

## L10. *ggplot* (4)

Preview on more functionality

Sim, Min Kyu, Ph.D., [mksim@seoultech.ac.kr](mailto:mksim@seoultech.ac.kr)



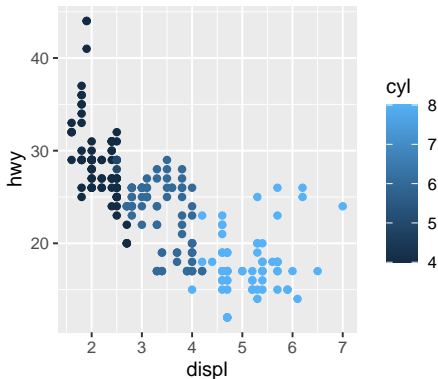
서울과학기술대학교 데이터사이언스학과

*position = "jitter"*

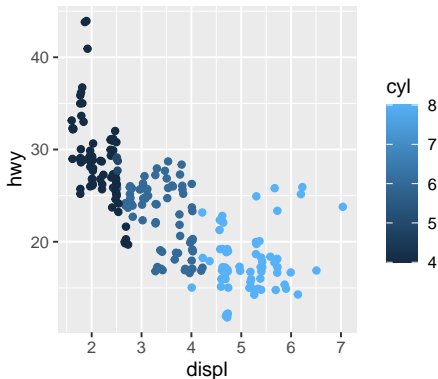
데이터들이 중첩되는 경우에 노이즈를 넣어주는 기능

```
a <- ggplot(mpg) + geom_point(aes(displ, hwy, color = factor(cyl)))  
b <- ggplot(mpg) + geom_point(aes(displ, hwy, color = factor(cyl)), position = "jitter")
```

a



b



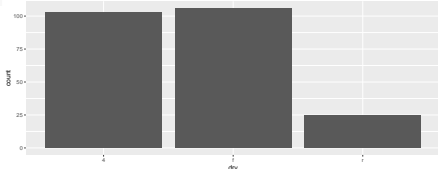
## Barchart (막대그래프)

- X축은 Categorical 변수, Y축은 X변수의 count 또는 density
- `geom_bar()`의 aesthetics: `x`, `alpha`, `color`, `fill`, `linetype`, `size`, `weight`

- 1 변수로 그리는 경우

- ▶ y의 aesthetic으로 x변수를 count하여 사용한다.
- ▶ 즉, `geom_bar()`의 default stat은 count이다.
- ▶ 즉, `geom_bar()`의 인수로 `stat="count"`가 포함되어 있다.

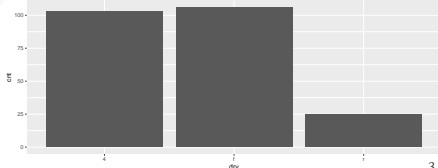
```
ggplot(mpg) +  
  geom_bar(aes(x=drv))
```



- 2 변수로 그리는 경우

- ▶ Frequency table이 있는 상황에서 사용한다.
- ▶ y의 aesthetic을 사용하기 때문에 `stat = "identity"`을 명시해준다.

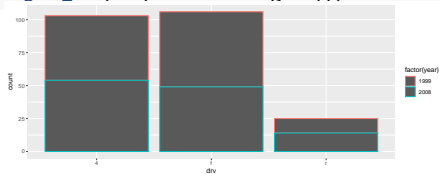
```
df <- mpg %>%  
  group_by(drv) %>%  
  summarise(cnt = n())  
ggplot(df) +  
  geom_bar(aes(x=drv, y=cnt), stat = "identity")
```



## `geom_bar()`에 다른 Categorical 변수로 차원을 추가

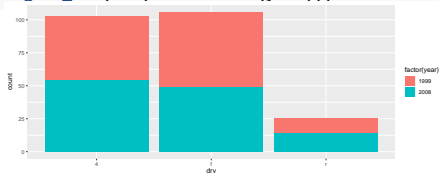
`color`는 테두리에 관여하는 aesthetic 요소이다.

```
ggplot(mpg, aes(x=drv)) +  
  geom_bar(aes(color=factor(year)))
```



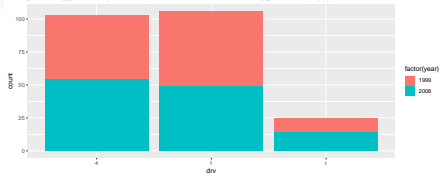
`fill`은 색을 채우는 aesthetic 요소이다.

```
ggplot(mpg, aes(x=drv)) +  
  geom_bar(aes(fill=factor(year)))
```

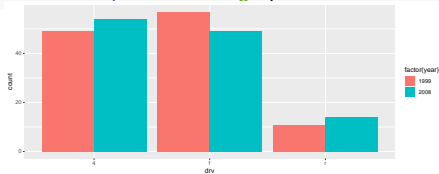


*position="dodge"*으로 비껴서 배치

```
ggplot(mpg, aes(x=drv)) +  
  geom_bar(aes(fill=factor(year)))
```

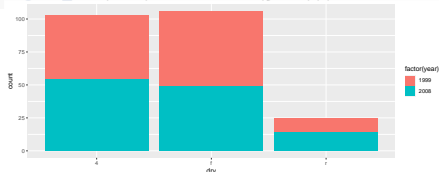


```
ggplot(mpg, aes(x=drv)) +  
  geom_bar(aes(fill=factor(year)),  
    position="dodge")
```

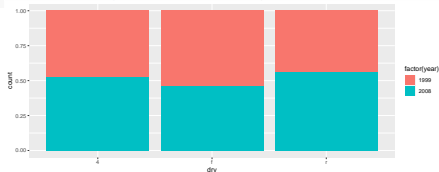


*position = "fill"*로 갯수가 아닌 비율에 초점을 맞추기

```
ggplot(mpg, aes(x=drv)) +  
  geom_bar(aes(fill=factor(year)))
```



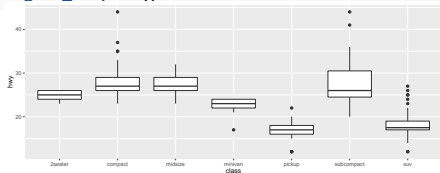
```
ggplot(mpg, aes(x=drv)) +  
  geom_bar(aes(fill=factor(year)),  
    position="fill")
```



## geom\_boxplot() vs geom\_violin()

geom\_boxplot()

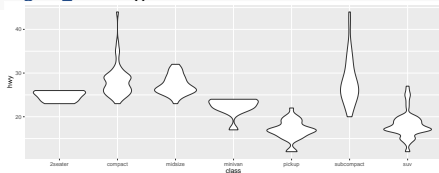
```
ggplot(mpg, aes(x = class, y = hwy)) +  
  geom_boxplot()
```



- 1 박스의 가운데 직선은 'median'
- 2 박스의 상하단은 'hinges'
- 3 선의 상하단은 'whiskers'
- 4 점으로 표현된 것들은 'outliers'

geom\_violin()

```
ggplot(mpg, aes(x = class, y = hwy)) +  
  geom_violin()
```

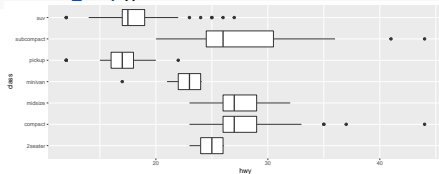


- Density curve를 회전시킨 모양

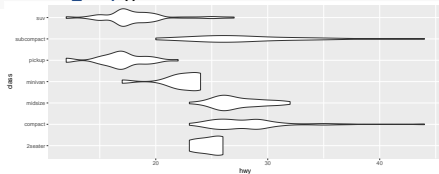
두 차트에서 X축의 공간이 부족하다면, 이런 경우에는 어떻게 할까?

## coord\_flip()

```
ggplot(mpg, aes(x = class, y = hwy)) +  
  geom_boxplot() +  
  coord_flip()
```



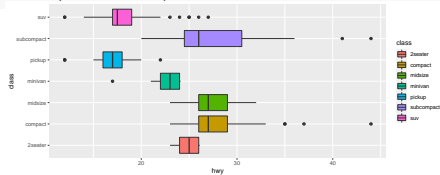
```
ggplot(mpg, aes(x = class, y = hwy)) +  
  geom_violin() +  
  coord_flip()
```



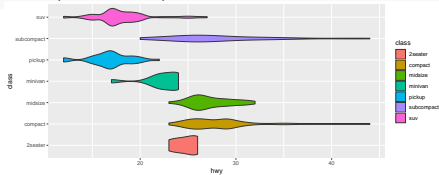


## aes()를 나중에 더해도 될까? 된다!

```
ggplot(mpg, aes(x = class, y = hwy)) +  
  geom_boxplot() +  
  coord_flip() +  
  aes(fill=class)
```

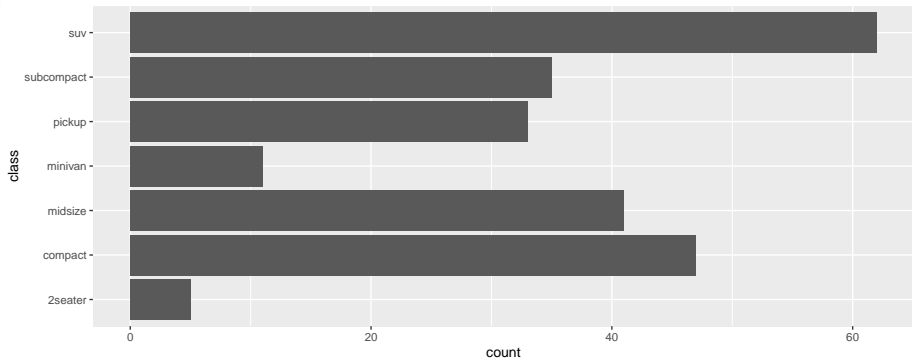


```
ggplot(mpg, aes(x = class, y = hwy)) +  
  geom_violin() +  
  coord_flip() +  
  aes(fill=class)
```



`coord_flip()`은  $x$ 축과  $y$ 축을 뒤집는다.

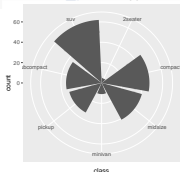
```
ggplot(mpg, aes(x = class)) +  
  geom_bar() + coord_flip()
```



파이차트는 어떻게 만들어야 할까? 바차트와 파이차트의 차이는 무엇인가??

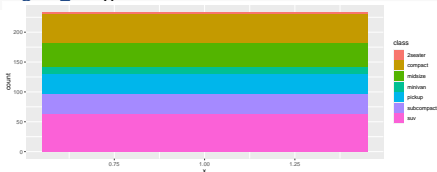
`coord_polar()`는 바차트를 파이차트로 만든다.

```
ggplot(mpg, aes(x = class)) +  
  geom_bar() + coord_polar()
```

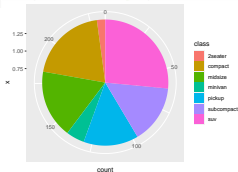


`coord_polar()`에서 aesthetic이 되는 `theta`는 y축에 대응된다. 그러므로...

```
ggplot(mpg, aes(x = 1, fill = class)) +  
  geom_bar()
```



```
ggplot(mpg, aes(x = 1, fill = class)) +  
  geom_bar() +  
  coord_polar(theta = "y")
```



- axis를 지우면 파이차트 완성!
- ggplot 개발의 motivation이 되었다고 한다.
- ~~그에 비해서 ggplot에서 만드는 것이 간단하지 않다.~~
- 파이차트는 구시대적이다?

```
"Tantum videmus quantum scimus."
```

```
## [1] "Tantum videmus quantum scimus."
```