AEA PAPERS AND PROCEEDINGS MAY 2026

Online Supplement to:

A New Perspective on Spatial Heterogeneity in African Development

by Pascaline Dupas, Nicolas Suarez, Zhongyi Tang

TABLE OF CONTENTS

APPENDIX FIGURES

Figure A1. : Location of Afrobarometer Round 6 Enumeration Areas

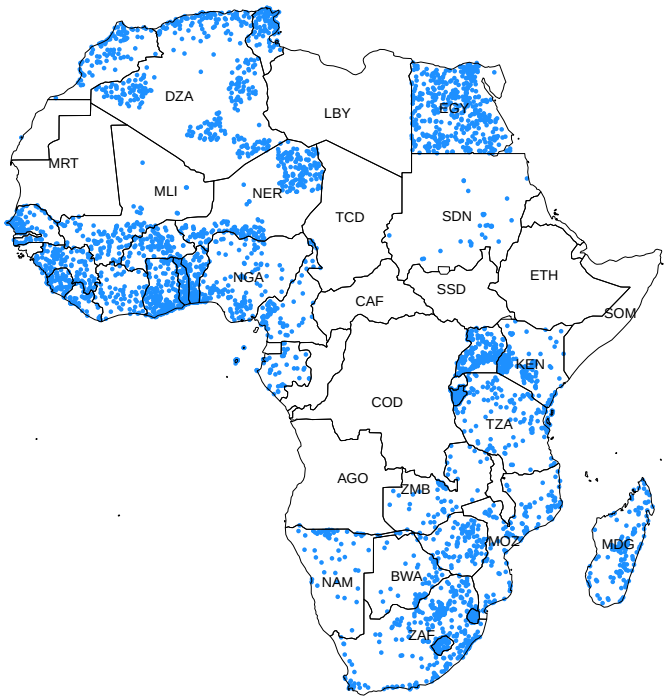


Figure A2. : Predictions: Zooming in on Cote d'Ivoire, Ghana and Togo

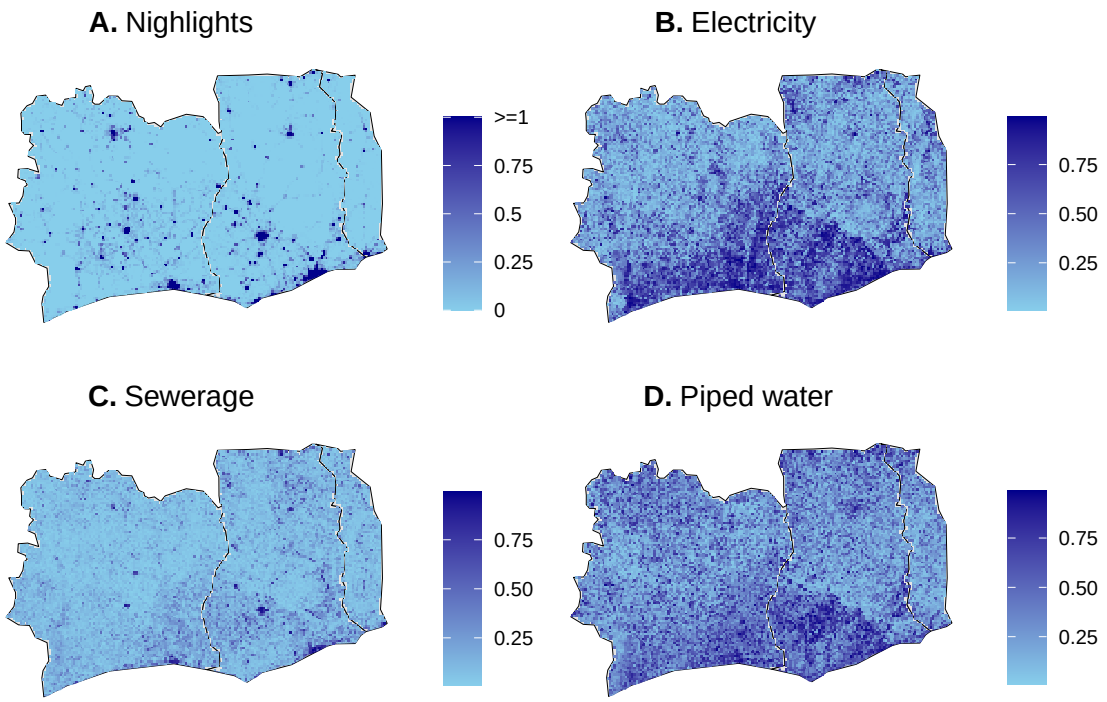


Figure A3. : Comparison of outcome distributions

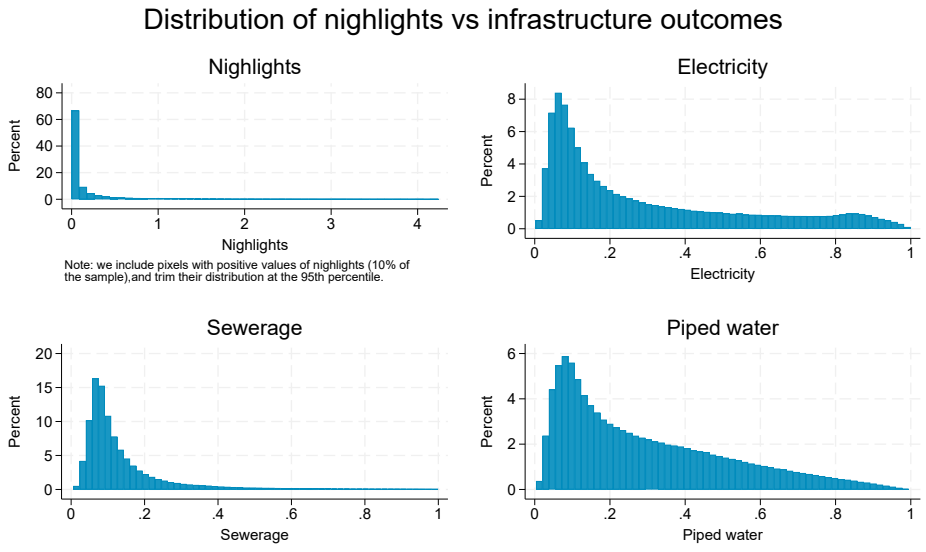


Figure A4. : National Institutions: Index of institutions

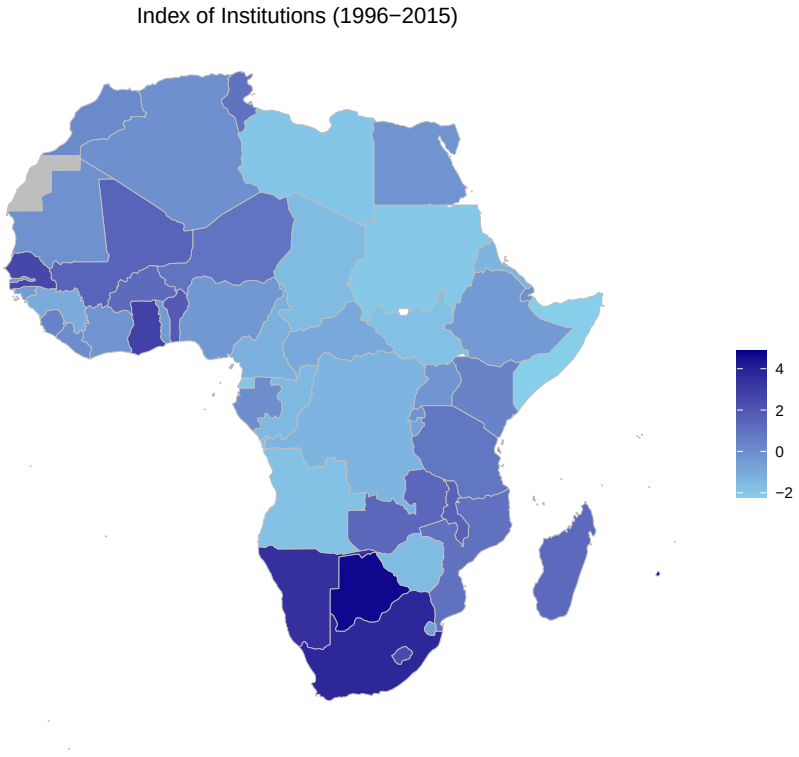
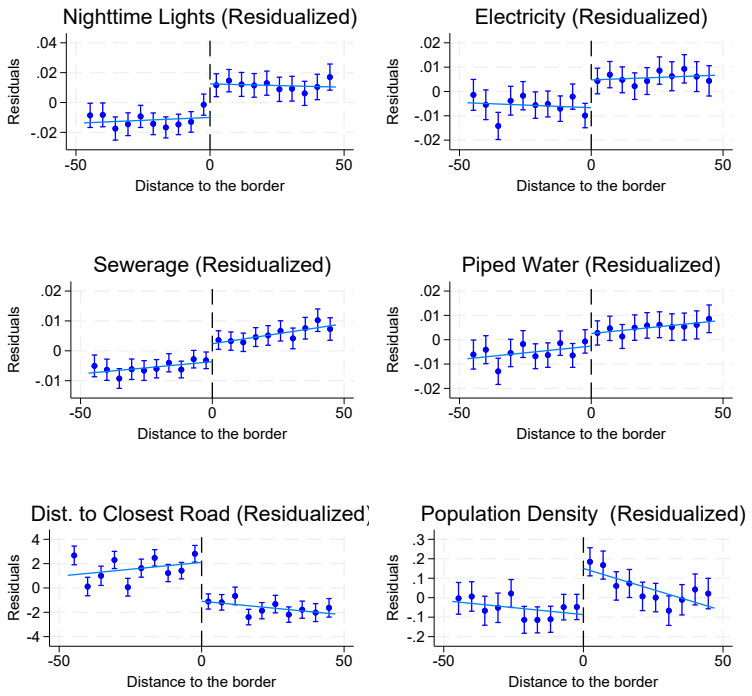
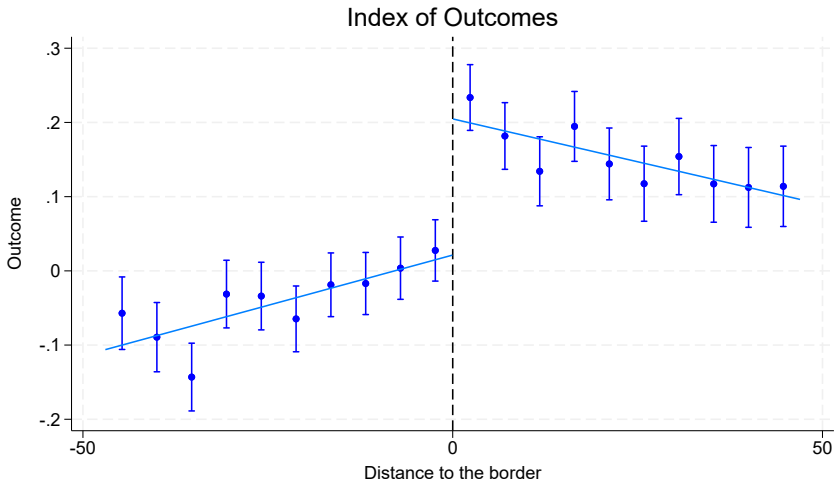


Figure A5. : Discontinuity around national borders, by outcome (residualized)



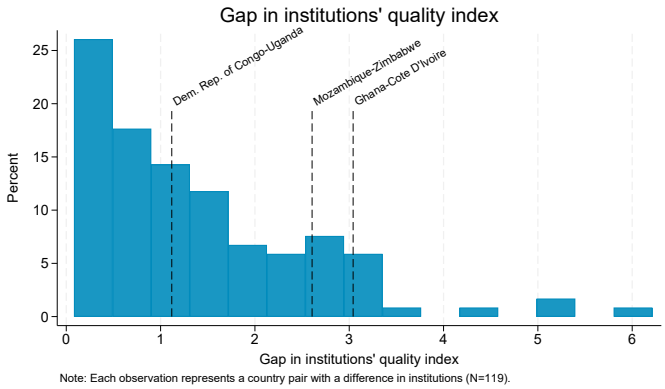
Note: The country with the better institutions is on the right (positive distance to the border).

Figure A6. : Discontinuity around national borders (not residualized)



Note: The country with the better institutions is on the right (positive distance to the border).

Figure A7. : Gaps in index of institutions across national borders



Note: Each observation represents a country pair with a difference in institutions (N=119).

APPENDIX TABLES

Table B1—: Summary Statistics: Patch-Level Outcomes

Variable	Mean	Std. Dev	Min	P25	Median	P75	Max	Obs
Electricity	0.29	0.26	0.00	0.08	0.18	0.45	1.00	746,475
Sewerage	0.14	0.14	0.01	0.07	0.10	0.16	1.00	746,475
Water	0.29	0.22	0.00	0.11	0.23	0.44	0.99	746,475
Nightlights	0.10	1.61	0.00	0.00	0.00	0.00	427.53	746,605
Distance to the closest road	17.05	35.63	0.00	0.00	3.41	18.56	3,625.39	731,145
Population Density	33.19	278.19	0.00	0.00	0.00	7.35	48,962.05	747,158

Table B2—: Institutions and Development: RD results by measure of institutional quality

Outcome	(1) Electricity (predicted)	(2) Sewerage (predicted)	(3) Piped water (predicted)	(4) Nightlights 2015	(5) Distance to closest road	(6) Pop. Density	(7) Index of outcomes
Panel B:							
Rule of Law	0.0073	0.0081**	0.0104**	0.0358***	6.0892**	0.1667	0.0862*
	(0.007)	(0.004)	(0.004)	(0.009)	(2.454)	(0.212)	(0.045)
	{0.30}	{0.05}	{0.02}	{0.00}	{0.02}	{0.44}	{0.06}
Control of Corruption	0.0062	0.0069*	0.0082*	0.0291***	-1.1136	0.0470	0.0997**
	(0.006)	(0.004)	(0.004)	(0.009)	(2.309)	(0.200)	(0.048)
	{0.34}	{0.06}	{0.07}	{0.00}	{0.63}	{0.81}	{0.04}
Polity IV Score	0.0170**	0.0064*	0.0066	0.0323***	-6.2464**	-0.0764	0.1382**
	(0.007)	(0.004)	(0.004)	(0.012)	(2.471)	(0.268)	(0.061)
	{0.02}	{0.09}	{0.14}	{0.01}	{0.01}	{0.78}	{0.03}
Electoral democracy	0.0170**	0.0063	0.0098*	0.0402***	-4.5525*	0.5955***	0.2138***
	(0.008)	(0.004)	(0.005)	(0.012)	(2.678)	(0.175)	(0.051)
	{0.03}	{0.12}	{0.06}	{0.00}	{0.10}	{0.00}	{0.00}
Observations	78,696	78,696	78,696	78,702	78,702	78,702	78,696

Notes: We include ethnicity level fixed effects, standard errors are clustered at the country level and are reported in parentheses, p-values are reported in curly brackets.

Table B3—: Distribution of Infrastructure Measures Across Countries

Outcome	(1) Electricity (predicted)	(2) Sewerage (predicted)	(3) Piped water (predicted)	(4) Nightlights 2015	(5) Distance to closest road	(6) Pop. Density	(7) Index of outcomes
Mean	0.37	0.16	0.38	0.20	56.15	0.02	0.35
Std. Dev.	0.19	0.06	0.15	0.22	284.30	2.94	2.45
Minimum	0.09	0.08	0.12	0.00	0.05	-4.61	-15.64
Percentile 25	0.24	0.12	0.29	0.05	0.91	-2.58	-0.09
Median	0.36	0.14	0.37	0.10	4.00	0.26	0.62
Percentile 75	0.46	0.17	0.41	0.29	15.29	2.24	1.26
Maximum	0.89	0.36	0.84	0.96	2101.24	5.45	3.93
Interquartile Range	0.22	0.05	0.13	0.24	14.37	4.82	1.35
Coefficient/IQR	0.08	0.25	0.08	0.14	-0.34	0.06	0.14

Notes: We report country level summary statistics. Coefficient/IQR ratios are defined relative to coefficients in Panel A of Table 1.

Table B4—: Optimal bandwidth for different polynomial degrees

Polynomial degree	Optimal bandwidth (km)
1	47.1
2	94.0
3	262.6
4	508.5

Table B5—: Accounting for multiple hypothesis testing

Outcome	(1) Electricity (predicted)	(2) Sewerage (predicted)	(3) Piped water (predicted)	(4) Nightlights 2015	(5) Distance to closest road	(6) Pop. Density
Panel A. Index of institutions:						
Coefficient	0.017	0.012	0.010	0.035	-4.844	0.293
Normal p-value	0.000	0.000	0.000	0.000	0.000	0.000
Clustered p-value	0.035	0.015	0.035	0.003	0.083	0.191
Clustered-FDR p-value	0.046	0.040	0.046	0.018	0.056	0.068
Panel B. Rule of Law:						
Coefficient	0.007	0.008	0.010	0.036	6.089	0.167
Normal p-value	0.004	0.000	0.000	0.000	0.000	0.000
Clustered p-value	0.301	0.048	0.021	0.000	0.017	0.437
Clustered-FDR p-value	0.137	0.043	0.036	0.002	0.036	0.171
Panel C. Control of Corruption:						
Coefficient	0.006	0.007	0.008	0.029	-1.114	0.047
Normal p-value	0.017	0.000	0.001	0.000	0.001	0.161
Clustered p-value	0.336	0.056	0.067	0.003	0.632	0.815
Clustered-FDR p-value	0.338	0.125	0.125	0.017	0.611	0.688
Panel D. Polity IV Score:						
Coefficient	0.017	0.006	0.007	0.032	-6.246	-0.076
Normal p-value	0.000	0.000	0.005	0.000	0.000	0.020
Clustered p-value	0.018	0.090	0.143	0.010	0.015	0.777
Clustered-FDR p-value	0.039	0.073	0.094	0.039	0.039	0.208
Panel E. Electoral Democracy:						
Coefficient	0.017	0.006	0.010	0.040	-4.553	0.596
Normal p-value	0.000	0.000	0.000	0.000	0.000	0.000
Clustered p-value	0.034	0.119	0.059	0.002	0.096	0.001
Clustered-FDR p-value	0.048	0.073	0.064	0.005	0.073	0.005

APPENDIX C: TECHNICAL NOTES ON TRAINING DATASET, CNNS, CALIBRATION AND MODEL PERFORMANCE

C1. Training data

We use the Afrobarometer Round 6 survey to identify communities with data on infrastructure access. Afrobarometer is an Africa-wide effort to collect household-level data every 4-5 years. Round 6, which we focus on for training, was collected between 2014 and 2015, and contains information from 36 African countries. Data on infrastructure access is recorded by the enumerator at the Enumeration Area (EA) level, which contains on average 8 households. The dataset we use contains 7,022 enumeration areas, with 150 to 300 EAs per country. We exploit the fact that the Afrobarometer data was geocoded by BenYishay et al. (2017) to match each EA to the corresponding satellite images.[5] We match the centroid of the EA to the centroid of the satellite image.

As Afrobarometer surveys, by construction, only collect data from populated areas, we do not have ground truth for locations where no one lives. To map infrastructure over the entire continent, we however need the model to learn to recognize areas where there are no households. Indeed, if the model did not learn that some places are not suitable to live and therefore, no infrastructure is built there, the risk of false positives could be high. To avoid this problem, we add to the training dataset 696 uninhabited patches from the Sahara Desert with "zero" labels for all infrastructure measures.

C2. Satellite Imagery

We use multispectral imagery from the Landsat 7 and 8 Surface Reflectance Tier 1 Collection and from the VIIRS Stray Light Corrected Nighttime Day/Night Band Composites Version 1. Landstat images are available at a 30 meter resolution, and we use 6 of its bands: 3 RGB bands and 3 infrared bands. We use the quality assessment band to remove pixels that are marked as having cloud, cloud shadow and snow in every image, and then take a median composite of every cloud-free pixel available between 2014 and 2015 for 6.72km by 6.72km patches around the centroid of the ground-truth data points (Afrobarometer enumeration areas and desert patches) described in section **??**.

The VIIRS Nighttime Light (NTL) imagery has resolution of 15 arc seconds, approximately 500 meters per pixel, and it only contains a single band recording the radiance at night. The data product for NTL data throughout this paper has been preprocessed to filter out clouds and correct for stray light. We generate 2014-2015 median composites based on monthly data (because of the possible lack of good quality data coverage for some months). The nightlight band is then reprojected and concatenated to the landsat images as the seventh band.[6] When we are training the model, the input imagery has size 224 pixels by 224 pixels (6.72km by 6.72km) and 7 bands, a dimension of $224 \times 224 \times 7$.

C3. Machine Learning Model

We build on the approach developed in Oshri et al. (2018). The first task consists in solving a multi-label binary classification problem: Using as a input a satellite image X, we want to predict whether a given enumeration area (our observation level) has or not access to infrastructure defined by a set of binary label $Y_1, Y_2, \ldots, Y_k \in \{0, 1\}$. For each infrastructure access label Y_i we train a

[5]The survey was geocoded ex-post, so some EAs do not have accurate coordinates. In particular, 3,053 EAs share geocoordinates with another EA. We drop these enumeration areas to focus on those for which the geolocation is accurate (N=3,969 EAs)

[6]VIIRS Nightlight Light data has different resolution from Landsat so we re-projected it to 30-meters.

separate binary classifier $h_i(X) = \{\hat{Y}_i, \hat{P}_i\}$, where $\hat{Y}_i$ is a class prediction and $\hat{P}_i$ is the confidence associated to that prediction.

To tackle our binary image classification problem, we use deep learning techniques—algorithms that combine a large amount of linear and non-linear functions in a hierarchical way, making the algorithms very flexible, and thus allowing them to learn very complicated functional forms. Convolutional Neural Networks (CNNs) are a class of deep learning algorithms used for image classification tasks, with their main advantages being that they preserve the spatial structure of images, and can learn patterns contained in images with fewer parameters than a regular neural network. The hierarchical structure of CNNs allows the algorithms to extract basic or low-level features from data, and then to aggregate these basic features to learn more complex features. As the algorithms identify features without supervision, researchers no longer need to program complex routines to identify edges or shapes.

We train a Residual Network (ResNet), a state-of-art CNN that has been shown to have good performance for binary classification problems (He et al., 2016). We modify the structure of the initial 18-layer Residual Network in 2 ways: we start by increasing the number of input channels in the first convolutional layer, given that the original model was trained to classify 3 band images on ImageNet and we are using satellite imagery with six bands for the Landsat data and 1 band for the VIIRS Nighttime data. We also modify the final fully connected softmax layer of the model, that was originally meant to classify an image into 1,000 categories, to work with a binary classification problem and predict the probability of an enumeration area having access to an infrastructure variable. We initialize the weights for the RGB bands of the Landstat images using the the ResNet pre-trained weights, and we use Xavier initialization to initialize the weights of remaining Landsat and VIIRS Nighttime bands, given that drawing initial values from the uniform distribution described in Glorot and Bengio (2010) has been shown to lead to faster convergence. We follow the same strategy regarding data processing, regularization and optimization as Oshri et al. (2018).

C4. Calibration

Calibration consists in modifying the "confidence" ($\hat{p}_i$) produced by the model, in order for the distribution of generated confidences to resemble the empirical distribution in the training dataset, without affecting the accuracy of the model. Current deep learning models tend to generate confidence levels that are very close to 0 or 1 for each observation, so they can get a small loss function. As economists or political scientists, however, we care about the *likelihood* that a given area has access to specific amenities—rather than the binary prediction, to the extent that it is correlated with the level of development. In other words, if the model predicts that area A has electricity with probability 55%, area B has electricity with probability 65% and area C with 85%, while all three areas would have a binary prediction of "1", we want to be able to interpret the A-C gap of 30 percentage points as having some cardinal interpretation: area C is"more developed" than area A, and the gap in development between C and A is greater than that between B and A. Calibration helps ensure that the confidence (the predicted likelihood) can be interpreted in that way.

We use a post-estimation calibration technique called "temperature scaling" (Guo et al., 2017), that works in the following way: After estimating the parameters in the Residual Network $\hat{h}_i$, we compute scores $\hat{s}_i$ with the data in our validation set. The scores are the inputs to the final softmax layer of our network that generates the confidence or predicted probability $\hat{p}_i$. We then run a logistic regression of the true classes in our validation set, $\hat{y}_i$, against our score, of the form $\hat{y}_i = \sigma(\frac{\hat{s}_i}{T} + \varepsilon)$, where σ is the logistic function. The newly estimated temperature parameter is the logistic function. The newly estimated temperature parameter $\hat{T}$ is then used to adjust the confidence, so now we get the new calibrated prediction $\tilde{p}_i \equiv \sigma(\frac{\hat{s}_i}{\hat{T}})$

C5. Model Performance

Out of 4,665 observations, we use 80% to train our model and 20% to test its performance, which means that we use 3,738 observations to train our model, and 927 observations to assess the performance of our final model.

Table C2 reports on the performance of the model. For each infrastructure considered, we classify a patch as 0 (no access) if its predicted probability is below 0.5, and as 1 (access) if the predicted probability is above 0.5. In the first column of Table C2, we report the mean of the ground truth outcomes in our test set: Around 55% of patches have access to electricity, 52% to piped water, and 28% to sewerage. The second column of Table C2 displays the accuracy of the model over the test set, where we calculate the number of correct predictions divided by the sample size. We can see that the model predictions are correct between 77% and 84% of the time, depending on the infastructure considered.

Next, we report *recall*—the number of correct positive predictions (when we predict there is access to infrastructure) divided by the number of actual positive observations, and *specificity*—the number of correct negative predictions (when we predict there is no access to infrastructure) divided by the number of actual negative observations. Recall ranges between 68% and 82%, and specificity is between 74% and 91%, which shows that the model performs well at predicting both instances of access and lack of access to infrastructure.[7]. Finally, we report AUCROC scores, which represent the area under the ROC curve. These scores show how well a classifier can distinguish between 2 classes, i.e., predicting 0 when the ground truth is 0 and predicting 1 when the ground truth is 1. This metric takes values ranging between 0 and 1, with 1 representing a perfect classifier. The area under the ROC curve ranges between 77% and 83% depending on the infrastructure considered.

Table C1 further analyses the performance of the models separately for patches in rural vs. urban areas. Urban enumeration areas are more likely to have access to electricity, piped water, and sewerage than rural enumeration areas. It is therefore important to make sure the model performance is not simply driven by the model's ability to separate urban from rural areas. Reassuringly, Table C1 shows that the model performs well in both rural and urban areas. As expected, recall is slightly higher in urban areas, and specificity is higher for rural areas: the model is better at identifying when there is access to infrastructure in urban areas, and at identifying when rural areas lack access to basic infrastructure.

Our model's performance is on par with similar models in this literature: Oshri et al. (2018) 's model presents accuracy levels between 67.3% and 83.2% when predicting our same three infrastructure outcome, achieving the same accuracy when predicting access to electricity, but lower accuracy when predicting access to sewerage and piped water.

C6. External validation with other datasets

We test whether our predicted measures are valid using the ground truth from new datasets with precise geolocations. We export satellite images at the corresponding locations with the same procedures as before, feed the images into our model, and generate predictions. Then, we compare our predictions with the ground truth data to check whether the model performs well in other settings. We also compare the performance of our model in these datasets against a "naive model" using nightlights only. The NL model will predict that an EA has access to electricity, piped water, and sewerage if there is any lit patch, and no access to all infrastructure of the area is completely dark.

[7]Breaking down the performance of by ground truth class (0 or 1) is important because of the class imbalance in some outcomes like sewerage. For instance, we could predict that no household has access to sewerage, and the accuracy of that prediction would be 72.4%. By looking at recall and specificity we can see that the high accuracy levels come from good predictions for both classes, and it is not being affected by class imbalances.

Table C1—: Model Performance for Infrastructure Outcomes: Urban vs. Rural

	Mean	Accuracy	Recall	Specificity	AUCROC
Electricity					
Urban	0.934	0.845	0.866	0.545	0.706
Rural	0.328	0.824	0.753	0.859	0.806
Piped water					
Urban	0.863	0.818	0.865	0.522	0.693
Rural	0.324	0.745	0.667	0.782	0.725
Sewerage					
Urban	0.642	0.713	0.721	0.700	0.710
Rural	0.069	0.922	0.488	0.955	0.721

Note: All performance metrics were computed in our test sample.

Table C2—: Model Performance for Infrastructure Outcomes

	Mean	Accuracy	Recall	Specificity	AUCROC
Electricity	0.547	0.832	0.822	0.843	0.833
Piped Water	0.519	0.771	0.786	0.756	0.771
Sewerage	0.276	0.847	0.684	0.909	0.796

Note: All performance metrics were computed in our test sample.

Our validation datasets are divided into two categories: datasets from the same time period and a different period. As the model is built to make predictions from 2014-2015 imagery, it is possible that it is particularly accurate in predicting infrastructure outcomes in the same period. We use four contemporaneous datasets at various levels (household, community) to test for within-period validity. We also test how well the model performs in earlier or later data. Specifically, for earlier periods, we use Afrobarometer Round 5, collected in 2011-2013, matched with Landast images from that same period. For later periods, we use the second round of the Greater Addis Ababa survey collected by Stanford's AUDRI initiative in 2018, paired with imagery from 2018. Details on the datasets are in Table C3.

We present the validation results in Table C4. We see that proposed model always significantly outperforms the NL model in predicting piped water and sewerage access. In term of predicting electricity, our model outperforms almost always, especially in rural regions. Our model performs overall better than the nighttime lights benchmark when we validate against the Afrobarometer Round 5. In particular, it does substantially better for patches in rural areas and near national borders, though it performs worse in urban areas. In the validation dataset of high schools in Ghana, nighttime lights perform slightly better when predicting access to electricity, but our model performs considerably better when predicting access to piped water. In Addis Ababa surveys, the naive nighttime lights model performs slightly better when predicting electricity only in Addis Ababa, but it performs significantly worse in piped water and sewerage. In Abidjan and rural Malawi, our model outperforms in every outcome.

Table C3—: Datasets for the validation exercises

Survey	Year	Landsat Period	Nightlight Period	Exporting Method
Training Dataset				
Afrobarometer R6 (multi-country)	2014-2015	2014-2015	2015-2016	centroid
Validation datasets: Same Time Period, Different Levels				
Household Access				
Neno & Mwanza, Malawi (rural)	2015	2014-2015	2015-2016	patch level
Greater Abidjan, Côte d'Ivoire (urban)	2014	2014-2015	2015-2016	patch level
Greater Addis Ababa, Ethiopia (urban)	2016	2014-2015	2015-2016	patch level
Institutional Access				
Ghana schools (country-wide)	2015	2014-2015	2015-2016	centroid
Validation datasets: Different Time Periods				
Afrobarometer R5 (multi-country)	2011-2013	2011-2013	2012-2014	centroid
Greater Addis Ababa, Ethiopia (urban)	2018	2018	2018	patch level

Table C4—: Validation Results

Dataset	**Benchmark-NL**			**Model - R6**		
	Electricity	Piped Water	Sewerage	Electricity	Piped Water	Sewerage
Afrobarometer R5						
All	0.59	0.56	0.28	0.74	0.66	0.79
Urban	0.92	0.85	0.61	0.84	0.76	0.66
Rural	0.40	0.40	0.09	0.69	0.59	0.86
Border	0.42	0.48	0.15	0.69	0.60	0.83
Ghana						
All	0.89	0.50		0.88	0.58	
Urban	0.97	0.58		0.94	0.62	
Rural	0.83	0.44		0.82	0.55	
Border	0.89	0.28		0.85	0.34	
Other surveys						
Addis - W1	0.84	0.65	0.22	0.71	0.52	0.71
Addis - W2	0.75	0.47		0.67	0.58	
Abidjan	0.80	0.54	0.05	0.75	0.58	0.76
Malawi (Rural)	0.56	0.56	0.53	0.95	0.95	1.00

Note: On the first 3 columns, we display the accuracy on different validation datasets of a "naive" model based in nighttime lights: we predict that there is access to infrastructure if nighttime lights are positive. The next 3 columns show the accuracy of our convolutional neural network model trained with Afrobarometer R6 groundtruth and satellite imagery.

APPENDIX D: DATASET BUILDING

D1. Grid of patches, national borders, valid patches

To build our dataset, we start by generating a grid of 6.72 by 6.72 km patches over all the African continent. We built a map of the continent of Africa by aggregating country shapefiles obtained from GADM (GADM, 2018) and GISCO (Eurostat, 2021), and then we divided it in 747,165 patches using the SF package for R (Pebesma, 2018).

We add to our grid the shapefiles containing the borders of the ethnical ethnicities of Africa described in Murdock (1959), and we also add individual shapefiles for every African country from GADM. We intersect both maps to identify ethnicities that are partitioned by national borders.

Not every patch of land is useful for our analysis. First, we exclude patches that are in the border of 2 ethnicities or 2 countries. We also use a water cover raster from the MOD44W MODIS dataset (Carroll et al., 2017) and a forest cover raster from the Dynamic Land Cover map from (Buchhorn et al., 2020). Patches marked as part of a forest or a water body are not considered for our analysis, since they will mechanically have no access to infrastructure, and this is caused by geographical factor rather than by national or pre-colonial institutions.

When studying ethnicities that are partitioned by national borders, it might be the case that more than one country can be present in a partitioned ethnicity (for instance, the territory of the Malinke ethnicity falls within 6 countries), and it there might also be discontinuities in the distribution of the distance to the border within certain ethnicities caused by ethnicities and countries that intersect in 2 separate areas, or parts of the territory of a ethnicity being mostly covered by a forest or a water body, so in order to overcome this, we define a set of criteria to assign patches in each ethnicity to the main border within the ethnicity and compute the distance to the border to that border. To do this, we use the following iterative procedure:

1) We identify all the feasible borders within a ethnicity.
2) We define the main border as the border separating 2 adjacent countries that combined have the biggest share of the area of the ethnicity.
3) For each patch we compute the distance to the main border.
4) We check for discontinuities in the distribution of distance to the border. If there is a gap of more than 10 km, all the patches outside the gap are no longer considered as belonging to our previous main border, but they can be considered as a part of a different border.
5) If one of the sides of the ethnicity has less than 10 patches, or if after removing patches due to a discontinuity we have another border covering a bigger area, we replace our main border with the new biggest border and go back to step 3, and we repeat this until there are no changes and we have a stable main border.

D2. Outcomes and controls

We intersect our grid with several rasters of data, and added other data sources. The final dataset includes the following variables:

- **National Institutions Quality Measures**: We use 4 measures.
  - **Rule of Law** from the Worldwide Governance Indicators (Kaufmann, Kraay and Mastruzzi, 2011). Rule of Law captures perceptions of the extent to which agents have confidence in and abide by the rules of society, and in particular the quality of contract enforcement, property rights, the police, and the courts, as well as the likelihood of crime and violence.

  - **Control of Corruption** from the Worldwide Governance Indicators (Kaufmann, Kraay and Mastruzzi, 2011). Control of Corruption captures perceptions of the extent to which public power is exercised for private gain, including both petty and grand forms of corruption, as well as "capture" of the state by elites and private interests.
  - **Polity IV** score from the Center for Systemic Peace (Marshall, Gurr and Jaggers, 2019). The Polity score of a country captures this regime authority spectrum on a 21-point scale ranging from -10 (hereditary monarchy) to +10 (consolidated democracy).
  - **The Electoral Democracy index** from Varieties of Democracy Dataset version 11 (Coppedge et al., 2021). The Electoral Democracy index captures to what extent the ideal of electoral democracy in its fullest sense is achieved, with politicians that are responsive to citizens, competitive and fair elections, and freedom of expression and press.

- **Malaria temperature suitability**: We obtained data for a temperature suitability for P. Falciparum and P. Vivax transmission for 2010 from Gething et al. (2011). Their data is at a resolution of 1 km by 1 km, so for each of our patches we take the average value of all the pixels that fall inside them using Google Earth Engine (Gorelick et al., 2017).

- **Land suitability**: We obtained a measure of land suitability for agriculture from The Atlas of the Biosphere (Ramankutty et al., 2002). This dataset has information about the fraction of each cell that is suitable to be used for agriculture. The original raster file is at a resolution of 0.5 degrees (approximately 55 kilometers around the Equator), so we intersect our patches with this raster, and take the average suitability for the cases where our patches lie within more than one cells of the original raster.

- **Elevation**: We use elevation data from the Shuttle Radar Topography Mission from NASA (Farr et al., 2007) with a 30 meters resolution. For each of our patches of land we take the average value of the raster with elevation using Google Earth Engine.

- **Distance to the equator**: To control for the distance to the equator, we take the absolute value of the latitude of each patch.

- **Distance to the capital**: We compute our distance to the capital for each country by retrieving the geographical coordinates of their capital using the R package CoordinateCleaner (Zizka et al., 2019), and then we compute the distance to the capital from each patch using the SF package (Pebesma, 2018).

- **Distance to the coast**: To define our distance to the coast, we combine the GADM maps of all the African countries to obtain an outline of the African continent, and then we compute the distance from each patch to the coast using the SF package.

- **Distance to petroleum sources**: We identify the location of petroleum sources with the PETRODATA v1.2 dataset from the Peace Research Institute Oslo (Lujala, Ketil Rod and Thieme, 2007). For each patch, we again use the SF package to find the closest petroleum sources inside and outside of the country the patch belongs to, and then we compute the distance to these 2 petroleum sources.

- **Distance to diamond sources**: We identify the location of diamond sources by using the DIADATA v1a dataset from the Peace Research Institute Oslo (Gilmore et al., 2005). For each patch, we use the SF package to identify the closest diamond sources inside and outside of the country the patch belongs to, and then we compute the distance to these 2 diamond sources.

- **Distance to the closest road**: We use the GRIP global roads database to compute the distance to the closest road for each patch of land in our grid (Meijer et al., 2018). This dataset contains maps for all the type of roads (highways, primary, secondary, tertiary and local roads) for all the continents, so we intersect this with country maps for Africa, and for each patch we compute the distance to the closest road within their respective country using the SF package in R.

- **Nightlights**: We use VIIRS Nighttime Light data (Elvidge et al., 2017). Specifically, we use annual composites for the year 2015 of the version 1, where they preprocess imagery to remove stray lights and outliers, average the annual data using only cloud-free luminosity measurements, and sets the background lights to 0. These images have a resolution of 15 arc seconds (around 500m at the Equator), so we use Google Earth Engine to compute the average luminosity for each of our patches of land.

- **Population density**:We use the Data For Good at Meta's High Resolution Population Density Maps (Tiecke et al., 2017), a product based on the Gridded Population of the World v4 dataset (Center for International Earth Science Information Network, CIESIN, 2016) that uses computer vision algorithms to re-distribute population density within territories according to the volume of the buildings in said territory. These predictions have a 30 meter resolution, so for each patch of land in our grid we take the average value of the pixels in the population density raster using Google Earth Engine.

APPENDIX E: ROBUSTNESS CHECKS

Table E1 shows a battery of robustness checks. We focus on the the specification using the PCA index of institutions. Panel A reproduces the results from Table 1 for benchmarking. Panel B shows the results excluding controls. The results are unchanged. Panel C focuses on borders with large discontinuities (defined as a difference in institutional quality across the two sides of the border above the median). Unsurprisingly, the coefficients increase, consistent with a "dose response" effect of institutional quality. Panel D focuses on the subsample of borders with the same colonizer on both sides.[8] The sample size shrinks, but most of the results stand. Finally, Panel E shows that the results mostly hold when we focus on patches far away from the capital of their respective countries.[9]

Table E1—: Alternative RD specifications

Outcome	(1) Electricity (predicted)	(2) Sewerage (predicted)	(3) Piped water (predicted)	(4) Nightlights 2015	(5) Distance to closest road	(6) Pop. Density	(7) Index of outcomes
Panel A. Main specification:							
Index of institutions (PCA)	0.0173**	0.0117**	0.0104**	0.0346***	-4.8445*	0.2934	0.1957***
	(0.008)	(0.005)	(0.005)	(0.011)	(2.734)	(0.221)	(0.053)
	{0.04}	{0.02}	{0.04}	{0.00}	{0.08}	{0.19}	{0.00}
Observations	78,696	78,696	78,696	78,702	78,702	78,702	78,696
Panel B. No controls:							
Index of institutions (PCA)	0.0111**	0.0065**	0.0076**	0.0276***	-2.4324	0.5583**	0.1653***
	(0.005)	(0.003)	(0.004)	(0.010)	(2.721)	(0.225)	(0.050)
	{0.03}	{0.02}	{0.04}	{0.01}	{0.38}	{0.02}	{0.00}
Observations	78,696	78,696	78,696	78,702	78,702	78,702	78,696
Panel C. Large discontinuities:							
Index of institutions (PCA)	0.0124	0.0174**	0.0109	0.0583***	-2.6512	0.2410	0.2169**
	(0.012)	(0.007)	(0.008)	(0.017)	(2.073)	(0.215)	(0.081)
	{0.32}	{0.02}	{0.20}	{0.00}	{0.21}	{0.27}	{0.01}
Observations	41,449	41,449	41,449	41,455	41,455	41,455	41,449
Panel D. Ethnicities with the same colonizer:							
Index of institutions (PCA)	0.0329***	0.0182**	0.0079	0.0468**	-0.4940	-0.5941***	0.1357**
	(0.011)	(0.008)	(0.009)	(0.020)	(3.967)	(0.213)	(0.065)
	{0.00}	{0.03}	{0.36}	{0.02}	{0.90}	{0.01}	{0.04}
Observations	28,490	28,490	28,490	28,496	28,496	28,496	28,490
Panel E. Neither patch within 300 km of the capital:							
Index of institutions (PCA)	0.0109	0.0072**	0.0092	0.0284***	-7.7320**	0.2646	0.1748***
	(0.007)	(0.004)	(0.006)	(0.010)	(3.445)	(0.251)	(0.056)
	{0.10}	{0.05}	{0.13}	{0.01}	{0.03}	{0.30}	{0.00}
Observations	43,036	43,036	43,036	43,042	43,042	43,042	43,036

Notes: We include ethnicity level fixed effects, standard errors are clustered at the country level and are reported in parentheses, p-values are reported in curly brackets.

Figure E1 checks whether the results are affected by the RD specification choices. Panel A shows that the estimated impact of institutional quality on outcomes is unaffected by the polynomial degree choice. Panel B shows that the effect, if anything, increases with bandwidth size. Finally,

[8] In appendix F we present information on the colonizer both (a) following the "scramble for Africa" and (b) at the time of independence, and present results using one or the other.

[9] A straightforward way to define patches away from capital cities would be to just exclude patches within a certain radius of their respective capitals. A potential problem with this is that it could lead to ethnicities with a very unbalanced number of patches between the two sides of the border, if only one side is close to the capital of a country. To solve this, for each ethnicity we assign each patch a "match" on the other side of the border (its symmetric), so each patch has a unique counterpart on the other side of the border that is at a very similar distance from the border. To match them, we use the Hungarian Algorithm. We then focus on the sample of paired patches and estimate the previous specification only for pairs of patches where neither patch is within 300 km of the capital of their respective country.

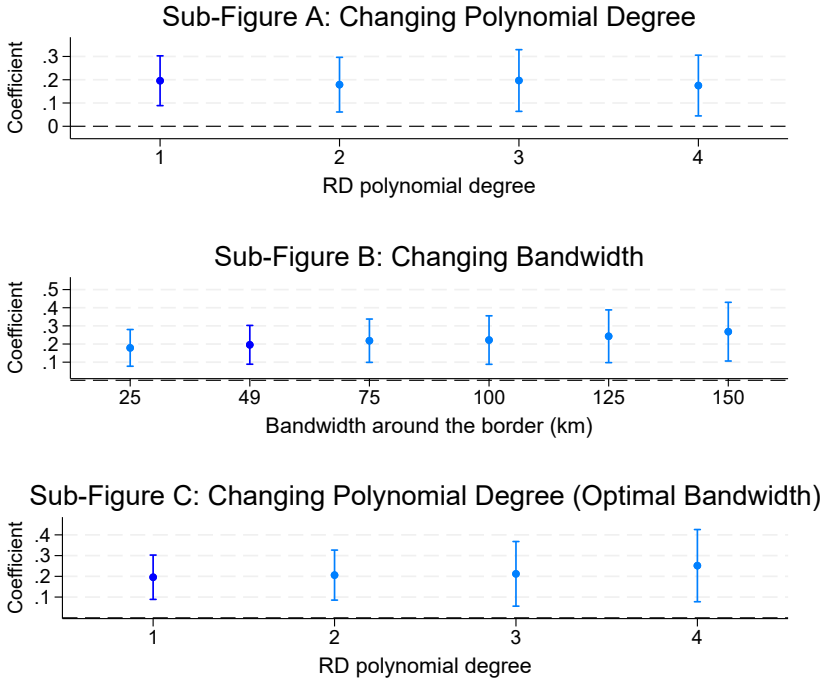
Table E2—: Endogenous Borders? Results for straight vs. non-straight line borders

Outcome	(1) Electricity (predicted)	(2) Sewerage (predicted)	(3) Piped water (predicted)	(4) Nightlights 2015	(5) Distance to closest road	(6) Pop. Density	(7) Index of outcomes
Panel A. Main specification:							
Index of institutions (PCA)	0.0173**	0.0117**	0.0104**	0.0346***	-4.8445*	0.2934	0.1957***
	(0.008)	(0.005)	(0.005)	(0.011)	(2.734)	(0.221)	(0.053)
	{0.04}	{0.02}	{0.04}	{0.00}	{0.08}	{0.19}	{0.00}
Observations	78,696	78,696	78,696	78,702	78,702	78,702	78,696
Panel B. Straight line (Primary feature):							
Index of institutions (PCA)	0.0368***	0.0167***	0.0381***	0.0580**	-15.1458***	0.5520**	0.4296***
	(0.011)	(0.003)	(0.008)	(0.023)	(4.452)	(0.219)	(0.082)
	{0.00}	{0.00}	{0.00}	{0.02}	{0.00}	{0.02}	{0.00}
Observations	29,278	29,278	29,278	29,278	29,278	29,278	29,278
Panel C. Non-straight line (Primary feature):							
Index of institutions (PCA)	0.0167*	0.0079*	0.0082	0.0306**	-5.1083	0.1645	0.1612***
	(0.009)	(0.005)	(0.006)	(0.014)	(3.125)	(0.309)	(0.055)
	{0.06}	{0.10}	{0.15}	{0.04}	{0.11}	{0.60}	{0.01}
Observations	44,644	44,644	44,644	44,650	44,650	44,650	44,644
Panel D. Straight line (Any feature):							
Index of institutions (PCA)	0.0078	0.0049	0.0088	0.0164	-6.8856*	0.2219	0.1361**
	(0.006)	(0.003)	(0.006)	(0.011)	(3.583)	(0.251)	(0.058)
	{0.24}	{0.15}	{0.12}	{0.14}	{0.06}	{0.38}	{0.02}
Observations	57,215	57,215	57,215	57,221	57,221	57,221	57,215
Panel E. Non-straight line (Any feature):							
Index of institutions (PCA)	0.0328**	0.0162**	0.0194	0.0784**	-2.2534	0.5216	0.3213***
	(0.012)	(0.007)	(0.012)	(0.031)	(2.514)	(0.421)	(0.106)
	{0.01}	{0.02}	{0.13}	{0.02}	{0.38}	{0.23}	{0.01}
Observations	16,707	16,707	16,707	16,707	16,707	16,707	16,707

Notes: We include ethnicity level fixed effects, standard errors are clustered at the country level and are reported in parentheses, p-values are reported in curly brackets. Data on border straightness from Paine, Qiu and Ricart-Huguet (2025).

in Panel C we vary the polynomial degree and re-compute the optimal bandwidth (our optimal bandwidth is increasing in the polynomial degree, as reported in Table B4). The magnitude of the coefficient is very stable.[10]

Figure E1. : Robustness of results to RD polynomial degree and bandwidth



Another possible concern is that we are considering six outcomes, so at a 5% significance level we have a 26% probability ($1-0.95^6$) of a false rejection for each one of our institutional measures. In Table B5 we show that adjusting the p-values to account for multiple hypotheses testing does not alter the key finding that institutional quality matters for infrastructure access. Specifically, for each institutional variable we compute the sharpened False Discovery Rate p-values described in Anderson (2008), using the main RD specification with fist degree RD polynomials, the optimal bandwidth, controls, fixed effects and clustered standard errors. For each specification shown in Table 1, we report the regression coefficients, p-values computed without clustering, clustered p-values, and sharpened False Discovery Rate clustered p-values.

[10]These robustness checks are shown for each outcome separately in Figure E2, Figure E3 and Figure E4.

Figure E2. : National Institutions: Robustness to polynomial degree choice

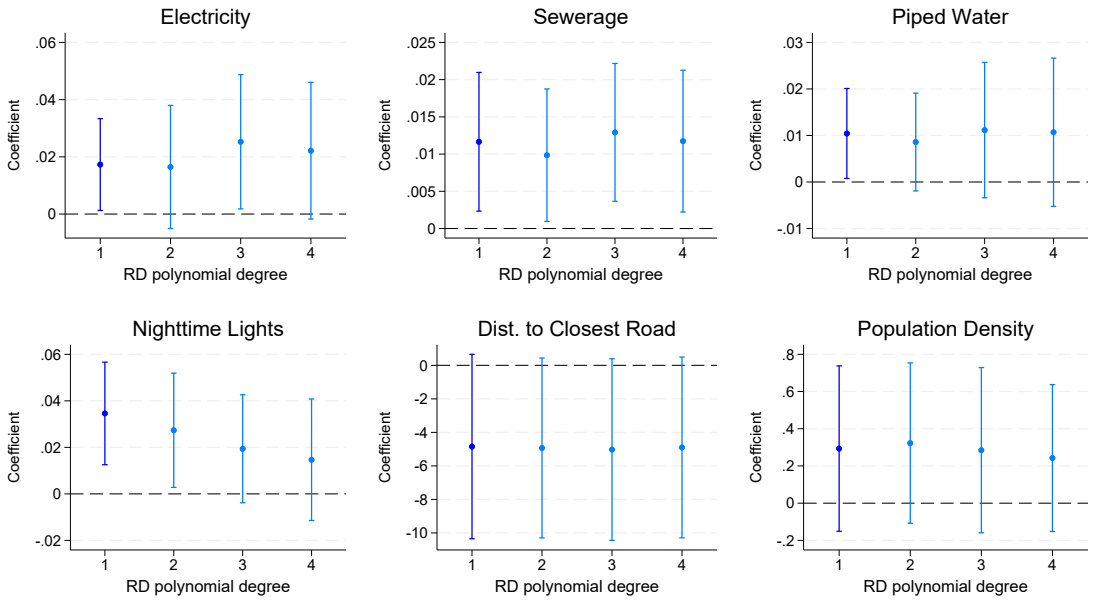


Figure E3. : National Institutions: Robustness to bandwidth choice

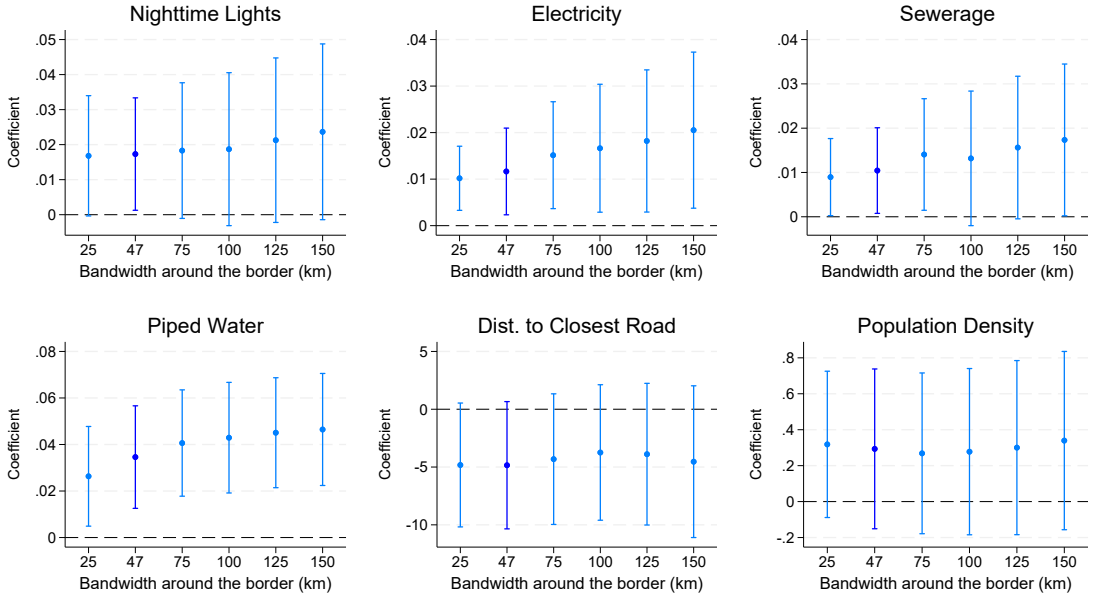
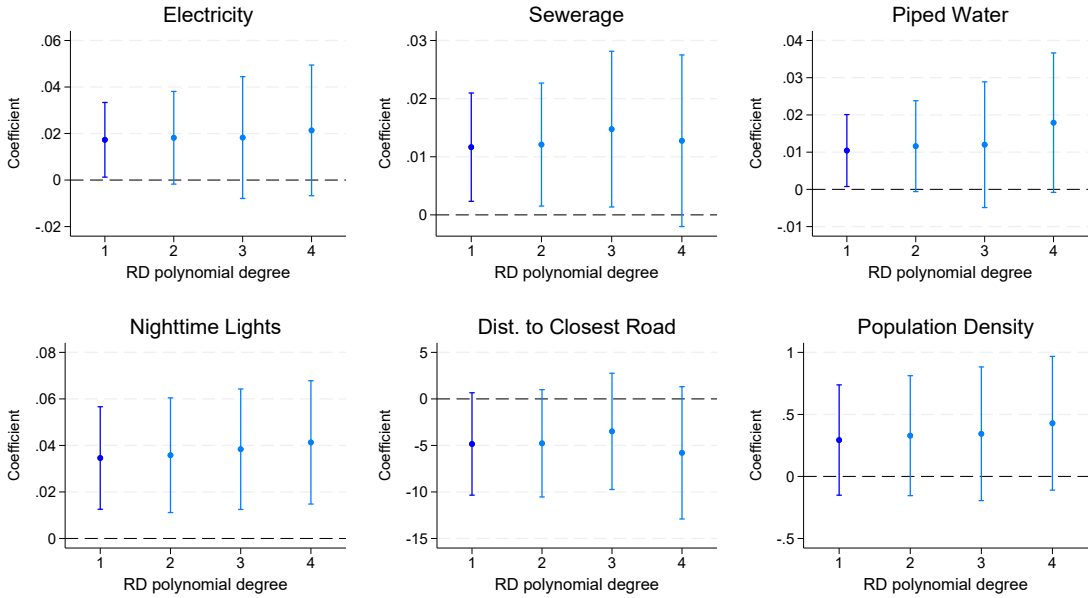


Figure E4. : National Institutions: Robustness to polynomial degree with optimal bandwidth choice



APPENDIX F: COLONIZERS AND RULERS AT INDEPENDENCE

The identification strategy in our first application relies on working with ethnicities that were partitioned by national borders, so that we can compare territories that have common ethnic origins and should not be too different apart from the different national institutions they are exposed to. Despite this, we could also think that European colonizers shaped the early institutions of modern African countries, so ethnicities that were separated into countries with different colonizers could have underlying differences that can be confounded for differences in national institutions.

To overcome this issue we produced a list of colonizers for each African country, and the we use this to re-estimate our results using a sample of ethnicities that were partitioned into countries with the same colonizers, or adding colonizer fixed effects. After World War I the allies re-distributed the control of some African colonies, so as an extra robustness check we also compiled a list of rulers at independence for most African countries, because rulers in the late colonial days might have been more important than early colonizers. In table F1 we present a list of African countries, their colonizers and their rulers at the time of their independence. For some countries we couldn't establish a single colonizer or ruler at independence: for instance, Somalia was controlled by Great Britain and Italy, so we cannot assign it to a single colonizer or ruler at independence. We code these colonizer and ruler into 3 categories:

Table F1—: Colonizers and rulers at independence in Africa

Country	Colonizer	Ruler at independence	Country	Colonizer	Ruler at independence
Algeria	France	France	Madagascar	France	France
Angola	Portugal	Portugal	Malawi	Great Britain	Great Britain
Benin	France	France	Mali	France	France
Botswana	Great Britain	Great Britain	Mauritania	France	France
Burkina Faso	France	France	Mauritius	France	Great Britain
Burundi	Germany	Belgium	Mozambique	Portugal	Portugal
Côte d'Ivoire	France	France	Namibia	Germany	Germany
Cameroon	Germany	-	Niger	France	France
Cape Verde	Portugal	Portugal	Nigeria	-	Great Britain
Central African Republic	France	France	Republic of Congo	France	France
Chad	France	France	Rwanda	Germany	Belgium
Comoros	France	France	São Tomé and Príncipe	Portugal	Portugal
Democratic Republic of the Congo	Belgium	Belgium	Saint Helena	Great Britain	Great Britain
Djibouti	France	France	Senegal	France	France
Egypt	Great Britain	Great Britain	Seychelles	Great Britain	Great Britain
Equatorial Guinea	Spain	Spain	Sierra Leone	Great Britain	Great Britain
Eritrea	Italy	Great Britain	South Africa	Great Britain	Great Britain
Ethiopia	Italy	Italy	South Sudan	Great Britain	Great Britain
Gabon	France	France	Sudan	Great Britain	Great Britain
Gambia	-	Great Britain	Swaziland	Great Britain	Great Britain
Ghana	-	Great Britain	Tanzania	Germany	Great Britain
Guinea	France	France	Togo	Germany	-
Guinea-Bissau	Portugal	Portugal	Tunisia	France	France
Kenya	-	Great Britain	Uganda	Great Britain	Great Britain
Lesotho	Great Britain	Great Britain	Zambia	Great Britain	Great Britain
Liberia	United States	United States	Zimbabwe	Great Britain	Great Britain
Libya	Italy	-			

With these definitions we replicate our main results restricting our sample to ethnicities with the same colonizers or the same rule at independence, or using our full sample with colonizer or ruler at independence fixed effects. We only include fixed effects for France and Great Britain, since all other colonizers and rulers combined controlled less than 25% of the territory in Africa. In Table F2 we present the results of these experiments: our results remain fairly similar to the main results presented in Panel A of Table 1, with the coefficients for Nighttime lights, Electricity, Sewerage, Piped ater and our PCA index being significant in most specifications, and presenting coefficients with similar magnitudes to the ones we observed in our main results.

Table F2—: Effects of colonizers/rulers at independence

Outcome	Index	Nightlights	Electricity	Sewerage	Piped water	Dist. Road	Pop. Density
Panel A. Ethnicities with the same colonizer:							
Index of institutions (PCA)	0.0329***	0.0182**	0.0079	0.0468**	-0.4940	-0.5941***	0.1357**
	(0.011)	(0.008)	(0.009)	(0.020)	(3.967)	(0.213)	(0.065)
	{0.00}	{0.03}	{0.36}	{0.02}	{0.90}	{0.01}	{0.04}
Observations	28,490	28,490	28,490	28,496	28,496	28,496	28,490
Panel B. Colonizer fixed effects:							
Index of institutions (PCA)	0.0185**	0.0122***	0.0118**	0.0387***	-5.4584**	0.2627	0.2087***
	(0.008)	(0.005)	(0.005)	(0.010)	(2.571)	(0.211)	(0.049)
	{0.02}	{0.01}	{0.01}	{0.00}	{0.04}	{0.22}	{0.00}
Observations	78,696	78,696	78,696	78,702	78,702	78,702	78,696
Panel C. Ethnicities with the same ruler at independence:							
Index of institutions (PCA)	0.0318***	0.0172**	0.0016	0.0420**	3.1163	-0.4978**	0.0977
	(0.010)	(0.008)	(0.009)	(0.019)	(3.822)	(0.217)	(0.066)
	{0.00}	{0.04}	{0.86}	{0.04}	{0.42}	{0.03}	{0.15}
Observations	29,524	29,524	29,524	29,530	29,530	29,530	29,524
Panel D. Ruler at independence fixed effects:							
Index of institutions (PCA)	0.0176**	0.0119**	0.0109**	0.0369***	-4.7494*	0.2953	0.2008***
	(0.008)	(0.005)	(0.005)	(0.010)	(2.723)	(0.212)	(0.052)
	{0.04}	{0.01}	{0.03}	{0.00}	{0.09}	{0.17}	{0.00}
Observations	78,696	78,696	78,696	78,702	78,702	78,702	78,696

Notes: We include ethnicity level fixed effects, standard errors are clustered at the country level and are reported in parentheses, p-values are reported in curly braces.

APPENDIX G. APPLICATION II: DISTRIBUTIVE POLITICS

Distributive politics, or how politicians and policy makers allocate public goods and services, is a major determinant of welfare, especially when it comes to infrastructure, because connecting an area to the electrical grid, the sewage system, improving access to piped water, or building roads to improve their connectivity can also lead to further opportunities for economic development. Starting with Bates (1974), ethnicity has been the focus of analyses of distributive allocations in Sub-Saharan Africa: see for example Burgess et al. (2015) and more recently Dreher et al. (2019). This literature typically focuses on one type of good in one country. But, as shown by Kramon and Posner (2013), there are many goods and services allocated simultaneously by governments, and if one group receives more roads, another may receive more health care. Golden and Min (2013) thus call for studies that consider multiple goods before clear patterns of political favoritism can be established.

We use our dataset to study the extent to which the allocation of infrastructure across geographical units within a country is consistent with ethnic favoritism, a phenomenon where politicians prioritize the geographies where their ethnicity resides when allocating public resources. We compute average infrastructure access at the second-level administrative unit (the next smallest unit after a state or region), and merge the dataset with information on the ethnicity of presidents since independence (or 1960 for countries that were never colonized)[11]. Despite the non-causal nature of our results, our dataset can be used to produce evidence the effects of ethnic favoritism over all African countries, and over the allocation of multiple types of infrastructure.

To do this, we estimate the following equation:

$$\text{Dev}_{g,c} = \alpha + \beta\, \text{Ties}_{g,c} + X'_{g,c}\Psi + \delta_c + \varepsilon_{g,c} \tag{G1}$$

Where $\text{Dev}_{g,c}$ represents the average development outcome at second-level administrative unit g in country c, $\text{Ties}_{g,c}$ is an indicator of geographical unit g containing the birthplace of any president ever elected in country c (a proxy for their ethnic homeland), which means one of these presidents has political ties to unit g, $X'_{g,c}$ is a vector of geographic controls, and δ_c are country fixed effects. Standard errors are clustered at the country level.

In Panel A of Table G1 reports the result of estimating Equation (G1), where we find a strong positive correlation between political ties and infrastructure access across all infrastructure measures considered. In Panel B we control for areas being far from the capital and we interact it with our political tie variable. Our findings suggest areas political ties affect more the allocation of infrastructure in areas closer to national capitals compared to our baseline, geographies that are close to the capital but have no political ties. Areas that are far from the capital and have no political ties have less access to infrastructure, but areas that are far from the capital that have ties to a president don't experience reduced access to infrastructure (their net effect is not statistically significant for any outcome). This means that political ties prevent geographies from suffering the negative effects of being far away from capital cities.

Studying sub-national allocation of roads in Kenya, Burgess et al. (2015) find that democracy constrains ethnic favoritism: expenditures on roads are twice larger in districts that share the ethnicity of the president during non-democratic periods, but there is no difference during periods of democracy. Building on this, Panel A of Table G2 examines whether parochialism is mitigated by good institutions. In other words, we ask: does having higher quality institutions keep presidents in check? For this, we add an interaction term between political tie and the country-level index of institutional quality described in section III. The estimates in column imply thatone standard deviation increase in institutional quality reduces the role of political ties in the allocation of in-

[11] We use the dataset built by Dreher et al. (2019) covering the period between 2000 to 2012, and add hand-collected data between the years 1960 to 2000.

Table G1—: Allocation of basic infrastructure

Outcome	(1) **Electricity (predicted)**	(2) **Sewerage (predicted)**	(3) **Piped water (predicted)**	(4) **Nightlights 2015**	(5) **Distance to closest road**	(6) **Index of outcomes**
Panel A:						
Political tie	0.0990***	0.1209***	0.1081***	0.8428***	-0.4414**	1.1637***
	(0.021)	(0.032)	(0.024)	(0.267)	(0.210)	(0.238)
	{0.00}	{0.00}	{0.00}	{0.00}	{0.04}	{0.00}
Outcome mean	0.46	0.23	0.44	0.48	1.35	1.88
Observations	4,633	4,633	4,633	4,633	4,633	4,633
$\boldsymbol{R^2}$	0.53	0.64	0.48	0.59	0.27	0.58
Panel B:						
Political tie	0.0992***	0.1500***	0.1226***	1.0315***	-0.4522*	1.2912***
	(0.027)	(0.034)	(0.029)	(0.278)	(0.263)	(0.264)
	{0.00}	{0.00}	{0.00}	{0.00}	{0.09}	{0.00}
Far from capital	-0.0473**	-0.0283**	-0.0370**	-0.1127	0.0623	-0.4074**
	(0.018)	(0.011)	(0.017)	(0.080)	(0.367)	(0.159)
	{0.01}	{0.02}	{0.04}	{0.16}	{0.87}	{0.01}
Political tie × Far from capital	-0.0288	-0.1133***	-0.0701*	-0.6915**	0.0731	-0.6657**
	(0.038)	(0.033)	(0.041)	(0.273)	(0.501)	(0.322)
	{0.46}	{0.00}	{0.09}	{0.01}	{0.88}	{0.04}
Outcome mean	0.46	0.23	0.44	0.48	1.35	1.88
Observations	4,633	4,633	4,633	4,633	4,633	4,633
$\boldsymbol{R^2}$	0.54	0.66	0.50	0.61	0.27	0.60
p-value of sum of coefficients	0.38	0.56	0.48	0.20	0.43	0.28

Notes: The unit of observation is a subnational geographic unit. The sample contains administrative level-2 (ADM2) units from 39 countries and administrative level-1 units from 7 countries for which ADM2 was not available. We include country-level fixed effects, standard errors are clustered at the country level and are reported in parentheses, and p-values are reported in curly brackets.

frastructure by about 20%. These mitigating effects range between 9.8% for access to roads to 27% for electricity.

Table G2—: Political Ties, Institutional Quality and International Aid

Outcome	(1) **Electricity (predicted)**	(2) **Sewerage (predicted)**	(3) **Piped water (predicted)**	(4) **Nightlights 2015**	(5) **Distance to closest road**	(6) **Index of outcomes**
Panel A: Institutional quality						
Political tie	0.1144***	0.1483***	0.1230***	1.0350***	-0.4005	1.3746***
	(0.026)	(0.035)	(0.028)	(0.298)	(0.240)	(0.272)
	{0.00}	{0.00}	{0.00}	{0.00}	{0.10}	{0.00}
Political tie × Index of institutions	-0.0309**	-0.0298	-0.0215*	-0.1381	0.0396	-0.3053**
	(0.014)	(0.018)	(0.012)	(0.130)	(0.126)	(0.144)
	{0.03}	{0.10}	{0.09}	{0.30}	{0.75}	{0.04}
Outcome mean	0.46	0.23	0.44	0.48	1.35	1.88
Observations	4,633	4,633	4,633	4,633	4,633	4,633
R^2	0.53	0.65	0.48	0.60	0.27	0.58
Panel B: International aid						
Political tie	0.1066	0.1317*	0.0904	0.5253	-0.6154	1.2054*
	(0.072)	(0.075)	(0.056)	(0.328)	(0.436)	(0.659)
	{0.14}	{0.09}	{0.11}	{0.12}	{0.16}	{0.07}
Political tie × High total aid	-0.0066	0.0055	0.0308	0.5890	0.3012	0.0553
	(0.073)	(0.086)	(0.064)	(0.506)	(0.511)	(0.717)
	{0.93}	{0.95}	{0.63}	{0.25}	{0.56}	{0.94}
Outcome mean	0.46	0.23	0.44	0.48	1.35	1.88
Observations	4,633	4,633	4,633	4,633	4,633	4,633
R^2	0.52	0.65	0.48	0.61	0.27	0.58
p-value of sum of coefficients	0.00	0.00	0.00	0.00	0.23	0.00

Notes: The unit of observation is a subnational geographic unit. The sample contains administrative level-2 (ADM2) units from 39 countries and administrative level-1 units from 7 countries for which ADM2 was not available. We include country-level fixed effects, and standard errors are clustered at the country level and are reported in parentheses. We control for distance to the capital, distance to the coast, closest diamond and petroleum sources inside and outside the country, temperature suitability for Malaria, land suitability for agriculture, elevation, distance from the equator, and area of the administrative unit. ***, **, and * indicate statistical significance at 1%, 5%, and 10% significance levels, respectively.

Our last analysis, shown in Panel B of Table G2, considers the role of foreign aid in the allocation of infrastructure within countries. Countries that are receiving aid from foreign organizations could be under more scrutiny, and their leaders might have a harder time allocating infrastructure to geographic areas with political ties to them. To test this hypothesis, we start by splitting countries into two groups, those receiving aid above the median, and those who receive aid below the median in terms of total aid received between 2000 and 2014, according to the OECD's Creditor Reporting System (CRS), to then interact our proxy of ethnic favoritism with an indicator variable of receiving high total international aid. We find not only that areas with political ties have on average better access to infrastructure, but that international aid further exacerbates this effect for all outcomes, except access to roads. We report that the p-value of the sum of both coefficients is 0 for almost all variables, so the effect of political connections for countries that receive high aid is positive and statistically significant.

SUPPLEMENTAL REFERENCES

Anderson, Michael L. 2008. "Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects." *Journal of the American statistical Association*, 103(484): 1481–1495.

Bates, Robert H. 1974. "Ethnic competition and modernization in contemporary Africa." *Comparative political studies*, 6(4): 457–484.

BenYishay, A., R. Rotberg, J. Wells, Z. Lv, S. Goodman, L. Kovacevic, and D. Runfola. 2017. "Geocoding Afrobarometer Rounds 1-6: Methodology & Data Quality."

Buchhorn, Marcel, Myroslava Lesiv, Nandin-Erdene Tsendbazar, Martin Herold, Luc Bertels, and Bruno Smets. 2020. "Copernicus global land cover layers—collection 2." *Remote Sensing*, 12(6): 1044.

Burgess, Robin, Remi Jedwab, Edward Miguel, Ameet Morjaria, and Gerard Padró i Miquel. 2015. "The value of democracy: evidence from road building in Kenya." *American Economic Review*, 105(6): 1817–51.

Carroll, ML, CM DiMiceli, MR Wooten, AB Hubbard, RA Sohlberg, and JRG Townshend. 2017. "MOD44W MODIS/Terra Land Water Mask Derived from MODIS and SRTM L3 Global 250m SIN Grid V006 [Data set]." *NASA EOSDIS Land Processes DAAC*.

Center for International Earth Science Information Network, CIESIN. 2016. "Gridded Population of the World, Version 4 (GPWv4): Population Density." Accessed August 9, 2021.

Coppedge, Michael, John Gerring, Carl Henrik Knutsen, Staffan I. Lindberg, Jan Teorell, Nazifa Alizada, David Altman, Michael Bernhard, Agnes Cornell, M. Steven Fish, Lisa Gastaldi, Haakon Gjerlow, Adam Glynn, Allen Hicken, Garry Hindle, Nina Ilchenko, Joshua Krusell, Anna Luhrmann, Seraphine F. Maerz, Kyle L. Marquardt, Kelly McMann, Valeriya Mechkova, Juraj Medzihorsky, Pamela Paxton, Daniel Pemstein, Josefine Pernes, Johannes von Romer, Brigitte Seim, Rachel Sigman, Svend-Erik Skaaning, Jeffrey Staton, Aksel Sundstrom, Ei tan Tzelgov, Yi ting Wang, Tore Wig, Steven Wilson, and Daniel Ziblatt. 2021. "V-Dem Country-Year/Country-Date Dataset v11.1."

Dreher, Axel, Andreas Fuchs, Roland Hodler, Bradley C Parks, Paul A Raschky, and Michael J Tierney. 2019. "African leaders and the geography of China's foreign assistance." *Journal of Development Economics*, 140: 44–71.

Elvidge, Christopher D, Kimberly Baugh, Mikhail Zhizhin, Feng Chi Hsu, and Tilottama Ghosh. 2017. "VIIRS night-time lights." *International journal of remote sensing*, 38(21): 5860–5879.

Eurostat. 2021. "GISCO: Geographic Information System of the COmmission." Accessed August 9, 2021, `https://ec.europa.eu/eurostat/web/gisco`.

Farr, Tom G, Paul A Rosen, Edward Caro, Robert Crippen, Riley Duren, Scott Hensley, Michael Kobrick, Mimi Paller, Ernesto Rodriguez, Ladislav Roth, et al. 2007. "The shuttle radar topography mission." *Reviews of geophysics*, 45(2).

GADM. 2018. "Database of Global Administrative Areas." Accessed August 9, 2021, `https://gadm.org/data.html`.

Gething, Peter W, Thomas P Van Boeckel, David L Smith, Carlos A Guerra, Anand P Patil, Robert W Snow, and Simon I Hay. 2011. "Modelling the global constraints of temperature on transmission of Plasmodium falciparum and P. vivax." *Parasites & vectors*, 4(1): 1–11.

Gilmore, Elisabeth, Nils Petter Gleditsch, Päivi Lujala, and Jan Ketil Rod. 2005. "Conflict diamonds: A new dataset." *Conflict Management and Peace Science*, 22(3): 257–272.

Glorot, Xavier, and Yoshua Bengio. 2010. "Understanding the difficulty of training deep feedforward neural networks." 249–256, JMLR Workshop and Conference Proceedings.

Golden, Miriam, and Brian Min. 2013. "Distributive politics around the world." *Annual Review of Political Science*, 16(1): 73–99.

Gorelick, Noel, Matt Hancher, Mike Dixon, Simon Ilyushchenko, David Thau, and Rebecca Moore. 2017. "Google Earth Engine: Planetary-scale geospatial analysis for everyone." *Remote Sensing of Environment*.

Guo, Chuan, Geoff Pleiss, Yu Sun, and Kilian Q Weinberger. 2017. "On calibration of modern neural networks." 1321–1330, PMLR.

He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. "Deep residual learning for image recognition." 770–778.

Kaufmann, Daniel, Aart Kraay, and Massimo Mastruzzi. 2011. "The worldwide governance indicators: methodology and analytical issues1." *Hague journal on the rule of law*, 3(2): 220–246.

Kramon, Eric, and Daniel N Posner. 2013. "Who benefits from distributive politics? How the outcome one studies affects the answer one gets." *Perspectives on Politics*, 11(2): 461–474.

Lujala, Päivi, Jan Ketil Rod, and Nadja Thieme. 2007. "Fighting over oil: Introducing a new dataset." *Conflict Management and Peace Science*, 24(3): 239–256.

Marshall, Monty G, Ted Robert Gurr, and Keith Jaggers. 2019. "Polity IV project: Political regime characteristics and transitions, 1800–2018." *Center for Systemic Peace*.

Meijer, Johan R, Mark AJ Huijbregts, Kees CGJ Schotten, and Aafke M Schipper. 2018. "Global patterns of current and future road infrastructure." *Environmental Research Letters*, 13(6): 064006.

Murdock, George Peter. 1959. "Africa its peoples and their culture history."

Oshri, Barak, Annie Hu, Peter Adelson, Xiao Chen, Pascaline Dupas, Jeremy Weinstein, Marshall Burke, David Lobell, and Stefano Ermon. 2018. "Infrastructure quality assessment in africa using satellite imagery and deep learning." 616–625.

Paine, Jack, Xiaoyan Qiu, and Joan Ricart-Huguet. 2025. "Endogenous Colonial Borders: Pre-colonial States and Geography in the Partition of Africa." *American Political Science Review*, 119(1): 1–20.

Pebesma, Edzer. 2018. "Simple Features for R: Standardized Support for Spatial Vector Data." *The R Journal*, 10(1): 439–446.

Ramankutty, Navin, Jonathan A Foley, John Norman, and Kevin McSweeney. 2002. "The global distribution of cultivable lands: current patterns and sensitivity to possible climate change." *Global Ecology and biogeography*, 11(5): 377–392.

Tiecke, Tobias G, Xianming Liu, Amy Zhang, Andreas Gros, Nan Li, Gregory Yetman, Talip Kilic, Siobhan Murray, Brian Blankespoor, Espen B Prydz, et al. 2017. "Mapping the world population one building at a time." *arXiv preprint arXiv:1712.05839*.

Zizka, Alexander, Daniele Silvestro, Tobias Andermann, Josue Azevedo, Camila Duarte Ritter, and Daniel Edler. 2019. "Package 'CoordinateCleaner'."