

R로 하는 시각화

# 목차

1. 기본 제공 함수 시각화

2. ggplot 시각화

3. ggpubr 시각화

4. PPT로 저장하기

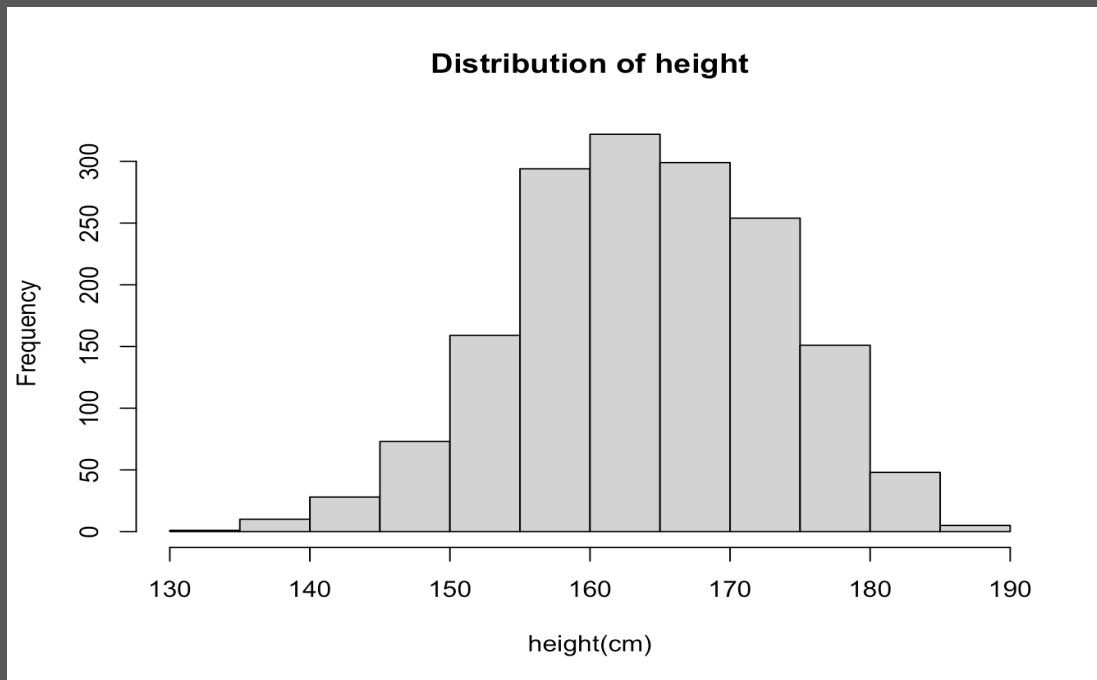
# 1. 기본 제공 함수 시각화

$\mathbb{R}$ 에서 기본적으로 제공하는 함수로  
간단한 그래프를 그릴 수 있다.

# 1. 기본 제공 함수 시각화

## Histogram - 연속형 데이터 시각화

```
hist(data$HGHT, main="Distribution of height", xlab="height(cm)")
```



# 1. 기본 제공 함수 시각화

## Histogram - 연속형 데이터 시각화

```
hist(data$HGHT, main="Distribution of height", xlab="height(cm)",  
      breaks = 30, freq=F, col="grey", border="white")
```

**breaks=n** 옵션을 이용해서 계급구간의 수를 설정하고,

**freq=F** 옵션을 이용하면 y축을 빈도수가 아닌 확률밀도로 표시 가능.

그래프의 색도 간단하게 설정할 수 있다.

# 1. 기본 제공 함수 시각화

Histogram - 연속형 데이터 시각화

실습!!

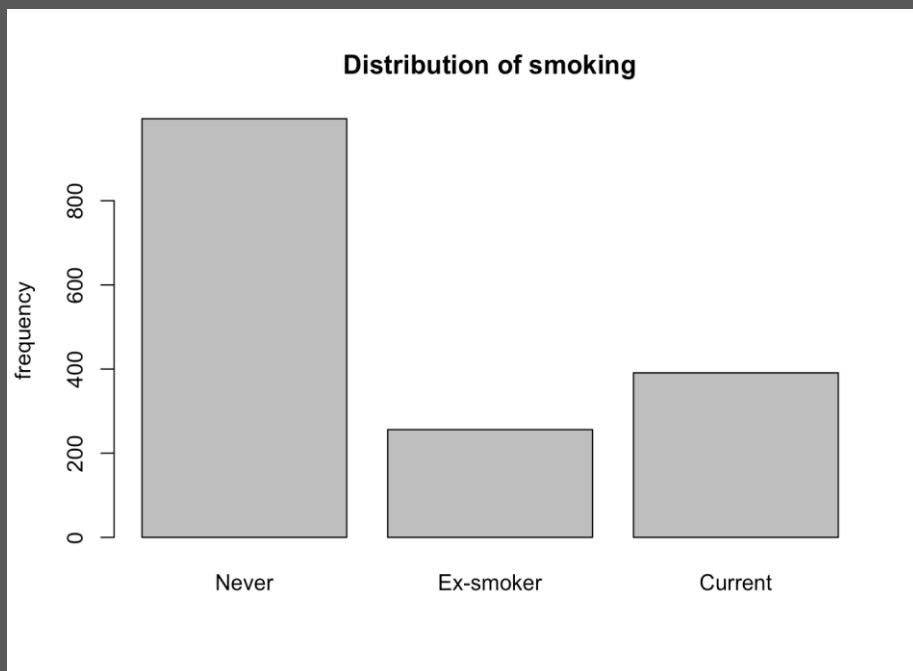
# 1. 기본 제공 함수 시각화

Barplot - X축에 표현하고자 하는 변수가 범주형 변수일 때는 빈도수를 바 그래프로 나타낼 수 있다. `table()` 함수를 이용해 빈도표를 만들고, 바 그래프로 나타낸다.

# 1. 기본 제공 함수 시각화

## Barplot

```
table <- table(data$Q_SMK_YN)  
barplot(table, main="Distribution of smoking",  
names.arg=c("Never", "Ex-smoker", "Current"), ylab="frequency")
```





# 1. 기본 제공 함수 시각화

## Barplot

```
table <- table(data$Q_SMK_YN, data$EXMD_BZ_YYYY)
barplot(table, main="Distribution of smoking by year", ylab="frequency",
        legend=c("Never", "Ex-smoker", "Current"), beside=T)
```

연도에 따른 흡연 여부의 분포를 하나의 그래프로 나타낼 수 있다. **beside=T** 옵션을 사용하면 적층형 그래프가 그룹형 그래프로 바뀐다.

# 1. 기본 제공 함수 시각화

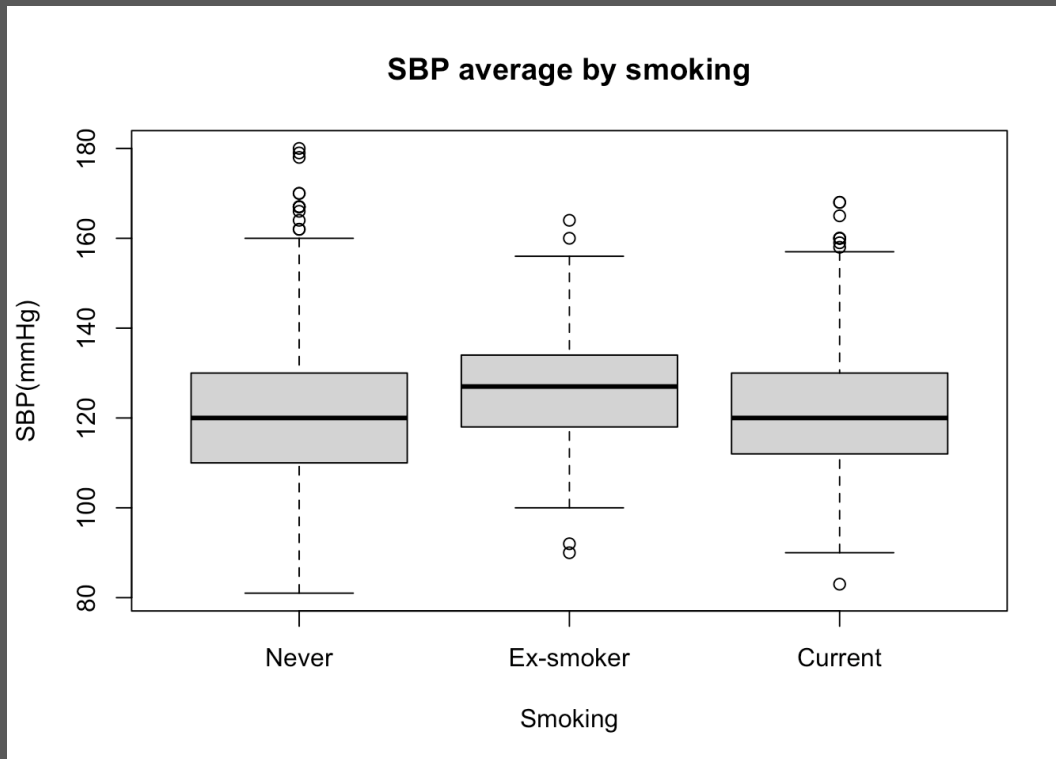
Barplot

실습!!

# 1. 기본 제공 함수 시각화

Boxplot - 범주형 변수에 따른 연속형 변수의 분포를 나타내는 데 이용

```
boxplot(BP_SYS ~ Q_SMK_YN, data = data, names=c("Never", "Ex-smoker", "Current"),  
        main="SBP average by smoking", ylab="SBP(mmHg)", xlab="Smoking")
```



# 1. 기본 제공 함수 시각화

Boxplot

실습!!

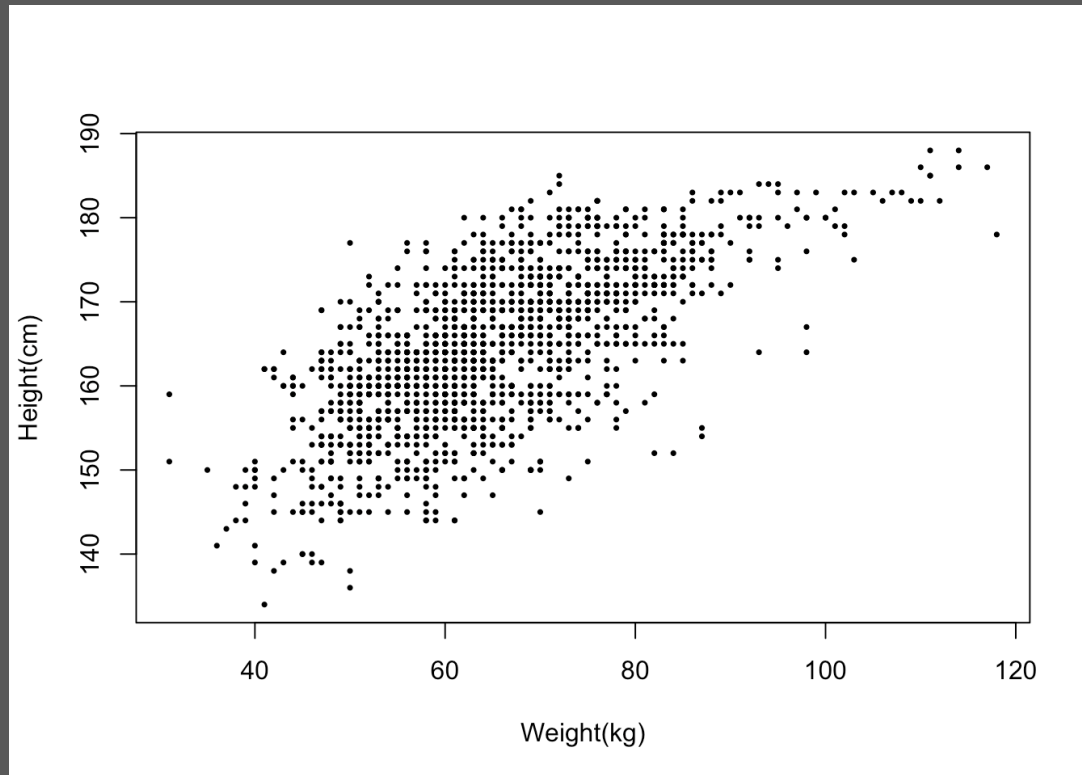
# 1. 기본 제공 함수 시각화

Scatter plot - 두 연속형 변수 간의 관계 시각화

```
plot(HGHT ~ WGHT, data=data, ylab="Height(cm)", xlab="Weight(kg)", pch=16, cex=0.5)
```

pch: 점의 모양

cex: 점의 크기



# 1. 기본 제공 함수 시각화

Scatter plot - 두 연속형 변수 간의 관계 시각화

범주형 변수에 따른 점의 분포를 표현하고자 할 때는 **점의 색깔(col= 옵션)**로 구분해서 표현할 수 있다.

또, **legend()** 함수를 이용하면 범례에 사용될 옵션을 따로 설정할 수 있다.

```
data2 <- data %>% filter(EXMD_BZ_YYYY %in% c(2009, 2015))  
plot(HGHT ~ WGHT, data=data2, col=factor(EXMD_BZ_YYYY), ylab="Height(cm)",  
      xlab="Weight(kg)", pch=16, cex=0.5)  
legend(x="bottomright", legend=c("2009", "2015"), col=1:2, pch = 19)
```

# 1. 기본 제공 함수 시각화

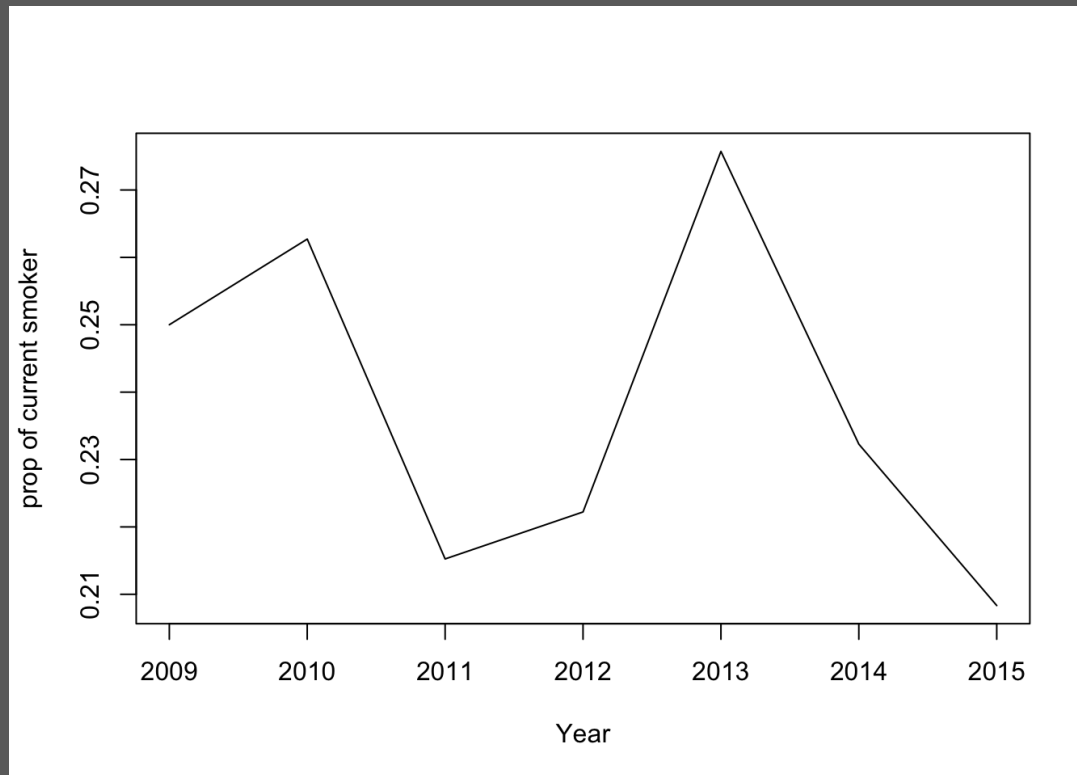
Scatter plot

실습!!

# 1. 기본 제공 함수 시각화

line plot - type="l" 옵션을 사용하면 선 그래프를 그릴 수 있다.

```
plot(table$EXMD_BZ_YYYY, table$smoker, type="l", xlab="Year", ylab="prop of current smoker")
```





# 1. 기본 제공 함수 시각화

line plot

실습!!

## 2. ggplot2 시각화

R을 이용한 데이터 시각화에  
가장 흔히 사용 되는 ggplot2 패키지이다.

## 2. ggplot2 시각화

### ggplot2 기본 문법

레이어를 하나씩 쌓는다고 생각하면 쉽다.

```
ggplot(data=data, aes(x=..., y=..., col=, ...))
```

시각화 할 데이터, x축과 y축 변수, col에 분류하고 싶은 변수 할당 등..

+ geom\_point() : 산점도

+ geom\_boxplot(): 박스플롯

+ geom\_bar(): 바 그래프

+ geom\_line(): 선 그래프

‘+’ 연산자를 이용하여 그리고 싶은 그래프 할당

# 2. ggplot2 시각화

## ggplot2 기본 문법

ggtitle(), xlab(), ylab() 함수를 이용해 그래프 제목, X축 라벨과 Y축 라벨을 추가할 수 있다.  
scale\_color\_manual() 함수를 이용하면 범례에 사용될 옵션들을 지정할 수 있다.

```
ggplot(data=data, aes(x=..., y=..., col=,...))
```

```
+ geom_point()
```

```
+ ggtitle("Title name") + xlab("x label name") + ylab("y label name")
```

```
+ scale_color_manual(
```

```
  values = c("원하는 색상 이름1", "원하는 색상 이름1", ...),
```

```
  labels = c("지정할 변수명1", "지정할 변수명2", ...),
```

```
  name = c("범례 제목")
```

## 2. ggplot2 시각화

ggplot2

히스토그램& 산점도 실습!!

## 2. ggplot2 시각화

### ggplot2 boxplot

```
ggplot(data=data2, aes(x=factor(x), y=y)) +  
  geom_boxplot() +  
  scale_x_discrete(labels=c("x1", "x2", ...))
```

범주형 변수 x에 따른 y값을 boxplot으로 보여주기 때문에 `scale_x_discrete`로 label 값을 매겨준다.

\* 연속형 변수일 경우에는 `scale_x_continuous`로 x축 범위 조절 가능

## 2. ggplot2 시각화

### ggplot2 facet

`facet_wrap()`, `facet_grid()`로 그래프 구분/분할 가능.

`facet_wrap`( ~ 나누길 원하는 변수, `labeller=...`)

`facet_grid`(나누길 원하는 변수 1 ~ 나누길 원하는 변수 2, `labeller = ...`)

## 2. ggplot2 시각화

ggplot2

boxplot , facet 실습!!



## 2. ggplot2 시각화

### ggplot2 barplot – 연도별 흡연유무 시각화

```
ggplot(data=data, aes(x=연도, fill=factor(흡연유무))) +
```

```
  geom_bar(position="fill", color="grey") +
```

```
  scale_x_continuous(breaks=2009:2015) *실습 데이터의 연도가 2009~2015년 까지 존재하기 때문에
```

table()을 사용하지 않고 바로 ggplot으로 그래프 만들기 가능

geom\_bar의 **position** 파라미터로 바 그래프 유형 변경 가능

position = “fill”: 100% 누적 비율 바 그래프

position = “stack”: 적층형 바 그래프

position = “dodge”: 그룹형 바 그래프

## 2. ggplot2 시각화

ggplot2 coord\_flip()

+coord\_flip() 함수로 x축과 y축의 위치를 서로 바꿔준다.

축의 scale과 label을 다시 설정하지 않아도 되기 때문에 편리하다.

## 2. ggplot2 시각화

ggplot2

barplot , coord\_flip 실습!!

### 3. ggpubr 시각화

ggpubr 패키지는 ggplot2의 문법을 활용해 논문 작성을 위한 데이터 시각화에 최적화된 기능을 제공하는 패키지이다.

ggpubr 패키지를 이용하면 그래프 위에 자동으로 p-value나 통계적 유의성을 표시할 수 있고, 여러 종류의 그래프를 한 페이지로 보여주도록 배열할 수 있다.

### 3. ggpubr 시각화

ggpubr – histogram (고혈압 병력에 따른 체중 차이)

```
gghistogram(data=data, x='체중', color='고혈압',  
             fill='고혈압', add='mean', xlab, ylab...)
```

add = “mean”을 사용하여 고혈압군에 따른 체중의 평균을  
histogram 그래프에 표시

### 3. ggpubr 시각화

ggpubr – boxplot (고혈압 병력에 따른 체중 차이)

```
ggboxplot(data=data, x='고혈압', y='체중',  
          color='고혈압', xlab, ylab,...) +
```

```
stat_compare_means(method="t.test", label.x.npc = "middle")
```

`stat_compare_means`를 사용하여 고혈압군에 따른 체중의 평균에 통계적으로 유의한 차이가 있는지 표시 가능

### 3. ggpubr 시각화

ggpubr – boxplot (세개 이상의 범주)

```
ggboxplot(data=data, x='범주', y='체중', color='범주')
```

```
+ stat_compare_means(comparisons = list(c("1", "2"), ...))
```

```
+ stat_compare_means(label.y=150) * label을 y축 150 값에 위치한다는 뜻
```

```
+ scale_x_discrete(labels=c("범주1", "범주2", "범주3", ...))
```

세개 이상의 범주도 통계적 유의성 검정이 가능하다.

`comparisons` 파라미터에 비교하고 싶은 범주를 list에 할당

### 3. ggpubr 시각화

ggpubr – scatter plot

```
ggscatter(data=data, x=x, y=y, color='범주',
```

```
add = "reg.line" + conf.int = T) * 추세선의 신뢰구간 표시
```

```
+ stat_cor(aes(color=범주))
```

add= "reg.line" 옵션을 이용하여 산점도 그래프에 추세선을 그릴 수 있다.

stat\_cor() 함수로 두 변수 간의 상관계수와 p-value를 구할 수 있다.



### 3. ggpubr 시각화

ggpubr – ggarrange

```
ggarrange(plot1, plot2
```

```
  labels = c("A", "B"),
```

```
  ncol = 2, nrow = 1)
```

ggarrange 함수로 여러 개의 그래프를 한 페이지에 배열하여 볼 수 있다.

ncol은 열의 개수, nrow는 행의 개수이다.

# 3. ggpubr 시각화

ggpubr

실습!!

## 4. 그래프 저장하기

svg, officer 패키지

svg 패키지는 그래프를 편집 가능한 그림 형태로 만들어주고,  
officer 패키지는 그림을 PPT 파일로 저장하는 기능을 제공한다.

## 4. 그래프 저장하기

### 실습!!

plot\_file이라는 이름의 pptx 파일을 생성하고,

순서대로 슬라이드를 추가하고,

제작한 그래프를 dml(ggobj =) 함수를 이용해 편집 가능한 그림  
형태로 슬라이드에 삽입하고,

plot\_file.pptx 파일로 로컬 드라이브에 저장하는 일련의 과정을  
실행해 보자.

수고하셨습니다.

