

Using R for data analysis

2024-01-08



# Contents

<b>About this course</b>	<b>5</b>
Teachers . . . . .	5
Goals & Topics . . . . .	5
Prerequisites . . . . .	5
Materials . . . . .	5
Programme . . . . .	5
<b>1 Basics of R</b>	<b>7</b>
1.1 Data sets . . . . .	9
1.2 Introduction . . . . .	9
1.3 Basics . . . . .	9
1.4 Projects and scripts . . . . .	9
1.5 Data frames (basics) . . . . .	9
1.6 Basic data types . . . . .	9
1.7 Functions and help files . . . . .	10
<b>2 Data types, part 1</b>	<b>11</b>
2.1 Basic data types (cont.) . . . . .	11
2.2 Vectors . . . . .	12
2.3 Data frames . . . . .	12
2.4 Matrices . . . . .	13
<b>3 Data types, part 2</b>	<b>15</b>
3.1 R scripts and reports (Rmarkdown) . . . . .	17
3.2 Lists . . . . .	17
3.3 Basic statistical tests . . . . .	17
3.4 Regression and formula objects . . . . .	17
3.5 Factors (advanced) . . . . .	17
<b>4 Functions</b>	<b>19</b>
4.1 User-defined Functions . . . . .	20
4.2 (*) Control flow constructs . . . . .	20
4.3 apply family: apply, lapply, sapply, tapply . . . . .	20

4.4	Type checking . . . . .	20
4.5	(*) R programming . . . . .	20
<b>Appendix</b>		<b>21</b>
	Character string processing & Pattern matching . . . . .	22
	S3 and S4 classes . . . . .	22
	Making errors the right way . . . . .	22
	Useful links . . . . .	22

# 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 &

#### 2.1.4.2.3 OR operator |

#### 2.1.4.3 Relational operators

##### 2.1.4.3.1 Equality operator ==

##### 2.1.4.3.2 Inequality operators

#### 2.1.4.4 Comparison of two vectors

#### 2.1.5 Type conversions

### 2.2 Vectors

#### 2.2.1 Square brackets operator

##### 2.2.1.1 By numbers

##### 2.2.1.2 By names

##### 2.2.1.3 By condition (logical indices)

#### 2.2.2 Other useful functions

### 2.3 Data frames

#### 2.3.1 What is a data frame

#### 2.3.2 Creating

##### 2.3.2.1 From manually provided vectors

##### 2.3.2.2 Import from files

###### 2.3.2.2.1 TSV file

###### 2.3.2.2.2 CSV file

###### 2.3.2.2.3 (\*) Microsoft Excel file

###### 2.3.2.2.3.1 (\*) With package readxl

###### 2.3.2.2.3.2 (\*) With package gdata

**2.3.2.3 (\*) SPSS files**

**2.3.3 Properties**

**2.3.3.1 Dimensions**

**2.3.3.2 Columns/rows names**

**2.3.4 Content**

**2.3.4.1 Dollar operator**

**2.3.4.2 Square brackets operator, single argument**

**2.3.4.3 Square brackets operator, two arguments**

**2.4 Matrices**

**2.4.1 What is a matrix**

**2.4.2 Creation**

**2.4.3 Dimensions**

**2.4.4 Setting/getting names of the columns and rows**

**2.4.5 Getting matrix elements**

**2.4.6 Useful matrix functions**

**2.4.6.1 Row/columns means/sums**

**2.4.6.2 (\*) Transposition**

**2.4.6.3 (\*) Matrix multiplication**

**2.4.6.4 (\*) Element-wise multiplication**



## 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

#### 3.4.2.3 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