National Research University Higher School of Economics

Master's Programme 'Data Analytics and Social Statistics (DASS)'

Exploratory Data Analysis (prof. Batagelj)

Project 1: The World Factbook

Contents

The project was prepared by **Timofei Korovin***, DASS student*

Short formulation of the task:

Using *The CIA World Factbook* it is needed to construct a data frame in which units (rows) are world countries with names from the "book" and variables (columns). Next, we should save the created data frame as a CSV file. Finally, we have to explore the collected data.

Creation date: 10/03/2025

The last change date: 21/04/2025

1. Research question

To what extent is the accessibility of medical services (measured by physician density and hospital bed density) associated with maternal and infant mortality rates across countries?

1.2. Studies on that topic

The relationship between access to medical services and mortality rates among women and children has been widely studied. Many researchers support the statement that improved healthcare accessibility, particularly in terms of healthcare infrastructure and workforce, is strongly correlated with lower maternal and infant mortality rates.

Some studies have highlighted the significant role of physician density in improving health outcomes. For instance, Anand and Bärnighausen (2004) found that a higher density of physicians was positively associated with reduced maternal and child mortality, especially in low- and middle-income countries.

Hospital bed density is another important factor of health system capacity. A comprehensive study examining China's healthcare system from 2004 to 2016 by J.Tian, F.Pan (2021) found that increased hospital bed supply was significantly associated with decreased maternal mortality rates. Specifically, the study reported that an increase of one hospital bed per 1,000 population corresponded to an average reduction of 18.21% in the maternal mortality ratio (MMR).

Overall, the existing literature strongly supports that strengthening healthcare infrastructure and expanding access to skilled healthcare providers significantly reduces mortality among women and children.

2. Variables selection

For our analysis we have chosen the following variables: **Maternal mortality rate, Infant mortality rate, Physicians density, and Hospital bed density.**

3.Data extraction

First, we use The CIA World Factbook to extract variables of our interest.

Maternal mortality rate

```r
mat_mort <- rep(0,n)

for(ci in 1:n){
   mmr <- J$countries[[ci]]$data$people$maternal_mortality_rate
   v <- mmr$deaths_per_100k_live_births[1]; d <- mmr$date[1]
   mat_mort[ci] <- ifelse(is.null(v), NA, v)
   year <- ifelse(is.null(d), "0000", d)
}
```

Infant mortality rate

```
inf_mor <- rep(0,n)

for(ci in 1:n){
   imr <- J$countries[[ci]]$data$people$infant_mortality_rate$total
   v1 <- imr$value[1]; d1 <- imr$date[1]
   inf_mor[ci] <- ifelse(is.null(v1), NA, v1)
   year <- ifelse(is.null(d1), "0000", d1)
}
```

Physicians density

```
phy_den <- rep(0,n)

for(ci in 1:n){
   pd <- J$countries[[ci]]$data$people$physicians_density
   v2 <- pd$physicians_per_1000_population[1]; d2 <- pd$date[1]
   phy_den[ci] <- ifelse(is.null(v2), NA, v2)
   year <- ifelse(is.null(d2), "0000", d2)
}
```

Hospital bed density

```
bed_den <- rep(0,n)

for(ci in 1:n){
   bd <- J$countries[[ci]]$data$people$hospital_bed_density
   v3 <- bd$beds_per_1000_population[1]; d3 <- bd$date[1]
   bed_den[ci] <- ifelse(is.null(v3), NA, v3)
   year <- ifelse(is.null(d3), "0000", d3)
}
```

ISO, Countries, Region

```
countries <- rep(0,n)
  for(ci in 1:n){
    countries[ci] <- C[ci]
  }
country_names <- countryname(countries, "country.name.en")

iso <- countryname(countries, destination = 'iso2c')

region <- rep(0,n)
  for(ci in 1:n){
    regn <- J$countries[[ci]]$data$geography$map_references
    region[ci] <- ifelse(is.null(regn), NA, regn)
  }
```

Constructing Data Frame

After extracting our variables we are ready to construct a data frame that will contain 6 variables: Names of the countries, ISO codes, Maternal_Mortality_rate, Infant_mortality_rate, Physicians_density, Hospital_bed_density and Region.

```r
df1 <- data.frame("Country" = country_names,
                  "V1" = iso,
                  "V2" = mat_mort,
                  "V3" = inf_mor,
                  "V4" = phy_den,
                  "V5" = bed_den,
                  "V6" = region
                  )

df_countries <- df1[!is.na(df1$V1),]
row.names(df_countries) <- c(1:237)
df_countries$V6 <- gsub("; $", "", df_countries$V6)
df_countries$V6 <- gsub(";$", "", df_countries$V6)

summary(df_countries)
```

```
##    Country               V1                    V2              V3
##  Length:237         Length:237         Min.   :   2.0    Min.   :  1.70
##  Class :character   Class :character   1st Qu.:  12.5    1st Qu.:  5.55
##  Mode  :character   Mode  :character   Median :  52.0    Median : 11.50
##                                        Mean   : 163.5    Mean   : 20.03
##                                        3rd Qu.: 205.5    3rd Qu.: 29.50
##                                        Max.   :1150.0    Max.   :104.30
##                                        NA's   :54        NA's   :14
##        V4               V5                V6
##  Min.   :0.010    Min.   : 0.100    Length:237
##  1st Qu.:0.390    1st Qu.: 1.300    Class :character
##  Median :1.570    Median : 2.500    Mode  :character
##  Mean   :1.855    Mean   : 3.126
##  3rd Qu.:2.910    3rd Qu.: 4.250
##  Max.   :8.300    Max.   :14.000
##  NA's   :38       NA's   :58
```

2.EDA

Overall, our data set contains 6 variables, lets explore those that are numeric.

We start with our second variable (Maternal_mortality_rate) its takes values in range from 2 to 1150. The mean equals to 163.5 and median - 52. It also contains 54 missing values. Now lets visualize the distribution of our variable.

```r
library(dplyr)

df_regions <- df_countries %>%
  group_by(V6) %>%
  drop_na() %>% #
```

```
  summarise(across(where(is.numeric), mean, na.rm = TRUE, .names = "mean{.col}"))

print(df_regions)
```

```
## # A tibble: 11 x 5
##    V6                                meanV2 meanV3 meanV4 meanV5
##    <chr>                              <dbl>  <dbl>  <dbl>  <dbl>
##  1 Africa                             349.   41.3  0.397   1.42
##  2 Arctic Region                        4     2.1  3.98    3.1
##  3 Asia                              88.6    23.9  2.40    4.86
##  4 AsiaEurope                          19      7.4  2.99    7.5
##  5 Central America and the Caribbean 93.6    14.5  1.89    2.29
##  6 Europe                             7.17    4.51 3.52    4.98
##  7 Middle East                       30.5    12.7  1.93    1.83
##  8 North America                     21.5     7.5  2.34    2
##  9 Oceania                           49.5    11.4  1.49    2.4
## 10 South America                     126.    16.5  2.19    2.1
## 11 Southeast Asia                    104.    20.8  0.96    1.84
```

Distributions of our variables

Maternal Mortality rate

Maternal Mortality rate (2017)


Number of female deaths per 100,000 live births
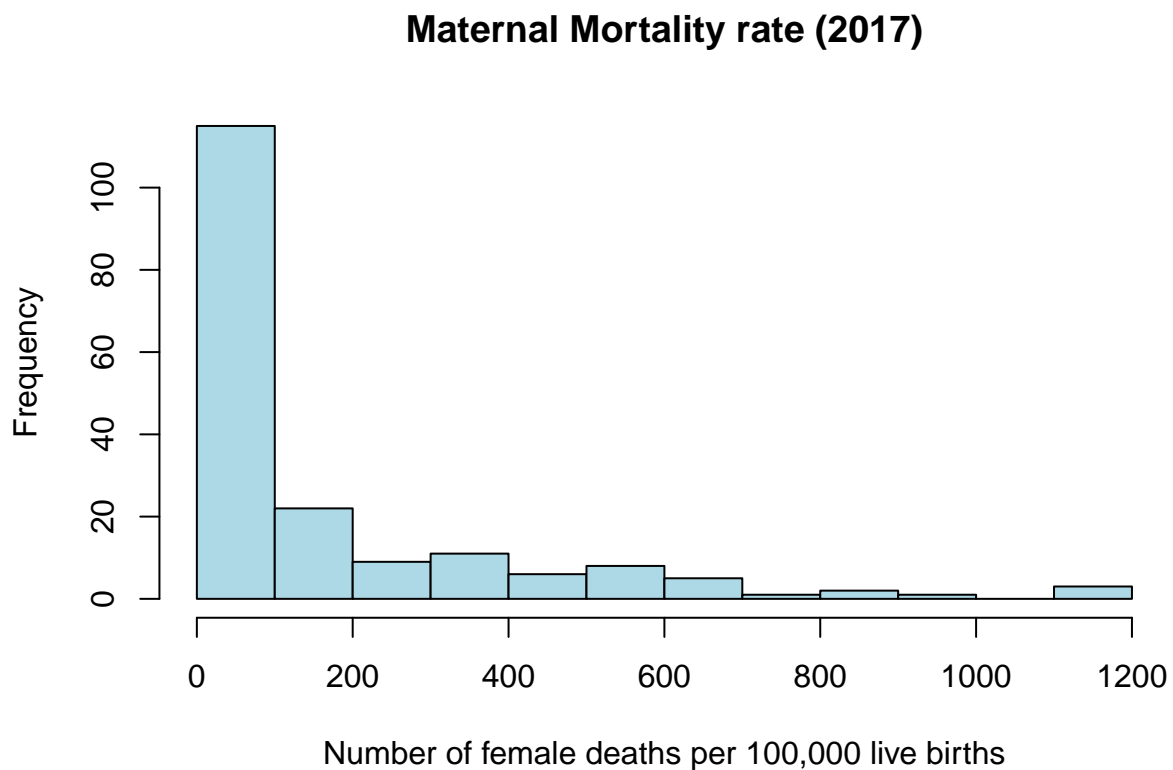
According to the graph below, the distribution is highly right-skewed, which indicates that most countries

have low maternal mortality rates (in the range from 0 to 100). However, there is a small number of countries that have relatively high values of that factor, which reflects the long tail on the right side of the plot.

Moving forward, we can visualize the Maternal Mortality rate by region. For this purpose we will calculate the mean value of the maternal mortality rate for each region and plot it on the same graph, so it will be easier to estimate the most problematic regions.

Maternal Mortality rate by region
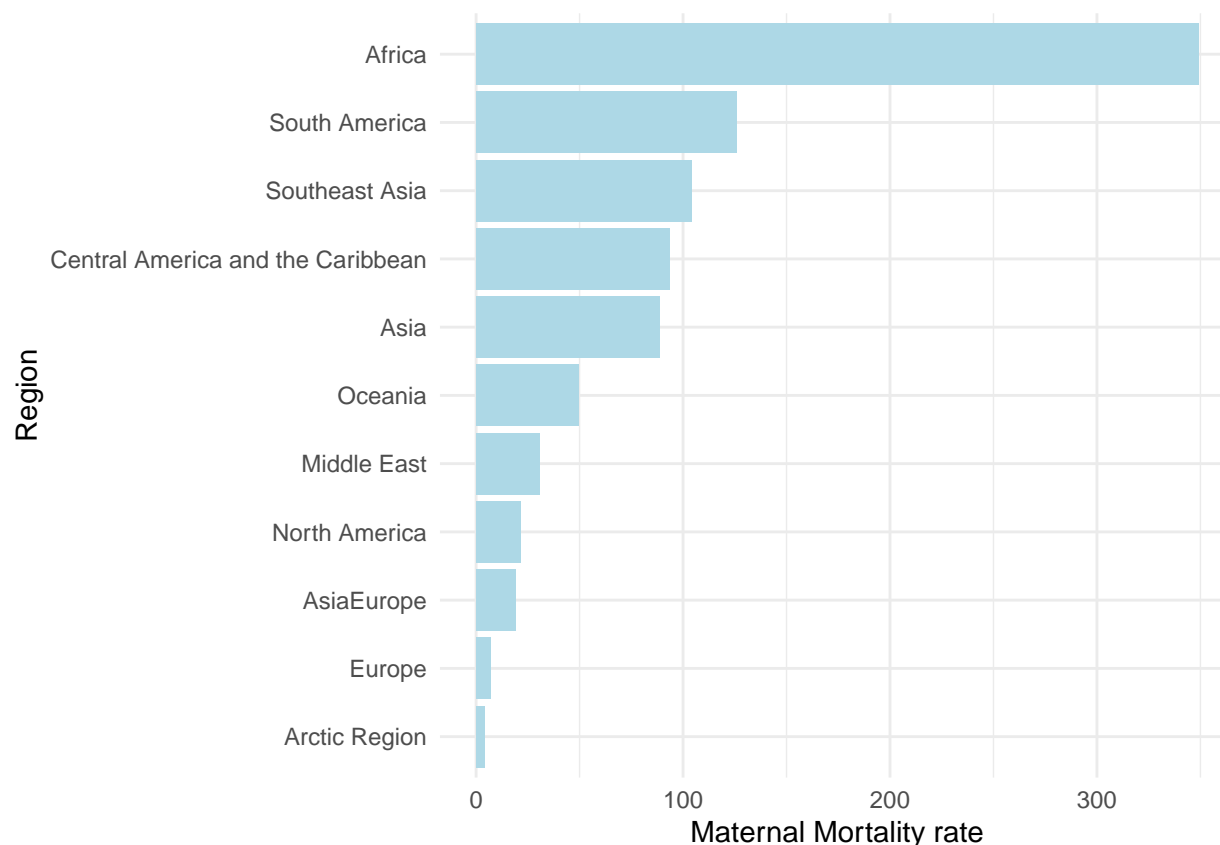
As seen in the graph, the region with the highest maternal mortality rate is Africa (approximately 350 per 100.000 live births). This is not surprising, as Africa is considered one of the poorest regions in the world, which often means limited access to quality healthcare services and infrastructure.

It is followed by South America and Southeast Asia, which demonstrate significantly lower but still high maternal mortality rates (more than 100). It can be explained by the presence of numerous developing countries in these regions, where healthcare systems may still be underdeveloped or underfunded.

Central America and the Caribbean, along with Asia, demonstrate elevated rates, although slightly lower than the previously mentioned regions (sligtly less than 100). In contrast, Oceania and the Middle East have noticeably lower maternal mortality rates(less than 50).

The Arctic Region has the lowest maternal mortality rate, which could be influenced by its small population. Europe also shows very low rates, likely due to advanced healthcare systems, widespread access to maternal care, and overall high standards of living.

Moving forward, lets conduct the same exploratory analysis for "Infant mortality rate" variable.
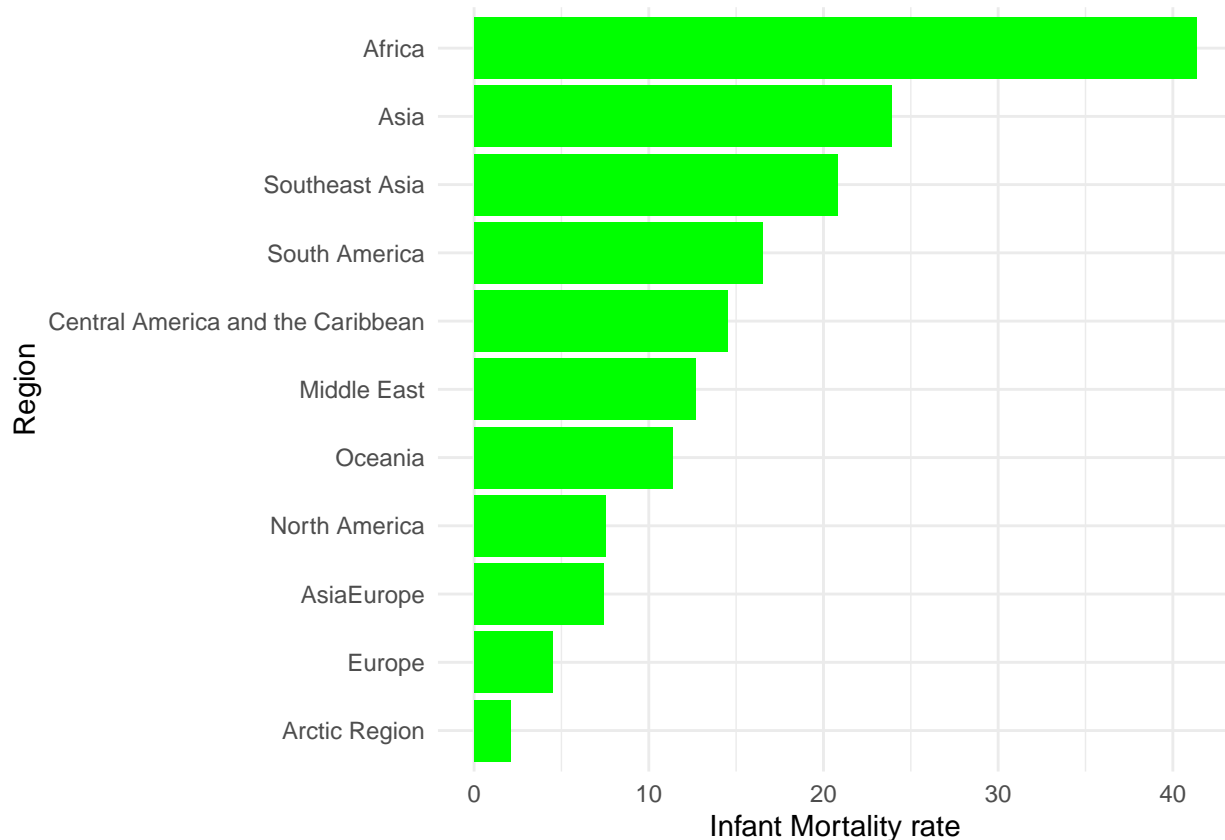
Infant Mortality rate

Infant Mortality rate takes values in the range from 1.7 to 104.3. Also it contains 14 missing values. The mean for that variable is 20 and median equals to 11.5. We will visualize its distribution using histogram.

Infant Mortality rate (2017)

Number of total deaths per 1000 live births
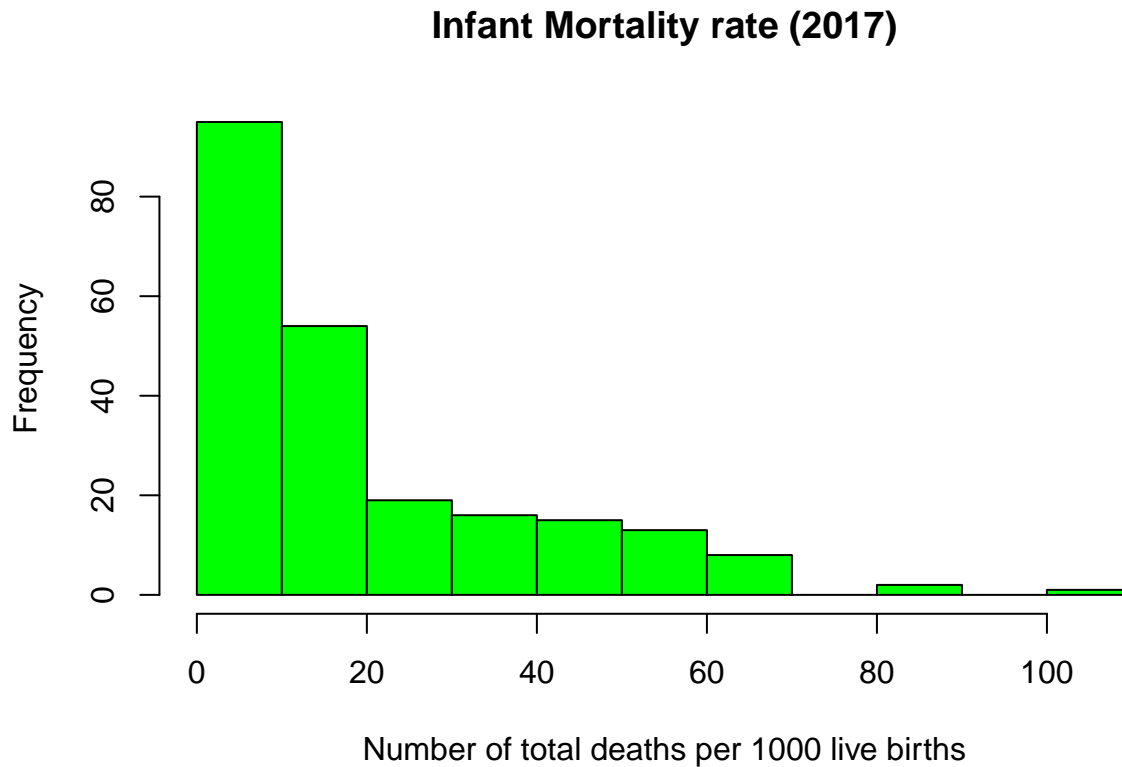
Again we encounter highly right-skewed distribution, which indicates that most of the countries have relatively low rate of Infant Mortality. The biggest share of countries demonstrate about 10 to 20 death per 1000 live birth. However, there is a significant proportion of countries that lag far behind. Countries that have more than 80 deaths per 1000 live birth are outlires in our data set (which also indicates that a very little number of countries demonstrate extremely high infant mortality rate).

Infant Mortality rate by region

The graph display Infant Mortality rate across different regions.

Africa again occupy the top position on the rating demonstrating has the highest infant mortality rate, which exceeds 40 deaths per 1,000 live births.

Asia follows with a significantly lower, but still concerning, infant mortality rates (lightly above 30). Southeast Asia and South America also report relatively high rates (20-25), which is not unexpected considering the large number of developing countries in these areas where healthcare systems can be still underdeveloped.

Central America and the Caribbean, as well as the Middle East and Oceania, occupy a middle position on the char demonstrating almost the same rates (approximately 10-15).

North America and Asia Europe also demonstrate approximately the same infant mortality rates (about 6-7), likely due to better healthcare coverage and stronger public health systems. Europe and the Arctic Region stand out as having the lowest infant mortality rates, suggesting highly effective maternal and newborn care, as well as overall good living conditions. However, as it was on previous graphs, Arctic region shows good rated most likely due to its small population.

It is also worth noting , the Arctic Region again shows very low mortality, which may partly result from its small population, according to the internet resources about 4 millions of people are living in that region.

Lets move forward to the next variable of our interest.

Physicians density

Physicians density variable take values in the range from 0.01 to 8.3. The means equals to 1.855 and median to 1.570. Variable contains 38 missing values. We will also use histogram to visualize its distribution.

Physicians density (2016)

Number of physicians per 1,000 of the population
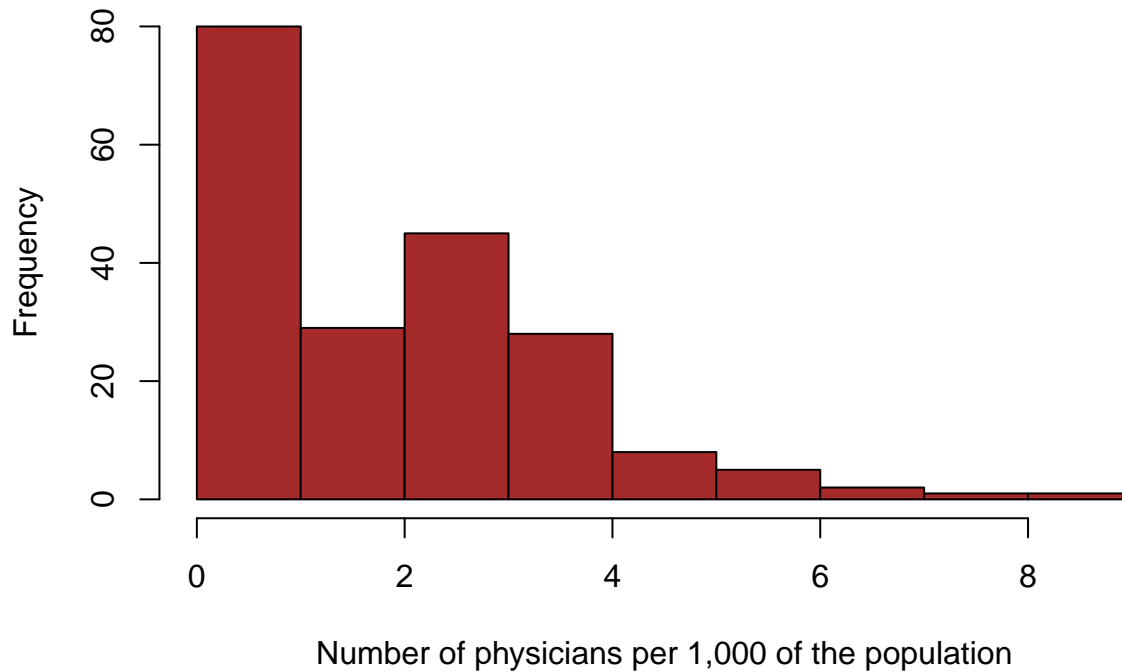
The histogram illustrates the distribution of physician density across different countries, measured by the number of physicians per 1,000 people.

The distribution is right-skewed, with a long tail extending to the right. This indicates that a majority of countries have a relatively low physician density, while a smaller number of countries have a significantly higher density.

Most of the countries fall in in the 0-1 bin, illustrating the significant proportion of countries fails to provide an adequate access to the healthcare.

Physicians density by region
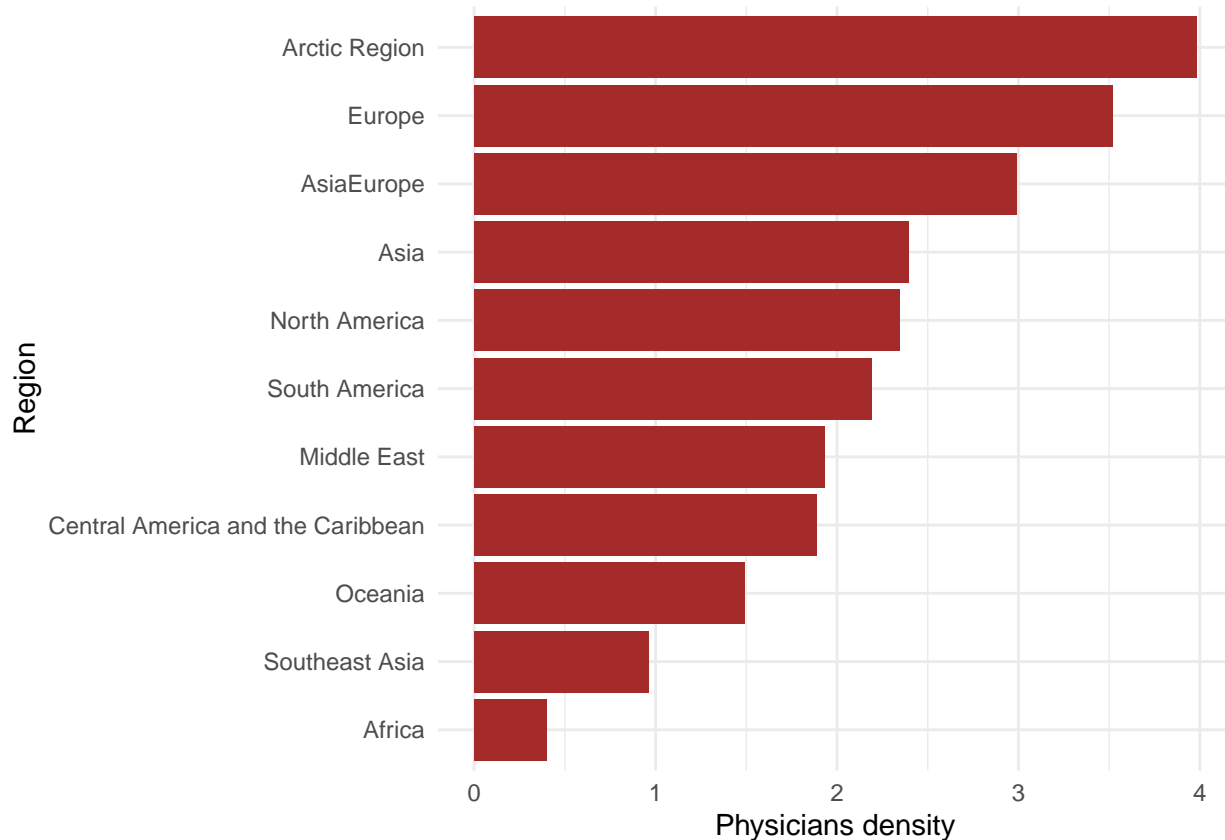
The graph displays the density of physicians across different world regions. The Arctic Region has the highest density of physicians, with a value exceeding 4. However, we think that Arctic region demonstrates such rates due to the small size of the region overall, as well as due to its small population.

Europe follows closely behind, with a physician density slightly below that of the Arctic Region (with 3.5 physicians per 1000 people). This reflects the region's well-developed healthcare infrastructure and adequate investments in the healthcare services. AsiaEurope and Asia also show relatively high densities (between 2 and 3 physicians).

North America demonstrates a moderate physician density (about 2.5), which aligns with its developed healthcare systems. South America, Middle East, and Central America and the Caribbean occupy middle positions on the chart (with approximately 2 physicians per 1000 people), suggesting this countries face challenges in terms of coverage and resources.

Oceania and Southeast Asia report lower physician densities, which could be linked to geographic isolation, limited medical workforce,and overall resource constraints. Africa has the lowest physician density among all regions, significantly below 1. This highlights severe shortages in medical personnel, often caused by economic limitations, brain drain, and inadequate investment in healthcare infrastructure due to the economic conditions.

Hospital Bed density

The last numerical variable, Hospital bed density, takes values in the range from 0.1 to 14. Its mean equals to 3.126, and median to 2.5. It contains 58 missing values. We will plot its distribution via histogram.
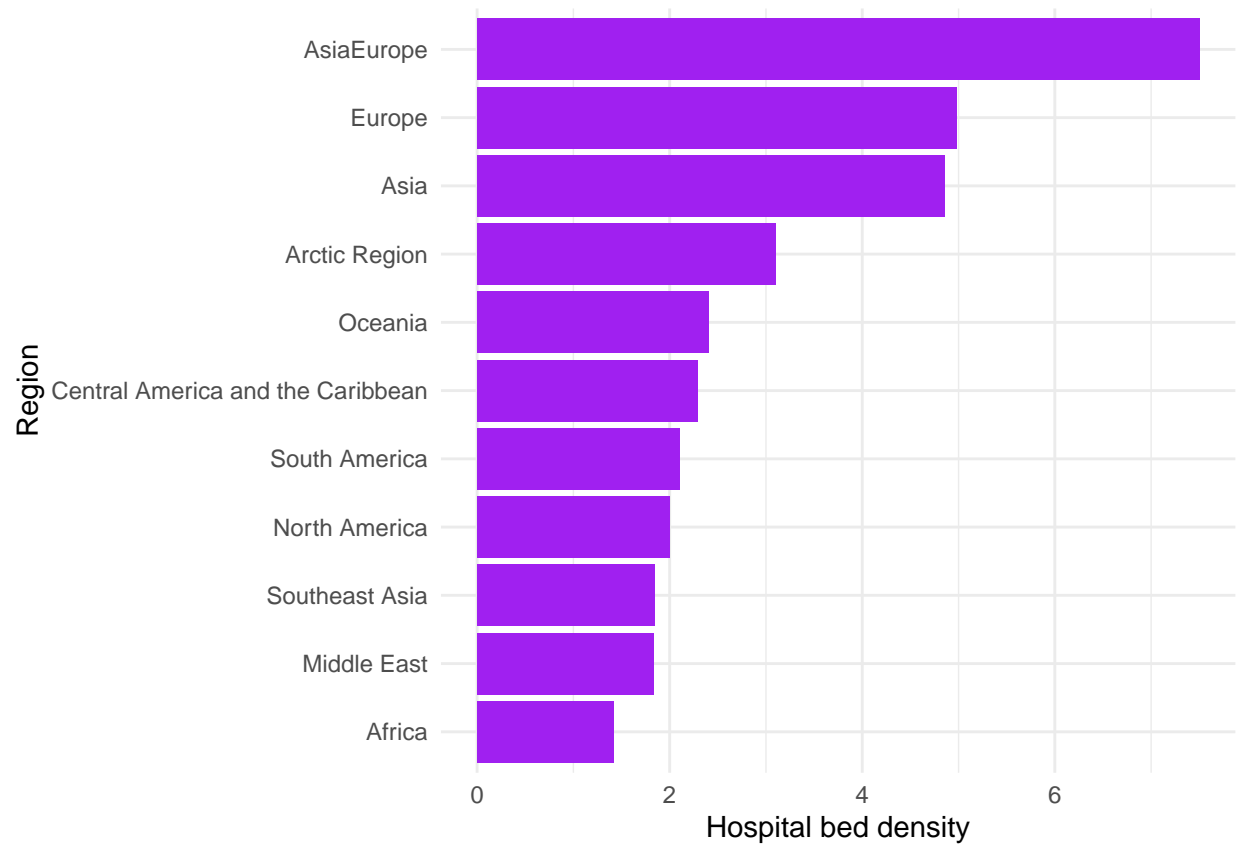
Hospital Bed density (2013)
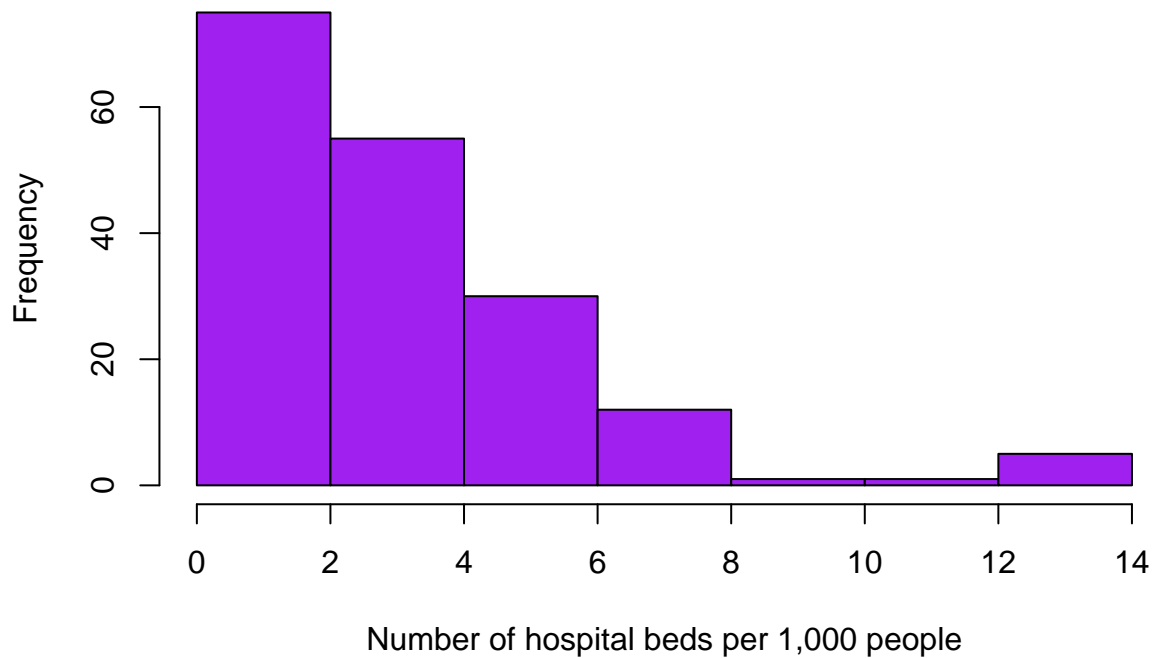
The histogram illustrates the distribution of hospital bed density in 2013, measured by the number of hospital beds per 1,000 people.

The data is right-skewed, as most countries have a low density of hospital beds. The highest frequency is observed in the range of 0 to 2 hospital beds per 1,000 people, indicating that many countries fall within this category. As hospital bed density increases, the frequency declines , with fewer countries having densities above 4 beds per 1,000 people.

A small number of countries exhibit high hospital bed densities (12-14 beds per 1,000 people), which are obvious outliers in the dataset.
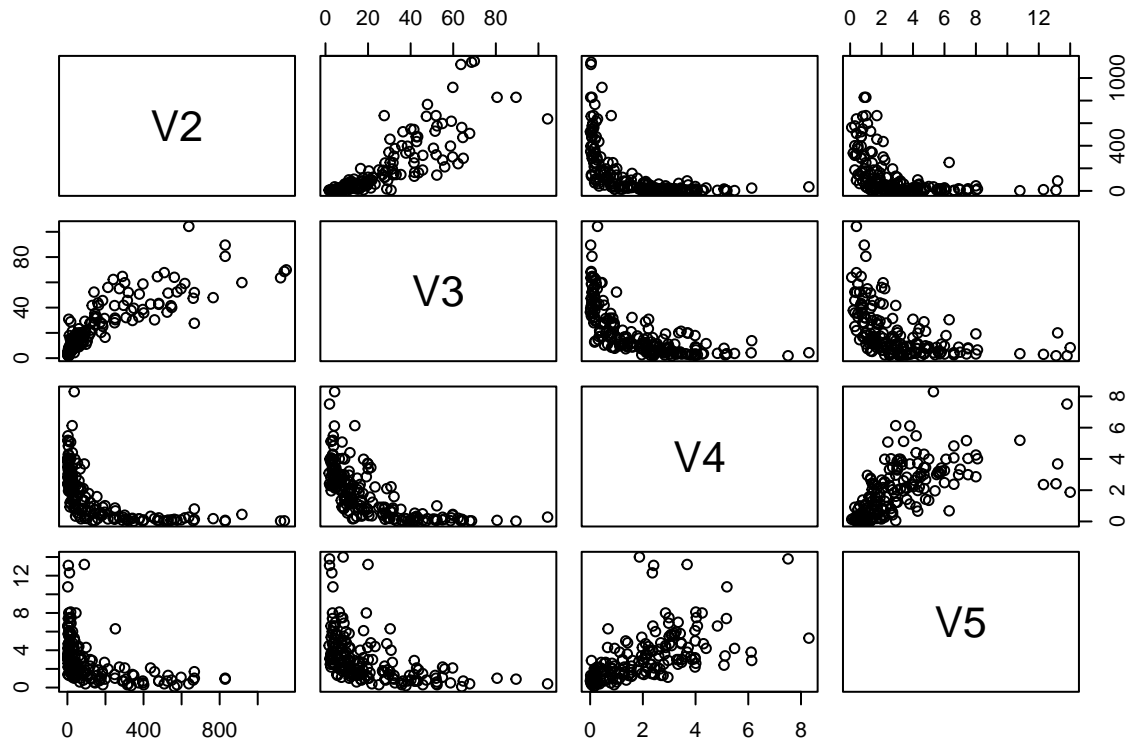
Hospital Bed density by region

The graph displays the density of hospital beds across different world regions. Asia Europe has the highest density of hospital beds, exceeding 6 beds per 1,000 people. It may reflect substantial investments in healthcare infrastructure in this region.

Europe and Asia follow with relatively high hospital bed densities (4.5 beds), indicating well-established healthcare systems in these regions. The Arctic Region demonstrates a moderate density (2.5), which may be adequate for its smaller and more dispersed population. Oceania, Central America and the Caribbean, South America, and North America occupy middle positions on the chart with approximetly 2 beds per 1000 people , suggesting mixed levels of healthcare infrastructure.

Southeast Asia and the Middle East report lower hospital bed densities, potentially linked to resource constraints or developmental challenges in healthcare systems. Africa has the lowest hospital bed density among all regions, significantly below 2 beds per 1,000 people. This highlights a critical shortage of hospital beds, often due to economic limitations and insufficient investment in healthcare.

Pairs diagram

According to the pairs graph, there is a clear structure in the relationships between maternal mortality (V2), infant mortality (V3), physician density (V4), and hospital bed density (V5). Maternal and infant mortality show a strong positive correlation, reflecting shared health factors between mother and child. Both mortality indicators are negatively associated with physician and hospital bed density, indicating that better access to healthcare resources contributes to lower mortality rates. Additionally, physician and hospital bed densities are moderately positively correlated, suggesting that regions with bigger number of physicians also tend to have more hospital infrastructure.

3.Map visualizations

For our visualizations on maps we will choose just Africa Region since it appears to be the most problematic region and thus requires further analysis.

Physicians density for Africa

Physicians density in Africa

Source: CIA World Factbook
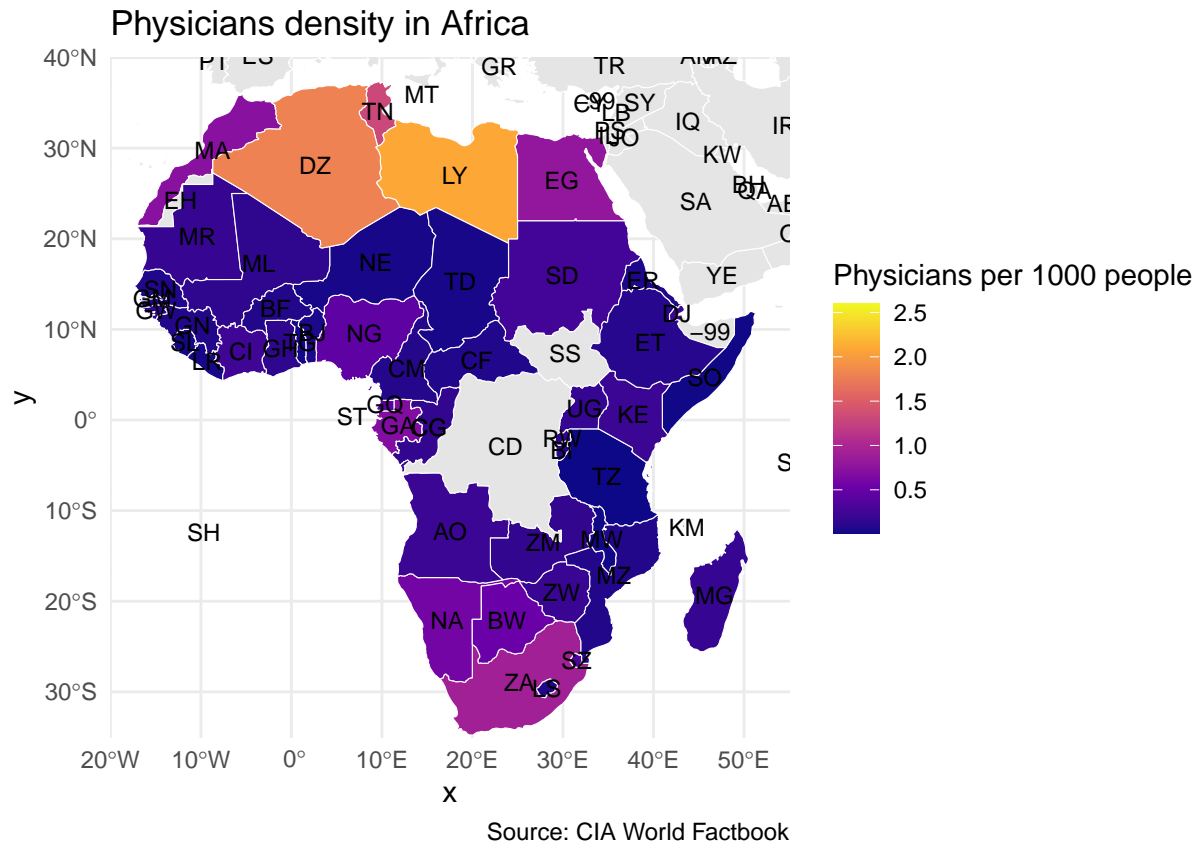
According to the graph below countries with the highest proportion of physicians (closer to 2-2.5 per 1,000 people) are Algeria (DZ) and Libya (LY). Followed by Tunisia (TN) with approximately 1.5 physicians per 1000 people. Morocco (MA), Egypt (EG), South Africa Republic/Zuid-Afrika (ZA), Gabon (GA), Namibia (NA), Botswana(BW) demonstrate moderate rate between 1-1.5 physicians per 1000 people.

It is also noticeable, that many countries have physician density below 1 per 1000 people. Countries with the lowest rate of physicians density within Africa region are Chad (TD), Niger (NE), Tanzania (TZ), Mozambique (MZ), Somali (SO).

Hospital bed density for Africa

Hospital bed density in Africa

Source: CIA World Factbook
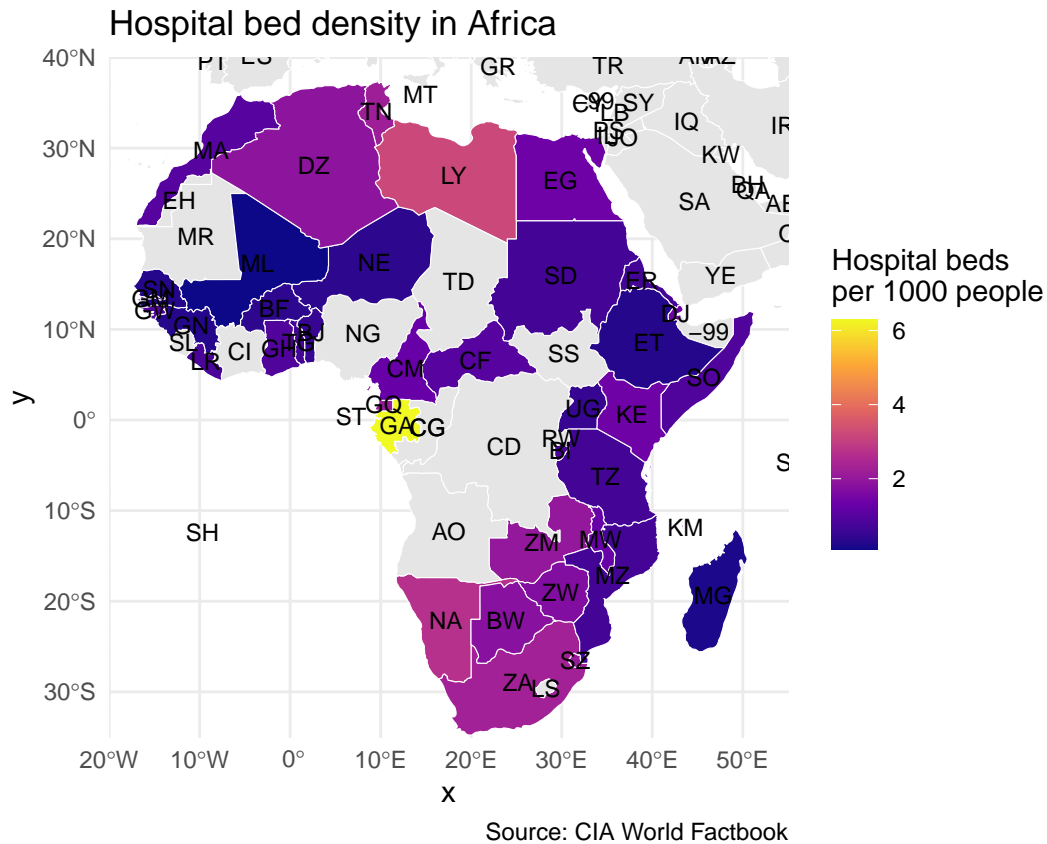
We have less data for Africa on hospital bed density, but what is notable on that graph is that Gabon (GA) appear to be the leading country within this category (approximately 6 beds per 1000 people), followed by Libya (LY) and Equatorial Guinea (GQ), with approximately 4 beds per 1000 people.

Algeria (DZ), Tunisia (TN), Namibia (NA), Egypt (EG), Botswana (BW), Zimbabwe (ZW), Zambia (ZM), South Africa (ZA), Kenya (KE) demonstrate moderate rate of hospital bed density (2-4 beds per 1000 people).

Mostly, countries in Central and West Africa demonstrate lowest rates of the hospital bed density (less than 2 per 1000 people) within Africa region.

Conclusion

To answer our question that was posed at the beginning of this project, there is indeed a negative relationship between maternal and child mortality and access to health services. Exploratory analysis provided proved that access to healthcare infrastructure measured by physician density and hospital bed density is closely related to maternal and infant health outcomes on a global scale. The regional disparities that we observed while exploring our variables highlights the significant role of healthcare accessibility in reducing mortality.

Regions with the lowest physician and hospital bed densities, most notably Africa and some parts of Asia, also demonstrated the highest maternal and infant mortality rates. Conversely, regions such as Europe and the Arctic, which show higher physician coverage and better healthcare infrastructure, report the lowest mortality rates. These findings align well with the existing studies that emphasize the importance of health system capacity in improving population health.