Tourism project

Lounès AÏT KARROU & Jinane AMAL

```r
library(here)
library(dplyr)
library(tidyr)
library(ggplot2)
library(knitr)
library(corrplot)
library(car)
theme_set(theme_bw())
here::i_am("TourismProject.Rproj")
```

I. Loading the data

```r
Arrivals<-read.csv("Datas/arrivals.csv",sep=";")
AvgExpenditures<-read.csv("Datas/avexpinttourists.csv",sep=";")
FoodExpenditures<-read.csv("Datas/foodexp.csv",sep=";")
GDP<-read.csv("Datas/gdppercap.csv",sep=";")
Gini<-read.csv("Datas/gini.csv",sep=";")
Happy<-read.csv("Datas/happinessladder.csv",sep=";")
TerrorDeath<-read.csv("Datas/terrorism-deaths.csv",sep=";")
Forest<-read.csv("Datas/propofprotectedforests.csv",sep=";")
```

II. Sources and GitHub

You can have access to the source of our data by clicking on the name of the variable :

- Average expenditures by tourist

- Gross Domestic Product per capita
- Gini Coefficient
- Number of arrivals
- Happiness and life satisfaction
- Food expenditure
- GDP per capita
- Terrorism death

Click here to access to our GitHub project.

III. Description of our sources

To get our data, we decide to use two different data banks.

First, we have used is the research publication called "Our World in Data". Founded by the economist Max Roser in 2011, "Our World in Data" is working in collaboration with thousands of researchers all around the world to try to answer and face the hardest problem our world is facing: poverty, diseases, hunger, climate change, war etc.
"Our World in Data" uses interactive charts and maps to illustrate the work of the researchers. In this website, we found the data for the following topics :

- The food expenditures per person from 2017 to 2021
- The self-reported life satisfaction from 2011 to 2022.
- The income inequalities measured by the Gini coefficient from 1967 to 2021
- The share of forest area within protected areas from 2000 to 2020
- The international tourist expenditures within the country they visit from 1995 to 2021
- The terrorism deaths from 1970 to 2021
- The GDP per capita from 1990 to 2021

All this data and their variables are going to be explained in the next part. Note that we are not going to keep those periods, we will deal with this issue later, during the data cleaning part.

We also went to the Data world bank to find our principal data set : the number of arrivals. The World Bank Group, established in 1944 along the International Monetary Fund at the Bretton Woods Conference, is one of the world's largest sources of funding and knowledge for developing countries. Its five institutions share a commitment to reducing poverty, increasing shared prosperity, and promoting sustainable development. This group is dividing in 5 institutions:

- The international bank for reconstruction

- The international development association

- The international finance corporation
- The multilateral investment guarantee agency
- The International Centre for Settlement of Investment Disputes.

Even if their main mission is, as they said themselves, to provide a wide array of financial products and technical assistance but also to help countries share, they also produce data that can be find in their site, where there is a whole page dedicated to a whole free collection of data.

IV. Description of our data

A) Number of arrivals:

Tourist arrivals, the primary dependent variable, stand as the pivotal metric impacted by chosen independent variables, shaping a country's overall attractiveness to tourism. Understanding and optimizing these influential factors can significantly influence the influx of visitors.

Data set A cleaning task 1 : Focus (2019 VS 2020)

```r
Arrivals19vs20 <- Arrivals |>
  select(1:2, 64:65)
```

Data set A cleaning task 2 : alphabetical order

```r
Arrivals19vs20 <- Arrivals19vs20 |>
  arrange(Country.Name)
```

Data set A cleaning task 3 : Renaming all 4 columns

```r
Arrivals19vs20 <- Arrivals19vs20|>
  rename(Country=Country.Name, Code= Country.Code, "2019" = X2019 , "2020" = X2020)
```

Cleaned Data set A Summary

Get number of rows and columns

```
nbrows <- nrow(Arrivals19vs20 %>% distinct(Country))
nbcol <- ncol(Arrivals19vs20)
```

The Cleaned Data set A Summary contains 266 number of columns and 4 number of rows

Cleaned Data set A Summary to be shown in html rendering:

```
print(summary(Arrivals19vs20))|>
knitr::kable()
```

```
   Country               Code                  2019                 2020          
 Length:266         Length:266         Min.   :3.600e+03   Min.   :      900  
 Class :character   Class :character   1st Qu.:1.209e+06   1st Qu.:   287550  
 Mode  :character   Mode  :character   Median :4.905e+06   Median :   877700  
                                       Mean   :9.191e+07   Mean   :  4685639  
                                       3rd Qu.:3.276e+07   3rd Qu.:  2902500  
                                       Max.   :2.403e+09   Max.   :117109000  
                                       NA's   :43          NA's   :134        
```

Country	Code	2019	2020
Length:266	Length:266	Min. :3.600e+03	Min. : 900
Class :character	Class :character	1st Qu.:1.209e+06	1st Qu.: 287550
Mode :character	Mode :character	Median :4.905e+06	Median : 877700
NA	NA	Mean :9.191e+07	Mean : 4685639
NA	NA	3rd Qu.:3.276e+07	3rd Qu.: 2902500
NA	NA	Max. :2.403e+09	Max. :117109000
NA	NA	NA's :43	NA's :134

Get number of rows and columns

```
nbrows <- nrow(Arrivals19vs20 %>% distinct(Country))
nbcol <- ncol(Arrivals19vs20)
```

The Cleaned Data set A Summary contains 266 number of columns and 4 number of rows.

B) Average Expenditures:

Average expenditures, as the second dependent variable, are intricately influenced by factors like cultural offerings and safety measures, shaping a country's overall appeal to tourism. These variables contribute to the financial decisions of tourists, impacting the level of spending and economic contributions within the destination.

Data set B cleaning task 1 : Focus (2019 VS 2020)

```
AvgExpenditures19vs20 <- AvgExpenditures|>
  filter(Year %in% c(2019, 2020))
```

Data set B cleaning task 2 : Pivoting

```
AvgExpenditures19vs20v1 <- AvgExpenditures19vs20|>
   pivot_wider(names_from= Year, values_from=Inbound_Tourism_Expenditure_adjusted)
```

Data set B cleaning task 3 : Renaming the first column

```
AvgExpenditures19vs20v1 <- AvgExpenditures19vs20v1|>
  rename(Country=Entity)
```

Data set B cleaning task 4 : Mutating the last 2 columns to be recognised as numerical values

```
AvgExpenditures19vs20v1 <- AvgExpenditures19vs20v1 |>
  mutate_at(vars("2019", "2020"), as.numeric)
```

```
Warning: There was 1 warning in `mutate()`.
i In argument: `2019 = .Primitive("as.double")(`2019`)`.
Caused by warning:
! NAs introduced by coercion
```

Cleaned Data set B Summary

```
print(names(AvgExpenditures19vs20v1))
```

```
[1] "Country" "Code"    "2019"    "2020"
```

```
print(dim(AvgExpenditures19vs20v1))
```

```
[1] 47  4
```

```
print(summary(AvgExpenditures19vs20v1))
```

```
   Country              Code                2019                2020          
 Length:47          Length:47          Min.   :1.607e+09   Min.   :5.425e+08  
 Class :character   Class :character   1st Qu.:5.556e+09   1st Qu.:2.050e+09  
 Mode  :character   Mode  :character   Median :1.257e+10   Median :6.716e+09  
                                       Mean   :2.139e+10   Mean   :1.078e+10  
                                       3rd Qu.:2.912e+10   3rd Qu.:1.331e+10  
                                       Max.   :7.218e+10   Max.   :7.589e+10  
                                       NA's   :2                              
```

Cleaned Data set B Summary to be shown in html rendering :

```
print(summary(AvgExpenditures19vs20v1))|>
knitr::kable()
```

```
   Country              Code                2019                2020          
 Length:47          Length:47          Min.   :1.607e+09   Min.   :5.425e+08  
 Class :character   Class :character   1st Qu.:5.556e+09   1st Qu.:2.050e+09  
 Mode  :character   Mode  :character   Median :1.257e+10   Median :6.716e+09  
                                       Mean   :2.139e+10   Mean   :1.078e+10  
```

```
                                         3rd Qu.:2.912e+10   3rd Qu.:1.331e+10
                                         Max.   :7.218e+10   Max.   :7.589e+10
                                         NA's   :2
```

Country	Code	2019	2020
Length:47	Length:47	Min. :1.607e+09	Min. :5.425e+08
Class :character	Class :character	1st Qu.:5.556e+09	1st Qu.:2.050e+09
Mode :character	Mode :character	Median :1.257e+10	Median :6.716e+09
NA	NA	Mean :2.139e+10	Mean :1.078e+10
NA	NA	3rd Qu.:2.912e+10	3rd Qu.:1.331e+10
NA	NA	Max. :7.218e+10	Max. :7.589e+10
NA	NA	NA's :2	NA

Get number of rows and columns

```r
nbrows <- nrow(AvgExpenditures19vs20v1 %>% distinct(Country))
nbcol <- ncol(AvgExpenditures19vs20v1)
```

The Cleaned Data set B Summary contains 47 number of columns and 4 number of rows.

C) Food Expenditures:

Food expenditures, as an independent variable, directly impact a country's tourism attractiveness by influencing the accessibility and affordability of diverse culinary experiences, shaping the overall appeal for travelers.

Data set C cleaning task 1 : Focus (2019 VS 2020)

```r
FoodExpenditures19vs20<- FoodExpenditures |>
  filter(Year %in% c(2019, 2020))
```

Data set C cleaning task 2: Pivoting

```
FoodExpenditures19vs20v1<- FoodExpenditures19vs20 |>
    pivot_wider(names_from= Year, values_from=Total.food.expenditure)
```

Data set C cleaning task 3 : Renaming the first column

```
FoodExpenditures19vs20v1 <- FoodExpenditures19vs20v1|>
  rename(Country=Entity)
```

Cleaned Data set C Summary

```
print(names(FoodExpenditures19vs20v1))
```

```
[1] "Country" "Code"    "2019"    "2020"
```

```
print(dim(FoodExpenditures19vs20v1))
```

```
[1] 104   4
```

```
print(summary(FoodExpenditures19vs20v1))
```

```
   Country              Code                2019              2020        
 Length:104         Length:104         Min.   :   1.608   Min.   :  27.34  
 Class :character   Class :character   1st Qu.: 711.413   1st Qu.: 737.59  
 Mode  :character   Mode  :character   Median :1228.930   Median :1217.91  
                                       Mean   :1387.151   Mean   :1427.79  
                                       3rd Qu.:1950.551   3rd Qu.:1976.76  
                                       Max.   :4318.868   Max.   :4181.52  
```

Cleaned Data set C Summary to be shown in html rendering :

```r
print(summary(FoodExpenditures19vs20v1))|>
knitr::kable()
```

```
   Country              Code                2019               2020     
 Length:104         Length:104         Min.   :   1.608   Min.   :  27.34  
 Class :character   Class :character   1st Qu.: 711.413   1st Qu.: 737.59  
 Mode  :character   Mode  :character   Median :1228.930   Median :1217.91  
                                       Mean   :1387.151   Mean   :1427.79  
                                       3rd Qu.:1950.551   3rd Qu.:1976.76  
                                       Max.   :4318.868   Max.   :4181.52  
```

Country	Code	2019	2020
Length:104	Length:104	Min. : 1.608	Min. : 27.34
Class :character	Class :character	1st Qu.: 711.413	1st Qu.: 737.59
Mode :character	Mode :character	Median :1228.930	Median :1217.91
NA	NA	Mean :1387.151	Mean :1427.79
NA	NA	3rd Qu.:1950.551	3rd Qu.:1976.76
NA	NA	Max. :4318.868	Max. :4181.52

Get number of rows and columns

```r
nbrows <- nrow(FoodExpenditures19vs20v1 %>% distinct(Country))
nbcol <- ncol(FoodExpenditures19vs20v1)
```

The Cleaned Data set C Summary contains 104 number of columns and 4 number of rows.

D) Forest:

The extent of protected forest surface in a country serves as an independent variable influencing its tourism attractiveness, signifying environmental conservation and offering unique natural attractions.

Data set D cleaning task 1 : change 0s into NA

```
Forest <- Forest|>
 mutate_all(~ifelse(. == 0, NA, .))
```

Data set D cleaning task 2 : Get rid of NA lines

```
Forest <- Forest|>
  filter(!is.na(Code))
```

Data set D cleaning task 3 : Focus (2019 VS 2020)

```
Forest19vs20 <- Forest |>
  filter(Year %in% c(2019, 2020))
```

Data set D cleaning task 4 : pivoting

```
Forest19vs20v1 <- Forest19vs20 |>
    pivot_wider(names_from= Year, values_from= Proportionofprotectedforests)
```

Data set D cleaning task 5 : Renaming the first column

```
Forest19vs20v1 <- Forest19vs20v1|>
  rename(Country=Entity)
```

Cleaned Data set D Summary

```
print(names(Forest19vs20v1))
```

```
[1] "Country" "Code"    "2019"    "2020"
```

```
print(dim(Forest19vs20v1))
```

```
[1] 153   4
```

```
print(summary(Forest19vs20v1))
```

```
   Country              Code                2019             2020      
 Length:153         Length:153         Min.   : 0.31   Min.   : 0.31  
 Class :character   Class :character   1st Qu.:10.23   1st Qu.:10.52  
 Mode  :character   Mode  :character   Median :18.45   Median :18.42  
                                       Mean   :24.15   Mean   :24.04  
                                       3rd Qu.:33.46   3rd Qu.:33.15  
                                       Max.   :98.66   Max.   :99.74  
                                       NA's   :8       NA's   :6      
```

Cleaned Data set C Summary to be shown in html rendering

```
print(summary(Forest19vs20v1))|>
knitr::kable()
```

```
   Country              Code                2019             2020      
 Length:153         Length:153         Min.   : 0.31   Min.   : 0.31  
 Class :character   Class :character   1st Qu.:10.23   1st Qu.:10.52  
 Mode  :character   Mode  :character   Median :18.45   Median :18.42  
                                       Mean   :24.15   Mean   :24.04  
                                       3rd Qu.:33.46   3rd Qu.:33.15  
                                       Max.   :98.66   Max.   :99.74  
                                       NA's   :8       NA's   :6      
```

Country	Code	2019	2020
Length:153	Length:153	Min. : 0.31	Min. : 0.31
Class :character	Class :character	1st Qu.:10.23	1st Qu.:10.52
Mode :character	Mode :character	Median :18.45	Median :18.42
NA	NA	Mean :24.15	Mean :24.04
NA	NA	3rd Qu.:33.46	3rd Qu.:33.15
NA	NA	Max. :98.66	Max. :99.74
NA	NA	NA’s :8	NA’s :6

Get number of rows and columns

```
nbrows <- nrow(Forest19vs20v1 %>% distinct(Country))
nbcol <- ncol(Forest19vs20v1)
```

The Cleaned Data set D Summary contains 153 number of columns and 4 number of rows.

E) Gross Domestic Product per capita :

High GDP signifies Economic prosperity which often translates into improved amenities, accessibility, and diverse attractions, making the destination more appealing to tourists.

Data set E cleaning task 1 : Focus (2019 VS 2020)

```
GDP19vs20 <- GDP |>
  select(1:2, 64:65)
```

Data set E cleaning task 2 : Renaming all 4 columns

```
GDP19vs20 <- GDP19vs20|>
  rename(Country= Country.Name, Code= Country.Code, "2019" = X2019 , "2020" = X2020)
```

Cleaned Data set E Summary

```
print(names(GDP19vs20))
```

```
[1] "Country" "Code"    "2019"    "2020"
```

```
print(dim(GDP19vs20))
```

```
[1] 266   4
```

```
print(summary(GDP19vs20))
```

```
  Country              Code                2019             2020       
Length:266         Length:266         Min.   :   217   Min.   :   216.8  
Class  :character  Class  :character  1st Qu.:  2185   1st Qu.:  2132.9  
Mode   :character  Mode   :character  Median :  6897   Median :  6249.0  
                                      Mean   : 17420   Mean   : 16195.0  
                                      3rd Qu.: 22438   3rd Qu.: 19409.9  
                                      Max.   :199383   Max.   :182537.3  
                                      NA's   :8        NA's   :8         
```

Cleaned Data set E Summary to be shown in html rendering :

```r
print(summary(GDP19vs20))|>
knitr::kable()
```

```
  Country              Code                2019             2020       
Length:266         Length:266         Min.   :   217   Min.   :   216.8  
Class  :character  Class  :character  1st Qu.:  2185   1st Qu.:  2132.9  
Mode   :character  Mode   :character  Median :  6897   Median :  6249.0  
                                      Mean   : 17420   Mean   : 16195.0  
                                      3rd Qu.: 22438   3rd Qu.: 19409.9  
                                      Max.   :199383   Max.   :182537.3  
                                      NA's   :8        NA's   :8         
```

	Country	Code	2019	2020
	Length:266	Length:266	Min. : 217	Min. : 216.8
	Class :character	Class :character	1st Qu.: 2185	1st Qu.: 2132.9
	Mode :character	Mode :character	Median : 6897	Median : 6249.0
	NA	NA	Mean : 17420	Mean : 16195.0
	NA	NA	3rd Qu.: 22438	3rd Qu.: 19409.9
	NA	NA	Max. :199383	Max. :182537.3
	NA	NA	NA’s :8	NA’s :8

Get number of rows and columns

```r
nbrows <- nrow(GDP19vs20 %>% distinct(Country))
nbcol <- ncol(GDP19vs20)
```

The Cleaned Data set E Summary contains 266 number of columns and 4 number of rows.

F) GINI Index :

The Gini Index measures income inequality within a country, with higher values indicating greater inequality. High Gini Index scores may negatively impact a country's attractiveness to tourism, as economic disparities can affect overall stability and inclusivity, influencing visitors' perceptions and experiences.

Data set F cleaning task 1 : Focus (2019 VS 2020)

```
Gini19vs20 <- Gini |>
  filter(Year %in% c(2019, 2020))
```

Data set F cleaning task 2 : Getting rid of the additional lines for urban or rural region for which we already have an average gini coefficient

```
Gini19vs20 <- Gini19vs20|>
  filter(!grepl(" - (rural|urban)$", Entity))
```

Data set F cleaning task 3 : pivoting

```
Gini19vs20v1 <- Gini19vs20 |>
    pivot_wider(names_from= Year, values_from= Gini.coefficient)
```

Data set C cleaning task 4 : Renaming the first column

```
Gini19vs20v1 <- Gini19vs20v1|>
  rename(Country=Entity)
```

Cleaned Data set F Summary

```
print(names(Gini19vs20v1))
```

```
[1] "Country" "Code"    "2019"    "2020"
```

```r
print(dim(Gini19vs20v1))
```

```
[1] 62  4
```

```r
print(summary(Gini19vs20v1))
```

```
   Country              Code                2019             2020       
 Length:62          Length:62          Min.   :0.2323   Min.   :0.2438  
 Class :character   Class :character   1st Qu.:0.2930   1st Qu.:0.3473  
 Mode  :character   Mode  :character   Median :0.3427   Median :0.4015  
                                       Mean   :0.3513   Mean   :0.3936  
                                       3rd Qu.:0.4044   3rd Qu.:0.4516  
                                       Max.   :0.5349   Max.   :0.5417  
                                       NA's   :2        NA's   :43      
```

Cleaned Data set F Summary to be shown in html rendering

```r
print(summary(Gini19vs20v1))|>
knitr::kable()
```

```
   Country              Code                2019             2020       
 Length:62          Length:62          Min.   :0.2323   Min.   :0.2438  
 Class :character   Class :character   1st Qu.:0.2930   1st Qu.:0.3473  
 Mode  :character   Mode  :character   Median :0.3427   Median :0.4015  
                                       Mean   :0.3513   Mean   :0.3936  
                                       3rd Qu.:0.4044   3rd Qu.:0.4516  
                                       Max.   :0.5349   Max.   :0.5417  
                                       NA's   :2        NA's   :43      
```

	Country	Code	2019	2020
	Length:62	Length:62	Min. :0.2323	Min. :0.2438
	Class :character	Class :character	1st Qu.:0.2930	1st Qu.:0.3473
	Mode :character	Mode :character	Median :0.3427	Median :0.4015
	NA	NA	Mean :0.3513	Mean :0.3936
	NA	NA	3rd Qu.:0.4044	3rd Qu.:0.4516
	NA	NA	Max. :0.5349	Max. :0.5417
	NA	NA	NA's :2	NA's :43

Country	Code	2019	2020

Get number of rows and columns

```
nbrows <- nrow(Gini19vs20v1 %>% distinct(Country))
nbcol <- ncol(Gini19vs20v1)
```

The Cleaned Data set F Summary contains 62 number of columns and 4 number of rows.

G) Happiness Ladder :

The Happiness Ladder, a measure of a country's overall well-being and life satisfaction, serves as a pivotal independent variable influencing its attractiveness to tourism. Higher rankings on the Happiness Ladder often correlate with a positive perception, encouraging tourists to seek enriching and joyful experiences in those destinations.

Data set G cleaning task 1 : Focus (2019 VS 2020)

```
Happy19vs20 <- Happy |>
  filter(Year %in% c(2019, 2020))
```

Data set G cleaning task 2 : pivoting

```
Happy19vs20v1 <- Happy19vs20 |>
    pivot_wider(names_from= Year, values_from= Cantril.ladder.score)
```

Data set G cleaning task 3 : Renaming the first column

```
Happy19vs20v1 <- Happy19vs20v1|>
  rename(Country=Entity)
```

Data set G cleaning task 4 : Switching the order of the last 2 columns representing our focused years of study

```
Happy19vs20v1 <- Happy19vs20v1|>
  select(-"2019", -"2020", "2019", "2020")
```

Cleaned Data set G Summary

```
print(names(Happy19vs20v1))
```

```
[1] "Country" "Code"    "2019"    "2020"
```

```
print(dim(Happy19vs20v1))
```

```
[1] 153   4
```

```
print(summary(Happy19vs20v1))
```

```
   Country              Code                2019            2020      
 Length:153         Length:153         Min.   :2.567   Min.   :2.523  
 Class :character   Class :character   1st Qu.:4.724   1st Qu.:4.852  
 Mode  :character   Mode  :character   Median :5.515   Median :5.534  
                                       Mean   :5.473   Mean   :5.533  
                                       3rd Qu.:6.229   3rd Qu.:6.255  
                                       Max.   :7.809   Max.   :7.842  
                                                       NA's   :4      
```

Cleaned Data set G Summary to be shown in html rendering :

```
print(summary(Happy19vs20v1))|>
knitr::kable()
```

```
   Country              Code                  2019            2020
 Length:153         Length:153         Min.   :2.567   Min.   :2.523
 Class :character   Class :character   1st Qu.:4.724   1st Qu.:4.852
 Mode  :character   Mode  :character   Median :5.515   Median :5.534
                                       Mean   :5.473   Mean   :5.533
                                       3rd Qu.:6.229   3rd Qu.:6.255
                                       Max.   :7.809   Max.   :7.842
                                                       NA's   :4
```

Country	Code	2019	2020
Length:153	Length:153	Min. :2.567	Min. :2.523
Class :character	Class :character	1st Qu.:4.724	1st Qu.:4.852
Mode :character	Mode :character	Median :5.515	Median :5.534
NA	NA	Mean :5.473	Mean :5.533
NA	NA	3rd Qu.:6.229	3rd Qu.:6.255
NA	NA	Max. :7.809	Max. :7.842
NA	NA	NA	NA’s :4

Get number of rows and columns

```
nbrows <- nrow(Happy19vs20v1 %>% distinct(Country))
nbcol <- ncol(Happy19vs20v1)
```

The Cleaned Data set G Summary contains 153 number of columns and 4 number of rows.

H) Terrorism Attacks

Terrorist attacks act as an independent variable negatively impacting a country’s attractiveness to tourism, creating safety concerns and deterring potential visitors. The frequency and severity of such incidents significantly influence the perceived security of a destination, shaping tourists’ decisions and preferences.

Data set H cleaning task 1 : Focus (2019 VS 2020)

```
TerrorDeath19vs20 <- TerrorDeath |>
  filter(Year %in% c(2019, 2020))
```

Data set H cleaning task 2 : Get rid 0’s into NA

```
TerrorDeath19vs20 <- TerrorDeath19vs20|>
 mutate_all(~ifelse(. == 0, NA, .))
```

Data set H cleaning task 3 : Get rid of NA lines

```
TerrorDeath19vs20 <- TerrorDeath19vs20|>
  filter(!is.na(Code))
```

Data set H cleaning task 4 : Since previous command changes the values into NA’s we should now revert the NA’S of the last column to 0 (mening no terrorist attack)

```
TerrorDeath19vs20[is.na(TerrorDeath19vs20[, ncol(TerrorDeath19vs20)]), ncol(TerrorDeath19v
```

Data set H cleaning task 5 : Renaming the first column

```
TerrorDeath19vs20 <- TerrorDeath19vs20|>
  rename(Country=Entity)
```

Data set D cleaning task 6 : pivoting

```
TerrorDeath19vs20v1 <- TerrorDeath19vs20 |>
    pivot_wider(names_from= Year, values_from= Terrorism.deaths)
```

Cleaned Data set H Summary

```
print(names(TerrorDeath19vs20v1))
```

```
[1] "Country" "Code"    "2019"    "2020"
```

```
print(dim(TerrorDeath19vs20v1))
```

```
[1] 199   4
```

```
print(summary(TerrorDeath19vs20v1))
```

```
   Country               Code                2019              2020        
 Length:199         Length:199         Min.   :   0.00   Min.   :    0.0  
 Class :character   Class :character   1st Qu.:   0.00   1st Qu.:    0.0  
 Mode  :character   Mode  :character   Median :   0.00   Median :    0.0  
                                       Mean   : 103.09   Mean   :  115.4  
                                       3rd Qu.:   4.75   3rd Qu.:    4.0  
                                       Max.   :8257.00   Max.   :10081.0  
                                       NA's   :1         NA's   :1        
```

Cleaned Data set H Summary to be shown in html rendering:

```
print(summary(TerrorDeath19vs20v1))|>
knitr::kable()
```

```
   Country               Code                2019              2020        
 Length:199         Length:199         Min.   :   0.00   Min.   :    0.0  
 Class :character   Class :character   1st Qu.:   0.00   1st Qu.:    0.0  
 Mode  :character   Mode  :character   Median :   0.00   Median :    0.0  
                                       Mean   : 103.09   Mean   :  115.4  
                                       3rd Qu.:   4.75   3rd Qu.:    4.0  
                                       Max.   :8257.00   Max.   :10081.0  
                                       NA's   :1         NA's   :1        
```

	Country	Code	2019	2020
	Length:199	Length:199	Min. : 0.00	Min. : 0.0
	Class :character	Class :character	1st Qu.: 0.00	1st Qu.: 0.0
	Mode :character	Mode :character	Median : 0.00	Median : 0.0
	NA	NA	Mean : 103.09	Mean : 115.4
	NA	NA	3rd Qu.: 4.75	3rd Qu.: 4.0
	NA	NA	Max. :8257.00	Max. :10081.0
	NA	NA	NA’s :1	NA’s :1

Get number of rows and columns

```
nbrows <- nrow(TerrorDeath19vs20v1 %>% distinct(Country))
nbcol <- ncol(TerrorDeath19vs20v1)
```

The Cleaned Data set H Summary contains 198 number of columns and 4 number of rows.

V. Description of our Research Question :

"What makes a country more attractive to tourists ?"

Travel holds profound importance in our lives, extending far beyond the mere act of moving from one place to another. It serves as a gateway to diverse cultures, broadening our perspectives and fostering a deeper understanding of the world. The importance of travel lies not only in the personal enrichment it offers but also in the role it plays in showcasing a country's distinctive charm. Every nation becomes a storyteller, enticing visitors with its rich history, cultural treasures, and breathtaking landscapes. Tourism becomes a bridge between cultures, a means to celebrate diversity and foster global understanding. As we explore different corners of the world, we contribute to a shared narrative of interconnectedness, where each country's unique appeal adds vibrancy to the collective tapestry of global tourism. Therefore, understanding the factors that make a destination appealing becomes a fascinating research endeavor. As we embark on this research trail, the pivotal question arises: "What makes a country more attractive to tourists?" This question will focus on a comparison study between 2 years 2019 & 2020 : 2019 which will represent the initial tourism trends before the pandemic and 2020 will show us how the covid pandemic impacted these tourism trends through the evolution of the chosen variables This query seeks to unravel the distinct elements that shape a nation's allure, delving into cultural, environmental, infrastructural and economy oriented facets that draw visitors. By exploring this question, we aim to not only enrich our comprehension of travel dynamics but also contribute valuable insights to the ongoing dialogue on global tourism trends. In evaluating a country's tourism success, we are focusing on the number of tourist arrivals. The number of arrivals serves as a primary indicator, reflecting the appeal and popularity of a destination, while average expenditure provides insights into the economic impact of tourism. To understand the influencing factors, we consider various variables. Gross Domestic Product (GDP) reflects the economic strength of a nation, potentially correlating with tourism appeal. Gini coefficient measures income inequality, influencing the distribution of tourist spending. The Happiness Ladder index gauges the overall well-being of a country's population, potentially affecting its attractiveness to visitors. The proportion of protected forests contribute to environmental considerations, influencing sustainable tourism. We also took the food expenditures as another indicator of the affordability experiencing a country's culinary traditions. Lastly, we chose as another variable the number of terrorist attacks per country reflecting the potential danger tourists will face if they visit a country .

VI. Graphical representation of the main variable

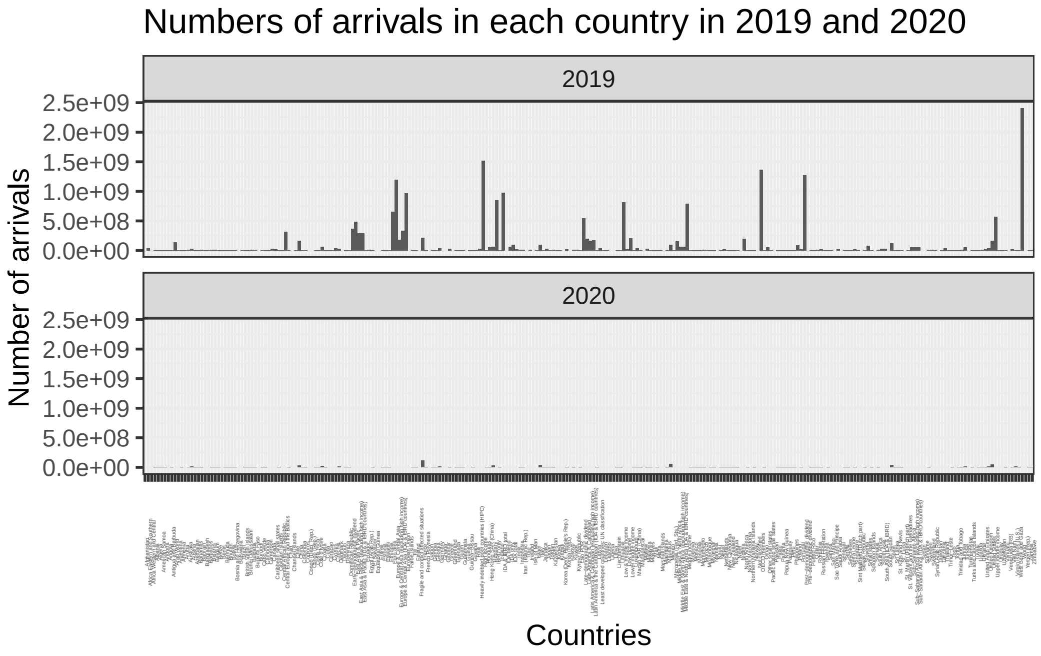
To have a proper representation of our main target variable, we will first transform our data into a long data, to have a column for the two different years.

```
datalong<-gather(Arrivals19vs20,key="Years",value="Values",-Country, -Code)
```

Now we can, using the ggplot2 library, have the graphical representation :

```
ggplot(datalong, aes(x = Country, y = Values)) +
  geom_bar(stat = "identity", position = "dodge") +
  facet_wrap(~Years, ncol=1) +
  labs(title = "Numbers of arrivals in each country in 2019 and 2020",
       x = "Countries",
       y = "Number of arrivals") +
  theme(axis.text.x = element_text(size = 2, angle = 90))
```

```
Warning: Removed 177 rows containing missing values (`geom_bar()`).
```



V. Merged Data set:

Before providing the correlation matrix and the linear regression, we merge our data set using the following code : We decided to merge our tables using the inner_joint function, joining by the country and the code so we can have the same countries for all the tables.
Noticed that, due to the numerous number of N.a on the Gini table, we decided to exclude this variable in our study.

```
TabA<-inner_join(Arrivals19vs20,AvgExpenditures19vs20v1,by=c('Code','Country'),suffix=c('_
TabB<-inner_join(FoodExpenditures19vs20v1,Forest19vs20v1,by=c('Code','Country'),suffix=c('
TabC<-inner_join(GDP19vs20,Happy19vs20v1,by=c('Code','Country'),suffix=c('gdp','_happy') )
TabD<-inner_join(TabA,TabB,by=c('Code','Country'))
Tab <- inner_join(TabD,TabC,by=c('Code','Country'))
print(Tab)|>knitr::kable()
```

```
          Country Code 2019_arrivals 2020_arrivals    2019_avg    2020_avg
1       Australia  AUS       9466000       1828000 43698405000 24566798000
2         Austria  AUT      31884000      15091000 26367382000 15201416000
3          Brazil  BRA       6353000            NA 10444931000  6715630000
4        Bulgaria  BGR      12552000       4973000  9880238000  3643277800
5          Canada  CAN      32430000            NA 31014547000 14142006000
6           Chile  CHL       5431000            NA  3508667600   688953340
7        Colombia  COL       4531000       1396000 12570884000  3808076800
8      Costa Rica  CRI       3366000       1146500  6241390600  2055008800
9         Croatia  HRV      60021000      21608000 20312027000  9460204000
10        Czechia  CZE      37202000            NA 12084634000  5837232000
11        Denmark  DNK      33093000      15595000  7577459000  3396685000
12        Estonia  EST       6103000       1695000  2540746200   847222100
13        Finland  FIN       3290000        896000  3676791600  1200951600
14         France  FRA     217877000     117109000 70561915000 35428737000
15        Germany  DEU      39563000      12449000 49050790000 25265594000
16        Hungary  HUN      61397000      31641000 12835703000  5836135400
17      Indonesia  IDN      16107000       4053000 48872790000  9884304000
18        Ireland  IRL      10951000            NA  5556050400  2019657900
19         Israel  ISR       4905000            NA  6417405400  2045350500
20          Italy  ITA      95399000      38419000 61096395000 24090290000
21         Latvia  LVA       8342000       3204000  1607296400  1239247000
22      Lithuania  LTU       6150000       2284000  2687526000  1001357600
23    Netherlands  NLD      20129000       7265000 20303493000 10216919000
24    New Zealand  NZL       3888000        996000 10587685000  5683342000
25         Norway  NOR       5879000       1397000  5055613400  1656521500
26         Poland  POL      88515000            NA 29119690000 16834577000
```

```
27        Romania  ROU      12815000       5023000  8246021600  3245932000
28   Saudi Arabia  SAU      20292000            NA 39394790000  9354371000
29       Slovenia  SVN       4702000       1216000  4490630700  1956867300
30   South Africa  ZAF      14797000       3886600 18801973000  6448314000
31         Sweden  SWE       7616000       1957000  9246004000  4245574700
32    Switzerland  CHE      11818000            NA 13572140000  7249521000
33 United Kingdom  GBR      40857000      11101000 61683315000 27754422000
34  United States  USA     165478000      45037000          NA 75888140000
   2019_food 2020_food 2019_forest 2020_forest   2019gdp   2020gdp 2019_happy
1  2546.8447 2722.4756       18.09       18.09 54941.066 51722.069     7.2228
2  2291.1020 2085.3740       22.63       22.63 50070.403 48809.227     7.2942
3   599.6366  655.7949       29.45       29.68  8845.324  6923.700     6.3756
4  1038.9020 1123.2827       18.37       18.37  9878.769 10153.477     5.1015
5  2098.3420 2294.0552        8.50        8.50 46374.153 43349.678     7.2321
6  1275.6061 1230.2683       21.62       21.62 14627.145 13165.386     6.2285
7   561.5499  555.1840       20.70       20.70  6436.509  5304.289     6.1634
8  2220.0195 2211.1170       44.08       44.08 12669.341 12179.257     7.1214
9  1699.2692 1762.2521        2.87        2.87 15086.212 14236.535     5.5047
10 1568.2462 1630.1895        5.48        5.48 23664.848 22992.879     6.9109
11 2806.4456 3019.4075        8.39        8.39 59592.981 60915.424     7.6456
12 2075.1401 2199.5994       21.99       21.99 23424.485 23595.244     6.0218
13 2563.4158 2809.4917       12.63       12.63 48629.858 49169.719     7.8087
14 2658.6528 2846.2764       23.18       23.29 40494.898 39055.283     6.6638
15 2267.8400 2317.7488       28.95       28.95 46793.687 46772.825     7.0758
16 1170.9205 1282.0813       22.47       22.54 16786.214 16125.609     6.0004
17  700.9293  693.1733       54.48       54.48  4151.228  3895.618     5.2856
18 1778.8379 1919.0405       19.14       19.24 80927.075 85420.191     7.0937
19 3373.9187 3008.4070       18.18       18.18 44452.233 44846.792     7.1286
20 2710.3584 2820.4473       35.12       35.12 33673.751 31918.693     6.3874
21 1764.8906 1816.6553       16.51       16.51 17945.222 18207.140     5.9500
22 2251.4084 2692.4365       31.18       31.28 19598.401 20363.924     6.2155
23 2404.7288 2481.7320       59.48       59.48 52476.273 52162.570     7.4489
24 3137.8190 3241.8650       36.22       36.22 42796.431 41760.595     7.2996
25 2814.1484 2733.5383        4.88        5.02 76430.589 68340.018     7.4880
26 1301.0120 1516.3815       32.82       32.82 15700.014 15816.820     6.1863
27 1932.8785 2018.2958       37.76       37.76 12957.999 13047.458     6.1237
28 1741.5045 1771.8440        0.31        0.31 23405.706 20398.061     6.4065
29 1811.9442 1828.0132       19.59       19.59 26016.079 25545.241     6.3634
30  616.8080  597.0115        1.31        1.31  6688.775  5741.641     4.8141
31 2588.6580 2601.9092        7.74        7.74 51939.430 52837.904     7.3535
32 3662.3967 3970.2050       17.73       17.73 84121.931 85656.323     7.5599
33 1831.6174 1951.0782        9.19        9.19 42747.080 40318.417     7.1645
34 2356.8203 2595.5752       10.23       10.23 65120.395 63528.634     6.9396
```

```
   2020_happy
1      7.1835
2      7.2678
3      6.3301
4      5.2655
5      7.1033
6      6.1719
7      6.0124
8      7.0694
9      5.8817
10     6.9647
11     7.6195
12     6.1888
13     7.8421
14     6.6899
15     7.1545
16     5.9916
17     5.3445
18     7.0853
19     7.1571
20     6.4831
21     6.0320
22     6.2554
23     7.4640
24     7.2766
25     7.3925
26     6.1661
27     6.1400
28     6.4940
29     6.4607
30     4.9564
31     7.3627
32     7.5715
33     7.0636
34     6.9515
```

Country	Code	2019_arrivals	2020_arrivals	2019_avg	2020_avg	2019_food	2020_food	2019_forest	2020_forest	2019_gdp	2020gdp	2019_happy	2020_happy
Australia	AUS	9466000	1828000	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	18.09	[illegible]	[illegible]	[illegible]	7.1835
Austria	AUT	31884000	15091000	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	22.63	[illegible]	[illegible]	[illegible]	7.2678
Brazil	BRA	6353000	NA	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	29.68	[illegible]	[illegible]	[illegible]	6.3301
Bulgaria	BGR	12552000	4973000	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	18.37	[illegible]	[illegible]	[illegible]	5.2655

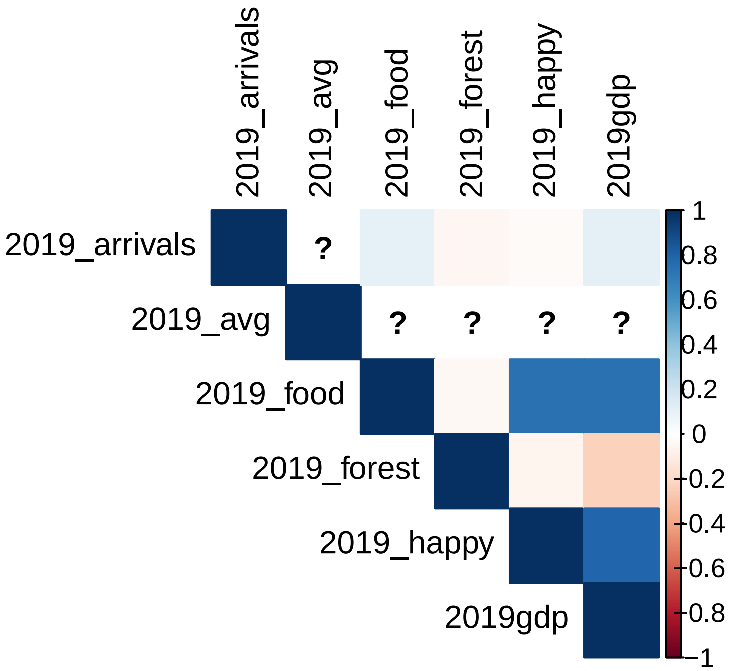
Country	Code	2019_arrivals	2020_arrivals	2019_avg	2020_avg	2019_food	2020_food	2019_forest	2020_forest	2019gdp	2020gdp	2019_happy	2020_happy
Canada	CAN	32430000	NA	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	8.50	[illegible]	[illegible]	[illegible]	7.1033
Chile	CHL	5431000	NA	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	21.62	[illegible]	[illegible]	[illegible]	6.1719
Colombia	COL	4531000	1396000	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	20.70	[illegible]	[illegible]	[illegible]	6.0124
Costa Rica	CRI	3366000	1146500	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	44.08	[illegible]	[illegible]	[illegible]	7.0694
Croatia	HRV	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	2.87	[illegible]	[illegible]	[illegible]	5.8817
Czechia	CZE	37202000	NA	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	5.48	[illegible]	[illegible]	[illegible]	6.9647
Denmark	DNK	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	8.39	[illegible]	[illegible]	[illegible]	7.6195
Estonia	EST	6103000	1695000	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	21.99	[illegible]	[illegible]	[illegible]	6.1888
Finland	FIN	3290000	896000	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	12.63	[illegible]	[illegible]	[illegible]	7.8421
France	FRA	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	23.29	[illegible]	[illegible]	[illegible]	6.6899
Germany	DEU	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	28.95	[illegible]	[illegible]	[illegible]	7.1545
Hungary	HUN	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	22.54	[illegible]	[illegible]	[illegible]	5.9916
Indonesia	IDN	16107000	4053000	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	54.48	[illegible]	[illegible]	[illegible]	5.3445
Ireland	IRL	10951000	NA	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	19.24	[illegible]	[illegible]	[illegible]	7.0853
Israel	ISR	4905000	NA	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	18.18	[illegible]	[illegible]	[illegible]	7.1571
Italy	ITA	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	35.12	[illegible]	[illegible]	[illegible]	6.4831
Latvia	LVA	8342000	3204000	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	16.51	[illegible]	[illegible]	[illegible]	6.0320
Lithuania	LTU	6150000	2284000	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	31.28	[illegible]	[illegible]	[illegible]	6.2554
Netherlands	NLD	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	59.48	[illegible]	[illegible]	[illegible]	7.4640
New Zealand	NZL	3888000	996000	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	36.22	[illegible]	[illegible]	[illegible]	7.2766
Norway	NOR	5879000	1397000	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	5.02	[illegible]	[illegible]	[illegible]	7.3925
Poland	POL	88515000	NA	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	32.82	[illegible]	[illegible]	[illegible]	6.1661
Romania	ROU	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	37.76	[illegible]	[illegible]	[illegible]	6.1400
Saudi Arabia	SAU	20292000	NA	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.31	[illegible]	[illegible]	[illegible]	6.4940
Slovenia	SVN	4702000	1216000	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	19.59	[illegible]	[illegible]	[illegible]	6.4607
South Africa	ZAF	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1.31	[illegible]	[illegible]	[illegible]	4.9564
Sweden	SWE	7616000	1957000	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	7.74	[illegible]	[illegible]	[illegible]	7.3627
Switzerland	CHE	11818000	NA	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	17.73	[illegible]	[illegible]	[illegible]	7.5715
United Kingdom	GBR	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	9.19	[illegible]	[illegible]	[illegible]	7.0636
United States	USA	[illegible]	[illegible]	NA	[illegible]	[illegible]	[illegible]	[illegible]	10.23	[illegible]	[illegible]	[illegible]	6.9515

VI. Correlation Matrix:

In this section, we want to quantify the strength of linear relationships between the variables we kept for our following regression through correlation matrices for the two comparison years 2019 vs 2020 .

$Focus\ Year$: 2019

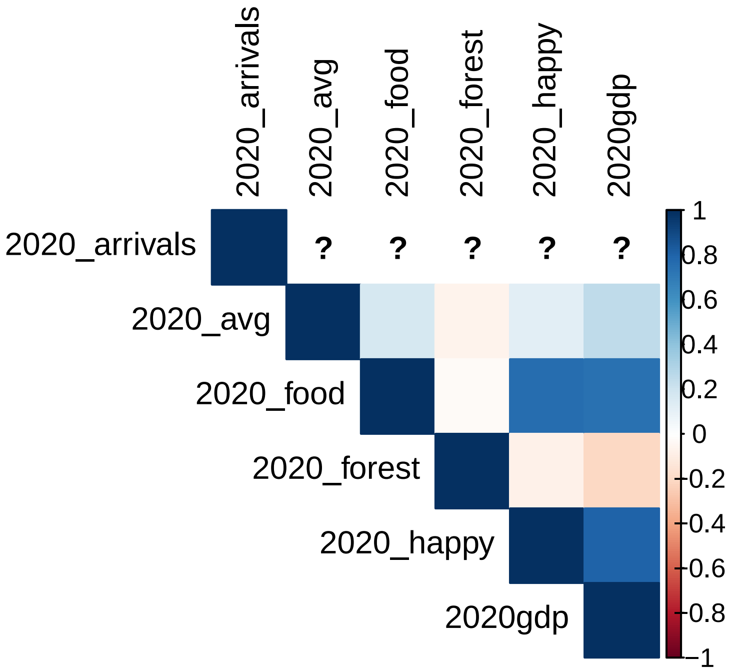
```
corr = cor(subset(Tab, select = c(`2019_arrivals`,`2019_avg`,`2019_food`, `2019_forest`,`2
corrplot(corr, type = "upper", method = "color" , tl.col = "black")
```



We notice two important things for year 2019: -First: 3 of the explanatory variables are positively correlated to each other: food, happiness, and the gdp indicators. -Second: The average expenditure indicator is the only indicator which we can't determine its correlation to other variables

$Focus\ Year$: 2020

```
corr = cor(subset(Tab, select = c(`2020_arrivals`,`2020_avg`,`2020_food`, `2020_forest`,`2
corrplot(corr, type = "upper", method = "color" , tl.col = "black")
```



We notice that for year 2020: The obtained correlation matrix slightly resembles the 2019 correlation matrix as much as the 3 positively correlated explanatory variables (food, happiness, and the gdp indicators) are concerned . We also have that 2020_arrivas is this time the only indicator which we can't determine its correlation to other variables .

VII. Linear regression:

In this section, we want to study the impact of all the variables on our main variable by using a linear regression. To do so, we will first focus on the impact in 2019 and after that in 2020.

We want to estimate the following model :

$$Arrivals = \beta_0 + \beta_1 AverageExp + \beta_2 Food + \beta_3 Forest + \beta_4 Happy + \beta_5 GDP + \varepsilon$$

With ε the error vector.

```
reg1 <- lm (Tab$'2019_arrivals'~ Tab$'2019_avg'+ Tab$'2019_food' + Tab$'2019_forest' + Tab
summary(reg1)
```

```
Call:
```

```
lm(formula = Tab$"2019_arrivals" ~ Tab$"2019_avg" + Tab$"2019_food" +
    Tab$"2019_forest" + Tab$"2019_happy" + Tab$"2019gdp")

Residuals:
      Min        1Q    Median        3Q       Max
-50726654 -12213414  -3160403   5360948 110575029

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)        4.307e+07  7.412e+07   0.581    0.566
Tab$"2019_avg"     1.421e-03  2.992e-04   4.748 5.99e-05 ***
Tab$"2019_food"    1.457e+04  1.174e+04   1.241    0.225
Tab$"2019_forest" -2.081e+05  4.254e+05  -0.489    0.629
Tab$"2019_happy"  -9.117e+06  1.354e+07  -0.673    0.506
Tab$"2019gdp"     -2.268e+02  4.944e+02  -0.459    0.650
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 32680000 on 27 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.4786,    Adjusted R-squared:  0.382
F-statistic: 4.956 on 5 and 27 DF,  p-value: 0.002396
```

We notice that only a single explanatory variable is significant : `2019_avg`.
In fact, looking at the value of the t-stat, we have $t_{\beta_1} = 4.748$ and since $t_{27}^{-1}(1 - \frac{0.05}{2}) = 2.052$, we get $t_{\beta_1} > t_{27}^{-1}$.
For all the others explenatory variable, we have the opposite, meaning that they are not significant.

We do the same but now for the second year of our study :

```
reg2 <- lm (Tab$'2020_arrivals'~ Tab$'2020_avg'+ Tab$'2020_food' + Tab$'2020_forest' + Tab
summary(reg2)
```

```
Call:
lm(formula = Tab$"2020_arrivals" ~ Tab$"2020_avg" + Tab$"2020_food" +
    Tab$"2020_forest" + Tab$"2020_happy" + Tab$"2020gdp")

Residuals:
      Min        1Q    Median        3Q       Max
-23912427  -8096168  -2559705   3839610  72613858
```

```
Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)        3.484e+07  6.880e+07   0.506   0.6185
Tab$"2020_avg"     9.286e-04  3.085e-04   3.010   0.0072 **
Tab$"2020_food"    1.604e+04  9.733e+03   1.648   0.1157
Tab$"2020_forest" -9.154e+04  3.180e+05  -0.288   0.7765
Tab$"2020_happy"  -8.027e+06  1.341e+07  -0.599   0.5564
Tab$"2020gdp"     -3.346e+02  5.550e+02  -0.603   0.5536
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 20930000 on 19 degrees of freedom
  (9 observations deleted due to missingness)
Multiple R-squared:  0.4323,    Adjusted R-squared:  0.283
F-statistic: 2.894 on 5 and 19 DF,  p-value: 0.0416
```

Once again, the only significant variable is `2020_avg`.

However, one should pay attention to the value of our adjusted R^2. In fact, in both cases, R^2 is not high enough (respectively 0.382 and 0.283). Thus, maybe this linear regression with thoses explanatory variable is not the best, maybe the unsignificant variable are downgrading the quality of our model, suggesting that those variables may not explain the number of arrivals in a country.

VIII. Durbin Watson Test:

Let's keep our initial model. Indeed, we wish to assess the extent to which this model violates the fundamental assumptions of simple regression.
To do this, we will carry out the Durbin-Watson test, a test allowing us to highlight a potential autocorrelation of the errors, in other words the errors are correlated. Mathematically, the Durbin-Watson test seeks to verify the significance of the coefficient ρ in the formula:

$$\varepsilon_i = \rho\varepsilon_j + u_i$$

To do so, we will use the `durbinWatsonTest` available on r :

```
durbinWatsonTest (reg1)
```

```
 lag Autocorrelation D-W Statistic p-value
   1      -0.1969869      2.250298   0.468
 Alternative hypothesis: rho != 0
```

```
durbinWatsonTest (reg2)
```

```
 lag Autocorrelation D-W Statistic p-value
   1      -0.1290652      2.126653   0.894
 Alternative hypothesis: rho != 0
```

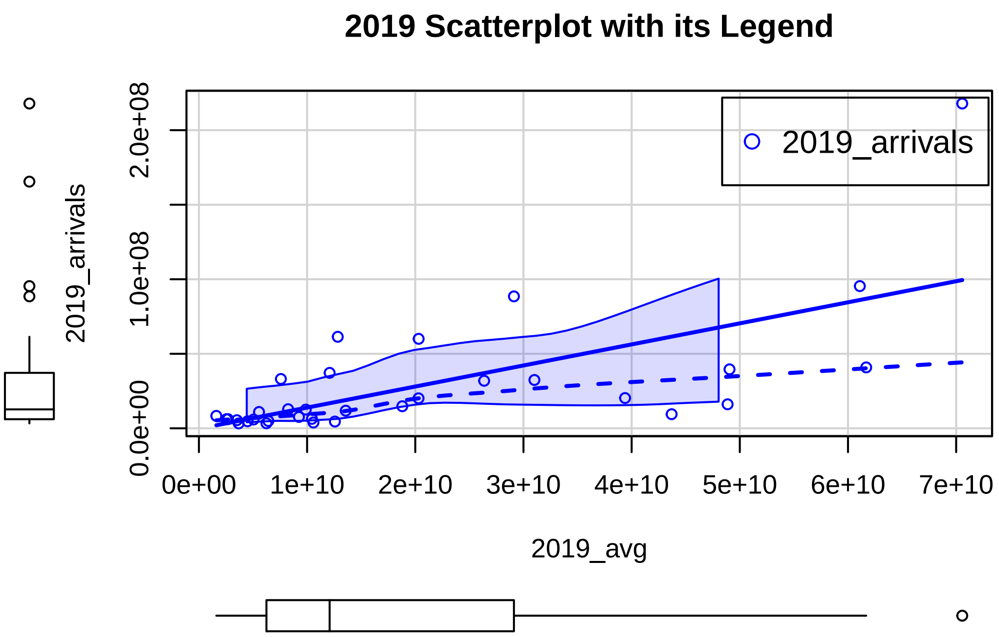
We observe that, for both regression, the p-value is higher than 0.05, we won't reject the null hypothesis. In fact, this result could have been more thant predictable since our regression is not temporal and it depends on different countries, so their results and data are not correlated to each other.

IX.Plots:

IX. Final Scatter plots:

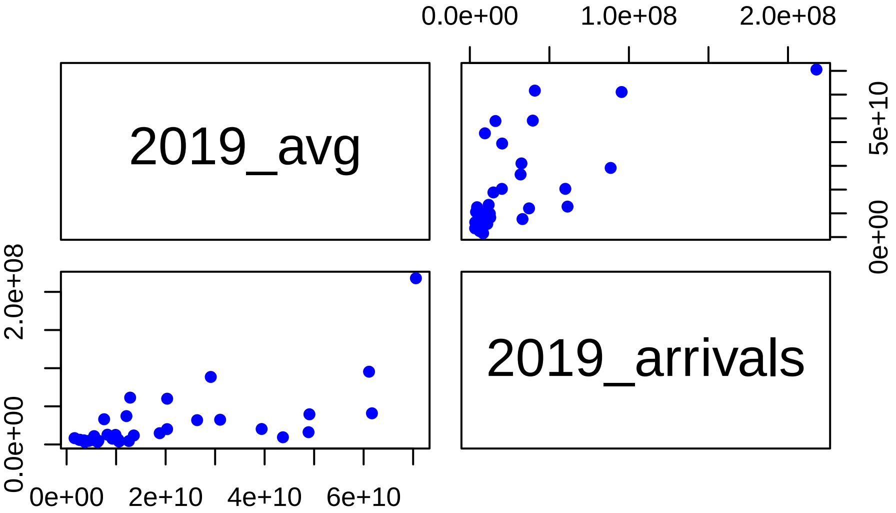
For the final part of the project, we will focus on the relation between the arrivals and the only significant explanatory variable we found on part VII : the average expenditures.
To do so, we are going to compare how it would have been if the two variables were perfectly correlated and what is happening in reality.

```
scatterplot(Tab$'2019_arrivals' ~ Tab$'2019_avg', data = Tab, main = "2019 Scatterplot wit
            xlab = "2019_avg", ylab = "2019_arrivals")
legend("topright", legend = c("2019_arrivals"), col = c("blue"), pch = 1)
```



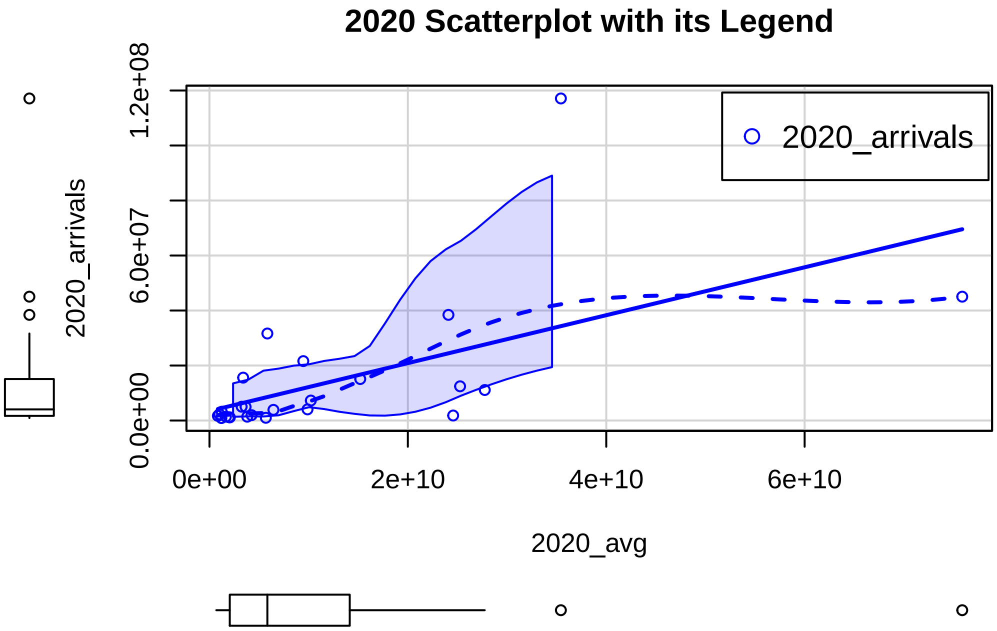
On this graph, the blue line represent the perfect correlation between arrivals and the average expenditures.\ The blue surface, for its part, represent the average dispertion of our results in real life, from our data, in 2019.
To observe this dispertion, we also provide below a pair of graph that represent those variable individually.

```
pairs(Tab[, c('2019_avg', '2019_arrivals')], col = "blue", pch = 16)
```

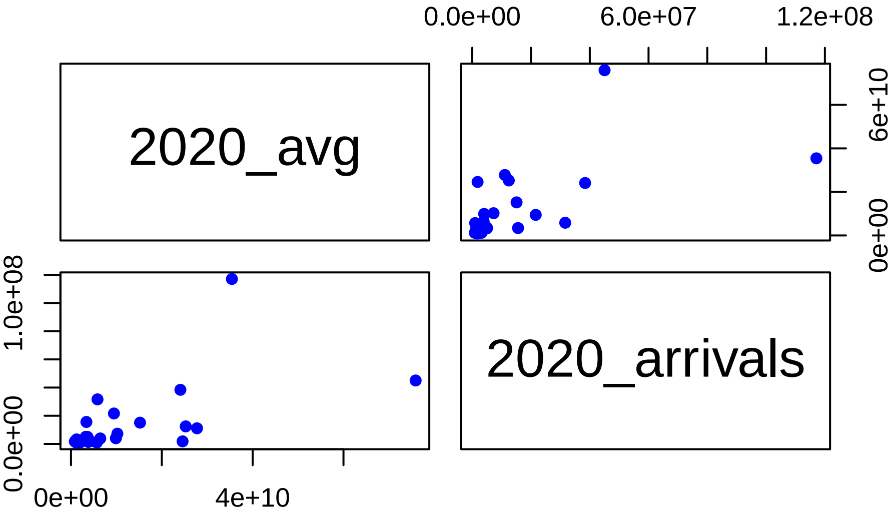


We do the same graphs in 2020.

```
scatterplot(Tab$'2020_arrivals' ~ Tab$'2020_avg', data = Tab, main = "2020 Scatterplot wit
            xlab = "2020_avg", ylab = "2020_arrivals")
legend("topright", legend = c("2020_arrivals"), col = c("blue"), pch = 1)
```



```
pairs(Tab[, c('2020_avg', '2020_arrivals')], col = "blue", pch = 16)
```



We can observe that, in both year, there is a correlation between those two variables. They are following the same pattern of growth as we can observe in all of those grpahs, even though at some point the difference between those two is not negligible.
Those observations are coherent with the obtained results from the two linear regression where the coefficient was positive, translating a positive realtion between the variables.

X. Conclusion :

These quantitative analysis has helped us understand what is the most important factor in a country's tourism attraction .

Our regression results followed by graphical representations proves that the most influential variable throughout both years 2019 and 2020 (covid year) is the average expenditures which is not surprising : Indeed a higher average expenditure in a country is synonymous to higher quality of services, luxurious experiences, cultural or business events . It can also be a sign of safety.

We can also say that this high rate of tourists arrivals in these countries is achieved via effective marketing and branding which is due to their high money allocated to research and business marketing .

In conclusion average expenditure is key to the success of the country's tourism objective and this won't change in the near future .