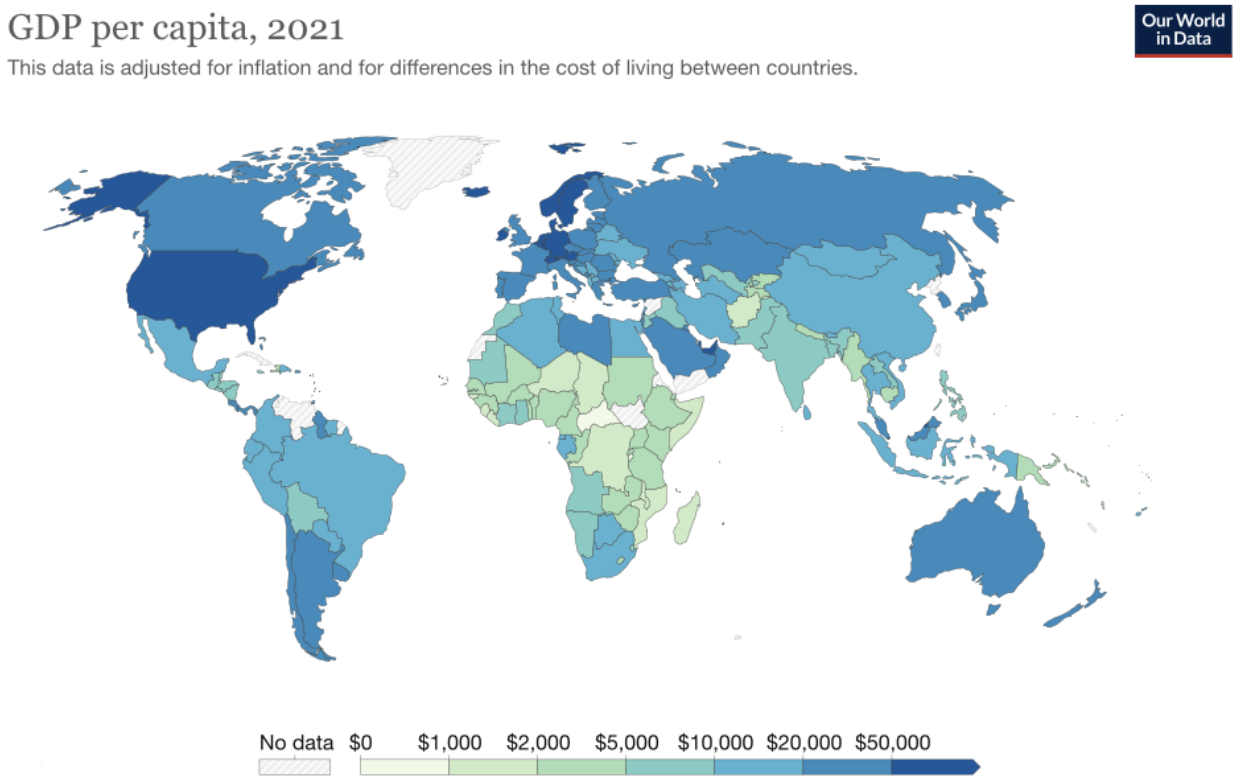
# The Puzzle of Wealth: Understanding Per Capita GDP Pieces

NARB




(GDP per capita, 2021)

## Introduction

GDP is a crucial indicator of a nation's financial well-being as it reflecting average incomes, levels of consumption, job availability, and other wealth indicators. In our model GDP per capita is measured in US dollars and represents the average economic output per person for each represented country. Understanding the factors influencing GDP helps countries know what key issues to focus on successfully promote economic growth. Our research questions are as followed:

- Does a high fertility rate correlate with lower average GDP?
- Is there a relationship between average population age and GDP?
- What factor most significantly predicts GDP?

Low fertility rates may signal higher parental incomes, allowing greater investment in quality education, potentially indicating a higher average GDP (Doepke, 2022). A higher average age might be a sign of increased labor productivity or, conversely, a higher dependency rate. If neither fertility rates nor average age emerges as the primary GDP predictor, our dataset, combining 2023 US Census Bureau and Focus Economics data, enables exploration of other influential factors. The Census Bureau gathers information through censuses and surveys, while Focus Economics compiles monthly data from diverse international forecasts (International Database, 2023; About Us, 2023). The data used for this project contained country-level information across 165 nations on quantitative and categorical factors like life expectancy, dependency, growth rates, income, etc. The target variable for prediction modeling was the log-transformed version of GDP per capita calculated in USD.

## Methods and Analysis

First, we cleaned the data by filtering out N/A values, converting the structures of the objects, and reformatting the position and ordering for easier access. Then, we conducted an initial exploratory analysis by plotting all quantitative variables with GDP on scatterplots to understand the relationships between these variables and GDP. We analyzed qualitative factors like country size, income level, and growth rate with boxplots and interaction plots. Multicollinearity and variable screening through stepwise regression helped identify highly correlated predictor variables. The removal and addition of quantitative predictors, like fertility and dependency, through variable screens improved model accuracy.

The remaining significant quantitative predictors were average age, life expectancy, and their interaction. After testing combinations of qualitative factors through comparison of model fit

statistics, we chose to include income because it was significant. Our final model contained average age, life expectancy, their interaction, and income level classified into four levels and three dummy variables: high income (the control), low income, lower middle income, upper middle income. The best model was chosen through a Global F test, Nested F test on income levels, t-tests on important predictors, R^2 values, and 2s values. The linear regression model was tested for violations of assumptions through residual analysis split into residual plots, influence analysis, and outlier detection. From the residual plots, we confirmed that the zero mean and constant variance assumptions were not violated. From the Q-Q plot and histogram, we confirmed that the normality assumption was not violated. However, after careful analysis of the influential observations, we noticed five influential points that can be categorized as outliers due to their influence on the data: Angola, San Marino, Seychelles, Sierra Leone, and Ukraine. With cook's distance, studentized residuals, and leverage plots, we used a subsetting technique to confirm that the removal of these data points sharpened the model by increasing the R^2 and F Statistics, and reducing the residual standard error and P-value. Lastly, we applied a box-cox transformation to check sensitivity to transformations and learned that the log transformation applied to the GDP was significant and contributed to the accurate linearization of the data.

## Results

We concluded that our model was adequate with an R^2 of 0.8733, meaning 87.33% of the variability observed in GDP per capita is explained by the regression model. Comparing this to the first model with a 0.6393 R^2 and the pre-residual analysis model with a 0.7396 R^2, our latest and final model accounts for the best parameters and fits our data most accurately. From this, we used our model for estimation and prediction and calculated the confidence and prediction intervals for E(y) and y, respectively. Using data within our frame, we accurately predicted multiple parameters such as the GDP, Average Age, and more of a country given values using 95% confidence intervals. Thus, our final regression model stands as a statistically sound and reliable tool for explaining and predicting GDP per capita variability between countries.

## Conclusions

From our data analysis, we found that fertility rate has a moderate relationship with GDP. As the average number of children the women in a country will have increases, GDP per capita decreases by a rate of 0.67. However, fertility rate displays extreme multicollinearity

with growth rate (0.86), average age (-0.85), and dependency (0.859). These relationships suggest that these explanatory variables provide similar information about GDP per capita, so it can't be determined that fertility rate alone is a significant predictor. Alternatively, while average age presents some concerns of multicollinearity, we confirmed through a t-test and tests for violations of assumptions that it is significant for predicting GDP. We concluded that for each one year increase in the log of average age, the log of GDP per capita decreases by 5.2341. Finally, we found that income level is also a significant predictor. If a country is classified as a low income country, the log of GDP per capita decreases by 3.0819, in lower middle income countries the log of GDP per capita decreases by 0.9872, in upper middle income countries the log of GDP per capita decreases by 0.9782, and in high income countries, the log of GDP per capita doesn't increase or decrease. Ultimately, an increase in income is associated with an expected decrease in the log of GDP per capita.

Our dataset has limitations, including the exclusion of some countries present in only one data source due to resource constraints, leading to a bias against smaller or less developed nations. Additionally, variations in census collection policies among countries pose challenges, with some, like the United States, updating data every 10 years, while others, such as Mauritania, may have outdated information since 2008. Moreover, concerns about source accuracy arise, The International Dataset relies partially on estimations and is subject to sample errors (Jacobsen, 2023). Additionally, the data collected in 2020 may be inaccurate or unreliable due to COVID-19 pandemic.

Future steps for our research include comparisons between other years' census data and considering other variables in our analysis. Because many countries collect census data every 10 years, we could revisit this model in 2030 to observe what trends might have continued or changed over time. Another benefit of this would be that many countries may be more resourced and able to collect data with ample time to recover from the effects of the pandemic. Similarly, we could look at data from 2010 to observe which countries have improved or worsened GDP growth rates. By viewing trends across a 20 year time span, we can make more accurate predictions about GDP. We could also try adding a country's capital as an explanatory variable to predict the trajectory of GDP. Capital in this case refers to the supply of things like buildings, machinery, software, etc (The Conference Board, 2022). An understanding of capital in combination with labor and total factor productivity, which refers to efficiency) will give a more comprehensive picture of GDP.

## Appendix A: Data Dictionary

| Variable Name | Measured | Units/Levels | Description | Data Source |
|---|---|---|---|---|
| GDP | GDP per capita by country | USD | Gross Domestic Product (GDP) per capita is the total economic output (the sum of all goods and services produced within a country's borders) divided by the population of the country. It represents the average income per person in the country. | World Bank |
| Life_Expectancy | Life expectancy by country | Years | Life expectancy is the average number of years a person is expected to live if current mortality rates remain the same throughout their lifetime. | US Census Bureau |

| Variable Name | Measured | Units/Levels | Description | Data Source |
|---|---|---|---|---|
| Density | Population density by country | People per square kilometer | Population density is the average number of people living per square kilometer of land. Population density provides a data point on how crowded or urban a country is. | US Census Bureau |
| Dependency | Dependency ratio by country | Percentage | The dependency ratio is the ratio of the dependent population to the working-age population. The dependent population is people aged 0-14 and 65+. The working-age population is people ages 15-64. | US Census Bureau |

| Variable Name | Measured | Units/Levels | Description | Data Source |
|---|---|---|---|---|
| Fertility | Total fertility ratio by country | Children per woman | The total fertility rate or TFR is the average number of children that a woman has in the country. | US Census Bureau |
| Sex_Ratio | Gender ratio by country | Ratio of males to females | The gender ratio is the number of males to the number of females in the country. | US Census Bureau |
| Debt | Debt as a percentage of GDP by country | Percentage of GDP | Debt as a percentage of GDP is the ratio of a country's total debt to its Gross Domestic Product/GDP. | Focus Economics |

| Variable Name | Measured | Units/Levels | Description | Data Source |
|---|---|---|---|---|
| Income | Income group of countries | High income, low income, middle income | Income group classification is determined by a country's Gross National Income (GNI) per capita. Lower Income: $1,135 or less of GNI per capita. Middle Income: between $1,136 and $13,845. High Income: $113,846 or more of GNI per capita. | World Bank |
| Growth | Population growth rate by country | Negative, Average, or Rapid | Population growth rate is the percentage change in a country's population from one year to the next. For reference, the World's population growth rate in 2022 was .9%. | US Census Bureau |

| Variable Name | Measured | Units/Levels | Description | Data Source |
|---|---|---|---|---|
| Size | Size of the country by population | Small, medium, or large | The size of a country by population is the total number of people in the country, including all land within its borders. | US Census Bureau |

## Appendix B: Data Rows

| | Name | Size | Growth | Density | Fertility | Life_Expectancy | Sex_Ratio |
|---|---|---|---|---|---|---|---|
| 1 | Albania | M | A | 113.2 | 1.55 | 79.7 | 0.97 |
| 2 | Algeria | L | R | 19.4 | 2.97 | 77.8 | 1.03 |
| 3 | Andorra | S | N | 182.6 | 1.46 | 83.6 | 1.05 |
| 4 | Angola | L | R | 28.9 | 5.76 | 62.5 | 0.96 |
| 5 | Antigua and Barbuda | S | R | 229.1 | 1.94 | 78.0 | 0.89 |
| 6 | Argentina | L | A | 17 | 2.17 | 78.6 | 0.98 |
| 7 | Armenia | M | N | 106 | 1.65 | 76.4 | 0.96 |
| 8 | Australia | L | R | 3.4 | 1.73 | 83.3 | 0.99 |
| 9 | Austria | L | A | 108.4 | 1.51 | 82.5 | 0.96 |
| 10 | Azerbaijan | L | A | 128.3 | 1.69 | 74.9 | 1.00 |

| | Dependency | Income | Debt | GDP |
|---|---|---|---|---|
| 1 | 48.1 | Upper middle income | 73.32 | 6802.805 |
| 2 | 60.5 | Lower middle income | 9.06 | 4273.922 |
| 3 | 46.7 | High income | 40.90 | 41992.793 |
| 4 | 98.1 | Lower middle income | 64.24 | 2998.501 |
| 5 | 47.2 | High income | 99.10 | 18745.174 |
| 6 | 56.7 | Upper middle income | 52.13 | 13686.009 |
| 7 | 48.1 | Upper middle income | 46.89 | 7014.207 |
| 8 | 54.1 | High income | 37.64 | 64491.430 |
| 9 | 53.6 | High income | 86.19 | 52131.447 |
| 10 | 45.4 | Upper middle income | 28.32 | 7736.701 |

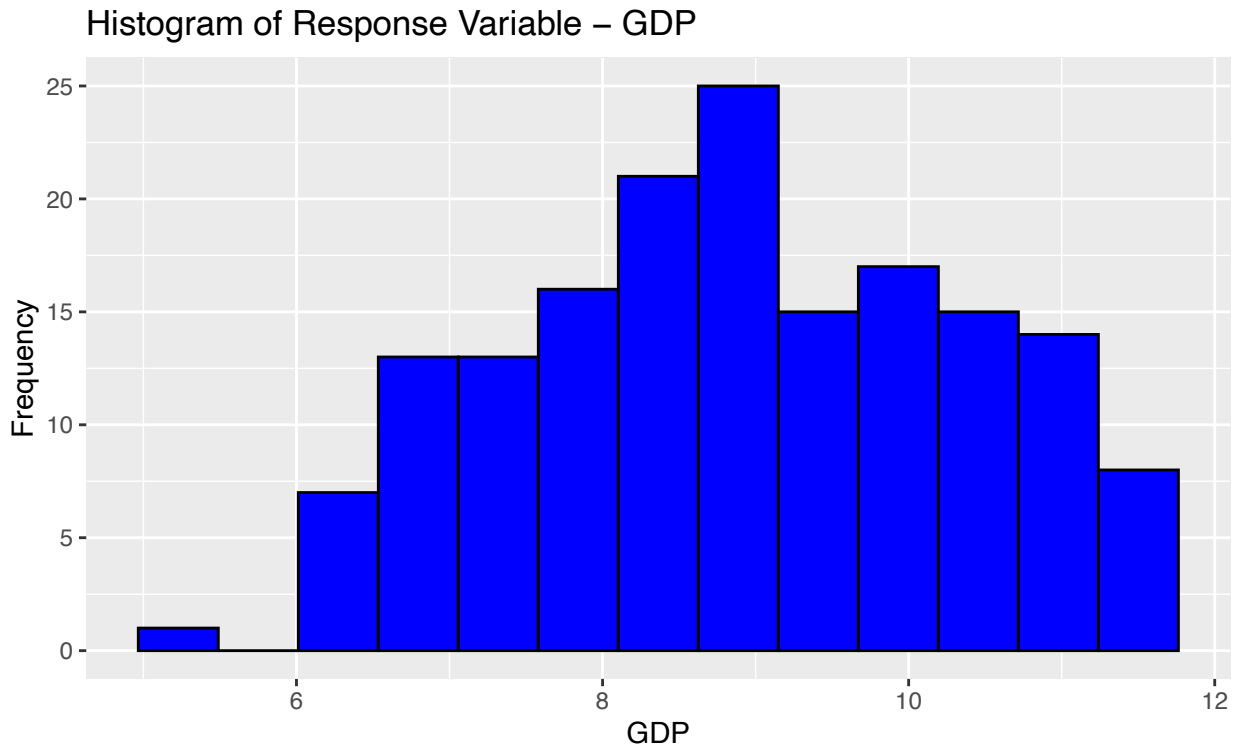
## Appendix C: Final Model Output and Plots

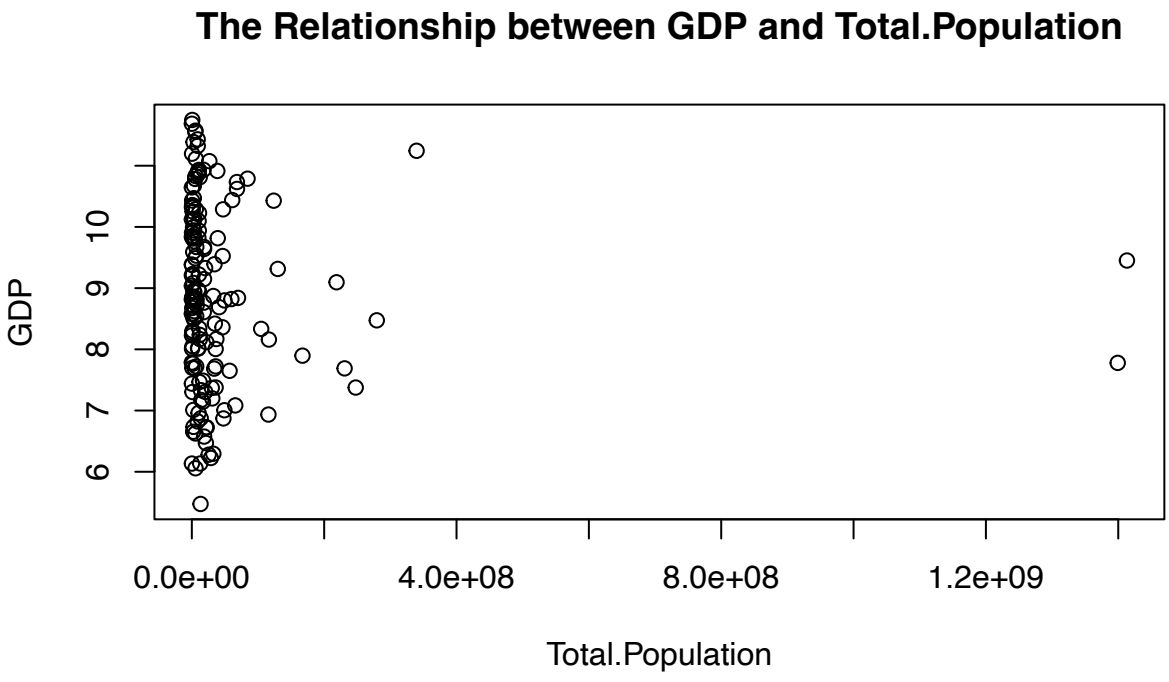
```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

## corrplot 0.92 loaded

##
## Attaching package: 'Hmisc'

## The following objects are masked from 'package:dplyr':
##
##     src, summarize

## The following objects are masked from 'package:base':
##
##     format.pval, units

## Loading required package: carData

##
## Attaching package: 'car'

## The following object is masked from 'package:dplyr':
##
##     recode

##
## Attaching package: 'olsrr'

## The following object is masked from 'package:datasets':
##
##     rivers

##
## Attaching package: 'MASS'

## The following object is masked from 'package:olsrr':
##
##     cement

## The following object is masked from 'package:dplyr':
##
##     select
```

# EDA(Plots)

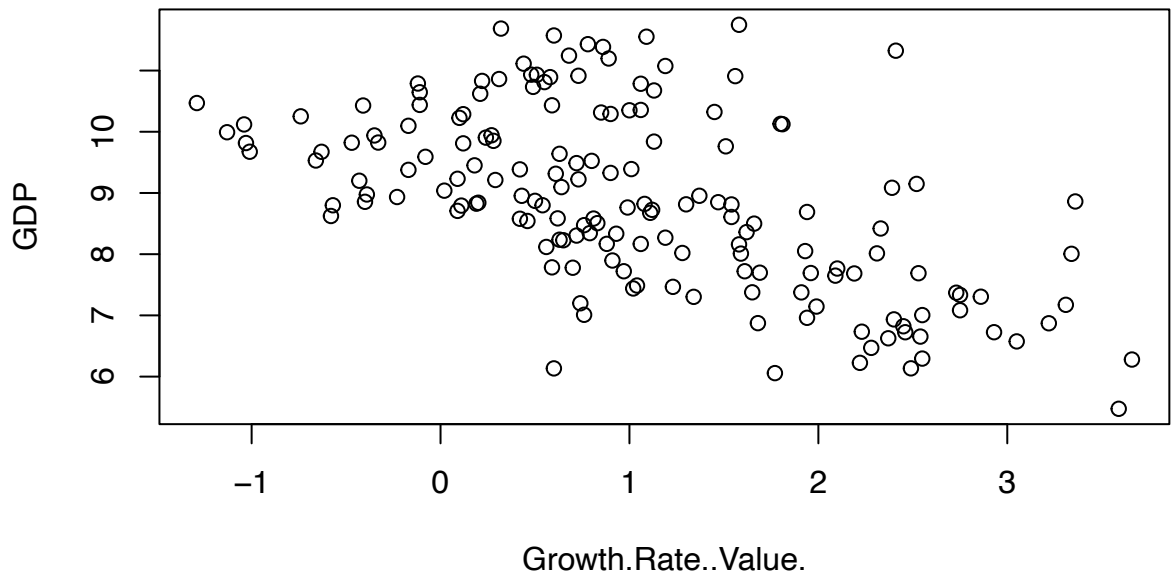
```
##                  Name Total.Population Growth.Rate..Value. Density Fertility
## 1             Albania          3101621                0.19   113.2      1.55
## 2             Algeria         46286076                1.62    19.4      2.97
## 3             Andorra            85468               -0.11   182.6      1.46
## 4              Angola         35981281                3.34    28.9      5.76
## 5 Antigua and Barbuda           101489                1.13   229.1      1.94
## 6           Argentina         46621847                0.80    17.0      2.17
##   Life_Expectancy Under.5.Mortality.Rate Sex_Ratio Dependency Average.Age  Debt
## 1            79.7                  12.3      0.97       48.1    3.577948 73.32
## 2            77.8                  21.7      1.03       60.5    3.363842  9.06
## 3            83.6                   4.1      1.05       46.7    3.873282 40.90
## 4            62.5                  87.1      0.96       98.1    2.785011 64.24
## 5            78.0                  16.8      0.89       47.2    3.514526 99.10
## 6            78.6                  10.8      0.98       56.7    3.496508 52.13
##   Size Growth              Income       GDP
## 1    M      A Upper middle income  8.825090
## 2    L      R Lower middle income  8.360287
## 3    S      N         High income 10.645253
## 4    L      R Lower middle income  8.005868
## 5    S      R         High income  9.838692
## 6    L      A Upper middle income  9.524129
```

```
## 'data.frame':    165 obs. of  15 variables:
##  $ Name                  : chr  "Albania" "Algeria" "Andorra" "Angola" ...
##  $ Total.Population      : num  3101621 46286076 85468 35981281 101489 ...
##  $ Growth.Rate..Value.   : num  0.19 1.62 -0.11 3.34 1.13 0.8 -0.4 1.19 0.31 0.43 ...
##  $ Density               : num  113.2 19.4 182.6 28.9 229.1 ...
##  $ Fertility             : num  1.55 2.97 1.46 5.76 1.94 2.17 1.65 1.73 1.51 1.69 ...
##  $ Life_Expectancy       : num  79.7 77.8 83.6 62.5 78 78.6 76.4 83.3 82.5 74.9 ...
##  $ Under.5.Mortality.Rate: num  12.3 21.7 4.1 87.1 16.8 10.8 13.5 3.6 3.8 13.4 ...
##  $ Sex_Ratio             : num  0.97 1.03 1.05 0.96 0.89 0.98 0.96 0.99 0.96 1 ...
##  $ Dependency            : num  48.1 60.5 46.7 98.1 47.2 56.7 48.1 54.1 53.6 45.4 ...
##  $ Average.Age           : num  3.58 3.36 3.87 2.79 3.51 ...
##  $ Debt                  : num  73.32 9.06 40.9 64.24 99.1 ...
##  $ Size                  : Factor w/ 3 levels "L","M","S": 2 1 3 1 3 1 2 1 1 1 ...
##  $ Growth                : Factor w/ 3 levels "A","N","R": 1 3 2 3 3 1 2 3 1 1 ...
##  $ Income                : Factor w/ 4 levels "High income",..: 4 3 1 3 1 4 4 1 1 4 ...
##  $ GDP                   : num  8.83 8.36 10.65 8.01 9.84 ...
```

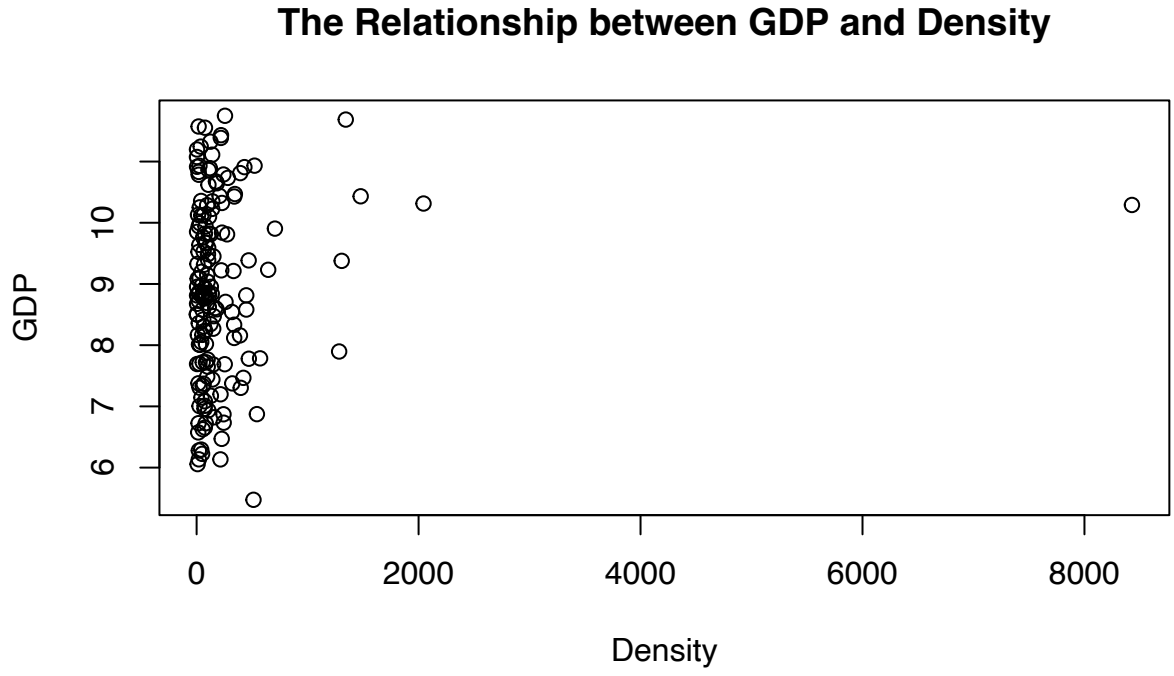

Histogram of Response Variable – GDP
Frequency
25
20
15
10
5
0
6
8
10
12
GDP


The Relationship between GDP and Total.Population
GDP
6
7
8
9
10
0.0e+00
4.0e+08
8.0e+08
1.2e+09
Total.Population

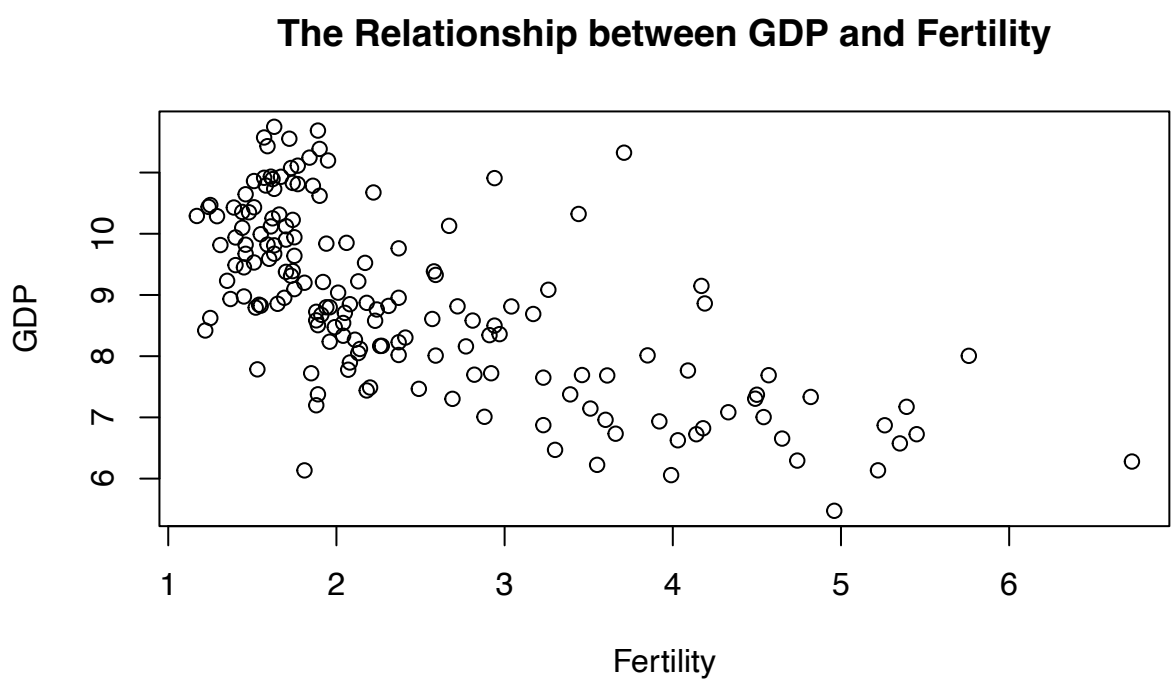
```
## [1] -0.03447192
```



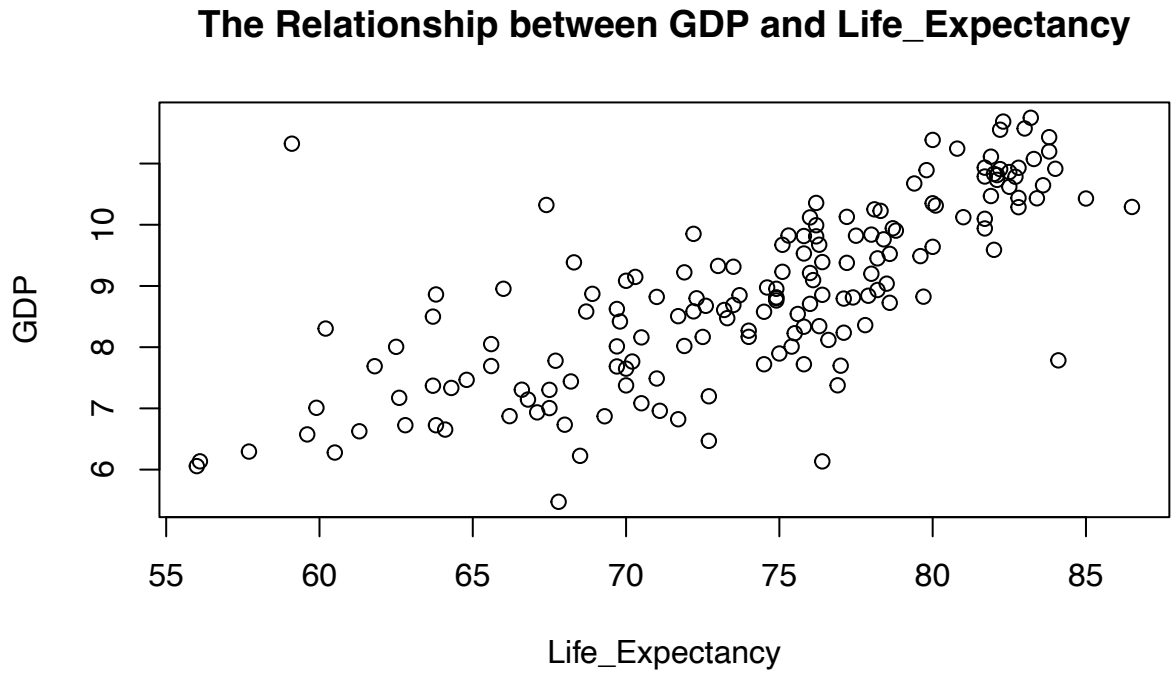

```
## [1] -0.580973
```



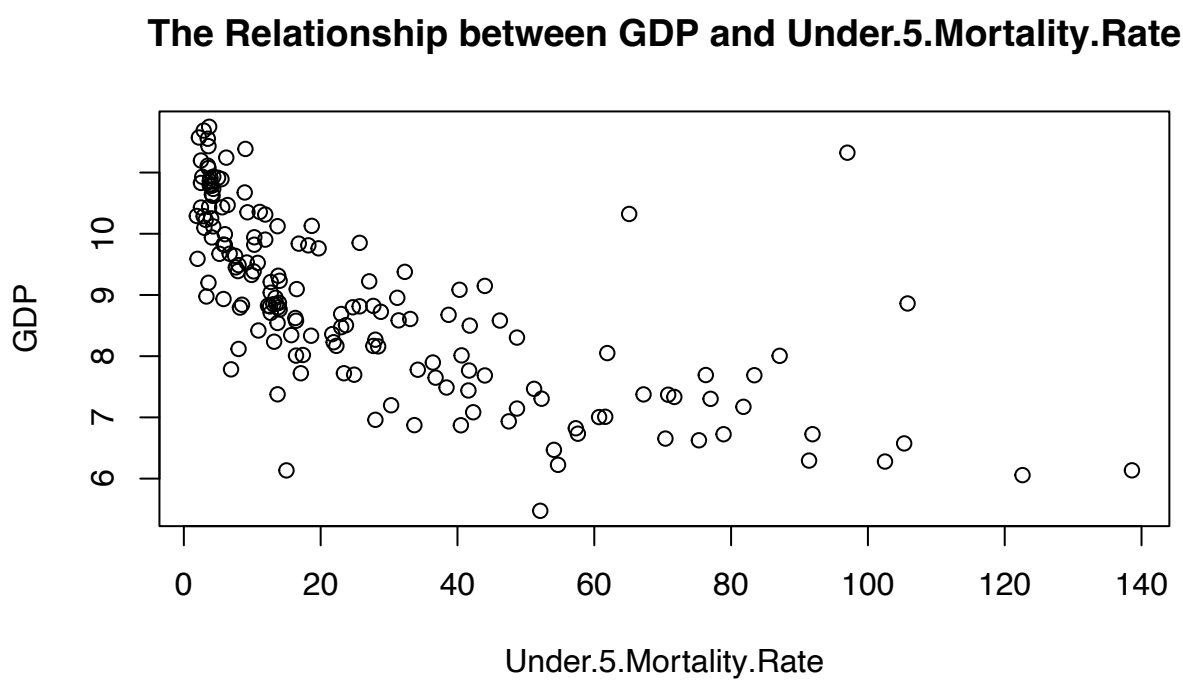

```
## [1] 0.119799
```



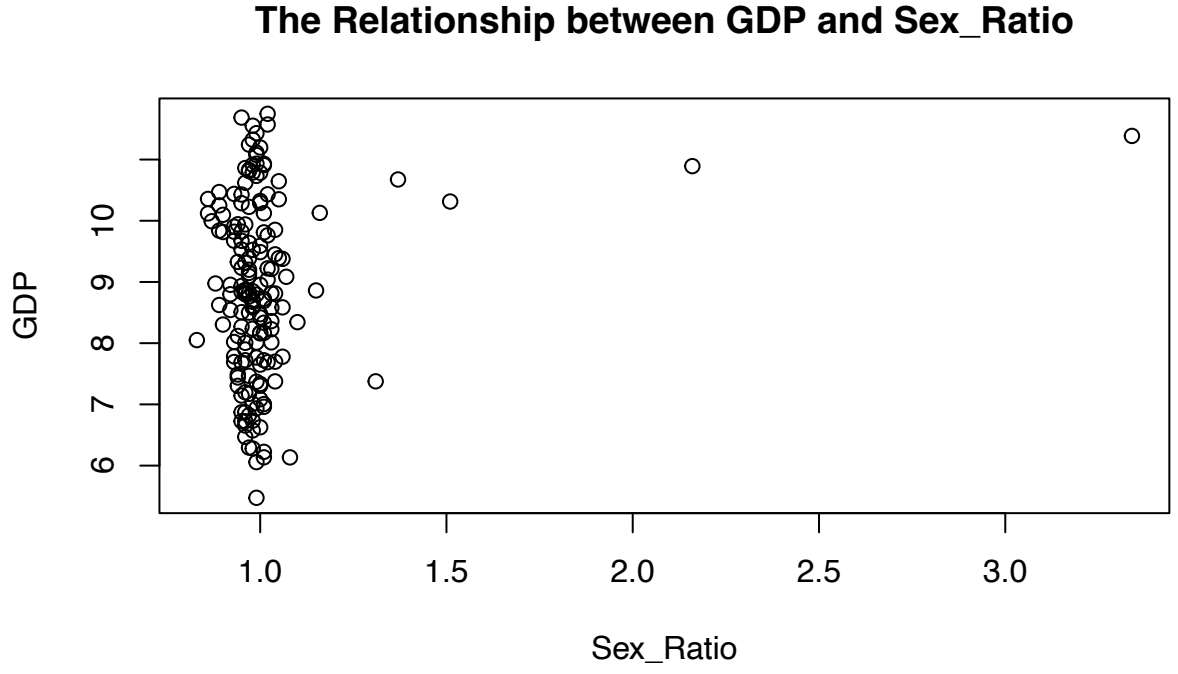

```
## [1] -0.6707076
```



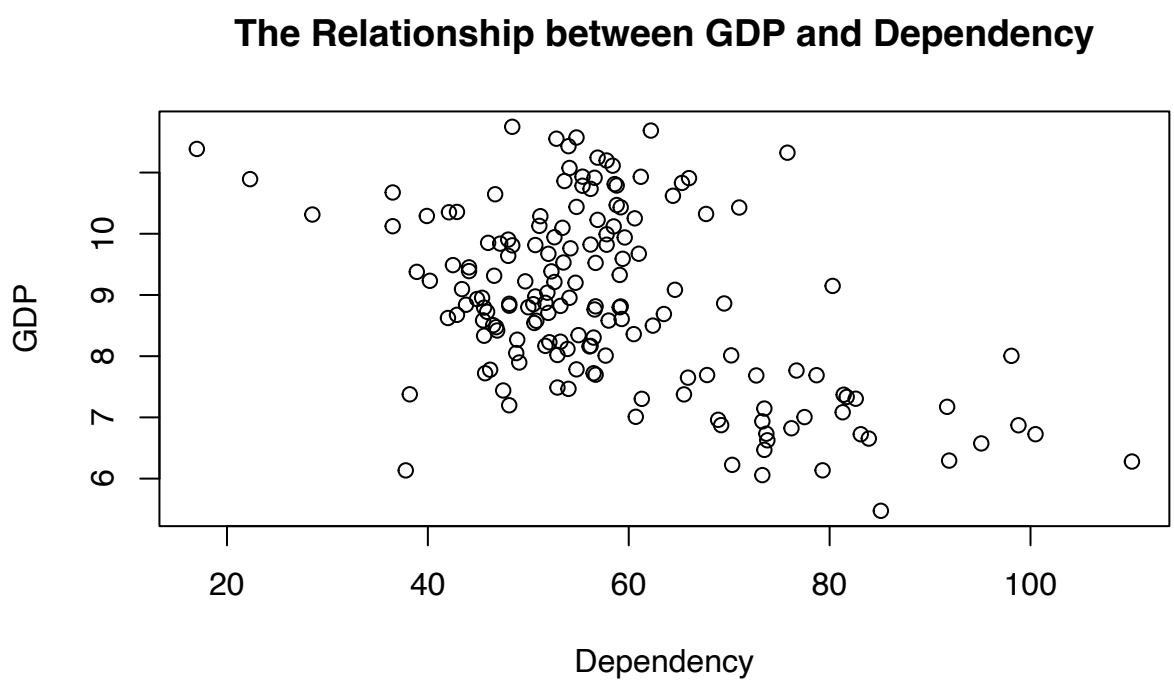

```
## [1] 0.7482856
```



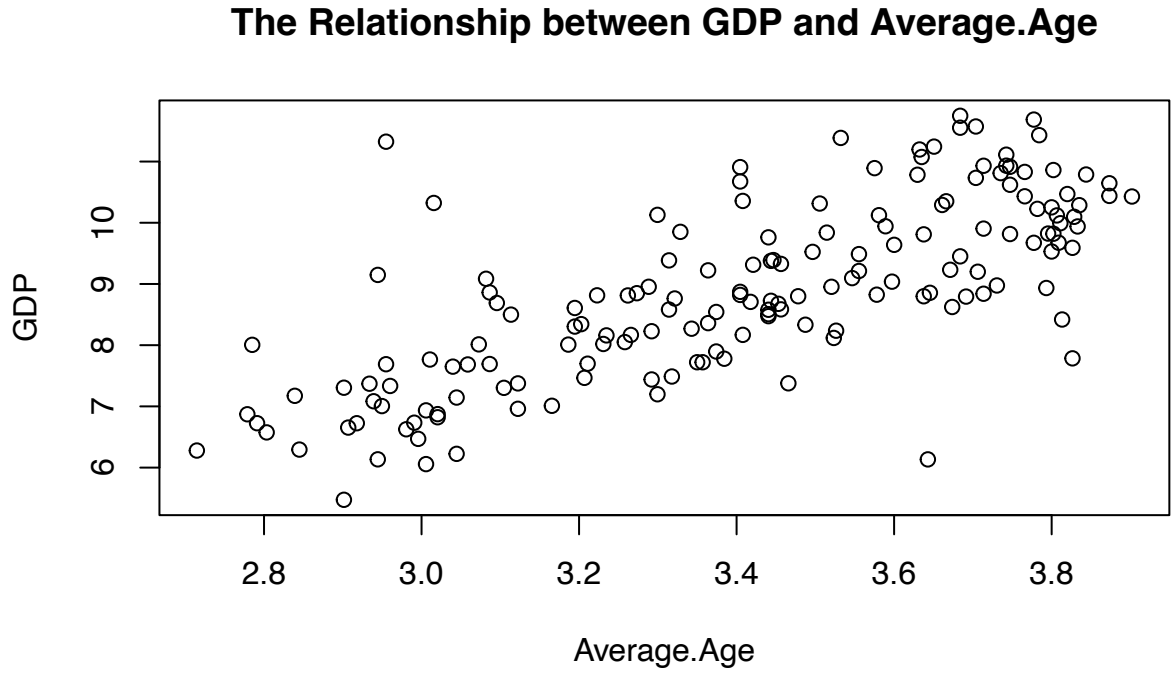

```
## [1] -0.6833988
```



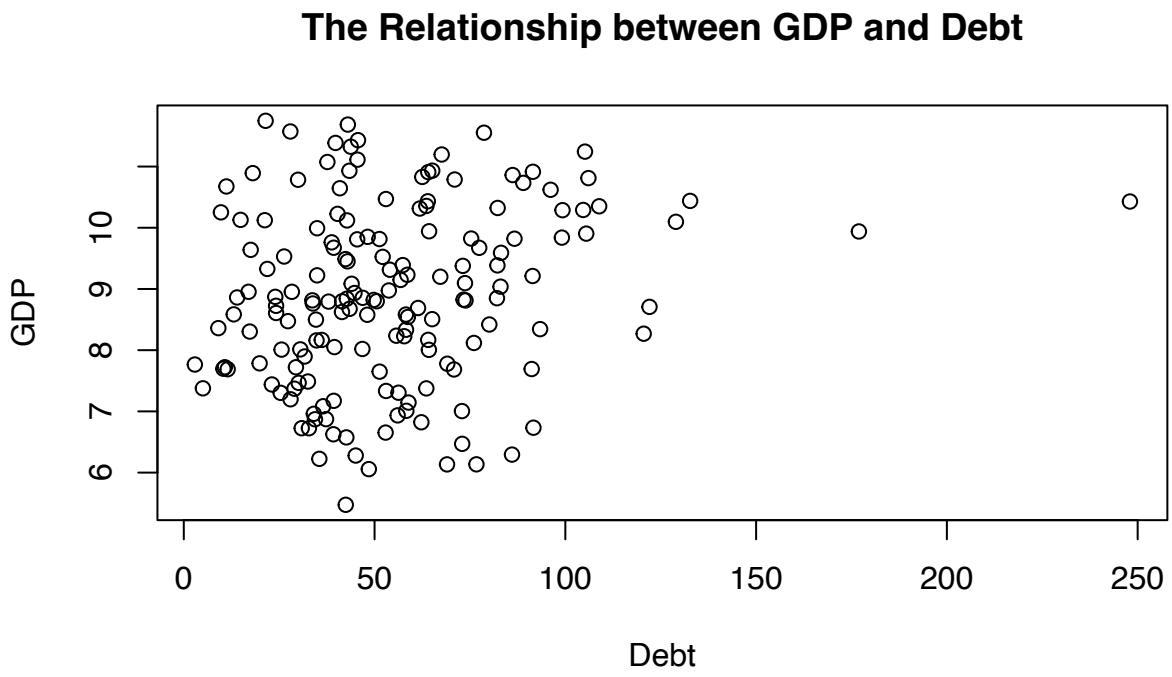

```
## [1] 0.168265
```




```
## [1] -0.4901131
```



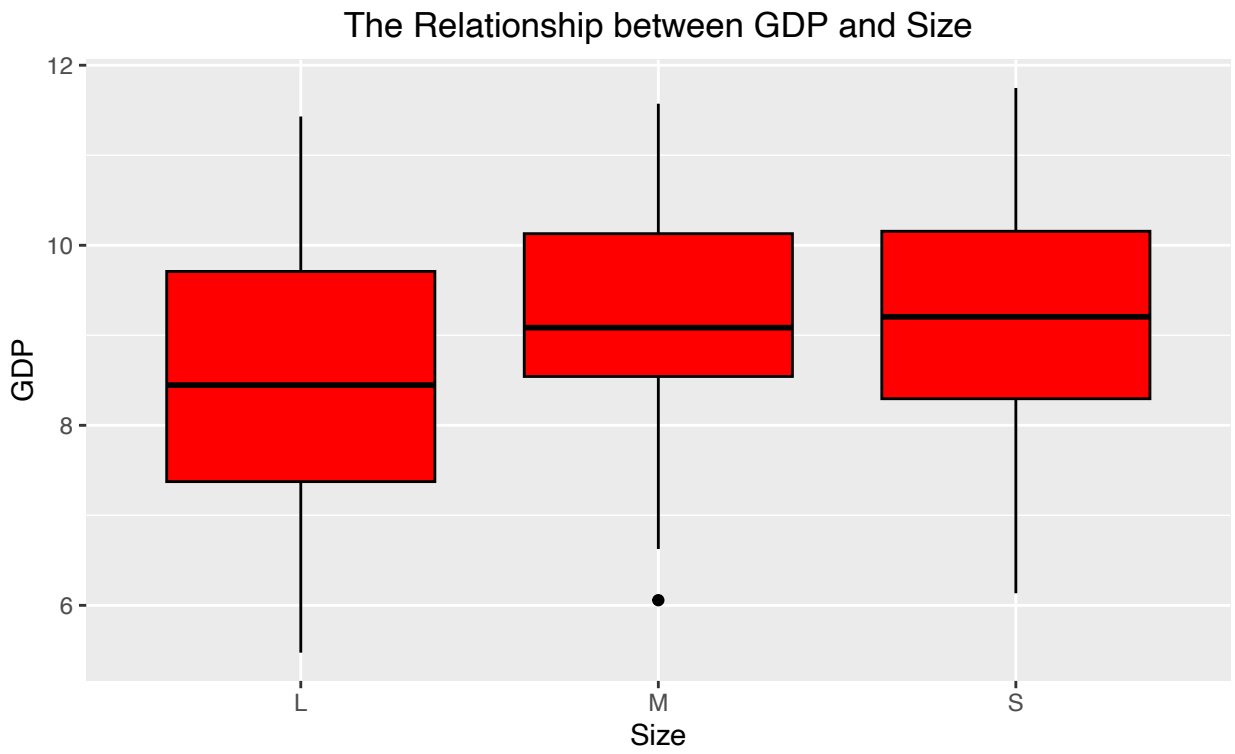

```
## [1] 0.7420579
```



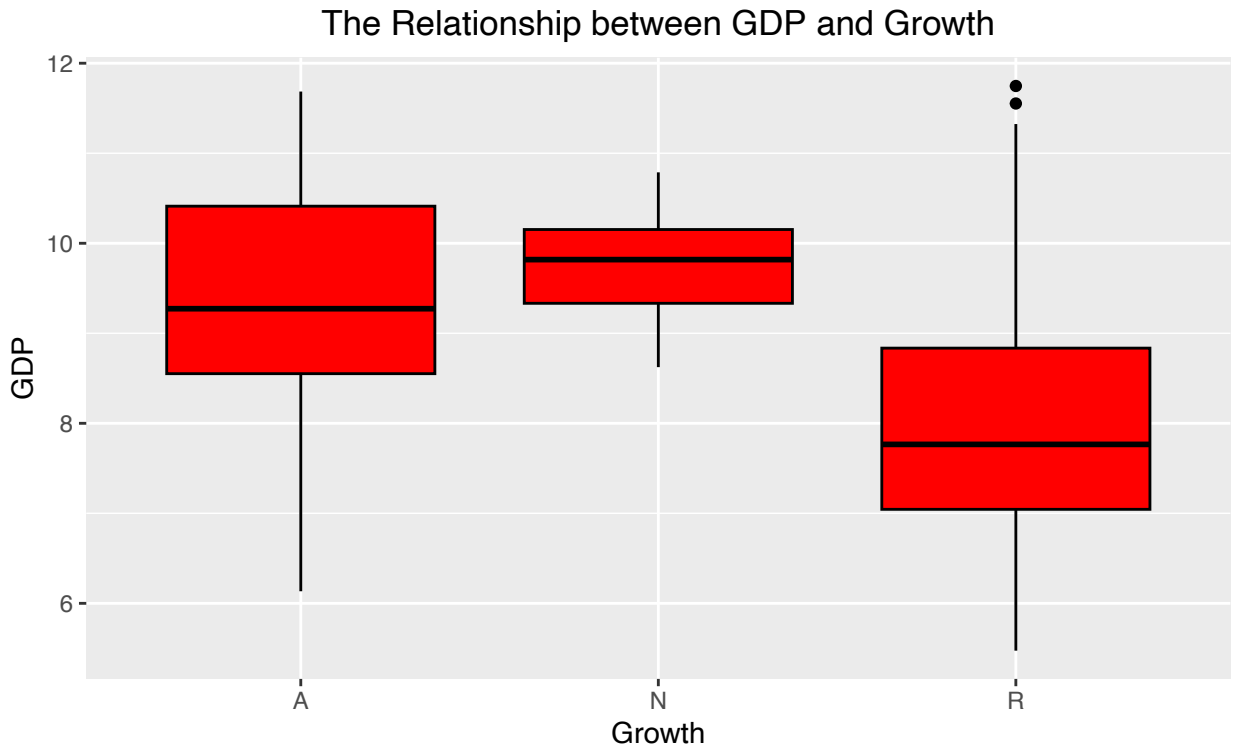

```
## [1] 0.2105487
```

```
## Warning: ‘aes_string()‘ was deprecated in ggplot2 3.0.0.
## i Please use tidy evaluation idioms with ‘aes()‘.
## i See also ‘vignette("ggplot2-in-packages")‘ for more information.
## This warning is displayed once every 8 hours.
## Call ‘lifecycle::last_lifecycle_warnings()‘ to see where this warning was
## generated.
```

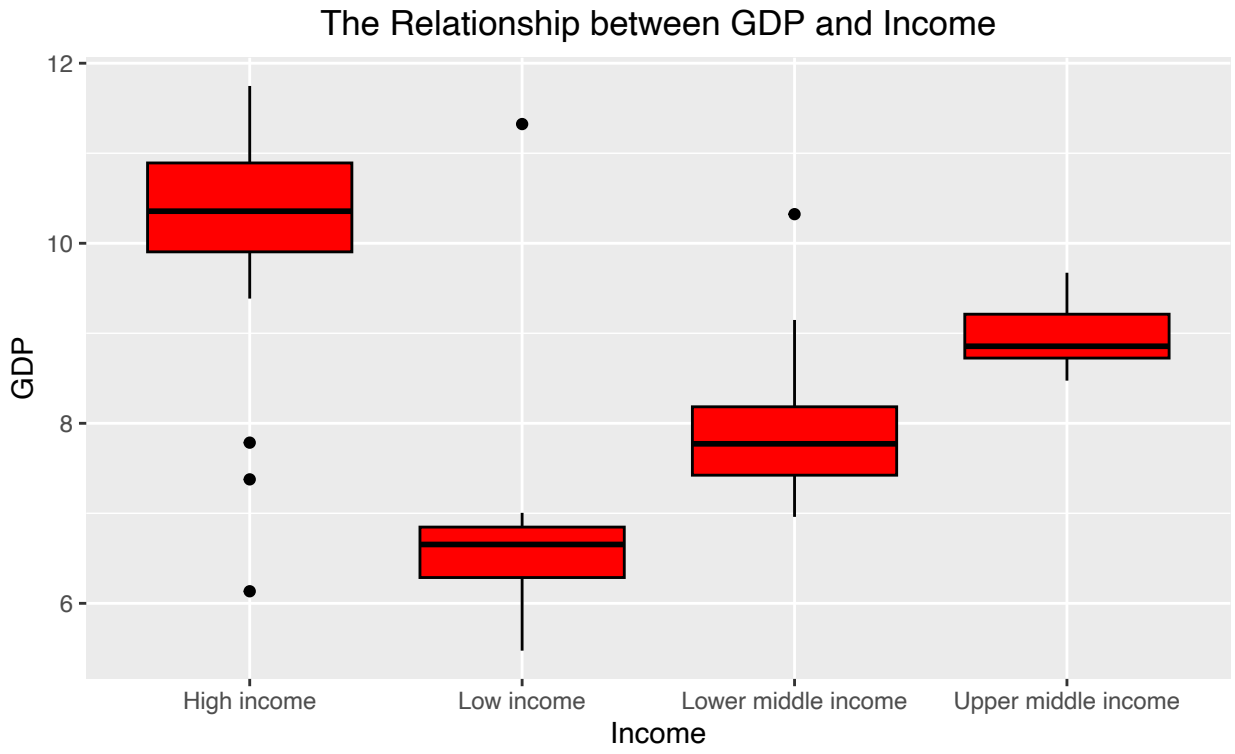
```
## $L
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   5.474   7.374   8.447   8.603   9.709  11.431
##
## $M
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   6.057   8.542   9.085   9.199  10.129  11.573
##
## $S
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   6.134   8.295   9.206   9.205  10.155  11.747
```

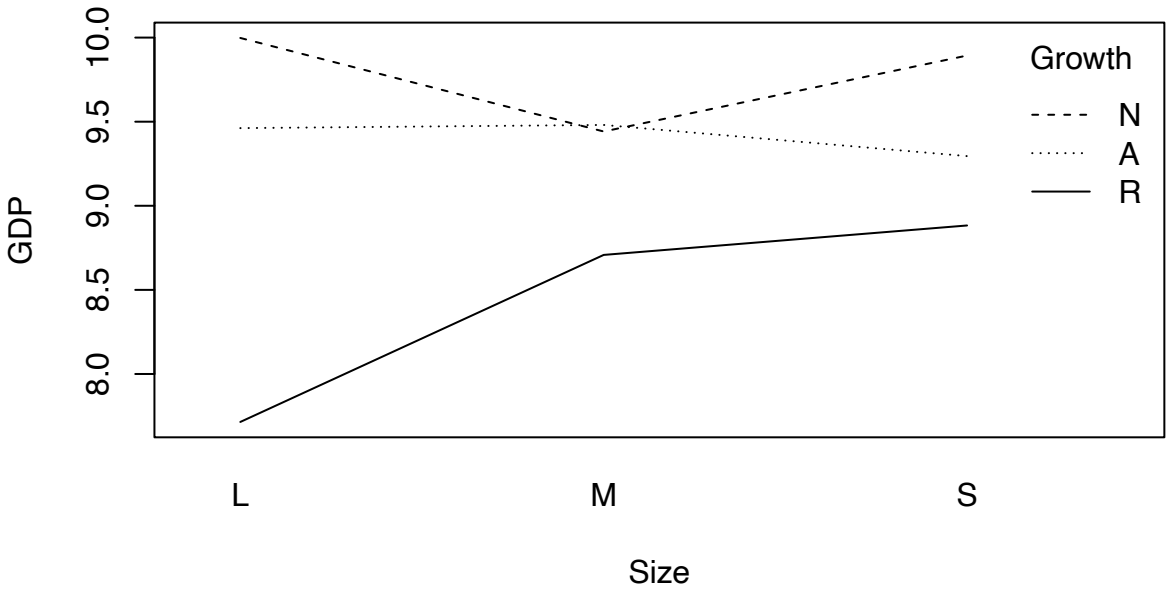



```
## $A
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   6.134   8.551   9.273   9.418  10.412  11.686
##
## $N
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   8.624   9.333   9.818   9.744  10.153  10.788
##
## $R
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   5.474   7.044   7.766   8.136   8.835  11.747
```



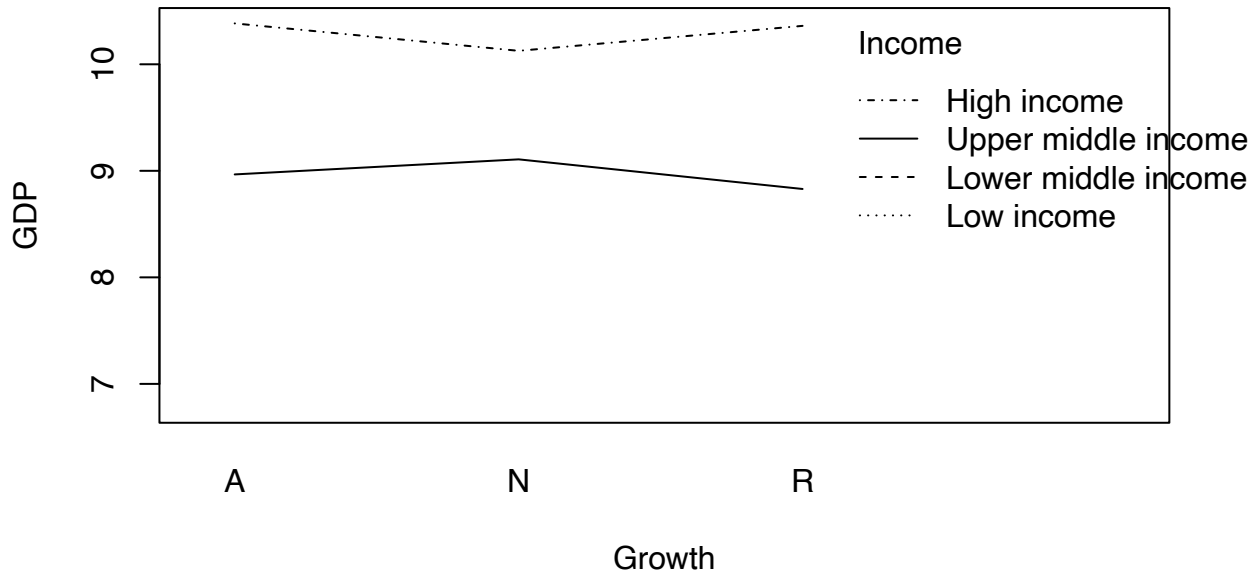

```
## $`High income`
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   6.134   9.904  10.356  10.311  10.892  11.747
##
## $`Low income`
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   5.474   6.286   6.654   6.779   6.847  11.324
##
## $`Lower middle income`
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   6.961   7.424   7.772   7.867   8.183  10.323
##
## $`Upper middle income`
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   8.474   8.725   8.856   8.955   9.212   9.672
```


The Relationship between GDP and Income
12
10
8
6
GDP
High income
Low income
Lower middle income
Upper middle income
Income

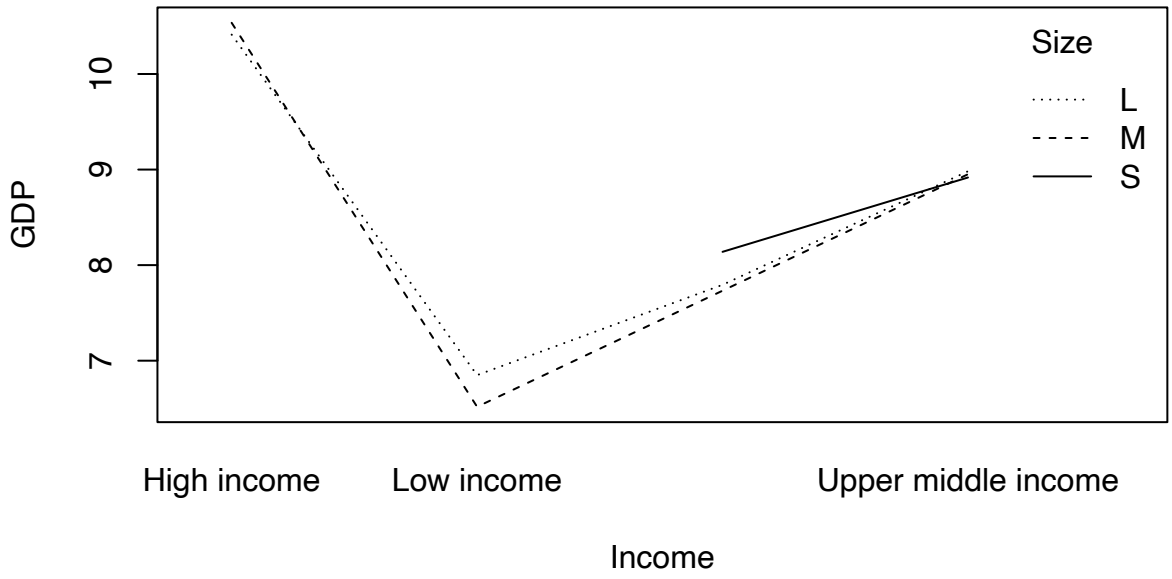

GDP
10.0
9.5
9.0
8.5
8.0
Growth
N
A
R
L
M
S
Size

```
##
##      A  N  R
##   L 33  9 46
```

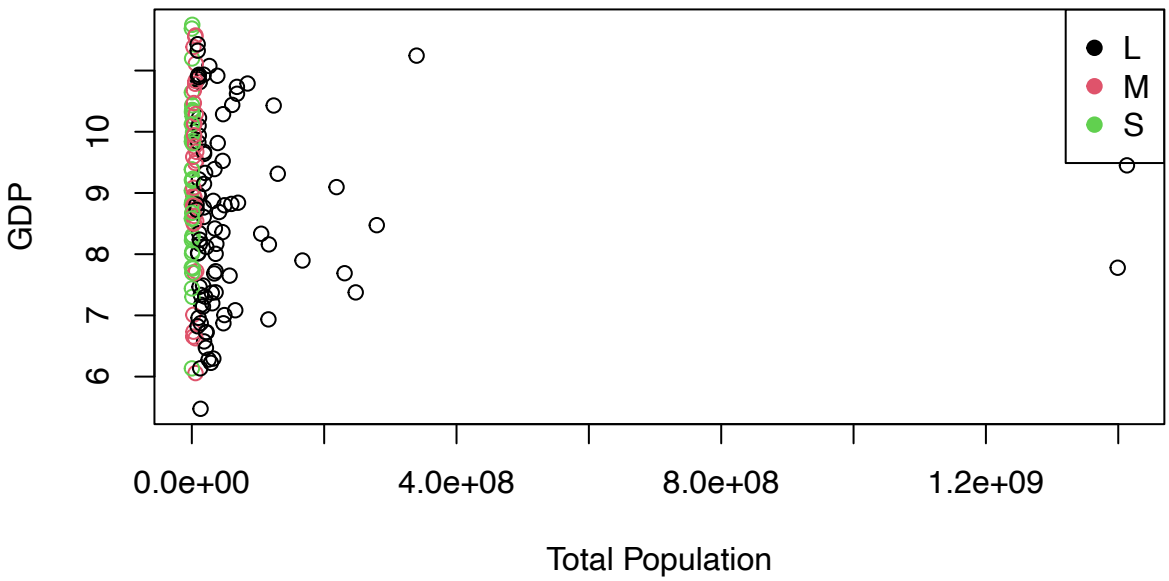
```
##   M 14 10 13
##   S 19  5 16
```

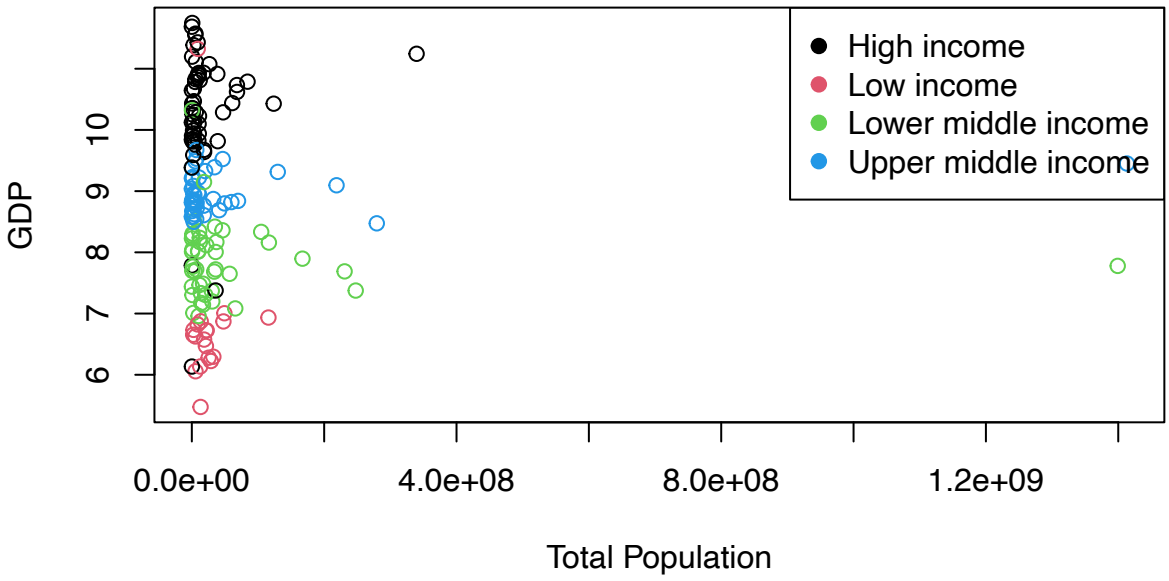


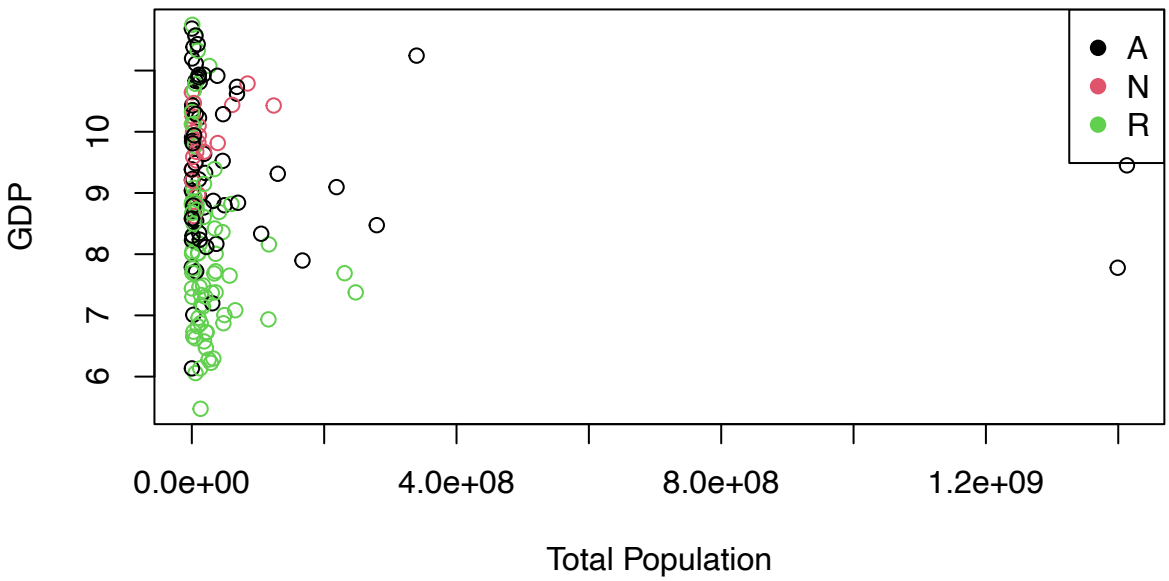

```
##
##     High income Low income Lower middle income Upper middle income
##   A          30          0                  13                  23
##   N          15          0                   0                   9
##   R          12         19                  31                  13
```

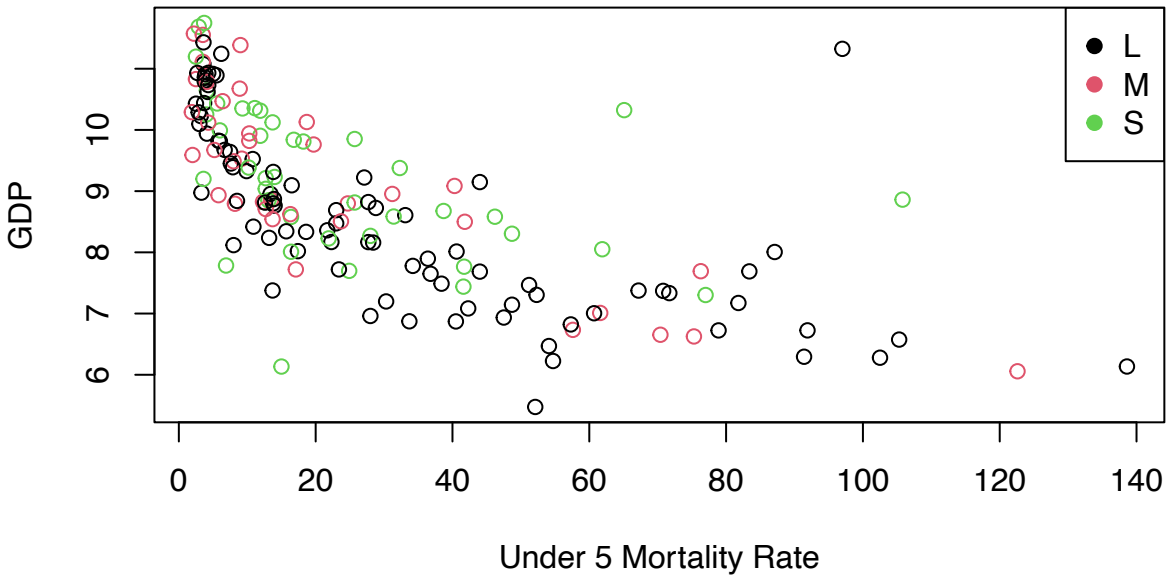


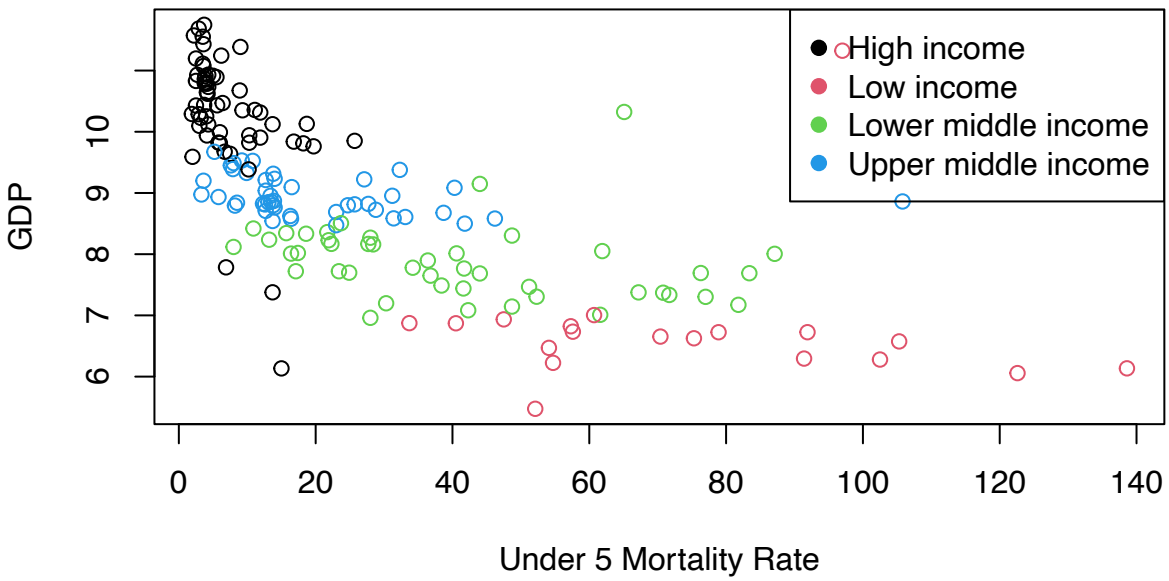

```
##
##     High income Low income Lower middle income Upper middle income
##   L          24         15                  30                  19
##   M          15          4                   4                  14
##   S          18          0                  10                  12
```

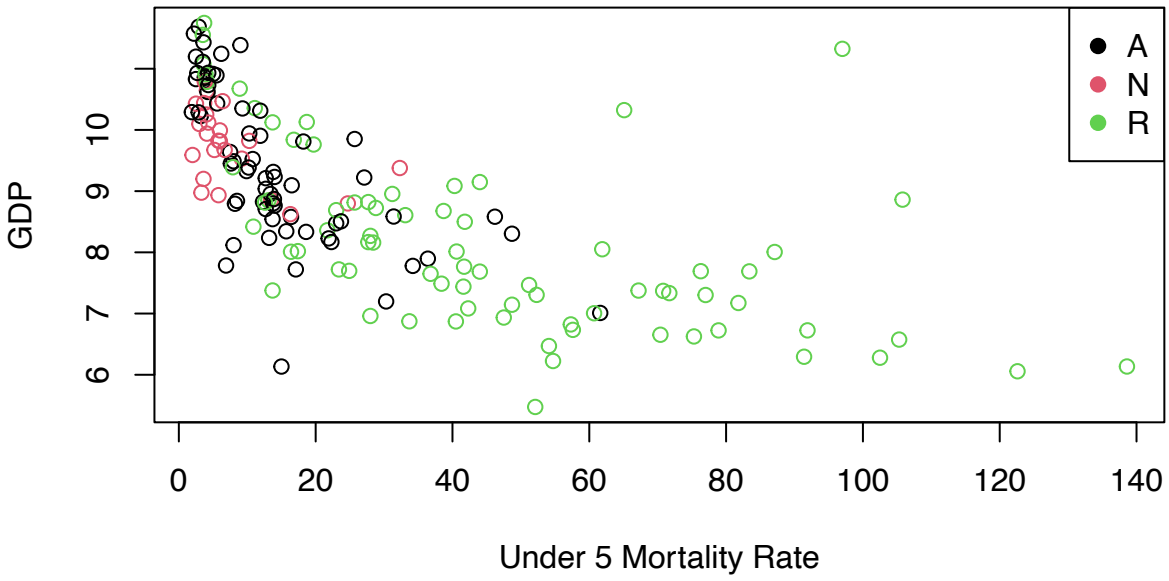

GDP
6
7
8
9
10
0.0e+00
4.0e+08
8.0e+08
1.2e+09
Total Population
L
M
S

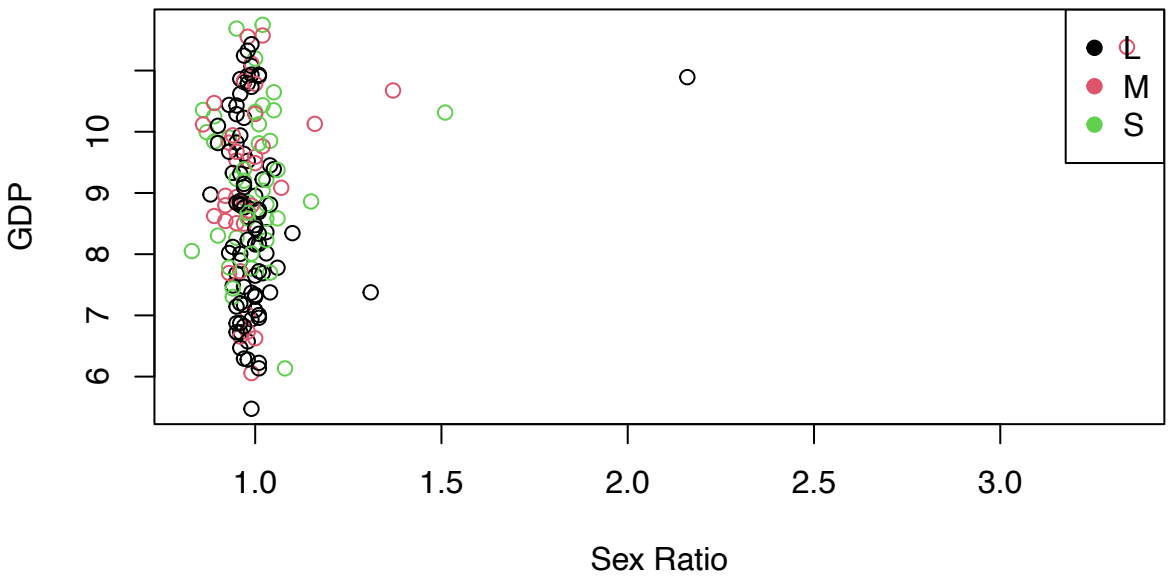

GDP
6
7
8
9
10
0.0e+00
4.0e+08
8.0e+08
1.2e+09
Total Population
High income
Low income
Lower middle income
Upper middle income

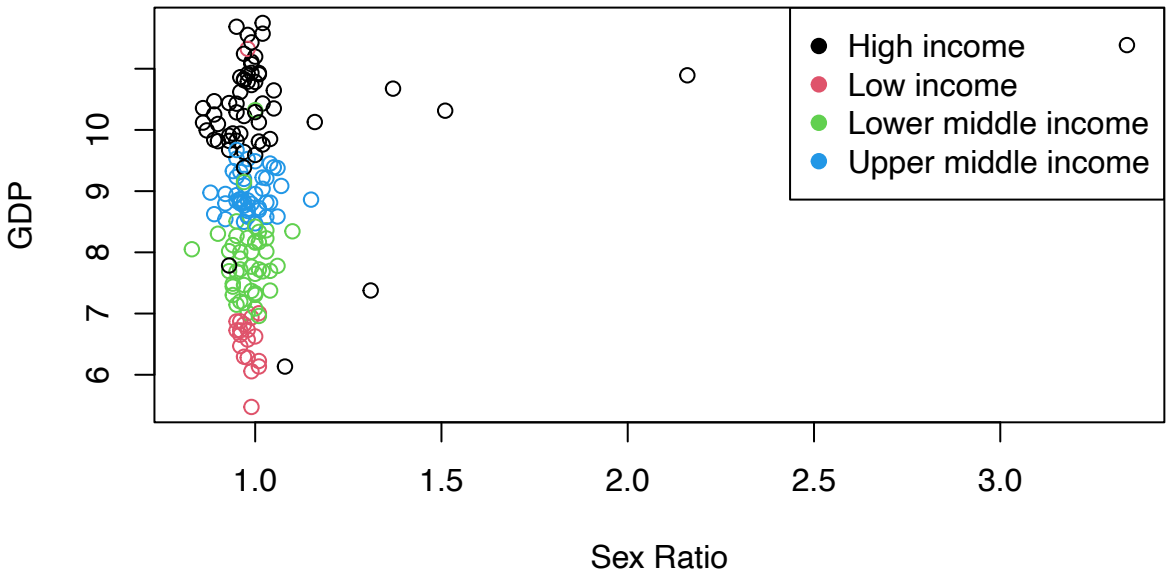

GDP
6
7
8
9
10
0.0e+00
4.0e+08
8.0e+08
1.2e+09
Total Population
A
N
R

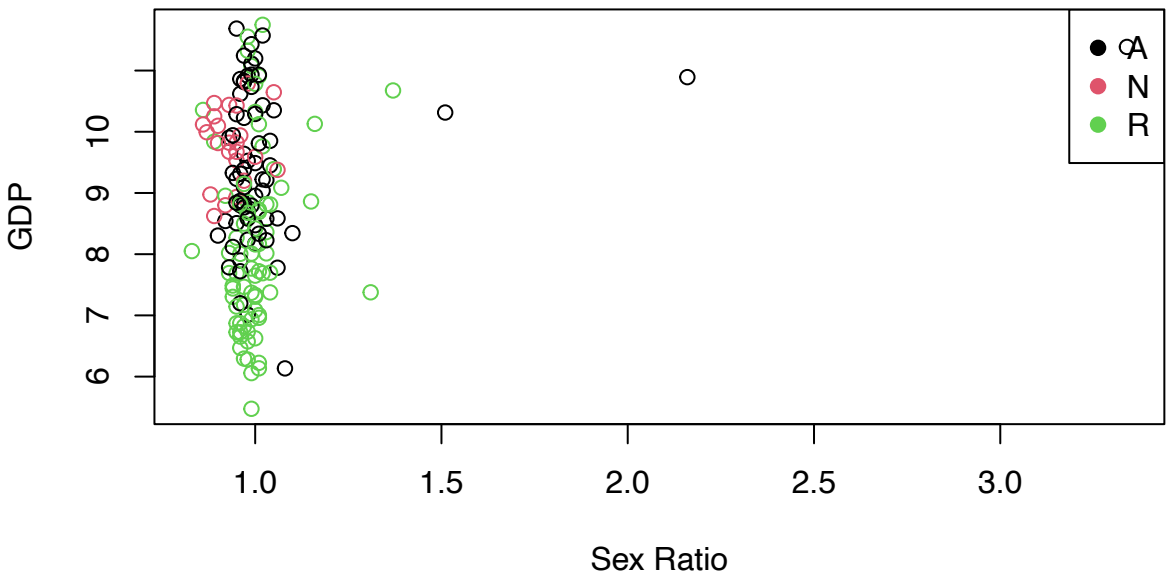

GDP
6
7
8
9
10
0
20
40
60
80
100
120
140
Under 5 Mortality Rate
L
M
S


GDP
6
7
8
9
10
0
20
40
60
80
100
120
140
Under 5 Mortality Rate
High income
Low income
Lower middle income
Upper middle income

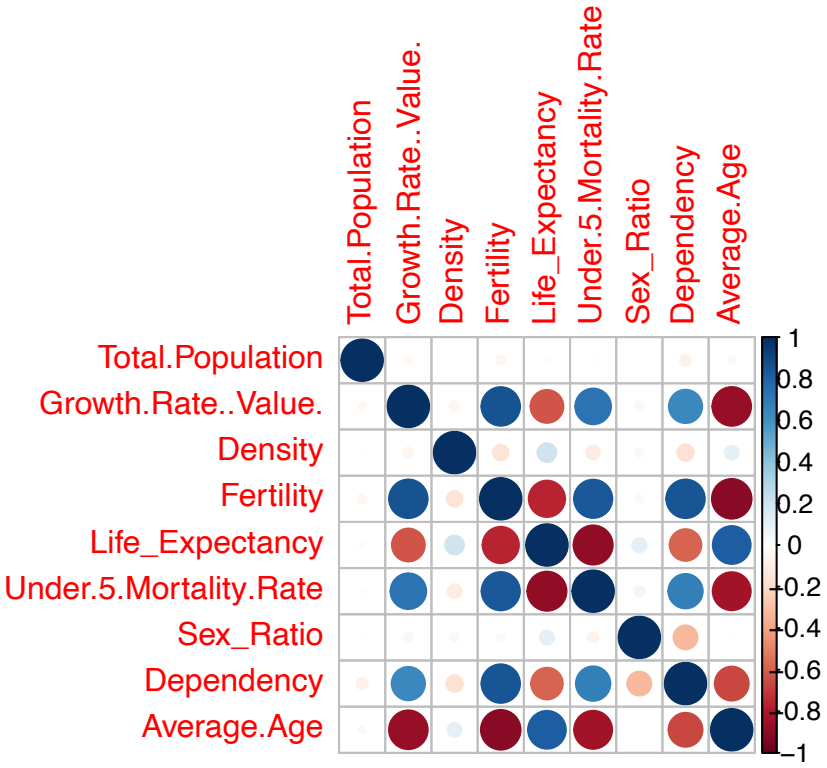

GDP
6
7
8
9
10
0
20
40
60
80
100
120
140
Under 5 Mortality Rate
A
N
R


GDP
6
7
8
9
10
1.0
1.5
2.0
2.5
3.0
Sex Ratio
L
M
S


GDP
6
7
8
9
10
1.0
1.5
2.0
2.5
3.0
Sex Ratio
High income
Low income
Lower middle income
Upper middle income


GDP
6
7
8
9
10
1.0
1.5
2.0
2.5
3.0
Sex Ratio
A
N
R

# Multicollinearity/Variable Screening




```
##
## Call:
## lm(formula = GDP ~ . - Name - Size - Growth - Income, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.9879 -0.4764 -0.0382  0.4332  4.0862
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)             -4.721e+00  2.668e+00  -1.769 0.078797 .
## Total.Population        -2.836e-10  4.413e-10  -0.643 0.521464
## Growth.Rate..Value.      1.761e-01  1.586e-01   1.110 0.268556
## Density                 -5.045e-06  1.059e-04  -0.048 0.962078
## Fertility               -7.422e-01  3.337e-01  -2.225 0.027569 *
## Life_Expectancy          7.808e-02  2.811e-02   2.778 0.006154 **
## Under.5.Mortality.Rate   3.649e-03  6.826e-03   0.535 0.593681
## Sex_Ratio                1.537e+00  4.388e-01   3.503 0.000602 ***
## Dependency               3.991e-02  1.552e-02   2.572 0.011067 *
## Average.Age              1.635e+00  8.069e-01   2.026 0.044451 *
## Debt                    -9.179e-04  2.423e-03  -0.379 0.705278
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8941 on 154 degrees of freedom
```

```
## Multiple R-squared:  0.642,  Adjusted R-squared:  0.6187
## F-statistic: 27.62 on 10 and 154 DF,  p-value: < 2.2e-16
```

```
##       Total.Population    Growth.Rate..Value.                Density
##                  1.020                  5.747                  1.127
##              Fertility        Life_Expectancy Under.5.Mortality.Rate
##                 29.577                  7.475                  7.463
##              Sex_Ratio             Dependency            Average.Age
##                  1.866                 10.413                 12.885
##                   Debt
##                  1.249
```

```
## [1] 7.8822
```

```
##
##                                 Stepwise Selection Summary
## ----------------------------------------------------------------------------------------
##                             Added/                   Adj.
## Step       Variable        Removed     R-Square    R-Square     C(p)        AIC        RMSE
## ----------------------------------------------------------------------------------------
##    1    Life_Expectancy    addition       0.560       0.557    28.2990    459.9667    0.9635
##    2      Average.Age      addition       0.608       0.603     9.5690    442.8290    0.9120
##    3       Sex_Ratio       addition       0.622       0.615     5.5970    438.8774    0.8985
##    4      Dependency       addition       0.629       0.619     4.7300    437.9425    0.8933
##    5       Fertility       addition       0.636       0.625     3.3950    436.4608    0.8867
## ----------------------------------------------------------------------------------------
```

```
##
##                               Selection Summary
## ------------------------------------------------------------------------------
##         Variable                        Adj.
## Step        Entered        R-Square    R-Square     C(p)        AIC        RMSE
## ------------------------------------------------------------------------------
##    1    Life_Expectancy      0.5599      0.5572    28.2994    459.9667    0.9635
##    2    Average.Age          0.6081      0.6033     9.5688    442.8290    0.9120
##    3    Sex_Ratio            0.6220      0.6150     5.5968    438.8774    0.8985
##    4    Dependency           0.6287      0.6194     4.7303    437.9425    0.8933
##    5    Fertility            0.6364      0.6250     3.3950    436.4608    0.8867
## ------------------------------------------------------------------------------
```

```
##
##
##                                 Elimination Summary
## ----------------------------------------------------------------------------------------
##         Variable                                Adj.
## Step           Removed             R-Square    R-Square     C(p)       AIC        RMSE
## ----------------------------------------------------------------------------------------
##    1    Density                       0.642      0.6212    9.0023    441.9169    0.8912
##    2    Debt                         0.6416      0.6233    7.1533    440.0786    0.8888
##    3    Under.5.Mortality.Rate       0.6411      0.6251    5.3859    438.3274    0.8866
##    4    Total.Population              0.640      0.6264    3.8475    436.8200    0.8851
##    5    Growth.Rate..Value.          0.6364       0.625    3.3950    436.4608    0.8867
## ----------------------------------------------------------------------------------------
```

```
##
## Call:
## lm(formula = GDP ~ . - Name - Size - Growth - Income - Fertility -
##     Density - Growth.Rate..Value. - Total.Population - Under.5.Mortality.Rate -
##     Debt - Dependency, data = data)
##
## Coefficients:
##     (Intercept)  Life_Expectancy        Sex_Ratio      Average.Age
##        -4.71826          0.08205          0.80012          1.97195
```

```
##
## Call:
## lm(formula = GDP ~ . - Name - Size - Growth - Income - Fertility -
##     Density - Growth.Rate..Value. - Total.Population - Under.5.Mortality.Rate -
##     Debt - Dependency, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4640 -0.4363 -0.0852  0.4354  4.5822
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)     -4.71826    0.85327  -5.530 1.27e-07 ***
## Life_Expectancy  0.08205    0.01871   4.386 2.08e-05 ***
## Sex_Ratio        0.80012    0.32903   2.432   0.0161 *
## Average.Age      1.97195    0.40628   4.854 2.84e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8985 on 161 degrees of freedom
## Multiple R-squared:  0.622,  Adjusted R-squared:  0.615
## F-statistic: 88.31 on 3 and 161 DF,  p-value: < 2.2e-16
```

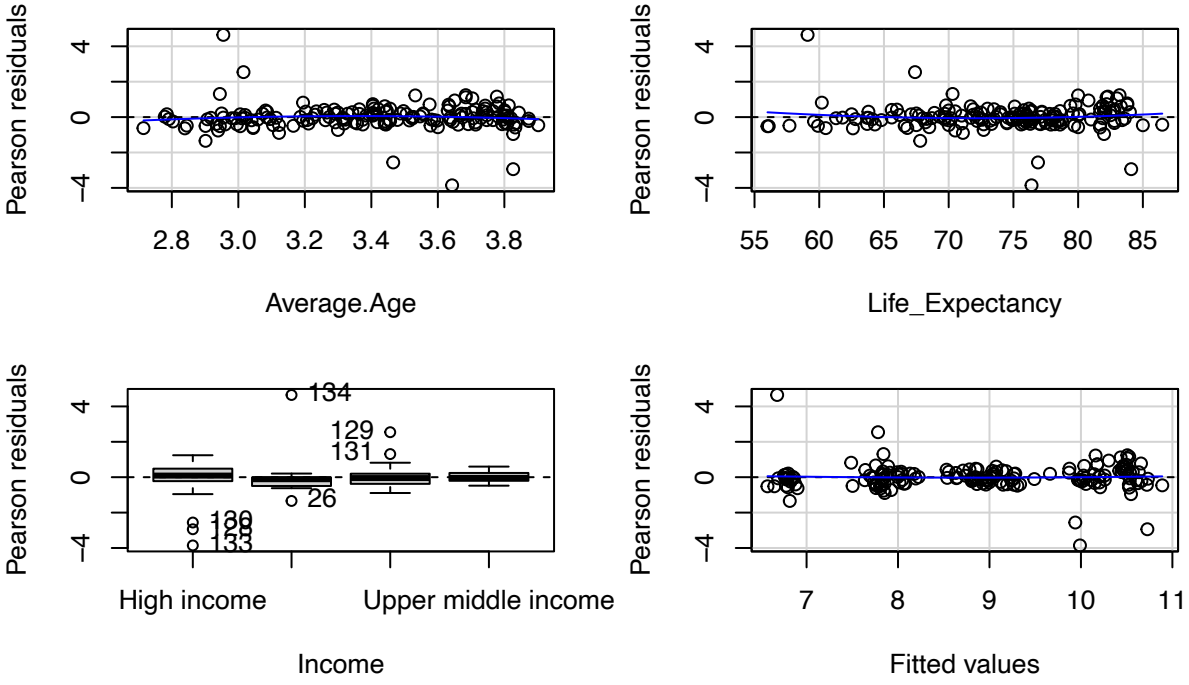
```
## Life_Expectancy       Sex_Ratio     Average.Age
##           3.278           1.039           3.235
```

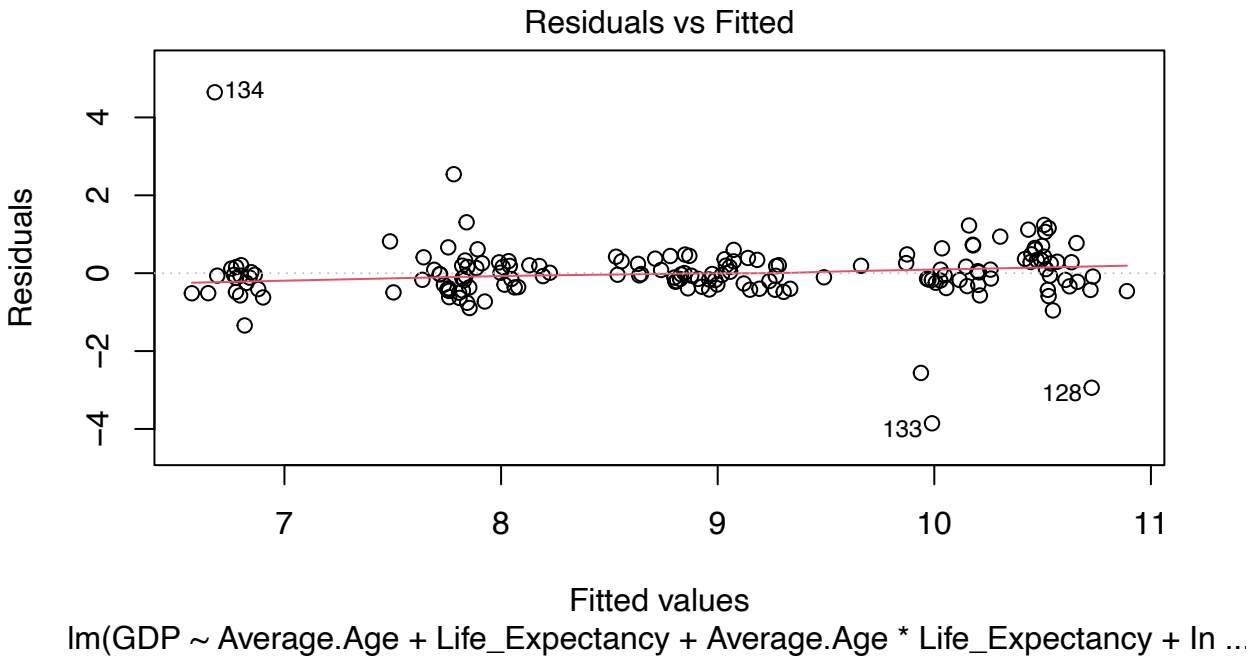
```
## [1] 2.517333
```

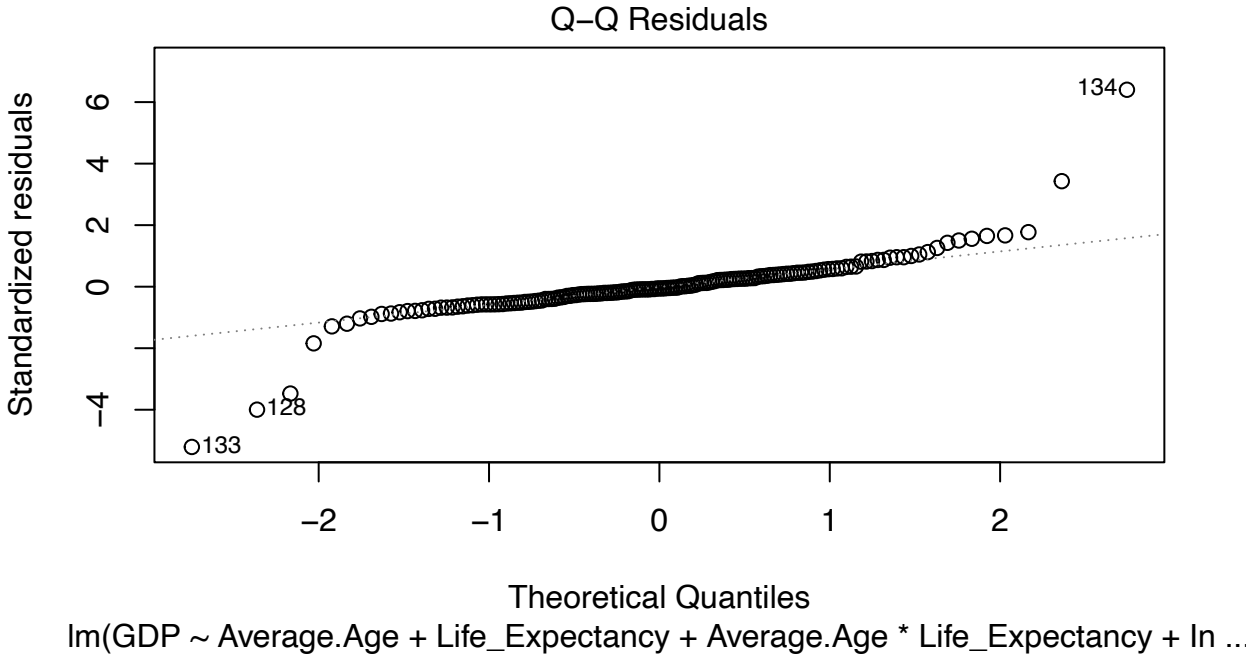
# Pre-Residual Removal Model

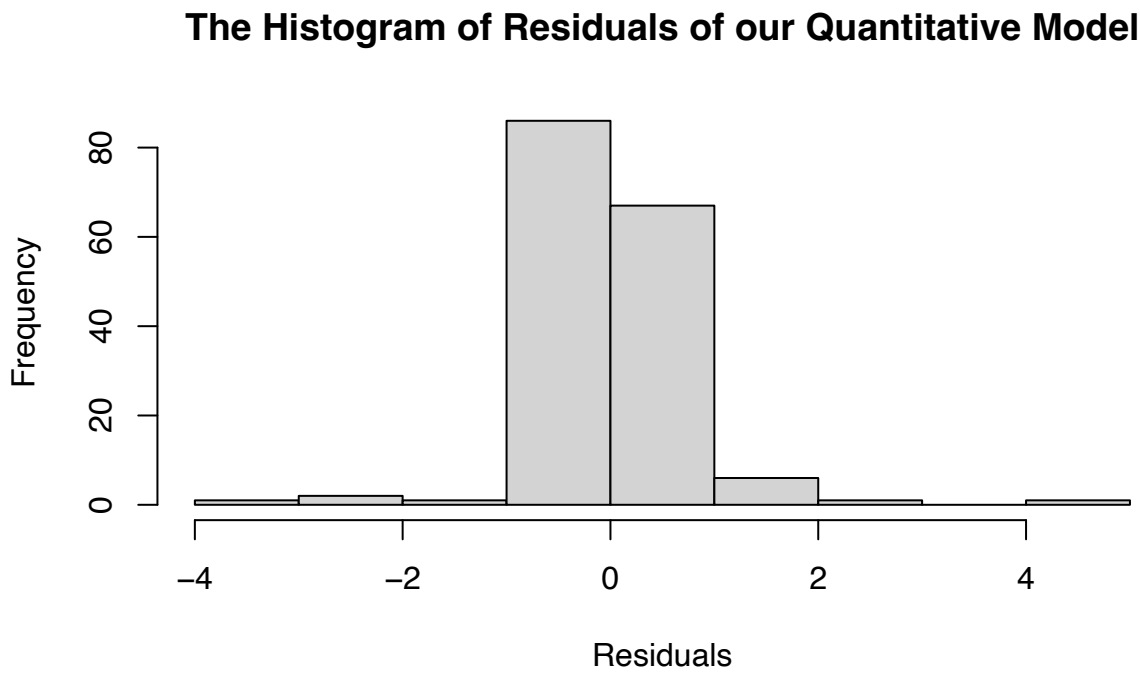
```
##
## Call:
## lm(formula = GDP ~ Average.Age + Life_Expectancy + Average.Age *
##     Life_Expectancy + Income, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.8552 -0.2987 -0.0421  0.2809  4.6460
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   25.59939    9.27481   2.760  0.00646 **
## Average.Age                   -5.75669    2.87551  -2.002  0.04700 *
## Life_Expectancy               -0.22559    0.12645  -1.784  0.07634 .
## IncomeLow income              -2.75660    0.34318  -8.033 2.07e-13 ***
## IncomeLower middle income     -1.75484    0.24629  -7.125 3.48e-11 ***
## IncomeUpper middle income     -0.87003    0.19900  -4.372 2.22e-05 ***
## Average.Age:Life_Expectancy    0.08119    0.03847   2.110  0.03640 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7528 on 158 degrees of freedom
## Multiple R-squared:  0.7396, Adjusted R-squared:  0.7297
## F-statistic: 74.79 on 6 and 158 DF,  p-value: < 2.2e-16
```

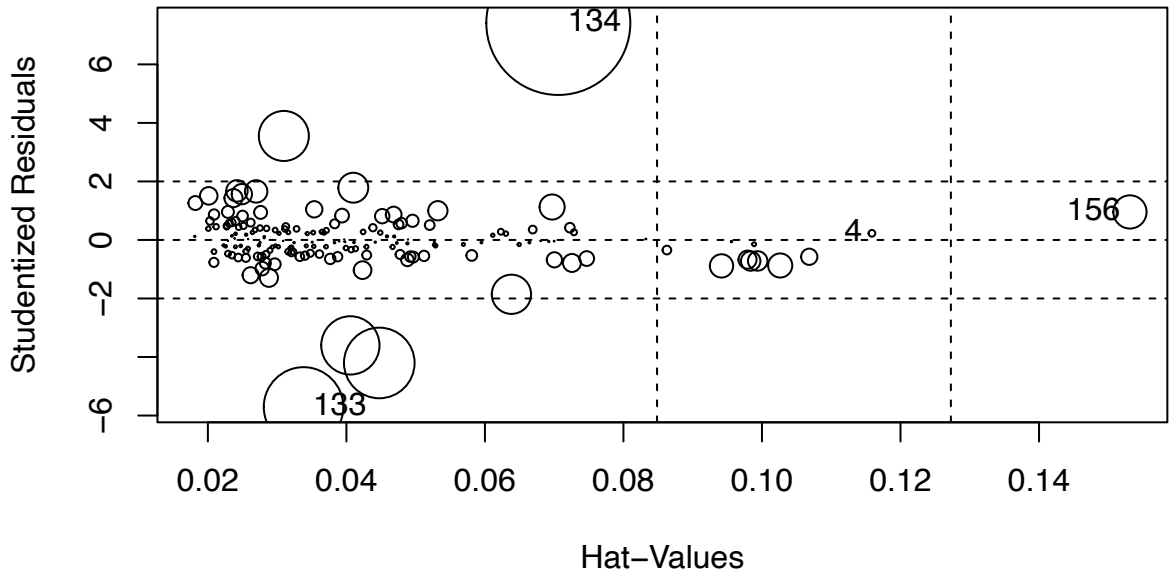
# Assumptions/Residual Analysis/Influential Observation Analysis




Residuals vs Fitted
Residuals
Fitted values
lm(GDP ~ Average.Age + Life_Expectancy + Average.Age * Life_Expectancy + In ...
134
133
128
7
8
9
10
11
−4
−2
0
2
4

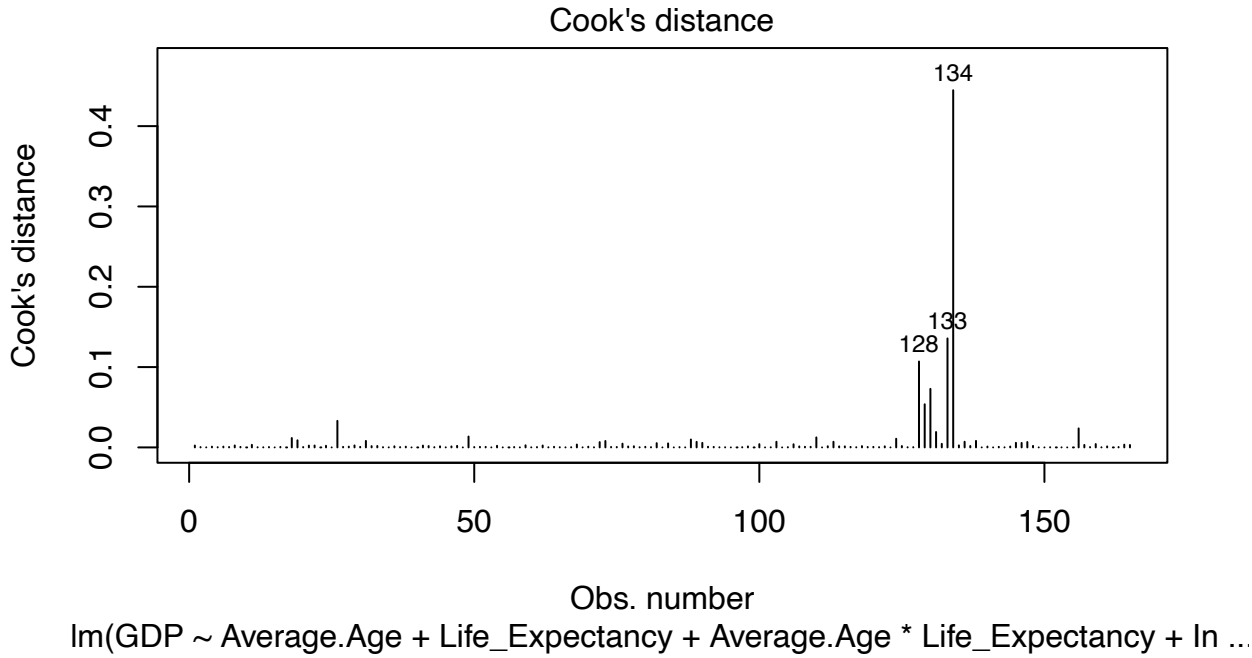

Q−Q Residuals
Standardized residuals
Theoretical Quantiles
lm(GDP ~ Average.Age + Life_Expectancy + Average.Age * Life_Expectancy + In ...
134
128
133
−2
−1
0
1
2
−4
0
2
4
6



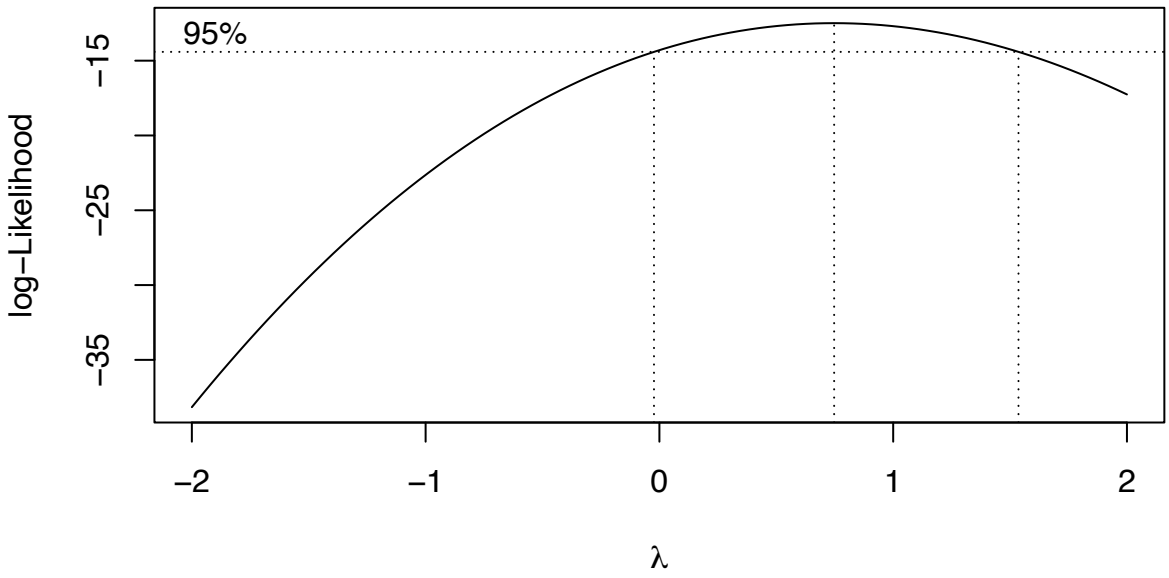

```
##         StudRes        Hat        CookD
## 4     0.2270134 0.11586985 0.0009706756
## 133 -5.7069476 0.03381706 0.1357298668
```

```
## 134  7.4152335 0.07059705 0.4447179793
## 156  0.9570428 0.15312495 0.0236713177
```




```
##           4         128         133         134         156
##  0.08218236 -0.90978337 -1.06768253  2.04369504  0.40695339

## [1]  0.06342613 -0.57164490  0.52941454  0.47810477 -0.19417639

## [1] -0.00269222
```

# Box Cox Plot




```
## [1] 0.3943296
```

# Confidence and Prediction Intervals

```
# We can therefore, employ this new and final model for estimation and
## prediction using the metrics below.

# E(y) confidence interval
conf_interval <- predict(finalModel, interval = "confidence", level = 0.95,
                         newdata = test_data)
conf_interval
```

```
          fit       lwr       upr
2    8.223635  8.002709  8.444562
3   10.833050 10.599054 11.067046
10   8.974368  8.818028  9.130709
15   8.972765  8.759713  9.185816
18   7.727090  7.385143  8.069037
20   7.913944  7.741419  8.086469
28   7.826610  7.644841  8.008378
29   7.684491  7.419507  7.949475
33  10.352846 10.201828 10.503864
45   8.951398  8.771367  9.131429
55   6.585890  6.314379  6.857402
56   8.781440  8.569424  8.993456
57  10.631667 10.449865 10.813469
58   7.821522  7.647344  7.995700
59  10.625109 10.447719 10.802499
66   7.886086  7.718331  8.053841
70   8.860997  8.702888  9.019107
73  10.395311 10.152000 10.638622
80   7.678842  7.485353  7.872331
85   6.377456  6.116422  6.638490
88  10.658407 10.493546 10.823268
96   7.652781  7.464847  7.840716
98   8.873747  8.715846  9.031648
100  7.851593  7.632793  8.070393
105  9.559761  9.230475  9.889048
```

```
108 10.580063 10.420816 10.739309
116 10.171166  9.960258 10.382074
119  8.586129  8.385028  8.787229
124 10.319898 10.148162 10.491634
126  6.512186  6.249504  6.774868
130 10.080140  9.869506 10.290774
134  6.334268  6.049178  6.619358
145 10.784487 10.584814 10.984159
```

```
confint(finalModel)
```

```
                                  2.5 %       97.5 %
(Intercept)                  9.70065800 37.0707114302
Average.Age                 -9.57727572 -1.0708742289
Life_Expectancy             -0.36891085 -0.0006319617
IncomeLow income            -3.56869768 -2.5950085314
IncomeLower middle income   -2.22729666 -1.5397335200
IncomeUpper middle income   -1.25379182 -0.7026959425
Average.Age:Life_Expectancy  0.01638531  0.1288613515
```

# Final Model and Summary

```
##
## Call:
## lm(formula = GDP ~ Average.Age + Life_Expectancy + Average.Age *
##     Life_Expectancy + Income, data = subdata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.70317 -0.30829 -0.02124  0.24556  2.56927
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                  23.38568    6.92706   3.376 0.000932 ***
## Average.Age                  -5.32407    2.15288  -2.473 0.014492 *
## Life_Expectancy              -0.18477    0.09321  -1.982 0.049229 *
## IncomeLow income             -3.08185    0.24643 -12.506  < 2e-16 ***
## IncomeLower middle income    -1.88352    0.17401 -10.824  < 2e-16 ***
## IncomeUpper middle income    -0.97824    0.13948  -7.014 7.01e-11 ***
## Average.Age:Life_Expectancy   0.07262    0.02847   2.551 0.011717 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5212 on 153 degrees of freedom
## Multiple R-squared:  0.8733, Adjusted R-squared:  0.8684
## F-statistic: 175.8 on 6 and 153 DF,  p-value: < 2.2e-16

## [1] 0.2716175

```

## Appendix D: References


About Us. (2023). Focus Economics. https://www.focus-economics.com

Andrew Chatzky. (2021, May 11). Why Does the Census Matter? Council on Foreign Relations. https://www.cfr.org/backgrounder/why-does-census-matter

Country Indicators. (n.d.). [dataset]. https://www.focus-economics.com/country-indicator/

International Database. (2023, July 12). [dataset]. United States Census Bureau. https://www.census.gov/programs-surveys/international-programs/about/idb.html#:~:text=Methodology,one%20of%20our%20primary%20resources.

GDP per capita. (2021). [Map]. Accessed November 28, 2023. Our World in Data. https://ourworldindata.org/grapher/gdp-per-capita-worldbank

How to Forecast GDP in the Long Run: A Primer. (2022, October 20). The Conference Board. https://www.conference-board.org/topics/recession/how-to-forecast-GDP-in-the-long-run#:~:text=Longer%2Dterm%20forecasts%20of%20GDP,buildings%2C%20machinery%2C%20or%20software.

Kunal Sen. (2021, January). Why should I care about economic growth? UNU-WIDER. https://www.wider.unu.edu/publication/why-should-i-care-about-economic-growth#:~:text=Growth%20creates%20wealth%2C%20some%20of,and%20gain%20improved%20living%20standards.

Linda Jacobsen. (2023, February 10). How Accurate Was the 2020 Census—And Why Should You Care? Population Reference Bureau. https://www.prb.org/resources/how-accurate-was-the-2020-census-and-why-should-you-care/#:~:text=The%20PES%20estimated%20a%20net,2010%20to%205.8%25%20in%202020.

Matthias Doepke. (2022, July 22). THE NEW ECONOMICS OF FERTILITY. International Monetary Fund. https://www.imf.org/en/Publications/fandd/issues/Series/Analytical-Series/new-economics-of-fertility-doepke-hannusch-kindermann-tertilt#:~:text=It%20suggests%20that%20as%20parents,across%20countries%20and%20over%20time.

Our Censuses. (2022, August 4). United States Census Bureau. https://www.census.gov/programs-surveys/censuses.html#:~:text=It%20helps%20the%20government%20decide,each%20State%20holds%20in%20Congress.&text=The%20U.S.%20census%20counts%20every%20resident%20in%20the%20United%20States.

Population and Economic Growth. (n.d.). Vaia. Retrieved October 19, 2023,

from https://www.hellovaia.com/explanations/macroeconomics/economic-performance/population-and-economic-growth/#:~:text=For%20example%2C%20an%20increase%20in,the%20country%20as%20a%20result!

Top Questions About the Survey. (2023, June 27). United States Census Bureau. https://www.census.gov/programs-surveys/acs/about/top-questions-about-the-survey.html