SEP
SECRETARÍA DE EDUCACIÓN PÚBLICA

TECNOLÓGICO NACIONAL DE MÉXICO

TIJUANA INSTITUTE OF TECHNOLOGY

ACADEMIC

DEPARTMENT OF SYSTEMS AND COMPUTATION COMPUTER

SYSTEMS ENGINEERING

SEMESTER FEBRUARY- JULY 2022

SUBJECT

BDD-1703SC9C Data Mining

Activity

Evaluation

Teacher

MC JOSE CHRISTIAN ROMERO HERNANDEZ

Student

17212327 -Aldarete Delgado Angel Esteban
17210659 - Villegas Carmona Damaris

Tijuana, BC March 22, 2022

Instructions

Develop the following problem with R and RStudio to extract the knowledge that the problem requires.

The World Bank was very impressed with your delivery on the previous assignment and they have a new project for you.

You must generate a scatter-plot showing the statistics
for life expectancy (Life expectancy - y-axis) and fertility rate (Fertility Rate -x-axis) by country (Country).

The scatterplot should also be classified by Country
Regions.

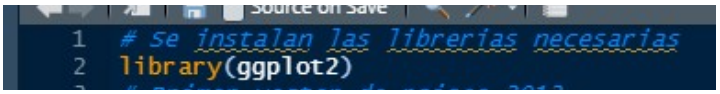
You have been given data for 2 years: 1960 and 2013 and you are required to produce a visualization for each of these years.
Some data has been provided in a CVS, some in R vectors.

file CVS contains combined data for both years. All data manipulation must be done in R (Not Excel) because this project can be audited at a later stage.

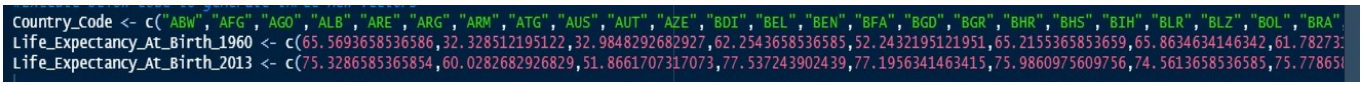
You have also been asked to provide information on how the
two periods compare. (Hint: Basically explaining your observations.)

Development of the exam

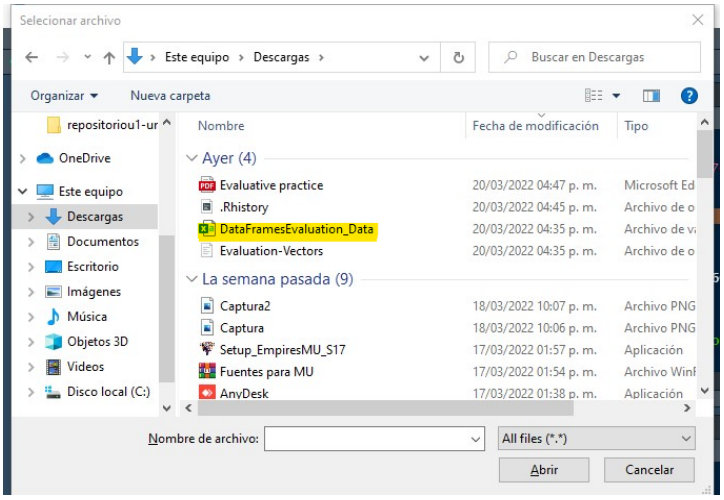
1.- We will start by installing the ggplot library, which is useful for scatter diagrams.

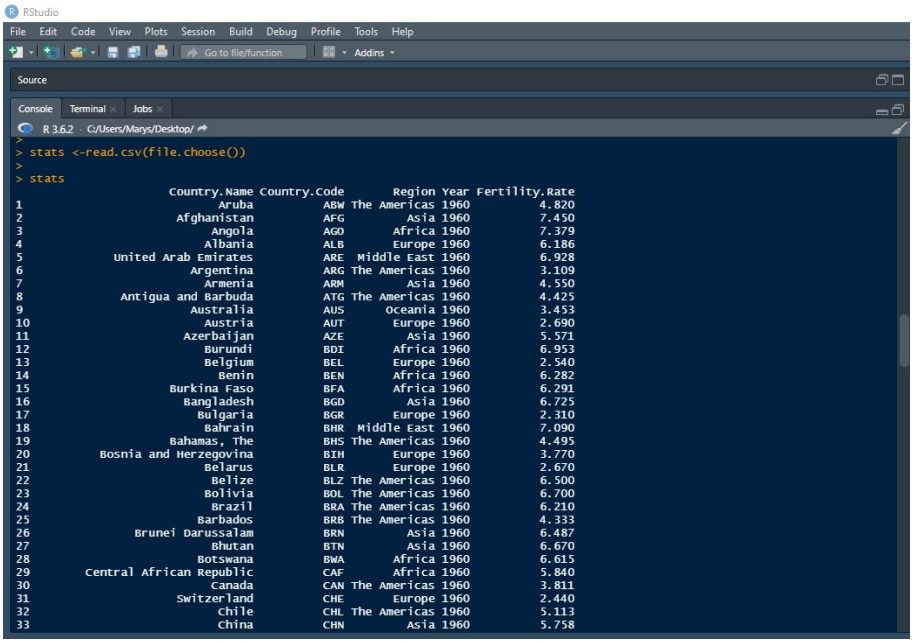


2.- Once the first step is done, we are going to execute the following code to be able to generate the 3 vectors that we need.

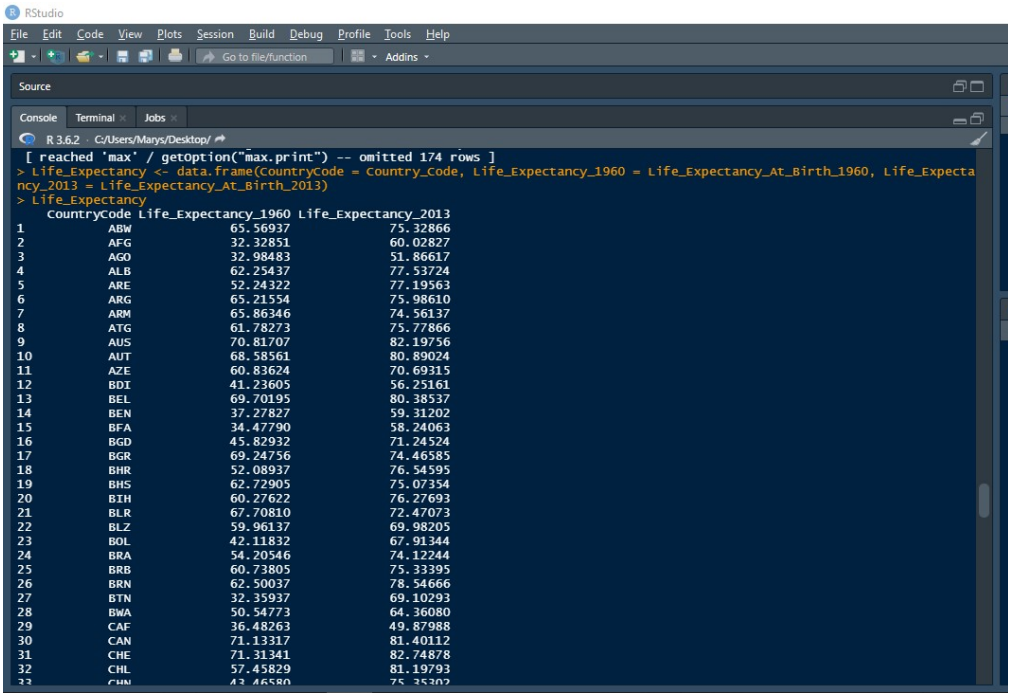


3.- We import the necessary data that was provided from the DataFrameEvaluation_Data.csv.



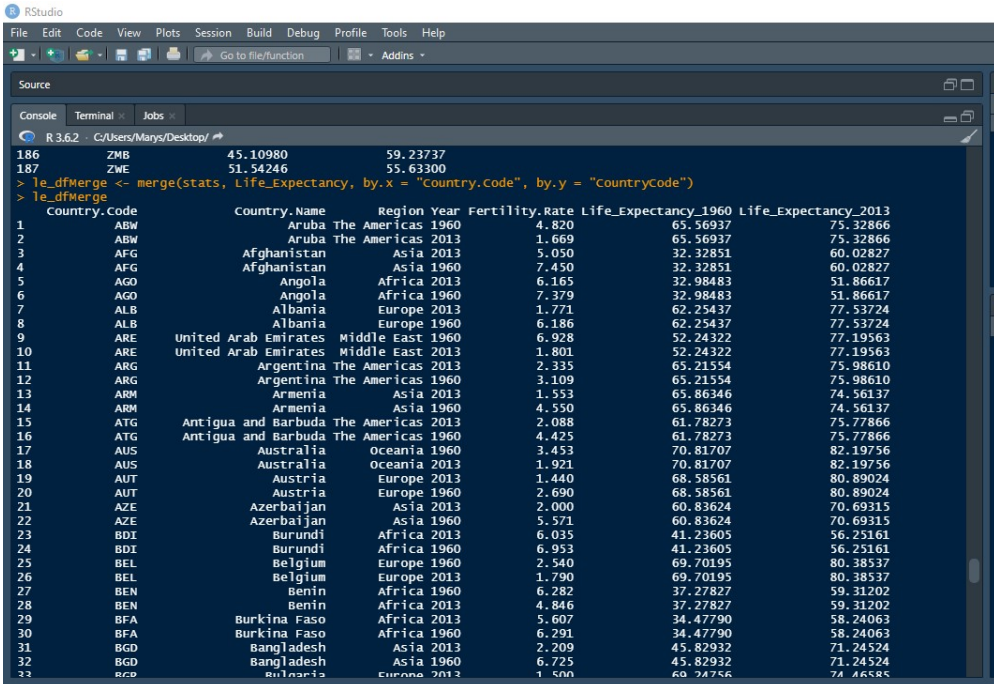


4.-# We generate a new dataframe with the new life expectancy data.



```
 [ reached 'max' / getOption("max.print") -- omitted 174 rows ]
> Life_Expectancy <- data.frame(CountryCode = Country_Code, Life_Expectancy_1960 = Life_Expectancy_At_Birth_1960, Life_Expecta
ncy_2013 = Life_Expectancy_At_Birth_2013)
> Life_Expectancy
    CountryCode Life_Expectancy_1960 Life_Expectancy_2013
1           ABW              65.56937              75.32866
2           AFG              32.32851              60.02827
3           AGO              32.98483              51.86617
4           ALB              62.25437              77.53724
5           ARE              52.24322              77.19563
6           ARG              65.21554              75.98610
7           ARM              65.86346              74.56137
8           ATG              61.78273              75.77866
9           AUS              70.81707              82.19756
10          AUT              68.58561              80.89024
11          AZE              60.83624              70.69315
12          BDI              41.23605              56.25161
13          BEL              69.70195              80.38537
14          BEN              37.27827              59.31202
15          BFA              34.47790              58.24063
16          BGD              45.82932              71.24524
17          BGR              69.24756              74.46585
18          BHR              52.08937              76.54595
19          BHS              62.72905              75.07354
20          BIH              60.27622              76.27693
21          BLR              67.70810              72.47073
22          BLZ              59.96137              69.98205
23          BOL              42.11832              67.91344
24          BRA              54.20546              74.12244
25          BRB              60.73805              75.33395
26          BRN              62.50037              78.54666
27          BTN              32.35937              69.10293
28          BWA              50.54773              64.36080
29          CAF              36.48263              49.87988
30          CAN              71.13317              81.40112
31          CHE              71.31341              82.74878
32          CHL              57.45829              81.19793
33          CHN              43.46580              75.25302
```

5.- # We generate a merge to complement both and create the table.

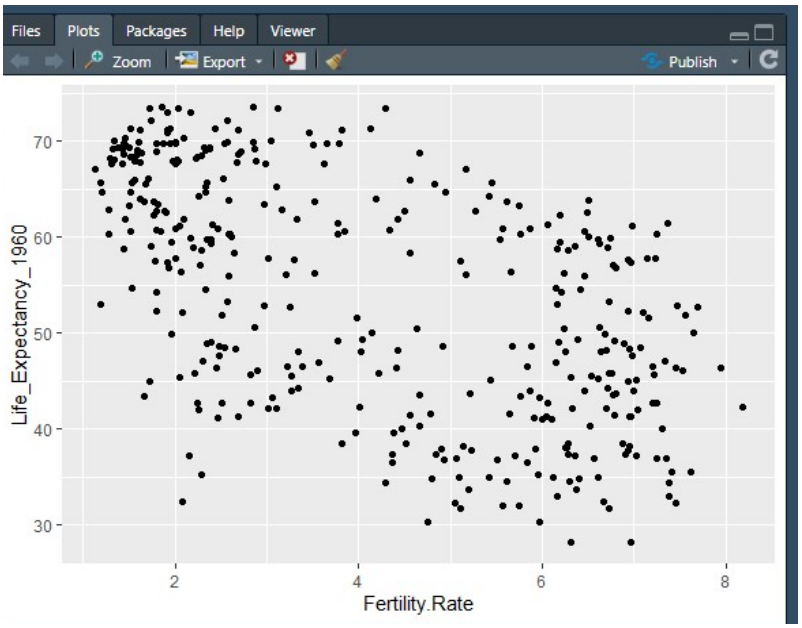


```
186         ZMB              45.10980              59.23737
187         ZWE              51.54246              55.63300
> le_dfMerge <- merge(stats, Life_Expectancy, by.x = "Country.Code", by.y = "CountryCode")
> le_dfMerge
    Country.Code             Country.Name       Region Year Fertility.Rate Life_Expectancy_1960 Life_Expectancy_2013
1            ABW                    Aruba The Americas 1960          4.820              65.56937              75.32866
2            ABW                    Aruba The Americas 2013          1.669              65.56937              75.32866
3            AFG              Afghanistan         Asia 2013          5.050              32.32851              60.02827
4            AFG              Afghanistan         Asia 1960          7.450              32.32851              60.02827
5            AGO                   Angola       Africa 2013          6.165              32.98483              51.86617
6            AGO                   Angola       Africa 1960          7.379              32.98483              51.86617
7            ALB                  Albania       Europe 2013          1.771              62.25437              77.53724
8            ALB                  Albania       Europe 1960          6.186              62.25437              77.53724
9            ARE     United Arab Emirates  Middle East 1960          6.928              52.24322              77.19563
10           ARE     United Arab Emirates  Middle East 2013          1.801              52.24322              77.19563
11           ARG                Argentina The Americas 2013          2.335              65.21554              75.98610
12           ARG                Argentina The Americas 1960          3.109              65.21554              75.98610
13           ARM                  Armenia         Asia 2013          1.553              65.86346              74.56137
14           ARM                  Armenia         Asia 1960          4.550              65.86346              74.56137
15           ATG      Antigua and Barbuda The Americas 2013          2.088              61.78273              75.77866
16           ATG      Antigua and Barbuda The Americas 1960          4.425              61.78273              75.77866
17           AUS                Australia      Oceania 1960          3.453              70.81707              82.19756
18           AUS                Australia      Oceania 2013          1.921              70.81707              82.19756
19           AUT                  Austria       Europe 2013          1.440              68.58561              80.89024
20           AUT                  Austria       Europe 1960          2.690              68.58561              80.89024
21           AZE               Azerbaijan         Asia 2013          2.000              60.83624              70.69315
22           AZE               Azerbaijan         Asia 1960          5.571              60.83624              70.69315
23           BDI                  Burundi       Africa 2013          6.035              41.23605              56.25161
24           BDI                  Burundi       Africa 1960          6.953              41.23605              56.25161
25           BEL                  Belgium       Europe 1960          2.540              69.70195              80.38537
26           BEL                  Belgium       Europe 2013          1.790              69.70195              80.38537
27           BEN                    Benin       Africa 1960          6.282              37.27827              59.31202
28           BEN                    Benin       Africa 2013          4.846              37.27827              59.31202
29           BFA             Burkina Faso       Africa 2013          5.607              34.47790              58.24063
30           BFA             Burkina Faso       Africa 1960          6.291              34.47790              58.24063
31           BGD               Bangladesh         Asia 2013          2.209              45.82932              71.24524
32           BGD               Bangladesh         Asia 1960          6.725              45.82932              71.24524
33           BGR                 Bulgaria       Europe 2013          1.500              69.24756              74.46585
```

6.- The life expectancy of the year 1960

```
> qplot(data = le_dfMerge, y = Life_Expectancy_1960, x = Fertility.Rate)
>
```

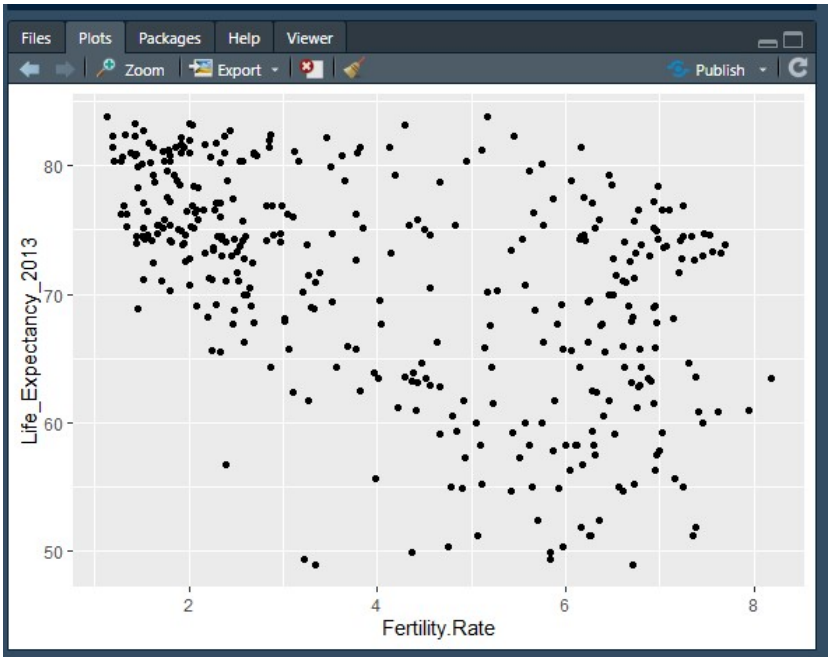
is shown Below you can see the dispersion diagram of the life expectancy of the year 1960.



7.- The life expectancy of the year 2013

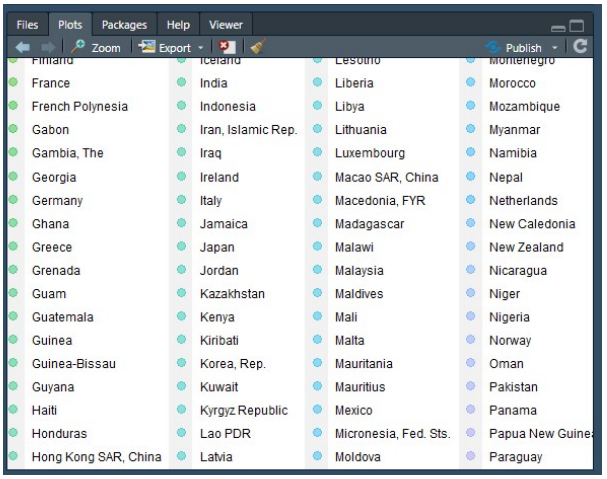
```
> qplot(data = le_dfMerge, y = Life_Expectancy_2013, x = Fertility.Rate)
>
```

is shown Below you can see the diagram of dispersion of life expectancy for the year 2013.



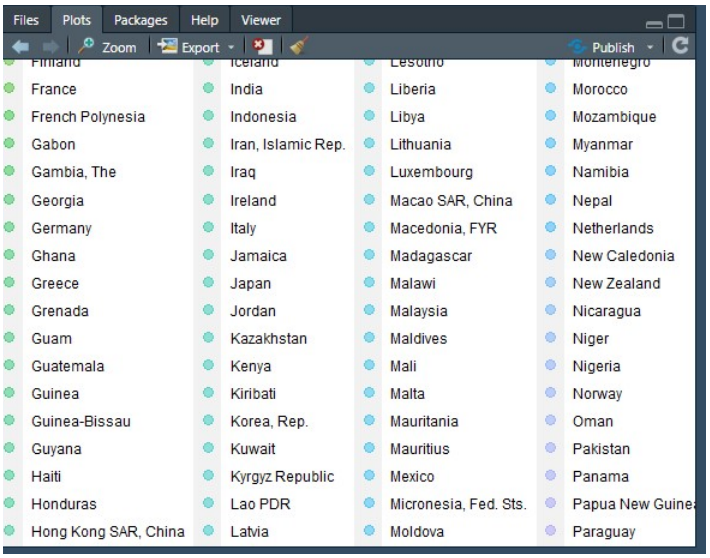
8.- Once this information is obtained, we can generate a graph that relates the percentage of fertility and life expectancy by country and for the year 1960

```
> qplot(data = le_dfMerge, x = Fertility.Rate, y = Life_Expectancy_1960, color = Country.Nam
e, size=I(3), shape=I(19), alpha =I(.4), main = "Fertility Rate vs Life Expectancy by Countr
y in 1960")
>
```



9.- Once the graph of 1960 we proceed to create the one for 2013 comparing the percentage of fertility and life expectancy for each country in the year 2013.

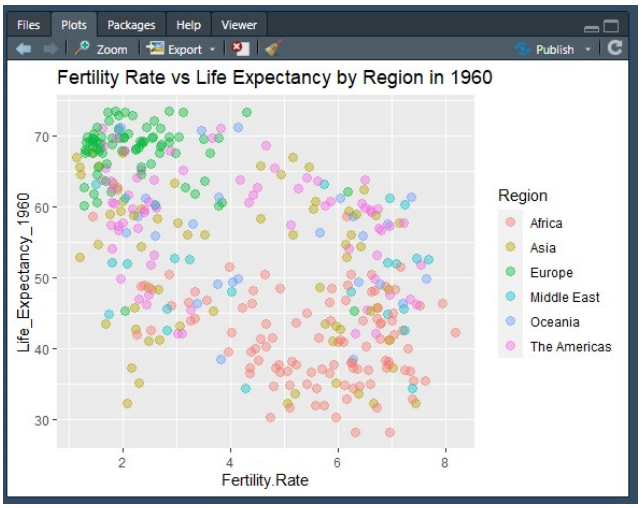
```
> qplot(data = le_dfMerge, x = Fertility.Rate, y = Life_Expectancy_2013, color = Country.Nam
e, size=I(3), shape=I(19), alpha =I(.4), main = "Fertility Rate vs Life Expectancy by Countr
y in 2013")
>
```



10.- Once the graphs for the years 1960 and 2013 for each country have been made, we proceed to create the one for each region for the same years beginning with 1960.

```
> qplot(data = le_dfMerge, x = Fertility.Rate, y = Life_Expectancy_1960, color = Region, siz
e=I(3), shape=I(19), alpha =I(.4), main = "Fertility Rate vs Life Expectancy by Region in 19
60")
>
```

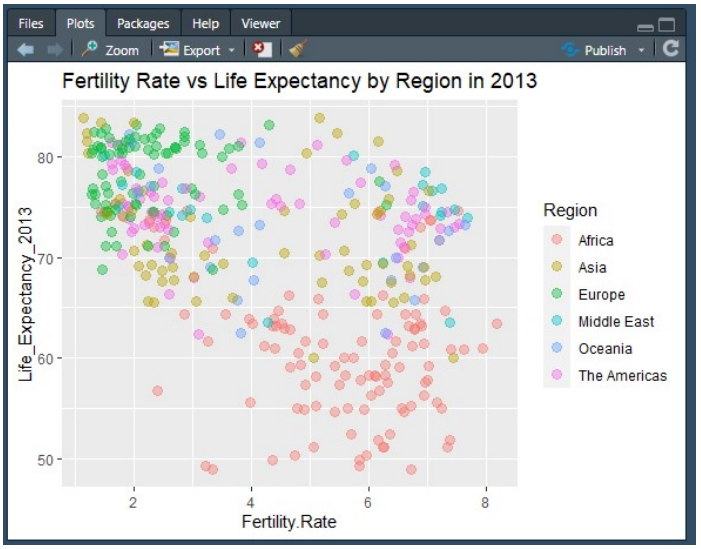
This is the graph of each region comparing the fertility rate with life expectancy in 1960.



11.- Once the graphs for the years 1960 and 2013 for each country have been made, we proceed to create the each of re region for the same years beginning with 2013.

```
60")
> qplot(data = le_dfMerge, x = Fertility.Rate, y = Life_Expectancy_2013, color = Region, siz
e=I(3), shape=I(19), alpha =I(.4), main = "Fertility Rate vs Life Expectancy by Region in 20
13")
>
```

This is the graph of each region comparing the fertility rate with life expectancy in 2013.



Conclusion

With this practice we were able to put into practice what we saw in class, which would be the creation of a dataframe , which we implement to generate new data. At the same time, they generate a merge to be able to complement the tables.

The dispersion diagrams were successfully generated in which the statistics of life expectancy and fertility rate of each country could be visualized.

In order to generate the scatter plots, we rely on the ggplot2 library, which serves as a system to create graphs.

Link del video: https://www.youtube.com/watch?v=ZDwU-jwgmGU
Link del repositorio:
https://github.com/AngelEsteban124020/repositoriou1/tree/unidad1