When Co-ethnicity Fails*

Alexander Kustov† Giuliana Pardelli‡

APSA Pre-print Draft: August 3, 2020

Abstract

Why do communities with larger shares of ethnic and racial minorities have worse public outcomes such as service provision? Many studies emphasize the role of diversity and other demographic variables, but the question of causality remains underexplored. We contribute to this debate by tracing the roots of contemporary racial demography and public goods provision to the uneven historical expansion of the state in Brazil. We show more remote municipalities with lower levels of state capacity in the past were more frequently selected by escaped slaves to serve as permanent settlements. Consequently, such municipalities have worse public services and larger shares of Afro-descendants today. These results indicate pervasive endogeneity of ethnic demography and public outcomes, which are both influenced by state development. Failure to account for this and other context-dependent historical confounders raises concerns over the validity of previous findings regarding the social costs and benefits of any particular demographic composition.

Keywords: Race and Ethnicity; Demographic Change; Public Goods; Historical Legacies

*The authors names appear in alphabetical order. An earlier version of the paper was presented at the APSA, EPSA, MPSA, and SPSA annual meetings, as well as the Yale political economy seminar. We would like to thank our colleagues who have read and commented on previous drafts of this article. We are especially grateful to René Flores, Egor Lazarev, and Michael Touchton. All replication materials will be available at Harvard University's Dataverse.

†Yale University, Jackson Institute for Global Affairs, alexander.kustov@yale.edu.

‡New York University Abu Dhabi, Social Science Division, giuliana.pardelli@nyu.edu.

Introduction

Why do communities with larger shares of ethnic or racial minorities have worse public outcomes, such as less efficient service provision?[1] To explain this empirical regularity, political scientists and economists have hypothesized the existence of a "diversity debit" (e.g., Habyarimana et al., 2009). Despite the scant evidence for a causal link between demographic patterns and socially inefficient outcomes, the sheer number of studies showing that diversity harms public goods provision, trust, and social cohesion suffices to convince the most skeptical reader. More recently, however, some of these findings have been challenged both empirically and theoretically. On the empirical side, scholars have demonstrated how negative effects of diversity can result from a statistical artifact that, when accounted for, shows *homogeneous* communities to have worse public outcomes than diverse ones (Kustov and Pardelli, 2018). On the theoretical side, studies have noted the potential endogeneity of this relationship by highlighting how both ethnic diversity and poor service provision result from historically weak national states (Wimmer, 2016; Singh and vom Hau, 2016).

We contribute to this growing literature by examining the causal underpinnings of the recently uncovered "homogeneity debit" and other alleged "effects" of ethnic demography *at the subnational level*. To do so, we investigate one important common determinant of local ethnic demography and public goods provision: historical levels of state capacity. Specifically, we argue that both the contemporary ethnic composition of local communities *and* their ability to provide public services, in part, trace their origins to the distribution of state presence across the national territory in the past. Accordingly, the relationship between current ethnic demography and public outcomes does not necessarily result from a causal association between these variables but stems from the fact that these are both consequences of state development.

[1] While the empirical focus of this paper is on public goods provision, we use the term "public outcomes" to emphasize that our argument is more general and can potentially be applied to other important measures of collective well-being (e.g., human development, social trust, conflict, etc.).

In our analysis, we focus on the case of Brazil and rely on a new purpose-built geocoded, historical dataset of 5,505 municipalities, including a variety of racial demography, public goods, state capacity, and economic geography variables. Overall, we show that municipalities that had a weaker state apparatus in 1920 have worse public services *and* larger shares of Afro-descendants in 2000. We then examine the relationship between the initial distribution of state presence and racial demographic change across municipalities throughout the 20th century.

To illustrate one of the important channels through which past state capacity influenced the geographic distribution of racial groups, we utilize the location of *quilombos* (communities of runaway slaves). Mirroring the fact that fugitive enslaved persons had strong incentives to self-select away from areas of strong state presence, our results show that homogeneous communities of Afro-descendants were more likely to form in remote, inaccessible areas of the country. Moreover, because these hard-to-reach communities inherited weaker state apparatuses, they have been less capable of providing public services relative to their counterparts–despite displaying comparable levels of public spending in the present.

As we demonstrate, local communities do not reach their current demographic structure and service provision levels abruptly–where localities are now depends critically on where they have been in the past. Our results thus cast doubt on the idea that Brazilian municipalities with larger shares of Afro-descendants have worse public outcomes due to intergroup coordination problems or other demographic mechanisms emphasized in the political economy literature. Rather, the findings in this paper indicate that this association is likely the product of path-dependent processes of state development, in general, and the "negative" selection of Afro-descendants into remote areas as a response to slavery, in this particular case. These results also show that the detrimental effects of slavery extend far beyond the localities that experienced it.

More broadly, our findings challenge the idea that an exogenous shift in the ethnic composition of a given community would, in and of itself, significantly change public outcomes. Ethnic diversity and other demographic compositions may indeed correlate with a variety of important social and economic indicators, but once basic structural variables are considered, the alleged causal relationship between these indicators becomes less compelling. If the current demographic makeup of localities is the outcome of historical forces deeply rooted in the territorial reach of the state and if these same forces conditioned the ability of local governments to provide public goods or promote development, the causal relationship between these two sets of variables cannot be identified without accounting for this consequential common antecedent factor.

In summary, the observed associations of contemporary ethnic demography and public outcomes may often be just that–associations. The failure to account for past state capacity and other relevant, context-specific, historical factors thus calls into question the validity of previous findings regarding the social costs and benefits of any particular (diverse or homogeneous) ethnoracial composition. In the best-case scenario, the literature that has neglected historical forces and focused exclusively on present-day covariates has overestimated the effects of ethnic demography; in the worst-case scenario, it may have mistakenly characterized a spurious relationship as causal.

Ethnoracial Demography and Public Outcomes

Three Waves of Research on Ethnic Demography

The first *descriptive* question that much of the political economy research on ethnic demography has been concerned with is whether the relationship between diversity and various public outcomes is positive or negative. Accordingly, a standard approach in this literature has been to examine local or national-level outcomes, such as social spending, as a function of (as-if exogenous) ethnic fractionalization (ELF) or similar

measures, after accounting for a number of confounding variables. Following this strategy, scholars have found support for the "diversity debit" hypothesis across a wide variety of regions and outcomes (for a review, see Stichnoth and Van der Straeten, 2013; Dinesen et al., 2020). Although this literature is understandably hesitant to make policy prescriptions, one implication of these findings is that having a homogeneous demographic structure is beneficial to societies and their governance. The absence of (potentially) politically salient societal cleavages–whether based on religion, race, or language–makes cooperation among different groups easier to sustain and social cohesion more likely to emerge (Easterly et al., 2006). The ease of intergroup coordination, in turn, makes communities more inclined to make investments with longer-term returns (Habyarimana et al., 2009).

More recently, however, the limitations of this approach have been brought to the fore. Scholars have emphasized the failure of previous work to account for the heterogeneous effects of diversity across different types of public goods (Gisselquist, 2014), units of analysis (Gerring et al., 2015), and institutional contexts (Lee et al., 2018; Charnysh, 2019). Furthermore, recent studies have problematized the adoption of the fractionalization index as the standard measure of diversity. In particular, by treating ethnic groups as equivalent, this variable fails to indicate which ones are represented in what proportions in the population and thus obscures important differences in the ways distinct groups relate to public outcomes (Rushton, 2008; Kustov and Pardelli, 2018). Altogether, what these studies have demonstrated is that even the (deceptively) simple exercise of determining the sign of the association between local demographic composition and public goods provision may produce contradictory findings, depending on the outcomes studied (Kramon and Posner, 2013), the samples considered, and the measures used.

A second *theoretical* question that scholars have grappled with revolves around the mechanisms behind the "diversity debit": why does it lead to socially inefficient

outcomes? The literature has proposed a number of potential channels to account for these negative effects. However, much like the ELF index itself, these explanations assume that ethnic groups are analogous–that is, given otherwise identical contexts, they face similar incentives and behave in comparable ways. According to the "in-group bias" mechanism, for instance, individuals benefit from the well-being of a fellow group member and attach lower (or even negative) utility to the welfare of the out-group (Alesina and La Ferrara, 2005). Although this channel helps to understand why more diverse communities may contribute less to the public welfare, it fails to clarify why one group of ethnically homogeneous localities systematically experiences worse outcomes than diverse localities. Likewise, mechanisms such as shared tastes and preferences, increased efficacy and findability, and facilitated social sanctions elucidate why more homogeneous communities might find it easier to work collectively (Habyarimana et al., 2007). However, these same mechanisms fail to explain the systematic divergence in the outcomes of equally homogeneous communities of different ethnic groups, further emphasizing the idea that groups are not interchangeable. Therefore, before identifying the precise channels through which ethnic demography affects public goods provision, some of the previously detected associations may have to be reevaluated.

Following this premise, a third stream of research has focused on directly addressing *endogeneity* concerns (e.g., see Portes and Vickstrom, 2011). Although scholars have long acknowledged the endogenous effects of ethnic demography on collective outcomes (see, e.g., Alesina and La Ferrara, 2005), few studies have taken into account the historical processes that have influenced the distribution of groups across space when trying to examine the effects of diversity (for a notable exception, see Charnysh, 2019). Because groups are rarely (if ever) randomly assigned to different territories, it is still unclear whether the relationships uncovered in previous observational research are in fact causal. One alternative interpretation is that public outcomes themselves cause particular demographic distributions to emerge. Another possibility is that a

third (omitted) variable, such as the strength of the state at the national level, has influenced both contemporary ethnic diversity and public goods provision (Singh and vom Hau, 2016). As evidenced by Wimmer (2016), states that have historically been more capable are able to provide public services more effectively while simultaneously being more successful at homogenizing their diverse populations.

This paper contributes to this growing literature by further investigating the extent to which the alleged causal relationship between ethnic demography and public outcomes can be attributed to common antecedent factors *at the subnational level*. Theoretically, the unit of analysis matters because the mechanisms at play differ from those that operate across countries (Gerring et al., 2015). Empirically, the study of subnational variation requires novel data, given that national level measures–such as Wimmer's pre-colonial state centralization–provide little information about the presence of the state across the territory (Soifer, 2008; Giraudy, 2012). Finally, investigating subnational patterns allows us to identify more precisely the ways in which the state may have influenced the initial distribution of groups across space, while keeping other important confounders, such as culture and institutional environment, constant.

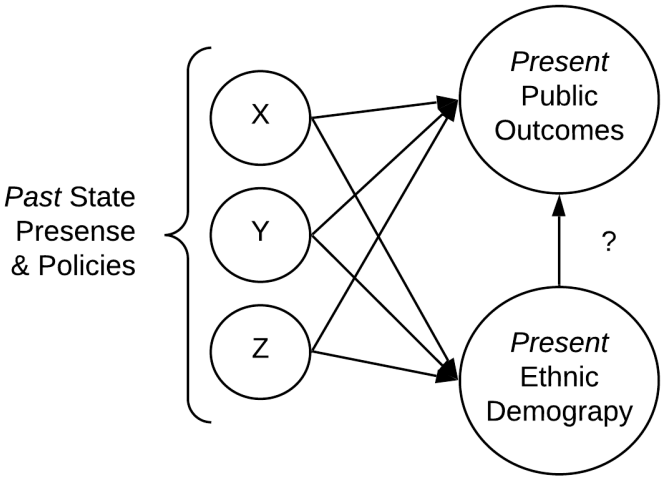
The Endogeneity of Ethnic Demography

Although the relationship between diversity and socially (in)efficient outcomes is generally described as causal, the type of demographic changes that would allow us to identify these effects are rare, making any relationship we observe between these variables likely to be spurious. Among the various historical forces that could concomitantly influence subnational variation in ethnic demography and public outcomes, in this paper, we focus on the territorial expansion of the state.

We argue that the willingness and ability of different ethnic groups to settle in specific areas across a territory can be constrained by their socioeconomic status and by the type of relationship they have with the state. Although these constraints might

be transitory, their effects are likely to unfold over time in a cumulative manner. Thus, accounting for the contemporary compositional characteristics of communities or for other important proximate factors, such as average levels of social spending, might not be sufficient to address endogeneity concerns. To examine whether the association we observe between current ethnic demography and public outcomes can be attributed to more basic structural factors (see Figure 1), we explore some of the historical forces that are known to have influenced the spatial patterns of settlement across Brazil.

Figure 1: The Endogenous Relationship of Ethnic Demography and Public Outcomes



X, Y, and Z represent the multiplicity of possible historical economic and political confounding variables that affect both ethnic demography and public outcomes.

While it is beyond the scope of this paper to build a theoretical framework that exhaustively elucidates the co-determinants of subnational ethnic demography and public outcomes, we highlight a number of factors that may play a salient confounding role. The first and one of the most consequential drivers of group-specific settlement patterns lies in *geographic and climatic factors*, which can favor certain types of economic activity and labor organization and encourage greater state penetration and population growth in some areas as opposed to others (Michalopoulos, 2012; Pardelli, 2019). Weather and crop characteristics, for example, have been shown to be significantly

associated with taxation and education spending across U.S. counties (Ramcharan, 2010). Moreover, suitability for the cultivation of cotton in certain regions of North America has been found to explain some of the spatial variation in the historical prevalence of slavery (Archarya et al., 2016). In other words, geographic characteristics have the potential to affect both the subnational distribution of ethnic groups and the levels of public goods provision.

Second, beyond the effects of climate and geography, *state policies* can differentially affect ethnic groups (i.e., engendering heterogeneous birth, death, intermarriage, and migration rates) in ways that are systematically related to the strength and penetration of the state across the territory. Most prominently, the state can deliberately relocate some ethnic groups to less developed regions through explicit "demographic engineering" measures (e.g., McNamee and Zhang, 2019). Alternatively, the adoption of less explicit directives and procedures may produce the same outcome. Governments have been shown to rely on land use regulation, for instance, to help property owners at the expense of the poor in ways that simultaneously intensify ethnic segregation and differentially affect public goods provision (Trounstine, 2018).[2]

More generally, even ostensibly "color-blind" state interventions (e.g., road construction, land reform, regional development programs, tax subsidies) may disproportionately impact some groups and, as a result, exacerbate any pre-existing ethnic and spatial disparities.[3] Whether deliberate or unintended, the overall effects of state policies are all the more consequential if we consider that even small initial compositional differences between groups (e.g., in their average income) give rise to heterogeneous responses that, in the long run, can magnify between-group disparities across multiple dimensions (see Abascal and Baldassarri, 2015). In other words, the ethnically biased effects of state policies may persist long after the policies themselves have disappeared.

[2]Another type of "demographic engineering" can result from official ethnic (re-)classification in censuses (Loveman et al., 2012) or administrative (re-)classification through border changes (Posner, 2004).

[3]State policies may also drive changes in individuals' ethnic (self-)identification, which in turn can spur shifts in *observed* demographic structures (Telles, 2014).

Although one can certainly identify multiple state actions that have had heterogeneous effects across ethnic groups and territories, we focus on local variation in the strength of the state or *state capacity* itself.[4] We argue that state capacity is one of the most important factors influencing the willingness (and ability) of specific ethnic groups to settle in certain areas of the country, especially when forced labor institutions and official policies of discrimination are in place. In doing so, we build on the growing literature conceiving of ethnic demography as a product of the same historical factors that make public goods provision more or less efficient in the present (Wimmer, 2016; Singh and vom Hau, 2016), and we further elucidate one of the mechanisms through which this process occurs. While the potential of state capacity to facilitate the provision of public goods is somewhat intuitive (Figure 2, *a*)[5], its relationship with ethnic demography outcomes (Figure 2, *b*) is less clear, as it depends on the context and level of analysis. This is what we examine next, focusing in particular on the case of Brazil.

Endogenous Racial Demography in Brazil

Deep social cleavages, along with the presence of significant social and economic differences across space, make Brazil particularly well-suited for studying the relationship between ethnoracial demography and public outcomes. Specifically, Brazil provides sufficient variation in the local predominance of racial groups to allow for clear empirical differentiation between different types of homogeneous communities. According to its last census, the country has a near-equal proportion of African and European de-

[4] According to Mann (1984, 189), the infrastructural power of the state rests in its ability to "actually penetrate civil society, and to implement logistically political decisions throughout the realm." Social scientists often refer to these capabilities as state capacity, and understand them as comprising the ability of the state to deploy authority, especially for revenue extraction and regulation of social relationships (for a review, see Berwick and Christia, 2018). Since fiscal resources constitute a necessary condition for the effective exercise of other state functions, we follow the literature in focusing on taxation as an indicator of local state capacity (Hendrix, 2010).

[5] While some scholars include public goods provision in the very definition of state capacity, it may be more fruitful to view public goods as state outputs (Lee et al., 2014).

scendants (50.74% *negros*[6] and 47.73% *brancos*), and almost as many majority white as majority black municipalities, which may display similar levels of racial diversity while having different demographic compositions (Figure A2). Furthermore, Brazilian municipalities offer a large number of comparable cases that share the same electoral rules and exhibit wide variation in public outcomes, while providing distinct historical sources of variation in racial demography and public goods provision.

As demonstrated by previous research in Brazil, more *homogeneous* Afro-descendant municipalities currently have worse public goods provision despite displaying largely similar levels of public spending as their counterparts (Kustov and Pardelli, 2018). These results highlight the limitations of intergroup behavioral strategies in explaining the patterns we observe, i.e., localities with similar levels of homogeneity but radically different outcomes. Furthermore, they emphasize the need to examine the determinants of spatial variation in racial demography and whether the latter is indeed causally related to public outcomes. To address these questions and test whether racial demography and public outcomes have common antecedent factors, we utilize the historical variation in local state capacity and its association with the subsequent trajectory of racial demography across Brazil.

Race-Based Selection into Low State Capacity and *Quilombos*

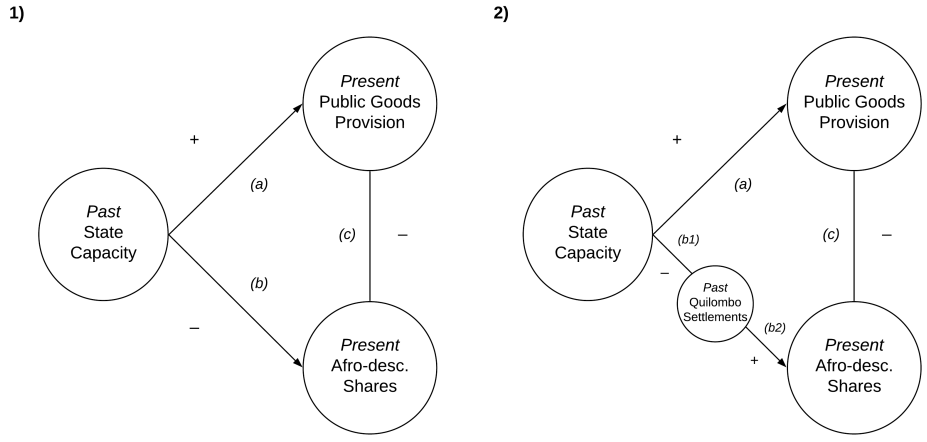
Among the numerous Afro-descendant communities that exist across Brazil's territory today, some were born from the donations of provision grounds to enslaved persons, others arose from migratory movements following the abolition of slavery, when ex-slaves fled from plantations to take up squatting claims on frontier lands. A number of the more recently formed communities emerged as a result of voluntary migration and involuntary displacement, themselves driven by a variety of forces that range from local economic growth to speculation and land grabs. Finally, many of today's majority

[6]This common classification includes both Brown (*pardos*, 43.13%) and Black (*pretos*, 7.61%) Census categories. Other categories include Asian (*amarelos*, 1.09%) and Indigenous (*indígenas*, 0.43%).

Afro-descendant communities originated in permanent settlements for escaped slaves that were established throughout the 17th-19th centuries.

In this paper, we focus on the location of runaway slave settlements, *quilombos* (or maroon communities), as one prominent mechanism through which past state capacity may have influenced both current demographic patterns and public goods provision (see Figure 2, 2).[7] In particular, we build on the fact that these autonomous settlements were established in hard-to-reach areas of low state presence precisely to avoid being "discovered and destroyed by punitive expeditions" (Bethell, 1984) (see Figure 2, *b1*). Historically weak state capacity, in turn, has made it harder for these localities to develop and provide public goods (a). At the same time, the fact that these communities were predominantly composed of escaped slaves resulted in relatively higher shares of Afro-descendant populations in the surrounding areas (*b2*). We thus argue that such race-based selection into low state capacity territories can give rise to a strong association between contemporary racial demography and public goods provision, even in the absence of a causal relationship between these variables (c).

Figure 2: Endogenous Racial Demography and Public Goods Provision in Brazil



The diagram illustrates how, as a result of past state capacity (e.g., via Quilombos), present racial demography and public goods provision can be strongly related, even in the absence of a causal link.

[7]Although every *quilombo* arguably represented a form of resistance to slavery, not all of them were established by runaway slaves. Indeed, a number of communities were formed by enslaved persons who inherited the land; others were created in territories that had been abandoned by farmers after an economic downturn; and some were established by freed slaves who purchased their own plots, invaded unclaimed lands, or received land in exchange for services rendered to the State (Moura, 2001).

Importantly, one of the reasons why *quilombola* communities were much more prevalent in Brazil than in other countries is due to the nature of its frontiers:

> "Although all slave societies had runaway communities, Brazil probably had the most numerous, longest lasting, and most widespread distribution of such quilombos (sometimes also called mocambos) communities in the Americas. Such settlements were in existence for well over a century, and others would be continually founded until the end of slavery in the late nineteenth century. The reasons for the intensity of quilombo activity in Brazil have a great deal to do with both the size of the slave labor force introduced into the country and the open nature of the frontier in all regions of plantation or slave activity. [...] Unlike the 19th-century United States, the slave zones of Brazil were neither blocked by a hostile Indian frontier nor surrounded by white agricultural settlements, but rather were accessible to open frontiers everywhere just a few miles from the coast" (Klein and Luna, 2009, pp. 196).

Predictably, these escapes provoked the fierce reaction of slave owners, who would frequently resort to local militia groups and paid mercenaries to recapture runaway slaves and destroy their communities. There is little question, however, that the assistance of the state was indispensable in safeguarding slaveowners' interests. Municipal police forces and local courts played a crucial role in guaranteeing the enforcement of contracts, while local authorities conducted investigations and organized frequent expeditions (Gomes, 2004). As a result, to increase their chances of survival, *quilombola* communities sought withdrawal as much as possible, settling in remote regions with dense forests and difficult access.[8]

In sum, due to such race-based selection into isolation, of which *quilombos* constitute but one example[9], we expect areas with lower state capacity in the past to display both greater shares of Afro-descendants and lower levels of public goods provision today.

[8] Although it made more sense for them to be isolated, cases of quilombos established in the vicinity of farms, villages, and cities are not rare. However, these "suburban quilombos" necessarily had to be mobile, as Reis (1996) notes, because the proximity of urban centers facilitated denunciation and repression. Since less remote or mobile communities were more likely to be found and destroyed by the anti-*quilombos* expeditions organized by landowners and local authorities, the distribution of *quilombos* we observe today is probably the result of both self-selection and biased attrition. The fact that more isolated communities were more likely to survive, however, does not affect our argument about the overlap of low state capacity areas and *quilombos'* locations.

[9] One can argue that freedmen, emancipated slaves, and their descendants faced similar, albeit not as stringent, constrains in their settlement choices due to discrimination, lower wealth, limited connections, and lack of access to credit. These factors likely contributed to their being systematically pushed away from more highly valued areas with greater state presence (Klein and Luna, 2009). We leave the detailed examination of these other channels for future research.

Data

We use an original dataset of 5,505 Brazilian municipalities, including contemporary and historical racial demography, public goods, state capacity, and geography variables. Given our main aim of highlighting the historical roots of the present variation in both the quality of public services and racial demography, these factors constitute our two main dependent variables. To measure racial demography (2000), we use individual-level census data and construct the (Afro-descendant) racial group shares at the level of municipalities–Brazil's smallest, politically relevant administrative unit.[10] Our main composite measure of public goods provision (2000) summarizes the population's access to basic infrastructure (e.g., piped water, electricity, sewage), as well as the quality of local education and healthcare. The public services indices are constructed using a variety of quantitative indicators that are described in further detail in the Appendix.[11]

Our first independent variable measures historical levels of local state capacity, as reflected one of the earliest records of tax revenue per capita across municipalities in 1920 (logged).[12] To test our argument about the self-selection of Afro-descendants into remote, low state capacity areas, we use the density of *quilombos*[13] within the boundaries of each municipality as our second independent variable. Our dataset includes all the communities officially recognized as *quilombola* descendants (Schwartz, 1996).[14]

[10] Our main indicator is based on the sum of two of five Census racial categories (*pardos* and *pretos*), but our results are robust to considering individual categories as separate groups. Some of our additional empirical specifications also include historical Afro-descendant shares from 1872, 1940, and 1980. Note that the 1920 Census did not include "color" or "race" categories.

[11] As a robustness check, we also use each of the individual variables that compose these indices as separate dependent variables and observe no substantive changes in the results (see the Appendix).

[12] The data were collected by Pardelli (2019) from the 1926 statistical yearbook of public finances. Because Brazilian *municípios* have changed significantly over time, we use 'area interpolation' methods to map the data from 1920 onto contemporary boundaries. The number of Brazilian municipalities increased from 1,304 in 1920 to 5,505 in 2000. Adopting an area-weighting method allows us to estimate past levels of local state capacity within modern-day localities. For details, see the Appendix.

[13] As specified in Decree 4887 (November 20, 2003) adjusting Article 68 of the Constitution, *quilombos* are defined as "ethnoracial groups, according to criteria of self-attribution, with its own historical trajectory, characterized by specific territorial relations, with the presumption of Black ancestry related to the historical resistance and endured oppression."

[14] The data were obtained from the Afro-Brazilian Communities Information System (SICAB).

To strengthen our confidence in the obtained results, we replicate our analyses using two alternative measures of state presence across the territory. In particular, we use the density of railroads[15] in 1920 (logged) and a direct measure of geographic remoteness–defined as the average travel time required to reach the nearest city from a particular municipality (see the Appendix).[16]

Our covariates include a set of geographic characteristics that can influence the ability of local governments to effectively provide public services–these include the size of the locality, altitude, rainfall, sunshine, distance from the coast, distance from the capital, latitude, and longitude (see Naritomi et al., 2012) (for summary statistics, see Table A1). All our model specifications also include state fixed effects and robust standard errors clustered at the level of 1920 municipalities, as reflected in the corresponding 'Minimum Comparable Areas' (AMCs) (see Ehrl, 2017). Given that empirical specifications with historical variables are prone to spatial autocorrelation (Kelly, 2019; Pepinsky et al., 2020), we also report Moran's I for each model. For better presentation of our results, all variables used in the analysis were standardized between 0 and 1.

Analysis and Results

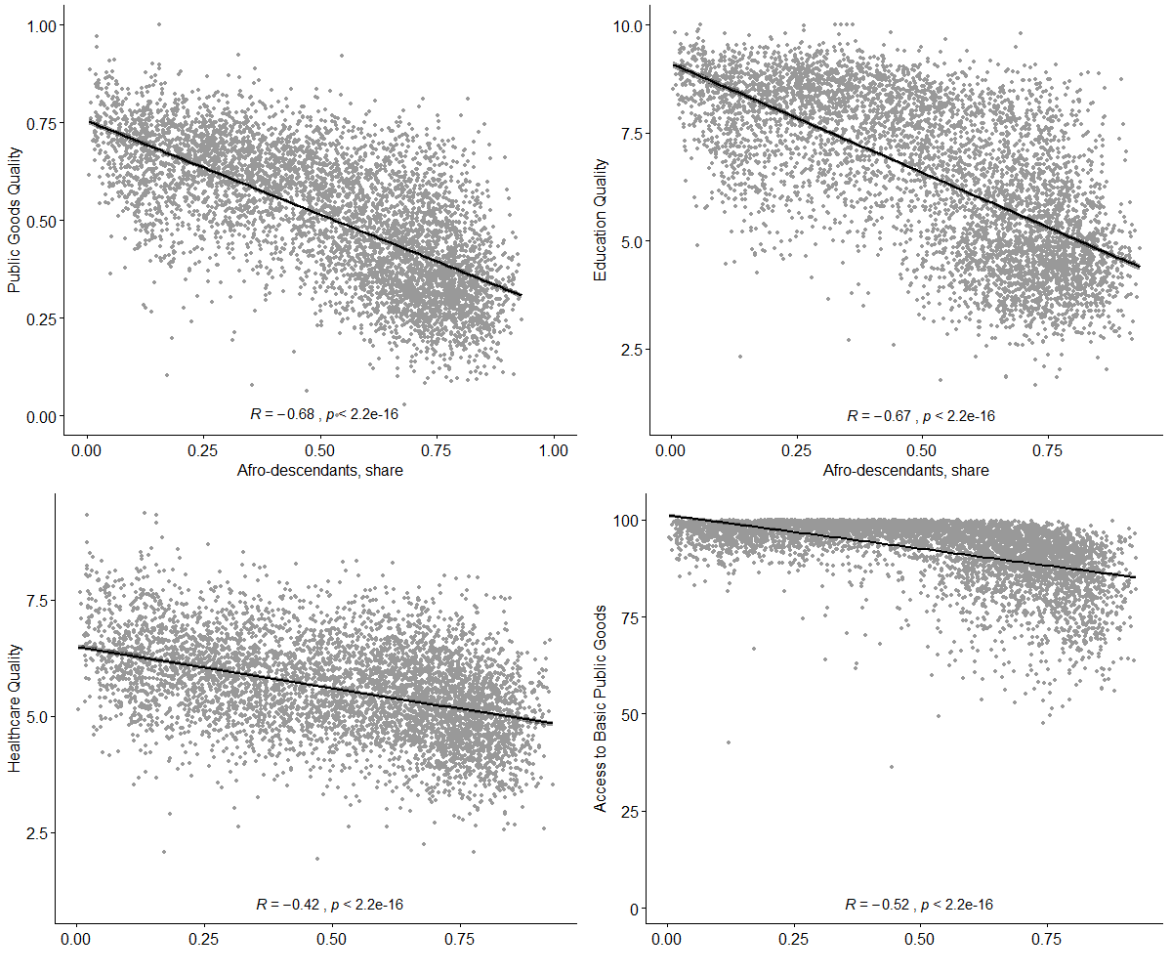
As a starting point for our analysis, we replicate the results of previous research showing the disadvantage of more homogeneous Afro-descendant municipalities in accessing public services. As shown in Figure 3 (and Table A3), the share of Afro-descendants is strongly and negatively correlated with public goods provision, explaining a staggering 20-50% of the variation in these outcome measures across Brazilian municipalities

[15] While this variable captures the relative penetration of state infrastructure across the country in the past, it may be more sensible to view it as a public good output rather than state capacity per se.

[16] The advantage of this measure, adapted from Poyart et al. (2018), is that it is based on new digital maps that take into account a number of salient geographic attributes in addition to distance, including roads, railways, rivers, water bodies, land cover types, and topography. The disadvantage is that, since it is based on present maps, it is more likely to be endogenous to past public outcomes.

(Figure 2, *c*). As previously documented, a significant part of this gap across groups remains even after accounting for all standard political economy and public finance covariates, including the level of government spending and average income levels (Kustov and Pardelli, 2018).

Figure 3: The "Homogeneity Debit" of Afro-descendant Municipalities in Brazil



Each dot represents a municipality. For variable descriptions, see the Appendix.

Our main empirical strategy is to model both present racial demography and public outcomes as a function of historical state capacity while accounting for stable geographic characteristics. Our main results in Table 1 show that tax revenues per capita in 1920 are strongly related to increased public goods provision *and* to lower Afro-descendant shares across Brazilian municipalities today. In particular, local taxation

100 years ago alone explains approximately 15-20% of the variation in these contemporary outcomes. Furthermore, these relationships persist even after we account for state fixed effects and a range of geographic factors.[17] Overall, this analysis confirms the role of past state capacity as a strong predictor of present-day outcomes, as illustrated in Figure 2 (*a* and *b*), and points to the endogeneity of the relationship between contemporary racial demography and public outcomes.

Table 1: Public Outcomes and Racial Demography as a Function of Past State Capacity

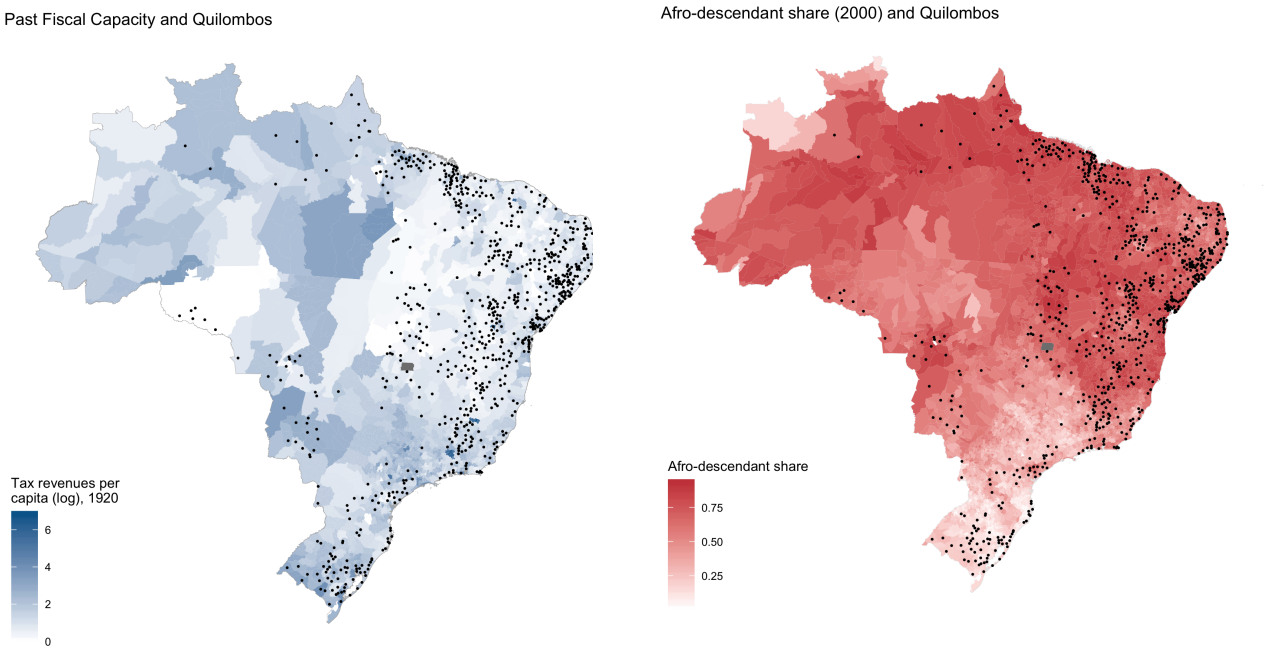
	Public Goods		Afro-descendants %	
	(1)	(2)	(3)	(4)
Tax Revenues (1920)	0.429***	0.062***	−0.921***	−0.128***
	(0.054)	(0.016)	(0.128)	(0.038)
State FE	No	Yes	No	Yes
Geographic controls	No	Yes	No	Yes
Moran's $\mathcal{I}$ Residuals	0.500***	0.002	0.558***	0.001
Observations	5,501	4,970	5,505	4,971
Adjusted R^2	0.146	0.739	0.198	0.816

All models are OLS regressions based on original data. For variable descriptions, see the Appendix.
Clustered standard errors are given in parentheses, *p<0.05; **p<0.01; ***p<0.001.
The Moran's $\mathcal{I}$ statistic indicates no spatial autocorrelation in the residuals of the specifications with the full set of covariates and FE.

While there are a number of mechanisms through which spatial variation in the strength of the state can differentially influence the demographic trajectories of ethnic groups (as outlined in our theoretical section), we focus on one that is specific to the Brazilian context, that is, the formation of *quilombola* settlements across the country. Specifically, we hypothesize that some part of the association we observe between public goods provision and Afro-descendant shares today stems from the fact that, historically, maroon communities had an incentive to self-select into hard-to-reach areas where state presence was minimal (for visualization of the relevant variables, see Figure 4).

[17] These estimates likely constitute a lower bound, given that our historical variables are less spatially disaggregated than contemporary ones due to changes in administrative boundaries over time.

Figure 4: Past State Capacity, the Location of *Quilombos*, and Racial Demography



We first test whether the locations of *quilombola* settlements are associated with lower levels of past state capacity, as measured by tax revenues in 1920–the earliest point in time for which we have national coverage and spatially disaggregated data (Figure 2, *b1*). This is indeed the case: the simple bivariate correlation between these variables is -0.22. As indicated in Table 2, this relationship continues to hold after we include state fixed effects and account for a variety of geographic factors.[18]

We then examine whether municipalities with a higher density of *quilombos* have greater shares of Afro-descendants in the population today (Figure 2, *b2*). As expected, the bivariate correlation is positive (0.24) and remains strong even after we account for the full set of additional covariates (see Table 3). In summary, Brazilian municipalities with more *quilombos* have a systematically different demographic composition that can itself be traced back to the historical distribution of state presence.

[18]Note that, while we assume that the (unobserved) initial absence of the state in a given area is causally prior to the establishment of *quilombos*, we are agnostic about the direction of the causal relationship between our empirical measures of state presence in 1920 and the presently recognized *quilombos*.

Table 2: Past State Capacity and Quilombos

	Number of Quilombos (log)	
	(1)	(2)
Tax Revenues (1920)	-0.098^{***}	-0.058^{**}
	(0.022)	(0.021)
State FE	No	Yes
Geographic controls	No	Yes
Moran's $\mathcal{I}$ Residuals	0.073^{***}	0.004
Observations	5,505	4,971

All models are OLS regressions based on original data. For variable descriptions, see the Appendix. Clustered standard errors are given in parentheses, $^{+}$p<0.1; *p<0.05; **p<0.01; ***p<0.001.
The Moran's $\mathcal{I}$ statistic indicates no spatial autocorrelation in the residuals of the specification with the full set of covariates and FE.

Table 3: Quilombos and Contemporary Share of Afro-descendants

	Afro-descendants %	
	(1)	(2)
Quilombos	0.525^{***}	0.093^{***}
	(0.059)	(0.020)
State FE	No	Yes
Geographic controls	No	Yes
Moran's $\mathcal{I}$ Residuals	0.636^{***}	-0.002
Observations	5,505	4,971
Adjusted R^2	0.055	0.815

All models are OLS regressions based on original data. For variable descriptions, see the Appendix. Clustered standard errors are given in parentheses, $^{+}$p<0.1; *p<0.05; **p<0.01; ***p<0.001.
The Moran's $\mathcal{I}$ statistic indicates no spatial autocorrelation in the residuals of the specification with the full set of covariates and FE.

Robustness of Race-based Selection Beyond *Quilombos*

These results, however compelling, raise an important question: what do the experiences of (recognized) runaway slave settlements tell us about the correlation we observe between demography and public outcomes in the broader sample of municipalities?

On the one hand, it is possible that our data only captures those *quilombos* that had a sufficiently high level of social cohesion or organizational capacity, which allowed them to seek official certification by the state. In this case, our results may underestimate the role of historical settlement choices on current outcomes for the broader sample of Afro-descendant communities. On the other hand, it is also possible that only those *quilombos* that encountered particularly critical challenges to their survival sought recognition from the state in an attempt to ameliorate their situation. If this is the case, our results using the sample of quilombo municipalities may instead overestimate the detrimental effects of self-selection into remote areas. Similarly, if we consider that being recognized as a *quilombo* allows communities to reach better economic outcomes[19], our results may underestimate the effects in the broader sample. At the same time, since some *quilombola* communities have taken active measures to preserve the past and resist urbanization, the opposite might be the case.

In practice, our data likely include cases of all these different trajectories, the relative prevalence of which cannot be easily ascertained. However, since we argue that *quilombos* is only one of the many channels through which past levels of state capacity have influenced racial demography, we can also examine the relationship of past state capacity and racial demography beyond *quilombola* communities. Accordingly, we replicate Table 1 by excluding municipalities with *quilombos* altogether with no change in our substantive results (see Table A8).

[19] As Lyons (2011, 118) emphasizes, "[r]ecognition as a quilombo is the precursor for services like roads, water, sanitation, education, and health care; becoming a quilombo bears the hope of being distinguished among the myriad poor, rural communities throughout Brazil in the eyes of mayors, municipal councils, government foundations, and non-government organizations."

As another robustness check, we also use geographic remoteness and the number of railroads within each municipality as alternative measures of state presence across the territory. As the estimates in Table A4 in the Appendix indicate, the results are statistically weaker but remain substantively unchanged. Municipalities that were more remote and inaccessible in the past have greater shares of Afro-descendants in the population and worse public outcomes today.

Finally, one may argue that there are other, more fundamental, historical factors at play affecting both state capacity and racial demography. In particular, the local prevalence of slavery may have influenced the strength and presence of the state across localities (e.g., Suryanarayan and White, 2019). Although this line of reasoning would still be consistent with our argument regarding the endogenous association between ethnic demography and public outcomes, it might lead to a different interpretation of the causal pathways at work. Therefore, we consider this possibility in detail (with no change in our substantive conclusions) by relying on the 1872 Census data on the share of (free and enslaved) Afro-descendants across municipalities in the Appendix.[20]

Persistence of Race-Based Selection Across Time

Our results suggest that municipalities where state presence was weaker in the past have historically come to have worse public goods provision and also, as a result of the hypothesized race-based selection, higher shares of Afro-descendant population. While the persistence of state capacity has been widely explored in the literature (for a review, see Dincecco, 2017), the question of whether and how racial demographic structures change thorough time remains unclear. Although this paper leaves the detailed exploration of these demographic trajectories for future research, in this section we provide some preliminary evidence regarding the stability and change of racial distributions in Brazil between 1940 and 2000 as a function of prior state capacity.

[20]Even though the slave trade ended in 1850, slavery lasted until 1888. Due to an active level of manumission and self-purchase arrangements, approximately 75% of Afro-descendants (or 43% of Brazil's total population) were free by the time of the 1872 Census. In comparison, only 6 percent of African Americans in the U.S. South were free prior to emancipation (e.g., see Schwartz, 1996).

To that end, here we examine additional data on racial demography based on our original digitization of the 1940 Population Census (complemented with data from the 1980 Census). As can be seen from Table 4 (which replicates Table 1 across 1940-2000), the share of Afro-descendants is consistently and negatively correlated with prior levels of local state capacity as captured by the tax revenues in 1920 across all available time periods (1940-2000).[21] These results also suggest that the selection of Afro-descendants into historically more remote areas peaked around the 1980s. While these static associations alone do not tell us how the initial distribution of state presence influenced the subsequent demographic structure of localities across the country, it does suggest that–despite the changes in self-identification over time and increased population mobility–these spatial demographic patterns have persisted across time.

Table 4: Racial Demography as a Function of Past State Capacity (1940-2000)

	Afro-des. % (1940)		Afro-des. % (1980)		Afro-des. % (2000)	
	(1)	(2)	(3)	(4)	(5)	(6)
Tax Revenues (1920)	-0.749^{***}	-0.065	-1.114^{***}	-0.135^{**}	-0.921^{***}	-0.128^{***}
	(0.123)	(0.049)	(0.136)	(0.050)	(0.128)	(0.038)
State FE	*No*	*Yes*	*No*	*Yes*	*No*	*Yes*
Geographic controls	*No*	*Yes*	*No*	*Yes*	*No*	*Yes*
Observations	5,503	4,970	5,505	4,971	5,505	4,971
Adjusted R^2	0.118	0.733	0.192	0.783	0.198	0.816

All models are OLS regressions based on original data. For variable descriptions, see the Appendix. Clustered standard errors are given in parentheses, $^{+}$p<0.1; *p<0.05; **p<0.01; ***p<0.001.

Along with the general population growth over the last century, the share of Afro-descendants in the median municipality has increased from 34% in 1940 to 56% in 2000. Nonetheless, during the same period, 19% of municipalities have not experience much change in their racial demographic composition, and 14% have seen a relative decrease in their Afro-descendant shares. If our argument is correct, we should expect municipalities with higher levels of state capacity in the past to experience relatively lower growth (or even an outright decline) in the share of Afro-descendants across time.

[21] Unfortunately, the reliable municipal-level data on race are not available in other Census decades.

To measure demographic change, we follow Hill et al. (2019) and construct two separate variables indicating the absolute and the relative change in the share of Afro-descendant population from 1940 to 2000 across municipalities. While the former variable simply denotes the difference in the Afro-descendant percentage points between 1940 and 2000 in a given municipality, the latter captures the relative growth rate of Afro-descendants as a fraction of the group's baseline share.[22] To account for different baseline Afro-descendant shares in 1940, we also include those as a control variable in some specifications (along with all other previously used geographic controls).

As shown in Table 5, both the absolute and the relative change in the proportion of Afro-descendant population from 1940 to 2000 is negatively correlated with state capacity levels in 1920. While this analysis cannot determine the relative role of migration and other demographic processes (including birth and death rates) behind these patterns, it suggests that state capacity per se can be an important predictor of racial demographic trajectories. As a robustness check, we also look at the relationship between past state capacity and the more proximate demographic change (1940-1980), which display coefficients of even larger magnitude (see Table A9).

Table 5: Racial Demographic Change as a Function of Past State Capacity (1940-2000)

	Abs. $\Delta(1940-2000)$			Rel. $\Delta(1940-2000)$		
	(1)	(2)	(3)	(4)	(5)	(6)
Tax Rev. (1920)	-0.172^{**}	-0.062	-0.109^{**}	1.110^{*}	-0.525	-0.808^{+}
	(0.057)	(0.046)	(0.033)	(0.514)	(0.456)	(0.417)
Afro-des. % (1940)			-0.709^{***}			-4.319^{***}
			(0.022)			(0.256)
State FE	*No*	*Yes*	*Yes*	*No*	*Yes*	*Yes*
Geographic controls	*No*	*Yes*	*Yes*	*No*	*Yes*	*Yes*
Observations	5,503	4,970	4,970	5,503	4,970	4,970
Adjusted R^2	0.015	0.324	0.654	0.007	0.311	0.448

All models are OLS regressions based on original data. For variable descriptions, see the Appendix. Clustered standard errors are given in parentheses, $^{+}$p<0.1; *p<0.05; **p<0.01; ***p<0.001.

[22] For example, if two municipalities had, respectively, 10% and 70% of Afro-descendants in 1940 and 20% and 80% in 2000, the absolute change would equal 10% (0.2-0.1 and 0.8-0.7) in both cases. However, the relative change would be 100% in the first case ($\frac{0.2-0.1}{0.1}$) but 14% in the second ($\frac{0.8-0.7}{0.1}$).

Discussion: Generalizability of the Argument

It is clear from the preceding analysis of Brazilian municipalities that the spatial distribution of racial groups and past state capacity are intimately related, with the latter accounting for a substantial part of the apparent effects of demographic patterns on contemporary public outcomes. To generalize the implications of this finding, one could argue that when a specific ethnic group is compelled to self-select away from the reach of the state, ethnic demography measures and public goods outcomes are likely to be systematically related even in the absence of an actual causal association. While the Brazilian case provides a particularly apt illustration of this phenomenon, we believe that it can also greatly inform the study of other settings. Below, we consider how our argument may apply to other types of outcomes, different contexts, and to levels of analysis beyond subnational units.

First, although this paper focuses on the association between state capacity and public goods provision in particular, our argument regarding the endogeneity of ethnic demography may be extended to other public outcomes. Another prominent example of a thesis that has attracted considerable attention and produced an abundance of research is that concerning the detrimental effects of ethnic diversity on social capital, trust, and cohesion.[23] Nonetheless, recent evidence also shows that prior investments in state capacity promote persistently higher levels of social capital (Ramey and Jensen, 2020). To the extent that ethnic diversity is itself influenced by state capacity, as demonstrated here, its previously uncovered effects on social capital may be spurious.

Second, if the logic of our argument is correct, it should be applicable to any country with significant local variation in state capacity where (1) the population

[23]Alesina and La Ferrara (2002) find that racial fragmentation has a significant negative effect on the proportion of trusting respondents in the U.S., even after controlling for inequality and individual characteristics. Stolle et al. (2008) report a negative association between diversity, and social trust and Fieldhouse and Cutts (2010) find a negative effect of neighborhood diversity on social capital (although they highlight that it is only a small fraction of the negative effect of poverty). For a recent meta-analysis corroborating these findings, see Dinesen et al. (2020). For critical reviews of this literature, see Portes and Vickstrom (2011); Abascal and Baldassarri (2015).

is differentiated along ethnic or other "visible" and "sticky" descent-based categories (Chandra, 2006); (2) there are strong status differences between groups, which coincide with durable socioeconomic and political inequalities (Tilly, 1999; Stewart, 2005); (3) at least one group has incentives to avoid areas of strong state presence, even if only temporarily. Given the countless historical episodes of intergroup conflict, state-backed discrimination, or outright subjugation of minorities around the world, there is little reason to believe that Brazil's geographic and racial inequalities constitute a unique case. A variety of historical legacies related to colonialism, slavery, feudalism, and caste may prompt the emergence of hierarchical distinctions among ethnicities that are highly persistent over time (Tilly, 1999). More generally, many countries around the world have a few main ethnic groups (Garcia-Montalvo et al., 2005) with a high degree of inequality between them (Baldwin and Huber, 2010).

Finally, our argument relies on mechanisms that are directly tied to subnational contexts, where boundaries tend to be more permeable than those at the national level. Nonetheless, future studies can examine whether, for example, stronger states tend to exhibit more restrictive immigration policies, which in turn systematically affect the willingness and ability of certain groups to migrate there (e.g., Rosenberg, 2019), confounding any association we might observe between diversity and public outcomes across countries. Overall, while our study does not rule out that ethnic demography can have an independent effect on important public outcomes such as service provision or social capital, it challenges the empirical findings of a vast body of work investigating contemporary associations while failing to account for fundamental historical forces.

Conclusion

In this paper, we revisit the prominent notion that ethnoracial diversity and public outcomes are causally related. We present evidence that questions the validity of such

claims and show that the observed negative association between the contemporary share of Afro-descendant population and public outcomes across Brazilian municipalities is in part due to a common antecedent factor: the uneven distribution of state presence more than 100 years ago. In fact, as a consequence of both the poverty and oppression that resulted from slavery, freedmen and runaway slaves were constrained in their movements across the territory. They had to seek and settle in isolated areas with weak state presence where they had a chance of forming viable frontier communities. Hence, the performance of these 'remote' majority-black localities today does not match that of their neighbors, which benefited from stronger state infrastructure early on, despite displaying similar levels of government spending.

Our results underline the role of historical legacies (e.g., slavery, state policies, group settlement choices) in shaping both the contemporary demographic composition and public outcomes of municipalities. Although we do not rule out that ethnic demography may be consequential and have an independent causal effect in certain contexts, our results highlight that many of the empirical findings in the literature likely overestimate the effects of diversity and other demographic variables on public outcomes.[24]

Our findings also speak to the literature that emphasizes the endogeneity of linguistic diversity and state capacity across countries (e.g., Wimmer, 2016). In particular, our argument aligns with the notion that ethnic demography is a consequence of state strength. However, it also complements this literature by elucidating the mechanisms that operate across *subnational* (rather than national) entities and in relation to *racial* (rather than linguistic) cleavages, which are arguably less susceptible to change.

In addition to questioning previous theoretical assumptions, the patterns uncovered here have practical implications for future research. They emphasize the need for scholars to consider the historical forces that have influenced the distribution of ethnic

[24] A priori, the true causal relationship between ethnic demography and public outcomes (when present) may be even stronger after we account for the relevant historical forces. However, this is unlikely to be the case in most empirical settings for the reasons described in the section above.

groups across the geography of the state and whether this spatial organization has persisted over time. Moreover, our results speak to the literature on the legacies of slavery (Nunn, 2008; Archarya et al., 2016; Dell and Olken, 2019). As indicated by our additional analysis, in some settings, the long-term effects of slavery may be obscured by the presence of significant countervailing effects. In fact, across Brazil, municipalities that had higher shares of slaves in the 1870s were also likely to experience massive inflows of immigration in the post-abolition period. As a result, the shares of enslaved do not add much analytical leverage in explaining long-term outcomes in this case.

Our paper is not without limitations. First, our measure of past state capacity may itself be endogenous. While we do corroborate our findings with additional measures of geographic remoteness, railway density, and the location of former slave communities, it is possible that local state capacity in 1920 was itself affected by racial politics (e.g., see Suryanarayan and White, 2019). Second, and related, our data do not allow us to explain how past levels of state capacity persisted over time or to identify all the ways in which it affected racial demography patterns–a number of different mechanisms are likely to be simultaneously at play.[25] Nonetheless, these limitations do not challenge our main argument regarding the biased distribution of ethnic groups across space, and the resulting pervasive endogeneity of ethnic demography and public outcomes.

One particularly consequential mechanism that may be explored in future research is endogenous administrative boundary changes. As a result of both federal reforms and local democratic initiatives, the number of municipalities in Brazil has increased from 642 in 1872 to 5,565 in 2010. Our preliminary evidence indicates that when new administrative boundaries are created, the "offspring" municipalities tend to be more racially homogeneous than the "parent" ones (not shown). Regardless of whether these effects are unintended or deliberate, they may help elucidate why the newly

[25] As discussed above, there are numerous ways in which greater capacity may have perpetuated and even exacerbated initial disparities (through prohibitively high taxes, land regulation, designation of protected areas, discouraged or subsidized immigration, etc.) that disproportionately affect one group.

created, more homogeneous Afro-descendant areas remain disadvantaged relative to more diverse ones.

Overall, this paper shows that the alleged effects of demographic variables can be, at least in part, a product of their histories. This suggests that previous claims about the benefits associated with ethnic homogeneity are likely overstated. More generally, our findings call for a greater focus on the role of the state in creating and reinforcing historical disparities between ethnic groups. Investigating the ways in which present outcomes depend upon past choices is therefore a critical task for future observational studies positing ethnic demography as an important determinant of social outcomes. Our findings can thus aid scholars in elaborating novel theoretical frameworks that delineate the scope conditions of previous theories and identify the specific mechanisms that might operate in different contexts.

References

Abascal, Maria and Delia Baldassarri (2015). Love Thy Neighbor? Ethnoracial Diversity and Trust Reexamined. *American Journal of Sociology 121*(3), 722–782.

Alesina, Alberto , Reza Baqir, and William Easterly (1999). Public goods and ethnic divisions. *The Quarterly Journal of Economics 114*(4), 1243–1284.

Alesina, Alberto and Eliana La Ferrara (2002). Who trusts others? *Journal of Public Economics 85*(2), 207–234.

Alesina, Alberto and Eliana La Ferrara (2005). Ethnic Diversity and Economic Performance. *Journal of Economic Literature 43*(3), 762–800.

Archarya, Avidit , Matthew Blackwell, and Maya Sen (2016). The Political Legacy of American Slavery. *Journal of Politics 78*(3), 621–641.

Baldwin, Kate and John D. Huber (2010). Economic versus Cultural Differences: Forms of Ethnic Diversity and Public Goods Provision. *American Political Science Review 104*(04), 644–662.

Berwick, Elissa and Fotini Christia (2018). State Capacity Redux: Integrating Classical and Experimental Contributions to an Enduring Debate. *Annual Review of Political Science 21*(1), 1–21.

Bethell, Leslie (Ed.) (1984). *The Cambridge History of Latin America Volume 2: Colonial Latin America*. Cambridge: Cambridge University Press.

Chandra, Kanchan (2006). What Is Ethnic Identity and Does It Matter? *Annual Review of Political Science 9*(1), 397–424.

Charnysh, Volha (2019). Diversity, Institutions, and Economic Outcomes: Post-WWII Displacement in Poland. *American Political Science Review*.

Dell, Melissa and Benjamin A Olken (2019). The Development Effects Of The Extractive Colonial Economy: The Dutch Cultivation System In Java. *The Review of Economic Studies*.

Dincecco, Mark (2017). *State Capacity and Economic Development Present and Past*. New York: Cambridge University Press.

Dinesen, Peter Thisted , Merlin Schaeffer, and Kim Mannemar Sønderskov (2020). Ethnic Diversity and Social Trust: A Narrative and Meta-Analytical Review. *Annual Review of Political Science 23*(1).

Easterly, William , Jozef Ritzen, and Michael Woolcock (2006). Social Cohesion, Institutions, and Growth. *Economics and Politics 18*(2), 103–120.

Ehrl, Philipp (2017). Minimum comparable areas for the period 1872-2010: An aggregation of Brazilian municipalities. *Estudos Economicos 47*(1), 215–229.

Fieldhouse, Edward and David Cutts (2010). Does Diversity Damage Social Capital? A Comparative Study of Neighbourhood Diversity and Social Capital in the US and Britain. *Canadian Journal of Political Science 43*(02), 289–318.

Garcia-Montalvo, J , M Reynal-Querol, and J Montalvo (2005). Ethnic Polarization, Potential Conflict, and Civil Wars. *American Economic Review 95*(3), 796–816.

Gerring, John , Strom C. Thacker, Yuan Lu, and Wei Huang (2015). Does diversity impair human development? A multi-level test of the diversity debit hypothesis. *World Development 66*, 166–188.

Giraudy, Agustina (2012). Conceptualizing state strength: Moving beyond strong and weak states. *Revista De Ciencia Política 32*(3), 599–611.

Gisselquist, Rachel M. (2014). Ethnic divisions and public goods provision, revisited. *Ethnic and Racial Studies 37*(9), 1605–1627.

Gomes, Flávio dos Santos (2004). Slavery, black peasants and post-emancipation society in Brazil (nineteenth century Rio de Janeiro). *Social Identities 10*(6), 735–756.

Habyarimana, James , Macartan Humphreys, Daniel N. Posner, and Jeremy M. Weinstein (2007). Why Does Ethnic Diversity Undermine Public Goods Provision? *American Political Science Review 101*(4), 709–725.

Habyarimana, James , Macartan Humphreys, Daniel N. Posner, and Jeremy M. Weinstein (2009). *Coethnicity: diversity and the dilemmas of collective action*. Russell Sage Foundation.

Hendrix, C.S. (2010, may). Measuring state capacity: Theoretical and empirical implications for the study of civil conflict. *Journal of Peace Research 47*(3), 273–285.

Hill, Seth , Daniel J. Hopkins, and Gregory Huber (2019). Local demographic changes

and US presidential voting, 2012 to 2016. *Proceedings of the National Academy of Sciences*.

Kelly, Morgan (2019). The Standard Errors of Persistence. *SSRN Working Paper (dx.doi.org/10.2139/ssrn.3398303)*.

Klein, Herbert S. and Francisco Vidal Luna (2009). *Slavery in Brazil*. Cambridge: Cambridge University Press.

Kramon, Eric and Daniel N Posner (2013). Who Benefits from Distributive Politics? How the Outcome One Studies Affects the Answer One Gets. *Perspectives on Politics 11*(02), 461–474.

Kustov, Alexander and Giuliana Pardelli (2018). Ethnoracial Homogeneity and Public Outcomes: The (Non)effects of Diversity. *American Political Science Review 112*(4), 1096–1103.

Lee, Melissa M. , Walter-Drop, and John Wiesel (2014). Taking the State (Back) Out? Statehood and the Delivery of Collective Goods. *Governance 27*(4), 635–654.

Lee, Neil , Katy Morris, and Thomas Kemeny (2018). Immobility and the Brexit vote. *Cambridge Journal of Regions, Economy and Society* (March).

Loveman, Mara , Jeronimo O. Muniz, and Stanley R. Bailey (2012). Brazil in black and white? Race categories, the census, and the study of inequality. *Ethnic and Racial Studies 35*(8), 1466–1483.

Lyons, Colleen M. Scanlan (2011). Spaces of Silence and Efforts toward Voice: Negotiation and Power Among Quilombo' Communities in Southern Bahia Brazil. *Afro-Hispanic Review 30*(2), 115–132.

Mann, Michael (1984). The autonomous power of the state: its origins, mechanisms and results. *European Journal of Sociology 25*(2), 185–213.

McNamee, Lachlan and Anna Zhang (2019). Demographic Engineering and International Conflict: Evidence from China and the Former USSR. *International Organization 73*(02), 291–327.

Michalopoulos, Stelios (2012). The origins of ethnolinguistic diversity. *American Economic Review 102*(4), 1508–1539.

Moura, C (2001). *Os quilombos na dinâmica social do Brasil*. EdUFAL.

Naritomi, Joana , Rodrigo R. Soares, and Juliano J. Assunção (2012). Institutional Development and Colonial Heritage within Brazil. *The Journal of Economic History 72*(2), 393–422.

Nunn, Nathan (2008). Slavery, Inequality, and Economic Development in the Americas: An Examination of the Engerman-Sokoloff Hypothesis. In E. Helpman (Ed.), *Institutions and Economic Performance*, pp. 148–180. Harvard University Press.

Pardelli, Giuliana (2019). *Financing the State: Inequality and Fiscal Capacity in Uneven Territories*. Dissertation, Princeton University.

Pepinsky, Thomas , Sara Wallace Goodman, and Conrad Ziller (2020). Does Proximity to Nazi Concentration Camps Make Germans Intolerant? Modeling Spatial Heterogeneity and Historical Persistence.

Portes, Alejandro and Erik Vickstrom (2011). Diversity, Social Capital, and Cohesion.

Posner, Daniel N. (2004). The Political Salience of Cultural Difference: Why Chewas and Tumbukas Are Allies in Zambia and Adversaries in Malawi. *American Political Science Review 98*(4), 529–545.

Poyart, E. , S. Bhatt, D. J. Weiss, H. S. Gibson, J. Rozier, B. Mappin, S. Peedell, A. Lieber, S. Y. Kang, W. Temperley, L. S. Tusting, K. E. Battle, E. Cameron, T. C. D. Lucas, M. Hancher, D. Bisanzio, N. Fullman, R. E. Howes, A. Nelson,

P. W. Gething, S. Belchior, and U. Dalrymple (2018). A global map of travel time to cities to assess inequalities in accessibility in 2015. *Nature 553*(7688), 333–336.

Ramcharan, Rodney (2010). Inequality and redistribution: Evidence from U.S. counties and states, 1890-1930. *Review of Economics and Statistics 92*(4), 729–744.

Ramey, Adam J and Jeffrey L Jensen (2020). Early investments in state capacity promote persistently higher levels of social capital. *Proceedings of the National Academy of Science*.

Rosenberg, Andrew S (2019). Measuring Racial Bias in International Migration Flows. *International Studies Quarterly 63*(4), 837–845.

Rushton, Michael (2008). A note on the use and misuse of the racial diversity index. *Policy Studies Journal 36*(3), 445–459.

Schwartz, Stuart B. (1996). *Slaves, peasants, and rebels: reconsidering Brazilian slavery*.

Singh, Prerna and Matthias vom Hau (2016). Ethnicity in Time: Politics, History, and the Relationship between Ethnic Diversity and Public Goods Provision. *Comparative Political Studies 49*(10), 1303–1340.

Soifer, Hillel (2008). State infrastructural power: Approaches to conceptualization and measurement. *Studies in Comparative International Development 43*(3-4), 231–251.

Stewart, Frances (2005). Horizontal inequalities: a neglected dimension of development. In UNU-WIDER (Ed.), *Wider Perspectives on Global Development*, pp. 101–135. Palgrave Macmillan.

Stichnoth, Holger and Karine Van der Straeten (2013). Ethnic diversity, public spending, and individual support for the welfare state: A review of the empirical literature. *Journal of Economic Surveys 27*(2), 364–389.

Stolle, Dietlind , Stuart Soroka, and Richard Johnston (2008). When does diversity erode trust? Neighborhood diversity, interpersonal trust and the mediating effect of social interactions. *Political Studies 56*(1), 57–75.

Suryanarayan, Pavithra and Steven White (2019). Slavery, Reconstruction, and Bureaucratic Capacity in the American South.

Telles, Edward E (2014). *Pigmentocracies: Ethnicity, Race, and Color in Latin America*. Chapel Hill: University of North Carolina Press.

Tilly, Charles (1999). *Durable inequality*. Berkeley: University of California Press.

Trounstine, Jessica (2018). *Segregation by Design*. Cambridge: Cambridge University Press.

Wimmer, Andreas (2016). Is diversity detrimental? Ethnic fractionalization, public goods provision, and the historical legacies of stateness. *Comparative Political Studies 49*(11), 1407–1445.

Online Appendix

Explanatory and Outcome Variables

Racial Demography

All demographic variables were constructed using microdata from the 2000, 1980, 1940, and 1872 Censuses from the Instituto Brasileiro de Geografia e Estatística (IBGE).

Public Goods

Our composite measure of public goods is an average of municipal education quality, healthcare quality, and access to basic public goods. The quality measures concern education and healthcare provision. These indices come from a comprehensive study conducted by Arretche, Fusaro and Vaughan in the Centro de Estudos da Metrópole (CEM) and take into account a set of more than 10 indicators per sector to build two overall measures of quality (one for education and one for health) at the municipal level.[26] Access to basic public goods indicates the proportion of the population that has access to garbage collection, electricity, sewage and piped water. These variables are built using microlevel data available in the 2000 Census.

Human Development Index

The municipal index is calculated as a geometric mean of income per capita, expected school years, and life expectancy. All the data come from the Instituto de Pesquisa Economica Aplicada (IPEA).

[26] For instance, the health indicator comprises, among other variables, infant mortality rate, hospitalization rate and vaccine coverage. The education index consists of coverage below 6 years of age, failure and abandonment rates, proportion of municipal schools with below-average grades on the national standardized test, etc. For more detailed information, see http://web.fflch.usp.br/centrodametropole.

Historical State Capacity

The data on tax revenues per capita and railroad number (logged) across municipalities in 1920 is taken from Brazil's 1926 statistical yearbook of public finances (Pardelli, 2019). To estimate the past levels of state capacity within modern-day administrative boundaries, we adopt the approach used by Archarya et al. (2016) to address changes in U.S. county boundaries between 1860 and 2000. Following this method, the total amount of taxes per capita collected in 1920 is divided among the relevant municipalities in 2000 such that the proportion of taxes from 1920 municipality i that is assigned to 2000 municipality j is based on the size of their overlapping areas (see Figure A1).

Geographic Remoteness

Geographic remoteness is defined as the average travel time required to reach the nearest city (>50000 people) via surface transport in a particular municipality. For details on methodology and calculations, see Poyart et al. (2018).

Geographic controls

The data come from the National Institute of Geology (INGEO) and include *Area size (logged)*, *Altitude*, *Distance to Coast (logged)*, *Distance from the Capital (logged)*, *Rainfall*, *Sunshine*, *Latitude* and *Longitude*. For details, see Naritomi et al. (2012).

Slavery Prevalence as a Competing Endogeneity Source

First, we examine whether historical shares (1872) of Afro-descendants in the population are related to contemporary ethnic demography measures (2000). Although the overall association is positive, the correlation is surprisingly weak (0.3) and, as can be seen from Table A5, it is not statistically significant when we include standard geographic covariates.[27] Furthermore, only the share of *free* Afro-descendants is positively associated with current Afro-descendant shares, whereas the proportion of *enslaved* Afro-descendants has a negative relationship. Finally, none of these historical demographic variables are predictive of present public goods outcomes.

Second, we investigate whether racial demography in 1872 is associated with the (temporally more proximate) levels of state capacity in 1920. The raw correlation between Afro-descendant shares and tax revenues per capita 50 years later is -0.2 (see Table A6), but it does not hold after including state fixed effects and geographic controls. The breakdown of the Afro-descendant population into free and enslaved reveals two opposing associations. On the one hand, the intensity of slavery has a *positive* (albeit weak) relationship with past state capacity. On the other hand, the proportion of free Afro-descendants is *negatively* related to local tax revenues.

Although at first glance these findings may seem counterintuitive, they are consistent with the series of migratory movements that followed the abolition of slavery. The significant negative association between the percent of enslaved persons in the population and the current proportion of Afro-descendants may be explained by two concurrent forces. First, as historians have often highlighted, once slavery was abolished, former slaves fled from plantation areas, leaving the estates en masse and occupying frontier lands (Klein and Luna, 2009). Second, this wholesale abandonment of farms by ex-slaves engendered an acute labor shortage, which planters sought to resolve by

[27] Archarya et al. (2016) find, for instance, the U.S. shares of Blacks in 1860 and 2000 have a correlation of 0.77.

resorting to a massive influx of European immigrants. Consequently, the areas where slavery was more prevalent in the last decades of the nineteenth century were precisely those that tended to become more racially diverse in later years. This may thus explain why the prevalence of slavery in 1872 does not correlate the tax revenues in 1920 (Table A6, column 3) or more contemporary public goods outcomes (Table A5, column 6).

But what about the areas with high shares of free Afro-descendants in 1872? These areas likely include the remote localities in which runaway slaves established their communities and also the rural areas in which Brazil's large freedmen population lived. Although the latter were not subject to the same necessity of runaway slaves to self-select into hidden, hard-to-reach regions, free Afro-descendants were also constrained in their choices of where to settle. This was to be expected given their ultimately poorer background and lack of capital. This is reflected in the negative association between this variable and the level of state capacity in the early twentieth century (Table A6, column 2) and in the fact that these communities are more likely to have remained majority-black to this day (Table A5, column 2). This interpretation is further corroborated by the results shown in Table A7, where we examine the relationship between the 1872 demographic variables and the concentration of *quilombos*. Namely, localities with a greater share of free Afro-descendants have a larger number of quilombos, whereas those where slavery was more prevalent have a lower number of such communities.

In summary, while the institution of slavery has undoubtedly had strong and lasting detrimental impacts on the nation, its role as a common antecedent factor of both contemporary *local* demographic structures and public goods outcomes is more ambiguous. In particular, the local prevalence of slavery is not only deeply intertwined with subsequent inflows of immigrants–which made localities more diverse–but it also coincides spatially with some of the zones that received the largest amounts of investment in subsequent decades. We hope future work will help elucidate the long-term consequences of these countervailing forces.

Tables and Figures

Figure A1: **Municipal Boundaries, 1920 and 2000**

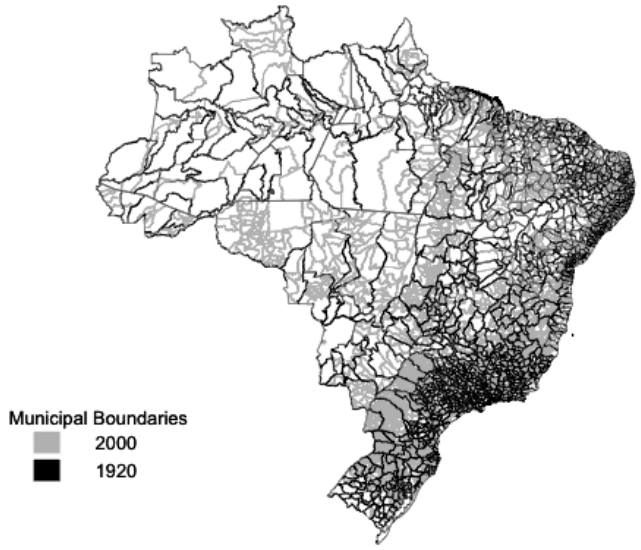
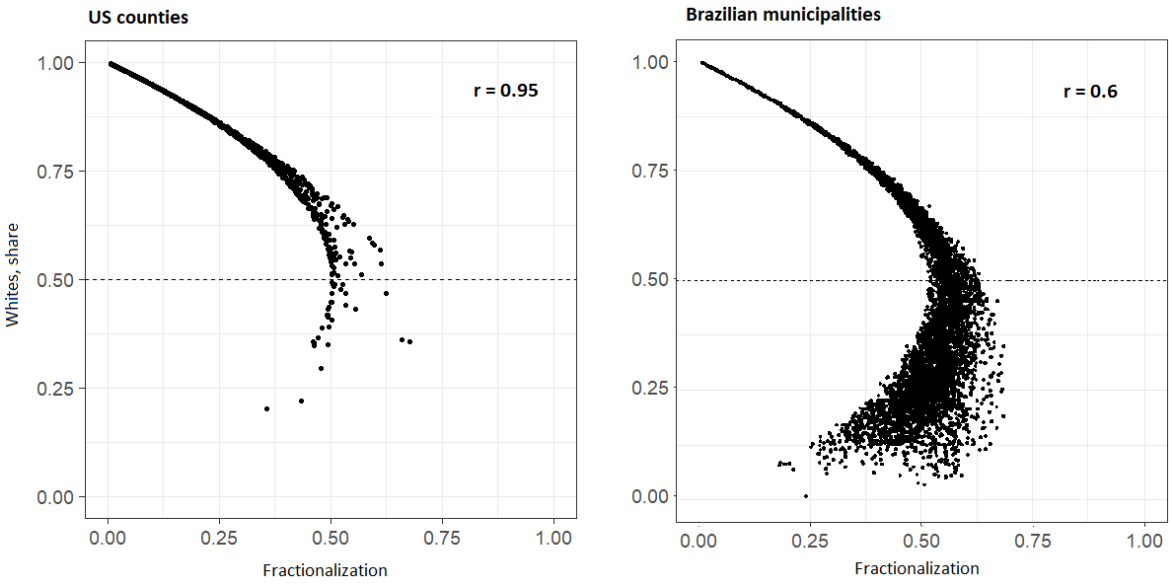


Figure A2: Racial Demography across Brazilian Localities (Compared to the U.S.)



Each dot represents racial demography in terms of fractionalization or group shares. The graph is based on the Brazil data from Kustov and Pardelli (2018) and the U.S. data from Alesina et al. (1999).

Table A1: Summary Statistics

Statistic	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Afro-descendants % (2000)	0.52	0.24	0.004	0.31	0.72	0.93
Afro-descendants % (1980)	0.48	0.29	0.001	0.19	0.75	1.00
Afro-descendants % (1960)	0.36	0.25	0.005	0.12	0.57	0.93
Education Quality	6.50	1.81	1.65	4.87	8.12	10.00
Healthcare Quality	5.57	1.00	1.92	4.89	6.24	9.35
Access to Basic Goods	92.29	7.99	36.21	89.20	98.08	100.00
Income PC (logged)	6.08	0.50	4.57	5.63	6.48	7.62
School Years (Expected)	9.46	1.10	4.34	8.74	10.21	12.83
Life Expectancy	73.07	2.68	65.30	71.13	75.15	78.64
Tax Revenues PC (logged)	1.24	0.79	0	0.6	1.7	7
Number of Quilombos (logged)	1.11	1.38	0.00	0.00	1.95	5.79
Railroad Number (logged)	0.25	0.47	0.00	0.00	0.32	4.68
Geographic Remoteness (logged)	3.99	1.01	0.00	3.47	4.53	8.28
Area Size (logged)	6.20	1.28	1.06	5.32	6.94	11.99
Altitude	4.12	2.93	0.00	1.53	6.32	16.28
Rainfall	12.80	3.96	3.00	12.00	15.00	33.00
Sunshine	21.33	3.26	12.00	18.00	24.00	30.00
Distance to Coast (logged)	5.37	1.33	0.003	4.76	6.24	7.91
Distance to Capital (logged)	5.27	0.87	0.00	4.81	5.88	7.30
Longitude	−46.17	6.40	−72.90	−50.79	−41.35	−34.81
Latitude	−16.40	8.27	−33.69	−22.80	−8.43	4.60
Afro-descendants % (1872)	0.55	0.18	0.03	0.41	0.67	0.96
Free Afro-descendants % (1872)	0.40	0.20	0.02	0.23	0.55	0.91
Enslaved Afro-descendants % (1872)	0.14	0.09	0.003	0.08	0.18	0.57

Full Sample (n = 5505)

Table A2: Public Goods Outcomes and Past State Capacity

	Education		Health		Access to Basic Goods	
	(1)	(2)	(3)	(4)	(5)	(6)
Tax Revenues (1920)	0.691***	0.065*	0.079+	−0.098***	0.412***	0.136***
	(0.093)	(0.028)	(0.041)	(0.027)	(0.045)	(0.020)
State FE	No	Yes	No	Yes	No	Yes
Geographic controls	No	Yes	No	Yes	No	Yes
Observations	5,505	4,971	5,505	4,971	5,501	4,970
Adjusted R^2	0.135	0.732	0.004	0.320	0.144	0.550

All models are OLS regressions based on original data. For variable descriptions, see the Appendix. The standard errors are given in parentheses, $^{+}$p<0.1; *p<0.05; **p<0.01; ***p<0.001.

Table A3: Public Goods Outcomes and Afro-descendant Shares (2000)

	Education		Health		Access to Basic Goods	
	(1)	(2)	(3)	(4)	(5)	(6)
Afro-descendants %	−0.606***	−0.348***	−0.240***	−0.122***	−0.271***	−0.109***
	(0.041)	(0.025)	(0.016)	(0.027)	(0.016)	(0.027)
State FE	No	Yes	No	Yes	No	Yes
Geographic controls	No	Yes	No	Yes	No	Yes
Observations	5,505	4,971	5,505	4,971	5,501	4,970
Adjusted R^2	0.446	0.758	0.180	0.325	0.269	0.550

All models are OLS regressions based on original data. For variable descriptions, see the Appendix. The standard errors are given in parentheses, $^{+}$p<0.1; *p<0.05; **p<0.01; ***p<0.001.

Table A4: Public Outcomes and Racial Demography (Robustness Checks)

	Public Goods		Afro-descendants %	
	(1)	(2)	(3)	(4)
Railroads (1920)	0.050	0.029^{*}	-0.078	-0.094^{**}
	(0.040)	(0.015)	(0.100)	(0.034)
Geographic Remoteness	-0.426^{***}	-0.088^{***}	0.717^{***}	0.075^{+}
	(0.050)	(0.020)	(0.087)	(0.045)
State FE	No	Yes	No	Yes
Geographic controls	No	Yes	No	Yes
Observations	5,207	4,711	5,208	4,711
Adjusted R^2	0.179	0.738	0.144	0.813

All models are OLS regressions based on original data. For variable descriptions, see the Appendix. The standard errors are given in parentheses, $^{+}$p<0.1; *p<0.05; **p<0.01; ***p<0.001.

Table A5: Past Racial Demography, Present Afro-Descendant Shares and Public Goods

	Afro-descendants % (2000)			Public Goods		
	(1)	(2)	(3)	(4)	(5)	(6)
Total Afro-descendants % (1872)	0.051^{+}			-0.014		
	(0.026)			(0.009)		
Free Afro-descendants % (1872)		0.073^{***}			-0.010	
		(0.020)			(0.009)	
Enslaved Afro-descendants % (1872)			-0.163^{***}			0.007
			(0.043)			(0.016)
State FE	Yes	Yes	Yes	Yes	Yes	Yes
Geographic controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4,955	4,955	4,955	4,954	4,954	4,954
Adjusted R^2	0.815	0.816	0.817	0.739	0.739	0.739

All models are OLS regressions based on original data. For variable descriptions, see the Appendix. The standard errors are given in parentheses, $^{+}$p<0.1; *p<0.05; **p<0.01; ***p<0.001.

Table A6: Slavery, Past Racial Demography and State Capacity

	Tax revenues per capita (1920)		
	(1)	(2)	(3)
Total Afro-descendants % (1872)	0.001		
	(0.021)		
Free Afro-descendants % (1872)		−0.036*	
		(0.016)	
Enslaved Afro-descendants % (1872)			0.056*
			(0.028)
State FE	Yes	Yes	Yes
Geographic controls	Yes	Yes	Yes
Observations	4,955	4,955	4,955
Adjusted R^2	0.464	0.467	0.466

All models are OLS regressions based on original data. For variable descriptions, see the Appendix. The standard errors are given in parentheses, $^{+}$p<0.1; *p<0.05; **p<0.01; ***p<0.001.

Table A7: Past Racial Demography and Quilombos

	Quilombos		
	(1)	(2)	(3)
Total Afro-descendants % (1872)	0.044***		
	(0.012)		
Free Afro-descendants % (1872)		0.058***	
		(0.014)	
Enslaved Afro-descendants % (1872)			−0.094**
			(0.031)
State FE	Yes	Yes	Yes
Geographic controls	Yes	Yes	Yes
Observations	4,955	4,955	4,955
Adjusted R^2	0.161	0.164	0.162

All models are OLS regressions based on original data. For variable descriptions, see the Appendix. The standard errors are given in parentheses, $^{+}$p<0.1; *p<0.05; **p<0.01; ***p<0.001.

Table A8: Public Goods Outcomes and Racial Demography as a Function of Past State Capacity (excluding *Quilombo* Municipalities)

	Public Goods (2000)		Afro-descendants % (2000)	
	(1)	(2)	(3)	(4)
Tax Revenues (1920)	0.419***	0.051**	−0.912***	−0.126**
	(0.054)	(0.016)	(0.130)	(0.041)
State FE	No	Yes	No	Yes
Geographic controls	No	Yes	No	Yes
Observations	4,721	4,251	4,723	4,252
Adjusted R^2	0.145	0.738	0.196	0.807

All models are OLS regressions based on original data. For variable descriptions, see the Appendix. Clustered standard errors are given in parentheses, *p<0.05; **p<0.01; ***p<0.001.

Table A9: Racial Demographic Change as a Function of Past State Capacity (1940-1980)

	Abs. $\Delta(1940-1980)$			Rel. $\Delta(1940-1980)$		
	(1)	(2)	(3)	(4)	(5)	(6)
Tax Rev. (1920)	-0.365^{***}	-0.070	-0.111^{*}	-0.618^{+}	-0.842^{**}	-1.062^{**}
	(0.049)	(0.052)	(0.045)	(0.326)	(0.327)	(0.333)
Afro-des. % (1940)			-0.633^{***}			-3.358^{***}
			(0.028)			(0.240)
State FE	*No*	*Yes*	*Yes*	*No*	*Yes*	*Yes*
Geographic controls	*No*	*Yes*	*Yes*	*No*	*Yes*	*Yes*
Observations	5,503	4,970	4,970	5,503	4,970	4,970
Adjusted R^2	0.051	0.324	0.524	0.004	0.182	0.332

All models are OLS regressions based on original data. For variable descriptions, see the Appendix. Clustered standard errors are given in parentheses, $^{+}$p<0.1; *p<0.05; **p<0.01; ***p<0.001.