

This is an **individual** graded event. While receiving help is permitted with proper documentation, over-reliance on the assistance from others may result in grade deduction.

Answer the questions in the spaces provided. Submit all code using the “Compile Report Document” button found at the top of your script in RStudio, ensure you select “MS Word”. Ensure each section of code has a committed label with its corresponding problem, example “#Problem1”. If you run out of room for an answer, you may attach additional sheets of paper.

Background: Since the industrial revolution, countries have developed at different rates. Several factors affect each country’s development, to include access to resources and being on the winning side of conflicts. Researchers hired by the United Nations Economic and Social Council collected data on selected countries’ life expectancy (years), Gross Domestic Product (GDP) per capita (International \$ per person), and total population from different years. The researchers are curious if certain continents have longer life expectancies. They are also interested potential relationships between the variables collected.

1. **Step 1. Ask a research question.** What is one research question being investigated in this study?

Data Analysis Lab 1

Name: _____

Data: For this data analysis lab you will use the file **Country.Data.csv**. The file contains the following variables:

Column (Variable)	Description	Units
Country	Country	N/A
Continent	Continent	N/A
Life.Expectancy.1940	Life Expectancy in 1940	Years
Life.Expectancy.1960	Life Expectancy in 1960	Years
Life.Expectancy.1980	Life Expectancy in 1980	Years
Life.Expectancy.2000	Life Expectancy in 2000	Years
Life.Expectancy.2015	Life Expectancy in 2015	Years
GDP.per.capita.1940	GDP per capita in 1940	International \$ per person
GDP.per.capita.1960	GDP per capita in 1960	International \$ per person
GDP.per.capita.1980	GDP per capita in 1980	International \$ per person
GDP.per.capita.2000	GDP per capita in 2000	International \$ per person
GDP.per.capita.2015	GDP per capita in 2015	International \$ per person
Total.Pop.1940	Total Population in 1940	People
Total.Pop.1960	Total Population in 1960	People
Total.Pop.1980	Total Population in 1980	People
Total.Pop.2000	Total Population in 2000	People
Total.Pop.2015	Total Population in 2015	People

2. Step 2. Design a study and collect data.

(a) Classify each variable as categorical or quantitative.

Variable	Classification	Variable	Classification
Country		Continent	
Life.Expectancy.1940		GDP.per.capita.1940	
Life.Expectancy.1960		GDP.per.capita.1960	
Life.Expectancy.1980		GDP.per.capita.1980	
Life.Expectancy.2000		GDP.per.capita.2000	
Life.Expectancy.2015		GDP.per.capita.2015	
Total.Pop.1940		Total.Pop.1960	
Total.Pop.1980		Total.Pop.2000	
Total.Pop.2015			

(b) What is the observational unit in this study?

Data Analysis Lab 1

Name: _____

3. Step 3. Explore the data.

- (a) Calculate the mean and standard deviation of life expectancy in 1940 and 2015 by continent. Additionally, determine the sample size by continent.
 - i. How have these statistics changed from 1940 to 2015?
 - ii. Have any continent's life expectancies grown faster? State which ones and provide quantitative support.
 - iii. Why do you believe Africa and Asia have much higher variability in life expectancy than the other regions?
- (b) Create a histogram of the Life Expectancy in 1940. Label your x-axis "1940 Life Expectancy (Years)" and your y-axis "Count." Additionally pick an appropriate figure title.
 - i. Describe the data distribution's shape, center, variability and identify any unusual observations of Life Expectancy in 1940.

Data Analysis Lab 1

Name: _____

- (c) Create a histogram of the life expectancy in 2015. Label your x-axis “2015 Life Expectancy (Years)” and your y-axis “Count.” Additionally pick an appropriate figure title.
 - i. Describe the data distribution’s shape, center, variability and identify any unusual observations of Life Expectancy in 2015.
 - ii. How has the life expectancy distribution’s shape, center, variability and unusual observations changed from 1940 to 2015?
- (d) Create a scatterplot that compares GDP per Capita (x-axis) to Life Expectancy (y-axis) for 1940.
 - i. For the scatterplot, color each point by continent, size each point by population size. Properly label the x and y-axis, title and legends. To change the size legend add the following layer: **scale_size_continuous()**.
 - ii. Describe the relationship between GDP per capita and life expectancy.

Data Analysis Lab 1

Name: _____

- (e) Create a scatterplot that compares GDP per Capita (x-axis) to Life Expectancy (y-axis) for 2015.
 - i. For the scatterplot, color each point by continent, size each point by population size. Properly label the x and y-axis, title and legends. To change the size legend add the following layer: `scale_size_continuous()`.
 - ii. How has the relationship between GDP per capita and life expectancy changed from 1940 to 2015?

- iii. Do the continents appear to be clustering together, if so do any continents appear to an advantage over others.

4. **Draw inferences beyond the data.** After exploring the data, do you believe certain continents have a larger life expectancy than others. Why and what is one potential reason?

5. **Formulate conclusions.** Is it reasonable for the researchers to conclude that as a country's GDP per capita increases, the country's life expectancy increases?

Data Analysis Lab 1

Name:

6. **Look back and ahead.** What other variables would you want to investigate if you had more time to collect data?