Module 2 Exercises

Brady Lamson

2/7/2022

3: Data Visualization (Graphics)

3.2: A Taxonomy for Data Graphics

Exercise 1:

For each graph indicate:

- The **visual cues** that are used
- The **coordinate system** that's used
- The **scales** that are used
- How **context** is provided

Diamond Price vs Weight

a)
- **Visual Cues:** *Position* along the x and y axis and *color*.
- **Coordinate System:** *Cartesian*
- **Scale:** *Numerical* for the axes scale *Categorical* for the colors
- **Context:** Legend, x and y-axis labels, title

`geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
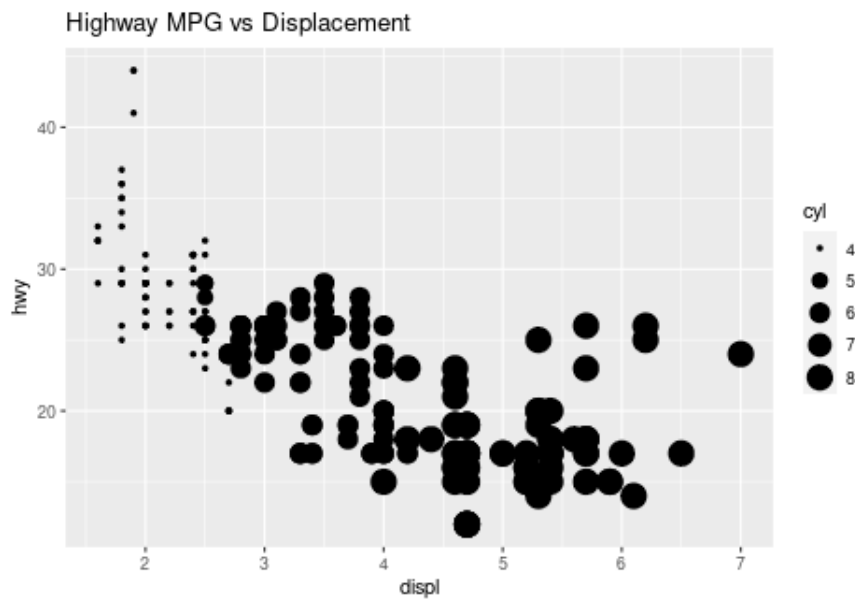
Diamond Price vs Weight

b)   • **Visual Cues:** *Position* along the x and y axis. *Angle / direction* of the line
      • **Coordinate System:** *Cartesian*
      • **Scale:** *Numerical*
   • **Context:** Legend, x and y-axis labels, title

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

Histogram of Diamond Prices
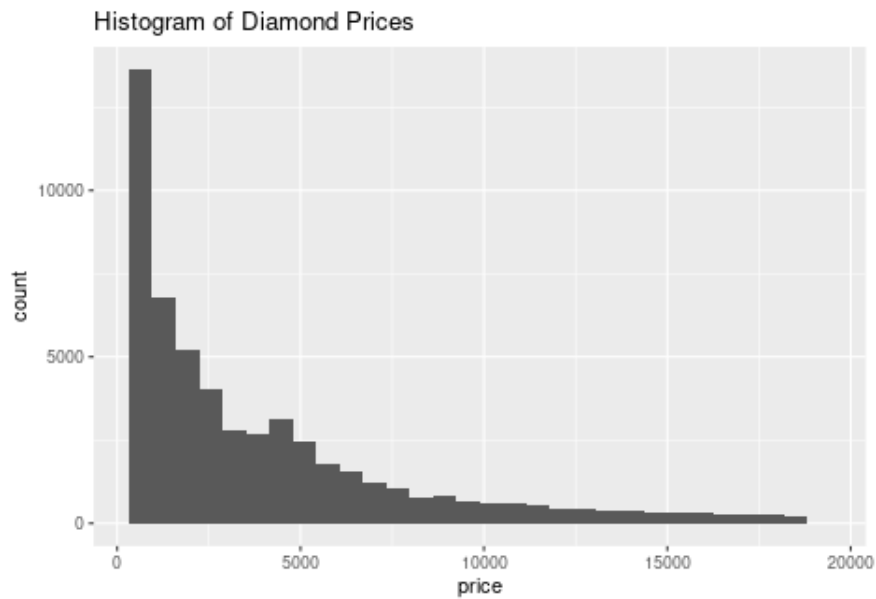
c)   • **Visual Cues:** *Position* along the x and y axis, *length* of the histogram bars. *Area* of the histogram bars.
       • **Coordinate System:** *Cartesian*
       • **Scale:** *Numerical*
       • **Context:** X and y-axis lables, title.

Exercise 2:

a)
- **Visual Cues:** *Position* along the x and y axis, *area* of the circles.
- **Coordinate System:** *Cartesian*
- **Scale:** *Numerical*
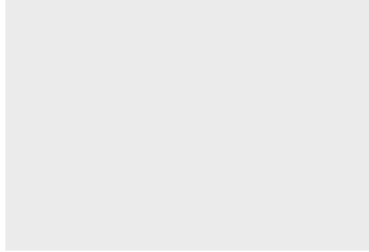- **Context:** X and y-axis lables, title, legend.
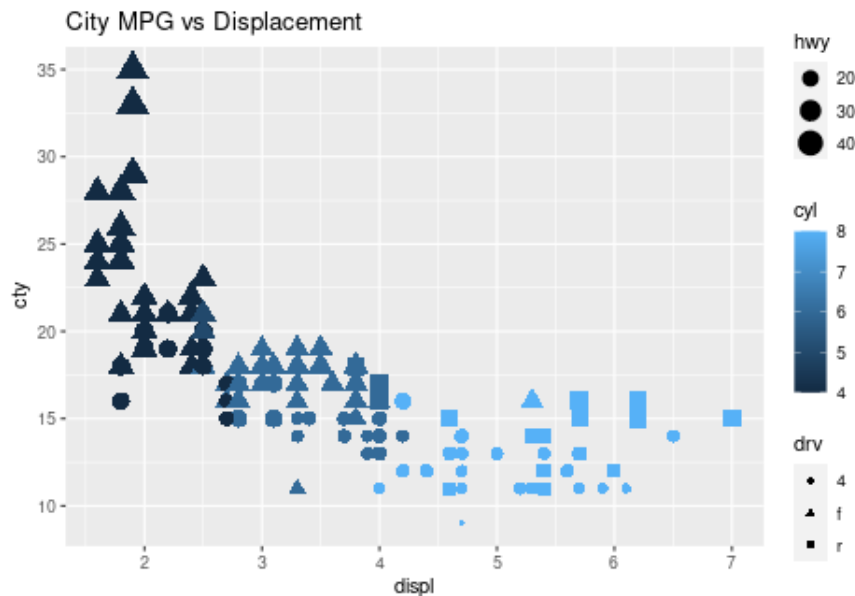
b)
- **Visual Cues:** *Position* along the x and y axis, *shade, shape, area.*
- **Coordinate System:** *Cartesian*
- **Scale:** *Numerical*
- **Context:** Title, x and y-axis, three different legends

4: A Grammar for Graphics with "ggplot2"

4.1: Introduction

Exercise 3:

This outputs a blank box, the box that will have more and more information added onto it once more things are specified.
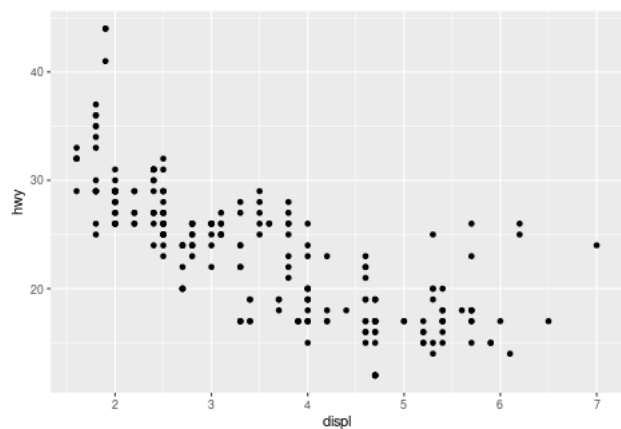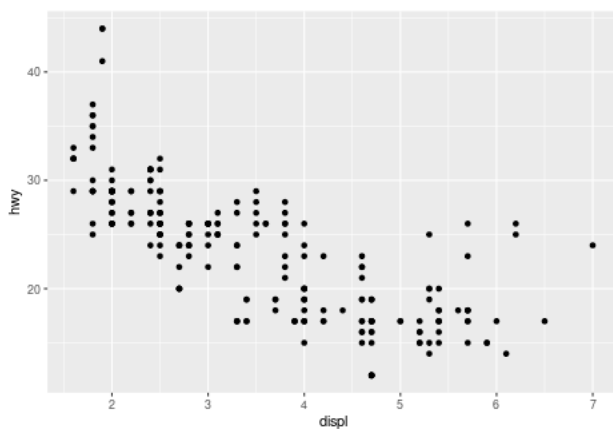
Exercise 4:

Guess whether the following commands both make the same scatterplot, then check your answer:

I would guess **yes**. These would be different *if* there was another geom_* used with a different dataset. These, I think, should be functionally equivelant.

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

## Specify data in geom_*() function:
ggplot() +
geom_point(data = mpg, mapping = aes(x = displ, y = hwy))
```
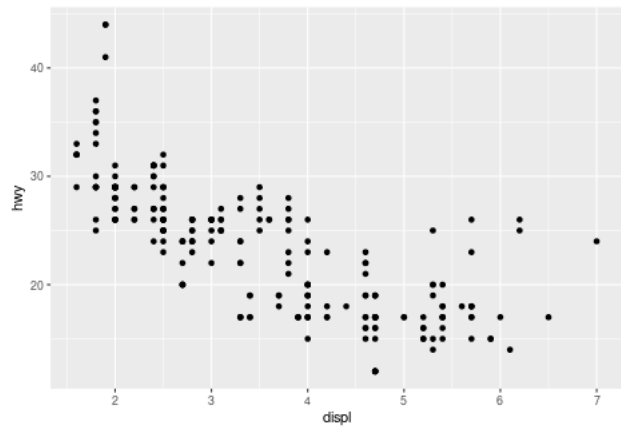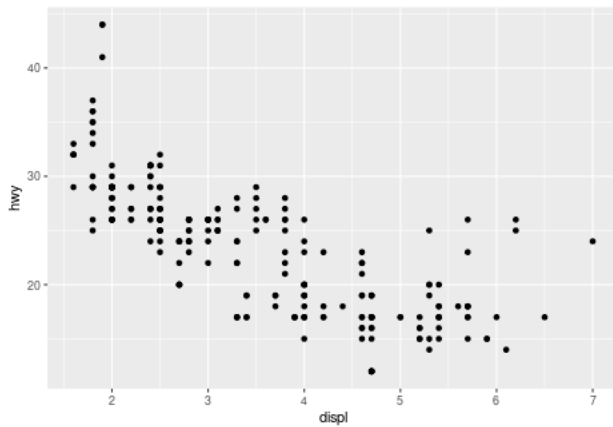
Exercise 5:

I would make the same guess that **yes**, these are equivalent expressions. Not enough is really going on to impact the graph.

```
## Specify aesthetics in geom_*() function:
ggplot(data = mpg) +
geom_point(mapping = aes(x = displ, y = hwy))
## Specify aesthetics in ggplot():
ggplot(data = mpg, mapping = aes(x = displ, y = hwy)) +
geom_point()
```

Exercise 6

   a) Guess what the **ggtitle(), xlab()**, and **ylab()** commands do to the scatterplot below and to the left. Then check your answers.

   - ggtitle will put a title at the top with the text "Highway MPG vs Displacement"
   - xlab and ylab will put text labels on the x-axis and y-axis respectively.

   b) Guess what the **labs()** command does to the scatterplot below and to the right. Then check your answer.

   - I'm not sure actually. col = drv means we have some categorical color usage here and my gut instinct says the only way a label would make sense here is with a legend. Let's go with that then! This will create a legend with the text label above it saying "Drivetrain".

```
# A
ggplot(data = mpg) +
geom_point(mapping = aes(x = displ, y = hwy)) +
ggtitle(label = "Highway MPG vs Displacement") +
xlab(label = "Displacement") +
ylab(label = "Highway MPG")

# B
ggplot(data = mpg) +
geom_point(mapping = aes(x = displ, y = hwy, color = drv)) +
labs(color = "Drivetrain")
```
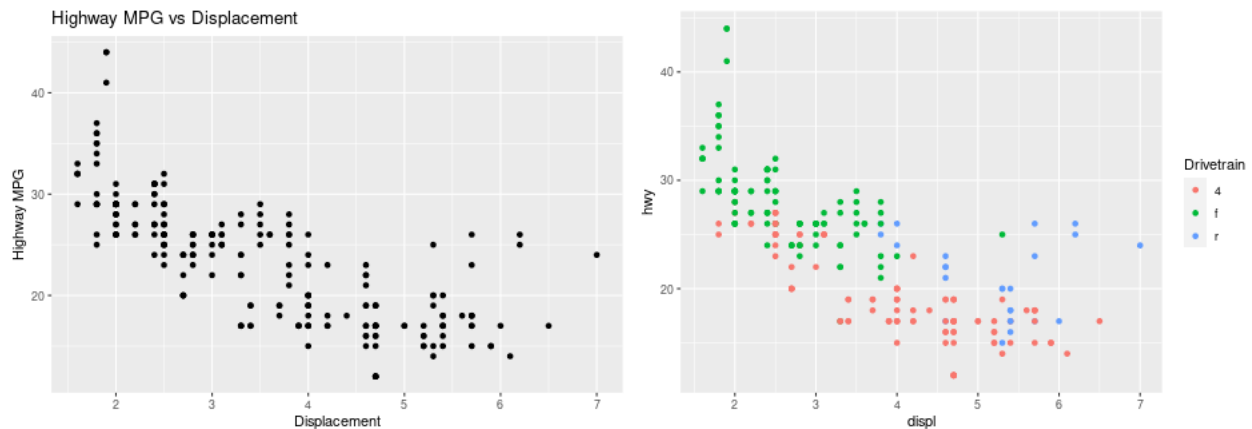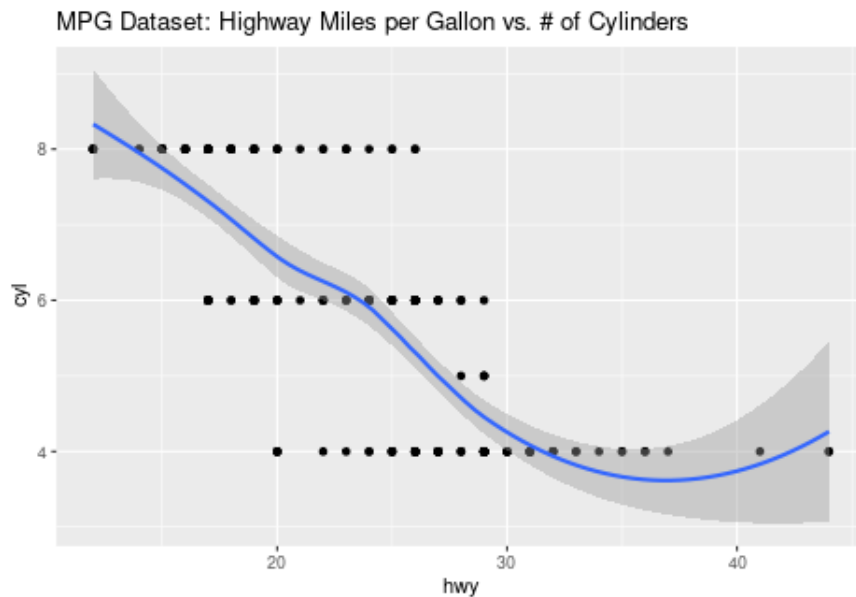
Exercise 7:

a) Make a scatterplot of hwy (on the y-axis) versus cyl (x-axis). Report your R commands.

MPG Dataset: Highway Miles per Gallon vs. # of Cylinders
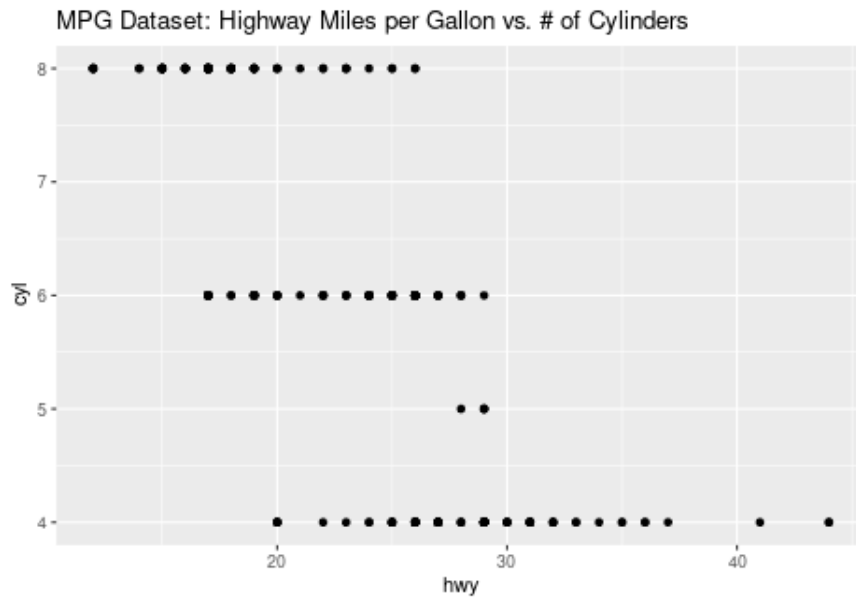
b) Reproduce the scatterplot of Part a, but now add a second layer to the plot using **geom_smooth()**. Report your R command(s).

```
## 'geom_smooth()' using method = 'loess' and formula 'y ~ x'
```

MPG Dataset: Highway Miles per Gallon vs. # of Cylinders

c) Make a scatterplot of class (y-axis) versus drv (x-axis)? What happens? Why is the plot not useful?
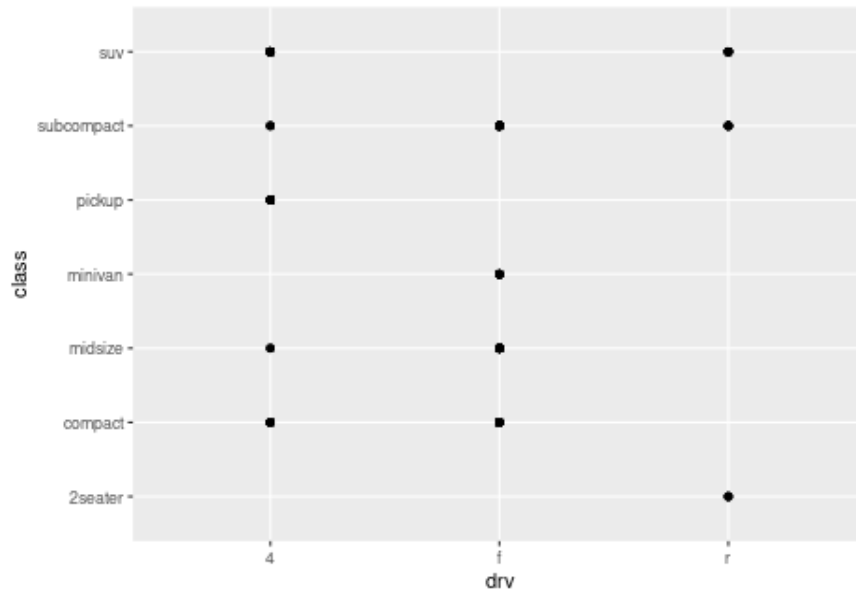
This plot is trying to use Cartesian coordinates for two different categorical values. The plot just doesn't make any sense as none of the visual cues carry any relevant information.

4.2: More on Aesthetic Mappings