SA-TIED Geospatial Analysis Workshop

Overview

Dr Justin van Dijk

j.t.vandijk@ucl.ac.uk

Objectives

We will explore the following topics:

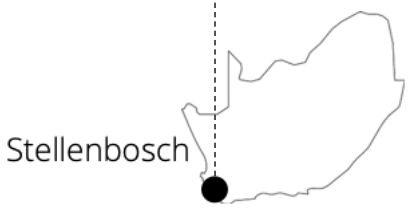
- Fundamentals of using R for data analysis.
- Creating thematic maps using R.
- Quantifying the degree of spatial dependence in a dataset.
- Incorporating space into statistical models.

Schedule

Day 1 – Morning	R for Data Analysis
Day 1 – Afternoon	R for Spatial Analysis
Day 2 – Morning	Spatial Autocorrelation
Day 2 – Afternoon	Spatial Models

Background

B.Sc. Human Geography and Planning
M.Sc.. Human Geography and Planning



Ph.D. Transport Economics

Lecturer in Social and Geographic Data Science

SA-TIED Geospatial Analysis Workshop

S01 – R for Data Analysis

This session

- What is a programming language?
- A gentle introduction to working with R.
- Why use R for data analysis?

Programming languages

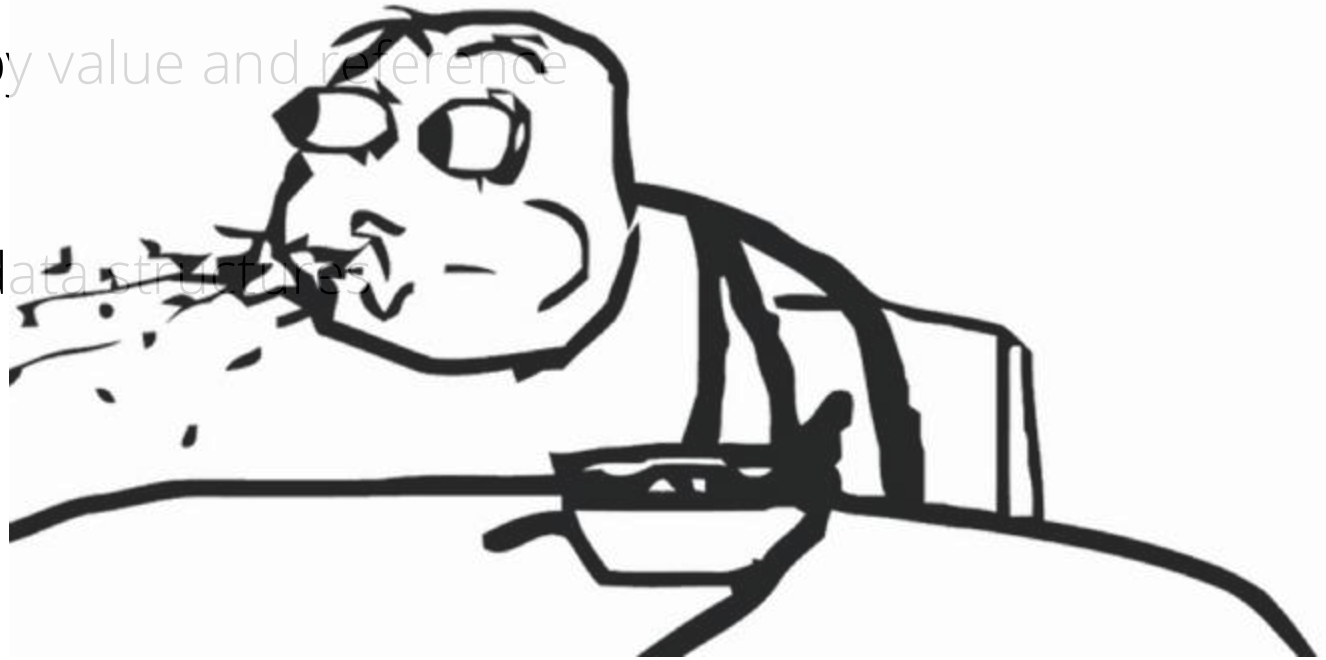
- Consist of a formal set of instructions that you can use to write software or perform computational tasks.
- Require users to write code, which involves typing commands in a text-based environment.
- Are highly flexible and powerful, allowing for custom solutions, automation, and complex operations.

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

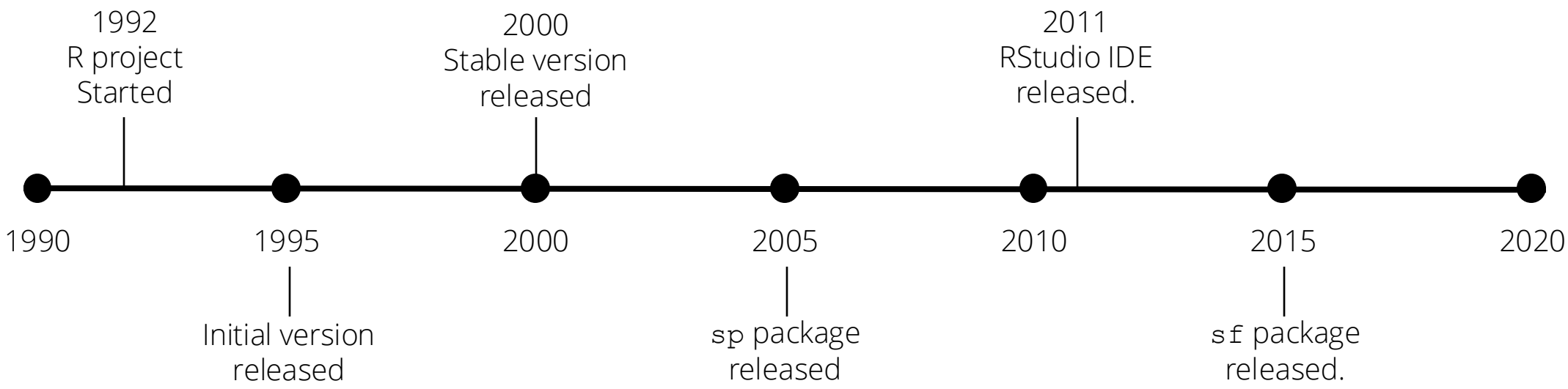
- Ιδεντιφιερς ανδ πριμιτιfε δατα τψπες
- Ασσιγνμεντ, αριτημετιξ, λογιξαλ ανδ ρελατιοναλ οπερατορς
- Εχπρεσσιον ανδ στατεμεντς, δεβυγγινγ
- Φλοω οφ ξοντρολ· σελεξτιον ανδ ρεπετιτιον
- Φυνξτιονς, παραμετερς πασσινγ, ξαλλ βψ αfλυε ανδ ρεφερενξε
- Οβσεξτ-οριεντεδ προγραμμινγ
- 1/2 διμενσιοναλ αρραψς, στρινγς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

The absolute basics

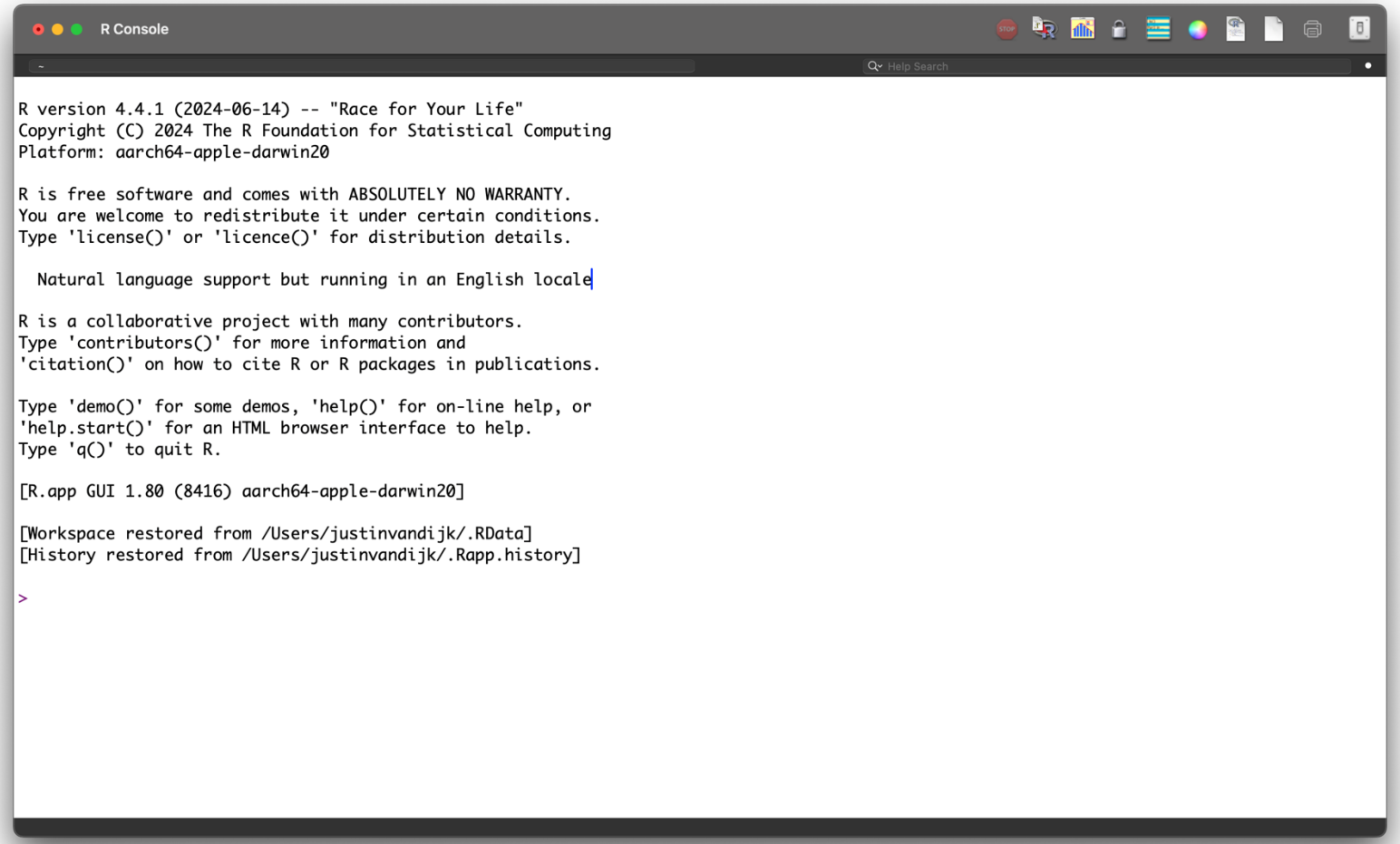
- R is primarily used through interactive command-line interfaces and scripts.
- R is the programming language, but we can interact with it using other software.
- R efficiently handles various data types, particularly vectors and tables.
- R's functionality can be extended through a vast ecosystem of packages.
- R is free and open source.

A little history

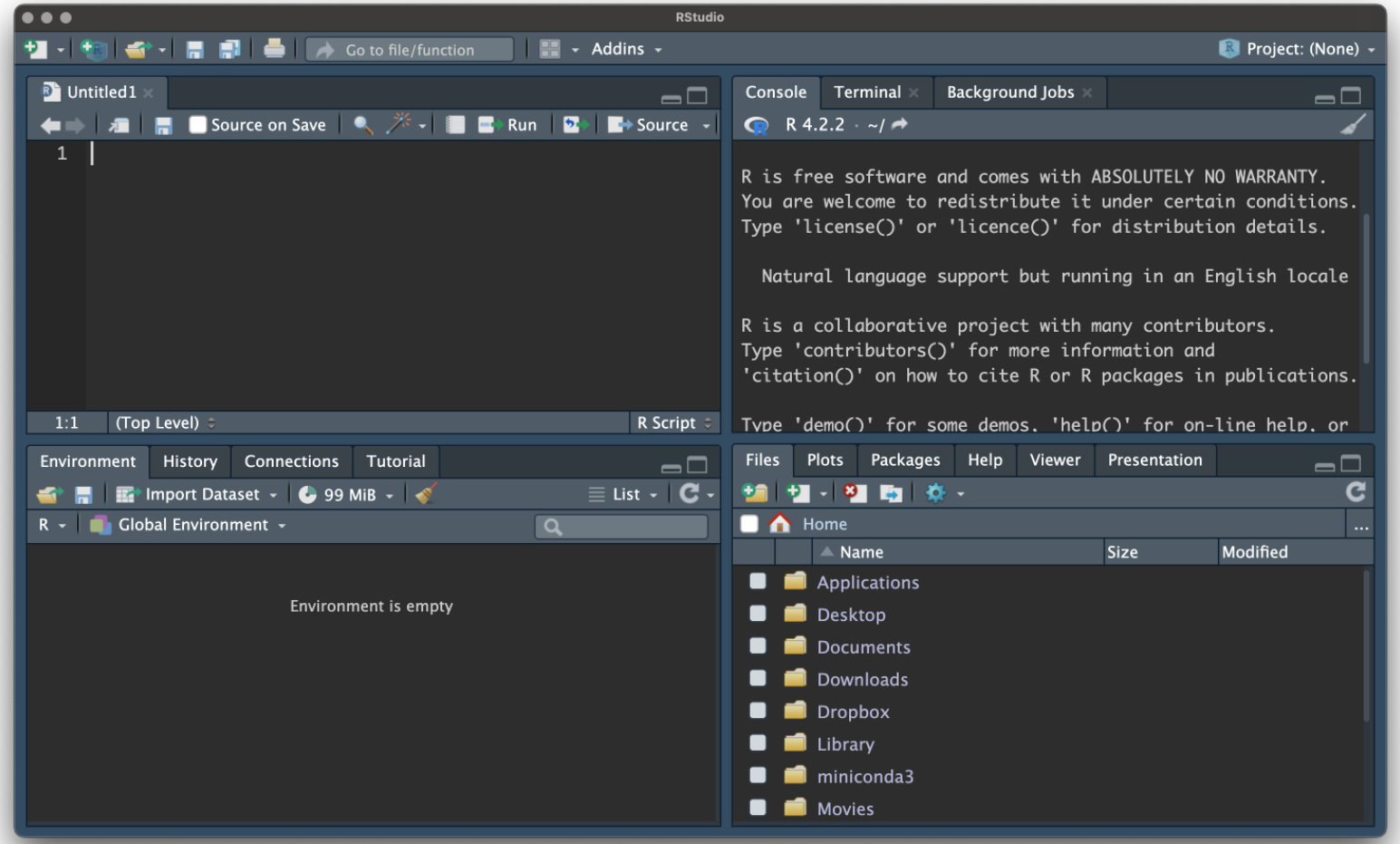
- R is programming language originally developed for statistical purposes.
- 1975: Bell Labs develops a language for Statistical Analysis (“S”).
- 1992: Ross Ihaka and Robert Gentleman develop opensource version of “S”.



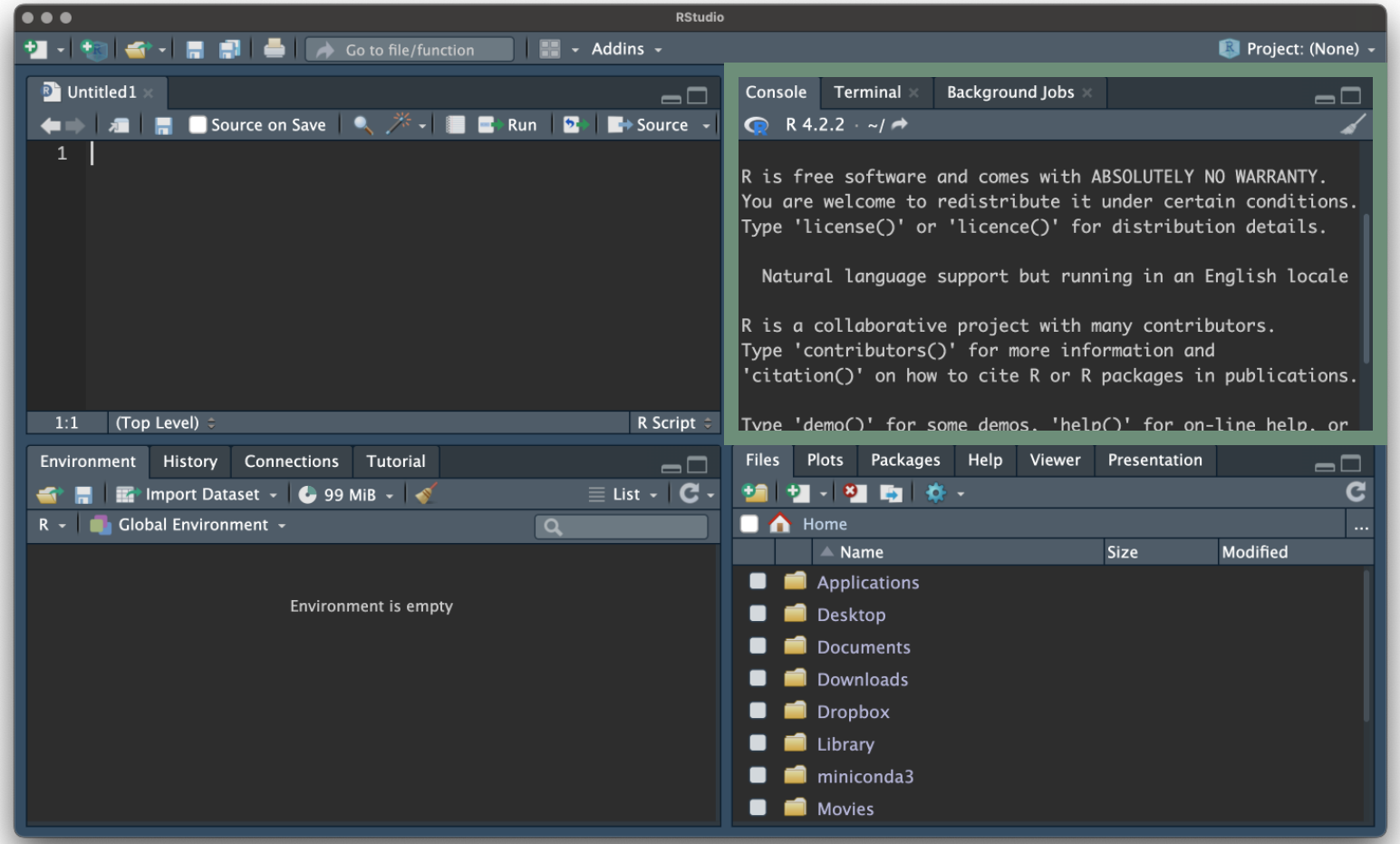
Working with R



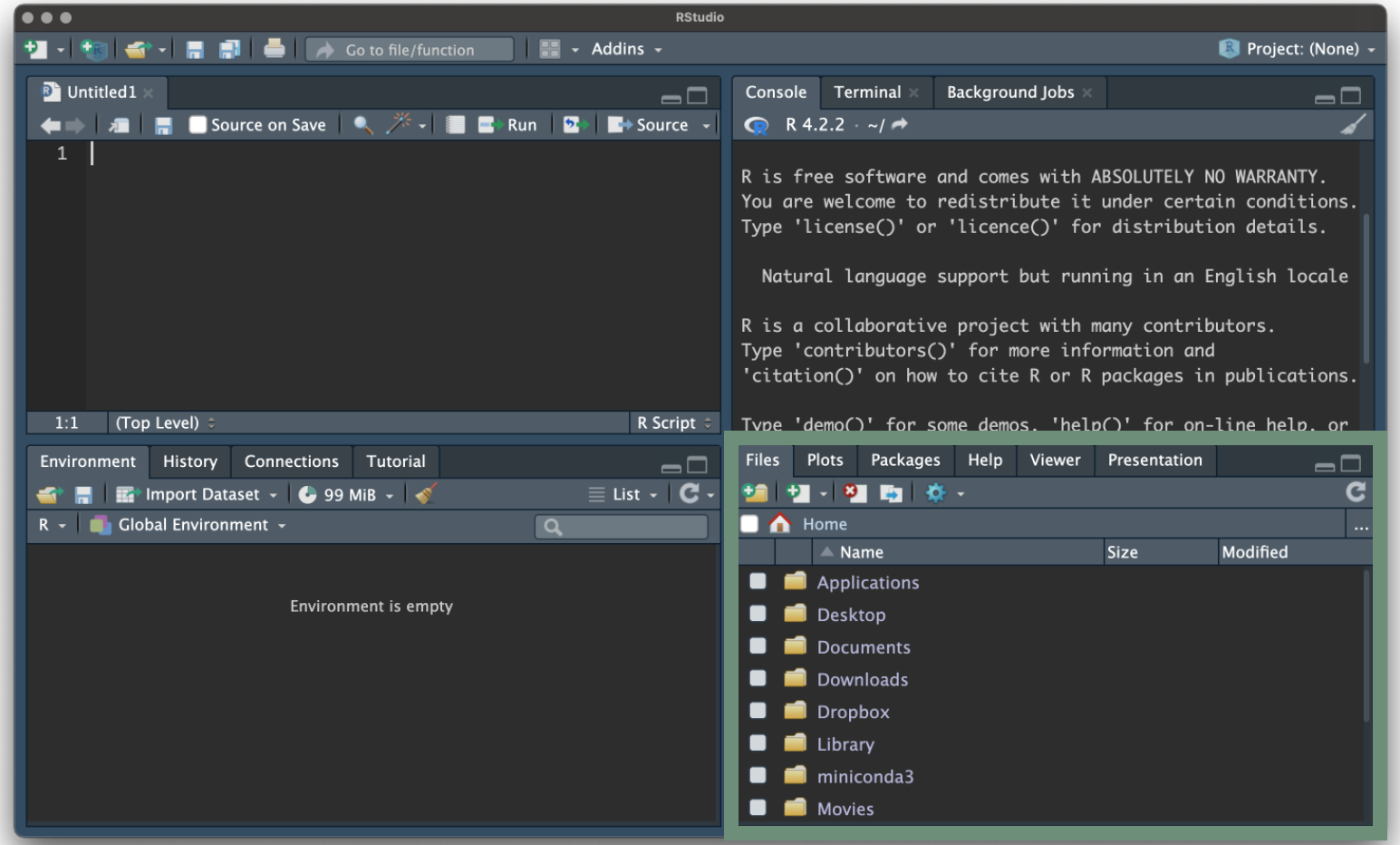
Working with R



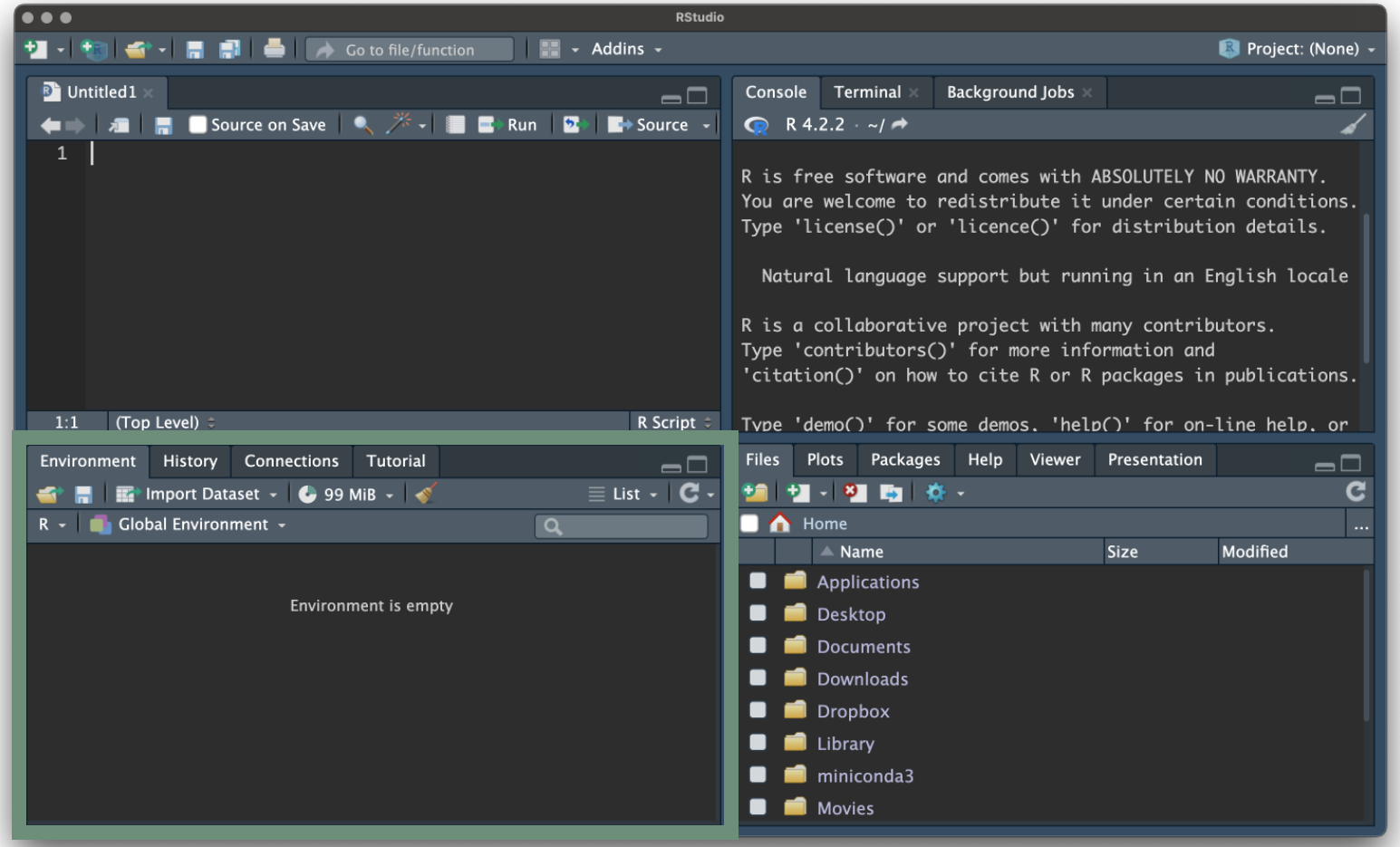
Working with R



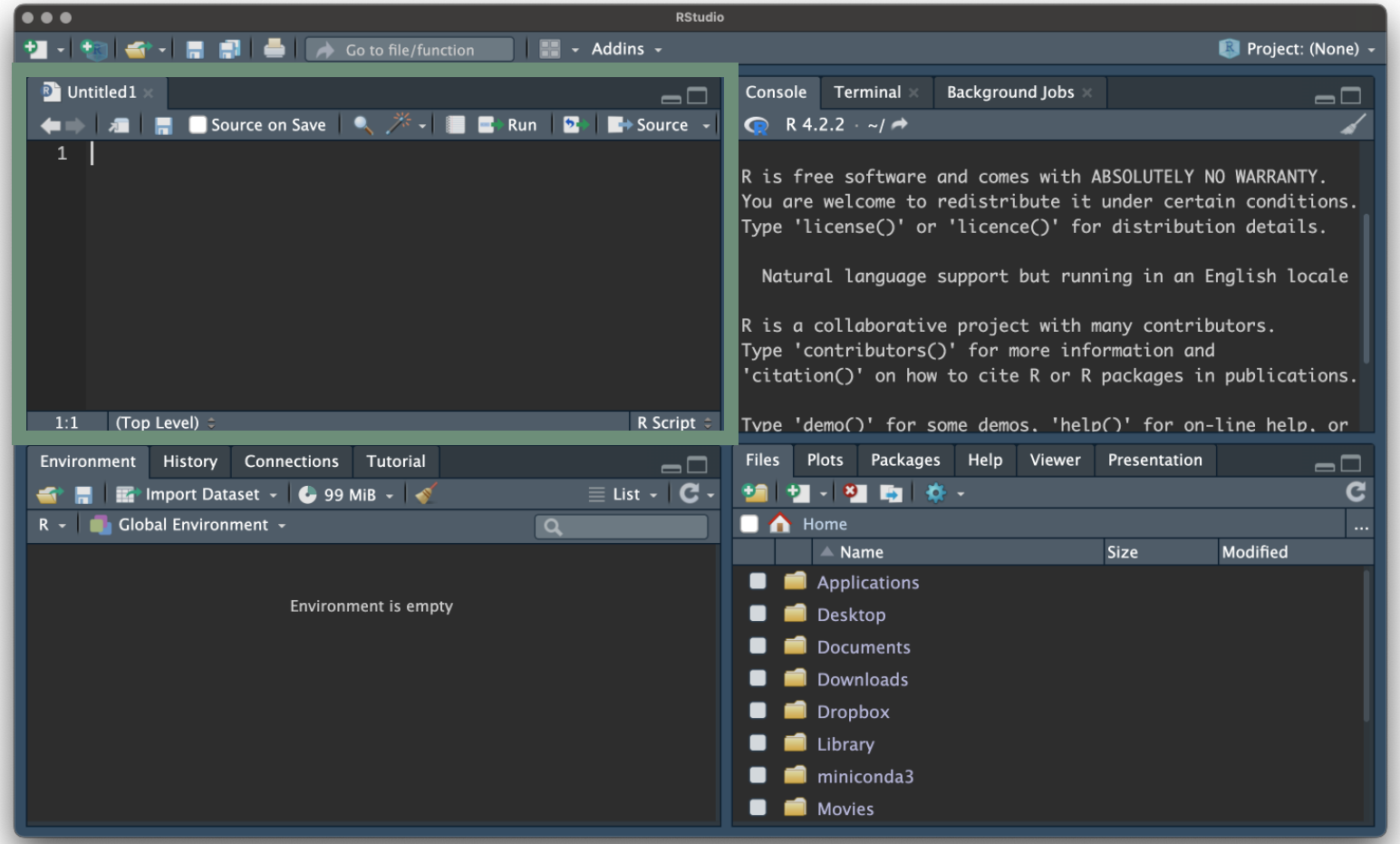
Working with R



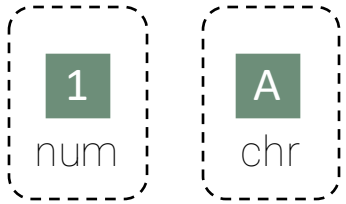
Working with R



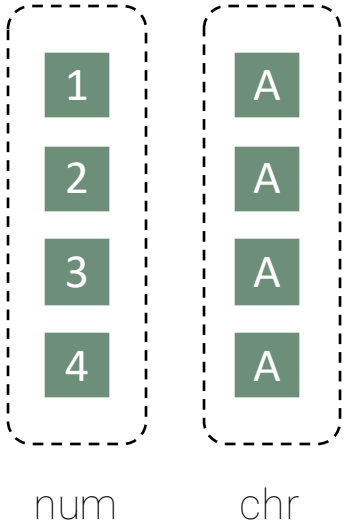
Working with R



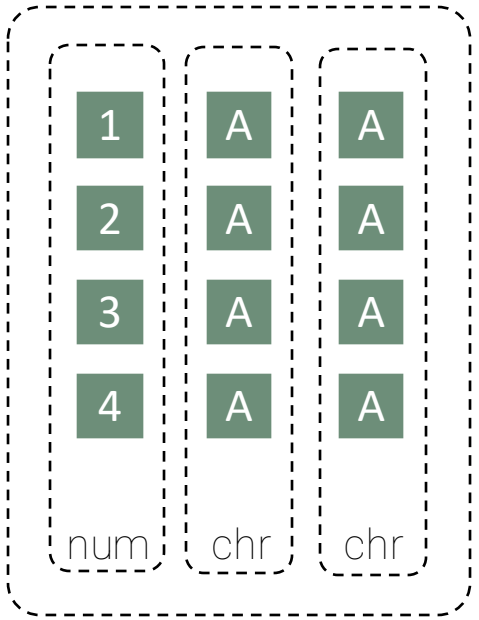
Principles of R: Data types



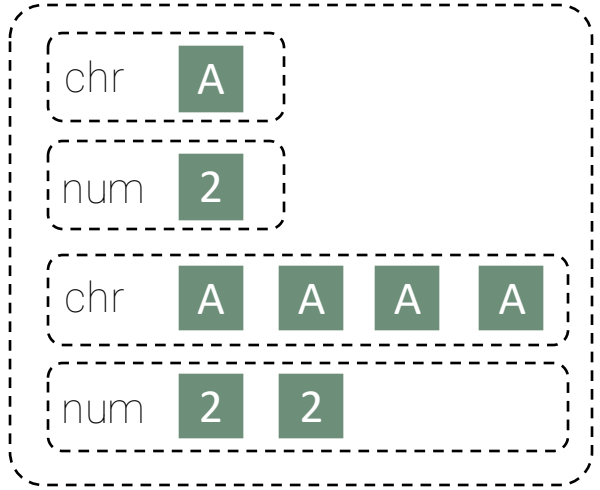
Scalar



Vector



Dataframe



List

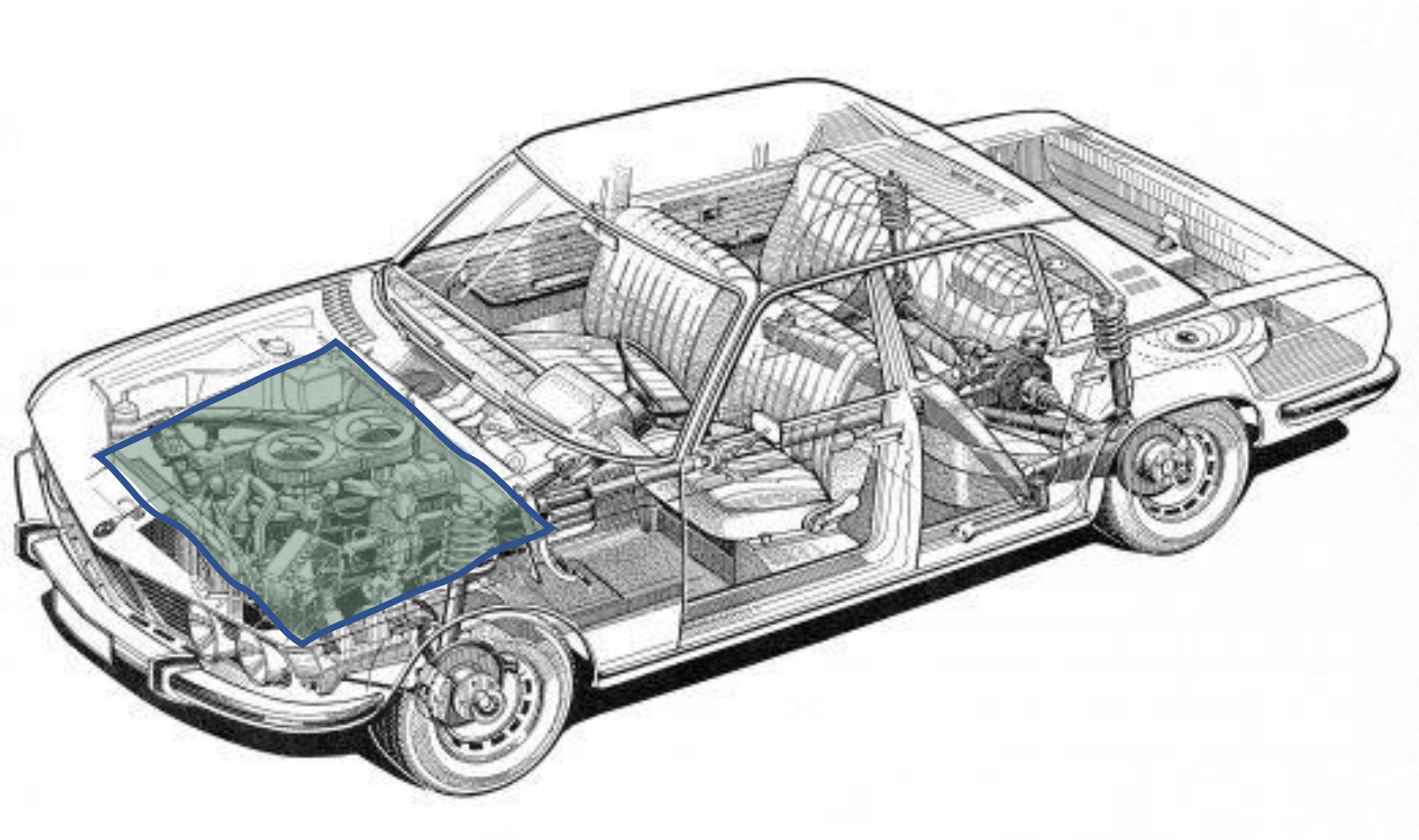
Principles of R: Variables

- 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.

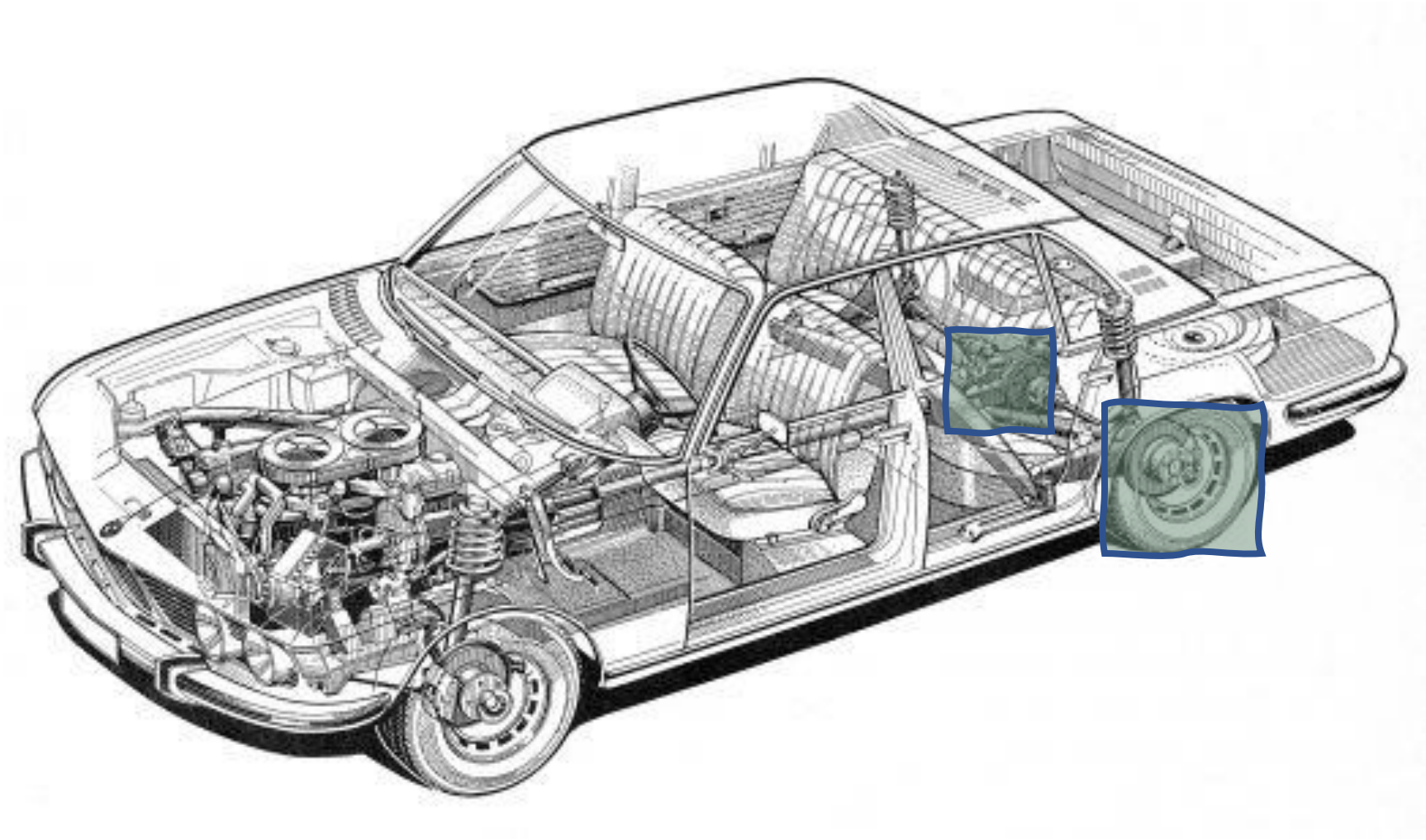
Principles of R: Functions

- Variables can be used as an input for functions.
- Functions are pieces of code designed to accomplish specific tasks.
- Once a function is written, it can be reused.

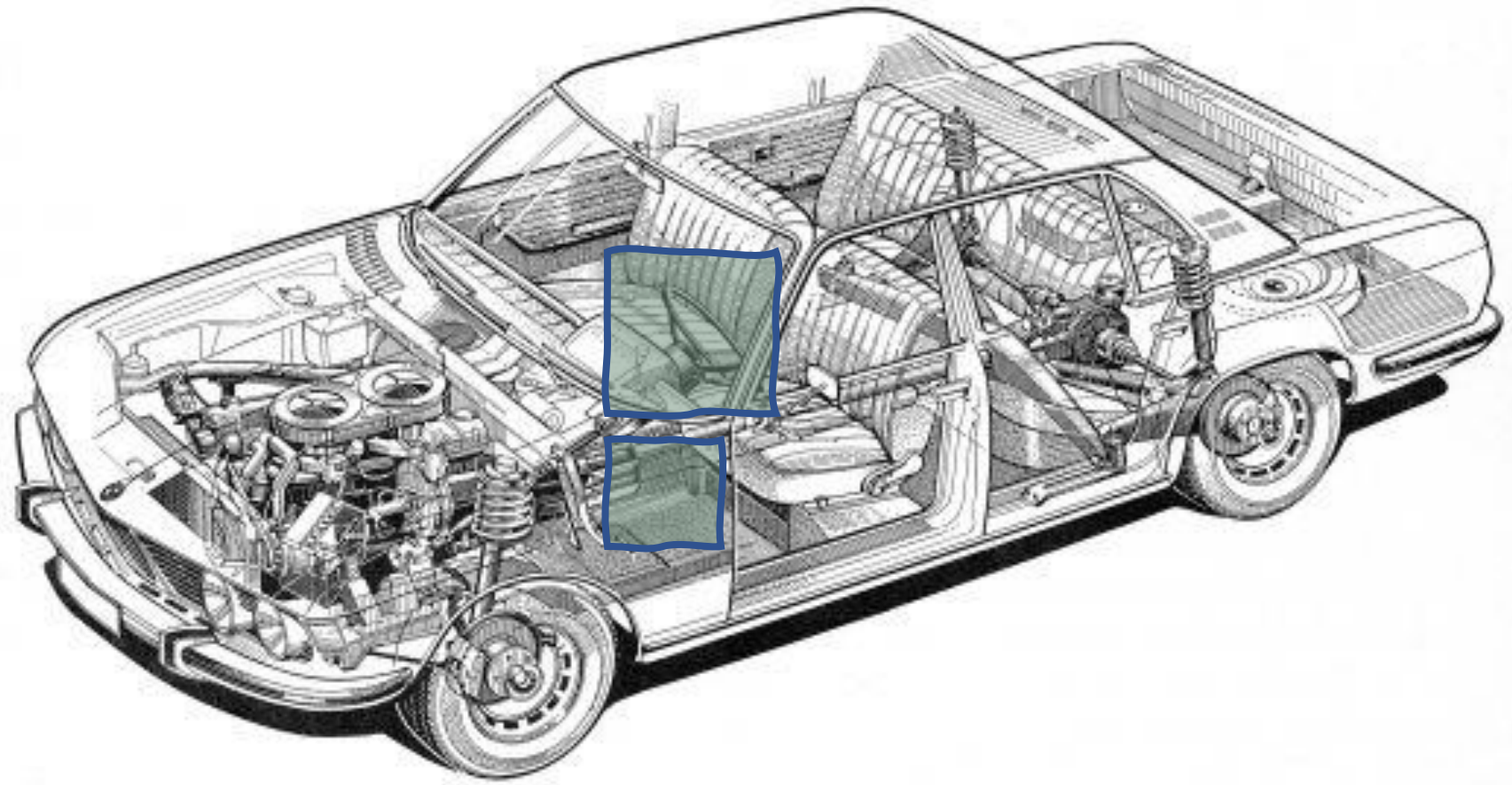
Principles of R: Extensions



Principles of R: Extensions



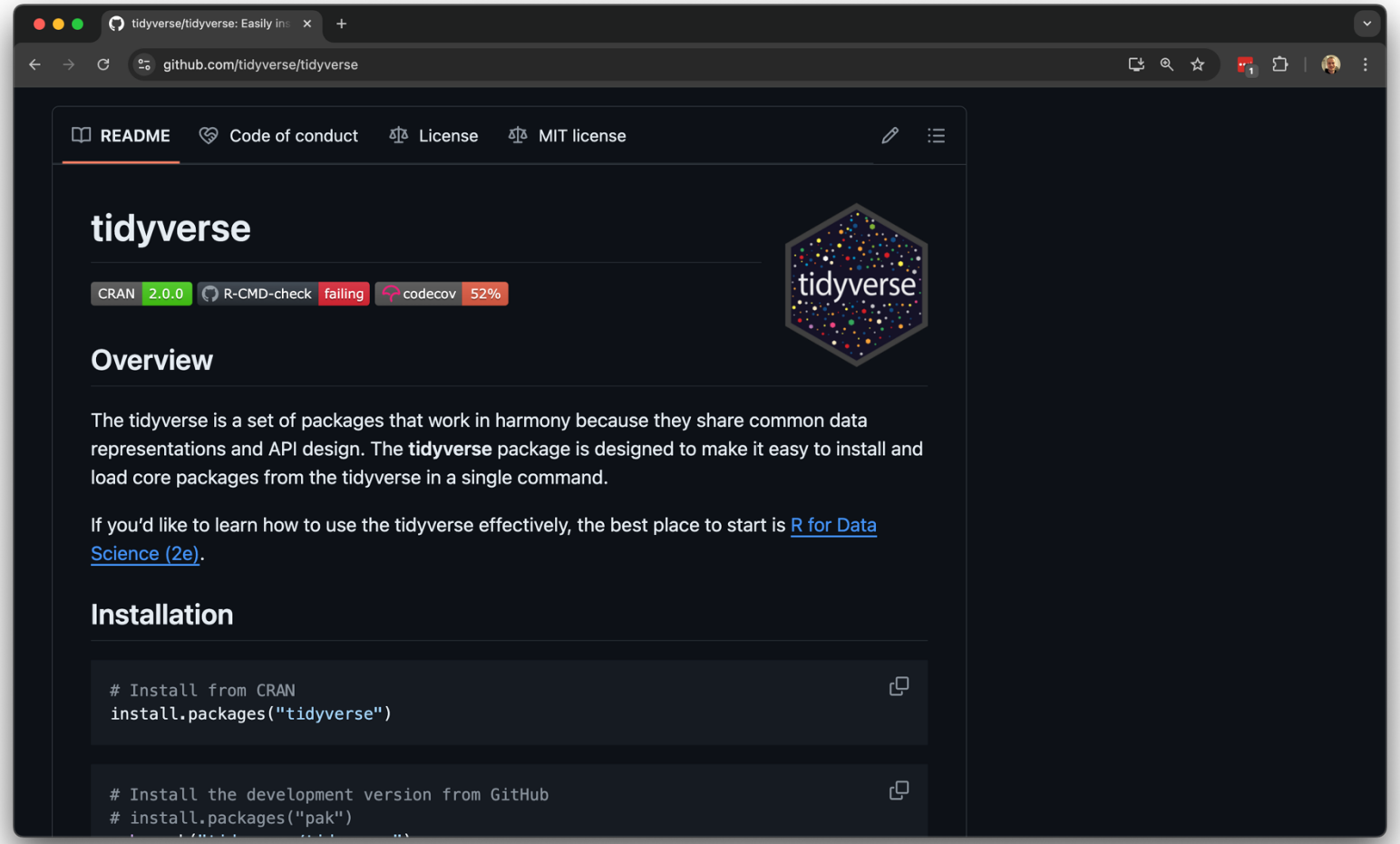
Principles of R: Extensions



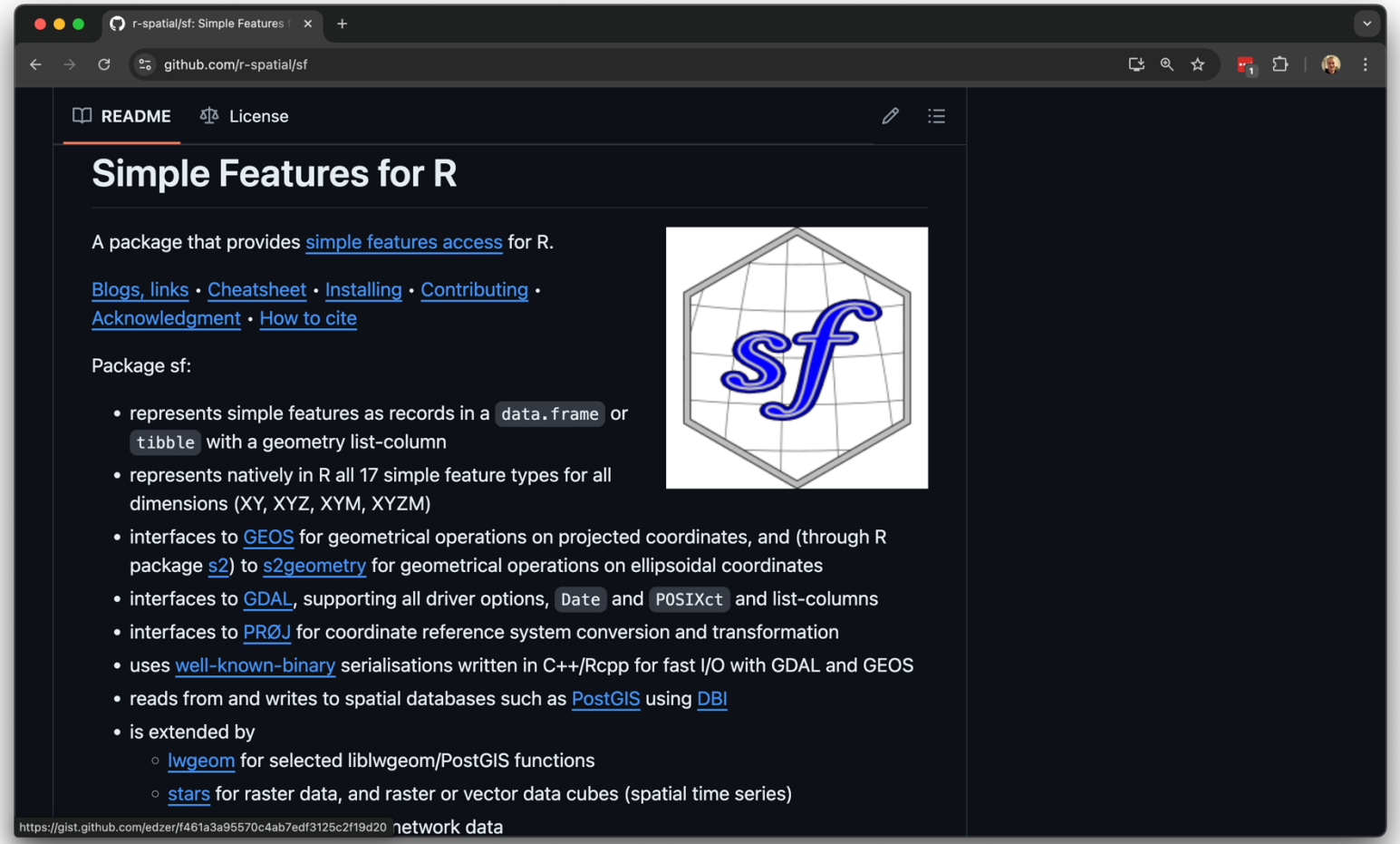
Principles of R: Extensions

- Available on The Comprehensive R Archive Network: CRAN
- CRAN is a repository for R packages and software.
- Currently CRAN features over 21,140 available packages (22/08/2024).
- Allows developers to submit and maintain their own R packages.

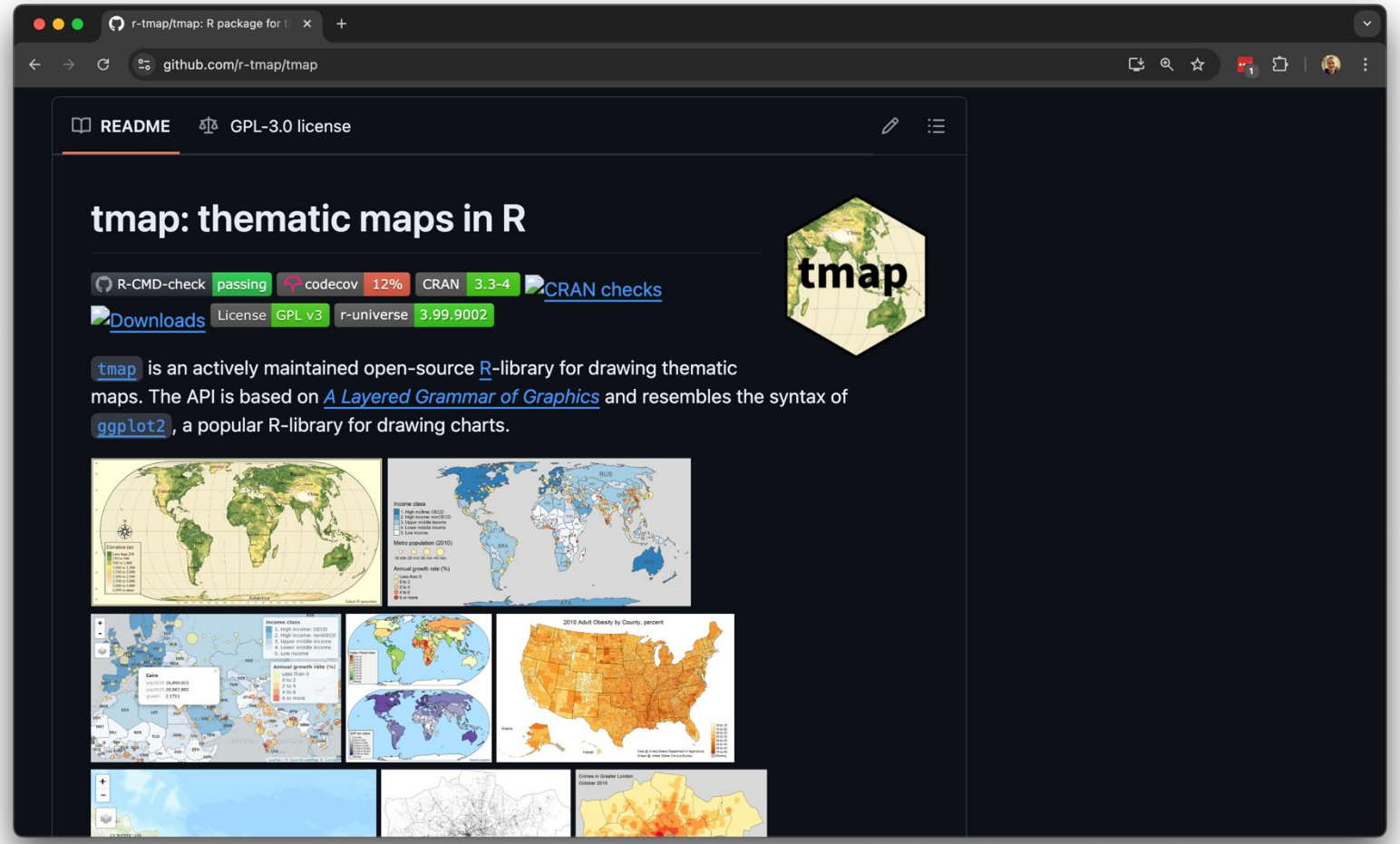
tidyverse


tidyverse/tidyverse: Easily ins
github.com/tidyverse/tidyverse
README
Code of conduct
License
MIT license
tidyverse
CRAN 2.0.0
R-CMD-check failing
codecov 52%
tidyverse
Overview
The tidyverse is a set of packages that work in harmony because they share common data representations and API design. The tidyverse package is designed to make it easy to install and load core packages from the tidyverse in a single command.
If you'd like to learn how to use the tidyverse effectively, the best place to start is R for Data Science (2e).
Installation
# Install from CRAN
install.packages("tidyverse")
# Install the development version from GitHub
# install.packages("pak")

Simple Features



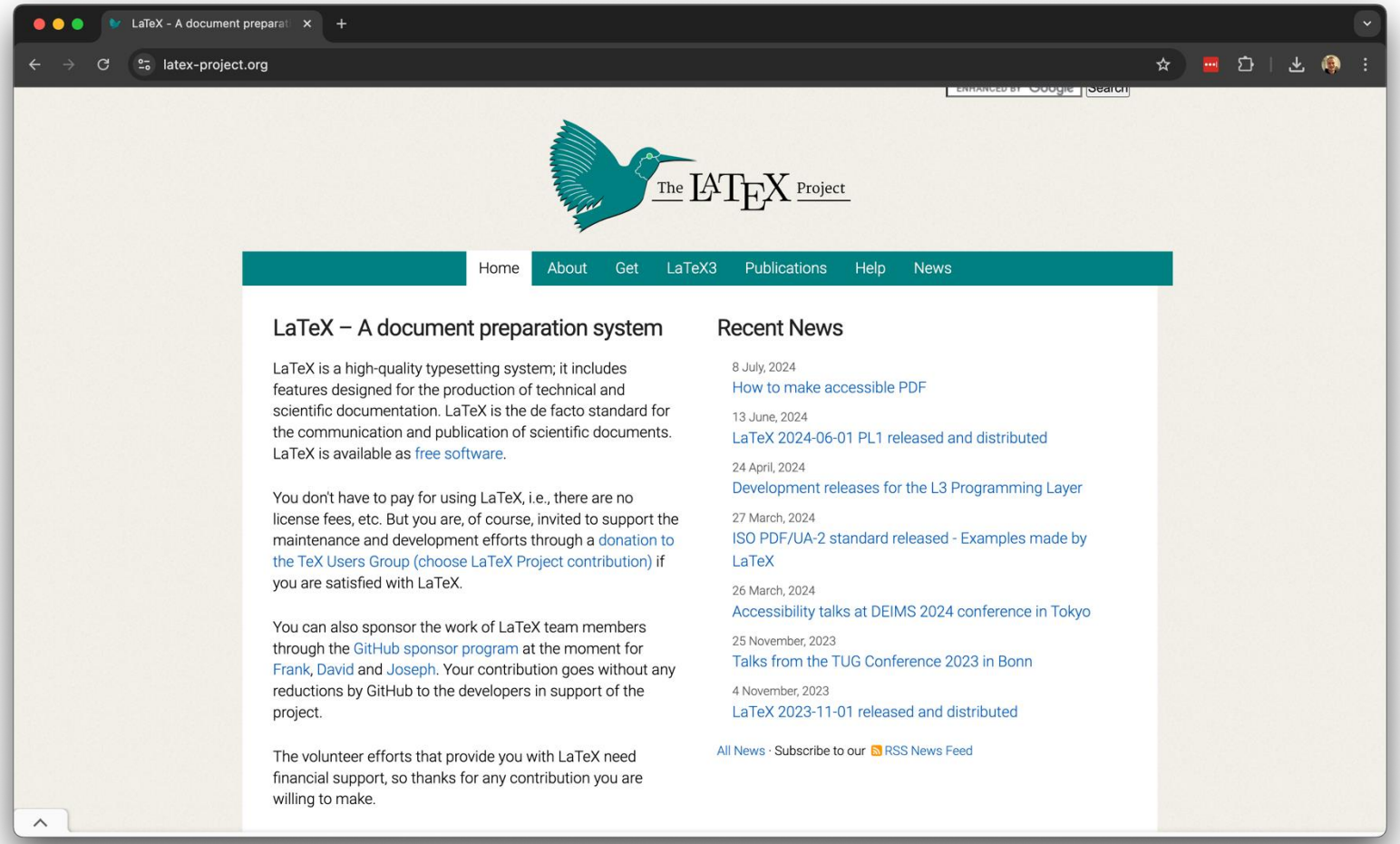
tmap


r-tmap/tmap: R package for t
github.com/r-tmap/tmap
README
GPL-3.0 license
tmap: thematic maps in R
R-CMD-check passing
codecov 12%
CRAN 3.3-4
CRAN checks
Downloads
License GPL v3
r-universe 3.99.9002
tmap
tmap is an actively maintained open-source R-library for drawing thematic maps. The API is based on A Layered Grammar of Graphics and resembles the syntax of ggplot2, a popular R-library for drawing charts.

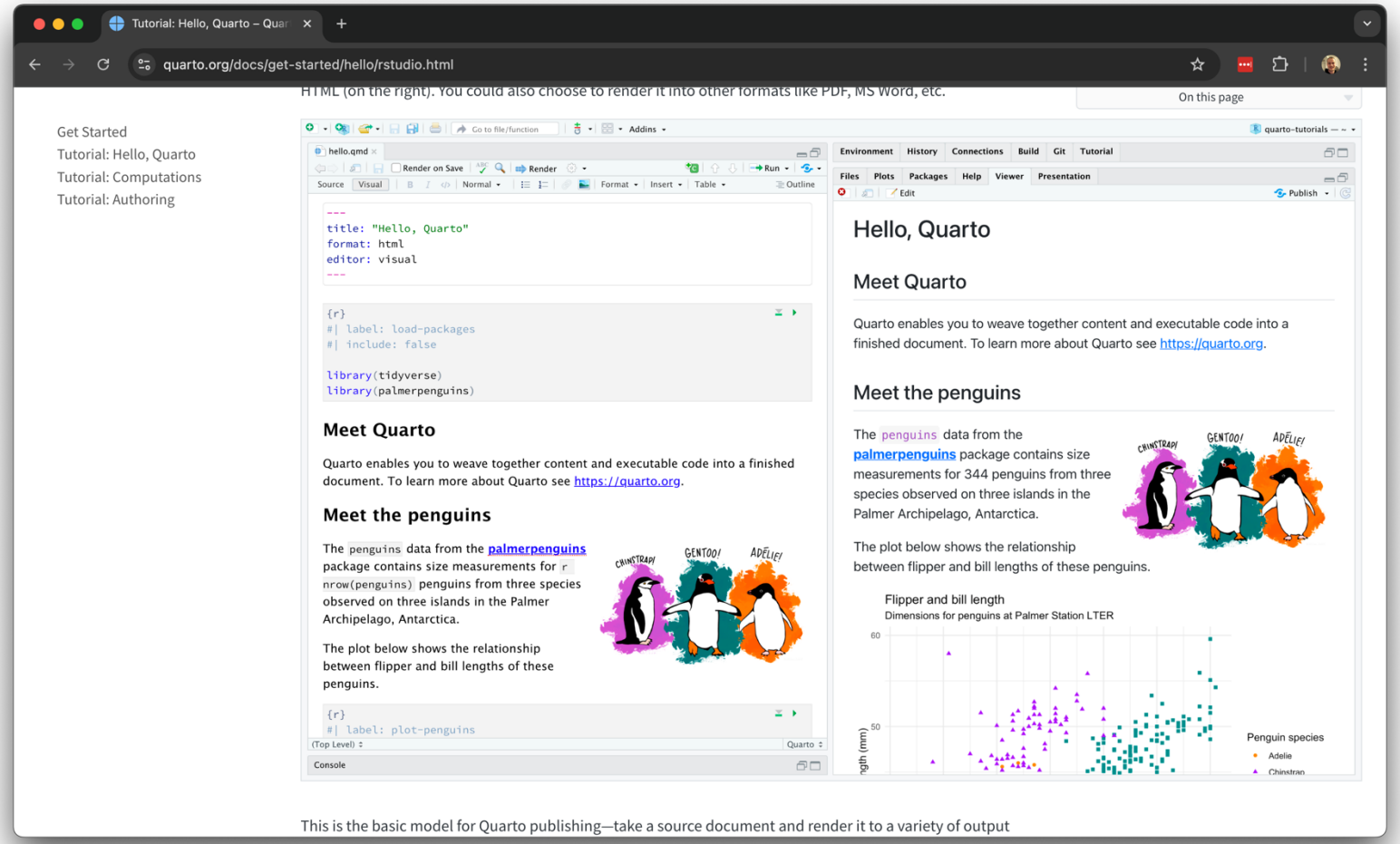
Reproducibility with R Markdown

- Markdown is a lightweight markup language designed for creating formatted text using a plain-text editor. It allows you to write in a simple, readable format that can be easily converted to HTML, PDF, or other formats.
- R Markdown extends Markdown by integrating R code with the text. This allows you to: embed code, execute code, create dynamic reports, format and typeset your content.

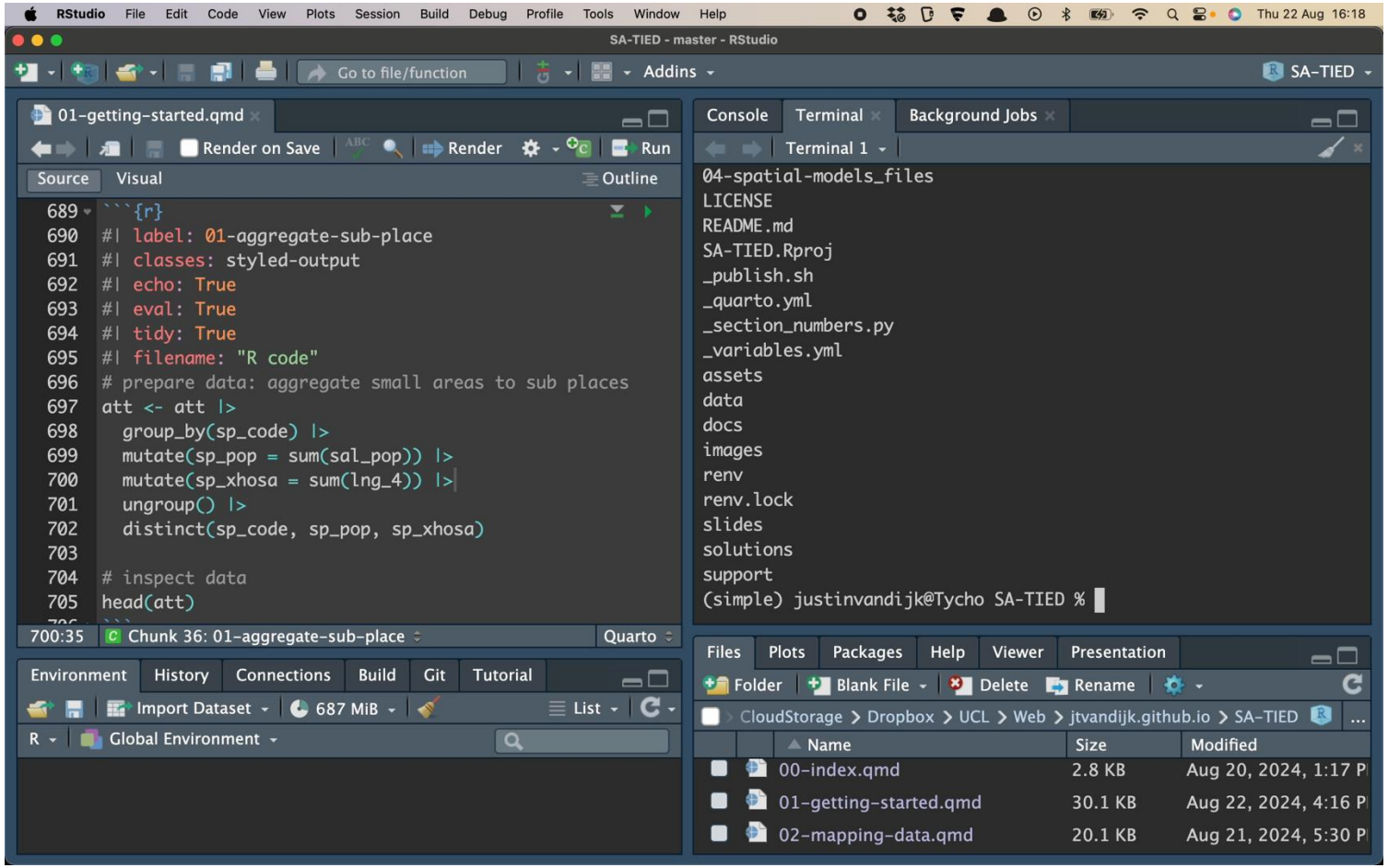
Reproducibility with R Markdown



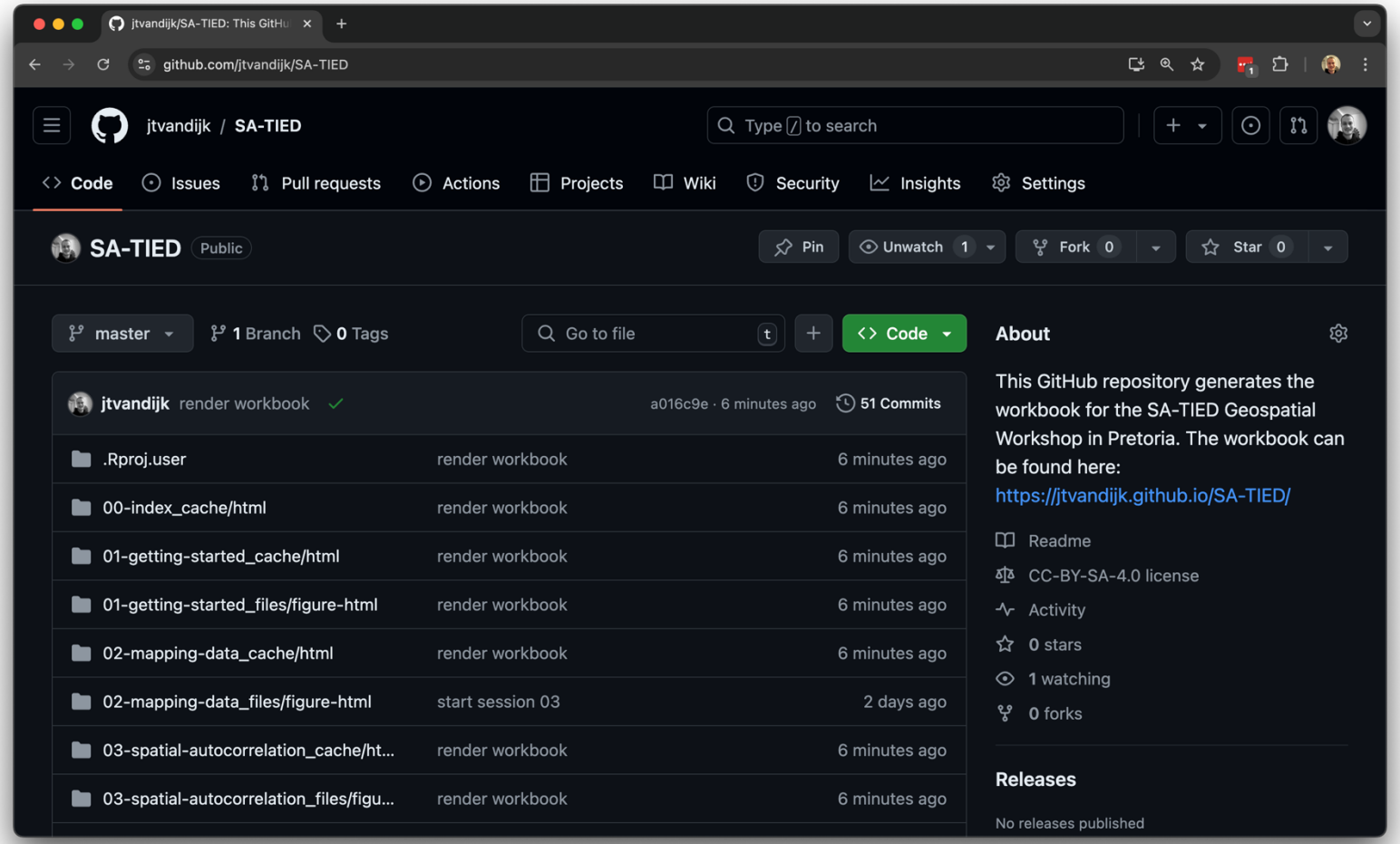
Reproducibility with R Markdown



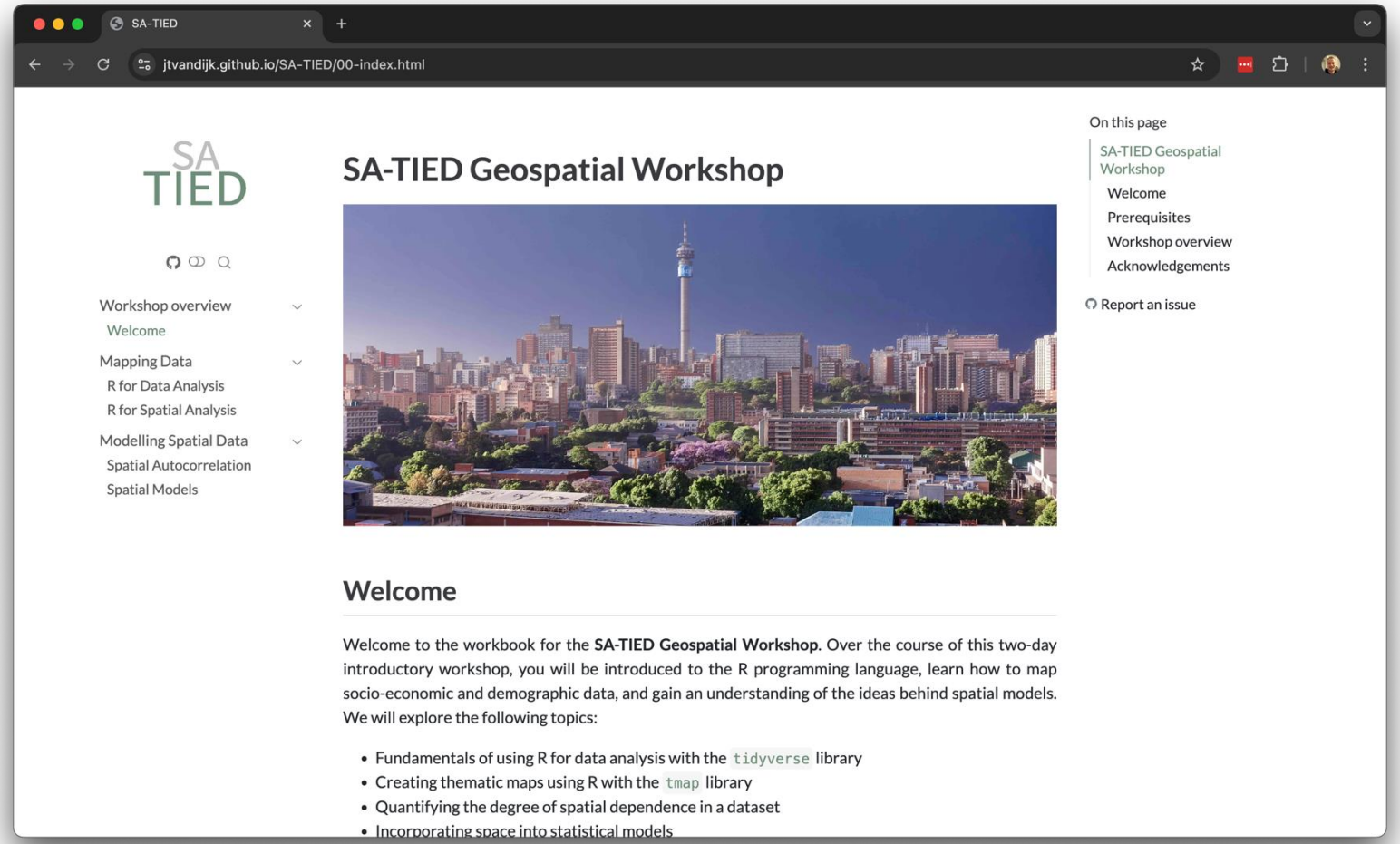
Reproducibility with R Markdown



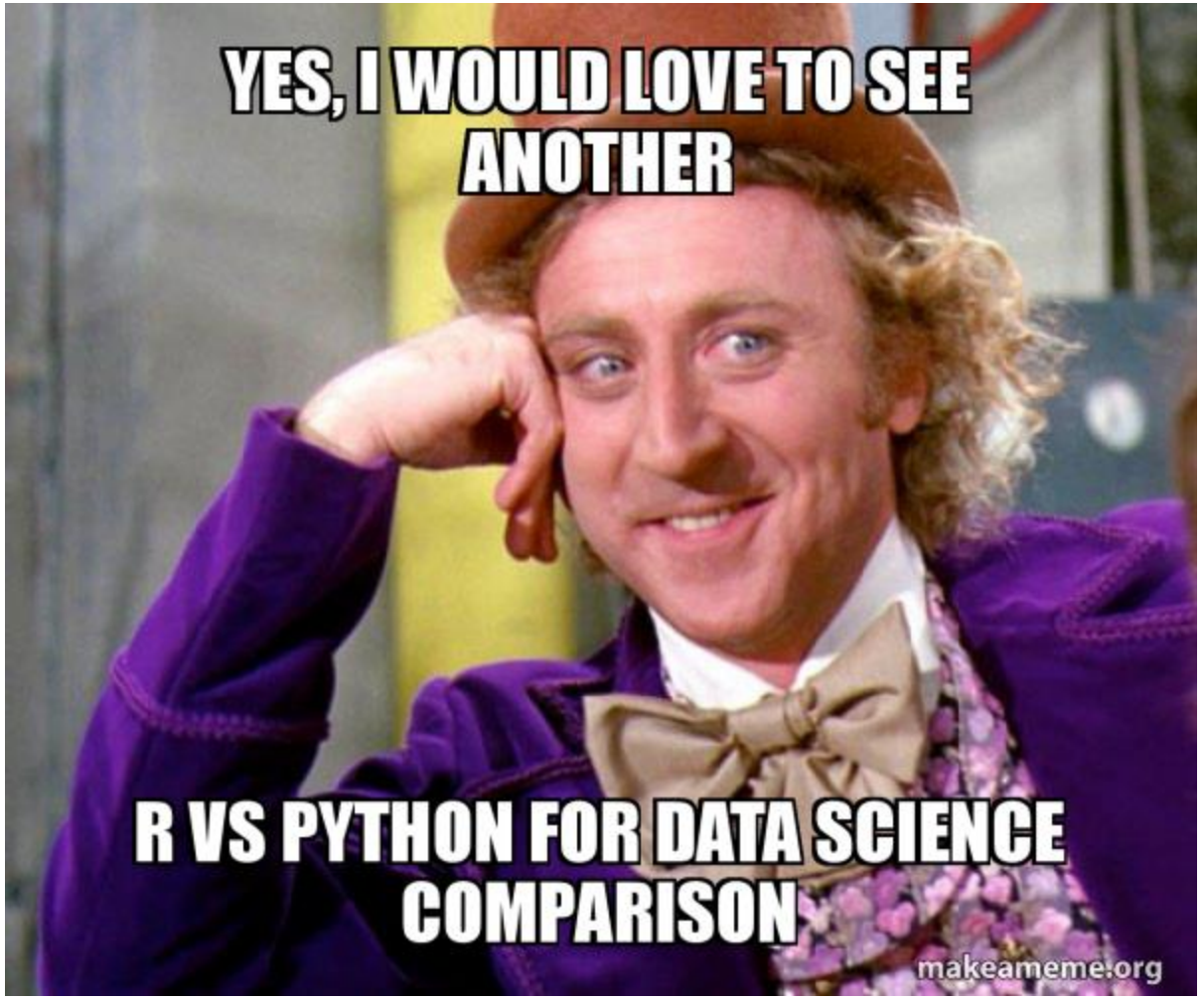
Reproducibility with R Markdown



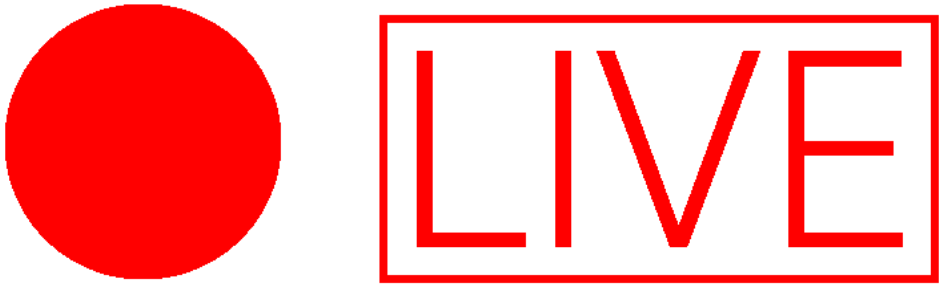
Reproducibility with R Markdown



Python?


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

Demonstration



Conclusion

- R is primarily used through interactive command-line interfaces and scripts.
- R is the programming language, but we can interact with it using other software.
- R efficiently handles various data types, particularly vectors and tables.
- R's functionality can be extended through a vast ecosystem of packages.
- R is free and open source.

Questions

Justin van Dijk

j.t.vandijk@ucl.ac.uk