

# hw03\_\_dplyr\_\_ggplot2

*Xuemeng Li*

*29/09/2019*

## Contents

<b>Overview</b>	<b>1</b>
<b>Task Option 2</b>	<b>1</b>
1. A table to show . . . . .	1
2. Visualize the data in a figure to get better comparison . . . . .	2
<b>Task Option 3</b>	<b>3</b>
Point spreading . . . . .	3
<b>Task Option 4</b>	<b>4</b>
Compare the mean for each continent with the trimmed mean . . . . .	4
<b>Task Option 5</b>	<b>5</b>
The mean of life expectancy of each continents are calculated and compared as follow: . . . . .	5

## Overview

The goal is to manipulate and explore a dataset with the dplyr package, complemented by visualizations made with ggplot2.

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

## Task Option 2

Get the maximum and minimum of GDP per capita for all continents.

### 1. A table to show

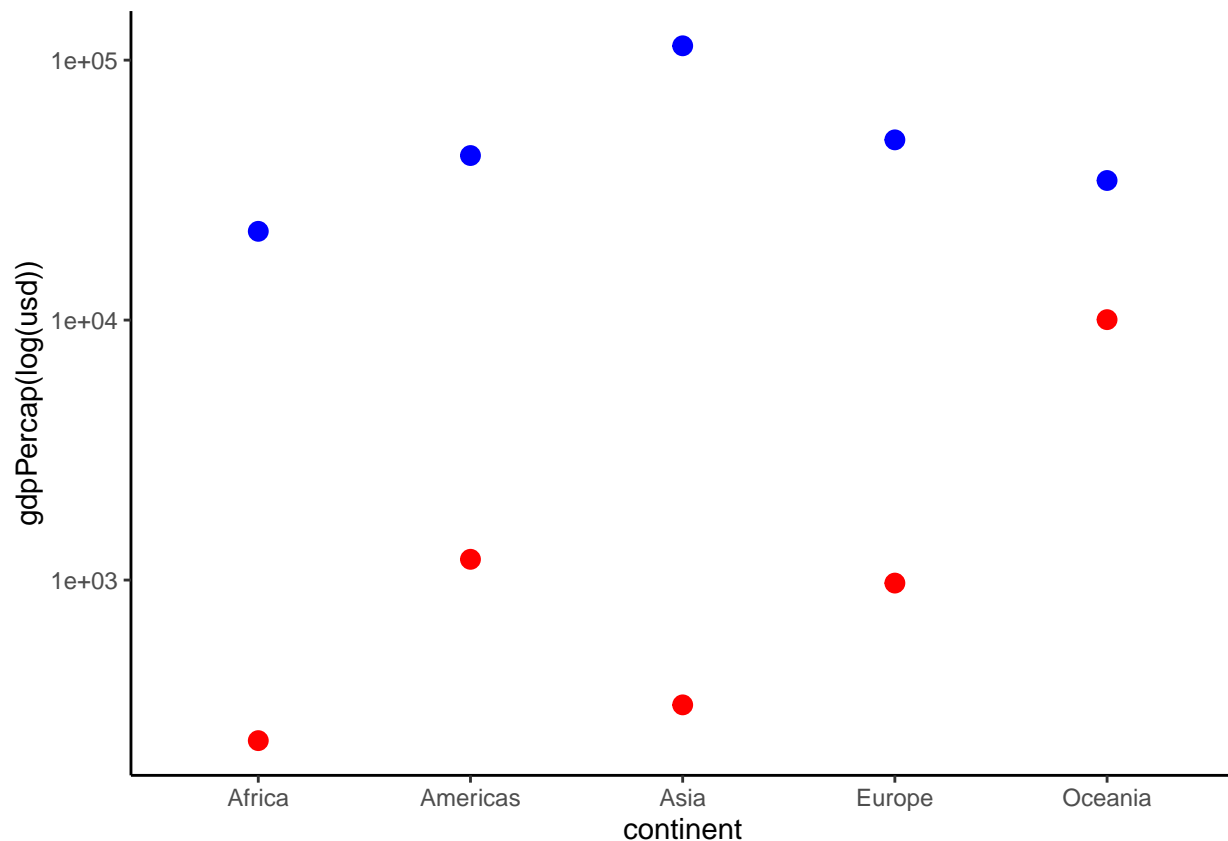
```
gapminder %>%
  select (gdpPercap, continent) %>%
  group_by (continent) %>%
  summarize(max_cap = max(gdpPercap), min_cap = min(gdpPercap), median_cap = median(gdpPercap))
```

```
## # A tibble: 5 x 4
##   continent max_cap min_cap median_cap
##   <fct>      <dbl>   <dbl>      <dbl>
## 1 Africa    21951.    241.      1192.
## 2 Americas  42952.   1202.     5466.
## 3 Asia     113523.    331      2647.
## 4 Europe    49357.    974.     12082.
## 5 Oceania   34435.  10040.    17983.
```

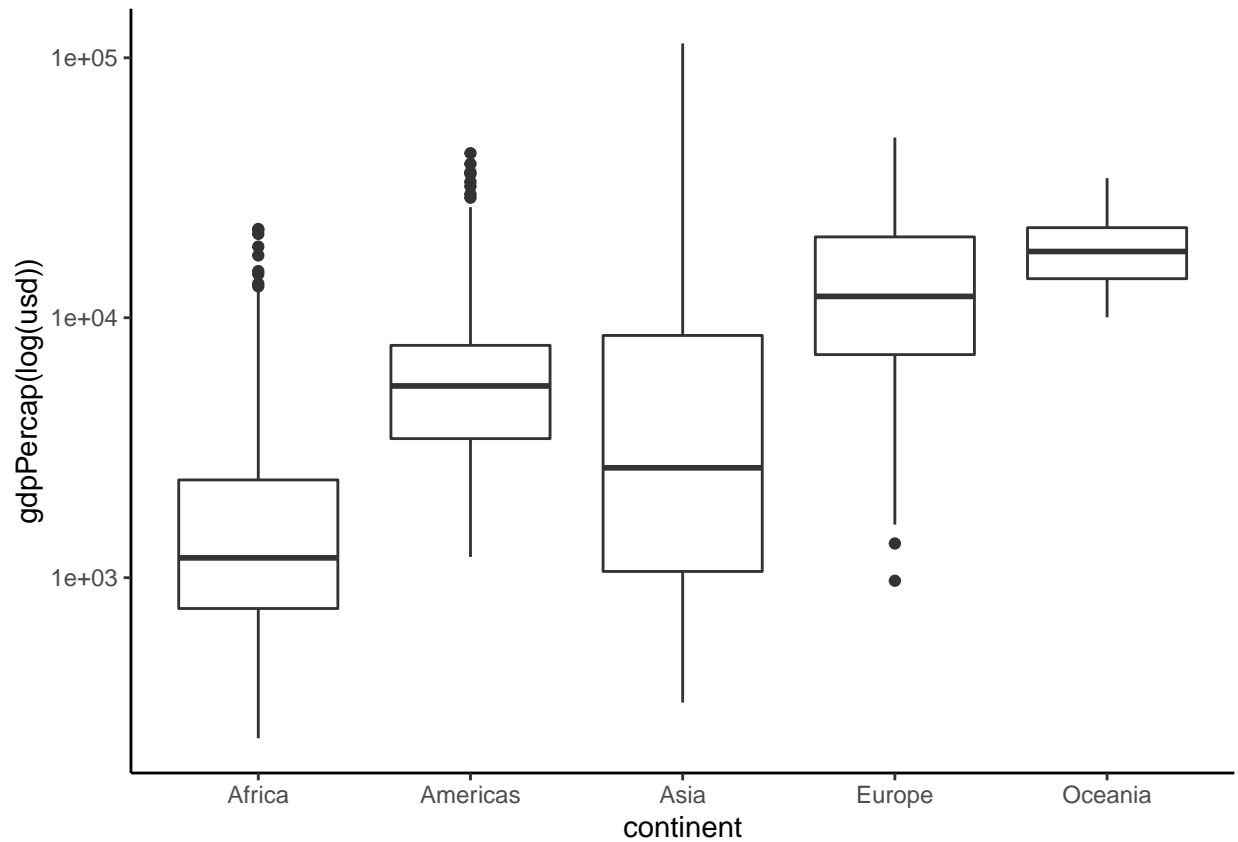
## 2. Visualize the data in a figure to get better comparison

```
m <- gapminder %>%
  select (gdpPercap, continent) %>%
  group_by (continent) %>%
  summarize(max_cap = max(gdpPercap), min_cap = min(gdpPercap), median_cap = median(gdpPercap))

ggplot(gapminder, aes(continent, gdpPercap)) +
  scale_y_log10("gdpPercap(log(usd))") +
  geom_point(data = m, aes(x = continent, y = max_cap), colour = "blue", size = 3) +
  geom_point(data = m, aes(x = continent, y = min_cap), colour = "red", size = 3) +
  theme_classic()
```



```
gapminder %>%
  select (gdpPercap, continent) %>%
  group_by (continent) %>%
  ggplot (aes (continent, gdpPercap)) +
  geom_boxplot() +
  scale_y_log10("gdpPercap(log(usd))") +
  theme_classic()
```

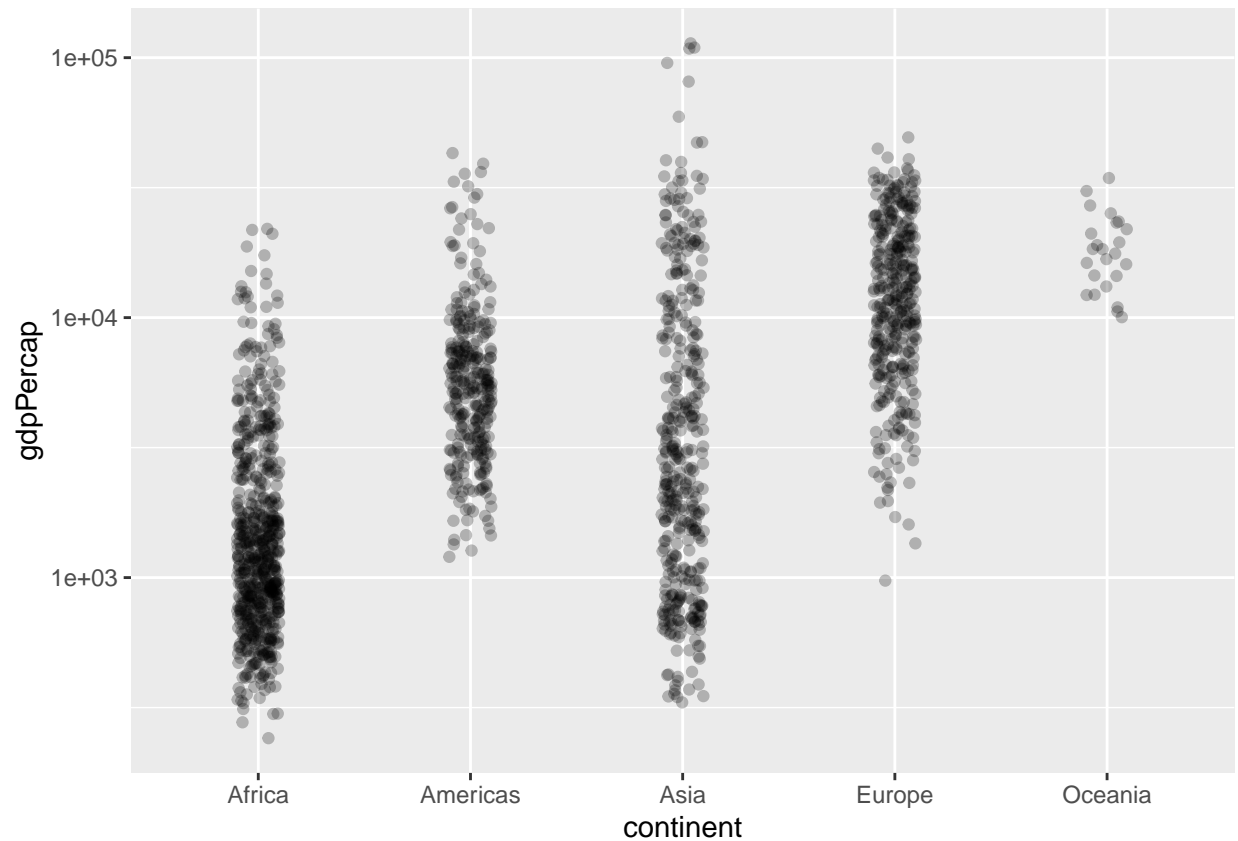


## Task Option 3

Look at the spread of GDP per capita within the continents.

### Point spreading

```
ggplot(gapminder, aes(x = continent, y = gdpPercap)) +
  scale_y_log10() +
  geom_jitter(position = position_jitter(width = 0.1, height = 0), alpha = 1/4)
```



## Task Option 4

Compute a trimmed mean of life expectancy for different years. Or a weighted mean, weighting by population. Just try something other than the plain vanilla mean.

Compare the mean for each continent with the trimmed mean

```
p1<-gapminder %>%
  select (lifeExp, year, continent) %>%
  group_by(continent, year) %>%
  mutate (mean_lifeExp = mean(lifeExp)) %>%
  mutate (trimmed_mean_lifeExp = mean(lifeExp, trim = 0.1)) %>%
  select (year, continent, mean_lifeExp, trimmed_mean_lifeExp) %>%
  ggplot(aes(x = year, y = mean_lifeExp, col = continent))+
  geom_point (size = 1) +
  geom_line (size = 0.5)

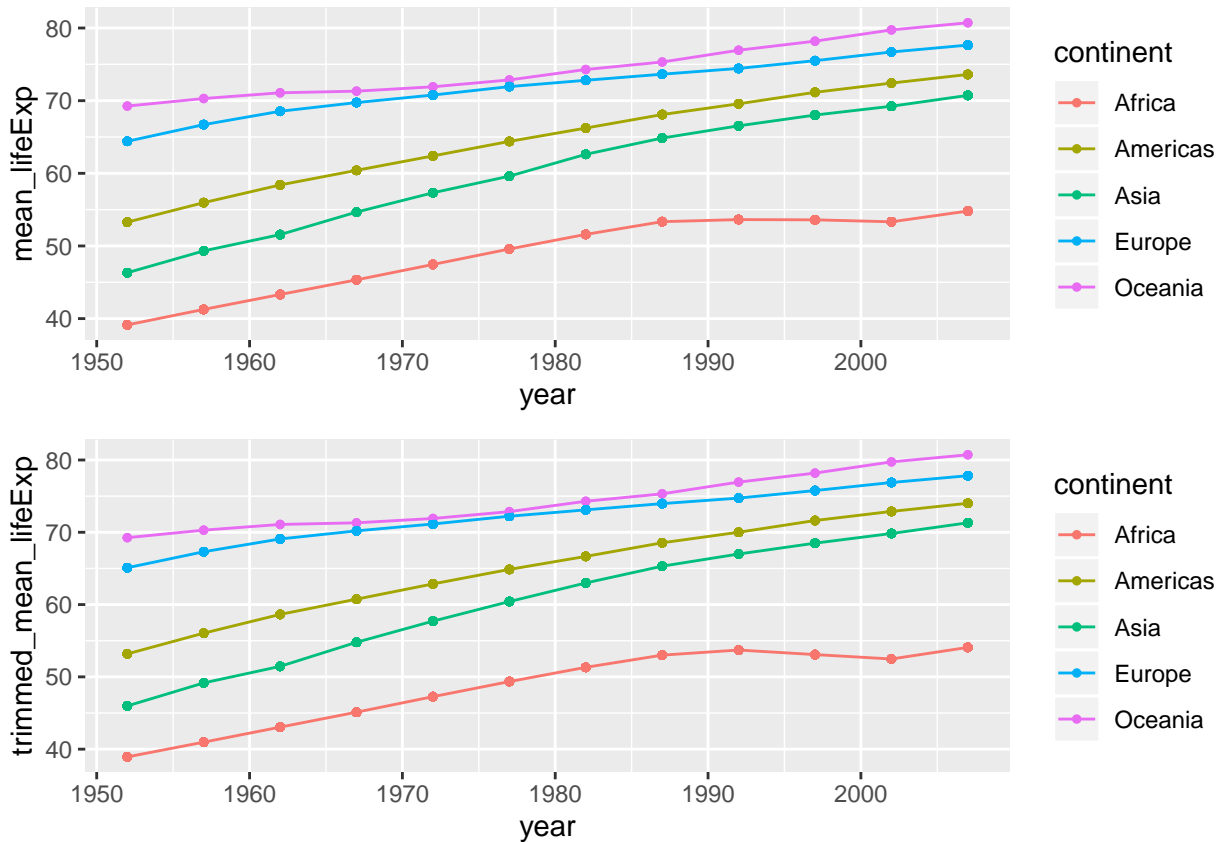
p2<-gapminder %>%
  select (lifeExp, year, continent) %>%
  group_by(continent, year) %>%
  mutate (mean_lifeExp = mean(lifeExp)) %>%
  mutate (trimmed_mean_lifeExp = mean(lifeExp, trim = 0.1)) %>%
```

```

select (year, continent, mean_lifeExp, trimmed_mean_lifeExp) %>%
  ggplot(aes(x = year, y = trimmed_mean_lifeExp, col = continent))+
  geom_point (size = 1) +
  geom_line (size = 0.5)

grid.arrange(p1, p2, nrow = 2)

```



## Task Option 5

How is life expectancy changing over time on different continents?

The mean of life expectancy of each continents are calculated and compared as follow:

```

gapminder %>%
  select (lifeExp, year, continent) %>%
  group_by(continent, year) %>%
  mutate (mean_lifeExp = mean(lifeExp)) %>%
  select (mean_lifeExp, year, continent) %>%
  ggplot(aes(x = year, y = mean_lifeExp, col = continent)) +
  geom_point (size = 1) +
  geom_line (size = 0.5)

```

