Created by JPWR, 20-Jan-2015

Data formats & how to import data

1. How do we import data into R?

Usually we work with *comma-separated values* - this format is simple, light-weight and can be read into lots of programs, including R.

Much of time we'll start with an excel spreadsheet (.xls format) - R can't handle this very well, and we'd rather work with the simpler **.csv** format.

If you do get an excel file to work with, simply open it up in Excel, click File -> save as and save as a comma-separated file (.csv).

But here we have a csv file already. Move it to your working directory (setwd() before you start) and type:

```
isle_dat<-read.csv(file="/Users/james/Dropbox/learnR/IsleRoyal_Data.csv")
```

Now isle_dat is loaded into your workspace and we can begin to explore the dataset, make plots, run analyses etc. Have a look at the first 5 rows...

```
head(isle_dat)
tail(isle_dat)
```

2. Data structures in the R workspace

There are many data types in R. 4 of the most commonly used are:

1. Vector
2. Matrix
3. Data frame
4. Lists

Here for more details: http://www.statmethods.net/input/datatypes.html

Vectors are sequences of data elements of one type.

```
a.vec<-c(1:10)          ### Create a vector named 'a.vec' of numbers from 1 to 10
mean(a.vec)  # Mean of all values of a.vec
```

Matrices must fit one data type (all numbers, or all words). These are most used to store modelling results (we'll use matrices for population models in mid-February)

```
a.mat<-matrix(1:50, nrow=10, ncol=5)
dim(a.mat)  ### what size is the matrix?
```

Data frames are combinations of numbers and words. Observational and experimental datasets should be kept as data frames.

```
animal=c("mouse", "mouse", "rat")
mass=c(10, 15, 13)
site=c(1,1,1)
treatment=c("TRUE", "TRUE", "FALSE")
a.data<-data.frame(animal, mass, site, treatment)
names(a.data)
```

Lists are a collection of objects - these are useful for storing data frames of different sizes. We won't use lists in this course (but they are very useful!).

3. Data formats in the R workspace

There are 4 major data attributes in R.

1. Numeric
2. Factor
3. String
4. Character
5. Logical

How do we check what our workspace objects are?

```
class(isle_dat)   ### What format is isle_dat?
is.numeric(a.vec) ### Logical statement. Is a.vec a numeric vector?
is.factor(a.vec)   ### Logical statement. Is a.vec a factor vector?
```

How do we change the data format of an object?

```
isle_dat_matrix<-as.matrix(isle_dat)  ### Turn data frame object into matrix
a.vec_factor<-as.factor(a.vec) ### Change numeric vector into factor vector
```

How do we explore the characteristics of an object?

```
dim(isle_dat)   ### What size is the data frame?
levels(a.vec_factor)   ### How many 'levels' in the factor - how many categories?
length(a.vec)   ### How many values in the vector?
str(isle_dat)   ### Size of dataframe, format of variables
```

4. What's a package?

Packages (or libraries) are like expansion packs for R - they offer new functions, new statistical approaches, and new plot types. Each package comes with a user manual that explains how to use the new functions.

To use a package *for the first time* we type

```
install.packages("vegan")
require(vegan)
```

Once the package is installed, we can just type:

```
require(vegan)
?vegan
```

Some useful packages:

- **ggplot2** - Data exploration and quick plot creation
- **plyr** - Useful for working with large datasets
- **vegan** - Community ecology functions