

# Hiển thị dữ liệu bằng ggplot2

**Tuan V. Nguyen**

Garvan Institute of Medical Research

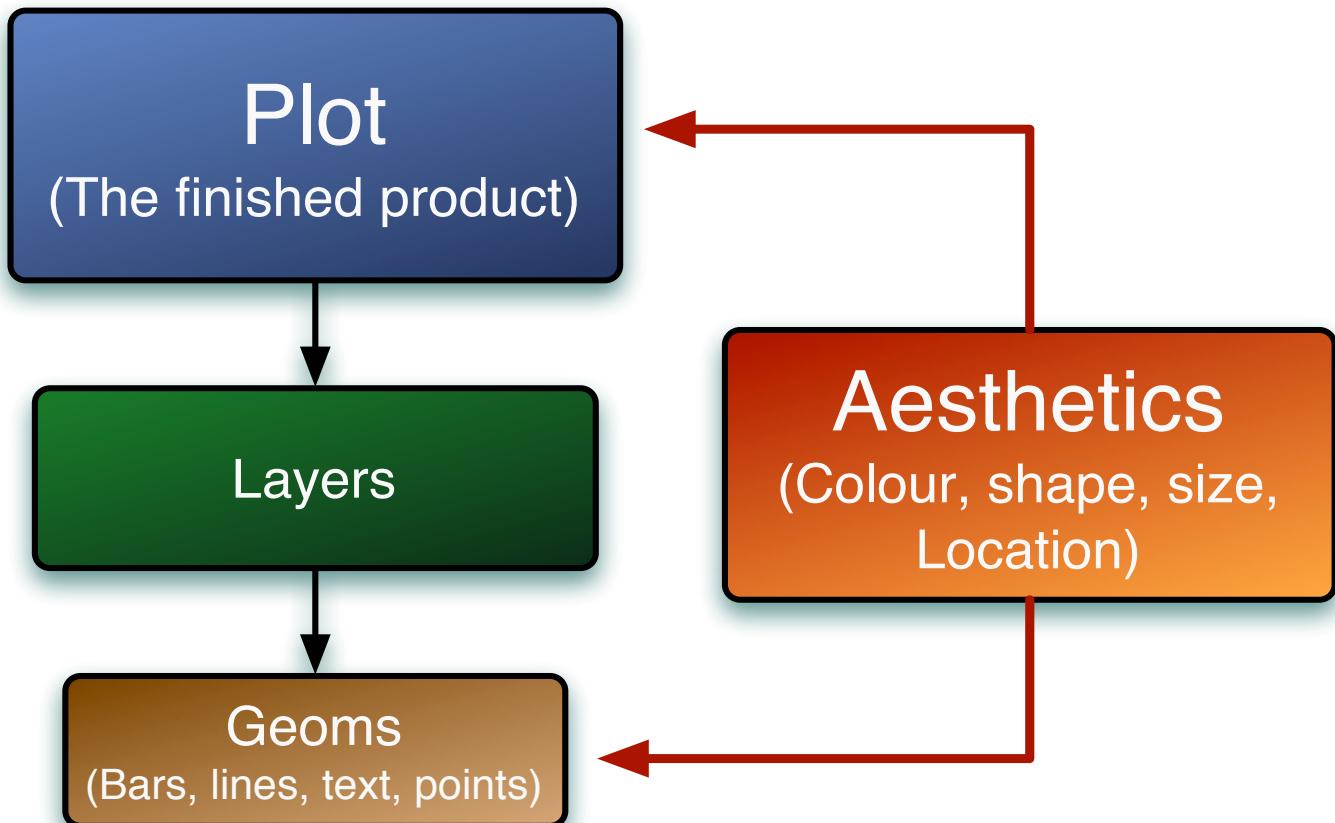
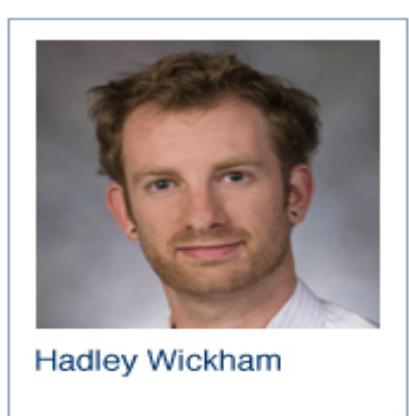
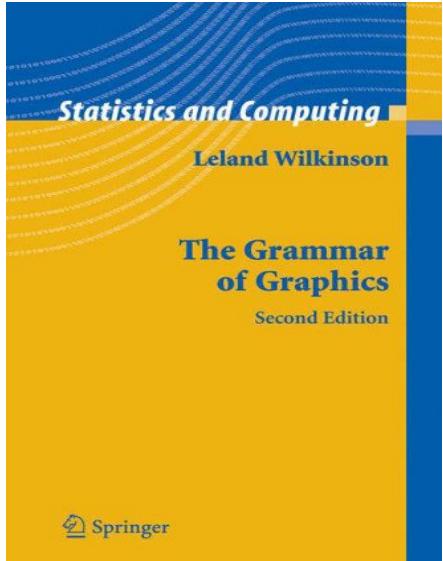
University of New South Wales (UNSW Sydney), Australia

University of Technology, Sydney (UTS), Australia

Ton Duc Thang University, Vietnam



# Mô hình *ggplot2*



# Mô hình nhiều lớp (layers)

- **Lớp 1** định nghĩa biến số để thể hiện: aes(x, y, ...)
- **Lớp 2** hình thức thể hiện **geom\_xxx()**
- **Lớp 3** trang trí (màu, kích thước, phông chữ): **theme\_xxx()**
- v.v.

# Dữ liệu thực tế: gapminder

- Dữ liệu gapminder (Hans Rosling) có trong package ‘gapminder’
- Biến số: country (quốc gia), continent (lục địa), year (năm), lifeExp (tuổi thọ), pop (dân số), gdpPercap (thu nhập bình quân) 1952 - 2007

```
> library(gapminder)
> data(gapminder)
> head(gapminder)
# A tibble: 6 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.
```

# Mục tiêu hiển thị dữ liệu gapminder

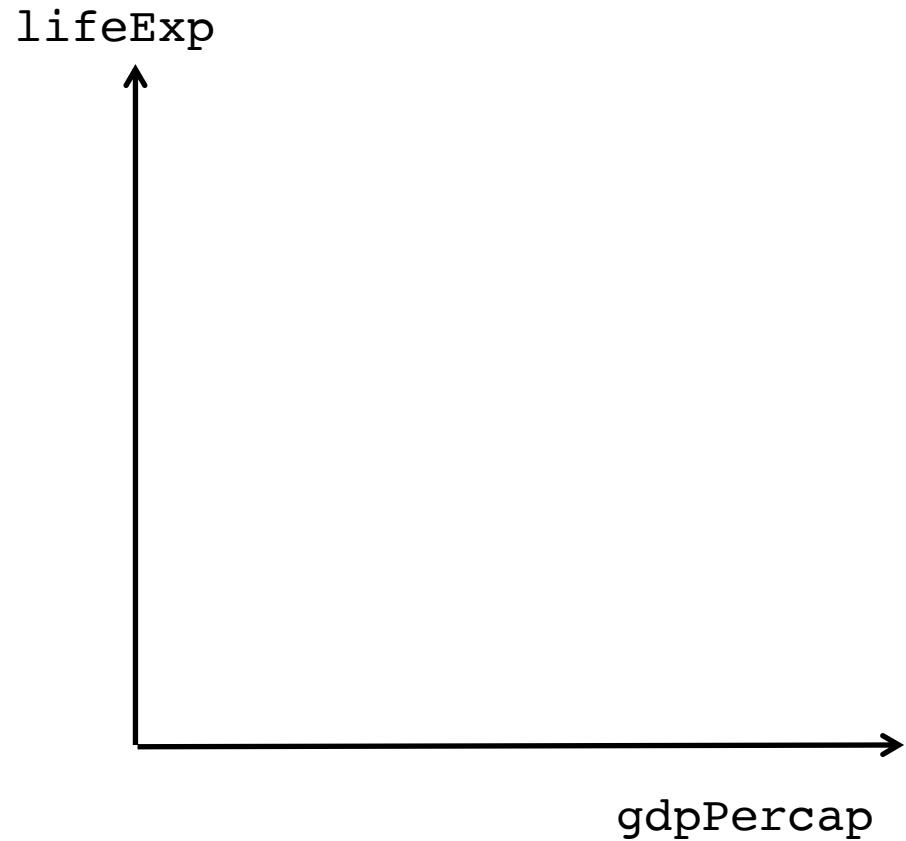
- Chúng tôi muốn thể hiện mối liên quan giữa tuổi thọ trung bình (lifeExp) và thu nhập bình quân đầu người (gdpPercap)
- Hình thức thể hiện: biểu đồ tương quan (scatter plot)
  - trục tung: lifeExp
  - trục hoành: gdpPercap
- Triển khai qua ggplot2

# Lớp 1: định nghĩa biến số và dữ liệu

- Xác định dataframe, biến x, y
- Tên object là p

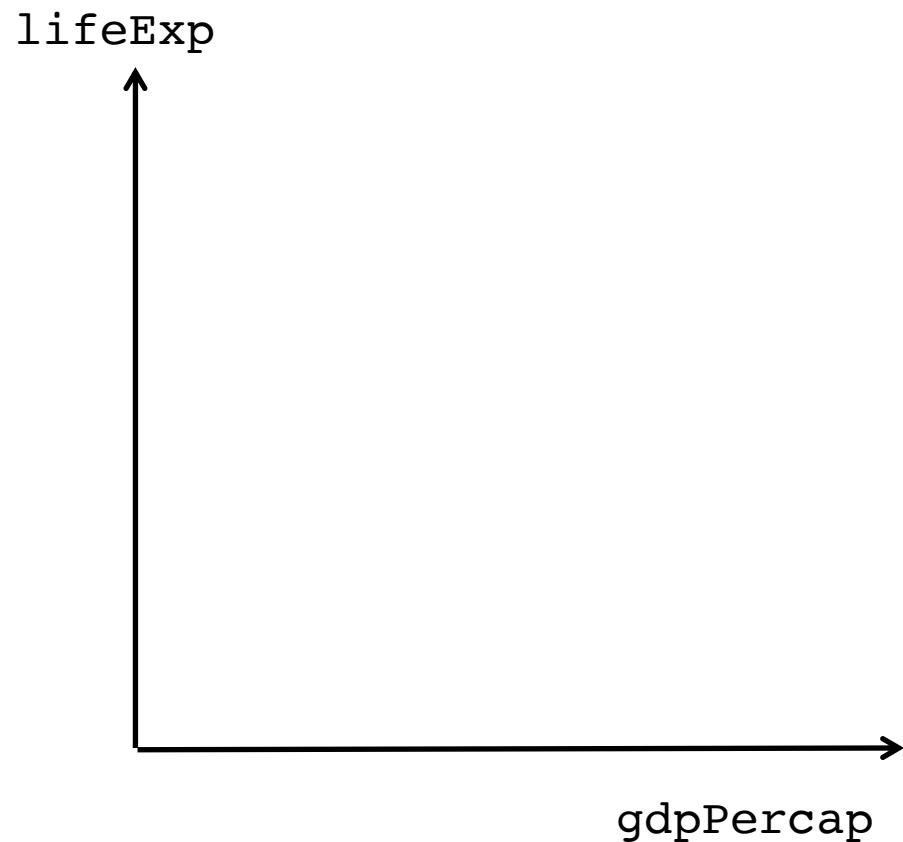
```
library(ggplot2)
p = ggplot(data=gapminder, aes(x=gdpPercap, y=lifeExp))
```

# Lớp 1: định nghĩa biến số và data frame

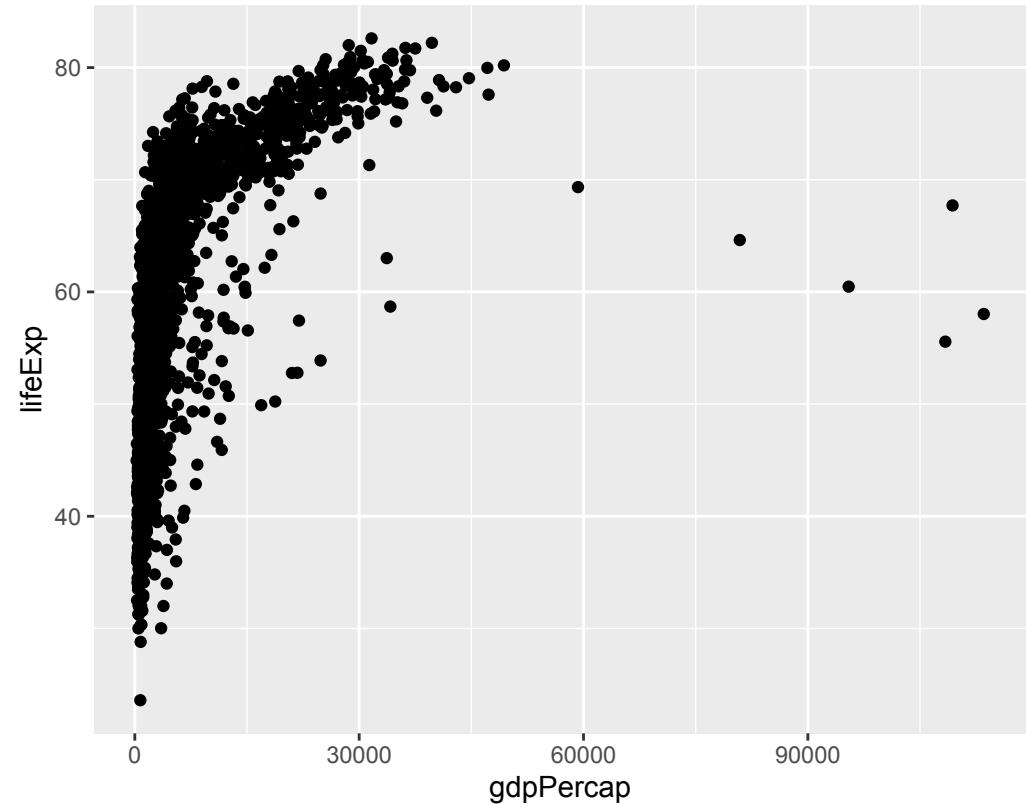


```
p = ggplot(data=gapminder,  
aes(x=gdpPercap, y=lifeExp))
```

# Lớp 2: hình thức thể hiện

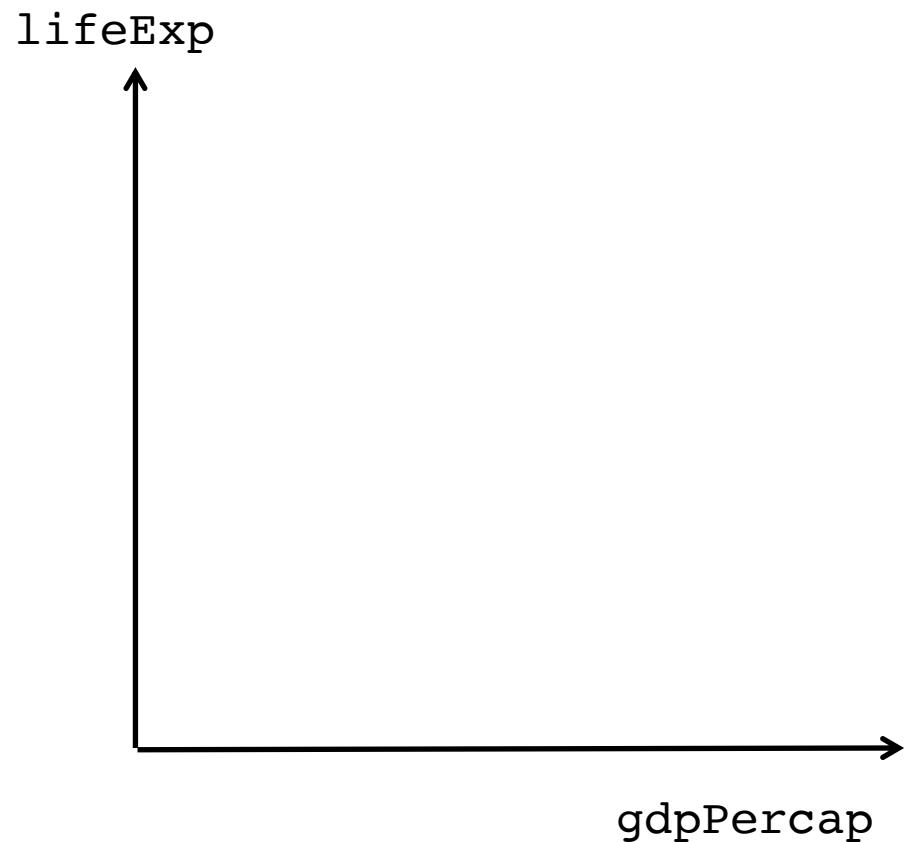


```
p = ggplot(data=gapminder,  
aes(x=gdpPerCap, y=lifeExp))
```

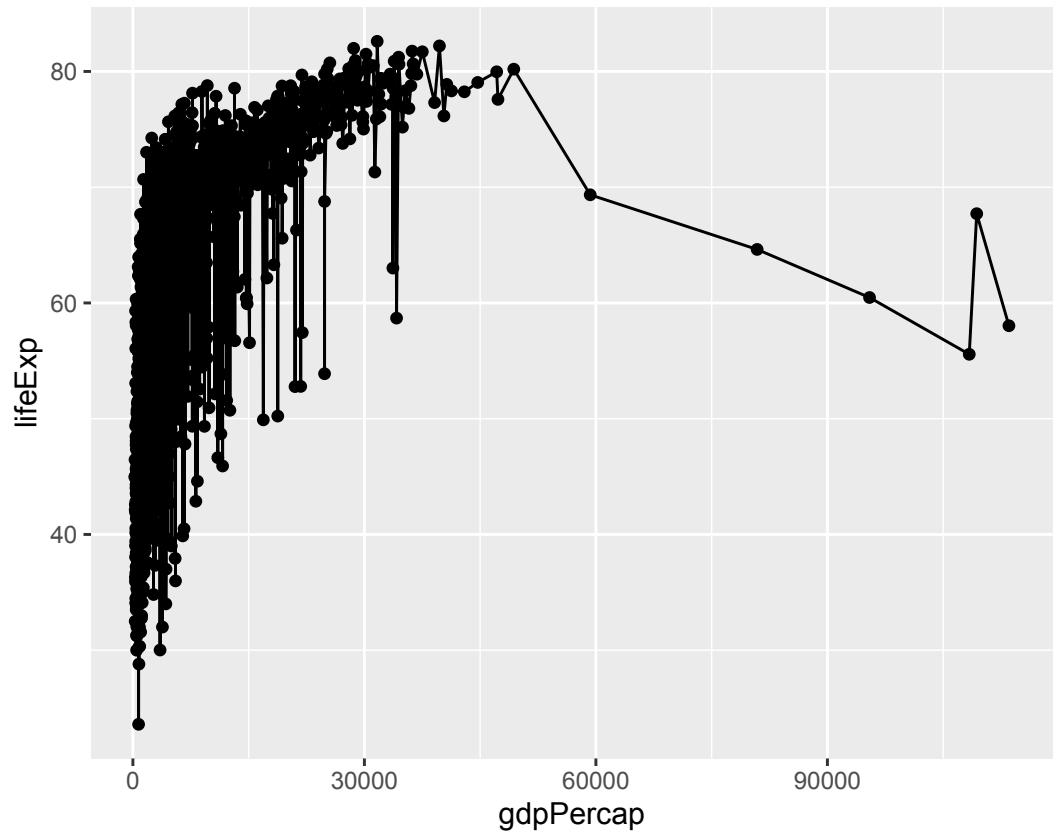


```
p = p + geom_point()  
p
```

# Lớp 2: hình thức thể hiện

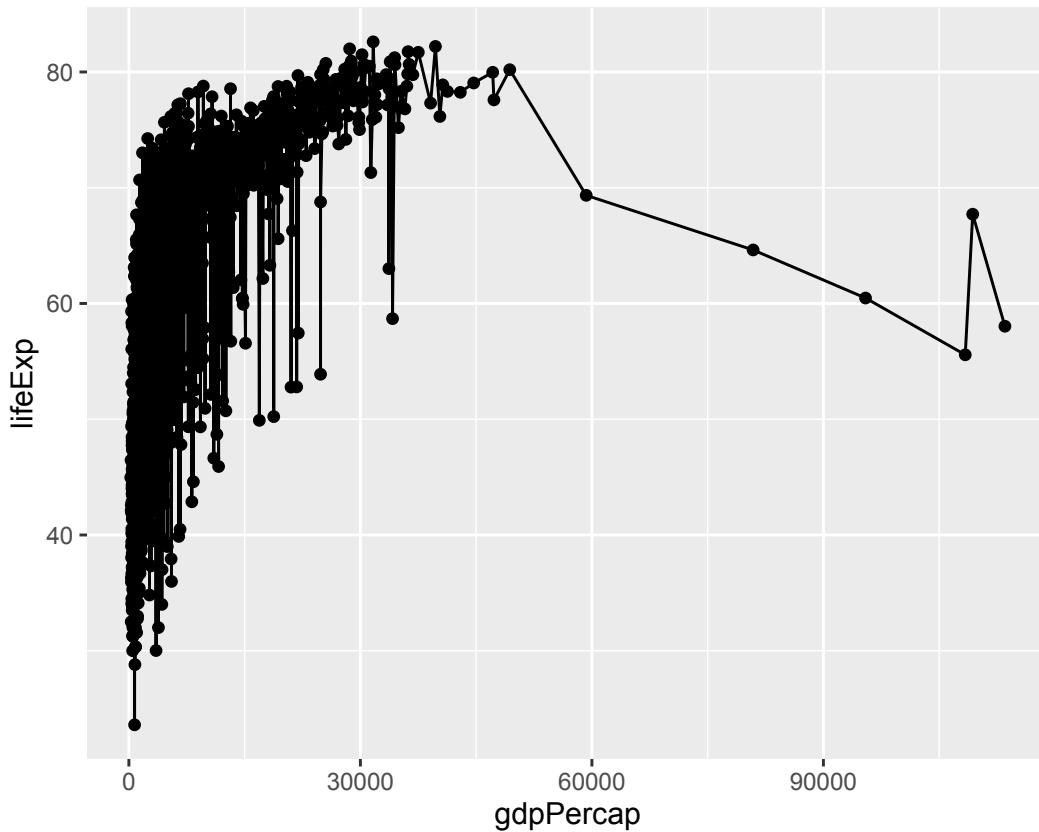


```
p = ggplot(data=gapminder,  
aes(x=gdpPerCap, y=lifeExp))
```

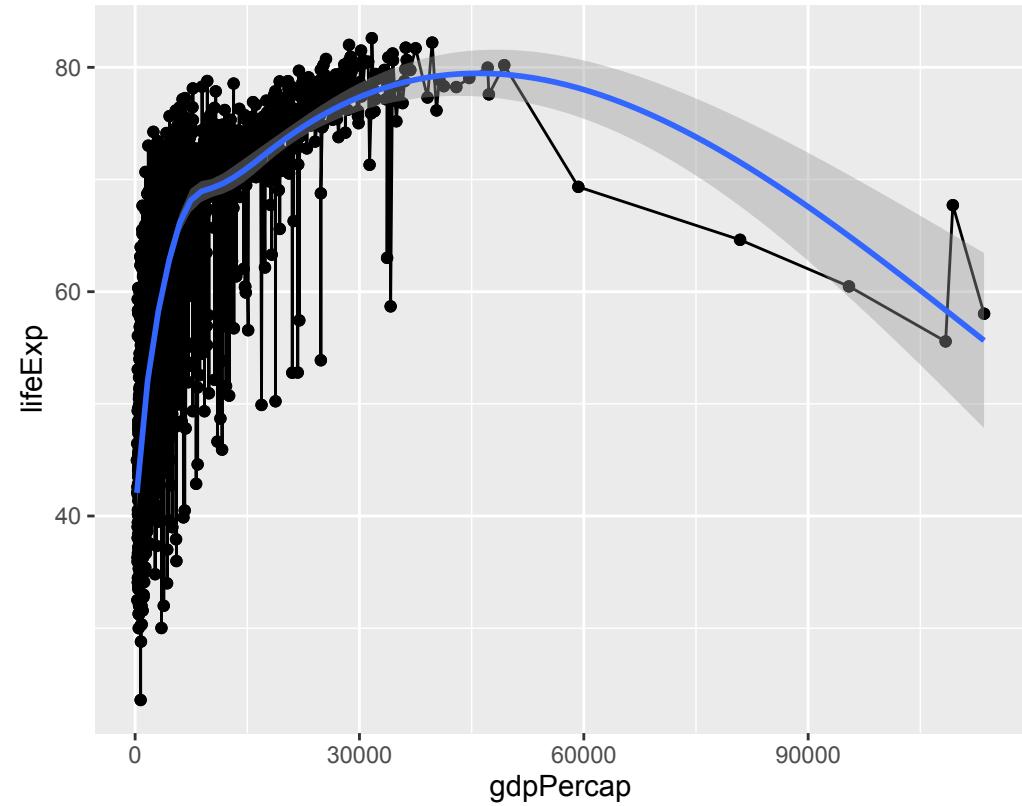


```
p = p + geom_line()  
p
```

# Lớp 2: hình thức thể hiện

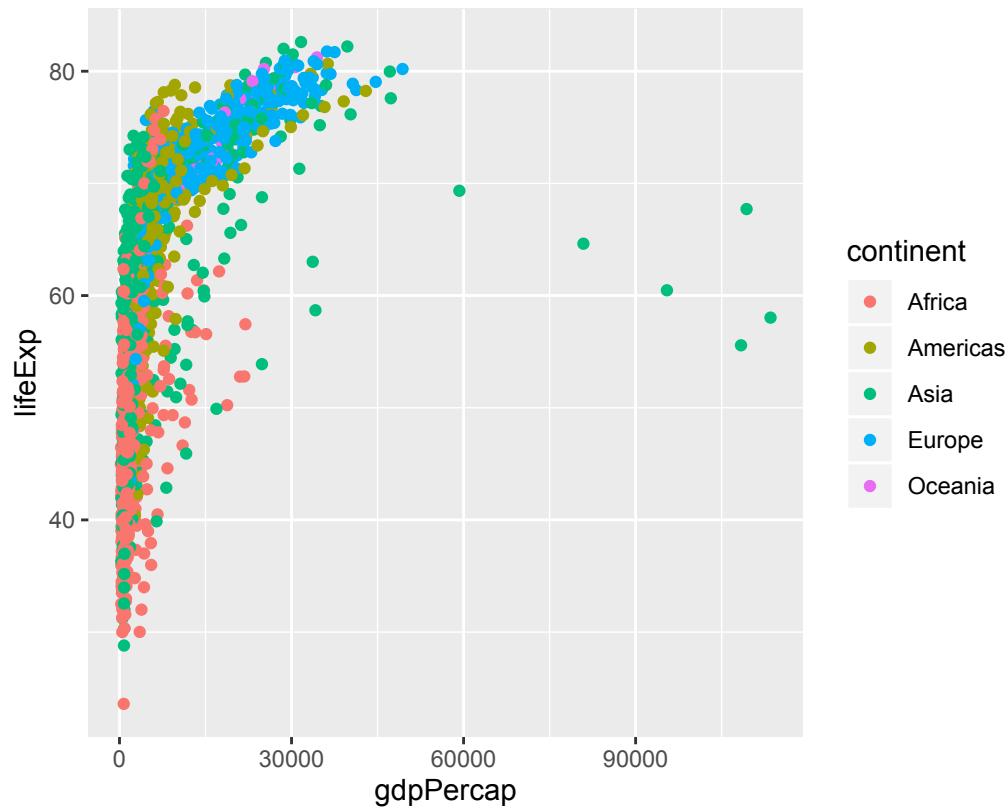


```
p = p + geom_line()  
p
```

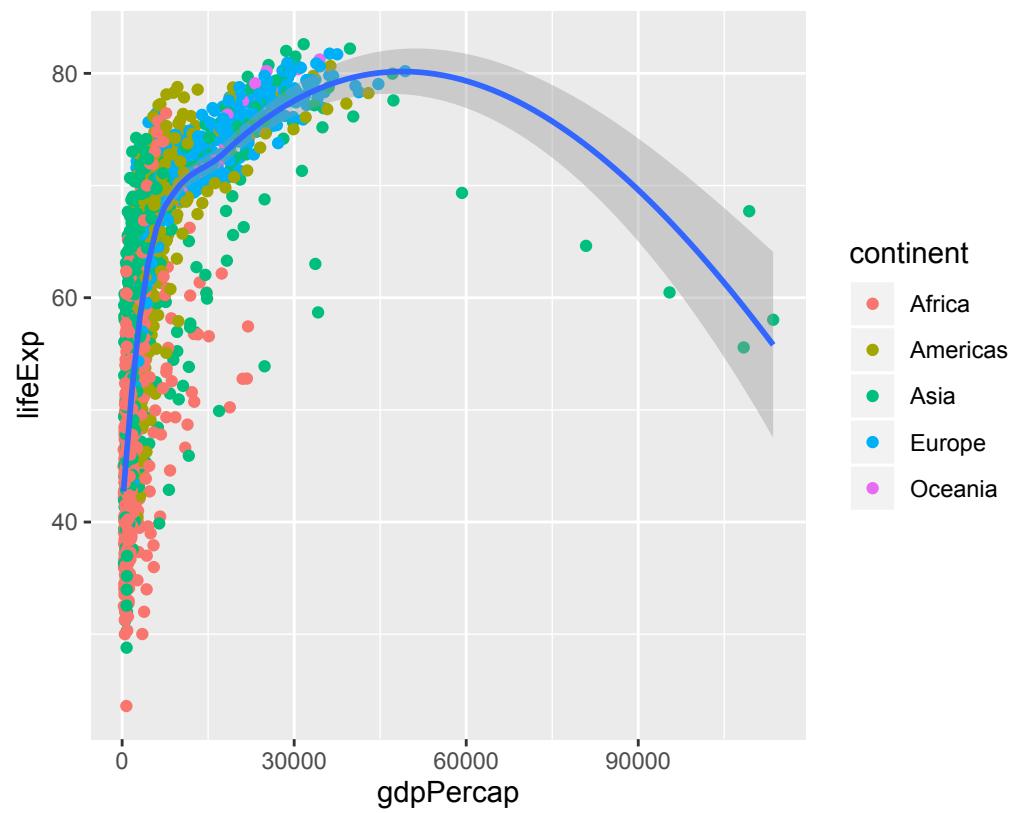


```
p = p + geom_point() + geom_smooth()  
p
```

# Lớp 3: thêm màu

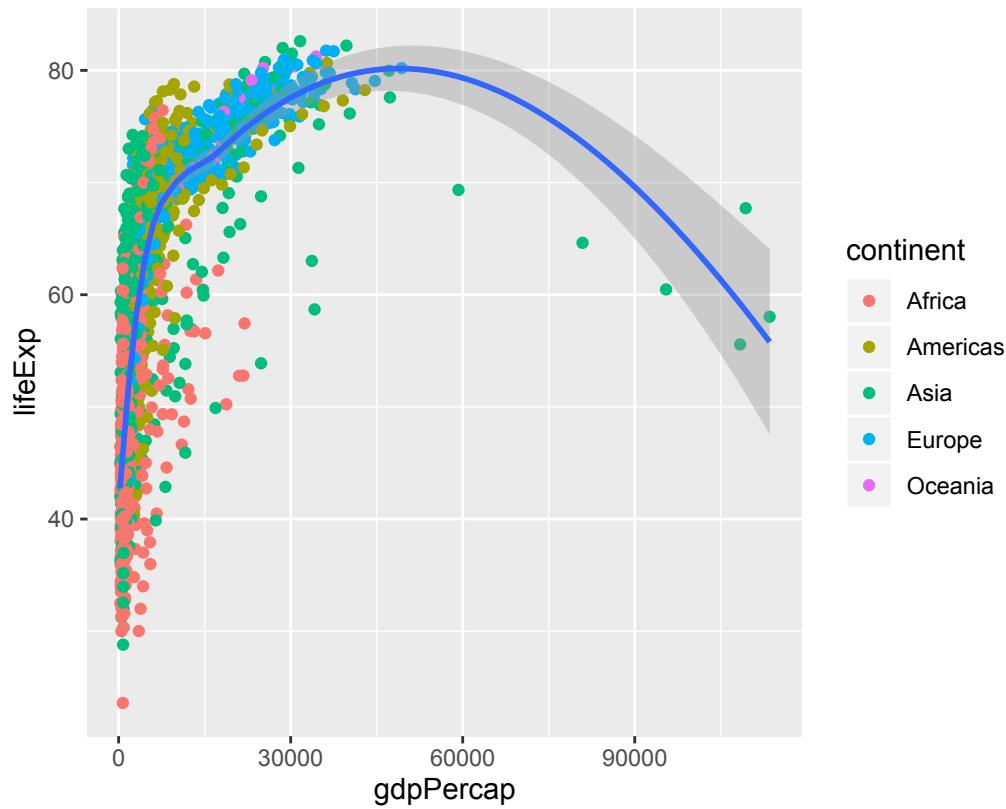


```
p = ggplot(data=gapminder,  
aes(x=gdpPerCap, y=lifeExp,  
color=continent))  
p = p + geom_point()  
p
```

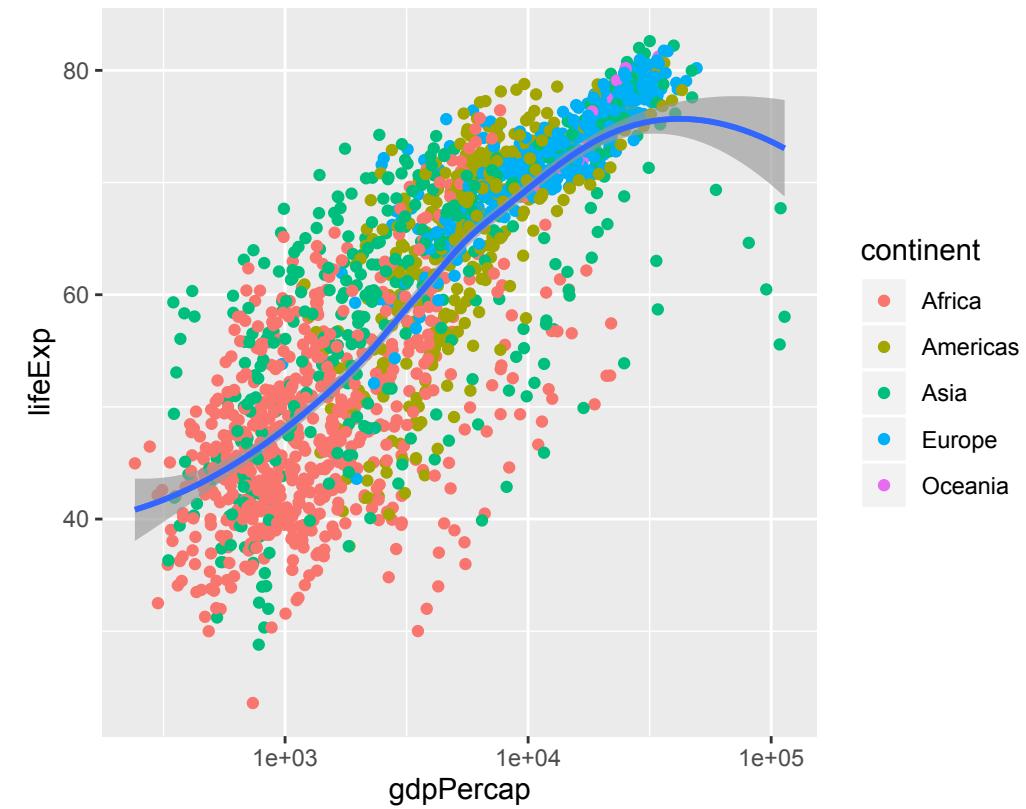


```
p = p + geom_smooth(method="loess")  
p
```

# Lớp 3: hoán chuyển trực hoành

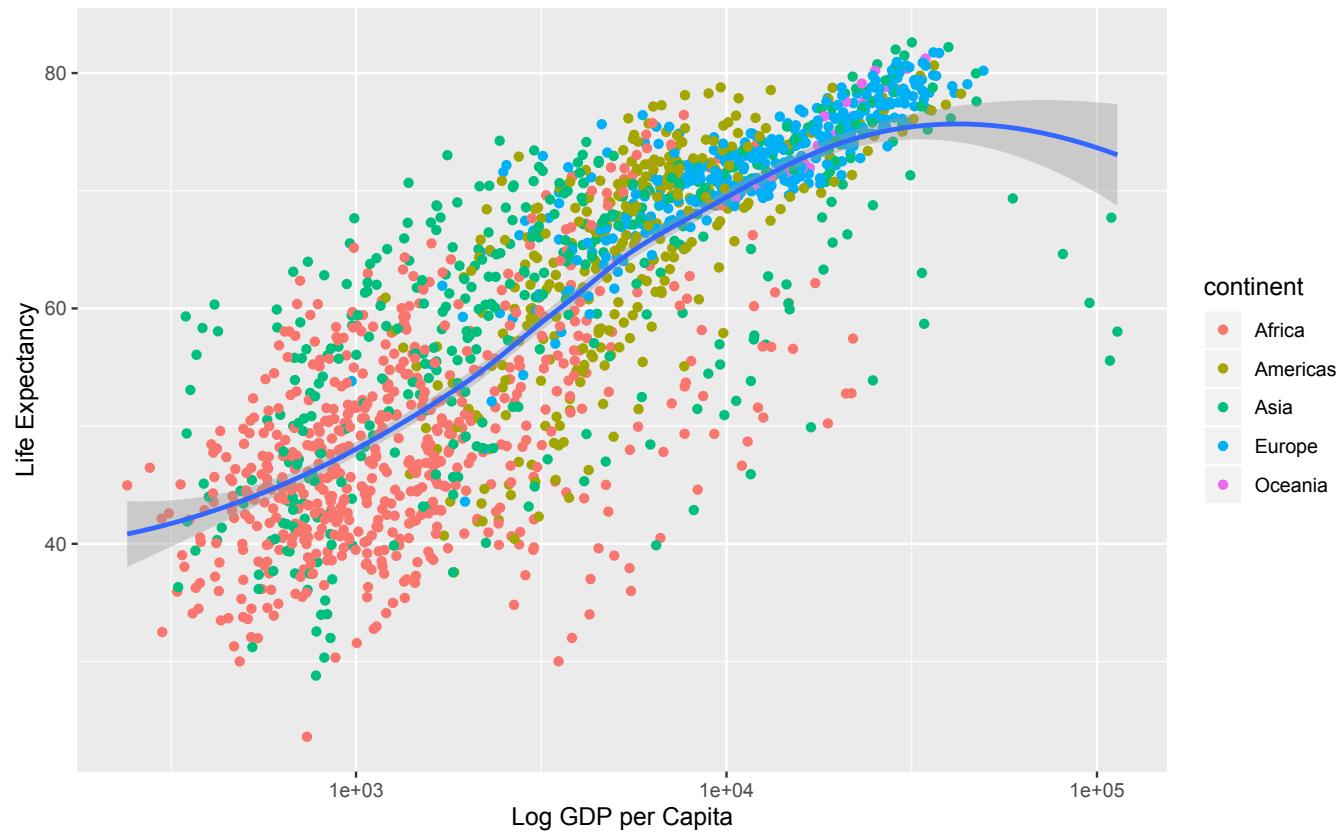


```
p = p + geom_smooth(method="loess")  
p
```



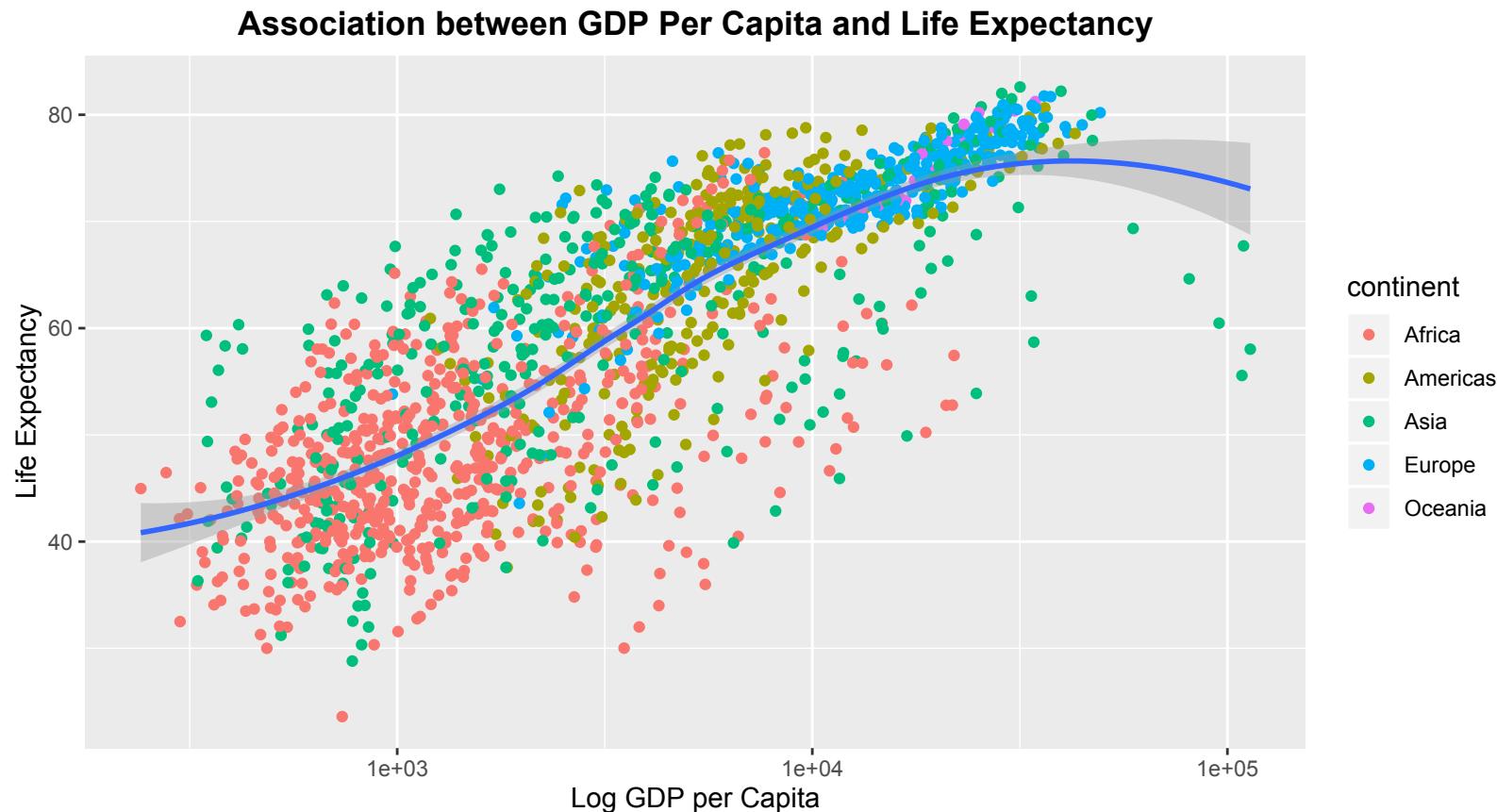
```
p = p + scale_x_log10()  
p
```

# Lớp 4: nhãn trục tung & hoành



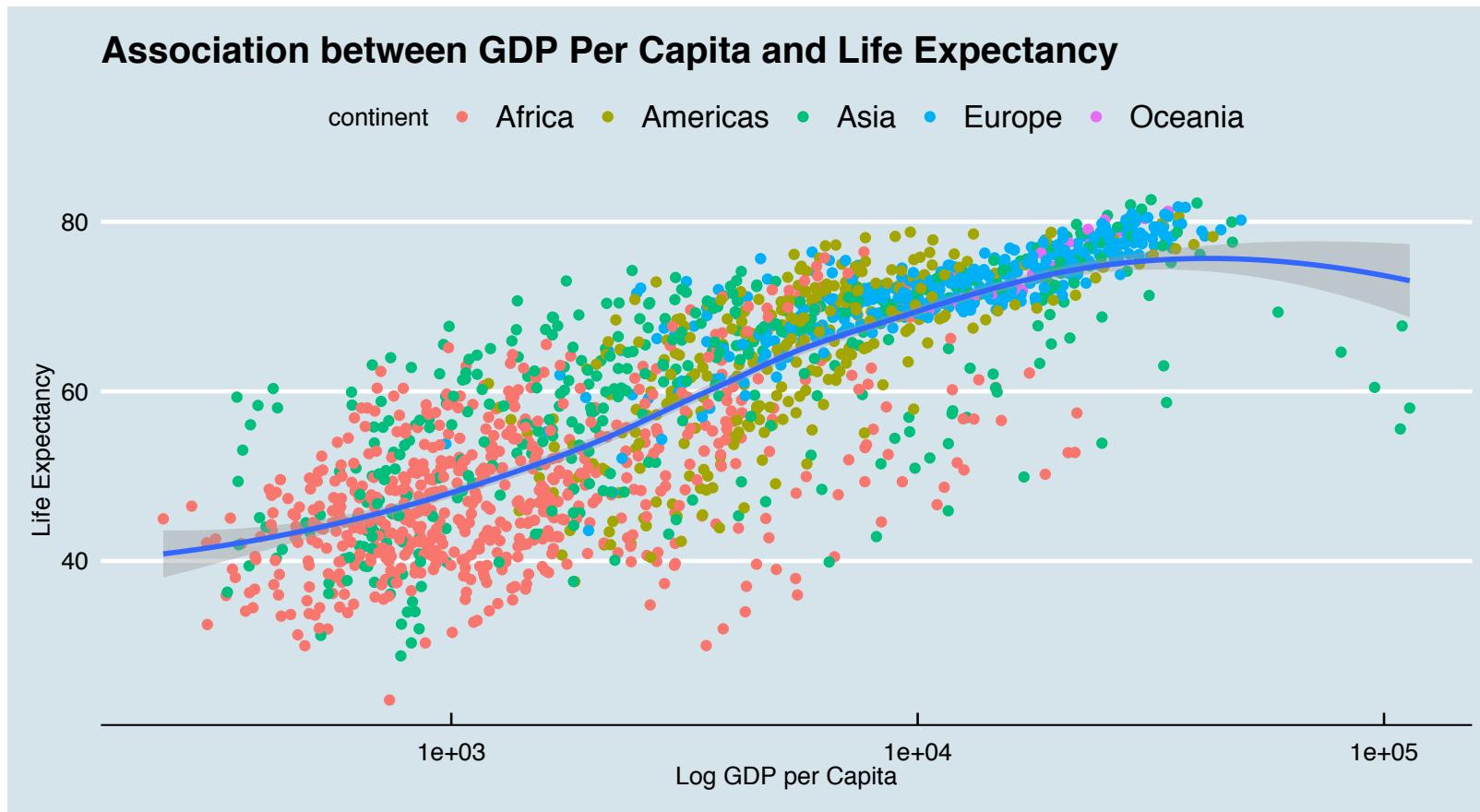
```
p = ggplot(data = gapminder, aes(x = gdpPercap, y = lifeExp))  
p = p + geom_point(aes(color=continent))  
p = p + geom_smooth(method="loess") + p + scale_x_log10()  
p = p + labs(x="Log GDP per Capita", y="Life Expectancy")  
p
```

# Lớp 4: tựa đề (title)



```
p = p + ggttitle("Association between GDP Per Capita and Life Expectancy") + theme(plot.title=element_text(lineheight=0.8, face="bold", hjust = 0.5))
```

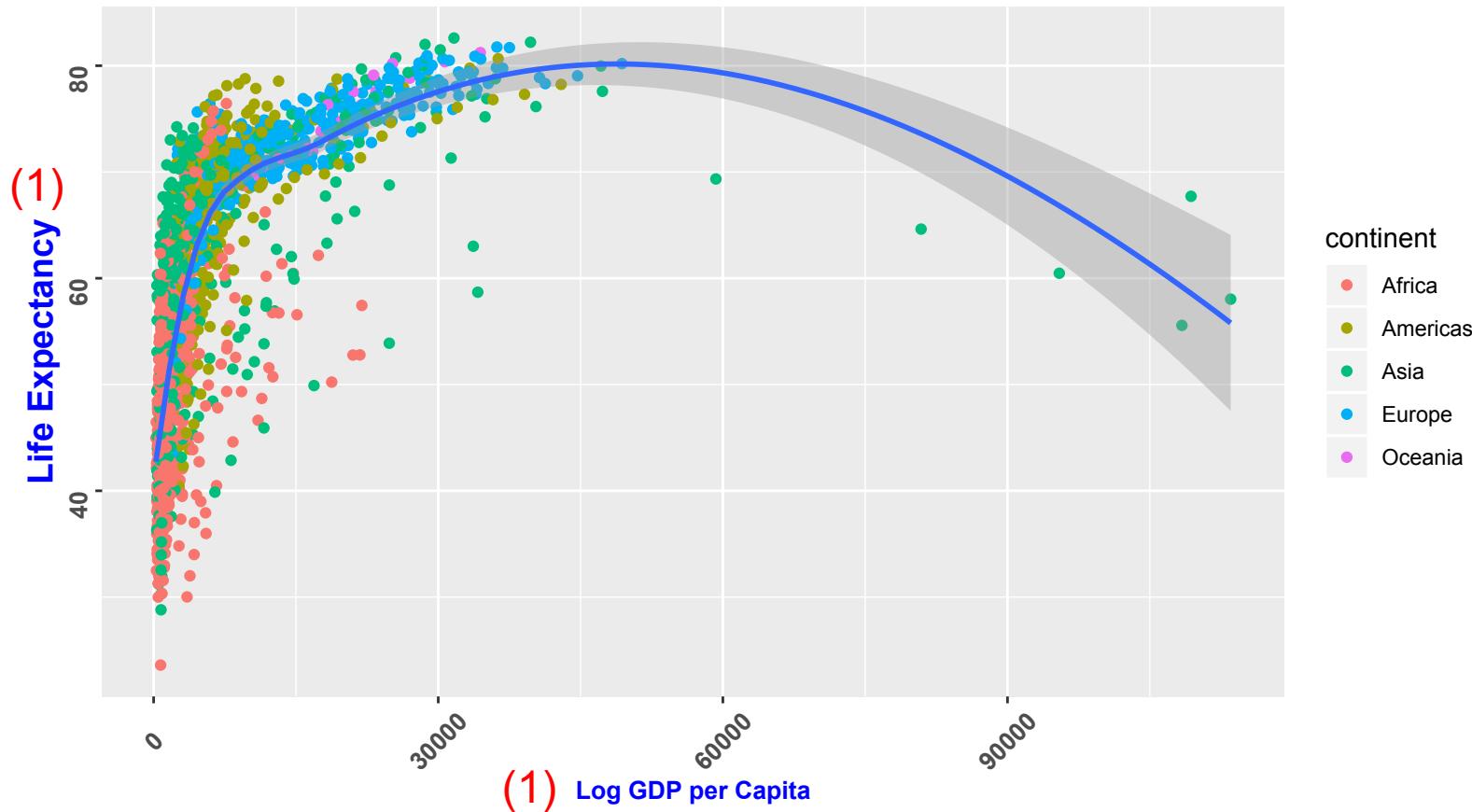
# Lớp 5: theme



```
library(ggthemes)
p + theme_economist()
```

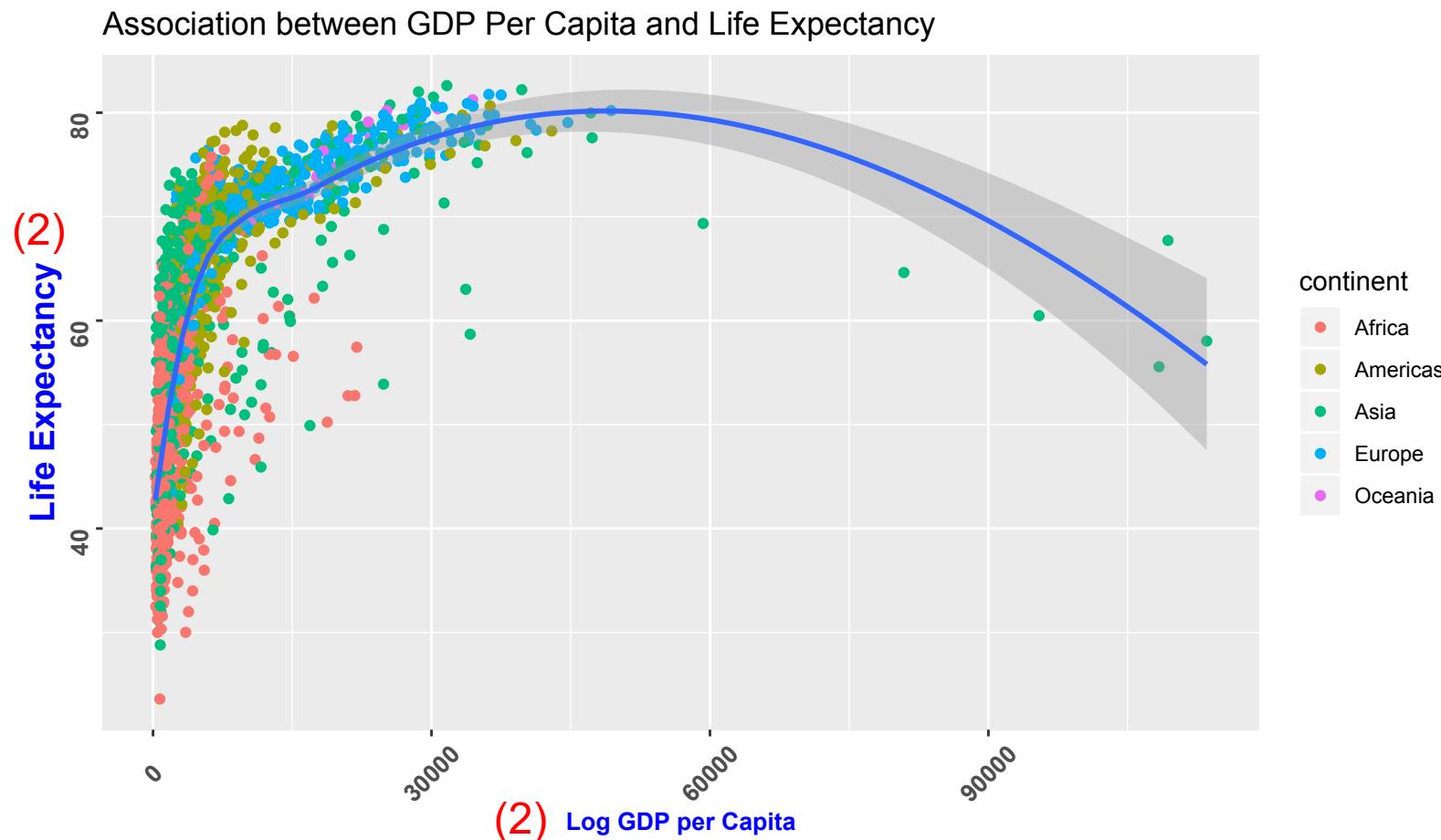
# Văn phạm trang trí ggplot2 : tóm tắt

(1) Association between GDP Per Capita and Life Expectancy



(1) `labs(title="Association between GDP Per Capita and Life Expectancy", x="Log GDP per Capita", y="Life Expectancy")`

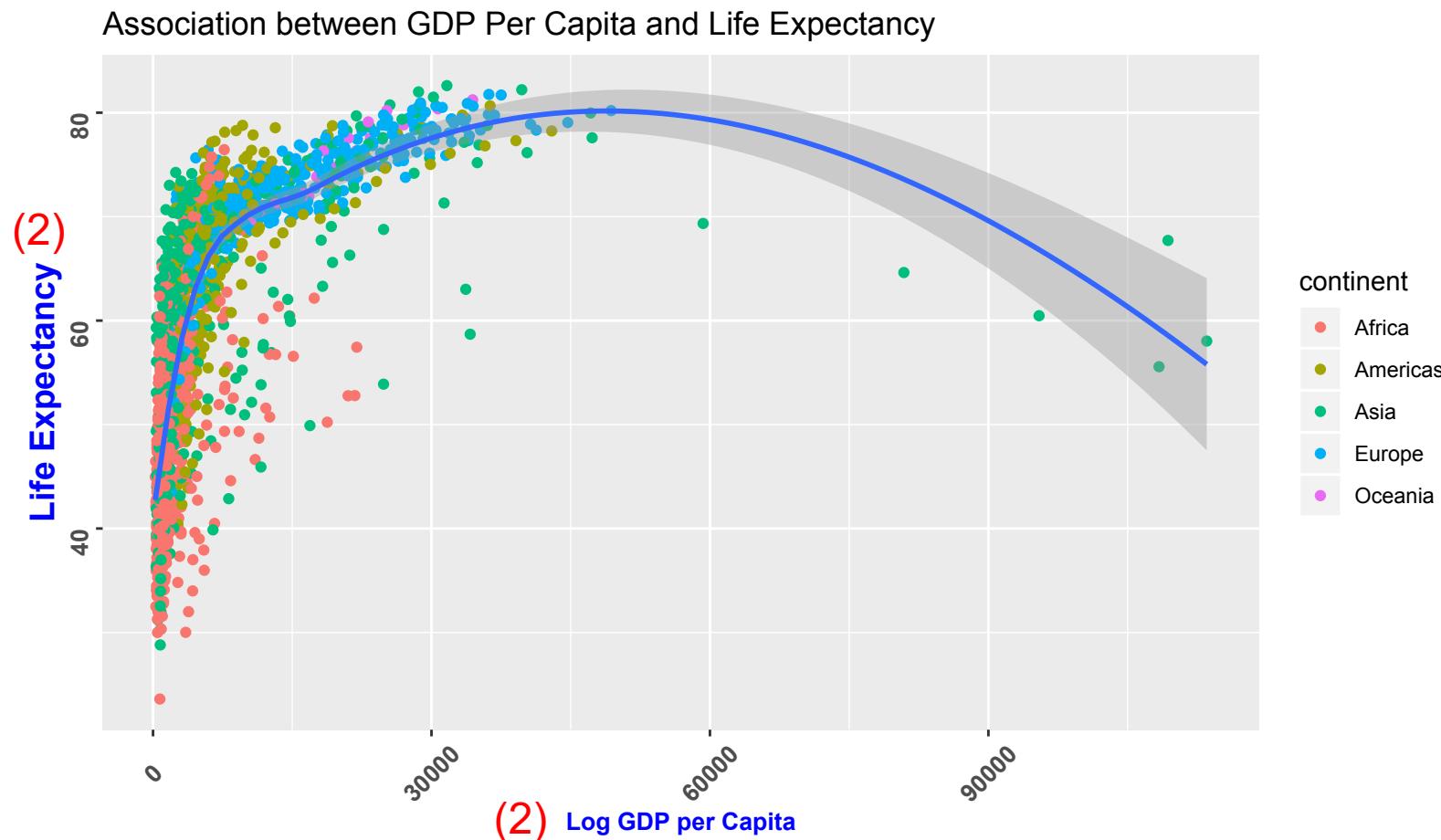
# Văn phạm trang trí ggplot2 : tóm tắt



(2)

```
theme(axis.title.x=element_text(color="blue", size=10, face="bold"),
axis.text.x=element_text(angle=45, vjust=0.5, size=10, face="bold"),
axis.title.y = element_text(color="blue", size=14, face="bold"),
axis.text.y=element_text(angle=90, vjust=0.5, size=10, face="bold"))
```

# Văn phạm trang trí ggplot2 : tóm tắt



(2)  
scale\_x\_continuous(breaks=seq(0, 90000, 30000)) +  
scale\_y\_continuous(breaks=seq(0, 100, 20))

# Hiển thị dữ liệu với ggplot2

- **ggplot2** – một phát triển quan trọng trong hiển thị dữ liệu
- Mô hình đa lớp của ggplot2
- Lớp 1: định nghĩa dữ liệu và tên biến số
- Lớp 2: hình thức thể hiện
- Lớp 3: hoán chuyển trực tung và trực hoành
- Lớp 4: màu sắc và trang trí
- Lớp 5: theme