Statistics for Oncology

A Course for Scottish Trainees

by… The Edinburgh Cancer Informatics Research Group

https://edin.ac/oncology-statistics

INTRODUCTION TO R

Plan for this session

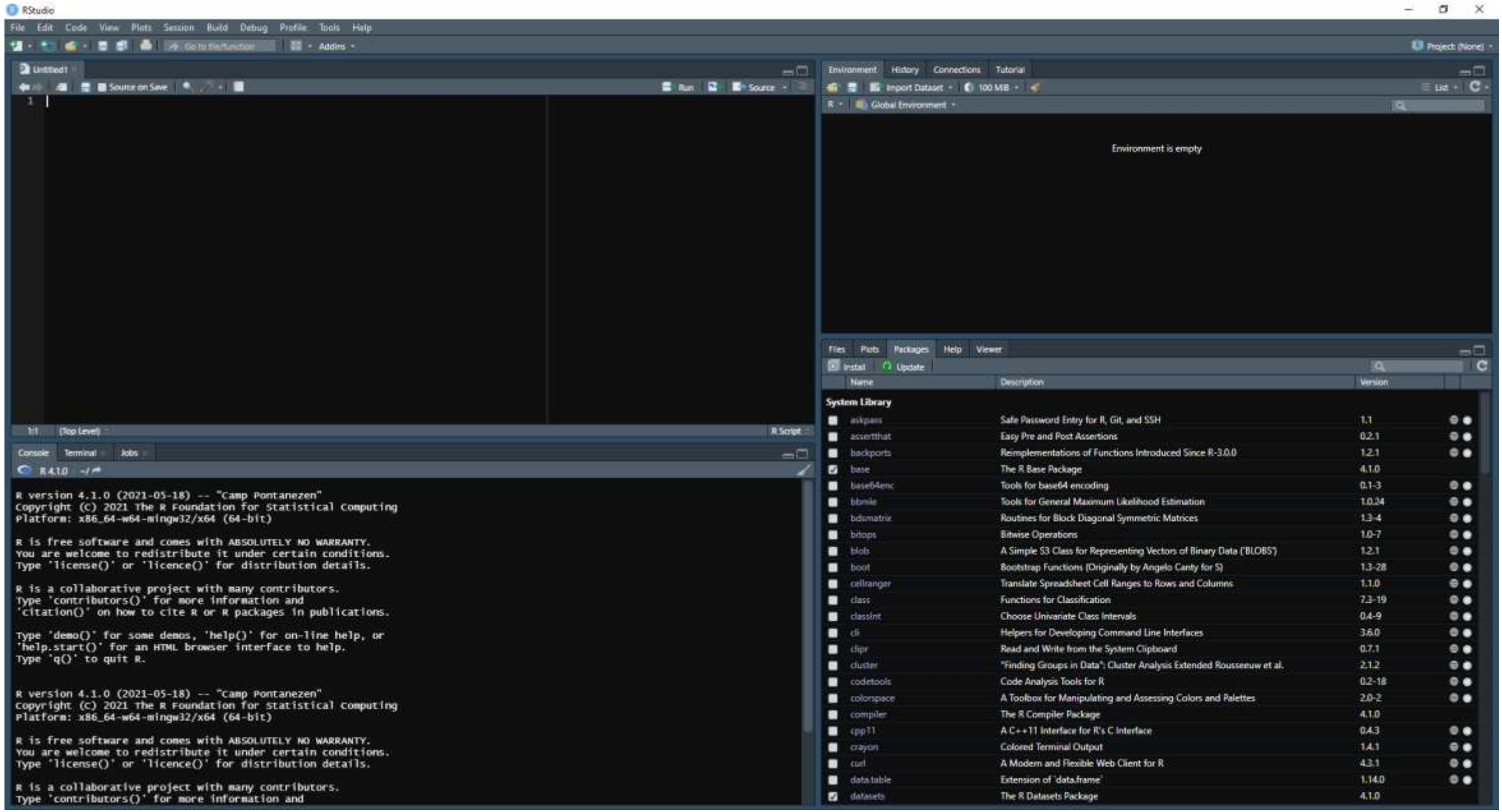
1. A compressed introduction to R in general
2. Advantages of doing analysis in R
3. Resources and Packages
4. A simple demonstration

What do we mean by R

R
Programming Language and Environment

CRAN
Comprehensive R Archive Network

RStudio
Integrated Development Environment


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Addins
Project: (None)
Untitled1
Source on Save
Run
Source
1
1:1 (Top Level)
R Script
Console Terminal Jobs
R 4.1.0 ~/
R version 4.1.0 (2021-05-18) -- "Camp Pontanezen"
Copyright (C) 2021 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)
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.
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 version 4.1.0 (2021-05-18) -- "Camp Pontanezen"
Copyright (C) 2021 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)
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.
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
Environment History Connections Tutorial
Import Dataset
100 MiB
List
R
Global Environment
Environment is empty
Files Plots Packages Help Viewer
Install
Update
Name
Description
Version
System Library
askpass Safe Password Entry for R, Git, and SSH 1.1
assertthat Easy Pre and Post Assertions 0.2.1
backports Reimplementations of Functions Introduced Since R-3.0.0 1.2.1
base The R Base Package 4.1.0
base64enc Tools for base64 encoding 0.1-3
bbmle Tools for General Maximum Likelihood Estimation 1.0.24
bdsmatrix Routines for Block Diagonal Symmetric Matrices 1.3-4
bitops Bitwise Operations 1.0-7
blob A Simple S3 Class for Representing Vectors of Binary Data ('BLOBS') 1.2.1
boot Bootstrap Functions (Originally by Angelo Canty for S) 1.3-28
cellranger Translate Spreadsheet Cell Ranges to Rows and Columns 1.1.0
class Functions for Classification 7.3-19
classInt Choose Univariate Class Intervals 0.4-9
cli Helpers for Developing Command Line Interfaces 3.6.0
clipr Read and Write from the System Clipboard 0.7.1
cluster "Finding Groups in Data": Cluster Analysis Extended Rousseeuw et al. 2.1.2
codetools Code Analysis Tools for R 0.2-18
colorspace A Toolbox for Manipulating and Assessing Colors and Palettes 2.0-2
compiler The R Compiler Package 4.1.0
cpp11 A C++11 Interface for R's C Interface 0.4.3
crayon Colored Terminal Output 1.4.1
curl A Modern and Flexible Web Client for R 4.3.1
data.table Extension of 'data.frame' 1.14.0
datasets The R Datasets Package 4.1.0

R essentials

The basic control of a process is through a script – a text file that lists the operations in their order of execution.
Within a script, we can assign values to variables, apply functions, load data tables, save files, call other scripts, and more.

Basic Data Types:

- Logical
- Integer
- Numeric
- Complex
- Character
- Raw

Other Types:

- Factor
- Double
- Date
- Difftime

Data Structures:

- Strings
- Lists
- Vectors
- Arrays
- Data Frames

Simple Operators:

- + addition
- - substraction
- * multiplication
- / division
- ^ exponent
- %% modulus
- %/% integer division

R essentials

Flow Control:

- if …. else
- while
- for
- break, next
- repeat

Logic:

- == equal to
- != not equal to
- >, <, >=, <=
- & and
- | or
- ! not

Utilities:

- <- assignment
- = assignment
- # comments
- "…" strings
- '…' strings
- `…` objects with white spaces
- [,:] indexing
- :: call to library
- $, @ access to attributes

Function definition:

- function(x) {}

Object Classes:

- S3 Class
- S4 Class
- Reference Class

Tidyverse: a dominant dialect

%>%
magrittr

Ceci n'est pas un pipe.

Useful Resources

- R code cheat sheets:
  - **base** https://github.com/rstudio/cheatsheets/blob/main/base-r.pdf
  - **dplyr** https://github.com/rstudio/cheatsheets/blob/main/data-transformation.pdf
  - **ggplot2** https://github.com/rstudio/cheatsheets/blob/main/data-visualization.pdf

- R code book: https://argoshare.is.ed.ac.uk/healthyr_book/

The book does a great job in getting you started (from scratch) in R using tidyverse, and teaches how to perform survival analysis and data visualization.