

데이터과학을 위한 **R**프로그래밍

4주차. 데이터시각화



이혜선 교수

포항공과대학교 산업경영공학과



목차

4주차. 데이터시각화

1차시

R 그래픽 I (히스토그램)

2차시

R 그래픽 II (상자그림, 산점도)

3차시

R 그래픽 III (ggplot2 활용)

4차시

R 그래픽 IV (공간지도분석)

An isometric illustration of a data visualization workshop. In the center, a large white trapezoidal platform contains the text. Surrounding it are various interactive displays: a large screen on the left with multiple charts, a curved screen at the top with a person interacting, a 3D bar chart on a pedestal, a large screen on the right with a grid of charts and gears, and a smaller screen on the far right with a bar chart. Several stylized human figures are shown interacting with these displays. The background is a light blue gradient.

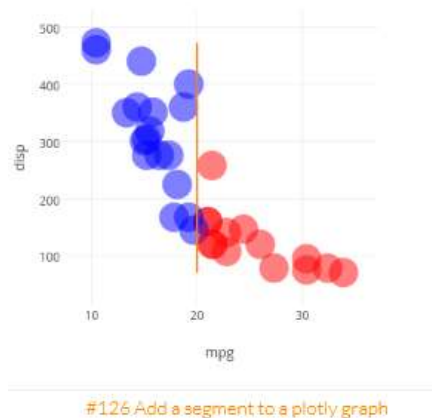
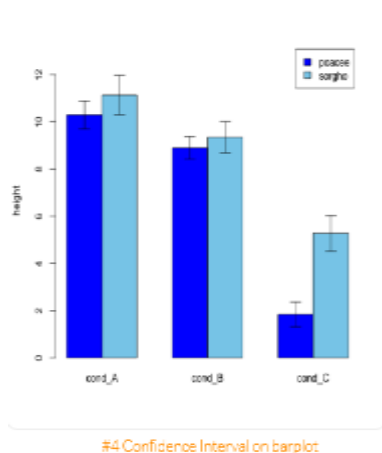
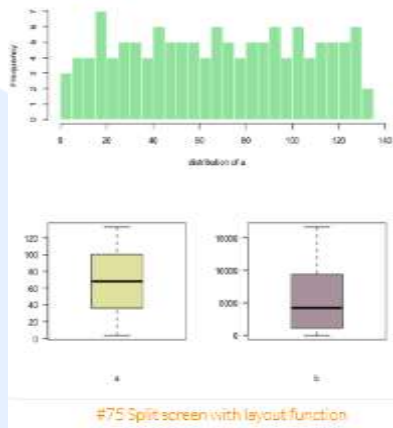
4주차

1차시

R 그래픽 I (히스토그램)

데이터의 시각화

- ✓ 데이터 시각화 : 정보의 요약된 형태를 그래프로 전달
- ✓ 빅데이터의 시각화를 통해 인과관계 발견
- ✓ 시각적 요약을 통해 인사이트 창출



● R 그래프 – 데이터의 분포

- ✓ 히스토그램 (histogram) : 1차원 (univariate, 일변량)
- ✓ 상자그림 (Boxplot) : 1차원 (데이터의 분포를 파악)
- ✓ 막대그림 (Bar plot) : 1차원 (범주형데이터의 빈도분포)
- ✓ 파이차트 (pie chart) : 1차원 (각 범주별 비율)
- ✓ 산점도 (scatterplot) : 2차원 (x와 y간의 관계를 해석)

R 그래픽 사이트 :

🔍 <http://www.r-graph-gallery.com/>

🔍 <http://www.cookbook-r.com/Graphs/>

● 그래프의 기본함수

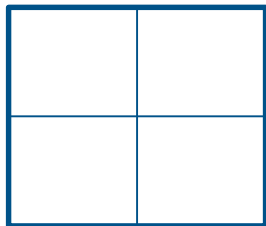
- ✓ 그래프 종류 : `plot()`, `barplot()`, `boxplot()`, `hist()`, `pie()`, `persp()`
- ✓ 그래프 구성시 조정사항 : 점, 선의 종류, 글자크기, 여백조정 등의 옵션을 조정
- ✓ 점그리기: `points()`
- ✓ 선그리기: `lines()`, `abline()`, `arrows()`
- ✓ 문자출력: `text()`
- ✓ 도형: `rect()`, `polygon()`
- ✓ 좌표축: `axis()`
- ✓ 격자표현: `grid()`

● 그래픽 옵션

- `par()` 그래프의 출력을 조정 - 그래프 화면의 분할, 마진, 글자 크기, 색상 등 설정
- `pty="s"` (x축과 y축을 동일비율로 설정, **s**quare) `pty="m"` (최대크기로 설정, **m**aximal)
- `legend = c("name1", "name2")`
- `bty="o"` (box type 그래프의 상자모양을 설정) o, l, 7, c, u
- `pch=1`(default) point character (1=동그라미, 2=세모, ..., 19=채운 동그라미)
- `lty=(solid가 default)` (line type, 1=직선, 2=점선)
- `lwd = 1, 2..` (선의 굵기)
- `cex=1`(default) (character expansion) 문자나 점의 크기, 숫자가 클수록 글자크기 커짐
- `mar` (아래, 왼쪽, 위쪽, 오른쪽)



`par(mfrow=c(2,2), mar=c(2,2,2,2))`



히스토그램 (1차원)

☑ Brain데이터를 이용한 그래프

```
# lec4_1.r : Basic Graphics I

# set working directory
# change working directory
setwd("D:/tempstore/moocr")

# Read brain data (lec3_1.R)
brain<-read.csv(file="brain2210.csv")

head(brain)
dim(brain)

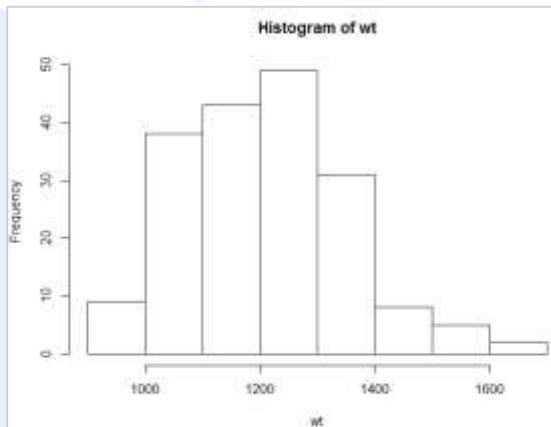
attach(brain)
```

데이터 불러들이기 단계 :
setwd, read.csv

히스토그램 (1차원)

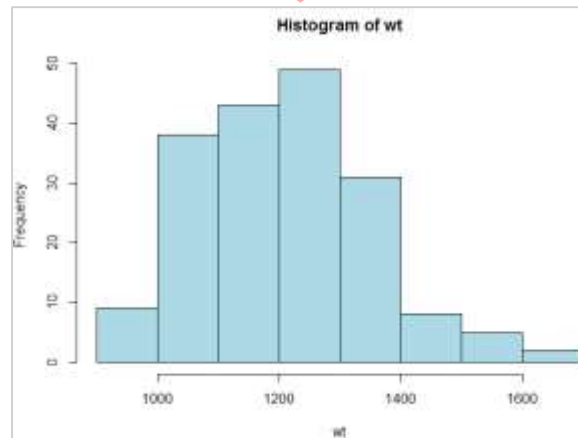
✓ 히스토그램 : hist(변수이름)

```
# 1. histogram  
  
# 1-1. histogram with no options  
# hist(brain$wt)  
hist(wt)
```



색상 선택

```
help(hist)  
  
hist(wt, col = "lightblue")
```



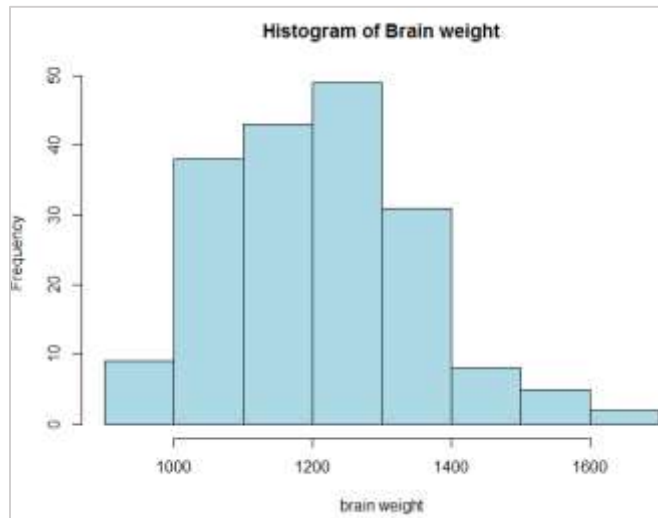
히스토그램 (1차원)

✓ 히스토그램 (색과 제목) : `hist(변수이름, col="colname", main=" ")`

```
# 1-2. histogram with color and title, legend  
hist(wt, breaks = 10, col = "lightblue", main="Histogram of Brain weight")
```

`col="colname", main="그림제목"`

예시 : `col="blue",
main="Histogram of weight"`



히스토그램 (1차원)

✓ 색 (657가지 색)

colors()

모든 색의 이름을 볼 수 있음

✓ grep("단어", colors(), value=TRUE) : '단어'가 포함된 색을 검색해 줌

```
# see rgb values for 657 colors, choose what you like
colors()

# select colors including "blue"
grep("blue", colors(), value=TRUE)
```

```
> # select colors including "blue"
> grep("blue", colors(), value=TRUE)
[1] "aliceblue"      "blue"           "blue1"
[4] "blue2"          "blue3"          "blue4"
[7] "blueviolet"     "cadetblue"      "cadetblue1"
[10] "cadetblue2"     "cadetblue3"     "cadetblue4"
[13] "cornflowerblue" "darkblue"        "darkslateblue"
[16] "deepskyblue"    "deepskyblue1"   "deepskyblue2"
[19] "deepskyblue3"   "deepskyblue4"   "dodgerblue"
```

히스토그램 (1차원)

✓ 밀도 함수 그려보기

```
# 1-3. fit function (find density function)  
par(mfrow=c(1,1))  
d <- density(brain$wt)  
plot(d)
```

