

## geom\_jitter ¶

**geom\_point**에서 각각의 점의 위치를 범위 내에서 무작위로 수평분산 시킨다.

```
In [2]: library(ggplot2)
```

**미국 중서부 5개 주 카운티를 대상으로 지역정보를 조사한 midwest 데이터**

```

poptotal      : 전체 인구
popdensity    : 인구 밀도
popwhite      : 백인의 수.
popblack      : 흑인의 수.
popamerindian : 아메리칸 인디언 수.
popasian      : 아시아 인의 수.
popother      : 다른 종족의 수.
percwhite     : 백인의 비율(백분율)
percblack     : 흑인의 비율(백분율)
percamerindan : 아메리카 인디언의 비율
percasian     : 아시아인의 비율
percother     : 다른 인종의 비율
popadults     : 성인의 수
perchsd
percollege    : 학사
대학교육을 받은 비율
percpof       : 전문인 비율
poppovertyknown
percpover tyknown
percbelowpoverty
percchildbelowpovert
percadultpoverty
percelderlypoverty
inmetro
In a metro area.

```

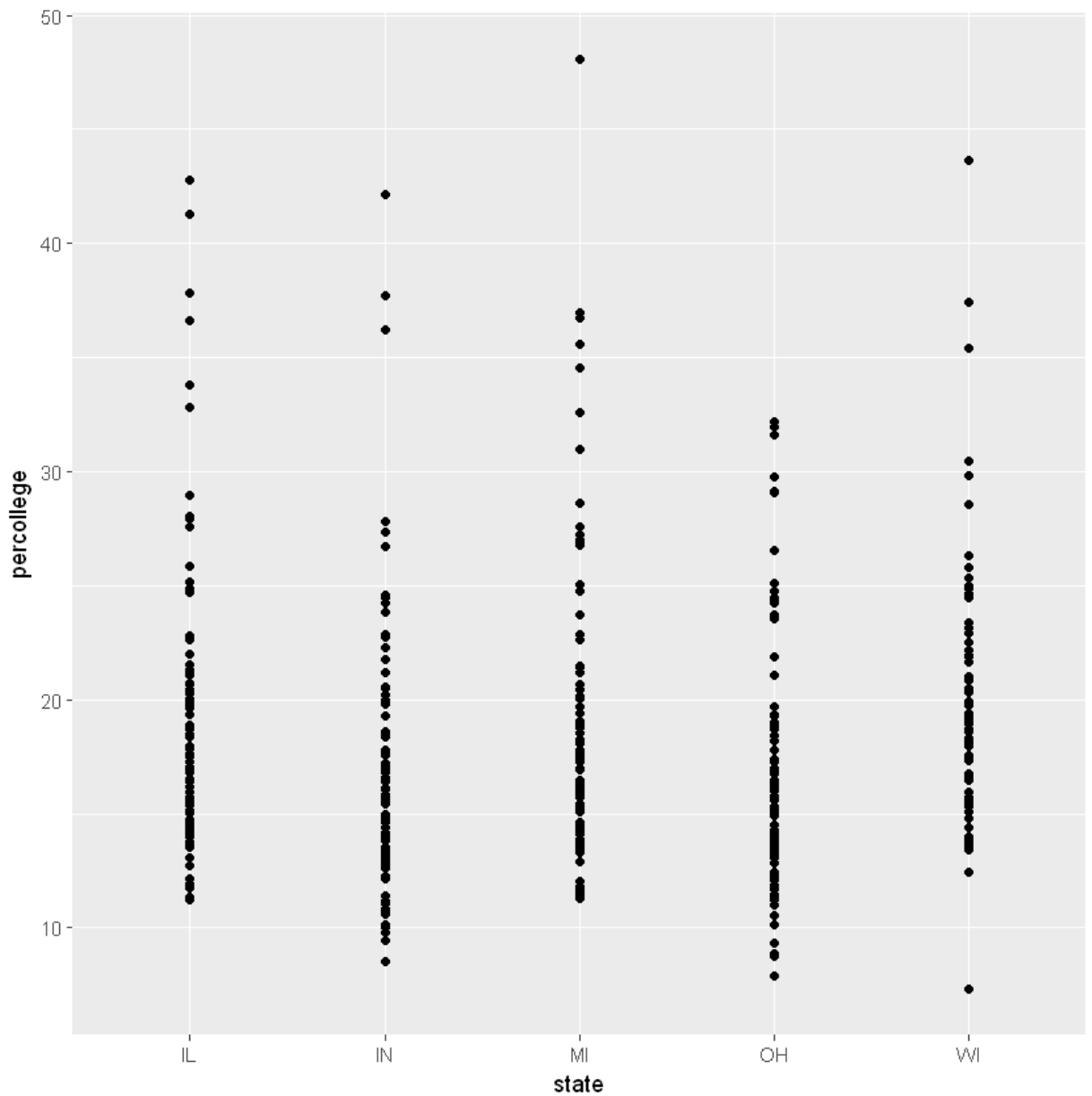
```
In [4]: head(midwest)
```

PID	county	state	area	poptotal	popdensity	popwhite	popblack	popamerindian	popasian	...	percollege	percpr
561	ADAMS	IL	0.052	66090	1270.9615	63917	1702	98	249	...	19.63139	4.3558
562	ALEXANDER	IL	0.014	10626	759.0000	7054	3496	19	48	...	11.24331	2.8703
563	BOND	IL	0.022	14991	681.4091	14477	429	35	16	...	17.03382	4.4885
564	BOONE	IL	0.017	30806	1812.1176	29344	127	46	150	...	17.27895	4.1978
565	BROWN	IL	0.018	5836	324.2222	5264	547	14	5	...	14.47600	3.3676
566	BUREAU	IL	0.050	35688	713.7600	35157	50	65	195	...	18.90462	3.2758

```
In [5]: head(midwest[,c('state', 'percollege')])
```

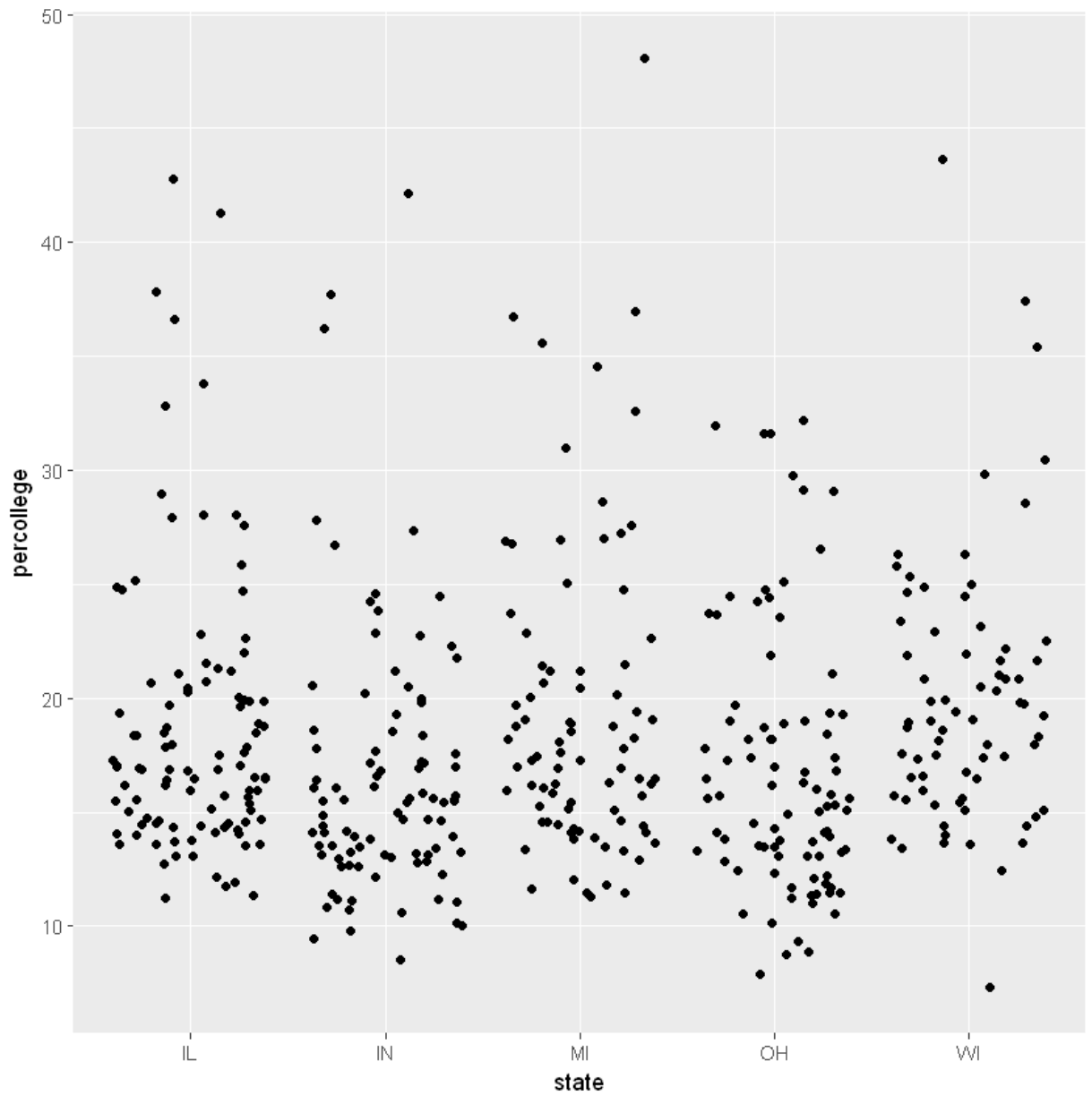
state	percollege
IL	19.63139
IL	11.24331
IL	17.03382
IL	17.27895
IL	14.47600
IL	18.90462

```
In [1]: # percollege 변수는 각각의 카운티에서 대학교육을 받은 사람의 비율을 나타내고,
# state는 중서부에 속한 다섯 개 주를 의미
ggplot(midwest, aes(state, percollege))+
  geom_point()
```



위와 같은 `geom_point` 함수의 가독성 문제를 보완하기 위해 `geom_jitter` 함수를 사용

```
In [6]: ggplot(midwest, aes(state, percollege))+  
        geom_jitter()
```



높이 조정

In [9]: # 총 넓어지는 범위는 height에 지정된 값의 두 배입니다. 디폴트 값은 0.4

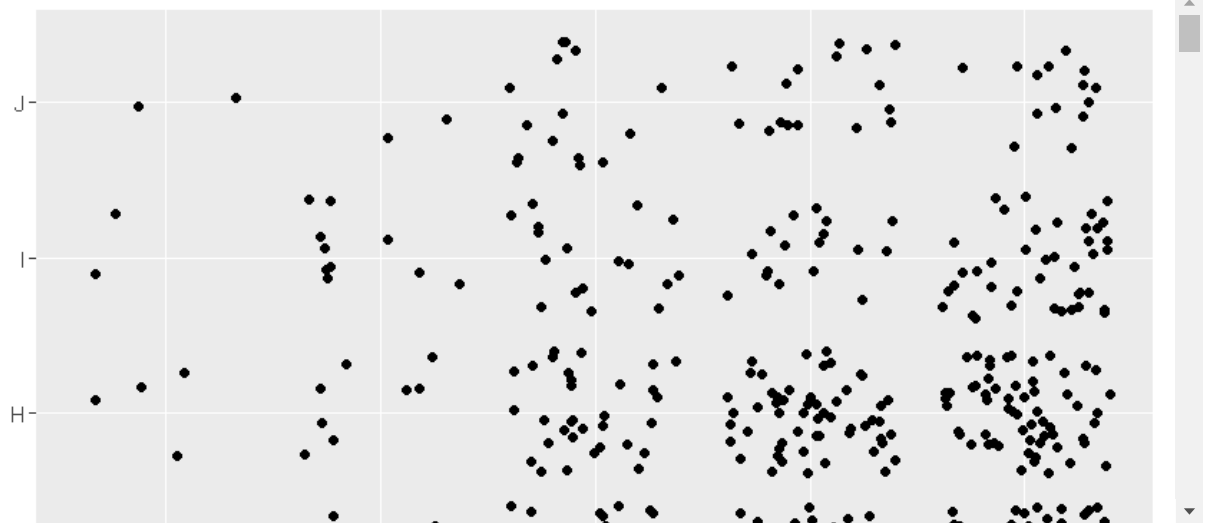
```
dia<-diamonds[sample(1:length(diamonds$carat), 1000), ]
```

```
ggplot(dia, aes(cut, color))+  
  geom_jitter()
```

```
ggplot(dia, aes(cut, color))+  
  geom_jitter(height = 0.1)
```

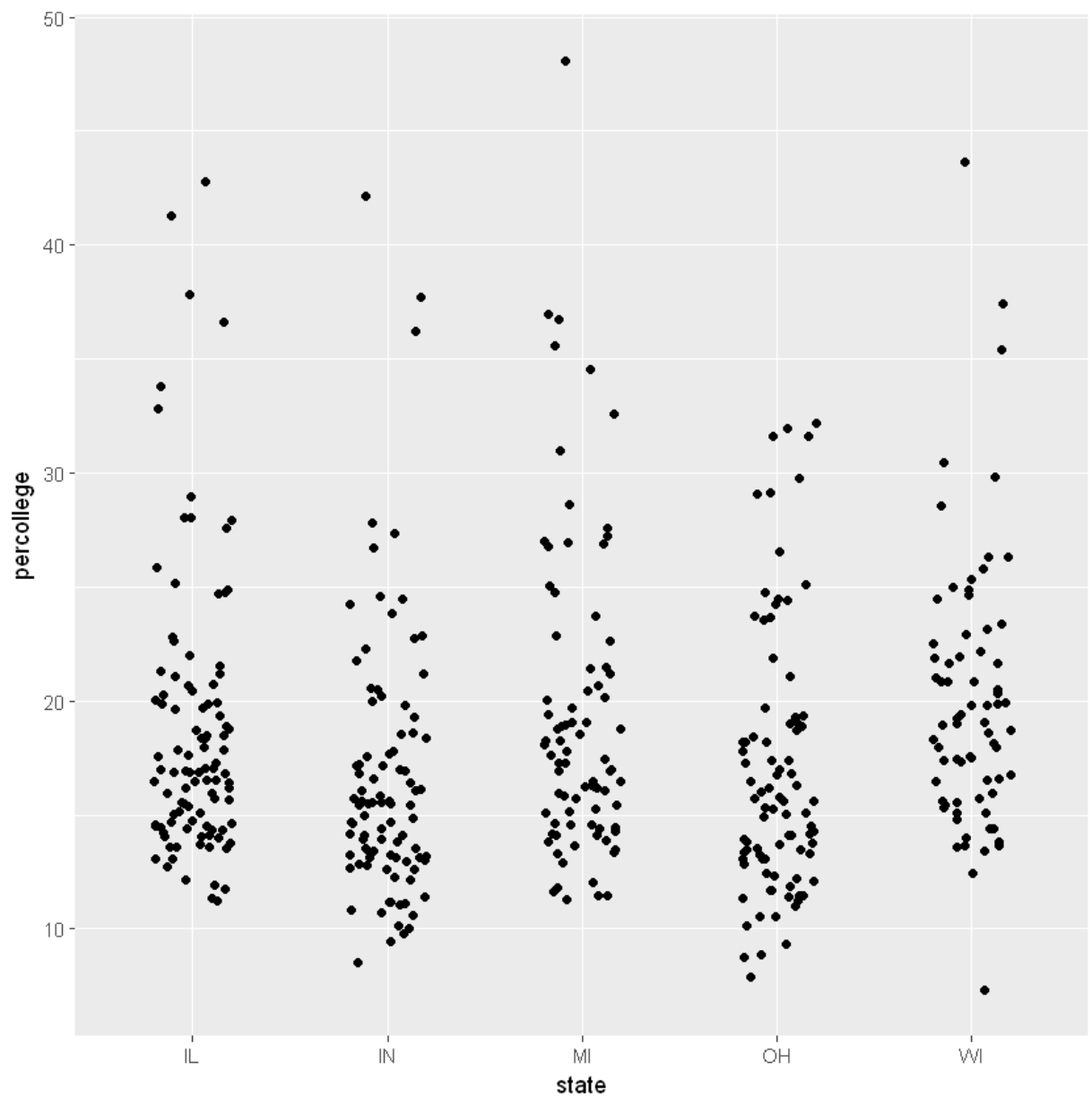
```
ggplot(dia, aes(cut, color))+  
  geom_jitter(height = 0.3)
```

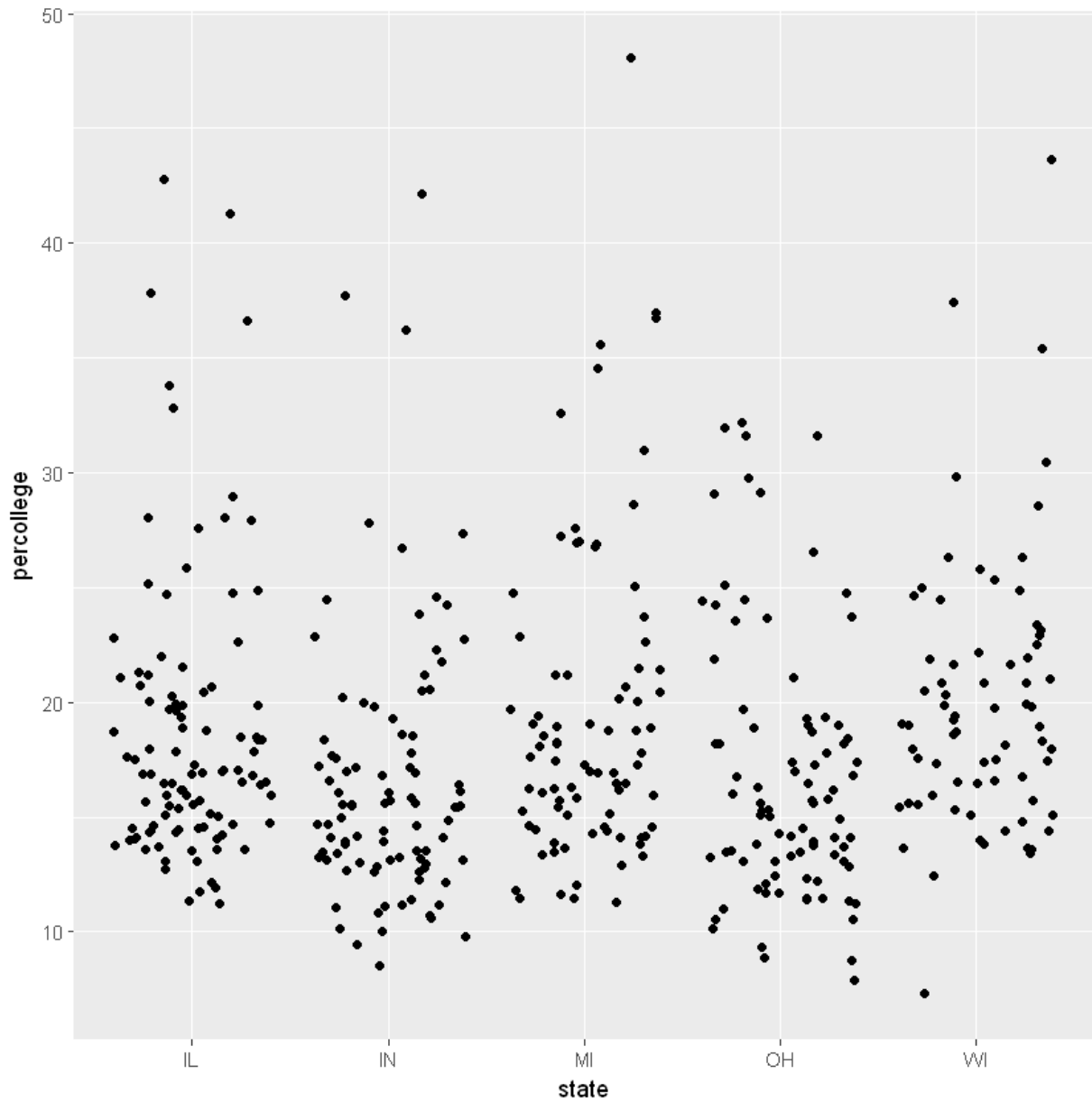
```
ggplot(dia, aes(cut, color))+  
  geom_jitter(height = 0.5)
```



너비 조정

```
In [12]: ggplot(midwest, aes(state, percollege))+  
  geom_jitter(width = 0.2)  
  
ggplot(midwest, aes(state, percollege))+  
  geom_jitter(width = 0.4)
```

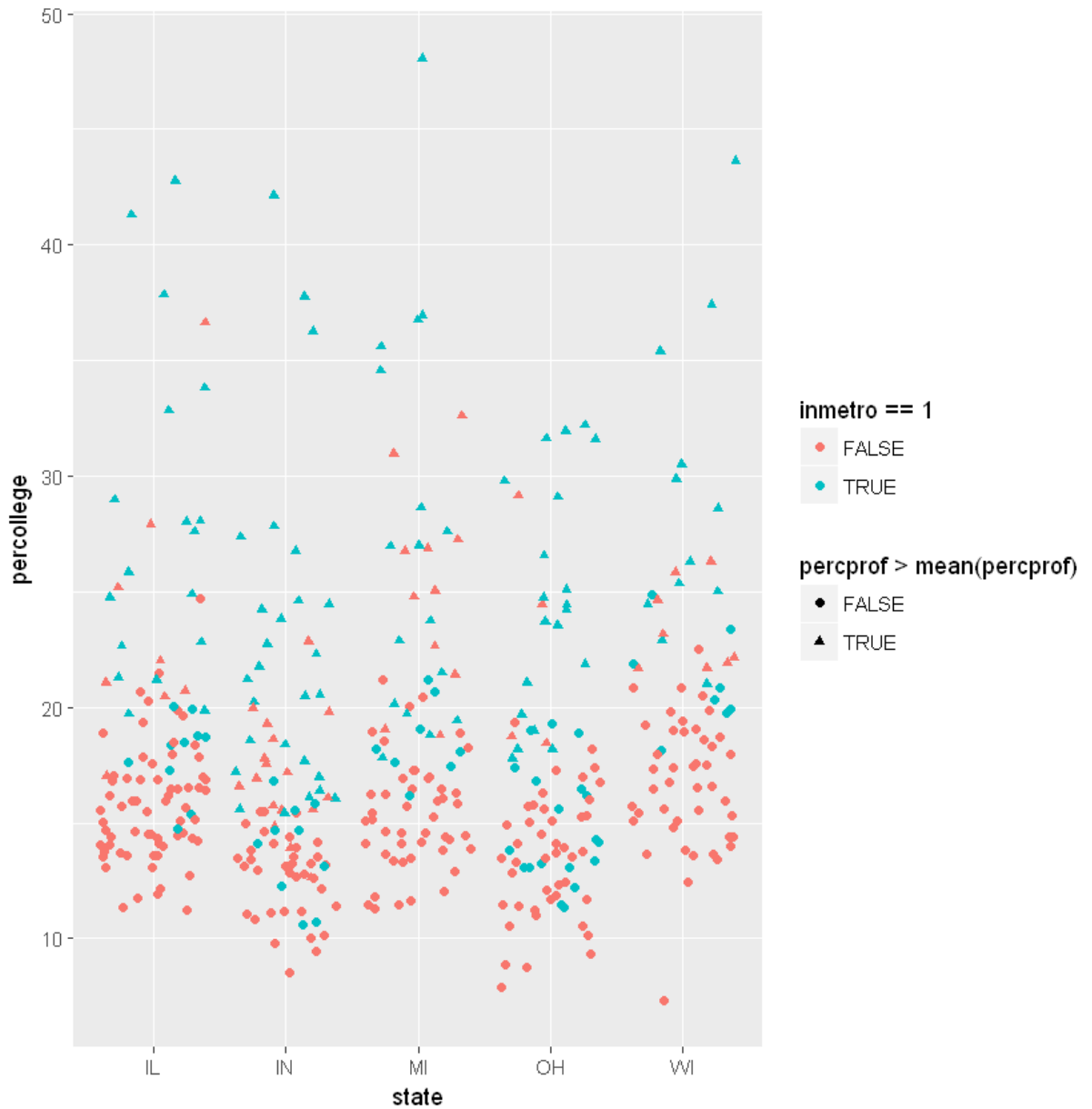




**color**

color 인수를 사용해 여러 조건들을 설정 가능

```
In [13]: ggplot(midwest, aes(state, percollege)) +
  geom_jitter(data=midwest, aes(color=inmetro==1,
                                shape=percprof>mean(percprof)))
```



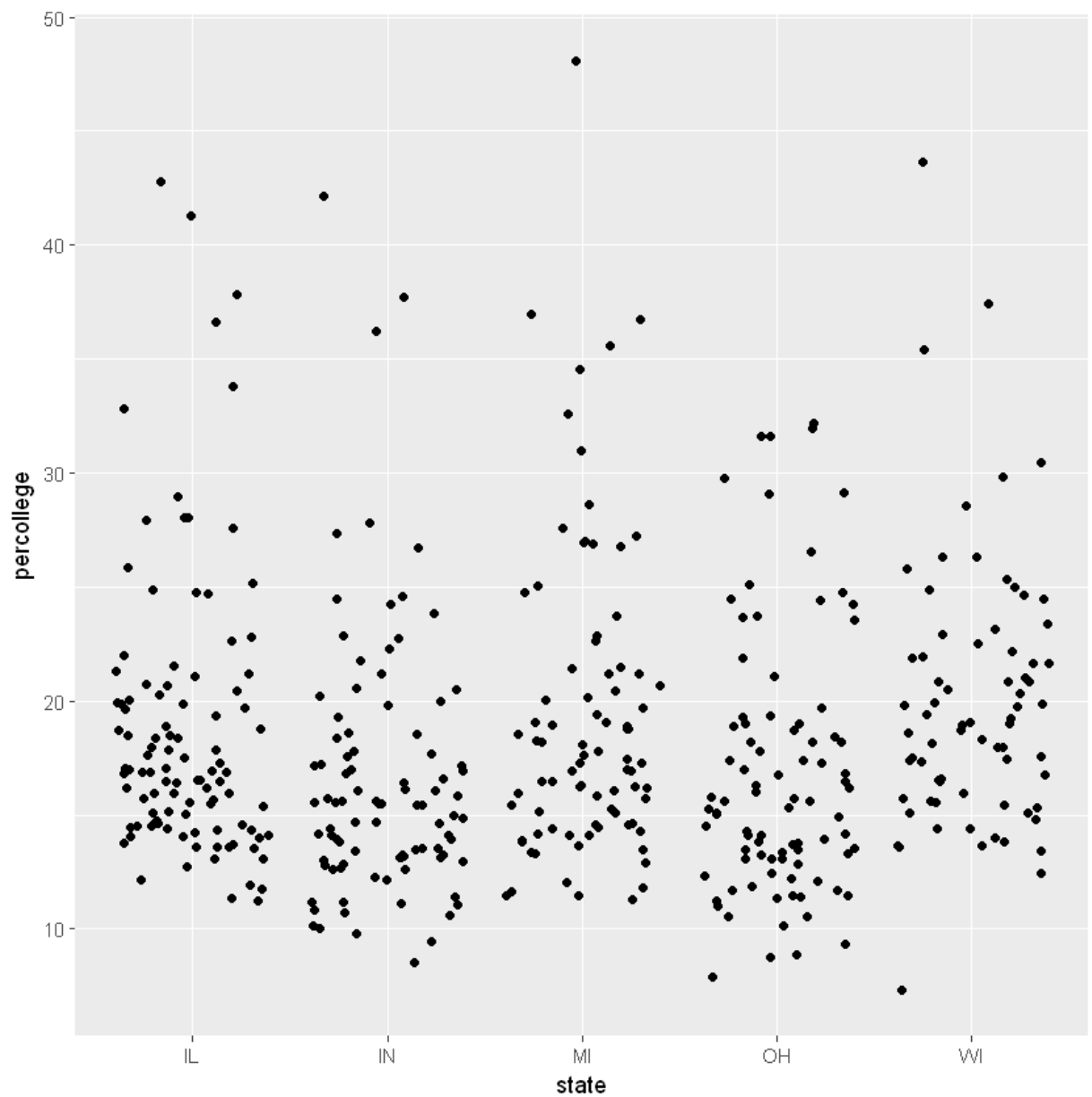
대도시권에 속해있는 카운티의 색을 다르게 지정하고

박사교육과정을 받은 비율이 평균보다 높은 카운티에는 모양을 변경

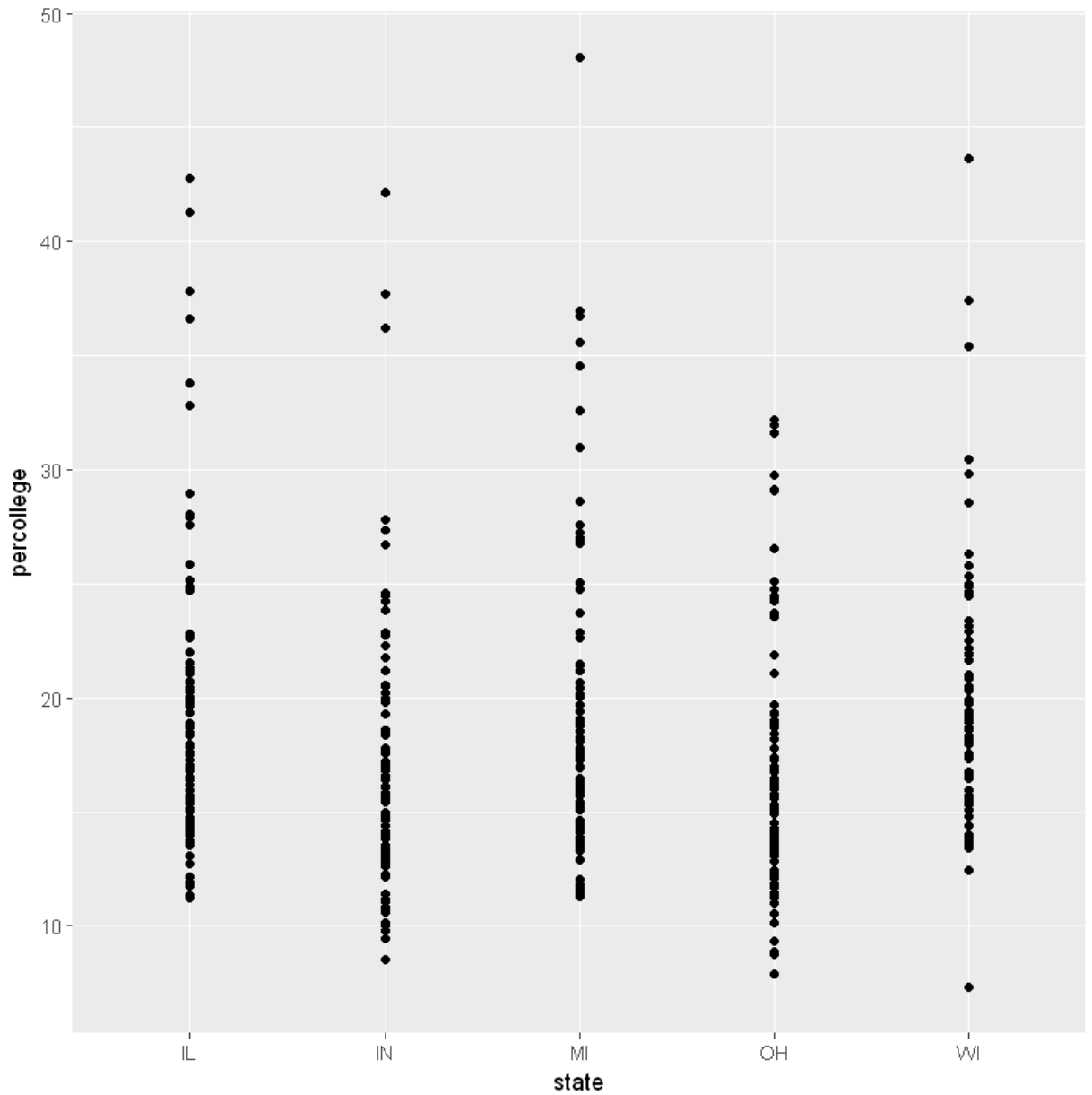
```
ggplot(midwest, aes(state, percollege)) + geom_jitter()
```

```
ggplot(midwest, aes(state, percollege)) + geom_point()
```

```
In [14]: ggplot(midwest, aes(state, percollege))+  
  geom_jitter()  
  
ggplot(midwest, aes(state, percollege))+  
  geom_point()
```







**geom\_point로 그래프를 출력한 경우**

두 번째 그래프처럼 점들의 간격이 좁을 경우 가독성이 떨어진다.

**geom\_jitter의 경우,**

일정 범위 내에 흩뿌려 준다

In [ ]: