Introduction to R for Biologists cheatsheet

glossary of vocabulary used in the course

Maria Doyle

03 October 2019

Packages

dplyr
tidyverse package for manipulating data, contains the `mutate()`, `filter()`, `select()`, `full_join()` functions and the `%>%` operator

ggplot2
tidyverse package for data visualisation

readr
tidyverse package for reading data into R, contains the `read_tsv()` function

tidyr
tidyverse package for data tidying, contains the `gather()` function

tidyverse
a collection of packages that work together for data reading, tidying, manipulating and visualising

Functions

`c()`
combine values (from base R)

`case_when()`
test multiple conditions (from dplyr), useful inside `mutate()` when creating columns

`colnames()`
access column names (from base R)

`colours()`
see the built-in R colours (from base R)

`dev.off()`
turn off the R graphics device (from base R). Used after e.g. `pdf()`, `png()`.

`dim()`
retrieve the dimensions of an object, for example, the number of rows and columns (from base R)

`factor()`
convert values to factor data type (from base R)

`filter()`
choose rows (from dplyr)

`full_join()`
joins 2 tables returning all rows and all columns from both tables.

`gather()`
function that enables converting from wide to long (tidy) format (from dplyr)

`getwd()`
find out what the working directory is (get working directory)

`ggplot()`
function used to create a ggplot (from ggplot2)

`head()`
selecting the first part of an object (from base R). Default is to show the first 6 items.

`labs()`
modify title, axis and legend labels on a ggplot (from ggplot2)

`levels()`
retrieve the levels (category names) of a factor (from base R)

`library()`
load packages (from base R)

`log2()`
compute the log2 (base 2) logarithms (from base R)

`mutate()`
add columns (from dplyr)

`pdf()`
create a pdf, used with `dev.off()` (from base R)

`pull()`
extract values e.g. out of a column (from dplyr)

`read_csv()`
read a comma-separated file into R (from readr)

`read_tsv()`
read a tab-separated file into R (from readr)

`select()`
choose columns (from dplyr)

`str()`
showing the structure of an object (from base R). Useful for checking data types.

`summary()`
producing a summary of an object (from base R). Useful for getting summary statistics of numeric columns (min, max, mean, median)

`tail()`
selecting the last part of an object (from base R). Default is to show the last 6 items.

`View()`
invoke a spreadsheet-like viewer on an R object (from base R)

Terms

argument
an input to a function

assignment operator
`<-` assigns values to objects, assigns a value on the right to an object on the left (from base R)

character
a data type in R, used to represent character strings, quotes indicate the data type is character

console
a window where you can interactively type in commands and the output is returned

double
a data type in R, used to represent numbers containing a decimal point (integer is the data type for numbers without decimal point)

factor
a data type in R, used to represent categories

function
a pre-defined set of commands used to perform a task, can be loaded in from packages or user-created

geom
type of ggplot e.g. `geom_line()`, `geom_point()`, `geom_jitter()`, `geom_boxplot()`, `geom_violin()`

object
everything in R is an object. The assignment operator `<-` can be used to create objects. Note that what R calls objects are called variables in other languages such as Python.

package
a package is a collection of functions and usually includes code, documentation, tests and example datasets.

pipe
`%>%` operator chains together tidyverse commands (from dplyr)

scales
scale_fill_manual(), scale_colour_manual(), scale_fill_brewer(), scale_colour_brewer(). Use to specify colours.

script
a text file containing commands, in R a script filename ends with .R

themes
the non-data components of a ggplot e.g. background, grid lines, font size and font type

working directory
the location (path) where R looks to read in data and save files

Symbols

`>`
prompt in console, means R is ready to take a command

`+`
used to add layers to a ggplot. Also the prompt symbol R uses when the command is not complete, such as missing a `)`

`<-`
assignment operator, see Terms above

`#`
comment, to add notes to a script

`$`
way to access a single column with base R e.g. `counts$gene_symbol`

`%in%`
operator used to test if a value is in a set of values

`%>%`
dplyr pipe operator, see Terms above

`~`
symbol to use when faceting in ggplot2, used to indicate the column to use to facet