Exercícios Cap 03

Rodrigo Giannotti

Capitulo 3

Inicialização

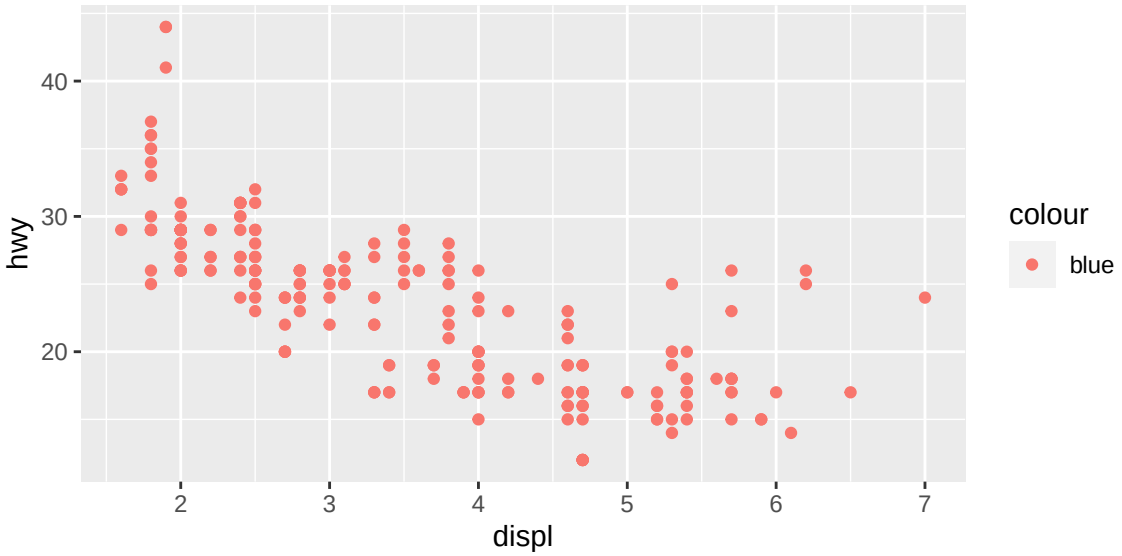
Exercícios

3.3 Aesthetic mappings

3.3.1

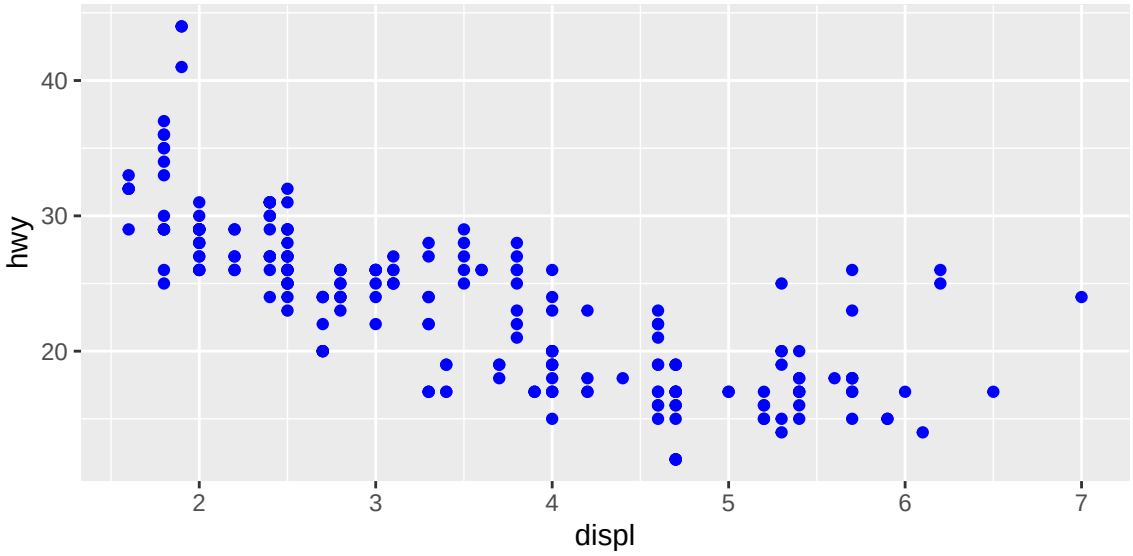
What's gone wrong with this code? Why are the points not blue?

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



Pois o blue, como não está variando de acordo com os grupos, mas sim tem a intenção de colorir todos os pontos deveria estar fora do parenteses aes, como o seguinte:

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



3.3.2

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?

```
# ?mpg
mpg
```

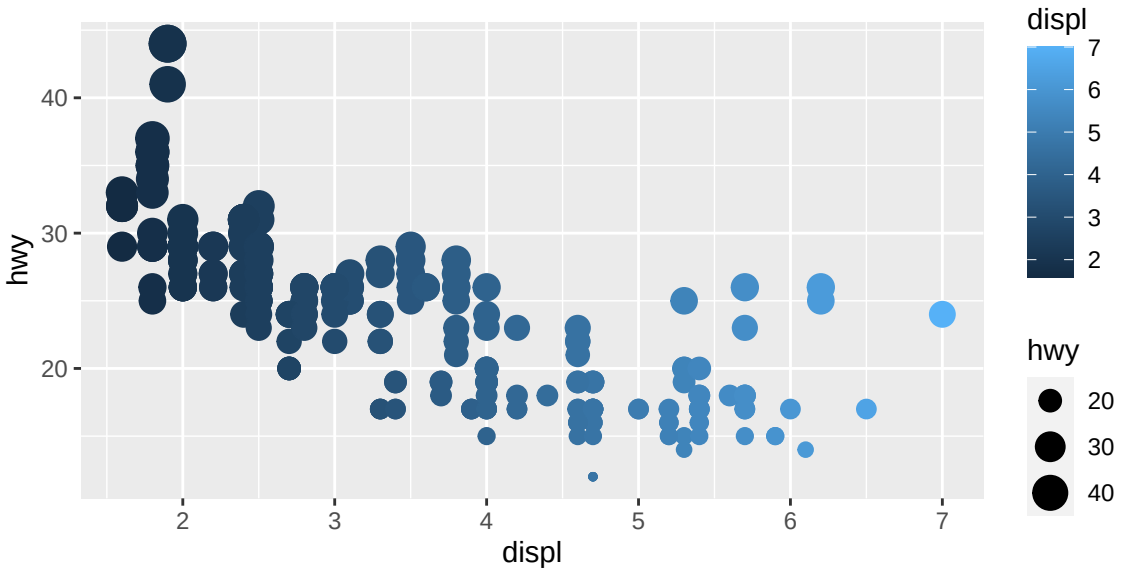
```
## # A tibble: 234 x 11
##    manufacturer model    displ  year   cyl trans   drv     cty   hwy fl    class
##    <chr>        <chr>    <dbl> <int> <int> <chr>   <chr> <int> <int> <chr> <chr>
##  1 audi         a4         1.8  1999     4 auto(l~ f        18    29 p     comp~
##  2 audi         a4         1.8  1999     4 manual~ f        21    29 p     comp~
##  3 audi         a4         2    2008     4 manual~ f        20    31 p     comp~
##  4 audi         a4         2    2008     4 auto(a~ f        21    30 p     comp~
##  5 audi         a4         2.8  1999     6 auto(l~ f        16    26 p     comp~
##  6 audi         a4         2.8  1999     6 manual~ f        18    26 p     comp~
##  7 audi         a4         3.1  2008     6 auto(a~ f        18    27 p     comp~
##  8 audi         a4 quat~   1.8  1999     4 manual~ 4        18    26 p     comp~
##  9 audi         a4 quat~   1.8  1999     4 auto(l~ 4        16    25 p     comp~
## 10 audi         a4 quat~   2    2008     4 manual~ 4        20    28 p     comp~
## # ... with 224 more rows
```

As variáveis marcadas como são com certeza categoricas, e podemos inferir que provavelmente todas as marcadas como também tem uma grande chance de serem categóricas. Uma forma melhor seria utilizar funções como summary() para ter um nível melhor de informações.

3.3.3

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,
              color = displ,
              size = hwy
            ))
```



Em primeiro lugar, mapear uma variavel continua para shape retorna um erro. As outras características porém criam escalas ao invés de níveis diferentes. O gráfico acima por exemplo pode ter uma quantidade alta de redundâncias, mas é bastante claro por usar das escalas de cor e tamanho para enfatizar os dados dos eixosa cartesianos.

3.3.4

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

Acabamos de ver isso no exemplo anterior.

3.3.5

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

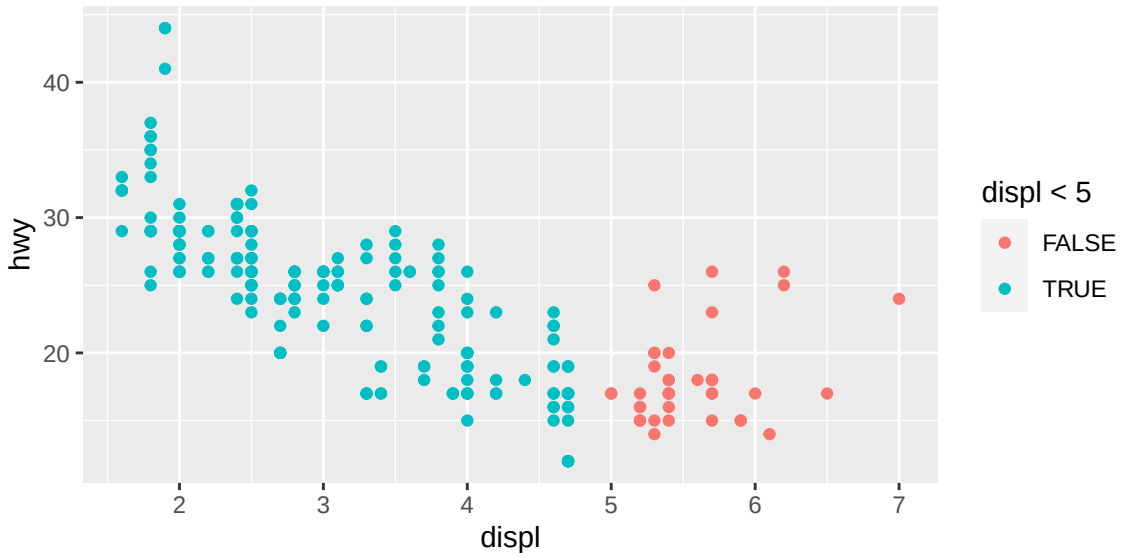
```
# ?geom_point ## explicativon, mas bom mesmo é o exemplo abaixo
# vignette("ggplot2-specs")
```

"Note that shapes 21-24 have both stroke colour and a fill. The size of the filled part is controlled by size, the size of the stroke is controlled by stroke. Each is measured in mm, and the total size of the point is the sum of the two. Note that the size is constant along the diagonal in the following figure." ou seja, para algumas das formas disponiveis no ggplot stroke determina o tamanho do contorno dessas formas.

3.3.6

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,
              color = displ < 5
              ))
```



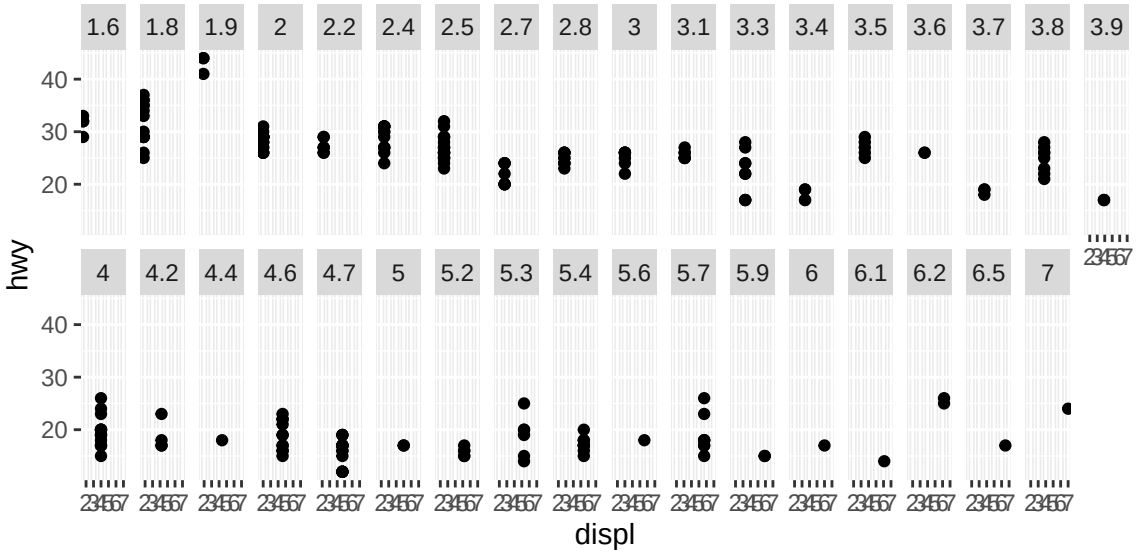
Como podemos ver no gráfico acima utilizar condições envolvendo as variaveis nas aesthetics faz com que ela trate essa separação como trataria uma variavel categorica.

3.5 Facets

3.5.1

What happens if you facet on a continuous variable?

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

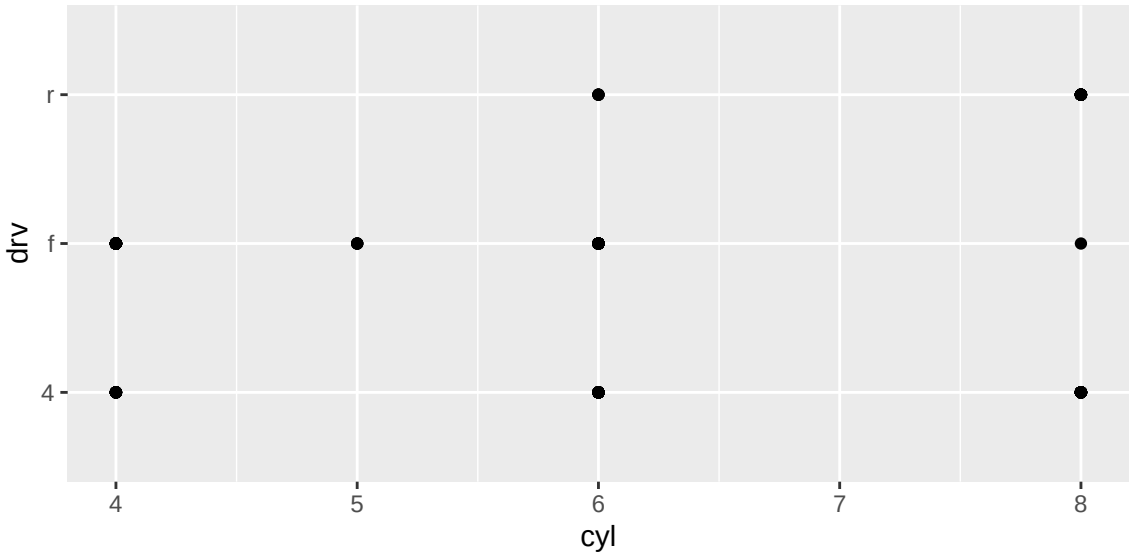


Em resumo, facet não se entende bem com variaveis continuas, retornando coisas estranhas ao tentar se adequar.

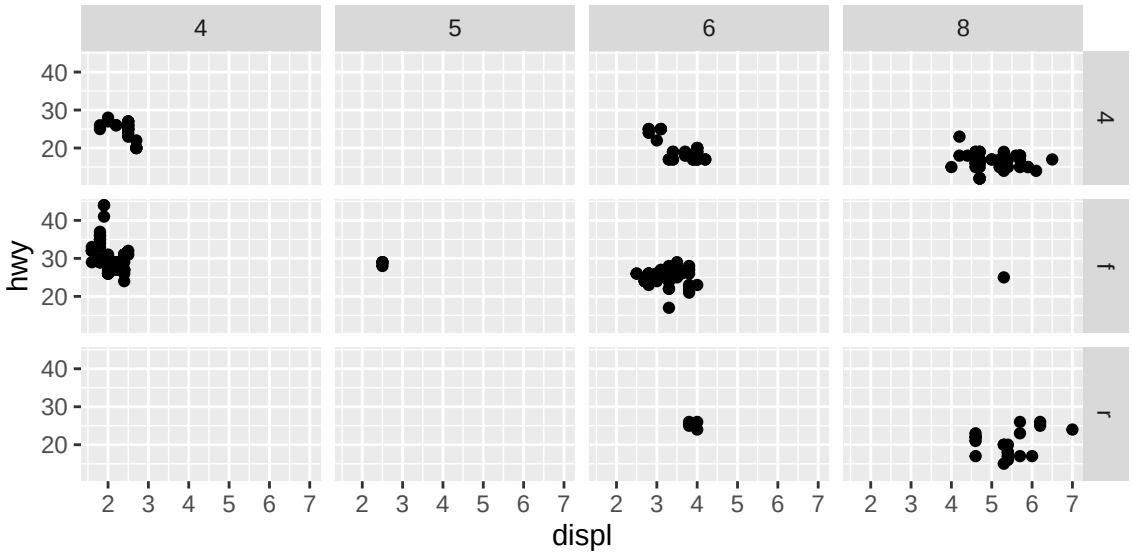
3.5.2

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

```
ggplot(data = mpg) +
  geom_point(mapping = aes(y = drv, x = cyl))
```



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

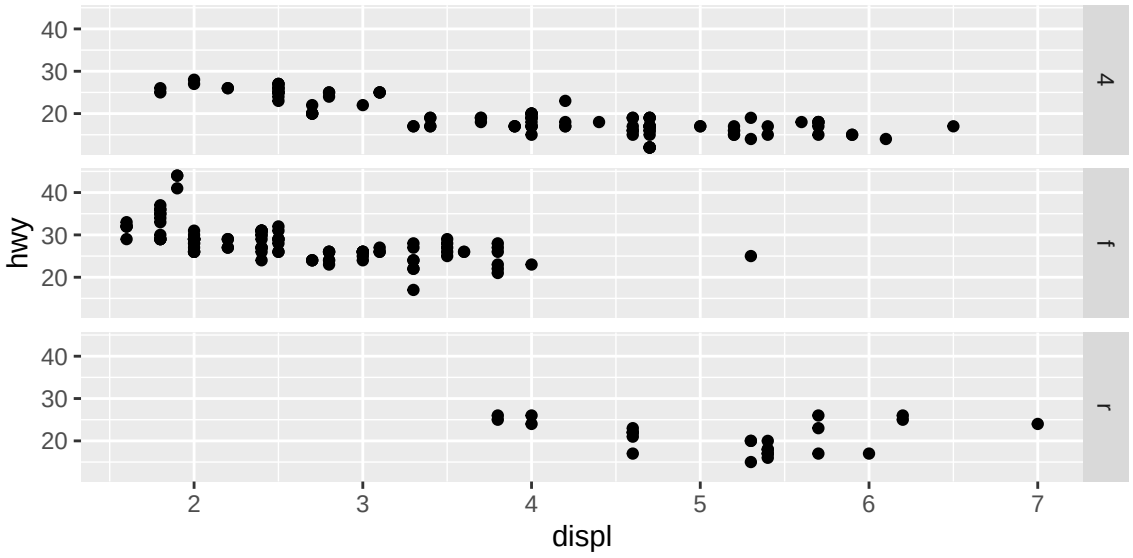


POdemos facilmente entender que as regioes vazias do segundo exemplo correspondem a combinações de drv e cyl que não existem, isso fica mais claro no primeiro grafico.

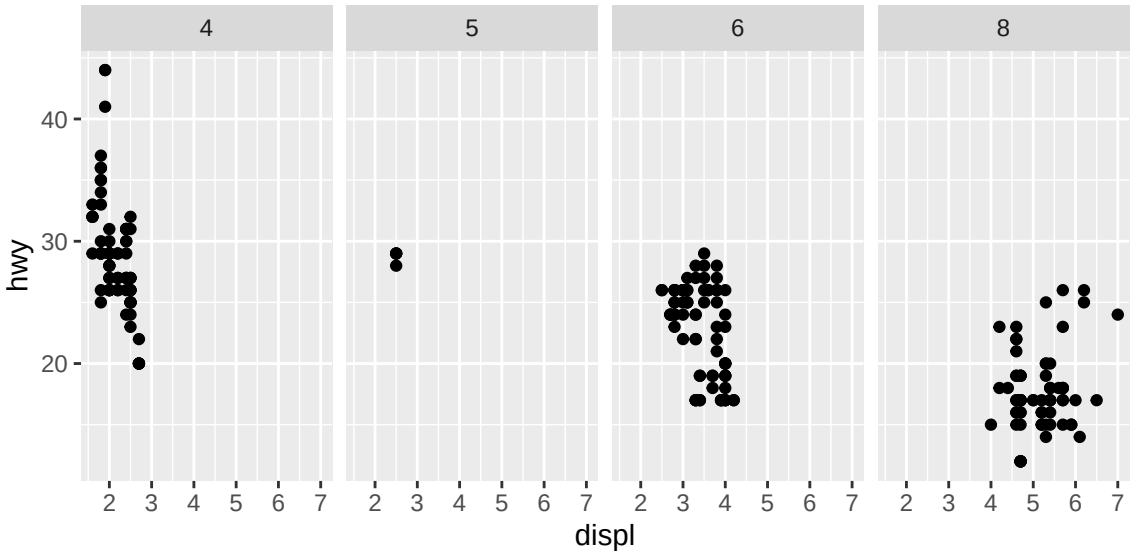
3.5.3

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)
```

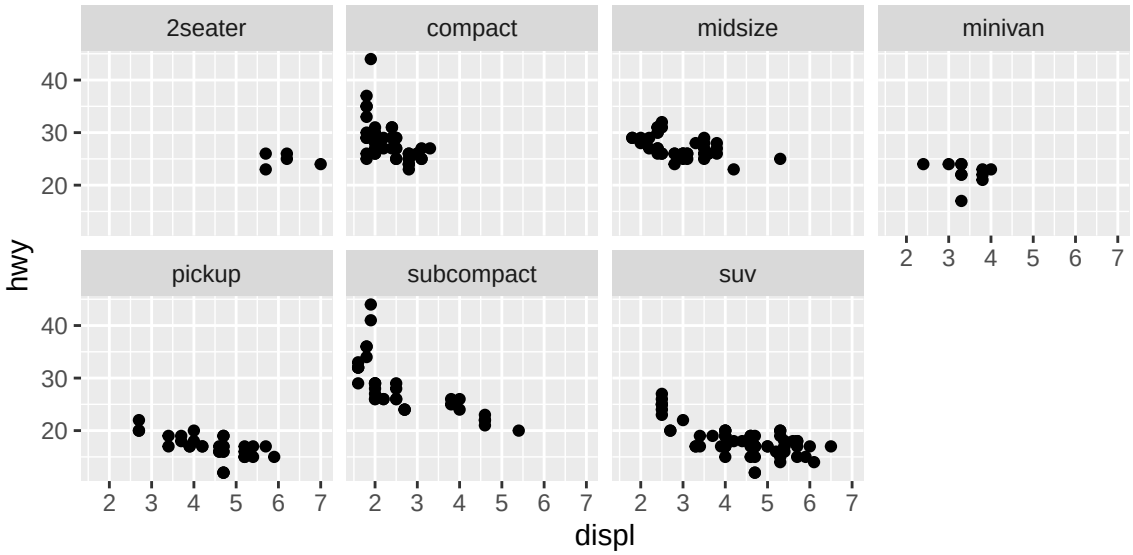


O '.' é utilizado para denotar todas as outras variaveis disponiveis, e neste caso gera só facets relacionadas a variavel que foi explicitada. o ponto estar do lado esquerdo ou direito da expressao configura se as facets estarao no eixo x ou y

3.5.4

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 colour aesthetic? What are the disadvantages? How might the balance change if you had a larger dataset?

Com as facetfica muito mais fácil perceber que os 2 seaters tem caracteristicas muito especificas na relação hwy x displ, usando cores, formatos de pontos ou outros metodos dificilmente conseguiriam mostrar tão bem isso devido a pequena quantidade de 2 seaters em relação as outras classes, escondendo esses pontos.

3.5.5

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 arguments?

```
# ?facet_wrap
# ?facet_grid
```

nrow e ncol são parametros para definir o formato da saida destes graficos, para caso se deseje algum arranjo especifico. dir e drop são outros parametros referentes ao layout e determinam a ordem na qual se sequenciam os graficos nas facets, podendo ser horizontal ou vertical. drop é usado para despresar graficos vazios. facet_grid nao possui argumento de numero de linhas e colunas pois estes deveriam ser definidos pelas variaveis assignadas para essas dimensões.

3.5.6

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

Pois normalmente as pessoas usam monitores nas posição horizontal, dessa forma sendo mais confortável ler gráficos mais longos do que altos.

3.6 Geometric objects

3.6.1

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

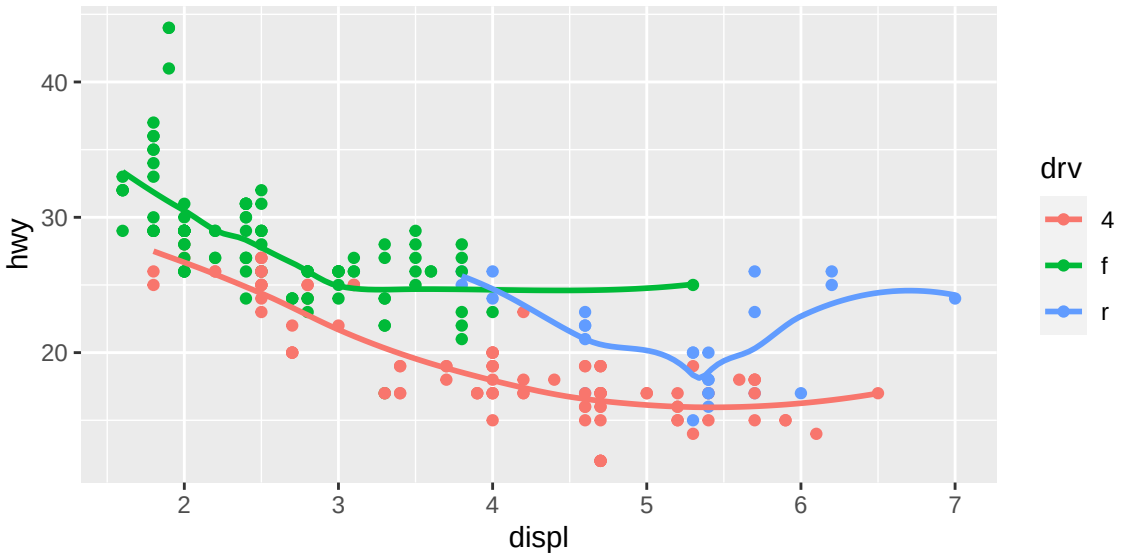
em ordem: geom_smooth; geom_boxplot; geom_histogram; geom_area

3.6.2

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(se = FALSE)
```

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



Um grafico com pontos e linhas com os pontos ilustrando as linhas que tentam explicá-los

3.6.3

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

Como eu esqueci de usar show.legend = FALSE no começo destes exercicios fica facil percever o que ele faz. Além de ser bastante óbvio pelo próprio nome.

3.6.4

What does the se argument to geom_smooth() do?

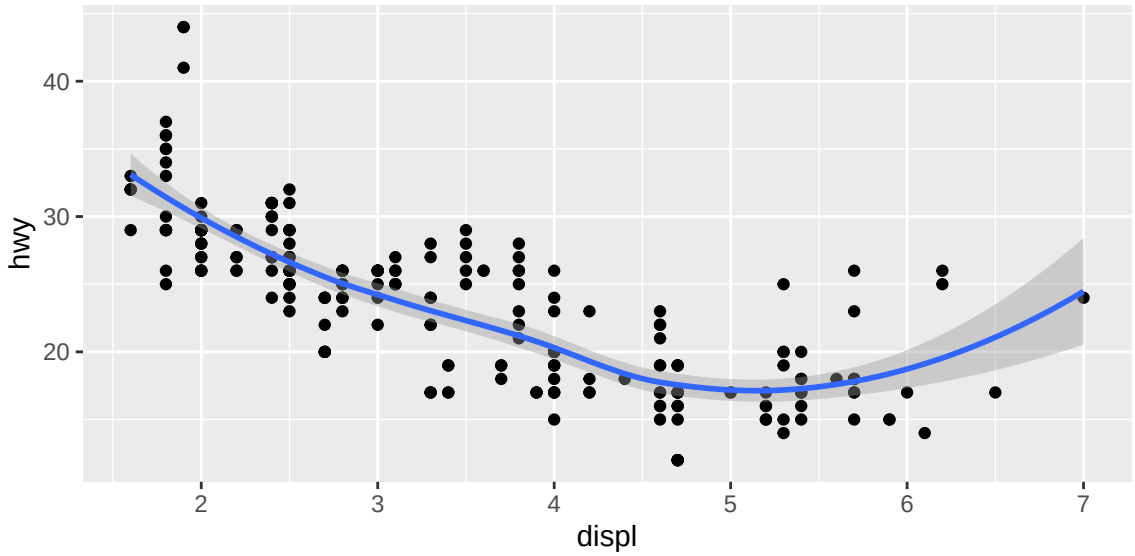
determina se as curvas virão acompanhadas de uma área ilustrando os intervalos de confiança

3.6.5

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

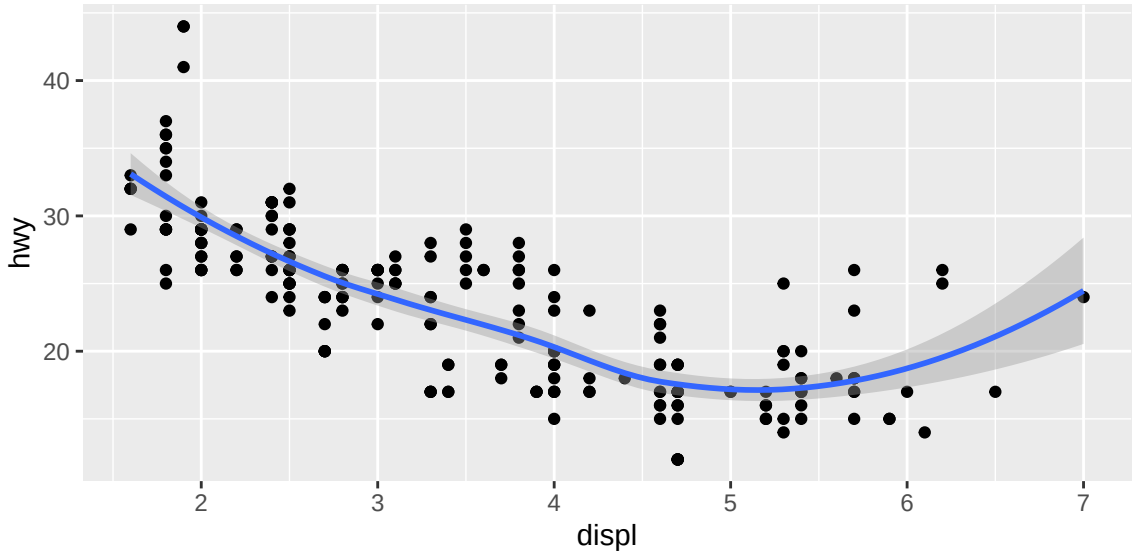
```
ggplot(data = mpg, mapping = aes(x = displ, y = hwy)) +
  geom_point() +
  geom_smooth()
```

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




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

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



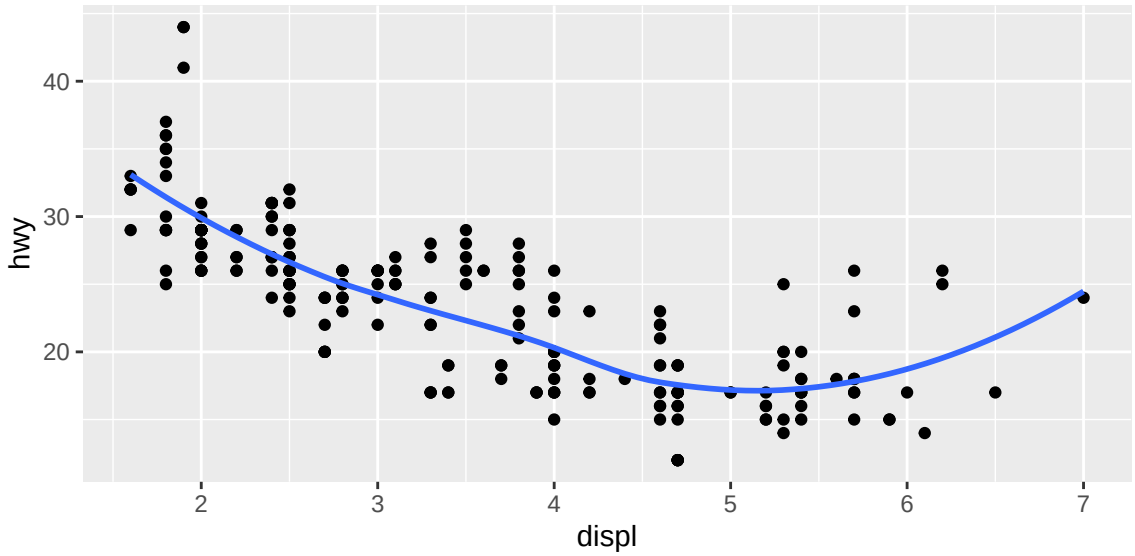
Deveriam ser identicos, pois para o ggplot2 onde você define as aesthetics não afeta o resultado final.

3.6.6

Recreate the R code necessary to generate the following graphs.

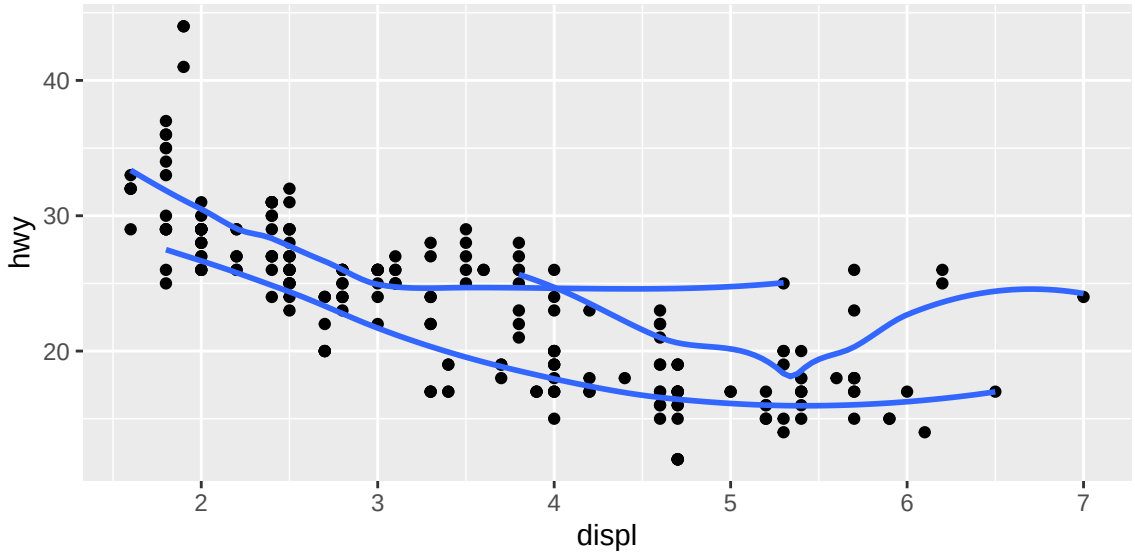
```
ggplot(data = mpg, mapping = aes(x = displ, y = hwy)) +
  geom_point() +
  geom_smooth(se = FALSE)
```

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



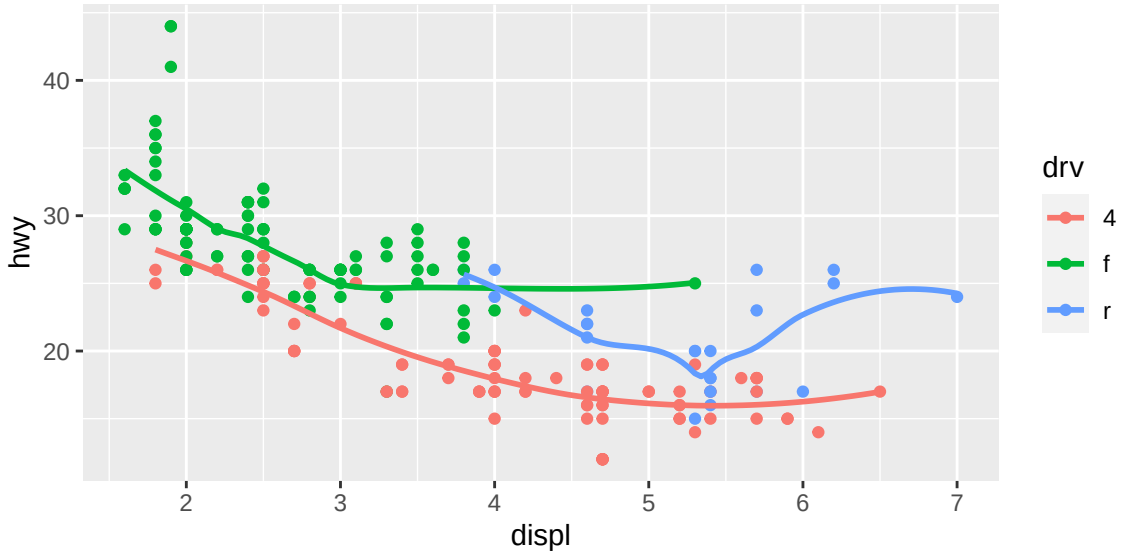

```
ggplot(data = mpg, mapping = aes(x = displ, y = hwy, group = drv)) +
  geom_point() +
  geom_smooth(se = FALSE)
```

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



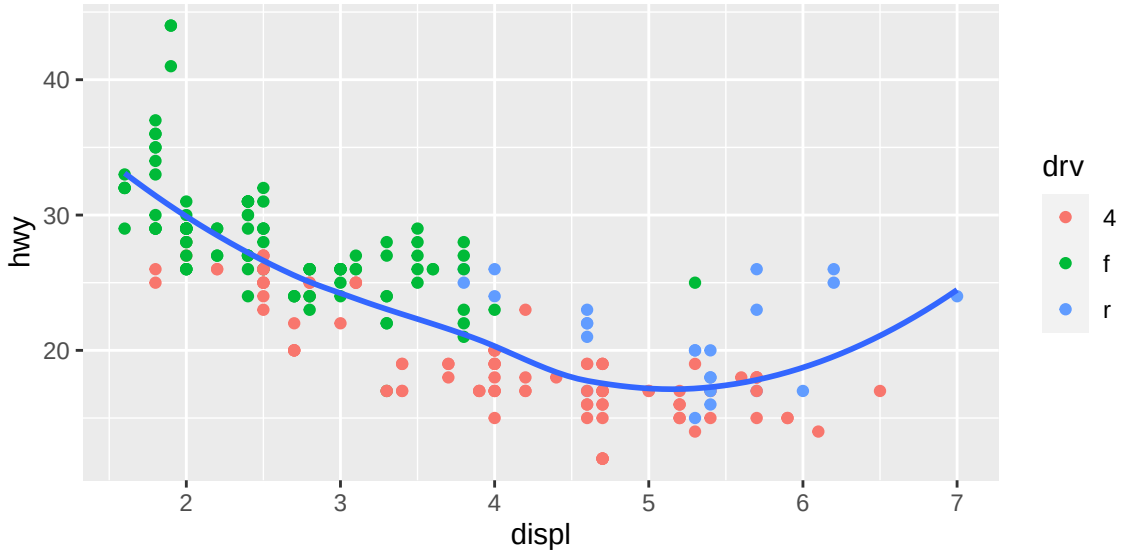

```
ggplot(data = mpg, mapping = aes(x = displ, y = hwy, color = drv)) +
  geom_point() +
  geom_smooth(se = FALSE)
```

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



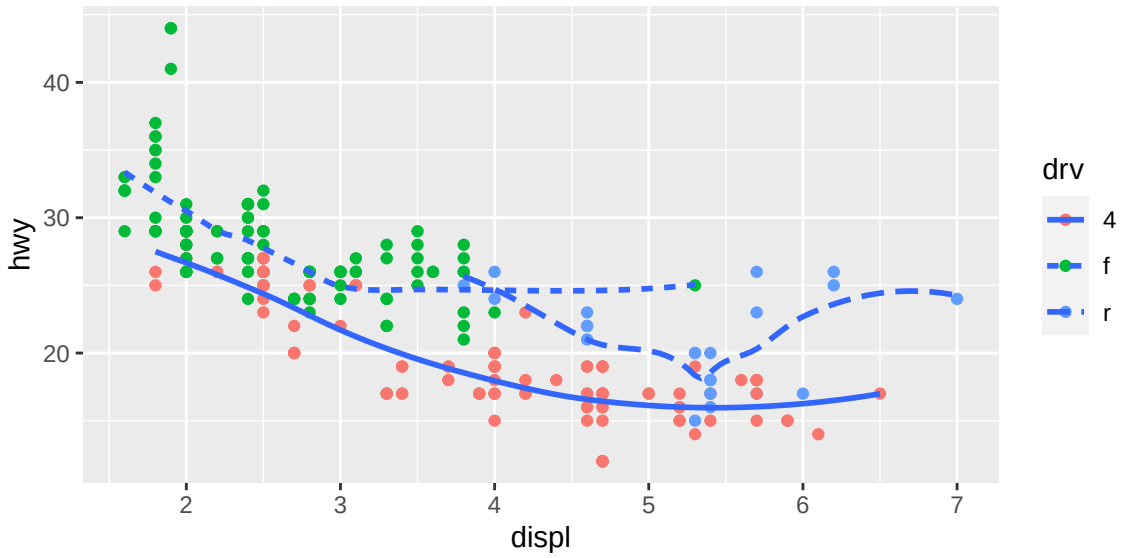

```
ggplot() +
  geom_point(data = mpg, mapping = aes(x = displ, y = hwy, color = drv)) +
  geom_smooth(data = mpg, mapping = aes(x = displ, y = hwy) , se = FALSE)
```

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

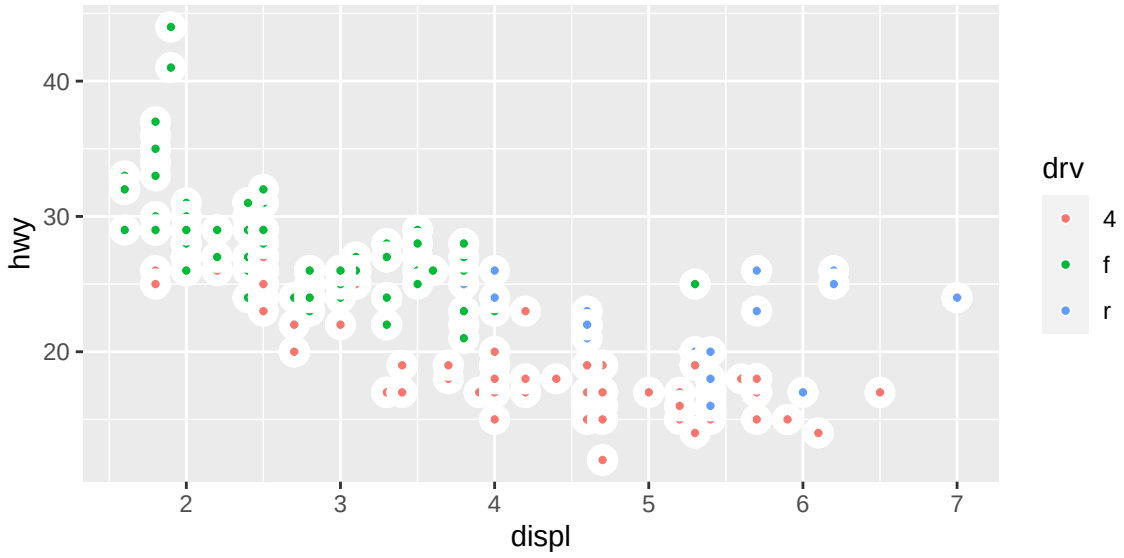



```
ggplot() +
  geom_point(data = mpg, mapping = aes(x = displ, y = hwy, color = drv)) +
  geom_smooth(data = mpg, mapping = aes(x = displ, y = hwy, linetype = drv) , se = FALSE)
```

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




```
ggplot() +
  geom_point(data = mpg, mapping = aes(x = displ, y = hwy, fill = drv, stroke = 3), color = "white" , sh
```



3.7 Statistical Transformations

3.7.1

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?

3.7.2

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

3.7.3

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?

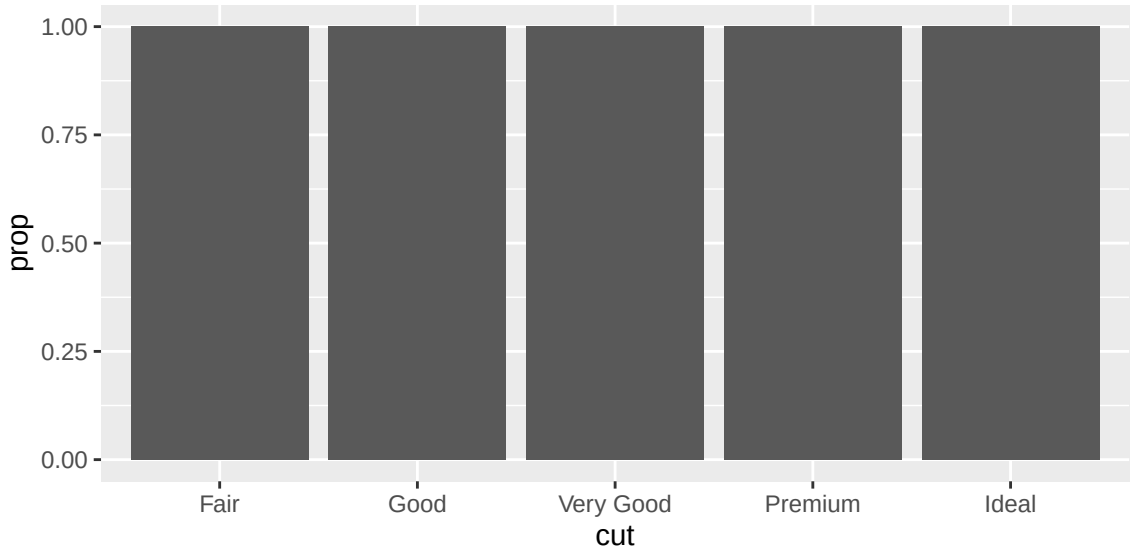
3.7.4

What variables does stat_smooth() compute? What parameters control its behaviour?

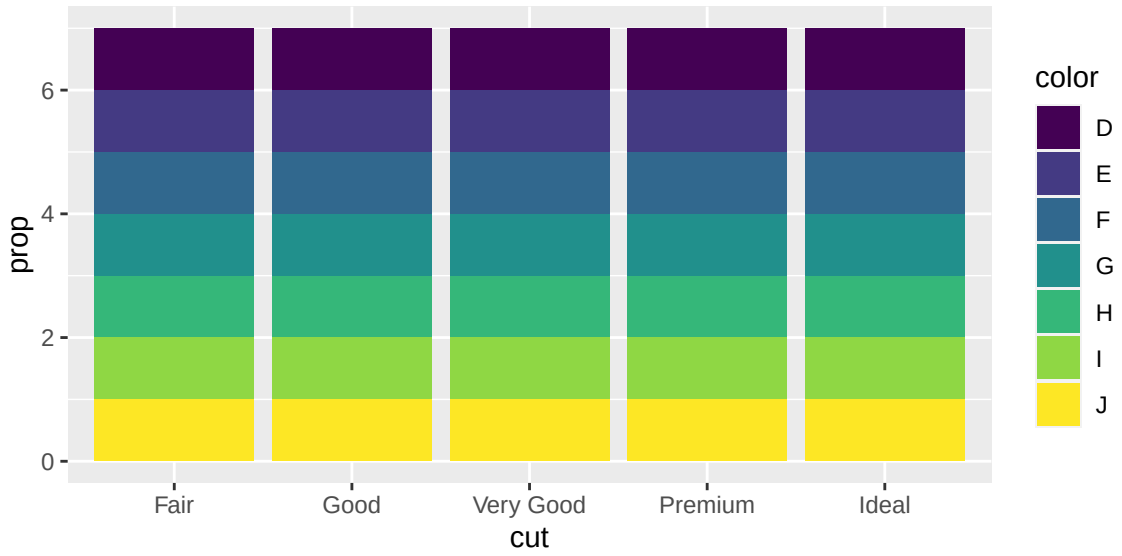
3.7.5

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..))
```

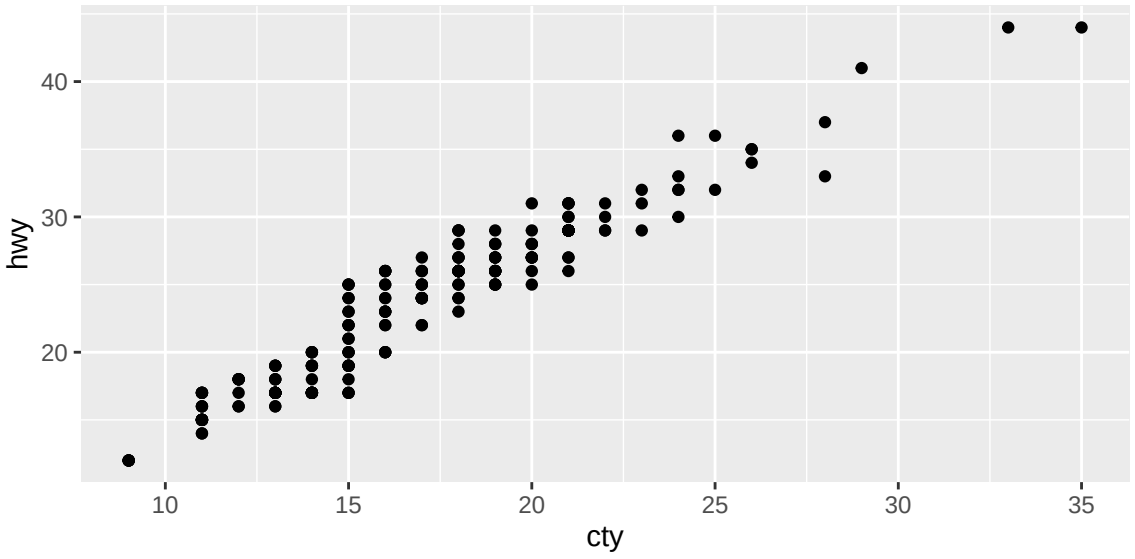

prop
6
4
2
0
Fair
Good
Very Good
Premium
Ideal
cut
color
D
E
F
G
H
I
J

3.8 Position sdjustments

3.8.1

What is the problem with this plot? How could you improve it?

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



3.8.2

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

3.8.3

Compare and contrast geom_jitter() with geom_count().

3.8.4

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

3.9 Coordinate Systems

3.9.1

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

3.9.2

What does labs() do? Read the documentation.

3.9.3

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

3.9.4

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

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