

Data Visualisation: Principles and practice

GM Analyst Network
17 May 2018

Henry Partridge
Data Lab Manager



Trafford Data Lab



<https://www.trafforddatalab.io/>

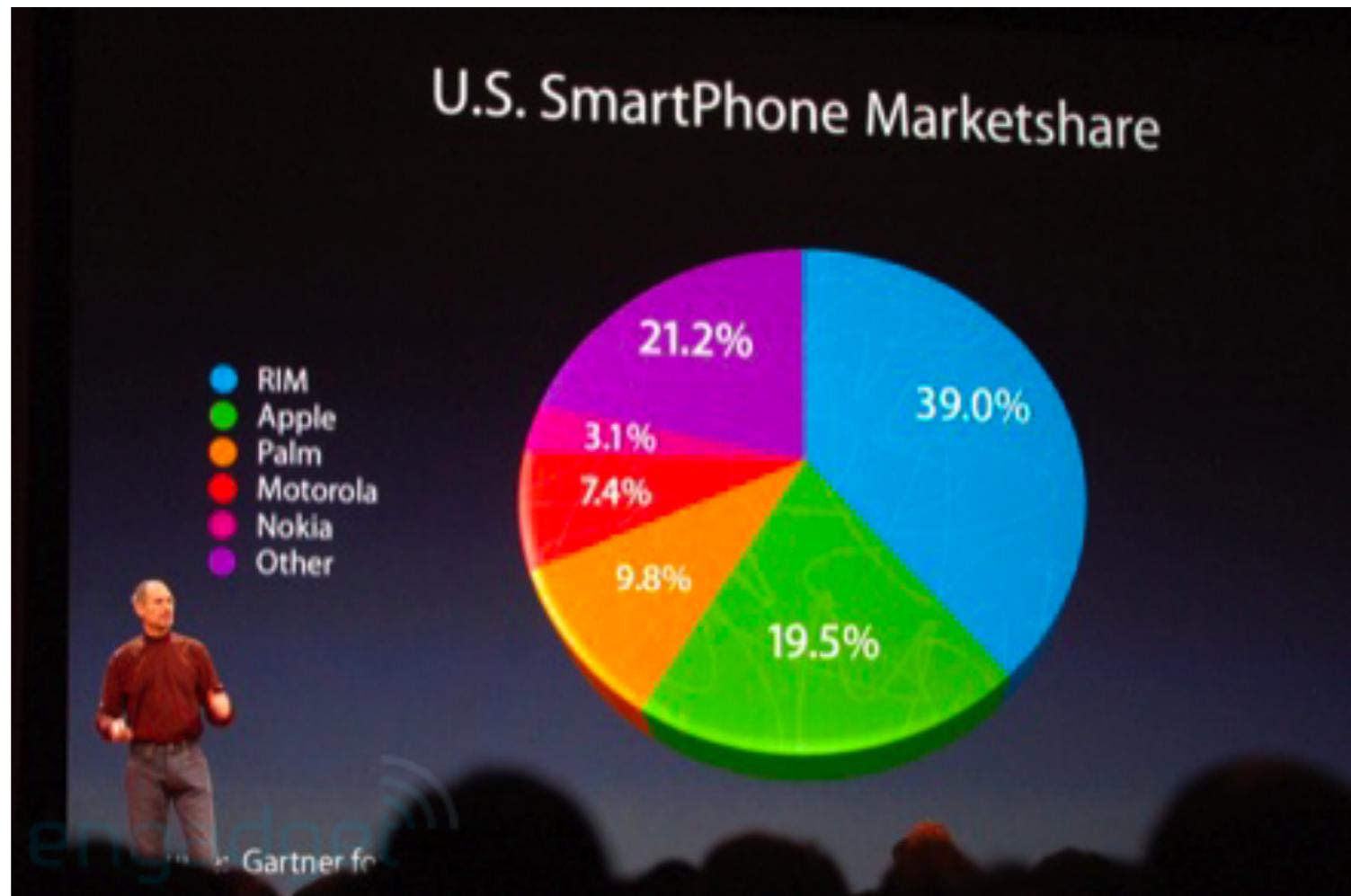
Principles

- Clarity
- Openness
- Reproducibility

Clarity

We strive for an economy of style
in all of our data visualisations.

Macworld 2008 keynote



"Chart time! It's first Q shipping it garnered 19% market share, Palm 9.8%, RIM 39%, other 20%, Nokia 7% ..."

Cleveland's perceptual hierarchy

Accuracy ↑	1 Position along a common scale	Scatter plots
	2 Positions along non-aligned, identical	Small multiples
	3 Length, direction, angle	Waterfall chart
	4 Area	Treemap
	5 Volume, curvature	3D bar charts
	6 Shading, colour saturation	Continuous color scale

Pie chart vs bar chart

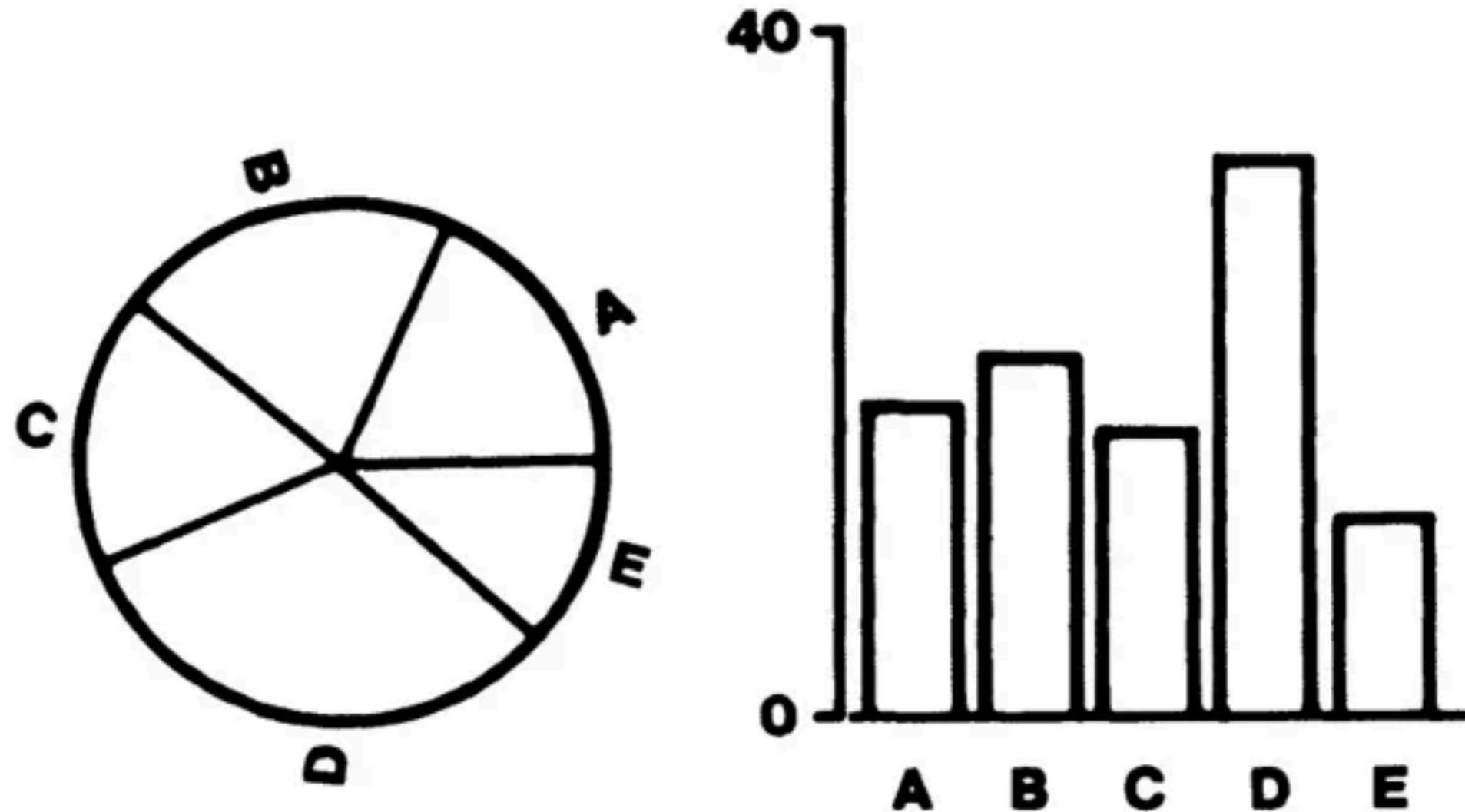
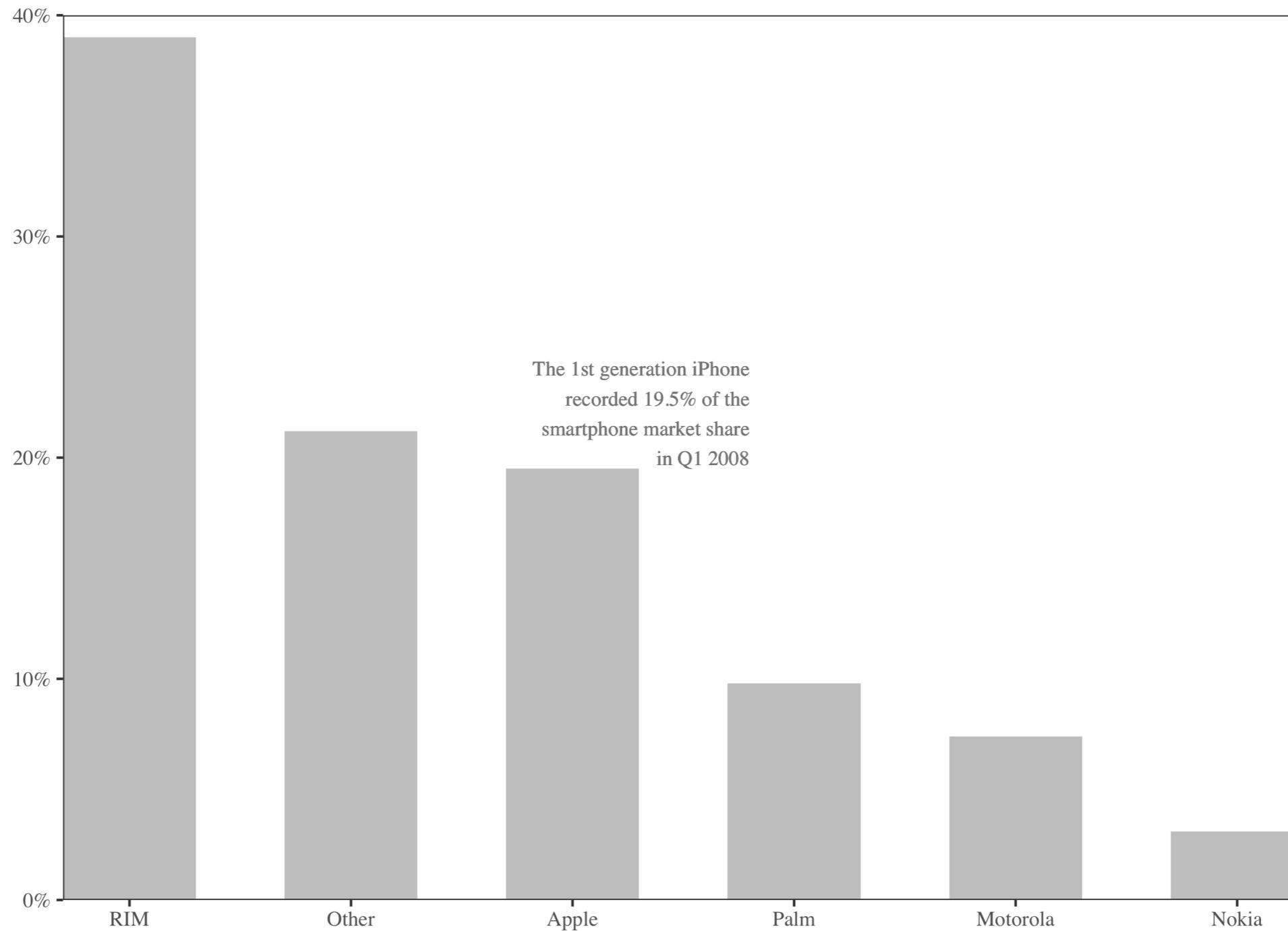


Figure 3. Graphs from position–angle experiment.

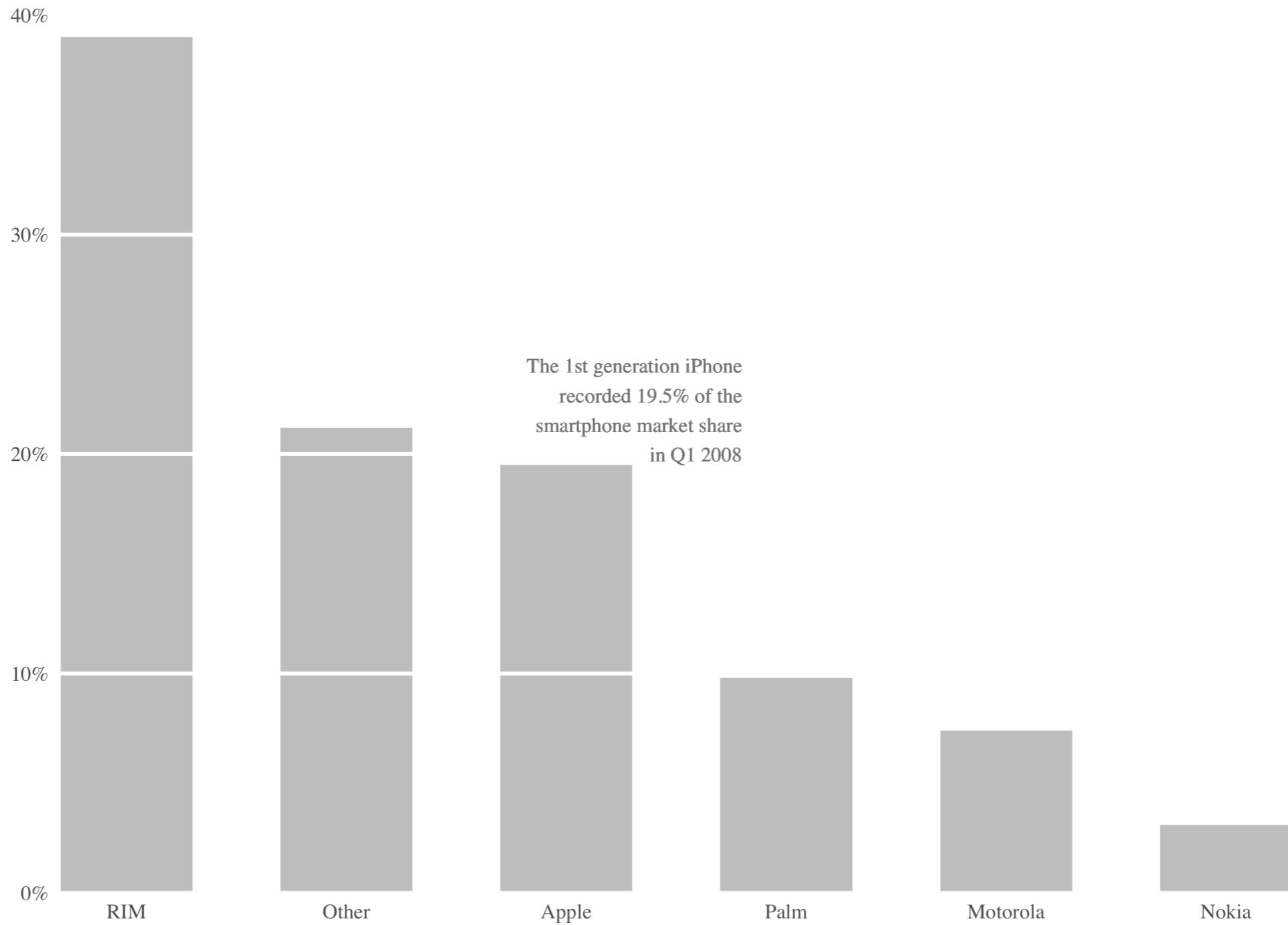
“A table is nearly always better than a dumb pie chart; the only worse design than a pie chart is several of them, for then the viewer is asked to compare quantities located in spatial disarray both within and between charts [...] Given their low density and failure to order numbers along a visual dimension, pie charts should never be used.”

— Edward Tufte (1983)

Tufte's aesthetic guidelines



Tufte's aesthetic guidelines

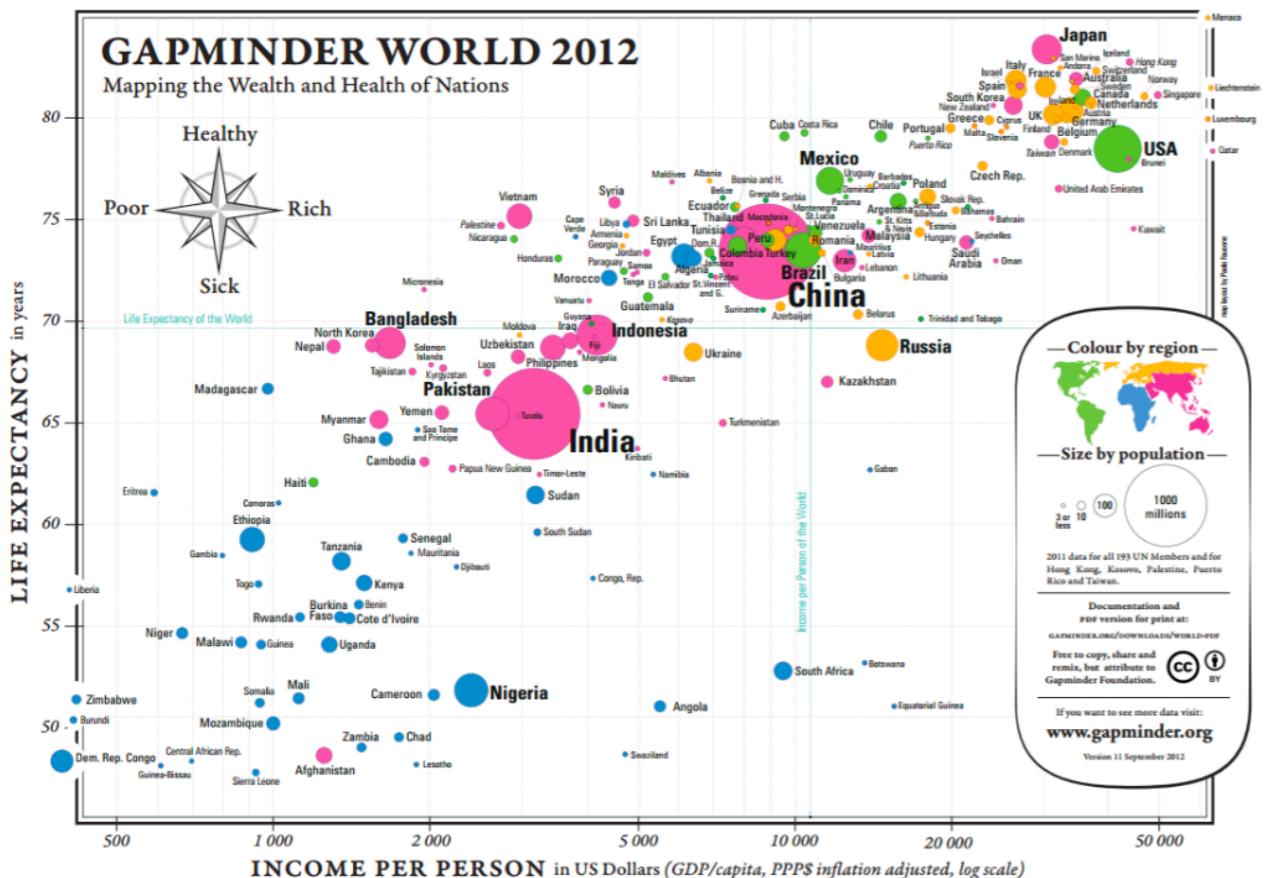


Wilkinson's Grammar of Graphics

"A statistical graphic is a mapping from data to aesthetic attributes (colour, shape, size) of geometric objects (points, lines, bars)."

— Hadley Wickham (2016)

Rosling's bubble charts



Variable

GDP per capita (\$)

Life Expectancy

Population

Continent

Geometry

point

point

point

point

Aesthetic

x-position

y-position

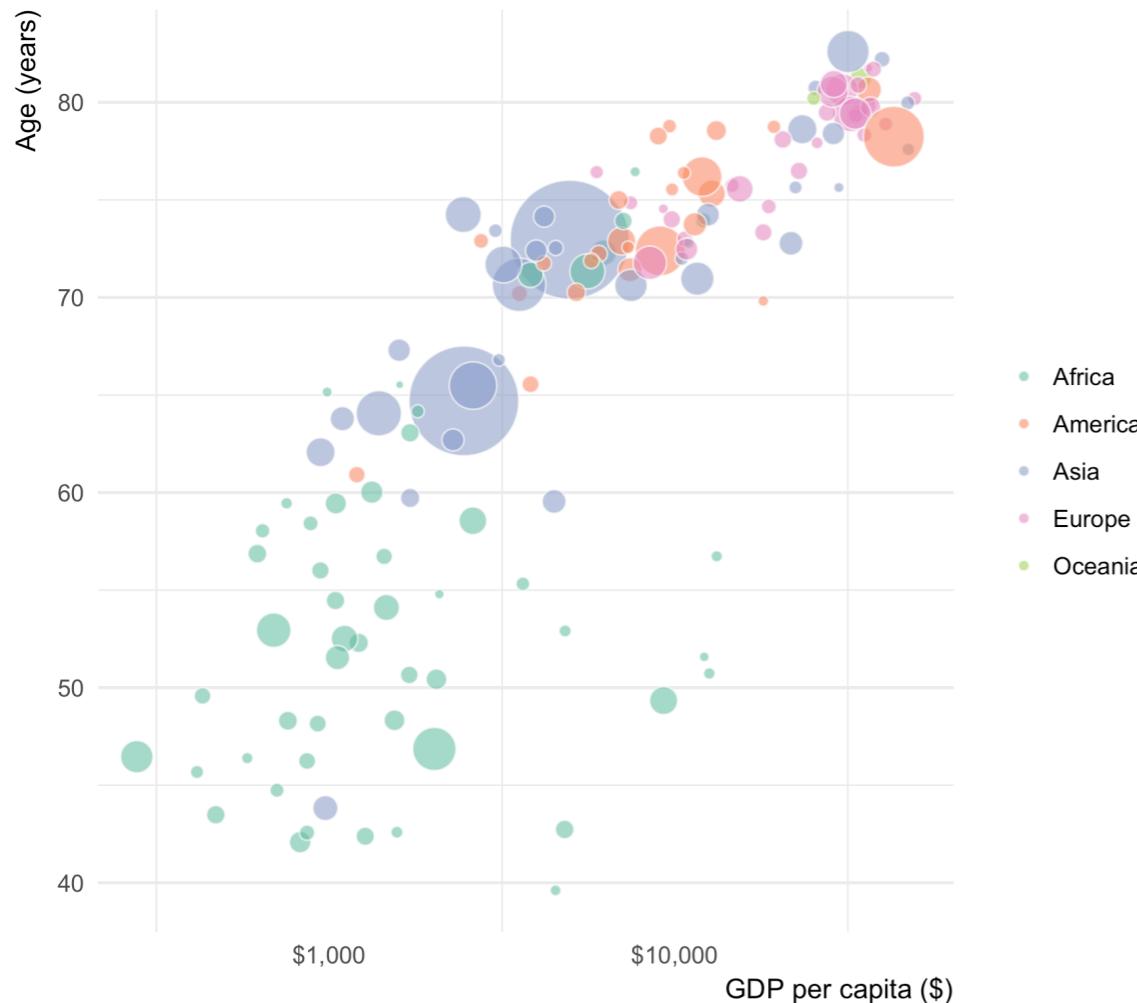
size

fill

Rosling's bubble charts in ggplot2

Life expectancy increases with income

Relationship between life expectancy and income, 2007



Source: Gapminder.org | @traffordDataLab

```
library(tidyverse) ; library(gapminder) ; library(scales)

ggplot(filter(gapminder, year == 2007), aes(x = gdpPercap, y = lifeExp)) +
  scale_x_log10(labels = scales::dollar) +
  geom_point(aes(size = pop, fill = continent), shape = 21, colour = "white", alpha = 0.6) +
  scale_fill_brewer(palette = "Set2") +
  scale_size_continuous(range = c(1, 20)) +
  labs(title = "Life expectancy increases with income",
       subtitle = "Relationship between life expectancy and income, 2007",
       caption = "Source: Gapminder.org | @traffordDataLab",
       x = "GDP per capita ($)",
       y = "Age (years)") +
  guides(size = FALSE) +
  theme_minimal() +
  theme(panel.grid.major.x = element_blank(),
        plot.caption = element_text(size = 10, hjust = 1, margin = margin(t = 15)),
        axis.title = element_text(size = 10, hjust = 1),
        legend.position = "right",
        legend.title = element_blank())

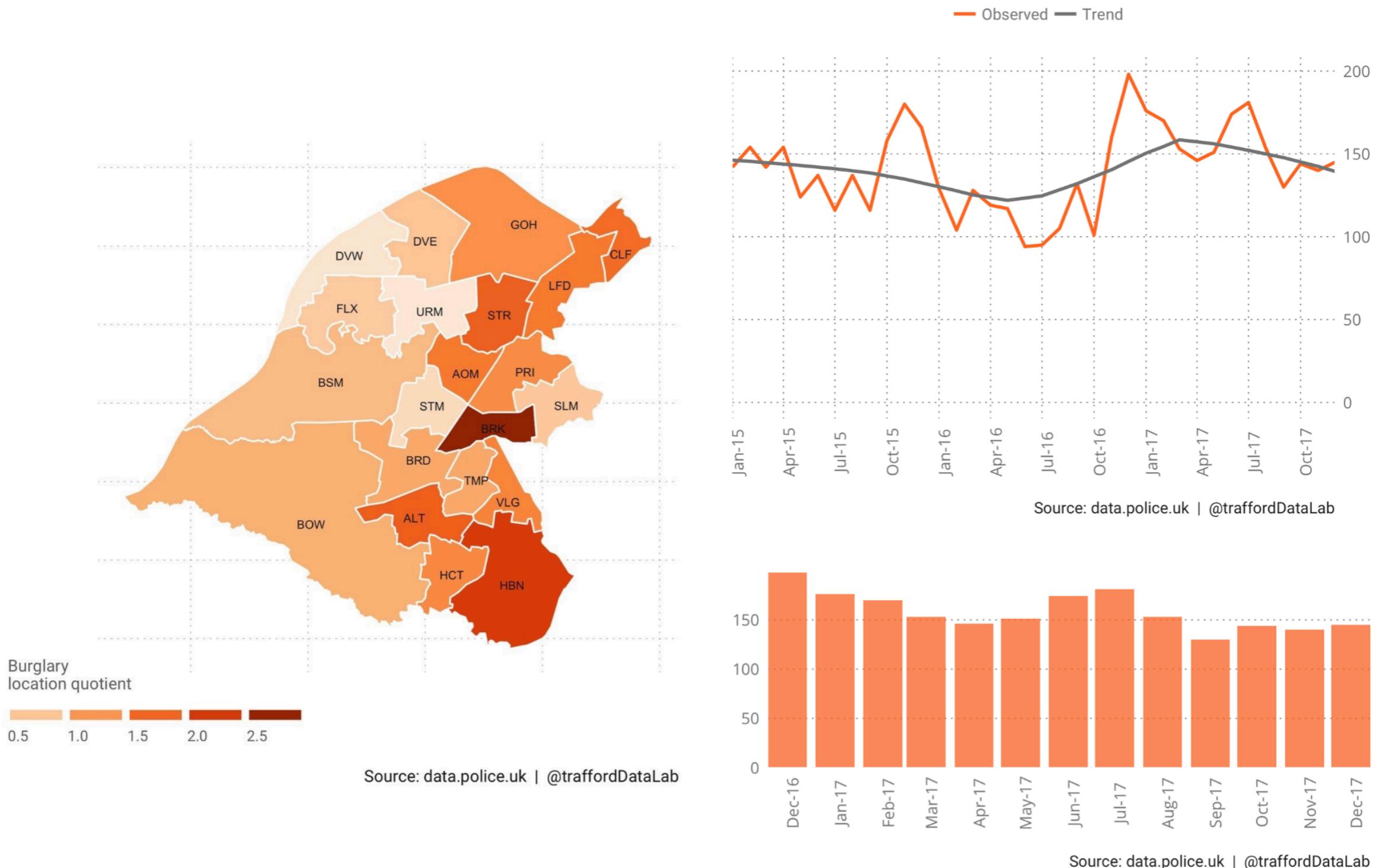
ggsave(file = "figures/rosling_bubble_chart.png", width = 6, height = 6)
```

Source: gapminder.org

Lab visualisation style

```
theme_lab <- function () {  
  theme_grey(base_size = 11.5, base_family = "Roboto") %+replace%  
  theme(  
    # add padding to the plot  
    plot.margin = unit(rep(0.5, 4), "cm"),  
    # remove the plot background and border  
    plot.background = element_blank(),  
    panel.background = element_blank(),  
    panel.border = element_blank(),  
    # make the legend and strip background transparent  
    legend.background = element_rect(fill = "transparent", colour = NA),  
    legend.key = element_rect(fill = "transparent", colour = NA),  
    strip.background = element_rect(fill = "transparent", colour = NA),  
    # add light, dotted major grid lines only  
    panel.grid.major = element_line(linetype = "dotted", colour = "#757575", size = 0.3),  
    panel.grid.minor = element_blank(),  
    # remove the axis tick marks and hide axis lines  
    axis.ticks = element_blank(),  
    axis.line = element_line(color = "#FFFFFF", size = 0.3),  
    # modify the bottom margins of the title and subtitle  
    plot.title = element_text(size = 18, colour = "#757575", hjust = 0, margin = margin(b = 4)),  
    plot.subtitle = element_text(size = 12, colour = "#757575", hjust = 0, margin = margin(b = 10)),  
    # add padding to the caption  
    plot.caption = element_text(size = 10, colour = "#212121", hjust = 1, margin = margin(t = 15)),  
    # change to Open Sans for axes titles, tick labels, legend title and legend key, and strip text  
    axis.title = element_text(family = "Open Sans", size = 11, colour = "#757575", face = "plain", hjust = 1),  
    axis.text = element_text(family = "Open Sans", size = 10, colour = "#757575", face = "plain"),  
    legend.title = element_text(size = 12, colour = "#757575"),  
    legend.text = element_text(size = 10, colour = "#757575"),  
    strip.text = element_text(family = "Open Sans", size = 12, colour = "#757575", face = "plain")  
  )  
}
```

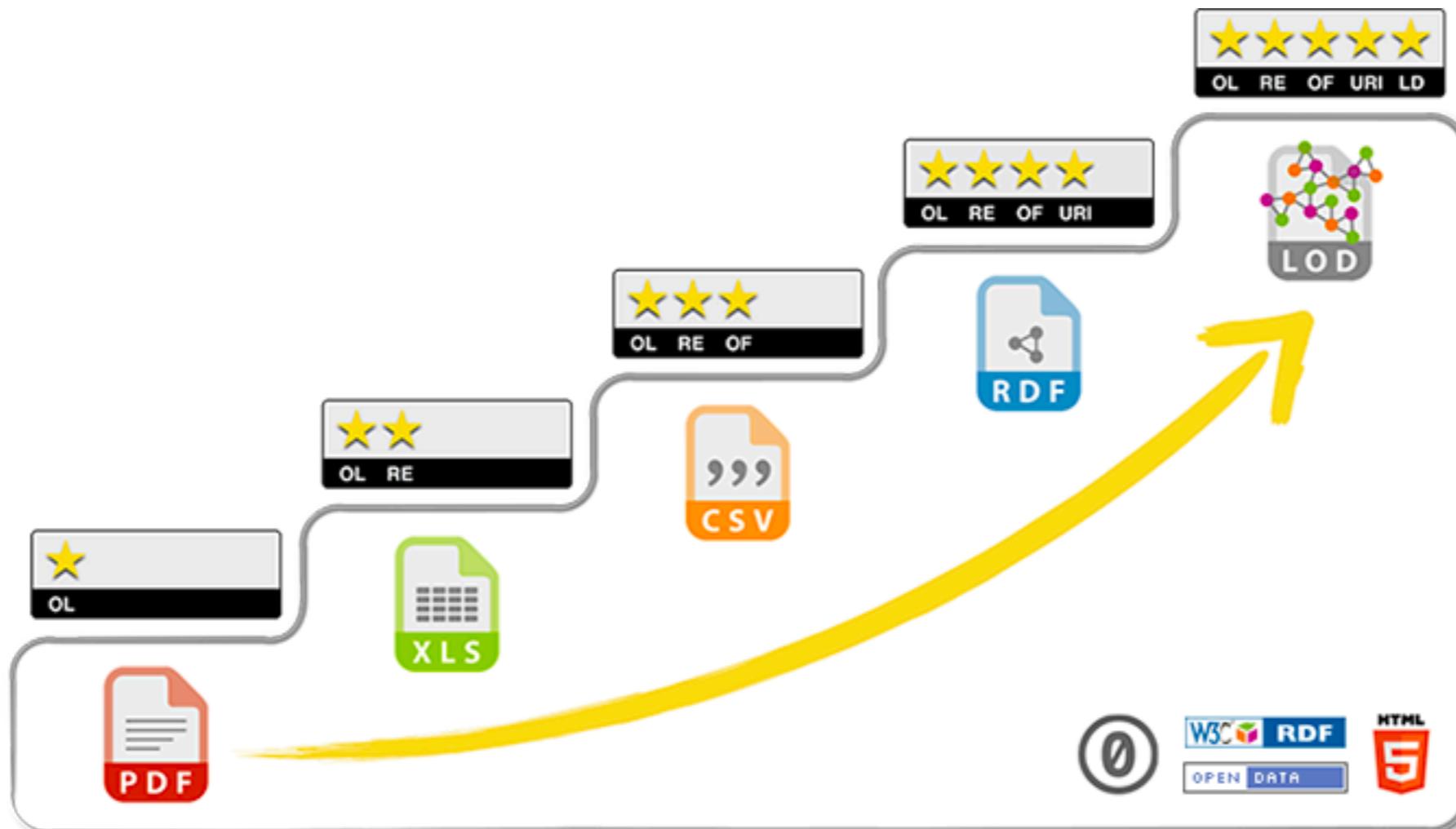
Lab visualisation style



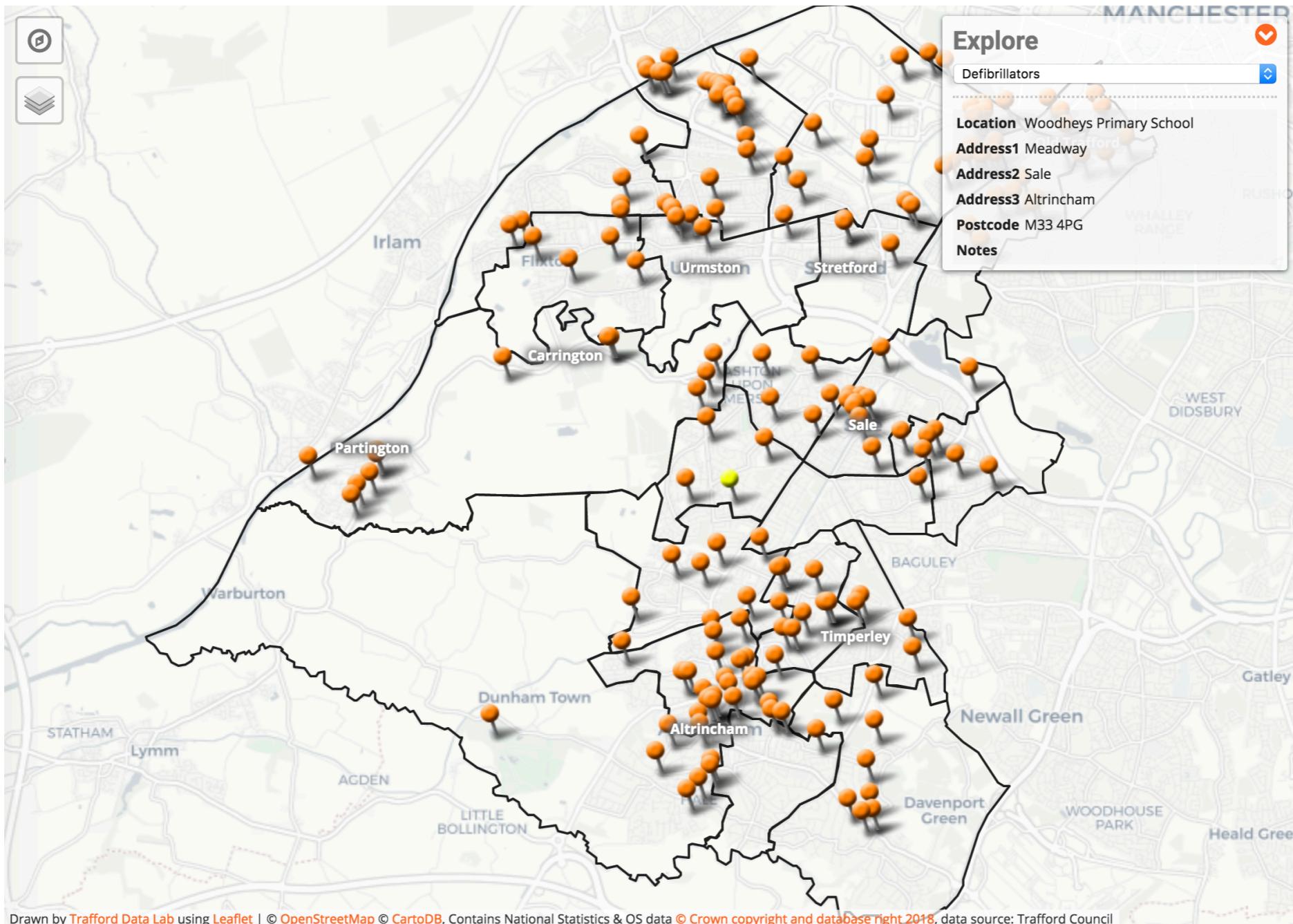
Openness

We are committed to the sourcing,
preparation and publication of
open data.

5-star Open Data



Defibrillators



Open data portal

3-s
to
Clic
on

D

We

Dataset name Defibrillators

Dataset description Defibrillator locations within Trafford

Source Councillor Butt's Mayoral Charity, [Hand on Heart](#) and crowd sourcing

Publisher Trafford Council

Publisher URL https://github.com/traffordDataLab/open_data/tree/master/defibrillators

Geography Point data

Geographic coverage Trafford

Temporal coverage Correct as of January 2017

Update frequency Not known

Licence [Open Government Licence](#)

Format CSV, GeoJSON, JSON

Openness rating ★★★☆☆ Structured data in open format (e.g. CSV)

Last updated January 2017

Notes Please note that these defibrillators may have restrictions on public availability and may not be available 24/7. The locations may not be exact as they are derived from postcodes.

[CSV](#) [GeoJSON](#) [JSON](#)

Info Maps Projects Data About

Ashton upon Mersey

Location Address1 Address2 Address3 Postcode Notes

Woodheys Primary School Meadoway Sale Altrincham M33 4PG

AECOM Aecom House 179 Moss Lane Altrincham W

Airkikx 9 Trafford Way Trafford Park Manchester M

All Saints Catholic Primary Cedar Road Sale M

trafford_defibrillators_styled.geojson rendered with ❤ by GitHub

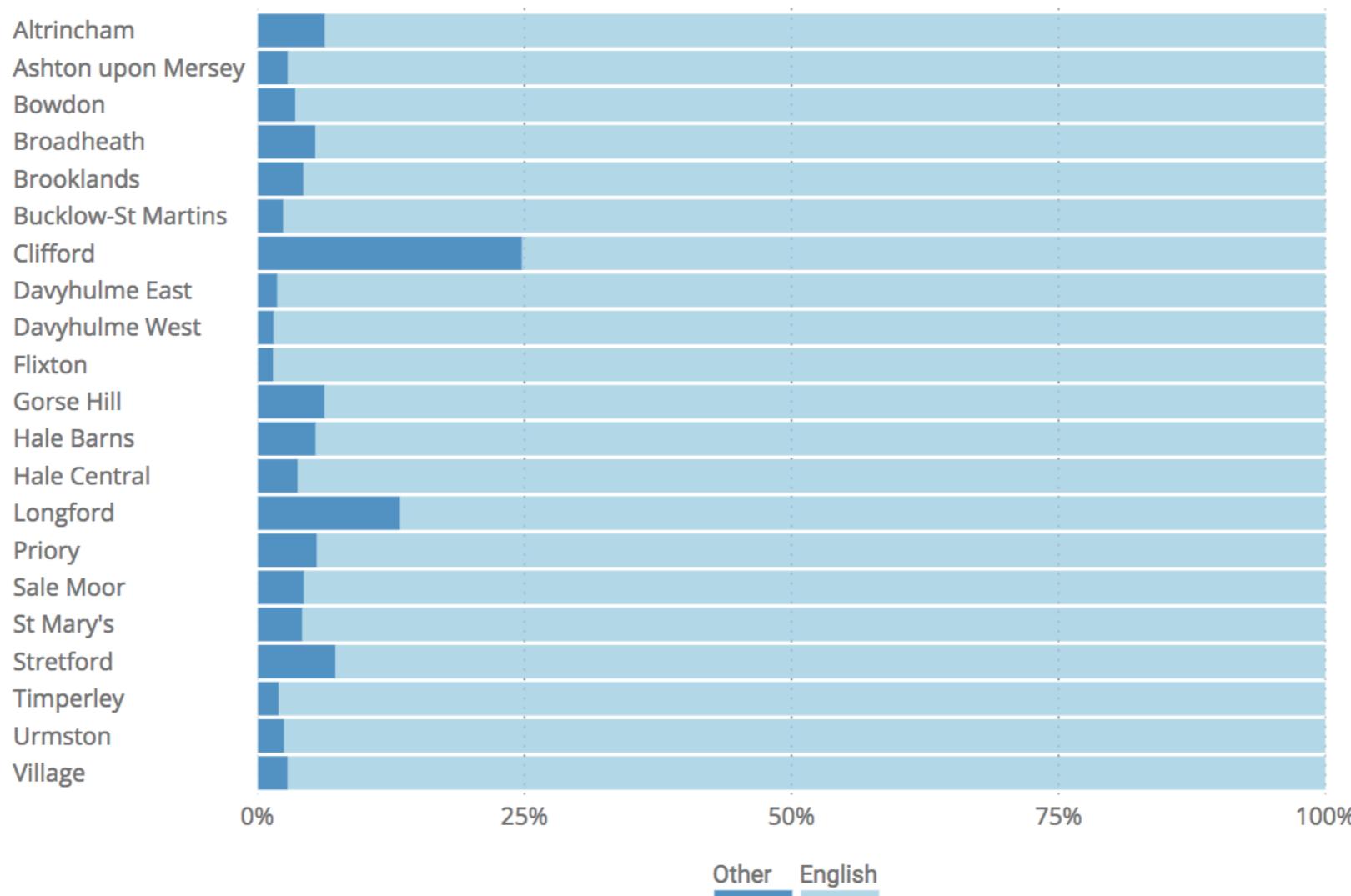
CSV JSON CSV JSON CSV JSON

Reproducibility

We will publish the data and code
that drive our data visualisations.

Example visualisation

Figure 1: Main language by ward, 2011

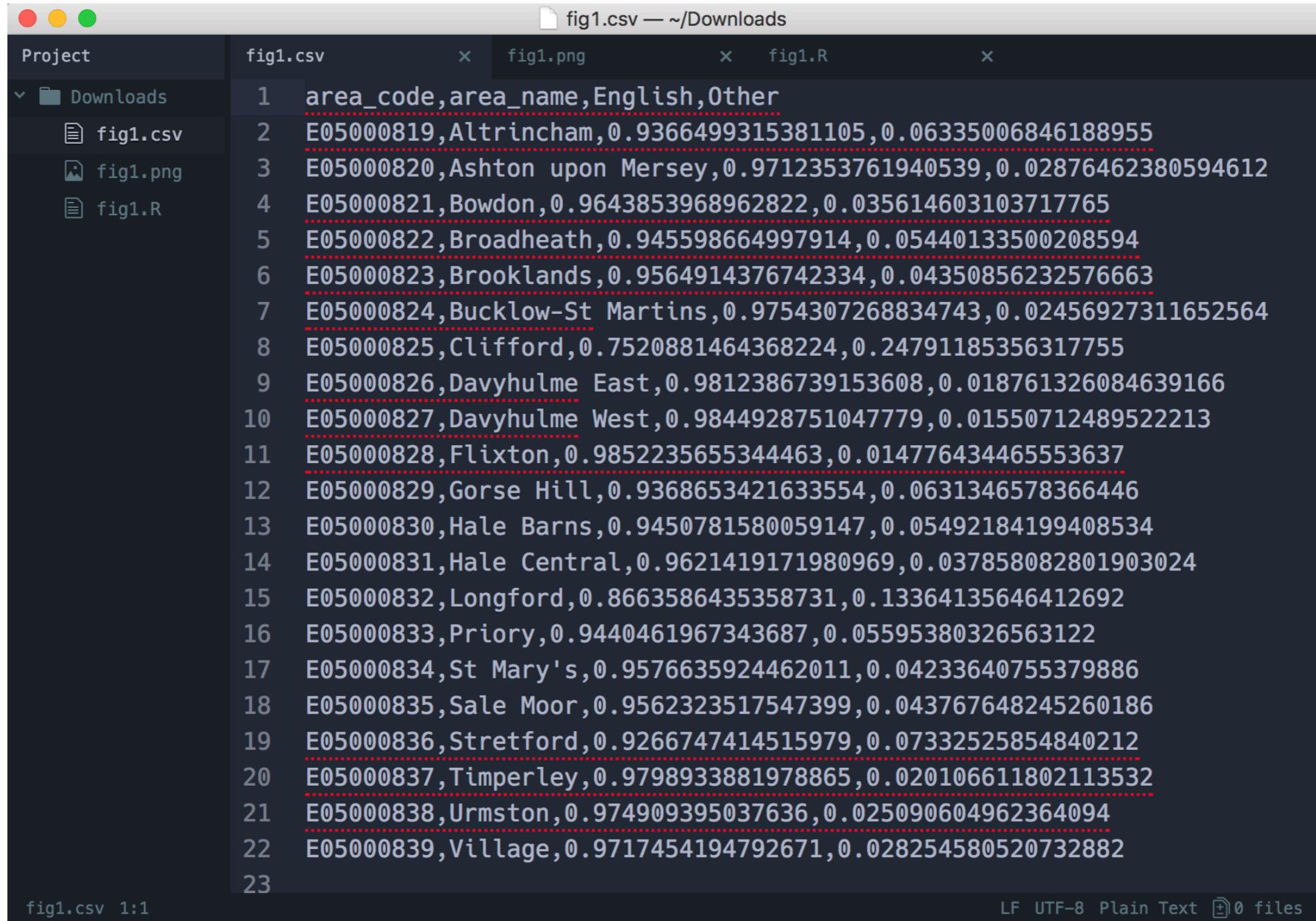


Source: 2011 Census | @traffordDataLab

Download: [Data](#) | [Image](#) | [R code](#)

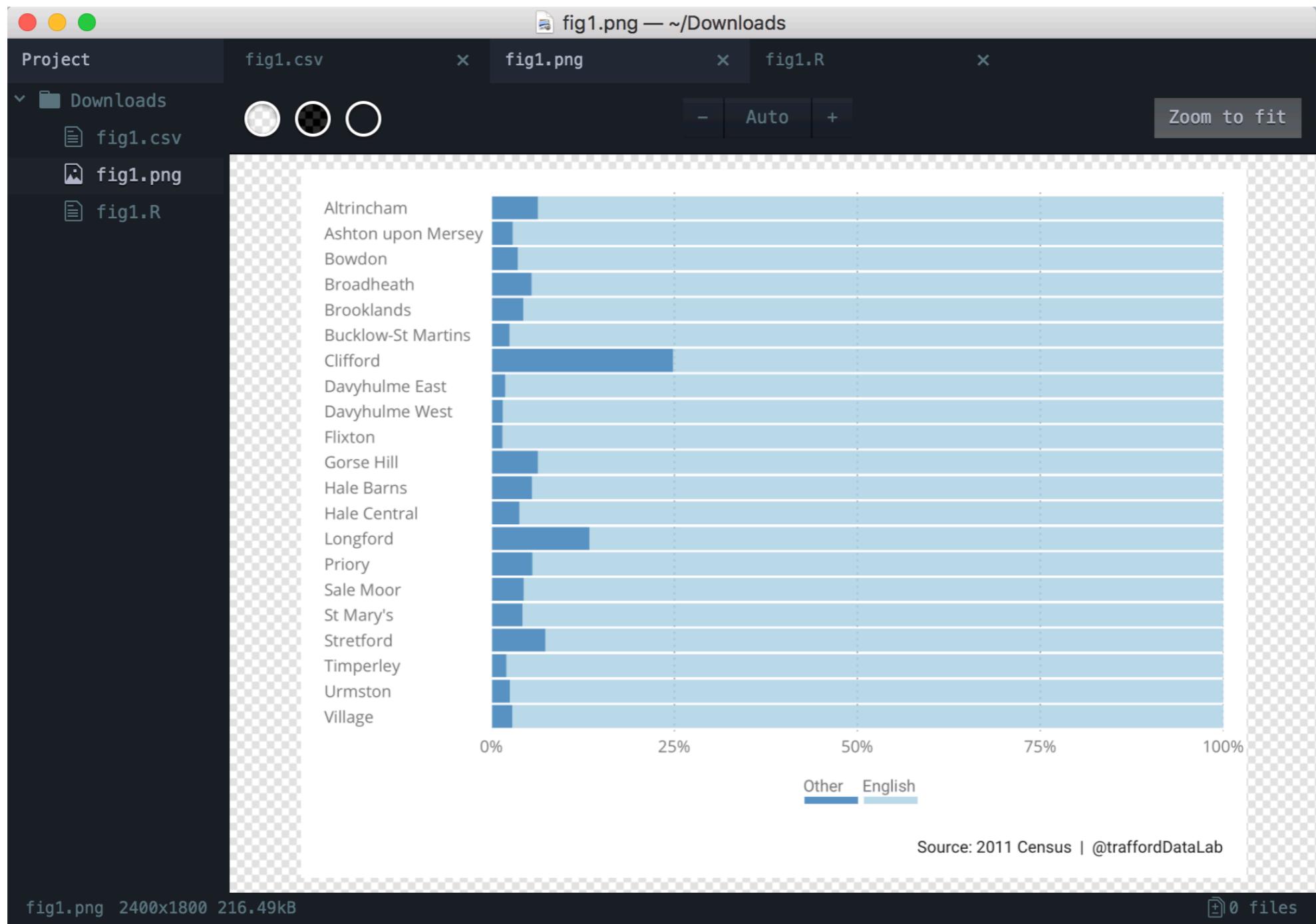
</> <https://www.trafforddatalab.io/info/demographics/languages/R/fig1.R>

Data

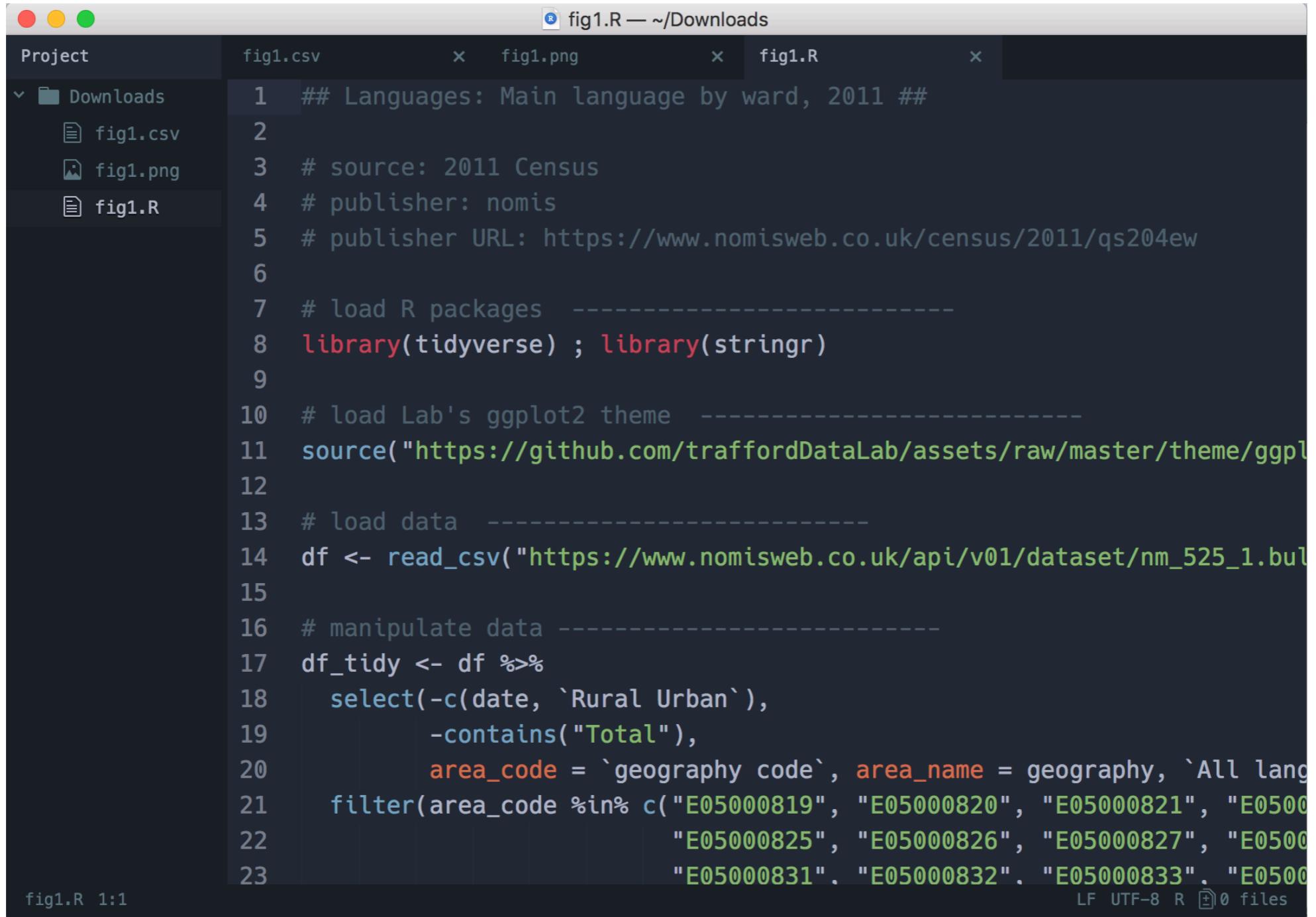


```
fig1.csv — ~/Downloads
Project fig1.csv fig1.png fig1.R
Downloads
  fig1.csv
  fig1.png
  fig1.R
1 area_code,area_name,English,Other
2 E05000819,Altrincham,0.9366499315381105,0.06335006846188955
3 E05000820,Ashton upon Mersey,0.9712353761940539,0.02876462380594612
4 E05000821,Bowdon,0.9643853968962822,0.035614603103717765
5 E05000822,Broadheath,0.945598664997914,0.05440133500208594
6 E05000823,Brooklands,0.9564914376742334,0.04350856232576663
7 E05000824,Bucklow-St Martins,0.9754307268834743,0.02456927311652564
8 E05000825,Clifford,0.7520881464368224,0.24791185356317755
9 E05000826,Davyhulme East,0.9812386739153608,0.018761326084639166
10 E05000827,Davyhulme West,0.9844928751047779,0.01550712489522213
11 E05000828,Flixton,0.9852235655344463,0.014776434465553637
12 E05000829,Gorse Hill,0.9368653421633554,0.0631346578366446
13 E05000830,Hale Barns,0.9450781580059147,0.0549218419940834
14 E05000831,Hale Central,0.9621419171980969,0.037858082801903024
15 E05000832,Longford,0.8663586435358731,0.13364135646412692
16 E05000833,Priory,0.9440461967343687,0.05595380326563122
17 E05000834,St Mary's,0.9576635924462011,0.04233640755379886
18 E05000835,Sale Moor,0.9562323517547399,0.043767648245260186
19 E05000836,Stretford,0.9266747414515979,0.07332525854840212
20 E05000837,Timperley,0.9798933881978865,0.020106611802113532
21 E05000838,Urmston,0.974909395037636,0.025090604962364094
22 E05000839,Village,0.9717454194792671,0.028254580520732882
23
```

Image



R code

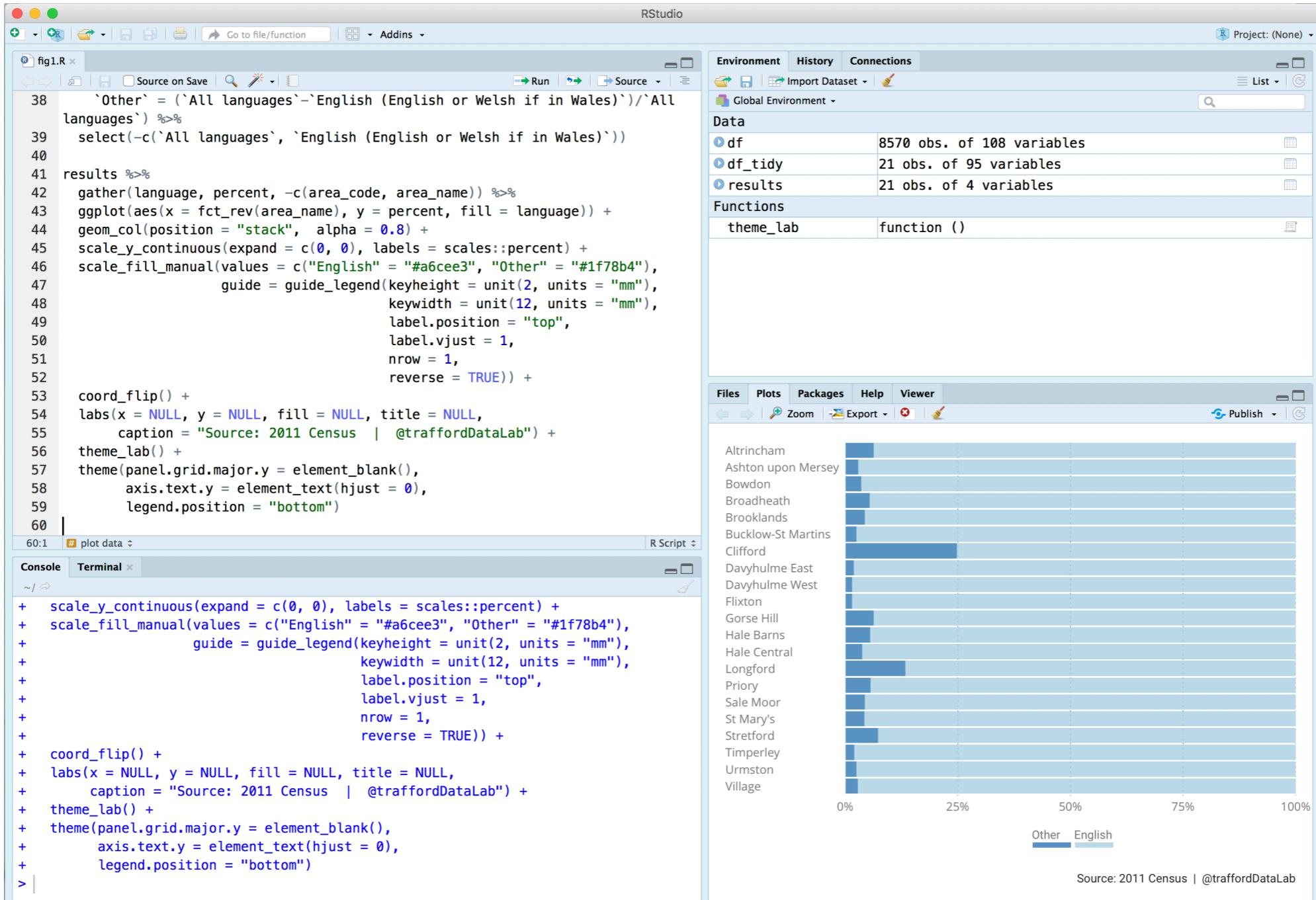


The screenshot shows an RStudio interface with a dark theme. The title bar says "fig1.R — ~/Downloads". The left sidebar shows a "Project" section with "Downloads" expanded, containing files "fig1.csv", "fig1.png", and "fig1.R". The main code editor area contains the following R code:

```
1 ## Languages: Main language by ward, 2011 ##
2
3 # source: 2011 Census
4 # publisher: nomis
5 # publisher URL: https://www.nomisweb.co.uk/census/2011/qs204ew
6
7 # load R packages -----
8 library(tidyverse) ; library(stringr)
9
10 # load Lab's ggplot2 theme -----
11 source("https://github.com/traffordDataLab/assets/raw/master/theme/ggplot2.R")
12
13 # load data -----
14 df <- read_csv("https://www.nomisweb.co.uk/api/v01/dataset/nm_525_1.bul")
15
16 # manipulate data -----
17 df_tidy <- df %>%
18   select(-c(date, `Rural Urban`),
19         -contains("Total"),
20         area_code = `geography code`, area_name = geography, `All languages` = `language code`),
21   filter(area_code %in% c("E05000819", "E05000820", "E05000821", "E05000822",
22                           "E05000825", "E05000826", "E05000827", "E05000828",
23                           "E05000831", "E05000832", "E05000833", "E05000834"))
24
25
```

At the bottom of the code editor, it says "fig1.R 1:1" and "LF UTF-8 R 0 files".

R and (RStudio)



Reproducibility in R

- Re-run code or adapt with different data
- R is a language so it is readable as text
- Post R code via Twitter, reddit, email etc

Graphics companion

Code ▾

ggplot2 graphics companion

- Setup
- Change over time
- Correlation
- Deviation
- Distribution
- Flow
- Magnitude
- Part-to-whole
- Ranking
- Spatial
- Useful resources

Trafford Data Lab

ggplot2 graphics companion

Last updated: 1 March 2018

The Graphics Companion provides the R code for different data visualisations that can be created using the `ggplot2` package.

The Companion adopts the structure of the [Financial Times' Visual Vocabulary](#) by categorising different chart types by the data relationships that they best illustrate.

The data used throughout the Companion derive from a subset of Hans Rosling's [Gapminder World](#) which are available in the `gapminder` R package. Data on life expectancy at birth, GDP per capita and total population are provided for 142 countries between 1952 and 2007.

Setup

You need to install - but only once - the `tidyverse` package and load it into your R session. `ggplot2` is part of the tidyverse suite of R tools for data science.

Hide

```
# install.packages('tidyverse')
library(tidyverse)
```

All of the example plots below use data contained in the `gapminder` R package which also needs to be installed / loaded:

Git and GitHub



Git and GitHub

This organization Search Pull requests Issues Marketplace Explore

Trafford Data Lab

Supporting decision-making in Trafford by revealing patterns in data through visualisation #opendata #opensource

Trafford http://www.trafforddata... infotrafford@trafford.gov.uk

Repositories 16 People 3 Teams 0 Projects 0 Settings

Pinned repositories

Customize pinned repositories

open_data projects info

Open data provided in various formats. Information and outputs regarding projects involving Trafford Data Lab. Report and profile based content for the Trafford Data Lab website.

R ★ 2 R ★ 2 R

Search repositories... Type: All Language: All New

info

Report and profile based content for the Trafford Data Lab website.

R Updated 19 hours ago

Top languages

R HTML

assets

Content such as templates, themes, logos etc. giving our outputs a consistent appearance.

HTML ★ 1 MIT Updated 19 hours ago

People 3 >

IGRiosTC James Austin

itsozz James Austin

Thank you!

[@trafforddatalab](#)