

ggplot2

```
library(tidyverse) # ggplot2 and friends
```

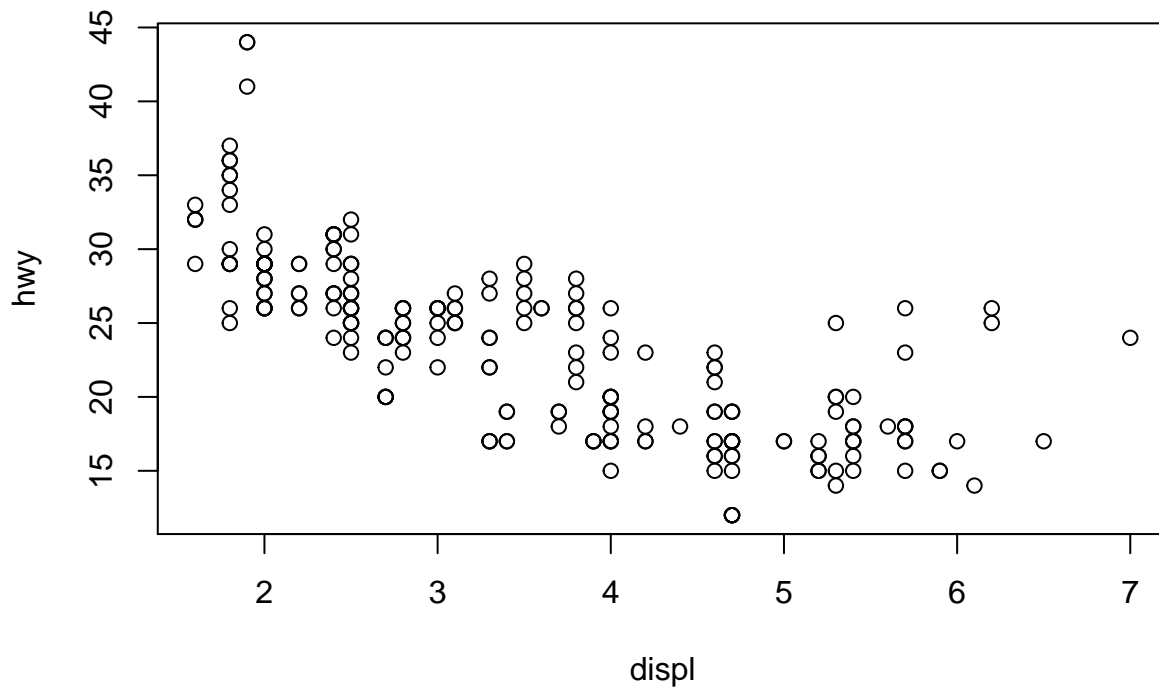
Focuses on when data is given.

Visualisation

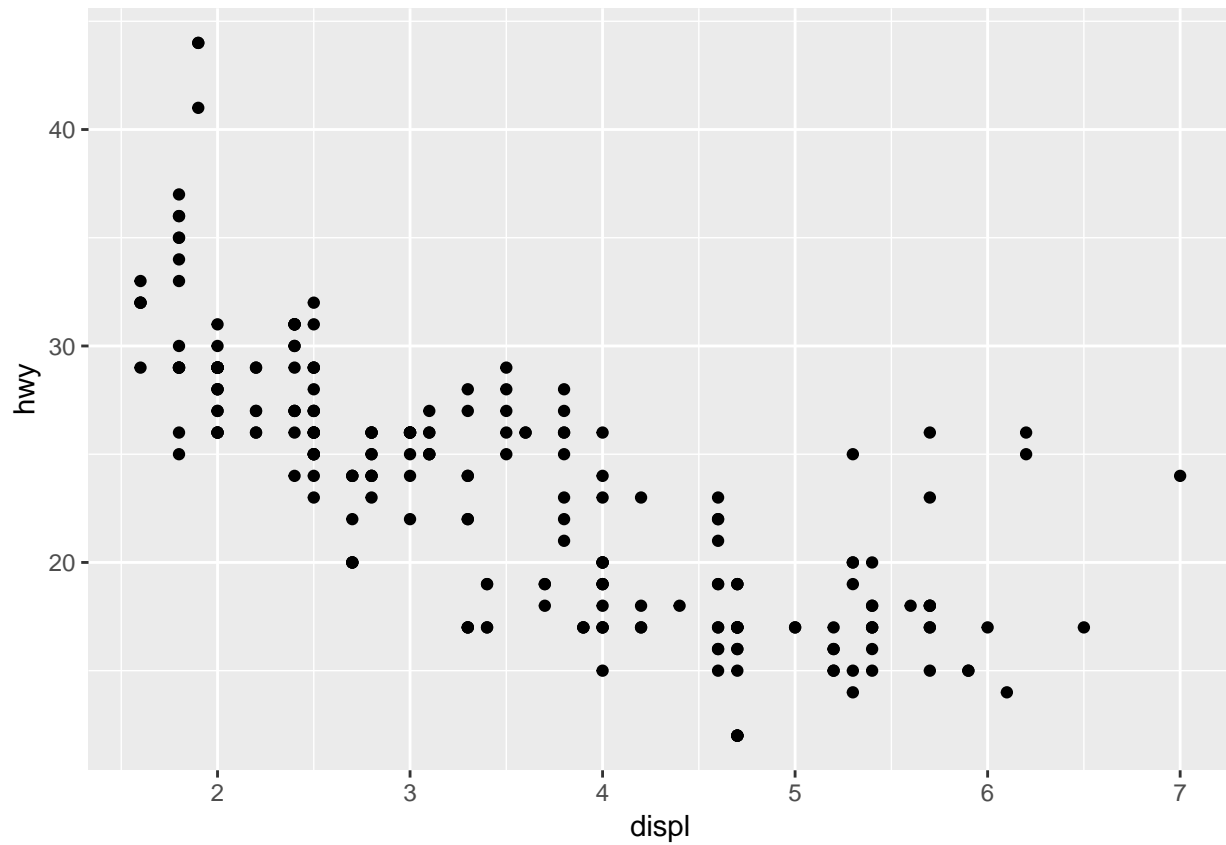
```
mpg
```

```
## # A tibble: 234 x 11
##   manufacturer model displ  year  cyl trans drv   cty   hwy fl  cla~
##   <chr>          <chr> <dbl> <int> <int> <chr> <chr> <int> <int> <chr> <ch>
## 1 audi          a4     1.8  1999    4 auto~ f    18    29 p  com~
## 2 audi          a4     1.8  1999    4 manu~ f    21    29 p  com~
## 3 audi          a4     2    2008    4 manu~ f    20    31 p  com~
## 4 audi          a4     2    2008    4 auto~ f    21    30 p  com~
## 5 audi          a4     2.8  1999    6 auto~ f    16    26 p  com~
## 6 audi          a4     2.8  1999    6 manu~ f    18    26 p  com~
## 7 audi          a4     3.1  2008    6 auto~ f    18    27 p  com~
## 8 audi          a4 q~  1.8  1999    4 manu~ 4    18    26 p  com~
## 9 audi          a4 q~  1.8  1999    4 auto~ 4    16    25 p  com~
## 10 audi         a4 q~  2    2008    4 manu~ 4    20    28 p  com~
## # ... with 224 more rows
```

```
plot(hwy ~ displ, mpg)
```

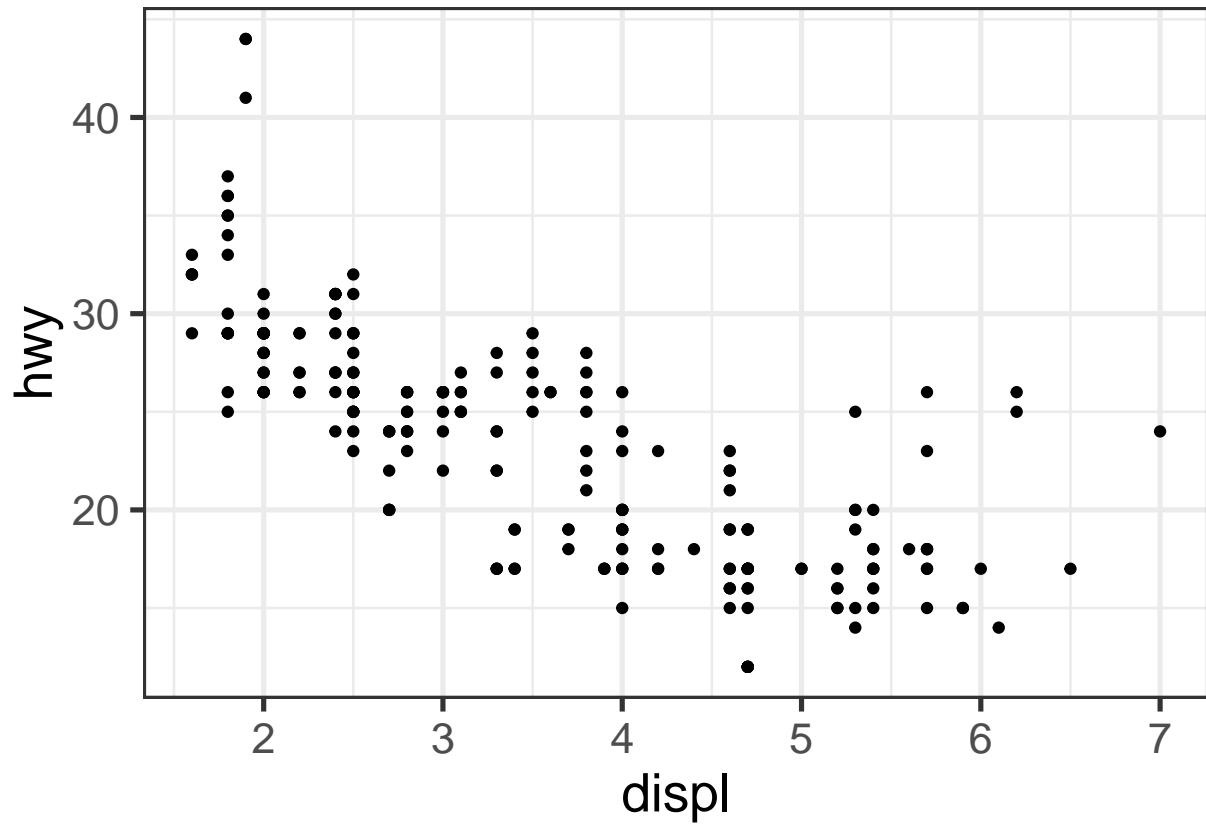


```
ggplot(data = mpg) +
  geom_point(mapping = aes(x = displ, y = hwy))
```

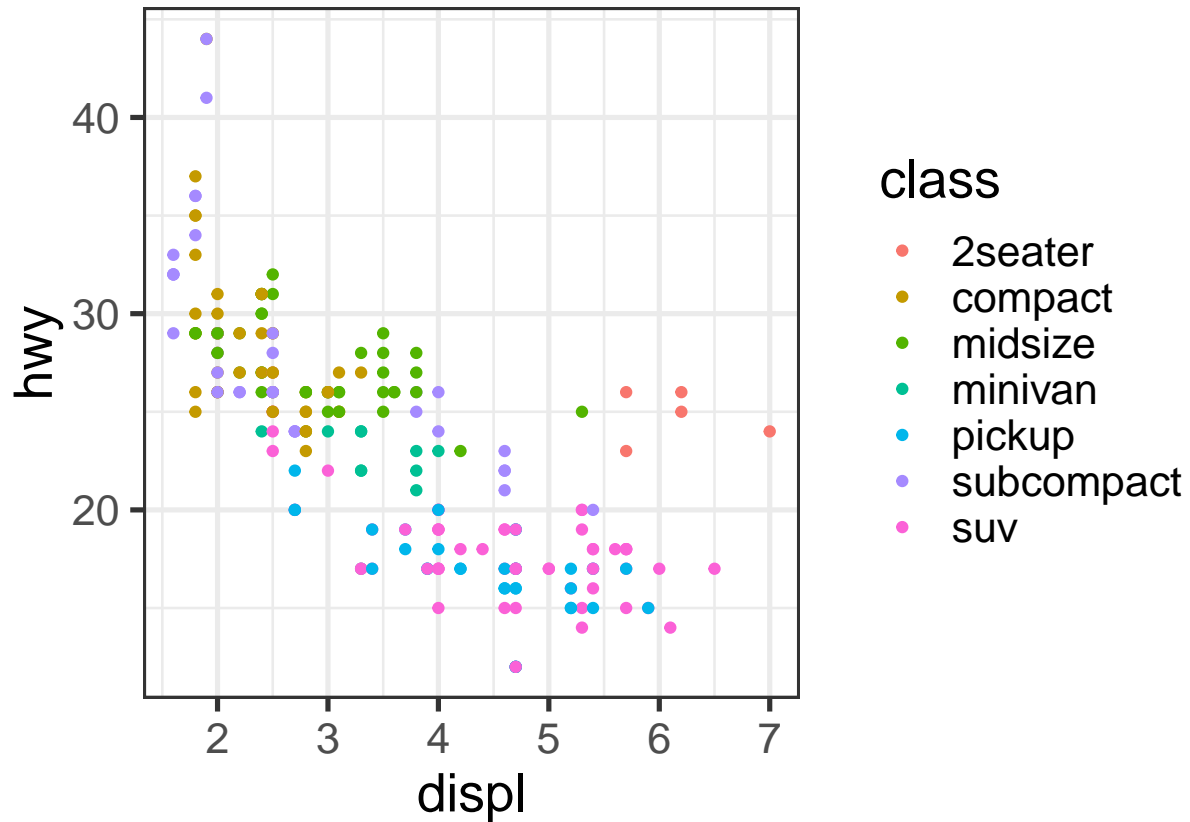


```
# Set black/white theme with larger font size than default  
theme_set(theme_bw(base_size = 20))
```

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



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



ggplot2 (part of tidyverse) based on (layered) grammar of graphics.

```
ggplot(data = <DATA>) +
  <GEOM_FUNCTION>(mapping = aes(<MAPPINGS>))
```

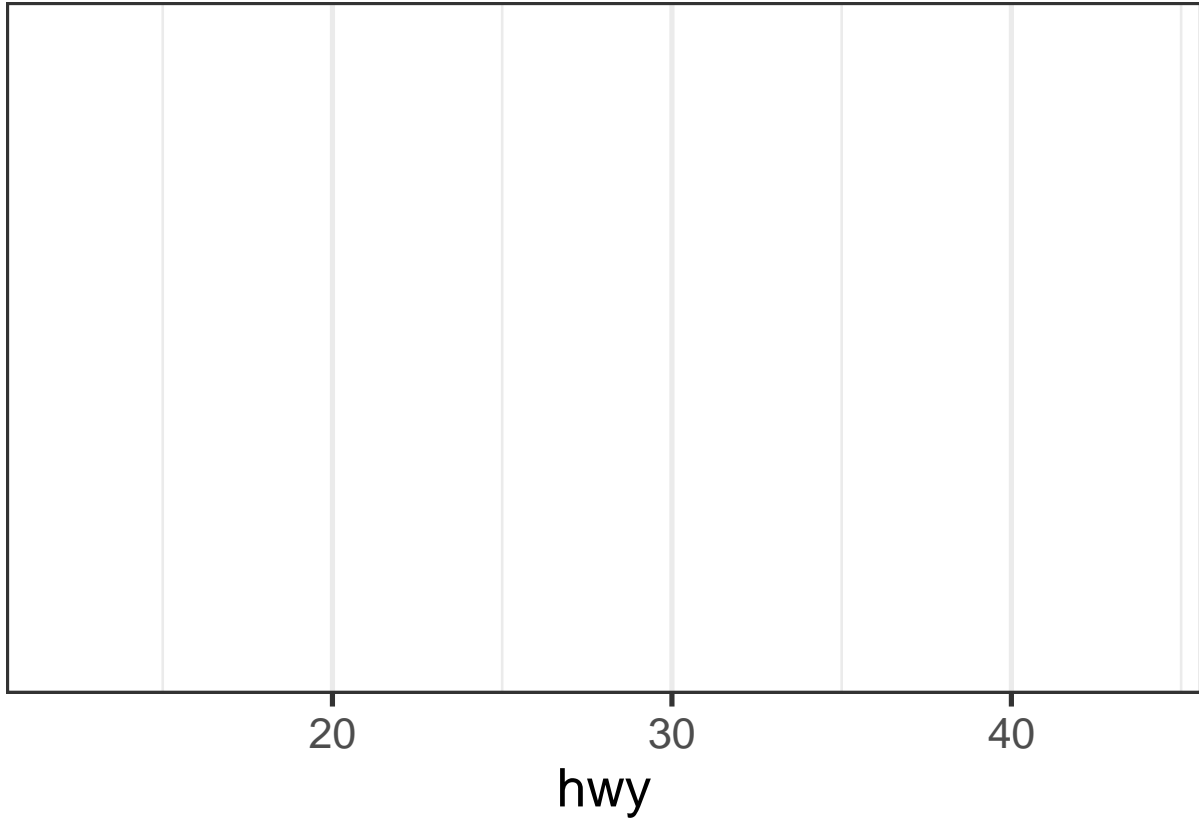
```
ggplot(data = <DATA>, mapping = aes(<MAPPINGS>)) +
  <GEOM_FUNCTION>()
```

- Data + aesthetic mapping (aes) of how to perceive data
- Layers: geom_* + stat_*
- Scales (color, size, ...)
- Coordinate (cartesian, polar, log10, ...)
- Faceting (conditioning or latticing/trellising)
- Theme (font size, background color, ...)
- No suggestions of which plots to use.
- No interactivity, only static graphics.
- Reference: <http://r4ds.had.co.nz/data-visualisation.html>
- Documentation: # <http://ggplot2.tidyverse.org/reference/>

Exercises 3.2.4

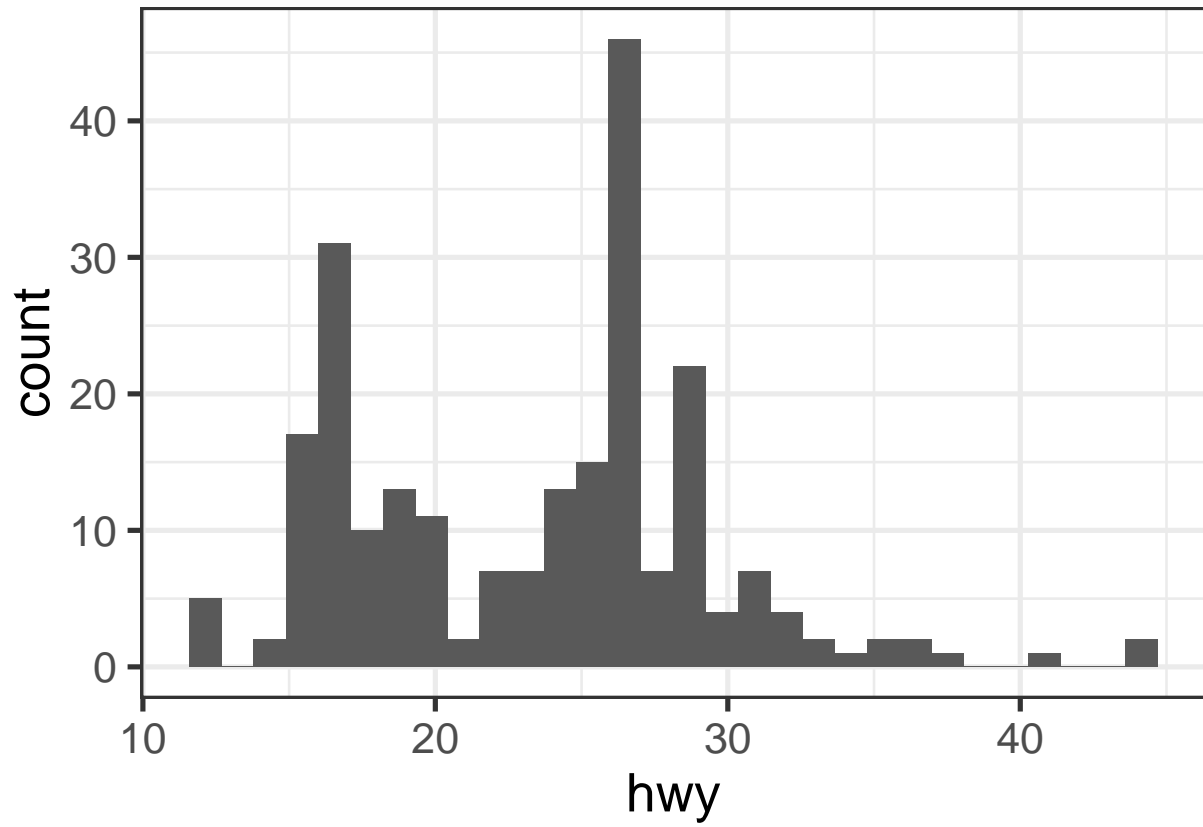
```
ggplot(data = mpg) +
  geom_histogram()
```

```
ggplot(data = mpg, mapping = aes(x = hwy))
```

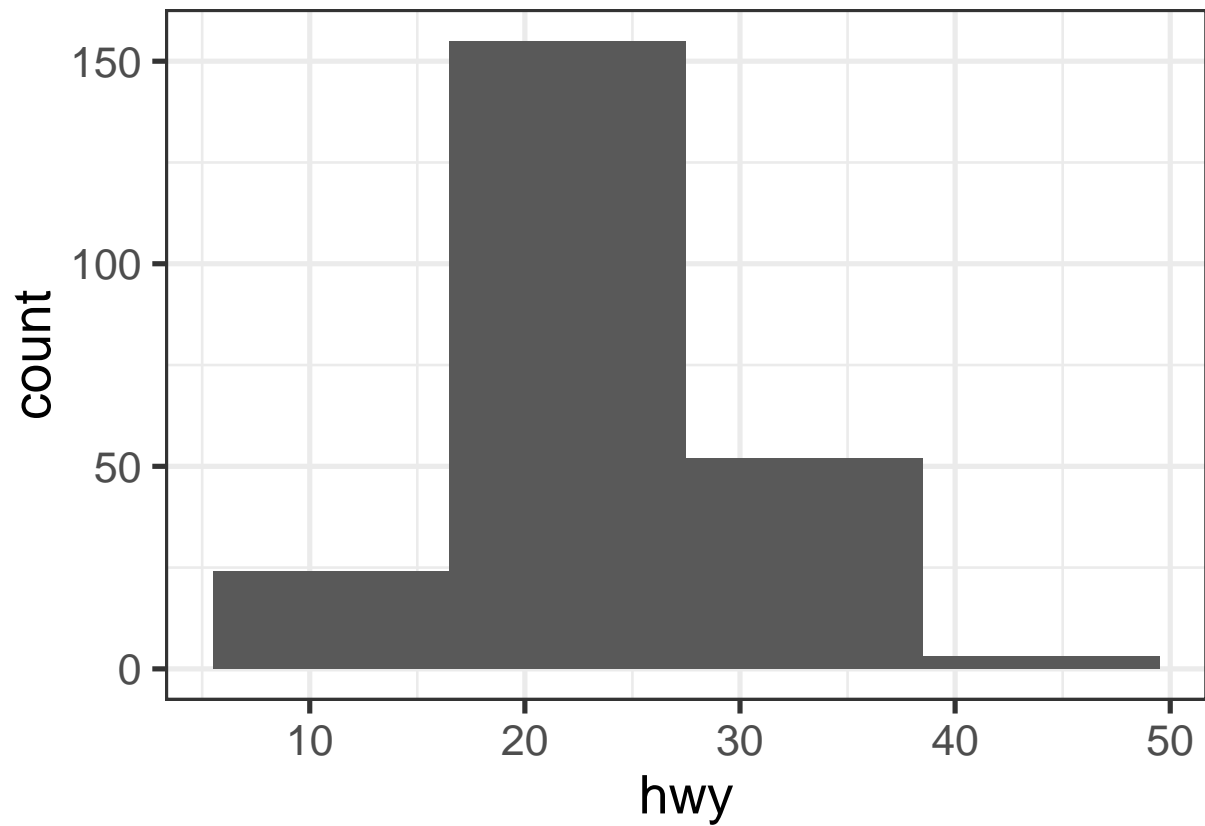


```
ggplot(data = mpg, mapping = aes(x = hwy)) +  
  geom_histogram()
```

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

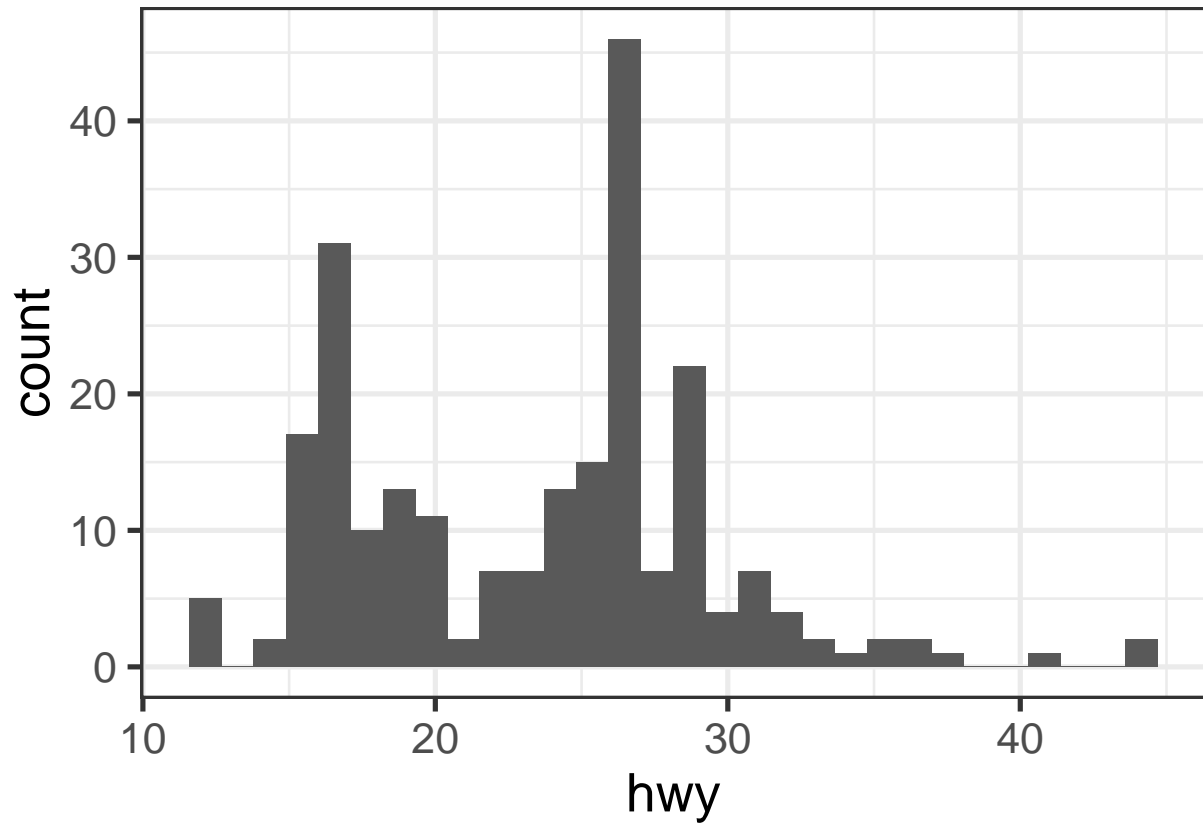


```
ggplot(data = mpg, mapping = aes(hwy)) +  
  geom_histogram(binwidth = nclass.FD)
```



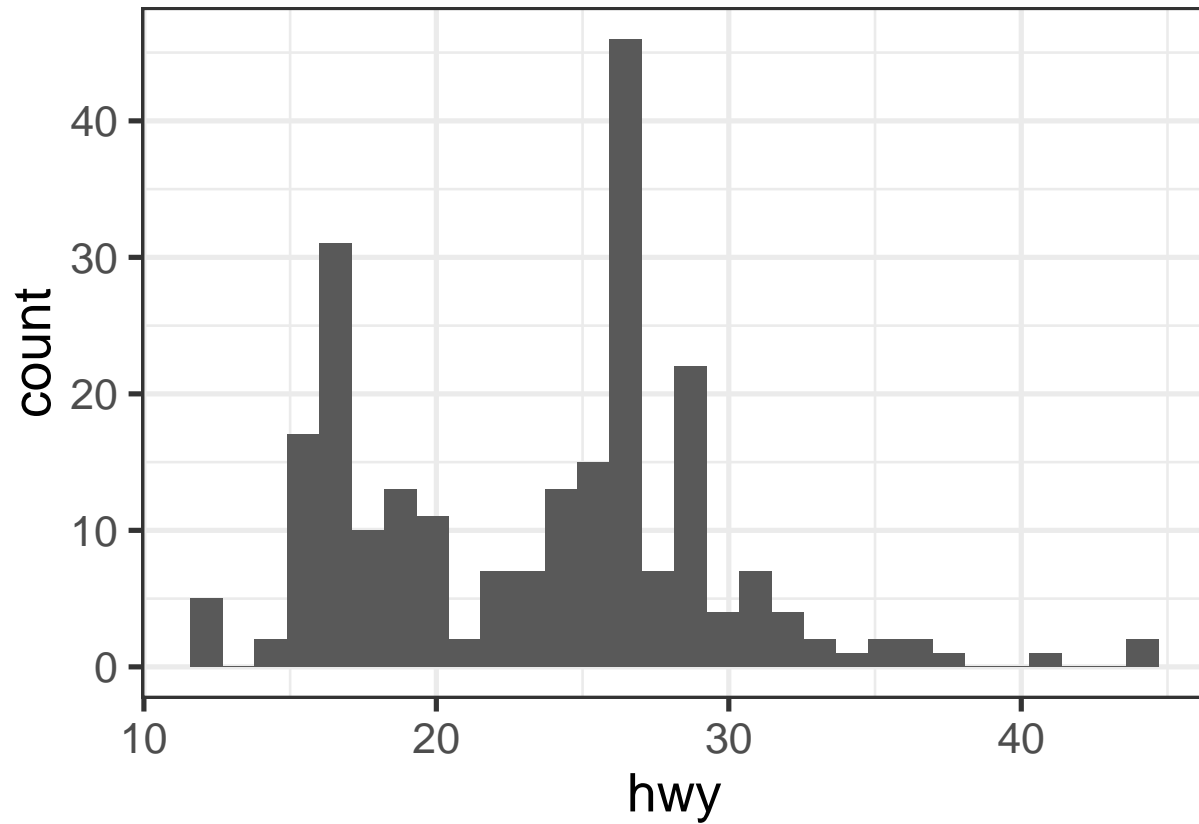
```
ggplot(mpg, aes(hwy)) +  
  geom_histogram()
```

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



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

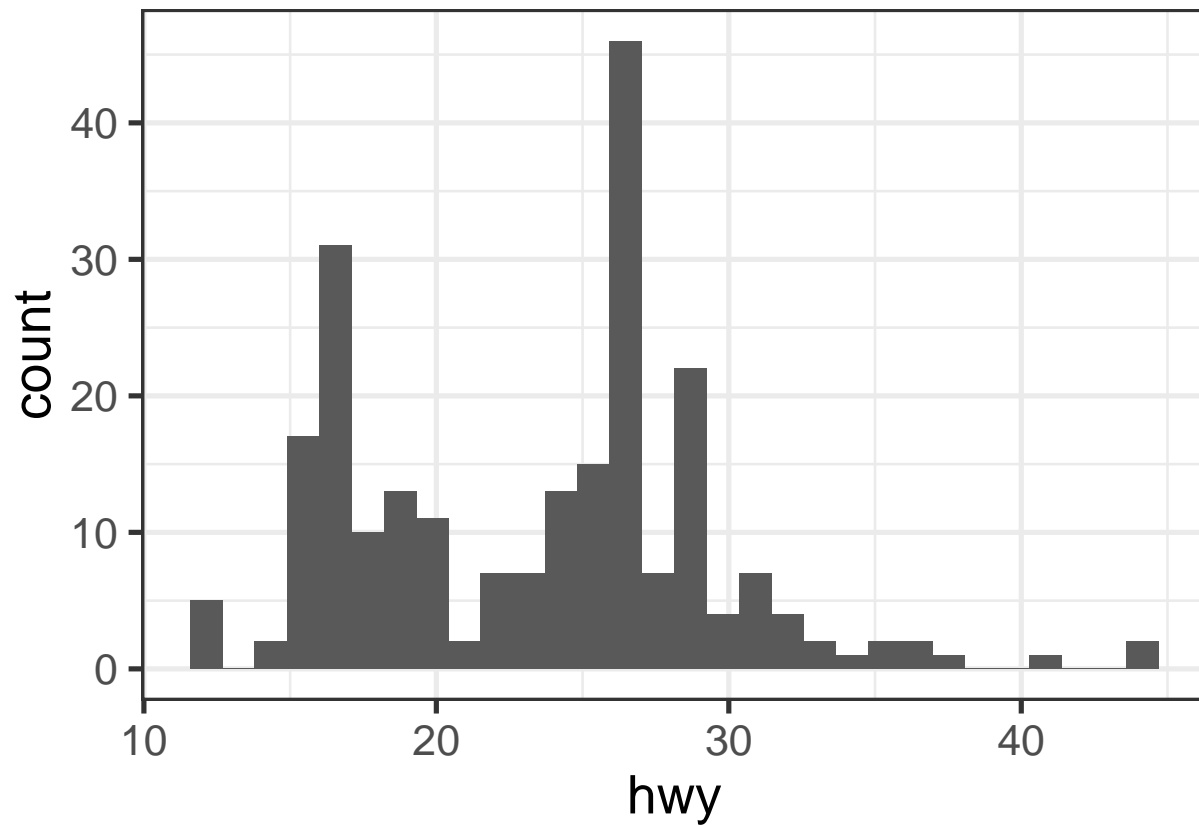
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
# Error:  
#ggplot() +  
# geom_histogram(mpg, aes(hwy))
```

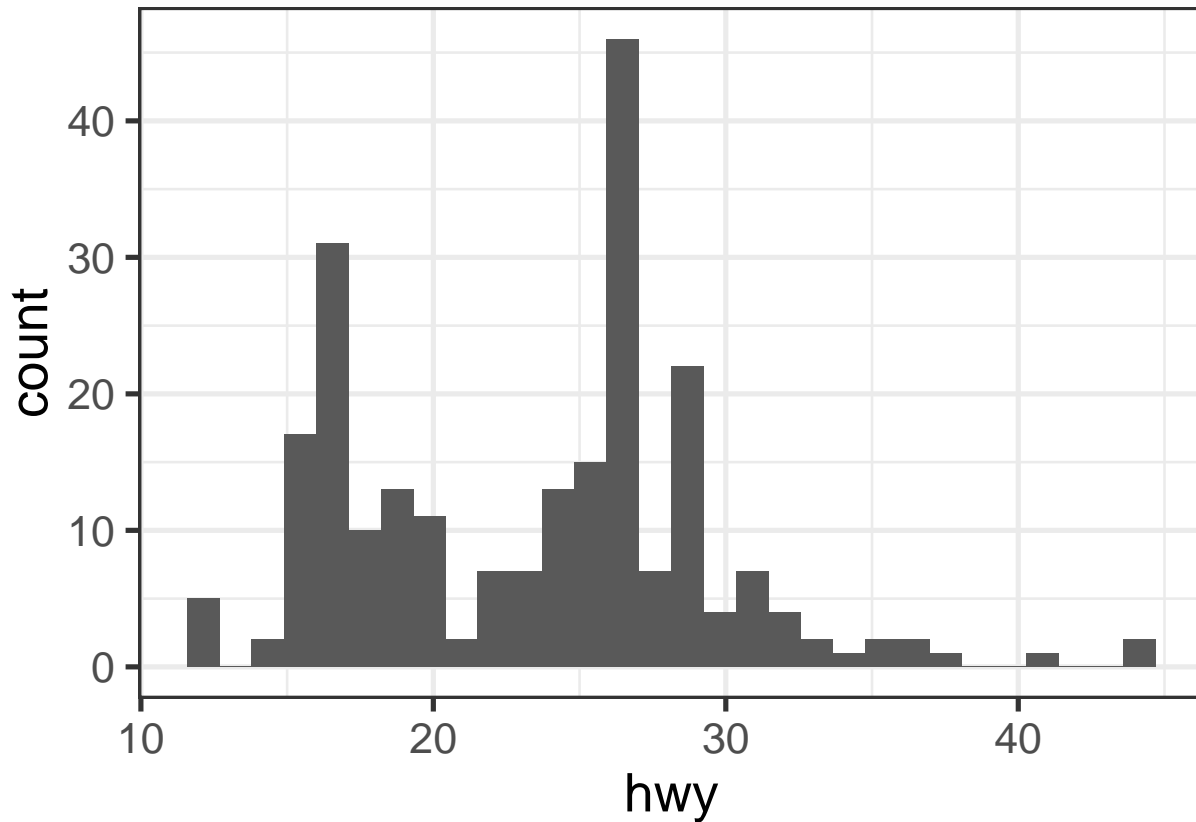
```
ggplot() +  
  geom_histogram(aes(hwy), mpg)
```

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



```
ggplot() +  
  geom_histogram(data = mpg, aes(hwy))
```

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



Generic template:

```
ggplot(data = <DATA>) +
  <GEOM_FUNCTION>(mapping = aes(<MAPPINGS>))
```

Mapping (`aes()`) in `ggplot()` is inherited to layers (`geom_*()`/`stat_*()`) if not overridden.

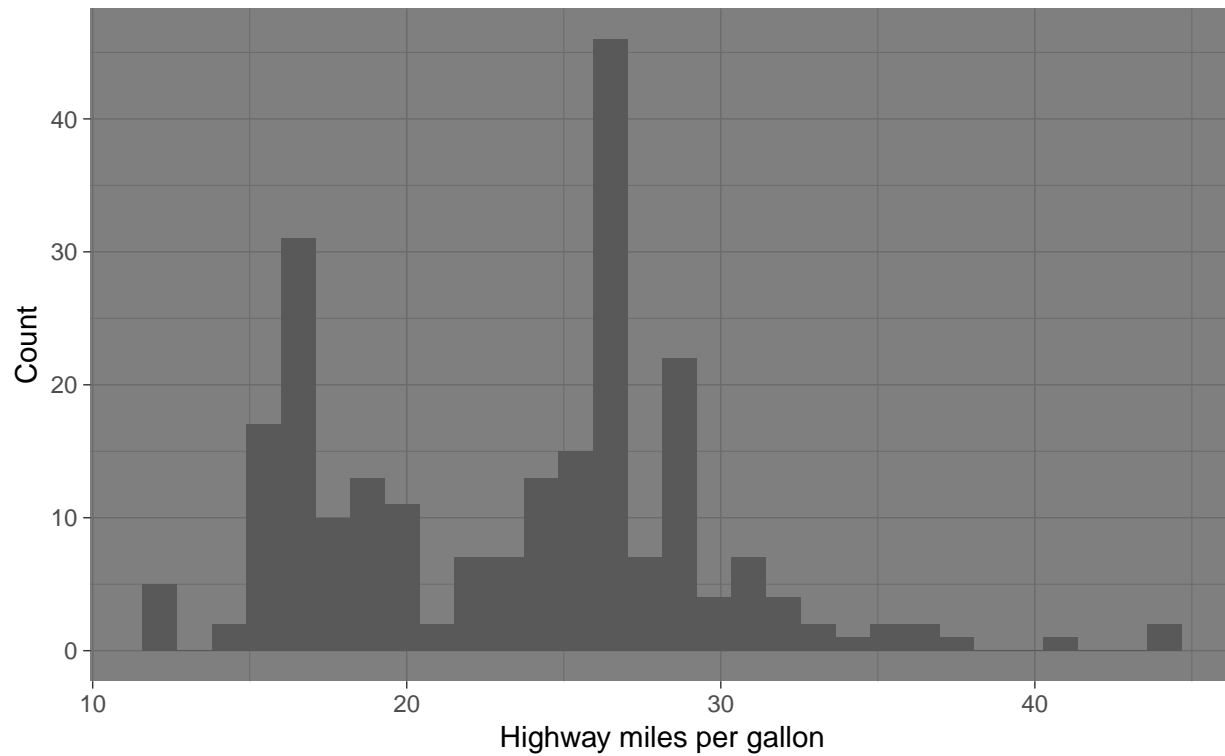
Captions, labels, themes

```
ggplot(mpg, aes(hwy)) +
  geom_histogram() +
  labs(title = "Fuel economy data",
       subtitle = "1999 and 2008 for 38 popular models of car",
       x = "Highway miles per gallon",
       y = "Count") +
  theme_dark()
```

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

Fuel economy data

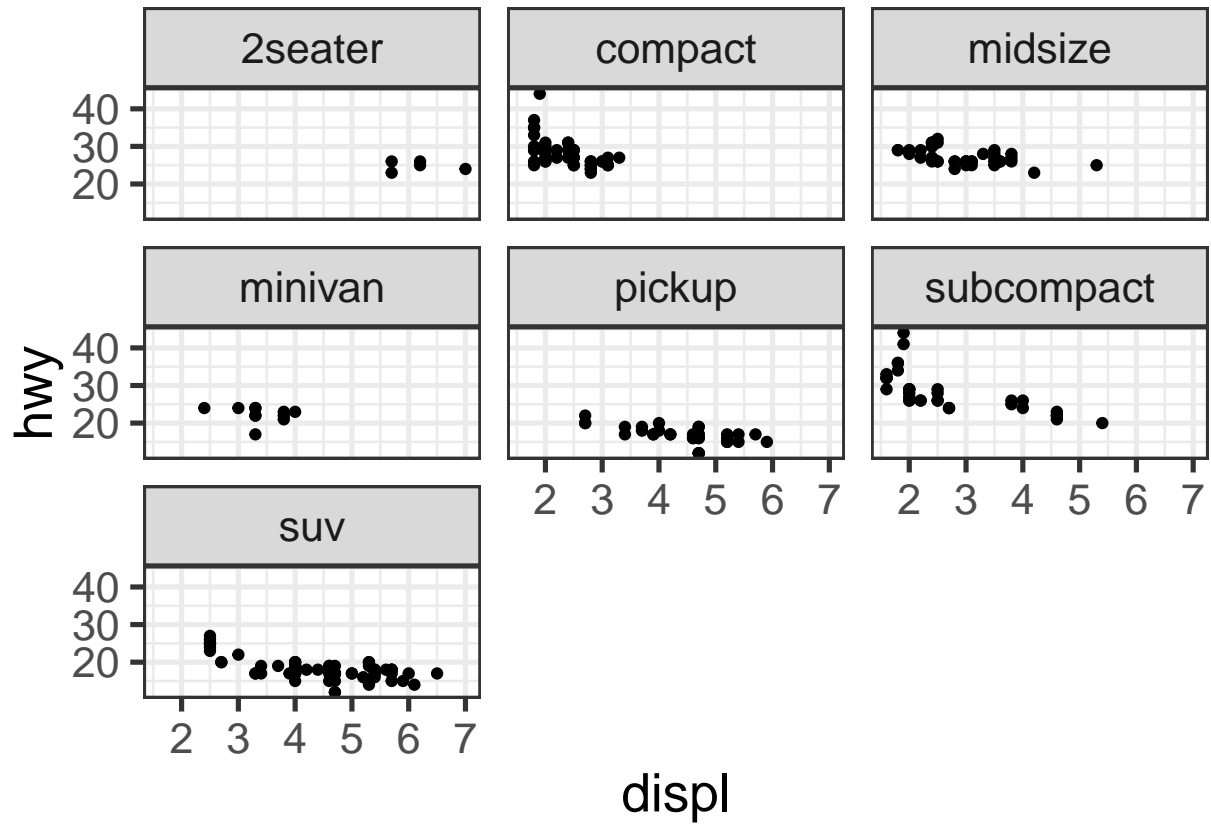
1999 and 2008 for 38 popular models of car



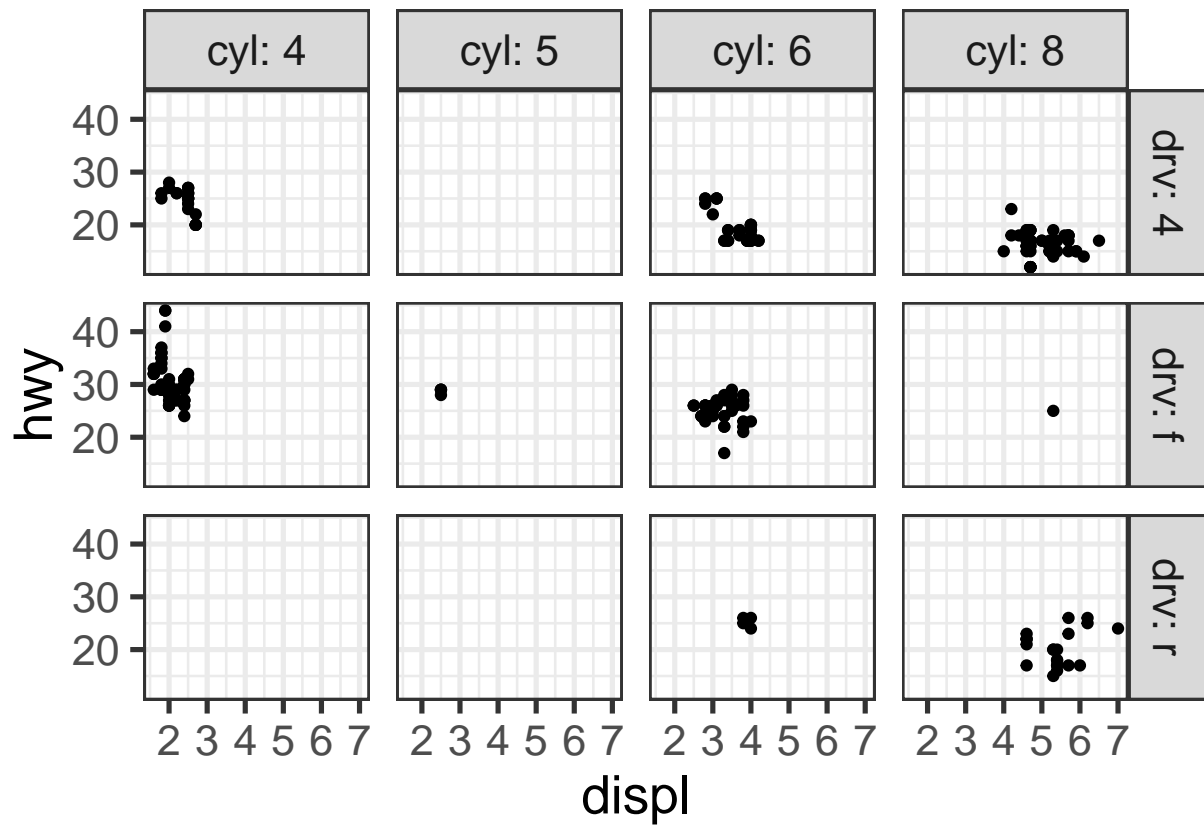
3.3.1 Exercises

Faceting/conditioning/latticing/trellising

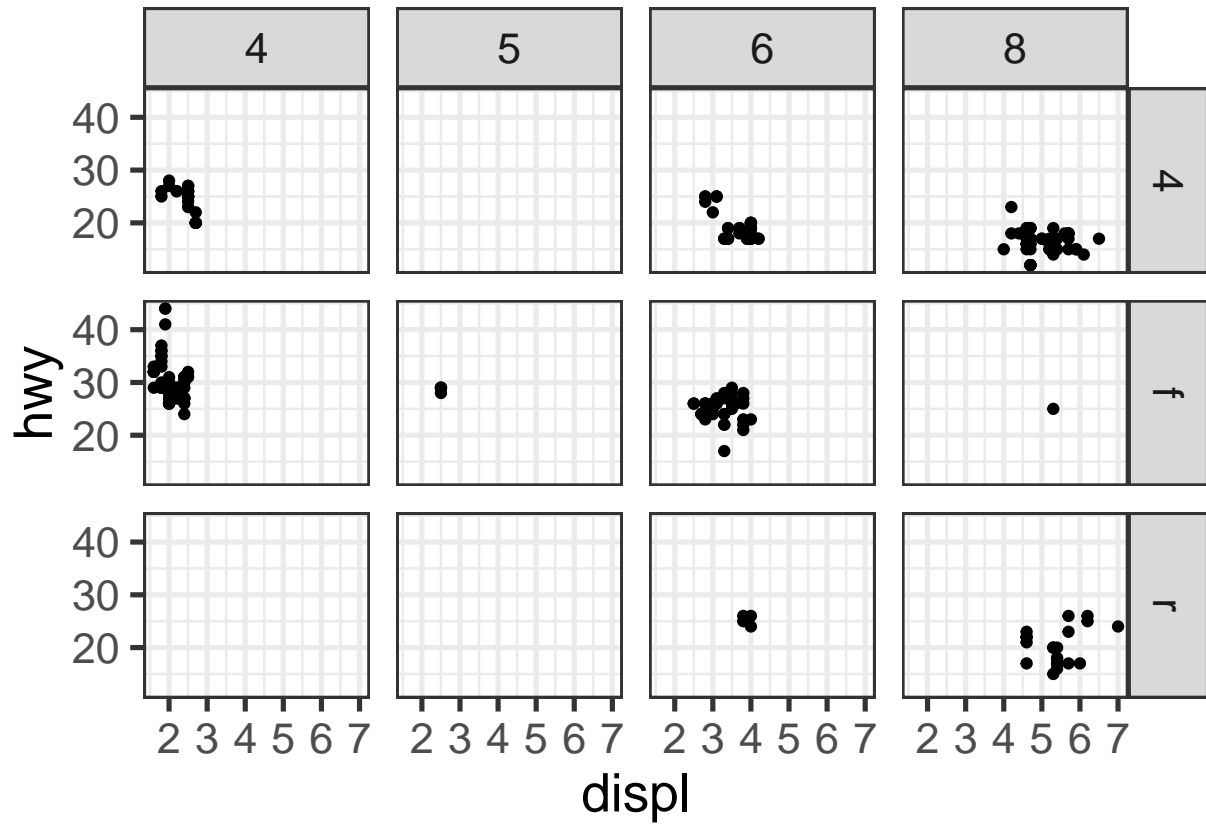
```
ggplot(data = mpg) +  
  geom_point(mapping = aes(x = displ, y = hwy)) +  
  facet_wrap(~ class, nrow = 3)
```



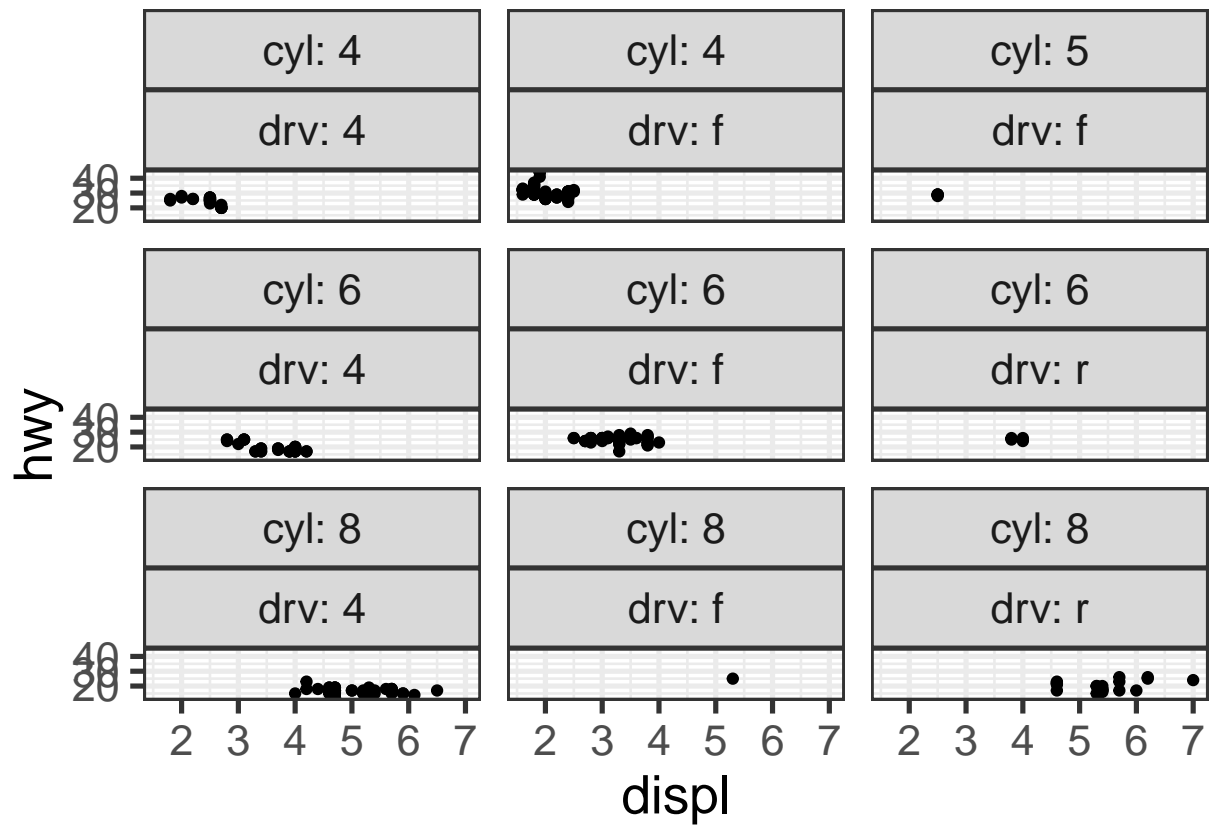
```
ggplot(data = mpg) +
  geom_point(mapping = aes(x = displ, y = hwy)) +
  facet_grid(drv ~ cyl, labeller = label_both)
```



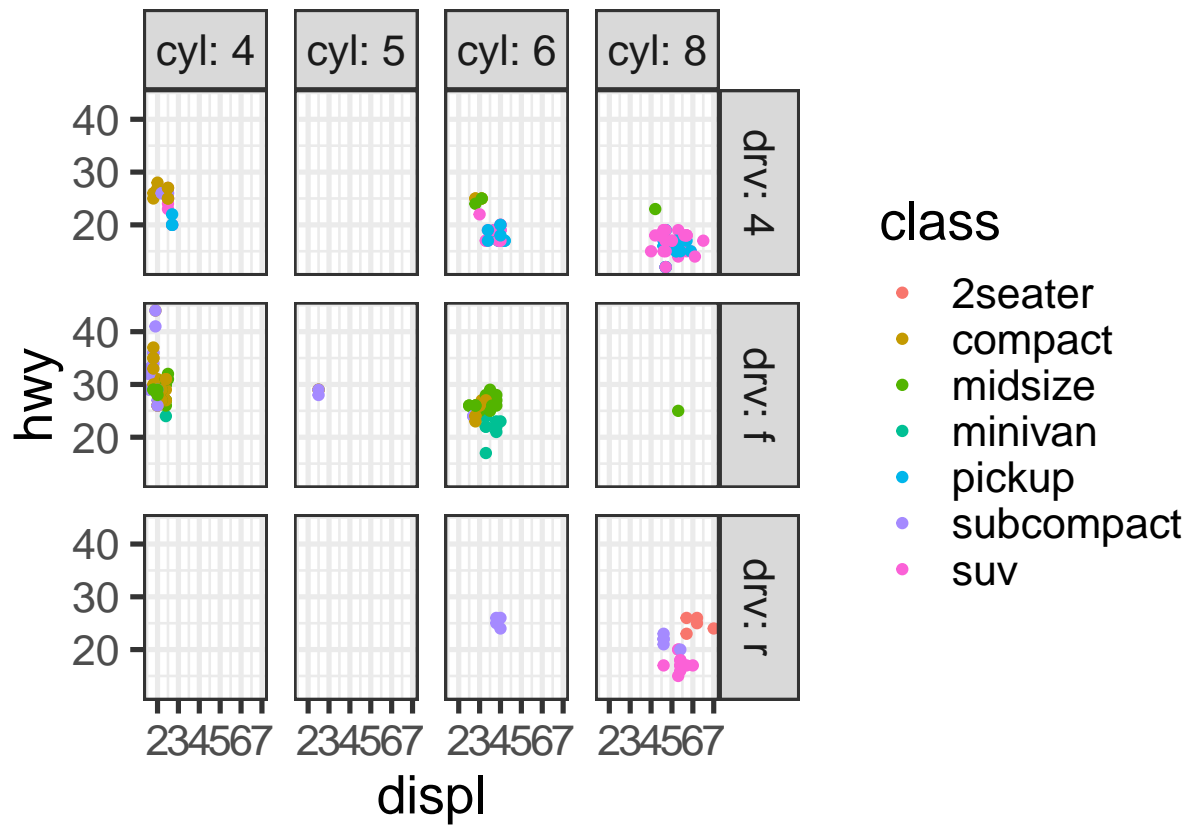
```
ggplot(data = mpg) +
  geom_point(mapping = aes(x = displ, y = hwy)) +
  facet_grid(drv ~ cyl)
```



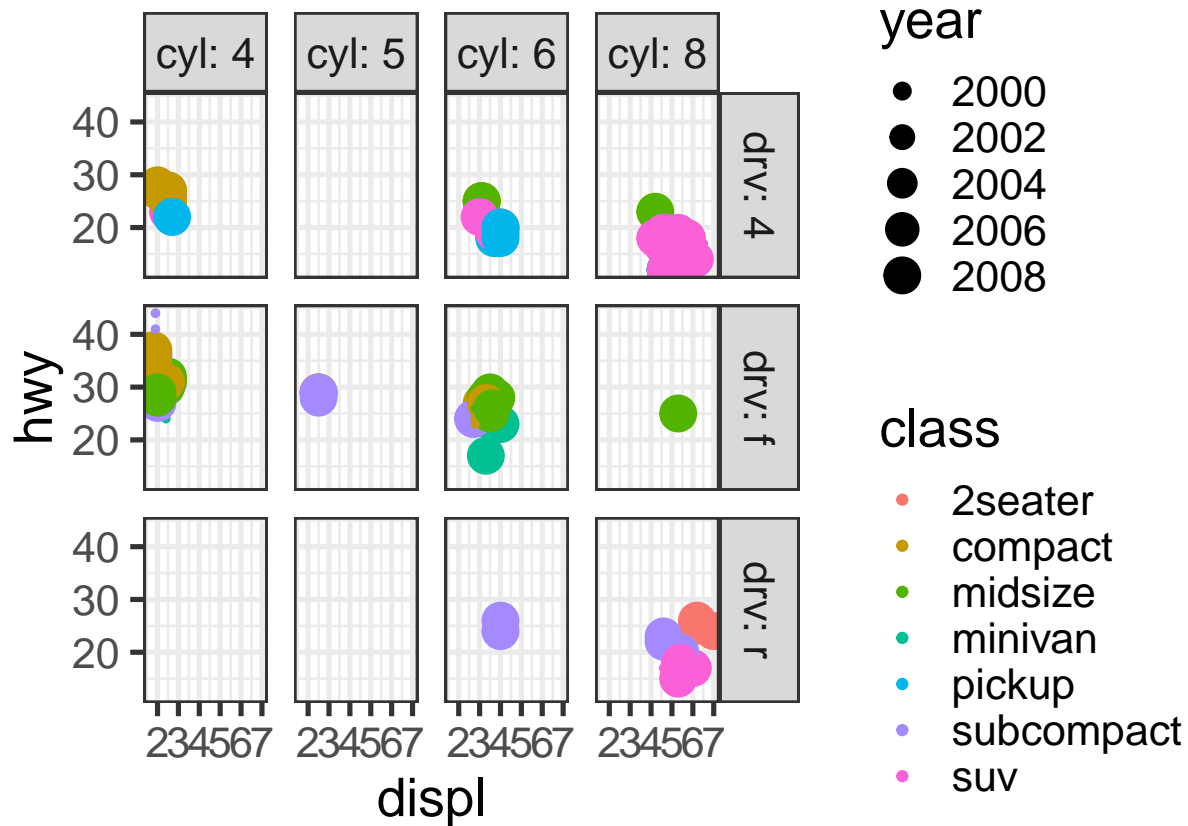
```
ggplot(data = mpg) +
  geom_point(mapping = aes(x = displ, y = hwy)) +
  facet_wrap(~ cyl + drv, labeller = label_both)
```



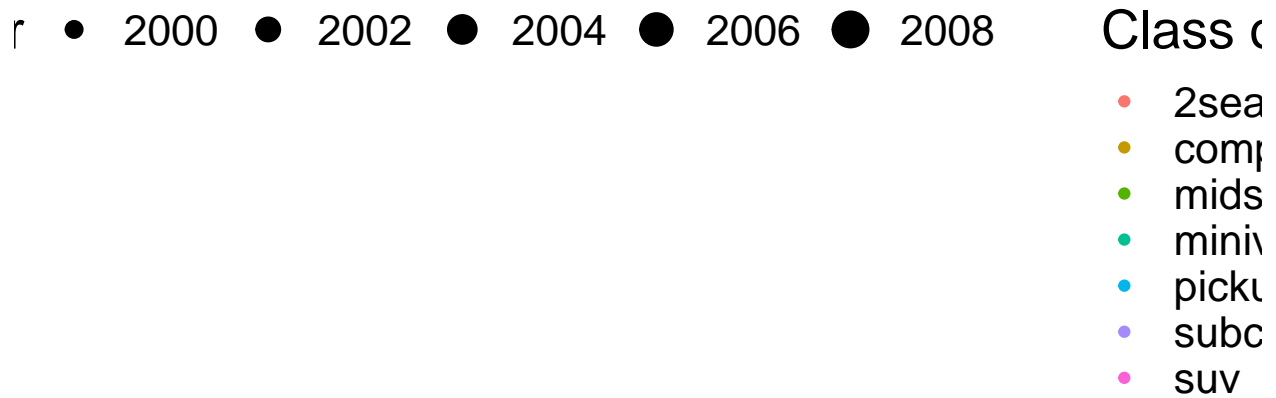
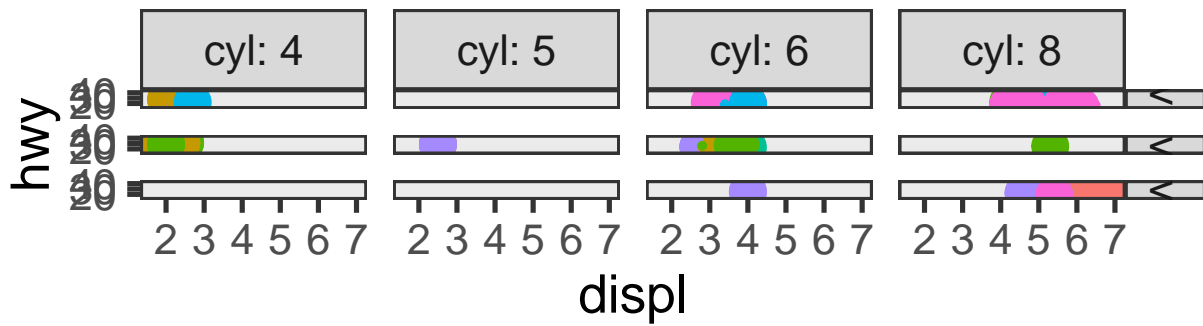
```
ggplot(data = mpg) +
  geom_point(mapping = aes(x = displ, y = hwy, color = class)) +
  facet_grid(drv ~ cyl, labeller = label_both)
```

```
ggplot(data = mpg) +
  geom_point(mapping = aes(x = displ, y = hwy, size = year, color = class)) +
  facet_grid(drv ~ cyl, labeller = label_both)
```



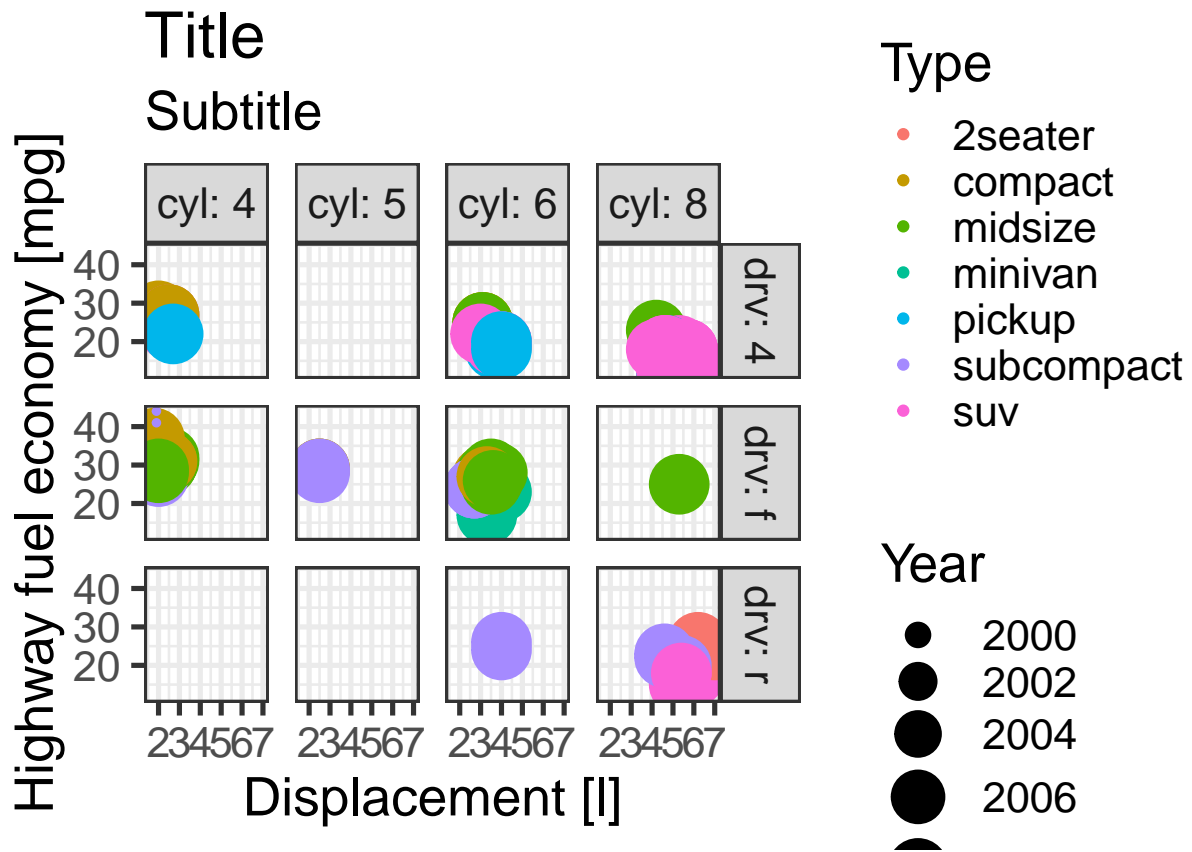
```
ggplot(data = mpg) +
  geom_point(mapping = aes(x = displ, y = hwy, size = year, color = class)) +
  facet_grid(drv ~ cyl, labeller = label_both) +
  theme(legend.position = "bottom") +
  scale_color_discrete("Class of car") +
  guides(color = guide_legend(title.position = "top",
                              ncol = 1))
```



```

ggplot(data = mpg) +
  geom_point(mapping = aes(x = displ, y = hwy, size = year, color = class)) +
  facet_grid(drv ~ cyl, labeller = label_both) +
  scale_size_continuous("Year", range = c(1, 10)) +
  scale_color_discrete("Type") +
  labs(title = "Title",
       subtitle = "Subtitle",
       x = "Displacement [l]",
       y = "Highway fuel economy [mpg]")

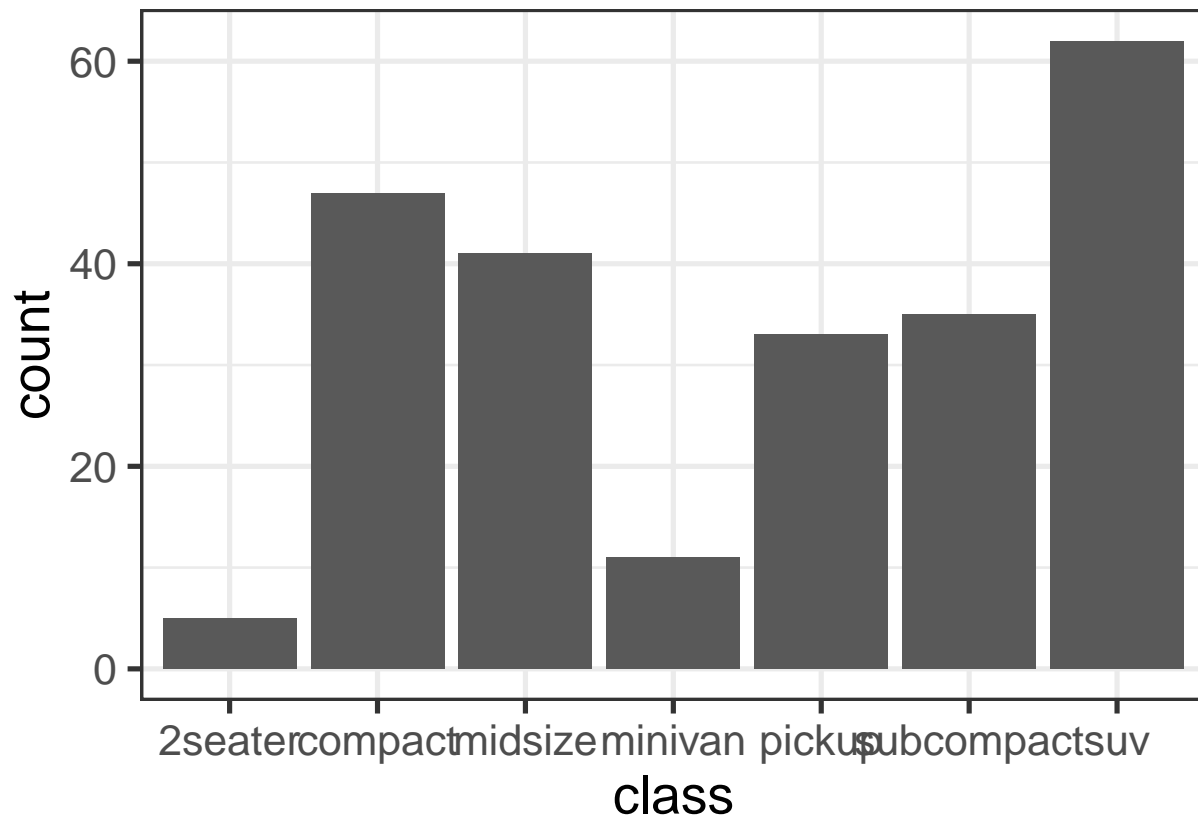
```



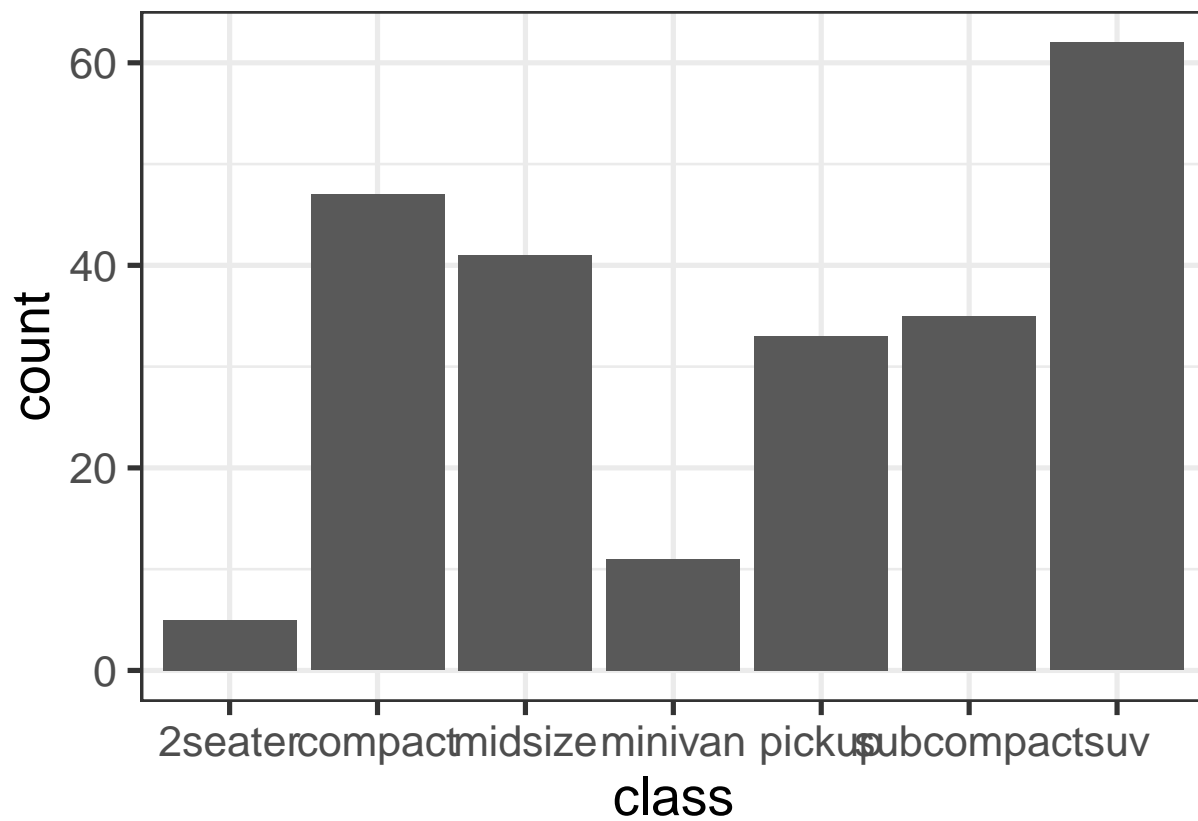
3.5.1 Exercises

Bar plots

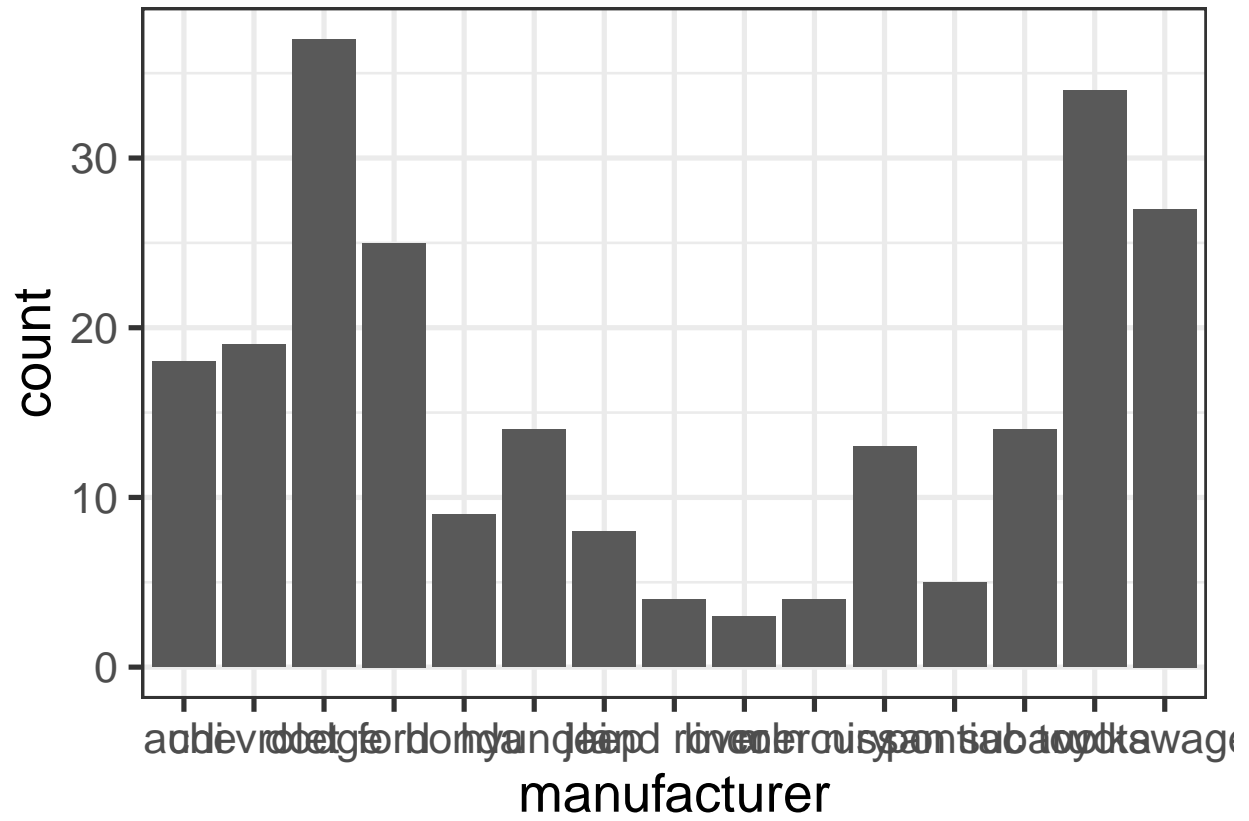
```
ggplot(data = mpg, mapping = aes(x = class)) +
  geom_bar()
```



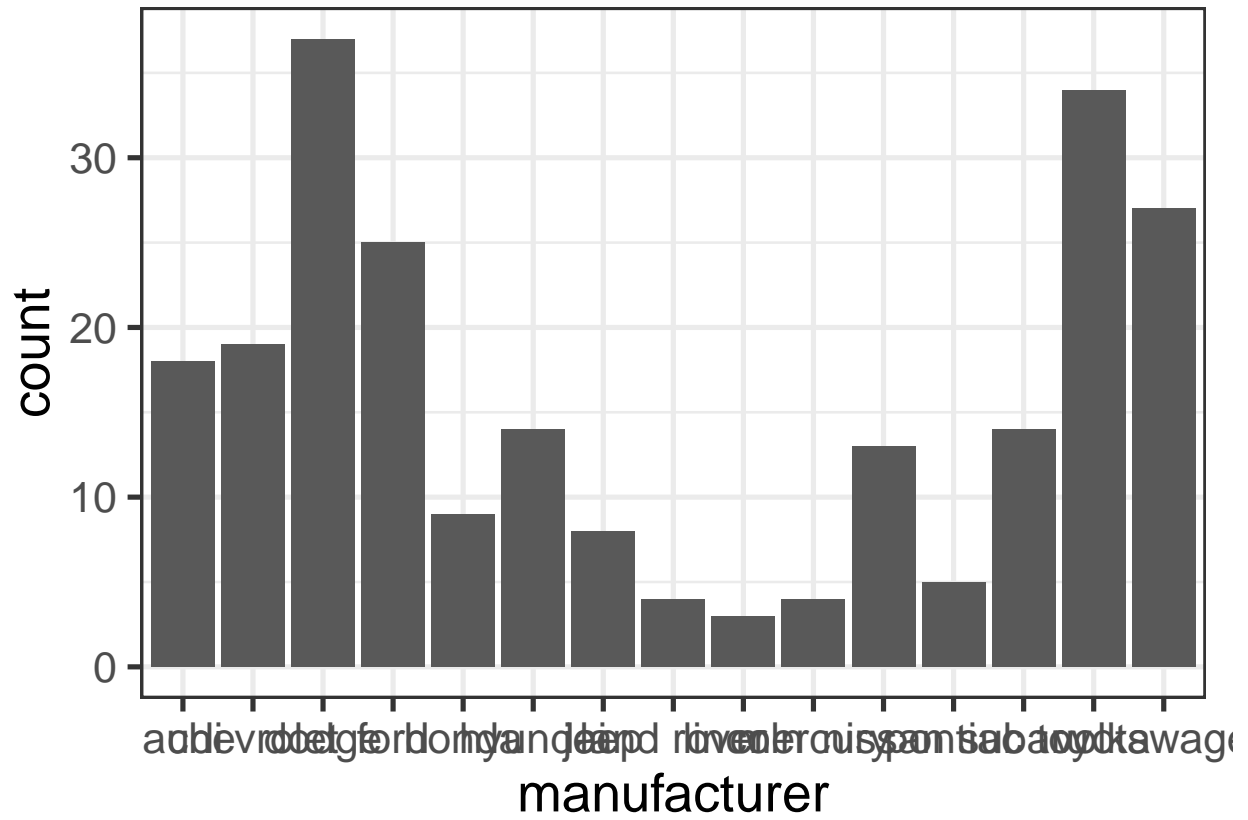
```
ggplot(mpg, aes(class)) +
  geom_bar()
```



```
p <- ggplot(mpg, aes(manufacturer)) +  
  geom_bar()  
p
```



```
print(p)
```



```
1:5
```

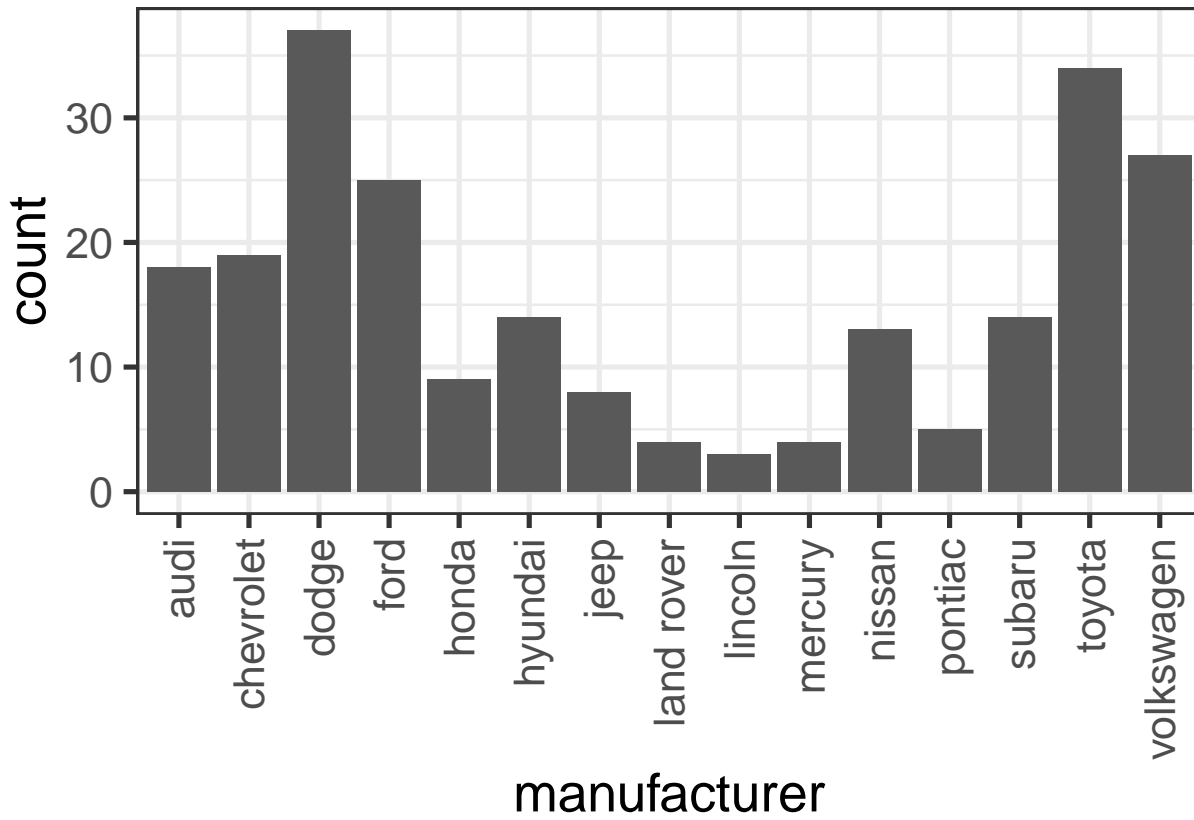
```
## [1] 1 2 3 4 5
```

```
print(1:5)
```

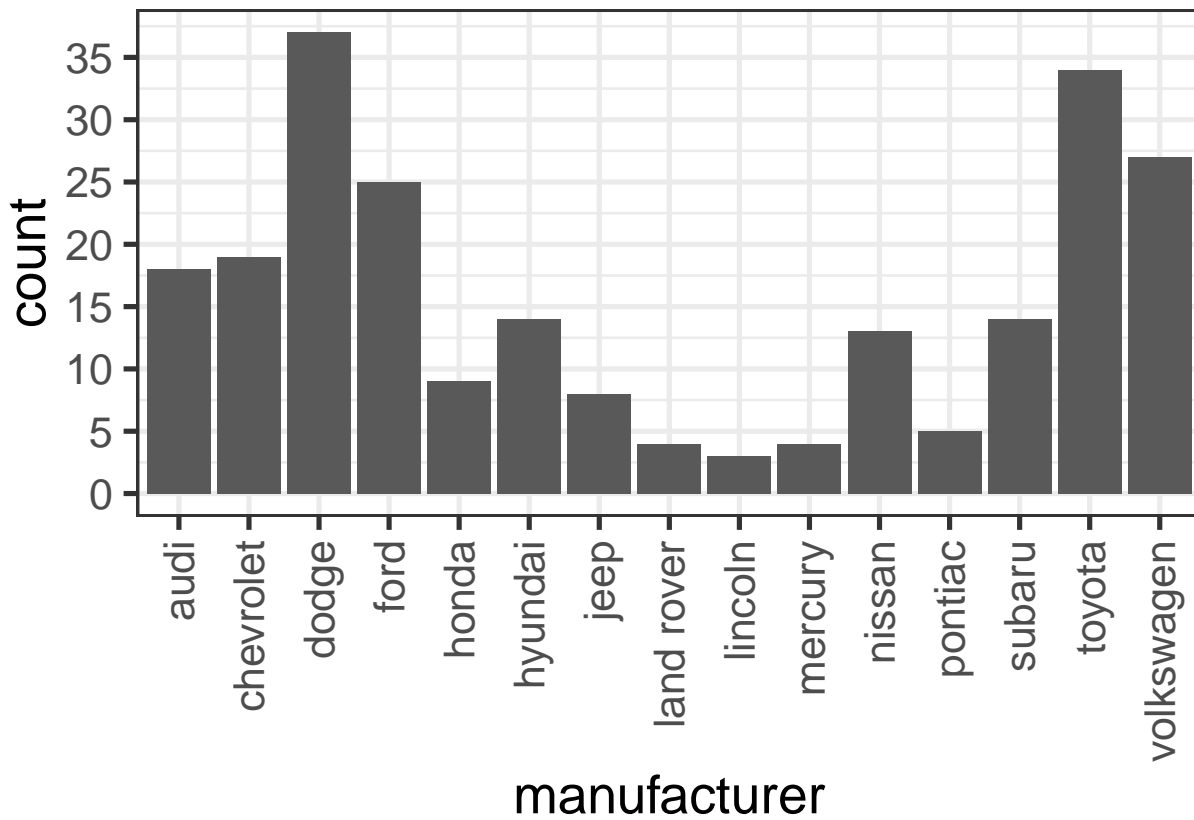
```
## [1] 1 2 3 4 5
```

```
p2 <- p +
  theme(axis.text.x = element_text(angle = 90,
                                    hjust = 1,
                                    vjust = 0.5))
```

```
p2
```



```
p2 + scale_y_continuous(breaks = scales::pretty_breaks(n = 10))
```

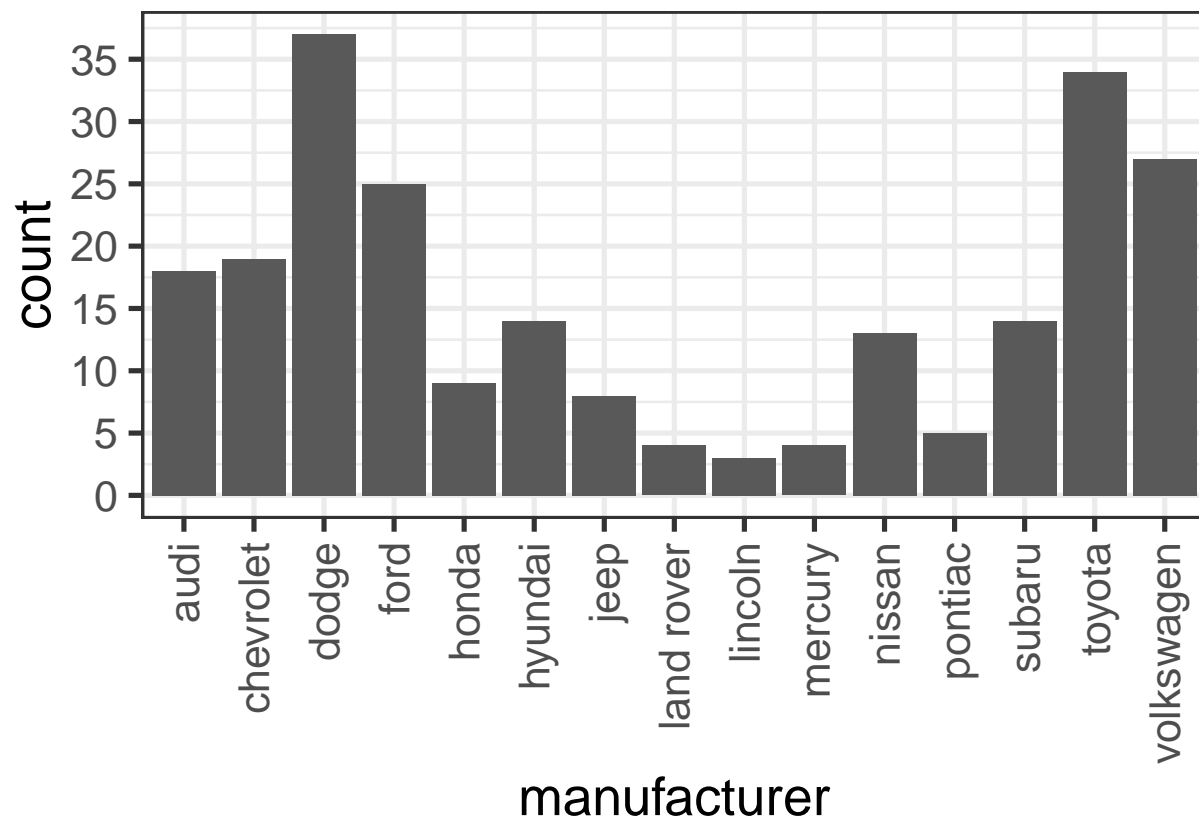



```
#p2 + scale_y_continuous(breaks = c(0, 7, 17))
```

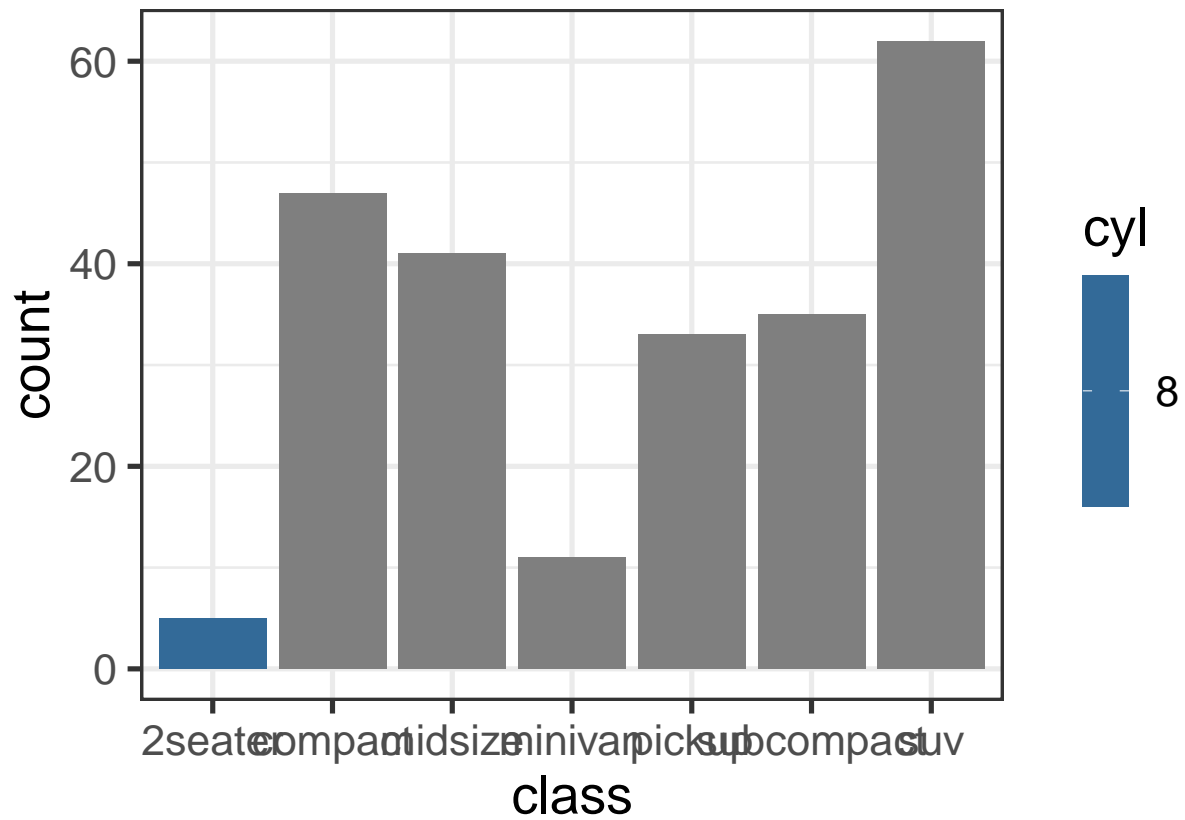
```
library(scales)
```

```
##  
## Attaching package: 'scales'  
## The following object is masked from 'package:purrr':  
##  
##   discard  
## The following object is masked from 'package:readr':  
##  
##   col_factor
```

```
p2 + scale_y_continuous(breaks = pretty_breaks(n = 10))
```



```
ggplot(mpg, aes(class, fill = cyl)) +  
  geom_bar()
```



```
str(mpg$cyl)
```

```
## int [1:234] 4 4 4 4 6 6 6 4 4 4 ...
```

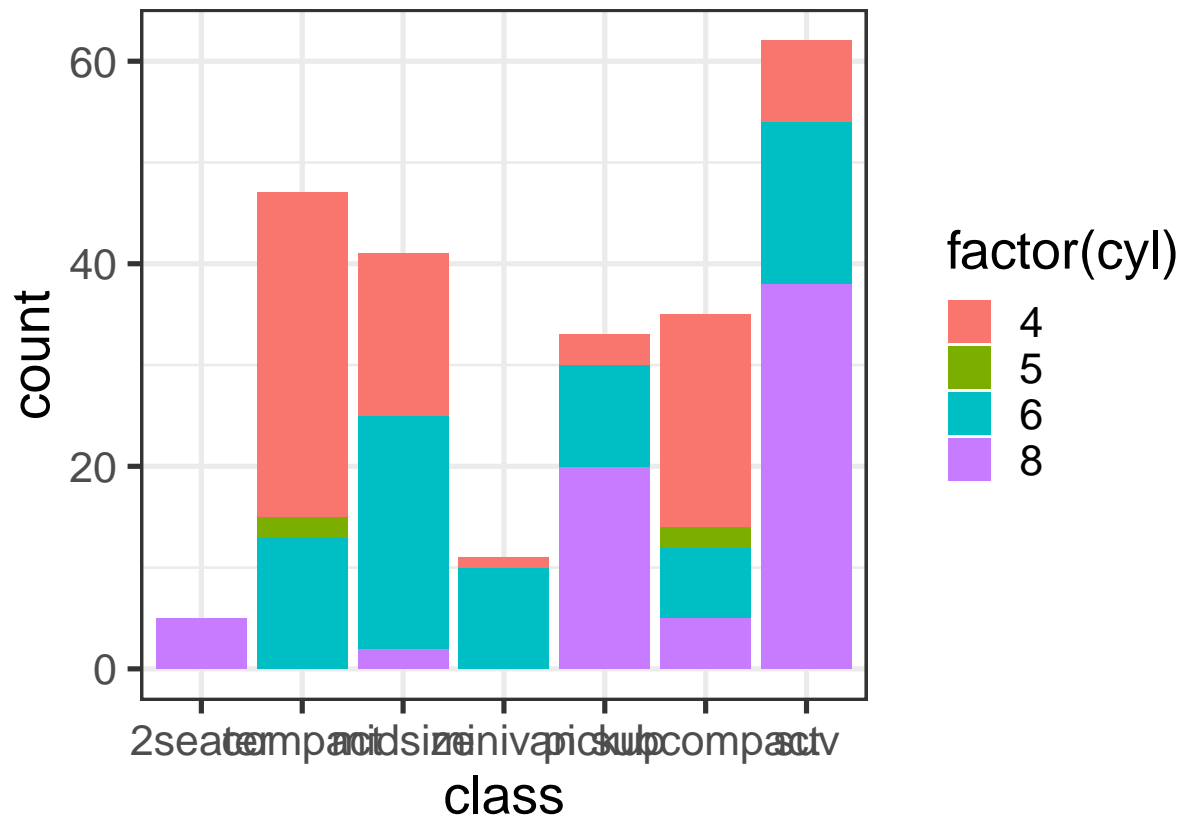
```
table(mpg$cyl)
```

```
##
```

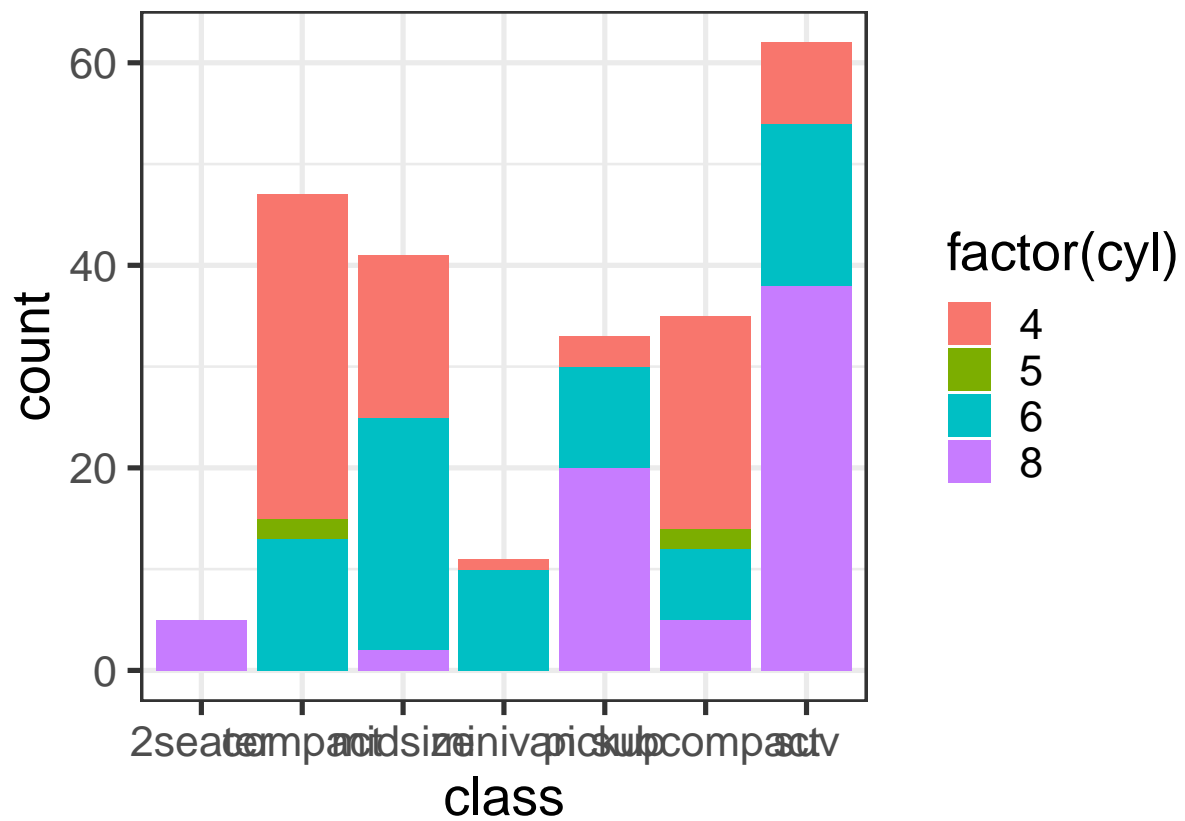
```
## 4 5 6 8
```

```
## 81 4 79 70
```

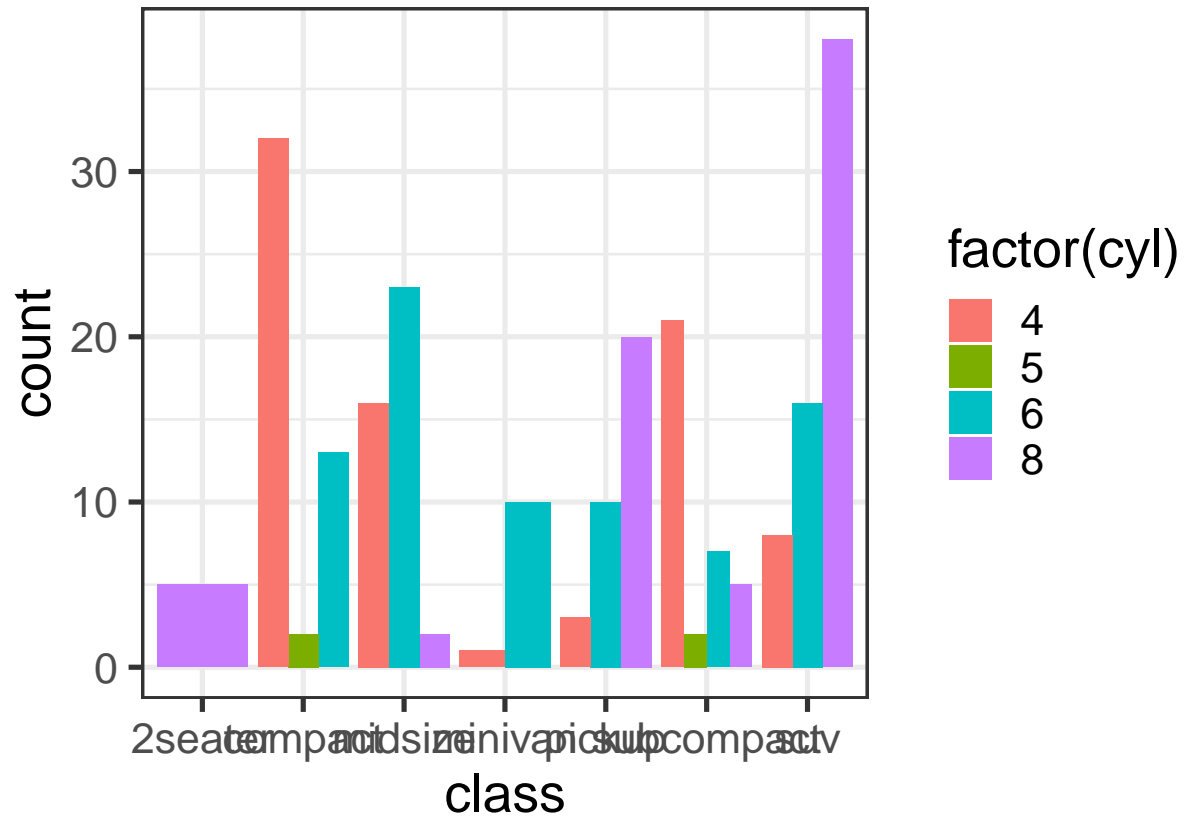
```
ggplot(mpg, aes(class, fill = factor(cyl))) +  
  geom_bar()
```



```
ggplot(mpg, aes(class, fill = factor(cyl))) +
  geom_bar()
```

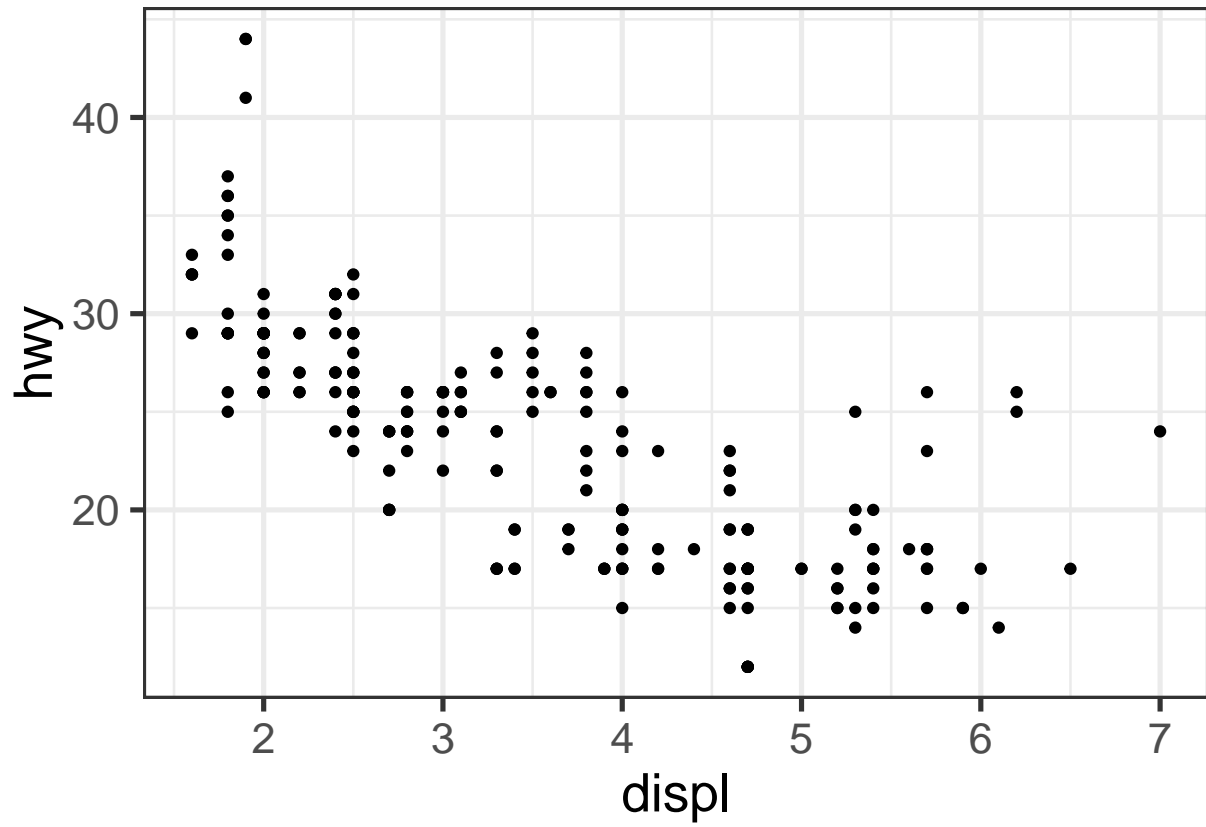


```
ggplot(mpg, aes(class, fill = factor(cyl))) +
  geom_bar(position = position_dodge())
```

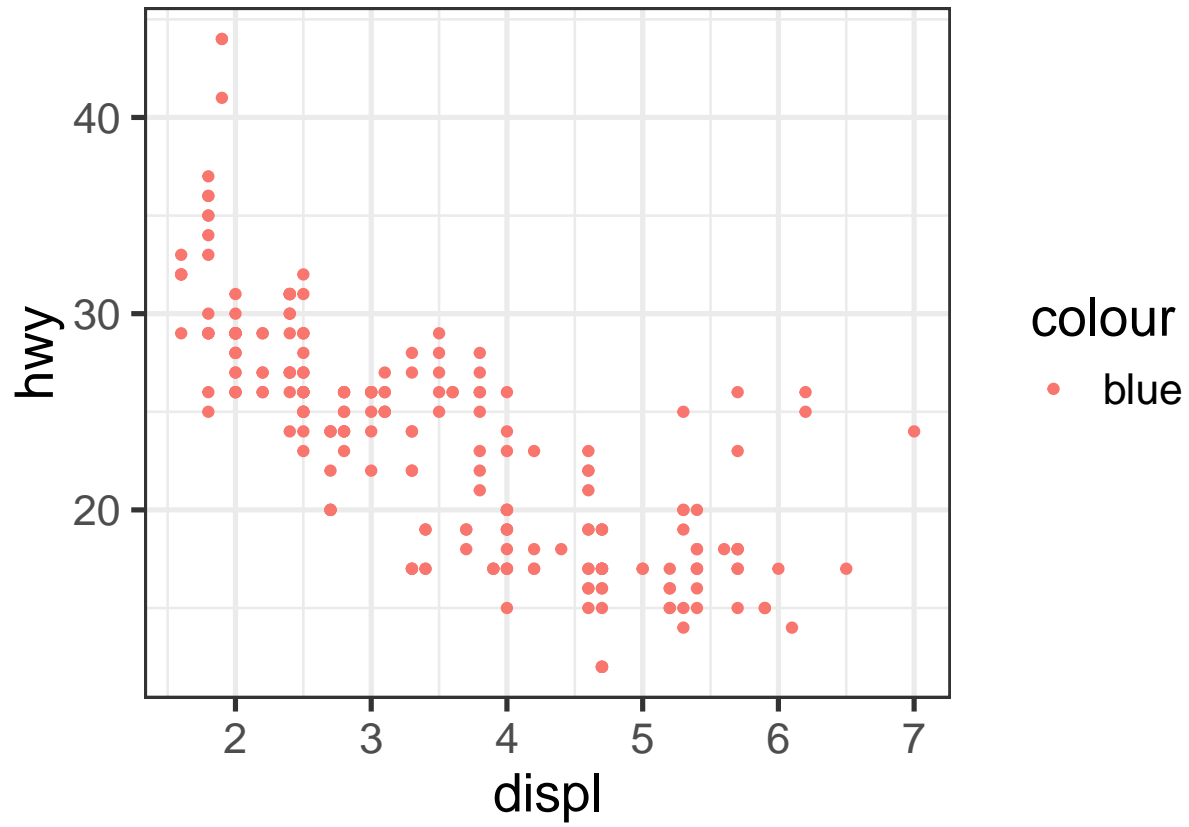


Points / scatter plots

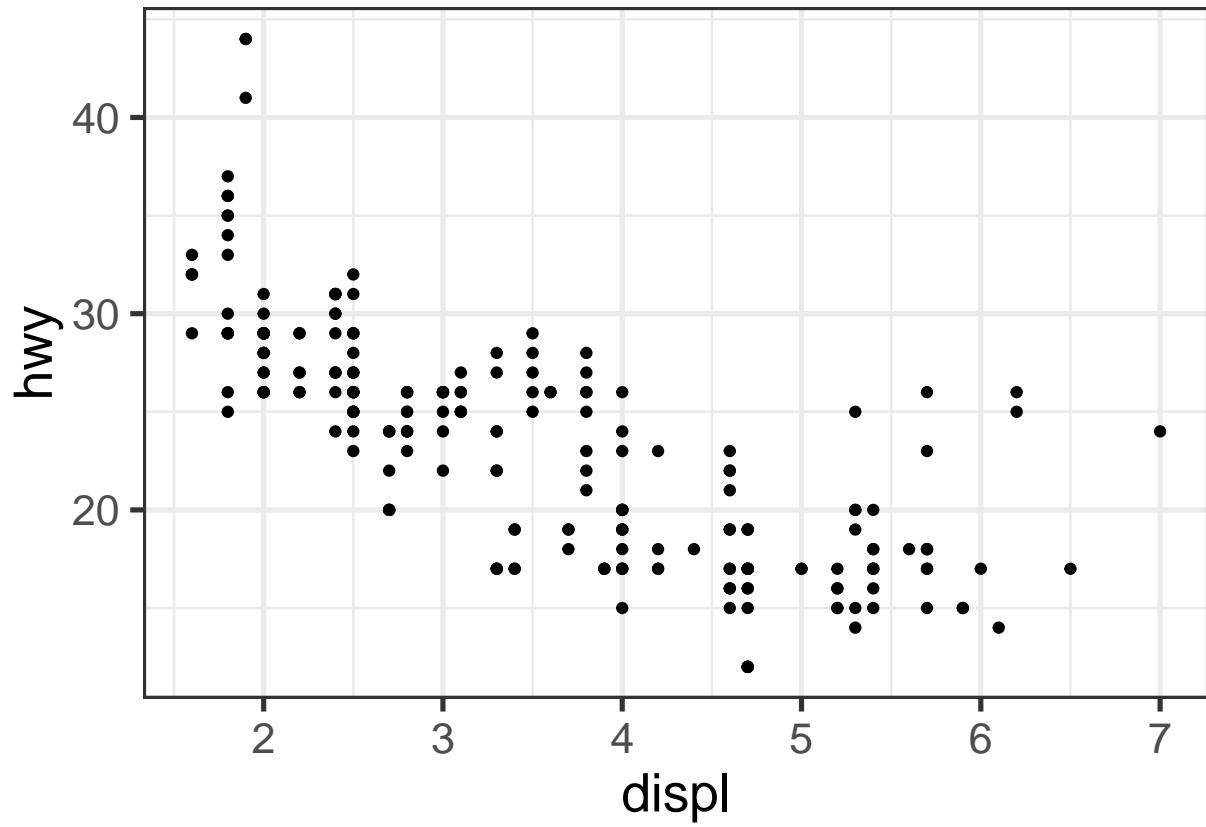
```
ggplot(mpg, aes(x = displ, y = hwy)) +
  geom_point()
```



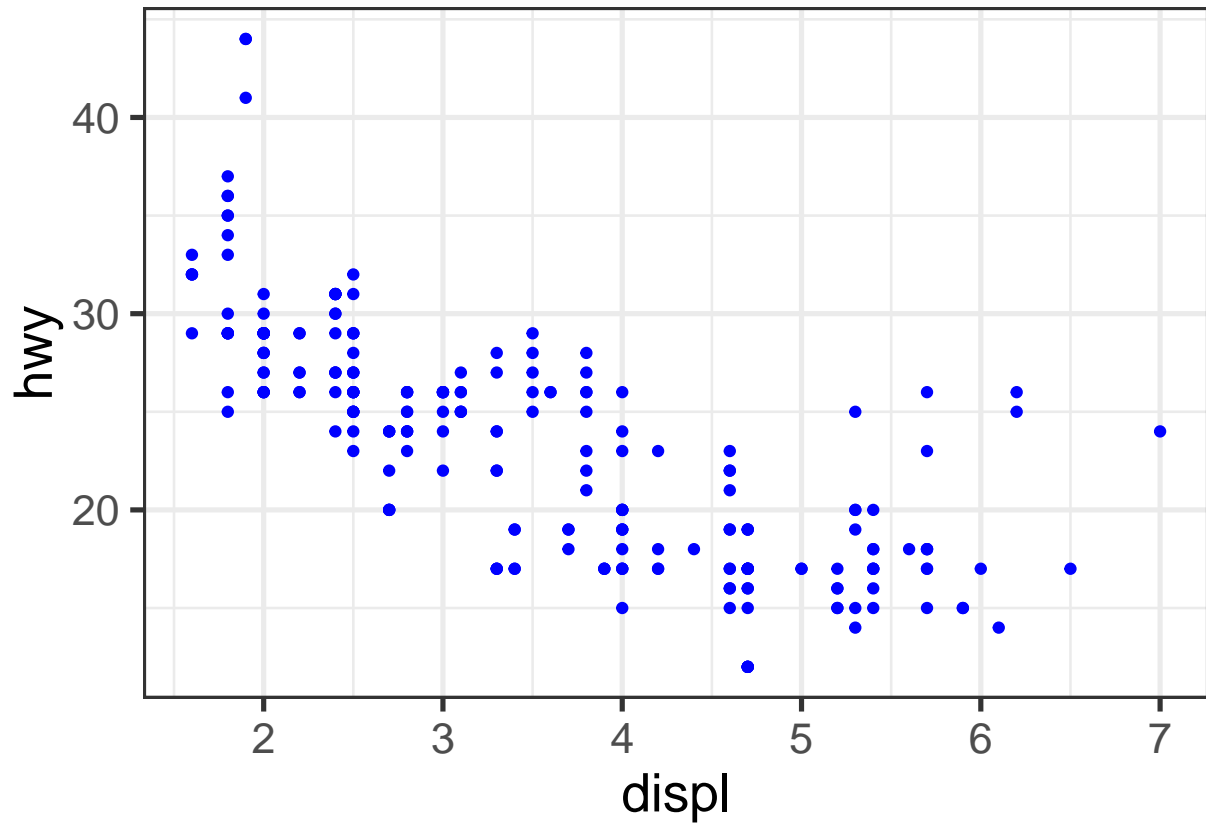
```
ggplot(mpg, aes(x = displ, y = hwy, color = "blue")) +  
  geom_point()
```



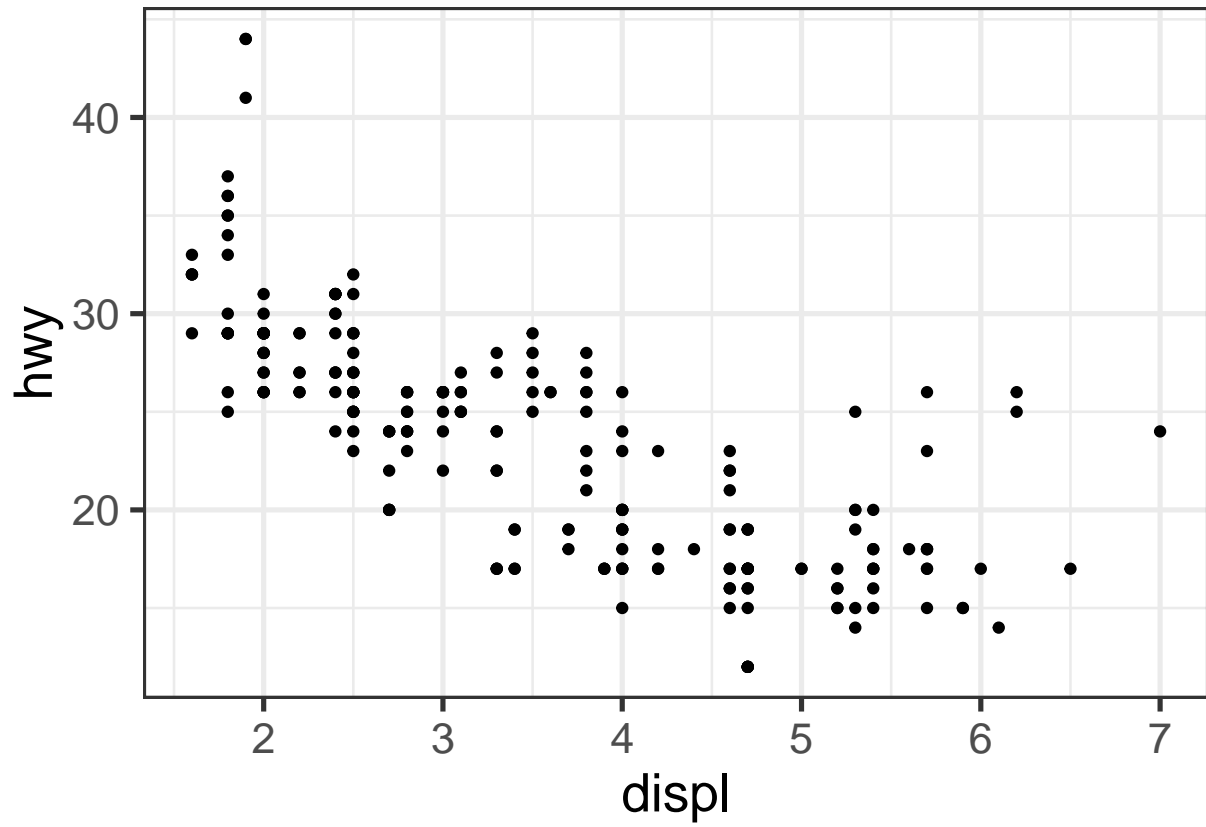
```
ggplot(mpg, aes(x = displ, y = hwy), color = "blue") +  
  geom_point()
```



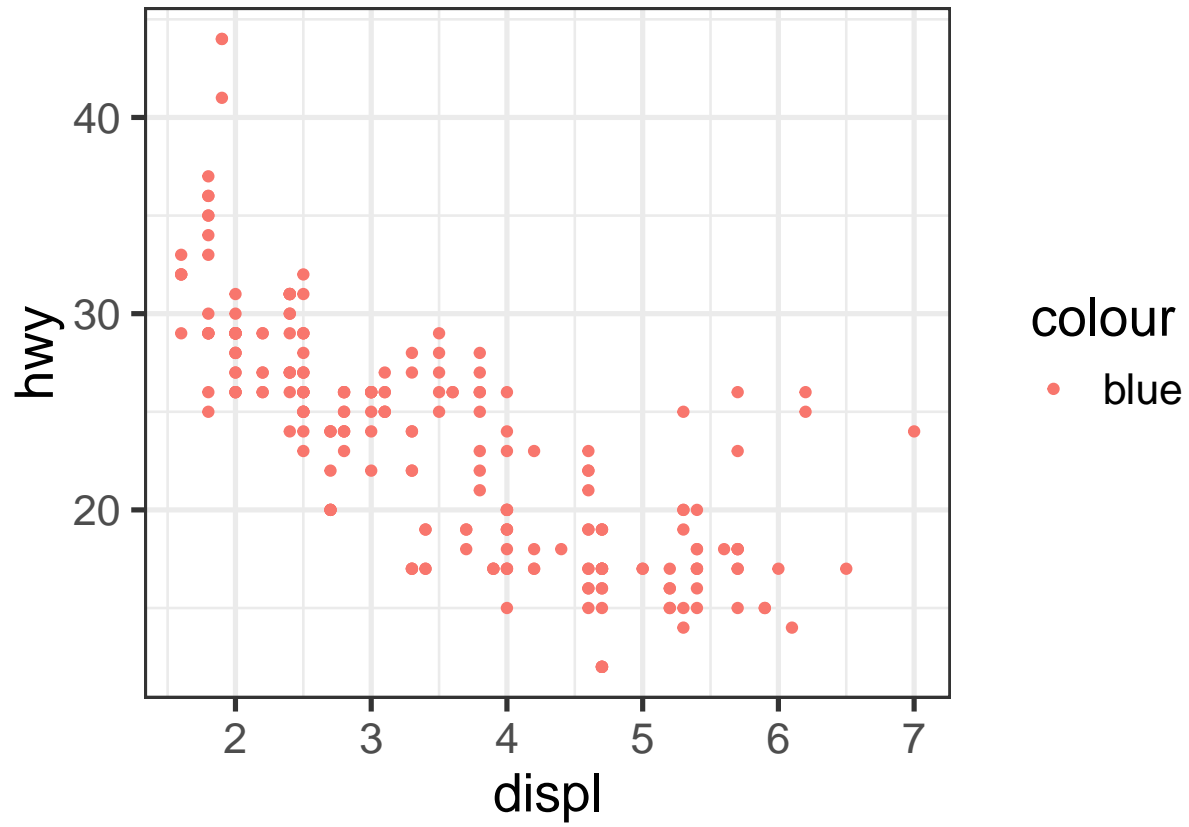
```
ggplot(mpg, aes(x = displ, y = hwy)) +  
  geom_point(color = "blue")
```



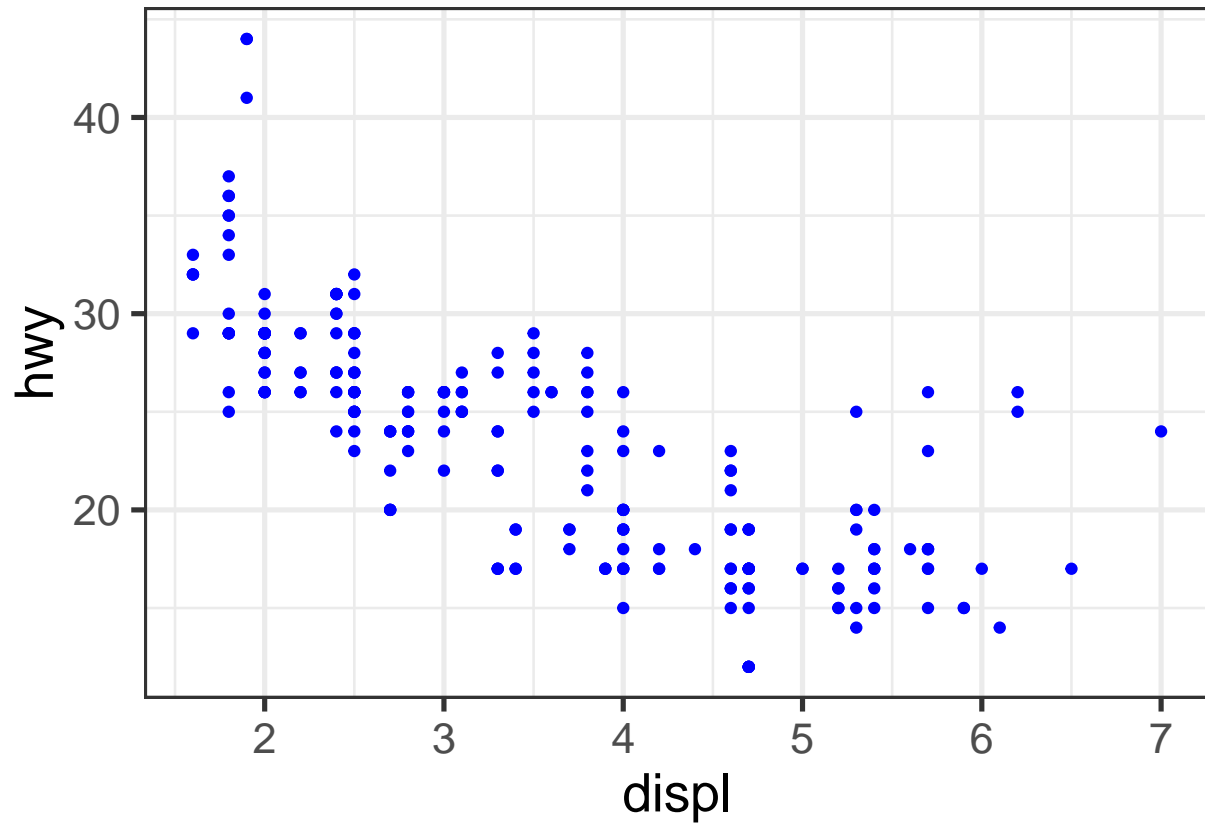
```
ggplot(mpg) +  
  geom_point(aes(x = displ, y = hwy))
```

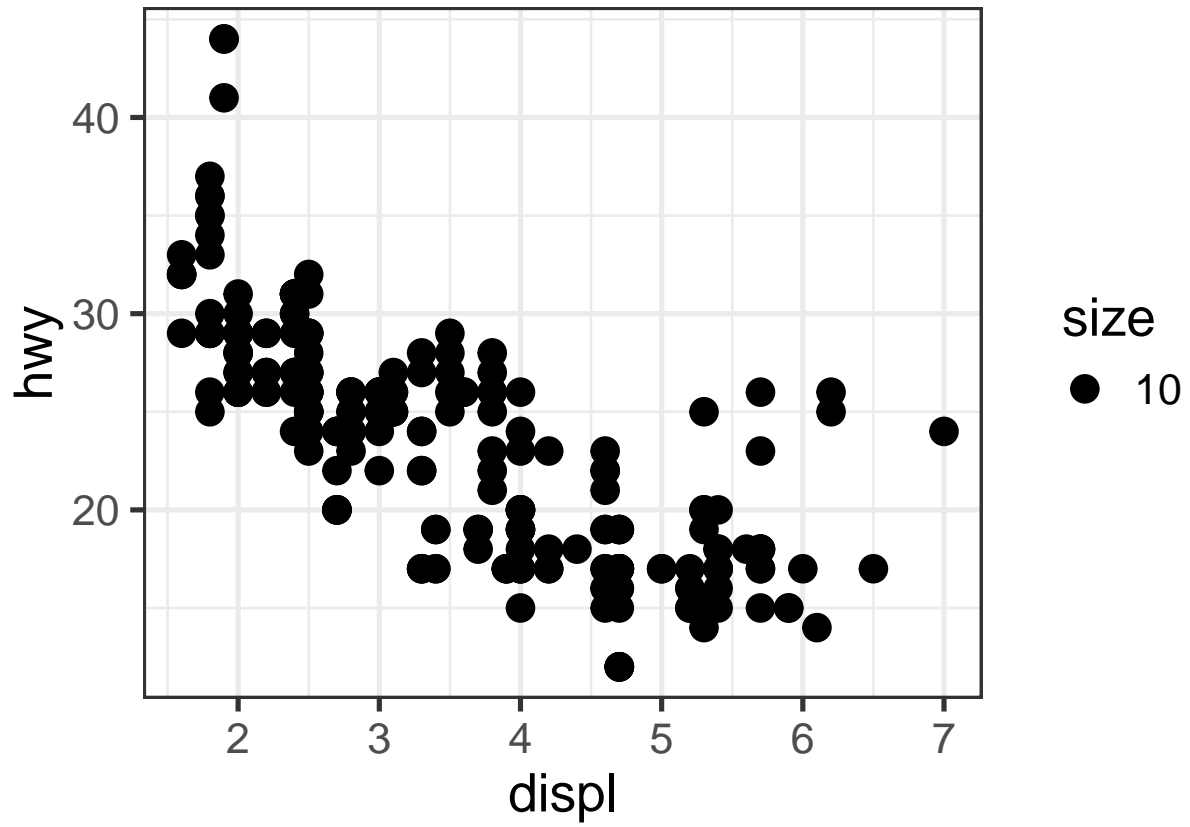
```
ggplot(mpg) +  
  geom_point(aes(x = displ, y = hwy, color = "blue"))
```



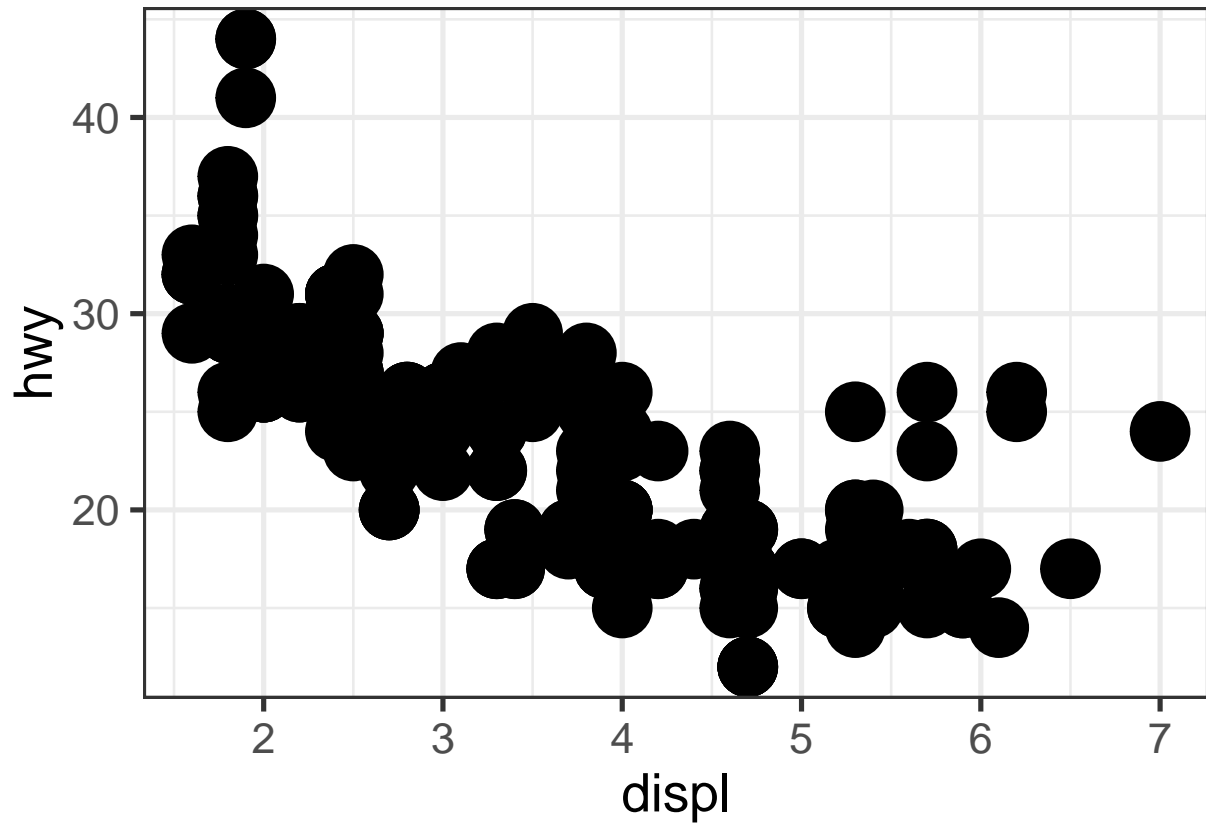
```
ggplot(mpg) +  
  geom_point(aes(x = displ, y = hwy), color = "blue")
```



```
ggplot(mpg) +  
  geom_point(aes(x = displ, y = hwy, size = 10))
```

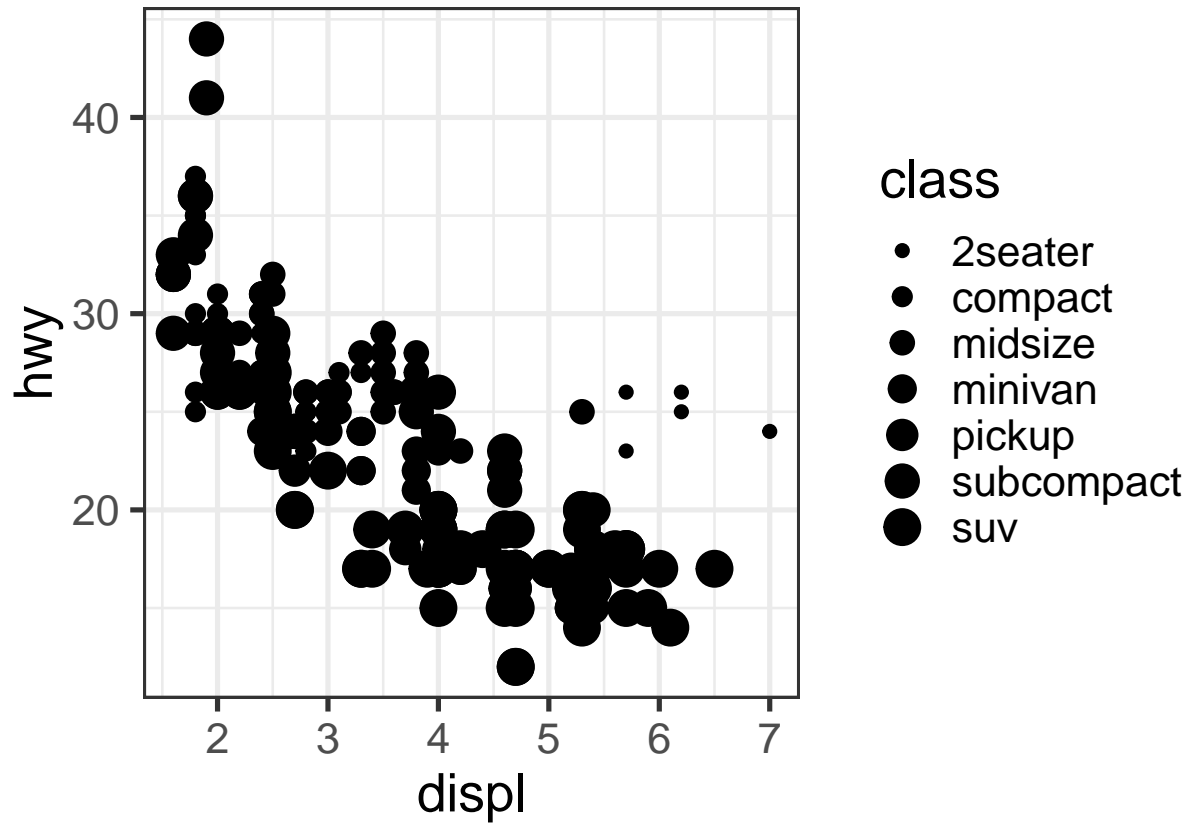


```
ggplot(mpg) +  
  geom_point(aes(x = displ, y = hwy), size = 10)
```

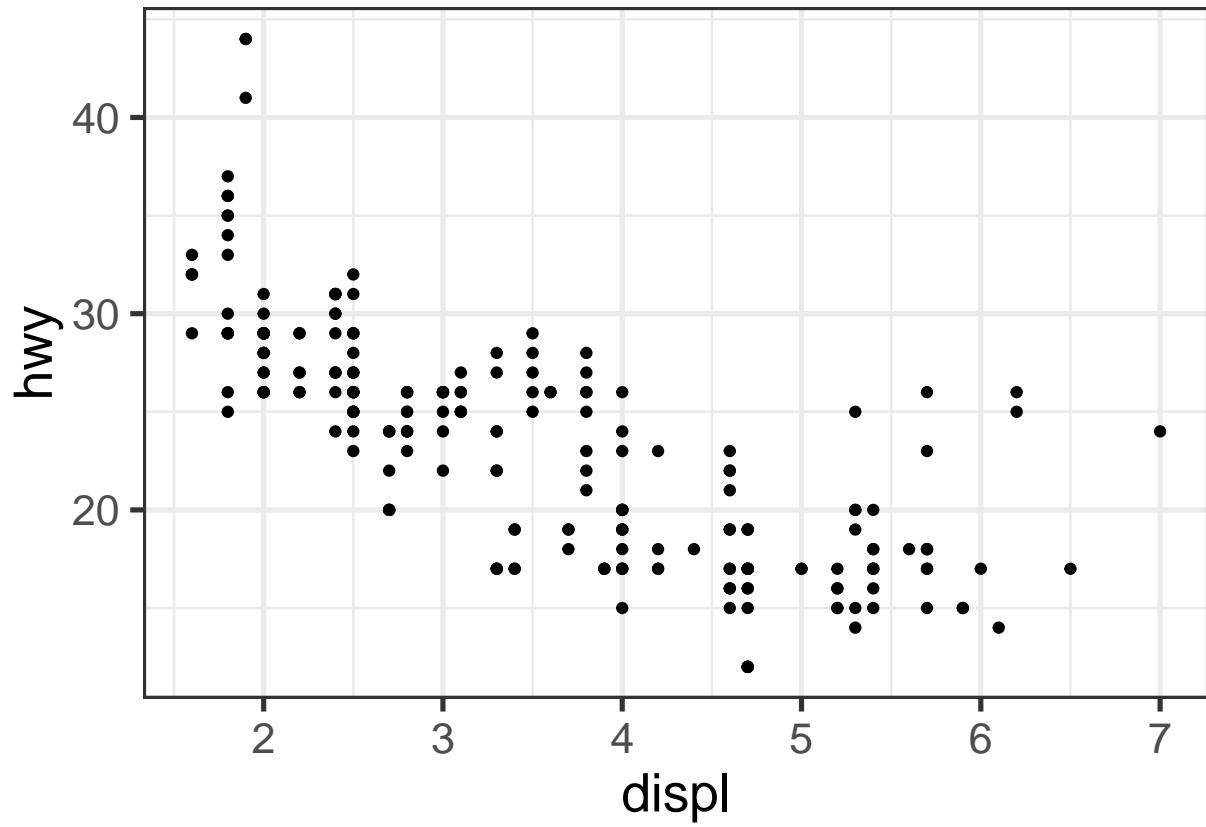


```
ggplot(mpg) +  
  geom_point(aes(x = displ, y = hwy, size = class))
```

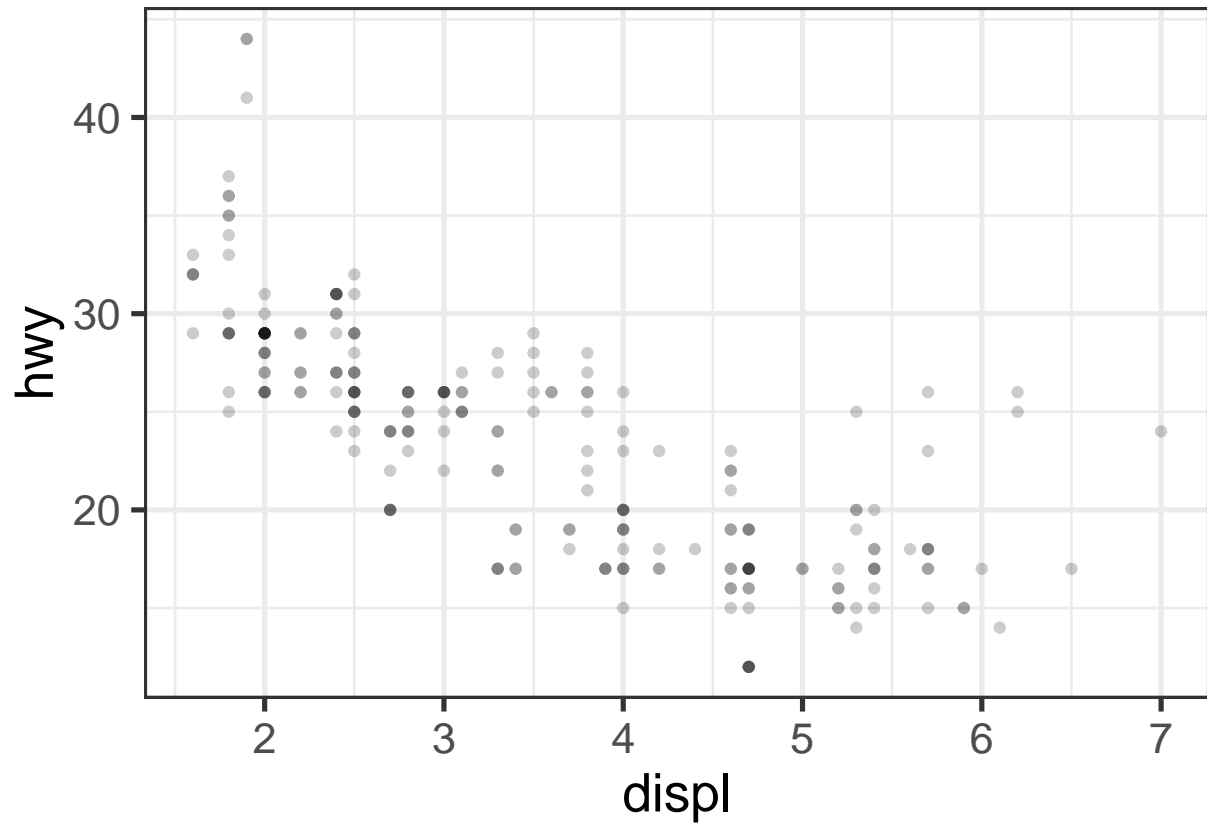
Warning: Using size for a discrete variable is not advised.



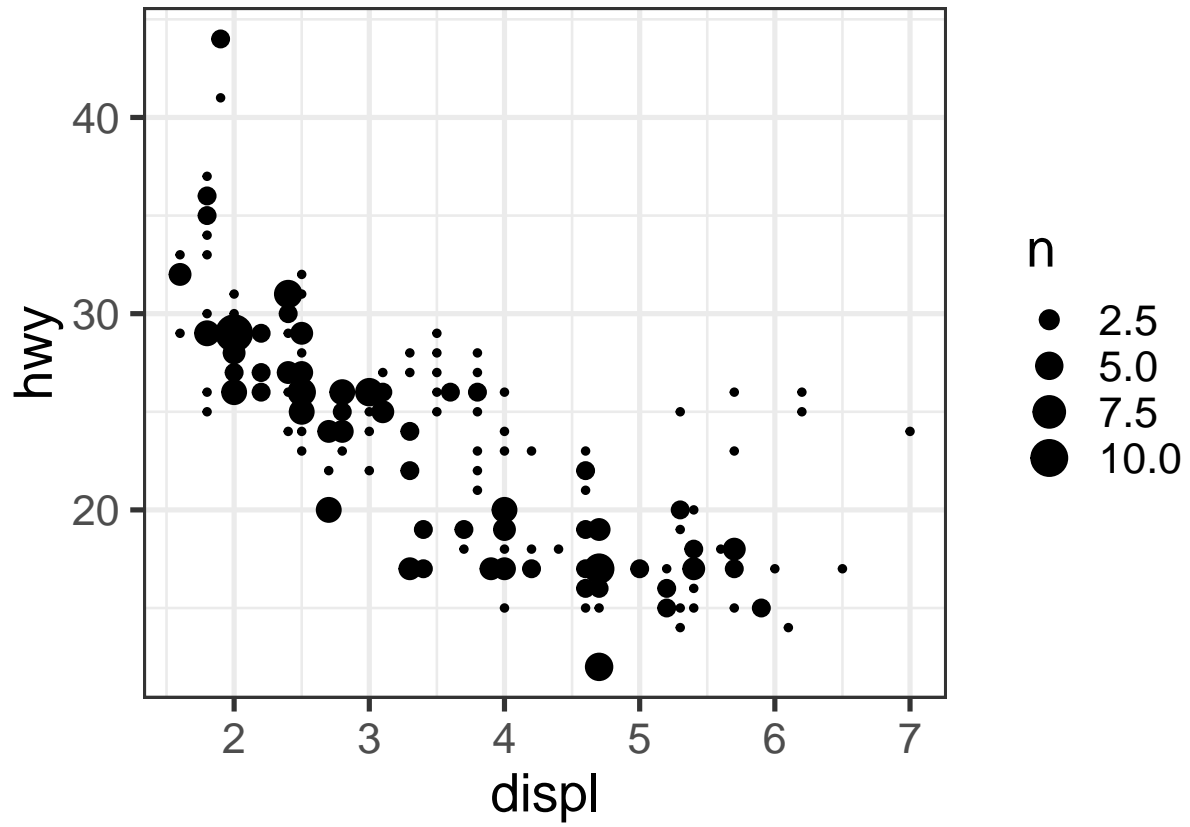
```
ggplot(mpg, aes(x = displ, y = hwy)) +  
  geom_point()
```



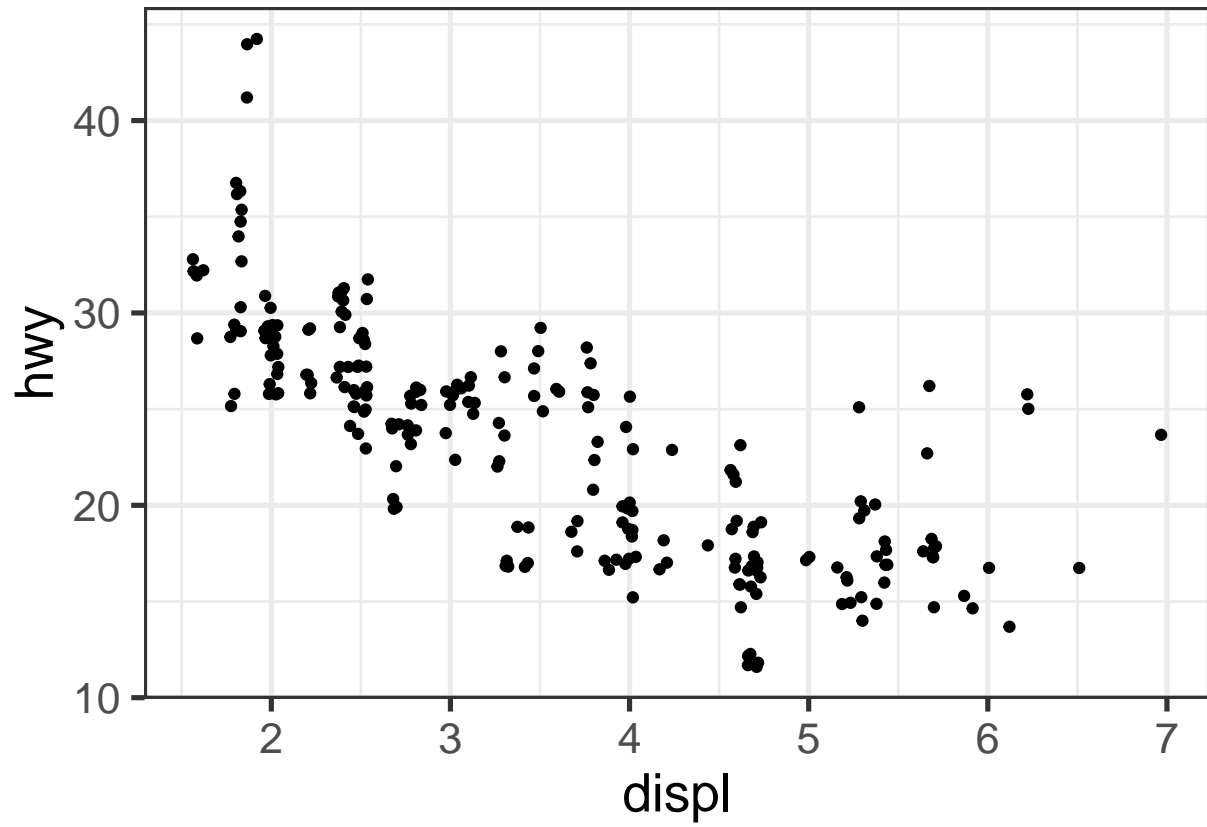
```
ggplot(mpg, aes(x = displ, y = hwy)) +  
  geom_point(alpha = 0.2)
```



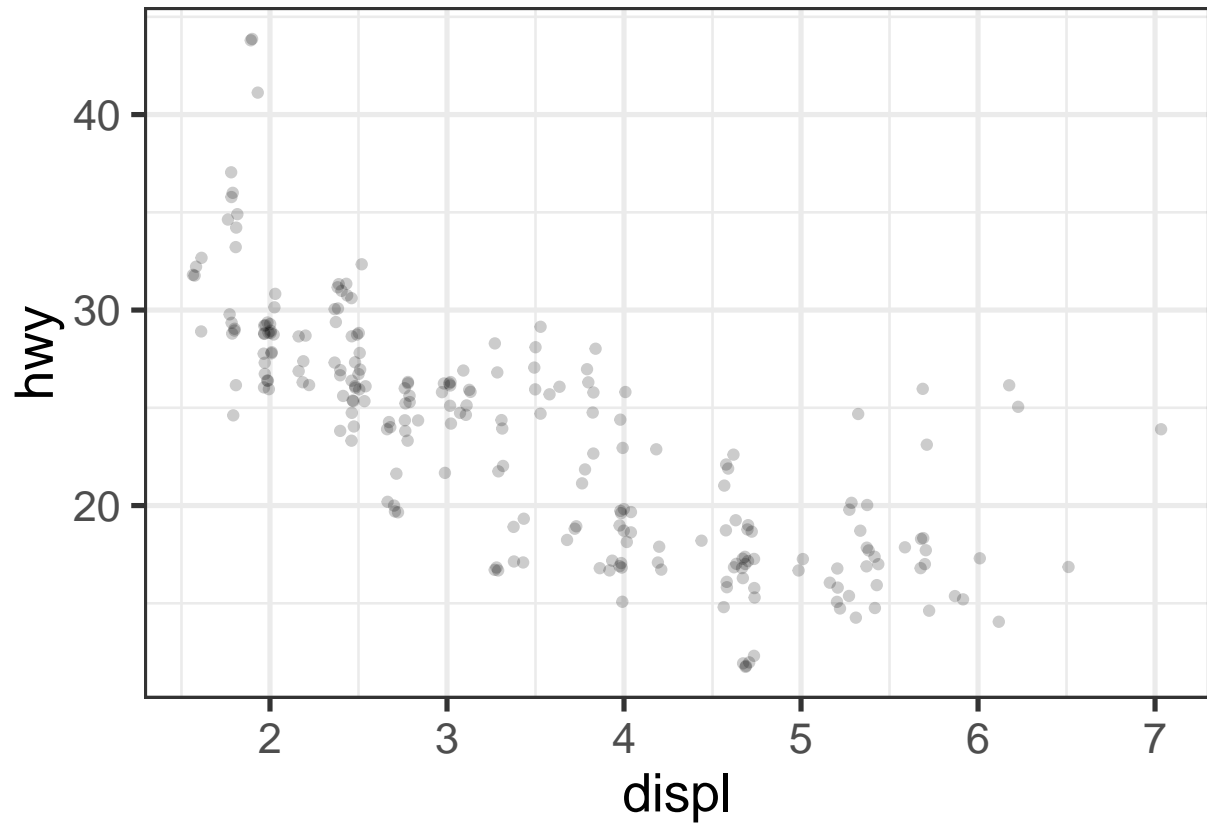
```
ggplot(mpg, aes(x = displ, y = hwy)) +  
  geom_count()
```

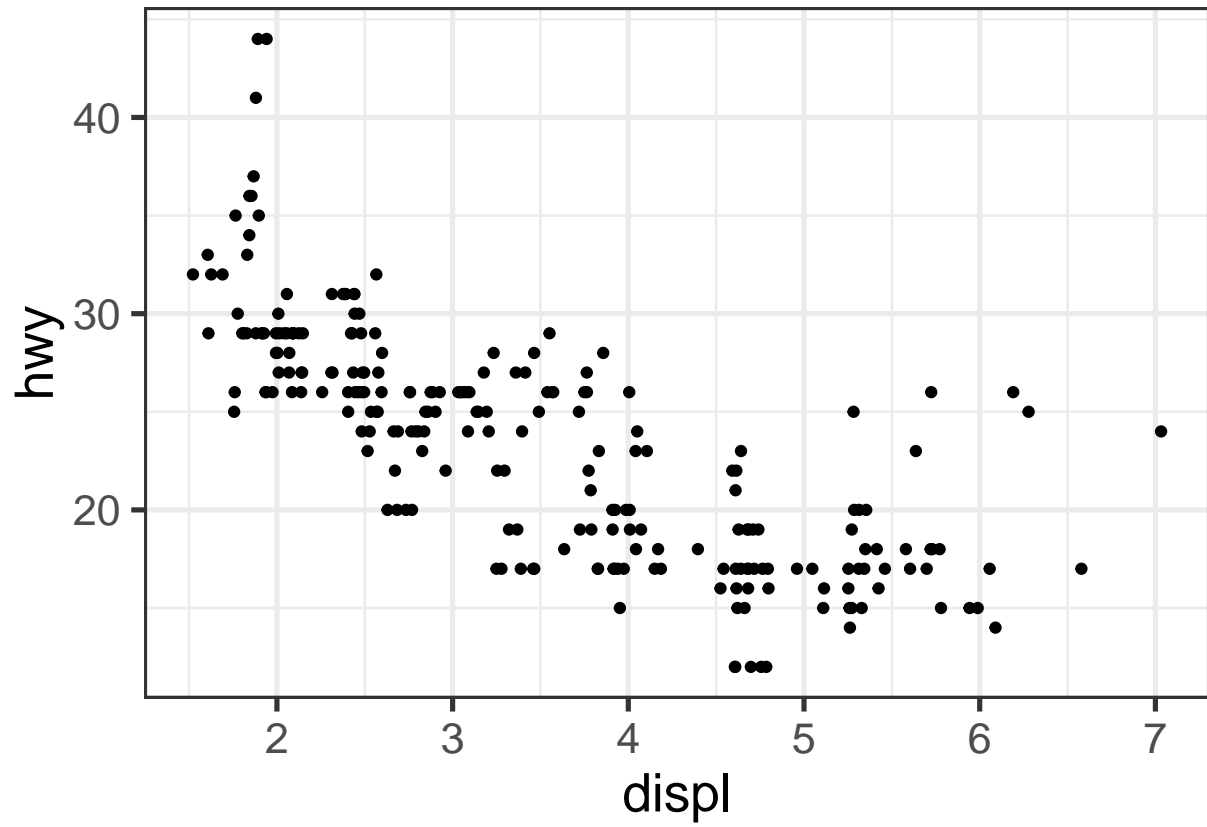
```
ggplot(mpg, aes(x = displ, y = hwy)) +  
  geom_jitter()
```



```
ggplot(mpg, aes(x = displ, y = hwy)) +  
  geom_jitter(alpha = 0.2) # control jitter with width and height parameters
```

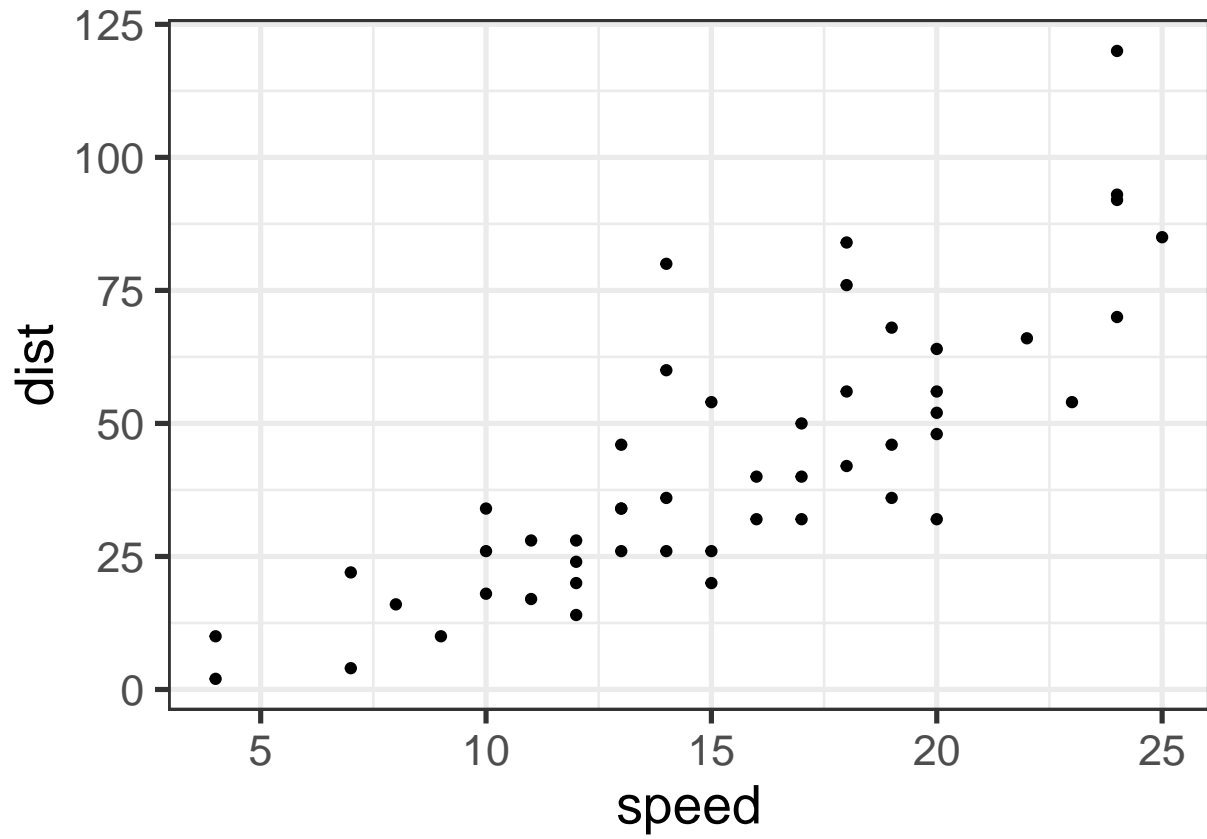


```
ggplot(mpg, aes(x = displ, y = hwy)) +  
  geom_jitter(height = 0, width = 0.1)
```

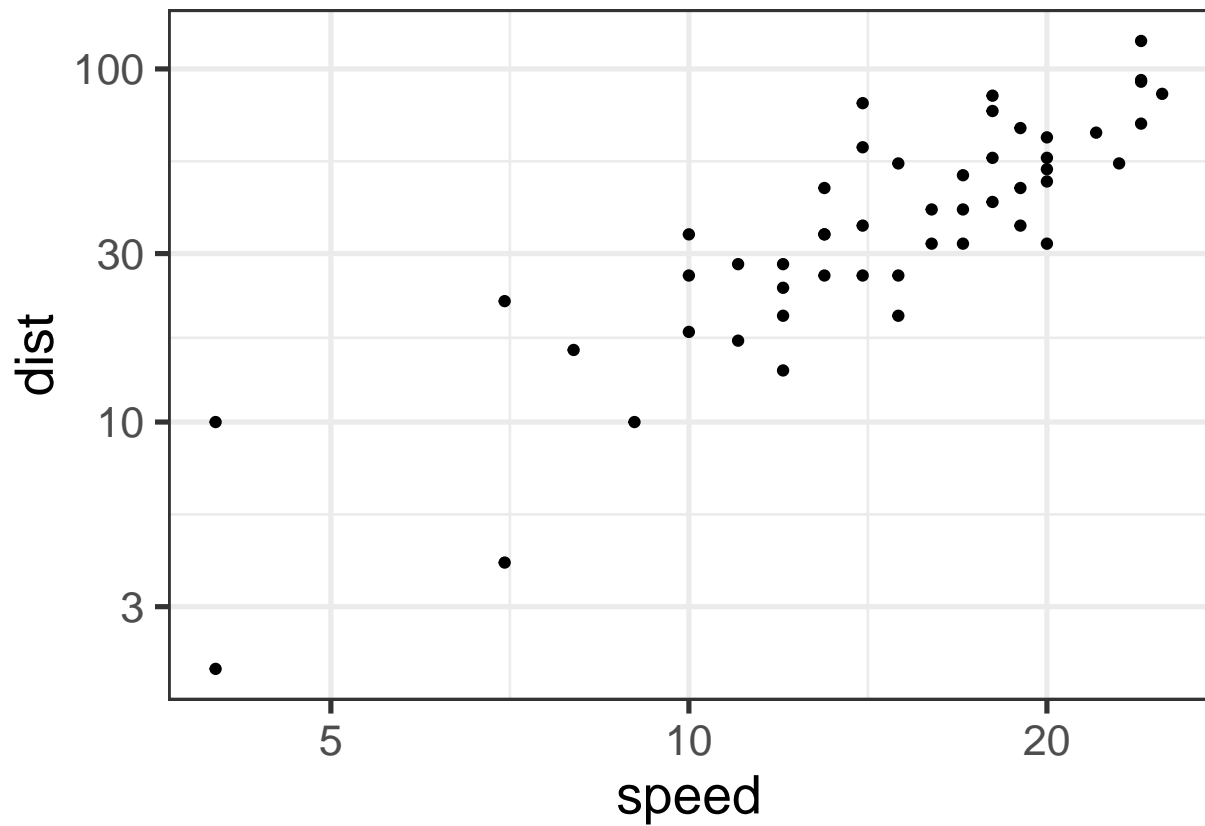


log-scale

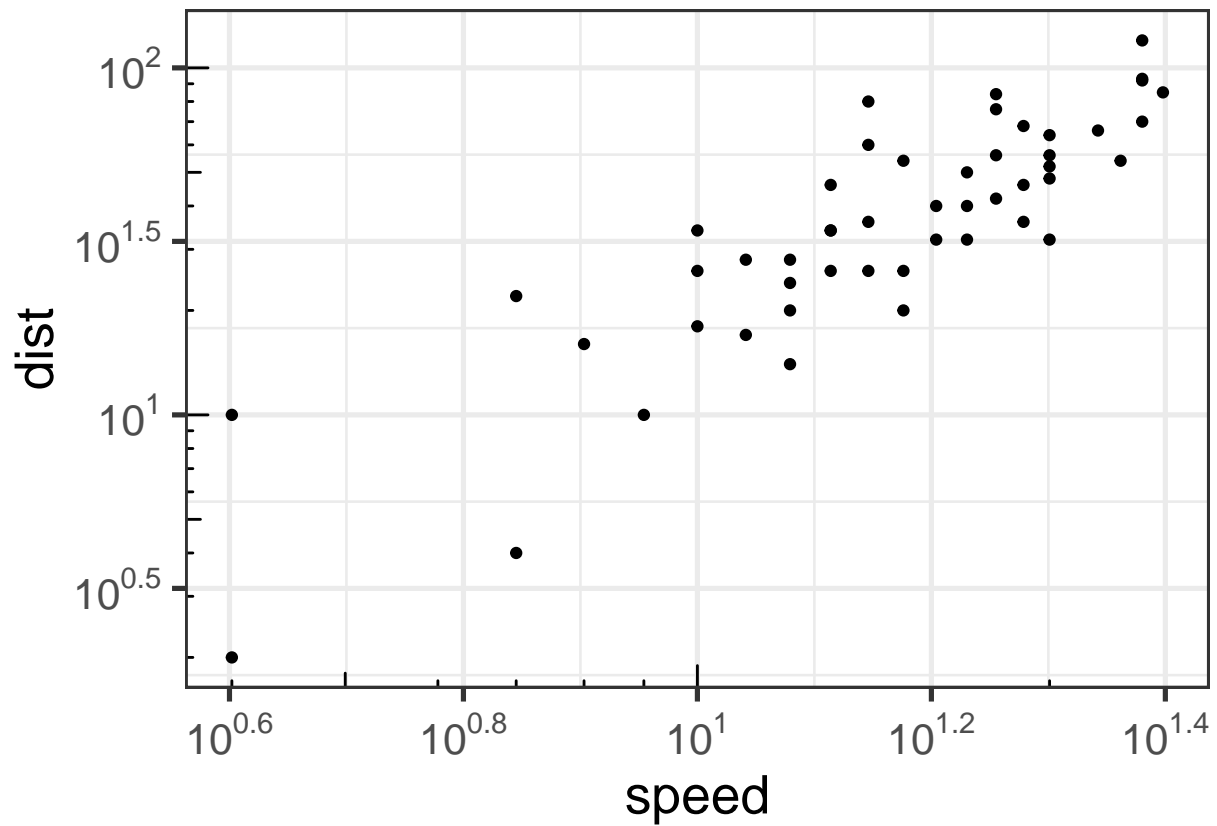
```
ggplot(cars, aes(x = speed, y = dist)) +  
  geom_point()
```



```
ggplot(cars, aes(x = speed, y = dist)) +  
  geom_point() +  
  scale_x_log10() +  
  scale_y_log10()
```



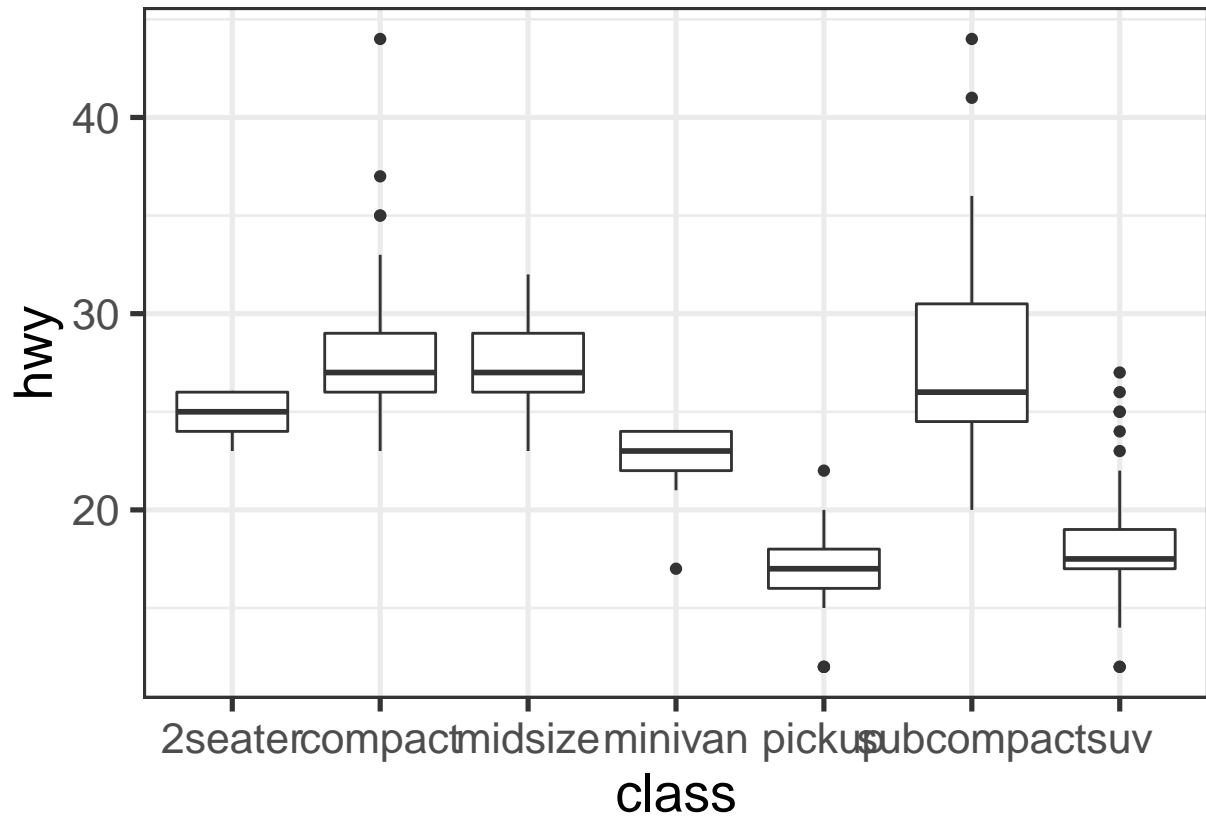
```
library(scales)
ggplot(cars, aes(x = speed, y = dist)) +
  geom_point() +
  scale_x_log10(breaks = trans_breaks("log10", function(x) 10^x),
               labels = trans_format("log10", math_format(10^.x))) +
  scale_y_log10(breaks = trans_breaks("log10", function(x) 10^x),
               labels = trans_format("log10", math_format(10^.x))) +
  annotation_logticks()
```



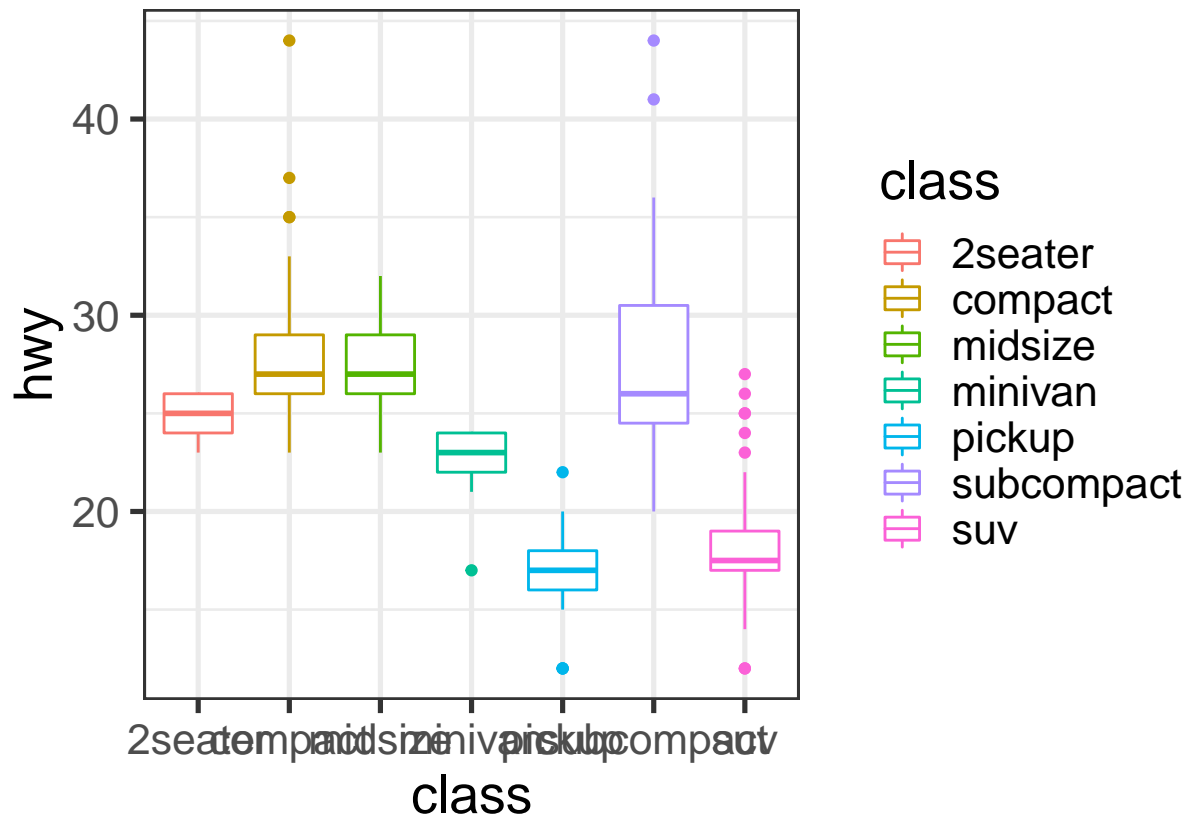
http://ggplot2.tidyverse.org/reference/annotation_logticks.html

Boxplots

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

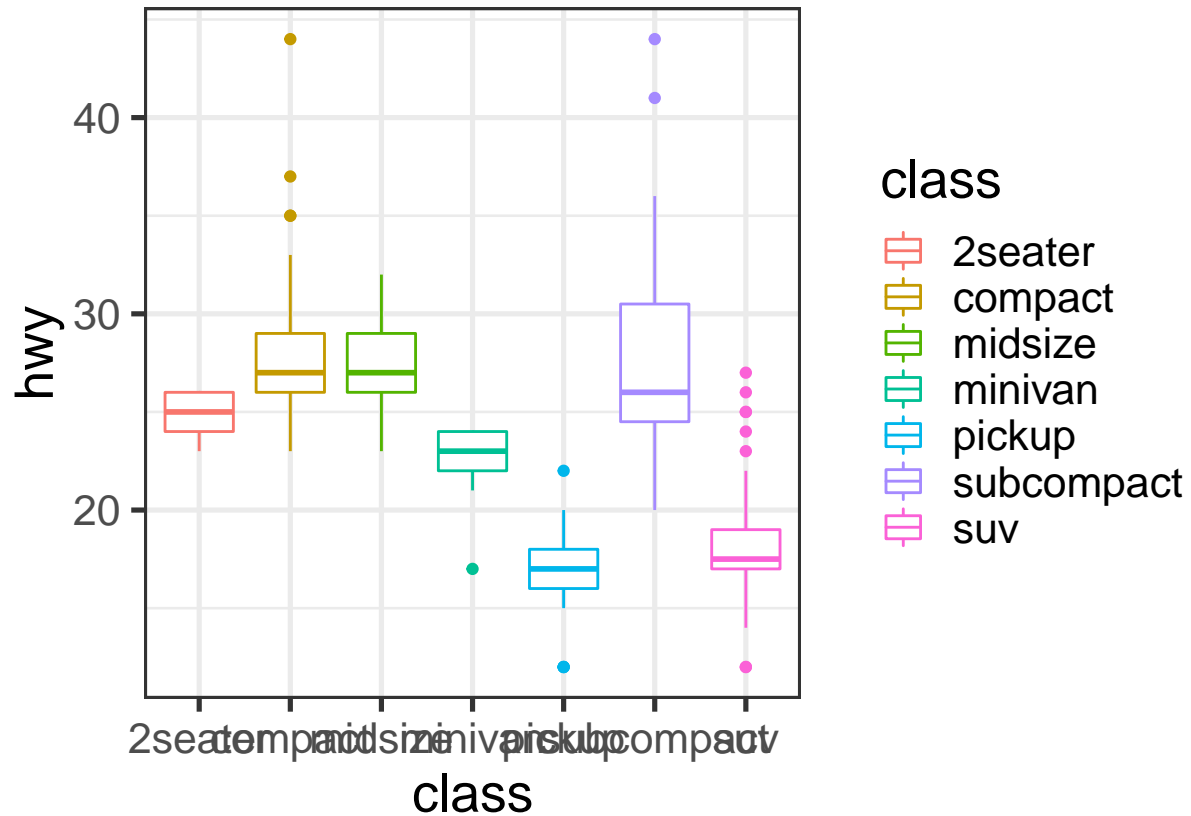


```
ggplot(mpg, aes(x = class, y = hwy, color = class)) +
  geom_boxplot()
```




```
p <- ggplot(mpg, aes(x = class, y = hwy, color = class)) +
  geom_boxplot()
```

p



```
pdf("fig1.pdf")
print(p)
dev.off()
```

```
## pdf
## 2
```

```
ggsave("fig2.pdf", p)
```

```
## Saving 6.5 x 4.5 in image
```

```
ggsave("fig2.pdf", p, width = 8, height = 6)
```

3.8.1 Exercises