



**National Technology of Mexico
Technological Institute of Tijuana**

ACADEMIC SUBDIRECTION
Systems and Computing Department

SEMESTER
February - June 2020

ACADEMIC CAREER
Information and Communication Technologies Engineer

SUBJECT AND KEY:
Data Mining BDD-1703TI9A

STUDENT'S NAME AND REGISTRATION:
Velázquez Farrera César Alejandro 17212937
Camacho Manabe Juan Daniel 17210534

NAME OF THE JOB:
Practice #2

UNIT TO BE EVALUATED
Unit I

TEACHER'S NAME:
Mc José Christian Romero Hernández

Due date: April 12, 2021

Practice #5

#Context

You are employed as a data scientist at the World Bank and are working on a project. To analyze the demographic trends of the world.

#Select data file

```
# Method 1: Select the file manually
#Demographic-Data.csv
stats <- read.csv(file.choose())
stats
```

Import the library to graph.

```
#install.packages("ggplot2")
library(ggplot2)
```

You should produce a scatterplot illustrating the birth rate and Internet usage by country.

```
#-----Visualizing With new Split
qplot(data = stats, x = Internet.users, y = Birth.rate,
      color = Country.Name, size=I(3), shape=I(19), alpha =I(.4),
      main = "Birth Rate vs Internet Users by Country")
```

It could also be graphed as follows, but due to the amount of data the chart is very broad, it would have to be subdivided into countries by continent.

```
ggplot(stats, aes( x = Country.Name, y = Birth.rate,
                  color = Internet.users, size=I(3), shape=I(19), alpha =I(.4),
                  main = "Birth Rate vs Internet Users by Country")) + geom_point()
```

The scatter plot should also be classified by income groups of the countries.

```
qplot(data=stats, x=Internet.users, y=Birth.rate, color=Income.Group,
      size=I(1))
```

#Practice #5

We install the dplyr library to be able to use the filter () function

```
install.packages('dplyr')  
library("dplyr")
```

The filter function is used to create subsets of a data frame. The function requires two parameters:

1. data - The data to filter.
2. Condition - Used to filter the data.

#Filter countries by low income

```
filter(stats, Income.Group == "Low income")
```

Filter countries by lower middle income

```
filter(stats, Income.Group == "Lower middle income")
```

Filter countries by upper middle income

```
filter(stats, Income.Group == "Upper middle income")
```

Filter by country Malta

```
filter(stats, Country.Name == "Malta")
```

Filter by country Qatar

```
filter(stats, Country.Name == "Qatar")
```

Filter by country Netherlands

```
filter(stats, Country.Name == "Países Bajos")
```

Filter by country Norway

```
filter(stats, Country.Name == "Noruega")
```

You should produce a second illustration of the Birth Rate and Internet Usage Statistics by Country scatterplot.

```
qplot(data = stats, x = Internet.users, y = Birth.rate,  
      color = Country.Name, size=I(3), shape=I(19), alpha =I(.4),  
      main = "Birth Rate vs Internet Users by Country")
```