R: A Hitchhikers Guide to Reproducible Research

- Welcome to the ggungle

Brendan Palmer,
Clinical Research Facility - Cork &
School of Public Health
@B_A_Palmer

CRF-C
HRB Clinical Research Facility Cork

UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

School of
Public Health

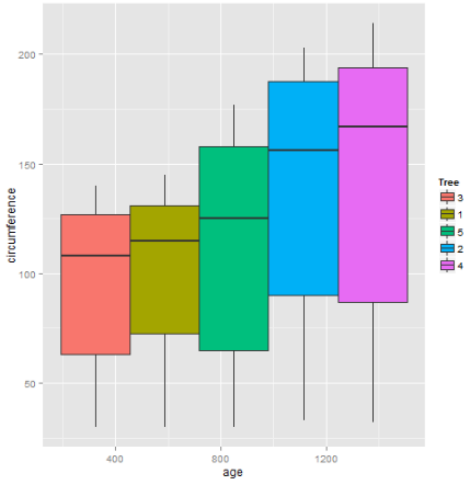
ggplot2

- Data visualisation based on "The Grammar of Graphics"

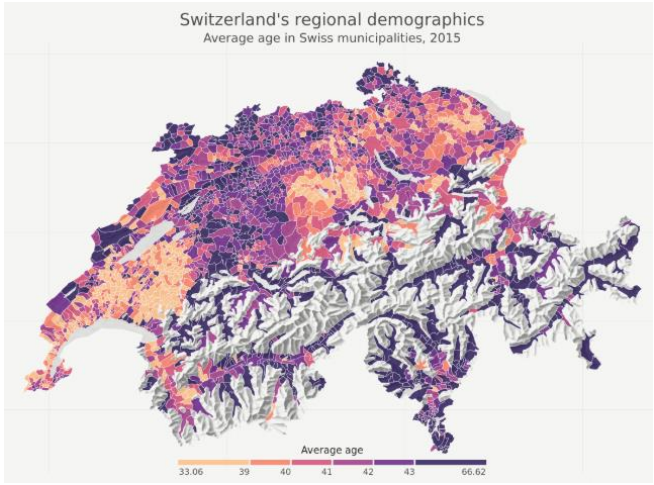
```
ggplot(data = <DATA>) +

    <GEOM_FUNCTION>(mapping = aes(<MAPPINGS>))    +

     linear model  +

     axes formatting  +

     legend formatting  +

     title  + etc. etc.
```

ggplot2

- Very versatile
- Allows you to go from

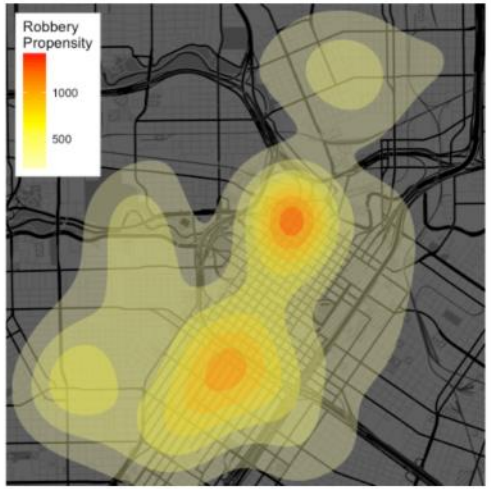


to

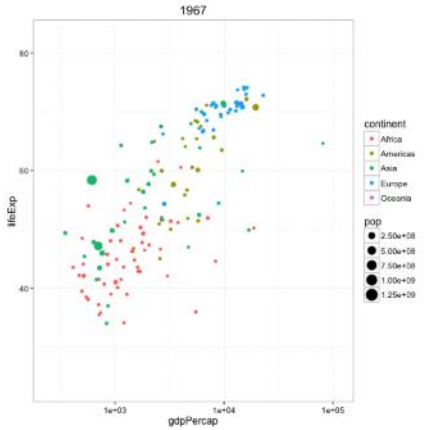


- Lots of add-on packages

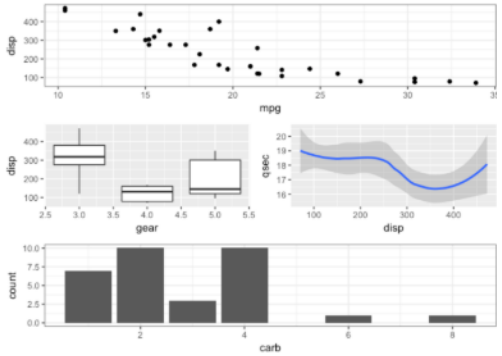
ggmap



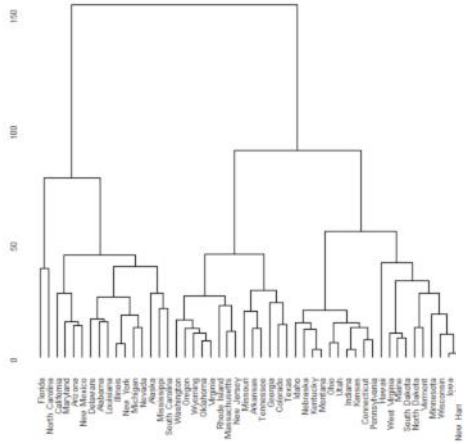
gganimate



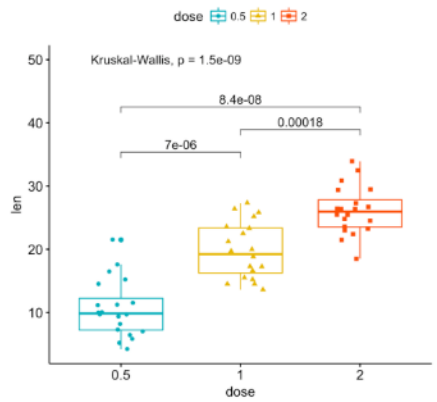
patchwork



ggdendro

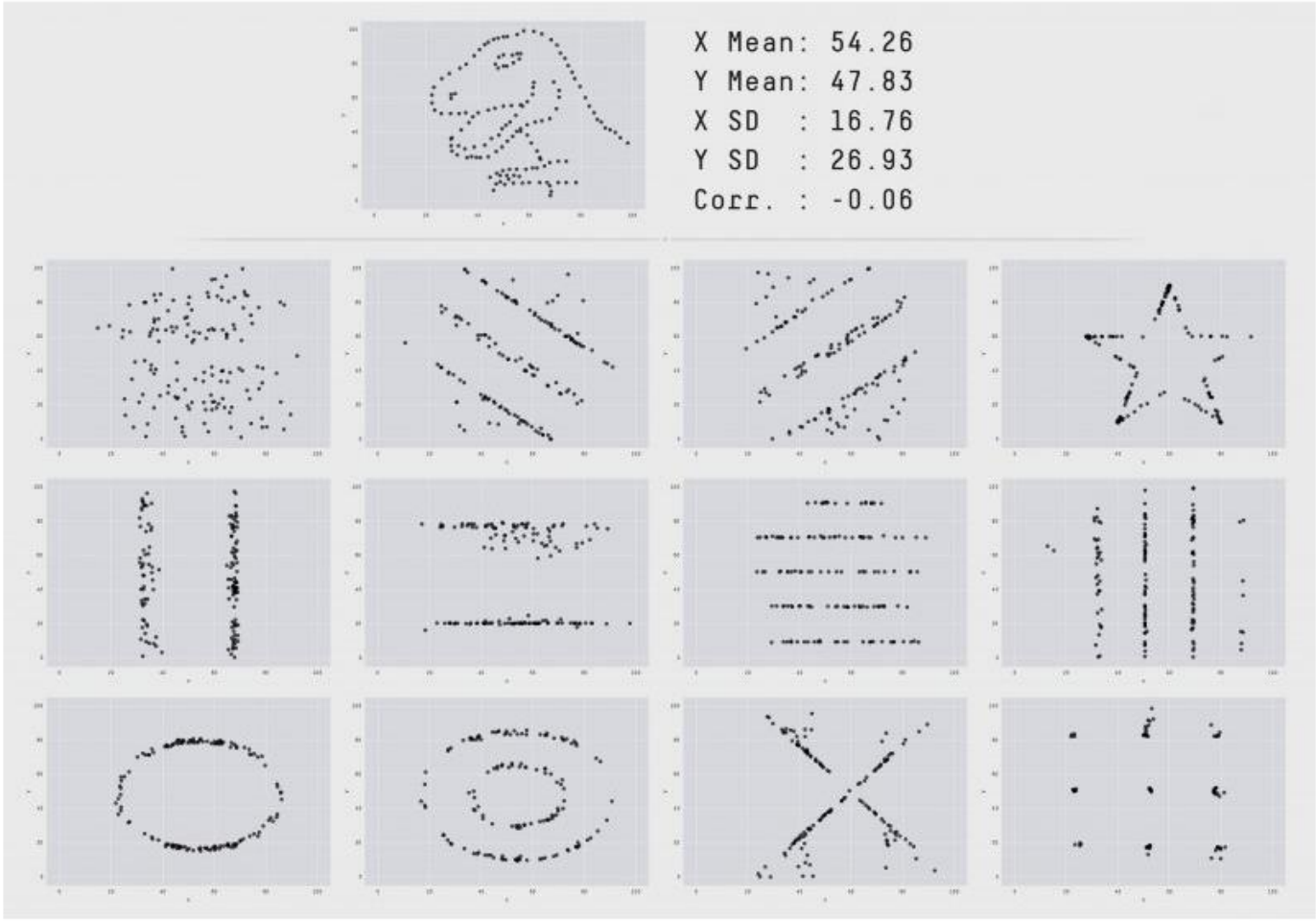


ggpubr



Always visualise your data

- Once you have tidied your data, you should always generate some visual outputs to check;
    - distribution
    - variance
    - subgroups
    - anomalies



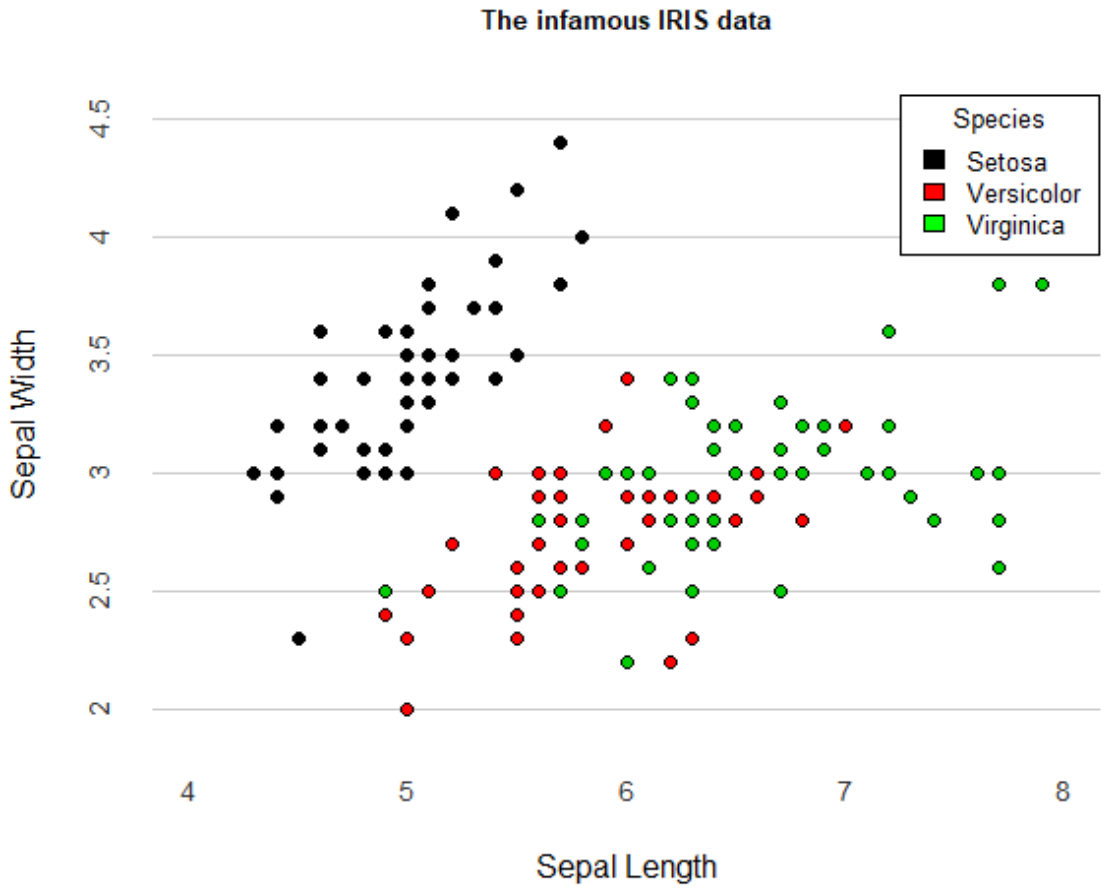
The Datasaurus Dozen. While different in appearance, each dataset has the same summary statistics (mean, standard deviation, and Pearson's correlation) to two decimal places.

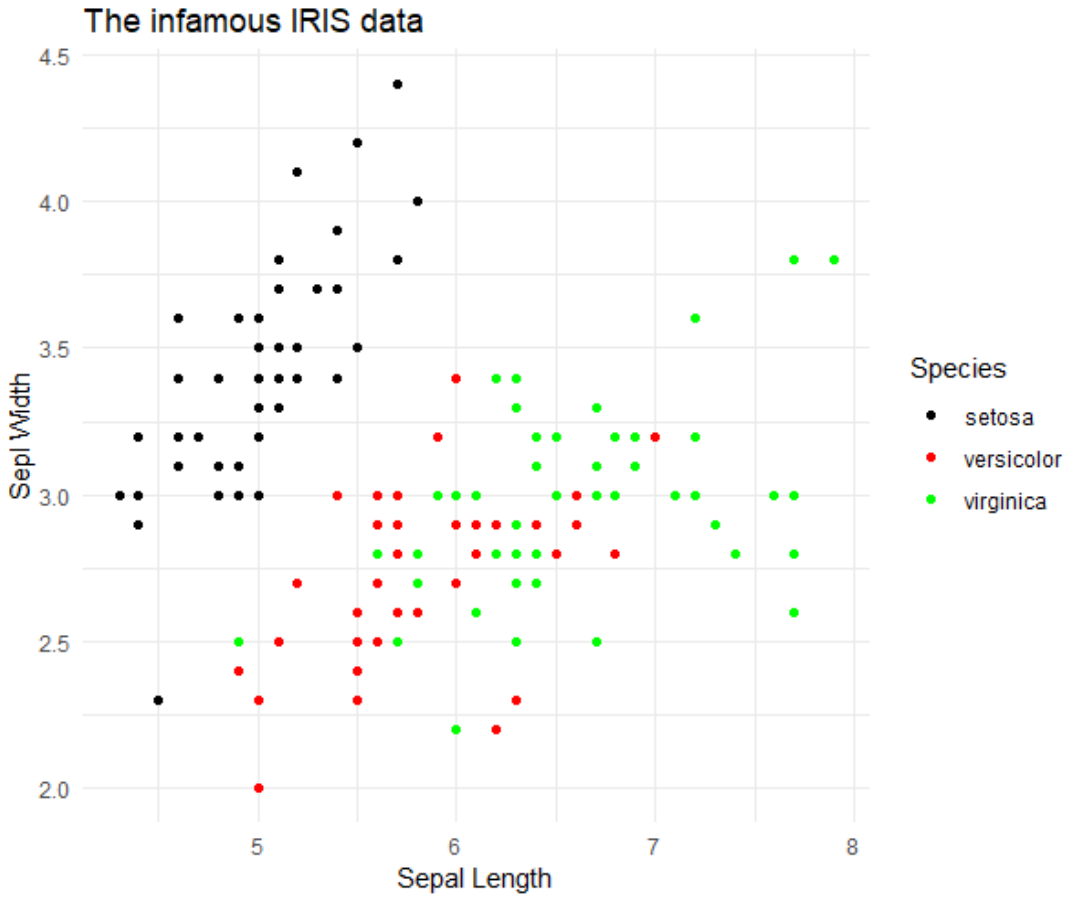
Plotting using base R graphics vs ggplot2

```r
# Here's an example using the graphics packages that comes with base R
plot(iris$Sepal.Length, iris$Sepal.Width,
     bg = iris$Species, # Fill colour
     pch = 21, # Shape: circles that can filed
     xlab = "Sepal Length", ylab = "Sepal Width", # Labels
     axes = FALSE, # Don't plot the axes
     frame.plot = FALSE, # Remove the frame
     xlim = c(4, 8), ylim = c(2, 4.5), # Limits
     panel.first = abline(h = seq(2, 4.5, 0.5), col = "grey80"))

at = pretty(iris$Sepal.Length)
mtext(side = 1, text = at, at = at,
      col = "grey20", line = 1, cex = 0.9)

at = pretty(iris$Sepal.Width)
mtext(side = 2, text = at, at = at, col = "grey20", line = 1, cex = 0.9)

legend("topright", legend = c("Setosa", "Versicolor", "Virginica"),
       title = "Species", fill=c("black", "red", "green"), cex=0.8)

title("The infamous IRIS data",
      cex.main = 0.8, font.main = 2, col.main = "black")

```

```r
# Now lets view the ggplot2 version
# library(tidyverse)
ggplot(iris, aes(x = Sepal.Length, y = Sepal.Width, colour = Species)) +
  geom_point() +
  scale_colour_manual(values = c("black", "red", "green")) +
  theme_minimal() +
  labs(title = "The infamous IRIS data",
       x = "Sepal Length",
       y = "Sepl Width")

```

Plotting using base R graphics vs ggplot2





- Open the script 08_graphics_example.R to see for yourself

Whistle-stop tour of ggplot2

Main features:
1. The data
2. The geoms
3. The mappings (x, y, colour, shape etc.)
4. Legends
5. Labels
6. Themes

and many many more

- Open the script 09_ggplot2.R

- Open the script 10_practise_plots.R

Huge support and resources out there

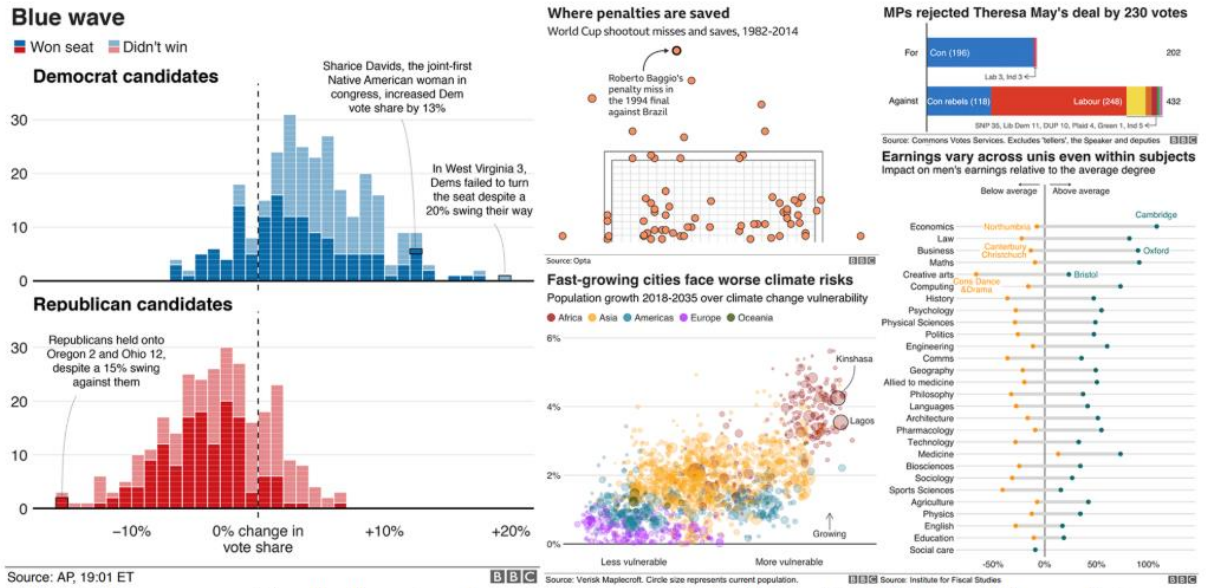
BBC Visual and Data Journalism cookbook for R graphics

Last updated: 2019-01-24

How to create BBC style graphics

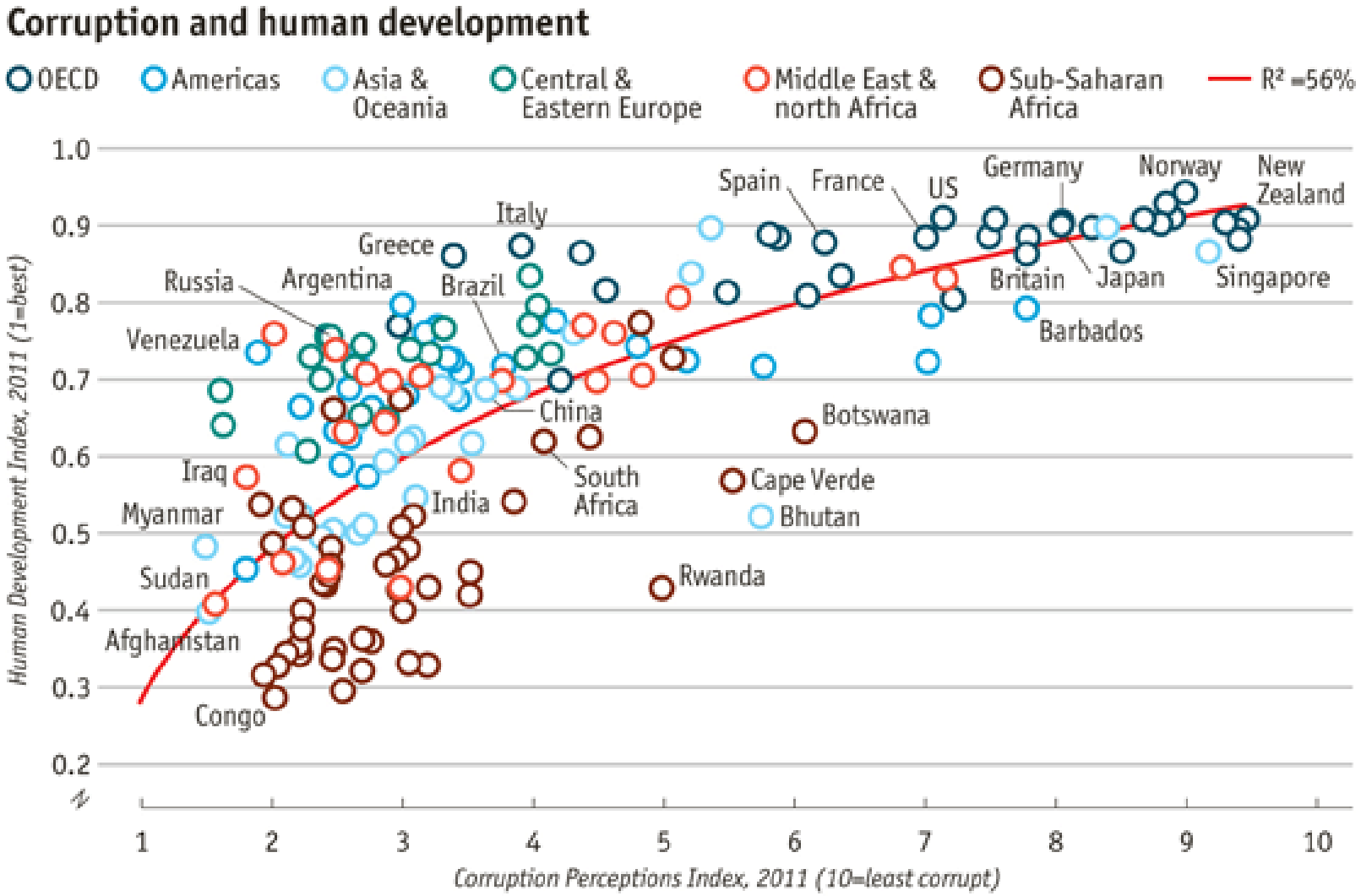
At the BBC data team, we have developed an R package and an R cookbook to make the process of creating publication-ready graphics in our in-house style using R's ggplot2 library a more reproducible process, as well as making it easier for people new to R to create graphics.

The cookbook below should hopefully help anyone who wants to make graphics like these:



We'll get to how you can put together the various elements of these graphics, but **let's get the admin out of the way first...**

Reproduce this graphic from the Economist



- Open the script 11_plotting_at_the_next_level.R