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

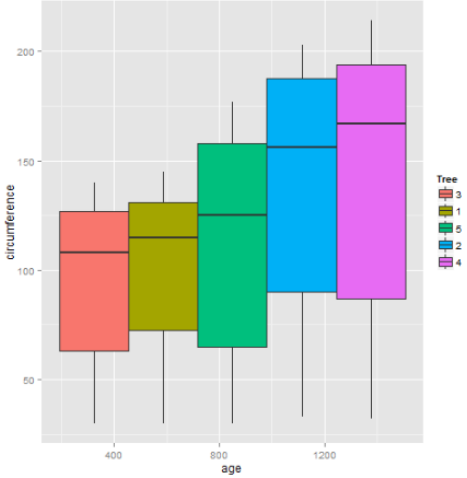
ggplot2

- Data visualisation based on "The Grammar of Graphics"

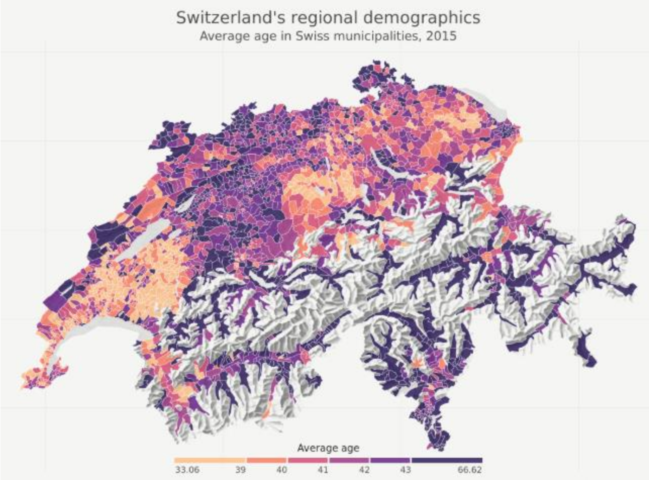
```
ggplot(data = <DATA>) +

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

      linear model  +

      axes formatting  +

      legend formatting  +

      title  + etc. etc.
```

ggplot2

- Very versatile
- Allows you to go from

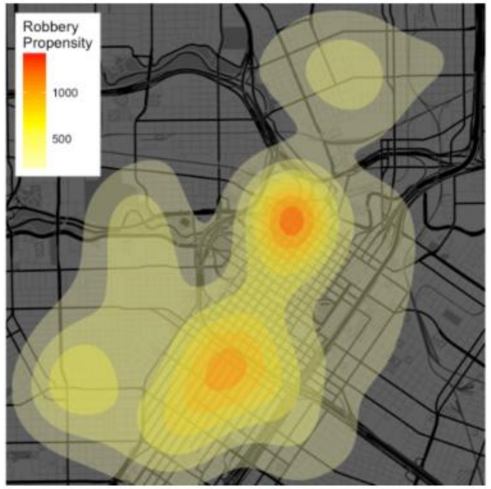


to

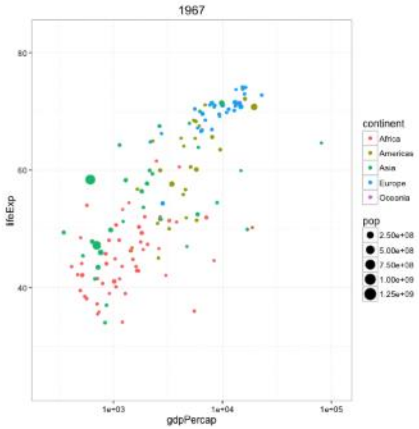

Switzerland's regional demographics
Average age in Swiss municipalities, 2015
Average age
33.06
39
40
41
42
43
66.62

- Lots of add-on packages

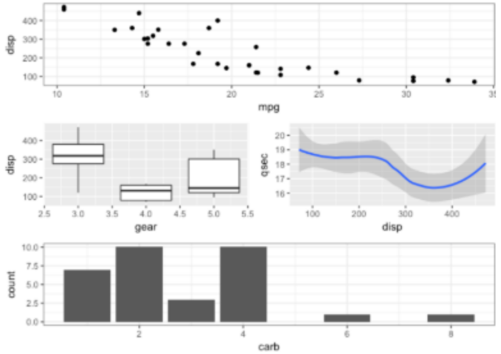
ggmap



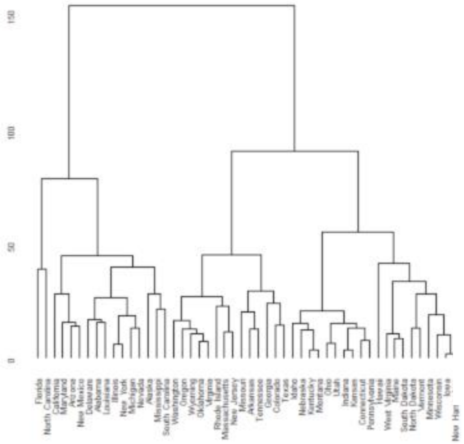
gganimate



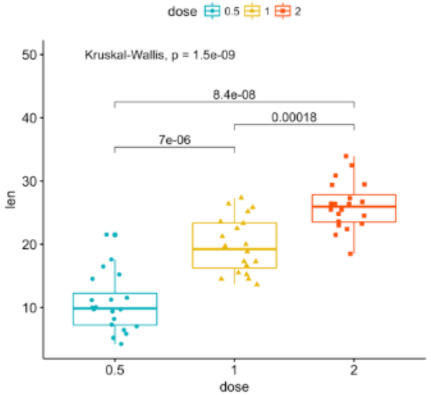
patchwork



ggdendro

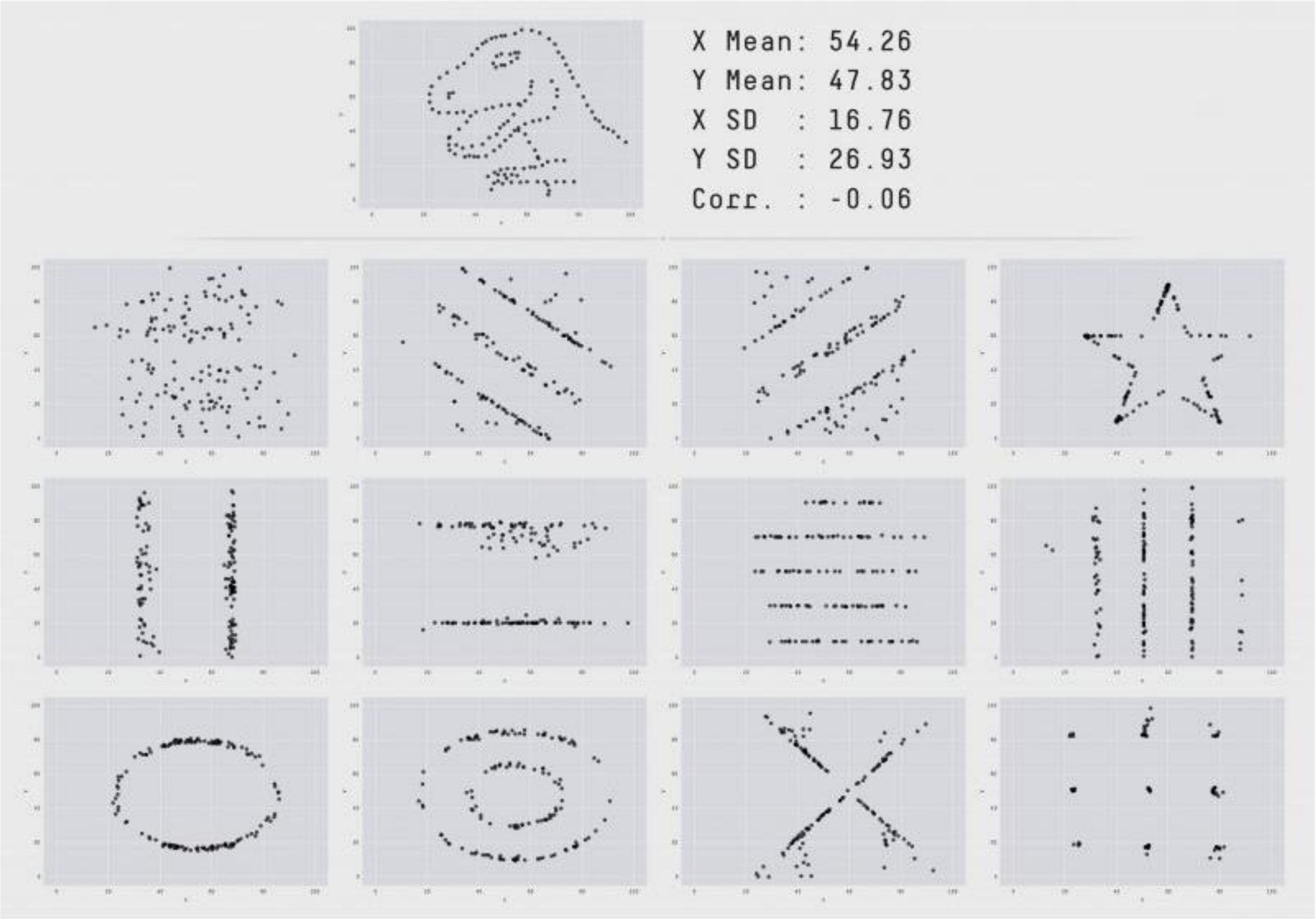


ggpubr



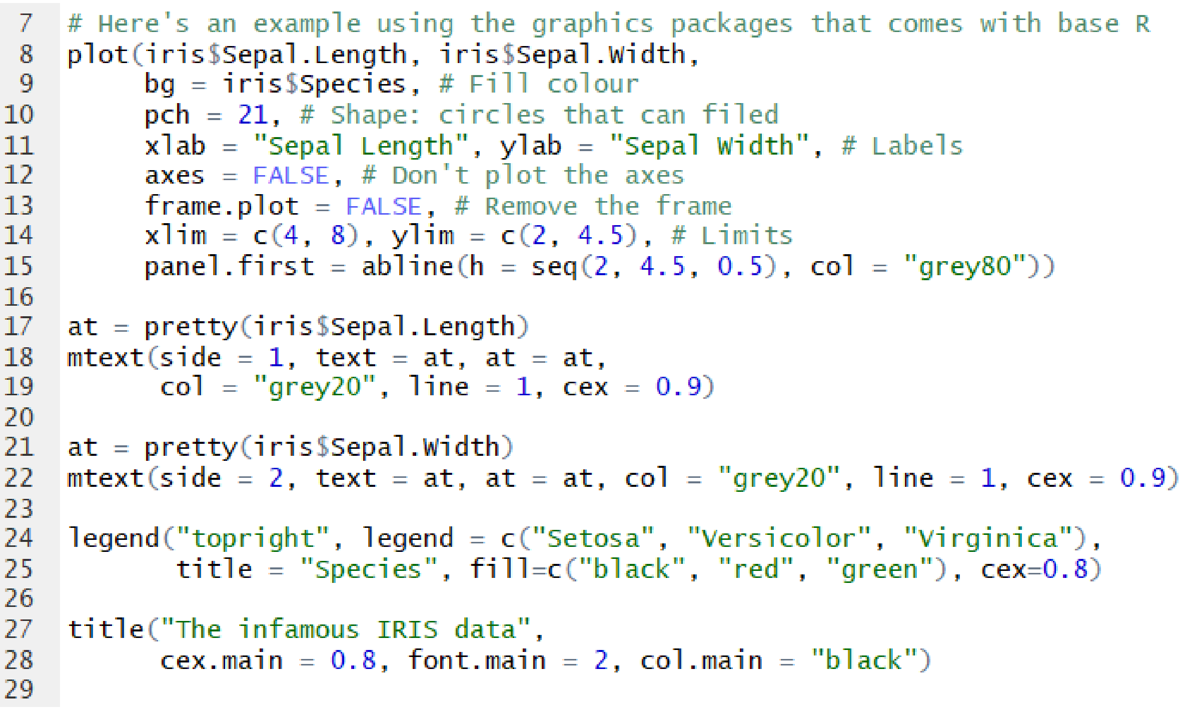
Always visualise your data

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

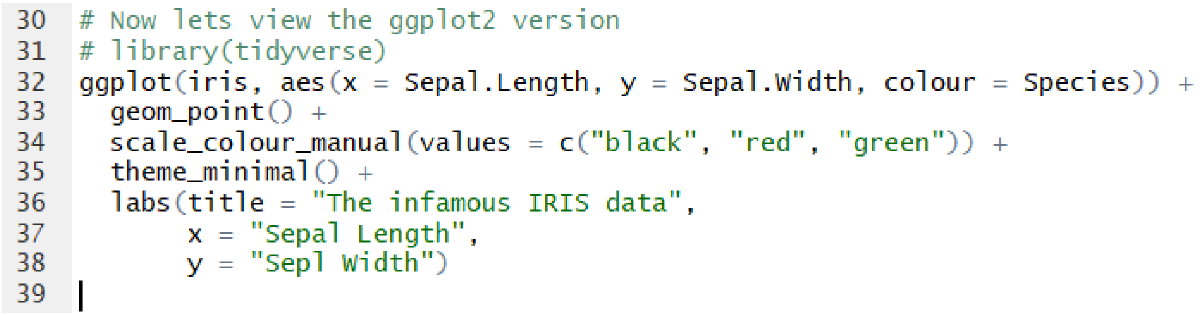


The Datasaurus Dozen. While different in appearance, each dataset has the same summary statistics (mean, standard deviation, and Pearson's correlation) to two decimal places.

Plotting using base R graphics vs ggplot2

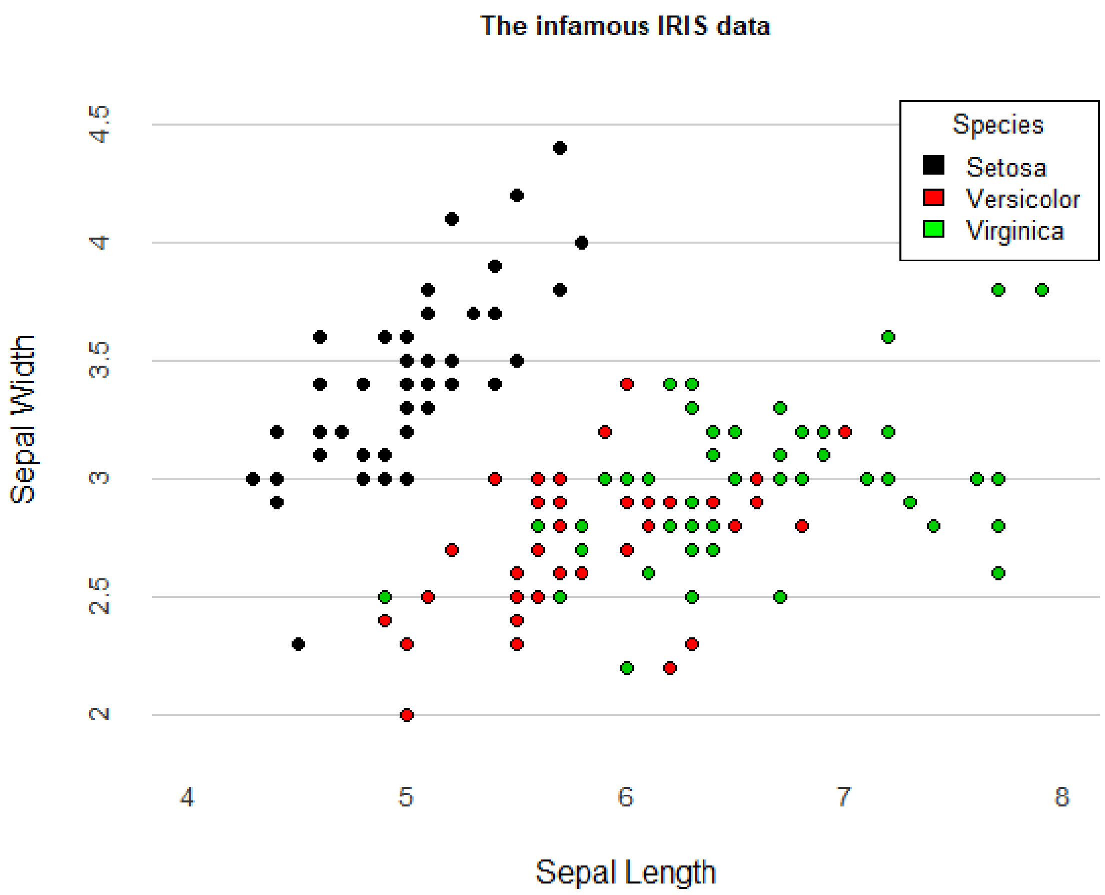


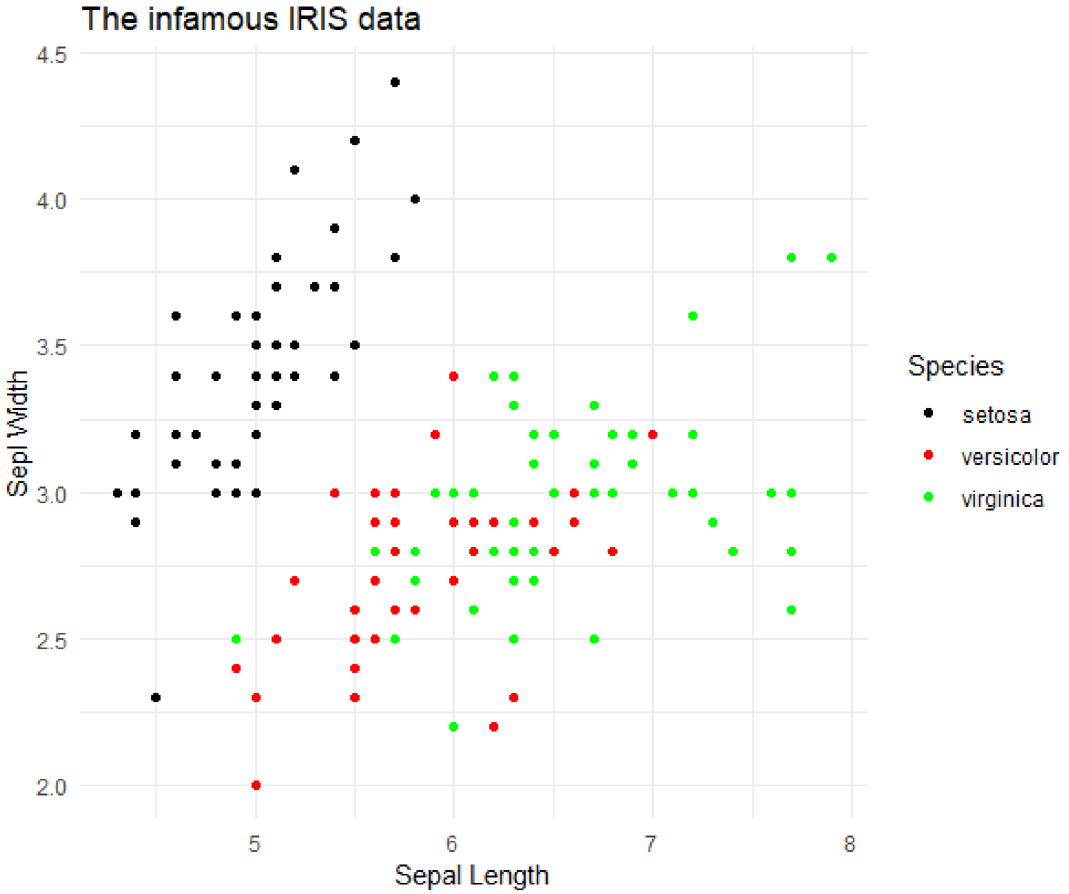
```r
# Here's an example using the graphics packages that comes with base R
plot(iris$Sepal.Length, iris$Sepal.Width,
     bg = iris$Species, # Fill colour
     pch = 21, # Shape: circles that can filed
     xlab = "Sepal Length", ylab = "Sepal Width", # Labels
     axes = FALSE, # Don't plot the axes
     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:

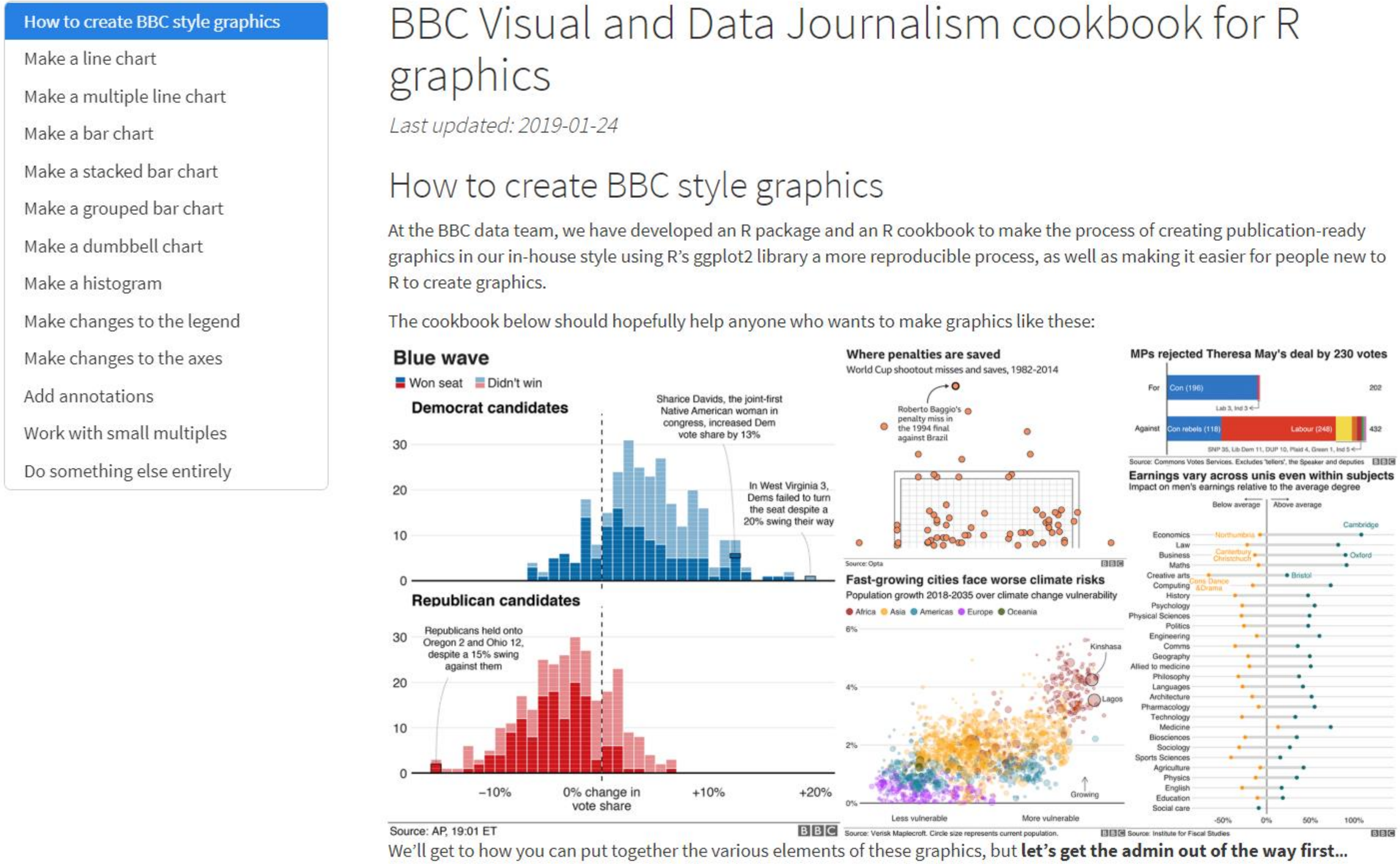
1. The data
2. The geoms
3. The mappings (x, y, colour, shape etc.)
4. Legends
5. Labels
6. Themes

and many many more

- Open the script 09_pm_ggplot2.R

- Open the script 10_pm_practise_plots.R

Huge support and resources out there



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.

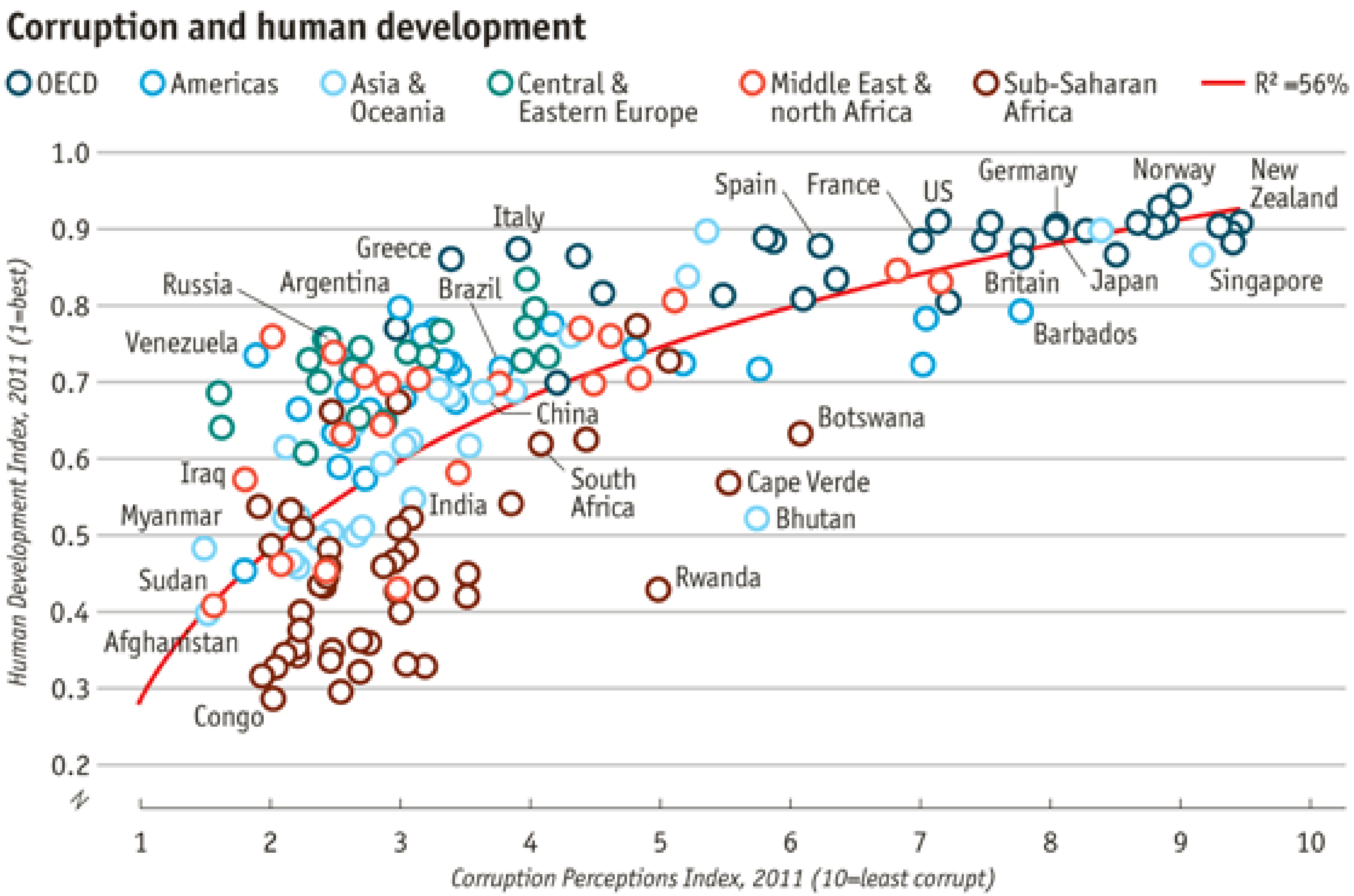
The cookbook below should hopefully help anyone who wants to make graphics like these:

Blue wave
Won seat
Didn't win
Democrat candidates
Sharice Davids, the joint-first Native American woman in congress, increased Dem vote share by 13%
In West Virginia 3, Dems failed to turn the seat despite a 20% swing their way
30
20
10
0
Republican candidates
Republicans held onto Oregon 2 and Ohio 12, despite a 15% swing against them
−10%
0% change in vote share
+10%
+20%
Source: AP, 19:01 ET
BBC
Where penalties are saved
World Cup shootout misses and saves, 1982-2014
Roberto Baggio's penalty miss in the 1994 final against Brazil
Source: Opta
Fast-growing cities face worse climate risks
Population growth 2018-2035 over climate change vulnerability
Africa
Asia
Americas
Europe
Oceania
6%
4%
2%
0%
Kinshasa
Lagos
Growing
Less vulnerable
More vulnerable
Source: Verisk Maplecroft. Circle size represents current population.
MPs rejected Theresa May's deal by 230 votes
For
Con (196)
202
Lab 3, Ind 3
Against
Con rebels (118)
Labour (248)
432
SNP 35, Lib Dem 11, DUP 10, Plaid 4, Green 1, Ind 5
Source: Commons Votes Services. Excludes 'tellers', the Speaker and deputies
Earnings vary across unis even within subjects
Impact on men's earnings relative to the average degree
Below average
Above average
Cambridge
Northumbria
Oxford
Canterbury Christchurch
Bristol
Cons Dance &Drama
Economics
Law
Business
Maths
Creative arts
Computing
History
Psychology
Physical Sciences
Politics
Engineering
Comms
Geography
Allied to medicine
Philosophy
Languages
Architecture
Pharmacology
Technology
Medicine
Biosciences
Sociology
Sports Sciences
Agriculture
Physics
English
Education
Social care
-50%
0%
50%
100%
Source: Institute for Fiscal Studies
BBC

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...**

`- Open the script 11_plotting_at_the_next_level.R`

Reproduce this graphic from the Economist



```
- Open the script 11_plotting_at_the_next_level.R
```

Home worksheet

Open ws5_script5_generate_this_graph.R

http://tutorials.iq.harvard.edu/R/Rgraphics/Rgraphics.html