

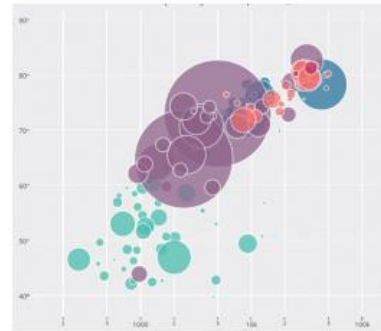
# 그래프 만들기



# 그래프 종류

# R로 만들 수 있는 그래프

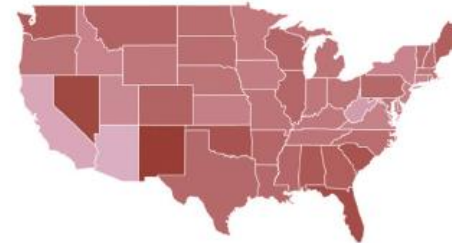
- ❖ 2차원 그래프, 3차원 그래프
- ❖ 지도 그래프
- ❖ 네트워크 그래프
- ❖ 모션 차트
- ❖ 인터랙티브 그래프



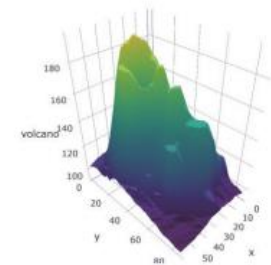
버블 차트(출처: [bit.ly/2rbbag](http://bit.ly/2rbbag))



네트워크 그래프(출처: [bit.ly/2rw3dh](http://bit.ly/2rw3dh))



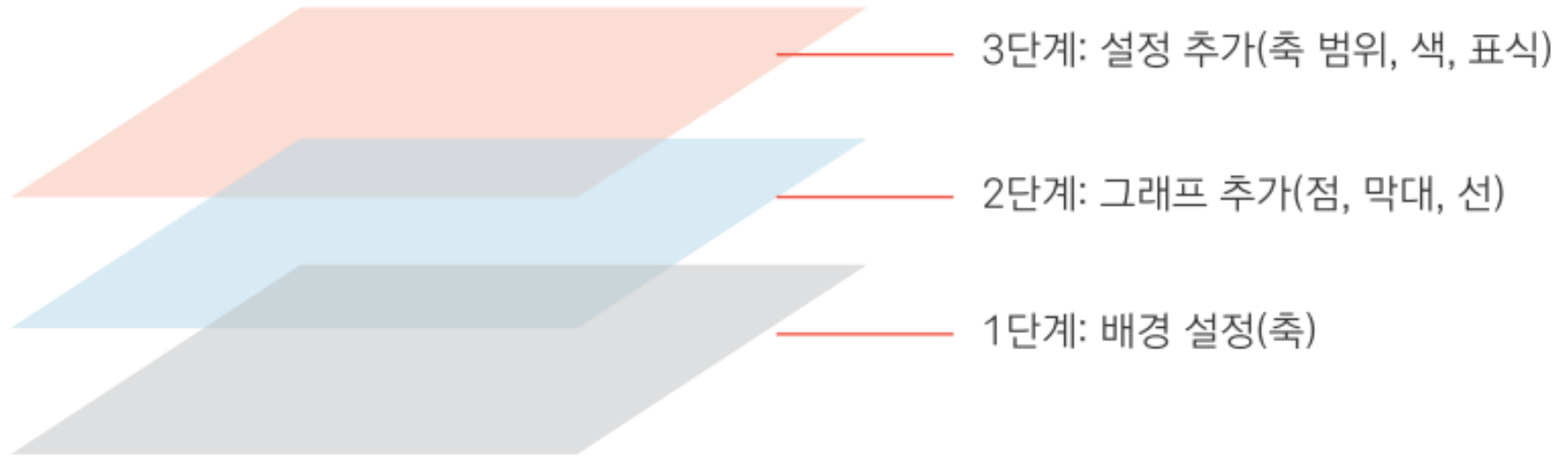
지도 그래프(출처: [bit.ly/2qqmkk9](http://bit.ly/2qqmkk9))



3D 그래프(출처: [bit.ly/2rmkrmn](http://bit.ly/2rmkrmn))

# 산점도

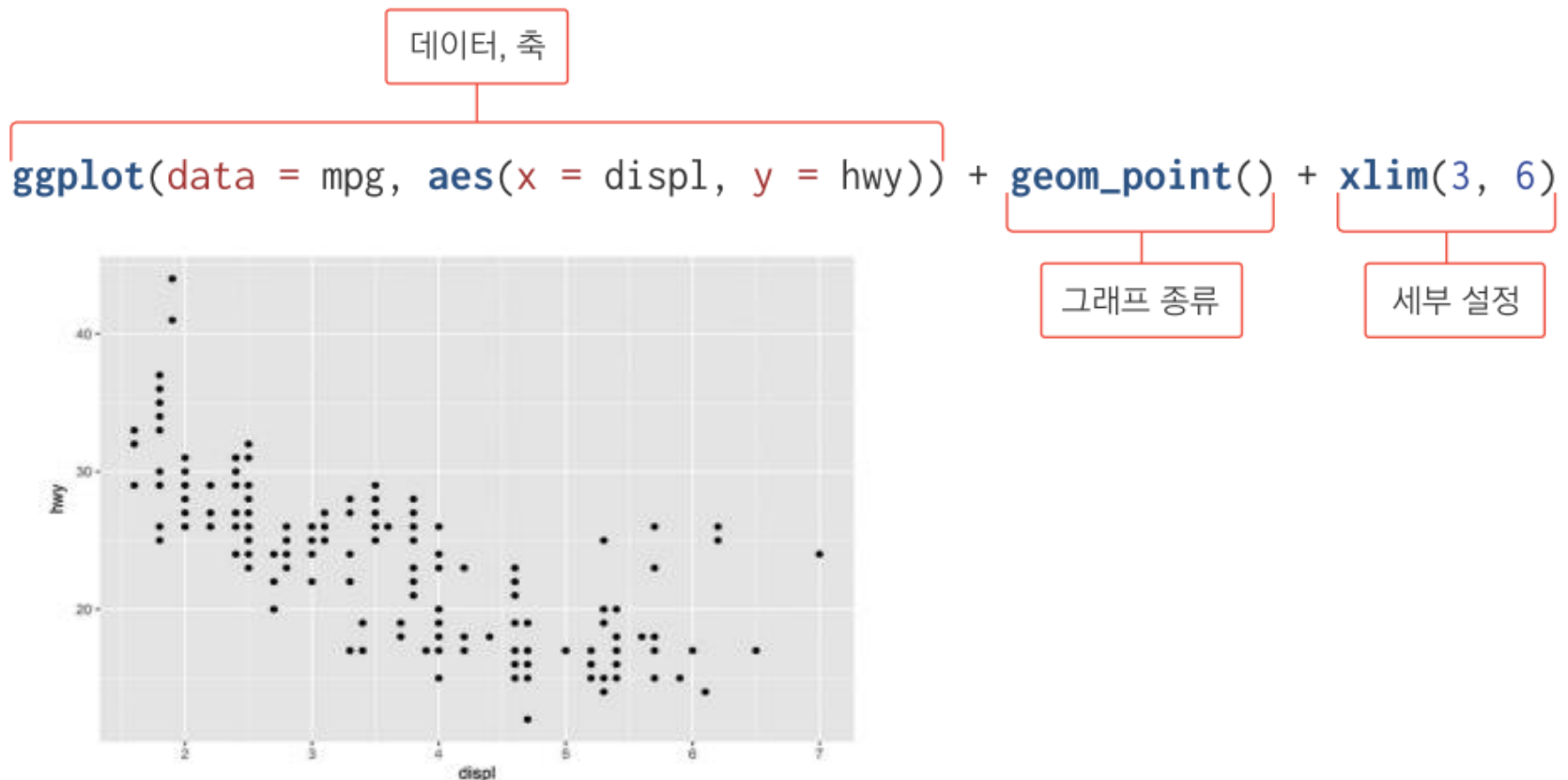
# Ggplot2 레이어 구조 이해



ggplot2 레이어 구조

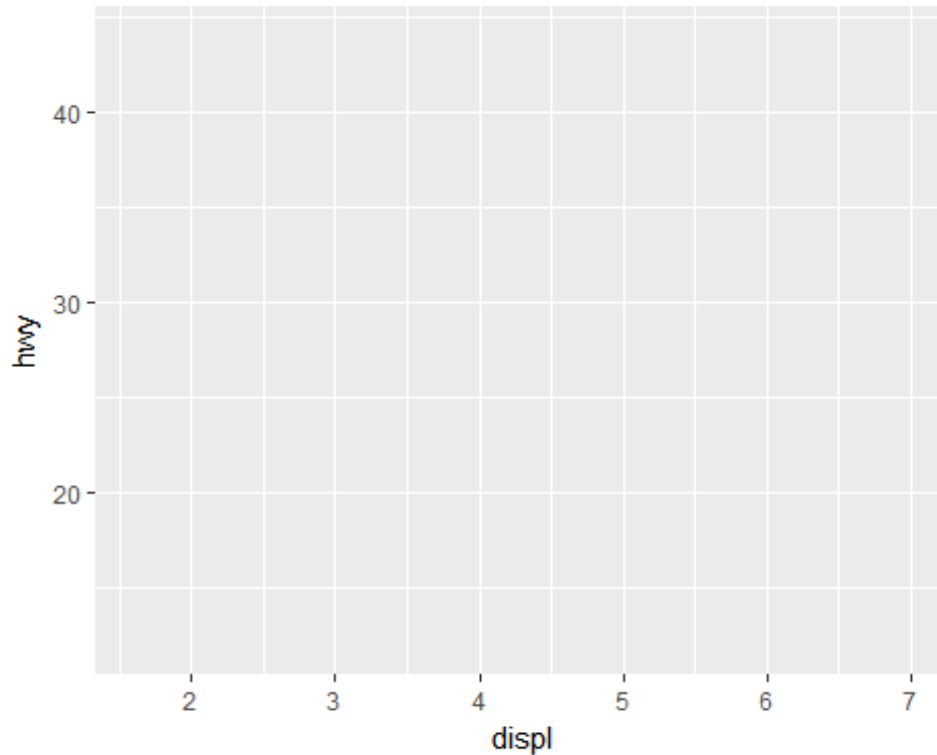
# 이해하기

- ❖ 산점도(Scater Plot) : 데이터를 x축과 y축에 점으로 표현한 그래프
- ❖ 나이와 소득처럼, 연속 값으로 된 두 변수의 관계를 표현할 때 사용



# 배경 설정하기

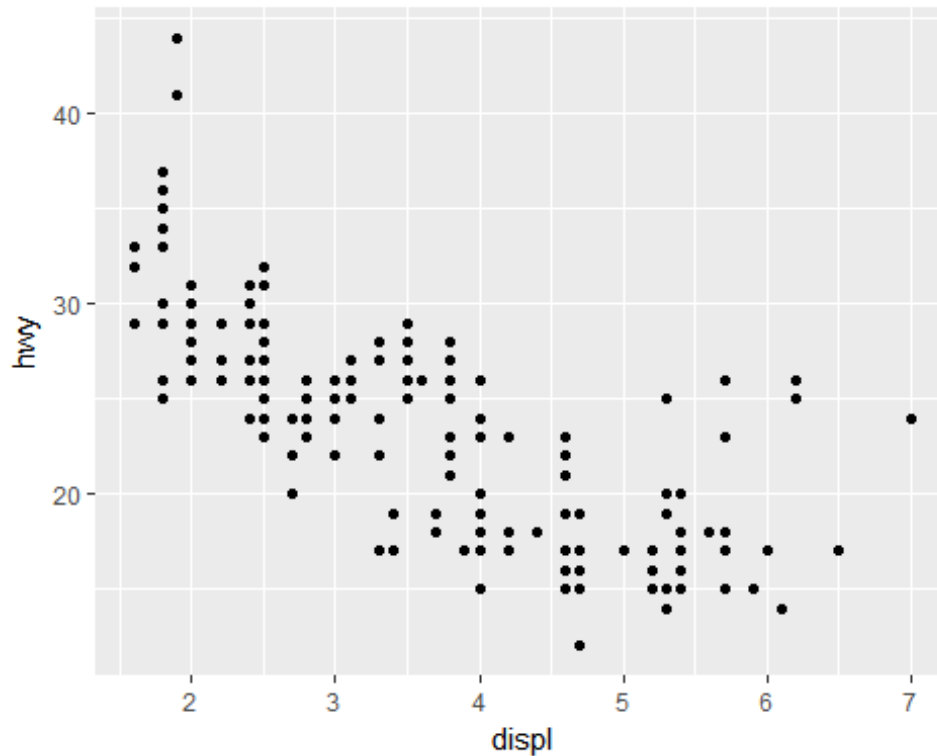
```
#라이브러리 등록  
library(ggplot2)  
# x축 displ, y축 hwy로 지정해 배경 생성  
ggplot(data = mpg, aes(x = displ, y = hwy))
```



# 그래프 추가하기

*# 배경에 산점도 추가*

```
ggplot(data = mpg, aes(x = displ, y = hwy)) + geom_point()
```

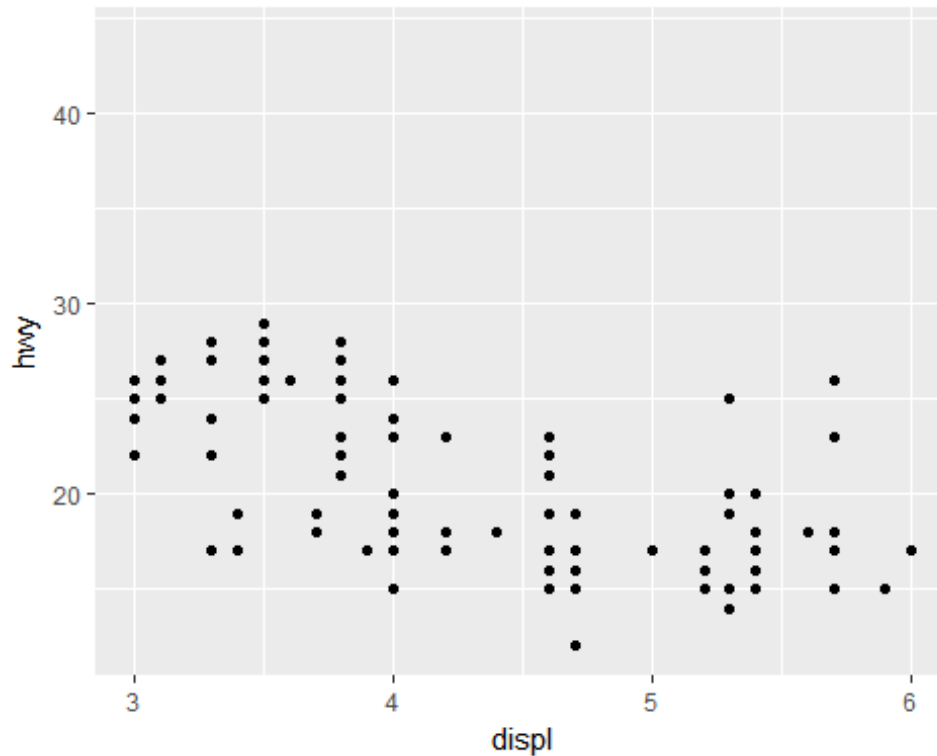




## 축 범위 조정

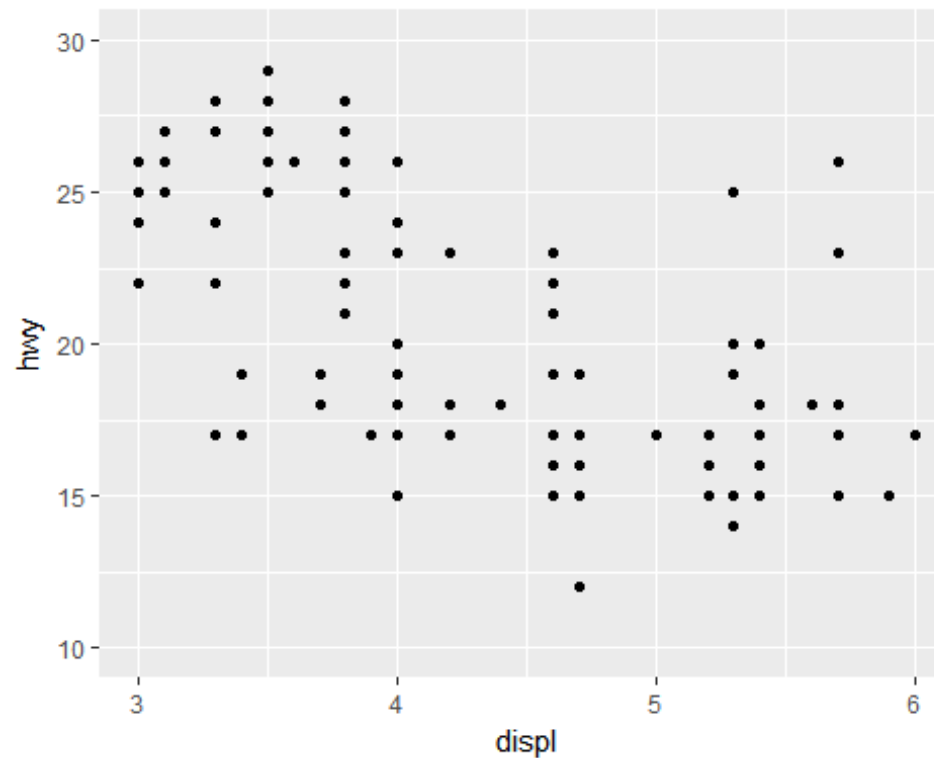
*# x축 범위 3~6으로 지정*

```
ggplot(data = mpg, aes(x = displ, y = hwy)) + geom_point() + xlim(3, 6)
```



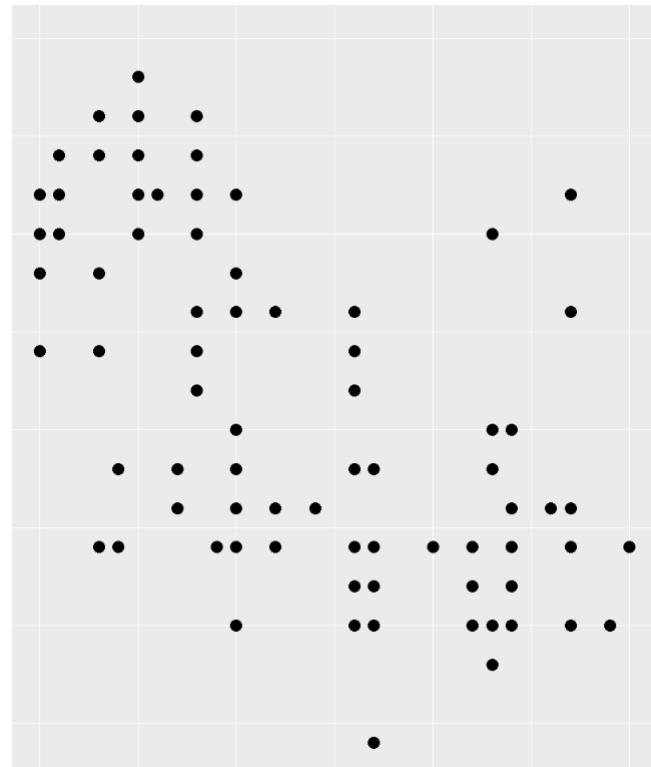
## 축 범위 조정

*# x축 범위 3~6, y축 범위 10~30으로 지정*  
`ggplot(data = mpg, aes(x = displ, y = hwy)) +  
 geom_point() +  
 xlim(3, 6) +  
 ylim(10, 30)`



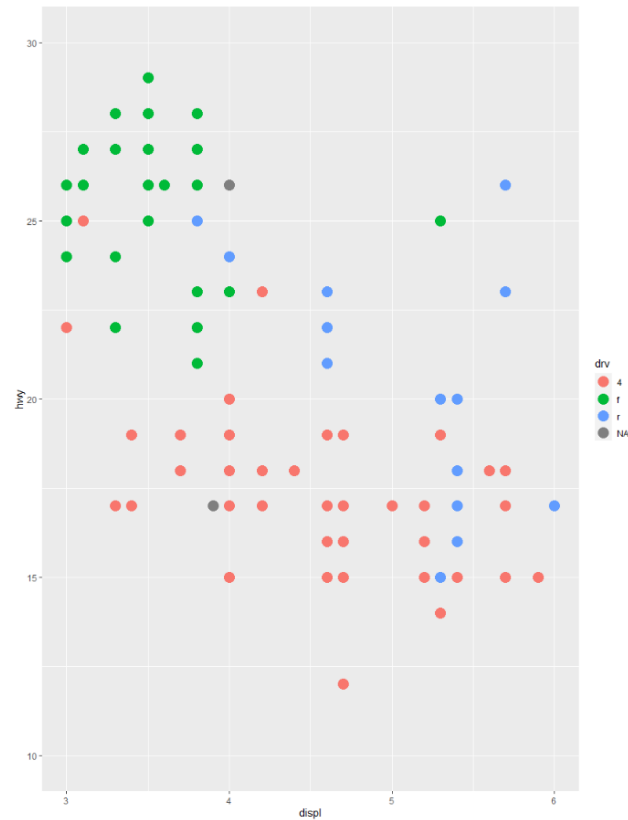
# Size 지정

```
library(ggplot2)
ggplot(data = mpg, aes(x = displ, y = hwy)) +
  #점의 크기 지정
  geom_point(size=5) +
  xlim(3, 6) +
  ylim(10, 30)
```



# Color=drv

```
library(ggplot2)
ggplot(data = mpg, aes(x = displ, y = hwy, color=drv)) +
  geom_point(size=5)
```



# Quiz

❖ mpg 데이터와 midwest 데이터를 이용해서 분석 문제를 해결해 보세요.

- Q1. mpg 데이터의 cty(도시 연비)와 hwy(고속도로 연비) 간에 어떤 관계가 있는지 알아보려고 합니다. x축은 cty, y축은 hwy로 된 산점도를 만들어 보세요.
- Q2. 미국 지역별 인구통계 정보를 담은 ggplot2 패키지의 midwest 데이터를 이용해서 전체 인구와 아시아인 인구 간에 어떤 관계가 있는지 알아보려고 합니다. x축은 poptotal(전체 인구), y축은 popasian(아시아인 인구)으로 된 산점도를 만들어 보세요. 전체 인구는 50만 명 이하, 아시아인 인구는 1만 명 이하인 지역만 산점도에 표시되게 설정하세요.

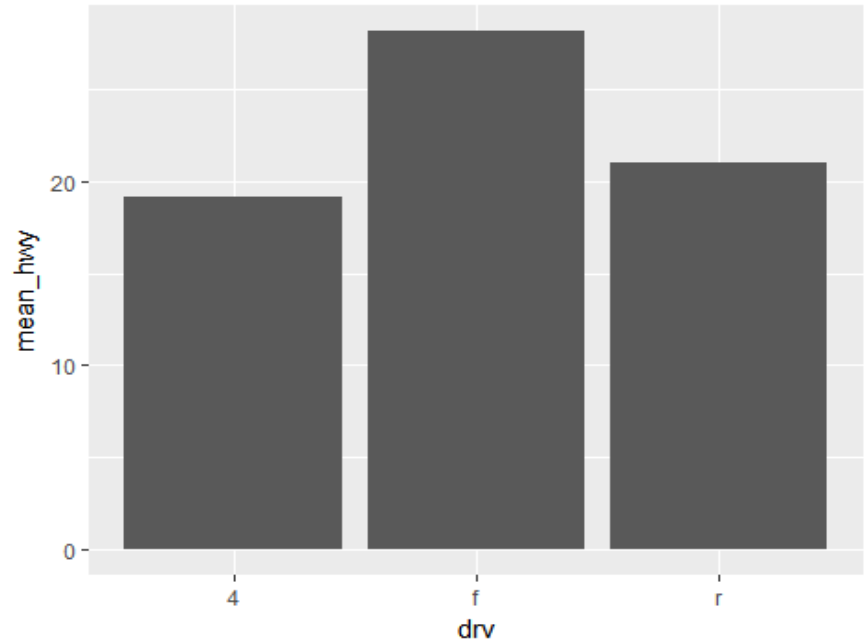
# 막대그래프

# 이해하기

- ❖ 막대 그래프(Bar Chart) : 데이터의 크기를 막대의 길이로 표현한 그래프
- ❖ 성별 소득 차이처럼 집단 간 차이를 표현할 때 주로 사용

# 평균 막대 그래프 만들기

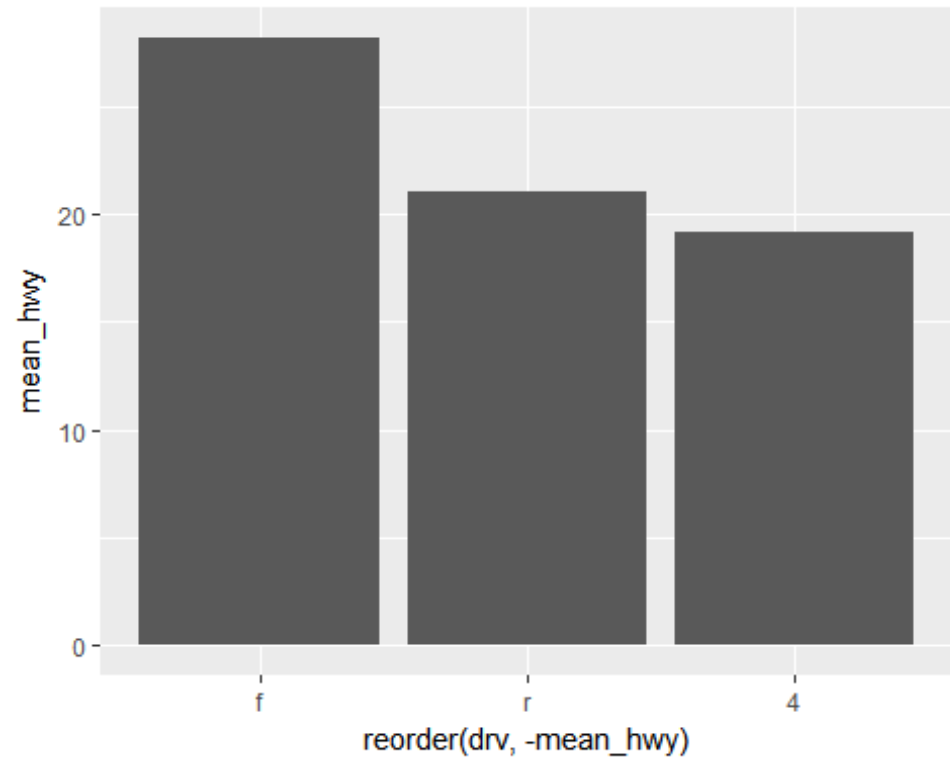
```
#라이브러리 등록
library(dplyr)
#평균 데이터 만들기
df_mpg <- mpg %>%
  group_by(drv) %>%
  summarise(mean_hwy = mean(hwy))
#확인
df_mpg
#그래프 생성
ggplot(
  data = df_mpg,
  aes(x = drv, y = mean_hwy)
) +
  geom_col()
```





# 크기순으로 정렬

```
ggplot(  
  data = df_mpg,  
  aes(  
    #reorder를 이용하여 정렬, 두 번째 데이터가 음수임으로 내림차순  
    x = reorder(drv, -mean_hwy),  
    y = mean_hwy)  
  ) +  
  geom_col()
```

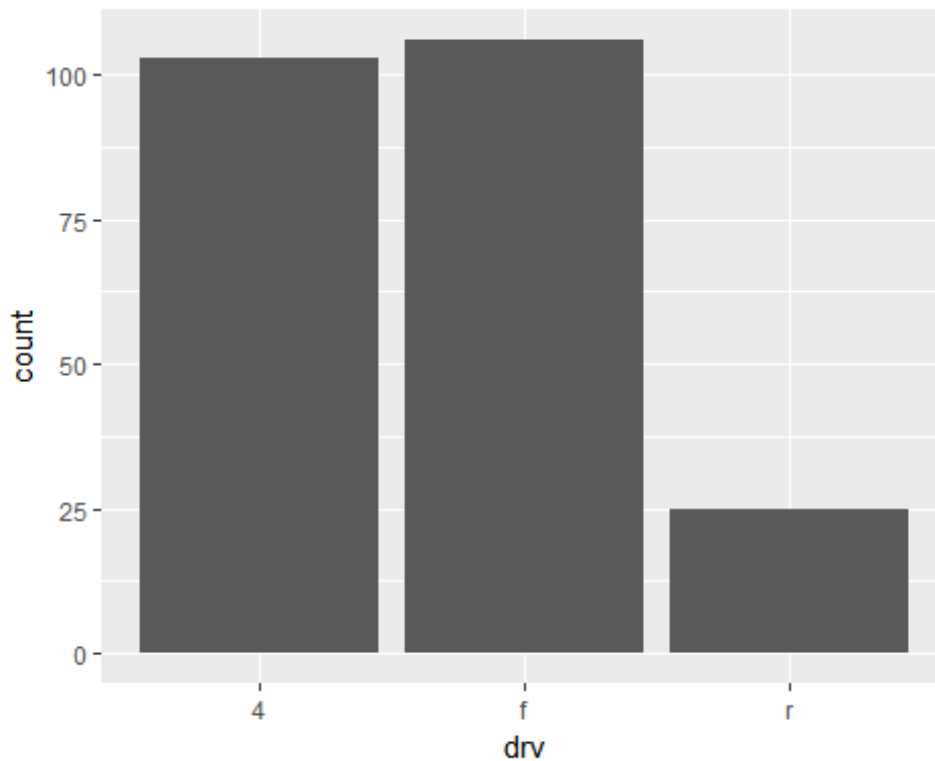


# 빈도 막대 그래프

❖ 값의 개수(빈도)로 막대의 길이를 표현한 그래프

# x축 범주 변수, y축 빈도

```
ggplot(data = mpg, aes(x = drv)) + geom_bar()
```

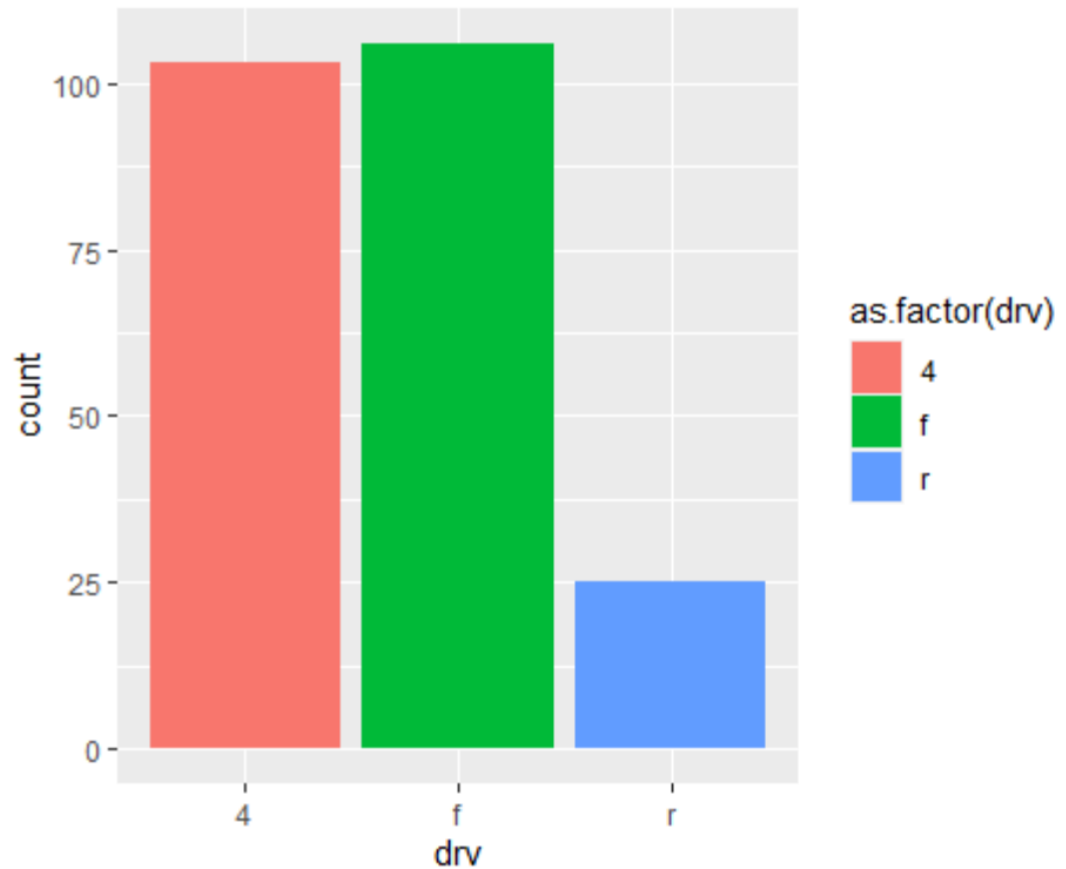


# 범주

```
ggplot(data = mpg, aes(x = drv)) +  
  geom_bar()
```

#표현할 범주 표시

```
mapping = aes(  
  x=drv,  
  fill=as.factor(drv)  
)
```

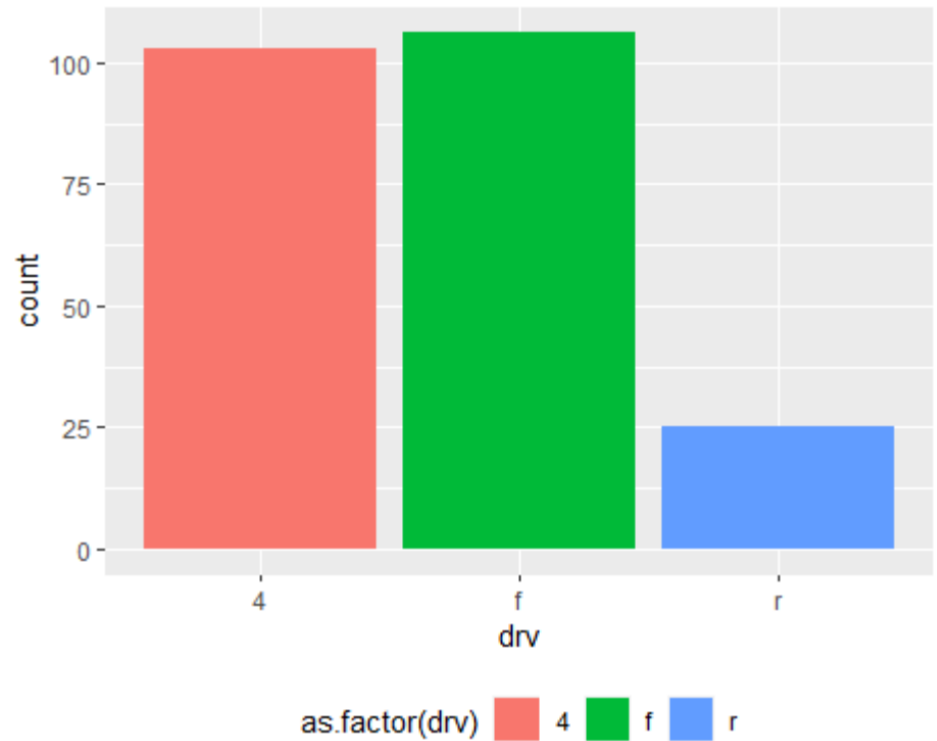


# Theme 활용

# theme를 이용한 범례 위치 변경

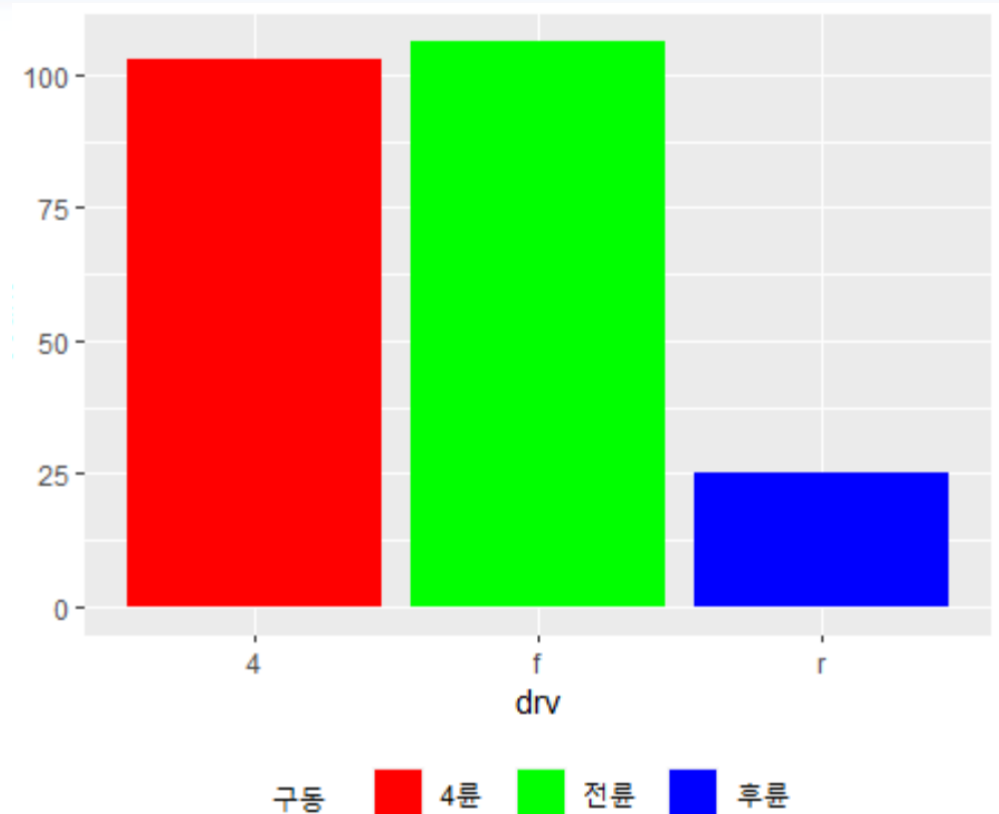
```
ggplot(data = mpg, aes(x = drv)) +  
  geom_bar(  
    mapping = aes(  
      x=drv,  
      fill=as.factor(drv)  
    )  
  )+  
  theme(legend.position = "bottom")
```

theme(legend.position = "bottom")



# scale\_fill\_manual

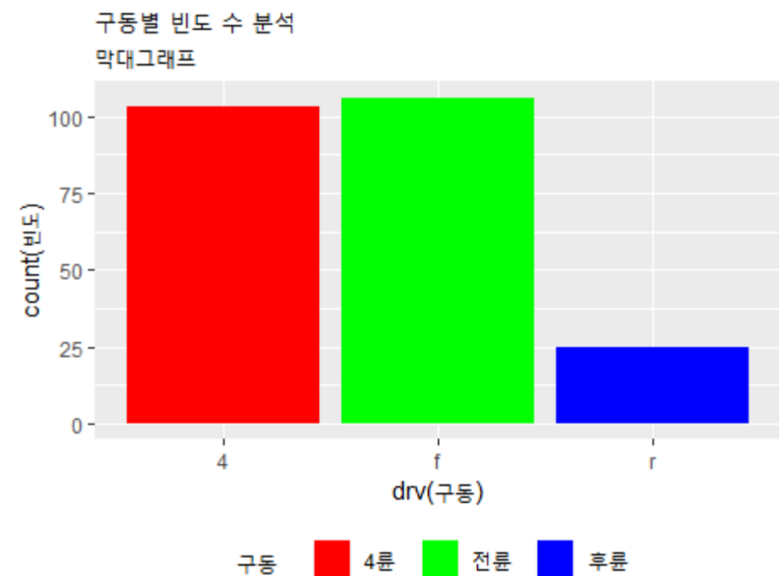
```
ggplot(data = mpg, aes(x = drv)) +  
  geom_bar(  
    mapping = aes(  
      x=drv,  
      fill=as.factor(drv)  
    )  
  )+  
  theme(legend.position = "bottom")+  
  scale_fill_manual(  
    values = c("red", "green", "blue"),  
    name = "구동",  
    labels=c("4륵", "전륵", "후륵")  
  )
```



# labs

```
ggplot(data = mpg, aes(x = drv)) +  
  geom_bar(  
    mapping = aes(  
      x=drv,  
      fill=as.factor(drv)  
    )  
  )+  
  theme(legend.position = "bottom")+  
  scale_fill_manual(  
    values = c("red", "green", "blue"),  
    name = "구동",  
    labels=c("4륜", "전륜", "후륜")  
  )+
```

```
labs(  
  x="drv(구동)",  
  y="count(빈도)",  
  title="구동별 빈도 수 분석",  
  subtitle="막대그래프",  
  caption = "출처 : 제작팀"  
)
```

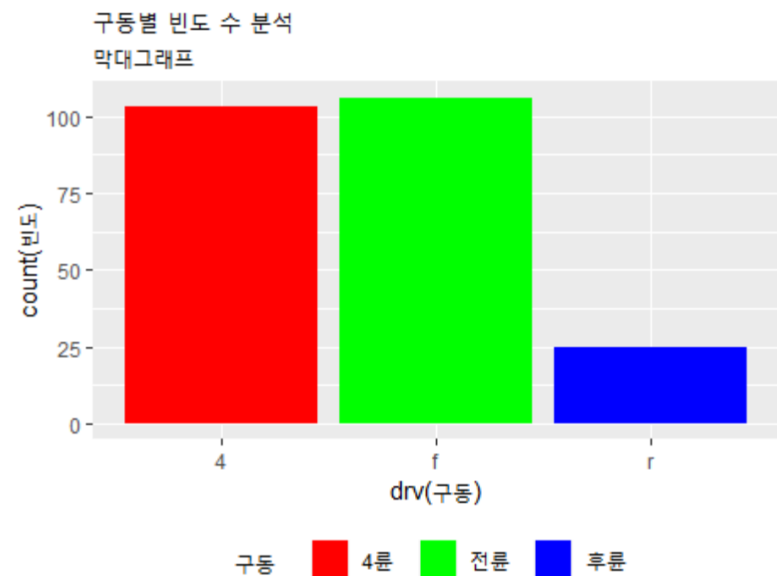


출처 : 제작팀

# labs

```
ggplot(data = mpg, aes(x = drv)) +  
  geom_bar(  
    mapping = aes(  
      x=drv,  
      fill=as.factor(drv)  
    )  
  )+  
  theme(legend.position = "bottom")+  
  scale_fill_manual(  
    values = c("red", "green", "blue"),  
    name = "구동",  
    labels=c("4륜", "전륜", "후륜")  
  )+
```

```
labs(  
  x="drv(구동)",  
  y="count(빈도)",  
  title="구동별 빈도 수 분석",  
  subtitle="막대그래프",  
  caption = "출처 : 제작팀"  
)
```

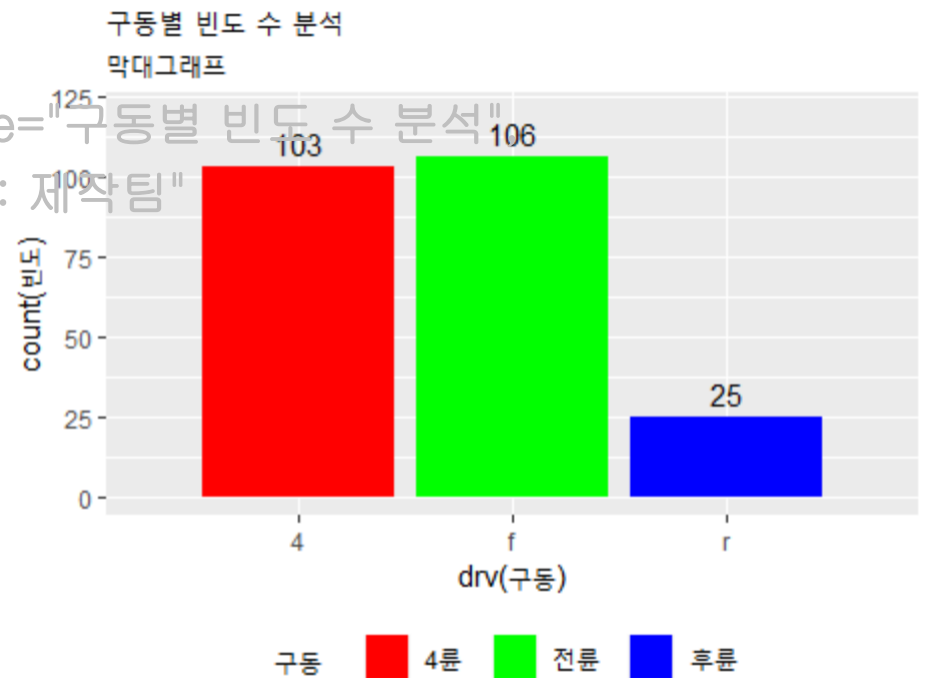


출처 : 제작팀

# geom\_text

```
ggplot(data = mpg, aes(x = drv)) +  
  geom_bar(mapping = aes(x=drv,fill=as.factor(drv))) +  
  theme(legend.position = "bottom")+  
  scale_fill_manual(values = c("red", "green", "blue"), name = "구동",  
    labels=c("4륜", "전륜", "후륜"))  
)+  
  labs( x="drv(구동)", y="count(빈도)", title="구동별 빈도 수 분석",  
    subtitle="막대그래프", caption = "출처 : 제작팀")  
)+  
  geom_text(  
    stat = "count",  
    aes(label=..count..),  
    position=position_dodge(width=1.8),  
    vjust=-0.5  
  )+  
  ylim(c(0, 120))
```

Label값이 계산된 y값을  
갖을 경우 stat를  
생략해도 된다.



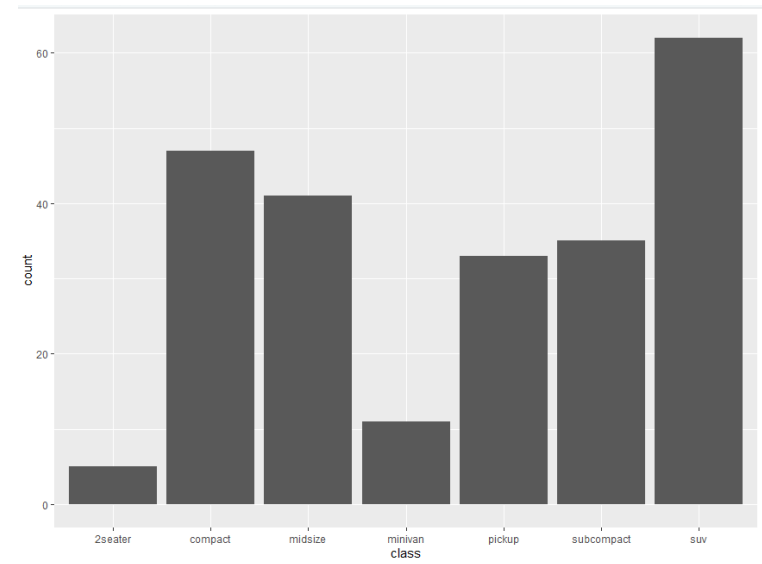
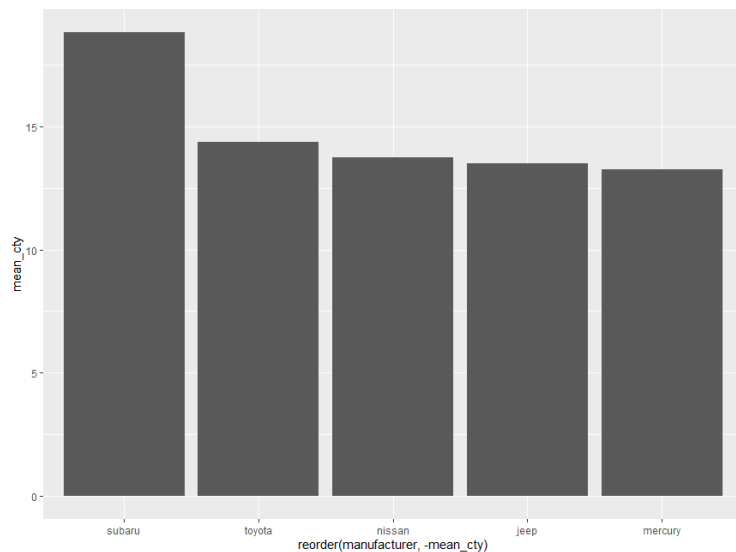
출처 : 제작팀



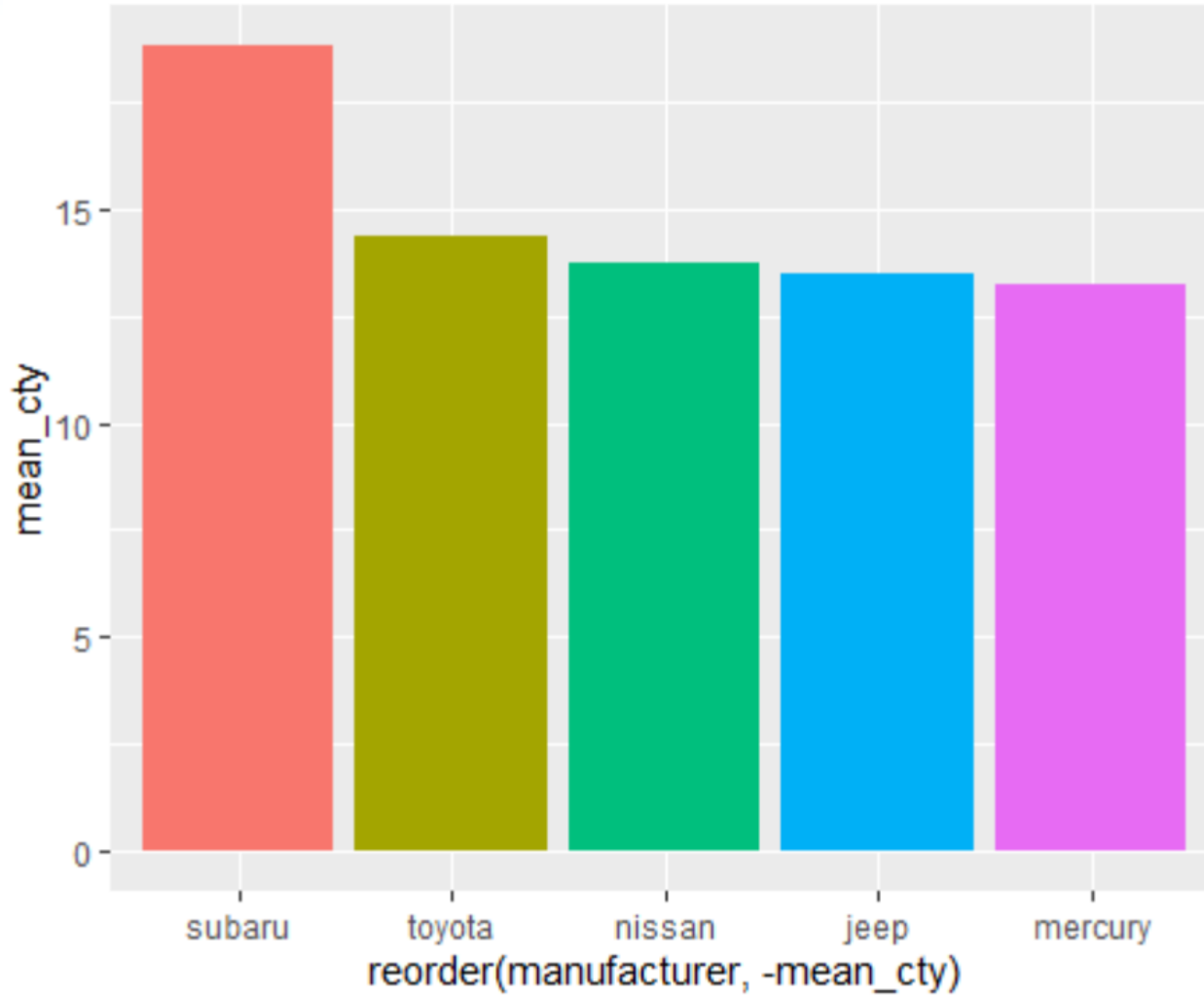
# Quiz

❖ mpg 데이터를 이용해서 분석 문제를 해결해 보세요.

- Q1. 어떤 회사에서 생산한 "suv" 차종의 도시 연비가 높은지 알아보려고 합니다. "suv" 차종을 대상으로 평균 cty(도시 연비)가 가장 높은 회사 다섯 곳을 막대 그래프로 표현해 보세요. 막대는 연비 가 높은 순으로 정렬하세요.
- Q2. 자동차 중에서 어떤 class(자동차 종류)가 가장 많은지 알아보려고 합니다. 자동차 종류별 빈도를 표현한 막대 그래프를 만들어 보세요.

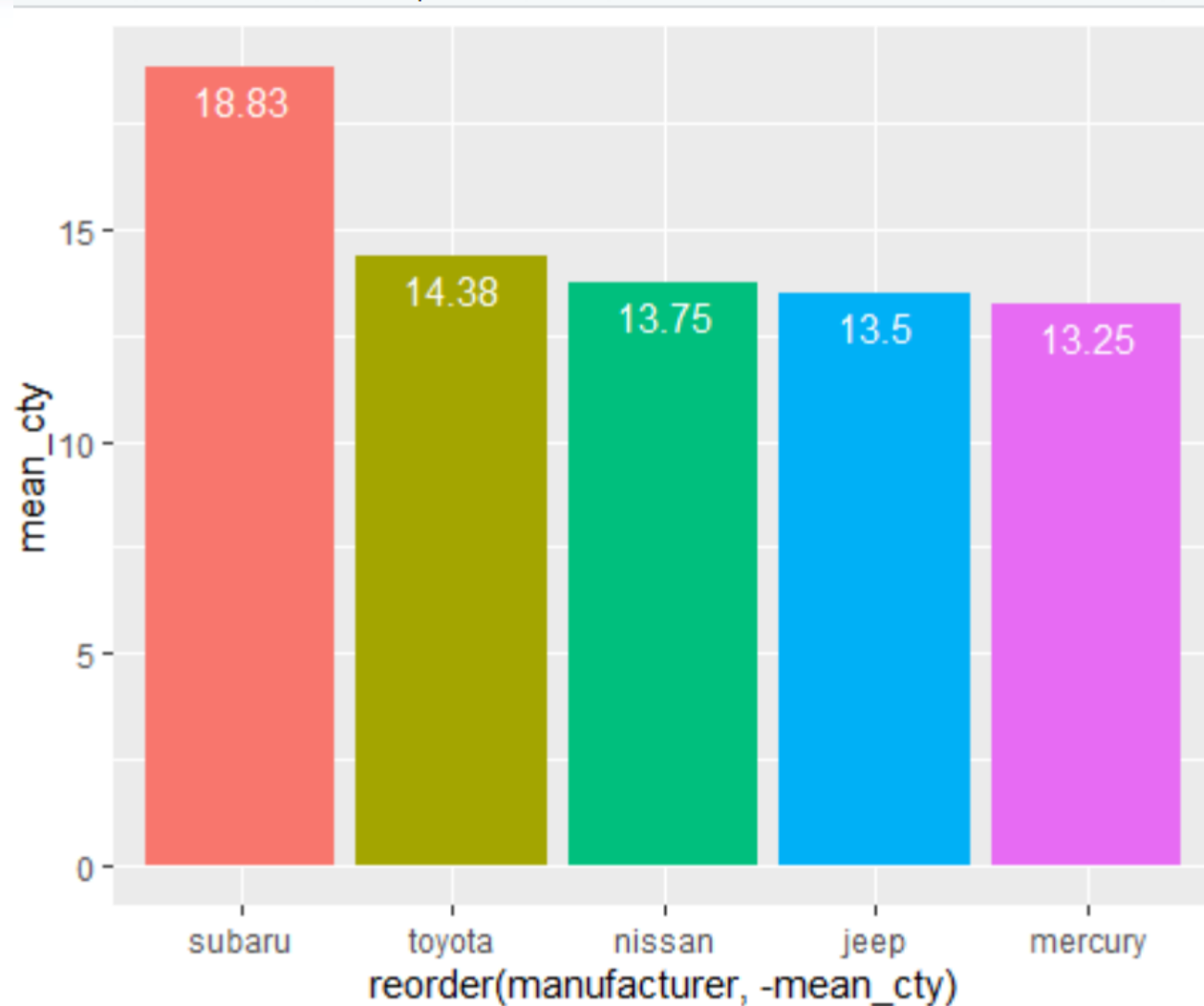


## Quiz1을 다음과 같이 변경하시오

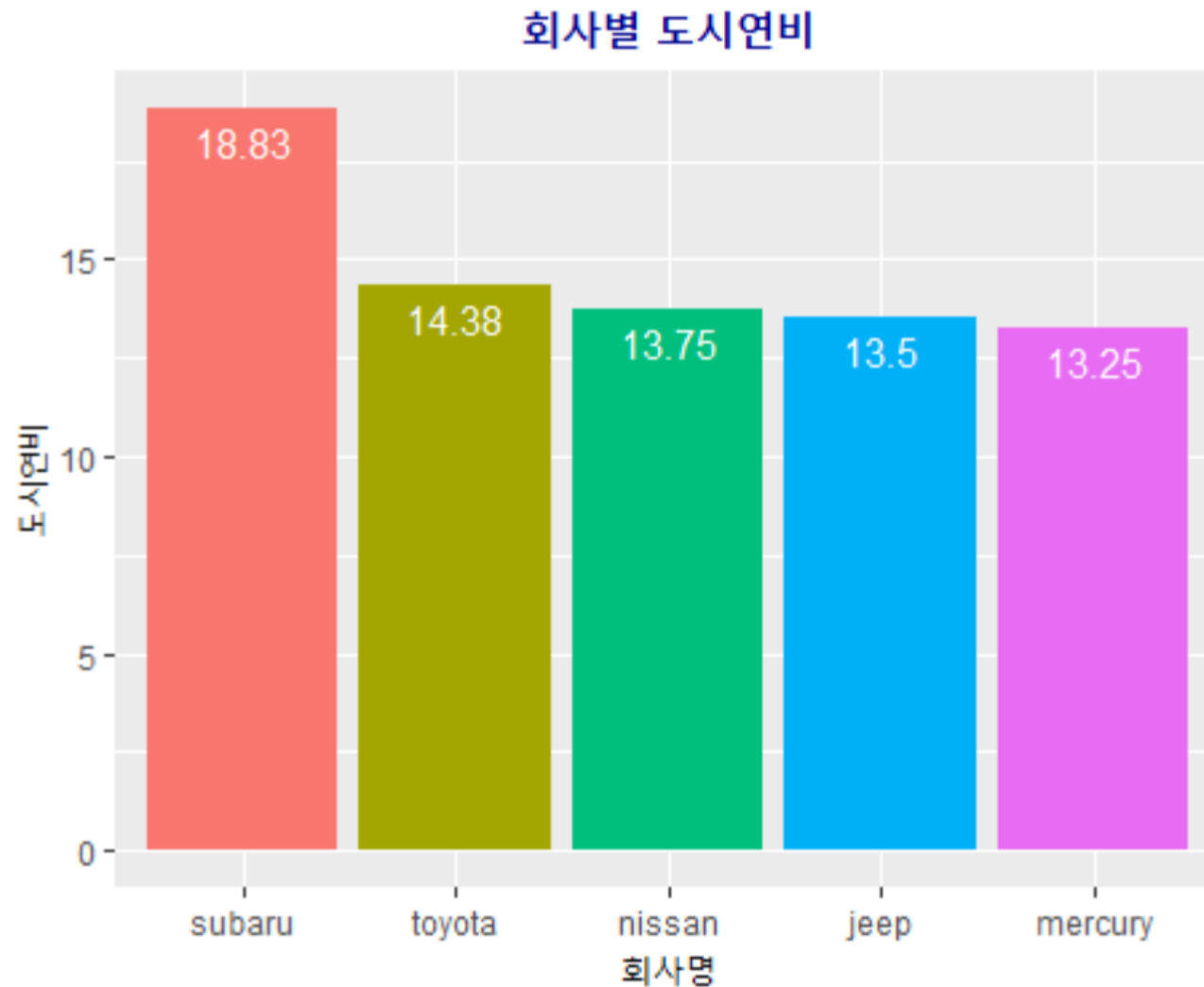


`guides(fill="none")`  
활용

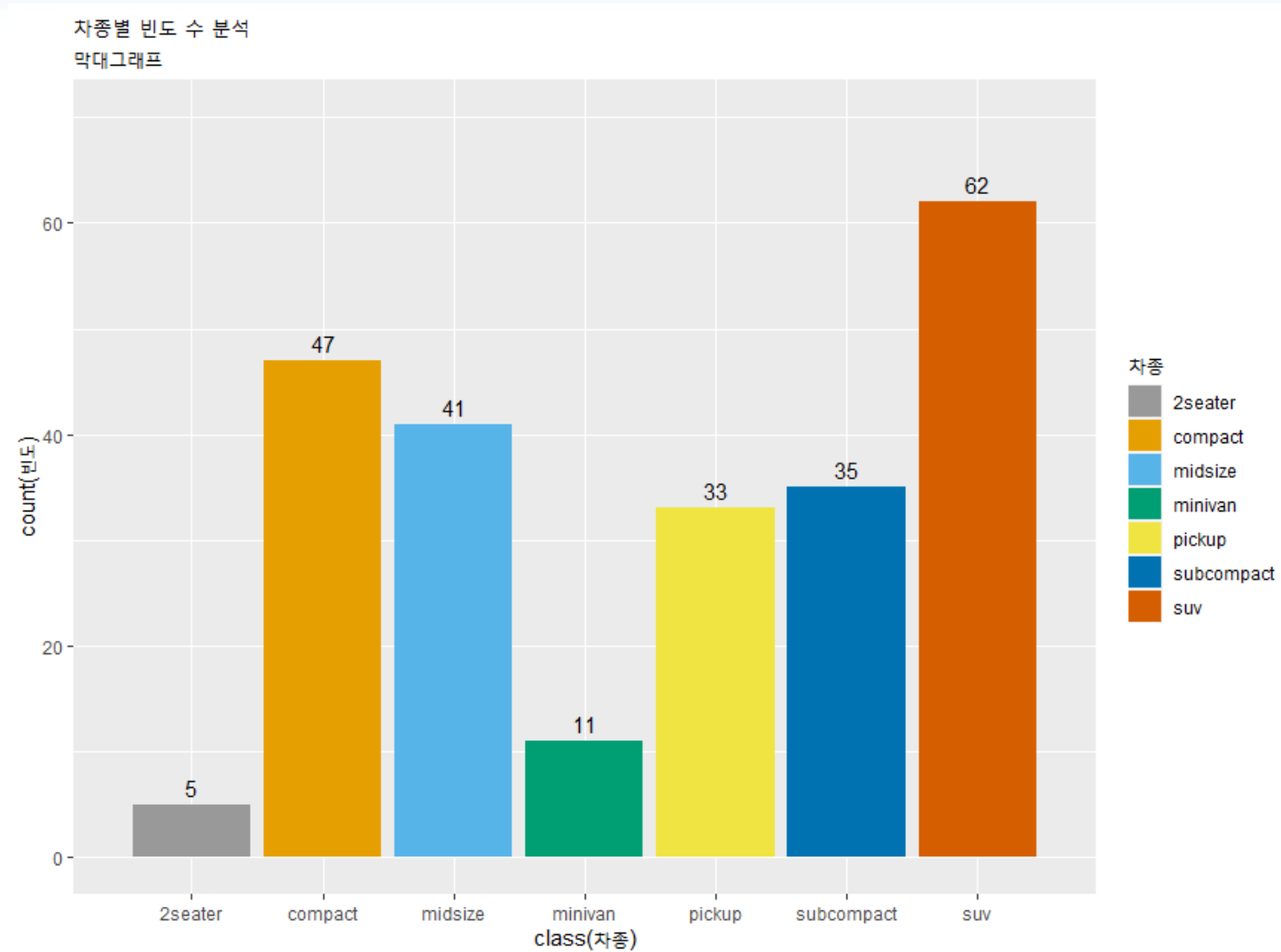
## 막대 그래프에 수치를 추가하시오



다음과 같이 제목 및 축제목을 출력하시오



# Quiz2를 다음과 같이 변경하시오



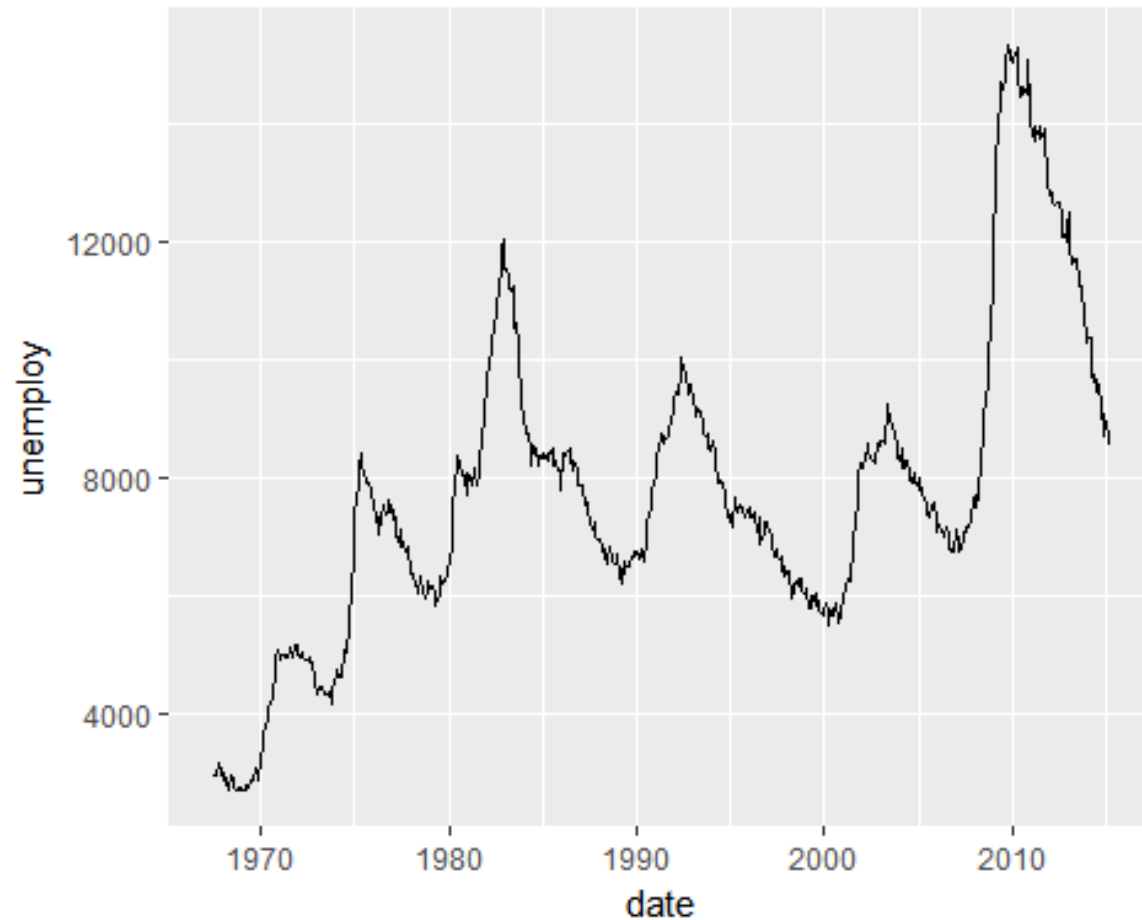
# 선그래프

# 이해하기

- ❖ 선 그래프(Line Chart) : 데이터를 선으로 표현한 그래프
- ❖ 시계열 그래프(Time Series Chart) : 일정 시간 간격을 두고 나열된 시계열 데이터(Time Series Data)를 선으로 표현한 그래프. 환율, 주가지수 등 경제 지표가 시간에 따라 어떻게 변하는지 표현할 때 활용

# 시계열 그래프 만들기

❖ `ggplot(data = economics, aes(x = date, y = unemploy)) + geom_line()`





# Quiz

❖ economics 데이터를 이용해서 분석 문제를 해결해 보세요.

- Q1. psavert(개인 저축률)가 시간에 따라서 어떻게 변해왔는지 알아보려고 합니다. 시간에 따른 개인 저축률의 변화를 나타낸 시계열 그래프를 만들어 보세요.

# 나무의 나이별 테두리

```
head(Orange)  
table(Orange$Tree)
```

```
library(dplyr)  
library(ggplot2)
```

```
Orange%>%  
  filter(Tree==1)%>%  
  ggplot(aes(age, circumference))+  
  geom_line()
```

## 색상 지정을 통한 분류

```
ggplot(Orange, aes(age, circumference))+  
  geom_line()
```

#색을 지정하면 여러 선을 표현할 수 있다.

```
ggplot(Orange, aes(age, circumference, color=Tree))+  
  geom_line()
```

#색 지정 다른 방법

```
ggplot(Orange, aes(age, circumference))+  
  geom_line(aes(color = Tree))
```

#선모양 변경

```
Ggplot(Orange, aes(age, circumference, color = Tree))+  
  geom_line(linetype = 6)
```

#배경 제거

```
ggplot(Orange, aes(age, circumference, color = Tree))+  
  geom_line(linetype = 6)+  
  theme(panel.background = element_blank())
```

#Tree가 숫자로 되어 있으므로 선종류 변경 가능

```
ggplot(Orange, aes(age, circumference, color = Tree))+  
  geom_line(aes(linetype = Tree))+  
  theme(panel.background = element_blank())
```