

Graphics with ggplot2

NHS R Conference 2020

Rhian Davies (@trianglegirl)



Introduction to {ggplot2}

- <http://bit.ly/nhs-ggplot2>

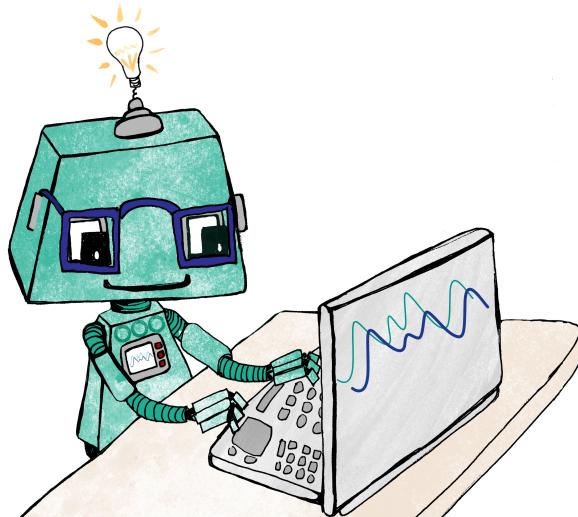




Who am I?



Jumping Rivers



- On-site training
- R and python consultancy
- Dashboard creation
- Code review
- Questionnaire design
- R Package development
- Predictive analytics
- Grant applications



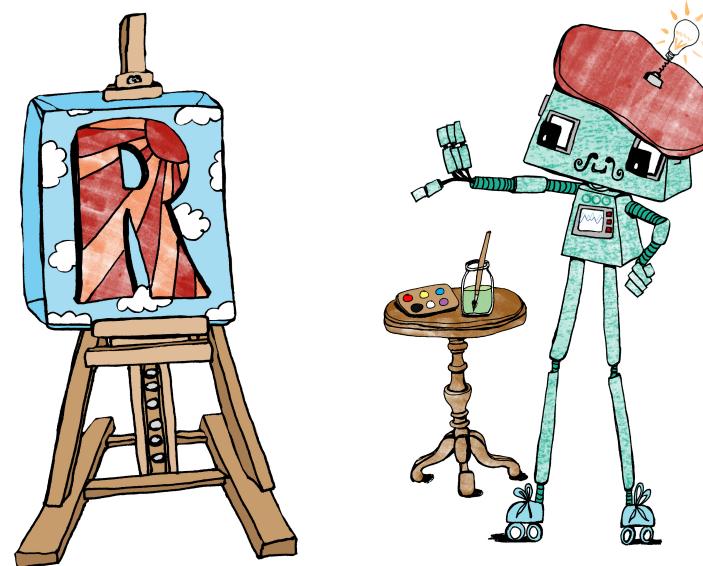
Our clients





Introduction

- *Many* different ways to make graphs in R
- {ggplot2} started in 2005 and follows the "Grammar of Graphics"
- Many companies have adopted {ggplot2} for graphics, including the BBC and the FT
- Think about graphics in terms of **layers**





The basic plot object

- Load the package with

```
library("ggplot2")
```

- Create an initial ggplot object, using `ggplot()`
- This function has two arguments:
- **data**: this must be a data frame (or tibble)
- an aesthetic **mapping**: this tells `{ggplot2}` how to map data to the graphical elements



Setting up the plot

```
movies = readRDS("data/movies.rds")
```

- The function `aes()` maps our data to the graph
- Here, duration is mapped to the x-axis

```
g = ggplot(data = movies,  
            mapping = aes(x = duration, y = rating))
```

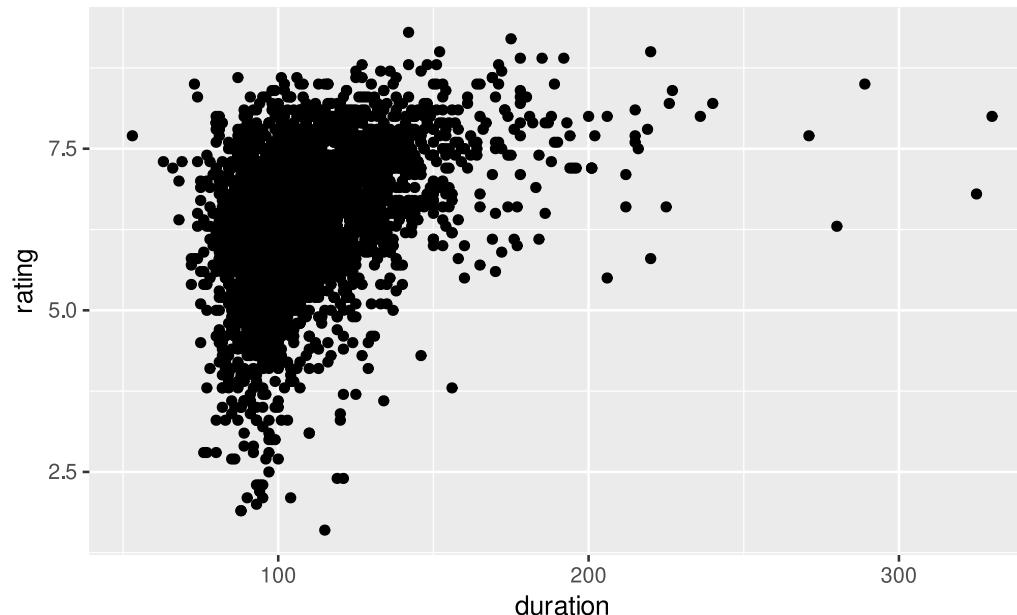
Notice we can store graphs in variables.



Scatter plots

- To add information we need to add a `geom`.
- Can have multiple `geoms` on a graph

```
h = ggplot(movies, aes(x = duration, y = rating))  
h + geom_point()
```

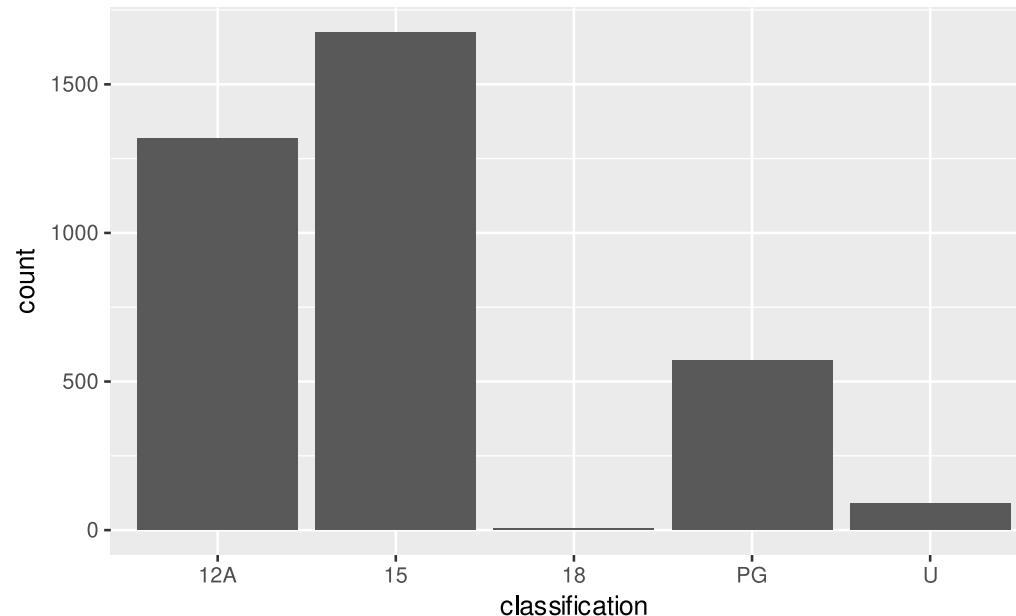




Bar charts

- Great for displaying qualitative data.
- Length of the bar represents the frequency

```
ggplot(movies, aes(x = classification)) +  
  geom_bar()
```



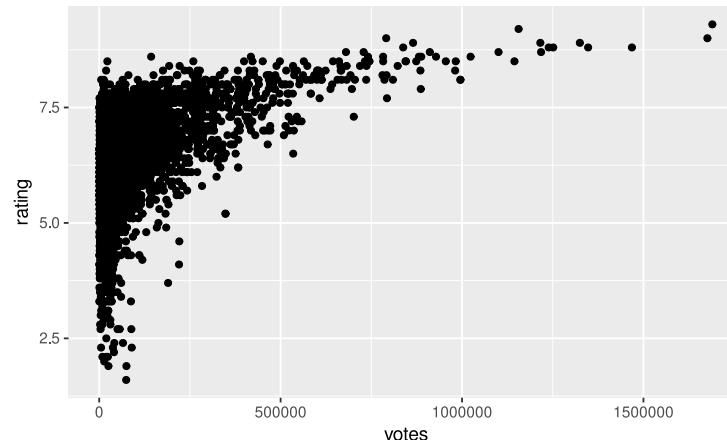


Quick Quiz!



Spot the mistakes

Sarah is trying to recreate the following plot, but her code doesn't work



Spot three errors in her code and think about how you would fix them.

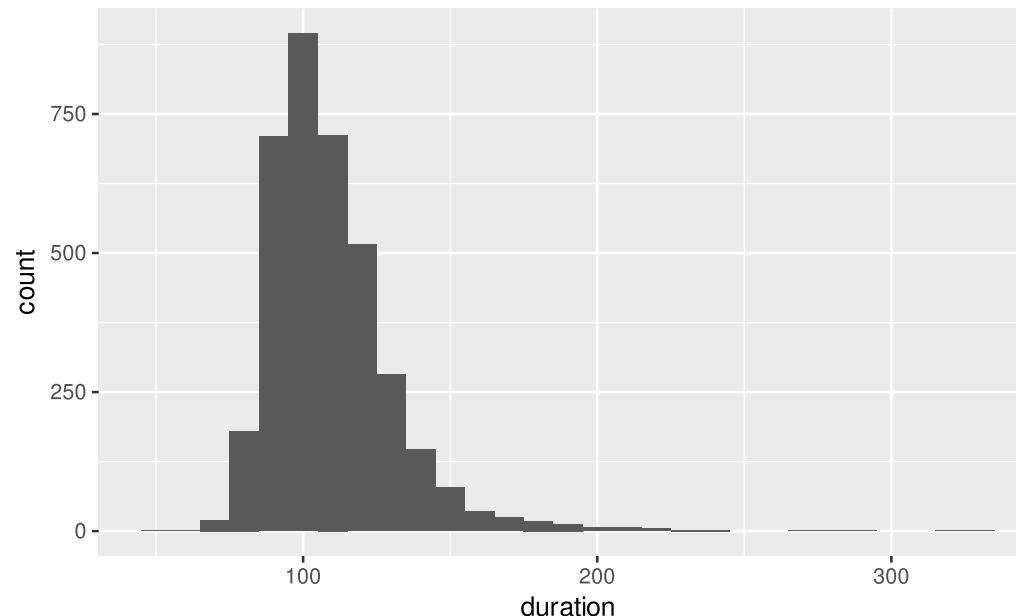
```
ggplot(data = movies, x = votes, y = rating) %>%  
  geom_scatter()
```



Histograms

- Good for plotting continuous variables.
- Data is split up into intervals called **classes**.
- Area of the columns represent the frequencies in the classes.

```
ggplot(movies, aes(x = duration)) + geom_histogram(binwidth = 10)
```

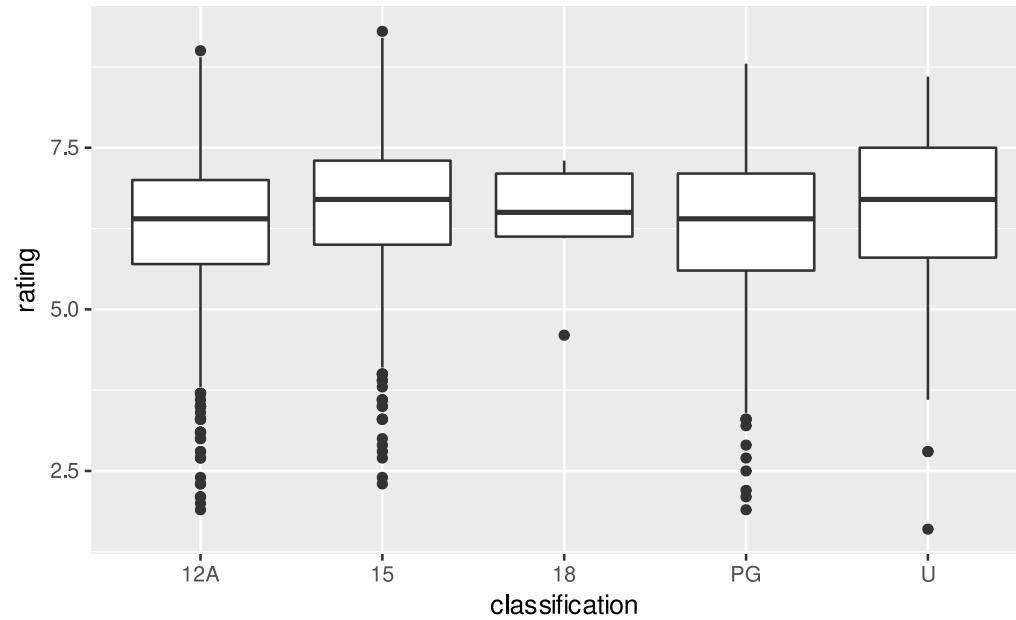




Box and whisker plots

- Central bar represents the median
- Top and bottom of the box represents the lower and upper quartiles

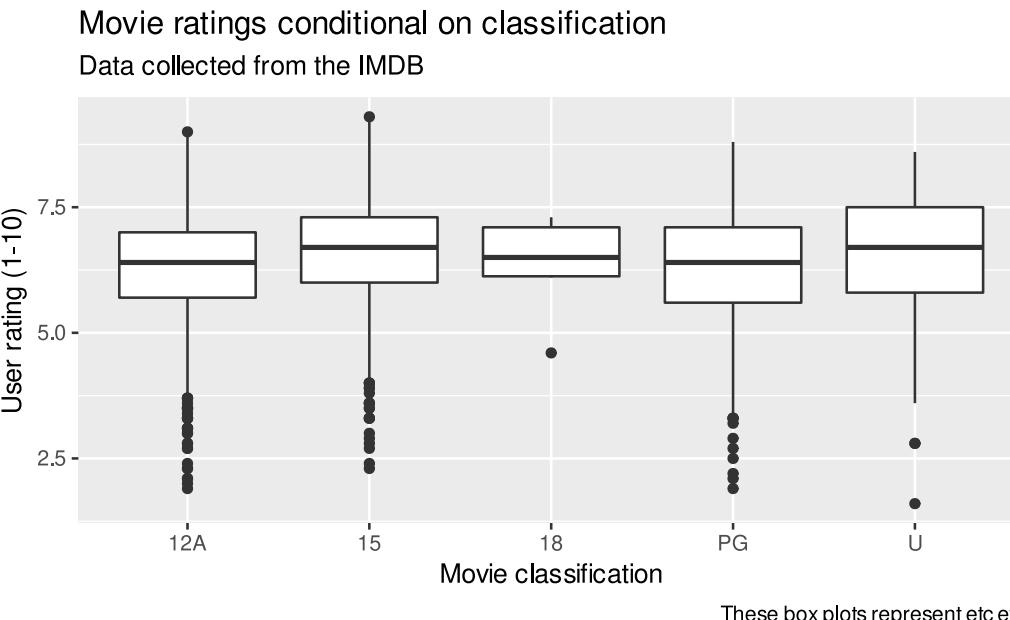
```
boxplot = ggplot(movies, aes(x = classification, y = rating)) +  
  geom_boxplot()
```





Titles and legends

```
boxplot + labs(x = "Movie classification",
                y = "User rating (1-10)",
                title = "Movie ratings conditional on classification",
                subtitle = "Data collected from the IMDB",
                caption = "These box plots represent etc etc")
```





Saving graphs to files

- So how do you save all your fantastic plots?
- RStudio has the *export* button
- It's also possible to export using code (better for reproducibility)



Practical



Aesthetics

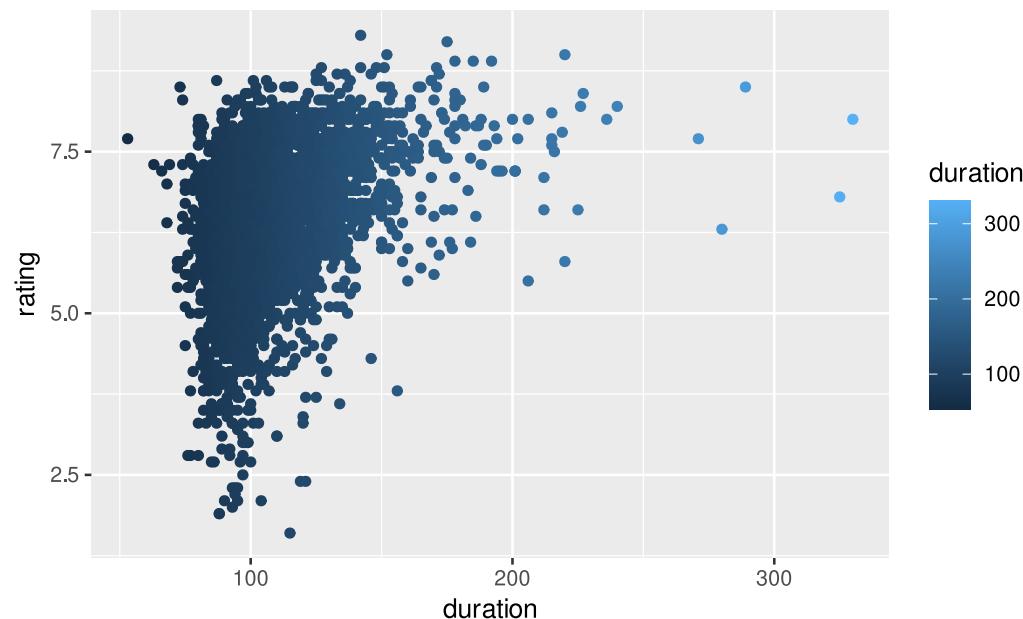
There are many different types of aesthetics

- Colour/fill 
- Shape 
- Size 
- Linetype 
- Alpha 



Colours

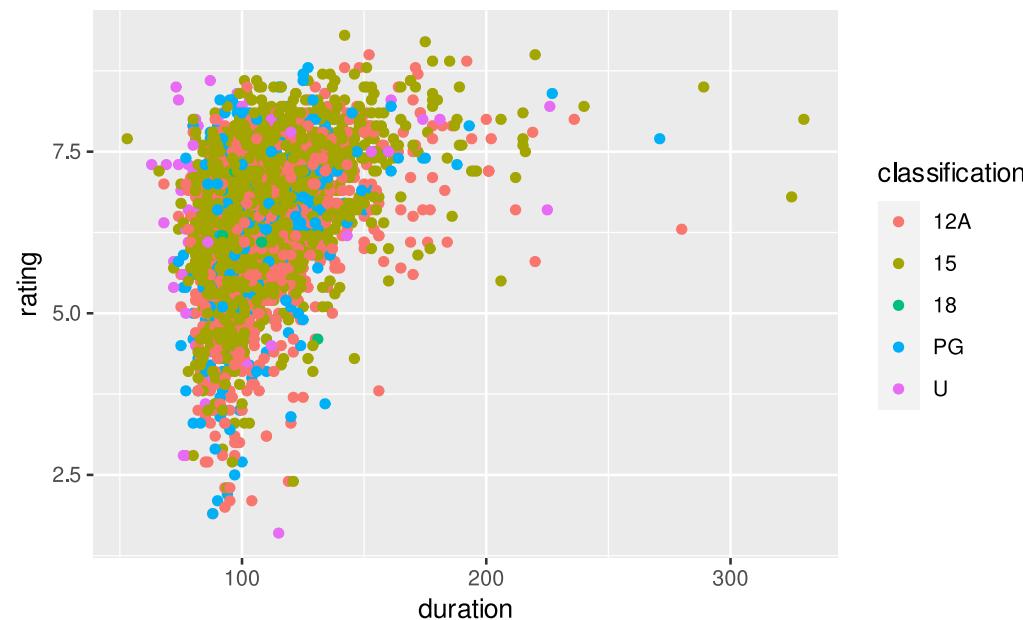
```
ggplot(movies, aes(x = duration, y = rating)) +  
  geom_point(aes(colour = duration))
```





Aesthetics

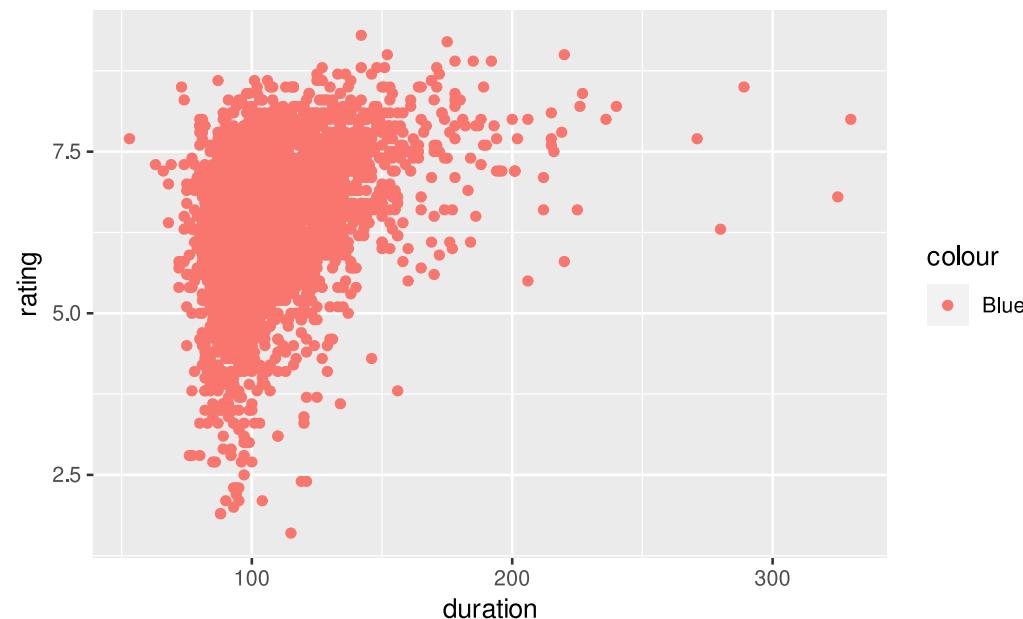
```
ggplot(movies, aes(x = duration, y = rating)) +  
  geom_point(aes(colour = classification))
```





Aesthetics

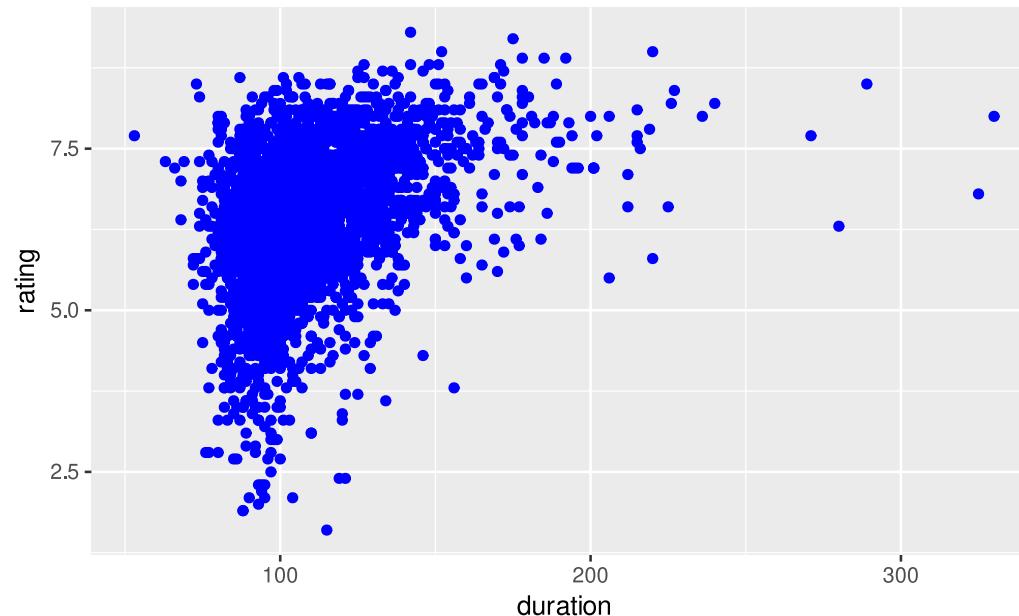
```
ggplot(movies, aes(x = duration, y = rating)) +  
  geom_point(aes(colour = "Blue"))
```





Aesthetics

```
ggplot(movies, aes(x = duration, y = rating)) +  
  geom_point(colour = "Blue")
```





Remember!

Inside aes():

Using columns of the data to style. E.g. `aes(colour = classification)`

Outside aes():

Fixed styling. E.g. `colour = "blue"`



Shapes

□ 0

◇ 5

⊕ 10

■ 15

■ 22

○ 1

▽ 6

⊗ 11

● 16

● 21

△ 2

⊗ 7

田 12

▲ 17

▲ 24

+ 3

* 8

⊗ 13

◆ 18

◆ 23

× 4

◇ 9

田 14

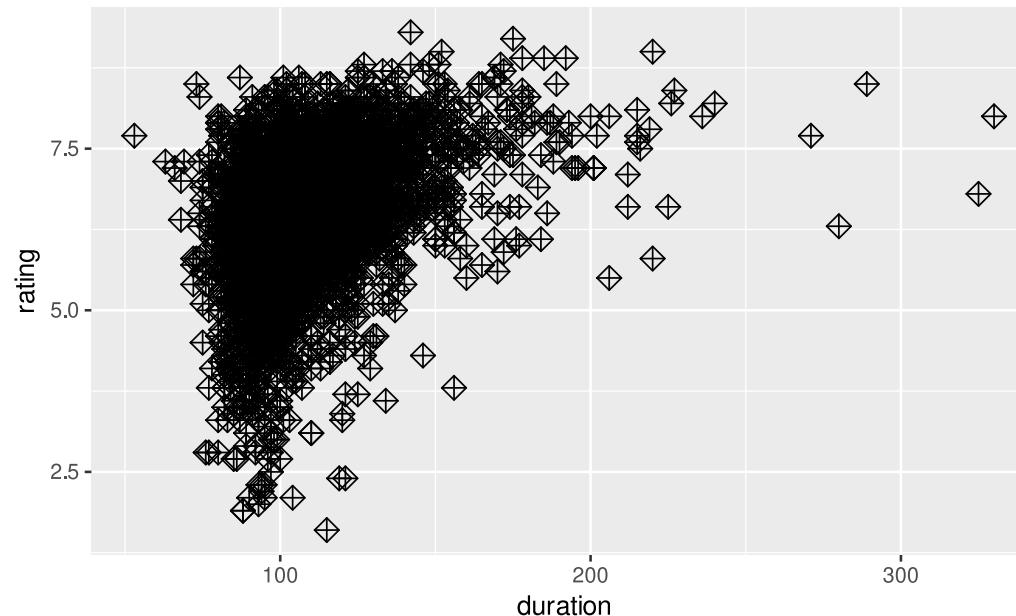
● 19

● 20



Change shape for *all* points

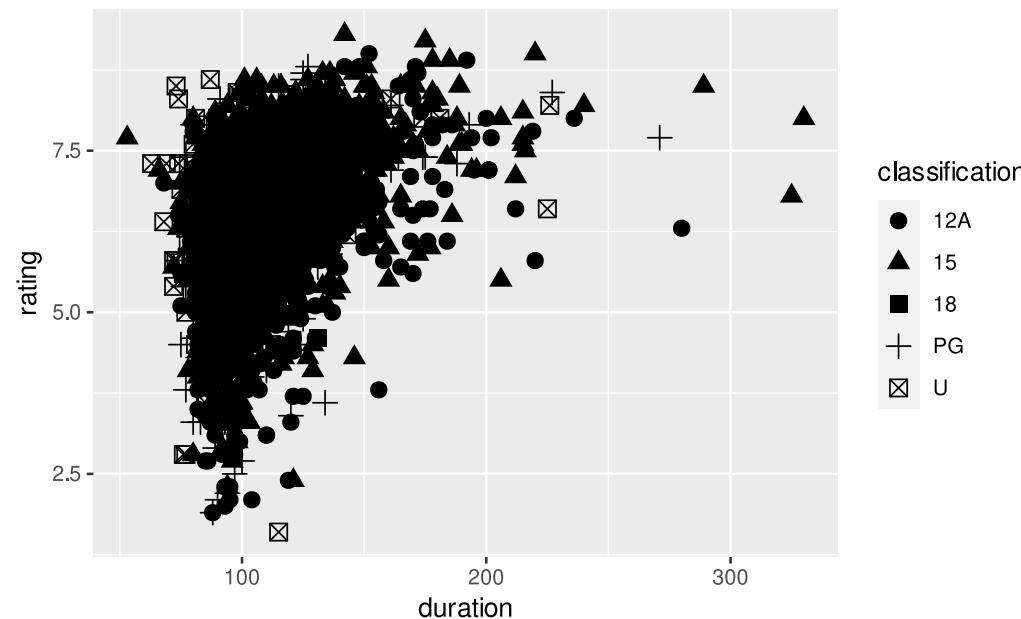
```
ggplot(movies, aes(x = duration, y = rating)) +  
  geom_point(shape = 9, size = 3)
```





Change shape based on the data

```
ggplot(movies, aes(x = duration, y = rating)) +  
  geom_point(aes(shape = classification), size = 3)
```





Quick Quiz!



Fill in the blanks

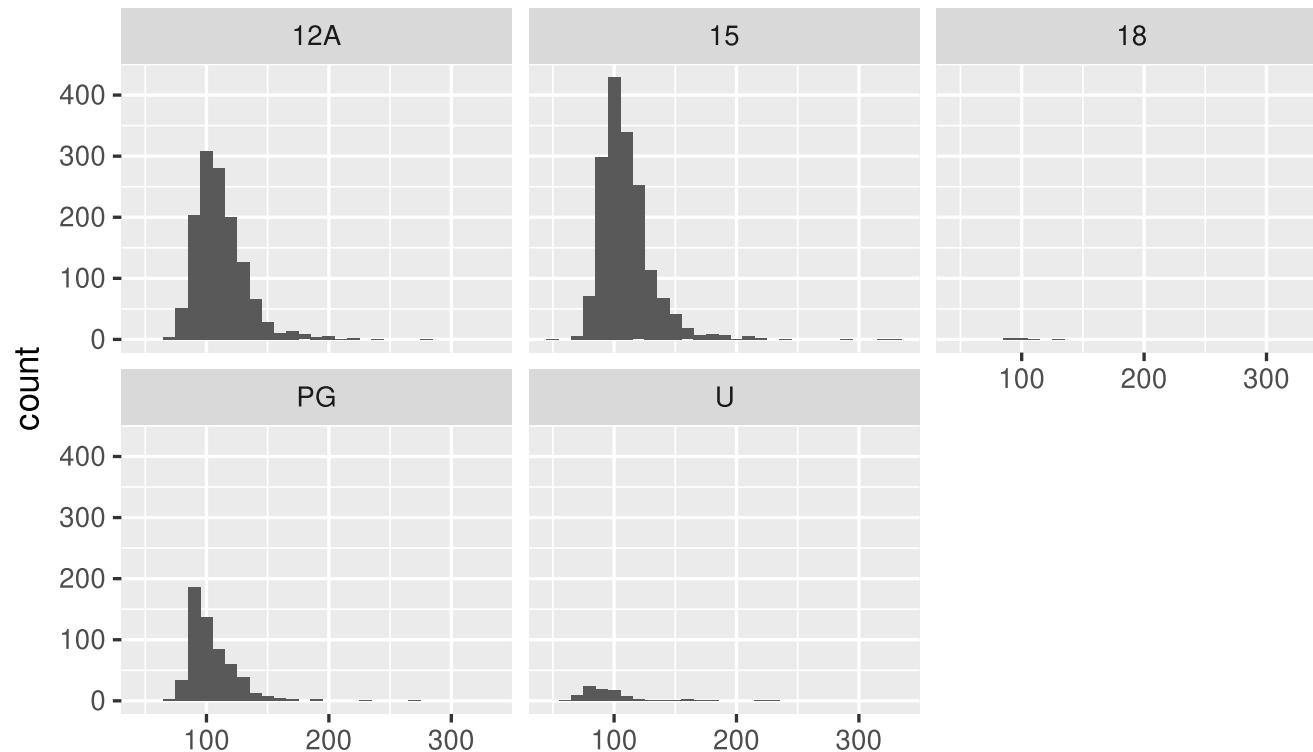
```
ggplot(movies, aes(x = year, y = rating)) +  
  geom_point(aes(___ = classification,  
                size = ___),  
             ___ = 0.3)
```





Facets - mini plots!

```
ggplot(movies, aes(x = duration)) +  
  geom_histogram(binwidth = 10) +  
  facet_wrap(~classification)
```

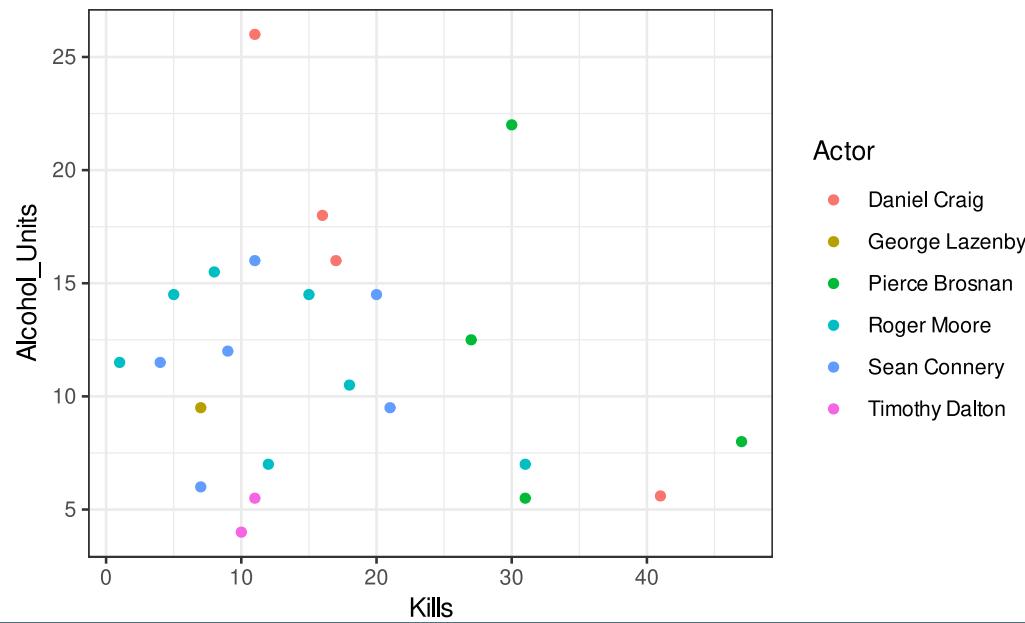




Themes

- Quick and easy way to add style to your plots.
- Eight in-built themes with {ggplot2}

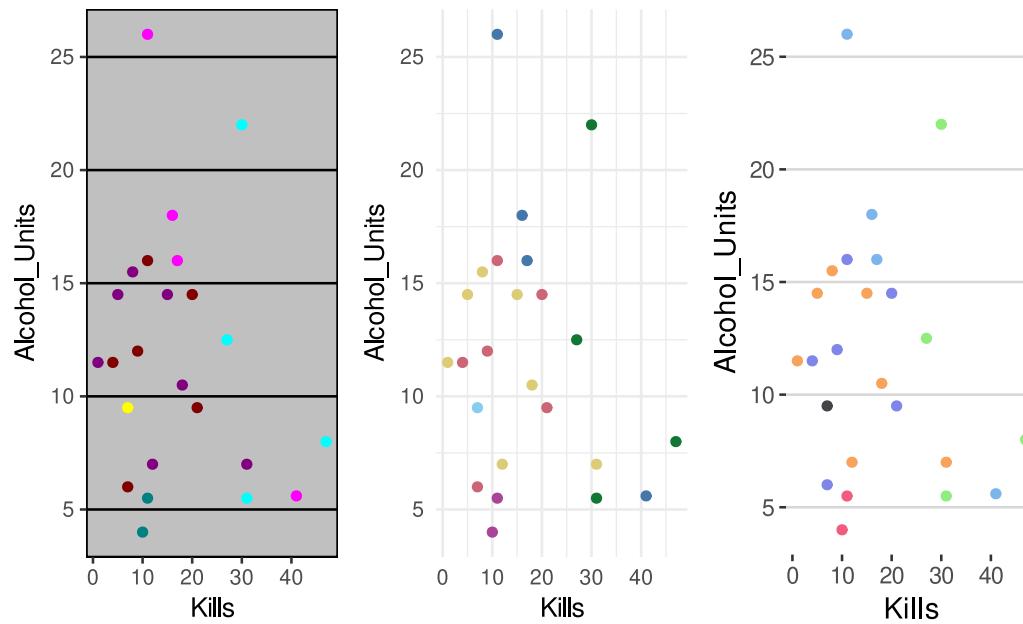
```
ggplot(bond, aes(Kills, Alcohol_Units)) +  
  geom_point(aes(colour = Actor)) +  
  theme_bw()
```





Other themes: ggthemes

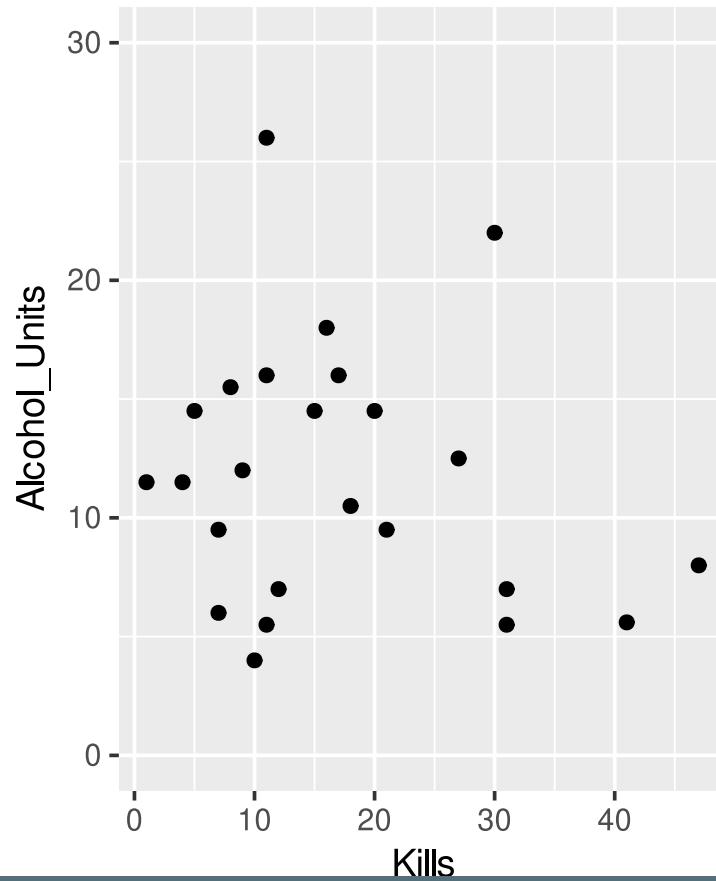
```
library("ggthemes")  
gg + theme_excel() + scale_colour_excel()  
gg + theme_minimal() + scale_color_ptol()  
gg + theme_hc() + scale_colour_hc()
```





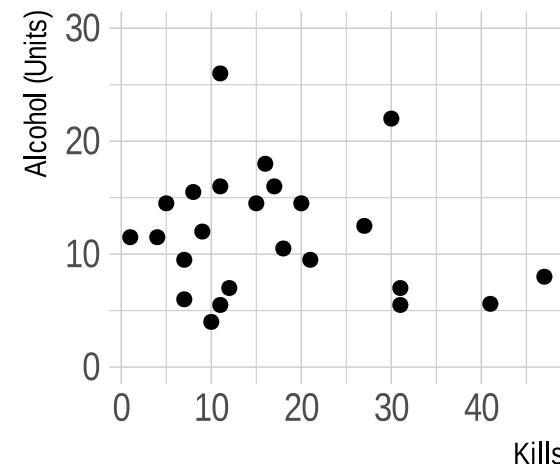
Other themes: hrbrthemes

See <https://github.com/hrbrmstr/hrbrthemes>



The Bond data set

Kills versus Alcohol consumption

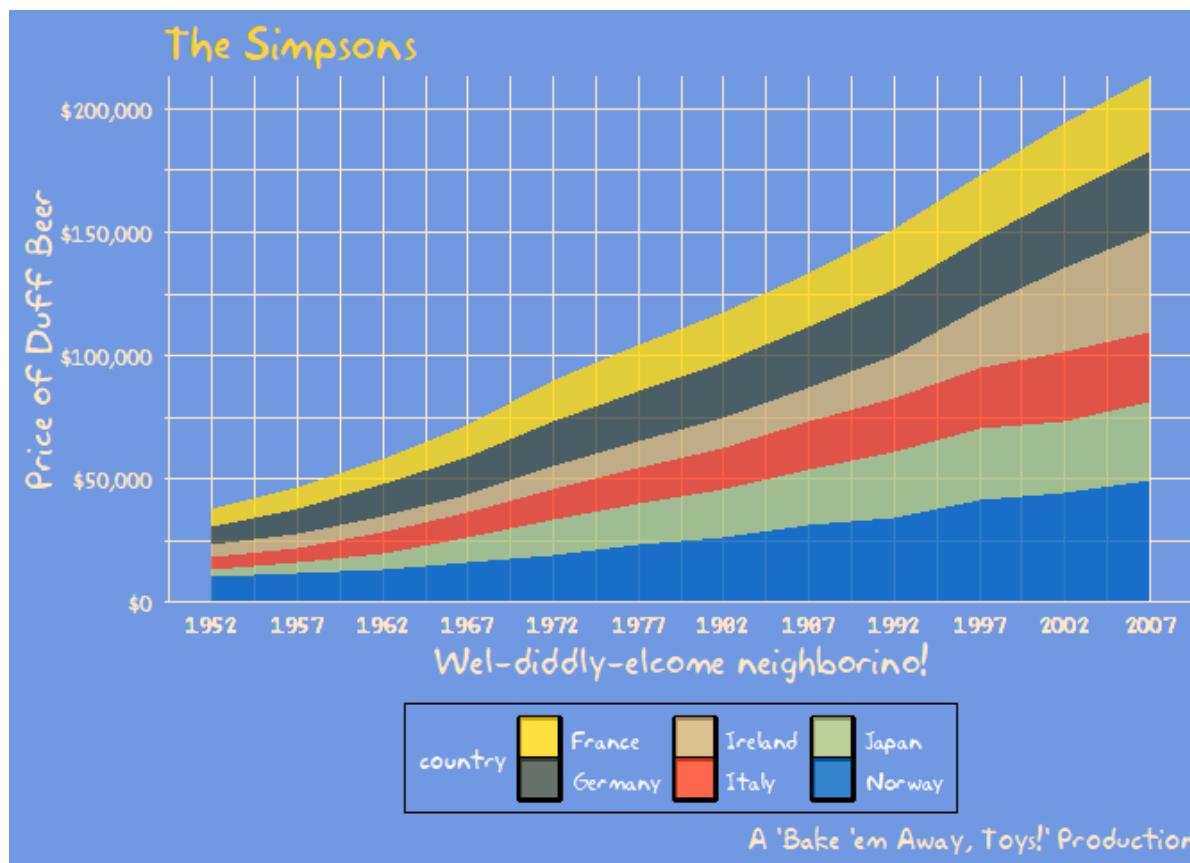


The name's Bond, James Bond



Other themes: {tvthemes}

You can even style your plot based on your favourite TV show with `{tvthemes}`





Practical



Top troubleshooting tips



- {ggplot2} is the , ggplot() is the function
- colour = "blue" vs aes(colour = gender)
- Round brackets! Geoms are functions. E.g. geom_point()
- Remember the +
- Are you plotting something sensible (is your data categorical?)
- Try manipulating your data **before** plotting it
-  Just x - geom_bar(), x and y - geom_col()



Links & resources



- R4DS book chapter 3
- R graphics cookbook
- RStudio {ggplot2} cheatsheet
- #TidyTuesday
- Our training courses

Share any cool plots you make on twitter!

@jumping_uk @trianglegirl #NHSR #NHSRconf2020

