Lab 2: Make a plot

In this lab, you will get some practice using `ggplot`'s syntax for mapping variables in our data to aesthetic elements in a scatter plot.

Load libraries

```
library(tidyverse)
library(gapminder)
```

Examine the Gapminder Data

Print out the gapminder data below to remind yourself of the data format and variable names (remember that you can also pull up the help file by typing `? gapminder` in the console below).

```
gapminder
```

```
## # A tibble: 1,704 x 6
##    country     continent  year lifeExp      pop gdpPercap
##    <fct>       <fct>     <int>   <dbl>    <int>     <dbl>
##  1 Afghanistan Asia       1952    28.8  8425333      779.
##  2 Afghanistan Asia       1957    30.3  9240934      821.
##  3 Afghanistan Asia       1962    32.0 10267083      853.
##  4 Afghanistan Asia       1967    34.0 11537966      836.
##  5 Afghanistan Asia       1972    36.1 13079460      740.
##  6 Afghanistan Asia       1977    38.4 14880372      786.
##  7 Afghanistan Asia       1982    39.9 12881816      978.
##  8 Afghanistan Asia       1987    40.8 13867957      852.
##  9 Afghanistan Asia       1992    41.7 16317921      649.
## 10 Afghanistan Asia       1997    41.8 22227415      635.
## # i 1,694 more rows
```

Write code to create each of the following plots described below. You will need to make a new code chunk for each plot.

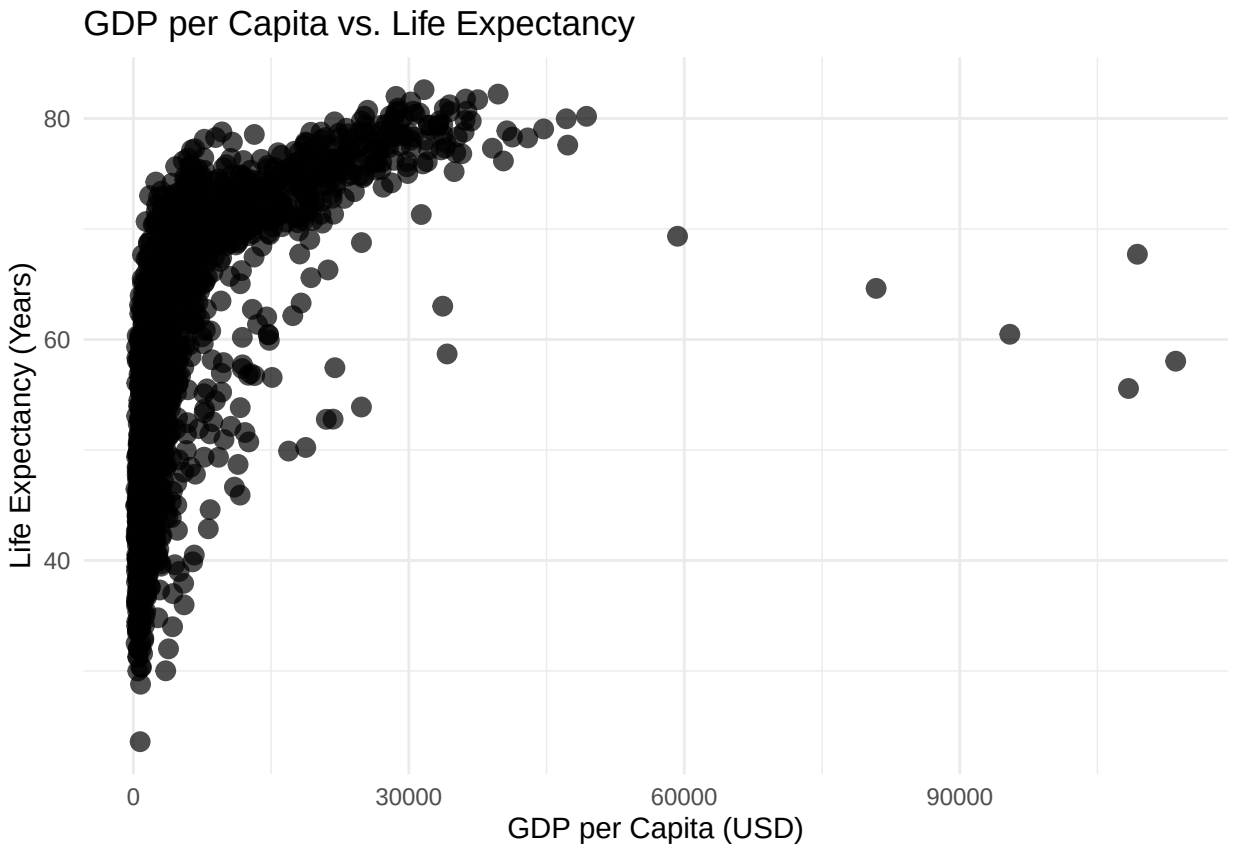
> Remember to work in small chunks and test your code (and whether your .Rmd will knit as you expect) periodically and often!

1. GDP and Life Expectancy

Create a scatter plot (using `geom_point()`) of GDP per Capita (on the x-axis) and Life Expectancy (on the y-axis).

```
ggplot(gapminder, aes(x = gdpPercap, y = lifeExp)) +
  geom_point(size = 3, alpha = 0.7) +
  labs(
    title = "GDP per Capita vs. Life Expectancy",
    x = "GDP per Capita (USD)",
    y = "Life Expectancy (Years)"
```

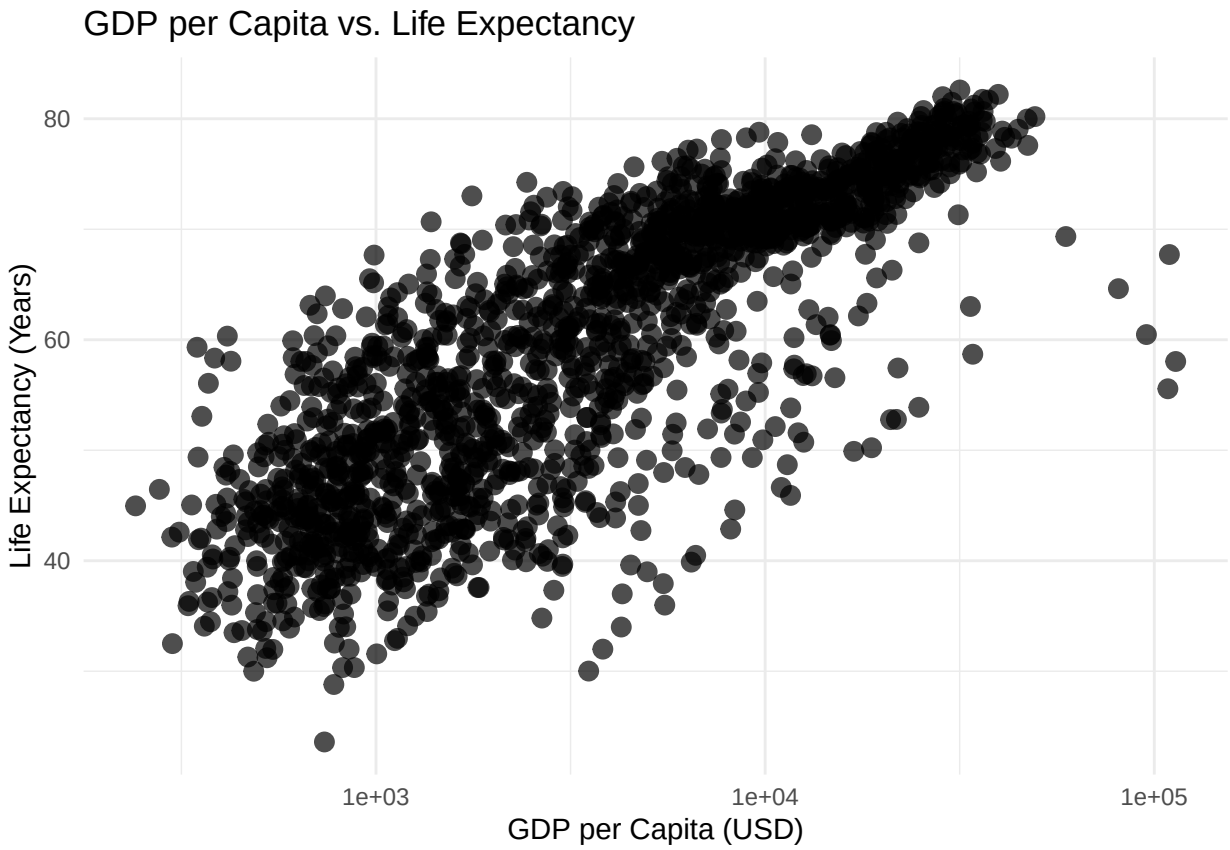
```
  ) +
  theme_minimal()
```



2. Log scale (scatterplot)

Adjust the plot so that the x-axis scale is in log units rather than raw GDP, and properly label the axes.

```
ggplot(gapminder, aes(x = gdpPercap, y = lifeExp)) +
  geom_point(size = 3, alpha = 0.7) +
  labs(
    title = "GDP per Capita vs. Life Expectancy",
    x = "GDP per Capita (USD)",
    y = "Life Expectancy (Years)"
  ) +
  theme_minimal() +
  scale_x_log10()
```



3. Narrow in on the Oceanic countries

Using the `filter()` function, narrow the scope of your scatter plot to only include countries in the Oceania continent. *Hint:* Use `==` and not `=` when asking `filter` to pick a particular value of `continent`.

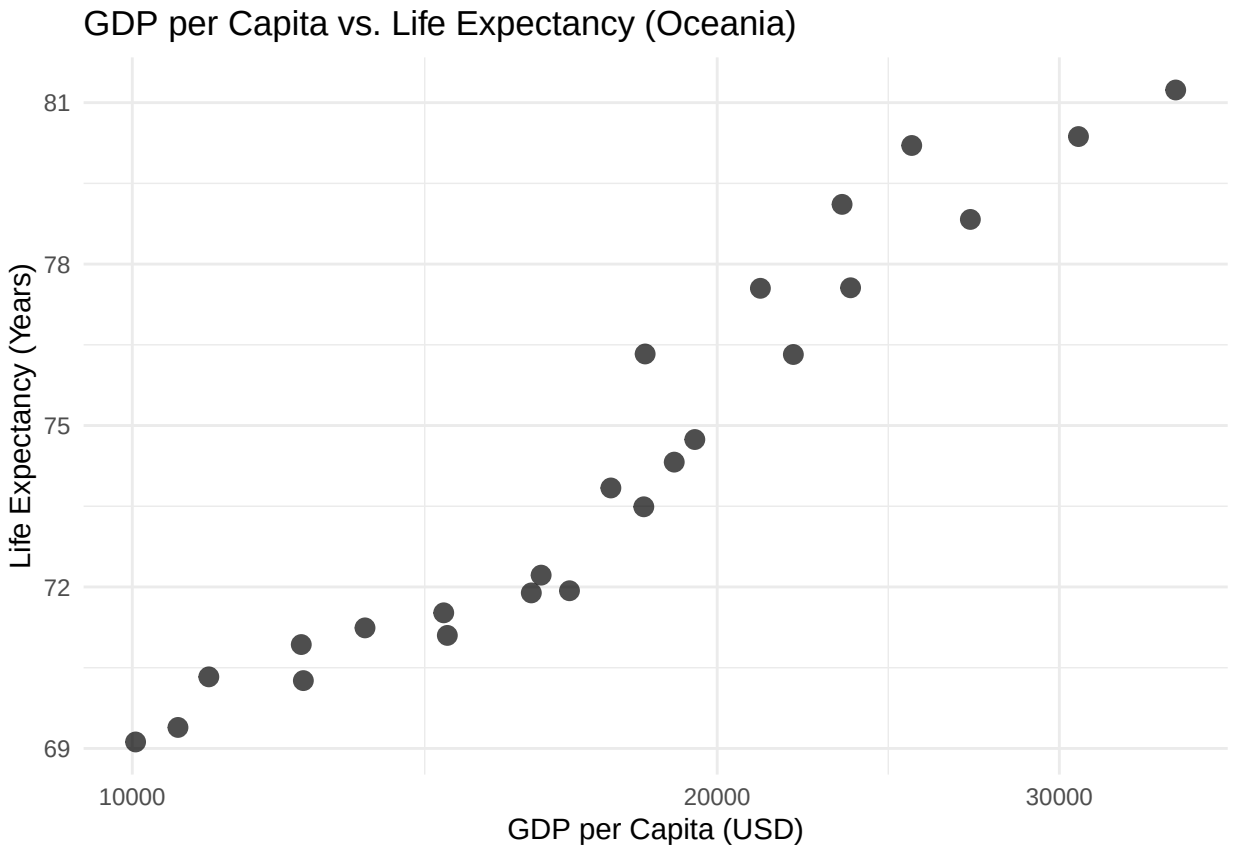
Here's some example code using the `filter` function:

```
filter( gapminder, country == "United States")
```

or equivalently (using a pipe operator from the `dplyr` package),

```
gapminder %>%
  filter( country == "United States")
```

```
ggplot(filter(gapminder, continent == "Oceania")
, aes(x = gdpPercap, y = lifeExp)) +
  geom_point(size = 3, alpha = 0.7) +
  labs(
    title = "GDP per Capita vs. Life Expectancy (Oceania)",
    x = "GDP per Capita (USD)",
    y = "Life Expectancy (Years)"
  ) +
  theme_minimal() +
  scale_x_log10()
```

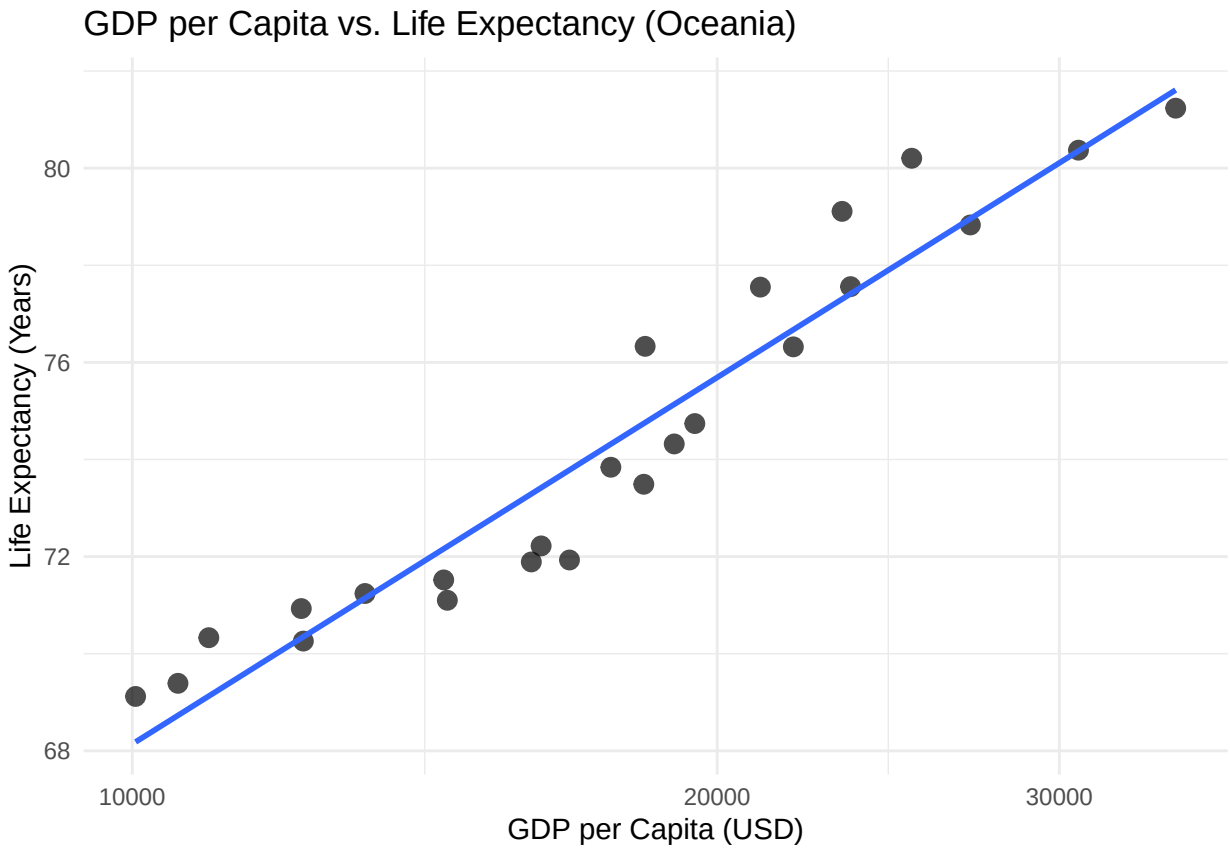


4. Overall trend

Add an overall *linear* trend line to your plot.

```
ggplot(filter(gapminder, continent == "Oceania")
, aes(x = gdpPercap, y = lifeExp)) +
  geom_point(size = 3, alpha = 0.7) +
  labs(
    title = "GDP per Capita vs. Life Expectancy (Oceania)",
    x = "GDP per Capita (USD)",
    y = "Life Expectancy (Years)"
  ) +
  theme_minimal() +
  scale_x_log10() +
  geom_smooth(method = "lm",se=FALSE)
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

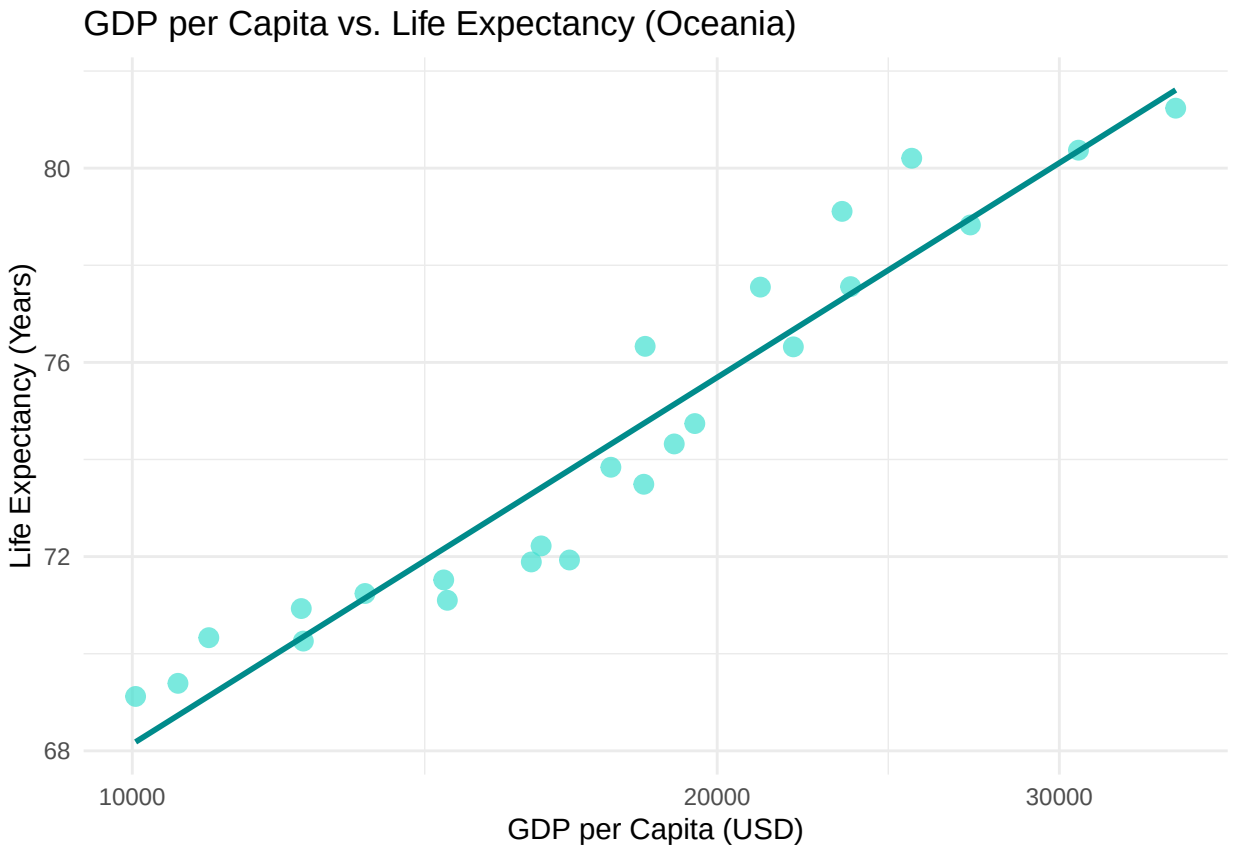


5. Oceanic colors

Change the color of points and trend line in your plot to colors of your choosing (don't worry about selecting these colors too carefully for this question).

```
ggplot(filter(gapminder, continent == "Oceania")
, aes(x = gdpPercap, y = lifeExp)) +
  geom_point(size = 3, alpha = 0.7, color = "turquoise") +
  labs(
    title = "GDP per Capita vs. Life Expectancy (Oceania)",
    x = "GDP per Capita (USD)",
    y = "Life Expectancy (Years)"
  ) +
  theme_minimal() +
  scale_x_log10() +
  geom_smooth(method = "lm",se=FALSE,color = "darkcyan")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

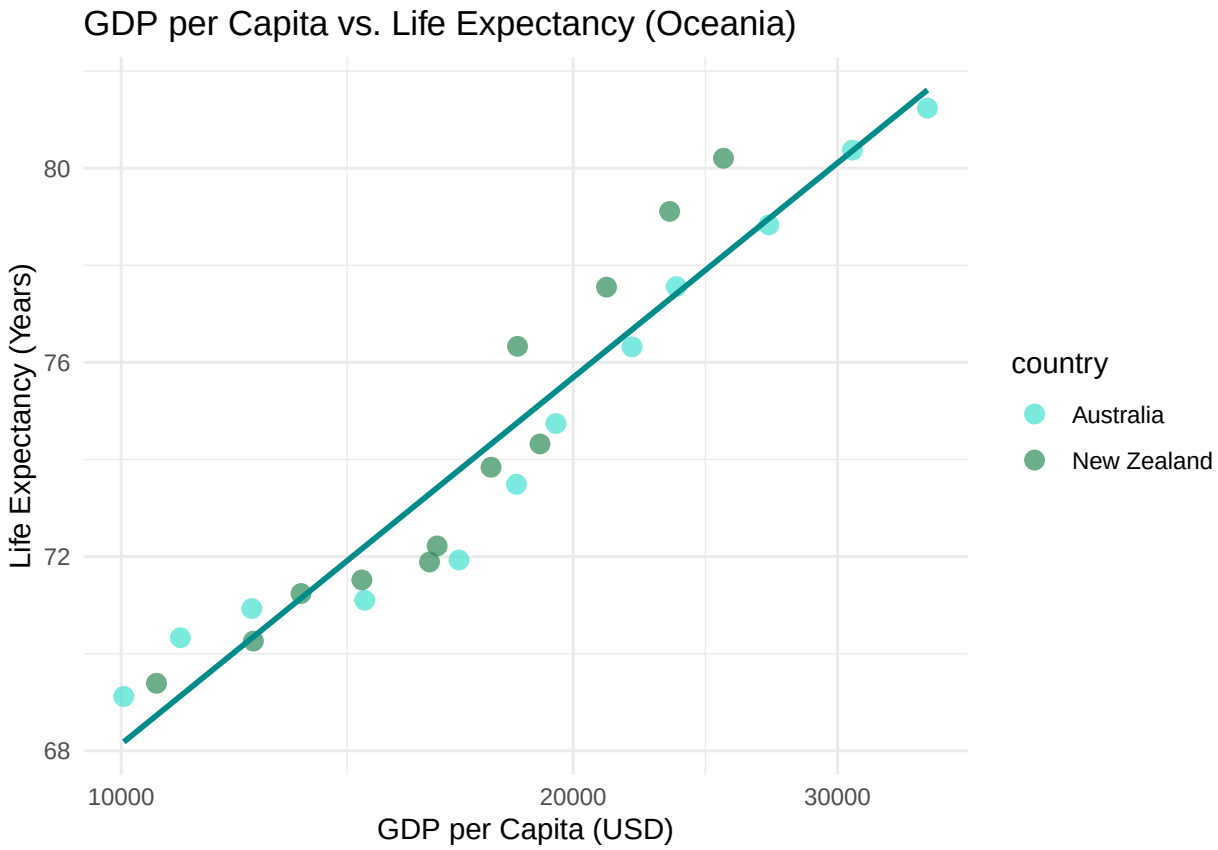


6. Colors by country

Indicate different countries in your plot by different colors but keep a single overall trend line.

```
ggplot(filter(gapminder, continent == "Oceania")
, aes(x = gdpPercap, y = lifeExp, color = country)) +
  geom_point(size = 3, alpha = 0.7) +
  labs(
    title = "GDP per Capita vs. Life Expectancy (Oceania)",
    x = "GDP per Capita (USD)",
    y = "Life Expectancy (Years)"
  ) +
  scale_color_manual(
    values = c("Australia" = "turquoise", "New Zealand" = "seagreen")
  ) +
  theme_minimal() +
  scale_x_log10() +
  geom_smooth(method = "lm",se=FALSE,color = "darkcyan")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



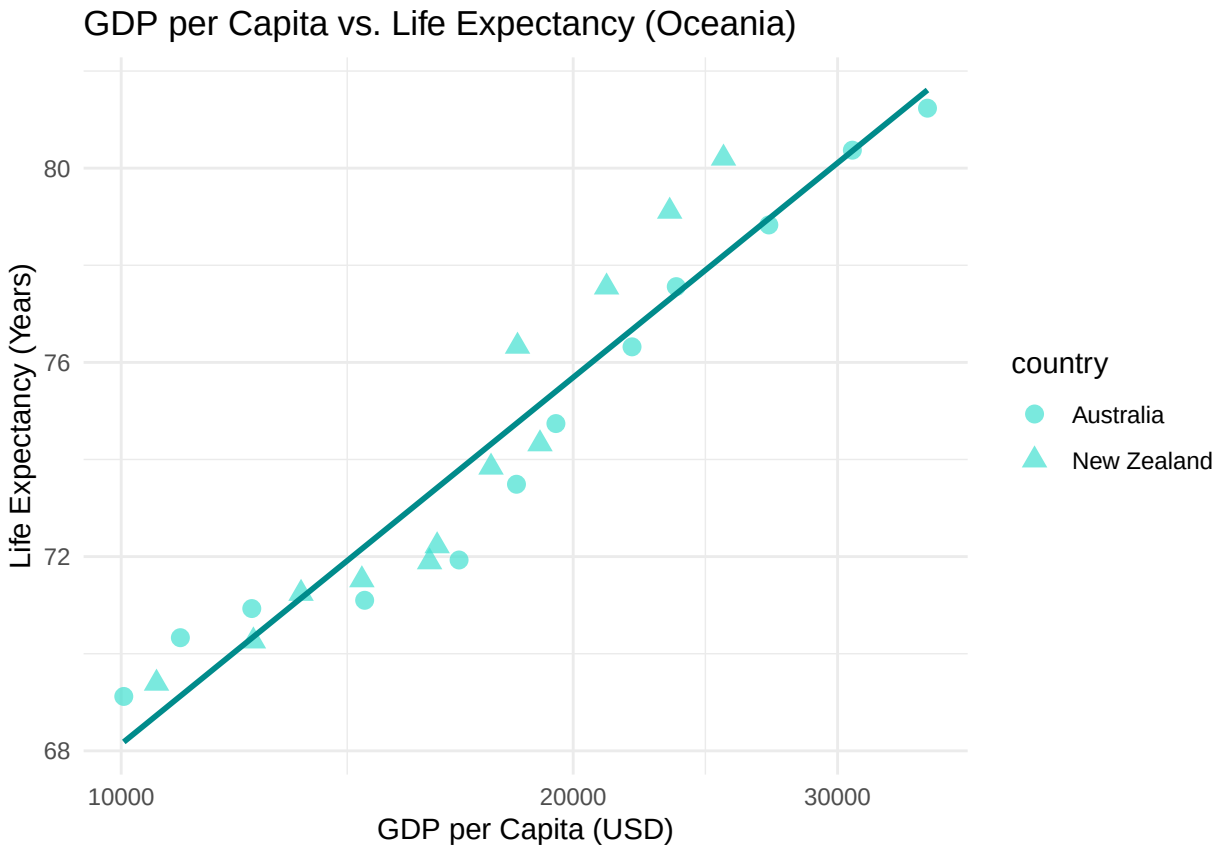
7. Country shapes

Indicate different countries in your plot by different shapes (all points should be the same color).

```
ggplot(filter(gapminder, continent == "Oceania")
, aes(x = gdpPercap, y = lifeExp, shape = country)) +
  geom_point(size = 3, alpha = 0.7, color = "turquoise") +
  labs(
    title = "GDP per Capita vs. Life Expectancy (Oceania)",
    x = "GDP per Capita (USD)",
    y = "Life Expectancy (Years)"
  ) +
  theme_minimal() +
  scale_x_log10() +
  geom_smooth(aes(group = 1),method = "lm",se=FALSE,color = "darkcyan")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

```
## Warning: The following aesthetics were dropped during statistical transformation: shape.
## i This can happen when ggplot fails to infer the correct grouping structure in
##   the data.
## i Did you forget to specify a `group` aesthetic or to convert a numerical
##   variable into a factor?
```



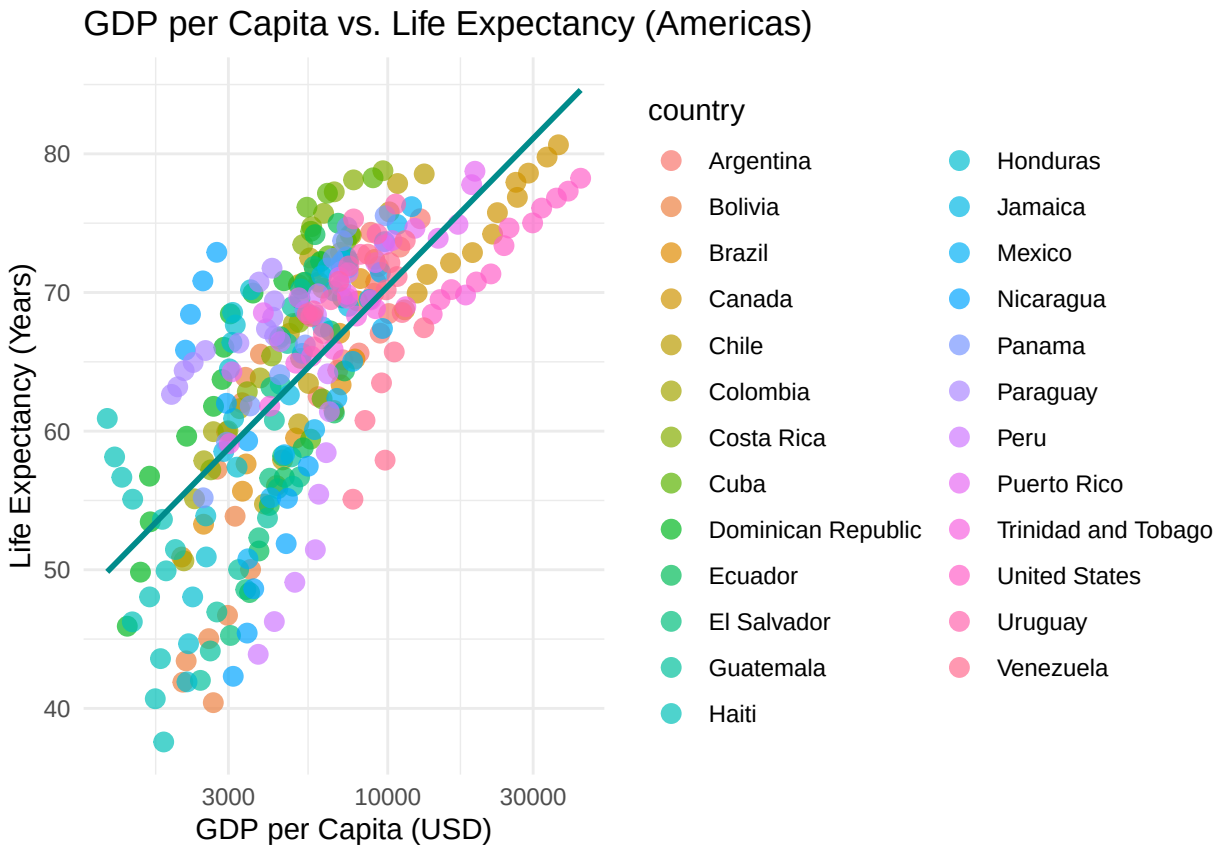
8. In the Americas

Recreate your scatter plot from Part 6 for countries in the Americas continent.

How many colors does ggplot include in its legend? Is it easy to tell the difference between the countries in your plot? A ton of colors, and its not easy to differentiate.

```
ggplot(filter(gapminder, continent == "Americas")
, aes(x = gdpPercap, y = lifeExp, color = country)) +
  geom_point(size = 3, alpha = 0.7) +
  labs(
    title = "GDP per Capita vs. Life Expectancy (Americas)",
    x = "GDP per Capita (USD)",
    y = "Life Expectancy (Years)"
  ) +
  theme_minimal() +
  scale_x_log10() +
  geom_smooth(method = "lm",se=FALSE,color = "darkcyan")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



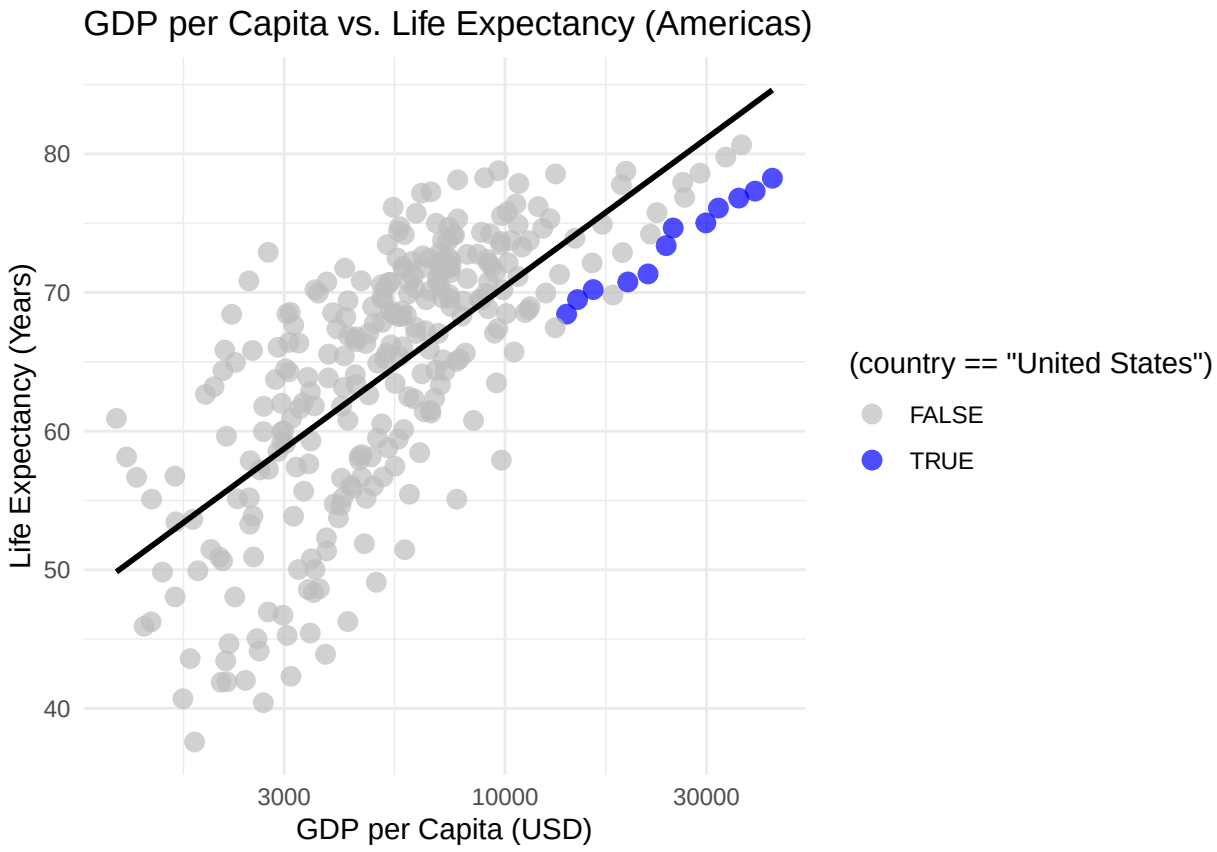
9. US vs. the Americas

Let's simplify this plot by singling out the US. Use the following code, which singles out Europe in a scatterplot of life expectancy by population, as an example.

```
ggplot( gapminder,
        aes( pop,
             lifeExp )) +
  geom_point( aes(color = (continent == "Europe") )) +
  scale_x_log10()
```

```
ggplot(filter(gapminder, continent == "Americas")
, aes(x = gdpPercap, y = lifeExp)) +
  geom_point(size = 3, alpha = 0.7, aes(color = (country == "United States") )) +
  labs(
    title = "GDP per Capita vs. Life Expectancy (Americas)",
    x = "GDP per Capita (USD)",
    y = "Life Expectancy (Years)"
  ) +
  scale_color_manual(
    values = c("TRUE" = "blue", "FALSE" = "grey")
  ) +
  theme_minimal() +
  scale_x_log10() +
  geom_smooth(method = "lm",se=FALSE,color = "black")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

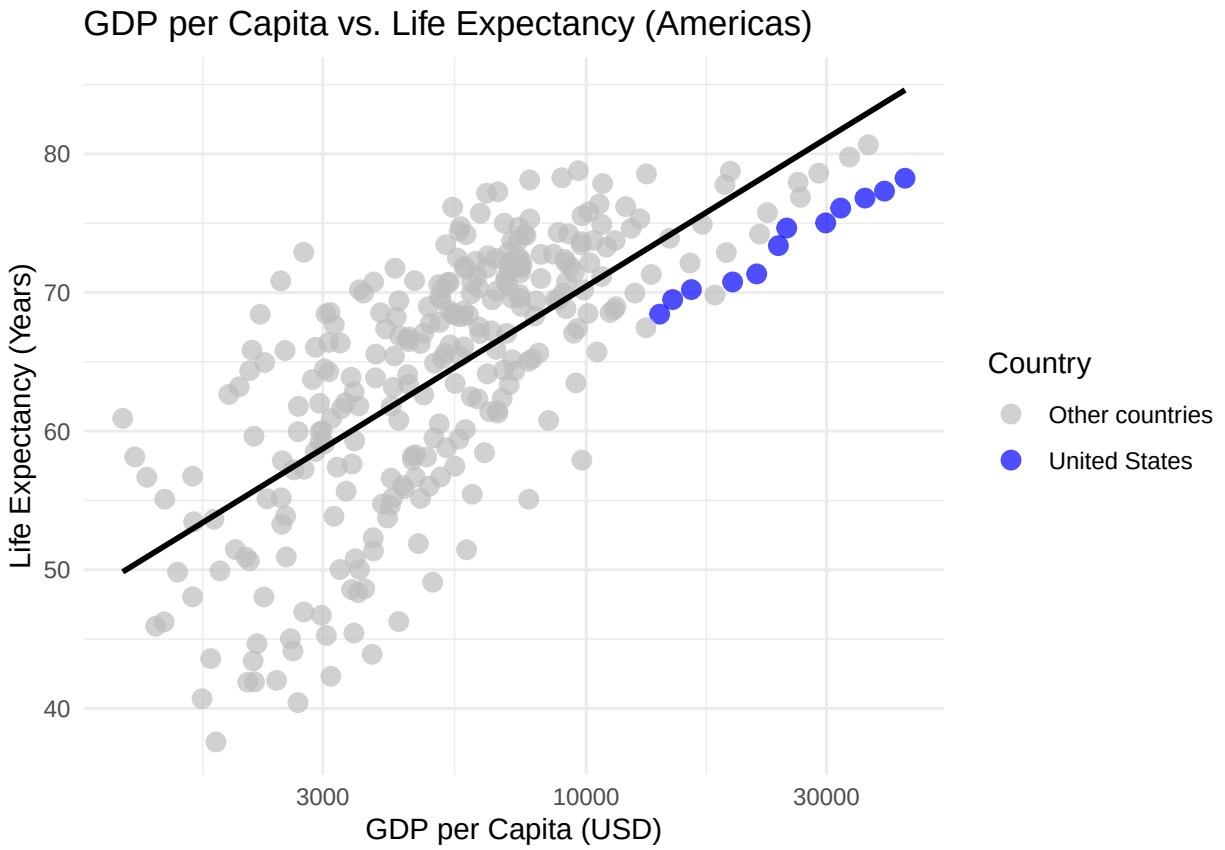


10. Clean up

Tidy up this plot as best you can: e.g., give it nice x- and y- labels, as well as a title and caption.

```
ggplot(filter(gapminder, continent == "Americas")
, aes(x = gdpPercap, y = lifeExp)) +
  geom_point(size = 3, alpha = 0.7, aes(color = (country == "United States") )) +
  labs(
    title = "GDP per Capita vs. Life Expectancy (Americas)",
    x = "GDP per Capita (USD)",
    y = "Life Expectancy (Years)",
    color = "Country"
  ) +
  scale_color_manual(
    values = c("TRUE" = "blue", "FALSE" = "grey"),
    labels = c("TRUE" = "United States", "FALSE" = "Other countries")
  ) +
  theme_minimal() +
  scale_x_log10() +
  geom_smooth(method = "lm",se=FALSE,color = "black")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



Are there any elements of the plot you would like to change but don't know how? What are they? nope

Submission Instructions

Knit the completed R Markdown file as a PDF document (use the "Knit" button at the top of the script editor window) and upload it to the appropriate assignment on Canvas.