

EDLD652_Lab_5

Michelle Cui

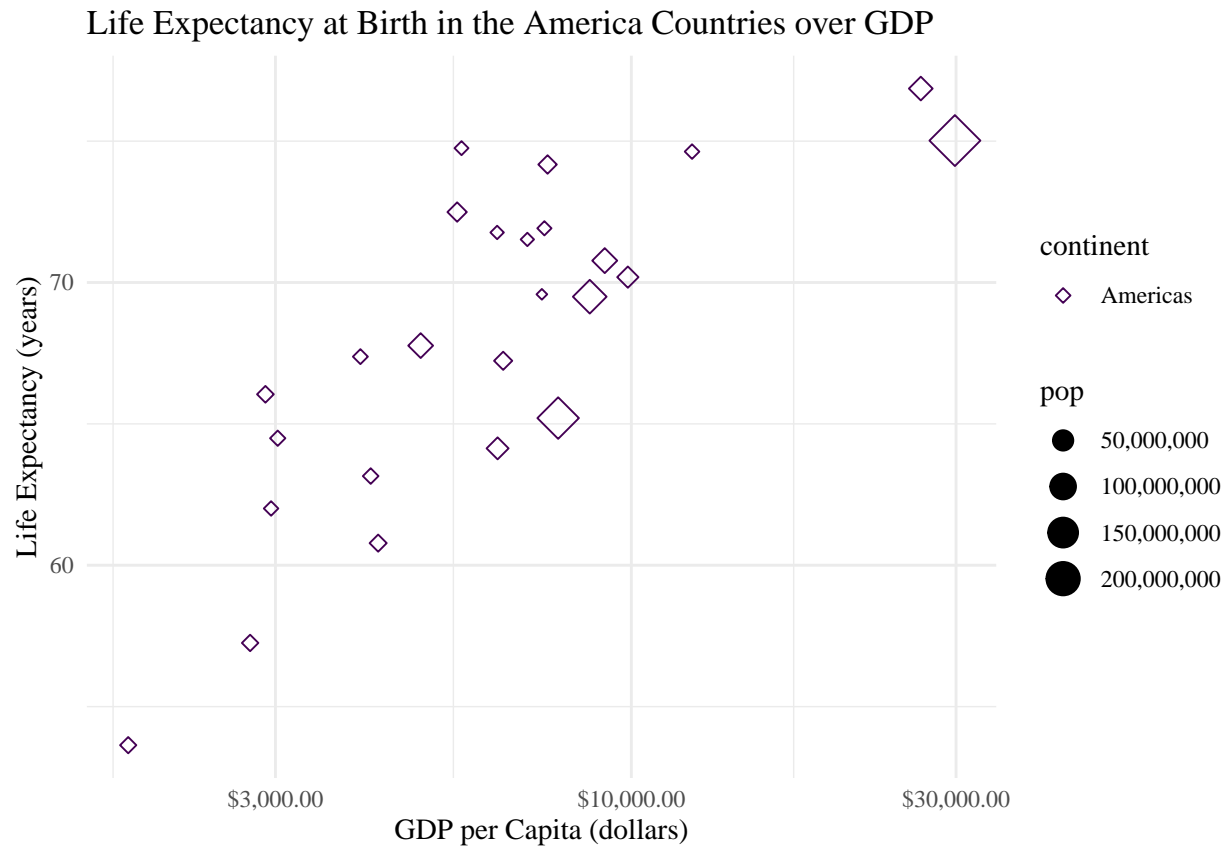
2025-02-05

```
gap <- read_csv(here("data/gapminder.csv"))
```

Lab Tasks

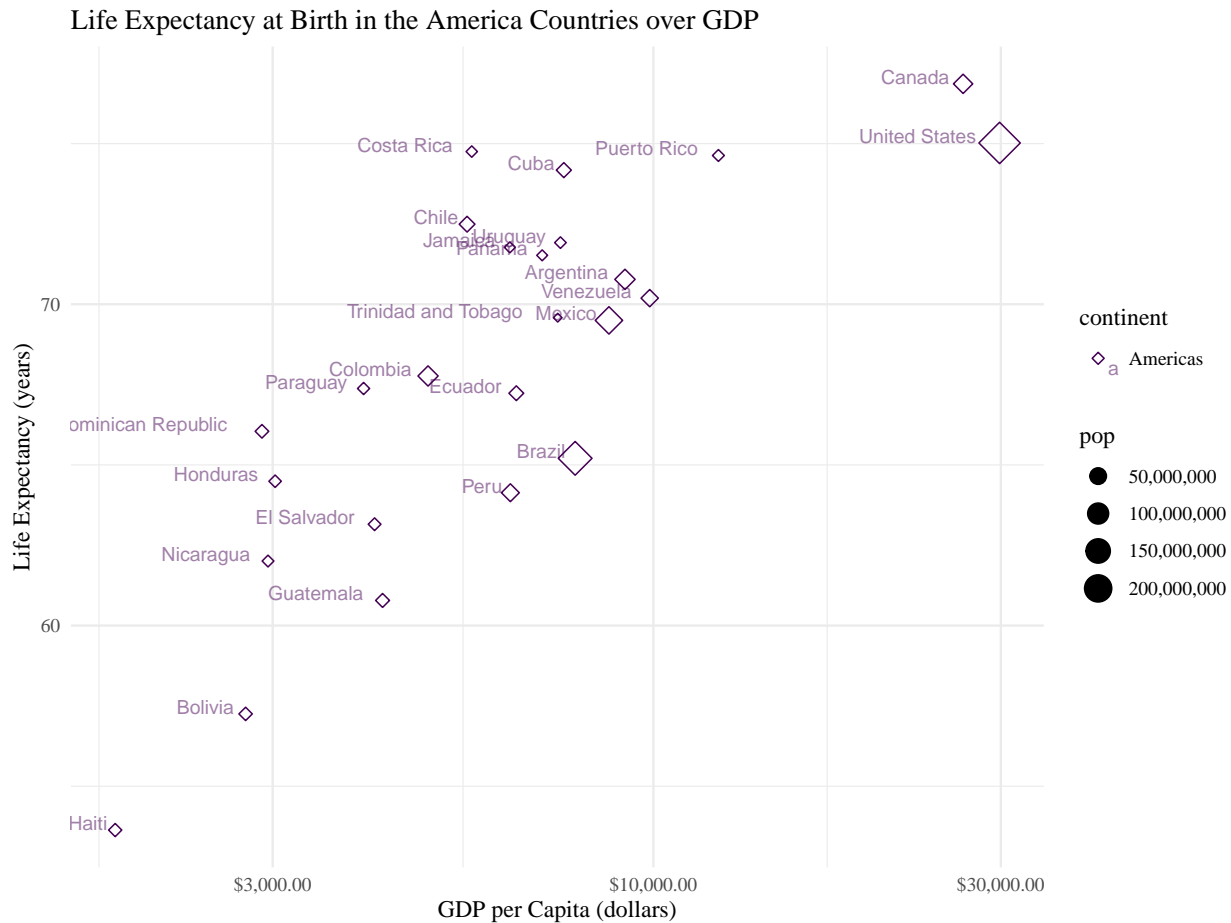
1. Recreate the final plot you generated in Lab PS-2 with just the countries in the American Continent

```
gap_1987_am <- gap %>%  
  filter(year == "1987", continent == "Americas")  
  
gap_1987_am %>%  
  ggplot(aes(x = gdpPercap, y = lifeExp, color = continent, shape = continent, size = pop)) +  
  geom_point() +  
  scale_color_viridis_d() +  
  scale_shape_manual(values = c(5)) +  
  scale_size_continuous(labels = label_comma()) +  
  scale_x_log10(labels = label_dollar()) +  
  labs(title = "Life Expectancy at Birth in the America Countries over GDP",  
       x = "GDP per Capita (dollars)",  
       y = "Life Expectancy (years)")
```



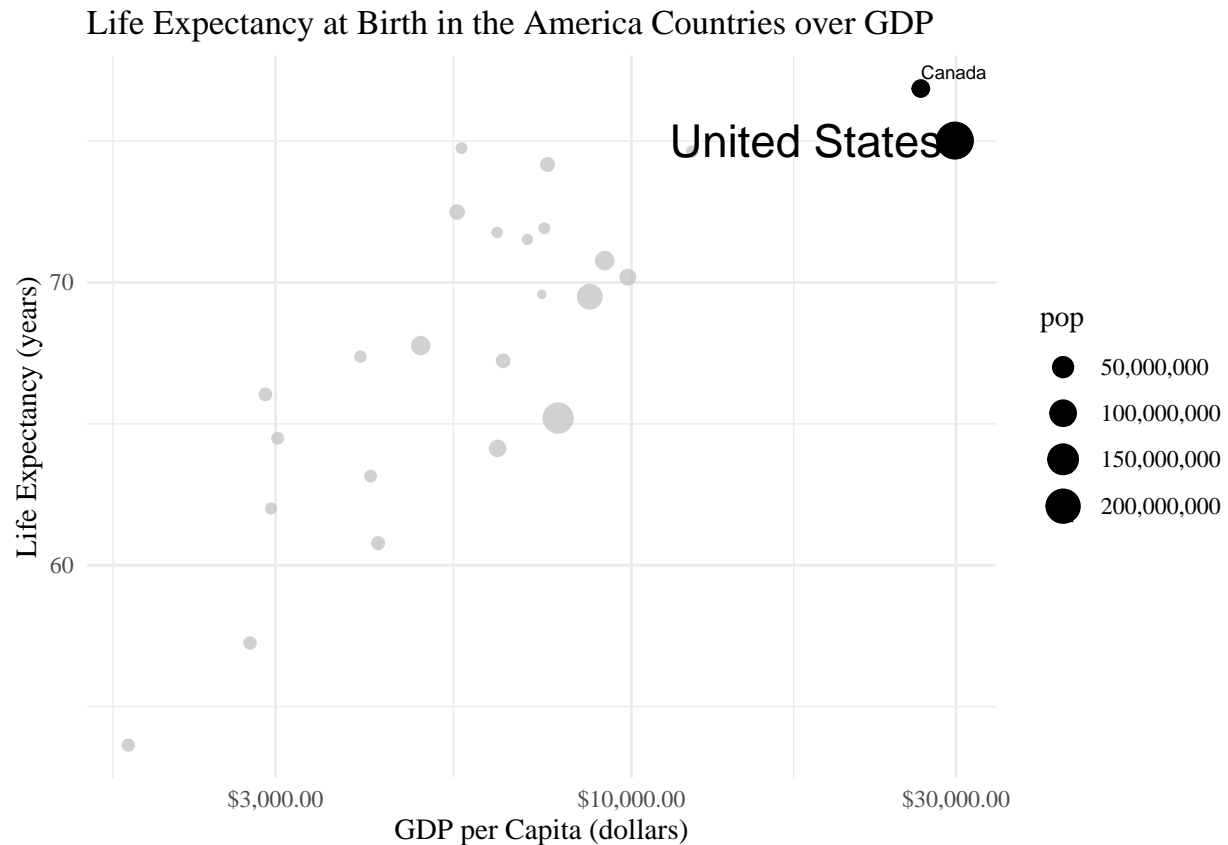
2. Add labels to the countries as there are fewer categories to compare

```
gap_1987_am %>%
  ggplot(aes(x = gdpPercap, y = lifeExp, color = continent, shape = continent, size = pop)) +
  geom_point() +
  scale_color_viridis_d() +
  scale_shape_manual(values = c(5)) +
  scale_size_continuous(labels = label_comma()) +
  scale_x_log10(labels = label_dollar()) +
  geom_text(aes(label = country),
            size = 3, alpha = 0.5, hjust = 1.2, vjust = 0) + labs(title = "Life Expectancy at Birth in -",
  x = "GDP per Capita (dollars)",
  y = "Life Expectancy (years)")
```



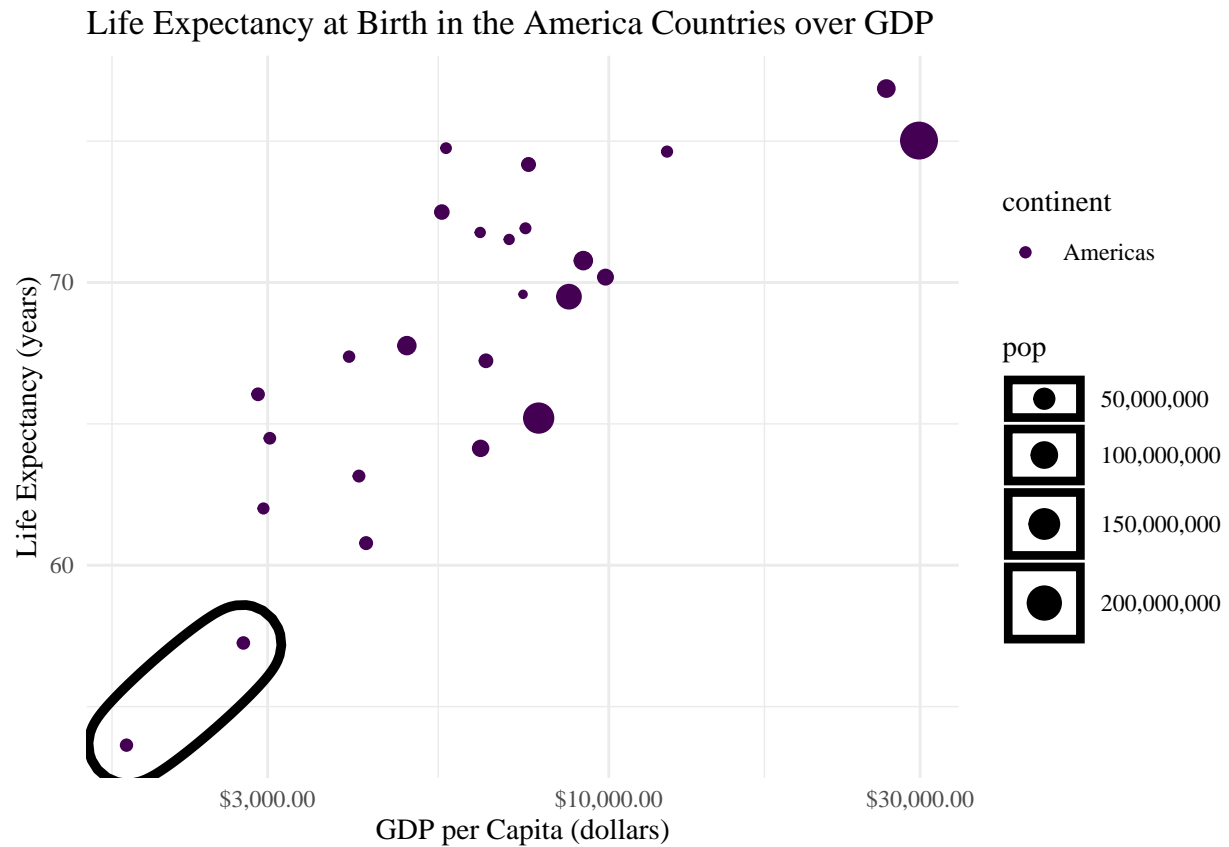
3. Highlight (hint: use `gghighlight`) only those countries that have life expectancy greater than 75.

```
gap_1987_am %>%
  ggplot(aes(x = gdpPercap, y = lifeExp, size = pop)) +
  geom_point() +
  scale_color_viridis_d() +
  scale_shape_manual(values = c(2)) +
  scale_size_continuous(labels = label_comma()) +
  scale_x_log10(labels = label_dollar()) +
  gghighlight(lifeExp > 75) +
  geom_text_repel(data = filter(gap_1987_am, lifeExp > 75),
    aes(label = country)) +
  labs(title = "Life Expectancy at Birth in the America Countries over GDP",
    x = "GDP per Capita (dollars)",
    y = "Life Expectancy (years)")
```



4. Try a marked up ellipse(hint: use ggforce library) and `geom_mark_ellipse` to highlight the two outliers on the SW grid of the plot instead of `gghighlight`

```
gap_1987_am %>%
  ggplot(aes(x = gdpPercap, y = lifeExp, size = pop)) +
  geom_point(aes(color = continent)) +
  scale_color_viridis_d() +
  scale_shape_manual(values = c(2)) +
  scale_size_continuous(labels = label_comma()) +
  scale_x_log10(labels = label_dollar()) +
  geom_mark_ellipse(
    data = filter(gap_1987_am, lifeExp < 60 & gdpPercap < 3000),
    aes(x = gdpPercap, y = lifeExp) +
  labs(title = "Life Expectancy at Birth in the America Countries over GDP",
       x = "GDP per Capita (dollars)",
       y = "Life Expectancy (years)")
```



Which one do you prefer between #3 and #4?

Answer: I like plot #3 more than #4 because the visualization on #3 is much clear and neat. It presents the outliers with a simple and direct approach. It might also because the outliers are not large in the dataset.