STATSOFT2021-ch02v2

패키지 관련

```
library(dplyr)
```

```
##
## 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)
library(gridExtra)
```

```
## Warning: 패키지 'gridExtra'는 R 버전 4.1.2에서 작성되었습니다
```

```
##
## 다음의 패키지를 부착합니다: 'gridExtra'
```

```
## The following object is masked from 'package:dplyr':
##
##     combine
```

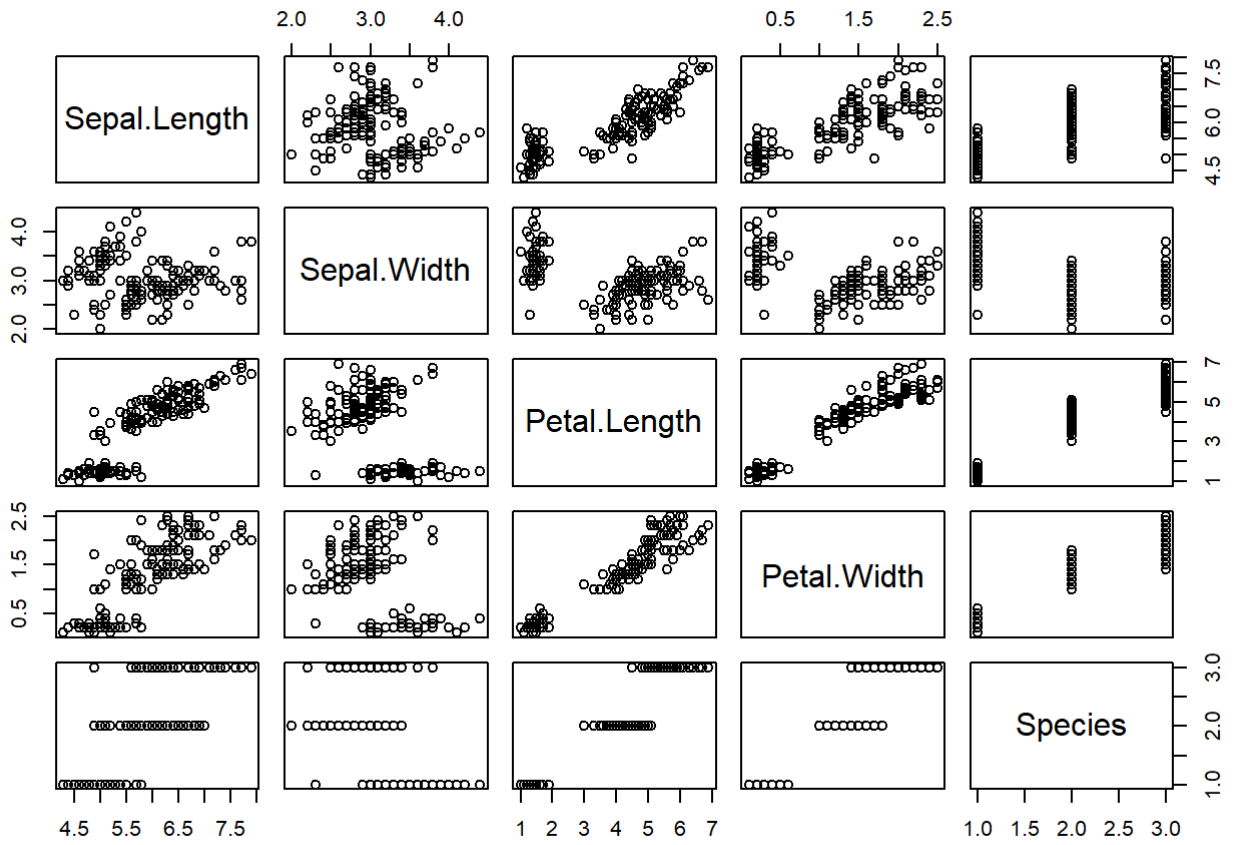
```
str(iris)
```

```
## 'data.frame':    150 obs. of  5 variables:
##  $ Sepal.Length: num  5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...
##  $ Sepal.Width : num  3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...
##  $ Petal.Length: num  1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...
##  $ Petal.Width : num  0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...
##  $ Species     : Factor w/ 3 levels "setosa","versicolor",..: 1 1 1 1 1 1 1 1 1 1 ...
```

```
head(iris, 10)
```

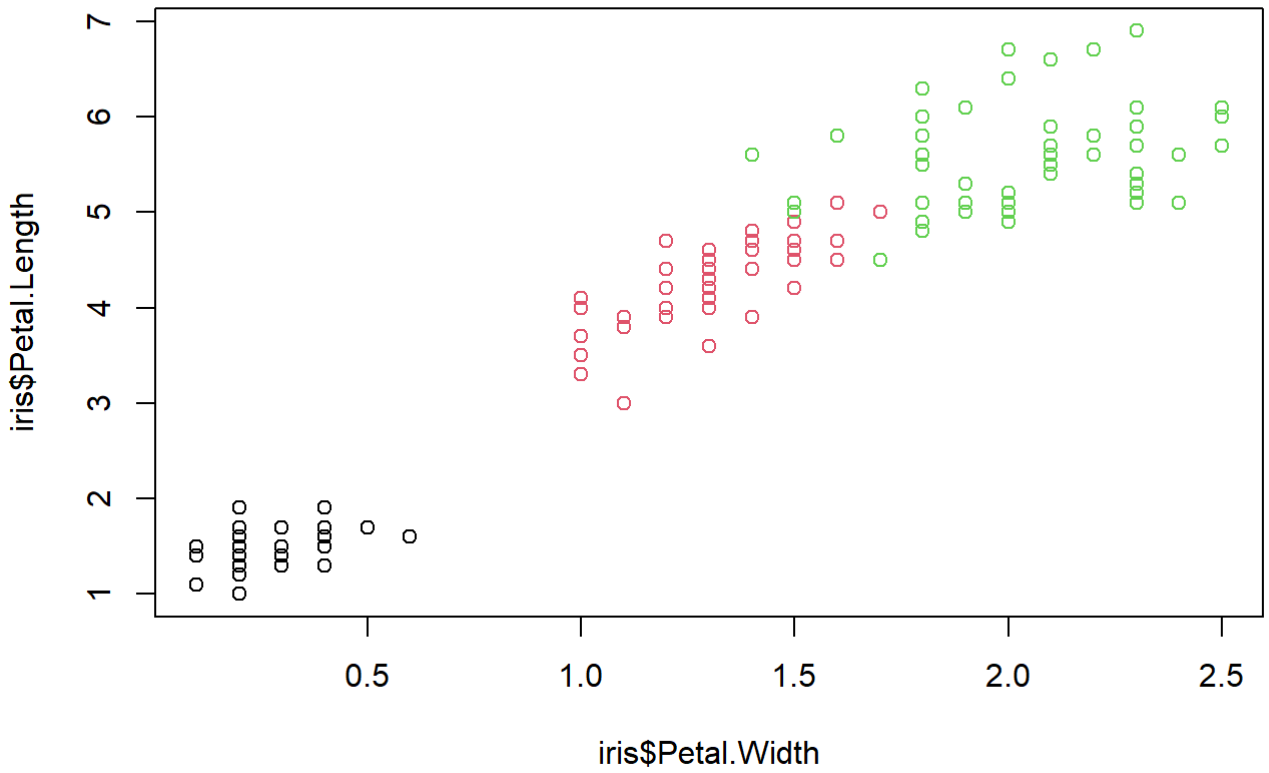
```
##    Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 1           5.1         3.5          1.4         0.2  setosa
## 2           4.9         3.0          1.4         0.2  setosa
## 3           4.7         3.2          1.3         0.2  setosa
## 4           4.6         3.1          1.5         0.2  setosa
## 5           5.0         3.6          1.4         0.2  setosa
## 6           5.4         3.9          1.7         0.4  setosa
## 7           4.6         3.4          1.4         0.3  setosa
## 8           5.0         3.4          1.5         0.2  setosa
## 9           4.4         2.9          1.4         0.2  setosa
## 10          4.9         3.1          1.5         0.1  setosa
```

```
plot(iris)
```



read.csv

```
plot(iris$Petal.Width, iris$Petal.Length, col=iris$Species)
```




```
tips <- read.csv('C:WWWORKWWdataWWtips.csv', stringsAsFactors=TRUE)
str(tips)
```

```
## 'data.frame':    244 obs. of  7 variables:
##  $ total_bill: num  17 10.3 21 23.7 24.6 ...
##  $ tip       : num  1.01 1.66 3.5 3.31 3.61 4.71 2 3.12 1.96 3.23 ...
##  $ sex       : Factor w/ 2 levels "Female","Male": 1 2 2 2 1 2 2 2 2 2 ...
##  $ smoker    : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1 1 1 1 1 ...
##  $ day       : Factor w/ 4 levels "Fri","Sat","Sun",..: 3 3 3 3 3 3 3 3 3 3 ...
##  $ time      : Factor w/ 2 levels "Dinner","Lunch": 1 1 1 1 1 1 1 1 1 1 ...
##  $ size      : int  2 3 3 2 4 4 2 4 2 2 ...
```

```
head(tips, 10)
```

```
##    total_bill  tip    sex smoker day   time size
## 1       16.99 1.01 Female     No Sun Dinner    2
## 2       10.34 1.66   Male     No Sun Dinner    3
## 3       21.01 3.50   Male     No Sun Dinner    3
## 4       23.68 3.31   Male     No Sun Dinner    2
## 5       24.59 3.61 Female     No Sun Dinner    4
## 6       25.29 4.71   Male     No Sun Dinner    4
## 7        8.77 2.00   Male     No Sun Dinner    2
## 8       26.88 3.12   Male     No Sun Dinner    4
## 9       15.04 1.96   Male     No Sun Dinner    2
## 10      14.78 3.23   Male     No Sun Dinner    2
```

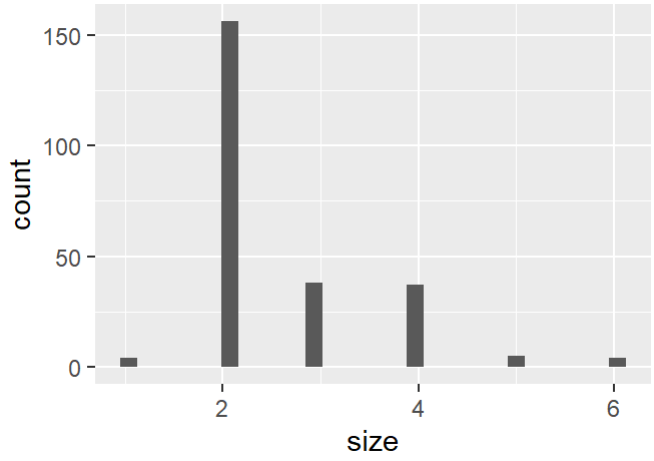
```
summary(tips)
```

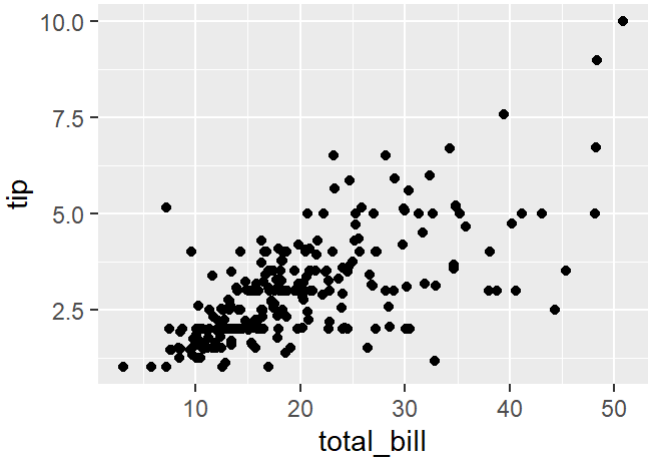
```
##    total_bill        tip             sex      smoker      day         time
##  Min.   : 3.07   Min.   : 1.000   Female: 87   No :151   Fri :19   Dinner:176
##  1st Qu.:13.35   1st Qu.: 2.000   Male  :157   Yes: 93   Sat :87   Lunch : 68
##  Median :17.80   Median : 2.900                          Sun :76
##  Mean   :19.79   Mean   : 2.998                          Thur:62
##  3rd Qu.:24.13   3rd Qu.: 3.562
##  Max.   :50.81   Max.   :10.000
##       size
##  Min.   :1.00
##  1st Qu.:2.00
##  Median :2.00
##  Mean   :2.57
##  3rd Qu.:3.00
##  Max.   :6.00
```

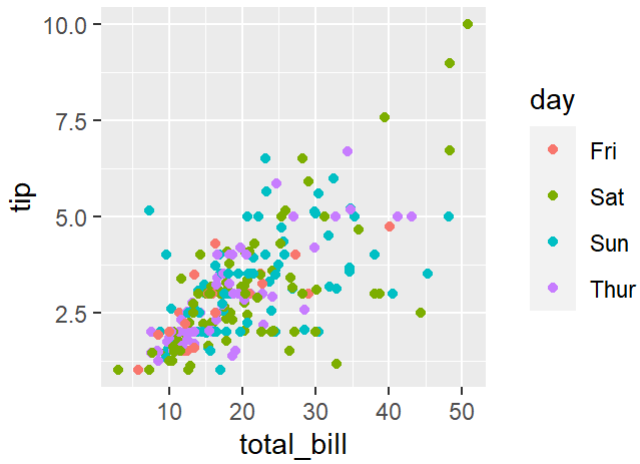
dplyr와 ggplot 사용 예

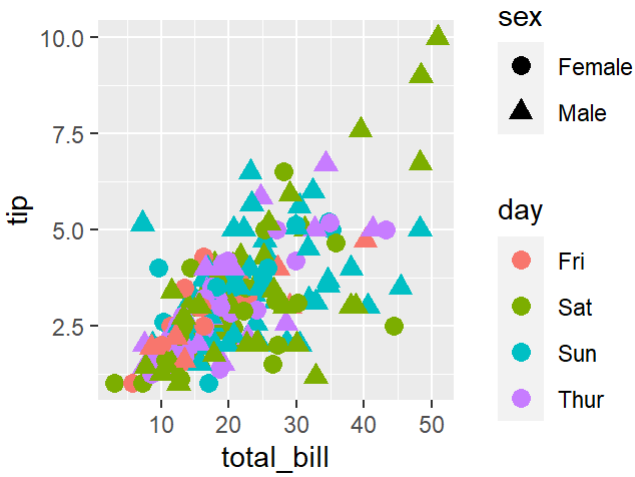
```
g1 <- tips %>%
  ggplot(aes(size)) + geom_histogram()
g2 <- tips %>%
  ggplot(aes(total_bill, tip)) + geom_point()
g3 <- tips %>%
  ggplot(aes(total_bill, tip)) + geom_point(aes(col = day))
g4 <- tips %>%
  ggplot(aes(total_bill, tip)) +
  geom_point(aes(col = day, pch=sex), size=3)
grid.arrange(grobs=list(g1, g2, g3, g4), nr=2)
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```


count
150
100
50
0
2
4
6
size


tip
10.0
7.5
5.0
2.5
10
20
30
40
50
total_bill


tip
10.0
7.5
5.0
2.5
10
20
30
40
50
total_bill
day
Fri
Sat
Sun
Thur


tip
10.0
7.5
5.0
2.5
10
20
30
40
50
total_bill
sex
Female
Male
day
Fri
Sat
Sun
Thur