Hello ggplot2!

Heewon Jeon

http://freesearch.pe.kr

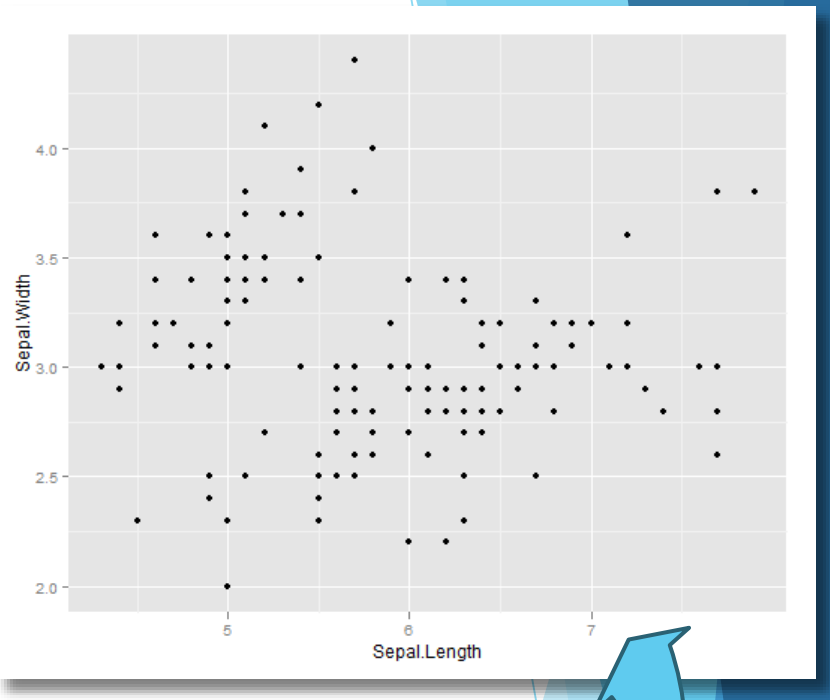
2013.9.14

About the ggplot2 Package

- Grammar of Graphics의 구현체
  - 미적 매핑
  - 통계적인 변환(stat)
  - 기하객체에 적용(geom)
  - 위치 조정(position adjustment)

- “기본 그래픽 시스템은 그림을 그리기 위해 좋은 툴이지만, ggplot2는 데이터를 이해하는 데 좋은 시각화 툴이다.”
  - *Hadley Wickham* -

Basic Graph Functions

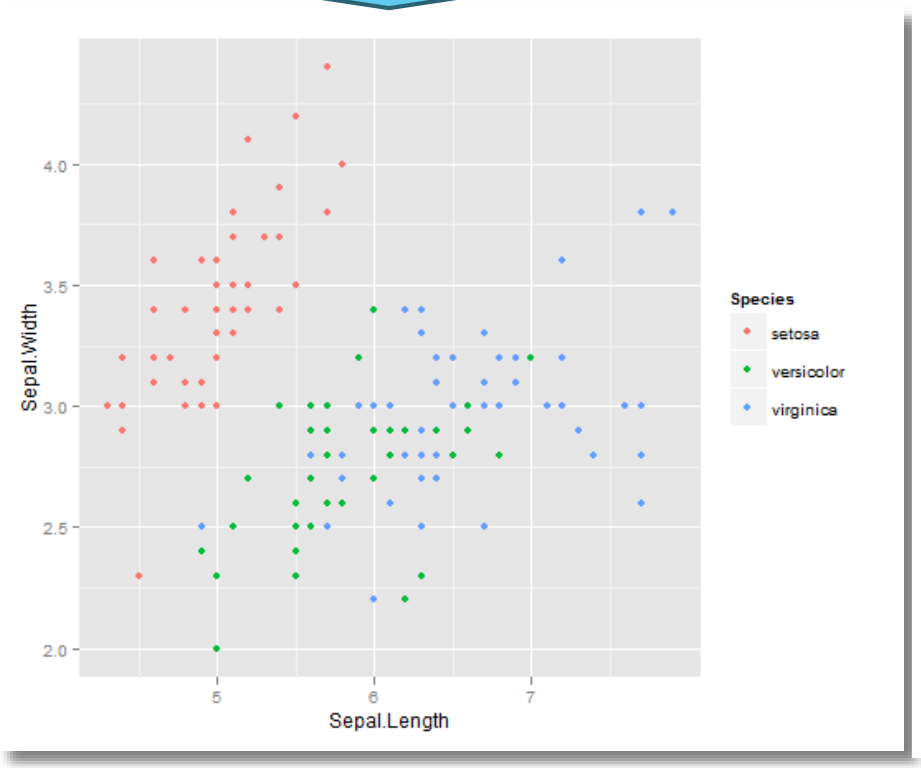


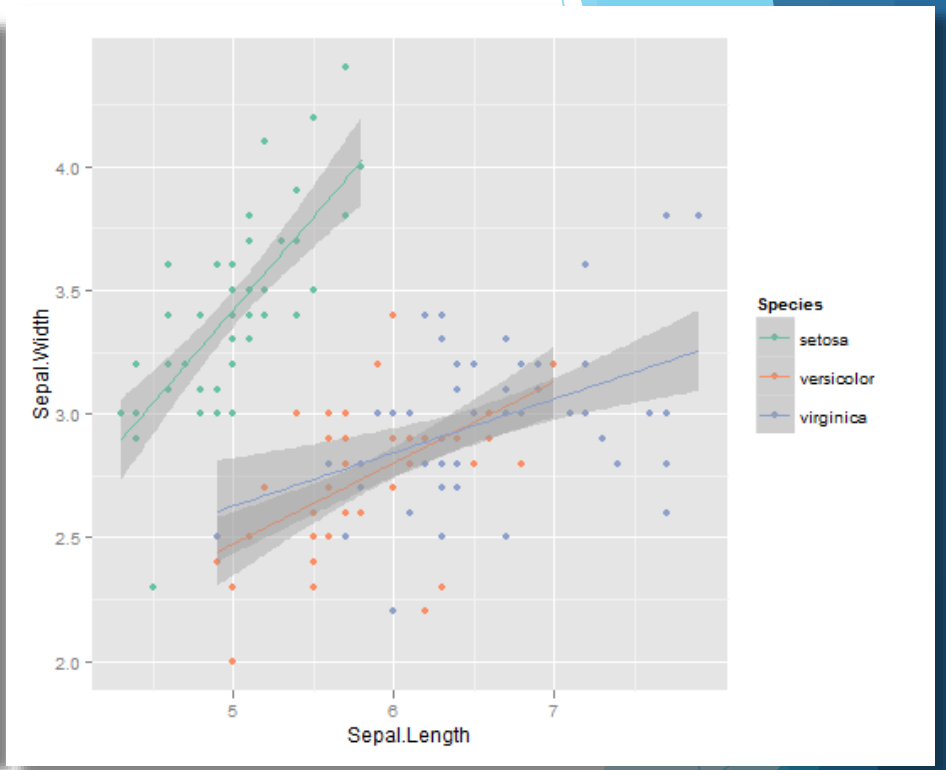

```
install.package("ggplot2")
library(ggplot2)
```

- qplot - 빠른 플로팅을 위한 함수
- ggplot - 문법 기준, 상세한 설정을 할 수 있는 함수

```
qplot(data=iris, x=Sepal.Length, y=Sepal.Width, geom="point")

ggplot(iris,aes(Sepal.Length,Sepal.Width))+geom_point()
```

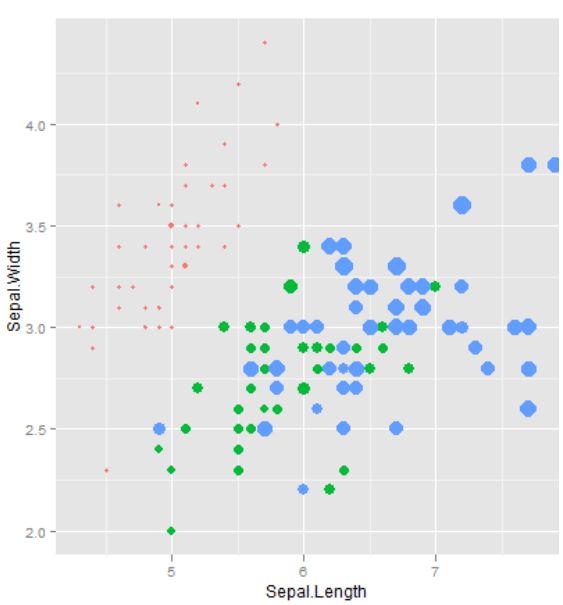
#아이리스 종류별로 컬러링
ggplot(iris,aes(Sepal.Length,Sepal.Width)) + **geom_point**(aes(colour=Species))

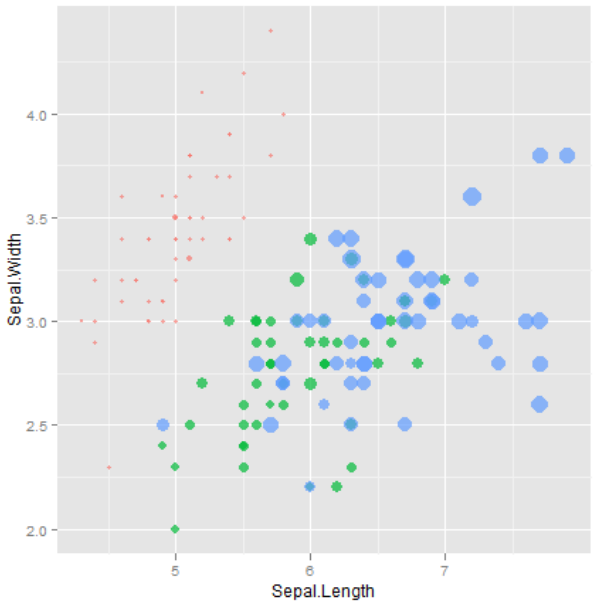


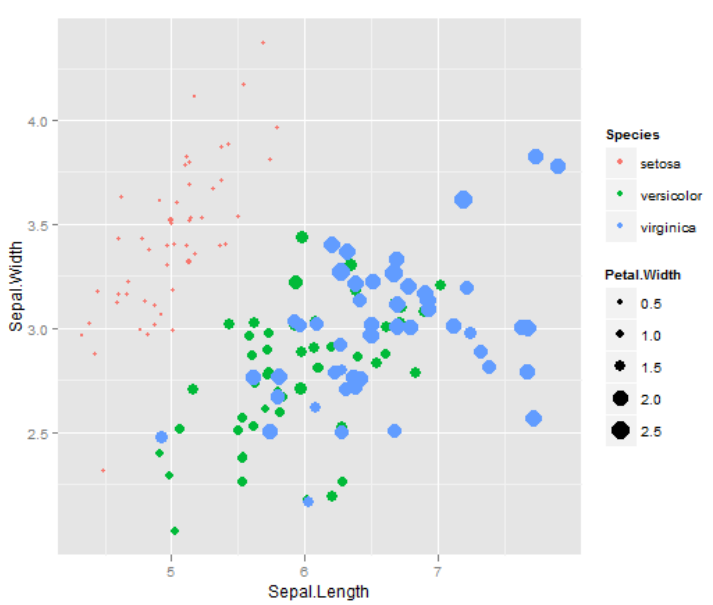


#아이리스 종류별로 regression 라인 피팅
ggplot(iris,aes(Sepal.Length,Sepal.Width)) +
geom_point(aes(colour=Species)) +
geom_smooth(aes(colour=Species), method=rlm)

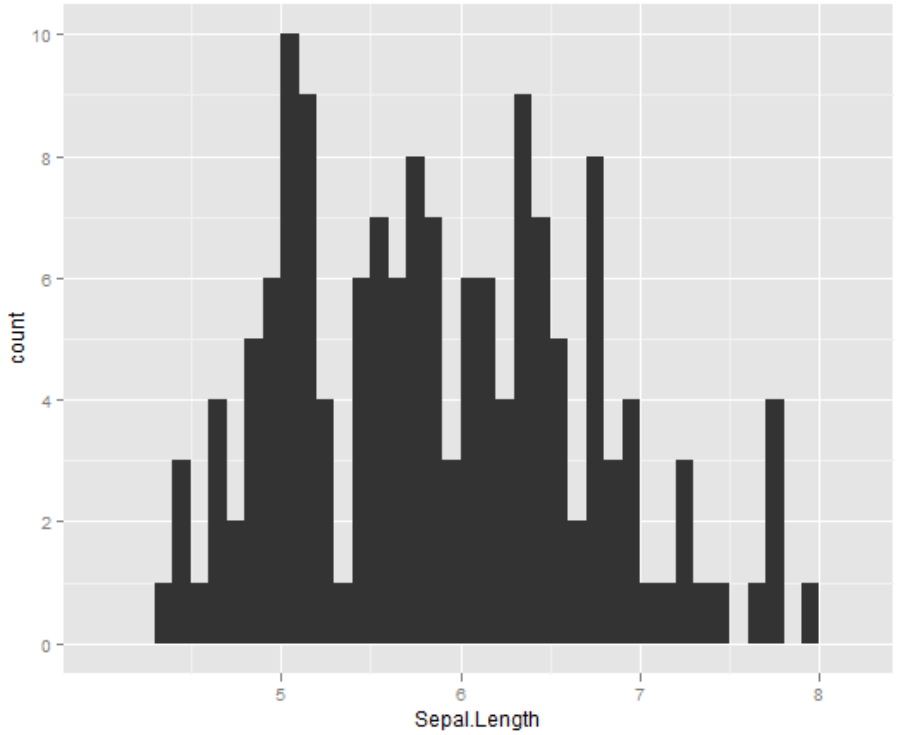
```
#아이리스 종류별로 컬러링 / 꽃잎 길이에 따른 포인트 크기 표현
ggplot(iris,aes(Sepal.Length,Sepal.Width)) +
  geom_point(aes(colour=Species, size=Petal.Width))


#중복된 점 표현 방법 1
ggplot(iris,aes(Sepal.Length,Sepal.Width)) +
  geom_point(aes(colour=Species,
    size=Petal.Width), alpha=I(0.7))

#중복된 점 표현 방법 2
ggplot(iris,aes(Sepal.Length,Sepal.Width)) +
  geom_point(aes(colour=Species, size=Petal.Width), position="jitter")
```

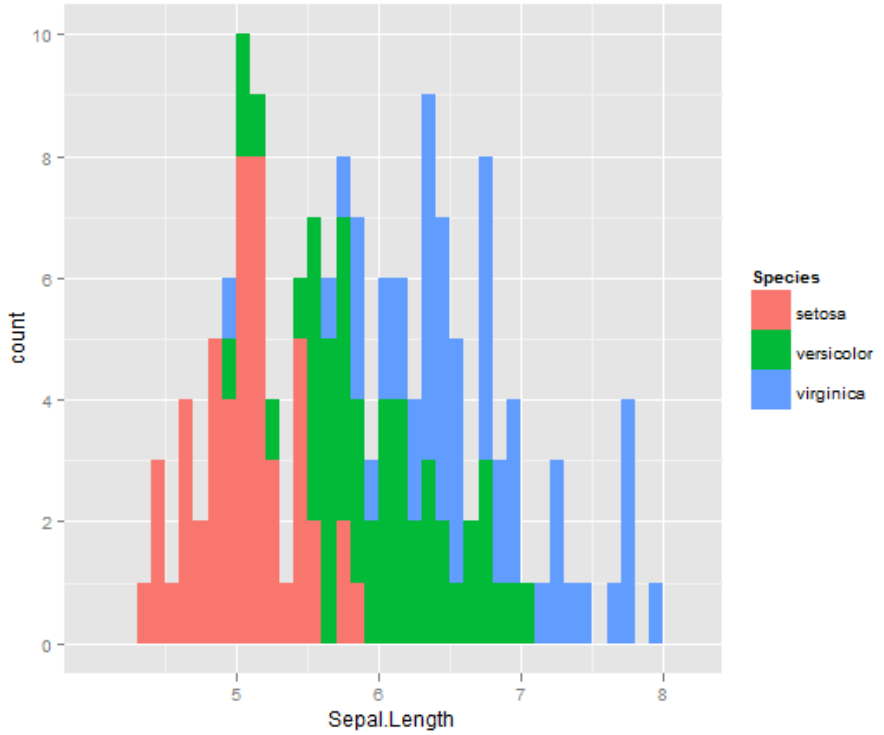






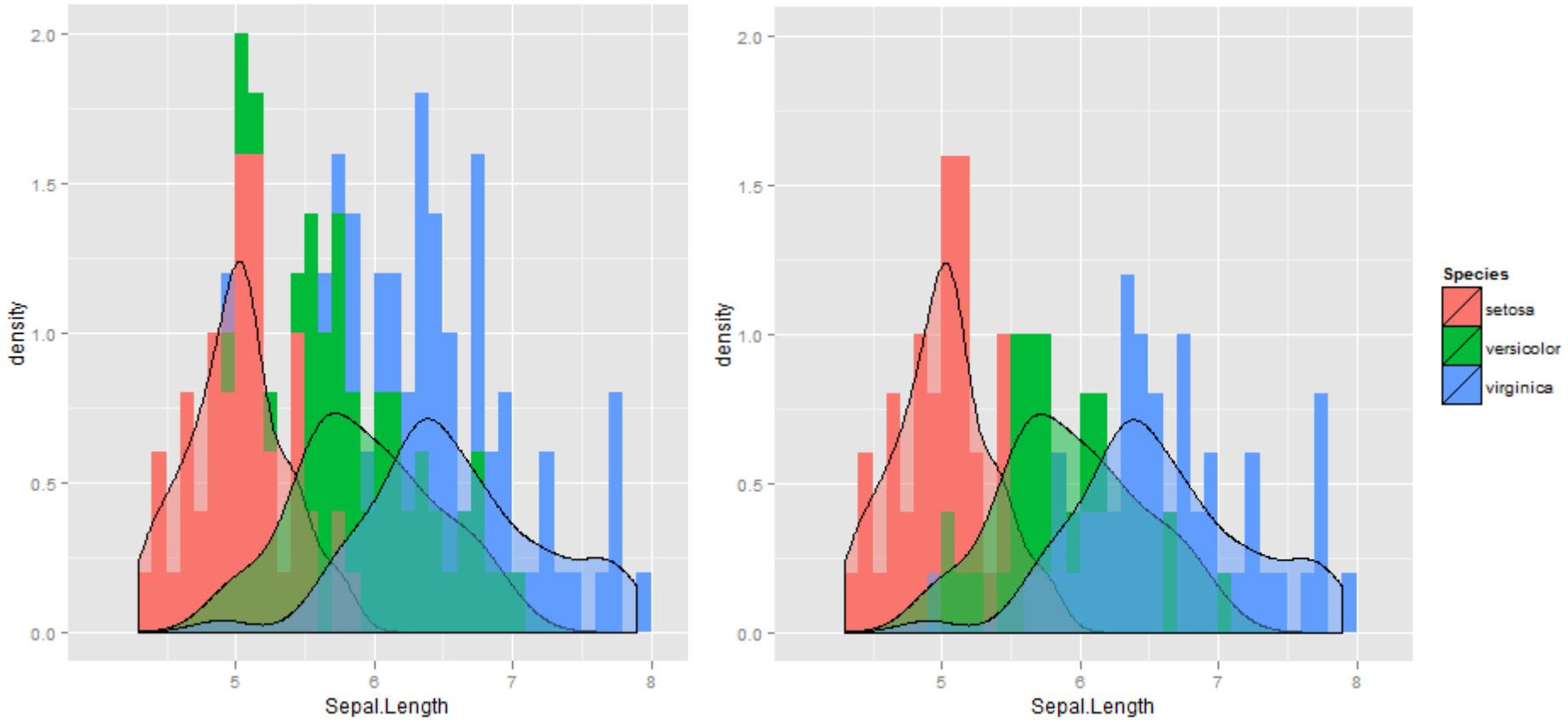
#꽃받침 길이에 대한 히스토그램
ggplot(iris,aes(Sepal.Length))+**geom_histogram**(binwidth=0.1)





#종류에 따른 꽃받침 길이에 대한 히스토그램(stack)
ggplot(iris,aes(Sepal.Length))+**geom_histogram**(binwidth=0.1, aes(fill=Species))

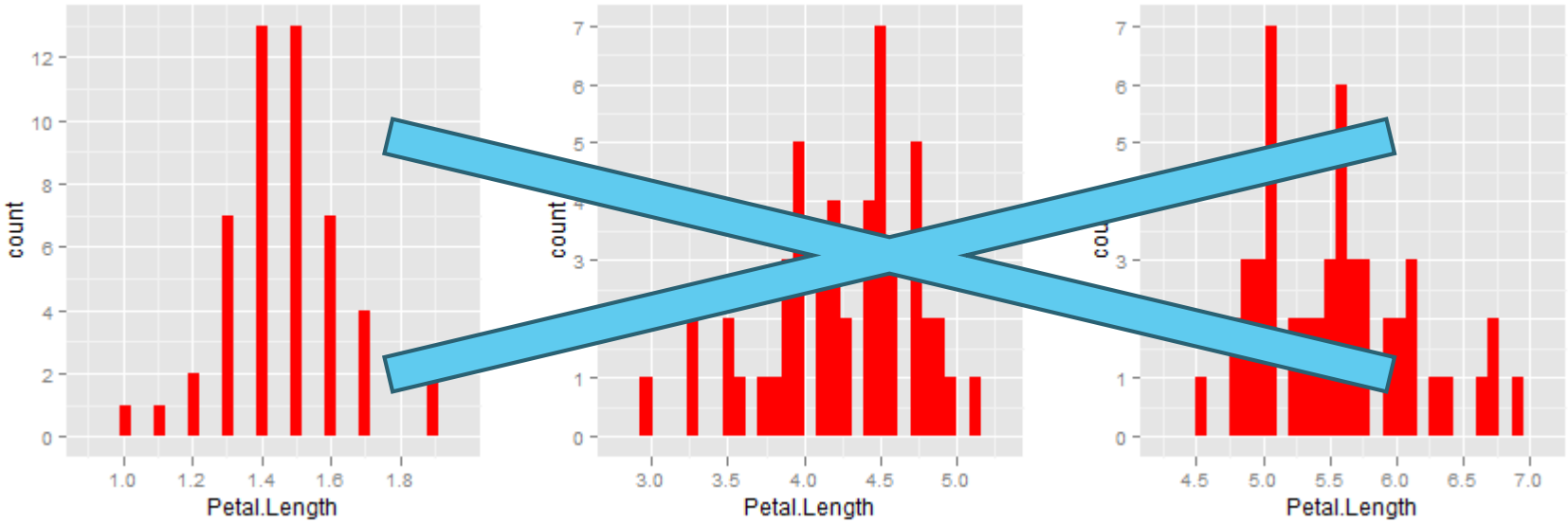
```
#density plot?
ggplot(iris,aes(Sepal.Length)) +
  geom_histogram(binwidth=0.1, aes(y=..density..,fill=Species)) +
  geom_density(aes(fill=Species), alpha=I(0.5))
```



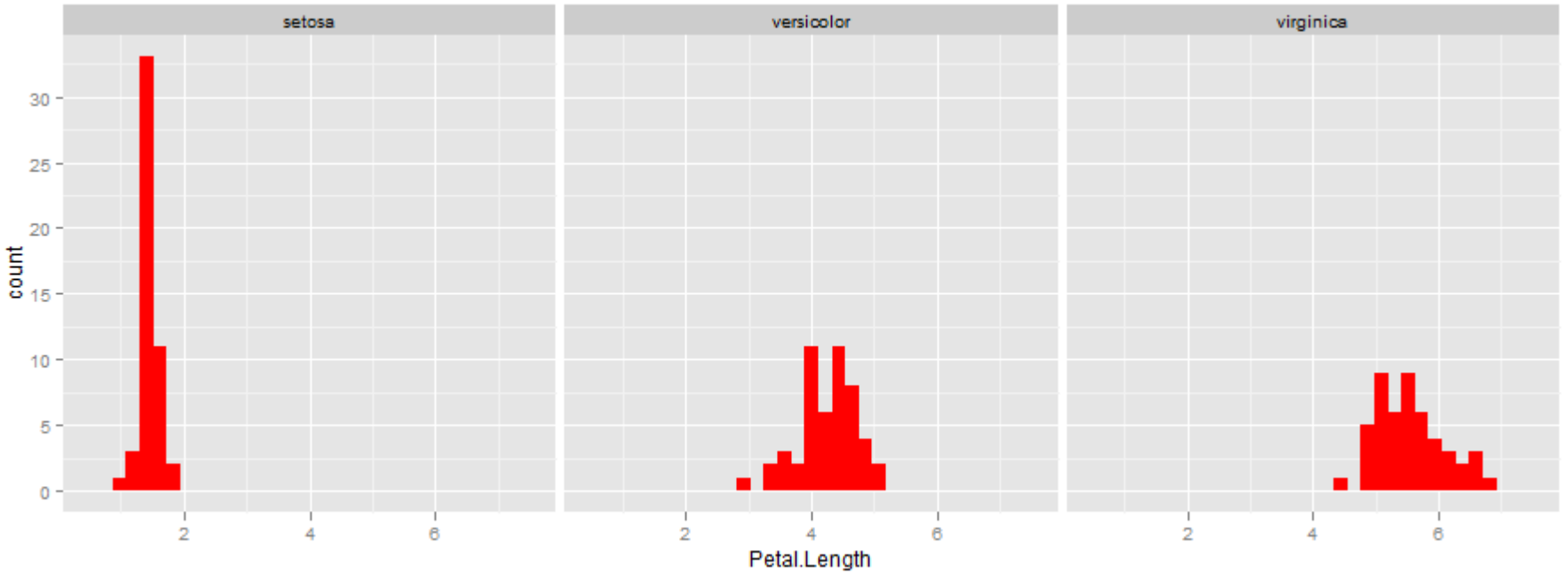

```
#density plot!
ggplot(iris,aes(Sepal.Length)) +
  geom_histogram(binwidth=0.1, aes(y=..density..,fill=Species),
position="identity") +
  geom_density(aes(fill=Species), alpha=I(0.5)) +  ylim(0,2.0)
```

그래프 객체 재사용

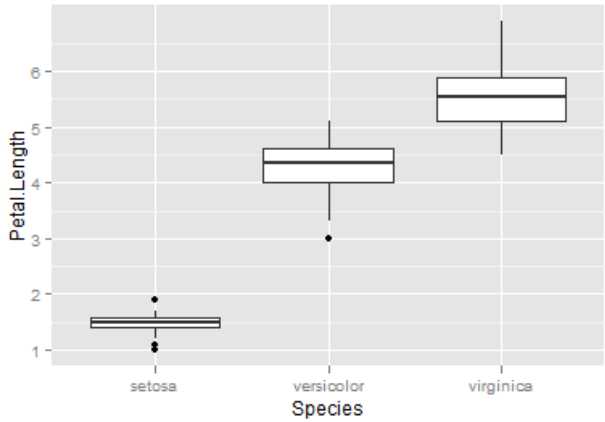
```
library(gridExtra)
library(scales)
setosaData <- subset(iris, Species=="setosa")
versicolorData <- subset(iris, Species=="versicolor")
virginicaData <- subset(iris, Species=="virginica")

setosaDataG <- ggplot(setosaData, aes(Petal.Length)) +
  geom_histogram(fill=alpha("red", 0.5))
setosaDataG
versicolorDataG <- setosaDataG %+% versicolorData
virginicaDataG <- setosaDataG %+% virginicaData

grid.arrange(setosaDataG, versicolorDataG, virginicaDataG, ncol=3)
```



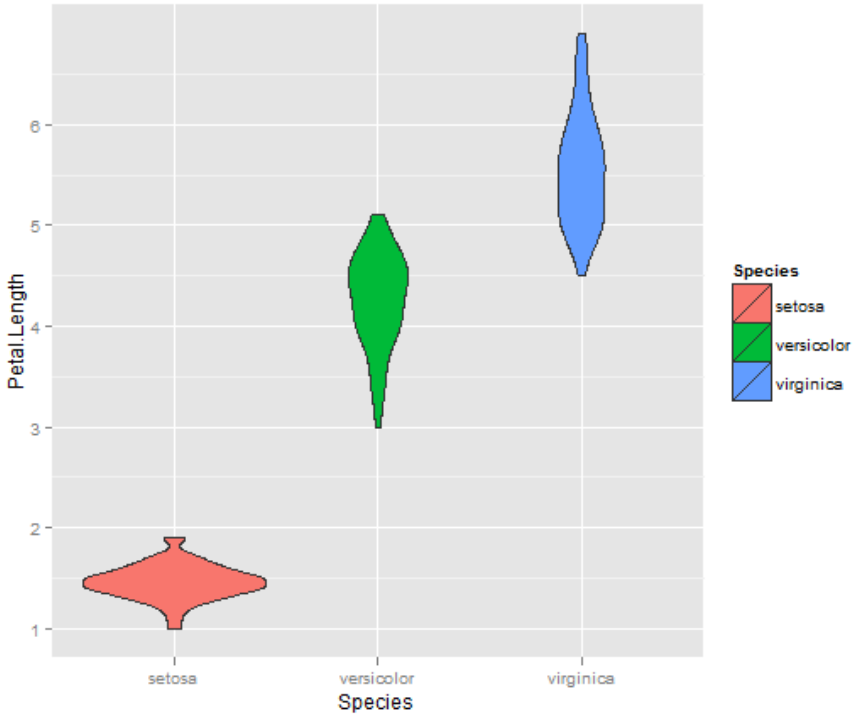
```
ggplot(iris,aes(Petal.Length)) +
  geom_histogram(fill = alpha("red", 0.5)) +
  facet_grid(.~Species, scales="fixed")
```

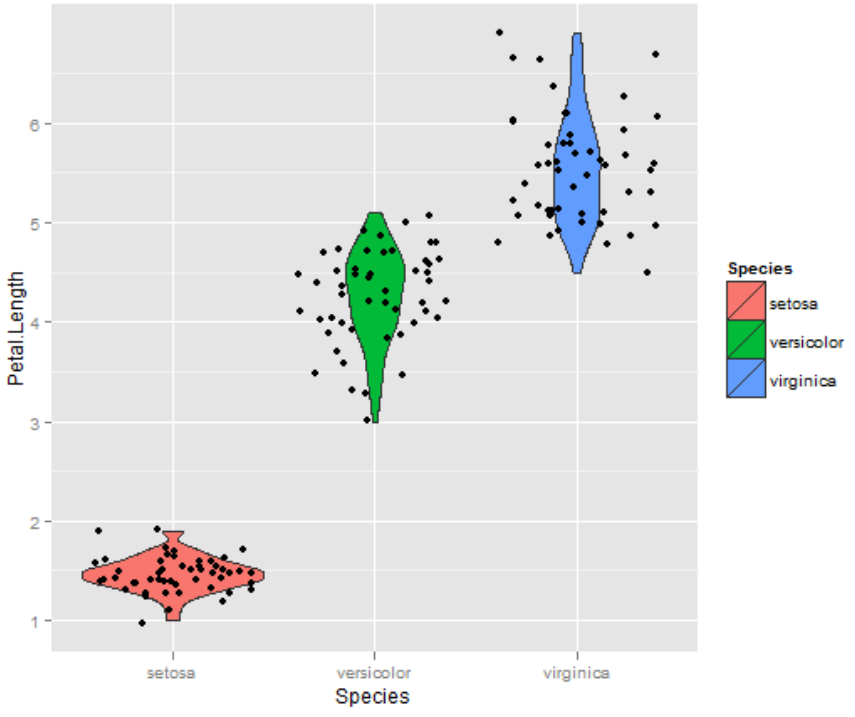



```
#boxplot
ggplot(iris) +
  geom_boxplot(aes(y=Petal.Length, x=Species))
```



```
ggplot(iris, aes(Species, Petal.Length)) +
  geom_violin(aes(fill=Species))
```

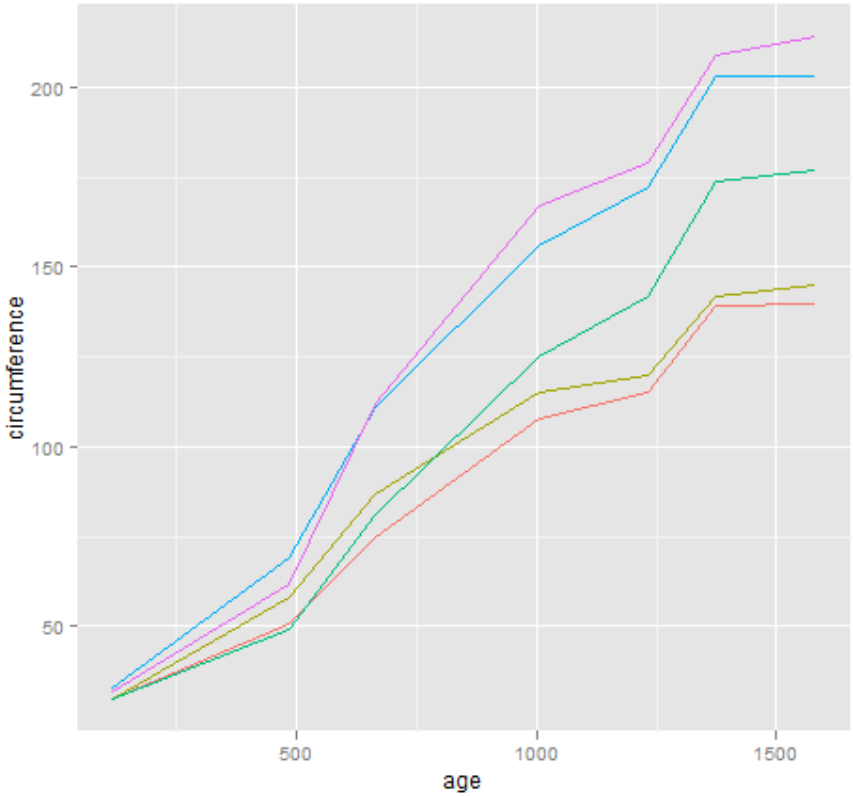


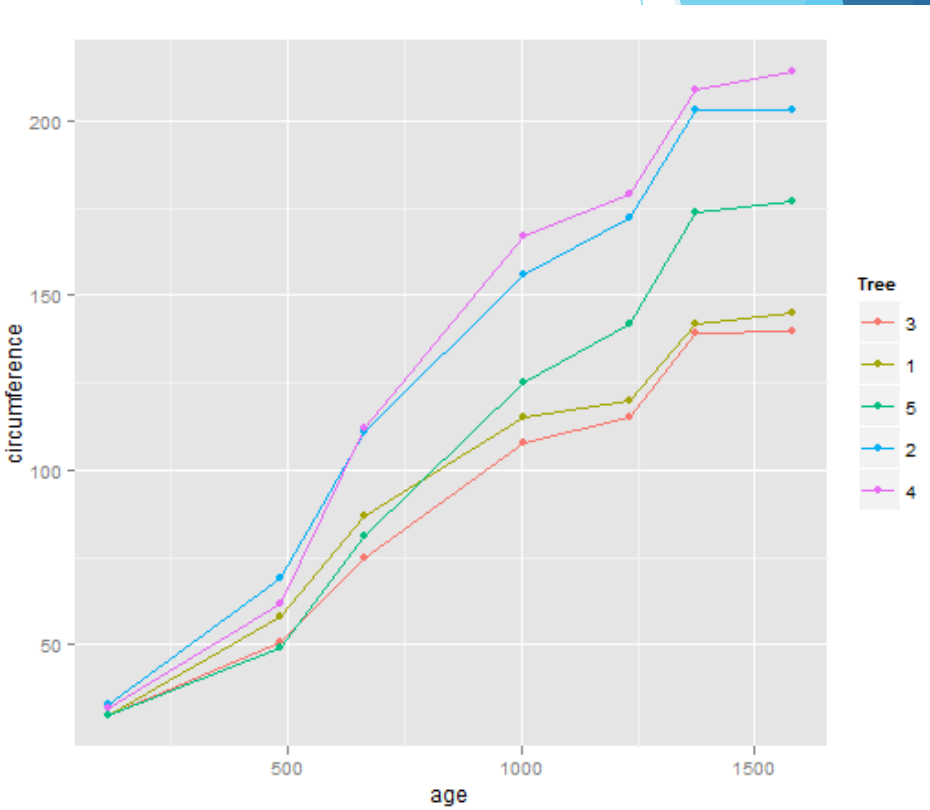



```
ggplot(iris, aes(Species, Petal.Length)) +
  geom_violin(aes(fill=Species)) + geom_jitter(height=0)
```

ggplot(Orange, aes(age, circumference)) + **geom_line**(aes(colour=Tree))

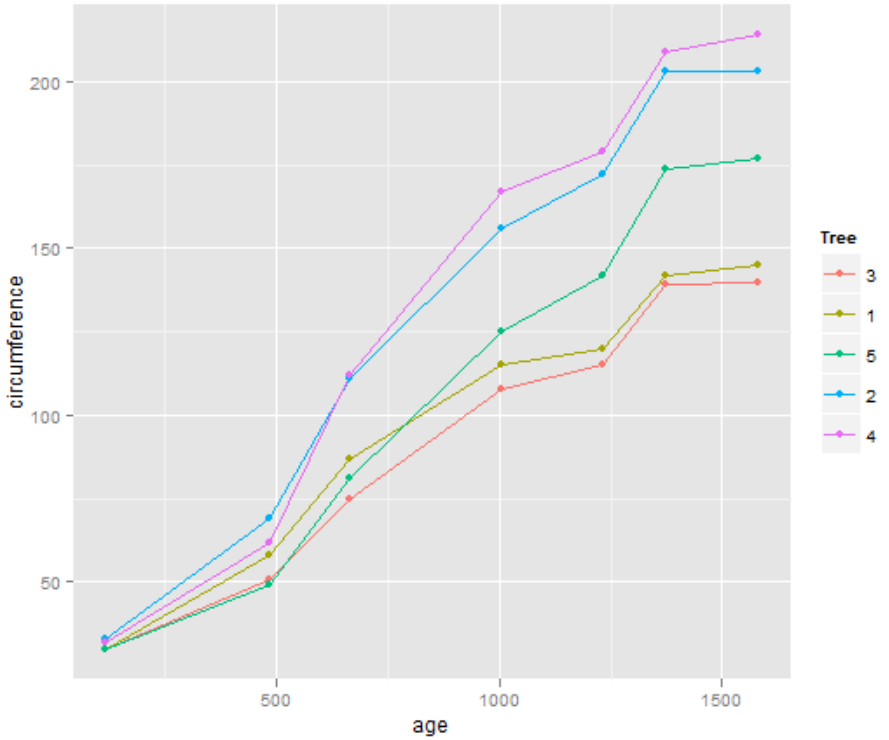
ggplot(Orange, aes(age, circumference, ***colour=Tree***)) + **geom_line**() + **geom_point**()



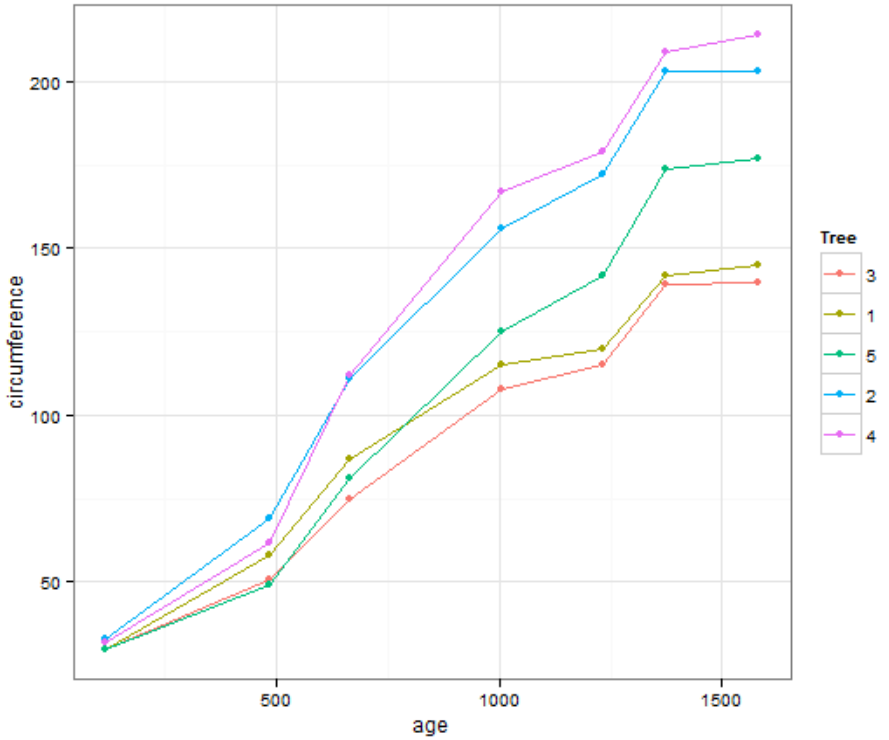


```
p <- ggplot(Orange, aes(age, circumference, colour=Tree)) + geom_line() +
    geom_point()
p  #(1)
old_theme <- theme_set(theme_bw())
p  #(2)
theme_set(old_theme)
p  #(1)
p + theme_bw() #(2)
```

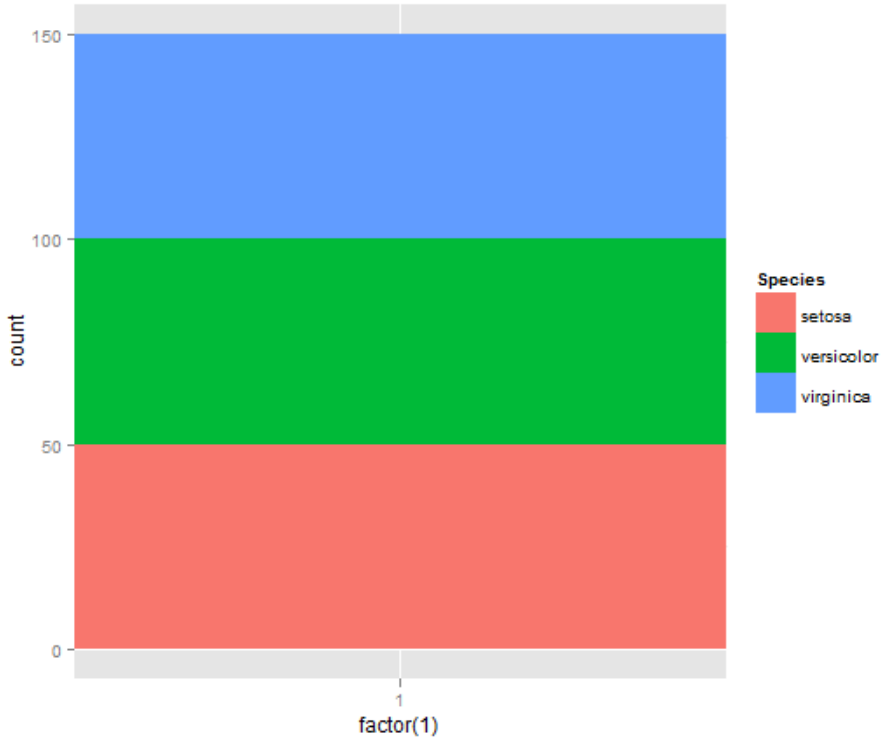
(1)

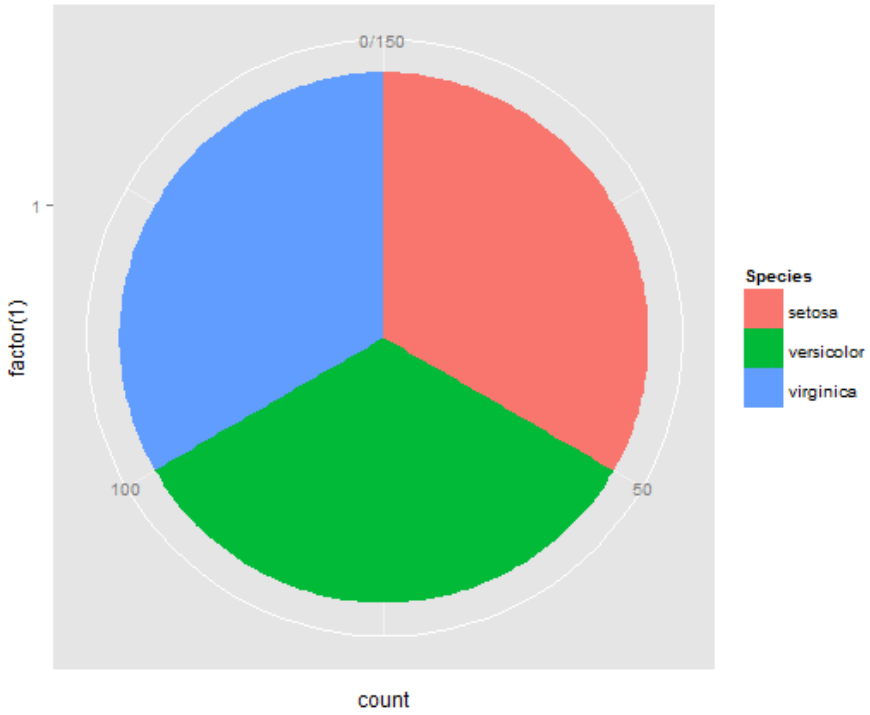


(2)

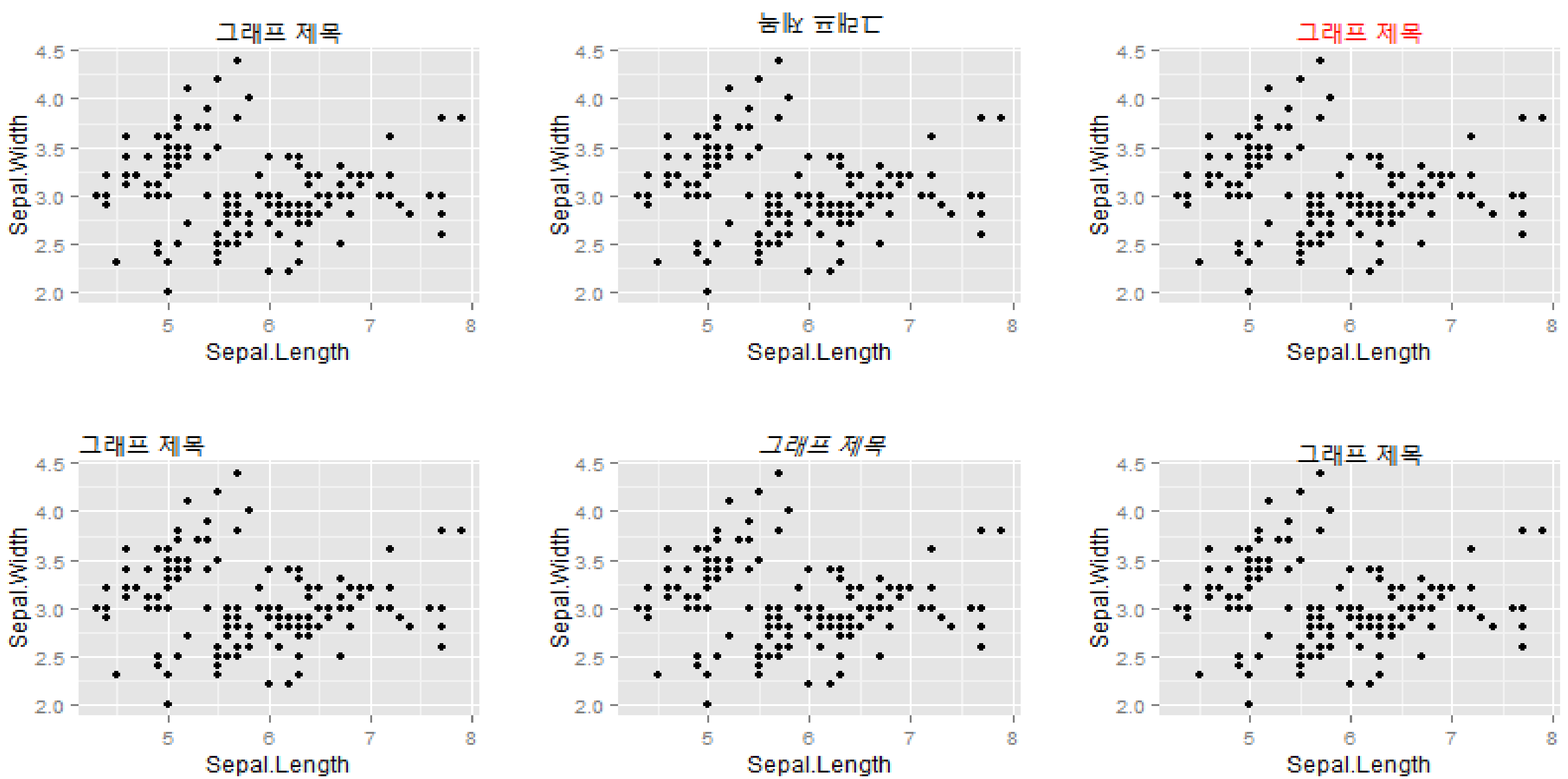


```
#stack bar chart
ggplot(iris,aes(x=factor(1), fill=Species)) + geom_bar(width=1)

#pie chart
ggplot(iris,aes(x=factor(1), fill=Species)) + geom_bar(width=1) +
coord_polar(theta="y")
```

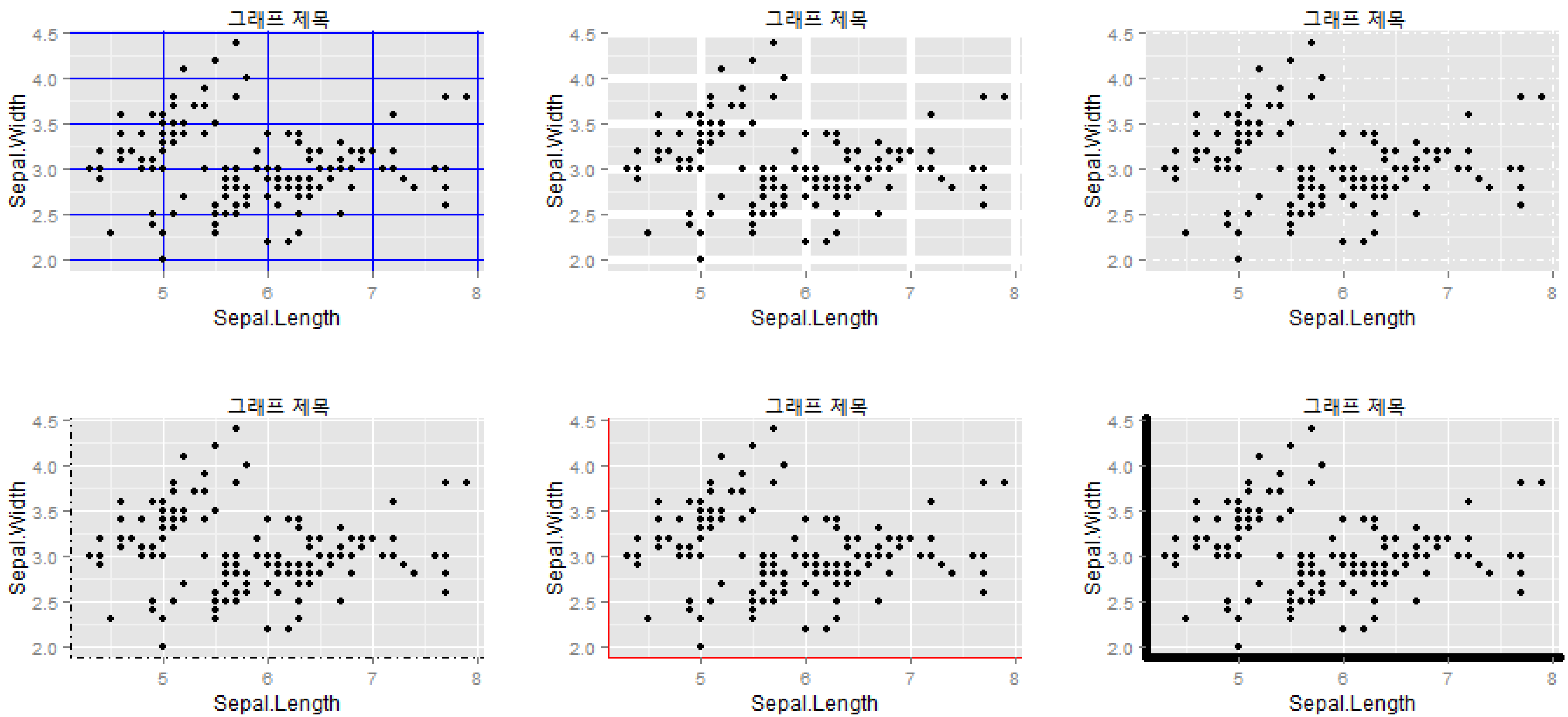




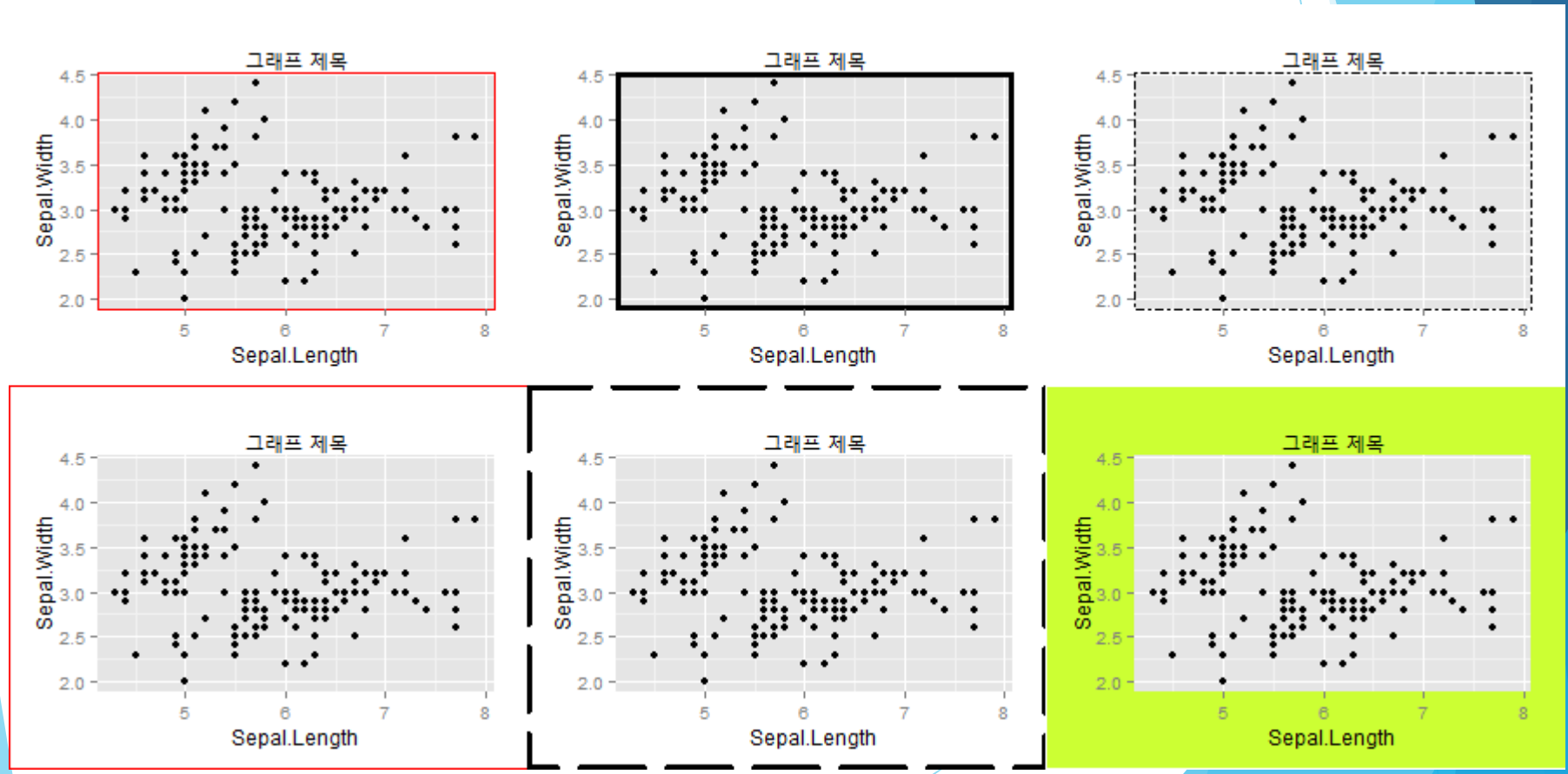
```
d <- ggplot(iris, aes(Sepal.Length, Sepal.Width)) + geom_point() + ggtitle("그래프 제목")
g1 <- d + theme(plot.title=element_text(size=15))
g2 <- d + theme(plot.title=element_text(size=15, angle=180))
g3 <- d + theme(plot.title=element_text(size=15, color="red"))
g4 <- d + theme(plot.title=element_text(size=15, hjust=0))
g5 <- d + theme(plot.title=element_text(size=15, face="italic"))
g6 <- d + theme(plot.title=element_text(size=15, vjust=0))
```



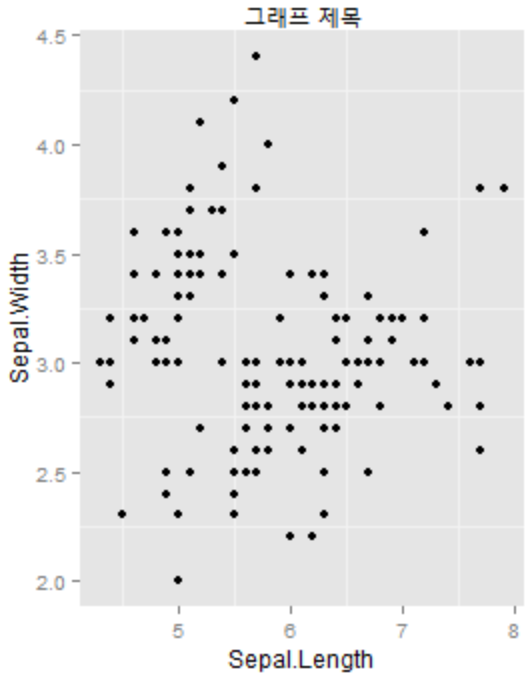
```
d <- ggplot(iris, aes(Sepal.Length, Sepal.Width)) + geom_point() + ggtitle("그래프 제목")
g1 <- d + theme(panel.grid.major=element_line(colour="blue"))
g2 <- d + theme(panel.grid.major=element_line(size=2))
g3 <- d + theme(panel.grid.major=element_line(linetype="dotdash"))
g4 <- d + theme(axis.line=element_line(linetype="dotdash"))
g5 <- d + theme(axis.line=element_line(colour="red"))
g6 <- d + theme(axis.line=element_line(lineend="round", size=2))
```

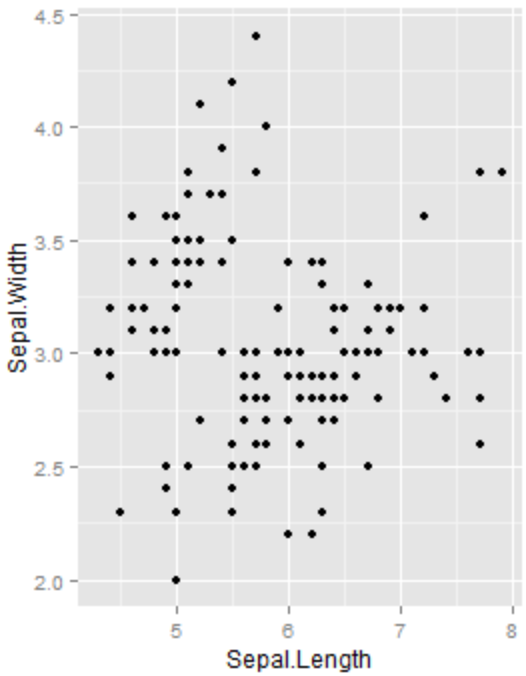


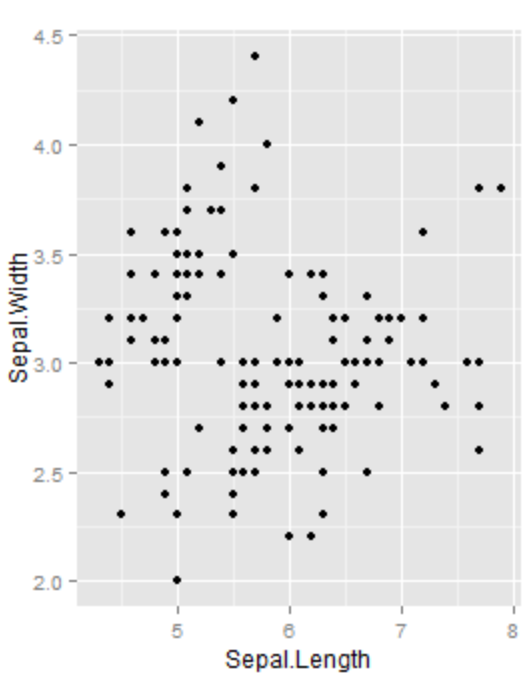
```
d <- ggplot(iris, aes(Sepal.Length, Sepal.Width)) + geom_point() + ggtitle("그래프 제목")
g1 <- d + theme(panel.border=element_rect(colour="red", fill=NA))
g2 <- d + theme(panel.border=element_rect(size=2, fill=NA))
g3 <- d + theme(panel.border=element_rect(linetype="dotdash", fill=NA))
g4 <- d + theme(plot.background=element_rect(colour="red"))
g5 <- d + theme(plot.background=element_rect(size=2, colour="black", linetype="longdash"))
g6 <- d + theme(plot.background=element_rect(fill="#CCFF33"))
```



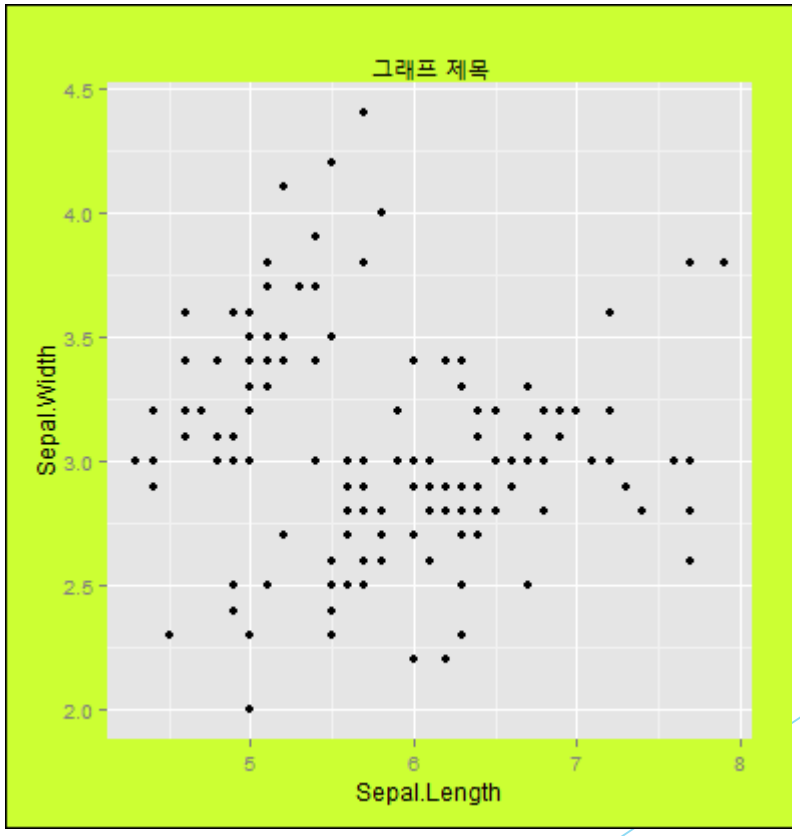
```
d <- ggplot(iris, aes(Sepal.Length, Sepal.Width)) + geom_point() + ggtitle("그래프 제목")
g1 <- d + theme(panel.grid.major=element_blank())
g2 <- d + theme(plot.title=element_blank())
g3 <- d + theme(plot.title=element_text(colour=NA))
```







```
#플롯의 배경색만 다르게 설정
old_theme <- theme_update(plot.background=element_rect(fill="#CCFF33"))
d
#다시 예전 테마로
theme_set(old_theme)
```



#ggplot에서 한글 폰트 사용하기

http://kldp.net/projects/unfonts/ 에서 은폰트 설치

#윈도우 폰트 확인
```
head(names(windowsFonts()))
```

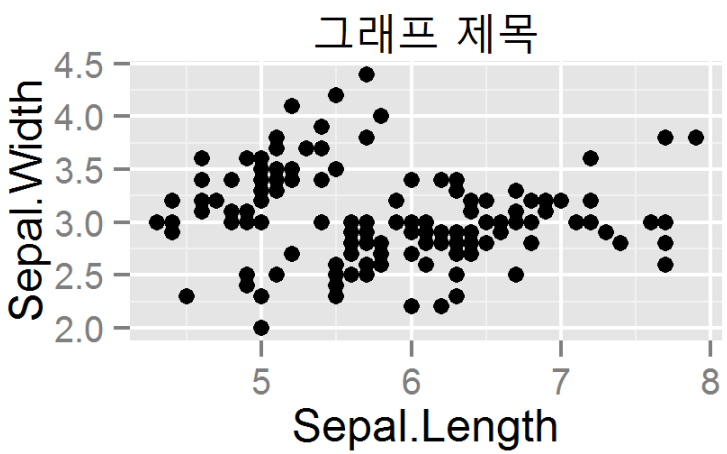
Postscript 폰트 확인
```
head(names(postscriptFonts()))
```

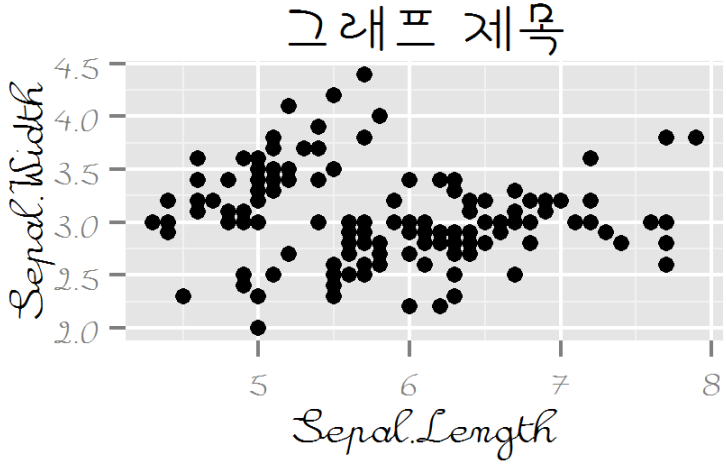
PDF 폰트 확인
```
head(names(pdfFonts()))
```

#패키지 설치
```
install.packages("extrafont")
library(extrafont)
font_import()
```

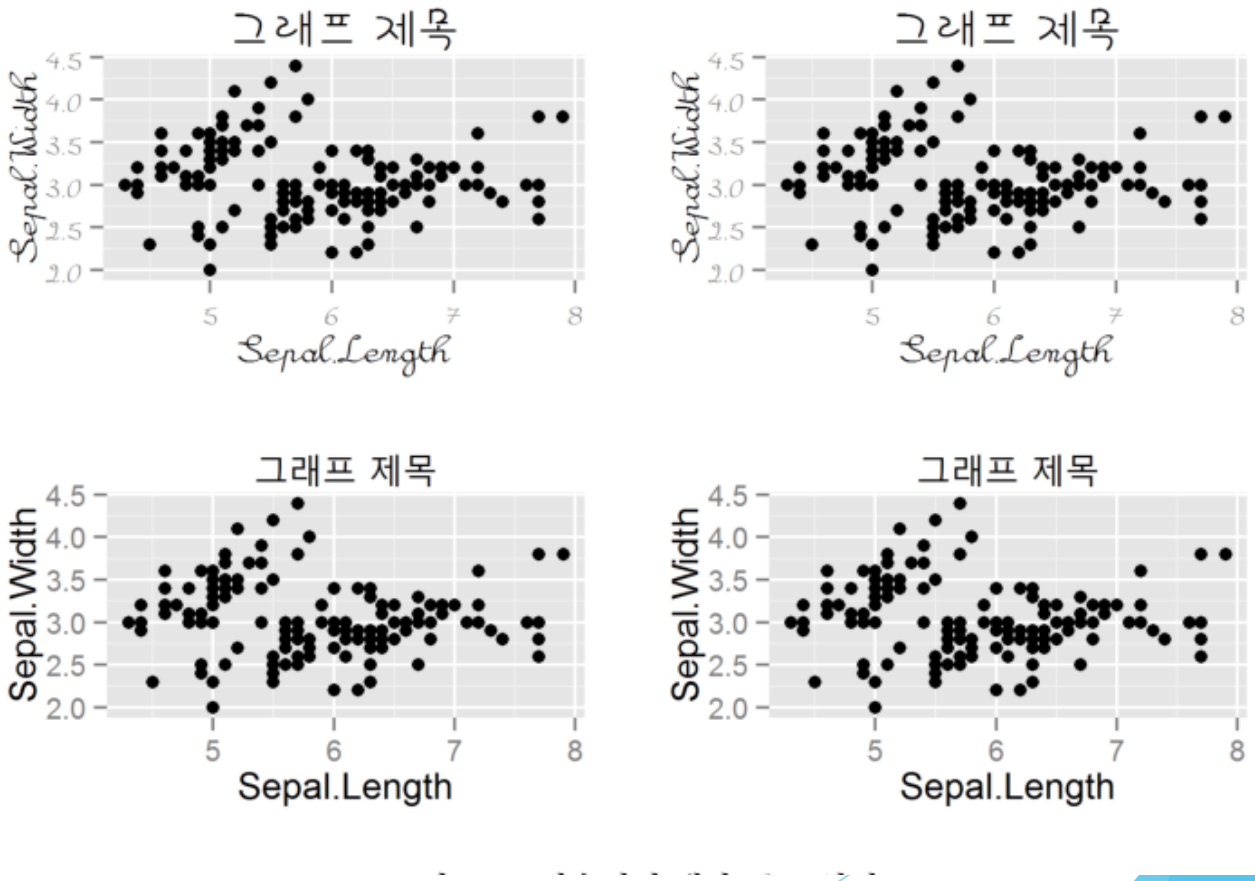
```
Importing fonts may take a few minutes, depending on the number of fonts and the speed of the system.
Continue? [y/n] y
Scanning ttf files in C:\Windows\Fonts ...
Extracting .afm files from .ttf files...
C:\Windows\Fonts\ahn_b.ttf => E:/R/R-3.0.1/library/extrafontdb/metrics/ahn_b
C:\Windows\Fonts\ahn_l.ttf => E:/R/R-3.0.1/library/extrafontdb/metrics/ahn_l
C:\Windows\Fonts\ahn_m.ttf => E:/R/R-3.0.1/library/extrafontdb/metrics/ahn_m
C:\Windows\Fonts\ahronbd.ttf : Aharoni-Bold already registered in fonts database. Skipping.
C:\Windows\Fonts\aldhabi.ttf : Aldhabi already registered in fonts database. Skipping.
C:\Windows\Fonts\ALGER.TTF : Algerian already registered in fonts database. Skipping.
C:\Windows\Fonts\Along.ttf
```

```
#폰트 설치 확인
grep( " UnPilgi " , fonttable()$FamilyName)

d <- ggplot(iris, aes(Sepal.Length, Sepal.Width)) + geom_point() + ggtitle("그래프 제목")
d
d + theme(text=element_text(family="UnPilgi"))
```

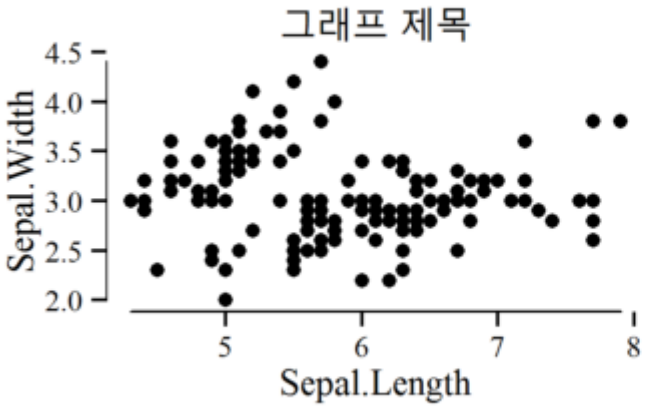


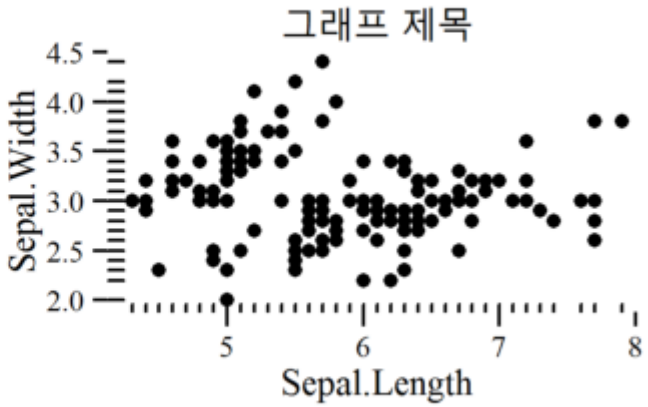


```
old_theme <- theme_set(theme_gray(base_family="UnPilgi"))
d
d
theme_set(old_theme)
d
d
```

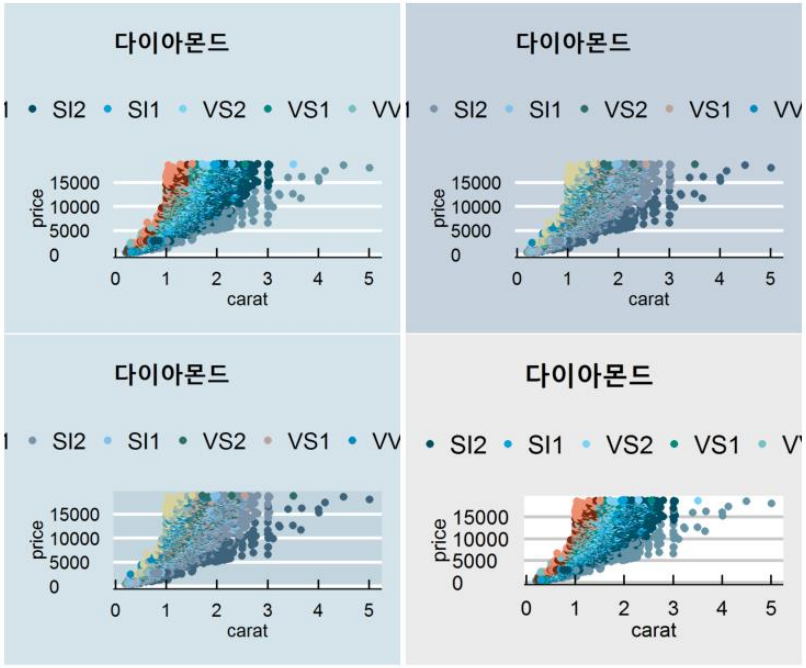


```
library(ggthemes)

d + geom_rangeframe() + theme_tufte()
d + geom_rug() + theme_tufte(ticks = TRUE)
```





```
dsamp <- diamonds[sample(nrow(diamonds), 1000), ]
q <- ggplot(diamonds, aes(carat, price, colour = clarity)) + geom_point() +
ggtitle("다이아몬드")
## Standard
q + theme_economist() + scale_colour_economist()
## Stata colors
q + theme_economist(stata = TRUE) + scale_color_economist(stata=TRUE)
## Darker plot region
q + theme_economist(dkpanel = TRUE) +
scale_colour_economist(stata=TRUE)
## White panel/light gray background
(q + theme_economist_white() + scale_colour_economist())
```



References

- http://docs.ggplot2.org/current/