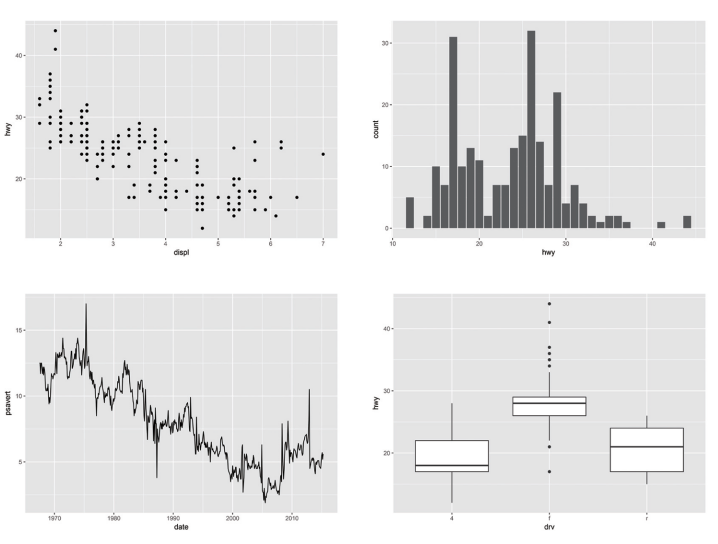
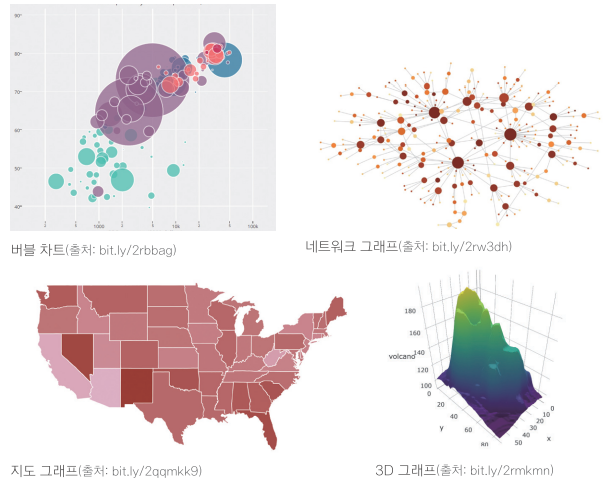
# 그래프 만들기



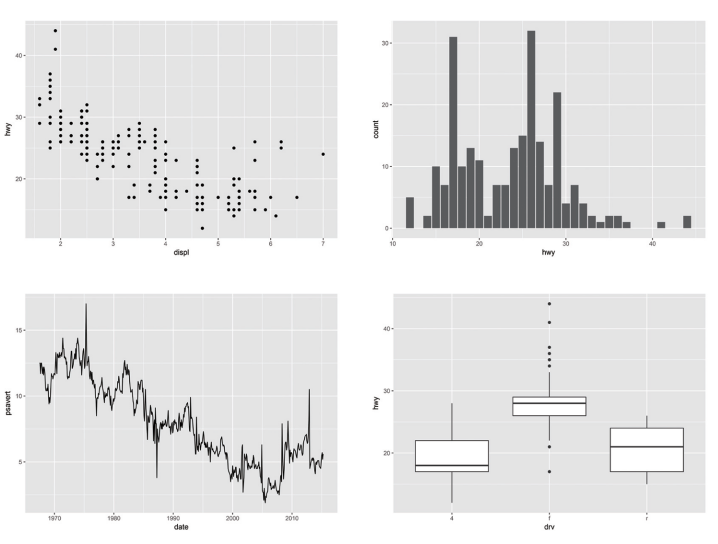
##### NP

## 08-1. R로 만들 수 있는 그래프 살펴보기

* 2차원 그래프, 3차원 그래프
* 지도 그래프
* 네트워크 그래프
* 모션 차트
* 인터랙티브 그래프

##### NP

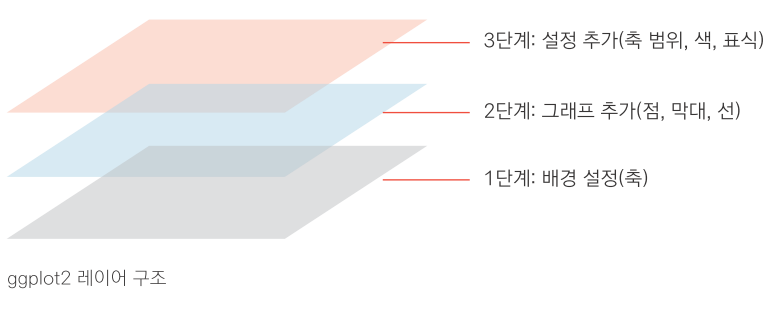
#### 쉽게 그래프를 만들 수 있는 ggplot2 패키지



##### NP

## 08-2. 산점도 - 변수 간 관계 표현하기

#### ggplot2 레이어 구조 이해하기



##### NP

### 산점도 만들기

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

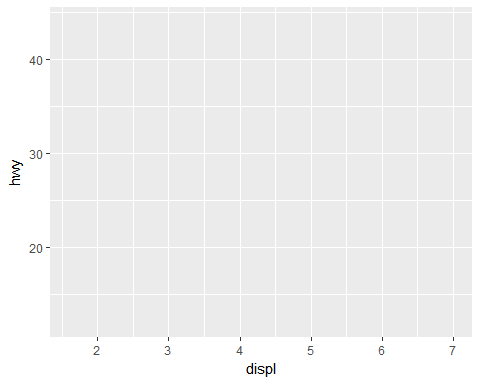
#### ggplot2 로드

library(ggplot2)

##### NP

#### 1. 배경 설정하기

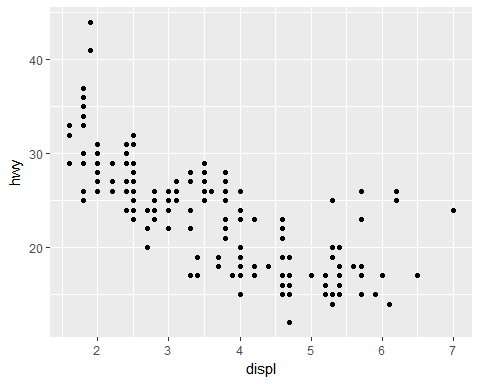
# x축 displ, y축 hwy로 지정해 배경 생성  
ggplot(data = mpg, aes(x = displ, y = hwy))



##### NP

#### 2. 그래프 추가하기

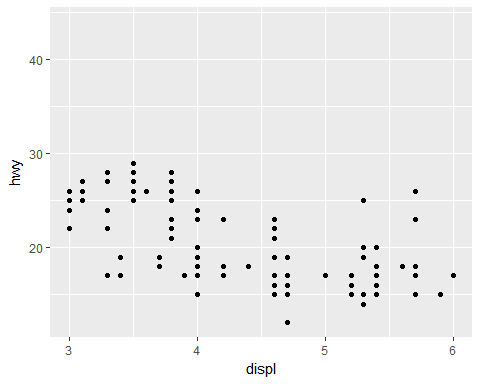
# 배경에 산점도 추가  
ggplot(data = mpg, aes(x = displ, y = hwy)) + geom\_point()



##### NP

#### 3. 축 범위를 조정하는 설정 추가하기

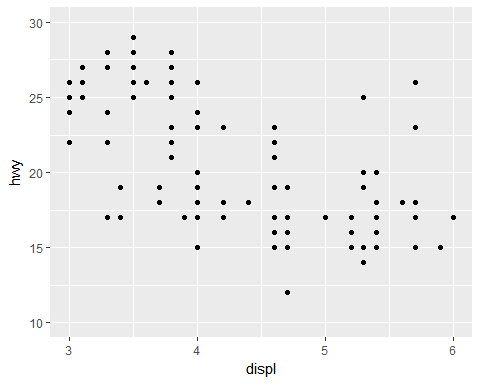
# x축 범위 3~6으로 지정  
ggplot(data = mpg, aes(x = displ, y = hwy)) + geom\_point() + xlim(3, 6)



##### NP

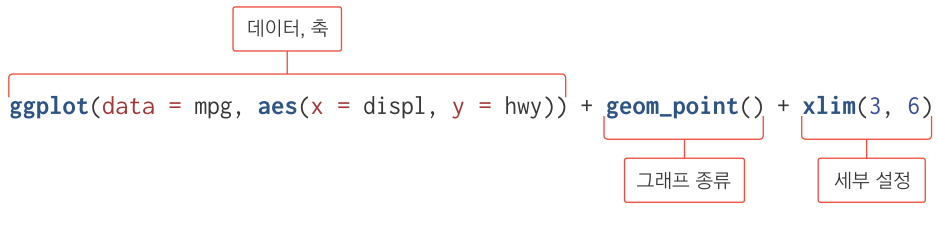
#### 3. 축 범위를 조정하는 설정 추가하기

# 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)



##### NP

#### ggplot 함수 구조



##### NP

#### ggplot2 코드 가독성 높이기

* 한 줄로 작성

ggplot(data = mpg, aes(x = displ, y = hwy)) + geom\_point() + xlim(3, 6) + ylim(10, 30)

* + 뒤에서 줄 바꾸기

ggplot(data = mpg, aes(x = displ, y = hwy)) +  
 geom\_point() +  
 xlim(3, 6) +  
 ylim(10, 30)

##### NP

#### ggplot() vs qplot()

* qplot() : 전처리 단계 데이터 확인용 문법 간단, 기능 단순
* ggplot() : 최종 보고용. 색, 크기, 폰트 등 세부 조작 가능

##### NP

##### NP

## 08-3. 막대 그래프 - 집단 간 차이 표현하기

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

##### NP

### 막대 그래프 1 - 평균 막대 그래프 만들기

* 각 집단의 평균값을 막대 길이로 표현한 그래프

#### 1. 집단별 평균표 만들기

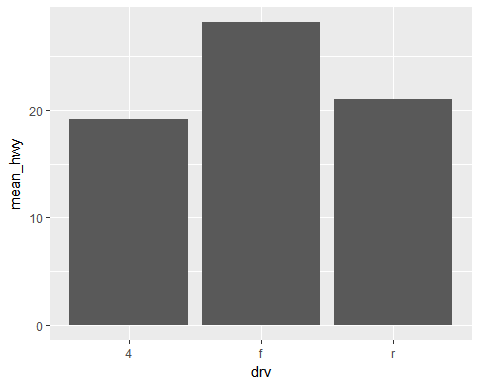
library(dplyr)  
  
df\_mpg <- mpg %>%  
 group\_by(drv) %>%  
 summarise(mean\_hwy = mean(hwy))  
  
df\_mpg

## # A tibble: 3 x 2  
## drv mean\_hwy  
## <chr> <dbl>  
## 1 4 19.17476  
## 2 f 28.16038  
## 3 r 21.00000

##### NP

#### 2. 그래프 생성하기

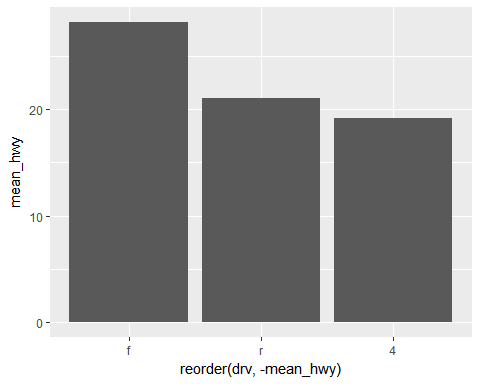
ggplot(data = df\_mpg, aes(x = drv, y = mean\_hwy)) + geom\_col()



##### NP

#### 3. 크기 순으로 정렬하기

ggplot(data = df\_mpg, aes(x = reorder(drv, -mean\_hwy), y = mean\_hwy)) + geom\_col()

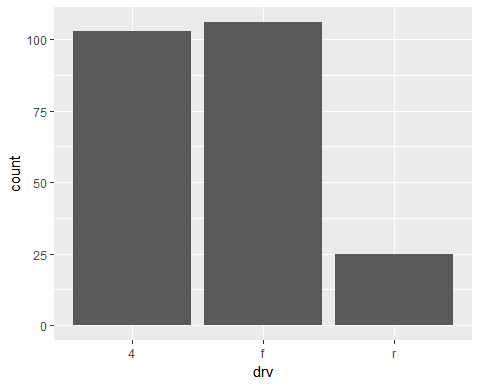


##### NP

### 막대 그래프 2 - 빈도 막대 그래프

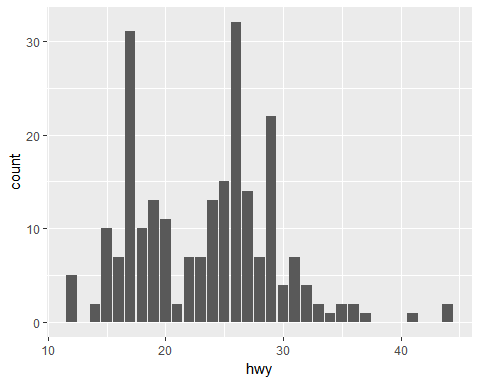
* 값의 개수(빈도)로 막대의 길이를 표현한 그래프

# x축 범주 변수, y축 빈도  
ggplot(data = mpg, aes(x = drv)) + geom\_bar()



##### NP

# x축 연속 변수, y축 빈도  
ggplot(data = mpg, aes(x = hwy)) + geom\_bar()



##### NPP

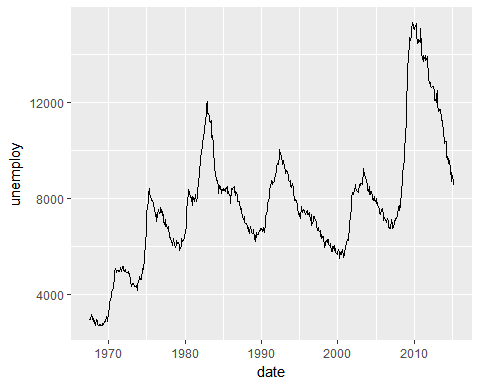
## 08-4. 선 그래프 - 시간에 따라 달라지는 데이터 표현하기

* 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

##### NP

### 시계열 그래프 만들기

ggplot(data = economics, aes(x = date, y = unemploy)) + geom\_line()



##### P

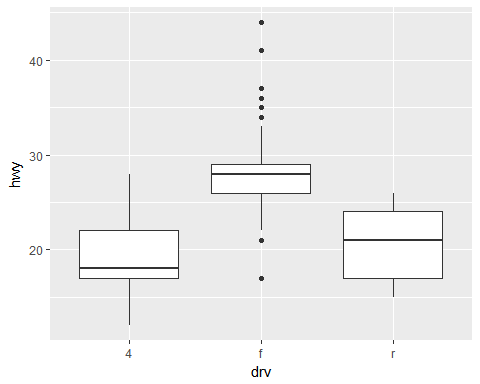
## 08-5. 상자 그림 - 집단 간 분포 차이 표현하기

* 상자 그림(Box Plot) : 데이터의 분포(퍼져 있는 형태)를 직사각형 상자 모양으로 표현한 그래프
* 분포를 알 수 있기 때문에 평균만 볼 때보다 데이터의 특성을 좀 더 자세히 이해할 수 있음

##### NP

### 상자 그림 만들기

ggplot(data = mpg, aes(x = drv, y = hwy)) + geom\_boxplot()



##### NP

|  |  |  |
| --- | --- | --- |
| 상자 그림 | 값 | 설명 |
| 상자 아래 세로선 | 아래 수염 | 하위 0~25% 내에 해당하는 값 |
| 상자 밑면 | 1사분위수(Q1) | 하위 25% 위치 값 |
| 상자 내 굵은 선 | 2사분위수(Q2) | 하위 50% 위치 값(중앙값) |
| 상자 윗면 | 3사분위수(Q3) | 하위 75% 위치 값 |
| 상자 위 세로선 | 윗수염 | 하위 75~100% 내에 해당하는 값 |
| 상자 밖 점 표식 | 극단치 | Q1, Q3 밖 1.5 IQR을 벗어난 값 |

참고 1.5 IQR: 사분위 범위(Q1~Q3간 거리)의 1.5배

##### NPNP

#### 앞에서 다룬 ggplot2 함수들

|  |  |
| --- | --- |
| 값 | 내용 |
| geom\_point() | 산점도 |
| geom\_col() | 막대 그래프 - 요약표 |
| geom\_bar() | 막대 그래프 - 원자료 |
| geom\_line() | 선 그래프 |
| geom\_boxplot() | 상자 그림 |

##### NP

### 정리하기

# 1.산점도  
ggplot(data = mpg, aes(x = displ, y = hwy)) + geom\_point()  
  
# 축 설정 추가  
ggplot(data = mpg, aes(x = displ, y = hwy)) +  
 geom\_point() +  
 xlim(3, 6) +  
 ylim(10, 30)  
  
# 2.평균 막대 그래프  
  
# 1단계.평균표 만들기  
df\_mpg <- mpg %>%  
 group\_by(drv) %>%  
 summarise(mean\_hwy = mean(hwy))  
  
# 2단계.그래프 생성하기, 크기순 정렬하기  
ggplot(data = df\_mpg, aes(x = reorder(drv, -mean\_hwy), y = mean\_hwy)) + geom\_col()  
  
# 3.빈도 막대 그래프  
ggplot(data = mpg, aes(x = drv)) + geom\_bar()

# 4.선 그래프  
ggplot(data = economics, aes(x = date, y = unemploy)) + geom\_line()  
  
# 5.상자 그림  
ggplot(data = mpg, aes(x = drv, y = hwy)) + geom\_boxplot()