Session 1: Introduction to R and RStudio

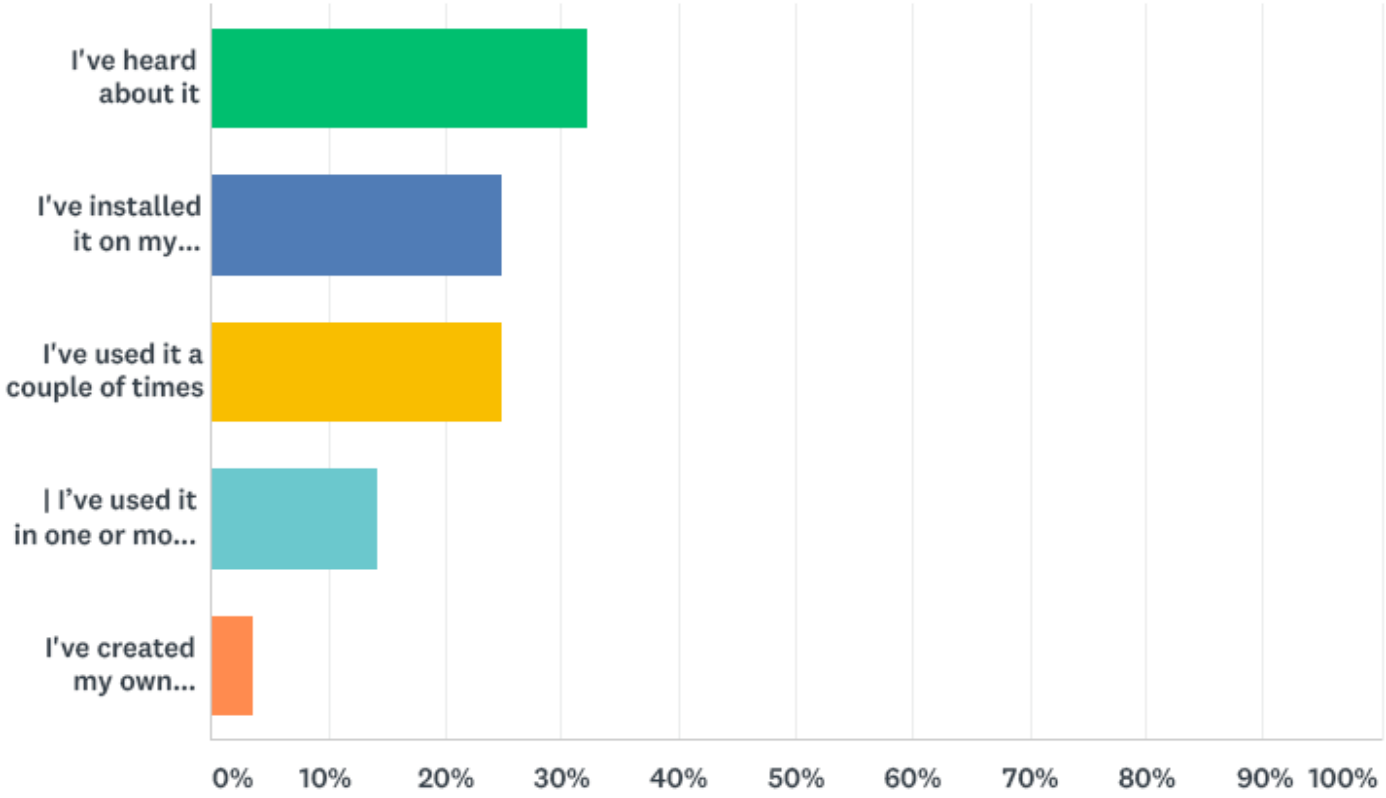
Andrew Jones

andrew.jones40@nhs.net

Healthcare Analyst | The Strategy Unit

Q5 Which one of these options best describes your familiarity with R?

Answered: 28 Skipped: 0



Agenda – Day 1

Intro + Setup

Refreshments
11:20 – 11:35 →

Graphics with ggplot2

Data wrangling with dplyr

Lunch (13:30 – 14:30)

Objects in R | Import data

Review and questions

Agenda – Day 2

Relational data | Tidy data

EDA exercise

Refreshments
11:20 – 11:35

Data structures | R resources

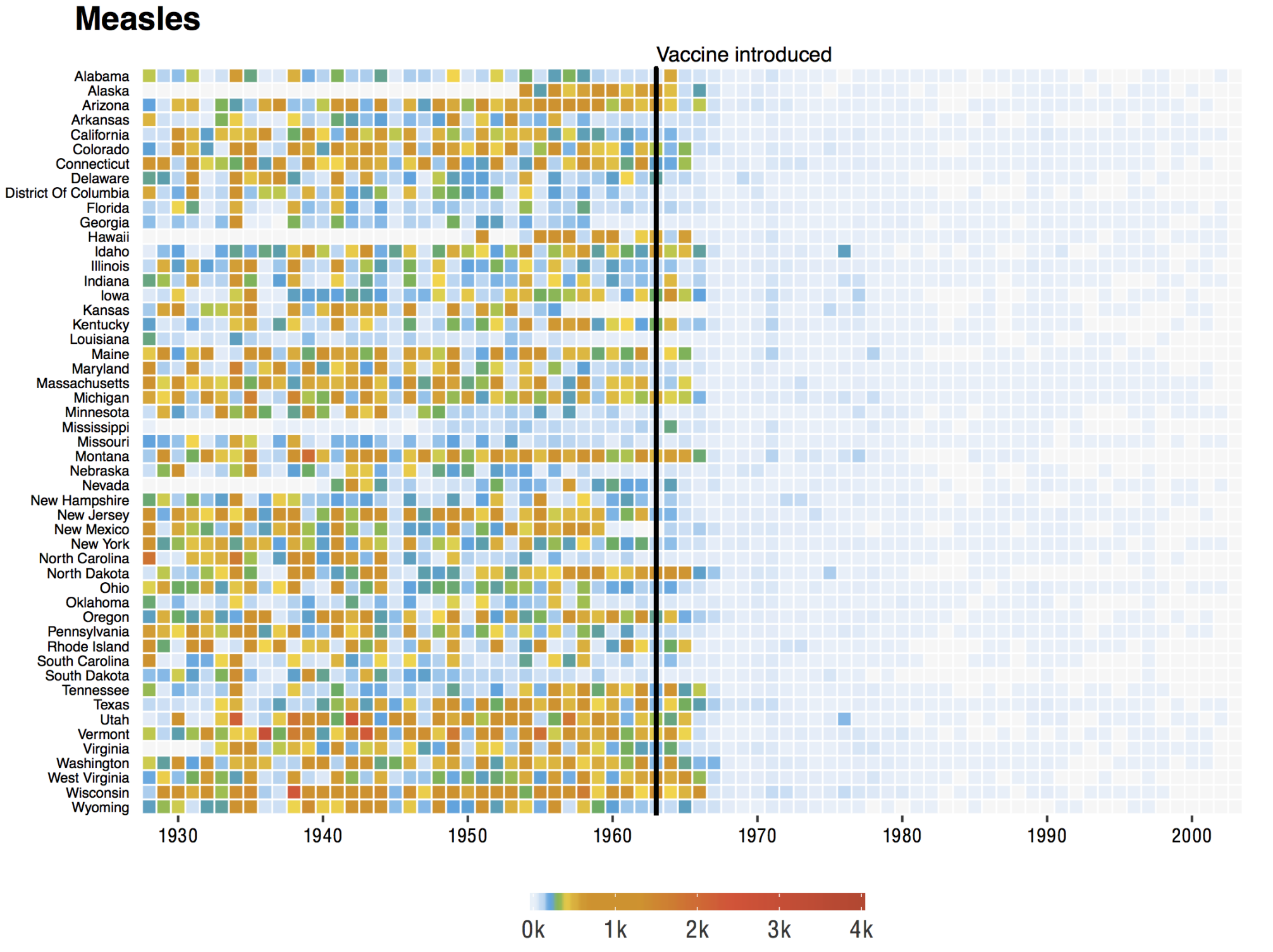
Lunch (13:30 – 14:30)

Exercise | Self-directed learning

Review and questions

Course Aims

1. We can show you (some of) the possibilities:


Measles
Vaccine introduced
Alabama
Alaska
Arizona
Arkansas
California
Colorado
Connecticut
Delaware
District Of Columbia
Florida
Georgia
Hawaii
Idaho
Illinois
Indiana
Iowa
Kansas
Kentucky
Louisiana
Maine
Maryland
Massachusetts
Michigan
Minnesota
Mississippi
Missouri
Montana
Nebraska
Nevada
New Hampshire
New Jersey
New Mexico
New York
North Carolina
North Dakota
Ohio
Oklahoma
Oregon
Pennsylvania
Rhode Island
South Carolina
South Dakota
Tennessee
Texas
Utah
Vermont
Virginia
Washington
West Virginia
Wisconsin
Wyoming
1930
1940
1950
1960
1970
1980
1990
2000
0k
1k
2k
3k
4k
https://benjaminlmoore.wordpress.com/tag/ggplot2/

Stats

OPEN ACCESS

Changes in admission thresholds in English emergency departments

Steven Wyatt,[1] Kieran Child,[1] Andrew Hood,[1] Matthew Cooke,[2] Mohammed A Mohammed[3]

► Additional material is published online only. To view please visit the journal online (http://dx.doi.org/10.1136/emermed-2016-206213).

[1]Strategy Unit, NHS Midlands and Lancashire Commissioning Support Unit, West Bromwich, UK
[2]Warwick Medical School, University of Warwick, Coventry, UK
[3]Faculty of Health Studies, University of Bradford, Bradford, UK

Correspondence to
Steven Wyatt, Strategy Unit - NHS Midlands and Lancashire Commissioning Support Unit Kingston House 438-450 High Street West Bromwich B70 9LD UK; swyatt@nhs.net

Received 16 July 2016
Revised 22 June 2017
Accepted 28 June 2017

ABSTRACT

Background The most common route to a hospital bed in an emergency is via an Emergency Department (ED). Many recent initiatives and interventions have the objective of reducing the number of unnecessary emergency admissions. We aimed to assess whether ED admission thresholds had changed over time taking account of the casemix of patients arriving at ED.

Methods We conducted a retrospective cross-sectional analysis of more than 20 million attendances at 47 consultant-led EDs in England between April 2010 and March 2015. We used mixed-effects logistic regression to estimate the odds of a patient being admitted to hospital and the impact of a range of potential explanatory variables. Models were developed and validated for four attendance subgroups: ambulance-conveyed children, walk-in children, ambulance-conveyed adults and walk-in adults.

Results 23.8% of attendances were for children aged under 18 years, 49.7% were female and 30.0% were conveyed by ambulance. The number of ED attendances increased by 1.8% per annum between April 2010–March 2011 (year 1) and April 2014–March 2015 (year 5). The proportion of these attendances that

Key messages

What is already known on this subject?

- The most common route to a hospital bed in an emergency is via an ED.
- Many recent initiatives and interventions have the objective of reducing the number of unnecessary emergency admissions.
- Several studies have identified patient and attendance characteristics that are associated with increased risk of admission.

What this study adds?

- The casemix-adjusted odds of admission via ED to NHS hospitals in England have decreased since April 2010.
- The number of attendances that have a low probability of admission has reduced since April 2010.

Many recent NHS policy initiatives and commissioning interventions have been designed to avoid

Statistical methods

We used mixed-effects logistic regression to estimate the association between the odds of a patient being admitted to hospital and the impact of each of the potential explanatory variables. Initial univariate analysis indicated markedly different unadjusted odds of admission for children and adults and between patients conveyed by ambulance and those who arrived by some other means (henceforth referred to as walk-ins). These differences remained substantial having adjusted for other covariates. Given that the records available to build models were plentiful, to minimise the reliance on interaction terms and to increase the model accuracy, we developed four separate models for ambulance-conveyed children, walk-in children, ambulance-conveyed adults and walk-in adults.

All candidate predictor variables were included in the model on the basis of adequately strong univariate association with the outcome variable or because inclusion improved the fit of the multivariate model. The hospital trust (provider) of the ED was included as a random effect to reflect clustering of attendances within hospital trusts. Final model fit was measured using the C-statistic (area under the receiver operating characteristic curve), calibration plots and the Hosmer-Lemeshow goodness-of-fit test.

The impact of adding interaction terms was tested as a sensitivity analysis. All two-way interaction terms between the fixed effects (other than arrival year) were tested individually and in combination, for their impact on the coefficients for the arrival year variable. In addition, the coefficients of interaction terms between arrival year and each of the other fixed effects were assessed.

To further explore changes in admission thresholds over time, we applied the casemix-adjusted odds of admission from year 1 to attendances in years 2–5.

Data processing was conducted in Microsoft SQL Server 2012 and analysis in R V.3.2.3 statistical software package.

Machine Learning

Automated reports

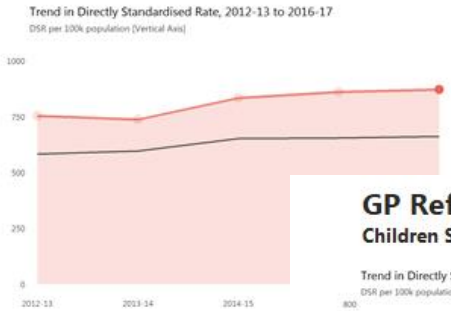
Inpatient Summary

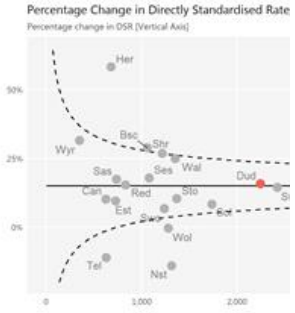
Opportunity	Admissions	2016-17 Spend (000s)	Rate	Rate of Change
ACS Acute	2,750	£3,445	High	High
ACS Chronic	2,680	£4,856	Low	-
ACS Vaccine	1,770	£5,085	High	-
Alcohol (wholly)	1,220	£1,724	High	High
Alcohol (partially - chronic)	4,970	£7,775	High	Low
Alcohol (partially - acute)	1,930	£2,506	High	High
End of Life Care (3-14 days)	320	£1,030	High	-
End of Life Care (0-2 days)	200	£305	High	-
Falls	3,620	£8,873	High	-
Frail Elderly (occasional)	1,090	£2,349	High	-
Frail Elderly (usual)				
Medically Unexplained				
Medicines - Explicit				
Medicines - Implicit AntiDiab				
Medicines - Implicit Benzo				
Medicines - Implicit Diuretics				
Medicines - Implicit NSAIDs				
Obesity (largely)				
Obesity (marginal)				
Obesity (somewhat)				
PLCV Cosmetic				
PLCV Alternative				
PLCV Ineffective				
PLCV Risks				
Mental Health Admissions from ED				
Self-harm				
Smoking (large)				
Smoking (somewhat)				
Zero Length of Stay (adult)				
Zero Length of Stay (child)				

Notes: Rate and rate of change comp

Frail Elderly Admissions

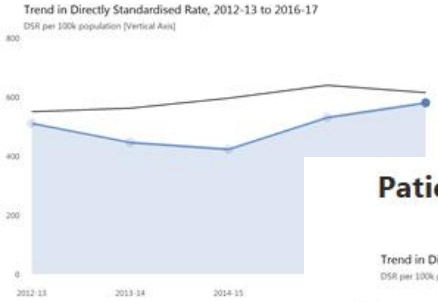
Could Usually be Managed in a Non-Acute Setting



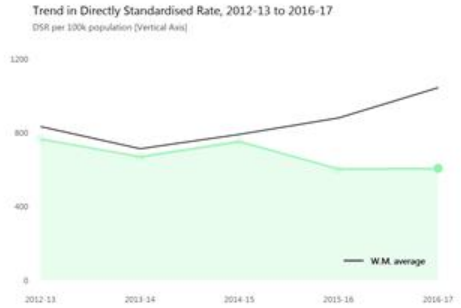


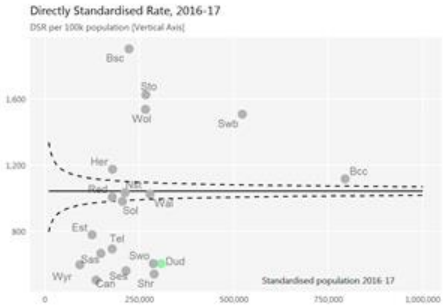
GP Referred First Outpatient Attendances

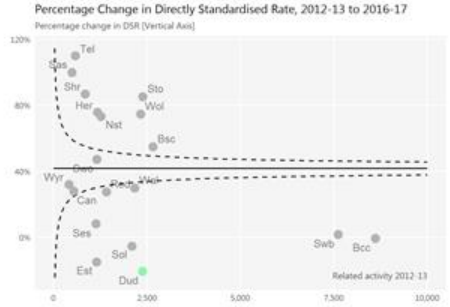
Children Surgical Specialties



Patients Leaving ED Before Being Seen

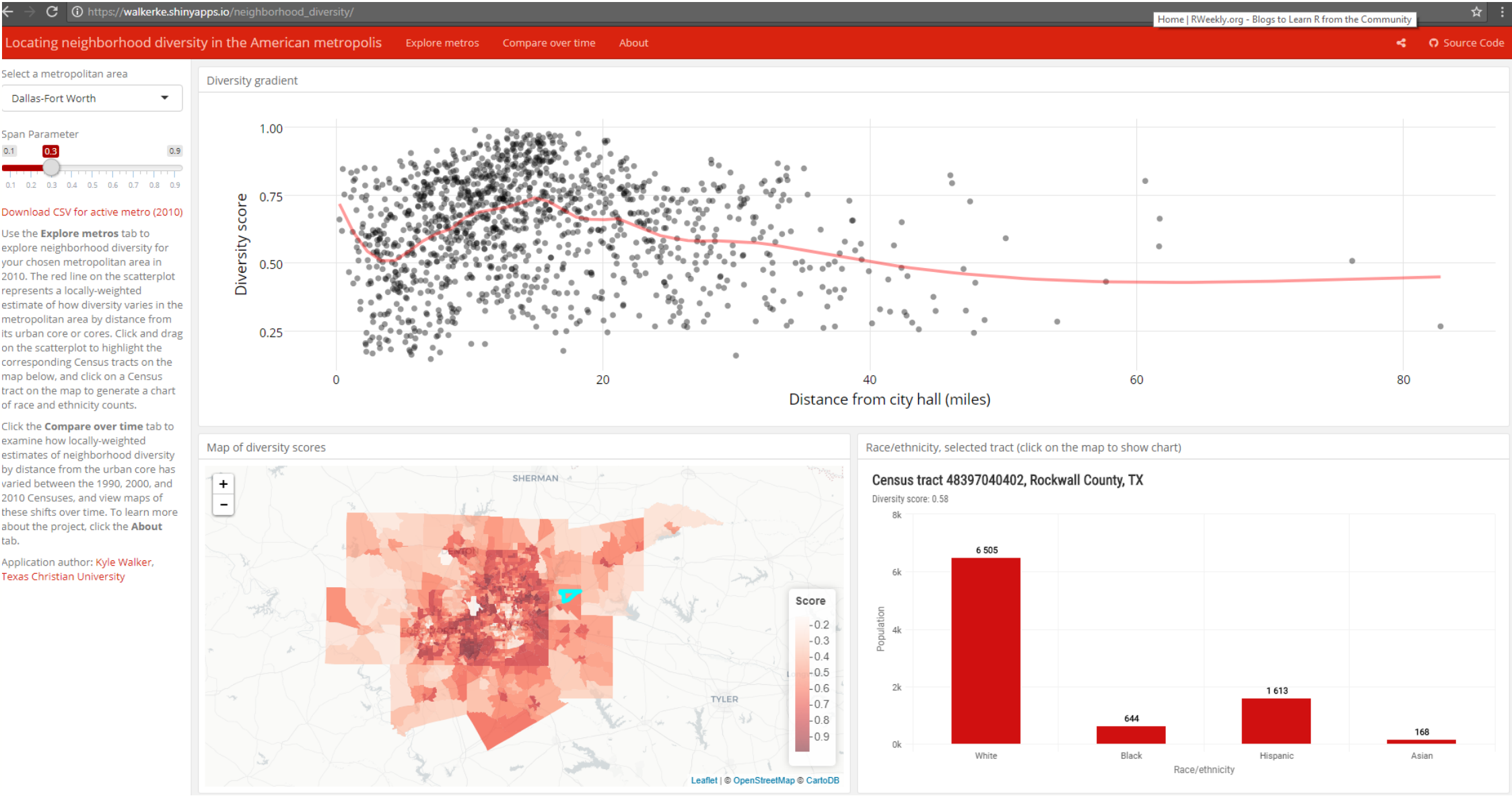






1,860 attendances	**£0.1M** spent
2.1% of all attendances	**£70** per unit activity

(Interactive) Dashboards



https://walkerke.shinyapps.io/neighborhood_diversity/

(Interactive) Dashboards

Diversity gradient

```
# Here, we draw the diversity gradient with ggplotly
output$scatter <- renderPlotly({

  key <- metro()$tractid # This will uniquely identify tracts for Plotly

  p1a <- ggplot(metro()@data) +
    geom_point(alpha = 0.4, aes(Distance, Score, key = key)) +
    theme_minimal(base_size = 14) +
    stat_smooth(aes(Distance, Score),
                color = 'red', method = 'loess', span = input$span, se = FALSE) +
    xlab('Distance from city hall (miles)') + ylab('')

  g <- ggplotly(p1a, source = 'source') %>%
    layout(dragmode = 'lasso',
           yaxis = list(title = 'Diversity score'),
           margin = list(l = 100),
           font = list(family = 'Open Sans', size = 16))

  # Need to manually set the hoverinfo to avoid the key appearing in it
  build <- plotly_build(g)
```

https://walkerke.shinyapps.io/neighborhood_diversity/

R-bloggers

R news and tutorials contributed by (750) R bloggers

RcppCNPy 0.2.9

March 22, 2018

By Thinking inside the box

Another minor maintenance release of the RcppCNPy package arrived on CRAN this evening. RcppCNPy provides R with read and write access to NumPy files thanks to the cnpy library by Carl Rogers. There is only small code change: a path is now checked be...

Read more »

The most prolific package maintainers on CRAN

March 22, 2018

By David Smith

During a discussion with some other members of the R Consortium, the question came up: who maintains the most packages on CRAN? DataCamp maintains a list of most active...

Read more »

Comparing additive and multiplicative regressions using AIC in R

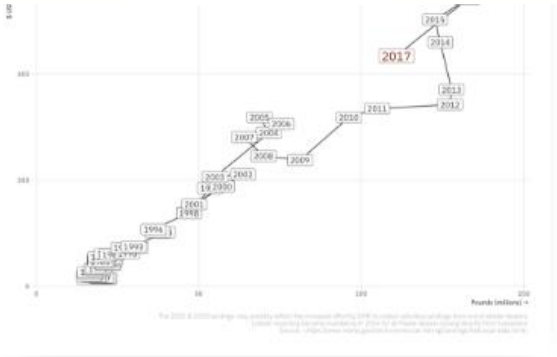
March 22, 2018

By Ivan Svetunkov

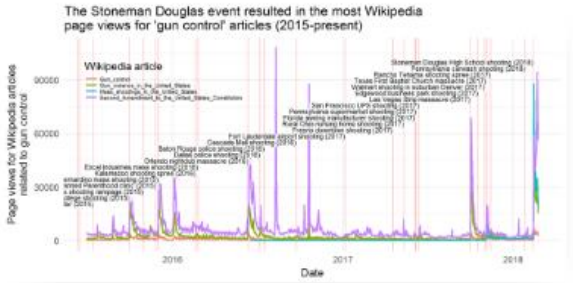
One of the basic things the students are taught in statistics classes is that the comparison of models using information criteria can only be done when the models have...

Read more »

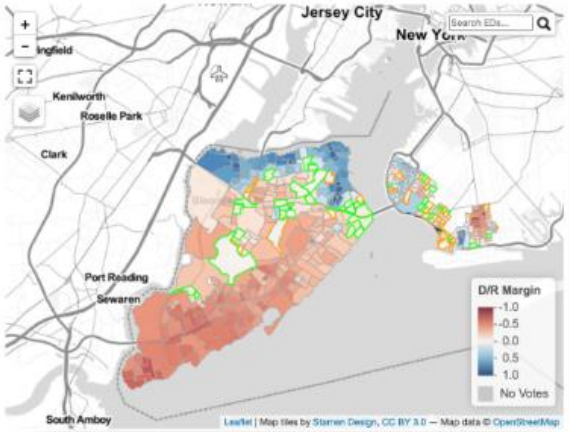
RWeekly.org



- Investigating how NVIDIA's new TITAN V performs in R and a few other applications! (brucemeng.ca)
- Persistence of Public Interest in Gun Control After Mass Shooting Events in the US: Is the Stoneman Douglas High School Event Different? (rawgit.com)

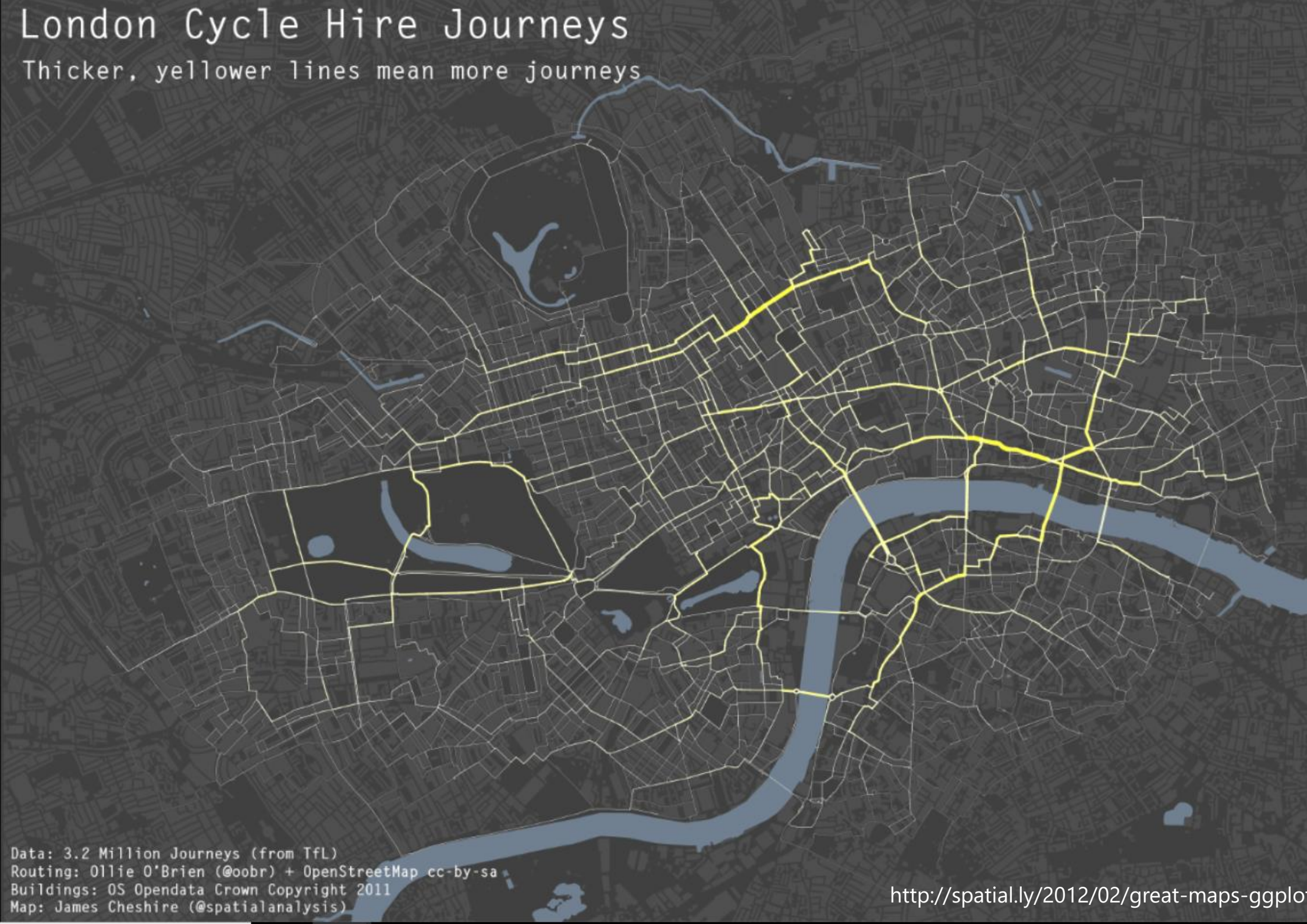


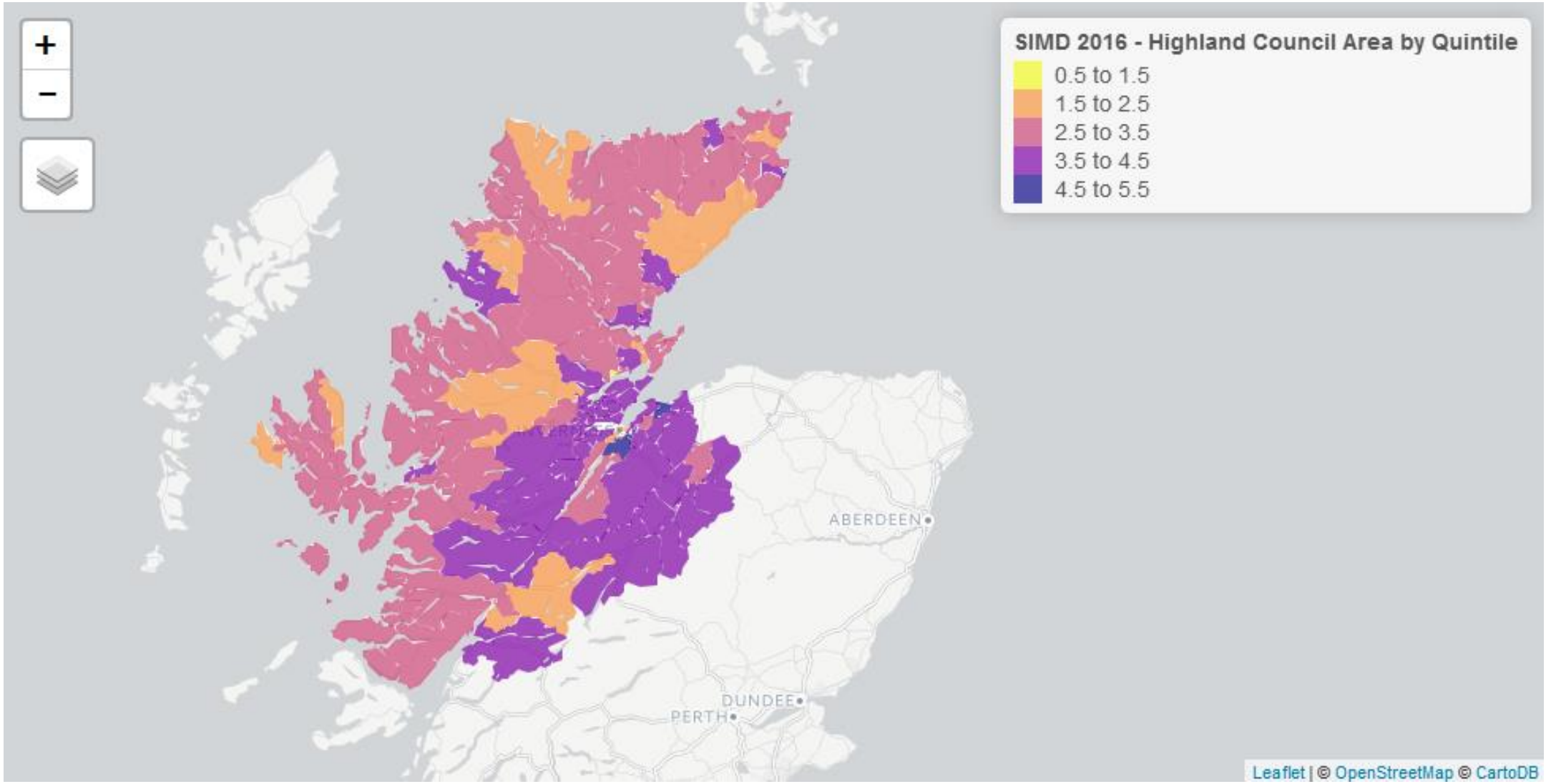
- Stability testing: How do you know whether your single-cell clusters are 'real'? (jef.works)
- EDA using fiftystater (pradeepadhokshaja.wordpress.com)
- Mapping how Staten Island voted in 2016 (and what it might mean for 2018) (mattherman.info)



Version Control

```
82  97
83  98  # Functions ------------------------------------------------------------------
84  99  "Careful with summary functions <- many old parameters exist here"
    @@ -176,7 +191,7 @@ plot_trend(plotTrendActive,
176 191             plotTrendActive$DSRate,
177 192             plotTrendComparators$DSRate)
178 193
179     -# plot_cost  <- function(df){
    194 +plot_cost  <- function(df){
180 195    ggplot(df) +
181 196      geom_bar(aes(x = ShortName, y = DSCostsPerHead, fill = IsActiveCCG), stat = "identity") +
182 197      geom_text(
    @@ -210,6 +225,14 @@ plot_fun   <- function(df_funnels, df_units){
210 225                          , yend = target))+
211 226      #geom_hline(aes(yintercept = target)) +
212 227      geom_point(data = df_units, aes(x = DerivedPopulation, y = DSRate, colour = IsActiveCCG), size = 3)+
    228 +    geom_text(data = df_funnels
    229 +              , aes(x = max(n), y = min(fnlLow), label = "Standardised population 2016/17")
    230 +              , vjust = "bottom"
    231 +              , hjust = "right"
    232 +              , family = "Segoe UI"
    233 +              , size   = 3
    234 +              , fontface  = "plain"
    235 +               )+
213 236      # geom_text_repel(data = df_units
214 237      #   , aes(x = DerivedPopulation
215 238      #         , y = DSRate
    @@ -229,7 +252,7 @@ plot_fun   <- function(df_funnels, df_units){
229 252      scale_y_continuous(labels = scales::comma)+
230 253      theme_strategy()+
231 254      theme(legend.position = "none"
232     -          , axis.title.y = element_blank()
    255 +          , axis.title = element_blank()
233 256            #, plot.subtitle = element_text(face = "italic")
234 257            )+
235 258      labs(
```

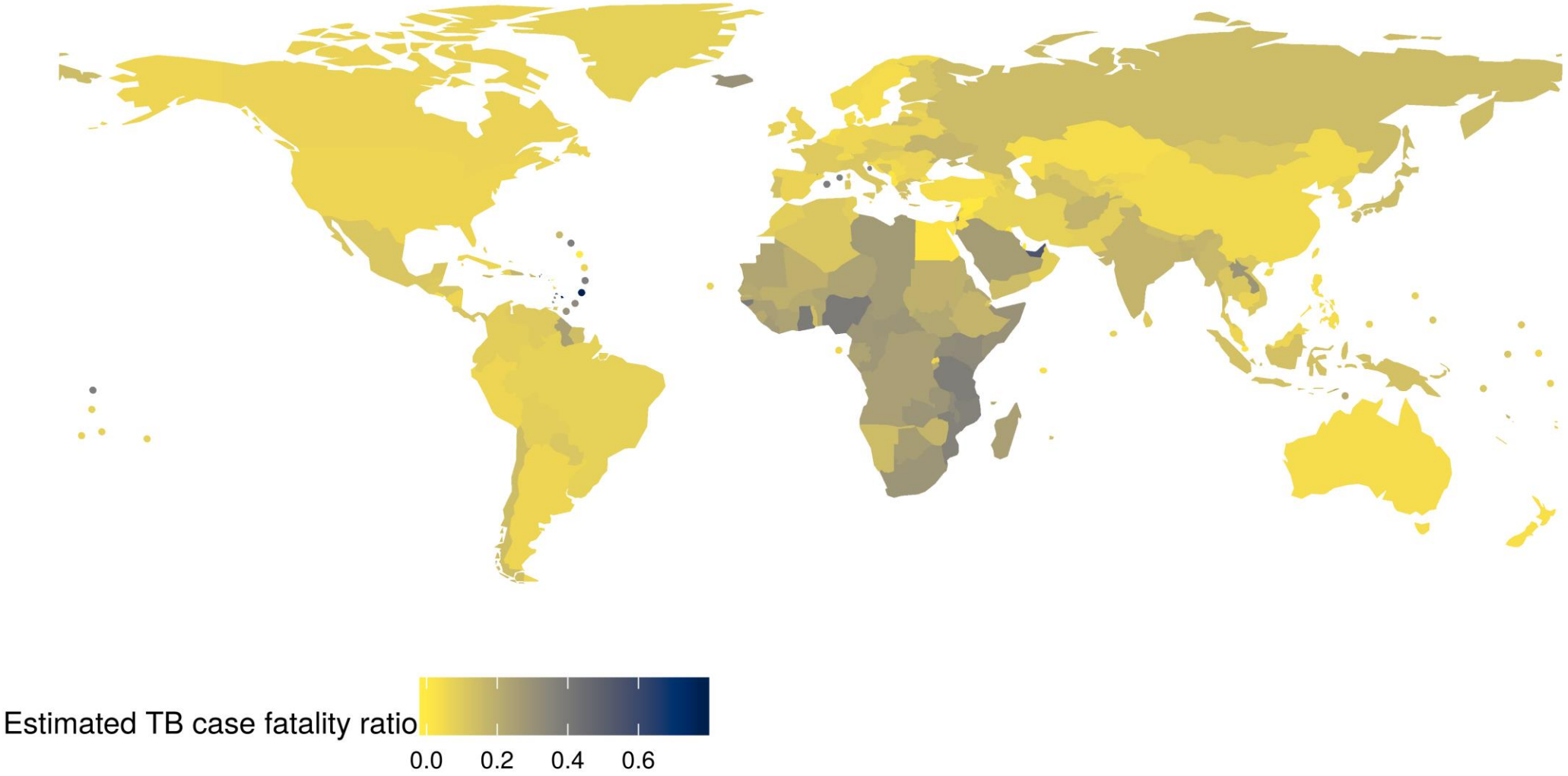

London Cycle Hire Journeys
Thicker, yellower lines mean more journeys
Data: 3.2 Million Journeys (from TfL)
Routing: Ollie O'Brien (@oobr) + OpenStreetMap cc-by-sa
Buildings: OS Opendata Crown Copyright 2011
Map: James Cheshire (@spatialanalysis)
http://spatial.ly/2012/02/great-maps-ggplo



https://www.johnmackintosh.com/2017-09-01-easy-maps-with-tmap/

Map of Tuberculosis Case Fatality Ratio - 2016

Case fatality rate estimated by the WHO



@seabbs Source: WHO
https://www.samabbott.co.uk/post/est-cfr-gettbinr/

Scraping data from the web

Connect R directly to a database

Create books/blogs/websites

README.md

blogdown

build passing downloads 17K

An open-source (GPL-3) R package to generate static websites based on R Markdown and Hugo. You can install the package via CRAN or GitHub:

```
## Install from CRAN
install.packages('blogdown')

## Or, install from GitHub
devtools::install_github('rstudio/blogdown')
```

You may create a new site via the function `blogdown::new_site()` under an *empty* directory. It will create a skeleton site, download a Hugo theme from Github, add some sample content, launch a web browser and you will see the new site. The sample blog post `hello-world.Rmd` should be opened automatically, and you can edit it. The website will be automatically rebuilt and the page will be refreshed after you save the file.

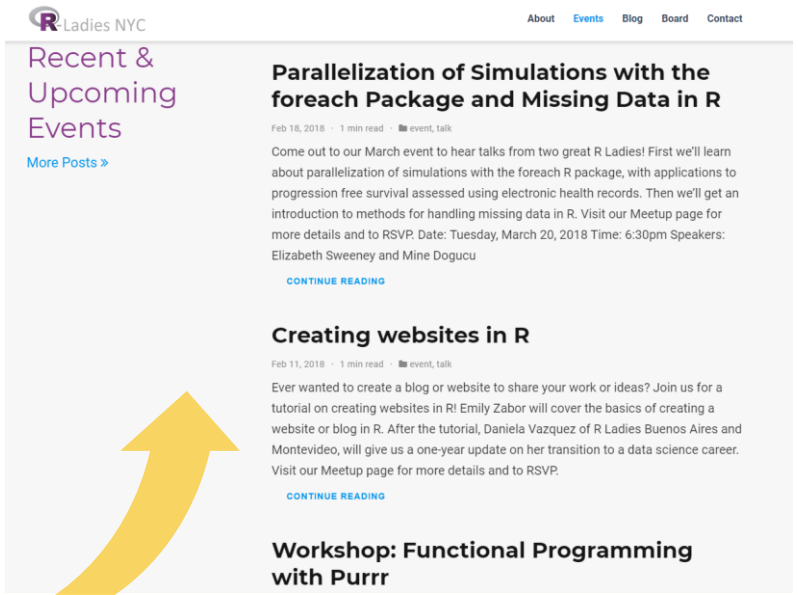
If you use RStudio, you can create a new RStudio project for your website from the menu `File -> New Project -> New Directory -> Website using blogdown`.

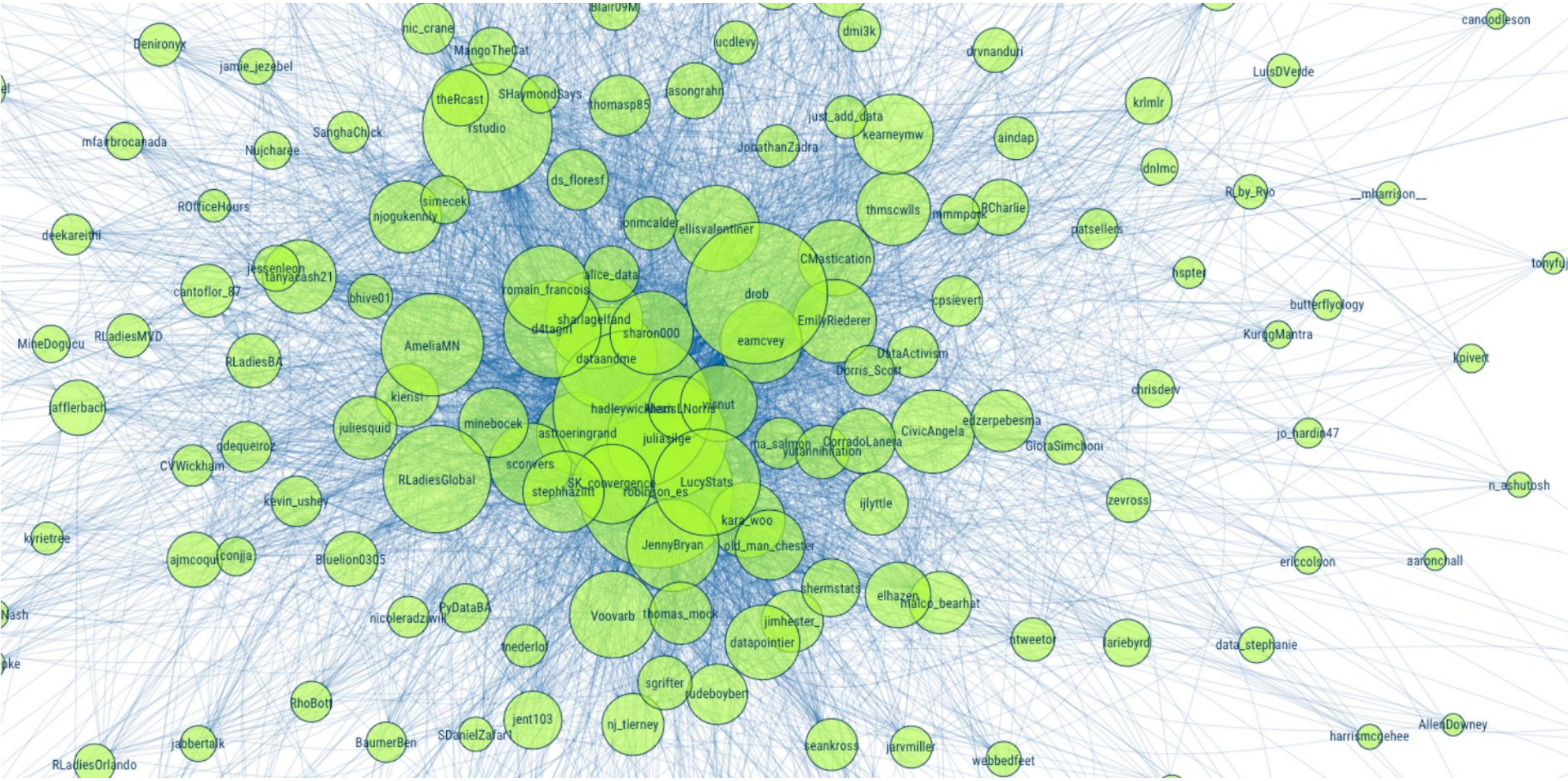
The function `blogdown::serve_site()` may be the most frequently used function in this package. It builds the website, loads it into your web browser, and automatically refreshes the browser when you update the Markdown or R Markdown files. Do not use the command line `hugo server` to build or serve the site. It only understands plain Markdown files, and cannot build R Markdown.

You may not be satisfied with the default site created from `new_site()`. There are two things you may want to do after your first successful experiment with **blogdown**:

1. Pick a Hugo theme that you like from http://themes.gohugo.io. All you need is its Github user and repository name, to be passed to the `theme` argument of `new_site()`.
2. Add more content (pages or posts), or migrate your existing website.

The full documentation is the **blogdown** book freely available at https://bookdown.org/yihui/blogdown/. You are expected to read at least the first chapter. You are welcome to send us feedback using Github issues or ask questions on StackOverflow with the `blogdown` tag.





Community

https://github.com/mkearney/rstudioconf_tweets

stackoverflow Questions Developer Jobs Tags Users Search...

I have a data.frame (df) with 17 rows and 40 columns. I would like to plot all those columns like this:

0

```
windows()
plot(NULL,xlim=c(0,17),ylim=c(5000,90000),xaxt='n',xlab="", ylab="")
points(df$c1,type="b",pch=15,col="gold3")
points(df$c2,type="b",pch=15,col="gold3")
.
.
points(df$c40,type="b",pch=15,col="gold3")
```

I would like to create a loop inside the plot to not have to write all the lines for the 40 columns. I tried different things without success. Thanks in advance!

1 Answer

active oldest votes

Here is an example using standard `plot` and `points` as well as a `ggplot2` example.

This answer is useful

```
ta.frame(x=1:10,
       y1=rnorm(10),
       y2=rnorm(10),
       y3=rnorm(10))
plot(df$x, df$y1)
# points(df$x, df$y2)
# points(df$x, df$y3)
for(i in 3:4) {
  points(df$x, df[[i]])
}


library(reshape2)
library(ggplot2)
melt_df <- melt(df, 'x')
ggplot(melt_df, aes(x, value)) +
  geom_point()
```

share improve this answer

answered 1 hour ago

drmariod

3,503 ● 19 ● 53

Supports Diversity

Tweets 3,551 | Following 2,410 | Followers 4,957 | Likes 4,959 | Lists 3

R-Ladies Global
@RLadiesGlobal

Promoting gender diversity in the #rstats community via meetups, mentorship & global collaboration! 60+ groups worldwide. #RLadies

The World
rladies.org
Joined August 2016
237 Photos and videos

Tweets | Tweets & replies | Media

Pinned Tweet

R-Ladies Global @RLadiesGlobal · 30 Aug 2016
Interested in starting an #RLadies meetup in your city? We'd love to help! Send us an email at info@rladies.org to get in touch. #rstats

17 80 117

R-Ladies Global Retweeted

Forwards @R_Forwards · 3h
Our advice to a black woman who has told her mentor "the tech industry is not for people like me" - what would you say?

And much more...

Biggest Benefits

Community -> Using R broadens your outlook

Must seek out and learn from others' approaches

Encourages the best form of analysis:
precise, transparent, and communicative

Benefits both you and the NHS

Course Aims

1. We can show you the possibilities.

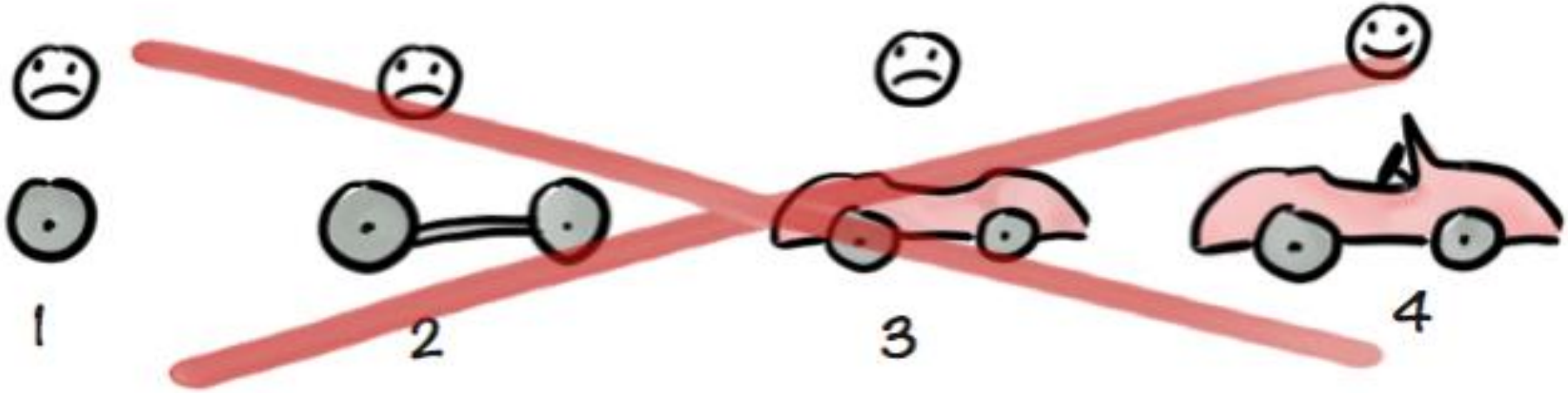
2. We can teach you enough so you can start teaching yourself (myriad blogs, worked examples, free books, and videos available – we will cover resources tomorrow)

Course philosophy

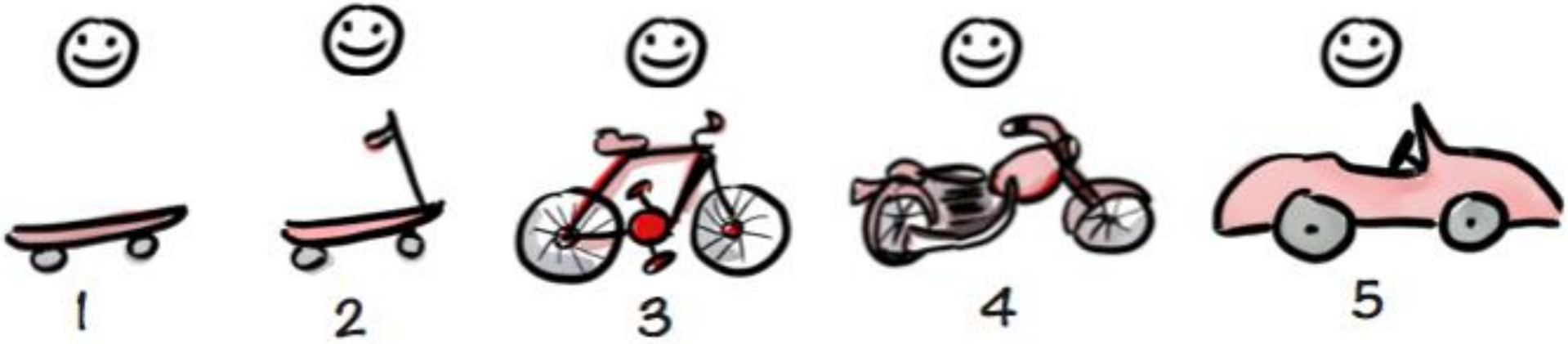
(Mostly) top-down approach:

Learn how now, then add more theory as you progress.

Not like this....



Like this!



Henrik Kniberg

https://blog.crisp.se/2016/01/25/henrikkniberg/making-sense-of-mvp

R and RStudio

R is a programming language

RStudio is an IDE

a software application with tools to improve your programming experience

R

Credit: http://www.rollcentre.com/project--onuk-sazan-


RStudio
Photo by Roberto Nickson (@g) on Unsplash

RStudio

The dashboard (and appealing interior)

Many excellent features to help you with your analyses.

Never have to think about R and RStudio as separate:
Opening R-Studio opens an R session.

Analogy from the book Modern Dive: www.moderndive.com

Open RStudio

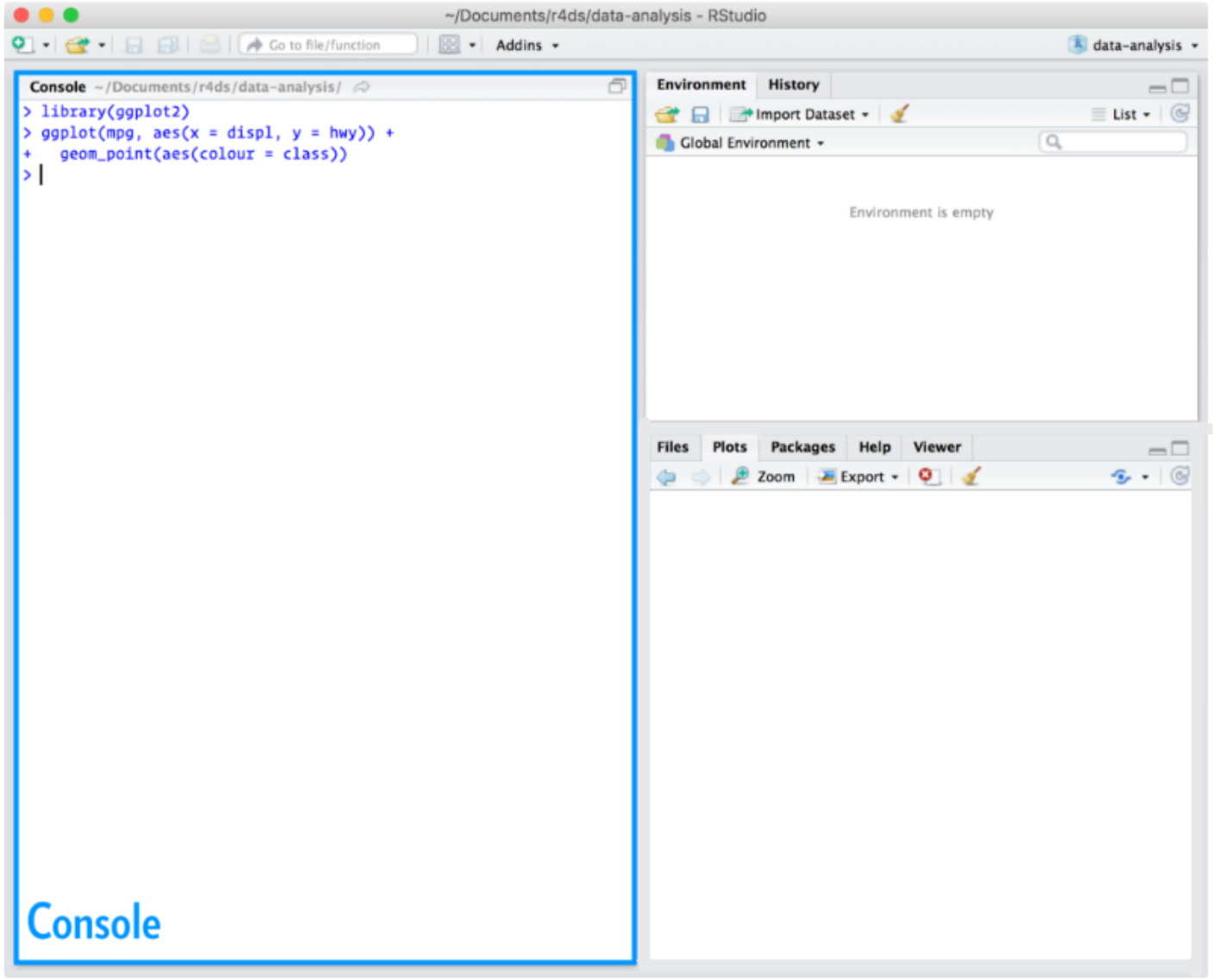
Never have to think about R and RStudio as separate:
Opening R-Studio opens an R session.

The **Console** is your window to R.

You can code directly in the console...

```
pi*2     [Enter]
37/12 [Enter]
```

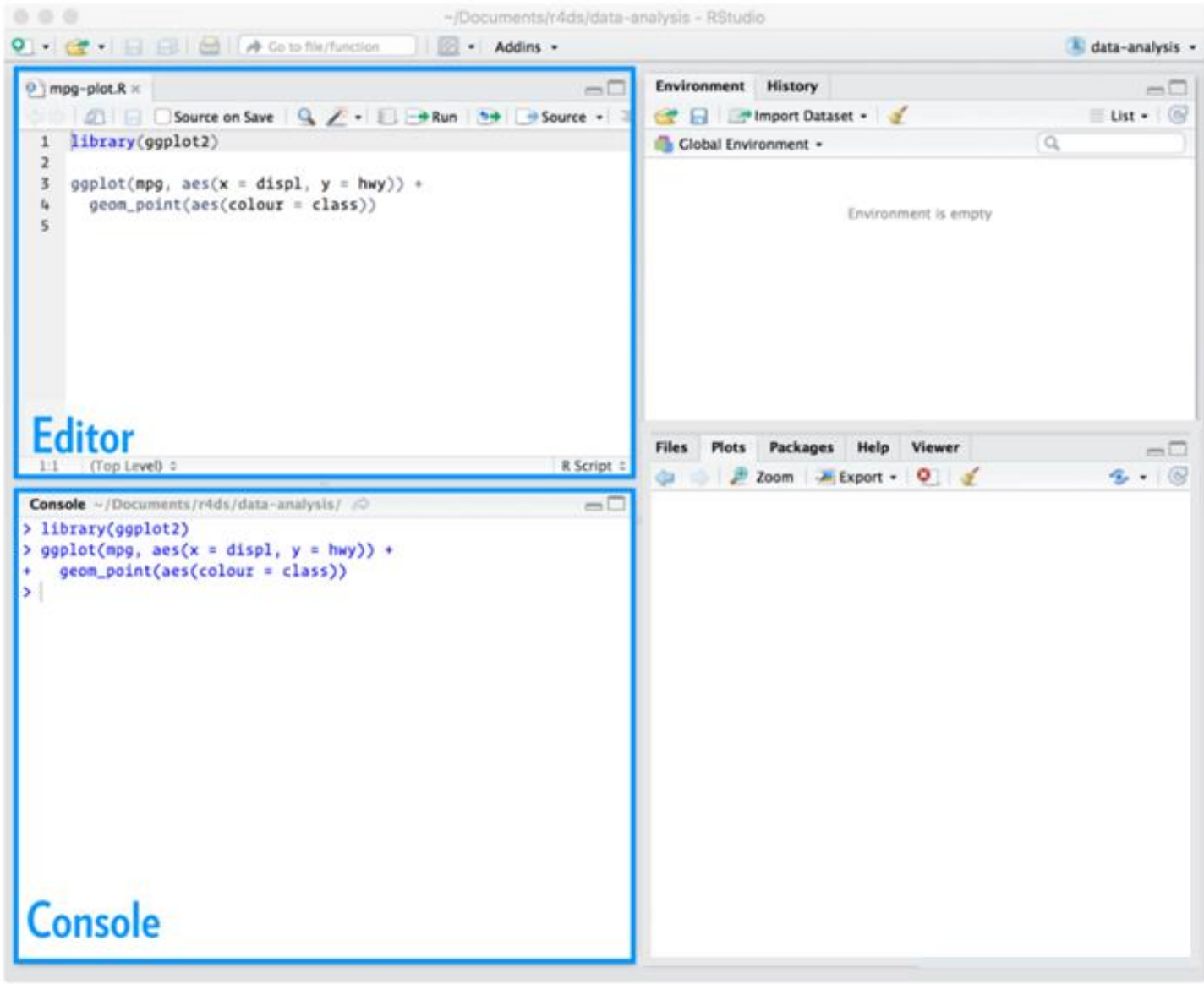
... but there is a better way...



The Editor.

If you don't see the **Editor** pane, go to the toolbar and click:

View →
Panes →
Show All Panes



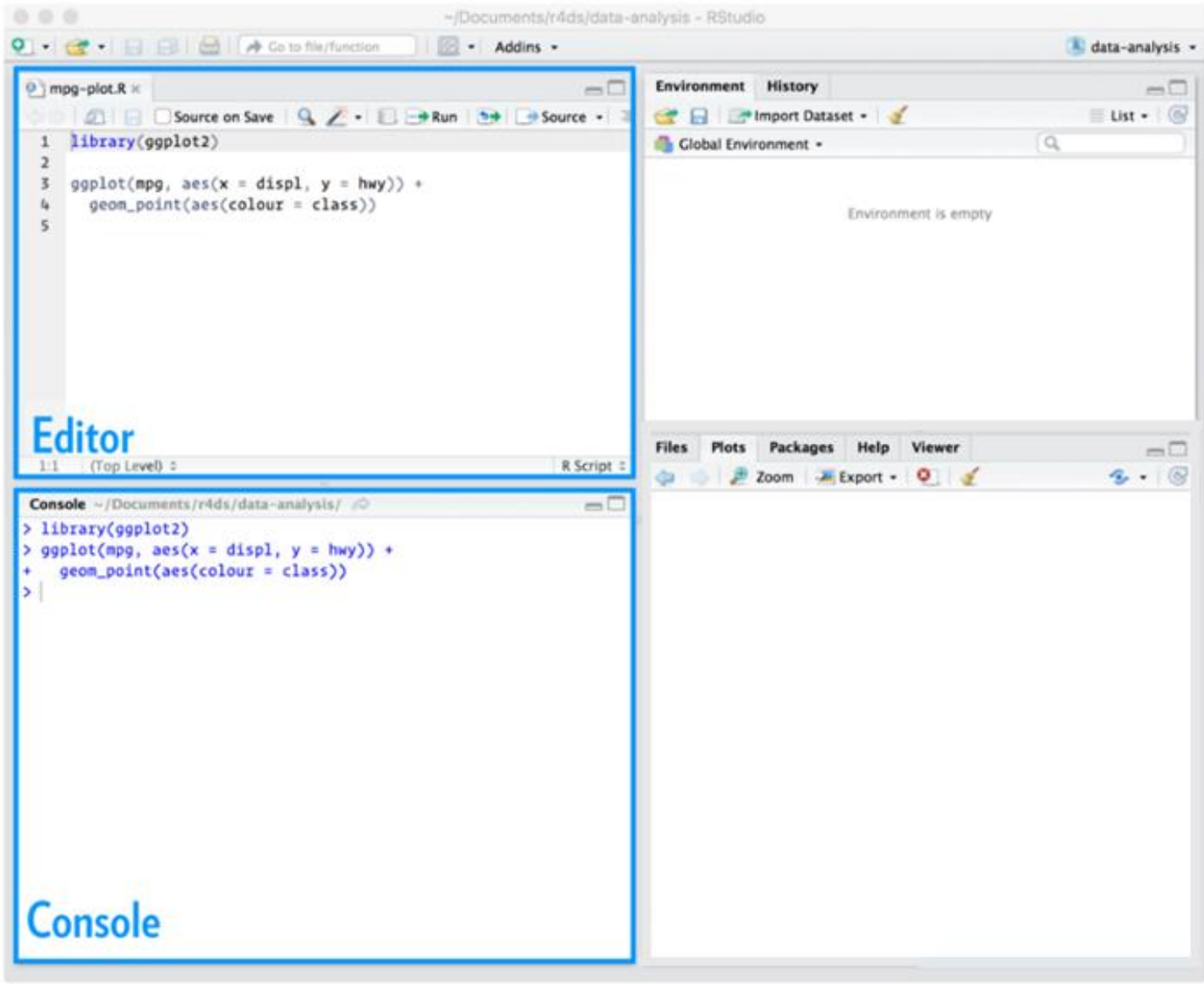
The **Editor** is just like any other text editor (you can copy, paste, and save the text)

More forgiving

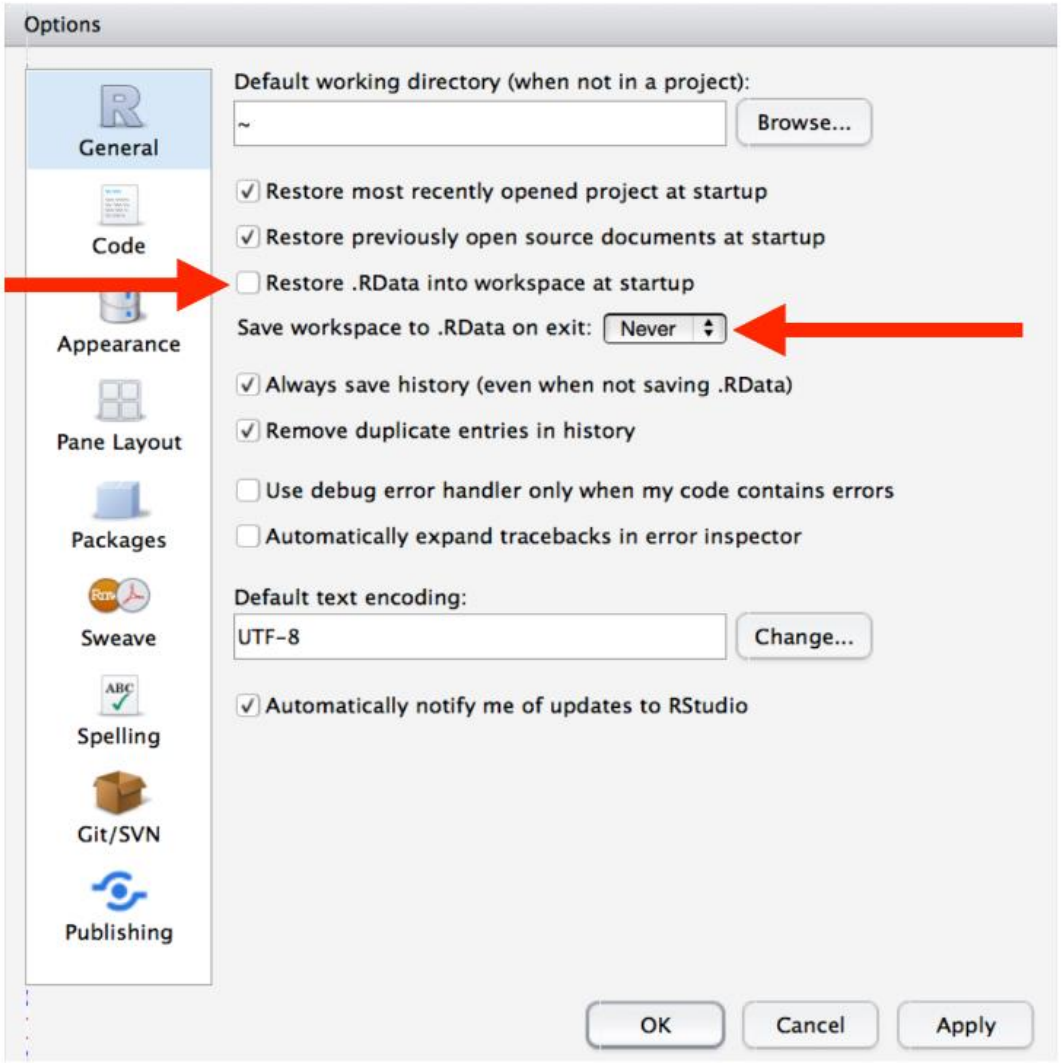
R Syntax highlighting

Autocomplete

Ctrl + Enter (sends line of code to Console)



Tools → Global Options



Credit: R for Data Science. See Chapter 6 for discussion of why this is important

Packages

Packages

R packages are like apps for your phone:

"base R"

Extend the capabilities of out of the box R

with extra functions, datasets, documentation

Packages

Download App

Open the App

Happens just once

Every new session

Download (`install.packages`)

"Open" `library()`

Phone detail from photo by Rahul Chakraborty on Unsplash

Packages

```
install.packages("gapminder")
```

will download a package to your personal library. Then:

```
library(gapminder)
```

tells R to load the gapminder package from your personal library. (Needed once every session)

CRAN repository

12,000+ packages. Free. Peer reviewed.

(Manifold possibilities) eg. machine learning, mine twitter data, create PowerPoint docs, maps.

Other ways to get packages (eg. GitHub, …)

Packages:

The tidyverse

What is the tidyverse?

The tidyverse package collects (some of) the most popular R packages into one.

All developed with the same underlying principles:

Simple tools (with consistent structure) to solve complex problems

What is the tidyverse?

During the workshop we will use the ggplot2, dplyr, and readr packages. These are all bundled up in the tidyverse package. We load it by running:

```
library(tidyverse)
```

This work is licensed as

Creative Commons

Attribution-ShareAlike 4.0

International

To view a copy of this license, visit

https://creativecommons.org/licenses/by-sa/4.0/

End