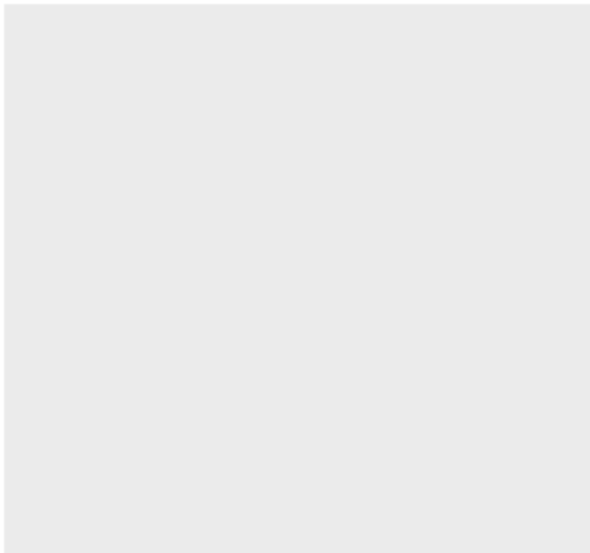**STOR 320**
Homework 1
Amanda Liu (PID:730042603)

Section 3.2.4: Exercises 1-5

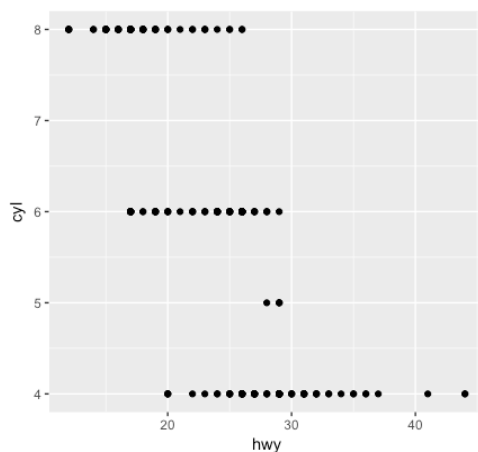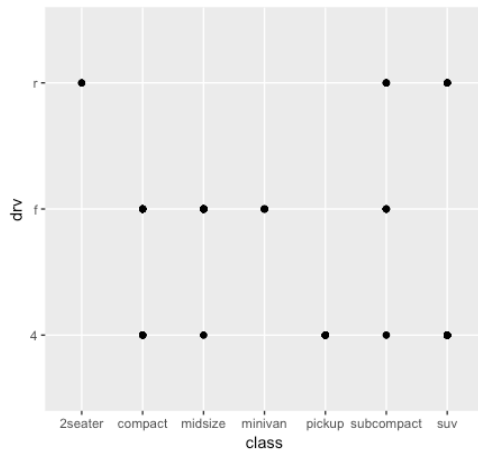1. ggplot(data = mpg)

2. [1] 234   11 : 234 rows and 11 columns.
3. drv variable describes whether the vehicle is front-wheel drive, rear wheel drive, or 4 wheel drive.
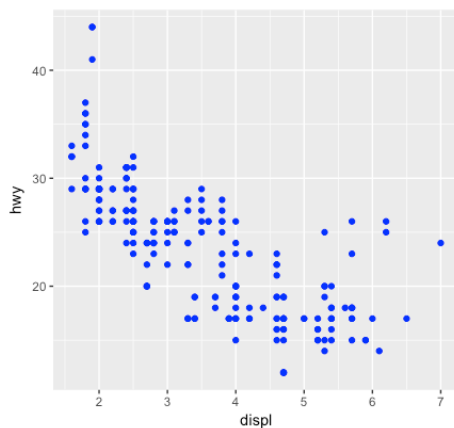4. ggplot(mpg) + geom_point(aes(hwy, cyl))

5. ggplot(mpg) + geom_point(aes(class, drv))

Because both variables are categorical, the points will overlap.

Section 3.3.1: Exercises 1,2,5

1. Because the color argument is within aes(), not geom_point(). It should be "ggplot(data = mpg) + geom_point(mapping = aes(x = displ, y = hwy), color = "blue")" to have the points blue.

2.

```
> mpg
# A tibble: 234 x 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
7          audi          a4   3.1  2008     6   auto(av)      f    18    27     p
8   audi a4 quattro         1.8  1999     4 manual(m5)      4    18    26     p
9   audi a4 quattro         1.8  1999     4   auto(l5)      4    16    25     p
10  audi a4 quattro         2.0  2008     4 manual(m6)      4    20    28     p
# ... with 224 more rows, and 1 more variables: class <chr>
```

Categorical: manufacturer, model, trans, drv, fl, class
Continuous: displ

5. Stroke controls the thickness of the border of certain shapes. It only works for the shapes which have borders: 21-24.

Section 3.6.1: Exercises 1-5

1. Line chart - geom_line(); Boxplot - geom_boxplot(); Histogram - geom_histogram(); Area chart - geom_area()
2. There are both plotted dots and smooth lines. The X-axis will be "displ" and the Y-axis will be "hwy". And the color is defined by different drv variables in the graph.
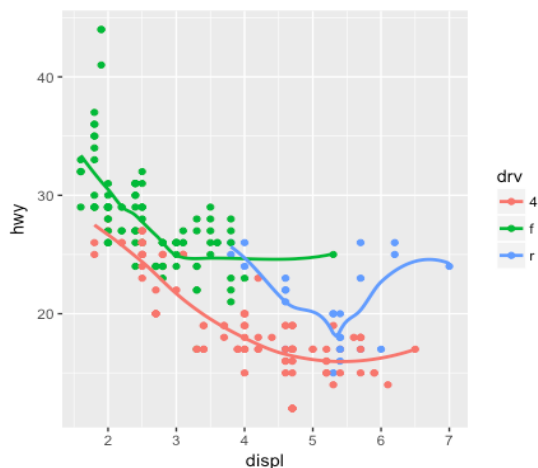

3. It removes the legend. The aesthetics are still mapped and plotted, but it gives a cleaner plot.
4. It determines whether to remove the confidence intervals from the smoothed lines or not
5. They are exactly the same. Because they use the same data and mapping settings

Open response

1. This is a graph which shows the relationship between price, carat and clarity. The X-axis is carat and the Y-axis is price. And the color is

determined by clarity. We can see from the graph, with the same carat, the clarity "IF" is more expensive than other clarity's. And the clarity "IF" has a limited quantity of carat. And the "I1" is the most cheapest kind of clarity and have a large quantity of carat.

ggplot(data=diamonds)+geom_point(mapping = aes(x = carat, y = price, color = clarity))
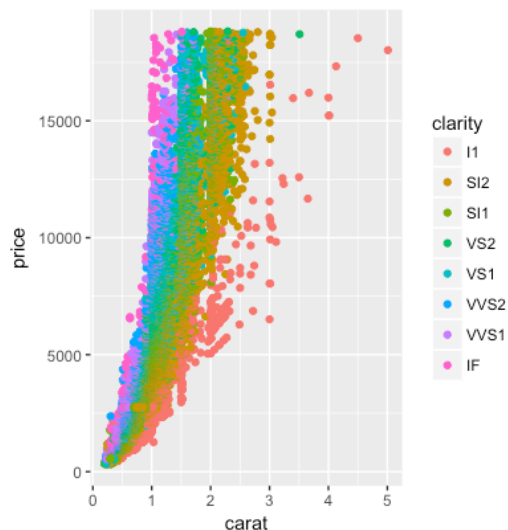
2. This is a graph which shows the relationship between price, carat and clarity. The X-axis is carat and the Y-axis is price. And the color is determined by clarity. The smooth line shows that larger quantity of diamonds will be more expensive.

ggplot(data=diamonds)+geom_point(mapping = aes(x = carat, y = price, color = clarity))+geom_smooth(mapping = aes(x=carat, y = price))