

R TUTORIAL, #6: IMPORTING DATA and DATA FRAMES

The (>) symbol indicates something that you will type in.

A bullet (•) indicates what the R program should output (and other comments).

IMPORTING DATA FILES FROM THE WEB

- We will import the text file “RMidterm.txt” from the web directory <http://www.kkuniyuk.com/RFiles>. (Excel files can be saved as text files.)

> Type: midterms =

```
read.table(url("http://www.kkuniyuk.com/RFiles/RMidterm.txt"), header=T)
```

- ‘midterms’ will be a data frame containing the imported data.
- ‘header=T’ indicates that it is “True” that the file has headers.

> Type: midterms

- You will see the data frame ‘midterms’.

DATA FRAMES: ENTRIES, ROWS, AND COLUMNS

- What are the dimensions of the data frame ‘midterms’?

> Type: dim(midterms)

- We see there are four rows and four columns.

- What did the second student get on Midterm 1?

> Type: midterms[2, 3]

- We obtain the entry in Row 2, Column 3.

- Instead of the column number, we could also use its column header.

> Type: midterms[2, "Mid1"]

- Let’s get the information for the fourth student.

> Type: midterms[4,]

- Let’s get the scores for Mid1.

> Type: midterms[,3]

- We could have also typed: midterms[,"Mid1"]

DATA FRAMES: ACCESSING DATA USING ‘ATTACH’

- To analyze data in the data frame, we must first “attach” the data frame.
- > Type: attach(midterms)

DATA FRAMES: OPERATIONS

- Find the mean of ‘Mid1’ and the mean of ‘Mid2’.
- > Type: mean(Mid1)
> Type: mean(Mid2)
- In fact, ‘mean’ can be applied to a data frame.
- > Type: mean(midterms)
- It doesn’t make sense to take the mean of ‘Mid1’ or ‘Mid2’, so we get “NA”s for them.
- Sort the values in ‘Mid1’ and the values in ‘Mid2’.
- > Type: sort(Mid1)
> Type: sort(Mid2)
- Sort the rows of the data frame by ‘Mid1’ scores.
- > Type: midterms[order(Mid1),]
- We can also sort in decreasing order.
- > Type: midterms[order(Mid1, decreasing=T),]
- Do side-by-side boxplots of ‘Mid1’ and ‘Mid2’.
- > Type: boxplot(Mid1, Mid2, names=c("Mid1", "Mid2"))
- To avoid confusion with other work, we will “detach” the data frame.
- > Type: detach(midterms)

IMPORTING DATA FILES FROM WORD, EXCEL, ETC.

- Type data into Word, Excel, etc. Use TABs or spaces as separators in Word.
 - Use Row 1 for headers such as “Name.”
 - Save the file as a “tab-delimited” text file; this will place in TABs as spaces.
 - Sometimes, you have to use Word or a text editor to make sure that “RETURN” or “ENTER” has been used after the last line.
 - If the data is separated by commas (,), then use the argument sep = “,” below.
- > Type: x = read.table(file.choose(), header=T)
> Type: x