Instituto Tecnológico de Tijuana
Ingeniería en Informática

Subject Name:

Data Mining

Exercise:

Evaluative Practice - Unit 1

Teacher:

Romero Hernandez Jose Christian

Students:

Castillo Ramirez Guadalupe 17213043

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 of 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 provided 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 file, some in R vectors. The CVS file contains combined data from 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 the explanation of his observations)

Note:

The files to carry out this practice, which are CountryRegionVectors.R and DataFrameEvaluation_Data.csv, are located in the shared Google Drive folder Resources / Information-UI:

Evaluation instructions

- Delivery time September 28, 2021

- At the end put the code and the documentation with its explanation in the corresponding branch of your github, likewise make your explanation of the solution in your google drive in google document (Cover, Introduction, Development, etc). - Finally defend its development in a 6-8 min video explaining its solution and comments, this will serve to give your rating of this evaluation practice, this video must be uploaded to YouTube to be shared by a public link (Use google meet with the cameras lit and record your defense to build the video).

```
# import the CSV resource, we assign the csv to an object
Datos <- read.csv(file.choose())

# countries Vector

Countries_2012_Dataset <- c("Aruba","Afghanistan","Angola","Albania","United Arab
Emirates","Argentina","Armenia","Antigua and
Barbuda","Australia","Austria","Azerbaijan","Burundi","Belgium","Benin","Burkina
Faso","Bangladesh","Bulgaria","Bahrain","Bahamas, The","Bosnia and
Herzegovina","Belarus","Belize","Bermuda","Bolivia","Brazil","Barbados","Brunei
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","Micronesia,
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","Kazakhstan","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","Mauritania","Mauri
tius","Malawi","Malaysia","Namibia","New
Caledonia","Niger","Nigeria","Nicaragua","Netherlands","Norway","Nepal","New
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-Leste","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","BRA","BRB","BRN","BTN","B
WA","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","GBR","GEO","GHA","GIN","GMB","GNB","GNQ","GRC","GRD","GRL","G
TM","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","LBN","LBR",
"LBY","LCA","LIE","LKA","LSO","LTU","LUX","LVA","MAC","MAR","MDA","MDG","MDV","MEX","M
KD","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","POL","PRI","PRT",
"PRY","PYF","QAT","ROU","RUS","RWA","SAU","SDN","SEN","SGP","SLB","SLE","SLV","SOM","S
RB","SSD","STP","SUR","SVK","SVN","SWE","SWZ","SYC","SYR","TCD","TGO","THA","TJK","TKM
","TLS","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","Asia","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","Oc
eania","Africa","Europe","Asia","Africa","Africa","Africa","Africa","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","Africa","Europe"
,"Africa","Asia","The
Americas","Europe","Africa","Europe","Asia","Europe","Asia","Africa","Africa","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","BRB","BRN","BTN","BWA","C
AF","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","GHA","GIN","GMB","GNB","GNQ","GRC","GRD","GTM","GUM","GUY","H
KG","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","LKA","LSO",
"LTU","LUX","LVA","MAC","MAR","MDA","MDG","MDV","MEX","MKD","MLI","MLT","MMR","MNE","M
NG","MOZ","MRT","MUS","MWI","MYS","NAM","NCL","NER","NGA","NIC","NLD","NOR","NPL","NZL
","OMN","PAK","PAN","PER","PHL","PNG","POL","PRI","PRT","PRY","PYF","QAT","ROU","RUS",
"RWA","SAU","SDN","SEN","SGP","SLB","SLE","SLV","SOM","SSD","STP","SUR","SVK","SVN","S
WE","SWZ","SYR","TCD","TGO","THA","TJK","TKM","TLS","TON","TTO","TUN","TUR","TZA","UGA
","UKR","URY","USA","UZB","VCT","VEN","VIR","VNM","VUT","WSM","YEM","ZAF","COD","ZMB",
"ZWE")

Life_Expectancy_At_Birth_1960 <-
c(65.5693658536586,32.328512195122,32.9848292682927,62.2543658536585,52.2432195121951,
65.2155365853659,65.8634634146342,61.7827317073171,70.8170731707317,68.5856097560976,6
0.836243902439,41.2360487804878,69.7019512195122,37.2782682926829,34.4779024390244,45.
8293170731707,69.2475609756098,52.0893658536585,62.7290487804878,60.2762195121951,67.7
080975609756,59.9613658536585,42.1183170731707,54.2054634146342,60.7380487804878,62.50
03658536585,32.3593658536585,50.5477317073171,36.4826341463415,71.1331707317073,71.313
4146341463,57.4582926829268,43.4658048780488,36.8724146341463,41.523756097561,48.58163
41463415,56.716756097561,41.4424390243903,48.8564146341463,60.5761951219512,63.9046585
365854,69.5939268292683,70.3487804878049,69.3129512195122,44.0212682926829,72.17658536
58537,51.8452682926829,46.1351219512195,53.215,48.0137073170732,37.3629024390244,69.10
92682926829,67.9059756097561,38.4057073170732,68.819756097561,55.9584878048781,69.8682
926829268,57.5865853658537,39.5701219512195,71.1268292682927,63.4318536585366,45.83146
34146342,34.8863902439024,32.0422195121951,37.8404390243902,36.7330487804878,68.163902
4390244,59.8159268292683,45.5316341463415,61.2263414634146,60.2787317073171,66.9997073
170732,46.2883170731707,64.6086585365854,42.1000975609756,68.0031707317073,48.64031707
31707,41.1719512195122,69.691756097561,44.945512195122,48.0306829268293,73.42865853658
54,69.1239024390244,64.1918292682927,52.6852682926829,67.6660975609756,58.367585365853
7,46.3624146341463,56.1280731707317,41.2320243902439,49.2159756097561,53.0013170731707
,60.3479512195122,43.2044634146342,63.2801219512195,34.7831707317073,42.6411951219512,
57.303756097561,59.7471463414634,46.5107073170732,69.8473170731707,68.4463902439024,69
.7868292682927,64.6609268292683,48.4466341463415,61.8127804878049,39.9746829268293,37.
2686341463415,57.0656341463415,60.6228048780488,28.2116097560976,67.6017804878049,42.7
363902439024,63.7056097560976,48.3688048780488,35.0037073170732,43.4830975609756,58.74
```

```
52195121951,37.7736341463415,59.4753414634146,46.8803902439024,58.6390243902439,35.515
0487804878,37.1829512195122,46.9988292682927,73.3926829268293,73.549756097561,35.17082
92682927,71.2365853658537,42.6670731707317,45.2904634146342,60.8817073170732,47.691585
3658537,57.8119268292683,38.462243902439,67.6804878048781,68.7196097560976,62.80892682
92683,63.7937073170732,56.3570487804878,61.2060731707317,65.6424390243903,66.055292682
9268,42.2492926829268,45.6662682926829,48.1876341463415,38.206,65.6598292682927,49.381
7073170732,30.3315365853659,49.9479268292683,36.9658780487805,31.6767073170732,50.4513
658536585,59.6801219512195,69.9759268292683,68.9780487804878,73.0056097560976,44.23378
04878049,52.768243902439,38.0161219512195,40.2728292682927,54.6993170731707,56.1535365
853659,54.4586829268293,33.7271219512195,61.3645365853659,62.6575853658537,42.00975609
7561,45.3844146341463,43.6538780487805,43.9835609756098,68.2995365853659,67.8963902439
025,69.7707317073171,58.8855365853659,57.7238780487805,59.2851219512195,63.73021951219
51,59.0670243902439,46.4874878048781,49.969512195122,34.3638048780488,49.0362926829268
,41.0180487804878,45.1098048780488,51.5424634146342)

Life_Expectancy_At_Birth_2013 <-
c(75.3286585365854,60.0282682926829,51.8661707317073,77.537243902439,77.1956341463415,
75.9860975609756,74.5613658536585,75.7786585365854,82.1975609756098,80.890243902439,70
.6931463414634,56.2516097560976,80.3853658536585,59.3120243902439,58.2406341463415,71.
245243902439,74.4658536585366,76.5459512195122,75.0735365853659,76.2769268292683,72.47
07317073171,69.9820487804878,67.9134390243903,74.1224390243903,75.3339512195122,78.546
6585365854,69.1029268292683,64.3608048780488,49.8798780487805,81.4011219512195,82.7487
804878049,81.1979268292683,75.3530243902439,51.2084634146342,55.0418048780488,61.66639
02439024,73.8097317073171,62.9321707317073,72.9723658536585,79.2252195121951,79.256390
2439025,79.9497804878049,78.2780487804878,81.0439024390244,61.6864634146342,80.3024390
243903,73.3199024390244,74.5689512195122,75.648512195122,70.9257804878049,63.177878048
7805,82.4268292682927,76.4243902439025,63.4421951219512,80.8317073170732,69.9179268292
683,81.9682926829268,68.9733902439024,63.8435853658537,80.9560975609756,74.07951219512
2,61.1420731707317,58.216487804878,59.9992682926829,54.8384146341464,57.2908292682927,
80.6341463414634,73.1935609756098,71.4863902439024,78.872512195122,66.3100243902439,83
.8317073170732,72.9428536585366,77.1268292682927,62.4011463414634,75.2682926829268,68.
7046097560976,67.6604146341463,81.0439024390244,75.1259756097561,69.4716829268293,83.1
170731707317,82.290243902439,73.4689268292683,73.9014146341463,83.3319512195122,70.45,
60.9537804878049,70.2024390243902,67.7720487804878,65.7665853658537,81.459756097561,74
.462756097561,65.687243902439,80.1288780487805,60.5203902439024,71.6576829268293,74.91
27073170732,74.2402926829268,49.3314634146342,74.1634146341464,81.7975609756098,73.980
4878048781,80.3391463414634,73.7090487804878,68.811512195122,64.6739024390244,76.60260
97560976,76.5326585365854,75.1870487804878,57.5351951219512,80.7463414634146,65.654097
5609756,74.7583658536585,69.0618048780488,54.641512195122,62.8027073170732,74.46,61.46
6,74.567512195122,64.3438780487805,77.1219512195122,60.8281463414634,52.4421463414634,
74.514756097561,81.1048780487805,81.4512195121951,69.222,81.4073170731707,76.841048780
4878,65.9636829268293,77.4192195121951,74.2838536585366,68.1315609756097,62.4491707317
073,76.8487804878049,78.7111951219512,80.3731707317073,72.7991707317073,76.33407317073
17,78.4184878048781,74.4634146341463,71.0731707317073,63.3948292682927,74.177634146341
5,63.1670487804878,65.878756097561,82.3463414634146,67.7189268292683,50.3631219512195,
72.4981463414634,55.0230243902439,55.2209024390244,66.259512195122,70.99,76.2609756097
```

```
561,80.2780487804878,81.7048780487805,48.9379268292683,74.7157804878049,51.19148780487
81,59.1323658536585,74.2469268292683,69.4001707317073,65.4565609756098,67.522365853658
5,72.6403414634147,70.3052926829268,73.6463414634147,75.1759512195122,64.2918292682927
,57.7676829268293,71.159512195122,76.8361951219512,78.8414634146341,68.2275853658537,7
2.8108780487805,74.0744146341464,79.6243902439024,75.756487804878,71.669243902439,73.2
503902439024,63.583512195122,56.7365853658537,58.2719268292683,59.2373658536585,55.633
)









# Method: Set Working Directory and Read Data

getwd()

#windows

setwd("C:/Users/Cylon/Documents")







# 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(Datos)

Datos$Year %in% 1960

filterYear1960 <- Datos$Year %in% 1960

filterYear1960

Datos[filterYear1960,]
```

```
Datos$Year %in% 2013

filterYear2013 <- Datos$Year %in% 2013

filterYear2013

Datos[filterYear2013,]



# Merge two data frames by common columns



# dataframes to vectors 1960

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

head(MCountriesData1960)

# Data frame de vector 2013

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

head(MCountriesData2013)



# Merge of cvs with dataframes per year 1960

MergeContries1960F <- merge(Datos[filterYear1960,],MCountriesData1960, by.x =
"Country.Code", by.y = "Code" )

head(MergeContries1960F)

tail(MergeContries1960F)
```

```
# Merge of cvs with dataframes per year 2013

MergeContries2013F <- merge(Datos[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",

    )
```

Conclusión

Se implementaron todos los temas vistos en clase, además de haber investigado un poco sobre

los diagramas de euroconector para la representación de gráficos y su funcionamiento, pudiendo comprobar el

datos de los marcos de datos o conjunto de datos y hacer combinaciones de columnas o fusionados, con el fin de

visualizar mejor la tasa de fecundidad y la esperanza de vida que en los años 1960 y 2013,

Enlace GitHub: https://github.com/CastilloRmz/Data-Mining.git