Child Mortality and Access to Cities

A Geospatial Analysis

Will Merrow
INFO 609
Final Project
5/14/20

INTRODUCTION

Child mortality, defined as the probability of a child dying before reaching age five, has long been a focus of global development and forms a core part of the UN's Sustainable Development Goals. In 2016, 5.6 million children died before their fifth birthday. This represents a significant decline from previous years.

This report uses granular geospatial data to map the **relationship between child mortality and access to cities**, an important correlate of child mortality given its relationship with income, access to healthcare, and other socioeconomic variables.

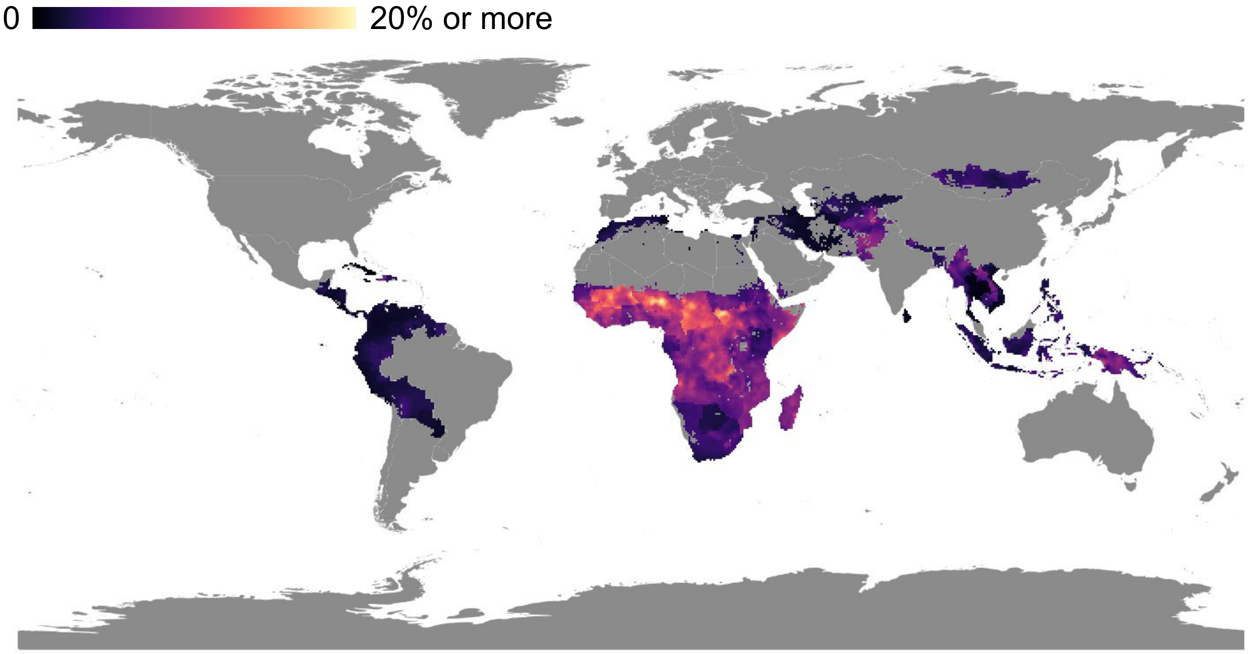
This analysis takes a global scope, examining child mortality in 99 low- and middle-income countries, and focusing on particular countries to look at sub-national trends.

PRIMARY DATA

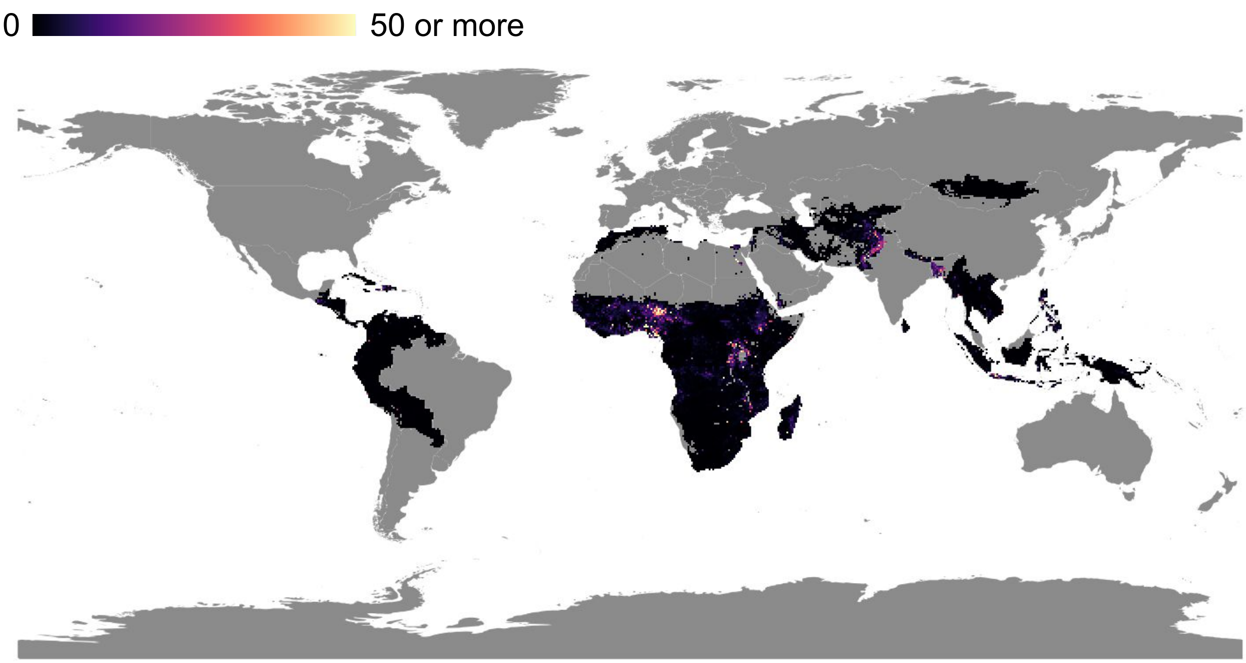
The child mortality data comes from the Institute for Health Metrics and Evaluation, and is summarized and visualized on their website: http://www.healthdata.org/research-article/mapping-123-million-neonatal-infant-and-child-deaths-between-2000-and-2017

The data includes total deaths as well as rates for the years 2000 to 2017. The data is available as gridded data at 5km resolution, as well as aggregated to the country level, first administrative level, and second administrative level. My analysis uses the gridded 5km data.

Under-5 mortality rates, 2017



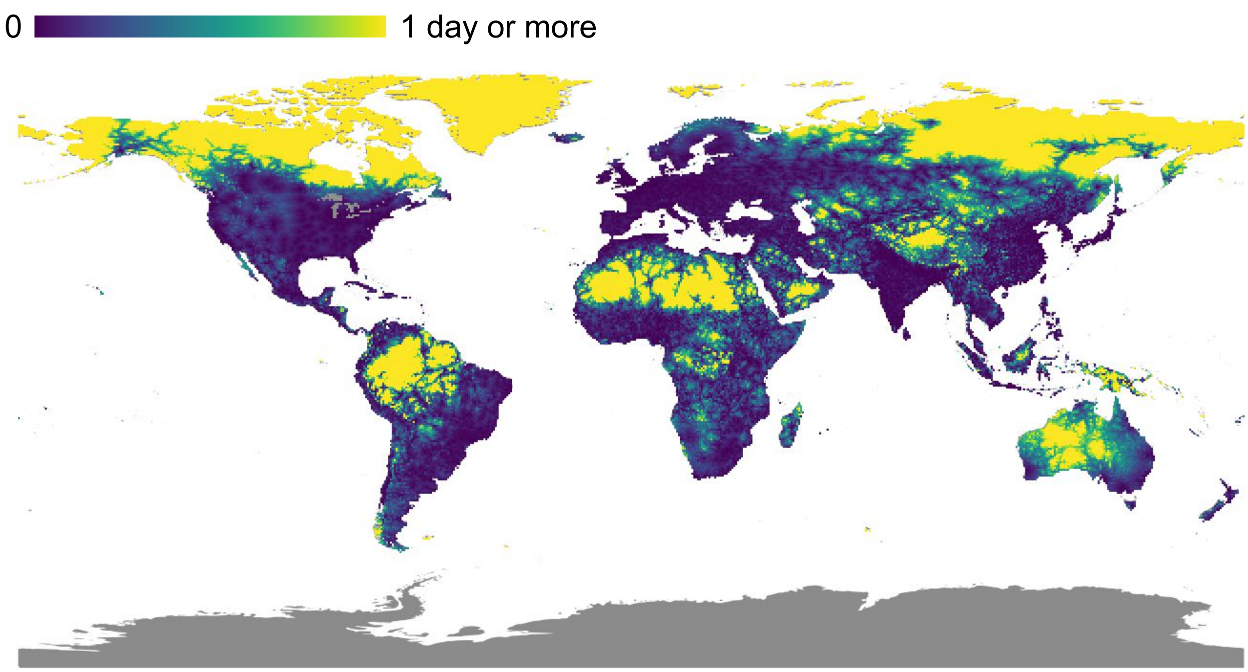
Under-5 deaths, 2017



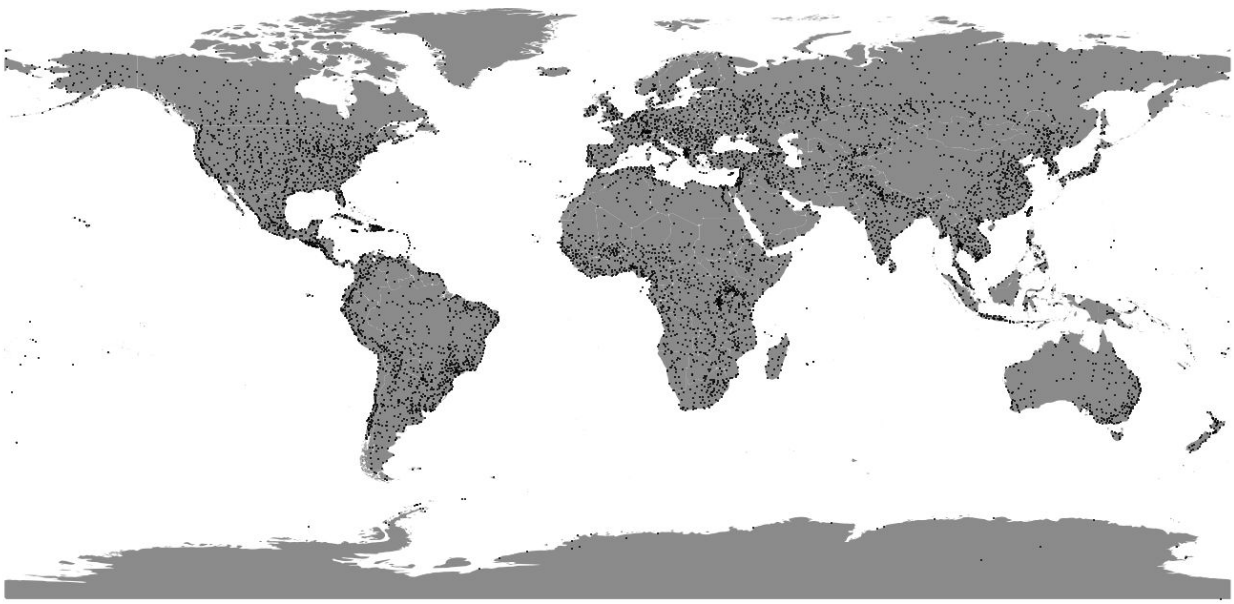
The data for access to cities comes from the Malaria Atlas Project, and maps the travel time using surface transportation to the nearest city in 2015. The data is gridded at 1km resolution for all countries.

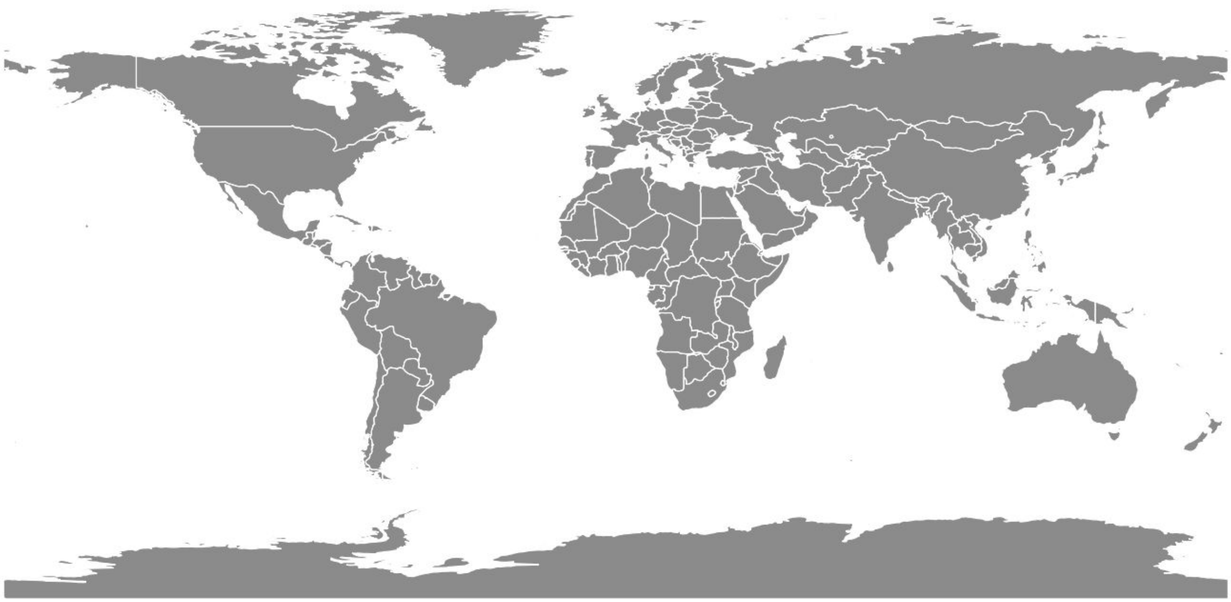
The dataset is based off of the Google Roads database and Open Street Map. Travel time is far more useful than simply using distance to cities, because the distribution of people and transportation around urban centers is uneven, and because of differences in the quality of transportation infrastructure. A city is defined as an urbanization of 50,000 people or more.

Access to cities, 2015



Other datasets used in my analysis include populated places and country borders, both from Natural Earth:





ANALYSIS

There are many possible ways to investigate the correlation between child mortality and access to cities. I am most interested in looking at this relationship within individual countries, and at the most granular extent possible. To do this, I overlay data for access to cities (and other datasets) on top of a choropleth of child mortality data, in the following steps.

1. **Align raster files**

    We must first downscale the access to cities data from 1km to 5km resolution using the **align rasters** tool. I use “average” as the method of aggregation in order to include all 25 of the input cells.

    This produces values of -9999 where there was previously no data (e.g. cells representing water). To change these values back to no data, we can use the **raster calculator**, using conditional statements to change all negative values to no data, and keep all other values the same:

```
("Access_to_cities_2015_5km@1" >= 0) / ("Access_to_cities_2015_5km@1" >= 0) *
"Access_to_cities_2015_5km@1"
```

    We can confirm the alignment worked successfully by overlaying the child mortality data (purple) over the access data (blue) and zooming in on a border area along the edge of the child mortality data:

Before: not aligned (1km and 5km)

After: aligned (5km)

2. Create categories for different levels of access to cities

In order to simplify the access to cities data for easier analysis, I segment the data into six categories based on time to city. In my final presentation, I delineate only one category at a time.

- 5 minutes or less
- 6-60 minutes
- 61-360 minutes
- 361-720 minutes
- 721-1440 minutes
- Greater than 1440 minutes

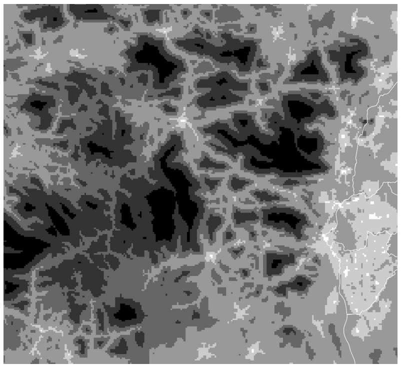
I accomplish this using the **raster calculator**, first creating new raster layers where cells are either 0 or 1:

```
"Access_to_cities_2015_5km@1" <= 5
("Access_to_cities_2015_5km@1" > 5) AND ("Access_to_cities_2015_5km@1" <= 60)
("Access_to_cities_2015_5km@1" > 60) AND ("Access_to_cities_2015_5km@1" <= 360)
("Access_to_cities_2015_5km@1" > 360) AND ("Access_to_cities_2015_5km@1" <= 720)
("Access_to_cities_2015_5km@1" > 720) AND ("Access_to_cities_2015_5km@1" <= 1440)
"Access_to_cities_2015_5km@1" > 1440
```

I then use the raster calculator to combine these into one layer with values of 1, 2, 3, 4, 5, and 6:

```
"Access_5km-Urban-5@1" * 6 + "Access_5km-Suburban-5to60@1" * 5 +
"Access_5km-Rural-60to360@1" * 4 + "Access_5km-Rural-360to720@1" * 3 +
"Access_5km-Rural-720to1440@1" * 2 + "Access_5km-Rural-1440@1" * 1
```

This is the result:



3. **Create contour layer**

   Because access to cities is a continuous variable, we can visualize it using contour lines. This would be possible to do directly from the original raster file, although due to computer processing time, I do this on the categorized raster layer as a way of drawing lines in between the six different access categories. I don't use all categories in my final presentation, although all are useful in analyzing the data.

4. **Filter cities**

   I then filter the cities to only those used in creating the original access to cities data. Although the data does not include a list of these cities, we can approximate this by filtering a separate list of cities to just those that are located at the "0" cells from the access data. I use the list of populated places from Natural Earth, which includes city locations as points, with associated data such as city name. We could filter by population, but filtering by location will provide a more accurate approximation of the author's original list given their data file is from 2015. This involves a few steps:
   - Create a new raster layer of the 1km level access to cities data filtered to just those cells with a value of zero (cells that are presumably cities), again using the **raster calculator**.
   - **Polygonize** the layer.
   - Add a **buffer** of 3km.
   - Filter the Natural Earth shapefile of cities to just those within the resulting areas using **select by location**.

5. **Calculate change in child mortality over time**

   In order to visualize change over time, we can calculate the percentage point change in the child mortality rate from 2000 to 2015 using the raster calculator to subtract the 2000 band from the 2015 band.

6. **Calculate population**

   We can calculate total population using the raster calculator, using the data for child mortality rates and total child deaths.

7. **Create a table of raster cells**

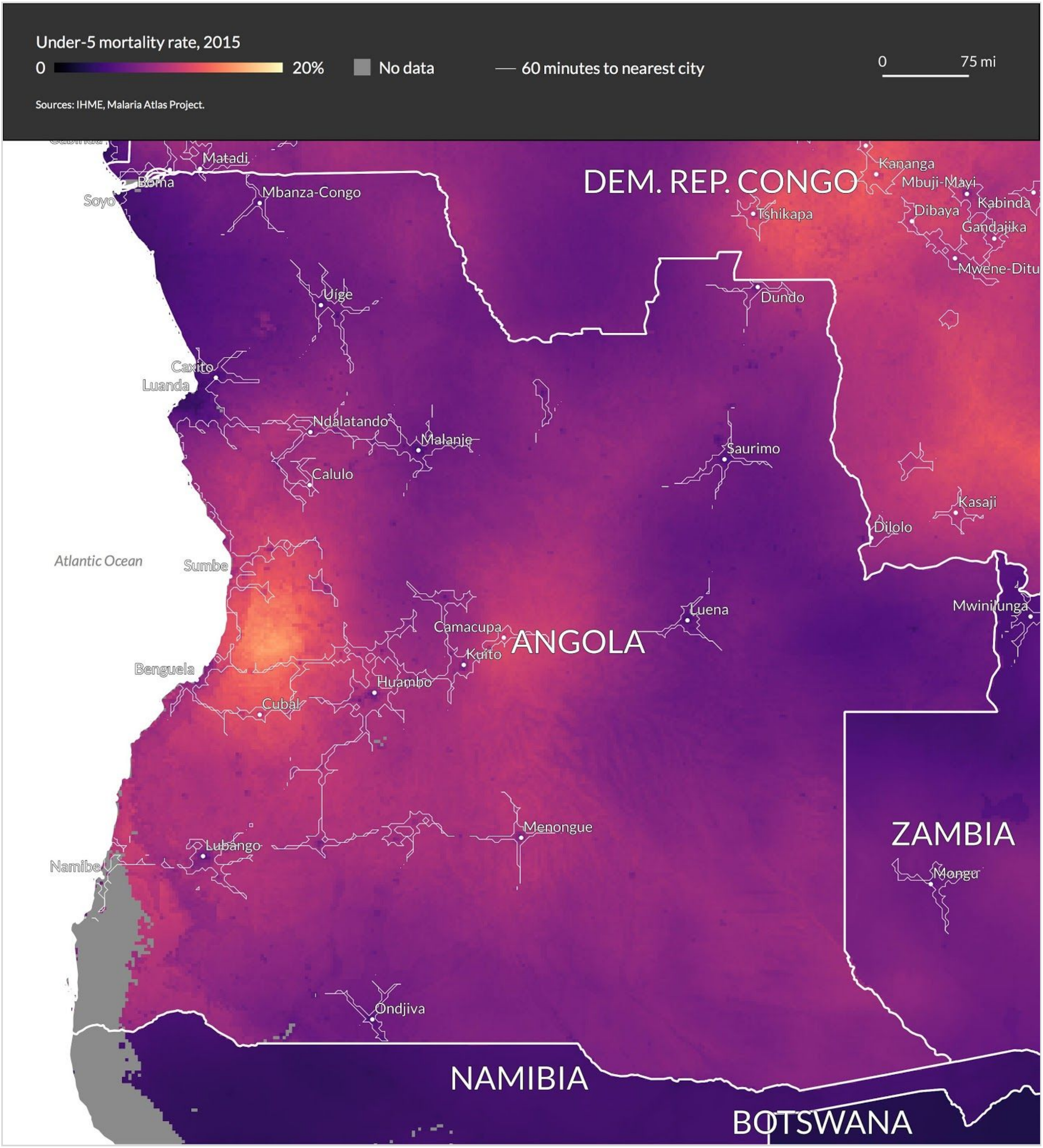
   To conduct further analysis, we can bring the raster data into R and convert it to a table where each row is a cell, and each column is a raster layer. Because this involves millions of points, this is not possible in Excel. R code is included in the appendix.

   The result is a dataframe with a row for each 5km cell and columns for access to cities, child mortality rate, and child deaths, which allows for further quantitative analysis of the relationship between the variables.

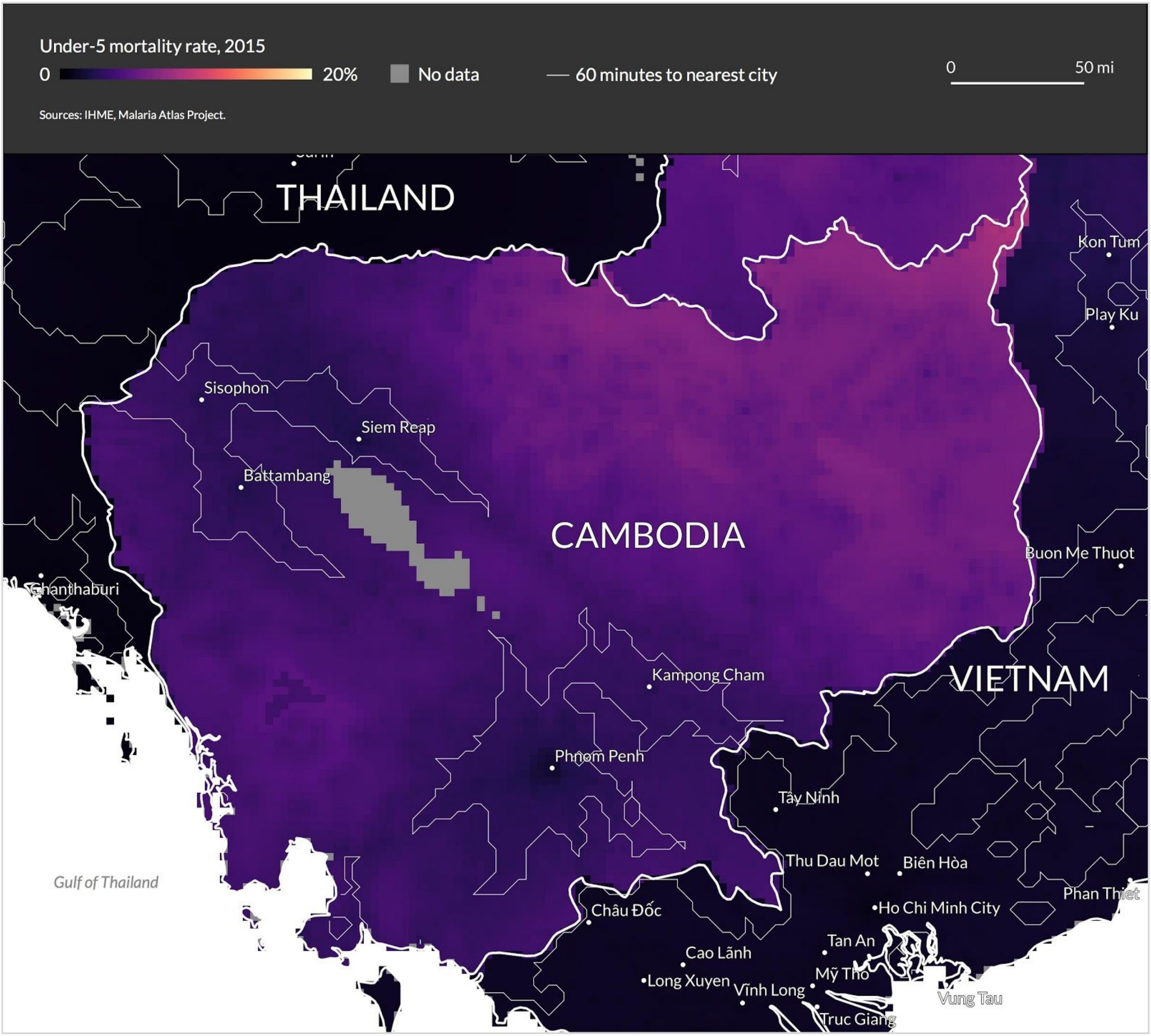
RESULTS

We can overlay the contour lines for access to cities to analyze the correlation with child mortality in 2015. For readability, I overlay one contour line at a time, starting with an access level of 60 minutes time to the nearest city.

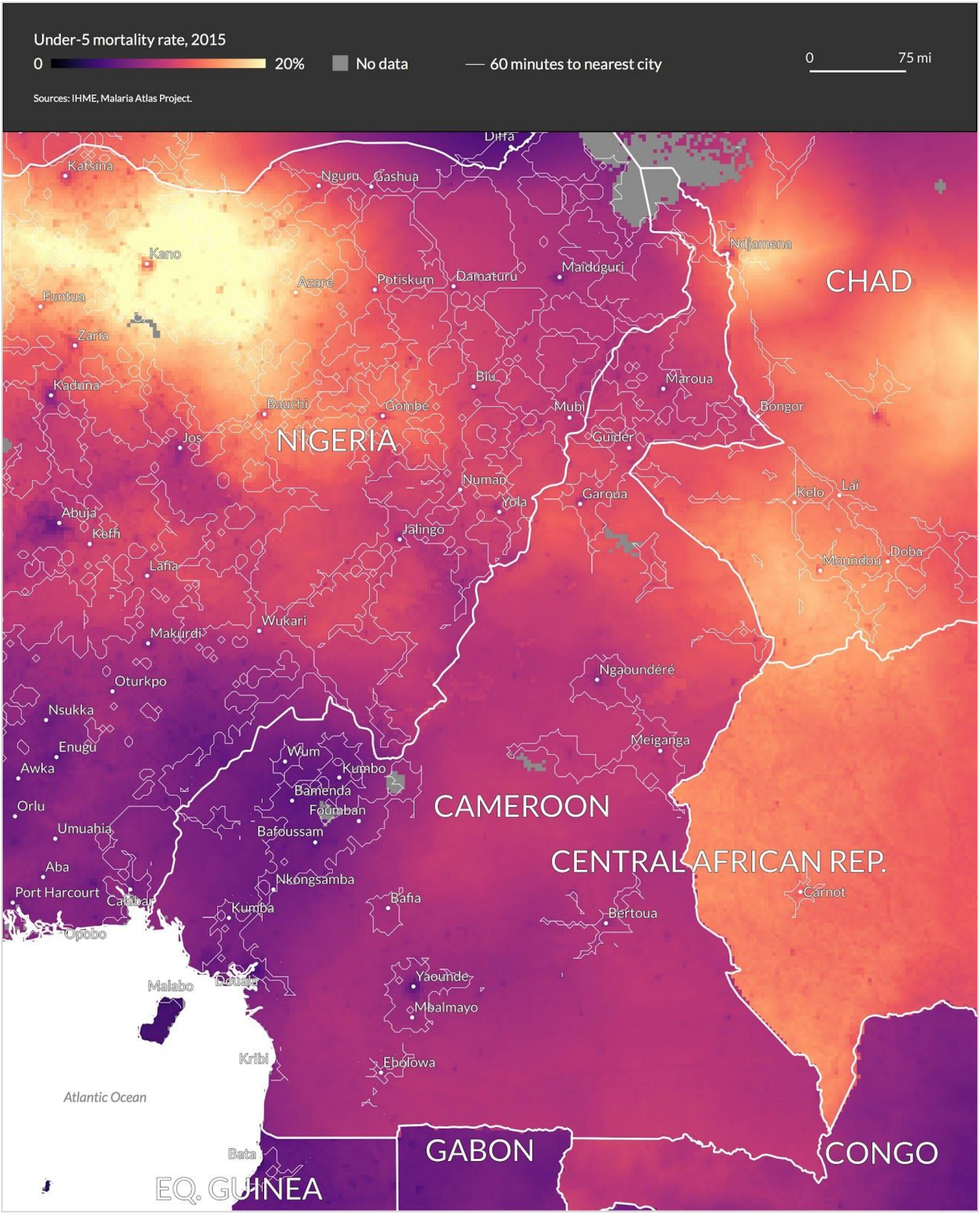
The result is that we can see that **access to cities is strongly correlated with child mortality rates**: cities generally have lower child mortality rates than surrounding areas, and suburban areas generally have lower rates than surrounding rural areas, although not in all cases.



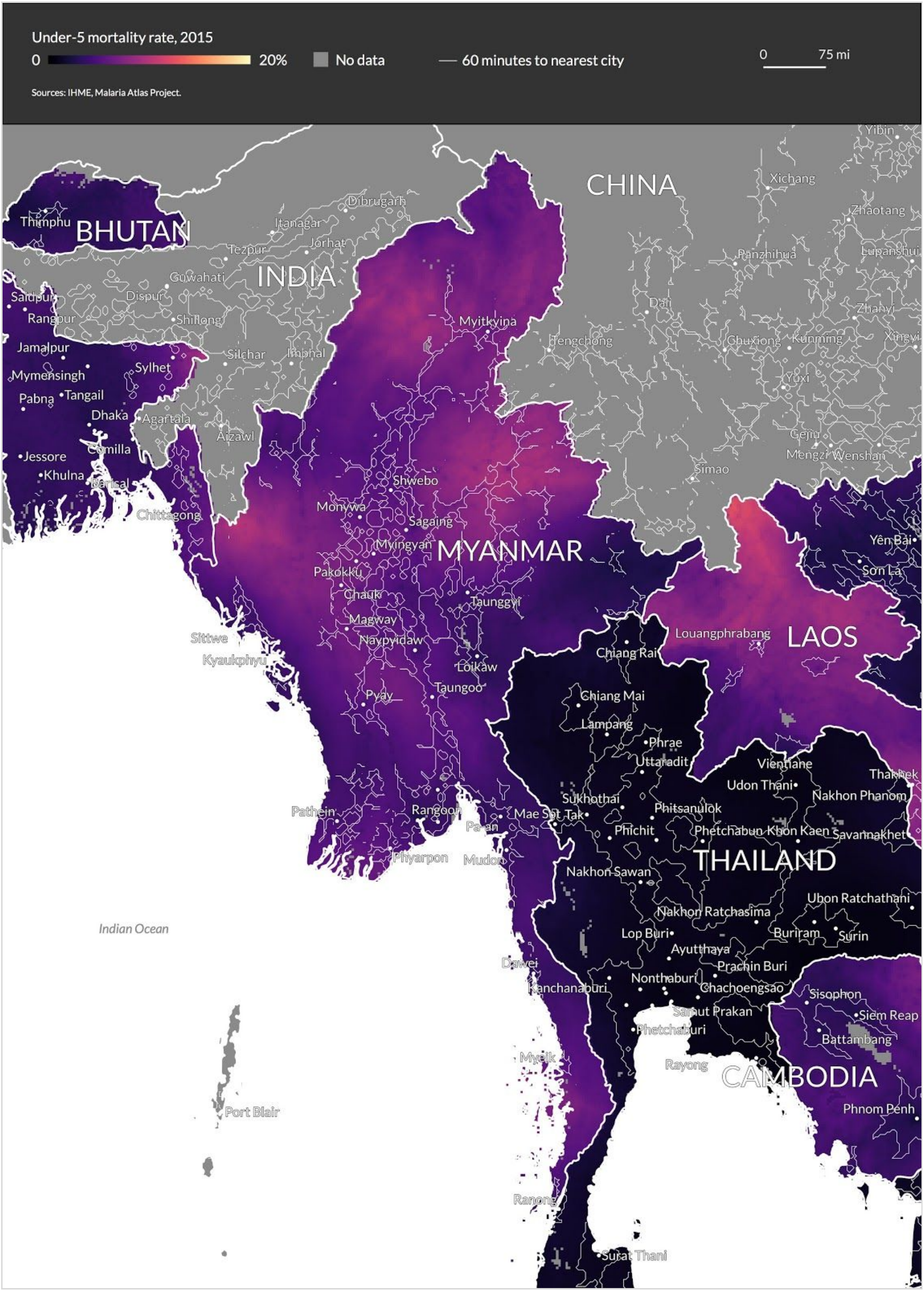
In Angola, we can see that areas with high access to cities almost always have significantly lower child mortality rates than more rural areas, as is the case for Luanda and Caxito, and yet there are also regional variations in child mortality within the country that do not correlate with access to cities. This is the pattern for most countries.



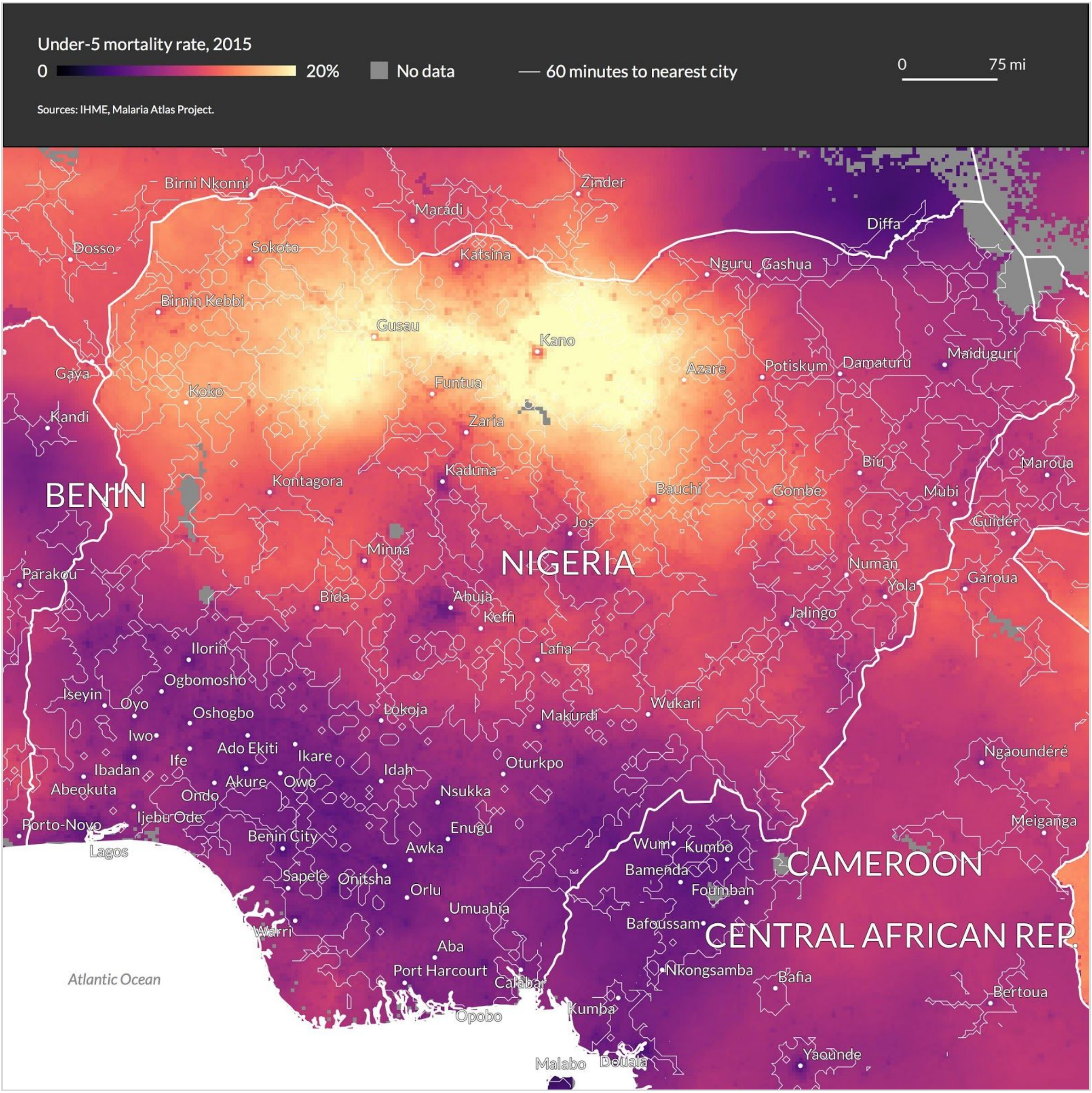
In Cambodia, we see a clear pattern, with the areas with greater access to cities having significantly lower mortality rates. We can also see the stark differences between countries, as child mortality rates in Thailand and Vietnam are near zero, regardless of the level of access to cities.



Cameroon shows a similar pattern, although the suburban areas surrounding cities don’t always differ markedly from the more rural areas.

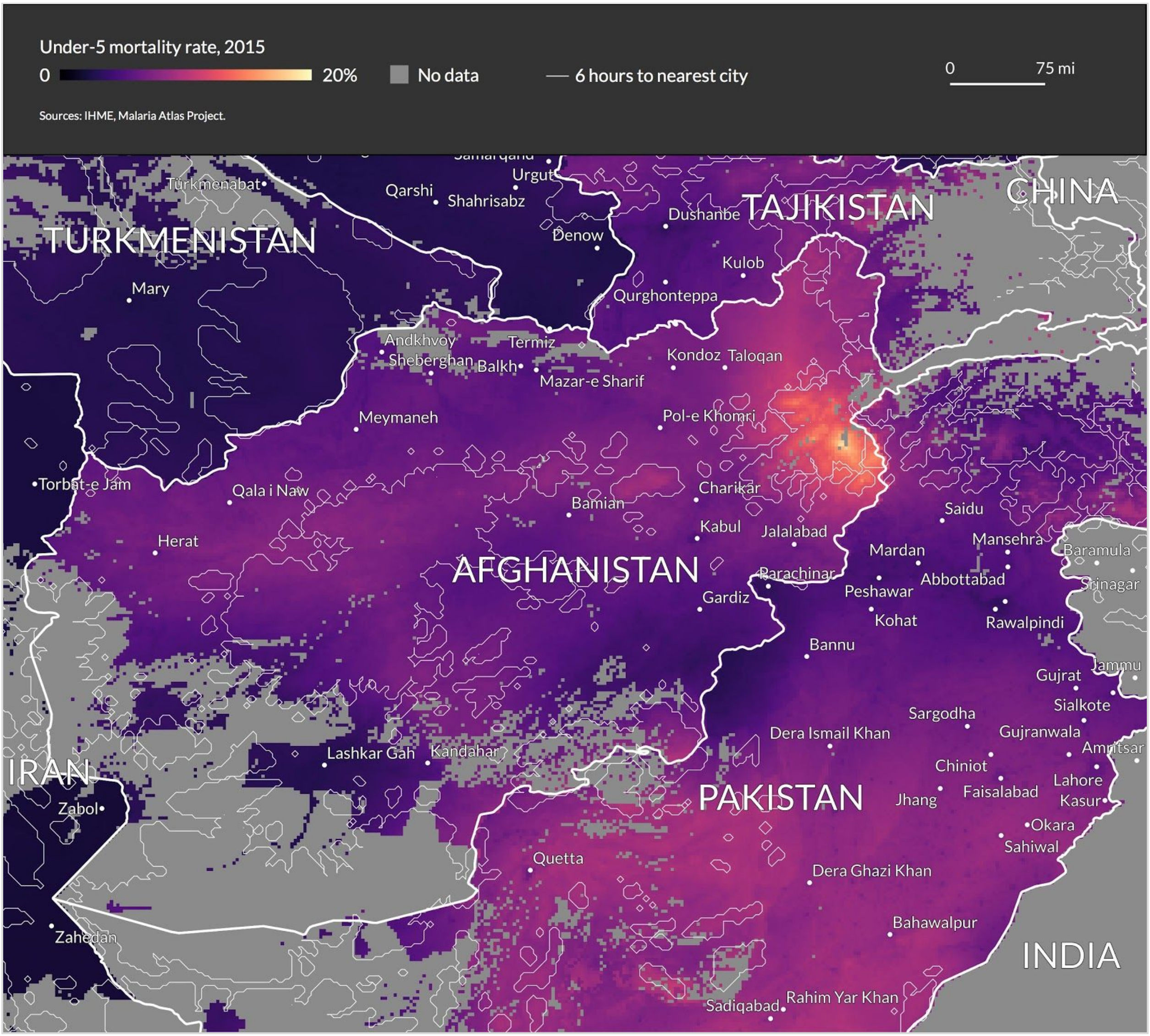

Under-5 mortality rate, 2015
0
20%
No data
60 minutes to nearest city
0
75 mi
Sources: IHME, Malaria Atlas Project.
CHINA
BHUTAN
INDIA
MYANMAR
LAOS
THAILAND
CAMBODIA
Indian Ocean
Thimphu
Dibrugarh
Itanagar
Jorhat
Tezpur
Guwahati
Dispur
Shillong
Saidpur
Rangpur
Jamalpur
Mymensingh
Sylhet
Pabna
Tangail
Dhaka
Agartala
Silchar
Imphal
Aizawl
Jessore
Comilla
Khulna
Barisal
Chittagong
Myitkyina
Tengchong
Shwebo
Monywa
Sagaing
Myingyan
Pakokku
Chauk
Taunggyi
Magway
Naypyidaw
Sittwe
Kyaukphyu
Loikaw
Taungoo
Pyay
Pathein
Rangoon
Pa-an
Mae Sot
Tak
Phyarpon
Mudon
Dawei
Myeik
Ranong
Port Blair
Surat Thani
Xichang
Zhaotang
Panzhihua
Liupanshui
Dali
Zhanyi
Chuxiong
Kunming
Xingyi
Yuxi
Gejiu
Mengzi
Wenshan
Simao
Yên Bái
Sơn La
Louangphrabang
Chiang Rai
Chiang Mai
Lampang
Phrae
Uttaradit
Vientiane
Udon Thani
Thakhek
Nakhon Phanom
Sukhothai
Phitsanulok
Phichit
Phetchabun
Khon Kaen
Savannakhet
Nakhon Sawan
Ubon Ratchathani
Nakhon Ratchasima
Lop Buri
Buriram
Surin
Ayutthaya
Prachin Buri
Nonthaburi
Kanchanaburi
Chachoengsao
Sisophon
Samut Prakan
Siem Reap
Phetchaburi
Battambang
Rayong
Phnom Penh

Myanmar shows a similar pattern, and stark differences with its neighbors.

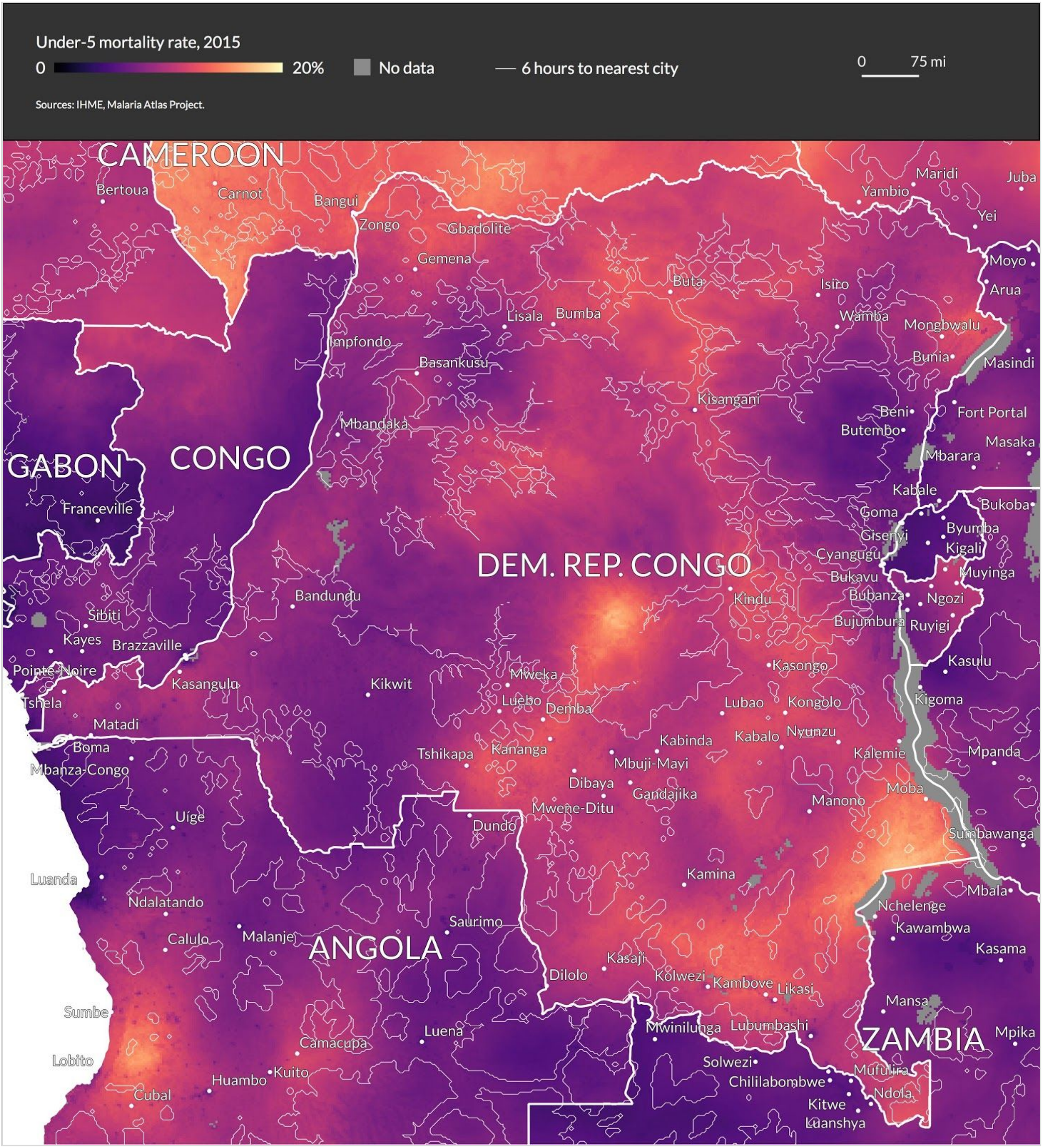


Differences in child mortality within Nigeria are large, with the north of the country showing some of the highest rates in the world, although we still see the pattern of lower rates in cities. Nigeria is also one of the most urban countries in the study, with a large portion of the country living within an hour of a city.

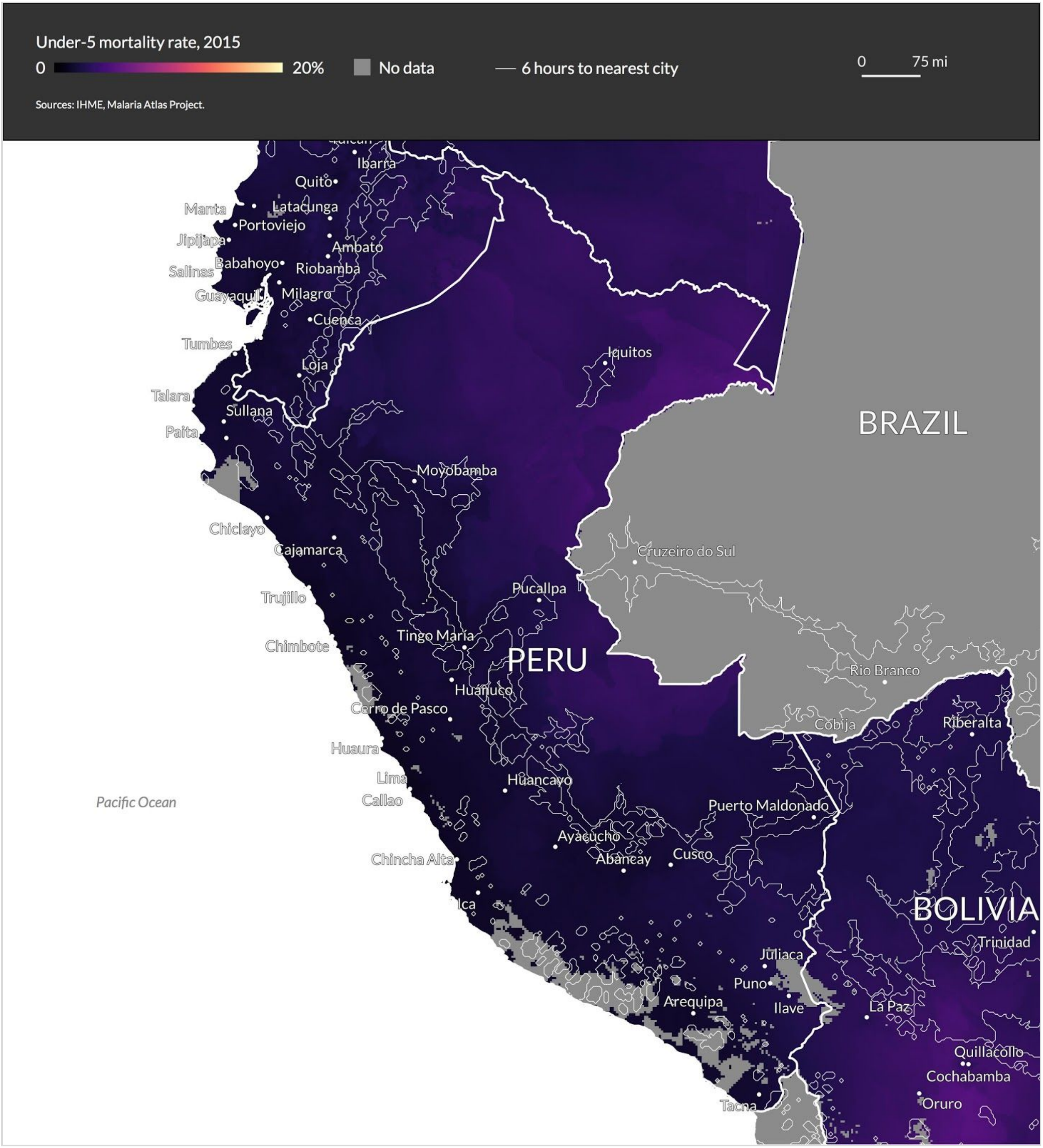
In more rural countries, it is more helpful to overlay a lower level of access to cities. The following maps delineate **6 hours** of travel time to the nearest city.



In Afghanistan, the six hour delineation often lines up nearly exactly with jumps in the child mortality rate, with higher rates in the rural areas in the middle and northeast of the country.

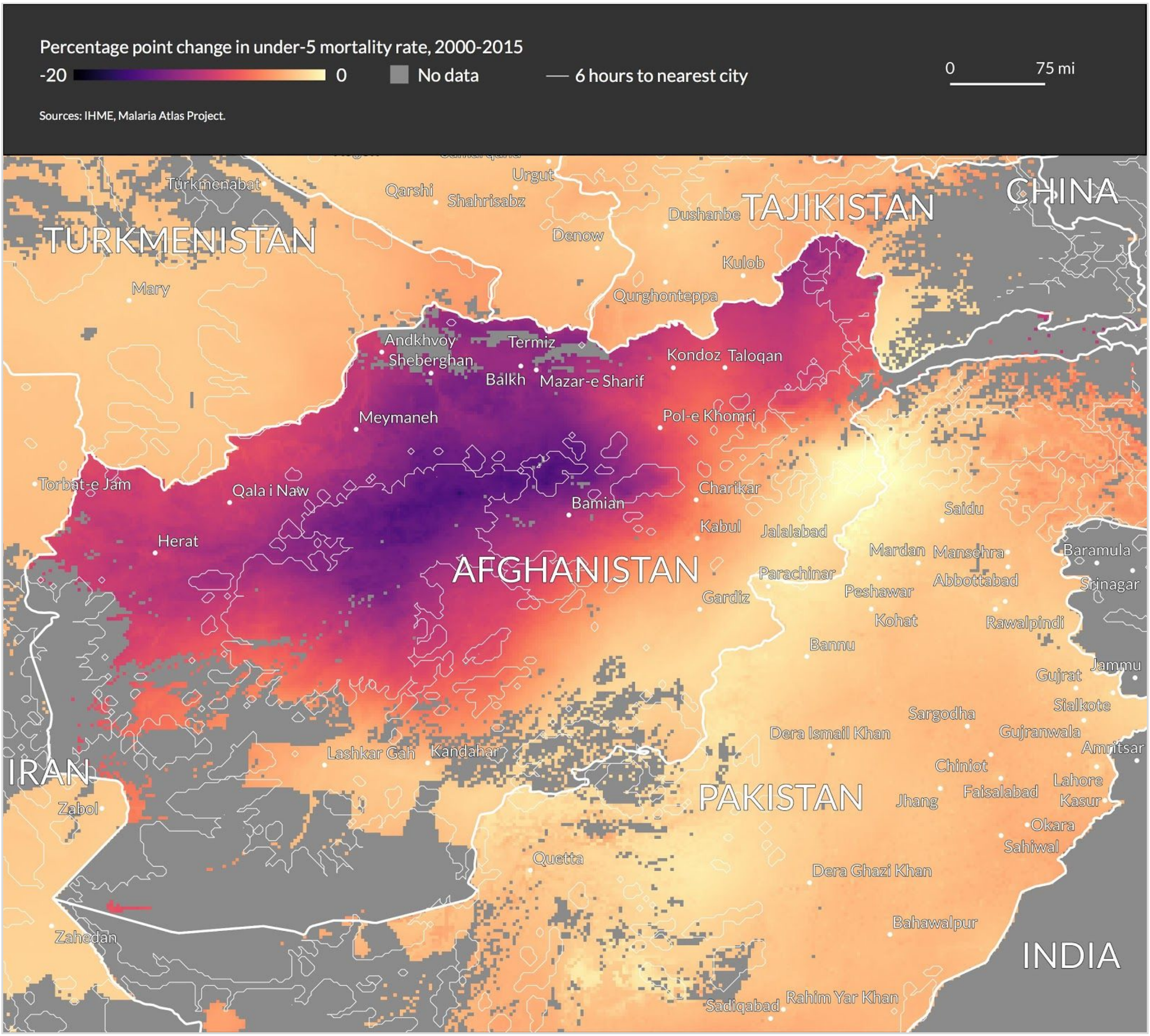


The Democratic Republic of the Congo shows a number of interesting patterns. Unlike other countries, some areas with higher access to cities show **higher** rates of child mortality, and throughout the country, there are regional patterns that do not correlate with access to cities.

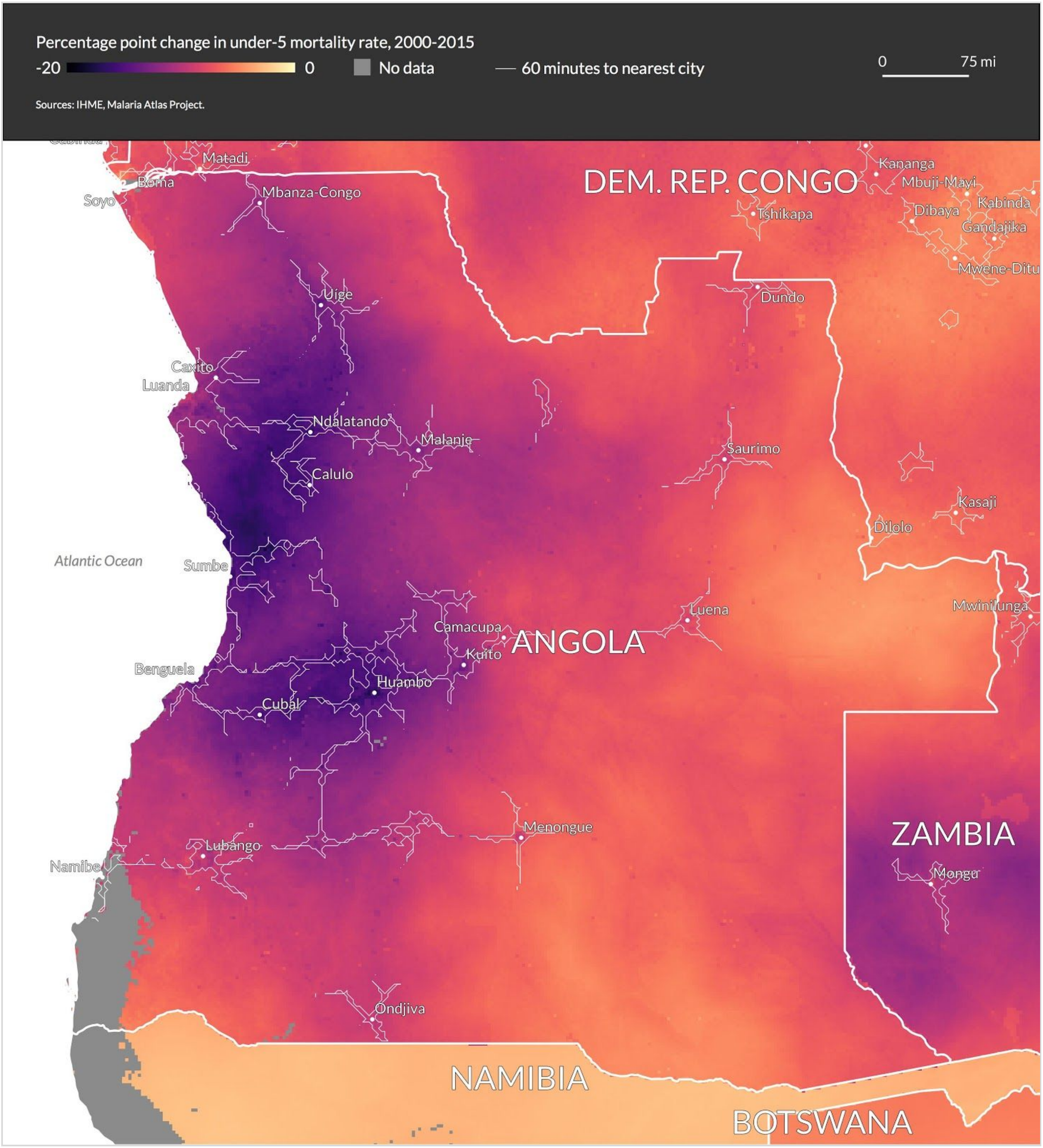


Peru has relatively low rates throughout the country, although we still see higher rates in the rural areas (areas in the Amazon).

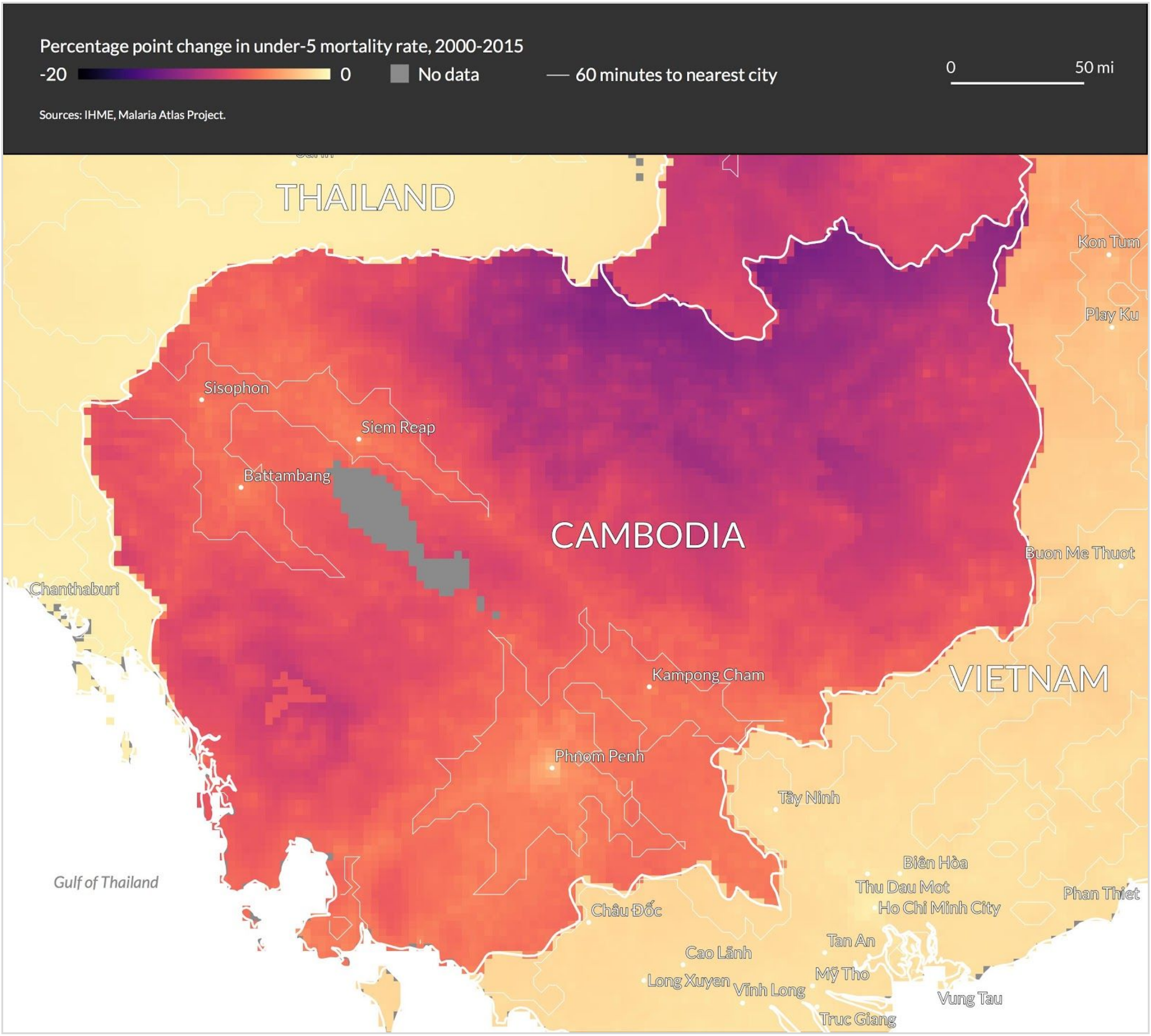
We can also examine the **change in child mortality rate from 2000-2015**, and how it correlates to access to cities in 2015. Interestingly, the relationship does not always go in the same direction.



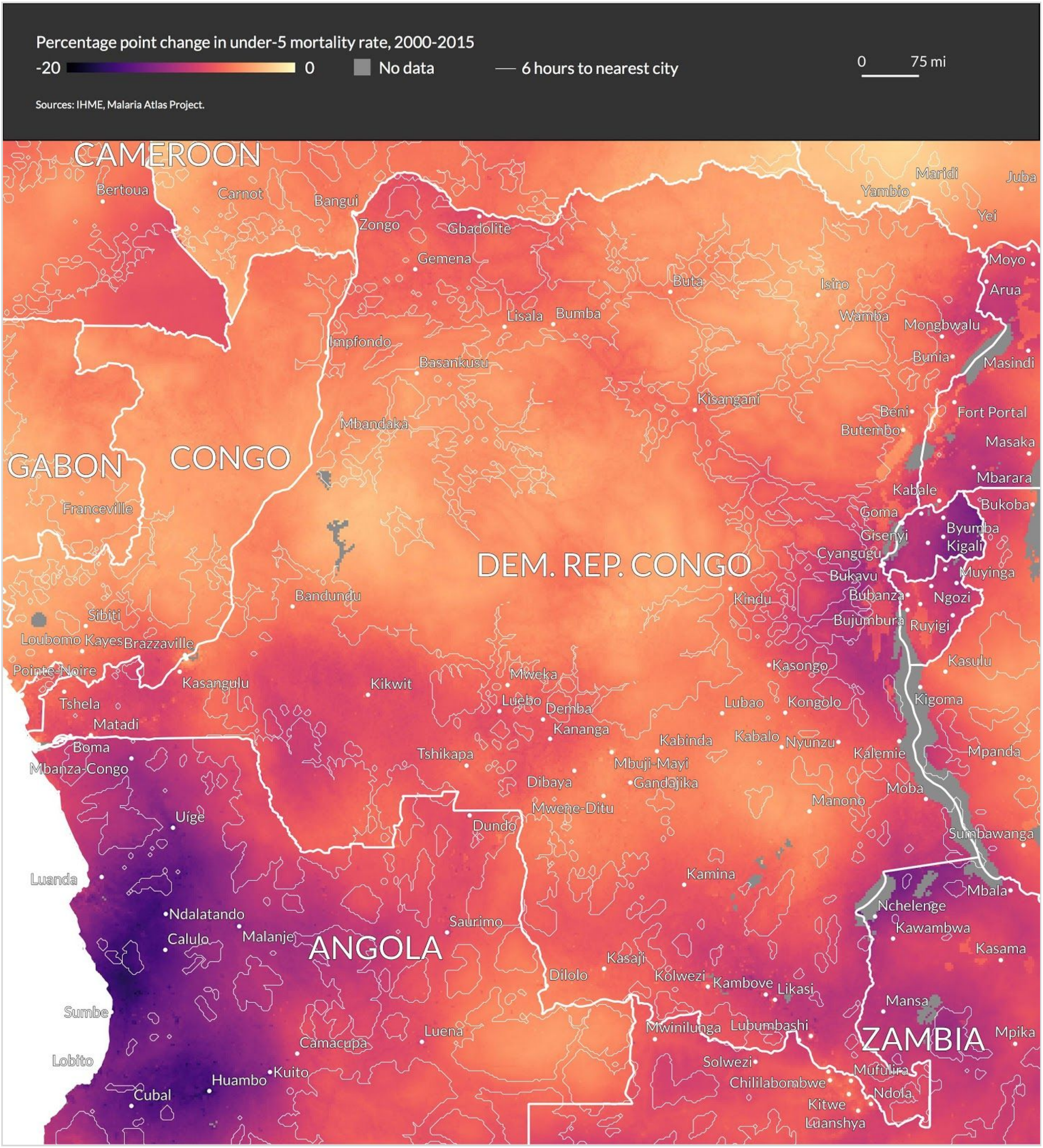
In Afghanistan, the rural areas in the middle of the country show huge improvement from 2000 to 2015 (a more than 10 percentage point drop in mortality rate). However, the rural area in the northeast of the country (which we saw earlier has a very high mortality rate) does not.



Unlike Afghanistan, in Angola the areas with higher access to cities show the largest progress.

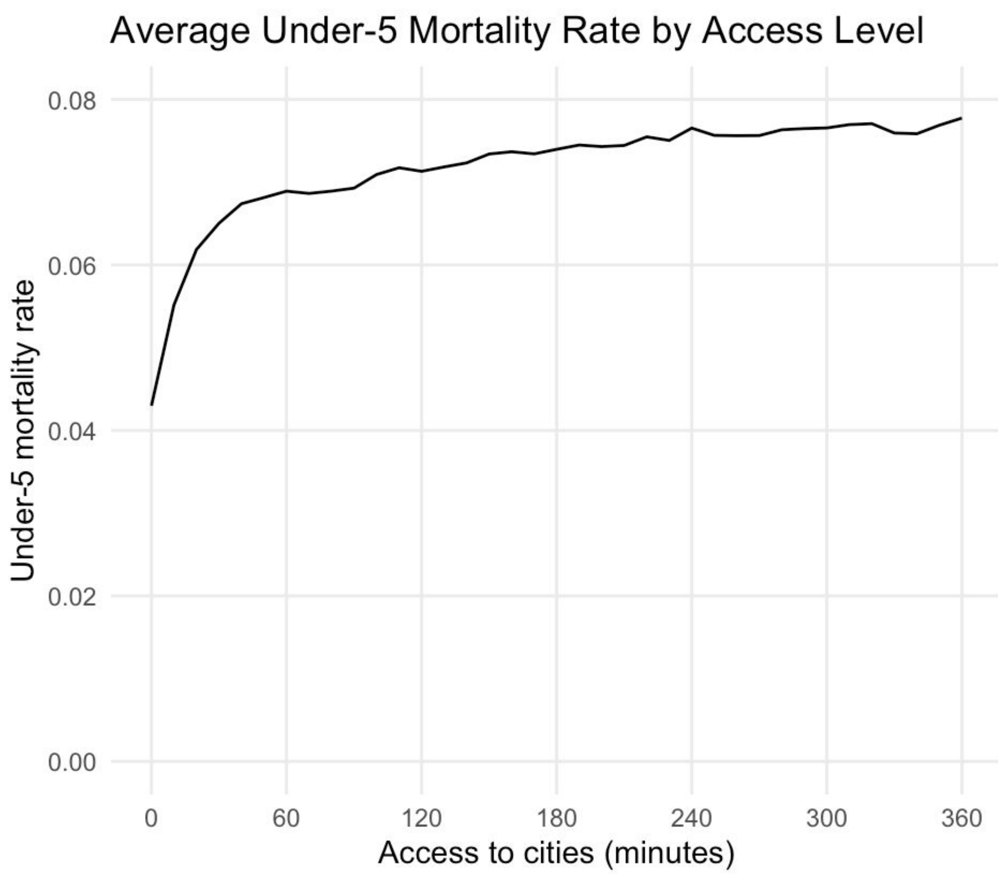


By contrast, in Cambodia the rural areas show the largest improvement, and the cities themselves show a smaller change.

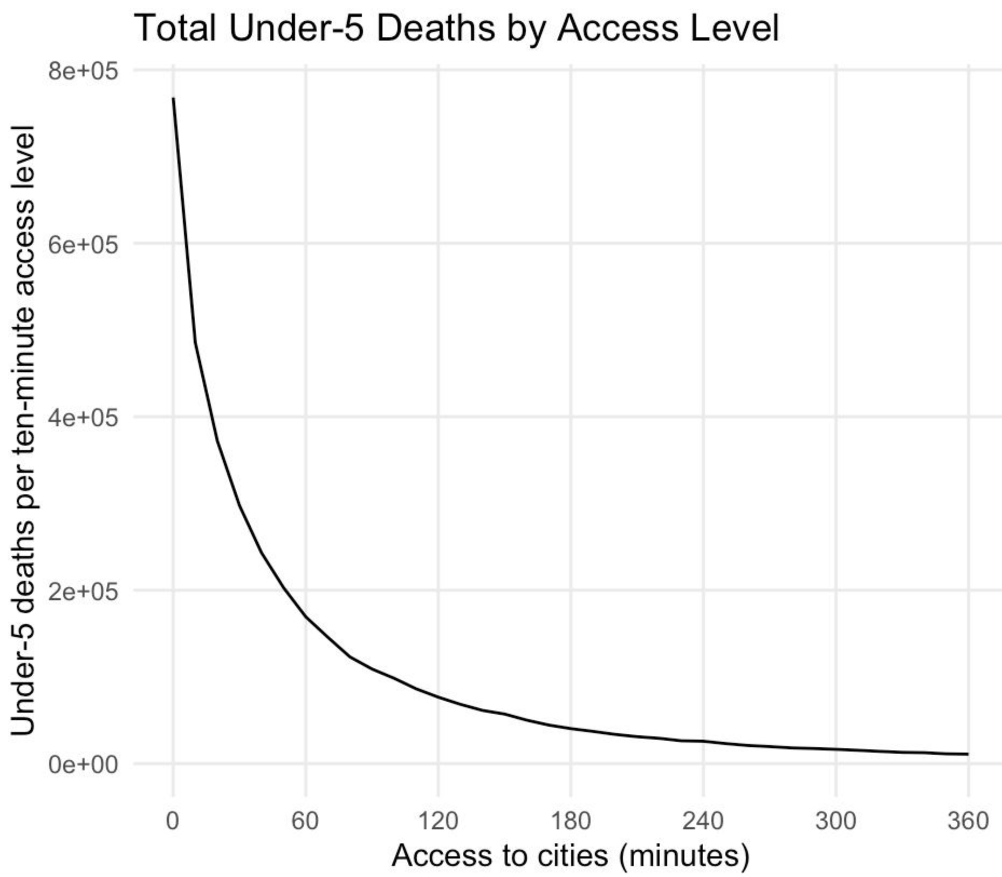


In the Democratic Republic of the Congo, the areas with greater access to cities often show slightly more improvement than surrounding areas, although as with the general child mortality rate, there are other regional patterns.

Finally, we can use R to analyze the aggregated mortality rate by access level, as well as the total number of deaths by access level, dividing access into bins of ten minutes.



As we would expect, we can see that areas with greater access to cities have lower child mortality rates. Interestingly, increased access results in increasingly better mortality rates: moving from 60 to 30 minutes results in a modest improvement, while moving from 30 to 0 minutes results in a much larger improvement.



Looking at total child deaths, we can see that despite lower rates, urban areas have much larger numbers of deaths due to their large populations.

As a summary, we can sum the number of deaths for three different levels of access, and see that most deaths occur in areas within 60 minutes of a city, although a large portion do occur outside this range:

Access	**Deaths**
< 5 minutes	831,034
6-60 minutes	1,644,501
> 60 minutes	1,617,107

This suggests that all levels of access require attention from policymakers.

APPENDIX

R code:

```
library(raster)
library(rgdal)
library(ggplot2)

access <- raster("Access_5km_Rraster.grd")
deaths <- raster("U5_deaths_2015.grd")
#rate <- raster("U5_rate_2015.grd")
pop <- raster("population_2015_derived.grd")

# we will aggregate to get rates for different access levels,
# therefore we bring in population (averaging rates without
# including population would lead be an unweighted average)

# clip access to extent of other layers
access_cropped <- crop(x = access, y = deaths)

# stack layers
st <- stack(list(deaths=deaths, pop=pop, access=access_cropped))

# convert to dataframe
df <- as.data.frame(st, xy=FALSE)

# round to nearest 10 to create bins
# df$access_round <- round(df$access, digits = -1)
# custom round function, because R's native round function rounds 5s to nearest even
number
round2 = function(x, n) {
  posneg = sign(x)
  z = abs(x)*10^n
  z = z + 0.5
  z = trunc(z)
  z = z/10^n
  z*posneg
}

df$access_bin <- round2(df$access, -1)
head(df, 70)

# change to character in order to aggregate by bin
df$access_bin_char <- as.character(df$access_bin)
head(df, 70)

# aggregate all columns by rounded access level
```

```
agg <- aggregate(.~access_bin_char, df, sum, na.rm=TRUE)
str(agg)

# change back to numeric for line chart
agg$access_bin_num <- as.numeric(agg$access_bin_char)
head(agg)
str(agg)

# calculate rate for each access bin
agg$rate <- agg$deaths / agg$pop
head(agg)

# Rate line chart
ggplot(data = agg, aes(x = access_bin_num, y = rate)) +
  geom_line() +
  scale_x_continuous(breaks = seq(0, 360, by = 60), limits = c(0, 360)) +
  ylim(0, 0.08) +
  theme_minimal() +
  theme(panel.grid.minor = element_blank()) +
  labs(y= "Under-5 mortality rate",
       x = "Access to cities (minutes)",
       title = "Average Under-5 Mortality Rate by Access Level")

# Deaths line chart
ggplot(data = agg, aes(x = access_bin_num, y = deaths)) +
  geom_line() +
  scale_x_continuous(breaks = seq(0, 360, by = 60), limits = c(0, 360)) +
  theme_minimal() +
  theme(panel.grid.minor = element_blank()) +
  labs(y= "Under-5 deaths per ten-minute access level",
       x = "Access to cities (minutes)",
       title = "Total Under-5 Deaths by Access Level")
```