



Non-Programming Introduction to R

Legal Stuff



sheepsqueezers.com

This work may be reproduced and redistributed, in whole or in part, without alteration and without prior written permission, provided all copies contain the following statement:

Copyright ©2011 sheepsqueezers.com. This work is reproduced and distributed with the permission of the copyright holder.

This presentation as well as other presentations and documents found on the sheepsqueezers.com website may contain quoted material from outside sources such as books, articles and websites. It is our intention to diligently reference all outside sources. Occasionally, though, a reference may be missed. No copyright infringement whatsoever is intended, and all outside source materials are copyright of their respective author(s).



R Lecture Series

*Non-
Programming
Introduction*

*Programming
I*

*Programming
II*

*Graphics
I*

*Advanced
Topics*



Charting Our Course

- Goal of this Presentation
- What is R?
- Installing the Software
- Installing *R Commander* and *Rattle* Packages
- Installing *RExcel* for *R Commander*
- Installing *RGGobi*
- Installing Additional Packages
- Introducing *Rterm* and *RGui*
- Introducing *R Commander*
- Introducing *Rattle*
- Introducing *RExcel*
- Introducing *RGGobi*
- Appendix A: References
- Appendix B: R-Related Websites

Goal of this Presentation



The goal of this presentation is to show you how to use R through its many graphical user interfaces (GUIs). We decided to start with this instead of charging directly into programming to ease you into R rather than throwing you to the wolves.

Many of you will be casual R users and will not need or not want to learn the R programming language itself and it is our hope that this presentation will be to your benefit. This presentation will allow the casual R user to use R and will allow future R programmers to relate back to the casual R users.

There are a number of R GUIs, but we will not show you all of them. Some of these GUIs are still under development and will be added to this presentation once they are out of beta. As a casual R user, you will have your choice of GUI to use, but you may need to use several GUIs in order to produce the results you are looking for.

What is R?



R is a freely available programming language for statistical computing as well as the creation of publication-ready graphics. R is an implementation of the S programming language which itself was created by John Chambers, Rick Becker and Allan Wilks of Bell Laboratories. S is sold as a commercial product under the name *S-Plus*.

R is part of the GNU project which means that its source code is freely available under the GNU General Public License. The source code as well as pre-compiled versions of R are available at the R Project website (www.r-project.org) for a variety of platforms such as Windows and Linux.

R can be extended through the use of user-submitted packages (think SAS modules such as SAS/STAT and SAS/OR) and, at present, there are over two thousand freely available packages for download at the Comprehensive R Archive Network (CRAN) website (cran.r-project.org). A complete list of packages is located at cran.r-project.org/web/packages/index.html and topics cover basic and advanced statistics, genetic algorithms, ODBC/OLEDB database access, time series, financial computations, etc.

R comes with excellent documentation along with executable examples.

R comes with two interfaces: Rterm, a command line interface, and RGui, a GUI interface.

What is R?



```
ca. Rterm
C:\Program Files (x86)\R\R-2.10.1\bin>R

R version 2.10.1 (2009-12-14)
Copyright (C) 2009 The R Foundation for Statistical Computing
ISBN 3-900051-07-0

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> -
```

```
RGui
File Edit View Misc Packages Windows Help

R Console

R version 2.10.1 (2009-12-14)
Copyright (C) 2009 The R Foundation for Statistical Computing
ISBN 3-900051-07-0

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> |
```



Installing the Software

Installing the Software



sheepsqueezers.com

There are several ways you can install and run the R software depending on where you want to run it.

If you want to install R at home, visit the R Project website (www.r-project.org) and install R using one of the pre-compiled binaries for your operating system. To run R on Windows, you click Start...All Programs...R (folder)...R (executable) to start the software. Once started, you should see the RGui as shown on the previous slide. To exit the software, either click File...Exit and select No on the Save Workspace Image? dialog box, or type `q("no")` at the command line and hit the Enter key.

Note that on the following slides, the Fat Kids dataset is referred to several times. You can obtain this data (called `fatkids.csv`) on the Downloads tab of the sheepsqueezers.com website.

Installing the *R Commander* and *Rattle* Packages



Besides the RGui provided with R, we will explore four additional GUIs: *R Commander*, *Rattle*, *RExcel* and *RGGobi*.

R Commander is a menu-driven GUI based on the tcltk windowing software. You can find out more on R Commander at: <http://cran.r-project.org/web/packages/Rcmdr/index.html>.

Rattle is a menu-driven GUI based on the more modern RGtk2 windowing software and was specifically designed for data mining. You can find out more on Rattle at: <http://cran.r-project.org/web/packages/rattle/index.html>.

RExcel is a menu-driven add-in to Microsoft Excel which allows you to run R. We discuss this option below.

When installing each package, be aware that several additional packages will be installed along with these four *main* packages. During the installation process, you will see messages like the following cross your screen, so don't be alarmed:

```
trying URL 'http://lib.stat.cmu.edu/R/CRAN/bin/windows/contrib/2.10/timeDate_2120.90.zip'  
Content type 'application/zip' length 1084749 bytes (1.0 Mb)  
opened URL  
downloaded 1.0 Mb
```

```
trying URL 'http://lib.stat.cmu.edu/R/CRAN/bin/windows/contrib/2.10/timeSeries_2120.89.zip'  
Content type 'application/zip' length 1196866 bytes (1.1 Mb)  
opened URL  
downloaded 1.1 Mb
```

Installing the *R Commander* and *Rattle* Packages

To install R Commander or Rattle, you can either use the menus provided by RGui or you can type in the following at the R command line (don't type the greater-than symbol):

```
> install.packages("Rcmdr", dependencies=TRUE)
> install.packages("rattle", dependencies=TRUE)
```

Note that you may be asked to pick a *mirror* before the installation process begins. A *mirror* is a website hosting the R software. Choose the site nearest to you geographically (or psychologically...tee-hee) then click OK.

As mentioned on the previous slide, several other packages will be installed at the same time, so you will see a lot of messages flying across your screen. Once complete, you can then use the R Commander and Rattle packages.

Referring to Rattle, you may be asked to install the Gtk+ software before you use Rattle. Go ahead and install the software. Once Gtk+ is installed, you will have to exit out of R and then restart it.



Installing RExcel for R Commander



One additional GUI interface we will look at is RExcel. This package allows R to be used from within Excel as an add-in. With RExcel, you can transfer data between R and Excel; run R code from Excel ranges; write Excel macros calling R to perform calculations without exposing R to the user; R Commander's menus are available from with Excel.

There are two ways to install RExcel. The method outlined below assumes that R is already installed. If R is not installed, then you can download a *bundle* at <http://rcom.univie.ac.at/download.html> and run the RAndFriends.exe installer.

To install RExcel, at the R command prompt, enter in the following code:

```
> install.packages(c("RExcelInstaller",  
                    "RthroughExcelWorkbooksInstaller",  
                    "RcmdrPlugin.HH"), dependencies=TRUE)
```

Select the closest CRAN mirror if asked. Once installed, enter the following at the command prompt to start additional installations:

```
> library(RExcelInstaller)  
> installRExcel()
```

Installing RExcel for R Commander

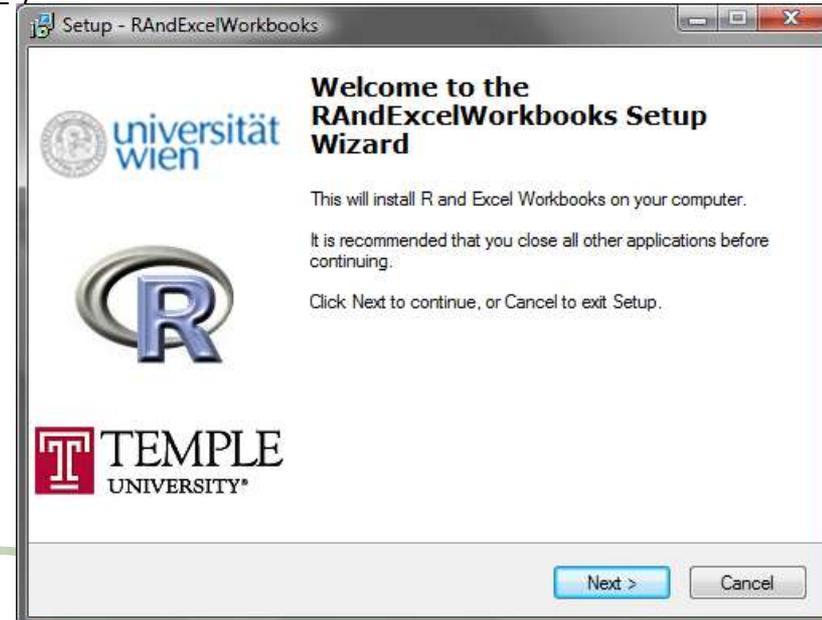


sheepsqueezers.com

At this point you will see the RExcel installer wizard. Follow the instructions to install RExcel and answer Yes and OK to dialog boxes that popup.



```
> library(RthroughExcelWorkbooksInstaller)
> installRthroughExcel()
```



Installing *RExcel* for *R Commander*



sheepsqueezers.com

Next, you will have to install statconnDCOM by entering the following at the R command line:

```
> library(rcom)
> installstatconnDCOM()
```

Following the instructions when the wizard appears and ensure that Local Server Installation is chosen later on in the installation process:



Finally, issue the following command at the R command prompt:

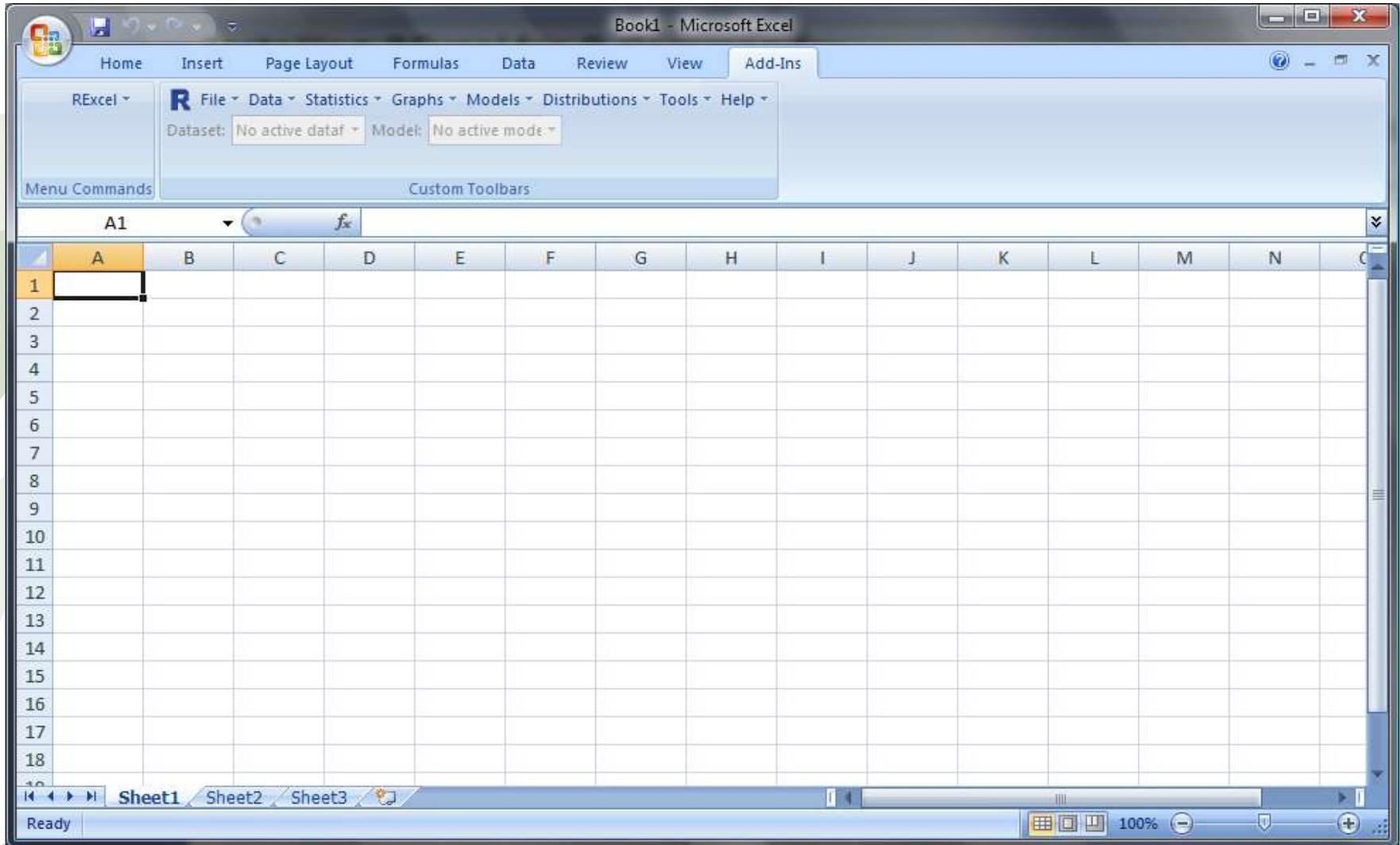
```
> library(RcmdrPlugin.HH)
```

Installing RExcel for R Commander



sheepsqueezers.com

You will notice that the installation software has placed an icon on your desktop entitled *RExcel#### with RCommander*. Double-click this icon to start RExcel. You will notice that not only does Excel start, but R Commander and R itself. Take note of the Add-Ins tab in Excel. Click this and you will see R-related actions.



Installing *RGGobi*



sheepsqueezers.com

RGGobi is an R interface into GGobi, an interactive and dynamic graphics package. In order to install the RGGobi package, you need to first install the GGobi software.

The creators of R/GGobi have made the install simple by providing just one command to run from within R:

```
> source("http://www.ggobi.org/downloads/install.r")
```

You will be asked to install Gtk+, but if you already installed Rattle then you don't need to install Gtk+.

You will also be asked to install GGobi. Click OK to begin the installation wizard for this software. Follow the prompts and then click Finish.

You will have to restart R.

To start RGGobi, enter the following at the R command line:

```
> library(rggobi); ggobi();
```

Installing Additional Packages



sheepsqueezers.com

If you have found an additional package you'd like installed, at the RGui command line type in the `install.packages()` command just like you did to install R Commander and Rattle. For example, on the CRAN website, there is a complete list of all of the packages available to you. Here's a very brief list of some of them:

SampleSizeMeans	Sample size calculations for normal means
SampleSizeProportions	Calculating sample size requirements when estimating the difference between two binomial proportions
SciViews	SciViews GUI API - Main package
ScottKnott	The ScottKnott Clustering Algorithm
SemiPar	Semiparametric Regression
SenSrivastava	Datasets from Sen & Srivastava
StatMatch	Statistical Matching

If you want to download the StatMatch package, say, type in the following at the command line:

```
> install.packages("StatMatch", dependencies=TRUE)
```

To find out more about the StatMatch package, click on the StatMatch link (not shown above, but shown on the CRAN website) and you will be transported magically to the webpage shown on the next slide. Take note of the Reference Manual linte item. This contains additional useful information about the StatMatch package as well as all of its available functions.

Installing Additional Packages



sheepsqueezers.com

The screenshot shows a Windows Internet Explorer browser window. The address bar contains the URL <http://cran.r-project.org/web/packages/StatMatch/inc>. The page title is "CRAN - Package StatMatch". The main content area displays the following information:

StatMatch: Statistical Matching

This package provides some R functions to perform statistical matching between two data sources sharing a number of common variables. These functions can also be used to impute missing values in data sets through hot-deck methods.

Version: 0.8
Depends: R ($\geq 2.7.0$), [proxy](#), [lpSolve](#)
Suggests: [optmatch](#)
Published: 2009-09-13
Author: Marcello D'Orazio
Maintainer: Marcello D'Orazio <madorazi at istat.it>
License: [GPL \(\$\geq 2\$ \)](#)
CRAN checks: [StatMatch results](#)

Downloads:

Package source: [StatMatch 0.8.tar.gz](#)
MacOS X binary: [StatMatch 0.8.tgz](#)
Windows binary: [StatMatch 0.8.zip](#)
Reference manual: [StatMatch.pdf](#)
Old sources: [StatMatch archive](#)

The status bar at the bottom of the browser window shows "Done", "Internet | Protected Mode: On", and a zoom level of "100%".



Running the Software

Introducing *Rterm* and *RGui*



sheepsqueezers.com

Rterm is a command line interface to R whereas RGui is graphical user interface (GUI) with menus to R.

```
ca. Rterm
C:\Program Files (x86)\R\R-2.10.1\bin>R

R version 2.10.1 (2009-12-14)
Copyright (C) 2009 The R Foundation for Statistical Computing
ISBN 3-900051-07-0

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> _
```

```
RGui
File Edit View Misc Packages Windows Help

R Console

R version 2.10.1 (2009-12-14)
Copyright (C) 2009 The R Foundation for Statistical Computing
ISBN 3-900051-07-0

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

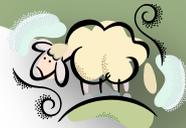
Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> |
```

Introducing *Rterm* and *RGui*



sheepsqueezers.com

Rterm is available on both windowing operating systems (like Windows and Linux with X) as well as non-windowing systems (like DOS and Linux command line only). Rterm only allows you to enter commands *a la* the R programming language. Note that RGui as well as R Commander, Rattle and RExcel are not available on non-windowing machines.

By double-clicking the R icon, you will start the RGui (see previous slide for image):



← my, what a fabulous icon!

Introducing *Rterm* and *RGui*



sheepsqueezers.com

You start the command line version of R by typing the letter R at the command line and hitting the enter key. Below is a what R looks like in a non-windowing environment for me:

```
$ R

R version 2.10.1 (2009-12-14)
Copyright (C) 2009 The R Foundation for Statistical Computing
ISBN 3-900051-07-0

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

>
```

Type your commands at the command line and be amazed!!

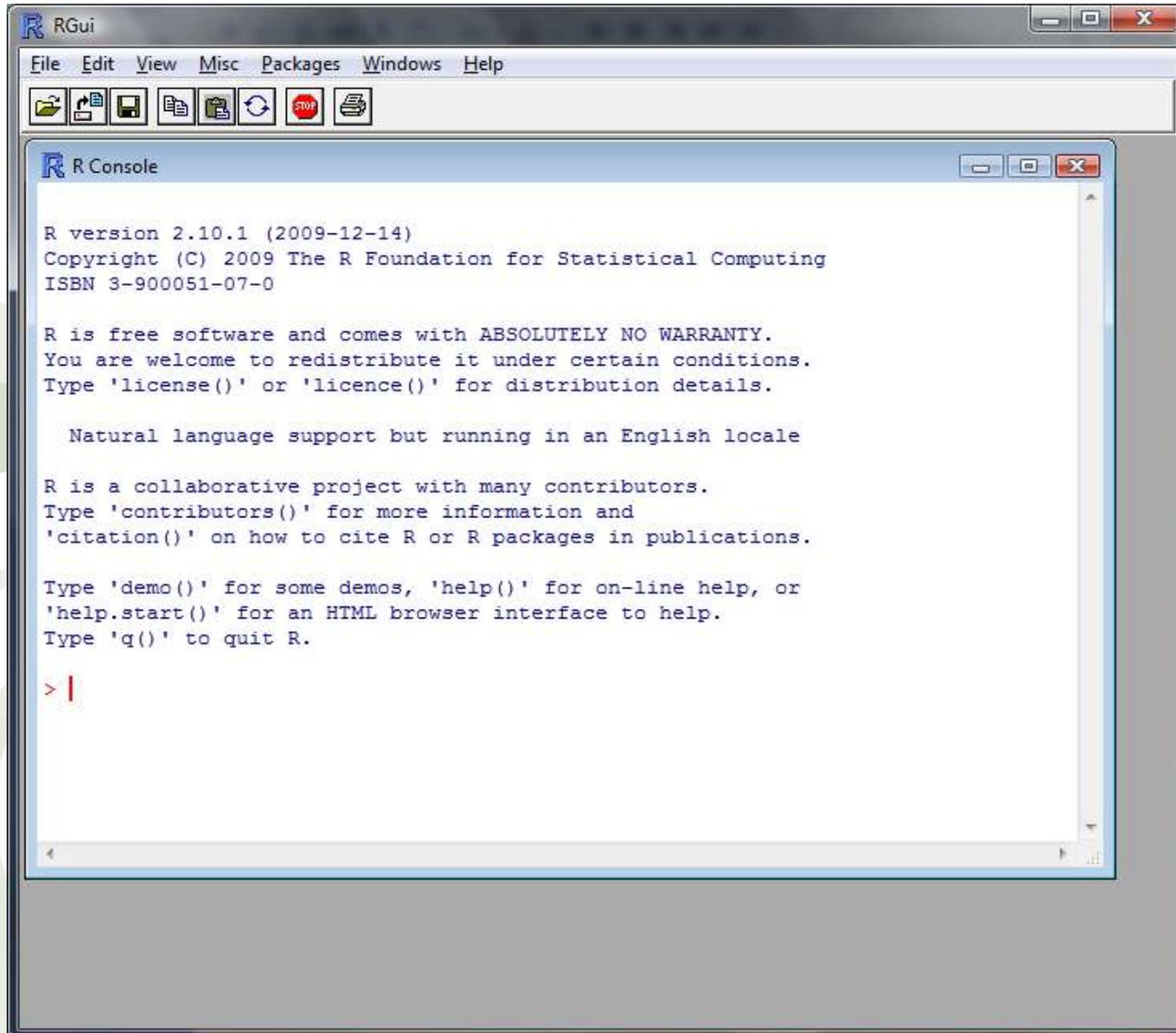
To exit out of the command line version of R, type in `q("no")` at the command line and hit the enter key.

Introducing *Rterm* and *RGui*



sheepsqueezers.com

In a windowing environment, you will see the RGui instead:



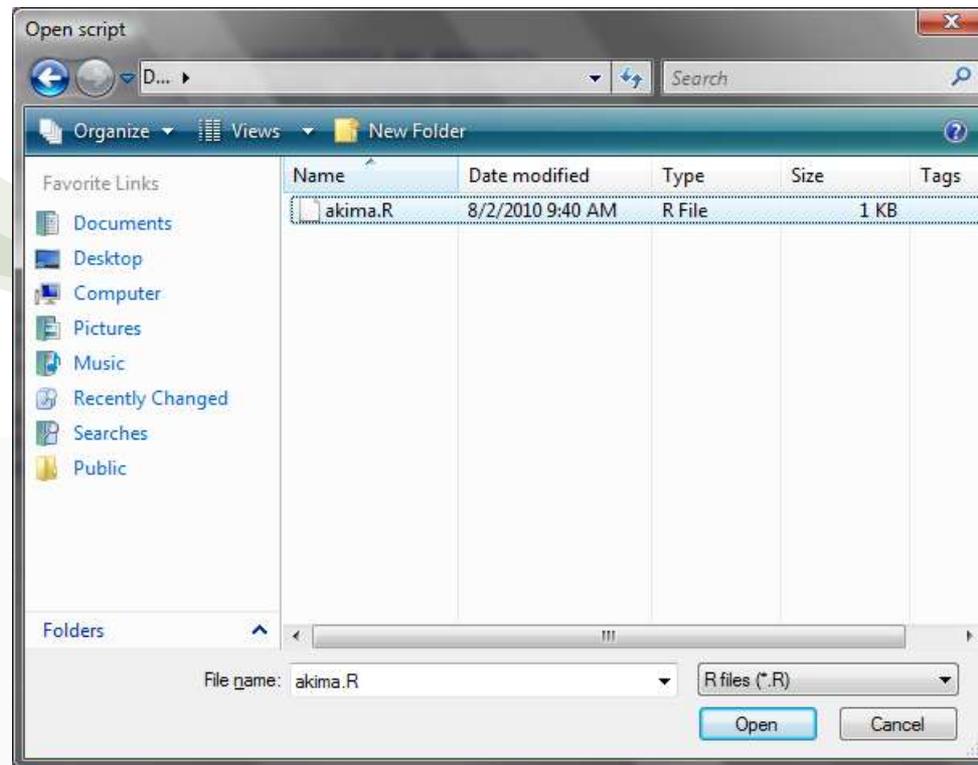
Introducing *Rterm* and *RGui*



sheepsqueezers.com

Unfortunately, the RGui is just a windowing version of Rterm, so you still get a command line, but it's a lot fancier! Woo-hoo!

In both Rterm and RGui, you can copy and paste R code or you can open up a file with the `.R` extension (like `ComputeStuff.R`) by using the File...Open... dialog box:



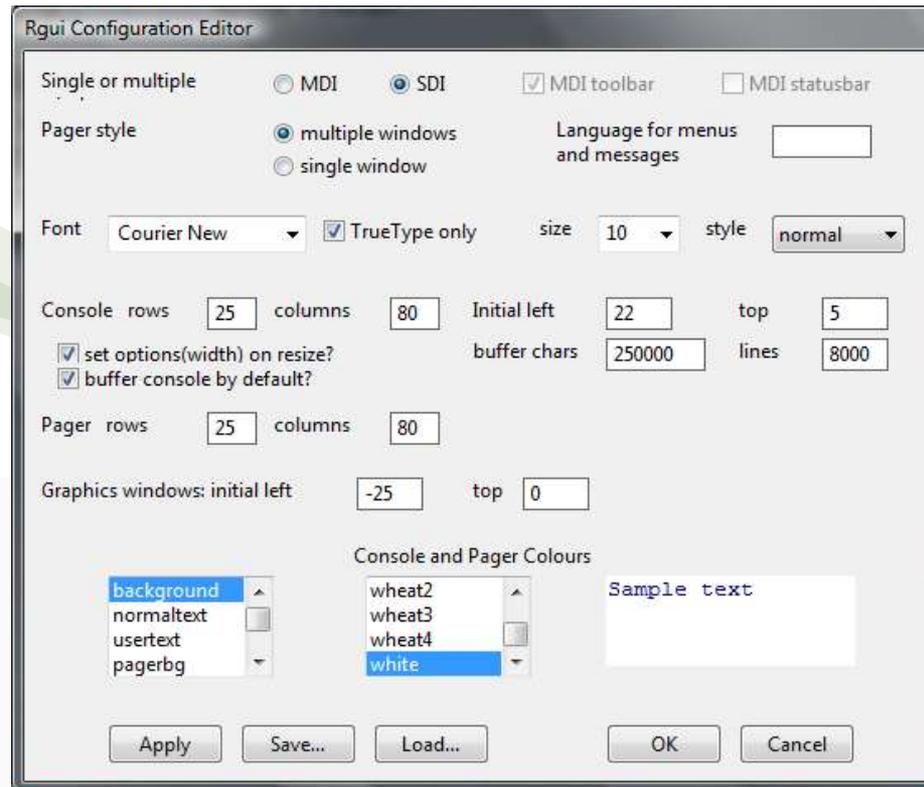
But, again, the goal of this presentation is to avoid learning the R programming language.

Introducing *Rterm* and *RGui*



sheepsqueezers.com

Initially, RGui is set up as a Multiple Document Interface (MDI). R Commander, which we describe in the next section, works best with RGui running as a Single Document Interface. To switch RGui to a single document interface, click on the menu Edit...GUI Preferences and select the SDI radio button on the Rgui Configuration Editor dialog box (shown below):



Click the Save... button, then click Save again. Click OK. You will have to restart R.

Introducing R Commander



sheepsqueezers.com

R Commander is an R package that provides basic statistics in a GUI interface employing menus, buttons, etc. which allows you to avoid issuing programming language statements in RGui or Rterm. According to its creator, John Fox of McMaster University:

The design objectives of the R Commander were as follows: to support, through an easy-to-use, extensible, cross-platform GUI, the statistical functionality required for a basic-statistics course (though its current functionality has grown to include support for linear and generalized-linear models, and other more advanced features); to make it relatively difficult to do unreasonable things; and to render visible the relationship between choices made in the GUI and the R commands that they generate.

While we will not describe all of the functionality of R Commander, we will show you some basics. You can find more on R Commander at the following websites:

<http://socserv.mcmaster.ca/jfox/> - John Fox's home page

<http://socserv.mcmaster.ca/jfox/Misc/Rcmdr/> - R Commander home page

<http://www.jstatsoft.org/v14/i09/paper> - Paper on R Commander by John Fox

<http://socserv.mcmaster.ca/jfox/Misc/Rcmdr/Getting-Started-with-the-Rcmdr.pdf> -
Introduction to R Commander

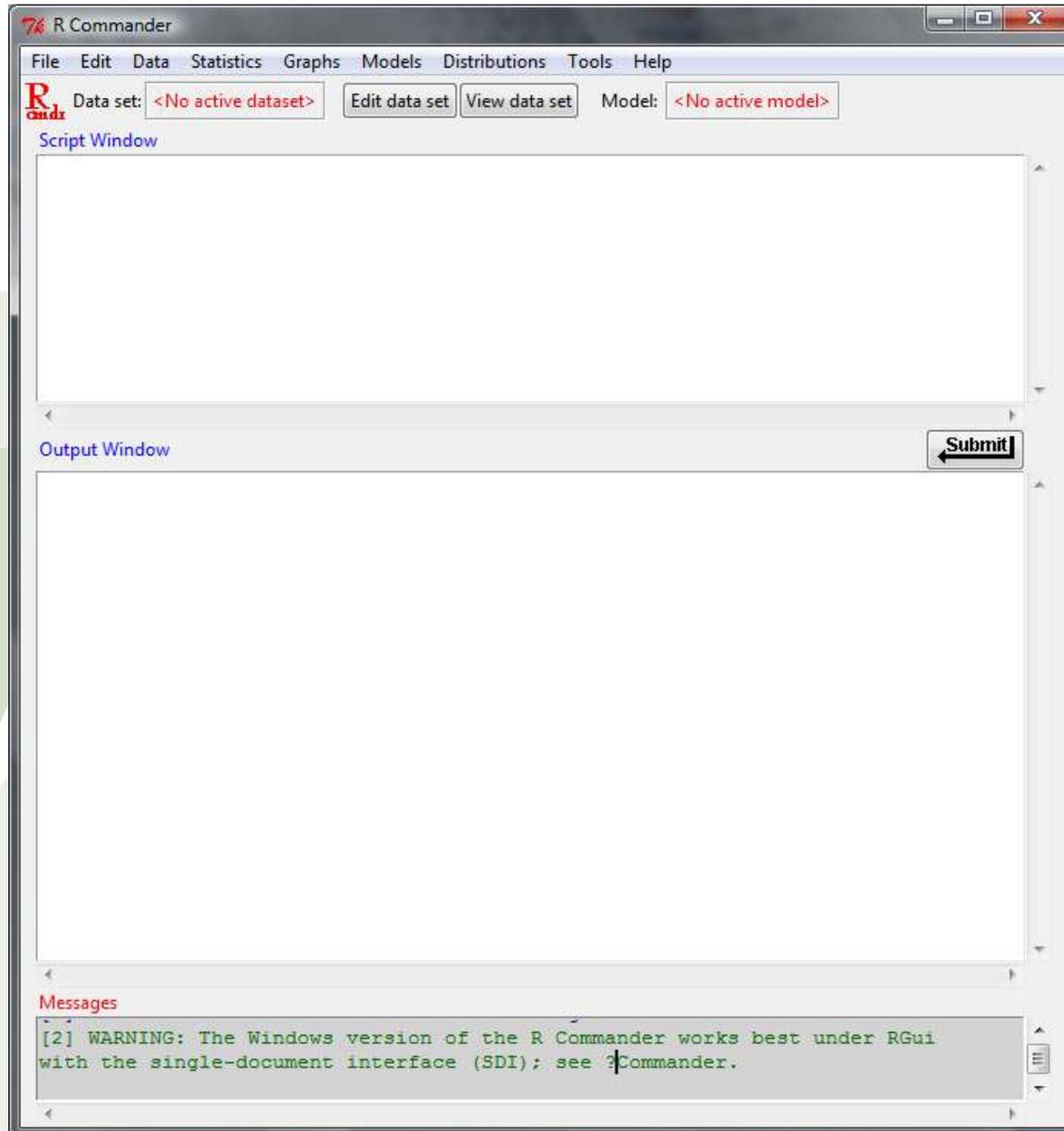
In order to start R Commander, start R first by clicking on the R icon or navigating to START → All Programs → R → R 2.10.1. At the command prompt, type in the following to bring up the R Commander GUI:

```
> library(Rcmdr)
```



Introducing R Commander

At this point you should see the R Commander GUI (shown below):



The window labeled *Script Window* shows you the R commands generated based on the menu items you have chosen.

The window labeled *Output Window* shows the results of executing the commands shown in the *Script Window*.

The *Messages* window shows you notes, warnings or error messages.

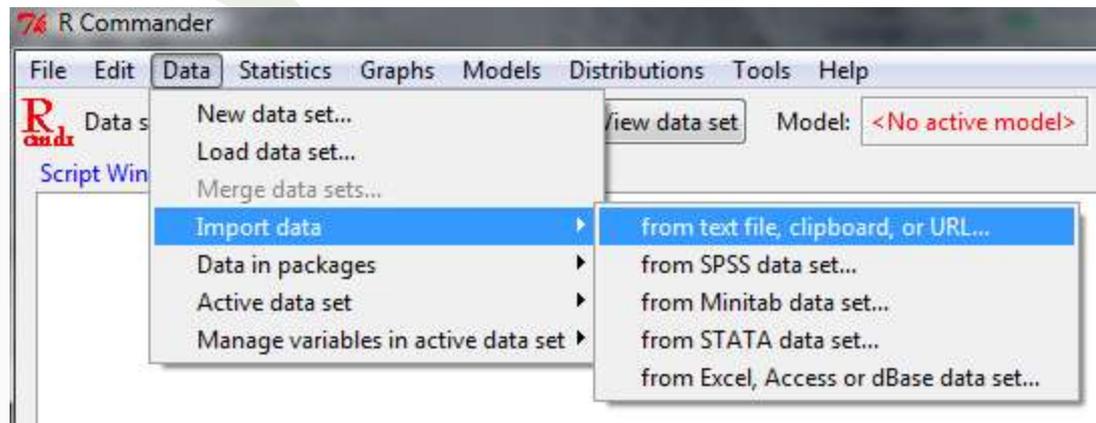
Introducing R Commander



The first task is to read in some data. R Commander can read in text files, delimited or fixed width files, Excel spreadsheets, Access database tables and much more.

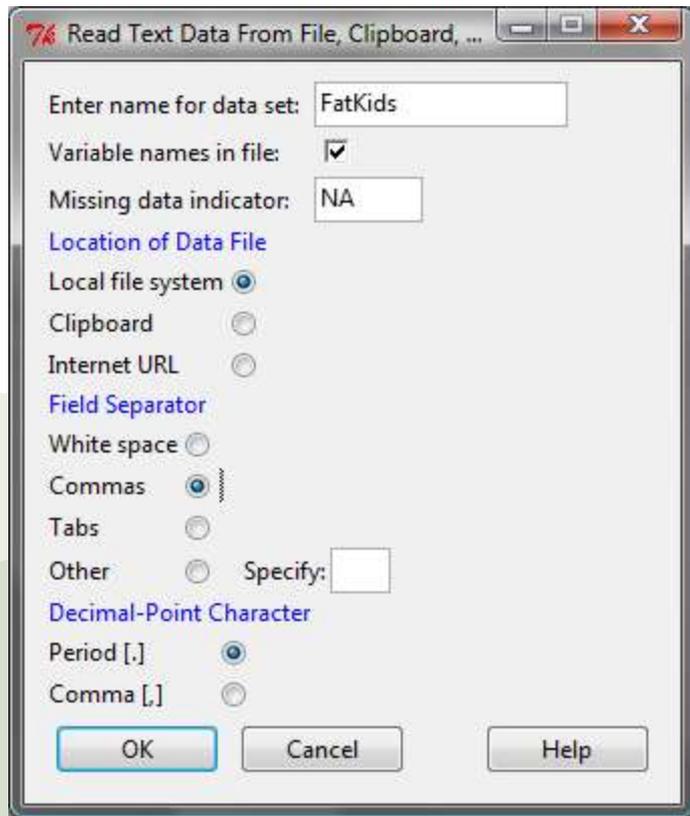
One thing it does not do well is read in SAS datasets. The R language itself can read in older SAS datasets (.*ssd*), but not the newer (.*sas7bdat*) datasets. See the packages `foreign` and `Hmisc`. R can read in SAS XPORT files, though.

So, let's read in the politically incorrect Fat Kids dataset which contains a child's first name, height (inches), weight (pounds) and the Fatty Mass Index. To read in this comma-delimited data, click on Data → Import Data → from text file, clipboard or URL...



You will see the following dialog box:

Introducing R Commander



Fill in a name for your dataset, ensure that Variable names in file is checked, ensure that Local file system is checked, ensure that Commas is checked. Click OK.

The familiar Open File Dialog box will appear. Navigate to the location of your file and click Open.

As you will see on the next slide, the R programming language statement used to read in the dataset appears in the *Script Window* and a message about how many rows were read in is printed in the *Messages* window.

```
FirstName,Height,Weight,FattyIndex
ALBERT,45,150,3.3333
ROSEMARY,35,123,3.5143
TOMMY,78,167,2.1410
BUDDY,12,189,15.7500
FARQUAR,76,198,2.6053
SIMON,87,256,2.9425
LAUREN,54,876,16.2222
```

Introducing R Commander



sheepsqueezers.com

Take note that the text to the right of the words Data set: has changed from <No active dataset> to FatKids.

To view the data in the dataset, click on the View data set button.

The screenshot shows the R Commander application window. At the top, the menu bar includes File, Edit, Data, Statistics, Graphs, Models, Distributions, Tools, and Help. Below the menu bar, there is a toolbar with buttons for 'Data set: FatKids', 'Edit data set', 'View data set', and 'Model: <No active model>'. The 'Data set: FatKids' button is highlighted with a red box. The main area is divided into two windows: the 'Script Window' and the 'Output Window'. The Script Window contains the following R code:

```
FatKids <- read.table("C:/Users/Scott/Desktop/R Series/FatKids.csv", header=TRUE, s
```

. The Output Window shows the execution of the code:

```
> FatKids <- read.table("C:/Users/Scott/Desktop/R Series/FatKids.csv", header=TRUE,
```

. A 'Submit' button is located to the right of the Output Window. At the bottom, there is a 'Messages' window showing the following text:

```
with the single-document interface (SDI); see ?Commander.  
[3] NOTE: The dataset FatKids has 7 rows and 4 columns.
```

The screenshot shows a window titled 'FatKids' displaying a table with 7 rows and 4 columns. The columns are labeled 'FirstName', 'Height', 'Weight', and 'FattyIndex'. The data is as follows:

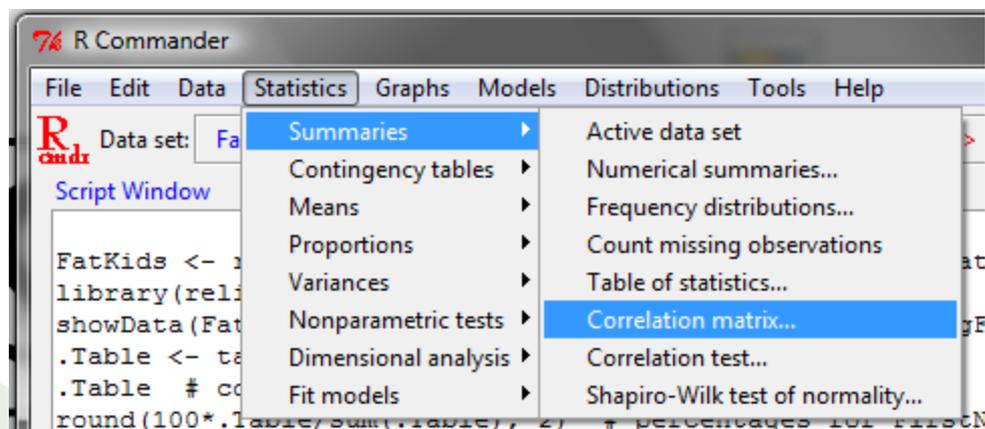
	FirstName	Height	Weight	FattyIndex
1	ALBERT	45	150	3.3333
2	ROSEMARY	35	123	3.5143
3	TOMMY	78	167	2.1410
4	BUDDY	12	189	15.7500
5	FARQUAR	76	198	2.6053
6	SIMON	87	256	2.9425
7	LAUREN	54	876	16.2222

Introducing R Commander



sheepsqueezers.com

Next, let's perform a correlation between the Height and the Weight across the dataset. Click on Statistics → Summaries → Correlation matrix...



From the dialog box that appears, choose two variables to correlate and then click OK. The results will appear in the *Output Window*:



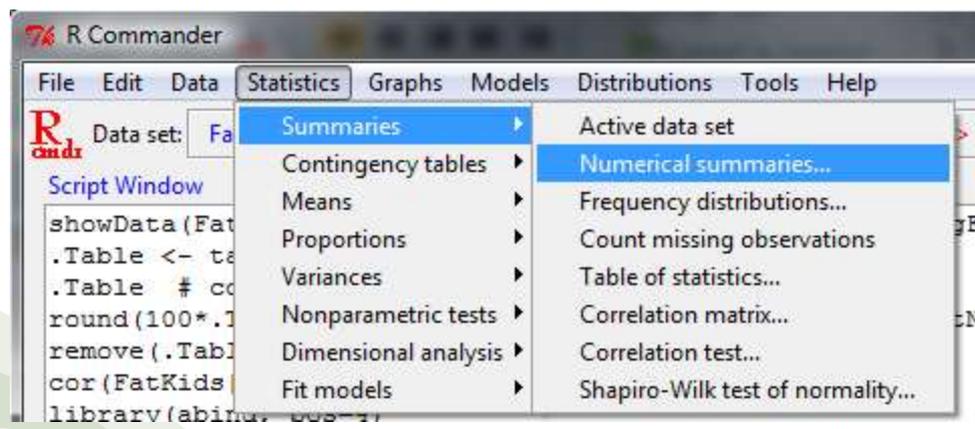
```
> cor(FatKids[,c("Height", "Weight")], use="complete.obs")  
           Height      Weight  
Height 1.00000000 0.06209589  
Weight 0.06209589 1.00000000
```

With a correlation of .062 we can conclude that fat kids have no concept of basic statistics.

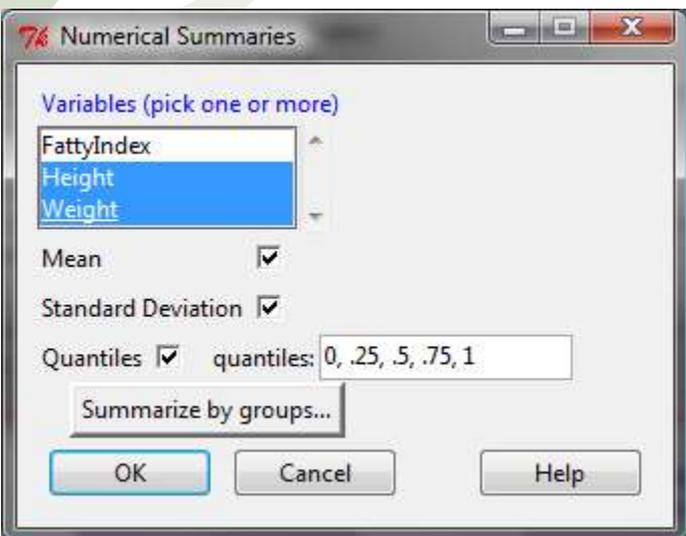


Introducing R Commander

Next, let's compute the mean, median, minimum, maximum, first quartile and third quartile of the Height and Weight variables. Click Statistics → Summaries → Numerical summaries...



From the dialog box that appears, choose the Height and Weight variables and then click OK. The results will appear in the *Output Window*:



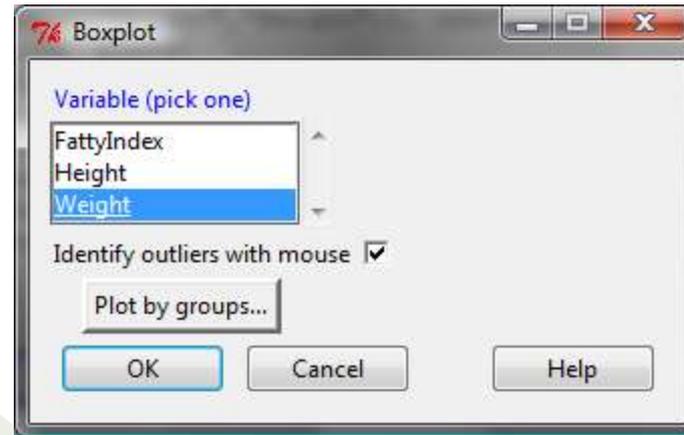
	mean	sd	0%	25%	50%	75%	100%	n
Height	55.28571	26.90548	12	40.0	54	77	87	7
Weight	279.85714	266.18128	123	158.5	189	227	876	7

Introducing R Commander

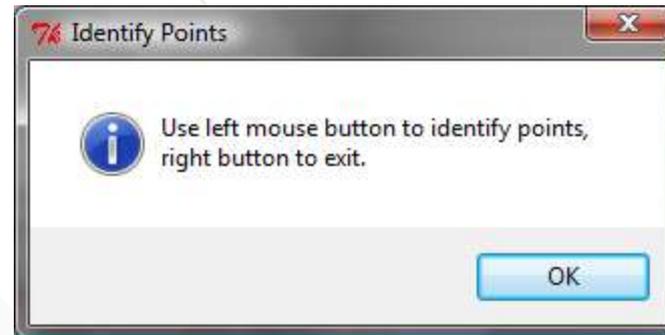


sheepsqueezers.com

R Commander also produces graphs. Let's produce a boxplot of the Weight. Click on Graphs → Boxplot... Choose Weight and ensure that the Identify outliers with mouse is checked (see below):



In order to identify the points, use the left mouse button. To exit, click the right mouse button.

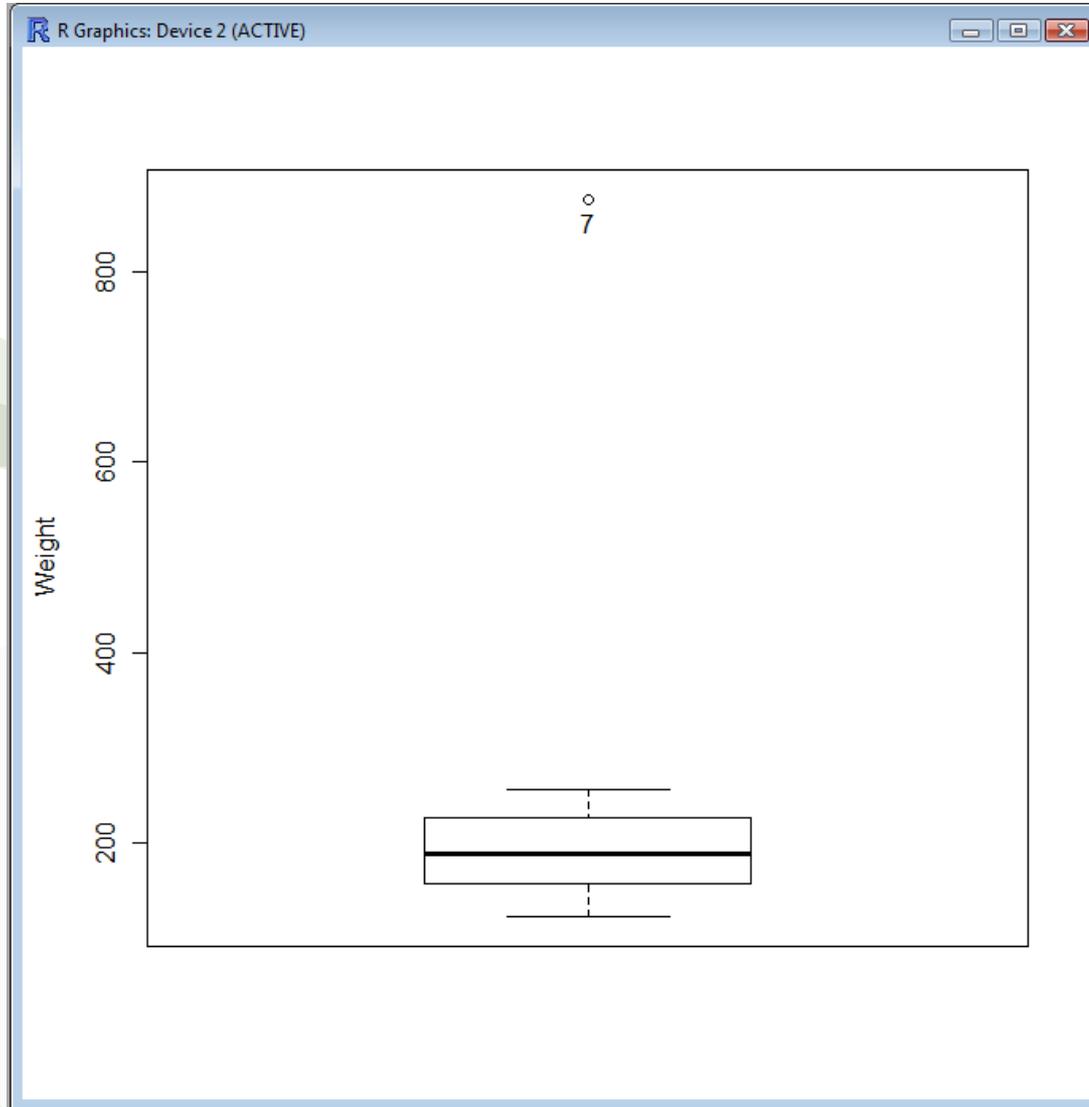


Introducing R Commander



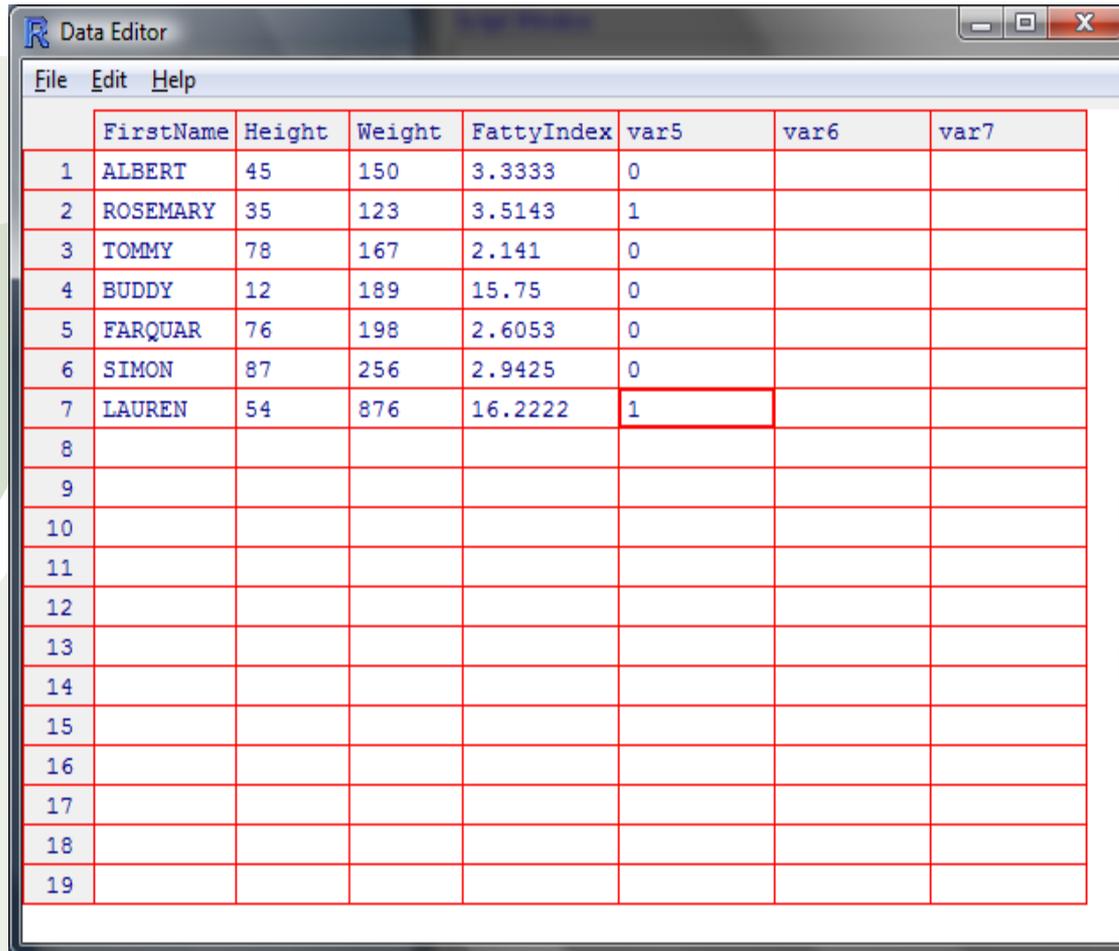
sheepsqueezers.com

When I clicked on the outlier, I was presented with the number 7 which indicates that Lauren is the outlier (with an overall best-in-class weight of 876 pounds):



Introducing R Commander

Although the correlation is abysmal, let's compute a linear regression on the Fatty Index (as dependent variable) and height and weight (as independent variables). Let's also include a gender indicator (as an additional independent variable) with 0 indicating Males and 1 indicating Females. In order to add a new variable, click on the `Edit data set` button at the top of R Commander and you will see the Data Editor grid (shown below):



	FirstName	Height	Weight	FattyIndex	var5	var6	var7
1	ALBERT	45	150	3.3333	0		
2	ROSEMARY	35	123	3.5143	1		
3	TOMMY	78	167	2.141	0		
4	BUDDY	12	189	15.75	0		
5	FARQUAR	76	198	2.6053	0		
6	SIMON	87	256	2.9425	0		
7	LAUREN	54	876	16.2222	1		
8							
9							
10							
11							
12							
13							
14							
15							
16							
17							
18							
19							

Under var5, enter in a 0 for the male kids and a 1 for the female kids. To rename the variable, click on the text `var5` and a dialog box will pop up allowing you to enter in a new variable name:



Variable editor

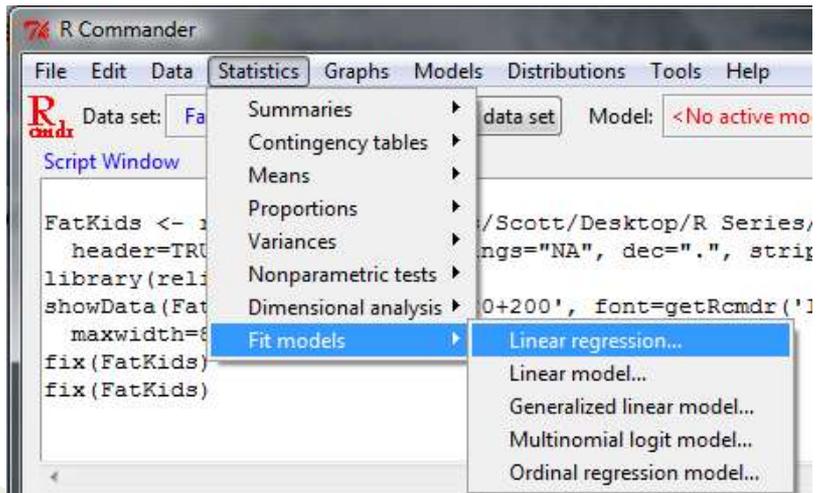
variable name:

type: numeric character

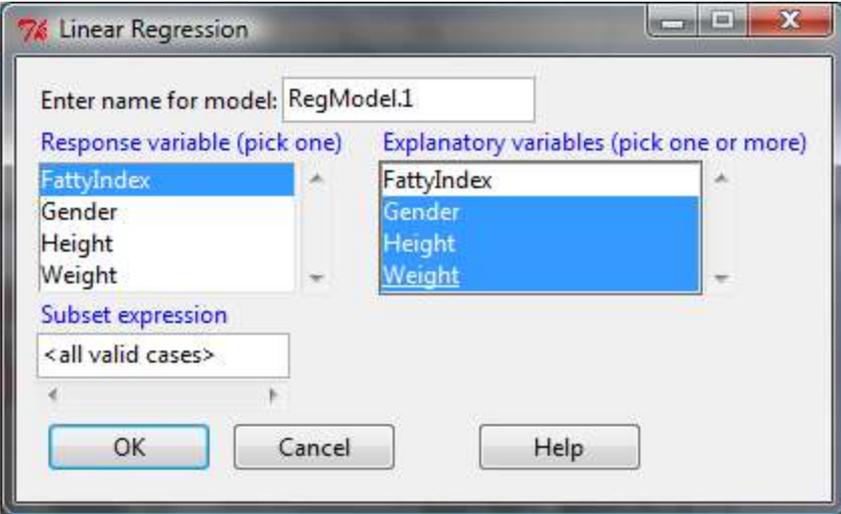
To close this dialog box, hit the Enter key. To close the Data Editor, click File...Close.

Introducing R Commander

Next, to perform a linear regression, click on Statistics → Fit models → Linear regression...



On the Linear Regression dialog box, choose FattyIndex as the dependent variable and Gender, Height and Weight as the independent variables. Click OK.



Introducing R Commander



sheepsqueezers.com

The results will appear in the *Output Window*:

```
lm(formula = FattyIndex ~ Gender + Height + Weight, data = FatKids)
```

Residuals:

```
      1      2      3      4      5      6      7
-3.71083  0.27093  0.64007  1.92384  0.05985  1.08707 -0.27093
```

Coefficients:

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  11.797573   2.648168   4.455  0.0211 *
Gender       -4.997188   2.764762  -1.807  0.1684
Height      -0.179372   0.041496  -4.323  0.0228 *
Weight       0.022122   0.004884   4.530  0.0201 *
```

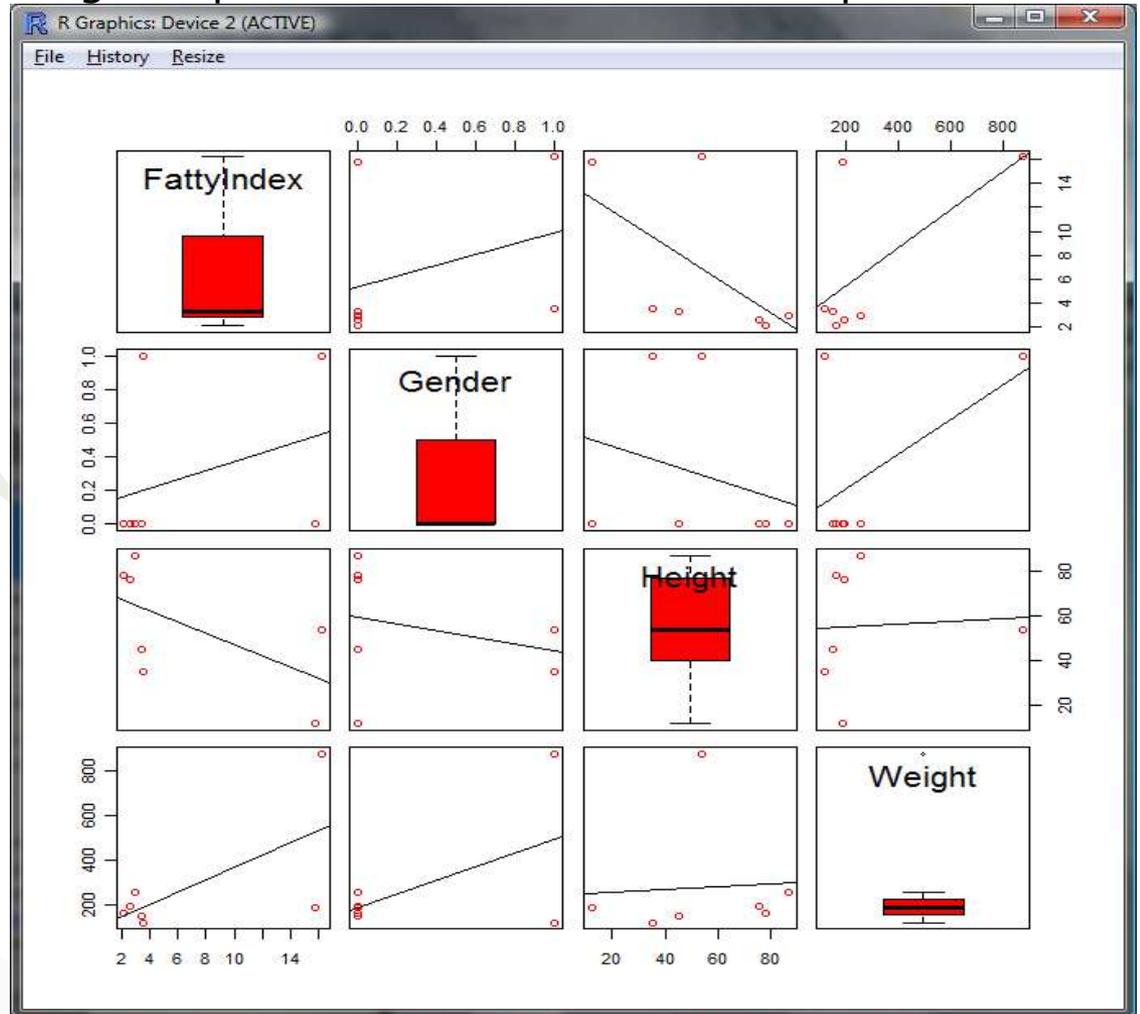
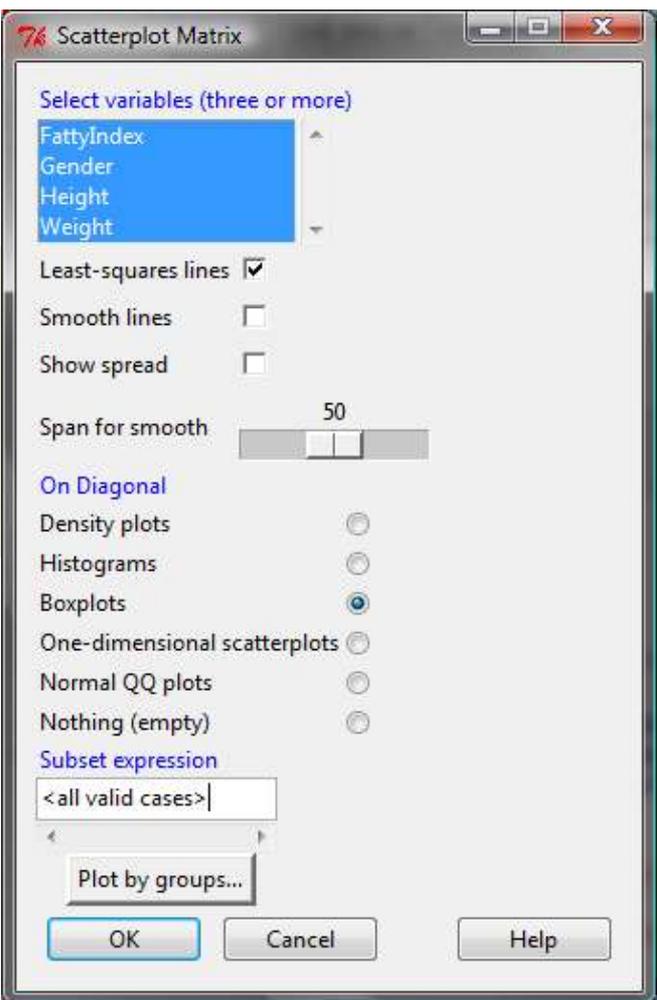
```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 2.531 on 3 degrees of freedom
Multiple R-squared:  0.9218,    Adjusted R-squared:  0.8436
F-statistic: 11.79 on 3 and 3 DF,  p-value: 0.03624
```



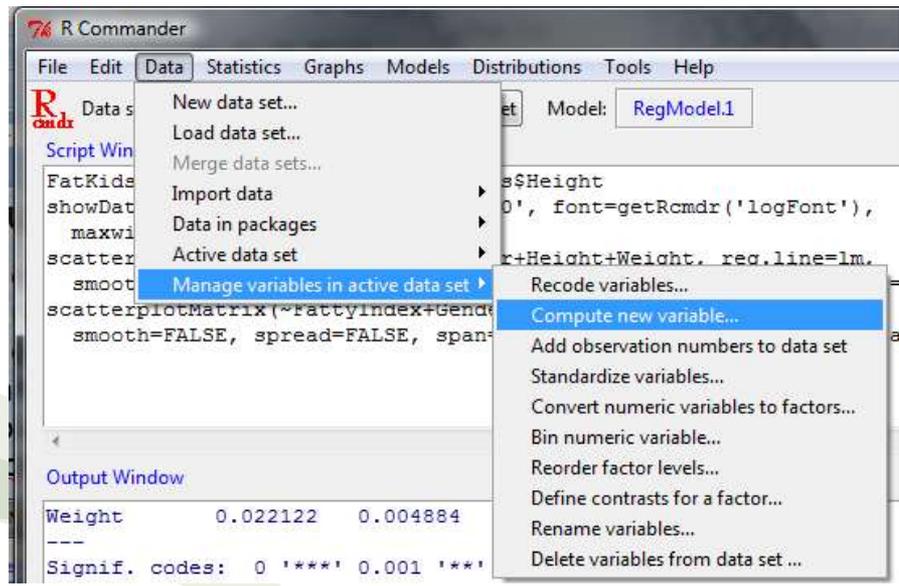
Introducing R Commander

Another nice graph R Commander produces is a scatterplot matrix. This graph contains a series of graphs based on each combination of pairs of variables. To produce a scatterplot matrix, click on Graphs → Scatterplot matrix... Choose all of the variables (not necessarily two) to graph. Below, I have chosen all four numeric variables for the scatterplot, and I've requested that the main diagonal will contain boxplots and the off-diagonal plots will also show a least squares line:

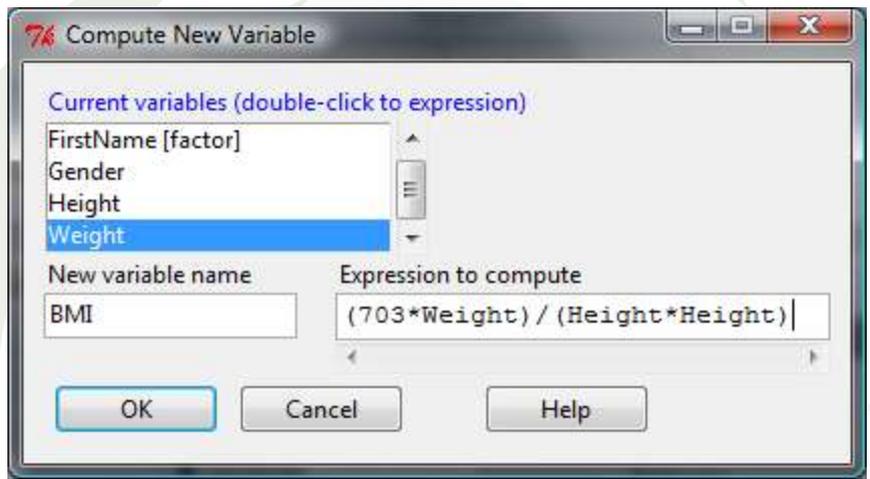


Introducing R Commander

Finally, to compute a new variable, you can click on Data → Manage variables in active data set → Compute new variable... in active data set → Compute new variable...



Let's compute the Body Mass Index. Fill in the new name of the variable along with the formula and click OK.



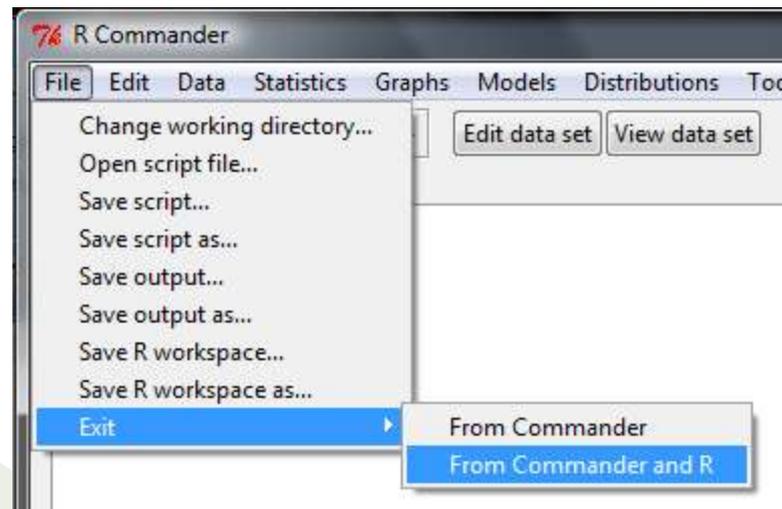
	FirstName	Height	Weight	FattyIndex	Gender	BMI
1	ALBERT	45	150	3.3333	0	52.07407
2	ROSEMARY	35	123	3.5143	1	70.58694
3	TOMMY	78	167	2.1410	0	19.29668
4	BUDDY	12	189	15.7500	0	922.68750
5	FARQUAR	76	198	2.6053	0	24.09868
6	SIMON	87	256	2.9425	0	23.77699
7	LAUREN	54	876	16.2222	1	211.18930

Introducing R Commander



sheepsqueezers.com

To exit from R Commander (as well as R), chose File → Exit → From Commander and R.



Note that you can get additional graphics and statistics by loading in the HH plugin to R Commander. Instead of starting R Commander using `library(Rcmdr)`, start it using `library(RcmdrPlugin.HH)`. You can learn more about this at

<http://cran.r-project.org/web/packages/RcmdrPlugin.HH/index.html>.

These additional options were written by Richard Heilberger and Burt Holland to support an introductory statistics course at Temple University.

There's a lot more to R Commander than what I've shown, so have at it!

Introducing *Rattle*



sheepsqueezers.com

Rattle is a GUI for data mining with R and is similar to R Commander, but not as extensive in its data management capabilities.

To run rattle, type the following at the R command prompt:

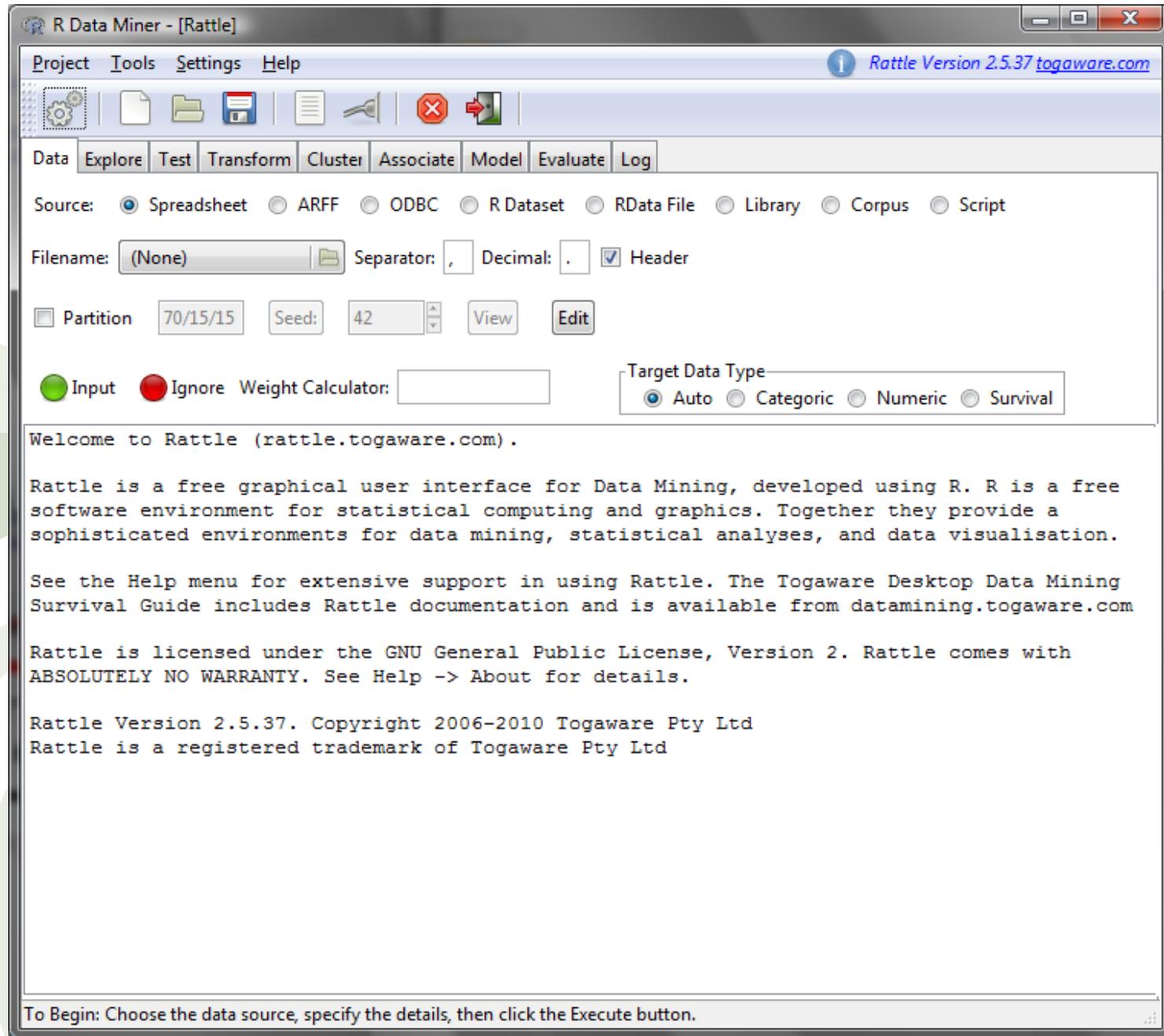
```
> library(rattle)
> rattle()
```

The following display appears (see next slide).

Introducing *Rattle*



sheepsqueezers.com





Introducing *Rattle*

In order to read in the FatKids data with Rattle, ensure that the Source radio button is set to Spreadsheet, and ensure that the Separator is a comma and that the Header is checked. Next, click on the button to the right of the word Filename:.. This will display the File Open dialog box. Find your CSV data and click the Open button. In order for anything to happen in Rattle, you will have to click on the execute button at the top left, or hit the F2 button.

R Data Miner - [Rattle (FatKids.csv)]
Rattle Version 2.5.37 togaware.com

Project Tools Settings Help

Data Explore Test Transform Cluster Associate Model Evaluate Log

Source: Spreadsheet ARFF ODBC R Dataset RData File Library Corpus Script

Filename: FatKids.csv Separator: , Decimal: . Header

Partition 70/15/15 Seed: 42 View Edit

Input Ignore Weight Calculator: Target Data Type: Auto Categorical Numeric Survival

No. Variable	Data Type	Input	Target	Risk	Ident	Ignore	Comment
1	FirstName	Categoric	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Unique: 7
2	Height	Numeric	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Unique: 7
3	Weight	Numeric	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Unique: 7
4	FattyIndex	Numeric	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Unique: 7

Roles noted. 7 observations and 4 input variables. No target thus no predictive modelling nor sampling.

Introducing *Rattle*



sheepsqueezers.com

Click the View button to see the data.

Warning: Changes will not be saved.

Warning: Changes	FirstName	Height	Weight	FattyIndex
1	ALBERT	45	150	3.333
2	ROSEMARY	35	123	3.514
3	TOMMY	78	167	2.141
4	BUDDY	12	189	15.75
5	FARQUAR	76	198	2.605
6	SIMON	87	256	2.942
7	LAUREN	54	876	16.22
	NA			

Next, click the Explore tab and select all of the information you want to see about your data. Don't forget to hit the F2 button to execute the tab.

Introducing *Rattle*



sheepsqueezers.com

Below we summarise the dataset.

The data is limited to the training dataset.

Data frame: crs\$dataset[crs\$sample,] 4 observations and 4 variables Maximum # NAs: 0

```
Levels Storage
FirstName    7 integer
Height      integer
Weight      integer
FattyIndex  double
```

```
+-----+-----+
|Variable |Levels                                     |
+-----+-----+
|FirstName|ALBERT, BUDDY, FARQUAR, LAUREN, ROSEMARY, SIMON, TOMMY|
+-----+-----+
```

For the simple distribution tables below the 1st and 3rd Qu. refer to the first and third quartiles, indicating that 25% of the observations have values of that variable which are less than or greater than (respectively) the value listed.

FirstName	Height	Weight	FattyIndex
ALBERT :0	Min. :12.00	Min. :123.0	Min. : 2.942
BUDDY :1	1st Qu.:29.25	1st Qu.:172.5	1st Qu.: 3.371
FARQUAR :0	Median :44.50	Median :222.5	Median : 9.632
LAUREN :1	Mean :47.00	Mean :361.0	Mean : 9.607
ROSEMARY:1	3rd Qu.:62.25	3rd Qu.:411.0	3rd Qu.:15.868
SIMON :1	Max. :87.00	Max. :876.0	Max. :16.222
TOMMY :0			

Introducing *Rattle*



sheepsqueezers.com

Click the Test tab and click on the radio button to the left of the word Correlation. Select Height as Sample 1 and Weight as Sample 2. Click F2.

Correlation Test

The paired sample correlation test is performed on the two specified samples. The two samples are expected to be paired (two observations for the same entity). The null hypothesis is that the two samples have no (i.e., 0) correlation. Pearson's product moment correlation coefficient is used.

If the p-value is less than 0.05 then we reject the null hypothesis and accept the alternative hypothesis that the samples are correlated, at the 95% level of confidence.

The two samples being compared are the two variables, 'Height' and 'Weight'

Title:

Pearson's Correlation Test

Test Results:

PARAMETER:

Degrees of Freedom: 5

SAMPLE ESTIMATES:

Correlation: 0.0621

STATISTIC:

t: 0.1391

P VALUE:

Alternative Two-Sided: 0.8948

Alternative Less: 0.5526

Alternative Greater: 0.4474

CONFIDENCE INTERVAL:

Two-Sided: -0.7249, 0.7787

Less: -1, 0.7087

Greater: -0.6412, 1

Introducing *Rattle*



sheepsqueezers.com

Briefly, the Transform tab allows you to scale or rank your data.

The Cluster tab allows you to perform hierarchical, Kmeans, Clara and BiCluster clustering on numeric variables.

The Associate tab allows you to perform Association Rules Analysis.

The Model tab allows you to create Decision Tree Models.

The Evaluate tab allows you to create Error Matrices (aka, confusion matrices).

The Log tab shows you the Rattle Log File similar to a SAS Log file.

To exit Rattle, click on Project...Quit.

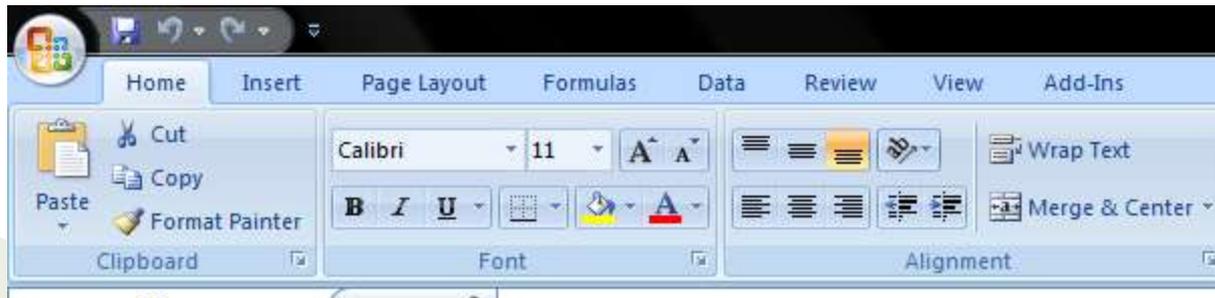
There's more to this package, so go forth and play.

Introducing RExcel

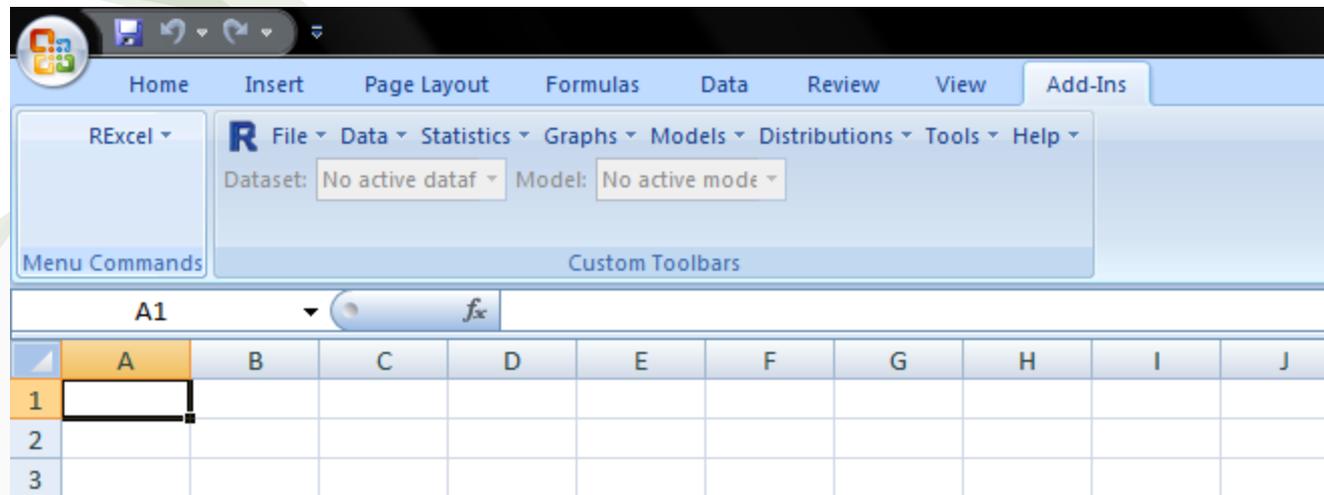


sheepsqueezers.com

RExcel is an Excel interface to the R software. RExcel is responsible for starting Excel, the R Console (RGui) as well as R Commander. To start RExcel, double-click on the RExcel icon on your desktop. Wait for R, R Commander and Excel to start. RExcel creates an additional tab labeled Add-Ins on the menu bar.



If you click on the Add-Ins menu, you will see the following:



The Add-Ins menu shows the following menus: File, Data, Statistics, Graphs, Models, Distributions, Tools and Help. These are the R Commander menus, so RExcel is using R Commander to do its dirty work.

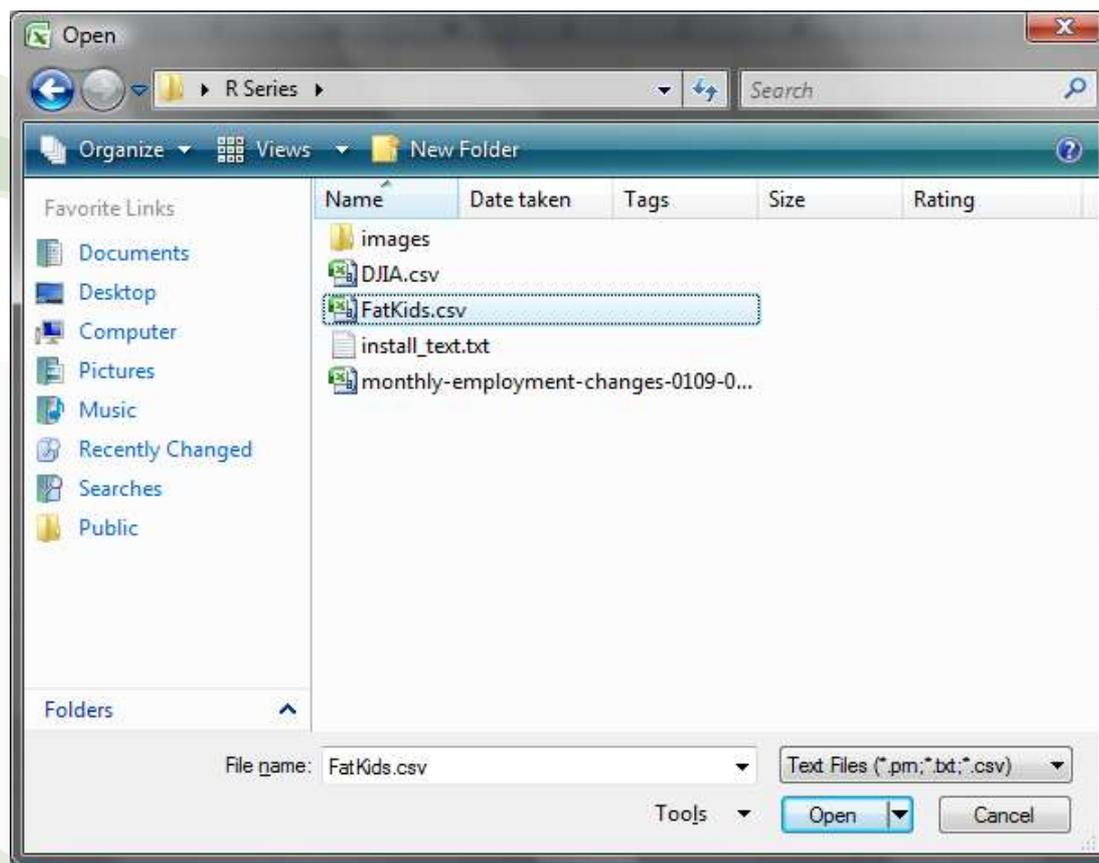
Introducing RExcel



sheepsqueezers.com

Yes, you can perform all of the R Commander functions within the Add-Ins menu, but the beauty of RExcel is to be able to provide data to R and R Commander *from within Excel itself*.

Open up an Excel workbook or CSV file. Here, we will open up the FatKids CSV file in Excel using the File Open dialog box ensuring that we select Text Files as the filetype. Click on FatKids.csv then click the Open button.



Introducing RExcel



sheepsqueezers.com

As you see below, the FatKids data has been read into Excel, but at this point, R has no idea that this data exists. In order to tell R about it – that is, to have R read in this data and create an R dataset (aka, dataframe) from it in R – you have to highlight the data, right-click and select Put R DataFrame to load the data into R (see next slide).

The screenshot shows the Microsoft Excel interface with the RExcel add-in ribbon active. The ribbon includes a menu with options: File, Data, Statistics, Graphs, Models, Distributions, Tools, and Help. Below the menu, there are dropdowns for 'Dataset: No active dataf' and 'Model: No active mode'. The main spreadsheet area shows a table with columns A through I and rows 1 through 9. The data is as follows:

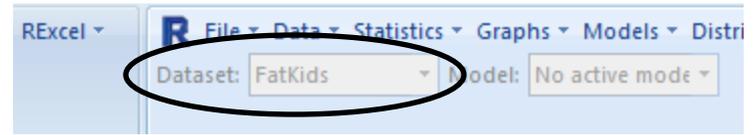
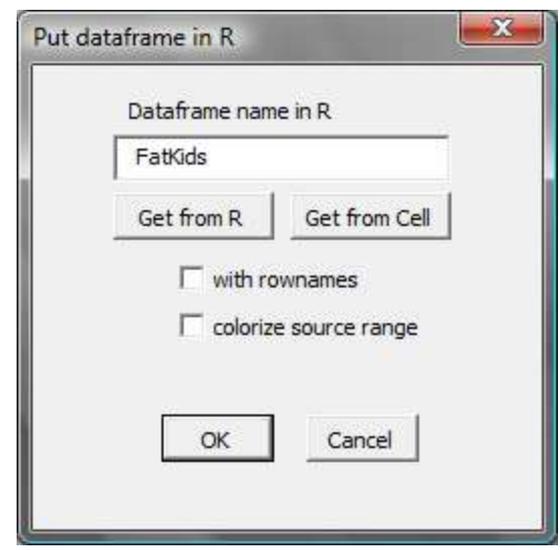
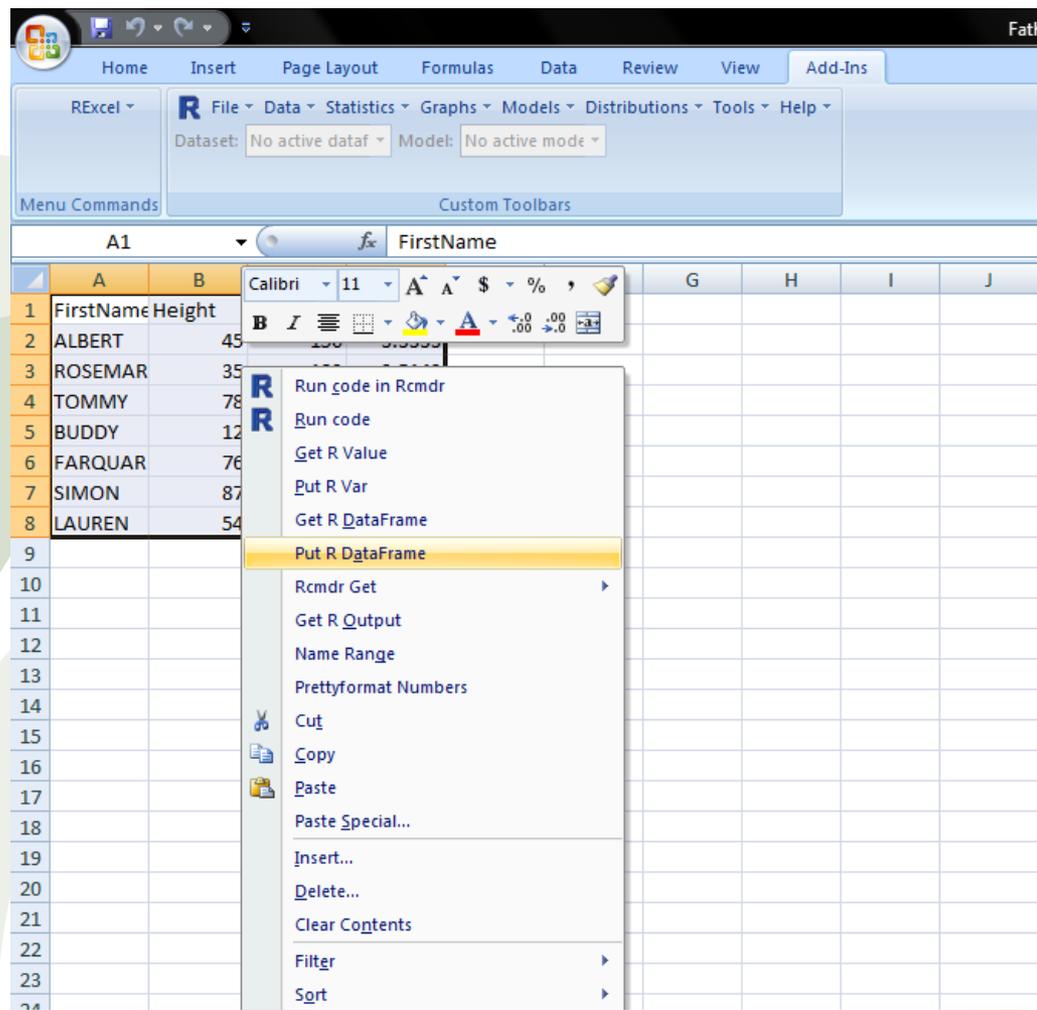
	A	B	C	D	E	F	G	H	I
1	FirstName	Height	Weight	FattyIndex					
2	ALBERT	45	150	3.3333					
3	ROSEMAR	35	123	3.5143					
4	TOMMY	78	167	2.141					
5	BUDDY	12	189	15.75					
6	FARQUAR	76	198	2.6053					
7	SIMON	87	256	2.9425					
8	LAUREN	54	876	16.2222					
9									

Introducing RExcel



sheepsqueezers.com

As you can see below left, there are several additional actions available to you through this popup menu. Now, once you click on Put R DataFrame, you will see the Put dataframe in R dialog box (shown right). Give a name to your dataset and click OK. Take note that the dataset name has changed to FatKids.





Introducing RExcel

At this point, you can run any R Commander-type analysis you want on that data. For example, let's perform a linear regression with FattyIndex as our dependent variable and both Height and Weight as independent variables. To do this, click on Statistics → Fit Models → Linear Regression... Fill in the Linear Regression dialog box and click OK.

The screenshot shows the RExcel interface with the 'Statistics' menu open. The 'Fit models' option is selected, and the 'Linear regression...' option is highlighted. The background shows a spreadsheet with the following data:

	A	B	C	F	G	H	I
1	FirstName	Height	Weight				
2	ALBERT	45	1				
3	ROSEMAR	35	1				
4	TOMMY	78	167	2.141			
5	BUDDY	12	189	15.75			
6	FARQUAR	76	198	2.6053			
7	SIMON	87	256	2.9425			
8	LAUREN	54	876	16.2222			

The 'Linear Regression' dialog box is shown with the following settings:

- Enter name for model: RegModel.1
- Response variable (pick one): FattyIndex
- Explanatory variables (pick one or more): FattyIndex, Height, Weight
- Subset expression: <all valid cases>

Buttons: OK, Cancel, Help



Just as when we used R Commander, the results show up in the R Commander *Output Window* and do NOT show up in Excel.

```
Call:
lm(formula = FattyIndex ~ Height + Weight, data = FatKids)

Residuals:
    1      2      3      4      5      6      7
-2.6875 -3.5661  0.8269  4.0827  0.4684  1.4963 -0.6207

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 10.310812   3.150720   3.273   0.0307 *
Height      -0.151273   0.048157  -3.141   0.0348 *
Weight       0.016782   0.004868   3.448   0.0261 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.168 on 4 degrees of freedom
Multiple R-squared:  0.8367,    Adjusted R-squared:  0.755
F-statistic: 10.24 on 2 and 4 DF,  p-value: 0.02668
```

Now, if you already have data loaded into R as a dataframe, you can bring that data from R into Excel. R comes with a lot (!) of dataframes for you to play with (see the `datasets` package for more...type in `library(help="datasets")` to see a complete list). Let's load in the `cars` dataset which contains the speed and stopping distance of cars. Type in `data("cars")` at the R command prompt to load in the data, then type in `cars` to see the data. This dataset contains two variables, speed and dist.

Introducing RExcel



sheepsqueezers.com

To pull this data from R into Excel, right-click on an empty cell and click Get R DataFrame. Click the Get from R button and select cars.

The screenshot shows the Excel interface with the RExcel add-in. The RExcel ribbon is active, showing the 'R' tab with a 'File' menu. The 'Dataset' is set to 'FatKids' and the 'Model' is 'RegModel'. A right-click context menu is open over cell A1, with 'Get R DataFrame' selected. Two dialog boxes are overlaid: 'Get dataframe from R' and 'Select dataframe'. The 'Get dataframe from R' dialog has 'Get from R' selected, 'with rownames' checked, and 'colorize target range' unchecked. The 'Select dataframe' dialog shows 'cars' selected in the list.

Click OK after selecting cars and the cars data shows up in Excel.

Introducing RExcel



sheepsqueezers.com

As you can see below, the cars data shows up on the spreadsheet. But, take note that the Dataset indicator still says FatKids. Change this by clicking the down-arrow and selecting cars. This way, you can then perform analysis on the cars data.

The screenshot shows the RExcel interface with a spreadsheet containing car data. The Dataset indicator is set to FatKids. The spreadsheet has columns A, B, C, D, and E, and rows 1 through 25. The data in the spreadsheet is as follows:

	A	B	C	D	E
1		speed	dist		
2	1	4	2		
3	2	4	10		
4	3	7	4		
5	4	7	22		
6	5	8	16		
7	6	9	10		
8	7	10	18		
9	8	10	26		
10	9	10	34		
11	10	11	17		
12	11	11	28		
13	12	12	14		
14	13	12	20		
15	14	12	24		
16	15	12	28		
17	16	13	26		
18	17	13	34		
19	18	13	34		
20	19	13	46		
21	20	14	26		
22	21	14	36		
23	22	14	60		
24	23	14	80		
25	24	15	20		

Introducing *RExcel*



sheepsqueezers.com

Another nice feature is that you can place an R command in an Excel cell as long as you know the command line syntax for the function. For example, you can compute the mean using the mean function, but a nice feature of the mean function in R is that you can trim the mean. That means a certain percentage from the top and bottom will be removed. For example, to compute the 10% trimmed mean on the distance variable `dist` in the `cars` dataframe using the R mean function, you can issue the following command at the R command prompt:

```
> mean(cars$dist, 0.1)
```

In Excel, you can use the `RApply` function to perform this calculation. Type in the following in a blank cell (on the `cars` spreadsheet, say):

```
=RApply("mean", C2:C51, 0.1)
```

You should see the following (I entered additional examples) on the next slide...

Introducing RExcel



	A	B	C	D	E	F	G	H
1		speed	dist					
2	1	4	2					
3	2	4	10					
4	3	7	4					
5	4	7	22					
6	5	8	16		Trimming	R Mean Function	Excel AVERAGE Function	
7	6	9	10		0	42.98	42.98	
8	7	10	18		0.1	40.875		
9	8	10	26		0.2	39.46666667		
10	9	10	34					
11	10	11	17					
12	11	11	28					

As long as you know the syntax of the R command you want to run, you can use it from within an Excel cell. That's a-nice!!

Introducing *RExcel*

Clearly, there is a lot more to RExcel than what I've discussed. I'll leave the rest up to you, my friends...*Vaya Con RExcel!*



sheepsqueezers.com

Introducing *RGGobi*

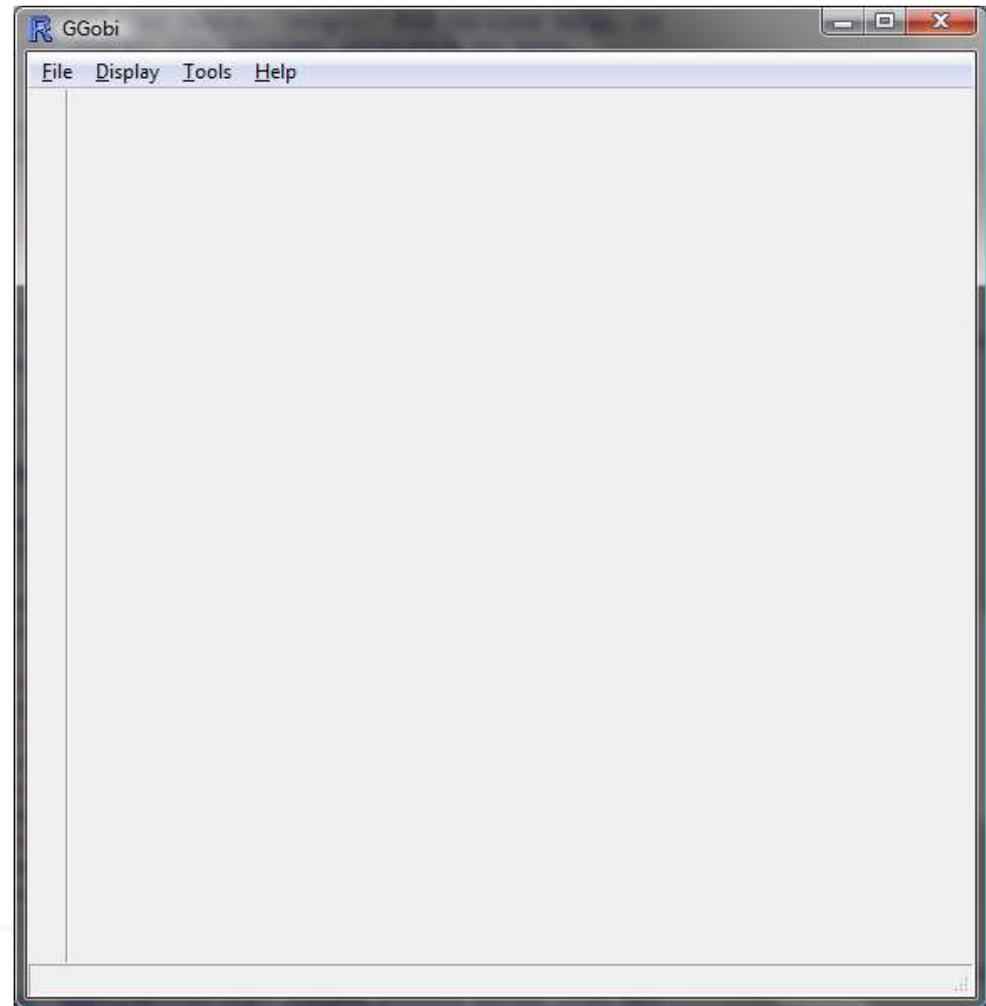


sheepsqueezers.com

The RGGobi package provides an interface to GGobi, an interactive and *dynamic* graphics package. The graphics you've seen so far have been *static* graphics.

To start RGGobi, enter the following commands from the R command line and hit the Enter key:

```
> library(rggobi)
> ggobi()
```

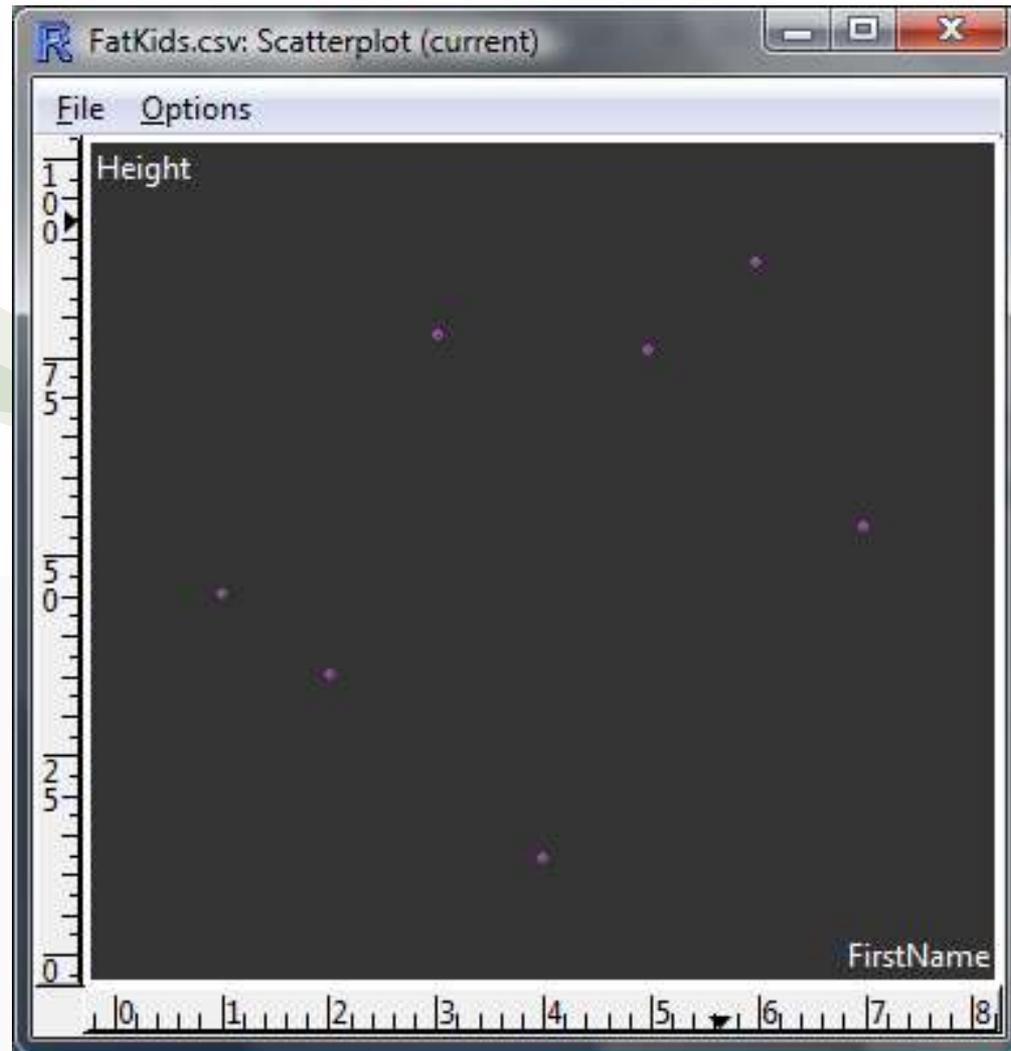


Installing *RGGobi*



sheepsqueezers.com

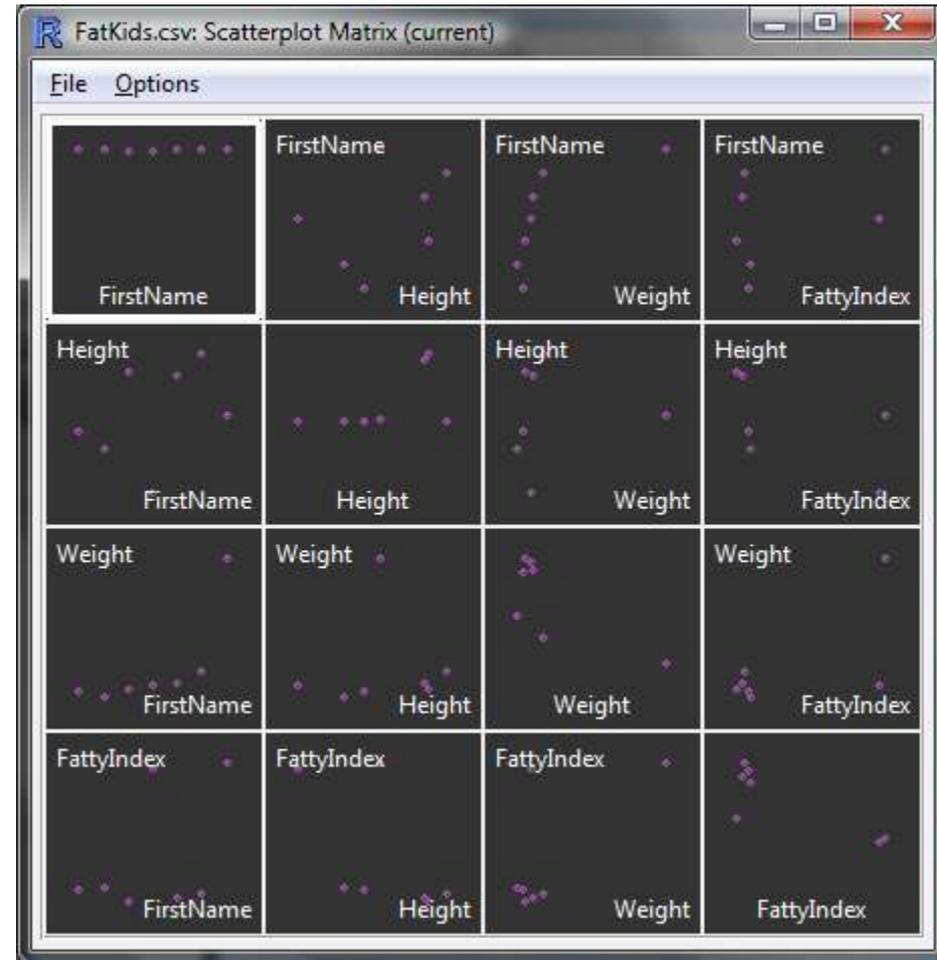
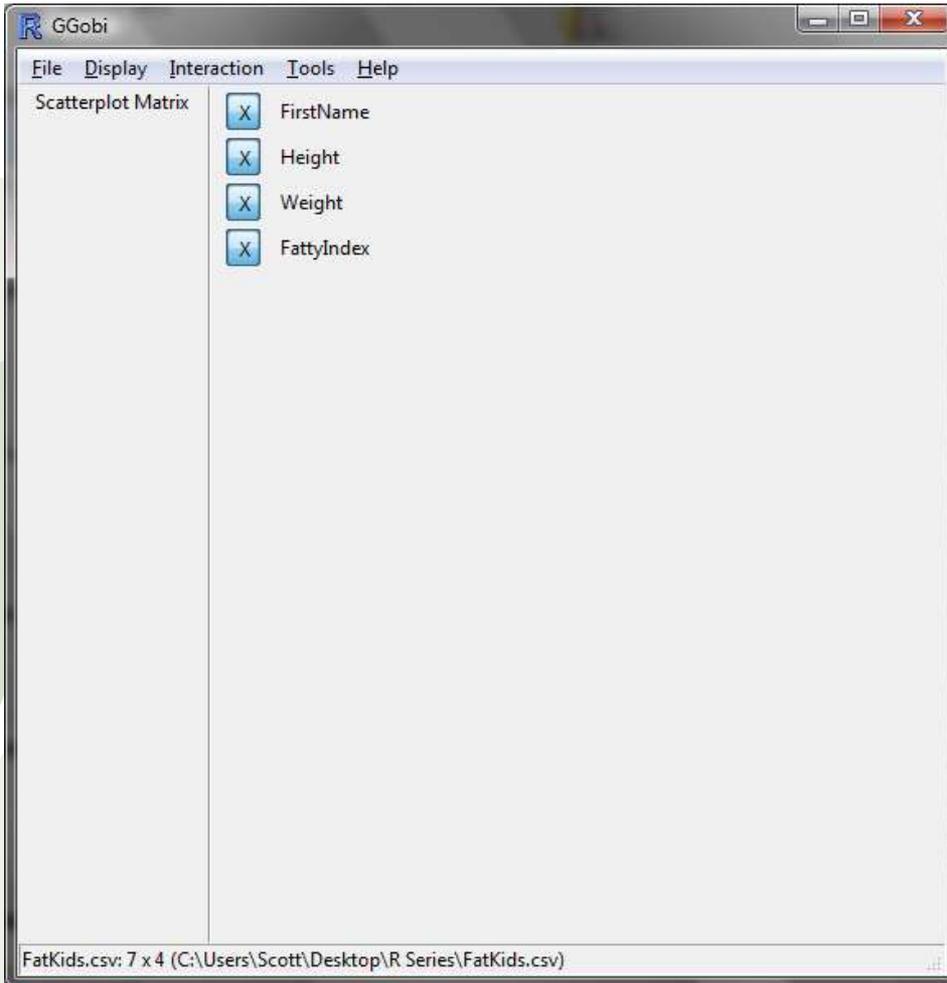
In order to open up the FatKids CSV file, click on File...Open, navigate to the FatKids.csv file and click the Open button. Once open, GGobi will open a second window as shown below.





Installing *RGGobi*

You can close the additional window by clicking File...Close. Then, to do more interesting things, go back to the GGobi main window and click on Display...New Scatterplot Matrix. You will see the following display (both windows displayed):

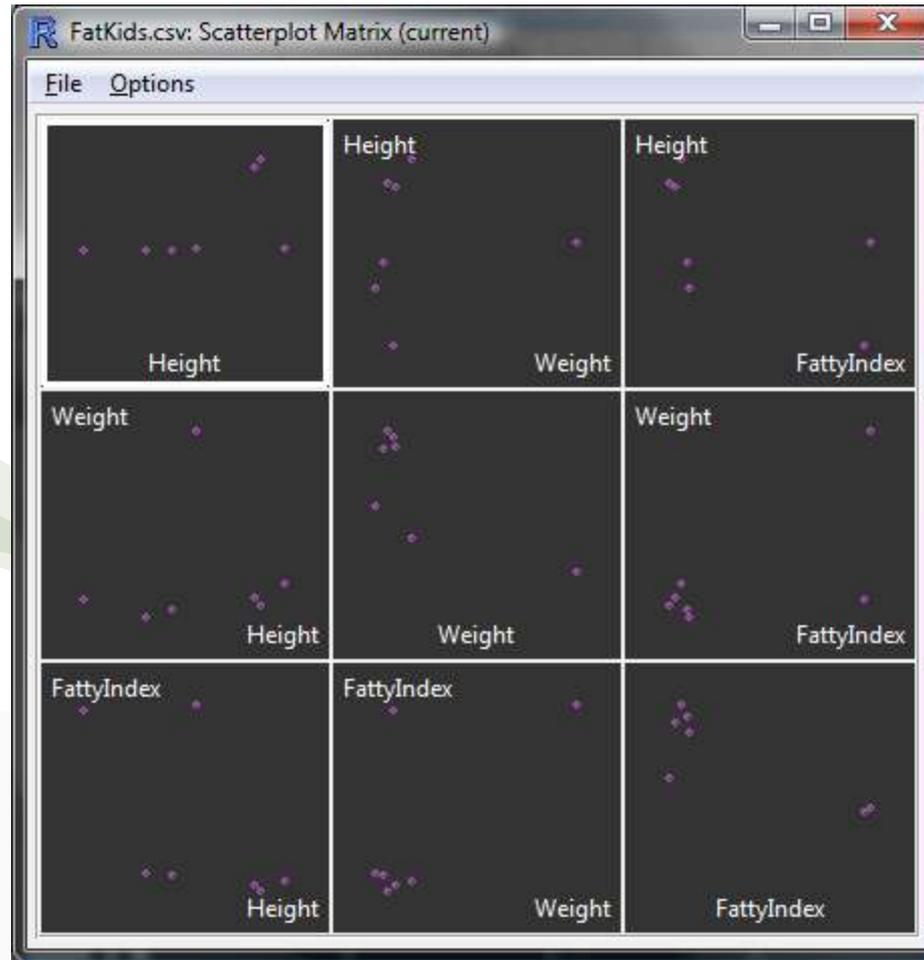


Click the X button to the left of FirstName to remove it from the scatterplot matrix.



Installing *RGGobi*

You will now see the following display:



There is nothing *interactive* about this graph. To make it slightly more useful, click on Interaction...Identify on the main GGobi window. You will see the following screen (next slide, left). Click on the variable `FirstName` on the left side to help identify each point.



Installing RGGobi

You will now see the following displays. As you move your mouse pointer over each point, the FirstName will be displayed. Note that you must first click on one of the scatterplots in order to activate it and THEN you can hover your mouse over a point.

GGobi

File Display Interaction Tools Help

Identify

Points

Record Label

Record Number

FirstName

Height

Weight

FattyIndex

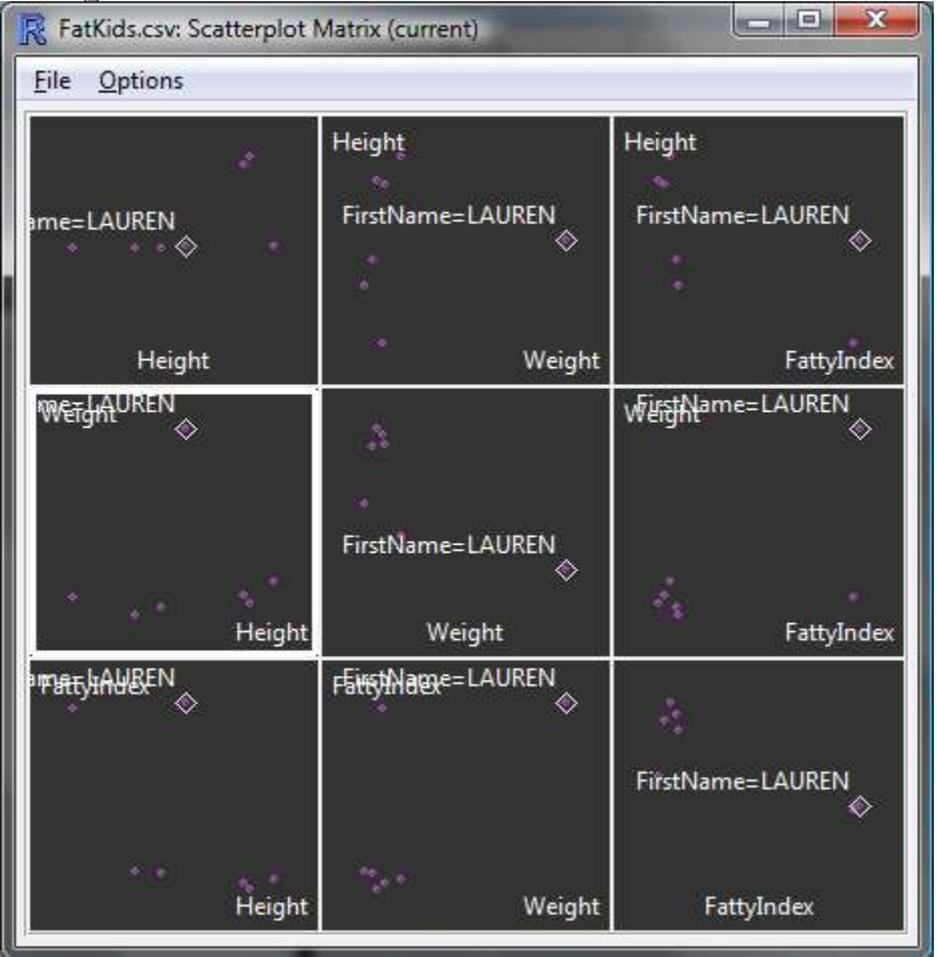
Remove labels

Label all

Recenter data

Recenter

FatKids.csv: 7 x 4 (C:\Users\Scott\Desktop\R Series\FatKids.csv)



Installing *RGGobi*



sheepsqueezers.com

There's much more the GGobi than just this...visit the GGobi website or download some manuals on this package.

Note that you can use R Commander to load the FatKids dataset into R, quit out of R Commander and then start GGobi with the FatKids dataset. This prevents having to read in the dataset using GGobi which has limited "read" functionality as compared to R Commander.

```
>library(rggobi); ggobi(FatKids);
```



Appendix

Appendix A: References



sheepsqueezers.com

Click the titles below to be taken to Amazon.com's website.

- [SAS and R](#), 1st Edition, Ken Kleinman and Nicholas J. Horton (ISBN: 9781420070576)
- [R for SAS and SPSS Users](#), 1st Edition, Robert A. Muenchen (ISBN: 9780387094175)
- [Data Manipulation with R](#), 1st Edition, Phil Spector (ISBN: 9780387747309)
- [R In A Nutshell](#), Joseph Adler (ISBN: 9780596801700)
- [Software for Data Analysis: Programming with R](#), John M. Chambers (ISBN: 9780387759357)
- [R Through Excel](#), Richard Heiberger and Erich Neuwirth, (ISBN: 9781441900517)
- [The R Book](#), Michael J. Crawley (ISBN: 9780470510247)
- [Interactive and Dynamic Graphics for Data Analysis with R and GGobi](#), Dianne Cook and Deborah F. Swayne (ISBN: 9780387717616)
- [Lattice: Multivariate Data Visualization with R](#), Deepayan Darkar (ISBN: 9780387759685)
- [ggplot2: Elegant Graphics for Data Analysis](#), Hadley Wickham (ISBN: 9780387981406)
- [Introductory Statistics with R](#), 2nd Edition, Peter Dalgaard (ISBN: 9780387790541)
- [Modern Applied Statistics with S](#), W.N. Venables and B.D. Ripley (ISBN: 9781441930088)
- Manuals Provided with R Software: [An Introduction to R](#), [R Data Import/Export](#), [R Language Definition](#), [R Installation and Administration](#), [R Internals](#), [Writing R Extensions](#)
- GGobi Manual (<http://www.ggobi.org/rggobi/introduction.pdf>), Deborah F. Swayne, Hadley Wickham, et. al.
- cran.r-project.org/web/packages/RcmdrPlugin.HH/RcmdrPlugin.HH.pdf: Documentation for the RcmdrPlugin.HH plug-in to R Commander.

Appendix B: R-Related Websites



sheepsqueezers.com

- ❑ www.r-project.org: This is the main R Software Site and contains the software itself for various platforms as well as documentation.
- ❑ cran.r-project.org: This is the website of the Comprehensive R Archive Network (CRAN) and it houses all of the R packages you could ever possibly want.
- ❑ journal.r-project.org: This is the website of the R Journal which is a refereed journal of the R Project and contains articles on a variety of topics related to R.
- ❑ wiki.sciviews.org/doku.php: This is the R Wiki website and houses a variety of searchable content.
- ❑ sourceforge.net: This is the main SourceForge website and contains free software not necessarily related to R (but does contain a lot of R-related software).
- ❑ r-forge.r-project.org: This is the main Rforge website and contains R-related development software such as packages, graphical user interfaces, etc.
- ❑ rcom.univie.ac.at: R and Friends website with RCOM package and also R bundled with RCOM.
- ❑ www.ggobi.org/downloads: GGobi website with free download of GGobi.
- ❑ www.ggobi.org/demos: GGobi demo movies and screen shots.
- ❑ cran.r-project.org/web/packages/RcmdrPlugin.HH: Website for the RcmdrPlugin.HH plug-in to R Commander.



Support sheepsqueezers.com

If you found this information helpful, please consider supporting sheepsqueezers.com. There are several ways to support our site:

- Buy me a cup of coffee by clicking on the following link and donate to my PayPal account: [Buy Me A Cup Of Coffee?](#).
- Visit my Amazon.com Wish list at the following link and purchase an item:
<http://amzn.com/w/3OBK1K4EIWIR6>

Please let me know if this document was useful by e-mailing me at comments@sheepsqueezers.com.