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

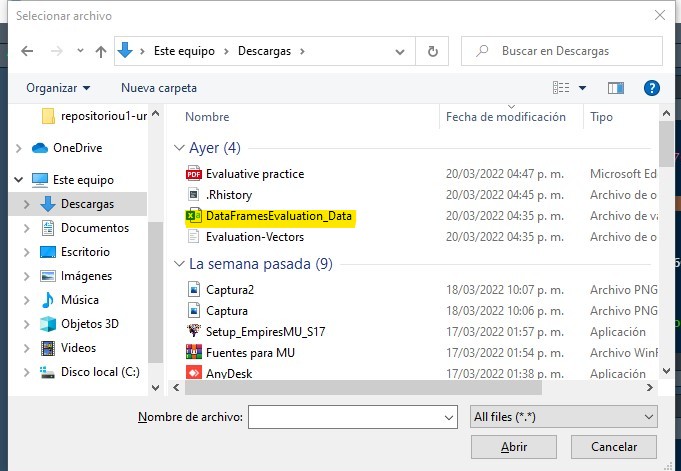
two periods compare. (Hint: Basically explaining your observations.)

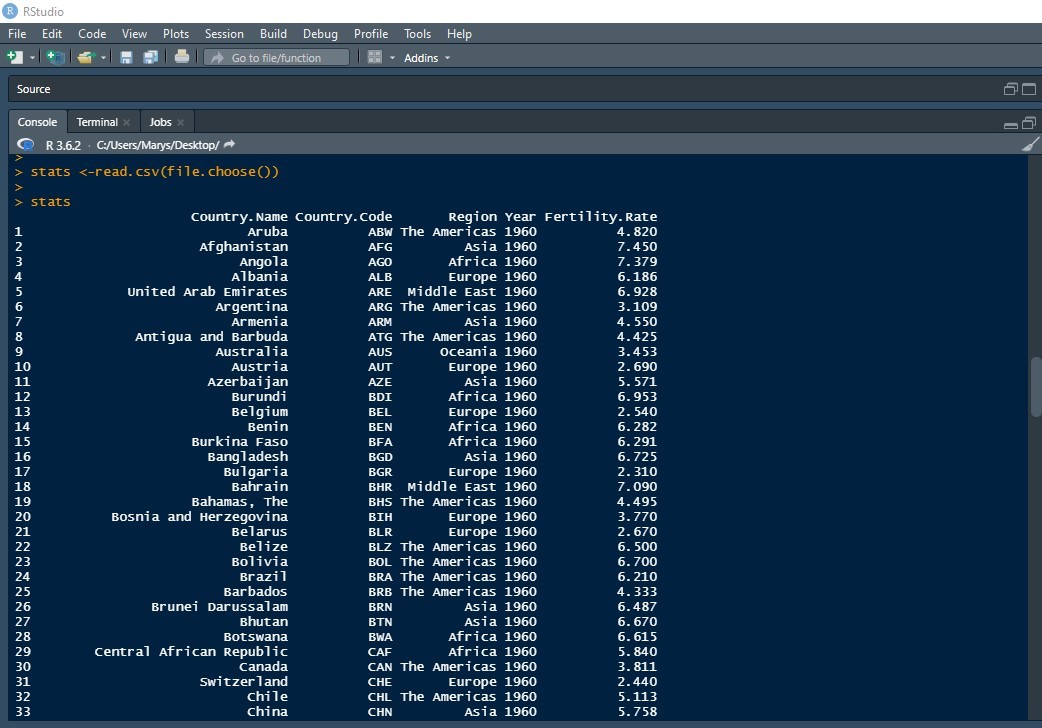
### Development of the exam

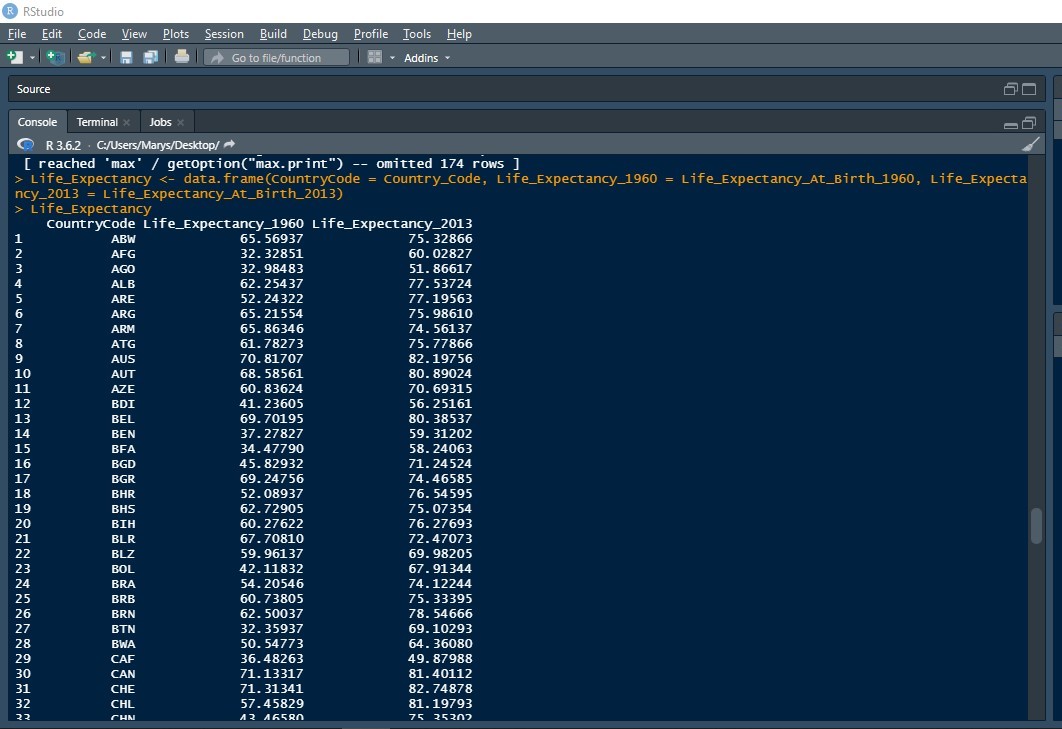
1.- We will start by installing the ggplot library, which is useful for scatter diagrams.

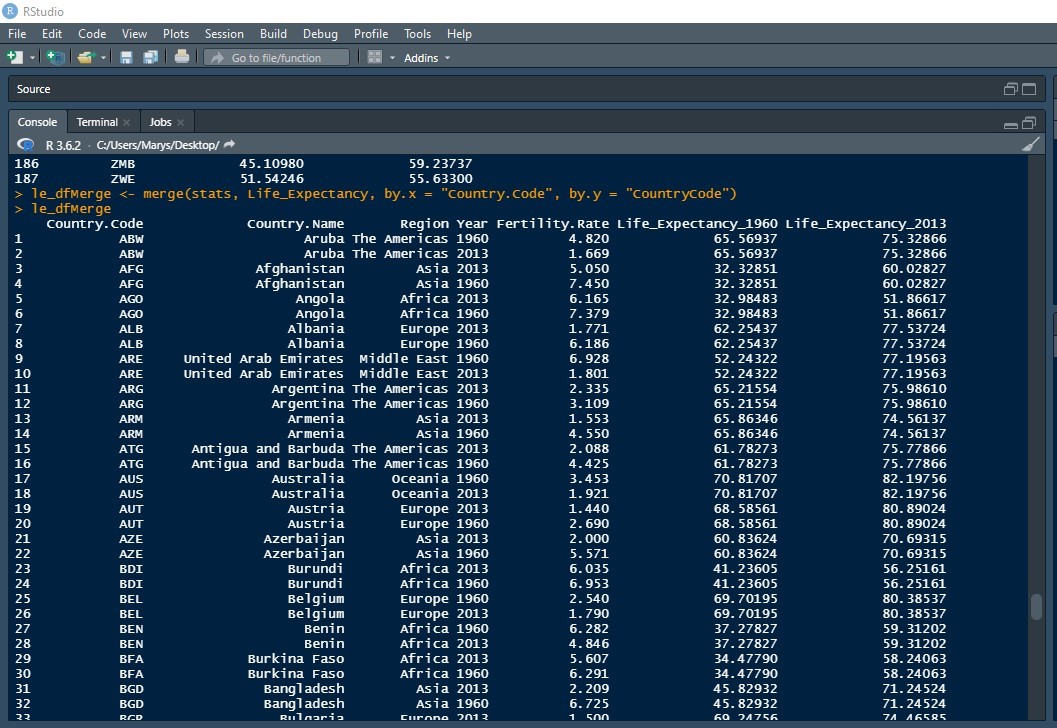


2.- Once the first step is done, we are going to execute the following code to be able to generate the 3 vectors that we need.

3.- We import the necessary data that was provided from the DataFrameEvaluation\_Data.csv.

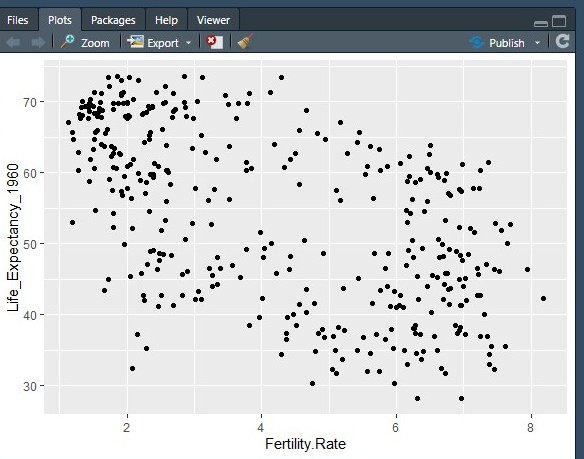


4.-# We generate a new dataframe with the new life expectancy data.

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

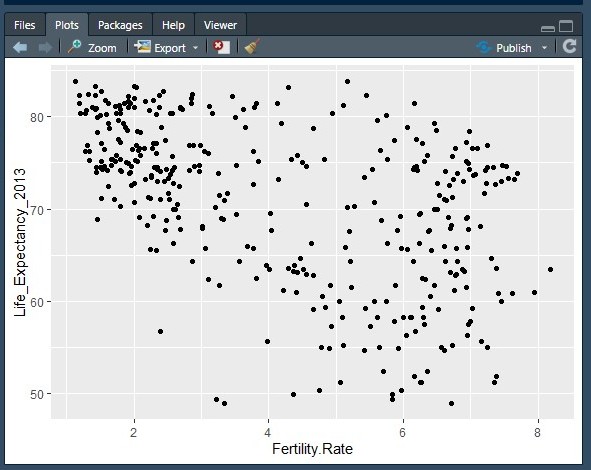
6.- The life expectancy of the year 1960

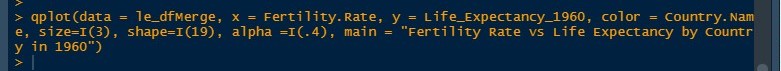
is shown Below you can see the dispersion diagram of the life expectancy of the year 1960.

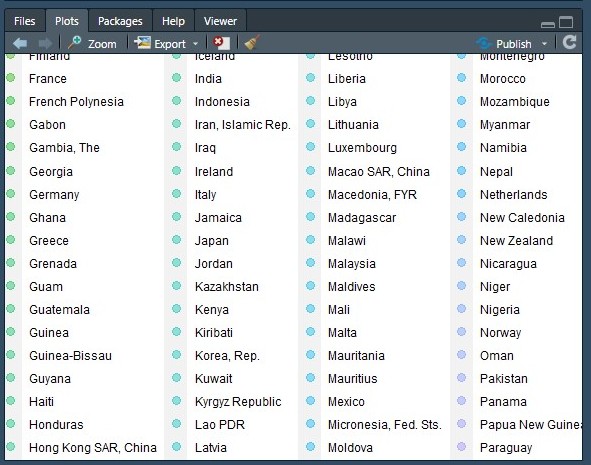


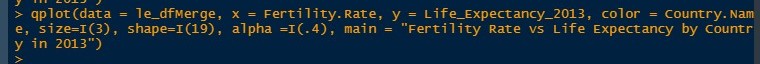
7.- The life expectancy of the year 2013

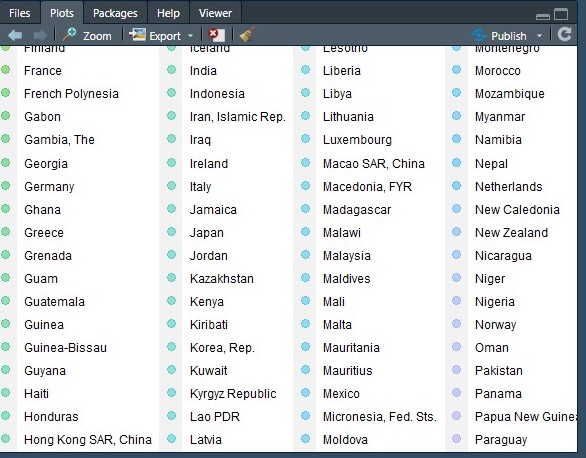
is shown Below you can see the diagram of dispersion of life expectancy for the year 2013.



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

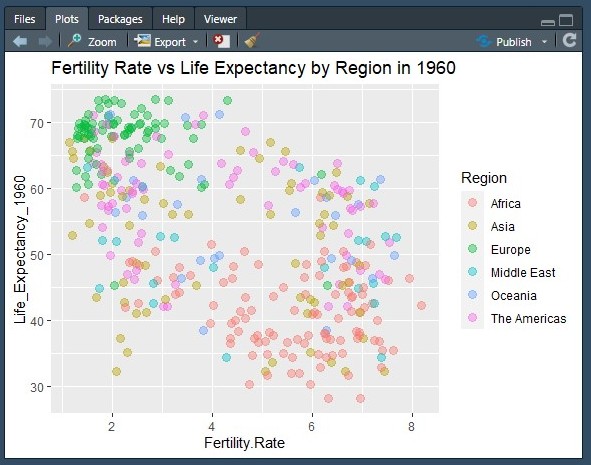


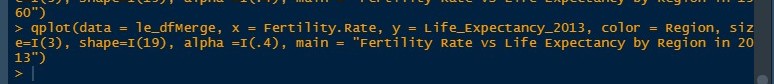
9.- Once the graph of 1960 we proceed to create the one for 2013 comparing the percentage of fertility and life expectancy for each country in the year 2013.

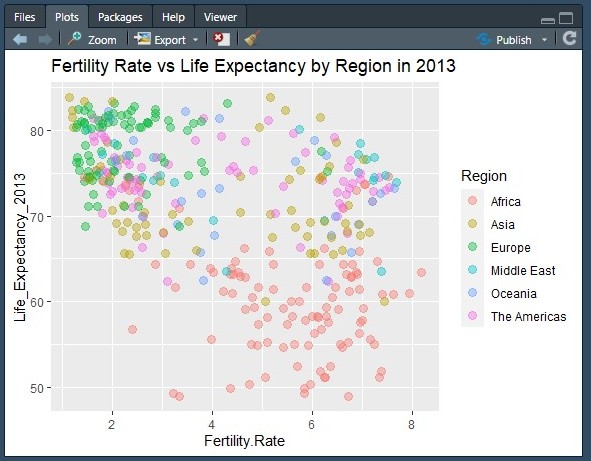


10.- Once the graphs for the years 1960 and 2013 for each country have been made, we proceed to create the one for each region for the same years beginning with 1960.

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.

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**](https://www.youtube.com/watch?v=ZDwU-jwgmGU) **Link del repositorio:** [**https://github.com/AngelEsteban124020/repositoriou1/tree/unidad1**](https://github.com/AngelEsteban124020/repositoriou1/tree/unidad1)