



# Tecnológico Nacional de México

## Instituto Tecnológico de Tijuana

---

Subdirección Académica  
Departamento de Sistemas y Computación  
Ingeniería en Sistemas Computacionales  
Semestre: AGOSTO-DICIEMBRE 2021

### MINERÍA DE DATOS

*BDD-1703SC9A*

#### *“Evaluación 1”*

Jiménez Ramírez Julio Fabián 17212147

Flores González Luis Diego C16211486

**MC. JOSE CHRISTIAN ROMERO HERNANDEZ**

Campus Tomas Aquino

---

*“Por una juventud integrada al desarrollo de México”*

Tijuana B.C. a 28 de Septiembre del 2021

## Assessment 1

The World Bank was very impressed with your delivery of the previous assignment and they have a new project for you.

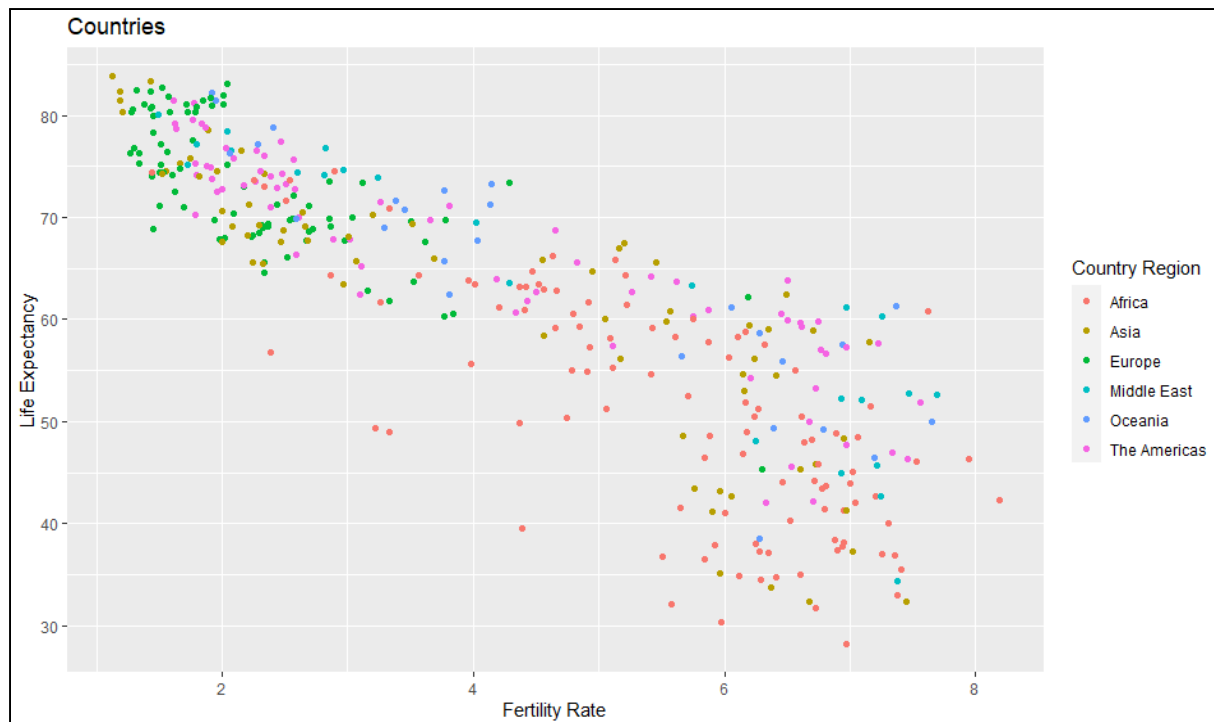
You should generate a scatterplot showing the statistics on life expectancy (life expectancy - y-axis) and fertility rate (fertility rate-x-axis) by country.

You have been provided data for 2 years: 1960 and 2013 and you are required to produce a visualization for each of these years.

```
# We extreme the file browser data and create a new vector with the 1960 and
2013 life expectancy data. Finally a new data frame is created with a new
column for the life expectancy.
df <- read.csv(file.choose())
y <- c(Life_Expectancy_At_Birth_1960, Life_Expectancy_At_Birth_2013)
df["Life.expectancy"] <- y

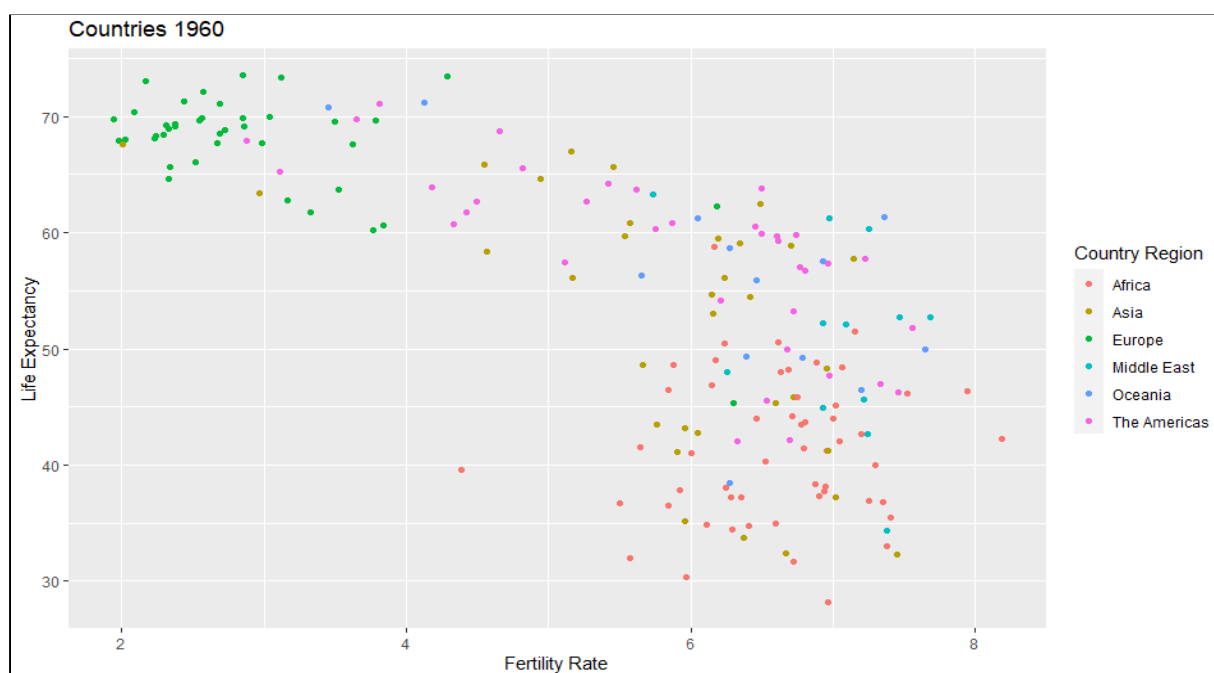
# We create new data frames to visualize the scatterplots for the dates of
1960 and 2013. First the records with the specific dates are filtered.
df_2 <- df$Year == "1960"
df_3 <- df$Year == "2013"
df_1960 = df[df_2,]
df_2013 = df[df_3,]

# The library is exported to create graphs
library(ggplot2)
# The values for "x" and "y" of the data frame created are determined to
visually see the dispersion of the total set with their respective titles.
ggplot(
  df,
  aes(x=df$Fertility.Rate, y=df$Life.expectancy, colour=df$Region)) +
  geom_point() +
  labs(title="Countries", x="Fertility Rate", y="Life Expectancy",
colour="Country Region")
```



# In the same way as the previous one, the data frame created from the year 1960 is used.

```
ggplot(
  df_1960,
  aes(x=df_1960$Fertility.Rate, y=df_1960$Life.expectancy,
  colour=df_1960$Region)) +
  geom_point() +
  labs(title="Countries 1960", x="Fertility Rate", y="Life Expectancy",
  colour="Country Region")
```



```
# In the same way as the previous one, the data frame created from the year 2013 is used.
```

```
ggplot(  
  df_2013,  
  aes(x=df_2013$Fertility.Rate, y=df_2013$Life.expectancy,  
  colour=df_2013$Region)) +  
  geom_point() +  
  labs(title="Countries 2013", x="Fertility Rate", y="Life Expectancy",  
  colour="Country Region")
```

