Strength in Numbers: Ethnic Group Size and Economic Development

Victoria Gerenutti Klarosk* Rogerio Santarrosa†

August 2, 2022

Abstract

Ethnic identity – and its associated history, norms, and institutions – is a fundamental determinant of political and economic outcomes in Africa. We use the partition of ethnic groups across different countries, caused by the Scramble for Africa, to study the effect of group size on economic development. The arbitrariness of the border design caused exogenous variation in the population size of split groups and thus works as a natural experiment. We combine information on the historical homeland of ethnic groups, gridded geo-referenced estimates of population before the colonization, and a measure of ethnic contemporary economic performance, as proxied by satellite images of light density at night. Using within group variation plus country fixed effects, we find that larger groups are currently more developed. We exploit several channels that can potentially mediate these effects. Larger groups are more likely to have natural endowments (as access to lake and diamond mines); they are closer to the capital city; they are linguistically more similar to the rest of the population; and they are more likely to be politically organized and be represented in the highest political positions in the country.

Keywords: Population size, economic development, ethnicity, Africa

*Insper, Department of Economics.

†Insper, Department of Economics.

1 Introduction

Ethnic diversity has been strongly associated with Africa's underdevelopment (Easterly and Levine (1997), Alesina and Ferrara (2005)). However, there are still significant challenges in interpreting this relationship as causal. Most of these empirical patterns have been established in cross-country analyses and must assume that ethnic fractionalization is not correlated to any other (non-observable) determinant of national economic performance. This concern is reinforced by recent evidence that ethnic classifications have been shaped by natural endowments, as land quality (Michalopoulos (2012)), and historical events, such as the slave trade (Nunn (2008)), making it partly result of the fundamental roots of long-run development.

In this paper, we investigate a different dimension, and in some degree an implied consequence, of this phenomenon. We test whether ethnic group size matters for its economic development. If ethnic diversity has a negative effect on aggregate outcomes, we would expect that small groups would perform relatively worse than their larger counterparts.

There are several reasons (which we cover in detail in the next section) why larger groups might have an economic advantage over small ones – many of them are relevant candidates as fundamental causes of development for which there is no conclusive evidence in the literature. For instance, in environments where trust and taste preferences follow ethnic lines, individuals from larger groups will be more likely to have favorable interactions and less likely to be discriminated. Moreover, as long cultural and linguistic barriers exist, bigger groups can benefit from economies of scales implying more intense division of labor and transmission of knowledge. Ethnic size can be also an important determinant of political capital and allocation of public goods.

To establish the causal effect of group size on economic development, we use an historical natural experiment: the Scramble for Africa. This event started with the Berlin Conference of 1884/85 and was completed by the turn of the 20th century, when Europeans partitioned the continent into colonies, protectorates, and spheres of influence. Many evidences indicate that

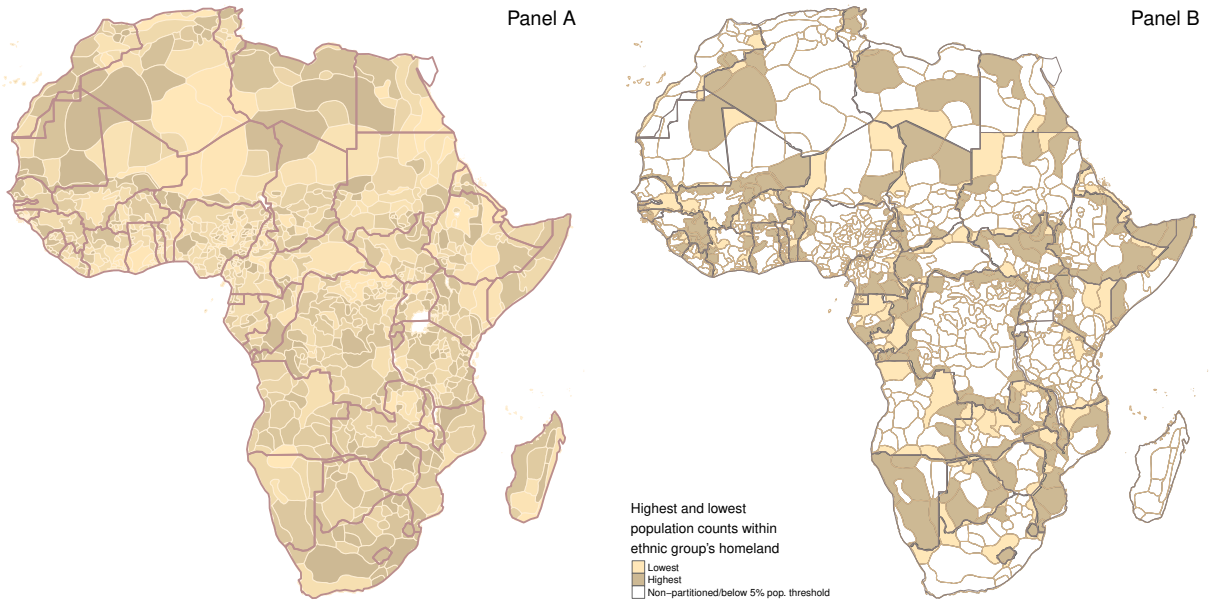
the borders were designed with little information or regard for the geographic distribution of native institutions already in place in the continent. In fact, 80% of African borders follow latitudinal and longitudinal lines (Alesina et al. 2011). This caused many ethnic homelands to be partitioned into two or more countries. In practice, this made a larger share of one split group (higher population) fall in one country while a lower share of the same tribe became part of another state. Intuitively, our identification strategy compares the economic outcomes of these two parts to establish the causal effect of group size.

Michalopoulos and Papaioannou (2016) relies on a similar identification strategy to test the effect of ethnic partition (comparing split versus non-split groups). They provide compelling evidence that country boundaries were set exogenously to the local political geography of the region, by testing an extensive set of pre-colonial characteristics across split and non-split ethnic groups. The results support the largely arbitrary nature of the borders' drawing, at least with respect to ethnic organization. Here instead, we use the within group variation also caused by the same natural experiment.

We use several data sources to implement our empirical design. First, we use the Ethnolinguistic Map constructed by Murdock (1967) that depicts the spatial distribution of the continent's ethnic groups at the time of the colonization, building on sources covering the period 1860-1940. We overlay Murdock's map with current country borders to obtain a map at the country-tribe level. Second, we use gridded historical population data (HYDE) to construct estimates of population size at each country-tribe in 1880. Importantly, these are estimates of the population just before the national borders were defined. Figure 1 shows the overlap of national borders and the pre-existing tribes (Panel A). Panel B specifically highlights the division of groups across countries in segments with higher versus lower population count. Third, since no granular, comparable measures of economic development are available for the whole African continent, we proxy economic activity by average light density at nighttime (Michalopoulos and Papaioannou 2013, Alesina et al. 2016).

The arbitrariness of the border design creates a natural experiment which allows us to assess how the exogenous size of split ethnic groups affects the groups' development. For this, we leverage within-ethnicity cross-country variation in population size, as of 1880 (pre-scramble), while holding country and ethnic characteristics constant – our main specification includes country fixed effects[1] and ethnic fixed effects. This gives the causal intention-to-treat effect of ethnic group size on development. In our preferred specification, we find that a 10% increase in the group size, light density increases by 1.7%. This result is robust to a battery of geography and agriculture controls.[2]

Figure 1: Traditional ethnic homelands, country borders, and sample of partitioned ethnic groups



Note: Panel A displays the overlay of the ethnic groups' homelands in Murdock (1959, 1967), as colored areas, and current country borders, as lines. Panel B emphasizes the tribes identified as split according to the details discussed in section 4.1, and uses two shades to distinguish parts with highest and lowest population sizes within split ethnic groups. Areas of tribes partitioned into more than one country have lighter shade if their population is lower than the median population count, and darker if it is greater.

Next we assemble data from a variety of sources to examine potential channels through which group size can affect economic development. First, we examine the role of natural

[1]We are able to include country fixed effects because several split ethnic groups might be in a single country

[2]Geography controls are distance to coast (log), indicator for being landlocked, terrain ruggedness index, average slope, average precipitation, and malaria suitability index. Agriculture controls are share of irritated area in 1880, share of grazing area in 1880, share of pasture area in 1880, share of cropland area in 1880, mean altitude, maximum altitude, crop suitability index, land suitability level, share of water sites, dominant soil, agro-zones, and hydro-basin class.

endowments. A larger group might be more likely to have access to natural resources which can facilitate agricultural production (water) or even mines/oil fields that can boost some economic activity. We find that larger groups are more likely to have lakes and diamond mines in their territory. We do not find a higher likelihood of having oil fields or other gem mines.

Second, we investigate group's capacity to attract investments in infrastructure to their homeland. We do find some evidence that larger groups have more access to railway. However, we do not find any effect on the presence or easier access to roads or colonial investments.

Third, we show that even though larger groups are not more likely to have the country's capital in their homeland, the distance is in fact shorter. This is relevant because economic mobility in Africa is highly associated with migration to urban centers (mostly comprised by the capital). Moreover, distance to capital is associated with access to markets which can be used either to buy important inputs (e.g fertilizers) or sell agricultural surplus. Another relevant economic factor is the language distance to other co-citizens in the country – which can affect communication, transmission of technology and human capital accumulation. The results is consistent with larger groups having lower linguistic distance to their co-citizens.

Finally, we combine Murdock's tribes to the Ethnic Power Relations Dataset (EPR). This last dataset consists of a selected list of politically relevant groups. They are groups for which there is at least one organization claiming to representing them at national politics. EPR also has information on the degree of political power that these groups possess. We find that larger groups are more likely to become politically relevant, and as a result, occupy dominant positions in the national government.

This paper is structured as follows. In the next section, we review part of the literature on the links between ethnic group size and economic development, focusing on papers that have investigated the African context. In section 3, we discuss the historical background of the Scramble for Africa. In section 4, we describe the data sources and the way we have used them to carry out our analysis. Section 5 shows the empirical strategy and the main results. The

potential channels are investigated in Section 6. We end with concluding remarks in Section 7.

2 Related literature

Setting aside country and institutional specifics, why would a group benefit from being a majority? Or, equivalently, what kind of obstacles does a minority group face that could harm its economic success?

An initial explanation is borrowed from the Social Psychology literature and relates to the social identity theory. Previous work has argued that patterns of intergroup behavior are consistent with individuals attributing positive utility to the well-being of members of their own group and negative utility to that of other groups (Tajfel et al. 1971). This could imply that minorities engage less in community social activities due to their status and become segregated (Alesina and La Ferrara 2000). But beyond the explicit *taste* for interacting primarily with members of one's group, group affiliation may be important for trade in the presence of market imperfections. Greif (1993) points that, in Medieval times, alliances along ethnic lines were common across traders in order to exchange information about opportunistic behavior of others. Group membership according to ethnicity, therefore, may have helped sustain a reputation mechanism (Greif 1993). Put differently, ethnic affiliation may be instrumental for constructing *trust*.

High level of trust may be mapped to various desirable economic outcomes. When individuals trust each other, transaction costs are reduced, organizations run better, the need for formal regulation reduces, governments provide services more efficiently, policy promises become more credible, and financial systems develop better (Finseraas et al. 2019). Moreover, lack of trust in out-group individuals may penalize proportionally more minorities if it is detrimental to inter-group interaction. At the workplace, co-ethnicity is linked to efficiency in contact across employees (Hjort 2014), and thus minority employees may be less productive due to the need to interact more frequently with out-group individuals. By the same reason, minorities can be

more penalized if lack of trust deters technology transmission across groups. This is because a larger group of individuals potentially innovates more simply due to a mechanic effect of its size and complexity.

Although *taste* and *trust* may be important mechanisms liking group size and economic activity, cultural traits, such as norms, religion, and language, could have similar effects on contact, by obstructing out-group and facilitating in-group interaction across individuals. Additionally, Laitin and Ramachandran (2022) shows that individuals have more years of education and higher reading proficiency when their indigenous language is the language of instruction at school in African countries. Evidently, if minorities have lower bargaining power to choose the language of instruction at schools, they will be disproportionately affected by this channel. Although taste, trust and cultural traits may have similar effects on interaction, and may actually be very entangled, trust and group identity related to taste may be more easily malleable (Lowe 2021, Depetris-Chauvin et al. 2020), Finseraas et al. 2019) than cultural practices. By identifying this distinction, one can argue by which of these channels group size has greatest effect on economic development.

Until this point we have focused on explanations that favor an *economic mechanism* for linking the two objects. But the relationship between them may be complicated by concurring and complementary explanations related to *political capital*. Comparative politics works have pointed that ethnic favoritism in politics exist throughout the world and despite of the system of government (De Luca et al. 2018), but is particularly important in African countries (Dickens 2018). Politicians often engage in redistribution towards their own personal homelands or co-ethnics through public services investments, such as in educational inputs (Kramon and Posner 2016) and roads (Burgess et al. 2015).

This political channel may impact group development if its size influences its access to formal politics. Francois et al. (2015) shows suggestive evidence that elites in the African continent go through some degree of bargaining over political power, as political coalitions share

ministerial positions highly proportionally to ethnic groups' population sizes. Therefore, ethnic groups that have members belonging to the political elites may be favored for redistribution. Investigating whether minority individuals are able to access the political elites may be key in understanding this link.

Finally, if minority groups face more obstacles to be formally represented and to have their preferences met by institutional politics, they may also rely more on customary institutions. This dynamic, in turn, may affect property rights and the rule of law. While these may have both positive and negative consequences for development - for example, by relying in informal institutions, individuals may have more trust in each other, but also lack sufficient rule of law and juridical complexity to foster innovation -, this is still a understudied subject.

3 Historical background

What started in the 1860 with the French and the British exploring systematically West Africa was followed by 40 years of Europeans signing treaties and agreements to partition the largely unexplored African continent into free-trade zones, protectorates, and colonies. The Berlin Conference took place from November 1884 until February 1885, and although its concrete goal was to set the boundaries of Central Africa, it is regarded among historians and political scientists as the landmark of ethnic partitioning in the continent, as it set down the guidelines that would be used to delimit the zones.

Previous works have discussed this event in great detail. For instance, Michalopoulos and Papaioannou (2016) argue that the key principle followed by the European powers was to preserve the *status quo* and prevent conflict over Africa, which resulted in borders being designed with little regard for local conditions. In fact, many drawing decisions were made before information from explorers, geographers, and missionaries could arrive to Europe (Michalopoulos and Papaioannou 2016),[3] while in the Berlin conference "there was no African representation, and African concerns were, if they mattered at all, completely marginal to the basic economic,

[3]This is evident in a speech delivered by British prime minister Robert Cecil in 1890:

strategic, and political interests of the negotiating European powers"(Asiwaju, 1985, p. 25).

Several explanations surfaced in African historiography to rationalize the largely accidental border design, among which have gathered relevance the following. First, it is emphasized that the continent had been little explored by Europeans at the time of the partitioning, and it was not uncommon for these powers to swipe pieces of land back and forth with limited idea of what they were worth. Similarly, neither locals or Europeans were foreseeing that these zones would became *states*: they were simply delimiting regions of influence. Moreover, demarcation on the ground was poor, and where it existed, locals could freely move across colonial borders in a way that resistance from the bottom does not seem to have been relevant or to have decidedly and systematically shaped the borders. With these accounts, Michalopoulos and Papaioannou (2016) runs a battery of regressions of an ethnic partitioning indicator (as consequence of border design) over a set of geographical, ecological, and some documented pre-colonial characteristics in linear probability models. It shows that there is no systematic association between these variables and the probability of an ethnic group being partitioned into two or more countries, which gives a formal foundation to the argument of the artificiality of border design, at least with respect to the tribes homelands.

As the literature of different areas of research have nearly settled on this argument, the debate over the independent African states' conformity with the borders is more nuanced and empirically less convincing due to data limitations. Yet, Michalopoulos and Papaioannou (2016) show suggestive evidence that in the aftermath of the borders' drawing, split and non-split ethnic groups remained very balanced along all relevant geographical, ecological, and natural resources characteristics. In fact, with the independence of 40 African states happening within a short period of time over the 1950s and first half of the 1960s, nearly all of them signed the Charter of the Organization of the African Union (OUA) in 1964 accepting colonial

"*We have been engaged in drawing lines upon maps where no white man's feet have ever tord; we have been giving away mountains and rivers and lakes to each other, only hindered by the small impediment that we never knew exactly where the mountains and rivers and lakes were.*"(Times, August 7, 1890)

borders (with exception of Somalia and Morocco). From the departing Europeans side, with interest in maintaining the special rights and commerce deals with their former colonies, the issue of border design was a timely one to be avoided, while African leaders could have found it threatening that border realignment would imperil their position. Therefore, it is largely accepted in the literature that borders have, for the most part of the continent, remained unchanged since colonial times.

4 Data description

4.1 Location of historical ethnic homelands and delimitation of partitioned groups

The sources of ethnic-level data are the Murdock map of ethnoliguistic regions (Murdock 1959) and the Ethnolinguistic Atlas (Murdock 1967), which together compose a anthropological database that is largely used across social sciences. Murdock (1959, 1967) compiled rich and systematic portraits of diverse societies with information on various aspects of life, such as subsistence, political organization, or social norms, based on numerous ethnographies from the 19th century. For each of the 843 ethnic groups identified in the African continent, Murdock (1959, 1967) provides the geographical coordinates and maps which have been digitized and made available by Nunn and Wantchekon (2011). The compiled map portrays ethnic homelands as partitions of the whole continent, including uninhabited regions, just before the Scramble for Africa for the near entirety of the tribes. As we cannot perfectly track how the spatial distribution of these ethnicities changed since then, in all exercises that will follow we use the ethnic population figures as of 1880 (Pre-Scramble). Reassuringly, other papers have shown that the distribution of ethnic groups from Murdock (1959, 1967) has not changed drastically in the last 150 years. Using data from the Afrobarometer, Nunn and Wantchekon (2011) report a correlation of 0.55 between the location of the respondents in 2005 and the historical homeland of their ethnicity as identified in Murdock's map. Recently, Bahrami-Rad et al. (2021) made an effort to link ethnicities from Murdock's map to corresponding ones reported in Standard Demographic and Health Surveys (DHS), using the complete version of historical

societies in Murdock's map (around 1200 groups). Using self-reported data from 790,000 individuals across 43 countries, it then validates the use of Murdock's map by showing that there are positive associations between the historical measures collected by ethnographers and the current practices reported in DHS by the descendants of these historical societies.

To identify partitioned ethnicities and assign each area to the respective country, we intersect Murdock's map with the Global Administrative Database (GADM), that portrays contemporary national boundaries. With the overlaying regions, we are able to generate polygons of the intersections and identify split ethnic groups. We take a conservative approach and restrict the sample of split ethnic groups to the ones which had at least 5% of their population laying in one side of the border, according to the historical (gridded) population estimates provided by the History database of the Global Environment (HYDE), the main data source for population estimates. This choice accounts for potential imprecision of border lines in Murdock (1959, 1967), as well as the possibility that individuals that ended up in countries where their ethnic group is a small share may have more easily migrated to bordering countries where their ethnic group was larger. Figure 1 panel A displays the simple overlay of both maps, and panel B emphasizes the partitioned tribes whose population shares are above the 5% threshold.

4.2 Satellite light density at night

With the aim of comparing subnational levels, we require detailed spatial data on economic development. To the best of our knowledge, geocoded high-resolution measures of economic development spanning all of Africa are not available. To overcome this limitation, we use satellite light density at night to proxy for local economic activity. The luminosity data come from the Visible Infrared Imaging Radiometer Suite (VIIRS) sensor on board the polar-orbiting Suomi National Polar-orbiting Partnership (Suomi NPP) and NOAA-20 weather satellite, and made available by the Earth Observation Group Elvidge et al. (2021). We use the product Annual VNL V2 for 2019, produced from monthly cloud-free average radiance for every 15 arc second grid cells (approximately 500 meters at the Equator line), in the form of raster files.

The satellite detects lights from human settlements, fires, gas flares, lightning, and the aurora. We construct average light density per square kilometer for 2019, averaging across pixels at the desired level of aggregation.

The measure available is expressed in terms of nano Watts per steradian per square centimeter, a measure of radiance per area. Therefore, at the pixel level (the finest grid), this measure reflects not only light availability (e.g. a pixel is lit or not), but *light density* (the intensity of light use), aggregated for the whole year in question.

Several other papers have used luminosity data to proxy for development. Earlier contributions and diffusion of this method can be credited to the work of Henderson et al. (2011), along with previous work that has shown that light density is a robust proxy of economic activity (see Elvidge et al. (1997), Doll et al. (2006), and Pinkovskiy (2011)). These studies establish a strong within-country correlation between light density at night and GDP levels and growth rates. Even Chen et al. (2016), who stressed some of the problematic issues with using satellite image data, argued that luminosity can be quite useful for regional analysis in countries with poor-quality income data.

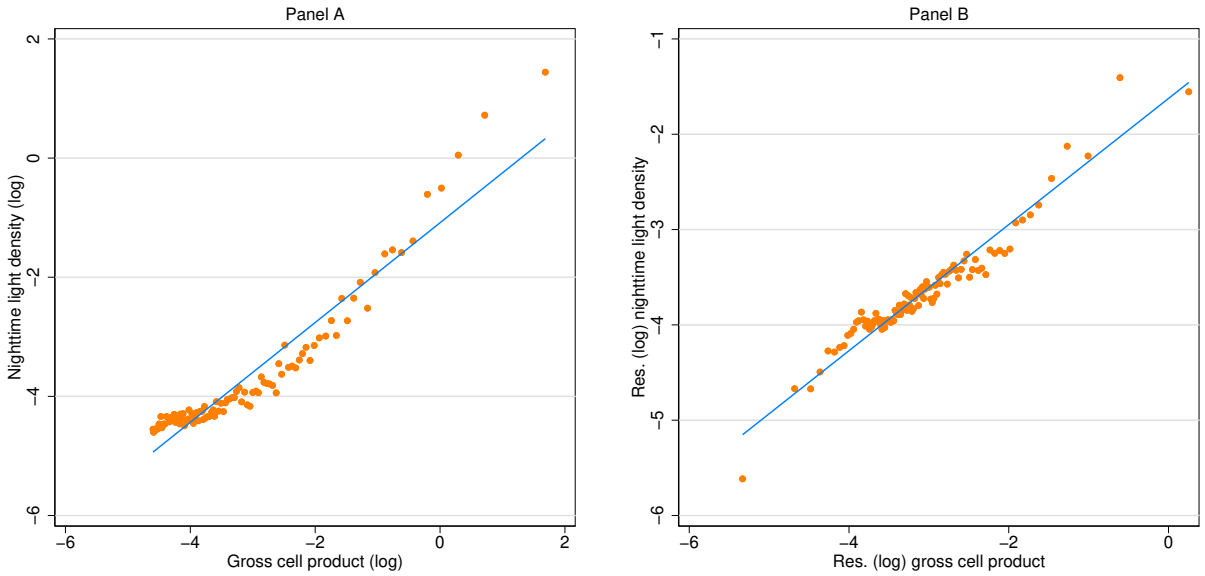
4.3 Cross-validation: satellite light density and regional development

Nordhaus et al. 2006 developed a geophysically scaled economic data set (hence, G-Econ), that consists of 1:1 degree grid cells for all terrestrial regions and is available for four years, 1990, 1995, 2000, and 2005. Its main purpose is to provide consistent economic activity data for subnational levels, comparable throughout the world. Importantly, its main innovation, the gross cell product (GCP), a version compatible with a regional gross product for the grid cell unit, takes into account the geographical intensity of economic activity without directly using measures of nighttime light intensity, as it uses primarily harmonized measures of the System of National Accounts. As a measure of gross value added, we can validate the use of light intensity by aggregating it at the 1:1 grid cell level and comparing it to the available GCP.

We do this exercise in figure 2, which uncovers the highly positive relationship between light

luminosity and the GCP in African countries, both unconditionally (panel A) and conditionally on country and year fixed effects and other country characteristics available at the grid level (panel B). These results reassure our choice of proxy for economic development, since we cannot use the GCP directly due to its level of aggregation (around 100 km by 100 km) being higher than some of the partitioned ethnic homelands.

Figure 2: Plots of light density versus gross cell product at the grid-cell level for Africa



Note: Panel A is a binned plot of grid-cell averages of nighttime light density and gross cell product (Nordhaus et al. 2006) for years 2000 and 2005 (last available) for the whole African continent calculated by the Peace Research Institute Oslo (PRIO). Panel B displays the binned residualized variables after controlling for year, country, existence of diamond mine and petroleum exploration, distance to the country's capital, and population count. Because of the skewed distribution of both measures and because of 0 values, we follow the literature and transform them using the operator $\ln(\cdot + 0.01)$.

4.4 Other data sources

Other data included in the exercises come mainly from georreferenced files, such as shapefiles and raster images, and are aggregated at the geographic level of the analysis. These are the Historical Database of the Global Environment - HYDE (historical population estimates, irrigated soil, pasture land, cropland, and others); the Global Agro-Ecological Zoning - GAEZ/FAO (soil crop suitability, water availability, altitude, etc); The Malaria Project (malaria suitability); the Peace Research Institute Oslo - PRIO (occurrence of onshore petroleum and precious gemstones); and NASA (such as the global roads database, gROADS, and distance from the coast). We complement these data with information made available by Nunn

and Wantchekon (2011), such as colonial railways, and Nunn and Puga (2012), such as terrain ruggedness and slope. In order to aggregate the variables at the desired level, we use the weights given by the grid cell area definition in each of the data sets, when available.

5 Effects of ethnic group size on economic development

5.1 Empirical strategy

The source of variation explored is the population size across partitioned groups. Therefore, we estimate the following regression model at the level of the tribe-country unit in the sample of partitioned ethnic groups

$$Y_{e,c,2019} = \delta_e + \delta_c + \beta_1 Size_{e,c,1880} + X_{e,c}\beta_2 + \varepsilon_{e,c} \tag{1}$$

where indices e, c denote, respectively, ethnicity and country, and δs are fixed effects. The outcome of interest is the average nighttime light intensity at the country-tribe level. The variable of interest, $Size$, is calculated from gridded population estimates from the 1880s in order to capture effects disregarding endogenous migration that has happened since then. As discussed in section 4.1, this approach is endorsed by previous works that have documented a significant persistence in location decisions since the Scramble for Africa throughout the continent. Therefore, our estimates can be interpreted as lower bounds or as intention-to-treat. The inclusion of ethnicity- and country-specific effects allows us to overcome omitted variable bias stemming from group-specific characteristics (e.g. precolonial institutions, precolonial development, etc), and country attributes, such as overall economic development, and country-specific policies and institutions, allowing us to emulate a quasi-natural experiment.

Table 1 presents descriptive statistics for the main variables at the ethnicity-country level. Although the light density at pixel level varies from 0 to 50664.59 $nW/cm^2/sr$ in Africa, the aggregated values by the mean at the country-tribe level are quite low (around 0.11 $nW/cm^2/sr$), while a direct measure of quantity, such as the share of lit pixels in the continent, is around 3%. Although consistent with low development levels, these values uncover large heterogeneity

across regions. Both the distribution of light density and population size are highly skewed to smaller values. To smooth the light density distribution and given that it has some null values, we follow the literature and apply a logarithmic transformation after the addition of a small number, set to 0.01 as in Michalopoulos (2012) and Alesina et al. (2016).

Table 1: Summary statistics at the ethnicity-country group level

	Mean	SD	Min	Max	N
Average luminosity in 2019, $nW/cm^2/sr$	0.11	0.64	0.00	13.25	1314
Share of lit pixels in 2019	0.03	0.08	0.00	0.98	1314
Ethnic group is split, unconditional	0.64	0.48	0.00	1.00	1314
Ethnic group is split — population share $\geq 1\%$ in one country	0.53	0.50	0.00	1.00	1314
Ethnic group is split — population share $\geq 5\%$ in one country	0.41	0.49	0.00	1.00	1314
Ethnic group is split — population share $\geq 10\%$ in one country	0.33	0.47	0.00	1.00	1314
Surface area (km^2)	22,636.33	43,081.62	0.01	494,958.34	1314
Share of country's area	0.04	0.08	0.00	0.94	1314
Average population count in 1880 by km^2	601.73	1,389.34	0.00	35,662.68	1314
Average population count in 2017 by km^2	6,662.73	19,750.33	0.00	544,643.94	1314
Population share of ethnic homeland in country (1880 population)	0.04	0.09	0.00	0.97	1314
Population share of ethnic homeland in country (2017 population)	0.04	0.09	0.00	0.98	1314
Total population in 1880, base 1000	82.76	262.07	0.00	5,932.42	1314
Total population in 2017, base 1000	882.20	2,950.21	0.00	81,717.84	1314

5.2 Results

First, given the exogenous nature of the event of border design with respect to the tribes' homelands, one should not expect to find association across ethnic population size and geography and land endowment variables that are invariant to land mass. Accordingly, in appendix tables A1 and A2 we show that, conditional on ethnicity and country, the tested outcomes do not exhibit systemic association with the size of the population, with the (marginal) exception of measures that reflect the median altitude of the area. This potentially uncovers the fact that larger land areas (positively associated with population size) have their centroids more distant from the coast, which could explain these differences in elevation. Moreover, as distance to coast is only marginally significant, we do not draw much conclusion from it, but it is convenient to note that the positive coefficient implies that larger groups are more distant from the shore, and this could only act in he opposite direction of the expected effect of group size on development (only making it smaller). Also, there are no consistent differences across

land use and suitability. Nonetheless, we test the stability of the estimates to the inclusion of these variables.

Table 2 displays the effects of the log of population size on the proxy for economic outcomes, while the columns test slightly different specifications that combine ethnicity and country fixed effects with two sets of control variables for geography and agricultural resources. As the sample is restricted to partitioned groups, we are comparing economic outcomes across the *same* groups residing on both sides of country borders. In other words, in matching groups that share the same unobservable characteristics, and purging specific effects of ethnicity and country, we compare the level of development for the same group on both sides of the border, and therefore the estimator of the parameter of interest will represent a sort of quasi-experiment as it flushes away ethnic and country characteristics that might affect estimates.

The complete specification in column (7) documents that a 10% increase in population size increases nighttime light density by about 1.7%. This point estimate is overall stable over the inclusion of fixed effects, as columns (1) to (4) show, and robust to the inclusion of the controls tested in the appendix tables (columns (5) to (7)).

Table 2: Effects of ethnic group size on economic development

	Light density (log)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Population in 1880 (log)	0.193***	0.130***	0.244***	0.179***	0.181***	0.150***	0.172***
	(0.038)	(0.043)	(0.032)	(0.032)	(0.034)	(0.039)	(0.044)
Ethnic FE		✓		✓	✓	✓	✓
Country FE			✓	✓	✓	✓	✓
Geography					✓		✓
Agricultural resources						✓	✓
# Observations	540	540	537	535	528	535	528
R^2	0.070	0.789	0.480	0.877	0.886	0.904	0.913
Mean Dep. Var.	-3.767	-3.767	-3.780	-3.789	-3.808	-3.789	-3.808
S.D. Dep. Var.	1.106	1.106	1.093	1.085	1.065	1.085	1.065

* p <0.10 ** p <0.05 *** p <0.01. The table presents the estimates for equation 1 in the sample of partitioned ethnic groups whose population shares across borders are above 5% (see section 4.1). Geography controls are distance to coast (log), indicator for being landlocked, terrain ruggedness index, average slope, average precipitation, and malaria suitability index. Agriculture controls are share of irritated area in 1880, share of grazing area in 1880, share of pasture area in 1880, share of cropland area in 1880, mean altitude, maximum altitude, crop suitability index, land suitability level, share of water sites, dominant soil, agro-zones, and hydro-basin class. Standard errors double-clustered for ethnicity and country.

6 Channels

We now turn to the analysis of potential channels through which the size of partitioned groups may affect development. In principle there are several competing theories of why the size of ethnic groups may matter for development. We group these theories in three categories and provide suggestive evidence for each.

6.1 The role of natural endowments

The first approach we discuss relates to the persistence in development associated with the supply or availability of natural endowments. Early settlement decisions may have been shaped by specific favorable endowments in place, thus more naturally suitable places may have concentrated the biggest part of the ethnic homeland population, implying that the subsequent border design, while partitioning areas into different sizes, may be just capturing these effects, which could have persisted throughout time. The literature on development economics commonly agrees that geography and natural endowments can have important consequences for economic outcomes. For instance, access to large bodies of water likely facilitates agricultural production, while the existence of reserves of tradable natural resources directly affects economic activity. However, the debate over the precise role of natural conditions still persists, with some stressing its indirect role on current economic development through its effect on the types of institutions that were ultimately attracted and persisted in time (Acemoglu et al. 2001, Acemoglu et al. 2002), and others emphasizing direct effects of natural environment on economic outcomes (Diamond and Renfrew (1997)).

We assemble natural endowments data from many sources to extend the appendix tables A1 and A2, and test the association between population size and availability of natural amenities, repeating the exercise from equation 1 on these new outcome variables. Table 3 shows that larger groups are more likely to have lakes and diamond mines in their territory, while there is no increased likelihood of having oil fields or other gems mines, or greater supply of river water.

Table 3: Ethnic group size and natural endowment supply

	Natural endowments							
	(1) River extension per km^2 of land	(2) Lake indicator	(3) Onshore oil fields indicator	(4) Diamond mines indicator	(5) Other gems mines indicator	(6) Quantity of oil fields	(7) Quantity of diamond mines	(8) Quantity of gems mines
Population in 1880 (log)	-0.002	0.040**	0.013	0.074***	0.010	-0.004	0.243**	0.066
	(0.001)	(0.016)	(0.015)	(0.021)	(0.015)	(0.030)	(0.095)	(0.043)
Ethnic FE	✓	✓	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓	✓	✓
Geography and agricultural resources	✓	✓	✓	✓	✓	✓	✓	✓
# Observations	528	528	528	528	528	528	528	528
R^2	0.782	0.818	0.817	0.758	0.809	0.849	0.702	0.828
Mean Dep. Var.	0.078	0.063	0.057	0.108	0.063	0.108	0.373	0.136
S.D. Dep. Var.	0.022	0.242	0.232	0.311	0.242	0.576	1.929	0.664

* p <0.10 ** p <0.05 *** p <0.01. The table presents the estimates for equation 1 in the sample of partitioned ethnic groups whose population shares across borders are above 5% (see section 4.1). Geography controls are distance to coast (log), indicator for being landlocked, terrain ruggedness index, average slope, average precipitation, and malaria suitability index. Agriculture controls are share of irritated area in 1880, share of grazing area in 1880, share of pasture area in 1880, share of cropland area in 1880, mean altitude, maximum altitude, crop suitability index, land suitability level, share of water sites, dominant soil, agro-zones, and hydro-basin class. Standard errors double-clustered for ethnicity and country.

6.2 Investment attraction and economic relevance

If population size matters for economic activity, regions occupied by larger groups may turn into economic hubs. These, in turn, attract a series of infrastructure investments and stand out in other measures of economic relevance. We assess these possibilities in the next two tables. First, table 4 shows evidence that larger groups have access to a larger railway network per area of land, while no effect is found on railways constructed during colonial times, which may suggest that such investments were made post-independence (likely in the form of expansion of the colonial investments (see Jedwab and Storeygard (2019))). There is no evidence of larger groups attracting road infrastructure, using a common data source for the whole continent.[4]

[4]We use the Global Roads Open Access Data Set (gROADS), v.1 (1980–2010) provided by SEDAC/NASA. This database combines available roads data by country into a global roads coverage using a common model. The major limitations of this database is that roads' sources vary for each country, with potentially big differences in precision. Most of the roads in Africa could not be classified into a functional class (explored in the table 4), and there is no information on the operation status of each road, although the data set has this information for other countries. Yet, this is the only database that covers the entire continent, and integrates cross-border roads.

Table 4: Ethnic group size and infrastructure investment

	Distance in log		**Extension per km^2 of land**						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Distance to nearest road	Distance to nearest railway	All roads	Highways	Highways and primary roads	Highways, primary, secondary, and tertiary roads	Local and trail roads	Railways (base 10^{-3})	Colonial railways (base 10^{-3})
Population in 1880 (log)	-0.091	-0.035	0.004	-0.000	0.000	0.002	0.004	0.001**	0.000
	(0.068)	(0.040)	(0.004)	(0.000)	(0.001)	(0.001)	(0.002)	(0.000)	(0.000)
Ethnic FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Geography and agricultural resources	✓	✓	✓	✓	✓	✓	✓	✓	✓
# Observations	528	528	528	528	528	528	528	528	528
R^2	0.803	0.926	0.929	0.777	0.908	0.906	0.969	0.781	0.900
Mean Dep. Var.	8.153	11.609	0.114	0.001	0.005	0.018	0.038	0.002	0.001
S.D. Dep. Var.	1.263	1.164	0.115	0.003	0.017	0.048	0.100	0.006	0.003

* p <0.10 ** p <0.05 *** p <0.01. The table presents the estimates for equation 1 in the sample of partitioned ethnic groups whose population shares across borders are above 5% (see section 4.1). Geography controls are distance to coast (log), indicator for being landlocked, terrain ruggedness index, average slope, average precipitation, and malaria suitability index. Agriculture controls are share of irritated area in 1880, share of grazing area in 1880, share of pasture area in 1880, share of cropland area in 1880, mean altitude, maximum altitude, crop suitability index, land suitability level, share of water sites, dominant soil, agro-zones, and hydro-basin class. Standard errors double-clustered for ethnicity and country. Distances are measured from a polygon's centroids to the nearest lines in meters.

Table 5 shows that even though larger groups are not more likely to have the country's capital within their homeland (column (2)), the distance to the country's capital is in fact shorter (column (1)). This is relevant because economic mobility in Africa is highly associated with migration to urban centers (mostly comprised by the capital). Moreover, distance to capital is associated to access to markets which can be used either to buy important inputs (e.g fertilizers) or sell agricultural surplus. Another relevant economic factor is the language distance to other co-citizens in the country – which can affect communication, transmission of technology and human capital accumulation. We link ethnic groups from Murdock's map to the language trees available in Ethnologue (Gordon Jr 2005) using the matching algorithm LEDA (Müller-Crepon et al. 2022), described in the next section. After matching each ethnic group to the language tree, the linguistic distance between two groups i, j is given by

$$distance_{i,j} = 1 - \left(\frac{d(L_i, R) + d(L_j, R) - d(L_i, L_j)}{d(L_i, R) + d(L_j, R)} \right)^{\delta} \tag{2}$$

where the function $d(L, R)$ denotes the distance between language or dialect L to the root of the tree R in terms of segments of line between them, and $d(L_i, L_j)$ does the same for the two languages, and δ is a factor used to discount distances further away from the root and is usually set to 0.5. Therefore, each split ethnicity is matched to all other ethnic groups that are identified to have their ethnic homelands within the same country, and the distances are

calculated according to equation 2. We then calculate the weighted average distance for each ethnic group-country combination, using weights given by each ethnic group's population count in the same country as of 1880. The results displayed in columns (3) and (4) are consistent with larger groups having lower linguistic distance to their co-citizens, either when averaging the index including the linguistic distance to itself (always 0) as in column (4), or excluding the distance to itself as in column (3).

Table 5: Effects of ethnic group size on economic relevance

	Proxies for economic relevance			
	(1) Distance to country capital	(2) Country capital within homeland	(3) Linguistic distance to co-citzens (with self)	(4) Linguistic distance to co-citzens (without self)
Population in 1880 (log)	-0.082***	0.016	-0.017**	-0.017**
	(0.029)	(0.010)	(0.007)	(0.008)
Ethnic FE	✓	✓	✓	✓
Country FE	✓	✓	✓	✓
Geography and agricultural resourcs	✓	✓	✓	✓
# Observations	525	525	525	525
R^2	0.901	0.777	0.925	0.923
Mean Dep. Var.	12.890	0.029	0.466	0.483
S.D. Dep. Var.	0.795	0.167	0.236	0.241

* p <0.10 ** p <0.05 *** p <0.01. The table presents the estimates for equation 1 in the sample of partitioned ethnic groups whose population shares across borders are above 5% (see section 4.1). Geography controls are distance to coast (log), indicator for being landlocked, terrain ruggedness index, average slope, average precipitation, and malaria suitability index. Agriculture controls are share of irritated area in 1880, share of grazing area in 1880, share of pasture area in 1880, share of cropland area in 1880, mean altitude, maximum altitude, crop suitability index, land suitability level, share of water sites, dominant soil, agro-zones, and hydro-basin class. Standard errors double-clustered for ethnicity and country. Distance to capital is in log and is measured from the polygon's centroid to the coordinates of the capital city in meters.

6.3 Political capital acquisition

In order to assess whether larger groups benefit from greater access to political power, we match each of the tribe-country groups to the ethnic-country groups defined in the Ethnic Power Relations family data sets (Vogt et al. 2015). We describe the this collection of data and the matching procedure below in the next two subsections.

6.3.1 The Ethnic Power Relations Family

The EPR Core dataset provides annual data for *politically relevant* ethnic groups and their access to state power from 1946 to 2019. This core data set is complemented by geo-spatial information on ethnic groups' settlement patterns, data on ethnic groups' links to rebel organizations, on the trans-border relations of ethnic groups, and on intraethnic cleavages, composing the EPR Family of data sets. EPR data provide a comprehensive selection of politically relevant ethnic groups, including minority and majority, and discriminated as well as state-controlling groups.

The EPR data set defines ethnicity "as a subjectively experienced sense of commonality based on a belief in common ancestry and shared culture" (Vogt et al. 2015). Therefore, different markers may be used to indicate such shared ancestry and culture in each country, such as a common language, similar phenotypical features, or adherence to the same faith. Moreover, an ethnic group is considered *politically relevant* if at least one political organization has claimed to represent its interests at the national level or if its members are subjected to state-led political discrimination. The data set uses the following nomenclatures to classify politically relevant actors: Monopoly or Dominance, when the group rules alone; Senior or Junior Partner, when the group shares power; and Powerless, Discrimination, or Self-Exclusion, when the group is excluded from power.[5]

We use the whole ethnicity-country yearly time series of EPR data until 2017. Among the many measures of an ethnic group's political status, we use the ones most directly associated with access to power and representation in formal politics, which are (i) whether the ethnic group occupies highest political positions in the country in 2017, having either monopoly or dominance over political power, or being senior partner; (ii) whether the group is powerless or discriminated (in which we include self-exclusion); (iii) the share of years the ethnic group occupies highest political positions since the country's independence, which is calculated as

[5]For more details about these categories, we refer to EPR's codebook (2021).

the ratio between the number of years a group is coded as in (i) and the difference between 2017 and the year of the country's independence (i.e. the year the country is coded into EPR).

6.3.2 The Linking Ethnic Data from Africa algorithm

As each data set employs a different definition for ethnic group, matching different data sets at the ethnicity level is challenging. The LEDA algorithm (Müller-Crepon et al. (2022)) allows one to link ethnic groups from different data sets based on the groups' match to the linguistic trees available in Ethnologue (Gordon Jr 2005). To match ethnicities from Murdock's map and EPR, we employ the algorithm's option to link ethnic groups based on the highest overlap in the language tree departing from the dialect level (e.g. how many nodes two groups share on the tree) and respecting country boundaries. For each group from Murdock, we assign the corresponding EPR group or groups, concluding that groups with no match (e.g. those whose overlap to any other group is zero) are not represented in EPR, and are therefore not politically represented in the national level. Doing this, the average language tree overlap that we consider in the matching procedure is 0.9 (in a scale from 0 to 1).

6.3.3 Results

Table 6 shows that larger groups are more likely to be politically relevant in the present (column (1)), and to access political power and be formally represented at the national level (column(3)). As a result, groups with largest populations tend to occupy dominant positions in the national government both in the present (column (2)) and throughout the years since independence (column (4)). Additionally, larger groups seem to be more likely to engage in political conflict (column (5)), and this association can be attributed mainly to conflict over territory (column (6)), while involvement in conflict against the government does not seem relevant (column (7)). These last findings are overall consistent with population size being positively associated with power in national politics.

Table 6: Ethnic group size and political capital

	Measures of political capital						
	(1) Politically relevant in country	(2) Occupies highest political positions	(3) Powerless or discriminated	(4) Share of years in highest positions since independence	(5) Involved in conflict	(6) Involved in conflict over territory	(7) Involved in conflict with government
Population in 1880 (log)	0.118***	0.073***	-0.095***	0.054***	0.037***	0.023**	0.014
	(0.028)	(0.025)	(0.028)	(0.019)	(0.012)	(0.010)	(0.009)
Ethnic FE	✓	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓	✓
Geography and agricultural resources	✓	✓	✓	✓	✓	✓	✓
# Observations	525	525	525	525	525	525	525
R^2	0.787	0.786	0.771	0.797	0.725	0.696	0.751
Mean Dep. Var.	0.505	0.189	0.621	0.166	0.025	0.017	0.008
S.D. Dep. Var.	0.500	0.392	0.486	0.328	0.156	0.130	0.087

* p <0.10 ** p <0.05 *** p <0.01. The table presents the estimates for equation 1 in the sample of partitioned ethnic groups whose population shares across borders are above 5% (see section 4.1). Geography controls are distance to coast (log), indicator for being landlocked, terrain ruggedness index, average slope, average precipitation, and malaria suitability index. Agriculture controls are share of irritated area in 1880, share of grazing area in 1880, share of pasture area in 1880, share of cropland area in 1880, mean altitude, maximum altitude, crop suitability index, land suitability level, share of water sites, dominant soil, agro-zones, and hydro-basin class. Standard errors double-clustered for ethnicity and country.

7 Conclusion

We have studied the effect of group size on the ethnic-level economic performance. Our exercise is an indirect way to test the claim that ethnic diversity causes lower development due to heterogeneity of preferences, taste discrimination, political competition or lack of economies of scale (due to shared culture, values, language, among others).

We take advantage of the Scramble for Africa that in practice divided the African continent between the European powers with negligible considerations about the location of pre-existing ethnic tribes. This caused a natural experiment where groups were split across countries and generated exogenous variation on their size in each country.

We document that large groups have an economic advantage over their smaller counterparts. We find suggestive evidence in favor of the economies of scale and political power mechanism. Larger groups are more likely to have access to natural endowments, like water and diamond mines. In addition, larger groups are linguistically closer to the rest of country's population, which can facilitate interactions, trade, and transmission of knowledge. They are also closer to the capital city, which can facilitate profitable migration and access to markets. Last but not least, larger groups are more likely to be politically relevant, indicating the existence of a political organization claiming to represent them. As a result, they are also more likely to have co-ethnics in the executive power. In an environment plagued with ethnic favoritism, this

can imply advantage in patronage and allocation of public goods.

All together, our results highlight the relevance of ethnic barriers for the development process. The evidence also sheds light to another detrimental effect of the Scramble for Africa. By not taking into account the pre-existing settlement of ethnic groups, their partitioning have caused the presence of smaller groups within countries, who were eventually discriminated or less favored by economies of scale or political representation.

References

D. Acemoglu, S. Johnson, and J. A. Robinson. The colonial origins of comparative development: An empirical investigation. *American economic review*, 91(5):1369–1401, 2001. 16

D. Acemoglu, S. Johnson, and J. A. Robinson. Reversal of fortune: Geography and institutions in the making of the modern world income distribution. *The Quarterly journal of economics*, 117(4): 1231–1294, 2002. 16

A. Alesina and E. L. Ferrara. Ethnic diversity and economic performance. *Journal of economic literature*, 43(3):762–800, 2005. 1

A. Alesina and E. La Ferrara. Participation in heterogeneous communities. *The Quarterly Journal of Economics*, 115(3):847–904, 2000. 5

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

A. Alesina, S. Michalopoulos, and E. Papaioannou. Ethnic inequality. *Journal of Political Economy*, 124(2):428–488, 2016. 2, 14

A. I. Asiwaju. *Partitioned Africans: Ethnic relations across Africa's international boundaries, 1884-1984*. C. Hurst & Co. Publishers, 1985. 8

D. Bahrami-Rad, A. Becker, and J. Henrich. Tabulated nonsense? testing the validity of the ethnographic atlas. *Economics Letters*, 204:109880, 2021. 9

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

D. L. Chen, T. J. Moskowitz, and K. Shue. Decision making under the gambler's fallacy: Evidence from asylum judges, loan officers, and baseball umpires. *The Quarterly Journal of Economics*, 131 (3):1181–1242, 2016. 11

G. De Luca, R. Hodler, P. A. Raschky, and M. Valsecchi. Ethnic favoritism: An axiom of politics? *Journal of Development Economics*, 132:115–129, 2018. 6

E. Depetris-Chauvin, R. Durante, and F. Campante. Building nations through shared experiences: Evidence from african football. *American Economic Review*, 110(5):1572–1602, 2020. 6

J. Diamond and C. Renfrew. Guns, germs, and steel: The fates of human societies. *Nature*, 386 (6623):339–339, 1997. 16

A. Dickens. Ethnolinguistic favoritism in african politics. *American Economic Journal: Applied Economics*, 10(3):370–402, 2018. 6

C. N. Doll, J.-P. Muller, and J. G. Morley. Mapping regional economic activity from night-time light satellite imagery. *Ecological Economics*, 57(1):75–92, 2006. 11

W. Easterly and R. Levine. Africa's growth tragedy: policies and ethnic divisions. *The quarterly journal of economics*, pages 1203–1250, 1997. 1

C. D. Elvidge, K. E. Baugh, E. A. Kihn, H. W. Kroehl, E. R. Davis, and C. W. Davis. Relation between satellite observed visible-near infrared emissions, population, economic activity and electric power consumption. *International Journal of Remote Sensing*, 18(6):1373–1379, 1997. 11

C. D. Elvidge, M. Zhizhin, T. Ghosh, F.-C. Hsu, and J. Taneja. Annual time series of global viirs nighttime lights derived from monthly averages: 2012 to 2019. *Remote Sensing*, 13(5):922, 2021. 10

H. Finseraas, T. Hanson, Å. A. Johnsen, A. Kotsadam, and G. Torsvik. Trust, ethnic diversity, and personal contact: A field experiment. *Journal of Public Economics*, 173:72–84, 2019. 5, 6

P. Francois, I. Rainer, and F. Trebbi. How is power shared in africa? *Econometrica*, 83(2):465–503, 2015. 6

R. G. Gordon Jr. Ethnologue, languages of the world. *http://www. ethnologue. com/*, 2005. 18, 21

A. Greif. Contract enforceability and economic institutions in early trade: The maghribi traders' coalition. *American Economic Review*, pages 525–548, 1993. 5

V. Henderson, A. Storeygard, and D. N. Weil. A bright idea for measuring economic growth. *American Economic Review*, 101(3):194–99, 2011. 11

J. Hjort. Ethnic divisions and production in firms. *The Quarterly Journal of Economics*, 129(4): 1899–1946, 2014. 5

R. Jedwab and A. Storeygard. Economic and political factors in infrastructure investment: evidence from railroads and roads in africa 1960–2015. *Economic History of Developing Regions*, 34(2): 156–208, 2019. 17

E. Kramon and D. N. Posner. Ethnic favoritism in education in kenya. *Quarterly Journal of Political Science*, 11(1):1–58, 2016. 6

D. D. Laitin and R. Ramachandran. Linguistic diversity, official language choice and human capital. *Journal of Development Economics*, 156:102811, 2022. 6

M. Lowe. Types of contact: A field experiment on collaborative and adversarial caste integration. *American Economic Review*, 111(6):1807–44, 2021. 6

S. Michalopoulos. The origins of ethnolinguistic diversity. *American Economic Review*, 102(4):1508–39, 2012. 1, 14

S. Michalopoulos and E. Papaioannou. Pre-colonial ethnic institutions and contemporary african development. *Econometrica*, 81(1):113–152, 2013. 2

S. Michalopoulos and E. Papaioannou. The long-run effects of the scramble for africa. *American Economic Review*, 106(7):1802–48, 2016. 2, 7, 8

C. Müller-Crepon, Y. Pengl, and N.-C. Bormann. Linking ethnic data from africa (leda). *Journal of Peace Research*, 2022. 18, 21

G. P. Murdock. Africa, its peoples and their culture history. 1959. 3, 9, 10

G. P. Murdock. Ethnographic atlas: a summary. *Ethnology*, 6(2):109–236, 1967. 2, 3, 9, 10

W. Nordhaus, Q. Azam, D. Corderi, K. Hood, N. M. Victor, M. Mohammed, A. Miltner, and J. Weiss. The g-econ database on gridded output: methods and data. *Yale University, New Haven*, 6, 2006. 11, 12

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

N. Nunn and D. Puga. Ruggedness: The blessing of bad geography in africa. *Review of Economics and Statistics*, 94(1):20–36, 2012. 13

N. Nunn and L. Wantchekon. The slave trade and the origins of mistrust in africa. *American Economic Review*, 101(7):3221–52, 2011. 9, 12

M. Pinkovskiy. Economic discontinuities across borders: Evidence from satellite data on lights at night'. report, 2011. 11

H. Tajfel, M. G. Billig, R. P. Bundy, and C. Flament. Social categorization and intergroup behaviour. *European Journal of Social Psychology*, 1(2):149–178, 1971. 5

M. Vogt, N.-C. Bormann, S. Rüegger, L.-E. Cederman, P. Hunziker, and L. Girardin. Integrating data on ethnicity, geography, and conflict: The ethnic power relations data set family. *Journal of Conflict Resolution*, 59(7):1327–1342, 2015. doi: 10.1177/0022002715591215. URL `http://dx.doi.org/10.1177/0022002715591215`. 19, 20

A Balance checks across partitioned groups

Table A1: Balance of natural endowment supply

	Geography variables							
	(1) Distance to coast (log)	(2) Median altitude	(3) Rugged terrain index	(4) Average slope (percent)	(5) Yearly precipitation (1970-2000)	(6) Rainy days (year)	(7) Average temperature (year)	(8) Malaria suitability
Population in 1880 (log)	0.019*	21.313***	0.028	0.077	-0.700	0.005	-0.100**	-0.001
	(0.010)	(7.018)	(0.024)	(0.064)	(0.473)	(0.429)	(0.040)	(0.002)
Ethnic FE	✓	✓	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓	✓	✓
# Observations	535	535	535	535	535	535	535	528
R^2	0.988	0.963	0.904	0.904	0.992	0.994	0.972	0.973
Mean Dep. Var.	5.946	603.793	0.640	1.716	84.970	152.752	26.020	0.203
S.D. Dep. Var.	1.206	421.828	0.724	1.945	52.249	77.315	2.823	0.158

* p <0.10 ** p <0.05 *** p <0.01. The unit of analysis is the tribe-country group, and the sample is restricted to partitioned ethnic groups with at least 5% of its population as of 1880 falling at one side of the border. Each column reports a different regression of a geographic variable against the logarithmic of the population count in 1880. All regressions use fixed effects for ethnic group and country. Standard errors double-clustered for ethnicity and country.

Table A2: Balance of natural endowment supply

	Land use and endowments								
	(1) Irrigated area in 1880 (percent)	(2) Rain-fed area in 1880 (percent)	(3) Grazing area in 1880 (percent)	(4) Pasture area in 1880 (percent)	(5) Rangeland area in 1880 (percent)	(6) Cropland area in 1880 (percent)	(7) Crop suitability index	(8) Land suitability index	(9) Water site (percent)
Population in 1880 (log)	0.006	-0.096	0.302	0.369**	-0.029	-0.090	0.075*	60.836	-1.077
	(0.006)	(0.152)	(0.290)	(0.172)	(0.308)	(0.152)	(0.038)	(84.437)	(0.690)
Ethnic FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
# Observations	535	535	535	535	535	535	535	535	535
R^2	0.700	0.906	0.902	0.852	0.895	0.907	0.860	0.947	0.664
Mean Dep. Var.	0.028	3.157	14.283	3.267	9.647	3.185	2.445	4770.193	3.485
S.D. Dep. Var.	0.185	4.437	11.878	5.674	11.715	4.451	0.951	3598.435	14.451

* p <0.10 ** p <0.05 *** p <0.01. The unit of analysis is the tribe-country group, and the sample is restricted to partitioned ethnic groups with at least 5% of its population as of 1880 falling at one side of the border. Each column reports a different regression of a land endowment variable against the logarithmic of the population count in 1880. All regressions use fixed effects for ethnic group and country. Standard errors double-clustered for ethnicity and country.