

Lab 01: Data visualization

Due: Tu, Sep 28 at 11:59pm

Aicha Sidiya - S20106146

27 Feb 2023

Packages

```
library(tidyverse)
my_mpg <- mpg
```

Exercise 1

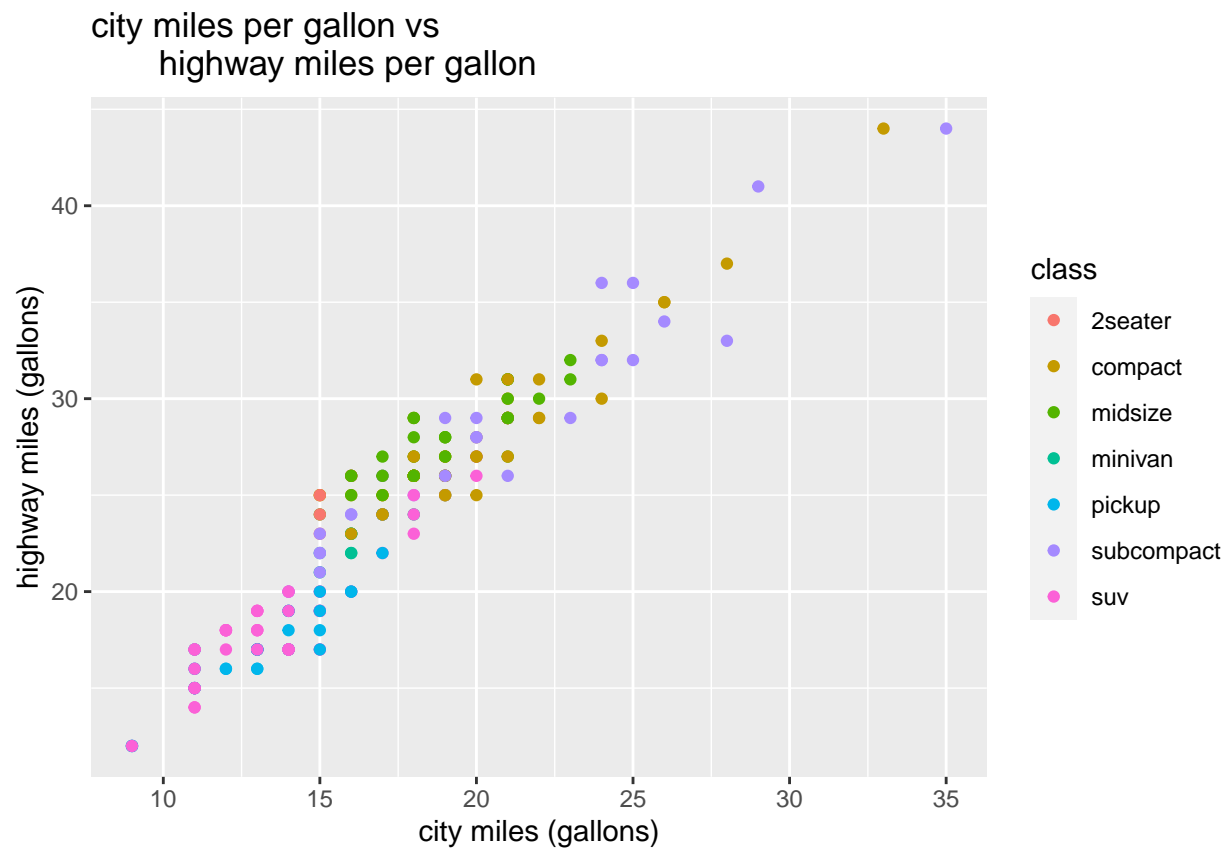
```
glimpse(my_mpg)
```

```
## Rows: 234
## Columns: 11
## $ manufacturer <chr> "audi", "audi", "audi", "audi", "audi", "audi", "audi", "~
## $ model        <chr> "a4", "a4", "a4", "a4", "a4", "a4", "a4", "a4 quattro", "~
## $ displ        <dbl> 1.8, 1.8, 2.0, 2.0, 2.8, 2.8, 3.1, 1.8, 1.8, 2.0, 2.0, 2.~
## $ year         <int> 1999, 1999, 2008, 2008, 1999, 1999, 2008, 1999, 1999, 200~
## $ cyl          <int> 4, 4, 4, 4, 6, 6, 6, 4, 4, 4, 4, 6, 6, 6, 6, 6, 6, 8, 8, ~
## $ trans        <chr> "auto(l5)", "manual(m5)", "manual(m6)", "auto(av)", "auto~
## $ drv          <chr> "f", "f", "f", "f", "f", "f", "f", "f", "4", "4", "4", "4", "4~
## $ cty          <int> 18, 21, 20, 21, 16, 18, 18, 18, 16, 20, 19, 15, 17, 17, 1~
## $ hwy          <int> 29, 29, 31, 30, 26, 26, 27, 26, 25, 28, 27, 25, 25, 25, 2~
## $ fl           <chr> "p", "p", "p", "p", "p", "p", "p", "p", "p", "p", "p", "p", "p~
## $ class        <chr> "compact", "compact", "compact", "compact", "compact", "c~
```

MPG has 234 rows and 11 columns the variables included in the data frame are: manufacturer, model, displacemnt, year, cyl (number of cylinders), trans (transmission type), drv (drive type), cty (city milage per gallon), hwy (highway millage per gallon), fl (fuel type), class.

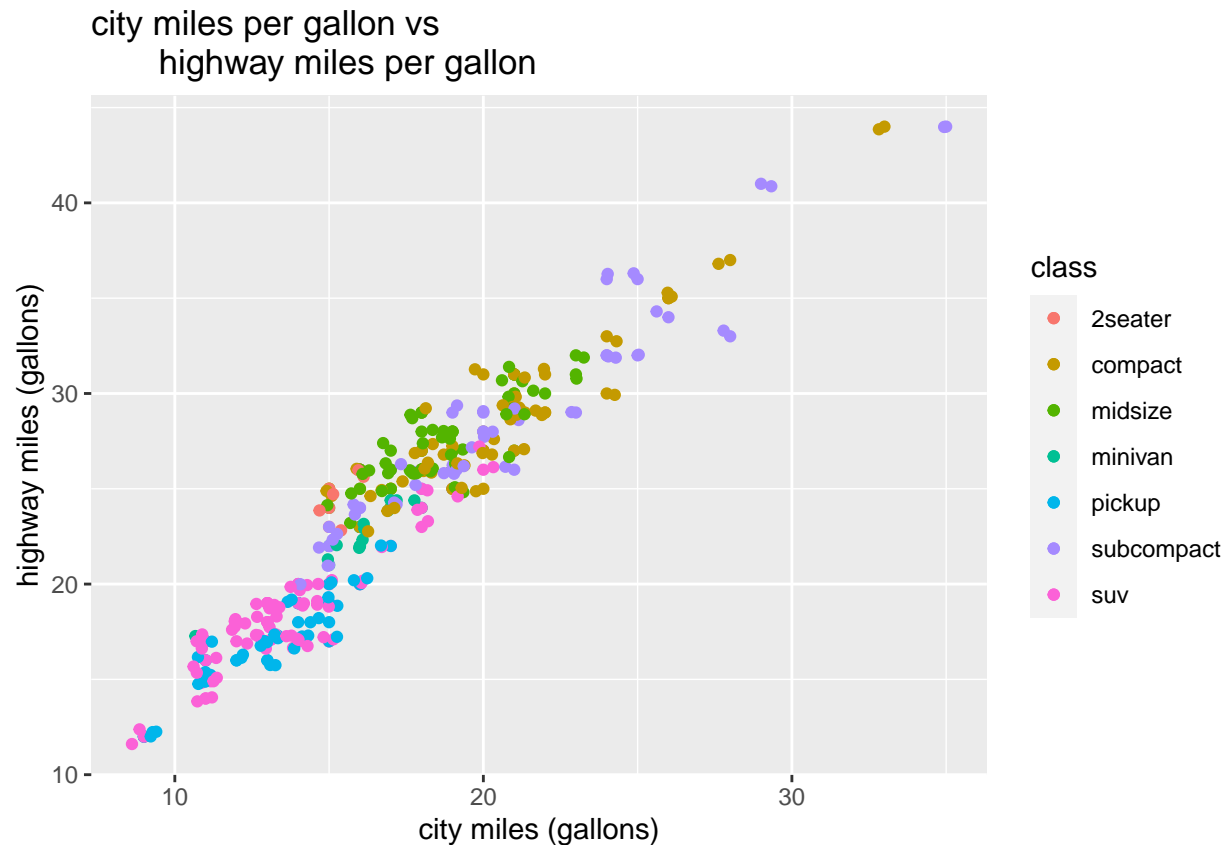
Exercise 2

```
ggplot(data=my_mpg,
       mapping=aes(x=cty, y=hwy,
                   color=class)) +
  geom_point() +
  labs(title = "city miles per gallon vs
             highway miles per gallon",
       x="city miles (gallons)",
       y="highway miles (gallons)")
```



Exercise 3

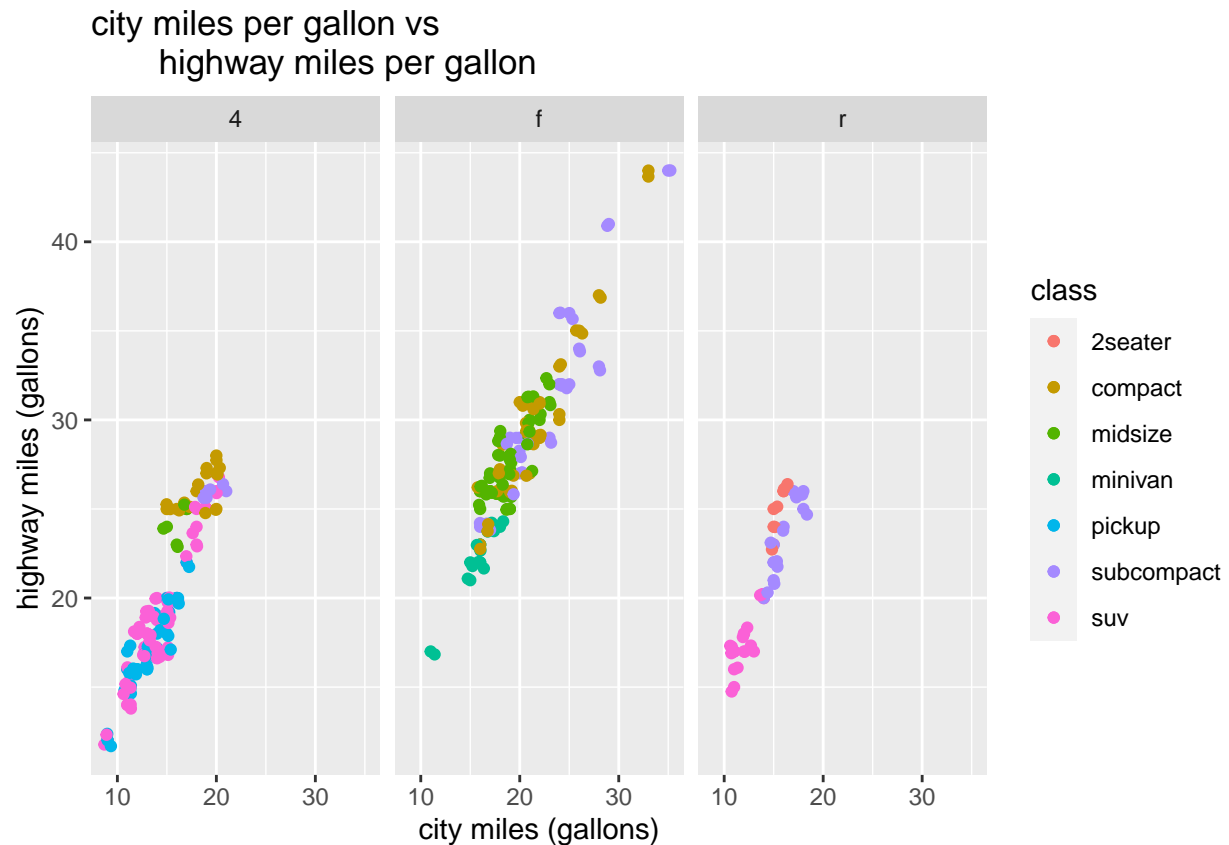
```
ggplot(data=my_mpg,
       mapping=aes(x=cty, y=hwy,
                   color=class)) +
  geom_point() + geom_jitter() +
  labs(title = "city miles per gallon vs
             highway miles per gallon",
       x="city miles (gallons)",
       y="highway miles (gallons)")
```



The first plot done in exercise 2 provided the direct relation with the exact results stored in the dataframe, which helps in understanding the data we have in our dataframe. In exercise 3 we added some noise to the data. By adding noise we were able to visualize all point because in exercise 2 some points were on top of each other. However, this plot does not capture the exact values as in exercise 2.

Exercise 4

```
ggplot(data=my_mpg,
       mapping=aes(x=cty, y=hwy,
                    color=class)) +
  geom_point() + geom_jitter() +
  labs(title = "city miles per gallon vs
               highway miles per gallon ",
       x="city miles (gallons)",
       y="highway miles (gallons)") +
  facet_wrap(~drv)
```

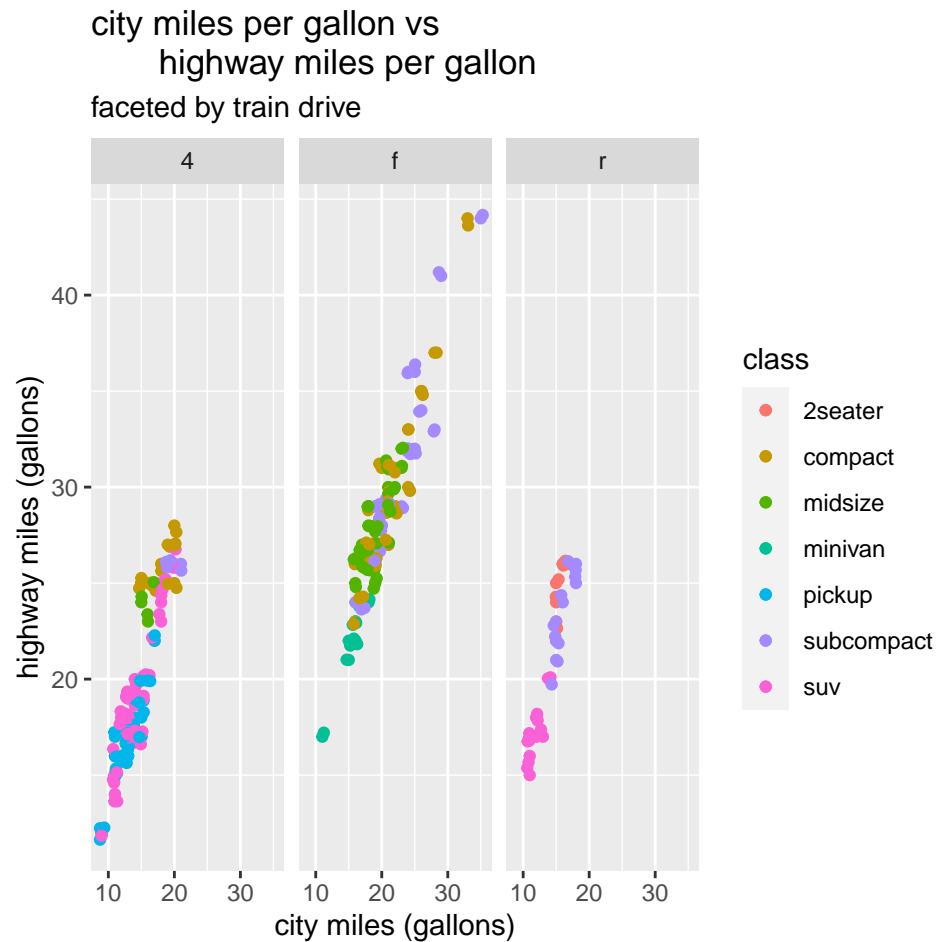


Comment here on what you notice about the relationship between class and city miles per gallon.

The suv, subcompact, pickup, 2seater tend to run between 10 to 15 miles per gallon in the city, while the minivan, midsize, and compact tend to run between 15 to 25 miles per gallon in the city.

Exercise 5

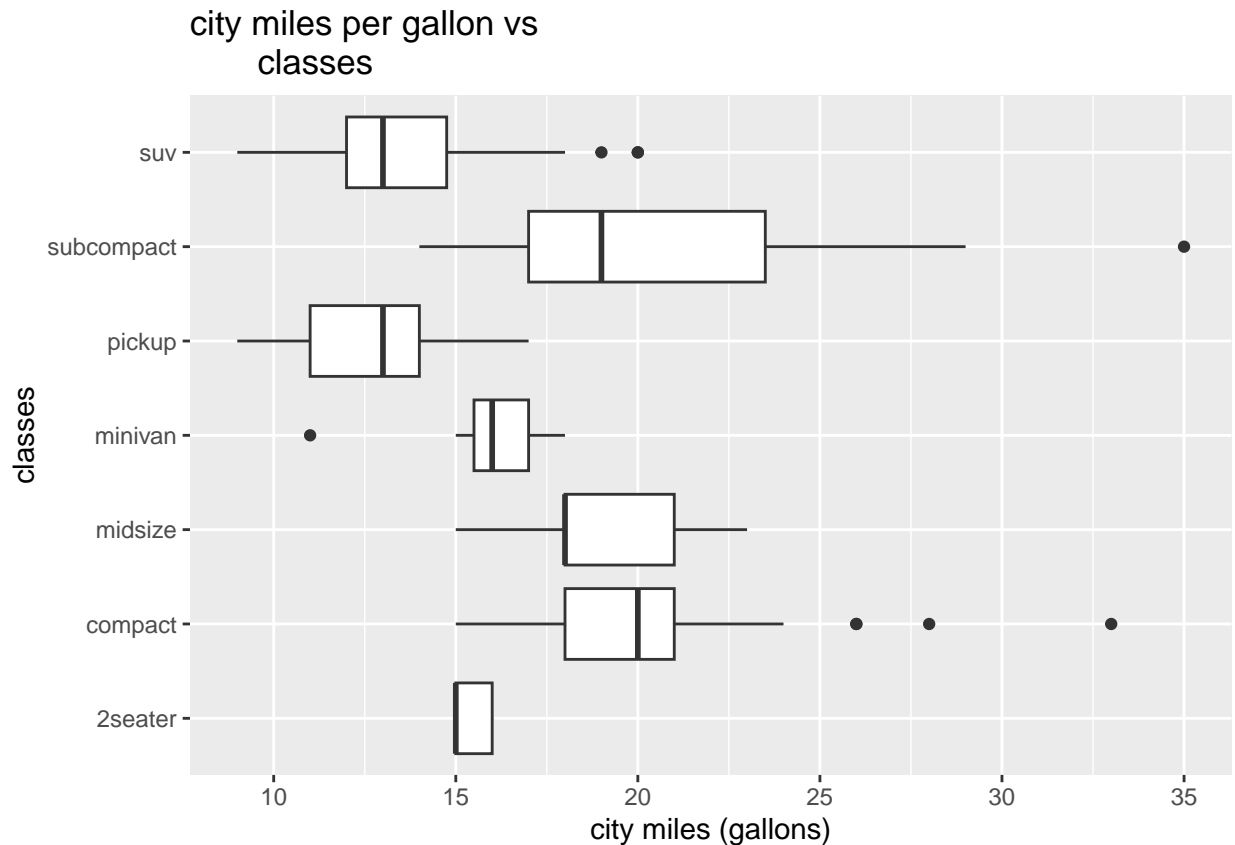
```
ggplot(data=my_mpg,
       mapping=aes(x=cty, y=hwy,
                   color=class)) +
  geom_point() + geom_jitter() +
  labs(title = "city miles per gallon vs
             highway miles per gallon",
       subtitle = "faceted by train drive",
       x="city miles (gallons)",
       y="highway miles (gallons)") +
  facet_wrap(~drv)
```



Upon faceting the point graph by train drive we notice that the pickup mostly use a 4 wheel drive, while the suv uses both a 4 wheel and rear wheel drive. The majority of midsize class use a front wheel drive, along with minivans, subcompact, and compact. Some compact use 4 wheels, and some subcompact use a rear wheel drive.

Exercise 6

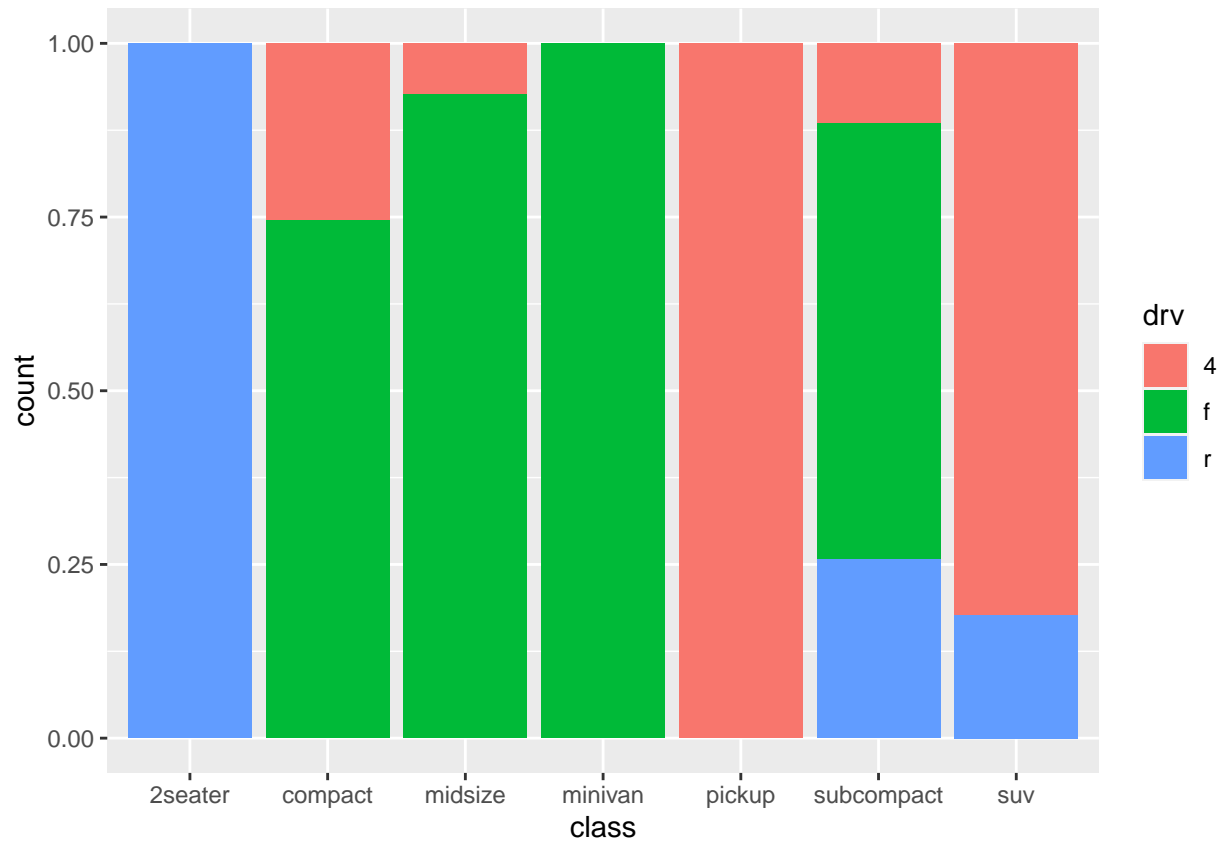
```
ggplot(data=my_mpg, mapping=aes(x=cty, y=class)) +
  geom_boxplot() +
  labs(title = "city miles per gallon vs
    classes",
    x="city miles (gallons)",
    y="classes")
```



The 2easter and minivan city mileages are clustered between 15 ti 17 miles per gallon. The mileage for the suv and pickup is also fairly similar ranging from 12 to 14 miles per gallon. Similarly, the midsize nad compact have a mileage ranging from 18 to 21 miles per gallon. The subcompact mileage is the most spread out ranging from 14 to 24 miles per gallon with a midean of 18 miles per gallon.

Exercise 7

```
ggplot(data=my_mpg,
       mapping=aes(x=class, fill=drv)) +
  geom_bar(position = "fill")
```



What do you notice from the bar chart above?

The 2seater uses a rear wheel drive 100% of the time, the minivan uses the front wheel drive 100% of the time, the pickup uses the 4 wheel drive 100% of the time. The subcompact uses all types of drives the front wheel drive 50% of the time. The compact, midsize, and suv use two types of drive with one dominating type at approximately 75 to 90% of the time. The compact and midsize mostly use the front wheel drive. The suv uses the rear wheel drive only 25% of the time. All three use the 4 wheel drive.

Exercise 8

```
ggplot(data=my_mpg,
       mapping=aes(x=displ, y=hwy)) +
  geom_point(size=0.50) +
  labs(title = "Highway mpg vs displacement",
       subtitle = "faceted by class",
       x="Displacement (litres)",
       y="Highway mpg") +
  facet_wrap(~class, nrow=2) + theme_bw()
```

Highway mpg vs displacement
 faceted by class

