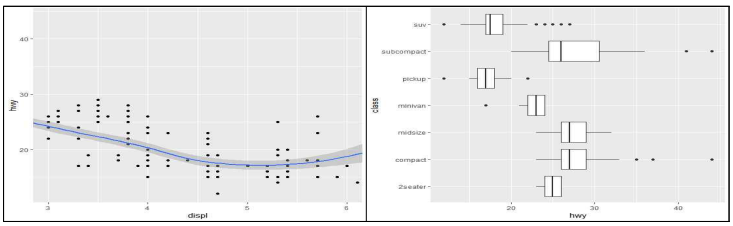
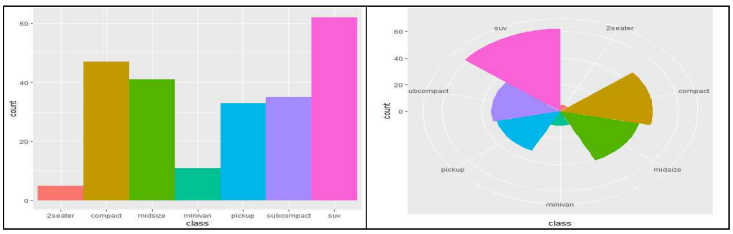
좌표계 : coord_cartesian & coord_flip

```
p <- ggplot(mpg, aes(x=displ,y=hwy)) + geom_point() + geom_smooth()
p + coord_cartesian(xlim=c(3,6))

ggplot(mpg, aes(x=class,y=hwy)) + geom_boxplot() + coord_flip()
```



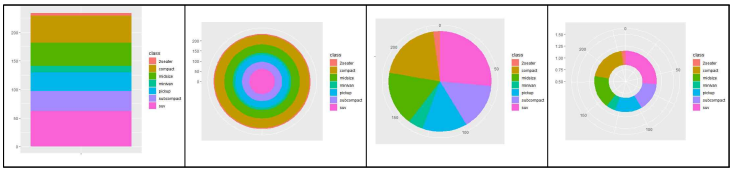
좌표계 : coord_polar()

```
b <- ggplot(mpg, aes(x=class,fill=class)) + geom_bar(show.legend=FALSE,width=1)
b + coord_polar()
```



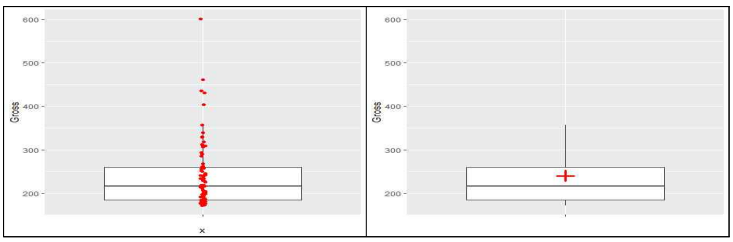
파이그래프

```
b2 <- ggplot(mpg, aes(x="",fill=class)) + geom_bar(width=1) + labs(x="",y="")
b2 + coord_polar()
b2 + coord_polar(theta ="y")
b3 <- ggplot(mpg, aes(x=1,fill=class)) + geom_bar(width=0.3) + labs(x="",y="") + xlim(0.5,1.5)
b3 + coord_polar(theta="y")
```



Boxplot

```
library(UsingR)
bp1 <- ggplot(alltime.movies, aes(x="",y=Gross)) + geom_boxplot(outlier.shape=NA)
bp1 + geom_jitter(col="red",width=0.01)
bp1 + stat_summary(fun.y="mean",geom="point",col="red",shape=3,size=4,stroke=2) + xlab("")
```

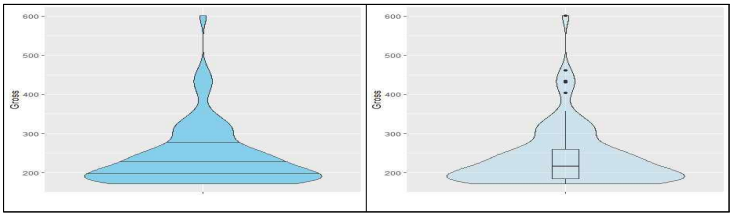


Outlier

```
my_box <- boxplot(alltime.movies$Gross, plot=FALSE)
alltime <- as_tibble(alltime.movies) %>% rownames_to_column(var="Movie.Title")
top_movies <- alltime %>% filter(Gross %in% my_box$out)

# A tibble: 5 x 3
  Movie.Title                          Gross Release.Year
  <chr>                                <dbl>        <dbl>
1 "Titanic                             "   601         1997
2 "Star Wars                           "   461         1977
3 "E.T.                                "   435         1982
4 "Star Wars: The Phantom Menace       "   431         1999
5 "Spider-Man                          "   404         2002
```

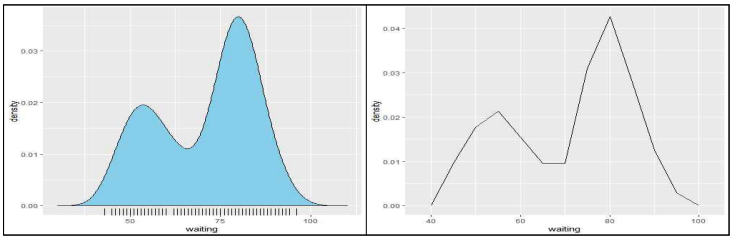
Violin plot

```
vio <- ggplot(alltime.movies, aes(x="",y=Gross)) + xlab("")
vio + geom_violin(draw_quantiles=c(0.25,0.5,0.75),fill="skyblue")
vio + geom_violin(alpha=0.3,fill="skyblue") + geom_boxplot(fill=NA,width=0.1)
```



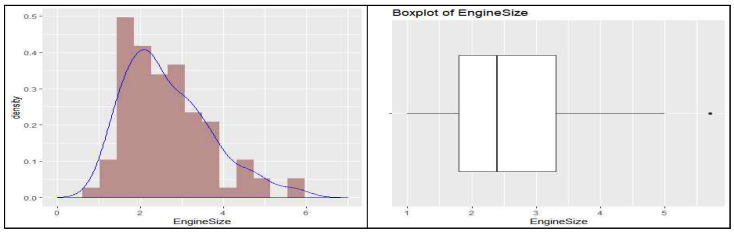
확률밀도함수 & 상대도수분포다각형

```
p <- ggplot(faithful, aes(x=waiting)) + geom_density(fill="skyblue")
p + xlim(30,110) + geom_rug()

pp <- ggplot(faithful, aes(x=waiting))
pp + geom_freqpoly(aes(y=..density..),binwidth=5)
```



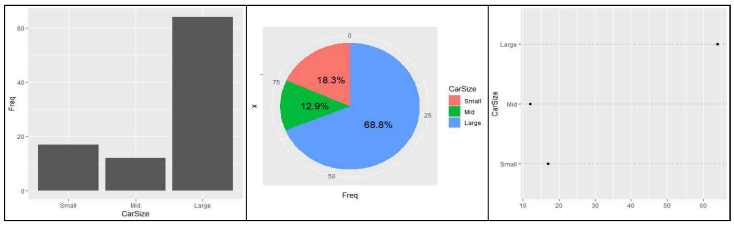
일변량 연습문제 1

```
ggplot(Cars93, aes(x=EngineSize,y=..density..)) +
  geom_histogram(bins=18,fill="rosybrown") +
  geom_density(color="blue") + xlim(0,7)

ggplot(Cars93, aes(x="",y=EngineSize)) + geom_boxplot() +
  coord_flip() + labs(x="",title="Boxplot of EngineSize")
```



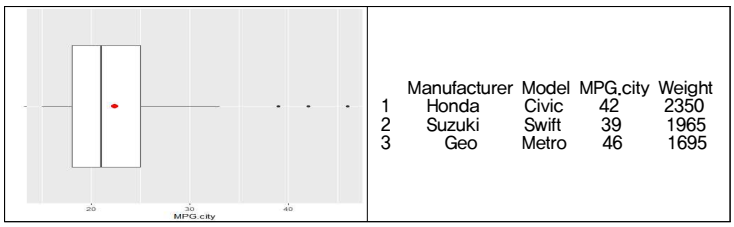
```
Cars93 %>% filter(EngineSize>5) %>% dplyr::select(Manufacturer,Model)

     Manufacturer     Model
1           Buick Roadmaster
2       Chevrolet  Corvette


Cars93 %>% filter(EngineSize<=5) %>% summarise(EngineSize_m=mean(EngineSize),EngineSize_sd=sd(EngineSize)) %>% round(,2)
  EngineSize_m EngineSize_sd
1          2.6          0.94
```

```
library(scales)
data(Cars93, package="MASS")
p <- mutate(Cars93,CarSize=cut(EngineSize,breaks=c(min(EngineSize)-1,1.6,2.0,max(EngineSize)+1),labels=c("Small","Mid","Large")))
counts <- table(p$CarSize) %>% as.data.frame() %>% mutate(pct=percent(Freq/sum(Freq)))
names(counts) <- c("CarSize","Freq","pct")

ggplot(counts,aes(x=CarSize,y=Freq)) + geom_bar(stat="identity")
ggplot(counts,aes(x="",y=Freq,fill=CarSize)) + geom_bar(stat="identity",width=1) + coord_polar(theta="y") +
  geom_text(aes(label=pct),size=5,position=position_stack(vjust=0.5))
ggplot(counts,aes(x=Freq,y=CarSize)) + geom_point() + xlab("") +
  theme(panel.grid.major.y=element_line(linetype=2,color="darkgray"))
```



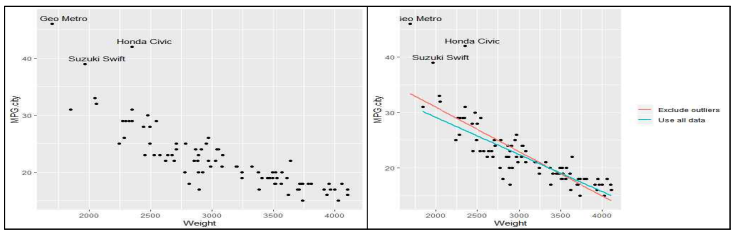
일변량 연습문제 2

```
ggplot(Cars93,aes(x="",y=MPG.city)) + geom_boxplot() +
  stat_summary(fun.y="mean",geom="point",col="red",shape=20,size=5) +
  coord_flip()

Cars93 %>% filter(MPG.city>35) %>% select(Manufacturer,Model,MPG.city,Weight) %>% arrange(desc(Weight))
```



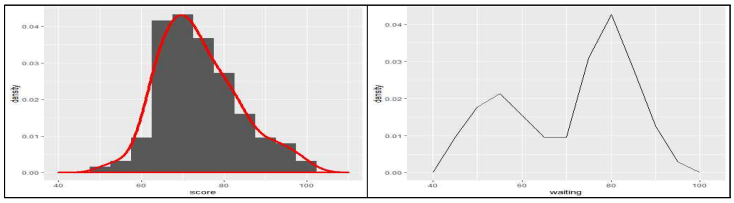
	Manufacturer	Model	MPG.city	Weight
1	Honda	Civic	42	2350
2	Suzuki	Swift	39	1965
3	Geo	Metro	46	1695

```
a <- ggplot(Cars93, aes(x=Weight,y=MPG.city)) + geom_point() +
  geom_text(data=filter(Cars93,MPG.city>35),
            aes(label=paste(Manufacturer,Model)),nudge_x=100,nudge_y=1)

a + geom_smooth(aes(col="Exclude outliers"),method="lm",se=FALSE) +
  geom_smooth(data=filter(Cars93,MPG.city<35),aes(col="Use all data"),method="lm",se=FALSE) + labs(col="")
```



일변량 연습문제 3

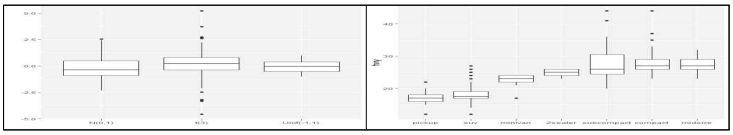
```
set.seed(1234)
score <- rnorm(125,mean=75,sd=10)
score <- as.data.frame(score) %>% mutate(score=if_else(score>100,100,score)) %>% round()
ggplot(score, aes(x=score,y=..density..)) + geom_histogram(binwidth=5) +
  geom_density(col="red",size=1.5) + xlim(40,110)
```



```
score <- mutate(score, grade=cut(score,breaks=c(max(score)+1,quantile(score,0.8),quantile(score,0.5),
                                                quantile(score,0.2),quantile(score,0.05),min(score)-1),
                                 labels=c("A","B","C","D","F")))
table(score$grade) %>% prop.table() %>% round(2)

   A    B    C    D    F
0.06 0.14 0.36 0.26 0.18
```

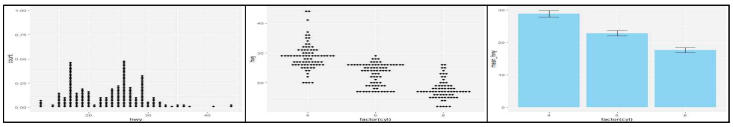
이변량 상자그림

```
set.seed(1234)
x1 <- rnorm(100)
x2 <- rt(100,df=3)
x3 <- runif(100,min=-1,max=1)
data.frame(x=c(rep(1:3,each=100)),y=c(x1,x2,x3)) %>% ggplot(aes(factor(x),y)) +
  geom_boxplot() +
  scale_x_discrete(labels=c("N(0,1)","t(3)","Unif(-1,1)")) +
  labs(x="",y="")
ggplot(mpg,aes(x=reorder(class,hwy,FUN=median),y=hwy)) +
  geom_boxplot() + labs(x="")
```



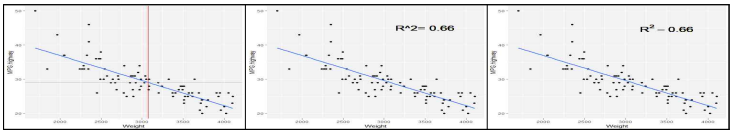
다중 점 그래프 & # 평균막대그래프와 error bar

```
ggplot(mpg, aes(x=hwy)) + geom_dotplot(binwidth=0.5)
ggplot(filter(mpg,cyl!=5), aes(x=factor(cyl),y=hwy)) +
  geom_dotplot(binaxis="y",binwidth=0.5,stackdir="center")

hwy_stat <- filter(mpg,cyl!=5) %>% group_by(cyl) %>%
  summarise(mean_hwy=mean(hwy),sd_hwy=sd(hwy),n_hwy=n(),
            ci_low=mean_hwy-qt(0.975,df=n_hwy-1)*sd_hwy/sqrt(n_hwy),
            ci_up=mean_hwy+qt(0.975,df=n_hwy-1)*sd_hwy/sqrt(n_hwy))
ggplot(hwy_stat,aes(x=factor(cyl),y=mean_hwy)) +
  geom_col(fill="skyblue") +
  geom_errorbar(aes(ymin=ci_low,ymax=ci_up),width=0.3)

filter(mpg,cyl!=5) %>% ggplot(aes(x=factor(cyl),y=hwy)) +
  stat_summary(fun.y="mean",geom="bar",fill="skyblue") +
  stat_summary(fun.data="mean_cl_normal",geom="errorbar",width=0.3)
```



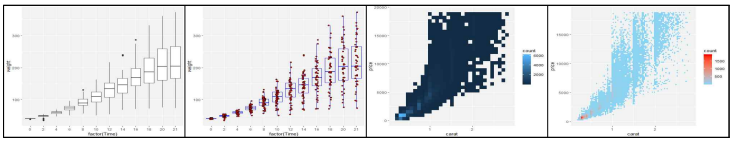
산점도

```
data(Cars93, package="MASS")
ggplot(Cars93,aes(x=Weight,y=MPG.highway)) + geom_point() +
  geom_smooth(method="lm",se=FALSE) +
  geom_vline(aes(xintercept=mean(Weight)),col="red") +
  geom_hline(aes(yintercept=mean(MPG.highway)),col="darkgray")

fit <- lm(MPG.highway~Weight,Cars93)
r2 <- round(summary(fit)$r.squared,2)
pp <- ggplot(Cars93,aes(x=Weight,y=MPG.highway)) + geom_point() +
  geom_smooth(method="lm",se=FALSE)
pp + geom_text(x=3500,y=45,size=7,label=paste("R^2=",r2))
pp + geom_text(x=3500,y=45,size=7,label=paste("R^2==",r2),parse=TRUE)
```



산점도과 boxplot & 2차원 히스토그램

```
ggplot(ChickWeight, aes(x=factor(Time),y=weight)) + geom_boxplot()
ggplot(ChickWeight, aes(x=factor(Time),y=weight)) +
  geom_boxplot(outlier.shape=NA,fill=NA,col="blue") +
  geom_jitter(width=0.1,shape=21, fill="red")

a <- ggplot(filter(diamonds,carat<3),aes(x=carat,y=price))
a + geom_bin2d()
a + geom_bin2d(bins=100) + scale_fill_gradient(low="skyblue",high="red")
```

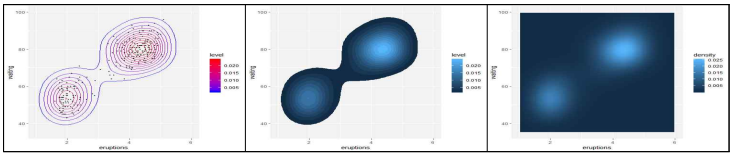


연속형 변수를 범주형 변수로 변환

```
cut_width(x, width, boundary) : 동일간격으로 구분, boundary는 시작점 지정
cut_number(x, number=n) : n개의 구간으로 구분, 각 구간에 속한 자료의 개수는 동일
cut_interval(x, n, length) : n개의 구간으로 구분하되, 구간의 길이는 동일
```

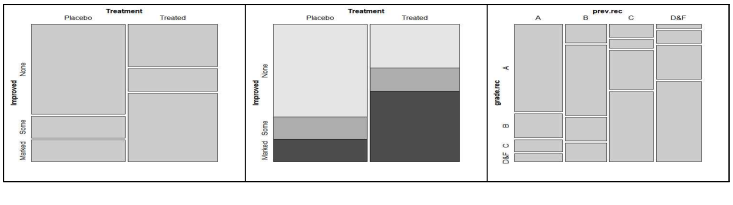
등고선

```
a <- ggplot(faithful, aes(x=eruptions,y=waiting)) + xlim(1,6) + ylim(35,100)
a + geom_density_2d(aes(color=..level..)) +
  scale_color_gradient(low="blue",high="red") + geom_point(shape=20)
a + stat_density2d(aes(fill=..level..),geom="polygon")
a + stat_density_2d(aes(fill=..density..),geom="raster",contour=FALSE)
```



Mosaic plot

```
library(vcd)
my_table <- with(Arthritis, table(Treatment,Improved))
mosaic(my_table, direction="v")
mosaic(~ Treatment + Improved, data=Arthritis, direction="v")
mosaic(Improved ~ Treatment, data=Arthritis, direction="v")
mosaic(grade, direction="v")
```

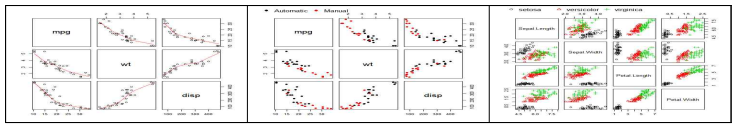


Outlier

```
a <- ggplot(alltime.movies,aes(x="",y=Gross)) + geom_boxplot()
ggplot_build(a)[[1]]
my_out <- ggplot_build(a)[[1]][[1]]$outliers[[1]]
alltime <- as_tibble(alltime.movies) %>% rownames_to_column(var="Movie.Title")
top_moives <- alltime %>% filter(Gross%in%my_out)
```

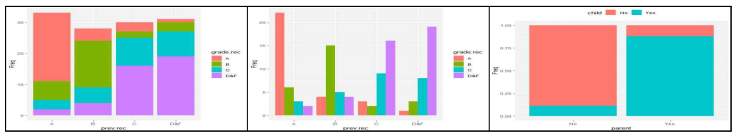
산점도 행렬

```
pairs(~mpg+wt+disp,mtcars,panel=panel.smooth)
my_panel_1 <- function(x,y) points(x,y,col=mtcars$am+1,pch=16)
pairs(~mpg+wt+disp,mtcars,panel=my_panel_1)
legend("topleft",c("Automatic","Manual"),pch=16,col=c(1,2),
       xpd=TRUE,horiz=TRUE,bty="n",y.intersp=-1)

iris.panel <- function(x,y){ points(x, y, col=iris$Species, pch=as.numeric(iris$Species)) }
pairs(iris[1:4], panel=iris.panel)
legend("topleft", as.character(unique(iris$Species)), pch=unique(iris$Species), col=unique(iris$Species),
       bty="n", horiz=TRUE, xpd=TRUE, y.intersp=-2)
```



이변량 연습문제 1,2

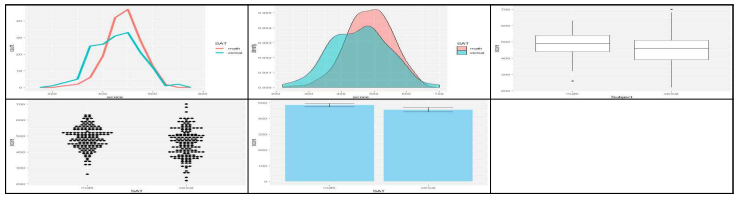
```
data(grades,package="UsingR")
with(grades, table(prev,grade))
grade <- mutate(grades,prev.rec=factor(prev, labels=c("A","A","B","B","B","C","C","D&F","D&F")),
                       grade.rec=factor(grade, labels=c("A","A","B","B","B","C","C","D&F","D&F"))) %>%
  with(., table(prev.rec,grade.rec))
addmargins(grade)
round(prop.table(grade,1),2)

ggplot(as.data.frame(grade),aes(x=prev.rec,y=Freq,fill=grade.rec)) + geom_col()
ggplot(as.data.frame(grade),aes(x=prev.rec,y=Freq,fill=grade.rec)) + geom_col(position="dodge")
ggplot(df_1,aes(x=parent,y=Freq,fill=child)) + geom_col(position="fill") + theme(legend.position="top")
```



```
belt <- matrix(c(58,8,2,16), nrow=2, ncol=2)
dimnames(belt) <- list(parent=c("Yes","No"),child=c("Yes","No"))
df_1 <- data.frame(parent=c("Yes","Yes","No","No"), child=c("Yes","No","Yes","No"), Freq=c(58,8,2,16))
```

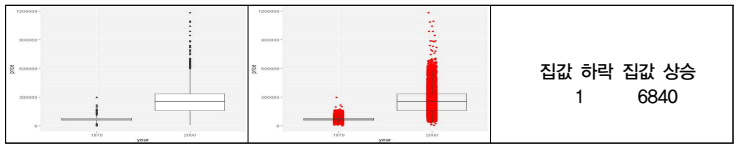
이변량 연습문제 1,2

```
data(stud.recs, package="UsingR")
stud.recs.df <- data.frame(SAT=rep(c("math","verbal"),each=160),
                           score=rep(c(stud.recs$sat.m,stud.recs$sat.v)))
ggplot(stud.recs.df,aes(x=score,col=SAT)) + geom_freqpoly(binwidth=50,size=2)
ggplot(stud.recs.df,aes(x=score,fill=SAT)) + geom_density(binwidth=50,alpha=0.5)
ggplot(stud.recs.df,aes(x=SAT,y=score)) + geom_boxplot() + xlab("Subject")
ggplot(stud.recs.df,aes(x=SAT,y=score)) + geom_dotplot(binaxis="y",binwidth=10,stackdir="center")
ggplot(stud.recs.df,aes(x=SAT,y=score)) +
  stat_summary(fun.y="mean", geom="bar", fill="skyblue") +
  stat_summary(fun.data="mean_cl_normal", geom="errorbar",width=0.5)
```



13장 연습문제 4번

```
data(homedata, package="UsingR")
homedata.df <- data.frame(year=rep(c("1970","2000"),each=6841),
                          price=c(homedata$y1970,homedata$y2000))
ggplot(homedata.df,aes(x=year,y=price)) + geom_boxplot()
ggplot(homedata.df,aes(x=year,y=price)) +  geom_jitter(col="red",width=0.05) +
  geom_boxplot(outlier.shape=NA,fill=NA)

homedata_t <- mutate(homedata,group=if_else(y2000-y1970<0,"집값 하락","집값 상승"),
                              group=factor(group))
levels(homedata_t$group)=list("집값 하락"=1,"집값 상승"=2)
table(homedata_t$group)
```



```
집값 하락 집값 상승
        1       6840
```

iris 산점도

```
# 대각선 위아래 패널 : 옵션 upper, lower
upper=list(continuous=, combo=, discrete=)
lower=list(continuous=, combo=, discrete=)
continuous : "points","smooth","smooth_loess","density","cor","blank"
combo : "box","dot","facethist","facetdensity","denstrip","blank"
discrete : facetbar,"ratio","blank"
# 대각선 패널 : 옵션 diag
diag=list(continous=, discrete=)
continous : "densityDiag","barDiag","blankDiag"
discrete : "barDiag","blankDiag"

library(GGally)
mtcars %>% dplyr::select(mpg,wt,disp,cyl,am) %>%
  mutate(cyl=factor(cyl),am=factor(am)) %>%
  ggpairs(aes(color=am),lower=list(continuous=wrap("smooth",se=FALSE),combo=wrap("facethist",bins=10)))

# iris 활용하기
iris %>% ggpairs(aes(color=Species),upper=list(continuous="blank",combo="blank"),lower=list(combo=wrap("facethist",bins=20)))
```