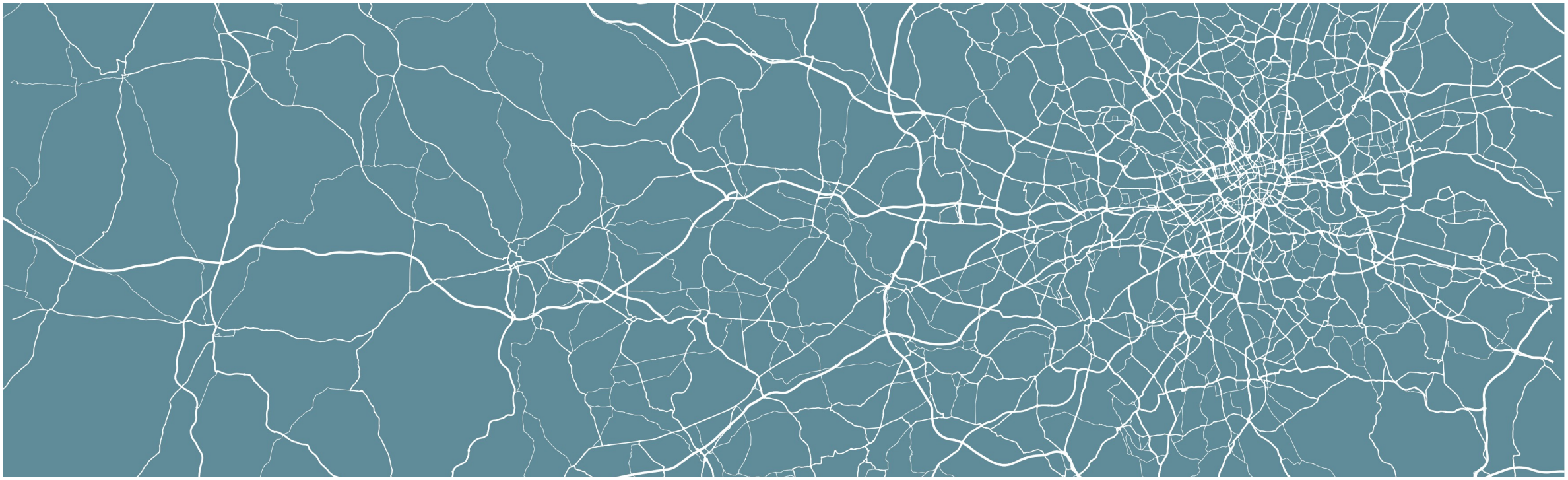
Geocomputation

W4 – Programming for Data Analysis



UCL

Where are we at?

Part I: Foundational Concepts

W1	Geocomputation: An Introduction	
W2	GIScience and GIS Software	QGIS
W3	Cartography and Visualisation	
W4	**Programming for Data Analysis**	R
W5	Programming for Spatial Analysis	

This week

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

Before we start

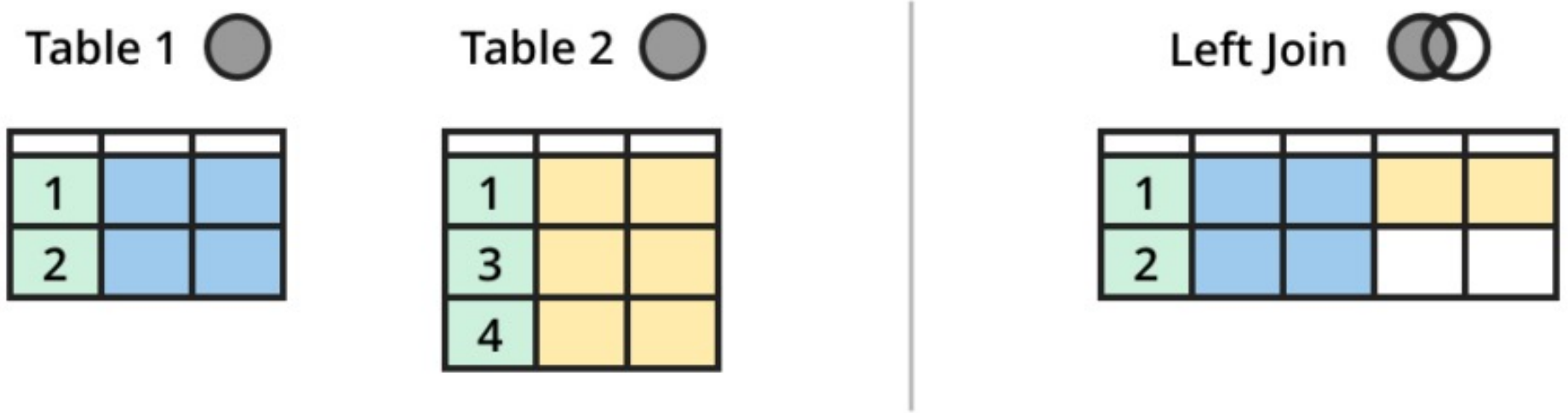
- Go to www.menti.com
- Use code: 9211 2788

Recap

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

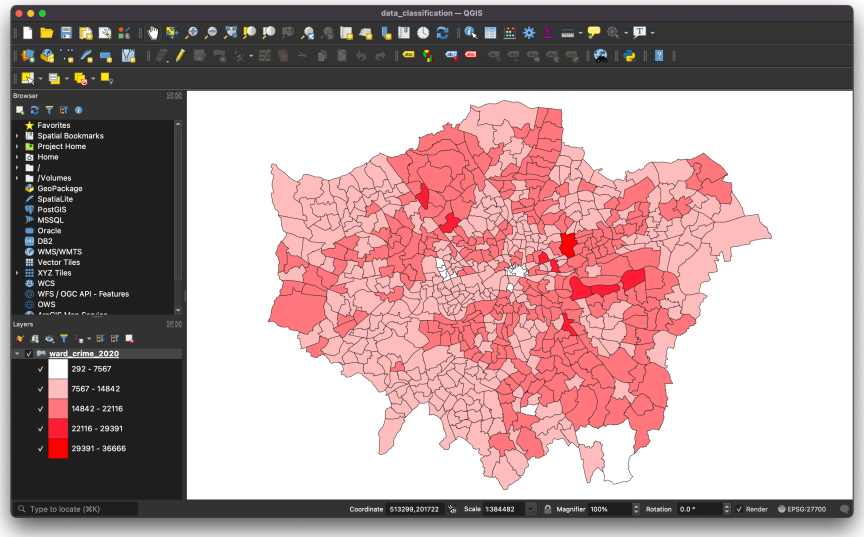
8	9	9	10	0	10	10	10	0	0	0	7	5	3	0	0	0	0	1
8	9	9	10	10	0	10	9	9	0	0	5	3	0	0	0	0	0	0
8	8	9	9	10	0	0	9	8	7	5	0	0	0	1	0	0	0	0
5	8	8	9	10	10	0	9	7	5	0	0	5	5	5	0	0	0	1
3	5	8	9	9	10	0	0	3	0	0	0	5	0	0	1	0	0	2
2	5	8	8	9	9	10	0	0	0	1	5	0	0	0	0	0	0	1
2	4	6	8	8	9	0	0	0	1	5	0	0	5	5	5	0	0	1
0	3	6	8	8	0	0	0	0	5	0	5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

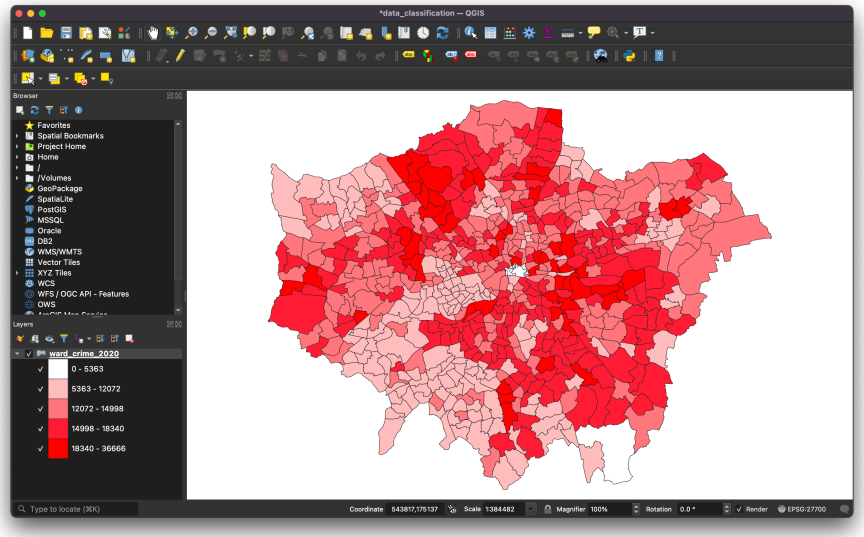
Attribute join

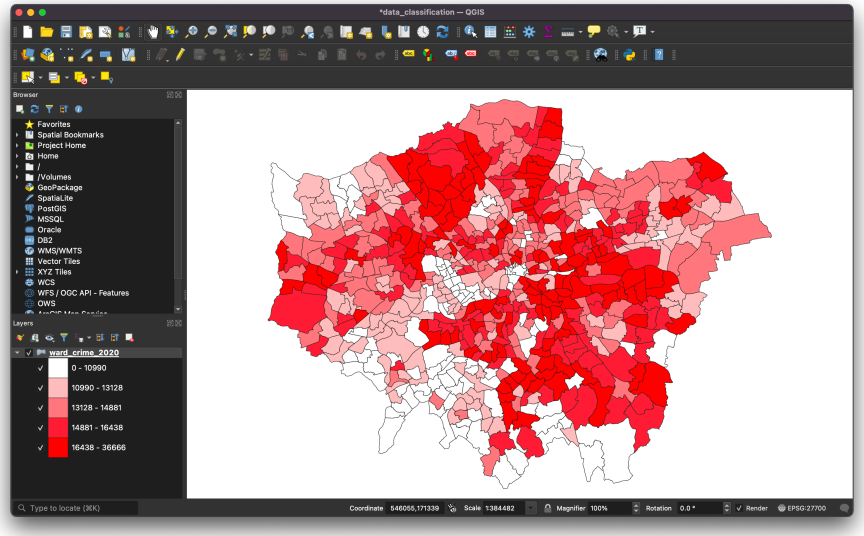


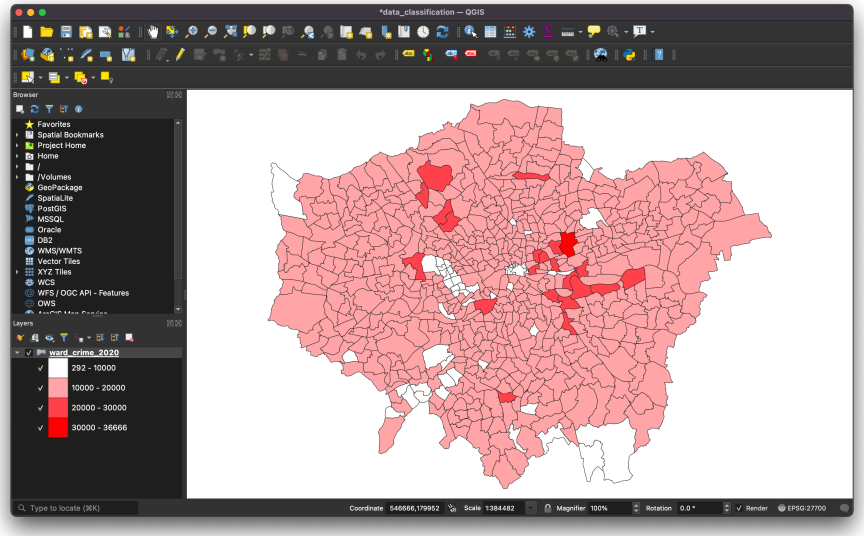
The Data School. 2019. *SQL Join Types Explained Visually*.

Classifying data

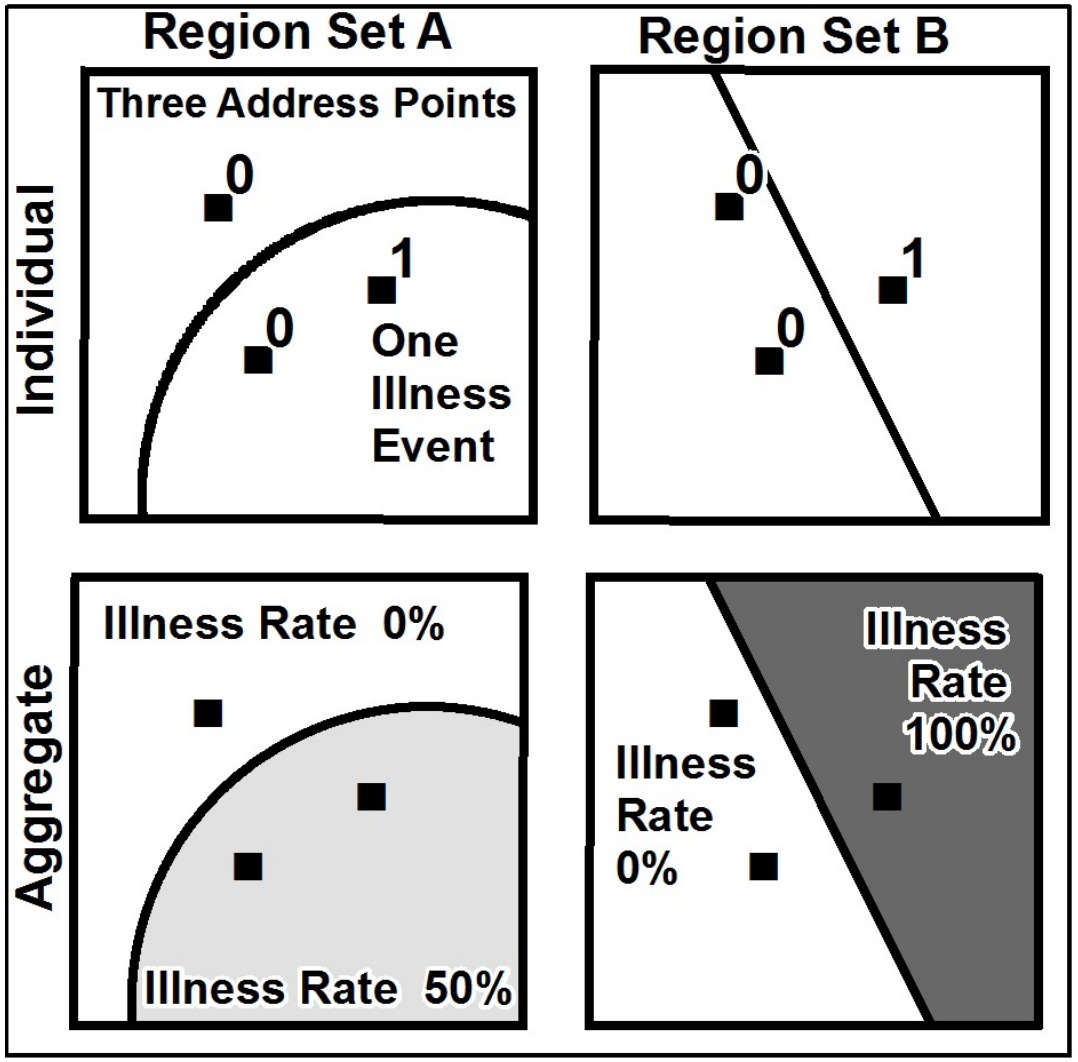




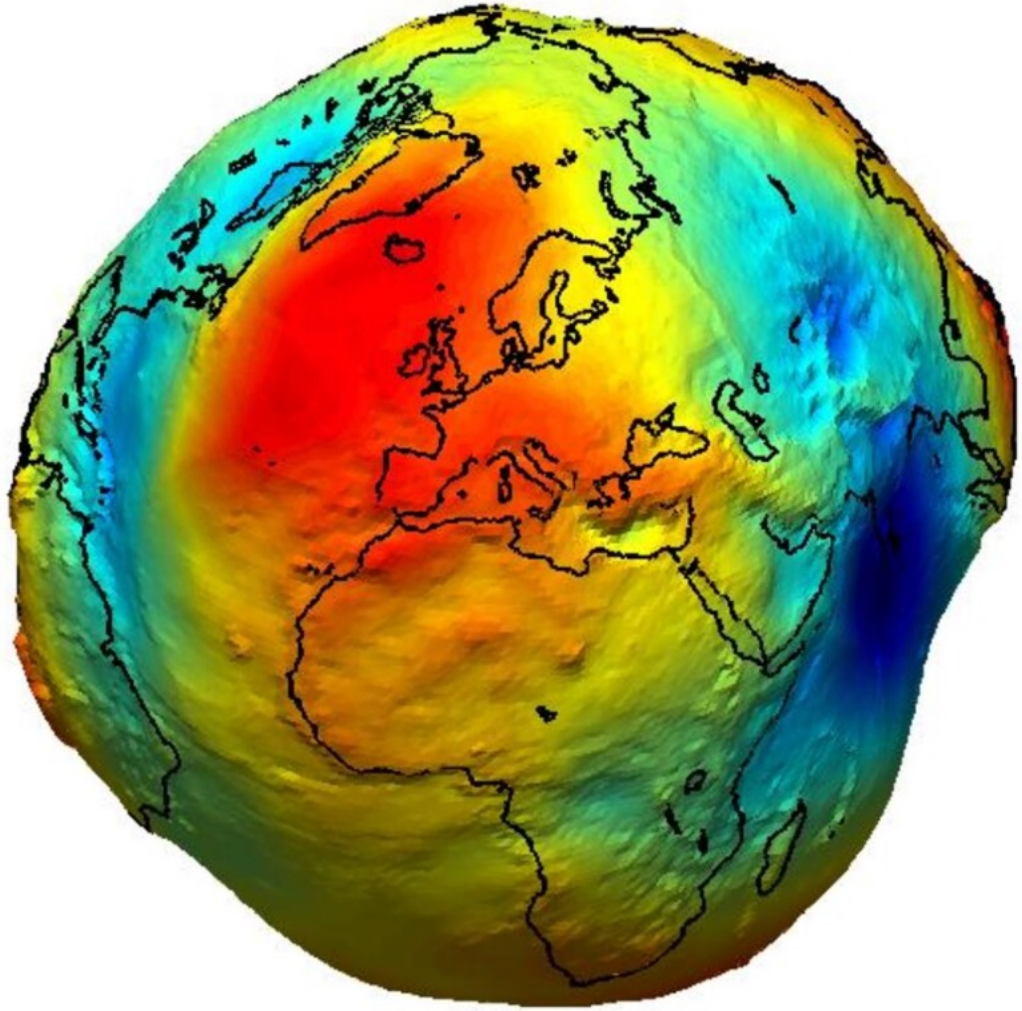




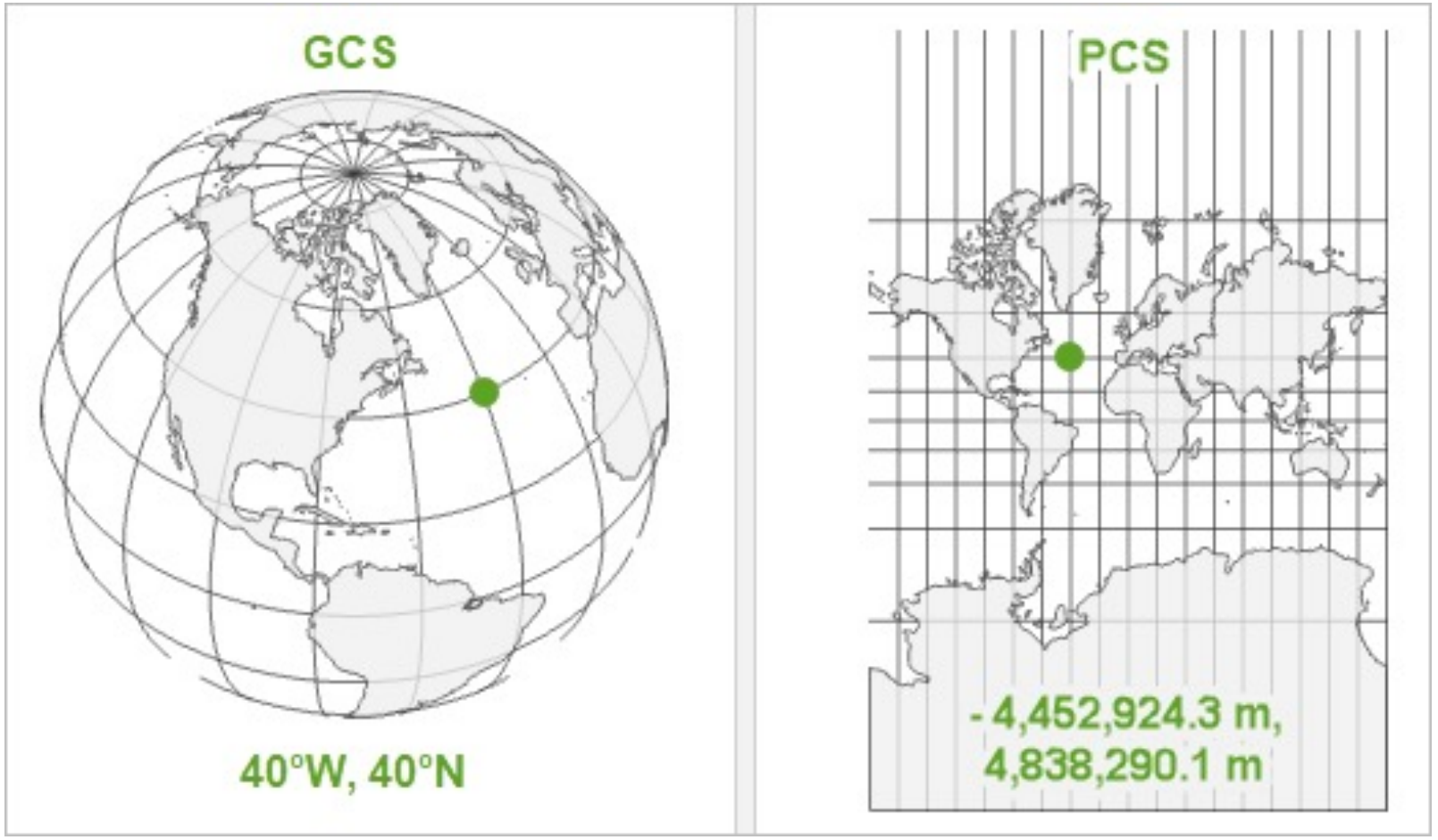
Modifiable Areal Unit Problem



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

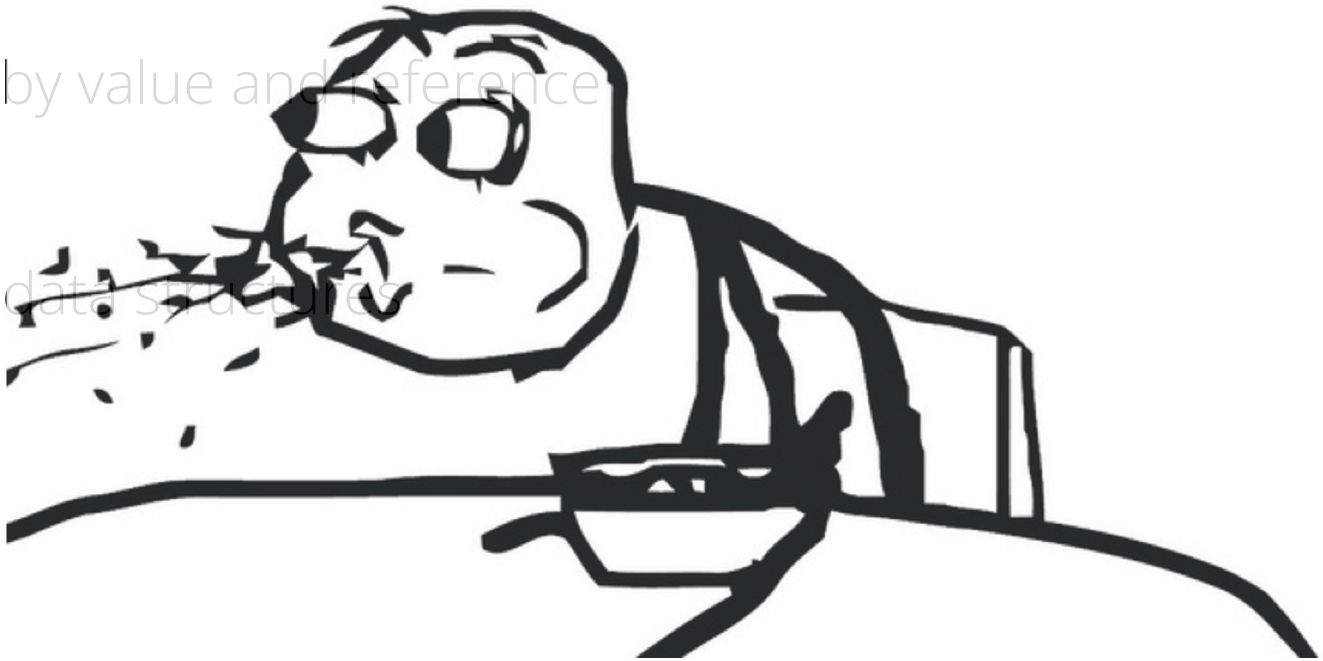
- From ”point-and-click” to “writing commands”.
- In our case: writing commands to execute various data management and analysis tasks.
- We will be using a programming language called R.

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 πριμιτιfε δατα τψπες
- Ασσιγνμεντ, αριτημετιξ, λογιξαλ ανδ ρελατιοναλ οπερατορς
- Εχπρεσσιον ανδ στατεμεντς, δεβυγγινγ
- Φλοω οφ ξοντρολ· σελεξτιον ανδ ρεπετιτιον
- Φυνξτιονς, παραμετερς πασσινγ, ξαλλ βψ αfλυε ανδ ρεφερενξε
- Οβσεξτ-οριεντεδ προγραμμινγ
- 1/2 διμενσιοναλ αρραψς, στρινγς ανδ δατα στρυξτυρες

Principles of R

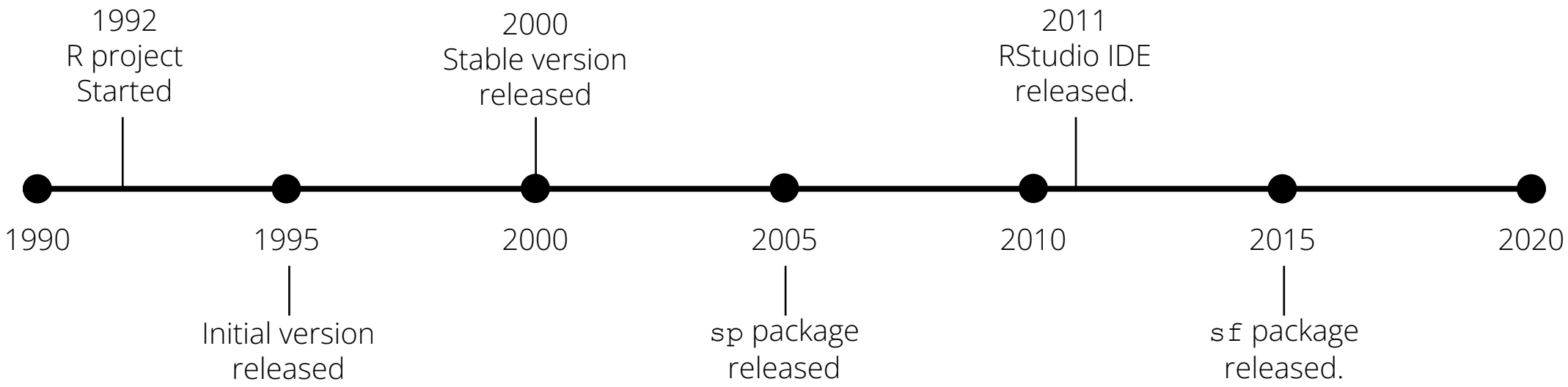
- What is R?
- R is used primarily through interactive command-line.
- R can create and use different types of data but works well with a linear collection of things (vector) and tables.
- R is extremely extendable through packages.
- R is brilliant when it comes to static graphics and dealing with spatial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



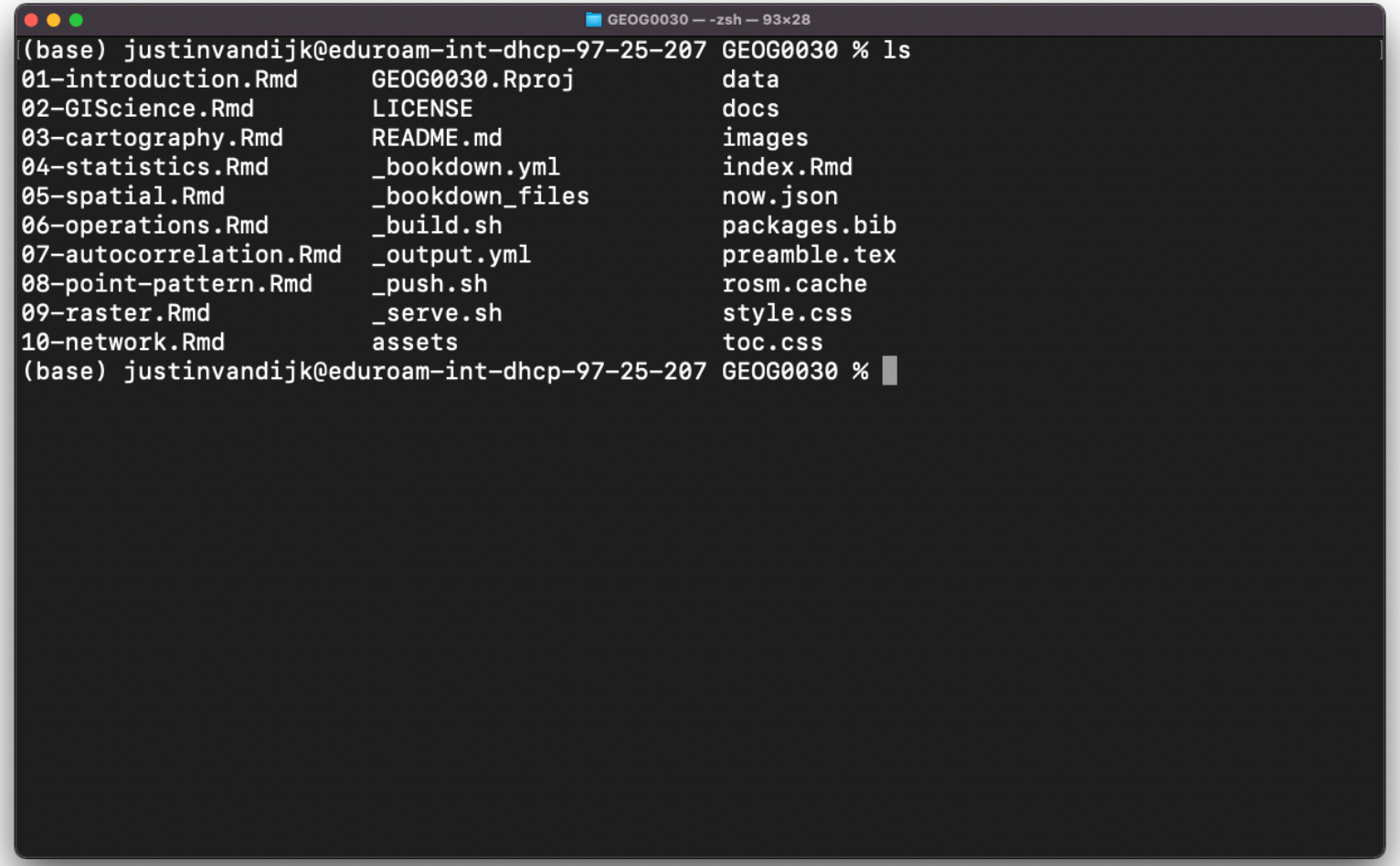
Principles of R

- 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

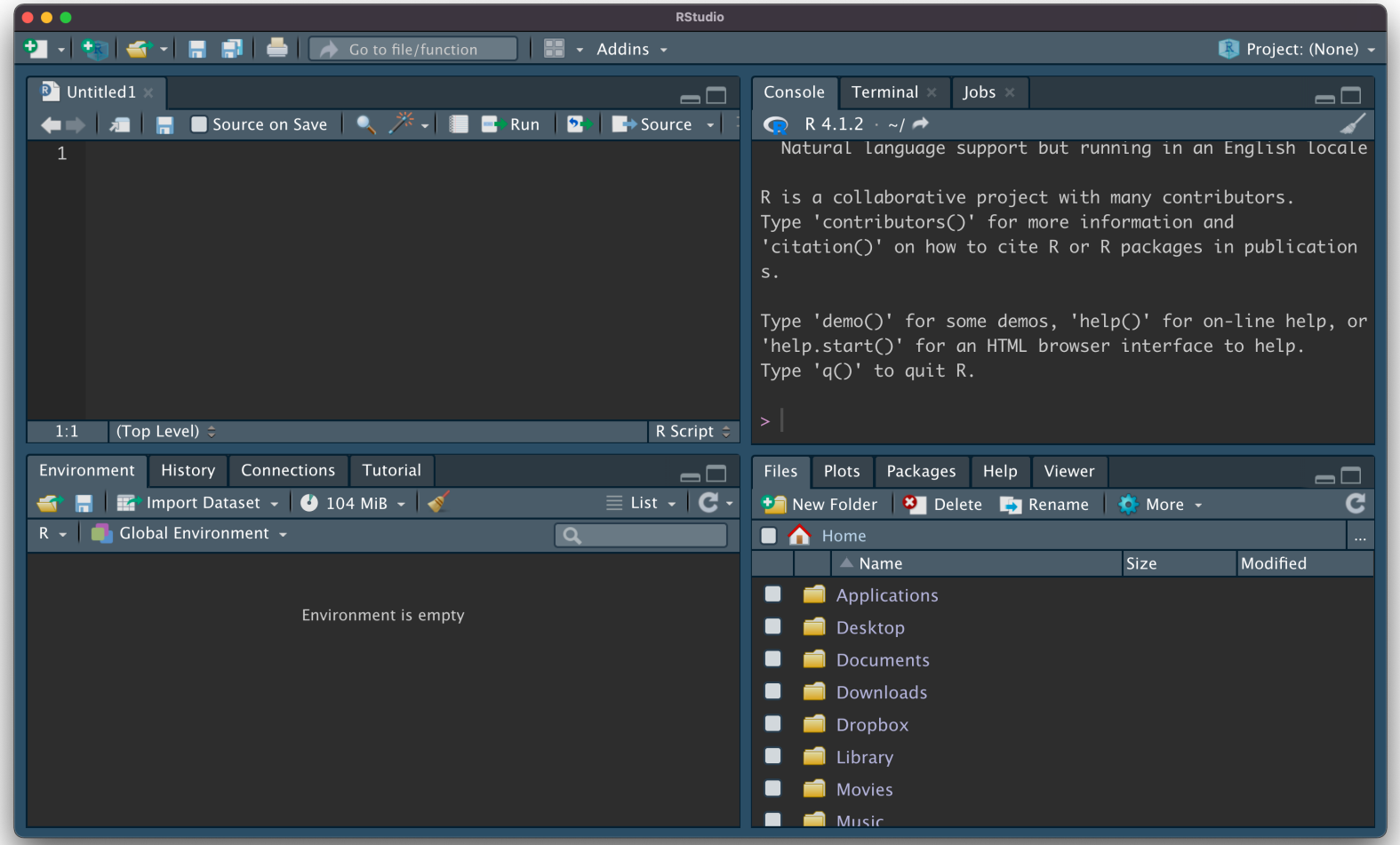


Graphical User Interface

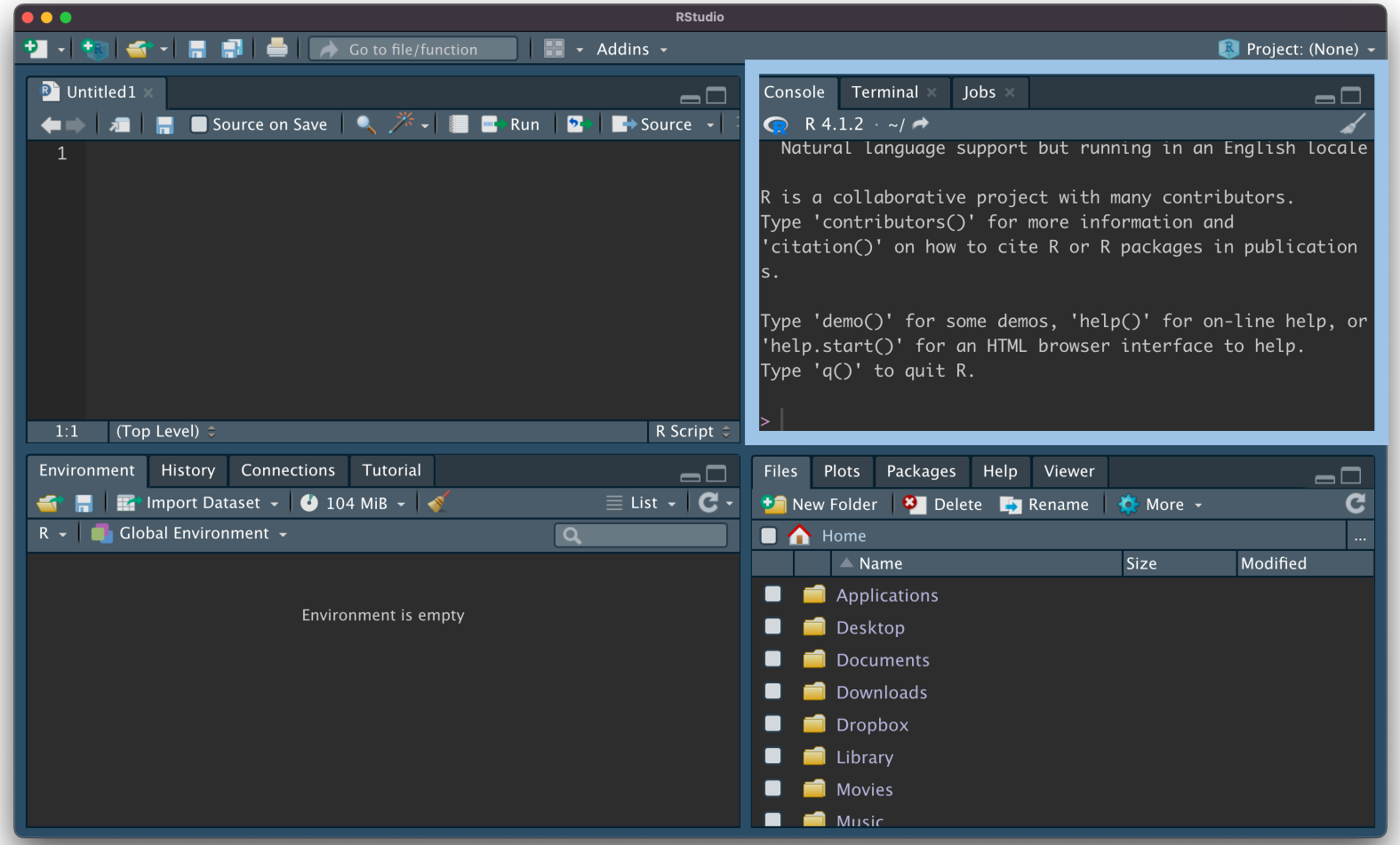
Interaction through language

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

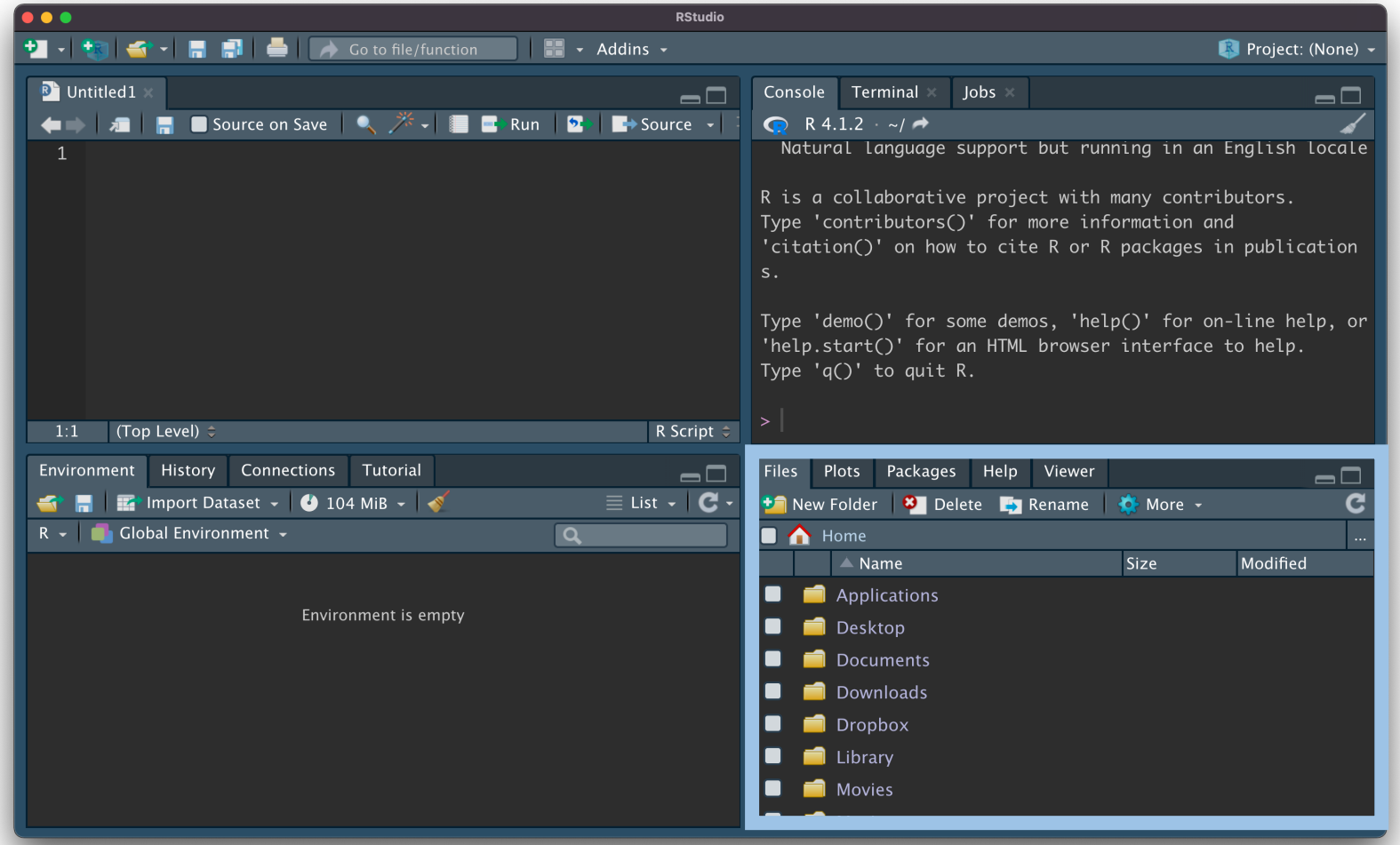
How does this work in R



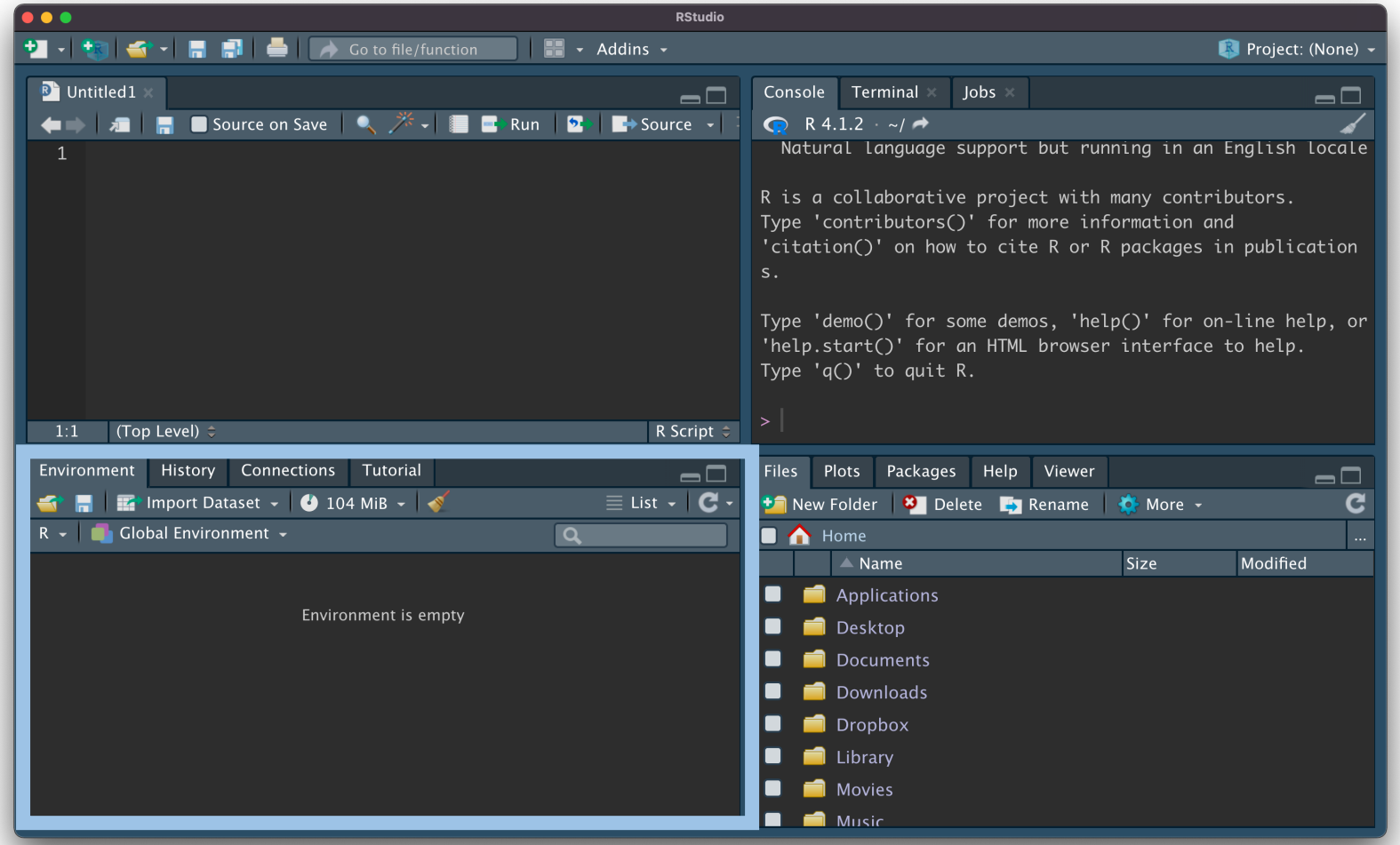
How does this work in R



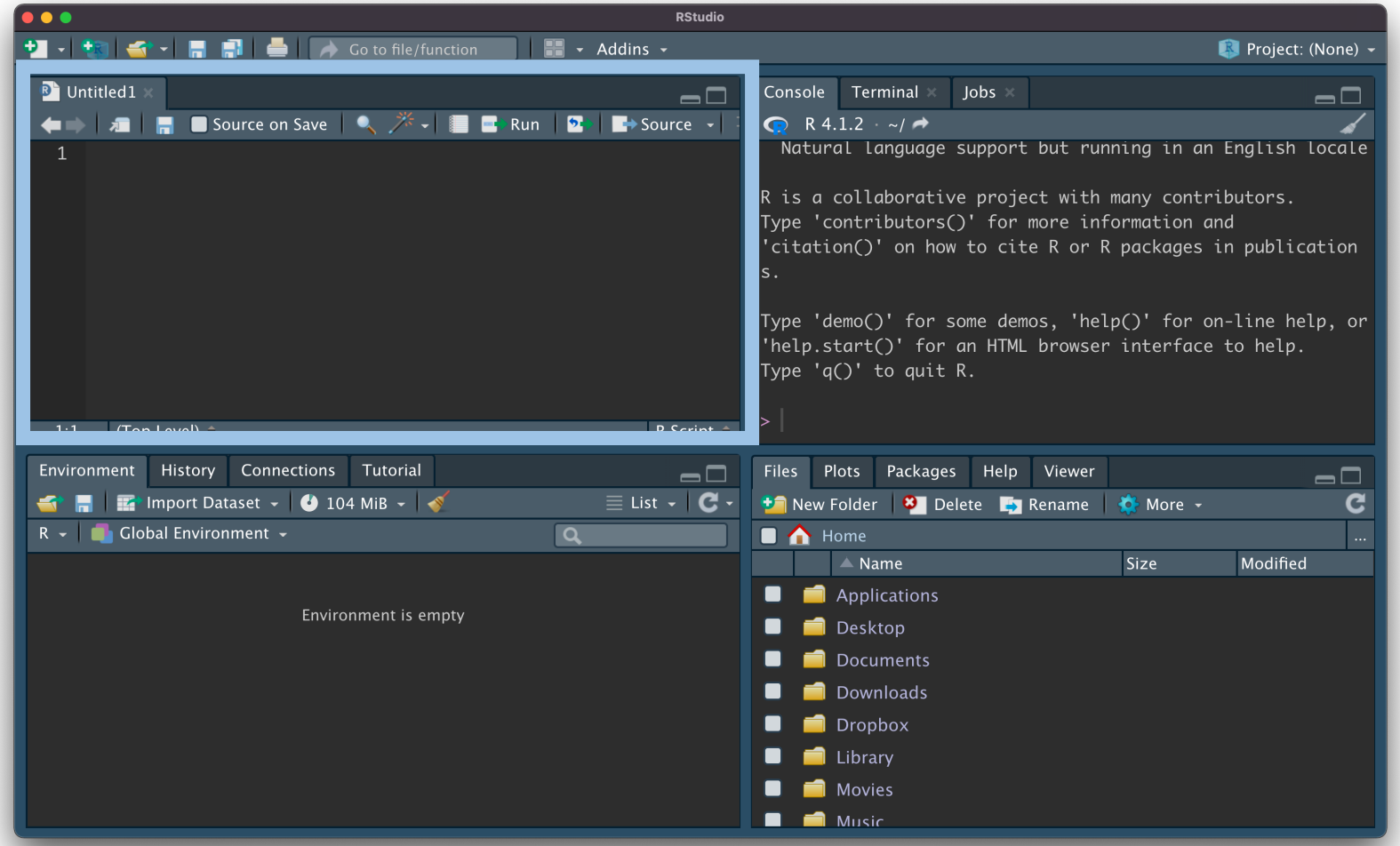
How does this work in R



How does this work in R



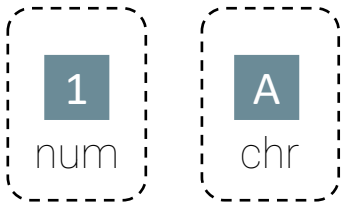
How does this work in R



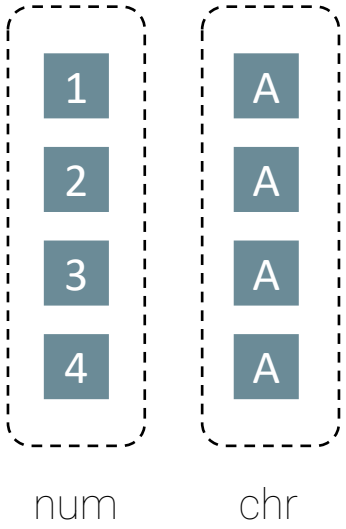
Principles of R

- 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.

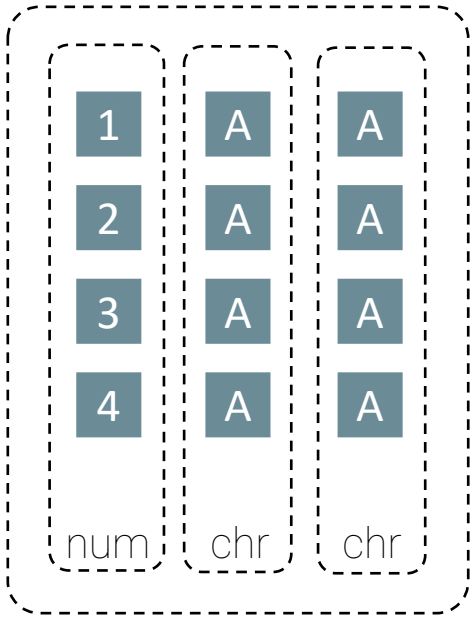
Data types



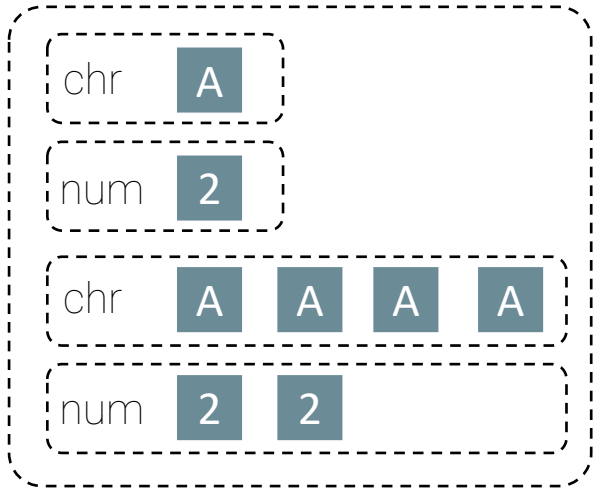
Scalar



Vector



Dataframe



List

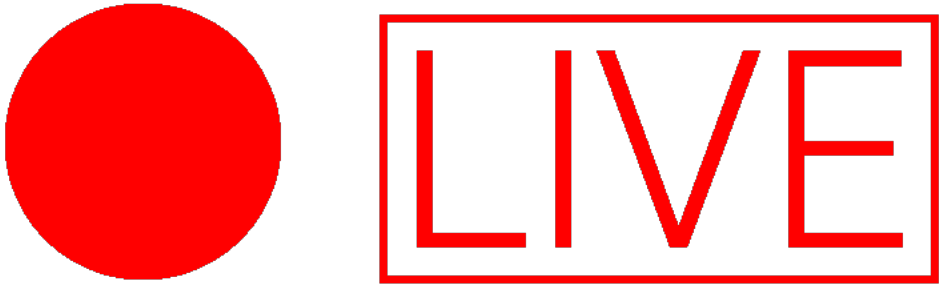
Variables

- 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.

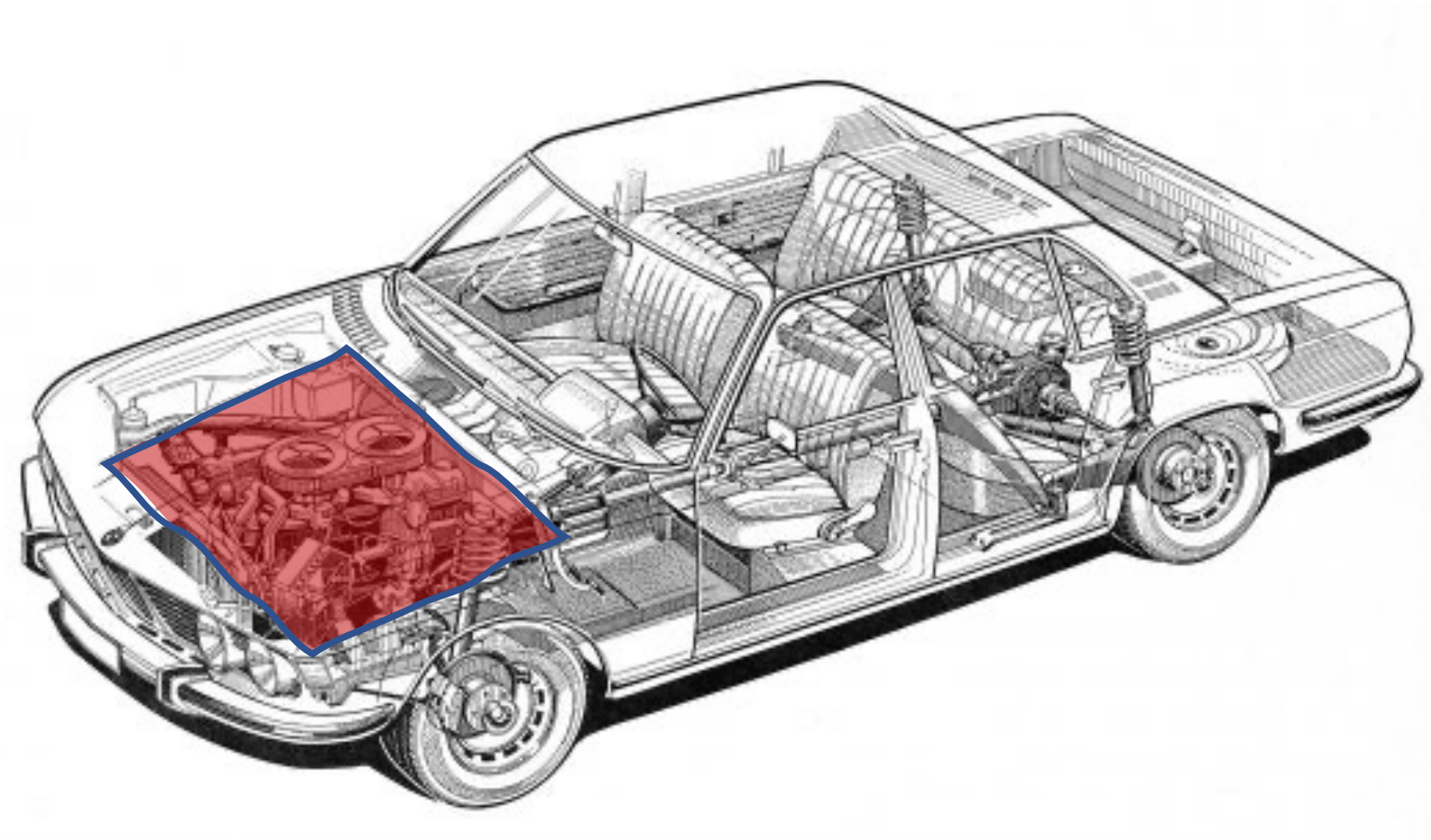
RStudio



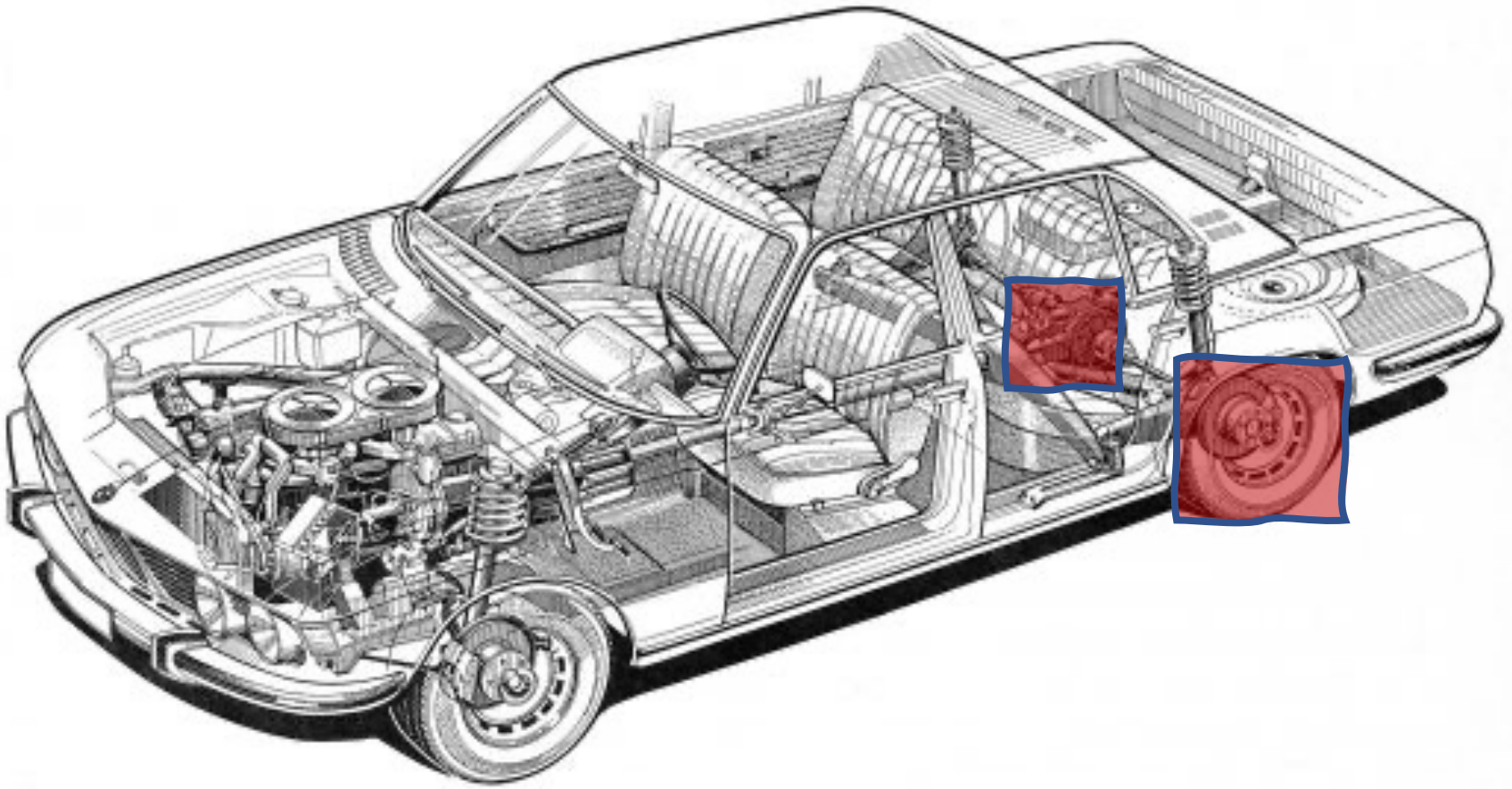
Principles of R

- 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.

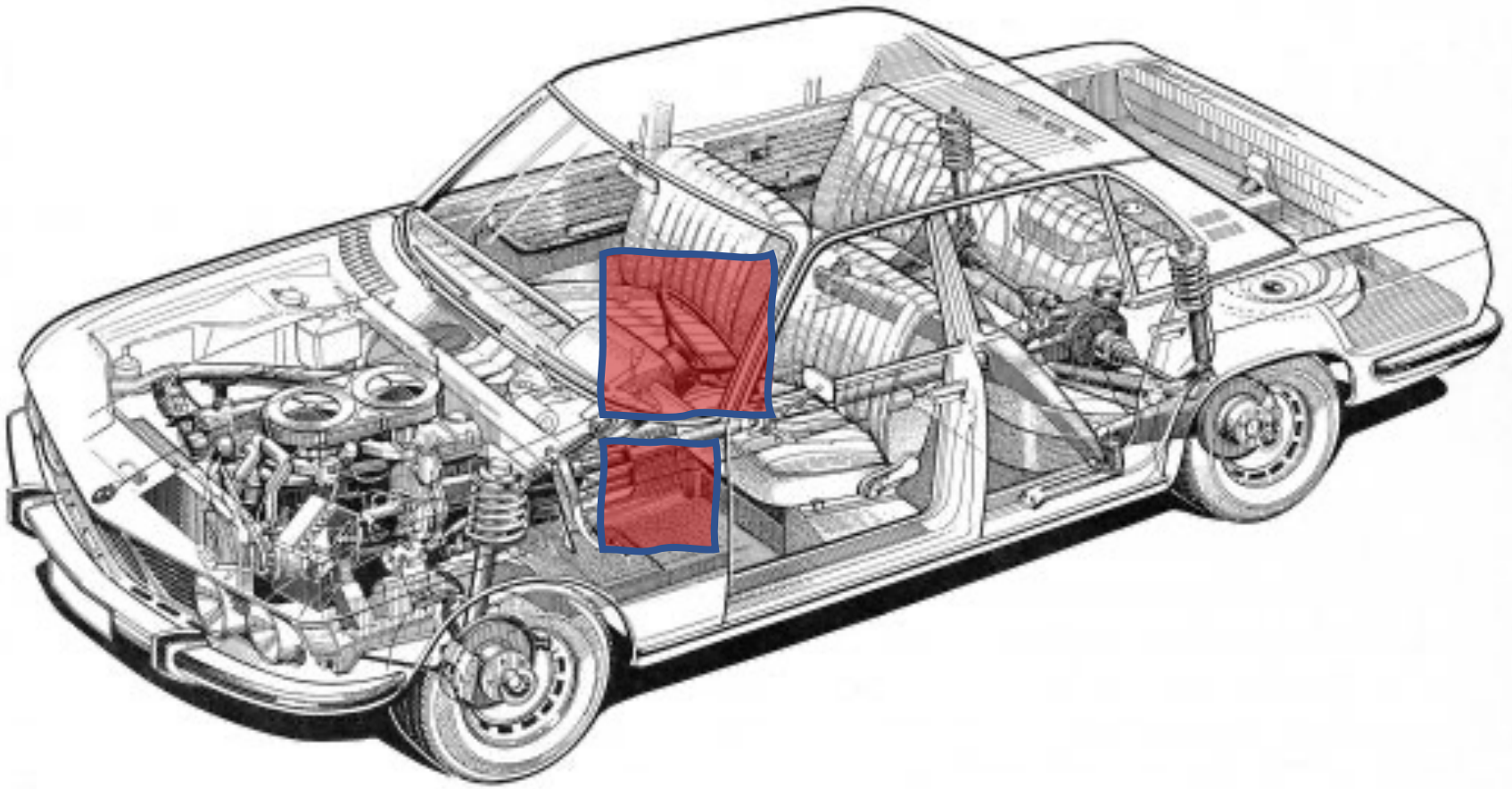
Packages



Packages



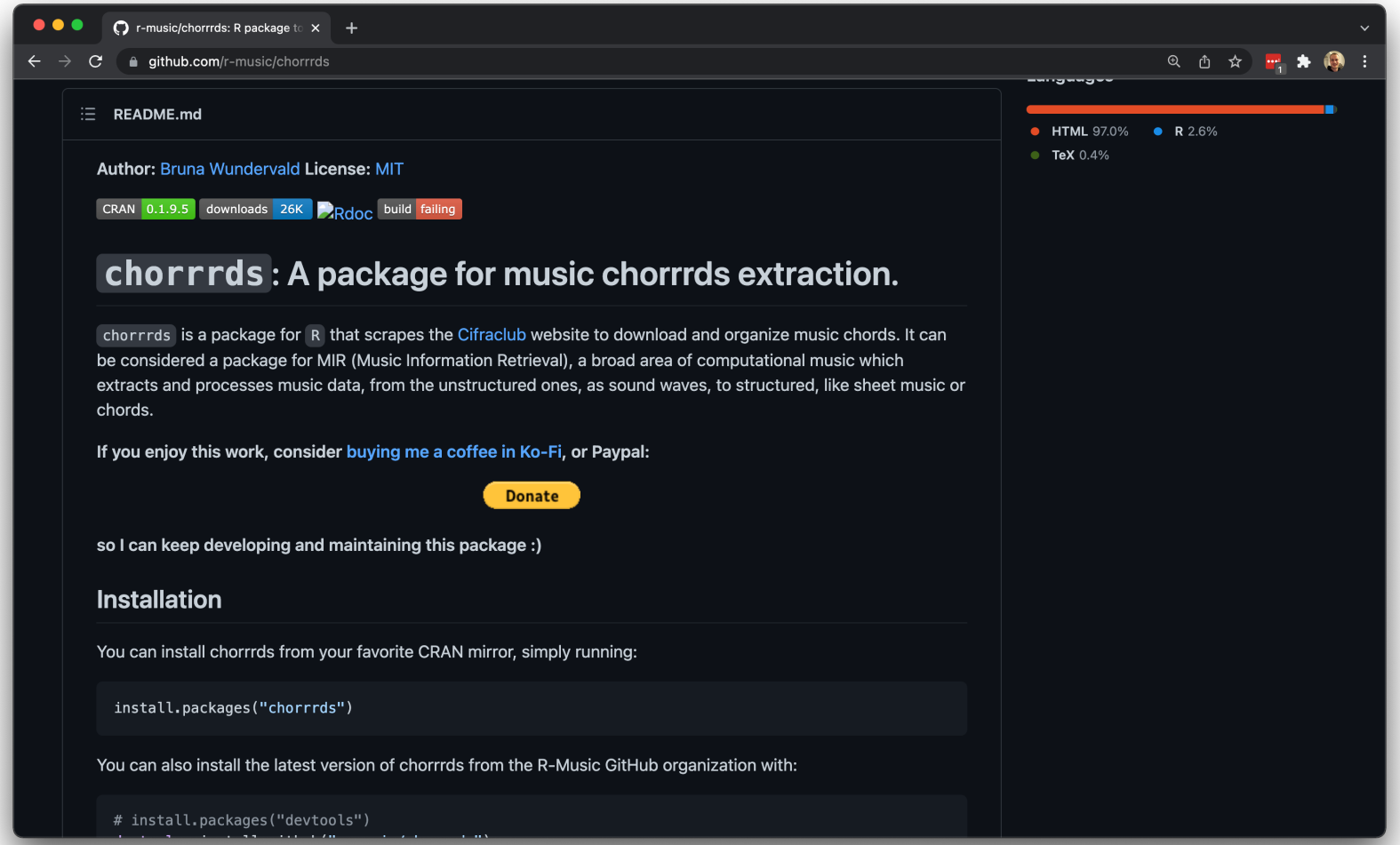
Packages



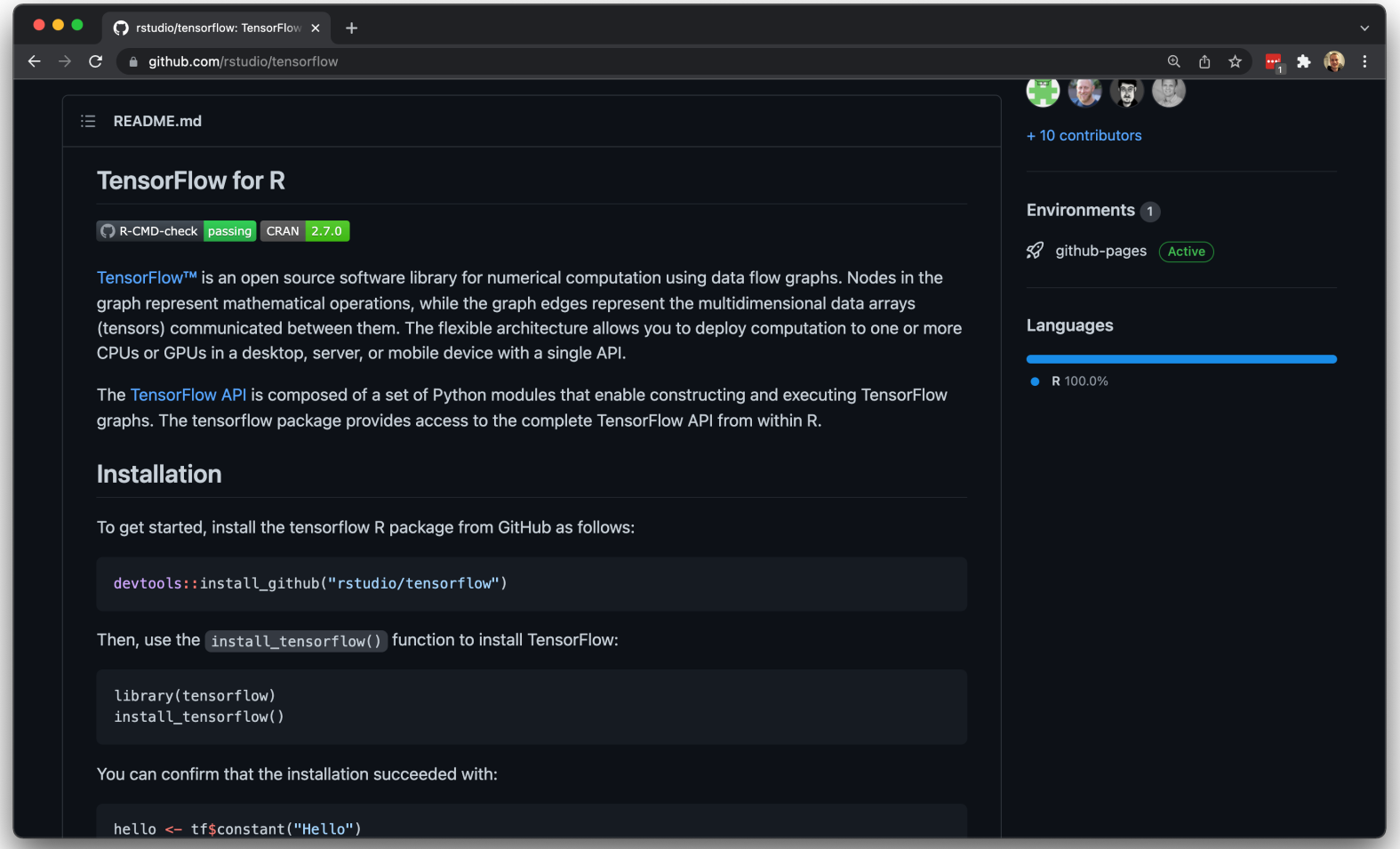
Packages

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

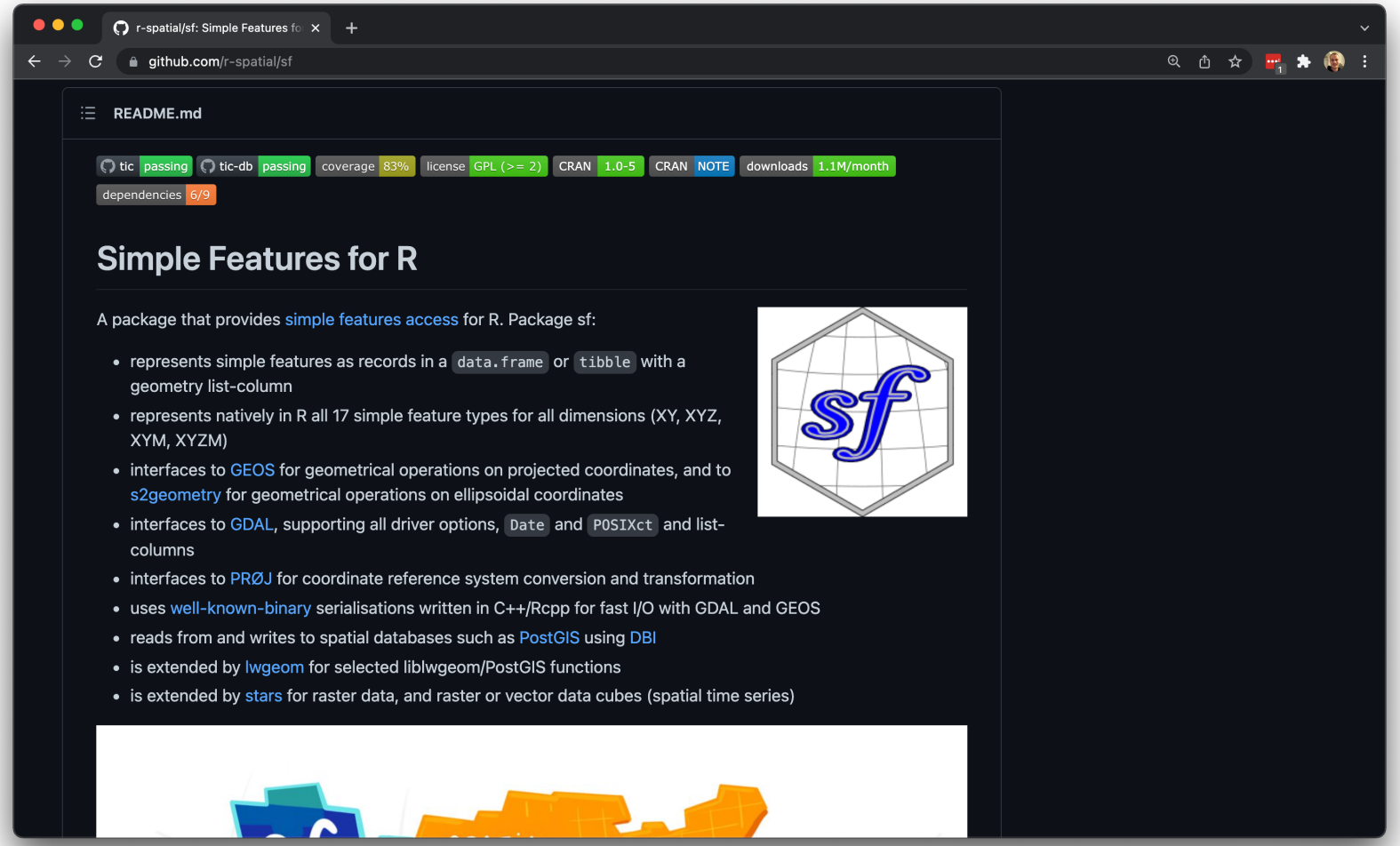
chorrrds


r-music/chorrrds: R package
github.com/r-music/chorrrds
README.md
Author: Bruna Wundervald 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")
HTML 97.0%
R 2.6%
TeX 0.4%

tensorflow


rstudio/tensorflow: TensorFlow
github.com/rstudio/tensorflow
README.md
TensorFlow for R
R-CMD-check passing
CRAN 2.7.0
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")
+ 10 contributors
Environments 1
github-pages Active
Languages
R 100.0%

sf

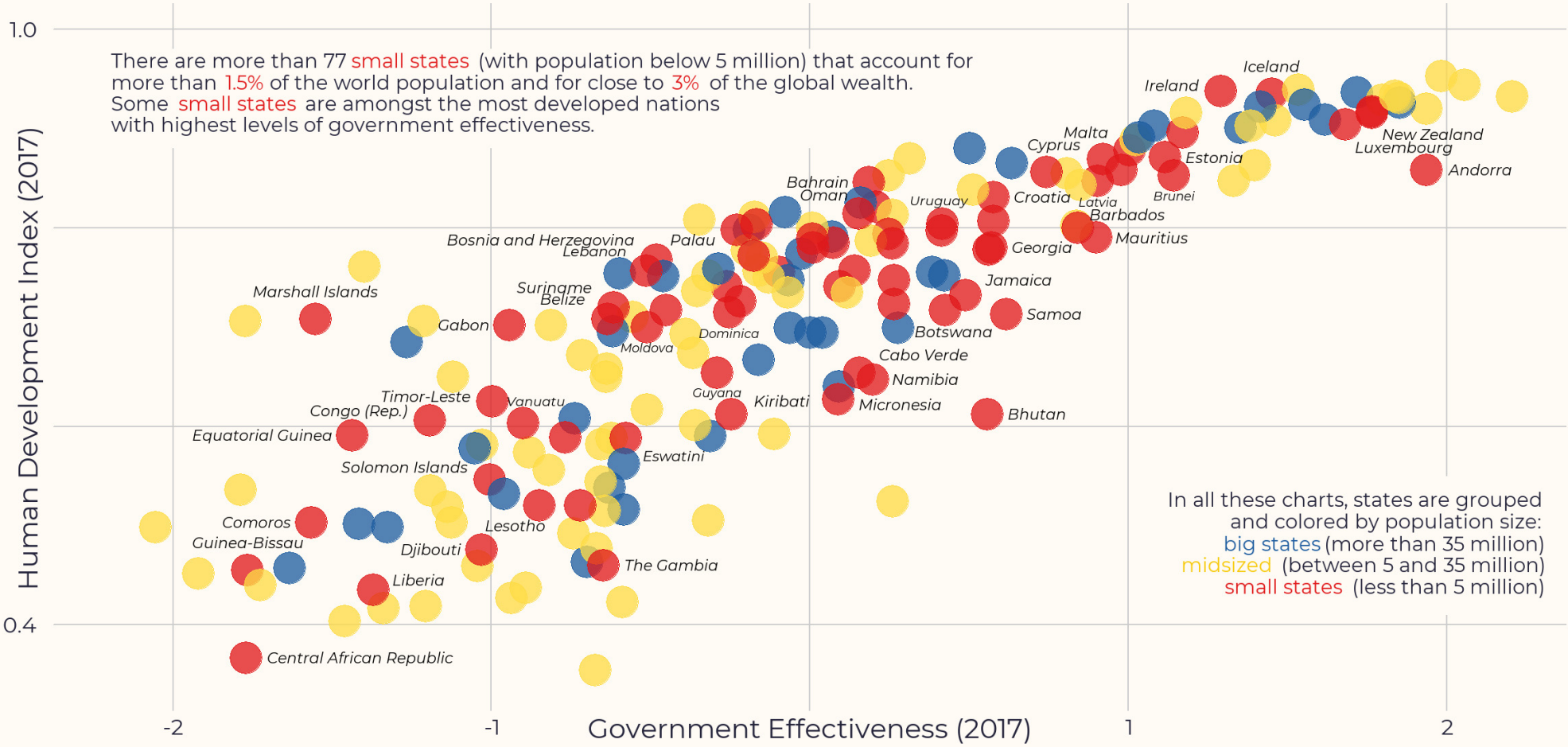

r-spatial/sf: Simple Features fo
github.com/r-spatial/sf
README.md
tic passing tic-db passing coverage 83% license GPL (>= 2) CRAN 1.0-5 CRAN NOTE downloads 1.1M/month
dependencies 6/9
Simple Features for R
A package that provides simple features access for R. Package sf:
represents simple features as 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 PRØJ 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

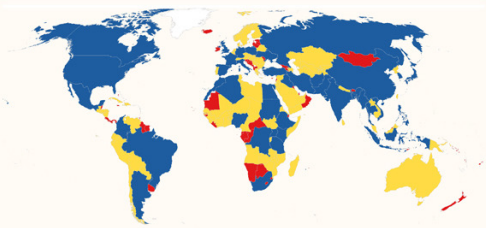
- What is R?
- R is used primarily through interactive command-line.
- R can create and use different types of data but works well with a linear collection of things (vector) and tables.
- R is extremely extendable through packages.
- R is brilliant when it comes to static graphics and dealing with spatial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

Small States Can Be Big Players in Development and Good Governance

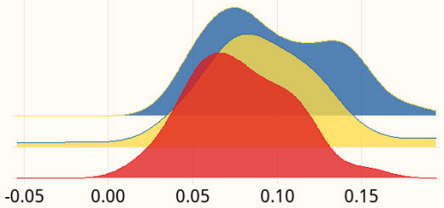
There are more than 77 small states (with population below 5 million) that account for more than 1.5% of the world population and for close to 3% of the global wealth. Some small states are amongst the most developed nations with highest levels of government effectiveness.



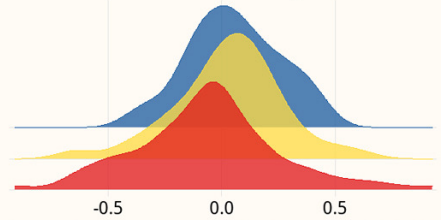
In all these charts, states are grouped and colored by population size:
big states (more than 35 million)
midsized (between 5 and 35 million)
small states (less than 5 million)



But small states have improved less

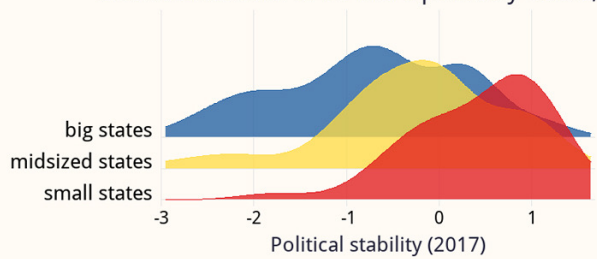


and have even lost some ground.

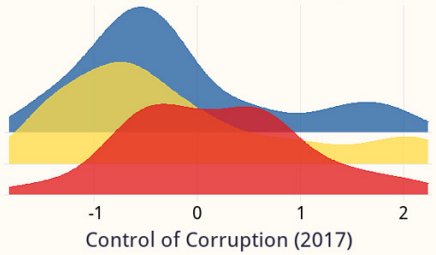


These charts show the distribution densities of different variables by three groups based on state size:

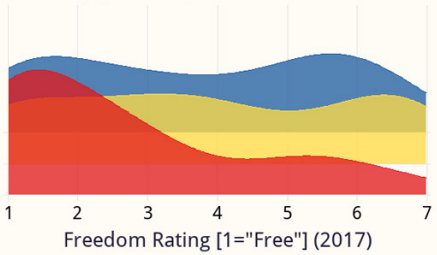
Small states tend to be more politicaly stable,



and have lower levels of corruption,



and enjoy more political freedoms.

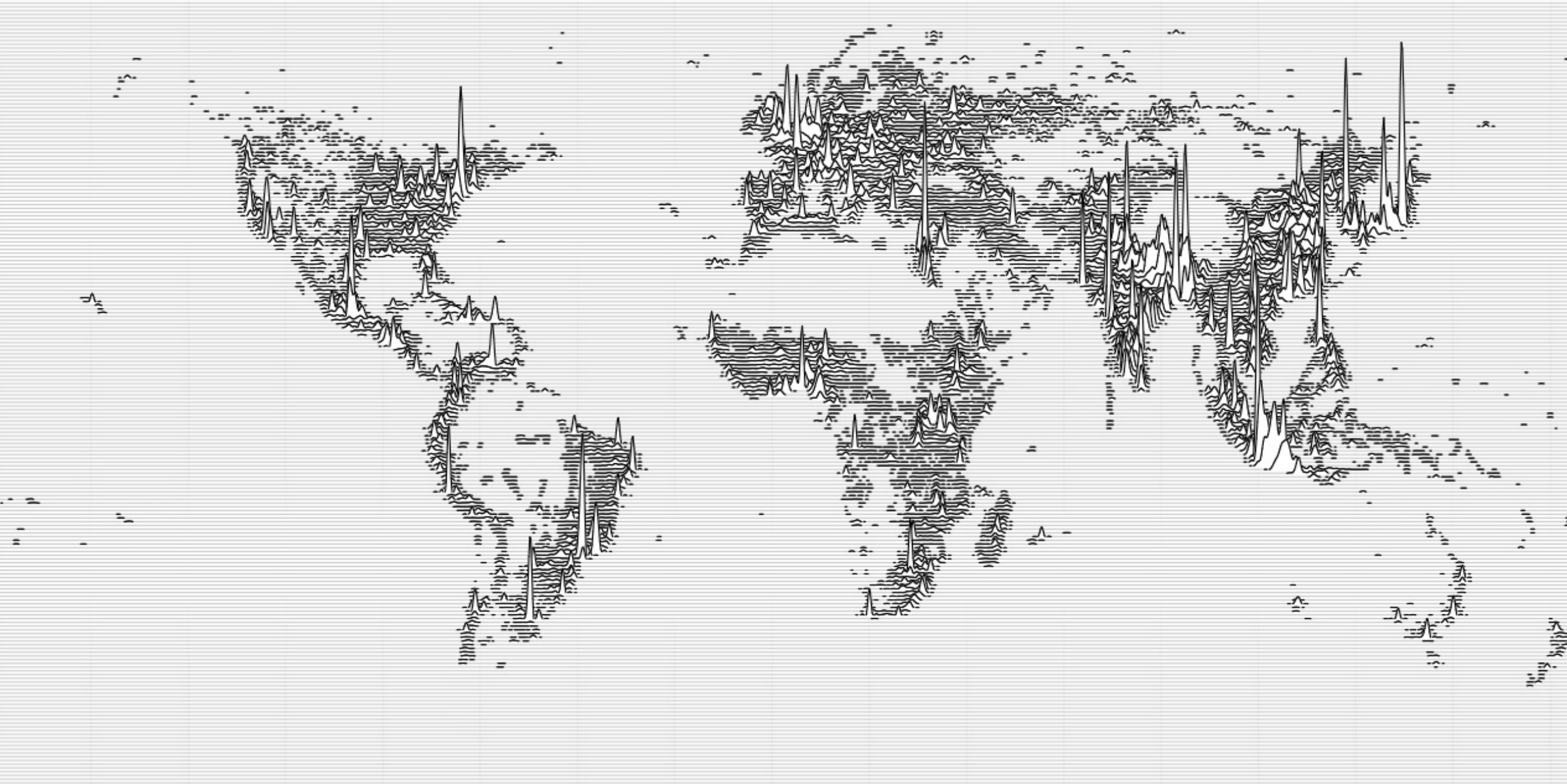


الجائزة العالمية لفن عرض للبيانات
WORLD DATA VISUALIZATION PRIZE

Data: Human Development Index from Human Development Reports. Government Effectiveness, Control of Corruption, and Political Stability from Worldwide Governance Indicators. Freedom Rating from Freedom House.

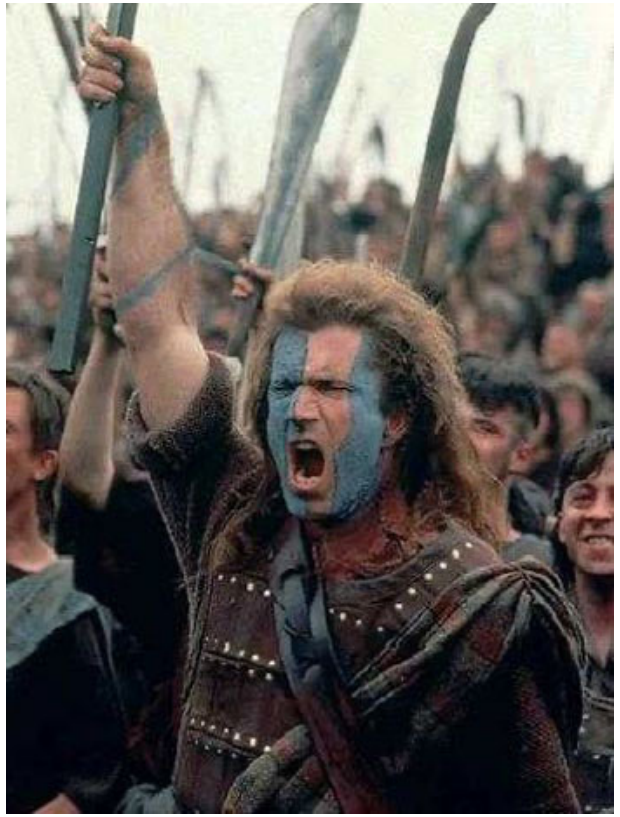
London Cycle Hire Journeys

Thicker, yellower lines mean more journeys



5 more reasons on why you should use R

R is free as in

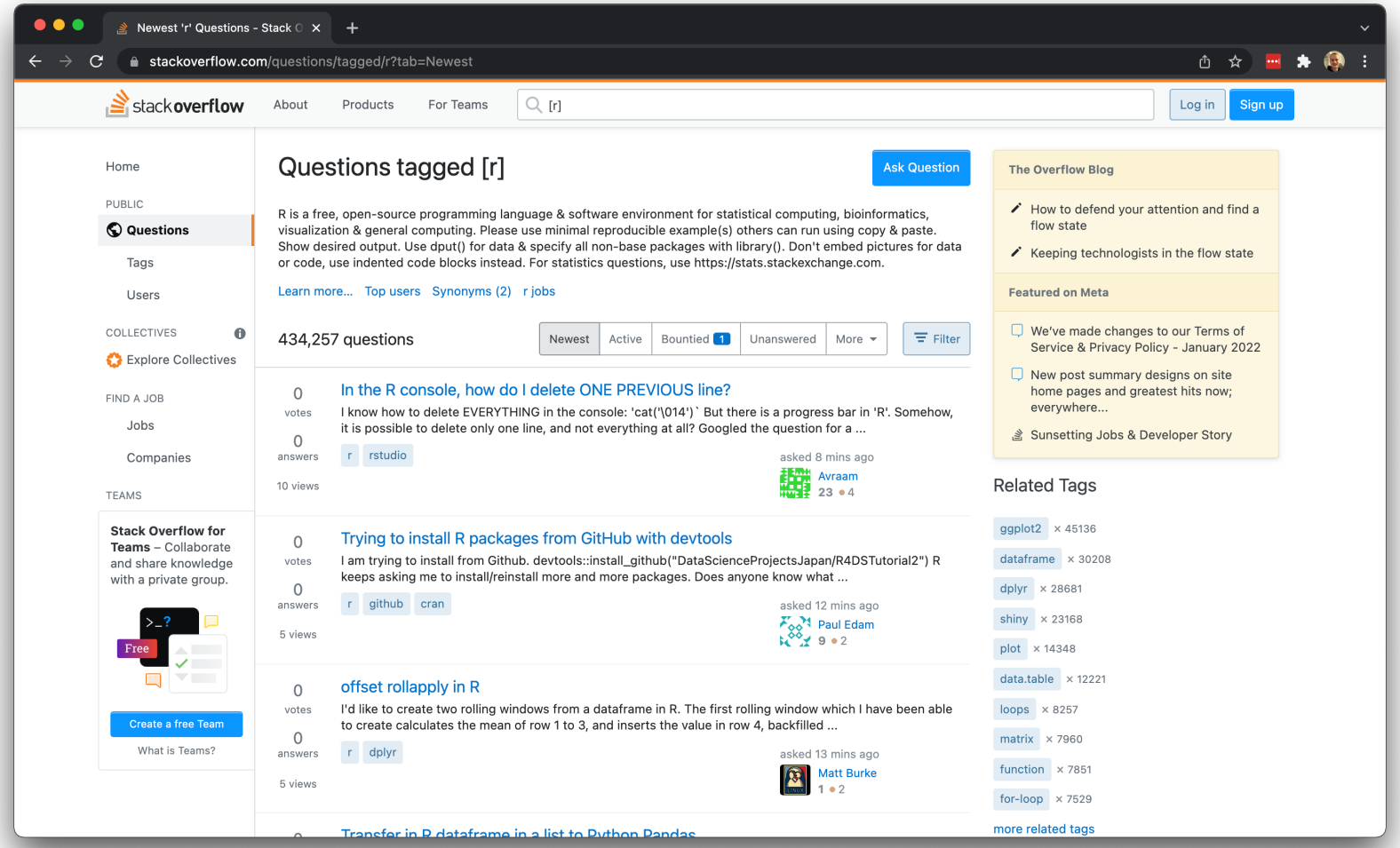


Freedom

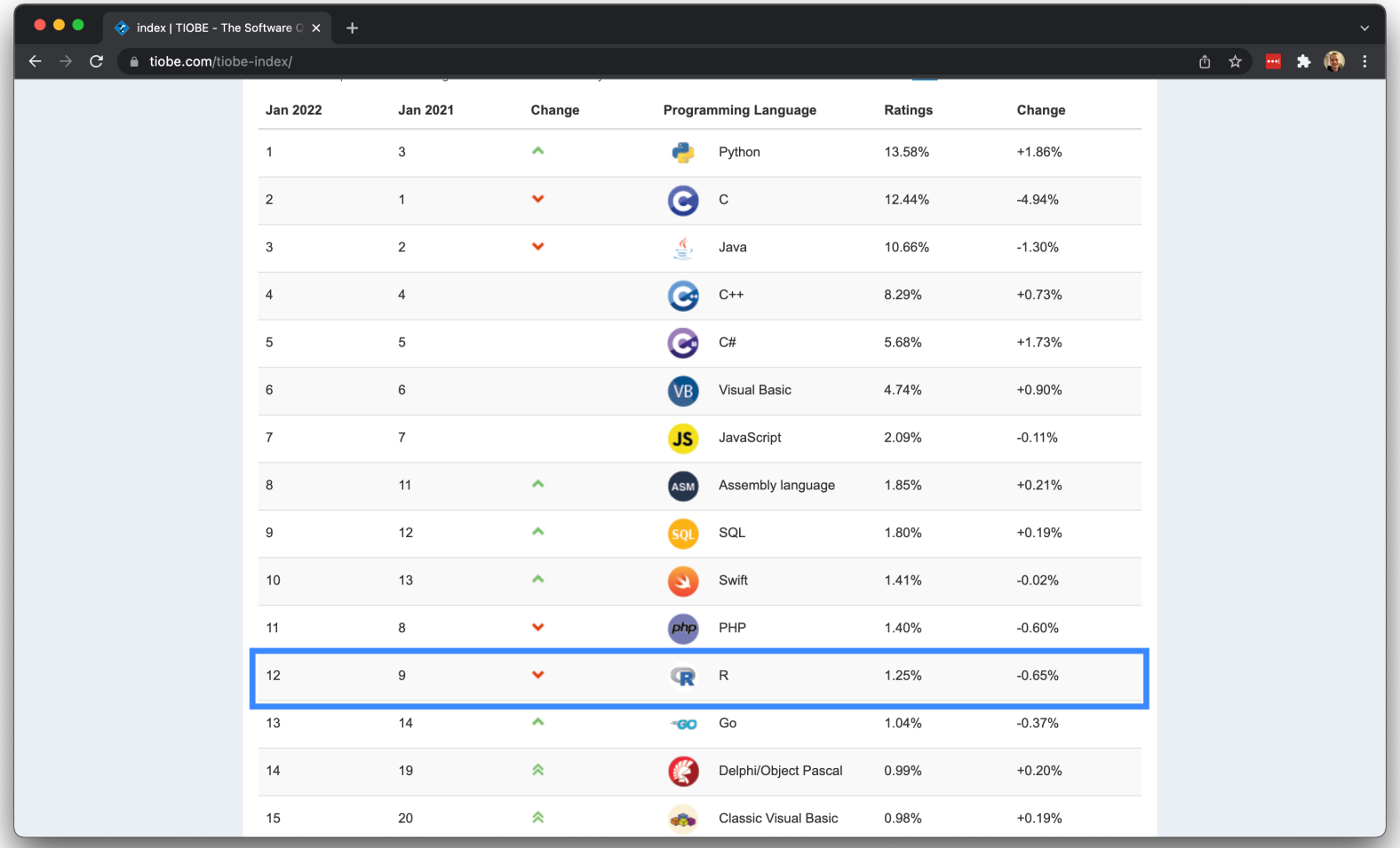
Free beer

R allows you to produce your outputs programmatically

R is supported by a large vibrant community



R is popular and in-demand



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

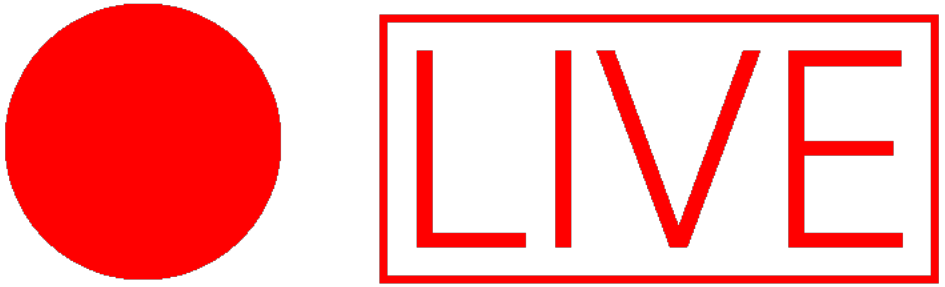
Python?


YES, I WOULD LOVE TO SEE ANOTHER
R VS PYTHON FOR DATA SCIENCE COMPARISON
makeameme.org

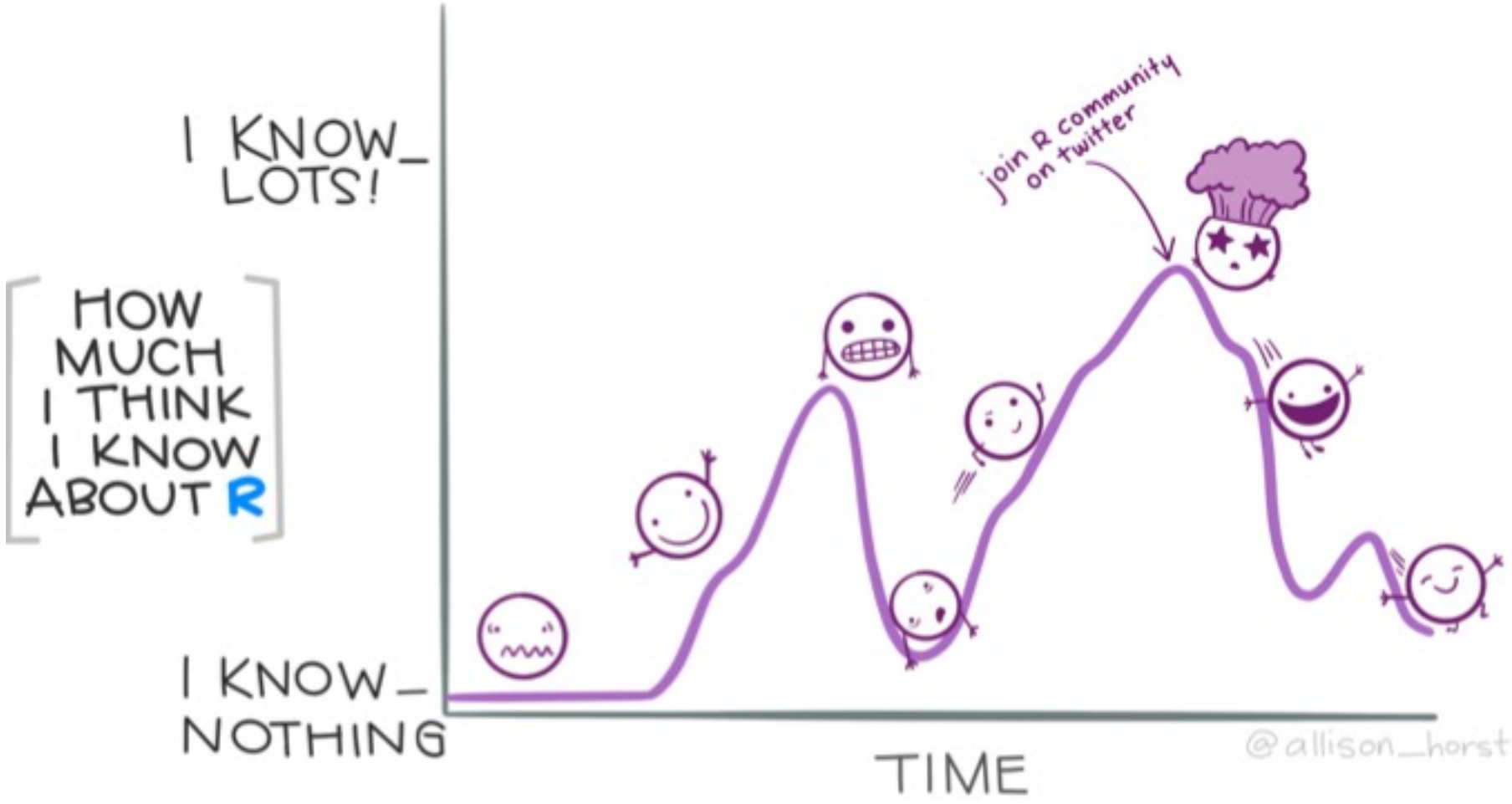
R Markdown

- 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



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

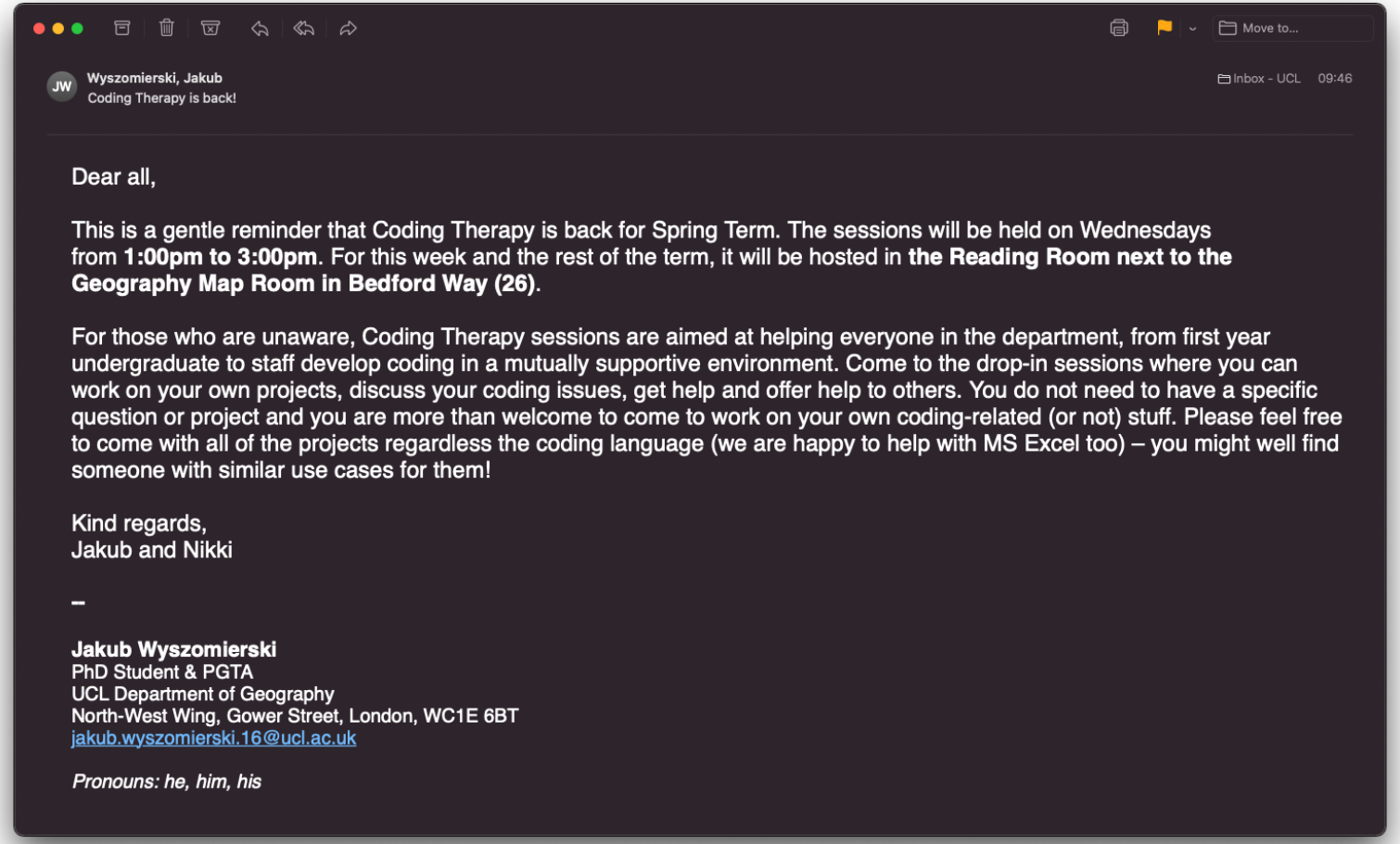
Computer tutorial

- 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



Move to...

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