Using R for data analysis

2024-01-08

Contents

About this course

Placeholder

Teachers

Goals & Topics

Prerequisites

Materials

Programme

0.0.1 Fifth NKI edition, June 5th, 6th, 8th, and 9th 2023

Chapter 1

Basics of R

Placeholder

1.1 Data sets

1.1.1 Pulse

1.1.2 Survey

1.2 Introduction

1.2.1 Why this course?

1.2.2 What is R?

1.2.2.1 *R*: a short history

1.2.2.2 Open source

1.2.2.3 Obtaining R

1.2.3 R and RStudio

1.2.4 Course overview

1.2.4.1 What we teach

1.2.4.2 Course structure

1.3 Basics

1.3.1 R as a calculator

1.3.1.1 Calculations

1.3.1.2 Parentheses

1.3.1.3 Arithmetic functions

1.3.1.4 Multiline commands

1.3.1.5 (*) Integer division and remainders

1.3.2 Variables

1.3.2.1 Variable names

1.3.2.2 Legal variable names

1.3.2.3 Assignment

1.3.2.4 (*) The workspace

1.3.3 Vectors

1.3.3.1 Vector basics

1.3.3.2 Simple sequences

1.3.3.3 Simple selection

1.4 Projects and scripts

1.4.1 Projects

1.4.1.1 Creating and opening a new project

1.4.1.2 Creating a folder

1.4.1.3 Quick exercises

1.6.2.1.2 Relationships between two variables

1.6.3 Character data

1.6.3.1 Creating character data

1.6.3.2 Names

1.6.3.3 (*) Useful premade character vectors

1.6.3.4 (*) Calculating with character vectors

1.6.4 Factor - categorical data

1.6.4.1 Making a factor

1.6.4.2 Factor basics

1.6.4.3 Turning factor (back) into character

1.6.4.4 Table

1.6.4.5 (*) Table of table

1.7 Functions and help files

1.7.1 Function arguments

1.7.1.1 Functions may have multiple arguments

1.7.1.2 Getting help for a function

1.7.1.3 Specifying arguments by name

1.7.1.4 Default arguments

1.7.1.5 (*) the ... argument

1.7.2 Other help file aspects

1.7.2.1 Help search

Chapter 2

Data types, part 1

Placeholder

2.1 Basic data types (cont.)

2.1.1 Missing values

2.1.2 Numeric vectors

2.1.2.1 Single number (numeric vector with a single element)

2.1.2.2 Multiple numbers

2.1.2.3 Sequence of numbers (one by one)

2.1.2.4 Sequence of numbers

2.1.2.5 Combine (several) vectors of numbers

2.1.3 Character vectors

2.1.3.1 Simple single text (character vector with a single element)

2.1.3.2 Multiple texts

2.1.4 Logical vectors

2.1.4.1 Elementary logical values

2.1.4.2 Logical operators

2.1.4.2.1 Negation

2.1.4.2.2 AND Operator `&`

Chapter 3

Data types, part 2

Placeholder

3.1 R scripts and reports (Rmarkdown)

3.1.1 Markdown : R code + Markdown text formating language

3.1.1.1 Short session

3.1.2 Features of R Markdown

3.1.2.1 Markdown examples

3.1.2.2 Including R code in markdown

3.1.2.3 Chunk options

3.1.2.4 R Studio and RMarkdown

3.1.2.5 Try this!

3.2 Lists

3.2.1 What is a list

3.2.2 Creation

3.2.3 Length

3.2.4 Getting names of the elements

3.2.5 Getting a single element

3.2.6 (*) Getting multiple elements as a list

3.2.7 (*) Removing an element

3.2.8 (*) Adding an element

3.3 Basic statistical tests

3.3.1 Statistical methods with R

3.3.1.1 Statistical tests - the t test

3.3.1.2 Accessing the result

3.3.1.3 Wilcoxon test

3.3.1.4 Other statistical tests

3.4 Regression and formula objects

3.4.1 Formula objects

3.4.1.1 Basics of formula objects

3.4.1.2 Use of formula objects in statistics

3.4.1.3 (*) The formula class

3.4.2 Simple linear regression

3.4.2.1 The `lm` function

3.4.2.2 Visualizing a regression

3.4.3 Multiple regression and prediction

Chapter 4

Functions

Placeholder

4.1 User-defined Functions

4.2 (*) Control flow constructs

4.2.1 Sequential execution

4.2.2 Conditional execution: if

4.2.3 Conditional execution: if/else

4.3 `apply` family: apply, lapply, sapply, tapply

4.3.1 apply(X, MARGIN, FUN, ...)

4.3.2 sapply(X, FUN, ...)

4.3.3 lapply(X, FUN, ...) : apply a function to a list/vector

4.3.4 tapply(X, INDEX, FUN, ...) : apply a function to a vector, according to groups of INDEX

4.4 Type checking

4.4.1 is.* family functions

4.4.2 (*) stop(...)/warning(...)

4.5 (*) R programming

4.5.1 General coding conventions

4.5.2 Finding, installing and loading packages

Appendix

Placeholder

Character string processing & Pattern matching

Learning goals

Quotes and escape characters (characters with special meaning)

Character string

paste(. . . , sep = ” “, collapse = NULL):

Pulse data set

tolower, toupper, nchar

Split character string : strsplit(x, split, . . .):

rbind(. . .), cbind(. . .) : combine by rows or columns

strsplit(x, split, . . .)

do.call(what, args, . . .) : execute function on list of arguments

grep(pattern, x, ignore.case = FALSE, value = FALSE, . . .)

grepl(pattern, x, ignore.case = FALSE, . . .)

substr(x, start, stop) : extract/replace substrings

gsub/sub(pattern, replacement, x, ignore.case = FALSE, . . .)

Extra exercises

S3 and S4 classes

Learning goals

Objects

S3

S4

Making errors the right way

Annoying errors

How to correct (annoying) errors

Dangerous errors

How to avoid (dangerous) errors

Useful links