

Importing the data in R Studio using the read.csv() function.  
Create a data frame to store the data.

```
> mydata <- read.csv("/Users/sadie/Documents/R_Ex/gapminderDataFiveYear.csv")
> |
```

Check the structure of the data frame by using the str() function.

```
> mydata <- read.csv("/Users/sadie/Documents/R_Ex/gapminderDataFiveYear.csv")
> str(mydata)
'data.frame': 1704 obs. of 6 variables:
 $ country : chr "Afghanistan" "Afghanistan" "Afghanistan" "Afghanistan" ...
 $ year : int 1952 1957 1962 1967 1972 1977 1982 1987 1992 1997 ...
 $ pop : num 8425333 9240934 10267083 11537966 13079460 ...
 $ continent: chr "Asia" "Asia" "Asia" "Asia" ...
 $ lifeExp : num 28.8 30.3 32 34 36.1 ...
 $ gdpPercap: num 779 821 853 836 740 ...
```

Provide the summary statistics of the data frame using the summary() function.

```
> summary(mydata)
country      year      pop      continent
Length:1704  Min.   :1952  Min.   :6.001e+04  Length:1704
Class :character 1st Qu.:1966 1st Qu.:2.794e+06  Class :character
Mode :character  Median:1980  Median:7.024e+06  Mode :character
                Mean  :1980  Mean  :2.960e+07
                3rd Qu.:1993 3rd Qu.:1.959e+07
                Max.   :2007 Max.   :1.319e+09

lifeExp      gdpPercap
Min.   :23.60  Min.   : 241.2
1st Qu.:48.20 1st Qu.: 1202.1
Median :60.71 Median : 3531.8
Mean   :59.47 Mean   : 7215.3
3rd Qu.:70.85 3rd Qu.: 9325.5
Max.   :82.60 Max.   :113523.1
```

For the visualization, create a scatter plot of life expectancy versus GDP per capita using the ggplot2 package.

Add a color scale to the scatter plot based on the continent.

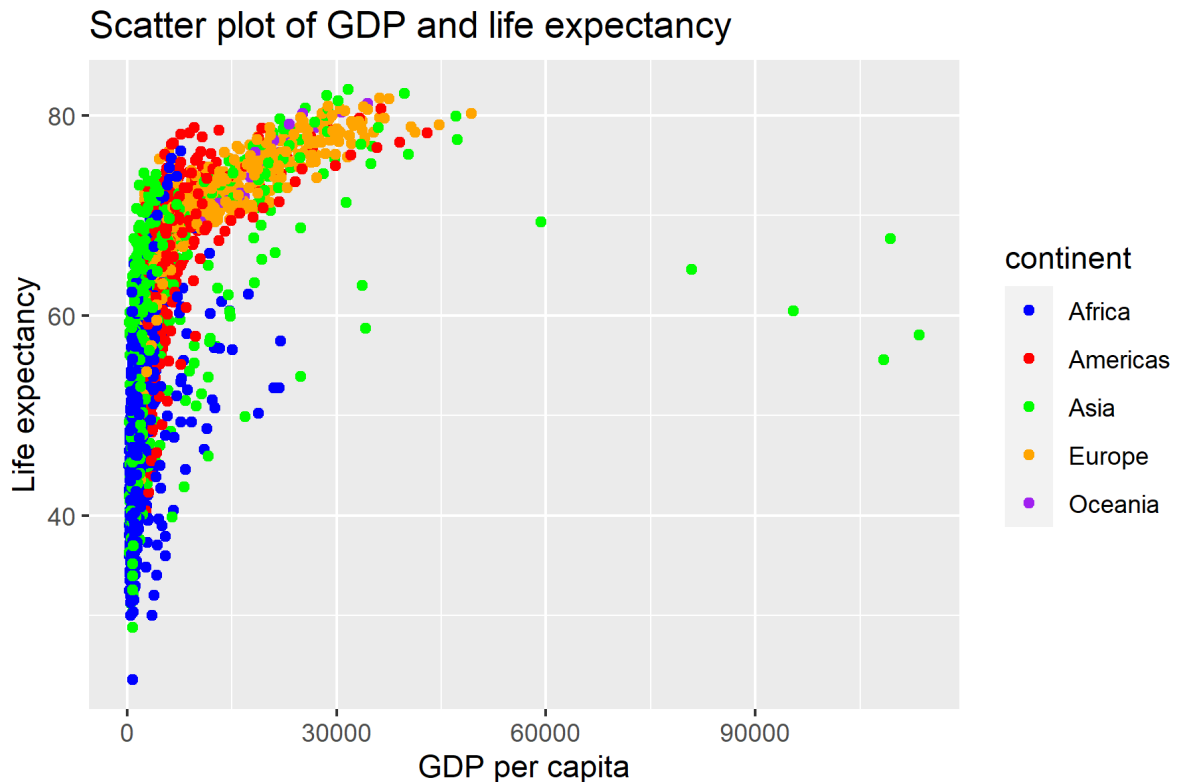
```
> library(ggplot2)
Error in library(ggplot2) : there is no package called 'ggplot2'
> install.packages("ggplot2")
WARNING: Rtools is required to build R packages but is not currently installed. Please download and install the appropriate version of Rtools before proceeding:

https://cran.rstudio.com/bin/windows/Rtools/
Installing package into 'C:/Users/sadie/AppData/Local/R/win-library/4.2'
(as 'lib' is unspecified)
also installing the dependencies 'colorspace', 'utf8', 'farver', 'labeling', 'munsell', 'R6', 'RColorBrewer', 'viridisLite', 'fansi', 'magrittr', 'pillar', 'pkgconfig', 'cli', 'glue', 'gtable', 'isoband', 'lifecycle', 'rlang', 'scales', 'tibble', 'vctrs', 'withr'
```

```

> library(ggplot2)
>
>
> ggplot(mydata, aes(x = gdpPercap, y = lifeExp, color = continent)) +
+   geom_point() +
+   scale_color_manual(values = c("blue", "red", "green", "orange", "purple")) +
+   labs(x = "GDP per capita", y = "Life expectancy", title = "Scatter plot of GDP and lif
e expectancy")
> |

```

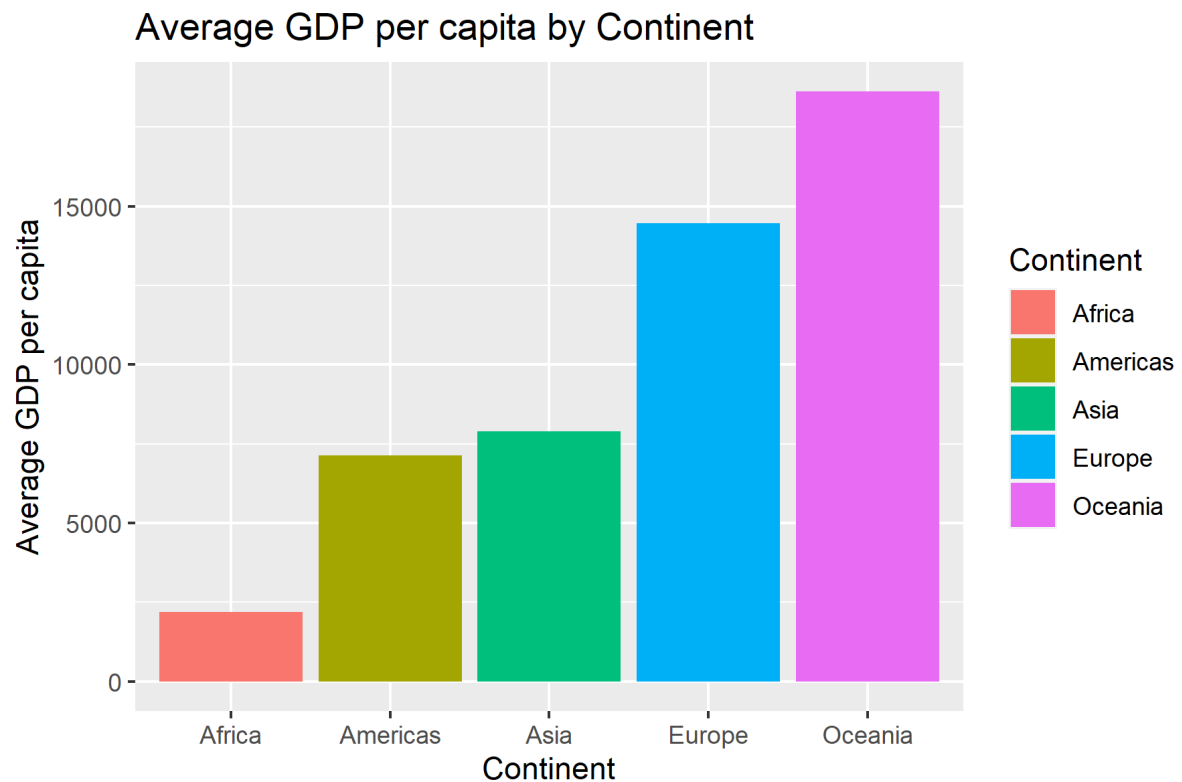


Create a bar chart of the average GDP per capita for each continent using the ggplot2 package.

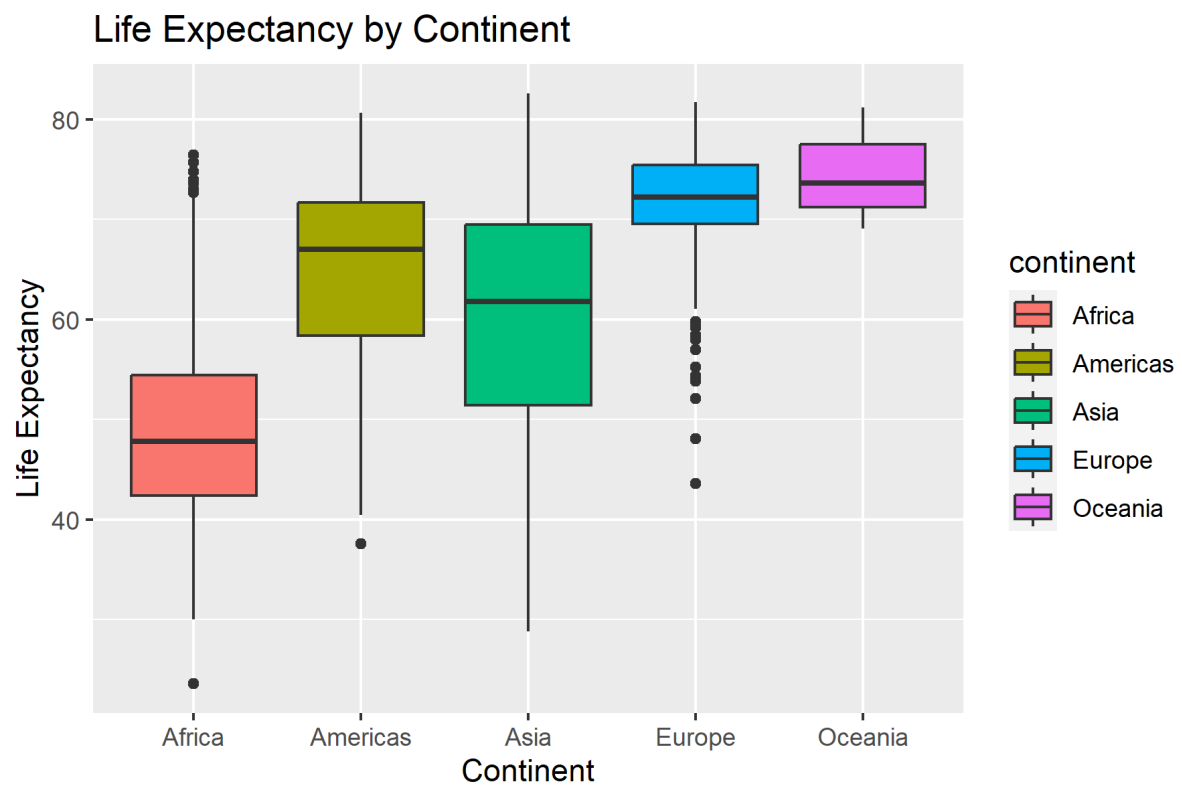
```

> ggplot(mydata, aes(x = gdpPercap, y = lifeExp, color = continent)) +
+   geom_point() +
+   scale_color_manual(values = c("blue", "red", "green", "orange", "purple")) +
+   labs(x = "GDP per capita", y = "Life expectancy", title = "Scatter plot of GDP and lif
e expectancy")
> # Calculate the average GDP per capita for each continent
> avg_gdp <- aggregate(mydata$gdpPercap, by = list(mydata$continent), FUN = mean)
> names(avg_gdp) <- c("Continent", "Average_GDP")
>
> # Create a bar chart of the average GDP per capita for each continent
> ggplot(avg_gdp, aes(x = Continent, y = Average_GDP, fill = Continent)) +
+   geom_bar(stat = "identity") +
+   labs(x = "Continent", y = "Average GDP per capita", title = "Average GDP per capita by
Continent")

```







Create a box plot of the life expectancy for each continent using the ggplot2 package.



Export the plots as image files using the `ggsave()` function.

```
> ggsave("LifeExpectancy.png", plot = last_plot(), width = 6, height = 4, dpi = 300)
> ggsave("averageGDP.png", plot = last_plot(), width = 6, height = 4, dpi = 300)
> ggsave("scatterplot.png", plot = last_plot(), width = 6, height = 4, dpi = 300)
~ |
```

 AlbertoVelasco_M2FA1.1	22/03/2023 6:40 am	R File	2 KB
 scatterplot	22/03/2023 6:40 am	PNG File	48 KB
 averageGDP	22/03/2023 6:40 am	PNG File	47 KB
 LifeExpectancy	22/03/2023 6:40 am	PNG File	36 KB

## Discussion:

For this exercise, we have explored methods of visualizing the data in RStudio. We first start by importing the CSV file to the R file by using the `read.csv` function and storing it in a data frame. Initial data exploration consists of using the `str()` function to check the structure of the imported data set columns' datatype with the sample values alongside it, and the `summary()` function which gives the statistical information per column such as the mean and median. After exploring the data we now get the gist of the data that we have and therefore can now perform visualization. The `ggplot2` library must be installed first and imported to the file. The first visualization we created is the scatter plot of the GDP and life expectancy which indicated where each continent lies on the graph. We see some outliers belonging to the Asia continent, while the points generally indicate that there is a somehow direct correlation between life expectancy and the GDP per capita. We see that as the points generally cluster together until the 70s in the y-axis, there is some increase in life expectancy that goes beyond around 75 starting on the 15000 mark of the GDP per capita. We also see that most of the low life expectancy and GDP per capita are populated by the Africas data points, and the Asias data points start to cluster starting past the 60 mark of the y-axis. The next visualization is the bar chart of the average GDP per capita for each continent where we see that the poorly represented Oceania (probably due to population) has the highest score, following Europe, and then Asia in the Middle which is then followed shortly after the Americas, and then by a significant margin drop, the Africas. The life expectancy is then represented with a box plot which is suitable as it represents the ranges. We see that the order of the continents is still similar to the previous GDP per capita in this chart of Life Expectancy. We see that the life expectancy ranges of the leading continents, Europe and Oceania, are significantly higher with short ranges, unlike the rest; and the Africas having significantly shorter life expectancy. We can generally surmise the presence of inequality here, especially with the continents with a large range of life expectancy as the resources are not distributed well for the average person to have a high life expectancy. Lastly, we use the `ggsave()` function to save the plots to our local storage.