Business and Local Government
Data Research Centre

Introduction to R

EXPLORING DATA
ENHANCING KNOWLEDGE
EMPOWERING SOCIETY

E·S·R·C
ECONOMIC
& SOCIAL
RESEARCH
COUNCIL

University of Essex

Business and Local Government Data Research Centre

Cutting-edge training delivered by leading experts in the field of data analytics brought to you by the Business and Local Government Data Research Centre

- Exploring data
- Enhancing knowledge
- Empowering society

Offering grant funded data analytics research projects, training, events, webinars and data consultation services.

To find out more please contact Laura.brookes@essex.ac.uk

Agenda

1. Introduction
2. R-Syntax, Data structures & types
3. Data import
4. Creating, amending, exporting data frames
5. Type cohersion
6. Loops and conditions

Introduction to the session

- For the hands-on session visit the following web page: https://philippbroniecki.github.io/suffolk2019/

What is R?

- An environment for statistical computing and graphics
- It is free
- It packs powerful graphical facilities
- It is a simple and effective programming language
- Most statistical models are already implemented
- New models are often implemented in R first

Visualization

Open source

Data science

Platform agnostic

Computational
statistics

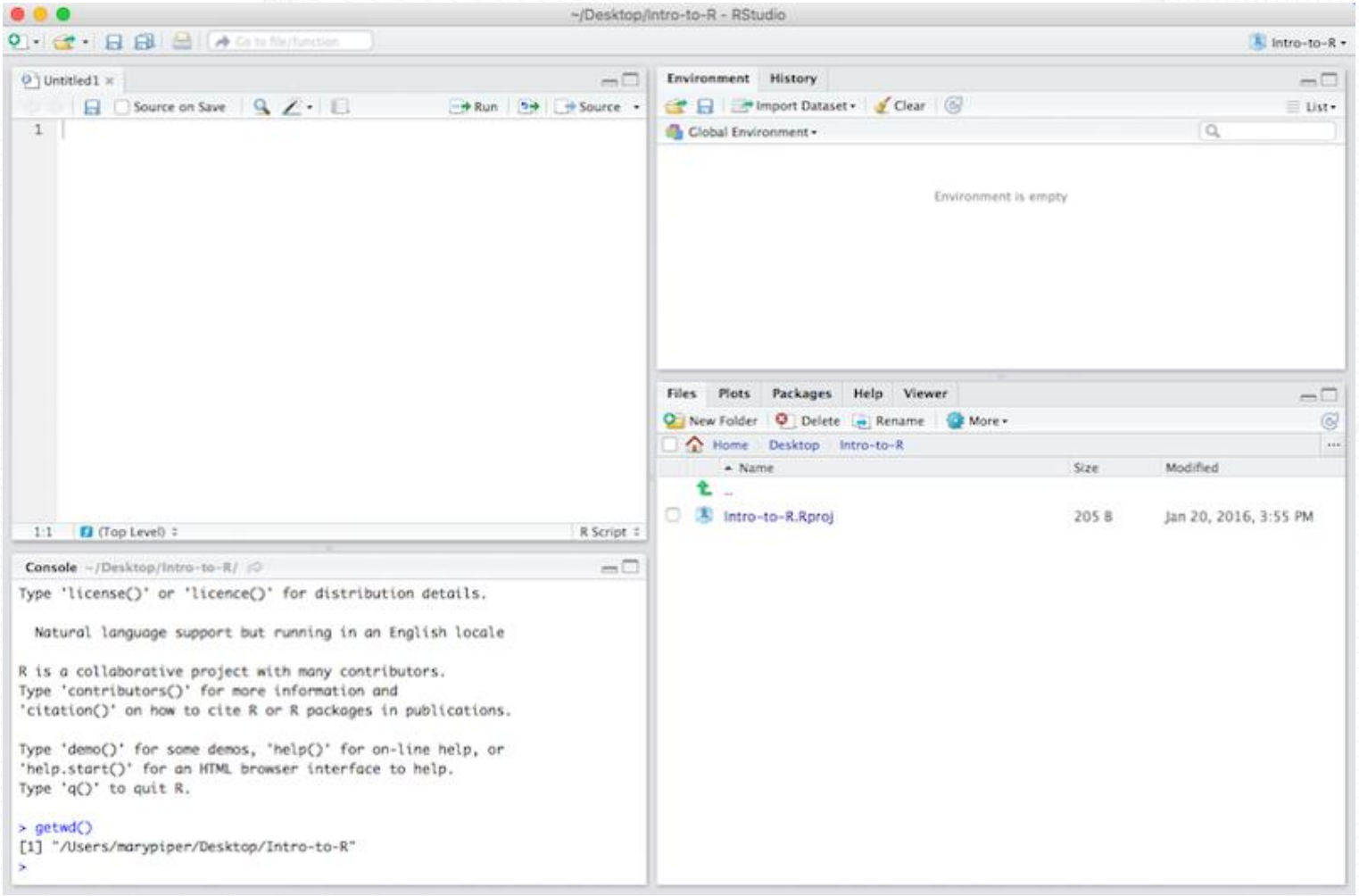
Home of R and help on R questions

- Home of all things R: https://cran.r-project.org/
  - Keep R updated – check for new versions

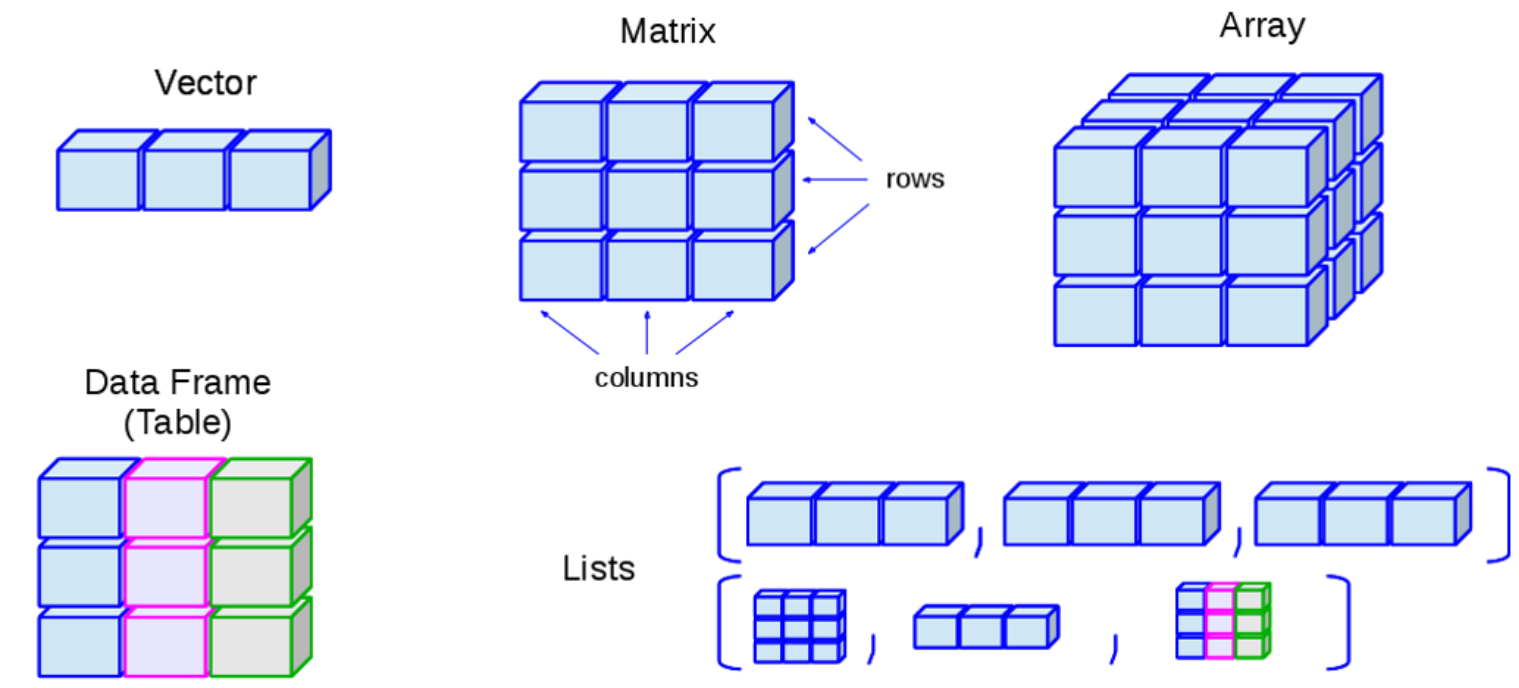
- To get help:
  - Google the question or error message is always a good start
  - https://stackoverflow.com/questions/tagged/r

R Studio

- Working environment R
  - While you do not need R Studio to run R, it makes working with R much easier
  - Just like R, it works on PC, Mac & Linux


~/Desktop/Intro-to-R - RStudio
Intro-to-R
Untitled1
Source on Save
Run
Source
1
1:1
(Top Level)
R Script
Environment
History
Import Dataset
Clear
List
Global Environment
Environment is empty
Files
Plots
Packages
Help
Viewer
New Folder
Delete
Rename
More
Home
Desktop
Intro-to-R
Name
Size
Modified
..
Intro-to-R.Rproj
205 B
Jan 20, 2016, 3:55 PM
Console ~/Desktop/Intro-to-R/
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.
> getwd()
[1] "/Users/marypiper/Desktop/Intro-to-R"
>

R-Syntax, data structures and types



http://venus.ifca.unican.es/RIntro/dataStruct.html

Data structures and types

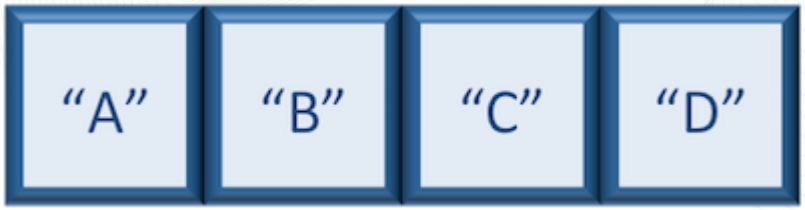
- Vectors: one-dimensional; store collection data of same mode; one coordinate to identify a unique vector element):
  - Numeric vectors (mode: numeric)
  - Logical vectors (mode: logical)
  - Character vector or vector of strings (mode: character)
  - Factors are vectors for categorical variables that assign a label to each category
- Matrices: two dimensional, store data of same mode; two coordinates to identify a unique matrix element
- Arrays: n-dimensional; store data of same mode; n coordinates needed to identify a unique array element
- Lists: ordered collection of objects, elements can be of different types
- Data frames: Matrices that can store data of different modes
  - For Analysis: Always place observations in rows & variables in columns
  - Call missing values for numeric mode NA and for string mode "”

Vectors

- Numeric vectors
  - a <- 5
  - a <- c(1, 50, 9, 42)

- Logical vectors
  - b <- a < 10

- Character vectors
  - a <- "this is text"



Matrix

```
> # create a matrix with 3 rows and 4 columns
> a <- matrix( data = c(90, 5, 137, 9, 87, 40, 2, 52, 4, 102, 32, 41),
+              nrow = 3,
+              ncol = 4,
+              byrow = TRUE
+              )
> a
     [,1] [,2] [,3] [,4]
[1,]   90    5  137    9
[2,]   87   40    2   52
[3,]    4  102   32   41
> # check the dimensions of a
> dim(a)
[1] 3 4
```

90	5	137	9
87	40	2	52
4	102	32	41

Array

```
> # create an array with 2 rows, 3 columns & 4 layers
> a <- array(data = 1:24, dim = c(2, 3, 4))
> a
, , 1

     [,1] [,2] [,3]
[1,]    1    3    5
[2,]    2    4    6

, , 2

     [,1] [,2] [,3]
[1,]    7    9   11
[2,]    8   10   12

, , 3

     [,1] [,2] [,3]
[1,]   13   15   17
[2,]   14   16   18

, , 4

     [,1] [,2] [,3]
[1,]   19   21   23
[2,]   20   22   24

> # check dimensions of a
> dim(a)
[1] 2 3 4
```

Data frame

- Each row is a vector
- Each column is a vector
- All rows must have equal length
- All columns must have equal length

“A”	102	“Hela”	TRUE
“B”	40	“BHK”	F
“C”	12	“hESC”	T

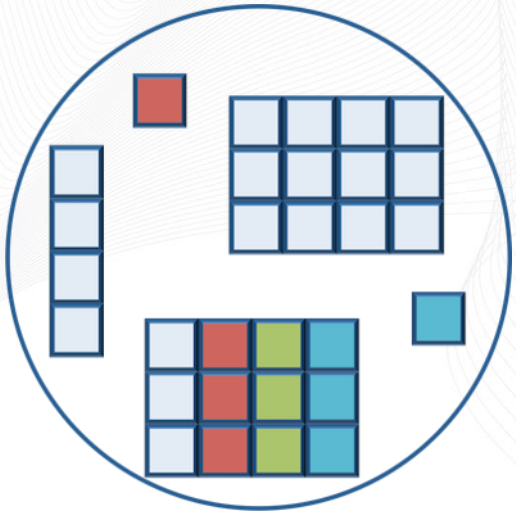
Data frame

- We will use data frames for analysis
- Data frames can store different data modes (numeric, factor, character)
- Data frames are rectangular
  - All rows have the same length
  - All columns have the same length

List

- Lists hold any number of any types of other data structures
- If you have variables of different data structures you wish to combine you can put all of those into one list object by using the list() function and placing all the items you wish to combine within parentheses:

- a <- list(object1, object2, object3)



Math functions

`log(x)`	Natural log.	`sum(x)`	Sum.
`exp(x)`	Exponential.	`mean(x)`	Mean.
`max(x)`	Largest element.	`median(x)`	Median.
`min(x)`	Smallest element.	`quantile(x)`	Percentage quantiles.
`round(x, n)`	Round to n decimal places.	`rank(x)`	Rank of elements.
`signif(x, n)`	Round to n significant figures.	`var(x)`	The variance.
`cor(x, y)`	Correlation.	`sd(x)`	The standard deviation.

Business and Local Government Data Research Centre

Thank you

Join in the conversation online:

@BLGDataResearch #Data2Life

LinkedIn: ESRC Business and Local Government Data Research Centre

YouTube: ESRC Business and Local Government Data Research Centre

Email: BLGDataResearch@essex.ac.uk

EXPLORING DATA
ENHANCING KNOWLEDGE
EMPOWERING SOCIETY

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

University of Essex