

R Basics

A Brief Introduction to R for APMA6430

Xiaofeng Wang
Department of Systems and Information Engineering
University of Virginia
xw4u@virginia.edu

October 16, 2011

1 What is R?

R is a free software environment and programming language for statistical computing and graphics.
For more information about R, please refer to:

- R Project website: <http://www.r-project.org/>
- R (programming language)-Wikipedia:[http://en.wikipedia.org/wiki/R_\(programming_language\)](http://en.wikipedia.org/wiki/R_(programming_language))

2 Installation of R

R can be downloaded from the *R project website*(<http://www.r-project.org/>).

From the homepage, click "download R" and then choose any mirror to the download page (you may open it directly at:<http://cran.OpensourceResources.org/>).

Under the section "Download and Install R", choose the precompiled binary distributions based on your operation systems:

- If you use Windows: click *Windows* → *base* → *Download R 2.13.2 for Windows (or the latest version)*;
- If you use Mac: click *MacOS X* → *R-2.13.2.pkg* (or the latest version)

With the downloaded install package, just run it and follow the instructions.

If you encounter any problem during installation, you should find solutions under section "FAQs" at <http://cran.OpensourceResources.org/>.

3 Basic R Commands with Examples

3.1 How to Run Commands or Functions?

After opening R program, we will see the "R Console" window. We can type commands or functions after the command line prompt '>' to use R interactively. Each time we press Enter key, R executes the command.

For example, we type $2*3$ and press Enter key:

```
> 2*3
[1] 6
```

Alternatively, we can type all the commands and functions into a script and then run them. To do this, open "File" → "New script". We will see a new window "Untitled-R Editor". Type commands and functions line by line into the script, and run them by clicking "Edit" → "Run all"; or select certain lines and click "Edit" → "Run line or selection" (of course it's faster to use shortcut key "Ctrl+R").

3.2 Basic Commands and Functions

Here is a script to demonstrate basic operations in R. You can download this script on Collab, open it in R and run it line by line. The '#' is used to make comments. Anything following '#' is ignored by R.

The R script:

```
#####
##A Brief Introduction to R for APMA6430##
## by Xiaofeng Wang (xw4u@virginia.edu) ##
#####

##Basic computations:
2*3/(5+6)^2-1 #You can use R as a calculator
exp(2)         #Exponential function
log(10)        #Natural log function
pi             #=3.1415926....
sqrt(9)        #Square root

##Read data:
rm(list = ls()) #Clear workspace
setwd('c://APMA6430//RCode//')
                  #Set working directory
help(setwd)      #Help for 'setwd' command; it's a good way to learn functions
D<-read.table('Rdata_demo.csv',header=T,sep=',')
                  #Read a dataset; try 'help(read.table)' to learn more about it
D<-read.table('c://RCode//Rdata_demo.csv',header=T,sep=',')
                  #If your dataset is located at another directory
D
                  #Display values of 'D'
summary(D)       #A summary for 'D'

##Structures of datasets
D.R1<-D[1,]      #Get the first row of 'D'
```

```

D.R1
D.R1to10<-D[1:10,] #Get the first 10 rows
D.R1to10
D.C2<-D[,2] #Get the second column
D.C2
R.sample<-sample(1:length(D[,1]),10)
#Here are two functions: sample(), length()
C.sample<-c(3,5,6,7)#Assign a vector (3 5 6 7) to 'C.sample'
D.sample<-D[R.sample,C.sample]
#Get a subset of 'D'
D.sample<-D[sample(1:length(D[,1]),10),c(3,5,6,7)]
#You can write the three functions above together

D.sample
X<-rnorm(10) #Generate a random vector (from the normal distribution)
x<-2*rnorm(10) #Note R is case sensitive, so x is different from X
D.sample.new<-data.frame(D.sample,X,x)
#Add 'X' and 'x' to 'D.sample';
#explore two similar functions: 'cbind()' and 'rbind()'

D.sample.new

##Explore data
D.sample.new$Y1 #Access the variable 'Y1' in 'D.sample.new'
attach(D.sample.new) #Attach 'D.sample.new' to the R search path
Y1 #So, you can access 'Y1' easily
detach(D.sample.new) #If you won't work with 'D.sample.new', detach it
Y1 #You will see "Error: object 'Y1' not found"
attach(D.sample) #Let's work with 'D.sample'
summary(D.sample) #You can find basic statistics for each variable
plot(Y1) #Plot values of 'Y1'
plot(as.factor(Y1)) #If you know 'Y1' is a categorical variable
plot(D1,D2) #A scatterplot
boxplot(D3) #Boxplot
par(mfrow=c(1,2)) #Multiple plots
hist(D1,main="Histogram of D1",col="blue")
#Histogram plot
plot(density(D1),main="Density of D1")
#Density plot
par(mfrow=c(1,1)) #Set back to default

##To end, let's try a simple linear regression model:
lmodel<-lm(D1~D2+D3+x+X,data=D.sample.new)
summary(lmodel)
par(mfrow=c(2,2))
plot(lmodel)

```

```
par(mfrow=c(1,1))
```

4 Resources for Beginners

If you are new to R, you will find following resources helpful:

- *Introduction to R* by Professor Karl W. Broman: <http://www.biostat.wisc.edu/kbroman/Rintro/>
- *An Introduction to R* by W.N. Venables, D.M. Smith, and R Development Core Team: <http://cran.r-project.org/doc/manuals/R-intro.pdf>
- Chapter 1: A brief introduction to R in *Data Analysis and Graphics Using R-An Example-Based Approach, 3rd ed.* by J. Maindonald and J. Braun.