

# Lab 2: Fuel

Jake James

January 26, 2021

## Introduction

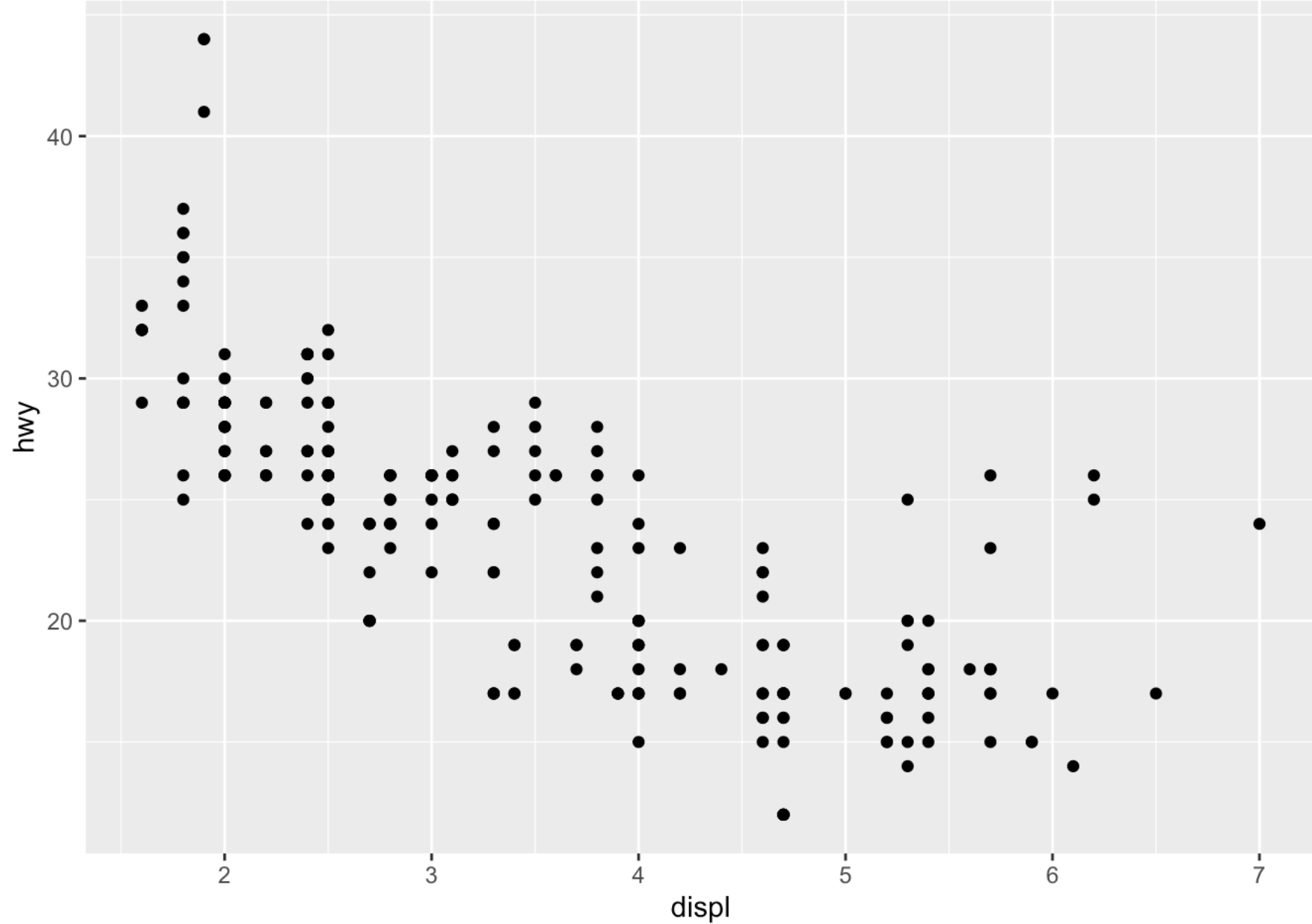
This lab assignment is based on `mpg` data from `ggplot2` package. This dataset contains a subset of the fuel economy data that the EPA makes available on <http://fuelconomy.gov>. Each row of the data frame represents a different car model and. There are 234 rows and 11 variables in the dataset.

## Exercises

### Part 1: Basic Plot

- a. Use scatterplot to visualize the relationship between `displ` (engine displacement) and `hwy` (highway miles per gallon) from `mpg` with `displ` on x-axis and `hwy` on y-axis.

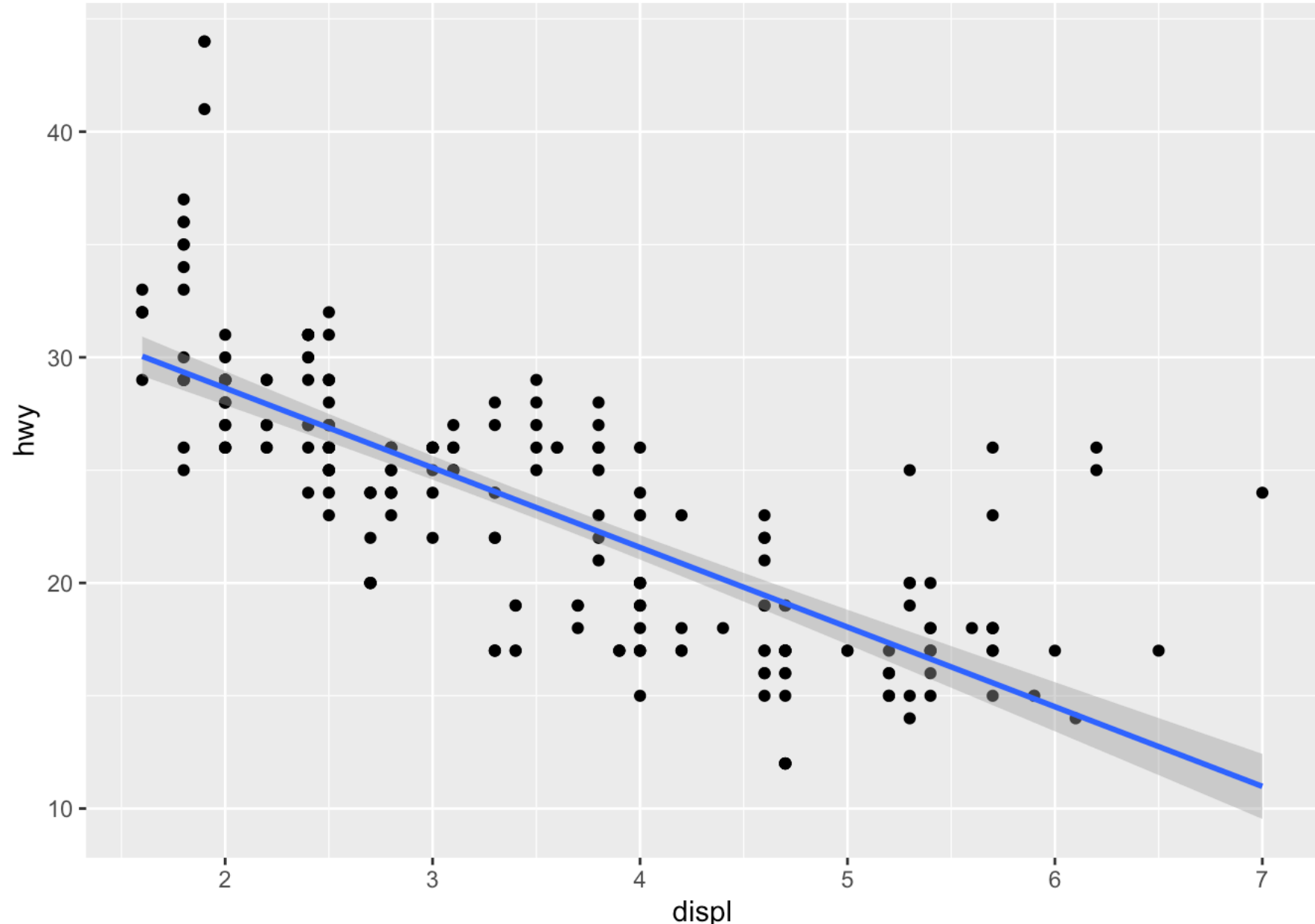
```
ggplot(data = mpg) +  
  geom_point(aes(x = displ, y = hwy))
```



- b. Add a smooth curve to the previous scatterplot with linear regression (`lm`) as smoothing method.

```
ggplot(data = mpg) +  
  geom_point(aes(x = displ, y = hwy)) +  
  geom_smooth(aes(x = displ, y = hwy),method = lm)
```

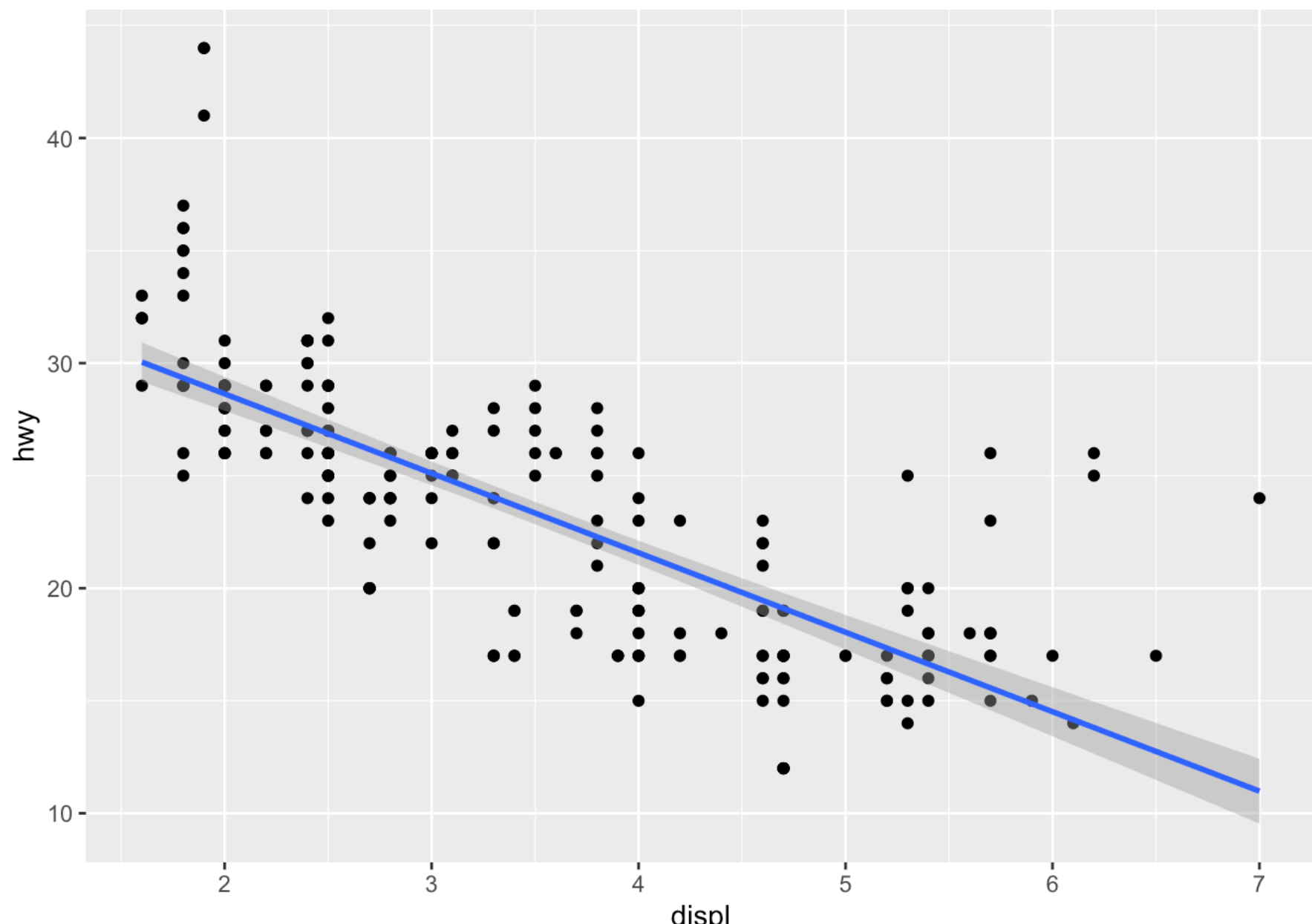
```
## `geom_smooth()` using formula 'y ~ x'
```



- c. Generate the same plot as in (b) but specify the aesthetic mappings in `ggplot()` function. Is there any difference between plot (c) and plot (b)?

```
ggplot(data = mpg, mapping = aes(x = displ, y = hwy)) +  
  geom_point(aes(x = displ, y = hwy)) +  
  geom_smooth(aes(x = displ, y = hwy),method = lm)
```

```
## `geom_smooth()` using formula 'y ~ x'
```



### No, it is the same graph.

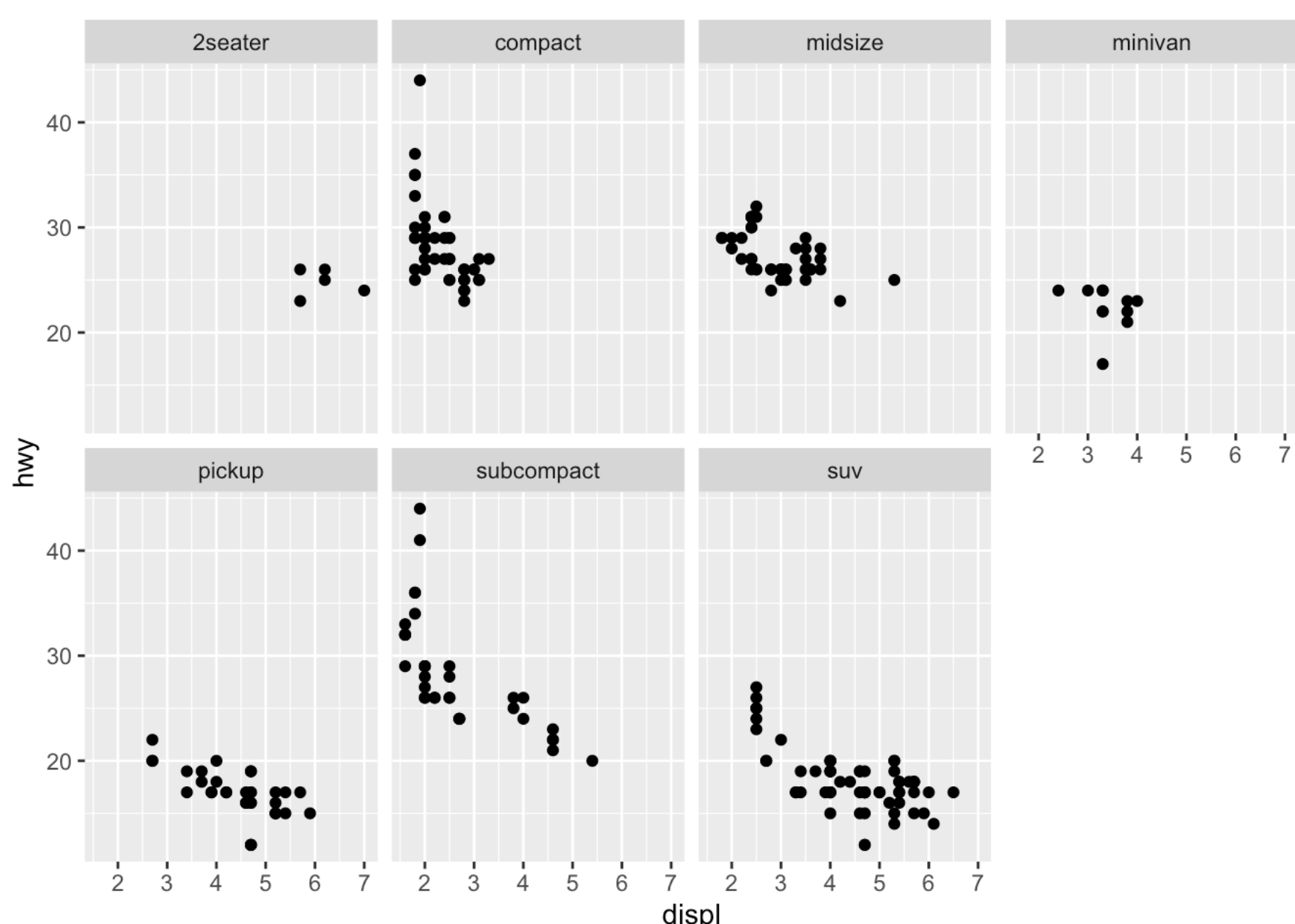
- d. Generate the same plot as in (b) but the color of scatterplot points are controlled by `class` (type of car) in `mpg`.

```
ggplot(data = mpg) +  
  geom_point(aes(x = displ, y = hwy, color = as.factor(class))) +  
  geom_smooth(aes(x = displ, y = hwy),method = lm)
```

### Part 2: Advacned Plot

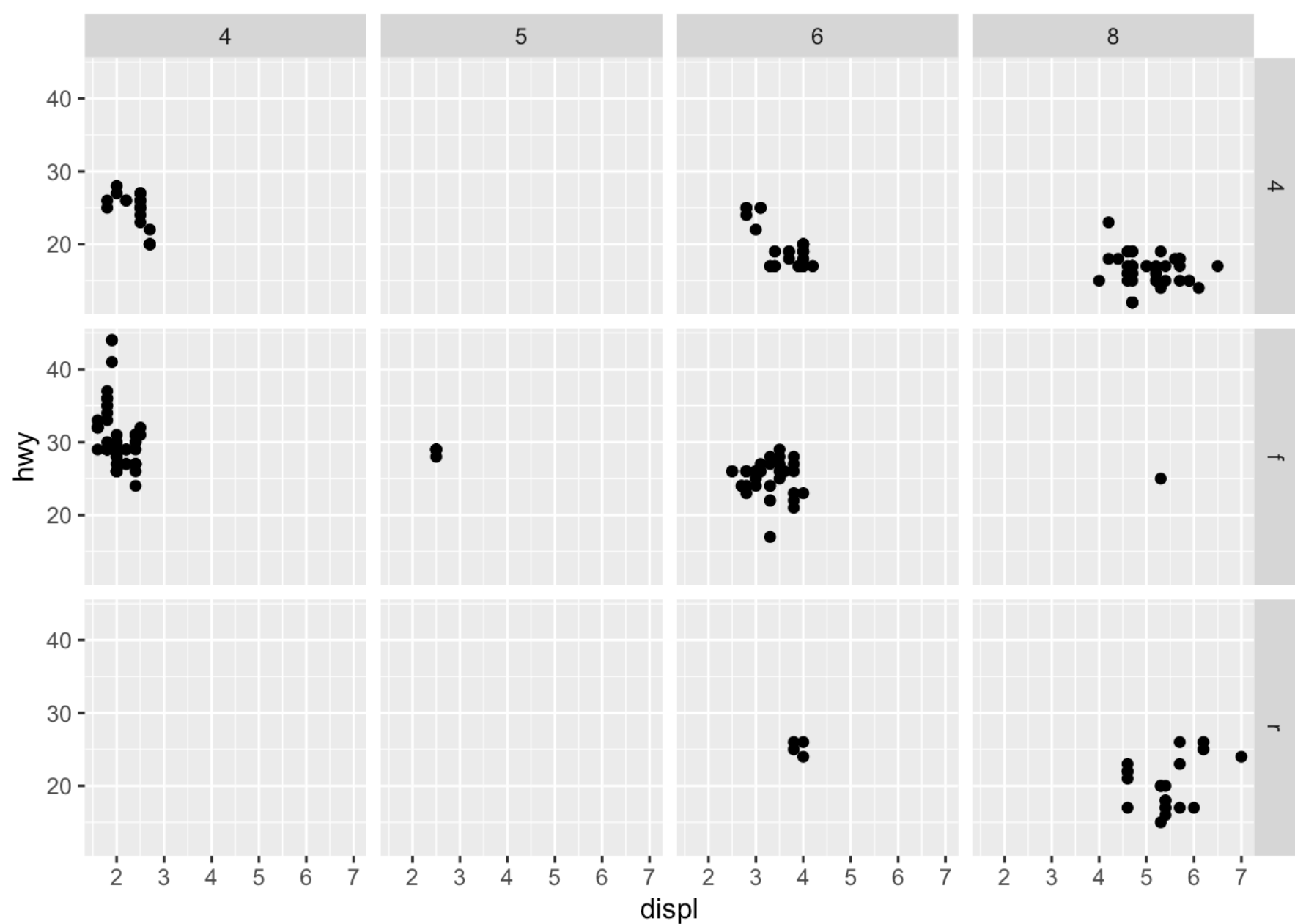
- e. Use `facet_wrap` to visualize the relationship between `displ` and `hwy` based on `class`.

```
ggplot(data = mpg) +  
  geom_point(aes(x = displ, y = hwy)) +  
  facet_wrap(class~., nrow=2)
```



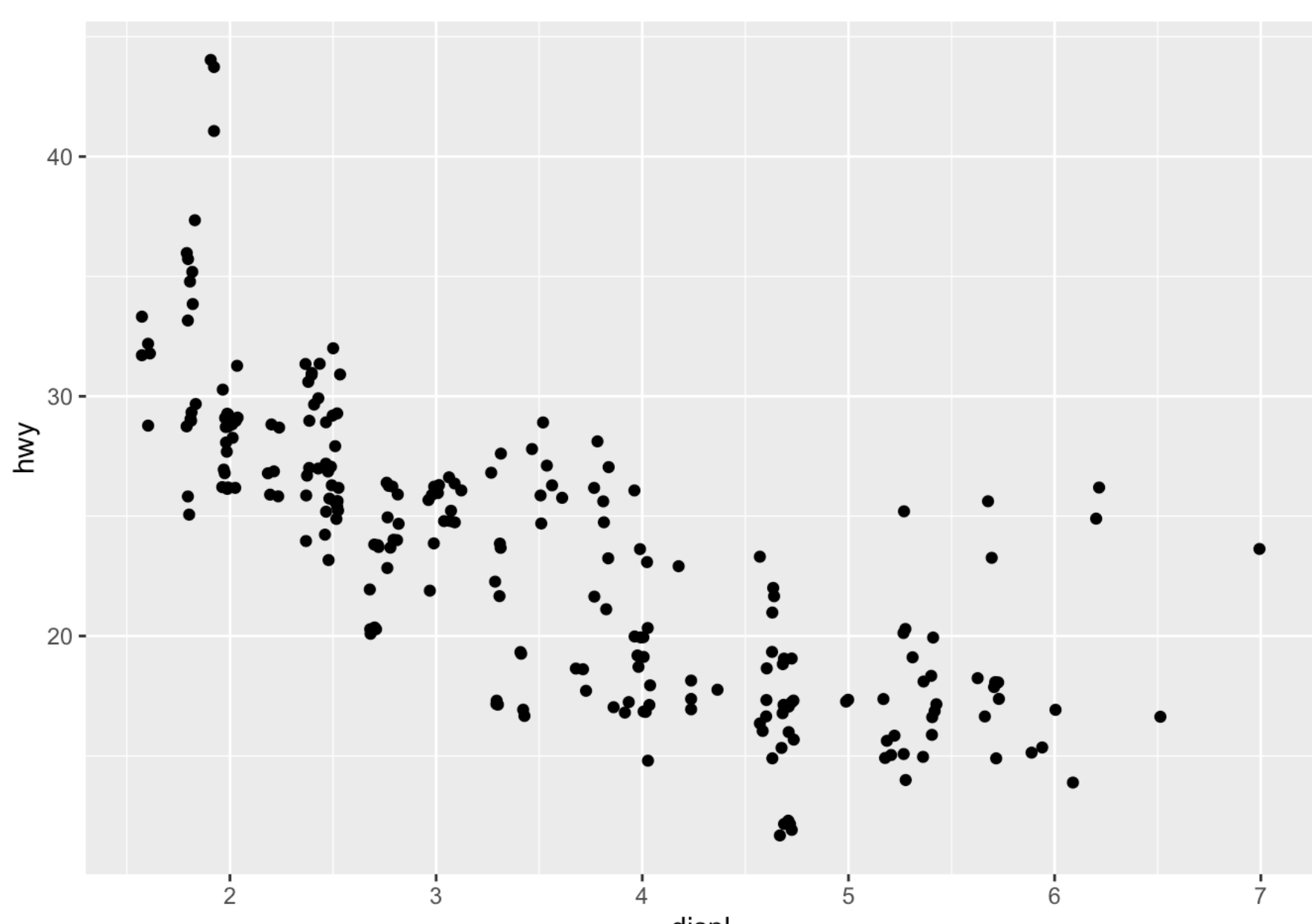
- f. Use `facet_grid` to visualize the relationship between `displ` and `hwy` based on the relationship between `drv` (type of drive train) and `cyl` (number of cylinders).

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



- g. Compare the following plot with the plot in (a), what is the difference?

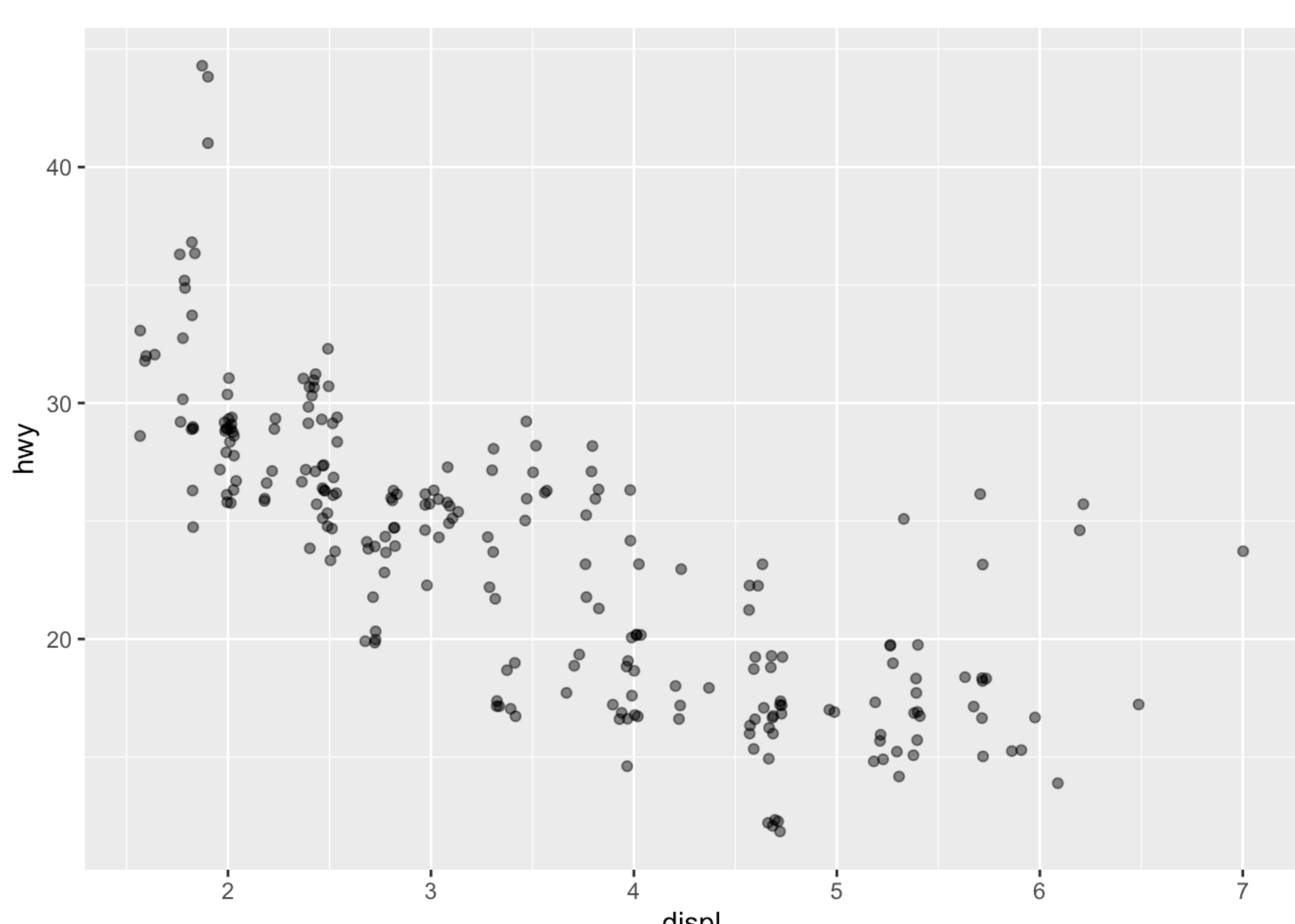
```
ggplot(data = mpg) +  
  geom_point(aes(x = displ, y = hwy), position = "jitter")
```



### The data is more jumbled so that you can see each point individually and they don't overlap.

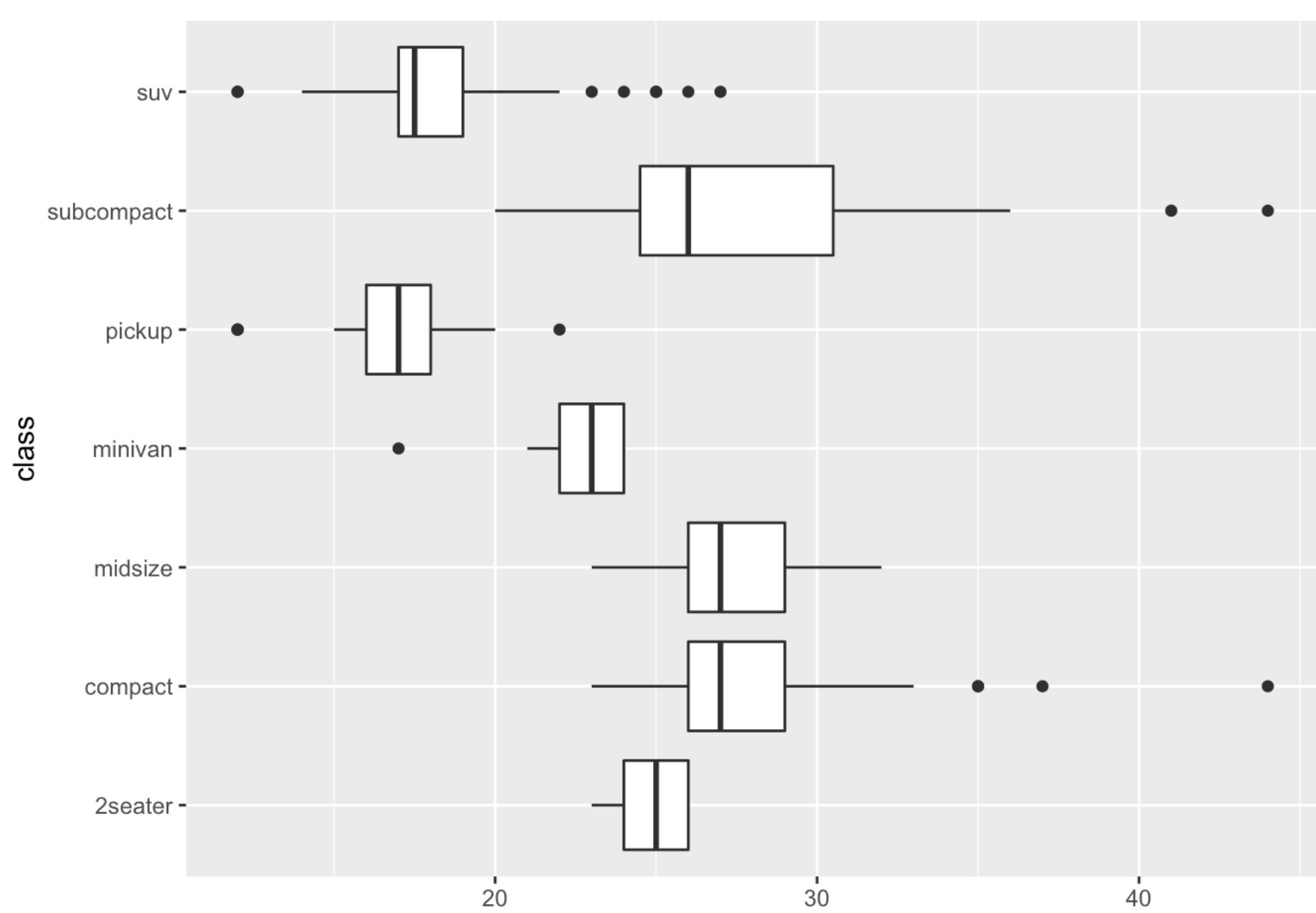
- h. `geom_jitter` is a convenient shortcut for `geom_point(position = "jitter")`. Generate the plot in (g) and set the points to be transparent with scale `.5`.

```
ggplot(mpg, aes(displ, hwy)) +  
  geom_jitter(alpha = 0.5)
```



- i. Generate boxplot of `hwy` based on `class`.

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



Flip the coordinates of the boxplot

with `coord_flip()`.

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

