

Introdução ao ggplot2

Tomás Barcellos

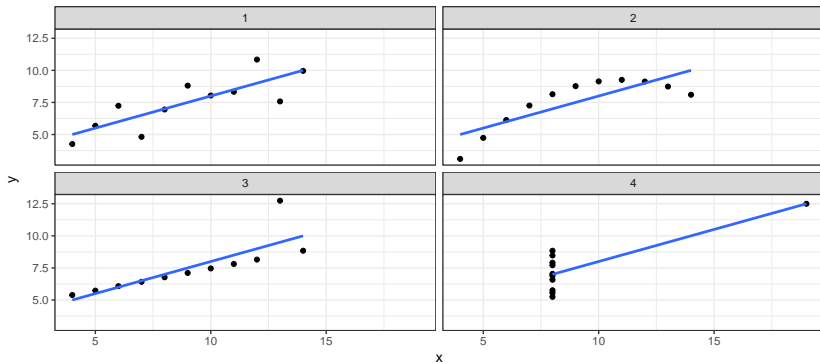
22 de maio de 2018

A gramática dos gráficos

Quarteto de Anscombe

x1	y1	x2	y2	x3	y3	x4	y4
10	8.04	10	9.14	10	7.46	8	6.58
8	6.95	8	8.14	8	6.77	8	5.76
13	7.58	13	8.74	13	12.74	8	7.71
9	8.81	9	8.77	9	7.11	8	8.84
11	8.33	11	9.26	11	7.81	8	8.47
14	9.96	14	8.10	14	8.84	8	7.04
6	7.24	6	6.13	6	6.08	8	5.25
4	4.26	4	3.10	4	5.39	19	12.50
12	10.84	12	9.13	12	8.15	8	5.56
7	4.82	7	7.26	7	6.42	8	7.91
5	5.68	5	4.74	5	5.73	8	6.89

Quarteto de Anscombe



Tipos de gráficos

O gráfico é um meio de comunicação e deve ser adequado ao seu *público* e *objetivo*. É diferente preparar um **gráfico de apresentação** para apresentar conclusões ou fazer um **gráfico exploratório** para você mesmo.

O ggplot2

O ggplot2 é mais do que um pacote para fazer gráficos; ele é a implementação da **gramática dos gráficos em camadas** no R.

Os gráficos criados pelo ggplot2 são compostos por 9 elementos e este tutorial está focado em dois deles: mapeamento de estéticas (aesthetic) e geometrias.

O ggplot2

Por que uma **gramática** dos gráficos?

Através dela podemos definir **sistematicamente** quais são os componentes de um gráfico e como eles se interrelacionam.

Veja mais informações em <http://docs.ggplot2.org/>.

A gramática dos gráficos

Fonte: www.datacamp.com



A gramática dos gráficos

elemento	exemplos
dados (informação)*	produção, temperatura
(a)estética*	cor, formato
geometrias*	barra, ponto
estatísticas	mediana, máximo
facetar	facetar
coordenadas	polar, cartesiana
t(h)emas	eixos, título

* aspetos estéticos imprescindíveis para criar um gráfico no ggplot2

Sintaxe do ggplot2

```
ggplot(um_data_frame,  
       aes(estética1 = variável1,  
           estética2 = variável2,  
           estética3 = variável3)) +  
  geometria(estética4 = "atributo1") +  
  facetas(~variável4) +  
  tema
```

Note que cada função cria uma (ou mais) camadas e que usamos o + para ir adicionando camadas.

Dados

```
# conhecer estrutura e campos dos dados usados  
mtcars$am <- as.factor(mtcars$am)  
mtcars$cyl_fct <- as.factor(mtcars$cyl)  
str(mtcars)
```

Conhecendo os dados

```
## 'data.frame':   32 obs. of  12 variables:
## $ mpg      : num  21 21 22.8 21.4 18.7 18.1 14.3 24.4 22
## $ cyl      : num  6 6 4 6 8 6 8 4 4 6 ...
## $ disp     : num  160 160 108 258 360 ...
## $ hp       : num  110 110 93 110 175 105 245 62 95 123 .
## $ drat     : num  3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3
## $ wt       : num  2.62 2.88 2.32 3.21 3.44 ...
## $ qsec     : num  16.5 17 18.6 19.4 17 ...
## $ vs       : num  0 0 1 1 0 1 0 1 1 1 ...
## $ am       : Factor w/ 2 levels "0","1": 2 2 2 1 1 1 1 1
## $ gear     : num  4 4 4 3 3 3 3 4 4 4 ...
## $ carb     : num  4 4 1 1 2 1 4 2 2 4 ...
## $ cyl_fct  : Factor w/ 3 levels "4","6","8": 2 2 1 2 3 2
```

Aspectos Estéticos

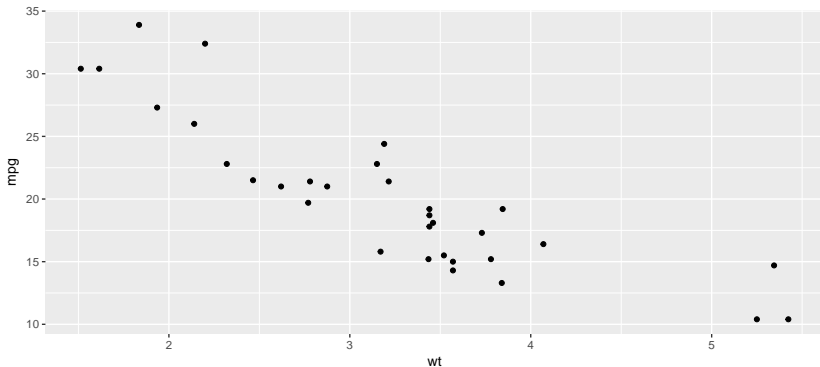
Aspectos Estéticos

Os principais aspectos estéticos são:

Estética	Descrição
x	Eixo horizontal
y	Eixo vertical
colour	Cor dos pontos ou das linhas das formas
fill	Cor de preenchimento
size	Diametro dos pontos e espessura das linhas
alpha	Transparência
linetype	Tipo (padrão) da linhas
labels	Texto no gráfico ou nos eixos
shape	Forma

Representando dados

```
ggplot(mtcars, aes(x = wt, y = mpg)) +  
  geom_point()
```



Representando dados

O computador precisa de critérios para decidir como representar os dados, como o peso de um veículo, em um gráfico.

O peso (`wt`) pode ser representado no eixo vertical ou faixas de valores podem aparecer como cores ou formas. Ao fazer isso estamos mapeando uma variável em uma estética.

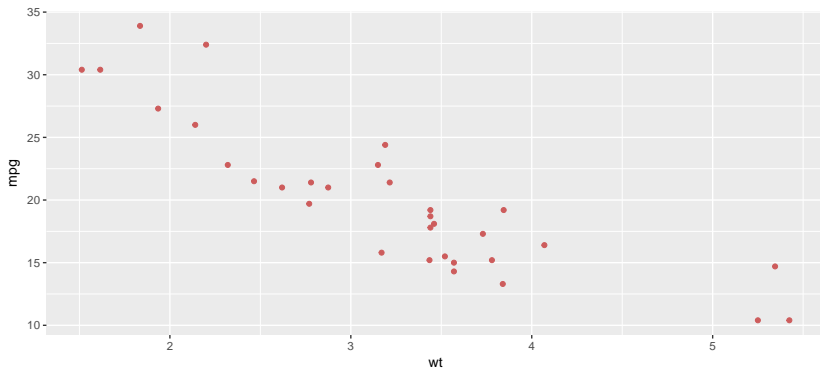
Maapeando variáveis em estéticas

Maapear uma variável em uma estética é determinar que a cor **vermelha** representa carros com câmbio automático e a cor **azul** carros manuais (dentro do `aes(...)`).

Isto é diferente de definir a cor de pontos ou barras como **verde** (fora do `aes(...)`).

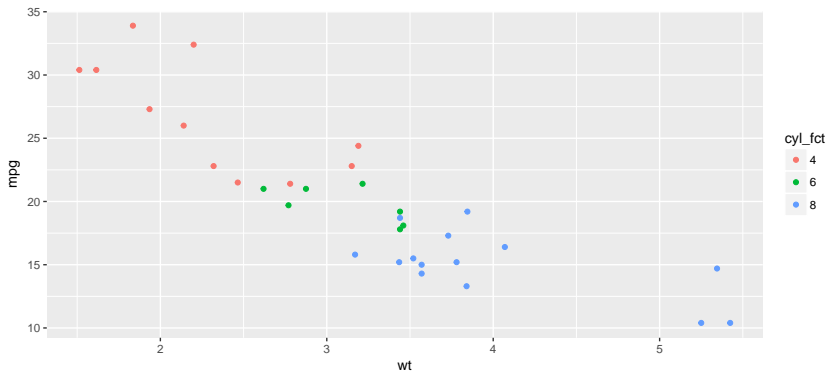
Atributos Estéticos - Atribuir cor à elemento estético

```
ggplot(mtcars, aes(x = wt, y = mpg)) +  
  geom_point(col = "indianred")
```



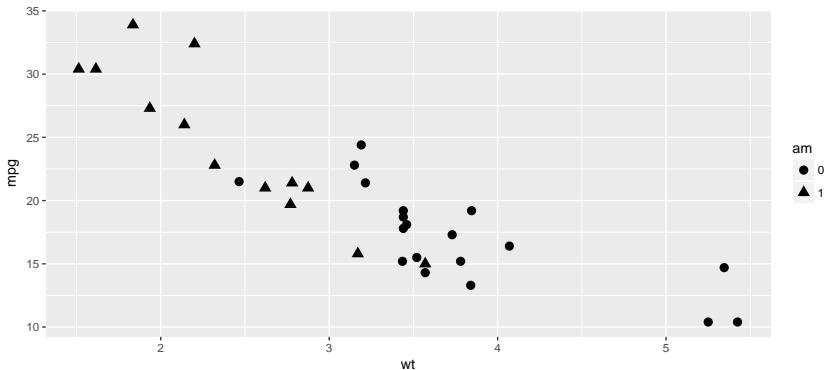
Aspectos Estéticos - Mapear os cilindros na cor

```
ggplot(mtcars, aes(x = wt, y = mpg, col = cyl_fct)) +  
  geom_point()
```



Aspectos Estéticos - Mapear o tipo de cambio na forma

```
ggplot(mtcars , aes(x = wt, y = mpg)) +  
  geom_point(aes(shape = am), size = 3)
```



Aspectos Estéticos - Variáveis contínuas

Estética	Descrição
x	Eixo horizontal
y	Eixo vertical
colour	Cor dos pontos ou das linhas das formas
fill	Cor de preenchimento
size	Diametro dos pontos e espessura das linhas
alpha	Transparência

Aspectos Estéticos - Variáveis contínuas

Fonte: www.datacamp.com

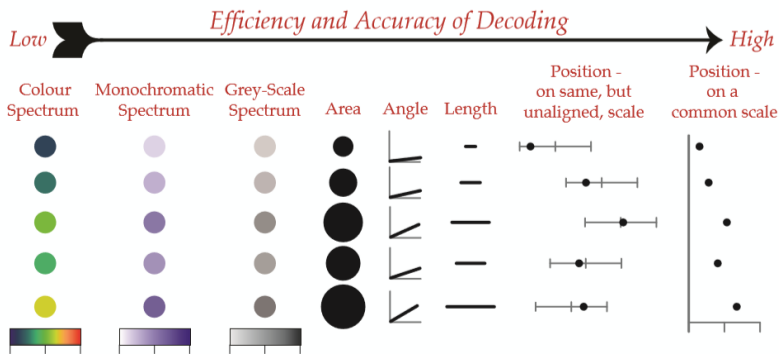
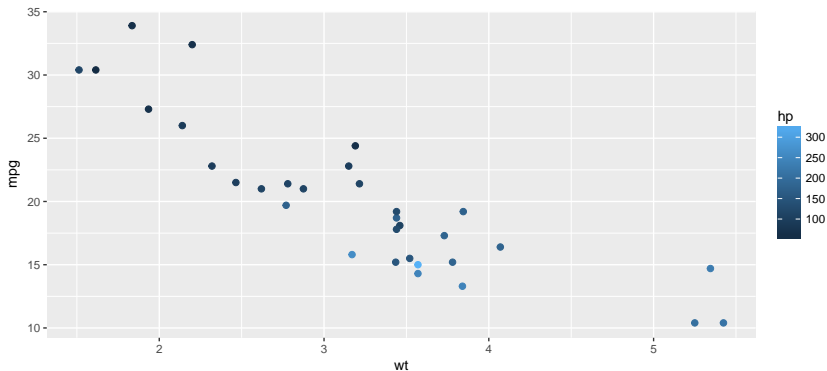


Figure 2:

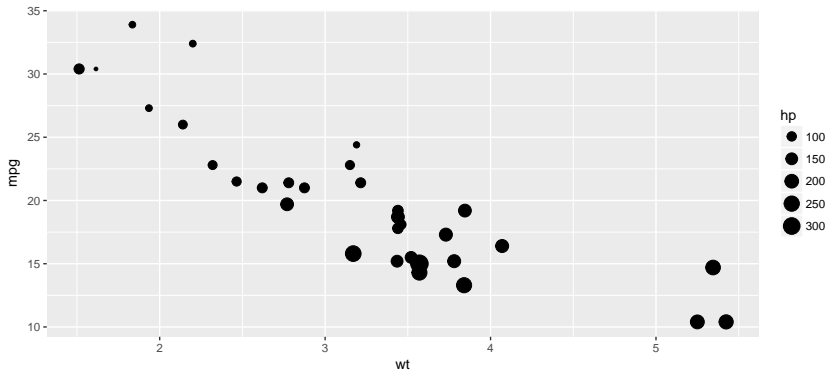
Aspectos Estéticos - Variáveis contínuas

```
ggplot(mtcars, aes(x = wt, y = mpg, col = hp)) +  
  geom_point(size = 2)
```



Aspectos Estéticos - Variáveis contínuas

```
ggplot(mtcars, aes(x = wt, y = mpg, size = hp)) +  
  geom_point()
```

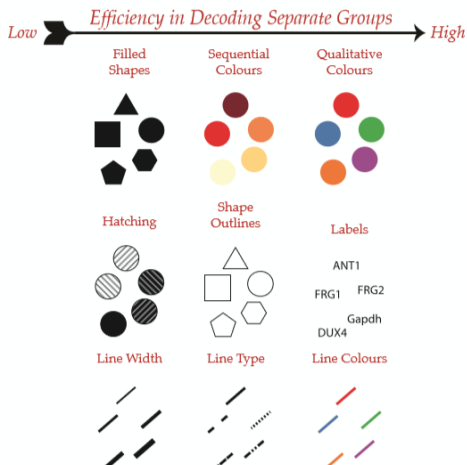


Aspectos Estéticos - Variáveis categóricas

Estética	Descrição
colour	Cor dos pontos ou das linhas das formas
fill	Cor de preenchimento
size	Diametro dos pontos e espessura das linhas
alpha	Transparência
linetype	Tipo (padrão) da linhas
labels	Texto no gráfico ou nos eixos
shape	Forma

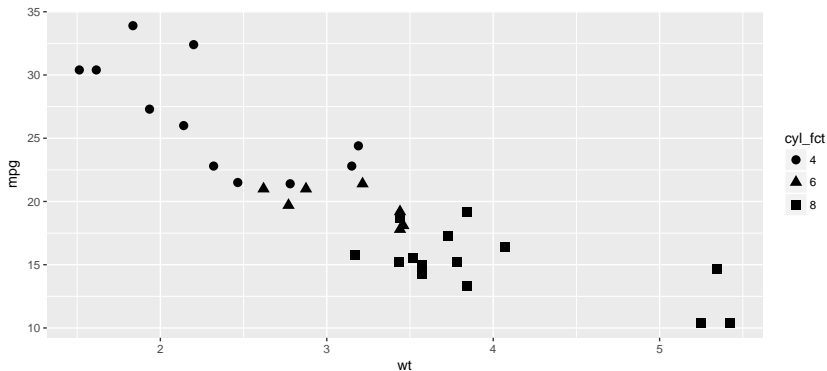
Atributos Estéticos - Variáveis categóricas

Fonte: www.datacamp.com



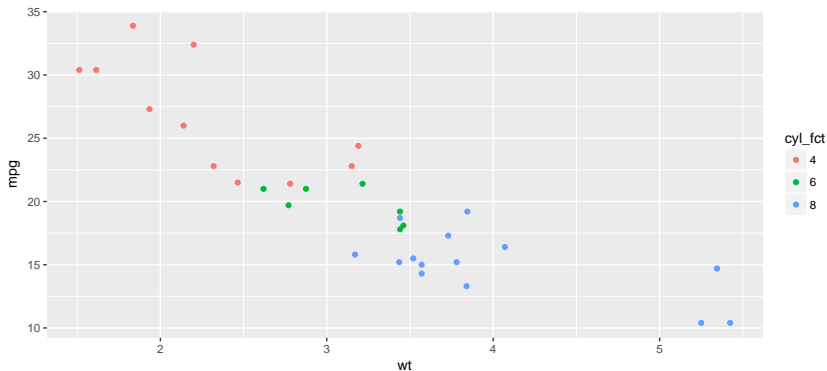
Aspectos Estéticos - Variáveis categóricas

```
ggplot(mtcars, aes(x = wt, y = mpg, shape = cyl_fct)) +  
  geom_point(size = 3)
```



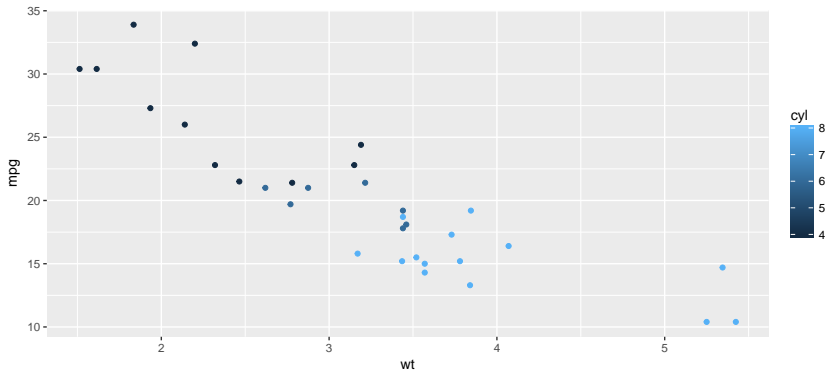
Aspectos Estéticos - Variáveis categóricas

```
ggplot(mtcars, aes(x = wt, y = mpg, col = cyl_fct)) +  
  geom_point()
```



Aspectos Estéticos - Variáveis categóricas

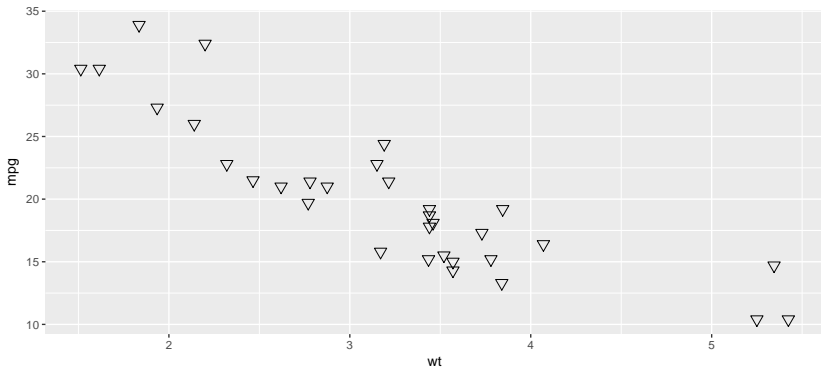
```
ggplot(mtcars, aes(x = wt, y = mpg, col = cyl)) +  
  geom_point()
```



Geometrias

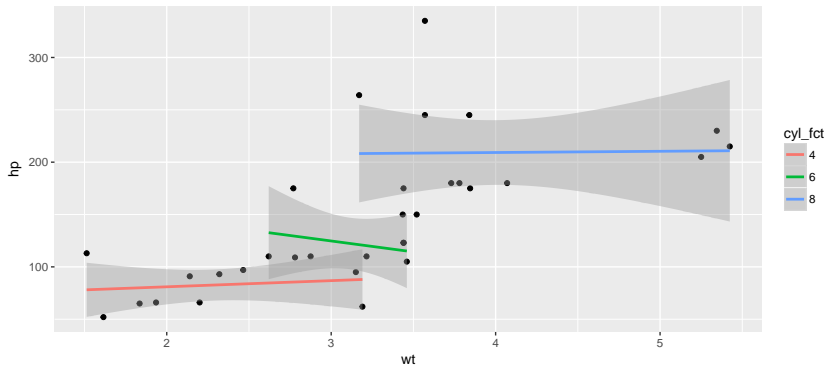
geom_point()

```
ggplot(mtcars, aes(x = wt, y = mpg)) +  
  geom_point(shape = 6, size = 3)
```



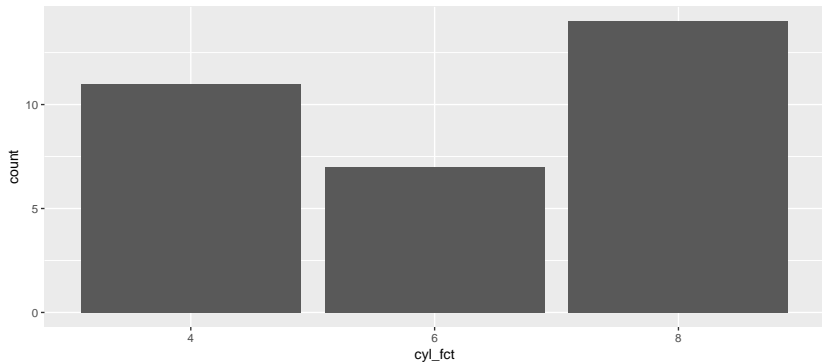
geom_smooth()

```
ggplot(mtcars, aes(x = wt, y = hp)) +  
  geom_point() +  
  geom_smooth(aes(col = cyl_fct), method = "lm")
```



geom_bar()

```
ggplot(mtcars, aes(x = cyl_fct)) +  
  geom_bar()
```

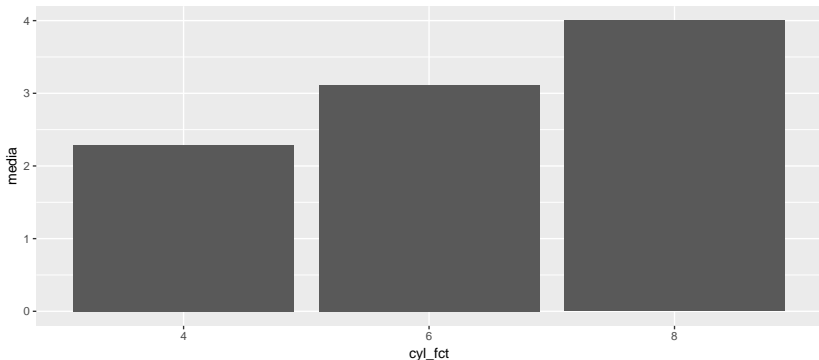


geom_col() ou geom_bar()

```
medias <- mtcars %>% group_by(cyl_fct) %>%  
  summarise(media = mean(wt, na.rm = TRUE))
```

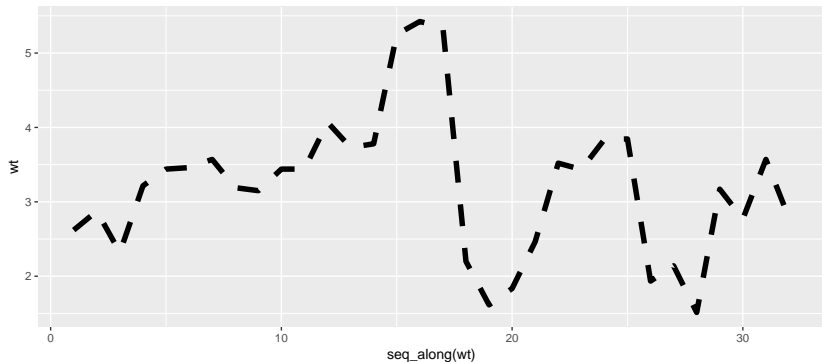
geom_col() ou geom_bar()

```
ggplot(medias, aes(x = cyl_fct, y = media)) +  
  geom_col() # ou geom_bar(stat = "identity")
```



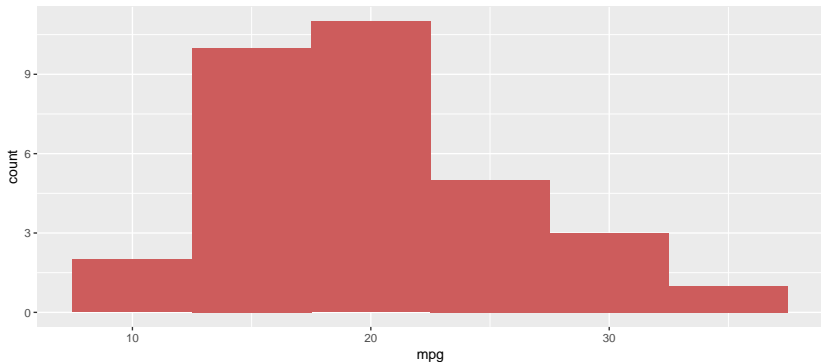
geom_line()

```
ggplot(mtcars, aes(seq_along(wt), wt)) +  
  geom_line(group = 1, size = 2, linetype = 2)
```



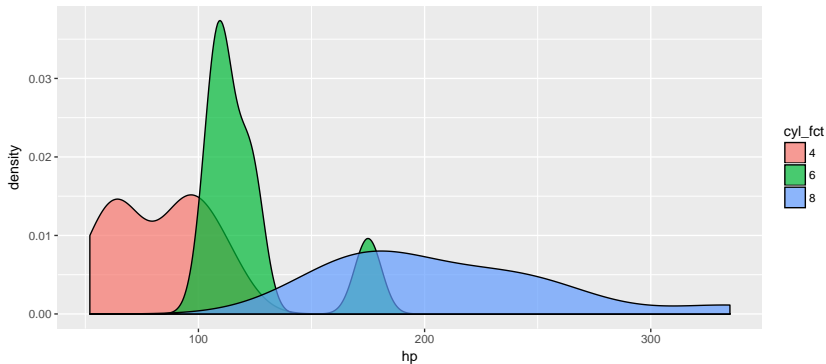
geom_histogram()

```
ggplot(mtcars, aes(x = mpg)) +  
  geom_histogram(fill = "indianred", binwidth = 5)
```



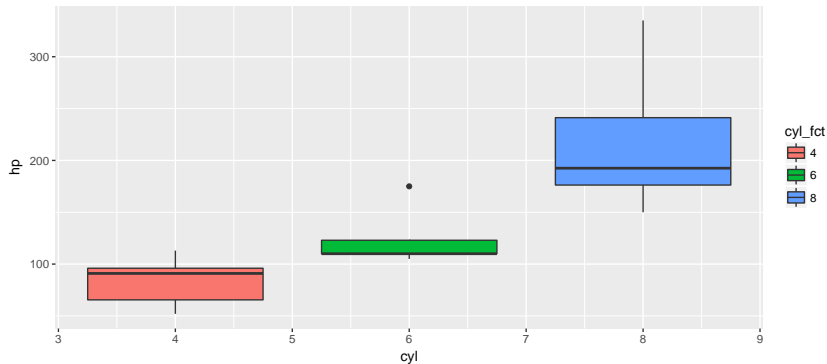
Ou então `geom_density()`

```
ggplot(mtcars, aes(x = hp, fill = cyl_fct)) +  
  geom_density(alpha= 0.7)
```



geom_boxplot()

```
ggplot(mtcars, aes(x = cyl, y = hp)) +  
  geom_boxplot(aes(fill = cyl_fct))
```



... ou `geom_violin()`

```
ggplot(mtcars, aes(x = cyl, y = hp, fill = cyl_fct)) +  
  geom_violin(draw_quantiles = 0.5)
```

