Instituto Tecnológico de Tijuana

Ingeniería en Sistemas Computacionales

Examen I

Materia: Minería de Datos

Unidad: Unidad I

Facilitador:

José Christian Romero Sánchez

Alumnos:

- Hernández Negrete Juan Carlos **- 16212021**
- Sifuentes Martinez Manuel Javier **- 17212934**

Fecha:

Tijuana Baja California a 13 de Abril del 2021.

Introduction.

For this evaluative practice, the topics seen in class will be applied, such as the introduction to the R language, and how to program in it, make use of variables, matrices, datasets, vectors, functions, etc. For this, the matplot installation will also be applied to be able to make the corresponding graphs and display the dataframes.

Development.

EVALUATION PRACTICE OF UNIT 1

Data

- Countries Vector

```
Countries_2012_Dataset <-
c("Aruba","Afghanistan","Angola","Albania","United Arab
Emirates","Argentina","Armenia","Antigua and
Barbuda","Australia","Austria","Azerbaijan","Burundi","Belgium","Benin","B
urkina Faso","Bangladesh","Bulgaria","Bahrain","Bahamas, The","Bosnia and
Herzegovina","Belarus","Belize","Bermuda","Bolivia","Brazil","Barbados","B
runei Darussalam","Bhutan","Botswana","Central African
Republic","Canada","Switzerland","Chile","China","Cote
d'Ivoire","Cameroon","Congo, Rep.","Colombia","Comoros","Cabo
Verde","Costa Rica","Cuba","Cayman Islands","Cyprus","Czech
Republic","Germany","Djibouti","Denmark","Dominican
Republic","Algeria","Ecuador","Egypt, Arab
Rep.","Eritrea","Spain","Estonia","Ethiopia","Finland","Fiji","France","Mi
cronesia, Fed. Sts.","Gabon","United
Kingdom","Georgia","Ghana","Guinea","Gambia,
The","Guinea-Bissau","Equatorial
Guinea","Greece","Grenada","Greenland","Guatemala","Guam","Guyana","Hong
Kong SAR,
China","Honduras","Croatia","Haiti","Hungary","Indonesia","India","Ireland
","Iran, Islamic
Rep.","Iraq","Iceland","Israel","Italy","Jamaica","Jordan","Japan","Kazakh
stan","Kenya","Kyrgyz Republic","Cambodia","Kiribati","Korea,
Rep.","Kuwait","Lao PDR","Lebanon","Liberia","Libya","St.
Lucia","Liechtenstein","Sri
Lanka","Lesotho","Lithuania","Luxembourg","Latvia","Macao SAR,
China","Morocco","Moldova","Madagascar","Maldives","Mexico","Macedonia,
FYR","Mali","Malta","Myanmar","Montenegro","Mongolia","Mozambique","Maurit
ania","Mauritius","Malawi","Malaysia","Namibia","New
Caledonia","Niger","Nigeria","Nicaragua","Netherlands","Norway","Nepal","N
ew Zealand","Oman","Pakistan","Panama","Peru","Philippines","Papua New
Guinea","Poland","Puerto Rico","Portugal","Paraguay","French
Polynesia","Qatar","Romania","Russian Federation","Rwanda","Saudi
Arabia","Sudan","Senegal","Singapore","Solomon Islands","Sierra Leone","El
Salvador","Somalia","Serbia","South Sudan","Sao Tome and
Principe","Suriname","Slovak
Republic","Slovenia","Sweden","Swaziland","Seychelles","Syrian Arab
Republic","Chad","Togo","Thailand","Tajikistan","Turkmenistan","Timor-Lest
e","Tonga","Trinidad and
```

```
Tobago","Tunisia","Turkey","Tanzania","Uganda","Ukraine","Uruguay","United
States","Uzbekistan","St. Vincent and the Grenadines","Venezuela,
RB","Virgin Islands (U.S.)","Vietnam","Vanuatu","West Bank and
Gaza","Samoa","Yemen, Rep.","South Africa","Congo, Dem.
Rep.","Zambia","Zimbabwe")
```

- Code Vector

```
Codes_2012_Dataset <-
c("ABW","AFG","AGO","ALB","ARE","ARG","ARM","ATG","AUS","AUT","AZE","BDI",
"BEL","BEN","BFA","BGD","BGR","BHR","BHS","BIH","BLR","BLZ","BMU","BOL","B
RA","BRB","BRN","BTN","BWA","CAF","CAN","CHE","CHL","CHN","CIV","CMR","COG
","COL","COM","CPV","CRI","CUB","CYM","CYP","CZE","DEU","DJI","DNK","DOM",
"DZA","ECU","EGY","ERI","ESP","EST","ETH","FIN","FJI","FRA","FSM","GAB","G
BR","GEO","GHA","GIN","GMB","GNB","GNQ","GRC","GRD","GRL","GTM","GUM","GUY
","HKG","HND","HRV","HTI","HUN","IDN","IND","IRL","IRN","IRQ","ISL","ISR",
"ITA","JAM","JOR","JPN","KAZ","KEN","KGZ","KHM","KIR","KOR","KWT","LAO","L
BN","LBR","LBY","LCA","LIE","LKA","LSO","LTU","LUX","LVA","MAC","MAR","MDA
","MDG","MDV","MEX","MKD","MLI","MLT","MMR","MNE","MNG","MOZ","MRT","MUS",
"MWI","MYS","NAM","NCL","NER","NGA","NIC","NLD","NOR","NPL","NZL","OMN","P
AK","PAN","PER","PHL","PNG","POL","PRI","PRT","PRY","PYF","QAT","ROU","RUS
","RWA","SAU","SDN","SEN","SGP","SLB","SLE","SLV","SOM","SRB","SSD","STP",
"SUR","SVK","SVN","SWE","SWZ","SYC","SYR","TCD","TGO","THA","TJK","TKM","T
LS","TON","TTO","TUN","TUR","TZA","UGA","UKR","URY","USA","UZB","VCT","VEN
","VIR","VNM","VUT","PSE","WSM","YEM","ZAF","COD","ZMB","ZWE")
```

- Region Vector

```
Regions_2012_Dataset <- c("The Americas","Asia","Africa","Europe","Middle
East","The Americas","Asia","The
Americas","Oceania","Europe","Asia","Africa","Europe","Africa","Africa","A
sia","Europe","Middle East","The Americas","Europe","Europe","The
Americas","The Americas","The Americas","The Americas","The
Americas","Asia","Asia","Africa","Africa","The Americas","Europe","The
Americas","Asia","Africa","Africa","Africa","The
Americas","Africa","Africa","The Americas","The Americas","The
Americas","Europe","Europe","Europe","Africa","Europe","The
Americas","Africa","The
Americas","Africa","Africa","Europe","Europe","Africa","Europe","Oceania",
"Europe","Oceania","Africa","Europe","Asia","Africa","Africa","Africa","Af
rica","Africa","Europe","The Americas","The Americas","The
Americas","Oceania","The Americas","Asia","The Americas","Europe","The
Americas","Europe","Asia","Asia","Europe","Middle East","Middle
East","Europe","Middle East","Europe","The Americas","Middle
```

```
East","Asia","Asia","Africa","Asia","Asia","Oceania","Asia","Middle
East","Asia","Middle East","Africa","Africa","The
Americas","Europe","Asia","Africa","Europe","Europe","Europe","Asia","Afri
ca","Europe","Africa","Asia","The
Americas","Europe","Africa","Europe","Asia","Europe","Asia","Africa","Afri
ca","Africa","Africa","Asia","Africa","Oceania","Africa","Africa","The
Americas","Europe","Europe","Asia","Oceania","Middle East","Asia","The
Americas","The Americas","Asia","Oceania","Europe","The
Americas","Europe","The Americas","Oceania","Middle
East","Europe","Europe","Africa","Middle
East","Africa","Africa","Asia","Oceania","Africa","The
Americas","Africa","Europe","Africa","Africa","The
Americas","Europe","Europe","Europe","Africa","Africa","Middle
East","Africa","Africa","Asia","Asia","Asia","Asia","Oceania","The
Americas","Africa","Europe","Africa","Africa","Europe","The Americas","The
Americas","Asia","The Americas","The Americas","The
Americas","Asia","Oceania","Middle East","Oceania","Middle
East","Africa","Africa","Africa","Africa")
```

- Life Expectancy Vectors

```
Country_Code <-
c("ABW","AFG","AGO","ALB","ARE","ARG","ARM","ATG","AUS","AUT","AZE","BDI",
"BEL","BEN","BFA","BGD","BGR","BHR","BHS","BIH","BLR","BLZ","BOL","BRA","B
RB","BRN","BTN","BWA","CAF","CAN","CHE","CHL","CHN","CIV","CMR","COG","COL
","COM","CPV","CRI","CUB","CYP","CZE","DEU","DJI","DNK","DOM","DZA","ECU",
"EGY","ERI","ESP","EST","ETH","FIN","FJI","FRA","FSM","GAB","GBR","GEO","G
HA","GIN","GMB","GNB","GNQ","GRC","GRD","GTM","GUM","GUY","HKG","HND","HRV
","HTI","HUN","IDN","IND","IRL","IRN","IRQ","ISL","ITA","JAM","JOR","JPN",
"KAZ","KEN","KGZ","KHM","KIR","KOR","KWT","LAO","LBN","LBR","LBY","LCA","L
KA","LSO","LTU","LUX","LVA","MAC","MAR","MDA","MDG","MDV","MEX","MKD","MLI
","MLT","MMR","MNE","MNG","MOZ","MRT","MUS","MWI","MYS","NAM","NCL","NER",
"NGA","NIC","NLD","NOR","NPL","NZL","OMN","PAK","PAN","PER","PHL","PNG","P
OL","PRI","PRT","PRY","PYF","QAT","ROU","RUS","RWA","SAU","SDN","SEN","SGP
","SLB","SLE","SLV","SOM","SSD","STP","SUR","SVK","SVN","SWE","SWZ","SYR",
"TCD","TGO","THA","TJK","TKM","TLS","TON","TTO","TUN","TUR","TZA","UGA","U
KR","URY","USA","UZB","VCT","VEN","VIR","VNM","VUT","WSM","YEM","ZAF","COD
","ZMB","ZWE")
```

```
Life_Expectancy_At_Birth_1960 <-
c(65.5693658536586,32.328512195122,32.9848292682927,62.2543658536585,52.24
32195121951,65.2155365853659,65.8634634146342,61.7827317073171,70.81707317
07317,68.5856097560976,60.836243902439,41.2360487804878,69.7019512195122,3
7.2782682926829,34.4779024390244,45.8293170731707,69.2475609756098,52.0893
658536585,62.7290487804878,60.2762195121951,67.7080975609756,59.9613658536
```

```
585,42.1183170731707,54.2054634146342,60.7380487804878,62.5003658536585,32
.3593658536585,50.5477317073171,36.4826341463415,71.1331707317073,71.31341
46341463,57.4582926829268,43.4658048780488,36.8724146341463,41.52375609756
1,48.5816341463415,56.716756097561,41.4424390243903,48.8564146341463,60.57
61951219512,63.9046585365854,69.5939268292683,70.3487804878049,69.31295121
95122,44.0212682926829,72.1765853658537,51.8452682926829,46.1351219512195,
53.215,48.0137073170732,37.3629024390244,69.1092682926829,67.9059756097561
,38.4057073170732,68.819756097561,55.9584878048781,69.8682926829268,57.586
5853658537,39.5701219512195,71.1268292682927,63.4318536585366,45.831463414
6342,34.8863902439024,32.0422195121951,37.8404390243902,36.7330487804878,6
8.1639024390244,59.8159268292683,45.5316341463415,61.2263414634146,60.2787
317073171,66.9997073170732,46.2883170731707,64.6086585365854,42.1000975609
756,68.0031707317073,48.6403170731707,41.1719512195122,69.691756097561,44.
945512195122,48.0306829268293,73.4286585365854,69.1239024390244,64.1918292
682927,52.6852682926829,67.6660975609756,58.3675853658537,46.3624146341463
,56.1280731707317,41.2320243902439,49.2159756097561,53.0013170731707,60.34
79512195122,43.2044634146342,63.2801219512195,34.7831707317073,42.64119512
19512,57.303756097561,59.7471463414634,46.5107073170732,69.8473170731707,6
8.4463902439024,69.7868292682927,64.6609268292683,48.4466341463415,61.8127
804878049,39.9746829268293,37.2686341463415,57.0656341463415,60.6228048780
488,28.2116097560976,67.6017804878049,42.7363902439024,63.7056097560976,48
.3688048780488,35.0037073170732,43.4830975609756,58.7452195121951,37.77363
41463415,59.4753414634146,46.8803902439024,58.6390243902439,35.51504878048
78,37.1829512195122,46.9988292682927,73.3926829268293,73.549756097561,35.1
708292682927,71.2365853658537,42.6670731707317,45.2904634146342,60.8817073
170732,47.6915853658537,57.8119268292683,38.462243902439,67.6804878048781,
68.7196097560976,62.8089268292683,63.7937073170732,56.3570487804878,61.206
0731707317,65.6424390243903,66.0552926829268,42.2492926829268,45.666268292
6829,48.1876341463415,38.206,65.6598292682927,49.3817073170732,30.33153658
53659,49.9479268292683,36.9658780487805,31.6767073170732,50.4513658536585,
59.6801219512195,69.9759268292683,68.9780487804878,73.0056097560976,44.233
7804878049,52.768243902439,38.0161219512195,40.2728292682927,54.6993170731
707,56.1535365853659,54.4586829268293,33.7271219512195,61.3645365853659,62
.6575853658537,42.009756097561,45.3844146341463,43.6538780487805,43.983560
9756098,68.2995365853659,67.8963902439025,69.7707317073171,58.885536585365
9,57.7238780487805,59.2851219512195,63.7302195121951,59.0670243902439,46.4
874878048781,49.969512195122,34.3638048780488,49.0362926829268,41.01804878
04878,45.1098048780488,51.5424634146342)
```

```
Life_Expectancy_At_Birth_2013 <-
c(75.3286585365854,60.0282682926829,51.8661707317073,77.537243902439,77.19
56341463415,75.9860975609756,74.5613658536585,75.7786585365854,82.19756097
56098,80.890243902439,70.6931463414634,56.2516097560976,80.3853658536585,5
9.3120243902439,58.2406341463415,71.245243902439,74.4658536585366,76.54595
12195122,75.0735365853659,76.2769268292683,72.4707317073171,69.98204878048
78,67.9134390243903,74.1224390243903,75.3339512195122,78.5466585365854,69.
```

```
1029268292683,64.3608048780488,49.8798780487805,81.4011219512195,82.7487804878049,81.1979268292683,75.3530243902439,51.2084634146342,55.0418048780488,61.6663902439024,73.8097317073171,62.9321707317073,72.9723658536585,79.2252195121951,79.2563902439025,79.9497804878049,78.2780487804878,81.0439024390244,61.6864634146342,80.3024390243903,73.3199024390244,74.5689512195122,75.648512195122,70.9257804878049,63.1778780487805,82.4268292682927,76.4243902439025,63.4421951219512,80.8317073170732,69.9179268292683,81.9682926829268,68.9733902439024,63.8435853658537,80.9560975609756,74.079512195122,61.1420731707317,58.216487804878,59.9992682926829,54.8384146341464,57.2908292682927,80.6341463414634,73.1935609756098,71.4863902439024,78.872512195122,66.3100243902439,83.8317073170732,72.9428536585366,77.1268292682927,62.4011463414634,75.2682926829268,68.7046097560976,67.6604146341463,81.0439024390244,75.1259756097561,69.4716829268293,83.1170731707317,82.290243902439,73.4689268292683,73.9014146341463,83.3319512195122,70.45,60.9537804878049,70.2024390243902,67.7720487804878,65.7665853658537,81.459756097561,74.462756097561,65.687243902439,80.1288780487805,60.5203902439024,71.6576829268293,74.9127073170732,74.2402926829268,49.3314634146342,74.1634146341464,81.7975609756098,73.9804878048781,80.3391463414634,73.7090487804878,68.811512195122,64.6739024390244,76.6026097560976,76.5326585365854,75.1870487804878,57.5351951219512,80.7463414634146,65.6540975609756,74.7583658536585,69.0618048780488,54.641512195122,62.8027073170732,74.46,61.466,74.567512195122,64.3438780487805,77.1219512195122,60.8281463414634,52.4421463414634,74.514756097561,81.1048780487805,81.4512195121951,69.222,81.4073170731707,76.8410487804878,65.9636829268293,77.4192195121951,74.2838536585366,68.1315609756097,62.4491707317073,76.8487804878049,78.7111951219512,80.3731707317073,72.7991707317073,76.3340731707317,78.4184878048781,74.4634146341463,71.0731707317073,63.3948292682927,74.1776341463415,63.1670487804878,65.878756097561,82.3463414634146,67.7189268292683,50.3631219512195,72.4981463414634,55.0230243902439,55.2209024390244,66.259512195122,70.99,76.2609756097561,80.2780487804878,81.7048780487805,48.9379268292683,74.7157804878049,51.1914878048781,59.1323658536585,74.2469268292683,69.4001707317073,65.4565609756098,67.5223658536585,72.6403414634147,70.3052926829268,73.6463414634147,75.1759512195122,64.2918292682927,57.7676829268293,71.159512195122,76.8361951219512,78.8414634146341,68.2275853658537,72.8108780487805,74.0744146341464,79.6243902439024,75.756487804878,71.669243902439,73.2503902439024,63.583512195122,56.7365853658537,58.2719268292683,59.2373658536585,55.633)
```

Method: Set Working Directory and Read Data

```
getwd()
#windows
setwd("C:\\Users\\jc_rc\\Data Mining Class\\DataMining")
```

Import the CSV resource, we assign the csv to an object

```
DataCountriescsv <- read.csv("DataFramesEvaluation_Data.csv", )
```

Creating Data Frames from ours vector

```
DfCoutriesData <- data.frame(Country= Countries_2012_Dataset,
                             Code= Codes_2012_Dataset,
                             Region= Regions_2012_Dataset)

head(DfCoutriesData)
```

Data frame to Life expectancy at 1960

```
DfLifeExpectancy1960 <- data.frame(Code = Country_Code,
                                   Life_Expectancy1960 =
Life_Expectancy_At_Birth_1960)

head(DfLifeExpectancy1960)
```

Data frame to Life expectancy at 2013

```
DfLifeExpectancy2013 <- data.frame(Code = Country_Code,
                                   Life_Expectancy2013 =
Life_Expectancy_At_Birth_2013)

head(DfLifeExpectancy2013)
```

Filter the data frame of the cvs document by year

```
head(DataCountriescsv)
DataCountriescsv$Year %in% 1960
filterYear1960 <- DataCountriescsv$Year %in% 1960
filterYear1960
DataCountriescsv[filterYear1960,]

DataCountriescsv$Year %in% 2013
filterYear2013 <- DataCountriescsv$Year %in% 2013
filterYear2013
DataCountriescsv[filterYear2013,]
```

Merge two data frames by common columns

Dataframes to vectors 1960

```
MCountriesData1960 <- merge(DfCoutriesData,DfLifeExpectancy1960, by.x =
"Code", by.y = "Code")
head(MCountriesData1960)
```

Dataframes to vectors 2013

```
MCountriesData2013 <- merge(DfCoutriesData,DfLifeExpectancy2013, by.x =
"Code", by.y = "Code")
head(MCountriesData2013)
```

Merge of cvs with dataframes per year 1960

```
MergeContries1960F <-
merge(DataCountriescsv[filterYear1960,],MCountriesData1960, by.x =
"Country.Code", by.y = "Code" )
head(MergeContries1960F)
tail(MergeContries1960F)
```

Merge of cvs with dataframes per year 1960

```
MergeContries2013F <-
merge(DataCountriescsv[filterYear2013,],MCountriesData2013, by.x =
"Country.Code", by.y = "Code" )
head(MergeContries2013F)
tail(MergeContries2013F)
```

Instal library to be able to make complex graphics

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

We use the qplot function to create the scatter plots

Creating the scatter plot for 1960

```
qplot(data=MergeContries1960F,
      x= Fertility.Rate,
      y=Life_Expectancy1960,
      size= Region.x ,
      color = Country.Name ,
      main="Fertility rate vs Life expectancy at 1960",
      xlab= "Fertility Rate",
      ylab="Life expectancy",

)
```

Creating the scatter plot for 2013

```
qplot(data=MergeContries2013F,
      x= Fertility.Rate,
      y=Life_Expectancy2013,
      size= Region.x ,
      color = Country.Name ,
      main="Fertility rate vs Life expectancy at 2013",
      xlab= "Fertility Rate",
      ylab="Life expectancy",
)
```

Conclusion

All the topics seen in class were implemented, in addition to having investigated a little about the scart plots for the representation of graphs and how they worked, being able to check the data of the dataframes or dataset and make combinations of columns or merged, in order to visualize better the fertility rate and life expectancy than it was in the years 1960 and 2013, with this we conclude unit 1.

Repository

https://github.com/JuanCarlos-Negrete/Data-Mining/tree/Unit_1/Unit_1/Evaluation

Video

https://youtu.be/1mtsslE5vMg