

산점도

📖 산점도와 plot 함수

📖 산점도 및 함수 그리기

Lorem ipsum dolor sit amet, ius an molestie facilisi erroribus, mutat natorum delectus ei vis. Has ornatus conclusionemque id, an vide molestatis sit. In etqui praesent sit. An vel agan porro comprehensan, ad ludus constituto nea, et ius utroque scaevola assuaverit.

Vis cu nodus nulla feugait, oratio facilisi ex usu, eilit vitae sea te. Ea fabulas accusamus dissonantia sea, facete tacinates definitiones et per. Nihil dicant mediocrem pro eu, no mei nostro sensibus platonem. Qui id sunno perpetua neglegantur. Vel ipsum novum copiosae ut. Quo et liber detracto probatus. Nam augue scribentur an. Sea oporteat percipitur incidere at. Qui viris nemore an.



산점도와 plot 함수



산점도(scatterplot)

- + $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ 의 짝으로 얻어진 자료에 (x_i, y_i) 대해 좌표평면에 x-축 값이 x_i , y-축 값이 y_i 인 곳에 점으로 표시하는 그림
- + n 개의 짝이 있으면 산점도엔 n 개의 점이 표시되며 이 점들에 의해 두 변수 x 와 y 사이의 관계를 직관적으로 확인 가능



산점도와 plot 함수



plot 함수



사용법

```
plot(x, y, type=, main=, sub=, pch=, lty=, lwd=, xlab=, ylab=, ...)
```



매개변수

➤ **x** : x-좌표값 x_1, x_2, \dots, x_n 으로 이루어진 벡터를 설정하거나 행렬, data.frame 등의 객체를 설정함

- x에 행렬을 설정하면 첫 번째 열이 가로축, 두 번째 열이 세로축인 산점도를 그림
- data.frame인 경우 matplot 함수를 사용한 모든 가능한 산점도를 그림

➤ **y** : y_1, y_2, \dots, y_n 을 원소로 갖는 벡터를 지정하며 x가 행렬 또는 plot 함수를 적용할 수 있는 객체인 경우 생략이 가능함

➤ $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ 좌표에 점을 그리는 것이 기본이나 선을 그릴 수 있음





산점도와 plot 함수



plot 함수



매개변수

- type : 산점도의 종류를 설정함

type에 설정 가능한 값

- "p" 좌표 (x_i, y_i) 에 점을 그림(points)
- "l" (x_i, y_i) 와 (x_{i+1}, y_{i+1}) 을 연결하는 선을 그림(line)
- "b" 점과 선을 모두 사용함. 이 때 점과 선은 연결되지 않고 점과 점 사이에 선분을 그림(both)
- "c" "b"를 사용한 경우에서 점만 제외함
- "o"는 "b"와 마찬가지로 점과 선을 모두 사용하나 점과 선이 겹치게 그림
- "n" : 점이나 선을 그리지 않고 그림 영역만 확보함. 주로 그림영역을 확보한 후 추가로 그림이나 문자 등을 넣고자 할 때 사용함

- main, sub : 주 제목과 보조제목에 해당하는 문자열을 지정함
- xlab, ylab : 각각 x축 및 y축 이름을 설정함





산점도 및 함수 그리기



예시



자료를 사용한 산점도 그리기 및 함수그리기

1. BMI 자료를 사용하여 키와 몸무게의 산점도
2. 함수 $y = f(x) = x^2$ 을 $-10 \leq x \leq 10$ 에서 그려보자.

사용할 자료

```
> BMI <- read.table(url("http://jupiter.hallym.ac.kr/ftpdata/data/bmi.txt"),  
  col.names=c("height", "weight", "year", "religion", "gender", "marriage"))
```



내용 : 2000년, 177명에 대한 조사 결과

- 키, 몸무게, 출생년도
- 종교(Bu=불교, C1=개신교, C2=가톨릭, No=없음)
- 성별(F=여자, M=남자)
- 결혼여부(N=미혼, Y=기혼)



산점도 및 함수 그리기



예시

+ plot 함수 사용

```
par(mfrow=c(2,2))  
plot(BMI$weight, BMI$height, main="키와 몸무게")  
plot(x <- seq(-10, 10, by=0.2), y <- x^2, xlab="x", ylab="", main="y=x^2")  
plot(x <- seq(-10, 10, by=0.2), y <- x^2, xlab="x", ylab="", main="y=x^2", type="l")  
plot(x <- seq(-10, 10, by=0.2), y <- x^2, xlab="x", ylab="", main="y=x^2", type="o")
```



산점도 및 함수 그리기

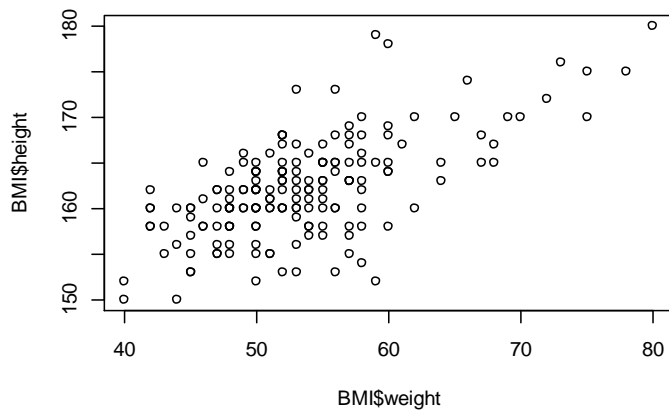


예시

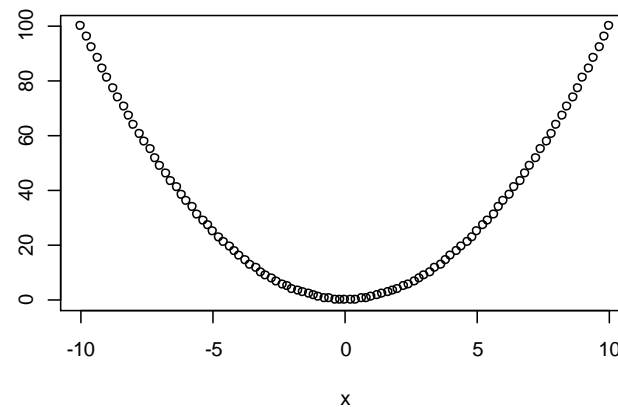


plot 함수 사용 결과

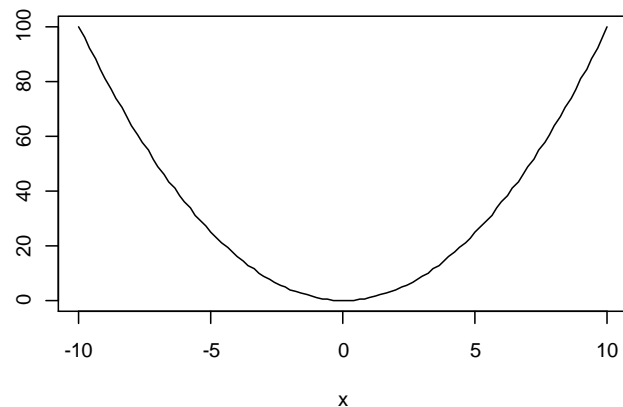
키와 몸무게



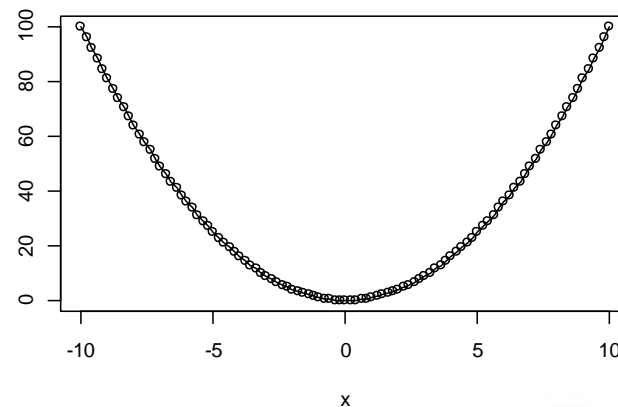
$y=x^2$



$y=x^2$



$y=x^2$



행렬산점도

📖 행렬산점도와 pairs 함수

📖 행렬산점도 그리기

Lorem ipsum dolor sit amet, ius an molestie facilisi erroribus, mutat natorum delectus ei vis. Has ornatus conclusionemque id, an videri molestatis sit. In etqui praesent sit. An vel agan porro comprehensan, ad ludus constituto nea, et ius utroque scaevola assumaverit.

Vis cu nodus nulla feugait, oratio facilisi in usu, eili vitae sea te. Ea fabulas accusamus dissonantia sea, facete tacinates definitiones et per. Nihil dicant mediocrem pro eu, no mei nostro sensibus platonem. Qui id sunno perpetua neglegentur. Vel ipsum novum copiosae ut. Quo et liber detracto probatus. Nam augue scribentur an. Sea oporteat percipitur incidereit at. Qui viris nemore an.





행렬산점도와 pairs 함수



행렬산점도

- + 두 개 이상의 변수에 대해, 모든 가능한 산점도를 그리는 것
- + 만일 변수가 $x1$, $x2$, $x3$ 가 있다면 $x1$ 과 $x2$, $x1$ 과 $x3$, $x2$ 와 $x3$ 의 산점도 및 이들의 역순 $x2$ 와 $x1$ (이 산점도는 $x1$ 과 $x2$ 의 산점도에서 x - y 축을 바꾼 것으로 결과로 얻는 산점도는 원래 산점도를 직선 $y = x$ 에 대해 대칭이동한 것이 됨), $x3$ 과 $x1$ 등의 산점도를 모두 그리는 것
- + 이 함수는 그림영역 전체를 내부적으로 두 부분으로 나누어 사용하므로 mfrow나 mfcol 등은 적용되지 않음





행렬산점도와 pairs 함수



pairs 함수



사용법

```
pairs(formula, data = NULL, subset, labels=, lower.panel=,  
      upper.panel=, diag.panel=..., )
```

```
pairs(x, labels, row1 at top ...)
```



매개변수

➤ x : 행렬 산점도를 그릴 자료를 포함한 행렬

- x의 i번째 열 x[,i]와 j번째 열 x[,j]의 산점도가 행렬 산점도의 (i,j)번째에 위치함
- x에 data.frame이 설정될 수도 있음
- 논릿값이나 factor 형식의 값은 숫자로 변환되어 산점도를 그림



행렬산점도와 pairs 함수



pairs 함수



매개변수

- formula : R-언어의 모형식(formula)를 설정함
 - $\sim x + y + z$ 로 설정하면 세 변수 x , y , 및 z 에 대한 행렬산점도가 만들어짐
 - 종속변수가 설정된 형태 $s \sim x + y + z$ 로 설정하면 종속변수와 독립변수 구별없이 네 변수 s , x , y , 및 z 의 산점도가 생성됨
- data : formula를 사용할 경우 formula에 사용한 변수가 포함된 data.frame 또는 list의 이름을 설정함
- subset : 자료의 전부가 아닌 일부만 사용하고자 할 때 자료의 선택 기준을 설정함
 - 만일 자료에서 여자인 경우만 선택할 경우 `subset=(gender == "F")`과 같이 조건문을 설정하면 됨
- labels : 대각선에 들어갈 각 변수의 이름을 설정함(기본값 : 변수의 이름)
- row1atop : TRUE/FALSE로 설정하며 T인 경우(기본값) 변수의 순서대로 행렬 산점도를 그리고 F인 경우 역순으로 그림(이 때 각 변수의 설명도 바뀜)
- lower.panel, upper.panel, diag.panel : 대각선 기준으로 아래와 위에 사용할 함수를 설정할 수 있음
 - NULL을 설정하면 해당 부분의 산점도/문자열은 생략됨



행렬산점도 그리기



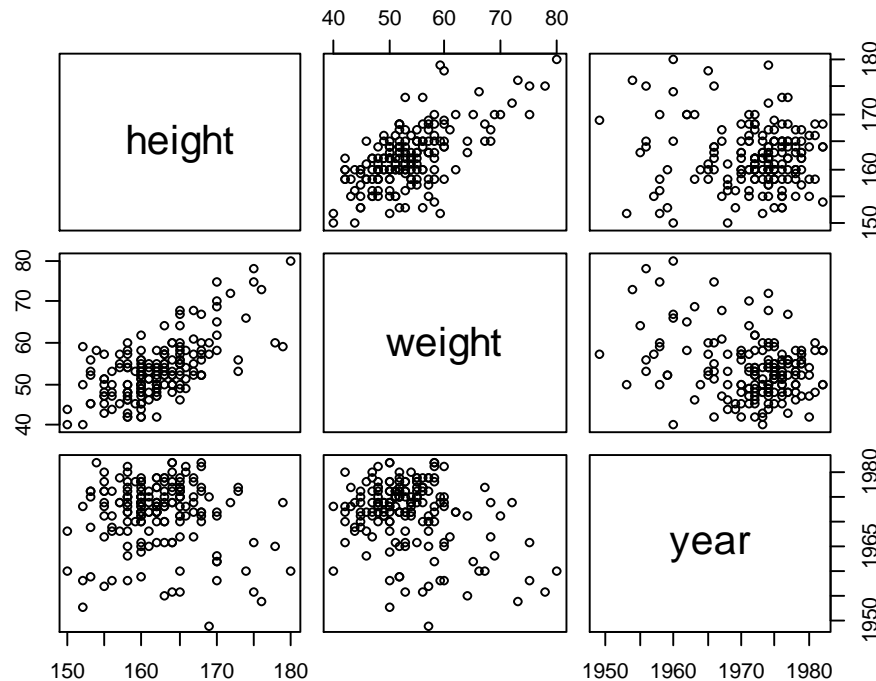
기본값에 의한 결과



자료

키 몸무게 자료에 BMI에서 키, 몸무게, 출생연도에 대한 모든 가능한 산점도를 그려보자.

```
> pairs(~height+weight+year, data=BMI)
```



- 모든 옵션이 기본값으로 설정된 결과
- 변수 이름이 순서대로 height, weight, year로 나타나며 변수이름은 왼쪽 위부터 오른쪽 아래로 내려감

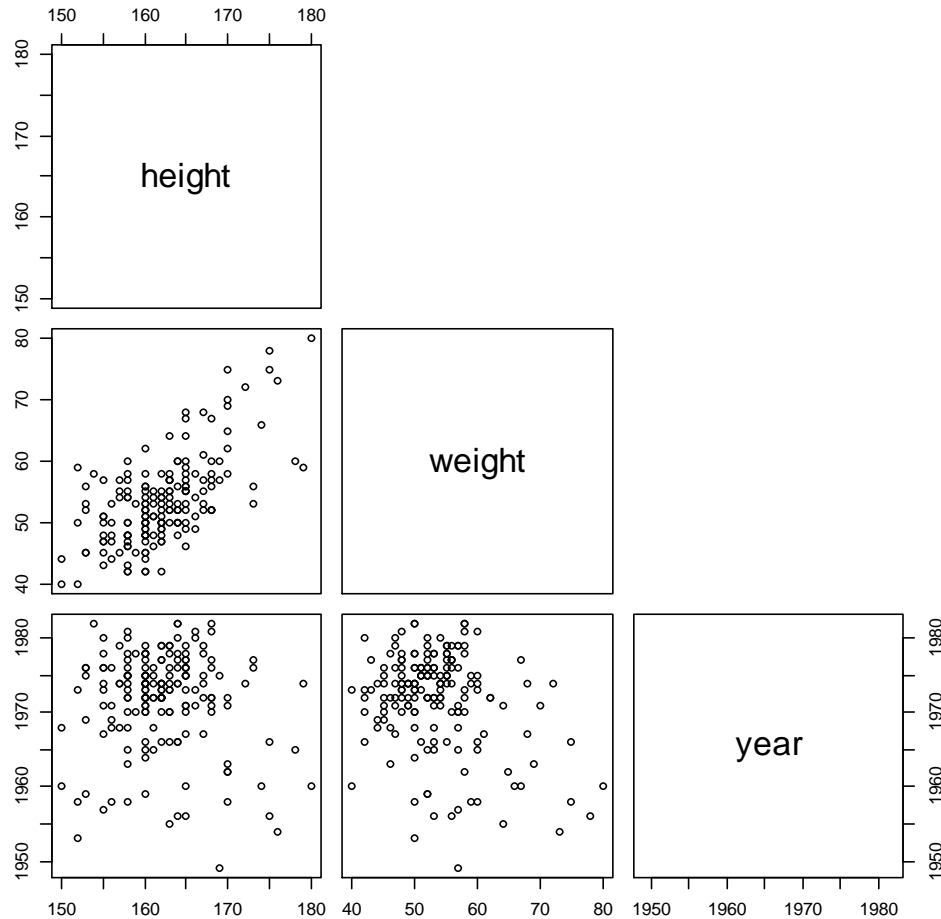


행렬산점도 그리기



row1atop=F 설정 결과

```
> pairs(~height+weight+year, data=BMI, upper.panel=NULL)
```



- upper.panel = NULL 옵션에 의해 대각선 위의 그림은 없음



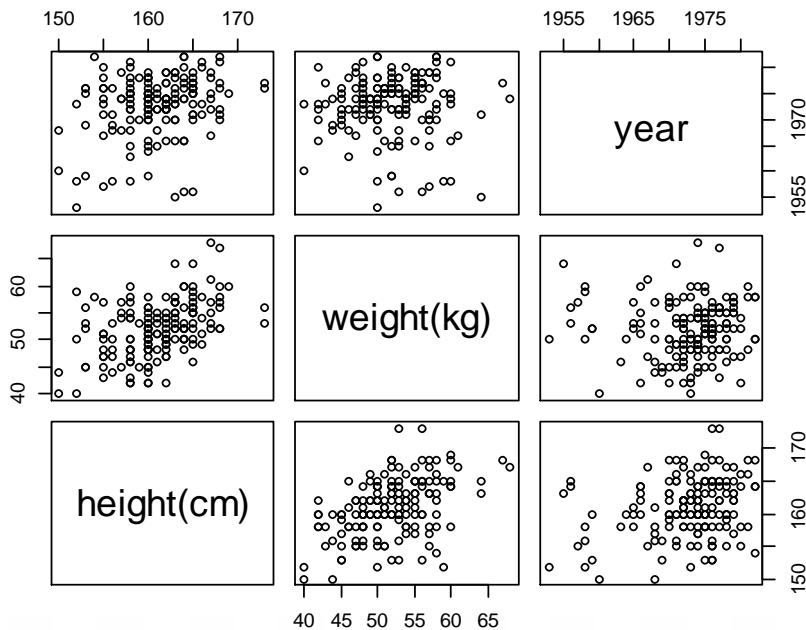


행렬산점도 그리기



pairs.test() 실행 결과

```
pairs.test <- function() {  
  pairs(~height+weight+year, data=BMI, subset=gender == "F",  
        labels=c("height(cm)", "weight(kg)", "year"), row1atop=F)  
} # end function  
> pairs.test()
```



- 자료는 subset 속성의 설정에 의해 성별이 여자인 경우만 산점도를 얻었음
- 반 대각선에 있는 이름은 labels에서 설정한 이름들임
- row1atop=F 옵션에 의해 대각선의 방향도 바뀜



행렬산점도 그리기



대각선에 함수를 사용하여 결과 인쇄하기



대각선에 함수를 사용하여 결과인쇄

앞의 결과를 보면 대각선엔 문자열이 들어가지만 함수의 결과(출력)를 대각선에 입력하는 다음의 방법을 생각해보자. (diag.panel에 사용자 작성 함수 지정)

```
my.panel.text <- function(x, ...) {  
  mx <- as.character(round(mean(x), 1))  
  mystr <- paste("평균 Wn", mx, sep="")  
  h <- hist(x, plot=F)  
  xloc <- mean(h$breaks);  
  yloc <- xloc  
  text(xloc, yloc, mystr, cex=2, col=2)  
}
```

```
pairs(~height+weight+year, data=BMI, diag.panel=my.panel.text)
```




행렬산점도 그리기



대각선에 함수를 사용하여 결과 인쇄하기



my.panel.text 함수는 매개변수 x 를 받아서

- x 의 평균을 소수점 1자리까지만 문자열로 변환하고
- “평균”이라는 문자열을 추가하여 mystr이라 함
- hist 함수를 호출하되 그림은 그리지 않고 출력만 받아 h라고 하고
- 히스토그램의 각 기둥의 x -좌표들의 평균을 xloc, yloc이라 함
- 이제 좌표 (xloc, yloc)에 문자열을 그림



pairs 함수에서 diag.panel에 my.panel.text 함수를 설정하여

- 이 함수의 결과인 평균을 빨간색으로 대각선에 인쇄함



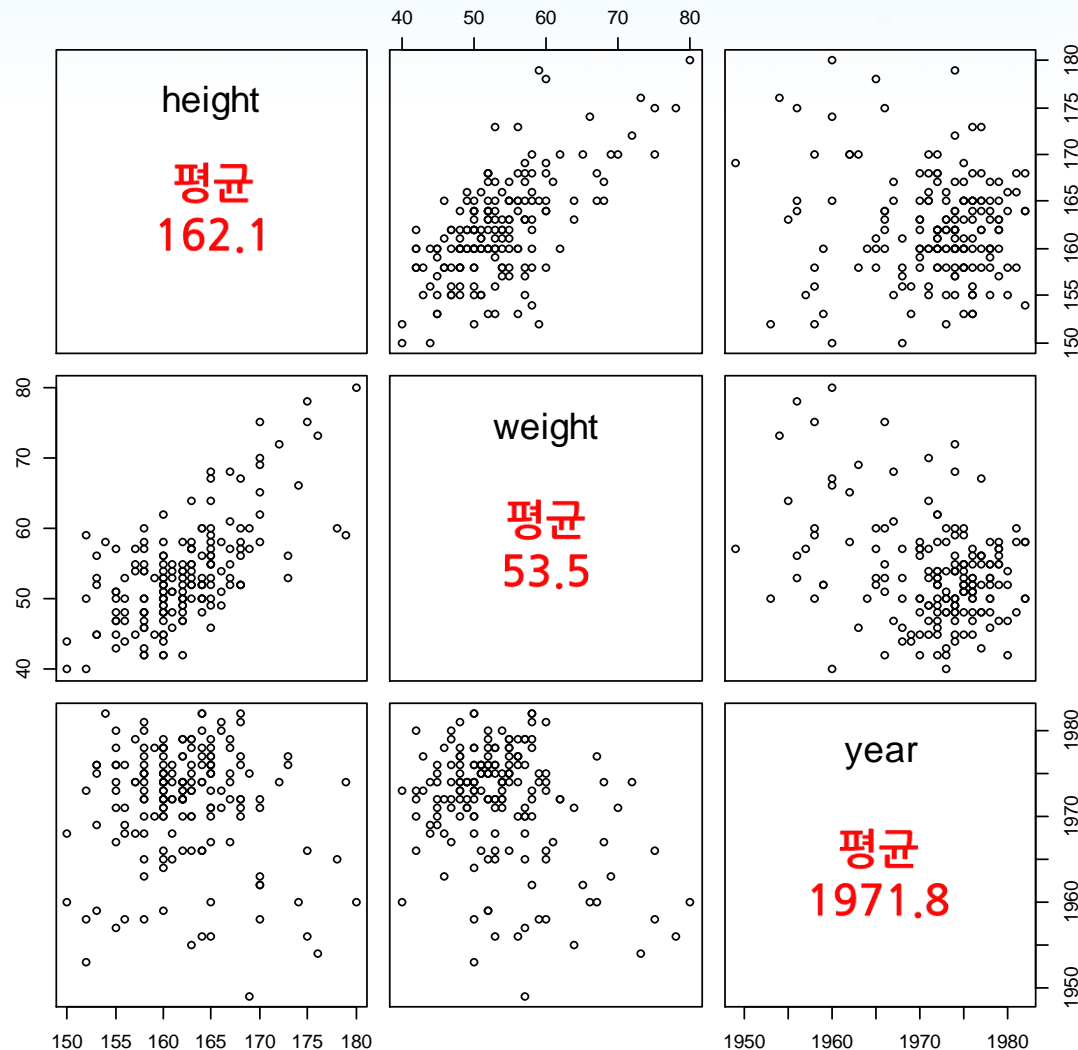
행렬산점도 그리기



대각선에 함수를 사용하여 결과 인쇄하기



함수 실행 결과



두 개 이상의 산점도 겹치기

📖 matplotlib 함수

📖 중복한 산점도 또는 함수 그리기

matplot

정의

- + 주어진 하나의 x 벡터 값에 대해서 두 개 이상의 y 벡터에 대한 산점도를 그리는 함수

사용법

```
matplot(x, y, type = "p", lty = 1:5, lwd = 1, pch = NULL, col = 1:6,  
        xlab = NULL, ylab = NULL, xlim = NULL, ylim = NULL, ...)
```

+ 매개변수

- x, y : 산점도를 그리기 위한 벡터 또는 행렬의 값이며 행의 개수는 모두 같아야 함
 - 만일 둘 중에 하나만 주면 주어진 값은 y로 해석하고 x는 1:n으로 설정하여 matplot을 그림
- type : 산점도의 종류로 점을 사용할 때는 "p", 선을 사용할 때는 "l", 점과 선을 모두 사용하면 "pl"을 의미하며 이 세 값들을 원소로 하는 벡터가 설정되면 각 산점에 대해서 해당되는 방법으로 산점도를 그림
- pch, lty, lwd 옵션 : 점이나 선의 모양을 설정하는 것

matplot 함수와 중복된 산점도 / 함수



예시 1



두 함수를 겹쳐 그리기

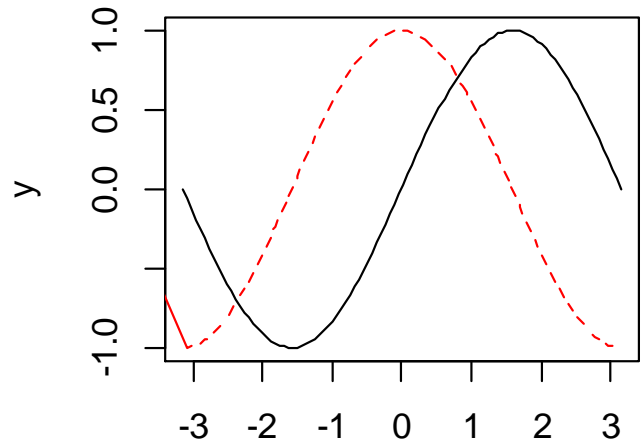
- x 가 $-\pi$ 부터 π 사이의 100개의 값일 때 $\sin(x)$ 와 $\cos(x)$ 를 한 그림에 겹쳐 그린 보기임.
- y 는 x 의 100개의 값을 $\sin(x)$ 와 $\cos(x)$ 의 값을 각각 첫 번째 및 두 번째 열로 갖는 행렬.

```
matplot.test <- function() {  
  par(mfrow=c(1,2))  
  x <- seq(-pi,pi, length=100); # -  $\pi$ 부터  $\pi$ 사이의 100개의 값을 벡터로  
  y <- cbind(sin(x), cos(x)) # x에서 sin 및 cos 함수의 값 계산  
  matplot(x, y, type="l") # x에서 sin과 cos함수 그리기  
  matplot(y, type="p", pch=16, main="한 값만 입력")  
  # y만 입력된 경우 x는 (1,100)  
} # end function
```

matplotlib 함수와 중복된 산점도 / 함수

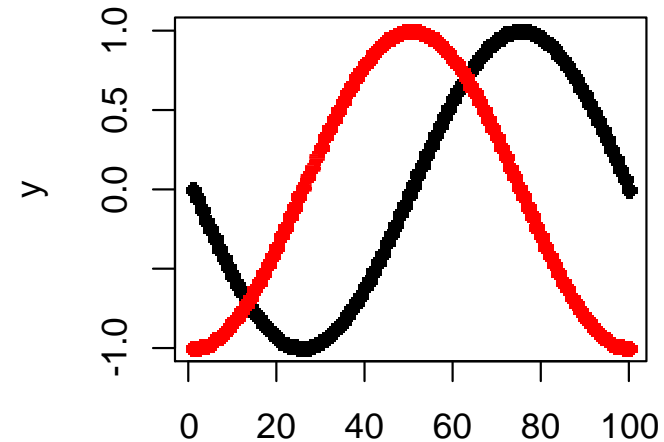
예시 1

+ matplotlib.test()로 호출한 결과



- 기본 설정의 그림

한 값만 입력



- pch를 16번으로 설정한 그림
- x의 입력을 생략하였으므로 x가 1부터 100사이인 것으로 간주됨

matplot 함수와 중복된 산점도 / 함수

예시 2



2006-2014년의 GRDP 변화 보기

- 지역별(16개 시도 2006-2014)까지의 GRDP자료를 사용하여 광역시만 GRDP 값을 연도별 그려보자.
- xlsx 패키지는 R 기초과정에서 다룬 내용으로 엑셀에 대한 여러 가지 함수를 제공하는 패키지로 이 패키지가 제공하는 함수 read.xlsx 는 엑셀 자료를 R로 읽어 들이는 함수
- 이 자료는 국가통계포털 KOSIS(<http://kosis.kr/>)에서 얻을 수 있음

```
library(xlsx)
```

```
df <- read.xlsx("c:/hwp/STI/grdp.xlsx", startRow=2, encoding="UTF-8",  
               sheetIndex = 1)
```

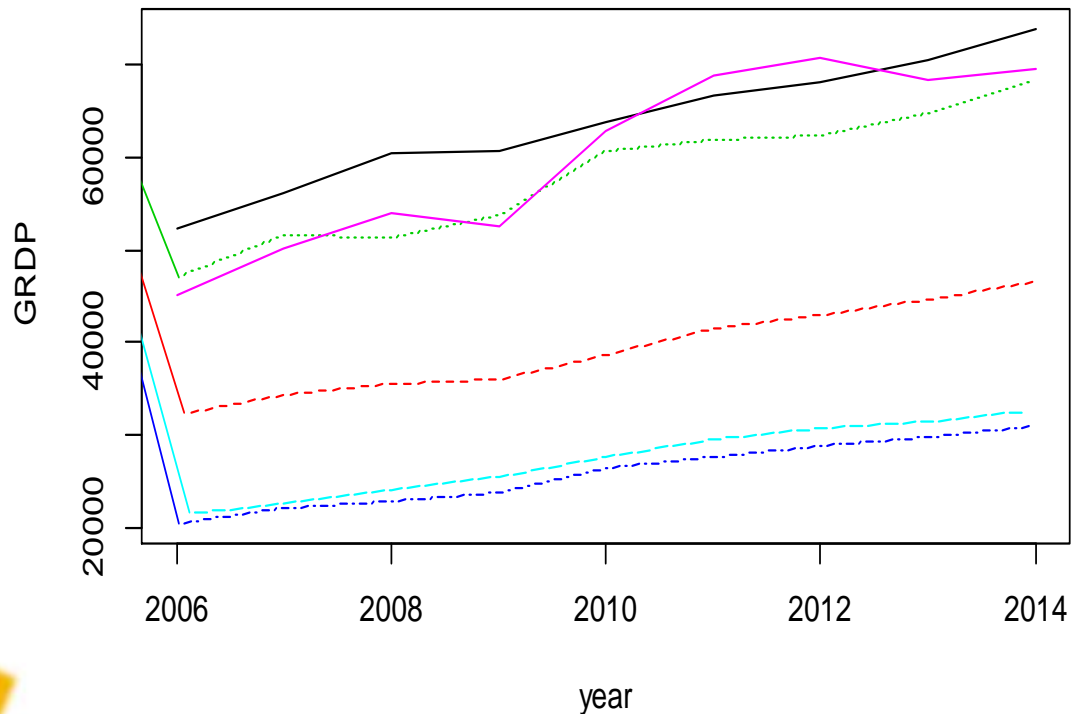
	NA	X2006	X2007	X2008	X2009	X2010	X2011	X2012	X2013	X2014
1 전국		966660	1043255	1105722	1151367	1265146	1330888	1377041	1430255	1484542
2 서울		231224	249485	263000	273199	289719	303813	313479	318607	327602
3 부산		52358	56193	60467	60695	63737	66648	67999	70338	73744
4 대구		32276	34388	35465	36017	38580	41448	43021	44754	46592
5 인천		47055	51638	51274	53796	60708	61854	62208	64654	68374
6 광주		20614	22310	22940	23834	26401	27789	28914	29763	30998

matplotlib 함수와 중복된 산점도 / 함수

예시 2

+ 패키지에서 제공하는 read.xlsx 함수 사용 결과

```
matplotlib(seq(2006,2014), t(df[3:8,-1]), type="l", xlab="year", ylab="GRDP")
```



- seq(2006, 2014)는 2006년부터 2014년의 연도임
- t(df[3:8],-1)은 위 엑셀자료에서 첫 번째 열은 지역이름이므로 제외하고 3:8번 행이 광역시만의 자료이며, 이 자료는 원래 모양이 아닌 행과 열이 바뀐 형태라야 하므로 t 함수(transpose)로 행과 열을 바꾸었음
- xlab, ylab은 각각 x축과 y축의 이름설정임