European Flights Data

2023-04-15

Table of Contents

Intorduction

First time in early 19's **Wright Brothers** began the Ariel age of flights by successful flights of a powered heavier-than-air flying machine. The best flight of the day, with Wilbur at the controls, covered 255.6 m (852 ft) in 59 seconds. for more information visit CLICK HERE Air Flights are the mode of transportation between cities, countries, and continents. It is the world fastest mode of transportation and many people use this transportation to reach at their destinations. Although this mode of transport is far more expensive then the other existing means of transport but beside that huge number of people are using it from decades.

As we know nowadays Air flights become one the best and most favorite medium among the people who travel from one country to an other country of the world. In this data set of **European Flights data** We are going to analyze the data of all different air flights of Europe. In this data we will discuss the trend of air flights before, during, and after COVID-19 by analyzing following variables which are **(YEAR)** which shows the year in which the flights did their operations, **(MONTH_MON)** Shows the month of flight operations,**(STATE_NAME)** from which state that flight belongs,**(FLT_DEP_1)** the departure time of a flight, and many more. This data is provided by the **jthomasmock** which gives you the record till 2022. For more details you can visit https://github.com/rfordatascience/tidytuesday/tree/master/data/2022/2022-07-12.

In this report we will discuss the major steps include in data analysis and figure out the answers of our research questions and prove our answers with the help of visualization of data in the form of charts or presentation. Here is the summary of the data.

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.2.3
```

```
## Warning: package 'ggplot2' was built under R version 4.2.3

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.
0.0 ──
## ✓ dplyr     1.1.0     ✓ readr     2.1.4
## ✓ forcats   1.0.0     ✓ stringr   1.5.0
## ✓ ggplot2   3.4.2     ✓ tibble    3.1.8
## ✓ lubridate 1.9.2     ✓ tidyr     1.3.0
## ✓ purrr     1.0.1
## ── Conflicts ────────────────────────────────────────── tidyverse_conflict
s() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
## ℹ Use the ]8;;http://conflicted.r-lib.org/conflicted package]8;; to force
all conflicts to become errors

flights <- read.csv("F:/Data Work/flights.csv")
data(flights)

## Warning in data(flights): data set 'flights' not found

##View(flights)
summary.data.frame(flights)

##       YEAR        MONTH_NUM       MONTH_MON           FLT_DATE
##  Min.   :2016   Min.   : 1.000   Length:688099      Length:688099
##  1st Qu.:2017   1st Qu.: 3.000   Class :character   Class :character
##  Median :2019   Median : 6.000   Mode  :character   Mode  :character
##  Mean   :2019   Mean   : 6.301
##  3rd Qu.:2020   3rd Qu.: 9.000
##  Max.   :2022   Max.   :12.000
##
##    APT_ICAO           APT_NAME          STATE_NAME          FLT_DEP_1
##  Length:688099      Length:688099      Length:688099      Min.   :  0.00
##  Class :character   Class :character   Class :character   1st Qu.:  5.00
##  Mode  :character   Mode  :character   Mode  :character   Median : 17.00
##                                                           Mean   : 63.24
##                                                           3rd Qu.: 71.00
##                                                           Max.   :847.00
##
##    FLT_ARR_1        FLT_TOT_1      FLT_DEP_IFR_2    FLT_ARR_IFR_2
##  Min.   :  0.00   Min.   :   0.0   Min.   :   0.0   Min.   :  0.0
##  1st Qu.:  5.00   1st Qu.:  10.0   1st Qu.:  38.0   1st Qu.: 38.0
##  Median : 17.00   Median :  35.0   Median :  91.0   Median : 91.0
##  Mean   : 63.28   Mean   : 126.5   Mean   : 143.7   Mean   :143.6
##  3rd Qu.: 71.00   3rd Qu.: 141.0   3rd Qu.: 195.0   3rd Qu.:195.0
##  Max.   :813.00   Max.   :1628.0   Max.   :1039.0   Max.   :817.0
##                                    NA's   :479785   NA's   :479785
##  FLT_TOT_IFR_2    Pivot.Label
##  Min.   :   0.0   Length:688099
##  1st Qu.:  76.0   Class :character
```

```
##  Median : 182.0   Mode  :character
##  Mean   : 287.3
##  3rd Qu.: 390.0
##  Max.   :1624.0
##  NA's   :479785
```

Research Questions

Q1. As we know on daily basis there are tones of flights that do operations form Europe to the different parts of the word and the operations only may disturb due some weather issues or some flight internal problems so our first question is base on what are the trends of flights before COVID-19 in the Europe?

Q2. During COVID-19 the world stopped the daily routine activities and in the world every where government imposed lock down due to which people were unable to travel from one country to another. So here we want to know about What are the trends of flights during COVID-19?

Q3. As in 2022 the world was successful to make a vaccination for COVID-19 so slowly all activities were on the way to normal routine. So in this research phase we want to know what are the trends of flights after the 1 year of COVID-19 vaccination?

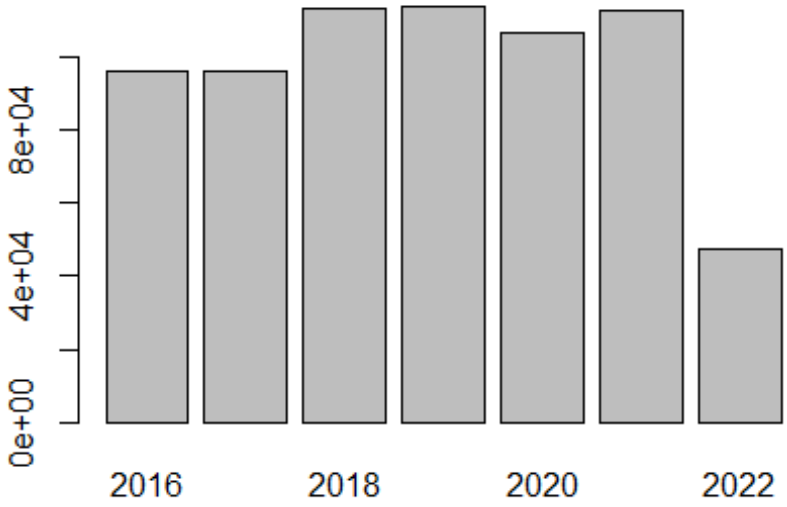
Variables of Interest

Following are the variables we are going to use in our analysis in order to find the answer of the above questions.

1. YEAR:

This variable is used to represent the data of the Years which shows the record of flights from the year 2016 to year 2022 like it contains **2016**, **2017**. Here is the code.

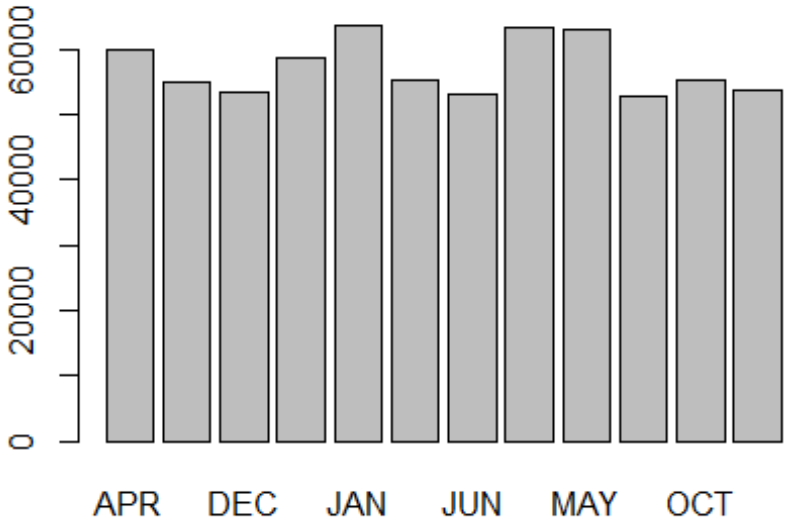
```
library(ggplot2)
barplot(table(flights$YEAR))
```



2. MONTH_MON:

This variable also works like the previous base on the date data. But in case of MONTH_MON it will return the complete name of the month in which the data is recorded for the flight. For example: it will return **MARCH** , **APRIL**. Here is the code.

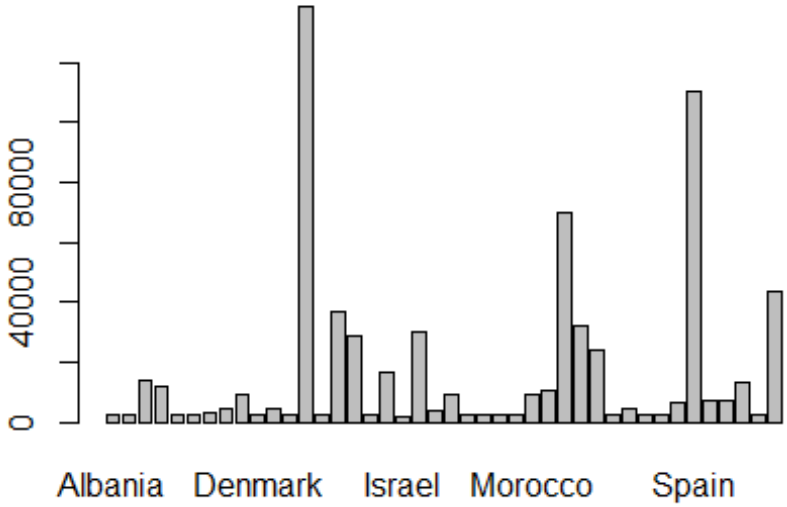
```
barplot(table(flights$MONTH_MON))
```



3. STATE_NAME:

This variable will give you detail info regarding the states of the Europe from which you get the data of flights. Here is the code.

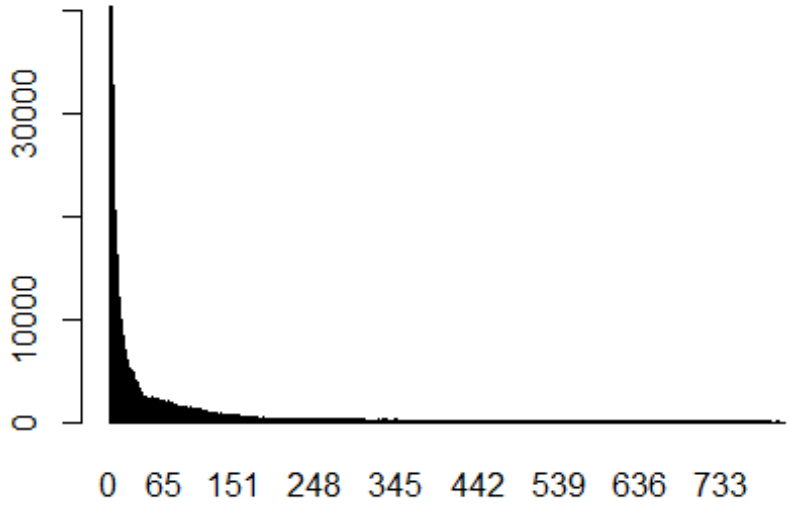
```
barplot(table(flights$STATE_NAME))
```



4. FLT_DEP_1:

Will give information regarding the first flight departure from 2016 to 2022 according to the data available in the data set. the minimum flight Departure number during analysis was 0 while the maximum latitude was 847. Here is the code.

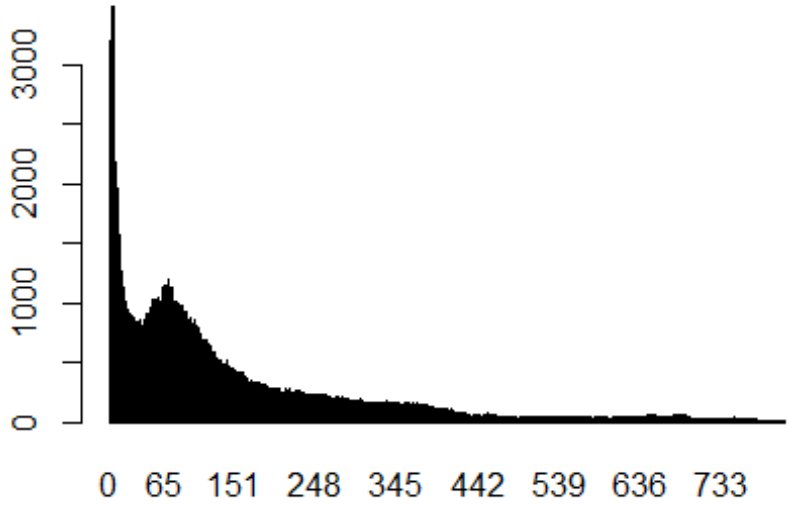
```
barplot(table(flights$FLT_DEP_1))
```



5. FLT_DEP_IFR_2:

Will give information regarding the second flight departure from 2016 to 2022 according to the data available in the data set. the minimum flight Departure number during analysis was 0 while the maximum latitude was 1039. Here is the code.

```
barplot(table(flights$FLT_DEP_IFR_2))
```



Data cleaning/wrangling

By using the function of **summary()** we will get to know the actual amount of observations in the data which is **688099**. Analyzing this amount of data is to hectic and so we will divide this data into small chunks to make it more readable and error free. Following are the methods we use in data cleaning/wrangling.

1. Filter Data: One of the most used function when it comes to data cleaning is **filter()** this will help you to filter the you need in you analysis. As our data is consist on the Air flight operations of Europe from year 2016 to year 2022, So make it more sense I divided this data on the basis of periods of flights which are **Before_Covid**, **In_Covid**, and **After_Covid**. and filter each era of data into a separate Data Frame. The code for the filter() is given below.

*Before_Covid Code:**

```
library(dplyr)
Before_Covid <- filter(flights, YEAR == "2016" | YEAR == "2017" | YEAR == "20
18" | YEAR == "2019" )
```

IN_Covid Code:

```
In_Covid <- filter(flights, YEAR == "2020" | YEAR == "2021")
```

After_Covid Code:

```
After_Covid <- filter(flights, YEAR == "2022")
```

2. Sorting

After dividing data into small data frames now it is time to sort data make it ready for our analysis. In sorting we have 2 ways to sort one sort data in ascending order and one in descending order the choose in yours. in this data we sort data in descending order. To arrange data in descending order we use arrange function with name of data set by placing **-** sign at the start of the data frame.

Here is the code to sort the data by using **arrange()** function.

```
##arrange(Before_Covid,desc(-FLT_DEP_1))
## I put this code in a comment as it generate and place all of the data of d
atabase in a word file
```

That's how you can arrange data in the remaining data frames too. Sometimes it's important to rearrange data but sometimes it doesn't matter. In our case to answer question 1 sorting of data doesn't matter. So it depends on you. If you want to sort it then above code will help you out in it.

Choice of data visualisations and rationale

Now it's time to visualize our data by using some functions and packages in this assignment I use **ggplot2** for data visualization the procedure to use it to first install the package by using command **install.packages("ggplot2")**. Once it installed then you need to import it by using **library(ggplot2)**. When it comes to visualization you have so many ways to visualize you data by using different kinds of graphs and Visualizations, So it's hard to decide which one is best for your visualization. So there are some tips to choose the best way to represent your data.

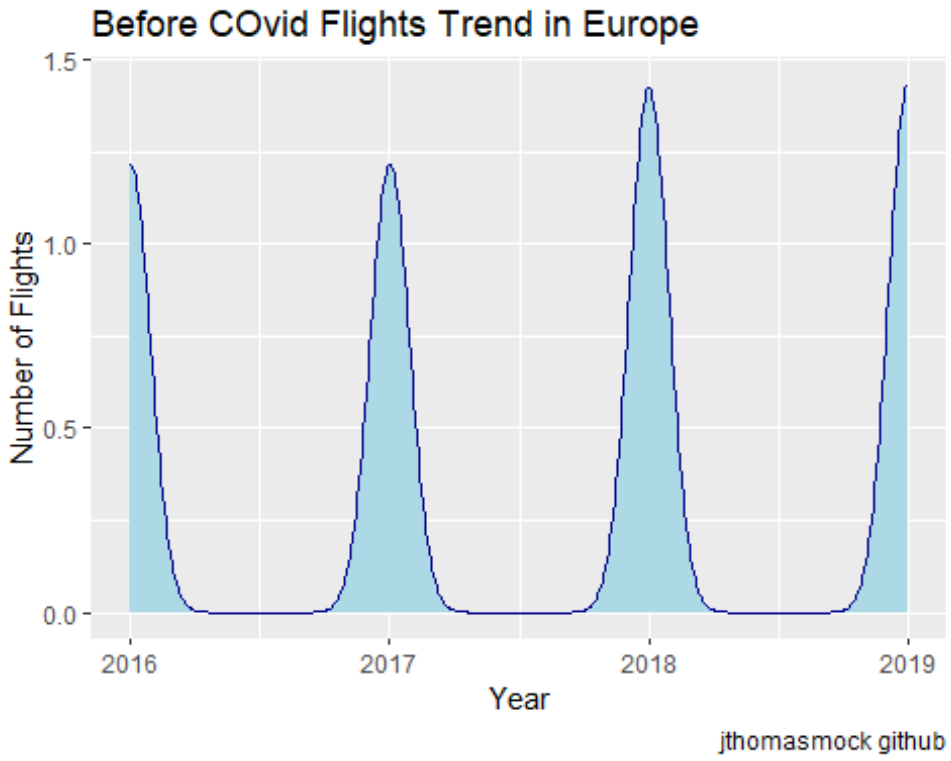
1. The visualization should be clear.
2. Should be easy to understand.
3. It should be fine and engaging.
4. The representation of data should be prominent.

As our data frames has data of thousands of rows so it's really hard to present it. So I chosen point graph to give the answer of my first question by making a visualization of every data frame I created. Here is the code of the visualization.

Visualizations For Question 1

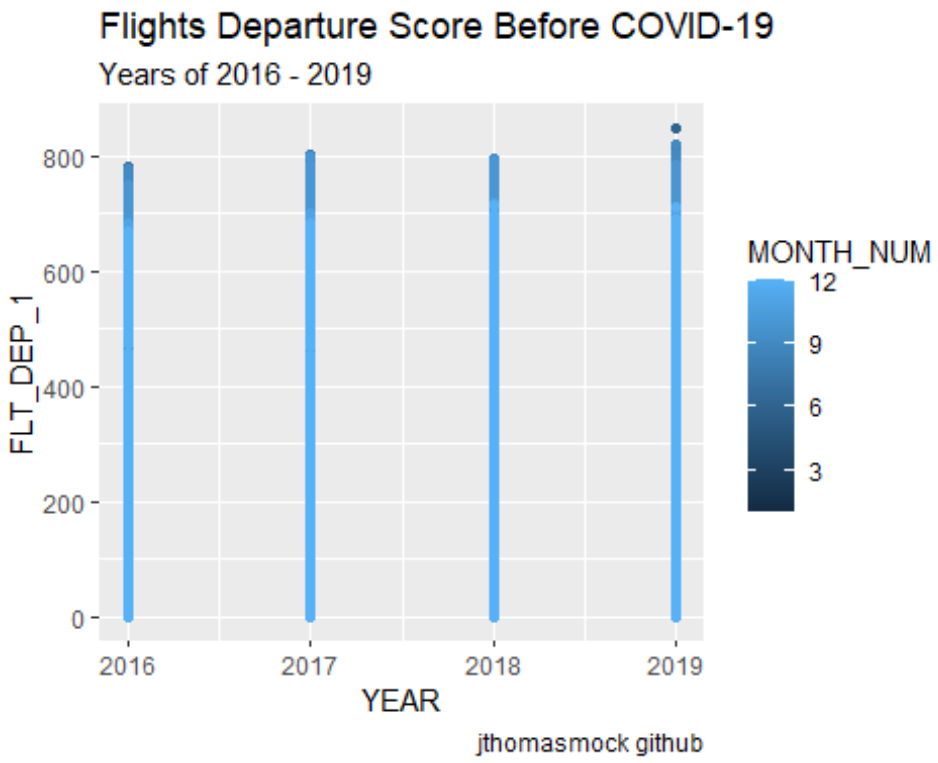
Before_Covid Visualization Code 1:

```
library(ggplot2)
ggplot(data = Before_Covid, aes(x = YEAR)) +
  geom_density(fill = "lightblue", color = "darkblue") +
  labs(x = "Year", y = "Number of Flights", title = "Before COvid Flights Tre
nd in Europe", caption = "jthomasmock github")
```



Before_Covid Visualization Code 2:

```
ggplot(data = Before_Covid) + geom_point(mapping = aes(x = YEAR, y = FLT_DEP_
1,
color = MONTH_NUM))+labs(title = "Flights Departure Score Before COVID-19",
subtitle = "Years of 2016 - 2019",
caption = "jthomasmock github")
```

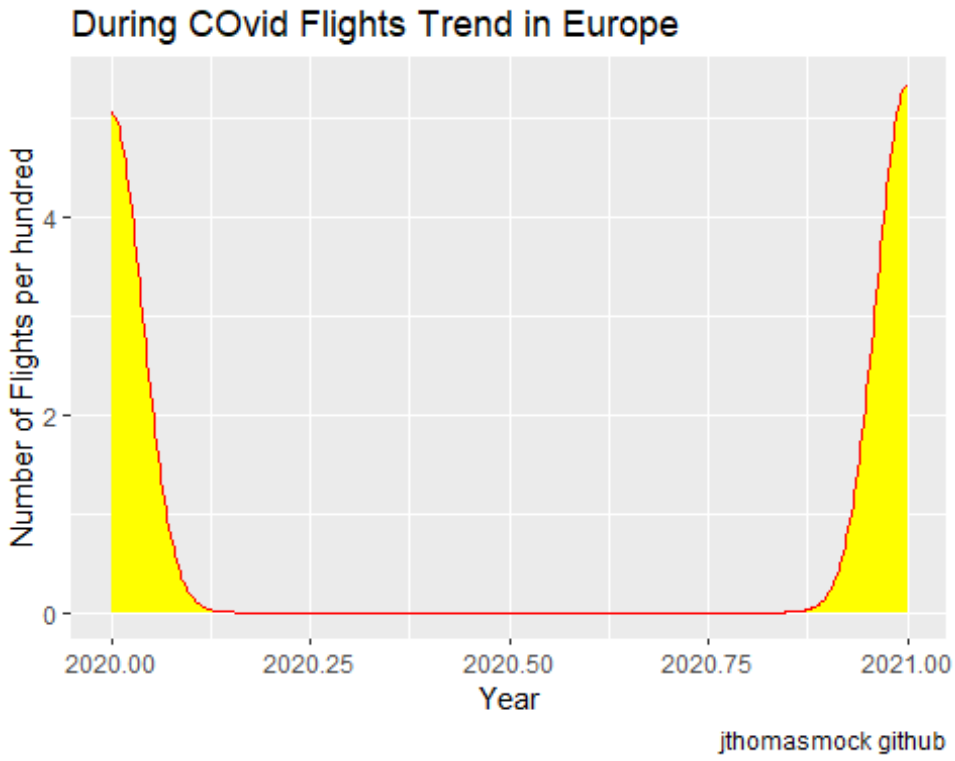


Visualization For Question 2

As for question 2 we need to find out the flights trends in the era of covid-19 that is 2020 to 2021 after that scientists discover a vaccination against virus so these visualizations shows the trends of flights in covid-19.

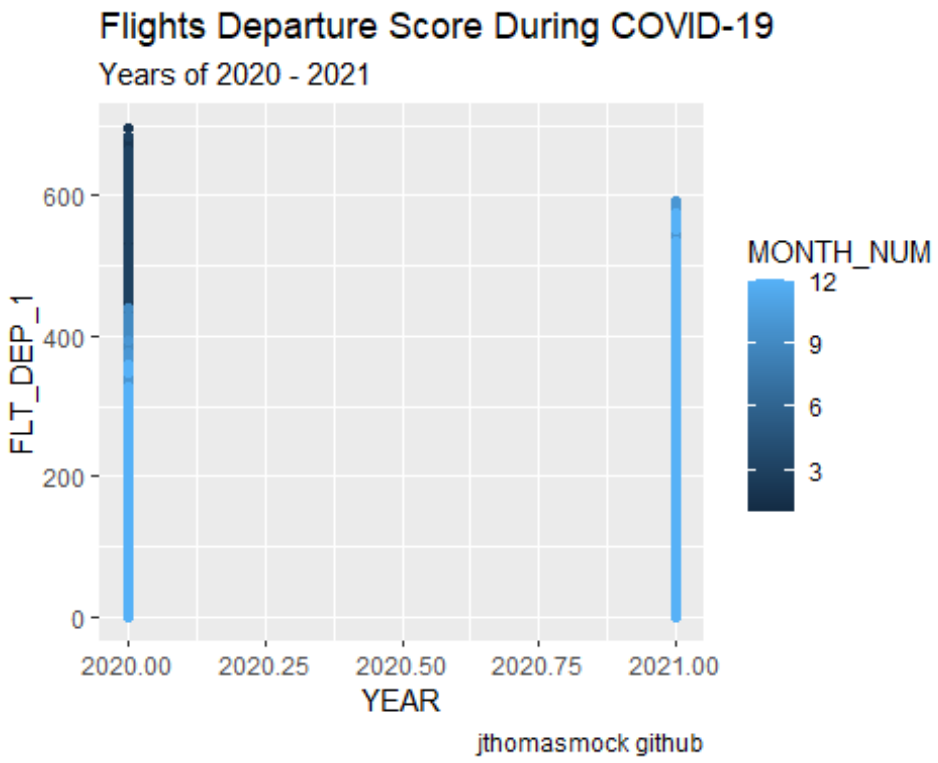
In_Covid Visualization Code 1:

```
ggplot(data = In_Covid, aes(x = YEAR)) +
  geom_density(fill = "yellow", color = "red") +
  labs(x = "Year", y = "Number of Flights per hundred", title = "During COvid
Flights Trend in Europe", caption = "jthomasmock github")
```



In_Covid Visualization Code 2:

```
ggplot(data = In_Covid) + geom_point(mapping = aes(x = YEAR, y = FLT_DEP_1,
color = MONTH_NUM))+labs(title = "Flights Departure Score During COVID-19",
subtitle = "Years of 2020 - 2021",
caption = "jthomasmock github")
```

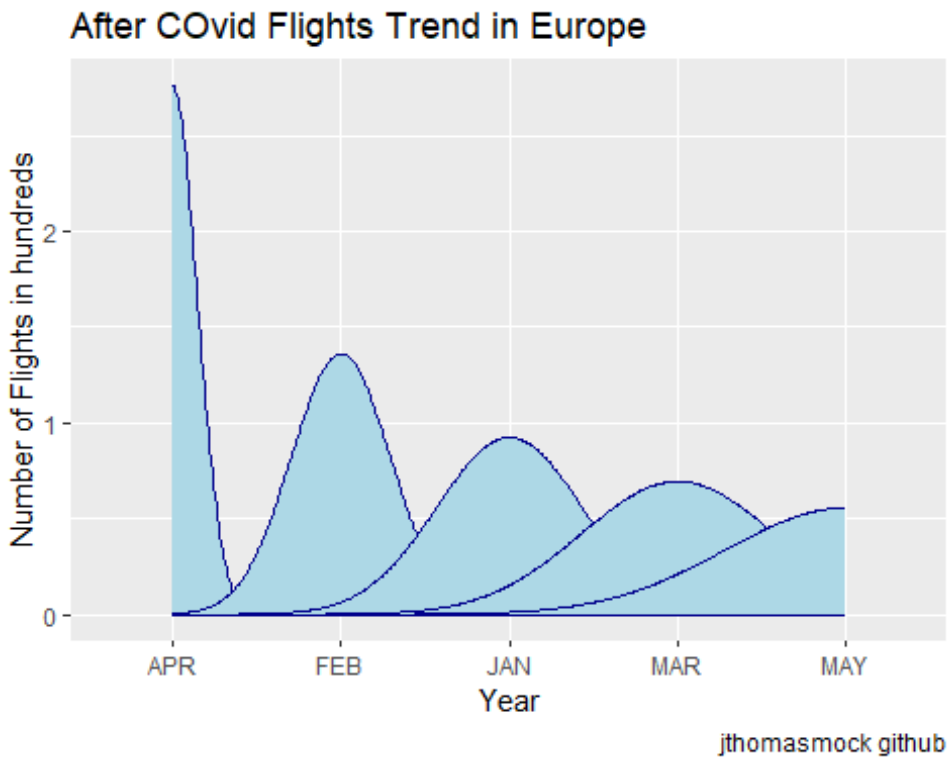


Visualization For Question 3

As for question 3 we need to find out the flights trends After the era of covid-19 that is 2022.SO these visualizations are only depend upon the era after covid-19.

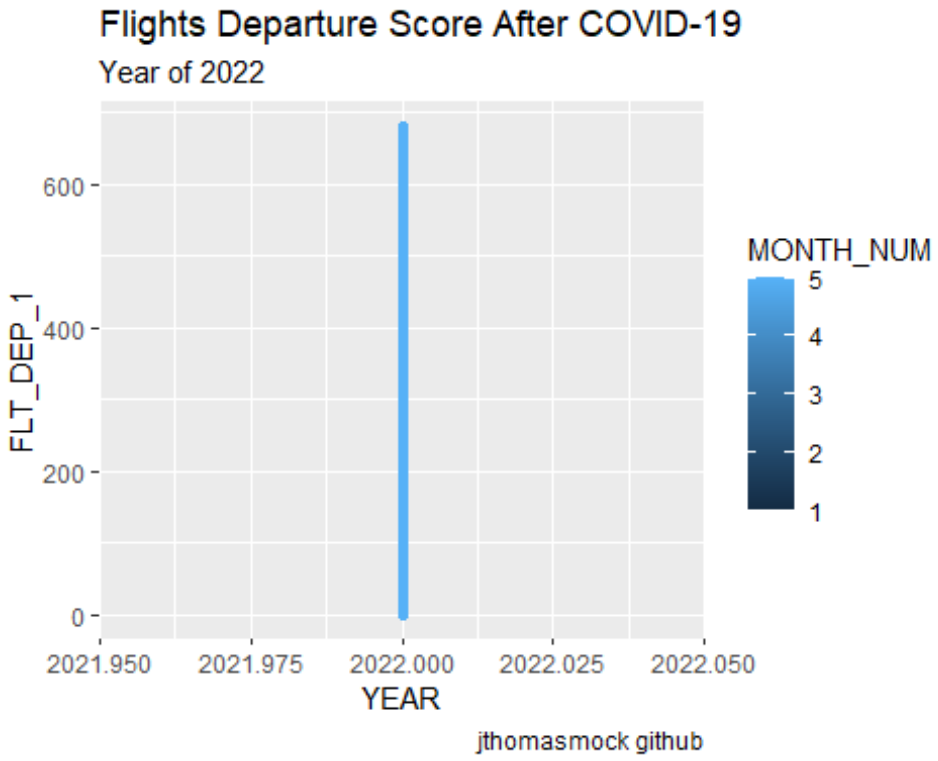
After_Covid Visualization Code 1:

```
ggplot(data = After_Covid, aes(x = MONTH_MON)) +
  geom_density(fill = "lightblue", color = "darkblue") +
  labs(x = "Year", y = "Number of Flights in hundreds", title = "After COvid
Flights Trend in Europe", caption = "jthomasmock github")
```



In_Covid Visualization Code 2:

```
ggplot(data = After_Covid) + geom_point(mapping = aes(x = YEAR, y = FLT_DEP_1
,
color = MONTH_NUM))+labs(title = "Flights Departure Score After COVID-19",
subtitle = "Year of 2022",
caption = "jthomasmock github")
```



Conclusions

for the final conclusion the data insights we get from our analysis is when it comes to **answer** the **Question 1** then in case 2016 to early 2019 the flights are on their usual routine following their schedule and people are travelling every where in the world with out any restriction but in November the first patient of COVID-19 virus was reported in the city of Wuhan and after that the virus took over the whole country in lesser then 3 months due to that reason the We saw a decline in trend in the late 2019 in the operations of flights as China impose emergency and so many flights were cancelled due to COVID-19. In the graph 2 of **Before Covid-19 Visualization** You can see a break which shows that the flights operations were stopped for a while and a few flights were flew after that.

Now in order to give the an answer to the **Question 2** we will take a reference from the answer of Question 1 and continue to answer. As at the end of 2019 the Pandemic took over all of the major countries of the world and in almost every country there was a state of emergency and every government was trying to make their country save and secure, So the flight operations among the high positive percentage COVID countries were suspended and flights operations were facing the effects of COVID. That's why when look on the visualizations of questions 2 you will see a little number of flights of operation throughout the years of 2020 and 2021. As 2020 was the early stage of COVID so in that year you can see more number of flights did their operations as compared to 2021 as In 2021 COVID changes it's form and came with more dangerous variants which Imposed more restrictions on traveling of people from one part of the world to another part of the world.

Now for the final answer of **Question 3** we will discuss the situation of COVID-19 in the year of 2022 as at the end of 2021 scientists were successful in the making of COVID-19 Vaccination and this vaccination was successful against the fight of COVID-19 so the government were ordered to vaccinate the people of their country so that the world can continue it's activity as usual. The people vaccination process was started at the end of 2021 and people got vaccinated at the start of 2022 and the flight operations were coming back on it's usual. when you will look at the graph of you will see the trend starting from to may and you will see a incline and decline from April to March.