

Tecnológico Nacional De México

Instituto Tecnológico De Tijuana

Subdirección Académica

Departamento de Sistemas y Computación

Semestre Enero - Junio 2022

Ingeniería Informática

Mineria De Datos

Práctica Evaluatoria

Unidad 1

Perez Ortega Victoria Valeria No.18210718

Díaz Ruiz Uriel No.18210839

JOSE CHRISTIAN ROMERO HERNANDEZ

Tijuana, B.C. a 22 de Marzo de 2022.





TECNOLÓGICO NACIONAL DE MÉXICO

INSTITUTO TECNOLÓGICO DE TIJUANA SUBDIRECCIÓN ACADÉMICA

Departamento de Sistemas y Computación **EXAMEN**

Carrera: Ingeniería En Sistemas Computacionales/ Tecnologías de la información/ Informática Período: Febrero-Junio 2022 Materia: Minería de datos Grupo: BDD-1703SC9C Salón: Unidad (es) a evaluar: Unidad 1 Tipo de examen:Practico Fecha: Catedrático: José Christian Romero Hernandez Firma del maestro: Calificación:

Alumno: <u>Diaz Ruiz Uriel y Perez Ortega Victoria Valeria</u> No. Control: <u>18210839</u>, <u>18210718</u>

Instrucciones Instrucciones

Desarrolle el siguiente problema con R y RStudio para la extracción de conocimiento que el problema requiere.

El Banco Mundial quedó muy impresionado con su entrega en la asignación anterior y tienen un nuevo proyecto para usted.

Debe generar un diagrama de dispersión (scatter-plot) que muestre las estadísticas de esperanza de vida (Life expectancy - eje y) y tasa de fertilidad (Fertility Rate -eje x) por país (Country).

El diagrama de dispersión también debe clasificarse por países Regiones (Country Regions).

Se le han proporcionado datos durante 2 años: 1960 y 2013 y se le exige que produzca una visualización para cada uno de estos años.

Algunos datos se han proporcionado en un archivo CVS, algunos en vectores R. El archivo CVS contiene datos combinados de ambos años. Toda manipulación de datos debe realizarse en R (No en Excel) porque este proyecto puede ser auditado en una etapa posterior.

También se le ha pedido que proporcione información sobre cómo se comparan los dos períodos. (Hint: Básicamente la explicación de sus observaciones)

Nota:

Los archivos para realizar esta práctica que son **Evaluation_Vectors.R y**





DataFrameEvaluation_Data.csv se encuentra en la carpeta de Google Drive compartida Recursos/Informacion-U1:

Instrucciones de evaluación

- Tiempo de entrega 22 de marzo 2022
- Al terminar poner el código y la documentación con su explicación en el branch correspondiente de su github, así mismo realizar su explicación de la solución en su google drive en documento de google (Portada, Introducción, Desarrollo, etc). Finalmente defender su desarrollo en un video de 6-8 min explicando su solución y observaciones, este servirá para dar su calificación de esta práctica evaluatoria, este video debe subirse a youtube para ser compartido por un link público (Utilicen google meet o MS Teams con las cámaras encendidas y graben su defensa para elaborar el video). Happy Coding:)!

Hay que importar primero nuestra biblioteca mediante el comando **ggplot2**, al correr esa línea de código se instalará automáticamente para así poder utilizar el comando de gráficos de RStudio.

que sería así:









A	U		U	L
3 Sweden	SWE	Europe	2013	1.91
7 Swaziland	SWZ	Africa	2013	3.334
3 Syrian Arab Republic	SYR	Middle East	2013	2.964
9 Chad	TCD	Africa	2013	6.264
Togo	TGO	Africa	2013	4.655
1 Thailand	THA	Asia	2013	1.524
2 Tajikistan	TJK	Asia	2013	3.514
3 Turkmenistan	TKM	Asia	2013	2.326
1 Timor-Leste	TLS	Asia	2013	5.2
5 Tonga	TON	Oceania	2013	3.768
Trinidad and Tobago	TTO	The Americas	2013	1.789
7 Tunisia	TUN	Africa	2013	2.25
3 Turkey	TUR	Europe	2013	2.041
7 Tanzania	TZA	Africa	2013	5.215
Uganda	UGA	Africa	2013	5.869
1 Ukraine	UKR	Europe	2013	1.506
2 Uruguay	URY	The Americas	2013	2.033
3 United States	USA	The Americas	2013	1.8695
1 Uzbekistan	UZB	Asia	2013	2.2
St. Vincent and the Grenadines	VCT	The Americas	2013	1.997
Venezuela, RB	VEN	The Americas	2013	2.391
7 Virgin Islands (U.S.)	VIR	The Americas	2013	1.77
3 Vietnam	VNM	Asia	2013	1.743
9 Vanuatu	VUT	Oceania	2013	3.383
) Samoa	WSM	Oceania	2013	4.147
1 Yemen, Rep.	YEM	Middle East	2013	4.284
2 South Africa	ZAF	Africa	2013	2.387





```
#Execute below code to generate three new vectors

Country_Code <- c("ABW", "AFG", "AGO", "ALB", "ARE", "ARG", "ARM", "ATG", "AUS", "AUT", "AZE", "BDI", "BEL", "BEN", "BFN", "BNN", "BNN", "BNN", "BN
```

Primer vector:





Country Name
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
Bolivia
Brazil
Barbados
Brunei Darussalam
Dhutan





CountriesDataset <- 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, <u>The</u> Bosnia and Herzegovina Belarus Belize Bolivia Brazil Barbados Brunei Darussalam Bhutan Botswana Central African Republic Canada

Switzerland Chile (Top Level) ±





```
CountriesDataset <- c(
|"Aruba", "Afghanistan", "Angola",
"Albania",
"United Arab Emirates",
"Argentina",
"Armenia",
"Antigua and Barbuda",
"Australia",
 "Austria",
"Azerbaijan",
"Burundi",
"Burundi",
"Belgium",
 "Benīn",
 "Burkina Faso",
"Bangladesh",
"Bulgaria",
"Bahrain",
"Bahamas",
"The Bosnia and Herzegovina",
"Belarus",
"Belize",
"Bolivia",
"Brazil",
"Barbados",
"Brunei Darussalam",
"Bhutan",
"Botswana",
"Central African Republic",
"Canada",
"Switzerland",
"Chile",
```





CountriesNames <- 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", "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", "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","Guatemala","Guam",
"Guyana","Hong Kong SAR, China","Honduras","Croatia","Haiti","Hungary","Indonesia","India","Ireland","Iran, Islamic Rep.","Iraq", "Iceland","Italy","Jamaica","Jordan","Japan","Kazakhstan","Kenya","Kyrgyz Republic","Cambodia","Kiribati","Korea, Rep.","Kuwait", "Lao PDR","Lebanon","Liberia","Libya","St. Lucia","Sri Lanka","Lesotho","Lithuania","Luxembourg","Latvia","Macao SAR, China", "Morocco","Moldova","Madagascar","Maldives","Mexico","Macedonia, FYR","Mali","Malta","Myanmar","Montenegro","Mongolia","Mozambique", "Mauritania", "Mauritius", "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", "South Sudan", "Sao Tome and Principe", "Suriname", "Slovak Republic", "Slovenia", "Sweden", "Swaziland", "Syrian Arab Republic", "Chad", "Togo", "Thailand", "Tajikistan", "Turkmenistan", "Timor-Leste", "Tonga", "Trinidad and Tobago", "Tunisia", "Turkey", "Tanzania", "Uganda", "Ukraine", "Uruquay", "United States", "Uzbekistan", "St. Vincent and the Grenadines", "Venezuela", "RB Virgin Islands (U.S.)", "Vietnam", "Vanuatu", "Samoa", "Yemen", "Rep. South Africa", "Congo", "Dem. Rep. Zambia", "Zimbabwe")

Segundo Vector:

Hacemos lo mismo que el anterior solo que ahora vamos a insertar los códigos de los países.

```
CodesCountry <- 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", "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", "GBR", "GEO", "GHA", "GIN", "GMB", "GNB", "GNB", "GNC", "GRC", "GRD", "GRU", "GU", "GW", "GU", "HKG", "HND", "HTI", "HUN", "IDN", "IND", "IRL", "IRN", "IRQ", "ISL", "ISR", "ITA", "JAM", "JOR", "JPN", "KAZ", "KEN", "KGZ", "KHM", "KIR", "KOR", "KWT", "LAO", "LBR", "LBR", "LEE", "LKA", "LSO", "LTU", "LUX", "LVA", "MAC", "MAR", "MDA", "MDG", "MDV", "MEX", "MKD", "MLI", "MLI", "MNF", "MNF", "MNE", "MNG", "MOZ", "MRT", "MUS", "MWI", "MYS", "NAM", "NCL", "NER", "NGA", "NIC", "NLD", "NOR", "NPL", "NZL", "OMN", "PAK", "PAN", "PER", "PHL", "PNG", "POL", "PRI", "PRY", "PYF", "QAT", "ROU", "RUS", "RWA", "SAU", "SDN", "SEN", "SLE", "SLE", "SLV", "SOM", "SRB", "SSD", "STP", "SUR", "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")
```

Tercer Vector:





Aquí agregamos para nuestro 3cer. vector las regiones de cada país.

RegionsDataset <- c("The Americas", "Asia", "Africa", "Europe", "Middle East", "The Americas", "Asia", "Europe", "The Americas", "Ceania",

"Europe", "Asia", "Africa", "Europe", "Africa", "Africa", "Asia", "Europe", "Middle East", "The Americas", "Europe", "Europe", "The Americas", "The Americas", "The Americas", "The Americas", "The Americas", "The Americas", "Africa", "The Americas", "The Americas", "The Americas", "Europe", "Ceania", "Europe", "Coeania", "Europe", "Africa", "Africa", "Africa", "Africa", "Africa", "Africa", "Europe", "The Americas", "The Americas", "Oceania", "The Americas", "Africa", "Europe", "Asia", "Africa", "Africa

Vector para poder conseguir la Esperanza de vida:

Country_Code <- c("ABW", "AFG", "AGO", "ALB", "ARE", "ARG", "ARM", "ATG", "AUS", "AUT", "AZE", "BDI", "BEL", "BEN", "BF





Vector para poder conseguir la Esperanza de vida del año de 1960:

Life_Expectancy_At_Birth_1960 <- c(65.5693658536586,32.328512195122,32.9848292682927,62.2543658536585,52.

```
Life_Expectancy_At_Birth_1960 <- c(65.5693658536586,32.328512195122,32.9848292682927,62.2543658536585,52.2432195121951,65.2155365853659,65.8634634146342,61.782731707317,70.81707317,708.58586097560976,60.836243902439,41.2360487804878,69.7019512195122,37.2782682926829,34.4779024390244,45.829317073177,69.2475609756098,52.0893658356585,62.7920487804878,60.2762195121951,67.7080975609756,59.9613658536585,62.7920487804878,60.2762195121951,67.7080975609756,59.9613658536585,42.1183170731707,54.2054634146342,60.7380487804878,62.5003658536585,32.3593658536585,50.5477317073171,36.4826341463415,76.716756097561,41.4424390243903,48.856414634165,60.761951219512,63.904658536585,42.1183170731773,71.3134146341463,57.4582926829,68.43.4658048780488,36.8724146344163,41.523756097561,48.5816341463415,56.716756097561,41.4424390243902,43,8.8564146341465,60.761951219512,63.904658536584,49.5939268292683,70.3487804878049,69.3129512195122,44.0212682926829,72.1765835365837,51.8452682926829,46.1351219512195,53.215,48.0137073170732,78.819756097561,53.215,48.0137073170732,78.3269024390244,59.3192682926829,67.9059756097561,38.4057073170732,68.819756097561,55.9584878048781,69.8682926829268,57.586585365837,39.5701219512195,71.1268292682927,63.4318536585366,45.8314634146342,48.865902439024,32.0422195121951,73.804390243902,36.7330487804878,61.693902439024,45.98159268292683,45.5316341463415,61.22634146341463,60.2787317073171,66.9997073170732,46.2883170731707,64.6086585365854,42.1000975609756,68.0031707317073,48.6403170731707,41.1719512195122,68.50682926829,67.6660975609756,58.3675853658537,34.866585365854,69.123902439024,49.2159756097561,53.0013170731707,68.440390244,69.1818292682927,52.68526829682,67.6660975609756,58.3675853658537,39.270195129512,34.8830975609756,68.7503756097561,53.0013170731707,68.4463902439024,69.78682926829,67.303756097561,53.0013170731707,68.4463902439024,98.836902439024,98.836902439024,98.368902439024,98.36902439024,98.36902439024,98.36902439024,98.36902439024,98.36902439024,99.355853658537,39.36902439024,99.28.69902439024,99.
```





Vector para poder conseguir la Esperanza de vida del año de 2013:

Life_Expectancy_At_Birth_2013 <- c(75.3286585365854,60.0282682926829,51.8661707317073,77.537243902439,77.

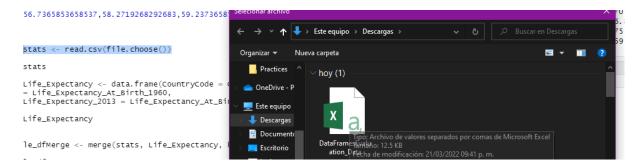
```
Life_Expectancy_At_Birth_2013 <- c(75.3286585365854,60.0282682926829,51.8661707317073,77.537243902439,77.1956341463415,75.9860975609756,74.5613658536585,75.778658536584,82.1975609756098,80.890243902439,70.693146341463415,75.986097560976,80.3853658536585,75.778658536584,82.1975609756098,80.890243902439,74.4658536585366,76.5459512195122,75.07353658536585,79.3120243902439,58.2406341463415,71.245243902439,74.465853658536586,76.5459512195122,75.0735365853659,76.2769268292683,72.4707317073171,69.9820487804878,69.19380243903,75.3339512195122,78.5466585365854,69.1029268292683,64.368046780488,69.8798780487805,81.4011219512195,82.7487804878049,81.1979268292683,75.3530243902439,51.2084654146342,55.0418048780488,61.6663902439024,73.809731707317,162.9321707317073,72.9723658536585,79.2252195121951,79.2563902439025,79.3497804878049,78.2780487804878,81.043902439024,61.68664634146342,80.3024390243903,73.3199024390224,74.5689512195122,75.648512195122,70.9257804878049,63.1778780487805,82.426829268292,76.4243902439024,74.568951219512,82.75.648512195122,70.9257804878049,63.177878780487805,82.426829268292,76.4243902439024,78.89926829268,89.333902439024,78.89926829268,89.333902439024,78.89926829268,89.333902439024,78.69926829268,89.333902439024,78.69926829268,89.333902439024,78.68926829268,89.333902439024,78.68926829268,89.3446341464,57.290829682927,80.634146341463,73.193609756098,71.4863902439024,78.689268292683,73.3901416634163,83.3319512195122,62.6207043002439024,78.689268292683,78.3901416634163,83.3319512195122,62.6207045,60.9537804878.049,70.2024390244,79.65786926833,73.9014166341633,83.319512195122,64.673902439024,78.6578604878,65.766585365854,75.1870487804878,65.766585366854,75.1870487804878,65.766585366854,75.1870487804878,65.766585366854,75.1870487804878,65.766585366854,75.1870487804878,65.766585366854,75.1870487804878,65.766585366854,75.1870487804878,75.351951219512,60.8733170731707,76.8410487804878,65.963682926893,77.41921951219512,66.8573902439024,78.657890249024,78.65780487804878,65.96368292683,78.441054166341664,71.0731
```

Para conseguir la tasa de fertilidad es necesario abrir con el siguiente comando:

```
stats <- read.csv(file.choose())
```







Nota: Mí Práctica Evaluativa - Unidad 1 la guarde en downloads.

	Country.Name	Country.Code	Region	Year [‡]	Fertility.Rate
1	Aruba	ABW	The Americas	1960	4.820
2	Afghanistan	AFG	Asia	1960	7.450
3	Angola	AGO	Africa	1960	7.379
4	Albania	ALB	Europe	1960	6.186
5	United Arab Emirates	ARE	Middle East	1960	6.928
6	Argentina	ARG	The Americas	1960	3.109
7	Armenia	ARM	Asia	1960	4.550
8	Antigua and Barbuda	ATG	The Americas	1960	4.425
9	Australia	AUS	Oceania	1960	3.453
10	Austria	AUT	Europe	1960	2,690
11	Azerbaijan	AZE	Asia	1960	5.571
12	Burundi	BDI	Africa	1960	6.953
13	Belgium	BEL	Europe	1960	2.540
14	Benin	BEN	Africa	1960	6.282
15	Burkina Faso	BFA	Africa	1960	6.291
16	Bangladesh	BGD	Asia	1960	6.725
17	Bulgaria	BGR	Europe	1960	2.310
18	Bahrain	BHR	Middle East	1960	7.090
19	Bahamas, The	BHS	The Americas	1960	4.495
20	Bosnia and Herzegovina	BIH	Europe	1960	3.770
21	Belarus	BLR	Europe	1960	2.670
22	Belize	BLZ	The Americas	1960	6.500
23	Bolivia	BOL	The Americas	1960	6.700
24	Brazil	BRA	The Americas	1960	6.210
25	Barbados	BRB	The Americas	1960	4.333
26	Brunei Darussalam	BRN	Asia	1960	6.487

Puntos requeridos para generar el diagrama:





Debe generar un diagrama de dispersión (scatter-plot) que muestre las estadísticas de esperanza de vida (Life expectancy - eje y) y tasa de fertilidad (Fertility Rate -eje x) por país (Country).

El diagrama de dispersión también debe clasificarse por países Regiones (Country Regions).

Se le han proporcionado datos durante 2 años: 1960 y 2013 y se le exige que produzca una visualización para cada uno de estos años.

Guardado de datos sobre la esperanza de vida:

```
Life_Expectancy <- data.frame(CountryCode = Country_Code, Life_Expectancy_1960 = Life_Expectancy_At_Birth_1960, Life_Expectancy_2013 = Life_Expectancy_At_Birth_2013)

Life_Expectancy
```

Generación del comando Merge para poder implementar la información compilada.

```
le_dfMerge <- merge(stats, Life_Expectancy, by.x = "Country.Code", by.y ="CountryCode")
le_dfMerge</pre>
```

Generación para poder implementar la información compilada sobre la esperanza de vida y fertilidad de los años 1960 y 2013 por separado para generar nuestra gráfica relacional.

```
qplot(data = le_dfMerge, y = Life_Expectancy_1960, x = Fertility.Rate)
qplot(data = le_dfMerge, y = Life_Expectancy_2013, x = Fertility.Rate)
```

Generación para poder generar los diagramas mediante la información compilada sobre la esperanza de vida y fertilidad de los años 1960 y 2013 por separado para poder generar nuestra gráfica relacional pero ahora por regiones.

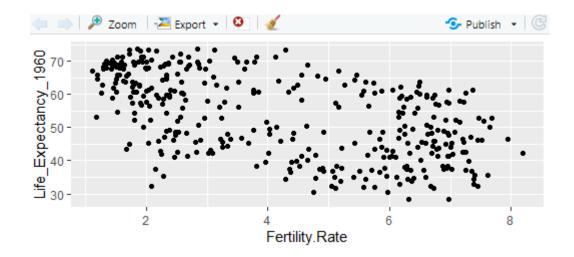


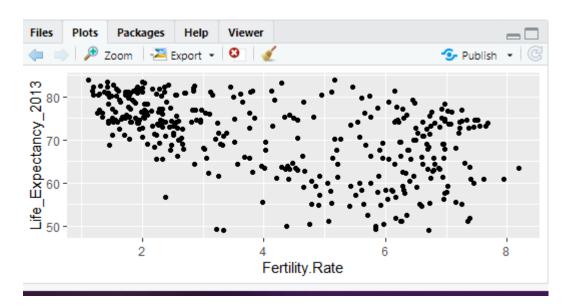


 ${\tt qplot(data = le_dfMerge, \ y = Life_Expectancy_1960, \ x = Fertility.Rate)}$

qplot(data = le_dfMerge, y = Life_Expectancy_2013, x = Fertility.Rate)

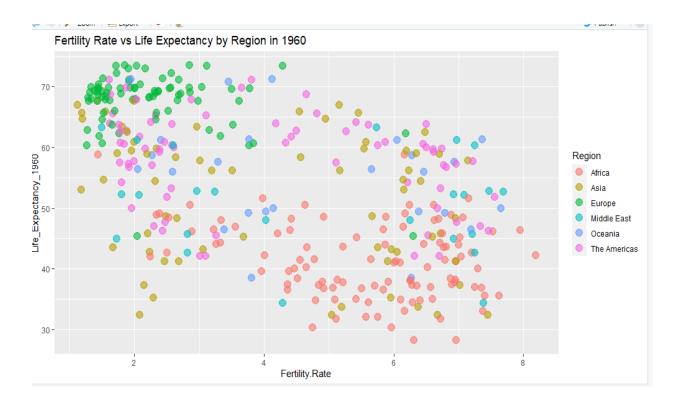
qplot(data = le_dfMerge, x = Fertility.Rate, y = Life_Expectancy_2013,
color = Region, size=I(6), shape=I(20), alpha =I(.6), main =
 "Fertility Rate vs Life Expectancy by Region in 2013")

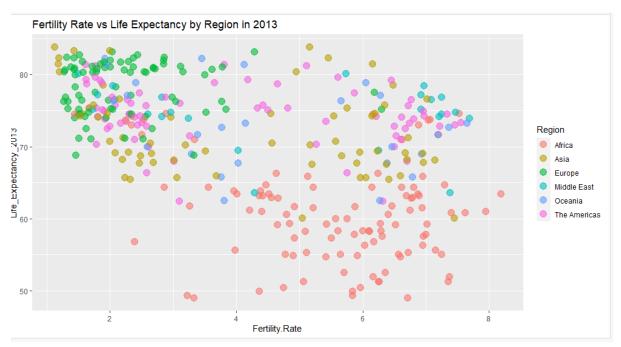












URL: https://youtu.be/36ztAKAMMPE