TECNOLÓGICO NACIONAL DE MÉXICO INSTITUTO TECNOLÓGICO DE TIJUANA

SUBDIRECCIÓN ACADÉMICA

DEPARTAMENTO DE SISTEMAS Y COMPUTACIÓN

SEMESTRE FEBRERO-JUNIO 2022

CARRERA
Ingeniería en informática

MATERIA
Minería de datos

TÍTULO
Práctica evaluatoria
Integrantes:
Munguía silva Edgar Geovanny #17212344
Lopez Higuera Saul Alfredo #18210493

NOMBRE DEL MAESTRO
Jose Christian Romero Hernadez

Tijuana Baja California 21 de marzo del 2022

Introduction.

In this unit, we are working with vectors, dataframes and other kinds of data. We will learn how to analyze, describe, find patterns and show relevant information in graphics for an easy output of information. The goal of this evaluation practice is to learn how to use basic R functions, including: creating/importing data frames in CSV format, convert different data type into another one(For example, convert a vector into a data frame),also we will learn to use another kind of functions that are also useful like merge, and last but not least, to learn how to show the patterns we found into a graphical representation. We will show the steps we followed to make the evaluation practice from unit 1 from the course of data mining.

Step #1: Install ggplot library.

```
#1 Instalar la libreria ggplot:
install.packages(
  "ggplot2",
  repos = c("http://rstudio.org/_packages",
            "http://cran.rstudio.com")
)
library(ggplot2)
```

Files	Plots	Packages	Help	Viewer		
🔵 Install	🟢 Update					
	Name		Description			V

User Library

	Name	Description	
☐	cli	Helpers for Developing Command Line Interfaces	2
☐	colorspace	A Toolbox for Manipulating and Assessing Colors and Palettes	2
☐	crayon	Colored Terminal Output	1
☐	digest	Create Compact Hash Digests of R Objects	0
☐	ellipsis	Tools for Working with ...	0
☐	fansi	ANSI Control Sequence Aware String Functions	0
☐	farver	High Performance Colour Space Manipulation	2
☑	ggplot2	Create Elegant Data Visualisations Using the Grammar of Graphics	3
☐	glue	Interpreted String Literals	1
☐	gtable	Arrange 'Grobs' in Tables	0
☐	isoband	Generate Isolines and Isobands from Regularly Spaced Elevation Grids	0
☐	labeling	Axis Labeling	0
☐	lifecycle	Manage the Life Cycle of your Package Functions	1
☐	magrittr	A Forward-Pipe Operator for R	2

Step #2: Import CSV file.

```
#2 Importación de archivo csv
values <- read.csv(file.choose())
```


Step #3: Generation of 3 new vectors:

```
#3 Generación de 3 nuevos vectores:
Country_Code <- c("ABW","AFG","AGO","ALB","ARE","ARG","ARM","ATG","AUS","AUT","AZE","BDI","BEL","BEN","BFA
Life_Expectancy_At_Birth_1960 <- c(65.5693658536586,32.328512195122,32.9848292682927,62.2543658536585,52.
Life_Expectancy_At_Birth_2013 <- c(75.3286585365854,60.0282682926829,51.8661707317073,77.537243902439,77.
```

Environment	History	Connections	Tutorial		

Import Dataset ▾   192 MiB ▾

R ▾   Global Environment ▾

Data

values	374 obs. of 5 variables

Values

Country_Code	chr [1:187] "ABW" "AFG" "AGO" "ALB" "ARE" "ARG" "ARM" "ATG" "AU...
Life_Expectancy_At_Bir...	num [1:187] 65.6 32.3 33 62.3 52.2 ...
Life_Expectancy_At_Bir...	num [1:187] 75.3 60 51.9 77.5 77.2 ...

Step #4: Convert the above vectors into a data frame for further analysis.

```
#4 Convertir los vectores anteriores en un data frame para su posterior análisis
df2<-data.frame(country=Country_Code,expectancy1960=Life_Expectancy_At_Birth_1960,expectancy2013=Life_Exp
```

	country	expectancy1960	expectancy2013
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

Step #5: Make use of the "merge" function to merge both data frames and create a new data frame.

```
#5 Hacer uso de la función "merge" para mezclar ambos y crear un nuevo data frame.
merge<-merge(values,df2, by.x = "Country.Code",by.y = "country")
merge
```

	Country.Code	Country.Name	Region	Year	Fertility.Rate	expectancy1960	expectancy2013
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

Step #6: Use of qplot function to show the life expectancy of the year 1960.

```
#6 Hacer uso de qplot para mostrar la expectativa de vida del año 1960
qplot(data= merge, y = expectancy1960, x = Fertility.Rate,color = Region, size=I(3), shape=I(19), alpha
```
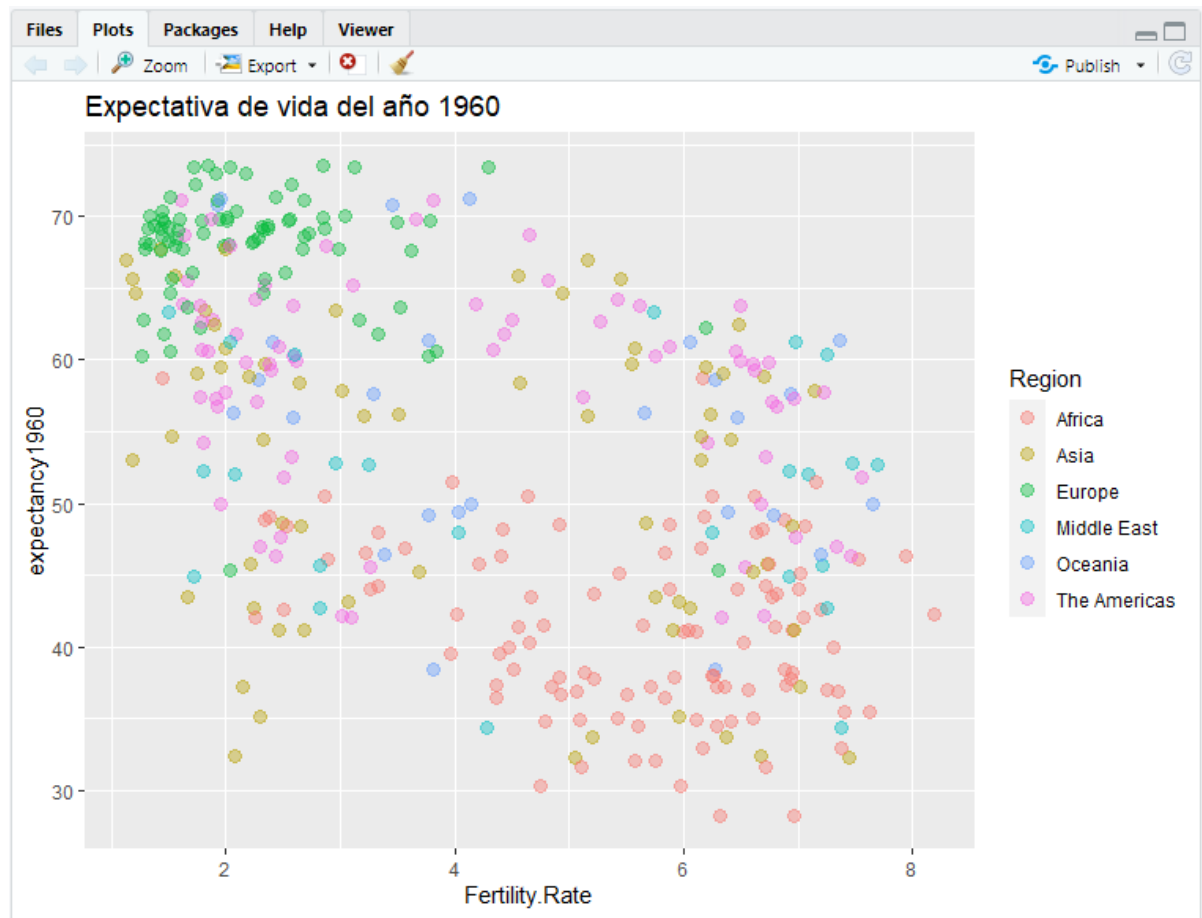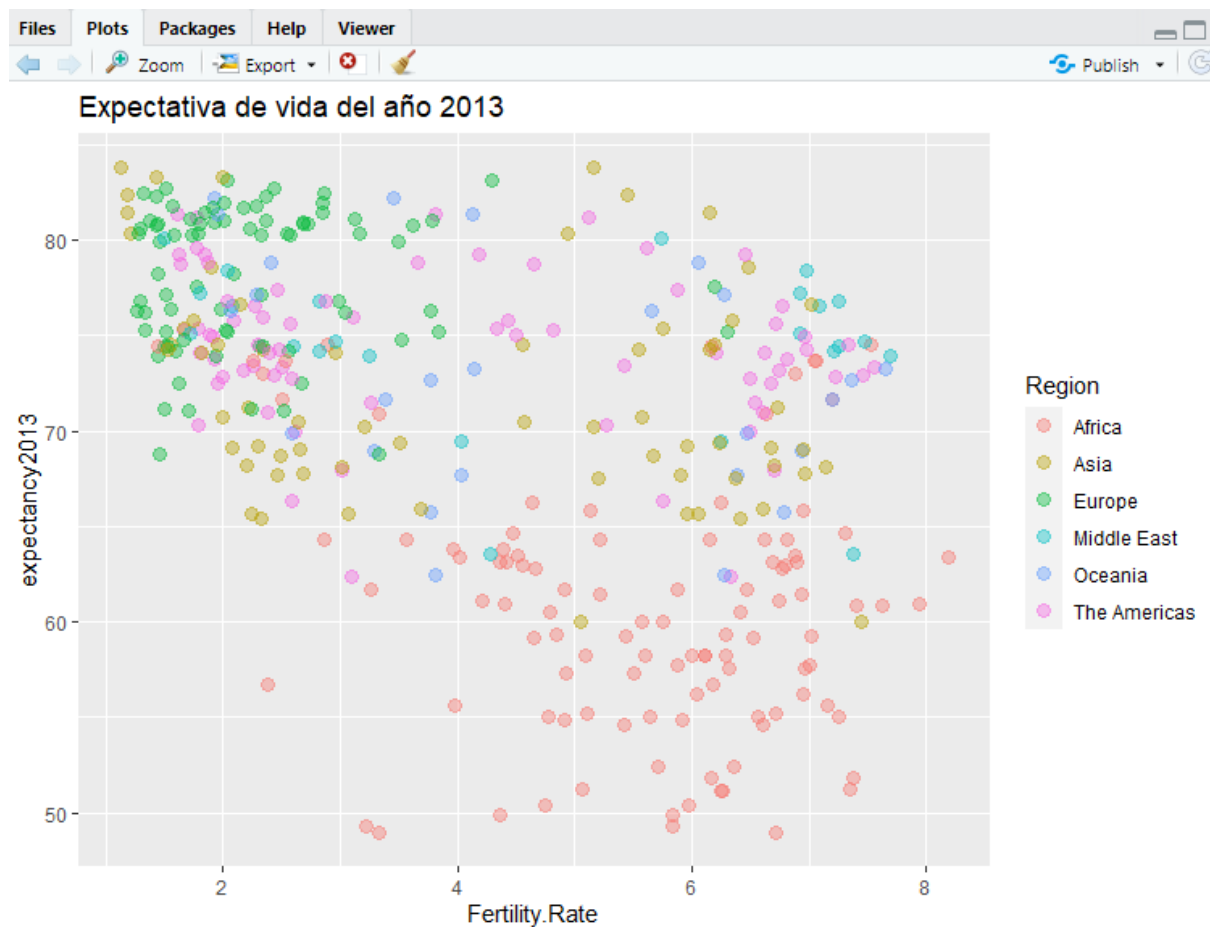
Step #7: Basically the same as above, but we are showing life expectancy from 2013.

```
#7 Lo mismo, pero ahora del año 2013
qplot(data= merge, y = expectancy2013, x = Fertility.Rate,color = Region, size=I(3), shape=I(19), alpha =
```


Expectativa de vida del año 2013

Conclusions.

Edgar Munguia: Well, it's incredible the things we can do using data mining with a lot of data that seems to be "random". We can sort, identify patterns and show relevant information about something. At first, when i saw the files, i thought the data in there was just about of random things, but after using the data mining software (R studio) i realized that i was wrong. I learned a lot of things this unit, basically, i learnt to work with dataframes, convert different data type into another data, i learnt also to analyse infomation, and show them in different graphical ways and a lot of thing more.

Saul Higuera: I find it very impressive what can be achieved in Rstudio, how we can create vectors and convert those same vectors into dataframes to be able to use

them in graphs, without a doubt in this unit I learned many new things, including how to play with the data and graph them.

Also, I had no idea that we could import csv files, or even install packages via console.

Github repository link:
https://github.com/Saul12344/mineria-de-datos

Evidence video link (Youtube): https://youtu.be/7W7lCVOGI64