```r
# Read the CSV file into a data frame in R
data <- read.csv("/kaggle/input/global-cost-of-living-rankingsaffordability-index/wikipedia_cost_of_living_indices3.csv")

# Display the first few rows of the data frame
head(data)
```

A data.frame: 6 × 7

	Country	Cost.of.Living.Index	Rent.Index	Cost.of.Living.Plus.Rent.Index	Groceries.Index	Restaurant.Price.Index	Local.Purchasing.Power.Index
	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	Switzerland	101.1	46.5	74.9	109.1	97.0	158.7
2	Bahamas	85.0	36.7	61.8	81.6	83.3	54.6
3	Iceland	83.0	39.2	62.0	88.4	86.8	120.3
4	Singapore	76.7	67.2	72.1	74.6	50.4	111.1
5	Barbados	76.6	19.0	48.9	80.8	69.4	43.5
6	Norway	76.0	26.2	52.1	79.0	73.5	114.7

+ Code + Markdown

[2]:
```r
str(data)
```

```
'data.frame':   121 obs. of  7 variables:
 $ Country                       : chr  "Switzerland" "Bahamas" "Iceland" "Singapore" ...
 $ Cost.of.Living.Index          : num  101.1 85 83 76.7 76.6 ...
 $ Rent.Index                    : num  46.5 36.7 39.2 67.2 19 26.2 26.4 59.4 41.7 33.4 ...
 $ Cost.of.Living.Plus.Rent.Index: num  74.9 61.8 62 72.1 48.9 52.1 50.2 65.3 56.6 52.5 ...
 $ Groceries.Index               : num  109.1 81.6 88.4 74.6 80.8 ...
 $ Restaurant.Price.Index        : num  97 83.3 86.8 50.4 69.4 73.5 81.3 46.2 67.2 62.5 ...
 $ Local.Purchasing.Power.Index  : num  158.7 54.6 120.3 111.1 43.5 ...
```

[3]:
```r
# Display the unique countries
unique_countries <- unique(data$Country)
cat("Unique Countries:\n")
print(unique_countries)
```

```
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"
```

[17]:
```r
# Calculate the average values grouped by Country
avg_values <- aggregate(. ~ Country, data = data, FUN = mean)
```

```
# Display the average values
cat("\nAverage Values:\n")
print(head(avg_values))
```

```
Average Values:
           Country Cost_of_Living_Index Rent_Index
1         Australia                 70.2       33.4
2           Bahamas                 85.0       36.7
3          Barbados                 76.6       19.0
4           Denmark                 72.3       26.4
5 Hong Kong (China)                 70.8       59.4
6           Iceland                 83.0       39.2
  Cost_of_Living_Plus_Rent_Index Groceries_Index Restaurant_Price_Index
1                           52.5            77.3                   62.5
2                           61.8            81.6                   83.3
3                           48.9            80.8                   69.4
4                           50.2            64.8                   81.3
5                           65.3            84.6                   46.2
6                           62.0            88.4                   86.8
  Local_Purchasing_Power_Index Affordability_Index
1                        127.4           0.5644661
2                         54.6           0.2072893
3                         43.5           0.1994498
4                        127.2           0.5711720
5                        109.3           0.4277886
6                        120.3           0.4352388
```

[18]:
```
print(tail(avg_values))
```

```
            Country Cost_of_Living_Index Rent_Index
5  Hong Kong (China)                 70.8       59.4
6           Iceland                  83.0       39.2
7            Norway                  76.0       26.2
8         Singapore                  76.7       67.2
9       Switzerland                 101.1       46.5
10    United States                  70.4       41.7
   Cost_of_Living_Plus_Rent_Index Groceries_Index Restaurant_Price_Index
5                            65.3            84.6                   46.2
6                            62.0            88.4                   86.8
7                            52.1            79.0                   73.5
8                            72.1            74.6                   50.4
9                            74.9           109.1                   97.0
10                           56.6            75.0                   67.2
   Local_Purchasing_Power_Index Affordability_Index
5                         109.3           0.4277886
6                         120.3           0.4352388
7                         114.7           0.4969671
8                         111.1           0.4203557
9                         158.7           0.4845802
10                        142.3           0.5916840
```

[5]:
```
# Calculate the affordability Index
```

[6]:
```
# Convert data to a data frame in R
df <- as.data.frame(data)
```

[7]:
```
head(df)
```

A data.frame: 6 × 7

	Country	Cost.of.Living.Index	Rent.Index	Cost.of.Living.Plus.Rent.Index	Groceries.Index	Restaurant.Price.Index	Local.Purchasing.Power.Index
	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	Switzerland	101.1	46.5	74.9	109.1	97.0	158.7
2	Bahamas	85.0	36.7	61.8	81.6	83.3	54.6
3	Iceland	83.0	39.2	62.0	88.4	86.8	120.3
4	Singapore	76.7	67.2	72.1	74.6	50.4	111.1
5	Barbados	76.6	19.0	48.9	80.8	69.4	43.5
6	Norway	76.0	26.2	52.1	79.0	73.5	114.7

[8]:
```
str(df)
```

```
'data.frame':   121 obs. of  7 variables:
 $ Country                       : chr  "Switzerland" "Bahamas" "Iceland" "Singapore" ...
 $ Cost.of.Living.Index          : num  101.1 85 83 76.7 76.6 ...
 $ Rent.Index                    : num  46.5 36.7 39.2 67.2 19 26.2 26.4 59.4 41.7 33.4 ...
 $ Cost.of.Living.Plus.Rent.Index: num  74.9 61.8 62 72.1 48.9 52.1 50.2 65.3 56.6 52.5 ...
 $ Groceries.Index               : num  109.1 81.6 88.4 74.6 80.8 ...
 $ Restaurant.Price.Index        : num  97 83.3 86.8 50.4 69.4 73.5 81.3 46.2 67.2 62.5 ...
 $ Local.Purchasing.Power.Index  : num  158.7 54.6 120.3 111.1 43.5 ...
```

[16]:
```
# Add the Affordability Index column to the data frame
df$Affordability_Index <- df$Local.Purchasing.Power.Index / (
  df$Cost.of.Living.Plus.Rent.Index +
  df$Rent.Index +
  df$Groceries.Index +
  df$Restaurant.Price.Index
)

# View the updated data frame
print(head(df))
```

```
     Country Cost.of.Living.Index Rent.Index Cost.of.Living.Plus.Rent.Index
1 Switzerland                101.1       46.5                           74.9
2    Bahamas                 85.0       36.7                           61.8
```

```
3     Iceland                83.0       39.2                         62.0
4   Singapore                76.7       67.2                         72.1
5    Barbados                76.6       19.0                         48.9
6      Norway                76.0       26.2                         52.1
  Groceries.Index Restaurant.Price.Index Local.Purchasing.Power.Index
1           109.1                   97.0                         158.7
2            81.6                   83.3                          54.6
3            88.4                   86.8                         120.3
4            74.6                   50.4                         111.1
5            80.8                   69.4                          43.5
6            79.0                   73.5                         114.7
  Affordability_Index
1           0.4845802
2           0.2072893
3           0.4352388
4           0.4203557
5           0.1994498
6           0.4969671
```

[10]:

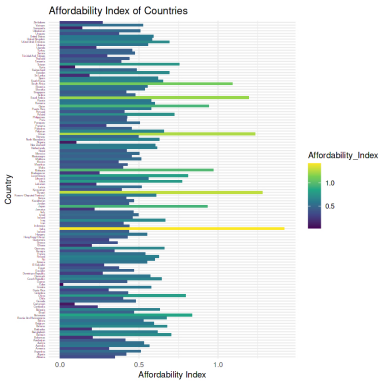
```
# Sort the data frame by Affordability Index in descending order
sorted_df <- df[order(-df$Affordability_Index), ]

# Display the sorted data frame with selected columns
cat("Sorted by Affordability Index:\n")
print(sorted_df[, c("Country", "Affordability_Index")])
```

```
Sorted by Affordability Index:
                      Country Affordability_Index
118                     India          1.41924399
52                     Kuwait          1.28209192
51                       Oman          1.23607427
48                Saudi Arabia          1.19592199
78                South Africa          1.09129512
94                    Malaysia          0.97201946
37                       Qatar          0.94100467
46                       Japan          0.93375898
93                    Botswana          0.83527886
20                  Luxembourg          0.81039076
85                       China          0.79587405
120                      Libya          0.77205882
49                      Taiwan          0.75476190
59                      Poland          0.72466216
34                     Bahrain          0.70422535
27        United Arab Emirates          0.69247428
25                      Sweden          0.69148336
110                     Belarus          0.67889908
76      Bosnia And Herzegovina          0.67497034
99                        Iraq          0.66666667
21                     Germany          0.66043956
121                   Pakistan          0.65688488
24                South Korea          0.65235540
44              Czech Republic          0.64318530
66                    Bulgaria          0.63274746
82             North Macedonia          0.62962963
17                    Finland          0.62533121
40                      Spain          0.61978466
117                 Bangladesh          0.61638734
18                 Netherlands          0.61375921
97  Kosovo (Disputed Territory)          0.60886571
13                 New Zealand          0.60228970
69                     Romania          0.60075330
75                        Fiji          0.60070671
23                     Belgium          0.59351351
9                United States          0.59168399
22              United Kingdom          0.58240647
16                 Puerto Rico          0.57821229
47                     Croatia          0.57781202
89                      Russia          0.57696693
7                       Denmark          0.57117198
10                    Australia          0.56446611
41                    Lithuania          0.55843195
111                     Ukraine          0.55790960
38                     Slovenia          0.55410959
55                      Hungary          0.55034722
36                      Estonia          0.54993160
15                       France          0.53894737
42                     Slovakia          0.53610675
11                      Austria          0.53084500
105                     Bolivia          0.52712100
103                     Vietnam          0.52665800
39                       Latvia          0.51540832
80                      Moldova          0.51321586
98                    Argentina          0.51234568
109                  Uzbekistan          0.50840880
113                    Paraguay          0.50658858
96                      Morocco          0.50260417
6                        Norway          0.49696707
19                       Israel          0.49625562
1                   Switzerland          0.48458015
12                       Canada          0.48076032
68                       Serbia          0.47718631
104                     Tunisia          0.47352941
91                       Brazil          0.46792453
84                    Kazakhstan          0.46757322
83                       Ecuador          0.46689113
26                         Italy          0.46607143
14                       Ireland          0.45808790
70                        Turkey          0.45666356
56                     Palestine          0.45374016
62                    Montenegro          0.45252352
107                    Indonesia          0.43905817
79                      Thailand          0.43823845
57                        Jordan          0.43750000
3                         Iceland          0.43523878
102                      Colombia          0.42929936
8             Hong Kong (China)          0.42778865
74                     Mauritius          0.42645607
28                        Cyprus          0.42500000
88                    Philippines          0.42420382
95                          Peru          0.42130751
112                        Nepal          0.42122720
4                      Singapore          0.42035566
92                         Kenya          0.42014742
101                      Algeria          0.41818182
87                    Azerbaijan          0.41379310
50                      Portugal          0.40730136
```

108	Iran	0.40247678
31	Malta	0.40035273
63	Chile	0.39840989
106	Kyrgyzstan	0.39058524
116	Tanzania	0.38830585
29	Uruguay	0.37309476
119	Egypt	0.37174721
60	Mexico	0.36910569
35	Greece	0.36351166
53	Albania	0.34605377
81	Georgia	0.34439024
58	Armenia	0.30948553
64	Guatemala	0.30858676
33	Costa Rica	0.30575778
32	Trinidad And Tobago	0.29931507
45	Panama	0.29224377
61	El Salvador	0.27690972
73	Zimbabwe	0.27102804
67	Dominican Republic	0.27049953
114	Madagascar	0.24880383
71	Cambodia	0.23947896
100	Uganda	0.22997416
54	Lebanon	0.22924188
30	Jamaica	0.21843687
2	Bahamas	0.20728929
90	Ghana	0.20000000
5	Barbados	0.19944979
77	Sri Lanka	0.18597237
65	Venezuela	0.13907285
86	Nigeria	0.10261194
115	Syria	0.09532062
72	Cameroon	0.09186352
43	Cuba	0.02077687

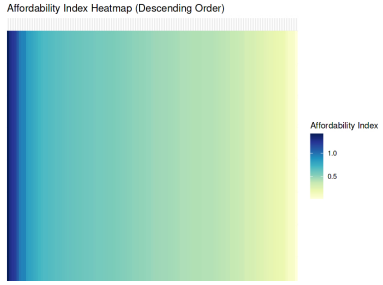
[11]:

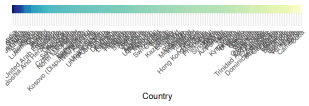
```r
# Load necessary library
library(ggplot2)

# Create a bar plot using ggplot2
ggplot(sorted_df, aes(x = Affordability_Index, y = Country, fill = Affordability_Index)) +
  geom_bar(stat = "identity") +
  scale_fill_viridis_c() +  # Use a color palette (viridis)
  labs(
    x = "Affordability Index",
    y = "Country",
    title = "Affordability Index of Countries"
  ) +
  theme_minimal() +
  theme(axis.text.y = element_text(size = 3))  # Adjust text size if needed
```



[12]:

```r
# Reorder the data frame by Affordability Index in descending order
heatmap_data <- sorted_df[order(-sorted_df$Affordability_Index), ]

# Create a heatmap using ggplot2
ggplot(heatmap_data, aes(x = reorder(Country, -Affordability_Index), y = 1, fill = Affordability_Index)) +
  geom_tile() +
  scale_fill_gradientn(colors = scales::brewer_pal(palette = "YlGnBu")(9)) +
  labs(
    x = "Country",
    y = NULL,
    fill = "Affordability Index",
    title = "Affordability Index Heatmap (Descending Order)"
  ) +
  theme_minimal() +
  theme(
    axis.text.y = element_blank(),       # Remove y-axis labels
    axis.ticks.y = element_blank(),      # Remove y-axis ticks
    axis.text.x = element_text(angle = 45, hjust = 1)  # Rotate x-axis labels
  )
```

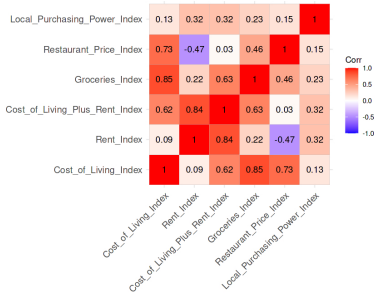




[13]:
```r
# Create the data
data <- data.frame(
  Country = c("Switzerland", "Bahamas", "Iceland", "Singapore", "Barbados",
              "Norway", "Denmark", "Hong Kong (China)", "United States", "Australia"),
  Cost_of_Living_Index = c(101.1, 85.0, 83.0, 76.7, 76.6, 76.0, 72.3, 70.8, 70.4, 70.2),
  Rent_Index = c(46.5, 36.7, 39.2, 67.2, 19.0, 26.2, 26.4, 59.4, 41.7, 33.4),
  Cost_of_Living_Plus_Rent_Index = c(74.9, 61.8, 62.0, 72.1, 48.9, 52.1, 50.2, 65.3, 56.6, 52.5),
  Groceries_Index = c(109.1, 81.6, 88.4, 74.6, 80.8, 79.0, 64.8, 84.6, 75.0, 77.3),
  Restaurant_Price_Index = c(97.0, 83.3, 86.8, 50.4, 69.4, 73.5, 81.3, 46.2, 67.2, 62.5),
  Local_Purchasing_Power_Index = c(158.7, 54.6, 120.3, 111.1, 43.5, 114.7, 127.2, 109.3, 142.3, 127.4)
)

# View the data frame
print(data)
```

```
            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
```

[14]:
```r
# Correlation matrix
correlation_matrix <- cor(data[,-1])  # Exclude Country column

# Visualize correlation
library(ggcorrplot)
ggcorrplot(correlation_matrix, lab = TRUE)
```



[15]:
```r
# Create Affordability Index as a derived variable
data$Affordability_Index <- data$Local_Purchasing_Power_Index /
  (data$Cost_of_Living_Plus_Rent_Index + data$Rent_Index + data$Groceries_Index + data$Restaurant_Price_Index)

# Fit a linear regression model
model <- lm(Affordability_Index ~ Cost_of_Living_Index + Rent_Index +
            Groceries_Index + Restaurant_Price_Index + Local_Purchasing_Power_Index, data = data)

# Summary of the model
summary(model)
```

```
Call:
lm(formula = Affordability_Index ~ Cost_of_Living_Index + Rent_Index +
    Groceries_Index + Restaurant_Price_Index + Local_Purchasing_Power_Index,
    data = data)

Residuals:
       1        2        3        4        5        6        7        8
```

```
-0.001415  0.013055 -0.005724 -0.004452 -0.013396  0.007853 -0.003349  0.002196
        9        10
-0.006960  0.012193

Coefficients:
                             Estimate Std. Error t value Pr(>|t|)
(Intercept)                 0.4830300  0.0402632  11.997 0.000277 ***
Cost_of_Living_Index       -0.0023175  0.0019017  -1.219 0.289931
Rent_Index                 -0.0018186  0.0006531  -2.784 0.049594 *
Groceries_Index            -0.0022796  0.0008908  -2.559 0.062694 .
Restaurant_Price_Index     -0.0007077  0.0009343  -0.758 0.490907
Local_Purchasing_Power_Index  0.0040277  0.0001751  23.006 2.12e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.01303 on 4 degrees of freedom
Multiple R-squared:  0.9961,    Adjusted R-squared:  0.9912
F-statistic: 203.6 on 5 and 4 DF,  p-value: 6.679e-05
```

The regression model demonstrates a strong overall fit, explaining 99% of the variability in the Affordability Index (Adjusted R-squared = 0.9912). Among the predictors, Local Purchasing Power Index stands out as the most significant variable, with a positive effect (Estimate = 0.0040, p < 0.001). This suggests that higher local purchasing power strongly enhances affordability. Conversely, Rent Index (Estimate = -0.0018, p = 0.0496) and Groceries Index (Estimate = -0.0023, p = 0.0627) negatively impact affordability, indicating that higher rent and grocery costs are associated with lower affordability, although groceries have a marginal significance. Other predictors, such as Cost of Living Index and Restaurant Price Index, do not significantly contribute to the model, implying a weaker relationship with affordability in the context of the given data.

In conclusion, affordability is primarily driven by purchasing power, with living expenses like rent and groceries acting as significant detractors. Policymakers or stakeholders seeking to enhance affordability should focus on improving local purchasing power while addressing housing and grocery costs to mitigate their adverse effects. Non-significant predictors, such as general cost of living or restaurant prices, may have less relevance in this context, but further analysis could explore their indirect effects or regional variations. The model's robustness, with minimal residual errors, supports the reliability of these insights.