Visualizing African Societies with Data

Name: Masixole Boya

Student number: 1869204

Ranking the African continent relative to other continents

The data gives the average income per person for all countries of the world, and I group these countries into continents. The generalizability of the data, and consequently, of the analysis, is limited to only the year 2021. The relative ranking of the African continent is what I expected to see, to some extent (see **Figure 1**). This is because the African continent is home to the most underdeveloped countries, in terms of economic growth and other factors. Income per person, also GDP per capita, tend to be one of those factors that are used to gauge economic growth. Continents such as Europe, Asia, and North America have more developed infrastructure and industrialization, and so would rank higher than Africa. Africa ranks more than 8 times lower the Europe. However, I did not expect South America, which is home to some very poor countries such as Bolivia, to rank more than double - 2.67 higher, Africa which has diverse economies. The closest continent to Africa still ranks more than double – 2.17, the average income of Africa.

What we deduce from this is that Africa, as much as Africans political leaders tend to claim that Africa is rich in minerals and has the highest natural diversity of both plants and animals, and that it houses the most amount of free land in the world, we are still far behind in economic development compared to other continents. Therefore, you as the youth of the land of Africa, still have a big duty ahead of you, if you wish to realize the true potential of Africa.

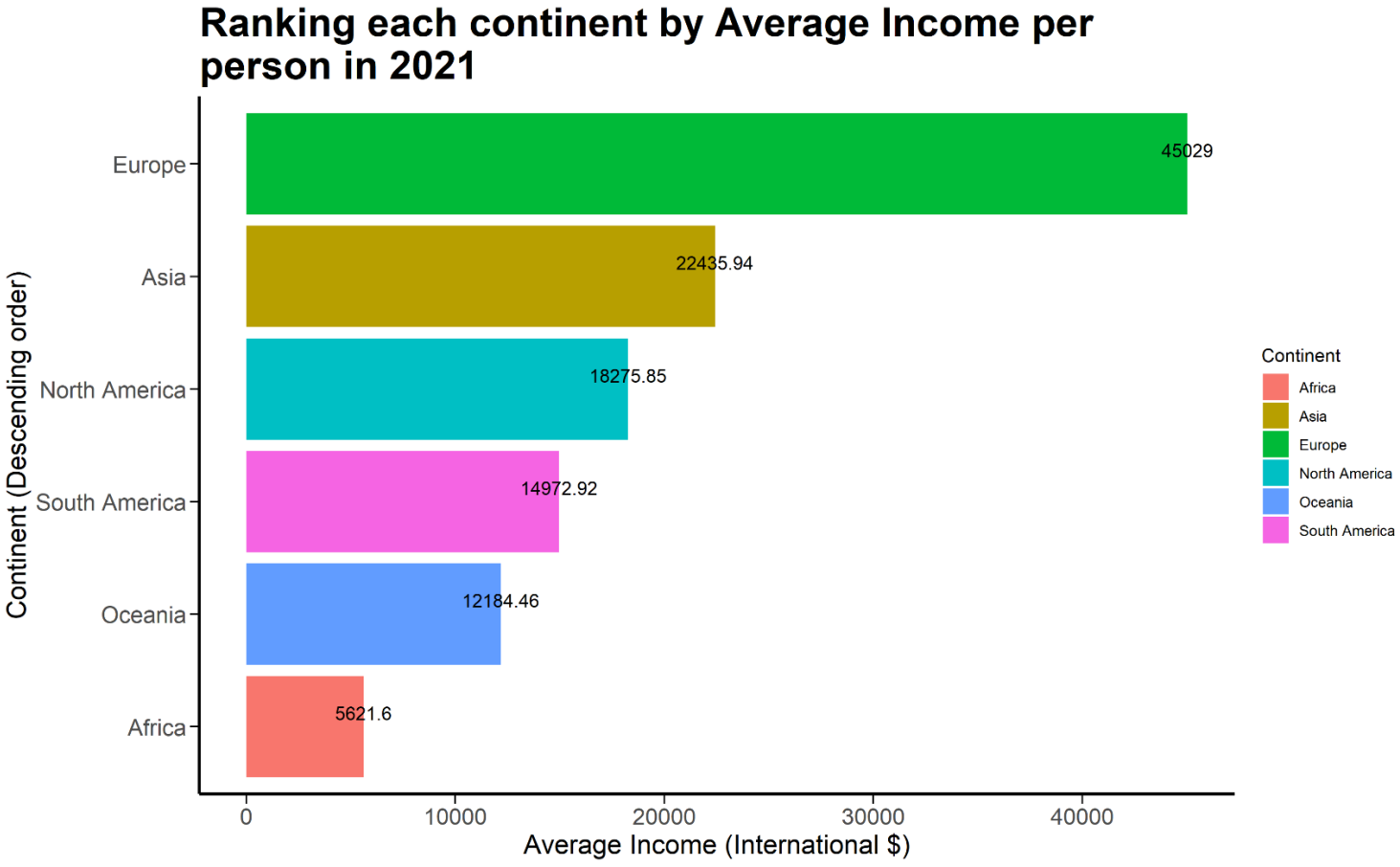

Figure 1: Continents by average income in 2021

R code

The R syntax is written in blue (including comments for the parts that may not be too obvious) and in a different font so that it may be easy for you, as the reader, to see exactly where it starts and end.

```
#Imported the two datasets for the two variables

df_income <- read.csv("C:/Users/STAFF/Desktop/Data Science/Viz/Assignment/Actual_data/cleaned_gdp_pcap.csv")

df_population <- read.csv("C:/Users/STAFF/Desktop/Data Science/Viz/Assignment/Actual_data/urban_population_percent_of_total.csv")


#Installed all necessary packages

install.packages('dplyr')

install.packages('ggplot2')

install.packages('viridis')

install.packages('ggpubr')

install.packages('stringr')


library(dplyr)

library(ggplot2)

library(viridis)

library(ggpubr)

library(stringr)


#Hanlding missing values

missing_values_income <- sapply(df_income, function(x) sum(is.na(x)))

cat("The Sum of missing of the missing values in the Income dataset columns: ", missing_values_income, "\n")


missing_values_population <- sapply(df_population, function(x) sum(is.na(x)))

cat("The sum of missing values in the Population dataset: ", missing_values_population, "\n")


class(df_income) #ensuring it is a dataframe

names(df_income)


#Extracting only the columns for the year 2021

df1 <- df_income[c("country", "X2021")]

df2 <- df_population[c("country", "X2021")]


#merging the two dataframes so I can plot easily
```

```
merged_df <- merge(df1, df2, by = "country")


names(merged_df)
names(merged_df) <- c("country", "df_income$2021", "df_population$2021")
View(merged_df)

# Summary statistics
summary_stats <- summary(merged_df)
summary_stats

# Created a data frame to allow me to easily map the countries to its continents
country_to_continent <- c(
  'Afghanistan' = 'Asia',
  'Angola' = 'Africa',
  'Albania' = 'Europe',
  'Andorra' = 'Europe',
  'UAE' = 'Asia',
  'Argentina' = 'South America',
  'Armenia' = 'Asia',
  'Antigua and Barbuda' = 'North America',
  'Australia' = 'Oceania',
  'Austria' = 'Europe',
  'Azerbaijan' = 'Asia',
  'Burundi' = 'Africa',
  'Belgium' = 'Europe',
  'Benin' = 'Africa',
  'Burkina Faso' = 'Africa',
  'Bangladesh' = 'Asia',
  'Bulgaria' = 'Europe',
  'Bahrain' = 'Asia',
  'Bahamas' = 'North America',
  'Bosnia and Herzegovina' = 'Europe',
  'Belarus' = 'Europe',
  'Belize' = 'North America',
  'Bolivia' = 'South America',
  'Brazil' = 'South America',
```

'Barbados' = 'North America',

'Brunei' = 'Asia',

'Bhutan' = 'Asia',

'Botswana' = 'Africa',

'Central African Republic' = 'Africa',

'Canada' = 'North America',

'Switzerland' = 'Europe',

'Chile' = 'South America',

'China' = 'Asia',

"Cote d'Ivoire" = 'Africa',

'Cameroon' = 'Africa',

'Congo, Dem. Rep.' = 'Africa',

'Congo, Rep.' = 'Africa',

'Colombia' = 'South America',

'Comoros' = 'Africa',

'Cape Verde' = 'Africa',

'Costa Rica' = 'North America',

'Cuba' = 'North America',

'Cyprus' = 'Asia',

'Czech Republic' = 'Europe',

'Germany' = 'Europe',

'Djibouti' = 'Africa',

'Dominica' = 'North America',

'Denmark' = 'Europe',

'Dominican Republic' = 'North America',

'Algeria' = 'Africa',

'Ecuador' = 'South America',

'Egypt' = 'Africa',

'Eritrea' = 'Africa',

'Spain' = 'Europe',

'Estonia' = 'Europe',

'Ethiopia' = 'Africa',

'Finland' = 'Europe',

'Fiji' = 'Oceania',

'France' = 'Europe',

'Micronesia, Fed. Sts.' = 'Oceania',

'Gabon' = 'Africa',

'UK' = 'Europe',

'Georgia' = 'Asia',

'Ghana' = 'Africa',

'Guinea' = 'Africa',

'Gambia' = 'Africa',

'Guinea-Bissau' = 'Africa',

'Equatorial Guinea' = 'Africa',

'Greece' = 'Europe',

'Grenada' = 'North America',

'Guatemala' = 'North America',

'Guyana' = 'South America',

'Hong Kong, China' = 'Asia',

'Honduras' = 'North America',

'Croatia' = 'Europe',

'Haiti' = 'North America',

'Hungary' = 'Europe',

'Indonesia' = 'Asia',

'India' = 'Asia',

'Ireland' = 'Europe',

'Iran' = 'Asia',

'Iraq' = 'Asia',

'Iceland' = 'Europe',

'Israel' = 'Asia',

'Italy' = 'Europe',

'Jamaica' = 'North America',

'Jordan' = 'Asia',

'Japan' = 'Asia',

'Kazakhstan' = 'Asia',

'Kenya' = 'Africa',

'Kyrgyz Republic' = 'Asia',

'Cambodia' = 'Asia',

'Kiribati' = 'Oceania',

'St. Kitts and Nevis' = 'North America',

'South Korea' = 'Asia',

'Kuwait' = 'Asia',

'Lao' = 'Asia',

'Lebanon' = 'Asia',

'Liberia' = 'Africa',

'Libya' = 'Africa',

'St. Lucia' = 'North America',

'Sri Lanka' = 'Asia',

'Lesotho' = 'Africa',

'Lithuania' = 'Europe',

'Luxembourg' = 'Europe',

'Latvia' = 'Europe',

'Morocco' = 'Africa',

'Monaco' = 'Europe',

'Moldova' = 'Europe',

'Madagascar' = 'Africa',

'Maldives' = 'Asia',

'Mexico' = 'North America',

'Marshall Islands' = 'Oceania',

'North Macedonia' = 'Europe',

'Mali' = 'Africa',

'Malta' = 'Europe',

'Myanmar' = 'Asia',

'Montenegro' = 'Europe',

'Mongolia' = 'Asia',

'Mozambique' = 'Africa',

'Mauritania' = 'Africa',

'Mauritius' = 'Africa',

'Malawi' = 'Africa',

'Malaysia' = 'Asia',

'Namibia' = 'Africa',

'Niger' = 'Africa',

'Nigeria' = 'Africa',

'Nicaragua' = 'North America',

'Netherlands' = 'Europe',

'Norway' = 'Europe',

'Nepal' = 'Asia',

'Nauru' = 'Oceania',

'New Zealand' = 'Oceania',

'Oman' = 'Asia',

'Pakistan' = 'Asia',

'Panama' = 'North America',

'Peru' = 'South America',

'Philippines' = 'Asia',

'Palau' = 'Oceania',

'Papua New Guinea' = 'Oceania',

'Poland' = 'Europe',

'North Korea' = 'Asia',

'Portugal' = 'Europe',

'Paraguay' = 'South America',

'Palestine' = 'Asia',

'Qatar' = 'Asia',

'Romania' = 'Europe',

'Russia' = 'Europe',

'Rwanda' = 'Africa',

'Saudi Arabia' = 'Asia',

'Sudan' = 'Africa',

'Senegal' = 'Africa',

'Singapore' = 'Asia',

'Solomon Islands' = 'Oceania',

'Sierra Leone' = 'Africa',

'El Salvador' = 'North America',

'Somalia' = 'Africa',

'Serbia' = 'Europe',

'South Sudan' = 'Africa',

'Sao Tome and Principe' = 'Africa',

'Suriname' = 'South America',

'Slovak Republic' = 'Europe',

'Slovenia' = 'Europe',

'Sweden' = 'Europe',

'Eswatini' = 'Africa',

'Seychelles' = 'Africa',

'Syria' = 'Asia',

'Chad' = 'Africa',

```
  'Togo' = 'Africa',
  'Thailand' = 'Asia',
  'Tajikistan' = 'Asia',
  'Turkmenistan' = 'Asia',
  'Timor-Leste' = 'Asia',
  'Tonga' = 'Oceania',
  'Trinidad and Tobago' = 'North America',
  'Tunisia' = 'Africa',
  'Turkey' = 'Asia',
  'Tuvalu' = 'Oceania',
  'Taiwan' = 'Asia',
  'Tanzania' = 'Africa',
  'Uganda' = 'Africa',
  'Ukraine' = 'Europe',
  'Uruguay' = 'South America',
  'USA' = 'North America',
  'Uzbekistan' = 'Asia',
  'St. Vincent and the Grenadines' = 'North America',
  'Venezuela' = 'South America',
  'Vietnam' = 'Asia',
  'Vanuatu' = 'Oceania',
  'Samoa' = 'Oceania',
  'Yemen' = 'Asia',
  'South Africa' = 'Africa',
  'Zambia' = 'Africa',
  'Zimbabwe' = 'Africa'
)

#Putting the "continent" column into the merged dataframe
merged_df_with_continent <- left_join(merged_df, data.frame(country = names(country_to_continent), continent = country_to_continent), by = "country")
names(merged_df_with_continent)

names(merged_df_with_continent)[2] <- "Income_2021"
names(merged_df_with_continent)[3] <- "Population_2021"

View(merged_df_with_continent)
```

```
continent_data <- summarise(
  group_by(merged_df_with_continent, continent),
  Average_Income = mean(`Income_2021`),
  Average_Population = mean(`Population_2021`)
)
continent_data <- continent_data[!is.na(continent_data$continent), ]

# Arranging the data in terms of the Average income, descending order
continent_data <- arrange(continent_data, desc(Average_Income))

# Created a bar plot to rank the continents
ggplot(data = continent_data, aes(x = reorder(continent, Average_Income), y = Average_Income, fill = continent)) +
  geom_bar(stat = "identity", position = "dodge") +
  geom_text(aes(label = round(Average_Income, 2)), vjust = -0.5, size = 4) +
  labs(x = "Continent (Descending order)", y = "Average Income (International $)", fill = "Continent") +
  theme_minimal()+
  theme(panel.grid = element_blank(),axis.line = element_line(color = "black", linewidth = 1),
        plot.title = element_text(size = 24, face = "bold"),
        axis.title = element_text(size = 16),
        axis.text.x = element_text(size = 14, hjust = 0.5),
        axis.text.y = element_text(size = 14, vjust = 0.5),
        axis.ticks = element_line(color = "black"),
        axis.ticks.length = unit(0.2, "cm"))+
  coord_flip() + ggtitle(str_wrap("Ranking each continent by Average Income per person in 2021", width = 50))
```

The relationship between income and urban population for African Countries

The relationship between income per person and people going to live in urban areas is not random but is rather significant and moderately strong one (see **Figure 2**). The data represents only these given African countries and so is not generalized to countries of other continents. The relationship observed is also about African countries. The “income per person” refers to the average income earned in the year 2021 per person in a certain African country. The urban population refers to the percentage of the population that leaves in urban areas. About 42 in every 100 people who move into urban areas do so due to the increase in income. This goes with what I had expected to see, but not completely. The reason I expected this is because of the common behavior of many people in South Africa, most especially in the Eastern Cape where I stayed for many years, who have lived in rural areas move from rural into urban living as soon as they can afford a better life. I inferred this observation to the whole country and in fact, the whole continent.

I expected to see about 80 in every 100 people, because almost everyone who is not in urban areas wishes they were living in urban areas. There are countries that do not seem to follow this trend, the ones with high average income but less than 50% of urban population and the ones with high urban population (up to 70%) but have low average income. I believe that for the ones with high average income and low urban population, they may have either very expensive urban areas that most people do not afford or has a high percentage of elderly people who tend to find peace when away from the business of some urban districts. What we deduce from this is that in most countries, people living in non-urban areas (rural or informal settlements) would definitely move as soon as they get a chance. Therefore, in the quest of the South African government to end informal settlements like shacks, the real answer might be to give people jobs so they can afford to move into urban areas.

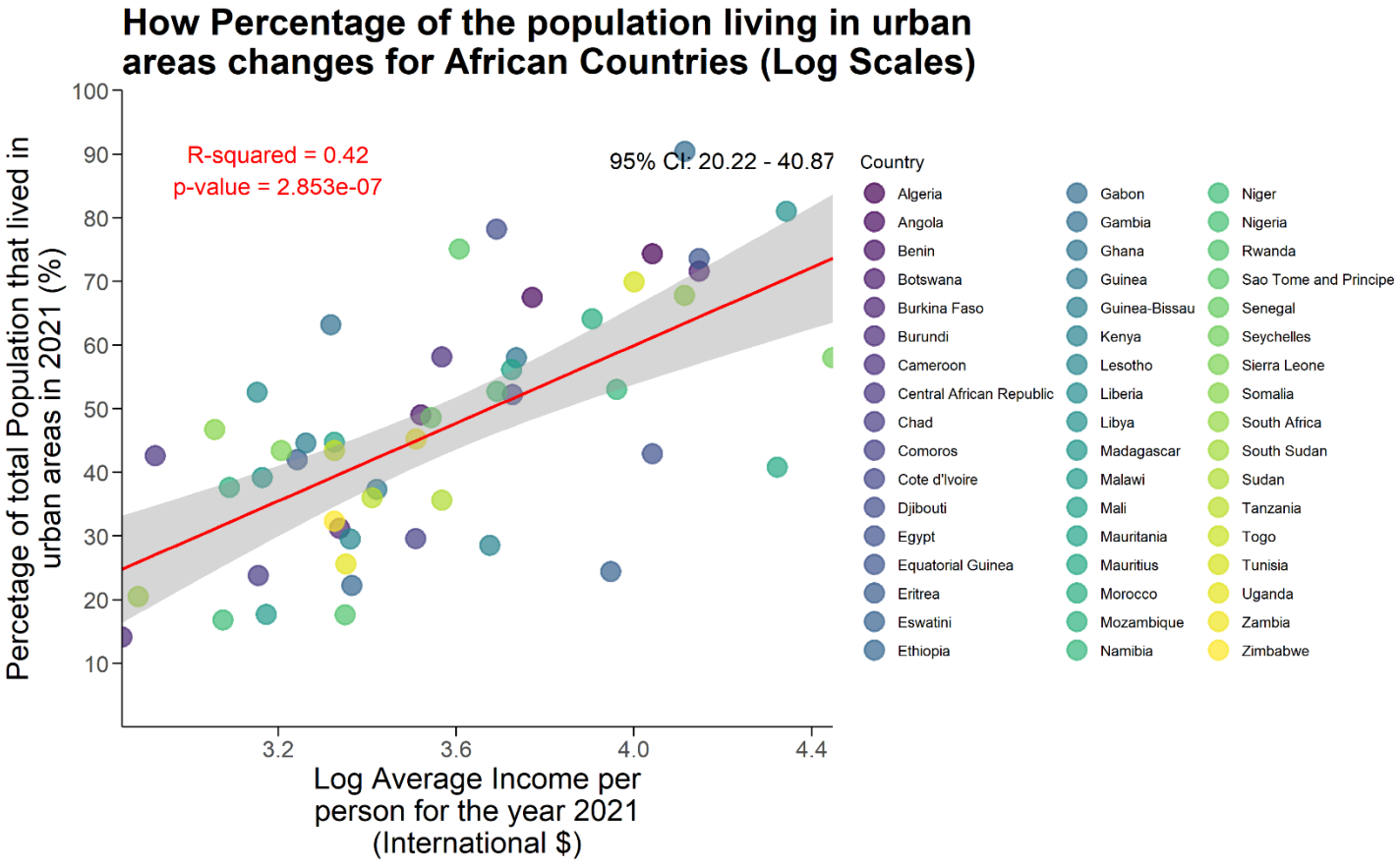

Figure 2: The relationship between urban population and average income per person in Africa

R code

```
install.packages('dplyr')

install.packages('ggplot2')

install.packages('viridis')

install.packages('ggpubr')

install.packages('stringr')

install.packages("RColorBrewer")


library(dplyr)

library(ggplot2)

library(viridis)
```

```
library(ggpubr)
library(stringr)
library(RColorBrewer)

names(merged_df) <- c('country', 'Income_2021', 'Population_2021')

merged_df$Income_2021 <- log10(merged_df$Income_2021)
merged_df$Population_2021 <- merged_df$Population_2021

african_countries <- c(
  'Algeria', 'Angola', 'Benin', 'Botswana', 'Burkina Faso', 'Burundi', 'Cabo Verde', 'Cameroon', 'Central African Republic', 'Chad',
  'Comoros', 'Congo, Dem. Rep.', 'Congo, Rep.', 'Cote d\'Ivoire', 'Djibouti', 'Egypt', 'Equatorial Guinea', 'Eritrea', 'Eswatini',
  'Ethiopia', 'Gabon', 'Gambia', 'Ghana', 'Guinea', 'Guinea-Bissau', 'Kenya', 'Lesotho', 'Liberia', 'Libya', 'Madagascar', 'Malawi',
  'Mali', 'Mauritania', 'Mauritius', 'Morocco', 'Mozambique', 'Namibia', 'Niger', 'Nigeria', 'Rwanda', 'Sao Tome and Principe', 'Senegal',
  'Seychelles', 'Sierra Leone', 'Somalia', 'South Africa', 'South Sudan', 'Sudan', 'Tanzania', 'Togo', 'Tunisia', 'Uganda', 'Zambia',
  'Zimbabwe'
)

african_data_2021 <- merged_df[merged_df$country %in% african_countries, ]

lm_model <- lm(Population_2021 ~ Income_2021, data = african_data_2021)
rsquared <- summary(lm_model)$r.squared
pvalue <- format(summary(lm_model)$coefficients[2, 4], digits = 4)

ci <- confint(lm_model)

# Extracted the lower and upper bounds of the confidence intervals for Population_2021
ci_lower <- ci["Income_2021", "2.5 %"]
ci_upper <- ci["Income_2021", "97.5 %"]



ggplot(data = african_data_2021, aes(x = Income_2021, y = Population_2021, color = country)) +
  geom_point(alpha = 0.7, size = 5, stroke = 1) +
  geom_smooth(method = 'lm', se = TRUE, color = 'red', size = 1) + # Add correlation line
  scale_color_viridis(discrete = TRUE) +
  labs(title = str_wrap('How Percentage of the population living in urban areas changes for African Countries (Log Scales)',width = 50),
```

```
    x = str_wrap('Log Average Income per person for the year 2021 (International $)',width = 30),
    y = str_wrap('Percetage of total Population that lived in urban areas in 2021 (%)',width = 45)) +
theme_minimal() +
theme(legend.position = 'right',axis.line = element_line(color = "black"), plot.title = element_text(size = 22, face = "bold"),
    panel.grid.major = element_blank(),panel.grid.minor = element_blank(),
    axis.text.x = element_text(size = 14, hjust = 0.5),
    axis.text.y = element_text(size = 14, vjust = 0.5),
    axis.ticks = element_line(color = "black"),
    axis.ticks.length = unit(0.2, "cm")) +
guides(color = guide_legend(title = 'Country')) +
scale_x_continuous(limits = c(min(african_data_2021$Income_2021), max(african_data_2021$Income_2021))) +
scale_y_continuous(limits = c(0,100), breaks = c(10,20,30,40,50,60,70,80,90,100)) +
coord_cartesian(expand = FALSE) +
theme(axis.title = element_text(size = 18), axis.text = element_text(size = 10))+
annotate("text", x = 3.2, y = 90, label = paste("R-squared =", round(rsquared, 2)), size = 5, color = 'red') +
annotate("text", x = 3.2, y = 85, label = paste("p-value =", pvalue), size = 5, color = 'red')+
annotate("text", x = 4.2, y = 85, label = paste("95% CI:", round(ci_lower, 2), "-", round(ci_upper, 2)), size = 5, vjust = -1)
```

Spatial dimensions of urban population for African countries

The data file for plotting and the data file containing the values of the income had a few mismatches where some countries would be present in one file and not on the other. Such countries appear on the map as white spaces. The first catergory range, as given by the key, also appear white and so the distinction between the first category countries and the “empty” countries might not be easy. However, an easy way to distinguish between these is by looking if a country is named. If it is not named then there is no data for that country but if it is named but appears just white, then it simply falls within the first catergory range.

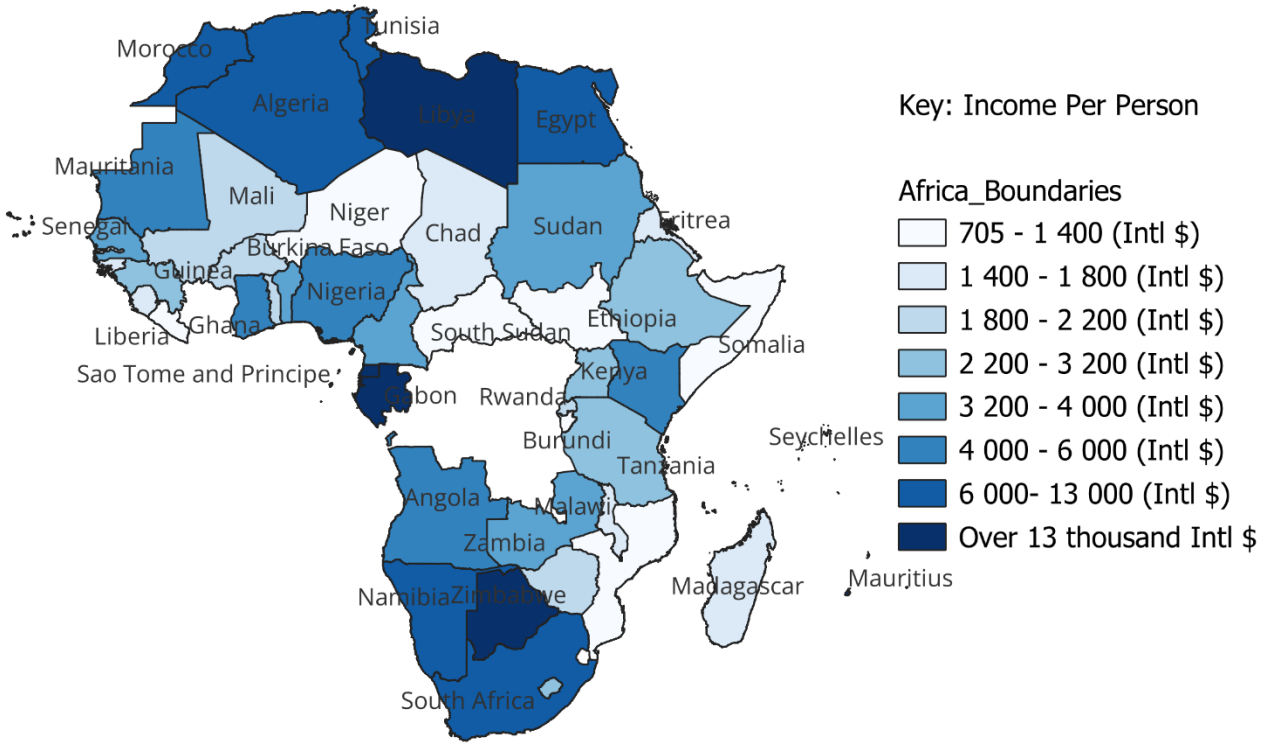

Figure 3: The map of Africa showing the spatial dimensions of the Income per person variable. The corresponding range for each country is given by the key in the figure.