Data Entry in R

Jim Bentley

Data may be entered into R in a number of ways. Three commonly used methods will be discussed.

1 Manual Entry

Perhaps the easiest way to enter small datasets is to enter each variable indvidually and then combine them into a `data frame`. Using the data from BPS5 problem 4.9, this might look like:

```
> sex = c(rep("Female",12),rep("Male",7))
> mass = c(36.1, 54.6, 48.5, 42.0, 50.6, 42.0, 40.3, 33.1, 42.4,
+          34.5, 51.1, 41.2, 51.9, 46.9, 62, 62.9, 47.4, 48.7, 51.9)
> rate = c(995, 1425, 1396, 1418, 1502, 1256, 1189, 913, 1124, 1052,
+          1347, 1204, 1867, 1439, 1792, 1666, 1362, 1614, 1460)
> gender = c(rep(1,12),rep(2,7))
> bps5.4.9 = data.frame(sex, mass, rate, gender)
```

We can now check to see if the data frame has been created by entering

```
> ls()

[1] "bps5.4.9" "gender"   "mass"     "rate"     "sex"
```

Note that the listing also shows the individual variables that were used to create the data frame. These can be deleted by using rm().

```
> rm("sex", "mass", "rate", "gender")
> ls()

[1] "bps5.4.9"
```

The attributes of the data frame and some summary statistics can be computed using the `attributes` and `summary` functions.

```
> attributes(bps5.4.9)

$names
[1] "sex"    "mass"   "rate"   "gender"

$row.names
 [1]  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19

$class
[1] "data.frame"

> summary(bps5.4.9)
```

```
    sex          mass            rate          gender
 Female:12   Min.   :33.10   Min.   : 913   Min.   :1.000
 Male  : 7   1st Qu.:41.60   1st Qu.:1196   1st Qu.:1.000
             Median :47.40   Median :1396   Median :1.000
             Mean   :46.74   Mean   :1370   Mean   :1.368
             3rd Qu.:51.50   3rd Qu.:1481   3rd Qu.:2.000
             Max.   :62.90   Max.   :1867   Max.   :2.000
```

Notice that while `sex` was treated as a categorical variable, `gender` was treated as if it was cardinal. R is smart in that it recognizes the difference between cardinal and categorical (which it calls "factor") variables. To make `gender` a factor variable we can enter

```
> bps5.4.9$gender = factor(bps5.4.9$gender,levels=c(1,2),labels=c("F","M"))
```

Using `summary` we can see that `gender` is treated as a factor, or categorical, variable.

```
> summary(bps5.4.9)
```

```
    sex          mass            rate      gender
 Female:12   Min.   :33.10   Min.   : 913   F:12
 Male  : 7   1st Qu.:41.60   1st Qu.:1196   M: 7
             Median :47.40   Median :1396
             Mean   :46.74   Mean   :1370
             3rd Qu.:51.50   3rd Qu.:1481
             Max.   :62.90   Max.   :1867
```

2 Using Rcmdr

The package `Rcmdr` allows us to import data created in a number of packages. While the Windows version of R will import Excel (.XLS) files, the Mac version of R does not. However, both versions will import SPSS transport files.

To use Rcmdr we first need to load the package. This can be accomplished using menus or by using the `library` function. Assuming that `Rcmdr` is installed we enter

```
> library(Rcmdr)
```

If everything is working correctly, the `Rcmdr` GUI interface should start. After selecting **Data** – **Import Data** – **from Excel, Access, or dBase data set**, R will ask us for a name for our data set. Enter something descriptive but easy to type (*e.g.* `HtWt`). Remeber that R is case sensitive.

Next, you will have to select the Excel file that contains your data. R will then ask which sheet in the Excel file you wish to import. Once you have selected a sheet, R will complete the import and the data set/frame will be created.

`Rcmdr` will indicate that the data frame has been created and selected by showing `Data set: HtWt` above the script window. You can now view the data by clicking on `View data set`.

Noting that the `Group` variable (which is really a sex variable) is coded as a numeric (1 or 2), we should probably recode it as a factor variable. `Rcmrd` makes this easy. Click on **Data** – **Manage variables in active data set** – `Convert numeric variables to factors`. Select the variable we wish to change — in this case `Group`. We will supply level names and use the same variable for the factor recoding. Click on **OK**. We are going to overwrite `Group` so click on **Yes**. In this case a 1 is a Male and a 2 is a Female. Once the level names have been entered, click on **OK**.

Clicking on **View data set** we see that the `Group` variable is now coded as Female and Male. R now recognizes `Group` as a factor/categorical variable.

Data that is stored in SPSS portable or save formats can be imported in a similar manner. The files that come with BPS5e are actually in the portable format so you can use the menus to create a new data frame.

3 Reading Comma Seperated Value (CSV) Files

R has a utility for reading comma seperated value (CSV) ascii files. These files can reside on the host machine or on a server. If the files are in standard CSV format, either of

```
> HtWt = read.csv("c:/stat/ncssdata/htwt.csv")
> htwt = read.csv(
+ "http://newton.uor.edu/facultyfolder/jim_bentley/downloads/math111/htwt.csv")
```

will create a data frame that contains the NCSS Sample data set's height and weight data. Note the use of forward slashes instead of backslashes.

The group variable will be imported as a numeric. To help R function efficiently, it will need to be converted to a factor variable using one of the methods from above.

4 Saving and Loading Data Frames

Regardless of how they were created, data frames may be saved in R as part of the R workspace. The workspace contains all of the variables, data frames, and functions that you have defined. A workspace is a snapshot of your work to the point of the save.

To save a workspace click on **File** – **Save Workspace**. Select the folder to which you wish to save the file and a file name and then click on **Save**. Your workspace is now safely tucked away on your drive. This file can later be **Load**ed or you can open it by double clicking on the file.

History files store the commands that you used during your R session. These can be saved ans loaded in a manner similar to that of workspaces. These files are are text files and can be edited using Wordpad or something similar.