ggplot2 & Data analysis

Geonwoo Ban

2021 6 30

Visualization with mpg data set

```
data(mpg)
#?mpg
mpg %>% summary()
```

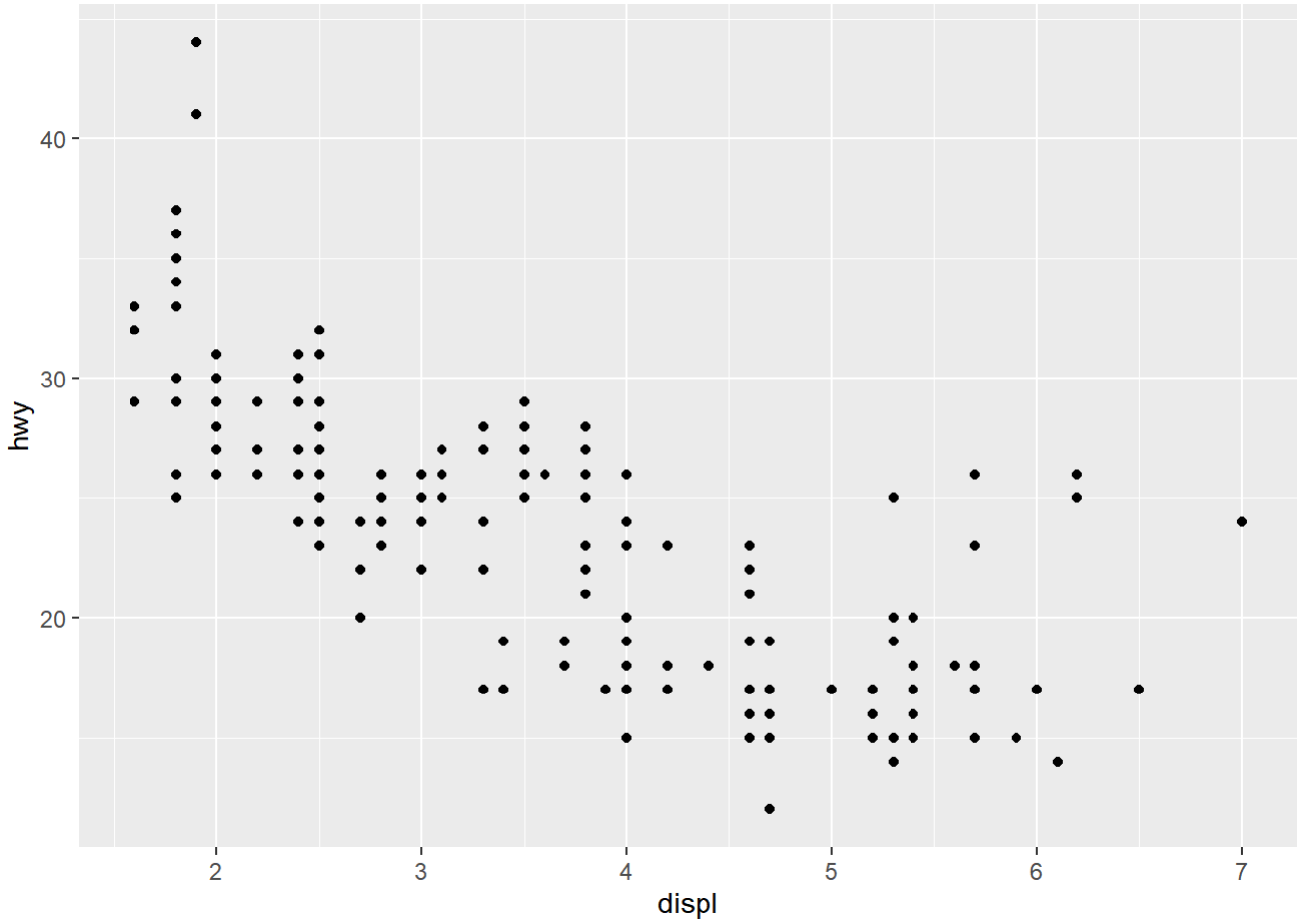
```
##  manufacturer          model               displ            year
##  Length:234         Length:234         Min.   :1.600   Min.   :1999
##  Class :character   Class :character   1st Qu.:2.400   1st Qu.:1999
##  Mode  :character   Mode  :character   Median :3.300   Median :2004
##                                        Mean   :3.472   Mean   :2004
##                                        3rd Qu.:4.600   3rd Qu.:2008
##                                        Max.   :7.000   Max.   :2008
##       cyl           trans               drv                 cty
##  Min.   :4.000   Length:234         Length:234         Min.   : 9.00
##  1st Qu.:4.000   Class :character   Class :character   1st Qu.:14.00
##  Median :6.000   Mode  :character   Mode  :character   Median :17.00
##  Mean   :5.889                                         Mean   :16.86
##  3rd Qu.:8.000                                         3rd Qu.:19.00
##  Max.   :8.000                                         Max.   :35.00
##       hwy             fl                class
##  Min.   :12.00   Length:234         Length:234
##  1st Qu.:18.00   Class :character   Class :character
##  Median :24.00   Mode  :character   Mode  :character
##  Mean   :23.44
##  3rd Qu.:27.00
##  Max.   :44.00
```

- 변수 type 확인하기
- 변수의 성질을 고려하지 않은 그림은 오히려 역효과
- 성질을 잘 파악하고 적재적소에 맞는 시각화를 하는 것이 매우 중요
- 항상 마인드는 비전공자가 본다는 생각으로 최대한 직관적으로 시각화하기

conti vs conti

1. displ : 차의 엔진 사이즈
2. hwy : 고속도로 연비

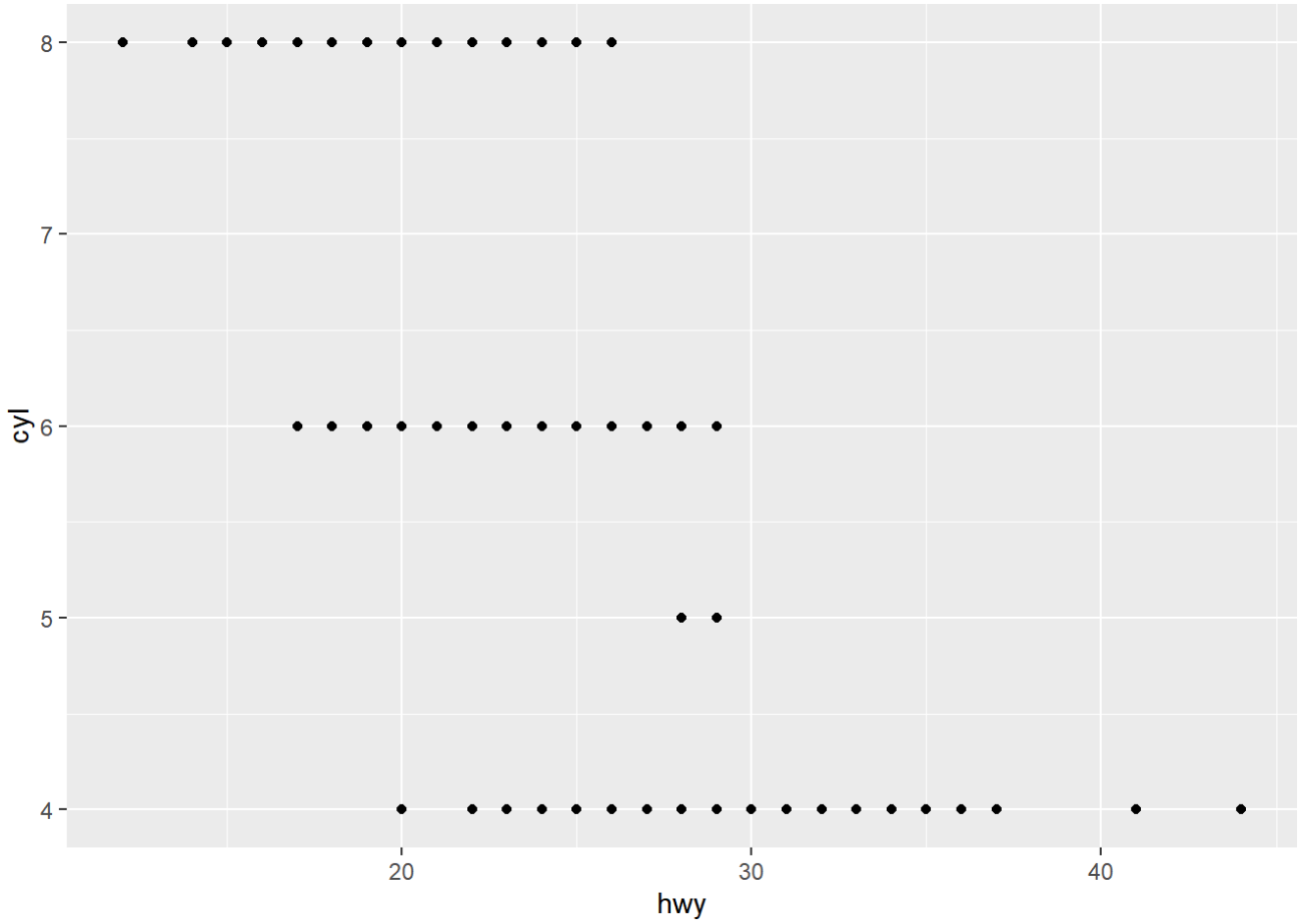
```
mpg %>%
  ggplot()+geom_point(aes(x=displ, y=hwy))
```



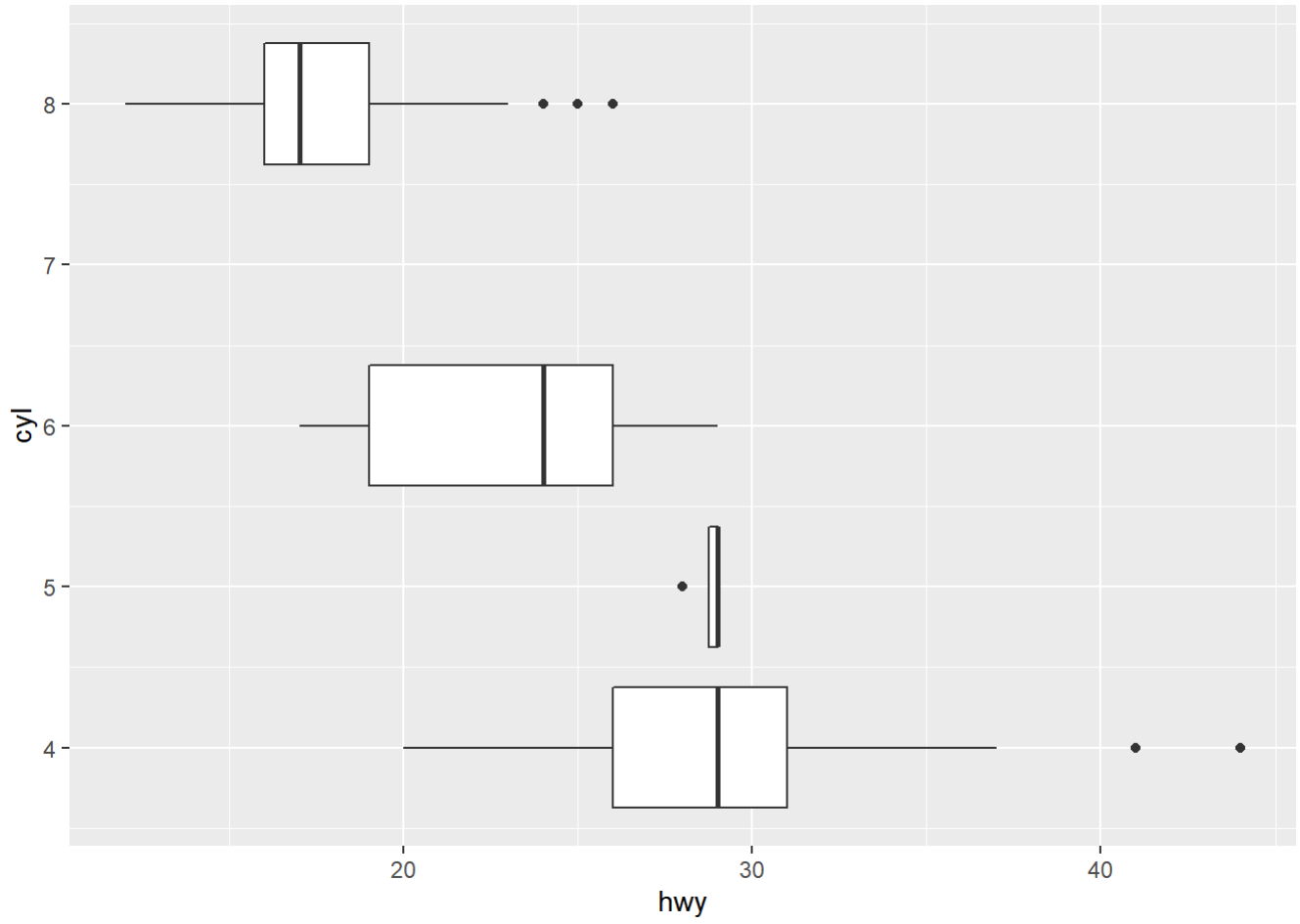
conti vs categorical

1. hwy : 고속도로 연비
2. cyl : 실린더 개수

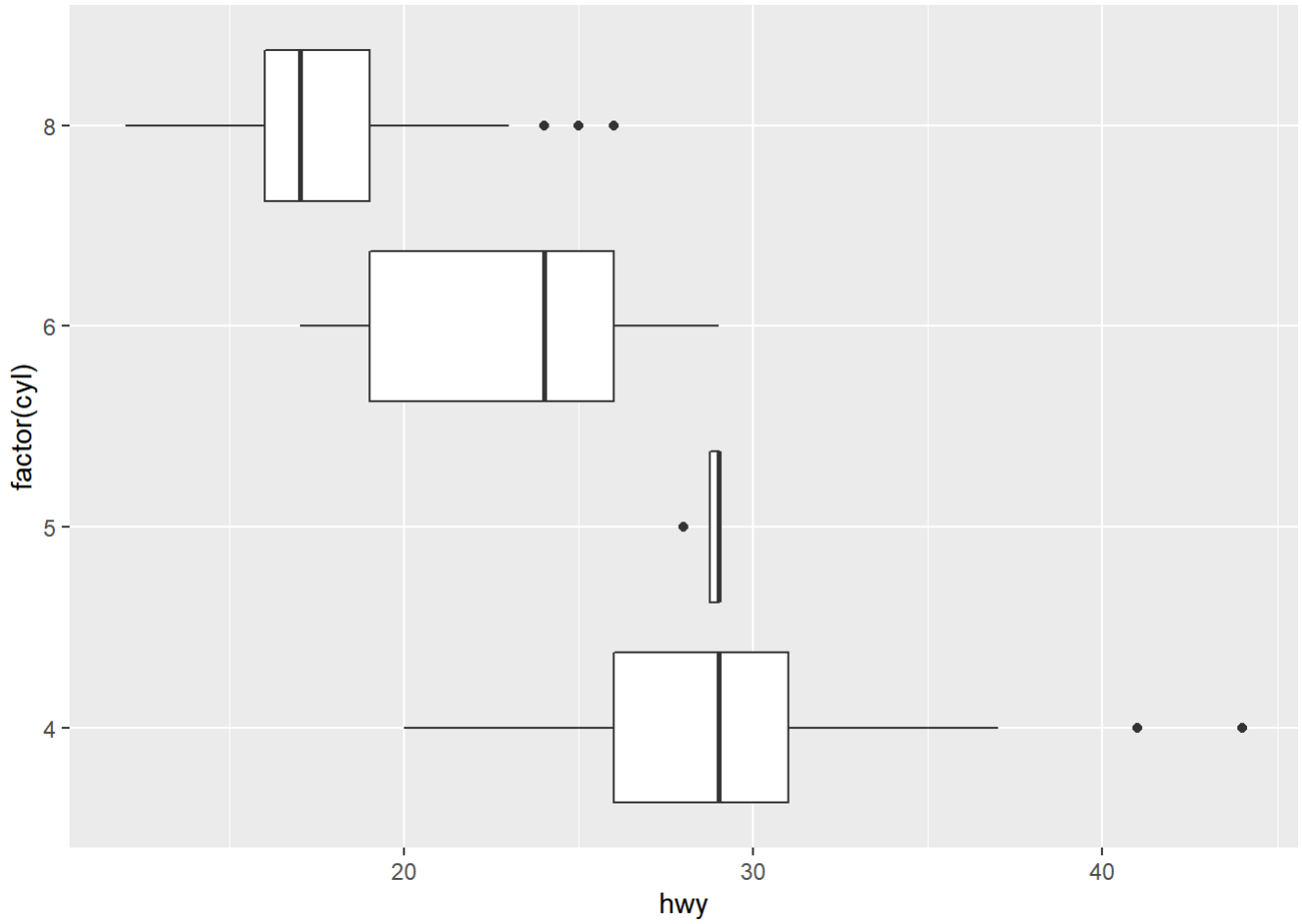
```
mpg %>%
  ggplot()+geom_point(aes(x=hwy, y=cyl))
```




```
mpg %>%
  ggplot()+geom_boxplot(aes(x=hwy, y=cyl, group=cyl))
```



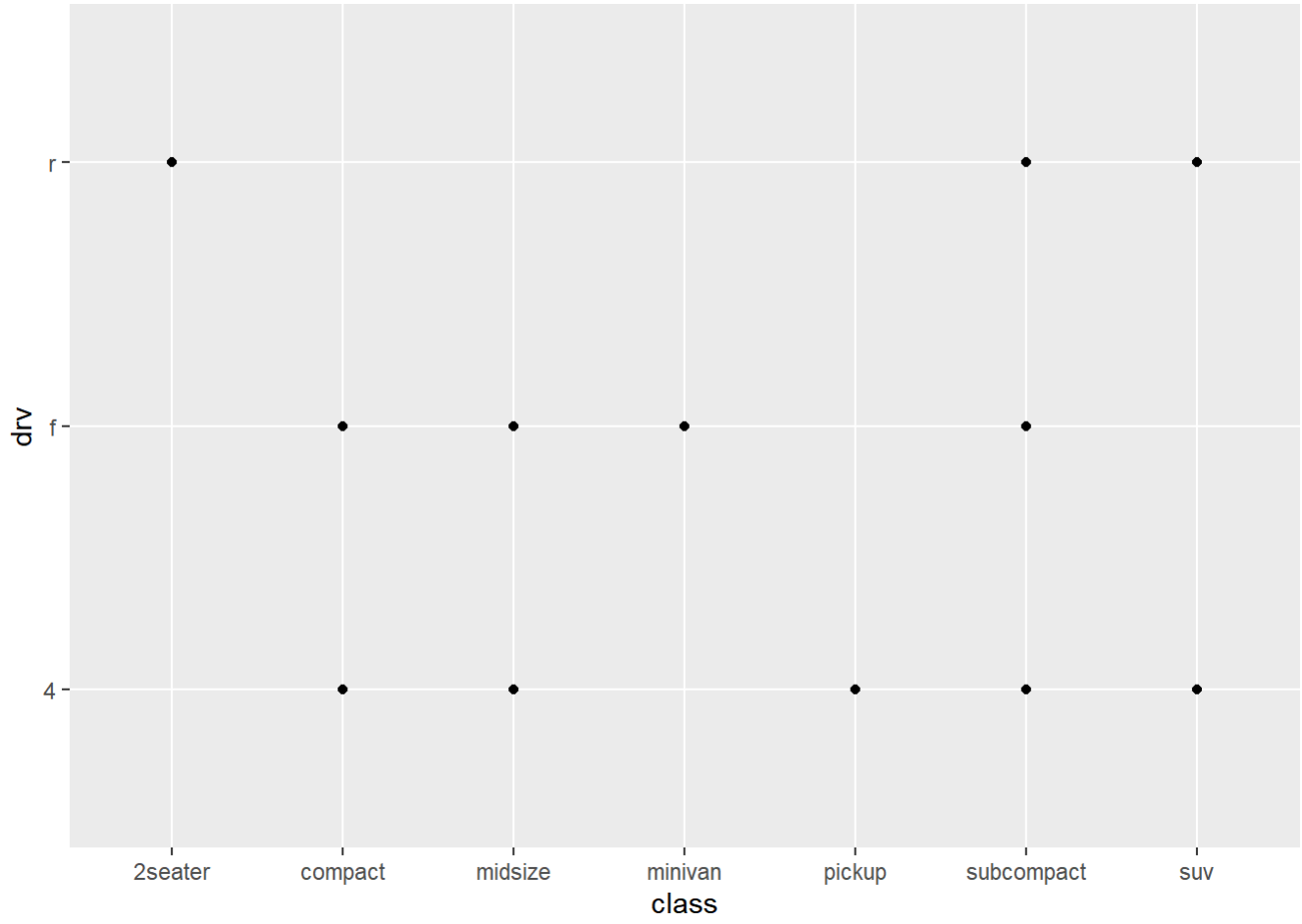

```
mpg %>%
  ggplot()+geom_boxplot(aes(x=hwy, y=factor(cyl), group=factor(cyl)))
```


8
6
5
4
factor(cyl)
20
30
40
hwy

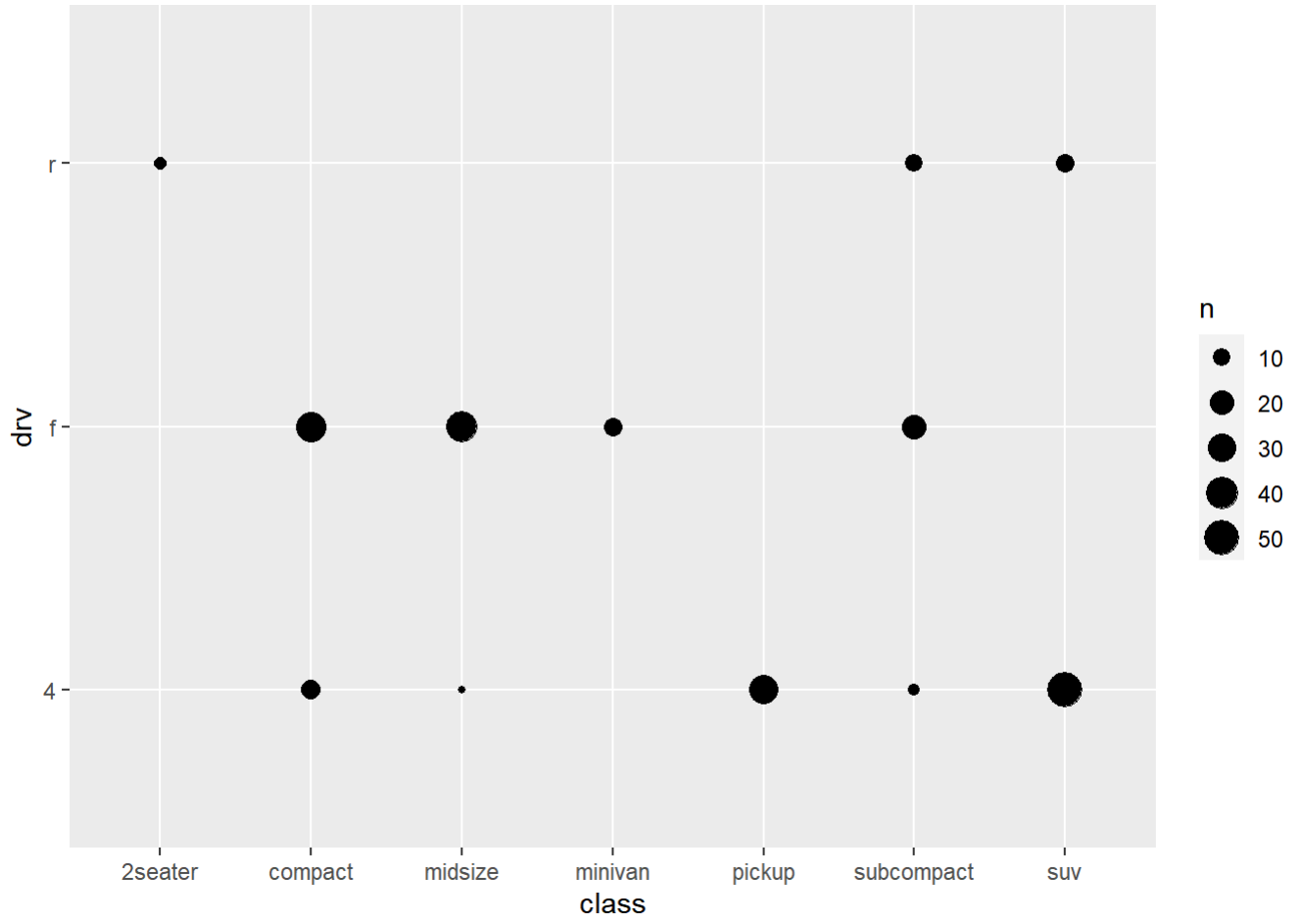
cate vs cate

1. class : 차종
2. drv : 전륜, 후륜, 사륜

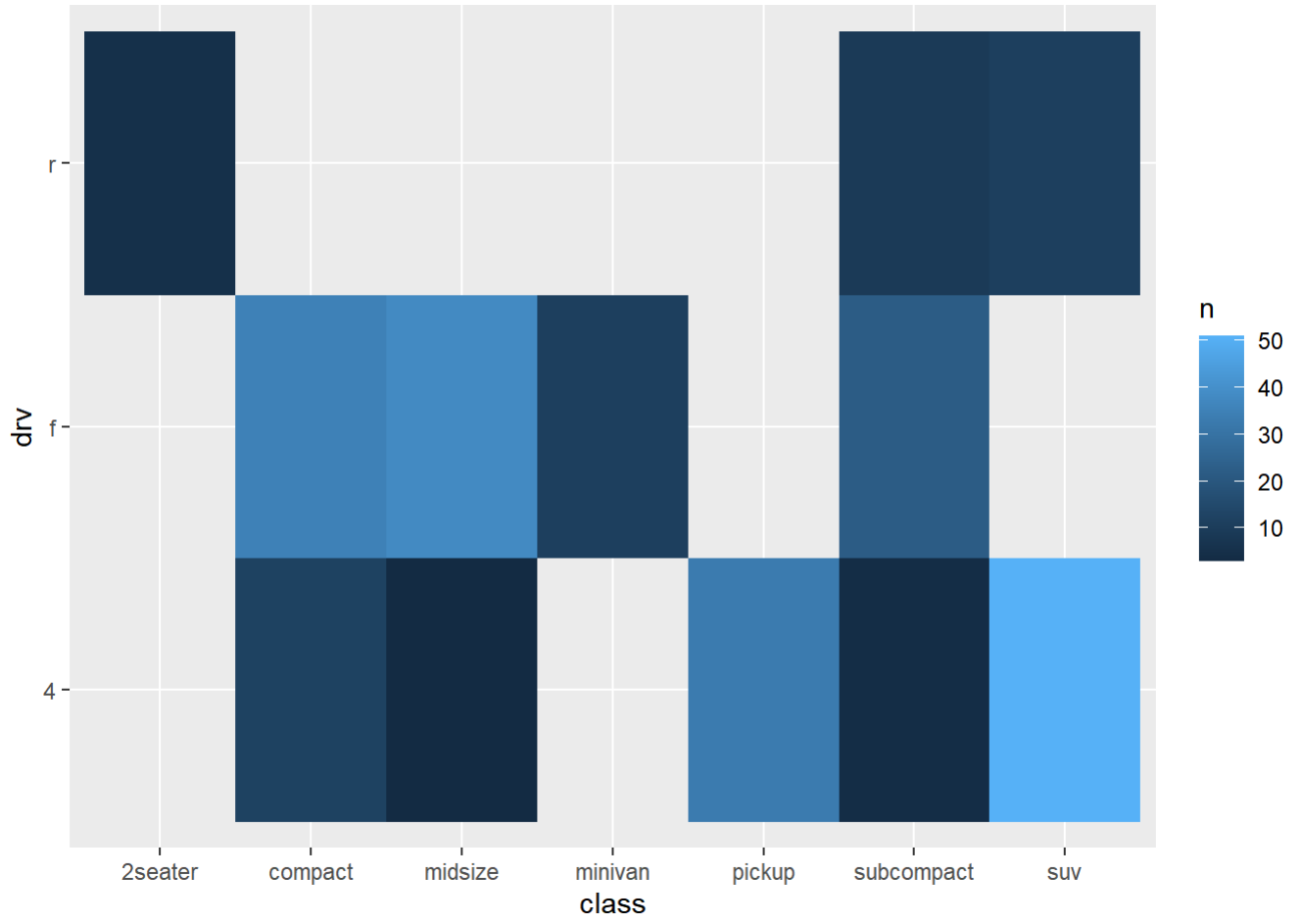
```
mpg %>%
  ggplot()+geom_point(aes(x=class, y=drv))
```




```
mpg %>%
  ggplot(aes(x=class, y=drv))+
  geom_count()
```

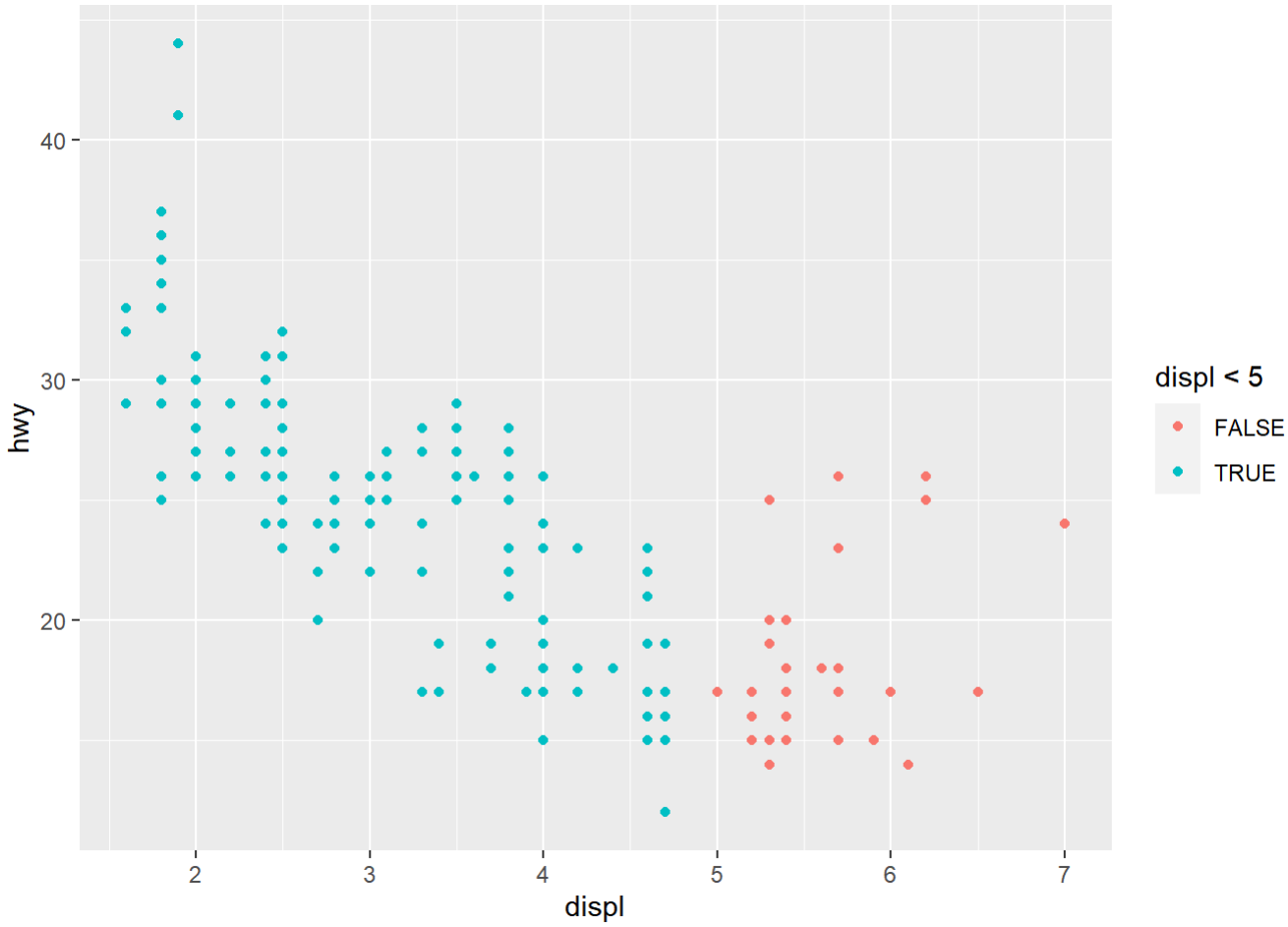



```
mpg %>%
  count(class, drv) %>%
  ggplot(aes(x=class, y=drv))+
  geom_tile(aes(fill=n))
```

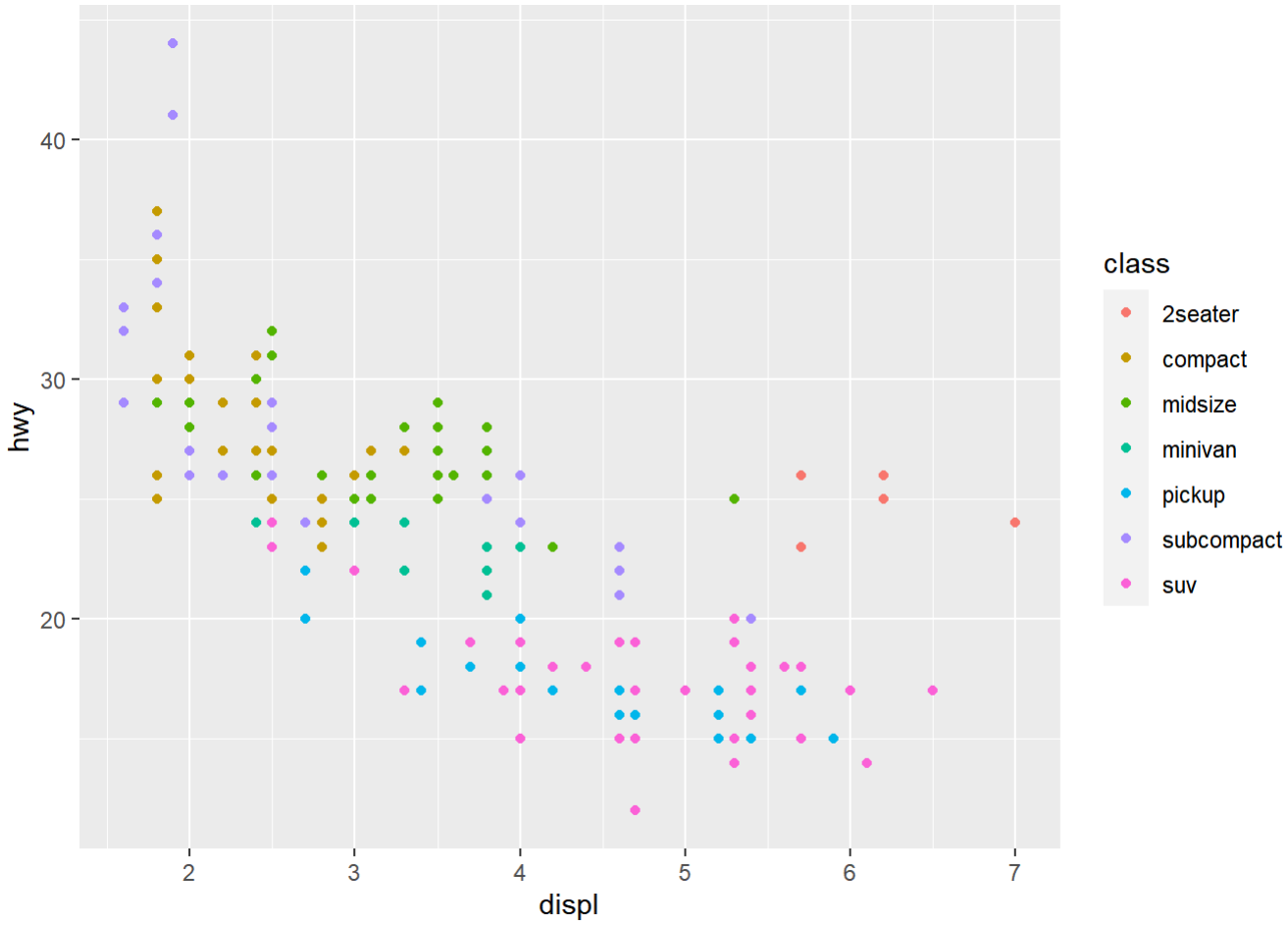

drv
r
f
4
2seater
compact
midsize
minivan
pickup
subcompact
suv
class
n
50
40
30
20
10

coloring

```
mpg %>%
  ggplot()+geom_point(aes(x=displ, y=hwy, color=displ<5))
```

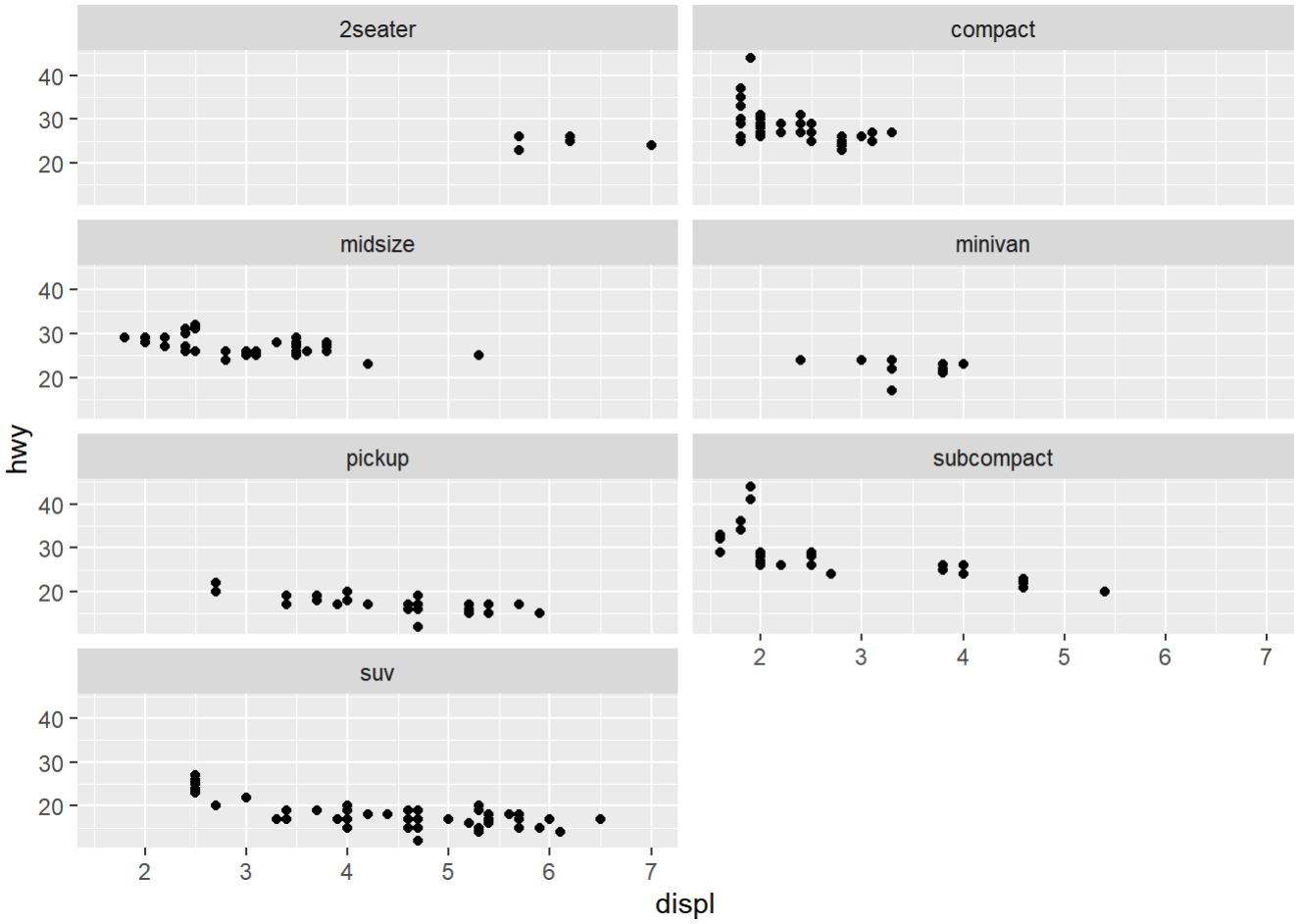



```
mpg %>%
  ggplot()+
  geom_point(aes(x=displ, y=hwy, color=class))
```

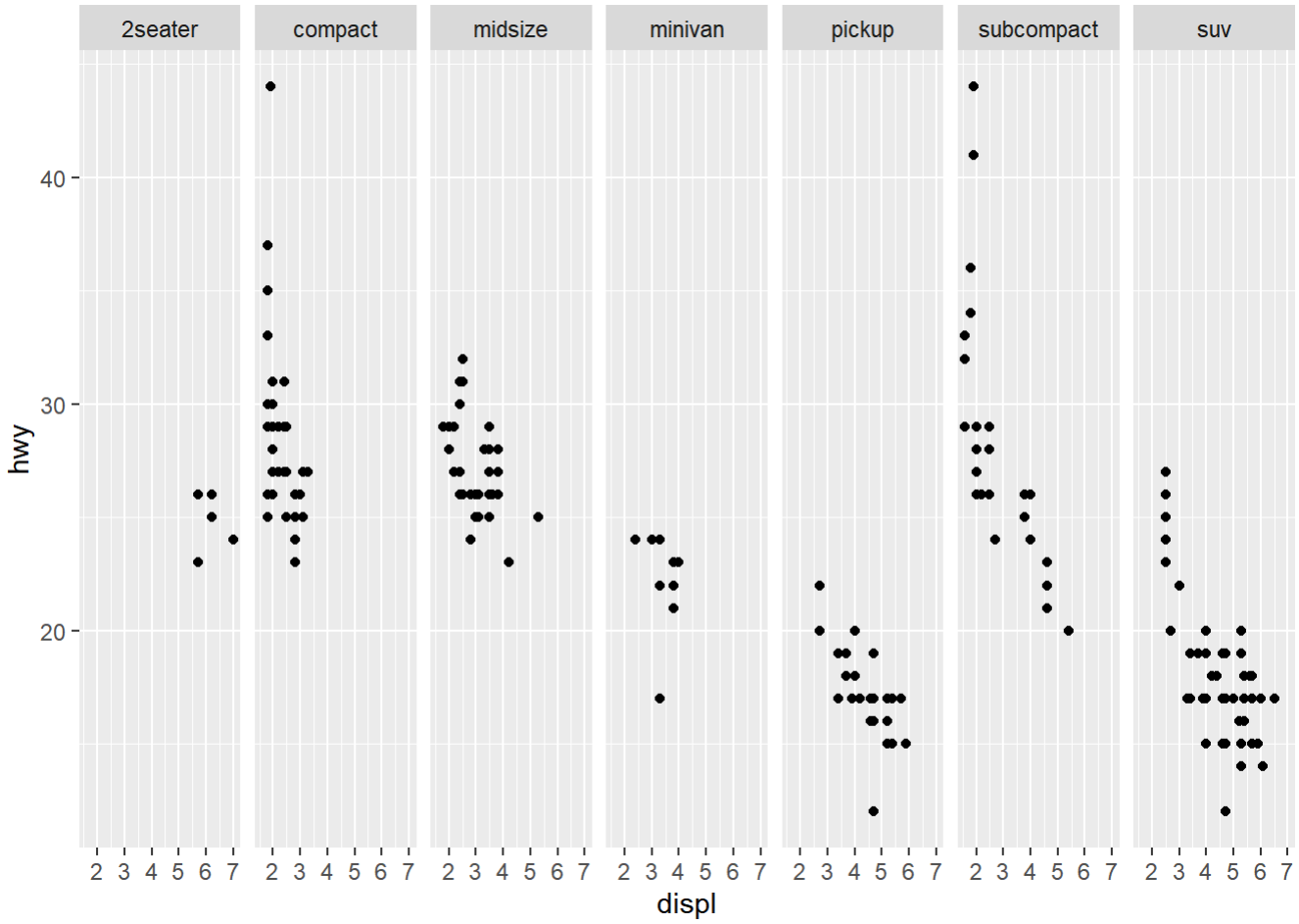

hwy
40
30
20
2
3
4
5
6
7
displ
class
2seater
compact
midsize
minivan
pickup
subcompact
suv

facet

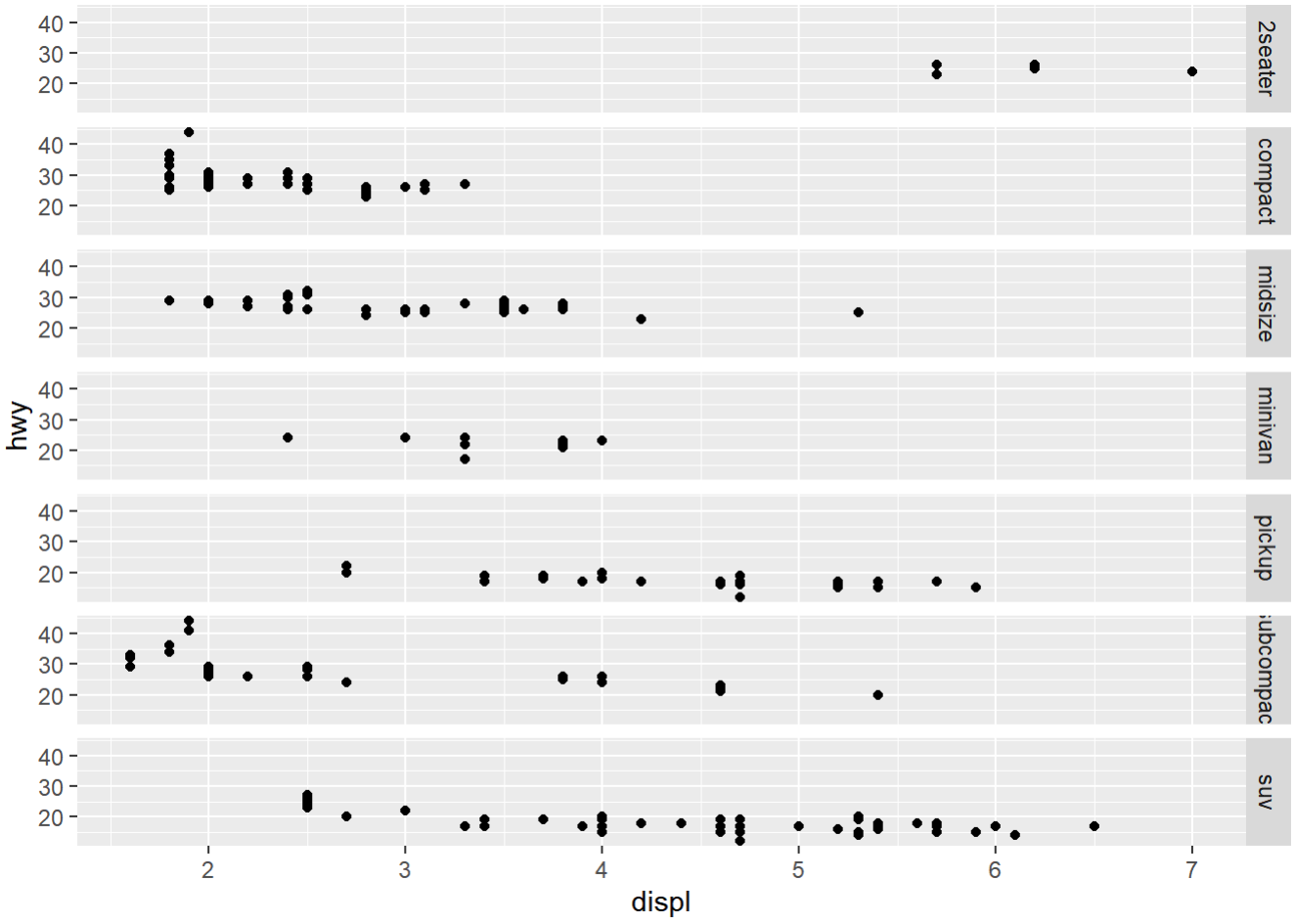
```
mpg %>%
  ggplot()+
  geom_point(aes(x=displ, y=hwy))+
  facet_wrap(~class, nrow=5)
```



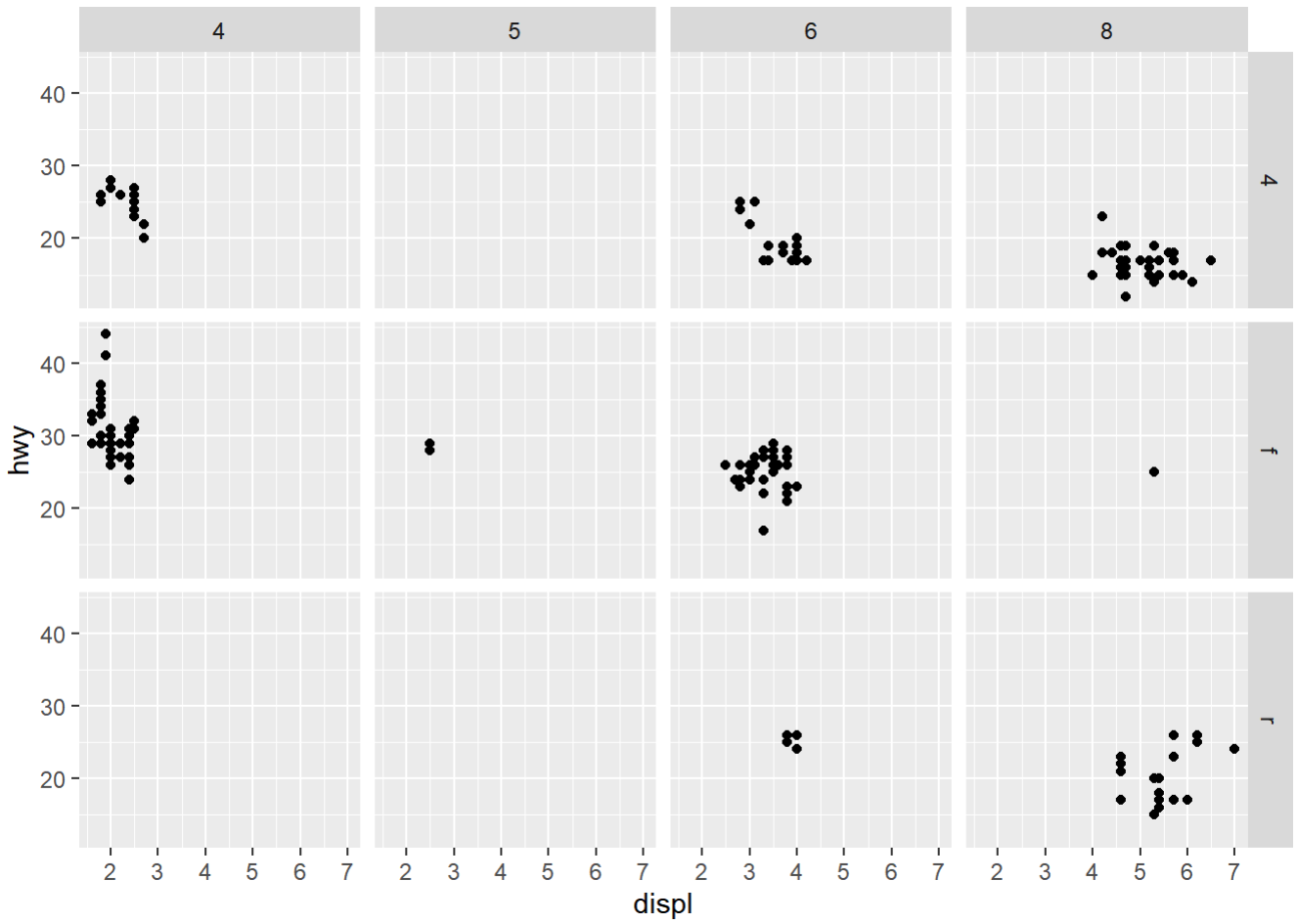

```
mpg %>%
  ggplot()+
  geom_point(aes(x=displ, y=hwy))+
  facet_grid(~class)
```




```
mpg %>%
  ggplot()+
  geom_point(aes(x=displ, y=hwy))+
  facet_grid(class~.)
```




```
mpg %>%
  ggplot()+
  geom_point(aes(x = displ, y = hwy))+
  facet_grid(drv~cyl)
```


4
5
6
8
4
f
r
hwy
displ
40
30
20
2
3
4
5
6
7

soomth line

```
ggplot(data = mpg, mapping = aes(x = displ, y = hwy, colour = drv)) +
  geom_point() +
  geom_smooth(se = FALSE)
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```