Local majorities: How administrative divisions shape comparative development*

Richard Bluhm[†] Roland Hodler[‡] Paul Schaudt[§]

October 2021

Abstract

We study the role of subnational borders and the importance of local majorities for local economic development. We exploit that France imposed a particular administrative structure on its Sub-Saharan African possessions in the early 20th century. The French government had little interest in pre-colonial political units. As a result, their colonial districts cut across ethnic homelands in a way that led to plausibly exogenous variation in an ethnic group's population share across colonial districts. We find that ethnic groups who were a local majority in most colonial districts, in which they were present, are more economically developed today. Furthermore, we show that the parts of ethnic homelands with a higher district-level population share are more economically developed today than other parts of the same homeland. We also provide evidence that the effects are persistent for various reasons, including the stickiness of subnational borders and higher infrastructure investments during colonial times.

Keywords: Ethnic politics, local majorities, administrative borders, colonialism, regional development, persistence.

JEL classification: D72, F54, H54, H75, N97, O10, R12, R50, Z13.

*We are thankful for helpful comments by Matthias Thoenig, Stelios Michalopoulos, Samuel Bazzi, Gedeon Lim, Miriam Artiles, David Weil, Kai Gehring, Christian Ochsner, Charles Gottlieb; conference participants at PACDev, European Meeting of the UEA, EPCS, German Development Conference, Swiss Development Network, 3rd International Conference on Globalization and Development, EEA; and seminar participants at the University of Bergen, Brown University, University of Fribourg, and University of California San Diego. Richard Bluhm has benefited from generous financial support by the Alexander von Humboldt foundation for this project and thanks the University of California San Diego for hospitality during parts of this research.

†Institute of Macroeconomics, Leibniz University Hannover; e-mail: *bluhm@mak.uni-hannover.de*

‡Department of Economics, University of St. Gallen; CEPR, London; CESifo, Munich; e-mail: *roland.hodler@unisg.ch*

§Department of Economics, University of St. Gallen; e-mail: *paul.schaudt@unisg.ch*

1. Introduction

We lack a good understanding of how administrative borders shape the economic success of different ethnic, religious, or cultural groups. The main empirical challenge is that politically powerful and economically successful groups typically influence the design these borders, making causal identification difficult. As a result, most of what we know is based on the—often arbitrary—drawing of country (or protectorate) borders by the former colonial powers in Sub-Saharan Africa (e.g., Alesina et al., 2011). For example, recent studies document that ethnic groups split by an international border are mired in conflict (Michalopoulos and Papaioannou, 2016). Conversely, groups with a sizeable population exercise more political power and are more economically successful today (Francois et al., 2015; Dimico, 2017; Hodler and Raschky, 2017). Subnational borders are rarely studied. However, from a policy perspective, it is particularly interesting to understand the link between subnational borders and comparative development precisely because these borders can be changed more easily.

In this paper, we study how subnational borders shape the economic fortunes of ethnic groups. We exploit variation in local group shares resulting from the early administrative-territorial structures that the European colonial powers imposed on their colonies and protectorates in Sub-Saharan Africa.[1] We obtain these structures from official maps drawn from around 1900 to the 1920s, when Europeans drew the first administrative borders in Africa.[2] The borders within French colonial Africa are particularly suitable for our purpose. Contrary to the British, the French had little interest in pre-colonial political units when dividing their protectorates into districts. Instead, they relied on a centralized structure and direct rule. As a result, the extent to which colonial district borders crosscut an ancestral homeland of a given size was as good as random in French protectorates. We exploit this plausibly exogenous variation to study how historical local majorities shape long-run economic outcomes both across and within ethnic homelands.

Our primary units of analysis are the ancestral homelands of ethnic groups within each protectorate and a partition of these homelands into small grid cells of 5×5 arc minutes (corresponding to 9.3×9.3 km at the equator). To measure the influence of local majorities, we first compute the population share of every group in each colonial district. We then compute the expected population share of co-ethnics living in the same colonial district as a randomly chosen member of each group. This share is high if most members of an ethnic group lived in a colonial district where the group was in the majority, but low if no or few group members were in the majority in their colonial district. This measure

[1] For simplicity, we subsequently use the terms "protectorates" and "colonies" interchangeably.

[2] The Berlin conference of 1884–1885 established the principle of effective occupation, which stipulated that rights over colonial lands required effective occupation or possession of these lands. As a consequence, the Scramble for Africa accelerated and the European colonizers claimed almost all Sub-Saharan African territory by 1900. However, it typically took them another decade or two to establish an administrative-territorial structure in their protectorates (Crowder, 1968).

summarizes the experience of ethnic groups across ethnic homelands. Within the 13 former French protectorates of Sub-Saharan Africa, balance tests based on a large set of geographical, historical, and group-specific variables confirm that the expected population share of an ethnic group is as good as random (conditional on group size). Leveraging this natural experiment, we find that differences in the expected population share on the onset of colonization can explain contemporary differences in economic development, as measured by nighttime light emissions or survey data across ethnic homelands. Moreover, we show that the district-level population share of an ethnic group can explain differences in comparative development both within homelands and within colonial districts.

Our focus on subnational borders sets this paper apart from much of the literature. Studies of the consequences of European colonization for African societies typically focus on colonial policies and practices that are immutable.[3] However, many African countries have reformed their administrative-territorial structures since independence, and these reforms routinely change the composition of ethnic groups within and across subnational units. We exploit this variation to test if local majorities matter for current economic development because subnational borders are persistent and current local majorities matter, or because local majorities during colonial times directly affect current economic development. We find strong evidence for persistence in subnational borders and, therefore, in local majorities at the level of both ethnic homelands and grid-cells. Moreover, when running a horse race between colonial and current (expected) population shares, we find their effects are very similar in magnitude. These results suggest that the effect of local majorities is persistent but that administrative-territorial reforms can weaken the link between colonial history and current development outcomes.

Going one step further, we show that the historical presence of local majorities benefits both those residing in ancestral homelands today and groups identifying with these homelands elsewhere in the country. Using data from the Demographic and Health Surveys (DHS), we provide evidence that people residing in the ancestral homelands of local majorities are wealthier and better educated than people living elsewhere. Moreover, we find that people identifying with the ethnic group of those homelands are wealthier, more educated, and have lower infant mortality rates than people identifying with other groups (independent of their place of residence within the country). These results have important implications for potential channels through which local majorities in colonial times impact current economic development. They suggest that there are multiple channels, some of which are place-based and others identity-based. Regarding place-based channels, we find evidence that ethnic groups and homelands with local majorities

[3] See Michalopoulos and Papaioannou (2020) for a recent overview of this literature. Some of these contributions focus on the effects of specific colonial institutions and politics (e.g., Acemoglu et al., 2001, 2002; Michalopoulos and Papaioannou, 2014; Ali et al., 2019; Müller-Crepon, 2020) or colonial investments (e.g., Huillery, 2009; Jedwab and Moradi, 2016; Jedwab et al., 2017; Ricart-Huguet, 2020) which are closely related to our paper.

in French colonial Africa are more likely to host the colonial district capitals and, hence, more likely to benefit from economies of agglomeration and colonial infrastructure investments. They also have access to a better health infrastructure today. Turning to people residing in those homelands, we document that majority groups seem to have better utilized early investments into education and health. Moreover, we find strong evidence that people sort in space and that members of majority groups, on average, do so towards better locations than minority groups.

We contribute to several strands of the literature on ethnic politics. First, we complement the recent literature on the importance of the size of an ethnic group for its role in national politics. Francois et al. (2015), Dimico (2017), and Hodler and Raschky (2017) find that national-level population shares matter for the composition of the cabinet, the allocation of public goods, and economic development.[4] We show that the size of an ethnic group within an administrative unit matters for local economic development both across and within ethnic homelands. Second, we contribute to the recent literature analyzing differences in the contemporary development of ethnic groups (e.g. Michalopoulos and Papaioannou, 2013, 2014, 2016; Alesina et al., 2016) by showing that local majorities cause differences in the economic success of these groups. Third, more indirectly, we also contribute to the old debate about the optimal design of the territorial structure in ethnically segregated countries. Some scholars argue that creating ethnically homogenous administrative units reduces ethnic tensions and leads to a more efficient provision of public goods (e.g., Lijphart, 1977; Horowitz, 1985; Alesina et al., 1995). Others suggest that administrative borders should deliberately cut across ethnic homelands to create regional cleavages that differ from ethnic cleavages (e.g., Lipset, 1960). Both of these arguments rely on assumptions about the distribution of power within ethnically divided subnational units. We contribute to this literature by establishing that local majorities matter.

The remainder of the paper is structured as follows. Section 2 discusses the administrative design of French colonies. Section 3 introduces our data on colonial districts, measures of local majorities, and proxies for economic development. Section 4 present our identification strategy and main results. Section 5 examines the persistence of subnational borders and studies the role of current and past local majorities. Section 6 uses survey data to study whether the main benefits of local majorities are place-based or identity-based. Section 7 investigates several mechanisms. Section 8 provides evidence for the external validity of our results. Section 9 concludes.

[4] These findings may well be the result of a common pattern characterizing the national politics of many Sub-Saharan African countries: citizens vote along ethnic lines (e.g., Horowitz, 1985; Banerjee and Pande, 2007; Huber, 2012; Ichino and Nathan, 2013; Bluhm et al., 2021), and political leaders and cabinet members favor their co-ethnics (e.g., Posner, 2005; Franck and Rainer, 2012; Burgess et al., 2015; De Luca et al., 2018; Dickens, 2018; Widmer and Zurlinden, 2019; Beiser-McGrath et al., 2020).

2. Administrative-territorial structures in French colonial Africa

France followed a centralized and direct rule approach within its protectorates in Sub-Saharan Africa (Crowder, 1964; Herbst, 1989). This approach had two key elements. First, the French showed little respect for ancestral homelands when dividing their protectorates into colonial districts (*cercles*) of similar populations or areas. The resulting districts "frequently cut across pre-colonial political boundaries" (Crowder, 1964, p. 199), leading to "the break up of traditional political units into smaller units, or the amalgamation of disparate smaller units into large groups so that there was some uniformity in the political units to be administered" (Crowder, 1968, p. 175). This cross-cutting of ancestral homelands lead to administrative units that had "no roots in tradition" (Crowder, 1968, p. 191). Prominent examples include the Fouta Djallon and Kissi groups in French Guinea (the Republic of Guinea after independence). The Imamate of Fouta Djallon was a centralized theocratic state which was broken up into several districts (Crowder, 1968), while the independent villages of the Kissi were grouped into arbitrarily combined units (Conklin, 1997). It was even at some point discussed whether villages could serve as the primary administrative entity to completely break the power of larger chiefdoms (Crowder, 1964).

The second key element of the French approach was that the traditional leaders were placed "in an entirely subordinate role of the political officer" (Crowder, 1964, p. 199). All power derived from the *commandant de cercle* who was simultaneously the head of government, the primary judicial authority, and the chief of police. Defiant chiefs were arrested, exiled or killed. Former kings were demoted and at best kept on as religious figure heads. As a result, "people became terrified of the *commandant*" (Edwards and Roberts, 1986, p. 338). The Governor General of French West Africa, Van Voellenhoven, summarized this view in a circular on the position of the chiefs in 1917. In this document, he emphasized that the traditional chiefs posses no personal power of any kind and are only an instrument of the state (Edwards and Roberts, 1986). Moreover, the allocation of districts officers in the early period was based primarily on vacancy, not selection, which introduced another component of randomness (Cohen, 1971).

All colonial powers initially had limited knowledge of the areas they aimed to govern, especially when it concerned areas further away from the coast. As a result, the colonizers tried, by and large, to follow geographical markers if such markers existed and were known to them.[5] However, historians and geographers agree that ethnic groups in Sub-Saharan Africa had been less clearly divided by geography than ethnic groups in Europe, leading to unintentional cross-cutting when the colonizers used natural markers to draw borders.

[5] Zartman (1965) documents that even a quarter of national boundaries established during the scramble for Africa follow natural markers, such as rivers.

Bening (1984), for example, documents that even in cases where colonizers used natural markers, such as rivers, in present day Ghana, they often split ethnic groups.[6] We study how "natural" colonial district borders and the borders of ancestral homelands were in different colonies in Online Appendix C. We find that colonial district borders geographical markers more frequently than the boundaries of ethnic homeland. Moreover, we document that district borders located within ethnic homelands tend to be more natural than the homeland borders encompassing them.[7] This tells us that exogenously determined geography played a large role in how the colonial powers approached the first territorial designs in Sub-Saharan Africa.

In summary, the centralized administrative-territorial structure that French bureaucrats implemented in the form of *cercles* and their all powerful *commandants* deliberately broke pre-colonial institutions was uniformly imposed across all French possessions south of the Sahara (the two federations known as *Afrique-Occidentale française*, or AOF, and *Afrique-Équatoriale française*, or AEF, as well as Madagascar). Moreover, Europeans seem to have followed natural markers when creating administrative divisions, which often cut across historical boundaries of ethnic groups. In other words, territorial structure imposed in the French colonial empire scrambled the pre-colonial power dynamics between different ethnic groups.

3. Data and measurement

In this section, we first present our database of the early administrative-territorial structures and the corresponding colonial district borders. We then introduce our measures of local majorities and our primary proxy for economic development. Other data and measures are introduced when they are first used in the analysis.

3.1. Colonial district and ethnic homelands

There are no authoritative data on the spatial extent of subnational administrative units (districts or provinces) in Sub-Saharan Africa during the colonial period. To fill this gap, we collected and digitized historical maps containing protectorate and district borders from libraries and archives worldwide.

We use Geographic Information Systems (GIS) to geo-reference these maps and digitize the relevant districts. We have taken various steps to ensure our geo-referenced maps are of the highest possible quality and precision. First, we primarily rely on high-

[6]In other cases different groups claimed dominion over the same area between two rivers, as in the Dagomba-Mamprusi land dispute (Bening, 1986).

[7]We also find no systematic deviations from the use of natural borders for groups that become more often local majorities compared to other, suggesting that the use of natural borders was not strategic with respect to general group characteristics (see Online Appendix C for details).

resolution military maps and maps from the colonial survey offices, which we complement with colonial records and historical atlases. Second, whenever possible, we use the latest published map containing the earliest known colonial districts. Third, we have verified each district's history and its borders within a protectorate (see Online Appendix D for a list of the source materials). The result is a comprehensive spatial database of colonial districts across Sub-Saharan Africa for the period between 1900 and the 1920s (with the earliest maps for some countries dating back to the 1890s).

Figure I offers an illustration. Panel A shows an administrative map of the Colony and Protectorate of French Guinea, published by the Colonial Office of French West Africa in 1922. Panel B shows the colonial districts (their borders and names) that we have extracted from this map.

Figure I about here.

Our data include 40 Sub-Saharan African protectorates and a total of 578 colonial districts. Panel A of Table I provides summary statistics on the area and historical population of all protectorates, and singles out the 13 French protectorates.[8] France was the first colonizer of eight protectorates in French West Africa, four protectorates in French Equatorial Africa, and Madagascar.[9] Panel B provides summary statistics for the colonial districts across the French and all Sub-Saharan Africa protectorates. We see that French colonial districts were less populous and had considerably less variation in their population size compared to all colonial districts, supporting the notion that France implemented a relatively uniform administrative-territorial structure across its colonies.

Table I about here.

A key premise underlying our analysis is that colonial districts often cross-cut pre-colonial ethnic homelands. Hence, we need a map of ancestral homelands prior to colonization. We use Murdock's *Ethnolinguistic Map* (1959), which shows the spatial distribution of ethnic groups across Africa around 1900.[10] Panel C of Figure I shows the ethnic homelands of French Guinea, while Panel C of Table I provides summary statistics for all ancestral homelands across Sub-Saharan Africa and those in French colonies.[11] In

[8]The historical population is computed using the History Database of the Global Environment (HYDE, version 3.2) by Klein Goldewijk et al. (2010), which will be discussed in more detail shortly. Table D-2 in Section D-1 lists all the protectorates in our sample and documents their first colonizer (and also the second colonizer if there was a change). The map in Panel A of Figure A-1 shows the districts by first colonizers.

[9]French Somaliland (known as Djibouti since independence) was a French colony as well, but is omitted from our analysis because the French established districts only much later.

[10]Alternative ethnographic maps focus on later periods, e.g., the maps by the Ethnologue (Eberhard et al., 2020) or GREG (Weidmann et al., 2010) or are available for just a single or a few countries rather than all of Africa. Murdock coded many different characteristics, e.g., pre-colonial centralization, for more than half of these ethnic groups. The map is widely used in the literature on colonialism in Africa (see e.g., Nunn, 2008; Michalopoulos and Papaioannou, 2013, 2016, and many more since)

[11]Figure A-1 in Section A-1 shows all ethnic homelands included in our sample.

French colonies, homelands tend to be smaller than districts, both in terms of their area and population, so many districts will consist of more than one territory traditionally claimed by an ethnic group.

3.2. Measuring local majorities

We now introduce two measures of local majorities. For this purpose, let us denote the historical population of ethnic group e in colonial district d of protectorate p by N_{edp}. Hence, the protectorate-wide population of group e and district d are $N_{ep} = \sum_d N_{edp}$ and $N_{dp} = \sum_e N_{edp}$, respectively.

The population share of group e in district d is $PS_{edp} = N_{edp}/N_{dp}$. We view this share as a good proxy for group e's district-level majority status. One could think of alternative, more direct proxies of local majorities, e.g., indicators for whether group e is the largest group in district d or whether its population share exceeds 50 percent. We prefer the district-level population share, PS_{edp}, for three reasons. First, data on the historical population distribution is imperfect, and small changes in N_{edp} only lead to small changes in PS_{edp}, but potentially large changes in these binary indicators. Second, the population share of the largest group may matter, i.e., it may make a difference whether its share is 60% or 90%. Third, differences in population sizes of minority groups may matter as well (say 40% vs. 5%).

To measure local majorities at the level of ethnic groups, we use the expected population share of co-ethnics living in the same district as a randomly chosen member of group e. The probability that a randomly chosen member of group e lives in a particular district d is equal to N_{edp}/N_{ep}. Aggregating over districts, we can write the expected population share as

$$EPS_{ep} := \sum_d \frac{N_{edp}}{N_{ep}} PS_{edp} = \sum_d \frac{(N_{edp})^2}{N_{ep}N_{dp}}. \quad (1)$$

The sole building block of our measures is N_{edp}, i.e., the historical population of ethnic group e in colonial district d of protectorate p. We empirically measure N_{edp} by the estimated historical population residing in the spatial intersection of group e's homeland and colonial district d. This approach is based on the assumption that ethnic homelands were ethnically homogenous at the time of colonization.[12] We use our map of early colonial district borders and Murdock's ethnolingusitic map to identify all homeland-district intersections. To compute the historical population in these areas, we use population density estimates for 1900 from the History Database of the Global Environment (HYDE, version 3.2) by Klein Goldewijk et al. (2010), which provides historical information on population density for grid cells of 0.5×0.5 arc minutes (corresponding to 9.3×9.3 km

[12]This assumption may be violated for many homelands. We return to this issue in our robustness tests and investigate areas in French colonial Africa, where we suspect this assumption to be violated in a systematic manner.

near the equator).[13]

Let us consider a single colony – French Guinea – to illustrate the computation of our measures for local majorities. Figure II shows the ancestral homelands (with borders and names in black) and colonial districts (with borders and names in blue) in French Guinea.

Figure II about here.

Panel A illustrates the district-level population share PS_{edp} of group e in each district d. This share varies across homeland-district intersections. Looking at the country's west, we see, e.g., that the Fouta Djalon had a higher population share in Koumbia district than in Boko district. Panel B illustrates the group-level expected population share EPS_{ep}, which only varies across ethnic homelands. We see, e.g., that this share is higher for the Fouta Djalon than the Koranko. The reason is that many members of the Fouta Djalon lived in a district (e.g., Koumbia or Labe) where their population share was large, while the Koranko were a minority in all districts in which they lived.

Averaging across all the homeland-district combinations, we find that the population share PS_{edp} of ethnic groups within their colonial district is 0.24 in French protectorates and 0.21 in the entire Sub-Saharan Africa sample (with the corresponding standard deviations being 0.30 and 0.29). The homeland-level expected population share EPS_{ep} is on average 0.28 in French protectorates and 0.22 in the entire Sub-Saharan Africa sample (with the corresponding standard deviations being 0.25 and 0.26).[14]

3.3. Proxying for economic development

We use nighttime light emissions to proxy for economic development at the subnational level. The underlying idea is that most forms of consumption and production in the evening require light and that public infrastructure too is often lit at night. Henderson et al. (2012) and Hodler and Raschky (2014) indeed find a high correlation between nighttime light emissions and GDP at the level of countries and provinces, respectively.[15] The main advantages of nighttime lights data are that they are available for all locations across Sub-Saharan Africa and have high spatial resolution, which allows us to proxy for economic development at various levels of spatial aggregation.

We primarily use the 2015 nighttime lights data from the Visible Infrared Imaging

[13]We validate the quality of the HYDE's population density estimates in Online Appendix B. In particular, we show that the HYDE population data are useful to predict the following three types of historical population data: ethnic group-level population estimates by Murdock, administrative data for French West Africa for 1935, and census data for British East Africa from 1948.

[14]On average, EPS_{ep} is higher than PS_{edp}. The reason is that EPS_{edp} is based on the expected population share of a randomly selected member of group e and that most group members live in districts where the group is relatively well represented, i.e., where PS_{edp} is relatively high.

[15]Bruederle and Hodler (2018) find positive correlations between nighttime lights and various measures of human development even at the very local level.

Radiometer Suite (VIIRS) sensors rather than the more commonly used data from the Defense Meteorological Satellite Program-Operational Linescan System (DMSP-OLS). The VIIRS data are considerably more accurate than the DMSP data in several dimensions, e.g., they have little overglow and improved low and top light detection.[16]

4. Local majorities and development

4.1. Empirical strategy and identification

For our main specification, the units of analysis are ethnic homelands or, more precisely, ethnic homeland-protectorate intersections.[17] We use the following specification to estimate the effect of the expected population share of an ethnic group in the colonial districts where its members lived on current economic development within the ancestral homeland of a group:

$$\ln \textsc{Lights}_{ep} = \beta\, EPS_{ep} + FE_p + \mathbf{z}'_{ep}\boldsymbol{\gamma} + \epsilon_{ep}, \tag{2}$$

where $\ln Lights_{ep}$, is the natural logarithm of nighttime light density in a given ethnic homeland. To avoid losing observations with reported nighttime light emissions of zero, we follow the literature in adding 0.01 before taking logs (e.g., Michalopoulos and Papaioannou, 2013; Hodler and Raschky, 2014). We usually include protectorate fixed effects FE_p, which control for all potential confounders at the protectorate (or country) level, as well as control variables $\mathbf{z}_{ep}$ for the geography within homelands and the pre-colonial history of each group.[18] We cluster the standard errors at the level of ethnolinguistic families (called culture groups in Murdock's ethnolinguistic map).[19]

The identifying assumption is that the distribution of colonial districts across ethnic homelands and therefore our continuous treatment (EPS_{ep}) are conditionally independent of a group's potential outcomes. As discussed earlier, this is ex-ante plausible because "the French made little concession to indigenous political units in dividing up their African territories for administrative purposes" (Crowder, 1964, p. 203). The French split many ethnic homelands and often formed colonial districts combining fragments of various homelands. The degree to which colonial district borders cut across different ethnic homelands was primarily the product of chance. The territories of some ethnic groups

[16]The disadvantage of the VIIRS data is that it is available for a shorter time period (2015–2019) than the DMSP-OLS data (which is available from 1992–2013). This potential disadvantage is inconsequential for us, as we exploit spatial rather than inter-temporal variation.

[17]We subsequently use the term homeland to imply homeland-protectorate intersections. That is, we treat ethnic homelands that overlap, say, two protectorates as two different homelands.

[18]Section D-2 provides detailed descriptions and the sources of all control variables.

[19]This follows other studies investigating differences in contemporary outcomes across homelands (e.g., Michalopoulos and Papaioannou, 2013, 2016).

were cut so frequently that most of its members were in the minority in the colonial district where they ended up living. Other similarly-sized groups were "lucky" in the sense that most of their members belonged to the majority in the colonial districts where they lived. There is one important non-random element: populous groups occupying large territories were naturally more likely to be in the majority in at least some of the resulting district(s) located inside their traditional homelands (see Michalopoulos and Papaioannou, 2016, for similar results with respect to probability of being separated by an international border). Therefore, we have good reasons to believe that the variation in EPS_{ep} across ethnic groups within French protectorates has been randomly assigned conditional on the size and population of a group.

Table II presents balancing tests supporting this notion. We systematically check whether an ethnic group's expected population share across colonial districts is independent of the geography of the homeland and its pre-colonial history.

Table II about here.

There are three important findings: First, as expected, an ethnic group's expected population share increases in both the historical population density of a group and the area of its homeland. Second, we find no evidence suggesting that EPS_{ep} is systematically related to a large set of geographical variables: the presence of rivers and lakes, elevation, ruggedness, crop suitability for agriculture, distance to the coast, the estimated malaria burden, and tsetse fly suitability.[20] Third, EPS_{ep} also appears to be unrelated to variables capturing important aspects of the local history prior to colonization: the presence of pre-colonial conflict, the prevalence of slave trade, the existence of early European explorer routes, and the presence of pre-colonial cities or pre-colonial kingdoms.[21] In addition, Table A-1 in Online Appendix A confirms balancedness along settlement patterns as well as the extent of class stratification and political centralization around the time of colonization (using data from Murdock, 1967, which is available for slightly more than half of the ethnic groups in French colonial Africa).[22] We conclude that the sample of ethnic groups and homelands in French colonial Africa is balanced among all these important geographical and historical dimensions. Hence, these results agree with the historical narrative about French administrative design in colonial Africa and support our research design.

[20]These geographical variables are all based on spatial data and aggregated at the homeland level. Section D-2 provides detailed descriptions and the sources of all the variables used in Table II.

[21]There is one exception: the coefficient on pre-colonial conflicts is negative and statistically significant at the 10 percent level in column (3), but not when controlling for local geography in column (4).

[22]Table A-2 and Table A-3 in Online Appendix A enter all these variables separately and confirm that none of them has an effect on EPS_{ep} when controlling for historical population density and area.

4.2. Across homeland evidence

We now turn to our main results on the long-run effect of an ethnic group's expected district-level population share, our proxy for local majorities, on current economic development in its homeland. Table III presents our estimates of eq. 2 phasing in the full set of controls and protectorate fixed effects. Columns (1) only includes our key controls (i.e., log of historical population density, log of area, and dummy variables for the presence of rivers and lakes), column (2) adds all control variables used in the balance tests, and column (3) includes protectorate fixed effects. Column (4) uses all control variables as well as the protectorate fixed effect.

Table III about here.

We find a positive, statistically significant effect of EPS_{ep} on nighttime light intensity throughout all specifications. Moreover, the point coefficients are relatively similar across specifications, highlighting the uniform approach with whom France created its administrative-territorial structure across its vast colonial empire south of the Sahara. The coefficient estimate in column (4), the most restrictive specification, implies that an increase in EPS_{ep} by one standard deviation (0.28) increases nighttime light emissions in homelands of former French protectorates by around 64 percent. Henderson et al. (2012) and Hodler and Raschky (2014) study the relation between nighttime light emissions and GDP at the level of countries and provinces, respectively. They both report an elasticity of around 0.3. Assuming the same elasticity at the level of ethnic homelands and the new VIIRS data, this increase in nighttime light itensity corresponds to an increase in local GDP by around 19 percent.[23] We conclude that ethnic groups whose members were part of local majorities in most of the colonial districts where they lived are more economically developed today. This effect is both, statistically and economically significant.

Robustness tests: The cross-homeland results are not sensitive to changes in the specification, measures, or sample. We find qualitatively unchanged effects if we discretize the treatment into simple and absolute majorities (see Table A-4). Our results do not depend on the specific source of nighttime lights data (see Table A-5) or, as we explore further below, whether we use nighttime lights as a development outcome at all.

We also find that the effect of local majorities is not just a function of national group shares (see Table A-6), even though there is a high correlation between a group's local

[23]These effects are also economically meaningful compared to other prominent determinants explaining differences in development across homelands. For example, Michalopoulos and Papaioannou (2013) report that a one-standard deviation increase in pre-colonial hierarchies is associated with an increase in nighttime light emissions of about 0.12 standard deviations, while our own results imply that a one-standard deviation increase in EPS_{ep} is associated with an increase in nighttime light emissions of about 0.3 standard deviations.

EPS_{ep} and the national population share (0.69).

Given the mechanical correlation between a group's population or the size of its homeland and its status as a local majority, we also investigate if our results are driven by our log-transformed controls of historical population and area. In Table A-7, we choose a semi-parametric approach to control for historical population density and area. We use percentiles to construct dummy variables for ten population density bins and ten area bins, respectively. The results remain qualitatively unchanged.

We also find no evidence that our results are driven by a particular set of districts or homelands. In Table A-8, we test the robustness of our results to four changes in our sample. First, we base our measure (EPS_{ep}) only on diverse districts, i.e., we disregard the effect of homogenous districts. Second, we omit ethnic homelands hosting the country's colonial national capital. Third, we omit ethnic homelands hosting the country's current national capital. Fourth, we omit ethnic homelands that are split by country borders (as in Michalopoulos and Papaioannou, 2016). We find that the effect size is relatively stable throughout these sample perturbations. These four sets of excluded districts arguably represent areas where our assumption that ethnic homelands were homogenous during colonization is most likely to be violated. Therefore, we interpret the relatively stable effect sizes as evidence that this homogeneity assumption is not particularly problematic. Similarly, our findings are not driven by any particular group or culture group (see Figure A-2 in which we exclude one group and culture group at the time). This result also supports our claim that France cared little about individual ethnic groups when drawing colonial district borders.[24] Finally, we document that the statistical significance of our main results is robust to alternative forms of clustering of the regression errors and spatial autocorrelation (see Figure A-3).

4.3. Within district and homeland evidence

We now turn to differences in local economic development *within* ethnic homelands to substantiate our claim that local majorities are important for understanding comparative development. We continue to make use of the fact that the ancestral homelands of most ethnic groups in French colonial Africa were cut by colonial district borders in a manner that disregarded their local circumstance.[25] However, we now focus the historical population share of an ethnic group within a particular colonial district (PS_{edp}).

[24] Local majorities are also not an expression of particular forms of colonial agricultural production, i.e. cash crop production (Roessler et al., 2020). Table C-3 interacts our local majority measure with proxies for the intensity of cash crop suitability (based on GAEZ data) on the respective homelands for eight different cash crops. We find no systematic evidence that cash crop production drives our results, but some evidence that suggest that homeland which are particular suited for the production of groundnuts capitalize somewhat less on their local majorities. Details on how we proxy for the different cash crop suitability is provided in Online Appendix D.

[25] The median ethnic homeland intersects two colonial districts in French protectorates, and the median colonial district intersects four ethnic homelands.

To exploit this local variation and obtain a unique mapping of which ethnic group occupies which part of a district, we use grid cells of 5 × 5 arc minutes (equivalent to 9.3 × 9.3 km at the equator) as our unit of analysis.[26] For each grid cell c, we determine both the ethnic homeland e and the colonial district d in which the cell's centroid is located. We then assign the corresponding district-level population share PS_{edp} to each cell within any given homeland-district intersection (indexed by ed). Figure III illustrates the grid cells in French Guinea and the values of PS_{edp} that we assign to each of these cells.

Figure III about here.

For the grid-level analysis, our main empirical specification is

$$\ln Lights_{cedp} = \beta\, PS_{edp} + FE_{ep} + FE_{dp} + \mathbf{z}'_{cedp}\gamma + \epsilon_{cedp}, \tag{3}$$

where $Lights_{cedp}$ is the log of nighttime light density in grid cell c. The more disaggregated unit of analysis also allows including district fixed effects FE_{dp} and homeland (i.e., homeland-protectorate) fixed effects FE_{ep}. The district fixed effects absorb all district-specific characteristics, such as district-level ethnic fractionalization and polarization, the personality of the first district officer, and so on. The homeland fixed effects absorb all homeland-specific and group-specific characteristics, such as pre-colonial centralization, the initial level of development, and history of violent conflict. Finally, we control for the geography of each cell via $\mathbf{z}_{cedp}$.[27]

Given this particular fixed-effects structure, we now leverage variation within ancestral homelands that were split over multiple ethnically heterogenous districts. Our estimate of interest, β, tells us how much more (or less) light intensity we observe in a cell that is traditionally occupied by a group that happens to have higher population share in this cell compared to other cells (in heterogenous districts) it traditionally occupies. We expect these estimates to be smaller than the homeland-level estimates, as they stand for considerably more local differences in economic development.

Table IV about here.

Table IV shows estimates of the effect of local majorities, measured by the district-level population share, on current economic development at the grid-level. Column (1) documents a positive relation in French protectorates in the absence of any fixed effects beyond the level of protectorates. Column (2) shows that the relation remains positive and statistically significant when adding district fixed effects. Hence, within colonial districts, the area belonging to the homeland of a majority group is more economically

[26]This matches the resolution of the HYDE grid.

[27]Once we include these two types of fixed effects, there is no longer any need for district-level or homeland-level control variables.

developed today than areas belonging to the homelands of minority groups. Column (3) includes homeland but no district fixed effects, so that it compares outcomes of the same group across different districts. Column (4) includes both district and homeland-protectorate of fixed effects. This final specification purges differential investment levels and other factors specific to colonial districts by using the variation from ethnically diverse districts alluded to above. The coefficients of interest are remain positive and statistically significant. Hence, within ethnic homelands, the areas intersecting colonial districts in which a group had a higher population share are more economically developed today than areas intersecting colonial districts in which the same group was a minority. The coefficient estimate in column (4) implies that a one-standard deviation increase in a group's district-level population share (0.32) increases nighttime light emissions by four percent and, consequently, GDP by slightly more than one percent.[28] Overall, these are consistent with our earlier results and suggest that local majorities drive the aggregate homeland effect and not some other factor that correlates with the expected population share of these groups. Local majority groups are more prosperous today compared to minorities within the same district, and groups are more prosperous in the majoritarian areas of their homelands, even if we partial out all district-specific characteristics.

Robustness tests: The within district and homeland results are robust to battery of perturbations. We obtain qualitatively similar results using alternative measures of local majorities discussed earlier (see Table A-10) or using the DMSP luminosity data on the left hand side (see Table A-11). The statistical significance of the effect also carries over to predicting extensive margin using a "cell is lit" dummy on the left hand side (similar to Michalopoulos and Papaioannou, 2013). Moreover, we document that our results are not driven by *(i)* ethnic homelands being more economically developed in their center than their periphery and *(ii)* district-level population shares PS_{edp} being mechanically higher in the center of ethnic homelands than in their periphery. In Table A-12, we show that the distance of a grid-cell to the homeland centroid is predictive of neither population density, nor PS_{edp}. Furthermore, we do not find any evidence suggesting that our within-district and within-homeland results are amplified or mitigated by the national population share of an ethnic group or their homeland-level expected population share (see Table A-13). Finally, we show that the results are not driven by grid cells in close proximity to colonial district capitals (see Table A-14).

[28]Given the two types of fixed effects, this result is driven neither by ethnically homogenous colonial districts, nor by ethnic homelands that are entirely within a single colonial district.

5. Persistent borders or persistent advantages?

Having established that local majorities during early colonization shape the geography of economic development today, both across and within ethnic homelands, raises the question of what drives this persistence. Do colonial local majorities matter *(i)* because subnational borders and, therefore, local population shares are persistent, or *(ii)* because the advantages of colonial local majorities persist even if the borders change? In this section, we first look at the persistence of subnational borders.[29] We then investigate the extent to which the long-run effect of local majorities in colonial times on current economic development could result from a combination of persistent subnational borders and a short-run effect of current local majorities on current economic development.

We return to the example of French Guinea to illustrate our approach. Figure IV shows the current and past district-level population shares and EPS_{ep}, where the current districts reflect the primary subnational level (ADM1) and obtained from the Database of Global Administrative Areas (GADM, 2015).[30] The figure highlights that most of the current border segments are still from the colonial area or are, at least, in very close proximity to colonial district borders. Nonetheless, Guinea has slightly reduced the number of subnational units by abolishing some colonial district borders and merging several former districts into larger provinces. We observe that the three largest groups managed to hold on to their early dominance, or even managed to expand it, while most smaller groups have lower expected population shares today than they had in the past.

Figure IV about here.

To study this issue more systematically, we recompute the population shares PS_{edp} of each ethnic group e in each subnational administrative unit d as well as the group/homeland-level expected population share EPS_{ep} using current GADM-ADM1 borders. Note that we deliberately use the same historical population density maps to compute our colonial measures to isolate changes in our measures due to border reforms from changes due to migration and differences in population growth (see Bazzi and Gudgeon, 2020, for a similar approach).[31] Migration flows into and out of majority homelands are, of course, a key aspect of how ethnic groups sort in space after independence, which is why we analyze them separately further below.

[29] Craib (2017) documents that colonial borders no matter how arbitrary became the reference point for reforms even after independence. Moreover, Carter and Goemans (2011) find that "new" international borders also tend to follow previous administrative borders.

[30] We use GADM version 2.8, which presents borders valid as of November 2015. GADM does not provide information for how long the depicted borders are already valid.

[31] Table A-15 in Online Appendix A uses current population values based on the GHSL to construct our measures. The results stay qualitatively the same.

Table V about here.

Table V shows that both past and current local majorities matter for current economic development within and across ancestral homelands. Column (1) reports a positive, large and precisely estimated effect of EPS_{ep} from colonial times (EPS_{ep} in the 1920s) on current EPS_{ep} (in the 2010s), using country fixed effects (akin to protectorate fixed effects) as well as our full set of homeland-level control variables. The correlation between the expected population share during colonial times and the same measure in the 2010s is about 65%. Next, test whether the current or the colonial version of EPS_{ep} is a better predictor of current economic development across ethnic homelands. Column (2) highlights that both, colonial and current EPS_{ep}, matter for economic development, and we cannot reject that their effects are the same. Columns (3) and (4) replicate the exercise on the grid-cell level employing the most stringent set of fixed effects (corresponding to column 4 of Table IV). We find a slightly weaker, but still substantial persistence in the district-level population share of a group in the 1920s and the 2010s. Moreover, we again find that both measures matter and have comparable effect sizes.

The persistence results have two important implications. First, they suggest that we can undo the adverse effects of French administrative policy by creating more homogenous administrative units (and hence units with a clear local majority) today. However, the high degree of persistence documented in both measures and the lasting effect of local majorities from colonial times also suggests that this rarely occurs. The second implication is that local majorities in colonial times shape comparative development today beyond their direct effect on current local majorities. These channels are at work even if subnational administrative borders today do not resemble those of the past.

6. Place-based or identity-based advantages?

It is a priori unclear whether the benefits of local majorities accrue to people identifying with an ethnic group who happened to be a local majority during colonial times or the people currently living in the corresponding ancestral homelands of such a group. We cannot discriminate among these two scenarios by relying on the light intensity of a particular location alone. This section uses individual-level data from the Demographic and Health Surveys (DHS) to address this question. Using this micro-data has the added benefit that we can study more disaggregated measures of economic well-being, such as household wealth, and broader measures of human development, such as education and health outcomes.

The DHS are large, nationally representative household surveys that include questions on health, fertility, education, household assets, and ethnicity. We use data from 40 geo-referenced DHS surveys in 12 Sub-Saharan African countries which used to be part of French colonial Africa (see Table D-3 in Online Appendix A for details).[32] Hence, the sample differs the data underlying our previous analyses in two ways. First, some Sub-Saharan African countries were never surveyed, and, second, even if a survey was undertaken, the survey may have not sampled any cluster in a particular ethnic homeland or former colonial district. Figure A-4 in Online Appendix A shows the spatial distribution of DHS cluster locations. We observe that DHS clusters are concentrated in a few regions of French colonial Africa.

To test if the effects of local majorities occur at the homeland or the group level, we now calculate two new variants of the expected population share. EPS_{ep} *place* is the expected population share of the ancestral homeland on which a DHS respondent resides, no matter the respondent's ethnicity. EPS_{ep} *identity* is the expected population share of the ethnic group a DHS respondent identifies with, independent of the actual place of residence. To calculate the identity-based measure, we match the ethnicity reported in the DHS surveys with Murdock groups following Hodler et al. (2020). We the use these two measures as explanatory variables in individual-level specifications in which the dependent variables are the DHS wealth index (as computed by Bruederle and Hodler, 2018), years of schooling, or infant mortality.[33] All our individual-level specifications include protectorate wave-fixed effects, the set of geography controls (now at the cluster level) plus a series of individual-level controls (age, age squared, an urban indicator, a gender indicator, and a non-indigenous indicator for individuals living outside their ethnic homeland).[34]

Table VI around here.

Table VI presents results that suggest the effects of local majorities in the early colonial period run through place and identity. Columns (1)–(4) of Table VI show that both EPS_{ep} *place* and EPS_{ep} *identity* have a positive effect on household wealth, as well as on years of schooling. Columns (5) and (6) report a statistically negative effect on infant mortality, which is statistically significant only for EPS_{ep} *identity*. These results show that our main finding is not restricted to nighttime light emissions, but also holds for other measures of economic and human development. In Online Appendix A, we add that our within-homeland results are also robust to the use these alternative measures of economic

[32]In all geo-referenced DHS, urban clusters are displaced by up to 2 km and rural clusters by up to 5 km (and 1% of rural clusters by up to 10 km). We also collect data on all other former colonies resulting in 86 geo-referenced DHS from 25 Sub-Saharan African countries, which we use when we study the external validity of our results.

[33]The sample size differs for these three dependent variables. It is largest for infant mortality, as we get this information from the full birth history of each female respondent. It is smallest for years of schooling, which is missing for a considerable share of the respondents.

[34]See Michalopoulos and Papaioannou (2016) for a similar approach to using the DHS data.

and human development (see Table A-17).[35] While we find that both place-based and the identity-based effects matter, the latter is significantly stronger in four out of these six specifications. This result tells us that members of local minorities are unlikely to close the gap to members of local majorities simply by migrating to the homelands of local majority groups.

7. Mechanisms

We have already established the persistence of subnational borders as one crucial mechanism through which local majorities in colonial times affect current economic development. However, we have also seen that the effect of past majorities remains even if border reforms change the ethnic composition of contemporary districts. Moreover, we just documented that some benefits of local majorities during the early colonial period are place-based, while others are identity-based. All of these insights together suggest that there are multiple mechanisms at work. This section studies a set of potentially influential mechanisms from early agglomeration over colonial infrastructure or health and education to migration.

7.1. Early agglomerations and colonial communication infrastructure

The long-run effects of early agglomerations and early infrastructure investments on contemporary development are well-documented. Many colonial district capitals have become important urban centers in today's Sub-Saharan Africa (Bairoch, 1991, p. 508). Colonial infrastructure investments directed where agglomeration took place (Jedwab and Moradi, 2016; Jedwab et al., 2017). An important driver of why local majorities in colonial times are more developed today may be that they were more likely to host one or several colonial district capitals. If so, then they disproportionately benefited from early infrastructure investments and economies of scale.

We collect data on the locations of colonial district capitals across Sub-Saharan Africa. Fortunately, district capitals are usually depicted on the same maps that we use to digitize district borders. In the case of French West Africa our source maps also include information on the presence of important communication infrastructure, such as post offices and telegraph stations.[36] We generally observe a clustering of early communication infrastructure around district capitals, which is consistent with the French government's approach of direct rule. We also have information on the location of capital cities and the

[35] We can only verify the within-homeland results based on the respondents place of residence but not based on their identity. The reason is their identity refers to the level of ethnic groups rather than the level of ethnic group-district intersections.

[36] Figure A-5 in Online Appendix A shows these data for French Guinea in 1922.

presence of communication infrastructure in 1935.[37] Using this data, we test whether local majority groups are more likely to obtain district capitals and communication infrastructure in 1922, and whether those initial investments lead to more investments later on (in 1935).

Table VII about here.

Table VII presents a number of interesting results at the cross-homeland level. Column (1) shows that ethnic groups with high expected population shares EPS_{ep} host considerably more capitals than minorities (after conditioning on the full set of control variables and protectorate fixed effects). Columns (2) and (3) add that the number of district capitals is positively associated with the number of post and telegraph stations. Early communication infrastructure clusters mostly in and around district capitals, and groups "receive" infrastructure primarily via their proximity to the capital. In columns (4) to (5), we take the number of post and telegraph stations in 1935 as dependent variables and additionally control for the number of such stations in 1922. Not surprisingly, we find that lagged investments predict subsequent investments in 1935. However, we also find that ethnic homelands with more district capitals gain additional communication infrastructure over the period from 1922–1935. Even when we account for both of these sources of persistence, we still find a positive association of EPS_{ep} with communication infrastructure in 1935 (although this effect is only significant at the 10% level).

In Online Appendix A, we present the corresponding within-homeland results (Table A-19). These results corroborate the overall findings at the homeland level. They show that district capitals are considerably more likely to be located in the parts of an ethnic homeland where the group is a local majority (where the group's district-level population share PS_{edp} is high) and that these capital-cells are more likely to be connected to the colonial communications network. The results also confirm that path dependence investments and historically high population shares both predict communication infrastructure in 1935.

Important as this channel may be, early agglomeration and infrastructure investments alone cannot explain the effect of colonial local majorities on comparative development. No matter if we include colonial district capitals or control for early infrastructure in the homeland-level and grid-level estimations, we still estimate economically meaningful and statistically precise effects of our measures of local majorities on current nighttime light intensity (see Table A-20).

[37] In Online Appendix A, these are shown in Panel C of Figure A-5.

7.2. Health

Next we study whether members of local majority groups have better health outcomes because they benefited more from colonial investments in public health. To proxy for colonial investments in health infrastructure, we aggregate the number of health personnel within colonial districts in 1922 collected by Huillery (2009) to the homeland level. Since a homeland is usually represented in more than once district, we use weighted-averages where the weights are the historical population share of the group in each district. To measure the current health infrastructure across ancestral homelands, we calculate the log of health facilities density based on data from Maina et al. (2019), who collect the coordinates of close to a 100,000 health facilities across 50 African countries.[38]

Table VIII presents our results.

Table VIII about here.

Columns (1)–(3) show that ancestral homelands belonging to colonial local majorities have indeed a higher density of health facilities today and that our imperfect proxy of colonial health infrastructure also tends to be positively related to this density (but not a standard levels of statistical significant). This latter result is in line with the finding of Huillery (2009) that areas with higher colonial health investments have a better health infrastructure today.

The remaining columns of Table VIII are based on our DHS samples. We look at birth attendance as a measure of health services and infant mortality as a measure of health outcomes. We find that the density of health facilities is predictive of high birth attendance and low infant mortality rate. Hence, members of colonial local majority groups enjoy the place-based benefits of better current health infrastructure. In addition, consistent with results shown earlier, we find that these members also enjoy mobile identity-based benefits in the form of higher birth attendant and lower infant mortality compared to people identifying with colonial local minority groups.

7.3. Education

Since human capital is embodied, early access to education could be one of the key factors why we observe both place-based and identity-based differences in economic development today. It is well-established that differential investments in education are a key source of regional inequalities in Sub-Saharan Africa (e.g., Ricart-Huguet, 2020; Huillery, 2009). Moreover, historical differences in education appear to affect inter-generational mobility in Africa even if spatial sorting is taken into account (Alesina et al., 2021).

We use data from the sample DHS surveys in combination with data on colonial

[38]Details on the measure are provided in Section D-2.

investments into education from Huillery (2009) to investigate whether respondents identifying with historical local majorities have higher educational attainment than those identifying with minorities, after controlling for differences in colonial investments into education. Moreover, we test whether respondents residing on areas which historically received more investments are more educated compared to people residing elsewhere. To this end, we aggregate the historical educational investment data from Huillery (2009) to the homeland level. Specifically, we take the weighted average of teachers across a homelands district, in the same way as we aggregated health personal earlier on. Our dependent variable is the log years of schooling.

Table IX about here.

Table IX presents the results. Columns (1) and (2) show that respondents identifying with ethnic groups from homelands with high colonial investments into education are better educated today. Furthermore, respondents residing in homelands of colonial local majority groups are more educated as well. These results suggest that group identity and residence both matter. Columns (3) and (4) add that these effects are considerably smaller for respondents born after 1980 than for respondents prior to 1980.

7.4. Migration

Last but not least, we study whether selective migration may explain why the areas where colonial local majority groups lived are doing better today. After all, subnational borders are not hard borders, so that people can easily move towards areas with better infrastructure, health care, and educational opportunities, or towards places where they are part of the majority (Tiebout, 1956). We use two different approach: *i)* indirect evidence via current places of residence for all countries covered by the DHS program and *ii)* direct evidence of lifetime migration for a select group of countries.

We start with a cross-homeland (or cross-group) analysis based on the DHS sample and test whether respondents from ethnic groups that were colonial local majorities are more or less likely to live outside their ancestral homeland. Column (4) in Panel A of Table X shows that groups with a higher expected population share are less likely to live outside their homeland. We interpret this as indirect evidence of selective migration in the sense that respondents belonging to local minority groups must have moved away from their homeland more frequently after the end the colonial period. Panels B and C adds that this selectivity is more pronounced in migration towards rural areas and among older cohorts.

Table X about here.

While these results are very suggestive, they leave open a number of important questions, such as which generation actually migrated or is there selective migration within ethnic homelands (towards majority areas)? We cannot address these questions using DHS data, as they do not record the birthplace of survey respondents.

To address these questions, we rely on census data from IPUMS. The project harmonizes census microdata from around the world, some of which include all the information we need: the ethnic identity of a respondent, the name of the current subnational unit where the respondent resides, and the name of the subnational unit where the respondent was born. Current censuses in five former French colonies in Sub-Saharan Africa (Benin, Burkina Faso, Guinea, Mali and Senegal) contain these three pieces of information at sufficient granularity that we can match individuals to their historical ethnic group and colonial district.[39] We then build three measures of internal migration.[40] First, we define a general migration dummy which equals one for all respondents for which the birth district and current district of residence differ (roughly 17% of our sample of 2.5 million people). Second, we classify migration as "inside of homeland" if the district of current residence overlaps with the ancestral homeland of the ethnic group with whom the respondent identifies, and as "outside of homeland" otherwise.

In columns (1)–(3) of Panel A in Table X, we use these three migration dummies as dependent variables and regress them on the colonial population share PS_{edp} of the ethnic group with whom the individual identifies measured in the birth district. We see that individuals born in districts with a higher population share of their own groups are less likely to migrate. Column (1) suggests that one standard deviation increase in PS_{edp} implies a respondent is -3.1 percentage points less likely to migrate. This corresponds to a 17.5% reduction compared to the sample average of 17%. Moreover, columns (2) and (3) show the effect is mostly (but not exclusively) driven by a lower propensity of individuals to migrate into a district outside of their ancestral homeland. The weaker results in Panel B suggests that these findings are mostly driven by migration towards rural areas. Panel C adds that the propensity to migrate out of the ancestral homeland is larger in the sample of individuals born prior to independence.

We conclude that selective migration across and within homelands may help to explain why colonial local majorities shape aggregate development outcomes until today. In addition, we find that the selectivity of migration is more pronounced for migration towards rural areas. Considering that subsistence agricultural is still important in rural areas in Sub-Saharan Africa and that new arrivals in rural areas have typically no land of their own, this result may also contribute towards explaining why some of the disadvantages of being from a local minority group are tied to identity rather than place.

[39] IPUMS is not georeferenced.

[40] External migration is excluded.

8. External validity

How do the results from French colonial Africa translate to other colonies? So far, we have relied exclusively on former French colonies because the anecdotal historical record and the quantitative evidence provided in this paper strongly support the notion that the French paid little regard to local circumstances when they established the first districts in their colonies and protectorates. The British, in turn, are known to select specific groups to rule for them by proxy through the system of indirect rule. Nevertheless, there is evidence that this system was not uniformly implemented and differences among these systems are continuous instead of discrete.[41] Moreover, when the British drew the first districts, they too lacked reliable information on the ethnic makeup of their protectorates, so that early districts rarely followed ethnic boundaries as consistently as the districts of the 1950s (see, e.g., Burgess et al., 2012, on boundaries in colonial Kenya).

We find some evidence of selection on group characteristics within former British colonies, as well as in the sample of all former colonies in Sub-Saharan Africa in Online Appendix A. Figure A-6, for example, shows that groups which are more politically developed prior to colonization, as measured by class stratification, are more likely to become local majorities. As expected, the evidence favoring selection is not particularly strong for early district boundaries implemented by the British and other colonizers. While we can no longer interpret these results as a clean natural experiment, there is randomness in the design of the districts owed in part to a lack of information and a tendency of all colonizers to follow natural boundaries. We proceed by re-estimating our core homeland results based on the most restrictive specification but cautiously interpreting these results as no more than partial correlations.

Table XI about here.

Table XI reports three sets of results: light intensity across homelands, the DHS wealth index, and grid-cells within homelands. Columns (1) and (2) show that the point estimates for the British and full sample are well within a standard error of our core results. Columns (3) and (4) add that the results are also similar if we focus on wealth using the DHS micro-data. Although, the identity effects are somewhat smaller. To us this pattern suggests that even though colonial powers other that France sometimes took pre-colonial characteristics into account, these cases are swamped by many other instances when they did not.[42] Selection may play little role in the link between local majorities and

[41] In Southern Nigeria, for example, Lugard himself perceived that "most districts were populated by tribes in the lowest state of primitive savagery" (Kirk-Greene, 1968, p. 67). As a result, only "a few large tribes [. . .] were well adapted to indirect rule. Other groups [. . .] were arbitrarily lumped into provinces staffed by British administrative offices" (Whittlesey, 1937, p. 365).

[42] This holds across (rather than within) other colonizers as well. Portugal and Belgium, for example, are usually considered to have ruled their dependencies even more directly than the French.

contemporaneous differences in development across homelands in Sub-Saharan Africa. Columns (5) and (6) probe the relationship within homelands using data at the grid level together wit the most restrictive specification with district and homeland fixed effects. Here we observe some notable differences. The effect of the population share in British districts on contemporary development is now twice as large as our baseline estimate for former French colonies. The effect size for all of Sub-Saharan Africa is right in the middle of these two estimates. This hints hints at a positive selection problem creating an upward bias within British colonies, which fits the historical narrative. However, these results could also imply that local majorities wield greater power in more decentralized societies (most former French colonies are more centralized their British counterparts today).

9. Conclusion

This paper shows that the composition of ethnic groups within and across administrative units is intrinsically linked to the economic well-being of these groups. We exploit the first administrative-territorial structures that the European colonial powers imposed on their protectorates in Sub-Saharan Africa in the early 20th century as a natural experiment. More specifically, we focus on French protectorates, as the French showed little interest in pre-colonial political units when designing the administrative and territorial structure. Exploiting the plausibly exogenous variation in the population shares of an ethnic group across colonial districts, we find that ethnic groups which ended up being local majorities in most districts are more economically successful today. Moreover, parts of ethnic homelands with a higher district-level population share during colonial times are more economically developed today than other parts of the same homeland. Becoming a local majority during colonial times has long-run effect on current economic development for various reasons, including (i) the persistence of administrative borders, (ii) agglomeration economies and higher investments during colonial times, (iii) an ability to better utilize those investments in education and health, as well as (iv) a migration response that amplified these differences.

Neither the administrative units set up by the European colonizers, nor their effects on the spatial distribution of economic activity are destiny. Many African governments have reformed their administrative-territorial structures and we show that these reforms forge different majority (and minority) areas. This has implications for future research and policy. First, we need to a better understanding of the political economy of administrative reforms (in Africa and beyond). Second, from a policy perspective, it is important to understand that any such reform appears to have distributional consequences, so that one could identify those reforms that have the potential to reduce ethnic inequality at the level of countries (Alesina et al., 2016) or the local level (Hodler et al., 2020).

References

Acemoglu, D., S. Johnson, and J. A. Robinson (2001). The colonial origins of comparative development: An empirical investigation. *American Economic Review 91*(5), 1369–1401.

Acemoglu, D., S. Johnson, and J. A. Robinson (2002). Reversal of fortune: Geography and institutions in the making of the modern world income distribution. *Quarterly Journal of Economics 117*(4), 1231–1294.

Alesina, A., W. Easterly, and J. Matuszeski (2011). Artificial states. *Journal of the European Economic Association 9*(2), 246–277.

Alesina, A., S. Hohmann, S. Michalopoulos, and E. Papaioannou (2021). Intergenerational mobility in Africa. *Econometrica 89*(1), 1–35.

Alesina, A., S. Michalopoulos, and E. Papaioannou (2016). Ethnic inequality. *Journal of Political Economy 124*(2), 428–488.

Alesina, A., R. Perotti, and E. Spolaore (1995). Togetheror separately? Issues on the costs and benefits of political and fiscal unions. *European Economic Review 39*(3-4), 751–758.

Ali, M., O.-H. Fjeldstad, B. Jiang, and A. B. Shifa (2019). Colonial legacy, state-building and the salience of ethnicity in sub-Saharan Africa. *Economic Journal 129*(619), 1048–1081.

Bairoch, P. (1991). *Cities and economic development: From the dawn of history to the present.* Chicago, IL: University of Chicago Press.

Banerjee, A. V. and R. Pande (2007). Parochial politics: Ethnic preferences and politician corruption. CEPR Discussion Paper No. 6381.

Bazzi, S. and M. Gudgeon (2020). The political boundaries of ethnic divisions. *American Economic Journal: Applied Economics*, Forthcoming.

Beiser-McGrath, J., C. Müller-Crepon, and Y. I. Pengl (2020). Who benefits? How local ethnic demography shapes political favoritism in Africa. *British Journal of Political Science*, 1–19.

Bening, R. B. (1984). Internal colonial boundary problems of the Gold Coast 1907–1951. *The International Journal of African Historical Studies 17*(1), 81–99.

Bening, R. B. (1986). *The evolution of the administrative boundaries of northern Ghana 1898 to 1965.*

Bluhm, R., R. Hodler, and P. Schaudt (2021). Ethnofederalism and ethnic voting. CESifo Working Paper No. 9314.

Bruederle, A. and R. Hodler (2018). Nighttime lights as a proxy for human development at the local level. *PloS One 13*(9), e0202231.

Burgess, R., M. Hansen, B. A. Olken, P. Potapov, and S. Sieber (2012). The political economy of deforestation in the tropics. *The Quarterly Journal of Economics 127*(4), 1707–1754.

Burgess, R., R. Jedwab, E. Miguel, A. Morjaria, and G. Padró i Miquel (2015). The value of democracy: Evidence from road building in Kenya. *American Economic Review 105*(6), 1817–51.

Carter, D. B. and H. E. Goemans (2011). The making of the territorial order: New borders and the emergence of interstate conflict. *International Organization*, 275–309.

Cohen, W. B. (1971). *Rulers of Empire: The French Colonial Service in Africa.* Hoover Institution Press.

Conklin, A. L. (1997). *A mission to civilize*. Stanford University Press.

Craib, R. B. (2017). Cartography and decolonization. In J. R. Akerman (Ed.), *Decolonizing the Map*. University of Chicago Press.

Crowder, M. (1964). Indirect rule–French and British style. *Africa 34*(3), 197–205.

Crowder, M. (1968). *West Africa under colonial rule*. London: Hutchinson.

De Luca, G., R. Hodler, P. A. Raschky, and M. Valsecchi (2018). Ethnic favoritism: An axiom of politics? *Journal of Development Economics 132*, 115–129.

Dickens, A. (2018). Ethnolinguistic favoritism in African politics. *American Economic Journal: Applied Economics 10*(3), 370–402.

Dimico, A. (2017). Size matters: The effect of the size of ethnic groups on development. *Oxford Bulletin of Economics and Statistics 79*(3), 291–318.

Eberhard, D. M., G. F. Simons, and C. D. Fennig (2020). Ethnologue: Languages of the world. Twenty-third edition. *www.ethnologue.com*.

Edwards, E. and A. Roberts (1986). French black africa. In A. Roberts (Ed.), *The Cambridge history of Africa*, Volume 7, pp. 330–392. Cambridge University Press.

Franck, R. and I. Rainer (2012). Does the leader's ethnicity matter? Ethnic favoritism, education, and health in sub-Saharan Africa. *American Political Science Review 106*(2), 294–325.

Francois, P., I. Rainer, and F. Trebbi (2015). How is power shared in Africa? *Econometrica 83*(2), 465–503.

GADM (2015). Gadm database of global administrative areas, version 2.8. *www.gadm.org*.

Henderson, J. V., A. Storeygard, and D. N. Weil (2012). Measuring economic growth from outer space. *American Economic Review 102*(2), 994–1028.

Herbst, J. (1989). The creation and matintenance of national boundaries in Africa. *International Organization 43*(4), 673–692.

Hodler, R. and P. A. Raschky (2014). Regional favoritism. *Quarterly Journal of Economics 129*(2), 995–1033.

Hodler, R. and P. A. Raschky (2017). Ethnic politics and the diffusion of mobile technology in Africa. *Economics letters 159*, 78–81.

Hodler, R., S. Srisuma, A. Vesperoni, and N. Zurlinden (2020). Measuring ethnic stratification and its effect on trust in Africa. *Journal of Development Economics*, 102475.

Horowitz, D. L. (1985). *Ethnic groups in conflict*. University of California Press.

Huber, J. D. (2012). Measuring ethnic voting: Do proportional electoral laws politicize ethnicity? *American Journal of Political Science 56*(4), 986–1001.

Huillery, E. (2009). History matters: The long-term impact of colonial public investments in French West Africa. *American Economic Journal: Applied Economics 1*(2), 176–215.

Ichino, N. and N. L. Nathan (2013). Crossing the line: Local ethnic geography and voting in Ghana. *American Political Science Review 107*(2), 344–361.

Jedwab, R., E. Kerby, and A. Moradi (2017). History, path dependence and development: Evidence from colonial railways, settlers and cities in Kenya. *Economic Journal 127*(603), 1467–1494.

Jedwab, R. and A. Moradi (2016). The permanent effects of transportation revolutions in poor countries: Evidence from Africa. *Review of Economics and Statistics 98*(2), 268–284.

Kirk-Greene, A. H. M. (1968). *Lugard and the amalgamation of Nigeria: A documentary record.* Cass.

Klein Goldewijk, K., A. Beusen, and P. Janssen (2010). Long-term dynamic modeling of global population and built-up area in a spatially explicit way: HYDE 3.1. *The Holocene 20*(4), 565–573.

Lijphart, A. (1977). *Democracy in plural societies: A comparative exploration.* Yale University Press.

Lipset, S. M. (1960). *The social bases of politics.* The Johns Hopkins UniversityPress.

Maina, J., P. O. Ouma, P. M. Macharia, V. A. Alegana, B. Mitto, I. S. Fall, A. M. Noor, R. W. Snow, and E. A. Okiro (2019). A spatial database of health facilities managed by the public health sector in sub Saharan Africa. *Scientific data 6*(1), 1–8.

Michalopoulos, S. and E. Papaioannou (2013). Pre-colonial ethnic institutions and contemporary african development. *Econometrica 81*(1), 113–152.

Michalopoulos, S. and E. Papaioannou (2014). National institutions and subnational development in Africa. *Quarterly Journal of Economics 129*(1), 151–213.

Michalopoulos, S. and E. Papaioannou (2016). The long-run effects of the scramble for Africa. *American Economic Review 106*(7), 1802–48.

Michalopoulos, S. and E. Papaioannou (2020). Historical legacies and African development. *Journal of Economic Literature 58*(1), 53–128.

Müller-Crepon, C. (2020). Continuity or change? (In)direct rule in British and French colonial Africa. *International Organization*, 1–35.

Murdock, G. P. (1959). *Africa: Its peoples and their culture history.* McGraw-Hill.

Murdock, G. P. (1967). Ethnographic atlas: A summary. *Ethnology 6*(2), 109–236.

Nunn, N. (2008). The long-term effects of Africa's slave trades. *Quarterly Journal of Economics 123*(1), 139–176.

Posner, D. N. (2005). *Institutions and ethnic politics in Africa.* Cambridge University Press.

Ricart-Huguet, J. (2020). The origins of colonial investments in former British and French Africa. *British Journal of Political Science*, Forthcoming.

Roessler, P., Y. Pengl, R. Marty, K. S. Titlow, and N. van de Walle (2020). Extractive colonial economies and legacies of spatial inequality: Evidence from africa.

Tiebout, C. M. (1956). A pure theory of local expenditures. *Journal of Political Economy 64*(5), 416–424.

Weidmann, N. B., J. K. Rød, and L.-E. Cederman (2010). Representing ethnic groups in space: A new dataset. *Journal of Peace Research 47*(4), 491–499.

Whittlesey, D. (1937). British and French colonial technique in West Africa. *Foreign Affairs 15*(2), 362–373.

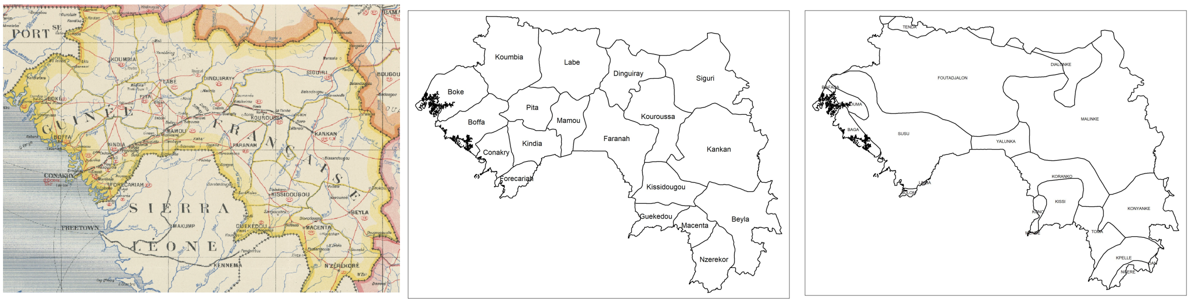
Widmer, P. and N. Zurlinden (2019). Born in the right place? Health ministers, foreign aid and infant mortality. University of St.Gallen Economics Working Paper No. 1911.

Zartman, W. I. (1965). The politics of boundaries in North and West Africa. *The International Journal of African Historical Studies 3*(2), 155–173.

Figures and Tables

Figure I
Illustration of colonial district borders and ethnic homelands

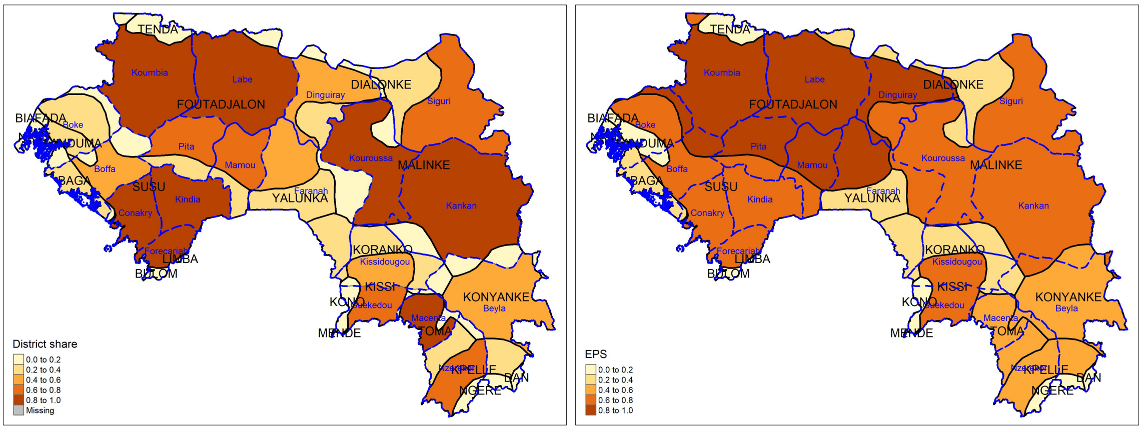
(A) Map (1922) (B) Circles of French Guinea (C) Homelands (Murdock)



Notes: Panel (A) of the figure depicts an administrative map of the Colony and Protectorate of French Guinea in 1922 published by the Colonial Office of French West Africa in 1922. Panel (B) shows the extracted district information, with the district names displayed at the district centroid. Panel (C) illustrates the ethnic composition of French Guinea based on Murdock's ethnolinguistic map.

Figure II
Ethnic groups across districts

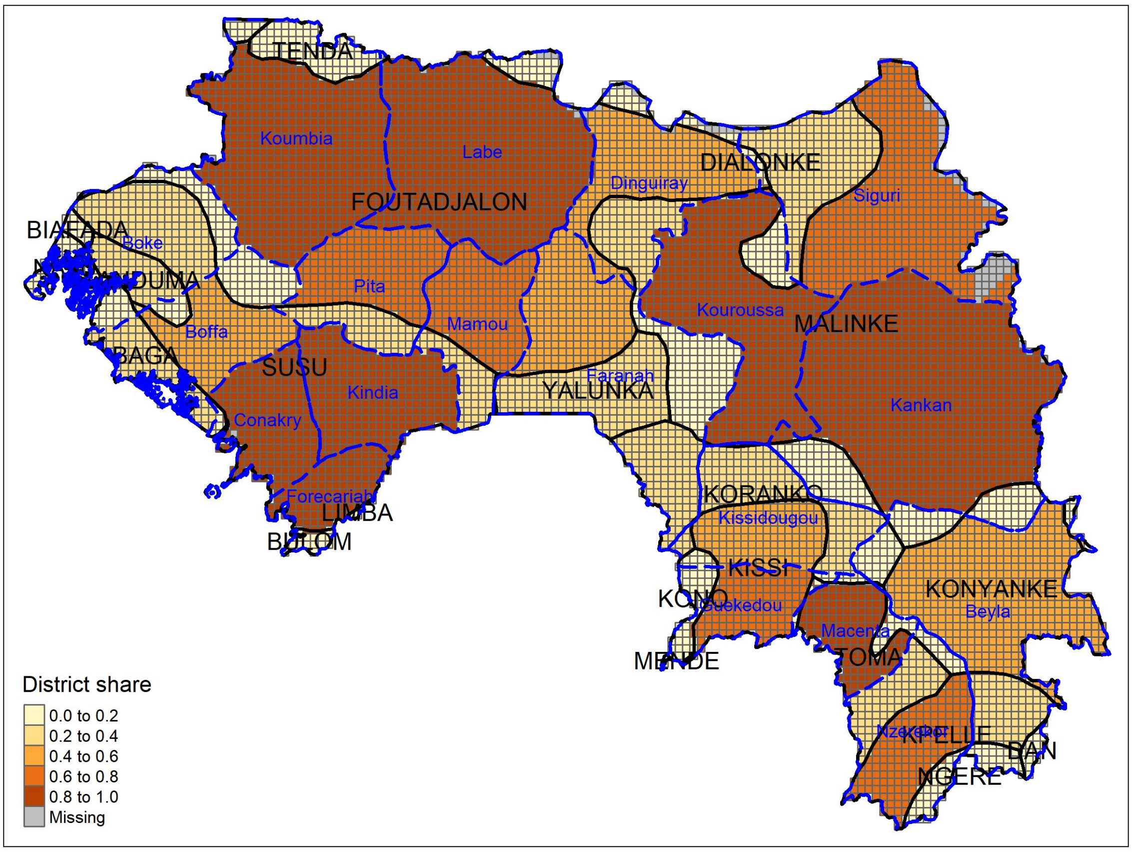
(A) District shares (B) EPS



Notes: Panel (A) of the figure depicts the distribution of ethnic groups across districts in the French Guinea around 1920. Murdock's homeland borders are highlighted in black (including the homeland name at homeland centroid), district borders in blue (name displayed at the district centroid). The population shares of each ethnic group are coloured in the homeland areas located within a specific district. Panel (B) plots the weighted averages of the district shares at the homeland level as defined in Section 3.

FIGURE III

Ethnic groups across districts and cells

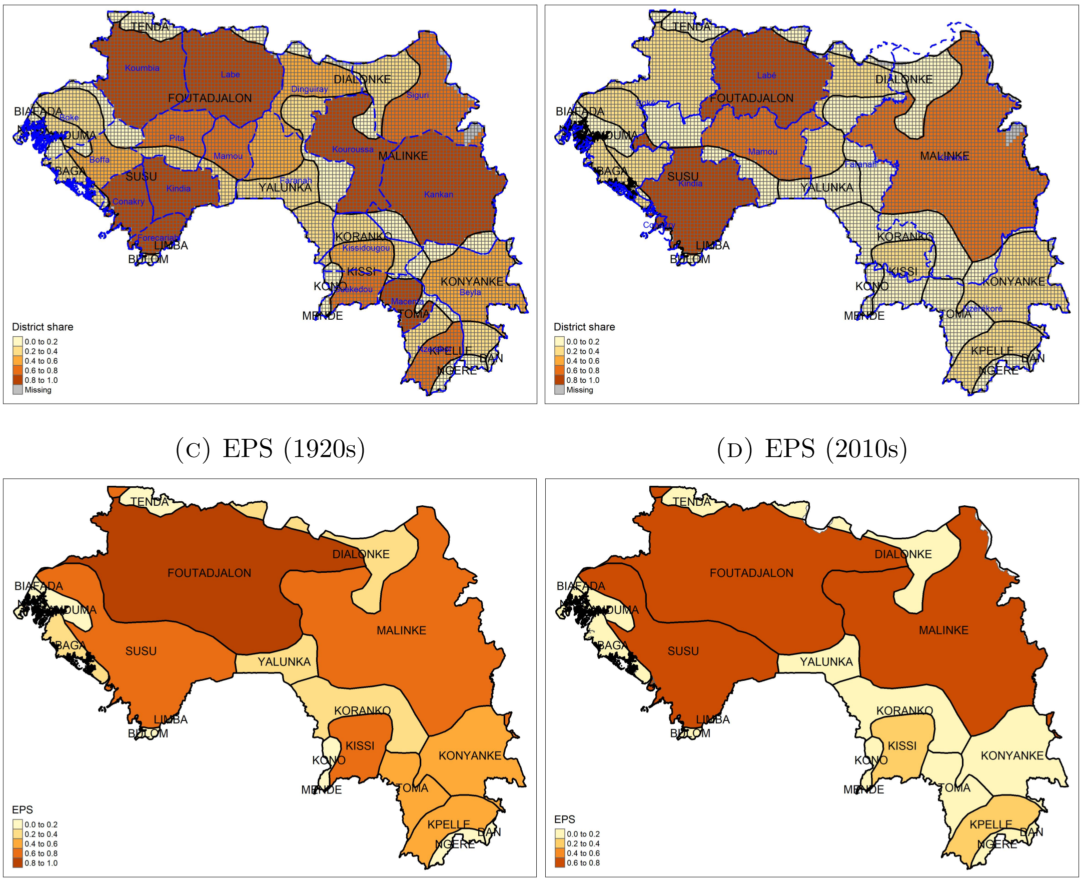


Notes: Figure depicts the distribution of ethnic groups across districts in French Guinea around 1920. Murdock's homeland borders are highlighted in black (including the homeland name at homeland centroid), district borders in blue (name displayed at the district centroid). The imposed grid represents has a 5 arc minute resolution.

Figure IV
Persistence: French Guinea

(A) District shares (1920s) (B) District shares (2010s)

(C) EPS (1920s) (D) EPS (2010s)



Notes: Panel (A) and (B) of the figure depict the district shares PS_{edp} of different ethnic groups across French Guinea in the 1920s and 2010s. Panels (C) and (D) plot the EPS_e for the different time periods across ethnic groups in French Guinea. Homeland borders are highlighted in black. District borders are highlighted in blue.

Table I
Summary statistics on protectorates, colonial districts, and ethnic homelands

Primary colonizer	Number of units	Area (in km^2): mean	std.dev.	Hist. population: mean	std.dev.
Panel A: Protectorates					
All	40	527581	579012	2158664	3939191
France	13	563395	423898	1288928	772441
Panel B: Colonial districts					
All	578	36510	56858	149388	248104
France	193	37948	59046	86818	110983
Panel C: Ethnic homelands within protectorates					
All	1108	19046	35179	78141	201764
France	299	24495	48915	56228	110744

Notes: Table summarizes the size and population of our different unites; Protectorates, districts, and homelands. The historical population estimates reported in Table I are based on the HYDE data introduced below.

Table II
Balancing tests: Expected population share (EPS_{ep}) of ethnic groups/homelands, 1900–1920

	(1)	(2)	(3)	(4)
Panel A: Geography, diseases, conflict and early settlements				
Log population density (HYDE)	0.061***	0.060***	0.059***	0.061***
	(0.013)	(0.016)	(0.013)	(0.015)
Log land area	0.078***	0.078***	0.077***	0.076***
	(0.008)	(0.008)	(0.008)	(0.008)
Rivers	-0.010	-0.011	-0.007	-0.006
	(0.020)	(0.023)	(0.023)	(0.026)
Lakes	0.030	0.034	0.036	0.038
	(0.038)	(0.038)	(0.032)	(0.033)
Elevation		0.001		0.003
		(0.007)		(0.007)
Ruggedness		0.229		0.153
		(0.216)		(0.275)
Crop suitability		0.178		0.146
		(0.318)		(0.322)
Distance to coast		0.001		0.002
		(0.006)		(0.007)
Malaria burden		-0.001		-0.004
		(0.004)		(0.005)
TseTse suitability		0.027		0.034
		(0.019)		(0.022)
Pre-colonial conflict			-0.066*	-0.060
			(0.036)	(0.038)
Log slaves per area			0.046	0.040
			(0.032)	(0.034)
Early explorer routes			0.002	0.009
			(0.024)	(0.023)
Pre-colonial city in 1400			0.023	0.031
			(0.045)	(0.050)
Pre-colonial kingdom			0.020	0.040
			(0.031)	(0.037)
Protectorates	13	13	13	13
Homeland-Protectorates	297	297	297	297
$Within - R^2$	0.498	0.493	0.502	0.498

Notes: The table reports regression results of the weighted district share of each homeland within a colony or protectorate (EPS_{ep}) on geographical, ecological and historical variables measured prior to European colonization. All columns include protectorate-level fixed effects. Homelands are defined as unique protectorate-ethnicity pairs. Standard errors clustered at ethnolinguistic families are provided in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

TABLE III
Baseline results: Homeland

Dependent variable: $\ln Lights_{ep}$	(1)	(2)	(3)	(4)
EPS_{ep}	1.665***	1.409***	1.956***	1.764***
	(0.514)	(0.327)	(0.441)	(0.400)
Protectorate FE	–	–	✓	✓
Full controls	–	✓	–	✓
Homeland-P	297	297	297	297
Within-R^2	0.370	0.505	0.415	0.468

Notes: The table reports the regression results of log light density (VIIRS) on the expected district share (EPS_{ep}) of each homeland within a colony/protectorate. All columns include the baseline controls log homeland population 1900 (HDYE), log area, a river and lake dummy. Full controls refers to the set of controls employed Table II. Standard errors clustered at the culture group in parenthesis. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

TABLE IV
Within homeland & district evidence: District share (ps^e_d)

	Dependent variable: $\ln$ LIGHT-DENSITY$_{iedp}$			
	(1)	(2)	(3)	(4)
PS_{edp}	0.094**	0.110***	0.131***	0.125***
	(0.036)	(0.037)	(0.042)	(0.042)
Homeland controls	✓	✓	–	–
District controls	✓	–	✓	–
Grid-cell controls	✓	✓	✓	✓
Protectorate FE	✓	✓	✓	✓
District FE	–	✓	–	✓
Homeland FE	–	–	✓	✓
Grid-cells	80881	80881	80881	80881
Within-R^2	0.0943	0.0529	0.0499	0.0445

Notes: The table reports the regression results of log light density (VIIRS) on the district share PS_{edp} of 5 arc minute grid-cells located within former French colonies. All columns include the following grid-cell level controls: log pop density (GHSL), log area, any river indicator, any lake indicator, mean elevation, mean ruggedness, mean crop suitability, mean distance to the coast, malaria burden, tse-tse suitability and an explorer indicator. Homeland controls are those reported in Panel A of Table II. District controls mirror the geographic controls of Panel A of Table II calculated for each district. Two-way clustered standard errors clustered at the protectorate specific homeland and district in parenthesis. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table V
Persistence

	Dependent variable:			
	EPS_{ep} 2010s	ln Light-density$_{ep}$	PS_{edp} 2010s	ln Light-*density*$_{cedp}$
	(1)	(2)	(3)	(4)
EPS_{ep} 1920s	0.649***	0.965**		
	(0.067)	(0.452)		
EPS_{ep} 2010s		1.200*		
		(0.607)		
PS_{edp} 1920s			0.430***	0.074*
			(0.127)	(0.044)
PS_{edp} 2010s				0.076**
				(0.031)
Full controls controls	✓	✓	✓	✓
Country-FE	✓	✓	✓	✓
Homeland-FE	–	–	✓	✓
District-FE	–	–	✓	✓
Obs	260	260	60478	60478
Within-R^2	0.770	0.471	0.133	0.0519

Notes: Column 1 of the table reports the regression results of the EPS_{ep} based on current borders on the EPS_{ep} based on 1920 districts. Columns 2 reports the regressions results of log light density (VIIRS) on both the current and the colonial EPS_{ep}. Column 3 & 4 report the corresponding gridcell specifications. Columns 1 & 2 include the full set of controls reported in Table II. Columns 3 & 4 include the following grid-cell level controls: log pop density (GHSL), log area, any river indicator, any lake indicator, mean elevation, mean ruggedness, mean crop suitability, mean distance to the coast, malaria burden, tse-tse suitability and an explorer indicator. Standard errors clustered at the culture group in parenthesis (columns 1 & 2) and are clustered two-way at homeland and district in columns 3 & 4. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

.

Table VI
Ethnic homeland vs. group: Individual level evidence

	Dependent variable:					
	WEALTH-INDEX$_{iep}$		YEARS-SCHOOLING$_{iep}$		INFANT-MORTALITY$_{iep}$	
	(1)	(2)	(3)	(4)	(5)	(6)
EPS_{ep} *place*	0.224***	0.251***	0.105**	0.122***	-1.923	-3.219
	(0.057)	(0.058)	(0.045)	(0.045)	(3.680)	(3.693)
EPS_{ep} *identity*	0.383***	0.408***	0.079***	0.093***	-15.979***	-13.983***
	(0.033)	(0.034)	(0.028)	(0.029)	(2.211)	(2.285)
P-val T-test EPS	.0010	.0088	.7237	.6595	.0009	.01089
Cluster controls	–	✓	–	✓	–	✓
Individual controls	✓	✓	✓	✓	✓	✓
Children controls	–	–	–	–	✓	✓
Protectorate-Wave FE	✓	✓	✓	✓	✓	✓
Homeland-P	213103	208275	98444	93964	625162	616056
Within-R^2	0.450	0.442	0.0223	0.0234	0.0241	0.0247

Notes: Columns 1 & 2 of the table report the results of regressing the DHS wealth index (Bruederle and Hodler, 2018) on EPS_{ep} (both homeland and identity of the respondent). Included controls are geographic characteristics of the DHS cluster: Log population density, log area, dummies for the presence of rivers and lakes, elevation, crop suitability, distance to coast, malaria burden and tsetse fly suitability. Individual level controls: Age, age squared, an urban indicator, a gender dummy, a non-indigenous indicator. Columns 3 & 4 switches the dependent variable to years of schooling, columns 5 & 6 to infant mortality. Note that the level of analysis in columns 5 & 6 are the respondent (mother) children. In columns 5 & 6 we further add children level controls: A gender dummy for the child, a indicator for multiples (e.g., twins or triplets), a set of birth order indicators and a set of decade of birth indicators. Standard errors clustered at the DHS cluster in parenthesis. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table VII

Communication infrastructure: Homeland level

	Dependent variables: No. of				
	$Capital_{1922}$	$Post_{1922}$	$Telegraph_{1922}$	$Post_{1935}$	$Telegraph_{1935}$
	(1)	(2)	(3)	(4)	(5)
EPS_{ep}	2.474***	0.736	0.428	0.607*	0.640*
	(0.407)	(0.453)	(0.680)	(0.317)	(0.335)
$Capital_{1922}$ $(No.)$		0.356***	0.363***	0.148**	0.159**
		(0.091)	(0.099)	(0.067)	(0.064)
$Post_{1922}$				0.060***	
				(0.017)	
$Telegraph_{1922}$					0.066***
					(0.018)
Baseline controls	✓	✓	✓	✓	✓
Geographic controls	✓	✓	✓	✓	✓
Homeland controls	✓	✓	✓	✓	✓
Protectorate FE	✓	✓	✓	✓	✓
Homeland-P	182	182	182	182	182

Notes: Table reports the results of regressions the number of district capitals (column 1) and several counts of communication infrastructure proxies within the homeland on the EPS. All specifications are estimated using poission pseudo maximum likelihood estimators. Full controls refers to the set of controls employed Table II. Standard errors, clustered at the culture group level in parenthesis. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table VIII
Health

	Homeland level			Individual level					
	Dependent variables:								
	ln Hospital-density$_{ep}$			Birth-attendant$_{iep}$			Infant-mortality$_{iep}$		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
EPS_{ep} (*place*)	0.834**		0.986***			0.160***			-3.339
	(0.304)		(0.300)			(0.022)			(4.060)
EPS_{ep} (*identity*)						0.032*			-15.058***
						(0.018)			(3.148)
$ln\ HealthPersonal_{1920}$ (homeland)		0.165	0.186*		-0.005	-0.008		0.032	-1.027
		(0.113)	(0.106)		(0.005)	(0.005)		(0.985)	(1.020)
$ln\ HealthPersonal_{1920}$ (identity)					0.014***	0.011**		-1.316*	1.344
					(0.004)	(0.005)		(0.753)	(0.998)
$ln\ Hospital-density_{ep}$				0.056***	0.055***	0.056***	-2.139**	-2.096**	-2.240**
				(0.006)	(0.006)	(0.006)	(1.057)	(1.058)	(1.056)
Protectorate (Wave) FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Full controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
Cluster controls	–	–	–	✓	✓	✓	✓	✓	✓
Individual controls	–	–	–	✓	✓	✓	✓	✓	✓
Children controls	–	–	–	–	–	–	✓	✓	✓
Obs.	182	182	182	131439	131439	131380	548445	548445	548118
Within-R^2	0.870	0.872	0.877	0.176	0.176	0.179	0.0254	0.0254	0.0255

Notes: Columns 1 to 3 of the table report the regression results of the *EPS* homeland on log hospital density. Columns 1 to 3 include the baseline controls log homeland population 1900 (HDYE), log area, a river and lake dummy. Full controls refers to the set of controls employed Table II. Standard errors clustered at the culture group in parenthesis. In columns 4 to 9 we report the results of regressing either a professional birth attendant dummy (columns 4 to 6) or the log of infant mortality (columns 7 to 9) on the expected population share of both homelands (*EPS homeland*)and groups (*EPS identity.*) Included controls are geographic characteristics of the DHS cluster: Log population density, log area, dummies for the presence of rivers and lakes, elevation, crop suitability, distance to coast, malaria burden and tse-tse fly suitability. Individual level controls: Age, age squared, an urban indicator, a gender dummy, a non-indigenous indicator. Note that the level of analysis in columns 4 to 6 are the respondent (mother) children. In columns 4 to 6 we further add children level controls: A gender dummy for the child, a indicator for multiples (e.g., twins or triplets), a set of birth order indicators and a set of decade of birth indicators. Standard errors in columns 4to 9 are clustered at the DHS cluster in parenthesis. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table IX
Education

	Dependent variable: ln Years-schooling$_{iep}$			
	(1)	(2)	(3)	(4)
EPS_{ep} (*place*)		0.106**		0.138**
		(0.050)		(0.064)
EPS_{ep} (*identity*)		0.054		-0.024
		(0.044)		(0.056)
Avg. teacher (*place*)	0.005	0.005	0.010**	0.009*
	(0.004)	(0.004)	(0.005)	(0.005)
Avg. teacher (*identity*)	0.014***	0.010*	0.022***	0.023***
	(0.004)	(0.005)	(0.005)	(0.006)
EPS_{ep} (*place*) $\times$*late cohort*				-0.059
				(0.072)
EPS_{ep} (*identity*) $\times$ *late cohort*				0.167**
				(0.067)
Avg. teacher (*place*) $\times$ *late cohort*			-0.007	-0.002
			(0.005)	(0.006)
Avg. teacher (*identity*) $\times$ *late cohort*			-0.015**	-0.028***
			(0.006)	(0.008)
Homeland controls	✓	✓	–	–
District controls	✓	–	✓	–
Grid-cell controls	✓	✓	✓	✓
Protectorate FE	✓	✓	✓	✓
District FE	–	✓	–	✓
Homeland FE	–	–	✓	✓
Grid-cells	80881	80881	80881	80881
Within-R^2	0.0943	0.0529	0.0499	0.0445

Notes: The table reports the regression results of years of schooling on the expected population share of both homelands (*EPS homeland*) and groups (*EPS identity*), as well as the avg. number of teachers in the 1920s of both homeland and the homeland an respondent identifies with. In column 3 and 4 we include interaction of those variables with the *latecohort* defined as people being born after the mean birth-year, which is 1980 in our sample. Included controls are geographic characteristics of the DHS cluster: Log population density, log area, dummies for the presence of rivers and lakes, elevation, crop suitability, distance to coast, malaria burden and tse-tse fly suitability. Individual level controls: Age, age squared, an urban indicator, a gender dummy, a non-indigenous indicator. Standard errors in columns 4to 9 are clustered at the DHS cluster in parenthesis. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table X
Migration

	Dependent variables:			
	Prob. Migrate			$Non-Indigenous$
	All	Inside of homeland	Outside of homeland	
	(1)	(2)	(3)	(4)
Panel A: All				
PS_{ed} Birthplace	-0.083***	-0.026*	-0.057***	
	(0.029)	(0.015)	(0.016)	
EPS_{ep} *identity*				-0.806***
				(0.063)
Individuals	3.077e+06	3.077e+06	3.077e+06	253470
Panel B: Urban only				
PS_{ed}	-0.022*	0.005	-0.028***	
	(0.012)	(0.009)	(0.009)	
EPS_{ep} *identity*				-0.396***
				(0.083)
Individuals	3.077e+06	3.077e+06	3.077e+06	86561
Panel C: Born before 1960				
PS_{ed} Birthplace	-0.083	-0.005	-0.078**	
	(0.050)	(0.025)	(0.033)	
EPS_{ep} *identity*				-0.847***
				(0.089)
Individuals	244212	244212	244212	31027
Individual controls	✓	✓	✓	✓
Country-Group FE	✓	✓	✓	✓

Notes: Columns 1 to 3 of the table reports the results of a regression predicting the probability that an individual migrated (defined as not living in the district in which the respondent was born) across different dimensions on the district share of he persons birth district. The sample consists of IPUMS international respondents for Benin, Burkina Faso, Guinea, Mali, and Senegal. Standard errors clustered at the homeland-group level. Columns 4 present results from regressing a dummy variable for residing outside of ones homeland on the expected population share of a persons group (*EPS identity*) using the DHS data. Included controls in columns 1 to 3 are a sex indicator, as well as birth year dummies. Included controls in columns 4 are geographic characteristics of the DHS cluster: Log population density, log area, dummies for the presence of rivers and lakes, elevation, crop suitability, distance to coast, malaria burden and tsetse fly suitability. Individual level controls: Age, age squared, an urban indicator, a gender dummy, a non-indigenous indicator, a indicator if the respondents identifies with the homeland on which the capital city is located. Standard errors are clustered the ethnic group level (columns 1 to 3) and at the DHS cluster (columns 4) in parenthesis.

TABLE XI
External validity

	Dependent variables:					
	ln LIGHT-DENSITY$_{ep}$		WEALTH-INDEX$_{iep}$		ln LIGHT-DENSITY$_{iedp}$	
	(1)	(2)	(3)	(4)	(5)	(6)
EPS_{ep}	1.895***	1.623***				
	(0.450)	(0.253)				
EPS_{ep} *(place)*			0.266***	0.252***		
			(0.045)	(0.033)		
EPS_{ep} *(identity)*			0.287***	0.318***		
			(0.034)	(0.023)		
PS_{edp}					0.240***	0.172***
					(0.059)	(0.029)
Colonizer	GBR	All	GBR	All	GBR	All
Controls	✓	✓	✓	✓	✓	✓
Protectorate FE	✓	✓	✓	✓	✓	✓
District FE	–	–	–	–	✓	✓
Homeland FE	–	–	–	–	✓	✓
Protectorate -Wave FE	–	–	✓	✓	–	–
Obs	410	1121	333484	596892	83244	261442
Within-R^2	0.455	0.435	0.380	0.405	0.0849	0.0659

Notes: Columns 1 and 2 replicate column 4 of Table III, columns 3 and 4 replicate column 2 of Table VI, and columns 5 and 6 replicate column 4 of Table IV for the for the British and full sample respectively. Standard errors in parenthesis are clustered at the culture group level in columns 1 & 2, at the DHS cluster in columns 3 & 4 and two-way clustered at the district and homeland in columns 5 & 6. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Online appendix

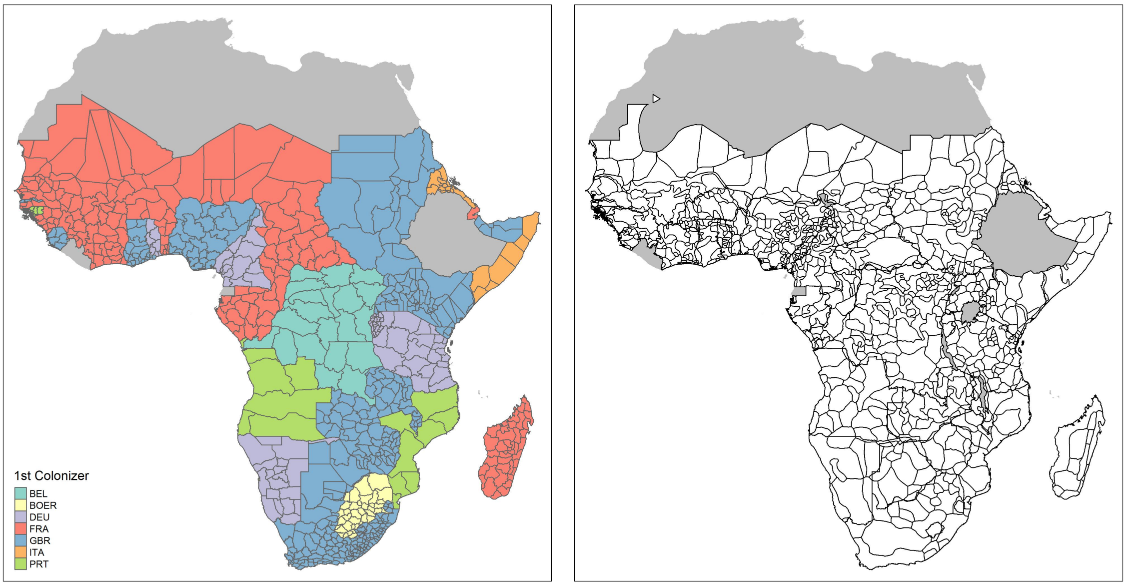
Contents

A. Additional Figures and Tables

A-1. Figures

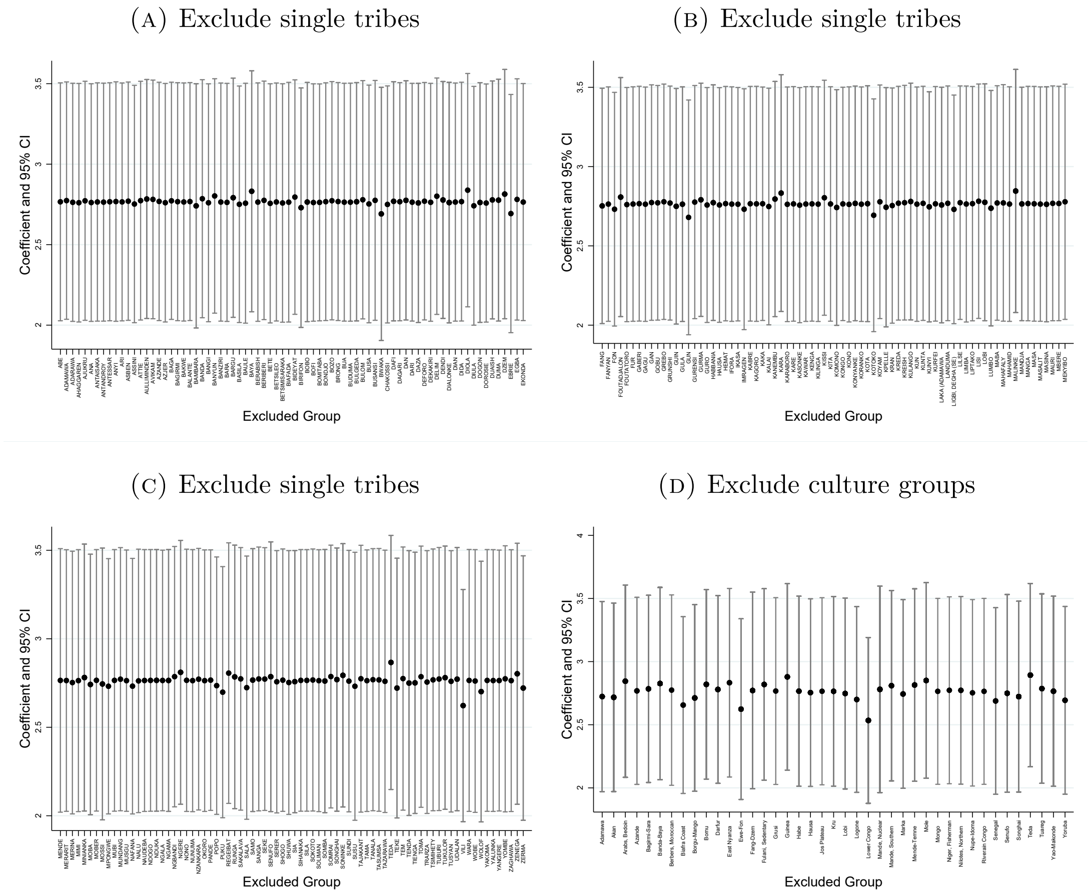
FIGURE A-1
Colonial districts and ancestral homelands (Murdock)

(A) Colonial districts (B) Ancestral homelands



Notes: Panel (A) plots our sample of colonial districts, by their initial colonizer. Panel (B) plots the set of ethnic homelands based on (Murdock, 1959) in our sample.

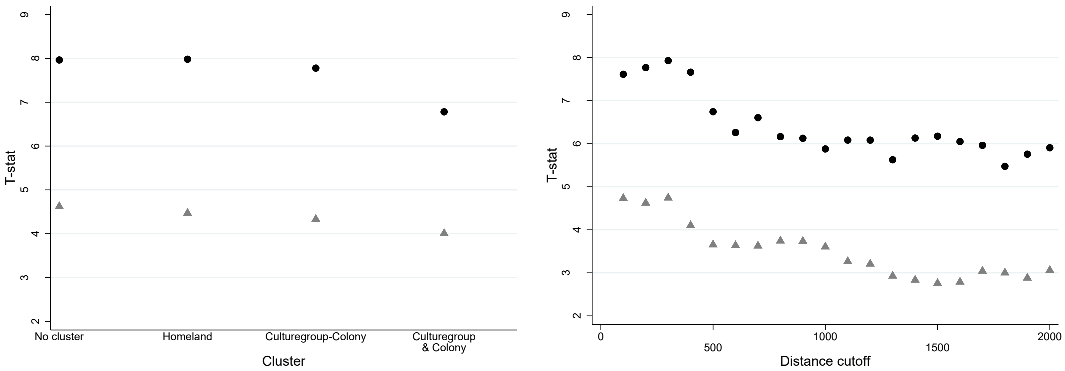
Figure A-2
Exclude: Groups and culture groups



Notes: The figure reports point coefficients and 95% confidence intervals for the effect of the expected district population share on luminosity based on our main homeland specification. Panel (A) to (C) replicate column (4) of Table III excluding one Murdock group at the time. Panel (D) replicates column 4 of Table III excluding one culture group at the time. 95% CI are based on standard errors clustered at the culture group level.

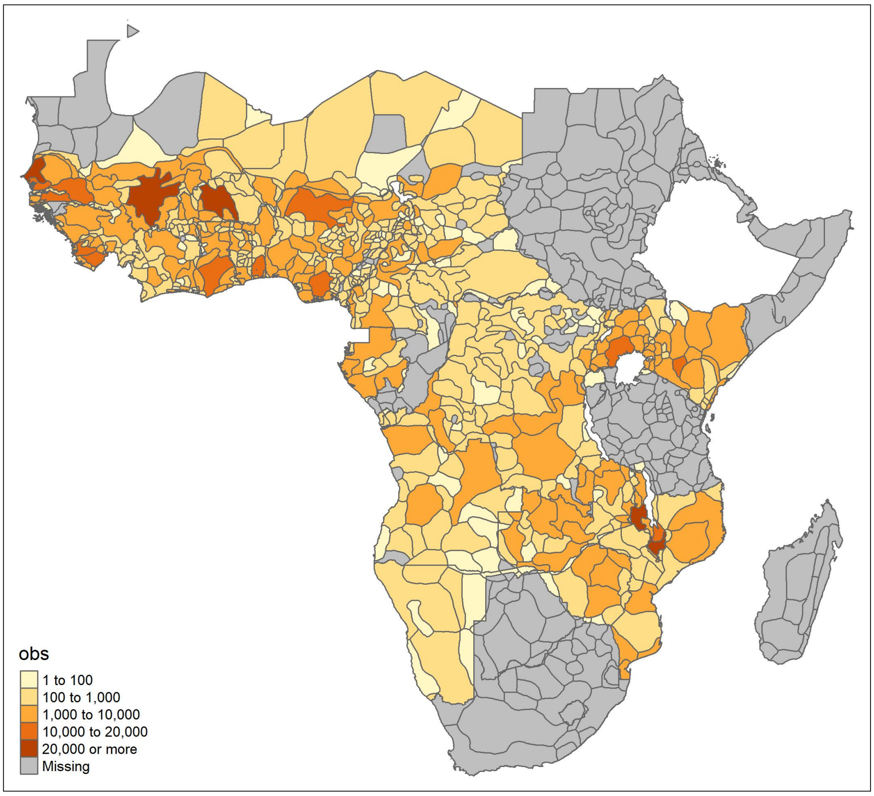
Figure A-3
Baseline results: Alternative standard error clustering

(A) Clustering on different units (B) Arbitrary spatial clustering



Notes: Panel (A) of the figure reports the t-stats for columns 1 (black dots) and 4 (grey triangles) of Table III using different cluster units. Panel (B) reports the t-stats for columns 1 (black dots) and 4 (grey triangles) of Table III employing different distance cutoffs for spatial clustering.

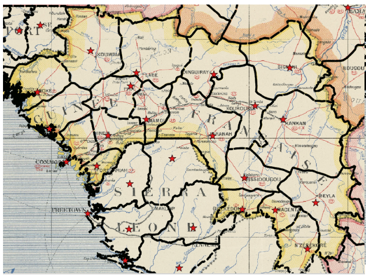
Figure A-4
DHS clusters across homeland-protectorates



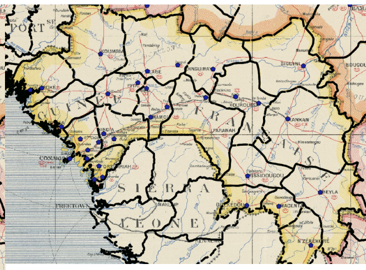
Notes: The figure depicts the distribution of DHS respondents across homelands within protectorates.

Figure A-5
French Guinea: Colonial capitals and infrastructure

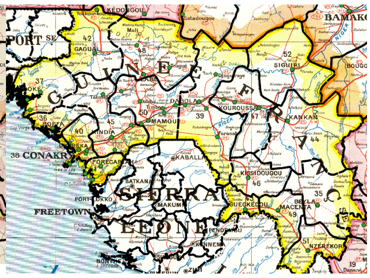
(A) Capitals 1922



(B) Infrastructure 1922



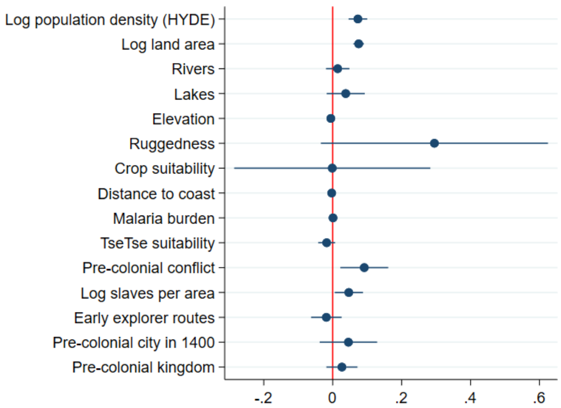
(C) Infrastructure 1935



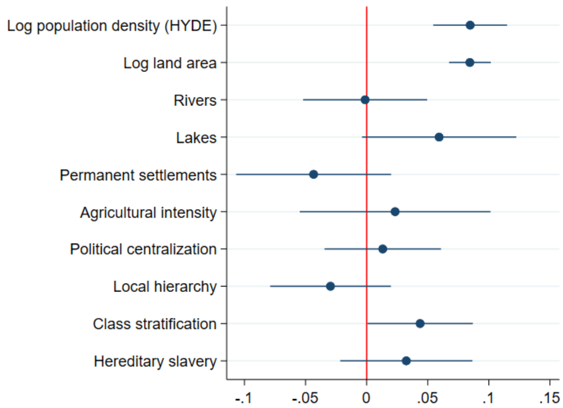
Notes: Panel (A) to (C) depict the distribution of capitals and communication infrastructure in French Guinea as shown on official survey maps of French West Africa published in 1922 and 1935.

Figure A-6
Bi-variate balancing: British and all former colonies

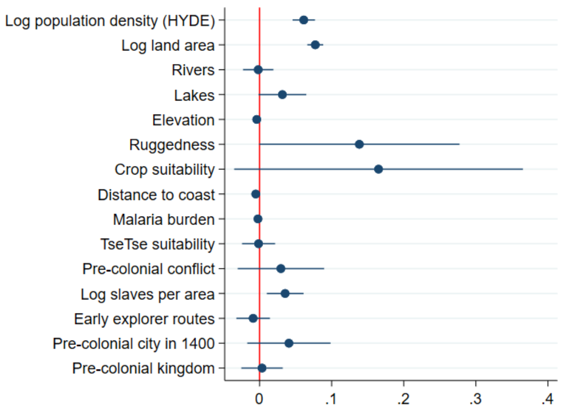
(A) Geography and colonial experience



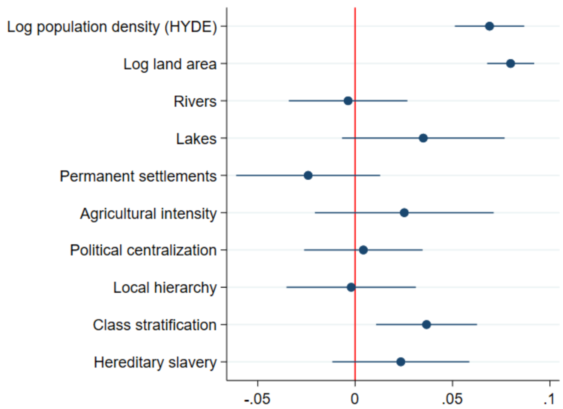
(B) Ethnographic Atlas



(C) Geography and colonial experience



(D) Ethnographic Atlas



Notes: The figure depicts the results of bi-variate balancing test for British (Panel A and B) and all former colonies within Sub-Saharan Africa, similar to the results represented for the former French colonies in Table A-2 and Table A-3. Confidence intervals are 95% based on clustered standard errors at the culture group level.

A-2. Tables

Table A-1
Balancing tests: Expected population share (EPS_{ep}) of ethnic groups/homelands, 1900–1920

	(1)	(2)	(3)	(4)
Pre-colonial settlement patterns and political organization				
Log population density (HYDE)	0.051***	0.056***	0.057***	0.052***
	(0.017)	(0.015)	(0.016)	(0.018)
Log land area	0.081***	0.079***	0.078***	0.079***
	(0.009)	(0.008)	(0.008)	(0.009)
Rivers	-0.039	-0.035	-0.031	-0.048
	(0.040)	(0.036)	(0.036)	(0.038)
Lakes	0.017	0.016	0.019	0.014
	(0.047)	(0.046)	(0.045)	(0.047)
Permanent settlements	0.038			0.013
	(0.037)			(0.060)
Agricultural intensity	0.031			0.050
	(0.040)			(0.041)
Political centralization		0.033		0.030
		(0.031)		(0.038)
Local hierarchy		0.027		0.016
		(0.042)		(0.053)
Class stratification			-0.008	-0.022
			(0.029)	(0.027)
Hereditary slavery			0.041	0.040
			(0.032)	(0.033)
Protectorates	13	13	13	13
Homeland-P	168	168	168	168
Within-R^2	0.500	0.501	0.501	0.495

Notes: The table reports regression results of the weighted district share of each homeland within a colony or protectorate (EPS_{ep}) on binary variables derived from Murdock's (1967) Ethnographic Atlas. Settlement patterns refers to homelands with permanent settlements, ranging from 'neighborhoods of dispersed family homesteads' to 'complex settlements'. Agricultural intensity indicates homelands whose agricultural activities range from 'extensive or shifting agriculture' to 'intensive irrigated agriculture'. Political centralization is unity for communities whose jurisdictional hierarchy beyond the local community are classified as paramount chiefdoms or part of large states. Local hierarchy indicates whether the Ethnographic Atlas reports at least three levels of local jurisdictional hierarchies (e.g. nuclear family, extended family, clan/barrio, and village). Class stratification indication whether the ethnic group has classes based on 'wealth distinctions' or a range of more complex structures. Hereditary slavery is a binary variable for whether slavery was 'hereditary and socially significant'. All columns include protectorate-level fixed effects. Homelands are defined as unique protectorate-ethnicity pairs. Standard errors clustered at ethnolinguistic families are provided in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

TABLE A-2

Balancing test: French Africa – single covariates

	DV: EPS_{ep}										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Log pop density (HYDE)	0.063	0.061	0.059	0.060	0.062	0.061	0.062	0.058	0.061	0.061	0.062
	[0.013]***	[0.013]***	[0.015]***	[0.016]***	[0.013]***	[0.014]***	[0.013]***	[0.013]***	[0.014]***	[0.013]***	[0.013]***
Log land area	0.078	0.078	0.078	0.078	0.078	0.079	0.078	0.078	0.078	0.078	0.077
	[0.008]***	[0.008]***	[0.008]***	[0.008]***	[0.008]***	[0.008]***	[0.008]***	[0.008]***	[0.008]***	[0.008]***	[0.008]***
Rivers	-0.009	-0.010	-0.008	-0.012	-0.009	-0.014	-0.009	-0.008	-0.010	-0.010	-0.010
	[0.020]	[0.020]	[0.021]	[0.021]	[0.021]	[0.021]	[0.021]	[0.021]	[0.021]	[0.021]	[0.020]
Lakes	0.030	0.032	0.031	0.030	0.030	0.031	0.036	0.026	0.030	0.029	0.029
	[0.038]	[0.038]	[0.038]	[0.038]	[0.038]	[0.037]	[0.034]	[0.035]	[0.038]	[0.039]	[0.038]
Elevation	0.004										
	[0.006]										
Ruggedness		0.261									
		[0.185]									
Crop suitability			0.201								
			[0.293]								
Distance to coast				-0.002							
				[0.005]							
Malaria burden					-0.087						
					[0.312]						
TseTse suitability						0.030					
						[0.018]					
Pre-colonial conflict							-0.033				
							[0.041]				
Log slaves per area								0.040			
								[0.028]			
Early explorer routes									0.004		
									[0.020]		
Pre-colonial city in 1400										0.002	
										[0.043]	
Pre-colonial kingdom											0.024
											[0.024]
Protectorate FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Protectorates	13	13	13	13	13	13	13	13	13	13	13
Homeland-P	297	297	297	297	297	297	297	297	297	297	297
Within-R^2	0.497	0.498	0.497	0.496	0.496	0.500	0.497	0.504	0.496	0.496	0.498

Notes: The table reports regression results of the weighted district share of each homeland within a colony or protectorate (EPS_{ep}) on geographical, ecological and historical variables measured prior to European colonization. All columns include protectorate-level fixed effects. Homelands are defined as unique protectorate-ethnicity pairs. Standard errors clustered at ethnolinguistic families are provided in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A-3
Balancing test: French Africa − single covariates (Murdock Atlas)

	DV: EPS_{ep}					
	(1)	(2)	(3)	(4)	(5)	(6)
Log pop density (HYDE)	0.051	0.053	0.057	0.055	0.057	0.057
	[0.017]***	[0.016]***	[0.015]***	[0.015]***	[0.016]***	[0.015]***
Log land area	0.081	0.080	0.079	0.079	0.079	0.078
	[0.009]***	[0.009]***	[0.008]***	[0.008]***	[0.008]***	[0.009]***
Rivers	-0.038	-0.036	-0.031	-0.032	-0.027	-0.030
	[0.039]	[0.039]	[0.034]	[0.039]	[0.036]	[0.036]
Lakes	0.017	0.019	0.017	0.020	0.020	0.018
	[0.047]	[0.046]	[0.046]	[0.045]	[0.044]	[0.045]
Permanent settlements	0.058					
	[0.044]					
Agricultural intensity		0.057				
		[0.046]				
Political centralization			0.036			
			[0.031]			
Local hierarchy				0.032		
				[0.043]		
Class stratification					0.012	
					[0.031]	
Hereditary slavery						0.038
						[0.031]
Protectorate FE	✓	✓	✓	✓	✓	✓
Protectorates	13	13	13	13	13	13
Homeland-P	168	168	168	168	168	168
Within-R^2	0.503	0.503	0.503	0.501	0.500	0.504

Notes: The table reports regression results of the weighted district share of each homeland within a colony or protectorate (EPS_{ep}) on binary variables derived from Murdock's (1967) Ethnographic Atlas. Settlement patterns refers to homelands with permanent settlements, ranging from 'neighborhoods of dispersed family homesteads' to 'complex settlements'. Agricultural intensity indicates homelands whose agricultural activities range from 'extensive or shifting agriculture' to 'intensive irrigated agriculture'. Political centralization is unity for communities whose jurisdictional hierarchy beyond the local community are classified as paramount chiefdoms or part of large states. Local hierarchy indicates whether the Ethnographic Atlas reports at least three levels of local jurisdictional hierarchies (e.g. nuclear family, extended family, clan/barrio, and village). Class stratification indication whether the ethnic group has classes based on 'wealth distinctions' or a range of more complex structures. Hereditary slavery is a binary variable for whether slavery was 'hereditary and socially significant'. All columns include protectorate-level fixed effects. Homelands are defined as unique protectorate-ethnicity pairs. Standard errors clustered at ethnolinguistic families are provided in parentheses.* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

TABLE A-4

Baseline results: Alternative independent variables

	DV - $\ln Lights_{ep}$			
	(1)	(2)	(3)	(4)
Expected Majority	1.001***	0.960***		
	(0.293)	(0.272)		
Expected largest group			0.647***	0.593**
			(0.234)	(0.222)
Protectorate-FE	✓	✓	✓	✓
Full controls	–	✓	✓	✓
Homeland-P	297	297	297	297
Within-R^2	0.400	0.462	0.377	0.438

Notes: The table reports the results of regressing the log of light density (VIIRS) on the expected local majority (columns 1 & 2) and the expected local largest group (columns 3 & 4). Both the local majority and largest groups are dummy variables for groups having a PS_{ed} above 50% or the largest pop share within a district. All columns include the baseline controls log homeland population 1900 (HDYE), ln area, a river and lake dummy. Full controls refers to the set of controls employed Table II. Standard errors clustered at the culture group in parenthesis. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

TABLE A-5

Baseline results: Alternative dependent variable

	Dependent variable: $\ln Lights_{ep}$ *DMSP*			
	(1)	(2)	(3)	(4)
EPS_{ep}	1.871***	1.531***	1.975***	1.717***
	(0.504)	(0.334)	(0.416)	(0.358)
Protectorate FE	–	–	✓	✓
Full controls	–	✓	–	✓
Homeland-P	297	297	297	297
Within-R^2	0.382	0.533	0.438	0.502

Notes: This table replicates the main results table using the averaged log of light density based on the DMSP-OLS data (1992-2013) as the deponent variables. As before, all columns include protectorate fixed effects and the baseline controls log homeland population 1900 (HDYE), ln area, a river and a lake dummy. Full controls refers to the set of controls employed Table II. Standard errors clustered at the culture group level in parenthesis. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A-6

Baseline results: National vs. local majorities

	DV - ln Light-density$_{ep}$			
	(1)	(2)	(3)	(4)
EPS_{ep}	1.727***	1.430***	1.587***	1.357***
	(0.504)	(0.461)	(0.518)	(0.478)
National pop share	1.367	1.993*	-4.884	-1.671
	(1.205)	(1.179)	(3.179)	(3.904)
EPS_{ep} × *National pop share*			7.995*	4.690
			(4.090)	(4.932)
Protectorate-FE	✓	✓	✓	✓
Full controls	–	✓	–	✓
Homeland-P	297	297	297	297
Within-R^2	0.416	0.472	0.418	0.472

Notes: The table reports the regression results of log light density (VIIRS) on the expected district share (EPS) of each homeland within a colony/protectorate. All columns include the baseline controls log homeland population 1900 (HDYE), log area, a river and lake dummy. Full controls refers to the set of controls employed Table II. Standard errors clustered at the culture group in parenthesis. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A-7

Baseline results: Population and area percentiles

	DV - ln Light-density$_{ep}$			
	Population bins controls		*Area bins controls*	
	(1)	(2)	(3)	(4)
EPS_{ep}	1.193**	1.116**	2.535***	2.248***
	(0.456)	(0.422)	(0.493)	(0.486)
Protectorate-FE	✓	✓	✓	✓
Full controls	–	✓	–	✓
Homeland-Protectorates	297	297	297	297
Within-R^2	0.543	0.568	0.421	0.468

Notes: The table reports the regression results of log light density (VIIRS) on the *EPS* of each homeland within a colony or protectorate. All columns include dummies for each population density percentile (the 1st percentile is the omitted category) alongside the baseline controls the log land area, a river and lake dummy (columns 1 & 2). Columns 3 & 4 uses area bins instead of population density bins alongside the baseline controls log population density, and the river and lake dummy. Full controls refers to the set of controls employed Table II. Standard errors clustered at the culture group in parenthesis. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A-8
Baseline results: Different samples

	DV - ln Light-density$_{ep}$							
	Diverse districts		*Exl. colonial capital*		*Exl. current capital*		*Exl. split homelands*	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
EPS_{ep}	1.768***	1.651***	1.315***	1.102***	1.325***	1.127***	2.689***	2.592***
	(0.427)	(0.385)	(0.430)	(0.366)	(0.420)	(0.355)	(0.443)	(0.438)
Protectorate FE	✓	✓	✓	✓	✓	✓	✓	✓
Full controls	–	✓	–	✓	–	✓	–	✓
Homeland-P	297	297	285	285	284	284	142	142
Within-R^2	0.403	0.461	0.322	0.402	0.323	0.402	0.440	0.448

Notes: Columns 1 & 2 of the table reports the regression results of log light density (VIIRS) on the expected district share (EPS) based exclusively on diverse districts (districts with at least two groups) of each homeland within a colony/protectorate. Columns 2 & 3 exclude homelands on which the colonial capital is located. Columns 4 & 5 exclude homelands on which the current capital is located. Columns 7 & 8 exclude homelands split by a national border (folloing Michalopoulos and Papaioannou, 2016). All columns include the baseline controls log homeland population 1900 (HDYE), log area, a river and lake dummy. Full controls refers to the set of controls employed Table II. Standard errors clustered at the culture group in parenthesis. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A-9
Local majorities & cash-crops

	Dependent Variable: $\ln Lights_{ep}$								
Cash-crop interaction:	– (1)	Bananas (2)	Cocoa (3)	Coffee (4)	Cotton (5)	Groundnuts (6)	Palmoil (7)	Tea (8)	Tobacco (9)
EPS_{ep}	1.607***	1.711***	1.693***	1.698***	1.679***	1.709***	1.736***	1.683***	1.774***
	(0.381)	(0.375)	(0.372)	(0.377)	(0.383)	(0.407)	(0.372)	(0.358)	(0.403)
Cash crop		0.178	0.186	0.162	0.209*	0.244**	0.017	0.409***	0.116
		(0.154)	(0.125)	(0.144)	(0.109)	(0.094)	(0.157)	(0.146)	(0.105)
EPS_{ep} × Cash crop		-0.405	-0.404	-0.444	-0.119	-0.432**	-0.519	-0.358	-0.286
		(0.315)	(0.327)	(0.347)	(0.230)	(0.178)	(0.355)	(0.309)	(0.197)
Protectorate FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Baseline controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
Geographic controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
Homeland controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
Homeland-P	297	297	297	297	297	297	297	297	297
Within-R^2	0.493	0.469	0.470	0.470	0.474	0.477	0.473	0.486	0.467

Notes: The table replicates baseline results (column 4 of Table III) controlling for cash crops suitability in column 1, and adding interactions of single cash crop suitability in columns 2 to 9. Cash crop interactions are standardized (cashcrop - mean(cashcrop) /sd(cashcrop)). The cash crop data are obtained from the GAEZ and described in detail in Section D-2. Standard errors clustered at the culture group in parenthesis. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A-10

Within homeland & district evidence: Majority & largest group

	Dependent variable: ln Light-density$_{iedp}$			
	(1)	(2)	(3)	(4)
Panel (A)	Majority treatment ($PS_{edp} > 0.5$)			
Majority	0.032	0.046*	0.060**	0.057**
	(0.022)	(0.024)	(0.026)	(0.026)
Within-R^2	0.0937	0.130	0.0493	0.0443
Panel (B)	Largest group treatment (within district)			
Largest group	0.035	0.042**	0.061**	0.042**
	(0.023)	(0.019)	(0.026)	(0.017)
Within-R^2	0.0938	0.130	0.0493	0.0442
Grid-cells	80949	80949	80949	80949
Homeland controls	✓	✓	–	–
District controls	✓	–	✓	–
Grid-cell controls	✓	✓	✓	✓
Protectorate FE	✓	✓	✓	✓
District FE	–	✓	–	✓
Homeland FE	–	–	✓	✓

Notes: The table reports the regression results of log light density (VIIRS) on the district share PS_{edp} of 5am grid-cells within former French colonies. All columns include the following grid-cell level controls: log pop density (GHSL), log area, any river indicator, any lake indicator, mean elevation, mean ruggedness, mean crop suitability, mean distance to the coast, malaria burden, tse-tse suitability and an explorer indicator. Homeland controls are those reported in Panel A of Table II. District controls mirror the geographic controls of Panel A of Table II calculated for each district. Two-way clustered standard errors clustered at the protectorate specific homeland and district in parenthesis. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

TABLE A-11
Within homeland & district evidence: Alternative DV

	(1)	(2)	(3)	(4)
Panel (A)	*Dependent variable:* LIT_{iedp}			
PS_{edp}	0.025**	0.028**	0.033***	0.023***
	(0.010)	(0.012)	(0.010)	(0.007)
Grid-cells	80881	80881	80881	80881
Within-R^2	0.109	0.115	0.0516	0.0475
Panel (B)	*Dependent variable:* $\ln \text{LIGHT-DENSITY}_{iedp}(DMSP)$			
PS_{edp}	0.129**	0.151***	0.138**	0.158***
	(0.055)	(0.056)	(0.060)	(0.055)
Grid-cells	80881	80881	80881	80881
Within-R^2	0.110	0.111	0.0430	0.0386
Homeland controls	✓	✓	–	–
District controls	✓	–	✓	–
Grid-cell controls	✓	✓	✓	✓
Protectorate FE	✓	✓	✓	✓
District FE	–	✓	–	✓
Homeland FE	–	–	✓	✓

Notes: Panel (A) of the table reports the regression results of a is lit dummy (VIIRS) on the district share PS_{edp} of 5am grid-cells within former French colonies. Panel(B) used the log lof light density (DMSP). Panel (C) uses the DHS wealth index. All columns include the following grid-cell level controls: log pop density (GHSL), log area, any river indicator, any lake indicator, mean elevation, mean ruggedness, mean crop suitability, mean distance to the coast, malaria burden, tse-tse suitability and an explorer indicator. Homeland controls are those reported in Panel A of Table II. District controls mirror the geographic controls of Panel A of Table II calculated for each district. Two-way clustered standard errors clustered at the homeland and district in parenthesis. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A-12
Within homeland & district evidence: Distance from homeland centroid

	Dependent variable: ln Population-density$_{iedp}$			
	(1)	(2)	(3)	(4)
Log distance to homeland centroid	0.227	0.236	0.239	0.227
	(0.190)	(0.187)	(0.182)	(0.190)
Grid-cells	81297	80949	81297	81297
Within-R^2	0.0925	0.202	0.145	0.0925
Homeland controls	✓	✓	–	–
District controls	✓	–	✓	–
Grid-cell controls	✓	✓	✓	✓
Protectorate FE	✓	✓	✓	✓
District FE	–	✓	–	✓
Homeland FE	–	–	✓	✓

Notes: The table reports the regression results of log population density (HYDE) on the log distance from the homeland centroid of 5am grid-cells within former French colonies. All columns include the following grid-cell level controls: log area, any river indicator, any lake indicator, mean elevation, mean ruggedness, mean crop suitability, mean distance to the coast, malaria burden, tse-tse suitability and an explorer indicator. Two-way clustered standard errors clustered at the protectorate specific homeland and district in parenthesis. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

TABLE A-13

Within homeland & district evidence: National majorities

	Dependent variable: ln LIGHT-DENSITY$_{ep}$			
	(1)	(2)	(3)	(4)
Panel (A): National population share:				
PS_{edp}	0.063**	0.098***	0.086***	0.104***
	(0.030)	(0.030)	(0.031)	(0.028)
PS_{edp} × *National population share*	0.236	0.458**	0.328**	0.380
	(0.152)	(0.179)	(0.165)	(0.247)
National population share	0.139	-0.086		
	(0.100)	(0.076)		
Within-R^2	0.0988	0.0543	0.0508	0.0448
Panel (B): EPS:				
PS_{edp}	0.094**	0.119***	0.128***	0.137***
	(0.045)	(0.041)	(0.044)	(0.035)
$PS_{edp} \times EPS_{ep}$	-0.015	-0.016	0.005	-0.023
	(0.042)	(0.047)	(0.048)	(0.066)
EPS_{ep}	0.014	0.005		
	(0.019)	(0.018)		
Within-R^2	0.0943	0.0529	0.0499	0.0445
Grid-cells	80881	82844	260444	80881
Grid-cell controls	✓	✓	✓	✓
Homeland controls	✓	✓	–	–
District controls	✓	–	✓	–
Grid-cell controls	✓	✓	✓	✓
Protectorate FE	✓	✓	✓	✓
District FE	–	✓	–	✓
Homeland FE	–	–	✓	✓

Notes: The table reports the regression results of log light density (VIIRS) on the district share PS_{edp} of 5am grid-cells. All columns include the following grid-cell level controls: log pop density (GHSL), log area, any river indicator, any lake indicator, mean elevation, mean ruggedness, mean crop suitability, mean distance to the coast, malaria burden, tse-tse suitability and an explorer indicator. All interactions of the district share PS_{edp} with the interaction variable $\tilde{z}$ are standardized such that $\tilde{z} \equiv (z - \bar{z})/\sigma_z$. Two-way clustered standard errors clustered at the protectorate specific homeland and district in parenthesis. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A-14
Exclude grid cells close to capitals

	Dependent variable: ln LIGHT-DENSITY$_{iedp}$							
	Within 10km of district capital				*Within 20km of national capital*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
PS_{edp}	0.072**	0.073**	0.106***	0.112***	0.051*	0.048*	0.083**	0.094**
	(0.033)	(0.032)	(0.039)	(0.041)	(0.030)	(0.028)	(0.036)	(0.037)
Homeland controls	✓	✓	–	–	–	✓	–	–
District controls	✓	–	✓	–	–	–	✓	–
Cell controls	✓	✓	✓	✓	✓	✓	✓	✓
District FE	–	✓	–	✓	–	✓	–	✓
Homeland FE	–	–	✓	✓	–	–	✓	✓
Obs	80395	80395	80395	80853	78993	78993	78993	80769
Within-R^2	0.0360	0.0430	0.0417	0.0427	0.0343	0.0403	0.0396	0.0415

Notes: Table reports the results of regressing the log of light density (VIIRS) on the district share PS_{edp}, excluding grid-cells within 10km of a colonial capitals (columns 1 to 5) and within 20km of a colonial capital (columns 5 to 8). All columns include the following grid-cell level controls: log pop density (GHSL), log area, any river indicator, any lake indicator, mean elevation, mean ruggedness, mean crop suitability, mean distance to the coast, malaria burden, tse-tse suitability and an explorer indicator. Two-way-clustered standard errors, clustered at the district and homeland in parenthesis. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

TABLE A-15

Persistence: *EPS* & *PS* based on current population

	Dependent variable:			
	EPS_{ep} 2010s	ln LIGHT-DENSITY$_{ep}$	PS_{edp} 2010s	ln LIGHT-*density*$_{cedp}$
	(1)	(2)	(3)	(4)
EPS_{ep} 1920s	0.696***	1.182***		
	(0.064)	(0.364)		
EPS_{ep} 2010s		0.763*		
		(0.411)		
PS_{edp} 1920s			0.458***	0.085**
			(0.110)	(0.042)
PS_{edp} 2010s				0.049**
				(0.025)
Full controls controls	✓	✓	✓	✓
Country-FE	✓	✓	✓	✓
Homeland-FE	–	–	✓	✓
District-FE	–	–	✓	✓
Obs	260	260	60478	60478
Within-R^2	0.734	0.490	0.138	0.0519

Notes: Columns 1 of the table reports the regression results of the *EPS* based on current borders on the *EPS* based on 1920 districts. Columns 2 reports the regressions results of log light density (VIIRS) on both the current and the colonial *EPS*. Column 3 & 4 report the corresponding grid-cell specifications using the district share PS_{edp}. Columns 1 & 2 include the full set of controls reported in Table II. Columns 3 & 4 include the following grid-cell level controls: log pop density (GHSL), log area, any river indicator, any lake indicator, mean elevation, mean ruggedness, mean crop suitability, mean distance to the coast, malaria burden, tse-tse suitability and an explorer indicator. Standard errors clustered at the culture group in parenthesis (columns 1 & 2) and are clustered two-way at homeland and district in columns 3 & 4. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

.

TABLE A-16
Persistence: EPS & PS averages

	Dependent variable:					
	EPS_{ep} avg.	ln LIGHT-DENSITY$_{ep}$		PS_{edp} avg.	ln LIGHT-*density*$_{cedp}$	
	(1)	(2)	(3)	(4)	(5)	(6)
EPS_{ep} 1920s	0.643***		1.120*			
	(0.064)		(0.573)			
EPS_{ep} (1960-2015)		1.901***	0.971			
		(0.638)	(0.867)			
PS_{edp} 1920s				0.464***		0.043
				(0.117)		(0.043)
PS_{edp} (1960-2015)					0.152***	0.137***
					(0.044)	(0.041)
Full controls controls	✓	✓	✓	✓	✓	✓
Country-FE	✓	✓	✓	✓	✓	✓
Homeland-FE	–	–	–	✓	✓	✓
District-FE	–	–	–	✓	✓	✓
Obs	260	260	260	60436	60436	60436
Within-R^2	0.804	0.458	0.466	0.164	0.0521	0.0521

Notes: Columns 1 of the table reports the regression results of the average EPS based on all border between 1960 and 2015 on the EPS based on 1920 districts. Columns 2 & 3reports the regressions results of log light density (VIIRS) on both the average and the colonial EPS. Column 4 to 6 report the corresponding grid-cell specifications using the district share PS_{edp}. Columns 1 to 3 include the full set of controls reported in Table II. Columns 4 to 6 include the following grid-cell level controls: log pop density (GHSL), log area, any river indicator, any lake indicator, mean elevation, mean ruggedness, mean crop suitability, mean distance to the coast, malaria burden, tse-tse suitability and an explorer indicator. Standard errors clustered at the culture group in parenthesis (columns 1 to 3) and are clustered two-way at homeland and district in columns 4 to 6. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

.

TABLE A-17

Within homeland & district evidence: DHS

	DV - WEALTH-INDEX$_{iedp}$			
	(1)	(2)	(3)	(4)
PS_{edp}	0.473***	0.552***	0.530***	0.717***
	(0.047)	(0.060)	(0.047)	(0.074)
Respondents	209103	211564	211564	211564
Within-R^2	0.0943	0.0529	0.0499	0.0445
Homeland controls	✓	✓	–	–
District controls	✓	–	✓	–
DHS cluster controls	✓	✓	✓	✓
Individual controls	✓	✓	✓	✓
Protectorate FE	✓	✓	✓	✓
District FE	–	✓	–	✓
Homeland FE	–	–	✓	✓

Notes: The table reports the results of regressing the DHS wealth index (Bruederle and Hodler, 2018) on the district share PS_{edp} of 5am grid-cells. Included controls are geographic characteristics of the DHS cluster: Log population density, log area, dummies for the presence of rivers and lakes, elevation, crop suitability, distance to coast, malaria burden and tsetse fly suitability. Individual level controls: Age, age squared, an urban indicator, a gender dummy, a non-indigenous indicator. Homeland controls are those reported in Panel A of Table II. District controls mirror the geographic controls of Panel A of Table II calculated for each district. Two-way clustered standard errors clustered at the protectorate specific homeland and district in parenthesis. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A-18

Communication infrastructure: Homeland level

	Dependent variables:				
	$Capital_{1922}$	$Post_{1922}$	$Telegraph_{1922}$	$Post_{1935}$	$Telegraph_{1935}$
	(1)	(2)	(3)	(4)	(5)
Panel A: LPM - DV as dummies					
EPS_{ep}	1.294***	0.482***	0.237	-0.055	-0.041
	(0.165)	(0.172)	(0.156)	(0.199)	(0.141)
$Capital_{1922}$ (any)		0.472***	0.567***	0.322***	0.320***
		(0.107)	(0.106)	(0.102)	(0.113)
$Post_{1922}$				0.303**	
				(0.118)	
$Telegraph_{1922}$					0.351**
					(0.130)
Baseline controls	✓	✓	✓	✓	✓
Geographic controls	✓	✓	✓	✓	✓
Homeland controls	✓	✓	✓	✓	✓
Protectorate FE	✓	✓	✓	✓	✓
Within-R^2	0.438	0.531	0.525	0.558	0.597
Homeland-P	182	182	182	182	182

Notes: Panel A replicates Table VII using an LPM on the homeland level. The capital and infrastructure variables are dummies, coded unity if any, district capital, post ,telegraph or phone station is present. Panel B replicates Table VII using poisson regressions on the count of the capitals and infrastructure variables. Full controls refers to the set of controls employed Table II. Standard errors, clustered at the culture group level in parenthesis. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

TABLE A-19
Communication infrastructure in French West Africa

	Dependent variables:				
	$Capital_{1922}$	$Post_{1922}$	$Telegraph_{1922}$	$Post_{1935}$	$Telegraph_{1935}$
	(1)	(2)	(3)	(4)	(5)
PS_{edp}	0.875***	0.007**	0.005*	0.006***	0.007***
	(0.154)	(0.003)	(0.003)	(0.002)	(0.002)
$Capital_{1922}$		0.343***	0.332***	0.433***	0.429***
		(0.051)	(0.052)	(0.054)	(0.054)
$Post_{1922}$				0.125***	
				(0.029)	
$Telegraph_{1922}$					0.146***
					(0.028)
Protectorate FE	✓	✓	✓	✓	✓
District FE	✓	✓	✓	✓	✓
Homeland FE	✓	✓	✓	✓	✓
Obs	53301	53301	53301	53301	53301
Within-R^2	0.306	0.0591	0.0572	0.101	0.111

Notes: Column 1 reports the result of regressing a district capital is located on homeland e dummy, which equals 1 for all 5am gridcells belonging to homeland e on the district share (PS_{edp}). Columns 2 and 3 show regress grid cell dummies for the presence of post and telegraph stations in 1922 on the district share and the capital dummy. Columns 4 to 5 regress communication infrastructure dummies in 1935 on the district share and district capital on different types of communications infrastructure in 1922. All columns include the following grid-cell level controls: log pop density (GHSL), log area, any river indicator, any lake indicator, mean elevation, mean ruggedness, mean crop suitability, mean distance to the coast, malaria burden, tse-tse suitability and an explorer indicator. The sample corresponds to all colonies located in French West Africa. Two-way-clustered standard errors, clustered at the district and homeland in parenthesis. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

TABLE A-20

Infrastructure and economic development

	DV - $\ln Lights_{ep}$				DV - $\ln Lights_{iedp}$			
	Homeland evidence				*Grid cell evidence*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
EPS_{ep}	0.954**	0.861**	0.821*	0.859*				
	(0.445)	(0.416)	(0.468)	(0.431)				
No. of capitals	0.092	-0.110	0.002	-0.109				
	(0.110)	(0.141)	(0.120)	(0.138)				
PS_{edp}					0.144***	0.137***	0.134***	0.129***
					(0.049)	(0.047)	(0.047)	(0.045)
$Capital$					3.962***	3.576***	3.310***	3.039***
					(0.204)	(0.196)	(0.187)	(0.185)
$Post_{1922}$		-0.047		-0.033		0.458		0.460
		(0.069)		(0.089)		(0.285)		(0.284)
$Telegraph_{1922}$		0.187***		0.198***		0.689***		0.505*
		(0.054)		(0.054)		(0.248)		(0.260)
$Post_{1935}$			0.100	0.157			0.491	0.485
			(0.337)	(0.308)			(0.501)	(0.483)
$Telegraph_{1935}$			-0.039	-0.185			0.876*	0.769
			(0.345)	(0.311)			(0.515)	(0.496)
Obs.	182	182	182	182	53301	53301	53301	53301
Within-R^2	0.496	0.520	0.513	0.521	0.115	0.124	0.133	0.139
Controls	✓	✓	✓	✓	✓	✓	✓	✓
Protectorate FE	✓	✓	✓	✓	✓	✓	✓	✓
District FE	–	–	–	–	✓	✓	✓	✓
Homeland FE	–	–	–	–	✓	✓	✓	✓

Notes: Columns 1 to 4 of the table report the regression results of log light density (VIIRS) on the EPS , the number of capitals located on a homeland within a colony/protectorate, as well as the counts of our infrastructure investment proxies. Columns 1 to 4 include the baseline controls log homeland population 1900 (HDYE), log area, a river and lake dummy, as well as the full set of controls presented in Table II. Columns 5 to 8 , repeat the analysis on the grid cell level using a the PS_{edp} , a dummy for grid cells located within 5km of a colonial capital, and dummies for the presence of our infrastructure proxies. Columns 5 to 8 include the usual grid cell controls: log pop density (GHSL), log area, any river indicator, any lake indicator, mean elevation, mean ruggedness, mean crop suitability, mean distance to the coast, malaria burden, tse-tse suitability and an explorer indicator. Standard errors clustered at the culture group (columns 1 to 4) and two-way on the on the homeland and district level (columns 5 to 8) in parenthesis. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A-21
Migration inside (outside) of homelands

	Dependent variables:			
	PS_{edp} District of residence			
	1920		Gadm level 2	
	(1)	(2)	(3)	(4)
	Born prior 1960	All	Born prior 1960	All
PS_{edp} (birth district)	0.073**	0.029**	0.042**	0.013
	(0.029)	(0.013)	(0.019)	(0.013)
Country-Group FE	✓	✓	✓	✓
Individual controls	✓	✓	✓	✓
Individuals	28367	233768	28367	233768
Within-R^2	0.0132	0.00341	0.00366	0.000711

Notes: Table present results form regressing the district population share of the current district on residence on the district population share of the birth district using the IPUMS data. Included controls are a sex indicator, as well as birth year dummies. Standard errors are clustered the ethnic group level in parenthesis. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A-22
External validity: Early agglomerations

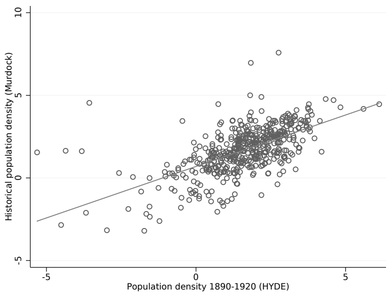
	Dependent variables:					
	Capital		ln Light-density$_{ep}$		ln Light-density$_{iedp}$	
	(1)	(2)	(3)	(4)	(5)	(6)
PS_{edp}	1.007***	0.989***	0.213***	0.153***		
	(0.091)	(0.064)	(0.058)	(0.029)		
Capital			3.771***	3.628***		
			(0.176)	(0.117)		
EPS_{ep}					1.774***	1.668***
					(0.456)	(0.268)
Capital (No.)					0.036	-0.013
					(0.022)	(0.017)
Colonizer	GBR	All	GBR	All	GBR	All
Controls	✓	✓	✓	✓	✓	✓
Protectorate FE	✓	✓	✓	✓	✓	✓
District FE	✓	✓	✓	✓	–	–
Homeland FE	✓	✓	✓	✓	–	–
Obs	83244	261442	83244	261442	410	1121
Within-R^2	0.373	0.336	0.127	0.106	0.455	0.434

Notes: Columns 1 and 2 report the result of regressing a district capital is on homeland dummy on the district share PS_{edp} on 5am gridcells, for former British colonies as well as all colonies within Sub-Saharan Africa. Columns 3 and 4 regress the log light density within 5am gridcells on the capital indicator from columns 1 and 2 and the district population share (PS_{edp}). Columns 5 and 6 run the corresponding homeland specifications using the expected population district share (EPS_{ep}) and the count of district capitals located on the homeland as independent variables. Columns 1 to 4 include the following grid-cell level controls: log pop density (GHSL), log area, any river indicator, any lake indicator, mean elevation, mean ruggedness, mean crop suitability, mean distance to the coast, malaria burden, tse-tse suitability and an explorer indicator. Columns 5 and 6 include the baseline controls log homeland population 1900 (HDYE), log area, a river and lake dummy. Full controls refers to the set of controls employed Table II. Two-way-clustered standard errors, clustered at the district and homeland in (columns 1 to 4) and clustered at the culture group in parenthesis. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

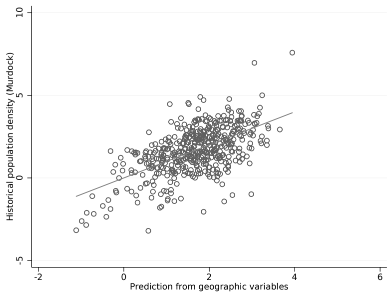
B. Appendix: Validating HYDE with historical data

Figure B-1
Historical data and HYDE

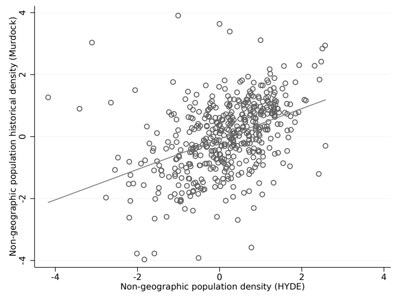
(a) Murdock: HYDE



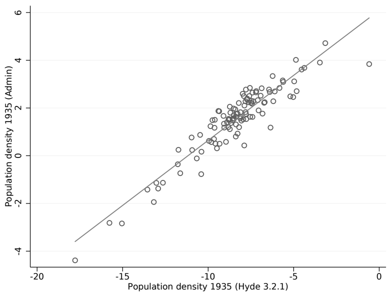
(b) Murdock: Geography



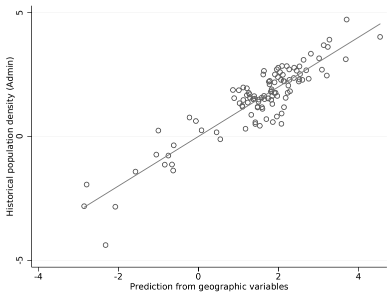
(c) Murdock: Added HYDE



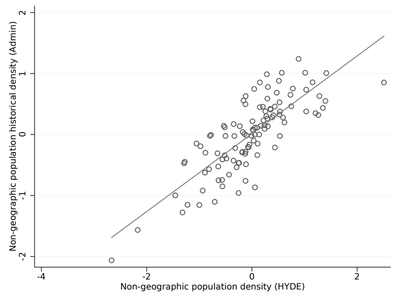
(d) W. Africa: HYDE



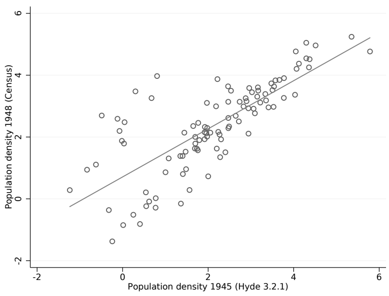
(e) W. Africa: Geography



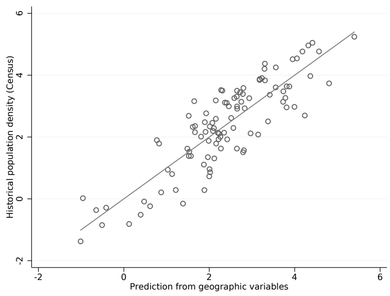
(f) W. Africa: Added HYDE



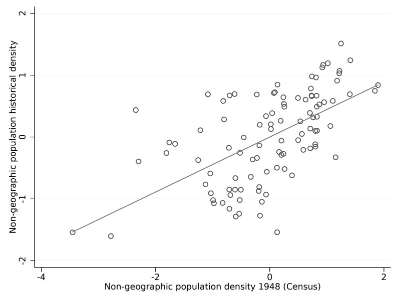
(g) E. Africa: HYDE



(h) E. Africa: Geography



(i) E. Africa: Added HYDE



Notes: The figures illustrate how the HYDE raster data and geographic variables help to predict historical population densities. The first column illustrates the tight fit obtained when running the historical data on HYDE densities. The second column uses a large set of geographic variables to predict the observed densities. The third column shows an added variable plot, i.e. it partials out the geographic variation on both sides and shows that the HYDE data is strongly correlated with the residual variation.

TABLE B-1
Validating HYDE at the homeland and district level

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent variable: Log of population density					
	Panel a) Murdock Homelands, pre-1900, rough estimates					
Pop. density 1905	0.623***		0.672***	0.506***		0.646***
	(0.063)		(0.106)	(0.064)		(0.102)
Pop. density 2000		0.495***	0.046		0.307***	-0.054
		(0.067)	(0.105)		(0.074)	(0.094)
Geo. controls	No	No	No	Yes	Yes	Yes
Adjusted R^2	0.375	0.329	0.447	0.449	0.413	0.493
Observations	478	473	473	467	462	462
	Panel b) French West Africa, 1935, administrative data					
Pop. density 1935	0.545***		0.345***	0.639***		0.459***
	(0.031)		(0.054)	(0.053)		(0.087)
Pop. density 2000		0.828***	0.325***		0.638***	0.235**
		(0.058)	(0.088)		(0.073)	(0.101)
Geo. controls	No	No	No	Yes	Yes	Yes
Adjusted R^2	0.866	0.843	0.879	0.928	0.912	0.932
Observations	111	111	111	110	110	110
	Panel c) British East Africa, 1948, census					
Pop. density 1945	0.778***		-0.048	0.445***		0.057
	(0.078)		(0.068)	(0.066)		(0.053)
Pop. density 2000		1.031***	1.076***		0.930***	0.871***
		(0.055)	(0.088)		(0.086)	(0.102)
Geo. controls	No	No	No	Yes	Yes	Yes
Adjusted R^2	0.587	0.855	0.854	0.811	0.900	0.899
Observations	101	101	101	101	101	101

Notes: The table shows the results from regressions of the log of actual population density—at different points in time and for different levels of aggregation—on population densities obtained using raster data sets and geographic controls. Geographic controls are log of homeland size, crop suitability, distance to coast, ruggedness, elevation, malaria burden, temperature, precipitation, TseTse suitability, and the fraction of inhabitable land. Standard errors are robust to heteroskedasticity. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

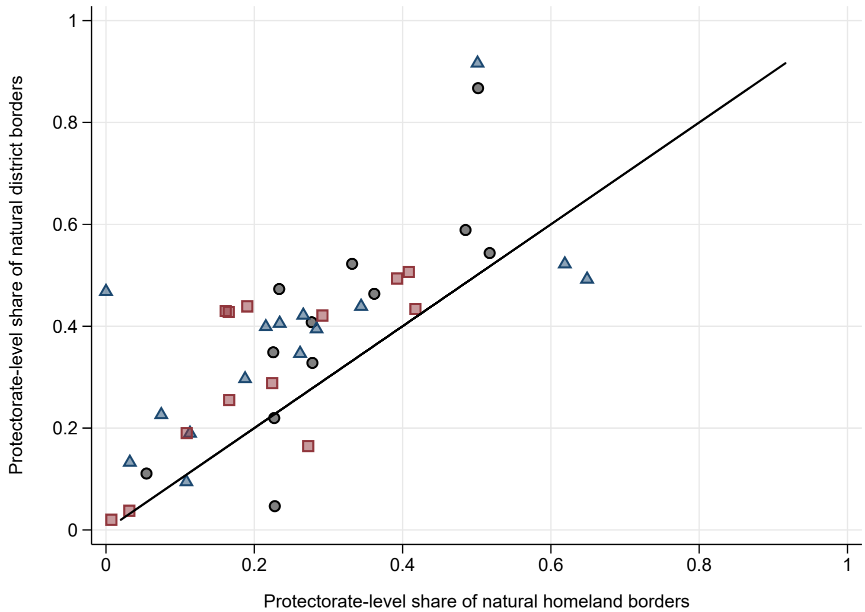
C. Appendix: Natural borders and border reforms

C-1. Natural borders

In this appendix we systematically test if the notion that colonizers draw borders mostly following natural borders if possible, as well as the claim that traditional homelands are less well separated by physical geography.

We start by setting up a 10km × 10km grid, and define each grid cell as being a natural choice for a border if it either has a elevation distance of 117m (Riley et al., 1999), host a natural water basin, a river. In a next step, we intersect our homeland and district borders with the grid, and define them either as natural if they fall within a "natural" grid cell or not natural otherwise. Finally, we weight the natural and non-natural border segments by the length of border contained within a grid-cell. This last steps accounts for the fact that some borders only cut through a small part of a grid cell, while others even turn within them. The "naturalness" of district and homeland borders is then simply the weighted sum of those classified border segments.

FIGURE C-1
Naturalness of colonial district and ethnic homeland borders



Notes: The figure plots the protectorate share of natural homeland borders (the fraction of homeland border km (Murdock, 1959) that fall within gridcells classified as natural border candidates) over the protectorate share of natural district borders (the fraction of district border km that fall within gridcells classified as natural border candidates). French protectorates are represented by red squares, British protectorates by blue triangles, and the protectorates of other colonizers by grey circles. The black line is the 45 degree line.

We plot the protectorate averages of natural homeland and district border segments in Figure C-1. The figure offers three insights. First, it provides evidence in support of

Table C-1
Naturalness of colonial district and ethnic homeland borders

	Primary colonizer		
	France	Britain	Any
	(1)	(2)	(3)
Naturalness district borders	0.316	0.383	0.369
	(0.047)	(0.051)	(0.031)
Naturalness homeland borders	0.218	0.259	0.261
	(0.037)	(0.051)	(0.026)
Difference	0.098***	0.124**	0.108***
Correlation	0.758	0.669	0.729
Observations	13	15	40

Notes: Table presents results the share of natural district borders, homeland borders, and the difference between the two across former French colonies (column 1), former British colonies (column 2), as well as all former colonies within Sub-Saharan Africa (column 3). Standard errors in parenthesis. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

the notion that colonial district borders are more natural than ethnic homeland borders in most protectorates. Second, it suggests that this pattern holds for both French and British protectorates as well as protectorates of other colonizers. Third, it reveals a high correlation (0.73) between the naturalness of ethnic homeland borders and the naturalness of colonial district borders, partly reflecting that some protectorates have more numerous and more prominent geographical markers than others. Table C-1 reports the averages by colonizers as well as the difference between the naturalness of district and homeland borders.

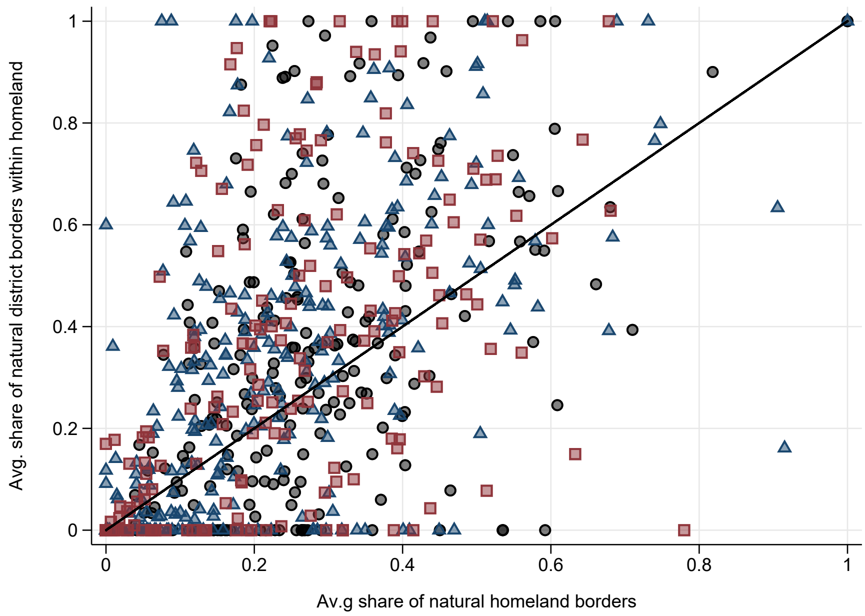
In a second step we zoom into the specific homelands. If the argument is correct that colonizers followed if possible natural markers and had little information on the exact distribution of ethnic groups in space, than actually implemented district borders within an existing homeland should be more natural compared to the actual homeland borders that they intersect. Hence, we calculate the average naturalness of the homeland borders for each homeland, and the average naturalness of the district borders located within each homeland.

Figure C-2 plots the resulting data again distinguishing between colonizers. We do indeed observe that most colonial district borders that are implemented tend to be more natural than the encompassing homeland borders, indicating by the mass of points above the 45 degree line. In fact only a clear minority seems to have less natural district borders compared to the encompassing homeland borders. As before we do not observe striking differences between colonizers.

Finally, we regress the share of natural district borders within ethnic homelands on the the share of homeland borders classified as natural for each homeland (see Table C-2). We control for the log of homeland population and the log of the homeland area, in

Figure C-2

Naturalness of colonial district and ethnic homeland borders



Notes: The figure plots the average share of district borders crosscutting ancestral homelands that are classified as natural borders over the average share of natural borders of the homelands the crosscut. French protectorates are represented by red squares, British protectorates by blue triangles, and the protectorates of other colonizers by black circles. The black line is the 45 degree line.

addition to the EPS. Note that ex ante the EPS, while an outcome of the subnational territorial structure is itself not determined by the share of borders which are classified as natural. However, it could very well be the case that homelands that receive a high EPS, have less natural borders cross-cutting them. This would be evidence of preferential treatment for some groups. We cannot include our full set of geographic variables, since some of them are used to classify borders as natural.

Table C-2 shows that there is a strong correlation between the share of homeland borders classified as natural and the share of district borders intersecting the homeland that are classified as natural as well, while none of the control variables seems to consistently matter for outcome. The effect is sizeable as the average naturalness of within homeland borders increases between 0.8 and 1.1 percentage point for each percentage point increase in the naturalness of the average homeland border. Taken together, we take this as evidence that the anecdotal evidence suggesting that colonizers followed natural markers when possibles is plausible and that we can assume sufficient local randomness in the border design.

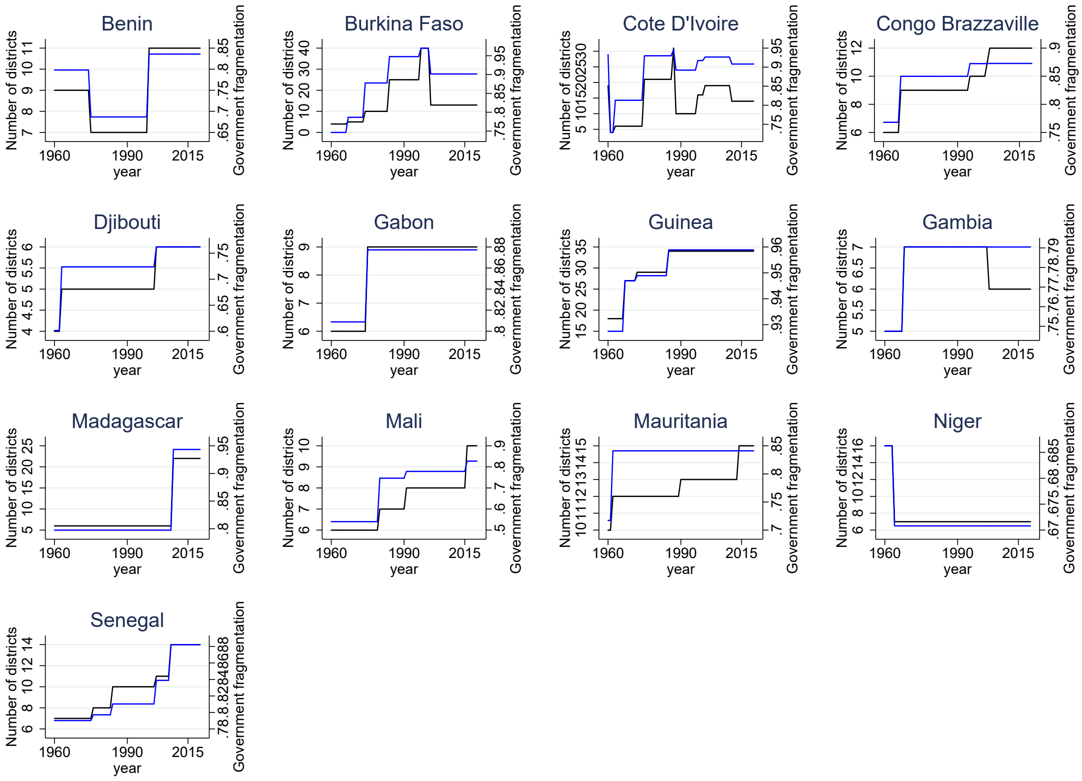
Table C-2
Naturalness within homeland district borders

	Dependent variable: NATURALNESS-DB_{ep}		
	Primary colonizer		
	France	Britain	All
	(1)	(2)	(3)
Naturalness of homeland borders	1.069***	0.799***	0.852***
	(0.191)	(0.112)	(0.080)
EPS	-0.070	0.076	-0.028
	(0.118)	(0.112)	(0.066)
Log population (HYDE)	0.028*	-0.001	0.011
	(0.017)	(0.020)	(0.012)
Log land area	-0.008	0.028	0.021*
	(0.018)	(0.017)	(0.012)
Protectorate FE	✓	✓	✓
Homeland-P	195	278	694
$Within - R^2$	0.231	0.219	0.189

Notes: Table reports the results of regressing the share of natural district borders located within a homeland on the fraction of borders classified as natural of the respective homelands, and the expected district population share EPS_{ep}. Standard errors clustered at the culture group in parenthesis. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

C-2. District reforms

Figure C-3
Territorial reforms in sample (1960-2015)



Notes: The figure plots the number of districts (black line) and government fragmenting (blue line) for the countries in former French colonial Africa from independence until 2015. Government fractionalization is the area fractionalization across subnational units within a country.

TABLE C-3

Local majorities & cash-crops

	Dependent Variable: $\ln Lights_{ep}$							
	EPS (ini)	$EPS_w(ini)$	EPS avg.	$EPS_w(curr)$		$\ln Lights_{ep}$		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
EPS 1920s	0.622***	0.213***	0.643***	0.315***				
	(0.077)	(0.056)	(0.064)	(0.060)				
EPS avg. (1960-2015)					1.901***			
					(0.638)			
EPS (ini)						1.078**		
						(0.429)		
EPS (2015)							1.746**	
							(0.663)	
EPS (ini) weighted						2.213***		3.606***
						(0.707)		(0.977)
EPS (interim) weighted								1.522**
								(0.729)
EPS (2015) weighted							0.200	0.867
							(1.085)	(0.668)
Protectorate FE	✓	✓	✓	✓	✓	✓	✓	✓
Baseline controls	✓	✓	✓	✓	✓	✓	✓	✓
Geographic controls	✓	✓	✓	✓	✓	✓	✓	✓
Homeland controls	✓	✓	✓	✓	✓	✓	✓	✓
Homeland-C	260	260	260	260	260	260	260	260
Within-R^2	0.764	0.155	0.804	0.492	0.458	0.472	0.463	0.476

Notes: Columns 1 to 4 of the table report EPS persistent estimates from later EPS on the EPS in the 1920s. Column 1 uses the EPS at independence (mostly 1960) as the dependent variable. Column 2 weights the initial EPS by the share of years between 1960 and 2015 the initial borders have been in place. Column (3) reports the persistent estimates on the avg. EPS, i.e., the potential different post independence EPS weighted by the time they have been in place. Column 4 uses the current EPS weighted by the time share they have been in place between 1960 and 2015 as the dependent variable. Columns (5) to (8) report estimates of the different post independence EPS measures on current day luminosity. Colöumn 5 uses the average EPS as a treatment, column 6 include the weighted and non-weighted initial EPS, while column 7 uses the current and weighted current EPS as treatments. Column (8) uses both the current and initial weighted EPS, as well as the interim EPS (the avg. EPS - the inital and current EPS, each weighted). All columns include country fixed effects and the full set of controls used in (column 4 of Table III). Standard errors clustered at the culture group in parenthesis. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

D. Appendix: Data sources

D-1. District border sources by protectorate

Table D-1
Colonies, and subnational borders

Colony	Border age	Source
Angola	1908	Atlas Colonial Portuguese 1908-1911
Dahomey	1922	Carte Administrative De L'Afrique Occidential Francaise 1922
Bechuanaland Protectorate	1926	Reichsamt für Landesaufnahmen 1940 reprint of Africa in 1926
Ruanda-Urundi	1938	Atlas du Congo Belge
Haute Volta	1922	Carte Administrative De L'Afrique Occidential Francaise 1922
Cameroun	1927	Carte du Cameroun 1927
British Kaffraria	1906	Millers Map of South Africa 1904
Oubangui-Chari	1919	Reichsamt für Landesaufnahmen 1940 reprint of Africa of 1933
Tchad	1925	Afrique Equatorial Francaise Colonie du Chad 1925
Congo français (Moyen-Congo)	1920	French Congo 1920
Congo, Free State	1927	Congo offical map 1927
Colonia Eritrea	1909	Colonia Eritrea Colle
		divisioni Administrative 1909
Gabon	1929	Afrique Equatoriale Francaise Colonie du Gabon 1929
Gambia Colony & Protectorate	1935	AMS fist edition, Corpt of Engineers, 1955
British Togoland	1913	Deutsch Togoland 1913
Gold Coast	1906	General Map of the Gold Coast 1906, Survey Department Gold Coast
Guinee	1922	Carte Administrative De L'Afrique Occidential Francaise 1922
Portuguese Guinea	< 1940	AMS 1955 1:500000
Cote d'Ivoire	1922	Carte Administrative De L'Afrique Occidential Francaise 1922
East Africa Protectorate / Kenya Colony	1920	Harmsworth 1920
Basutoland	1920	Harmsworth World Atlas 1920
Colonie de Madagascar	1931	Atlas Colonial Français. Colonies, protectorats, et pays sous mandat; cartes et texte du Commandant P. Pollachi. 2nd edition 1931
Nyasaland	1920	Harmsworth World Atlas 1920
Soudan français	1922	Carte Administrative De

Continued on next page

Table D-1 – *Continued from previous page*

Colony	**Border age**	**Source**
		L'Afrique Occidential Francaise 1922
Mauritanie	1922	Carte Administrative De L'Afrique Occidential Francaise 1922
Portuguese Mozambique	1903	Carta de Mocambique 1903
South West Africa	1912	Sprigade and Mosel 1912 Atlas
Niger	1922	Carte Administrative De L'Afrique Occidential Francaise 1922
Colony and Protectorate of Nigeria (N+S)	1929	Nigeria official map 1929, Survey Department Lagos
British Cameroons	1929	Nigeria official map 1929, Survey Department Lagos
Ruanda-Urundi	1938	Atlas du Congo Belge
Sénégal	1922	Carte Administrative De L'Afrique Occidential Francaise 1922
Sierra Leone Colony and Protectorate	1898	Sierra Leone, War Office 1898 Army Maps
British Somaliland	1926	War Office Reprint General Staff 1926
Protectorate		
Italian Somaliland	1925	Stielfers Atlas 1925
Cape of Good Hope (Cape Colony)	1906	Millers Map of South Africa 1904
Orange River Colony	1906	Millers Map of South Africa 1904
Union of South Africa	1906	Millers Map of South Africa 1904
Anglo-Egyptian Sudan	1915	Geographical Section General l Staff No. 2692
Swaziland Protectorate	1922	Map of Swaziland 1922
Tanganyika Territory	1920	Harmsworth World Atlas 1920
Zanzibar Protectorate	1920	Harmsworth World Atlas 1920
Togo	1913	Deutsch Togoland 1913
Uganda Protectorate	1920	Harmsworth World Atlas 1920
Northern Rhodesia (Chartered)	1926	Reichsamt für Landesaufnahmen 1940 reprint of Africa of 1926
Southern Rhodesia	1926	Reichsamt für Landesaufnahmen 1940 reprint of Africa of 1926

Notes: The border age is the earliest date for which we an verify the existence of the borders form the corresponding maps in the source column.

Table D-2
Colonies & colonizers

Colony	**Country**	*2nd* **colonizer**	*1st* **colonizer**
Angola	Angola	PRT	–
Dahomey	Benin	FRA	–
Bechuanaland Protectorate	Botswana	GBR	–
Ruanda-Urundi	Burdundi	BEL	DEU
Haute Volta	Burkina Faso	FRA	–
Cameroun	Cameroon	FRA	DEU
British Kaffraria	Cape of Good Hope	GBR	–
Oubangui-Chari	Central African Republic	FRA	–
Tchad	Chad	FRA	–
Congo français (Moyen-Congo)	Congo, Republic	FRA	–
Congo, Free State	Congo, Republic Dem.	BEL	–
Colonia Eritrea	Eritrea	ITA	–
Gabon	Gabon	FRA	–
Gambia Colony and Protectorate	Gambia	GBR	–
British Togoland	Ghana	GBR	DEU
Gold Coast	Ghana	GBR	–
Guinee	Guinea	FRA	–
Portuguese Guinea	Guinea Bissau	PRT	–
Cote d'Ivoire	Ivory Coast	FRA	–
East Africa Protectorate / Kenya Colony	Kenya	GBR	–
Basutoland	Lesotho	GBR	–
Colonie de Madagascar	Madagascar + Islands	FRA	–
Nyasaland	Malawi	GBR	–
Soudan français	Mali	FRA	–
Mauritanie	Mauritania	FRA	–
Portuguese Mozambique	Mozambique	PRT	–
South West Africa	Namibia	ZAF/ GBR	DEU
Niger	Niger	FRA	–
Colony and Protectorate of Nigeria (N + S)	Nigeria	GBR	–
British Cameroons	Nigeria/ Cameroon	GBR	DEU
Ruanda-Urundi	Rwanda	BEL	DEU
Sénégal	Senegal	FRA	–
Sierra Leone Colony and Protectorate	Sierra Leone	GBR	–
British Somaliland Protectorate	Somalia	GBR	–
Italian Somaliland	Somalia	ITA	–
Cape of Good Hope (Cape Colony)	South Africa	GBR	–
Orange River Colony	South Africa	GBR	Settlers

Continued on next page

Table D-2 – *Continued from previous page*

Colony	**Country**	*2nd* **colonizer**	*1st* **colonizer**
Union of South Africa	South Africa/ Namibia	GBR/ Self-gov'ed	– –
Anglo-Egyptian Sudan	Sudan, Egypt, Libya, S. Sudan	GBR/ EGY	–
Swaziland Protectorate	Swaziland	GBR	–
Tanganyika Territory	Tanzania	GBR	DEU
Zanzibar Protectorate	Tanzania	GBR	–
Togo	Togo	FRA	DEU
Uganda Protectorate	Uganda	GBR	–
Northern Rhodesia (Chartered)	Zambia	GBR	–
Southern Rhodesia	Zimbabwe	GBR	–

Notes: If the 1*st* colonizer is left empty it coincides with the 2*nd* / final colonizer. Note, that we only consider European colonizers for first colonizers. Hence, we omit previous Ottoman, Greek and Roman colonization experiences especially in current day Sudan, as well as Arabic colonization in east Africa.

D-2. Controls variables and further outcomes

Rivers dummy indicating that a major river passes through a homeland are taken from Michalopoulos and Papaioannou (2016).

Lakes dummy indicating that a lake is crosscutting or located within a traditional homeland is taken from Michalopoulos and Papaioannou (2016).

Elevation averages are calculates based on the elevation information contained in the SRTM CSI CGIAR 250m raster (Jarvis et al., 2008).

Ruggedness averages are calculates based on elevation differences of 250 meters based on the SRTM CSI CGIAR 250m data (Jarvis et al., 2008).

Crop suitability averages (1890-1920) are taken from the "Historic Croplands Dataset" provided by Ramankutty et al. (2002).

Distance to coast averages are calculated by averaging the the geographic distance of each grid cells centroid located within a homeland to the nearest coastline. Coastlines shapes are based on "Natural Earth" (https://www.naturalearthdata.com),

Malaria burden suitability is measured as the historical malaria suitability measure developed by Depetris-Chauvin and Weil (2018). We recreate their measure using the same inputs.

TseTese suitability is estimated using the procedure employed by Alsan (2015). We standardize the measure to out sample.

Pre-colonial conflict indicator is constructed based on the replication data of Michalopoulos and Papaioannou (2016), who code the distance to a pre-colonial conflict event. We instead use the centroids of pre-colonial wars occurring between 1400–1700 coded originally by Besley and Reynal-Querol (2014).

Log slaves per area (exports) are calculates based on the data collected by Nunn (2008). We take the exports reported by Nunn and divide them by the total homeland area. Note that the slave trade exports are not reported by homeland-protectorate but by the entire homeland.

Early explorer routes indicator is a dummy if any explorer passed either through or within 50km of a homeland. Source is Nunn (2008).

Pre-colonial city in 1940 indicator is taken from the replication data of Michalopoulos and Papaioannou (2016). It indicates that a pre-colonial city is located within the homeland. The data is based on Chandler (1987).

Pre-colonial kingdom / Empire indicator is unity if a homeland falls within the boundaries of a large pre-colonial kingdom or empire. The indicator is taken from the replication dataset of Michalopoulos and Papaioannou (2016). The original source are Besley and Reynal-Querol (2014).

Cash crop suitability proxies for bananas, cocoa, coffee, cotton, groundnuts, palm oil, tea, and tobacco are based on the GAEZ suitability raster estimates (https://gaez.fao.org/). We take the average suitability values for each homeland within the protectorates of our sample. Note that we use the average suitability estimates based on rain-fed agriculture between 1960 and 2000 to keep the measure es exogenous as possible. Using single years reduces the sample of gridcells dramitically.

Permanent settlements indicator equals unity for ethnic homelands in which groups live in some from of permanent settlement (V30 $\geq$ 4) in the Murdock (1967) data. The indicator is zero for ethnic groups recorded as having nomadic, semi-nomadic, and semi-sedentary groups settlement patterns (see Michalopoulos and Papaioannou, 2016, for a similar approach).

Agricultural intensity indicator equals unity for ethnic groups depending on agriculture to at least 45%, V42 $>$ 5 in the Murdock (1967) data (see Michalopoulos and Papaioannou, 2016, for a similar approach). Michalopoulos and Papaioannou (2016) note that the source variable V42, is based on information on ".. penetration of the soul, planting, tending the growing crops and harvesting, but not subsequent food preparation"(Murdock, 1967).

Political centralization indicator is based on the political centralization index included in the Michalopoulos and Papaioannou (2016) data. It is originally taken from the Murdock (1967) data, where information on the jurisdictional hierarchy of communities is provided in a 4 level index. Our dummy takes unity if V33 ≥ 2, indicating that their is at least one political layer above the local one.

Local hierarchy indicator is unity if the V32 from the Murdock (1967) data > 2. The variable is taken from the replication dataset of Michalopoulos and Papaioannou (2016). The dummy indicates that there is any form of hierarchy at the local level.

Class stratification indicator is constructed following Michalopoulos and Papaioannou (2016). It equals zero if V66 $= 1$ in the Murdock (1967) data, indicating the "absence of significant class distinctions among freemen, ignoring variations in individual repute achieved through skill, valor, piety, or wisdom." The presence of class stratification indicators that there are forms of such distinctions (V66 > 1). The data is taken again from the replication dataset of Michalopoulos and Papaioannou (2016).

Hereditary slavery indicator equals one if slavery is coded as heredity and socially significant (v70=4 based on Murdock (1967)). Data is taken from the replication dataset of Michalopoulos and Papaioannou (2016).

Health professionals per 100000 in 1920s is obtained from the replication data of Huillery (2009). The replication data reports the average number of teachers from 1910-1928 divided by the district population in 1925 for district in French West Africa. The estimates are based on colonial budgets (see Huillery (2009) for details). We calculate the weighted average of the measure using the share of a group that lives within a district as the weights.

Teachers per 100000 in 1920s is obtained from the replication data of Huillery (2009). The replication data reports the average number of doctors, nurses and medical auxiliaries from 1910-1928 divided by the district population in 1925 for district districts in French West Africa. The estimates are based on colonial budgets (see Huillery (2009) for details). We calculate the weighted average of the measure using the share of a group that lives within a district as the weights.

Hospital density is the number of health facilities divided by the homeland area. The measure is created by matching geocoded health facilities to sample of Murdock homelands within protectorates. The geocoded healht facitliy data is provided by Maina et al. (2019) and can be publicly accessed via :https://data.humdata.org/dataset/health-facilities-in-sub-saharan-africa?force_layout=desktop.

D-3. DHS surveys

Wealth index is either the DHS wealth index (v190 *Source*: DHS). Or the extended sample version by Bruederle and Hodler (2018).

Infant mortality is calculated based on an indicator variable for each life birth of a respondent that is unity if the child passed away during the first years (b6 *Source*: DHS). The indicator variable is than divided by 1000 (see Bruederle and Hodler (2019) for a similar approach).

Professional birth attendants is an indicator taking unity for all children of a respondents at witch's birth a professional birth attended has been present. Professional birth attendants are either doctors, professional midwifes or professional nurse (m3a - m3n) in the DHS. *Source*: DHS.

Years of schooling is the count of finished school years as provided by the DHS (v107 *Source*: DHS).

Age in years of the respondent (v012 and mv012) in the DHS. *Source*: DHS.

Female indicator taking unity for all respondents in the IR dataset of the DHS and zero for all respondents in the MR dataset of the DHS. *Source*: DHS.

Urban indicator is unity if a DHS survey cluster in the geocoded dataset is defined as being located in an urban area and zero otherwise. *Source*: DHS.

Non-indigenous indicator is unity if the self-reported ethnicity within the DHS is different from the ancestral homeland on which the respondent is residing. *Source*: DHS.

Multiple birth indicator is unity if a respondents child was born either as a twin or multiple ($b0_{01}$–$b0_{20}$). *Source*: DHS.

Sex indicator for respondents children, takes unity if the respondent child is female ($b4_{01}$–$b4_{20}$). *Source*: DHS.

Birth order indicators are created from the the birth order variables ($bord_{01}$–$bord_{20}$), they indicate if a respondents child is the $1st$, $2nd$ or up to the $20th$ child of a respondent. *Source*: DHS.

Year of birth child indicators are a set of dummies constructed from the year of birth information provided in the DHS for each child of a respondent ($b2_{01}$–$b2_{20}$). *Source*: DHS.

TABLE D-3
DHS survey sample

ISO	Interview year	Respondents	Share female
AGO	2015	3,261.00	0.73
AGO	2016	4,204.00	0.73
BFA	1992	1,815.00	0.77
BFA	1993	4,759.00	0.77
BFA	1998	2,560.00	0.71
BFA	1999	5,411.00	0.71
BFA	2003	11,632.00	0.78
BFA	2010	16,326.00	0.70
BFA	2014	5,144.00	1.00
BEN	1996	2,322.00	0.78
BEN	2001	7,103.00	0.69
BEN	2011	1,938.00	0.76
BEN	2012	15,928.00	0.76
COD	2007	2,186.00	0.67
COD	2013	2,252.00	0.69
COD	2014	267.00	0.69
CAF	1994	2,678.00	0.77
CAF	1995	1,162.00	0.79
CIV	1998	142.00	1.00
CIV	1999	203.00	1.00
CIV	2011	1,119.00	0.69
CIV	2012	8,209.00	0.66
CMR	2004	12,179.00	0.67
CMR	2011	7,128.00	0.68
GAB	2012	9,995.00	0.62
GHA	1993	4,015.00	0.78
GHA	1994	174.00	0.72
GHA	1998	2,300.00	0.75
GHA	1999	1,998.00	0.75
GHA	2003	4,642.00	0.50
GHA	2008	4,132.00	0.50
GHA	2014	5,966.00	0.68
GHA	2016	2,160.00	1.00

Continued on next page

Table D-3 – *Continued from previous page*

ISO	Interview year	Respondents	Share female
GIN	1999	5,579.00	0.77
GIN	2005	2,728.00	0.72
GIN	2012	3,031.00	0.68
KEN	2003	7,906.00	0.69
KEN	2008	3,517.00	0.71
KEN	2009	3,672.00	0.70
KEN	2014	31,532.00	0.71
KEN	2015	3,596.00	1.00
MLI	1995	1,743.00	0.81
MLI	1996	3,299.00	0.79
MLI	2001	12,398.00	0.84
MLI	2006	9,942.00	0.77
MLI	2012	5,918.00	0.70
MLI	2013	4,080.00	0.70
MLI	2015	4,248.00	1.00
MWI	2000	12,754.00	0.81
MWI	2004	8,544.00	0.78
MWI	2005	2,791.00	0.78
MWI	2010	22,028.00	0.76
MWI	2012	2,301.00	1.00
MWI	2014	2,214.00	1.00
MWI	2015	15,567.00	0.77
MWI	2016	8,054.00	0.76
MOZ	2011	14,701.00	0.77
NGA	2008	31,126.00	0.68
NGA	2010	5,464.00	1.00
NGA	2013	47,693.00	0.69
NGA	2015	6,535.00	1.00
NER	1992	5,755.00	0.80
NER	1998	8,751.00	0.68
NAM	2000	7,067.00	0.69
SLE	2008	8,650.00	0.70
SLE	2013	20,305.00	0.70
SLE	2016	5,021.00	1.00
SEN	1992	2,069.00	0.81
SEN	1993	4,246.00	0.82
SEN	1997	11,833.00	0.66
SEN	2005	15,514.00	0.79
SEN	2008	9,847.00	1.00
SEN	2009	6,402.00	1.00
SEN	2010	7,492.00	0.78
SEN	2011	10,784.00	0.75
SEN	2012	3,003.00	1.00
SEN	2013	4,826.00	1.00
SEN	2015	11,372.00	0.70
TCD	2014	3,652.00	0.75
TCD	2015	8,691.00	0.76
TGO	1998	2,303.00	1.00

Continued on next page

Table D-3 – *Continued from previous page*

ISO	Interview year	Respondents	Share female
TGO	1998	8,437.00	0.69
TGO	2013	3,832.00	0.68
TGO	2014	5,600.00	0.66
UGA	2009	3,142.00	1.00
UGA	2010	24.00	1.00
UGA	2011	17,183.00	0.95
UGA	2014	3,002.00	1.00
UGA	2015	1,508.00	1.00
UGA	2016	19,874.00	0.78
ZMB	2007	12,540.00	0.52
ZMB	2013	16,734.00	0.53
ZMB	2014	10,801.00	0.53
ZWE	2010	6,553.00	0.55
ZWE	2011	5,388.00	0.53

Notes: The table depicts the DHS survey included in our sample. The survey years, the number of respondents in each survey that we can match to our data as well as the share of female respondents within each DHS survey.

D-4. IPUMS data

Ethnic matches between Murdock and IPUMS are obtained in two steps. First we use the LEDA R package (Müller-Crepon et al., 2020) which provides matches between various datasets classifying ethnic groups. Specifically, we conduct a match between the Afrobarometer classification, which mostly overlaps with the ethnicities reported in IPUMS and the Murdock atlas. In a second step we qualitatively check each single match and correct errors that results from the imperfect match between IPUMS and Afrobarometer. The resulting match is reassuring, we are able to match most meaningful groups. The maximum number of respondents belonging to an IPUMS group we are unable to match is 23 (median 3). We report the final match in Table D-4.

Table D-4
IPUMS census & survey sample

Country	IPUMS ethnicity	Murdock ethnicity	Respondents	Share female
Benin	Bariba	BARGU	88,268.00	0.51
Benin	Boo	BUSA	7,961.00	0.51
Benin	Dendi	DENDI	26,161.00	0.50
Benin	Yoruba	EGBA	15,478.00	0.52
Benin	Adja	FON	83,477.00	0.53
Benin	Fon	FON	172,345.00	0.52
Benin	Goun	GUN	52,944.00	0.51
Benin	Mina	POPO	8,619.00	0.52

Continued on next page

Table D-4 – *Continued from previous page*

Country	IPUMS ethnicity	Murdock ethnicity	Respondents	Share female
Benin	Ditamari	SOMBA	22,504.00	0.51
Burkina Faso	Bobo	BOBO	1.00	1.00
Burkina Faso	Bissa	BUSANSI	36,995.00	0.55
Burkina Faso	Dafing	DAFI	12,846.00	0.49
Burkina Faso	Kassena	GRUNSHI	8,032.00	0.52
Burkina Faso	Senoufo	KARABORO	1.00	0.00
Burkina Faso	Senoufo	MINIANKA	1.00	1.00
Burkina Faso	Moore	MOSSI	600,519.00	0.53
Burkina Faso	Senoufo	SENUFO	16,668.00	0.52
Burkina Faso	Bobo	SIA	17,697.00	0.51
Guinea	Baga	BAGA	3,288.00	0.52
Guinea	Djalonke	DIALONKE	8,539.00	0.51
Guinea	Poular	FOUTADJALON	323,147.00	0.53
Guinea	Kissi	KISSI	38,209.00	0.52
Guinea	Kono	KONO	8,009.00	0.53
Guinea	Koniaka	KONYANKE	41,634.00	0.51
Guinea	Kouranko	KORANKO	18,029.00	0.52
Guinea	Kpele	KPELLE	37,072.00	0.52
Guinea	Landouma	LANDUMA	4,297.00	0.53
Guinea	Maninka	MALINKE	233,553.00	0.51
Guinea	Nalou	NALU	548.00	0.50
Guinea	Soussou	SUSU	167,188.00	0.51
Guinea	Bassari	TENDA	930.00	0.54
Guinea	Toma	TOMA	13,358.00	0.52
Mali	Arabic	KUNTA	1.00	0.00
Mali	Samogo	SAMO	5,536.00	0.52
Mali	Arabic	ZENEGA	1,573.00	0.48
Senegal	Balante	BALANTE	8,042.00	0.50
Senegal	Bainouk	BANYUN	1,898.00	0.47
Senegal	Diola	DIOLA	44,508.00	0.50
Senegal	Pulaar	FOUTATORO	1.00	0.00
Senegal	Bambara	MALINKE	15,752.00	0.50
Senegal	Mandinka	MALINKE	33,534.00	0.50
Senegal	Serer	SERER	170,915.00	0.50
Senegal	Soninke	SONINKE	8,044.00	0.53
Senegal	Bassari	TENDA	1,094.00	0.48
Senegal	Pulaar	TUKULOR	294,304.00	0.49
Senegal	Wolof	WOLOF	426,891.00	0.51

Notes: The table depicts the IPUMS census & surveys included in our sample. The survey years, the number of respondents in each survey that we can match to our data as well as the share of female respondents within each matched Murdock group.

Additional references

Alsan, M. (2015). The effect of the tsetse fly on African development. *American Economic Review 105*(1), 382–410.

Besley, T. and M. Reynal-Querol (2014). The legacy of historical conflict: Evidence from Africa. *American Political Science Review*, 319–336.

Bruederle, A. and R. Hodler (2018). Nighttime lights as a proxy for human development at the local level. *PloS One 13*(9), e0202231.

Bruederle, A. and R. Hodler (2019). Effect of oil spills on infant mortality in Nigeria. *Proceedings of the National Academy of Sciences 116*(12), 5467–5471.

Chandler, T. (1987). *Four thousand years of urban growth: An historical census*. Mellen.

Depetris-Chauvin, E. and D. N. Weil (2018). Malaria and early African development: Evidence from the sickle cell trait. *Economic Journal 128*(610), 1207–1234.

Huillery, E. (2009). History matters: The long-term impact of colonial public investments in French West Africa. *American Economic Journal: Applied Economics 1*(2), 176–215.

Jarvis, A., H. I. Reuter, A. Nelson, E. Guevara, et al. (2008). Hole-filled srtm for the globe version 4. *srtm.csi.cgiar.org*.

Maina, J., P. O. Ouma, P. M. Macharia, V. A. Alegana, B. Mitto, I. S. Fall, A. M. Noor, R. W. Snow, and E. A. Okiro (2019). A spatial database of health facilities managed by the public health sector in sub Saharan Africa. *Scientific data 6*(1), 1–8.

Michalopoulos, S. and E. Papaioannou (2016). The long-run effects of the scramble for Africa. *American Economic Review 106*(7), 1802–48.

Müller-Crepon, C., Y. Pengl, and N.-C. Bormann (2020). Linking ethnic data from Africa. *Journal of Peace Research*, Forthcoming.

Murdock, G. P. (1959). *Africa: Its peoples and their culture history*. McGraw-Hill.

Murdock, G. P. (1967). Ethnographic atlas: A summary. *Ethnology 6*(2), 109–236.

Nunn, N. (2008). The long-term effects of Africa's slave trades. *Quarterly Journal of Economics 123*(1), 139–176.

Ramankutty, N., J. A. Foley, J. Norman, and K. McSweeney (2002). The global distribution of cultivable lands: Current patterns and sensitivity to possible climate change. *Global Ecology and Biogeography 11*(5), 377–392.

Riley, S. J., S. D. DeGloria, and R. Elliot (1999). Index that quantifies topographic heterogeneity. *Intermountain Journal of Sciences 5*(1-4), 23–27.