

Preston Curve

Team A

2/16/2022



```
library(tidyverse)
library(ggrepel)
library(plotly)
library(styler)
```

(1) Import CSV as tibble

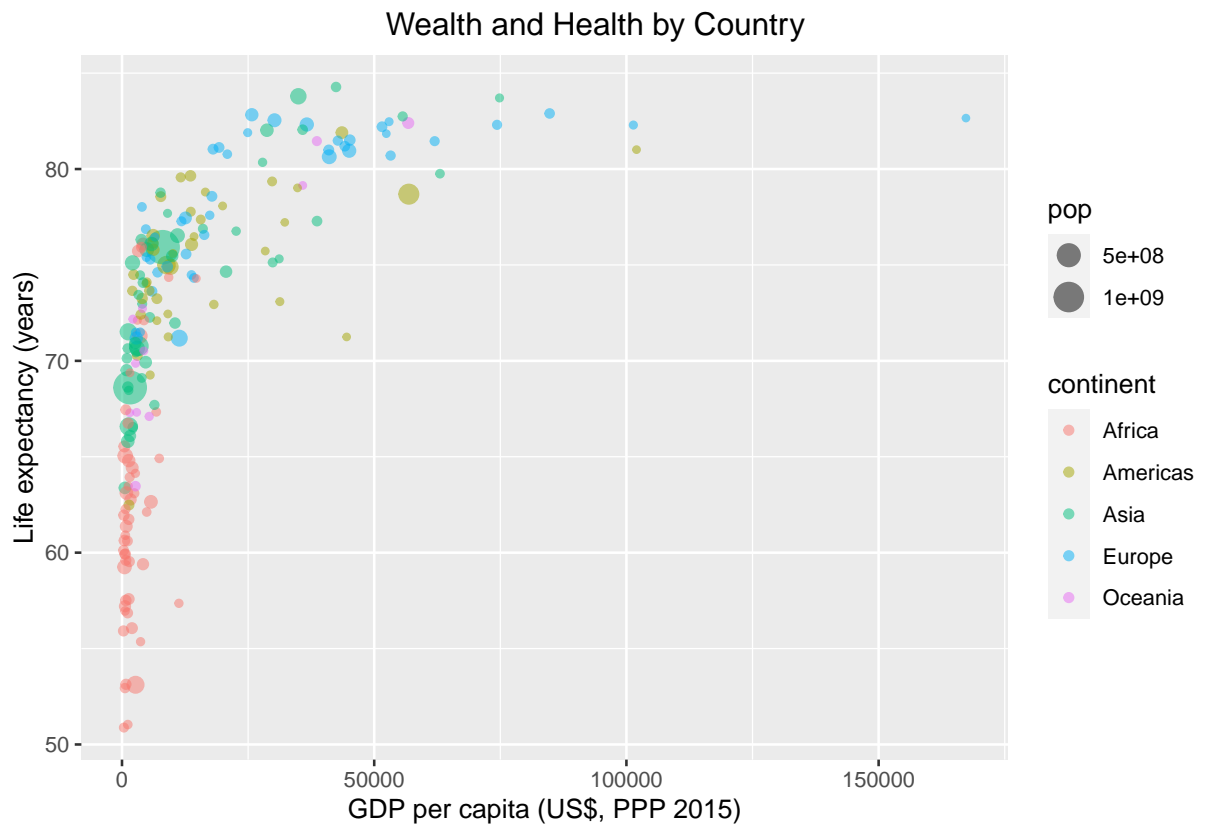
```
life_quality <- read_csv("life_quality.csv")
```

(2) Make a bubble chart with ggplot2

```
Q2 <- ggplot(
  data = life_quality,
  aes(
    x = gdp_per_capita,
    y = life_expectancy,
    label = country_name
  )
) +
  geom_point(aes(
    size = pop,
    colour = continent
  ),
  # make the points semi-transparent
  alpha = 0.5
) +
  # add titles
  labs(
    x = "GDP per capita (US$, PPP 2015)",
    y = "Life expectancy (years)",
    title = "Wealth and Health by Country",
    caption = "Source : World Bank"
  ) +
  # adjust the title to the center
  theme(plot.title = element_text(hjust = 0.6))
```

Q2



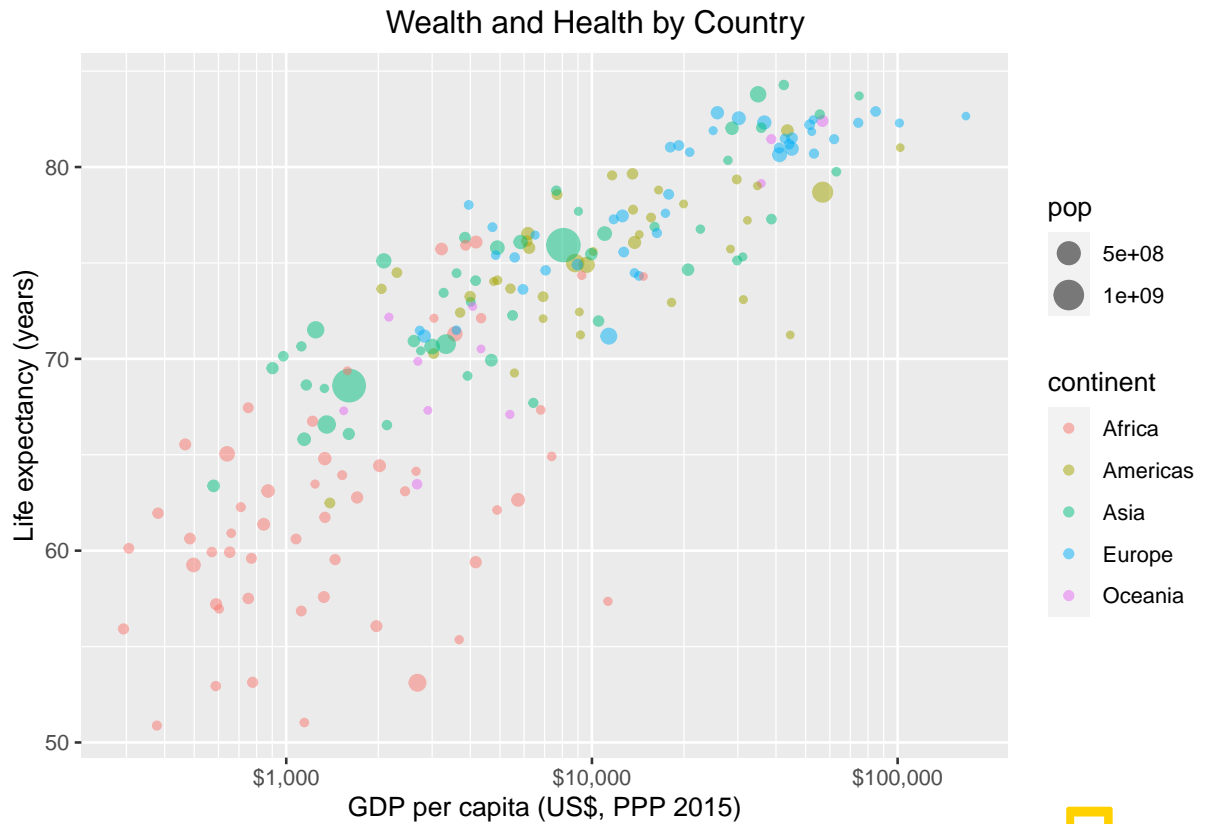


Source : World Bank

(3) Change the x-coordinates to a logarithmic scale

```
Q3 <- Q2 +
  scale_x_continuous(
    trans = "log10",
    breaks = c(1000, 10000, 100000),
    # add labels
    labels = c("$1,000", "$10,000", "$100,000"),
    # add minor ticks
    minor_breaks = c(
      seq(0, 1000, 100),
      seq(1000, 10000, 1000),
      seq(10000, 100000, 10000)
    )
  )
```

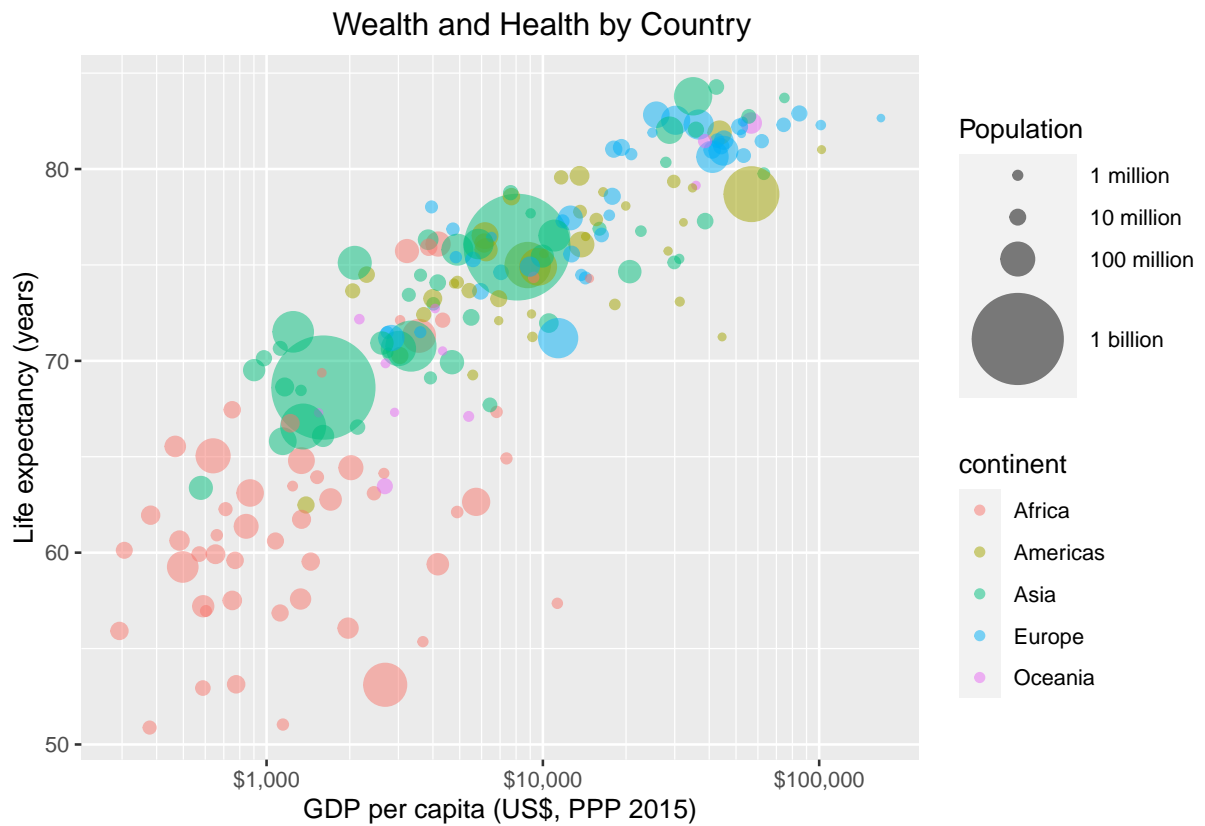
Q3



Source : World Bank

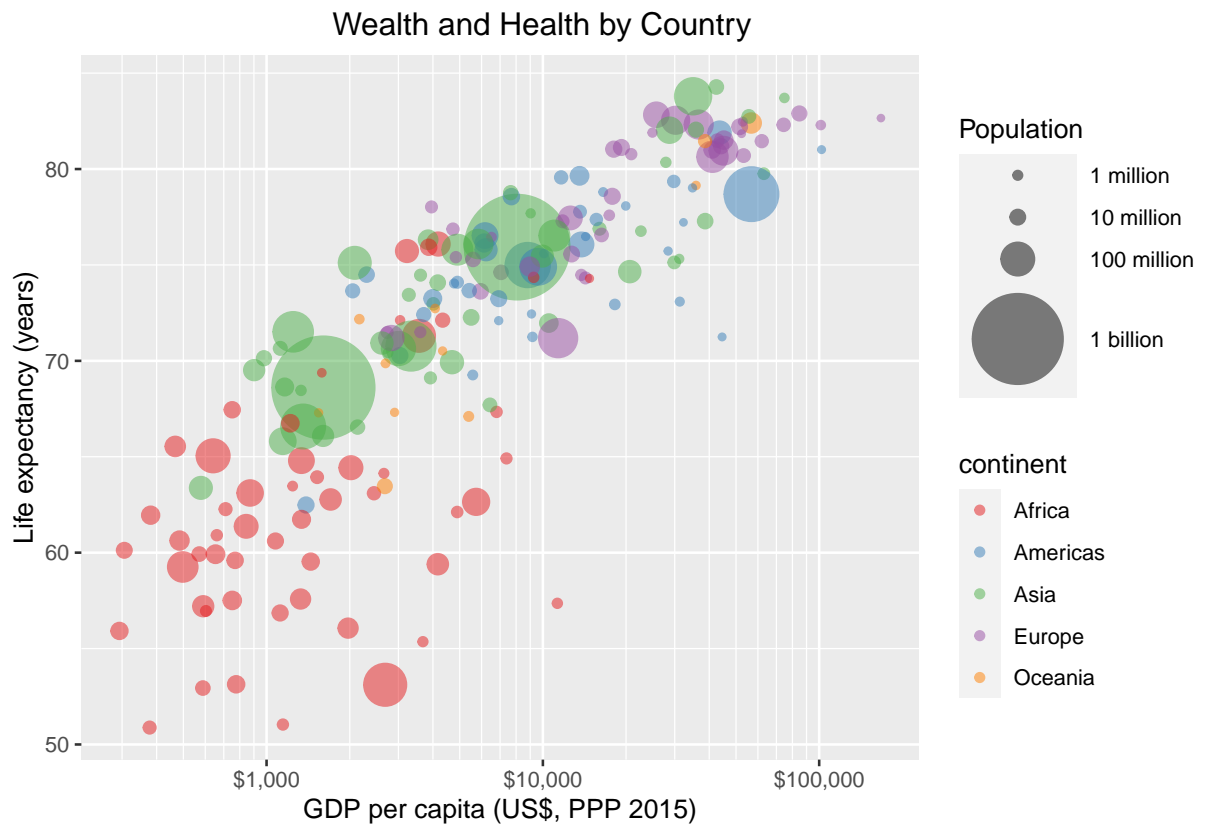
- (4) Use `scale_size_area()` so that the areas of the bubbles in the legend represent populations of 1 million, 10 million, 100 million and 1 billion. Change the numbers in the legends from '1e+06', '1e+07', '1e+08', '1e+09' to the reader-friendlier strings '1 million', '10 million', '100 million' and '1 billion'. Increase `max_size` so that the bubble areas are approximately the same as in the Gapminder figure.

```
Q4 <- Q3 +
  scale_size(
    range = c(1, 1000)
  ) +
  scale_size_continuous(
    name = "Population",
    range = c(1, 20),
    breaks = 10^(6:9),
    labels = c(
      "1 million",
      "10 million",
      "100 million",
      "1 billion"
    )
  )
Q4
```



(5) Change the colour scale to the ColorBrewer palette 'Set1'.

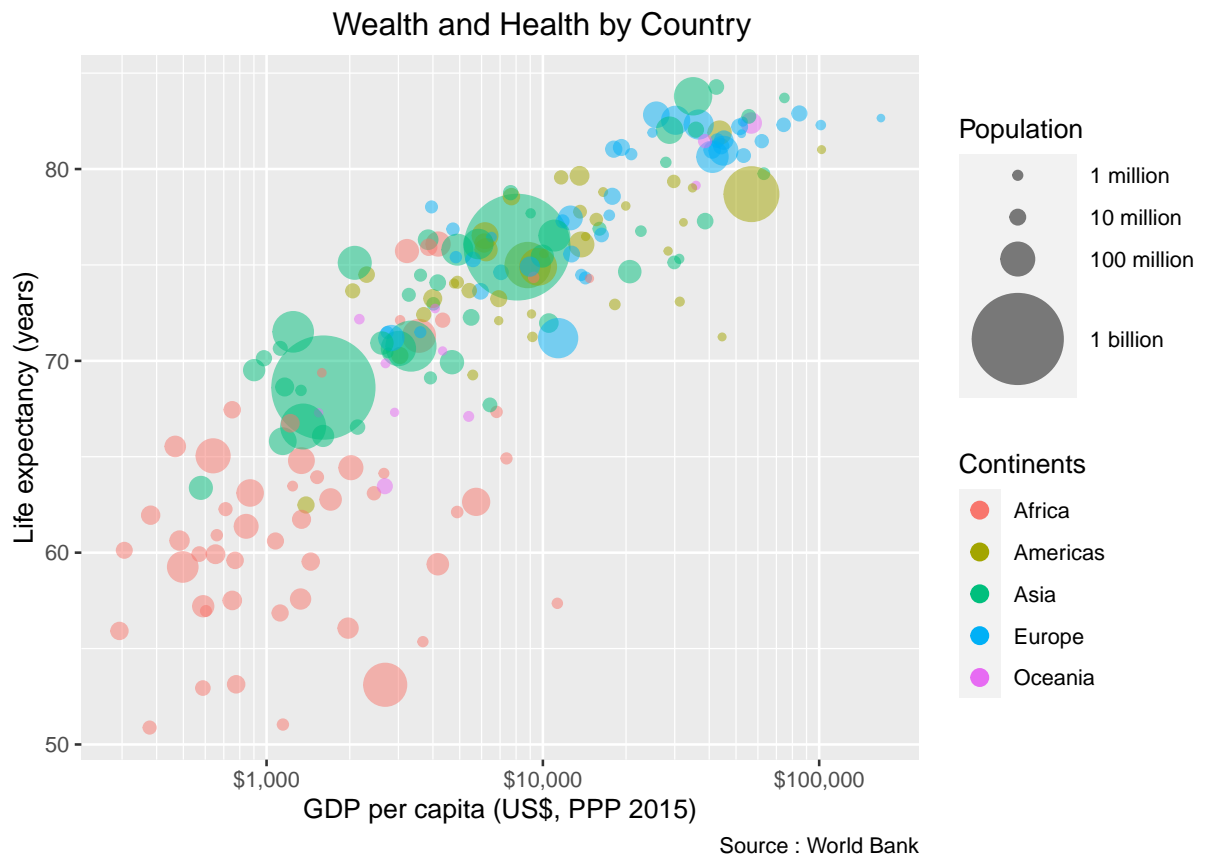
```
Q5 <- Q4 +
  scale_colour_brewer(palette = "Set1")
Q5
```



- (6) Override the alpha aesthetic in the legend. Also increase the sizes of the circles in the colour legend so that the colours are easier to read.

```
Q6 <- Q5 +
  guides(colour = guide_legend(override.aes = list(alpha = 1, size = 3))) +
  # Change the legend title
  scale_colour_discrete("Continents")
Q6
```

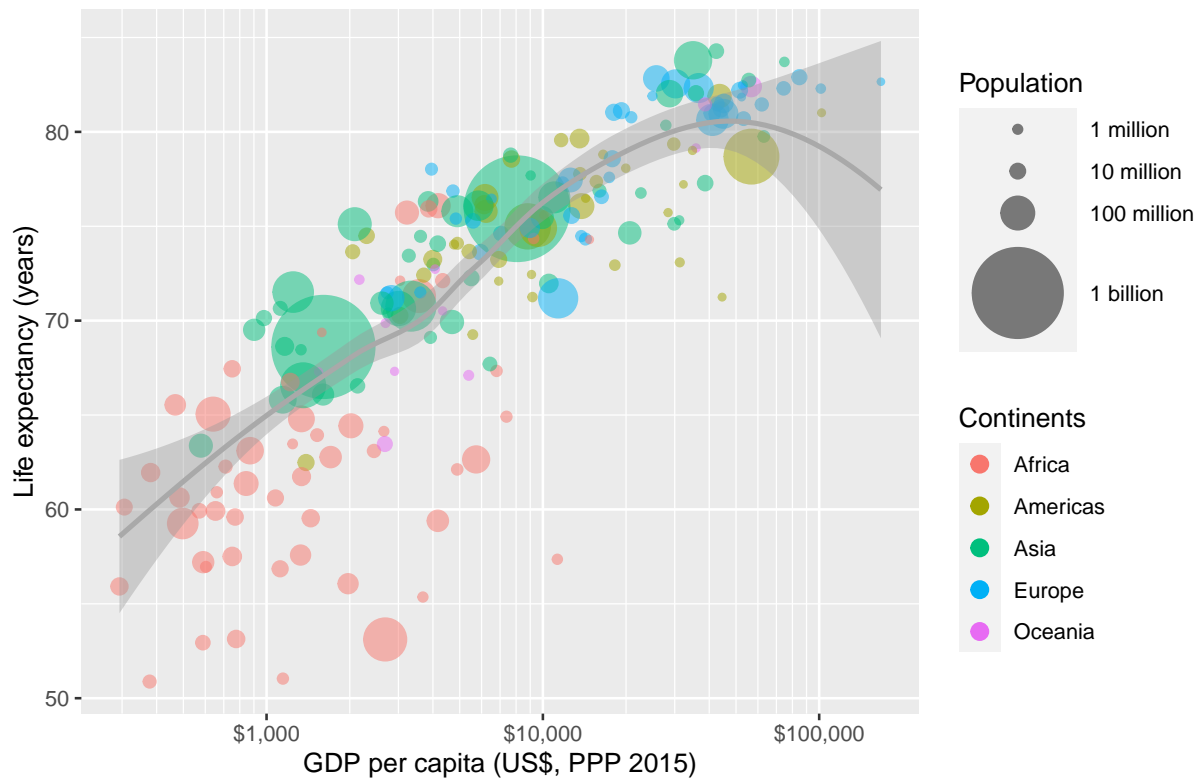




(7) Fit a single LOESS curve to all data points.

```
Q7 <- Q6 +
  geom_smooth(
    method = "loess",
    aes(weight = pop),
    color = "darkgrey"
  )
Q7
```

Wealth and Health by Country



Source : World Bank



(8) Using `ggrepel::geom_text_repel()` to add country names as labels

```
long_lived <- life_quality |>
  slice_max(life_expectancy, n = 5)

short_lived <- life_quality |>
  slice_min(life_expectancy, n = 5)

wealthiest <- life_quality |>
  slice_max(gdp_per_capita, n = 5)

poorest <- life_quality |>
  slice_min(gdp_per_capita, n = 5)

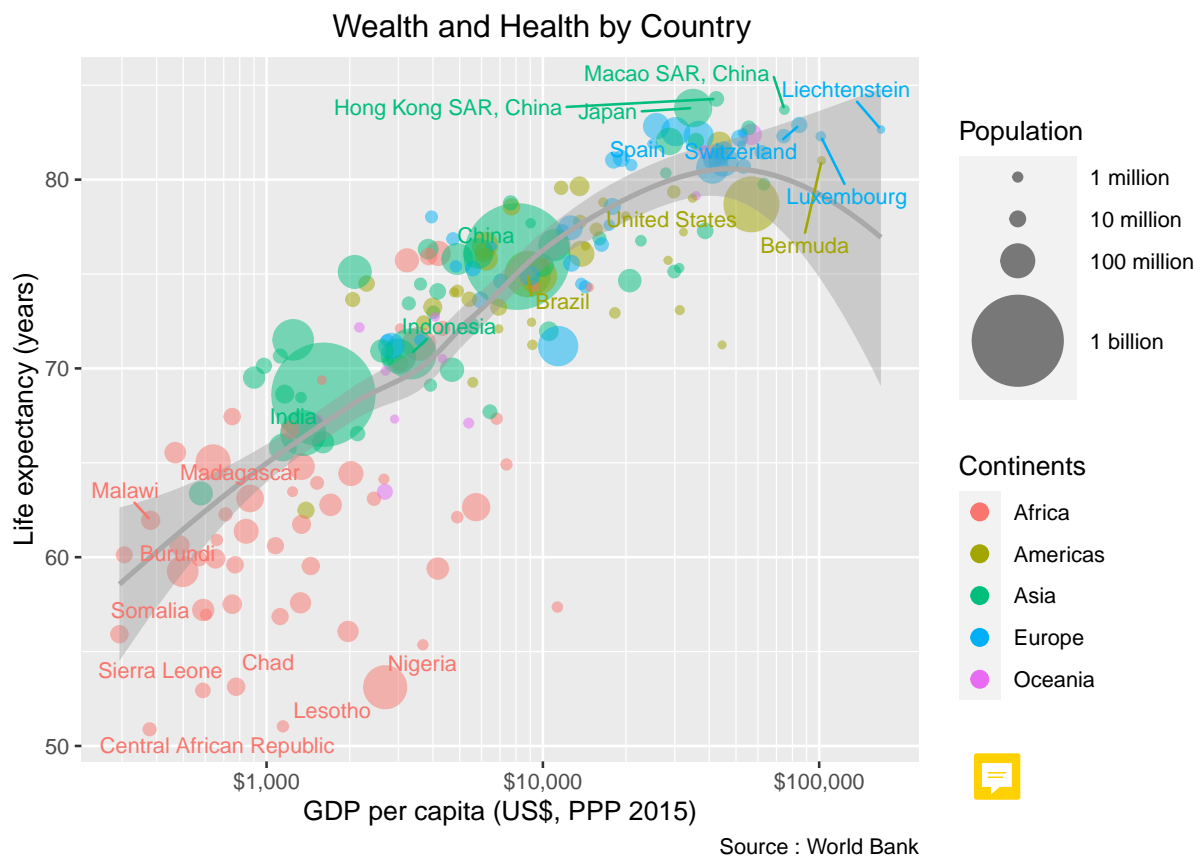
fertile <- life_quality |>
  slice_max(pop, n = 5)

# combine all and remove overlapped countries
life_quality_selected <-
  distinct(
    rbind(
      long_lived,
      short_lived,
      wealthiest,
      poorest,
      fertile
    )
  )
```

```

)
Q8 <- Q7 +
  geom_text_repel(
    data = life_quality_selected,
    aes(
      label = country_name,
      color = continent
    ),
    # set the point size
    size = 3,
    # reduce the overlap between labels
    force = 100
  )
Q8

```



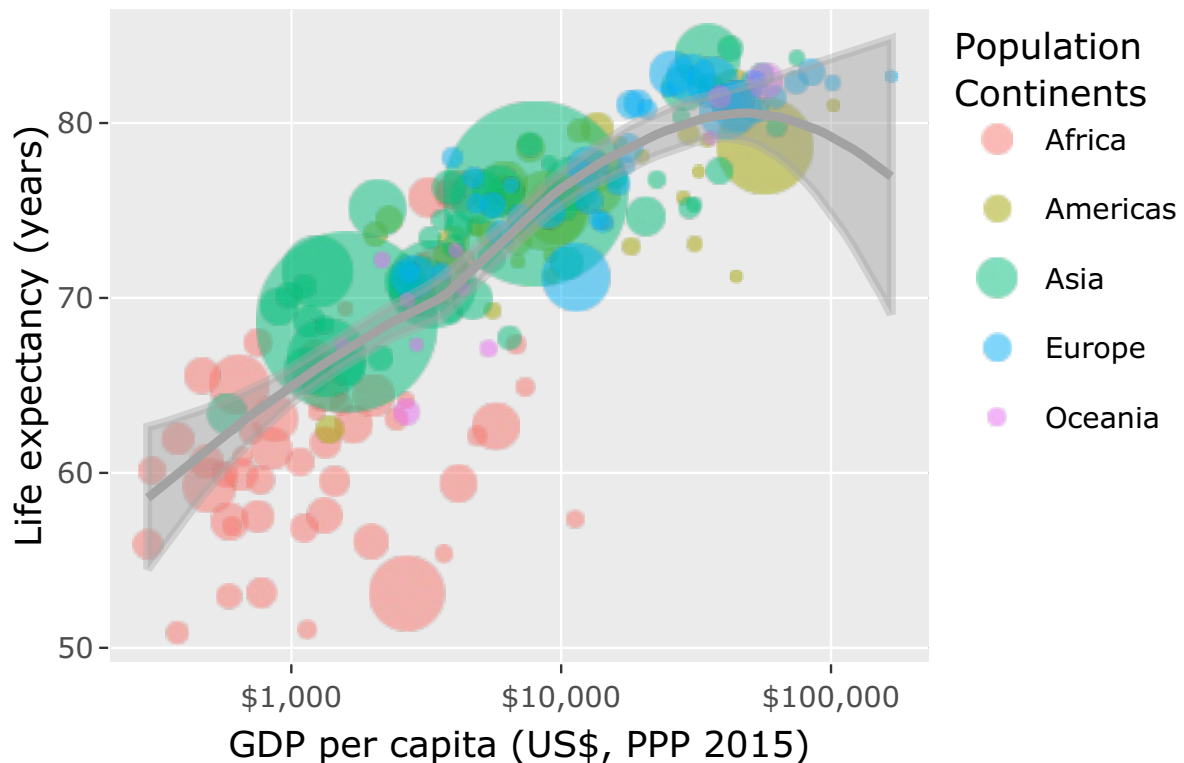
(9) make more adjustments if you think they improve the quality of the plot.

```

Q9 <- ggplotly(Q7)
Q9

```


Wealth and Health by Country



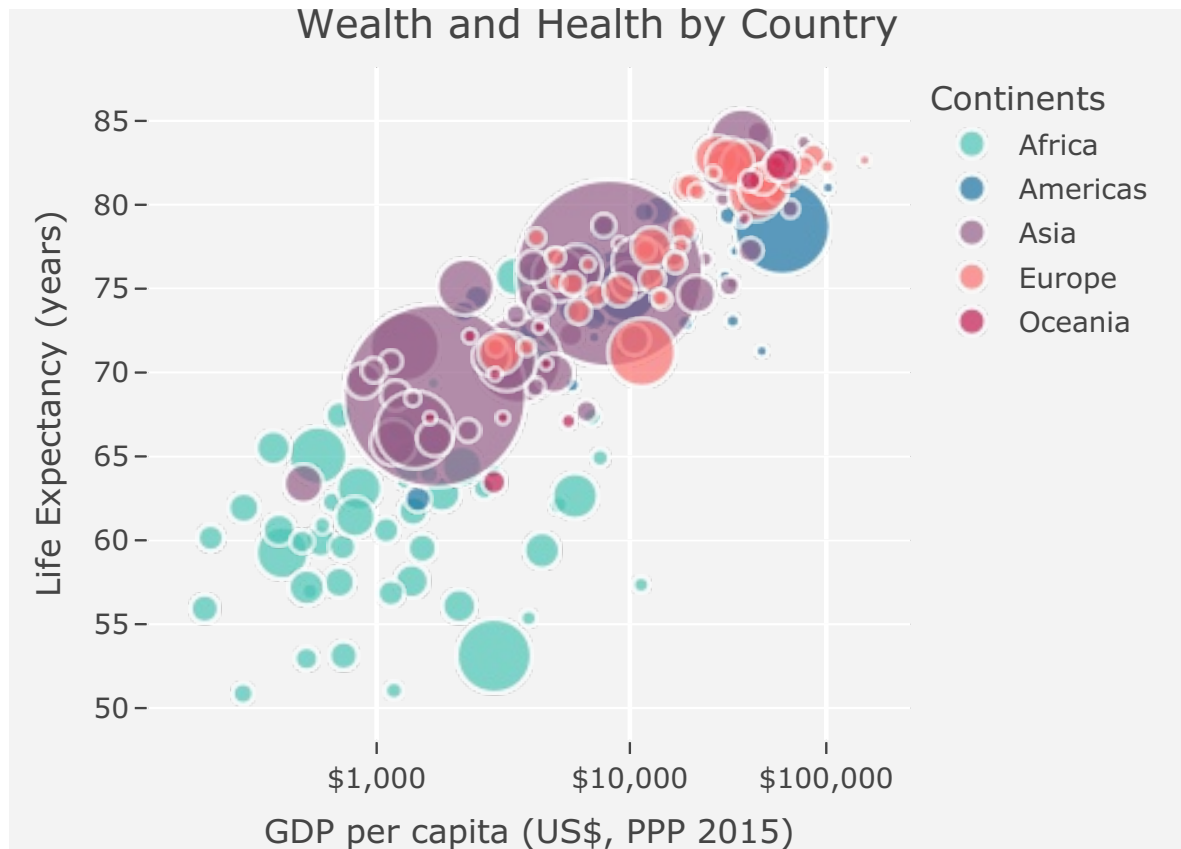
Here, we plotted an interactive plot using the function `ggplotly()`. By hovering the pointer over a country in the graph, an individual will be able to see all the information of a country. However, the labels in the hovertext are untidy and thus we can use the following code in the bottom:

```
data <- life_quality
data$size <- sqrt(data$pop * 1e-03) # assigning a "size" to each country proportional to their population
colors <- c("#4AC6B7", "#1972A4", "#965F8A", "#FF7070", "#C61951") # 5 colours for 5 continents
fig <- plot_ly(data,
  x = ~ log10(gdp_per_capita),
  y = ~ life_expectancy,
  size = ~ size, # cannot use population directly as difference is too big between countries, thus population is scaled
  type = "scatter",
  mode = "markers",
  color = ~ continent, # grouping by continent
  colors = colors, # assigning colours to each continent
  sizes = c(5, 80),
  marker = list(
    symbol = "circle",
    sizemode = "diameter",
    line = list(
      width = 2,
      color = "#FFFFFF"
    )
  ),
  text = ~ paste(
    "Country:", country_name, # labelling the labels in hovertext
    "<br>Life Expectancy:", life_expectancy,
```

```

    "<br>GDP:", gdp_per_capita,
    "<br>Pop.:", pop,
    "<br>Continent:", continent
)
) |>
layout(
  title = "Wealth and Health by Country",
  xaxis = list(
    title = "GDP per capita (US$, PPP 2015)",
    type = "log",
    gridcolor = "rgb(255, 255, 255)",
    zerolinewidth = 1,
    ticklen = 5,
    gridwidth = 2,
    ticktext = list(
      "$1,000",
      "$10,000",
      "$100,000"
    ), # break
    tickvals = list(3, 4, 5)
  ),
  yaxis = list(
    title = "Life Expectancy (years)",
    gridcolor = "rgb(255, 255, 255)",
    zerolinewidth = 1,
    ticklen = 5,
    gridwidth = 2
  ),
  paper_bgcolor = "rgb(243, 243, 243)",
  plot_bgcolor = "rgb(243, 243, 243)",
  legend = list(
    itemsizing = "constant",
    title = list(text = "<b> Continents <b>")
  )
) |>
animation_opts(2000,
  redraw = FALSE
)
fig

```



The code is written with reference to the website Plotly ¹. While the size of the points representing the countries are smaller, they are more proportional to their population. To resolve the issue of the points being too small, this plot enables users to zoom into a specific area of the plot that they want to investigate.

(10) Write a few sentences about the data. What does the plot reveal about the data?

The plot shows that there is an overall trend where as the Gross Domestic Product (GDP) per capita increases, the life expectancy of the country increases. From the graph, we also see that the countries part of the Africa continent tend to have lower life expectancy and lower GDP per capita while countries from the Europe continent tend to have higher life expectancy and higher GDP per capita.

¹<https://plotly.com/python/bubble-charts/>. Accessed 21st February 2022.