UNIVERSITY OF WESTMINSTER

INFORMATICS INSTITUTE OF TECHNOLOGY

5BUIS006C Data Visualisation and Communication

PORTFOLIO

Name: Tania G. Motha

UOW ID: w2052145

IIT No: 20230908

Contents

0

ACKNOWLEDGEMENTS 2

INTRODUCTION 3

DATA PREPARATIONS 4

EXPLORATORY DATA ANALYSIS 6

Univariate Analysis 6

Multivariate Analysis 10

DATA STORYTELLING 14

ACKNOWLEDGEMENTS

I would like to extend my gratitude to all individuals involved in preparing this college report. A special thanks go to Ms Yaalini Balathasan and Mr Fouzol Hassan for their invaluable support throughout the research process and for assisting in answering the course questions and completing the report. I am also grateful to my classmates for sharing their insights and for their help during the coursework. Thank you to everyone who contributed to the success of this report.

INTRODUCTION

Research Question

How does GDP per capita correlate with the ecological footprint across different countries?

Relevance: In an era where the balance between economic development and environmental sustainability is more critical than ever, understanding the relationships and connections between various socio-economic and ecological factors becomes vital. This analysis will provide insights on how ecological growth impacts environmental health. This knowledge is essential for those who aim to promote sustainable development practices that would ensure economic growth without compromising the planet's ecological integrity.

Data Used: The dataset used for this analysis includes variables such as GDP per Capita, Total Ecological Footprint, Population (millions), Region, Human Development Index (HDI), Cropland Footprint, Grazing Footprint, Forest Footprint, Carbon Footprint, Fish Footprint, Total Biocapacity, Biocapacity Deficit or Reserve, Earths Required, and Countries Required.

This analysis is aimed to uncover patterns and relationships that can contribute to a deep understanding of ecological and environmental factors at play.

Dataset Sourcing:

- Dataset name- 2016 Global Ecological Footprint
- Link- https: //www.kaggle.com/datasets/footprintnetwork/ecological-footprint

DATA PREPARATIONS

Data preparations for this analysis were performed using R. Commands like read.csv(), na.omit(), and gsub() were used to clean and prepare the data effectively.

First, the csv file was loaded into R using the method outlined below, and the data frame was assigned to a variable

```
> dataset <- read.csv("C:/Users/User/Desktop/Data VIS/countries.csv")
> class(dataset)
[1] "data.frame"
> head(dataset)
              Country                  Region Population  HDI GDP.per.Capita Cropland.Footprint Grazing.Footprint
1         Afghanistan Middle East/Central Asia      29.82 0.46       $614.66               0.30               0.20
2             Albania  Northern/Eastern Europe       3.16 0.73     $4,534.37               0.78               0.22
3             Algeria                  Africa      38.48 0.73     $5,430.57               0.60               0.16
4              Angola                  Africa      20.82 0.52     $4,665.91               0.33               0.15
5 Antigua and Barbuda           Latin America       0.09 0.78    $13,205.10                 NA                 NA
6           Argentina           Latin America      41.09 0.83    $13,540.00               0.78               0.79
  Forest.Footprint Carbon.Footprint Fish.Footprint Total.Ecological.Footprint Cropland Grazing.Land Forest.Land
1             0.08             0.18           0.00                       0.79     0.24         0.20        0.02
2             0.25             0.87           0.02                       2.21     0.55         0.21        0.29
3             0.17             1.14           0.01                       2.12     0.24         0.27        0.03
4             0.12             0.20           0.09                       0.93     0.20         1.42        0.64
5               NA               NA             NA                       5.38       NA           NA          NA
6             0.29             1.08           0.10                       3.14     2.64         1.86        0.66
  Fishing.Water Urban.Land Total.Biocapacity Biocapacity.Deficit.or.Reserve Earths.Required Countries.Required
1          0.00       0.04              0.50                         -0.30            0.46               1.60
2          0.07       0.06              1.18                         -1.03            1.27               1.87
3          0.01       0.03              0.59                         -1.53            1.22               3.61
4          0.26       0.04              2.55                          1.61            0.54               0.37
5            NA         NA              0.94                         -4.44            3.11               5.70
6          1.67       0.10              6.92                          3.78            1.82               0.45
  Data.Quality
1            6
2            6
3            5
4            6
5            2
6            6
> str(dataset)
'data.frame':   188 obs. of  21 variables:
 $ Country                       : Factor w/ 188 levels "Afghanistan",..: 1 2 3 4 5 6 7 8 9 10 ...
 $ Region                        : Factor w/ 7 levels "Africa","Asia-Pacific",..: 5 7 1 1 4 4 5 4 2 3 ...
 $ Population                    : num  29.82 3.16 38.48 20.82 0.09 ...
```

```
> #Remove Duplicated Rows
> dataset <- dataset[!duplicated(dataset), ]
> #Remove Unwanted Columns
> columns_to_keep <- c("Country", "Region", "Population", "HDI", "GDP.per.Capita",
+                      "Cropland.Footprint", "Grazing.Footprint", "Forest.Footprint",
+                      "Carbon.Footprint", "Fish.Footprint", "Total.Ecological.Footprint",
+                      "Total.Biocapacity", "Biocapacity.Deficit.or.Reserve", "Earths.Required",
+                      "Countries.Required")
> dataset <- dataset[, columns_to_keep]
> # Convert GDP per Capita to numeric by removing $ symbol and commas
> dataset$GDP.per.Capita <- as.numeric(gsub("[\\$,]", "", dataset$GDP.per.Capita))
> # Save the Cleaned Data
> write.csv(dataset, "C:/Users/User/Desktop/Data VIS/cleaned_countries.csv", row.names = FALSE)
> #View the first few rows of the cleaned data
> head(dataset)
     Country                  Region Population  HDI GDP.per.Capita Cropland.Footprint Grazing.Footprint
1 Afghanistan Middle East/Central Asia      29.82 0.46         614.66               0.30               0.20
2     Albania  Northern/Eastern Europe       3.16 0.73        4534.37               0.78               0.22
3     Algeria                  Africa      38.48 0.73        5430.57               0.60               0.16
4      Angola                  Africa      20.82 0.52        4665.91               0.33               0.15
6   Argentina           Latin America      41.09 0.83       13540.00               0.78               0.79
7     Armenia Middle East/Central Asia       2.97 0.73        3426.39               0.74               0.18
  Forest.Footprint Carbon.Footprint Fish.Footprint Total.Ecological.Footprint Total.Biocapacity
1             0.08             0.18           0.00                       0.79              0.50
2             0.25             0.87           0.02                       2.21              1.18
3             0.17             1.14           0.01                       2.12              0.59
4             0.12             0.20           0.09                       0.93              2.55
6             0.29             1.08           0.10                       3.14              6.92
7             0.34             0.89           0.01                       2.23              0.89
  Biocapacity.Deficit.or.Reserve Earths.Required Countries.Required
1                          -0.30            0.46               1.60
2                          -1.03            1.27               1.87
3                          -1.53            1.22               3.61
4                           1.61            0.54               0.37
6                           3.78            1.82               0.45
7                          -1.35            1.29               2.52
> |
```

All rows with missing values were cleaned to ensure data integrity, and then the dataset was checked for duplications and those were removed to avoid redundancy.

```
 $ HDI                          : num  0.46 0.73 0.73 0.52 0.78 0.83 0.73 NA 0.93 0.88 ...
 $ GDP.per.Capita               : Factor w/ 174 levels "","$1,016.83 ",..: 141 91 118 94 35 38 71 1 143 127 ...
 $ Cropland.Footprint           : num  0.3 0.78 0.6 0.33 NA 0.78 0.74 NA 2.68 0.82 ...
 $ Grazing.Footprint            : num  0.2 0.22 0.16 0.15 NA 0.79 0.18 NA 0.63 0.27 ...
 $ Forest.Footprint             : num  0.08 0.25 0.17 0.12 NA 0.29 0.34 NA 0.89 0.63 ...
 $ Carbon.Footprint             : num  0.18 0.87 1.14 0.2 NA 1.08 0.89 NA 4.85 4.14 ...
 $ Fish.Footprint               : num  0 0.02 0.01 0.09 NA 0.1 0.01 NA 0.11 0.06 ...
 $ Total.Ecological.Footprint   : num  0.79 2.21 2.12 0.93 5.38 ...
 $ Cropland                     : num  0.24 0.55 0.24 0.2 NA 2.64 0.44 NA 5.42 0.71 ...
 $ Grazing.Land                 : num  0.2 0.21 0.27 1.42 NA 1.86 0.26 NA 5.81 0.16 ...
 $ Forest.Land                  : num  0.02 0.29 0.03 0.64 NA 0.66 0.1 NA 2.01 2.04 ...
 $ Fishing.Water                : num  0 0.07 0.01 0.26 NA 1.67 0.02 NA 3.19 0 ...
 $ Urban.Land                   : num  0.04 0.06 0.03 0.04 NA 0.1 0.07 NA 0.14 0.15 ...
 $ Total.Biocapacity            : num  0.5 1.18 0.59 2.55 0.94 ...
 $ Biocapacity.Deficit.or.Reserve: num  -0.3 -1.03 -1.53 1.61 -4.44 ...
 $ Earths.Required              : num  0.46 1.27 1.22 0.54 3.11 1.82 1.29 6.86 5.37 3.5 ...
 $ Countries.Required           : num  1.6 1.87 3.61 0.37 5.7 ...
 $ Data.Quality                 : Factor w/ 7 levels "2","3B","3L",..: 7 7 6 7 1 7 2 1 6 6 ...
> #indentifying the columns of the dataset
> names(dataset)
 [1] "Country"                       "Region"                        "Population"
 [4] "HDI"                           "GDP.per.Capita"                "Cropland.Footprint"
 [7] "Grazing.Footprint"             "Forest.Footprint"              "Carbon.Footprint"
[10] "Fish.Footprint"                "Total.Ecological.Footprint"    "Cropland"
[13] "Grazing.Land"                  "Forest.Land"                   "Fishing.Water"
[16] "Urban.Land"                    "Total.Biocapacity"             "Biocapacity.Deficit.or.Reserve"
[19] "Earths.Required"               "Countries.Required"            "Data.Quality"
> #cleaning and preperation of the dataset
> dataset <- na.omit(dataset)
> #Remove Duplicated Rows
> dataset <- dataset[!duplicated(dataset), ]
> #Remove Unwanted Columns
> columns_to_keep <- c("Country", "Region", "Population", "HDI", "GDP.per.Capita",
+                      "Cropland.Footprint", "Grazing.Footprint", "Forest.Footprint",
+                      "Carbon.Footprint", "Fish.Footprint", "Total.Ecological.Footprint",
+                      "Total.Biocapacity", "Biocapacity.Deficit.or.Reserve", "Earths.Required",
```

After studying the research question, relevant variables were identified, and the unnecessary variables were cleaned. The selected variable for this study gives a comprehensive view on the relationship between economic and ecological impacts. Key variables include GDP per Capita, Human Development Index (HDI), and various ecological footprints (Cropland, Grazing, Forest, Carbon, Fish), which measure the environmental impact of human activities. Population size and regional context is needed to understand the scale and geographical distribution of the impacts. Total Biocapacity and Biocapacity Deficit or Reserve are crucial to access the sustainability of resource use, while Earths Required and Countries Required offer stark illustrations of global resource consumption. Together, all the variables enable a analysis that would provide straightforward insights into the research.

The Variable GDP was recognized to be the independent variable, and this column was converted to a numeric column by removing the $ symbol and commas to facilitate numerical analysis.

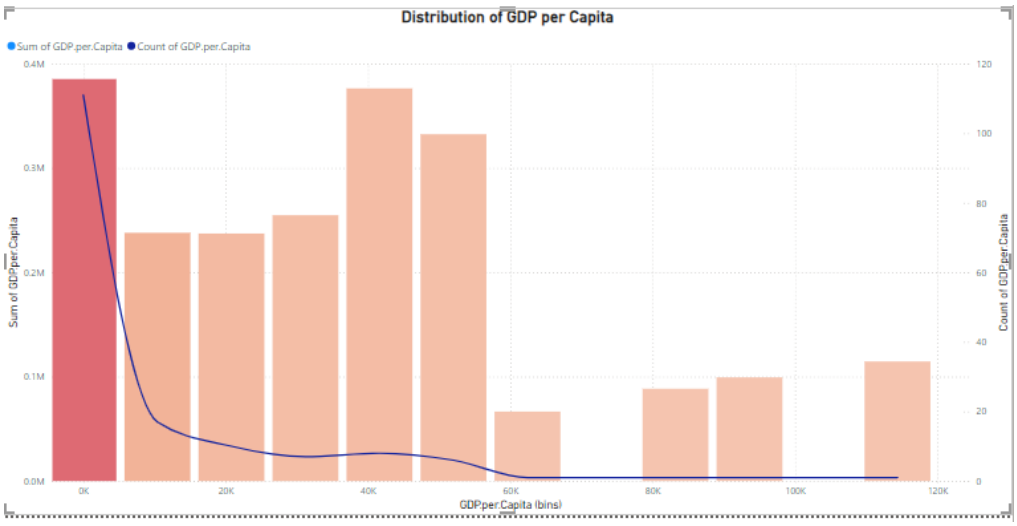
The cleaning process was completed, and the cleaned data was saved into a new file.

EXPLORATORY DATA ANALYSIS

Univariate Analysis

This approach demonstrates the various patterns and trends of a single variable, and gives insights.

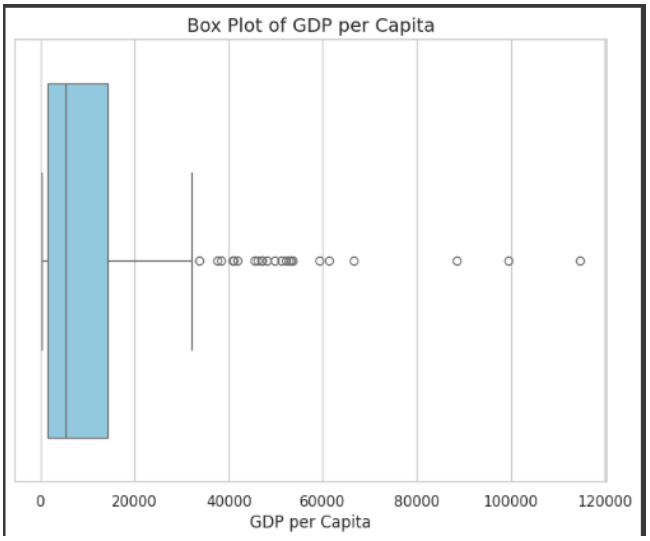
1. **Distribution of GDP per capita**



The above graph demonstrates the distribution of GDP per capita. It is noticeable that most countries fall within the lower GDP per capita bins, indicating that most countries have relatively modest economic output. The curve shows a downward trend, suggesting that as the GDP increases the number of countries decreases, highlighting the economic disparity. While a small number of countries enjoy a good economy majority of other nations operate within lower income.

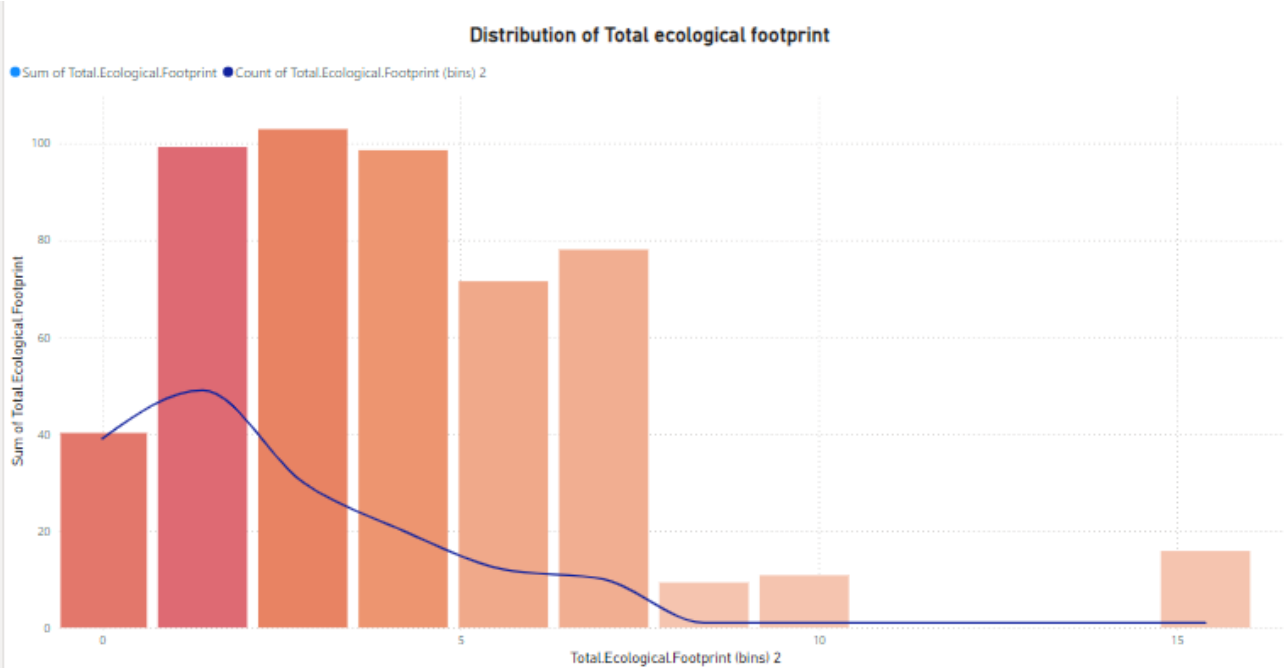
2. **Box plot of GDP per capita**

Note: Google Colab is used to generate the graph as some features weren't supported in my version of Power BI.



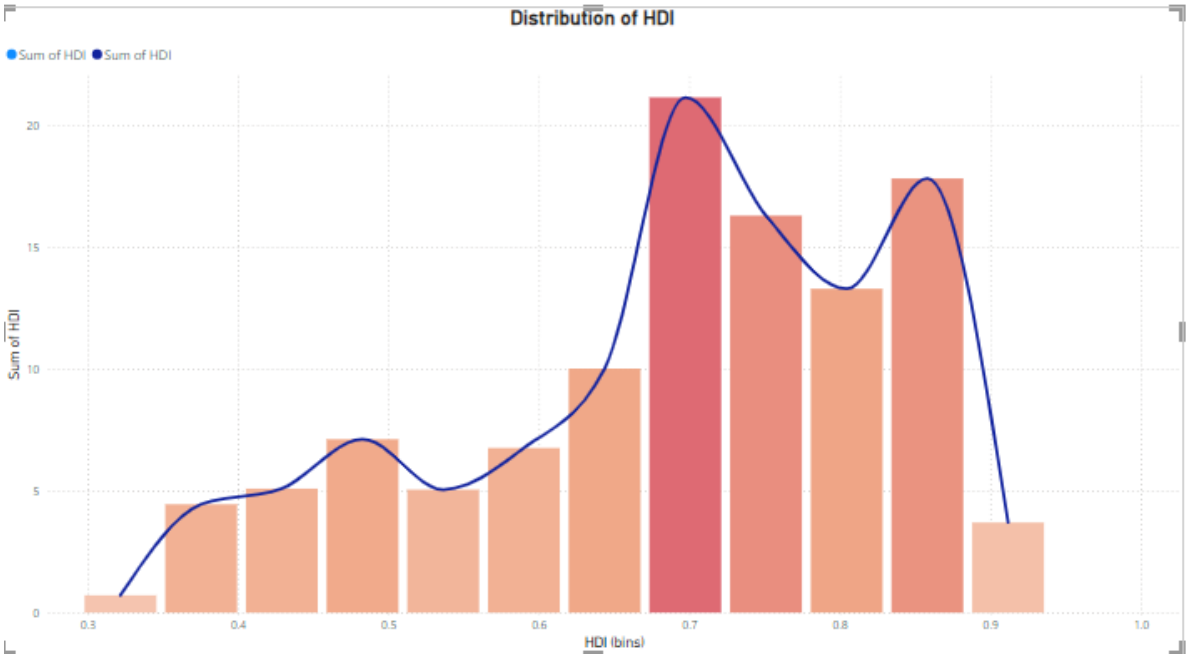
The distribution appears to be positively skewed, with a larger concentration of countries having low GDP values. Additionally, the length of the whiskers indicates variability among the countries.

3. Distribution of total ecological footprint



The Graph is skewed to the right, this suggests that only a few countries contribute substantially to the total ecological footprint. This reflects the uneven global distribution of ecological demand. It is likely that those countries with higher industrial activities contribute to the disproportion to the ecology.

4. Sum of HDI (Bins) vs. HDI

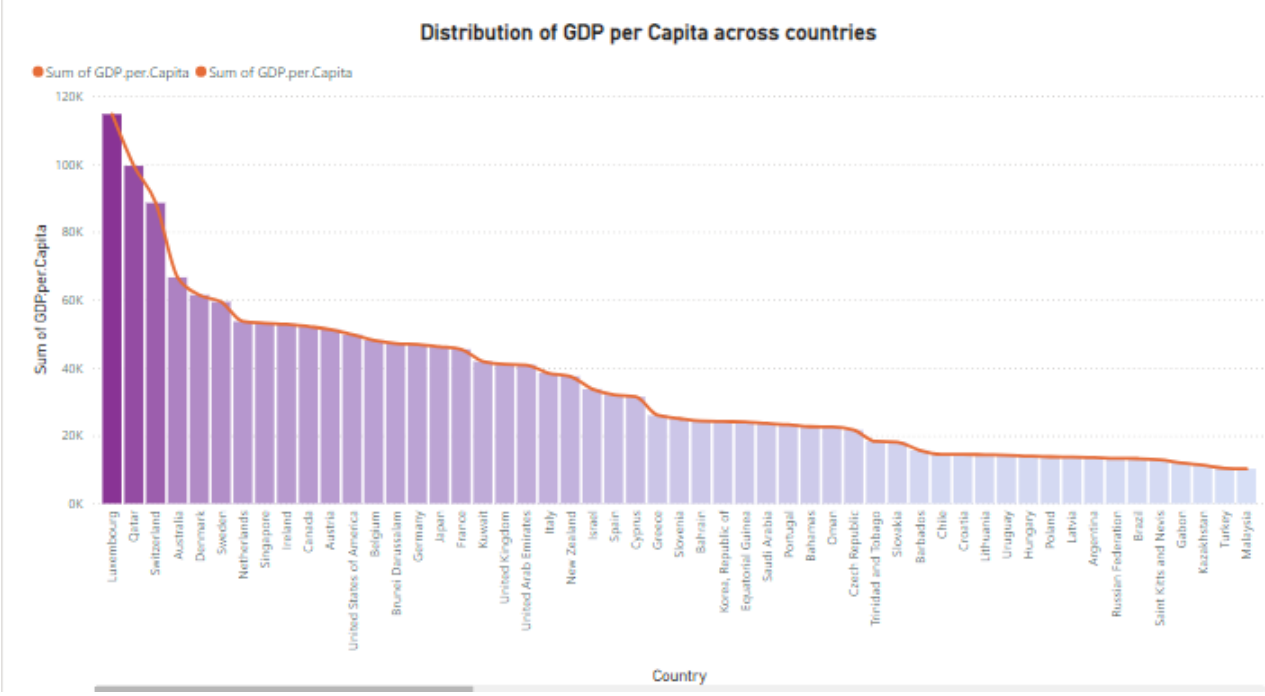


In this graph it is noticeable that there is a drop within the lower HDI bins, this indicates that few countries experience a low level of development. Conversely, the 0.7 – 0.9 there's a moderate representation. Overall, the graph highlights the global push towards human development.

Multivariate Analysis

Multivariate analysis will provide deeper insights uncovering trends and patterns by analysing the relationship between multiple variables.

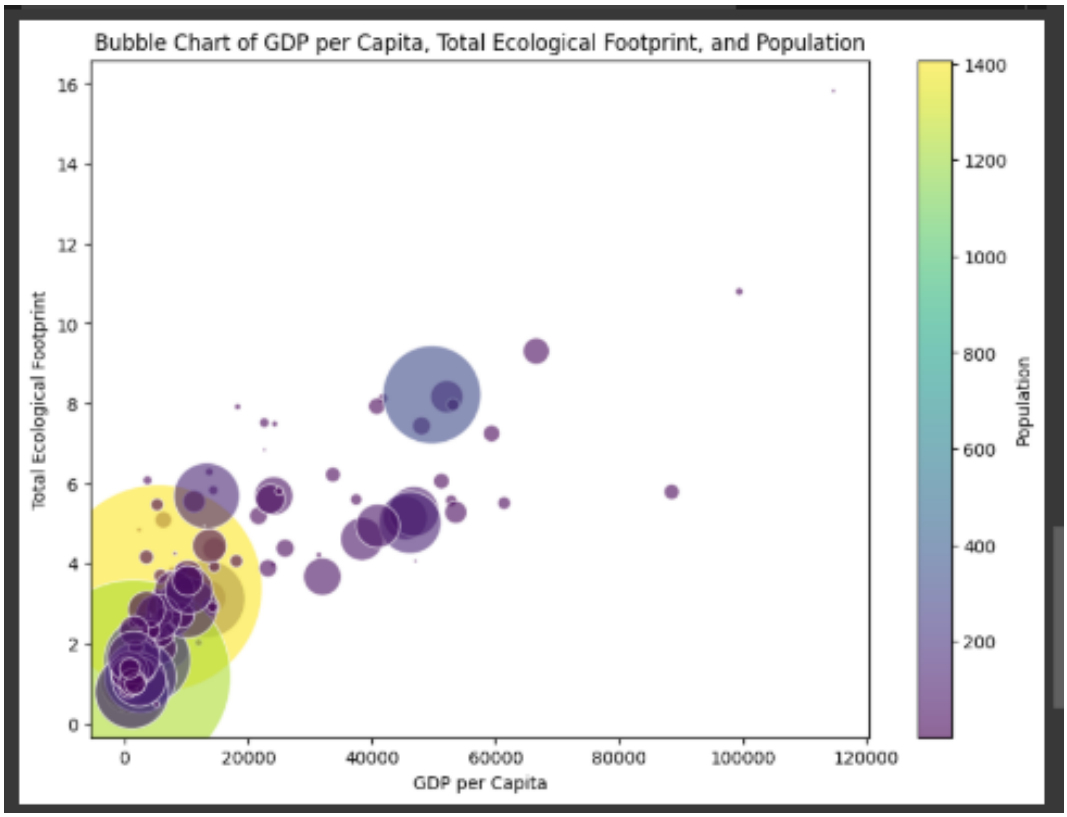
1. **Distribution of GDP per capita across countries**



This graph displays the distribution of GDP per capita across counties, providing an understanding of the range and frequency of GDP per capita values. We can see that, while some countries like Luxembourg, Qatar, Switzerland have a very high sum of GDP per capita, countries like Burundi, Bangladesh have a very low sum of GDP per capita. Showing the economic gap between the countries.

2. 3D Scatter Plot of GDP per Capita, Total Ecological Footprint, and HDI

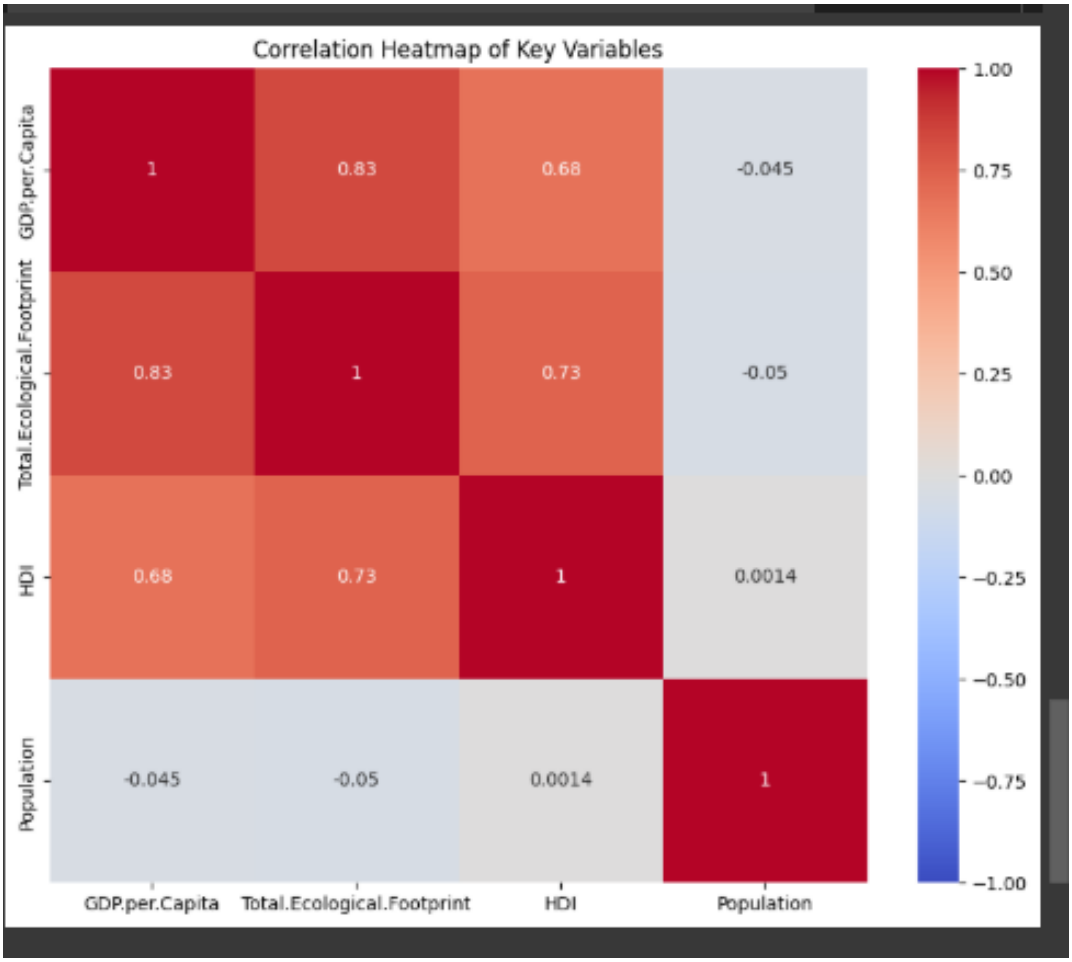
Note: Google Colab is used to generate the graph as some features weren't supported in my version of Power BI.



This bubble chart reveals key insights into the relationship between ecological development, ecological impact and population size. The chart visually displays that, countries with higher GDP per capita is mostly exhibiting larger ecological footprint, this indicates a correlation between economic prosperity and resource consumption. However, some outliers seem to display high ecological footprint regardless of low GDP per capita, suggesting inefficient resource use or due to high population. Large yellow green bubbles representing countries with high populations, are distributed across a range of GDP per capita and ecological footprint, this highlights the diversity of populated nations.

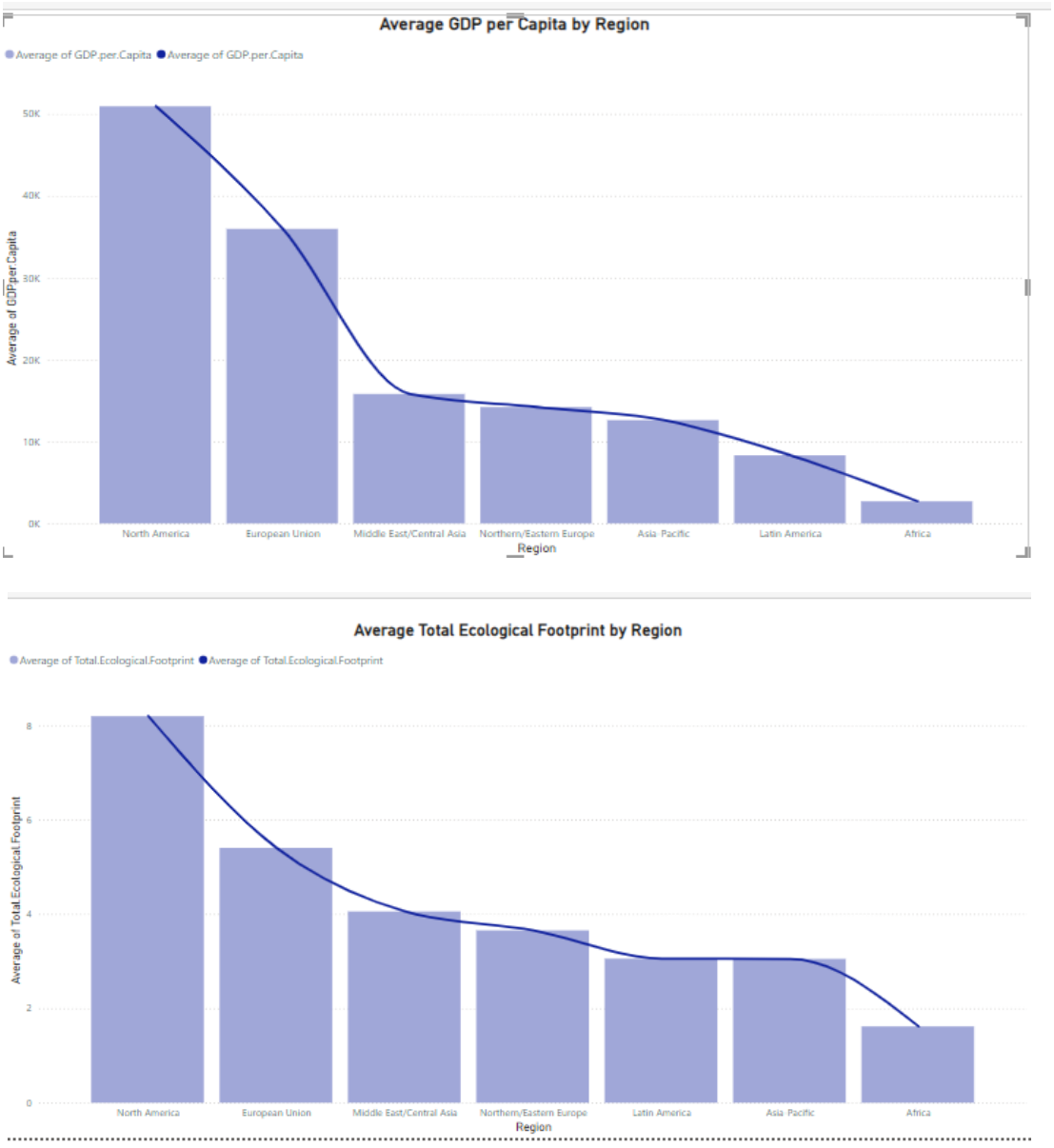
3. Correlation Heatmap of Key Variables

Note: Google Colab is used to generate the graph as some features weren't supported in my version of Power BI.



The heatmaps shows a quantitative overview of the interconnection between the variables 'GDP per Capita', 'Total Ecological Footprint', 'HDI', 'Population'. Notably, GDP per capita shows a positive correlation between HDI depicting that higher economic development often aligns with improvement in human development. Similarly, GDP per capita and total ecological footprint have a strong correlation, reflecting that economic growth is associated with resource consumption. It is noticeable that the relationship between population and other variables is weaker, this suggests that the population size is not the only factor that directly drives economical or ecological metrics.

4. **Combined analysis of Average GDP per capita by region and Average Total ecological footprint by region**

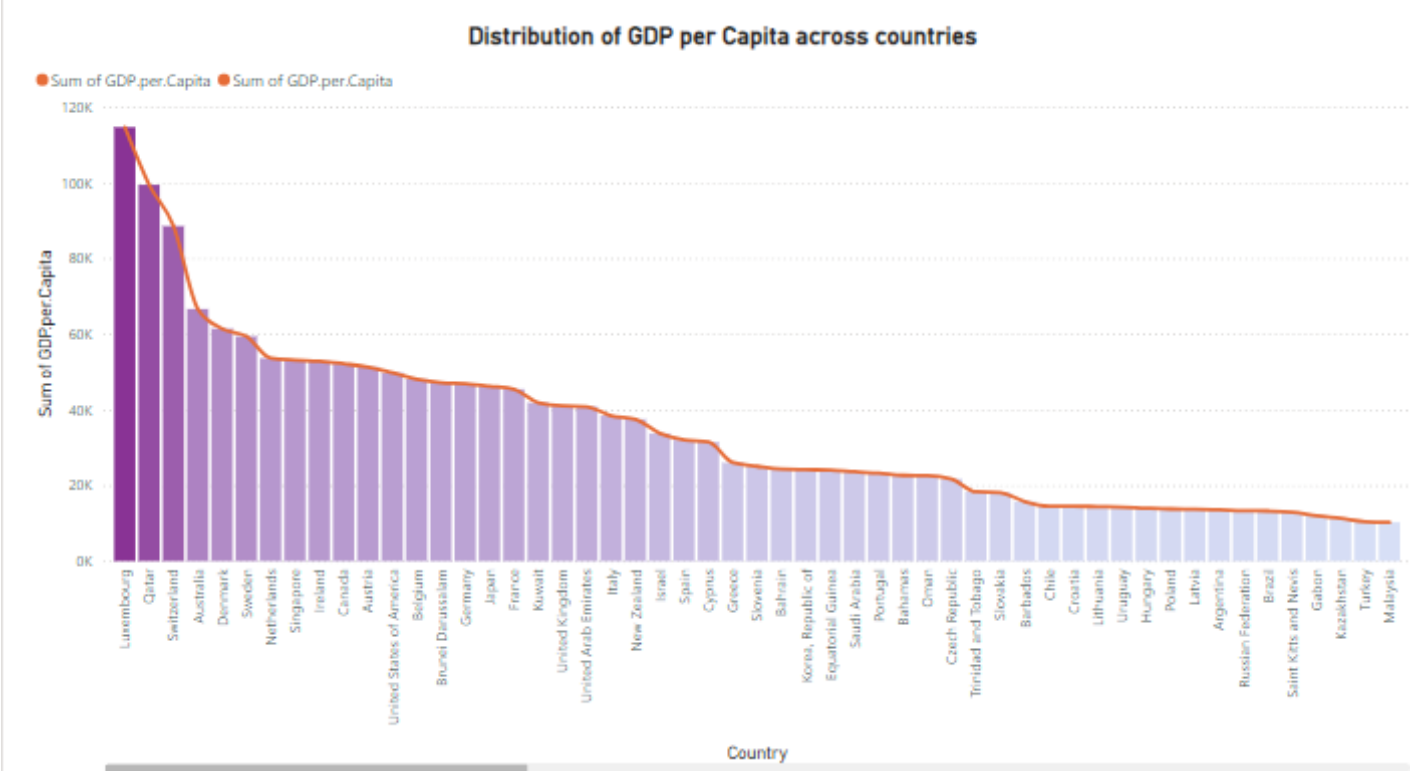


This graph provides a clear comparison of economic performance across different Region. North American region exhibit the highest average GDP, this shows that this region has significant economic development and resource availability. In contrast African regions have a low GDP average, highlighting the disparities in economic prosperity. Also, the regions with higher GDP average have higher average total ecological footprints, reflecting the resource- intensive nature of economically developed regions. By noting the similarity of the curve on both graphs, we can say that wealthier regions; while contributing significantly to global GDP, they also bear a larger share of the ecological footprint.

DATA STORYTELLING

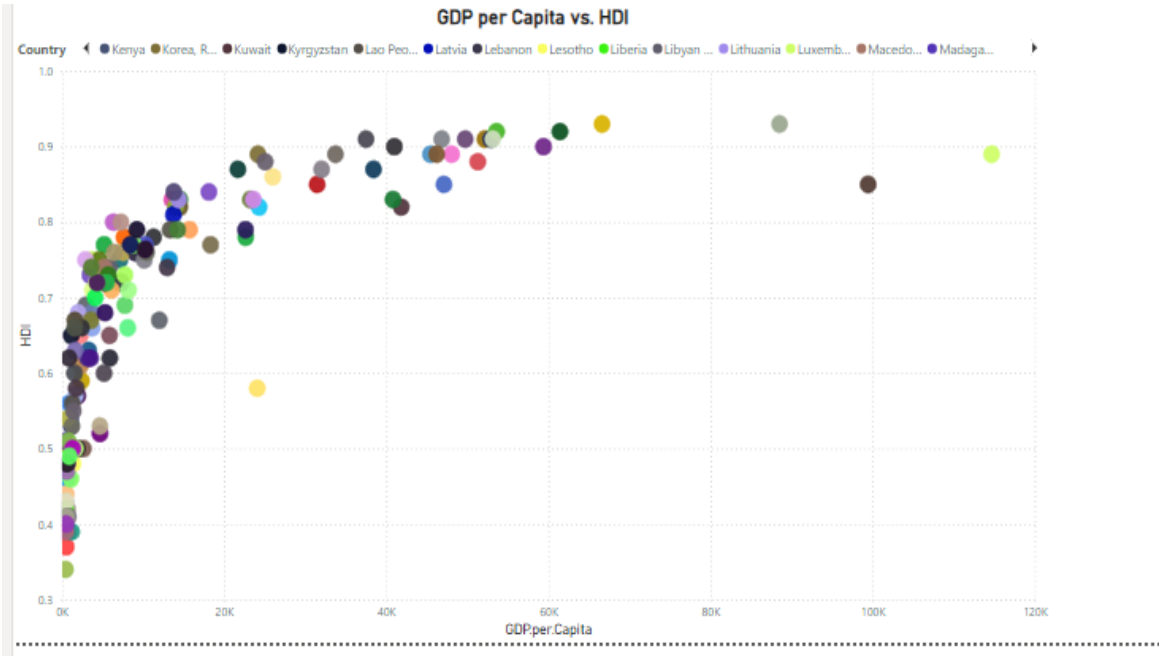
A Global Perspective on Development and Sustainability

We're in a world where the wealth of certain nations can power entire continents while other countries struggle a great deal to meet their basic needs. This analysis travels into the intricate relationships between GDP per capita, human development, ecological sustainability and population. Through the insights we gained from the visualisations, we uncover the disparities and challenges shaping our global landscape.

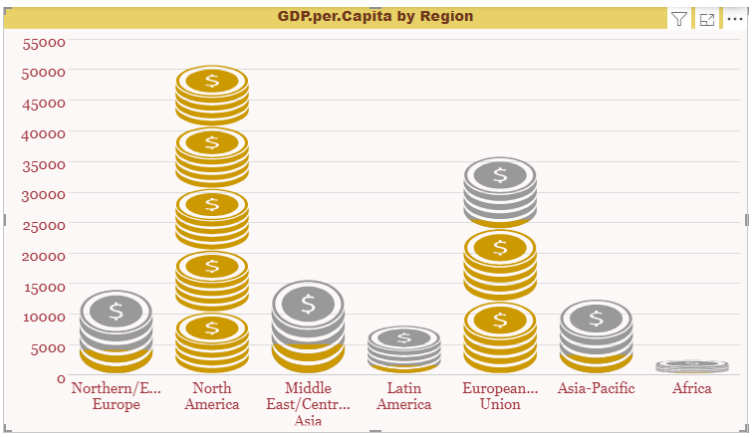


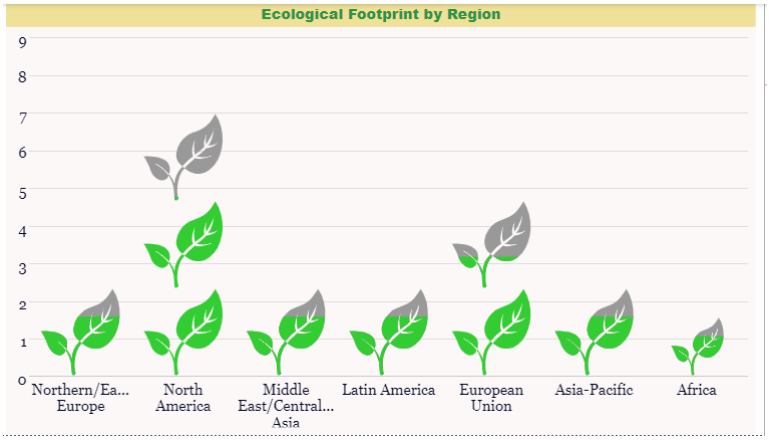
Rising prosperity: GDP per capita and development

Let's look at how nations fare economically. For instance, Luxembourg, has a GDP per capita above $100,000, with a HDI of almost 0.9. All while countries like Burundi have a GDP per capita below $300 and a HDI near 0.4, reflecting severe developmental challenges. This allows the wealthier nations to invest towards better healthcare, education, and quality of life and keep improving their development while the low economic countries struggle to develop.

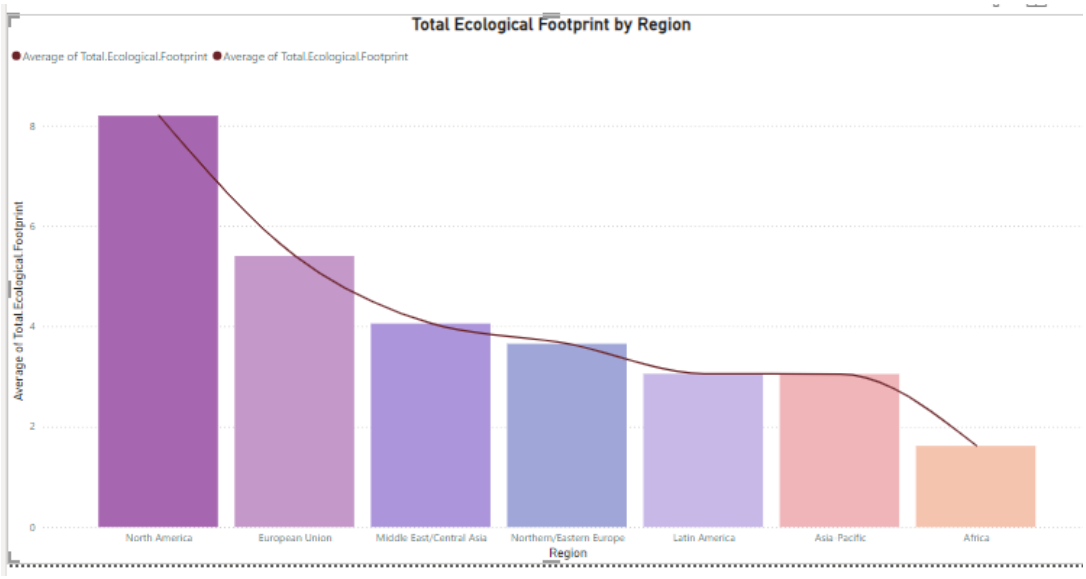


The price of wealth: Ecological footprint by region.

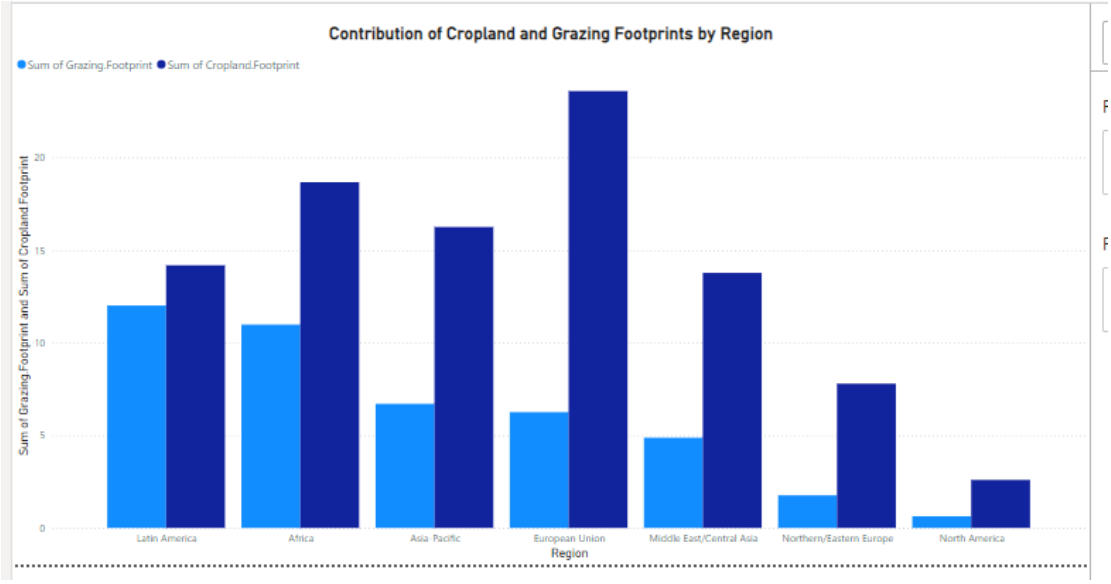




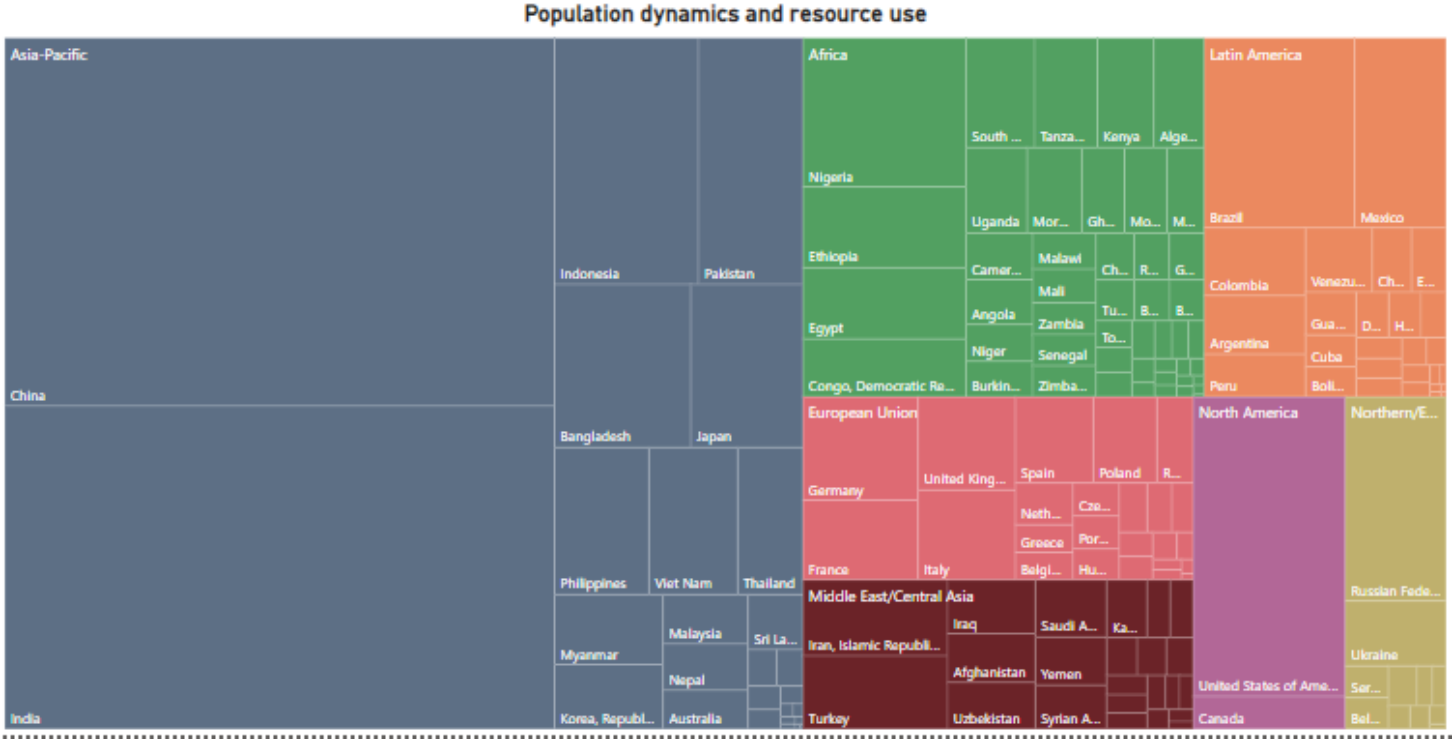
There's an environmental cost for economic prosperity. North American region for instance, has the highest carbon emission rate, with countries such as the US accounting for a high share in that. Meanwhile, we know that African region's ecological footprint largely stems from agricultural practices, such as cropland and grazing, driven by farming. These observations show how different economies have different types of demands on the planet's resources, this raises concern on sustainability.

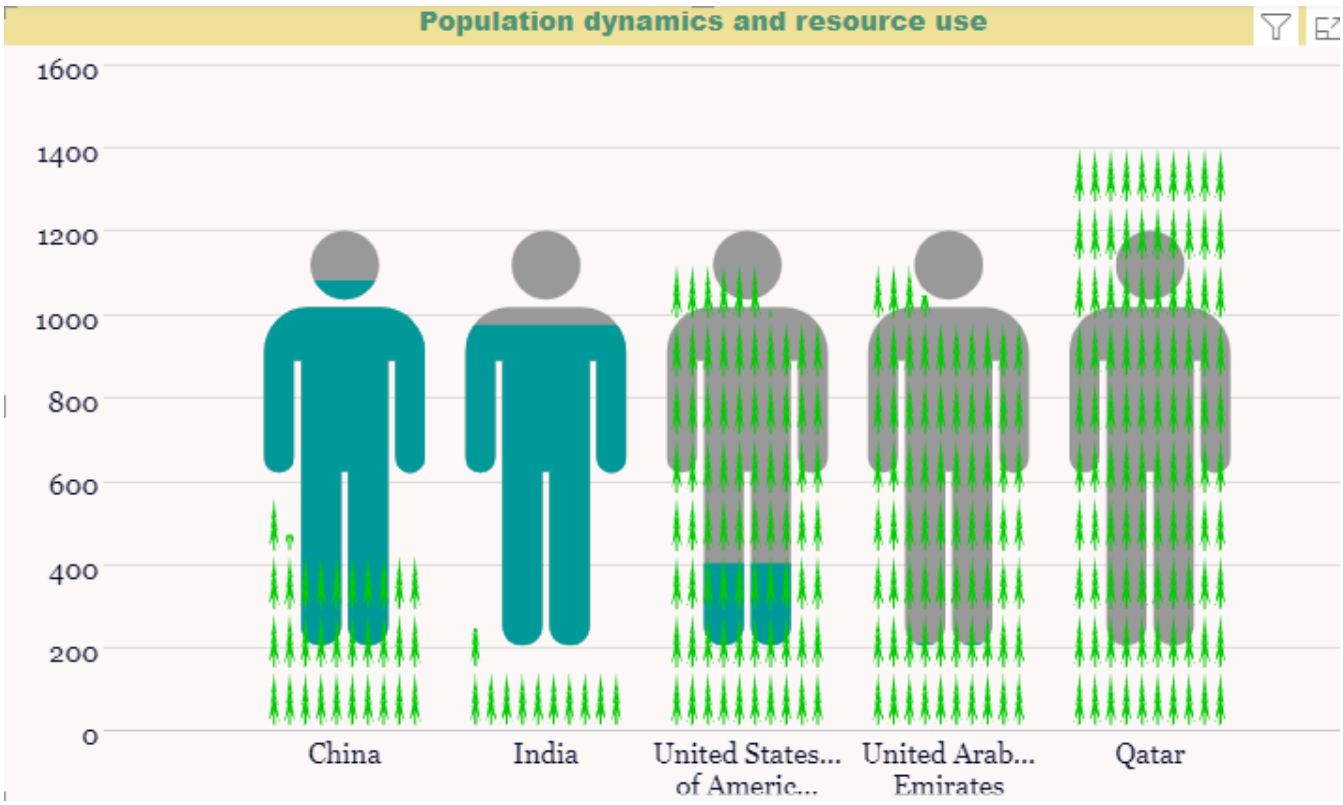


The graph below confirms that Africa has a high contribution of cropland and grazing footprints. The cropland footprint dominating, highlighting the relience on agriculture.



But are these findings relative to the population size of countries or regions? Let's explore on the matter further.


Population dynamics and resource use
Asia-Pacific
China
India
Indonesia
Pakistan
Bangladesh
Japan
Philippines
Viet Nam
Thailand
Myanmar
Malaysia
Nepal
Korea, Republ...
Australia
Sri La...
Africa
Nigeria
Ethiopia
Egypt
Congo, Democratic Re...
South ...
Tanza...
Kenya
Alge...
Uganda
Mor...
Gh...
Mo...
M...
Camer...
Angola
Niger
Burkin...
Malawi
Mali
Zambia
Senegal
Zimba...
Ch...
R...
G...
Tu...
B...
To...
Latin America
Brazil
Mexico
Colombia
Argentina
Peru
Venezu...
Gua...
Cuba
Boli...
Ch...
E...
D...
H...
European Union
Germany
France
United King...
Italy
Spain
Poland
R...
Neth...
Greece
Belgi...
Cze...
Por...
Hu...
North America
United States of Ame...
Canada
Northern/E...
Russian Fede...
Ukraine
Ser...
Bel...
Middle East/Central Asia
Iran, Islamic Republi...
Turkey
Iraq
Afghanistan
Uzbekistan
Saudi A...
Yemen
Syrian A...
Ka...


Population dynamics and resource use
1600
1400
1200
1000
800
600
400
200
0
China
India
United States... of Americ...
United Arab... Emirates
Qatar

A crowded world: Population dynamics and resource use

The tree map shows the contrast between populous nations like India and China and the smaller countries like Switzerland. In the map it is visible that despite the massive population in China and India, they display a moderate ecoligical footprint. On the other hand, countries like Qatar and the UAE, exhibit high ecological footprint eventhough they have relatively smaller populations. This observation highlights how resource-intensive lifestyles contribute more significantly to impact the ecology than the sheer number of people.

Finally, from the stability of Luxembourg to the challenges of Burundi, this analysis gives a picture of the inequality when it comes to resources, economy, sustainability and development

We observed that economic prosperity often comes with the cost of environmental health, with wealthier countries like the US bearing a significant amount of responsibility. However, since the recent times humans are focusing on sustainability more. So, there is hope that we can balance development along with sustainability.