Data Visualzation with ggplot2 exercises

Fechar Ourotcha

3/27/2020

EXERCISE 1

Question 1) Run ggplot(data = mpg). What do you see?

```
ggplot(data = mpg)
```


When I run the code, I get a blank graph.

Question 2) How many rows are in mtcars? How many columns?

```
dim(mpg)
```

There are 234 rows and 11 columns in this dataset

Question 3) What does the drv variable describe? Read the help for ?mpg ## to find out.

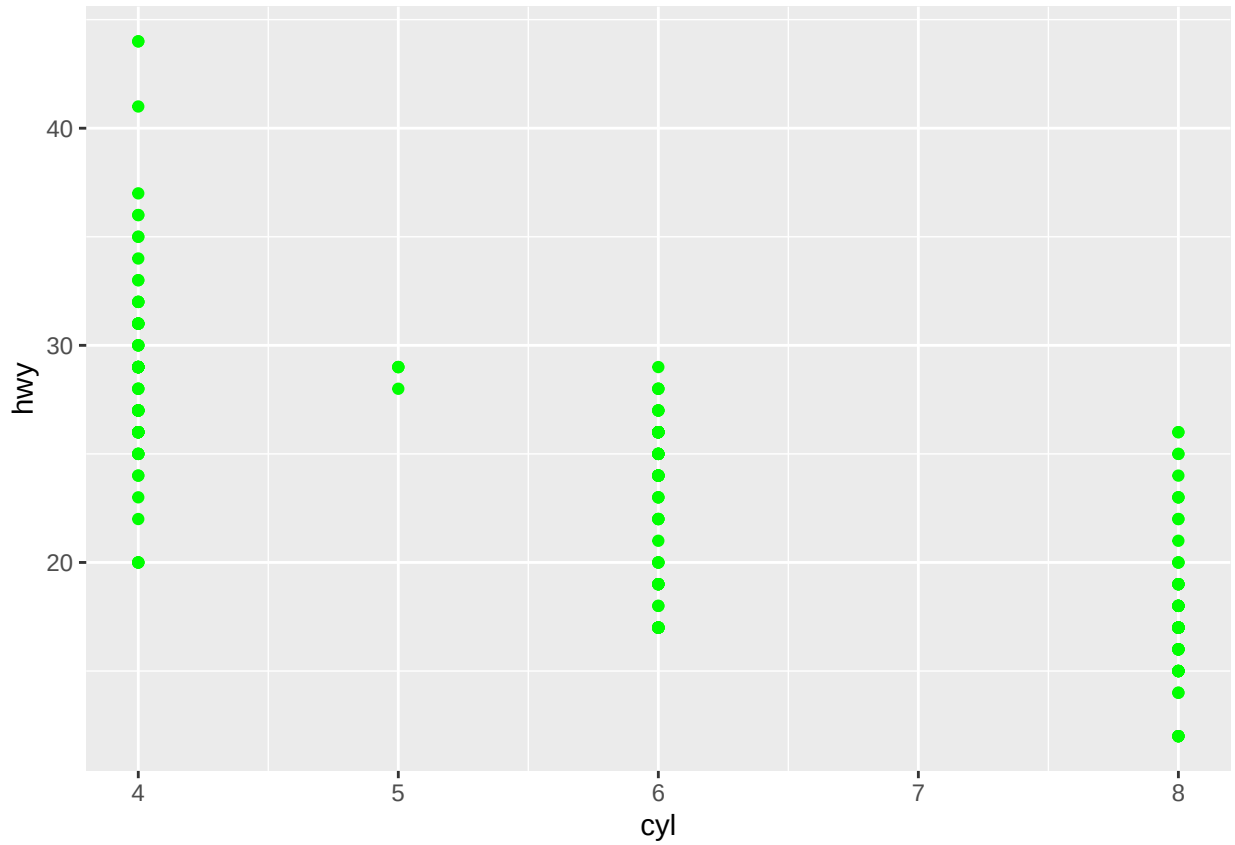
```
?mpg
```

```
## starting httpd help server ... done
```

the drv variable is a character class column; with information on whether the car model is either f, r, or 4 (f = front-wheel drive, r = rear wheel drive, 4 = 4w)

Question 4) Make a scatterplot of hwy versus cyl.

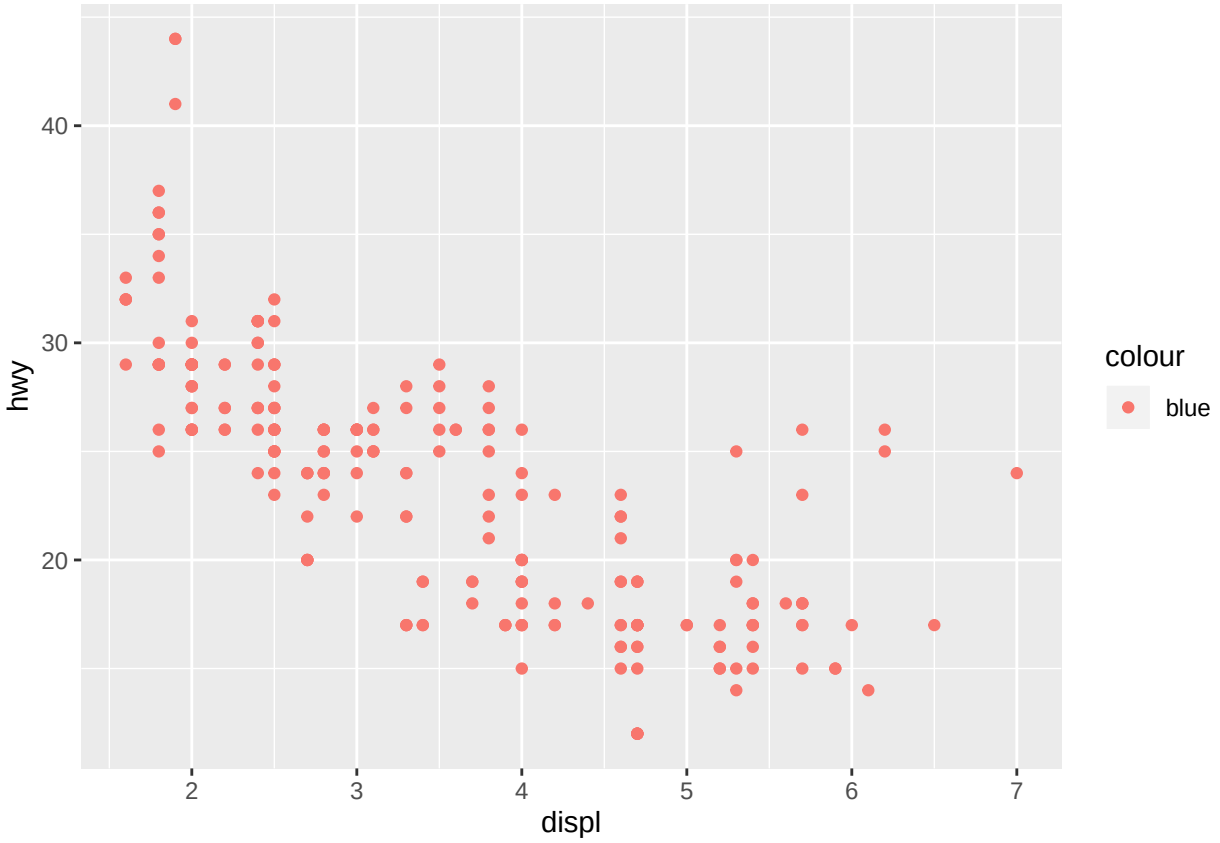
```
attach(mpg)
ggplot(data = mpg)+
  geom_point(mapping = aes( x = cyl, y = hwy ), color = ("green"))
```



EXERCISE 2

Question 1) What's gone wrong with this code? Why are the points not ## blue?

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



The color is not blue because the color argument is inside the aes, which is making r think "blue" is just a name. To change this bring the color argument outside of aes.

Question 2) 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?

```
glimpse(mpg)
```

Pretty much all the variables are categorical except for engine displacement, highway miles per gallon.

Question 3) Map a continuous variable to color, size, and shape. How do ## these aesthetics behave differently for categorical versus continuous ## variables?

Question 4) What happens if you map the same variable to multiple ## aesthetics?

Question 5) What does the stroke aesthetic do? What shapes does it work with? (Hint: use ?geom_point.)

```
?geom_point
```

the stroke aesthetic modifies the width of the border

EXERCISE 3

Quetsion 1) What happens if you facet on a continuous variable?

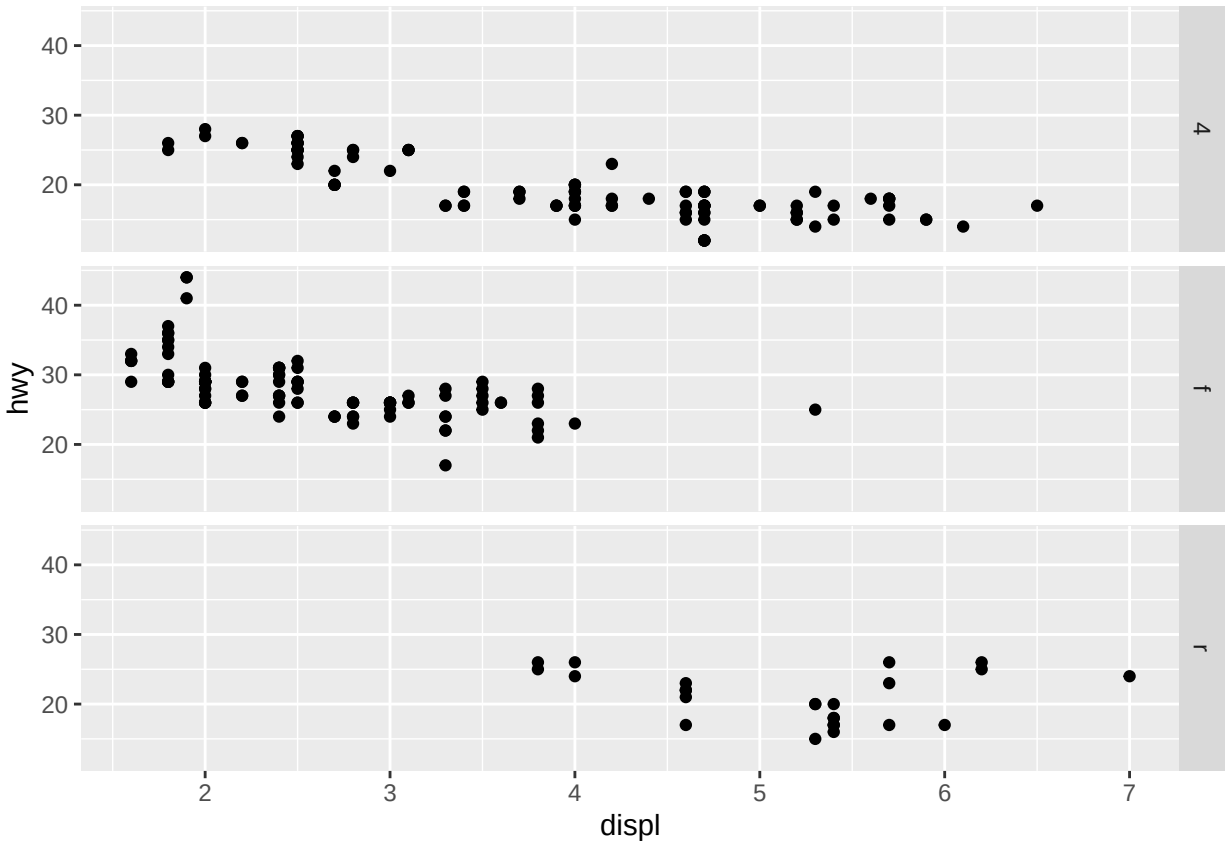
when you facet on a continuous variable, you basically breack the Variable into different categories that ## allows you to get a more detail insight about that variable.

Question 2) What do the empty cells in a plot with facet_grid(drv ~ ## cyl) mean? How do they relate to this plot?

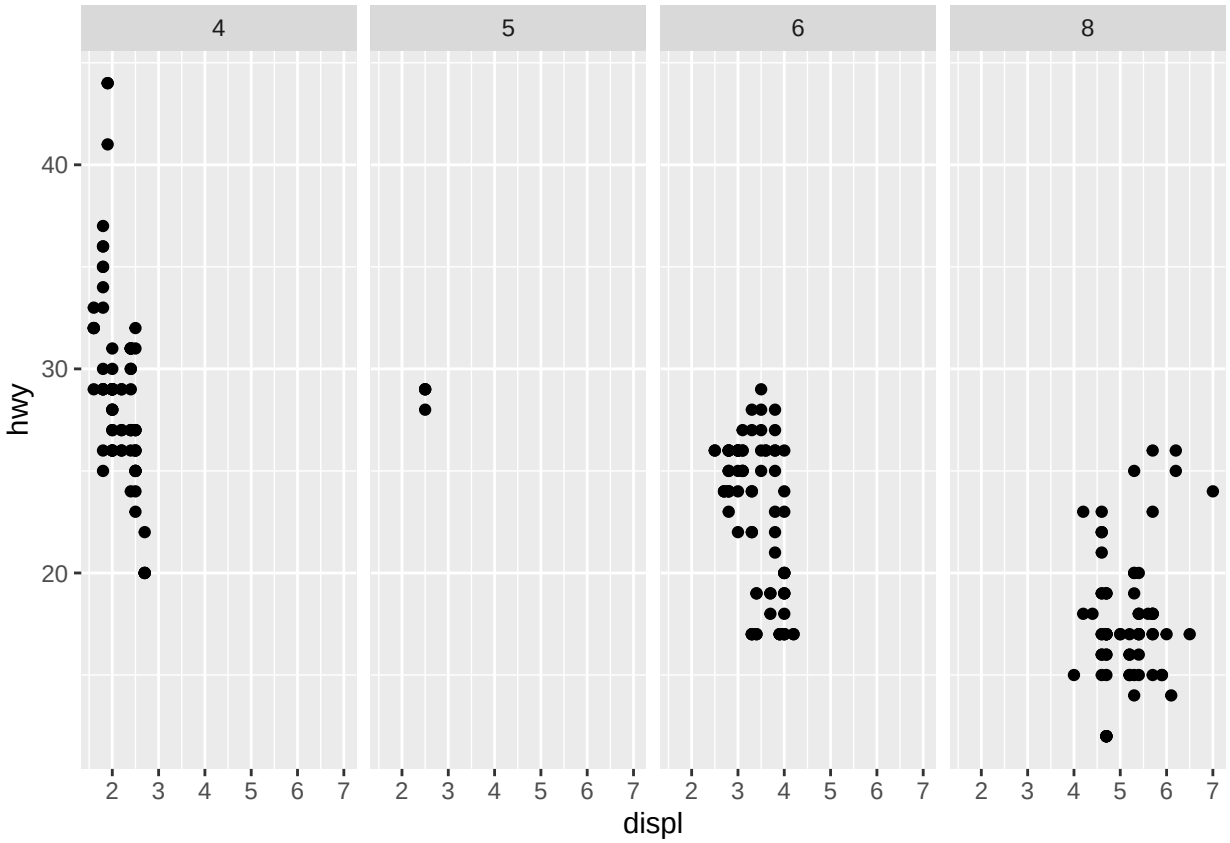
that means there are no value for the category. In the plot we see that there are no cars with 5 cylinders ## and 4 wheel drive. Also no cars with rear wheel drive and 4 or 5 cylinder.

Question 3) What plots does the following code make? What does . do?

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



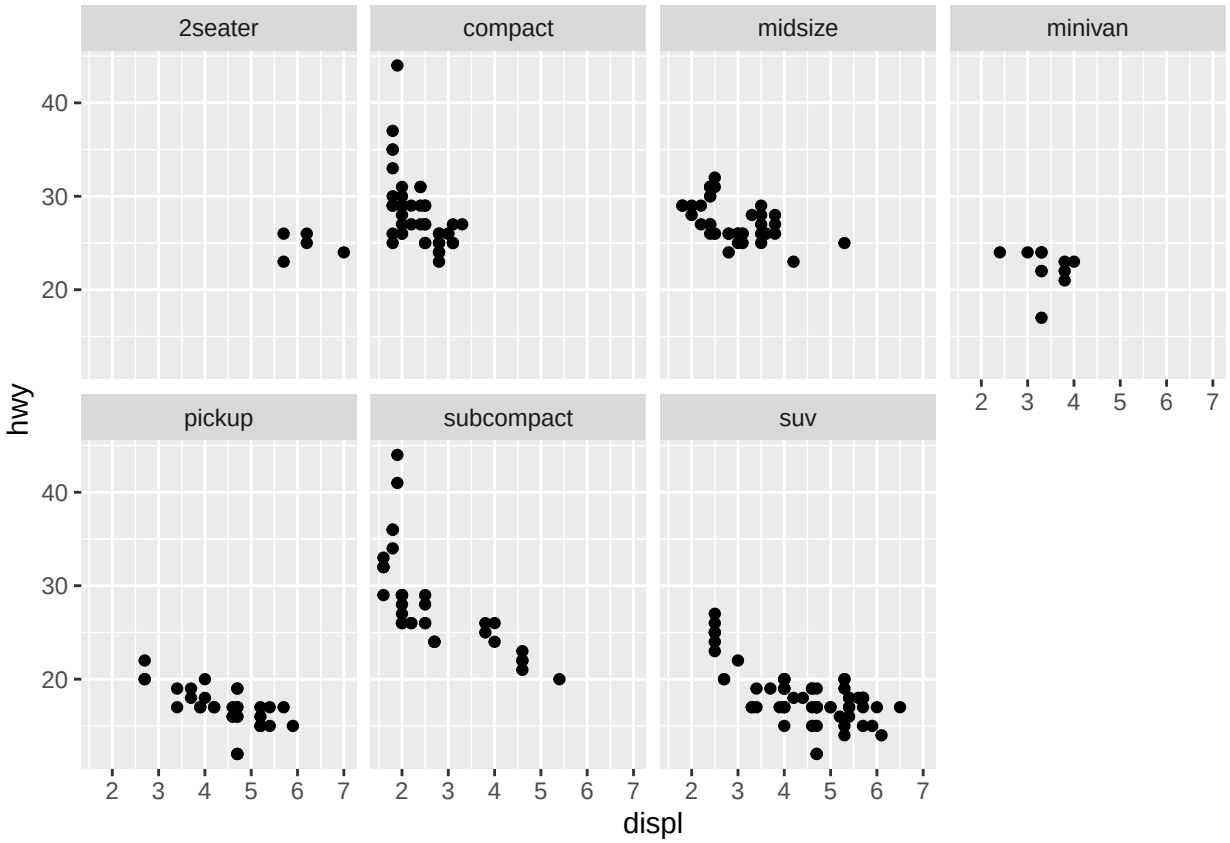

```
ggplot(data = mpg) +  geom_point(mapping = aes(x = displ, y = hwy)) +  facet_grid(. ~ cyl)
```



The foolowing plots create scatter plots. the dot is just a place holder

Question 4) Take the first faceted plot in this section:

```
ggplot(data = mpg) +  geom_point(mapping = aes(x = displ, y = hwy)) +  facet_wrap(~ class, nrow = 2)
```



What are the advantages to using faceting instead of the color aesthetic? What are the disadvantages? How might the balance change if you had a larger dataset?

faceting allows you to see the different categories within a variable independently

Question 5) Read ?facet_wrap.

What does nrow do?

What does ncol do?

What other options control the layout of the individual panels?

Why doesn't facet_grid() have nrow and ncol variables?

Question 6). When using facet_grid() you should usually put the variable with more unique levels in the columns. Why?

I guess because it looks better.

EXERCISE 4

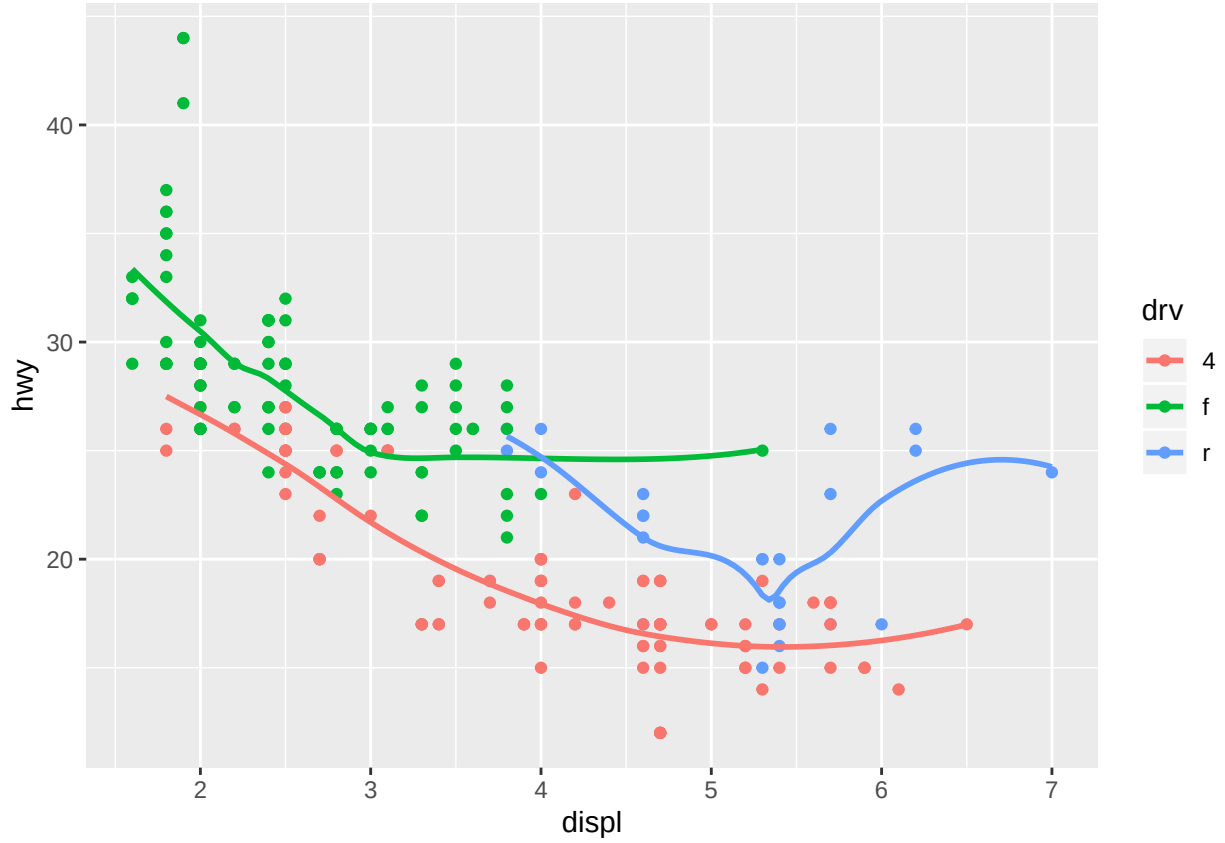
Question 1) What geom would you use to draw a line chart? A boxplot? A histogram? An area chart?

line geom, histgram geom, boxplot geom, and an area geom

Question 2) Run this code in your head and predict what the output will look like. Then, run the code in R and check your predictions:

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

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



it creates a scatterplot and then adds a smooth line on top

Question 3) What does show.legend = FALSE do? What happens if you remove it? Why do you think I used it earlier in ## the chapter?

it remove the legend labels off the graph. Youu did not use it because it was unnessary. there were only 3 categories within that variable, and the color argument identify them already.

Wuestion 4) What does the se argument to geom_smooth() do?

it display confidence interval around the smooth line

EXERCISE 5

Question 1) What is the default geom associated with stat_summary()?

How could you rewrite the previous plot to use that geom function instead of the stat function?

geom_pointrange is the one. I would write it like this:

```
ggplot(data = diamonds) +
  geom_pointrange( mapping = aes(x = cut,y = depth),
                   fun.ymin = min,
                   fun.ymax =max,
                   fun.y = median  )
```

```
## Warning: Ignoring unknown parameters: fun.ymin, fun.ymax, fun.y
```

```
## Error: geom_pointrange requires the following missing aesthetics: ymin, ymax
```

Question 2) What does geom_col() do? How is it different to geom_bar()?

geom_col make the height of the bar with the values in the dataset, where as geom_bar make the height base on the proportion of the data values

Question 3) Most geoms and stats come in pairs that are almost always used in concert. Read through the documentation and make a list of all the pairs. What do they have in common?

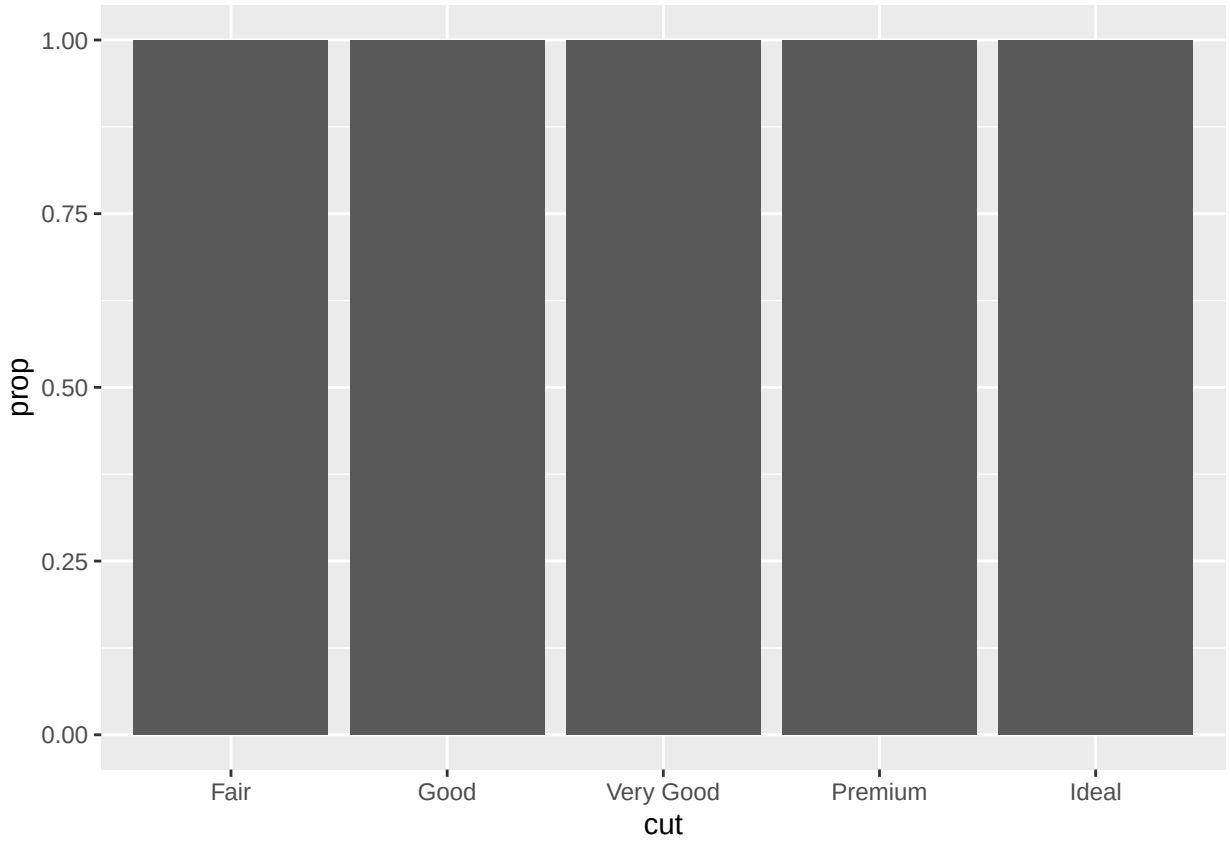
They do the same thing

Question 4) What variables does stat_smooth() compute? What parameters control its behavior?

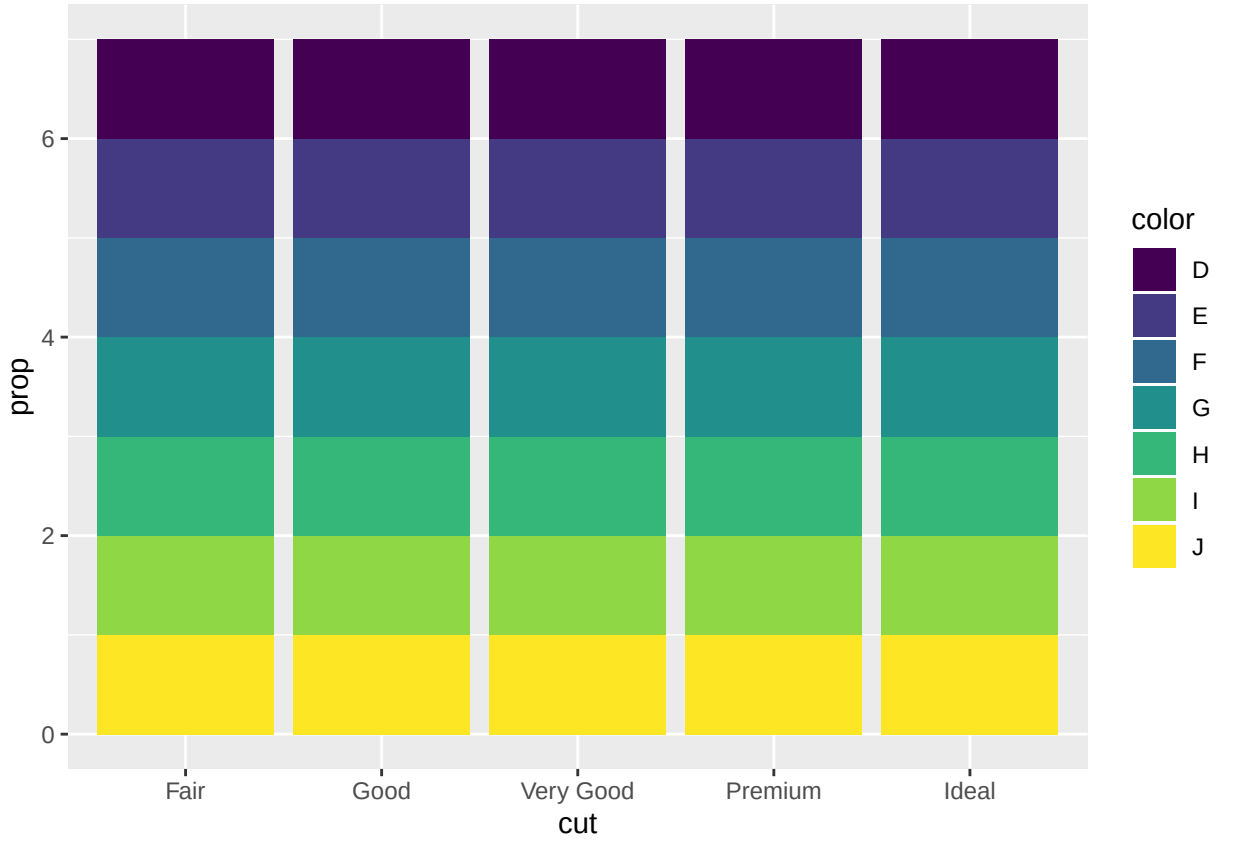
it computes the y variables. the mapping parameter, position and others things control its behaviors

Question 5) In our proportion bar chart, we need to set group = 1. Why? In other words what is the problem with these two graphs?

```
ggplot(data = diamonds) +  geom_bar(mapping = aes(x = cut, y = ..prop..))
```




```
ggplot(data = diamonds) +  geom_bar(mapping = aes(x = cut, fill = color, y = ..prop..))
```



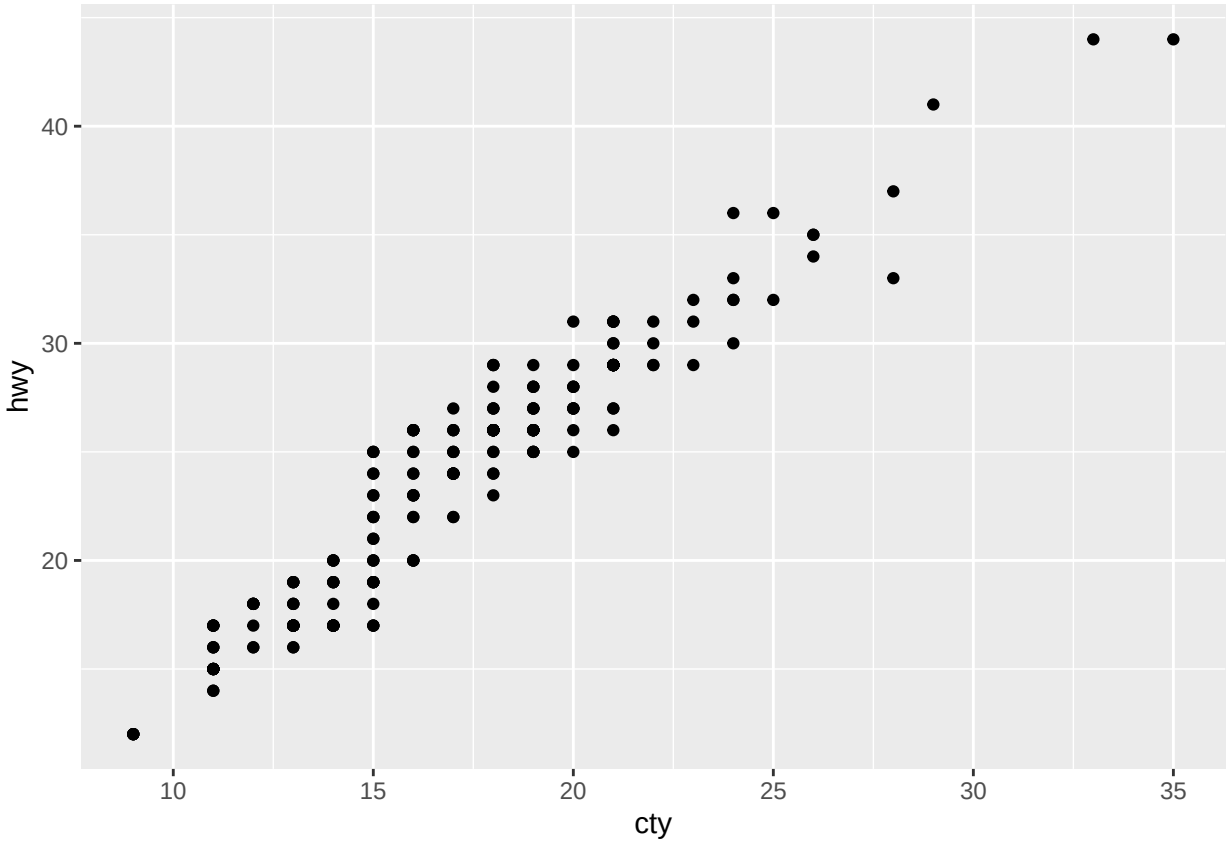
it's because proportion is a percentage and it cant go over 1

EXERCISE 6

Question 1) What is the problem with this plot?

How could you improve it?

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



This plot has no main title, x-axis and y-axis does not tell you anything meanful about what the graph is about. The fix it I would just add the things I mentioned.

Question 2) What parameters to geom_jitter() control the amount of jittering?

width and height

Question 3) Compare and contrast geom_jitter() with geom_count().

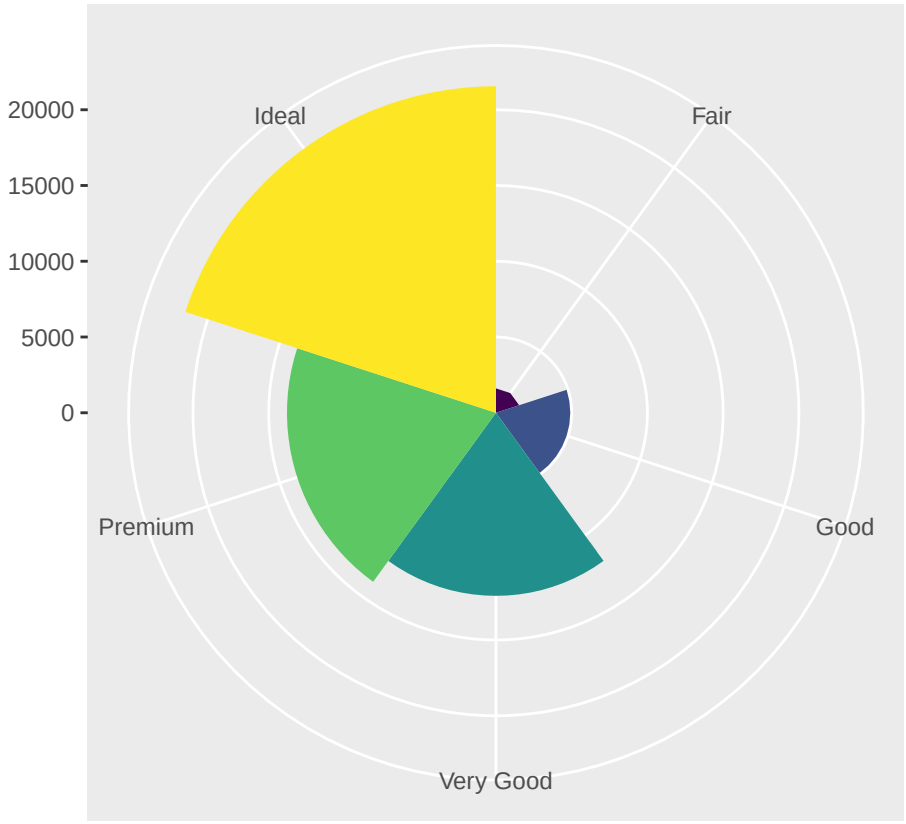
jitter It adds a small amount of random variation to the location of each point, where as geom_counts the number of observations at each location, then maps the count to point area

Question 4) What's the default position adjustment for geom_boxplot()? Create a visualization of the mpg dataset that demonstrates it.

EXERCISE 7

Question 1) Turn a stacked bar chart into a pie chart using coord_polar().

```
bar <- ggplot(data = diamonds) +  geom_bar(    mapping = aes(x = cut, fill = cut),    show.legend = FALS
bar + coord_polar()
```



Questionn 2) What does labs() do? Read the documentation.

They label your plots

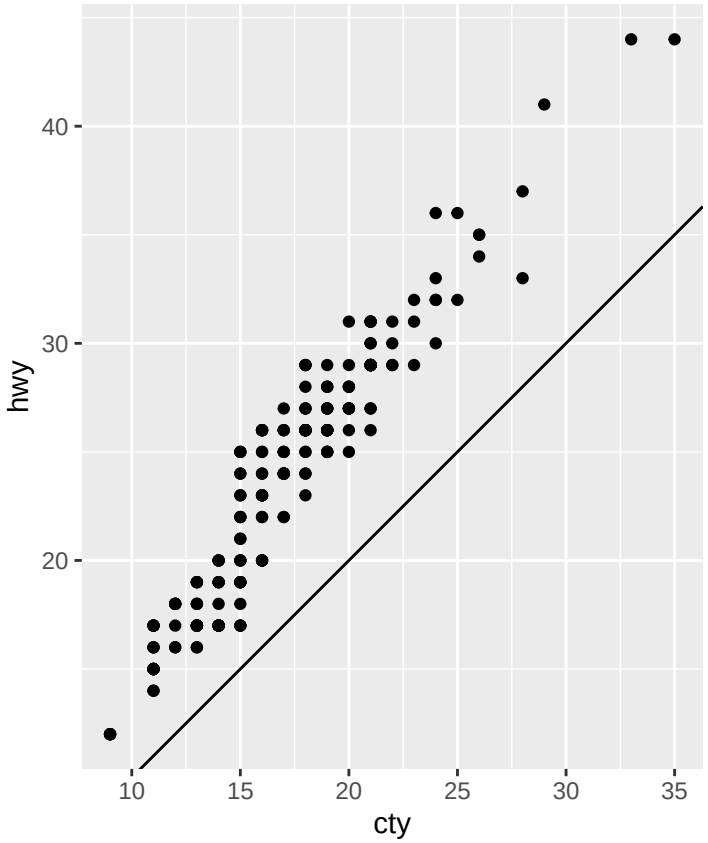
Question 3) What's the difference between coord_quickmap() and coord_map()?

coord_map projects a portion of the earth, which is approximately spherical, onto a flat 2D plane using any projection defined by the mapproj package. Map projections do not, in general, preserve straight lines, so this requires considerable computation. coord_quickmap is a quick approximation that does preserve straight lines. It works best for smaller areas closer to the equator.

Question 4) What does the following plot tell you about the relationship between city and highway mpg? Why is coord_fixed() important? What does geom_abline() do?

there is a positive relationship. geom_abline() adds a trendline t the scaterplot. coord_fixed() is important because it adjust your cordinate system acording to your graph.

```
ggplot(data = mpg, mapping = aes(x = cty, y = hwy)) +  geom_point() +  geom_abline() +  coord_fixed()
```


40
30
20
hwy
10
15
20
25
30
35
cty