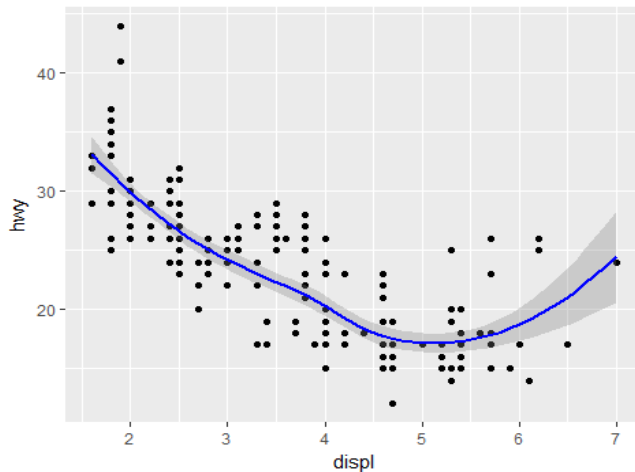


## 2020년 1학기 통계 그래픽스 중간고사 (4월 27일 14:00 ~ 5월 4일 00:00)

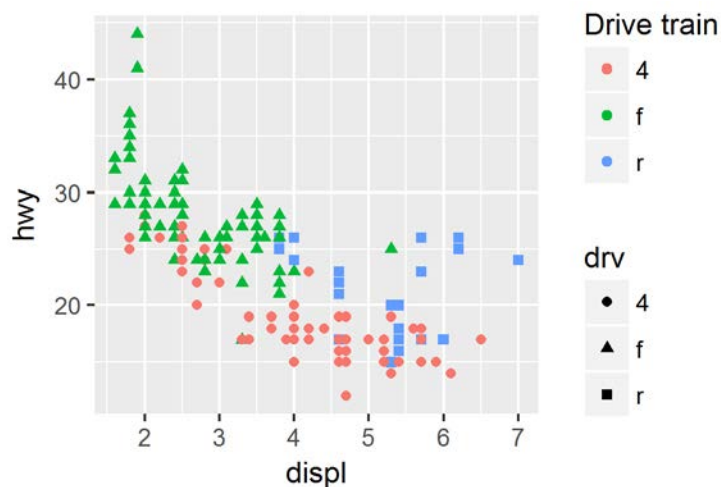
1. 아래의 code로 생성되는 자료를 이용하여 geom\_smooth를 이용하지 않고 아래의 그림과 같이 geom\_smooth의 효과를 갖는 그림을 그리시오.



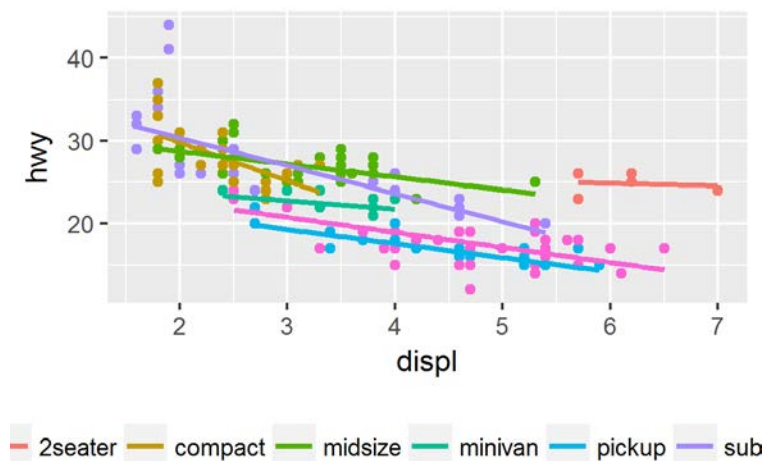
```
mod <- loess(hwy ~ displ, data = mpg)
smoothed <- data.frame(displ = seq(1.6, 7, length = 50))
pred <- predict(mod, newdata = smoothed, se = TRUE)
smoothed$hwy <- pred$fit
smoothed$hwy_lwr <- pred$fit - 1.96 * pred$se.fit
smoothed$hwy_upr <- pred$fit + 1.96 * pred$se.fit
```

2. 아래의 코드에 대한 그림이다. 잘못된 점을 찾고, 올바른 그림을 그리시오.

```
ggplot(mpg, aes(displ, hwy)) +
  geom_point(aes(colour = drv, shape = drv)) +
  scale_colour_discrete("Drive train")
```



3. mpg자료를 이용하여 아래의 그림을 그리는 code를 작성하시오.



4. Midterm-EDAdat1-2020.csv 파일은 2016년 사망자 자료를 성별, 나이별(ageG), 사망원인 별로 정리한 자료이다. 사망원인에 대한 code는 Midterm-D56.xlsx에 있다. 이 자료를 이용하여 자료의 특성을 살펴보고 자료의 특성을 가장 잘 나타낸다고 생각하는 그림 한 장을 그리고 그 그림을 선택하게 된 이유를 명시하시오. (자료를 살펴보는 과정도 나타낼 것)

5. Midterm-EDAdat2-2020.csv 파일에는 16개의 변수가 있다. 이 자료에 숨어있는 정보를 찾아 내시오. (숨어있는 정보를 찾아가는 과정에 사용한 code를 모두 제시해야합니다)

6. COVID-19-2020-04-20.csv 파일은 2019년 12월 31일부터 2020년 4월 20일까지의 전세계의 코로나 확진자와 사망자에 대한 자료이다. 이 자료를 이용하여 자료의 특성을 살펴보고 각 나라마다의 코로나 확진 추이의 차이를 살펴보시오.