

Training Materials
for the
One-day Training Workshop on R



by

Professor Valerie M. LeMay

Professor of Forest Measurements
Dept. of Forest Resources Management
Faculty of Forestry
University of British Columbia
Vancouver, Canada

*The Workshop was part of the activities during the **Conference on Forest Measurements in Complex Tropical Forests** held at the Federal University of Technology, Akure, Nigeria between 9th and 11th June, 2009.*

Introductory Remarks by Professor S. O. Akindele

I was privilege to attend a training workshop on R at the Faculty of Forestry, University of British Columbia (UBC), Vancouver, Canada in February 2006. The workshop was organized by the Inventory/Biometrics Research Group of the Faculty, and the instructor was Dr. Andrew Robinson of the Department of Statistics, University of Melbourne, Australia. My participation was facilitated by Professor Peter Marshall and Professor Valerie LeMay, who hosted me for my sabbatical leave at UBC. As a forest biometrician from a developing country where standard statistical software are very expensive to get, an open source software such as R presents a very good alternative to use. It is free and sophisticated enough to handle many statistical analyses we encounter on regular basis. It is also dynamic and constantly being improved upon by a network of users and developers across the globe. More packages are being incorporated into it to enhance its capability, and with some knowledge of programming, it can be customized to produce relevant results.

I discussed the possibility of having Professor LeMay visit us in Nigeria and conduct the training workshop on R for us. She readily obliged and started making preparations. She put together the training materials and even purchased some additional texts on the software. Much as she desired to come and conduct the training, other commitments made it impossible for her to come at this time. She then sent the training materials and additional resources to me to stand in for her in conducting the training.

The training workshop is aimed at introducing participants to the R statistical software. The software is available on the CD given to all participants. It can also be downloaded free from the internet (<http://www.r-project.org/>). The instruction on how to load the software and use it for common statistical analyses will be treated during this workshop.

Prof. S. O. Akindele

Professor of Forest Measurements

Deputy Coordinator, IUFRO 4.01.03 Working Group on Instruments and Methods for Forest Mensuration

Chairman, Organising Committee for the Conference on Forest Measurements in Complex Tropical Forests.

Table of Contents

[Module 1](#)

Introduction	1
1.1 Objectives	1
1.2 Background.....	1
1.3 Installing R on your home or laptop computer	1
1.4 Running R.....	2
1.5 Useful things to know	2
1.6 Graphs	3
1.7 Help	3
1.8 Expanding the R package	3
1.9 Learning R	3

[Module 2](#)

Basic Statistics and Regression Analysis Using R.....	5
2.1 Files Needed	5
2.2 Exercise	5
2.3 More Exercises	8

[Module 3](#)

Graphs Using R	10
3.1 Background.....	10
3.2 Files	10
3.3 Running the R Script.....	10

[Module 4](#)

Multiple Linear Regression Using R.....	11
4.1 Background.....	11
4.2 Objective.....	11
4.3 Files	11

[Module 5](#)

Extra Exercise on Multiple Linear Regression using R	12
5.1 Background.....	12
5.2 Objective.....	12
5.3 Exercise	12
5.4 Questions	12

[Module 6](#)

Using R and Stepwise Methods to Select Predictor Variables in a Regression Model.....	14
6.1 Background.....	14
6.2 Files	14
6.3 Exercise	14

[Module 7](#)

Experiments Using a Completely Randomized Design, One-Factor Using R	15
--	----

[Module 8](#)

Experiment Using a Completely Randomized Design, Two Factors.....	17
---	----

Contributed Documentation (Reference Materials) on R.....	19
Documents with more than 100 pages:.....	19
Documents with fewer than 100 pages:.....	19
Short Documents and Reference Cards:	20

Module 1

Introduction

1.1 Objectives

In this exercise you will be introduced to R, including how to get a copy of R and documentation that can be found on web sites. The exercises then use R to get basic statistics and a linear regression using tree data.

1.2 Background

R is a free software package that has been designed to analyze and graph data. A collection of people worldwide have developed libraries of functions that you can use to analyze data. Because this is freeware, there may be “bugs” in the software. However, many of the library functions have now been tested by many users, and compared to other commercially available software packages such as SAS and SPSS.

1.3 Installing R on your home or laptop computer

To run R, first you need to load the software. Generally, to load the software, you can go to R website directly. There are instructions for downloading software on the site. You will need to find a Cran near to your location, for faster service. A list of these can be found on the R website. This loads the standard package, with a number of libraries. Here are the steps you would follow:

Go to the R website: <http://www.r-project.org/>

Select CRAN, which is under Download, on the left hand side of the screen

From the list (centre and right side of the screen), select the closest location

Then, select Download and Install R, and select Windows

Under Subdirectories, select Base

R-2.9.0-win32.exe [or whatever the latest version of R is]

You will be asked if you want to run or to save this file. If you are ready to install the package, click on Run.

You will then be prompted for where you wish to save the file (pick a simple path on your home or laptop computer e.g. C:/R). Use the default settings.

When you are done, there should be the R icon on your desktop and R should be ready for your use.

1.4 Running R

After you install R for Windows, you will have an R icon on your desktop. To run R, click on the icon . You get a work session window. You could type in your commands here and they would run as you enter them. Instead, you can enter your commands into a separate file called *script*. The script is just R commands, organized and put into a text file. Then, you can open your script file while you are running R. You will get a separate window with the script. You can then run this all at once (only if you are very confident), or in segments (preferred).

The outputs from running your commands will also appear in the session window. At end time, you can use **Edit** and **Clear Console** to clean out the session window. However, the data you brought in, and any variables you created will still be there.

If part of your script involves getting a graph, another window opens up if you get a graph using R. If you run another graph, the original graph disappears and you get the new graph.

You can save the work session, the script, and the graph window anytime you wish, by using **File** and **Save** for the session window when that window is active, or to save the script when the script window is active, or to save the graph when the graph window is active.

1.5 Useful things to know

R is case sensitive. This means that the variable **Trees** is not the same as the variable **trees** for example. Also, R does not like spaces or special characters. Instead, use a '.' For example, **trees.pine** identifies a variable.

R uses two slashes instead of one to indicate a subfolder. For example, if your data in Windows are in: E:\measurements\trees.txt then in R you would use E:\\measurements\\trees.txt since the single slash has a different meaning.

Any R commands that start with # are just *comments* that you can add to explain what the script does.

1.6 Graphs

R is very good at graphs. The main way to make a graph is to use the function `plot()`, where there are a number of arguments in the brackets (i.e., the x variable, the y variable, labels, type of graph, etc). However, only one graph appears at once, in a separate window. When you graph in R, you should save the graph (e.g., as a .jpg file or as a metafile), before moving to the next graph. R can also do multiple graphs on the same graph window.

1.7 Help

The R website has a number of manuals that you might find useful, including:

1. An introduction to R (<http://cran.r-project.org/doc/manuals/R-intro.pdf>)
2. Using R for Data Analysis and Graphics: Introduction, Code and Commentary (<http://cran.r-project.org/doc/contrib/usingR.pdf>)
3. R for Beginners (http://cran.r-project.org/doc/contrib/Paradis-rdebuts_en.pdf)
4. icebreakR (<http://www.ms.unimelb.edu.au/~andrewpr/r-users/icebreakR.pdf>)
5. R: A Language and Environment for Statistical Computing – A Reference Index prepared by the R Development Core Team ([http://cran.r-project.org/doc/manuals/fullrefman\[1\].pdf](http://cran.r-project.org/doc/manuals/fullrefman[1].pdf))

There are also a number of very useful books published by Springer, and Chapman and Hall publishers. In addition, more reference materials have been listed at the end of the manual (Contributed Documentation).

At any time, you can also use `help()` where the function is given in the brackets. This help is a bit hard to follow, and is really meant to tell you the specific options for a function. However, there are also a few examples with the help that you might find useful as you are using R.

1.8 Expanding the R package

When you run R, only some of the functions are brought into the work session automatically to save memory. To add others, you can use `require()` where the package is given in brackets. Also, there are many other parts of R that are extra to the main package. To bring these in, you will need to access the website and get the software package. This then can be downloaded to the R directory in a sub-folder under library. For example, if you installed R in: C:\Program Files\R\R-2.9.0\, then you can add more software into C:\Program Files\R\R-2.9.0\library\ You can then use `library()` to bring in these other packages for your analysis.

1.9 Learning R

Many people have put documentation and examples using R code or script on the web. Examples are very helpful for reducing the time you spend in getting

R to do what you would like. However, the best way to learn R is really to use it. The course materials provided by Dr. Andrew Robinson (icebreakeR) are excellent to help you practice and learn R and become more comfortable with using it for your analyses. The exercises provided here are very brief and just give you a taste of using R for forestry problems.

Module 2

Basic Statistics and Regression Analysis Using R

2.1 Files Needed

You will need the files: *ht_dbh.xls* and *ht_dbh.txt* (the tree data) and *ht_dbh.R* (R commands, called R script).

2.2 Exercise

A forest land owner measures the outside bark diameters at 1.30 m above ground (dbh) and total tree height from ground to tree tip for a sample of 20 trees on a small piece of land. The trees are equally spaces over the land area. The measures are:

Tree Number	Dbh (cm)	Height (m)
1	10.1	14.2
2	11.2	15.1
3	19.7	25.3
4	20.5	21.2
5	17.8	21.5
6	17.0	18.0
7	11.0	12.1
8	4.1	5.2
9	6.0	6.3
10	8.0	9.1
11	2.3	10.1
12	20.1	19.2
13	18.0	16.0
14	22.1	26.3
15	16.3	17.3
16	20.5	19.8
17	17.0	20.1
18	18.0	22.3
19	17.0	19.5
20	19.7	18.6

Before we can do any analysis, we need to bring these data into the R environment. We can do this by:

1. Typing the data right into the R script (Parts I to IV of this Exercise)
2. Entering the data into EXCEL (eg., *ht_dbh.xls*) and then saving this as a tab delimited text file (e.g., *ht_dbh.txt*) or comma delimited file (e.g., *ht_dbh.csv*) (Part V of this Exercise).

Once the data are in the R environment, we can get basic statistics, fit models, get graphs, etc.

For this exercise, R script was provided as *ht_dbh.R*. The script is organized in parts using comments (the # denotes comments). To learn what the script is doing, you should run this in pieces and determine what the R code is doing before you move on to the next step.

To run this in segments, you can copy and past a part of the R script into the work session, and then running that part. Another way that we will use is to highlight a part of the script and using **Ctrl+R** to run that part of the script.

The work session window will include the R commands, and the outputs. At any time, you can copy and paste any part of the session window into a WORD file, or store the entire work session window.

1. First, start R, and bring the script in by using **File** and then **Open Script**. Browse until you find the *ht_dbh.R* file and click on it to bring it into R. You will see that there are comments added to the script to explain what each line of code does. Remember, comments begin with # .
2. Part I: Using the R script provided as *ht_dbh.R*, highlight Part I of the code that brings the data into R. This is done by 1) highlighting that part of the code, and 2) using **Ctrl+R** to run the code. You should see results in the “session” window.
What did each line do? Try to understand how each line of code was used to bring the data into the R environment.
3. Part II. Run the next part of the R code provided to calculate simple statistics for the heights. For each item in this list, 1) find the R code, highlight the code, and use Ctrl+R to run the code. Write down the answers you obtain.
 - a. The sample mean
 - b. The variance
 - c. The standard error of the mean
 - d. The mode
 - e. The median
 - f. The coefficient of variation as a percent
 - g. A 95% confidence interval for the true mean (all of the trees).
 - h. Given the sample data, and no assumptions about the probability distribution, what is the estimated probability that a tree will be more than 10.0 cm in dbh?

- i. Given the sample data, and the assumption that it follows a normal distribution, what is the estimated probability that a tree will be more than 10.0 cm in dbh?
3. Before running more of the provided R code, modify this to obtain the same statistics for dbh. To do this, use **File** and **New Script** to open a new window for your script that you will create. Then, copy and paste the code for the height basic statistics into the file, save it, and modify it for dbh instead of height. *Again, write down the answers as you get them OR copy and paste them from the console to a WORD file.*
4. Parts III and IV. Now, we would like a model to predict height from dbh, since height is harder to measure. The fitted model can then be used where only dbh was measured. Using the R code provided, locate and run the part of the code fits the model. *Run this in parts, as before and write down your answers as you go.*
 - a. Graph the height versus dbh for these sample data. NOTE: This will appear in a Graph window. Save the graph as picture for future reports.
 - b. Since this is not a linear relationship, transformations are needed to linearize the relationship before using linear regression. NOTE: Part III does height versus dbh (no transformations) whereas Part IV uses transformations.
 - c. Fit a simple linear regression of height versus your transformed dbh NOTE: There is no need to change units to be the same for both variables. Write down the answers that you get as you use the script to get:
 - i. The estimated intercept and slope. Use the estimated slope and intercept and overlay your equation over the selected graph in part c.
 - ii. Calculate the standard errors and 95% confidence intervals for the intercept and for the slope.
 - iii. The coefficient of determination (r^2) and the standard error of the estimate (SE_E), also called the root mean squared error (Root MSE). What do these mean?
 - iv. Graph the fitted line over the original points.
 - v. Based on the graph, are the assumptions that the line fits the data and that variances of y's around the x's are equal met for your selected equation? (i.e., you need the residual plot).
 - vi. Are errors normally distributed?
 - vii. How would you check the assumption that the observations are independent for these data?

5. In forestry, we sometimes measure height on photographs, or using LiDAR. In that case, dbh is the expensive variable. Using the same data, assume that the heights were measured using LiDAR and we then want an equation to predict dbh from LiDAR height. Use **File** and **Open Script** to open another window for some new script. Copy and paste the R code for the height vs dbh equation to New Script and modify the script to instead obtain an equation for dbh vs height. *Using your outputs, answer the same questions as in 4c but for this model.*
6. Before going to Part V, clean up your all of your work and remove all objects. This is done by using **Edit** and **Clear console** and also **MISC** and then **Remove all objects**. This allows you to start fresh, getting rid of any variables and data you brought in, and any outputs you have created. This can prevent errors, but you must bring in new data after clearing out all the objects.
7. Part V: In this part, the data come from an EXCEL file instead of being entered into the R code itself. These data were entered into EXCEL and then saved as a tab delimited text file to be used in R (*ht_dbh.txt*). You must give the full path for your data, and NOTE that the folders are given after \\ instead of the usual \ used by Microsoft Windows. *Run this other script, and again write down your answers as with Question 4 c.*

2.3 More Exercises

1. Close R to get rid of all script and datasets.
2. Open R again, and open the *ht_dbh.R* script.
3. Using **File** and **New Script** to open a new window for your script. Using the code provided in *ht_dbh.R* script as your model:

- a. Bring the *ht_dbh.txt* data into the R environment.
- b. Create two new variables and plot these by:

```
loght<-log(height)
logdbh<-log(dbh)
plot(loght,logdbh)
```

How strong is this relationship? Is it a linear relationship? Could you fit a linear regression to this relationship based on the graph? NOTE: You cannot compare the R square for this model to that where the y variable was height instead of loght.

- c. Using the R script as an example, get a linear regression of loght versus logdbh. *Does the residual plot indicate that this is a good regression (i.e., are the points balanced around zero across the range of predicted heights?)*

- d. Copy your regression results from the session window into WORD, and copy and paste any graphs to go with your regression results. Add a few points on why this model is a good model or not based on these outputs.
- e. Save your R script for future use.

Module 3

Graphs Using R

3.1 Background

R has some very useful graphics functions. These can be very helpful for conveying information to audiences in presentations and papers. We have already used histograms, and scatterplots for regression results.

3.2 Files

We will use the tree data found in *trees.txt* for this exercise. There are 250 *Populus* trees and 250 *Abies* trees in this dataset. We will run some simple plots to visualize this fairly large dataset. The script can be found in *graphs.R*.

3.3 Running the R Script

1. Start R
2. Use **File** and **Open script** to bring in the *graphs.R* script.
3. For graphs, a number of lines of the R script must be run together, to set up the graph, and then add data to the graph. These lines of R script are separated by blank lines. Run the R Script in parts by: 1) highlighting a part of the script and then 2) Using **Ctrl+R** to run the script.
4. As you run the script in parts, write down what each section does.
5. Also, click on the graph window and then **File** and **Save As** to save one or more of your graphs.

For discussion:

Which plot(s) did you find useful in visually describing these data?

Module 4

Multiple Linear Regression Using R

4.1 Background

Multiple linear regression uses more than one x-variable to predict the variable of interest, the y-variable. The x's can be several different variables that we have measured, or can be the originally measures variables, plus transformations of these variables. For example, we may use dbh and dbh squared to predict height, rather than just dbh or just dbh squared. In the case of the transformed variables, we are trying to meet the assumption that the linear model is correct.

4.2 Objective

Practice bringing in data that originally in an EXCEL file, and practice using R to get an equation with more than one predictor variable (x-variable) in a multiple linear regression.

4.3 Files

For this, you will use data gathered for a few African trees (provided by Prof. Akindele). The data can be found in *african_trees.xls*. There is also R script provided as *mlr.R*

1. *Getting the data into R:*
 - a. In EXCEL, bring up the data file.
 - b. Save this as a tab delimited text file called *african_trees.txt*.
 - c. Start R.
 - d. Bring in the R script found in *mlr.R*.
 - e. Modify the R script by correcting the path for the datafile.
2. Use the script to run a multiple linear regression to predict height (Ht) from dbh (Dbh) and transformations of dbh. As with Exercise 1, run this in segments and relate what happens to the R code that you have run (i.e., highlight a part of the code and use Ctrl+R to run that part). There are blank lines in the code to indicate a “part” of the code that should be run at the same time.

NOTE: In R, the variable *dbh* is different from the variable *Dbh* – capital letters matter.

Module 5

Extra Exercise on Multiple Linear Regression using R

5.1 Background

Data have been gathered on a number of plots in a forest. In each plot, the tree dbh and height, and the species were measured. An existing volume function was used to find the volume per tree. Then, each plot was summarized to obtain summary variables. The plot data are in *stand.txt*.

5.2 Objective

The objective is to find a good equation to estimate volume per ha, from variables that are easier to measure. Then, in future plots of a similar kind of forest, these other variables can be measured, summarized for the plot, and then used to estimate volume per ha by inputting them into the equation.

5.3 Exercise

Fit a model that predicts volume per ha from other variables. Consider X variables that are easier to measure first (e.g, average dbh). Use the *mlr.R* code as a guide and modify this for this new data and regression problem. Use any transformations you might need to meet the assumptions of multiple linear regression.

5.4 Questions

1. Which equations did you try? (Try at **most** two equations) Which ones met the assumptions of regression (i.e., normal distribution of residuals, even pattern of residuals around zero indicating that the model fits the data and that the variances are equal across the range of predicted values)
2. Of the equations where the ASSUMPTIONS were met, assess which equation is better in terms of:
 - a. The R square value (CAREFUL – can only compare those that had the SAME Y variable!!)
 - b. The Root MSE
 - c. The fitted line plot
 - d. Whether the equation is significant
 - e. Whether each variable is significant
 - f. The cost of measuring the X-variables (to use the equation).
3. Based on this assessment, which equation would you recommend for use?

NOTE: There is R script prepared, in *exercise_extra_MLR.R* if you do need help with setting up the R script. In the R script, you will find more code for graphs that you might find useful also!

Module 6

Using R and Stepwise Methods to Select Predictor Variables in a Regression Model

6.1 Background

Stepwise methods can be helpful for selecting some x'variables for predicting the y-variable. Methods can be forward (in only), backward (out only) or both (in and out). The resulting subset of x variables can be different, depending upon the method used. Once subsets of x variables are obtained using these selection methods, a full regression can be run, and the assumptions checked, etc.

6.2 Files

We will use the plot data found in *stand.txt* for this exercise. The data for each plot were compiled to obtain volume per ha, basal area per ha, stems per ha, top height, quadratic mean dbh, average age, site index. The script can be found in *stepwise.R*.

6.3 Exercise

Run the script in sections, as before, to be able to understand what the R code does. Then, using one of the subsets of selected variables, run a full regression analysis and check assumptions, etc.

For discussion:

How useful were these selection methods for choosing x variables to predict volume per ha? Did you obtain a good result with your full regression using the subset of x variables?

Module 7

Experiments Using a Completely Randomized Design, One-Factor Using R

A researcher wants to examine the impacts of thinning (tree removal) on growth of red pine trees in Ontario, Canada. There are three treatments: No removal (control), thinning (light – few trees are removed), heavy (many trees are removed). A plantation of 30 ha is selected, where trees are evenly spaced, with similar dbh's (diameter outside bark, measured at 1.3 m above ground) and are currently 15 years old. Fifteen areas are established in the plantation, each 1 ha in size (experimental unit). Each 1 ha area is then randomly assigned a treatment, resulting in five experimental units having each treatment. After 5 years, a number of 0.02 ha plots are established, systematically, over the each 1 ha area. The dbh's of all live trees are measured in each plot, and entered into an excel file. The average diameters are calculated for each 1 ha experimental unit resulting in the following values (data are in *crd.txt*):

Treatment	Exp_unit	AveDbh
None	10	7.50
None	4	6.70
None	1	7.20
None	14	8.20
None	3	8.60
light	13	9.60
light	8	8.40
light	5	8.90
light	2	9.60
light	12	11.10
heavy	11	11.40
heavy	9	9.90
heavy	6	10.60
heavy	7	12.70
heavy	15	13.50

Using the script found in *crd.R*:

1. Obtain a boxplot. Based on this boxplot, are there differences in AveDbh among the three treatments?
2. The null hypothesis is that there are no differences in mean of AveDbh among these three treatments?
 - a. Check the assumptions by getting a histogram and normality plot of the residual values.

- b. If assumptions are met, set up your hypothesis (H_0 and H_1), obtain the F test statistic, the F critical value (or p-value), and make your decision (reject H_0 ?), using the *lm* output. Use $\alpha=0.05$.
3. Use pairs of means t-tests to check for differences between pairs of treatments. Remember to correct this test using a Bonferroni correction (i.e., divide alpha by the number of pairs of means).

Discussion:

Are these tests reliable? Were assumptions of Analysis of Variance met, or are transformations needed?

If assumptions were met, what are the results of your tests? Are there differences in AveDbh? If so, which thinning methods differ?

Module 8

Experiment Using a Completely Randomized Design, Two Factors

In a second study, the impacts of thinning (tree removal) and fertilization on growth of red pine trees in Ontario are of interest. The three levels for the first factor, thinning, are: No removal (control), thinning (light – few trees are removed), heavy (many trees are removed). For the second factor, fertilization, there are two levels, from 1 (no fertilizer) to 2(fertilizer). In total, there are six treatments. Again, a plantation of 30 ha is selected, where trees are evenly spaced, with similar dbh's (diameter outside bark, measured at 1.3 m above ground) and are currently 15 years old. Twelve areas are established in the plantation, each 1 ha in size (experimental unit). Each 1 ha area is then randomly assigned a treatment, resulting in two experimental units having each treatment. After 5 years, a number of 0.02 ha plots are established, systematically, over the each 1 ha area. The dbh's of all live trees are measured in each plot, and entered into an excel file. The average diameters are calculated for each 1 ha experimental unit resulting in the following values (*crd_two_factors.txt*):

Exp_unit	Thinning	FertLevel	AveDbh
2	none	1	6.7
4	none	1	7.2
8	none	2	7.5
3	none	2	8.2
12	light	1	8.4
1	light	1	8.9
5	light	2	9.6
9	light	2	9.6
11	heavy	1	10.6
6	heavy	1	11.4
7	heavy	2	12.7
10	heavy	2	13.5

The researchers would like to know if there are differences in the mean of the AveDbh with different treatments.

1. Two analyses were run (*crd_two_factors.R*). The first used AveDbh as the *y- variable*, and the second analysis used the logarithm of AveDbh instead. Which of these should be interpreted (NOTE: meets assumptions of equal variance and normality of residuals)?
2. Based on the analysis that met the assumptions, is there an interaction between the two factors? Use alpha=0.05.

THEN:

3. If there is an interaction, which treatments differ? (Use pairs of means t-tests at the treatment level, and remember to use the Bonfereonni correction).

OR:

3. If there is no interaction, does thinning change the diameter? In what way (higher or lower average diameter across thinning levels? (Use the pairs of means t-test for thinning levels – remember to use the Bonferonni correction).
4. If there is no interaction, does fertilizer change the average diameter? In what way? (Use the pairs of means t-test for fertilization levels – remember to use the Bonferonni correction).

Discussion:

Was the transformation of *AveDbh* needed?

Was there an interaction? If yes, which treatments differed?

If there was no interaction, is there a difference in mean of *AveDbh* between thinning levels?

If there was no interaction, is there a difference in mean of *AveDbh* between fertilization levels?

Contributed Documentation (Reference Materials) on R

Documents with more than 100 pages:

- **“Using R for Data Analysis and Graphics - Introduction, Examples and Commentary”** by John Maindonald ([PDF](#), data sets and scripts are available at [JM's homepage](#)).
- **“Simple R”** by John Verzani ([PDF](#), data sets, various PDF, PS and a browsable HTML version are available at the [Simple R homepage](#)).
- **“Practical Regression and Anova using R”** by Julian Faraway ([PDF](#), data sets and scripts are available at the [book homepage](#)).
- The [Web Appendix](#) to the book “An R and S-PLUS Companion to Applied Regression” by John Fox contains information about using S (R and S-PLUS) to fit a variety of regression models.
- **“An Introduction to S and the Hmisc and Design Libraries”** by Carlos Alzola and Frank E. Harrell, especially of interest to SAS users, users of the Hmisc or Design packages, or R users interested in data manipulation, recoding, etc. ([PDF](#)).
- **“Statistical Computing and Graphics Course Notes”** by Frank E. Harrell, includes material on S, LaTeX, reproducible research, making good graphs, brief overview of computer languages, etc. ([PDF](#)).
- **“An Introduction to R: Software for Statistical Modelling & Computing”** by Petra Kuhnert and Bill Venables ([ZIP 3.8MB](#)): A 360 page PDF document of lecture notes in combination with the data sets and R scripts used in the manuscript.
- **“Introduction to the R Project for Statistical Computing for Use at the ITC”** by David Rossiter ([PDF](#)).
- **“Analysis of Epidemiological Data Using R and Epicalc”** by Virasakdi Chongsuvivatwong ([PDF](#)).
- **“Statistics Using R with Biological Examples”** by Kim Seefeld and Ernst Linder ([PDF](#)).
- **“Icebreaker”** by Andrew Robinson ([PDF](#), 2008-05-08).

Documents with fewer than 100 pages:

- **“R for Beginners”** by Emmanuel Paradis ([PDF](#)).
- **“Kickstarting R (version 1.6)”** compiled by Jim Lemon, a short introduction in English as HTML files: download as [gzipped TAR](#) or [ZIP](#); or [browse directly](#).

- **“Notes on the use of R for psychology experiments and questionnaires”** by Jonathan Baron and Yuelin Li ([PDF](#)). A browsable version is available at [JB's homepage](#).
- **“R for Windows Users (version 2.0)”** by Ko-Kang Wang ([PDF](#), [LaTeX source](#)). Updates, a Postscript version and a [browsable HTML version](#) are available at [KW's R Resources](#) page.
- **“Building Microsoft Windows Versions of R and R packages under Intel Linux”** by Jun Yan and A. J. Rossini ([PDF](#), associated [Makefile](#)).
- **“A Guide for the Unwilling S User”** by Patrick Burns ([PDF](#)).
- **“The R language — a short companion”** by Marc Vandemeulebroecke ([PDF](#)).
- **“Fitting Distributions with R”** by Vito Ricci ([PDF](#)).
- **“Econometrics in R”** by Grant Farnsworth ([PDF](#) | [LaTeX source](#)).
- **“The Friendly Beginners' R Course”** by Toby Marthews ([ZIP](#), 2009-06-05, 14 pages).
- **“An R companion to ‘Experimental Design’ ”** by Vikneswaran ([PDF](#)).
- **“The R Guide”** (version 2.3) by Jason Owen ([PDF](#), 2007-08-09).
- **“Multilevel Modeling in R”** by Paul Bliese ([PDF](#)), a brief introduction to R and the packages multilevel and nlme.
- **“Statistics with R and S-Plus”** by Hugo Quené ([PDF](#)).
- **“Using R for Scientific Computing”** by Karline Soetaert ([ZIP](#)): lecture notes and reference card for R beginners, including exercises.
- **“A (Not So) Short Introduction to S4”** by Christophe Genolini ([PDF](#), 2009-01-07, 68 pages).
- **“Creating R Packages: A Tutorial”** by Friedrich Leisch ([PDF](#), 2009-02-02, 19 pages).
- **“Creating R Packages, Using CRAN, R-Forge, And Local R Archive Networks And Subversion (SVN) Repositories ”** by Spencer Graves and Sundar Dorai-Raj ([PDF](#), [PPT](#), 2009-05-04, 45 slides).

Short Documents and Reference Cards:

- **“R reference card”** by Jonathan Baron ([PDF](#)).
- **“R and Octave”** by Robin Hankin ([Text](#)), a reference sheet translating between the most common [Octave](#) (or [Matlab](#)) and R commands.
- A **“time series reference card”** ([PDF](#)) and a **“regression reference card”** ([PDF](#)) by Vito Ricci.
- **“R reference card”** by Tom Short ([PDF](#), [LaTeX source](#)).