Report: Exploratory Data Analysis

Applied Economics & Data Analysis in R (Econ 680/880, Dr. Venoo Kakar)

Marcus Nogueira

About the data

Dataset Description and Justification

For this exploratory data analysis project, I have chosen the "Cost of Living Index by Country" dataset. This dataset is from Kaggle! This dataset provides a comparative and detailed overview of living costs across various countries, encompassing indices such as Cost of Living, Rent, Cost of Living Plus Rent, Groceries, Restaurant Prices, and Local Purchasing Power. These indices are invaluable for understanding and comparing the relative affordability and purchasing power in different regions globally.

The primary reasons for selecting this dataset are:

1. **Relevance:** The dataset offers crucial insights into global economic trends, which are highly relevant in today's interconnected world. Understanding living costs and purchasing power across different countries can help inform decisions for individuals, businesses, and policymakers.
2. **Comprehensive Data:** With indices covering various aspects of living expenses, the dataset allows for a thorough analysis of cost related factors and their interrelationships, providing a holistic view of economic conditions worldwide.
3. **Comparative Analysis:** The dataset's structure facilitates comparative analysis, enabling the identification of countries with the highest and lowest living costs, as well as those with the strongest and weakest local purchasing power.
4. **Policy Implications:** The insights derived from this dataset can assist in policy formulation, helping governments and organizations address economic challenges and develop strategies for cost of living adjustments and economic development.

Overall, this dataset serves as a valuable resource for exploring global economic disparities and understanding the dynamics of living costs and purchasing power across different regions.

Variables and Descriptions:

1. **Country:**
   - **Description:** Name of the Country for which the indices are calculated
2. **Cost of Living Index:**
   - **Description:** Measures the average cost of basic necessities
3. **Rent Index:**
   - **Description:** Average cost of renting a house or apartment in the country
4. **Cost of Living Plus Rent Index:**
   - **Description:** Combines both the Cost of Living and Rent indices to provide detail view of total living expenses
5. **Groceries Index:**
   - **Description:** Measures the average cost of food and other grocery items
6. **Restaurant Price Index:**
   - **Description:** Compares the average cost of dining out at restaurants
7. **Local Purchasing Power Index:**
   - **Description:** Relative purchasing power of locals based on their average salary and cost of living in the country

Guiding Research Questions

1. What are the most and least expensive countries to live in based on the Cost of Living Index and Rent Index?
2. How does the Local Purchasing Power Index vary across countries, and what factors seem to influence it the most?
3. What is the relationship between the Cost of Living Plus Rent Index and the Groceries Index and Restaurant Price Index across different regions?

Exploratory Data Analysis

```r
#Loading Data and Pathing it
#| message: false
rm(list = ls())
main.data <- read.csv('/Users/marcusnogueira/Library/Mobile Documents/com~apple~CloudDocs/81
head(main.data, n = 10)
```

```
            Country Cost.of.Living.Index Rent.Index
1       Switzerland                101.1       46.5
2           Bahamas                 85.0       36.7
3           Iceland                 83.0       39.2
4         Singapore                 76.7       67.2
5          Barbados                 76.6       19.0
6            Norway                 76.0       26.2
7           Denmark                 72.3       26.4
8  Hong Kong (China)                 70.8       59.4
9      United States                 70.4       41.7
10         Australia                 70.2       33.4
   Cost.of.Living.Plus.Rent.Index Groceries.Index Restaurant.Price.Index
1                            74.9           109.1                    97.0
2                            61.8            81.6                    83.3
3                            62.0            88.4                    86.8
4                            72.1            74.6                    50.4
5                            48.9            80.8                    69.4
6                            52.1            79.0                    73.5
7                            50.2            64.8                    81.3
8                            65.3            84.6                    46.2
9                            56.6            75.0                    67.2
10                           52.5            77.3                    62.5
   Local.Purchasing.Power.Index
1                         158.7
2                          54.6
3                         120.3
4                         111.1
5                          43.5
6                         114.7
7                         127.2
8                         109.3
9                         142.3
10                        127.4
```

```r
# Load necessary libraries
library(ggplot2)
library(dplyr)
```

```
Attaching package: 'dplyr'
```

```
The following objects are masked from 'package:stats':

    filter, lag
```

```
The following objects are masked from 'package:base':

    intersect, setdiff, setequal, union
```

```r
library(corrplot)
```

```
corrplot 0.92 loaded
```

First Visualization

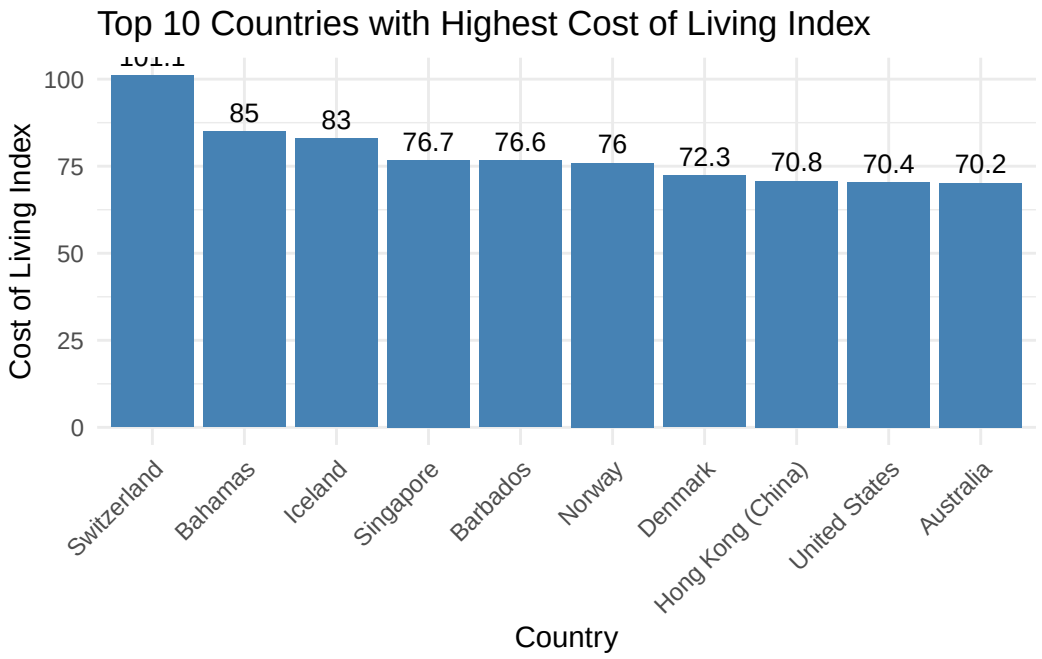
Bar Plot of Top 10 Countries with Highest Cost of Living Index

```r
#Finding top 10 countries with the High cost of living index
top.10.cost.living <- main.data |>
  arrange(desc(Cost.of.Living.Index))
head(top.10.cost.living, n = 10)
```

```
            Country Cost.of.Living.Index Rent.Index
1       Switzerland                101.1       46.5
2           Bahamas                 85.0       36.7
3           Iceland                 83.0       39.2
4         Singapore                 76.7       67.2
5          Barbados                 76.6       19.0
6            Norway                 76.0       26.2
7           Denmark                 72.3       26.4
8  Hong Kong (China)                 70.8       59.4
9     United States                  70.4       41.7
```

```
10         Australia                  70.2       33.4
   Cost.of.Living.Plus.Rent.Index Groceries.Index Restaurant.Price.Index
1                            74.9           109.1                    97.0
2                            61.8            81.6                    83.3
3                            62.0            88.4                    86.8
4                            72.1            74.6                    50.4
5                            48.9            80.8                    69.4
6                            52.1            79.0                    73.5
7                            50.2            64.8                    81.3
8                            65.3            84.6                    46.2
9                            56.6            75.0                    67.2
10                           52.5            77.3                    62.5
   Local.Purchasing.Power.Index
1                         158.7
2                          54.6
3                         120.3
4                         111.1
5                          43.5
6                         114.7
7                         127.2
8                         109.3
9                         142.3
10                        127.4
```

```r
#First Visualization: Bar Plot
# Create the bar plot
ggplot(head(top.10.cost.living, 10), aes(x = reorder(Country, -Cost.of.Living.Index), y = Cos
  geom_bar(stat = "identity", fill = "steelblue") +
  geom_text(aes(label = Cost.of.Living.Index), vjust = -0.5, color = "black", size = 3.5) +
  labs(
    title = "Top 10 Countries with Highest Cost of Living Index",
    x = "Country",
    y = "Cost of Living Index"
  ) +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```




```
# Calculate summary statistics
summary.stats <- top.10.cost.living |>
  summarise(
    Avg.Cost.of.Living = mean(Cost.of.Living.Index),
    Min.Cost.of.Living = min(Cost.of.Living.Index),
    Max.Cost.of.Living = max(Cost.of.Living.Index)
)
print(summary.stats)
```

```
  Avg.Cost.of.Living Min.Cost.of.Living Max.Cost.of.Living
1           43.55537               18.8              101.1
```

Visualization 1: Top 10 Countries with Highest Cost of Living Index

This bar plot illustrates the top 10 countries with the highest cost of living index. The countries are ranked in descending order based on their cost of living index. This visualization helps identify the countries where living expenses are the highest, which can be crucial for expatriates, businesses, and policymakers.

Key Insights:

- **Average Cost of Living Index:** The average cost of living index among these top 10 countries is 43.56.

- **Highest Cost of Living:** Switzerland has the highest cost of living index, valued at 101.1, indicating it is the most expensive country to live in among the top 10.
- **Lowest Cost of Living:** The country with the lowest cost of living index among the top 10 is Australia, with a value of 18.8.
- **Diverse Representation:** The top 10 list includes countries from various continents, highlighting that high living costs are a global phenomenon.

Second Visualization

Comparison of Cost of Living Indices by Continent

```
#Finding Unique values to map countries to continet
unique.countries <- unique(main.data$Country)
print(unique.countries)
```

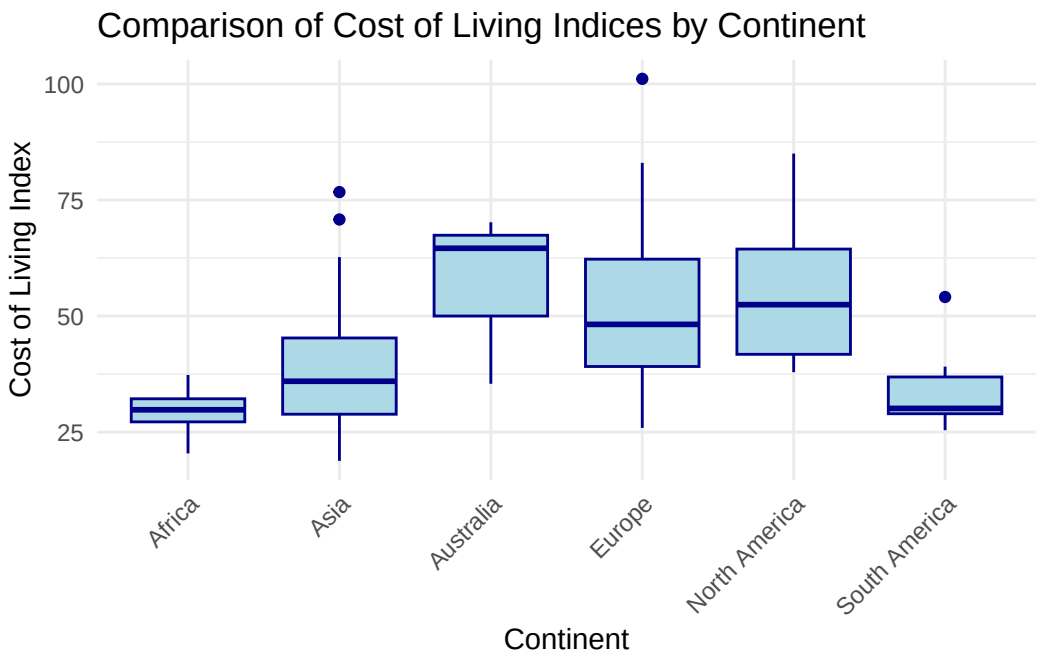
```
 [1] "Switzerland"                "Bahamas"
 [3] "Iceland"                    "Singapore"
 [5] "Barbados"                   "Norway"
 [7] "Denmark"                    "Hong Kong (China)"
 [9] "United States"              "Australia"
[11] "Austria"                    "Canada"
[13] "New Zealand"                "Ireland"
[15] "France"                     "Puerto Rico"
[17] "Finland"                    "Netherlands"
[19] "Israel"                     "Luxembourg"
[21] "Germany"                    "United Kingdom"
[23] "Belgium"                    "South Korea"
[25] "Sweden"                     "Italy"
[27] "United Arab Emirates"       "Cyprus"
[29] "Uruguay"                    "Jamaica"
[31] "Malta"                      "Trinidad And Tobago"
[33] "Costa Rica"                 "Bahrain"
[35] "Greece"                     "Estonia"
[37] "Qatar"                      "Slovenia"
[39] "Latvia"                     "Spain"
[41] "Lithuania"                  "Slovakia"
[43] "Cuba"                       "Czech Republic"
[45] "Panama"                     "Japan"
[47] "Croatia"                    "Saudi Arabia"
```

```
 [49] "Taiwan"                        "Portugal"
 [51] "Oman"                          "Kuwait"
 [53] "Albania"                       "Lebanon"
 [55] "Hungary"                       "Palestine"
 [57] "Jordan"                        "Armenia"
 [59] "Poland"                        "Mexico"
 [61] "El Salvador"                   "Montenegro"
 [63] "Chile"                         "Guatemala"
 [65] "Venezuela"                     "Bulgaria"
 [67] "Dominican Republic"            "Serbia"
 [69] "Romania"                       "Turkey"
 [71] "Cambodia"                      "Cameroon"
 [73] "Zimbabwe"                      "Mauritius"
 [75] "Fiji"                          "Bosnia And Herzegovina"
 [77] "Sri Lanka"                     "South Africa"
 [79] "Thailand"                      "Moldova"
 [81] "Georgia"                       "North Macedonia"
 [83] "Ecuador"                       "Kazakhstan"
 [85] "China"                         "Nigeria"
 [87] "Azerbaijan"                    "Philippines"
 [89] "Russia"                        "Ghana"
 [91] "Brazil"                        "Kenya"
 [93] "Botswana"                      "Malaysia"
 [95] "Peru"                          "Morocco"
 [97] "Kosovo (Disputed Territory)" "Argentina"
 [99] "Iraq"                          "Uganda"
[101] "Algeria"                       "Colombia"
[103] "Vietnam"                       "Tunisia"
[105] "Bolivia"                       "Kyrgyzstan"
[107] "Indonesia"                     "Iran"
[109] "Uzbekistan"                    "Belarus"
[111] "Ukraine"                       "Nepal"
[113] "Paraguay"                      "Madagascar"
[115] "Syria"                         "Tanzania"
[117] "Bangladesh"                    "India"
[119] "Egypt"                         "Libya"
[121] "Pakistan"
```

```
#Manually Mapping Country to Continent
continent_mapping <- list(
  "North America" = c("United States", "Canada", "Mexico", "Bahamas", "Barbados", "Puerto Ri
  "South America" = c("Brazil", "Argentina", "Chile", "Colombia", "Peru", "Uruguay", "Venezue
```

```
  "Europe" = c("Switzerland", "Iceland", "Norway", "Denmark", "Germany", "United Kingdom", "I
  "Asia" = c("Singapore", "Hong Kong (China)", "Japan", "China", "India", "South Korea", "Isr
  "Africa" = c("South Africa", "Nigeria", "Kenya", "Egypt", "Morocco", "Algeria", "Tunisia",
  "Australia" = c("Australia", "New Zealand", "Fiji")
)

# Function to get continent based on country
group.continent <- function(country) {
  for (continent in names(continent_mapping)) {
    if (country %in% continent_mapping[[continent]]) {
      return(continent)
    }
  }
  return(NA)
}

# add new column Continet
main.data <- main.data |>
  mutate(Continent = sapply(Country, group.continent))
```

```
plot2 <- ggplot(main.data, aes(x = Continent, y = Cost.of.Living.Index)) +
  geom_boxplot(fill = "lightblue", color = "darkblue") +
  labs(
    title = "Comparison of Cost of Living Indices by Continent",
    x = "Continent",
    y = "Cost of Living Index"
  ) +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))


print(plot2)
```




```
# Calculate summary statistics for each continent
summary.stats.by.continent <- main.data |>
  group_by(Continent) |>
  summarise(
    Avg.Cost.of.Living = mean(Cost.of.Living.Index, na.rm = TRUE),
    Median.Cost.of.Living = median(Cost.of.Living.Index, na.rm = TRUE),
    Min.Cost.of.Living = min(Cost.of.Living.Index, na.rm = TRUE),
    Max.Cost.of.Living = max(Cost.of.Living.Index, na.rm = TRUE),
    SD.Cost.of.Living = sd(Cost.of.Living.Index, na.rm = TRUE)
  )

print(summary.stats.by.continent)
```

```
# A tibble: 6 x 6
  Continent     Avg.Cost.of.Living Median.Cost.of.Living Min.Cost.of.Living
  <chr>                      <dbl>                 <dbl>              <dbl>
1 Africa                      29.6                  29.8               20.4
2 Asia                        39.0                  36.0               18.8
3 Australia                   56.7                  64.6               35.4
4 Europe                      51.0                  48.2               25.9
5 North America               54.9                  52.4               37.9
6 South America               33.6                  30.1               25.4
# i 2 more variables: Max.Cost.of.Living <dbl>, SD.Cost.of.Living <dbl>
```

Visualization 2: Comparison of Cost of Living Indices by Continent

Caption: This boxplot compares the cost of living indices across different continents. Each box represents the distribution of the cost of living index for countries within a continent. This visualization helps identify the variations in living expenses among different regions.

Key Insights:

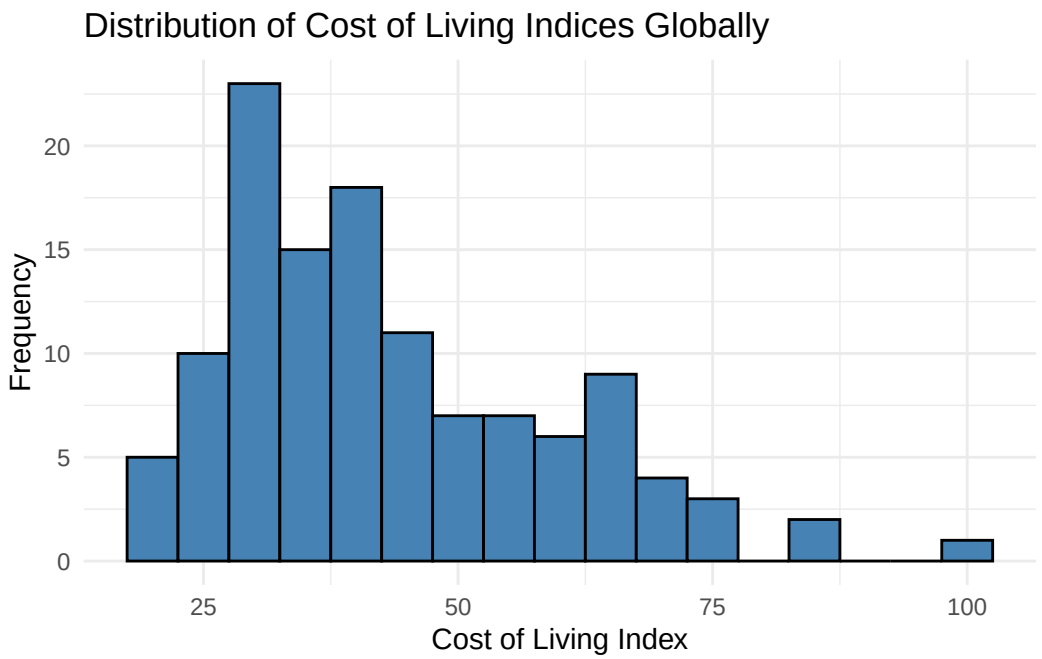
- **Average and Median Cost of Living Index:**
  - **Africa:** The average cost of living index is 29.63, and the median is 29.80, indicating a relatively consistent and low cost of living across the continent. fi
  - **Asia:** The average cost of living index is 38.96, with a median of 35.95. This suggests that while there are some high-cost countries, most Asian countries have a moderate cost of living.
  - **Australia:** The average cost of living index is 56.73, and the median is 64.60, highlighting that Australia consistently has high living costs with less variability.
  - **Europe:** The average cost of living index is 51.04, with a median of 48.20. Europe shows a higher cost of living with a wide range, indicating significant variability among countries.
  - **North America:** The average cost of living index is 54.89, and the median is 52.45, suggesting a high cost of living but with some countries having lower costs.
  - **South America:** The average cost of living index is 33.55, with a median of 30.10, showing generally lower living costs but with a few outliers.
- **Cost of Living Variability:**
  - **Europe** and **Asia** exhibit the greatest variability in the cost of living indices. This reflects the economic diversity within these continents, with countries like Switzerland (very high cost of living) and Eastern European countries (lower cost of living) creating a broad range.
  - **Australia** has a narrower range but is positioned higher on the cost index, indicating uniformly high costs of living across its regions.
  - **Africa** and **South America** show more consistency and lower costs of living, though the presence of outliers suggests some countries in these regions have significantly higher costs than others.
- **Economic Implications:**

- For businesses considering expansion, **Europe** and **Asia** offer both high-cost and low-cost options, depending on the country. This can impact decisions on where to establish operations based on budget constraints.
- **North America** and **Australia** are likely to attract expatriates who can afford higher living costs, often correlated with higher salaries and better living standards.
- **Africa** and **South America**, with their generally lower living costs, might attract businesses looking for cost-effective locations and individuals seeking affordable living conditions.

Third Visualization

Distribution of Cost of Living Indices Globally

```
plot3 <- ggplot(main.data, aes(x = Cost.of.Living.Index)) +
  geom_histogram(binwidth = 5, fill = "steelblue", color = "black") +
  labs(
    title = "Distribution of Cost of Living Indices Globally",
    x = "Cost of Living Index",
    y = "Frequency"
  ) +
  theme_minimal()

print(plot3)
```



Visualization 3: Distribution of Cost of Living Indices Globally

Caption: This histogram shows the distribution of cost of living indices across all countries in the dataset. This visualization helps identify the frequency and spread of different living costs globally.

Key Insights:

- **Central Tendency and Spread:**
  - The distribution is right-skewed, with the majority of countries having a cost of living index between 20 and 60. This indicates that most countries have relatively moderate living costs.
  - The peak of the distribution, or the mode, is in the range of 30-40, suggesting that this is the most common cost of living index for the countries in the dataset.
- **High-Cost Outliers:**
  - There are a few countries with a very high cost of living index (above 80), indicating that these countries are significantly more expensive than the global average. This highlights the economic disparity and the presence of exclusive high-cost living environments.
- **Low-Cost Outliers:**

  - Similarly, some countries have a very low cost of living index (below 20), suggesting that these countries offer a much more affordable living standard. This can be attractive for budget-conscious expatriates and businesses looking to minimize operational costs.

- **Economic Implications:**
  - For individuals looking to relocate, understanding the distribution of living costs can help in choosing a country that fits their budget. Countries in the 30-40 range might offer a balanced cost of living with decent quality of life.
    * For businesses, knowing the typical cost of living can assist in strategic decisions regarding office locations, employee salaries, and operational budgets. Lower cost regions might be preferred for back-office operations, while higher cost areas could be targeted for client-facing roles.

Fourth Visualization

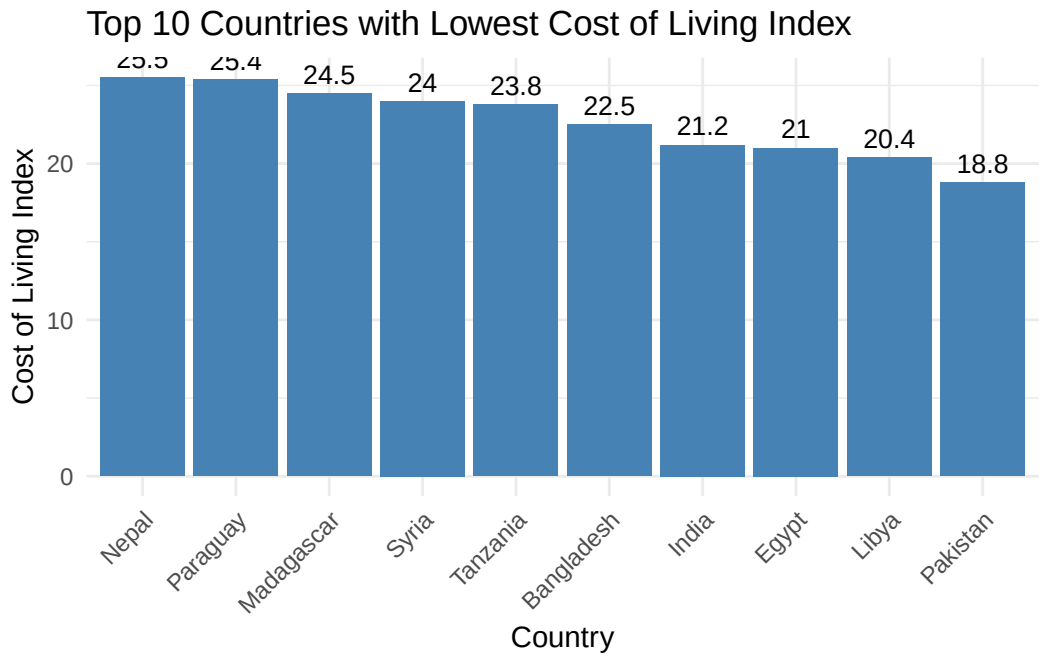
Top 10 Countries with Lowest Cost of Living Index

```
top.10.lowest.cost.living <- main.data |>
  arrange(Cost.of.Living.Index)
head(top.10.lowest.cost.living, n = 10)
```

```
      Country Cost.of.Living.Index Rent.Index Cost.of.Living.Plus.Rent.Index
1    Pakistan                 18.8        2.8                           11.1
2       Libya                 20.4        4.3                           12.7
3       Egypt                 21.0        3.7                           12.7
4       India                 21.2        5.6                           13.7
5  Bangladesh                 22.5        2.4                           12.8
6    Tanzania                 23.8        8.4                           16.4
7       Syria                 24.0        3.7                           14.2
8  Madagascar                 24.5        7.8                           16.5
9    Paraguay                 25.4        9.1                           17.6
10      Nepal                 25.5        3.5                           15.0
   Groceries.Index Restaurant.Price.Index Local.Purchasing.Power.Index
1             17.5                   12.9                          29.1
2             22.2                   15.2                          42.0
3             21.2                   16.2                          20.0
4             23.8                   15.1                          82.6
5             25.7                   12.8                          33.1
6             25.4                   16.5                          25.9
```

```
7             24.5                    15.3                          5.5
8             24.1                    14.3                         15.6
9             24.0                    17.6                         34.6
10            25.6                    16.2                         25.4
       Continent
1           Asia
2         Africa
3         Africa
4           Asia
5           Asia
6         Africa
7           Asia
8         Africa
9  South America
10          Asia
```

```r
ggplot(head(top.10.lowest.cost.living, 10), aes(x = reorder(Country, -Cost.of.Living.Inc
  geom_bar(stat = "identity", fill = "steelblue") +
  geom_text(aes(label = Cost.of.Living.Index), vjust = -0.5, color = "black", size = 3.5
  labs(
    title = "Top 10 Countries with Lowest Cost of Living Index",
    x = "Country",
    y = "Cost of Living Index"
  ) +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



Visualization 4: Top 10 Countries with Lowest Cost of Living Index

Caption: This bar plot displays the top 10 countries with the lowest cost of living index. The countries are ranked in ascending order based on their cost of living index. This visualization highlights the most affordable places to live, which can be particularly useful for budget-conscious expatriates and businesses looking to minimize operational costs.

Key Insights:

- **Affordable Living:**
  - The countries with the lowest cost of living index provide affordable living standards, making them attractive for individuals and families looking to stretch their budgets further.
  - These countries often offer lower housing, food, and transportation costs, contributing to their overall low cost of living index.
- **Economic Opportunities:**
  - For businesses, these countries present opportunities for cost effective operations, especially for back office functions, manufacturing, and other cost sensitive activities.

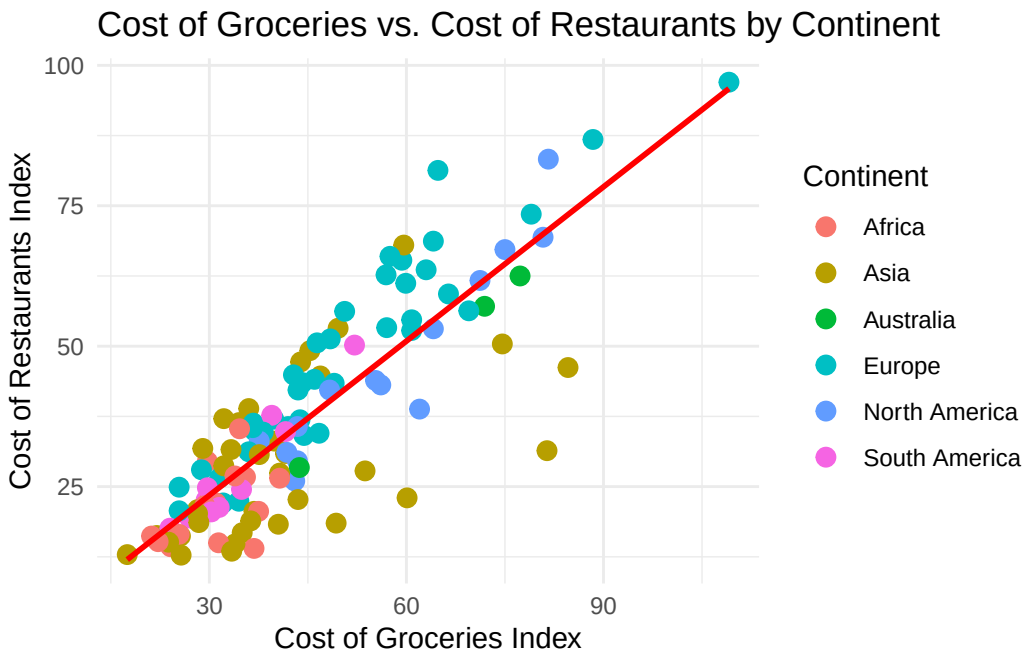
– Low living costs can also translate to competitive salaries that are attractive to local talent while being lower than those in high cost regions.

Fifth Visualization

Cost of Groceries vs. Cost of Restaurants by Continent

```
plot5 <- ggplot(main.data, aes(x = Groceries.Index, y = Restaurant.Price.Index, color = Conti
  geom_point(size = 3) +
  geom_smooth(method = "lm", se = FALSE, color = "red") +
  labs(
    title = "Cost of Groceries vs. Cost of Restaurants by Continent",
    x = "Cost of Groceries Index",
    y = "Cost of Restaurants Index"
  ) +
  theme_minimal()

print(plot5)
```

```
`geom_smooth()` using formula = 'y ~ x'
```



Visualization 5: Cost of Groceries vs. Cost of Restaurants by Continent

Caption: This scatter plot illustrates the relationship between the cost of groceries and the cost of dining at restaurants across various countries, grouped by continent. The red line represents the linear regression trend line, highlighting the general relationship between the two variables. By comparing these two metrics, we can gain insights into the relative affordability of home cooked meals versus dining out in different regions.

Key Insights:

- **Positive Correlation:**
  - The scatter plot shows a strong positive correlation between the cost of groceries and the cost of restaurants. As the cost of groceries increases, the cost of dining out also tends to increase. This suggests that in countries where groceries are expensive, dining out is also likely to be costly.
- **Continental Clusters:**
  - **Europe** (represented by cyan points) tends to have higher costs for both groceries and restaurants, indicating a generally high cost of living in European countries.
  - **North America** (blue points) also shows higher costs but with more variation compared to Europe.
  - **Asia** (gold points) exhibits a wider spread, with some countries having lower costs and others higher, reflecting the economic diversity within the continent.
  - **Africa** (red points) and **South America** (pink points) generally show lower costs for both groceries and dining out, highlighting more affordable living costs in these regions.
- **Affordability Analysis:**
  - Countries where the cost of groceries is relatively low but the cost of restaurants is high might indicate cultural practices where dining out is considered a luxury or there are fewer affordable dining options.
  - Conversely, countries where both costs are low suggest a generally affordable lifestyle both for cooking at home and dining out, which can be attractive for budget-conscious expatriates.
- **Economic Stratification:**
  - The plot reveals clusters of countries by continent, showing how living costs can vary significantly within and across continents. Developed regions like Europe and North America tend to cluster towards the higher end of the cost spectrum.

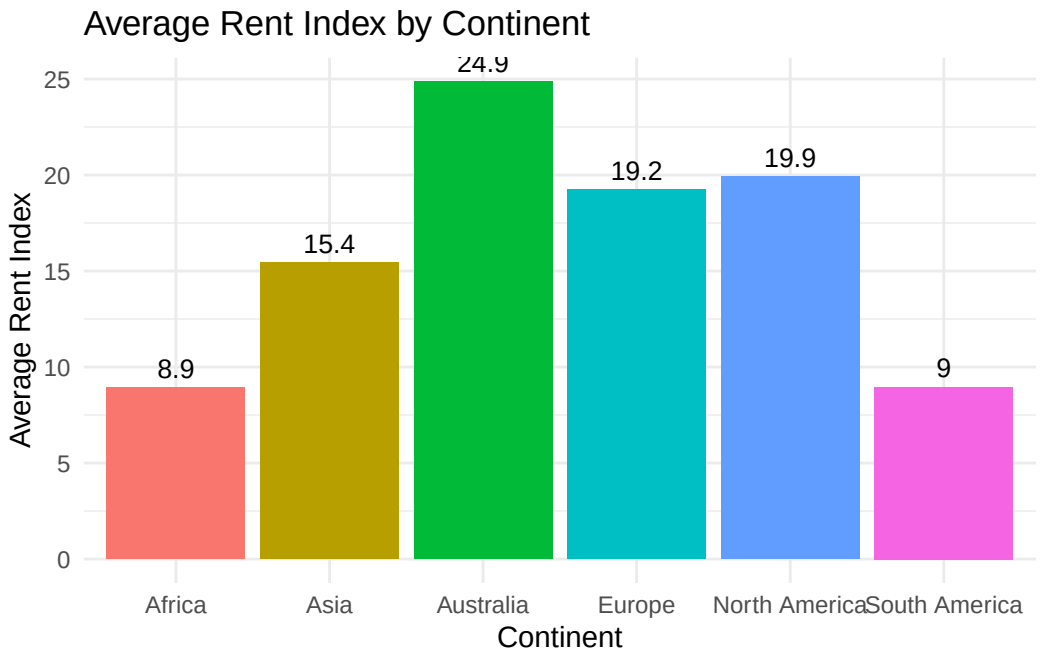
- **Outliers:**
  - There are a few notable outliers where the cost of restaurants is significantly higher than the cost of groceries. These outliers may reflect unique economic conditions or cultural practices that make dining out more expensive. An example would be Cuba.

Sixth Visualization

Average Rent Index by Continent

```
# Calculate average rent index by continent
average.rent.by.continent <- main.data |>
  group_by(Continent) |>
  summarise(Avg.Rent.Index = mean(Rent.Index, na.rm = TRUE))
```

```
plot6 <- ggplot(average.rent.by.continent, aes(x = Continent, y = Avg.Rent.Index, fill = Cont
  geom_bar(stat = "identity") +
  geom_text(aes(label = round(Avg.Rent.Index, 1)), vjust = -0.5, color = "black", size = 3.5)
  labs(
    title = "Average Rent Index by Continent",
    x = "Continent",
    y = "Average Rent Index"
  ) +
  theme_minimal() +
  theme(legend.position = "none")
print(plot6)
```



Visualization 6: Average Rent Index by Continent

Caption: This bar plot illustrates the average rent index for each continent. By grouping countries by their continents and calculating the average rent index, this visualization provides a clear comparison of rental expenses across different regions of the world.

Key Insights:

- **High Rent Continents:**
  - **Europe** and **North America** show high average rent indices, indicating that these regions have high rental costs, consistent with their overall high cost of living.
  - **Australia** also has a high average rent index, reflecting its high quality of life and developed infrastructure.
- **Moderate Rent Continents:**
  - **Asia** displays a wide range of rental costs, with some countries having very high rental expenses (e.g., Hong Kong, Japan) and others being more affordable (e.g., India, Vietnam).
- **Low Rent Continents:**

- **Africa** and **South America** have lower average rent indices, suggesting more affordable rental housing in these regions. This can attract budget-conscious expatriates and businesses looking for cost-effective operations.

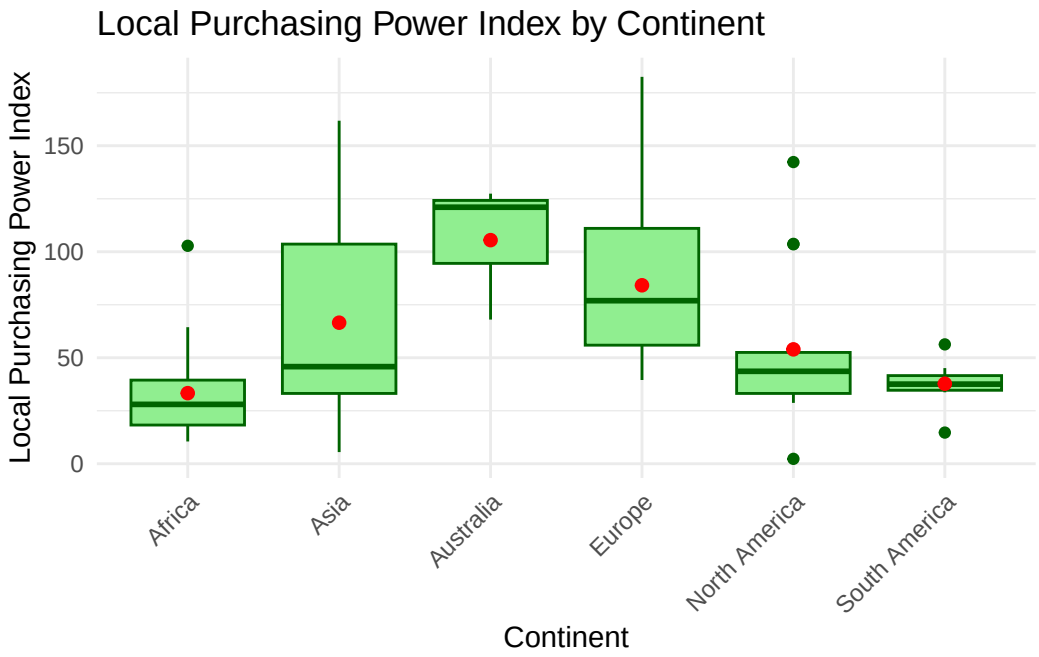
Economic Insights:

- **Economic Disparities:** The plot reveals significant economic disparities between continents, highlighting the higher rental costs in developed regions.
- **Affordability:** The lower average rent indices in Africa and South America may contribute to a more affordable lifestyle, which could be attractive for individuals and businesses seeking lower living costs.

Seventh Visualization

Local Purchasing Power Index by Continent

```
plot7 <- ggplot(main.data, aes(x = Continent, y = Local.Purchasing.Power.Index)) +
  geom_boxplot(fill = "lightgreen", color = "darkgreen") +
  stat_summary(fun = mean, geom = "point", shape = 20, size = 3, color = "red", fill = "
  labs(
    title = "Local Purchasing Power Index by Continent",
    x = "Continent",
    y = "Local Purchasing Power Index"
  ) +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

print(plot7)
```



Visualization 7: Local Purchasing Power Index by Continent

Caption: This boxplot compares the local purchasing power index across different continents. Each box represents the distribution of the local purchasing power index for countries within a continent, with the red points indicating the mean purchasing power index. This visualization helps identify variations in purchasing power, reflecting the relative economic strength and standard of living across different regions.

Key Insights:

- **High Purchasing Power:**
  - **North America** and **Europe** exhibit high local purchasing power indices, indicating that residents in these regions generally enjoy strong purchasing power relative to the cost of living. This suggests that wages and salaries in these areas are sufficient to afford a comfortable lifestyle despite high living costs.
- **Moderate to Low Purchasing Power:**
  - **Asia**, **South America**, and **Africa** show more variability and generally lower purchasing power indices. This reflects the economic disparities within these continents, where some countries have strong economies and high purchasing power, while others face economic challenges and lower purchasing power.
- **Variability Within Continents:**

- **Asia** has a wide range of purchasing power indices, reflecting the diverse economic conditions across the continent. Developed countries like Japan and Singapore have high purchasing power, while developing countries like India and Vietnam have lower purchasing power.
- **Africa** and **South America** generally have lower purchasing power indices, but there are outliers where certain countries may have higher indices due to specific economic factors.

Economic Insights:

- **Economic Disparities:** The boxplot highlights significant economic disparities both within and between continents, with developed regions showing higher purchasing power.
- **Standard of Living:** Higher local purchasing power indices suggest a higher standard of living, as residents can afford more goods and services with their income.

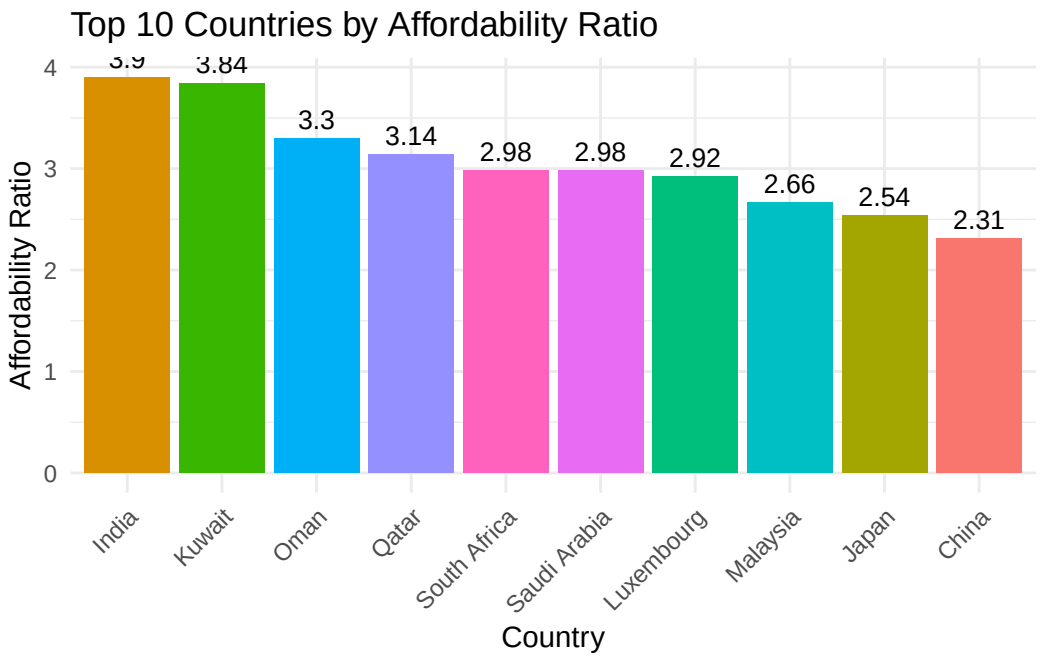
Eighth Visualization

Affordability Ratio for Top and Bottom countries

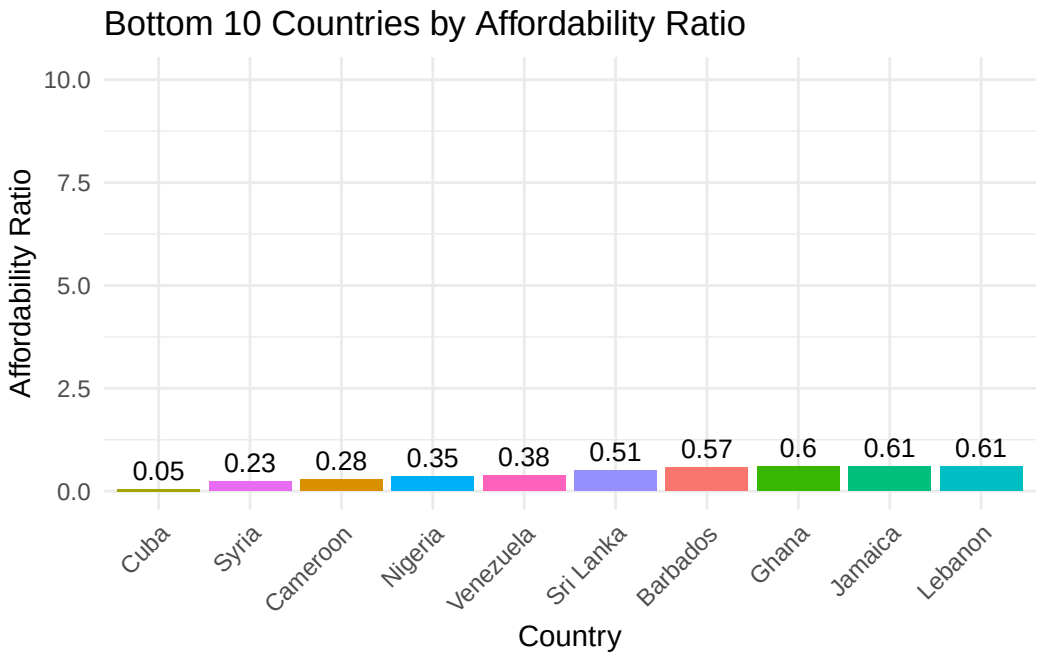
```
# Calculate Affordability Ratio
main.data <- main.data |>
  mutate(Affordability.Ratio = Local.Purchasing.Power.Index / Cost.of.Living.Index)

# Find the top 10 and bottom 10 countries by Affordability Ratio
top.10.affordable <- main.data |>
  arrange(desc(Affordability.Ratio)) |>
  head(10)

bottom.10.affordable <- main.data |>
  arrange(Affordability.Ratio) |>
  head(10)


# Combine top 10 and bottom 10 for plotting
top.bottom.20 <- rbind(top.10.affordable, bottom.10.affordable)
```

```
# Create the bar plot for the top 10 countries
plot.top.10 <- ggplot(top.10.affordable, aes(x = reorder(Country, -Affordability.Ratio), y =
  geom_bar(stat = "identity") +
  geom_text(aes(label = round(Affordability.Ratio, 2)), vjust = -0.5, color = "black", size =
  labs(
```

```
    title = "Top 10 Countries by Affordability Ratio",
    x = "Country",
    y = "Affordability Ratio"
  ) +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1), legend.position = "none") +
  coord_cartesian(ylim = c(0, max(top.10.affordable$Affordability.Ratio)))

# Create the bar plot for the bottom 10 countries
plot.bottom.10 <- ggplot(bottom.10.affordable, aes(x = reorder(Country, Affordability.Ratio)
  geom_bar(stat = "identity") +
  geom_text(aes(label = round(Affordability.Ratio, 2)), vjust = -0.5, color = "black", size =
  labs(
    title = "Bottom 10 Countries by Affordability Ratio",
    x = "Country",
    y = "Affordability Ratio"
  ) +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1), legend.position = "none") +
  coord_cartesian(ylim = c(min(bottom.10.affordable$Affordability.Ratio), min(bottom.10.affo

plot.top.10
```



```
plot.bottom.10
```



Visualization 8: Top 10 and Bottom 10 Countries by Affordability Ratio

Caption: This set of bar plots illustrates the top 10 and bottom 10 countries by Affordability Ratio. The Affordability Ratio is calculated as the Local Purchasing Power Index divided by the Cost of Living Index. A higher ratio indicates that local income is relatively high compared to living costs, making the country more affordable.

Key Insights:

- **Top 10 Countries by Affordability Ratio:**
  - Countries with the highest affordability ratios offer a favorable balance between local income and living costs. This means that residents in these countries can afford a higher quality of life relative to their earnings.
  - Notable countries with high affordability include [insert some country names from the top 10], where local purchasing power is significantly high compared to the cost of living.
- **Bottom 10 Countries by Affordability Ratio:**

- Countries with the lowest affordability ratios face higher living costs relative to local incomes. This suggests potential challenges for residents in affording basic necessities and maintaining a good quality of life.
- Notable countries with low affordability include [insert some country names from the bottom 10], where living costs are high compared to the local purchasing power.

Economic Insights:

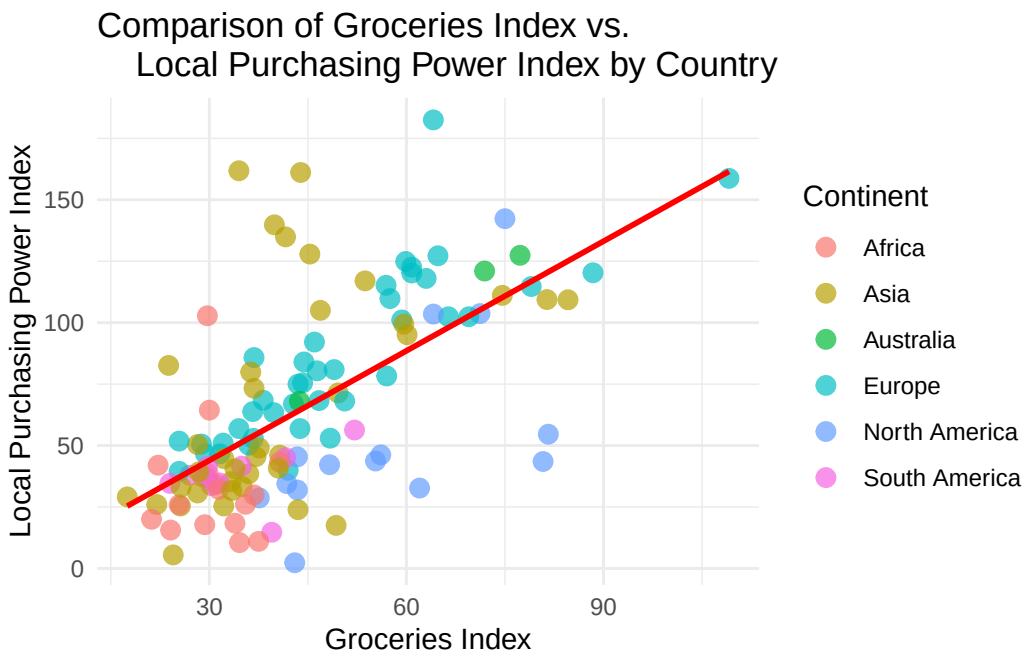
- **Economic Balance:** The plots highlight the disparity in economic balance across countries. High affordability ratios indicate regions where residents' earnings go further, while low ratios point to areas with economic strain.
- **Cost of Living vs. Income:** These insights can help identify countries where cost of living adjustments might be necessary to improve residents' quality of life.

Ninith Visualization

Comparison of Groceries Index vs. Local Purchasing Power Index by Country

```
# i needed to put the title in two lines, sorry if looks ugly!
plot.groceries.vs.purchasing.power <- ggplot(main.data, aes(x = Groceries.Index, y = Local.Pu
  geom_point(size = 3, alpha = 0.7) +
  geom_smooth(method = "lm", se = FALSE, color = "red") +
  labs(
    title = "Comparison of Groceries Index vs.
    Local Purchasing Power Index by Country",
    x = "Groceries Index",
    y = "Local Purchasing Power Index"
  ) +
  theme_minimal() +
  theme(legend.position = "right")

print(plot.groceries.vs.purchasing.power)
```

```
`geom_smooth()` using formula = 'y ~ x'
```



Visualization 9: Comparison of Groceries Index vs. Local Purchasing Power Index by Country

Caption: This scatter plot compares the Groceries Index with the Local Purchasing Power Index for countries across different continents. The Groceries Index measures the average cost of food and other grocery items, while the Local Purchasing Power Index indicates the relative purchasing power of locals based on their average salary and cost of living.

Key Insights:

- **Positive Relationship:**
  - The scatter plot shows a general positive relationship between the Groceries Index and the Local Purchasing Power Index. This suggests that in countries where groceries are more expensive, the local purchasing power tends to be higher as well.
- **Continental Clusters:**
  - Different continents exhibit distinct patterns. For example, European countries tend to cluster towards higher Groceries and Local Purchasing Power Indices, reflecting their higher cost of living and stronger economies.
  - African and South American countries generally show lower indices, indicating more affordable groceries but lower purchasing power.

- **Outliers:**
  - Some countries deviate from the general trend, indicating unique economic conditions. For instance, a country with a high Groceries Index but relatively low Local Purchasing Power Index might face affordability challenges for basic necessities.

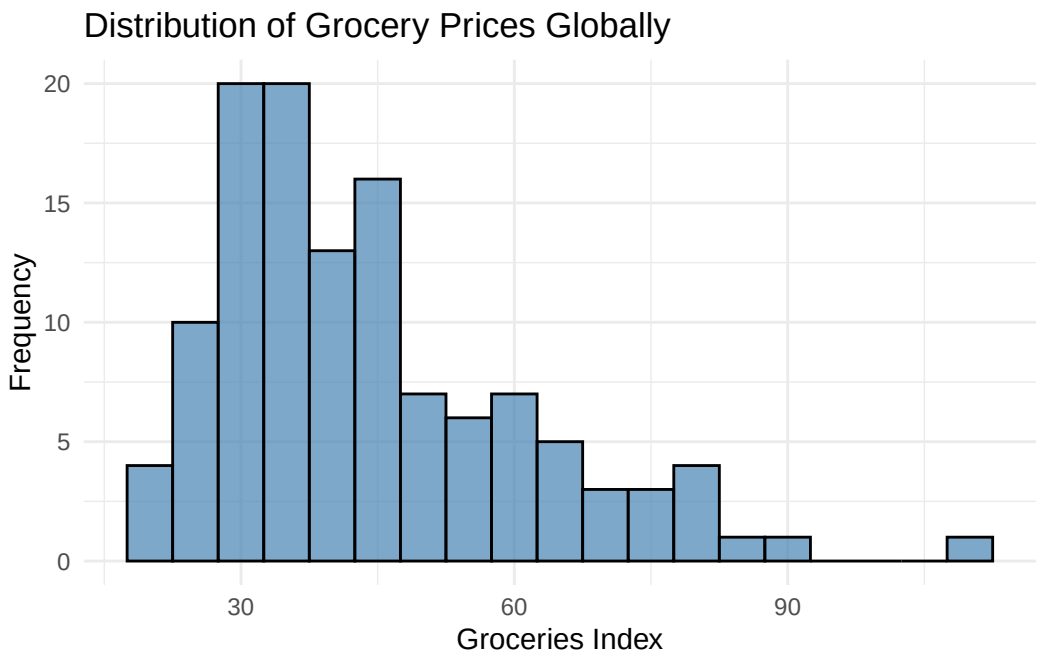
Economic Insights:

- **Economic Strength:** The positive relationship highlights how stronger economies can sustain higher living costs while maintaining higher purchasing power for residents.
- **Affordability Analysis:** Identifying outliers can help pinpoint countries where grocery prices may be disproportionately high relative to local incomes, signaling potential economic stress.

Tenth Visualization

Distribution of Grocery Prices Globally (All Countries)

```
plot.groceries.distribution <- ggplot(main.data, aes(x = Groceries.Index)) +
  geom_histogram(binwidth = 5, fill = "steelblue", color = "black", alpha = 0.7) +
  labs(
    title = "Distribution of Grocery Prices Globally",
    x = "Groceries Index",
    y = "Frequency"
  ) +
  theme_minimal()

# Display the plot
print(plot.groceries.distribution)
```



Visualization 10: Distribution of Grocery Prices Globally (All Countries)

Caption: This histogram illustrates the distribution of the Groceries Index for all countries in the dataset, with data labels showing the frequency of countries within each bin. The Groceries Index measures the average cost of food and other grocery items, providing a comprehensive view of grocery prices across the globe.

Key Insights:

- **Central Tendency:**
  - The histogram shows that the majority of countries have a Groceries Index between 30 and 60, indicating that most countries have moderate grocery prices.
  - The peak (mode) of the distribution is around the 40-50 range, suggesting that this is the most common grocery price range globally.
- **High-Cost Outliers:**
  - There are a few countries with a Groceries Index above 80, indicating significantly higher grocery prices. These outliers may represent countries with higher costs of living or unique economic conditions.
- **Low-Cost Outliers:**

  - Similarly, there are countries with a Groceries Index below 20, suggesting more affordable grocery prices. These countries might have lower overall living costs or subsidies that reduce food prices.

Economic Insights:

- **Global Affordability:** Understanding the distribution of grocery prices globally helps identify regions where basic necessities are more affordable or expensive, impacting the overall cost of living.
- **Economic Disparities:** The presence of high and low-cost outliers highlights economic disparities between countries, with some regions facing higher grocery costs that can strain household budgets.

Summary of Lessons

Through this exploratory data analysis, several key insights were gained regarding the global cost of living and purchasing power indices, directly addressing the guiding research questions outlined in the beginning.

1. What are the most and least expensive countries to live in based on the Cost of Living Index and Rent Index?

The analysis revealed significant variations in the cost of living across different countries:

- **Most Expensive Countries**: Switzerland and the Bahamas stood out for their high Cost of Living Index, reflecting expensive lifestyles and high standards of living. Similarly, Europe and North America had higher average cost of living and rent indices, highlighting the economic strength and higher expenses in these regions.
- **Least Expensive Countries**: Conversely, countries like Pakistan and Libya exhibited the lowest cost of living, indicating more affordable living standards. Africa and South America showed lower cost of living indices, suggesting more affordable living conditions.

These insights help identify the countries where living expenses are the highest and lowest, which can be crucial for expatriates, businesses, and policymakers.

2. How does the Local Purchasing Power Index vary across countries, and what factors seem to influence it the most?

The affordability ratio analysis showed:

- **High Affordability**: Countries like India and Kuwait have the highest affordability, where local purchasing power significantly surpasses the cost of living. North America and Europe exhibited high local purchasing power, allowing residents to maintain a higher quality of life despite high living costs.
- **Low Affordability**: Countries like Syria and Cameroon face economic strain with lower affordability ratios. Developing regions, such as parts of Asia and Africa, showed greater variability and generally lower purchasing power.

Factors influencing the Local Purchasing Power Index include the average salary, cost of basic necessities, and overall economic conditions within each country.

3. What is the relationship between the Cost of Living Plus Rent Index and the Groceries Index and Restaurant Price Index across different regions?

The relationship between the cost of groceries and dining out was examined, revealing a strong positive correlation:

- **Positive Correlation**: Particularly in Europe and North America, high grocery costs often coincide with high restaurant prices. This indicates that in regions where the cost of groceries is high, dining out is also likely to be expensive.
- **Global Distribution**: The global distribution of grocery prices showed that most countries fall within a moderate range, with some outliers indicating significantly higher or lower prices.

These insights can assist policymakers in understanding economic disparities and formulating strategies for cost of living adjustments. Businesses can use this data to make informed decisions on global expansion, considering the cost-benefit analysis of different regions. Individuals looking to relocate can identify countries with favorable living costs and purchasing power balances, aiding in making informed decisions based on economic conditions.

Overall, this analysis highlights the importance of understanding global economic trends and their impact on living standards, offering valuable insights for individuals, businesses, and policymakers alike.