Biostats 597 HW1

Bing Miu

A look at the data:

```
## # 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.
## # ... with 1,694 more rows
```

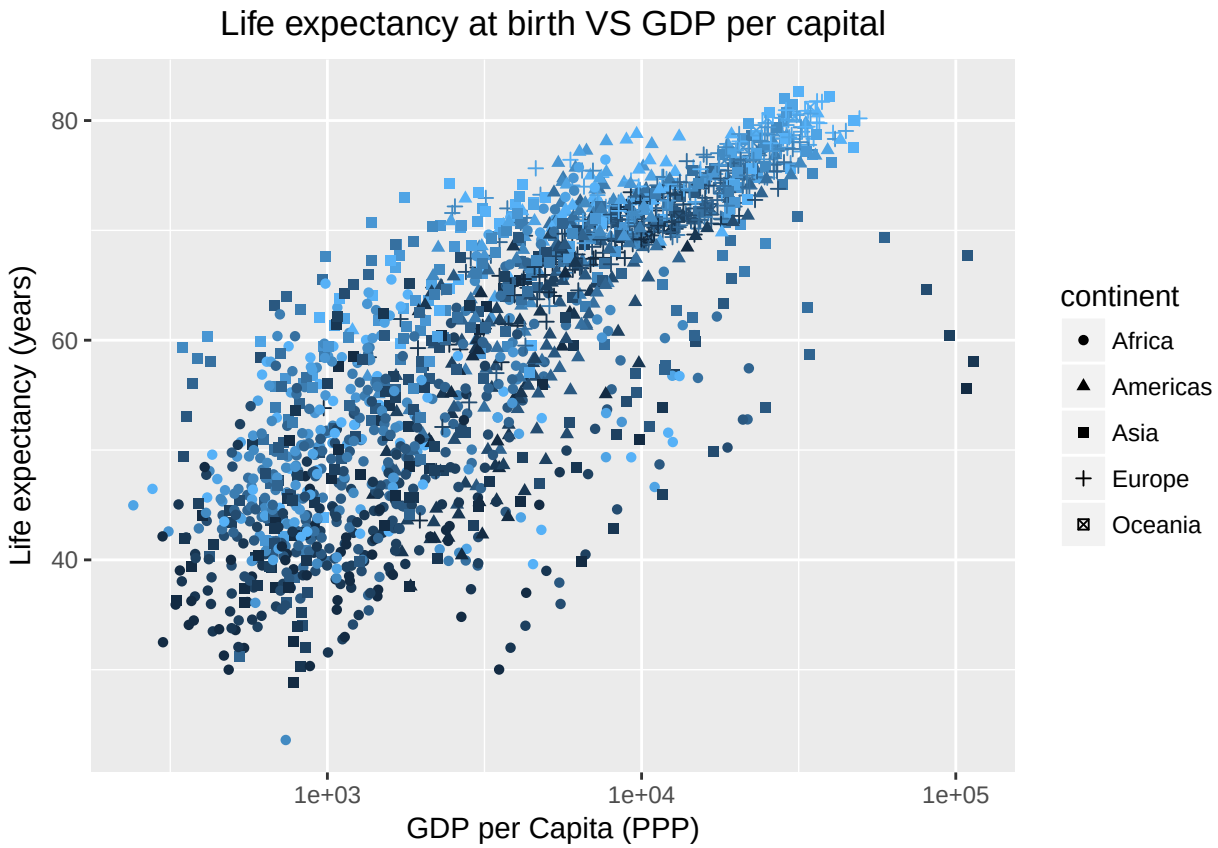
HW exercises:

use ggplot to make the graphs requested in (1) and (2). Throughout, avoid repetition of code. Goal: gain experience with ggplot and reading help files/using google searches to get help on (arguments of) R functions.

1. make one scatter plot using ggplot with

- Life expectancy on the y-axis, label axis as "Life expectancy"
- GDP per capita on the x-axis, use a log-transform of the x-axis while still displaying the (unlogged) GDP values
- shapes to indicate the continent, add a legend
- color to indicate the year, do NOT add a legend for that

```
ggplot(data = gapminder, mapping = aes(x = gdpPercap, y = lifeExp)) +
    labs(title = "Life expectancy at birth VS GDP per capital",
        x = "GDP per Capita (PPP)", y = "Life expectancy (years)") +
    scale_x_continuous(trans = "log10") + theme(plot.title = element_text(hjust = 0.5)) +
    geom_point(mapping = aes(color = year, shape = continent)) +
    scale_color_continuous(guide = "none")
```

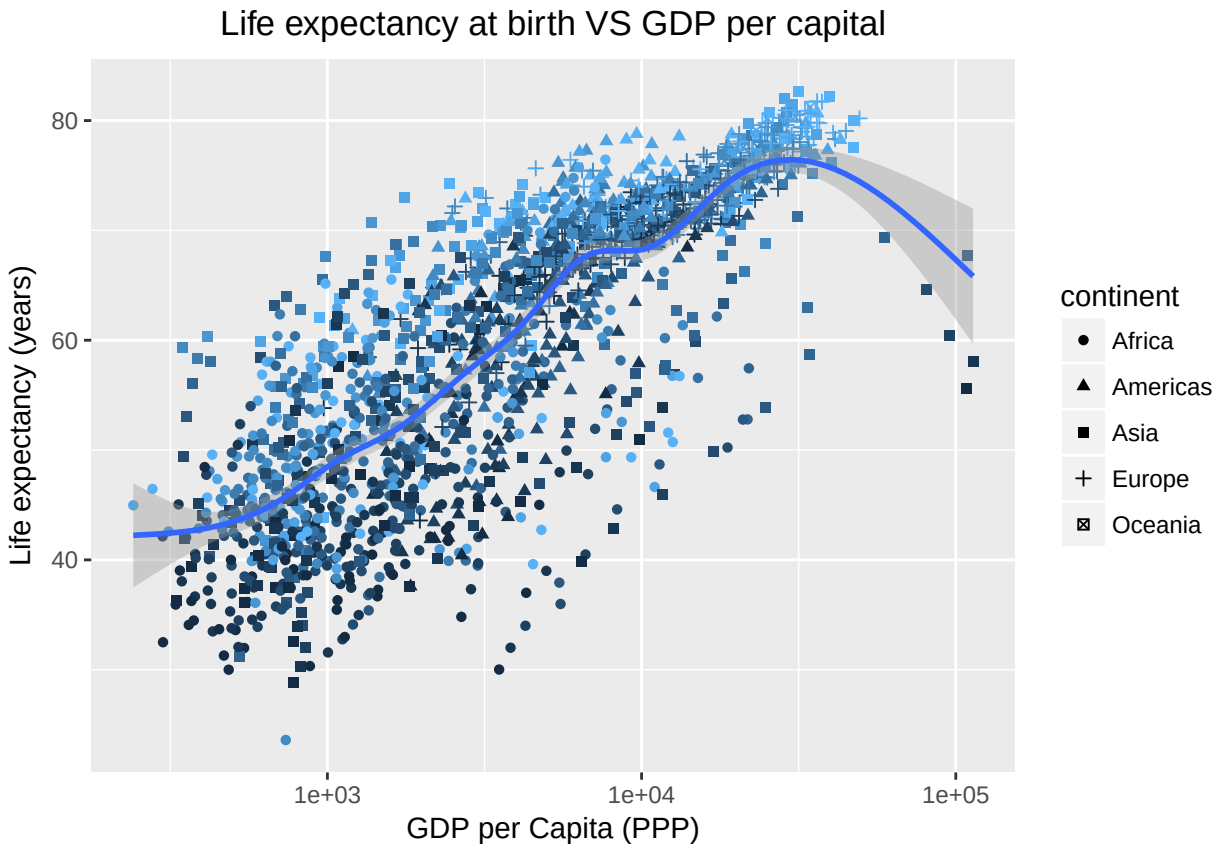


2. add two geom_smooth layers (a) and (b) to the plot you made in (1) where

(a) is fitted to the entire dataset and uses default settings for the smoother used and plotting settings

```
ggplot(data = gapminder, mapping = aes(x = gdpPercap, y = lifeExp)) +
    labs(title = "Life expectancy at birth VS GDP per capital",
        x = "GDP per Capita (PPP)", y = "Life expectancy (years)") +
    scale_x_continuous(trans = "log10") + theme(plot.title = element_text(hjust = 0.5)) +
    geom_point(mapping = aes(color = year, shape = continent)) +
    scale_color_continuous(guide = "none") + geom_smooth()
```

```
## `geom_smooth()` using method = 'gam'
```

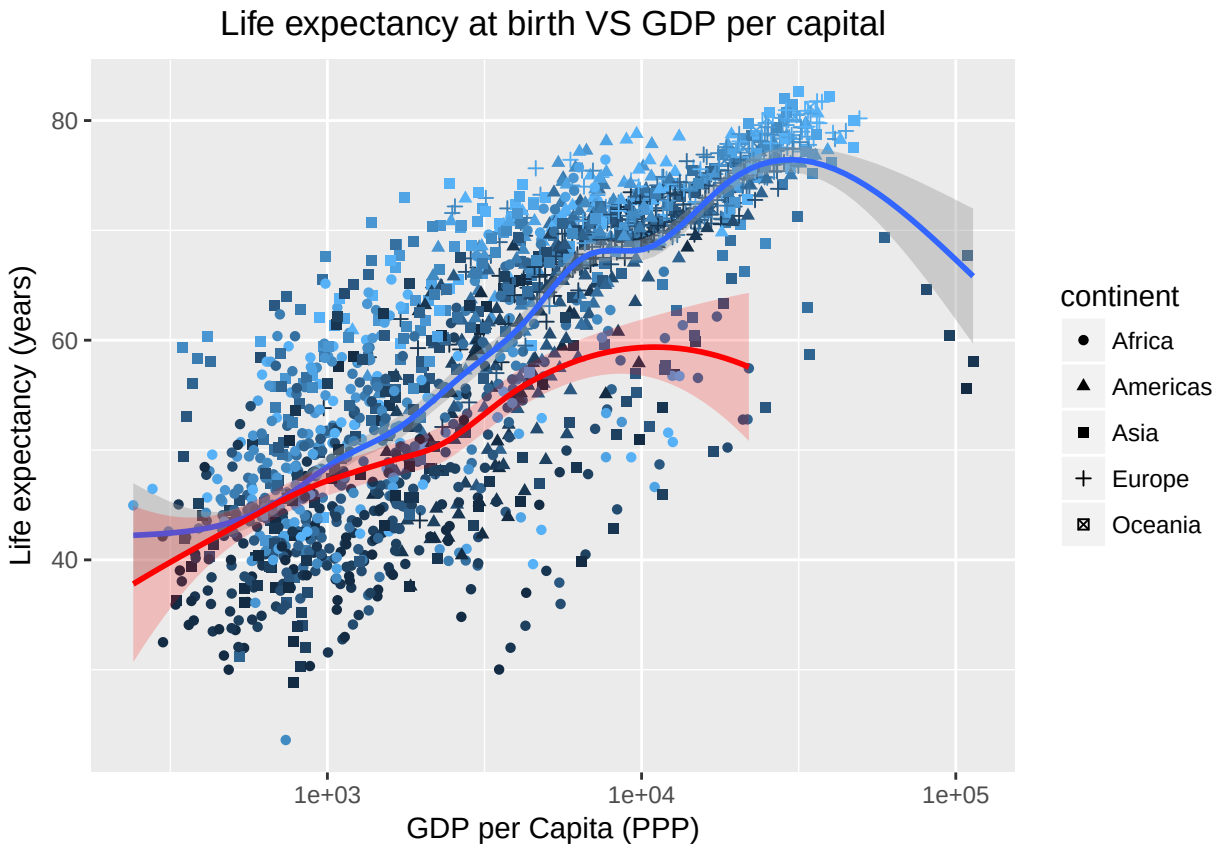


(b) is fitted to data in Africa only, provides a 99% confidence interval, is plotted in red (both point estimates as well as confidence interval), and is transparent (so if it overlaps with (a), (a) is still visible). Hint: to fit to data in Africa only, consider using "data = filter(gapminder, continent ==")Africa")"

```
ggplot(data = gapminder, mapping = aes(x = gdpPercap, y = lifeExp)) +
    labs(title = "Life expectancy at birth VS GDP per capital",
        x = "GDP per Capita (PPP)", y = "Life expectancy (years)") +
    scale_x_continuous(trans = "log10") + theme(plot.title = element_text(hjust = 0.5)) +
    scale_color_continuous(guide = "none") + geom_point(mapping = aes(color = year,
    shape = continent)) + geom_smooth() + geom_smooth(data = filter(gapminder,
    continent == "Africa"), color = "red", fill = "red", level = 0.99,
    alpha = 0.2)
```

```
## `geom_smooth()` using method = 'gam'
```

```
## `geom_smooth()` using method = 'loess'
```



Red = fitted to data in Africa only
Blue = fitted to data in all continents