통계 그래픽스 HW1

182STG15 이시영(통계학전공)

Use **mpg** data

P1. Scatterplot of **hwy** vs **cyl**.



P2. Scatterplot of **class** vs **drv.**



이 Plot이 유용하지 않는 이유: X, Y에 사용된 두 변수 모두 Categorical Variable이기에 이 산점도를 이용하여 두 변수 사이의 상관관계를 파악할 수 없다.

P3.

|  |  |
| --- | --- |
| 기존 Plot | 수정 Plot |
|  |  |
|  |  |

Color 지정이 aes() 안에 들어가 있기에 그래프가 ‘blue’라는 변수를 가지는 data에 대하여 고유 color 값을 지정한다. 하지만 데이터 안에 ‘blue’라는 값은 없기에 의도하지 않은 산점도가 출력된다.

P4.

|  |  |
| --- | --- |
|  |  |
|  |  |

**(Categorical) ~ .** 는 의 값에 따라서 행 방향으로 Plot을 그린다.

**. ~ (Categorical)** 는 의 값에 따라서 열 방향으로 plot을 그린다.

P5. 두 Plot이 차이를 보이는 이유

|  |  |
| --- | --- |
|  |  |
|  |  |

두 산점도에는 차이가 없다. geom\_point나 geom\_line 모두 동일한 변수를 사용하기에 ggplot에 한번에 입력해도 되고 각각의 명령문 안에 넣어도 된다.

P6.

|  |  |
| --- | --- |
|  |  |
| ggplot(mpg,aes(x= displ, y= hwy)) + geom\_point() +  geom\_smooth(aes(class = drv), se = F) | ggplot(mpg,aes(x= displ, y= hwy)) + geom\_point(aes(fill = drv), shape = 21, color = 'white', stroke = 1.5) |