university of groningen / faculty of spatial sciences

Association of informality and sprawl in rapidly growing African cities

Master's Thesis

submitted in partial fulfillment of the requirements for the degree

M.Sc. in Economic Geography

submitted by

Tamilwai Johannes Kolowa

Student Number: S4960858

supervised by

Dr. Michiel Daams

Groningen, August 2022

Abstract

Cities in Africa have been described as drivers of economic development but also as disconnected and sprawling, thus constraining their potential to reap agglomeration benefits. One of the urban expansion phenomena associated with shortcomings of urban form is informal urban growth. To find empirical evidence for the question, whether unplanned areas contribute more to sprawl than planned areas in growing African cities, this study will focus on the street accessibility dimension of sprawl. Connectivity metrics of urban locations will be used to compare unplanned and planned areas in Tanzanian secondary cities. A linear regression model with spatially clustered standard errors is used to trace the isolated effect of planning type on compactness. Results suggest that this effect is not homogenous across cities. Two types of cities are identified, where for one the effect of planning type is found to be significant. Further research on informal urban expansion in developing countries might be encouraged to take the methodology and results at hand into account when addressing the perceived shortcomings of informal urban morphologies.

Keywords: urban expansion, sprawl, informality, secondary cities, Tanzania, spatial regression

Contents

List of Tables

List of Figures

List of Abbreviations

AoI	Area of interest
CBD	Central Business District
CCI	Composite compactness index
DV	Dependent variable
EO4SD	Earth-Observation for Development
ESA	European Space Agency
GHG	Green-House-Gas
GHSL	Global Human Settlement Layer
GIS	Geographic Information System
GUI	Graphical User Interface
Obs.	Observations
OLS	Ordinary-least-squares
OSM	OpenStreetMap
Std. dev.	Standard deviation
SE	Standard Error
SSA	Sub-Saharan Africa
UN	United Nations
UNFPA	United Nations Population Fund
UNAT	Urban Network Analysis Toolkit
WBG	World Bank Group
WRI	World Ressources Institute
WSF-Evo	World Settlement Footprint Evolution

1. *Introduction*

Cities in Africa have been described simultaneously as a driver of economic development but also as disconnected and featuring the highest relative living costs in the world, thus constraining their potential. Economic Geography associates these trends with fragmented urban forms that lead to higher urban costs and inefficiencies (Lall et al. 2017). One of the urban expansion phenomena associated with inefficiency of urban form is *urban sprawl*. Sprawl is associated with decreasing accessibility and connectivity (Lee 2020).

Meanwhile, informal or "unplanned" urban expansion is commonly considered a hazard to sustainable urban development in developing countries both in policy (UNFPA 2007; Msuya et al. 2020) and research (Cobbinah and Amoako 2014; Hervé Tchékoté and Chrétien Ngouanet 2015; Jarah et al. 2019; Tchekoté and Ngouanet 2015), and not uncommonly labelled "sprawl" by default (Mosammam et al. 2017).

On the other hand, the limited empirical evidence on the association between informality and sprawling urban expansion is ambiguous or points in another direction. Comparing urban growth between 1990 and 2000 in 120 cities around the globe, Sheppard (2010) found that informal housing does not contribute more to urban expansion than formal housing. Expansion is here understood as the additional consumption of land per capita. Rather, urban expansion following planned areas is seen as causing increasing consumption of land. In Shepperd's analysis, a doubling in the share of informal housing in a city was associated with a 21% decrease in urban land use.

Mahendra et al. (2021) stress that haphazard, unplanned expansion is facilitated through both formally and informally constructed buildings alike, raising doubt to whether informal housing leads to "unplanned" areas more than formal housing does. At the same time, Rahman et al. (2021) find that unplanned, older areas are less connected and compact, although here the time-period of construction might be the main determinant of urban form.

From this juxtaposition of the developmental push for compact development and the consideration of informality as a potential hinderance to compactness, arises the following research question:

> *Are unplanned areas more than planned areas associated with sprawl in rapidly expanding secondary cities in Sub-Saharan Africa?*

Therefore, one goal of this study is to present further evidence about the quality of sprawl in unplanned development in developing countries, to potentially reevaluate the focus on informal sprawl in the literature on urban expansion. It is also worth mentioning that Egidi et al. (2020) ask for more context-specific typologies of sprawl to be developed for describing urban growth in different local context appropriately. The study at hand aims to contribute to this refinement.

To contribute empirical evidence for the question at hand, this study will focus on the accessibility dimension of sprawl, which is commonly measured in street network connectivity (Lowry and Lowry 2014). It will use street network connectivity metrics, namely the "reach" score of locations in the city following Sevtsuk and Mekonnen (2012), to compare unplanned and planned areas in seven Tanzanian secondary cities. Network connectivity is operationalized with cumulative opportunities that can be "reached" from a given point within a distance threshold (Bhat et al. 2000). In this study, the reachable building area in m^2 (reach) will proxy for these opportunities, assuming that the more building area any person can reach, the more connected to opportunities is the street network in general, thus the more compact the neighborhood.

Reachable building area has also been considered a good measure of urban form's ability to be conducive for economic development, specifically for retail choices (Sevtsuk 2010). In the local context of unplanned areas, reach may be considered a measure of potential for future development, even if economic activity at present is low. In addition, reach can be considered a good metric for the local context, as it can distinguish areas of low-built up density (sprawl) from high built-up density (compact) without being biased by population density (cf. Lall et al. 2021).

There are several dimensions of sprawl debated in the literature, which cannot all be addressed in this study. For example, Ewing et al. (2003) have proposed to quantitatively measure population density, land use mix, degree of centering, and street accessibility in the US-American context. One reason to exclusively choose street network centrality in this study is limited data on the urban built-up area in Sub-Saharan African (SSA) cities: data availability of detailed land-uses, service distribution, or places that function as urban centers are hardly available at scale. This is true in particular for secondary cities in the SSA region. In contrast, road network data and building data have increasingly become available through open-source, remote-sensing data and contributions from non-expert OpenStreetMap-participants, and is used for network analysis (Boeing 2017; Brandily and Rauch 2018). Participatory mapping has the great advantage that informal areas, which might not have appeared on official maps, are now mapped with accuracies of up to 73%, even though data quality ranges (Yeboah et al. 2021).

In the context of this study, lower levels of connectivity would mean that informality leads to an inherently less connected urban form that is unfit for future densification, is not conducive to agglomeration economies, and already provides less opportunities for residents by definition of the urban form. Manifested in urban forms, these conditions would be harder to counter with public intervention. As Tanzanian secondary cities are relatively monocentric (Huang et al. 2018), unplanned areas might also be less fit to develop polycentricity in the future, if their current development has been scattered.

Such finding would imply an additional policy imperative to monitor unplanned urban growth. Unplanned urban growth could then with more confidence be described as undesirable "sprawl". This would imply it had to be steered better, e.g. through sites-and-services development, a public intervention in greenfield development which has been neglected for decades (Gattoni 2009), but is gaining traction again (World Bank 2022a).

However, if unplanned and planned areas have the same levels of connectivity, or if the relationship is even reversed, this would mean that informality should not be the focus of attention when discussing urban form intervention in SSA cities. Without a doubt, informality impacts many planning outcomes that affect the life of residents (e.g., urban services, transit) (Msuya et al. 2020). However, these would not have to be engrained into close to irreversible urban morphologies (Duque et al. 2019) but could be feasibly addressed by public intervention.

This study suggests a novel quantitative approach of combining a remote-sensing-based informality classification from the Earth-Observation for Development (EO4SD) project, with openly accessible road and buildings data to investigate this association.

The study will focus on secondary cities in Tanzania, as exemplary for secondary cities' growth in developing countries. Despite their rapid growth and augmenting economic function in urban Tanzania, urban growth in secondary cities has been considered under-researched (Zhang et al. 2020). In Tanzania, 1/3 of the urban population lived in Dar es Salaam in 2021 (UN 2018b). This also means that 2/3 of the population lived in secondary cities. Secondary cities feature distinct urbanization, growth and migration dynamics thanks to their relation to often dominating prime cities in rapidly urbanizing developing countries (Cities Alliance 2022). Huang et al. (2018) selected secondary cities in Tanzania to investigate if urban planning had the ability to effectively steer their growth.

Tanzanian cities have been described as featuring mono-centric urban growth patterns, which persisted even as cities have grown faster (Huang et al. 2018), in contrast to the global phenomenon that large cities tend to become less monocentric over time (Bertaud 2004). For unplanned settlements with low connectivity, this would mean that they are even further away from any relevant center. On the other hand, unplanned settlements with high connectivity, might provide good conditions (in terms of network lay-out) to develop polycentricity later. One interpretation critical to this study could be that high connectivity at the neighborhood level might not be very meaningful, if not connected to any economically relevant center.

Results of the analysis suggest that there is not a homogenous effect of planning type across secondary cities in Tanzania. In four of the seven cities examined, planning type had a pronounced effect on connectivity (Kigoma, Mtwara, Mwanza, Tanga), while in three cities, no significant effect could be found (Arusha, Dodoma, and Mbeya). Both these uncovered effects provide fertile grounds for discussion.

The study at hand is structured as follows: Chapter 2 will outline the debate about informality and sustainable urban development, making definitory distinctions between sprawl, urban expansion, and connectivity in different local contexts. Further, the conceptual framework for linking network connectivity to sprawl is laid out. Lastly, important other factors in urban form development to be included in the analysis are theorized. Chapter 3 discusses the methodological approach to measuring connectivity and finding associations with informality. Chapter 4 introduces descriptive and regression results, which both suggest splitting cities in the sample in two distinct types to arrive at more conclusive results. In chapter 5, results are then discussed vis-à-vis their capacity to answer the research question, as well as to inform the debate on informality and sprawl. Chapter 6 offers concluding remarks.

2. *Literature Review and Theory*

Literature Review

One of the key challenges of research on compactness and sprawl is its definition. In the context of this study, compact urban development is understood in the sense of global development institutions such as the World Bank (WB) and the World Resources Institute (WRI), which favors high connectivity, vertical development and "sustainable densities" (see below). This is in opposition to "sprawl", which could be defined "in the most basic and objective way possible, as low-density, scattered, urban development without systematic large-scale or regional public land-use planning" (Bruegmann 2006, p. 18).

Global trends of urban expansion

From a global perspective it seems that compactness has decreased, or rather sprawl has increased: this general trend over the last decades, measured in decreasingly connected street grids, has been observed by Barrington-Leigh and Millard-Ball (2020).

Furthermore, through remote-sensing analysis, Schneider and Woodcock (2008) suggest that that there are four types of urban growth globally: "low-growth cities with modest rates of infilling; high-growth cities with rapid, fragmented development; expansive-growth cities with extensive dispersion at low population densities; and frantic-growth cities with extraordinary land conversion rates at high population densities" (Schneider and Woodcock 2008, p. 659). Distinguishing further, they conclude that none of the analyzed non-US cities show symptoms of the dispersed, fragmented growth patterns associated with US-American sprawl.

Determinants of sprawl or compactness

What causes sprawl has been up for debate. In the American context, Burchfield et al. (2006) identified a range of contextual causes of sprawl, such as slow population growth, uncertainty of population growth, surrounding topography, the presence of aquifers that developers can tap into, and others. Next to these environmental causes, their findings also point to some that are planning-related: cities sprawled more where there are unincorporated areas in the urban fringe, and municipal tax shares are low.

In the context of SSA, researchers have rather modelled urban growth processes to determine where and how urban growth occurs (Linard et al. 2013). For Tanzania, these models have also made a distinction of informal or unplanned growth. However, they mostly focus on Dar es Salaam (Augustijn-Beckers et al. 2011; Abebe 2011; Msuya et al. 2020). The EO4SD-project of the European Space Agency has analyzed how new settlements in Tanzanian secondary cities tend to be unplanned (Huang et al. 2018), yet this does not answer the question of whether they are associated with sprawl more than planned areas.

Informal urban expansion as a problem for sustainable development

In the above-mentioned discussions, unplanned urban growth is often described an obstacle for sustainable development. Studies investigating this, often focus on major and capital cities. For Dar es Salaam, Msuya et al. (2020) plainly describe a population shift from the central business district (CBD) to the center of the city, facilitated by improved transport infrastructure and rural-to-urban migration but see neighborhood sustainability threatened by limited government intervention. Similarly for the Cameroonian capital Yaoundé, Tchekoté and Ngouanet (2015) see the unplanned urban expansion as a threat to sustainable planning, considering non-building zones, or disputed easement rights as persistent conflicts that are often regulated in post by eviction. For Kumasi, Ghana, Cobbinah and Amoako (2014) lament that communities at the urban fringe get consumed through sprawl which has weakened effectiveness of urban management. While these effects of under-managed urban expansion most likely affect the lives of urban residents, the respective literature fails to address whether unplanned expansion affects urban form negatively, for instance producing and urban form outcome that could only be reversed at major costs.

Compact cities as a development trajectory for cities in developing countries

Addressing the challenge of unsustainable urban expansion is considered vital to finding avenues of achieving urban forms that support their (future) residents, and one proposal is to support compact urban development where possible (Lall et al. 2021). In the global debate on urban development policies, urban growth patterns have become a key issue for formulating policies to manage the global trend of urbanization. For instance, the United Nations put managing urban growth at the heart of sustainable urbanization globally when they state:

> *"Urban growth is closely related to the three dimensions of sustainable development: economic, social and environmental. Well-managed urbanization, informed by an understanding of population trends over the long run, can help to maximize the benefits of agglomeration while minimizing environmental degradation and other potential adverse impacts of a growing number of city dwellers." (UN 2018a, p. 1)*

In another example, both the World Bank and the WRI have issued research reports solely dedicated at incentivizing dense, compact urban development (Lall et al. 2021; Mahendra et al. 2021). WRI argues that unplanned, sprawling development will make it harder for cities to deliver on urban services, and exacerbate environmental degradation with potential impacts lasting for generations. The World Bank views promoting verticalization of urban growth trends in developing countries as a way to achieve "sustainable densities", i.e. for accommodating and concentrating an increasing urban population, while still reaping benefits of agglomeration economies (Lall et al. 2021, p. 6).

Evidence in favor of compact urban development in developing countries

Academic research also provides evidence in favor of compact development in rapidly urbanizing contexts. For the developed country context, there is a large body of literature in favor of compact development (Glaeser 2011; Speck 2013), while some authors question the gains for society from compact urban planning (Breheny 1996).

Limiting sprawl in the developed countries is viewed as a means of inter alia conserving energy, Green House Gas (GHG)-emissions, and natural and public resources (Floater et al. 2014). In developing countries, limiting sprawl is also discussed as a remedy against creating an increasing "urban service divide", where governments cannot provide urban services for fragmented development patters (Mahendra et al. 2021, p. 44).

Compact cities as a desirable outcome of growth for cities in developing countries have been discussed in academia as well (Jenks and Burgess 2000), yet as discussed, empirical approaches often focus on urban management challenges arising from rapid urban growth, rather than on the form that is constructed through this growth. They then tend to attribute the absence of planning, i.e. informality of urban expansion to these cities' growing pains (Jarah et al. 2019; Tchekoté and Ngouanet 2015).

With an innovative methodology, Harari (2020) shows that compact cities facilitate faster population growth and residents associate economic value with compactness. Building on Harari's model, Duque et al. (2019) show a significant association of compact urban form and urban productivity (proxied with night-time-lights) for Latin American cities

In connection with road networks in African cities, Brandily and Rauch (2018) show that road density influences population growth. Lower road density in centers is associated with slower population growth, therefore constraining the potential of agglomeration economies in cities.

Network connectivity and unplanned neighborhoods

Rahman et al. (2021) point out that connectivity measures at the neighborhood scale have been explored in the developed country context, but not in the developing one. This leaves a considerable research gap given the importance of understanding current urban form and trends of urban growth for future development. To fill this gap, the authors develop a composite compactness index (CCI) comprised of population density, evenness of development, clustering nature of development, land-use diversity, floor use mix, and road network connectivity. They apply this to eight neighborhoods in Dhaka, Bangladesh. Road network connectivity in this study is measured as percentage of cul-de-sac in the neighborhood (Rahman et al. 2021).

For road connectivity, Rahman et al. (2021) find that older traditional and unplanned areas in Dhaka feature more cul-de-sac, hence worse road connectivity, than recently constructed and planned neighborhoods. However, this metric is only looking at cul-de-sac which is a very different measure from the reach-metric proposed in this study.

For accessibility to urban services such as health facilities, schools, and urban parks, Huang et al. (2018) find that unplanned areas in Tanzanian cities scored lower than planned areas, in the seven cities investigated also in this study, with only the exception of Arusha.

With a similar research objective, Zhang et al. (2020) used a mixed-method approach to investigate the role of urban form in informal settlements for achieving sustainable development (in the sense of service provision, housing quality, and connectivity) in three Tanzanian cities (Dar es Salaam, Mwanza, and Kigoma, with one neighborhood for each city). The authors blame the informal, uncoordinated development of these neighborhoods for disordered, single land-uses, irregular road networks in poor condition, and low environmental conservation. An important finding here is that for informal neighborhoods buildings tended to be established first, with roads following the development.

For the measurement of urban form, Zhang et al. (2020) consider building metrics, such as size and shape, and road connectivity, i.e. circuitry, complexity, connectivity, and density of the road network. The authors conclude that informal neighborhoods have building densities too high and encourage policy to controlling them to improve living standards for residents. The assumption that high building densities are undesirable urban form is contrary to the assumption proposed in my study. They also conclude that road networks in the informal neighborhoods in Tanzania are not well connected, stating that:

> *"Roads in informal settlements are short and narrow and have few regular sections. Accessibility is poor, with a high proportions [sic!] of dead ends. Road networks are often irregular with changeable lengths, widths, and pavement types. Since unplanned roads are mixed amongst disorderly distributed buildings, blocks have no clear textural features, instead having a semi-natural, loosely distributed, and branch-shaped pattern. This has not been conducive to the accumulation of economic activities and the effective provisioning of infrastructure and social services." (Zhang et al. 2020, pp. 16–17)*

A critical shortcoming of the otherwise methodologically creative study by Zhang et al. (2020) is that it does not compare the urban form metrics, such as connectivity, between unplanned and planned areas. In contrast, it compares three informal settlements, to infer about characteristics of informal settlements in Tanzanian cities overall.

Conceptual framework

This study is building on the above approaches conceptually but wants to provide a direct comparison for unplanned and planned neighborhoods in Tanzanian secondary cities. Further, the normative assumption is that compact urban form is desirable, expressed in metrics such as high population and built-up densities. Network analysis is proposed to operationalize this compactness in an appropriate way for the local context.

Network Analysis

The analysis of networks has a rich tradition in regional planning and economic geography (Porta et al. 2006). With increasingly available network data for all global regions, researchers have started to apply these network analysis tools to investigate questions of urban form where it had not been applied before (Boeing 2017; Brandily and Rauch 2018). The network centrality metrics applied in this study originate from the "Multiple Centrality Assessment" approach first formulated by Porta et al. (2006), who argue that its representation of street networks as nodes and edges close to the original map centerlines were a better fit for urban analytics than the Space Syntax approach (Hillier 1996). Fleischmann et al. (2021) are criticizing that all of these approaches are relying too heavily on the street networks to describe urban form, and do not include other determinants such as buildings and land-uses.

Taking note of this debate, for the purpose of this study, the use of urban street network connectivity, including weights for building sizes, is considered an improvement over a simple consideration of population density (see below).

Precisely, connectivity is operationalized with cumulative opportunities that can be reached from a given point within a distance threshold (Bhat et al. 2000). In this study, the reachable building area will proxy for these opportunities, assuming that the more building area any person can reach, the more connected to opportunities is the street network in general.

Sevtsuk and Mekonnen (2012) proposed several measures to capture the cumulative opportunities in a network and developed the Urban Network Analysis Toolkit (UNAT) to compute these values for points in a city. This study will focus on the "reach" measure of this toolkit, which measures building surface area reachable at a specific distance on a network. Following the literature on walkable cities and transit-oriented development, 800m was chosen as the distance considered well-connected on foot for this research (Lang et al. 2020).

So, what does an increase in reachable building area ("reach" in the following) mean for a resident in the study area? Conceptually, this can be viewed as an 800m-service area being filled with more building surface accessible to a resident. Arguably, this is a quite crude way to capture connectivity and will not compare to advanced GIS-based methods for capturing service accessibility (cf.Yang et al. 2006). In contrast, I will consider this simple example: the median building size in the studied sample is approximately 35m^2. This means that a difference in 1,000m^2 of reach means that any resident is closely connected to 28 more or fewer median-sized houses, which could offer opportunities to him or her. For the sake of another example, the buildings in the 90th percentile in the sample are approximately 135m^2 in size. Buildings this size could be viewed as potentially fit for service delivery, such as small administrative or health related buildings. Again the 1,000m^2 difference in reach would equate to seven buildings more or less to be reached with an 800m walk from the resident in question.

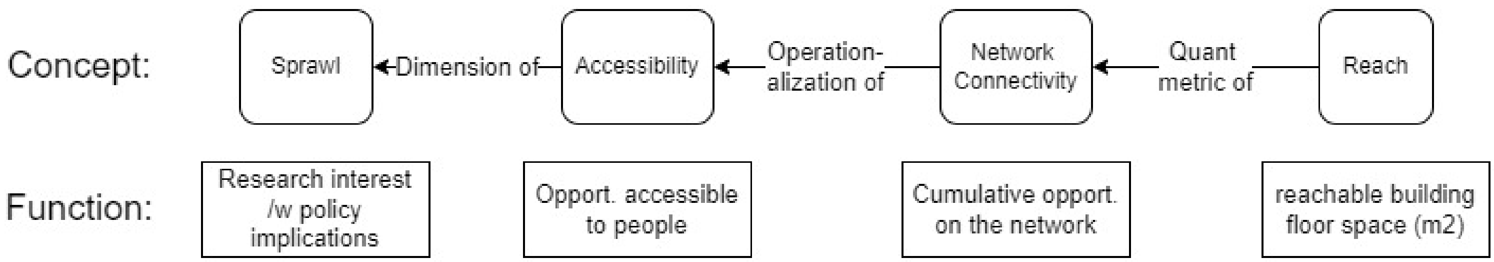
Figure 1 provides a conceptual illustration of the reach metric: it shows how many buildings on the network can be reached from the blue house given a limited walking distance (red line). The reach metric does not only capture the count of buildings, but weights for building surface area, resulting in "building area reachable" in square meters.

Figure 1: Conceptual illustration of reach metric

Source: Own illustration using OSM-data (OpenStreetMap contributors 2022).

The relation between sprawl, accessibility, street network connectivity, and reach is illustrated in Figure 2. The reach metric serves as a quantitative measure of network connectivity, and therefore further operationalizes accessibility in the framework of cumulative opportunities (Bhat et al. 2000). High network connectivity then proxies compact development, low network connectivity proxies more sprawling development. This is relevant as it will indicate for an urban form that is conducive for providing opportunities and services in the future, even if not available at present.

Figure 2: Conceptual Framework for measuring sprawl



Source: Own illustration

Potential other factors in determining connectivity

Reach will serve as the main variable under investigation in the study. Other, more intuitive metrics will be investigated as well, to cover a maximum of factors that might explain the compactness of urban form.

Distance to city center:

Distance to city center is considered an important control for connectivity. Following the mono-centric city model, lower densities and accessibility are expected further out from the central business district (CBD) (Alonso 1960). Accordingly, studies on urban morphologies usually investigate distance to the central district as well (Antos et al. 2016). In addition, Tanzanian cities have been described as surprisingly monocentric, despite their rapid growth. Notably in Tanzania, "[jobs], main services, and other central functions are still largely concentrated in CBDs or downtown areas of cities" (Huang et al. 2018, p. 12).

Density:

A common approach for investigating urban form questions for economists is to mainly consider population density and city size. Population density however ignores that urban form is the outcome of a complex bargaining process between households, business and the public sector, which culminate in street networks and land uses that persist over long periods of time (Duque et al. 2019). Assuming equal household density per building floor area, density might be the main determinant of the reach measure by definition. As population density then would proxy building area, high density in the 800m network area of a building would indicate that a high amount of building area can be reached. Under this assumption, no reach indicator was necessary for investigating the influence of unplanned areas on connectivity.

In reality, population is distributed unevenly over different neighborhoods. There might be high population densities in areas of lower building floor area (e.g., crowding in small buildings), or low population densities in more densely built-up areas (e.g., suburban areas with small household sizes). This ambiguity of the density metric has been identified as particularly difficult for analyzing cities in developing countries, as a distinction can be made between high densities through crowding in limited floor space, and high densities through vertical layering of floor space (Lall et al. 2021). Therefore, it is pertinent to control for density when looking at determinants of the building connectivity.

See Map 2 for the geospatial application of the density metric.

Built-up year:

Different time-periods likely have produced different urban forms. In the US-context, Lowry and Lowry (2014) make a distinction of different neighborhood types from three time-periods when they compare urban form metrics: pre-suburban (1891–1944), suburban (1945–1990), and late-suburban (1990–2007). In Tanzania, urban expansion has occurred over periods of more or less involved government planning and different degrees of master plan enforcement (Huang et al. 2018). The degree of planning per time-periods also varies by city (Table 1).

Additional factors on the form of urban growth dependent on time, might be the increasing urban population growth globally: sprawl has been observed as an increasing phenomenon on most countries, i.e. as urban areas and road networks grow, they tend to grow more dispersed (Barrington-Leigh and Millard-Ball 2020). Moreover, in the Tanzanian case, more recently constructed areas might tend to be informal due to recently increasing pressure on Tanzanian cities to provide housing (Kombe 2005).[1]

Therefore, the time-period that a building was constructed in, will serve as another control. In the vein of Lowry and Lowry (2014), but accounting to the recent acceleration in urban growth, this study will distinguish urban areas built before 1985, 1986-1995, 1996-2005, 2006-2015, and built after 2015 to account for unobserved temporal effects.

See Map 3 for the geospatial application of the classification.

[1] The share of people living in cities of the Tanzanian population doubled, from 15% in 1982 to 30% in 2013, to reach 35% in 2020 (UN 2018b).

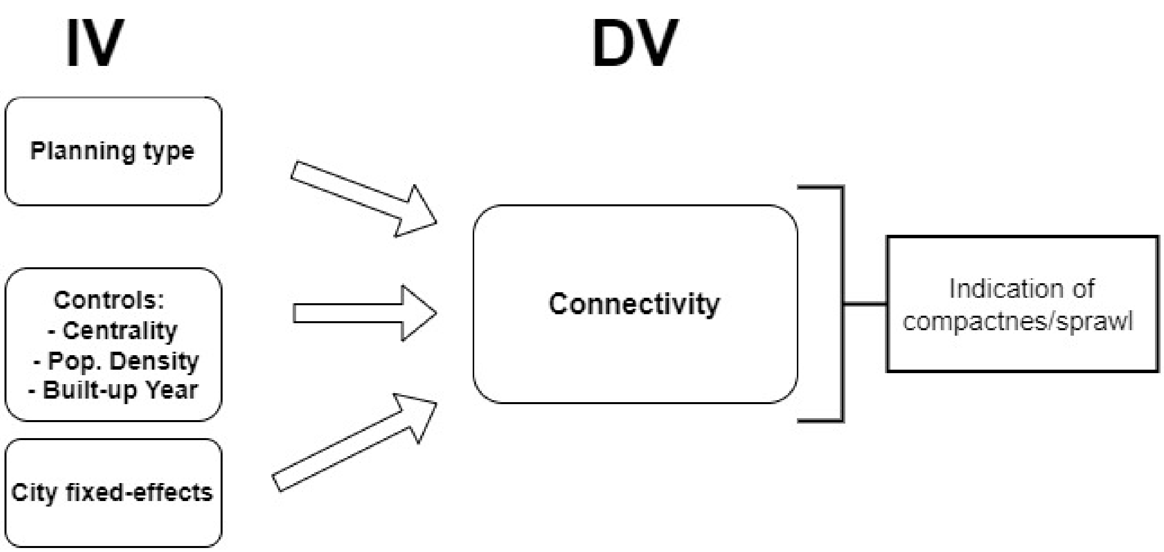
Table 1: Time-periods at least partially covered by Master Plan.

City	1970s	1980s	1990s	2000s	2010s
Mwanza					
Arusha					*
Dodoma					*
Mbeya					*
Tanga					*
Kigoma		*		*	
Mtwara				*	

* indicates new plan has been drafted. Adopted from Huang et al. (2018, p. 20)

Figure 3 illustrates the conceptual approach to the inference, including controls that might also influence connectivity. The question is whether planning type will have a significant effect on connectivity, after controlling for the other proposed factors, and including city-specific effects (Glaeser and Saiz 2003).

Figure 3: Framework for inference



Source: Own illustration

3. *Data and Methodology*

Data

The data used in this study were obtained from different, openly accessible sources. The time stamp for the different datasets ranges from 2015 to 2022 which is considered acceptable for the purpose of the analysis, given the persistence of urban form elements such as roads and buildings.

For distinguishing unplanned and planned areas, an earth-observation-based classification by EO4SD-program of the European Space Agency (ESA) has been used (European Space Agency 2018).[2] The classification layers range in their date from 2015 to 2018, which is considered acceptable for the purpose of this analysis. ESA provides a methodology for the informality classification in the respective city reports of the project (e.g.,European Space Agency 2018).

For the OSM-road network, an up to date version from geofabrik.de was used (updated 2022-06-27) as suggested by Karduni et al. (2016). Information on commercial activity in city centers was also sourced from OSM (OpenStreetMap contributors 2022). Note, that the OSM road data, is the only data used that does not rely on machine-learning detection, but manual remote-sensing.

Buildings data to feed into the network analysis is sourced from Google's "Open Buildings" project (Sirko et al. 2021).[3] Google buildings are produced with machine-learning algorithms and any analysis therefore depends on their accuracy. Mean confidence of buildings in the sample is 0.78 (SE 0.07). The gap of confidence levels for buildings between unplanned and planned areas is not large but significant (mean difference = 0.013, $t = 23.68$). Interestingly, confidence for buildings in unplanned areas is higher than in planned areas on average. What is considered important for this analysis, is that confidence levels for both planning types are at high levels and their confidence gap is not striking. Google buildings data is preferred over OSM-buildings for its exhaustiveness. Upon visual inspection, OSM-building data has remarkable data gaps in the targeted cities.

Population density at a 30x30m grid-cell level was sourced from a global dataset provided by Facebook/Meta, with a machine-learning-based methodology originally developed by Tiecke et al. (2017). With this high spatial resolution, the dataset seems better fit than other conventional global, or Tanzanian population density layers with coarser resolutions. An example would be WorldPop, which partially relies on census data and therefore rigid administrative boundaries (cf.WorldPop and Bondarenko 2020).

Data on built-up year of urban areas comes from the global World Settlement Footprint Evolution layer (WSF-Evo) developed by ESA (Mattia Marconcini et al. 2020) in the World

---

[2] A web-viewer is available here: https://urban-tep.eu/puma/tool/?id=743433804&lang=en#
[3] https://sites.research.google/open-buildings/

Settlement Footprint program . The global raster layer features annual built-up values from 1985 to 2015 at a 30m resolution. The spatial resolution is deemed well-fit for neighborhood-level analysis, and offers higher temporal granularity than comparable datasets such as the Global Human Settlement Layer (GHSL) (Corbane et al. 2018)

Method

Measuring connectivity to proxy sprawl

As outlined in the previous chapter, reach values were calculated using the UNAT for ArcMap by Sevtsuk and Mekonnen (2012). For the purpose of measuring residential connectivity, only reach values were calculated for buildings in classified unplanned and planned areas. Note that the reach values include buildings as destinations that are part of other land classifications as well (industrial, other urban uses). This is seen as beneficial for the analysis: the study is concerned with connectivity primarily of residential areas, but also as they can access areas of other land uses.

For a 2000m buffer area in the seven cities under investigation, over 1 million building values were considered from the building dataset. For computational efficiency, a sample of 10% has been taken from these building points, which is expected to be sufficient for statistical inference (Good 2005). Reach values were then computed for this sample using the UNAT in ArcMap.

Reach values were calculated for every city individually and then joined to a comprehensive dataset of 98,162 observed buildings with reach values. Datasets introduced above were spatially joined to building points. Descriptive statistics of variables and building confidence are reported in Table 2. Descriptive statistics for the key reach metric and different cities under investigation are reported in Table 3.

An exemplary output of this reach calculation for Dodoma can be seen in the annexed Map 1.

Table 2: Summary statistics

Variable	Obs.	Mean	Std. dev.	Min	Max
reach	98,162	14665.7	8292.6	6.2	59593.7
Distance to center(km)	98,162	4.33	2.53	0.03	13.29
Density (p. per cell)	98,162	17.9	17.9	0.00	110.45
Built-up year	98,162	1.69	1.06	1	5
Planning type	98,162	0.74	0.44	0	1
Confidence in buildings	98,162	0.78	0.07	0.6	0.91

Table 3: Descriptive about cities, area of interest, and distribution of reach metrics by city

	Population[4]	AoI (km2)[5]	% AoI of total AoI	Adm 2 area (km2)[6]	Reachable building area within 800m (m2)	Share of reachable building area in unplanned areas	Share of reachable building area in planned areas
Arusha	341,136	68.7	23%	267	1,359,797	87%	13%
Dodoma	180,541	51.7	17%	2,608	753,829	45%	55%
Kigoma	164,268	33.9	11%	93	409,889	33%	67%
Mbeya	291,649	52.2	17%	253	1,280,160	87%	13%
Mtwara	96,602	20.2	7%	170	238,894	45%	55%
Mwanza	436,801	52.8	17%	437	995,062	70%	30%
Tanga	224,876	22.8	8%	597	433,541	28%	72%
Total	1,735,873	302.3		4,424	5,471,171		

For density, missing values in the raster have been interpolated using "Focal Statistics" in ArcGIS Pro (Mitas, L., Mitasova, H. 1999). In particular, missing cells were assigned the mean of a circle with a five-cell, i.e., 150meter radius.

Further data diagnostics are reported in Annex III: Data Diagnostics. Detailed documentation of the spatial analysis workflow, as well as the STATA .do file is provided in Annex IV: Spatial Analysis Workflow Documentation and .do-file.

Determining city centers

Determining city centers remains a tricky exercise in absence of previously defined CBDs, or municipality-defined centers. For this study, central city areas were visually inspected for a combination of 1) an intersection of trunk roads, and b) a high density of commercial banking branches, following loosely classic approaches to determining CBDs (Murphy and Vance 1954). In the absence of the viable data, this study has to shy away from more

[4] https://worldpopulationreview.com
[5] Calculation made from EO4SD informal classification polygons
[6] Calculation from admin data https://data.humdata.org/dataset/cod-ab-tza

sophisticated approaches to determining the economic centers of cities (Borruso and Porceddu 2009; Afrose et al. 2019).

As, cities in Tanzania are considered to have developed in a monocentric fashion (Huang et al. 2018), a visual inspection of central banking density from OSM is considered appropriate for the purpose of this analysis. After determining city center points, distance to the respective city center points were calculated using the "Near" tool in ArcGIS Pro, i.e., each building was matched to the respective city center point and the distance calculated.

Creating a network dataset from OSM-roads

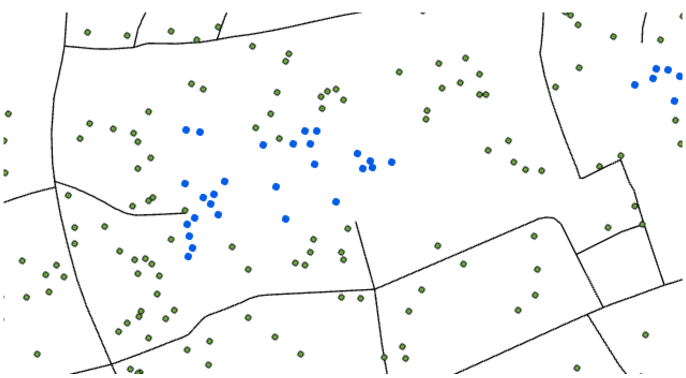
The ArcMap plugin GIS2FE was used to create an edge and node list for network building from OSM-data (Karduni et al. 2016).

2000m buffer applied to informal layer to cover sufficient street network to cover the 800m network analysis (2k Buffer, dissolve into one feature).

Treating reach values of zero

Approximately 1% of buildings in the sample received a reach value of 0 by the UNAT algorithm. In a few cases this might reflect reality for buildings on the fringe of the city that were still included in the informality classification by the EO4SD. Visual inspection reveals however, that in most cases this might be due to a network error, where a building could not be assigned to a road or path (Figure 4). This phenomenon occurs in all cities but is more likely to occur in unplanned areas (Table 4). Given that the reach value of 0 indicates a network error, these observations were deleted from further analysis.

Figure 4: Observations with reach of zero.



Buildings (in blue) with reach values of 0, not connecting to the network (lines in black). Own Illustration.

Table 4: Distribution of reach=0 in unplanned and planned areas

Planning type	Reach = 0
planned	5
unplanned	1,094
Total	1,099

Investigating the difference between unplanned and planned areas

The quantitative approach to determining whether unplanned areas are less connected than planned areas, is to build a regression model and estimate if there is a statistically significant difference in the dependent variable "reach" between these two types of neighborhoods.

The reach-values associated to buildings on the street network are then used in linear regression to model the association, with distance to center, population density and built-up year periods of the building as controls. The location-inherent differences between cities are captured with city-fixed effects, as is common in urban econometric research (Glaeser and Saiz 2003).

Ordinary-least-squares (OLS) assumes that errors are independent of each other and normally distributed (Mehmetoglu and Jakobsen 2017, p. 235). To address spatial autocorrelation, standard errors have been clustered following the criticized but common practice of clustering standard errors at the lowest level available with Stata default commands. This is under the assumption that variation using other cluster options tends to be small (Kelly 2020, p. 2). Clusters were sourced from the EO4SD's land-use classification layer, which has defined land-use polygons of broader classifications. The polygons relevant to this layer are classified as residential or commercial, and it is assumed that the buildings in one polygon share certain neighborhood characteristics. See Map 2 for an illustration of the clusters in Dodoma. This was done in absence of a better neighborhood area classification, where the values of one building are highly correlated with its neighboring building. The result are 14,368 clusters of buildings across the seven cases.

The model specification is as follows:

<u>Linear Model</u>

$$\text{reach} = \alpha + \beta_1 * (\text{unplanned/planned}) + \beta_2 * (\text{distance to center}) + \beta_3 * (\text{density}) + \beta_4 * (\text{year}) + e[\text{clustered}]$$

assuming that $E[e_i e_j] = 0$,

i.e. that errors are correlated within clusters but not across clusters (Colin Cameron and Miller 2015).

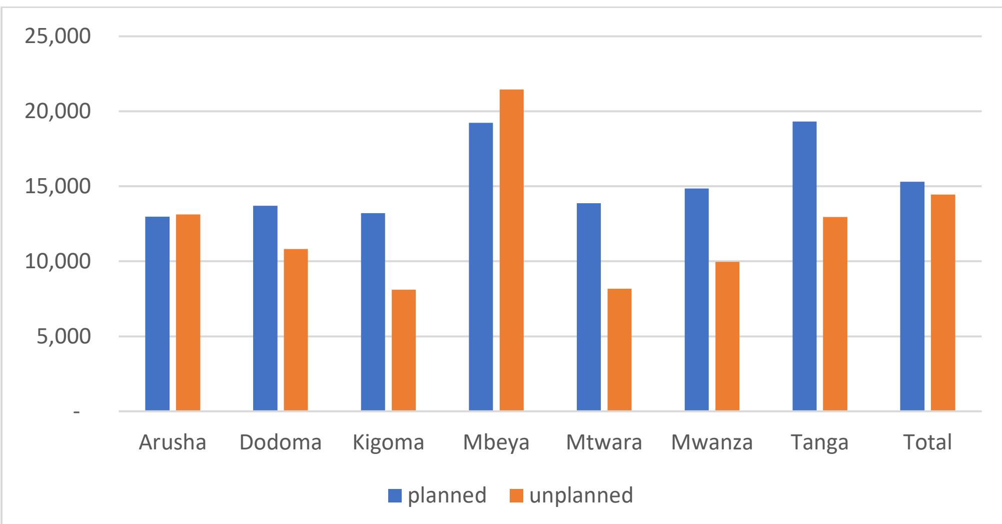
4. Results

Before building a regression model to evaluate the association between connectivity and planning type, some associations are explored with descriptive statistics to understand the phenomenon hat hand.

Descriptive Statistics

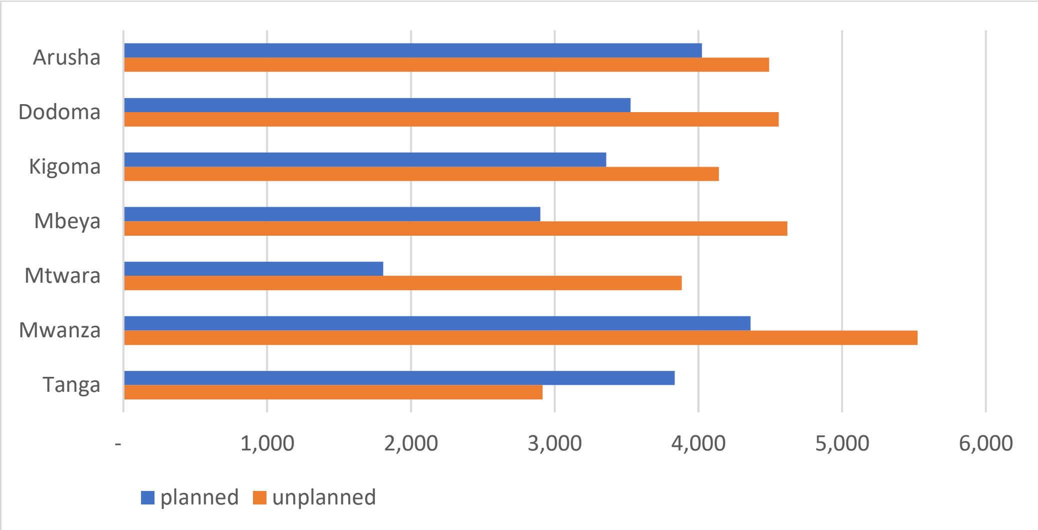
Comparing mean reach values indicate that there is not a big but significance difference in overall mean reach between unplanned and planned areas. For all cities, mean reach values are 6% lower in unplanned areas. This 6% difference is significant at the 99%-level (difference of 858m2, SE 60.5, t -14.18). However, for different cities, mean reach values differ more. Unplanned areas having lower reach values in all cities but Arusha and Mbeya. The extreme cases here are Mtwara with a 41% lower mean reach in unplanned areas, and Mbeya with a 11% higher mean reach in unplanned areas (Figure 5).

Figure 5: Mean reach values for different cities, by informality.



The next intuitive question is whether buildings in unplanned areas are generally further away from the city center. On average this is true for all cities but Tanga (Figure 6).

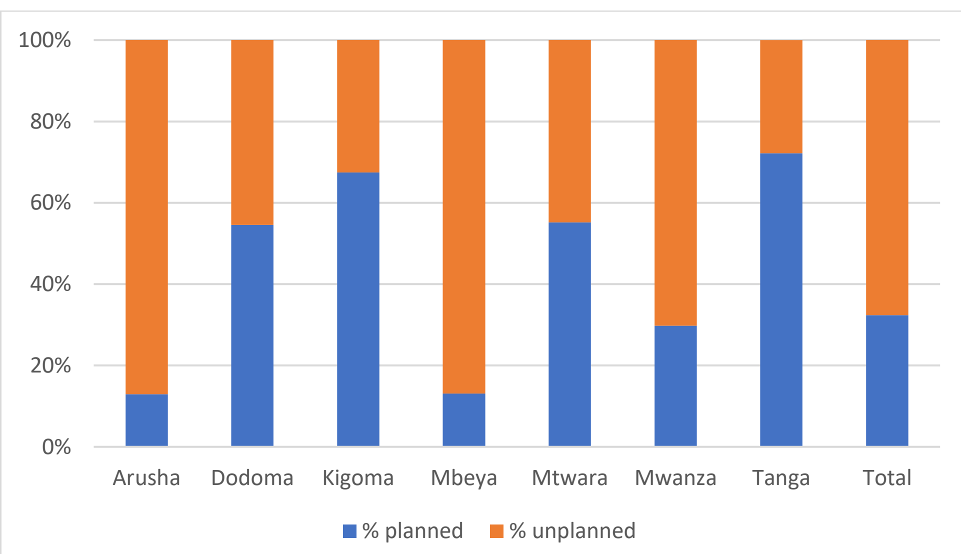
Figure 6: Mean distance to city center by planning type (m)



However, the metric above might be biased by the larger number of smaller buildings in unplanned areas. Looking at building area, the weight that will be form the reach value lastly, can bring more clarity. The distribution of building area in planned and unplanned areas varies remarkably among cities, which is a function of the informality classification. For all cities, one third of building surface area is in planned areas, and two thirds is in unplanned areas (Figure 7). Looking at the total number of buildings, regardless of size, 25% of them are in planned areas, vs. 3 quarters in unplanned, showing the bias of only considering building count.

In conclusion, buildings in unplanned areas are on average further out from the center and have lower reach values in most cities. The distribution of building area between unplanned and planned areas is quite uneven among cities.

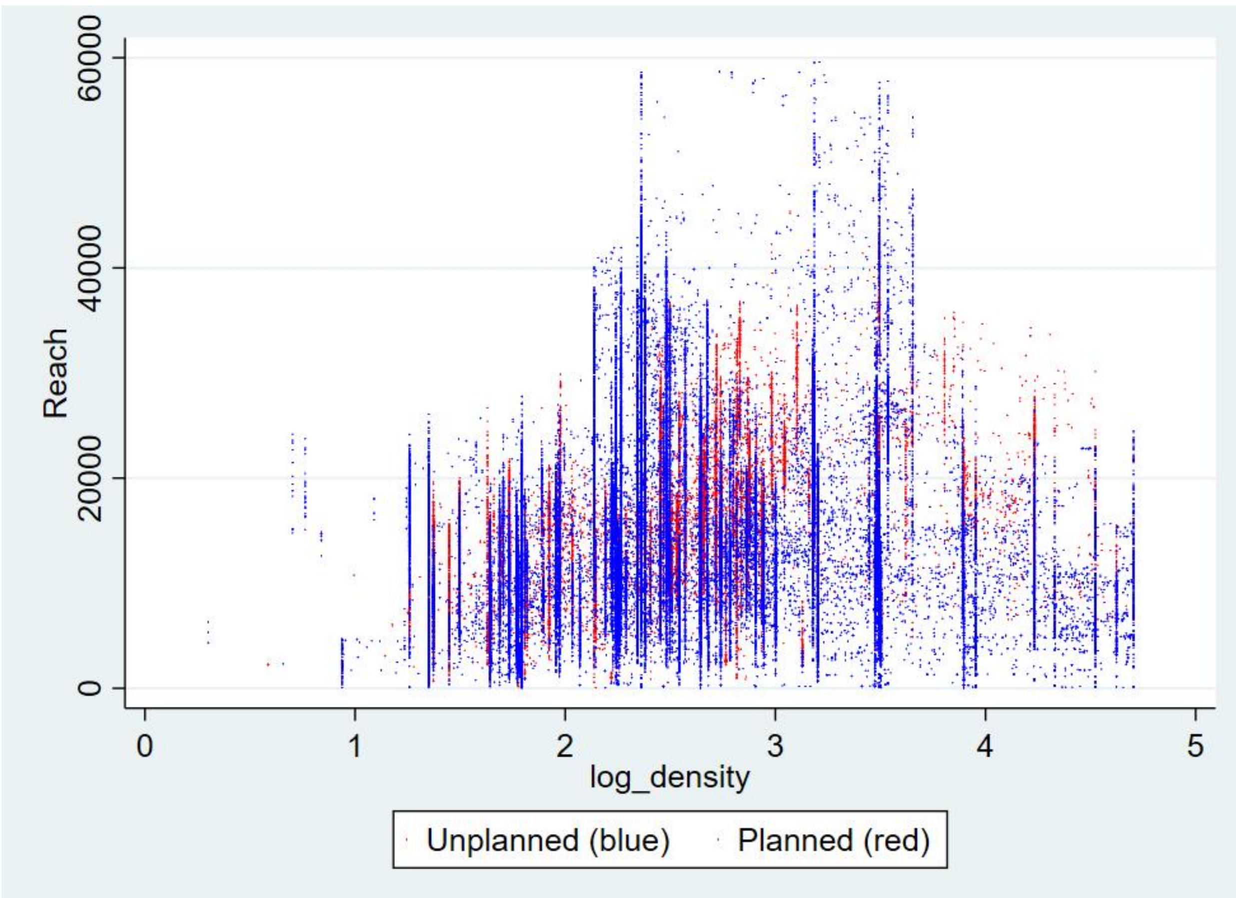
Figure 7: Distribution of building area between planned and unplanned areas, by city



Association of density and reach

Figure 8 shows that reach tends to be highest at medium densities for both unplanned and planned areas. It is also at these medium densities where unplanned areas seem to feature higher reach values. At the highest densities however (log_density >3.5), buildings in planned areas seem to have higher reach values. An important finding here is, that the highest reach values do not seem to coincide with the highest density values. This gives reason to consider the reach metric a useful addition to describing urban form, where density, especially in the local context of overcrowding can be misleading (cf. Lall et al. 2021). In line with this finding is the rather low correlation between log_density and reach of 0.12 for the overall sample.

Figure 8: Association of reach and density by planning type



Reach and density are correlated around 0.5 for cities individually Table 5. The crass exception being Mwanza, where correlation is close to 0, which might indicate a flaw in the density data for the city.

Table 5: Correlation of Reach, Distance, Density and Planning Type by city

	Reach vs Distance	Reach vs. Density	Density vs. Distance	Distance vs. unplanned	Reach vs. unplanned
Arusha	-0.43	0.47	-0.65	0.06	0.01
Dodoma	-0.42	0.58	-0.42	0.28	-0.26
Kigoma	-0.17	0.47	0.23	0.22	-0.55
Mbeya	-0.31	0.37	-0.39	0.16	0.07
Mtwara	-0.67	0.57	-0.51	0.67	-0.56
Mwanza	0.01	0.0004	-0.71	0.17	-0.36
Tanga	-0.53	0.54	-0.57	0.47	-0.41

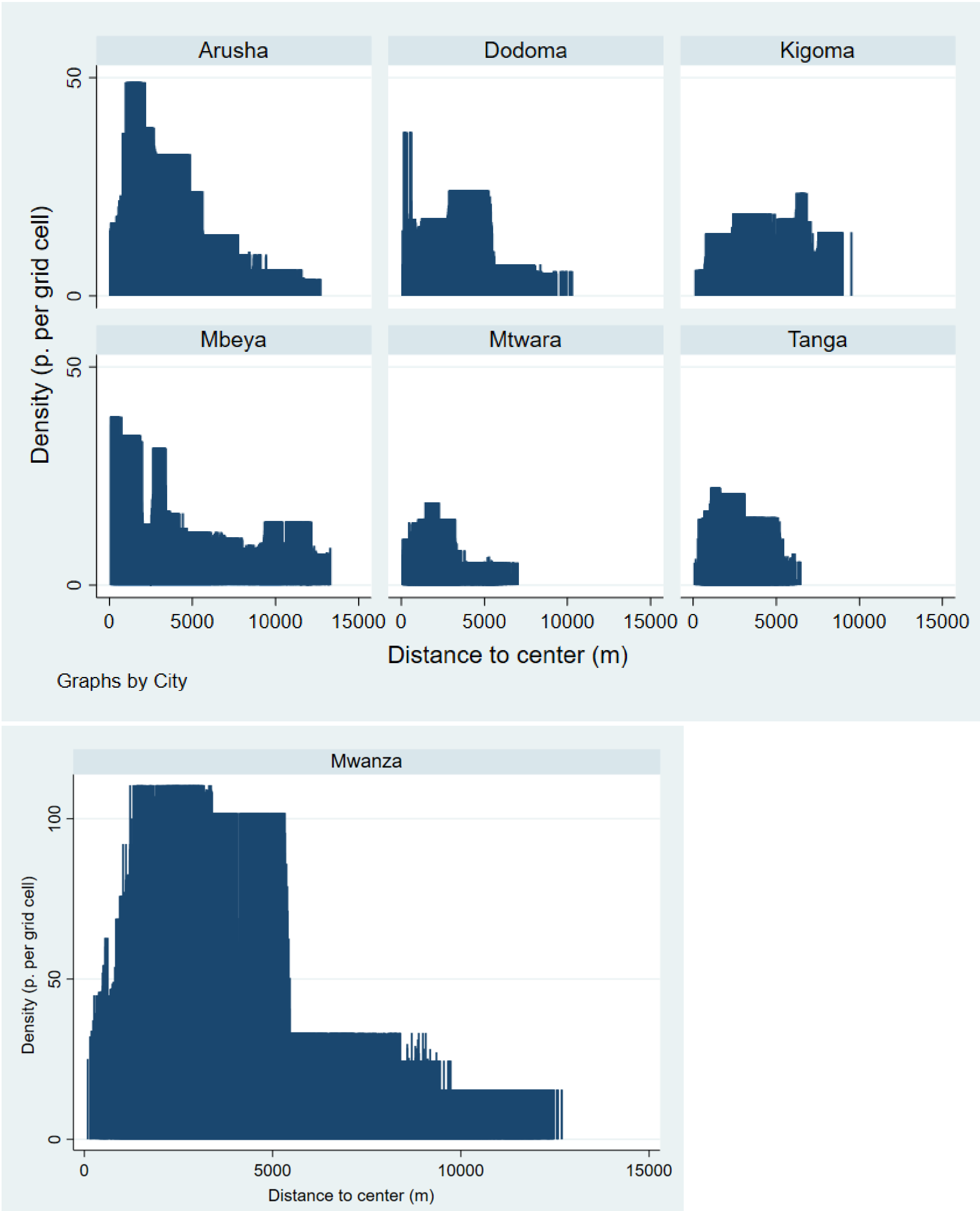
A further evaluation of the relationship between density and reach, reveals that the association is likely to be curvilinear, with different levels of density having different effects on reach. This might be, as the built-up structure of neighborhoods differing in density may vary substantially. Further justification for the inclusion of polynomials for density is provided in the annex, Figure 18.

Association of density and distance to center

As expected in the context of developing countries' cities varying densities through smaller floor space consumption, density gradients for the cities under investigation are ambiguous (Figure 9). The two largest cities show density gradients closer to what is expected in the monocentric model (Arusha, Mwanza). Other cities have a flatter density gradient, or experience higher densities further out from the city center (notably Dodoma, Mbeya, Kigoma). For Tanga and Mtwara it seems to be the case that residential densities simply start to be elevated slightly off the commercial CBD, as residential occupation there might be low. This can also be said for Arusha to a small extent. Their density gradient otherwise follows the monocentric model.

In line with this visual interpretation, Table 5 shows that correlation of density and distance are particularly high for Arusha and Mwanza, but less so for other cities. This ambiguity underlines that monocentric density gradients cannot be taken for granted in these cities, and investigation of other metrics of urban form, such as reach make sense.

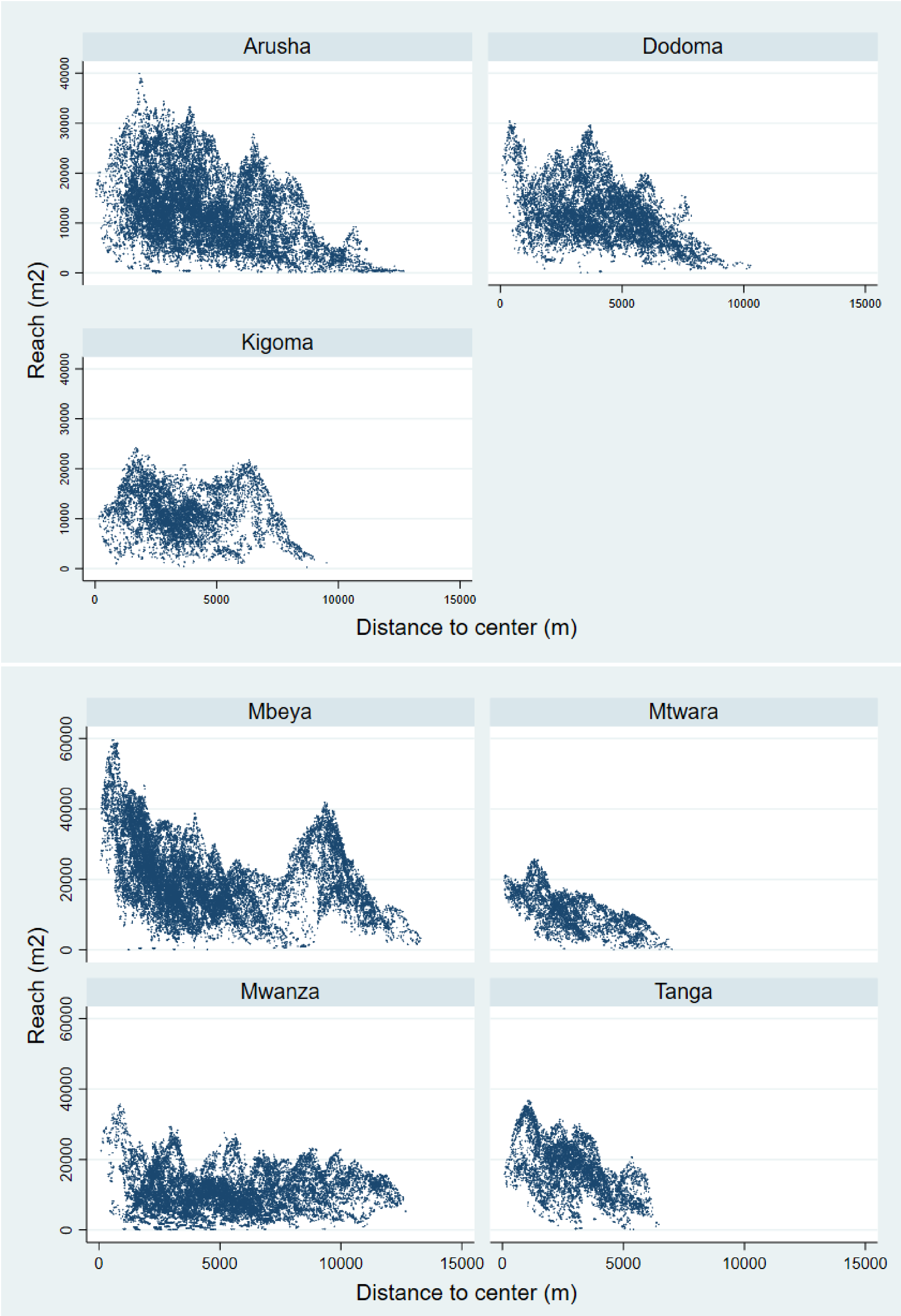
Figure 9: Density gradients for the seven cities



Association of distance to center and reach

In line with density gradients, “reach gradients” show to be non-linear for some cities (Figure 10). This is most notable for Mbeya with a peak second peak in reach almost 10km from the center. Particularly flat reach gradients can be seen for Kigoma and Mwanza. As indicated in the low correlation of reach and density for Mwanza (0.004) this can be noted graphically as the reach and density gradient for the city show stark contrasts. In Mwanza reach remains steady, where densities appear to fall dramatically.

Figure 10: Association of distance to center and reach

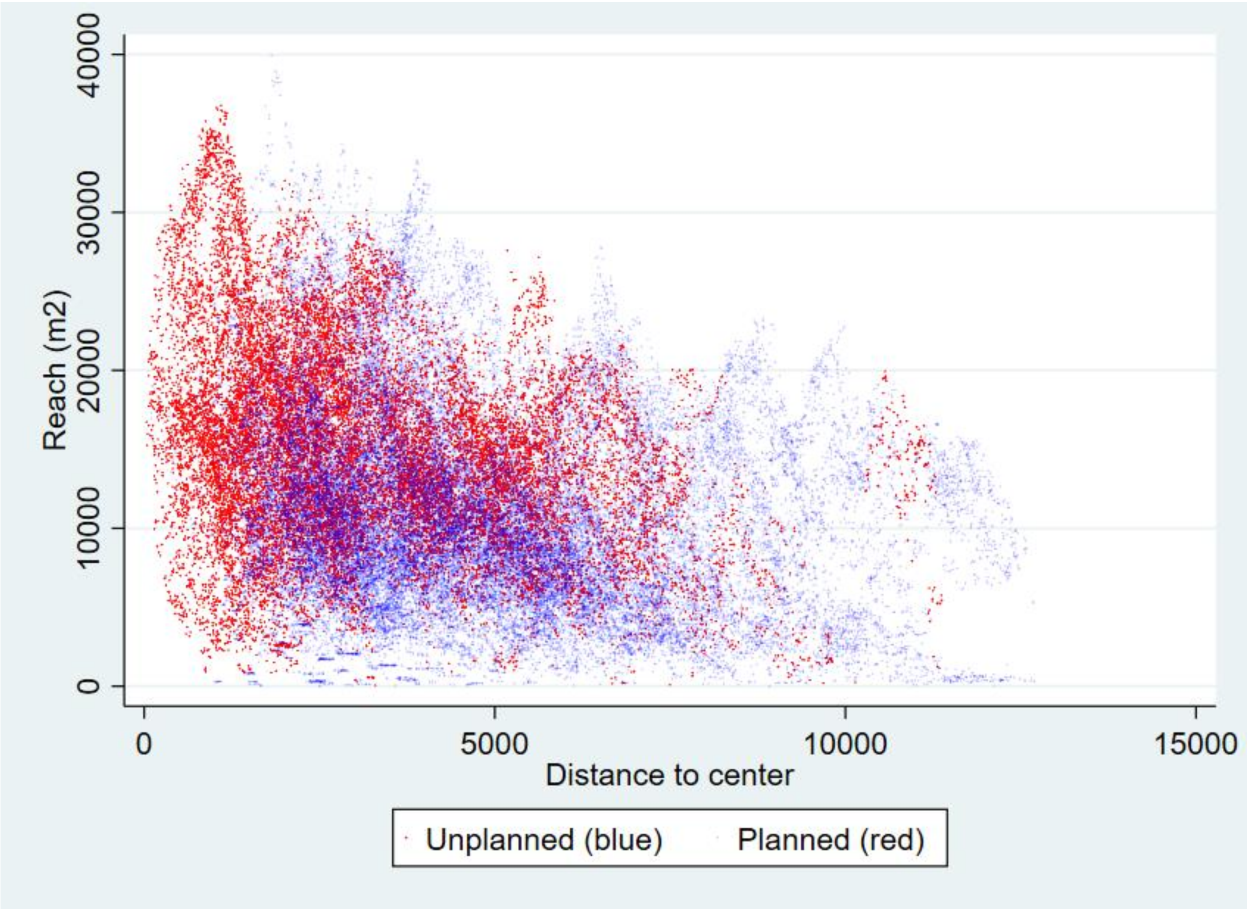


Association of distance to center, reach and density

In general, it appears that for the exception of Mwanza, reach and density gradients show graphic overlap from this bird's eye view. Yet, the varying correlation between cities presented in Table 5, indicates that the separate inclusion of the two metrics is desirable to capture the entire effect of planning type.

Lastly, the comparison of unplanned and planned buildings' reach values on a density gradient is difficult to interpret (Figure 11). Buildings closer to the center tend to be planned (as excepted for older city cores, and their reach tends to be highest (as expected from their centrality). 2km away from the center, the effect of planning type becomes less clear, as all the proposed regressors likely play a role in determining reach, along with planning type. Linear regression can help to evaluate the isolated effect of planning type (Mehmetoglu and Jakobsen 2017).

Figure 11: Association of reach and distance to center by planning type



Regression Results

Results of linear regression using OLS point to a significant negative association between the reach value and planning-type of the surrounding area. This would mean that buildings in unplanned areas area associated with lower reach values, even when controlling for distance to city center, density and built-up period. Results of OLS are reported in Table 6. As expected, distance to city center does play a significant role. A building one kilometer further away from the city center point has 274m^2 less reach. The effect of this relationship adds up of course if distance increases. For Mwanza however, the relationship seems to be reversed (Table 8 in annex).

The curvilinear relationship between reach and density turns out to be significant. At very low and very high densities, reach values seem to be affected negatively by density. At medium densities, the relationship is positive.

The effect of built-up year proves significant at the 95% or higher and to be increasingly strong the more recent the building has been built. This is interesting as the general correlation of built period and reach is merely 0.24. This indicates that there are two different effects at work here. Younger buildings have lower reach values, both in unplanned and planned areas, even when controlling for density.

Lastly, city fixed-effects are significant, with the exception of Dodoma and Mtwara. This reflects the general heterogeneity in reach levels between cities and incorporates other city-specific effects that are not observed.

Table 6: Linear Regression (DV: reach)

	(1) All cities	(2) High-coefficient	(3) Low- coefficient
Distance (km)	-273.9***	94.56*	-464.9***
	(-4.35)	(2.06)	(-5.46)
Planning type (Ref: Planned)			
Unplanned	-1648.9***	-4301.5***	473.4
	(-8.35)	(-25.05)	(1.41)
log_density	-25177.9***	1526.9	-24960.2**
	(-6.22)	(0.59)	(-2.90)
log_dens²	11493.4***	1198.6	9931.1**
	(7.71)	(1.28)	(2.74)
log_dens³	-1453.0***	-253.8*	-1040.6*
	(-8.64)	(-2.37)	(-2.13)
Built-Up Year (Ref: Pre 1985)			
1986-1995	-590.1*	-2507.4***	316.6
	(-2.26)	(-14.19)	(0.97)
1996-2005	-3377.8***	-2290.6***	-3364.5***
	(-15.27)	(-11.07)	(-12.50)
2006-2015	-4203.3***	-3445.9***	-3947.5***
	(-16.05)	(-15.79)	(-11.70)
2016+	-8580.5***	-8364.3***	-8469.6***
	(-17.11)	(-12.03)	(-16.42)
City (Ref: Arusha (Spec. 1+3)/ Kigoma (Spec. 2)			
Dodoma	258.1		914.5*
	(0.63)		(2.14)
Kigoma	-3676.5***		
	(-12.56)		
Mbeya	8357.0***		8689.6***
	(19.53)		(19.43)
Mtwara	-444.4	3484.7***	
	(-1.22)	(14.33)	
Mwanza	-5250.3***	376.8	
	(-11.77)	(1.33)	
Tanga	2154.6***	6216.6***	
	(6.80)	(28.95)	
Constant	30568.7***	5601.5*	31359.4***
	(8.87)	(2.43)	(4.82)
Observations	98,003	38,235	59,768
R²	0.41	0.37	0.41

t statistics in parentheses
* p<0.05, ** p<0.01, *** p<0.001

R^2 improves from below 0.1 to an acceptable 0.41 when adding city fixed-effects to the model (Mehmetoglu and Jakobsen 2017). R^2 values for cities' individual regressions vary from 0.19 (Mwanza) to 0.59 (Mtwara). The values compare to spatial regression R^2 on the neighborhood-level in the field. E.g., Pramanik et al. (2022) find an R^2 of 0.52 for spatial OLS at the neighborhood level.

Comparison connectivity of unplanned and planned areas

Comparing predicted margins of a categorical variable can help visualize its effect on the dependent variable well, making interpretation of model results more intuitive, holding all other regressors constant (Mehmetoglu and Jakobsen 2017). This is done here to illustrate the net effect of planning type on reach, while only varying the distance to center. Figure 12 shows that unplanned areas on average have considerably lower reach values at all distances from center, when holding other variables constant (namley density, built-up year and city fixed-effect). The effect is less pronounced in the city center (judging by 95%-confidence intervals) and most pronounced 5km away from the center. Starting from 11km from the center, confidence intervals overlap, rendering the prediction imprecise. Table 7 shows that in the distance beyond 9km from the center, observations of buildings in planned areas are simply too few to make good predictions.

Figure 12: Predicted reach values by planning type

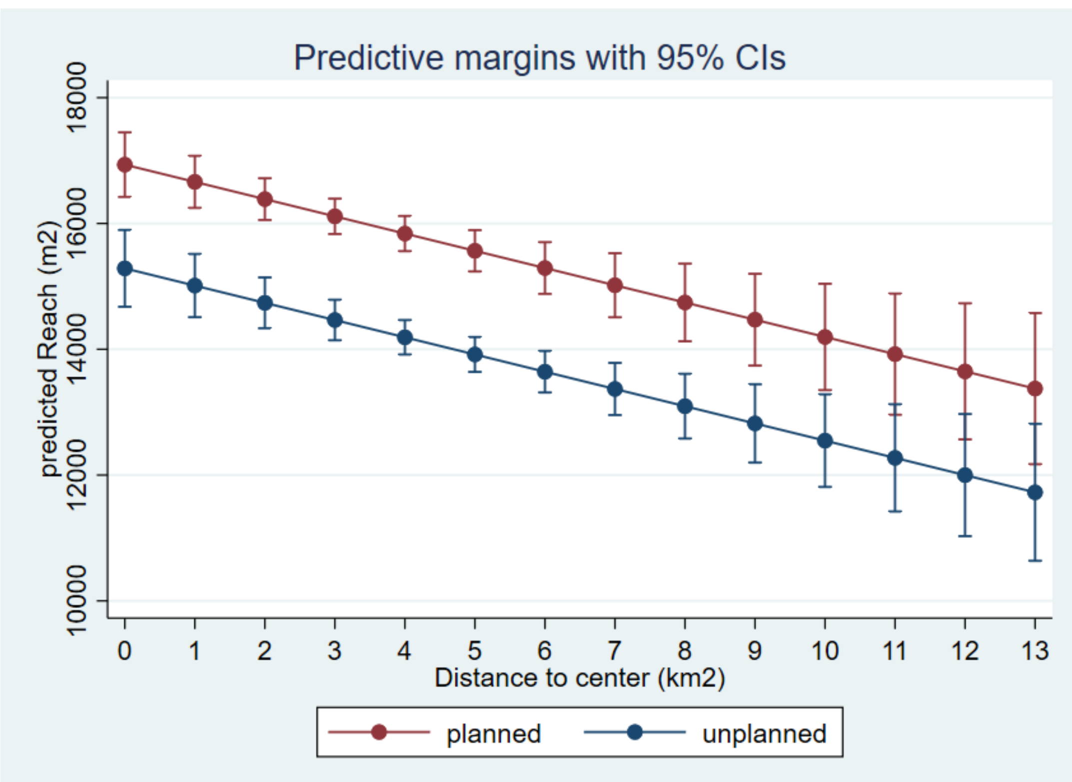


Table 7: Planned and unplanned observations by distance to center

Distance to center (km2)	*planned*	*unplanned*
0	*2,210*	*1,430*
1	*5,791*	*7,199*
2	*5,697*	*13,085*
3	*3,639*	*13,515*
4	*2,879*	*10,305*
5	*2,486*	*8,401*
6	*2,006*	*9,254*
8	*260*	*2,847*
9	*82*	*3,411*
10	*99*	*1,998*
11	*55*	*1,173*
12	*0*	*311*
13	*0*	*29*
Total	*25,204*	*72,958*

This result seems quite straightforward, indicating a significant effect of planning type on reach. However, as presented in Figure 5 above, the distribution of reach values between unplanned and planned areas varies considerably between cities. I therefore propose to split the model in two types of cities: A) those with a difference in mean reach between planned and unplanned areas above 25% (high coefficient cities), and B) those cities with mean reach difference below 25% (low coefficient cities). These coefficient differences are also observable when comparing the model results for each city individually as reported in the annexed Table 8.

High-coefficient cities then comprise of Kigoma, Mtwara, Mwanza, Tanga, and low-coefficient cities are Arusha, Dodoma, and Mbeya (with Mbeya showing a reverse relationship of planning type and reach). Regression results for these two groups vary dramatically (Table 6, specification 2 and 3). Similarly, predicted reach values in relation to city center shows indicate a high influence for high coefficient cities (Figure 13), but an insignificant influence for low coefficient cities (Figure 14).

In high-coefficient cities, the effect of distance to city center is surprisingly positive, meaning that further away from the city center reach values are higher on average. However, unplanned areas have a striking 4301m^2 less building area in their reach than planned areas. In low-coefficient cities, the negative effect of distance to center is even more pronounced, pointing to a more classical monocentric city model with declining connectivity in the sense of Alonso (1960). The effect of planning status is insignificant here, which can be inferred from the insignificant coefficient, and from the unanimous overlap of confidence levels in the predicted reach values (Figure 14).

Similarly to the first model specification for all cities, confidence intervals become larger further away from the center, in part due to a limited number of observations.

Figure 13: Predicted reach levels for high-coefficient cities (Kigoma, Mtwara, Mwanza, Tanga)

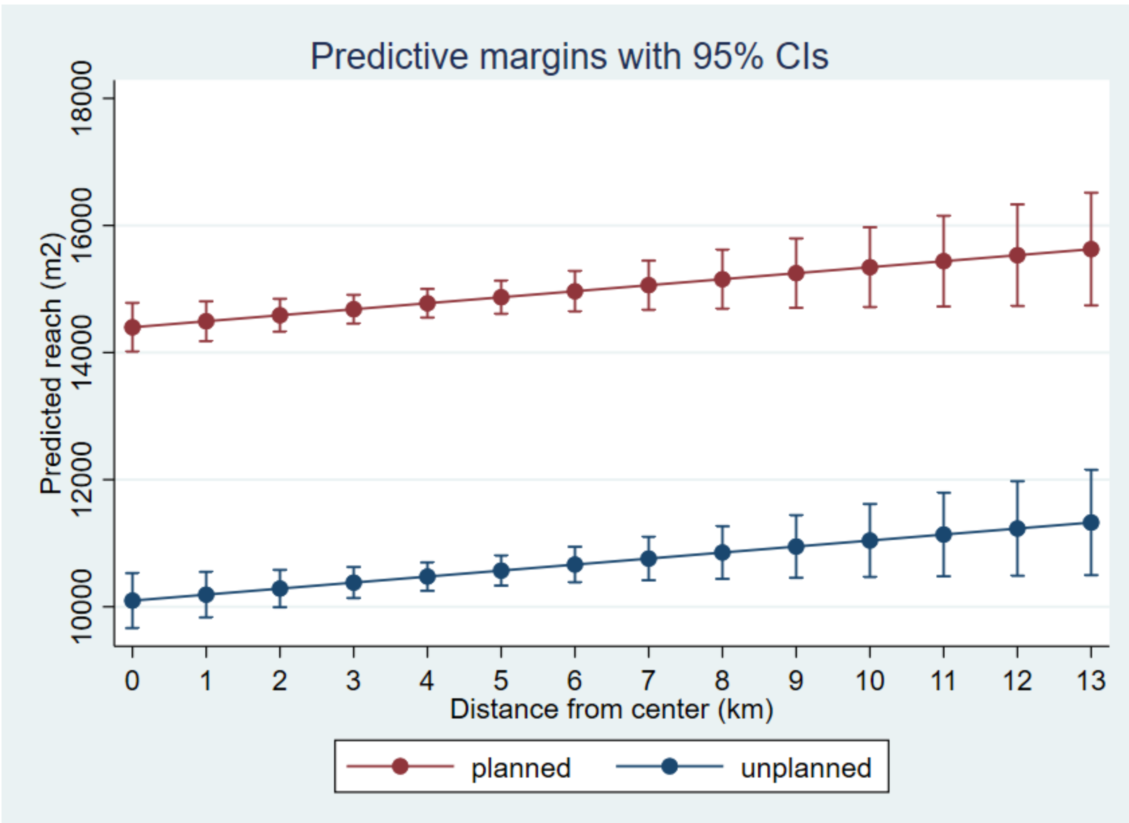
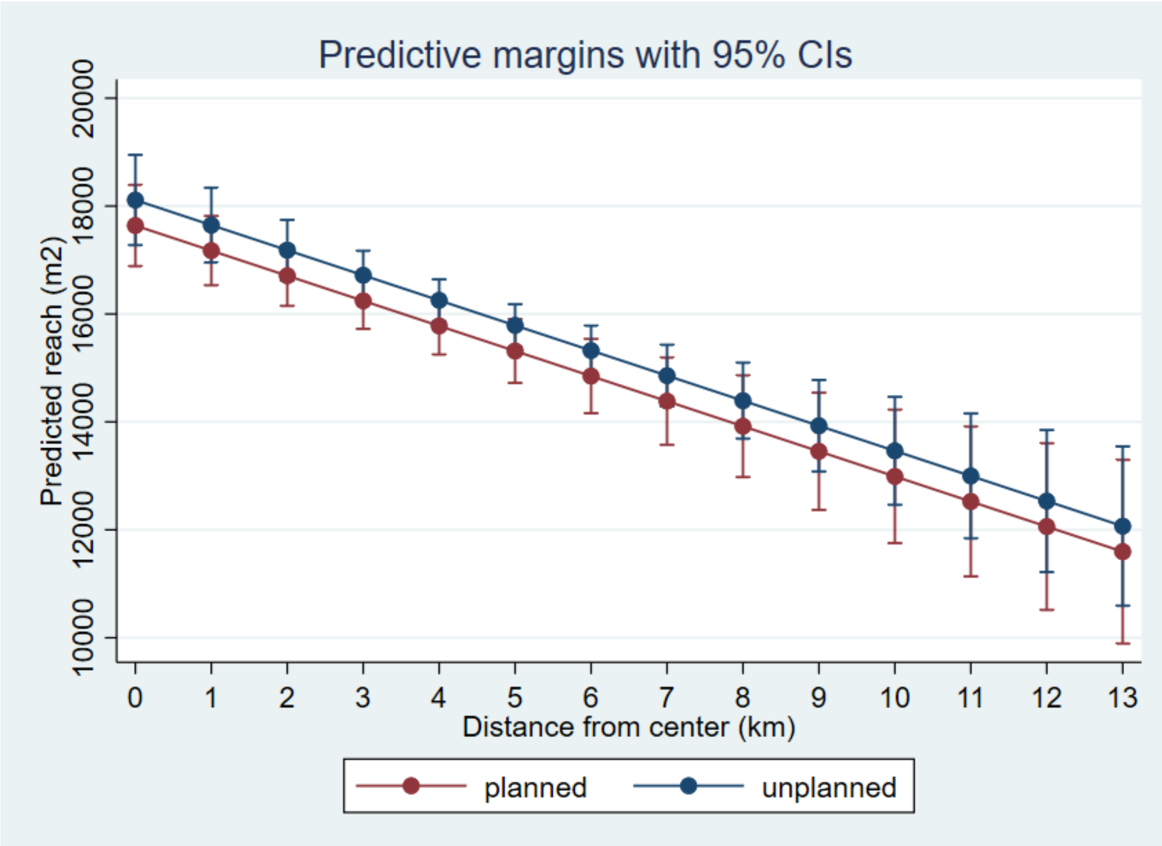


Figure 14: Predicted reach levels for low-coefficient cities (Arusha, Dodoma, Mbeya)



Limitations

The built model presented above provides interpretable results. However, there are a few caveats to the study.

The quality of OSM-data and particularly the role of foot paths might have an unmeasurable influence on the results. Particularly in unplanned areas connectivity may not be captured as well as in planned areas. Limited evidence is available of the accuracy of OSM-street data in informal areas. Yeboah et al. (2021) found that informal areas were mapped with an accuracy of up to 73% which might not be enough for a meaningly connectivity analysis, as footpaths could that are not mapped could increase reach in informal areas.

In the absence of building height, or floor level data, the study relies on the assumption that secondary cities in Tanzania are mostly flat, i.e., single-floor buildings. For smaller cities, such as Kigoma or Mtwara this might be truer than for larger cities, such as Dodoma, or Arusha, whose central areas certainly have multi-floor buildings. Central areas in such cities would actually have higher reach values than captured in the study, due to their vertical layering of floor space (Lall et al. 2021). However, in the light of the results, that unplanned areas, who seem more likely to be actually flat, have lower connectivity and not more, the assumption seemed not have disturbed results.

The base for the EO4SD-classification of unplanned areas, are a bit of a black box, but are based on building size, roof materials, and other remotely sensed criteria (European Space Agency 2018). There could be an endogeneity issue if this means measuring connectivity to buildings along the road network (measures in this study) to explain buildings and form of the road network (unplanned-classification). The unclear definition of the EO4SD-classification might also be the cause for widely different effects across cities.

Using an OLS model in spatial context can lead to biases due to spatial autocorrelation. Since spatial data is interdependent according to the first law of geography, OLS assumptions are violated which might lead to bias in OLS estimates. This could be avoided by accounting for spatial dependence in the dependent variable and introduce explicitly spatial models which build on OLS (Pramanik et al. 2022). The introduction of such models was beyond the scope of this analysis.

For modelling simplicity, it has been assumed that errors are correlated within clusters but not across clusters (Colin Cameron and Miller 2015). In reality, standard errors might well be correlated across the chosen clustered boundaries, therefore a more sophisticated modelling approach might be necessary to address this issue. One such could be the above mentioned explicit spatial models.

The study was partially motivated by the finding of Sheppard (2010) that informal areas consume less land per capita than formal areas. Unfortunatly, the study did not address land consumption per capita in the sense of Sheppard (2010), as it was beyond the scope of the analysis.

The determination of city centers surely was subjective, and more modern sophisticated approaches to determining CBDs exist (Borruso and Porceddu 2009).

Lastly. the quantitative analysis in this study heavily relied on GUI-based GIS-analysis, which decreases its reproducibility and transparency. An open-source programming language such as Python might have been better suited to achieve transparency and applicability. Not 100% of all rationales that led to analytical decisions might have been properly documented, due to the large amount of parameters to decide on in GUI-based analysis (Fleischmann et al. 2021). However, an attempt to reporting is provided in Annex IV: Spatial Analysis Workflow Documentation and .do-file.

5. *Discussion*

The results of the regression analyis indicate that there is considerable heterogeneity among Tanzanian secondary cities for the effect of planning type on connectivity. Descriptive analysis showed that conventional urban form indicators such as density gradients also point to a heterogeneity of the cities in question. This can be considered as a first important finding, reminding us that speaking of urban form and investigating potential interventions should be highly context specific, even for an assumingly homogenous group of cities, such as secondary cities in the light of Dar es Salaam's urban primacy.

The second and central finding is that for some cities the planning type did have a significant and considerably large effect on connectivity, namley for Kigoma, Tanga, and Mwanza, the second largest city in Tanzania. Assuming robustness of the OLS-model, this would mean that unplanned areas contribute more to sprawl than planned areas in these cities. In practice, the difference in 4301m2 of reach for a resident of an unplanned area would mean, he or she could reach within 800m, 150 fewer median homes (35m^2) on average. Surprisingly for these cities, connectivity even increased further away from the center (even though with a comparably small magnitude and lower significance level), questioning the applicability of conventional urban form models that assume some kind of gradient of urban activity (cf. Alonso 1960).

The atypical density gradient for Kigoma and the mismatch between density gradient and reach gradient for Mwanza, gives reason to believe that including the reach metric was crucial to arrive at this finding, as density alone would have not revealed these differences. A third finding therefore could be that the network analysis-based reach metric provides a useful addition to conventional metrics of urban form (Ewing et al. 2003).

The fourth central finding then would be that in some cities, including central drivers of Tanzanian urbanization such as Arusha and Dodoma, planning type did not significantly affect connectivity, and – in the framework of this analysis – cannot be attributed more to sprawl than planned areas. This finding might be informative for debates on the negative effects of sprawl in SSA cities for commentators such as Tchekoté and Ngouanet (2015), or Cobbinah and Amoako (2014). Most importantly, it could caution high-level policy goals to focus on informal sprawl containment at all costs in developing countries, when occurring urban expansion might not fit the description, even though it happens in an unplanned, haphazard way (UN 2018a, cf.).

The general negative effect of building period on connectivity aligns with the literature on increasing floor space consumption. When incomes increase, as they have steadily in Tanzania since the 1990s (World Bank 2022b), people consume more living space (Angel et al. 2011). In addition, built-up year might proxy for distance from center, although not true for all cities (e.g., Mwanza almost completely built-up pre-1986 in this sample) which puts emphasize on this trend.

6. *Conclusion*

This study set out to provide new evidence on the association between informality and urban sprawl. Sprawl here was conceptualized as the opposite of compact urban development in the understanding of international development actors, such as the World Bank. The results are meant to inform researchers' conception of unplanned urban growth and guide their research, as well as policy makers' understanding of the phenomenon. It is hoped this will lead to more context-specific recommendations for urban form interventions. Another aim of the study was to help refine local definitions of sprawl, in the light of a largely Euro-American centered debate on the term (Brown 2017).

The results of the study revealed that the answer to whether unplanned areas are associated more with sprawl than planned areas in rapidly expanding secondary cities in Tanzania, is not straightforward. Remarkable heterogeneity in the effect of planning type between cities in the country point to the conclusion that planning type should not be attributed to sprawling urban expansion *per se* in such cities. On the other hand, it does seem like it can exacerbate sprawl in some cities. City-specific investigations therefore would be needed, to arrive at tangible results to communicate to policy makers.

Further, the heterogenous results should be treated with caution when thinking of secondary cities in other countries. Those might feature again heterogenous association between planning type and urban form. This leads to the question, what would be more robust determinants of sprawling urban expansion in such cities? Ideally, such determinants would hold across cities, and potentially countries with similar urban development regimes. This study hopefully creates appetite for others to embark on such research trajectory. The main contribution of this study is, that informality is likely not the key determinant for compact and sprawling urban expansion.

References

Abebe, Fikresalassie K. (2011): Modelling informal settlement growth in Dar es Salaam, Tanzania. PhD Dissertation. Twente.

Afrose, Sadia; Riyadh, Ahmad Mojtoba; Haque, Afsana (2019): Cores of Dhaka city: area delimitation and comparison of their characteristics. In *Asia-Pac J Reg Sci* 3 (2), pp. 521–560. DOI: 10.1007/s41685-019-00112-z.

Alonso, William (1960): A Theory of the Urban Land Market. In *Papers in Regional Science* 6 (1), pp. 149–157. DOI: 10.1111/j.1435-5597.1960.tb01710.x.

Angel, Shlomo; Parent, Jason; Civco, Daniel; Blei, Alejandro (2011): Making room for a planet of cities. Cambridge, MA: Lincoln Institute of Land Policy (Policy focus report. Lincoln Institute of Land Policy).

Antos, Sarah E.; Lall, Somik V.; Lozano-Gracia, Nancy (2016): The Morphology of African Cities. Policy Research Working Paper;No. 7911. World Bank. Washington, DC.

Augustijn-Beckers, Ellen-Wien; Flacke, Johannes; Retsios, Bas (2011): Simulating informal settlement growth in Dar es Salaam, Tanzania: An agent-based housing model. In *Computers, Environment and Urban Systems* 35 (2), pp. 93–103. DOI: 10.1016/j.compenvurbsys.2011.01.001.

Barrington-Leigh, Christopher; Millard-Ball, Adam (2020): Global trends toward urban street-network sprawl. In *Proceedings of the National Academy of Sciences of the United States of America* 117 (4), pp. 1941–1950. DOI: 10.1073/pnas.1905232116.

Bertaud, Alain (2004): The Spatial Organization of Cities: Deliberate Outcome or Unforeseen Consequence? Background paper to World Development Report 2004. World Bank. Washington, DC.

Bhat, Chandra; Handy, Susan; Kockelman, Kara; Mahmassani, Hani; Chen, Qinglin; Weston, Lisa; University of Texas at Austin. Center for Transportation Research (2000): Development of an Urban Accessibility Index: Literature Review. University of Texas at Austin. Center for Transportation Research (Report No. TX-01/7-4938-1). Available online at https://rosap.ntl.bts.gov/view/dot/14882.

Boeing, Geoff (2017): OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. In *Computers, Environment and Urban Systems* 65, pp. 126–139. DOI: 10.1016/j.compenvurbsys.2017.05.004.

Borruso, Giuseppe; Porceddu, Andrea (2009): A Tale of Two Cities: Density Analysis of CBD on Two Midsize Urban Areas in Northeastern Italy. In Beniamino Murgante, Giuseppe Borruso, Alessandra Lapucci (Eds.): Geocomputation and Urban Planning, vol. 176. Berlin, Heidelberg: Springer Berlin Heidelberg (Studies in Computational Intelligence), pp. 37–56.

Brandily, P.; Rauch, F. G. (2018): Roads and urban growth. In *World Bank Working Paper* (859). Available online at https://collaboration.worldbank.org/content/usergenerated/asi/cloud/attachments/sites/collaboration-for-development/en/groups/research-partnership-for-sustainable-urban-development/documents/jcr:content/content/primary/blog/roads_and_urban_grow-bqme/3_roads_urban_growth.pdf.

Breheny, Michael (1996): Centrists, Decentrists and Compromisers: views on the future of urban form. In M. Jenks, Elizabeth Burton, Katie Williams (Eds.): The compact city. A sustainable urban form? / edited by Mike Jenks, Elizabeth Burton and Katie Williams. London: E & FN Spon, pp. 13–35.

Brown, Donald (2017): Challenging the conceptual boundaries of the compact city paradigm in sub-Saharan Africa: Towards Southern alternatives. In *DPU Working Paper* (187).

Bruegmann, Robert (2006): Sprawl. A compact history. 1st Pbk. ed. Chicago: University of Chicago Press.

Burchfield, M.; Overman, H. G.; Puga, D.; Turner, M. A. (2006): Causes of Sprawl: A Portrait from Space. In *Q J Econ* 121 (2), pp. 587–633. DOI: 10.1162/qjec.2006.121.2.587.

Cities Alliance (2022): Dynamics of Systems of Secondary Cities in Africa. Urbanization, Migration and Development. Cities Alliance. Brussels.

Cobbinah, P. B.; Amoako, Clifford (2014): Urban Sprawl and the Loss of Peri Urban Land in Kumasi, Ghana. In *International Journal of Humanities and Social Sciences* 8 (1), pp. 313–322. Available online at https://www.researchgate.net/profile/clifford-amoako/publication/232957200_urban_sprawl_and_the_loss_of_peri_urban_land_in_kumasi_ghana.

Colin Cameron, A.; Miller, Douglas L. (2015): A Practitioner's Guide to Cluster-Robust Inference. In *J. Human Resources* 50 (2), pp. 317–372. DOI: 10.3368/jhr.50.2.317.

Corbane, Christina; Florczyk, Aneta; Pesaresi, Martino; Politis, Panagiotis; Syrris, Vasileios (2018): GHS built-up grid, derived from Landsat, multitemporal (1975-1990-2000-2014), R2018A.

Duque, J. C.; Lozano Gracia, N.; Patino, J.; Restrepo, P. (2019): Urban Form and Productivity: What Is the Shape of Latin American Cities? In *World Bank Policy Research Working Paper* (8697).

Egidi, Gianluca; Cividino, Sirio; Vinci, Sabato; Sateriano, Adele; Salvia, Rosanna (2020): Towards Local Forms of Sprawl: A Brief Reflection on Mediterranean Urbanization. In *Sustainability* 12 (2), p. 582. DOI: 10.3390/su12020582.

European Space Agency (2018): EO4SD-Urban Project: Arusha City Report. GAF AG and Partners (No. 3.0).

Ewing, Reid; Pendall, Rolf; Chen, Don (2003): Measuring Sprawl and Its Transportation Impacts. In *Transportation Research Record* 1831 (1), pp. 175–183. DOI: 10.3141/1831-20.

Fleischmann, Martin; Feliciotti, Alessandra; Kerr, William (2021): Evolution of Urban Patterns: Urban Morphology as an Open Reproducible Data Science. In *Geogr Anal*. DOI: 10.1111/gean.12302.

Floater, Graham; Rode, Philipp; Robert, Alexis; Kennedy, Chris; Hoornweg, Dan; Slavcheva, Roxana; Godfrey, Nick (2014): Cities and the new climate economy: the transformative role of global urban growth. NCE Cities – Paper 01. London. Available online at http://eprints.lse.ac.uk/60775/1/NCE%20Cities%20Paper01.pdf.

Gattoni, George (2009): A case for the incremental housing process in sites-and-services programs and comments on a new initiative in Guyana. Inter-American Development Bank. Available online at http://web.mit.edu/incrementalhousing/articlesPhotographs/pdfs/IDB_Guyana_SS-Incremental.pdf.

Glaeser, Edward; Saiz, Albert (2003): The Rise of the Skilled City. Cambridge, MA (NBER WORKING PAPER SERIES, 10191).

Glaeser, Edward L. (2011): Triumph of the city. How our greatest invention makes us richer, smarter, greener, healthier, and happier. New York: Penguin Press.

Good, Phillip I. (2005): Increasing Computational Efficiency. In Phillip I. Good (Ed.): Permutation tests. A practical guide to resampling methods for testing hypotheses / Phillip Good. 3rd ed. New York: Springer (Springer Series in Statistics), pp. 233–254.

Harari, Mariaflavia (2020): Cities in Bad Shape: Urban Geometry in India. In *American Economic Review* 110 (8), pp. 2377–2421. DOI: 10.1257/aer.20171673.

Hervé Tchékoté; Chrétien Ngouanet (2015): Périurbanisation anarchique et problématique de l'aménagement du territoire dans le périurbain de Yaoundé. In Jan Bogaert, Jean-Marie Halleux (Eds.): Territoires périurbains. Développement, enjeux et perspectives dans les pays du Sud. Gembloux (Belgique): Les presses agronomiques de Gembloux, p. 12. Available online at https://www.researchgate.net/profile/tchekote-herve-

2/publication/312916710_periurbanisation_anarchique_et_problematique_de_l'amenagement_du_territoire_dans_le_periurbain_de_yaounde.

Hillier, Bill (1996): Space is the machine. A configurational theory of architecture / Bill Hillier. Cambridge: Cambridge University Press.

Huang, Chyi-Yun; Namangaya, Ally; Hassan Lugakingira, Mary Grace; Cantada, Isabel Margarita De Leon (2018): Translating Plans to Development : Impact and Effectiveness of Urban Planning in Tanzania Secondary Cities. World Bank Group. Washington, D.C. Available online at http://documents.worldbank.org/curated/en/300731546897829355/Translating-Plans-to-Development-Impact-and-Effectiveness-of-Urban-Planning-in-Tanzania-Secondary-Cities.

Jarah, Sivan; Zhou, Bo; Abdullah, Rebaz; Lu, Yawen; Yu, Wenting (2019): Urbanization and Urban Sprawl Issues in City Structure: A Case of the Sulaymaniah Iraqi Kurdistan Region. In *Sustainability* 11 (2), p. 485. DOI: 10.3390/su11020485.

Jenks, Mike; Burgess, Rod (2000): Compact cities. Sustainable urban forms for developing countries.

Karduni, Alireza; Kermanshah, Amirhassan; Derrible, Sybil (2016): A protocol to convert spatial polyline data to network formats and applications to world urban road networks. In *Scientific data* 3, p. 160046. DOI: 10.1038/sdata.2016.46.

Kelly, Morgan (2020): Direct standard errors for regressions with spatially autocorrelated residuals. In *UCD Centre for Economic Research Working Paper Series* (WP20/06). Available online at https://www.econstor.eu/handle/10419/228186.

Kombe, Wilbard Jackson (2005): Land use dynamics in peri-urban areas and their implications on the urban growth and form: the case of Dar es Salaam, Tanzania. In *Habitat International* 29 (1), pp. 113–135. DOI: 10.1016/S0197-3975(03)00076-6.

Lall, Somik; Lebrand, Mathilde; Park, Hogeun; Sturm, Daniel; Venables, Anthony (2021): Pancakes to Pyramids : City Form to Promote Sustainable Growth. World Bank. Washington, D.C. Available online at https://openknowledge.worldbank.org/handle/10986/35684.

Lall, Somik V.; Henderson, J. Vernon; Venables, Anthony (2017): Africa's cities. Opening doors to the world. Online-ausg. Washington, D.C: International Bank for Reconstruction and Development, The World Bank.

Lang, Wei; Hui, Eddie C.M.; Chen, Tingting; Li, Xun (2020): Understanding livable dense urban form for social activities in transit-oriented development through human-scale measurements. In *Habitat International* 104, p. 102238. DOI: 10.1016/j.habitatint.2020.102238.

Lee, Changyeon (2020): Metropolitan sprawl measurement and its impacts on commuting trips and road emissions. In *Transportation Research Part D: Transport and Environment* 82, p. 102329. DOI: 10.1016/j.trd.2020.102329.

Linard, Catherine; Tatem, Andrew J.; Gilbert, Marius (2013): Modelling spatial patterns of urban growth in Africa. In *Applied geography (Sevenoaks, England)* 44, pp. 23–32. DOI: 10.1016/j.apgeog.2013.07.009.

Lowry, John H.; Lowry, Michael B. (2014): Comparing spatial metrics that quantify urban form. In *Computers, Environment and Urban Systems* 44, pp. 59–67. DOI: 10.1016/j.compenvurbsys.2013.11.005.

Mahendra, Anjali; King, Robin; Du, Jillian; Dasgupta, Ani; Beard, Victoria; Kallergis, Achilles; Schalch, Kathleen (2021): Towards a More Equal City: Seven Transformations for More Equitable and Sustainable Cities. World Resources Institute. Washington, DC (World Resources Report, Towards a More Equal City).

Mattia Marconcini; Annekatrin Metz-Marconcini; Soner Üreyen; Daniela Palacios-Lopez; Wiebke Hanke; Felix Bachofer et al. (2020): World Settlement Footprint (WSF) 2015.

Mehmetoglu, Mehmet; Jakobsen, Tor G. (2017): Applied statistics using Stata. A guide for the social sciences / Mehmet Mehmetoglu, Tor Georg Jakobsen. 1st. Los Angeles: SAGE.

Mitas, L., Mitasova, H. (1999): Spatial Interpolation. In P. Longley, M. F. M.F. Goodchild, D. J. Maguire, D. W. Rhind (Eds.): Geographical information systems. Principles, techniques, management, and applications / edited by Paul A. Longley ... [et al.]. 1st ed. New York, Chichester: Wiley, pp. 481–492.

Mosammam, Hassan Mohammadian; Nia, Jamileh Tavakoli; Khani, Hadi; Teymouri, Asghar; Kazemi, Mohammad (2017): Monitoring land use change and measuring urban sprawl based on its spatial forms. In *The Egyptian Journal of Remote Sensing and Space Science* 20 (1), pp. 103–116. DOI: 10.1016/j.ejrs.2016.08.002.

Msuya, Ibrahim; Moshi, Irene; Levira, Francis (2020): Dar es Salaam: the unplanned urban sprawl threatening neighbourhood sustainability. Center for Sustainable Cities (SHLC) (SHLC RESEARCH SUMMARY, 11). Available online at http://www.centreforsustainablecities.ac.uk/research/dar-es-salaam-the-unplanned-urban-sprawl-threatening-neighbourhood-sustainability/.

Murphy, Raymond E.; Vance, J. E. (1954): Delimiting the CBD. In *Economic Geography* 30 (3), p. 189. DOI: 10.2307/141867.

OpenStreetMap contributors (2022): Planet dump 2022-06-27. Available online at https://planet.openstreetmap.org.

Porta, Sergio; Crucitti, Paolo; Latora, Vito (2006): The Network Analysis of Urban Streets: A Primal Approach. In *Environ Plann B Plann Des* 33 (5), pp. 705–725. DOI: 10.1068/b32045.

Pramanik, Suvamoy; Punia, Milap; Yu, Hanchen; Chakraborty, Saurav (2022): Is dense or sprawl growth more prone to heat-related health risks? Spatial regression-based study in Delhi, India. In *Sustainable Cities and Society* 81, p. 103808. DOI: 10.1016/j.scs.2022.103808.

Rahman, Md Hamidur; Islam, Md Hedaetul; Neema, Meher Nigar (2021): GIS-based compactness measurement of urban form at neighborhood scale: The case of Dhaka, Bangladesh. In *Journal of Urban Management*. DOI: 10.1016/j.jum.2021.08.005.

Schneider, Annemarie; Woodcock, Curtis E. (2008): Compact, Dispersed, Fragmented, Extensive? A Comparison of Urban Growth in Twenty-five Global Cities using Remotely Sensed Data, Pattern Metrics and Census Information. In *Urban Studies* 45 (3), pp. 659–692. DOI: 10.1177/0042098007087340.

Sevtsuk, Andres (2010): Path and Place: A Study of UrbanGeometry and Retail Activity in Cambridge and Somerville, MA. PhD Dissertation. Cambridge: MIT Department of Urban Studies and.

Sevtsuk, Andres; Mekonnen, Michael (2012): Urban network analysis. A new toolbox for ArcGIS. In *Revue internationale de géomatique* 22 (2), pp. 287–305. DOI: 10.3166/RIG.22.287-305.

Sheppard, Stephen (2010): Informal Sprawl? Williams College Working Paper (Williams College Working Paper). Available online at https://web.williams.edu/Economics/UrbanGrowth/Papers/InformalSprawl.pdf.

Sirko, Wojciech; Kashubin, Sergii; Ritter, Marvin; Annkah, Abigail; Bouchareb, Yasser Salah Eddine; Dauphin, Yann et al. (2021): Continental-Scale Building Detection from High Resolution Satellite Imagery. In *arXiv preprint* (arXiv:2107.12283). Available online at https://arxiv.org/pdf/2107.12283.

Speck, Jeff (2013): Walkable city. How downtown can save America, one step at a time. First paperback edition. New York: North Point Press a division of Farrar Straus and Giroux.

Tchekoté, Hervé; Ngouanet, Chrétien (2015): Périurbanisation anarchique et problématique de l'aménagement du territoire dans le périurbain de Yaoundé. In Jan Bogaert, Jean-Marie Halleux (Eds.): Territoires périurbains. Développement, enjeux et perspectives dans les pays du Sud. Développement, enjeux et perspectives dans les pays du Sud. Gembloux: Presses agronomiques de Gembloux (Liege University).

Tiecke, Tobias G.; Liu, Xianming; Zhang, Amy; Gros, Andreas; Li, Nan; Yetman, Gregory et al. (2017): Mapping the world population one building at a time. Available online at http://arxiv.org/pdf/1712.05839v1.

UN (2018a): World Urbanization Prospects: The 2018 Revision. Key Facts (Fact sheet). UN. Available online at https://www.un.org/development/desa/publications/2018-revision-of-world-urbanization-prospects.html#:~:text=Today%2C%2055%25%20of%20the%20world's,increase%20to%2068%25%20by%202050.

UN (2018b): World Urbanization Prospects: The 2018 Revision, Online Edition. United Nations, Department of Economic and Social Affairs, Population Division.

UNFPA (2007): Unleashing the potential of urban growth. United Nations Population Fund. New York. Available online at https://www.unfpa.org/sites/default/files/pub-pdf/695_filename_sowp2007_eng.pdf.

World Bank (2022a): Reconsidering Sites and Services : A Global Review. World Bank. Washington, DC. Available online at https://openknowledge.worldbank.org/handle/10986/37553.

World Bank (2022b): World Bank Data. World Bank. Available online at https://data.worldbank.org/.

WorldPop; Bondarenko, Maksym (2020): Individual Countries 1km Population Density (2000-2020).

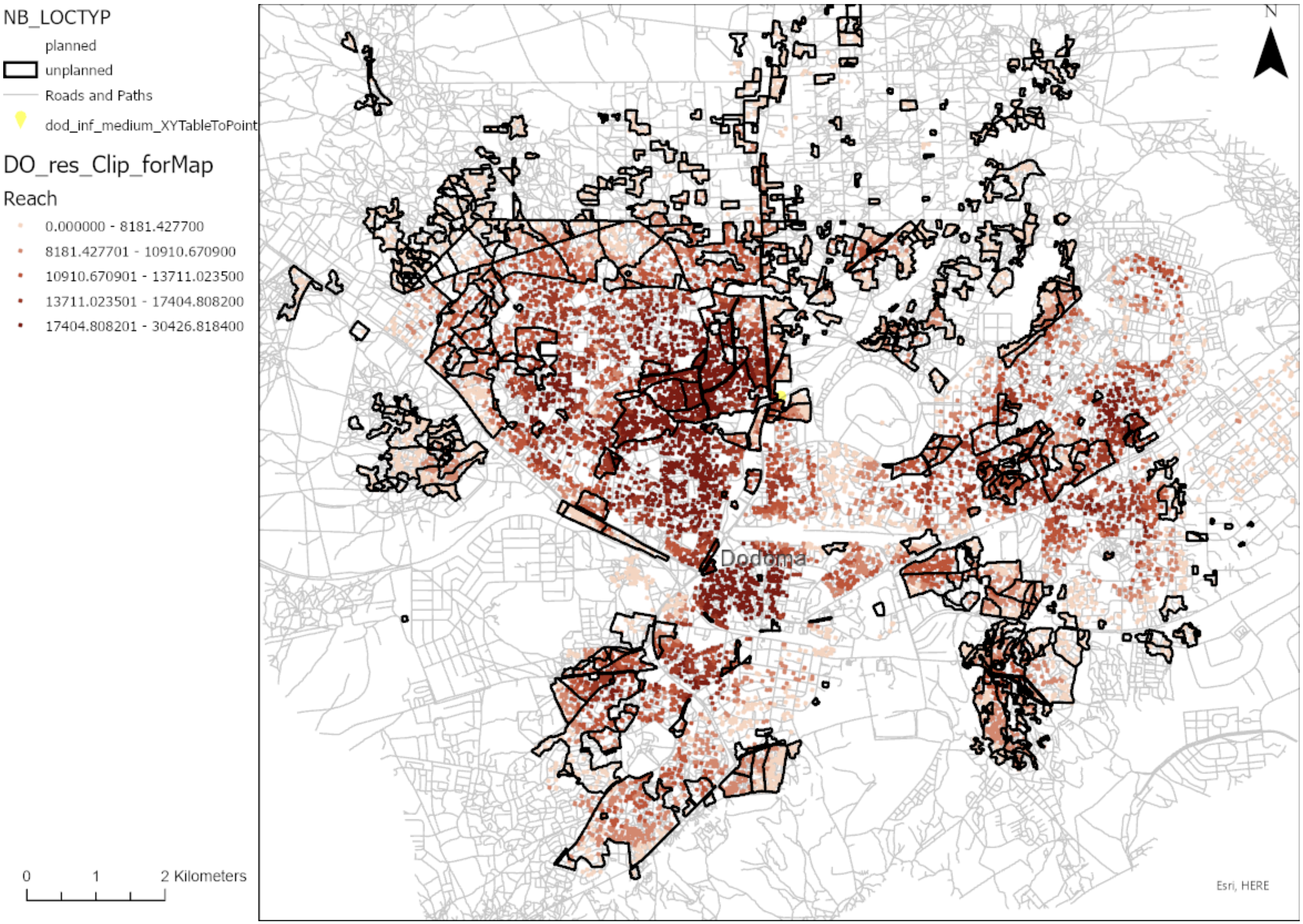
Yang, Duck-Hye; Goerge, Robert; Mullner, Ross (2006): Comparing GIS-based methods of measuring spatial accessibility to health services. In *Journal of medical systems* 30 (1), pp. 23–32. DOI: 10.1007/s10916-006-7400-5.

Yeboah, Godwin; Porto de Albuquerque, João; Troilo, Rafael; Tregonning, Grant; Perera, Shanaka; Ahmed, Syed A. K. Shifat et al. (2021): Analysis of OpenStreetMap Data Quality at Different Stages of a Participatory Mapping Process: Evidence from Slums in Africa and Asia. In *IJGI* 10 (4), p. 265. DOI: 10.3390/ijgi10040265.

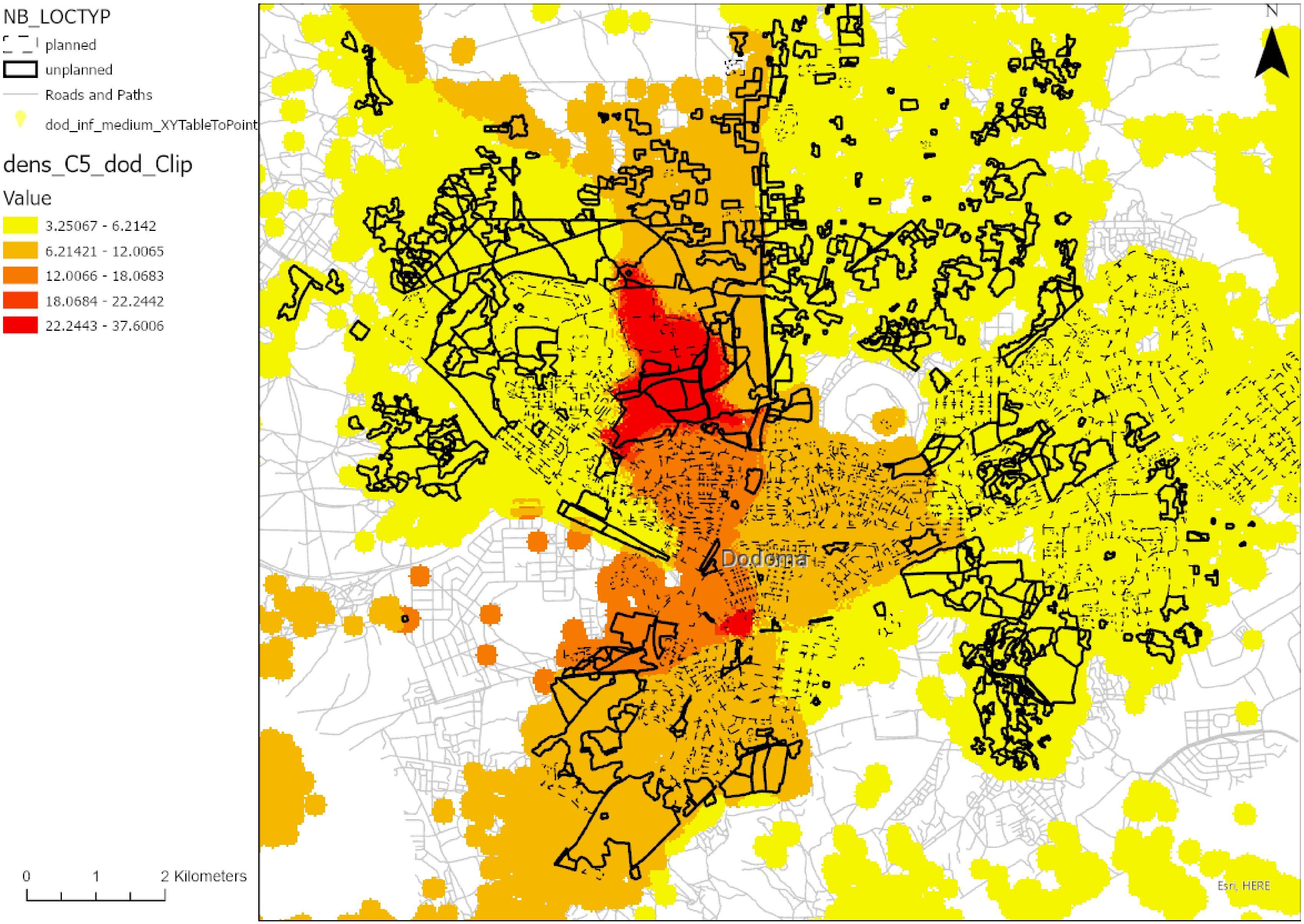
Zhang, Jiaqi; Shuang Chen, Sophia; Gao, Qun; Shen, Qiushi; Kimirei, Ismael Aaron; Mapunda, Damas William (2020): Morphological Characteristics of Informal Settlements and Strategic Suggestions for Urban Sustainable Development in Tanzania: Dar es Salaam, Mwanza, and Kigoma. In *Sustainability* 12 (9), p. 3807. DOI: 10.3390/su12093807.

Annex I: Maps

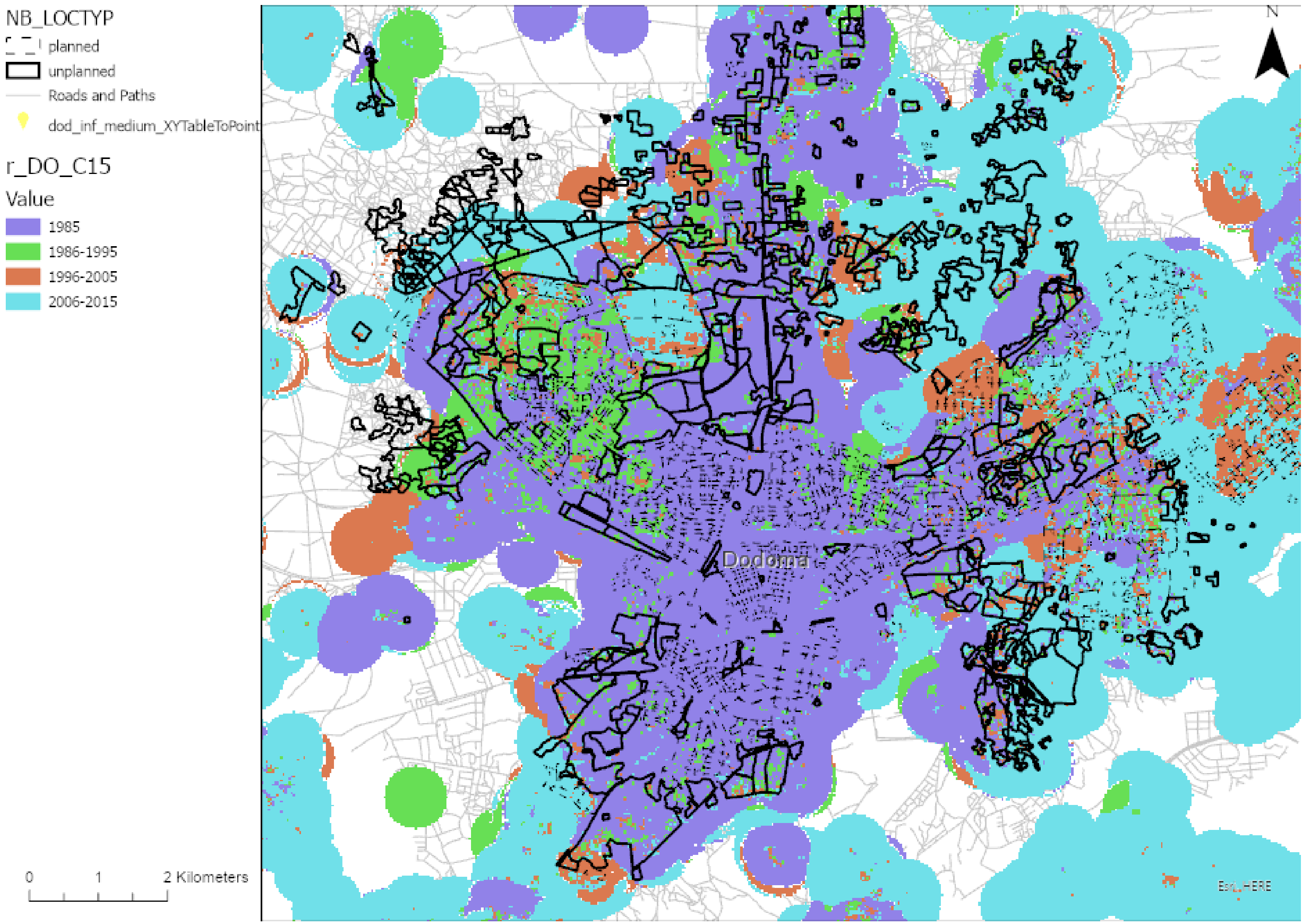
Map 1: Reach calculations for buildings in planned and unplanned areas in Dodoma, Tanzania



Map 2: Density values (people per 30x30m grid cell) for Dodoma. Data source: Facebook 2018 Data for Good population density



Map 3: Built-up year categories for study area in Dodoma. Data source: World Settlement Footprint evolution 2015



Annex II: Regression results individual city level

Table 8: OLS results DV reach for each city individually

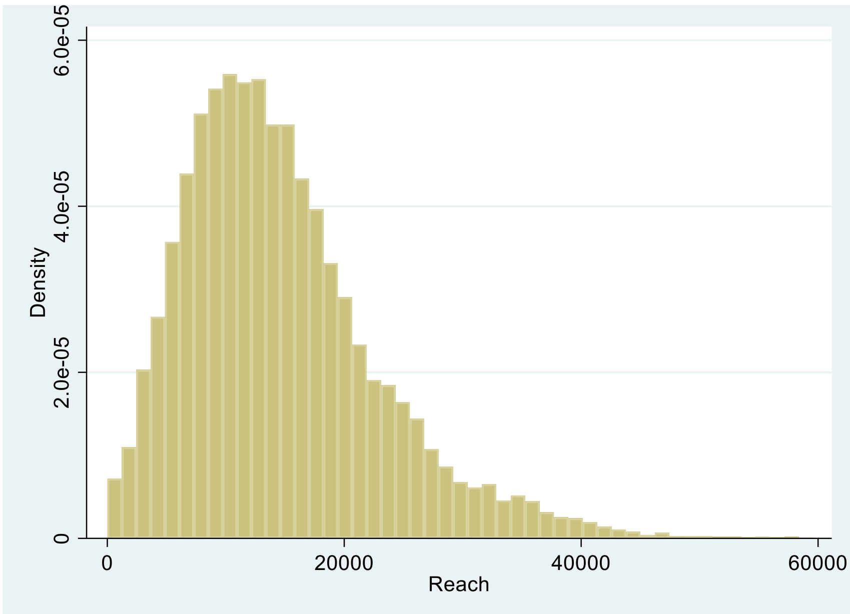
	Arusha Reach	Dodoma Reach	Kigoma Reach	Mbeya Reach	Mtwara Reach	Mwanza Reach	Tanga Reach
Distance to center	-684.3***	-493.1***	-454.9***	-471.7***	-1449.7***	559.5***	-1113.0***
	(-5.39)	(-4.45)	(-5.10)	(-4.34)	(-13.11)	(8.67)	(-6.56)
Planning type: Unplanned	325.5	-1908.8***	-2927.7***	3518.3***	-1243.7***	-4999.9***	-2829.7***
	(0.50)	(-4.90)	(-8.36)	(5.53)	(-3.94)	(-18.43)	(-5.40)
log_density	-64838.7***	42028.4	-173546.6**	-5213.1	90844.9***	-43308.2***	32998.5
	(-4.65)	(1.61)	(-3.01)	(-0.39)	(10.78)	(-6.57)	(1.19)
log_dens^2	24756.0***	-23354.8	74482.1**	3085.9	-53311.1***	14538.3***	-15464.4
	(4.43)	(-1.84)	(2.99)	(0.51)	(-10.70)	(6.63)	(-1.29)
log_dens^3	-2853.0***	4249.5*	-10182.4**	-225.2	9853.1***	-1502.7***	2675.9
	(-3.98)	(2.18)	(-2.90)	(-0.26)	(10.77)	(-6.52)	(1.59)
Built-up Year							
1986-1995	-421.9	-108.9	-544.4	344.7	-1602.4***	-2072.6***	-628.0*
	(-1.16)	(-0.34)	(-1.07)	(0.69)	(-5.79)	(-8.10)	(-2.01)
1996-2005	-2148.7***	-898.6**	-2346.6***	-5013.5***	-1105.6***	-2763.4***	-2845.4***
	(-5.91)	(-2.84)	(-4.17)	(-11.77)	(-4.28)	(-9.92)	(-5.98)
2006-2015	-2649.5***	-2042.5***	-3063.4***	-9866.4***	-2261.2***	-3828.4***	-2052.8***
	(-5.52)	(-5.52)	(-4.84)	(-20.98)	(-7.45)	(-11.22)	(-4.22)
2016+	-8118.1***	-5415.6***	-5698.0***	-15072.7***	-5054.7***	-8201.0***	-8836.8***
	(-11.70)	(-9.72)	(-4.22)	(-15.09)	(-9.69)	(-3.95)	(-3.68)
Constant	66672.1***	-10318.1	141704.8**	18846.3*	-32566.1***	51125.9***	-6772.3
	(5.96)	(-0.60)	(3.24)	(2.14)	(-7.45)	(8.41)	(-0.33)
Observations	23,010	12,145	7,558	24,613	4,785	18,914	6,978
R2	0.30	0.45	0.40	0.26	0.59	0.19	0.38

t statistics in parentheses
* p<0.05, ** p<0.01, *** p<0.001

Annex III: Data Diagnostics

An important assumption of linear regression is that the values of the dependent variable are close to normally distributed (Mehmetoglu and Jakobsen 2017). For the reach output it can be said that this assumption is met, despite a slight skewedness (Figure 15).

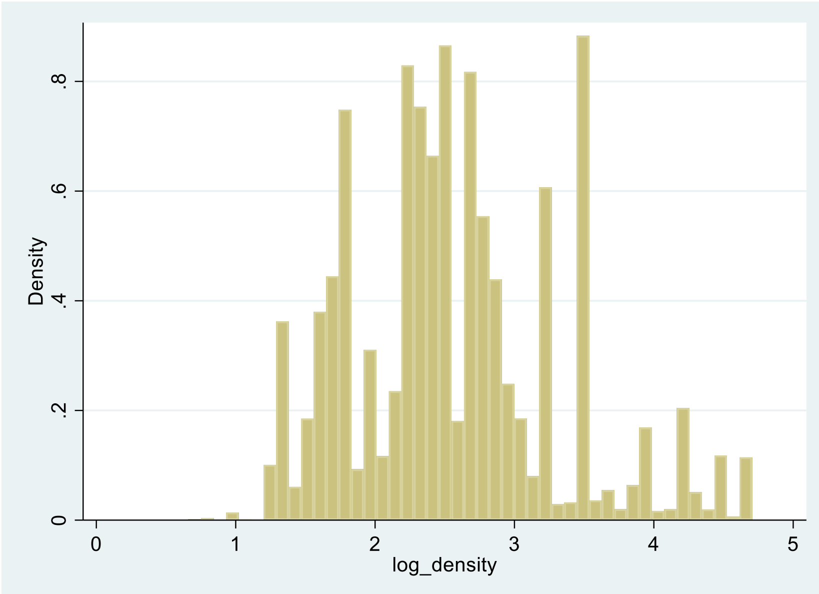
Figure 15: Count distribution of reach values for full sample.



The distribution of density could not be found to be evenly and normally distributed. A log transformation was therefore performed to approximate normality. Values of zero density were interpolated or omitted (n>300) (

Figure 16).

Figure 16: Count distribution of log_density



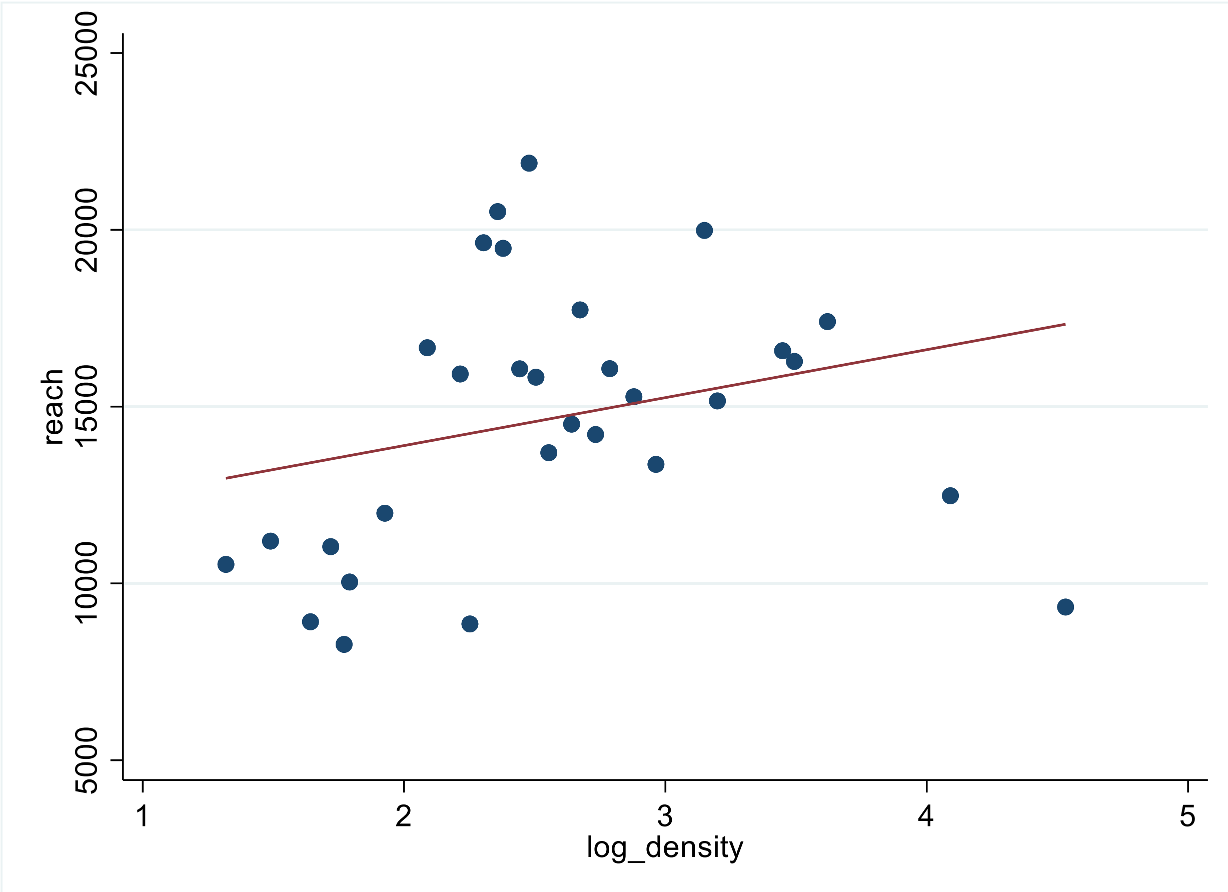
General correlation between variables is found to be rather low looking at the full sample (Figure 17). This changes however, when looking at cities individually, see Table 5.

Figure 17: Correlation matrix

	reach	un_pla~d	bu_year	dens_c5	near_d~t	city
reach	1.0000					
un_planned	-0.0452	1.0000				
bu_year	-0.2496	0.1590	1.0000			
dens_c5	-0.0119	0.0445	-0.2368	1.0000		
near_dist	-0.2602	0.2510	0.2691	-0.2372	1.0000	
city	0.0575	-0.1453	-0.2340	0.3249	0.0029	1.0000

The association between density and reach has been found to be slightly curvilinear. This has been tested by repeatedly adding polynomials of log_density to the regression model until their coefficients were rendered insignificant (at log_density to the power of 4). The *binscatter* sketching out the bends in the association of log_density and reach is presented in Figure 18.

Figure 18: Binscatter of log_density and reach describing a curvilinear relationship



Annex IV: Spatial Analysis Workflow Documentation and .do-file

Software used:

ArcMap 10.5.1
ArcGIS Pro 2.8.3
STATA SE 17
Microsoft Excel

Data download:

1. Informal layers from: respective City pages in the World Bank Data Catalogue, such as https://datacatalog.worldbank.org/search/dataset/0039353/Dodoma--Tanzania----Planned-and-Unplanned-Settlement-Areas--ESA-EO4SD-Urban-
2. OSM Roads from Geofabrik (https://download.geofabrik.de/africa/tanzania.html)
3. Google Buildings from https://sites.research.google/open-buildings/
4. Built-Up:
    1. Download WSF evolution. Tiles are named by coordinates of lower left corner
    - https://download.geoservice.dlr.de/WSF_EVO/files/
    - Meta: https://samapriya.github.io/awesome-gee-community-datasets/projects/wsf/

Density from https://dataforgood.facebook.com/dfg/tools/high-resolution-population-density-maps

ArcPro Data Preparation:

1. Informal 2016 layer buffer (dissolve) of 2km for every city (Batch Buffer)
    a. Output: 7 AoI buffers
    b. Merge 7 AoI buffers into 1 layer ' Buffer_2k_Workflow_EO4SD_7Cit'

2. Import Roads 'geofab TZ Road network'
3. Clip roads with the combined AoI buffer
    a. Output 1 raw road layer 'cit7_roads_clip'

4. Import google buildings from .tsv
5. Clip buildings with the 1 AoI buffer
    a. Output 7 urban-building layers
6. Subset features, create 10% sample for building layers

ArcMap Network Set-up:

1. Open a project, change projection to UTM 36S (right click on map)
2. Import raw road network - transform request
3. Create GISF2E nodes and edges - check projected CGS
4. Import sample Buildings - transform request (test for even smaller sample)

5. Create Network Dataset with roads
    a. Create .gdb , create feature dataset

b. Build Network dataset

6. Run Urban Network Analyst Tool Centrality (Sevtsuk and Mekonnen, 2012)
7. Export calculated reach values for ArcGis Pro

ArcPro Analysis:

1. Import reach buildings
2. Spatial Join with Informality layers

Workflow for smoothing controls (so that every building gets assigned a value from raster when using "Extract Values to Points":

Built-up:

1. Clip city extents from rasters
2. Filling in missing values w code in raster calculator

```
##filled = arcpy.sa.Con(arcpy.sa.IsNull(in_raster),arcpy.sa.FocalStatistics(in_raster,
            arcpy.sa.NbrRectangle(w, h),'MEAN'), in_raster)
```

Density:

1. Directly apply raster calculator mean "MEAN" Focal Statistics with below formular. No "Set Null" needed.

```
##filled = arcpy.sa.Con(arcpy.sa.IsNull( "Rasters\TZ_dens"),arcpy.sa.FocalStatistics( "Rasters\TZ_dens",
          arcpy.sa.NbrCircle(5, 'CELL'),'MEAN'),  "Rasters\TZ_dens")
```

Joining Rasters to produce final table:

1. Merge individual cities layers for a) Reach results and b) EO4SD informal
2. Spatial join of a and b to the reach points
3. "Extract Multi Values to Points" to join values of all built-up rasters + density raster in 1 go
4. Copy attribute table of final points layer that should contain reach, NB_localtype, built-up year, city, density, IDs (landuse polygons).

Adding Distance to city center:

1. City center points were determined by visual inspection of density of banks. Distance to respective center points were calculated using "Near" tool in ArcGis Pro, i.e. each building was matched to the respective city center point and the distance calculated.
2. Distance values were joined to Dataset in post

Data export:

1. Final dataset was exported to csv. for analyis in STATA
2. For STATA analysis see .do-file

STATA .do-file

```
*Install
*estat
*binscatter

clear all
cd C:\THESISDATA\STATA

*cleaning
import delimited "Results_6-29", varnames(1) rowrange(2)
*use CREATED dta-file
use "C:\THESISDATA\STATA\7cities.dta"

*### Importing Distance values and convert to km2
merge 1:1 objectid using Distance_values_matching_OID.dta, keepusing
(near_fid near_dist)
drop if _merge==2
gen distkm2 = near_dist/1000

global finalx distcat i.un_planned log_density log_dens2 log_dens3 i.bu_year
i.city
global finalxbycity distcat i.un_planned log_density log_dens2 log_dens3
i.bu_year

*#########################################################
* DATA PREP
generate un_planned = 1 if nb_loctyp=="unplanned"
replace un_planned = 0 if nb_loctyp=="planned"

*create land-use polygon ID
destring(id), replace
egen lu_id = group(city id)

drop if missing(un_planned)

*cleaning bu year raster outputs
replace r_mb_c15 = "0" if r_mb_c15=="<Null>"
replace r_mw_c15 = "0" if r_mw_c15=="<Null>"
replace r_mt_c15 = "0" if r_mt_c15=="<Null>"
replace r_ki_c15 = "0" if r_ki_c15=="<Null>"
replace r_ar_c15 = "0" if r_ar_c15=="<Null>"
replace r_ta_c15 = "0" if r_ta_c15=="<Null>"
replace r_do_c15 = "0" if r_do_c15=="<Null>"
```

```
destring(r_mb_c15 r_mw_c15 r_mt_c15 r_ki_c15 r_ar_c15 r_ta_c15 r_do_c15),
replace

gen bu_year = max(r_mb_c15, r_mw_c15, r_mt_c15, r_ki_c15, r_ar_c15,
r_ta_c15, r_do_c15)

*replacing missing data with value for after 2015-built
replace bu_year=5 if bu_year==0

*destring density and replace 0s with average of land-use ID

replace dens_c5 = "0" if dens_c5=="<Null>"
destring(dens_c5), replace

egen missdens = mean(dens_c5),by(lu_id)
replace dens_c5 = missdens if dens_c5==0
gen log_density = ln(dens_c5)

*reach cleaning
drop if reach == 0
gen log_reach = ln(reach)

* categorize distance
egen distcat = cut(distkm2), at(0,1,2,3,4,5,6,8,9,10,11,12,13,14)

*#################################
*Regression
* mean comparison
reg reach un_planned
reg reach un_planned i.city

*link of density and informal
reg log_density i.un_planned i.city
logistic un_planned log_density i.city

*##reg/logit with cluster
reg reach i.un_planned log_density i.bu_year, cluster(lu_id)
* city fixed effects
reg reach i.un_planned log_density i.bu_year i.city, cluster(lu_id)
* density interaction
reg reach i.un_planned log_density log_dens2 i.bu_year i.city,
cluster(lu_id)

     * --> dens^4 insifignicant ! reg reach i.un_planned log_density
log_dens2 log_dens3 log_dens4 i.bu_year i.city, cluster(lu_id)
```

```
logit un_planned log_reach log_density i.bu_year i.city, cluster(lu_id)
logistic un_planned reach log_density i.bu_year i.city, cluster(lu_id)

*## reg/logit with log_reach
reg log_reach i.un_planned log_density i.bu_year i.city, cluster(lu_id)
logistic un_planned log_reach log_density i.bu_year i.city, cluster(lu_id)

*##reg without cluster
reg log_reach i.un_planned

*obtaining residuals
predict resid_reach, residuals
hist resid_reach
sum resid_reach

*################
* clustered reg for each city individually
reg reach i.un_planned log_density i.bu_year if city==1, cluster(lu_id)
reg reach i.un_planned log_density i.bu_year if city==2, cluster(lu_id)
reg reach i.un_planned log_density i.bu_year if city==3, cluster(lu_id)
reg reach i.un_planned log_density i.bu_year if city==4, cluster(lu_id)
reg reach i.un_planned log_density i.bu_year if city==5, cluster(lu_id)
reg reach i.un_planned log_density i.bu_year if city==6, cluster(lu_id)
reg reach i.un_planned log_density i.bu_year if city==7, cluster(lu_id)
eststo
estout using finalreg1.txt, label replace

*city individually with final x spec
quietly reg reach $finalxbycity if city==1, cluster(lu_id)
quietly reg reach $finalxbycity if city==2, cluster(lu_id)
quietly reg reach $finalxbycity if city==3, cluster(lu_id)
quietly reg reach $finalxbycity if city==4, cluster(lu_id)
quietly reg reach $finalxbycity if city==5, cluster(lu_id)
quietly reg reach $finalxbycity if city==6, cluster(lu_id)
quietly reg reach $finalxbycity if city==7, cluster(lu_id)
eststo
esttab using fin_reg_bycity.txt, label replace

* reg for city sub-samples
* high coefficient cities
reg reach distkm2 i.un_planned log_density log_dens2 log_dens3 i.bu_year
i.city if city==3|city==6|city==7, cluster(lu_id)
*rech ffor low coefficient cities
```

```
reg reach distkm2 i.un_planned log_density log_dens2 log_dens3 i.bu_year i.city if city==1|city==2|city==4|city==5, cluster(lu_id)
*reg for all but Mbeya
reg reach distkm2 i.un_planned log_density log_dens2 log_dens3 i.bu_year i.city if city!=4, cluster(lu_id)

* reg with categorcial distance
reg reach distcat i.un_planned log_density log_dens2 log_dens3 i.bu_year i.city, cluster(lu_id)
eststo

* high coefficient cities (Kigoma, Mtwara, Mwanza, Tanga)
reg reach distcat i.un_planned log_density log_dens2 log_dens3 i.bu_year i.city if city==3|city==5|city==6|city==7, cluster(lu_id)
eststo
*reg for low coefficient cities (Arusha, Dodoma, Mbeya)
reg reach distcat i.un_planned log_density log_dens2 log_dens3 i.bu_year i.city if city==1|city==2|city==4, cluster(lu_id)
eststo

*marginsplot after this to show unplanned difference by distance
margins, at (distcat=(0(1)13) un_planned=(0 1))
marginsplot

*margins for high/low coefficient cities
** --> redo reg for high, then:
margins, at (distcat=(0(1)13) un_planned=(0 1))
marginsplot, xtitle("Distance from center (km)") ytitle("Predicted reach (m)")

*###############
*exploring linear vs. curvilinear relationship reach density
reg reach log_density
gen log_dens2=log_density*log_density
gen log_dens3=log_density*log_density*log_density
gen log_dens4=log_density*log_density*log_density*log_density
gen log_dens5=log_density*log_density*log_density*log_density*log_density
gen log_dens6=log_density^6
gen log_dens7=log_density^7

reg reach log_density log_dens2
reg reach log_density log_dens2 log_dens3
reg reach log_density log_dens2 log_dens3 log_dens4
reg reach log_density log_dens2 log_dens3 log_dens4 log_dens5
reg reach log_density log_dens2 log_dens3 log_dens4 log_dens5 log_dens6
```

```
reg reach log_density log_dens2 log_dens3 log_dens4 log_dens5 log_dens6
log_dens7

*################
* descriptive stat

sum reach dens_c5
sum reach distkm2 dens_c5 bu_year un_planned confidence
tab city
tab bu_year

table () un_planned, statistic(mean reach)
table city un_planned, statistic(mean reach)
table city un_planned, statistic(total area_in_me)
table city un_planned, statistic(percent area_in_me)

* dens reach scatter by informal
twoway (scatter reach log_density if un_planned==0, mcolor(red)) ///
       (scatter reach log_density if un_planned==1, mcolor(blue))
* distance reach scatter by informal
twoway (scatter reach near_dist if un_planned==0, mcolor(red)) ///
       (scatter reach near_dist if un_planned==1, mcolor(blue)) if city!=4

corr dens_c5 reach
*binscatter for dens^3 justification
binscatter reach log_density, nquantiles(30)

* density gradient, w Mwanza extra
twoway (bar dens_c5 near_dist) if city!=6, by(city)
twoway (bar dens_c5 near_dist) if city==6, by(city)
```