

# R: A Hitchhikers Guide to Reproducible Research

- Welcome to the ggungle

---

Brendan Palmer,  
Clinical Research Facility - Cork &  
School of Public Health  
 @B\_A\_Palmer

# ggplot2



- Data visualisation based on "[The Grammar of Graphics](#)"

`ggplot(data = <DATA>) +`

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

`linear model +`

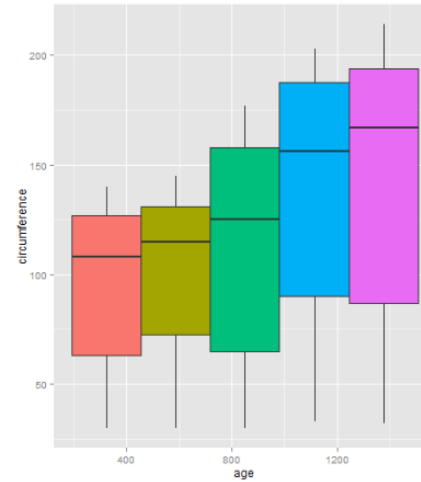
`axes formatting +`

`legend formatting +`

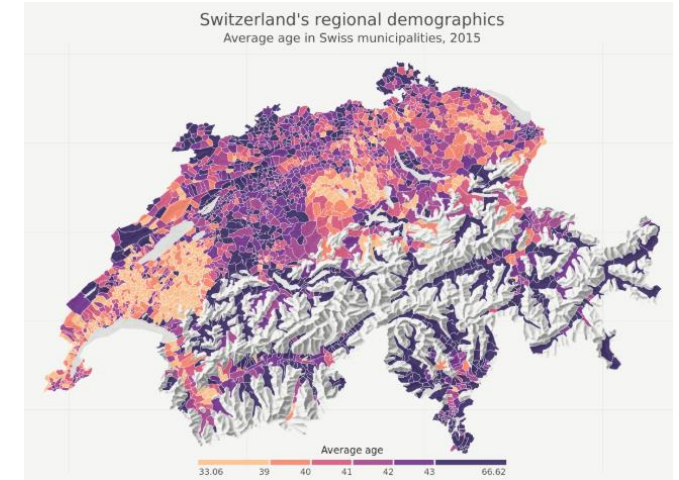
`title + etc. etc.`

# ggplot2

- Very versatile
- Allows you to go from

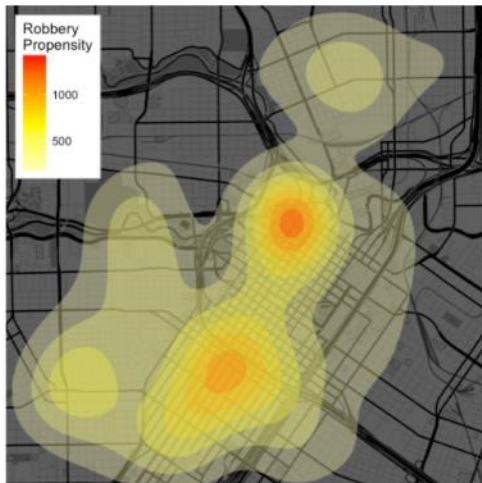


to

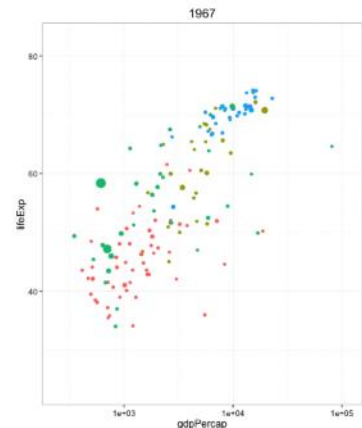


- Lots of add-on packages

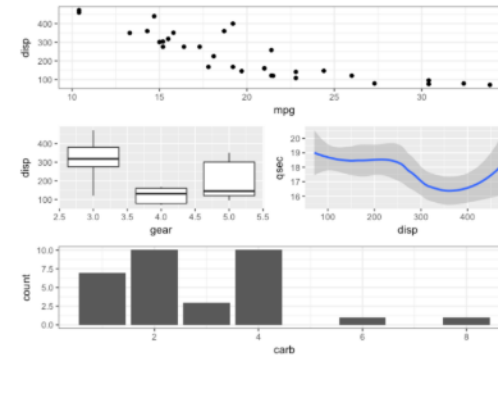
ggmap



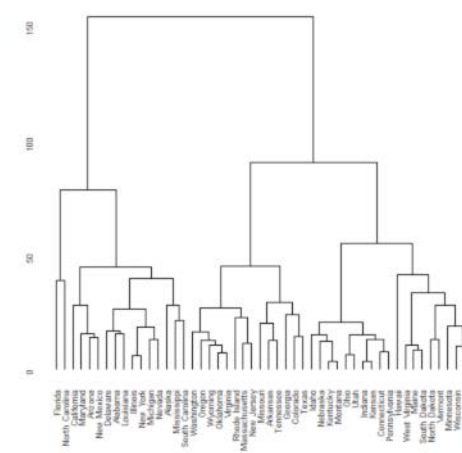
gganimate



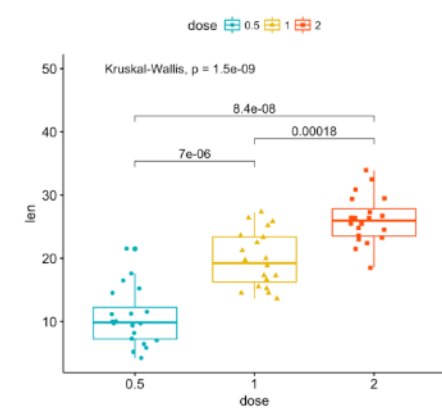
patchwork



ggdendro



ggpubr



# Lots of nice example code to be found

## How to create BBC style graphics

- Make a line chart
- Make a multiple line chart
- Make a bar chart
- Make a stacked bar chart
- Make a grouped bar chart
- Make a dumbbell chart
- Make a histogram
- Make changes to the legend
- Make changes to the axes
- Add annotations
- Work with small multiples
- Do something else entirely

## BBC Visual and Data Journalism cookbook for R graphics

*Last updated: 2019-01-24*

### How to create BBC style graphics

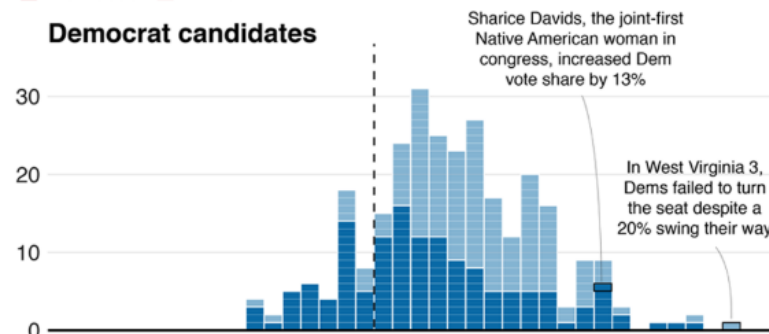
At the BBC data team, we have developed an R package and an R cookbook to make the process of creating publication-ready graphics in our in-house style using R's ggplot2 library a more reproducible process, as well as making it easier for people new to R to create graphics.

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

#### Blue wave

■ Won seat ■ Didn't win

##### Democrat candidates

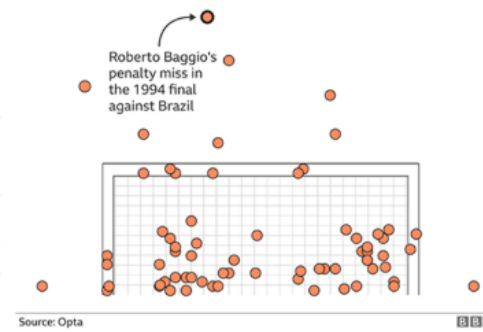


##### Republican candidates

Republicans held onto Oregon 2 and Ohio 12, despite a 15% swing against them

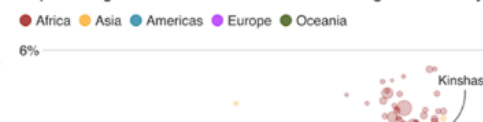
#### Where penalties are saved

World Cup shootout misses and saves, 1982-2014

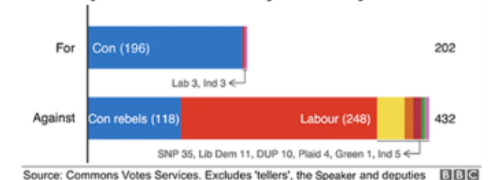


#### Fast-growing cities face worse climate risks

Population growth 2018-2035 over climate change vulnerability

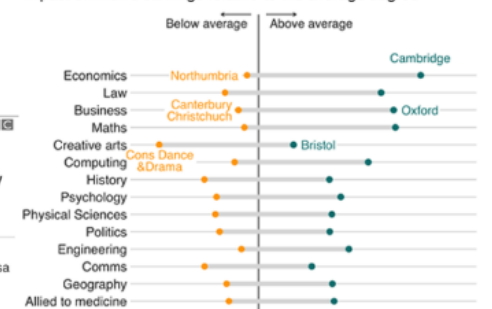


#### MPs rejected Theresa May's deal by 230 votes



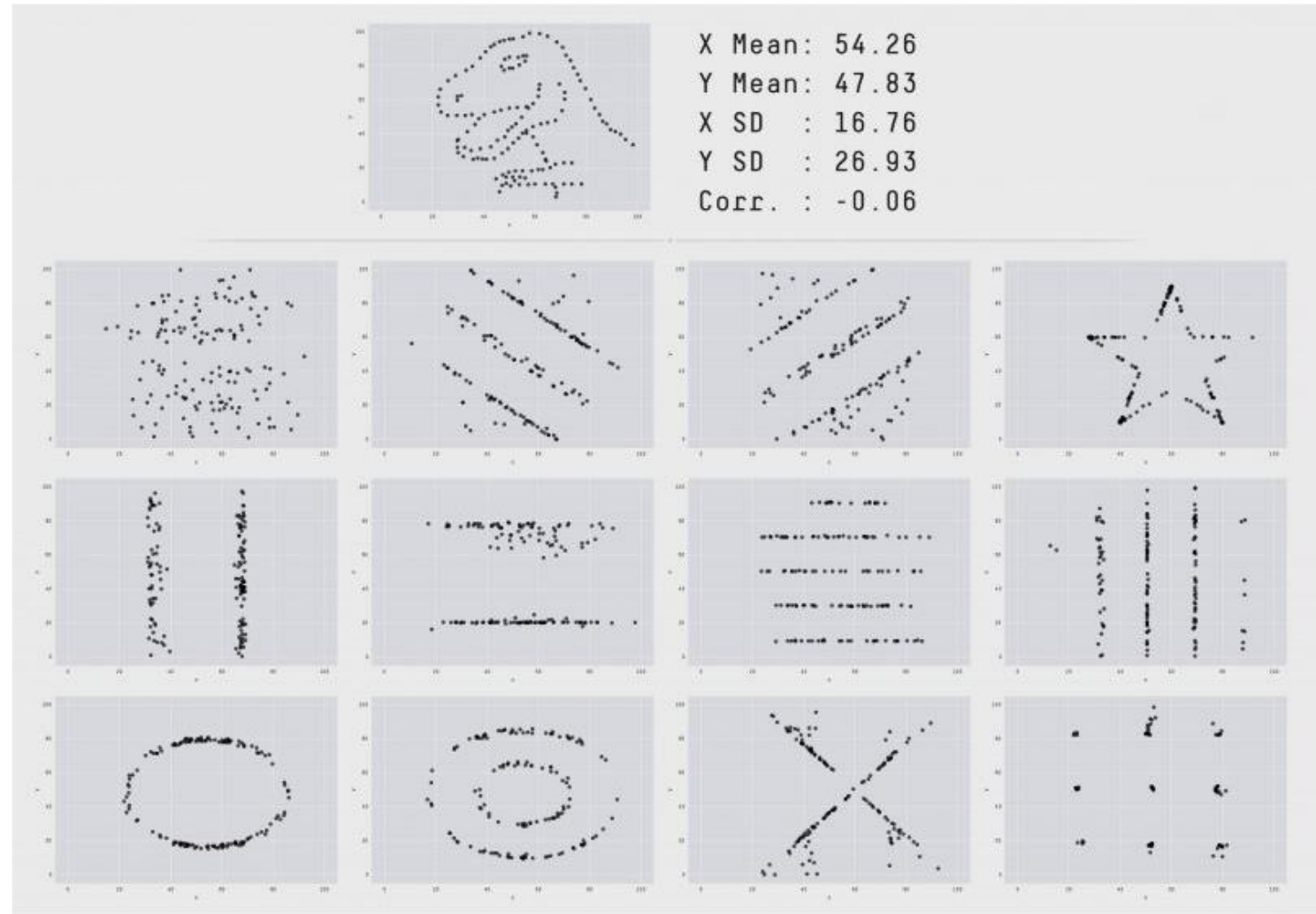
#### Earnings vary across unis even within subjects

Impact on men's earnings relative to the average degree



# Always visualise your data

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



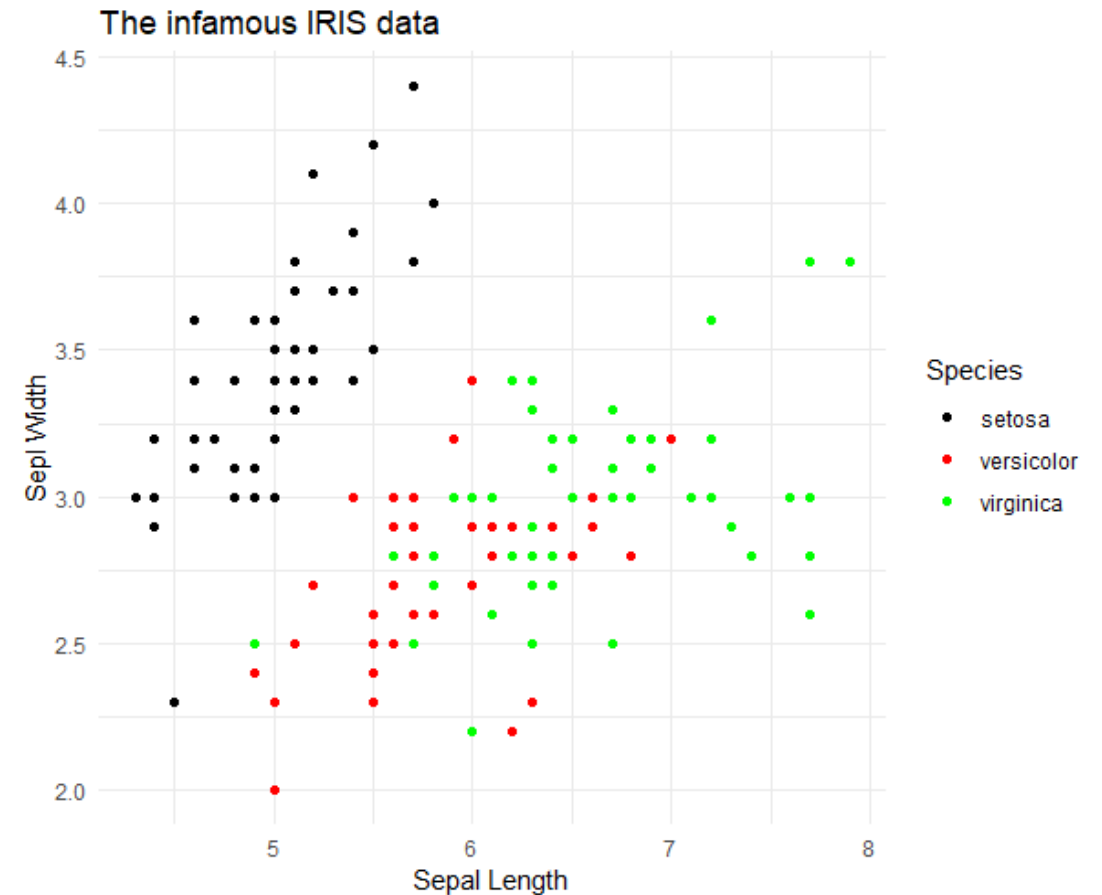
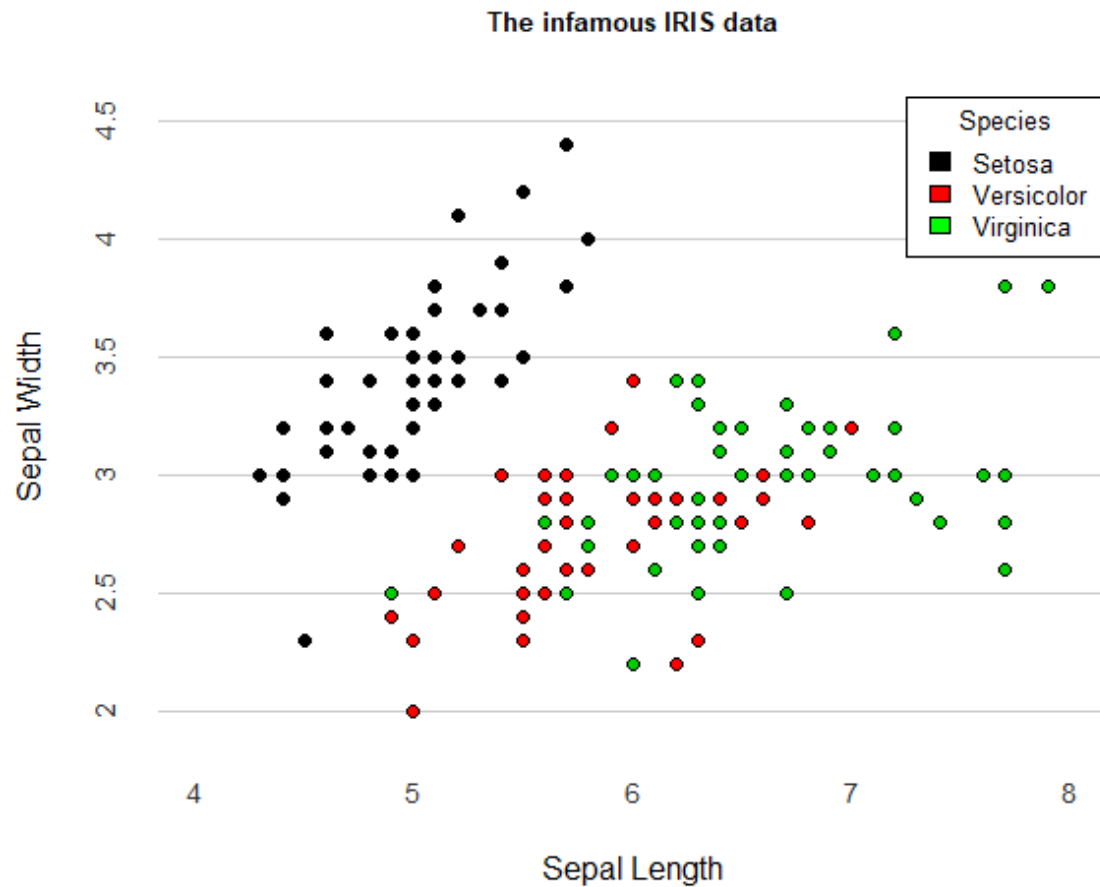
The Datasaurus Dozen. While different in appearance, each dataset has the same summary statistics (mean, standard deviation, and Pearson's correlation) to two decimal places.

# Plotting using base R graphics vs ggplot2

```
7 # Here's an example using the graphics packages that comes with base R
8 plot(iris$Sepal.Length, iris$Sepal.Width,
9      bg = iris$Species, # Fill colour
10     pch = 21, # Shape: circles that can filled
11     xlab = "Sepal Length", ylab = "Sepal Width", # Labels
12     axes = FALSE, # Don't plot the axes
13     frame.plot = FALSE, # Remove the frame
14     xlim = c(4, 8), ylim = c(2, 4.5), # Limits
15     panel.first = abline(h = seq(2, 4.5, 0.5), col = "grey80"))
16
17 at = pretty(iris$Sepal.Length)
18 mtext(side = 1, text = at, at = at,
19       col = "grey20", line = 1, cex = 0.9)
20
21 at = pretty(iris$Sepal.Width)
22 mtext(side = 2, text = at, at = at, col = "grey20", line = 1, cex = 0.9)
23
24 legend("topright", legend = c("Setosa", "Versicolor", "Virginica"),
25       title = "Species", fill=c("black", "red", "green"), cex=0.8)
26
27 title("The infamous IRIS data",
28       cex.main = 0.8, font.main = 2, col.main = "black")
29
```

```
30 # Now lets view the ggplot2 version
31 # library(tidyverse)
32 ggplot(iris, aes(x = Sepal.Length, y = Sepal.Width, colour = Species)) +
33   geom_point() +
34   scale_colour_manual(values = c("black", "red", "green")) +
35   theme_minimal() +
36   labs(title = "The infamous IRIS data",
37        x = "Sepal Length",
38        y = "Sepal Width")
39 |
```

# 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

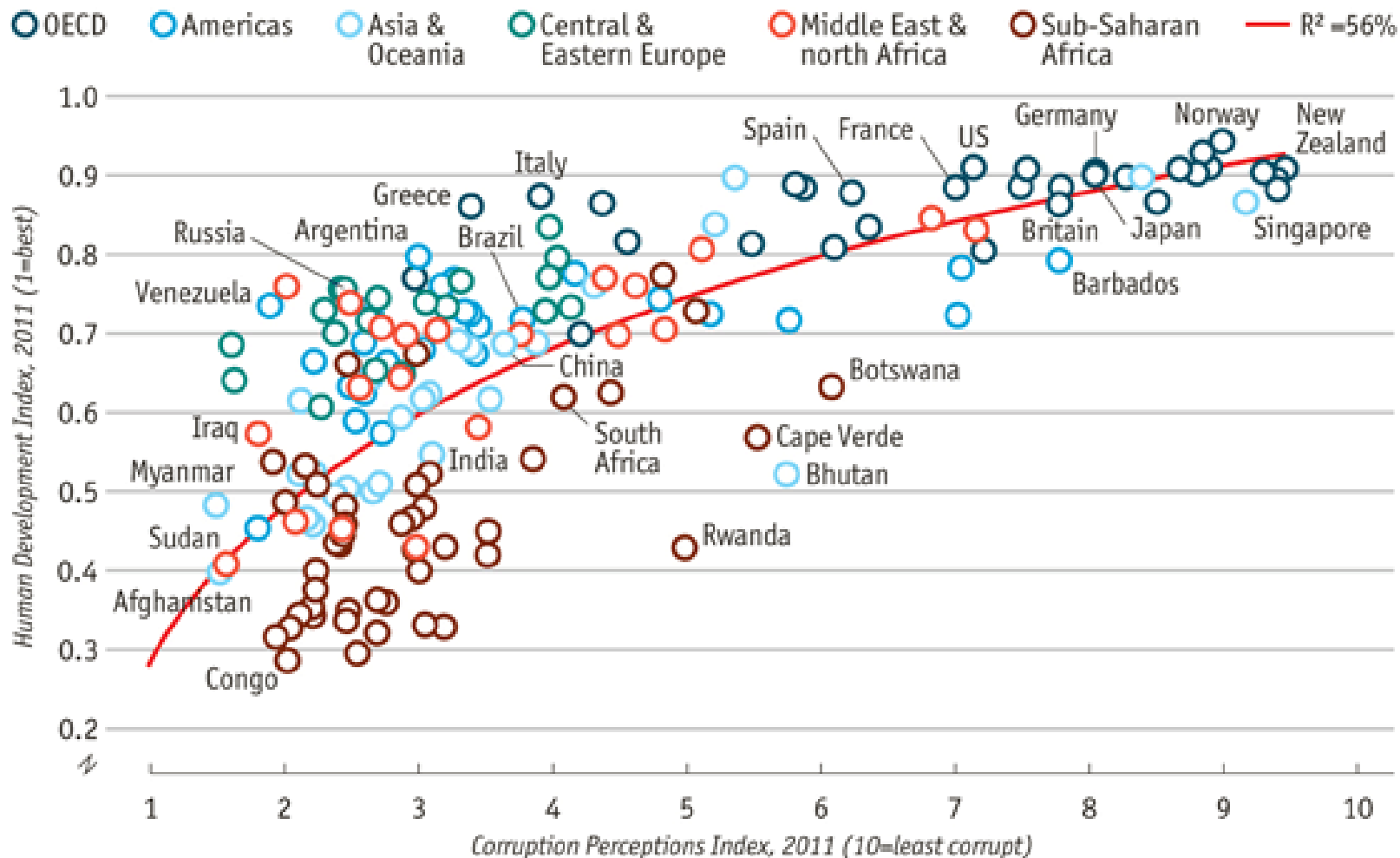
and many many more

- Open the script 09\_ggplot2.R
- Open the script 10\_practice\_plots.R



# Reproduce this graphic from the Economist

Corruption and human development



- Open the script `11_plotting_at_the_next_level.R`