CSE3505 FOUNDATION OF DATA ANALYTICS J COMPONENT REPORT

2022-11-16

TITLE: STATISTICAL ANALYSIS OF LIVING IN CANADA

TEAM MEMBERS:

20BPS1137 - KANCHARLA NAVEEN KUMAR

20BPS1142 - DIPANSU ROUT

PROBLEM STATEMENT:

Our main goal in analyzing the data sets comes down to see how multiple variables in a data set affects each other. As we already saw by analyzing cost of living and quality of living differs greatly cities to cities when compared on large scale data like healthcare, expenditure etc. Then we take the whole world for comparison on GDP basis while coming down to focus on a single country like Canada because many factors play role in its statistics being one of the highest in the world.

DATA SOURCE:

https://data.world/atbventures/qualityoflife (https://data.world/atbventures/qualityoflife)

https://data.world/atbventures/qualityoflife/workspace/file?filename=GDP_World.csv (https://data.world/atbventures/qualityoflife/workspace/file?filename=GDP_World.csv)

https://data.world/atbventures/qualityoflife/workspace/file?filename=HealthSpending_byprovince.csv (https://data.world/atbventures/qualityoflife/workspace/file?filename=HealthSpending_byprovince.csv)

https://data.world/atbventures/qualityoflife/workspace/file?filename=GDP_World_10yr_weightedaverage.csv (https://data.world/atbventures/qualityoflife/workspace/file?filename=GDP_World_10yr_weightedaverage.csv)

https://data.world/atbventures/qualityoflife/workspace/file?filename=stats_expenseindicators_canada_province.csv (https://data.world/atbventures/qualityoflife/workspace/file?filename=stats_expenseindicators_canada_province.csv)

OBJECTIVES:

GDP enables policymakers and central banks to judge whether the economy is contracting or expanding and promptly take necessary action. It also allows policymakers, economists, and businesses to analyze the impact of variables such as monetary and fiscal policy, economic shocks, and tax and spending plans.

Health-related quality of life (HRQL) focuses on the impact of health on a person’s ability to live a fulfilling life. HRQL represents a broad concepts of physical, psychological with social functioning and well-being that includes both positive and negative aspects.

Government spends a considerable portion of its income towards provision of social security benefits such as unemployment allowances old age pension, insurance benefits, sickness benefit, medical benefit, etc. Such benefits reduce the desire to work. In other words they act as disincentive to work.

ANALYSIS:

Gross domestic product tracks the health of a country's economy. It represents the value of all goods and services produced over a specific time period within a country's borders. Economists can use GDP to determine whether an economy is growing or experiencing a recession. Investors can use GDP to make investments decisions—a bad economy often means lower earnings and stock prices. Healthcare expenditure can result in better provision of health opportunities, which can strengthen human capital and improve the productivity, thereby contributing to economic performance. It is therefore important to assess the phenomenon of healthcare spending in a country.

IMPORTING THE NECESSARY LIBRARIRES:

```
library(dplyr)
```

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

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

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

```
library(ggplot2)
```

READING THE DATA AND GIVING A DESCRIPTION OF THE DATA.

Analyzing global per capita GDP provides insight into global economic prosperity and developments. Small,rich and more developed industrial countries tend to have the highest per capita GDP.

WHEN THE NA VALUES ARE PRESENT IT SHOULD BE CLEANED USING ANY OF THE CLEANING METHODS. AND HERE NA.OMIT IS USED.

```
df3 <- read.csv('GDP_World.csv')
dim(df3)
```

```
## [1] 8190    3
```

```
sum(is.na(df3)) #THERE ARE NA VALUES PRESENT USE COMPLETE.CASES AND CLEAN THE DATA
```

```
## [1] 681
```

```
df3_clean <- na.omit(df3)
sum(is.na(df3_clean)) #DATA IS CLEANED
```

```
## [1] 0
```

```
dim(df3_clean)
```

```
## [1] 7509    3
```

```
View(df3_clean)
```

STATISTICAL ANALYSIS SHOWS THE SITUATION OF MANY COUNTRIES COMPARED TO THE MEAN VALUE OF GDP AND GDP PER CAPITA

```
#STATISTICAL ANALYSIS

summary(df3_clean)
```

```
##    Country               Year      GDP_percapita
##  Length:7509        Min.   :1980   Min.   :    17.12
##  Class :character   1st Qu.:1993   1st Qu.:   968.15
##  Mode  :character   Median :2004   Median :  3200.97
##                     Mean   :2003   Mean   :  9675.70
##                     3rd Qu.:2014   3rd Qu.: 10981.92
##                     Max.   :2024   Max.   :132701.60
```

```
#YOU CAN SEE THAT THE MAX GDP PER CAPITA IS 132701 AND LUXEMBOURG IS HAVING THE HIGHEST GDP P
ER CAPITA AND MEAN GDP PER CAPITA IS 9675 WHICH SHOWS THATS THERE ARE MANY COUNTRIES LYING UN
DER THE MARGIN GDP PER CAPITA AND THE MINIMUM GDP PER CAPITA IS 17.12 WHICH SHOWS THAT COMBOD
IA IS HAVING THE MINIMUM GDP PER CAPITA
```

MANIPULATING THE DATA FOR THE VISUALISATION. ANALYZING AS PER YEAR AND COMPARING WITH PREVIOUS YEARS. STATISTICAL ANALYSIS SHOWS HOW IN RECENT YEARS THE MEAN GDP CHANGED EITHER INCREASED OR DECREASED. AND IT SHOWS THAT WHAT COUNTRY IS HAVING MAXIMUM AND MINIMUM GDP.

```
DF1 <- subset(df3_clean,df3_clean$Year==2018)
summary(DF1)
```

```
##    Country               Year      GDP_percapita
##  Length:181         Min.   :2018   Min.   :   307
##  Class :character   1st Qu.:2018   1st Qu.:  2038
##  Mode  :character   Median :2018   Median :  6013
##                     Mean   :2018   Mean   : 14498
##                     3rd Qu.:2018   3rd Qu.: 17464
##                     Max.   :2018   Max.   :115536
```

```
#MEAN VALUE IS 14498 WHICH IS AVERAGE GDP PER CAPITA COMPARED TO MAXIMUM GDP PER CAPITA

#WORLD GDP IN THE YEAR 2022 WITH THE YEAR 2018
df4 <- subset(df3_clean,df3_clean$Year==2022)
summary(df4)
```

```
##    Country               Year      GDP_percapita
##  Length:179         Min.   :2022   Min.   :   323.7
##  Class :character   1st Qu.:2022   1st Qu.:  2581.2
##  Mode  :character   Median :2022   Median :  6876.6
##                     Mean   :2022   Mean   : 16049.9
##                     3rd Qu.:2022   3rd Qu.: 20019.0
##                     Max.   :2022   Max.   :124467.1
```

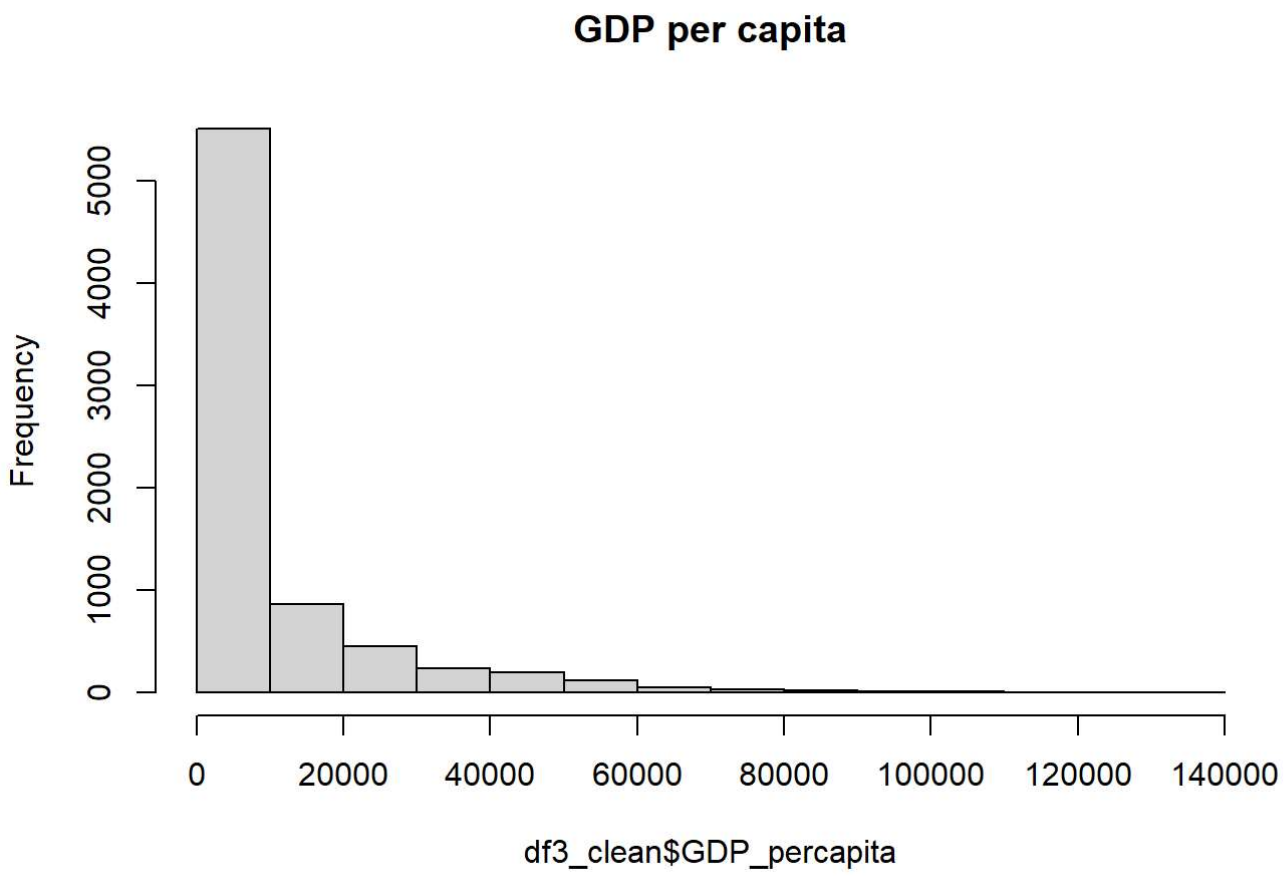
```
#YOU CAN SEE THAT MEAN GDP PER CAPITA IS INCREASED OVER THE YEARS

DF <- subset(df3_clean,df3_clean$Year==2024)
summary(DF)
```

```
##    Country               Year      GDP_percapita
##  Length:179         Min.   :2024   Min.   :   347.2
##  Class :character   1st Qu.:2024   1st Qu.:  2841.3
##  Mode  :character   Median :2024   Median :  7289.0
##                     Mean   :2024   Mean   : 17403.0
##                     3rd Qu.:2024   3rd Qu.: 21429.2
##                     Max.   :2024   Max.   :132701.6
```

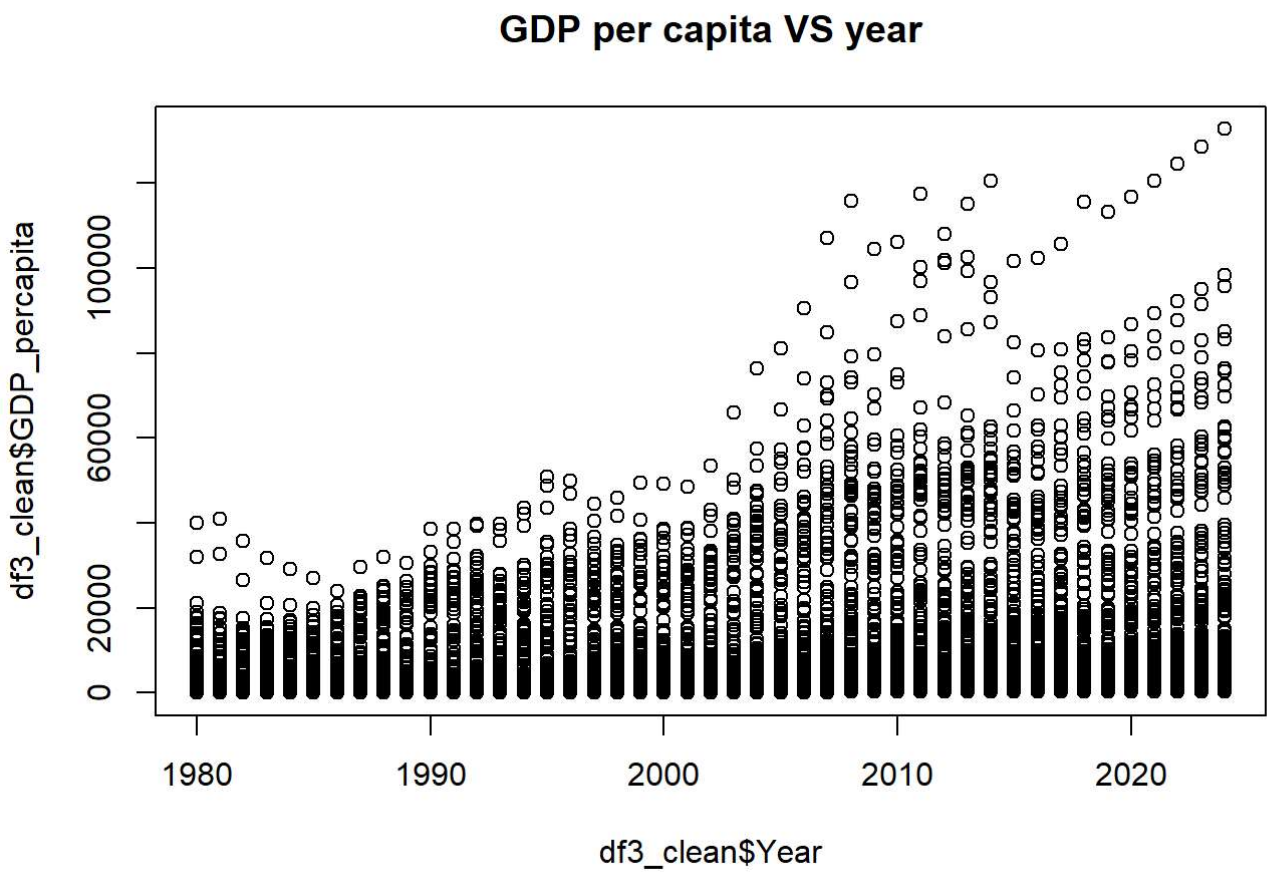
```
#IT SHOWS THAT THERE WILL BE A GRADUAL INCREASE IN MEAN GDP PER CAPITA WHICH SHOWS THAT THERE ARE SOME GOOD DEVELOPMENTS HAPPENING IN MANY COUNTRIES

df5 <- head(df4,15)
```

VISUALISATION ON THE DATA SET TAKEN. VISUALIZING USING HISTOGRAM AND SCATTER PLOT AND INFERING THE RESULTS

```
hist(df3_clean$GDP_percapita, main="GDP per capita")
```



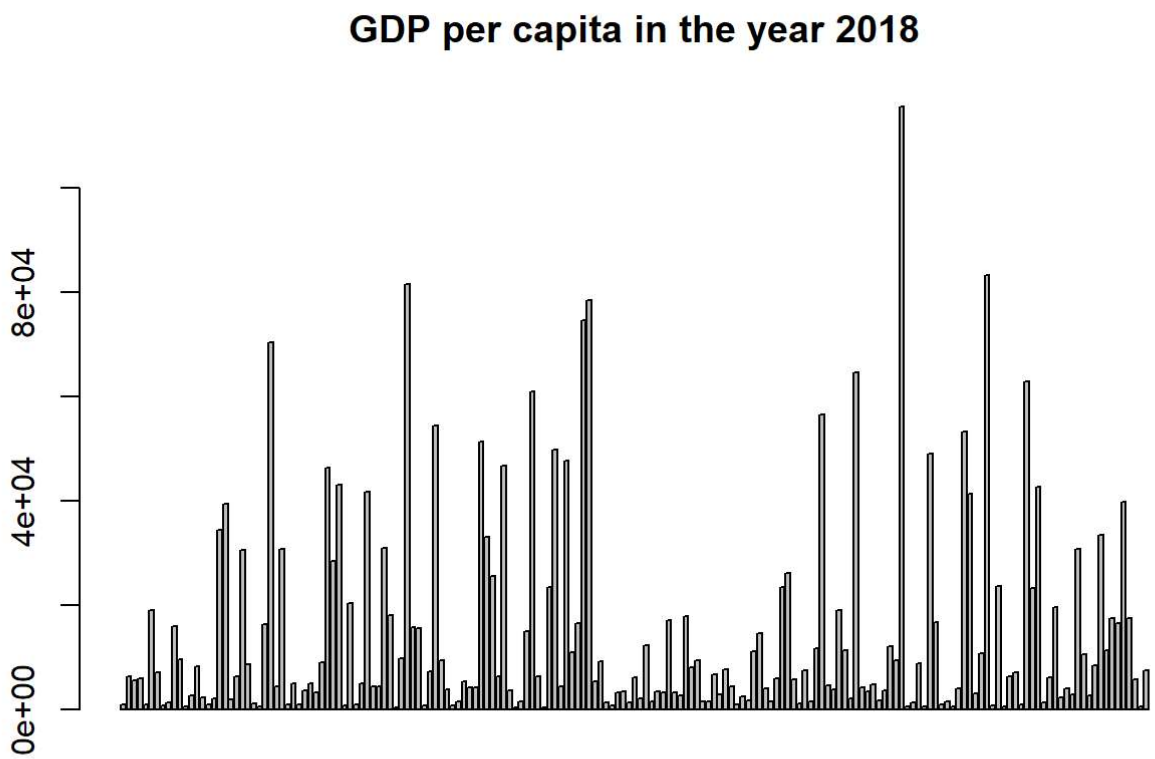

```
#THERE ARE MANY COUNTRIES UNDER 10000 GDP PER CAPITA AND FOLLOWED BY FEWER COUNTRIES IN THE R
ANGE OF 20000 TO 80000 WHICH SHOWS THAT MANY COUNTRIES ARE STILL IN DEVELOPING STAGE

plot(y=df3_clean$GDP_percapita,x=df3_clean$Year,main="GDP per capita VS year")
```



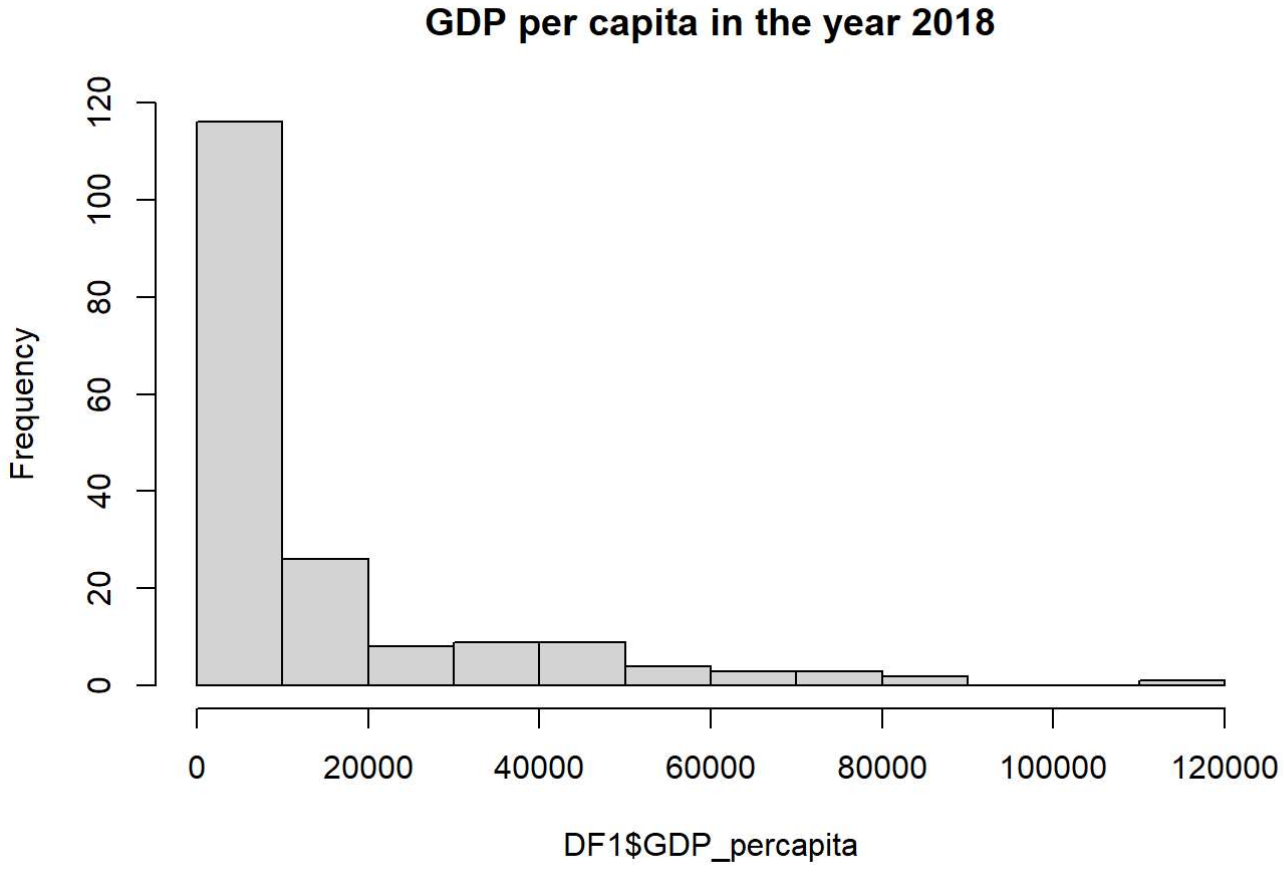

```
#UNDER THE RANGE OF 10000 GDP PER CAPITA FROM THE YEAR 1980 TO 2020 THERE ARE MANY COUNTRIES
BUT THERE ARE FEWER COUNTRIES HAVING MORE THAN 40000 GDP PER CAPITA AND IS IS INCREASING GRAD
UALLY PER EVERY YEAR THE RECENT GDP PER CAPITA IS HIGHEST AMONG ALL ANY OTHER YEARS
```

VISUALIZING FOR THE YEAR 2018 VISUALIZING USING BARPLOT AND HISTOGRAM AND INFERING THE RESULTS

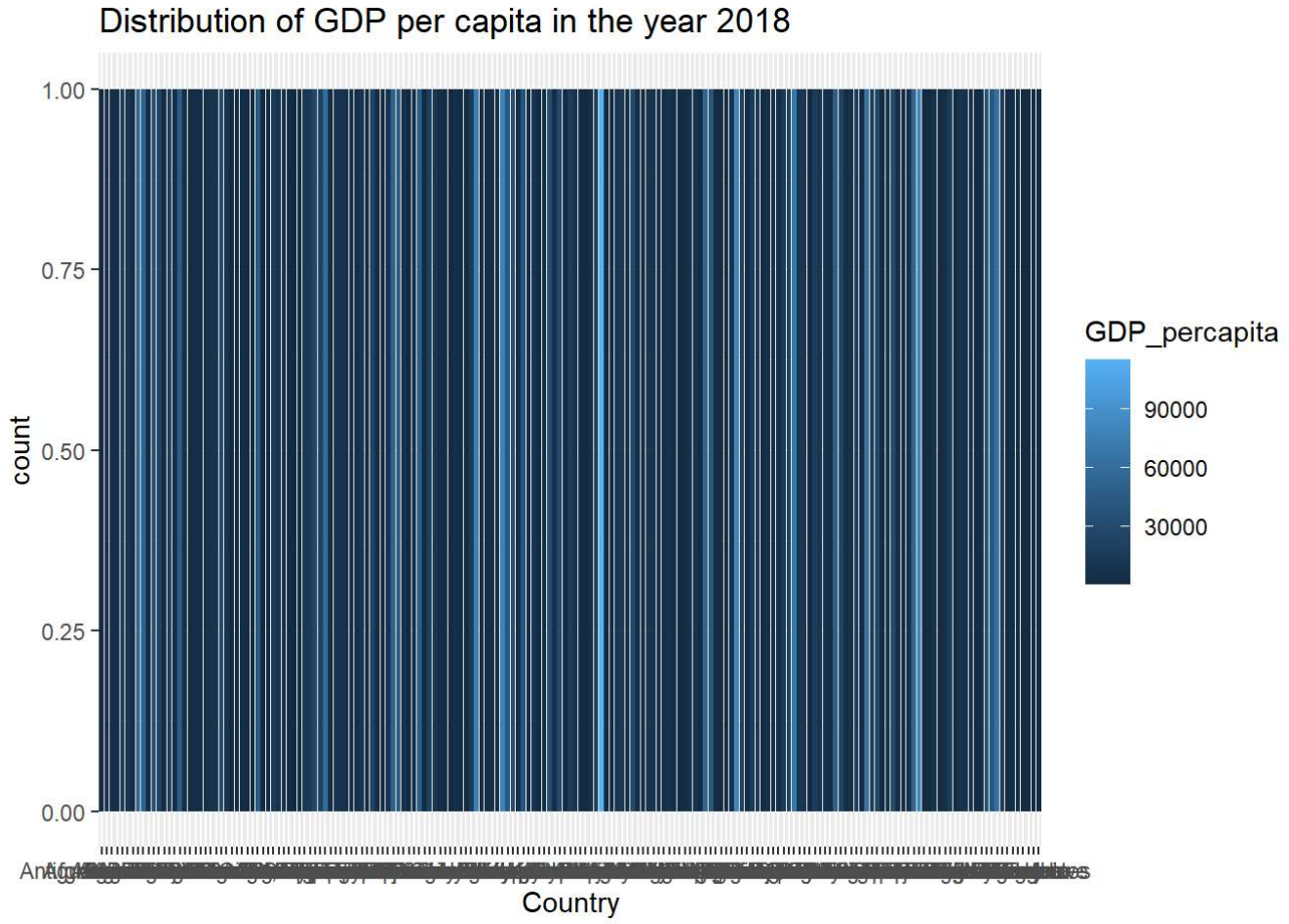
```
#FOR 2018
barplot(DF1$GDP_percapita, main="GDP per capita in the year 2018")
```




```
hist(DF1$GDP_percapita, main="GDP per capita in the year 2018")
```

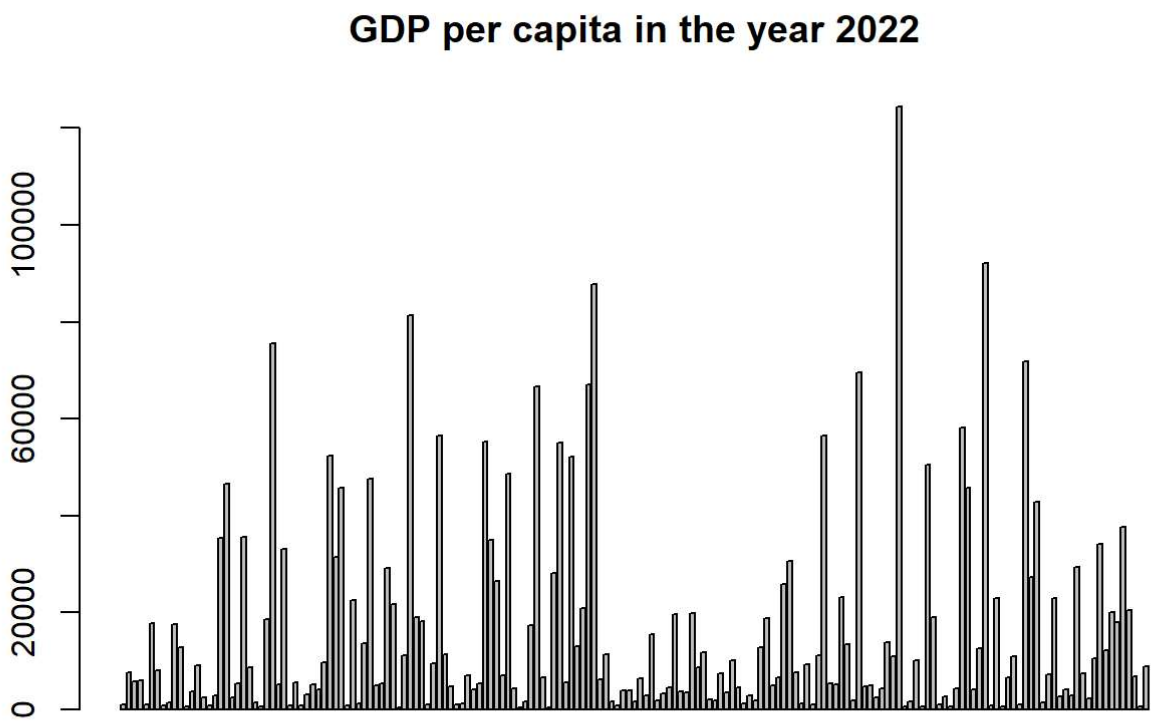


```
#THERE ARE MANY COUNTRIES IN THE RANGE OF 20000 GDP PER CAPITA AND MANY COUNTRIES IN THE RANG
E OF 20000 TO 100000 SHOWS THAT MANY COUNTRIES ARE NOT RICH AND NOT DEVELOPED

ggplot(DF1,aes(x=Country))+geom_bar(aes(fill=GDP_percapita))+ggtitle("Distribution of GDP per
capita in the year 2018")
```

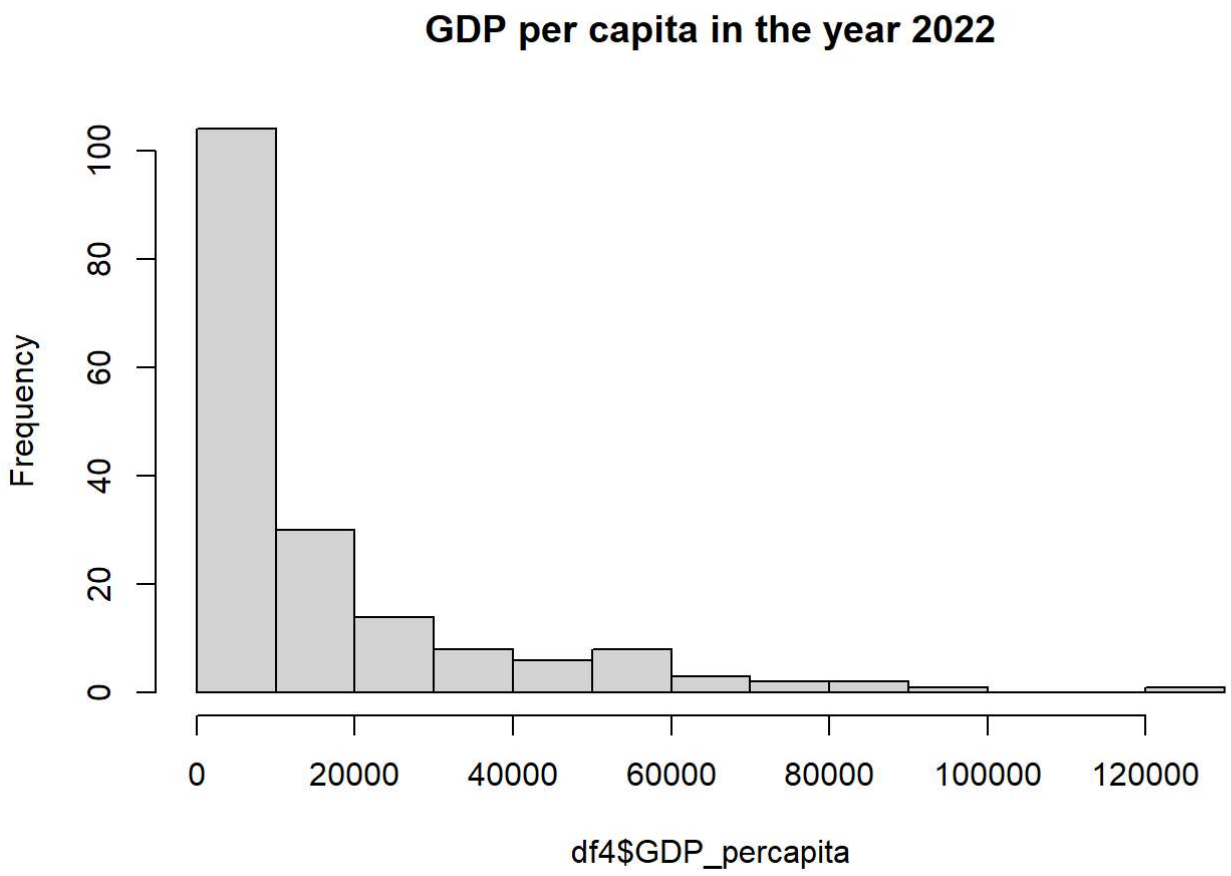


VISUALIZING FOR THE YEAR 2022. VISUALIZING USING BARPLOT AND HISTOGRAM AND INFERING THE RESULTS.

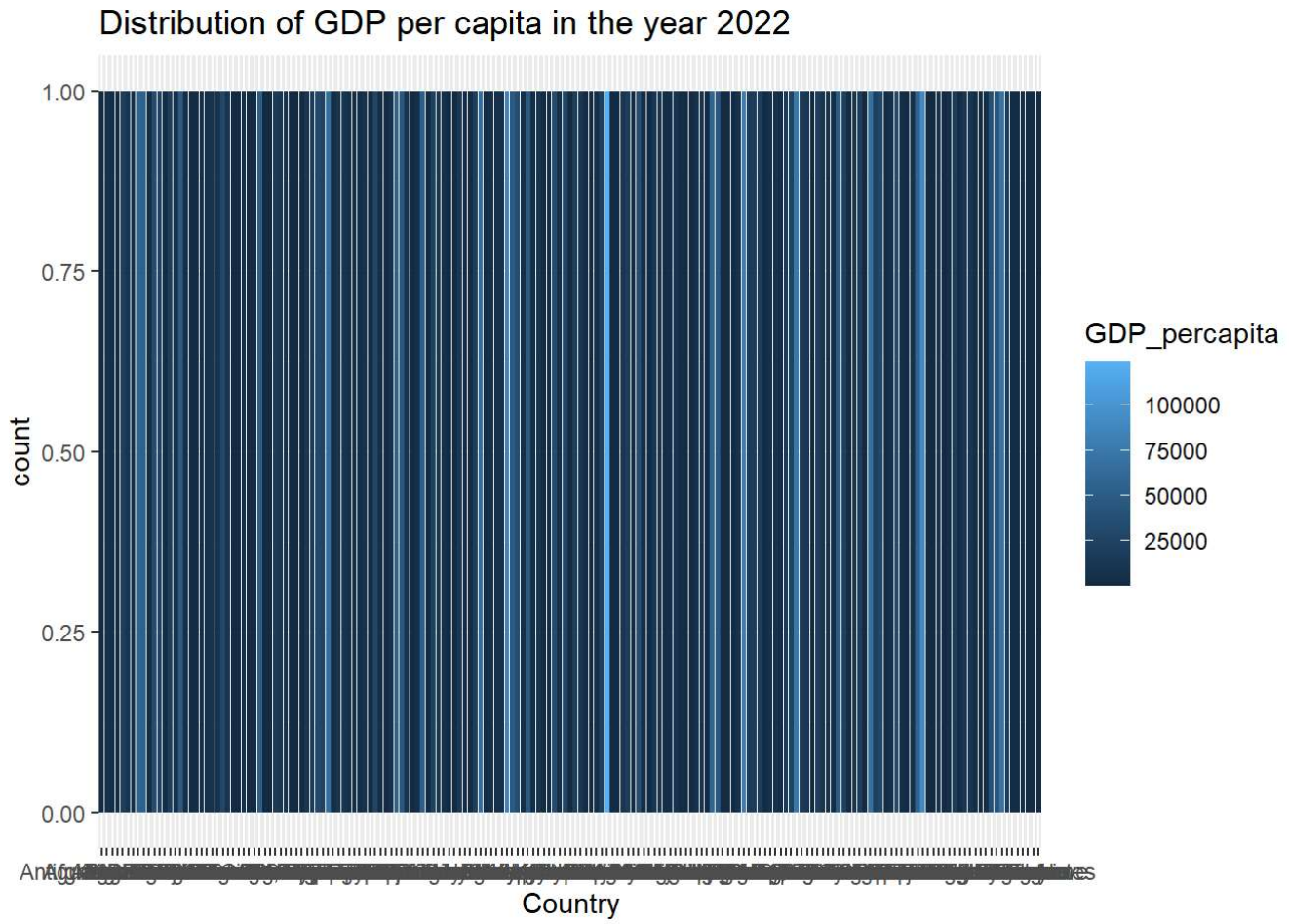
```
#FOR YEAR 2022
barplot(df4$GDP_percapita, main="GDP per capita in the year 2022")
```



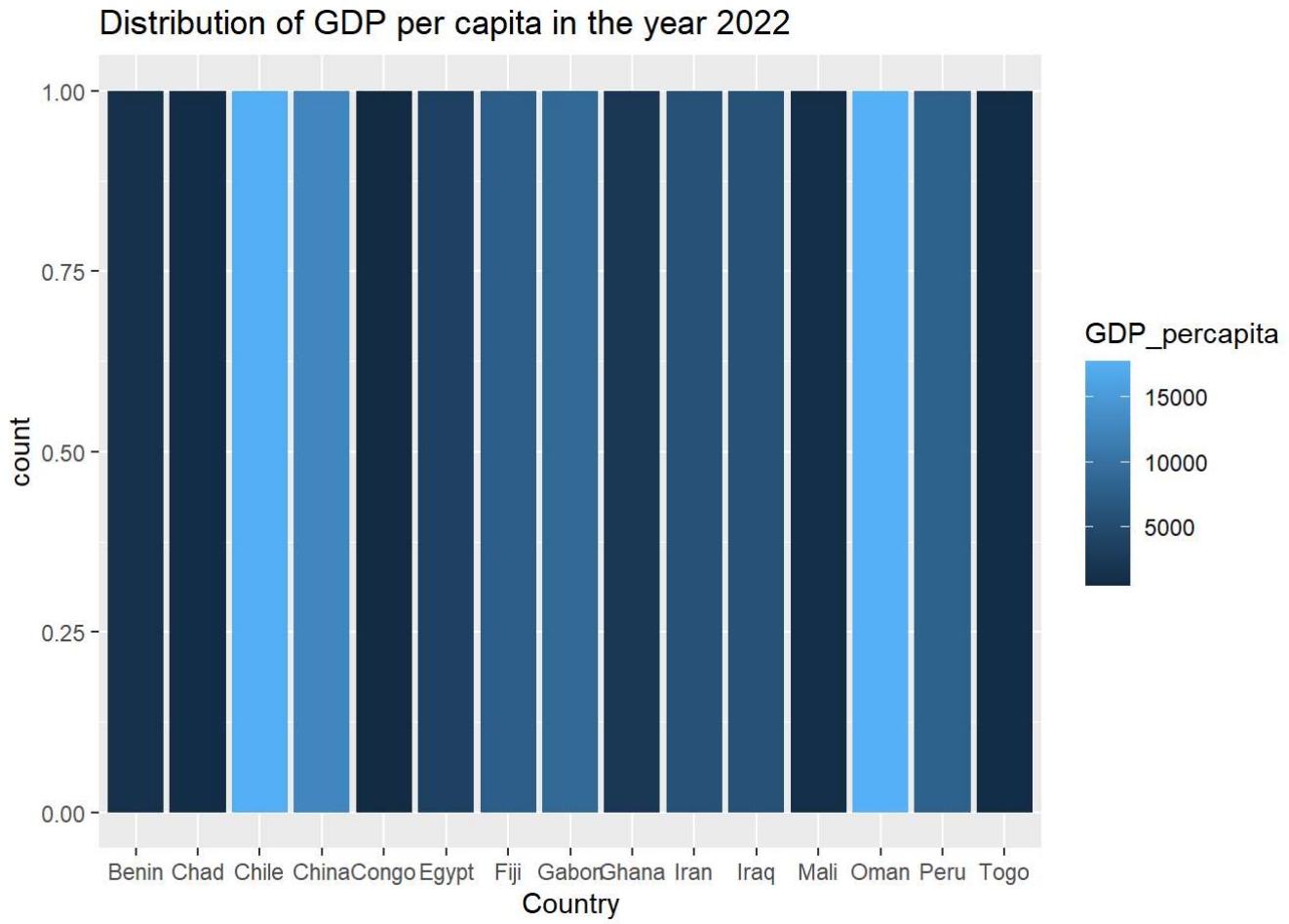

```
#RECENTLY THERE IS GRADUAL INCREASE IN GDP PER CAPITA SOME COUNTRIES ARE ABOVE 20000 WHICH SH
OWS THAT SOME DEVELOPMENTS HAVE HAPPENED IN RECENT YEARS

hist(df4$GDP_percapita, main="GDP per capita in the year 2022")
```




```
#IT SHOWS THAT MANY COUNTRIES ARE IN THE RANGE 0 TO 20000 AND FROM 20000 TO 100000 RANGE SOME COUNTRIES ARE HAVING GOOD GDP PER CAPITA

ggplot(df4,aes(x=Country))+geom_bar(aes(fill=GDP_percapita))+ggtitle("Distribution of GDP per capita in the year 2022")
```



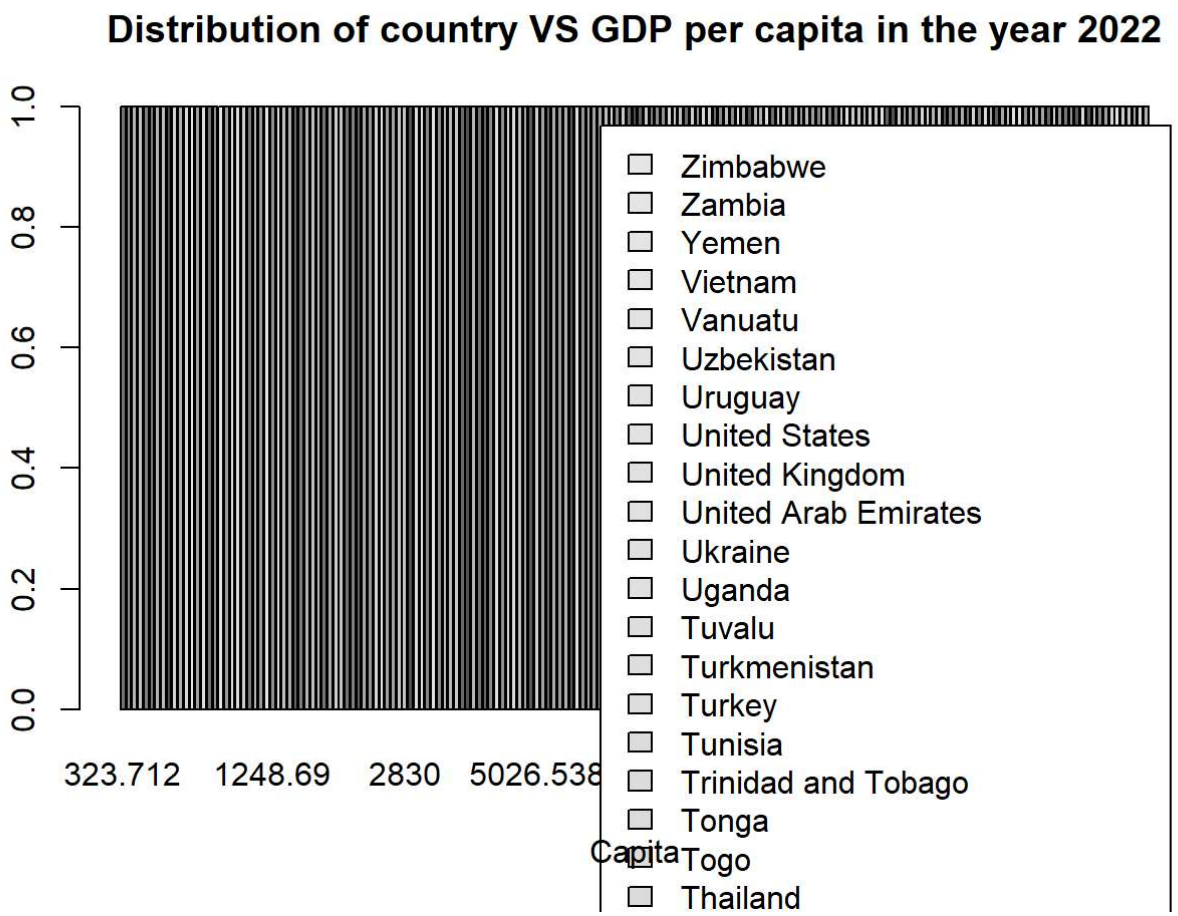

```
#AS YOU CAN SEE THAT THERE ARE FEWER LINES HAVING LIGHT BLUE COLOR SHOWING THAT ONLY SOME COU
NTRIES ARE HAVING HIGHEST GDP PER CAPITA

ggplot(df5,aes(x=Country))+geom_bar(aes(fill=GDP_percapita))+ggtitle("Distribution of GDP per
capita in the year 2022")
```




```
#TAKING THE FIRST 10 ROWS AND VISUALIZING THEM USING BAR PLOT. THERE ARE 2 LINES HAVING LIGHT BLUE WHICH SHOWS THAT THEY ARE HAVING HIGH GDP PER CAPITA

max(df5$GDP_percapita)
```

```
## [1] 17812.76
```

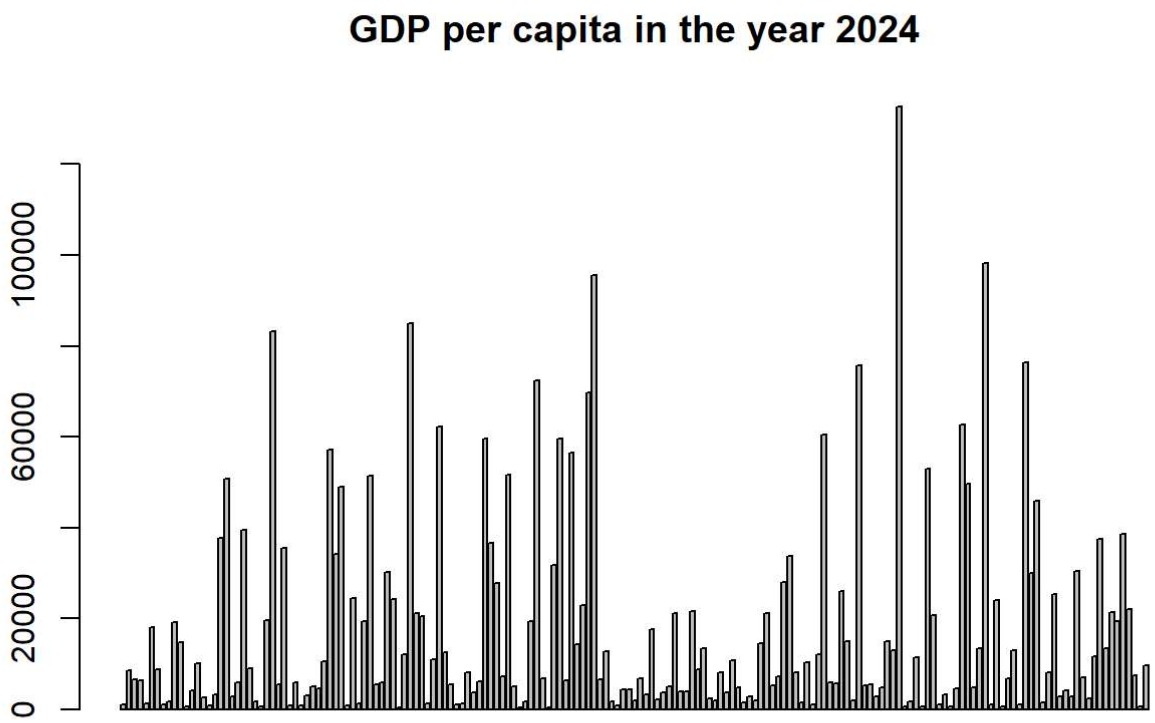
```
counts <- table(df4$Country, df4$GDP_percapita)
barplot(counts, main="Distribution of country VS GDP per capita in the year 2022",xlab="Capita",legend = rownames(counts))
```



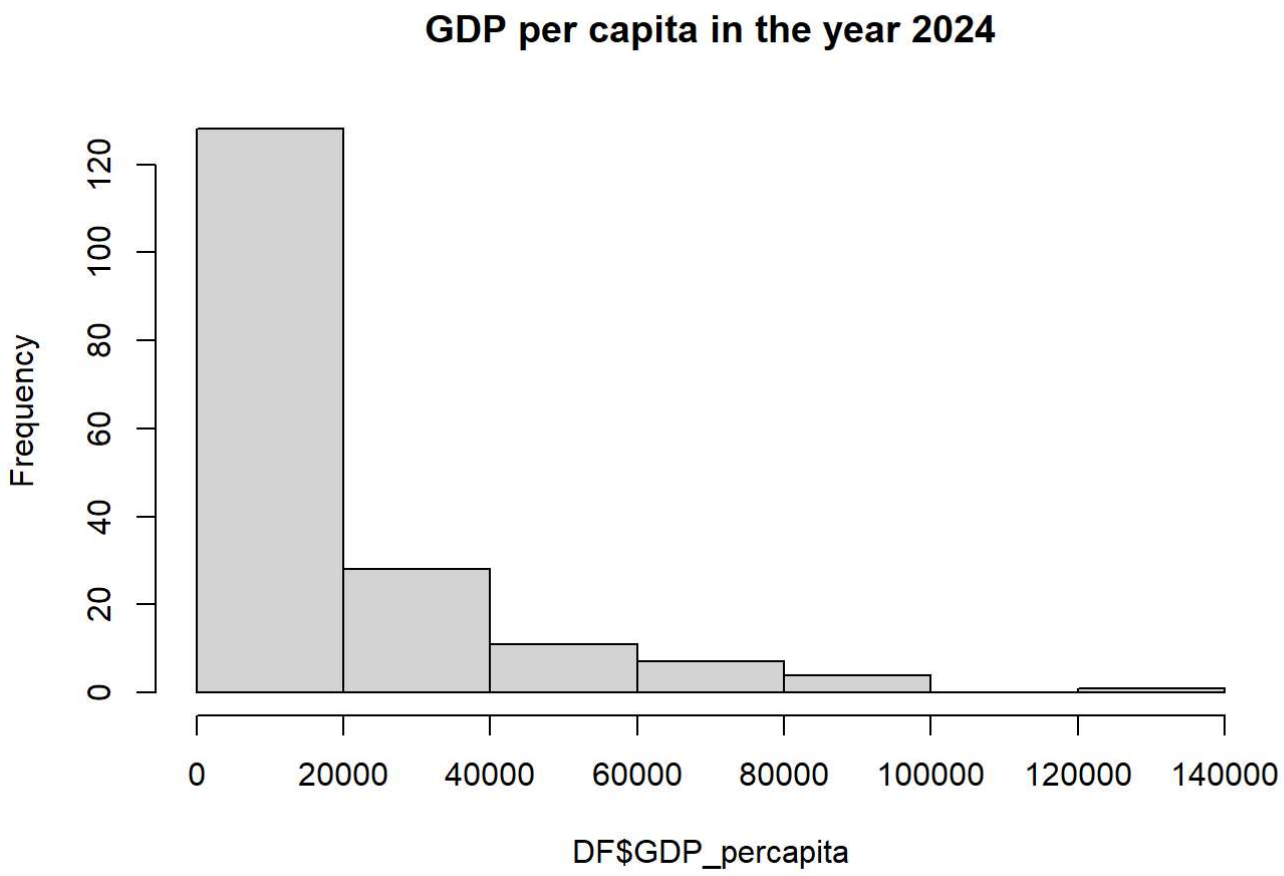

```
#SHOWS THE DISTRIBUTION OF GDP PER CAPITA COUNTRYWISE
```

PREDICTING WHAT WILL HAPPEN IN 2024 USING THE PREVIOUS DATA.

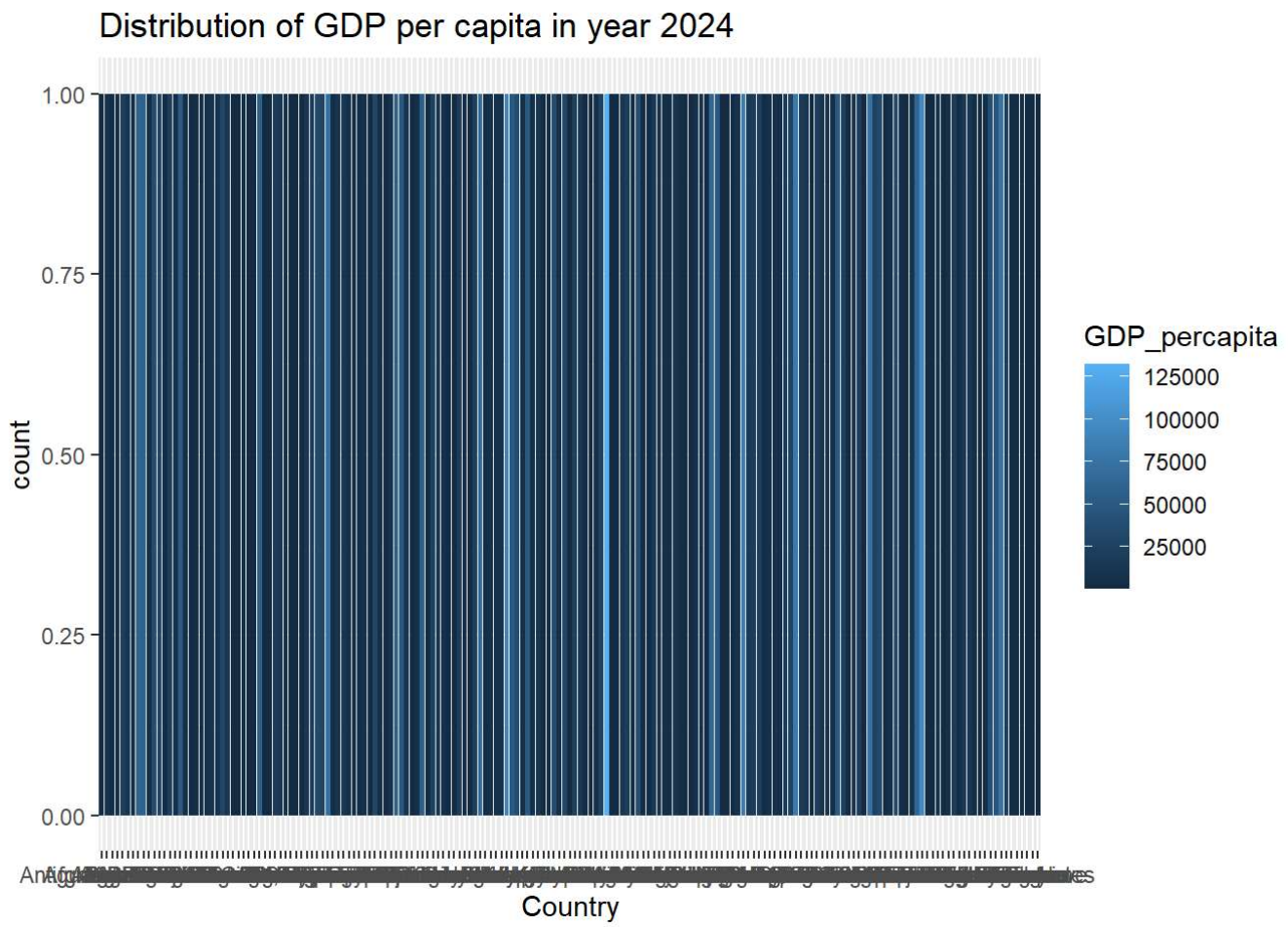
```
#FOR 2024
barplot(DF$GDP_percapita, main="GDP per capita in the year 2024")
```




```
#WHAT IS PREDICTED FOR 2024 IS THAT IT WILL REACH MAXIMUM GDP PER CAPITA EVER RECORDED IN ANY OTHER YEAR OUT OF 179 COUNTRIES THERE ARE MANY COUNTRIES UNDER 20000 GDP PER CAPITA AND FEWER COUNTRIES IN THE RANGE OF 20000 TO 80000

hist(DF$GDP_percapita, main="GDP per capita in the year 2024")
```




```
#THSI SHOWS THAT THE GDP PER CAPITA STILL WILL BE THE SAME AND MANY COUNTRIES WILL STILL IN B
E UNDER 20000 GDP PER CAPITA AND NOT LIKE THE PREVIOUS YEARS THERE ARE FAR MORE COUNTRIES IN
THE RANGE 20000 TO 100000

ggplot(DF,aes(x=Country))+geom_bar(aes(fill=GDP_percapita))+ggtitle("Distribution of GDP per
capita in year 2024")
```




```
#THIS SHOWS THAT THERE ARE MANY COUNTRIES THAT WILL REMAIN SAME IN THE RANGE OF 20000 GDP PER CAPITA
```

2. READING THE DATASET AND GIVING A DESCRIPTION OF THE DATA SET THIS DATASET IS CLEAN

```
df2 <- read.csv('GDP_World_10yr_weightedaverage.csv')
dim(df2)
```

```
## [1] 182   4
```

```
View(df2)
sum(is.na(df2)) #DATA IS CLEAN
```

```
## [1] 0
```

STATISTICAL ANALYSIS SHOWS THAT WHERE EXACTLY IS THE COUNTRY PLACED IN THE WORLDS POSITION.

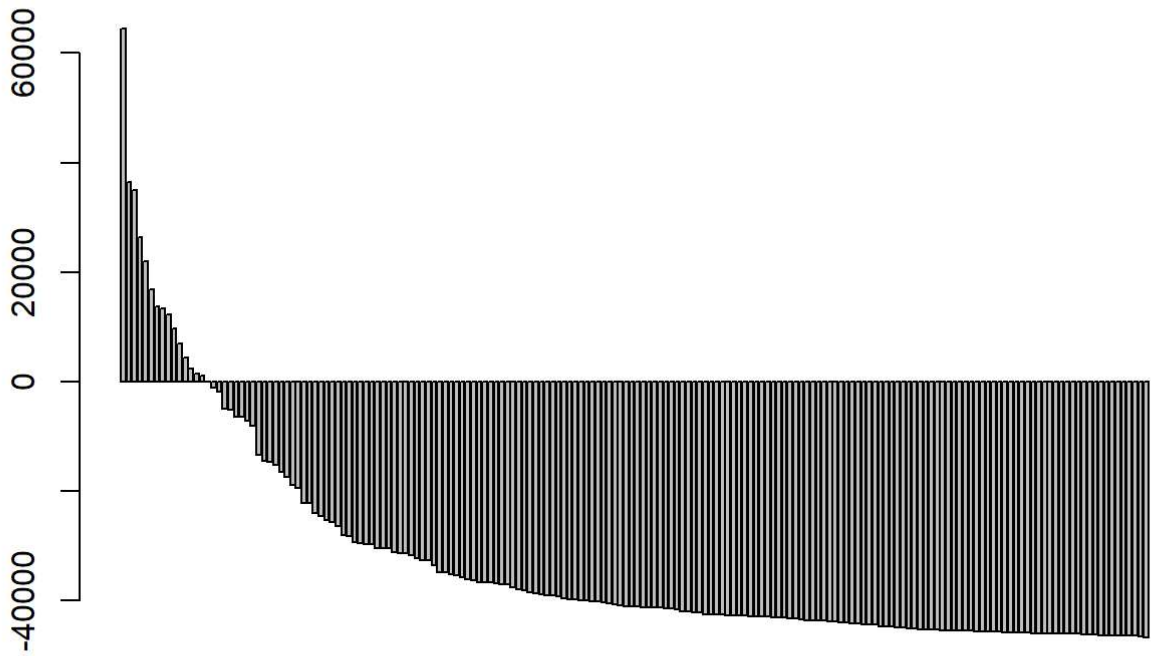
```
summary(df2)
```

```
##    Location         avgGDP_10year      Weighted_avgGDP_10year
##  Length:182         Min.   :   281.4   Min.   :    95.83
##  Class :character   1st Qu.:  2046.6   1st Qu.:  2001.03
##  Mode  :character   Median :  5818.1   Median :  5670.92
##                     Mean   : 13805.6   Mean   : 13938.47
##                     3rd Qu.: 15456.0   3rd Qu.: 16383.27
##                     Max.   :110552.3   Max.   :111232.72
##  WeightAvg_ComparedCanada
##  Min.   :-46666
##  1st Qu.:-44761
##  Median :-41091
##  Mean   :-32823
##  3rd Qu.:-30379
##  Max.   : 64471
```

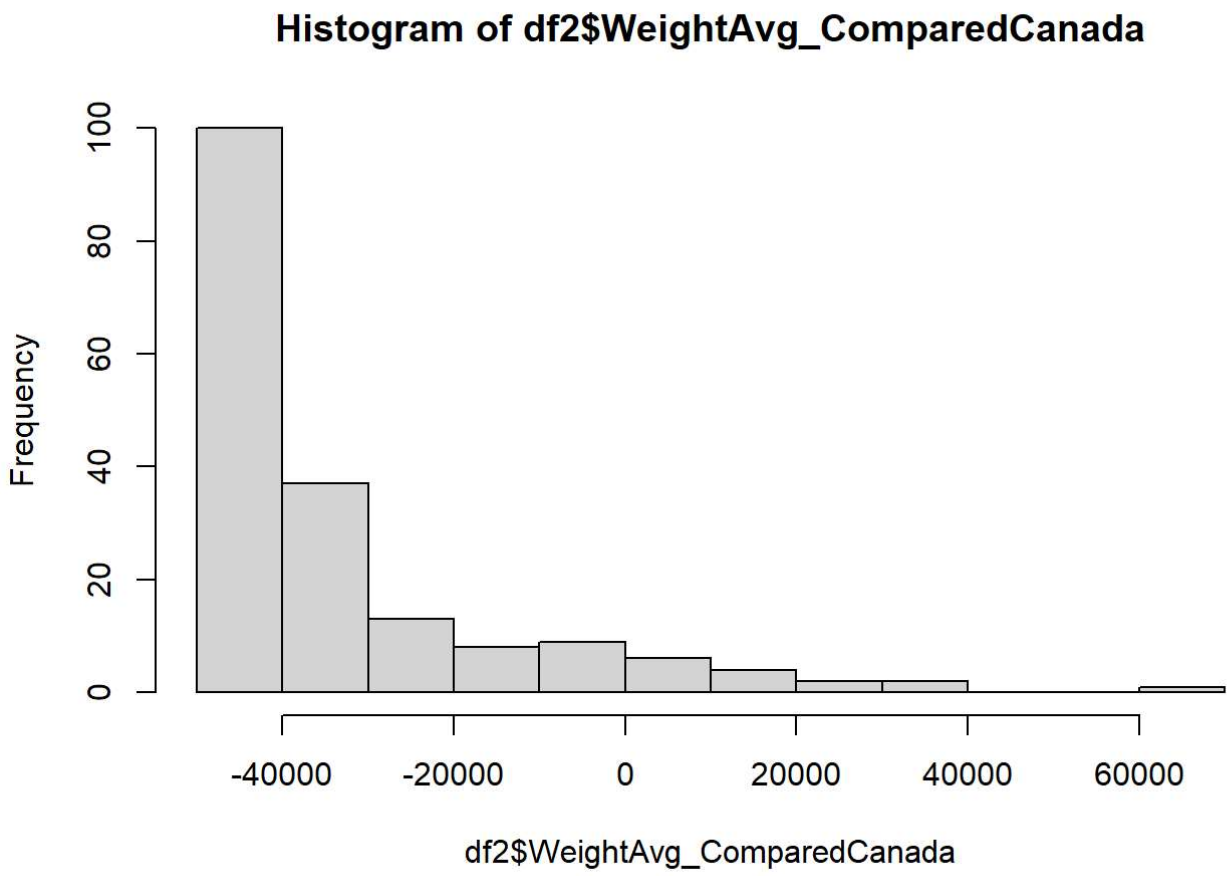
```
#THE MEAN VALUE FOR WEIGHTED AVG GDP FOR 10 YEARS IS 13938 AND THE COUNTRY HAVING IT IS LUXEM
BOURG AND MINIMUM VALUE IS 95 AND THE COUNTRY HAVING IT IS SYRIA SO IT SHOWS THAT THE AVERAGE
OF 10 YEARS SHOWS THAT STILL THERE ARE MANY COUNTRIES UNDER MEAN VALUE
#THE MEAN VALUE FOR WEIGHTED AVG COMPARED TO CANADA IS -32823 NEGATIVE VALUE WHICH TELLS THAT
POPULATION IS GROWING FASTER THAN ITS GDP AND COMPARED TO CANADA MANY COUNTRIES ARE NOT DEVEL
OPED AND THEY ARE NOT RICH COUNTRIES
```

VISUALIZATION ANALYSIS USING BARPLOT, HISTOGRAM, LINE PLOT AND SCATTER PLOT.

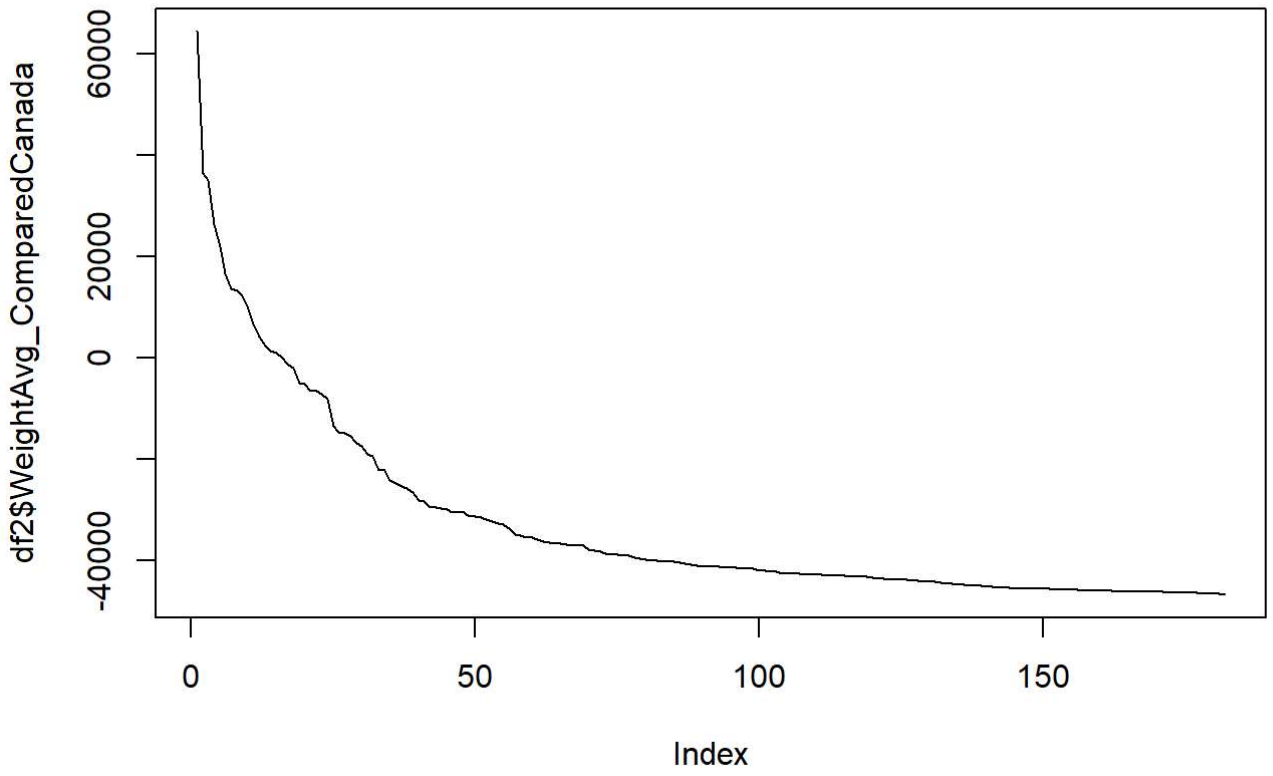
```
barplot(df2$WeightAvg_ComparedCanada)
```



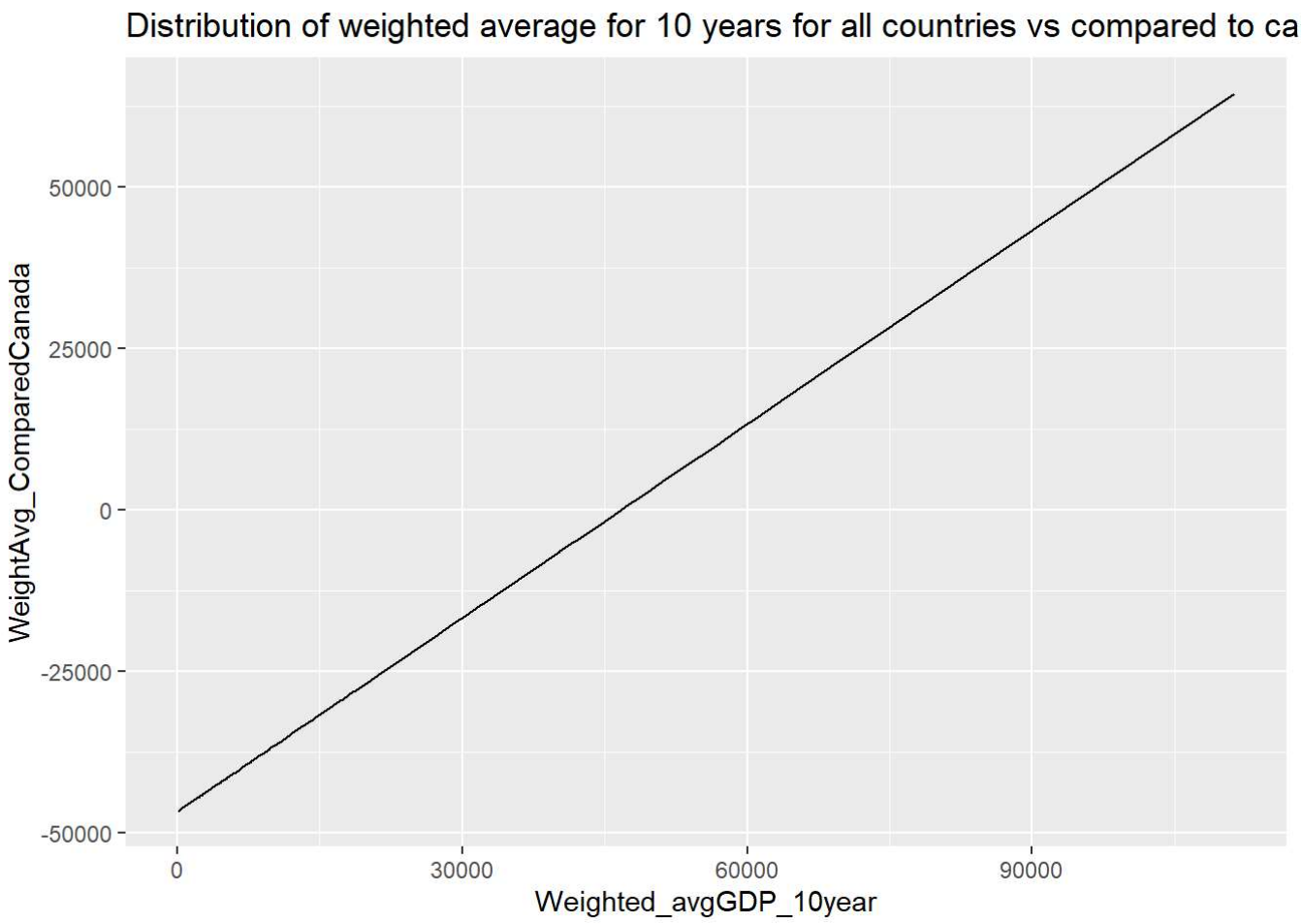
```
#FROM THE PLOT WE CAN SEE THAT MANY COUNTRIES ARE HAVING NEGATIVE VALUES SHOWING THAT COMPARE
D TO CANADA MANY COUNTRIES ARE NOT RICH AND NOT DEVELOPED

hist(df2$WeightAvg_ComparedCanada)
```



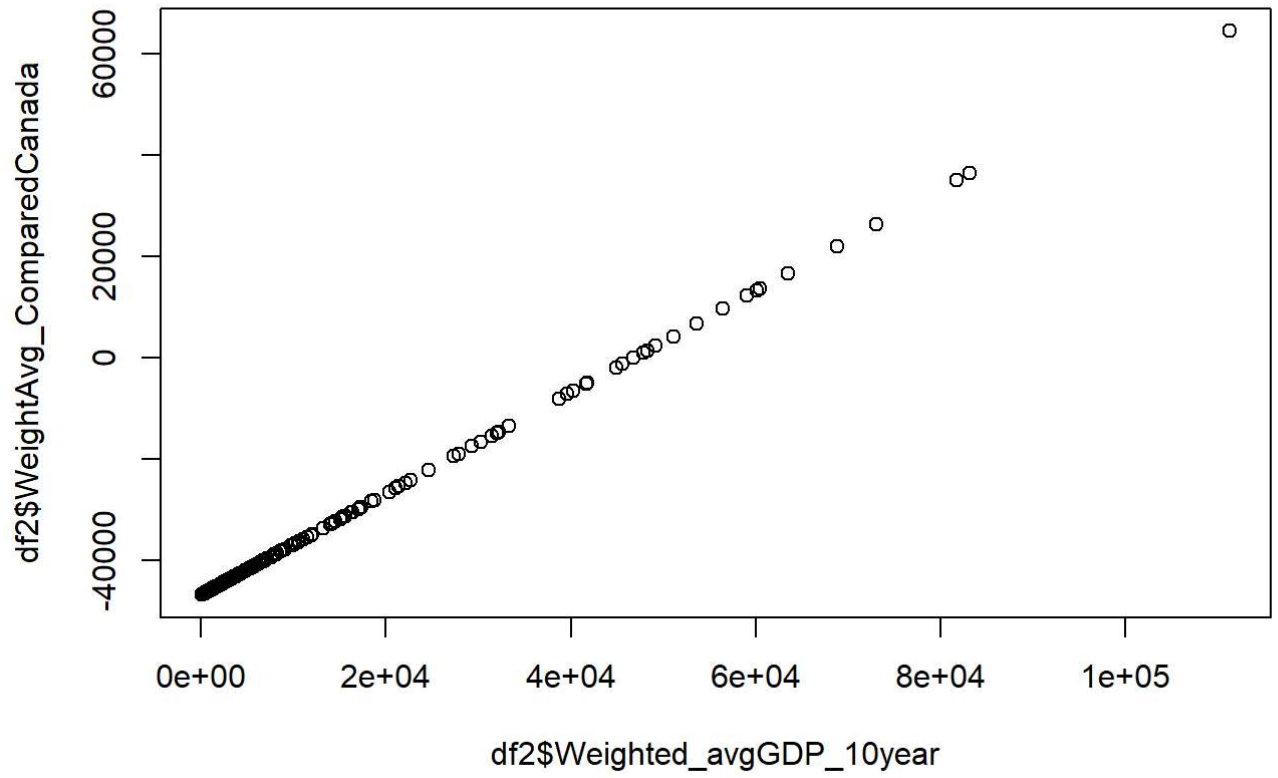

```
#THE DISTRIBUTION SHOWS THAT THERE ARE MANY VALUES IN RANGE -46666 TO -40000 AND MANY MORE CO
UNTRIES IN THE RANGE -40000 TO 0 COMPARED TO CANADA

plot(df2$WeightAvg_ComparedCanada,type="l",col=1)
```




```
#THE GRAPH IS GRADUALLY INCREASING UPTO 60000 BUT THERE ARE VALUES IN THE GRAPH HAVING NEGATI
VE GDP VALUES SHOWING THAT THE POPULATION IS GROWING FASTER COMPARED TO CANADA

ggplot(df2,aes(x=Weighted_avgGDP_10year,y=WeightAvg_ComparedCanada))+geom_line()+ggtitle("Dis
tribution of weighted average for 10 years for all countries vs compared to canada")
```




```
#THIS PLOT SHOWS THAT THE WEIGHTED AVG GDP FOR 10 YEARS AND WHEN COMPARED TO CANADA GIVES A S
TRAIGHT LINE SHOWING THAT THE GDP IS GROWING CONSISTENTLY WITH COMPARED TO CANADA

plot(df2$Weighted_avgGDP_10year,df2$WeightAvg_ComparedCanada,col=1)
```




```
#GIVES A SCATTER PLOT SHOWING THAT THE DISTRIBUTION BETWEEN THEM IS CONSISTENT
```

3. READING THE DATASET AND GIVING THE DESCRIPTION OF THE DATASET THERE ARE NA VALUES SO CLEANING IT USING NA.OMIT METHOD

```
df <- read.csv('stats_expenseindicators_canada_province.csv')
dim(df)
```

```
## [1] 32656     7
```

```
View(df)

sum(is.na(df)) #DATA IS NOT CLEAN
```

```
## [1] 4327
```

```
df <- df[complete.cases(df),] #CLEANING THE DATA
sum(is.na(df)) #DATA IS CLEAN
```

```
## [1] 0
```

STATISTICAL ANALYSIS

```
summary(df)
```

```
##   X_REF_DATE_       GEO               DGUID            Statistic
##  Min.   :2010   Length:28329       Length:28329       Length:28329
##  1st Qu.:2011   Class :character   Class :character   Class :character
##  Median :2013   Mode  :character   Mode  :character   Mode  :character
##  Mean   :2013
##  3rd Qu.:2015
##  Max.   :2017
##  Expenditures           UOM                VALUE
##  Length:28329       Length:28329       Min.   :  -193
##  Class :character   Class :character   1st Qu.:    68
##  Mode  :character   Mode  :character   Median :   203
##                                        Mean   :  1357
##                                        3rd Qu.:   689
##                                        Max.   :110024
```

```
str(df)
```

```
## 'data.frame':    28329 obs. of  7 variables:
##  $ X_REF_DATE_ : int  2010 2011 2012 2013 2014 2015 2016 2017 2010 2011 ...
##  $ GEO         : chr  "Canada" "Canada" "Canada" "Canada" ...
##  $ DGUID       : chr  "2016A000011124" "2016A000011124" "2016A000011124" "2016A000011124"
...
##  $ Statistic   : chr  "Average expenditure per household" "Average expenditure per househo
ld" "Average expenditure per household" "Average expenditure per household" ...
##  $ Expenditures: chr  "Total expenditure" "Total expenditure" "Total expenditure" "Total e
xpenditure" ...
##  $ UOM         : chr  "Dollars" "Dollars" "Dollars" "Dollars" ...
##  $ VALUE       : int  72075 73646 75695 79098 80727 82697 84489 86070 54013 55227 ...
```

```
DF1 <- unique(df$Expenditures)
#GIVES THE UNIQUE VALUES OF THE EXPENDITURES AND WE WILL FORMAT IT AS TABLE AND CONVERT TO DA
TAFRAME
DF <- table(df$Expenditures)
DF <- data.frame(DF)

#TAKING A SAMPLE OF 10 VALUES TO MAKE INFERENCE
lf <- sample_n(DF,10)
lf
```

```
##                                                                                Var1
## 1                                                                Household furnishings
## 2                                                                            Education
## 3                                                                     Restaurant meals
## 4                                                                  Automobiles (purchase)
## 5                                                               Total current consumption
## 6                                 Commissions for sale of real estate owned by the household
## 7                                              Water and sewage for principal accommodation
## 8  Non-prescribed medicines, pharmaceutical products, health care supplies and equipment
## 9                                                 Accident or disability insurance premiums
## 10                                                                          Personal care
##    Freq
## 1   104
## 2   104
## 3   104
## 4   104
## 5   104
## 6    75
## 7   104
## 8   103
## 9   104
## 10  104
```

```
cx <- unique(df$GEO)
#RETURNS ONLY UNIQUE VALUES BY REOMVING ALL THE DUPLICATES OF THE EXISTING FACTORS

cx <- data.frame(cx)
j <- 1
cnt <- 0
for(x in 1:dim(df)[1]){
  if(j>dim(cx)[1]){
    break
  }
  if (df$GEO[x]==cx$cx[j]){
    cnt <- cnt+df$VALUE[x]
  }
  else{
    print(df$GEO[x])
    print(cnt)
    cnt <- 0
    j <- j+1
  }
}
```

```
## [1] "Atlantic Region"
## [1] 1591875
## [1] "Newfoundland and Labrador"
## [1] 1286169
## [1] "Prince Edward Island"
## [1] 1351182
## [1] "Nova Scotia"
## [1] 1244839
## [1] "New Brunswick"
## [1] 1304451
## [1] "Quebec"
## [1] 1226129
## [1] "Ontario"
## [1] 1292024
## [1] "Prairie Region"
## [1] 1605535
## [1] "Canada"
## [1] 1690695
## [1] "Canada"
## [1] 0
## [1] "Canada"
## [1] 0
## [1] "Canada"
## [1] 0
## [1] "Canada"
## [1] 0
```

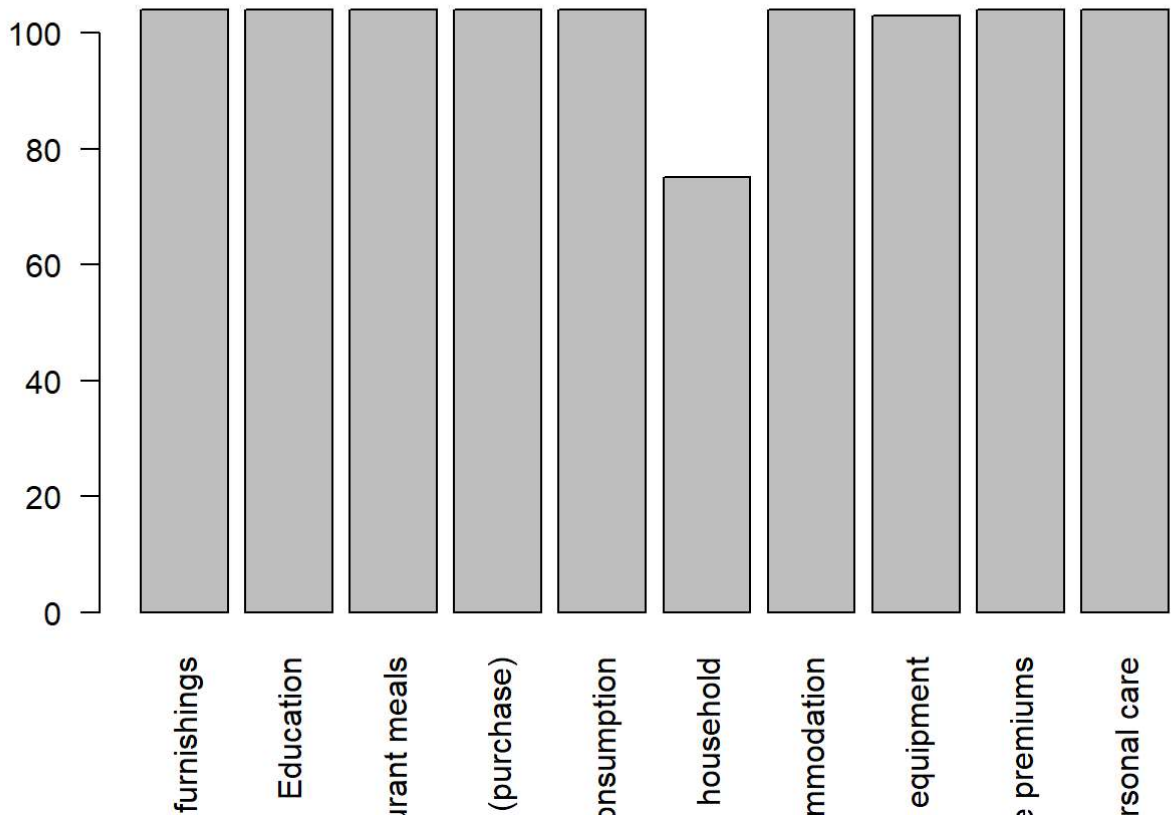
```
#PRINTS GEOGRAPHICAL LOCATIONS WITH THEIR RESPECTIVE DOLLAR SPENT
#CANADA IS SPENDING THE MOST DOLLARS COMPARED TO OTHER REGIONS

cx <- unique(df$Expenditures)
cx <- data.frame(cx)
j <- 1
cnt <- 0
for(x in 1:50){
  if(j>dim(cx)[1]){
    break
  }
  if (df$Expenditures[x]==cx$cx[j]){
    cnt <- cnt+df$VALUE[x]
  }
  else{
    print(df$Expenditures[x])
    print(cnt)
    cnt <- 0
    j <- j+1
  }
}
```

```
## [1] "Total current consumption"
## [1] 634497
## [1] "Food expenditures"
## [1] 415610
## [1] "Food purchased from stores"
## [1] 57534
## [1] "Food purchased from restaurants"
## [1] 40971
## [1] "Shelter"
## [1] 16561
## [1] "Principal accommodation"
## [1] 118721
```

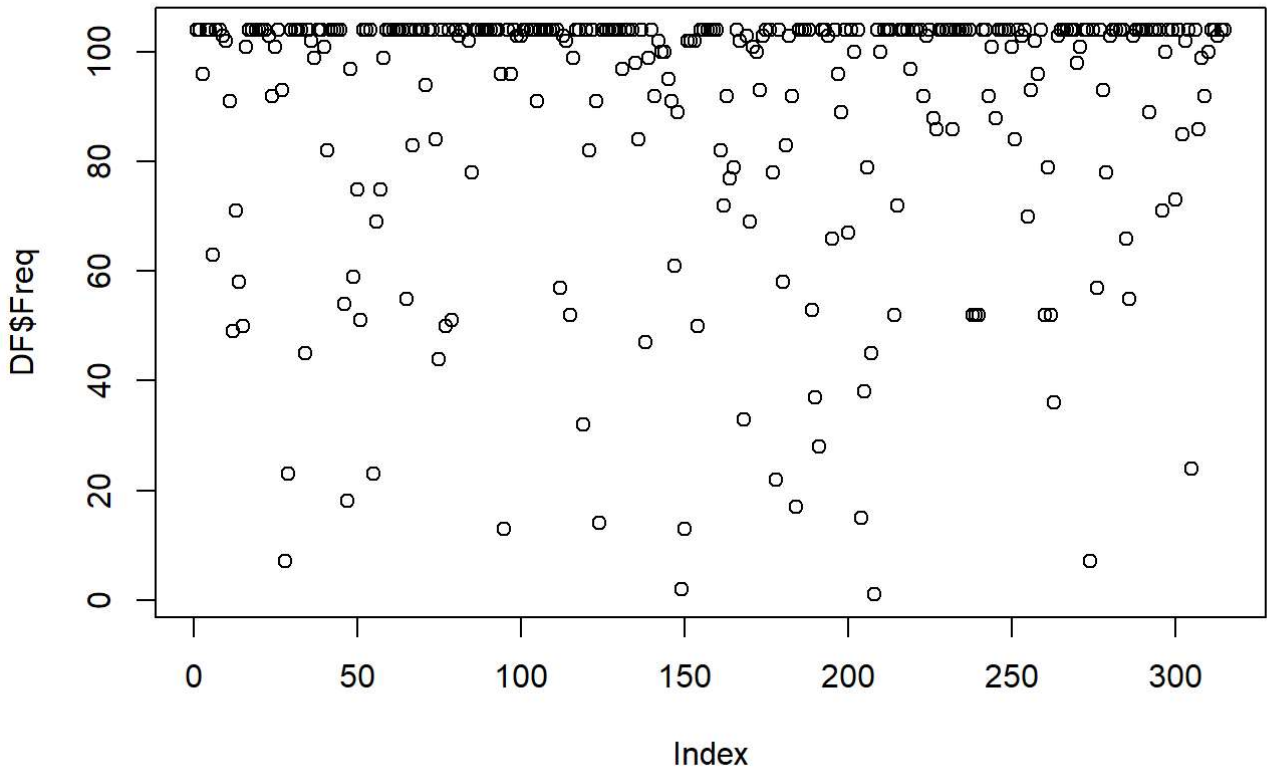
```
#PRINTS EXPENDITURES  WITH THEIR RESPECTIVE DOLLAR SPENT
#MORE DOLLARS ARE SPENT ON CURRENT CONSUMPTION FOLLOWED BY FOOD EXPENDITURES
#IT SHOWS THAT IN CANADA MORE COST IS SPENDING ON CURRENT CONSUMPTION AND LESS COST ON SHELTE
R
```

VISUALIZATION ANALYSIS

```
barplot(lf$Freq,names.arg = lf$Var1,las=2)
```



```
#THE FREQUENCY FOR VARIOUS EXPENDITURES ARE HAVING FREQUENCY AROUND 100 AND PARKING, HANGAR A
ND AIRPORT FEES, BOAT STORAGE AND HARBOUR DUES IS HAVING LOW FREQUENCY

plot(DF$Freq)
```




```
#FROM THE PLOT WE CAN SEE MORE FREQUENCY LIES AROUND 100
```

4.READING THE DATASET AND GIVING THE DESCRIPTION OF THE DATASET THERE ARE NA VALUES SO CLEANING IT USING COMPLETE.CASES METHOD

```
df1 <- read.csv('HealthSpending_byprovince.csv')
dim(df1)
```

```
## [1] 18900     6
```

```
View(df1)

sum(is.na(df1)) #DATA IS HAVING NA VALUES
```

```
## [1] 0
```

```
df <- df[complete.cases(df1),] #CLEANING DATA
sum(is.na(df1)) #DATA IS CLEAN
```

```
## [1] 0
```

STATISTICAL ANALYSIS

```
summary(df1)
```

```
##       Year        Province            Sector          Use.of.Funds
##  Min.   :1975   Length:18900       Length:18900       Length:18900
##  1st Qu.:1986   Class :character   Class :character   Class :character
##  Median :1997   Mode  :character   Mode  :character   Mode  :character
##  Mean   :1997
##  3rd Qu.:2008
##  Max.   :2019
##  Current.dollars    Current.dollars.per.capita
##  Length:18900       Length:18900
##  Class :character   Class :character
##  Mode  :character   Mode  :character
##
##
##
```

```
str(df1)
```

```
## 'data.frame':    18900 obs. of  6 variables:
##  $ Year                     : int  1975 1976 1977 1978 1979 1980 1981 1982 1983 1984 ...
##  $ Province                 : chr  "Newfoundland and Labrador" "Newfoundland and Labrado
r" "Newfoundland and Labrador" "Newfoundland and Labrador" ...
##  $ Sector                   : chr  "Provincial Government" "Provincial Government" "Provi
ncial Government" "Provincial Government" ...
##  $ Use.of.Funds             : chr  "Hospitals" "Hospitals" "Hospitals" "Hospitals" ...
##  $ Current.dollars          : chr  "111,121,750" "123,335,500" "137,070,750" "156,226,25
0" ...
##  $ Current.dollars.per.capita: chr  "200" "219" "242" "275" ...
```

VISUALIZATION ANALYSIS BY TAKING RANDOM VALUES AND PLOTTING IT USING BAR PLOT

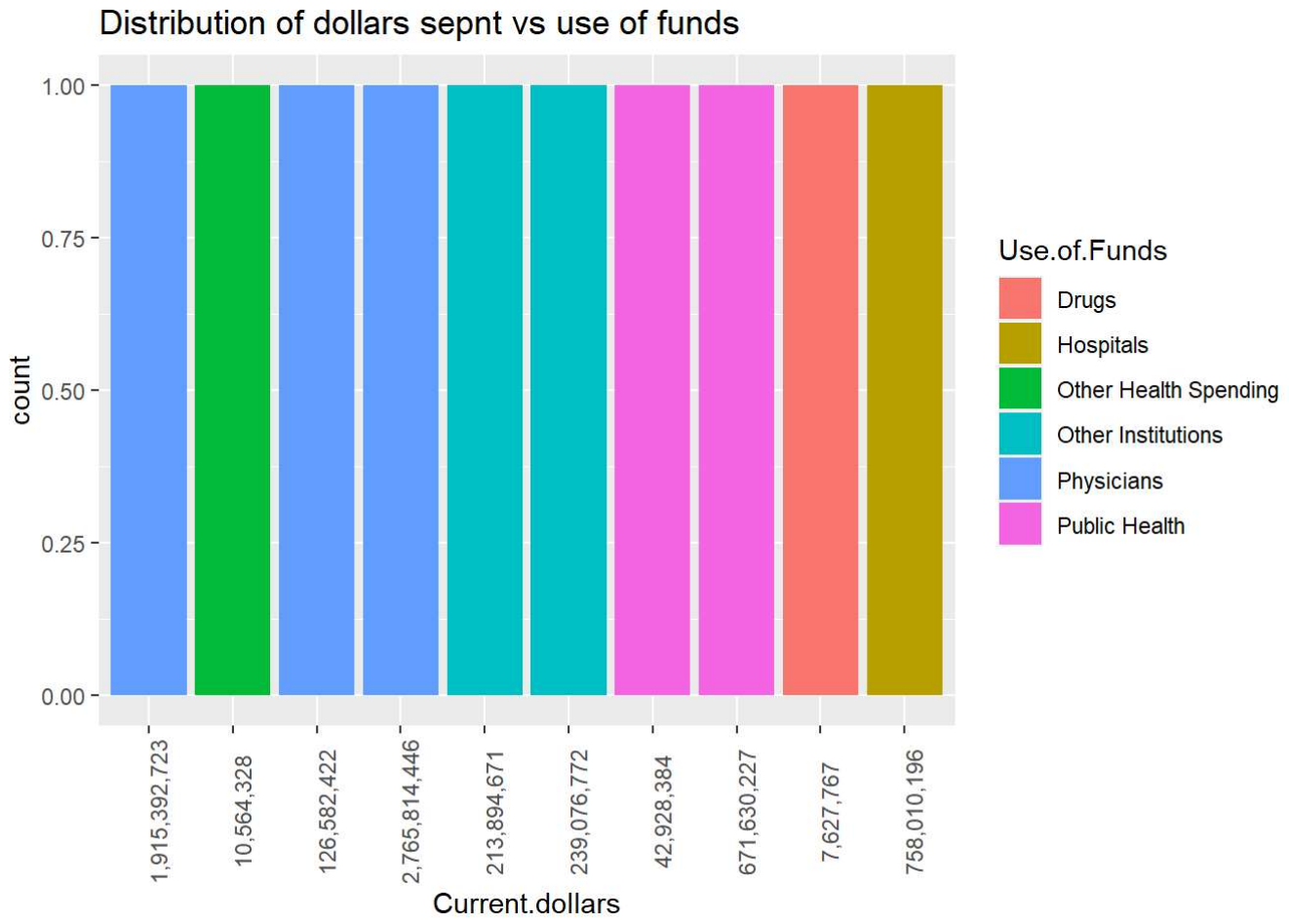
```
l <- unique(df1$Use.of.Funds)
l
```

```
##  [1] "Hospitals"             "Other Institutions"    "Physicians"
##  [4] "Other Professionals"   "Drugs"                 "Capital"
##  [7] "Public Health"         "Administration"        "Other Health Spending"
## [10] "Total"
```

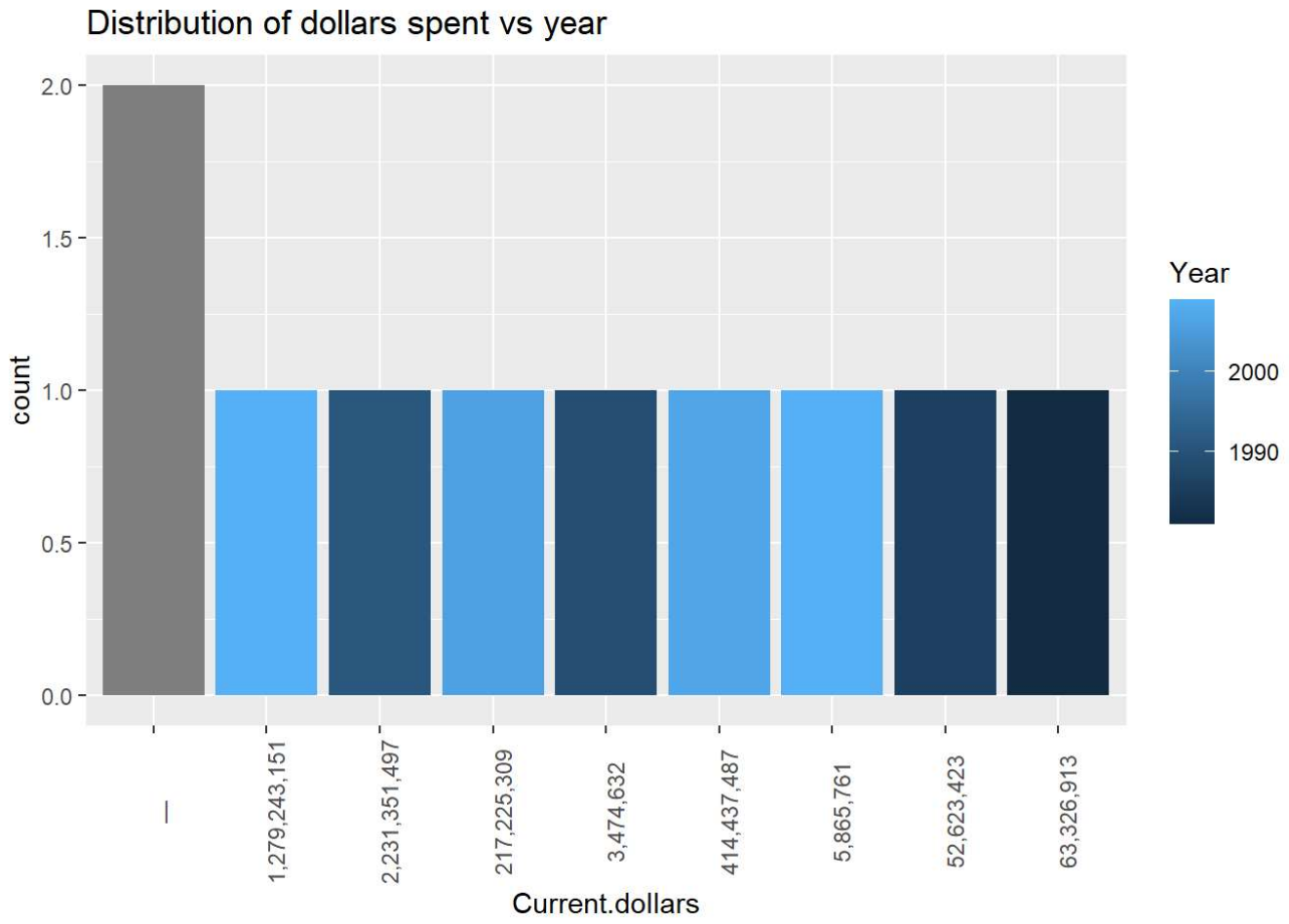
```
#SHOWS WHERE EXACTLY THE FUNDSA ARE BEING USED

lf <- sample_n(df1,size = 10)
lf
```

```
##    Year                  Province                Sector          Use.of.Funds
## 1  1988                   Alberta                Public    Other Institutions
## 2  2001                    Quebec                Public            Physicians
## 3  1999     Northwest Territories               Private                 Drugs
## 4  2005 Newfoundland and Labrador                Public             Hospitals
## 5  2000                     Yukon                Public Other Health Spending
## 6  1987              Saskatchewan                Public    Other Institutions
## 7  2018              Saskatchewan                Public         Public Health
## 8  1983                   Ontario Provincial Government            Physicians
## 9  1990 Newfoundland and Labrador                Public            Physicians
## 10 2019     Northwest Territories Provincial Government         Public Health
##    Current.dollars Current.dollars.per.capita
## 1      239,076,772                         97
## 2    2,765,814,446                        374
## 3        7,627,767                        188
## 4      758,010,196                      1,474
## 5       10,564,328                        347
## 6      213,894,671                        207
## 7      671,630,227                        578
## 8    1,915,392,723                        212
## 9      126,582,422                        219
## 10      42,928,384                        962
```

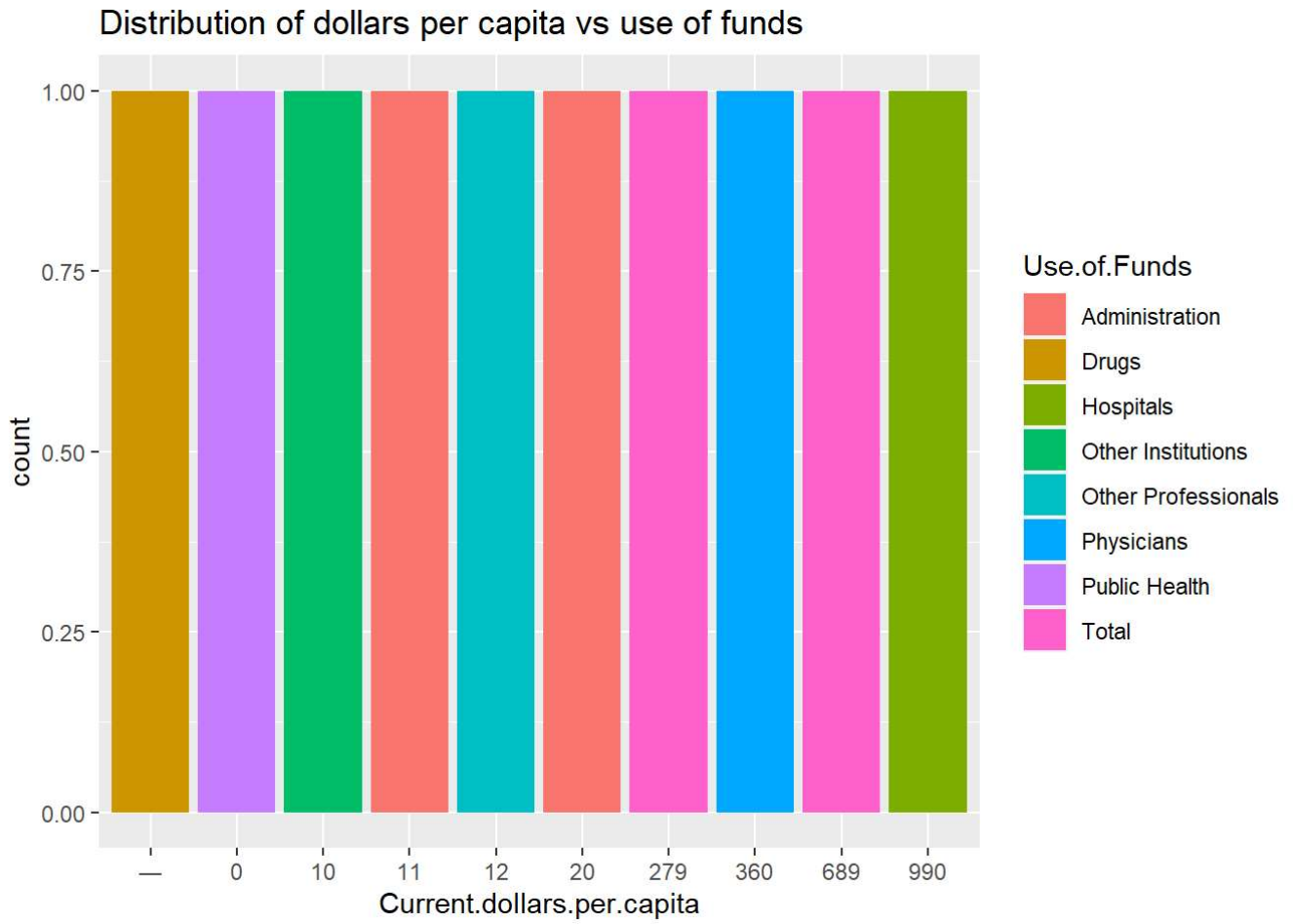
```
ggplot(lf,aes(x=Current.dollars),las=2)+geom_bar(aes(fill=Use.of.Funds))+ggtitle("Distributio
n of dollars sepnt vs use of funds")+theme(axis.text.x = element_text(angle = 90))
```



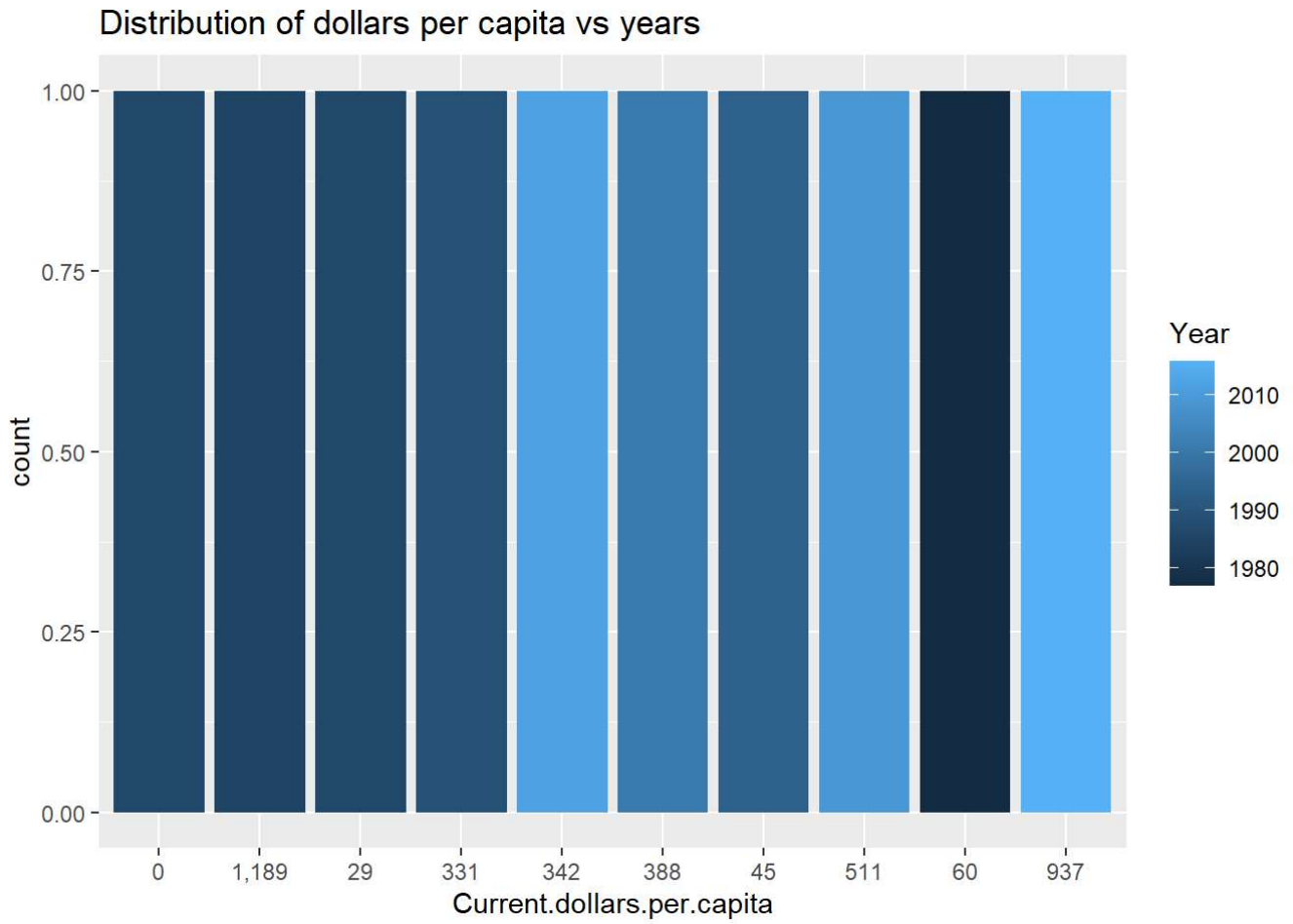

```
#THE PLOT SHOWS THAT HOW MANY DOLLARS IS BEING USED BY EACH SECTOR.


lf <- sample_n(df1,size = 10)
ggplot(lf,aes(x=Current.dollars))+geom_bar(aes(fill=Year))+ggtitle("Distribution of dollars s
pent vs year")+theme(axis.text.x = element_text(angle = 90))
```



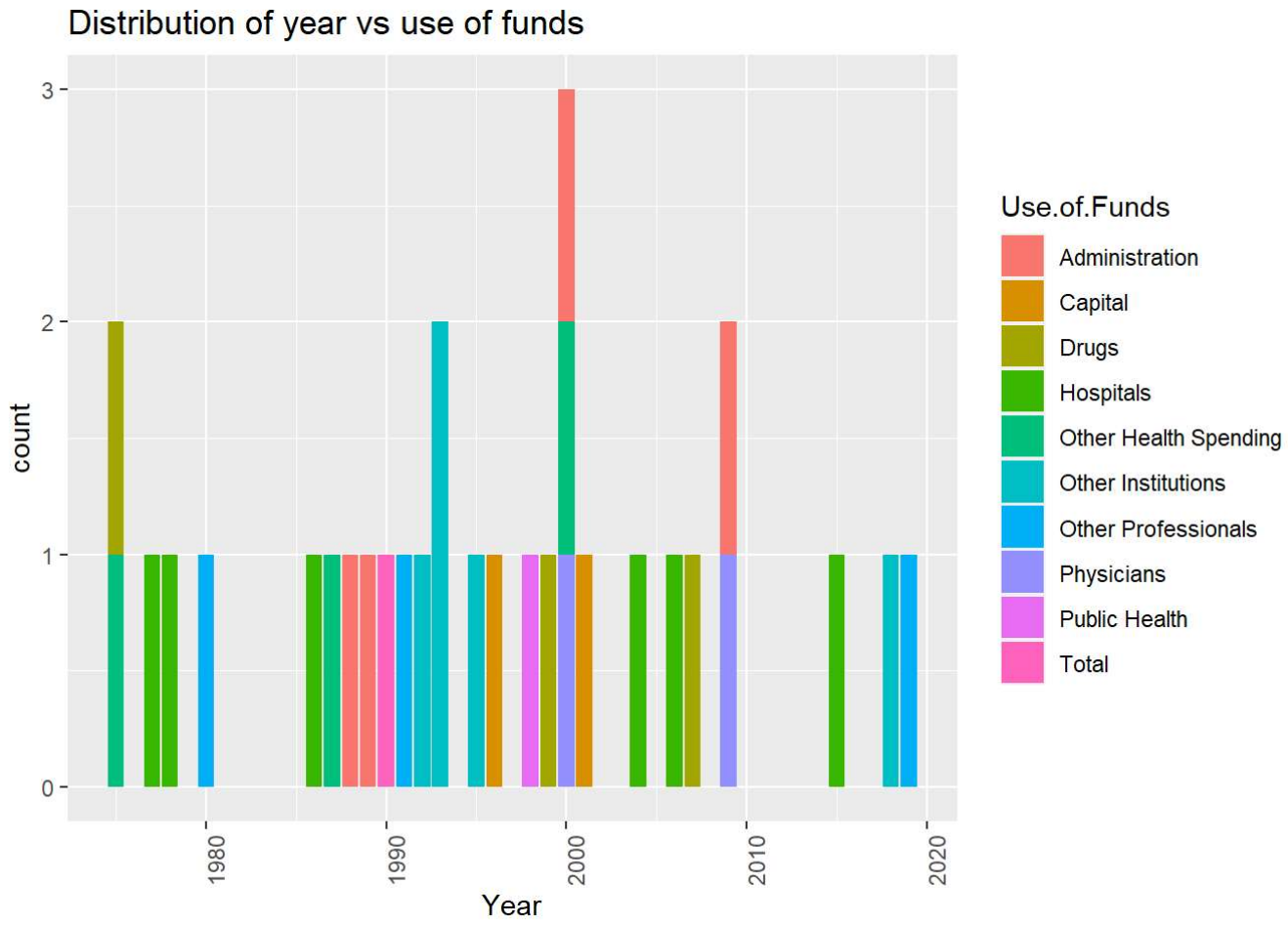

```
#THE PLOT SHOWS THAT IN WHICH YEAR HOW MANY DOLLARS ARE BEING SPENT


lf <- sample_n(df1,size = 10)
ggplot(lf,aes(x=Current.dollars.per.capita),las=2)+geom_bar(aes(fill=Use.of.Funds))+ggtitle
("Distribution of dollars per capita vs use of funds")
```



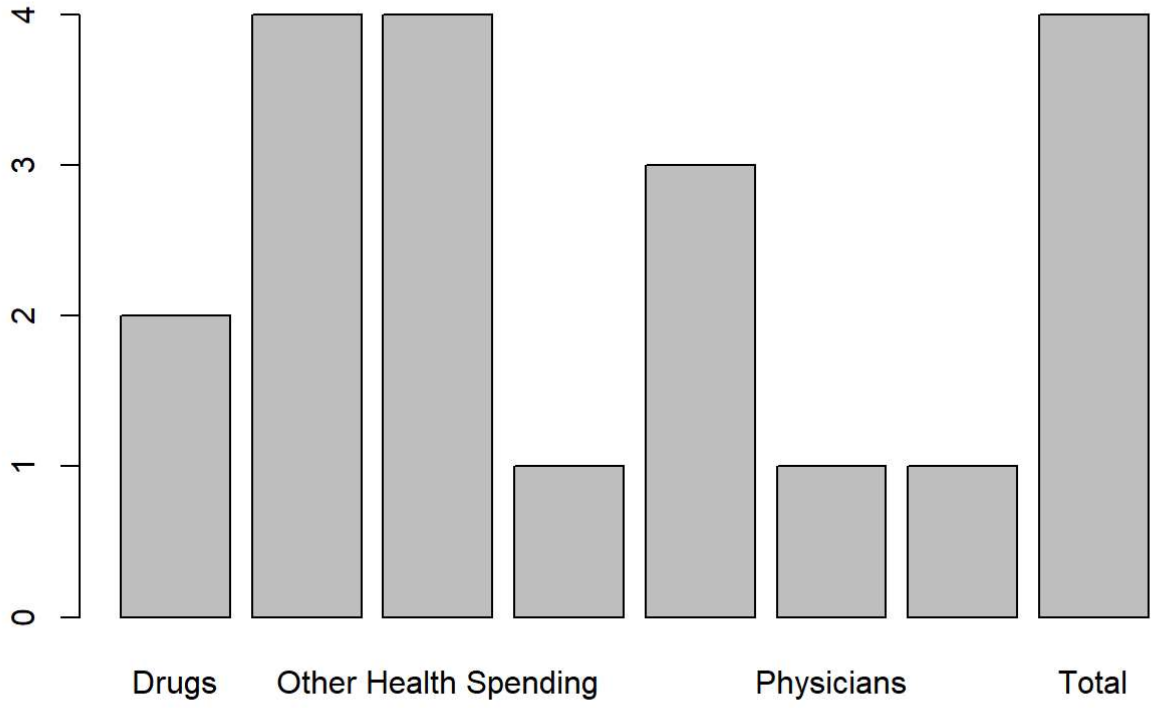

```
#THE PLOT SHOWS THAT WHICH SECTOR IS HAVING THE CURRENT DOLLARS PER CAPITA HIGHER AND LOWER


lf <- sample_n(df1,size = 10)
ggplot(lf,aes(x=Current.dollars.per.capita))+geom_bar(aes(fill=Year))+ggtitle("Distribution o
f dollars per capita vs years")
```




```
#THE PLOT SHOWS THAT IN WHICH YEAR HOW MUCH DOLLARS PER CAPITA IS SEPNT


lf <- sample_n(df1,size = 30)
ggplot(lf,aes(x=Year))+geom_bar(aes(fill=Use.of.Funds))+ggtitle("Distribution of year vs use
of funds")+theme(axis.text.x = element_text(angle = 90))
```

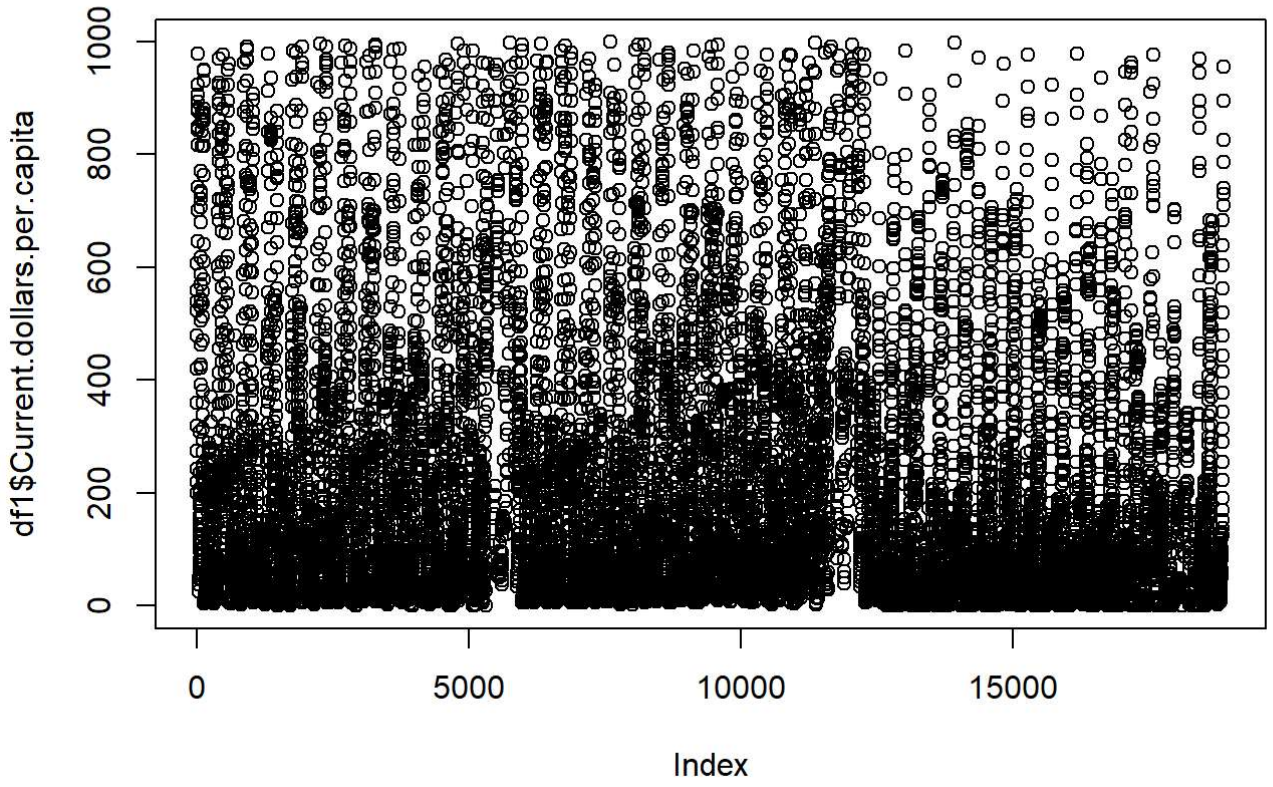



```
#THE PLOT SHOWS THAT IN WHICH YEAR WHERE THE FUNDS ARE BEING USED


lf <- sample(df1$Use.of.Funds,20)
x <- data.frame(table(lf))
barplot(x$Freq,names.arg = x$lf)
```



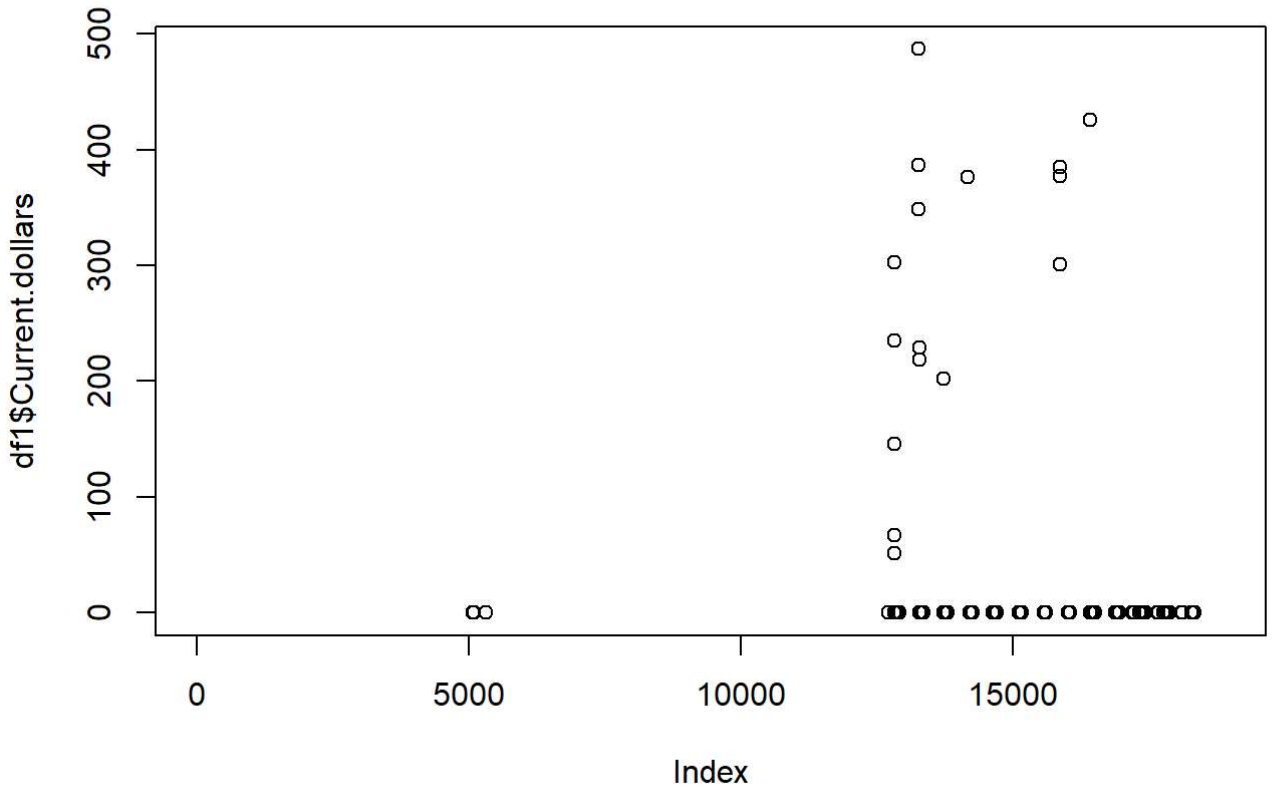

```
#THE PLOT SHOWS THAT WHAT SECTOR IS USING THE MOST FUNDS

plot(df1$Current.dollars.per.capita)
```

```
## Warning in xy.coords(x, y, xlabel, ylabel, log): NAs introduced by coercion
```




```
#WE CAN SEE THAT THE CURRENT DOLLARS PER CAPITA IS HIGHER BETWEEN 0 TO 100 AND ALSO FROM 100 TO 400 A LITTLE BIT HIGHER

plot(df1$Current.dollars)
```

```
## Warning in xy.coords(x, y, xlabel, ylabel, log): NAs introduced by coercion
```



CONCLUSION:

Our findings suggest that, in general, there is a positive association between healthcare spending and the economic indicators of labor productivity, personal income, per capita GDP, and other spending. Also, personal healthcare spending adversely impacts time spent on purchases of goods and services.Different states require varied investment in personal health expenditure, even if they have the same level of labor productivity. Overall, the study contributes to the growing literature on healthcare expenditure and economic performance. It outlines how the government can allocate healthcare expenditure in key dimensions that can stimulate economic growth while also improving the well-being of the population. It is also critical that policy makers implement appropriate policies at the macroeconomic level—targeted at public health expenditure and economic development. Overall, in light of the potential benefits of healthcare to the economy, universal access to healthcare is an area that warrants further research.