

# HW1

- What to hand in: soft copies of a notebook (Rmd file) with your answers (you can use this one as a template), and the corresponding knitted pdf.
- When: Due at noon, Thursday September 6, 2018.
- How: Information on how to submit the files will be given on moodle.

We will use data from the gapminder package, let's load it.

```
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.2.1
## √ ggplot2 3.0.0      √ purrr  0.2.5
## √ tibble  1.4.2      √ dplyr  0.7.6
## √ tidyr   0.8.1      √ stringr 1.3.1
## √ readr   1.1.1      √ forcats 0.3.0

## -- Conflicts ----- tidyverse_conflicts()
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
#install.packages("gapminder")
library(gapminder)
```

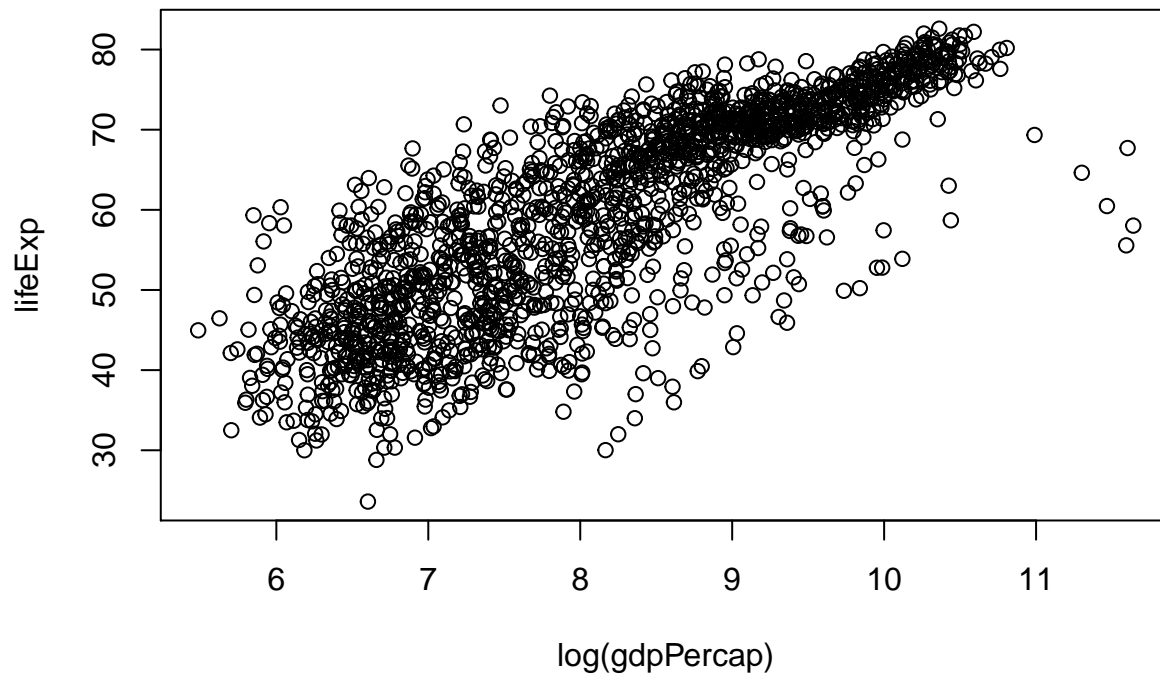
A look at the data:

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

A simple base R plot of life expectancy against log(GDP):

```
# Base R plot
plot(lifeExp ~ log(gdpPercap), data = gapminder)
```



## HW exercises:

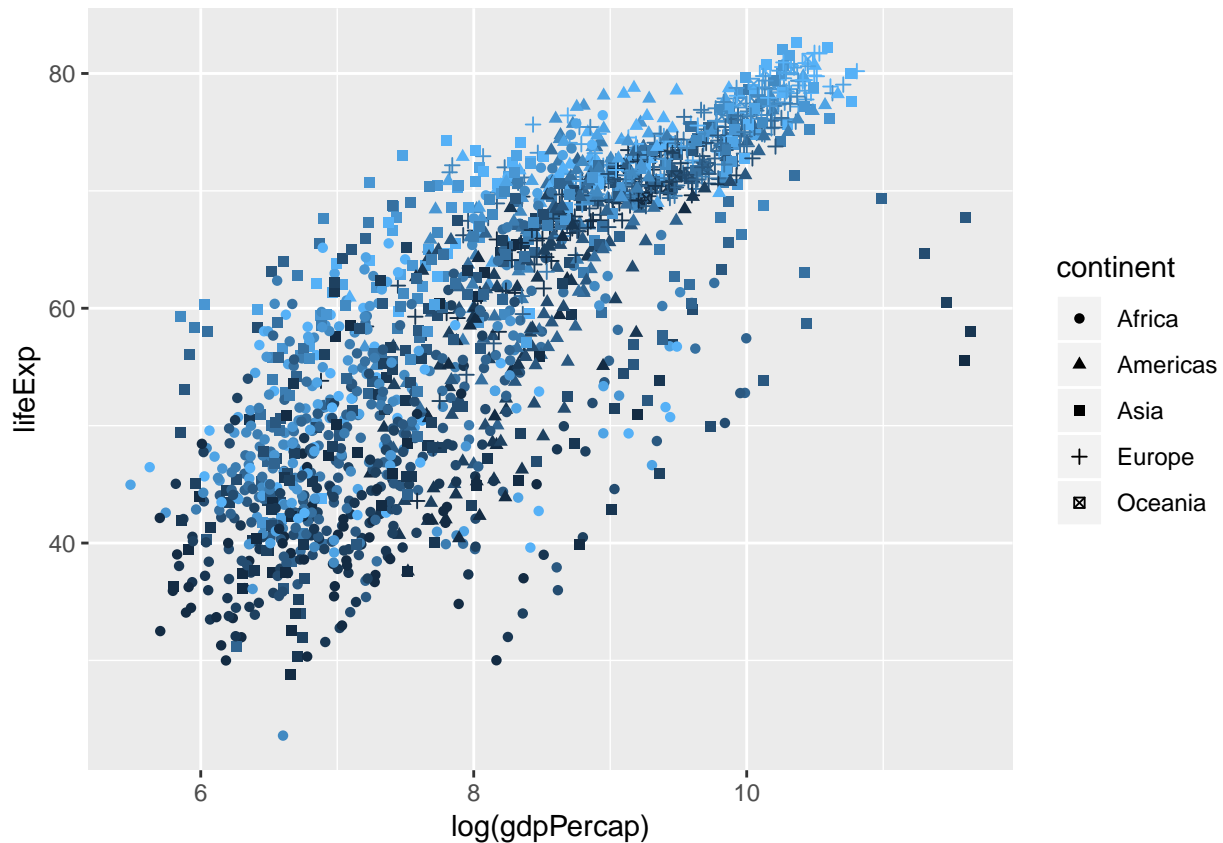
use ggplot to make the graphs requested in (1) and (2). Throughout, avoid repetition of code. Goal: gain experience with ggplot and reading help files/using google searches to get help on (arguments of) R functions.

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

Solution:

```
ggplot(gapminder) +  
  geom_point(mapping = aes(x = log(gdpPercap), y = lifeExp, color = year, shape = continent)) + guides()
```



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
- (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")`”

Solution:

```
ggplot(gapminder, mapping = aes(x = log(gdpPercap), y = lifeExp)) +
  geom_point(mapping = aes(color = year, shape = continent)) +
  geom_smooth(data = filter(gapminder, continent == "Africa"), se = T, level = 0.99, color = "red") + g
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

