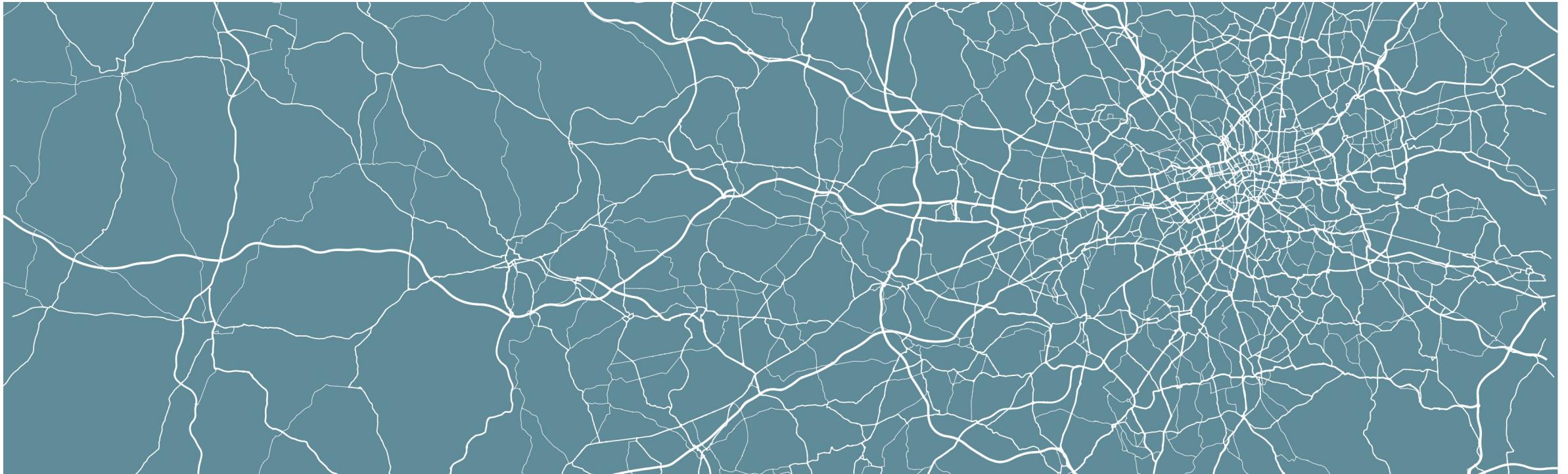


Geocomputation

W4 – Programming for Data Analysis



Where are we at?

Part I: Foundational Concepts

W1 Geocomputation: An Introduction

W2 GIScience and GIS Software

W3 Cartography and Visualisation



QGIS

W4 **Programming for Data Analysis**

W5 Programming for Spatial Analysis



R

This week

- Short recap
- What is a programming language?
- Principles of R

Before we start

- Go to www.menti.com
- Use code: 9211 2788



Recap

- Vector versus raster
- Joining attributes to spatial (vector layers)
- Data classification, idea of MAUP
- Geographic Coordinate Systems and Projected Coordinate Systems



8	9	9	10	0	10	10	10	0	0	0	7	5	3	0	0	0	0	1
8	9	9	10	10	0	10	9	9	0	0	5	3	0	0	0	0	0	0
8	8	9	9	10	0	0	9	8	7	5	0	0	0	1	0	0	0	0
5	8	8	9	10	10	0	9	7	5	0	0	5	5	5	0	0	0	1
3	5	8	9	9	10	0	0	3	0	0	0	5	0	0	1	0	0	2
2	5	8	8	9	9	10	0	0	0	1	5	0	0	0	0	0	0	1
2	4	6	8	8	9	0	0	0	1	5	0	0	5	5	5	0	0	1
0	3	6	8	8	0	0	0	0	5	0	5	5	5	5	5	0	0	0
2	2	5	8	0	0	0	0	0	0	5	5	5	5	5	5	3	0	0
0	2	5	0	0	1	2	3	4	4	4	4	4	4	4	5	0	0	0
0	0	0	0	1	1	1	1	4	4	4	4	4	4	4	5	0	0	0
0	0	1	1	2	2	2	2	3	3	3	3	3	3	3	4	0	3	0
1	1	1	1	2	2	3	3	3	3	1	1	1	1	1	2	3	4	3

Attribute join

Table 1



1		
2		

Table 2



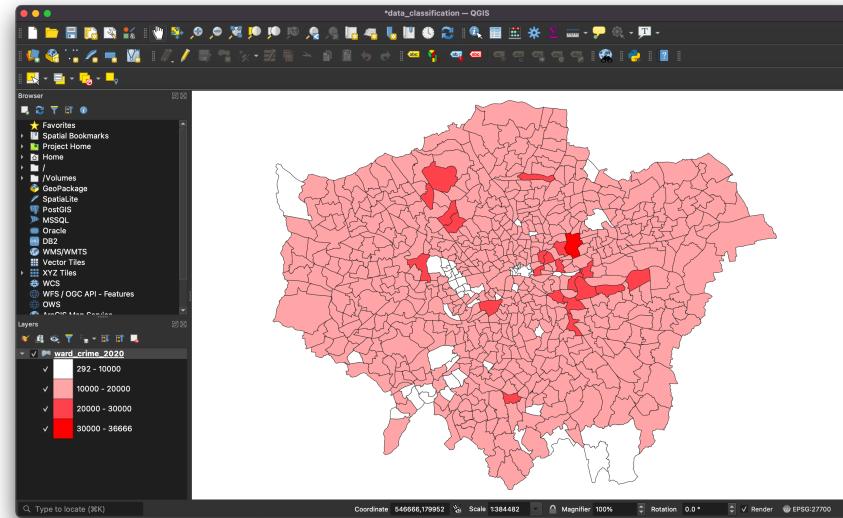
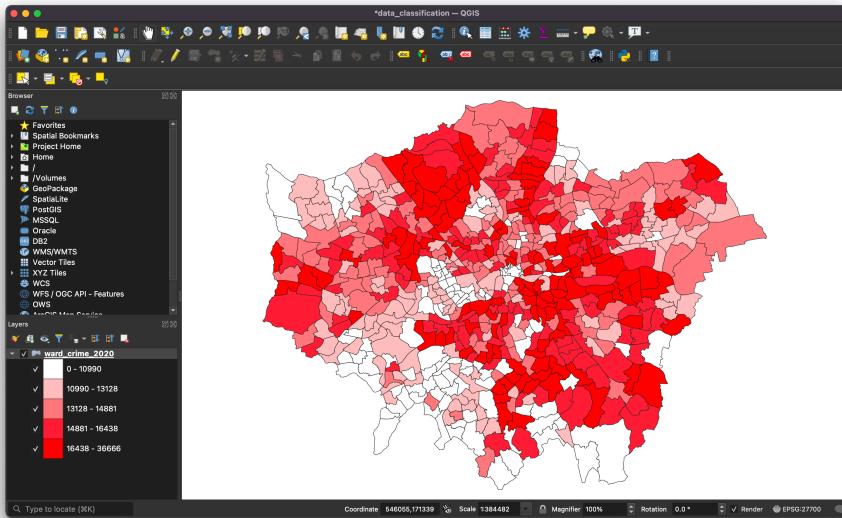
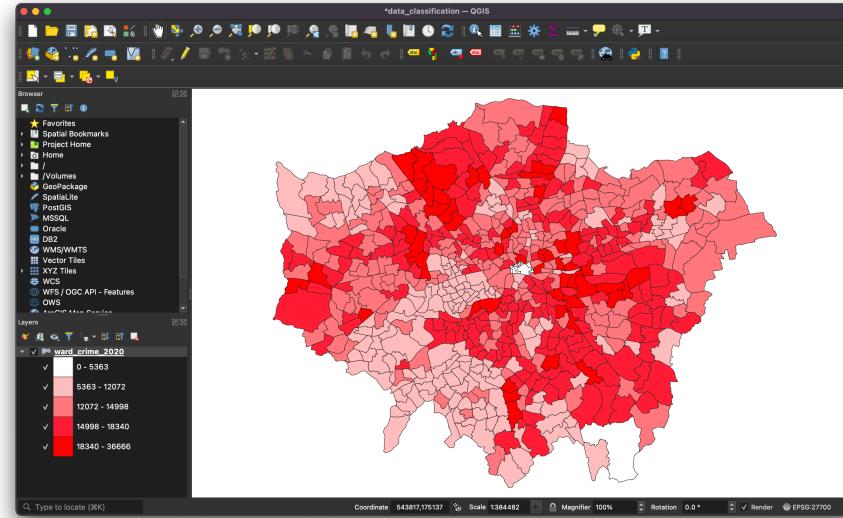
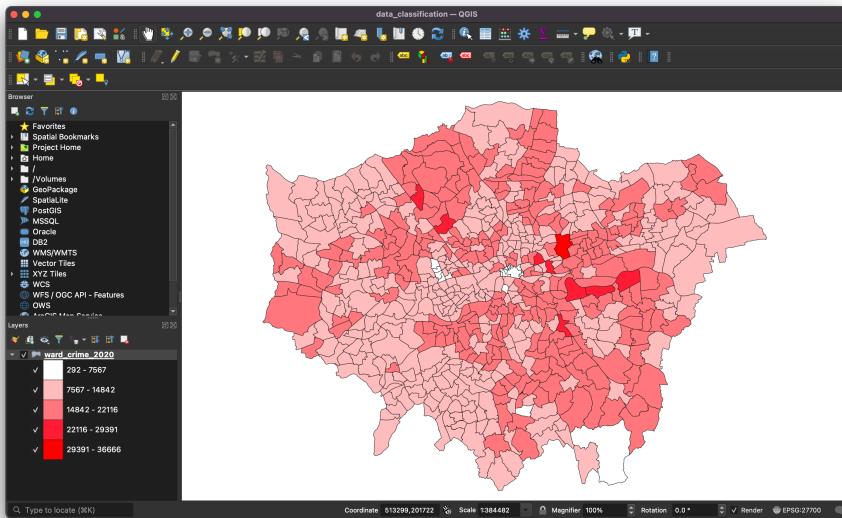
1		
3		
4		

Left Join

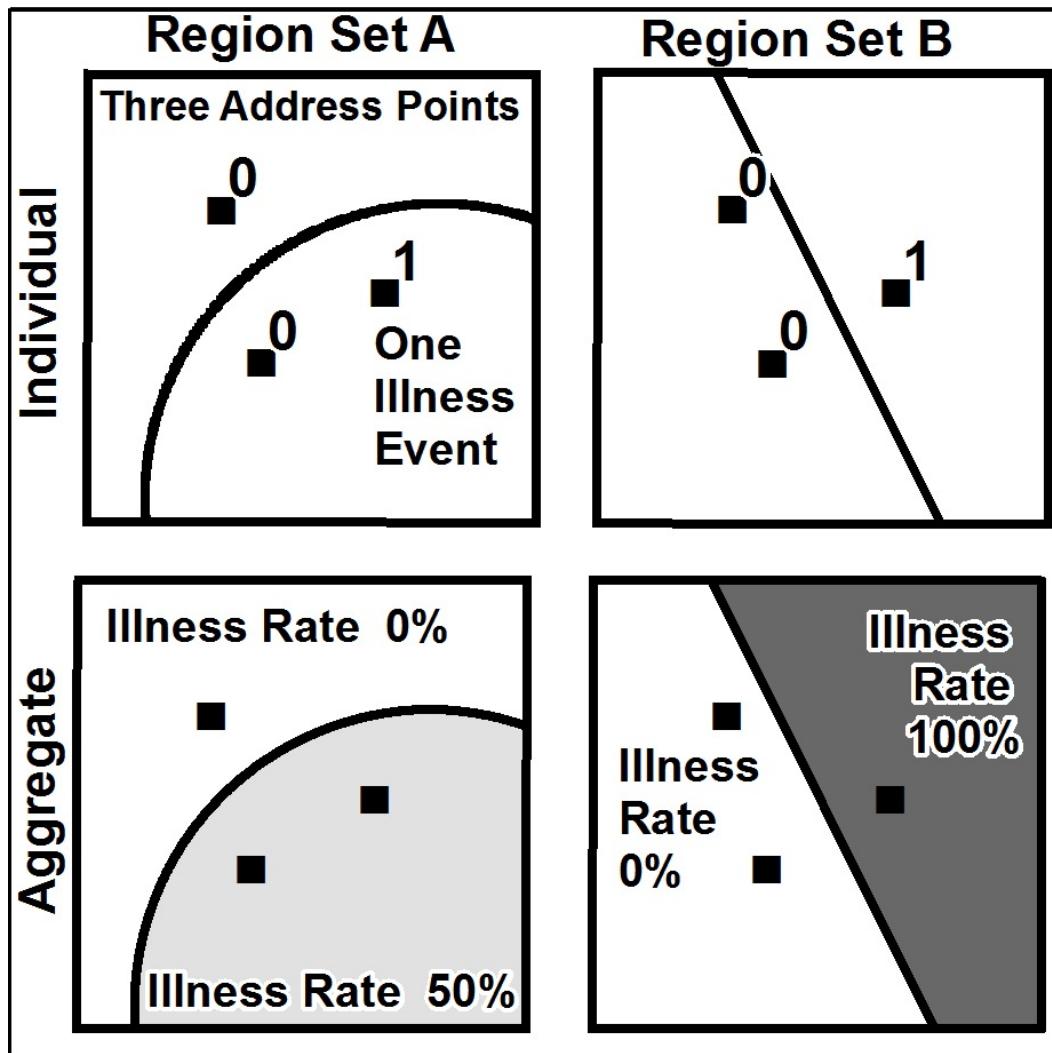


1			
2			

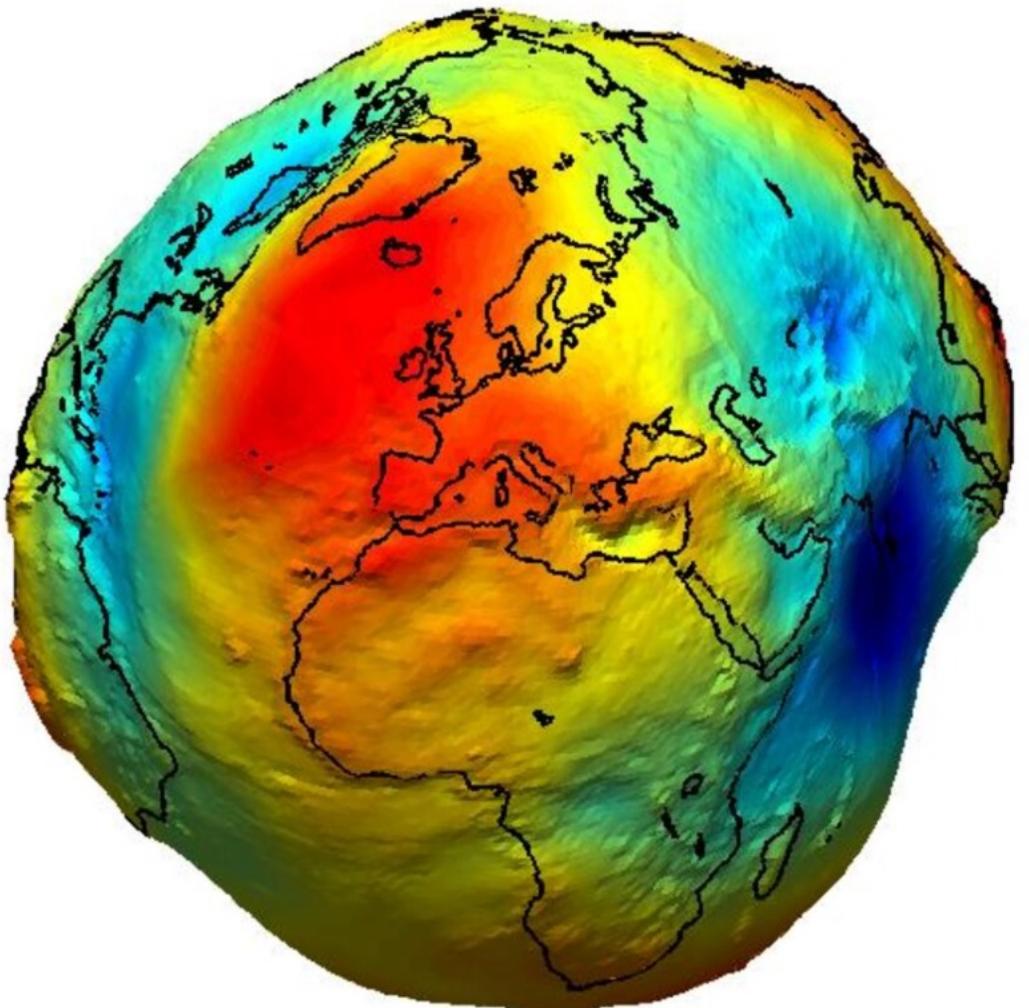
Classifying data



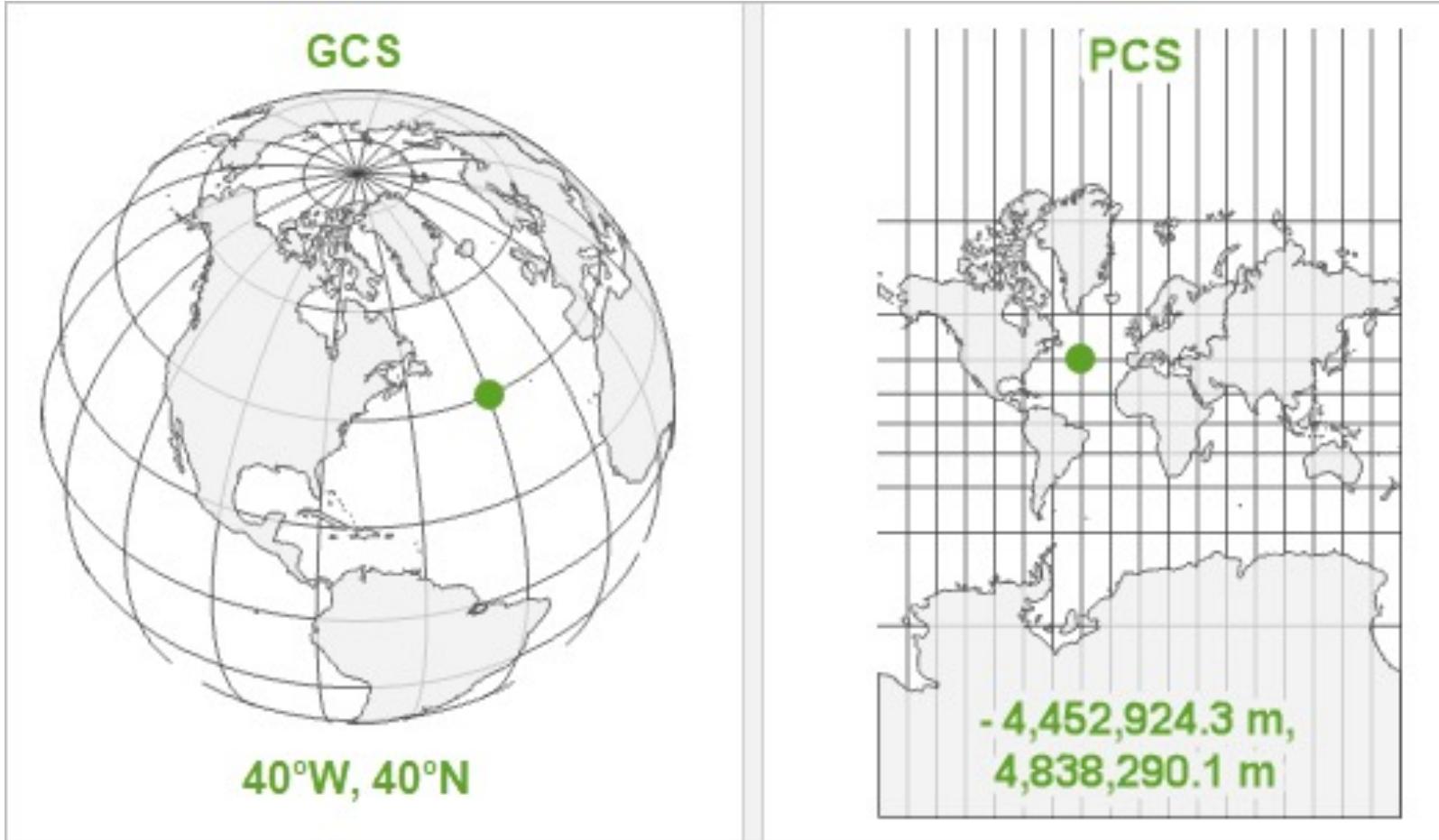
Modifiable Areal Unit Problem



Representing the globe



Moving from an ellipsoid to a plane



QGIS

- Powerful open-source GIS, widely used both within and outside of academia.
- Extensive documentation: [\[Link\]](#)
- However (1): QGIS can be a bit fiddly at times. *Aligning multiple maps on a canvas?*
- However (2): Graphical User Interface involves lots of manual actions, time-involved to repeat.

Programming languages

"Everyone does need to learn to code. It is no longer sufficient for a GI Scientists to just work with a standard GIS interface: menus, buttons and black boxes."

Brunsdon and Comber 2020

Programming languages

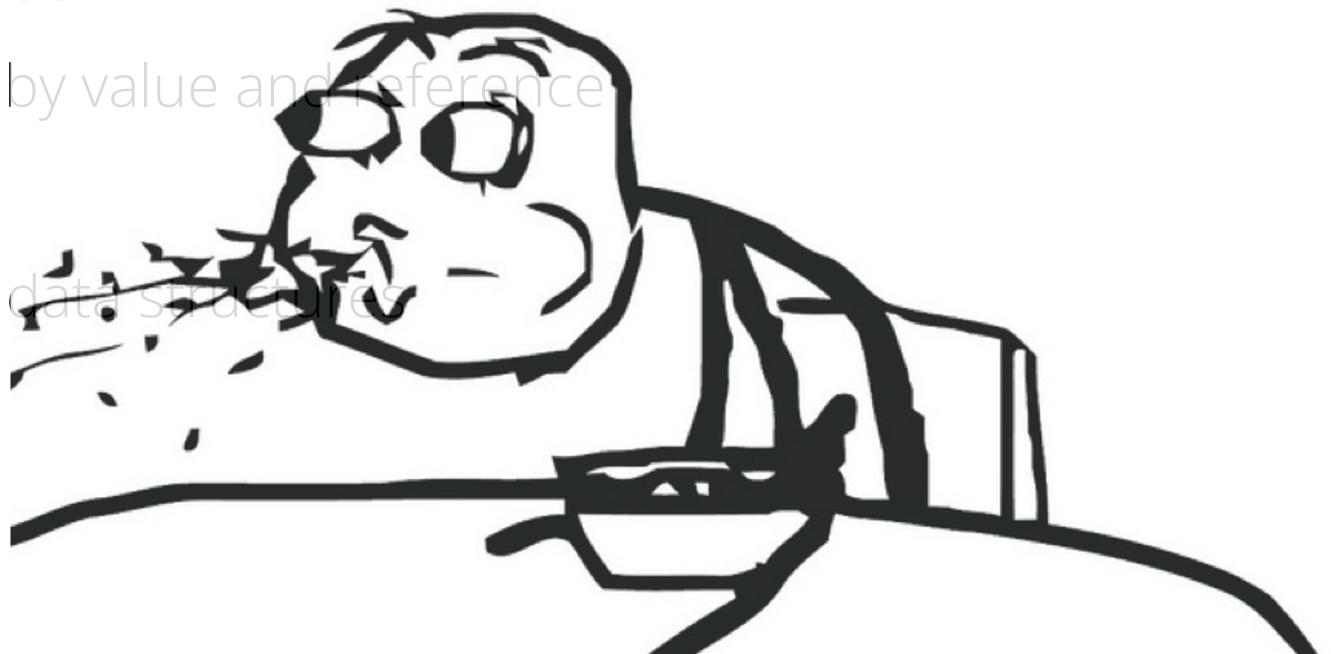
- From "point-and-click" to "writing commands".
- In our case: writing commands to execute various data management and analysis tasks.
- We will be using a programming language called R.

Programming languages

- Identifiers and primitive data types
- Assignment, arithmetic, logical and relational operators
- Expression and statements, debugging
- Flow of control: selection and repetition
- Functions, parameters passing, call by value and reference
- Object-oriented programming
- 1/2 dimensional arrays, strings and data structures

Programming languages

- Identifiers and primitive data types
- Assignment, arithmetic, logical and relational operators
- Expression and statements, debugging
- Flow of control: selection and repetition
- Functions, parameters passing, call by value and reference
- Object-oriented programming
- 1/2 dimensional arrays, strings and data structures



Programming languages

- Ιδεντιφιερς ανδ πριμιτιφε δατα τψπες
- Ασσιγνμεντ, αριτημετιξ, λογιξαλ ανδ ρελατιοναλ οπερατορς
- Εχπρεσσιον ανδ στατεμεντς, δεβυγγινγ
- Φλωρ οφ ξοντρολ· σελεξτιον ανδ ρεπετιτιον
- Φυνξτιονς, παραμετερς πασσινγ, ξαλλ βψ αφλυε ανδ ρεφερενξε
- Οβσεξτ-οριεντεδ προγραμμινγ
- 1/2 διμενσιοναλ αρραψ, στρινγς ανδ δατα στρυξτυρες

Principles of R

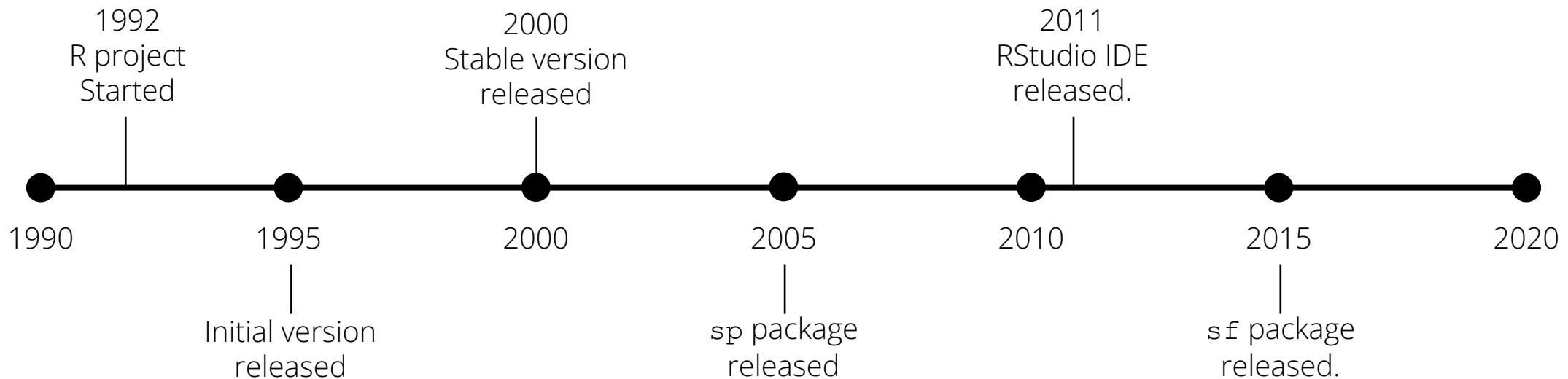
- What is R?
- R is used primarily through interactive command-line.
- R can create and use different types of data but works well with a linear collection of things (vector) and tables.
- R is extremely extendable through packages.
- R is brilliant when it comes to static graphics and dealing with spatial data.

Principles of R

- What is R?
- R is used primarily through interactive command-line.
- R can create and use different types of data but works well with a linear collection of things (vector) and tables.
- R is extremely extendable through packages.
- R is brilliant when it comes to static graphics and dealing with spatial data.

A little history

- R is programming language which specialises in statistics.
- 1975: Bell Labs develops a language for Statistical Analysis
- 1992: Ross Ihaka and Robert Gentleman develop opensource version of S



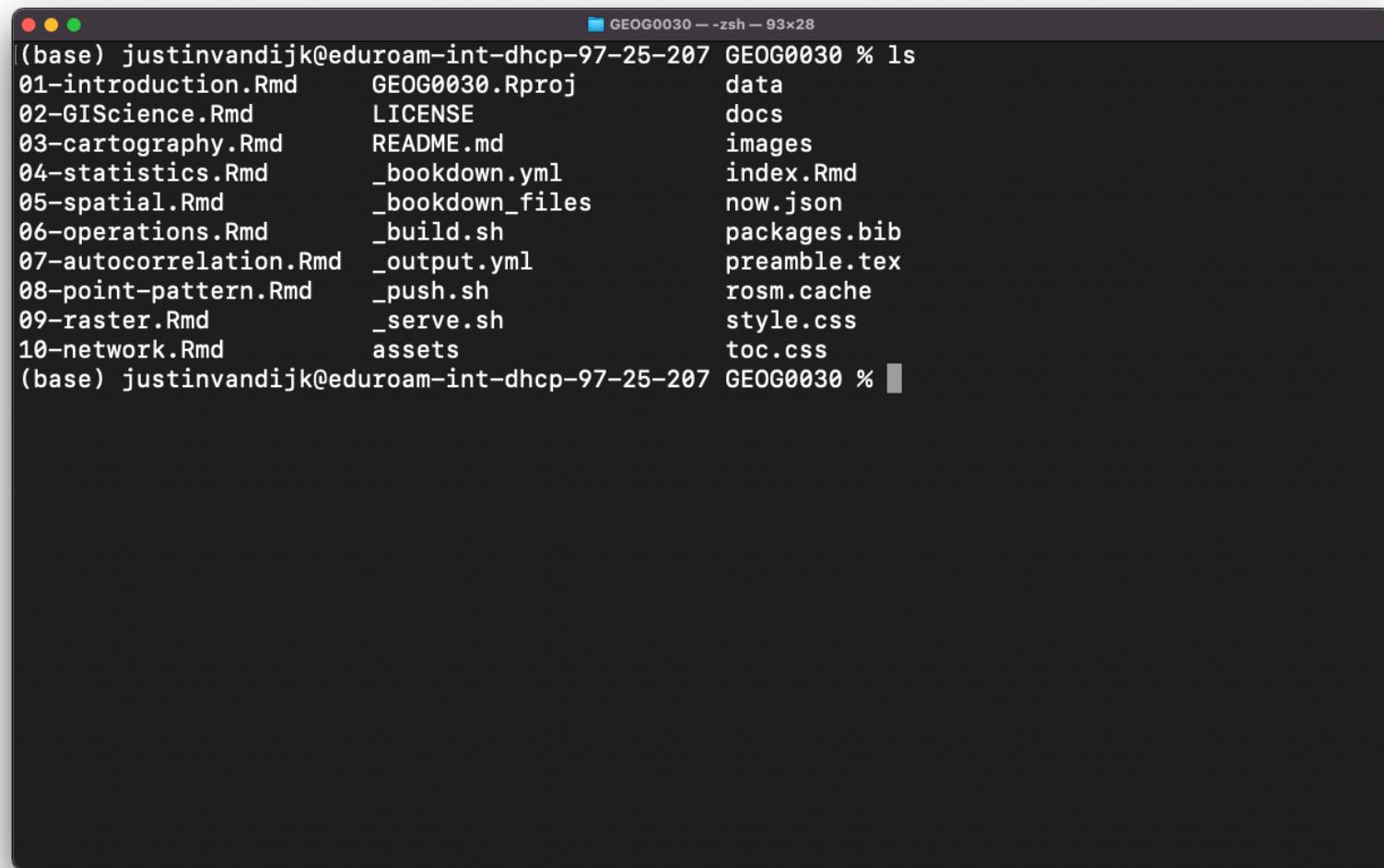
Principles of R

- What is R?
- R is used primarily through interactive command-line.
- R can create and use different types of data but works well with a linear collection of things (vector) and tables
- R is extremely extendable through packages.
- R is brilliant when it comes to static graphics and dealing with spatial data.

Command-Line Interface

- Command-Line Interface versus (CLI) Graphical User Interface (GUI)
- The GUI allows the user to interact with the system using graphical elements such as windows, icons, menus while the CLI allows the user to interact with the system using commands.

Command-Line Interface



A screenshot of a terminal window titled "GEOG0030 -- zsh -- 93x28". The window shows the output of the "ls" command, listing files and directories in a directory structure. The files are organized into several groups:

- Group 1: 01-introduction.Rmd, 02-GIScience.Rmd, 03-cartography.Rmd, 04-statistics.Rmd, 05-spatial.Rmd, 06-operations.Rmd, 07-autocorrelation.Rmd, 08-point-pattern.Rmd, 09-raster.Rmd, 10-network.Rmd.
- Group 2: GEOG0030.Rproj, LICENSE, README.md, _bookdown.yml, _bookdown_files, _build.sh, _output.yml, _push.sh, _serve.sh, assets.
- Group 3: data, docs, images, index.Rmd, now.json, packages.bib, preamble.tex, rosm.cache, style.css, toc.css.

The terminal prompt "(base)" appears twice at the bottom of the list, indicating the current working directory.

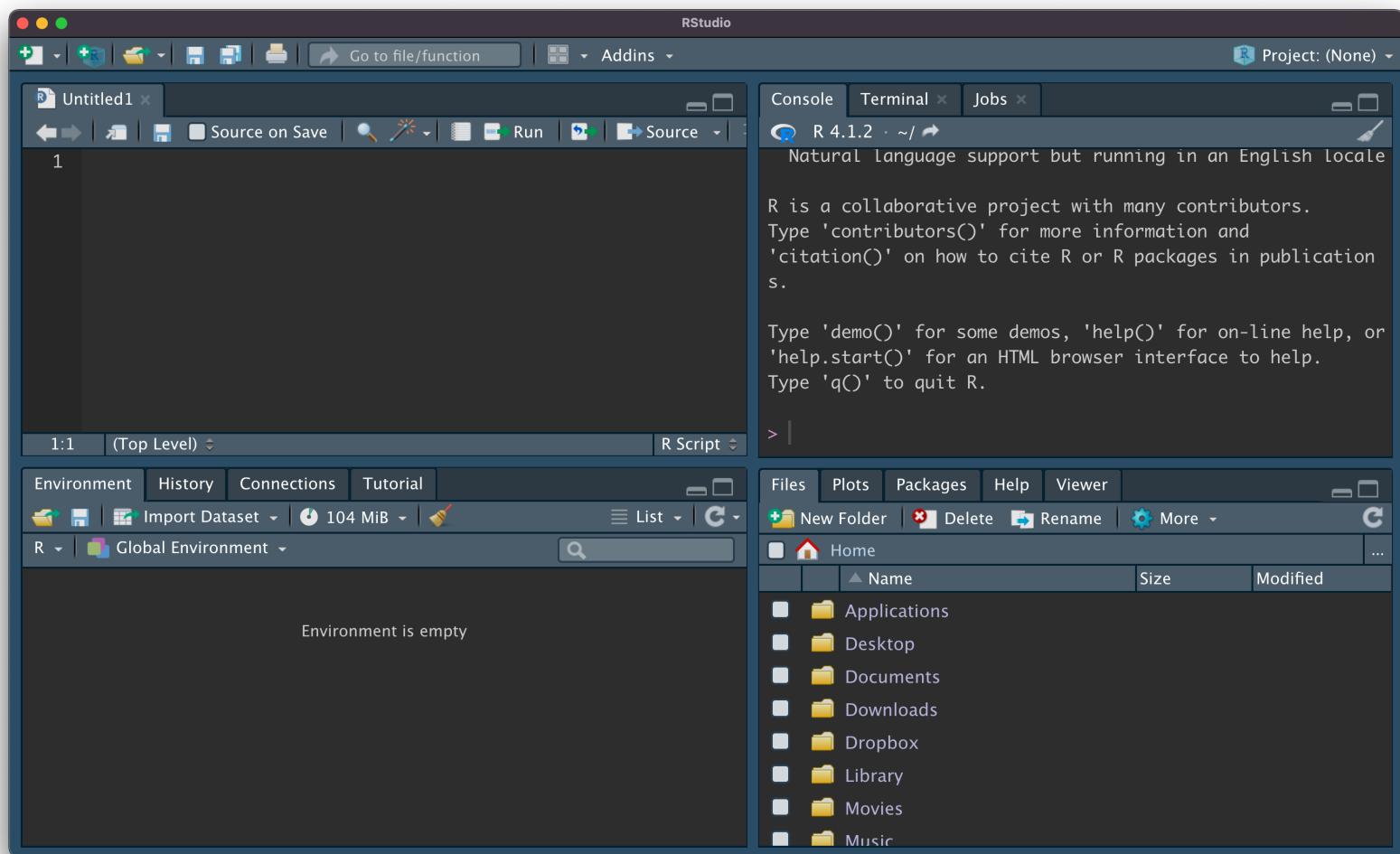
Graphical User Interface



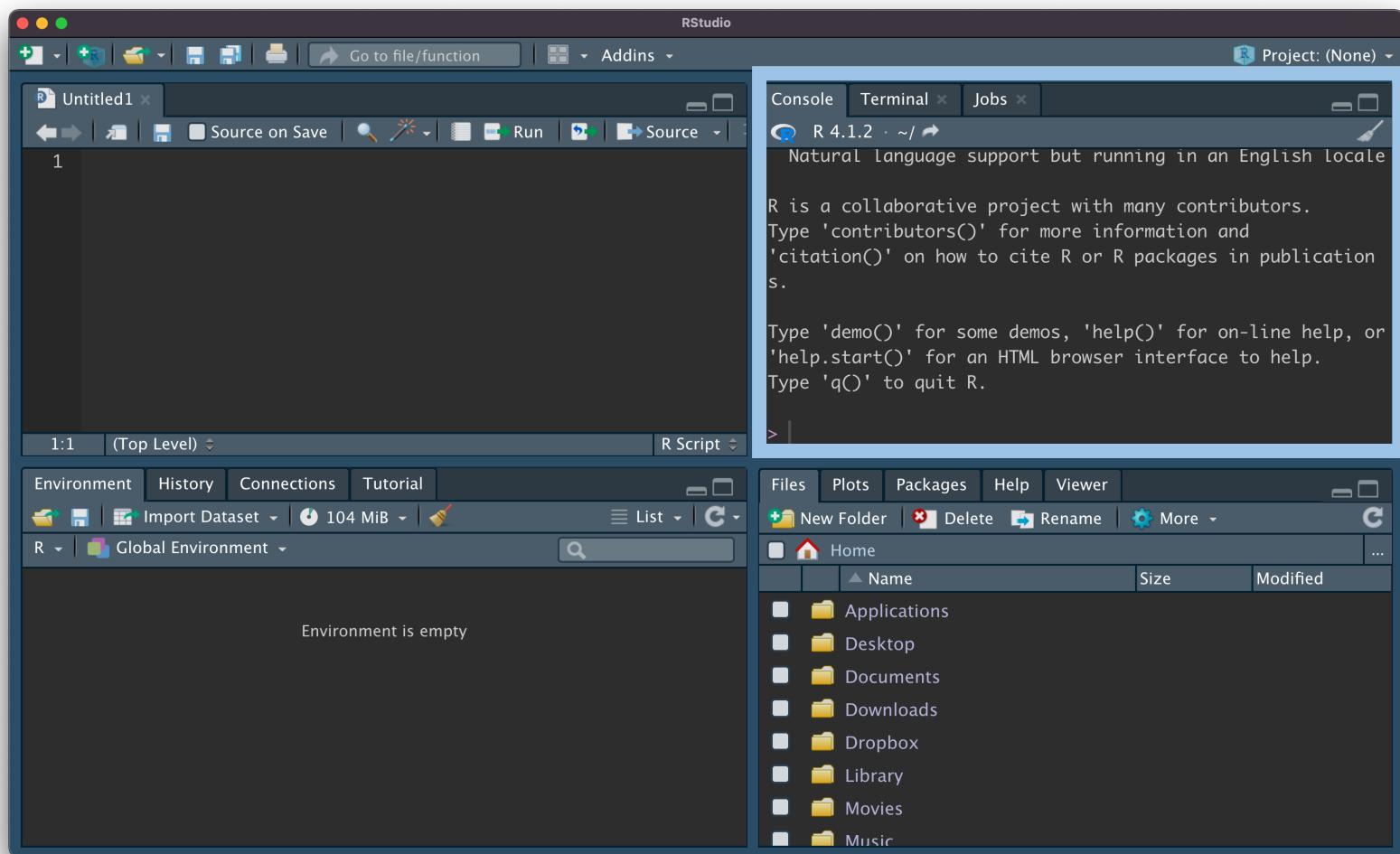
Interaction through language

- Capturing complex instruction with language is much easier than with skeuomorphism.
- Repeating stuff is easy.
- Much steeper learning curve but greater rewards as well.

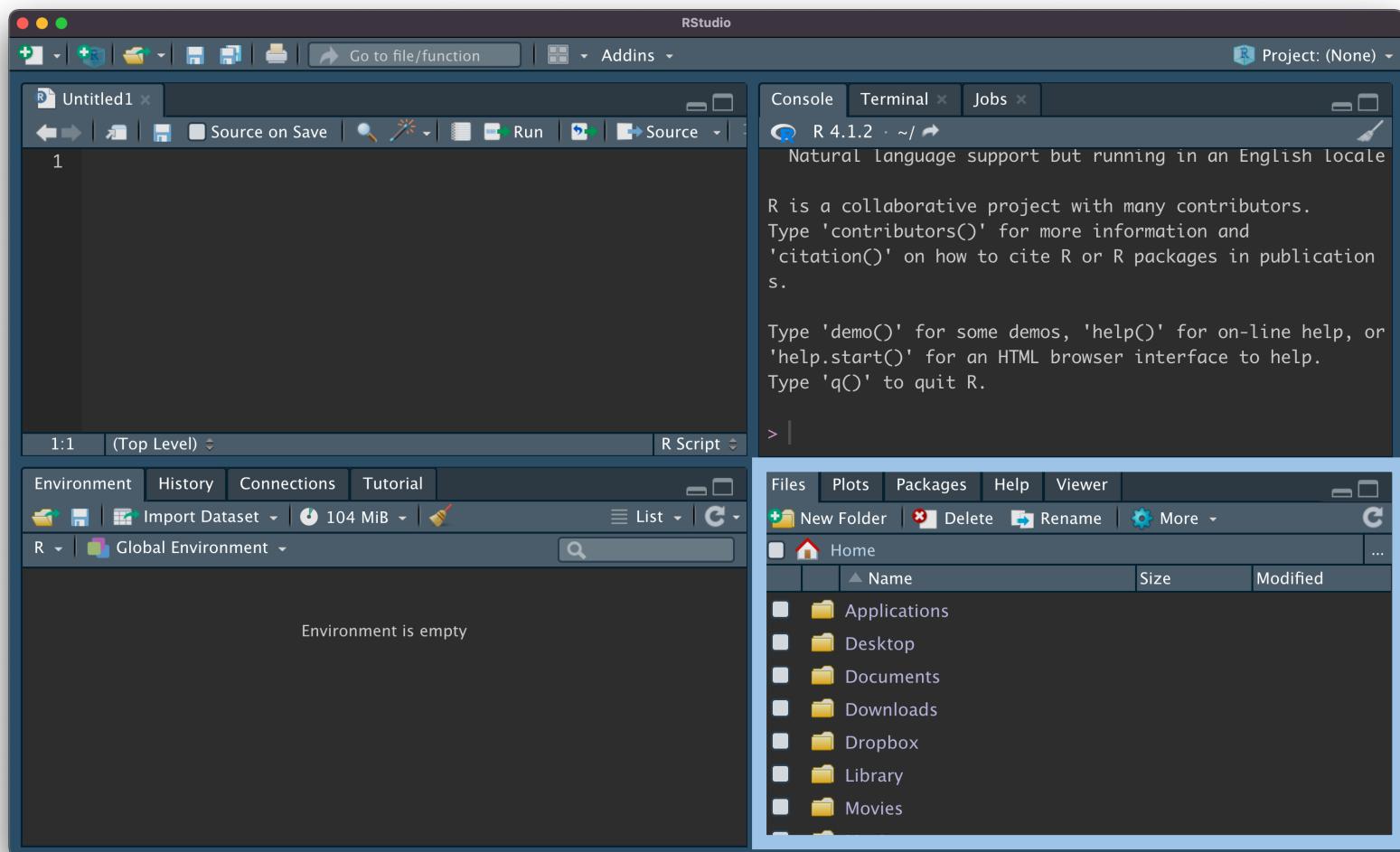
How does this work in R



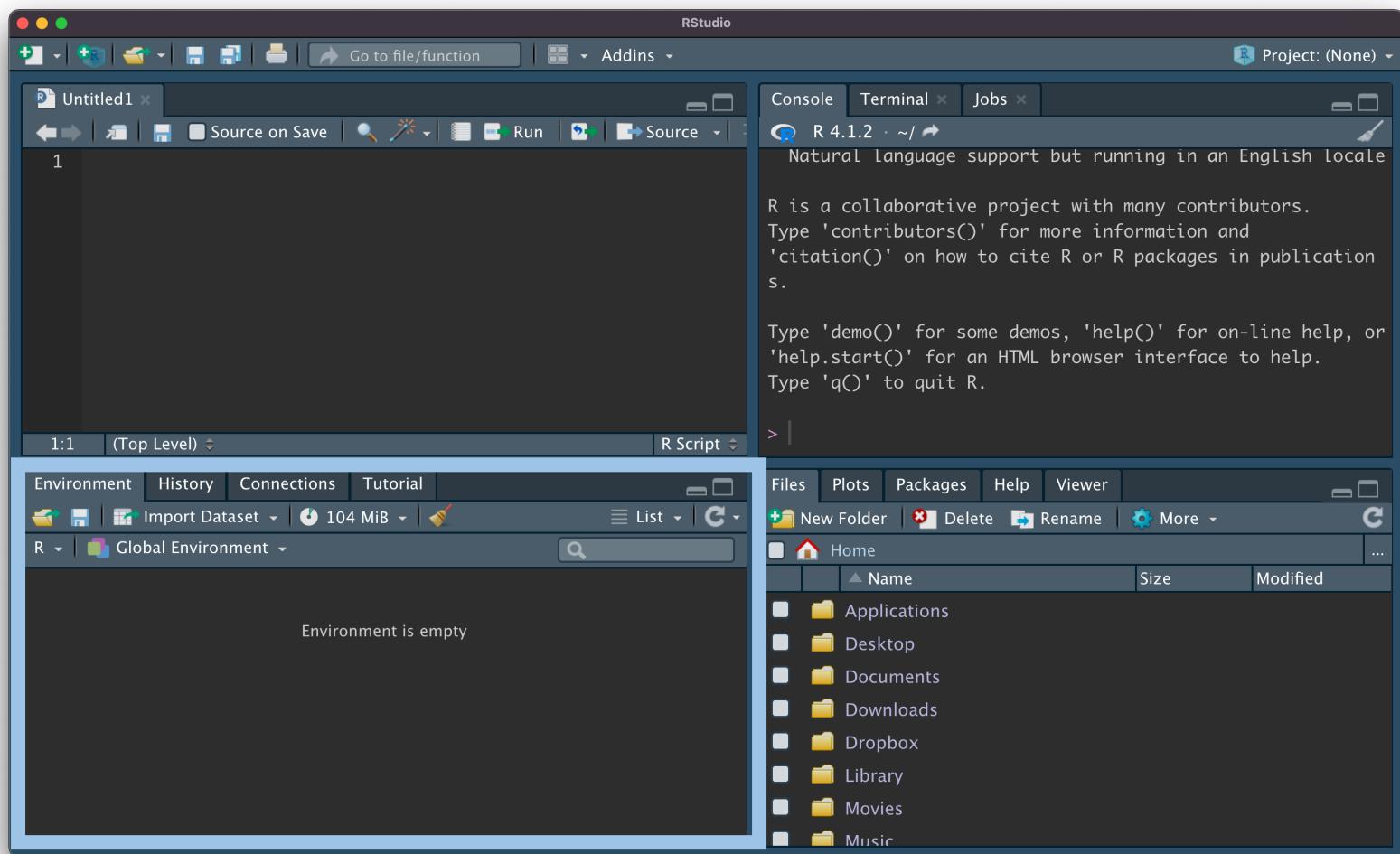
How does this work in R



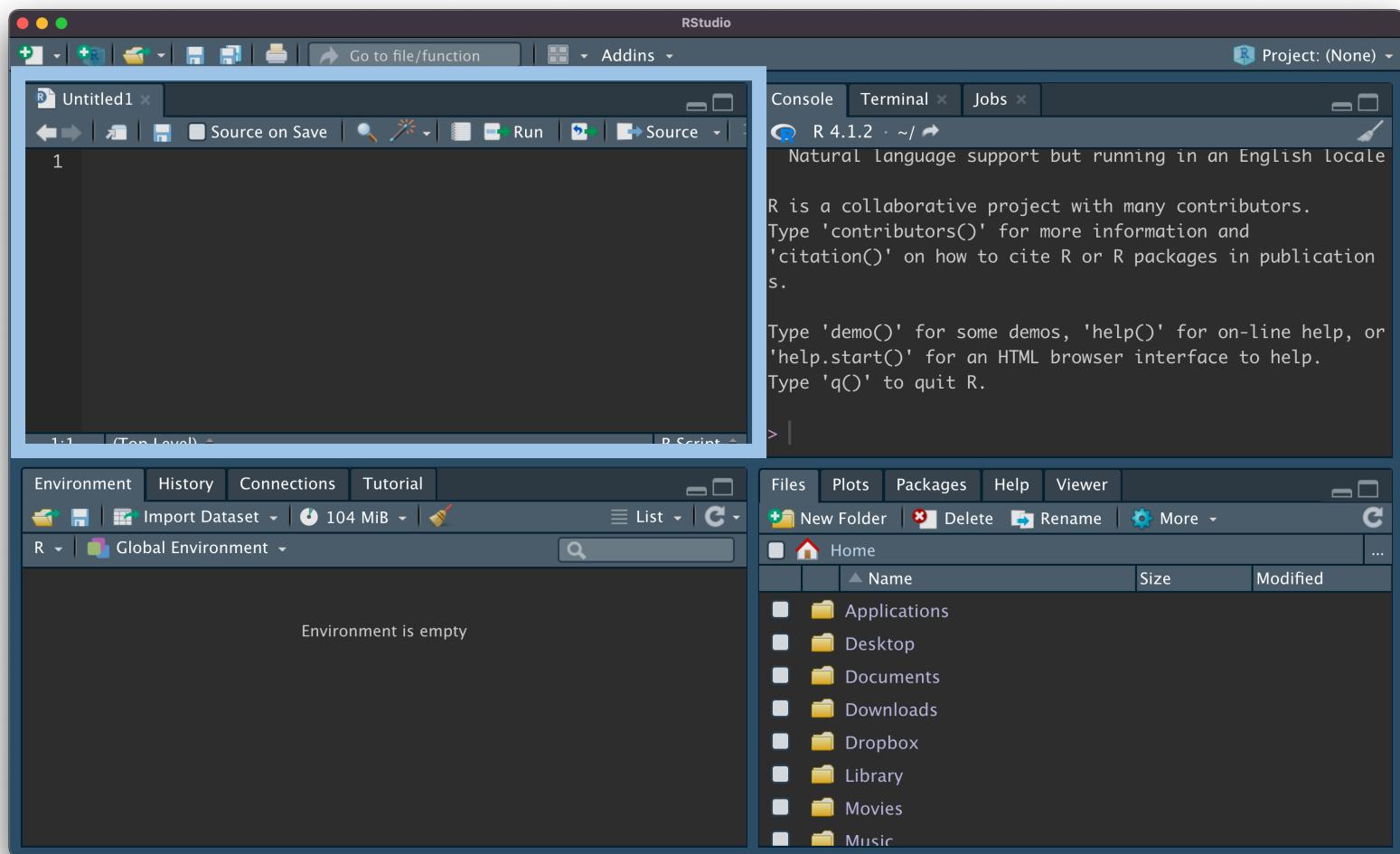
How does this work in R



How does this work in R



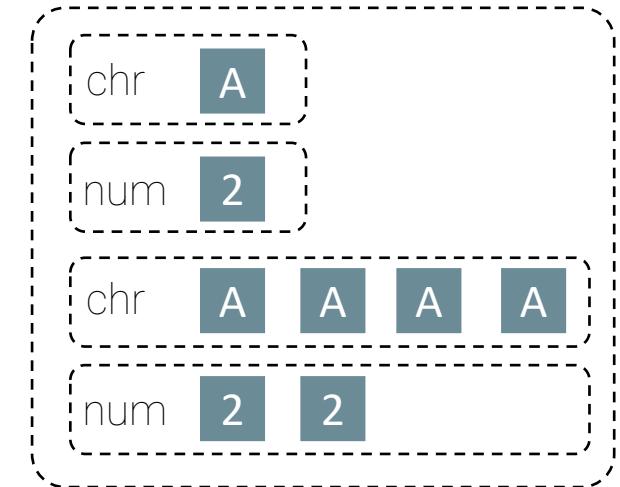
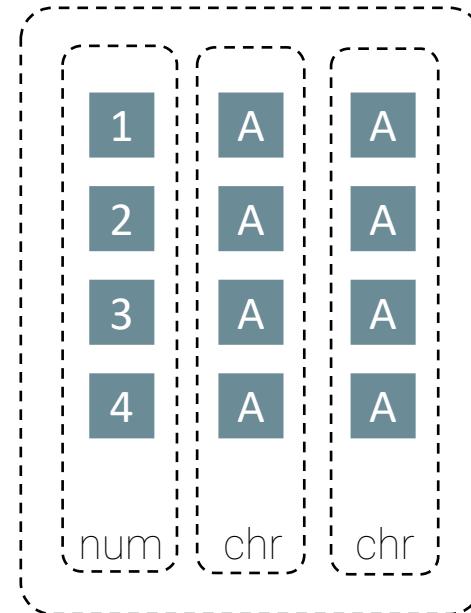
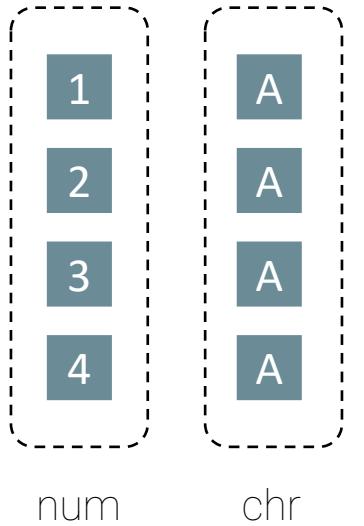
How does this work in R



Principles of R

- What is R?
- R is used primarily through interactive command-line.
- R can create and use different types of data but works well with a linear collection of things (vector) and tables.
- R is extremely extendable through packages.
- R is brilliant when it comes to static graphics and dealing with spatial data.

Data types



Scalar

Vector

Dataframe

List

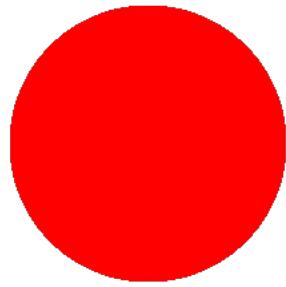
Variables

- All scalars, vectors, tables, and lists can be assigned to a variable.
- Variables are used to store information to be referenced and manipulated in a computer programme.

Functions

- Variables can be used as an input for functions.
- Functions are pieces of code that accomplish a specific task.
- Functions usually "take in" data, process it, and "return" a result.
- Once a function is written, it can be used over and over and over again.

RStudio

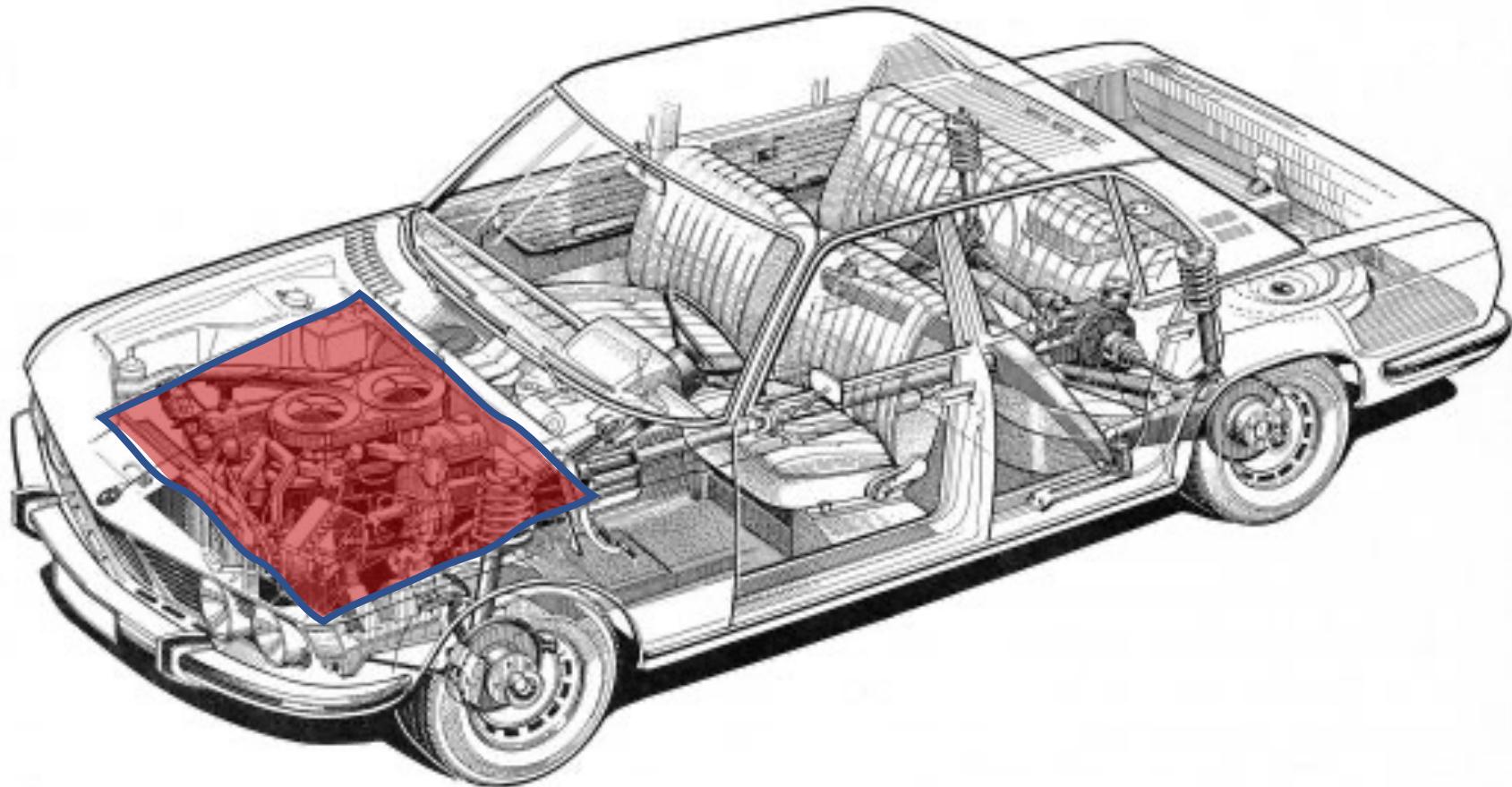


LIVE

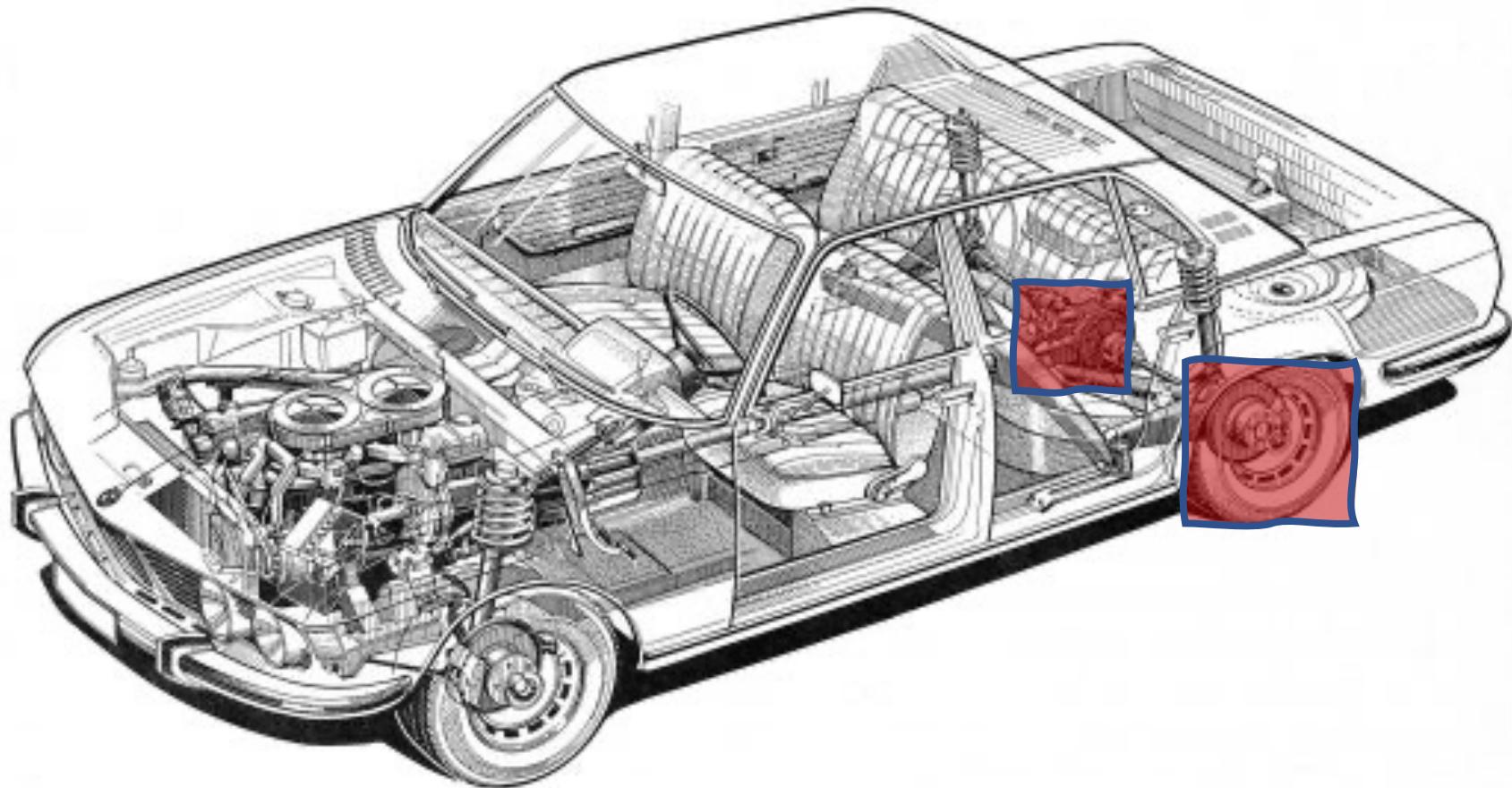
Principles of R

- What is R?
- R is used primarily through interactive command-line.
- R can create and use different types of data but works well with a linear collection of things (vector) and tables.
- R is extremely extendable through packages.
- R is brilliant when it comes to static graphics and dealing with spatial data.

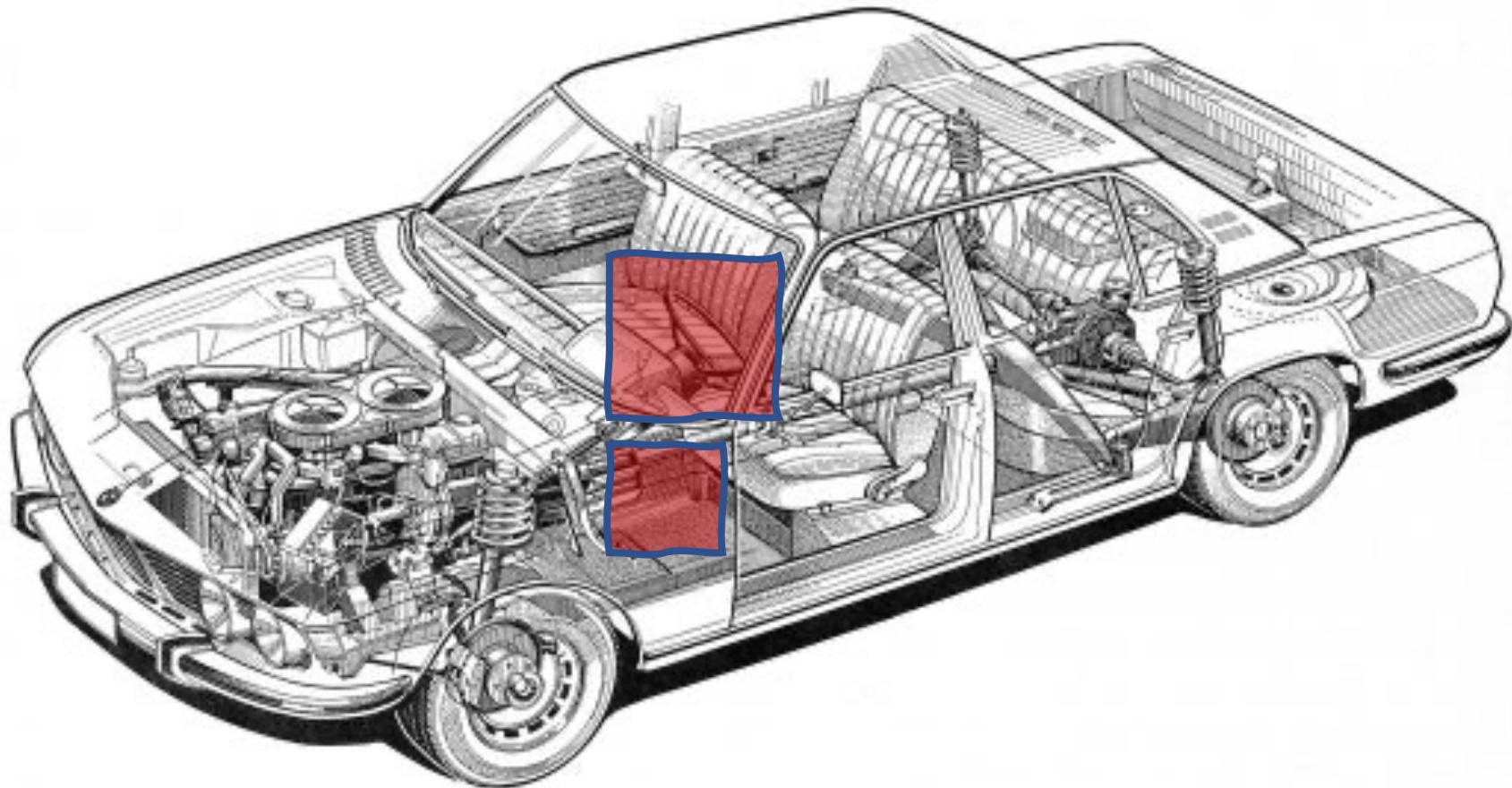
Packages



Packages



Packages



Packages

- Available on The Comprehensive R Archive Network ([CRAN](#)).
- Currently, the CRAN package repository features 18820 available packages (January 27).
- Packages provide extensions to R.

chorrrds

The screenshot shows the GitHub README page for the `chorrrds` package. The page has a dark theme. At the top, it displays the package name, version 0.1.9.5, 26K downloads, and an RDoc link. Below this, a large heading reads "chorrrds : A package for music chorrrds extraction." A paragraph describes the package as a tool for scraping chords from Cifraclub. It includes a "Donate" button and a note about maintaining the package. The "Installation" section provides instructions for CRAN installation and links to GitHub for the latest version. On the right side of the page, there is a "Languages" section showing a breakdown of the codebase: HTML (97.0%), R (2.6%), and TeX (0.4%).

Author: Bruna Wunderwald License: MIT

CRAN 0.1.9.5 downloads 26K RDoc build failing

chorrrds : A package for music chorrrds extraction.

`chorrrds` is a package for R that scrapes the [Cifraclub](#) website to download and organize music chords. It can be considered a package for MIR (Music Information Retrieval), a broad area of computational music which extracts and processes music data, from the unstructured ones, as sound waves, to structured, like sheet music or chords.

If you enjoy this work, consider [buying me a coffee in Ko-Fi](#), or [Paypal](#):

[Donate](#)

so I can keep developing and maintaining this package :)

Installation

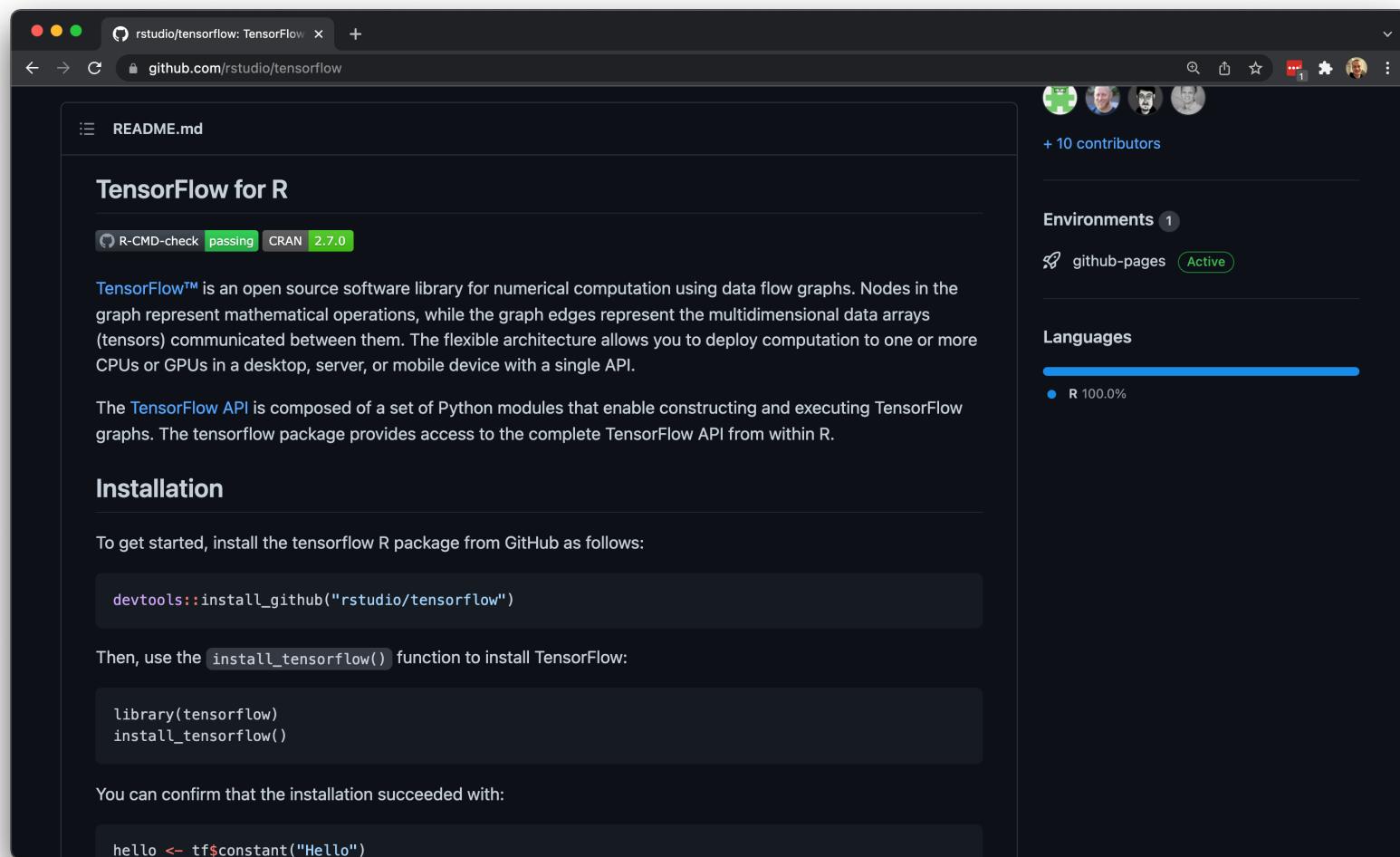
You can install `chorrrds` from your favorite CRAN mirror, simply running:

```
install.packages("chorrrds")
```

You can also install the latest version of `chorrrds` from the R-Music GitHub organization with:

```
# install.packages("devtools")
```

tensorflow



The screenshot shows the GitHub README page for the TensorFlow for R repository. The page has a dark theme. At the top, there's a navigation bar with icons for file, edit, and search, followed by a user profile icon and a three-dot menu. Below the header, the repository name "rstudio/tensorflow: TensorFlow" is displayed, along with a link to "github.com/rstudio/tensorflow". The main content area starts with a section titled "TensorFlow for R" which includes a "R-CMD-check" badge showing "passing" and a "CRAN" badge showing "2.7.0". A detailed description follows, explaining what TensorFlow is and how it can be used with R. Below this, there's a "Installation" section with instructions for installing the package from GitHub using devtools::install_github() and for installing TensorFlow using install_tensorflow(). Finally, there's a code snippet for confirming the installation with tf\$constant("Hello"). On the right side of the page, there are sections for "Contributors" (showing + 10 contributors), "Environments" (showing github-pages as active), and "Languages" (showing R at 100%).

TensorFlow™ is an open source software library for numerical computation using data flow graphs. Nodes in the graph represent mathematical operations, while the graph edges represent the multidimensional data arrays (tensors) communicated between them. The flexible architecture allows you to deploy computation to one or more CPUs or GPUs in a desktop, server, or mobile device with a single API.

The [TensorFlow API](#) is composed of a set of Python modules that enable constructing and executing TensorFlow graphs. The tensorflow package provides access to the complete TensorFlow API from within R.

Installation

To get started, install the tensorflow R package from GitHub as follows:

```
devtools::install_github("rstudio/tensorflow")
```

Then, use the `install_tensorflow()` function to install TensorFlow:

```
library(tensorflow)
install_tensorflow()
```

You can confirm that the installation succeeded with:

```
hello <- tf$constant("Hello")
```

sf

sf

Simple Features for R

A package that provides simple features access for R. Package sf:

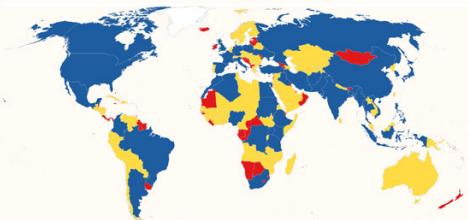
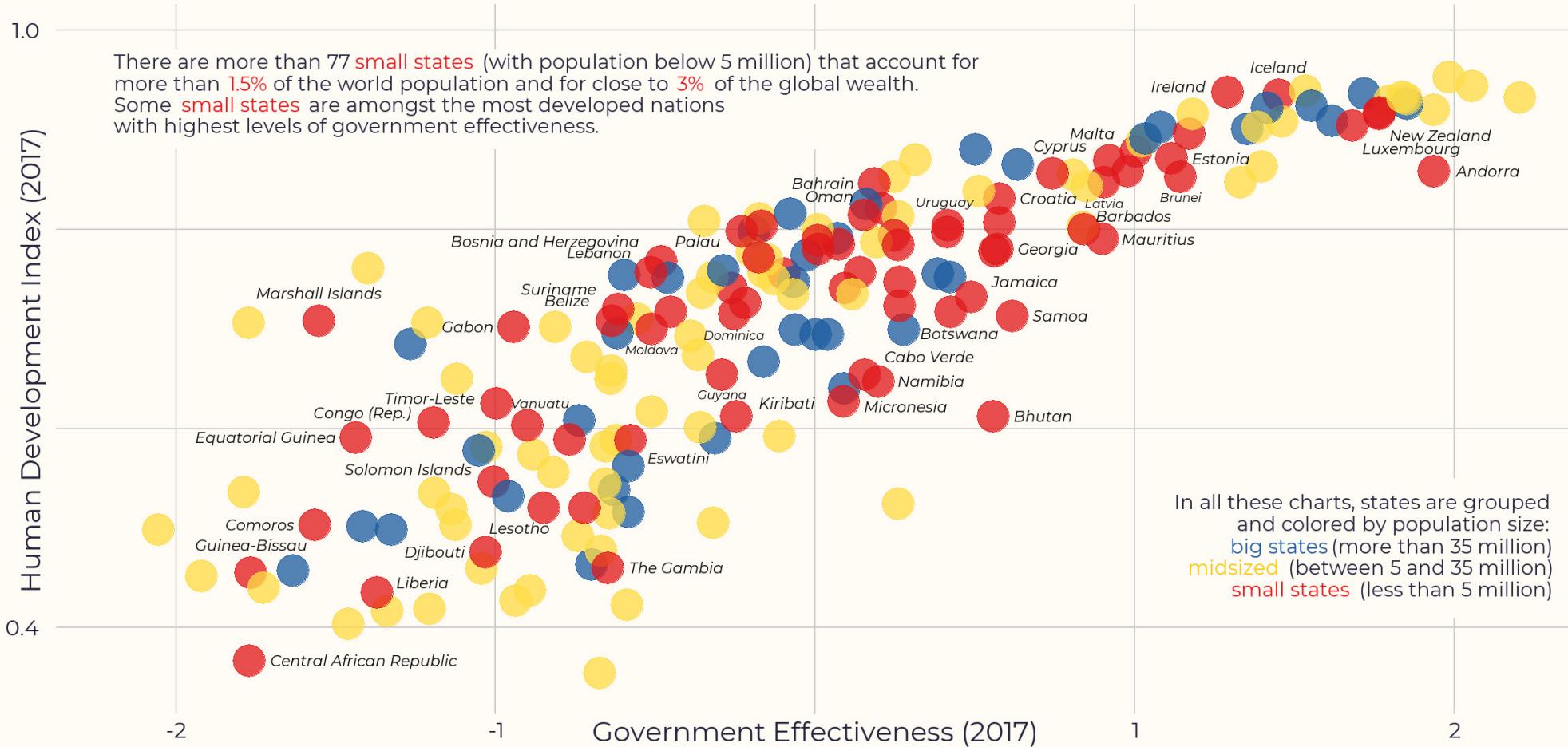
- represents simple features as records in a `data.frame` or `tibble` with a `geometry` list-column
- represents natively in R all 17 simple feature types for all dimensions (XY, XYZ, XYM, XYZM)
- interfaces to GEOS for geometrical operations on projected coordinates, and to s2geometry for geometrical operations on ellipsoidal coordinates
- interfaces to GDAL, supporting all driver options, `Date` and `POSIXct` and list-columns
- interfaces to PROJ for coordinate reference system conversion and transformation
- uses well-known-binary serialisations written in C++/Rcpp for fast I/O with GDAL and GEOS
- reads from and writes to spatial databases such as PostGIS using DBI
- is extended by lwgeom for selected liblwgeom/PostGIS functions
- is extended by stars for raster data, and raster or vector data cubes (spatial time series)



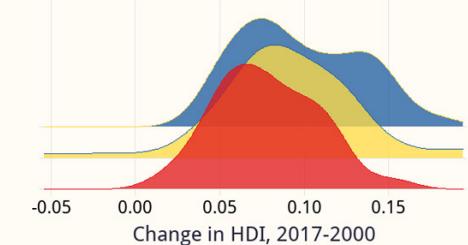
Principles of R

- What is R?
- R is used primarily through interactive command-line.
- R can create and use different types of data but works well with a linear collection of things (vector) and tables.
- R is extremely extendable through packages.
- R is brilliant when it comes to static graphics and dealing with spatial data.

Small States Can Be Big Players in Development and Good Governance

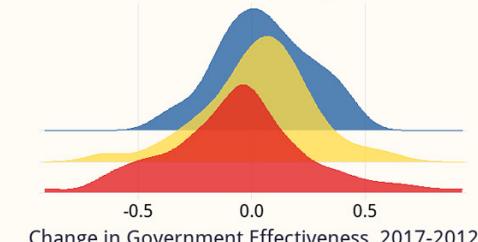


But small states have improved less

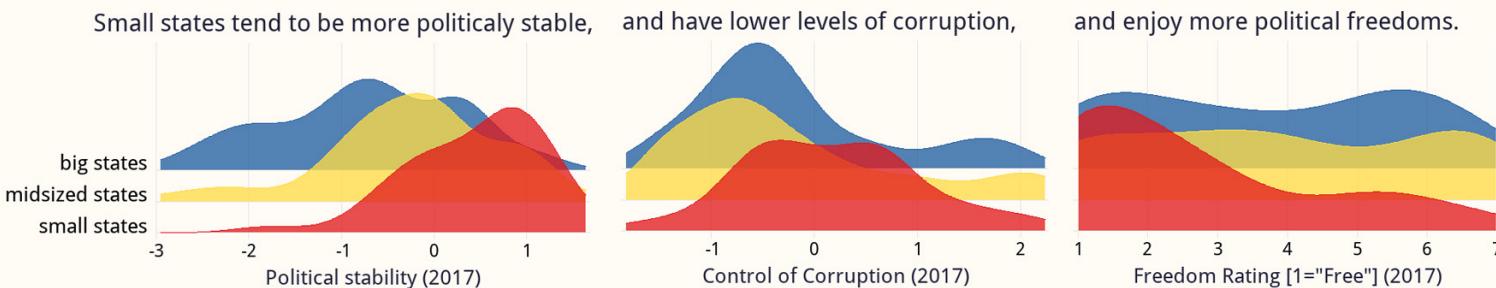


In all these charts, states are grouped and colored by population size:
big states (more than 35 million)
midsized (between 5 and 35 million)
small states (less than 5 million)

and have even lost some ground.



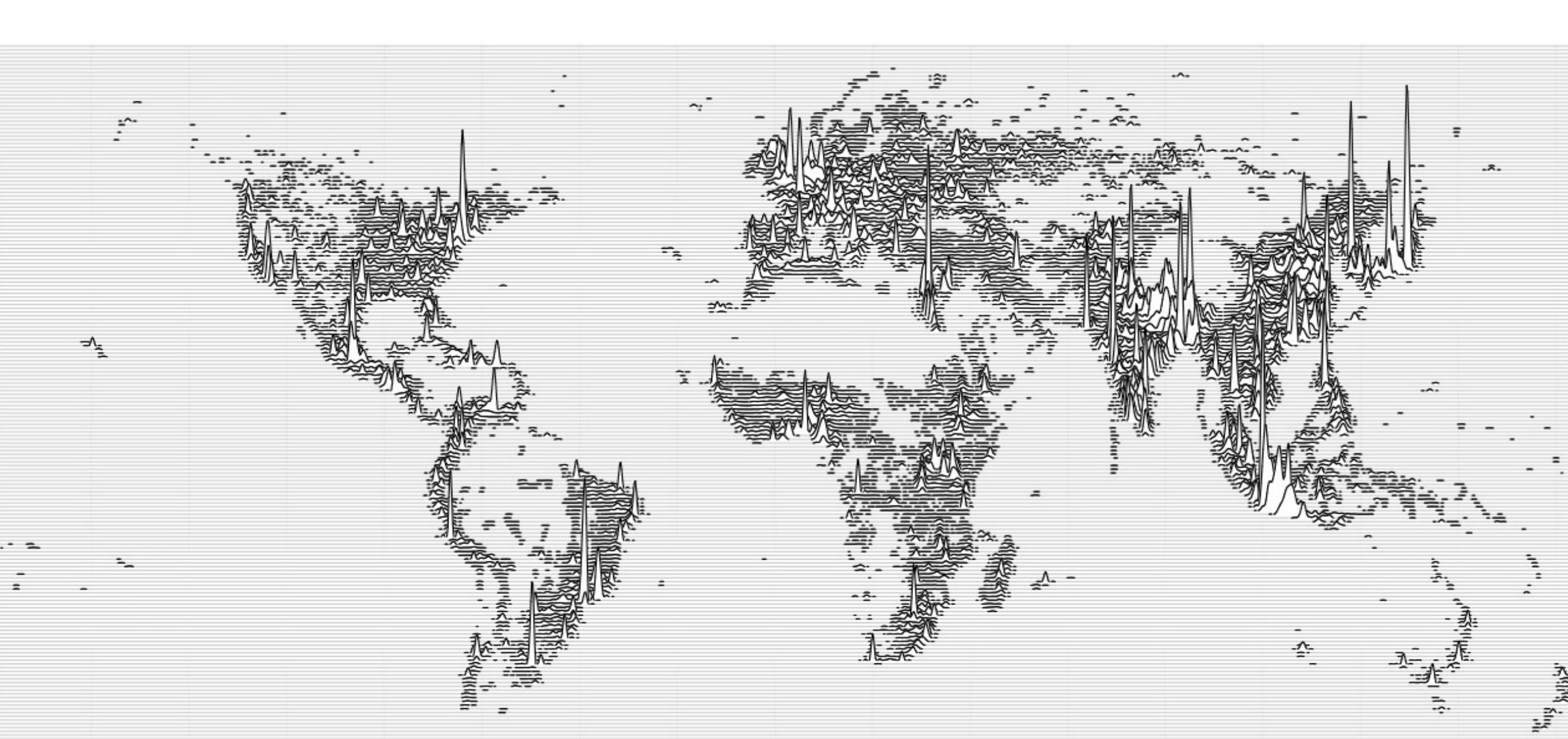
These charts show the distribution densities of different variables by three groups based on state size:



London Cycle Hire Journeys

Thicker, yellower lines mean more journeys





5 more reasons on why you should use R

R is free as in

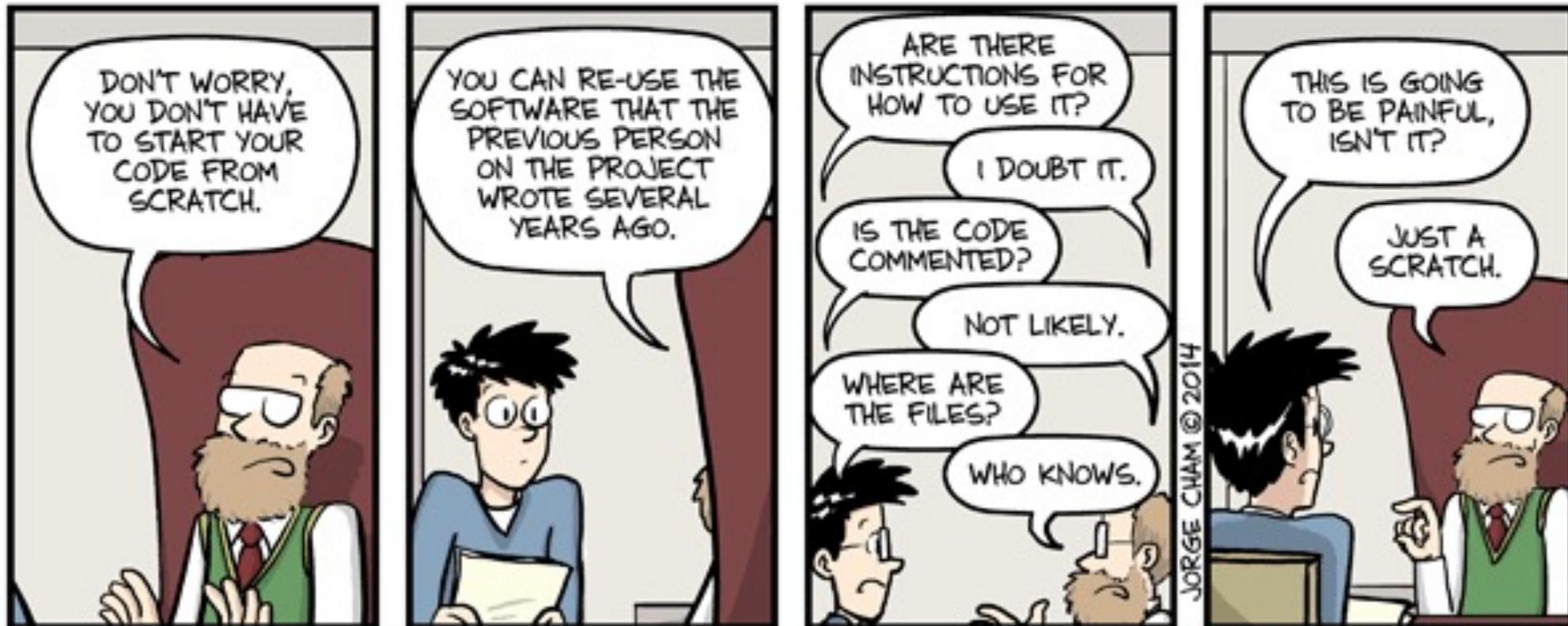


Freedom



Free beer

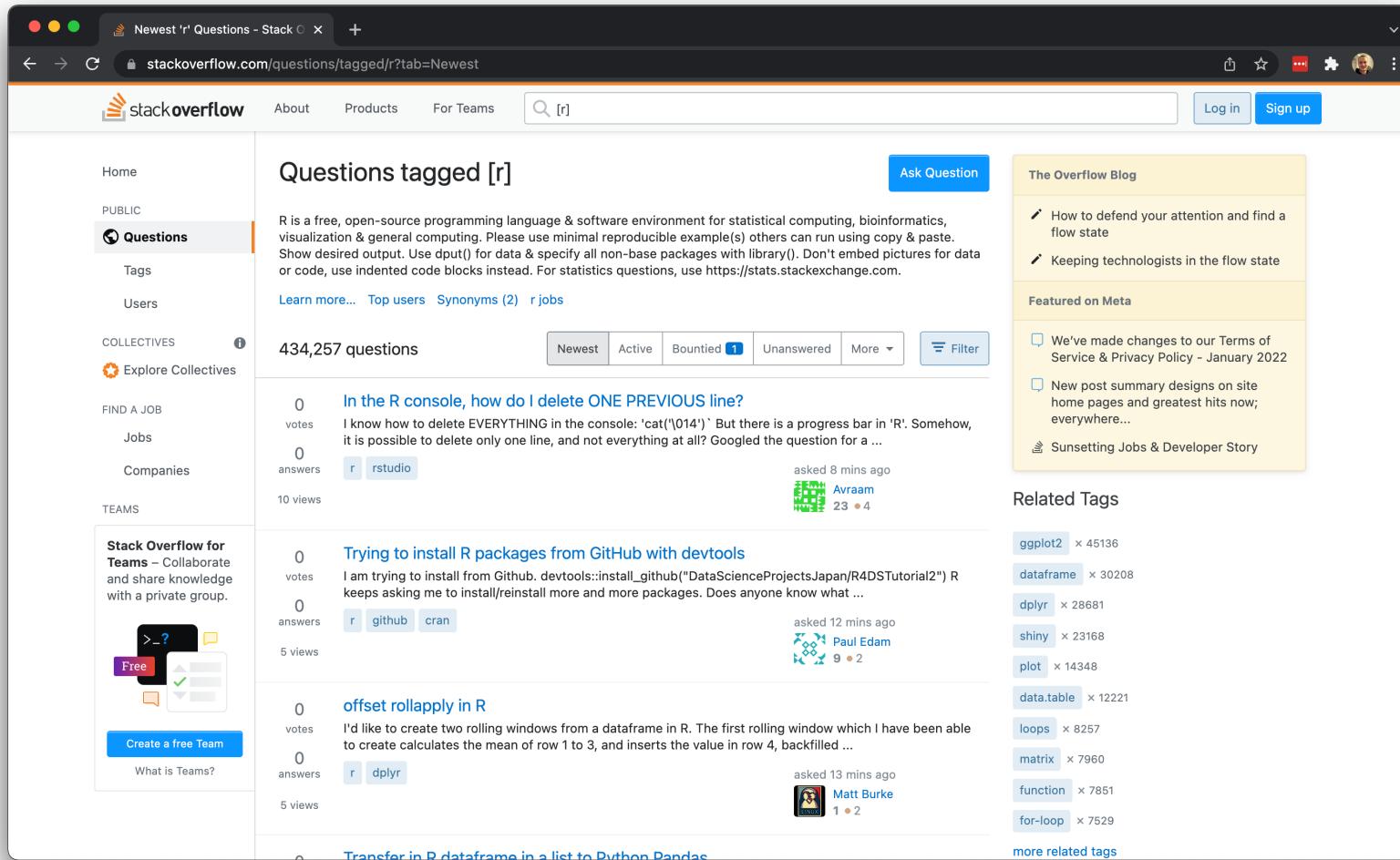
R allows you to produce your outputs programmatically



JORGE CHAM © 2014

WWW.PHDCOMICS.COM

R is supported by a large vibrant community



The screenshot shows the Stack Overflow website with the search term '[r]' entered in the search bar. The results are titled 'Questions tagged [r]'. The page displays several R-related questions, each with details like votes, answers, and tags. A sidebar on the right contains links to 'The Overflow Blog' and 'Related Tags'.

Questions tagged [r]

R is a free, open-source programming language & software environment for statistical computing, bioinformatics, visualization & general computing. Please use minimal reproducible example(s) others can run using copy & paste. Show desired output. Use `dput()` for data & specify all non-base packages with `library()`. Don't embed pictures for data or code, use indented code blocks instead. For statistics questions, use <https://stats.stackexchange.com>.

Learn more... Top users Synonyms (2) r jobs

434,257 questions

Newest Active Bountied 1 Unanswered More Filter

In the R console, how do I delete ONE PREVIOUS line?

I know how to delete EVERYTHING in the console: `'cat('014')` But there is a progress bar in 'R'. Somehow, it is possible to delete only one line, and not everything at all? Googled the question for a ...

0 votes 0 answers r rstudio asked 8 mins ago by Avraam 23 4

Trying to install R packages from GitHub with devtools

I am trying to install from GitHub. `devtools::install_github("DataScienceProjectsJapan/R4DSTutorial2")` keeps asking me to install/reinstall more and more packages. Does anyone know what ...

0 votes 0 answers r github cran asked 12 mins ago by Paul Edam 9 2

offset rollapply in R

I'd like to create two rolling windows from a data frame in R. The first rolling window which I have been able to create calculates the mean of row 1 to 3, and inserts the value in row 4, backfilled ...

0 votes 0 answers r dplyr asked 13 mins ago by Matt Burke 1 2

Transfer in R data frame in a list to Python Pandas

The Overflow Blog

- How to defend your attention and find a flow state
- Keeping technologists in the flow state

Featured on Meta

- We've made changes to our Terms of Service & Privacy Policy - January 2022
- New post summary designs on site home pages and greatest hits now; everywhere...

Sunsetting Jobs & Developer Story

Related Tags

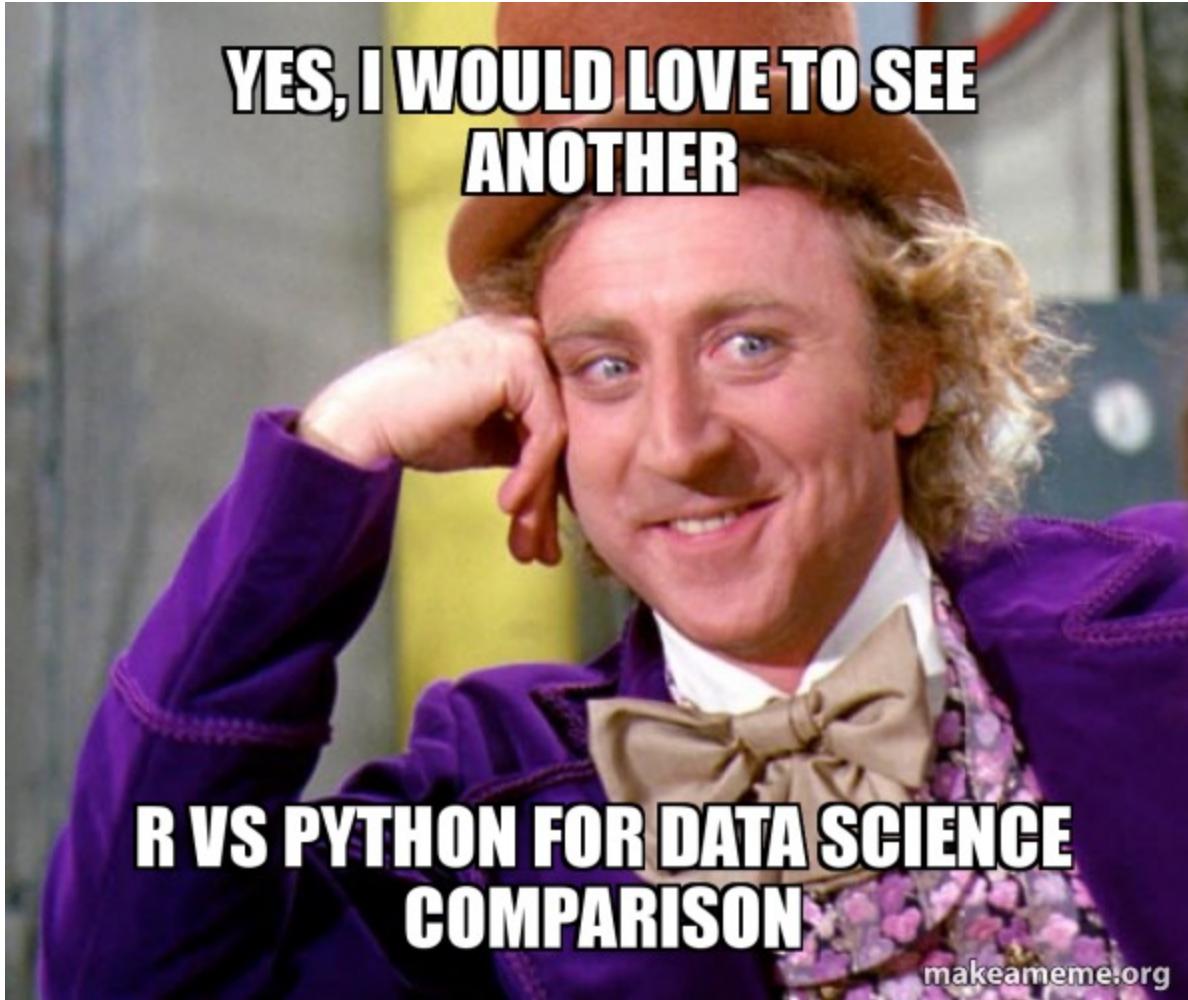
- ggplot2 x 45136
- dataframe x 30208
- dplyr x 28681
- shiny x 23168
- plot x 14348
- data.table x 12221
- loops x 8257
- matrix x 7960
- function x 7851
- for-loop x 7529

more related tags

R is popular and in-demand

	Jan 2022	Jan 2021	Change	Programming Language	Ratings	Change
1	3	1	▲	Python	13.58%	+1.86%
2	1	2	▼	C	12.44%	-4.94%
3	2	3	▼	Java	10.66%	-1.30%
4	4	5		C++	8.29%	+0.73%
5	5	6		C#	5.68%	+1.73%
6	6	7		Visual Basic	4.74%	+0.90%
7	7	8		JavaScript	2.09%	-0.11%
8	11	12	▲	Assembly language	1.85%	+0.21%
9	12	13	▲	SQL	1.80%	+0.19%
10	13	14	▲	Swift	1.41%	-0.02%
11	8	9	▼	PHP	1.40%	-0.60%
12	9	10	▼	R	1.25%	-0.65%
13	14	15	▲	Go	1.04%	-0.37%
14	19	20	▲	Delphi/Object Pascal	0.99%	+0.20%
15	20	21	▲	Classic Visual Basic	0.98%	+0.19%

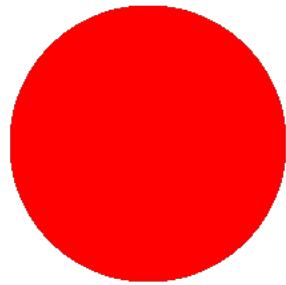
Python?



R Markdown

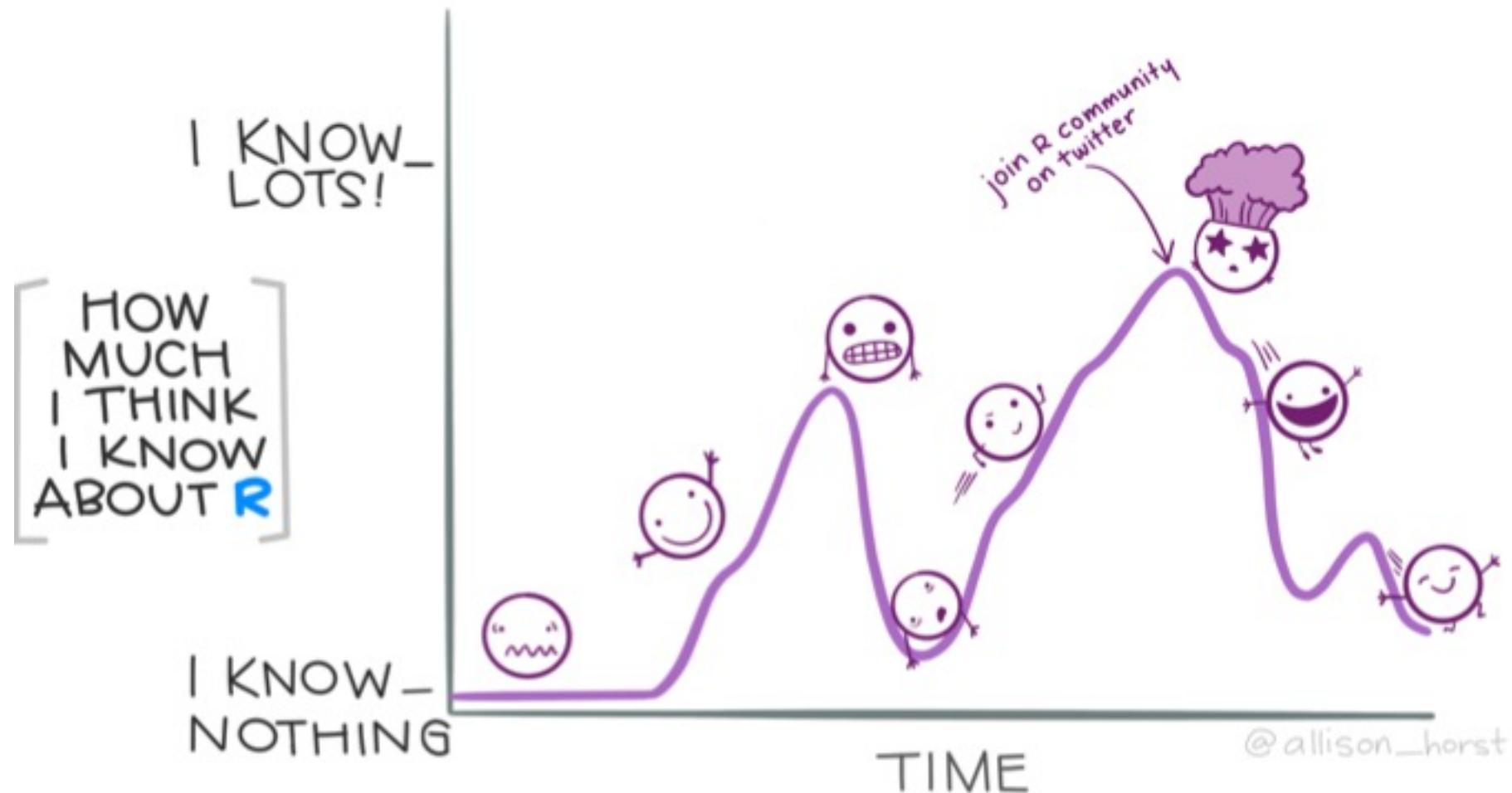
- Markdown is a lightweight markup language for creating formatted text using a plain-text editor.
- R Markdown is an extension of the markdown syntax that enables R code to be embedded in them in a way which can later be executed.
- Why do we want this? Typesetting.

RStudio



LIVE

Learning curve



Conclusion

- R comes from the rock-stars of the computer science industry.
- R is primarily command line based.
- R is extremely powerful, versatile and popular.
- R is free and Open Source.
- Plenty of tools and community around R.
- It is arguably of the best transferable skills you can learn.
- Supports reproducible academic research.

Further resources

Analysis in R

- [R for Data Science](#)
- [Advanced R](#)
- [Geocomputation in R](#)

R Markdown

- R Markdown [resources on Rstudio](#)
- [Definitive Guide to R Markdown](#)

Computer tutorial

- Today: general principles of programming and how to use R (and RStudio) effectively for data analysis.
- Carefully read the instructions although there *may* be small deviations in where to find certain menu buttons.
- Assignment: no need to hand-in but if you want to leave before the end of the computer tutorial you should be able to show your results.

Geocomputation Help Session

- Thursday from 14h00-15h00 in Foster Court 215.
- Bring your own laptop.

Coding Therapy

The screenshot shows an email window with a dark theme. The header includes the recipient's name 'Wyszomierski, Jakub' and a subject line 'Coding Therapy is back!'. The email body starts with 'Dear all,' followed by a reminder about the session times and location. It then describes the purpose of Coding Therapy sessions, emphasizing a supportive environment for everyone. The signature at the bottom provides contact information for Jakub Wyszomierski.

JW Wyszomierski, Jakub
Coding Therapy is back!

Inbox - UCL 09:46

Dear all,

This is a gentle reminder that Coding Therapy is back for Spring Term. The sessions will be held on Wednesdays from 1:00pm to 3:00pm. For this week and the rest of the term, it will be hosted in the **Reading Room next to the Geography Map Room in Bedford Way (26)**.

For those who are unaware, Coding Therapy sessions are aimed at helping everyone in the department, from first year undergraduate to staff develop coding in a mutually supportive environment. Come to the drop-in sessions where you can work on your own projects, discuss your coding issues, get help and offer help to others. You do not need to have a specific question or project and you are more than welcome to come to work on your own coding-related (or not) stuff. Please feel free to come with all of the projects regardless the coding language (we are happy to help with MS Excel too) – you might well find someone with similar use cases for them!

Kind regards,
Jakub and Nikki

—

Jakub Wyszomierski
PhD Student & PGTA
UCL Department of Geography
North-West Wing, Gower Street, London, WC1E 6BT
jakub.wyszomierski.16@ucl.ac.uk

Pronouns: he, him, his

Conclusion



Conclusion



Questions

Justin van Dijk

j.t.vandijk@ucl.ac.uk

