

데이터과학을 위한 **R**프로그래밍

4주차. 데이터시각화



이혜선 교수

포항공과대학교 산업경영공학과



목차

4주차. 데이터시각화

1차시

R 그래픽 I (히스토그램)

2차시

R 그래픽 II (상자그림, 산점도)

3차시

R 그래픽 III (ggplot2 활용)

4차시

R 그래픽 IV (공간지도분석)



4주차

3차시

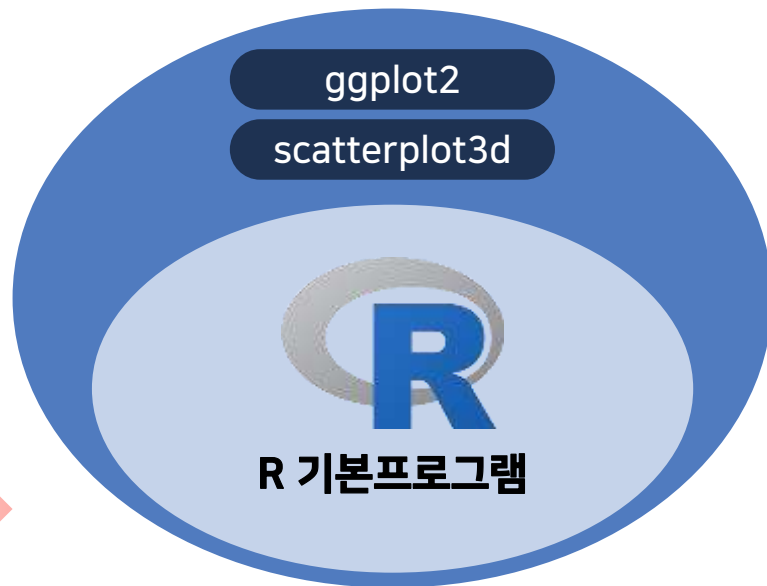
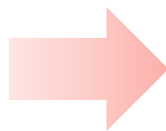
R 그래픽 III (ggplot2 활용)

● R 그래픽 : ggplot2 패키지

☑ 추가패키지 설치 (install.packages)

프로그램 편집 창

```
# Graphic using ggplot2  
  
# set working directory  
setwd("D:/tempstore/moocr")  
  
# library  
install.packages('ggplot2')  
library(ggplot2)
```



● R 그래픽 : ggplot2 패키지 구조

Grammar of graphics

(1) ggplot()이라는 기본 함수

+

(2-1) Layers : aes (Aesthetic) : 데이터를 어떻게 넣을지

(2-2) Layer : geom (Geometric objects) : point(점), line(선) 등

(2-3) Layer : coord (coordinate system)

- 1. **scale+coordinate system**은 그림을 그릴 때 캔버스로 생각하면 됨
- 2. 그 위에 **data+mapping+geom**을 추가
- 3. geom(기하학적 요소): **geom_point**, **geom_smooth**등을 이미 그려진 산점도에 추가하여(incremental) 그릴수 있음

R 그래픽 : ggplot2 그래프

✓ `ggplot(데이터이름, aes(x=x축변수, y=y축변수, color=factor변수, shape=factor변수))+geom_point(size=3)`

```
# lec4_3.r
# Graphic using ggplot2

# set working directory
setwd("D:/tempstore/mooocr")

# library
install.packages('ggplot2')
library(ggplot2)

# Read in R : autmpg data
car<-read.csv("autmpg.csv")
head(car)
str(car)

# subset of car : cyl (4,6,8)
car1<-subset(car, cyl==4 | cyl==6 | cyl==8)
attach(car1)

# 1. ggplot2 for scatterplot
# Color and shape display by factor (categorical variable)
# check the variable type(integer, numeric, factor) and define it
str(car1)
car1$cyl<-as.factor(car1$cyl)

# Now, you can draw one of plot using ggplot
par(mfrow = c(1, 1))
ggplot(car1, aes(x=wt, y=disp, color=cyl, shape=cyl)) +
  geom_point(size=3, alpha=0.6)
```

Step0 : 분석을 위한 설정(library, package, setwd)

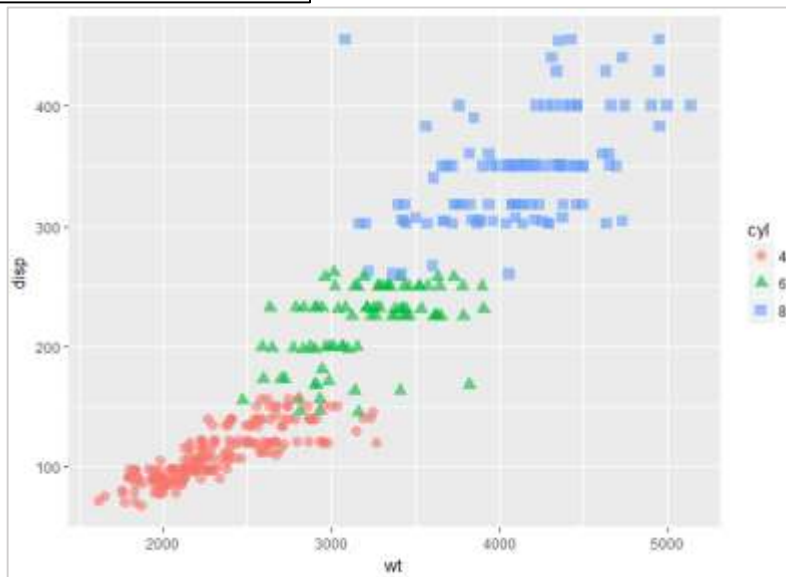
Step1 : 데이터핸들링
(불러들이기, subset, 변수값정의)

Step2 : 데이터탐색(그래픽)

● R 그래픽 : ggplot2 그래프

☑ ggplot2 : scatterplot (wt와 disp) by group (cyl)

```
par(mfrow = c(1, 1))  
ggplot(car1, aes(x=wt, y=disp, color=cyl, shape=cyl)) +  
  geom_point(size=3, alpha=0.6)
```



● R 그래픽 : ggplot2 그래프

☑ ggplot2 : ggplot객체들의 설명

```
# Color and shape display by factor (categorical variable)
# check the variable type(integer, numeric, factor) and def
str(car1)
car1$cyl<-as.factor(car1$cyl)
# Now, you can draw one of plot using ggplot
ggplot(car1, aes(x=wt, y=disp, color=cyl, shape=cyl)) +
  geom_point(size=3)
```

```
ggplot(car1, aes(x=wt, y=disp, color=cyl, shape=cyl))
```

1. ggplot함수에 데이터는 car1을 이용하고, x축에는 wt(차의 무게)를, y축에는 disp(배기량)의 산점도를 그리고, 점의 색상은 cyl(실린더 수)로 표현한다

```
geom_point(size=3, alpha=0.6)
```

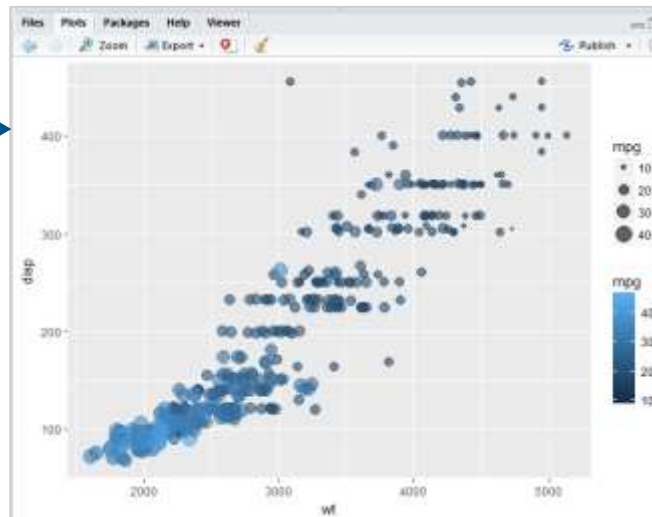
2. geom_point는 size=3(숫자 클수록 점 크기가 커짐)

● R 그래픽 : ggplot2 그래프

☑ mpg의 크기를 표시한 그래프

```
# mapping (continuous variable : mpg) on the scatterplot for wt and disp  
ggplot(car1, aes(x=wt, y=disp, color=mpg, size=mpg)) +  
  geom_point(alpha=0.6)
```

설명 : 차의 무게와 배기량의 산점도에
연비의 높고 낮음을 원의 사이즈와 색
으로 표시한 그래프



● R 그래픽 : ggplot2 구조

ggplot의 기본

ggplot : 새로운 ggplot을 생성

aes : aesthetic mapping을 구성 (데이터, 그래프구조)

qplot: 즉석 그림

geom 함수군

geom_abline, **geom_hline**, **geom_vline**

geom_bar

geom_point

geom_boxplot

geom_map

geom_smooth, **stat_smooth**

geom (geometric) 함수군

geom메뉴

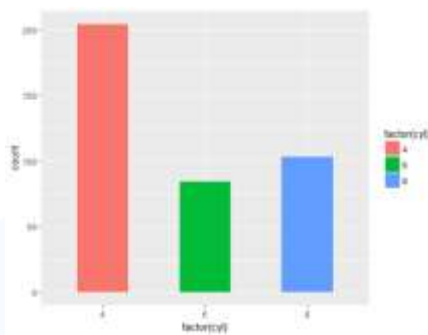
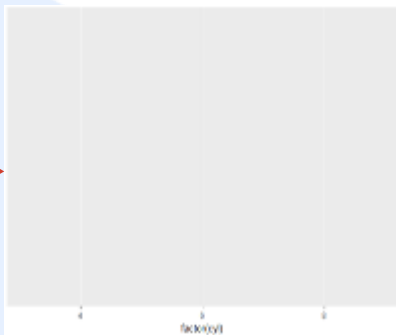
help.search("geom")

ggplot2::geom_abline	Reference lines: horizontal, vertical, and diagonal
ggplot2::geom_bar	Bars charts
ggplot2::geom_bin2d	Heatmap of 2d bin counts
ggplot2::geom_blank	Draw nothing
ggplot2::geom_boxplot	A box and whiskers plot (in the style of Tukey)
ggplot2::geom_contour	2d contours of a 3d surface
ggplot2::geom_count	Count overlapping points
ggplot2::geom_density	Smoothed density estimates
ggplot2::geom_density_2d	Contours of a 2d density estimate
ggplot2::geom_dotplot	Dot plot
ggplot2::geom_errorbarh	Horizontal error bars
ggplot2::geom_hex	Hexagonal heatmap of 2d bin counts
ggplot2::geom_freqpoly	Histograms and frequency polygons
ggplot2::geom_jitter	Jittered points
ggplot2::geom_crossbar	Vertical intervals: lines, crossbars & errorbars
ggplot2::geom_map	Polygons from a reference map
ggplot2::geom_path	Connect observations
ggplot2::geom_point	Points
ggplot2::geom_polygon	Polygons
ggplot2::geom_qq	A quantile-quantile plot

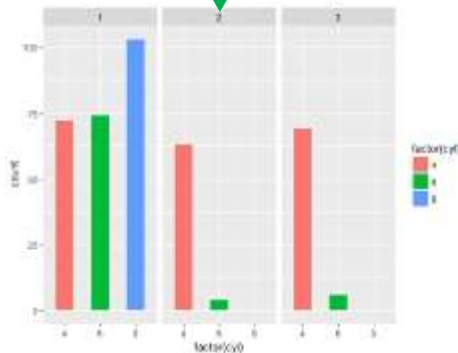
R 그래픽 : ggplot2 그래프

☑ geom_bar 을 이용한 단계별 그래프 설명

```
p1<-ggplot(car1, aes(factor(cyl), fill=factor(cyl)))  
p1  
# barplot define  
p1<-p1 + geom_bar(width=.5)  
p1  
# output by origin(1,2,3)  
p1<-p1 + facet_grid(. ~ origin)  
p1
```



1. grid를 그림-cyl에 따라서
2. geom_bar을 이용한 cyl 빈도의 막대그래프
3. cyl의 bar chart를 변수 'origin'에 따라 그림

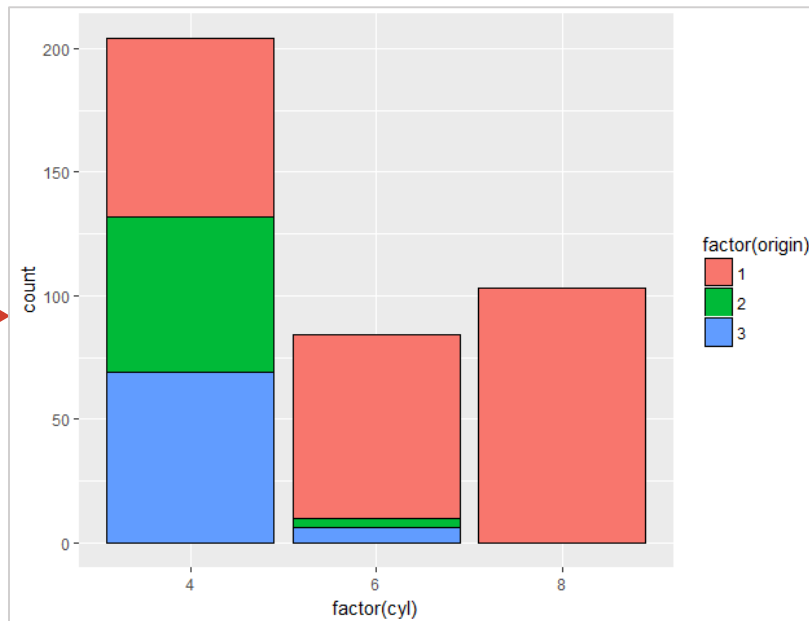


```
# 2-1. same plot with the above  
ggplot(car1, aes(factor(cyl), fill=factor(cyl)))+ geom_bar(width=.5)+ facet_grid(. ~ origin)
```

● R 그래픽 : ggplot2 그래프

☑ geom_bar 을 이용한 누적 막대그래프

```
# 5-2-2 : geom_bar : aesthetic mapping (4,6,8 cyl)  
p <- ggplot(data=car1, aes(factor(cyl)))  
p + geom_bar(aes(fill=factor(origin)), colour="black")
```



ggplot2 그래픽 사이트

✓ Ggplot2 활용

🔍 <https://www.rstudio.com/wp-content/uploads/2015/03/ggplot2-cheatsheet.pdf>

