03.Data Visualization with ggplot2

Youngjin\_Lee

2022 3 8

Table of Contents

## 3. Data Visualization with ggplot2

# <https://r4ds.had.co.nz/data-visualisation.html>

# tidyverse: ggplot2, purrr, tibble, dplyr,

# tidyr, stringr, readr, forcats

# 패키지설치 및 라이브러리 로드

# 01.데이터 불러오기

# p.100 설명

# 자동차 관련 데이터(내장 데이터)

# 데이터 확인하기

mpg

## # A tibble: 234 x 11  
## manufacturer model displ year cyl trans drv cty hwy fl class  
## <chr> <chr> <dbl> <int> <int> <chr> <chr> <int> <int> <chr> <chr>  
## 1 audi a4 1.8 1999 4 auto~ f 18 29 p comp~  
## 2 audi a4 1.8 1999 4 manu~ f 21 29 p comp~  
## 3 audi a4 2 2008 4 manu~ f 20 31 p comp~  
## 4 audi a4 2 2008 4 auto~ f 21 30 p comp~  
## 5 audi a4 2.8 1999 6 auto~ f 16 26 p comp~  
## 6 audi a4 2.8 1999 6 manu~ f 18 26 p comp~  
## 7 audi a4 3.1 2008 6 auto~ f 18 27 p comp~  
## 8 audi a4 quattro 1.8 1999 4 manu~ 4 18 26 p comp~  
## 9 audi a4 quattro 1.8 1999 4 auto~ 4 16 25 p comp~  
## 10 audi a4 quattro 2 2008 4 manu~ 4 20 28 p comp~  
## # ... with 224 more rows

# 데이터 확인하기

view(mpg)  
glimpse(mpg)

## Rows: 234  
## Columns: 11  
## $ manufacturer <chr> "audi", "audi", "audi", "audi", "audi", "audi", "audi", "~  
## $ model <chr> "a4", "a4", "a4", "a4", "a4", "a4", "a4", "a4 quattro", "~  
## $ displ <dbl> 1.8, 1.8, 2.0, 2.0, 2.8, 2.8, 3.1, 1.8, 1.8, 2.0, 2.0, 2.~  
## $ year <int> 1999, 1999, 2008, 2008, 1999, 1999, 2008, 1999, 1999, 200~  
## $ cyl <int> 4, 4, 4, 4, 6, 6, 6, 4, 4, 4, 4, 6, 6, 6, 6, 6, 6, 8, 8, ~  
## $ trans <chr> "auto(l5)", "manual(m5)", "manual(m6)", "auto(av)", "auto~  
## $ drv <chr> "f", "f", "f", "f", "f", "f", "f", "4", "4", "4", "4", "4~  
## $ cty <int> 18, 21, 20, 21, 16, 18, 18, 18, 16, 20, 19, 15, 17, 17, 1~  
## $ hwy <int> 29, 29, 31, 30, 26, 26, 27, 26, 25, 28, 27, 25, 25, 25, 2~  
## $ fl <chr> "p", "p", "p", "p", "p", "p", "p", "p", "p", "p", "p", "p~  
## $ class <chr> "compact", "compact", "compact", "compact", "compact", "c~

str(mpg)

## tibble [234 x 11] (S3: tbl\_df/tbl/data.frame)  
## $ manufacturer: chr [1:234] "audi" "audi" "audi" "audi" ...  
## $ model : chr [1:234] "a4" "a4" "a4" "a4" ...  
## $ displ : num [1:234] 1.8 1.8 2 2 2.8 2.8 3.1 1.8 1.8 2 ...  
## $ year : int [1:234] 1999 1999 2008 2008 1999 1999 2008 1999 1999 2008 ...  
## $ cyl : int [1:234] 4 4 4 4 6 6 6 4 4 4 ...  
## $ trans : chr [1:234] "auto(l5)" "manual(m5)" "manual(m6)" "auto(av)" ...  
## $ drv : chr [1:234] "f" "f" "f" "f" ...  
## $ cty : int [1:234] 18 21 20 21 16 18 18 18 16 20 ...  
## $ hwy : int [1:234] 29 29 31 30 26 26 27 26 25 28 ...  
## $ fl : chr [1:234] "p" "p" "p" "p" ...  
## $ class : chr [1:234] "compact" "compact" "compact" "compact" ...

# 02.그래프 그리기 (기본 함수 이용)

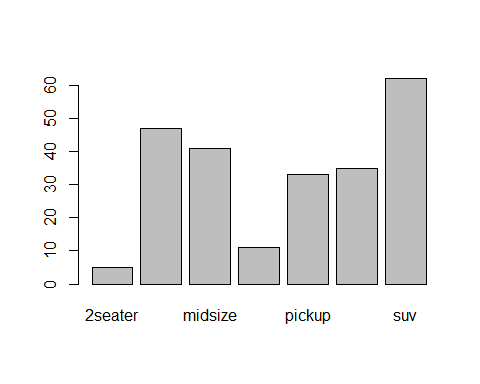
# Error in -0.01 \* height : non-numeric argument to binary operator

# 기본 함수로 그래프를 그리려면 table을 이용해 데이터를 정리한 후에 사용해야 함

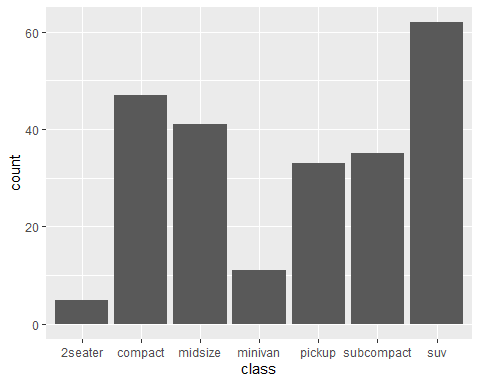
tb <- table(mpg$class)  
tb

##   
## 2seater compact midsize minivan pickup subcompact suv   
## 5 47 41 11 33 35 62

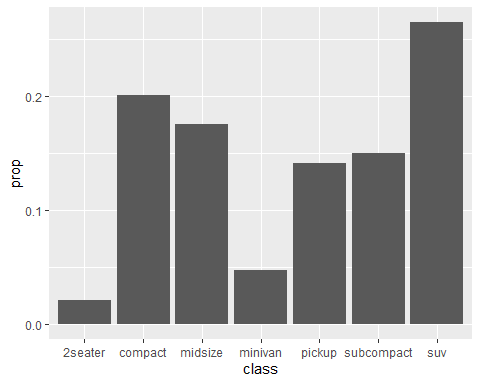
barplot(tb)

 # 03.ggplot 사용 # p.107 그림 설명 # ggplot을 이용하면 데이터를 자동으로 생성 # (Statistical transformations) # 빈도수 사용

ggplot(data = mpg,   
 mapping = aes(x = class)) +   
 geom\_bar()

 # 비율 사용

ggplot(data = mpg,   
 mapping = aes(x = class,  
 y = stat(prop),   
 group = 1)) +   
 geom\_bar()

 # 04.ggplot 기본문법 # p.111 설명 # 설명: Help -> Cheatsheets -> “Data Visualization with ggplot2.” # 1번과 2번 모두 사용가능하지만 1번을 이용 # geom\_function()의 기능을 사용해야 됨

ggplot(data = mpg,   
 mapping = aes(x = displ,   
 y = hwy)) +   
 geom\_point()

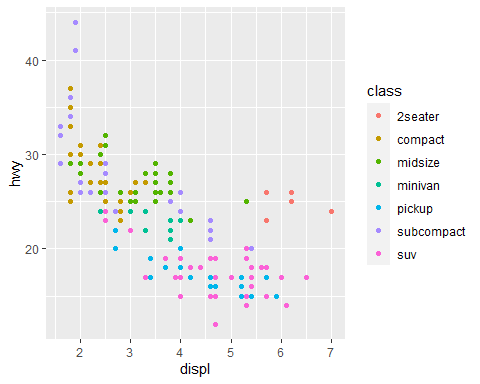


ggplot(data = mpg) +   
 geom\_point(mapping = aes(x = displ,   
 y = hwy))

 # 05.Aesthetic과 geom mappings 차이 # p.119 그림 설명 # aes는 변수에 매핑함 # geom은 플롯 전체에 매핑함

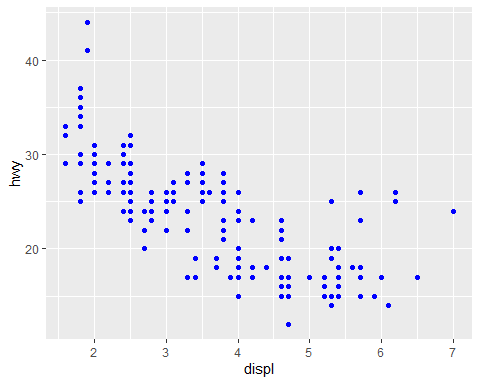
# Aesthetic mappings

ggplot(data = mpg,   
 mapping = aes(x = displ,   
 y = hwy,   
 color=class)) + # color = 변수명  
 geom\_point()



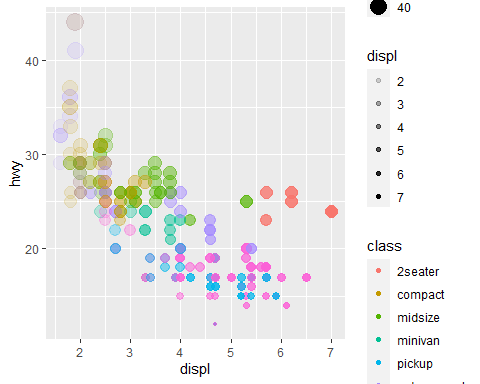
# geom mappings

ggplot(data = mpg,   
 mapping = aes(x = displ,   
 y = hwy)) +   
 geom\_point(color="blue") # color = 속성

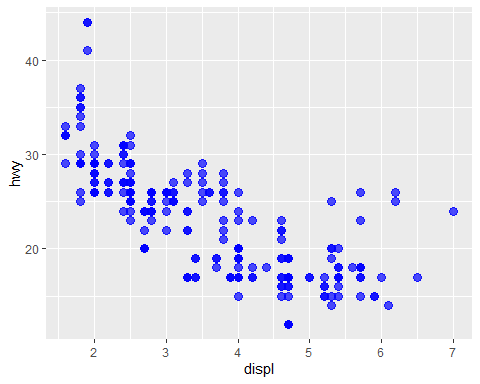


# Aesthetic mappings 옵션들

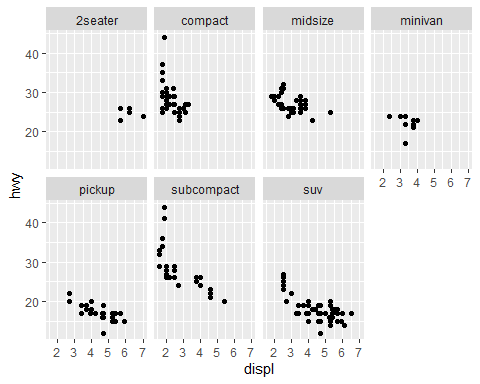
ggplot(data=mpg,   
 mapping=aes(x=displ, # 변수명 매칭  
 y=hwy, # 변수명 매칭  
 color=class, # 변수명 매칭  
 size=hwy, # 변수명 매칭  
 alpha=displ)) + #alpha = 투명도  
 geom\_point()

 # geom mappings 옵션들

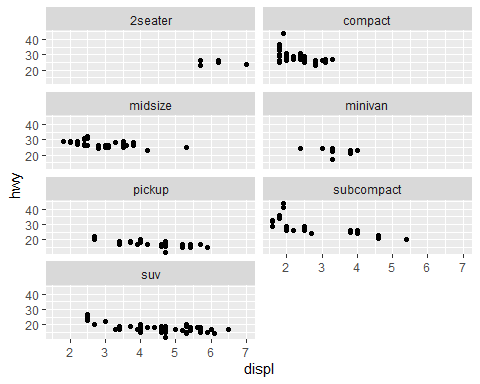
ggplot(data=mpg,   
 mapping=aes(x=displ,   
 y=hwy)) +  
 geom\_point(color="blue", # 속성 매칭  
 size=3,  
 alpha=0.7)

 # 06.Facet: 화면을 분리 # 구분변수가 1개일 때 # facet\_wrap ( ~ class)

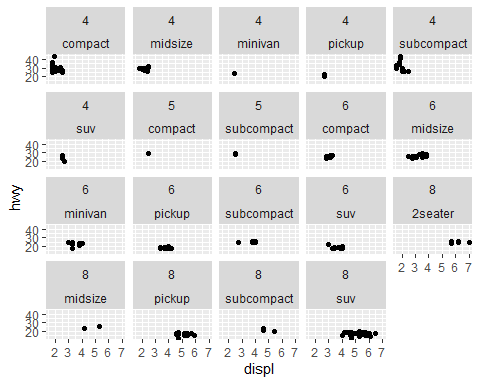
ggplot(data=mpg,   
 mapping=aes(x=displ,   
 y=hwy)) +  
 geom\_point() +  
 facet\_wrap(~ class,   
 nrow = 2) # nrow:가로, ncol: 세로



ggplot(data=mpg,   
 mapping=aes(x=displ,   
 y=hwy)) +  
 geom\_point() +  
 facet\_wrap(~ class,   
 ncol = 2) # nrow:가로, ncol: 세로

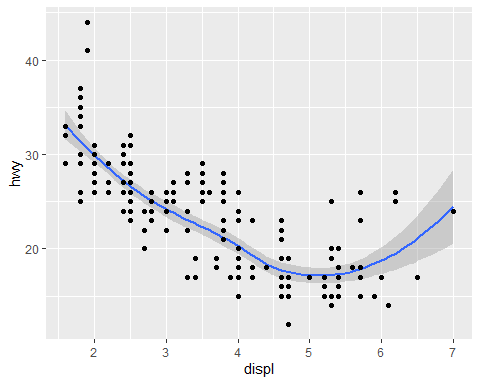
 # 구분변수가 2개일때

ggplot(data=mpg,   
 mapping=aes(x=displ,   
 y=hwy)) +  
 geom\_point() +  
 facet\_wrap(cyl ~ class)

 # 07.여러개 geom 함수 사용 # 여러 개 함수를 중복해서 사용 가능

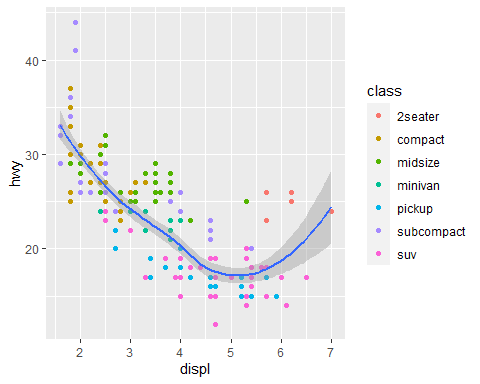
ggplot(data = mpg,  
 mapping = aes(x = displ,   
 y = hwy)) +   
 geom\_smooth() +  
 geom\_point()

## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'

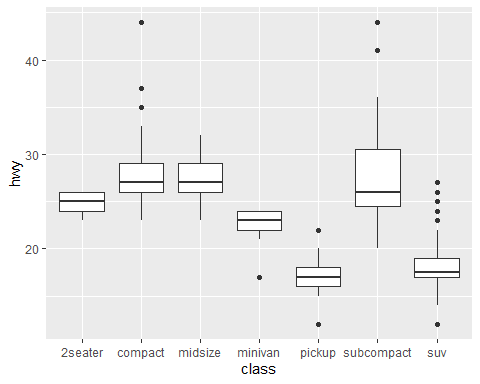
 # 변수에 매핑하기 위해서는 aes 사용

ggplot(data = mpg,   
 mapping = aes(x = displ,   
 y = hwy)) +   
 geom\_point(mapping = aes(color = class)) +  
 geom\_smooth()

## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'

 # 08.Coordinate systems # 수평 <-> 수직

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



ggplot(data = mpg,   
 mapping = aes(x = class,   
 y = hwy)) +   
 geom\_boxplot() +  
 coord\_flip()

