to the Americas, and we indeed have some migrants from Aragon in the first decade.

axis of genetic differentiation runs east to west, separating the Galician, Astur-Leonese, Castilian, Basque, Aragonese, and Catalan domains from each other. In contrast, genetic differences in the north-south direction within each of these domains are remarkably small. Bycroft et al. (2019) interpret this pattern as reflecting the *Reconquista*, during which the northern Christian kingdoms expanded southward from the eighth century onward, carrying their populations—and their genes—into newly conquered territory. The result is that modern-day individuals from Andalusia are genetically much closer to individuals from Old Castile, hundreds of kilometers to the north, than to individuals from Valencia or Catalonia at comparable distances to the east.

The correspondence between the genetic map of Spain and our map of migration to the Americas is remarkable (Figure 14). With the exception of the Basque country, which, as discussed elsewhere in the paper came to be linked to the migration via its shipbuilding industry, the remaining regions that sent migrants—Andalusia, Extremadura, and Old Castile—fall within the area of genetic homogeneity created by the north–south population movements of the *Reconquista*. The red dots in Figure 14.(a) extend to Cordoba and Granada in Andalusia, even if the area is colored yellow, indicating the presence of Leonese descendants in the area, despite Castilians being the majority in the 20^{th} century. The regions that did not participate—Galicia, Aragon, Catalonia, Valencia—are precisely those that are genetically differentiated from the Castilian core. This overlap is too close to be coincidental. It suggests that the social networks linking, say, a town in Extremadura to a town in Andalusia in 1500 were not an artifact of the migration to the Americas; they were the product of centuries of shared ancestry, intermarriage, and population exchange along the north–south corridor of the *Reconquista*. Two features of the genetic evidence are especially telling. First, within central and western Spain—the area corresponding to the old Kingdom of León and the Crown of Castile—Bycroft et al. (2019) find that individuals separated by hundreds of kilometers cannot be reliably distinguished on the basis of their DNA. This genetic homogeneity implies extensive mixing and population movement over centuries, precisely the conditions under which dense social networks would form across large distances. The towns of Extremadura were not isolated from those of Old Castile or Andalusia; they were connected by centuries of shared settlement, migration, kinship, and institutional ties going back to the *Reconquista*. When the opportunity to migrate to the Americas arose, these pre-existing connections provided the infrastructure through which information, assistance, and recruitment could flow. Second, and equally important, the regions that did *not* participate in the migration to the Americas are precisely those where the genetic evidence points to social isolation. Galicia, in the northwest, displays an extraordinary degree of fine-scale genetic differentiation, with distinct clusters appearing at scales below 10 kilometers,

often corresponding to individual river valleys (Bycroft et al. 2019). This is the finest-scale population structure observed anywhere in humans to date, and it implies that Galician communities were remarkably self-contained, with limited gene flow even between neighboring parishes. Such fragmentation is consistent with a society in which social networks were intensely local and did not extend far beyond the immediate community. Despite its large and dense population, its active seafaring tradition, and its Atlantic coastline, Galicia sent almost no one to the Americas during our period. The genetic evidence suggests a reason: the social networks that connected Galician villages to one another, let alone to distant regions, were too thin and too localized to support the kind of long-distance chain migration that linked southwestern Castile to the New World.[53]

Linguistic geography tells the same story from a different angle. The Astur-Leonese language originated in the Kingdom of Asturias—the Christian stronghold that was never conquered by the Moors—and spread southward as the Kingdom of León expanded during the *Reconquista*. As Menéndez Pidal (1906) first showed, Astur-Leonese developed from Latin independently of Castilian, and the two languages are distinguished by several well-known phonological features.[54] The linguistic domain of Astur-Leonese traces a vertical corridor from Asturias through León, Zamora, and Salamanca into Extremadura, where Extremaduran retains transitional features between Leonese and Castilian. This is precisely the north–south corridor that produced the bulk of migration to the Americas. Place names provide tangible evidence of this population movement. Throughout Extremadura and parts of western Andalusia, castellanized Astur-Leonese toponyms—town names that bear the phonological imprint of Leonese settlers even after centuries of linguistic replacement by Castilian—mark the trail of the *Reconquista* settlement. These names are absent east of the Castilian–Leonese linguistic boundary, just as migration to the Americas was almost absent from Eastern Castile.

The linguistic and genetic evidence thus converge on the same conclusion: the *Reconquista* created a specific geography of social connection within the Iberian Peninsula. As the Christian kingdoms pushed southward from the eighth century onward, they carried populations—and the networks binding those populations together—from the northern mountains into the newly conquered lowlands. The kingdom of León settled Extremadura with people from Asturias, León, and Zamora; Castile settled La Mancha and parts

[53]The Basque Country presents a similar pattern of genetic isolation from the rest of Spain, though its participation in the transatlantic migration—driven largely by its shipbuilding industry—followed a different mechanism rooted in the maritime economy rather than in the *Reconquista* settlement networks.

[54]By the late 15^{th} century, Astur-Leonese and Castilian had diverged on several well-known phonological features despite their common Latin origin: 1) Latin initial *f-*, which Castilian shifted to aspiration and eventually to silence (Latin *filium* > Castilian *hijo*), while Astur-Leonese preserved the labiodental (*filu*, *fiyu*); 2) Latin *-LJ-* and *-C'L-* yielded the palatal lateral /ʎ/ in Leonese (e.g., *ovella*, *ollo*) but /ʒ/ > /x/ in Castilian (*oveja*, *ojo*); Leonese retained the cluster *-mb-* (e.g., *palomba*) where Castilian simplified it to *-m-* (*paloma*) (Menéndez Pidal 1906; Penny 1990, 2002).

of Andalusia with people from Old Castile. When the opportunity to migrate to the Americas arose in 1492, these pre-existing corridors determined who could be reached by the networks radiating outward from Seville and who could not. The implication for our understanding of migration networks is important. The networks we document in the first decades of Spanish migration to the Americas were not built from scratch by the pioneers who crossed the Atlantic. They were the extension of a social infrastructure that had been accumulating for seven centuries of *Reconquista* settlement. The pioneers activated these networks for a new purpose—transatlantic migration—but they did not create them. This helps explain both why networks formed so quickly after Columbus (the underlying connections were already in place) and why they did not spread to new regions (the boundaries of the old kingdoms/genetics/language remained barriers to social connection). It also helps explain why the Seville monopoly was so effective at freezing the geography of migration: Seville sat at the southern terminus of the Leonese–Castilian settlement corridor, and the networks it could activate were precisely those that ran northward along that corridor. Regions outside this corridor—Galicia, Aragon, Catalonia—were separated from Seville not just by distance but by centuries of social and genetic differentiation that no amount of information about the Americas could easily bridge.

Finally, while genetic and linguistic evidence on the structure of social networks can explain why the geography of migrant origins remained constrained to Old Castile and Andalucia, it does not explain why migration was so widespread geographically already in the early years of the migration and why the geography was fixed quickly rather than originating in one area and expanding thereafter. The history of the *Reconquista*, however, provides a potential explanation. The initial voyages of exploration—after it became clear that they had were part of a project of conquest rather than simply trade—recruited individuals who had demonstrated their military ability in the conquest of Granada, which had concluded in 1492 (Szászdi León-Borja 2019, p. 13). Hillgarth (1978, pp. 377–378) discusses the manner in which the army that participated in this conquest was recruited, emphasizing that it was primarily Castilian. Thus, it is possible that the rapid establishment of the geography of emigration throughout Castile was the product of the use of existing structures of military recruitment that could be exploited to assemble crews for voyages to the Americas from throughout Castile. Moreover, the geographic extent of migration may have overlapped significantly with the geographic extent of the *Hermandad* (Ladero Quesada 2005), which was an important contributor to the mobilization for the conquest of Granada (Hillgarth 1978). Francisco Pinelo was a Sevilian merchant of Genoese origin. He was the *Hermandad*'s treasurer for 1490-1492, and he was in charge of organizing Columbus second (and largest) trip in 1493 (Ladero Quesada 2005). The regions in the northern coast of Spain, largely provided money and ships/sailors, not soldiers, to the work effort (Ladero Quesada 2005).

Moreover, by the end of the war in Granada, neighboring regions such as Murcia, Albacete, and Ciudad Real did not contribute soldiers through the *Hermandad*. Either they were contributing soldiers directly through their *Concejos*, or supplied money or mercenaries (Ladero Quesada 1967). Therefore, the area of migrants corresponds exactly with the catchment area of the *Hermandad* and the area around Seville. The connection of Francisco Pinelo, among others, between the *Hermandad* and the early voyages of exploration provides the final link between the *Reconquista* and the new conquests. Indeed, individuals from Old Castile were congregated in southern Spain in 1492 when the conclusion of the conquest of Granada led them to direct their attention elsewhere.[55] Individuals recruited in this manner would then have brought their friends and relatives into the migration, further perpetuating the geography of migration established by the first few voyages. Taken together, this evidence suggests that the *Reconquista* closed the frontier at home but forged the networks that carried the conquest across the empire's expanding American frontiers.

5.2 Discontinuous networks: From colonial to mass migration

Our findings raise a natural question: what happened to the migration networks we document after 1540? The map of Spanish migration to the Americas that we observe in our period—anchored in Andalusia, Extremadura, and the Basque Country, with little participation from Galicia, Aragon, or Catalonia—was both remarkably stable and yet strikingly different from the geography of later Spanish migration. Tracing how these patterns evolved over the following three centuries helps put our results in perspective and sheds light on the conditions under which migration networks persist, break down, or are displaced by new ones. García Hidalgo (2019) and Arroyo Abad and Espín-Sánchez (2024) document that the broad contours of migration established during the first half of the 16^{th} century persisted well into the 17^{th} century. Andalusia remained the leading source of migrants, followed by Extremadura, and the main destinations continued to be New Spain and Peru (see also Martínez Shaw 1994; Mörner 1975). Boyd-Bowman's (1976) figures for the second half of the 16^{th} century show that the five provinces of Seville, Badajoz, Cáceres, Toledo, and Valladolid together supplied over half of all identified emigrants through 1600. Jacobs (1995) confirms that during the reign of Philip III (1598–1621), Castilian dominance in the migration to the Americas was largely unchanged. In other words, the migration belt we identify in our data—running from Andalusia through Extremadura and into the northern Meseta—remained the backbone of Spanish transatlantic migration for at least a century after our period ends.

[55]The Christians secured a Papal Crusade, and enlisted the largest army in Medieval Europe, with up to 70,000 people (Quesada 2001). The Grenadine king, Boabdil, unexpectedly surrendered at the beginning of the siege. This means Muslims were allowed to keep all their assets, depriving the conquering soldiers of the expected spoils of war.

The persistence of these geographic origins during the 16^{th} and 17^{th} centuries is consistent with the strength of the network effects we document. The same regions that sent the first migrants after Columbus continued to supply disproportionate numbers of settlers throughout the colonial period. Altman's (1989) shows how Extremaduran conquistadors' hometowns remained tied to specific American destinations across generations, with families recruited relatives and townspeople to follow them. Altman's (2000) traces a similar chain between Brihuega (in New Castile) and Puebla (in Mexico) from 1560 to 1620. It illustrates how a single town could sustain a destination-specific flow for decades. These chains took shape in our period, and the qualitative evidence indicates they continued operating along essentially the same geographic lines for over a century. At the same time, the 17^{th} century brought important changes in the volume and composition of migration. Aggregate flows declined after the peak around 1590–1630, and the proportion of women increased substantially—reaching about one quarter of all emigrants in the 17^{th} century (Arroyo Abad and Espín-Sánchez 2024; García Hidalgo 2019, 2021). Arroyo Abad and Espín-Sánchez (2024) estimate nearly 83,000 migrants during the 17^{th} century.[56] The crisis of the Spanish economy, the fiscal strain of European wars, and recurrent epidemics all contributed to slowing the flow (Mörner 1991). But the geographic origins remained largely the same, suggesting that the networks established in our period were durable even as the scale of migration contracted.

The most dramatic transformation in the geography of Spanish migration came not from the gradual diffusion of networks—which, as we show, was notably absent during our period—but from a change in policy. In the 1760s, the Crown opened trade with the American colonies to nine peninsular ports, ending the centuries-old monopoly of Seville and later Cádiz. The *Reglamento para el Comercio Libre* of 1778 extended this reform further, authorizing thirteen Spanish ports to trade directly with the Americas: Alicante, Barcelona, Cádiz, Cartagena, Gijón, La Coruña, Málaga, Palma, Santa Cruz de Tenerife, Santander, Seville, and the *Alfaques* of Tortosa, among others (Márquez Macías 1995; Pérez-Artés 2024). For the first time, regions with their own ports could bypass Seville entirely and send migrants directly to the Americas. The effects of port liberalization on the geography of migration were substantial. Márquez Macías (1995) and (Arroyo Abad and Espín-Sánchez 2024) show that the origins of migrants shifted noticeably toward northern and eastern Spain for the period 1765–1824. Catalonia, which had been virtually absent from migration during our period and through most of the colonial era, began to send significant numbers of migrants to the Caribbean, particularly to Cuba, drawn by commercial opportunities in the textile trade and later the slave trade (Delgado Ribas 1982; Sonesson 1995). Galicia, which we find sent almost no migrants during

[56] Other estimates, such as Mörner (1975), place the total closer to 100,000 migrants during the 17^{th} century. Arroyo Abad and Espín-Sánchez (2024) show that these figures overstate the flow.

1492–1540 despite its large and dense population, emerged as one of the leading sources of emigration in the late colonial and especially the 19^{th} century (Eiras Roel and Guimerá Ravina 1991; Rodríguez Galdo 1993). Asturias and Cantabria followed a similar pattern (Barreiro Mallón 1991). It is also remarkable how some regions, like Murcia, which had not participated in the colonial migration, also did not experienced substantial migration growth after the port liberalization. This shift is striking in light of our results. We show that the map of migration was essentially fixed by 1510 and that the networks established in the first two decades after Columbus showed no sign of spreading to new regions. The geographic distribution of migration origins remained concentrated in the same areas for over 250 years. It was not the gradual diffusion of information or the extension of social networks that eventually brought Galicia, Asturias, or Catalonia into the transatlantic migration flow. Rather, it took a change in the institutional framework—the removal of the port monopoly—to break the geographic lock-in. This finding reinforces our interpretation that the persistence we observe during our period was not simply a function of time-invariant local characteristics favoring migration, but reflected the self-reinforcing nature of networks that, once established, channeled migration along fixed paths and did not extend to incorporate new regions.

The Galician case is particularly telling. As we note in our analysis, Galicia's absence from the early migration was not easily explained by geographic factors alone. Galicia had a dense population, high rates of internal migration, and an active seafaring tradition along its coast (Eiras Roel and Guimerá Ravina 1991). Yet during our period, Galician participation in transatlantic migration was negligible. The most likely explanation, consistent with our broader argument, is that Galicia lay outside the social networks that connected southwestern Castile to the American enterprise—networks rooted in the Reconquista, in the conquest of Granada, and in the administrative apparatus centered in Seville. Once the port monopoly was lifted and Galicians could embark from La Coruña and Vigo, emigration surged. By the 19^{th} century, Galicia became the single largest source of Spanish emigration to the Americas, with flows directed primarily to Cuba, Argentina, and Uruguay (De Cristóforis 2011; Eiras Roel and Guimerá Ravina 1991). The Galician case thus suggests that the absence of migration from certain regions during our period was not due to a lack of potential migrants or push factors, but to the absence of the network infrastructure needed to connect them to the transatlantic flow. The contrast between the colonial and mass migration periods also highlights a difference in the type of migrant. During our period, as we document, migration was an urban phenomenon, with migrants drawn from across the social hierarchy—not just the poor. By the late 18^{th} and especially the 19^{th} century, the new migrant streams from northern Spain were composed largely of rural laborers pushed by agricultural crisis, land fragmentation, and demographic pressure (Eiras Roel and Guimerá Ravina 1991;

Sánchez Alonso 1995). This shift is consistent with the broader transformation of European migration during the Age of Mass Migration, in which falling transport costs and the spread of information gradually made migration accessible to poorer and more rural populations (Hatton and Williamson 1998). In Spain, the opening of new ports played a role analogous to the reduction in transport costs: it lowered the effective cost of migration for populations that had previously been excluded by geography and by the network structure inherited from the 16^{th} century. Finally, even after the liberalization of ports, some of the patterns we document in the early period left lasting traces. Extremadura, which we identify as a major source of migrants in 1492–1540, continued to send emigrants to the Americas in the late colonial period (Márquez Macías 1995) and, in smaller numbers, during the Age of Mass Migration (Sánchez Rubio 1993). The Basque Country, which we find was an important source of crew members and migrants in our period, maintained a continuous presence in the Americas throughout the colonial era, with Basque merchants and administrators playing key roles in colonial commerce and governance (Alvarez Gila and Angulo Morales 2002). These long-run continuities suggest that, even as new regions entered the migration flow in the late 18^{th} century, the original networks established in the first decades after Columbus' maiden voyage were never fully displaced. They coexisted with—rather than being replaced by—the newer streams from northern Spain, creating a layered geography of migration that reflected successive waves of network formation.

In sum, we can draw three conclusions from this comparison. First, the migration networks established in the first decades after Columbus were remarkably durable, persisting with little geographic diffusion for over two centuries. Second, the eventual broadening of the migration's geographic origins was driven not by the gradual spread of networks but by an institutional shock—the liberalization of ports—that lowered barriers for previously excluded regions. Third, even after this shock, the old and new migration streams coexisted, suggesting that once established, migration networks are difficult to displace entirely. These findings speak to the broader literature on the role of networks in migration by showing that, in the absence of major institutional changes, network-driven migration can remain geographically concentrated for very long periods, and that the "ink in a white paper" model of gradual spatial diffusion is not a universal feature of migration dynamics.

5.3 Networks before networks: Early roots of mass migration Patterns

A large and growing literature in economics studies the role of social networks in shaping migration during the Age of Mass Migration (AMM), roughly 1840–1920. Our findings push the origins of the mechanisms identified in that literature back by more than three centuries, and offer a setting in which network formation

can be observed from its very beginning rather than inferred from an already mature process.[57] The most direct point of contact between our results and the modern literature is the "diffusion hypothesis" advanced by Spitzer and Zimran (2023). Studying Italian emigration between 1876 and 1920, they show that mass migration began in a small number of epicenters and spread gradually to neighboring areas in an orderly spatial pattern, "like an ink blot on paper" (Moya 1998, p. 113). Migration rates across Italian districts eventually converged as networks reached previously unconnected communities and sparked emigration from there. Spitzer (2021) documents a similar process for Jewish emigration from the Russian Empire's Pale of Settlement: migration spread from the northwestern provinces bordering Germany toward the east and south over the course of two decades, driven not by pogroms or local economic conditions but by the gradual arrival of chain-migration networks. In both cases, the key finding (with respect to our results) is that networks did not simply lower the cost of migration for a fixed set of communities; they *expanded* geographically over time, pulling new regions into the migration flow.

Our results present a sharp contrast. We find that the geographic origins of Spanish migration to the Americas were essentially fixed by 1510 and showed no sign of spatial expansion over the following three decades. Regions that did not participate in the earliest voyages—Galicia, Aragon, Catalonia—remained largely absent throughout our period, even as the volume of migration grew substantially. There is no evidence of the kind of gradual convergence that Spitzer and Zimran (2023) document for Italy or that Spitzer (2021) finds for the Jewish migration. Instead, the map of origins remained static. What explains this difference? We argue that institutions played a central role. During our period, all legal emigration to the Americas was channeled through a single port—first Seville, later Cádiz—under the close supervision of the *Casa de Contratación*. This institutional bottleneck meant that access to the migration flow depended on proximity to, and connections within, the Seville-centered administrative and commercial system. Regions without such connections—even those with large populations and strong push factors, like Galicia—were effectively locked out. By contrast, the institutional environment of the 19^{th} century was far more open. Italian migrants could leave from any number of ports, and there was no centralized licensing system controlling who could emigrate. In that setting, networks could spread through spatial proximity alone: a successful migrant from one village could pass information and assistance to relatives and acquaintances in nearby villages, and the process could repeat outward. In early modern Spain, the Seville monopoly short-circuited this mechanism. Networks could deepen within the regions already connected to Seville, but they could not easily spread to new regions, because the institutional framework did not allow it.

[57] Even if the first migrants from a particular place in the Age of Mass Migration can be identified, it is generally not possible to distinguish between true pioneers and those following in the footsteps of migrants from elsewhere.

This interpretation finds support in the fact that when the institutional barriers were eventually removed—through the port liberalization of 1765 and 1778, as we discuss in the preceding subsection—the geography of Spanish migration did broaden, and a process resembling Spitzer and Zimran's diffusion did take hold. Moya (1998), studying Spanish immigration to Buenos Aires (1850–1930), shows that Galician migration expanded from a handful of coastal towns and parishes in the mid-19^{th} century to encompass the entire Atlantic coast of Galicia by the early 20^{th} century, in a pattern of geographic spreading that closely matches the diffusion model. De Cristóforis (2011) documents a similar process for Galician emigration, noting that by 1850 the origins were already less restricted than is sometimes assumed, even though the movement continued to spread inward toward the provinces of Lugo and Orense well into the 20^{th} century. The contrast between the frozen geography of our period and the gradual diffusion of the mass migration era thus points to the importance of institutional context in determining whether networks expand spatially or remain locked in place.

The Spanish case during the AMM also offers a useful bridge between our period and the broader European experience. Sánchez-Alonso (2000) shows that regional variation in Spanish emigration rates between 1880 and 1914 was not well explained by wages or land inequality alone; instead, prior emigration history—a proxy for networks—was a strong predictor of current outflows. Sánchez-Alonso (2019), in a comprehensive survey of mass migration to Latin America, notes that about 13 million Europeans went to the region between 1870 and 1930, with Spain contributing more than four million. The geographic origins of these Spanish emigrants were concentrated in precisely the northern regions—Galicia, Asturias, Cantabria—that had been absent from the colonial migration we study, while the southern regions that dominated our period were much less prominent in the mass migration era. This reversal underscores the point that network geography may not be a fixed reflection of underlying regional characteristics. It is path-dependent, shaped by the specific historical circumstances under which networks first formed.

Hatton and Williamson (1998), in their influential treatment of the economics of the AMM emphasize the role of falling transport costs, rising incomes at origin (which relaxed liquidity constraints), and the stock of previous migrants at the destination (a proxy for network effects). Our findings suggest that all three mechanisms were already present in embryonic form during the early colonial period. Transport costs were enormous in the early 16^{th} century, but they were partly offset by institutional arrangements (Crown subsidies, encomienda grants, and financing by relatives) and by the commercial infrastructure centered on Seville. Liquidity constraints clearly mattered: migration in our period was not a mass phenomenon but was drawn disproportionately from the middling and upper ranks of society. And the stock of previous

migrants shaped subsequent flows in exactly the way the modern literature would predict—through the formation of town-to-destination chains that persisted over decades. One implication of our findings is that the network mechanisms documented in the AMM literature may have deeper historical roots than is commonly recognized. The AMM is often treated as the starting point for the systematic study of migration networks, in part because it is the earliest period for which large-scale quantitative data exist. But our evidence shows that the same forces—self-reinforcing chains, geographic persistence, network effects on destination choice—were at work in the very first years of European overseas expansion. This does not mean that the early modern and modern migration systems were identical; they obviously differed in scale, institutional context, and the composition of migrants. But it does suggest that the tendency for migration to cluster geographically and to channel through networks of kinship and community is a deep structural feature of human mobility, not an artifact of the particular economic and technological conditions of the nineteenth century.

6 Conclusion

The colonization of the Americas was arguably the most consequential migration episode in human history, reshaping the demography, culture, and political economy of three continents over five centuries. Yet we have shown that the migration that set this process in motion was remarkably small in volume and geographically concentrated. A few thousands of individuals over five decades from the Kingdom of León and the Crown of Castile initiated the flows we study. Pioneers were not systematically selected on wealth, or social status: *hidalgos* alongside artisans, merchants alongside servants appeared throughout the early crossings. What they had in common was falling into the social and administrative orbit of Seville. Their individual characteristics had little power to predict who migrated and who stayed. Within the constraints of Castilian origin and proximity to Seville, the initial draw was largely random.

This randomness makes the persistence all the more striking. Once these links were formed, they endured. The Seville monopoly over transatlantic trade froze the geography of migration where the first networks had formed, around 1510. For over 250 years, the same regions sent migrants while Galicia, Catalonia, and Aragon—regions with large populations and, in Galicia's case, strong seafaring traditions—remained locked out. Networks deepened within the connected regions but never spread outward, in sharp contrast with the spatial diffusion observed during the Age of Mass Migration in Italy or Eastern Europe. It took an institutional shock—the Bourbon port liberalizations of the 1760s—to break this pattern. The implication is

clear: institutions do not merely shape the cost of migration; they determine whether networks can expand or remain frozen in place. This interaction between networks and institutions is, we believe, among the most important and least understood dimensions of migration dynamics. The broader implication is that institutions do not merely shift migration costs at the margin.

The effects of this pattern were enormous and lasting on both sides of the Atlantic. At the destinations, the Andalusian and Extremaduran character of the first settlers shaped colonial dialects, legal customs, settlement patterns, and social hierarchies in ways still visible today. At the origins, participating regions became tied to the colonial economy through remittances, return migration, and commercial exchange. Transatlantic departures reshaped sending communities inside Spain, affecting their local labor supply. They may have induced internal relocation toward nodes such as Seville as well. Regions outside the Sevillian corridor such as Galicia and Catalonia followed different development trajectories, joining the transatlantic flow only centuries later under very different conditions. These divergent paths—both in Spain and in the Americas—trace back to the accident of which towns were connected to Seville's networks in the 1490s, not to deep structural differences between regions.

Several open questions follow. First, how did the networks we document shape colonial institutions and long-run development in the Americas? If the origins of settlers were largely random conditional on Seville proximity, then variation in migrant characteristics across American destinations provides leverage for studying how founding populations affected institutional formation. Second, did the similar network mechanisms operate within the regions connected to Seville? Our data show little spatial diffusion at the national level, but whether networks spread locally—from town to neighboring town within Extremadura or Andalusia—remains an open question that data for 1540-1640 could address.[58] Third, how general is the institutional lock-in we document? The Spanish case, characterized by a highly centralized port monopoly, may represent an extreme example; however, comparable administrative bottlenecks—such as visa regimes, colonial licensing systems, and restricted ports of entry—have constrained migration flows in a wide range of settings. Understanding when institutions freeze networks and when they allow diffusion is a question with relevance far beyond the sixteenth century. Finally, our finding that network mechanisms documented for the Age of Mass Migration were already operating three centuries earlier raises a broader point: the tendency for migration to cluster geographically and to channel through ties of kinship and community appears to be a deep structural feature of human mobility, not an artifact of 19^{th}-century conditions. *Caminante, sí hay camino*, but institutions do not offer it evenly. The early Spanish migration to the Americas shows how

[58]See Arroyo Abad and Espín-Sánchez (2024) for a general analysis on migration flows from Spain to the Americas during the entire colonial period.

the interaction between contingency, networks, and institutions can produce outcomes of world-historical significance from the most modest of beginnings.

References

Acemoglu, Daron, Simon Johnson, and James A. Robinson (2001). "The Colonial Origins of Comparative Development: An Empirical Investigation". *American Economic Review* 91:5, pp. 1369–1401.

Almorza Hidalgo, Amelia (2018). *"No se hace pueblo sin ellas" : mujeres españolas en el virreinato del Perú : emigración y movilidad social (siglos XVI-XVII)*. 1 online resource : illustrations vols. Colección Americana; Núm 67; Núm. 44; núm. 67.; no. 44. Madrid, Sevilla: Consejo Superior de Investigaciones Científicas ; Universidad de Sevilla.

Altman, Ida (1989). *Emigrants and Society: Extremadura and America in the Sixteenth Century*. University of California Press.

——— (1991). "A New World in the Old: Local Society and Spanish Emigration to the Indies". In *To Make America: European Emigration in the Early Modern Period*. Ida Altman and James Horn (ed.). Berkeley, CA: University of California Press.

——— (2000). *Transatlantic Ties in the Spanish Empire: Brihuega, Spain and Puebla, Mexico, 1560–1620*. Stanford: Stanford University Press.

——— (2021). *Life and Society in the Early Spanish Caribbean: The Greater Antilles, 1493–1550*. Baton Rouge, LA: LSU Press.

Alvarez Gila, Oscar and Alberto Angulo Morales (2002). *Las Migraciones Vascas En Perspectiva Histórica (Siglos XVI–XX)*. Bilbao: Universidad del País Vasco.

Archivo General de Indias (1940). *Catálogo de Pasajeros a Indias Durante Los Siglos XVI, XVII y XVIII*. 4. Sevilla: Imprenta editorial de la Gavidia.

Arroyo Abad, Leticia and José-Antonio Espín-Sánchez (2024). *The First Numbers of the Americas*. Working Paper DP20591. Centre for Economic Policy and Research.

Arroyo-Abad, Leticia and Nuno Palma (2021). "The Fruits of El Dorado: The Global Impact of American Precious Metals". In *The Fruits of the Early Globalization: An Iberian Perspective*. Rafael Dobado-González and Alfredo García-Hiernaux (ed.). Cham: Springer International Publishing, pp. 95–131.

Ballesteros Beretta, Antonio (1945). *Cristóbal Colón y el descubrimiento de América*. 2 vols. Vols. IV and V of *Historia de América y de los pueblos americanos*. Barcelona: Salvat.

Barreiro Mallón, Baudilio (1991). "Ritmo, Causas y Consecuencias de La Emigración Asturiana a América, 1700–1850". In *La Emigración Española a Ultramar, 1492–1914*. Antonio Eiras Roel (ed.). Colección Actas 1. Madrid: Ediciones Tabapress, pp. 41–58.

Boyd-Bowman, Peter (1964a). *Indice Geobiográfico de Cuarenta Mil Pobladores Españoles de América En El Siglo XVI, Tomo I, 1493-1519,* Bogotá: Instituto Caro y Cuervo.

——— (1964b). *Índice Geobiográfico de Más de 56 Mil Pobladores de La América Hispánica I. 1493-1519*. Mexico City: Fondo de Cultura Economica.

——— (1968). *Indice Geobiográfico de Cuarenta Mil Pobladores Españoles de América En El Siglo XVI, Tomo II, 1520-1539*. Mexico City: Editorial Jus, Academia Mexicana de Genealogía y Heráldica, A. C.

——— (1976). "Patterns of Spanish Emigration to the Indies 1579-1600". *The Americas* 33:1, pp. 78–95.

Bycroft, Clare, Ceres Fernandez-Rozadilla, Clara Ruiz-Ponte, Inés Quintela, Ángel Carracedo, Peter Donnelly, and Simon Myers (2019). "Patterns of genetic differentiation and the footprints of historical migrations in the Iberian Peninsula". *Nature Communications* 10:1, p. 551.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller (2011). "Robust Inference With Multiway Clustering". *Journal of Business & Economic Statistics* 29:2, pp. 238–249.

Carreras Monfort, Cèsar and Pau de Soto Cañamares (2010). *Historia de La Movilidad En La Península Ibérica: Redes de Transporte En SIG*. Barcelona: Editorial UOC.

Cervantes Saavedra, Miguel (1615). *Don Quijote de La Mancha*. Madrid: Editorial Verbum.

Chaunu, Huguette and Pierre Chaunu (1955). *Séville et l'Atlantique (1504-1650): Première Partie, Partie Statistique*. Paris: A. Colin.

De Cristóforis, Nadia Andrea (2011). "El Análisis de La Temprana Emigración Gallega a Buenos Aires". *Boletín Americanista* 61:62, pp. 177–200.

Delgado Ribas, Josep María (1982). "La Emigración Española a América Latina Durante La Época Del Comercio Libre (1765–1820). El Ejemplo Catalán". *Boletín Americanista* 24:32, pp. 115–137.

Donaldson, Dave and Richard Hornbeck (2016). "Railroads and American Economic Growth: A "Market Access" Approach". *The Quarterly Journal of Economics* 131:2, pp. 799–858.

Durán López, Gonzalo (1991). "La Emigración Española a Ultramar, 1492-1914". In. Antonio Elias Roel (ed.). Madrid: Fundación Española de Historia Moderna.

Eiras Roel, Antonio and Agustín Guimerá Ravina, eds. (1991). *La emigración española a ultramar, 1492–1914*. Colección Actas (Grupo Tabacalera) 1. Actas de la Reunión Científica de la Asociación Española de Historia Moderna. Madrid: Tabapress : Grupo Tabacalera.

Elliott, John Huxtable (1963). *Imperial Spain, 1469–1716*. London: Edward Arnold.

Esparza, José Javier (2015). *La Cruzada Del Océano: La Gran Aventura de La Conquista de América*. Madrid: La esfera de los libros.

Espín-Sánchez, José-Antonio, Salvador Gil-Guirado, W. Daniel Giraldo-Paez, and Chris Vickers (2019). "Labor Income Inequality in Pre-Industrial Mediterranean Spain: The City of Murcia in the 18th Century". *Explorations in Economic History* 73, p. 101274.

Espín-Sánchez, José-Antonio, Salvador Gil-Guirado, and Chris Vickers (2022). "La "Doña" è Mobile: The Role of Women in Social Mobility in a Pre-Modern Economy". *The Journal of Economic History* 82:1, pp. 1–41.

Esteban-Oliver, Guillermo, Eduard J. Álvarez-Palau, and Jordi Martí-Henneberg (2024). "Infraestructuras para la navegación en la España moderna y liberal: puertos, faros, canales y ríos navegables". *Investigaciones de Historia Económica*, pp. 18–25.

Fafchamps, Marcel and Flore Gubert (2007). "The Formation of Risk Sharing Networks". *Journal of Development Economics* 83:2, pp. 326–350.

Fernández, Rafael Diego and Rafael Diego-Fernández Sotelo (1987). *Capitulaciones Colombinas (1492-1506)*. Zamora, Mich.: El Colegio de Michoacán.

García Hidalgo, Palmira (2019). "La Emigración Española a América En La Época Moderna. Un Acercamiento al Estado de La Cuestión". *Naveg@merica* 23.

——— (2021). "La Emigración Española a América En El Siglo XVII. Mujeres Cruzando El Atlántico". PhD thesis. Huelva: Universidad de Huelva.

Gil, Juan (1985). "El rol del Tercer Viaje Colombino". *Suplemento de Anuario de Estudios Americanos, Sección Historiografía y Bibliografía* 29:1, pp. 83–110.

Gomellini, Matteo and Cormac Ó Gráda (2013). "Migrations". In *The Oxford Handbook of the Italian Economy since Unification*. Gianni Toniolo (ed.). Oxford: Oxford University Press. Chap. 10, pp. 271–302.

González Ochoa, José Ma (2003). *Quién Es Quién En La América Del Descubrimiento: 1492-1600*. Madrid: Acento Editorial.

Gould, Alice Bache (1984). *Nueva Lista Documentada de Los Tripulantes de Colón En 1492*. Madrid: Real Academia de la Historia.

Hamilton, Earl Jefferson (1934). *American Treasure and the Price Revolution in Spain, 1501-1650*. Vol. 1125. Cambridge, MA: Harvard University Press.

Haring, Clarence Henry (1918). *Trade and Navigation between Spain and the Indies in the Time of the Hapsburgs*. Vol. 19. Harvard Economic Studies. Cambridge, MA: Harvard University Press.

Hassig, Ross (2006). *Mexico and the Spanish conquest*. University of Oklahoma Press.

Hatton, Timothy J. and Jeffrey G. Williamson (1998). *The Age of Mass Migration: Causes and Economic Impact*. New York, NY: Oxford University Press.

Hennessy, Alistair (1993). "The Nature of the Conquest and the Conquistadors". In *Proceedings of the British Academy*. The Meeting of Two Worlds: Europe and the Americas 1492-1650. Warwick Bray (ed.). Vol. 81. Oxford University Press, pp. 5–36.

Hillgarth, Jocelyn N. (1978). *The Spanish Kingdoms: 1250–1516. Volume II: 1410–1516, Castilian Hegemony*. Oxford: Clarendon Press.

Huwise (2023). *Georef Spain Municipio*. Shapefiles of Spanish municipal boundaries.

Instituto Hispano-Cubano de Historia de América (1930–2024). *Catálogo de los Fondos Americanos del Archivo de Protocolos de Sevilla, Siglo XVI*. 16 vols. Tomos I–II (1930), III (1932), IV (1935), V (1937), VI (1986), VII (1990), VIII (2000), IX (2004), X (2006), XI (2009), XII (2011), XIII (2014), XIV (2017), XV (2020), XVI (2024). Tomos I–III originally published by Compañía Ibero-Americana de Publicaciones, Madrid. Sevilla: Instituto Hispano-Cubano de Historia de América.

Instituto Nacional de Estadística (2008). *Censo de Pecheros, Carlos I, 1528*. Vol. I-II. Madrid: Instituto Nacional de Estadística.

Jacobs, Auke Pieter (1995). *Los Movimientos Migratorios Entre Castilla e Hispanoamérica Durante El Reinado de Felipe III, 1598-1621*. Studia Ibero-Americana. Amsterdam and Atlanta, GA: Rodopi.

Jarvis, A., H. I. Reuter, A. Nelson, and E. Guevara (2008). *Hole-filled seamless SRTM data V4*.

Kingdon, Jack (2025). "3 Months Transit Time to Mars for Human Missions Using SpaceX Starship". *Scientific Reports* 15, p. 17764.

Kirkpatrick, Frederick A. (1934). *The Spanish Conquistadores*. Pioneer Histories. London: A. & C. Black.

Konetzke, Richard (1945). "Legislación Sobre Inmigración de Extranjeros En América Durante La Época Colonial". *Revista internacional de Sociología* 3:11, p. 269.

Ladero Quesada, Miguel Ángel (1967). *Castilla y la conquista del reino de Granada*. Valladolid: Universidad de Valladolid.

——— (2005). *La Hermandad de Castilla. Cuentas y Memoriales. 1480–1498*. Madrid: Real Academia de la Historia.

León Guerrero, María Montserrat et al. (2007). "Pasajeros del segundo viaje de Cristóbal Colón". *Revista de estudios colombinos* 3, pp. 29–60.

Lockhart, James (1968). *Spanish Peru 1532-1560: A Colonial Society*. Madison, WI: The University of Wisconsin Press.

——— (2013). *The men of Cajamarca: a social and biographical study of the first conquerors of Peru*. University of Texas Press.

Macías Domínguez, Isabelo (1999). *La Llamada Del Nuevo Mundo: La Emigración Española a América (1701-1750)*. Historia y Geografía 46. Sevilla: Universidad de Sevilla.

MacLeod, Murdo J (2010). *Spanish Central America: a socioeconomic history, 1520–1720*. University of Texas Press.

Márquez Macías, Rosario (1995). *La Emigración Española a América, 1765-1824*. Oviedo: Universidad de Oviedo.

Martínez, José Luis (1983). *Pasajeros de Indias: viajes transatlánticos en el siglo XVI*. Alianza universidad 355. Madrid: Alianza.

Martínez Shaw, Carlos (1980). "Cataluña y el comercio con América. El fin de un debate". *Boletín Americanista* 30, pp. 223–236.

——— (1994). *La emigración española a América, 1492-1824*. 1st ed. Vol. 11. Cruzar el charco. Colombres, Asturias: Archivo de Indianos.

Maurer, Noel and Carlos Yu (2010). *The Big Ditch: How America Took, Built, Ran, and Ultimately Gave Away the Panama Canal*. Course Book. Princeton, NJ: Princeton University Press.

Menéndez Pidal, Ramón (1906). *La España Del Cid*. Madrid: Tipografía de la Revista de Archivos.

Mira Caballos, Esteban (2000). *Nicolás de Ovando y los orígenes del sistema colonial español, 1502-1509*. Santo Domingo: Patronato de la Ciudad Colonial de Santo Domingo, Centro de Altos Estudios Humanísticos y del Idioma Español.

——— (2014). *La gran armada colonizadora de Nicolás de Ovando, 1501-1502*. Primera edición. Vol. 121. Academia Dominicana de la Historia. Santo Domingo: Academia Dominicana de la Historia.

Montana, Alberto (1943). *Descubrimientos, Exploraciones y Conquistas de Españoles y Portugueses En América y Oceanía*. Barcelona: Editorial Miguel A. Salvatella.

Mörner, Magnus (1975). "La Emigración Española al Nuevo Mundo Antes de 1810. Un Informe Del Estado de La Investigación". *Anuario de estudios americanos* 32, pp. 43–131.
——— (1991). "Migraciones a Hispanoamérica Durante La Época Colonial". *Suplemento de Anuario de Estudios Americanos* 48:2, pp. 3–26.
Moya, José C. (1998). *Cousins and Strangers: Spanish Immigrants in Buenos Aires, 1850–1930*. Berkeley: University of California Press.
Munshi, Kaivan (2003). "Networks in the Modern Economy: Mexican Migrants in the U.S. Labor Market". *Quarterly Journal of Economics* 118:2, pp. 549–599.
——— (2015). "Community Networks and Migration". In *The Oxford Handbook on the Economics of Networks*. Yann Bramoullé, Andrea Galeotti, and Brian W. Rogers (ed.). Oxford: Oxford University Press.
——— (2020). "Social Networks and Migration". *Annual Review of Economics* 12:1, pp. 503–524.
Munshi, Kaivan and Mark Rosenzweig (2016). "Networks and Misallocation: Insurance, Migration, and the Rural-Urban Wage Gap". *American Economic Review* 106:1, VERIFY.
Nickell, Stephen (1981). "Biases in Dynamic Models with Fixed Effects". *Econometrica* 49:6, pp. 1417–1426.
Núñez, Francisco Javier Gutiérrez (2002). "La Vinculación Americana de Fuente de Cantos: Pasajeros a Indias (Siglos XVI-XVII)". In *II Jornada de Historia de Fuente Cantos: Actas*. Asociación Cultural Lucerna, pp. 53–110.
Obolensky, Marguerite, Marco Tabellini, and Charles A Taylor (2024). *Homeward Bound: How Migrants Seek Out Familiar Climates*. Working Paper 24–031.
Odriozola Oyarbide, Lourdes (1998). "La construcción naval en Gipuzkoa. Siglos XVI-XVIII". *Itsas Memoria. Revista de Estudios Marítimos del País Vasco* 2, pp. 93–146.
Oropeza Chávez, Ana Brisa (2016). "Regulación y Práctica de La Extranjería En El Derecho Indiano: De Las Partidas a La Recopilación de 1860". *Anuario de Estudios Americanos* 75:1, pp. 223–57.
Ortiz Arza, Javier (2019). "Introducción al estudio de los navegantes vascos de la Carrera de Indias y su participación en la trata atlántica de esclavos del siglo XVI". *Brocar: Cuadernos de investigación histórica* 43, pp. 41–72.
Otte, Enrique (1993). *Cartas privadas de emigrantes a Indias, 1540-1616*. México: Fondo de Cultura Económica.
Penny, Ralph (1990). *La Lengua de Juan Del Encina*. Oxford: Oxford University Press.
——— (2002). *A History of the Spanish Language*. Cambridge: Cambridge University Press.
Pérez-Artés, María del Carmen (2024). "Numeracy Selectivity of Spanish Migrants in Colonial America (Sixteenth–Eighteenth Centuries)". *The Economic History Review* 77:2, pp. 503–522.
Pérez-Mallaína, Pablo Emilio (1998). *Spain's Men of the Sea: Daily Life on the Indies Fleets in the Sixteenth Century*. Trans. by Carla Rahn Phillips. Baltimore: Johns Hopkins University Press.
Pérez-Mallaína Bueno, Pablo Emilio (1998). *Spain's Men of the Sea: Daily Life on the Indies Fleets in the Sixteenth Century*. Baltimore, MD: Johns Hopkins University Press.
Quesada, Miguel Ángel Ladero (2001). *La Guerra de Granada, 1482–1491*. Granada, Spain: Diputación Provincial de Granada.
Richman, Irving Berdine (1921). *The Spanish Conquerors: A Chronicle of the Dawn of Empire Overseas*. Textbook ed. Chronicles of America Series ; v. 2. New Haven: Yale University Press.
Rodríguez Galdo, María Xosé (1993). *Galicia, País de Emigración. La Emigración Gallega a América Hasta 1930*. Colombres: Fundación Archivo de Indianos.
Rodríguez Lorenzo, Sergio M (2017). "El Contrato de Pasaje En La Carrera de Indias (1561-1622)". *Historia mexicana* 66:3, pp. 1479–1571.
Sánchez Alonso, Blanca (1995). *Las Causas de La Emigración Española, 1880–1930*. Madrid: Alianza.
Sánchez Rubio, Rocío (1993). *La Emigración Extremeña al Nuevo Mundo: Exclusiones Voluntarias y Forzosas de Un Pueblo Periférico En El Siglo XVI*. Badajoz: Siruela.
Sánchez-Alonso, Blanca (2000). "Those Who Left and Those Who Stayed Behind: Explaining Emigration from the Regions of Spain, 1880–1914". *The Journal of Economic History* 60:03, pp. 730–755.
——— (2019). "The Age of Mass Migration in Latin America". *The Economic History Review* 72:1, pp. 3–31.
Schafer, Ernesto (1946). "Comunicaciones Marítimas y Terrestres de Las Indias Españolas". *Anuario de Estudios Americanos* 3, pp. 969–983.

Slicher van Bath, B. H. (1986). "The Absence of White Contract Labour in Spanish America During the Colonial Period". In *Colonialism and Migration: Indentured Labour Before and After Slavery*. P. C. Emmer (ed.). Dordrecht: Springer Netherlands, pp. 19–31.

Sonesson, Birgit (1995). *Catalanes En Las Antillas. Un Estudio de Caso*. Colombres: Fundación Archivo de Indianos.

Spitzer, Yannay (2021). *Pogroms, Networks, and Migration: The Jewish Migration from the Russian Empire to the United States 1881–1914*. Pre-published.

Spitzer, Yannay, Gaspare Tortorici, and Ariell Zimran (2025). "International Migration Responses to Modern Europe's Most Destructive Earthquake: Messina and Reggio Calabria, 1908". *The Journal of Economic History* 85:3, pp. 617–663.

Spitzer, Yannay and Ariell Zimran (2023). *Like an Ink Blot on Paper: Testing the Diffusion Hypothesis of Mass Migration, Italy 1876-1920*. Working Paper 30847. National Bureau of Economic Research.

Stangl, Werner (2010). "Consideraciones Metodológicas Acerca de Las Cartas Privadas de Emigrantes Españoles Desde América, 1492–1824. El Caso de Las "Cartas de Llamada"". *Jahrbuch für Geschichte Lateinamerikas* 47:1, pp. 11–36.

Szászdi León-Borja, István (2019). "La Instrucción Virreinal fechada en la Isla Española el 9 de abril de 1494 y el poder otorgado por Cristóbal Colón a Mosén Pedro de Margarite". *Revista Cruz del Sur* 9:31, pp. 11–35.

Triano Milán, José Manuel (2018a). "¿ Un nuevo sistema de fiscalidad extraordinaria? La Santa Hermandad de los Reyes Católicos (1476-1498)". *Historia Medieval* 36, p. 171.

——— (2018b). *La llamada del rey y el auxilio del reino. Del pedido regio a las contribuciones de la Santa Hermandad (1406-1498)*. Universidad de Sevilla.

Vázquez de Prada Vallejo, Valentín (1991). "Algunas Consideraciones En Torno a La Emigración de Navarros al Nuevo Mundo". *Príncipe de Viana. Anejo* 13, pp. 31–40.

Verlinden, Charles (1970). *The Beginnings of Modern Colonization: Eleven Essays with an Introduction*. Ithaca N. Y: Cornell University Press.

Wegge, Simone A. (1998). "Chain Migration and Information Networks: Evidence from Nineteenth-Century Hesse-Cassel". *Journal of Economic History* 58:4, pp. 957–986.

Zavala, Silvio (1935). *La Encomienda Indiana*. Madrid: Centro de Estudios Históricos.

Zimran, Ariell (2020). "Transportation and Health in the Antebellum United States, 1820–1847". *The Journal of Economic History* 80:3, pp. 670–709.

Tables

Table 1: Descriptive statistics: municipalities

		Kingdoms				Provinces	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Variables	Spain	Castilla	Granada	Aragon	Navarra	Sevilla	Badajoz
GEOGRAPHY							
Elevation (km.)	0.711	0.762	0.853	0.584	0.588	0.150	0.413
	(0.367)	(0.324)	(0.416)	(0.433)	(0.205)	(0.155)	(0.124)
Ruggedness (SD)	−0.001	−0.152	1.354	0.126	0.183	−0.473	−0.430
	(0.997)	(0.861)	(1.566)	(1.041)	(0.728)	(0.429)	(0.327)
Wheat suitability (SD)	0.001	0.055	−0.847	0.091	−0.630	−0.315	−0.057
	(1.000)	(0.977)	(1.032)	(0.979)	(0.845)	(0.120)	(0.134)
Distance to Seville (1,000s km.)	0.518	0.449	0.227	0.703	0.692	0.036	0.152
	(0.195)	(0.159)	(0.071)	(0.127)	(0.031)	(0.022)	(0.037)
Travel cost to Seville, roads (SD)	−0.003	−0.446	0.236	0.994	0.340	−2.149	−1.525
	(0.999)	(0.785)	(1.029)	(0.744)	(0.304)	(0.256)	(0.243)
Travel cost to Seville, roads & water (SD)	−0.002	−0.196	0.440	0.241	1.080	−2.668	−1.624
	(0.999)	(0.848)	(1.429)	(1.130)	(0.436)	(0.423)	(0.404)
Navigable river	0.075	0.112	0.041	0.000	0.000	0.932	0.752
	(0.263)	(0.316)	(0.199)	(0.000)	(0.000)	(0.253)	(0.434)
Coast	0.233	0.124	0.674	0.436	0.171	0.081	0.000
	(0.423)	(0.329)	(0.470)	(0.496)	(0.377)	(0.275)	(0.000)
DEMOGRAPHICS							
Population (1,000s)	0.740	0.871	0.487	0.503	0.463	2.459	1.372
	(1.847)	(1.858)	(1.611)	(1.934)	(0.724)	(6.143)	(1.387)
Urban	0.014	0.017	0.011	0.008	0.008	0.081	0.021
	(0.119)	(0.131)	(0.106)	(0.090)	(0.090)	(0.275)	(0.145)
Share nobility (per 1,000)	36.225	36.288	5.203	21.079		7.732	8.636
	(128.044)	(128.198)	(8.007)	(18.354)		(19.962)	(21.895)
Share clergy (per 1,000)	2.074	2.069	3.142	5.278		2.805	1.590
	(4.942)	(4.945)	(2.450)	(3.845)		(2.440)	(2.759)
Assigned taxes (p.c. marv., 1528)	9.500	9.456	24.826	30.607		25.907	8.361
	(14.049)	(14.037)	(3.651)	(2.035)		(7.403)	(11.281)
Paid taxes (p.c. marv., 1528)	9.881	9.853	12.940	32.712		24.585	8.304
	(16.351)	(16.345)	(15.384)	(2.749)		(14.569)	(12.208)
Assigned taxes (p.c. marv., 1528)	9.500	9.456	24.826	30.607		25.907	8.361
	(14.049)	(14.037)	(3.651)	(2.035)		(7.403)	(11.281)
Paid taxes (p.c. marv., 1528, excl. 0)	9.881	9.853	12.940	32.712		24.585	8.304
	(16.351)	(16.345)	(15.384)	(2.749)		(14.569)	(12.208)
Assigned taxes (p.c. marv., 1528, excl. 0)	9.500	9.456	24.826	30.607		25.907	8.361
	(14.049)	(14.037)	(3.651)	(2.035)		(7.403)	(11.281)
MIGRATION (annual per 100,000)							
Migration 1492–1539	3.028	4.350	1.340	0.451	0.663	12.815	14.503
	(12.124)	(14.410)	(6.922)	(4.911)	(3.189)	(25.207)	(22.687)
Migration 1492–1512	0.905	1.266	0.654	0.201	0.039	3.787	3.111
	(7.985)	(8.656)	(7.446)	(6.813)	(0.547)	(9.260)	(6.266)
Migration 1513–1539	4.513	6.508	1.807	0.623	1.108	19.128	22.529
	(18.155)	(21.688)	(9.157)	(6.699)	(5.448)	(37.934)	(35.938)
Explorers	0.014	0.020	0.009	0.000	0.000	0.077	0.090
	(0.224)	(0.276)	(0.112)	(0.014)	(0.000)	(0.471)	(0.466)
Crew	0.600	0.880	0.201	0.059	0.095	1.313	2.091
	(4.051)	(4.967)	(1.402)	(0.756)	(0.527)	(2.482)	(4.674)
High-status	0.541	0.797	0.177	0.052	0.062	1.418	3.026
	(3.371)	(4.117)	(1.351)	(0.721)	(0.734)	(4.366)	(5.369)
Observations	6,388	4,148	267	1,727	246	74	141

Table 2: Cross section determinants, municipalities, binomial

Variables	(1)	(2)	(3)	(4)	(5)
Travel cost (roads+water)	-0.396^{a}	-0.503^{b}	-0.272	-0.747^{a}	-0.329^{b}
	(0.149)	(0.234)	(0.233)	(0.279)	(0.128)
Population (log.)	0.458^{a}	0.598^{a}	0.671^{a}	0.474^{a}	0.585^{a}
	(0.070)	(0.053)	(0.061)	(0.076)	(0.070)
Elevation	0.021	-0.158	-0.235	-0.059	-0.552^{c}
	(0.120)	(0.151)	(0.182)	(0.342)	(0.124)
Soil Quality	0.113	-0.052	-0.073	0.167	0.000
	(0.080)	(0.083)	(0.089)	(0.128)	(0.078)
Near River	-0.147	-0.043	-0.253	-0.412	-0.117
	(0.188)	(0.259)	(0.415)	(0.406)	(0.172)
Near Coast	-0.101	0.247	-0.038	0.269	-0.198
	(0.226)	(0.348)	(0.305)	(0.270)	(0.210)
Granada	-0.347			-1.228^{a}	-0.238
	(0.212)			(0.379)	(0.210)
Aragon	-2.357^{a}			-2.829^{a}	-2.267^{a}
	(0.305)			(0.463)	(0.280)
Navarra	-0.605^{c}			-1.662^{b}	-0.605^{b}
	(0.311)			(0.800)	(0.297)
Sevilla	0.645^{c}			0.124	0.745^{b}
	(0.344)			(0.375)	(0.342)
Badajoz	0.740^{a}			-0.007	0.863^{a}
	(0.223)			(0.319)	(0.212)
Observations	6,236	6,236	6,236	6,236	6,236
Pseudo R-squared	0.501	0.581	0.689		
FE	None	Prov	Dist	None	None
Mean rate mig.	0.030	0.030	0.030		
Period	Full	Full	Full	Before 1513	After 1513

Significance levels: ^{a} p<0.01, ^{b} p<0.05, ^{c} p<0.1

Notes: Dependent variable is the emigration rate to any destination. Standard errors clustered by district in parentheses. Columns 1–3 includes 1492–1539. Column 4 is before 1513 and column 5 is after 1513. Columns 1–5 include jurisdiction fixed effects. Unit of observation is a municipality. The sample includes all of Spain. Municipalities with population under 20 are omitted. Pseudo R-squared are McFadden's, with the baseline log-likelihood being that from regression on a constant over the regression's sample.

Table 3: Cross section determinants, binomial, municipalities, after 1513

Variables	(1)	(2)	(3)	(4)	(5)	(6)
Mig. (log p.c.) × mig. > 0 before 1513		0.819^a		0.715^a		0.632^a
		(0.087)		(0.054)		(0.072)
Mig. = 0 before 1513	−2.154^a	−5.048^a	−1.500^a	−4.412^a	−0.563^a	−3.482^a
	(0.304)	(0.379)	(0.236)	(0.289)	(0.194)	(0.373)
Observations	389	389	389	389	389	389
Pseudo R-squared	0.176	0.762	0.610	0.814	0.755	0.860
Controls			✓	✓	✓	✓
FE					Prov	Prov
Mean rate mig.	0.081	0.081	0.081	0.081	0.081	0.081

Significance levels: ^{a} p<0.01, ^{b} p<0.05, ^{c} p<0.1
Notes: This table presents the results of a binomial regression for migration to any destination since 1513. Unit of observation is a district. Dependent variable is the total number migrants to any destination. The explantory variable of interest is log. per capita migration before 1513, interacted with an indicator for at least one migrant prior to 1513; when migration before 1513 is zero, its value is zero. The coefficient on this variable could be interpreted as elasticity. Control variables are travel costs (water+land), share of urban population, elevation, soil quality, proximity to river and to coast, share of urban population, and indicators for Sevilla and Badajoz. Standard errors clustered by province in parentheses. Pseudo R-squared are McFadden's, with the baseline log-likelihood being that from regression on a constant over the regression's sample. The samples in columns 3 and 6 inlcude only Castilla.

Table 4: Networks regressions: district, lag

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Mig. lag (log. p.c.) × mig. lag > 0	0.823^a	0.569^a	0.409^a	0.034	0.822^a	0.569^a	0.409^a	0.034				
	(0.028)	(0.048)	(0.035)	(0.030)	(0.028)	(0.048)	(0.035)	(0.030)				
Mig. lag = 0	−4.927^a	−3.441^a	−2.427^a	−0.238	−4.924^a	−3.439^a	−2.425^a	−0.238				
	(0.217)	(0.277)	(0.180)	(0.158)	(0.217)	(0.277)	(0.180)	(0.158)				
Mig. lag									0.830^a	0.786^a	0.637^a	0.343^c
									(0.104)	(0.103)	(0.120)	(0.197)
Observations	3,790	3,760	3,760	3,790	3,790	3,760	3,760	3,790	3,790	3,760	3,760	3,790
R-squared									0.263	0.274	0.308	0.376
Pseudo R-squared	0.718	0.750	0.781	0.836	0.718	0.749	0.781	0.836				
Controls		✓	✓			✓	✓			✓	✓	
FE			Prov	Dist			Prov	Dist			Prov	Dist
Model	Binomial	Binomial	Binomial	Binomial	Poisson	Poisson	Poisson	Poisson	OLS	OLS	OLS	OLS
Mean rate mig.	0.035	0.035	0.035	0.035	0.035	0.035	0.035	0.035	0.035	0.035	0.035	0.035
Marginal effect ×10k	1.122	0.779	0.560	0.046	1.123	0.779	0.560	0.046				

Significance levels: ^{a} p<0.01, ^{b} p<0.05, ^{c} p<0.1
Notes: This table presents the results of regression for migration to any destination. Unit of observation is a district-lustro. Districts with population under 500 are omitted. Dependent variable in the binomial and Poisson regressions is the number of migrants to any destination. Dependent variable in the OLS regressions is the annual rate of migration to any destination per 10,000, and its mean is reported). Controls include standardized travel distance to Sevilla, share of urban population, elevation, ruggedness, and indicators for proximity to the coast and to navigable rivers. Migrant lag is defined as the annual rate of migration to any destination per 10,000 in the previous period. All specifications include period-fixed effects. Coefficients of binomial regressions can be interpreted as in a logit regression, as the effect on the log-odds ratio. Since the probability of migration is close to zero, this is close to semi-elasticities. Coefficients of Poisson regressions can be interpreted as semi-elasticities. Pseudo R-squared values for the binomial and Poisson regressions are McFadden's. Standard errors clustered by province in parentheses.

Table 5: Destination choice by network size

	(1)	(2)	(3)	(4)	(5)	(6)
Over-rep. of network (log)	0.242[a]	0.215[a]	0.168[a]	0.144[a]	0.150[a]	0.140[b]
	(0.032)	(0.036)	(0.041)	(0.051)	(0.055)	(0.059)
Interacting year var.	Trend	Trend	Trend	Trend	FE	FE
First year	1493	1530	1493	1530	1493	1530
Destinations	6	6	4	4	4	4
Passengers	17,260	9,459	16,429	8,689	16,429	8,689
Observations	82,549	50,398	57,307	34,756	57,307	34,756
Obs. with network > 0	55,121	38,286	46,043	30,305	46,043	30,305
Pseudo-R^2	0.006	0.005	0.002	0.002	0.002	0.002

Significance levels: [a] p<0.01, [b] p<0.05, [c] p<0.1

Notes: This table presents the results of conditional logit regressions, where the choice of the destination is a function of the over-representation of the district's network in the destination. Unit of observation is passenger-destination. Dependent variable is an indicator for passenger i choosing destination region d. The origin-destination-specific network is defined in terms of over-representation over the 5 years prior to the passenger's migration. For example, if half of the migrants from district of origin o travelled to destination region d over the years $[t-5, t-1]$, but only a third of all Spanish migrants did so, then the over-representation of the network of origin o in destination d in year t is 1.5. The main treatment variable is the log of the over-representation of the network when the network size is positive, and zero if its size is zero. The regressions control for an indicator for zero networks (not reported). Destination fixed effects are interacted with a time trend or with year fixed effects. The sample includes either the full period (1493–1539) or the final decade (1530–1539), when all destinations were established. The number of destinations is either 4, which includes only the primary destinations (Caribbean/Indies, Mexico, Tierra Firme, and Peru), or 6, which also includes secondary destinations (La Plata and Florida). Regressions with 4 destinations and destination-specific year fixed effects did not converge and are not reported. The table reports the number of observations with positive network values. The coefficient estimates capture the intensive margin effect of network strength among migrants with non-zero networks. Passengers who are known to be slaves, are coming from unidentified districts, or are coming from districts with population under 500, are omitted. Pseudo R-squared values are McFadden's, relative to the same regression without network variables. Standard errors clustered by district in parentheses.

Table 6: Voyage and Master Matching Using Notarial Contracts

Panel (a): Probability of Hiring the Same Master

	(1)	(2)	(3)	(4)
Same Surname	0.0295^a	0.0282^a	0.0289^a	0.0282^a
	(0.0029)	(0.0028)	(0.0029)	(0.0028)
Same Notary		0.0263^a		0.0260^a
		(0.0025)		(0.0025)
Same Office			0.0170^a	0.0003
			(0.0016)	(0.0007)
Year FE	✓	✓	✓	✓
Mean dep. var.	0.0051	0.0051	0.0051	0.0051
R^2	0.0048	0.0190	0.0133	0.0190
Observations	1,462,905	1,462,905	1,462,905	1,462,905
Migrants	1,711	1,711	1,711	1,711

Panel (b): Probability of Being on the Same Voyage

	(1)	(2)	(3)	(4)
Same Surname	0.0135^{***}	0.0125^{***}	0.0130^{***}	0.0125^{***}
	(0.0016)	(0.0015)	(0.0015)	(0.0015)
Same Notary		0.0220^{***}		0.0209^{***}
		(0.0021)		(0.0020)
Same Office			0.0146^{***}	0.0012^{***}
			(0.0014)	(0.0004)
Year FE	Yes	Yes	Yes	Yes
Mean dep. var.	0.0034	0.0034	0.0034	0.0034
R^2	0.0048	0.0198	0.0142	0.0198
Observations	1,462,905	1,462,905	1,462,905	1,462,905
Migrants	1,711	1,711	1,711	1,711

Notes: Dyadic regressions using notarial protocols. Panel (a) reports the probability that two migrants hire the same ship master. Panel (b) reports the probability that two migrants share the same voyage (same ship master within ±1 year).The unit of observation is a pair of migrants. All specifications include year fixed effects. Standard errors are multi-way clustered at the individual level (Cameron et al. 2011). $^*p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Table 7: Voyage and Master Matching Using Ships Registers

Panel (a): Probability of Hiring the Same Master

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Same Munic.		0.0011^{**}	0.0034^{***}			0.0016^{***}				-0.0003
		(0.0005)	(0.0003)			(0.0005)				(0.0005)
Same Province		0.0007^{***}		0.0018^{***}	0.0005^{**}	0.0009^{***}				0.0013^{***}
		(0.0002)		(0.0002)	(0.0002)	(0.0002)				(0.0002)
Same town	0.0043^{***}	0.0029^{***}			0.0038^{***}	0.0026^{***}	0.0062^{***}	0.0047^{***}	0.0063^{***}	0.0053^{***}
	(0.0004)	(0.0006)			(0.0004)	(0.0005)	(0.0004)	(0.0004)	(0.0004)	(0.0006)
Year FE										
Mean dep. var.										
R^2										
Observations										
Migrants										

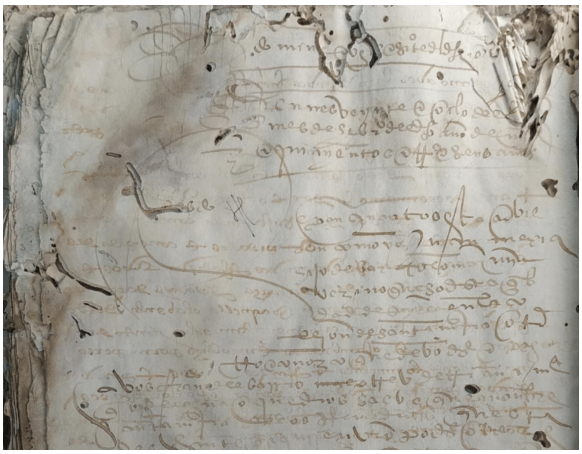
Panel (b): Probability of Being on the Same Voyage

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Same Munic.		0.0001	0.0030^{***}			0.0004				-0.0012^{***}
		(0.0004)	(0.0003)			(0.0003)				(0.0004)
Same Province		0.0008^{***}		0.0016^{***}	0.0004^{**}	0.0009^{***}				0.0013^{***}
		(0.0002)		(0.0002)	(0.0002)	(0.0002)				(0.0002)
Same town	0.0038^{***}	0.0034^{***}			0.0035^{***}	0.0034^{***}	0.0058^{***}	0.0042^{***}	0.0059^{***}	0.0058^{***}
	(0.0003)	(0.0004)			(0.0004)	(0.0004)	(0.0003)	(0.0003)	(0.0003)	(0.0004)
Year FE										
Mean dep. var.										
R^2										
Observations										
Migrants										

Notes: Dyadic regressions using Asientos. Panel (a) reports the probability that two migrants hire the same ship master. Panel (b) reports the probability that two migrants share the same voyage (same ship master within ±1 year).The unit of observation is a pair of migrants. All specifications include year fixed effects. Standard errors are multi-way clustered at the individual level (Cameron et al. 2011). $^{*}p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$.

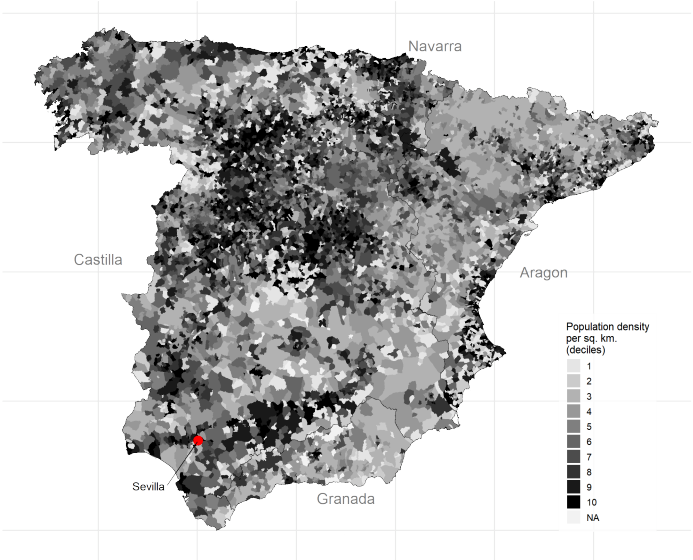
Figures

Figure 1: Contract of carriage



Notes: Juan Mejía to Puerto Rico in 1536: Originally from Villaverde del Rio, Mejía departed from the port of Sanlúcar de Barrameda to Puerto Rico in 1536. He promised to pay six ducats to the ship captain Francisco de Barrio for transportation at destination with a grace period of eight days after arrival. Sources: Catálogo de Pasajeros and AHPS SIGNATURA 9144, f. 193v 194r, 28/02/1536.

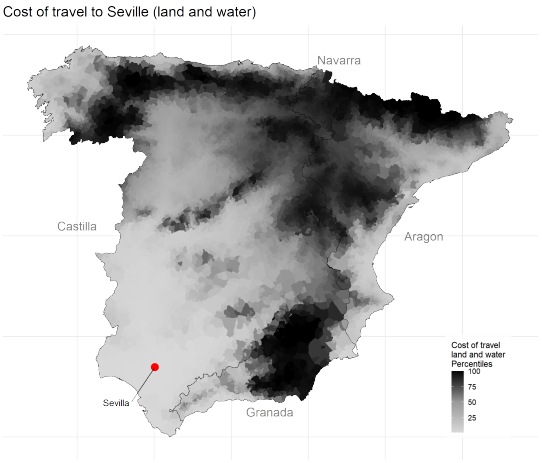
Figure 2: Population density



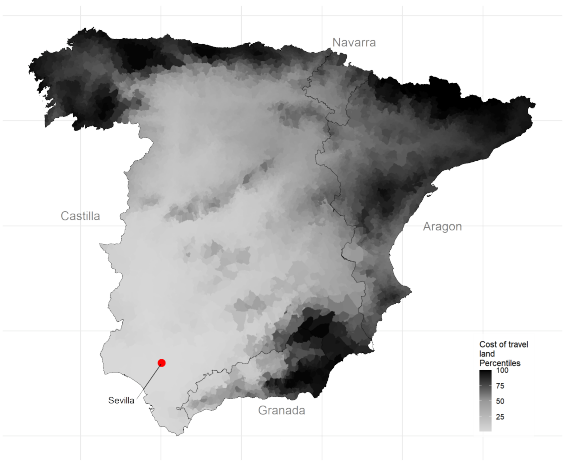
Notes: This map presents municipalities' population data, defined as the historical population assigned to the municipality divided by its modern area.

Figure 3: Travel cost to Seville

(a) Allowing water travel

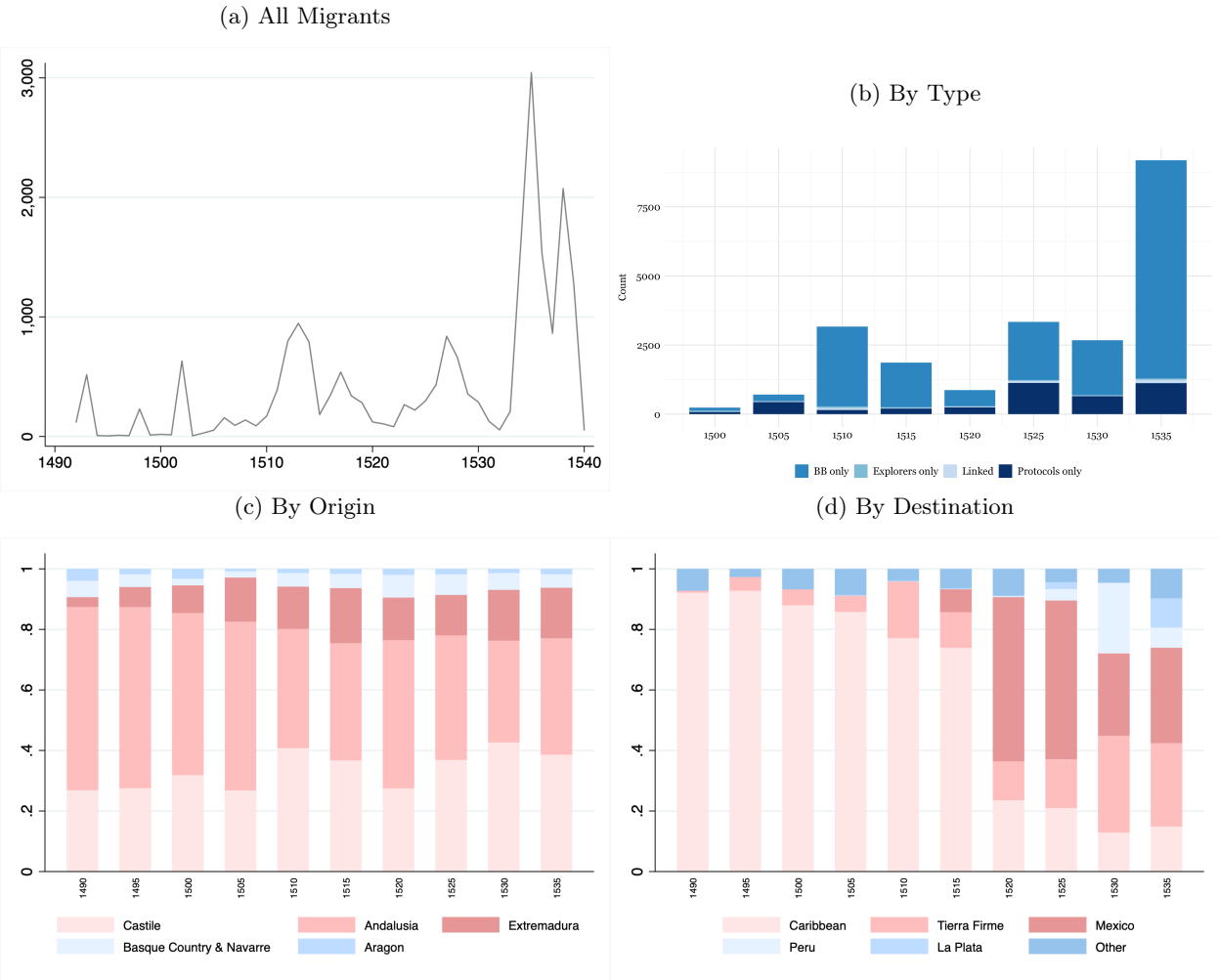


(b) Overland travel only



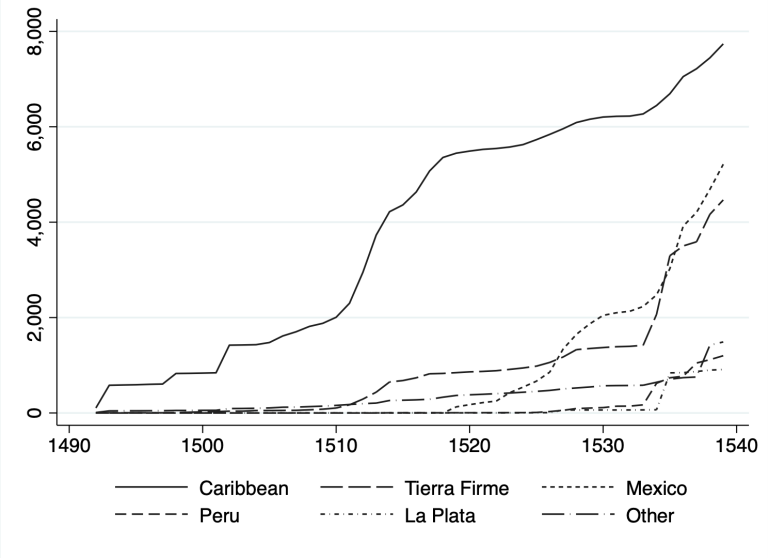
Notes: The maps report the percentile of the travel cost to Sevilla, while allowing and excluding water travel. Each unit is a municipality. See A.5 for details.

Figure 4: Migration counts



Notes: Panel (a) shows the total number of migrants by year. Panel (b) shows the number of migrants by lustrum and data source. The top blue color represents records from the Boyd-Bowman (BB) data only.“ Explorers only” represents the records that were added to the BB data through an exact match on identifiers. The lightest blue represents records that appear in both BB and Protocols (linked). Lastly, the darkest blue represents records that come exclusively from Protocols and that have no match in the BB data. Panel (c) shows the share of migrants by region of origin and lustrum. Panel (d) shows the share of migrants by lustrum and destination. For destination recorded as “Indias,” we reassigned migrants according to the shares of migrants with known origin.

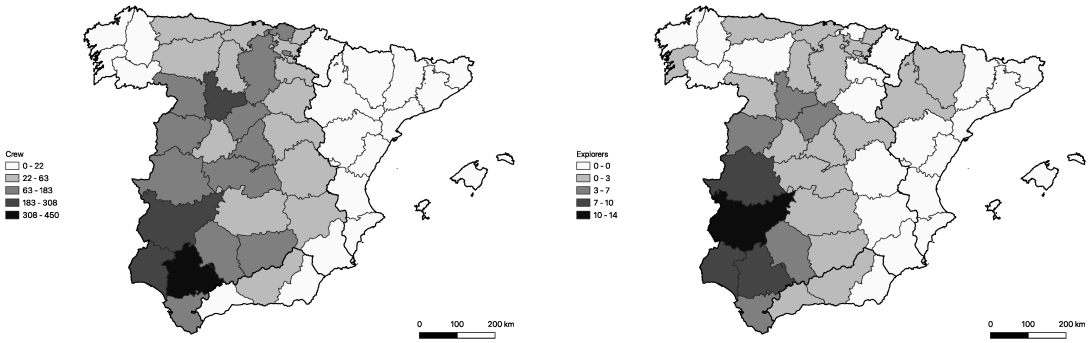
Figure 5: Cumulative migration by main destination, 1492-1540



Notes: This figure presents the cumulative number of migrants from all of our sources to each major destination group. For destination recorded as "Indias," we reassigned migrants according to the shares of migrants with known origin.

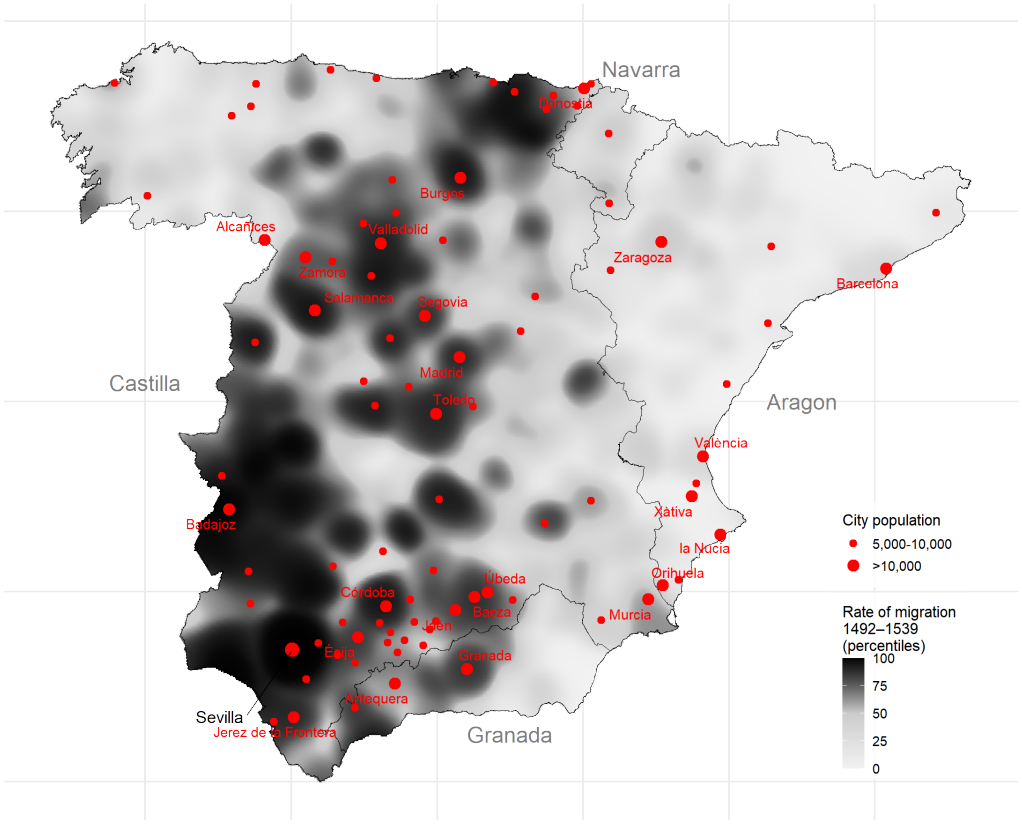
Figure 6: Crew and explorer counts by province

(a) Crew (b) Explorers



Notes: The figure presents a map of the number of crew members and explorers originating in each province, defining each group as discussed in the main text.

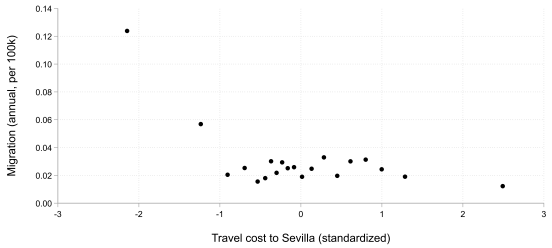
Figure 7: Emigration rate, 1492–1539, heat map



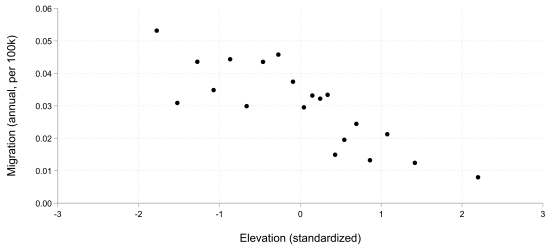
Notes: The figure presents a heat map of the geographic origins of Spanish migration, 1492–1540. Each cell's shade reflects the population-weighted average emigration rate in its vicinity, computed as $\bar{e}_i = \frac{\Sigma_j e_j p_j w_{ij}}{\Sigma_j p_j w_{ij}}$, where e_j and p_j are the emigration rate and population of municipality j, and $w_{ij} = \exp(-d_{ij}^2/2h^2)$ is a Gaussian weight based on distance d_{ij} between cell i and municipality centroid j, with bandwidth $h = 15$ km. For computational simplicity, only the 1,000 closest municipalities within four bandwidths are included; the effect of excluded municipalities is negligible. Darker shades indicate higher relative emigration rates. Cell sizes are 0.02×0.02 arc degrees, approximately 2×2 km. Larger population centers exert stronger influence on nearby cells and may produce darker shades if their emigration rates exceed those of neighboring areas.

Figure 8: Determinants of Emigration by Municipality

(a) Emigration and distance from Seville



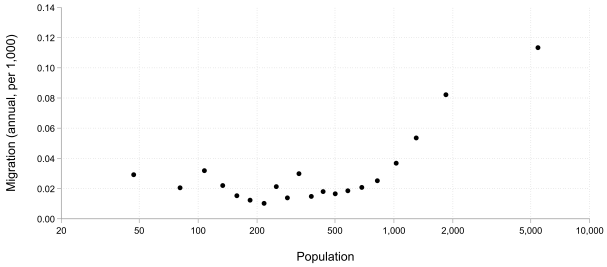
(b) Emigration and elevation



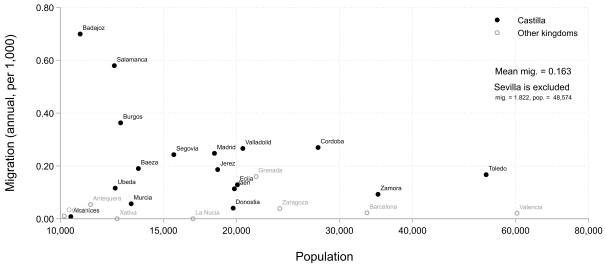
Notes: Panel (a): Average emigration rates of Spanish municipalities by travel cost to Seville, binned into 20 quantiles. Travel costs account for both land and water routes and are standardized. Each point represents the unweighted mean emigration rate within a bin. Panel (b): Emigration rates of Spanish municipalities by standardized elevation, binned into 20 quantiles. Each point represents the unweighted mean emigration rate within a bin.

Figure 9: Emigration and municipality population

(a) All municipalities



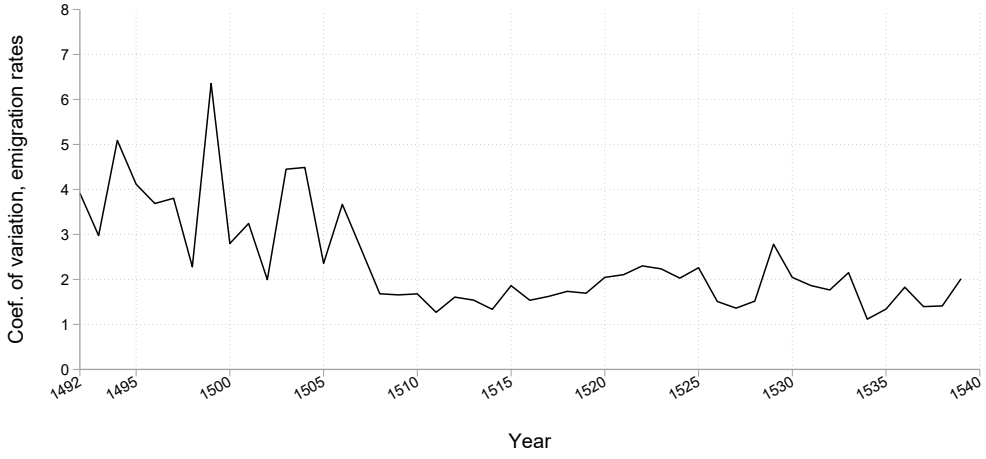
(b) Urban centers



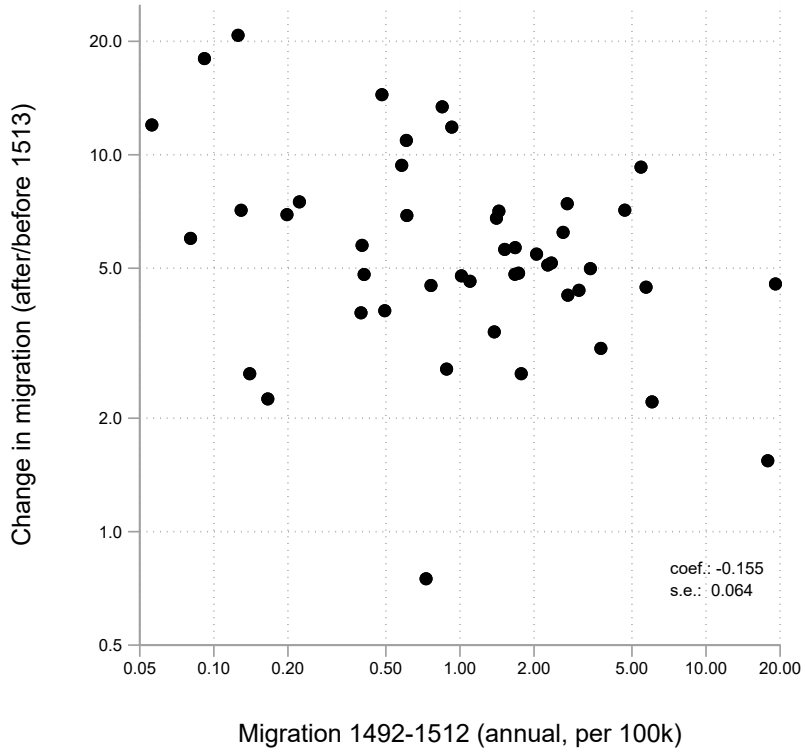
Notes: Emigration rates of Spanish municipalities (1492–1539) by population. Panel (a) shows all municipalities grouped into 20 quantiles (ventiles), with each point representing the unweighted mean emigration rate within a bin. Panel (b) shows the emigration rates of individual urban centers with populations over 10,000, excluding Seville. Cities within Castile and outside Castile are shown in different shades. Both plots use a logarithmic x-axis.

Figure 10: Convergence in emigration rates across provinces

(a) σ-convergence: annual



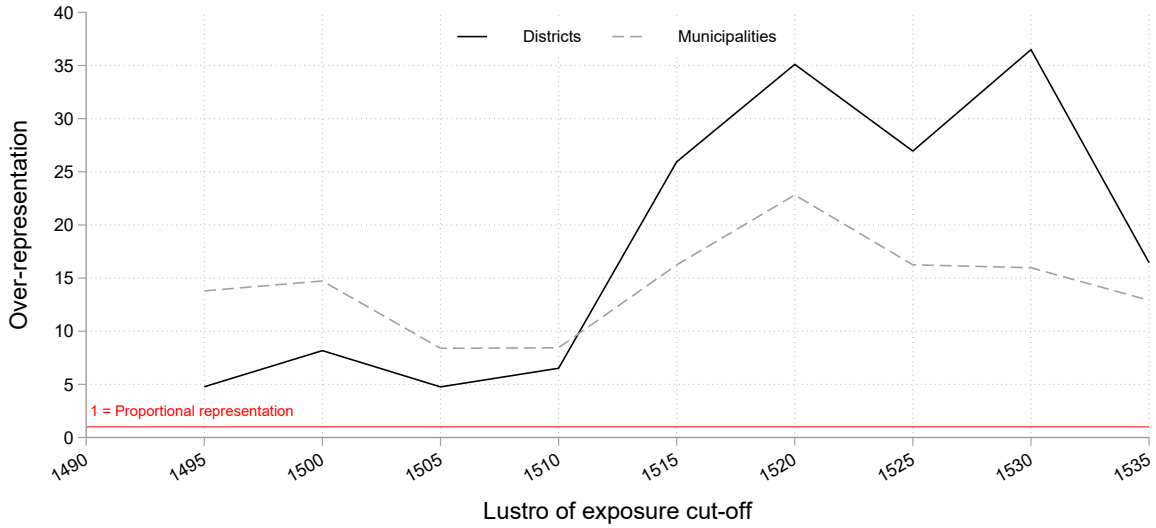
(b) β-convergence: before and after 1513



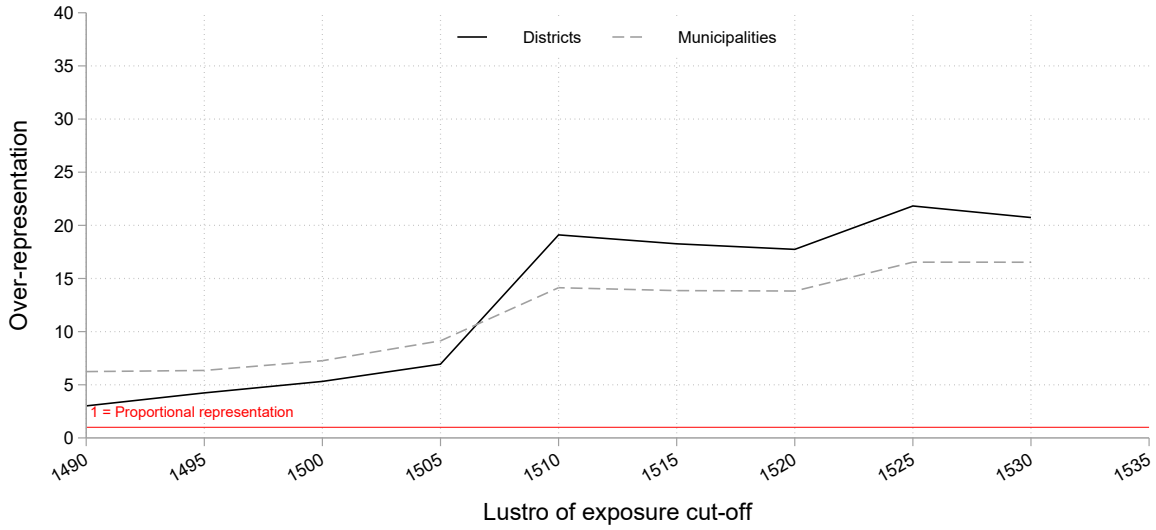
Notes: The plots report covergence in rates of emigration across Spanish modern-day provinces. Plot *a* reports the annual coefficient of variation of emigration. Plot *b* reports convergence in the rates of emigration between the periods 1492–1512 and 1513–1539; each marker is a single province. The vertical axis reports the ratio between emigration rates before and after 1513. Scales of both axes are logarithmic. One province with no pre-1513 migration is omitted. The coefficient and the standard error are from a univariate regression of the log of the ratio of emigration rates on the log of emigration rates before 1513.

Figure 11: Origin-Destination Overrepresentation

(a) Current

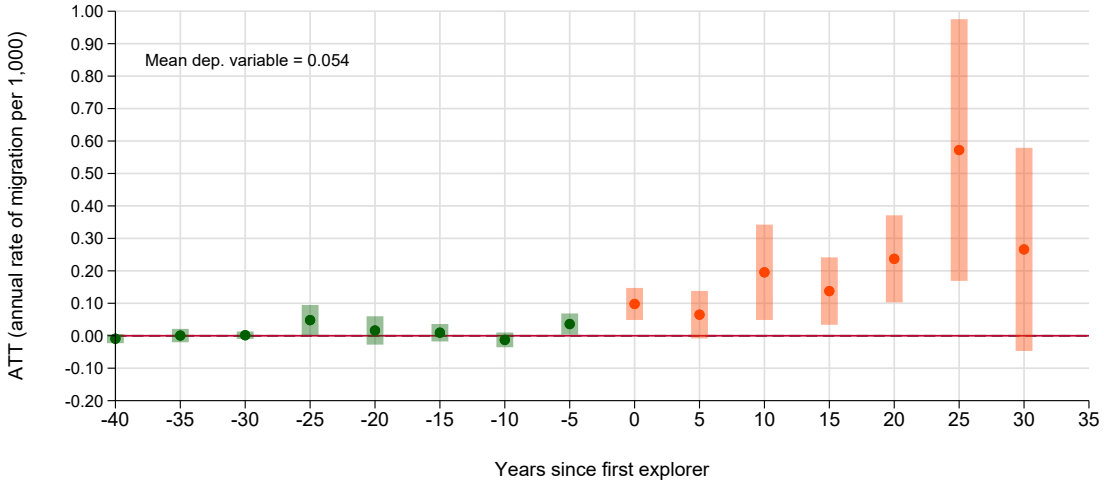


(b) Future



Notes: The figures show the over-representation in emigration of already-exposed places, where 1 is equal representation. In Panel (a), the exposed places are those that had at least one emigrant *prior* to each lustrum, and the over-representation is in migration during the current lustrum. For example, if the value in lustrum t is 2, then places that had already been exposed by lustrum $t-5$ (including), sent twice as many migrants as unexposed places during lustrum t. In Panel (b), the exposed places are those that had at least one emigrant *prior to or during* each lustrum, and the over-representation is in migration during all future periods, from the next lustrum to 1535–1539. For example, if the value in lustrum t is 2, then places that had already been exposed by lustrum t (including), sent twice as many migrants as unexposed places during the remaining lustrums, $t+5$–1535.

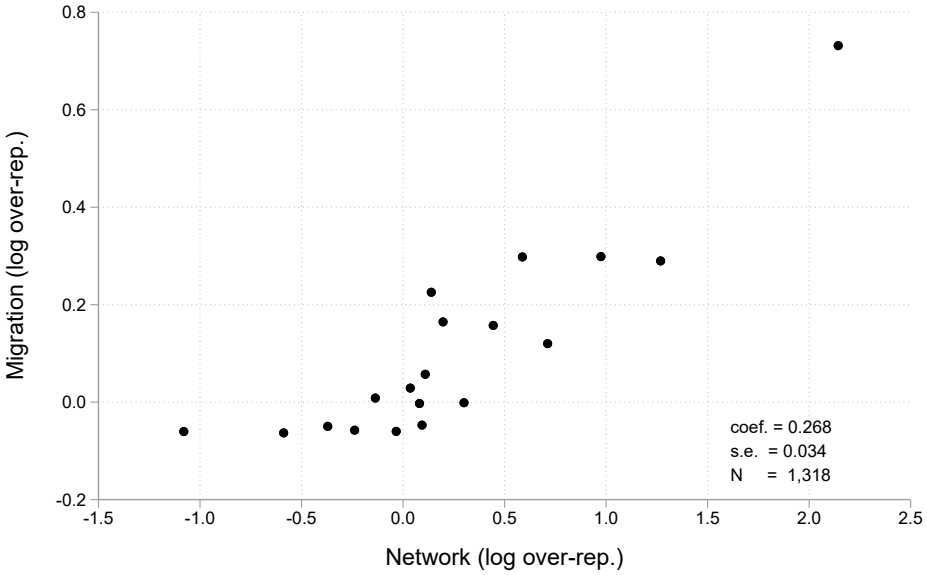
Figure 12: The effects of explorers: Event study, district-level



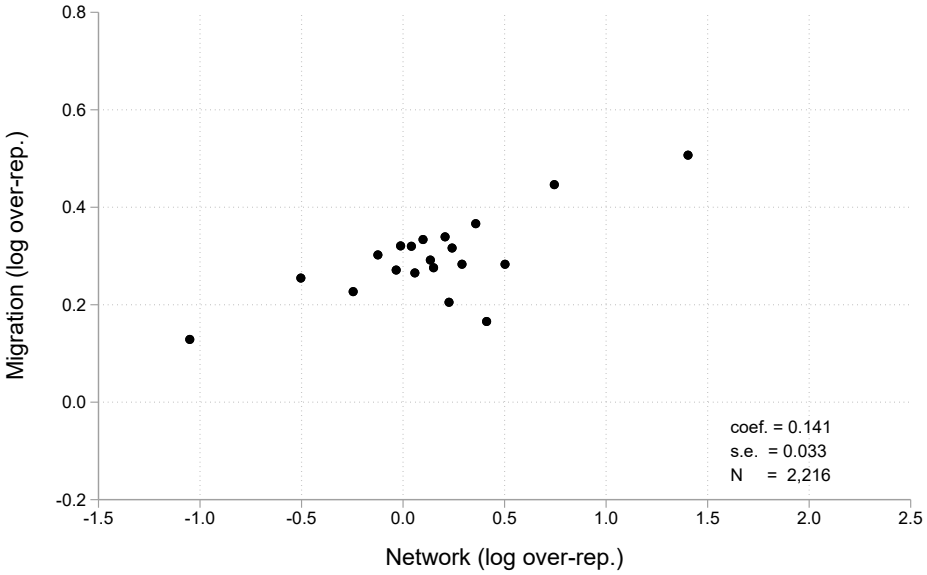
Notes: Event-study estimates of the effect of an explorer originating from a district on subsequent rate of migration from that district. The y-axis shows the average treatment effect on the treated (ATT) in terms of additional migrants per 1,000 inhabitants. Estimates are generated using the Callaway–Sant'Anna approach for staggered treatment timing with district fixed effects and lustrum periods. The horizontal line at 0 indicates no effect. Event times correspond to years since the first explorer in the district. The coefficient for the final period (35) is omitted from the figure due to an unusually large estimate based on few observations; it is reported in the text.

Figure 13: Destination-specific networks

(a) Unconditional



(b) Conditional

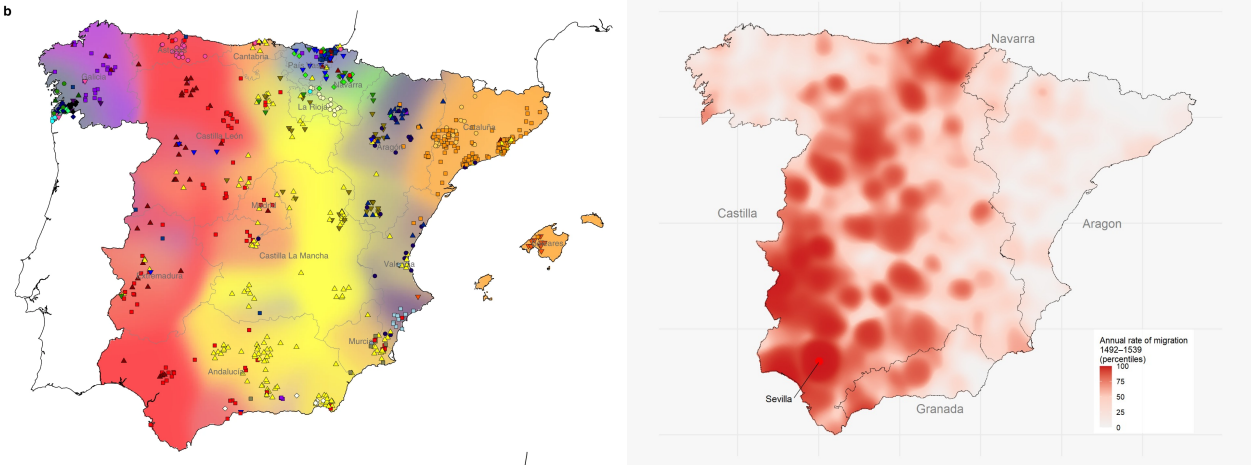


Notes: The binned scatter plots show the destination-specific correlation between migration and networks. Each observation is a district-destination-lustrum from 1495–1499 to 1535–1539. For each observation, the migration measure OR_{odt} is the log of the over-representation of migrants from district of origin o to destination d during lustrum t, relative to the entire pool of Spanish migrants during that period. The network measure is the migration measure of the same district-destination pair in the previous lustrum, $OR_{od,t-1}$. Plot (a) reports the unconditional correlation between migration and networks; due to the logarithmic transformation, observations with zero migration or zero networks are omitted. Plot (b) reports the residuals after controlling for destination × year fixed effects and an indicator for zero networks; due to the logarithmic transformation, observations with zero migration are omitted, but observations with zero networks are included. The coefficients and standard errors from log-log linear regressions are reported in the plots. Standard errors are clustered by district.

Figure 14: Migration and Genetic Variation

(a) Genetic variation

(b) Migration rates



Notes: Panel (a) displays Figure 1.b from Bycroft et al. (2019). Panel (b) shows a heat map of migration rates in our period. The region in red in panel (a) is the one identified as having primarily Leonese ancestry.

Online Appendix for

Caminante, sí hay Camino:
Networks in the First Transatlantic Migration, 1492-1540

Leticia Arroyo Abad
CUNY & CEPR

José Antonio Espín-Sánchez
Yale University

Yannay Spitzer
Hebrew University of Jerusalem & CEPR

Ariell Zimran
Vanderbilt University & NBER

February 2026

A Data Appendix

A.1 Sources for explorers and conquistadors

The voyages of exploration opened up territories for migration. Starting with Christopher Columbus, hundreds of courageous, adventurous, and in many cases deranged ventured to the Americas. As the first movers, they could be important in explaining the formation of networks back in Spain. We hand collected the names, place of origins, year of the expedition, destination, and fate of the most salient explorers. This exercise yielded 126 leading explorers for the period 1492-1540. To build this dataset, we consulted a number of secondary sources to obtain a list of the expeditions that pushed the migration frontier during this period (Esparza 2015; Fernández and Sotelo 1987; González Ochoa 2003; Kirkpatrick 2021; Montana 1943; Richman 1921). We crossed checked these names with the information in Boyd-Bowman (1964, 1968) to avoid duplicates. We also noted if the explorers attempted to settle in these new territories. The resulting list allows us to date the territories known to the Spanish over time.

In the same vein, the origin of the crew members of these expeditions may have explanatory power for the formation of networks. Typically, most crew members returned back to Spain and transmitting information about the "New World". We hand collected a list of crew members of the four Columbian voyages (1492, 1493, 1498, and 1502), the Ovando colonizing expedition (1502) from Gould (1984), Gil (1985), León Guerrero et al. (2007), and Mira Caballos (2000, 2014). Boyd-Bowman (1964, 1968) also compiled an index with a handful of crew members. From all these sources, we identified 3,786 crew members during this period with their corresponding location of origin and date of departure.

A.2 Occupations and Status Variables

We assign occupations and social status indicators through a multi-step process that combines automated matching with manual verification. The procedure begins by cleaning the raw occupation entries extracted from the original sources and moving descriptive text that does not correspond to actual occupations into a separate notes field. We then identify entries that contain more than one occupation by detecting common split patterns and their variations, and we separate these into two distinct occupation variables (`occupation1` and `occupation2`).

To standardize occupations, we match each cleaned entry against an updated crosswalk file that links original occupation strings to their corresponding HISCO and HISCLASS codes.[59] The matching proceeds in two stages: first through direct matching that prioritizes the longest possible match, and then through a conservative fuzzy match for entries that remain unmatched. Using the matched HISCO and HISCLASS codes, we assign each individual to an occupational class as follows: individuals with HISCLASS codes 1 through 5, or those flagged as *hidalgos*, are classified as managerial or professional; those with HISCLASS 6 or 7, or flagged as *vecinos*, are classified as skilled; HISCLASS 8, 9, or 10 as semi-skilled; and HISCLASS 11 or 12 as unskilled. Merchant-related and clergy-related occupations receive their own separate categories.

Alongside occupations, we construct a set of binary status indicators: `slave`, `clergy`, `military`, `hidalgo`, `nobility`, `merchant`, `artisan`, `caballero`, `state_admin`, `vecino`, and `servant`. These flags are initially assigned based on the matched occupation entries. We then update them using additional information from prefixes and titles found in individuals' full names and their fathers' names, again relying on the same crosswalk as reference. As a consistency check, we set `hidalgo`, `caballero`, `nobility`, and `vecino` to zero for any individual identified as a slave.

Because many individuals in the data travel as part of a household, we propagate status information across household members in a subsequent step. For households with more than one individual, we define kinship patterns to distinguish different family units within the same household—for instance, separating the household head's immediate family from a servant's family traveling with them. We then append the status of the household head to direct family members, and do the same for other identified subgroups within the household. The status variables updated through this propagation are `vecino`, `slave`, `hidalgo`, `nobility`,

[59] We use the crosswalk file from Espín-Sánchez et al. (2019). The matching procedure first attempts to find exact matches on full occupation strings before defaulting to shorter, more general versions.

and `servant`. Occupation is only propagated for family members of servants and slaves. In all cases, we do not overwrite any flag that has already been set to 1. A final round of updates ensures that occupations are shifted so that `occupation1` always contains the entry with a valid HISCO code when possible, and that the corresponding HISCO and HISCLASS fields are complete.

Table B.5 presents the average of each indicator by kingdom and region for all non-slave passengers arriving between 1492 and 1539. Throughout, non-positive indications are treated as negative, and passengers with unreported occupational class (HISCO) are classified as unskilled workers.

Several patterns emerge. First, the share of passengers explicitly classified as farmers is extremely small across all regions, never exceeding one percent. Castilians stand out in terms of status: 58 percent are classified as vecinos (a category that includes hidalgos), whereas this share is substantially lower in the other kingdoms. The Basque Country sent a disproportionately high share of merchants and white-collar workers (merchants fall within the white-collar category), with its white-collar share almost ten percentage points above the national average. Despite Sevilla's role as the principal port of embarkation, individuals originating from its province were only about half as likely as the average passenger to be listed as crew members. Granada, Aragon, and Navarra sent roughly twice the average share of artisans and, correspondingly, higher shares of skilled and semi-skilled workers, though fewer noblemen. Military personnel are fairly evenly distributed across kingdoms, but somewhat overrepresented in the Basque Country.

Figure B.5 shows the composition of occupational classes and its evolution over time, by lustrum. The main secular trend is a steady shift from skilled to unskilled workers. Whereas until 1504 unskilled workers accounted for roughly half of all passengers, by the 1530s they made up close to four out of five. Over the same period, the share of skilled and semi-skilled workers declined from about one-fifth to roughly one-tenth of the total. Farmers constituted only a very small fraction of migrants throughout the period. The sole exception occurs in the 1495–1499 lustrum, when individuals listed as *labradores* account for roughly 13 percent of passengers. This temporary spike is driven entirely by Columbus's third expedition in 1498; outside that convoy, the farmer share remains negligible.

A.3 Occupational status networks

We expand on the discussion in Section 4.5 and the analysis in Table 4 by adding occupational indicators to the network treatment. We ask whether the social class of previous migrants have a differential effect on current migration. For example, if merchants and other white-collar workers were more likely to sponsor fellow townsmen than unskilled workers or non–labor-force participants, then—conditional on network size—their share within the network should have a positive effect on current migration. To examine this, Table B.6 repeats the binomial and OLS specifications from Table 4, while adding five variables: four shares corresponding to the broad occupational groups (unskilled, skilled/semi-skilled, farmer, and white collar), and an additional variable capturing merchants, who are almost invariably coded as white-collar workers. When network size is positive, the first four shares sum to one. When network size is zero, all shares are set to zero by construction; all specifications include an indicator for a zero-sized network. As before, the unit of observation is the district-lustrum, and migration is measured using the lagged migration flow from the same district.

The results suggest that merchants may have had a greater capacity to draw their kinsmen and acquaintances across the Atlantic. Taking the coefficient reported in Column (4) at face value and interpreting it as a semi-elasticity, a 10 percentage point increase in the share of merchants within the network is associated with approximately 3.6 percent higher migration. Because the occupational shares sum to one when the network is positive, the coefficients are identified relative to the omitted category—unskilled workers—and therefore capture the effect of reallocating network composition away from unskilled migrants toward the indicated occupational group, holding network size constant. The corresponding OLS estimate (Column (8)) is positive and relatively large, especially compared to other occupational shares. The coefficients in the merchants after always very large, but noisily estimated when we add unit fixed-effects or a linear specification, probably due to the small sample size and the small number of merchants.

A.4 Sources for early sixteenth-Century Spanish population Counts

This appendix summarizes the principal population sources used to construct municipality- and district-level counts for Spain ca. 1530. Early modern population information is largely a by-product of fiscal administration; accordingly, the dataset relies on tax surveys, seigniorage documentation, and parish-based inquiries. Source coverage, the definition of the counted unit (typically households), and the treatment of exempt groups differ across territories.

Castile Our core source is the 1529 *Censo de Pecheros* (Instituto Nacional de Estadística 2008). The census was commissioned to revise the taxable base (*pecheros*) and provides household counts by locality, alongside information on exempt groups (e.g., nobility and clergy) (Instituto Nacional de Estadística 2008). The canvassing took several years, while revenues correspond to 1528–1529 (Instituto Nacional de Estadística 2008, p. VII).

Valencia We rely on the 1510 census and the *maridatje y coronatje* records. The 1510 census enumerated heads of household for fiscal purposes (Valldecabres Rodrigo 2002, p. 4) and is reported in the transcription used here (Garcia Carcel 1976). The *maridatje y coronatje* series provides complementary breakdowns of taxpayers and exempt groups (Viciano and Mira Jodar 2020, p. 124) and is used for social disaggregation (García Hidalgo 2019, p. 60). To convert households to population, we use the household-size guidance cited in the text: five persons per household in Capmany y de Montpalau (1807) (as reported by Pérez Puchal (1972, p. 8)) and 4–4.5 persons in (Chabás 1890).

Mallorca For Mallorca (including Menorca and Ibiza), we use the 1538 enumeration associated with the *morabatí* (Sastre Moll 1989, p. 164). Institutional background on the tax appears in Orcástegui Gros (1983, p. 113) and Baydal Sala (2008, p. 155), and the household-based character of the fogatge is noted in Sevillano Colom (1973, p. 234).

Catalonia We use the 1515 *fogatge*, the first of the available surveys (1496, 1515, 1553) to cover the Christian population as described in the text (Pagarolas i Sabaté 1981, pp. 48–9). The published transcription used is Iglesiés I Fort and Feliu I Montfort (1998).

Navarra Navarrese counts are drawn from seigniorage-related documentation and household surveys. The discussion of the monetary context and associated surveys follows (Carrasco Pérez 2011, pp. 80–1).

Aragon Aragon's main input is the 1495 household survey. The district-level enumeration structure is described in (Salas Auséns 2008, p. 694), and comparative remarks on late-fifteenth-century Aragonese population rely on (Falcón Pérez 1983, p. 263).

Granada For Granada, we use the *Servicio Extraordinario de Millones* survey (1589–1591), which extended beyond typical taxpayers to include clergy and nobility (Pintado y Rivera 1961, pp. 62–3).

Ciudad Real (Order of Calatrava) For Ciudad Real under the Calatrava Order, we use the *Relaciones Topográficas* material as cited in Pérez Puchal (1972).

Basque Country (Álava, Vizcaya, Guipúzcoa) Because fiscal exemptions complicate enumeration, we combine multiple inputs. Context on exemptions is drawn from (Angulo Morales 2003; Bilbao Bilbao 2010), and annexation context is cited to (Arregui Zamorano 1997). For Álava, institutional division into *Hermandades* is cited to (Martínez Díez 1973, p. 35), and taxpayer information is from Díaz de Durana Ortiz, Zarauz, et al. (2006, p. 66). For Vizcaya, town-level data are based on the 1591 Castilian material compiled in González (1829), with gap-filling from Enriquez Fernandex (1997). For Guipúzcoa, we use the town-level estimates in Piquero Zarauz (1988).

Harmonization and boundary limits Most sources record households rather than individuals. Following the procedure stated in the paper, we convert households to population using the 4.5 multiplier referenced to González (1829). Boundary uncertainty and the lack of accurate pre-19th-century cadastral mapping are discussed with reference to (Camarero Bullón and Aguilar Cuesta 2020; Moreno Bueno 2008), with additional caution on López (1804) and its limitations as discussed in San-Antonio-Gómez et al. (2011).

A.5 Travel cost construction

The network consists of all components described in Table A.1, which also shows the travel cost assigned to each component.

Table A.1: Travel cost details

Component	Figure	Cost
Main Roads and Mesta Roads	A.2	$\ell \times r$
Seas	A.3	$\frac{1}{6} \times \ell \times 100$
Rivers	A.3	$\frac{1}{3} \times \ell \times 100$
Ports	A.4	3000
Mode connections	A.5	$1.4 \times \ell \times 3 \times r$
Inter-municipality connections	A.6	$1.4 \times \ell \times 3 \times r$

Notes: This table shows the cost attributed to each component of the network, as well as the figure in which that component of the network is depicted.

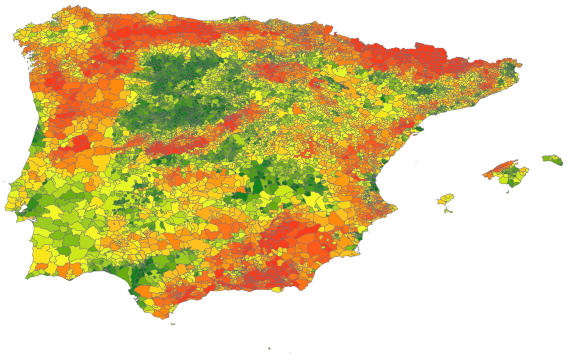
Notes:

1. ℓ denotes the length of a segment of the network in kilometers.
2. r denotes the ruggedness (standard deviation of elevation) of the municipality through which a segment passes. All segments are intersected with municipality boundaries such that each land segment lies within exactly one municipality. The average ruggedness across Spain is approximately 100. The map of ruggedness is presented in Figure A.1.
3. A factor of 1.4 is used to convert straight-line distances to true overland distances, following Fogel (1964) and Donaldson and Hornbeck (2016).
4. Access to water is available only at ports. Transshipping through a port has a cost of 3000 (i.e., about the cost of traveling 30 kilometers by road through the average municipality).
   (a) Spanish ports are the union of the medieval ports listed in Esteban-Oliver et al. (2024)
   (b) Portuguese ports are from Polónia (2008).
5. Rivers are connected directly to the sea.
6. Ports are connected directly to the sea.
7. The ranking of travel costs is as follows. These are all pretty arbitrary, but there is some logic to them.
   (a) The baseline component is the Main Roads and Mesta Roads.
      i. For Mesta Roads, we use only the "major" roads.
      ii. For the Main Roads, we use only the medieval roads from Carreras Monfort and de Soto Cañamares (2010).
   (b) Overland travel (mode connections and inter-municipality connections) costs 3 times the travel on a road, conditional on ruggedness.
   (c) Inter-municipality connections connect each municipality directly to any municipality within 30 kilometers.
   (d) Within-municipality travel distances to ports and roads are calculated using 200 random points, per Donaldson and Hornbeck (2016).
   (e) Travel by sea costs about one-sixth the cost of traveling through a municipality of average ruggedness by road.
   (f) Travel by river costs about 3 times the cost of travel by sea, or one-third the cost of overland travel through the average municipality by road, or one-ninth the cost of traveling through a municipality of average ruggedness off road.

8. All calculations include Portugal in order to allow movement through Portugal and the use of Portuguese ports.

9. The resulting minimum travel costs to Seville are presented in Figure 3 for municipalities and in Figure A.7 for districts. Given the arbitrariness of a lot of the numbers used in the calculation, I suggest that we focus on percentiles of distance rather than the distance measure itself.

Figure A.1: Ruggedness



Notes: Municipalities are divided into deciles of ruggedness

Figure A.2: Roads

(a) Mesta Roads

(b) Main Roads

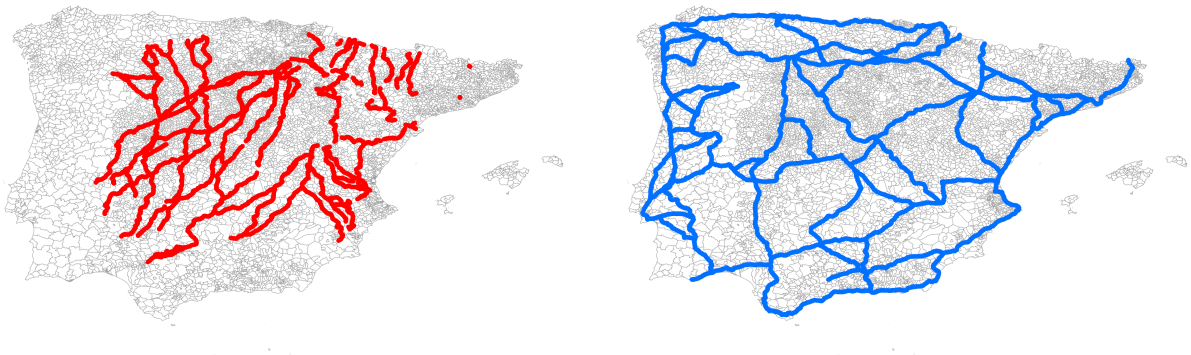
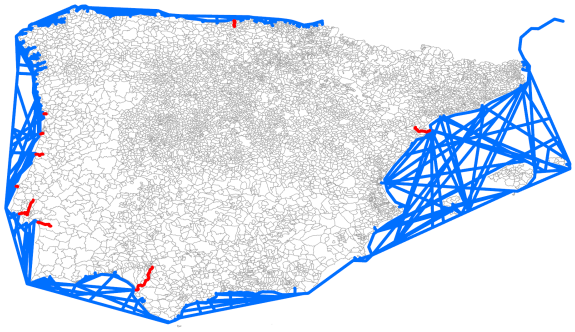


Figure A.3: Oceans and Rivers



Notes: Oceans are in blue; rivers are in red.

Figure A.4: Ports

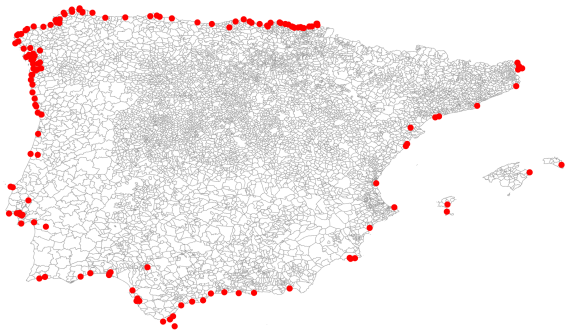
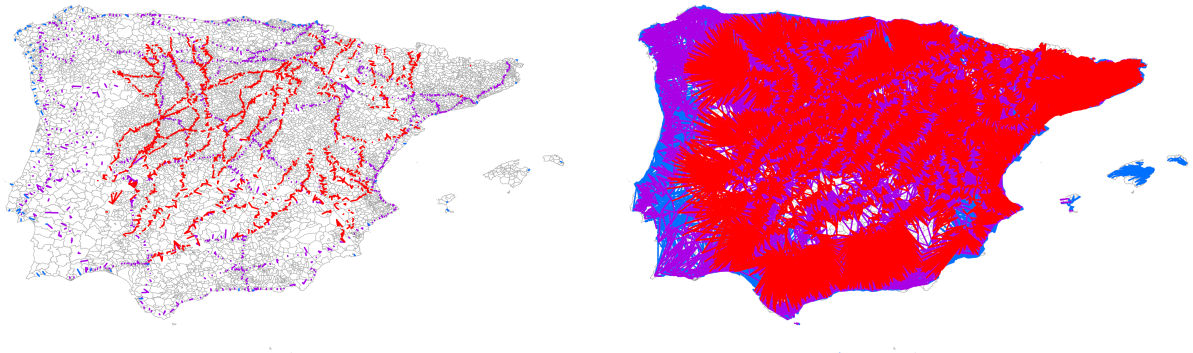


Figure A.5: Municipality-to-mode connections

(a) Within Municipalities (b) Between Municipalities



Notes: Blue lines are connections to ports. Red lines are connections to roads.

Figure A.6: Overland connections

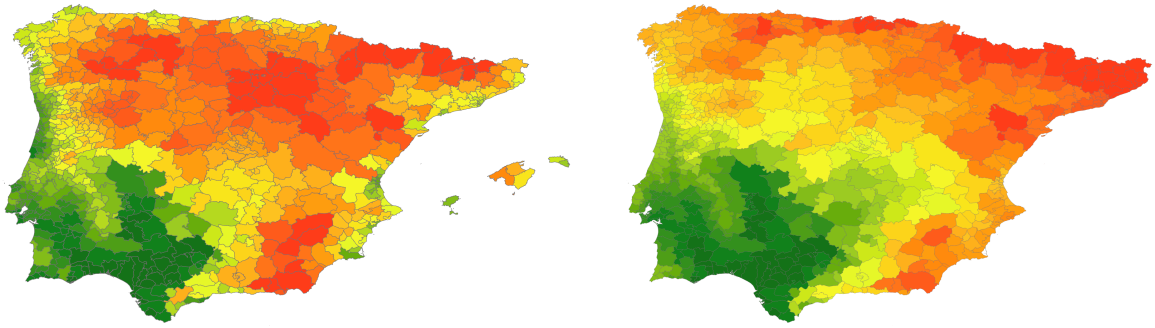


Notes: This figure shows the straight-line "wagon" routes drawn between nearby municipalities. Combined with an adjustment to convert straight lines into likely overland distances, these enable simple overland travel, though it is expensive.

Figure A.7: Travel costs to Sevilla, districts

(a) Permitting water travel

(b) Overland travel only



Notes: Each map is divided into ventiles of travel cost from Sevilla

B Econometric Appendix

In this section, we explain in more detail the econometric techniques used in the article.

B.1 Linking notarial records to the Boyd-Bowman dataset

We link the notarial protocols database to the Boyd-Bowman (BB) passenger list through a fuzzy merge that combines name similarity with origin and year proximity constraints. Both datasets underwent extensive cleaning before the merge, and several rounds of manual updates were applied afterward to resolve ambiguous cases and harmonize variables across sources (see Espín-Sánchez et al. (2022) for details on linking Spanish names).

We begin by appending 31 observations from the explorers literature that were missing from the original BB dataset. We then correct voyage years using a dedicated corrections file, add matched HISCO and HISCLASS occupational codes, add crew and title flags, and drop enslaved individuals. In a subsequent update, we resolve remaining year inconsistencies by cross-checking the corrected year variable against destination-year combinations and, where those combinations were implausible, extracting a valid four-digit or two-digit year from the original text fields. Where no valid year could be recovered, the original recorded year is retained. A final update cleans occupational strings by removing kinship information and titles unrelated to occupation, which reduces false positives in the occupational indicator variables.

The protocols dataset was constructed from a raw compiled file covering the full sixteenth century (Instituto Hispano-Cubano de Historia de América 1930–2024). Entries corresponding to payments or financial transactions rather than passenger travel were removed. The remaining entries were processed using an OpenAI-based variable extraction routine to recover structured information on names, origins, destinations, occupations, and voyage dates. Repeated entries arising from households that appeared in multiple contracts were removed using book-contract combinations, and a second deduplication pass used contract-name-year combinations to catch remaining duplicates. For entries with missing origin information, we assign the residency municipality when available, and the municipality of Seville otherwise.

We merge the cleaned protocols file with the BB dataset using a fuzzy name match subject to two additional constraints: the origin municipality must match, and the voyage years must lie within a window of ± 3 years. Name similarity is measured using a string similarity threshold of 0.85, so that only pairs whose names are at least 85 percent similar are considered candidate matches. When a single protocols entry produces more than one candidate BB record above this threshold, we resolve the ambiguity manually by comparing occupation, destination, and other details available in both sources, retaining only the match that is most consistent across variables. Candidate pairs that do not survive manual review are dropped.

After the initial merge, we carry out several rounds of updates. Destination variables in both datasets are re-extracted and standardized using a regular-expression matching function that maps raw destination strings to a common country-level taxonomy, replacing the original destination dummies. Origin variables are linked to a municipality crosswalk to recover modern municipality names, community names, province names, and provincial codes. Status and occupational class variables—including indicators for hidalgos, caballeros, nobility, military, clergy, state administrators, merchants, artisans, slaves, and servants—are constructed by combining information from both sources, using family members' information to fill gaps where the individual-level record is incomplete.

B.2 Binomial Maximum Likelihood

The binomial maximum likelihood regression expresses the migration rate of municipality i in period t, p_{it}, in the logit form

$$p_{it} = \frac{\exp(x'_{it}\beta)}{1 + \exp(x'_{it}\beta)},$$

or equivalently

$$\log\left(\frac{p_{it}}{1 - p_{it}}\right) = x'_{it}\beta,$$

where x_{it} are a set of regressors that can vary at the municipality-time level. After determining this logit migration demand, we use the binomial distribution to determine the probability that a given number of emigrants are observed from municipality i in time period t given p_{it} and the municipality's baseline population N_i, which enables us to estimate β by maximum likelihood using the log-likelihood function (after removing constants)

$$\mathfrak{L} = \sum_i \sum_t e_{it} N_i \log(p_{it}) + (1 - e_{it}) N_i \log(1 - p_{it}). \tag{B.1}$$

It is easy to see from equation (B.1) that the binomial maximum likelihood is identical to the likelihood for an individual-level logit regression for migration in which the regressors are all at the municipality-time level.

This method has two main advantages relative to a standard OLS regression of migration rates on local regressors. First, being derivable from an individual-level logit regression, it has microfoundations in a way that the OLS regression does not. Second, the BML regression assumes that effects are proportional to the level of migration. This can be achieved in an OLS regression by using log migration as the outcome, but then any location-period with no migration must be excluded from the analysis. The BML method captures these observations of zero migration by treating them as cases in which all individuals have a strictly positive migration probability, p_{it}, but the realization of every resident of the commune is to stay.

The main disadvantage of the BML is that, as a non-linear model, the inclusion of high-dimensional fixed effects in x_{it} creates an incidental parameters problem. To address this issue, we first establish that the results of linear regressions are not materially different from those of the BML, and then take advantage of the fact that the direction of bias in the case of a linear model is known to bound our estimates.

When the p_{it} are small, as in our case, two aspects of interpretation of the coefficients β are simplified. First, as in all logit specifications, the coefficients can be interpreted as they would be in a regression of $\log(p_{it})$ on the regressors—that is, rather than interpreting β as the effect of x on the log-odds of migration, β can be interpreted as the relationship between a unit change in the regressor and percent changes in migration probability. Second, it can be shown, and we have found that it is empirically true, that the coefficients and their interpretation are essentially identical to those that would arise from a Poisson regression using the number of emigrants as the outcome variable.

B.3 Dyadic

We want to test whether migrants who share geographic origins, surnames, or family ties were more likely to migrate together to the Americas, i.e., whether they sailed with the same captain, on the same voyage, or used the same notary to formalize their passage. The natural empirical framework for this question is a *dyadic regression*, where the unit of observation is a pair of migrants and the dependent variable captures whether the pair shares some migration outcome.

We draw on two archival sources covering migration from Spain to the Americas up to 1540. The first is the set of passenger registers (*asientos*) from the Archivo General de Indias (Contratacion, 5536), which record the migrant's place of origin, the name of the ship master, and the year of departure. The second is the set of notarial protocols, which additionally record the name of the ship, the notary (*escribano*), and the notarial office (*notaría*).

From a dataset of N migrants, we construct all $\binom{N}{2} = \frac{N(N-1)}{2}$ unique pairs. Each pair (i, j) constitutes one observation in the dyadic dataset. For each pair (i, j), we estimate a linear probability model of the form:

$$Y_{ij} = \alpha + \beta \, \text{SameOrigin}_{ij} + \gamma_t + \delta_p + \varepsilon_{ij} \tag{B.1}$$

where β is a vector of parameters and $SameOrigin_{ij}$ is a vector of regressors of interest. The variables are defined as follows. We consider several dependent variables Y_{ij}, each coded as an indicator:

- **Same voyage**: equals 1 if migrants i and j traveled with the same ship master in the same year.

- **Same master**: equals 1 if i and j traveled with the same master, regardless of year.
- **Same notary**: equals 1 if i and j used the same notary to register their passage (protocolos only).
- **Same notarial office**: equals 1 if i and j used the same notarial office (protocolos only).

The key regressors capture social proximity between migrants i and j:

- SameTown_{ij}: equals 1 if both migrants come from the same town/village of origin, or the same neighborhood within the city of Seville.
- $\text{SameMuniciaplity}_{ij}$: equals 1 if both migrants come from the same municipality.
- SameProvince_{ij}: equals 1 if both come from the same province.
- SameSurname_{ij}: equals 1 if both migrants share the same surname.
- FamilyTie_{ij}: equals 1 if migrant j's surname matches the surname of i's father or mother, or vice versa. This captures potential kinship ties beyond the nuclear surname match.

The term γ_t represents year fixed effects, defined over the minimum year in the pair, $t = \min(\text{year}_i, \text{year}_j)$. These absorb aggregate changes in migration flows and fleet composition over time. The term δ_p represents province fixed effects for both members of the dyad, absorbing broad regional patterns in migration.

B.3.1 Multi-Way clustering

Dyadic observations are not independent. Migrant i appears in $N-1$ pairs, so the error terms ε_{ij} and ε_{ik} for any $j \neq k$ are correlated through their shared dependence on i. Ignoring this dependence leads to severely understated standard errors. We address this by computing multi-way clustered standard errors following Cameron et al. (2011). The variance estimator takes the form:

$$\widehat{V}(\hat{\beta}) = \widehat{V}_i + \widehat{V}_j - \widehat{V}_{ij} \tag{B.2}$$

where $\widehat{V}_i$ clusters on the first member of the dyad, $\widehat{V}_j$ clusters on the second member, and $\widehat{V}_{ij}$ is the intersection. Since each pair (i, j) with $i < j$ is unique, the intersection reduces to the heteroskedasticity-robust (HC1) variance estimator. Each component takes the standard sandwich form. For clustering on the first member i:

$$\widehat{V}_i = (X'X)^{-1} \left(\sum_{c=1}^{N} \mathbf{u}_c \mathbf{u}_c' \right) (X'X)^{-1} \tag{B.3}$$

where $\mathbf{u}_c = \sum_{(i,j):i=c} X_{ij}' \hat{\varepsilon}_{ij}$ is the sum of score contributions across all dyads that include individual c as the first member.[60] The formula for $\widehat{V}_j$ is analogous, clustering on the second member.

The intersection term is:

$$\widehat{V}_{ij} = (X'X)^{-1} \left(\sum_{(i,j)} \hat{\varepsilon}_{ij}^2 \, X_{ij}' X_{ij} \right) (X'X)^{-1} \tag{B.4}$$

which is the standard HC1 robust variance estimator.

To ensure that the resulting variance-covariance matrix is positive semi-definite, we apply an eigenvalue adjustment: any negative eigenvalues of $\widehat{V}(\hat{\beta})$ are set to zero before recomposing the matrix.

[60] Dekker et al. (2007) propose permutation tests for network autocorrelation in dyadic regression, offering a nonparametric alternative to cluster-robust standard errors.

Interpretation The coefficient β on SameTown_{ij} gives the increase in the probability that two migrants share a migration outcome (e.g., same master) when they come from the same town, relative to pairs from different towns. If $\beta = 0.05$ in the same-master regression, this means that pairs from the same town are 5 percentage points more likely to travel with the same master than pairs from different towns. The family tie variable helps distinguish between two mechanisms. A large coefficient on SameSurname_{ij} could reflect either kinship networks or simply that common surnames are correlated with geographic origin. The FamilyTie_{ij} variable provides a more direct test of kinship by checking whether one migrant's surname matches the parental surnames of the other.

Practical considerations With N migrants, the dyadic dataset contains $N(N-1)/2$ observations. For $N = 1{,}000$, this yields about 500,000 dyads; for $N = 5{,}000$, it yields over 12 million. Computation of multi-way clustered standard errors requires looping over all N clusters, which can be slow for large samples (see also Sasaki and Zimran 2025). If N is very large, one option is to restrict the sample to migrants within the same decade or the same broad region. To speed up our estimation we divide the sample into blocks of 10 million pairs, and compute each block independently.

Appendix Tables

Table B.1: Cross section determinants, districts, binomial, full period

Variables	(1)	(2)	(3)	(4)	(5)	(6)
Travel cost (roads+water)	−0.916[a]				−0.684[b]	−1.436[a]
	(0.298)				(0.266)	(0.366)
Urban		1.368[b]			0.730[a]	1.025[a]
		(0.606)			(0.209)	(0.252)
Elevation			−0.491[b]		0.186	0.201
			(0.225)		(0.194)	(0.216)
Soil Quality				0.063	0.172[b]	0.118
				(0.054)	(0.077)	(0.108)
Near River					−0.244	−0.149
					(0.207)	(0.336)
Near Coast					−0.120	0.087
					(0.199)	(0.308)
Granada					−0.211	
					(0.301)	
Aragon					−2.238[a]	
					(0.255)	
Navarra					−1.011[a]	
					(0.273)	
Sevilla					0.980[a]	
					(0.322)	
Badajoz					0.498[c]	
					(0.276)	
Observations	389	389	389	389	389	389
FE	None	None	None	None	None	Prov

Significance levels: [a] p<0.01, [b] p<0.05, [c] p<0.1
Notes: Standard errors clustered by district in parentheses. Dependent variable is the emigration rate to any destination. The period covered is 1492–1539. Unit of observation is a district. The sample includes all of Spain. Districts with population under 500 are omitted.

Table B.2: Cross section determinants, municipalities, binomial, before 1513

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Travel cost (roads+water)	−1.031[a]				−0.747[a]	−0.567[b]	0.052
	(0.187)				(0.279)	(0.256)	(0.424)
Population (log.)		0.593[a]			0.474[a]	0.604[a]	0.658[a]
		(0.174)			(0.076)	(0.065)	(0.098)
Elevation			−0.860[a]		−0.059	−0.638[c]	−0.777
			(0.323)		(0.144)	(0.342)	(0.496)
Soil Quality				−0.044	0.167	−0.126	−0.136
				(0.074)	(0.108)	(0.128)	(0.138)
Near River					−0.412	−0.644	−0.082
					(0.406)	(0.464)	(0.465)
Near Coast					0.269	0.586	−0.523
					(0.270)	(0.425)	(0.664)
Granada					−1.228[a]		
					(0.379)		
Aragon					−2.829[a]		
					(0.463)		
Navarra					−1.662[b]		
					(0.800)		
Sevilla					0.124		
					(0.375)		
Badajoz					−0.007		
					(0.319)		
Observations	6,387	6,388	6,388	6,388	6,236	6,236	6,236
FE	None	None	None	None	None	Prov	Dist

Significance levels: [a] p<0.01, [b] p<0.05, [c] p<0.1

Notes: Standard errors clustered by district in parentheses. Dependent variable is the emigration rate to any destination. The period covered is 1492–1512. Columns 6–8 include jurisdiction fixed effects. Unit of observation is a municipality. The sample includes all of Spain. Municipalities with population under 20 are omitted.

Table B.3: Cross section determinants, municipalities, binomial, after 1513

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Travel cost (roads+water)	−0.828[a]				−0.329[b]	−0.488[b]	−0.310
	(0.214)				(0.128)	(0.243)	(0.231)
Population (log.)		0.585[a]			0.455[a]	0.594[a]	0.671[a]
		(0.161)			(0.070)	(0.053)	(0.063)
Elevation			−0.552[c]		0.025	−0.092	−0.173
			(0.300)		(0.124)	(0.136)	(0.173)
Soil Quality				0.000	0.104	−0.038	−0.060
				(0.075)	(0.078)	(0.084)	(0.092)
Near River					−0.117	0.066	−0.275
					(0.172)	(0.226)	(0.439)
Near Coast					−0.198	0.134	0.028
					(0.210)	(0.321)	(0.276)
Granada					−0.238		
					(0.210)		
Aragon					−2.267[a]		
					(0.280)		
Navarra					−0.605[b]		
					(0.297)		
Sevilla					0.745[b]		
					(0.342)		
Badajoz					0.863[a]		
					(0.212)		
Observations	6,387	6,388	6,388	6,388	6,236	6,236	6,236
FE	None	None	None	None	None	Prov	Dist

Significance levels: [a] p<0.01, [b] p<0.05, [c] p<0.1
Notes: Standard errors clustered by district in parentheses. Dependent variable is the emigration rate to any destination. The period covered is 1513–1539. Columns 6–8 include jurisdiction fixed effects. Unit of observation is a municipality. The sample includes all of Spain. Municipalities with population under 20 are omitted.

Table B.4: Geography regressions, binomial, with taxes

Variables	(1)	(2)	(3)	(4)	(5)	(6)
Assigned taxes (log. p.c.), 1528	−0.034			−0.169^{c}	−0.033	−0.128
	(0.079)			(0.095)	(0.086)	(0.085)
Population (log.)		0.579^{a}		0.556^{a}	0.678^{a}	0.511^{a}
		(0.170)		(0.132)	(0.112)	(0.115)
Travel cost (roads+water)			−0.876^{a}	−0.757^{a}	−0.575	−0.480
			(0.201)	(0.252)	(0.426)	(0.499)
Elevation				0.651^{a}	−0.077	−0.095
				(0.195)	(0.293)	(0.362)
Ruggedness				−0.464^{a}	−0.153	0.202
				(0.165)	(0.178)	(0.236)
Soil Quality				0.067	−0.080	−0.117
				(0.122)	(0.142)	(0.137)
Near River				−0.113	−0.311	−0.848^{b}
				(0.361)	(0.364)	(0.329)
Near Coast				0.407	0.257	−0.183
				(0.330)	(0.440)	(0.318)
Urban				−0.220	−0.280	0.234
				(0.287)	(0.283)	(0.309)
Sevilla				0.659		
				(0.414)		
Badajoz				0.345		
				(0.446)		
Observations	1,523	6,224	6,223	1,511	1,511	1,511
FE	None	None	None	None	Prov	Dist

Significance levels: ^{a} p<0.01, ^{b} p<0.05, ^{c} p<0.1

Notes: Standard errors clustered by district in parentheses. Dependent variable is the emigration rate to any destination. Unit of observation is a municipality. Municipalities with population under 20 are omitted.

Table B.5: Descriptive statistics of passengers

		Kingdoms				Provinces		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Variables	Spain	Castilla	Granada	Aragon	Navarra	Sevilla	Badajoz	Basque
STATUS								
Crew	0.167	0.166	0.177	0.197	0.192	0.084	0.146	0.201
Explorer	0.004	0.004	0.003	0.001	0.000	0.001	0.005	0.002
Vecino or Hidalgo	0.585	0.583	0.298	0.231	0.155	0.624	0.648	0.400
Hidalgo	0.010	0.010	0.009	0.008	0.005	0.007	0.008	0.019
Nobility	0.025	0.025	0.009	0.008	0.006	0.035	0.020	0.030
Clergy	0.021	0.021	0.023	0.027	0.025	0.019	0.021	0.022
Merchant	0.017	0.017	0.004	0.007	0.003	0.024	0.014	0.060
Artisan	0.034	0.034	0.065	0.070	0.079	0.052	0.021	0.020
Military	0.084	0.084	0.090	0.090	0.091	0.069	0.097	0.122
Caballero	0.002	0.002	0.002	0.001	0.000	0.002	0.002	0.000
OCC. CLASS								
Unskilled	0.758	0.756	0.695	0.684	0.672	0.748	0.750	0.676
Farmer	0.008	0.008	0.009	0.005	0.007	0.003	0.014	0.003
Skilled/semi-skilled	0.097	0.097	0.172	0.181	0.198	0.125	0.100	0.081
White collar	0.137	0.138	0.124	0.130	0.123	0.123	0.136	0.240
Observations	20,972	20,087	1,631	1,509	1,259	5,176	2,097	910

Notes: The table reports the averages of status indicators. The samples includes all 1492–1539 passengers who are not slaves. The Basque Country includes the provinces of Álava, Guipúzcoa, and Vizcaya. By design, the occupational class variables add up to 1 (passengers without definite information are counted as unskilled). Every hidalgo is also a vecino.

Table B.6: The role of occupational status in the networks

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mig. lag (log. p.c.) × mig. lag > 0	0.863^a	0.670^a	0.496^a	0.055^c				
	(0.057)	(0.033)	(0.046)	(0.030)				
Network					0.928^a	0.897^a	0.852^a	0.417
					(0.178)	(0.170)	(0.189)	(0.268)
Mig. lag = 0	−4.274^a	−3.428^a	−2.388^a	−0.293^b	−0.025^b	−0.016	−0.003	−0.011
	(0.206)	(0.180)	(0.160)	(0.146)	(0.012)	(0.013)	(0.013)	(0.015)
Farmer	0.140	−0.090	0.258	0.443^c	0.025	−0.005	0.012	−0.030
	(0.604)	(0.471)	(0.436)	(0.247)	(0.036)	(0.040)	(0.043)	(0.044)
Skilled/semi-skilled	0.453^b	0.214	0.327	−0.020	−0.004	−0.006	−0.006	−0.025
	(0.191)	(0.234)	(0.211)	(0.188)	(0.018)	(0.018)	(0.018)	(0.018)
White collar	0.071	0.220	0.177	−0.073	0.004	0.010	0.012	−0.005
	(0.166)	(0.141)	(0.160)	(0.177)	(0.019)	(0.018)	(0.016)	(0.020)
Merchant	1.249^a	0.989^a	0.883^a	0.363^c	0.070	0.070	0.054	0.050
	(0.430)	(0.282)	(0.235)	(0.207)	(0.050)	(0.048)	(0.062)	(0.058)
Observations	3,890	3,890	3,890	3,890	3,890	3,890	3,890	3,890
R-squared					0.391	0.401	0.417	0.489
Pseudo R-squared	0.743	0.762	0.792	0.842				
Controls		✓	✓			✓	✓	
FE			Prov	Dist			Prov	Dist
Model	Binomial	Binomial	Binomial	Binomial	OLS	OLS	OLS	OLS
Mean rate mig.	0.054	0.054	0.054	0.054	0.054	0.054	0.054	0.054
Marginal effect ×10k	3.490	2.709	2.003	0.221				

Significance levels: ^{a} p<0.01, ^{b} p<0.05, ^{c} p<0.1

Notes: This table presents the results of regression for migration to any destination. Unit of observation is a district-lustro. Districts with population under 500 are omitted. Dependent variable in the binomial regressions is the number of migrants to any destination. Dependent variable in the OLS regressions is the annual rate of migration to any destination per 10,000, and its mean is reported. Controls include standardized travel distance to Sevilla, share of urban population, elevation, ruggedness, and indicators for proximity to the coast and to navigable rivers. Network (migrant lag) is defined as the annual rate of migration to any destination per 10,000 in the previous period. Occupation classes are the shares of each category within the network. They add up to 1, unless the size of the network is zero, in which case they are all zero. Merchants are almost always white collar workers. All specifications control for an indicator for the size of the network being zero (Mig. lag. = 0). All specifications include period-fixed effects. Coefficients of binomial regressions can be interpreted as in a logit regression, as the effect on the log-odds ratio. Since the probability of migration is close to zero, this is close to semi-elasticities. Pseudo R-squared values for the binomial and Poisson regressions are McFadden's. Standard errors clustered by province in parentheses.

Appendix Figures

Figure B.1: Migration counts by destination over time

(a) Caribbean



(b) Mexico



(c) Tierra Firme



(d) Peru

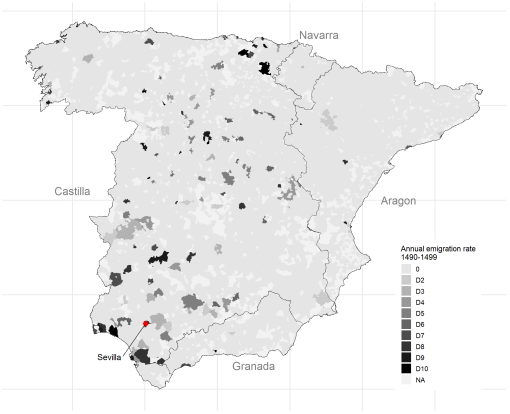


(e) La Plata

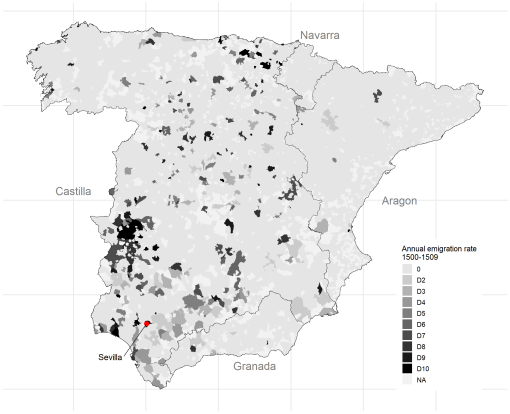


Notes: This figure presents the number of migrants per year to each major destination group from any place of origin.

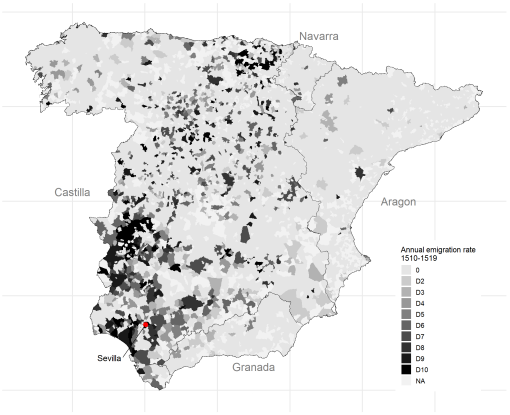
Figure B.2: Emigration rate by decade, 1490s–1530s



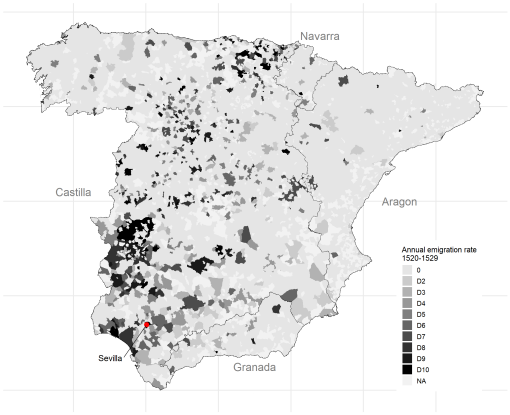
(a) Emigration 1490s



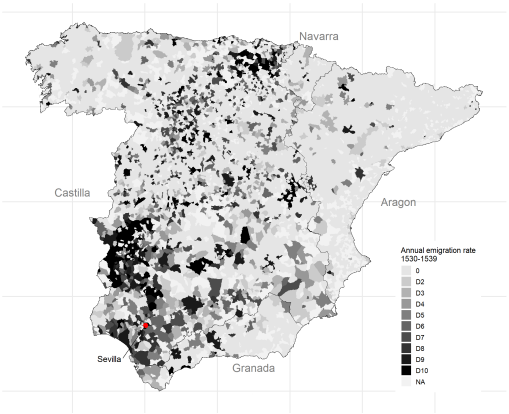
(b) Emigration 1500s



(c) Emigration 1510s



(d) Emigration 1520s



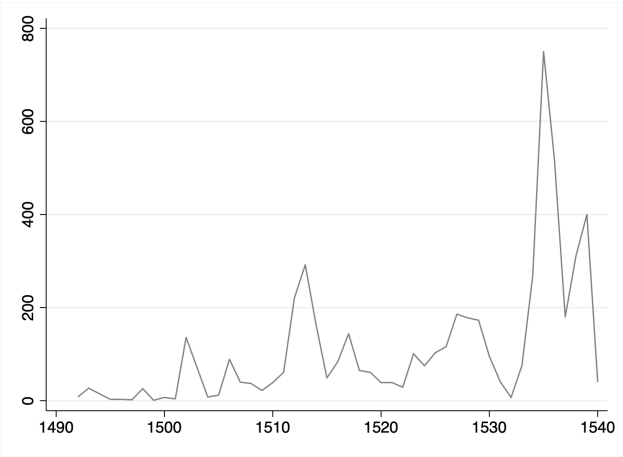
(e) Emigration 1530s

Figure B.3: Migration counts over time

(a) All places

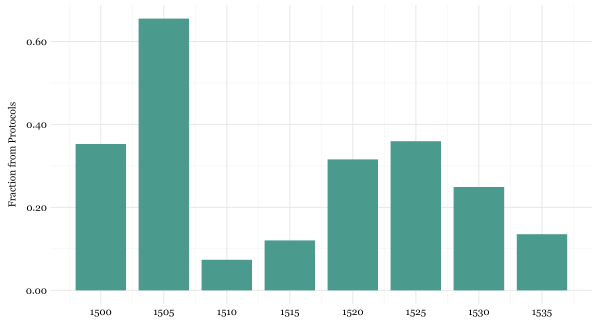


(b) Seville

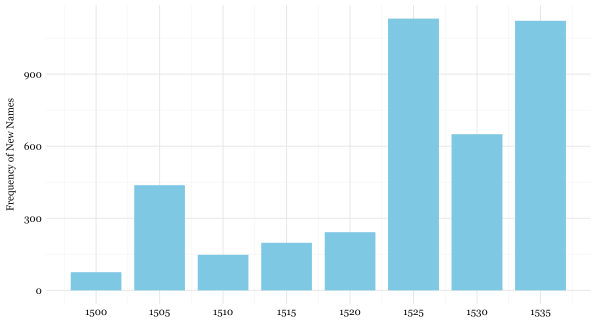


Notes: Panel (a) presents the count of the number of migrants in each year from all places of origin to all destinations. Panel (b) limits attention to the province of Seville.

Figure B.4: New Passengers from Notarial Protocols



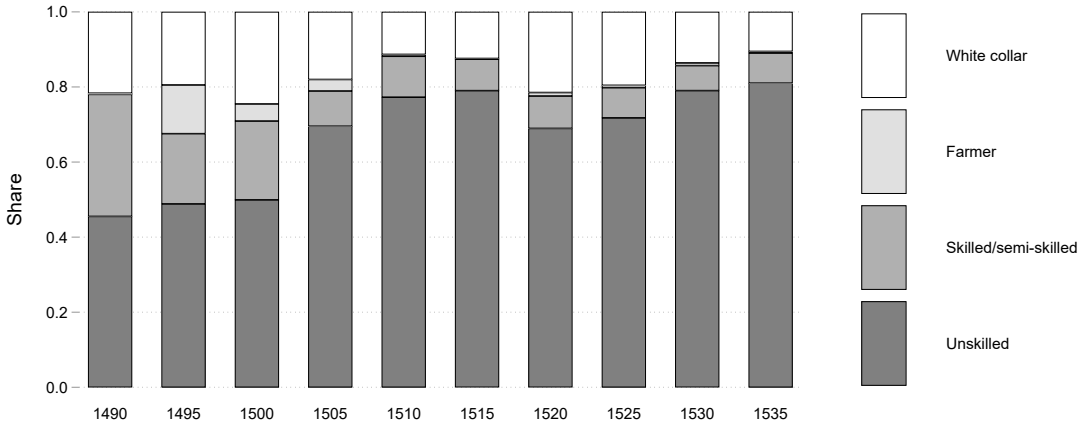
(a) New names as a fraction of existing BB records



(b) New passenger count

*Notes:*The sample covers notarial protocol records that could not be matched to any existing entry in the Boyd-Bowman dataset following the two-pass linking procedure described above. The conservative definition includes only those protocol entries for which no BB record satisfied both the Jaro-Winkler name distance threshold of 0.1 and the ±3 year window. The less-strict definition additionally reclassifies as non-matches those entries that were flagged as incorrectly matched during manual review of ambiguous cases. Panel (a) expresses the count of new primary passenger names as a fraction of the total number of BB passengers recorded in the same five-year interval. Panel (b) counts all passengers listed on the same notarial record as an unmatched primary passenger, including accompanying family members and dependents.

Figure B.5: Composition of occupational class (HISCO) over time



References

Angulo Morales, Alberto (2003). "Fiscalidad y economía en las provincias Exentas durante la Edad Moderna". *Estudis: Revista de historia moderna* 29, pp. 81–101.

Arregui Zamorano, Pilar (1997). "Álava en Castilla: apuntes sobre la relación existente entre el territorio de Álava y el adelantamiento castellano". *Anuario de Historia del Derecho Español*, pp. 1109–1130.

Baydal Sala, Vicent (2008). "El naixement de l'impost del morabatí al regne de València (1265-1266)". *Anales de la Universidad de Alicante: Historia Medieval* 15, pp. 141–.

Bilbao Bilbao, Luis (2010). "Treasury and Taxation in the Basque Province of Álava in the XVIth and XVIIth centuries". *Studia Historica. Historia Moderna* 32, pp. 83–124.

Boyd-Bowman, Peter (1964). *Indice Geobiográfico de Cuarenta Mil Pobladores Españoles de América En El Siglo XVI, Tomo I, 1493-1519,* Bogotá: Instituto Caro y Cuervo.

——— (1968). *Indice Geobiográfico de Cuarenta Mil Pobladores Españoles de América En El Siglo XVI, Tomo II, 1520-1539*. Mexico City: Editorial Jus, Academia Mexicana de Genealogía y Heráldica, A. C.

Camarero Bullón, Concepción and Ángel Ignacio Aguilar Cuesta (2020). "La Cartografía, instrumento para conocer el territorio, planificar y gestionar las reformas en la España del siglo XVIII". *Manuscrits: revista d'història moderna* 42, pp. 157–178.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller (2011). "Robust Inference With Multiway Clustering". *Journal of Business & Economic Statistics* 29:2, pp. 238–249.

Capmany y de Montpalau, Antonio (1807). *Cuestiones críticas sobre varios puntos de historia económica, política y militar*. Madrid: Imprenta Real.

Carrasco Pérez, Juan (2011). "El impuesto del monedaje en el reino de Navarra (ca. 1243-1355): fiscalidad, demografía, historia moderna". *Príncipe de Viana* 252, pp. 55–161.

Carreras Monfort, Cèsar and Pau de Soto Cañamares (2010). *Historia de La Movilidad En La Península Ibérica: Redes de Transporte En SIG*. Barcelona: Editorial UOC.

Chabás, Roque (1890). "Los moriscos de Valencia y su expulsión". *El Archivo, Revista de Ciencias Históricas* IV. Reprinted 1980, pp. 231–234.

Dekker, David, David Krackhardt, and Tom A. B. Snijders (2007). "Sensitivity of MRQAP Tests to Collinearity and Autocorrelation Conditions". *Psychometrika* 72:4, pp. 563–581.

Díaz de Durana Ortiz, José Ramón, Santiago Piquero Zarauz, et al. (2006). "Fiscalidad real, fiscalidad municipal y nacimiento de las haciendas provinciales en el País Vasco (ss. XIII al XV)". In *Fiscalidad de Estado y fiscalidad municipal en los reinos hispánicos medievales*. Casa de Velázquez, pp. 53–90.

Donaldson, Dave and Richard Hornbeck (2016). "Railroads and American Economic Growth: A "Market Access" Approach". *The Quarterly Journal of Economics* 131:2, pp. 799–858.

Enriquez Fernandex, et alli (1997). *Foguera de las villas de Vizcaya de 1514*. Bilbao: Eusko Ikaskuntza.

Esparza, José Javier (2015). *La Cruzada Del Océano: La Gran Aventura de La Conquista de América*. Madrid: La esfera de los libros.

Espín-Sánchez, José-Antonio, Salvador Gil-Guirado, W. Daniel Giraldo-Paez, and Chris Vickers (2019). "Labor Income Inequality in Pre-Industrial Mediterranean Spain: The City of Murcia in the 18th Century". *Explorations in Economic History* 73, p. 101274.

Espín-Sánchez, José-Antonio, Salvador Gil-Guirado, and Chris Vickers (2022). "La "Doña" è Mobile: The Role of Women in Social Mobility in a Pre-Modern Economy". *The Journal of Economic History* 82:1, pp. 1–41.

Esteban-Oliver, Guillermo, Eduard J. Álvarez-Palau, and Jordi Martí-Henneberg (2024). "Infraestructuras para la navegación en la España moderna y liberal: puertos, faros, canales y ríos navegables". *Investigaciones de Historia Económica*, pp. 18–25.

Falcón Pérez, María Isabel (1983). "La Población Aragonesa a Fines Del Siglo XV". *Aragón en la Edad Media* 5, pp. 249–272.

Fernández, Rafael Diego and Rafael Diego-Fernández Sotelo (1987). *Capitulaciones Colombinas (1492-1506)*. Zamora, Mich.: El Colegio de Michoacán.

Fogel, Robert William (1964). *Railroads and American Economic Growth: Essays in Econometric History*. Baltimore, MD: Johns Hopkins Press.

Garcia Carcel, Ricardo (1976). "El censo de 1510 y la población valenciana de la primera mitad del siglo XVI". *Cuadernos de geografía* 18, pp. 49–66.

García Hidalgo, Palmira (2019). "La Emigración Española a América En La Época Moderna. Un Acercamiento al Estado de La Cuestión". *Naveg@merica* 23.

Gil, Juan (1985). "El Rol Del Tercer Viaje Colombino". *Suplemento de Anuario de Estudios Americanos, Sección Historiografía y Bibliografía* 29:1, pp. 83–110.

González, Tomás (1829). *Censo de poblacion de las provincias y partidos de la corona de Castilla en el siglo XVI*. Madrid: Imprenta real.

González Ochoa, José Ma (2003). *Quién Es Quién En La América Del Descubrimiento: 1492-1600*. Madrid: Acento Editorial.

Gould, Alice Bache (1984). *Nueva Lista Documentada de Los Tripulantes de Colón En 1492*. Madrid: Real Academia de la Historia.

Iglesiés I Fort, J. and G Feliu I Montfort (1998). "El Fogatge de 1515". *Butlleti de la Societat Catalana d'Estudis Históric*s 9, pp. 23–85.

Instituto Hispano-Cubano de Historia de América (1930–2024). *Catálogo de los Fondos Americanos del Archivo de Protocolos de Sevilla, Siglo XVI*. 16 vols. Tomos I–II (1930), III (1932), IV (1935), V (1937), VI (1986), VII (1990), VIII (2000), IX (2004), X (2006), XI (2009), XII (2011), XIII (2014), XIV (2017), XV (2020), XVI (2024). Tomos I–III originally published by Compañía Ibero-Americana de Publicaciones, Madrid. Sevilla: Instituto Hispano-Cubano de Historia de América.

Instituto Nacional de Estadística (2008). *Censo de Pecheros, Carlos I, 1528*. Vol. I-II. Madrid: Instituto Nacional de Estadística.

Kirkpatrick, Frederick A. (2021). *Los Conquistadores Españoles*. Trans. by Rafael Vásquez Zamora. Madrid: Ediciones Rialp, S.A.

León Guerrero, María Montserrat et al. (2007). "Pasajeros Del Segundo Viaje de Cristóbal Colón". *Revista de estudios colombinos* 3, pp. 29–60.

López, Tomás (1804). *Atlas geográfico de España, que Comprehende el Mapa General del Reyno, y los Particulares de Sus Provincias*. Madrid: Se hallará en Madrid, calle de Atocha frente á la casa de los Gremios, y en la plazuela del Ángel núm. 19, n. 3, quarto principal, junto a la Librería de Lera.

Martínez Díez, Gonzalo (1973). "La Hermandad alavesa". *Anuario de Historia del Derecho Español* 43, pp. 5–112.

Mira Caballos, Esteban (2000). *Nicolás de Ovando y los orígenes del sistema colonial español, 1502-1509*. Santo Domingo: Patronato de la Ciudad Colonial de Santo Domingo, Centro de Altos Estudios Humanísticos y del Idioma Español.

——— (2014). *La gran armada colonizadora de Nicolás de Ovando, 1501-1502*. Primera edición. Vol. 121. Academia Dominicana de la Historia. Santo Domingo: Academia Dominicana de la Historia.

Montana, Alberto (1943). *Descubrimientos, Exploraciones y Conquistas de Españoles y Portugueses En América y Oceanía*. Barcelona: Editorial Miguel A. Salvatella.

Moreno Bueno, Tomás (2008). "Breve crónica de un siglo de Catastro en España (1906-2002)". *CT Catastro* 63, pp. 31–60.

Orcástegui Gros, Carmen (1983). "La reglamentación del impuesto del monedaje en Aragón en los siglos XIII-XIV". *Aragón en la Edad Media* 5, pp. 113–122.

Pagarolas i Sabaté, Laureà (1981). *El Fogatge de 1515: Transcripció i Estudi*. Barcelona: Fundació Salvador Vives Casajuana.

Pérez Puchal, Pedro (1972). "La población del País Valenciano hasta la época estadística". *Cuadernos de Geografía* 10, pp. 1–30.

Pintado y Rivera, José (1961). *El Servicio de Millones En El Reino de Granada (1589–1591)*. Granada: Universidad de Granada.

Piquero Zarauz, Santiago (1988). "Demografía guipuzcoana en el Antiguo Régimen (siglos XVI-XVIII)". PhD thesis. Bilbao: Universidad del País Vasco-Euskal Herriko Unibertsitatea.

Polónia, Amélia (2008). "The Northwestern Portuguese Seaport System in the Early Modern Age". *Tapio Bergholm, Lewis R. Fischer and Elisabetta Tonizzi, coord.-Making Global and Local Connections: Historical Perspectives on Ports*, pp. 113–136.

Richman, Irving Berdine (1921). *The Spanish Conquerors: A Chronicle of the Dawn of Empire Overseas*. Textbook ed. Chronicles of America Series ; v. 2. New Haven: Yale University Press.

Salas Auséns, José Antonio (2008). "Cuando las fuentes nos engañan: fogajes, vecindarios y demografía (SS. XIV-XVIII)". *Aragón en la Edad Media* 20, pp. 691–708.

San-Antonio-Gómez, Carlos, C Velilla, and Francisco Manzano-Agugliaro (2011). "Tomas Lopez's geographic atlas of Spain in the Peninsular war: A methodology for determining errors". *Survey Review* 43:319, pp. 30–44.

Sasaki, Yuya and Ariell Zimran (2025). *Using Discrepancies to Correct for False Matches in Historical Linked Data*. Working Paper.

Sastre Moll, Jaume (1989). "El Impuesto Del Morabatí En El Reino de Mallorca (1300-1349)." *Acta historica et archaeologica mediaevalia* 10, pp. 159–187.

Sevillano Colom, Francisco (1973). "La Demografía de Mallorca a Través Del Impuesto Del Morabatí: Siglos XIV, XV y XVI". *Bolletí de la Societat Arqueològica Lul·liana: revista d'estudis històrics* 34:820–1, pp. 233–272.

Valldecabres Rodrigo, Rafael (2002). *El Cens de 1510: Relació Dels Focs Valencians Ordenada per Les Corts de Montsó*. Vol. 6. Fonts Històriques Valencianes. València: Universitat de València.

Viciano, Pau and Antonio José Mira Jodar (2020). "Viure d'allò seu. Rendes i gestió del patrimoni reial al regne de València a la fi de l'edad Mitjana". *Recerques* 77, pp. 103–132.