

'R' is a powerful freeware (ie you can download it for free!) computer package that makes doing statistics easier - it does the boring numerical calculations which frees you up to do some *thinking*..

In this introductory session we will learn:

1. How to start 'R'.
2. How to insert data into 'R'.
3. Some of the many different ways of looking at data using 'R' graphics.
4. How to access one of the many strange data sets already inside 'R'.

I. Starting R

Once you've successfully logged in, starting 'R' is as simple as *clicking* on the 'R' icon on the desktop. Several windows will open up and you are ready to do some statistics!

You communicate with 'R' through the 'R Console Window'. All commands are entered here. This window will have a prompt given by a red > sign.

If you ask 'R' for help, the results will appear in the 'R Information Window'. When you construct graphs and plots, these will appear in a separate 'R Graphics Window'. Control of the different windows is available by *clicking* on them or selecting them from the 'Windows' pull down menu at the top of the screen.

Ia. Quitting R What? You want to leave so soon? Well its always good to know how to get out of a program once you are in. Most things in 'R' are *functions* which mean they take *arguements* inside parentheses. To exit 'R', type in the function `quit()`:

```
> quit()
```

That's it, now you are back to the safety of Windows.

IIa Entering data in R

Renter 'R' if you have quit. As we know, Statistics is all about *data* so we need to know how to enter some data we've collected into 'R' so we can manipulate, rearrange and graphically decribe it.

One easy way to insert data in 'R' is using the `c()` function. For example, if we had collected 'favorite number' data 3, 2, 1, 4, 3 from five people, we could store this data in a variable named `Numbers` by typing:

```
> Numbers = c(3, 2, 1, 4, 3)
```

Now 'R' has a copy of our data in `Numbers`. If we want to look at what is contained in `Numbers`, just type:

```
> Numbers  
[1] 3 2 1 4 3
```

and the results appear.

'R' is good at arithmetic. Suppose you wanted to add 22 to all the favorite numbers. This is easy:

```
> Numbers+22
[1] 25 24 23 26 25
```

Or, perhaps something more complicated like adding 22 and then taking the square root of the results:

```
> sqrt(Numbers+22)
[1] 5.000000 4.898979 4.795832 5.099020 5.000000
```

I**b**. Plotting data in **R**

Looking at raw numbers on a screen is never very exciting. To remedy this 'R' has a huge number of different plotting routines for graphically representing data.

The simplest graphing routine is a function called `plot()`. For example, to show our favorite number data, we could try:

```
> plot(Numbers)
```

This command produces no output in the Console Window, but it does give us a nice plot of our numbers in the Graphics Window. Try it.

How about histograms? 'R' is good at that too. The function for producing histograms is `hist()`. Try it on `Numbers`.

```
> hist(Numbers)
```

What do you get? Is it 'right'? Think about it!
Maybe this makes more sense:

```
> hist(Numbers,breaks=c(0.5:1:4.5))
```

Better? Why? What's this `breaks=` stuff? Think about it!

III. Existing Data Sets in **R**

R has a huge number of strange (and maybe interesting) data sets already installed and ready to look at statistically. To see a complete listing of **R**'s data collection type the command:

```
> data()
```

This should open a new window with a long list of data sets. To get more information about any data set, type:

```
> help( name of the dataset here )
```

For example, lets look at the data containing `iairquality` information in New York before you were born (1973). To load this data into **R**,

```
> data(airquality)
> attach(airquality)
```

The last command **attaches** the names of different data in `airquality` to the actual data.

To see what information is in the dataset, and what the names of the different pieces of data are:

```
> names(airquality)
```

We can now look at the data. For example, make a histogram of the OZONE measurements during 1973 - Ozone is bad for you, especially if you have lung problems to begin with. Try:

```
> hist(Ozone)
```

DO the following:

- Make a histogram of Ozone, Temperature and Wind Speed. Print each graph!
- How many days during 1973 was the air quality really bad (ozone count greater than 100)? Write the answer on the Ozone Histogram.
- What do you think the AVERAGE temperature was during the measurement period? How do you come up with this number? Write this on the Temperature Histogram. What about the AVERAGE ozone count?
- Explain, in words, the differences between the shapes of the three histograms. Which two are alike? Come up with a reason why this might be so.