Lecture 2 — Visualization I

```
library(ggplot2)
```

```
head(mpg)
```

```
## # A tibble: 6 × 11
##   manufacturer model displ  year   cyl      trans   drv   cty   hwy    fl
##          <chr> <chr> <dbl> <int> <int>      <chr> <chr> <int> <int> <chr>
## 1         audi    a4   1.8  1999     4   auto(l5)     f    18    29     p
## 2         audi    a4   1.8  1999     4 manual(m5)     f    21    29     p
## 3         audi    a4   2.0  2008     4 manual(m6)     f    20    31     p
## 4         audi    a4   2.0  2008     4   auto(av)     f    21    30     p
## 5         audi    a4   2.8  1999     6   auto(l5)     f    16    26     p
## 6         audi    a4   2.8  1999     6 manual(m5)     f    18    26     p
## # ... with 1 more variables: class <chr>
```

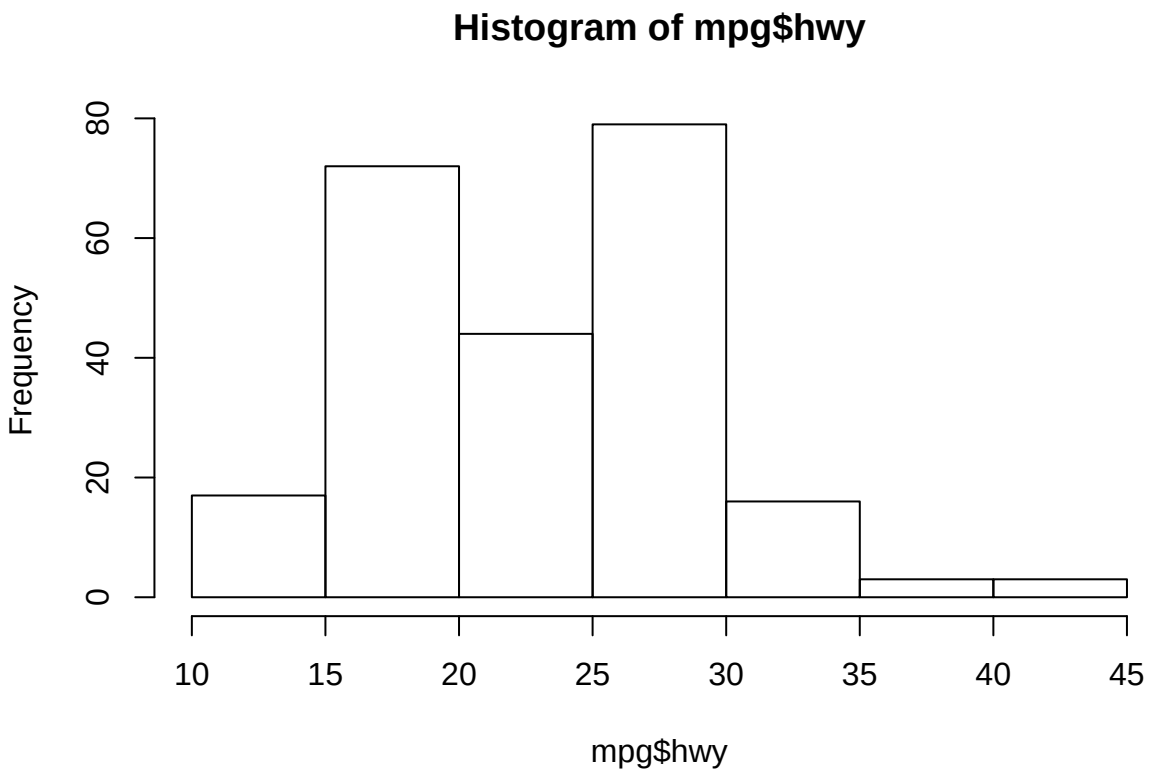
```
summary(mpg$hwy)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   12.00   18.00   24.00   23.44   27.00   44.00
```

```
summary(mpg$cty)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    9.00   14.00   17.00   16.86   19.00   35.00
```

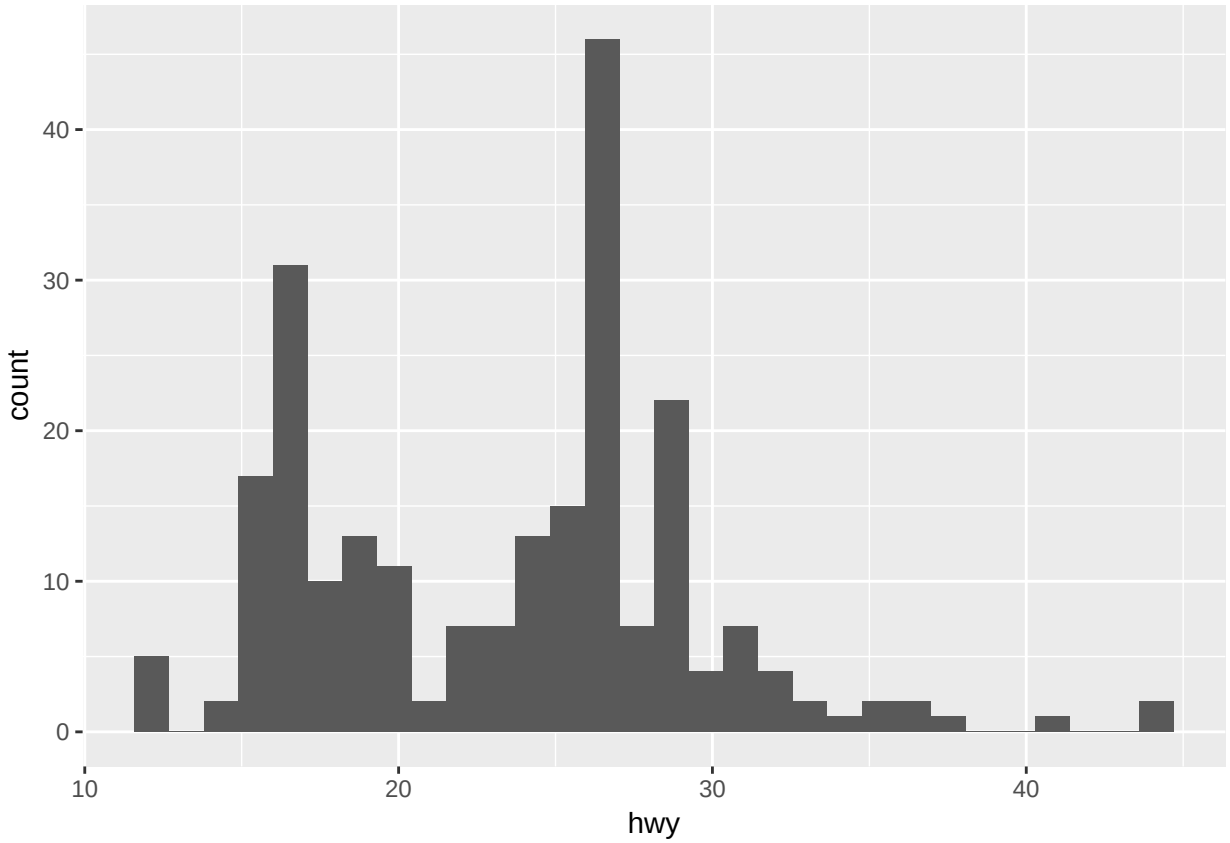
```
hist(mpg$hwy)
```



Let's use ggplot2:

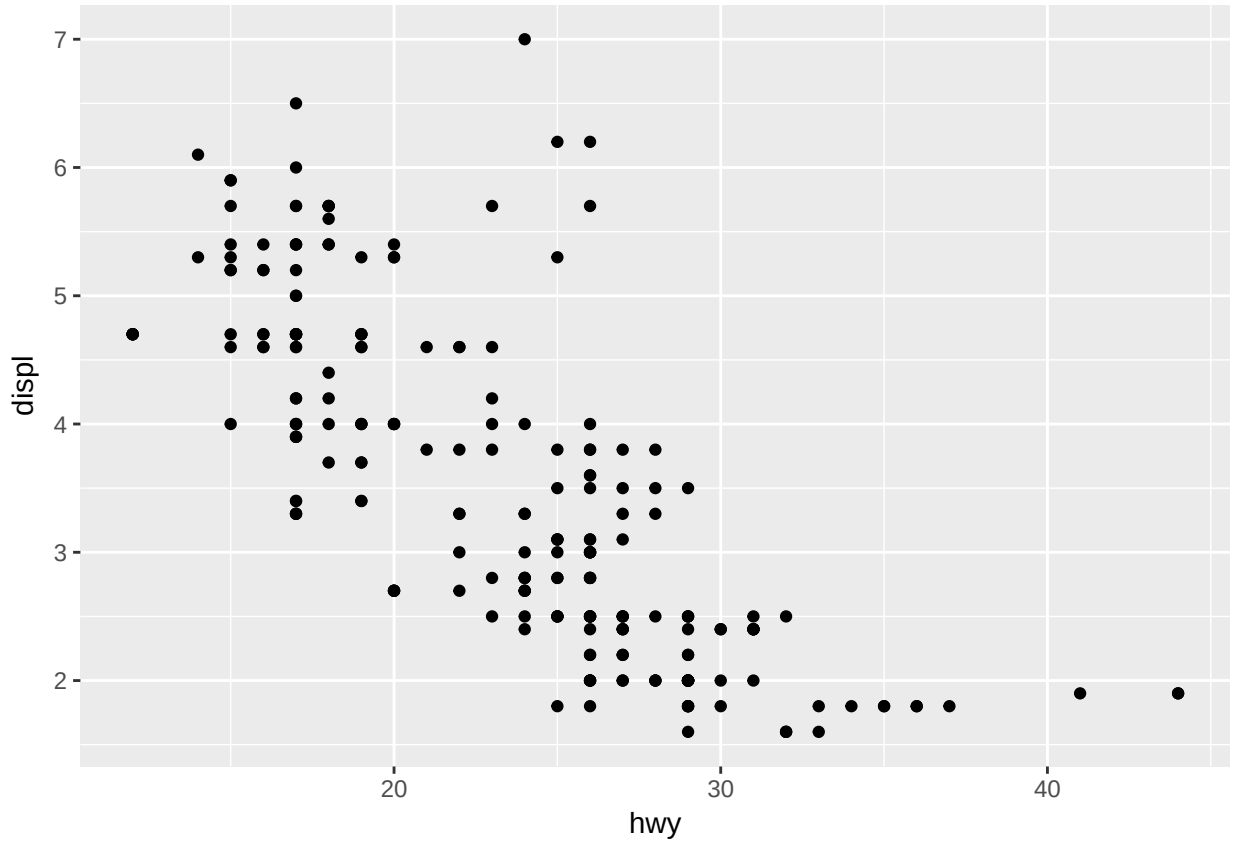
```
ggplot(mpg) + geom_histogram(aes(x=hwy))
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



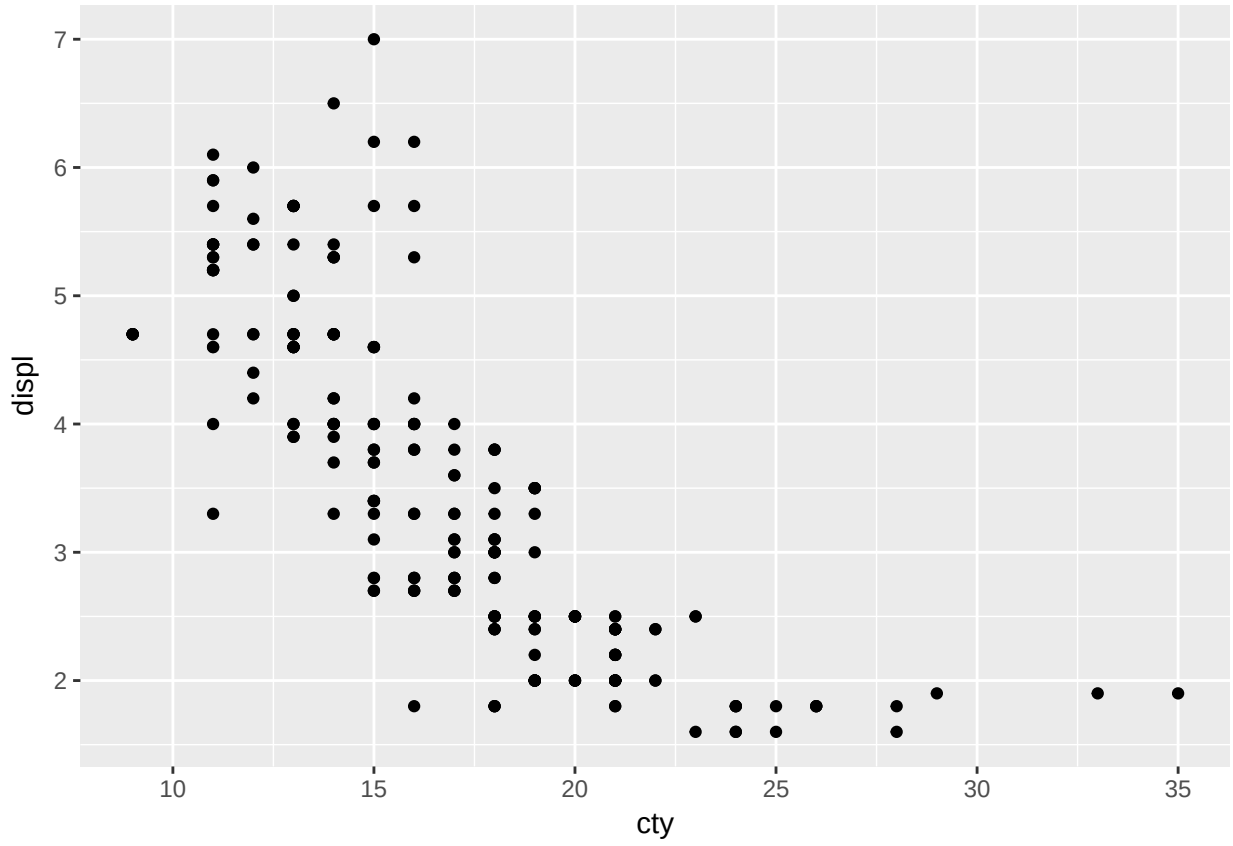
Let's look at `hwy` and `displ`:

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



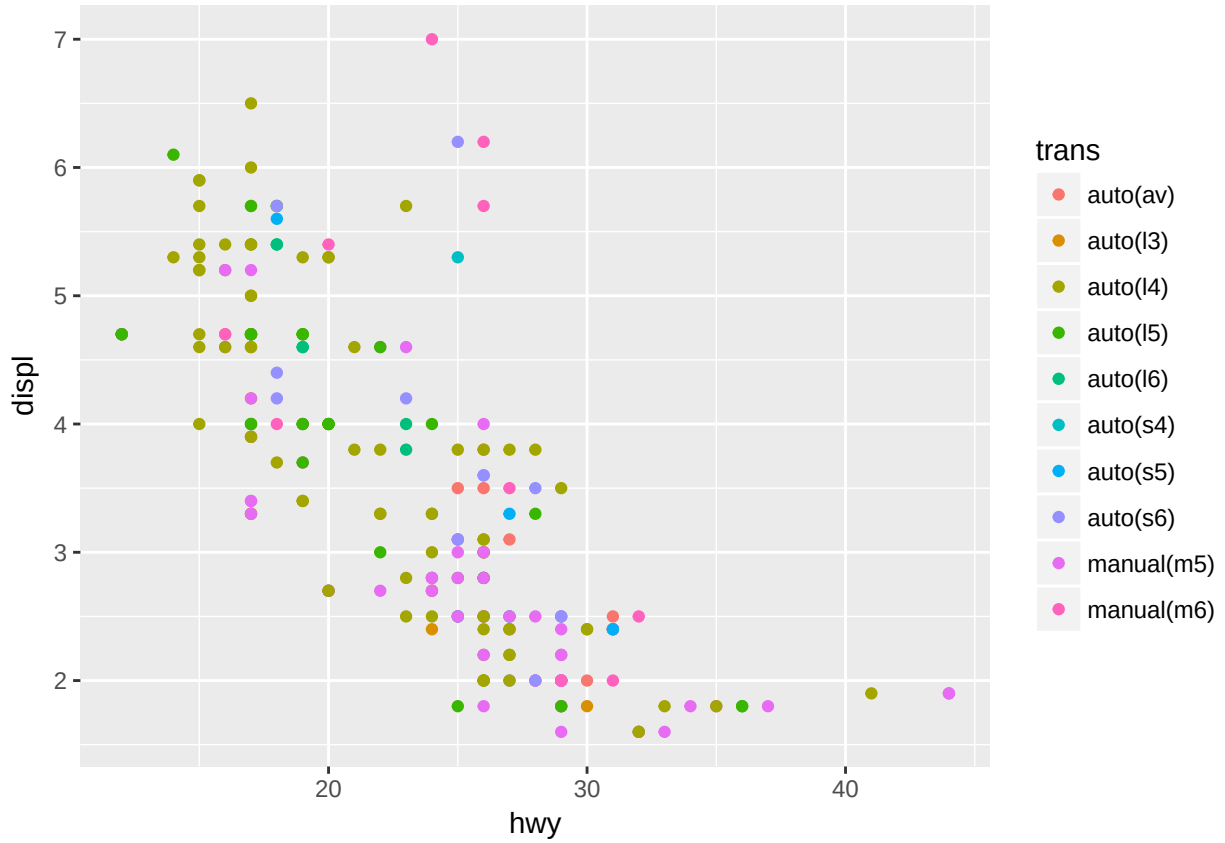
Does using `cty` make a difference? This is a different measure of fuel efficiency.

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

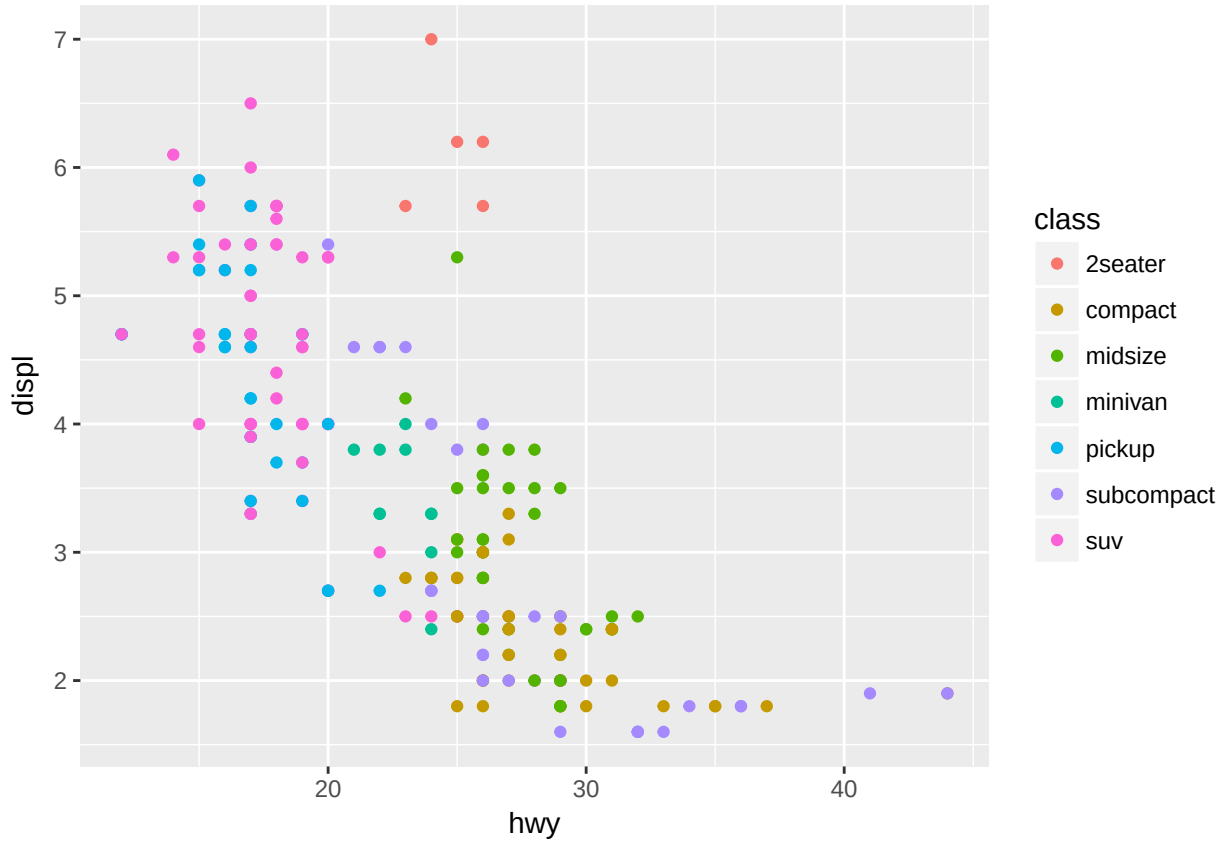


Nope, that doesn't make a difference

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

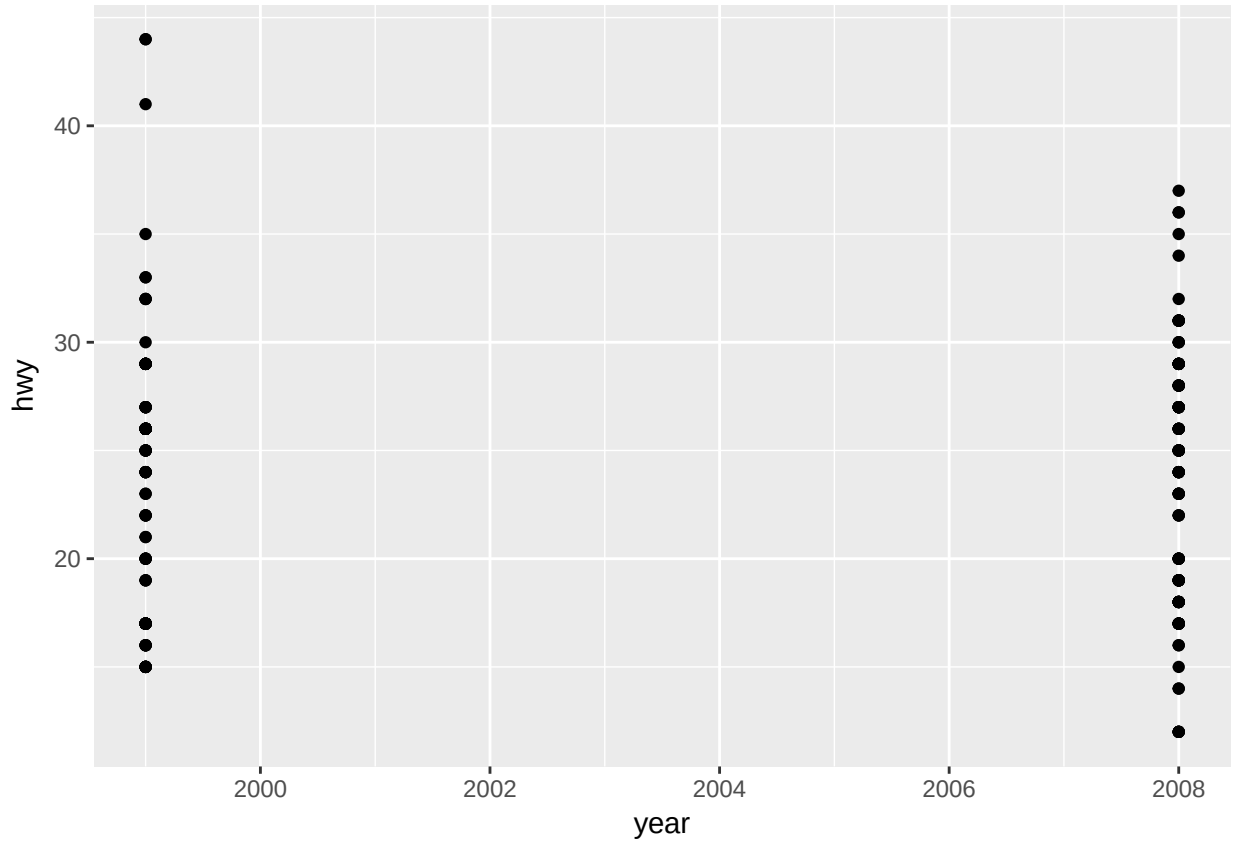



```
ggplot(mpg) + geom_point(aes(x=hwy, y=displ, color=class))
```



Fuel Efficiency By Year

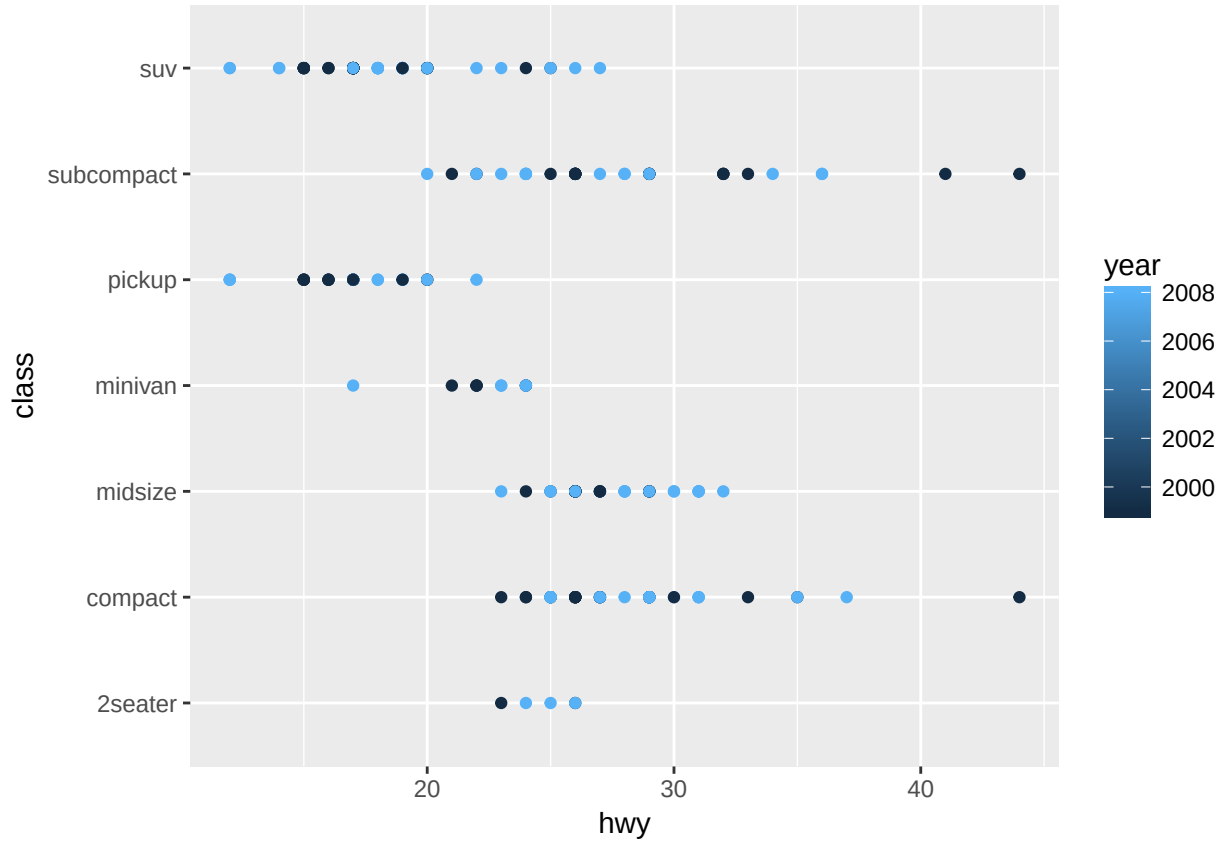
```
ggplot(mpg) + geom_point(aes(x=year, y=hwy))
```



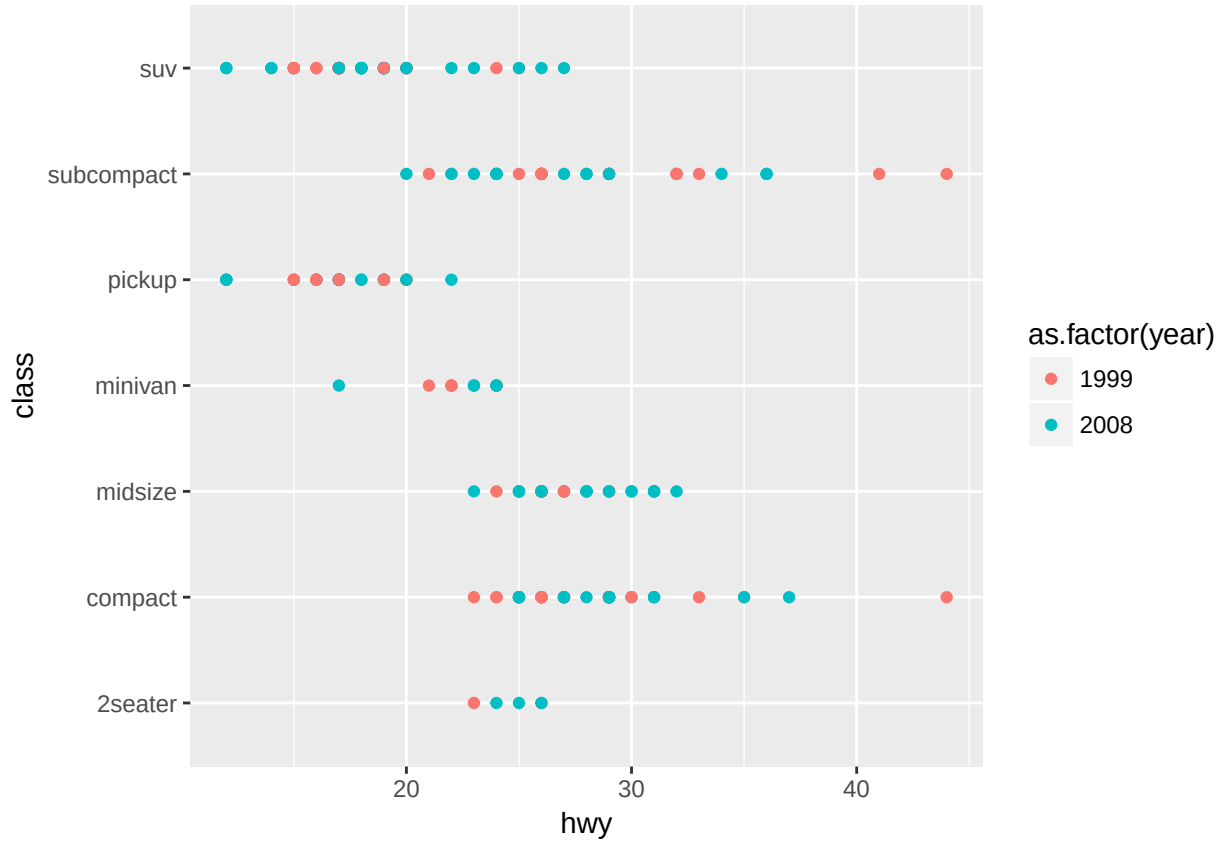

```
table(mpg$year)
```

```
##
## 1999 2008
##  117  117
```

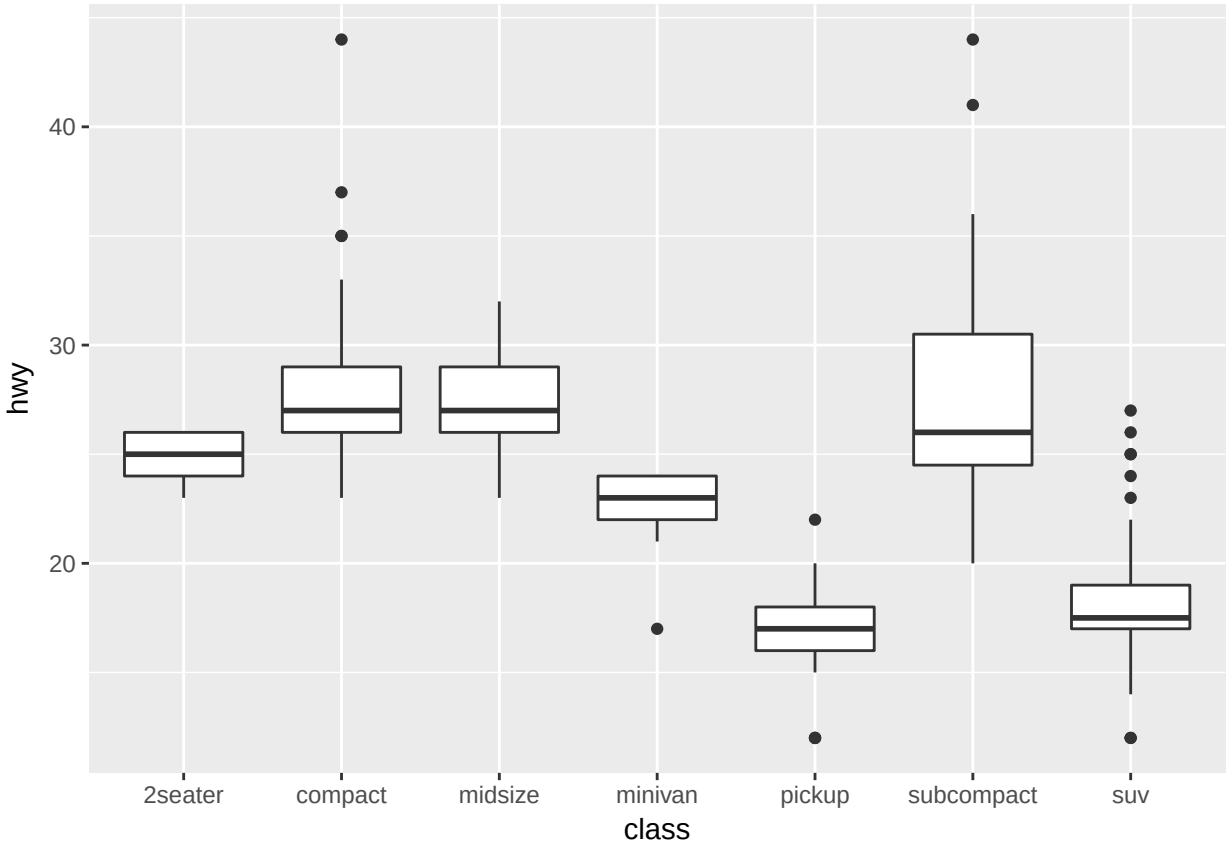
```
ggplot(mpg) + geom_point(aes(x=hwy, y=class, color=year))
```




```
ggplot(mpg) + geom_point(aes(x=hwy, y=class, color=as.factor(year)))
```

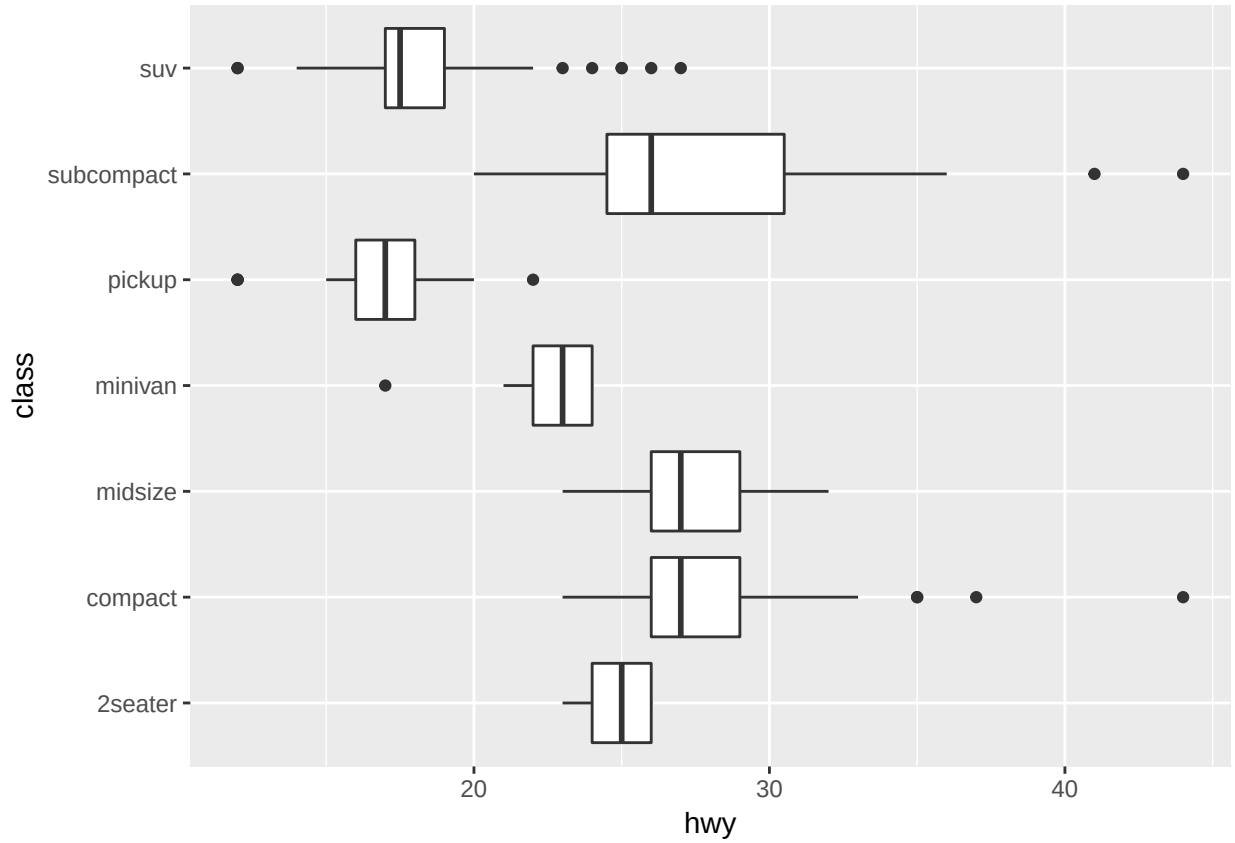



```
ggplot(mpg) + geom_boxplot(aes(y=hwy, x=class))
```

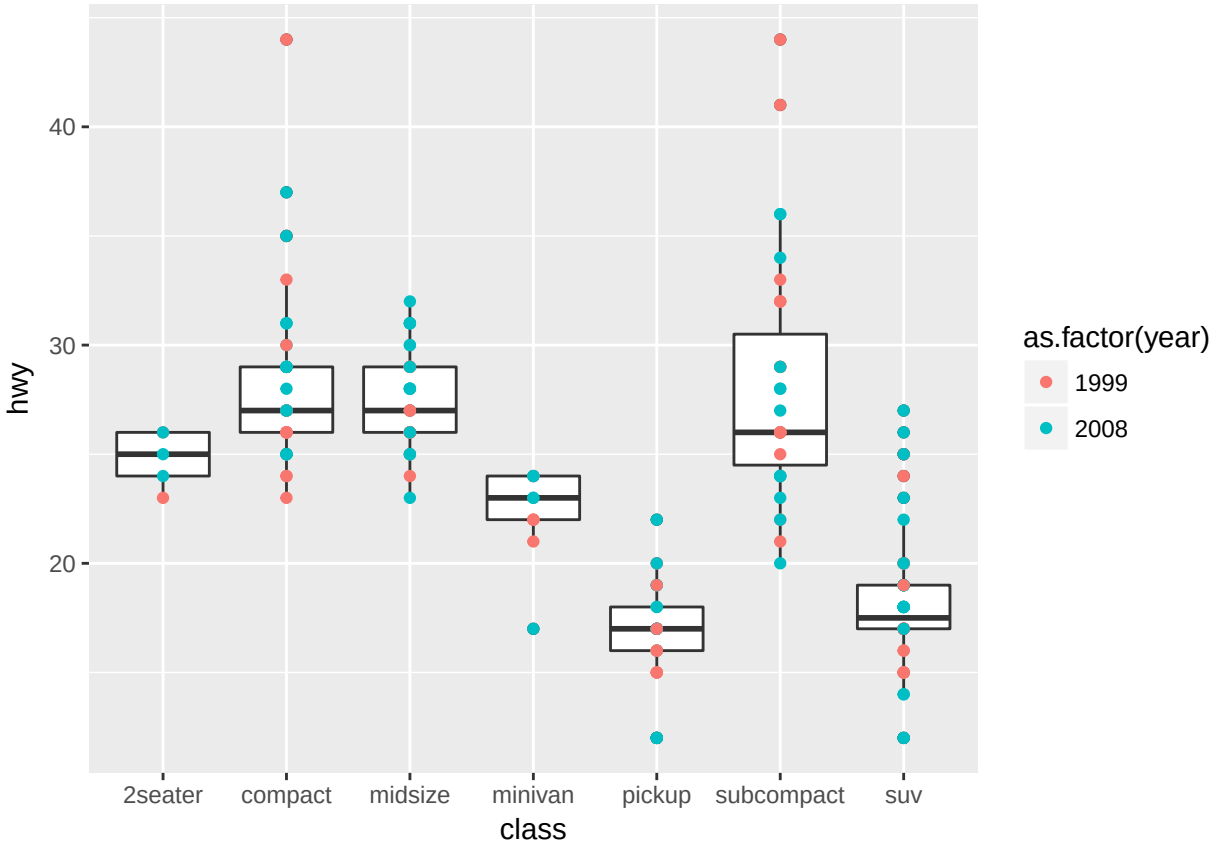


Note: If you wanted to flip the axes, use `coord_flip()`. For example:

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

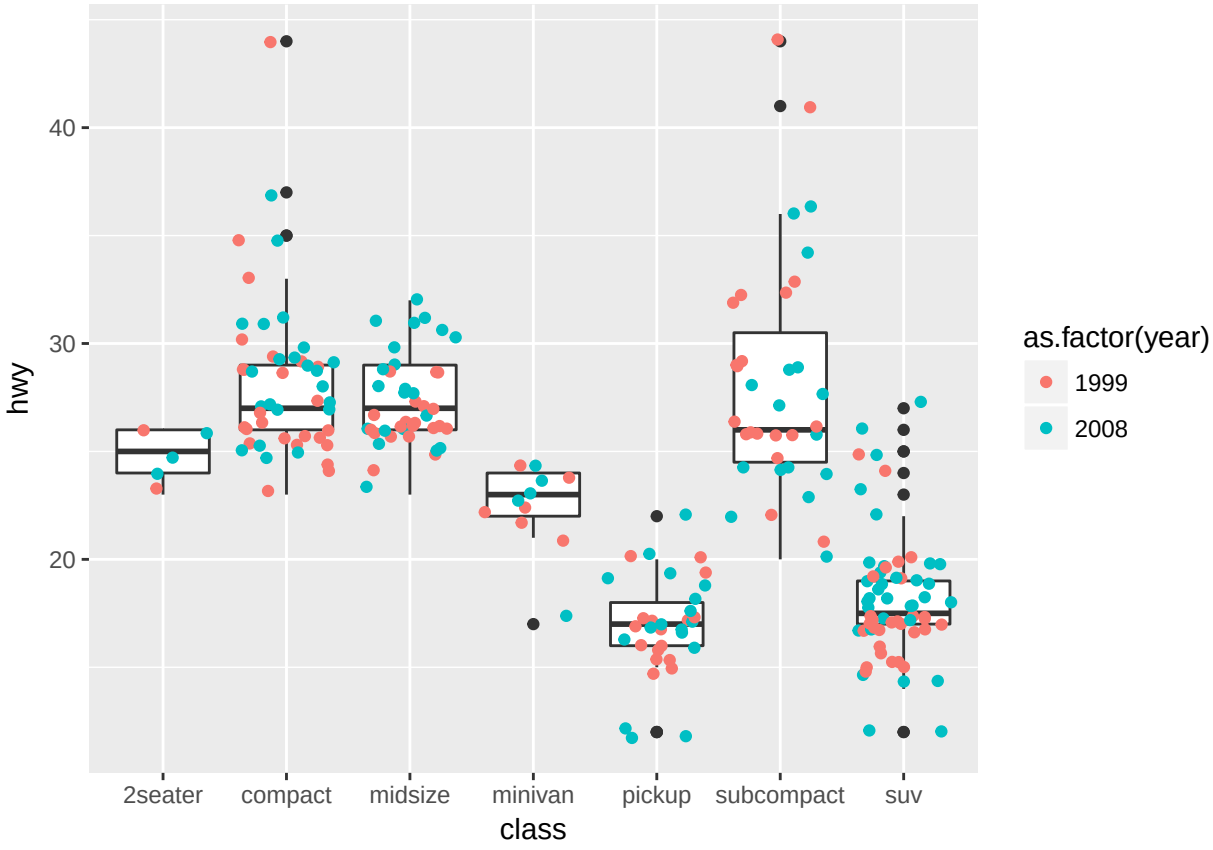



```
ggplot(mpg) + geom_boxplot(aes(y=hwy, x=class)) + geom_point(aes(x=class, y=hwy, color=as.factor(year))
```



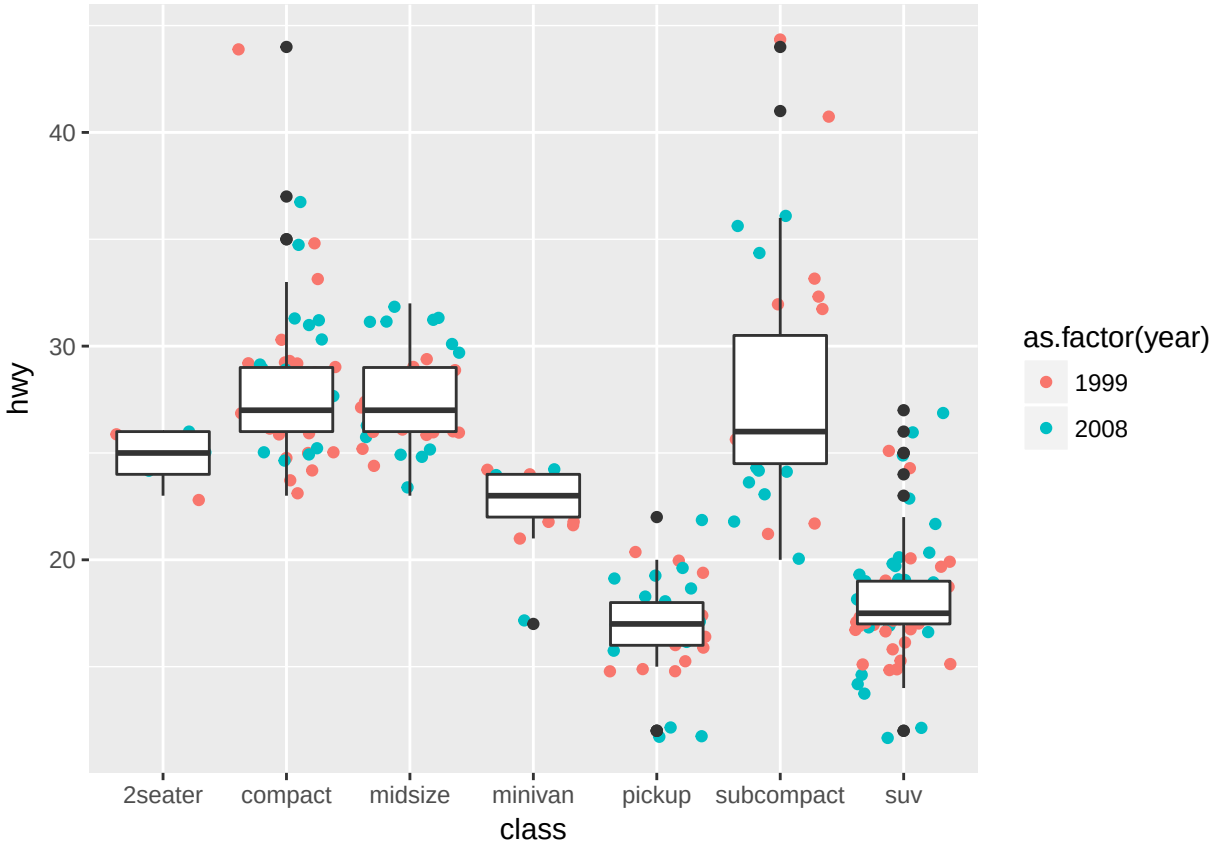
Due to overplotting: I am thinking about using some jitter to disperse the points.

```
ggplot(mpg) + geom_boxplot(aes(y=hwy, x=class)) + geom_jitter(aes(x=class, y=hwy, color=as.factor(year)
```



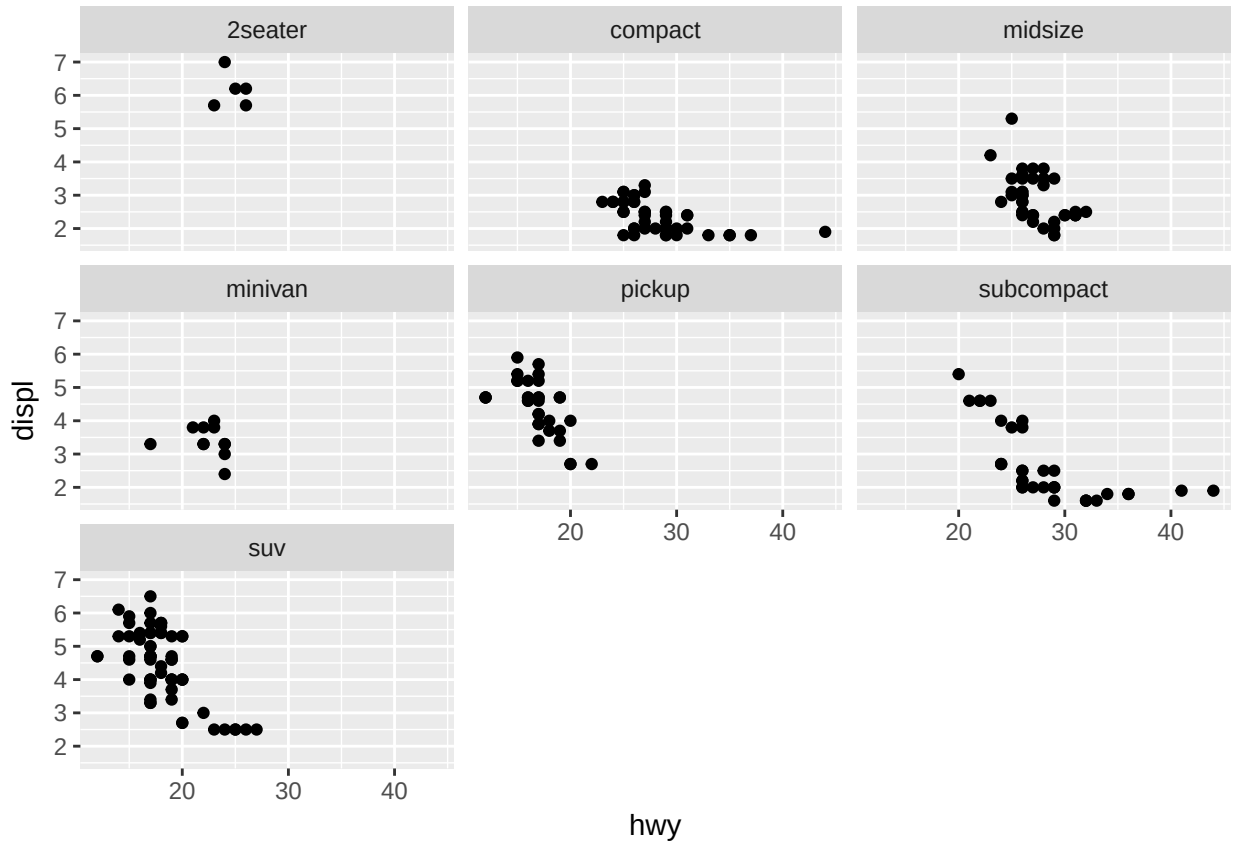
We could, if we liked this plot, customize it a bit.

```
ggplot(mpg) + geom_jitter(aes(x=class, y=hwy, color=as.factor(year))) + geom_boxplot(aes(y=hwy, x=class
```

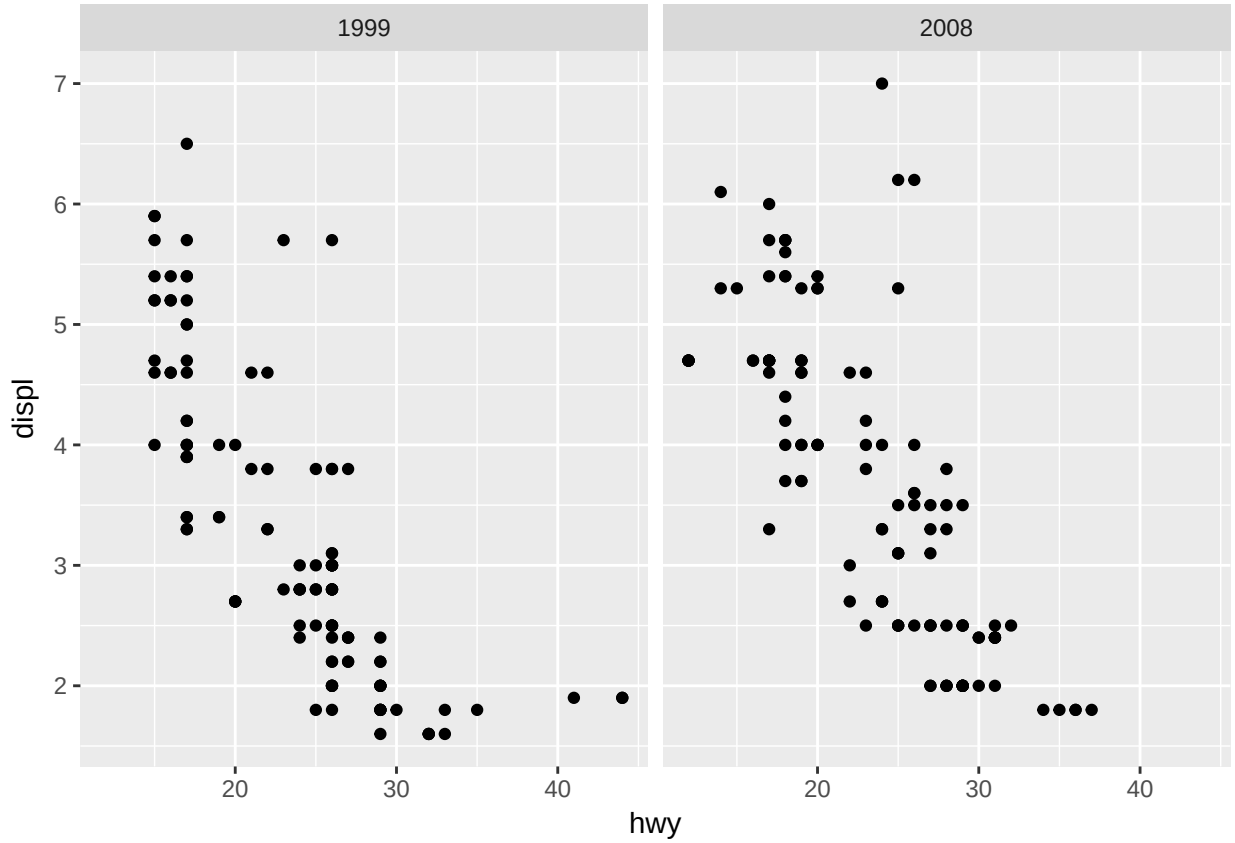


Facetting

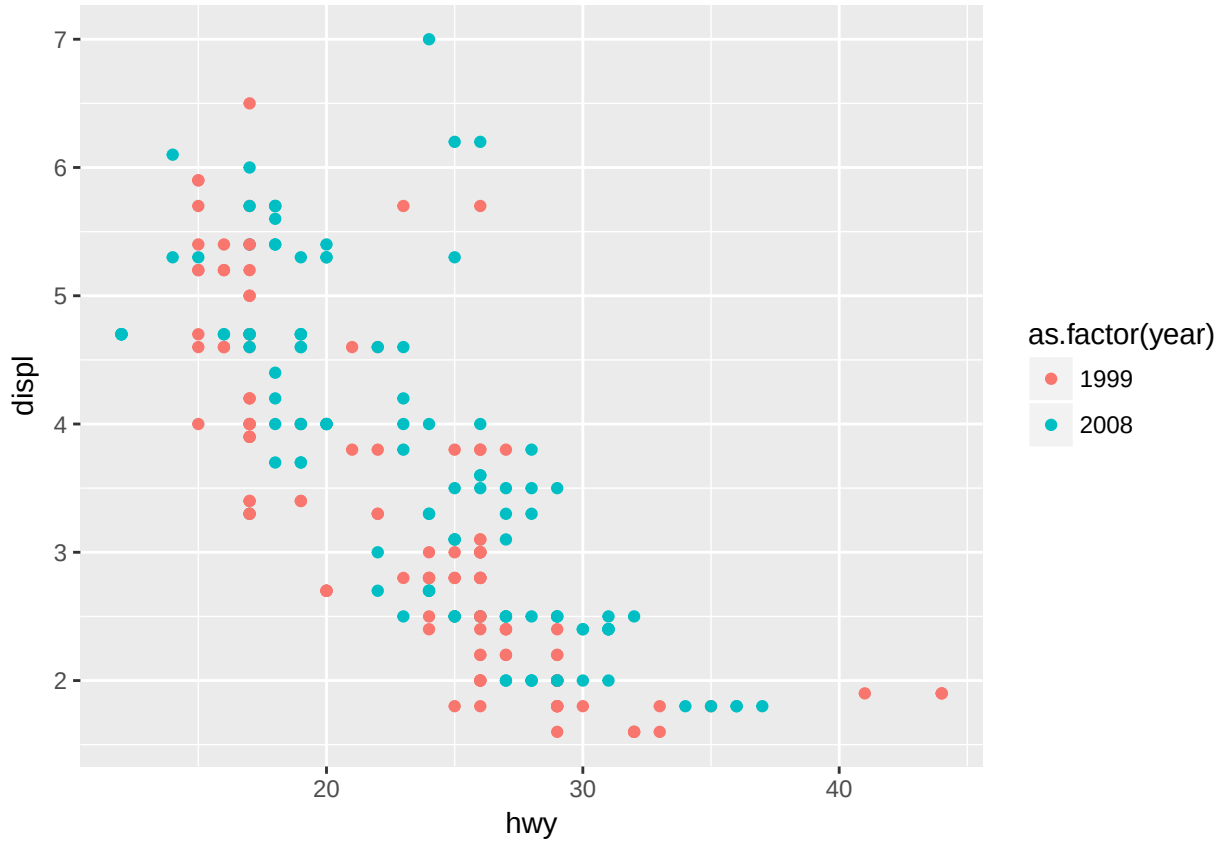
```
ggplot(mpg) + geom_point(aes(x=hwy, y=displ)) + facet_wrap(~ class)
```




```
ggplot(mpg) + geom_point(aes(x=hwy, y=displ)) + facet_wrap(~ year)
```



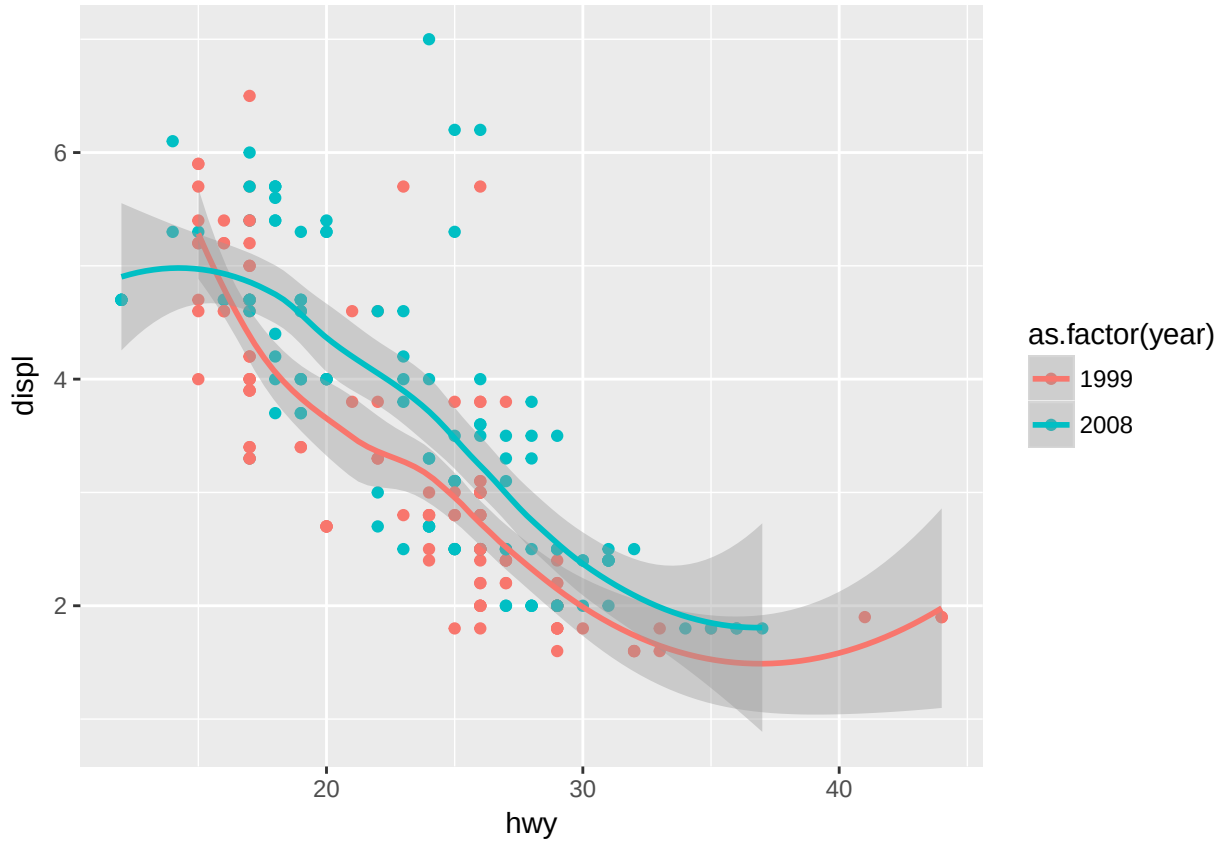

```
ggplot(mpg) + geom_point(aes(x=hwy, y=displ, color=as.factor(year)))
```



Some fun before next time:

```
ggplot(mpg) + geom_point(aes(x=hwy, y=displ, color=as.factor(year))) + geom_smooth(aes(x=hwy, y=displ,
```

```
## `geom_smooth()` using method = 'loess'
```



Exercises:

- Keep exploring the `mpg` data.
- Look up how to change the size of points.
- Look up how to change the shape of points.
- Make a violin plot of the hwy by class.