Untitled

```
# Load required libraries
library(readxl)
library(ggplot2)
library(corrplot)
```

```
corrplot 0.95 loaded
```

```
library(car)  # For VIF check
```

```
Loading required package: carData
```

```
library(dplyr)
```

```

Attaching package: 'dplyr'
```

```
The following object is masked from 'package:car':

    recode
```

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

    filter, lag
```

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

    intersect, setdiff, setequal, union
```

```
library(e1071)  # For skewness
```

```
# Set file path
file_path <- "~/Desktop/Econometrics/WDI_1970and2010headers.xlsx"

# Read the dataset
df <- read_excel(file_path)
```

```
# Display first few rows
head(df)
```

```
# A tibble: 6 × 67
  country     wbcode PopDnst10 Inflation10 PctPaved10 Cellper100_10 Telper100_10
  <chr>       <chr>      <dbl>       <dbl>      <dbl>         <dbl>        <dbl>
1 Afghanistan AFG         43.5       0.893       NA            41.4       0.0529
2 Albania     ALB        115.        3.55        NA            84.0      10.4
3 Algeria     DZA         15.6       3.91        77.1          92.4       8.24
4 American S… ASM        278.       NA           NA            NA        15.2
5 Andorra     ADO        166.       NA           NA            77.2      45.0
6 Angola      AGO         15.7      14.5         NA            46.7       1.59
# ℹ 60 more variables: MilPctGDP10 <dbl>, GDP2005D_10 <dbl>, GDPGrwth_10 <dbl>,
#   GDPPC2010 <dbl>, GDPPC2005D_10 <dbl>, GDPPCG10 <dbl>, YouthLit10 <dbl>,
#   AdultLit10 <dbl>, FtMPrimPct10 <dbl>, FtMSecPct10 <dbl>, FtMTerPct10 <dbl>,
#   PrimCmpt10 <dbl>, Grade5_10 <dbl>, SecSchool_10 <dbl>, EdSpending10 <dbl>,
#   PctImprovedWater10 <dbl>, BedsPer1000_10 <dbl>, DrPer1000_10 <dbl>,
#   LPart10 <dbl>, LForceFem_10 <dbl>, Unemp10 <dbl>, NetMig10 <lgl>,
#   LIFEEXP10 <dbl>, Fertility10 <dbl>, DR_Working10 <dbl>, PopGrwth10 <dbl>, …
```

```
# Get dataset structure
str(df)
```

```
tibble [214 × 67] (S3: tbl_df/tbl/data.frame)
 $ country           : chr [1:214] "Afghanistan" "Albania" "Algeria" "American Samoa" ...
 $ wbcode            : chr [1:214] "AFG" "ALB" "DZA" "ASM" ...
 $ PopDnst10         : num [1:214] 43.5 115 15.6 278.2 165.8 ...
 $ Inflation10       : num [1:214] 0.893 3.552 3.913 NA NA ...
 $ PctPaved10        : num [1:214] NA NA 77.1 NA NA NA NA NA NA NA ...
 $ Cellper100_10     : num [1:214] 41.4 84 92.4 NA 77.2 ...
 $ Telper100_10      : num [1:214] 0.0529 10.3944 8.2404 15.2002 44.979 ...
 $ MilPctGDP10       : num [1:214] 3.67 1.6 3.52 NA NA ...
 $ GDP2005D_10       : num [1:214] 1.02e+10 1.07e+10 1.17e+11 NA NA ...
 $ GDPGrwth_10       : num [1:214] 8.43 3.5 3.6 NA NA ...
 $ GDPPC2010         : num [1:214] 561 3764 4350 NA NA ...
 $ GDPPC2005D_10     : num [1:214] 361 3405 3147 NA NA ...
 $ GDPPCG10          : num [1:214] 5.8 3.53 1.7 NA NA ...
 $ YouthLit10        : num [1:214] NA NA NA NA NA ...
 $ AdultLit10        : num [1:214] NA NA NA NA NA ...
 $ FtMPrimPct10      : num [1:214] 67.8 NA 93.9 NA NA ...
 $ FtMSecPct10       : num [1:214] 49.1 NA 103.7 NA NA ...
 $ FtMTerPct10       : num [1:214] NA 138 144 NA NA ...
 $ PrimCmpt10        : num [1:214] NA NA 92.6 NA NA ...
 $ Grade5_10         : num [1:214] NA 97.9 94.6 NA 91.6 ...
 $ SecSchool_10      : num [1:214] NA 98.1 92 NA NA ...
 $ EdSpending10      : num [1:214] NA NA NA NA NA ...
 $ PctImprovedWater10: num [1:214] 57.1 94.9 83.8 100 100 52.6 97.9 99 98.6 97.4 ...
 $ BedsPer1000_10    : num [1:214] 0.4 NA NA NA NA NA 2.2 4.5 NA NA ...
 $ DrPer1000_10      : num [1:214] 0.194 1.153 1.207 NA 3.912 ...
 $ LPart10           : num [1:214] 48.1 55.2 43.1 NA NA ...
 $ LForceFem_10      : num [1:214] 15.9 41.2 16.8 NA NA ...
 $ Unemp10           : num [1:214] NA 14.2 10 NA NA ...
 $ NetMig10          : logi [1:214] NA NA NA NA NA NA ...
 $ LIFEEXP10         : num [1:214] 59.6 77 70.6 NA NA ...
 $ Fertility10       : num [1:214] 5.66 1.74 2.82 NA 1.22 ...
 $ DR_Working10      : num [1:214] 103.1 49.1 46.7 NA NA ...
 $ PopGrwth10        : num [1:214] 2.4584 -0.0331 1.8504 -1.0887 -0.9606 ...
 $ PctUrban10        : num [1:214] 23.2 52.3 72 93 87.8 ...
 $ TourismExp10      : num [1:214] NA 1.45e+09 7.37e+08 NA NA ...
 $ PopDnst70         : num [1:214] 16.89 78.06 6.17 136.46 51.66 ...
 $ Inflation70       : num [1:214] NA NA 6.6 NA NA ...
 $ PctPaved70        : logi [1:214] NA NA NA NA NA NA ...
 $ Cellper100_70     : num [1:214] 0 0 0 0 0 0 0 0 0 0 ...
 $ Telper100_70      : num [1:214] 0.134 8.101 0.697 NA 7.757 ...
 $ MilPctGDP70       : logi [1:214] NA NA NA NA NA NA ...
 $ GDP2005D_70       : num [1:214] NA NA 2.84e+10 NA 6.99e+08 ...
 $ GDPGrwth_70       : num [1:214] NA NA 8.86 NA NA ...
 $ GDPPC2005D_70     : num [1:214] NA NA 1935 NA 28806 ...
 $ GDPPCG70          : num [1:214] NA NA 5.87 NA NA ...
 $ YouthLit70        : logi [1:214] NA NA NA NA NA NA ...
 $ AdultLit70        : num [1:214] NA NA NA NA NA NA NA NA NA NA ...
 $ FtMPrimPct70      : num [1:214] 17.2 NA NA NA NA ...
 $ FtMSecPct70       : num [1:214] 16.1 NA NA NA NA ...
 $ FtMTerPct70       : num [1:214] 17.5 NA NA NA NA ...
 $ PrimCmpt70        : num [1:214] NA NA NA NA NA ...
 $ Grade5_70         : num [1:214] NA NA NA NA NA ...
 $ SecSchool_70      : num [1:214] NA NA NA NA NA ...
 $ EdSpending70      : num [1:214] NA NA NA NA NA ...
 $ PctImprovedWater70: logi [1:214] NA NA NA NA NA NA ...
 $ Bedsper1000_70    : num [1:214] 0.199 NA 2.842 NA NA ...
 $ DrPer1000_70      : num [1:214] 0.0649 0.7395 0.1235 NA NA ...
 $ LPart70           : logi [1:214] NA NA NA NA NA NA ...
 $ LForceFem_70      : logi [1:214] NA NA NA NA NA NA ...
 $ Unemp70           : logi [1:214] NA NA NA NA NA NA ...
 $ NetMig70          : logi [1:214] NA NA NA NA NA NA ...
 $ LIFEEXP70         : num [1:214] 36.7 66.9 50.3 NA NA ...
 $ Fertility70       : num [1:214] 7.67 5.05 7.64 NA NA ...
 $ DR_working70      : num [1:214] 90.9 89.3 101.5 NA NA ...
```

```
 $ PopGrwth70        : num [1:214] 2.63 2.58 2.79 2.51 5.15 ...
 $ PctUrban70        : num [1:214] 11 31.7 39.5 70.4 80.2 ...
 $ TourismExp70      : logi [1:214] NA NA NA NA NA NA ...

# Get summary statistics
summary(df)
```

```
   country             wbcode            PopDnst10          Inflation10
Length:214         Length:214         Min.   :    0.139   Min.   :-2.425
Class :character   Class :character   1st Qu.:   33.804   1st Qu.: 1.794
Mode  :character   Mode  :character   Median :   81.378   Median : 3.522
                                      Mean   :  401.431   Mean   : 4.966
                                      3rd Qu.:  198.257   3rd Qu.: 6.295
                                      Max.   :19093.786   Max.   :85.070
                                      NA's   :1           NA's   :34
   PctPaved10     Cellper100_10      Telper100_10        MilPctGDP10
Min.   :  6.80   Min.   :  1.238   Min.   :  0.05291   Min.   :0.1531
1st Qu.: 35.50   1st Qu.: 61.236   1st Qu.:  3.38013   1st Qu.:1.0811
Median : 70.40   Median : 91.921   Median : 16.29694   Median :1.5579
Mean   : 63.57   Mean   : 90.247   Mean   : 20.86383   Mean   :1.9970
3rd Qu.: 92.10   3rd Qu.:119.491   3rd Qu.: 32.52392   3rd Qu.:2.6737
Max.   :100.00   Max.   :206.429   Max.   :121.19072   Max.   :8.5884
NA's   :145      NA's   :14        NA's   :10          NA's   :71
 GDP2005D_10         GDPGrwth_10       GDPPC2010         GDPPC2005D_10
Min.   :2.381e+07   Min.   :-9.529   Min.   :   211.3   Min.   :  150.7
1st Qu.:4.030e+09   1st Qu.: 1.674   1st Qu.:  1405.3   1st Qu.:  949.4
Median :1.828e+10   Median : 3.926   Median :  4666.3   Median : 3400.6
Mean   :2.756e+11   Mean   : 4.139   Mean   : 13350.6   Mean   :10360.3
3rd Qu.:1.172e+11   3rd Qu.: 6.948   3rd Qu.: 15198.8   3rd Qu.:13001.0
Max.   :1.360e+13   Max.   :27.499   Max.   :145229.8   Max.   :80276.0
NA's   :30          NA's   :28       NA's   :24         NA's   :30
   GDPPCG10          YouthLit10      AdultLit10     FtMPrimPct10
Min.   :-12.1821   Min.   :31.41   Min.   :25.31   Min.   : 67.80
1st Qu.:  0.3417   1st Qu.:87.46   1st Qu.:72.63   1st Qu.: 96.33
Median :  2.4530   Median :97.74   Median :91.85   Median : 98.86
Mean   :  2.5586   Mean   :90.72   Mean   :83.21   Mean   : 97.17
3rd Qu.:  4.7809   3rd Qu.:98.63   3rd Qu.:94.62   3rd Qu.: 99.92
Max.   : 24.3962   Max.   :99.75   Max.   :99.68   Max.   :113.36
NA's   :28         NA's   :179     NA's   :179     NA's   :61
 FtMSecPct10      FtMTerPct10       PrimCmpt10       Grade5_10
Min.   : 42.31   Min.   : 17.48   Min.   : 31.32   Min.   : 0.00
1st Qu.: 96.26   1st Qu.: 87.64   1st Qu.: 82.25   1st Qu.:75.28
Median : 99.95   Median :122.09   Median : 94.78   Median :92.04
Mean   : 98.18   Mean   :120.25   Mean   : 88.53   Mean   :83.91
3rd Qu.:104.15   3rd Qu.:144.55   3rd Qu.:100.66   3rd Qu.:96.86
Max.   :139.77   Max.   :549.50   Max.   :115.75   Max.   :99.94
NA's   :71       NA's   :91       NA's   :82       NA's   :128
 SecSchool_10    EdSpending10    PctImprovedWater10 BedsPer1000_10
Min.   :36.75   Min.   : 6.363   Min.   : 29.30     Min.   :0.100
1st Qu.:85.46   1st Qu.:11.276   1st Qu.: 82.03     1st Qu.:1.400
Median :95.65   Median :14.000   Median : 95.00     Median :2.600
Mean   :88.90   Mean   :15.174   Mean   : 87.79     Mean   :3.107
3rd Qu.:98.53   3rd Qu.:18.259   3rd Qu.: 99.50     3rd Qu.:4.580
Max.   :99.98   Max.   :34.000   Max.   :100.00     Max.   :8.250
NA's   :109     NA's   :119      NA's   :24         NA's   :115
 DrPer1000_10      LPart10       LForceFem_10      Unemp10      NetMig10
Min.   :0.008   Min.   :38.60   Min.   :11.89   Min.   : 0.40   Mode:logical
1st Qu.:0.284   1st Qu.:56.40   1st Qu.:38.09   1st Qu.: 5.00   NA's:214
Median :1.595   Median :63.50   Median :43.70   Median : 7.65
Mean   :1.772   Mean   :63.57   Mean   :40.85   Mean   : 9.25
3rd Qu.:2.879   3rd Qu.:70.25   3rd Qu.:47.01   3rd Qu.:11.88
Max.   :6.723   Max.   :89.30   Max.   :53.01   Max.   :32.00
NA's   :67      NA's   :31      NA's   :31      NA's   :112
   LIFEEXP10      Fertility10     DR_Working10      PopGrwth10
Min.   :44.84   Min.   :1.003   Min.   : 16.54   Min.   :-3.3335
```

```
1st Qu.:64.69   1st Qu.:1.800   1st Qu.: 46.20   1st Qu.: 0.4673
Median :72.91   Median :2.344   Median : 52.96   Median : 1.2499
Mean   :69.95   Mean   :2.879   Mean   : 59.82   Mean   : 1.4000
3rd Qu.:76.61   3rd Qu.:3.736   3rd Qu.: 73.02   3rd Qu.: 2.2759
Max.   :83.16   Max.   :7.584   Max.   :110.12   Max.   :11.2153
NA's   :16      NA's   :14      NA's   :19
  PctUrban10      TourismExp10         PopDnst70          Inflation70
Min.   : 10.64   Min.   :5.500e+05   Min.   :    0.136   Min.   :-5.296
1st Qu.: 37.60   1st Qu.:1.445e+08   1st Qu.:   12.847   1st Qu.: 2.572
Median : 57.39   Median :5.915e+08   Median :   38.498   Median : 4.091
Mean   : 57.39   Mean   :5.683e+09   Mean   :  220.089   Mean   : 5.023
3rd Qu.: 76.58   3rd Qu.:3.670e+09   3rd Qu.:  112.699   3rd Qu.: 6.536
Max.   :100.00   Max.   :1.098e+11   Max.   :11742.000   Max.   :16.306
NA's   :4        NA's   :38          NA's   :10          NA's   :122
PctPaved70     Cellper100_70  Telper100_70      MilPctGDP70
Mode:logical   Min.   :0     Min.   : 0.02583   Mode:logical
NA's:214       1st Qu.:0     1st Qu.: 0.28901   NA's:214
               Median :0     Median : 1.57604
               Mean   :0     Mean   : 5.66580
               3rd Qu.:0     3rd Qu.: 7.38054
               Max.   :0     Max.   :44.79761
               NA's   :18    NA's   :74
 GDP2005D_70         GDPGrwth_70     GDPPC2005D_70        GDPPCG70
Min.   :8.358e+07   Min.   :-5.650   Min.   :  144.6   Min.   :-8.494
1st Qu.:1.780e+09   1st Qu.: 3.188   1st Qu.:  652.7   1st Qu.: 1.263
Median :6.301e+09   Median : 5.806   Median : 1905.7   Median : 3.380
Mean   :1.221e+11   Mean   : 6.690   Mean   : 6685.1   Mean   : 4.449
3rd Qu.:5.390e+10   3rd Qu.: 8.685   3rd Qu.: 8582.0   3rd Qu.: 6.565
Max.   :4.340e+12   Max.   :30.999   Max.   :75504.7   Max.   :25.953
NA's   :99          NA's   :104      NA's   :99        NA's   :105
YouthLit70       AdultLit70   FtMPrimPct70      FtMSecPct70
Mode:logical   Min.   :96    Min.   :  5.363   Min.   :  3.52
NA's:214       1st Qu.:96    1st Qu.: 70.080   1st Qu.: 53.29
               Median :96    Median : 92.651   Median : 82.89
               Mean   :96    Mean   : 83.331   Mean   : 78.00
               3rd Qu.:96    3rd Qu.: 98.786   3rd Qu.:102.12
               Max.   :96    Max.   :150.735   Max.   :136.84
               NA's   :213   NA's   :172       NA's   :177
 FtMTerPct70       PrimCmpt70       Grade5_70      SecSchool_70
Min.   : 0.927   Min.   : 23.98   Min.   :23.58   Min.   :28.64
1st Qu.:21.416   1st Qu.: 43.60   1st Qu.:43.16   1st Qu.:44.75
Median :48.001   Median : 64.10   Median :53.44   Median :49.71
Mean   :43.630   Mean   : 59.53   Mean   :52.25   Mean   :56.46
3rd Qu.:63.167   3rd Qu.: 75.71   3rd Qu.:67.80   3rd Qu.:77.03
Max.   :79.354   Max.   :101.95   Max.   :74.92   Max.   :86.50
NA's   :189      NA's   :197      NA's   :200     NA's   :203
 EdSpending70    PctImprovedWater70 Bedsper1000_70     DrPer1000_70
Min.   : 6.791   Mode:logical       Min.   : 0.1353   Min.   :0.01030
1st Qu.:11.206   NA's:214           1st Qu.: 1.3617   1st Qu.:0.07893
Median :15.885                      Median : 2.9955   Median :0.27110
Mean   :15.856                      Mean   : 4.2640   Mean   :0.49926
3rd Qu.:20.541                      3rd Qu.: 5.5112   3rd Qu.:0.72938
Max.   :28.433                      Max.   :26.7123   Max.   :3.20000
NA's   :183                         NA's   :80        NA's   :66
LPart70        LForceFem_70   Unemp70        NetMig70         LIFEEXP70
Mode:logical   Mode:logical   Mode:logical   Mode:logical   Min.   :32.39
NA's:214       NA's:214       NA's:214       NA's:214       1st Qu.:49.07
                                                           Median :60.24
                                                           Mean   :58.11
                                                           3rd Qu.:68.24
                                                           Max.   :74.65
                                                           NA's   :23
 Fertility70     DR_working70      PopGrwth70       PctUrban70
Min.   :1.828   Min.   : 45.25   Min.   :-2.796   Min.   :  2.382
1st Qu.:3.093   1st Qu.: 68.04   1st Qu.: 1.228   1st Qu.: 21.182
```

```
 Median :5.785   Median : 85.42   Median : 2.226   Median : 39.451
 Mean   :5.140   Mean   : 82.23   Mean   : 2.171   Mean   : 42.075
 3rd Qu.:6.697   3rd Qu.: 94.52   3rd Qu.: 2.819   3rd Qu.: 59.736
 Max.   :8.231   Max.   :120.76   Max.   :14.062   Max.   :100.000
 NA's   :24      NA's   :23       NA's   :4        NA's   :6
 TourismExp70
 Mode:logical
 NA's:214
```

```
# List all variable names
colnames(df)
```

```
 [1] "country"            "wbcode"             "PopDnst10"
 [4] "Inflation10"        "PctPaved10"         "Cellper100_10"
 [7] "Telper100_10"       "MilPctGDP10"        "GDP2005D_10"
[10] "GDPGrwth_10"        "GDPPC2010"          "GDPPC2005D_10"
[13] "GDPPCG10"           "YouthLit10"         "AdultLit10"
[16] "FtMPrimPct10"       "FtMSecPct10"        "FtMTerPct10"
[19] "PrimCmpt10"         "Grade5_10"          "SecSchool_10"
[22] "EdSpending10"       "PctImprovedWater10" "BedsPer1000_10"
[25] "DrPer1000_10"       "LPart10"            "LForceFem_10"
[28] "Unemp10"            "NetMig10"           "LIFEEXP10"
[31] "Fertility10"        "DR_Working10"       "PopGrwth10"
[34] "PctUrban10"         "TourismExp10"       "PopDnst70"
[37] "Inflation70"        "PctPaved70"         "Cellper100_70"
[40] "Telper100_70"       "MilPctGDP70"        "GDP2005D_70"
[43] "GDPGrwth_70"        "GDPPC2005D_70"      "GDPPCG70"
[46] "YouthLit70"         "AdultLit70"         "FtMPrimPct70"
[49] "FtMSecPct70"        "FtMTerPct70"        "PrimCmpt70"
[52] "Grade5_70"          "SecSchool_70"       "EdSpending70"
[55] "PctImprovedWater70" "Bedsper1000_70"     "DrPer1000_70"
[58] "LPart70"            "LForceFem_70"       "Unemp70"
[61] "NetMig70"           "LIFEEXP70"          "Fertility70"
[64] "DR_working70"       "PopGrwth70"         "PctUrban70"
[67] "TourismExp70"
```

```
# Choose final predictor variables and dependent variable
selected_vars <- c("GDPPC2005D_10", "PctImprovedWater10", "Fertility10", "DrPer1000_10", "LIFEEXP10")

# Subset the dataset
df_selected <- df[, selected_vars]

# Remove rows with missing values
df_selected <- na.omit(df_selected)

# Compute the correlation matrix
cor_matrix <- cor(df_selected, use = "complete.obs")

# Print correlation matrix
print(cor_matrix)
```

```
                  GDPPC2005D_10 PctImprovedWater10 Fertility10 DrPer1000_10
GDPPC2005D_10         1.0000000          0.4783199  -0.4510438    0.5321634
PctImprovedWater10    0.4783199          1.0000000  -0.7926809    0.6148819
Fertility10          -0.4510438         -0.7926809   1.0000000   -0.6928636
DrPer1000_10          0.5321634          0.6148819  -0.6928636    1.0000000
LIFEEXP10             0.6259871          0.7572874  -0.8332348    0.6804356
                   LIFEEXP10
GDPPC2005D_10      0.6259871
```

```
PctImprovedWater10  0.7572874
Fertility10        -0.8332348
DrPer1000_10        0.6804356
LIFEEXP10           1.0000000

# Visualize correlation matrix
corrplot(cor_matrix, method = "color", addCoef.col = "black", tl.col = "black", tl.srt = 45)
```

```
# Function to plot histograms
plot_histogram <- function(data, column_name) {
  ggplot(data, aes(x = .data[[column_name]])) +
    geom_histogram(fill = "blue", color = "black", bins = 30, alpha = 0.7) +
    labs(title = paste("Histogram of", column_name), x = column_name, y = "Frequency") +
    theme_minimal()
}

# Generate histograms for all selected variables
for (var in selected_vars) {
  print(plot_histogram(df_selected, var))
}
```

```
# Check skewness
skew_values <- sapply(df_selected, skewness, na.rm = TRUE)
print("Skewness of variables:")
```

```
[1] "Skewness of variables:"
```

```
print(skew_values)
```

```
     GDPPC2005D_10 PctImprovedWater10        Fertility10       DrPer1000_10
         1.7762897         -1.3553390           1.0418142          0.5902870
         LIFEEXP10
        -0.8284149
```

```
# Apply log transformation to GDP (adding 1 to avoid log(0))
df_selected$log_GDPPC2005D_10 <- log(df_selected$GDPPC2005D_10 + 1)

# Function to create scatterplots with regression lines
plot_scatter <- function(data, predictor, response = "LIFEEXP10") {
  ggplot(data, aes(x = .data[[predictor]], y = .data[[response]])) +
    geom_point(color = "blue", alpha = 0.6) +
    geom_smooth(method = "lm", color = "red", se = FALSE) +
    labs(title = paste("Scatterplot of", response, "vs", predictor),
         x = predictor, y = response) +
    theme_minimal()
}

# Generate scatterplots
```

```
for (var in selected_vars[-length(selected_vars)]) {  # Excluding LIFEEXP10
  print(plot_scatter(df_selected, var))
}
```

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

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

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

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

```
# Run the regression model with log-transformed GDP
model <- lm(LIFEEXP10 ~ log_GDPPC2005D_10 + PctImprovedWater10 + Fertility10 + DrPer1000_10, data =
         df_selected)

# Display summary of the regression model
summary(model)
```

```

Call:
lm(formula = LIFEEXP10 ~ log_GDPPC2005D_10 + PctImprovedWater10 +
    Fertility10 + DrPer1000_10, data = df_selected)

Residuals:
     Min       1Q   Median       3Q      Max
-25.4568  -1.5314   0.6091   2.1722   6.7855

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)        55.23635    4.75157  11.625  < 2e-16 ***
log_GDPPC2005D_10   2.26845    0.38133   5.949 2.42e-08 ***
PctImprovedWater10  0.04081    0.04132   0.988    0.325
Fertility10        -2.60788    0.45624  -5.716 7.28e-08 ***
DrPer1000_10        0.35331    0.34049   1.038    0.301
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.141 on 128 degrees of freedom
Multiple R-squared:  0.7916,    Adjusted R-squared:  0.7851
F-statistic: 121.6 on 4 and 128 DF,  p-value: < 2.2e-16

# Extract coefficients and p-values
coefficients <- summary(model)$coefficients

# Print significance levels of each variable
significance <- ifelse(coefficients[,4] < 0.05, "Significant", "Not Significant")
data.frame(Variable = rownames(coefficients), Estimate = coefficients[,1], P_Value =
         coefficients[,4], Significance = significance)
```

```
                           Variable   Estimate      P_Value    Significance
(Intercept)                (Intercept) 55.2363520 9.095436e-22     Significant
log_GDPPC2005D_10   log_GDPPC2005D_10  2.2684482 2.419849e-08     Significant
```

```
PctImprovedWater10 PctImprovedWater10  0.0408120 3.251201e-01 Not Significant
Fertility10               Fertility10 -2.6078824 7.282015e-08     Significant
DrPer1000_10             DrPer1000_10  0.3533059 3.013947e-01 Not Significant

# Check for multicollinearity using VIF
vif_values <- vif(model)
print(vif_values)

 log_GDPPC2005D_10 PctImprovedWater10        Fertility10       DrPer1000_10
          2.912445           3.192851           3.476835           2.168066

# Create a dummy variable for high GDP per capita (above median)
df_selected$High_GDP <- ifelse(df_selected$GDPPC2005D_10 > median(df_selected$GDPPC2005D_10, na.rm =
         TRUE), 1, 0)

# Create an interaction term between High GDP and Fertility Rate
df_selected$Interaction_Term <- df_selected$High_GDP * df_selected$Fertility10

# Run regression with interaction term
model_interaction <- lm(LIFEEXP10 ~ log_GDPPC2005D_10 + PctImprovedWater10 + Fertility10 +
         DrPer1000_10 + Interaction_Term, data = df_selected)

# Display regression results
summary(model_interaction)


Call:
lm(formula = LIFEEXP10 ~ log_GDPPC2005D_10 + PctImprovedWater10 +
    Fertility10 + DrPer1000_10 + Interaction_Term, data = df_selected)

Residuals:
     Min       1Q   Median       3Q      Max
-24.0181  -1.4560   0.2736   2.4961   7.4298

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)        50.76367    5.30705   9.565  < 2e-16 ***
log_GDPPC2005D_10   2.78782    0.47287   5.896 3.16e-08 ***
PctImprovedWater10  0.05144    0.04136   1.244    0.216
Fertility10        -2.49715    0.45617  -5.474 2.26e-07 ***
DrPer1000_10        0.29642    0.33886   0.875    0.383
Interaction_Term   -1.01862    0.55750  -1.827    0.070 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.104 on 127 degrees of freedom
Multiple R-squared:  0.797, Adjusted R-squared:  0.789
F-statistic: 99.71 on 5 and 127 DF,  p-value: < 2.2e-16

# Residual vs. Fitted plot (check homoscedasticity)
plot(model_interaction, which = 1)
```

```
# Q-Q plot (check normality of residuals)
plot(model_interaction, which = 2)
```

```
# Scale-Location plot (checks variance homogeneity)
plot(model_interaction, which = 3)
```

```
# Cook's Distance plot (identify influential points)
plot(model_interaction, which = 4)
```