

# Intro to R: Review

This document and the data in this example can be found at:

[https://software.rc.fas.harvard.edu/ngsdata/workshops/2015\\_March](https://software.rc.fas.harvard.edu/ngsdata/workshops/2015_March)

or on the cluster at:

`/n/ngsdata/workshops/2015_March`

## Getting Help

R has very good built-in documentation that describes what functions do.

To get help about a particular function, use `?` followed by the function name, like so:

`?read.table`

## Vectors

Create a vector with `c()`:

```
x<-c(1,2,3,4,5)
```

Type the name of an object (in this case `x`) to display it.

Some basic operations with vectors:

```
y = x+5  
z = 4  
y*z
```

```
## [1] 24 28 32 36 40
```

Below, `logic1` is a logical vector:

```
logic1<-c(TRUE, TRUE, FALSE, FALSE, TRUE)
```

Use this logical vector to select values from our vector `x`:

```
x[logic1]
```

```
## [1] 1 2 5
```

## Matrices

Create a matrix:

```
m = matrix( c(13, 42, 6, 3, 124, 40), nrow = 2, ncol = 3, byrow = TRUE)  
m
```

```
##      [,1] [,2] [,3]  
## [1,]  13  42   6  
## [2,]   3 124  40
```

Extract the second row as a vector:

```
m[2,]
```

```
## [1]   3 124  40
```

Extract the third column as a vector:

```
m[,3]
```

```
## [1] 6 40
```

Create a new matrix from all rows, but only the first two columns of m:

```
m[,1:2]
```

```
##      [,1] [,2]
## [1,]  13   42
## [2,]   3  124
```

Alternatively, we could have extracted all rows and columns 1 and 2 of m with:

```
m[,c(1,2)]
```

## Data frames

Data frames are like matrices, but where the columns are considered to be samples, and rows are considered to be the observations comprising each sample. Many functions which take a data frame as input will make this assumption about the meaning of the rows and columns.

Create a data frame from gene count data:

```
filePath='http://software.rc.fas.harvard.edu/ngsdata/workshops/2015_March/fruitfly.gene_counts.allsamples.tsv'
d = read.table(file = filePath, header = TRUE, row.names = 1, sep = '\t')
```

Use the following functions to view information about the data frame:

```
class(d)      #data type
dim(d)        #number of rows and columns
print(d)      #print the data frame to the screen
str(d)        #show the first few data points of each sample in the data frame
head(d)       #view the first six rows
summary(d)    #view some basic statistics (mean, median, etc) of each sample in the data frame.
```

Extract the second row of the data frame, as a vector:

```
d[2,]
```

Extract the fifth sample, as a data frame:

```
d[4]
```

Extract rows 10 to 20 of the third column, as a vector:

```
d[10:20,3]
```

## Write data to a file

Create a directory and write your data frame to a tab-separated file:

```
project.dir <- '~/My_R_Example'
dir.create(project.dir, showWarnings=FALSE)
write.table(d, file = file.path(project.dir,'Control_vs_Infected.tsv'), quote = FALSE, sep = '\t')
```