Vectors, data.frames, indexing and sub-setting

**20160530:23.00**

Preamble

* It is suggested that you use <ALT> + <TAB> to alternate between this Workbook module and RStudio
* In this module where a series of commands open drop down menus or dialogue boxes we will provide the sequence to follow separated by the pipe “|” character (e.g. **Files |New File | R Script** to open up a new R script)

This module is designed to provide you with:

* First intended learning outcome

This module assumes that you have a basic familiarity with the four panel layout of RStudio. If that is not the case then we recommend that you first study the module **“Introducing RStudio”.** You will also need to be able to reload the environment file **WBenv1.RData** that was created and then saved in the module **“Introducing RStudio”.** To recreate the environment *without* going back to the preceding module click [here](#Environment_set) and follow the instructions

## Initial setup

If you have not already loaded the environment file WBenv1.RData then do so now.

* Click on **Files** tab and navigate to the **RStud-master** folder
* Open this and set it as your Working Directory (hint – click on [**More**](#SetWorkingDir))
* Click on the file **WBenv1.RData** and then on **“Yes”** to confirm

## Starting and writing a Script

First we will create a script which we will then use to carry out a number of exercises. At the top left of the screen click on **Files |New File | R Script** (note that the short cut **<CTRL> + <SHIFT> + N** can also be used).

You should now have a blank script with the name Untitled1

In the margin of the script you will see the line number 1. Try entering “# My workscript” and then <ENTER>. Note:

1. The name of the script has turned red. The filename will always turn red if you make any change to the script
2. The cursor has moved on to line 2
3. Your first line started with a “#” and so will be treated by R as a comment

In the second line of the script try assigning the value 70 to a variable called “Weight”. You will need to type or paste “ Weight <- 70”

Run this line and note that a new object “Weight” has appeared in the Environment panel. This is a simple scalar variable of type number

To save the script either click on the blue floppy disk icon, or press **<CTRL> + S**, or do **File | Save**. The first time that you try to save a new script these will all result in a dialogue box expecting you to enter a filename of your choice. Enter “MyWorkScript” or any preferred name and click **Save**. Note that the Script filename has now changed to the name that you entered and is no longer red.

Vectors and the use of the colon operator and functions c and seq

In R a vector is a simple kind of data object which consists of a sequence of data elements of the same type. The individual data elements are known as the ‘components’ of a vector.

The variable “Weight” so far has only one component. It is in fact a **Vector**. We can assign a sequence of components to “Weight” in a number of ways. Try adding each of the following examples to the script and then run them one by one

**Use of colon operator**

This is a short hand way of creating a sequence of integer components in increments of 1 starting from the value to the left of the colon and ending at the value to the right of the colon. For example:

Weight <- 70:80

**Use of the c() “combine” function**

By using c() we can assign a series of any individual values in any order to a vector. For example:

Weight <- c(70, 69.3, 71, 68.7, 79, 77.4)

**Use of Sequence generator seq()**

By using seq() we can assign a series of components with start and end values that increment or decrement by a fixed amount. For example:

Weight <- seq(72,73,0.2)

**c() in more detail**

Note that the series of values handled by c() can include a combination of individual values, sequences declared with “:” and sequences declared with seq(). For example:

Weight <- c(65, 67, 70:74, 69, seq(75,77,0.4))

Also, as seen in the “Introducing RStudio” module, c() can be used to combine nominal components into a vector. For example:

colnames<-c("Id","Age","GenderMF","Hypert","SBP","DBP")

It will be useful to save the WBdata file as a CSV file into your working directory to save having to download it again. But first we need to fix two problems. Firstly the column names are not meaningful and secondly there is no agebands column. On the way we will get acquainted with the very useful RStudio command completion feature and also with the three functions listed.

**Command completion**

Into your script start typing in lower case – names. Notice that after a few letters a dialogue box pops up with suggestions. Click on names in the dialogue box and note that RStudio has completed the function name, added two brackets, and is prompting for a variable name to be typed inside the brackets. Start typing WBdata and notice that you are typing inside the brackets and that now RStudio offers you any variable name starting with those letters. Click on WBdata. These are two examples of the very useful completion feature of RStudio. You can also trigger this behaviour by starting to type a command and then pressing <TAB>. RStudio will help you by completing:

* Functions and their arguments
* Names of objects
* Filenames

Not only does this save time but also it helps to prevent mis-spelling of names (including problems with upper / lowercase letters) of functions and objects

**Assigning the output of names() to a variable and vice verse**

In your script you should have: **names (WBdata).** Run this line. Note that in the console R has responded with a list of the column names of WBdata which are currently not very informative. We can assign the output of the names function to a variable by using the arrow <- convention. Let’s send the names to a variable called OrigNames. Go back to your script and edit the line with the names command so that it starts with OrigNames <-

The line should look like: **OrigNames <- names(WBdata)**

Now run it

Note that a new object **OrigNames** has appeared in the Environment panel

Go to the next line in your script and start to type Or and then press <TAB>.

Note that RStudio offers you the variable name that you want – and gets the upper / lower case letters right. Click on this to complete the script line and run it.

Note that in the console you can see the values contained by OrigNames.

Suppose that we want to change the column names to: "Id","Age","GenderMF","Hypert","SBP","DBP". We can build this up in stages. On a fresh script line type “Id” and run the line. R simply responds with “Id”

On fresh script line try typing “Id”,“Age” and run the line. This returns with an error. We need a function **c()** to combine this series of elements. Delete the faulty line and instead type c(. Notice that RStudio adds the right bracket and also waits for you to enter the details inside the brackets. Type inside the brackets: c("Id","Age","GenderMF","Hypert","SBP","DBP"). Notice that each time you enter an open quote mark RStudio provides the close quote mark. Take care to advance the cursor beyond the second quote before adding a comma and then starting to enter the next name. Run the line. R echoes out all of the names in the console. As before, we can assign the output of **c()** to a variable. Add colnames <- to the start of the line so that it looks like:

colnames <- c("Id","Age","GenderMF","Hypert","SBP","DBP")

Run this to create a new variable **colnames** in the Environment panel

Earlier we assigned the output of the **names()** function to a variable. But it is also possible to assign a variable to the **names()** function. Thus we can assign the values in **colnames** to **names(WBdata)**.

Add a new line to your script:

**names(WBdata) <- colnames** and run it

Now if we do: **names(WBdata)** we can see that the column names of **WBdata** have changed. There is as yet no Agebands5 column

Script sitting on Home RStudio server. ??? introduce Str() and then add agebands?

## Vectors

## Instructions for setting up the environment

First download the data file WBdata.csv

1. **Go to the Environment panel (top right with Environment tab active)**
2. **First, clean out the environment by clicking on the broom icon and then clicking “Yes” in the “Confirm Remove Objects” dialogue box**
3. **Then, just below and to the Right of the Files tab find the “Import Dataset” tab**
4. **Click on the Import Dataset tab and select “From Web URL”**
5. **This will open the “Import from Web URL” box**
6. **Either copy and paste or type in the following URL**https://raw.githubusercontent.com/Lynd148/WBdata/master/WBdata.csv
7. **Click on ‘OK’**
8. **The Import dataset wizard box will open. Accept all of the default settings**
9. **Click on Import to create the data.frame WBdata**
10. **Close the WBdata view in the Source panel**

Next download the scripts and Rdata resources from the Course web repository. Select the appropriate version below

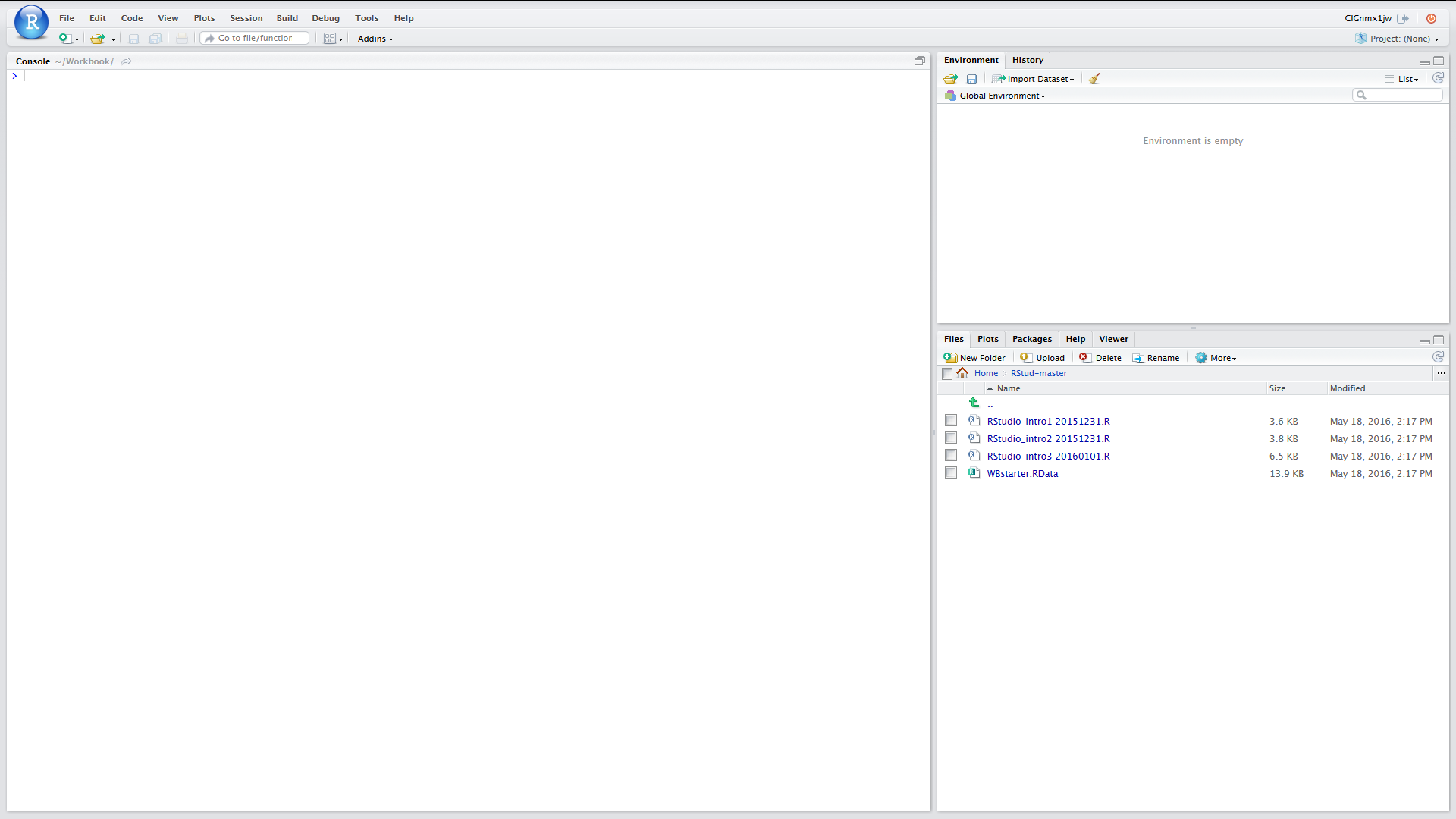
[Version for Desktop / notebook based RStudio](#FILEDLD_DESKTOP)

[Version for Server based RStudio](#FILEDLD_SERVER)

**File download for Desktop / notebook based RStudio**

1. **Open you browser and either copy and paste the following URL to the address bar or type it in:** [**https://github.com/Lynd148/RStud.git**](https://github.com/Lynd148/RStud.git)
2. **This should open up a github repository with the title “Temporary repository for Workbook download materials” which will look something like this and contain a number of files. Note the button marked “Clone or Download” below the blue line**



1. **Click on the “Clone or Download” button. A dialogue box will open with two options: “Open in Desktop” and “Download ZIP**
2. **Click on the Download ZIP button. Depending on your browser and settings this will either open a dialogue box asking whether you want to open or save a file called RStud-master.zip or else will result in this file being immediately downloaded to your download folder. In the former case select ‘open’ and if necessary change the “Open with” option to Windows Explorer then click “OK”. The folder RStud-master should appear after being downloaded  
     
   In the latter case (where the file has automatically downloaded) navigate to your download folder. Right click on the file and select ‘open with’ and the option Windows Explorer. The folder RStud-master should now appear**
3. **Copy the folder RStud-master to your C-drive or other easily found destination**
4. **Now return to RStudio. Make sure that the Files tab of the bottom right Files Plots Packages Help Viewer panel is active and click on the small box at the right edge of the screen containing three dots …**
5. **This should open a “Browse for folder” box. Browse to the folder RStud-master on your C-drive or wherever else you may have placed it and click OK.**
6. **The folder will now open under the Files tab which should look similar to the screen shot below**
7. **We now just need to set up the Working Directory. To do this click on the “More” tab (next to blue cogwheel icon just below the Files Plots Packages Help Viewer bar) and then Click on “Set as Working Directory”**
8. **All of the scripts and other files that we need in order to proceed should now be downloaded and also you have set the working directory**
9. **Click on the file WBstarter.RData to load it into the environment**
10. **Now save the environment to WBenv1 click on the Environment tab and then the blue floppy disk item, enter WBenv1 into the dialogue box and click Save**

Click [here](#Setup) to return to start of this module

**File download for server based RStudio**

1. **Open you browser and either copy and paste the following URL to the address bar or type it in:** [**https://github.com/Lynd148/RStud.git**](https://github.com/Lynd148/RStud.git)
2. **This should open up a github repository with the title “Temporary repository for Workbook download materials” which will look something like this and contain a number of files. Note the button marked “Clone or Download” below the blue line**



1. **Click on the “Clone or Download” button. A dialogue box will open with two options: “Open in Desktop” and “Download ZIP”**
2. **Click on the Download ZIP option. Depending on your browser and settings this will either open a dialogue box asking whether you want to open or save a file called RStud-master.zip or else will result in this file being immediately downloaded to your downloads folder. In the former case select ‘save’ and click “OK” because in this scenario we want a ZIP file to be placed in your download folder**
3. **Having ascertained the location of your downloads folder return to RStudio. Make sure that the Files tab of the bottom right Files Plots Packages Help Viewer panel is active.**
4. **Next you need to set your home directory to be the Working Directory**
   1. **Click on Home**
   2. **Click on More (next to blue cogwheel icon) and then Click on “Set as Working Directory”**
5. **Now look for the Upload button just below the Files Plots Packages Help Viewer bar. Click on this**
6. **This should open a “Upload Files” box. Within that box browse to the destination folder for your downloads and find the file RStud-master.zip and select / open this and then click on OK**
7. **This should result in the folder RStud-master being added to your home directory**
8. **Open this**
9. **Next you need to reset your home directory to this RStud-master directory. As before, click on the “More” tab (next to blue cogwheel icon) and then Click on “Set as Working Directory”**
10. **All of the scripts and other files that we need in order to proceed should now be downloaded and also you have set the working directory.**
11. **Click on the file WBstarter.RData to load it into the environment**
12. **Now save the environment to WBenv1 click on the Environment tab and then the blue floppy disk item, enter WBenv1 into the dialogue box and click Save**

Click [here](#Setup) to return to start of this module

## Lookup

Set Working directory

**File | More | Set as Working Directory**

Click **<ALT> + Left Arrow** <- to return