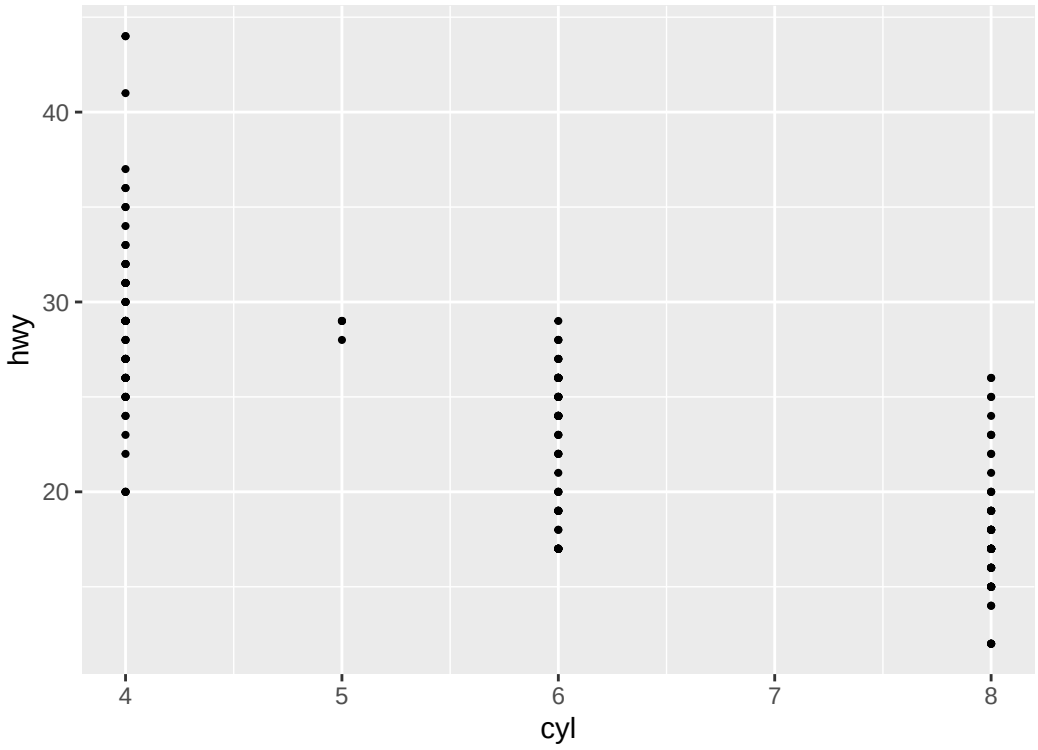
Strip Charts, Boxplots, Ridgelines

Problem 1: We will work with the `mpg` dataset provided by **ggplot2**. See here for details: https://ggplot2.tidyverse.org/reference/mpg.html

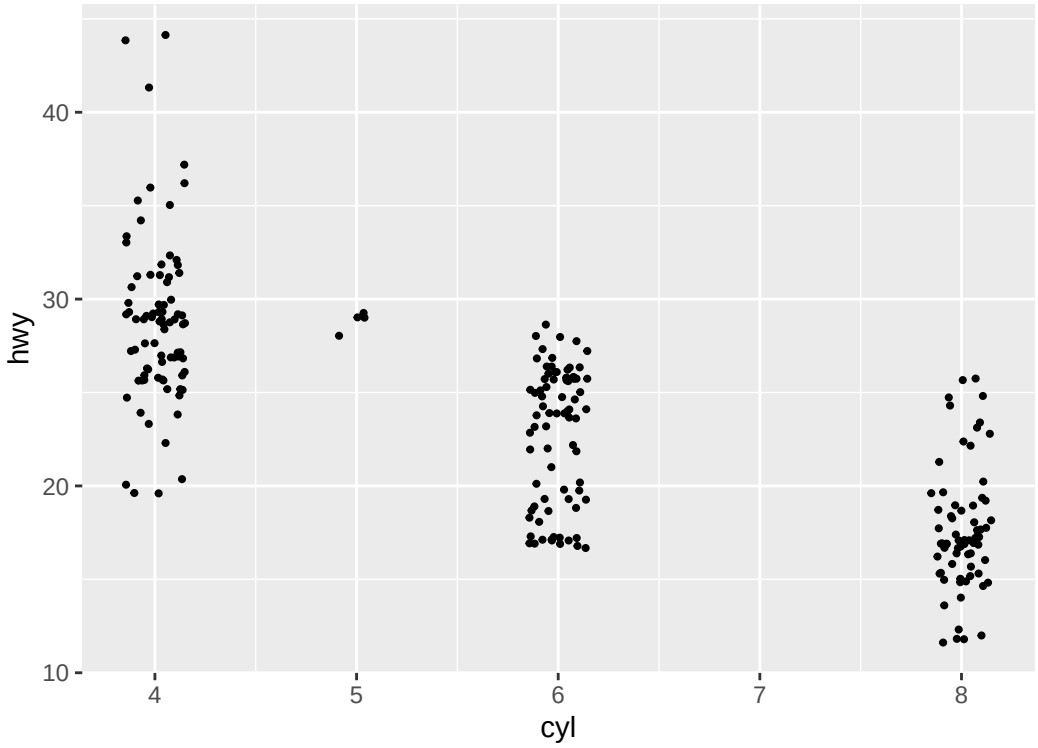
Make two different strip charts of highway fuel economy versus number of cylinders, the first one without horizontal jitter and second one with horizontal jitter. Explain in 1-2 sentences why the plot without jitter is highly misleading.

Hint: Make sure you do not accidentally apply vertical jitter. This is a common mistake many people make.

```
ggplot(mpg, aes(x = cyl, y = hwy)) +
  geom_point(size = 0.75)
```




```
ggplot(mpg, aes(x = cyl, y = hwy)) +
  geom_point(
    size = 0.75,
    position = position_jitter(width = 0.15)
  )
```

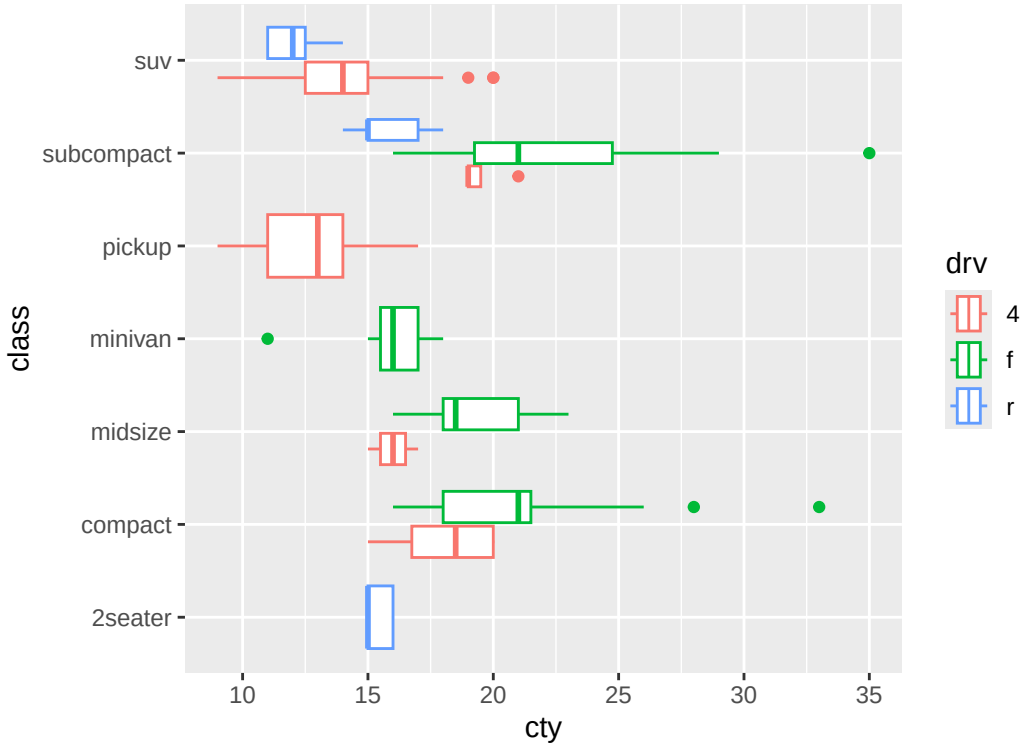


Explanation: *Without jittering, the data points are overlapping one another and it is not showing a clear picture of how many highway fuel economy data points are available for each type of cylinder. For example, there are 4 data points for cars with 5 cylinders, however, without jittering it only shows 2 data points.*

Problem 2: For this problem, we will continue working with the `mpg` dataset. Visualize the distribution of each car's city fuel economy by class and type of drive train with (i) boxplots and (ii) ridgelines. Make one plot per geom and do not use faceting. In both cases, put city mpg on the x axis and class on the y axis. Use color to indicate the car's drive train.

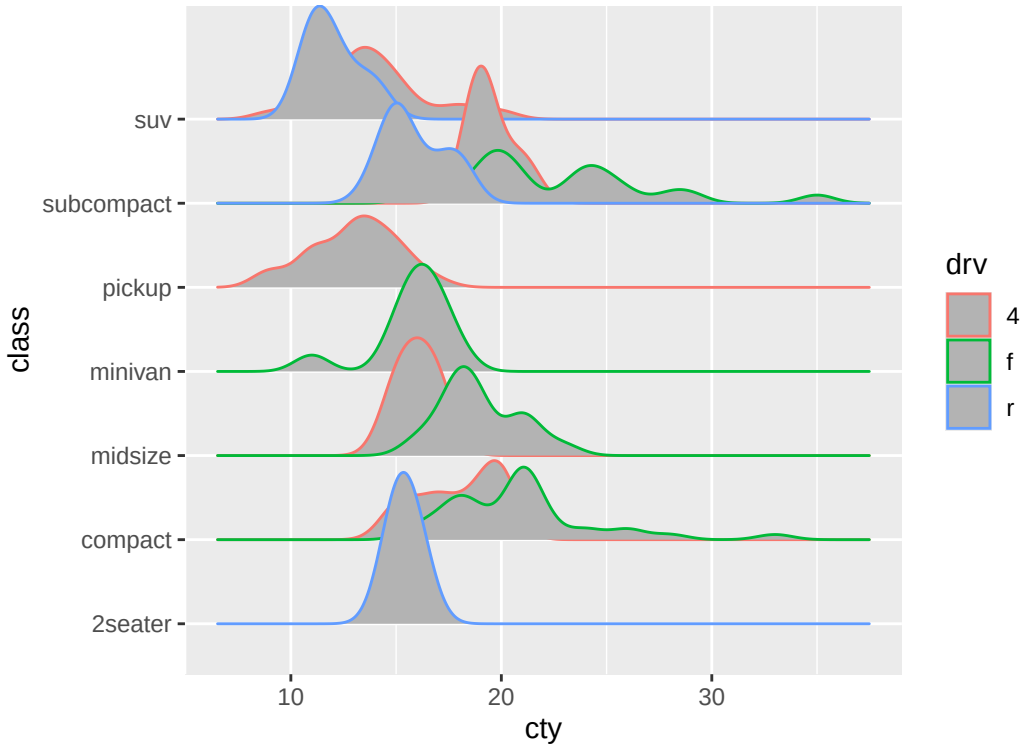
The boxplot ggplot generates will have a problem. Explain what the problem is. (You do not have to solve it.)

```
ggplot(mpg, aes(x = cty, y = class, color = drv)) +
  geom_boxplot()
```




```
ggplot(mpg, aes(x = cty, y = class, color = drv)) +
  geom_density_ridges()
```

```
## Picking joint bandwidth of 0.828
```



Explanation: *The boxplots has discrete categorical data on y-axis and continuous data on x-axis and that resulted in generating the boxplots horizontally for each class. This does not help visualizing the boxplots to understand the data distribution. Boxplots are already a crude way to visualize distribution and by generating them horizontally makes it harder to understand the data. This issue can be solved by putting class on the x-axis and city mpg on the y-axis.*