VISUAL AND STATISTICAL ANALYSIS ON HOMICIDE BETWEEN EUROPE AND LATIN AMERICAN AND CARRIBEAN

ARABA NGOZI BLESSING

@00675862

2022

Contents

Interactive Dashboard Design

I. Introduction

Top leading countries in relation to intentional homicide rates are mostly from countries in the Latin American and Caribbean region. Occurrence of homicide is exceptionally outrageous in El Salvador, Honduras, Jamaica, and Guatemala. Homicide rates in developed countries, say Europe, seems less compared to the world average rate, thereby proposing that variability in development indicators may describe the distinctness in homicide rates (Loureiro et al, 2012). I will focus on using series of visualization tools such as map, charts to aid the identification of the frequency of homicide rate in some selected countries of Europe and Latin America and Caribbean region. In contribution to the literatures on intentional homicide rates across countries, (ADELEYE, 2014), explored the possible factors affecting income inequality and in relations to crime. Using a similar approach, as in (ADELEYE, 2014) with the use of less development indicators and less panel of twelve countries in two geographical regions, we discover through the visuals that variability in some country's development indicator might infuse high effect in its homicide rate, while some indicators have mixed effect on homicide rate. The visualization analysis of these indicators in relation to homicide rate was accomplished with the use of Power Bi Desktop.

II. Background Studies

An interactive dashboard allows one to drill down and filter operational information for a clear view of data from different perspectives or in more detail.

Interactive dashboards enhance observation and interaction of data with the use of visualisations tools like maps, charts, tables or even text commonly on a lone visualization board (Alhamadi ,2020).

Even though dashboards are commonly used in many domains, (Alhamadi, 2020), identified and investigated dashboard related problems, insights on how to properly tackle the problems, also the modifications that can be used to lessen these problems. He aimed to investigate and assess range of adaptation methods which will enhance easy use of interactive dashboards

Ali et al. (2016), discussed immensely the expected associated big data visualization challenges and reviewed certain big data visualization tools.

To address the issue of the types of weapons used and the crime location proximity to the nearest police station, (Chui et al., n.d.), developed an interactive visual analytical dashboard on Baltimore Homicide Statistics which differs from other existing geospatial visualizations of crimes in US cities that mainly describes the cities with highest crime rates.

Diamond et al., (2021), identified the problems associated with collection, analysis, representation of various analytic data. They applied a dashboard interface for the improvement of user experience (UX), just how visualizations are considered as data cache, dashboard interface is also considered as visualization cache. They examined degree of activity and user engagement required for the data exploration which arises to the utilization of the interactive dashboard.

O'Brien, (2022), visualized differences in homicide age distributions in the United States, aimed to demonstrate understanding crime rate data with the help of visualization.

Silver et al. (2020), modelled a spatial analysis of homicides in the state of Pernambuco, Brazil, between the years of 2016 and 2019, by making us of a comprehensive analysed spatial homicide data with five variables which may possibly describe the incidence of homicide.

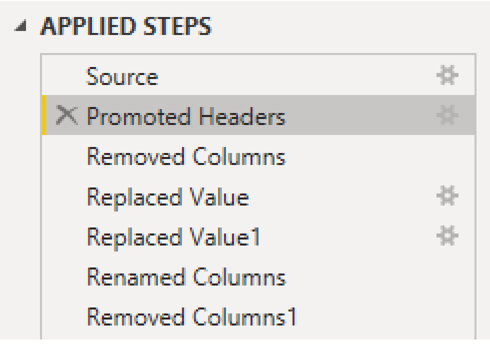
Sarikaya et al. (2019), identified the wide range of how dashboards can be used through exploring different examples of dashboards and documentation about their use. identified major types of dashboards and characterized dashboards by their design goals, levels of interaction and the practices around them.

Stamatel (2014), elaborated on the improvement of data visualization towards analytical reasoning in cross-national crime research with the exploration of spatial visualization and temporal visualization.

Walny et al. (2020), emphasized on how visual analytics research continues to look for approaches to illustrate inconsistencies in data. Identified potentials in prospective techniques for prototyping, testing, and communicating data-driven designs.

III. Dataset

Publicised Data are sometimes not appropriate for research or analysis due to the underreporting of the indicators. (ADELEYE, 2014). Due to the issues of missing values, twelve countries across Europe and Latin America were selected with missing values space filled with zeros. The compiled dataset which was gotten from www.worldbank.org World Development Indicators, initially consisted of sixteen attributes scaled down to twelve attributes in the success of reducing too much noise.



These attributes are described as

Gross Domestic Product Per Capital (GDP_PC); measures the country's economic output per person.

Gross Domestic Product Per Capital Growth Rate (GDP_PCG); shows the annual growth rate of the country's economy which is expected that as the growth rate increases, income inequality reduces then crime rates (in this case homicide rates) also reduces.

Gini Index: is a measure which is expected to represent the income inequality or the wealth inequality within a country.

Intentional Homicide Rate Per 100000 Inhabitants (Homicide): illegal death that is deliberately inflicted on an individual by another arising from disagreements at home, social disagreements, disagreements over inheritance and properties, and fights over power control between groups of gangs.

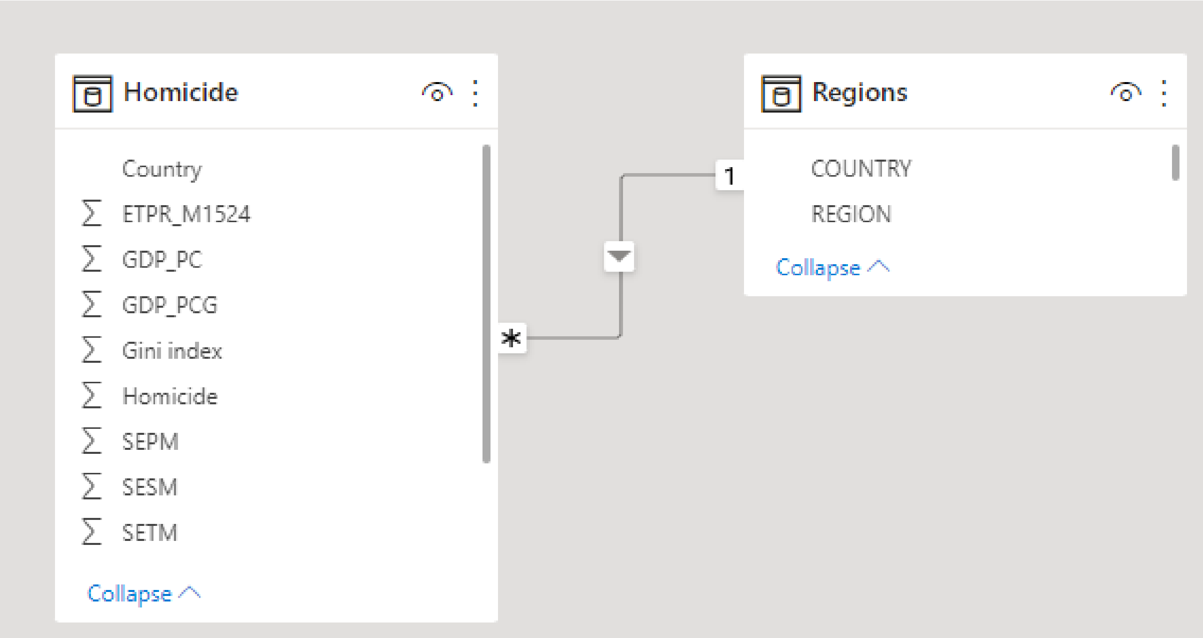
Employment to population ratio of male aged 15-24 years, Male enrolment rates for primary (SEPM), secondary (SESM) and tertiary (SETM) education: Male youths, by assumption, are susceptible to get involved with unlawful occurrences compared to other groups in the society, leading to increase in the involvement of unlawful offences in the adolescent age. A high ratio on employment shows the sizeable, employed proportion of youths while a low ratio might be seen as a positive sign, especially for young people, if caused by an increase in education.

Unemployment in male population (UNEM_M): unemployment, based on believe, is positively associated to wealth inequality and occurrence of homicide. A significant increase in unemployment, further widens income inequality which extends to increase in homicide rate.

Total Enrolled: This is a calculated column showing the total number of males enrolled for education using this DAX expression

```
Total Enrolled = Homicide[SEPM] + Homicide[SESM] + Homicide[SETM]
```

A different table grouping countries to respective regions was created. Here we dragged the Country field from the homicide table to the Country field in the Region table to join the tables and create a relationship between them.



IV. Data Workflows & Proposal for Design of Dashboard

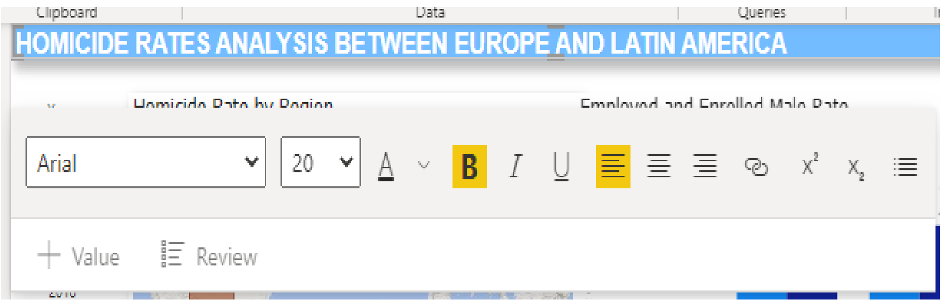
Selection of appropriate data visualizations can be challenging for beginners. Major hurdles faced by information visualization beginners were interpreting questions into data features, projection of visual layouts, and interpretation of the visualizations (Grammel, Tory, and Storey, 2010). The process of choosing which visuals to use begins from the types of attributes in the selected data and the magnitude of the dataset.

		as Dimension			as Measure	
		Sort order	Use e.g.	Do not use	Use e.g.	Do not use
Nominal		Alphabetical or by Measure	Bar chart, Pie chart, Scatter chart	Line chart	Count()	Average(), Median(), Sum()
Ordinal		by Intrinsic order	Bar chart, Line chart, Scatter chart	Pie chart	Count(), Median()	Average(), Sum()
Quantitative	Interval/ Coordinate	by Number	Bar chart, Line chart	Pie chart	Count(), Average(), Median()	Sum()
	Ratio/ Amount	by Number	Bar chart, Line chart, Histogram	Pie chart	Count(), Average(), Median(), Sum()	

Table 1.

Although there could be exceptions, community.qlik.com. (2014), summarized rules for visualization that could be connected to scale of measurement in *Table 1*.

These rules helped serve as guide while I was visualizing the analysis on homicide rates in Europe and Latin American and Caribbean countries. Insights on the different visualization types was elaborated by Mihart n.d. (2022).

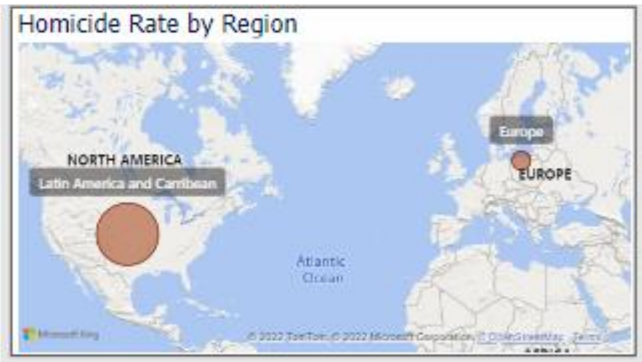


Visual 1.

Visual 1 is the title *Homicide Rate Analysis Between Europe and Latin America* this was achieved by selecting Text Box on the Insert ribbon, typed in uppercase letters, set the font size to 20 and Bold. Set background colour to blue, transparency at 41% and added the shadow effect.

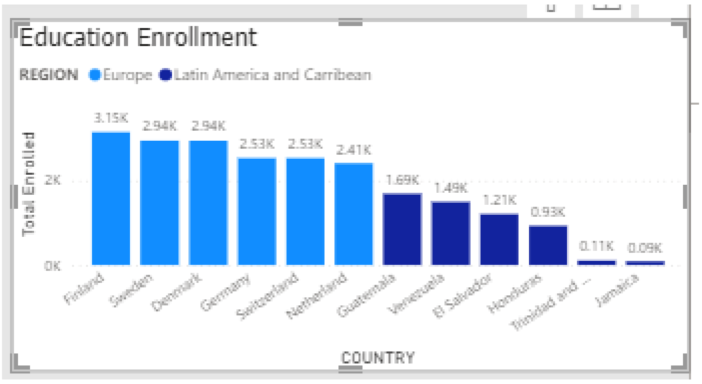
Visual 2.

Visual 2. Is a date table showing the time series used in the analysis in descending order. A selection on an individual year shows the analysis on other visuals.



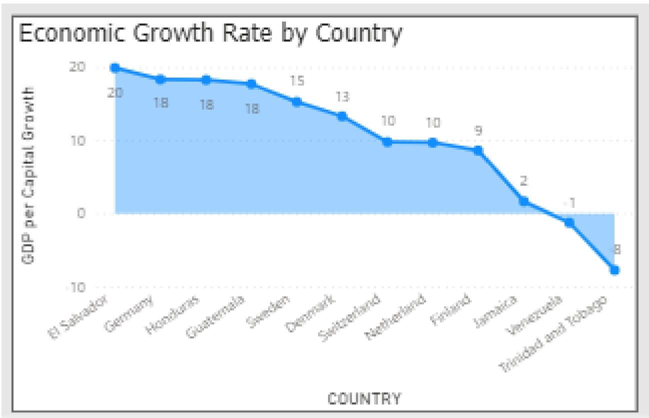
Visual 3.

Maps can be utilized when representing categorical and quantitative information with geographical positions (Mihart n.d.). Charting regions of the occurrence of homicide on the map a clear view of the specific region with high or low homicide rate. Visual 3 presents the geospatial map of the world with the Regions that are included in the analysis of homicide rate. The geospatial map was used to show the region-wise homicide rates. The size of the bubbles shows how high or low the homicide rate is in each region. The big bubble indicates high rates of homicide in the Latin American and Caribbean region while the small bubble indicates low rates of homicide in the European region.



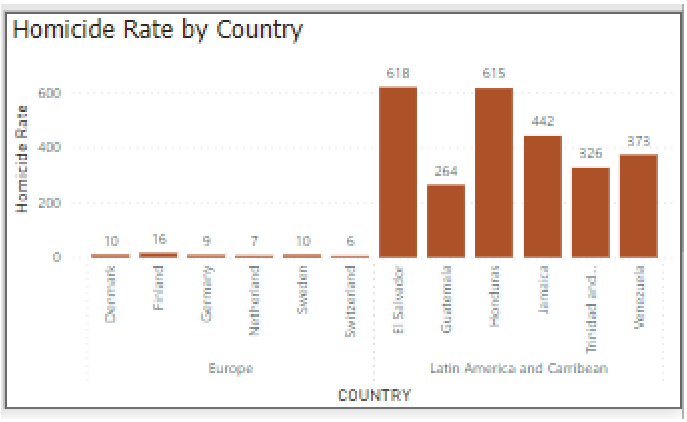
Visual 4.

Visual 4 is the stacked column chart showing the estimated level of enrolled male population into primary, secondary, and tertiary education in each country while encoded regionally by the different shades of blue. Also indicates the estimated value for each country. (Talagala & Shashikala, n.d.) used bar chart to look out certain values across various categories.



Visual 5.

Area charts highlights the rate of change over time. Visual 5 is an area chart which shows the variation between gross domestic product per capital growth rate and country. The area chart has markers indicating the data labels (growth rate) of each country. The growth rate area chart shows El Salvador having significant high growth rate as compared to Trinidad and Tobago whose growth rate is quite low.



Visual 6.

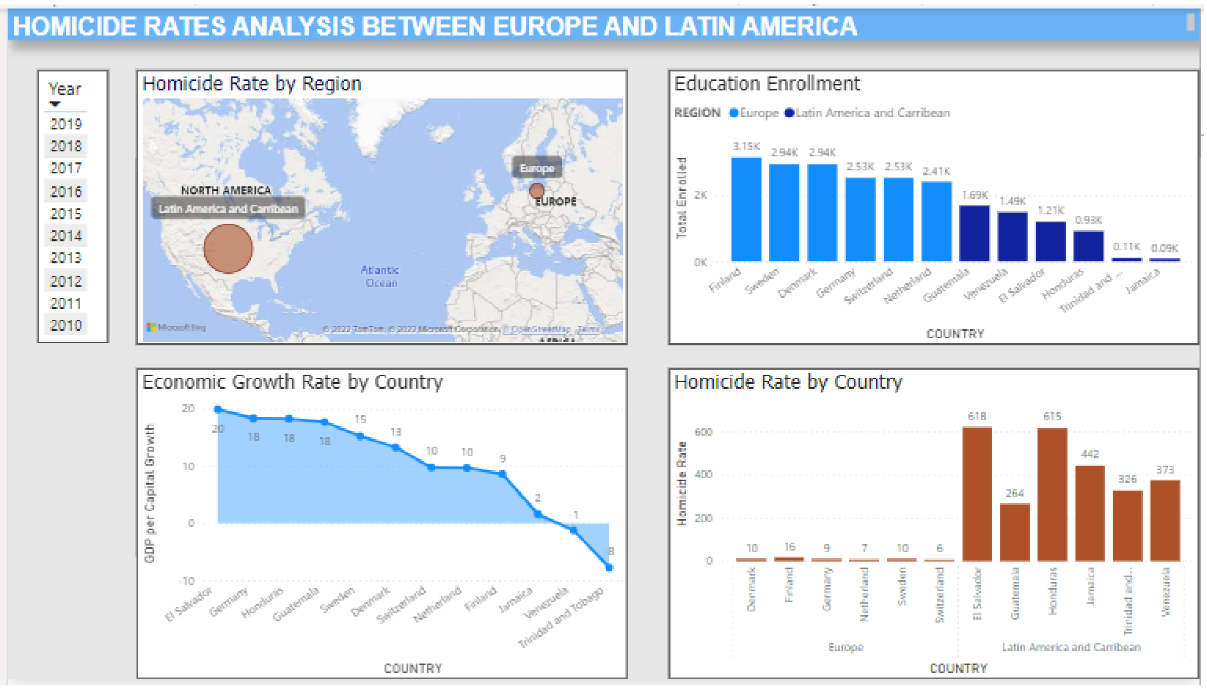
Visual 6 is the homicide rate clustered column chart by country which allows for region segmentation. Each bar shows the frequency of all the selected categorical features, also significantly shows the high rate of homicide in El Salvador and extreme low rate in the European region.

Since there is a linkage between charts and map on the dashboard, a random pick from any of the visual would automatically display a corresponding information on the other visuals. With this chart linking set up, the year table can serve as a supportive mechanism for the interactive visualization dashboard for the filtering of annual information on individual countries/region (Chui et al., n.d.).

V. Discussion

The different recent systematic software involves versatile, presentation of visualization in form of dashboard that views different levels of information at a glimpse (Chui et al., n.d.).

Hence, I have designed the visualization dashboard in an easy way of accessing the information of interest in a single view. For easy clear view of the variability in the country's development indicator (in this case gross domestic product per capital growth rate and estimated total number of male students) in relation to homicide rate by country, their respective charts are close together as seen in visual 7. Evaluating through each visualization, we can view from the map that there quite high rate in homicide from Latin America and Caribbean Region with about 98% in total compared to the European region (in respect to the dataset used for this visualization analysis).



Visual 7

The educational enrolment of male gender in total in Latin America and Caribbean region is undoubtedly poor. Normally, a higher educational procurement is likely to reduce wealth inequality, although might have a mixed effect on homicide rate (ADELEYE, 2014). Visual 3 does not show otherwise in this case, even though countries with less educational attainment have high rate in homicide, that is the Latin America and Caribbean region. Finland has the highest educational attainment, they still have high rate of homicide among the European regions, Guatemala with high rate in educational attainment has less homicide rate among the Latin American and Caribbean regions.

Gross domestic product per capital growth rate is an approximate for the level of economic prosperity of countries, high growth rate should indicate less wealth inequality also should have significant change in homicide rate. Viewing closely at the visualization board, I can see El Salvador having high growth rate in total but still have high homicide rate and low educational procurement. Guatemala could be understandable if they have such growth rate and still have low but not so low homicide rate. Meanwhile the Europe region are generally having high gross domestic product per capital, with low homicide rate in general as expected

VI. Conclusions

Studies from different literatures proposed the statistics of intentional homicide rates might be influenced by differences towards trauma care, precipitating different lethality of violent assaults, hence, intentional homicide rate must not automatically stipulate the overall degree of crime in the economy (Wikipedia Contributors, 2019). With the efficient use of the visualization dashboard, I was able to successfully visualize the variability in the development indicator used in relation to the homicide rate of the selected country. Which showed that the effect of these indicators on homicide rate in each individual country will depend on the policy of the country as its more of a mixed effect as we could see El Salvador having more homicide rate despite the high growth rate and fair education attainment.

Statistical Analysis

1. Introduction

Homicide has long been one of the human slayers globally, it is one of the cause of deaths in some countries also still being recorded as the leading cause of death in some countries like Venezuela, fifth leading cause of death in Guatemala and fourth in Honduras as of 2017. Across most countries of Europe, deaths by homicide is less than 0.5% as compared to Latin America countries like El Salvador more than 7%, Guatemala with more than 6% and 5% from Venezuela as of 2019 (Roser and Ritchie, 2013). In this study I aim to use variety of statistical analysis techniques such as correlation analysis, regression analysis, hypothesis testing, ANOVA, and time series analysis to analyse the different trends/patterns associated with homicide rate across some selected countries in Europe and Latin America and Caribbean.

2. Background Studies

Various research works has been developed on homicide rate analysis since the last half of the 20th century. Sociologist and criminologist were the pioneer researchers of this theme. The principal objectives for the research works were related to analysing if demographic, economic, ecological, and social variables maintained some correlations to the variation in homicide rates across time and space (Land *et al* 1990).

Using an international panel data set to study the determinants of intentional homicide, Loureiro (2012), demonstrated that intentional homicide is highly inert, counter-cyclical, declines with development and rises with violence and income inequality.

Trends and patterns can be predicted by the analysis of the time-series data based on past events.

Carter and Mahoney (2022) utilized a repeated time-series design to assess state-level behavioural changes from 1994–2003 on firearm homicide and suicide rates pre-, during, and post-enforcement of the federal regulation (i.e., The Brady Handgun Violence Prevention Act).

Machado Filho, da Silva and Zebende (2014), established the relationship between homicides and attempted homicides by a non-stationary time-series analysis, they identified a positive cross-correlation between homicides and attempted homicides with the use of Detrended Fluctuation Analysis and Detrended Cross-Correlation Analysis.

Vilalta and Muggah (2014), considered the length of correlation between certain demographic and socioeconomic factors and homicidal violence in the context of Mexico's "war" on organized crime and proposed a baseline model for testing spatial-temporal dynamics of organized violence in multiple settings.

Mohammadi et al. (2022), explored homicide patterns and their spatial associations with different socioeconomic and built-environment conditions in 140 neighbourhoods of Toronto, Canada. They analysed the neighbourhood-based indicator using spatial techniques such as Kernel Density Estimation, Global/Local Moran's *I* and Kulldorff's SatScan spatio-temporal methodology. Analysis of the spatially varying correlations between the homicide rate and independent variables were explored with Geographically weighted regression (GWR) and multi-scale GWR (MGWR).

3. Dataset

Statistics is not only concerned with just numbers; it is concerned with making the most use out of data of all kinds also with trying to find hidden patterns or trends in data. Most of the standard techniques used in statistics requires prior knowledge about the data attributes (Maltz, Gordon and Friedman, 2018).

Homicide rate data and certain independent variables were selected from (World Development Indicators | DataBank, n.d.). The variables of the dataset are described as follows:

- Gross Domestic Product Per Capital (GDP_PC); measures the country's annual economic output per person.
- Gross Domestic Product Per Capital Growth Rate (GDP_PCG); shows the annual growth rate of the country's economy which is expected that as the growth rate increases, income inequality reduces then crime rates (in this case homicide rates) also reduces.
- Gini Index: is a measure which is expected to represent the income inequality or the wealth inequality within a country. It is measured annually
- Intentional Homicide Rate Per 100000 Inhabitants (Homicide): this is the annual rate of illegal death purposefully inflicted on a person by another

person arising from disagreements at home, social disagreements, disagreements over inheritance and properties, and fights over power control between groups of gangs.

- Employment to population ratio of male aged 15-24 years(ETPR_M1524): this is the male youth population fraction that are considered employed annually. A high ratio on employment shows the sizeable proportion of youths are employed while a low ratio might be seen as a positive sign, especially for young people, if caused by an increase in education.
- Male enrolment rates for primary (SEPM), secondary (SESM) and tertiary (SETM) education: Male youths, by assumption, are susceptible to get involved with unlawful occurrences compared to other groups in the society, leading to increase in the involvement of unlawful offences in the adolescent age.
- Unemployment in male population (UNEM_M):this is the male group of labour force that are not employed but are available for work annually. A significant increase in unemployment, further widens income inequality which extends to increase in homicide rate
- Trade: this is the total of imports and exports of goods and services. It is measured annually as a share of GDP.
- Urban Population (Urban_Pop): statistically, these are persons living in urban areas annually.
- Government final consumption expenditure (GOV_EXP): this is governments' annual current expenses on the purchase of goods and services. It is measured as a percentage of GDP

The data was read into RStudio as

data <- read.csv("Homicide.csv", header = TRUE)

The issues of missing values make data not suitable for research works, on the selected dataset, 207 missing values was detected which was replaced with zeros

```
> sum(is.na(data))
[1] 207
> data[is.na(data)] = 0
> sum(is.na(data))
[1] 0
```

Using the *DataExplorer* package to plot boxplots, outliers were detected in some of the variables but were not removed.

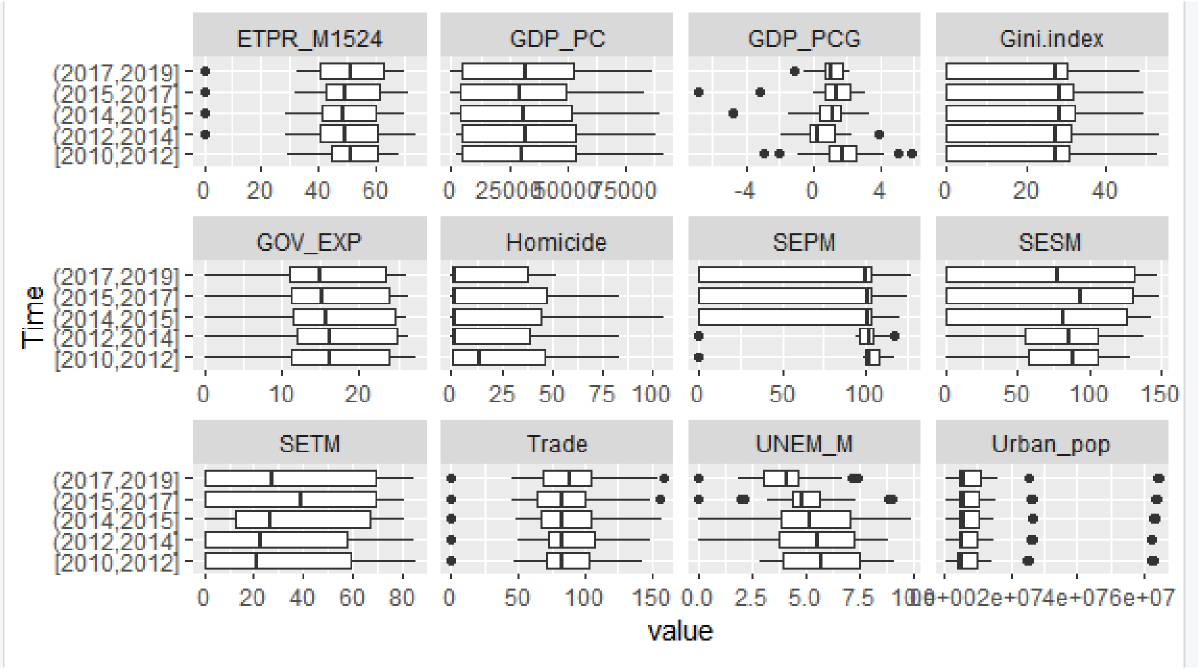


Fig 3.2 Detected outliers.

In the process of data preparation, all countries were assigned individually to the dependent variable (Homicide) on the RStudio.

4. Analysis

A detailed analysis is to be carried out on the homicide data that was read into R as 'data' to help discover patterns, relations and valuable information associated with homicide rates in Europe and Latin America and Caribbean

4.1. Descriptive Analysis

A comprehensive descriptive statistical analysis was done on the data. Names on the dataset are abbreviations of the original indicators as described in Table 3.1. The structure of the dataset is checked for understanding of the data. The dataset consists of two-integer variables, two-character variables and 11-numerical variables with a total of 120 observations and 15 variables.

The *DescTools* package was used for a comprehensive description of the data. It helped in describing each variable in the dataset.

```
1 - Time (integer)

    length         n       NAs    unique        0s      mean    meanCI'
       120       120         0        10         0  2'014.50   2'013.98
              100.0%      0.0%                0.0%             2'015.02

       .05       .10       .25    median       .75       .90        .95
  2'010.00  2'010.90  2'012.00  2'014.50  2'017.00  2'018.10   2'019.00

     range        sd     vcoef       mad       IQR      skew       kurt
      9.00      2.88      0.00      3.71      5.00      0.00      -1.25


    value  freq   perc  cumfreq  cumperc
1    2010    12  10.0%       12    10.0%
2    2011    12  10.0%       24    20.0%
3    2012    12  10.0%       36    30.0%
4    2013    12  10.0%       48    40.0%
5    2014    12  10.0%       60    50.0%
6    2015    12  10.0%       72    60.0%
7    2016    12  10.0%       84    70.0%
8    2017    12  10.0%       96    80.0%
9    2018    12  10.0%      108    90.0%
10   2019    12  10.0%      120   100.0%

' 95%-CI (classic)
```

```
2 - Country (character)

  length      n    NAs unique levels  dupes
     120    120      0     12     12      y
         100.0%   0.0%

                    level  freq  perc  cumfreq  cumperc
1                 Denmark    10  8.3%       10     8.3%
2             El Salvador    10  8.3%       20    16.7%
3                 Finland    10  8.3%       30    25.0%
4                 Germany    10  8.3%       40    33.3%
5               Guatemala    10  8.3%       50    41.7%
6                Honduras    10  8.3%       60    50.0%
7                 Jamaica    10  8.3%       70    58.3%
8             Netherlands    10  8.3%       80    66.7%
9                  Sweden    10  8.3%       90    75.0%
10            Switzerland    10  8.3%      100    83.3%
11    Trinidad and Tobago    10  8.3%      110    91.7%
12          Venezuela, RB    10  8.3%      120   100.0%

----------------------------------------------------------
3 - Region (character - dichotomous)

  length      n    NAs unique
     120    120      0      2
         100.0%   0.0%

               freq   perc  lci.95  uci.95'
Europe           60  50.0%   41.2%   58.8%
Latin America    60  50.0%   41.2%   58.8%

' 95%-CI (Wilson)
```

```
------------------------------------------------------------------------------
4 - GDP_PC (numeric)

      length          n        NAs      unique         0s        mean     meanCI'
         120        120          0         116          5  31'904.776  26'916.496
                 100.0%       0.0%                   4.2%              36'893.055

         .05        .10        .25      median        .75         .90         .95
   2'079.117  2'442.613  4'138.954  30'696.888  52'574.466  61'608.101  84'804.061

       range         sd      vcoef         mad        IQR        skew        kurt
  91'254.035 27'596.521      0.865  38'282.288  48'435.512       0.368      -1.149

lowest : 0.0 (5), 1'904.347, 2'088.316, 2'102.592, 2'144.342
highest: 86'388.405, 86'547.671, 88'109.487, 89'684.708, 91'254.035

' 95%-CI (classic)

------------------------------------------------------------------------------
5 - GDP_PCG (numeric)

      length          n        NAs     unique         0s       mean     meanCI'
         120        120          0        116          5  1.0274034   0.7257664
                 100.0%       0.0%                  4.2%              1.3290404

         .05        .10        .25     median        .75        .90        .95
  -1.5024032 -0.9165445  0.3586673  1.1670273  1.9463915  2.5208500  3.0979795

       range         sd      vcoef        mad        IQR       skew       kurt
  12.6789400  1.6687380  1.6242286  1.2106725  1.5877241 -1.1265685  4.4873842

lowest : -6.8093043, -4.7322748, -3.1380408, -2.9052634, -1.9699413
highest: 3.3901763, 3.9244988, 4.3396068, 5.0526786, 5.8696357

' 95%-CI (classic)
```

6 - Gini.index (numeric)

```
  length       n     NAs  unique      0s    mean  meanCI'
     120     120       0      55      41  22.137  19.015
          100.0%    0.0%           34.2%          25.258

     .05     .10     .25  median     .75     .90     .95
   0.000   0.000   0.000  27.800  31.625  43.410  49.420

   range      sd   vcoef     mad     IQR    skew    kurt
  53.400  17.269   0.780  16.160  31.625  -0.169  -1.274

lowest : 0.0 (41), 26.8, 27.1 (3), 27.2 (2), 27.3 (2)
highest: 49.9, 50.0, 52.6, 53.1, 53.4

heap(?): remarkable frequency (34.2%) for the mode(s) (= 0)

' 95%-CI (classic)
```

---

7 - Homicide (numeric)

```
      length            n          NAs       unique           0s         mean      meanCI'
         120          120            0          116            5   22.4628940  17.7248262
                   100.0%         0.0%                      4.2%               27.2009618

         .05          .10          .25       median          .75          .90          .95
   0.5305520    0.5857398    0.8485487    1.6010023   41.1310047   58.6833550   70.5156353

       range           sd        vcoef          mad          IQR         skew         kurt
 105.2311885   26.2122799    1.1669146    2.0032671   40.2824559    0.8194416   -0.3804811

lowest : 0.0 (5), 0.4996342, 0.5321793, 0.5354216, 0.5369983
highest: 74.9747489, 83.0026114, 83.0063921, 83.7669665, 105.2311885

' 95%-CI (classic)
```

---

8 - Trade (numeric)

```
     length          n        NAs     unique         0s       mean     meanCI'
        120        120          0        106         15   80.59146   73.36574
                100.0%       0.0%                 12.5%               87.81719

        .05        .10        .25     median        .75        .90        .95
    0.00000    0.00000   69.43564   83.62349  103.31852  123.19193  149.28246

      range         sd      vcoef        mad        IQR       skew       kurt
  158.82321   39.97468    0.49602   28.94341   33.88288   -0.48342    0.04566

lowest : 0.0 (15), 45.59652, 46.06942, 46.13689, 46.37274
highest: 150.05376, 155.27068, 156.02821, 157.81658, 158.82321

heap(?): remarkable frequency (12.5%) for the mode(s) (= 0)

' 95%-CI (classic)
```

---

9 - GOV_EXP (numeric)

```
    length         n       NAs    unique        0s      mean    meanCI'
       120       120         0       106        15  16.17063  14.69409
              100.0%      0.0%               12.5%            17.64718

       .05       .10       .25    median       .75       .90       .95
   0.00000   0.00000  11.27981  16.11486  24.32954  25.78690  26.08995

     range        sd     vcoef       mad       IQR      skew      kurt
  27.36585   8.16866   0.50515   7.80408  13.04973  -0.56206  -0.53562

lowest : 0.0 (15), 10.35848, 10.44035, 10.5135, 10.57282
highest: 26.24334, 26.36539, 26.47146, 26.56467, 27.36585

heap(?): remarkable frequency (12.5%) for the mode(s) (= 0)
```

10 - ETPR_M1524 (numeric)

```
   length         n       NAs    unique        0s      mean   meanCI'
      120       120         0       113         8   48.7063   45.6054
             100.0%      0.0%                6.7%             51.8072

      .05       .10       .25    median       .75       .90       .95
   0.0000   30.6580   41.4200   50.0350   61.2325   67.3380   68.7055

    range        sd     vcoef       mad       IQR      skew      kurt
  74.2100   17.1550    0.3522   14.6629   19.8125   -1.3115    1.7879

lowest : 0.0 (8), 28.32, 28.73, 29.46, 29.83
highest: 68.97, 69.9, 70.29, 71.76, 74.21

' 95%-CI (classic)
```

---

11 - SEPM (numeric)

```
    length          n       NAs     unique         0s       mean    meanCI'
       120        120         0         88         33    76.1240    67.5510
               100.0%      0.0%                 27.5%               84.6971

       .05        .10       .25     median        .75        .90        .95
    0.0000     0.0000    0.0000   101.7374   103.9403   112.4756   117.5792

     range         sd     vcoef        mad        IQR       skew       kurt
  127.5501    47.4284    0.6230     5.4361   103.9403    -0.9437    -1.0336

lowest : 0.0 (33), 92.6128, 94.2505, 96.7035, 98.2507
highest: 120.4872, 123.2888, 123.7832, 125.751, 127.5501

heap(?): remarkable frequency (27.5%) for the mode(s) (= 0)

' 95%-CI (classic)
```

---

12 - SESM (numeric)

```
     length           n         NAs      unique          0s        mean     meanCI'
        120         120           0          86          35    72.32545    62.77700
                 100.0%        0.0%                   29.2%                81.87391

        .05         .10         .25      median         .75         .90         .95
    0.00000     0.00000     0.00000    86.22078   120.96669   133.71603   141.63502

      range          sd       vcoef         mad         IQR        skew        kurt
  148.11768    52.82465     0.73037    60.58317   120.96669    -0.28541    -1.43007

lowest : 0.0 (35), 45.57154, 52.14148, 53.89228, 53.9577
highest: 144.94002, 146.77394, 147.00403, 147.37003, 148.11768

heap(?): remarkable frequency (29.2%) for the mode(s) (= 0)

' 95%-CI (classic)
```

---

13 - SETM (numeric)

```
    length          n        NAs     unique         0s       mean     meanCI'
       120        120          0         83         38   34.99609    29.43539
               100.0%       0.0%                 31.7%               40.55680

       .05        .10        .25     median        .75        .90         .95
   0.00000    0.00000    0.00000   26.53102   64.11553   75.96836    81.64710

     range         sd      vcoef        mad        IQR       skew        kurt
  85.30917   30.76333    0.87905   39.33490   64.11553    0.16935    -1.58040

lowest : 0.0 (38), 15.86603, 16.16997, 16.38831, 16.54838
highest: 82.81042, 84.18814, 84.27249, 84.81826, 85.30917

heap(?): remarkable frequency (31.7%) for the mode(s) (= 0)
```

```
------------------------------------------------------------
14 - UNEM_M (numeric)

  length       n     NAs  unique      0s    mean  meanCI'
     120     120       0     107       8  4.9962  4.5699
          100.0%    0.0%            6.7%          5.4225

     .05     .10     .25  median     .75     .90     .95
  0.0000  2.0550  3.6875  4.7500  6.7225  8.2010  8.8225

   range      sd   vcoef     mad     IQR    skew    kurt
  9.9000  2.3584  0.4720  2.2610  3.0350 -0.1574 -0.3494

lowest : 0.0 (8), 1.82, 1.93, 1.99, 2.01
highest: 9.04, 9.09, 9.33, 9.88, 9.9

' 95%-CI (classic)
```

```
------------------------------------------------------------------------------
15 - Urban_pop (integer)

      length            n          NAs       unique           0s      mean'
         120          120            0          = n            0  1.23e+07
                   100.0%         0.0%                      0.0%

         .05          .10          .25       median          .75       .90
  733'406.75 1'537'053.40 4'530'480.50 5'693'056.50     1.04e+07  2.63e+07

       range           sd        vcoef          mad          IQR      skew
    6.36e+07     1.67e+07         1.36 3'645'326.44 5'850'488.25      2.29

      meanCI
9'278'064.07
    1.53e+07

         .95
    6.25e+07

        kurt
        4.19

lowest : 717'530, 719'346, 721'775, 724'640, 727'665
highest: 6e+07, 6e+07, 6e+07, 6e+07, 6e+07

' 95%-CI (classic)
```

4.2 Correlation Analysis: Correlation analysis is a statistical technique which aims to measure the strength at which a pair of variables is related. A correlation analysis will be carried out on the homicide data to help examine the relation between homicide and other variables. The visual 4.2.1 is the Correlation matrix showing the correlations between homicide rate (dependent variable) and other independent variables. Most of the independent variables are negatively correlated with homicide rate while few shows weak to no correlation.

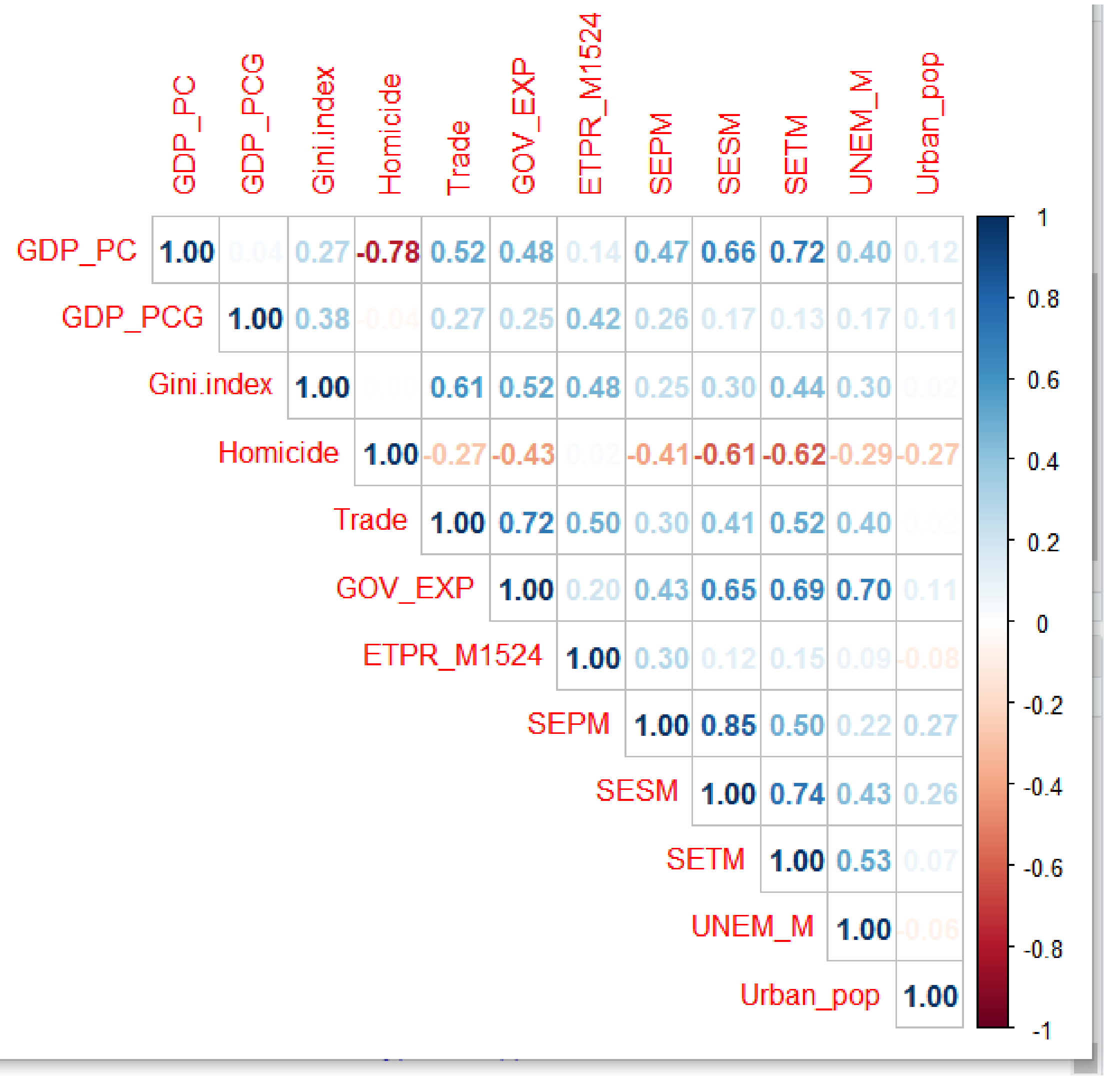


Figure 4.2.1 Correlation matrix

4.3. Regression Analysis

Regression analysis examines the possible linear relation between Homicide rate and other independent variables. Both the dependent variable and independent variables should be numerical, either continuous or discrete. To find out which independent variable that can possibly better explain the homicide rate variation, we will build a correlation matrix between all numerical variables of the data in figure 4.3.1

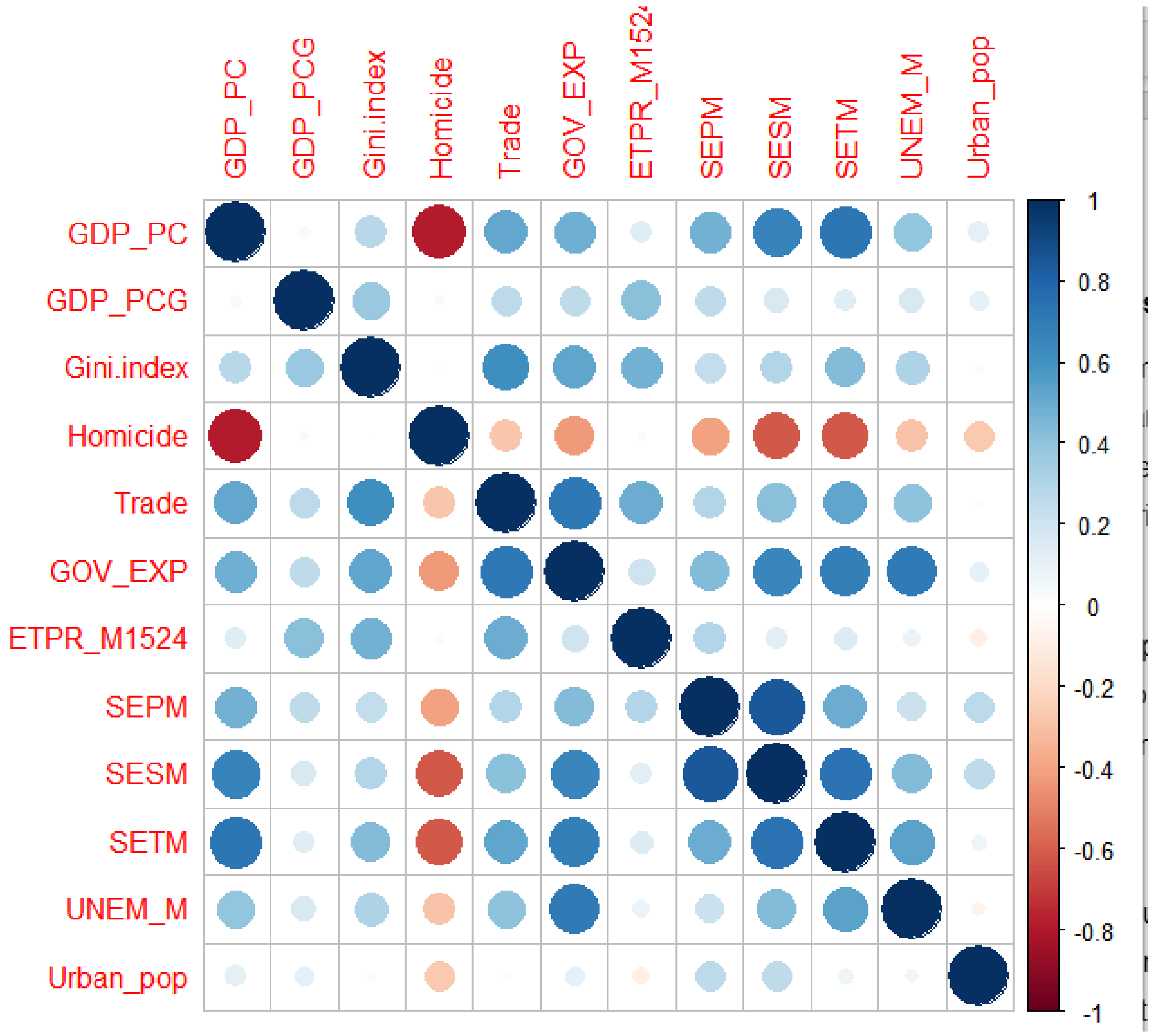


Figure 4.3.1 Correlation matrix for regression analysis

lm() function in default stats package of the RStudio was used to fit the regression model. From the correlation matrix it shows that Homicide is strongly negatively correlated to GDP Per Capital, that an increase in one variable should lead to a decrease in the other. A Simple Linear Regression model was tried between Homicide (Y) and GDP_PC (X). Forward Stepwise technique was used to run the regression analysis. The result is

```
Residuals:
    Min      1Q  Median      3Q     Max 
-46.137  -8.206  -2.336   8.876  61.844 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  4.614e+01  2.299e+00   20.07   <2e-16 ***
GDP_PC      -7.420e-04  5.459e-05  -13.59   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 16.43 on 118 degrees of freedom
Multiple R-squared:  0.6103,	Adjusted R-squared:  0.607 
F-statistic: 184.8 on 1 and 118 DF,  p-value: < 2.2e-16
```

Checking the Pr(>|t|) column we see the coefficients of intercept and GDP_PC are significant, and the simple linear regression equation will be:

Homicide = 4.614e+21 – 7.420e-04 × GDP_PC

R^2 is 0.61.

Also, SETM and SESM has strong negative correlation with Homicide, adding one of the variables to the Simple Linear Regression model to see the effect. Homicide(Y), GDP_PC(X1), SETM(X2), gives

```
Residuals:
    Min      1Q  Median      3Q     Max 
-46.971  -7.454  -2.765   8.488  63.139 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  4.697e+01  2.384e+00  19.705  < 2e-16 ***
GDP_PC      -6.689e-04  7.903e-05  -8.463 8.67e-14 ***
SETM        -9.051e-02  7.090e-02  -1.277    0.204    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 16.39 on 117 degrees of freedom
Multiple R-squared:  0.6156,	Adjusted R-squared:  0.6091 
F-statistic:  93.7 on 2 and 117 DF,  p-value: < 2.2e-16
```

Checking the Pr(>|t|) column we see the coefficients of intercept and GDP_PC are significant at 0.05 level while coefficient of SETM is not significant. R^2 is 0.62 which is not much difference compared to regression model on Homicide and Gross Domestic Product Per Capital (GDP_PC) whose R^2 is 0.61.

Adding SESM to the simple linear regression, Homicide(Y), GDP_PC(X1), SESM(X2), gives

```
Call:
lm(formula = Homicide ~ GDP_PC + SESM, data = dt)

Residuals:
    Min      1Q  Median      3Q     Max 
-48.760  -8.209  -3.053   8.145  58.833 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  4.876e+01  2.558e+00  19.064  < 2e-16 ***
GDP_PC      -6.375e-04  7.176e-05  -8.883 9.16e-15 ***
SESM        -8.240e-02  3.749e-02  -2.198   0.0299 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 16.17 on 117 degrees of freedom
Multiple R-squared:  0.6257,	Adjusted R-squared:  0.6193 
F-statistic:  97.8 on 2 and 117 DF,  p-value: < 2.2e-16
```

Checking the Pr(>|t|) column we see the coefficients of intercept and GDP_PC are significant at 0.05 level while coefficient of SESM is not significant. R^2 is 0.63 which still is not much difference compared to regression model on Homicide and Gross Domestic Product Per Capital (GDP_PC) whose R^2 is 0.61.

Hence, I will stop at the simple linear regression model. Visualizing the fitted regression line, I had to plot a scatter plot.

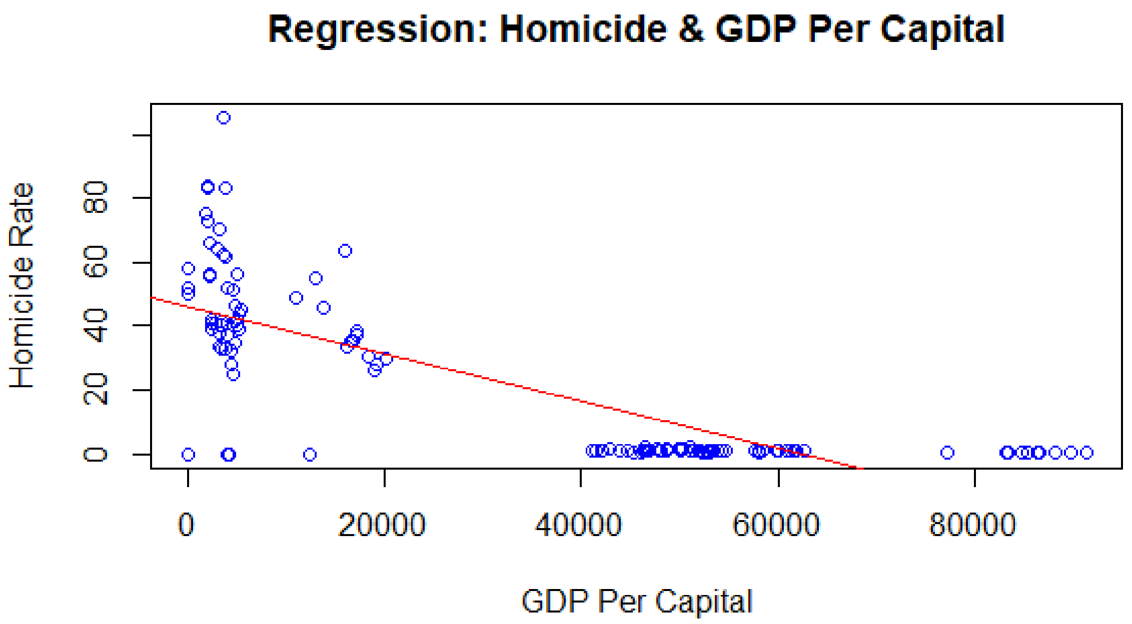


Figure 4.3.2 Regression line

I checked if the fitted model meets the following simple linear regression assumptions; Linearity between X and Y, Residuals' Independence, Normality of residuals, Equal variances of the residuals (Homoscedasticity).

Linearity: The relationship between X and Y must be linear. Check for this assumption by examining a scatterplot of Homicide rate and GDP_PC (Gross Domestic Product Per Capital).

Residuals' Independence: By examining a scatterplot of "residuals versus fitted"; the correlation should be approximately 0. The plot should not look like there is a relationship. Ideally, this plot would not have a pattern where the red line is approximately horizontal at zero (*8: Regression (General Linear Models Part I)*, n.d.).

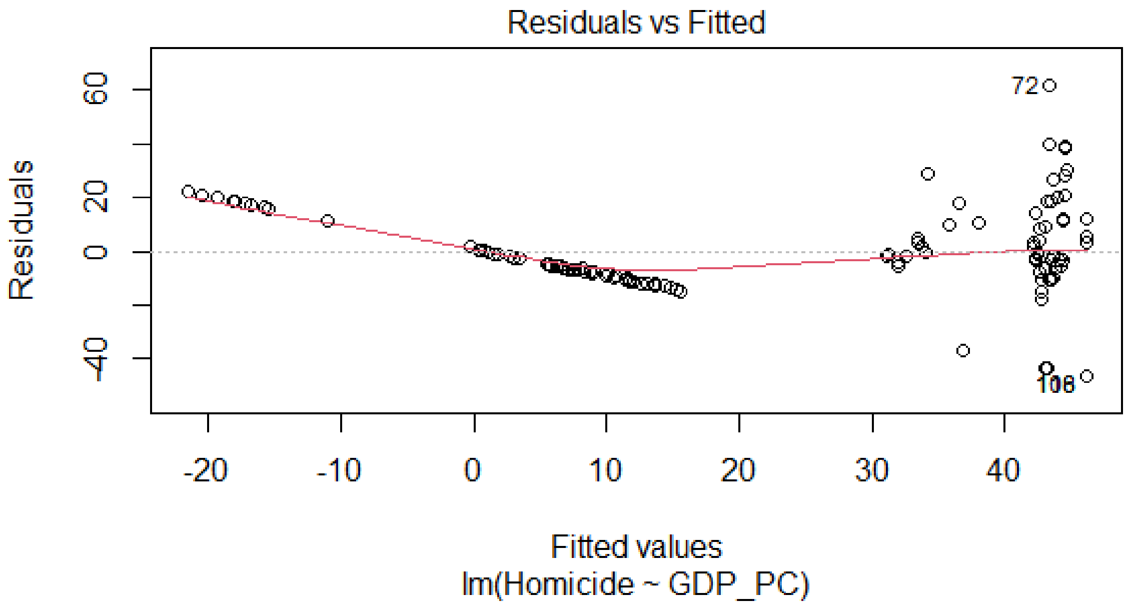


Figure 4.3.3 Residuals’ independence

Normality of Residuals: The residuals must be approximately normally distributed. This assumption is confirmed by examining a normal probability plot; the observations should be near the line (online.stat.psu.edu, n.d.)

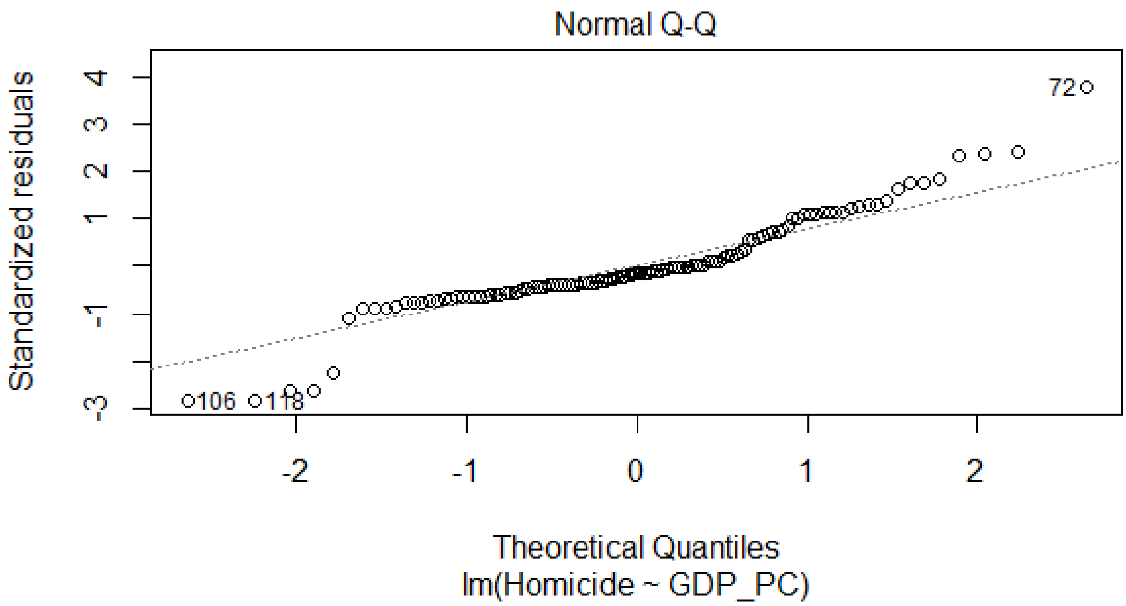


Figure 4.3.4 Normality of residuals

Equal variances of the residuals (Homoscedasticity): The Scale-Location plot is used to check the Homoscedasticity assumption. Homoscedasticity assumption says variance of the residuals are constant and not related to the fitted value (or even x)

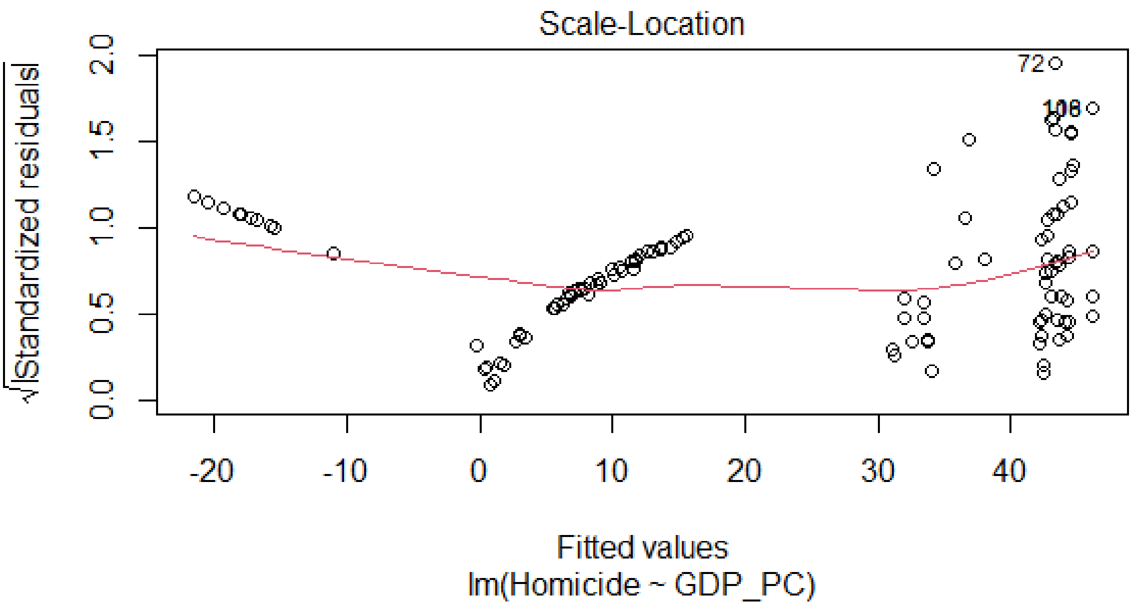


Figure 4.3.4 Homoscedasticity

4.4 Hypothesis Testing: This is a process of testing the hypothesis by comparing it with the null hypothesis. A check for normality should be done before the process of hypothesis testing. From the R studio, datarium, ggplot2, qqplotr and stats packages will be used for the execution. Here the Homicide dataset which is read into R as 'data' is transformed into two tables based on the region namely H_Europe and H_Latin.

Normality: Being unsure if the variable under consideration is normally distributed, these three methods can be used to check. The Q-Q plot (Visual Method), the Shapiro-Wilk Test (Statistical Test) or the Kolmogorov-Smirnov Test (Statistical Test).

The Shapiro-wilk test is used to test the normality of the numerical variable, Homicide, of both regions.

```
> shapiro.test(H_Europe$Homicide)

        Shapiro-Wilk normality test

data:  H_Europe$Homicide
W = 0.89558, p-value = 9.112e-05
```

```
> shapiro.test(H_Latin$Homicide)

        Shapiro-Wilk normality test

data:  H_Latin$Homicide
W = 0.95247, p-value = 0.02038
```

Result 4.4.1 test for normality

Since the p-value of both Shapiro-wilk test for normality is less than .05, it indicates that the data is not normally distributed.

Using Q-Q plot method to test the normality of the data visually, gives,

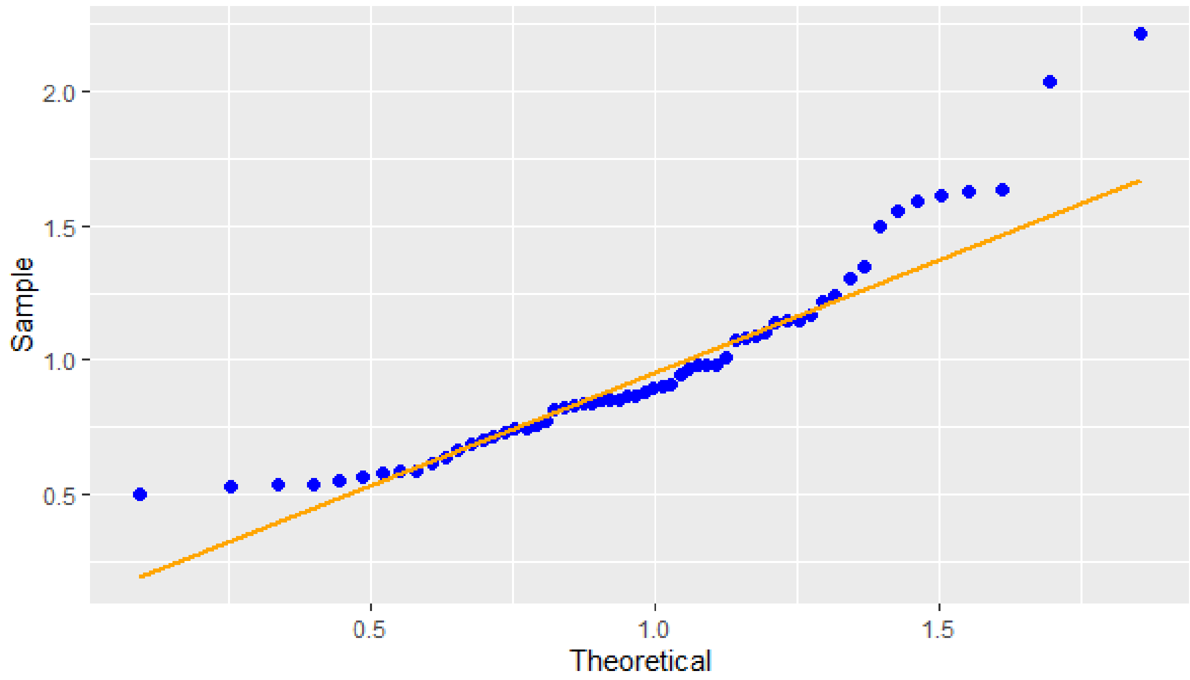


Figure 4.4.1 Q-Q plot normality result for Europe region

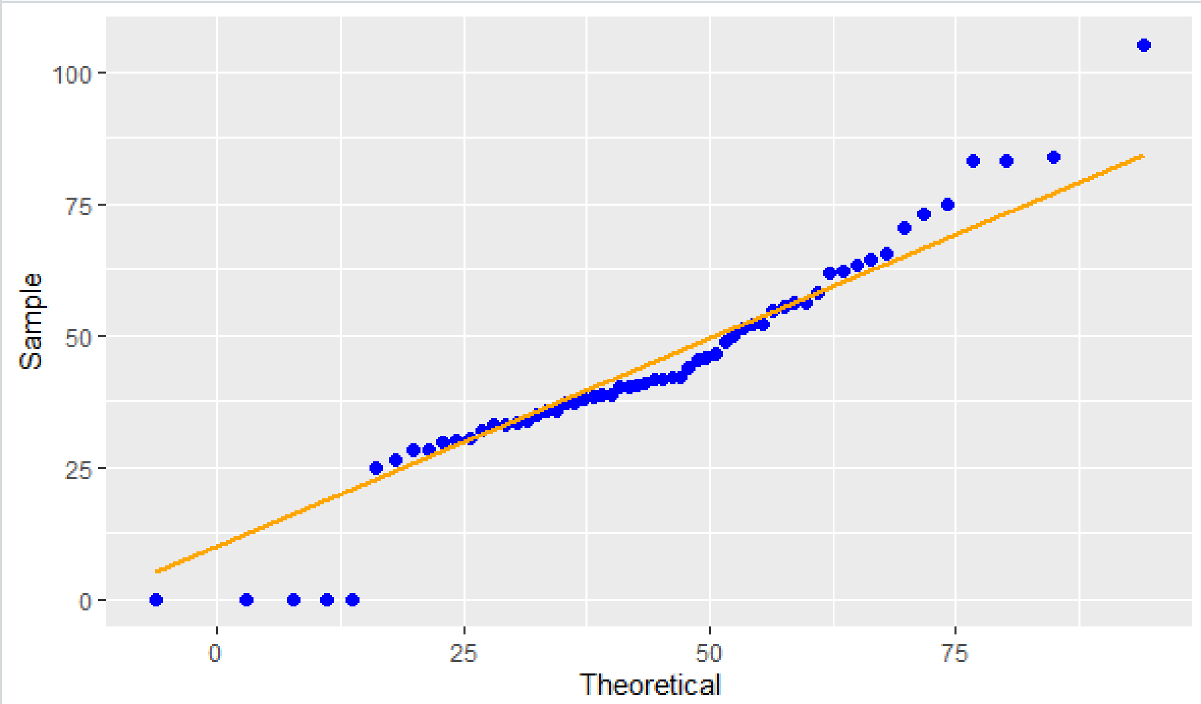


Figure 4.4.2 Q-Q plot normality result for Latin America region

From the plots above, it is seen that both transformed dataset are not normally distributed. Since both datasets are non-normal, I tried applying some transformations on them to see if they could become normal.

```
> log_h <- log10(H_Europe$Homicide)
> shapiro.test(log_h)

        Shapiro-Wilk normality test

data:  log_h
W = 0.9736, p-value = 0.2181

> log_hl <- log10(H_Latin$Homicide)
> shapiro.test(log_hl)

        Shapiro-Wilk normality test

data:  log_hl
W = NaN, p-value = NA

> sqrt_hl <- sqrt(H_Latin$Homicide)
> shapiro.test(sqrt_hl)

        Shapiro-Wilk normality test

data:  sqrt_hl
W = 0.79849, p-value = 1.238e-07

> cube_hl <- H_Latin$Homicide^(1/3)
> shapiro.test(cube_hl)

        Shapiro-Wilk normality test

data:  cube_hl
W = 0.66326, p-value = 1.849e-10
```

Result 4.4.2 Normality result after tranformation

From the above result, a change can be noticed on H_Europe$Homicide whose p-value was 9.112e-05 before the transformation became 0.2181 after the log transformation was applied which is greater than .05, therefore making it normally distributed from its normal state. But H_Latin$Homicide even became more non-normally distributed after the three different transformations were applied on the dataset.

Since the normality assumption has been violated, on non-parametric alternative to T-test, I can still run the hypothesis test using Wilcoxon rank sum test, this test will be performed on Homicide grouped by the Regions. Stating the null and alternative hypothesis as:

H_0 :The rate of homicide is equal in both populations

$H_{1:}$The rate of homicide is not equal in both populations

```
> wilcox.test(Homicide ~ Region, data=data)

        Wilcoxon rank sum test with continuity correction

data:  Homicide by Region
W = 300, p-value = 3.529e-15
alternative hypothesis: true location shift is not equal to 0
```

Result 4.4.3 Hypothesis test result

From the result in Result 4.4.3 we can see that the p-value is way smaller to .05, p-value is 3.529e-15. Hence, rejection of the null hypothesis while the alternative hypothesis is accepted. Therefore, the rate of homicide is not equal in both populations.

ANOVA : analysis of variance is used to examine the effect of two or more levels of one or two independent variables on a dependent variable, where the dependent variable will be numerical, and the independent variables will be categorical. It is a parametric test. The outcome of this analysis depends on some assumptions, stated thus:

- The dependent variable should be continuous
- The independent variables should be categorical with two or more categories
- Observations should be independent
- There should be no significant outliers
- The dependent variable should be approximately normally distributed for each of the independent variable
- The variances of the dependent variable within each category should be homogenous

Analysis of variance will be conducted on the homicide data; the variable Homicide serves as a dependent variable which is continuous while the variable Region serves as an independent variable with two categories also both variables have independent observations. Stating the null and alternative hypothesis as thus:

The null hypothesis is that homicide rate has similar effect on both regions, while the alternative hypothesis is that homicide rate does not have similar effect on both regions.

H_0: $\mu_E = \mu_L$

H_1: $\mu_E \neq \mu_L$

Outliers were detected through the boxplot were not significant as seen in figure 4.4.3 below

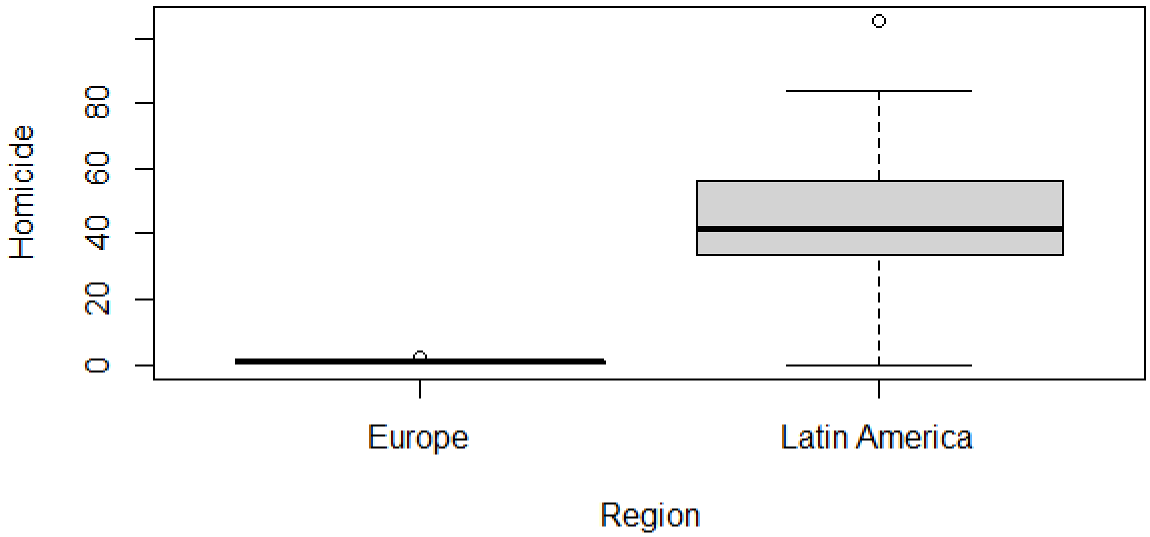


Figure 4.4.3 Outlier detection for ANOVA

From the normality test result in Result 4.4.4, it is observed that the dependent variable(homicide) is not normally distributed for each category of the independent variable since the p-values are less than 0.05.

```
> byf.shapiro(Homicide ~ Region, data=data)

        Shapiro-Wilk normality tests

data:  Homicide by Region

                   W   p-value
Europe        0.8956 9.112e-05 ***
Latin America 0.9525   0.02038 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Result 4.4.4 normality test

The last assumption, I used Bartlett test from **stats** in R to check for the homogeneity of variances in the two categories of the homicide variable. The result below was gotten

```
> bartlett.test(Homicide ~ Region, data=data)

        Bartlett test of homogeneity of variances

data:  Homicide by Region
Bartlett's K-squared = 391.92, df = 1, p-value < 2.2e-16
```

Result 4.4.5 homogeneity of variances

Since most of the assumptions were not meet by the data, I had to run the Welch ANOVA on the dataset where I set var.equal=FALSE to test the hypothesis whether effects of homicide rate are similar in both regions.

```
> oneway.test(Homicide ~ Region,data=data, var.equal = FALSE)

        One-way analysis of means (not assuming equal variances)

data:  Homicide and Region
F = 248.09, num df = 1.000, denom df = 59.036, p-value < 2.2e-16
```

Result 4.4.6 Welch ANOVA

From result 4.4.6, p-value is less than 0.05, the null hypothesis is rejected, thereby accepting the alternative hypothesis which states that homicide rate does not have similar effect on both regions.

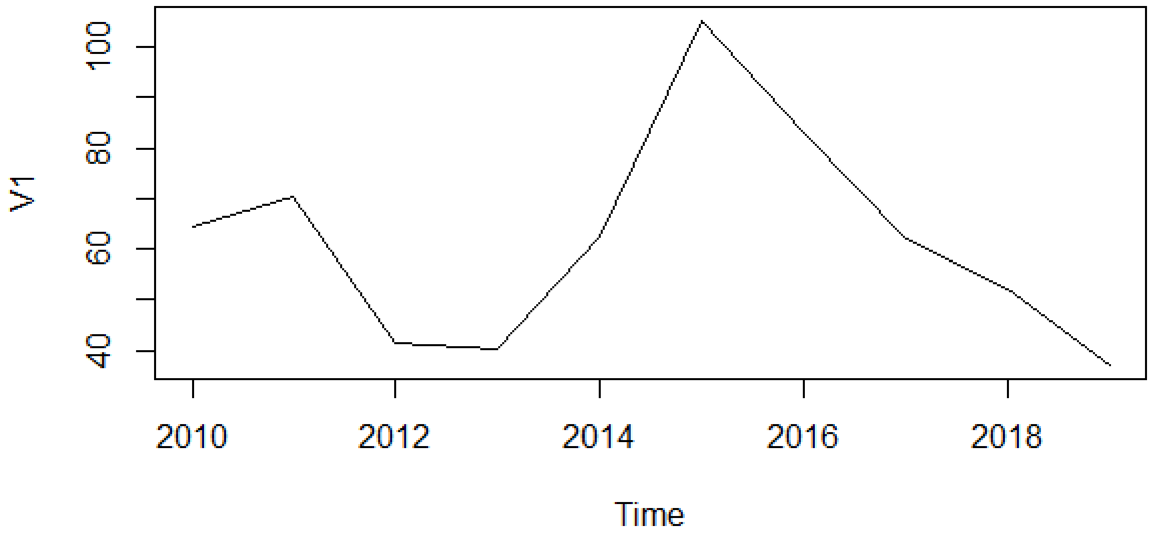
4.5. TIME SERIES ANALYSIS

The time series data of homicide rate in El Salvador over the annual period of 2010-2019 was used for this time series analysis section to forecast the pattern/trend of homicide rate in the next 10 years.

The data was read into the RStudio as:

```
> datatimeseries <- ts(dts, start = c(2010))
> datatimeseries
Time Series:
Start = 2010
End = 2019
Frequency = 1
             V1
 [1,]  64.47412
 [2,]  70.38005
 [3,]  41.58436
 [4,]  40.10484
 [5,]  62.28630
 [6,] 105.23119
 [7,]  83.00639
 [8,]  62.02134
 [9,]  52.11237
[10,]  37.15784
```

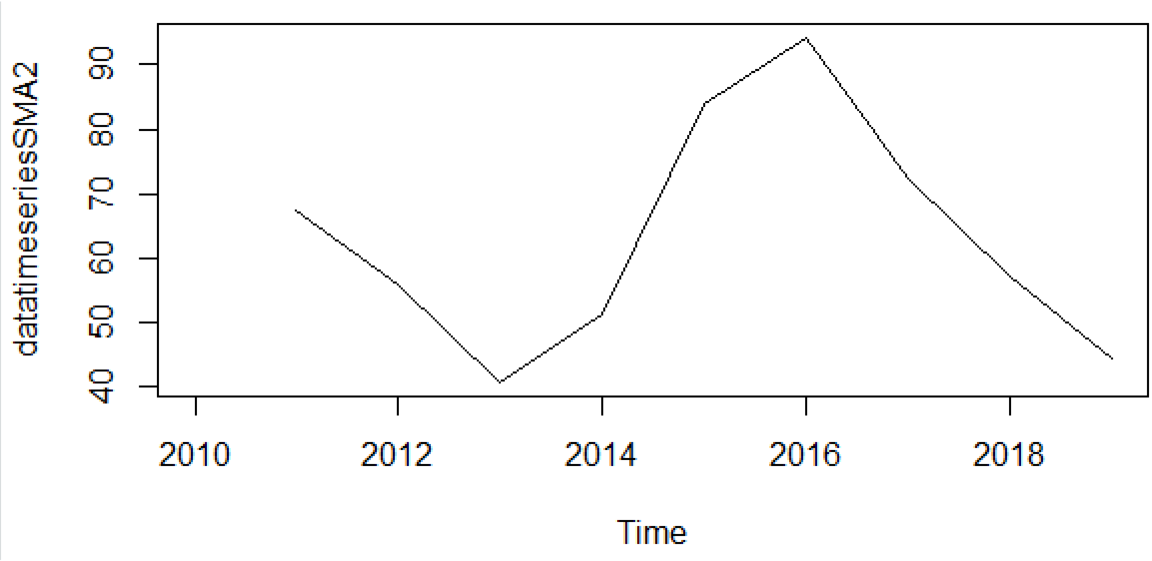
Plotting the time series data using plot.ts() function in R gives:



Plot 4.5.1. Time series plot

Since the random flunctuations are roughly constant in size over time, the time series from the time plot could probably be described using an additive model. It is observed to be a non-seasonal data with downward trend.

Next I decomposed the non-seasonal time series data, which tried to separate the time series into estimate of trend component and irregular component using smoothing method. SMA() function in the “TTR” R package was used to smooth the time series data using a simple moving average,which I used an order of 2 to set the simple moving average.



Plot 4.5.2 Plot after decomposing using smoothing method

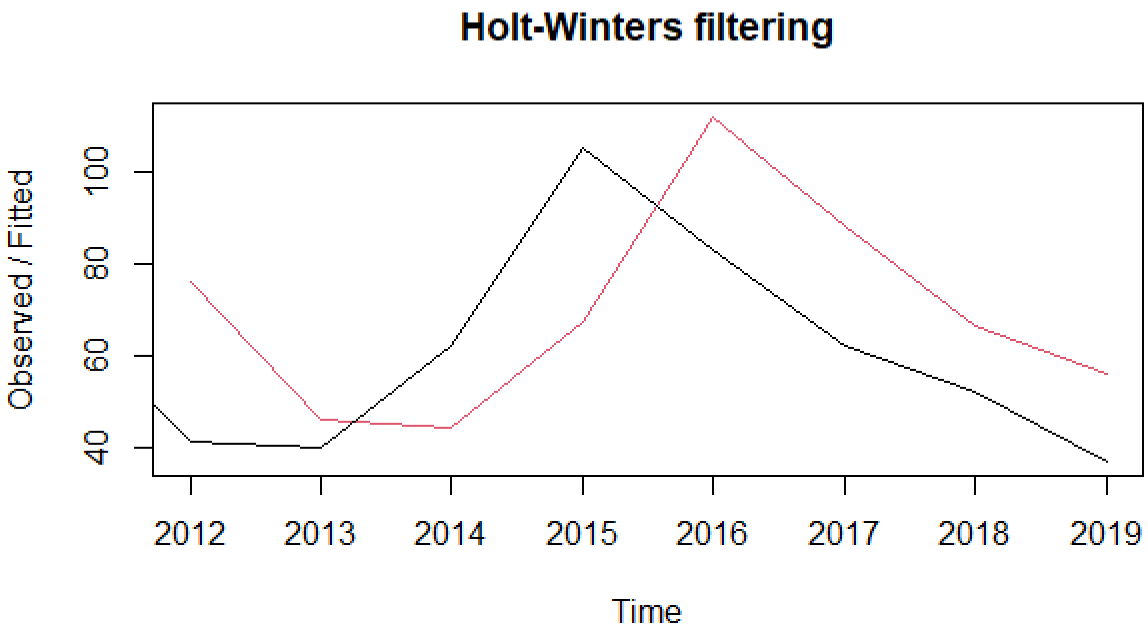
I made short term forecast using the Holt's exponential smoothing since the time series could be described as additive model with decreasing trend and no seasonality. To make this forecast I had to fit a predictive model using the HoltWinters() function in R as seen below:

```
> dtsforcast <- HoltWinters(datatimeseries, gamma=FALSE)
> dtsforcast
Holt-Winters exponential smoothing with trend and without seasonal component.

Call:
HoltWinters(x = datatimeseries, gamma = FALSE)

Smoothing parameters:
 alpha: 1
 beta : 0.0354036
 gamma: FALSE

Coefficients:
       [,1]
a 37.157843
b  3.298861
```

The output values of the forecasts using the HoltWinters() function is as seen below:

```
> dtsforcast$fitted
Time Series:
Start = 2012
End = 2019
Frequency = 1
          xhat     level    trend
2012  76.28598  70.38005 5.905932
2013  46.26173  41.58436 4.677370
2014  44.56424  40.10484 4.459394
2015  67.37312  62.28630 5.086819
2016 111.65832 105.23119 6.427131
2017  88.41914  83.00639 5.412749
2018  66.49951  62.02134 4.478172
2019  56.08119  52.11237 3.968815
```

Plotting the original time series against the forcast gives Plot 4.5.3, where the original time series is in black and the forecast is in red

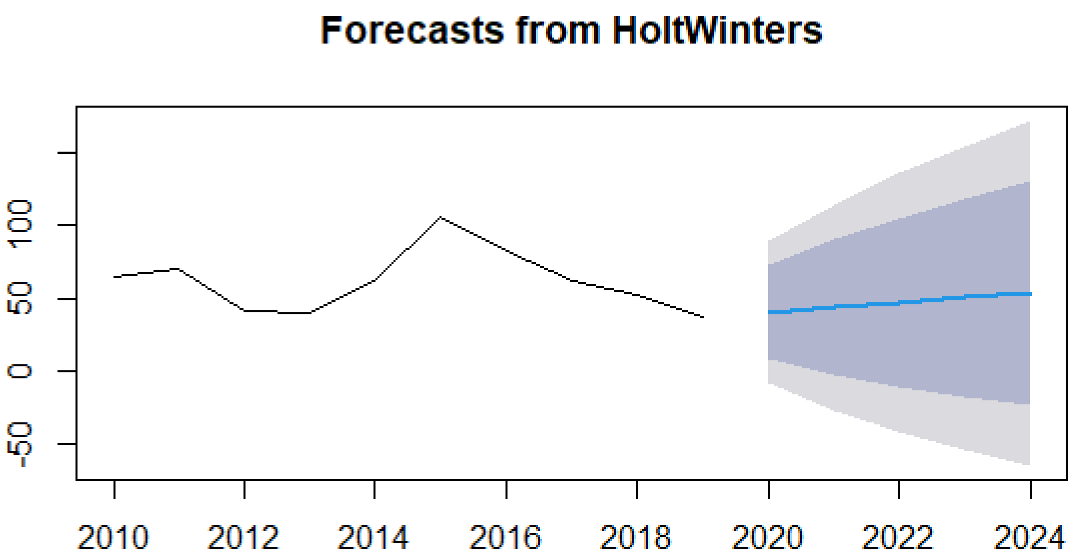


Plot 4.5.3

```
> dtsforcast$SSE
[1] 5072.274
```

The sum of squared errors for the in-sample forecast errors here is 5072.274

Predictions for 2020-2024 was obtained using the forecast() function from the forecast package in R



Plot 4.5.4

.

5. Discussion

The comprehensive descriptive analysis showed the characteristics of the individual variable of the dataset, thereby giving better understanding of the data. The correlation analysis of homicide being the dependent variable against the independent variables shows that Gross Domestic Product per Capital (GDP_PC) is highly negatively correlated to homicide followed by the gross rate of male enrolled in tertiary and secondary schools (SETM, SESM), which might indicate that an increase in these independent variables should effect a decrease in homicide rates. A simple regression model was considered on the dataset after the trial of a multiple regression model failed, and an estimated R-squared value of 0.61, this means Gross Domestic Product Per Capital (GDP_PC) can predict 61% of the entire variability in the Homicide rate. The simple regression model was then fitted and checked if it meets the simple linear regression assumptions giving a simple regression equation of

Homicide = 4.614e+21 – 7.420e-04 × GDP_PC which might be used to predict homicide rates based on Gross Domestic Product per Capital indicator.

Data transformation was done for the process of hypothesis testing since this test will be checked based on region. After the test for normality failed , the Wilcoxon test was used to carry out the hypothesis testing where the null hypothesis was rejected and the alternative hypothesis was accepted, which stated that the rate of homicide is not equal in both populations. In which based on the dataset collected the rates are not equal as the Latin American and Caribbean countries had high homicide rate records compared to the European countries. The analysis of variance test also resulted in that homicide rate does not have similar effect on both regions. Both regions have high distinctive rates based on their records, hence the low to no homicide rates in Europe might not have any effect on their economy unlike in Latin America and Caribbeans whose homicide rates are extremely high.

A consideration was made on El Salvador for time series data analysis, even despite having high GDP_PC they are still recording high homicide rates, for the period of 2010-2019. From the forecast plot obtained using smoothing model for the prediction of 2020-2024, El Salvador doesn't seem to be recording low homicide rates anytime soon.

6. Conclusions

Homicide and the process of eliminating it drastically should also be considered as its reduction in Europe is quite impressive. This study analysed statistically the different patterns, trends, and effect of homicide in Europe and Latin America and Caribbean. Negative correlation and a simple linear regression equation, for predicting a new sample, was obtained between Homicide and GDP_PC. Government needs to prioritize the sectors that are sources of the country's GDP_PC, as an increase in GDP_PC should possibly help curb crime rates (Shafiq & Ali, 2022). Although El Salvador might need to work on the role of public policies in conditioning the act of homicide (Cao & Zhang, 2016).

1 Part Three: References and Appendices

ADELEYE, Ngozi & Bosede, (2014). The Determinants of Income Inequality and the Relationship to Crime.

Alhamadi, M. (2020). Challenges, Strategies and Adaptations on Interactive Dashboards. *Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization*. https://doi.org/10.1145/3340631.3398678

Ali, S. M., Gupta, N., Nayak, G. K., & Lenka, R. K. (2016). Big data visualization: Tools and challenges. *2016 2nd International Conference on Contemporary Computing and Informatics (IC3I)*. https://doi.org/10.1109/ic3i.2016.7918044

Cao, L., & Zhang, Y. (2016). Governance and Regional Variation of Homicide Rates. International Journal of Offender Therapy and Comparative Criminology, 61(1), 25–45. https://doi.org/10.1177/0306624x15587278

Carter, C., & Mahoney, A. (2022). Employing a Repeated Time Series Design to Evaluate Waiting Periods on Firearm Homicide and Suicide Rates. *Behavior and Social Issues*. https://doi.org/10.1007/s42822-021-00089-5

Chui, C., Man, X., Chew, J., Kai, W., & Mun, P. (n.d.). *Baltimore Homicide Statistics: An Interactive Visual Analytical Dashboard for Analyzing Homicide Data*. https://wiki.smu.edu.sg/1516t1is428g1/img_auth.php/2/25/Research_paper_cc.pdf

Diamond, S., Basu, R., Cao, S., Hussain, A. (2021). The Canadian Cultural Diversity Dashboard: Data Storytelling and Visualization for the Cultural Sector. In: Soares, M.M., Rosenzweig, E., Marcus, A. (eds) Design, User Experience, and Usability: UX Research and Design. HCII 2021. Lecture Notes in Computer Science (), vol 12779. Springer, Cham. https://doi.org/10.1007/978-3-030-78221-4_25

Grammel, L., Tory, M., & Storey, M. (2010). How Information Visualization Novices Construct Visualizations. *IEEE Transactions on Visualization and Computer Graphics*, *16*(6), 943–952. https://doi.org/10.1109/tvcg.2010.164

Harris, Anthony R.; Stephen H. Thomas; Gene A. Fisher; David J. Hirsch (May 2002). "Murder and medicine: the lethality of criminal assault 1960–1999". Homicide Studies. **6** (2): 128–166. *doi*:*10.1177/1088767902006002003*.

K. C. Land, P. L. McCall, and L. E. Cohen, "Structural covariates of homicide rates: Are there any invariances across time and social space?", American journal of sociology, vol. 95, no. 4, pp. 922-963, 1990

Loureiro, P. R. A., & Silva, E. C. D. (2011). WHAT CAUSES INTENTIONAL HOMICIDE? *Journal of International Development*, *24*(3), 287–303. https://doi.org/10.1002/jid.1764

Machado Filho, A., da Silva, M. F., & Zebende, G. F. (2014). Autocorrelation and cross-correlation in time series of homicide and attempted homicide. *Physica A: Statistical Mechanics and Its Applications*, *400*, 12–19. https://doi.org/10.1016/j.physa.2014.01.015

Maltz, M.D., Gordon, A.C. and Friedman, W. (2018). Mapping Crime In Its Community Setting. Franklin Classics.

Mihart (n.d.). Visualization types in Power BI - Power BI. [online] learn.microsoft.com. Available at: https://learn.microsoft.com/en-us/power-bi/visuals/power-bi-visualization-types-for-reports-and-q-and-a.

Mohammadi, A., Bergquist, R., Fathi, G., Pishgar, E., de Melo, S. N., Sharifi, A., & Kiani, B. (2022). Homicide rates are spatially associated with built environment and socio-economic factors: a study in the neighborhoods of Toronto, Canada. *BMC Public Health*, *22*(1). https://doi.org/10.1186/s12889-022-13807-4

O'Brien, R. M. (2022). Visualizing Changes in the Age-Distributions of Homicides in the United States: 1964–2019. *Journal of Quantitative Criminology*. https://doi.org/10.1007/s10940-022-09543-y

8: Regression (General Linear Models Part I). (n.d.). Online.stat.psu.edu. Retrieved November 26, 2022, from https://online.stat.psu.edu/stat800/book/export/html/700

Roser, M., & Ritchie, H. (2013). Homicides. *Our World in Data*. https://ourworldindata.org/homicides?source=post_page---------------------------.

Sarikaya, A., Correll, M., Bartram, L., Tory, M., & Fisher, D. (2019). What Do We Talk About When We Talk About Dashboards? *IEEE Transactions on Visualization and Computer Graphics*, *25*(1), 682–692. https://doi.org/10.1109/tvcg.2018.2864903

Scales of Measurement. (2014, September 2). Community.qlik.com. https://community.qlik.com/t5/Design/Scales-of-Measurement/ba-p/1468193

Shafiq, N. B., & Ali, D. H. A. (2022). The Economic Determinants of Crime Rate in 7 Selected Countries: A Panel Data Analysis Approach. *Asian Journal of Empirical Research*, *12*(1), 44–53. https://doi.org/10.55493/5004.v12i1.4439

Stamatel, J. P. (2014). How data visualization can improve analytical thinking in cross-national crime research. *International Journal of Comparative and Applied Criminal Justice*, *39*(1), 31–46. https://doi.org/10.1080/01924036.2014.961495

Silva, C., Melo, S., Santos, A., Junior, P. A., Sato, S., Santiago, K., & Sá, L. (2020). Spatial Modeling for Homicide Rates Estimation in Pernambuco State-Brazil. *ISPRS International Journal of Geo-Information*, *9*(12), 740. https://doi.org/10.3390/ijgi9120740

Talagala, T., & Shashikala, R. (n.d.). *Interactive Dashboard to Monitor the COVID-19 Outbreak and Vaccine Administration*. Retrieved December 12, 2022, from https://arxiv.org/pdf/2205.07286.pdf

Vilalta, C., & Muggah, R. (2014). Violent disorder in Ciudad Juarez: a spatial analysis of homicide. *Trends in Organized Crime*, *17*(3), 161–180. https://doi.org/10.1007/s12117-014-9213-0

Walny, J., Frisson, C., West, M., Kosminsky, D., Knudsen, S., Carpendale, S., & Willett, W. (2020). Data Changes Everything: Challenges and Opportunities in Data Visualization Design Handoff. *IEEE Transactions on Visualization and Computer Graphics*, *26*(1), 12–22. https://doi.org/10.1109/TVCG.2019.2934538

Wikipedia Contributors (2019). *List of countries by intentional homicide rate*. [online] Wikipedia. Available at: https://en.wikipedia.org/wiki/List_of_countries_by_intentional_homicide_rate

World development indicators DataBank. Available at: https://databank.worldbank.org/source/world-development-indicators.