

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.

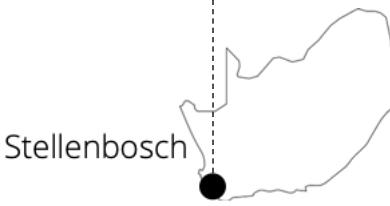
Schedule

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



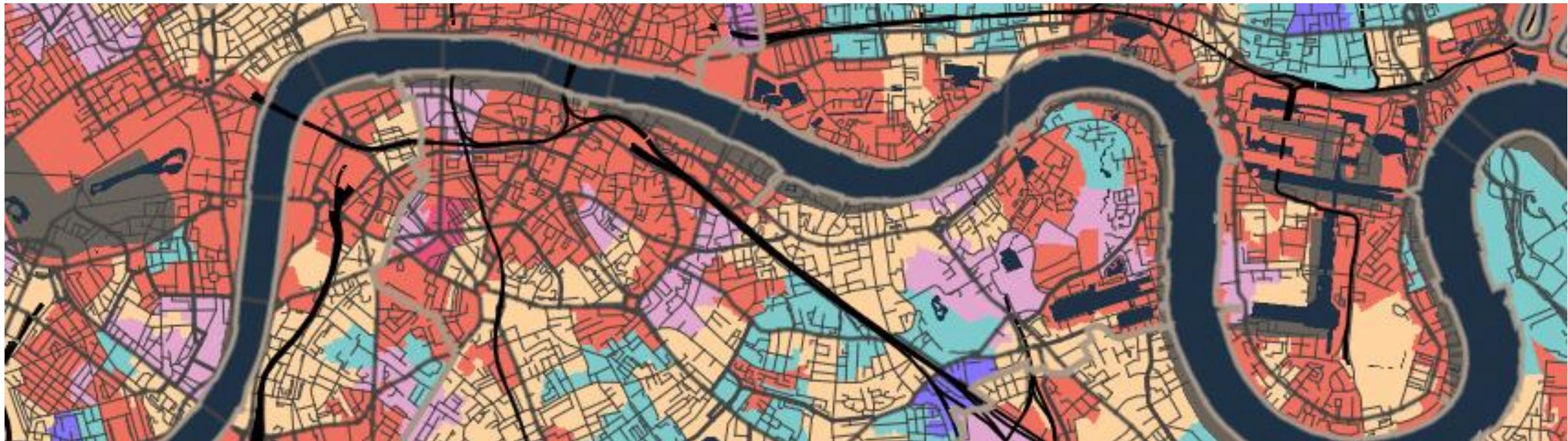
Ph.D. Transport Economics



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

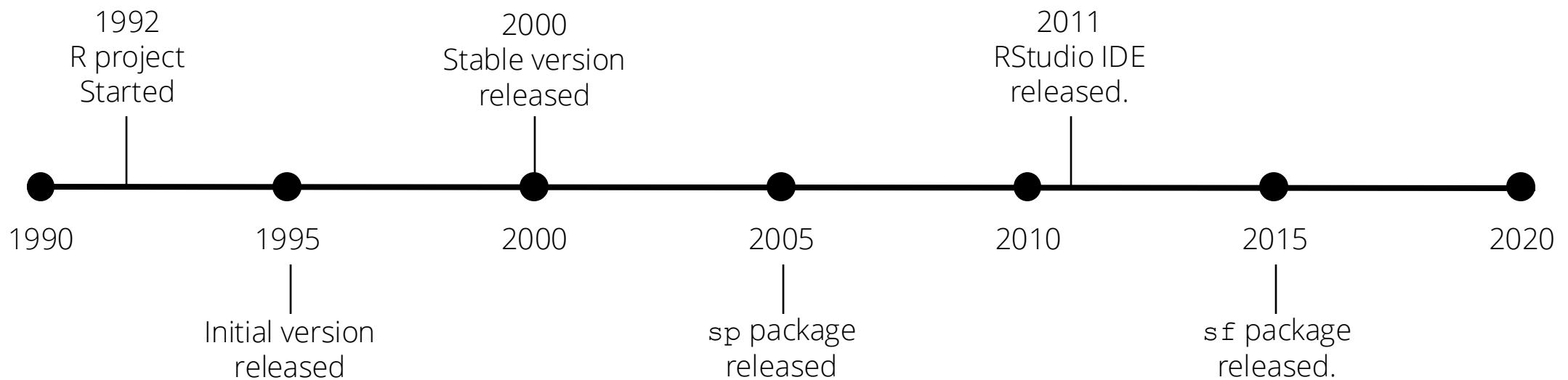
- Ιδεντιφιερς ανδ πριμιτι~~fe~~ δατα τψπες
- Ασσιγνμεντ, αριτημετιξ, λογιξαλ ανδ ρελατιοναλ οπερατορς
- Εχπρεσσιον ανδ στατεμεντς, δεβυγγινγ
- Φλω~~o~~ οφ ξοντρολ· σελεξτιον ανδ ρεπετιτιον
- Φυνξτιονς, παραμετερς πασσινγ, ξαλλ βψ α~~f~~λυε ανδ ρεφερενξε
- Οβσεξτ-οριεντεδ προγραμμινγ
- 1/2 διμενσιοναλ αρραψς, στρινγς ανδ δατα στρυξτυρες

The absolute basics

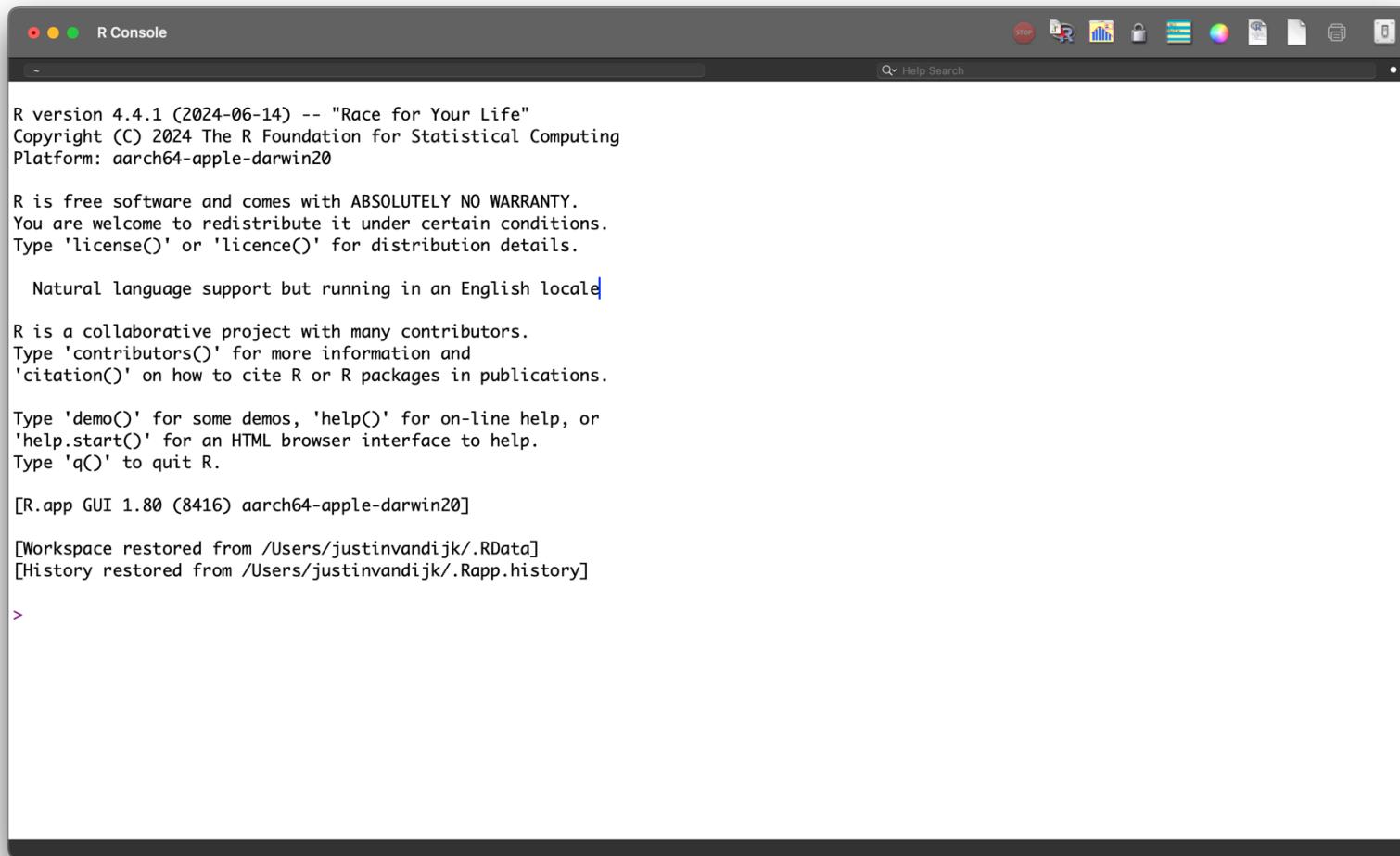
- R is primarily used through interactive command-line interfaces and scripts.
- R is the programming language, but we can interact with it using other software.
- R efficiently handles various data types, particularly vectors and tables.
- R's functionality can be extended through a vast ecosystem of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



R version 4.4.1 (2024-06-14) -- "Race for Your Life"
Copyright (C) 2024 The R Foundation for Statistical Computing
Platform: aarch64-apple-darwin20

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

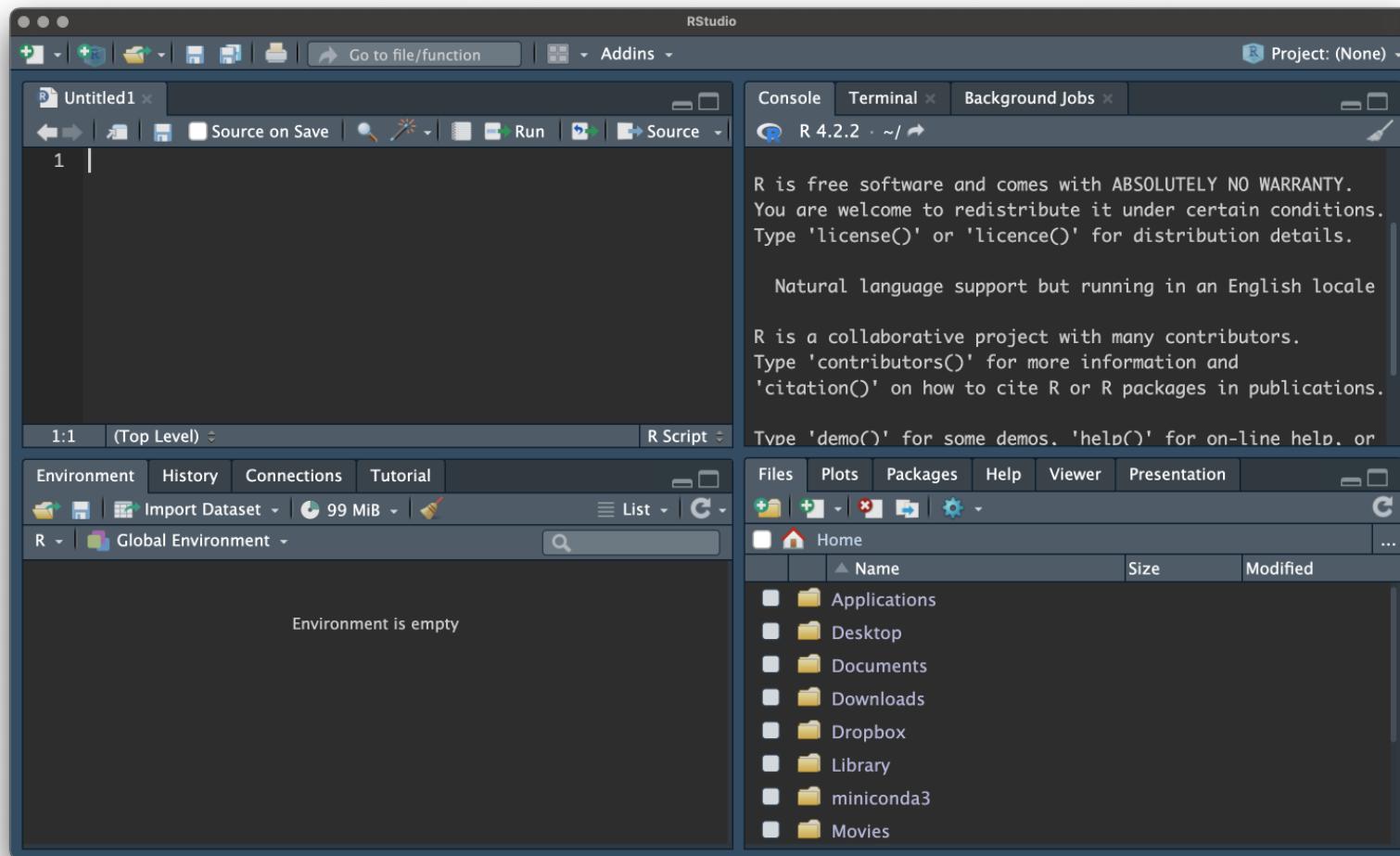
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[R.app GUI 1.80 (8416) aarch64-apple-darwin20]

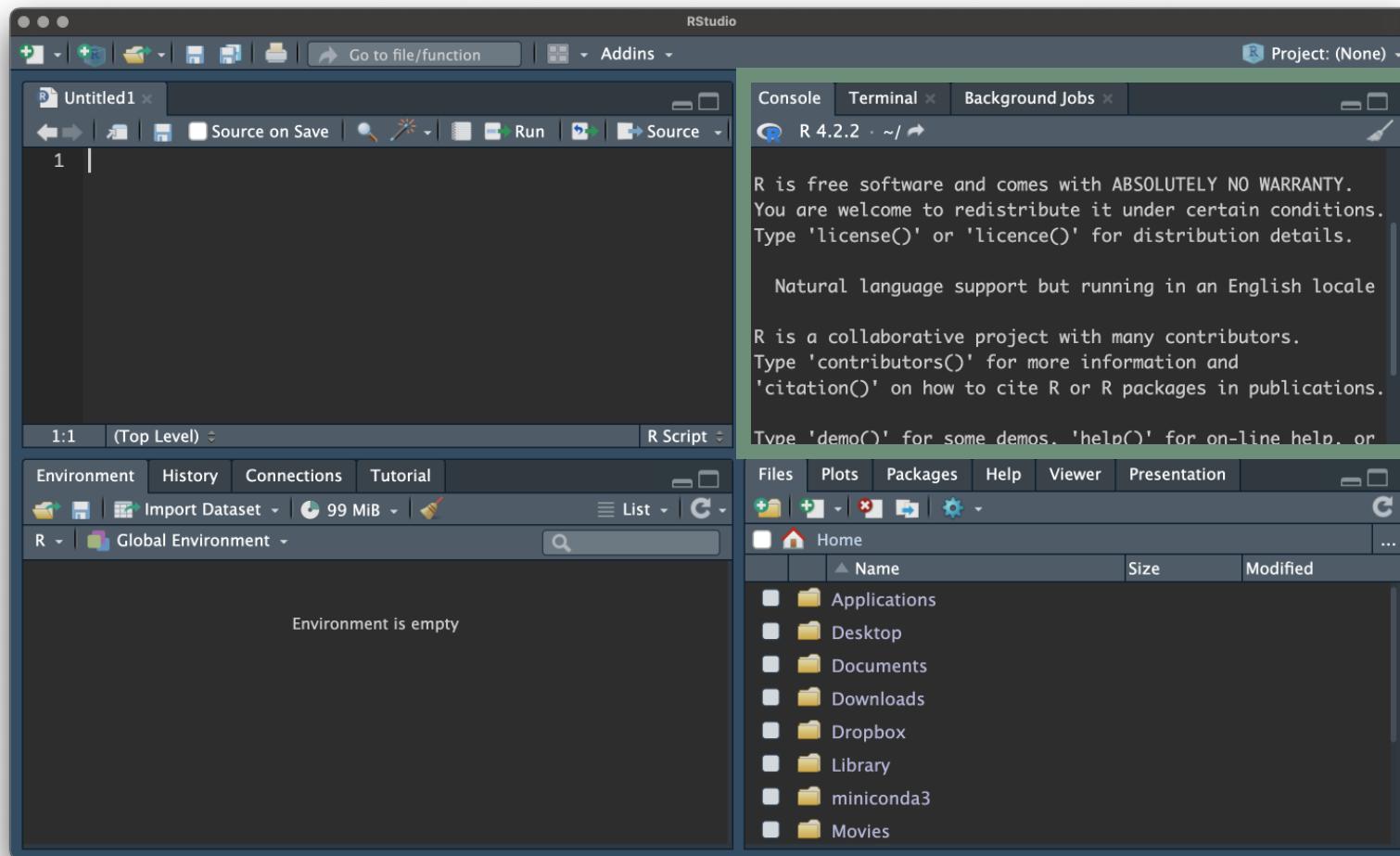
[Workspace restored from /Users/justinvandijk/.RData]
[History restored from /Users/justinvandijk/.Rapp.history]

>

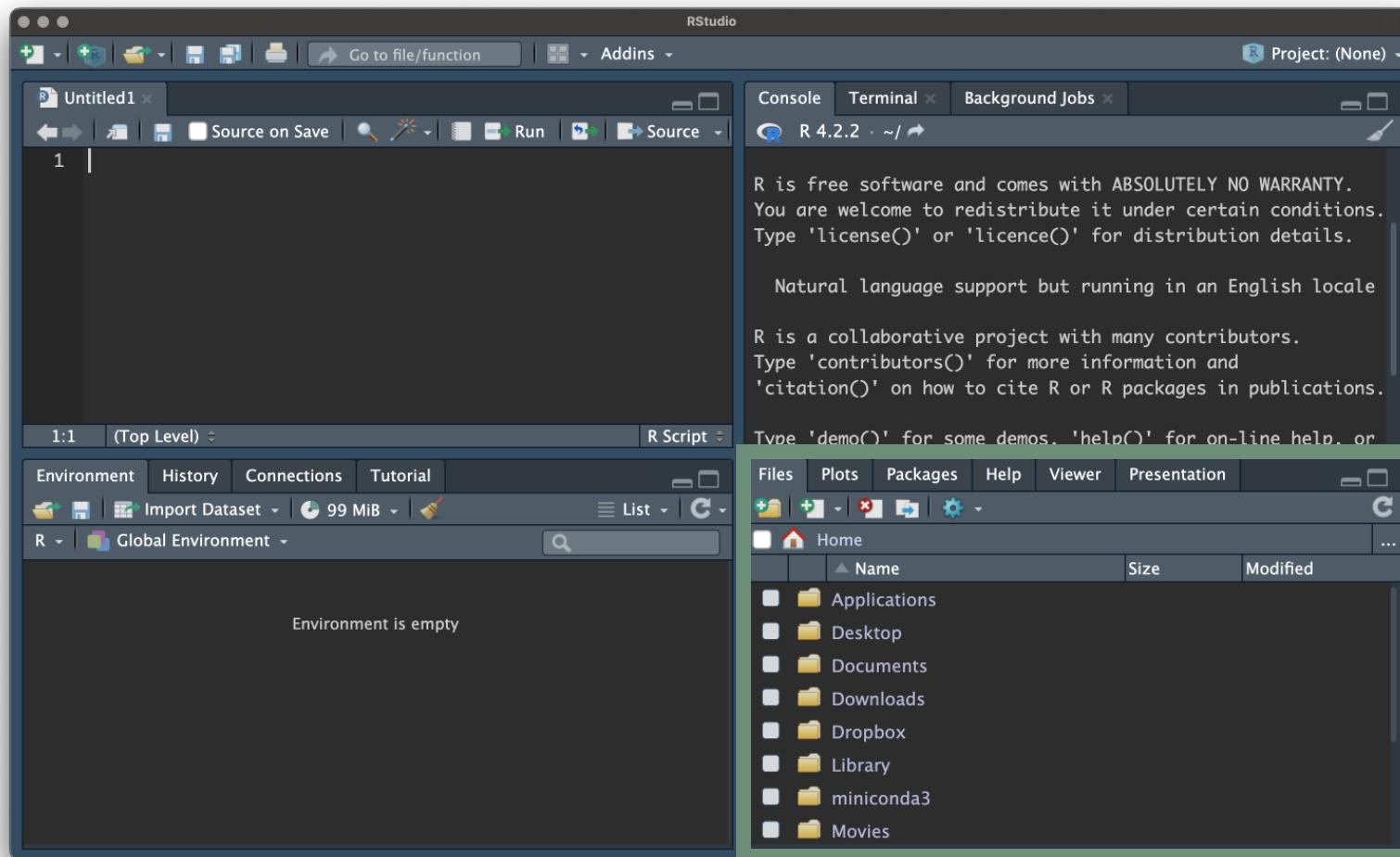
Working with R



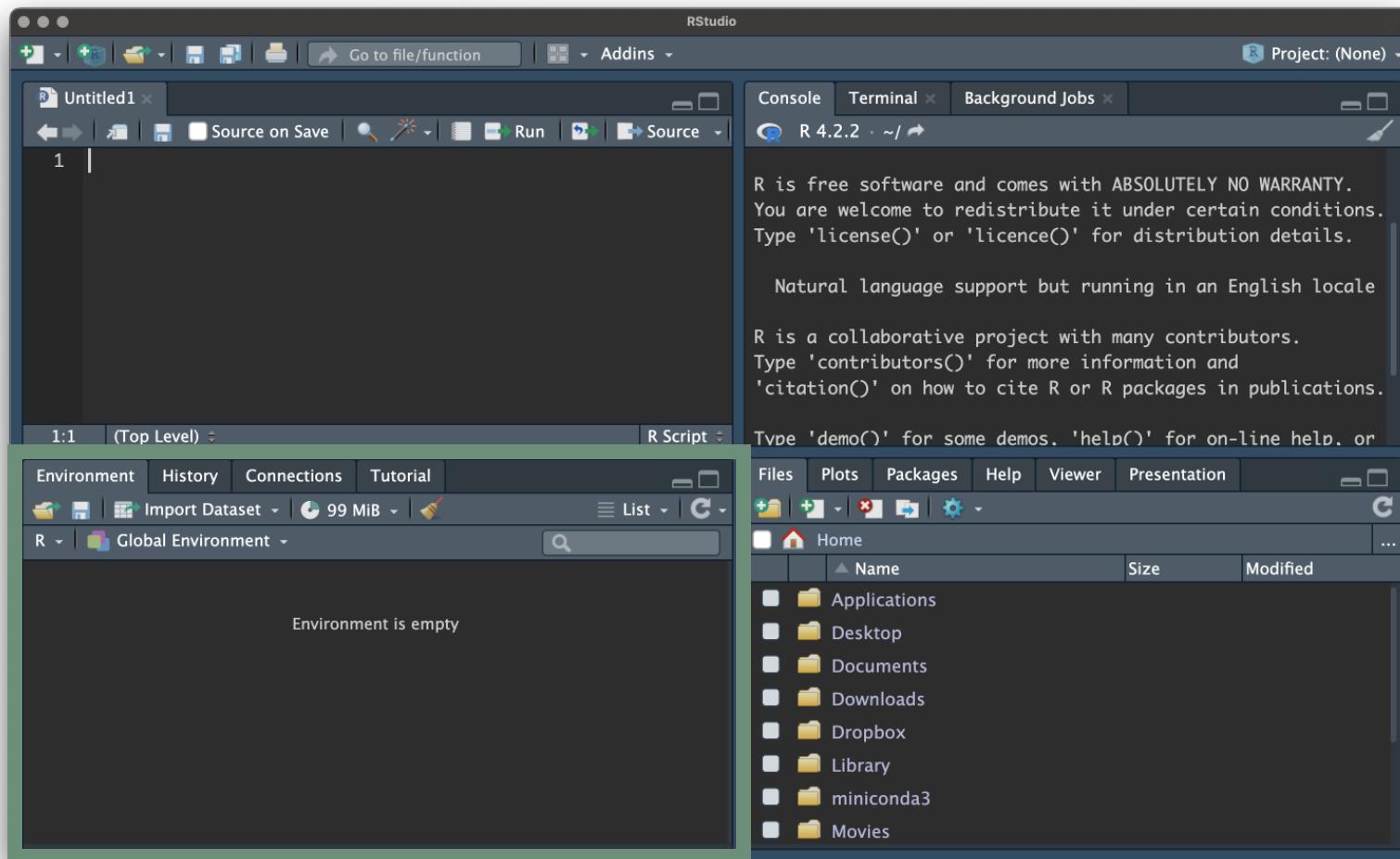
Working with R



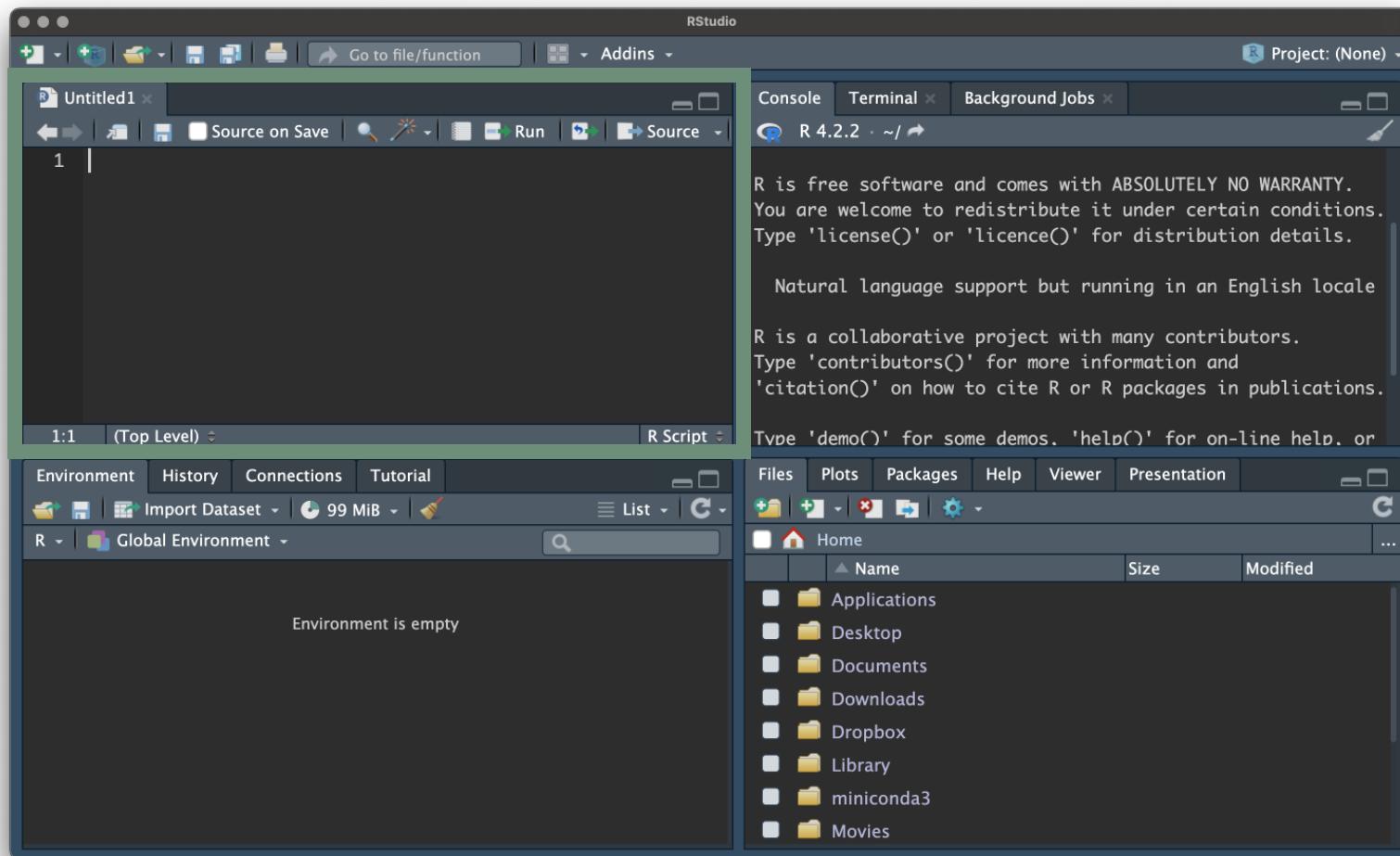
Working with R



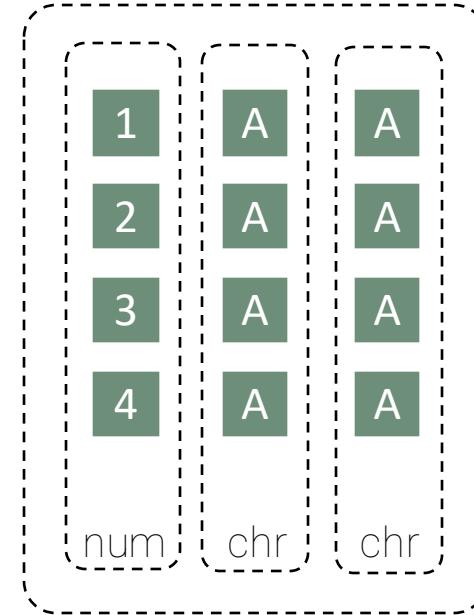
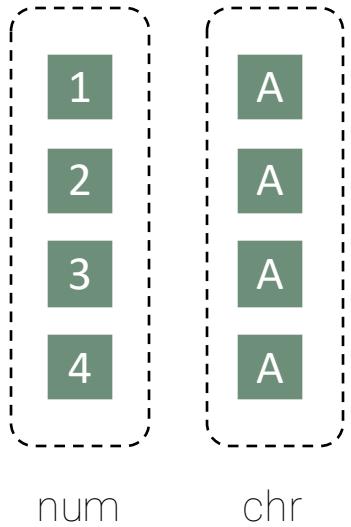
Working with R



Working with R



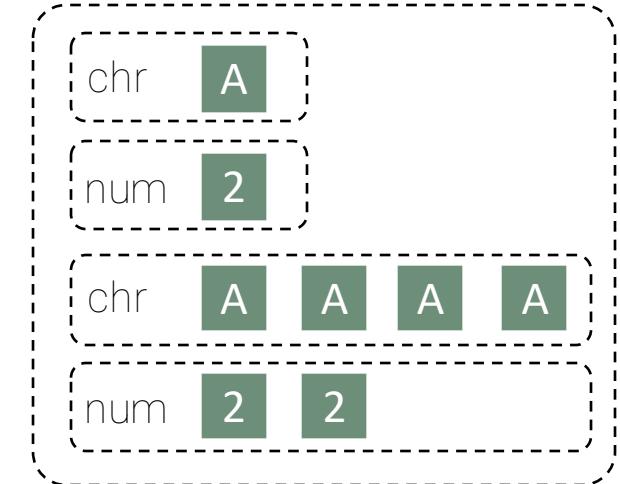
Principles of R: Data types



Scalar

Vector

Dataframe



List

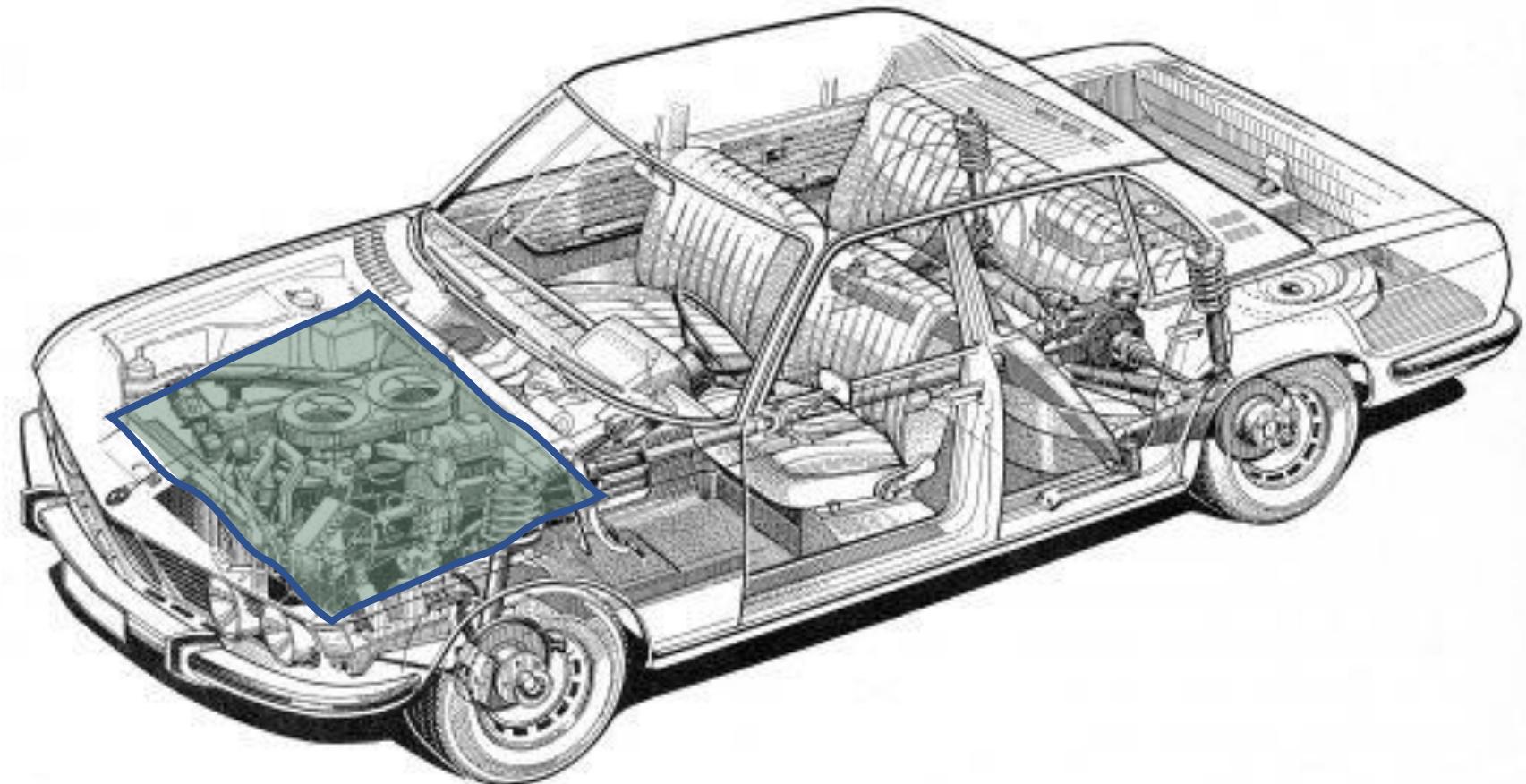
Principles of R: Variables

- 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.

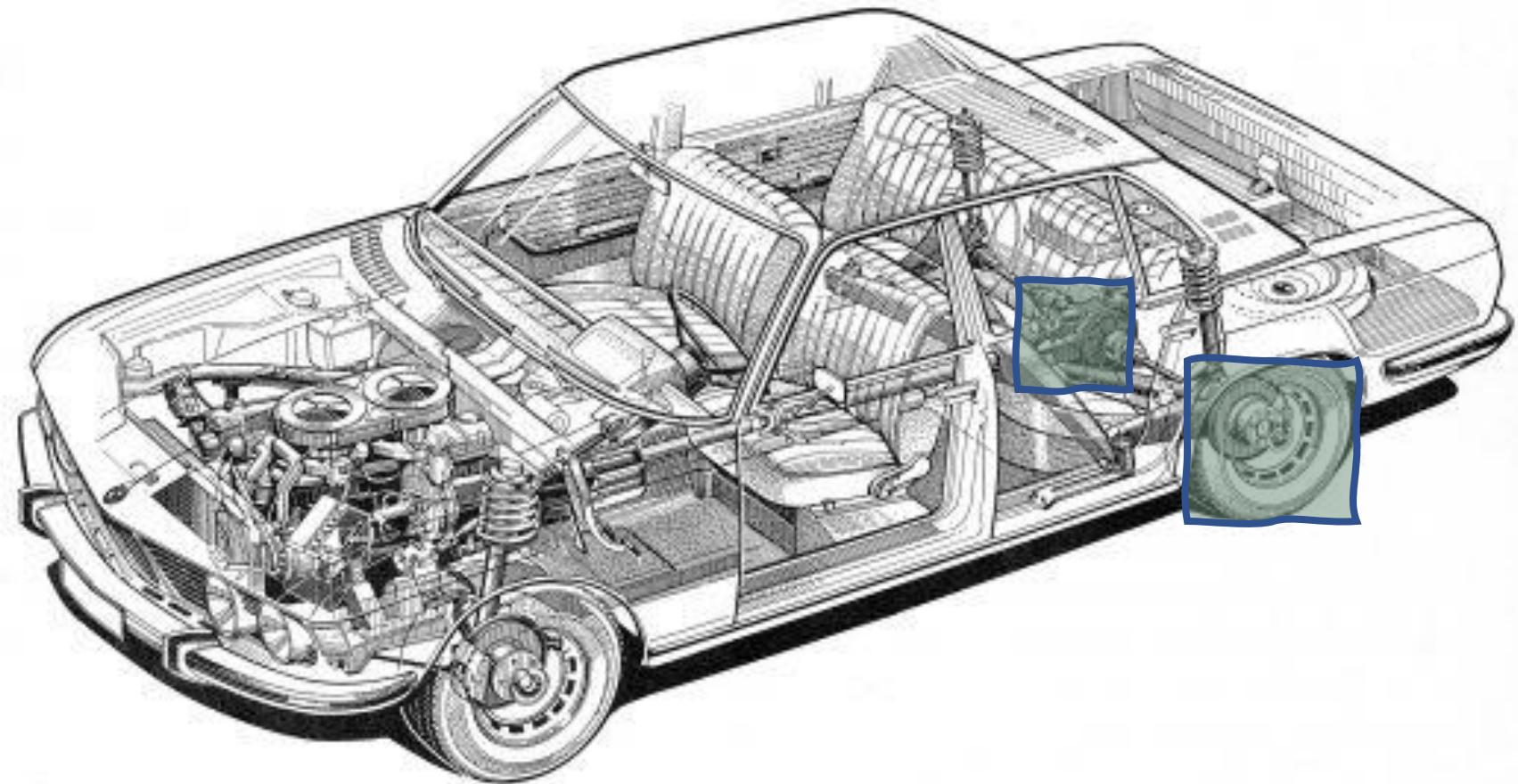
Principles of R: Functions

- 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.

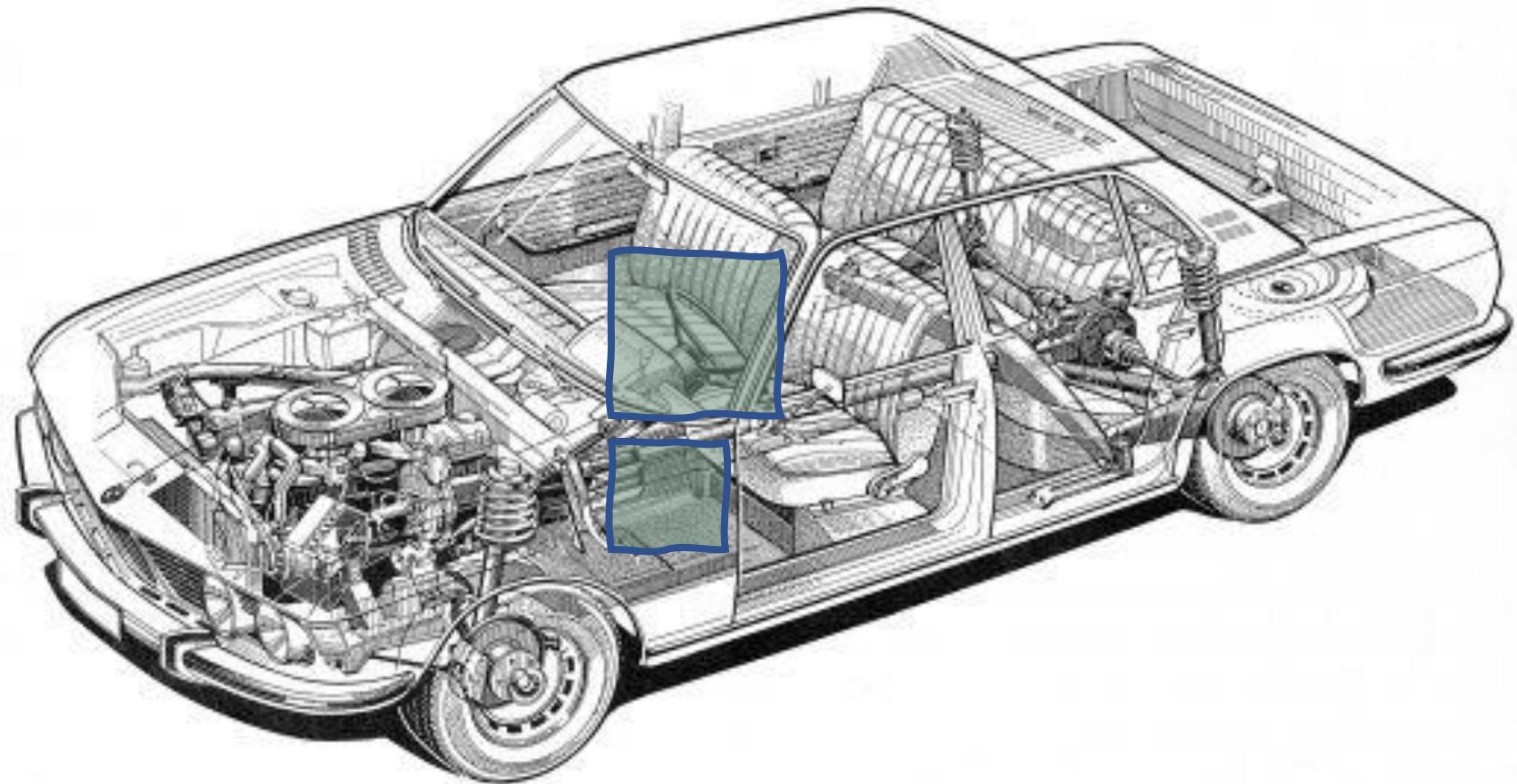
Principles of R: Extensions



Principles of R: Extensions



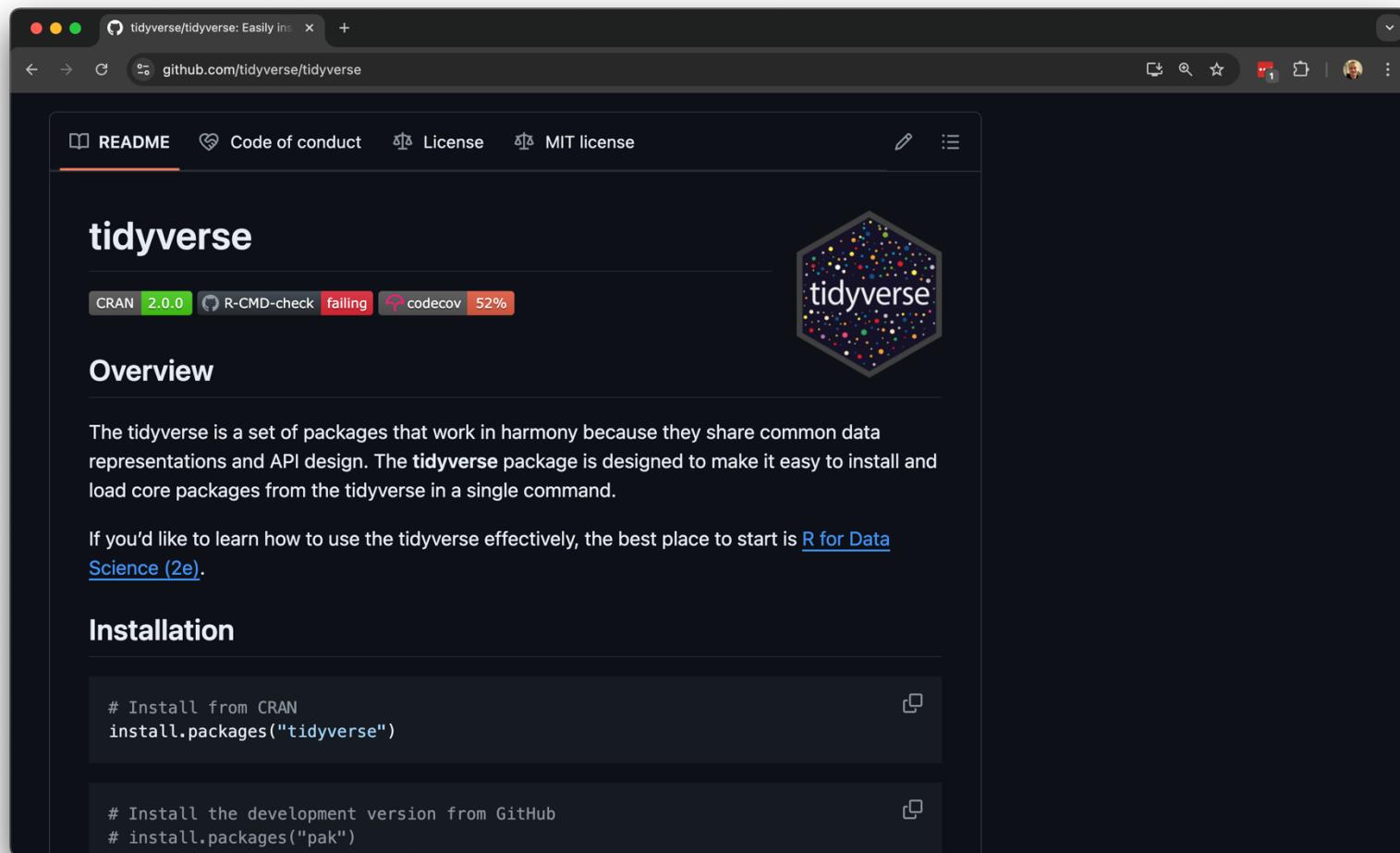
Principles of R: Extensions



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



A screenshot of a web browser displaying the GitHub repository for the tidyverse. The page has a dark theme. At the top, there's a navigation bar with links to 'README', 'Code of conduct', 'License', and 'MIT license'. Below this, the repository name 'tidyverse' is displayed in large white text. To the right of the name is a hexagonal logo filled with small colored dots. Below the name, there are three status badges: 'CRAN 2.0.0', 'R-CMD-check failing', and 'codecov 52%'. A section titled 'Overview' contains text about the tidyverse being a set of packages that work in harmony. It also mentions the 'tidyverse' package and provides a link to 'R for Data Science (2e)'. A 'Installation' section shows two code snippets for installing the package from CRAN and GitHub. The first snippet is:

```
# Install from CRAN  
install.packages("tidyverse")
```

The second snippet is:

```
# Install the development version from GitHub  
# install.packages("pak")
```

Simple Features

A screenshot of a web browser displaying the GitHub README page for the `r-spatial/sf` repository. The page title is "Simple Features for R".

The page content includes:

- A brief description: "A package that provides [simple features access](#) for R."
- Links to external resources: [Blogs](#), [links](#), [Cheatsheet](#), [Installing](#), [Contributing](#), [Acknowledgment](#), and [How to cite](#).
- A section titled "Package sf:" listing the package's features:
 - 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 (through R package [s2](#)) 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
 - [stars](#) for raster data, and raster or vector data cubes (spatial time series)

The page also features a logo consisting of the letters "sf" in blue, enclosed within a hexagonal grid pattern.

At the bottom of the page, there is a link to a gist: [network data](https://gist.github.com/edzer/f461a3a95570c4ab7edf3125c2f19d20).

tmap

github.com/r-tmap/tmap

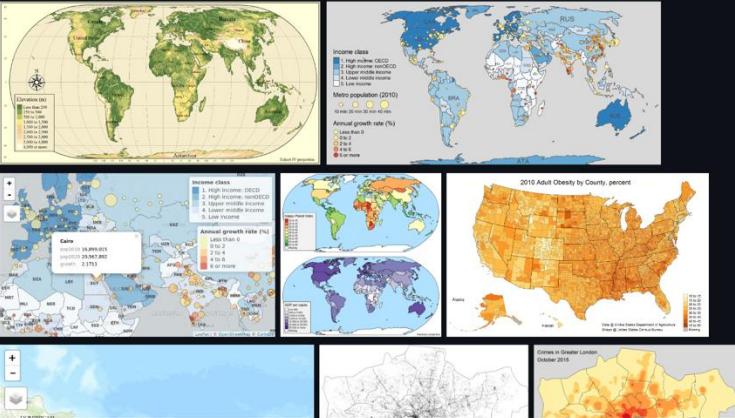
[README](#) [GPL-3.0 license](#)

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](#)



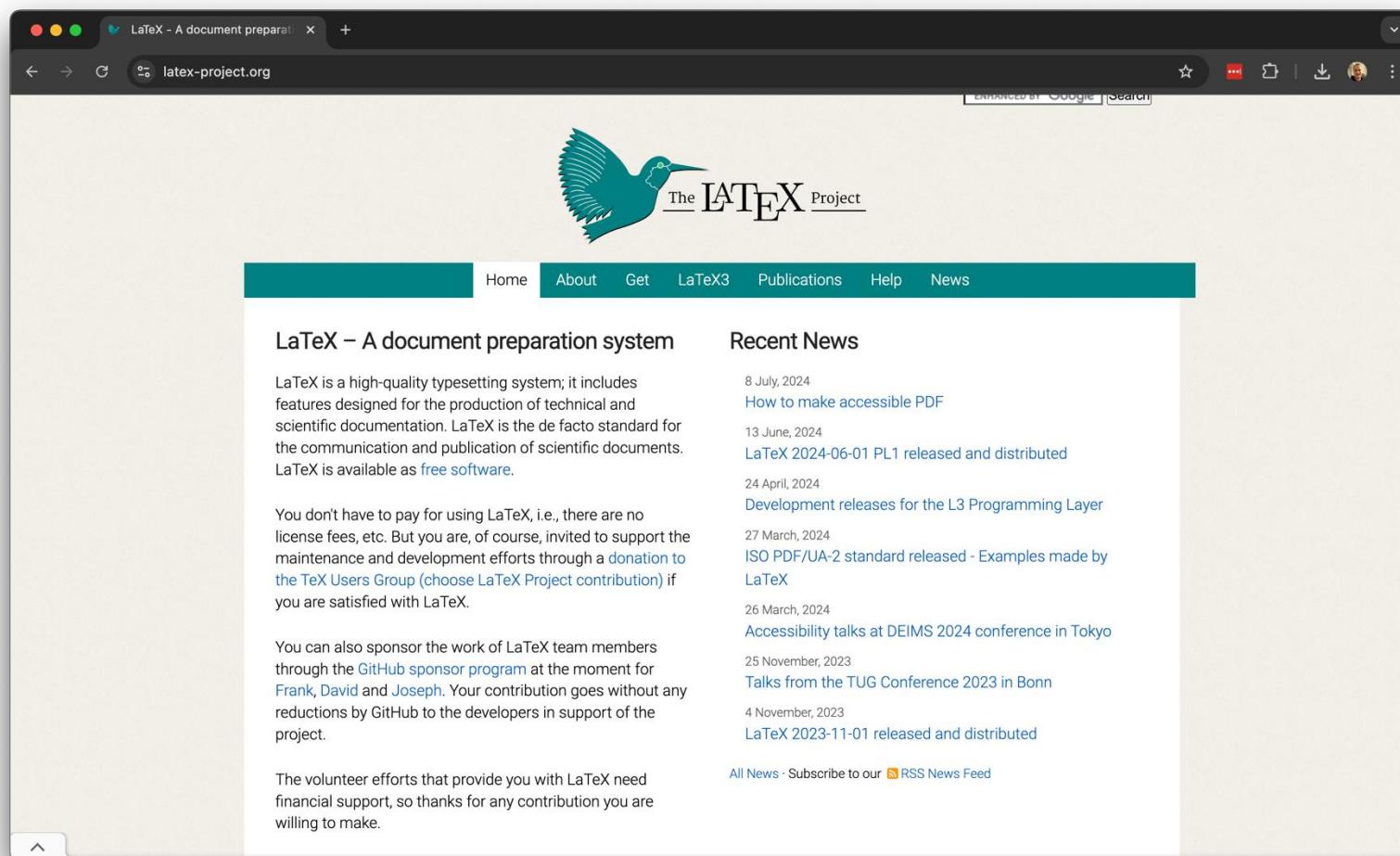
[tmap](#) is an actively maintained open-source [R](#)-library for drawing thematic maps. The API is based on [A Layered Grammar of Graphics](#) and resembles the syntax of [ggplot2](#), a popular R-library for drawing charts.



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



The screenshot shows a web browser window displaying the LaTeX Project homepage at [latex-project.org](https://www.latex-project.org). The page features a large teal bird logo above the text "The LATEX Project". A navigation bar at the top includes links for Home, About, Get, LaTeX3, Publications, Help, and News. The main content area has two columns. The left column contains text about LaTeX being a high-quality typesetting system for technical documentation, a note that it's free software, information about sponsorship through GitHub, and a message of thanks for volunteer contributions. The right column is titled "Recent News" and lists several items with dates and titles, such as "How to make accessible PDF" (8 July, 2024) and "LaTeX 2024-06-01 PL1 released and distributed" (13 June, 2024). At the bottom of the news section, there are links for "All News" and "Subscribe to our RSS News Feed".

LaTeX – A document preparation system

LaTeX is a high-quality typesetting system; it includes features designed for the production of technical and scientific documentation. LaTeX is the de facto standard for the communication and publication of scientific documents. LaTeX is available as [free software](#).

You don't have to pay for using LaTeX, i.e., there are no license fees, etc. But you are, of course, invited to support the maintenance and development efforts through a [donation to the TeX Users Group \(choose LaTeX Project contribution\)](#) if you are satisfied with LaTeX.

You can also sponsor the work of LaTeX team members through the [GitHub sponsor program](#) at the moment for [Frank](#), [David](#) and [Joseph](#). Your contribution goes without any reductions by GitHub to the developers in support of the project.

The volunteer efforts that provide you with LaTeX need financial support, so thanks for any contribution you are willing to make.

Recent News

8 July, 2024
[How to make accessible PDF](#)

13 June, 2024
[LaTeX 2024-06-01 PL1 released and distributed](#)

24 April, 2024
[Development releases for the L3 Programming Layer](#)

27 March, 2024
[ISO PDF/UA-2 standard released - Examples made by LaTeX](#)

26 March, 2024
[Accessibility talks at DEIMS 2024 conference in Tokyo](#)

25 November, 2023
[Talks from the TUG Conference 2023 in Bonn](#)

4 November, 2023
[LaTeX 2023-11-01 released and distributed](#)

[All News](#) · [Subscribe to our RSS News Feed](#)

Reproducibility with R Markdown

This is the basic model for Quarto publishing—take a source document and render it to a variety of output

The screenshot shows the Quarto RStudio extension interface. On the left, the RStudio sidebar displays navigation links: 'Get Started', 'Tutorial: Hello, Quarto', 'Tutorial: Computations', and 'Tutorial: Authoring'. The main workspace is split into two panes. The left pane shows the R Markdown source code for a file named 'hello.qmd':---

```
{r}<#| label: load-packages#| include: false}  
library(tidyverse)  
library(palmerpenguins)
```

Meet Quarto

Quarto enables you to weave together content and executable code into a finished document. To learn more about Quarto see <https://quarto.org>.

Meet the penguins

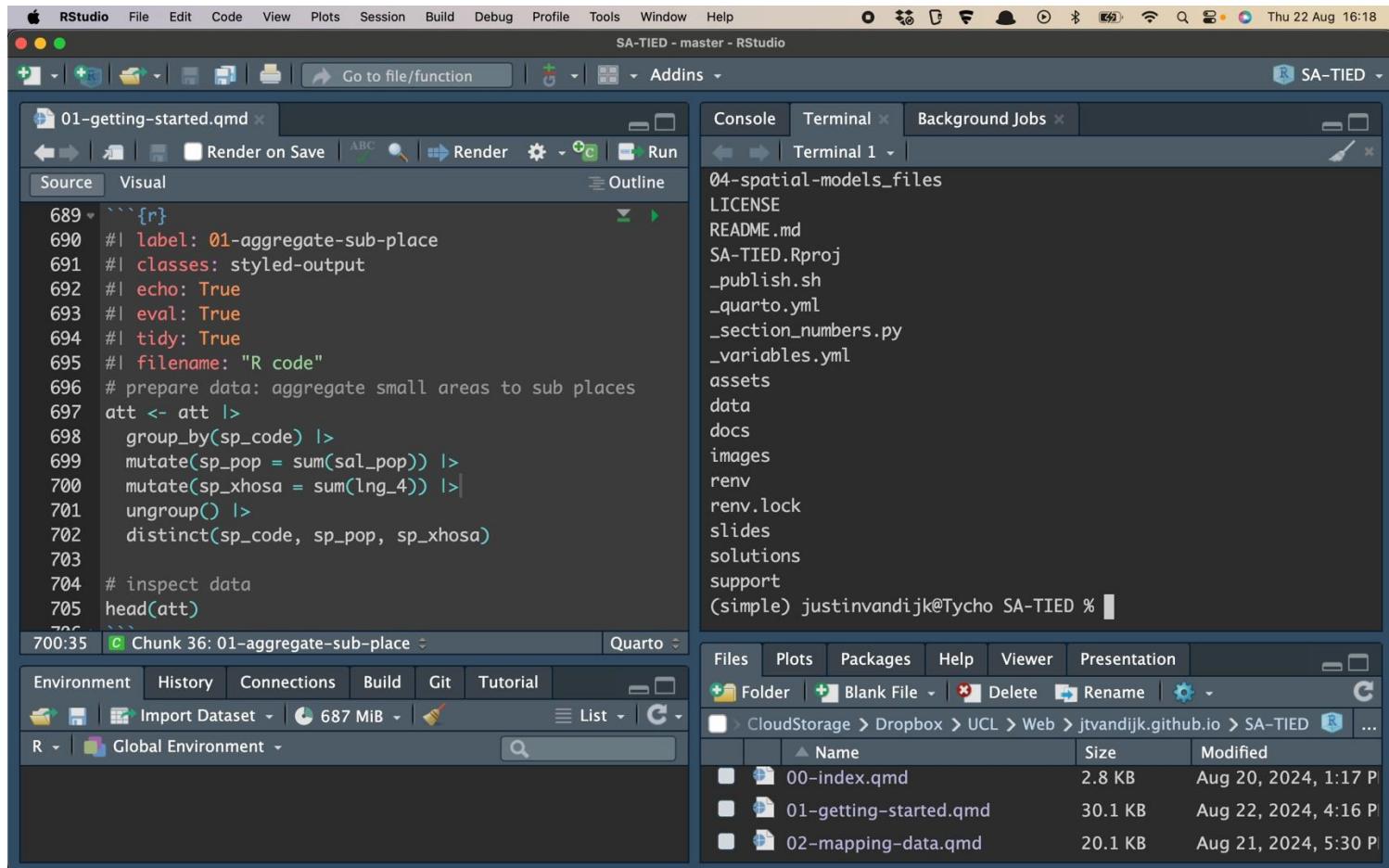
The `penguins` data from the `palmerpenguins` package contains size measurements for `nrow(penguins)` penguins from three species observed on three islands in the Palmer Archipelago, Antarctica.

The plot below shows the relationship between flipper and bill lengths of these penguins.

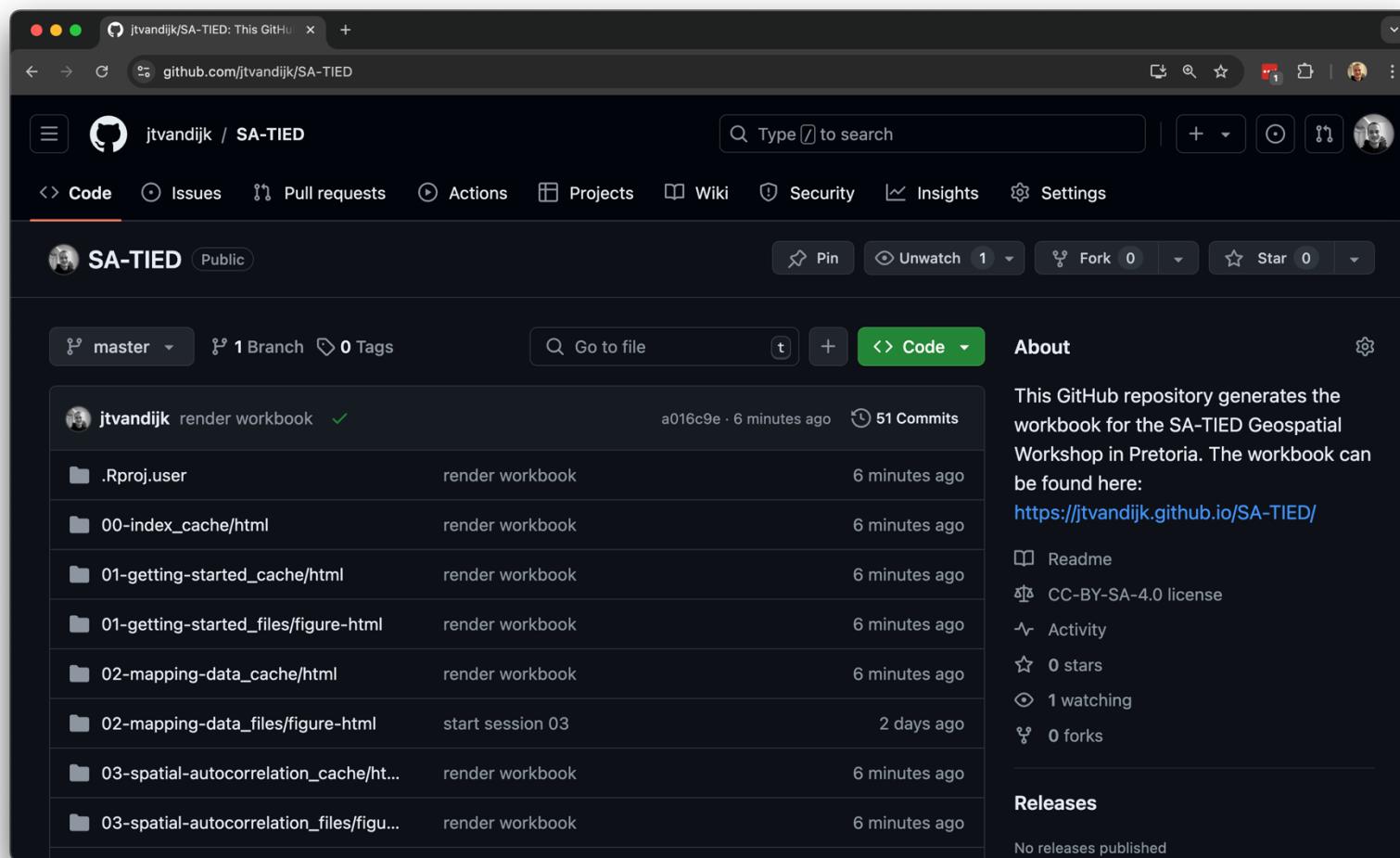
```
{r}<#| label: plot-penguins#| include: false}(Top Level) :  
Console
```

The right pane shows the rendered HTML output. It features a header 'Hello, Quarto' and a section 'Meet Quarto' with the same text as the source. Below it is a section 'Meet the penguins' with the same text and a small image of three penguins labeled 'CHINSTRAP', 'GENTOO', and 'ADELIE'. At the bottom is a scatter plot titled 'Flipper and bill length Dimensions for penguins at Palmer Station LTER'. The y-axis is 'bill (mm)' and the x-axis is 'flipper (mm)'. Data points are colored by species: Adelie (orange), Chinstrap (purple), and Gentoo (teal). A legend on the right identifies the species colors.

Reproducibility with R Markdown



Reproducibility with R Markdown



A screenshot of a GitHub repository page for "jtvdijk / SA-TIED". The repository is public and contains 1 branch and 0 tags. The "Code" tab is selected. A recent commit by "jtvdijk" titled "render workbook" is shown, along with several other commits related to rendering workbooks for different chapters of the workshop.

File	Action	Time
.Rproj.user	render workbook	6 minutes ago
00-index_cache/html	render workbook	6 minutes ago
01-getting-started_cache/html	render workbook	6 minutes ago
01-getting-started_files/figure-html	render workbook	6 minutes ago
02-mapping-data_cache/html	render workbook	6 minutes ago
02-mapping-data_files/figure-html	start session 03	2 days ago
03-spatial-autocorrelation_cache/ht...	render workbook	6 minutes ago
03-spatial-autocorrelation_files/figu...	render workbook	6 minutes ago

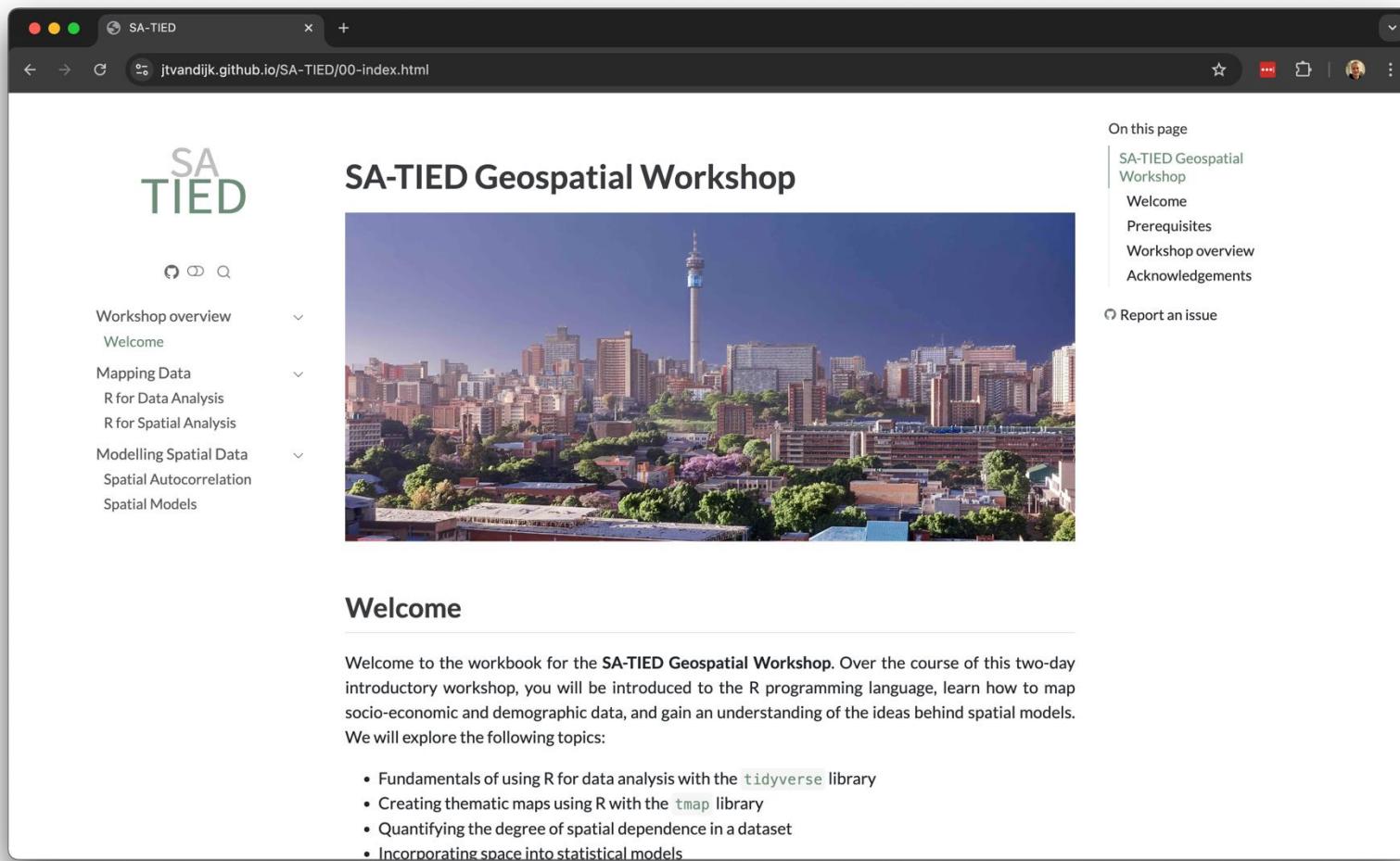
About

This GitHub repository generates the workbook for the SA-TIED Geospatial Workshop in Pretoria. The workbook can be found here:
<https://jtvdijk.github.io/SA-TIED/>

Readme
CC-BY-SA-4.0 license
Activity
0 stars
1 watching
0 forks

Releases
No releases published

Reproducibility with R Markdown



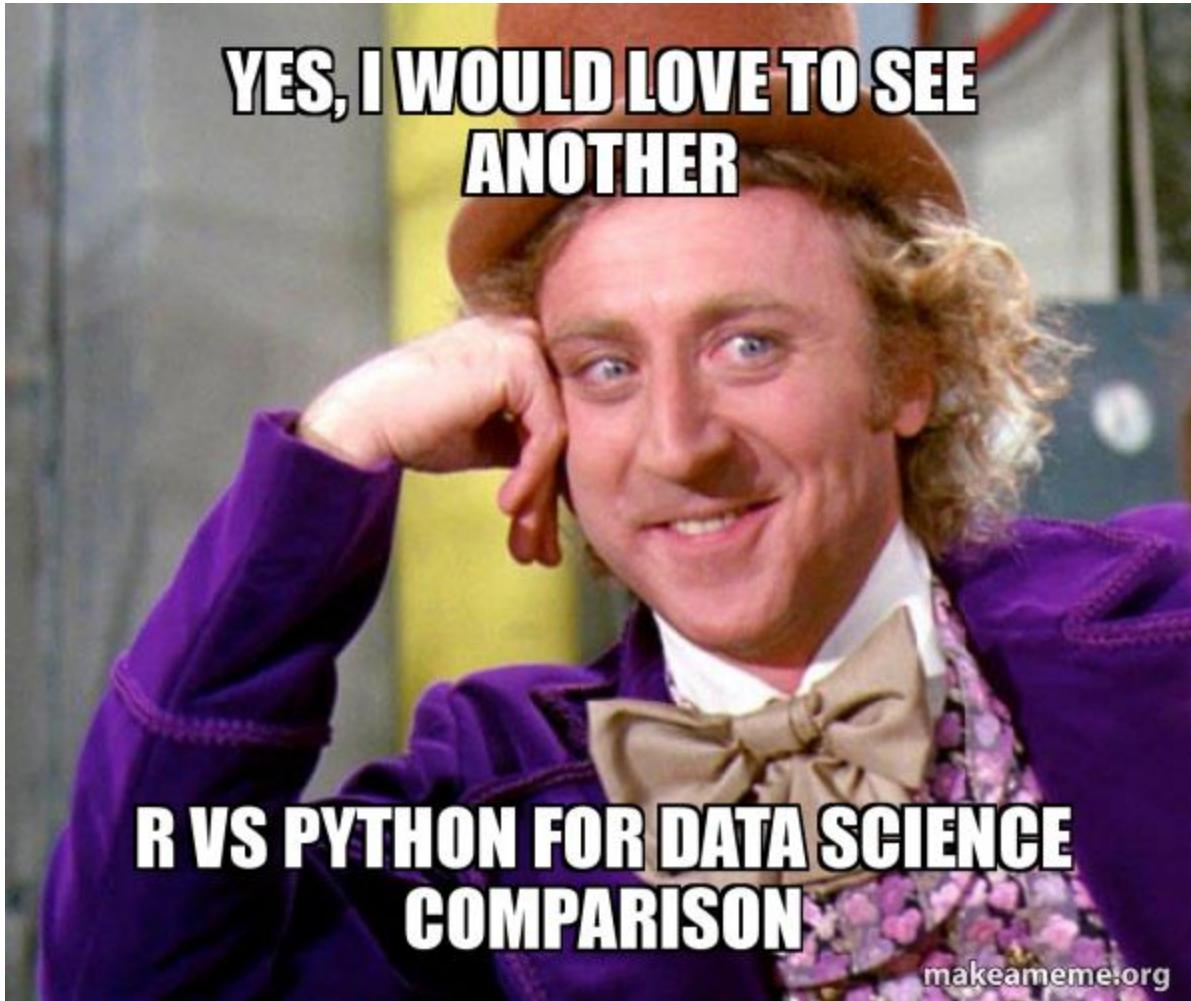
The screenshot shows a web browser window for the "SA-TIED Geospatial Workshop". The title bar says "SA-TIED" and the address bar shows "jtvdijk.github.io/SA-TIED/00-index.html". The page features a large header "SA-TIED Geospatial Workshop" with a city skyline image below it. To the left is a sidebar with navigation links for "Workshop overview", "Welcome", "Mapping Data", "R for Data Analysis", "R for Spatial Analysis", "Modelling Spatial Data", "Spatial Autocorrelation", and "Spatial Models". On the right, there's a "On this page" sidebar with links to "SA-TIED Geospatial Workshop", "Welcome", "Prerequisites", "Workshop overview", and "Acknowledgements". A "Report an issue" button is also present.

Welcome

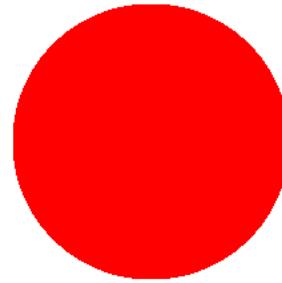
Welcome to the workbook for the **SA-TIED Geospatial Workshop**. Over the course of this two-day introductory workshop, you will be introduced to the R programming language, learn how to map socio-economic and demographic data, and gain an understanding of the ideas behind spatial models. We will explore the following topics:

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

Python?



Demonstration



LIVE

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

