R cheat sheet

Tips

- `?command` brings up the help page for the command.
- Command/control-return runs the current line/selection of your script. In the console, return runs your code.
- RStudio has tab-completion of function and variable names.

Basics

- `a <- b` computes `b` and assigns it to the variable `a`.
- `View(x)` (capital V!) displays `x` nicely.
- Basic math works as expected.
- `rm(x)` gets rid of variable `x`

Everything is a vector!

R's basic types are inherently "vectors," meaning some ordered list of things.

- `c(...)` joins the arguments into a vector
- `from:to` returns a sequence
- `typeof(x)` returns the type of `x`
- `length(x)` returns the length of `x`
- The index of the first element is 1!

`x[n]`	`n`-th element
`x[1:n]`	first `n` elements
`x[c(1,4,2)]`	specific elements, in a specific order
`x[x > 3]`	just the elements > 3
`x[order(x)]`	sorted
`rev(x)`	reversed
`unique(x)`	without duplicates

`as.integer()`, `as.double()`,[1] `as.character` convert types

NA

Missing data is encoded using the magical value `NA`. `is.na(x)` tests whether `x` is an `NA`.

Comparison and truth values

`TRUE` is true and `FALSE` is false.

`a == b`	$a = b$
`a != b`	$a \neq b$
`!p`	$\neg p$
`p & q`	$a \wedge b$
`p | q`	$a \vee b$
`x %in% vector`	whether or not `x` is an element of `vector`

These also apply to vectors, point-wise.

Data frames

A **data frame** has multiple *rows* of data. Each *column* can be named. If `dataframe` has a column named `Name`, `dataframe$Name` returns all the `Name` values in `dataframe` as a vector.

- `row.names(dataframe)` confusingly refers to the column names!
- `rbind(dataframe, row)` returns a copy of `dataframe` with vector `row` added to the bottom as a new row
- `nrow(dataframe)` = # of rows
- `length(dataframe)` = # of columns
- `subset(dataframe, [condition])` returns the rows of `dataframe` which satisfies the condition. The condition can refer to column names directly.

Manipulating and computing

- `aggregate(x, list(...), FUN)` groups elements of `x` by corresponding values in the list, and then runs the function FUN on each group.
- `table(x)` summarizes count data
- `merge(a,b)` merges two data frames by common columns names

Import/export

- `read.csv(filename)` reads `filename` CSV file and returns a data frame
- `write.csv(filename, x)` saves the data frame `x` to `filename` in CSV format

In RStudio, you can also import data in Tools > From Text File, which has a nice interface.

[1] A "double" is a number which can have decimals.

mitcho@mitcho.com