

Chapter 6,7 Review

2017100057 / 이영노

October 15, 2022

Chapter 6

Data visualization by procedure

```
library(dplyr)
```

```
## Warning: 패키지 'dplyr'는 R 버전 4.1.30에서 작성되었습니다
```

```
##
```

```
## 다음의 패키지를 부착합니다: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
```

```
##
```

```
##     filter, lag
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
##     intersect, setdiff, setequal, union
```

```
library(ggplot2)
```

```
## Warning: 패키지 'ggplot2'는 R 버전 4.1.30에서 작성되었습니다
```

```
library(dslabs)
```

```
## Warning: 패키지 'dslabs'는 R 버전 4.1.30에서 작성되었습니다
```

```
data(murders)
```

- example1 by procedure

```
data(murders) # load data murders from dslabs package  
murders %>% head
```

```
##      state abb region population total  
## 1    Alabama  AL   South    4779736   135  
## 2     Alaska  AK   West     710231    19  
## 3   Arizona  AZ   West    6392017   232  
## 4  Arkansas  AR   South    2915918    93  
## 5 California  CA   West   37253956  1257  
## 6 Colorado  CO   West    5029196    65
```

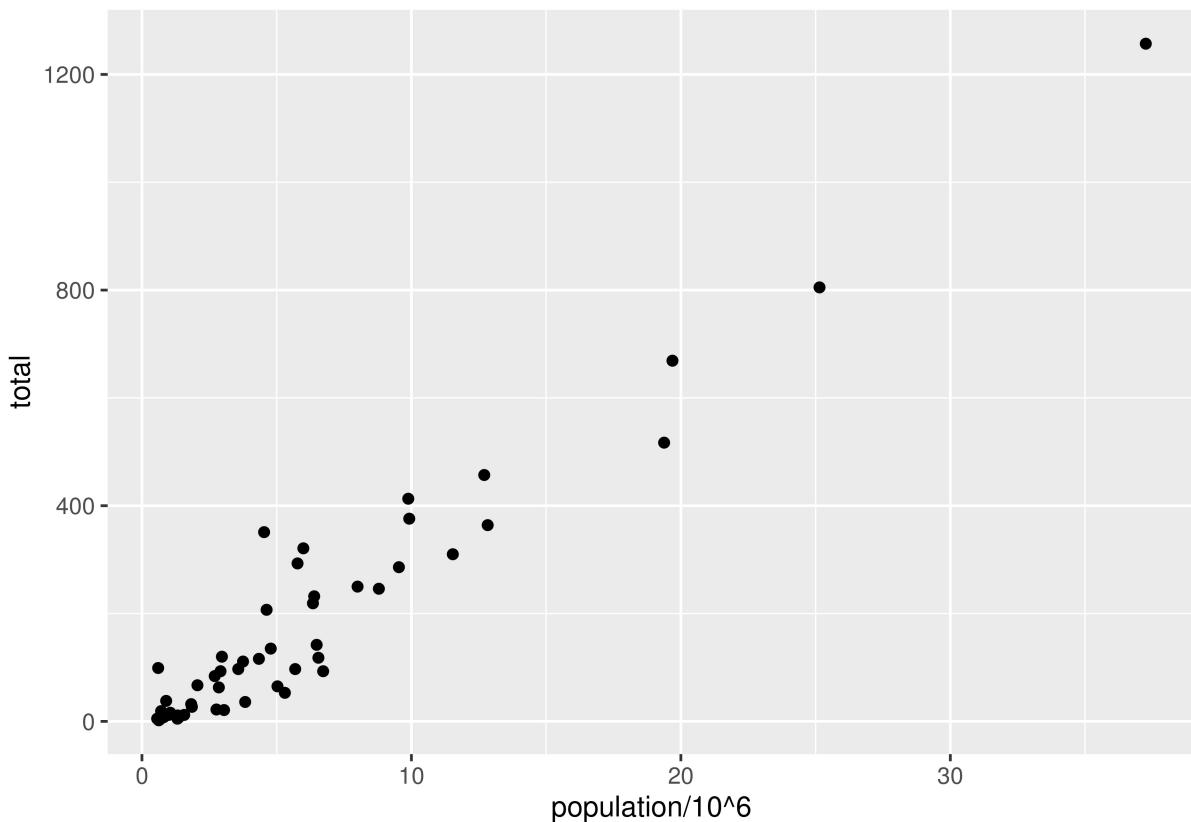
```

murders %>% str

## 'data.frame':   51 obs. of  5 variables:
## $ state      : chr "Alabama" "Alaska" "Arizona" "Arkansas" ...
## $ abb        : chr "AL" "AK" "AZ" "AR" ...
## $ region     : Factor w/ 4 levels "Northeast","South",...: 2 4 4 2 4 4 1 2 2 2 ...
## $ population: num  4779736 710231 6392017 2915918 37253956 ...
## $ total      : num  135 19 232 93 1257 ...

ggplot(murders,aes(x = population/10^6, y = total)) + # aes: variables mapped
  geom_point() # + sign is for adding LAYERS

```

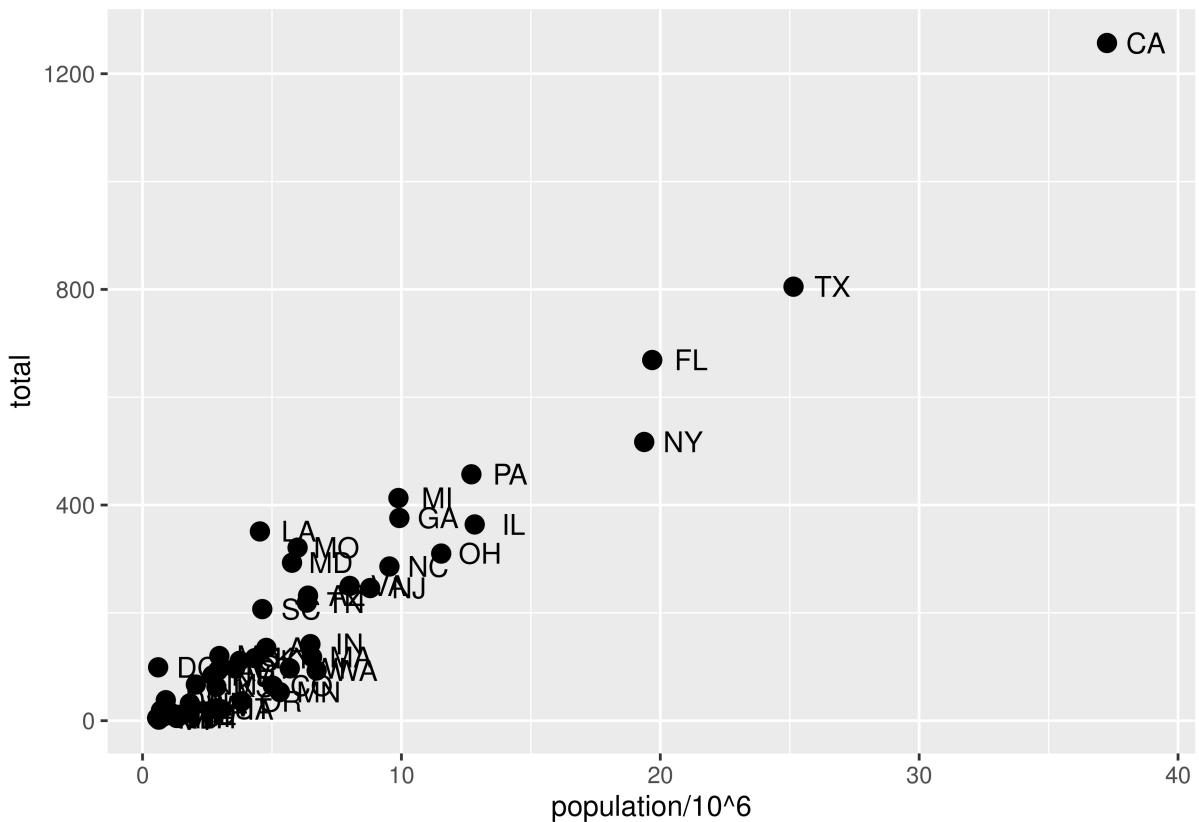


Or we can write like this directly by layers

```

ggplot(murders) +
  geom_point(aes(x = population/10^6, y = total), size = 3) +
  geom_text(aes(population/10^6, total, label = abb),nudge_x = 1.5) # 'label' inside aes() fct

```

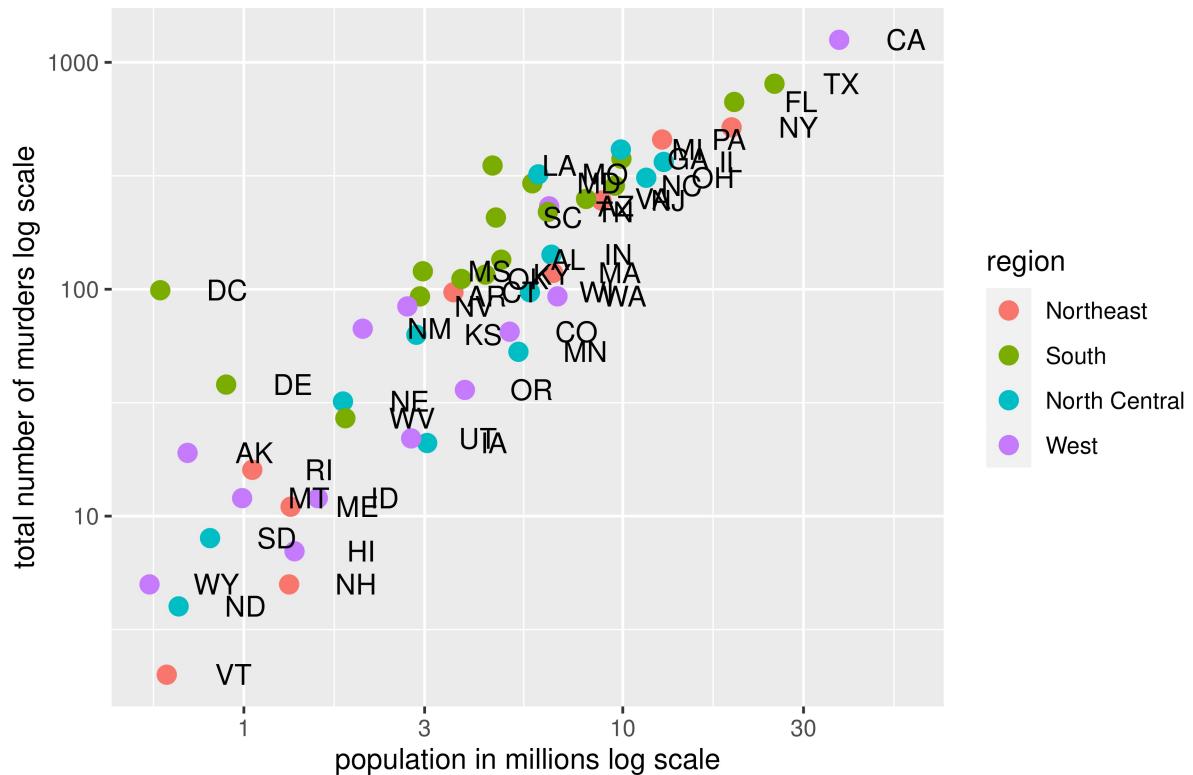


```
# if label=abb(local variable) outside aes(), it gives an error.
# aes() receives mapping options.
# Specially here, aes is local mapping
# size is not mapping so it should be outside aes() fct.
```

OR we can write it in a simple additive code.

```
# 'label' outside since ggplot is global
murders %>% ggplot(aes(population/10^6, total, label=abb)) +
  geom_point(size=3, aes(col=region)) +
  geom_text(nudge_x=log10(1.5))+
  geom_text(x=10,y=800,label = "Hello there!") +
  scale_x_continuous(trans="log10")+
  scale_y_continuous(trans="log10")+xlab("population in millions log scale")+
  ylab("total number of murders log scale")+ggtitle("US gun murders in 2010")
```

US gun murders in 2010



IN SUMMARY,

- POINT 1: global and local mapping에 따라 aesthetic option 다른(label)
- POINT 2: aesthetic option 여부 ~ size는 mapping이 아니라 그냥 option임
- example2: add data not in aes mapping.

```
r <- murders %>% summarize(rate=sum(total)/sum(population)*10^6) %>% pull(rate)
r
```

```
## [1] 30.34555
```

```
murders %>%
ggplot(aes(population/10^6, total, label = abb))+  
scale_x_log10() + scale_y_log10() +  
geom_point(aes(col=region), size=3)+  
geom_abline(intercept=log10(r))
```

