

## hw3\_2229027

JiminLee\_2229027

2023-09-23

```
install.packages('GGally', repos = "http://cran.us.r-project.org")

## Warning in readRDS(dest): lzma decoder corrupt data

##
## The downloaded binary packages are in
## /var/folders/zl/ljg24f_15m93w7f2rt14c2n80000gn/T//Rtmpj0UgZn/downloaded_p
ackages

library(GGally)

## Loading required package: ggplot2

## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg      ggplot2

library(tidyverse)

## — Attaching packages
## —————
## tidyverse 1.3.2 —

## ✓ tibble 3.2.1      ✓ dplyr 1.1.3
## ✓ tidyr 1.2.1      ✓ stringr 1.4.0
## ✓ readr 2.1.2      ✓ forcats 0.5.2
## ✓ purrr 1.0.2

## — Conflicts ————— tidyverse_conflict
s() —
## ✗ dplyr::filter() masks stats::filter()
## ✗ dplyr::lag() masks stats::lag()

library(ggmosaic)

##
## Attaching package: 'ggmosaic'
##
## The following object is masked from 'package:GGally':
##
## happy

data(mpg)
head(mpg, 5)
```

```
## # A tibble: 5 × 11
##   manufacturer model displ  year   cyl trans      drv   cty   hwy fl
class
##   <chr>          <chr> <dbl> <int> <int> <chr>      <chr> <int> <int> <chr>
<chr>
## 1 audi          a4      1.8  1999     4 auto(l5)  f       18    29 p
compa...
## 2 audi          a4      1.8  1999     4 manual(m5) f       21    29 p
compa...
## 3 audi          a4      2    2008     4 manual(m6) f       20    31 p
compa...
## 4 audi          a4      2    2008     4 auto(av)  f       21    30 p
compa...
## 5 audi          a4      2.8  1999     6 auto(l5)  f       16    26 p
compa...
```

```
nrow(mpg) # num of obseration : 234
```

```
## [1] 234
```

```
names(mpg)
```

```
## [1] "manufacturer" "model"          "displ"          "year"           "cyl"
```

```
## [6] "trans"         "drv"           "cty"           "hwy"           "fl"
```

```
## [11] "class"
```

```
help(mpg) # 변수의 세부적인 정보를 보기 위함
```

1) mpg 자료중 범주형 변수는 어떤 것들이 있는가?

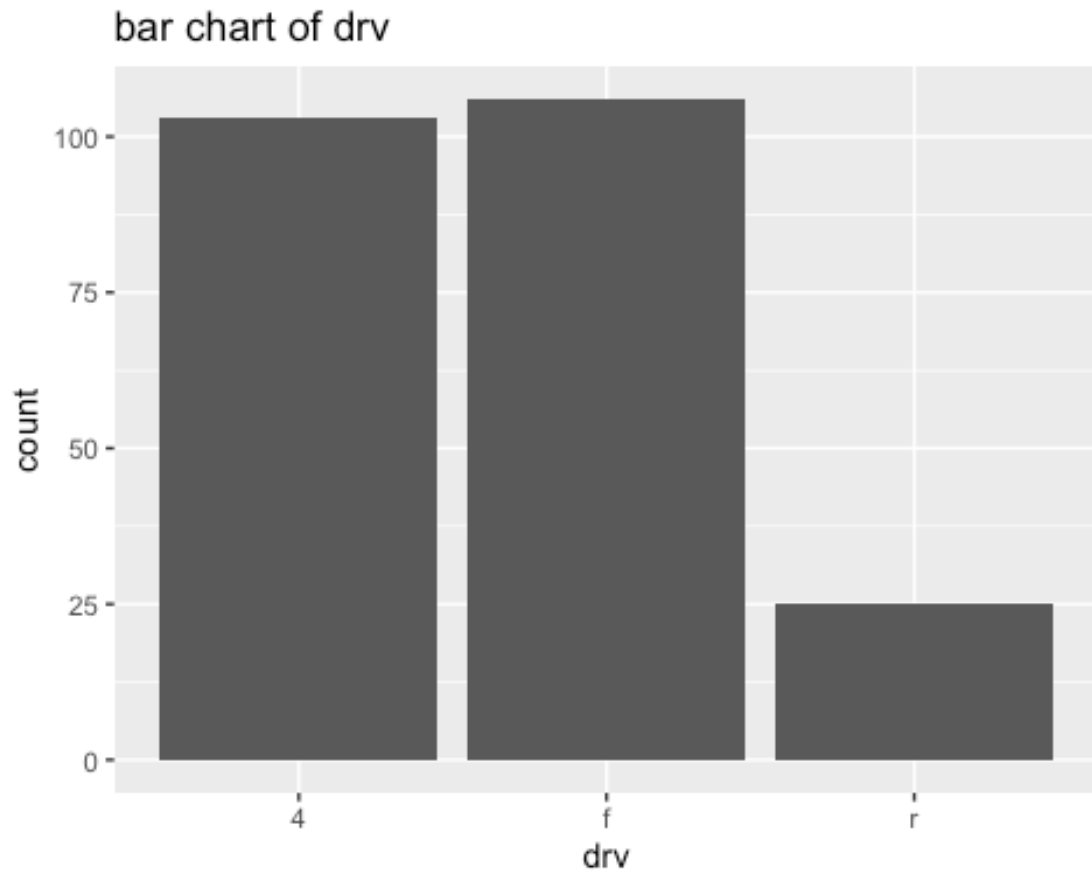
```
str(mpg)
```

```
## tibble [234 × 11] (S3: tbl_df/tbl/data.frame)
## $ manufacturer: chr [1:234] "audi" "audi" "audi" "audi" ...
## $ model       : chr [1:234] "a4" "a4" "a4" "a4" ...
## $ displ       : num [1:234] 1.8 1.8 2 2 2.8 2.8 3.1 1.8 1.8 2 ...
## $ year        : int [1:234] 1999 1999 2008 2008 1999 1999 2008 1999 1999
2008 ...
## $ cyl         : int [1:234] 4 4 4 4 6 6 6 4 4 4 ...
## $ trans       : chr [1:234] "auto(l5)" "manual(m5)" "manual(m6)" "auto(av)"
" ...
## $ drv         : chr [1:234] "f" "f" "f" "f" ...
## $ cty         : int [1:234] 18 21 20 21 16 18 18 16 20 ...
## $ hwy         : int [1:234] 29 29 31 30 26 26 27 26 25 28 ...
## $ fl          : chr [1:234] "p" "p" "p" "p" ...
## $ class       : chr [1:234] "compact" "compact" "compact" "compact" ...
```

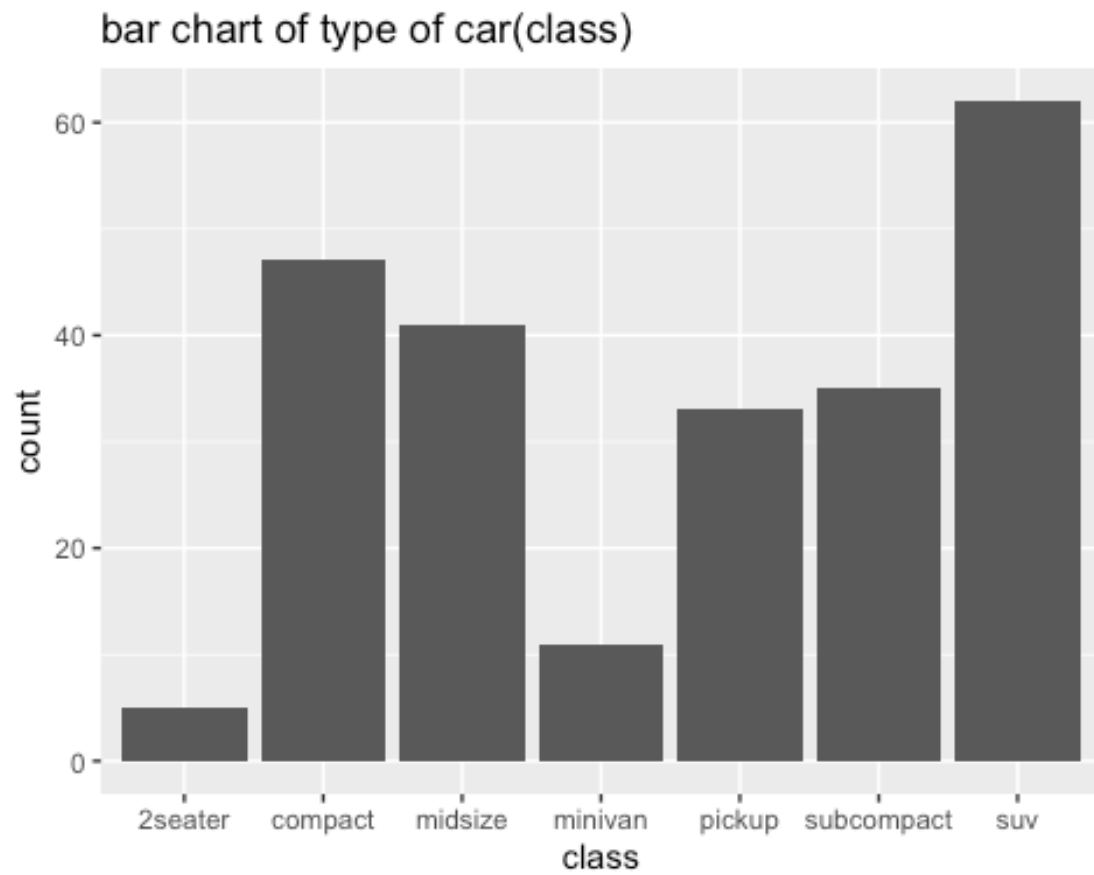
manufacturer, model, trans, drv, fl, class

2) 1)에서 확인한 범주형 변수들 중 *drv*, *class*, *fl* 을 수업시간에 배운 내용을 바탕으로 살펴보시오.

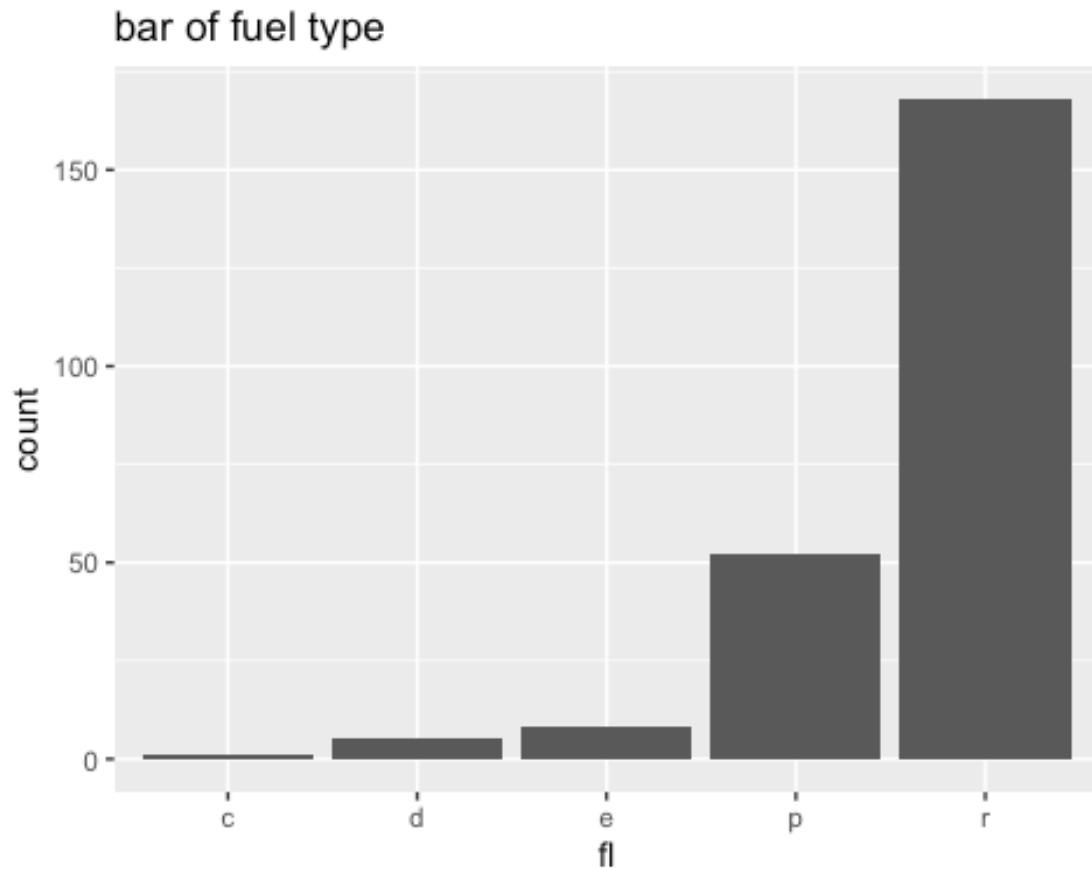
```
ggplot(mpg, aes(drv)) +  
  geom_bar() +  
  ggtitle('bar chart of drv')
```



```
ggplot(mpg, aes(class)) +  
  geom_bar() +  
  ggtitle('bar chart of type of car(class)')
```



```
ggplot(mpg, aes(f1)) +  
  geom_bar() +  
  ggtitle('bar of fuel type')
```

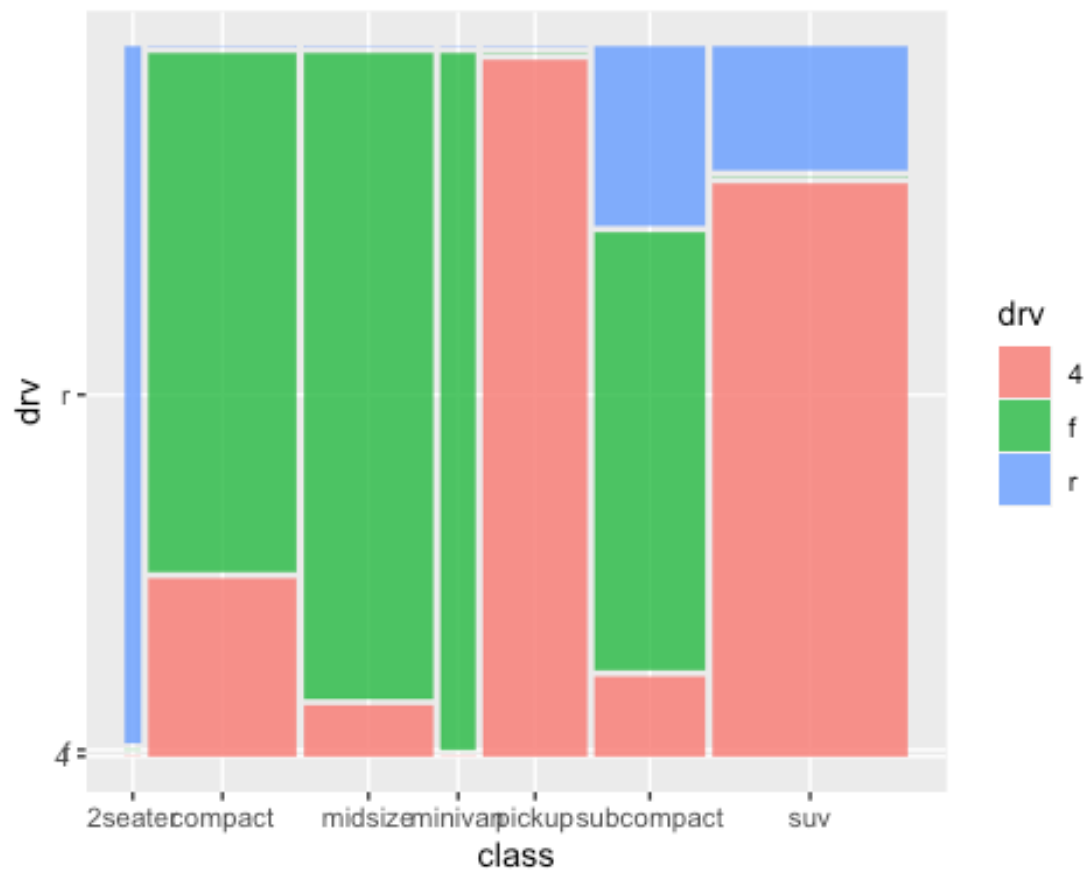


#### 3)

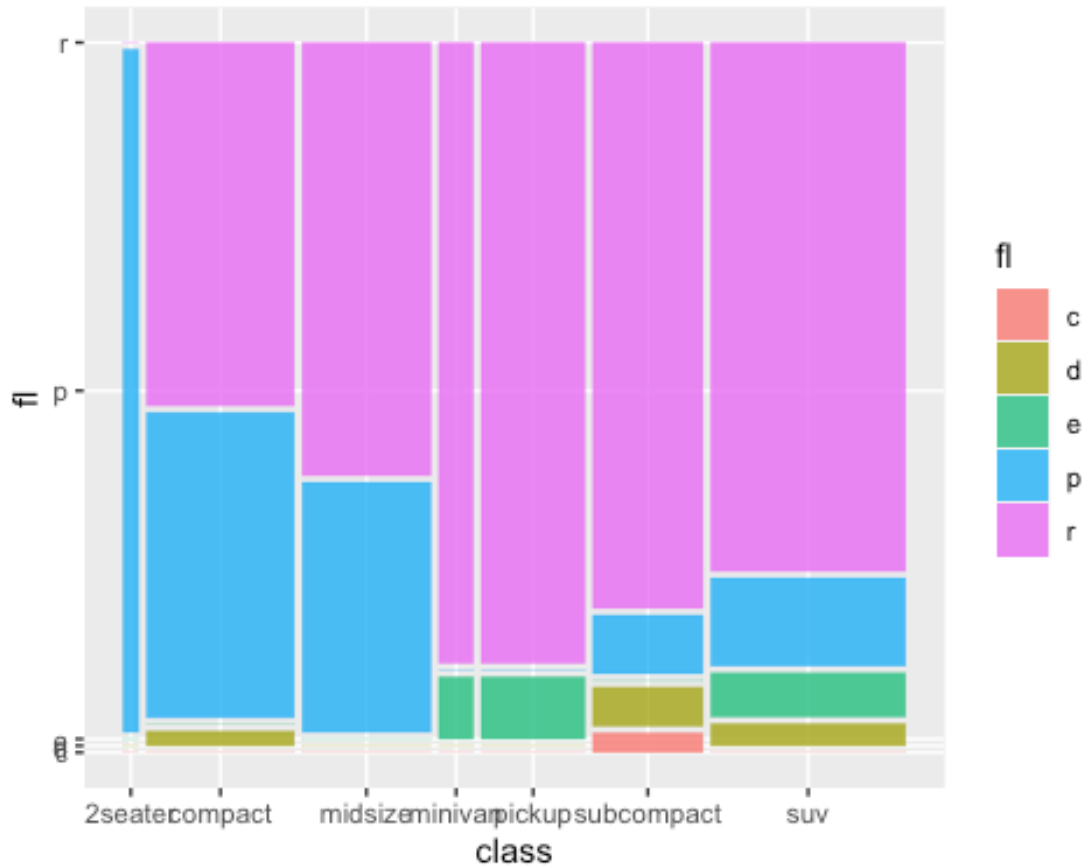
class 별 drv, fl 분포를 살펴보세요.

```
data1 = mpg[c('class', 'drv', 'fl')]
df1 = as.data.frame(data1)
ggplot(df1) + geom_mosaic(aes(x=product(class), fill=drv))

## Warning: `unite_()` was deprecated in tidyr 1.2.0.
## i Please use `unite()` instead.
## i The deprecated feature was likely used in the ggmosaic package.
## Please report the issue at <https://github.com/haleyjeppson/ggmosaic>.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```



```
ggplot(df1) + geom_mosaic(aes(x=product(class), fill=f1))
```



### 2.

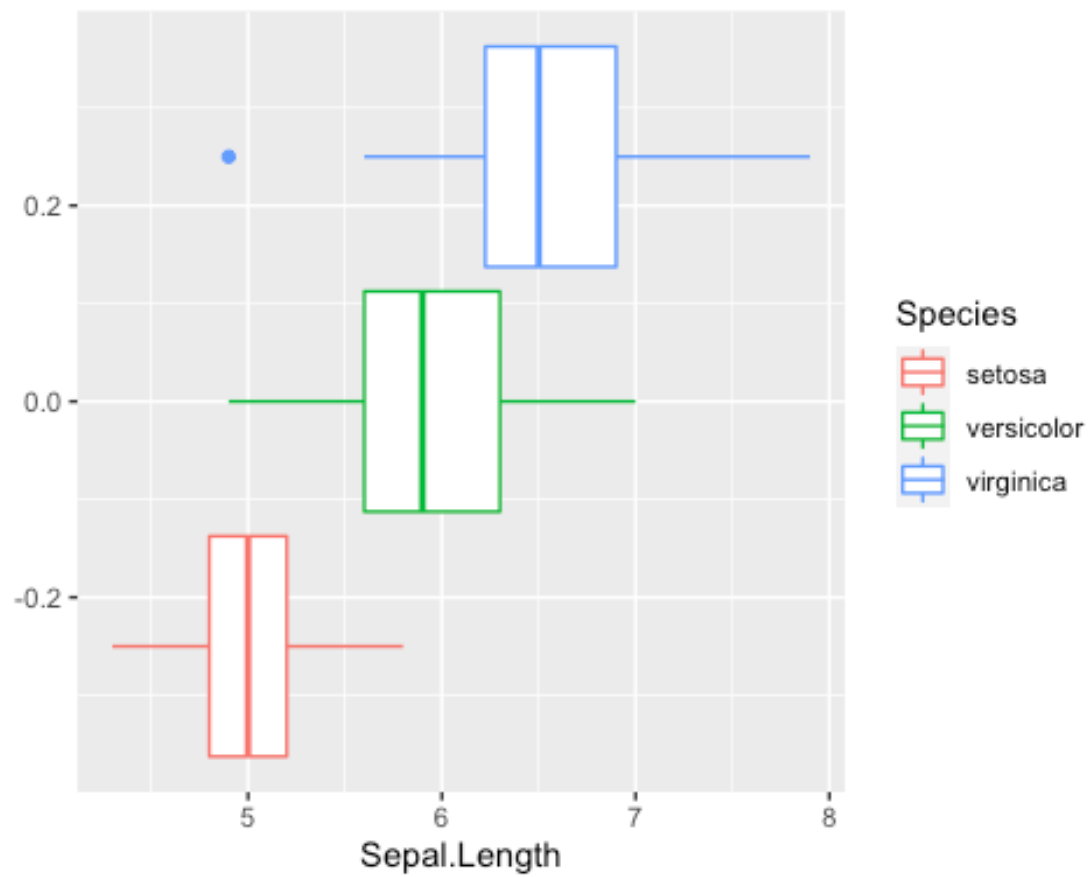
iris 자료는 R 에 내장되어 있는 자료이다.

1) Sepal.Length, Sepal.Width, Petal.Length, Petal.Width 의 분포가 Species 별로 어떻게 다른지 알아보려고 한다. 이에 알맞은 그림을 그리고 살펴보세요.

```
data(iris)
head(iris)

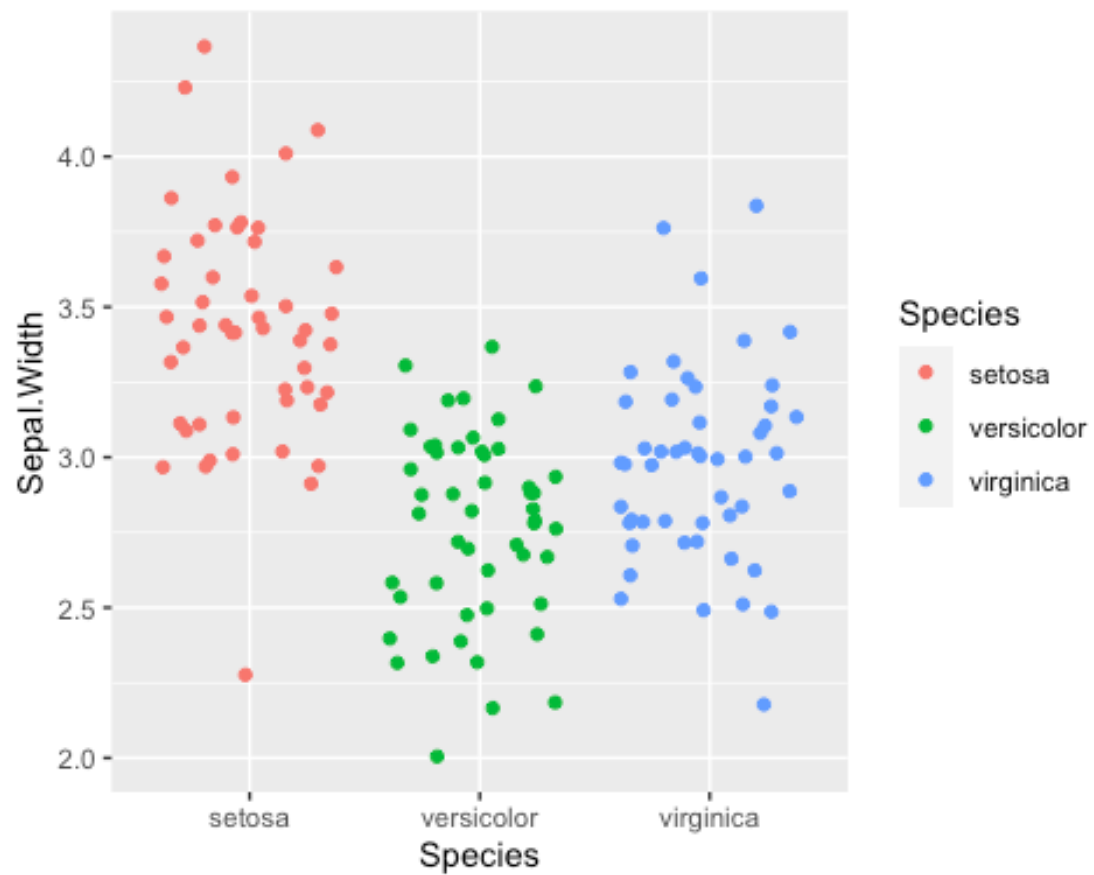
##   Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 1         5.1         3.5         1.4         0.2   setosa
## 2         4.9         3.0         1.4         0.2   setosa
## 3         4.7         3.2         1.3         0.2   setosa
## 4         4.6         3.1         1.5         0.2   setosa
## 5         5.0         3.6         1.4         0.2   setosa
## 6         5.4         3.9         1.7         0.4   setosa

ggplot(iris, aes(Sepal.Length, color=Species)) + geom_boxplot()
```

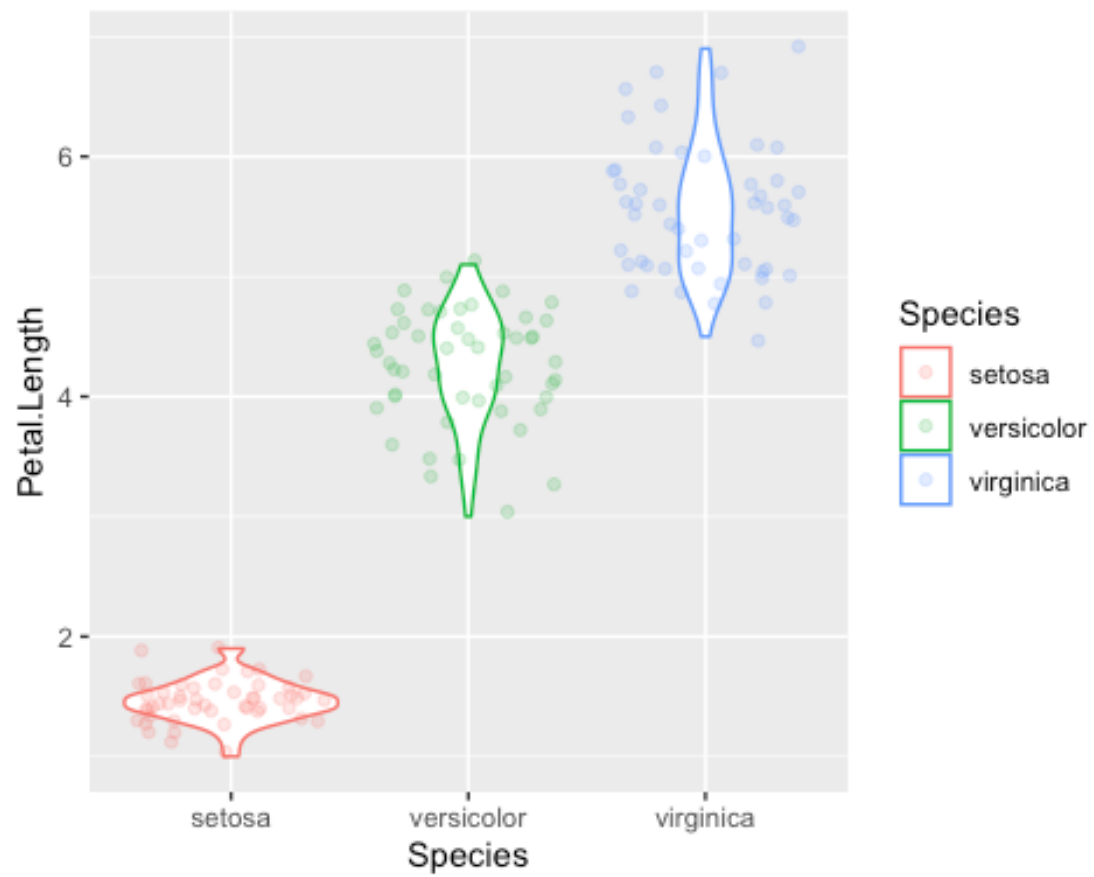


```
ggplot(iris, aes(Species, Sepal.Width, color=Species)) + geom_jitter()
```

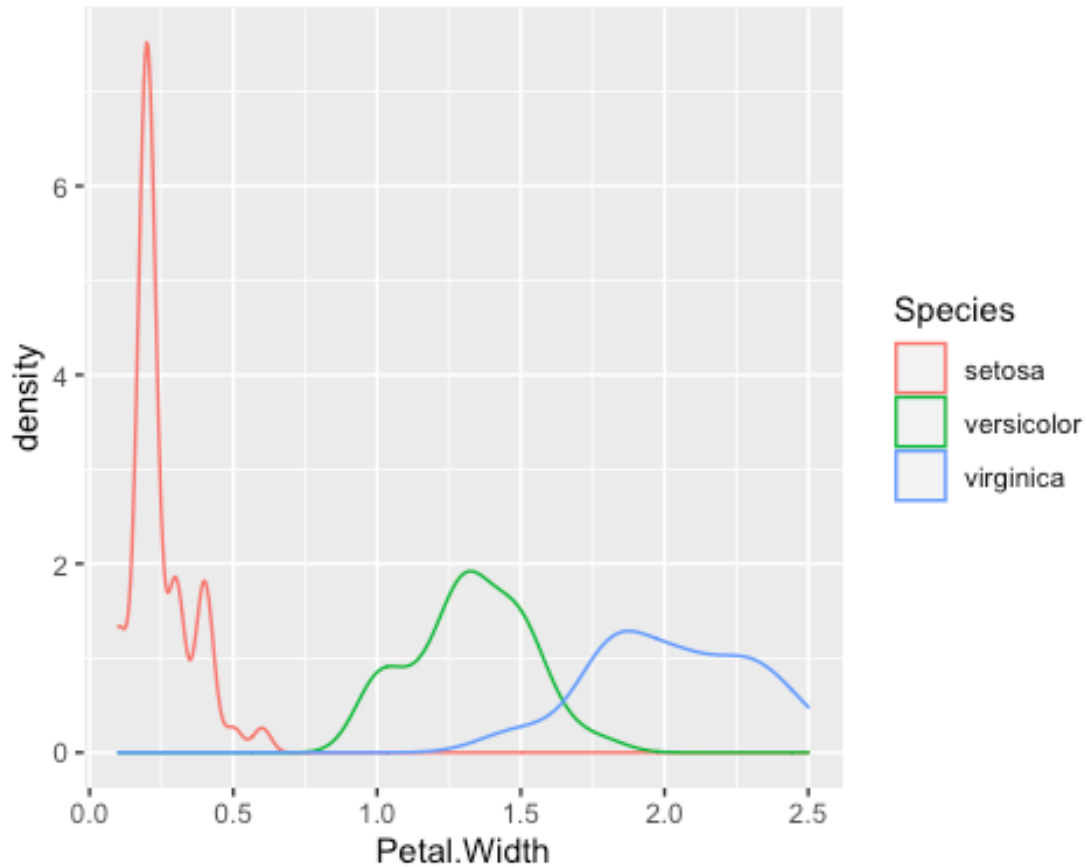




```
ggplot(iris, aes(Species, Petal.Length, color=Species)) + geom_violin() + geom_jitter(alpha=0.2, size=1.5)
```



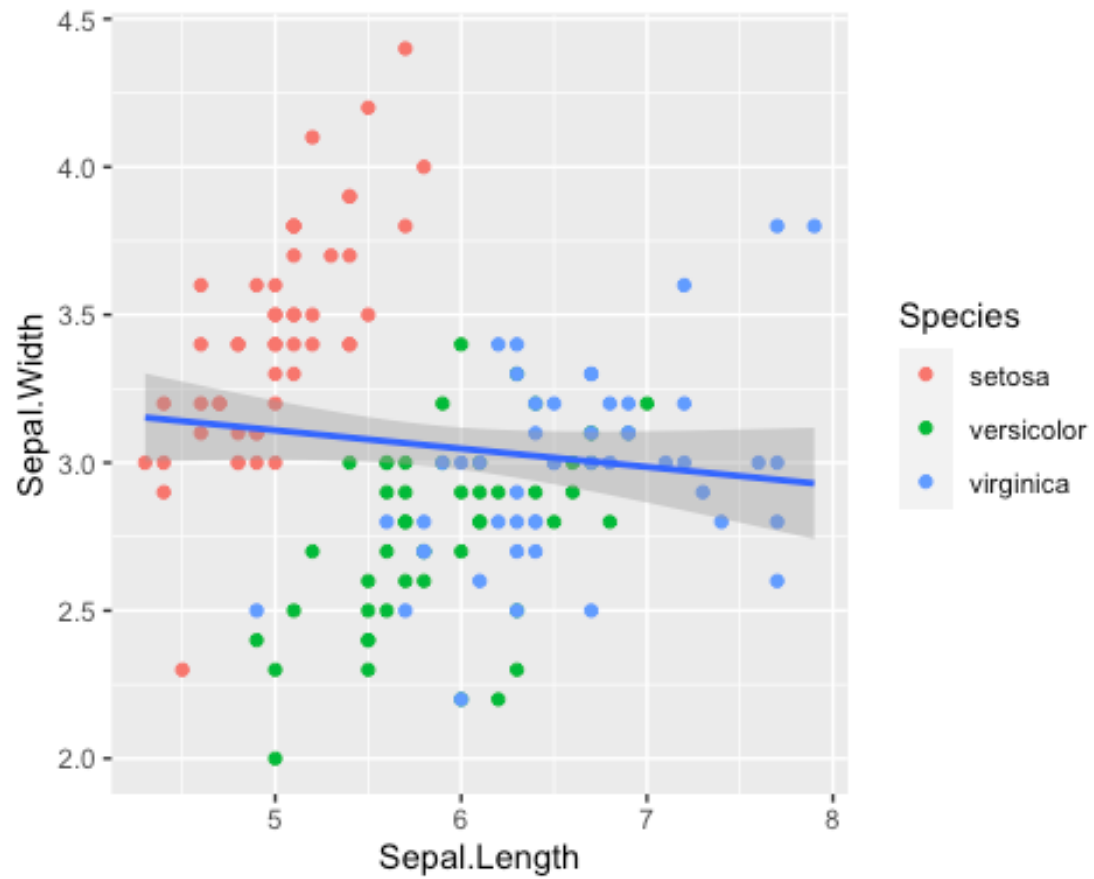
```
ggplot(iris, aes(Petal.Width, color=Species)) + geom_density()
```



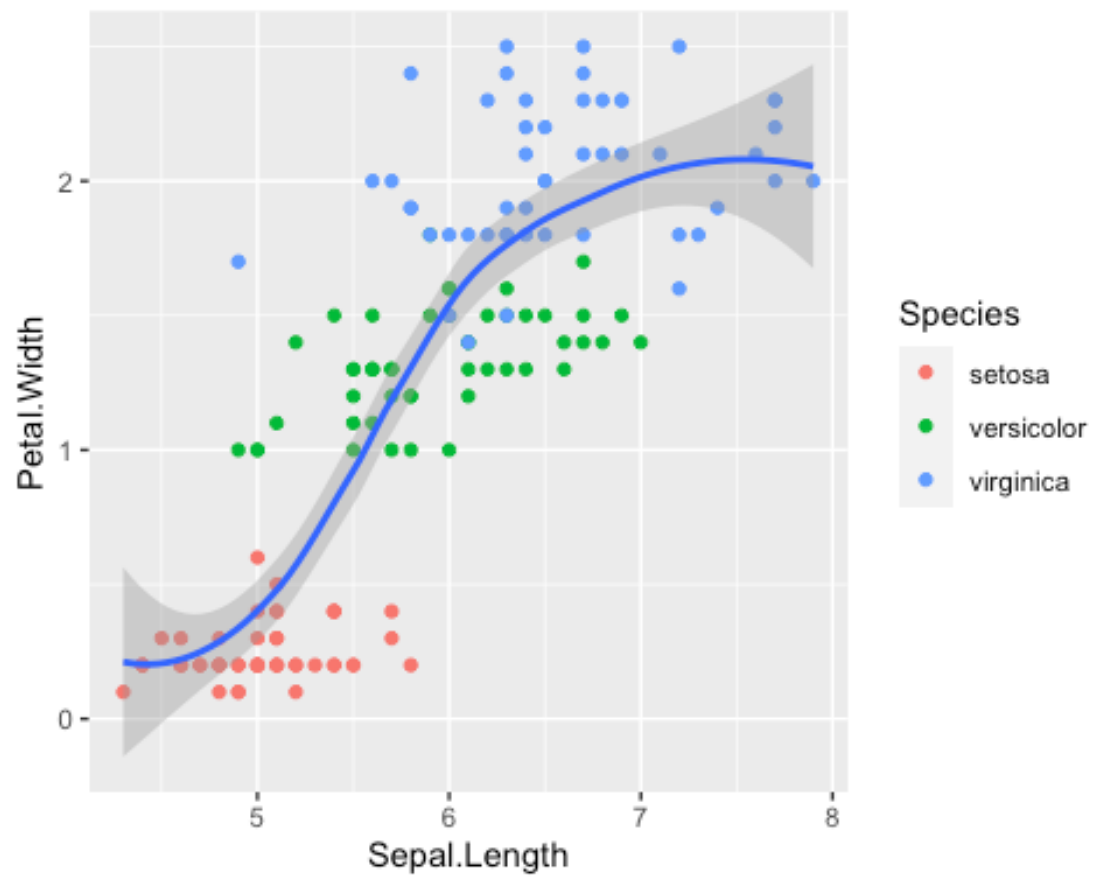
2) Sepal.Length 와 나머지 3 개의 연속변수들 (Sepal.Width, Petal.Length,

Petal.Width)의 관계를 알아보려고 한다. 수업시간에 배운 내용을 바탕으로 살펴 보시오.

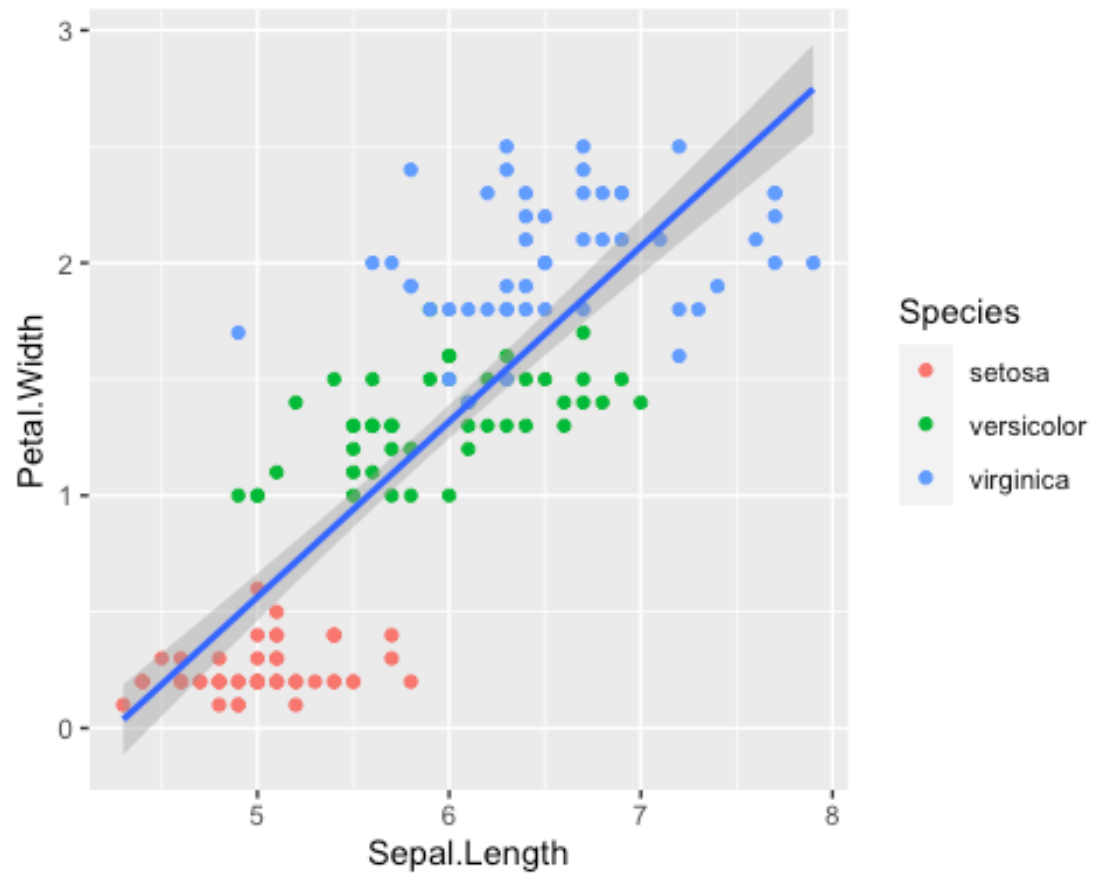
```
ggplot(iris, aes(Sepal.Length, Sepal.Width)) + geom_point(aes(color=Species))
+ geom_smooth(method="lm")
## `geom_smooth()` using formula 'y ~ x'
```



```
ggplot(iris, aes(Sepal.Length, Petal.Width)) + geom_point(aes(color=Species))  
+ geom_smooth()  
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



```
ggplot(iris, aes(Sepal.Length, Petal.Width)) + geom_point(aes(color=Species))  
+ geom_smooth(method="lm")  
## `geom_smooth()` using formula 'y ~ x'
```



3) 4 개의 연속변수를 산점도 행렬을 이용하여 살펴보시오.

```
ggpairs(iris, columns = 1:4, aes(color=Species))
```

