Lab4_ggplot2

Zhipeng Chen

```
rm(list = ls()) # clean-up workspace
library("tidyverse")
```

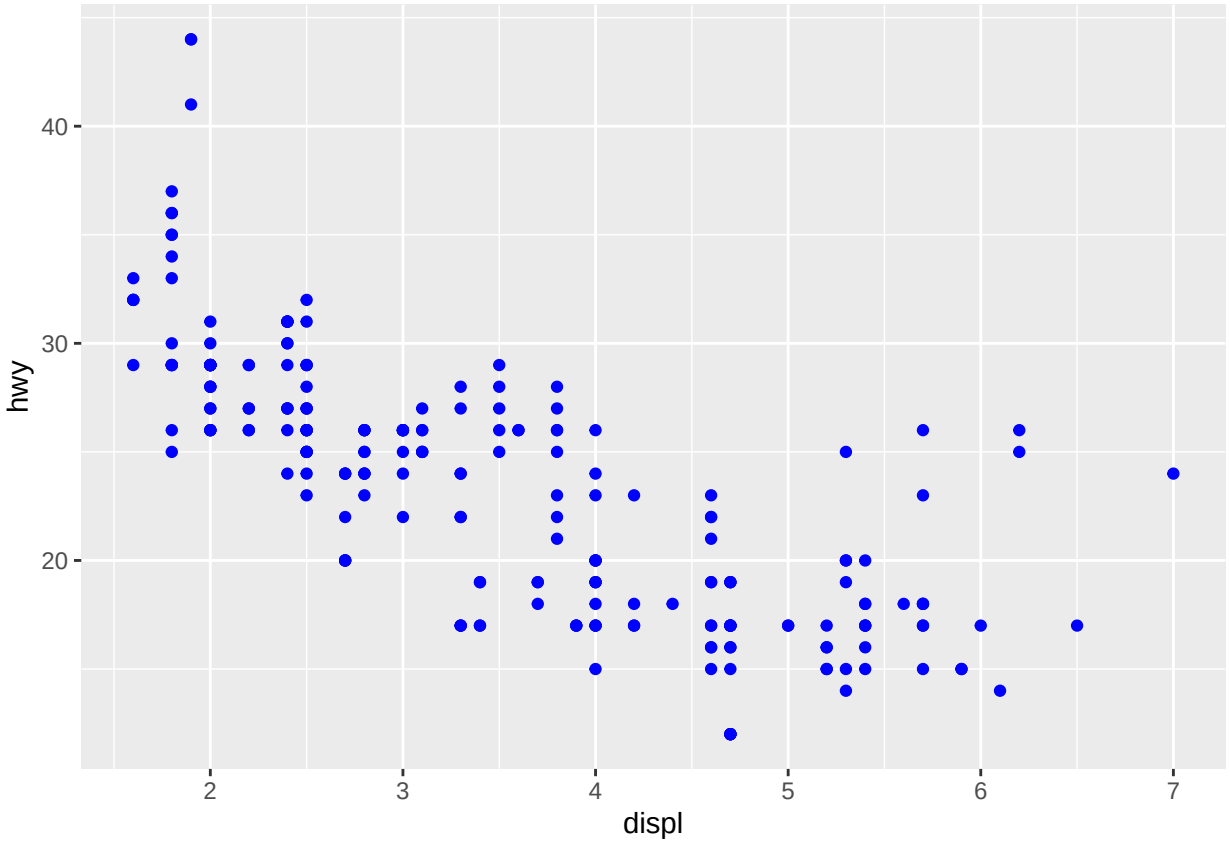
```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.4     v dplyr   1.0.7
## v tidyr   1.1.4     v stringr 1.4.0
## v readr   2.0.1     v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

Some exercise on ggplot2

Aesthetic mappings

Q1. What's gone wrong with this code?

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



The bracket should be placed after "y = hwy,".

Q2. Which variables in `mpg` are categorical? Which variables are continuous? (Hint: type `?mpg` to read the documentation for the dataset). How can you see this information when you run `mpg`?

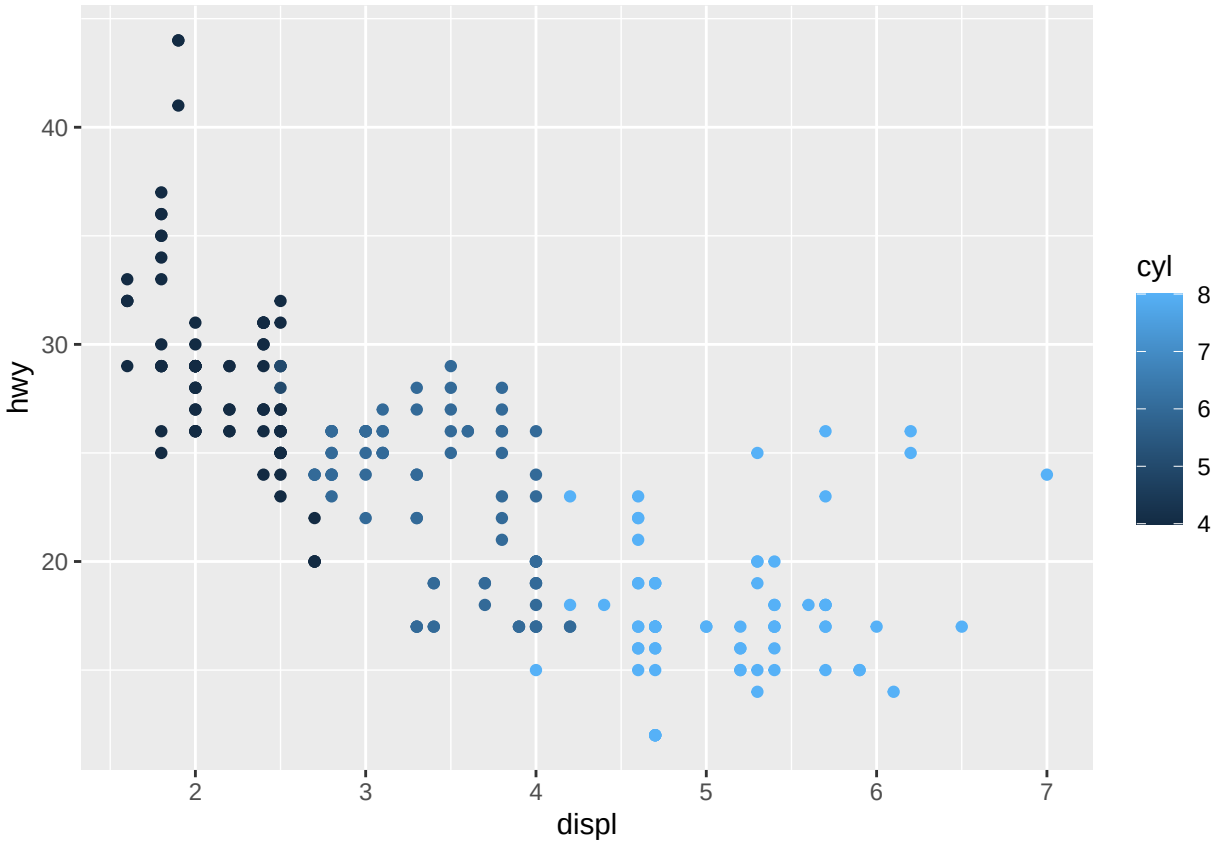
```
summary(mpg)
```

```
##  manufacturer          model               displ           year
##  Length:234         Length:234         Min.   :1.600   Min.   :1999
##  Class :character   Class :character   1st Qu.:2.400   1st Qu.:1999
##  Mode  :character   Mode  :character   Median :3.300   Median :2004
##                                        Mean   :3.472   Mean   :2004
##                                        3rd Qu.:4.600   3rd Qu.:2008
##                                        Max.   :7.000   Max.   :2008
##       cyl           trans               drv                cty
##  Min.   :4.000   Length:234         Length:234         Min.   : 9.00
##  1st Qu.:4.000   Class :character   Class :character   1st Qu.:14.00
##  Median :6.000   Mode  :character   Mode  :character   Median :17.00
##  Mean   :5.889                                         Mean   :16.86
##  3rd Qu.:8.000                                         3rd Qu.:19.00
##  Max.   :8.000                                         Max.   :35.00
##       hwy             fl               class
##  Min.   :12.00   Length:234         Length:234
##  1st Qu.:18.00   Class :character   Class :character
##  Median :24.00   Mode  :character   Mode  :character
##  Mean   :23.44
##  3rd Qu.:27.00
```

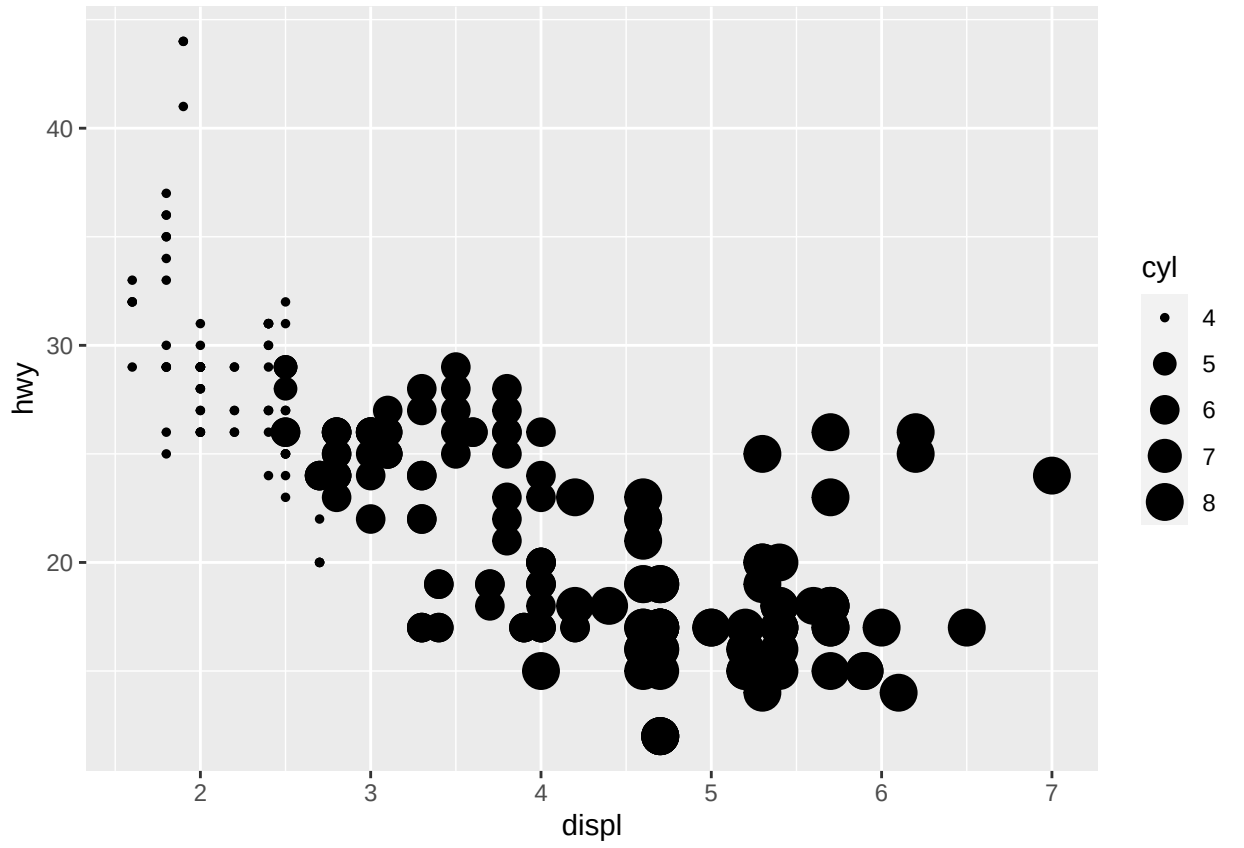
```
##  Max.   :44.00
```

Q3. Map a continuous variable to `color`, `size`, and `shape`. How do these aesthetics behave differently for categorical vs. continuous variables?

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




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




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

A continuous variable can not be mapped to shape.

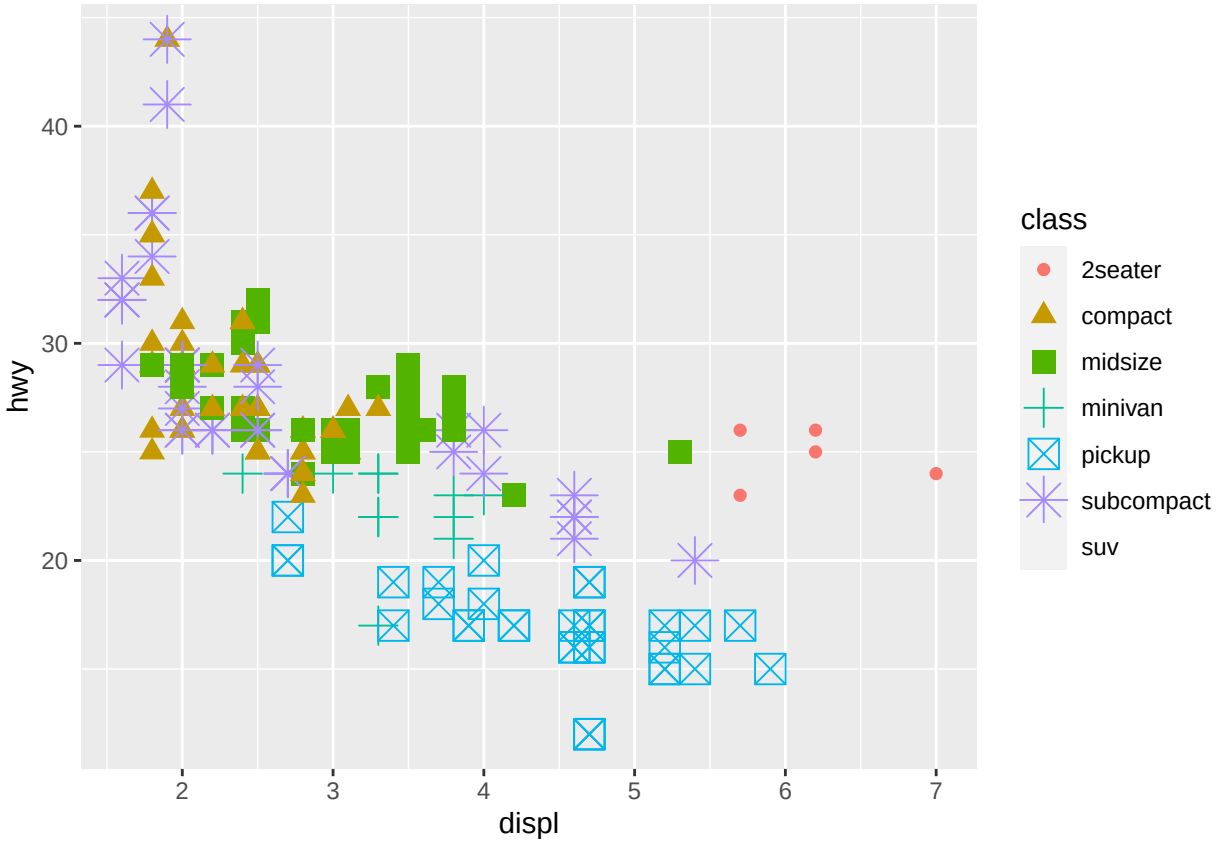
Q4. What happens if you map the same variable to multiple aesthetics?

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

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

```
## Warning: The shape palette can deal with a maximum of 6 discrete values because
## more than 6 becomes difficult to discriminate; you have 7. Consider
## specifying shapes manually if you must have them.
```

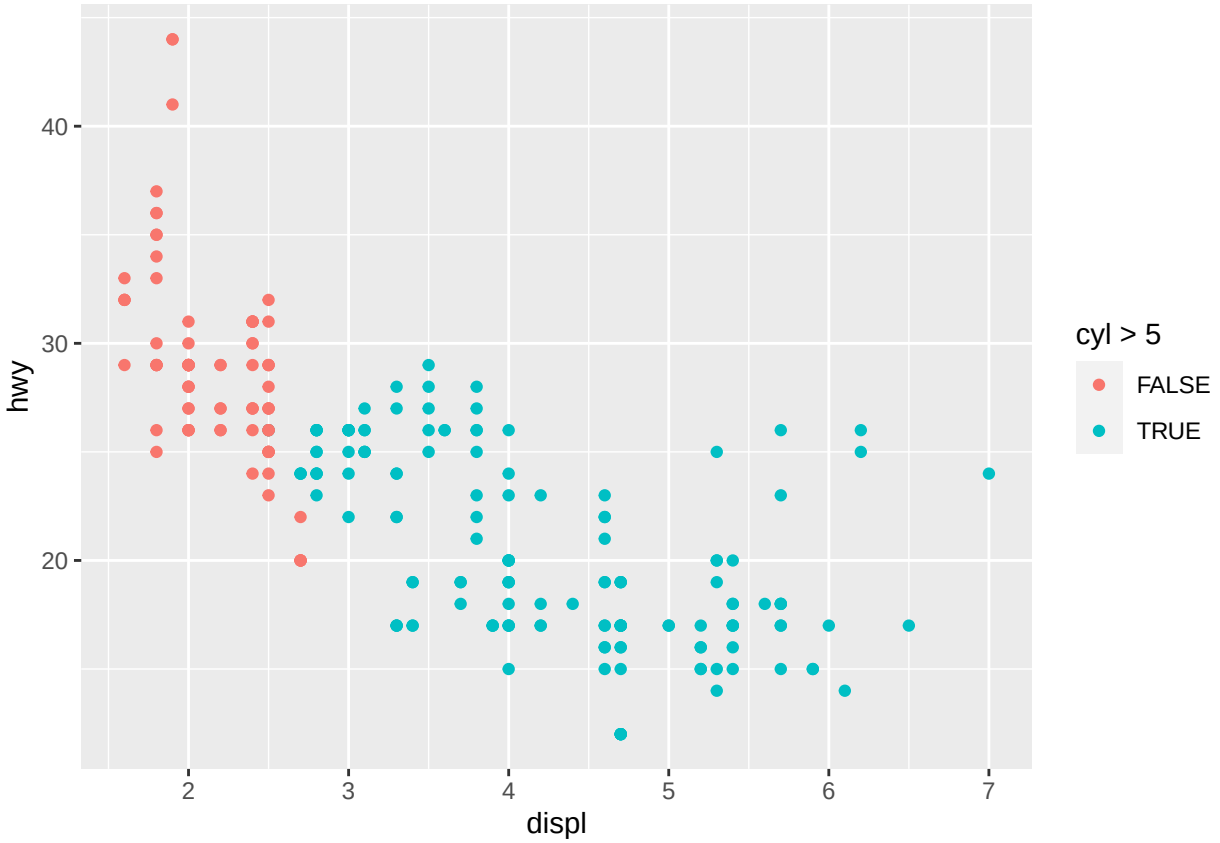
```
## Warning: Removed 62 rows containing missing values (geom_point).
```



The class will be different by all of them.

Q5. What happens if you map an aesthetic to something other than a variable name, like `aes(colour = displ < 5)`? Note, you'll also need to specify `x` and `y`.

```
ggplot(data = mpg) +
  geom_point(mapping = aes(x = displ, y = hwy, colour = cyl > 5))
```



There are different color shown up that differs from “True” or “False” by the condition.

Geometric objects

Q6. What geom would you use to draw a line chart? A boxplot? A histogram? An area chart?
1. geom_smooth 2. geom_boxplot 3. geom_bar 3. geom_polygon

Q7. Will these two graphs look different? Why/why not?

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

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

They are the same.