Exploratory Data Analysis

Exploratory Data Analysis

This is a document for exploring the global indicator dataset that was extracted from the combination of numerous datasets

Loading packages and dataset

The tidyverse package was loaded below.

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --
```

```
## v ggplot2 3.3.3      v purrr   0.3.4
## v tibble  3.1.0      v dplyr   1.0.5
## v tidyr   1.1.3      v stringr 1.4.0
## v readr   1.4.0      v forcats 0.5.1
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

The dataset is loaded below

```
df_world <- read.csv("~/Global-Analysis/Global_data.csv")
```

Data Comprehension

The datasets was reviewed below to improve our understanding of the dataset

```
head(df_world)
```

```
##        Country Year Life.Expectancy Income Population    Continent Overall.Score
## 1 Afghanistan 1995            53.3   1030   18100000         Asia           0.0
## 2      Albania 1995            74.6   4130    3110000       Europe          49.7
## 3      Algeria 1995            72.9   9340   28800000       Africa          55.7
## 4       Angola 1995            49.5   3410   13900000       Africa          27.4
## 5    Argentina 1995            73.3  14000   34800000 The Americas          68.0
## 6      Armenia 1995            69.9   2170    3220000         Asia           0.0
##   Property.Rights Government.Integrity Judicial.Effectiveness Tax.Burden
## 1               0                    0                      0        0.0
```

```
## 2              50          10              0       81.7
## 3              50          50              0       48.8
## 4              30          30              0       61.6
## 5              70          50              0       80.7
## 6               0           0              0        0.0
##   Government.Spending Fiscal.Health Business.Freedom Labor.Freedom
## 1                 0.0             0                0             0
## 2                34.3             0               70             0
## 3                69.5             0               70             0
## 4                 0.0             0               40             0
## 5                86.6             0               85             0
## 6                 0.0             0                0             0
##   Monetary.Freedom Trade.Freedom Investment.Freedom Financial.Freedom
## 1              0.0           0.0                  0                 0
## 2             22.1          59.0                 70                50
## 3             59.2          54.2                 50                50
## 4              0.0          25.0                 30                30
## 5             61.1          58.4                 70                50
## 6              0.0           0.0                  0                 0
##   Income.Index Average.years.of.Schooling Expected.years.of.Schooling
## 1        0.393                        1.9                         4.2
## 2        0.584                        8.0                        10.2
## 3        0.654                        4.7                         9.8
## 4        0.533                        0.0                         3.9
## 5        0.777                        8.3                        13.3
## 6        0.519                       10.4                        10.2
##   Education.Index Human.Development.Index
## 1           0.179                   0.331
## 2           0.550                   0.637
## 3           0.431                   0.595
## 4           0.000                   0.000
## 5           0.648                   0.741
## 6           0.631                   0.627
```

```
colnames(df_world)
```

```
##  [1] "Country"                    "Year"
##  [3] "Life.Expectancy"            "Income"
##  [5] "Population"                 "Continent"
##  [7] "Overall.Score"             "Property.Rights"
##  [9] "Government.Integrity"       "Judicial.Effectiveness"
## [11] "Tax.Burden"                 "Government.Spending"
## [13] "Fiscal.Health"              "Business.Freedom"
## [15] "Labor.Freedom"              "Monetary.Freedom"
## [17] "Trade.Freedom"              "Investment.Freedom"
## [19] "Financial.Freedom"          "Income.Index"
## [21] "Average.years.of.Schooling" "Expected.years.of.Schooling"
## [23] "Education.Index"            "Human.Development.Index"
```

Structure of the dataset

```
str(df_world)
```

```
## 'data.frame':    4014 obs. of  24 variables:
##  $ Country                  : Factor w/ 168 levels "Afghanistan",..: 1 2 3 4 5 6 7 8 9 10 ...
##  $ Year                     : int  1995 1995 1995 1995 1995 1995 1995 1995 1995 1995 ...
##  $ Life.Expectancy          : num  53.3 74.6 72.9 49.5 73.3 69.9 78.2 76.8 65 71 ...
##  $ Income                   : int  1030 4130 9340 3410 14000 2170 30400 33800 3430 42900 ...
##  $ Population               : int  18100000 3110000 28800000 13900000 34800000 3220000 18000000 799
##  $ Continent                : Factor w/ 6 levels "Africa","Asia",..: 2 3 1 1 6 2 5 3 2 4 ...
##  $ Overall.Score            : num  0 49.7 55.7 27.4 68 0 74.1 70 0 76.2 ...
##  $ Property.Rights          : num  0 50 50 30 70 0 90 90 0 60 ...
##  $ Government.Integrity     : num  0 10 50 30 50 0 70 90 0 70 ...
##  $ Judicial.Effectiveness   : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Tax.Burden               : num  0 81.7 48.8 61.6 80.7 0 59.6 46.3 0 99.4 ...
##  $ Government.Spending      : num  0 34.3 69.5 0 86.6 0 53.9 9.6 0 71.7 ...
##  $ Fiscal.Health            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Business.Freedom         : num  0 70 70 40 85 0 70 70 0 100 ...
##  $ Labor.Freedom            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Monetary.Freedom         : num  0 22.1 59.2 0 61.1 0 86.7 82.9 0 86.7 ...
##  $ Trade.Freedom            : num  0 59 54.2 25 58.4 0 77 81 0 78.4 ...
##  $ Investment.Freedom       : int  0 70 50 30 70 0 70 70 0 50 ...
##  $ Financial.Freedom        : int  0 50 50 30 50 0 90 90 0 70 ...
##  $ Income.Index             : num  0.393 0.584 0.654 0.533 0.777 0.519 0.872 0.907 0.513 0.926 ...
##  $ Average.years.of.Schooling : num  1.9 8 4.7 0 8.3 10.4 11.8 8.9 10.2 7.2 ...
##  $ Expected.years.of.Schooling: num  4.2 10.2 9.8 3.9 13.3 10.2 18.8 14.9 10 13.7 ...
##  $ Education.Index          : num  0.179 0.55 0.431 0 0.648 0.631 0.894 0.709 0.618 0.619 ...
##  $ Human.Development.Index   : num  0.331 0.637 0.595 0 0.741 0.627 0.888 0.825 0.604 0.778 ...
```

Renaming certain columns

```
df_world <- df_world %>%
  rename(c(Overall_Score = Overall.Score,
          Life_Expectancy = Life.Expectancy,
          Government_Integrity = Government.Integrity,
          Government_Spending = Government.Spending,
          Property_Rights = Property.Rights,
          Fiscal_Health = Fiscal.Health,
          Business_Freedom = Business.Freedom,
          Monetary_Freedom = Monetary.Freedom,
          Trade_Freedom = Trade.Freedom,
          Investment_Freedom = Investment.Freedom,
          Financial_Freedom = Financial.Freedom,
          Income_Index = Income.Index,
          Tax_Burden = Tax.Burden,
          Labor_Freedom = Labor.Freedom,
          Average_schooling = Average.years.of.Schooling,
          Expected_schooling = Expected.years.of.Schooling,
          Education_Index = Education.Index,
          Human_Development_Index = Human.Development.Index,
          Judicial_Effectiveness = Judicial.Effectiveness
          ))
```

Filtering out information

```
df_nigeria <- df_world %>%
  filter(Country == 'Nigeria') %>%
  select(Business_Freedom, Year)
df_nigeria
```

```
##     Business_Freedom Year
## 1               55.0 1995
## 2               55.0 1996
## 3               55.0 1997
## 4               55.0 1998
## 5               55.0 1999
## 6               55.0 2000
## 7               55.0 2001
## 8               55.0 2002
## 9               55.0 2003
## 10              55.0 2005
## 11              50.0 2006
## 12              58.6 2007
## 13              52.9 2008
## 14              55.1 2009
## 15              53.2 2010
## 16              51.6 2011
## 17              55.6 2012
## 18              55.7 2013
## 19              48.0 2014
## 20              48.3 2015
## 21              48.7 2016
## 22              48.9 2017
## 23              49.3 2018
## 24              51.2 2019
```

Charts

Scatter plot function

```
## Human Development Index vs Economic Freedom Index
scatter1 <- function(df, year) {
  df %>%
    filter(Year == year) %>%
    ggplot() +
    geom_point(aes(x = Overall_Score,
                   y = Human_Development_Index,
                   color = Continent), na.rm = TRUE) +
    labs(title = sprintf("Human Development Index vs Economic Freedom Index in %d", year),
         x = "Economic Freedom Index",
         y = "Human Development Index")

}
## Education Index vs Economic Freedom Index
scatter2 <- function(df, year) {
  df %>%
    filter(Year == year) %>%
```

```r
    ggplot() +
    geom_point(aes(x = Overall_Score,
                   y = Education_Index,
                   color = Continent), na.rm = TRUE) +
    labs(title = sprintf("Education Index vs Economic Freedom Index in %d", year),
         x = "Economic Freedom Index",
         y = "Education Index")
}
## Life Expectancy vs Economic Freedom Index
scatter3 <- function(df, year) {
  df %>%
    filter(Year == year) %>%
    ggplot() +
    geom_point(aes(x = Overall_Score,
                   y = Life_Expectancy,
                   color = Continent), na.rm = TRUE) +
    labs(title = sprintf("Life Expectancy vs Economic Freedom Index in %d", year),
         x = "Economic Freedom Index",
         y = "Life Expectancy")
}
## Government Integrity vs Government Spending
scatter4 <- function(df, year) {
  df %>%
    filter(Year == year) %>%
    ggplot() +
    geom_point(aes(x = Government_Spending,
                   y = Government_Integrity,
                   color = Continent), na.rm = TRUE) +
    labs(title = sprintf("Government Integrity vs Government Spending in %d", year),
         x = "Government Spending",
         y = "Government Integrity")
}
```

Human Development Index vs Economic Freedom Index

```r
scatter1(df_world, 2019)
```

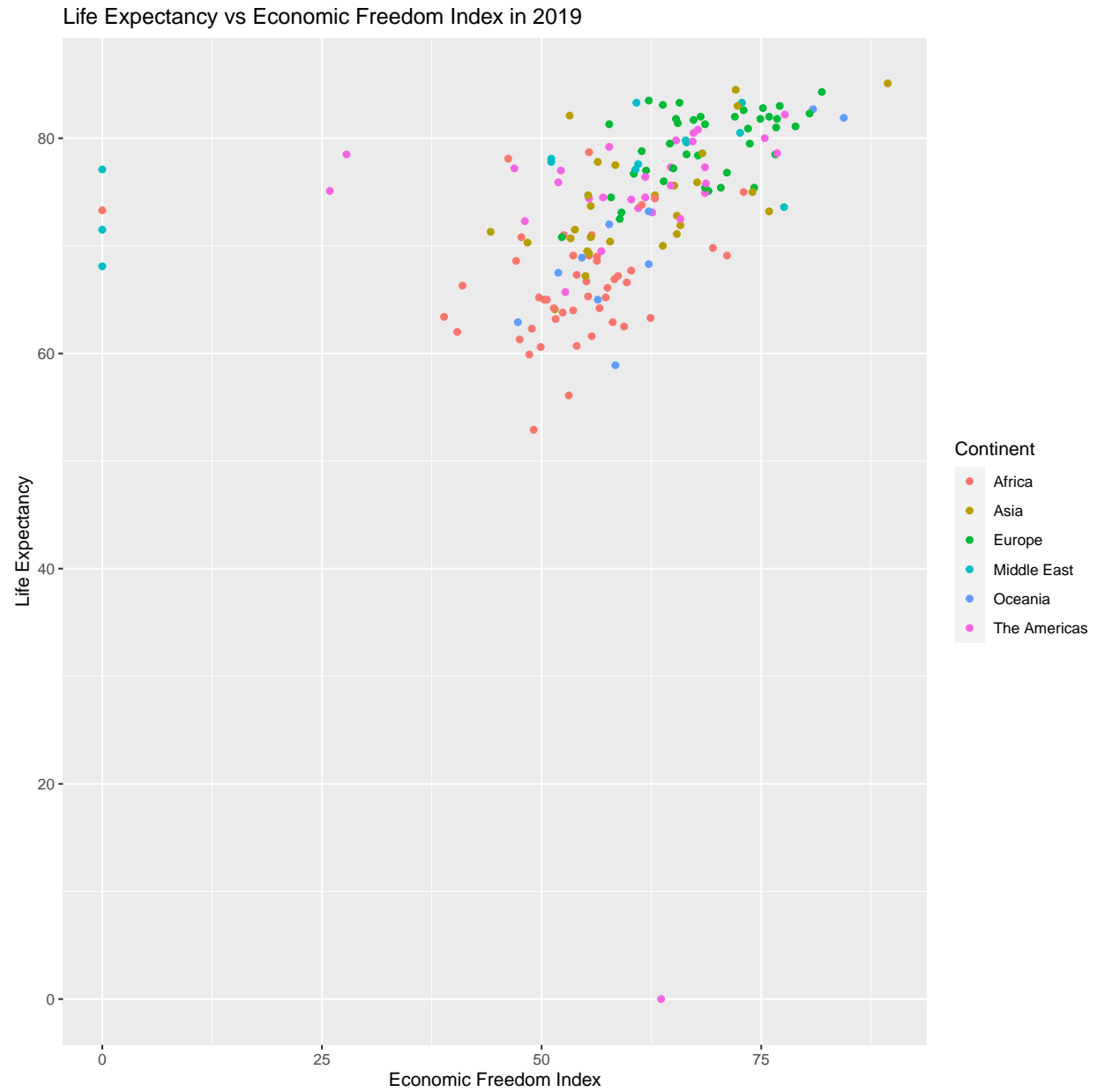Human Development Index vs Economic Freedom Index in 2019
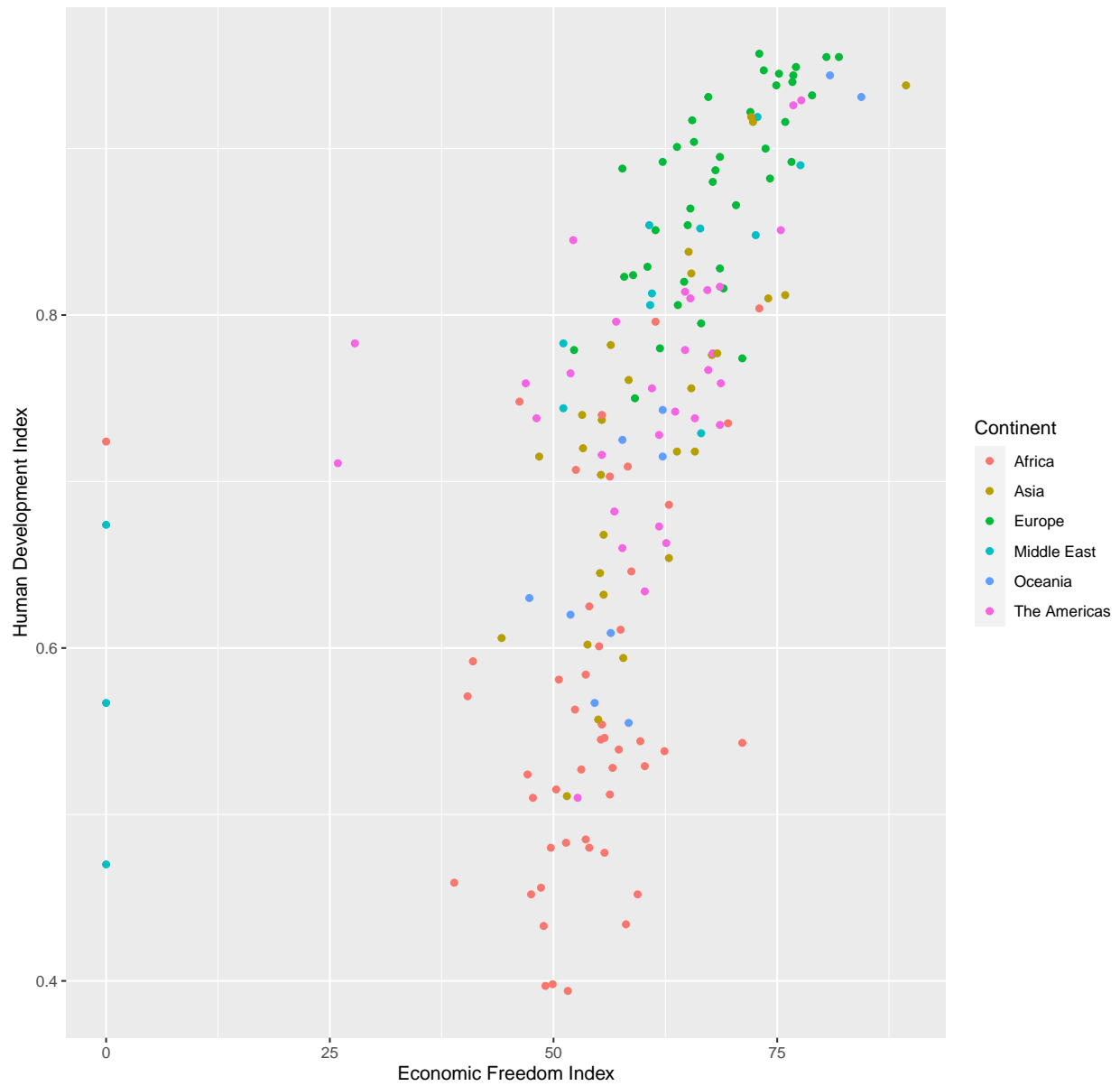
Education Index vs Economic Freedom Index
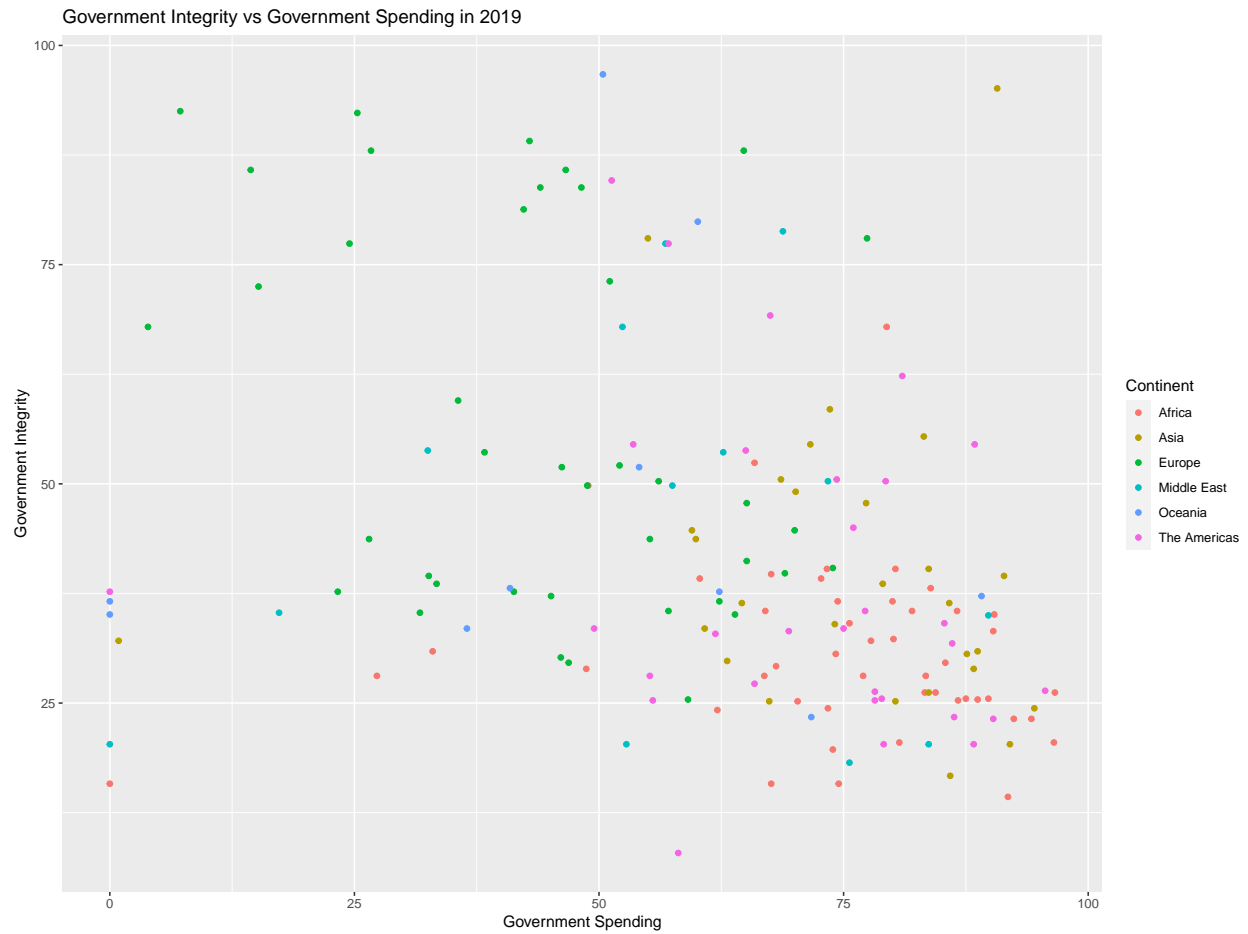
```
scatter2(df_world, 2019)
```

Education Index vs Economic Freedom Index in 2019
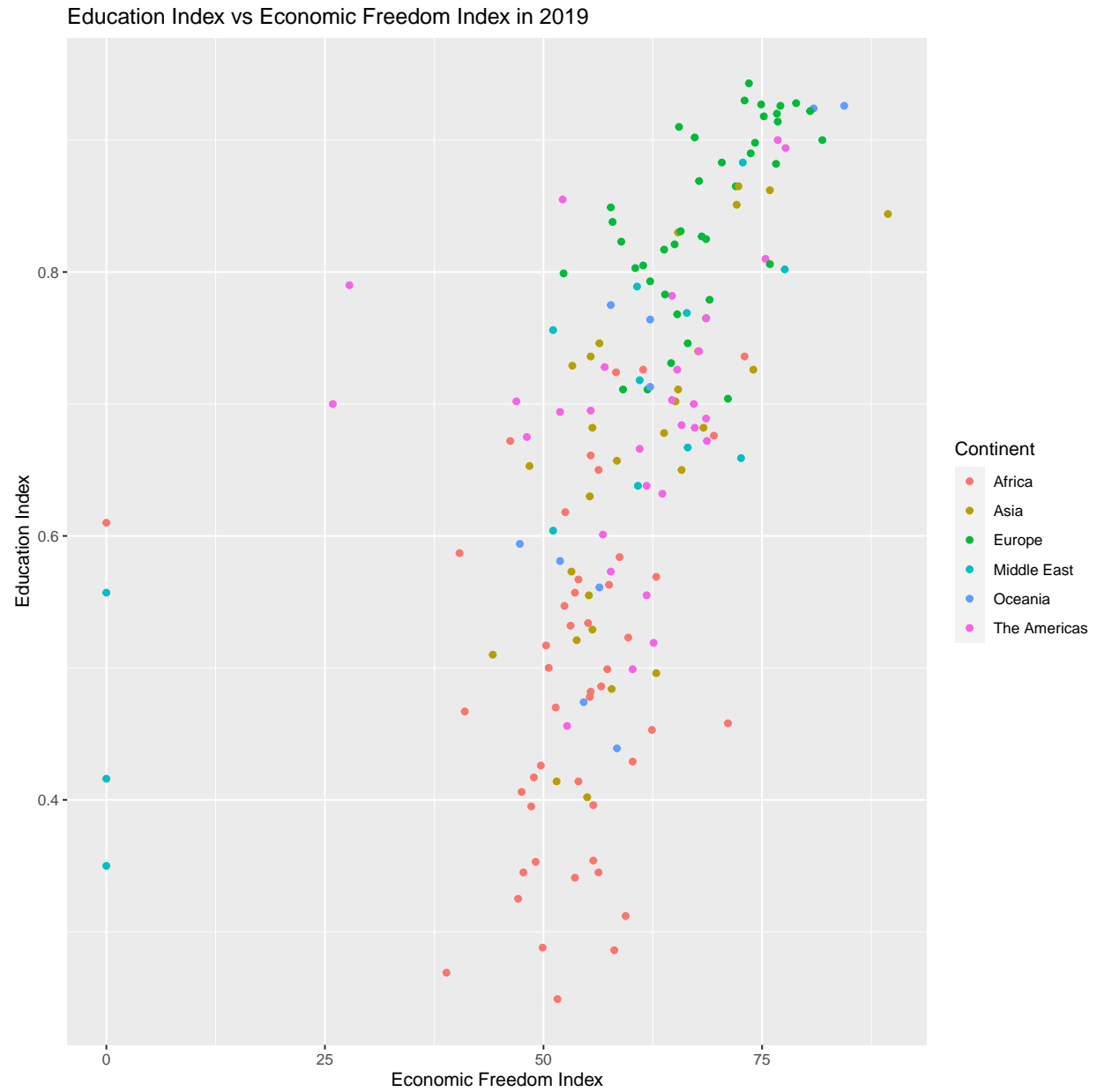
Life Expectancy vs Economic Freedom Index

```
scatter3(df_world, 2019)
```

Life Expectancy vs Economic Freedom Index in 2019

Government Integrity vs Government Spending

```
scatter4(df_world, 2019)
```
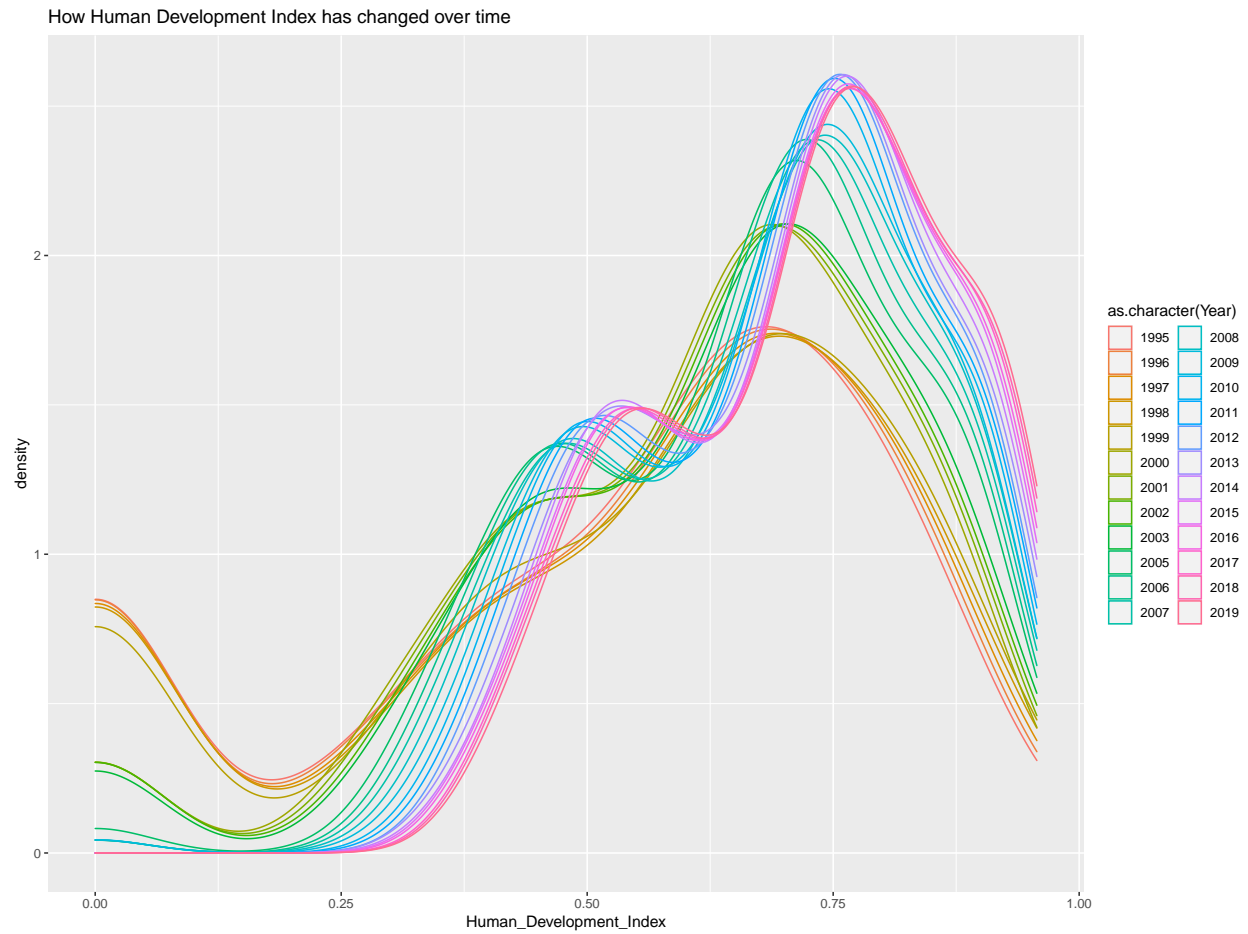
Government Integrity vs Government Spending in 2019

Kernel Density Estimation Plot

Human Development Index

```
ggplot(df_world) +
  geom_density(aes(x = Human_Development_Index,
                   color = as.character(Year)),
               show.legend = NA) +
  labs(title = "How Human Development Index has changed over time")
```

How Human Development Index has changed over time

Economic Freedom Index

```
ggplot(df_world) +
  geom_density(aes(x = Overall_Score,
                   color = as.character(Year)),
               show.legend = NA) +
  labs(title = "How Economic Freedom Index has changed over time")
```

How Economic Freedom Index has changed over time
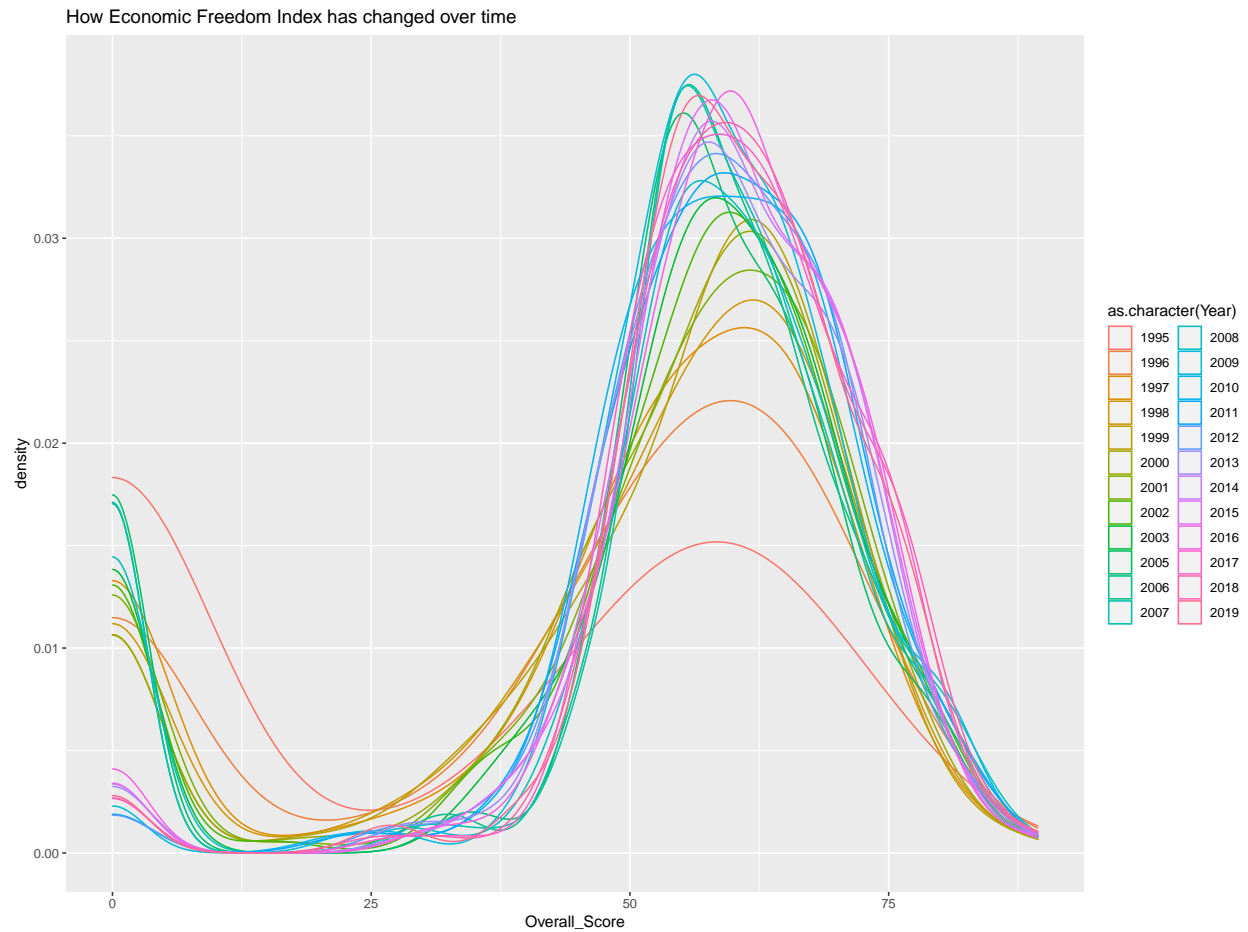

Box Plot

Continent Disparities for Human Development Index in 2019

```
df_world %>%
  filter(Year == 2019) %>%
  ggplot() +
  geom_boxplot(aes(x = Continent, y = Human_Development_Index,
                   color = Continent)) +
  labs(title = "How continents differ based on Human Development Index in 2019")
```

How continents differ based on Human Development Index in 2019
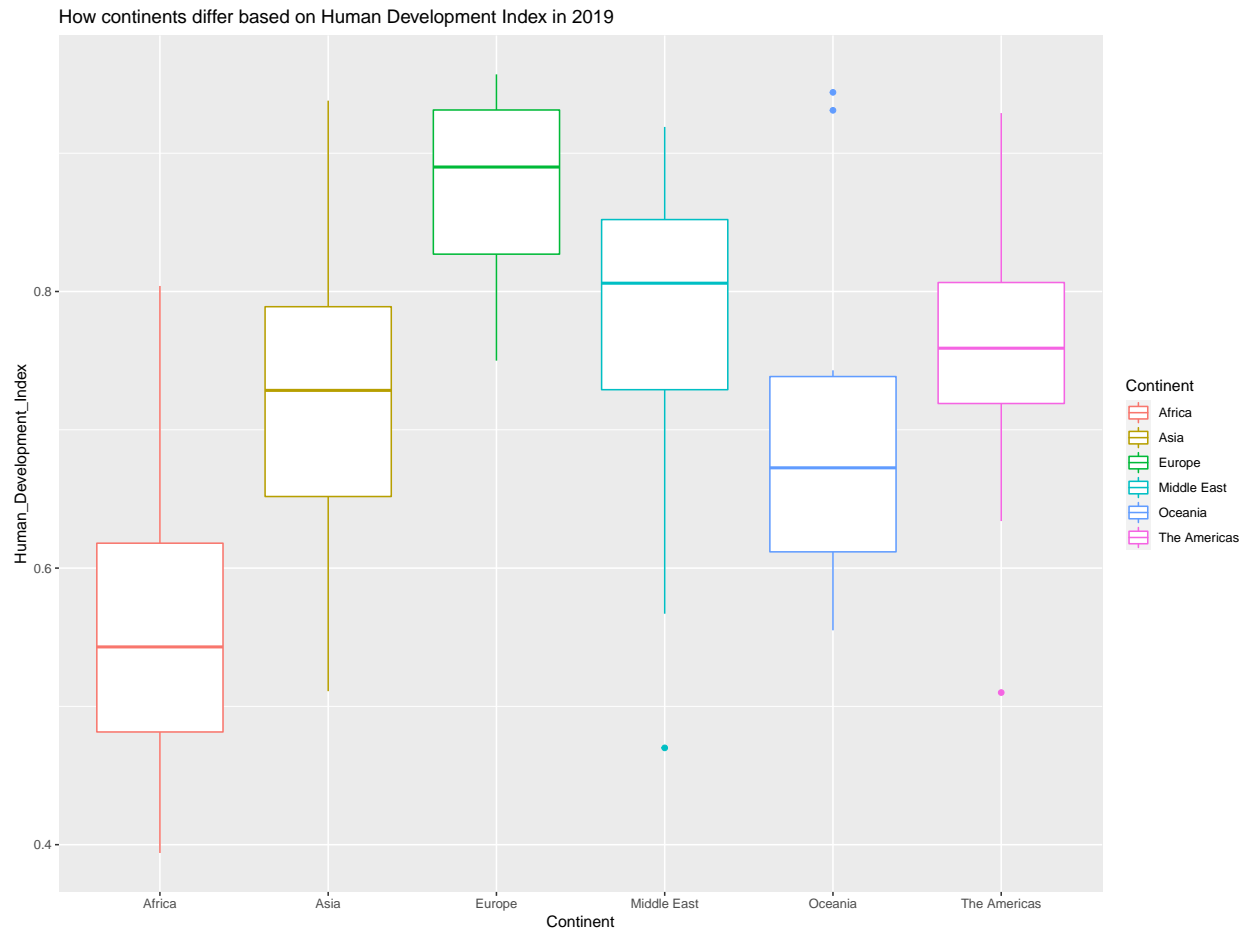
Continent Disparities for Economic Freedom Index in 2019

```
df_world %>%
  filter(Year == 2019) %>%
  ggplot() +
  geom_boxplot(aes(x = Continent, y = Overall_Score,
                   color = Continent)) +
  labs(title = "How continents differ based on Economic Freedom Index in 2019",
       y = "Economic Freedom Index")
```

How continents differ based on Economic Freedom Index in 2019
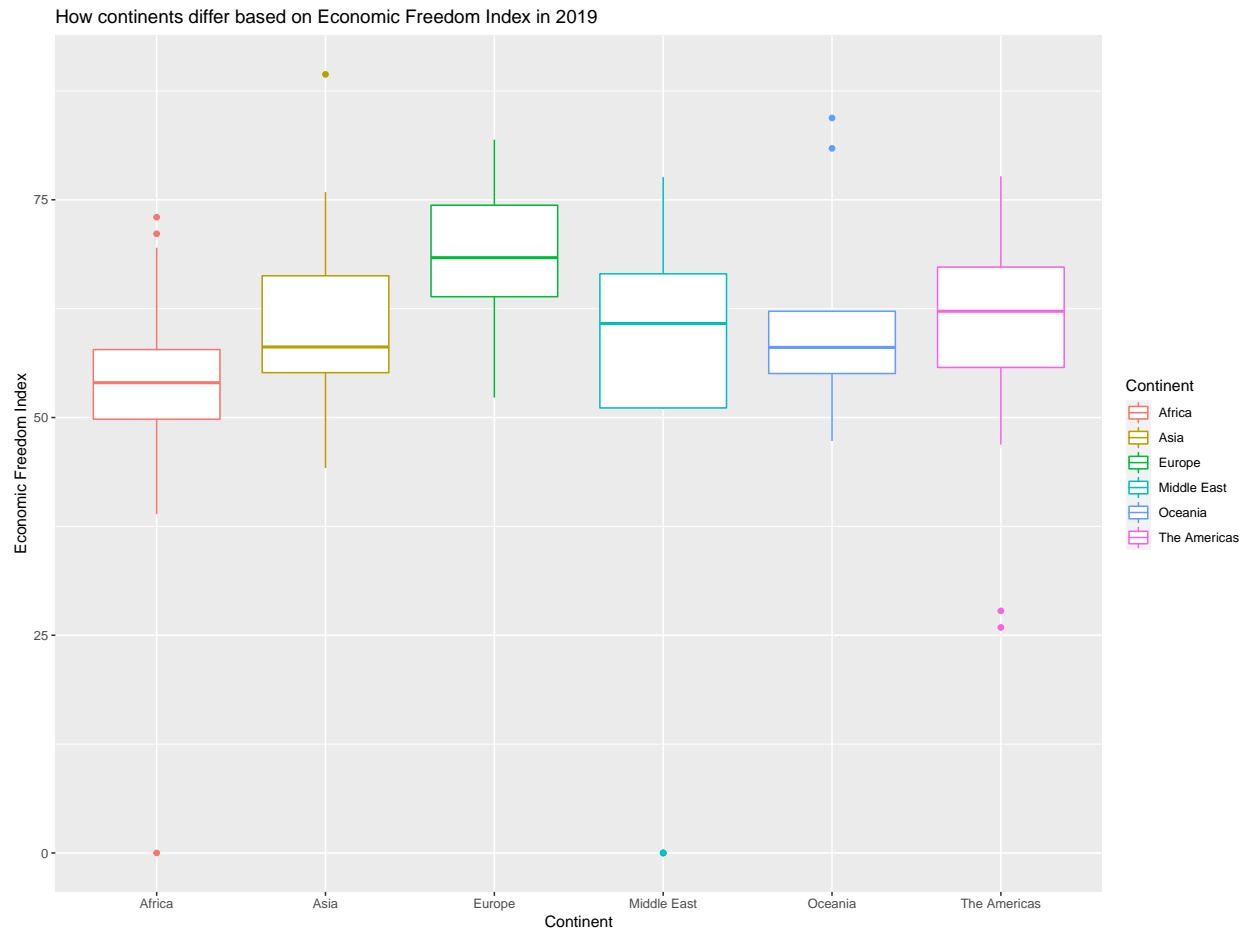
Bar Chart

Average schooling years across continents

```
df_world %>%
  filter(Year == 2019) %>%
  group_by(Continent) %>%
  summarize(mean_school = mean(Average_schooling, na.rm = TRUE)) %>%
  ggplot() +
  geom_bar(aes(x = Continent, y = mean_school,
               fill = Continent), stat = "identity") +
  coord_polar() +
  labs(title = "Average years of schooling across continents",
       y = "Average years of schooling")
```
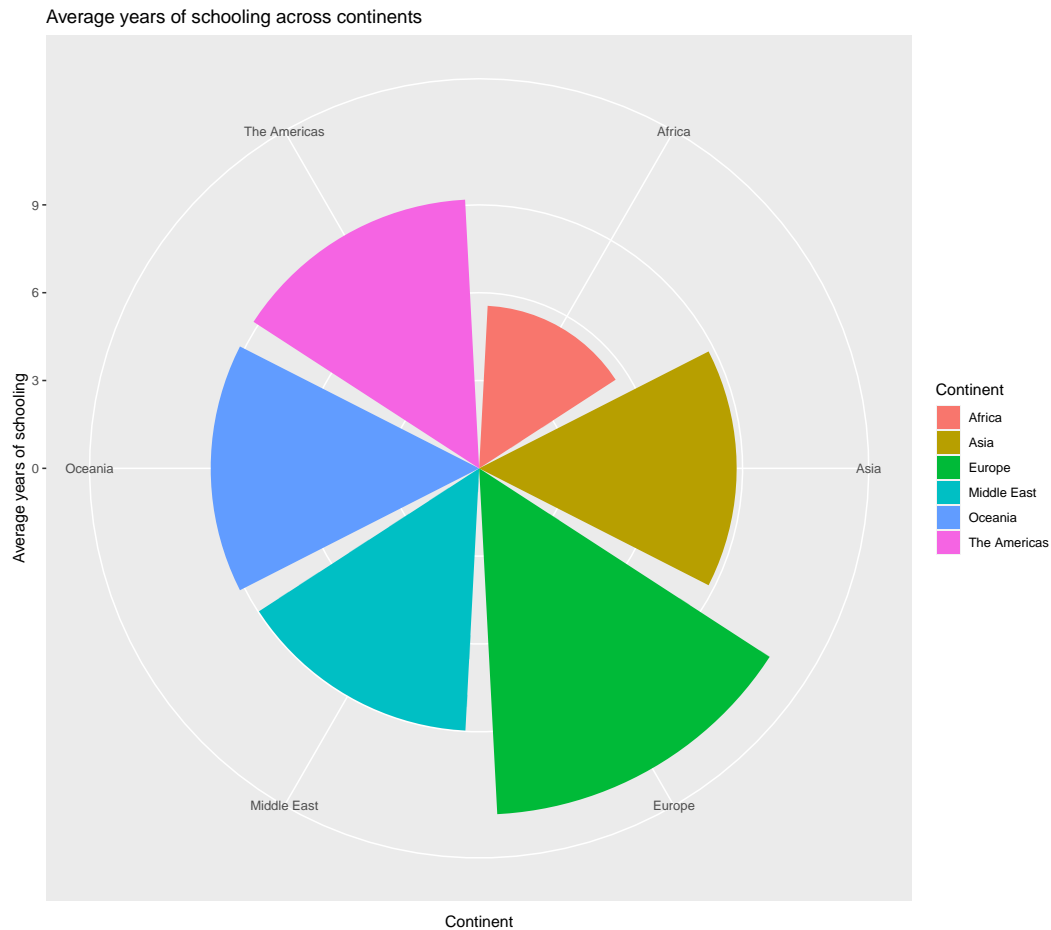
Average years of schooling across continents

Average life expectancy across continents

```
df_world %>%
  filter(Year == 2019) %>%
  group_by(Continent) %>%
  summarize(mean_life = mean(Life_Expectancy, na.rm = TRUE)) %>%
  ggplot() +
  geom_bar(aes(x = Continent, y = mean_life,
               fill = Continent), stat = "identity") +
  labs(title = "Average life expectancy across continents",
       y = "Average life expectancy") +
  theme_classic()
```

Average life expectancy across continents