

## Tutorial 09 - Variables, functions, data structures, and input/output

Do you have R installed?

Please see your TA as soon as possible if you do not have R and RStudio installed!

# When to use what data structure in R

To some extent this is based on personal preference or what a particular function in R requires as its input.

There are, however, some clear cases where one data structure is the right one for the job:

- ▶ a list is useful for information that is linked in someway (e.g. from the same analysis or observation), but differs a lot in its “shape” (vectors vs. matrices)
- ▶ a dataframe is probably most often used because it holds tabular data with different data types and is the default data structure used when data is read in from a file

## Terminal tab in RStudio

The Terminal tab in RStudio allows you to interact with `bash`.

This is helpful because you can use `git` and interact with github this way.

In OSX you should be all set by clicking on the *Terminal* tab.

In Windows you'll need to run `bash` (type `bash` and press Enter) to access the file system and `git` tools.

Try cloning the Exercise 7 repository using the Rstudio Terminal!

## Lecture Challenge - Starting with R

1. By default what data structure is created by `read.csv`?
2. Create a vector containing integers from 1 to 8. Now append the numbers 12, 14, 16 to the end of that vector.
3. Talk to at least two of your neighbors and find out their name, favorite color, and favorite number. Enter this information, along with your name, favorite color, and favorite number in an appropriate data structure.
4. Put your vector from #2 and data from #3 into a list.
5. What are the dimensions of the three data structures you created in challenges 2-4?

## Loading a file

Read in the iris data set from `iris.csv`

## Loading a file

Read in the iris data set from `iris.csv`

Don't forget to set your working directory:

```
setwd("~/Documents/biocomputing/tutorial09")
```

One can use `read.csv()` or `read.table()`

```
read.csv(file="iris.csv",header=TRUE)
```

```
read.table(file="iris.csv",header=TRUE,sep=",")
```

Remember to put the output of `read.csv()` into a variable!

## Write a file

- ▶ Create a 20-row and 5-column matrix with the numbers 1 to 100.
- ▶ Add column names “col\_one”, “col\_two”, “col\_three”, “col\_four”, and “col\_five” to the matrix.
- ▶ Write this matrix to a tab-delimited file.



## data modes and arithmetic

- ▶ Appending to a vector and data modes
  - ▶ Create a vector containing the numbers 10, 15, 32, 64
  - ▶ Find the sum of all elements in this vector
  - ▶ Create a new vector with the original vector and your name as the last element of your vector
  - ▶ Find the sum of all elements of the new vector. Did you get an error? Why?
  - ▶ Try to solve the error on your own using Google
- ▶ If you are stuck, try googling “how to force a character to numeric in R”

## Appending to data structures

- ▶ Create a vector containing the integers 1 to 5; now append the integers 5 to 1 to the end of this vector
- ▶ Create a  $5 \times 2$  matrix; now add a 3rd column with all zeroes to the matrix
- ▶ Create a data frame with a column containing four animals and a column containing the approximate weight of each animal in kg; now add a column to your animal data frame with a logical data mode, TRUE for rows containing a mammal and FALSE for rows containing a non-mammal