Introduction to R

March, 2022

A data science workflow

1. **Retrieve data**
- Generate
- Download

1. **Store data**
- Files
- Databases

3. Process data

- Preparation
- Analysis, Visualisation

4. Publish results

- Reports, Memos, Papers, Books
- Replication Data

What is programming?

Programming is about writing **programs**

- Sets of instructions (algorithms) to be executed by a computer
- written in a computer language (source code, or just code)
- stored in files (and organized in software projects)

Integrated Development Environments (IDEs)

Editor + Console + Tools

What is programming?

How many computer languages exist?

Why are there so many languages?

- Different needs: computation, visualization, robotisation...
- Different hardware: laptops, robots, fridges...
- Versions: improvements, new needs...
- Programming paradigms/ categorizations

Programming concepts are mostly language-agnostic

R

```r
library(readr)
mydata <- read_csv("myfile.csv")
```

Python

```python
import pandas as pd
mydata = pd.read_csv("myfile.csv")
```

Operators

Arithmetic Operators

- Addition, Subtraction, Multiplication, Division

Comparison Operators

- Equal, Not equal, Greater than, , Less than, Less or equal than

Boolean Operators

- and, or, not

Precedence of Operators

- Parentheses, Exponents and roots, Multiplications and divisions, Additions and subtractions
- Horizontal: left to right
- Vertical: top down

Data Types

Numeric

- integer, floating point, complex number

String

- sequence of characters

Boolean

- True/False, 1/0

Structured (combinations of the others)

- Vectors, Lists Arrays, Dictionaries, Dataframes

Functions are reusable chunks of code

- Each language has a set of **built-in** functions or **commands**

Control flow

Conditional statements

- if-then(-else)

Loop constructs

- for
- do while
- while

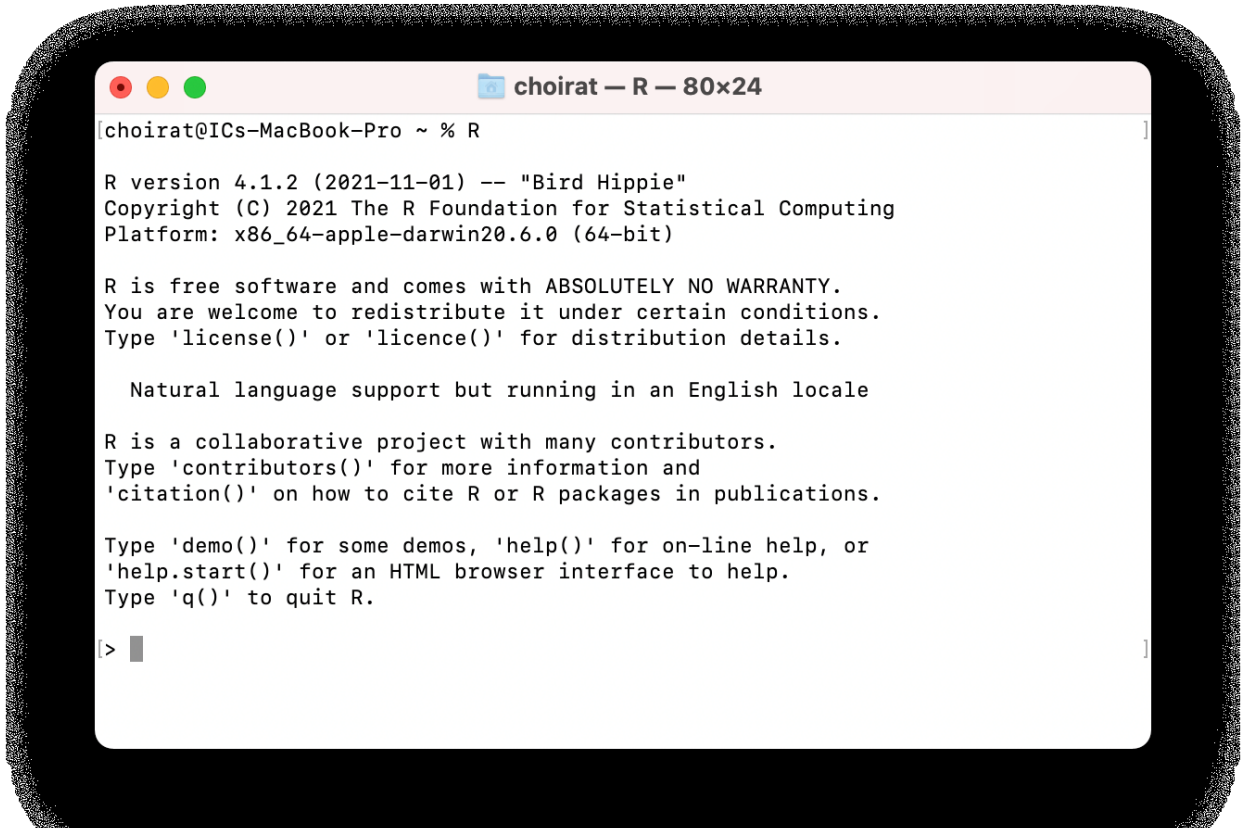
S, Splus, R

R is an implementation of the S programming language [...]. S was created by John Chambers in 1976 while at Bell Labs. A commercial version of S was offered as S-PLUS starting in 1988. Many codes written for S-PLUS run unaltered in R.

In 1991 Ross Ihaka and Robert Gentleman at the University of Auckland, New Zealand, embarked on an S implementation, independent of S-PLUS. They began publicizing it in 1993. It was named partly after the first names of the first two R authors and partly as a play on the name of S. In 1995, Martin Maechler convinced Ihaka and Gentleman to make R free and open-source software under the GNU General Public License.

(Source: Wikipedia)

What does R look like?



RStudio

A software company

RStudio

Software company

rstudio.com

Founder: Joseph J. Allaire

Founded: 2009

Headquarters location: Boston, Massachusetts, United States

Ports

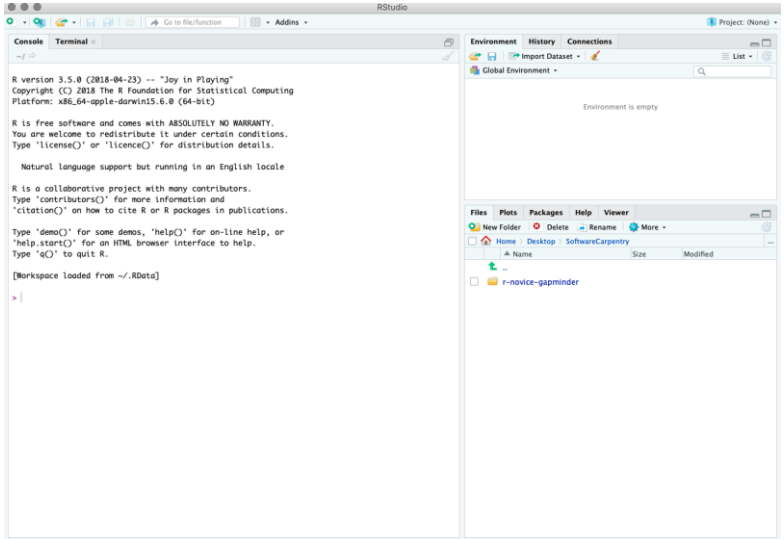
Revenue

Mission

RStudio's mission is **to create free and open-source software for data science, scientific research, and technical communication**.

https://www.rstudio.com › about

An Integrated Development Environment (IDE)



Alternatives to a local installation

+

- Computing power an storage are provided online by remote servers.
- Offers tools for state of the art research reproducibility and dissemination standards.

First steps with Renku

What is Renku?

Renku (連句 “linked verses”), is a Japanese form of popular collaborative linked verse poetry, written by more than one author working together.

The Renku Project is a platform that bundles together various tools for reproducible and collaborative data analysis projects. It is aimed at independent researchers and data scientists as well as labs, collaborations, and courses and workshops. Renku can be used by anyone who deals with data, whether they are a researcher, data analyst, project owner, or data provider.

(Source: Renku documentation)

SDSC

ETH zürich EPFL

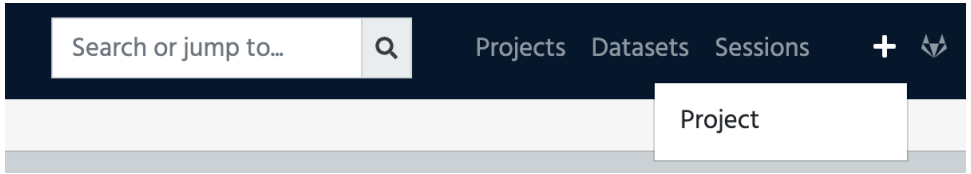
https://datascience.ch/

Demo

We give a quick tour, a demo, and then it's your turn!

https://renkulab.io/

1. Create an account
2. Login
3. Create an R project
4. Open RStudio



New project

Title

myfirstproject

There are a few reserved names you cannot use.

Namespace

cchoirat

Group namespaces may restrict the visibility options.

Identifier

cchoirat/myfirstproject

This is automatically derived from Namespace and Title.

Visibility

Public

Template source

RenkuLab | Custom

Template

Renku / Basic R (4.0.5) Project

The simplest R-4.0.5-based renku project with a basic directory structure and necessary supporting files.

Create project

myfirstproject

Public

cchoirat/myfirstproject

Start

Overview | Collaboration | Files | Datasets | Sessions | Settings

General

Stats

Commits

Status

(This project has no description. You can provide one on the settings tab.)

fork 0 | star 0 | View in GitLab

Updated 11 seconds ago.

README.md

myfirstproject

Introduction

This is a Renku project - basically a git repository with some bells and whistles. You'll find we have already created some useful things like `data` and `notebooks` directories and a `Dockerfile`.

Working with the project

The simplest way to start your project is right from the Renku platform - just click on the `Environments` tab and start a new session. This will start an interactive environment right in your browser.

To work with the project anywhere outside the Renku platform, click the `Settings` tab where you will find the git repo URLs - use `git` to clone the project on whichever machine you want.

myfirstproject

Public

cchoirat/myfirstproject

Start

Overview Collaboration Files Datasets Sessions Settings

Back to sessions list

Starting session

Checking existing sessions...

myfirstproject

Public

cchoirat/myfirstproject

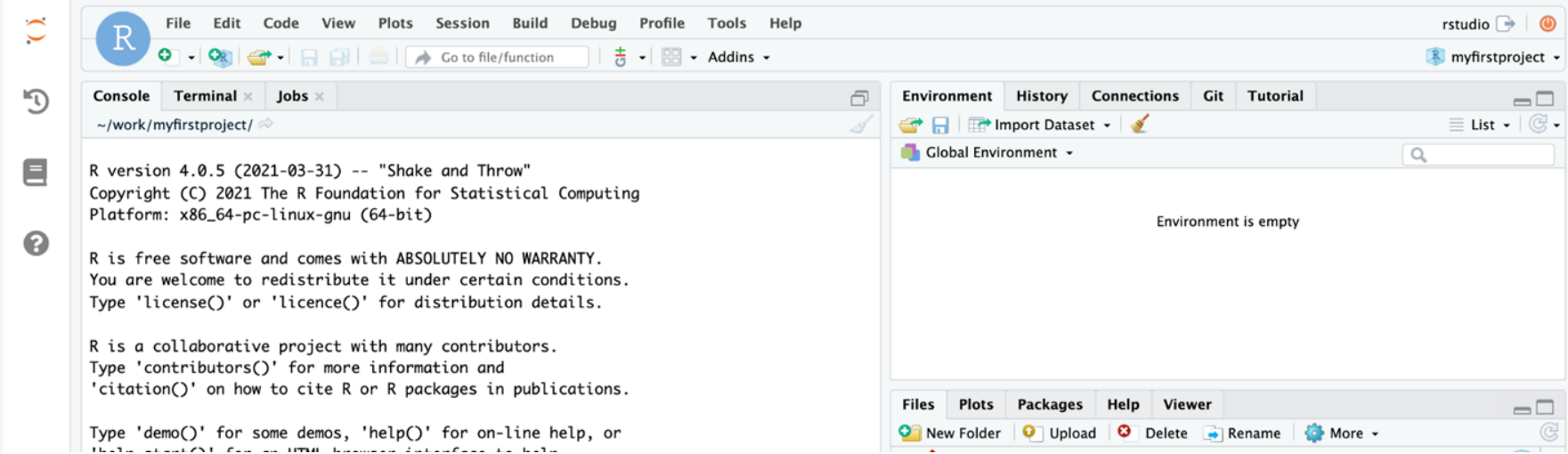
Start

Overview Collaboration Files Datasets Sessions Settings

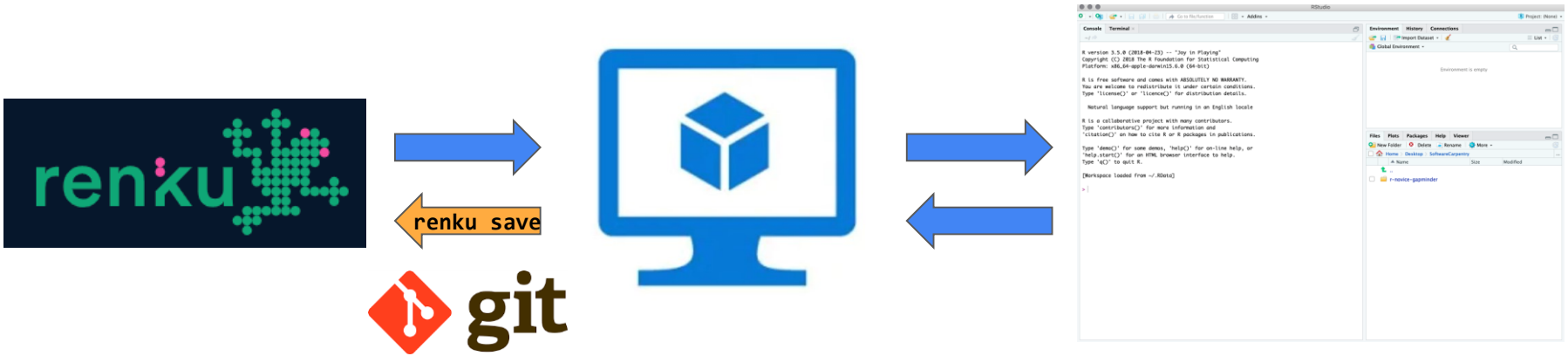
Back to sessions list

Branch master **Commit** 8fccde6b **Resources 0.25** cpu | **1G** memory | **1G** storage **Running since** 1 minute ago

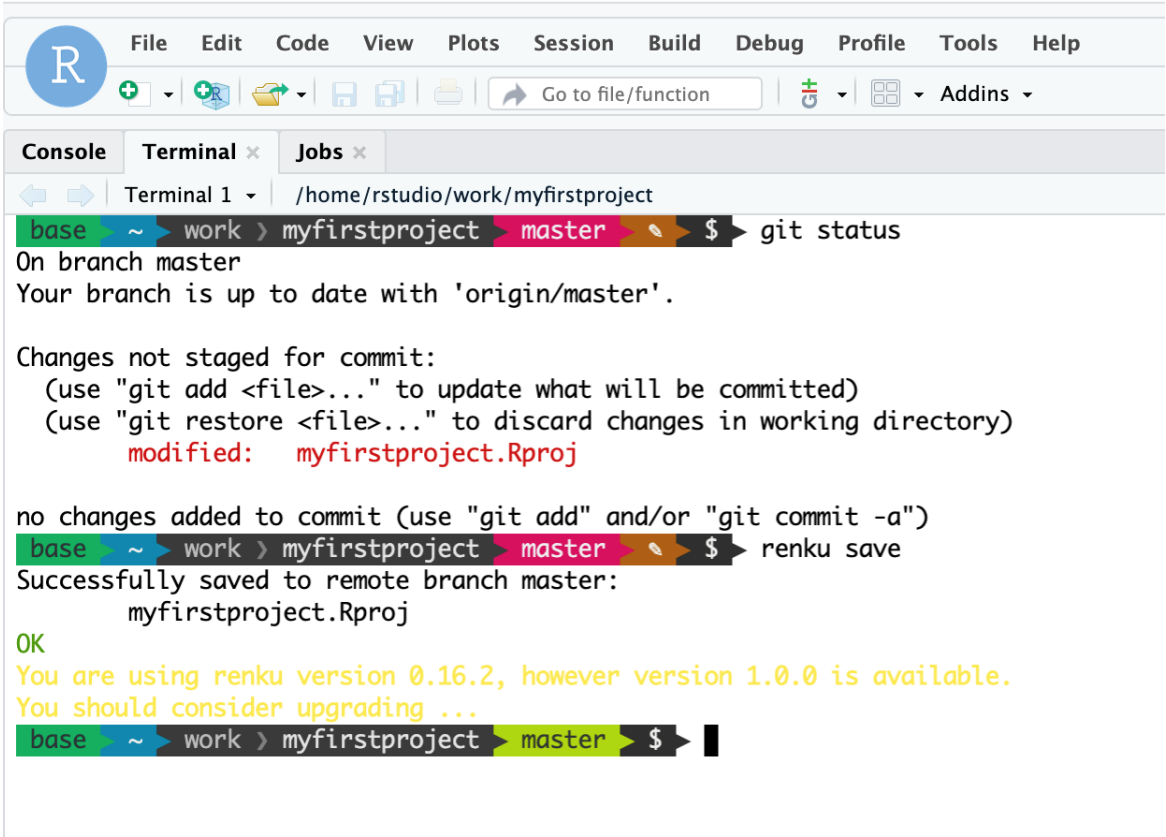
Open



`git` under the hood



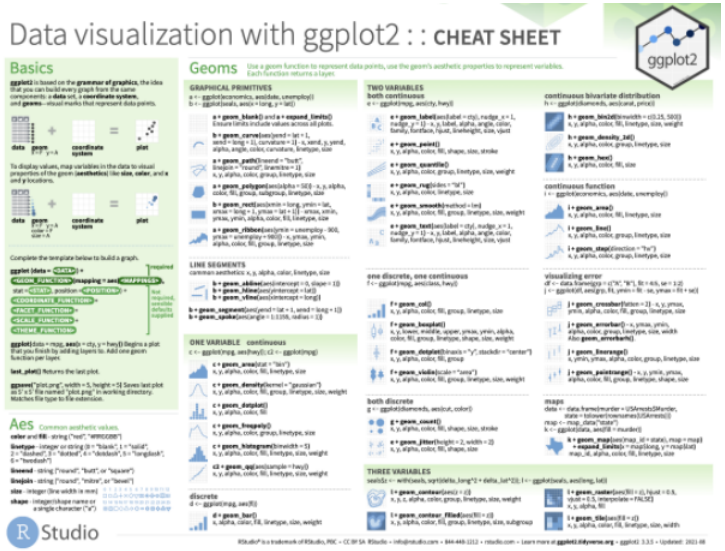
renku save


File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
Console
Terminal
Jobs
Terminal 1
/home/rstudio/work/myfirstproject
base ~ work ❯ myfirstproject master ✎ $ git status
On branch master
Your branch is up to date with 'origin/master'.
Changes not staged for commit:
(use "git add <file>..." to update what will be committed)
(use "git restore <file>..." to discard changes in working directory)
modified: myfirstproject.Rproj
no changes added to commit (use "git add" and/or "git commit -a")
base ~ work ❯ myfirstproject master ✎ $ renku save
Successfully saved to remote branch master:
myfirstproject.Rproj
OK
You are using renku version 0.16.2, however version 1.0.0 is available.
You should consider upgrading ...
base ~ work ❯ myfirstproject master $

Demo

Ressources and help

- Rstudio cheatsheets: https://www.rstudio.com/resources/cheatsheets/



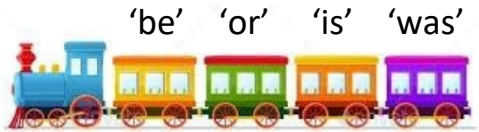
- Help from communities websites, e.g., Stackoverflow: https://stackoverflow.com/

Vectors

- Are a collection of multiple pieces of information. You can think of them as carriages linked together in a train.
- They can be made of numeric, character or factor data
- But each vector can only contain one type of data
- Creating a vector is easy – use the c function with brackets and separate the pieces of data using commas: x <- c(9,19,200,30,45)

numeric

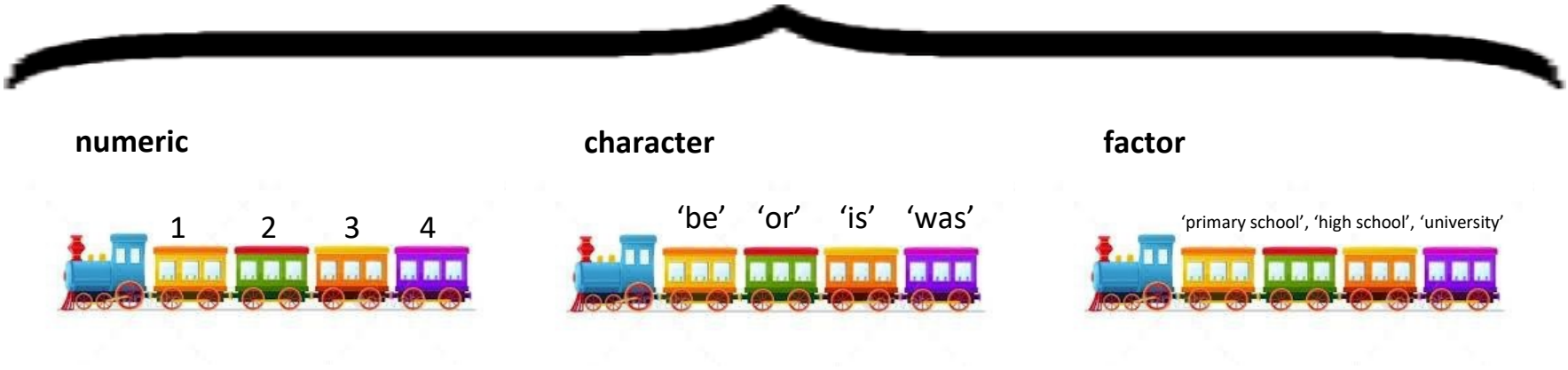
character



factor

Lists

- Like vectors are a collection of multiple pieces of information.
- They can be made up of individual pieces of data or collections of data: vectors, data frames, matrices, other lists
- Unlike vectors, lists can contain many types of data
- Example: one list made up of 3 vectors:



Matrices & Data Frames

Store data in rows and columns:
therefore are **2** dimensional (like tables – this means I need both the row and column number to find one piece of info!)

	Column 1	Column 2	Column 3	Column 4	Column 5
Row 1					
Row 2					
Row 3					
Row 4					