

National Oceanographic Atmospheric Administration Stock Assessments

Mentor:

Michael Prager, Ph.D.

Population Dynamics Team, NMFS SE Fisheries Science Center

NOAA Center for Coastal Fisheries and Habitat Research

Beaufort, North Carolina 28516

Authors:

Anthony Anderson and Kaiem Frink

Abstract - We developed an interface between two software packages we use, AD Model Builder (ADMB) and the statistics package R. Both packages offer high-level programming languages. We used ADMB language to fit models, and then used R to graph them. When fitting a model with ADMB, a mass of data is generated that must be graphed to understand the modeling results. Our interface contains code that allows an ADMB program to output data in a format readable by R, and it also contains a set of graphics functions in R that make dozens of standard graphs.

The benefits to NOAA and the Population Dynamics Team will be mainly an improved graphics function. We have increased familiarity with scientific programming in general, and programming in the R language in particular

I. Introduction

The Population Dynamics Team at the NOAA Beaufort Lab is heavily involved in numerical modeling of wild fish stocks, including the effects of fishing on population assessment. This modeling activity is known as stock assessment, and its results typically contribute to management of fish stocks in Federal

waters. Our models estimate, among other quantities, the fishing mortality rate in various fisheries (e.g., commercial fishery, hook-and-line fishery, etc.). To model fish populations we refine and implement most of the mathematical models ourselves, using a variety of computer languages.

R is a predecessor of S which was developed at Bell Laboratories, in the past AT&T, and now Lucent Technologies. The language and environment is used for statistical computing and graphics. R has been considered as a being very similar to S but with a different method of implementation of S.

The most important aspect of R is its ability to produce “well-designed publication-quality” plots. In addition to graphing capabilities, R can produce mathematical symbols and formulate where needed. The default settings are set for general use, but can be changed according to the user’s specifications.

II. Methodology

The scatterplot program required the proper software package considering the operation system. The program is

compatible with Microsoft Windows 95, 98, NT4, 2000, ME and XP, and the file size of 25.2 megabytes. A full installation takes up about 50Mb of disk space and a minimal one about 18Mb. There are no special requirements because the program is stored as a standard executable file. To improve the format of the R source code; Tinn-R was used to enhance readability and structure.

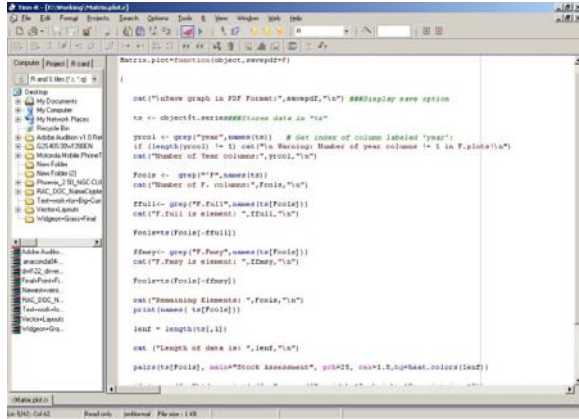


Figure1. Tinn-R

A scatterplot matrix allows the user to analyze data as a whole. The command “pairs” produces a pairwise scatterplot matrix of the variables defined by the columns of X that is, every column of X is plotted against every other column of X.

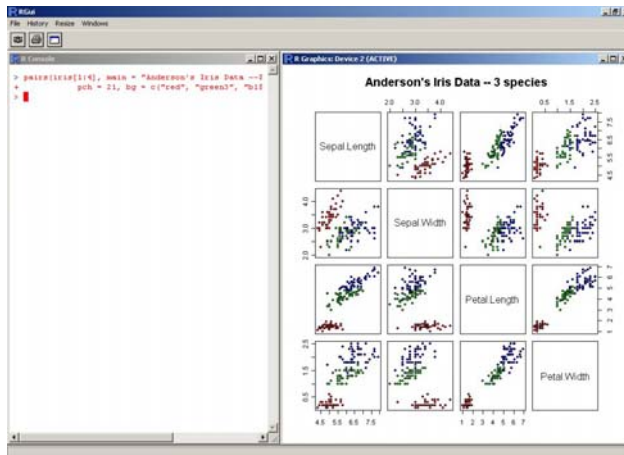


Figure2. Scatterplot Matrix

The default settings produce a basic graph with no color variation and minimal

details. R allows to user to add different plot shapes and colors to place emphasis on the data. Depending on the size of the graph the font and plots can be adjusted to provide clarity.

The data is introduced into R as an .rdat file which has been written in such a way that R can develop vectors, and matrixes.

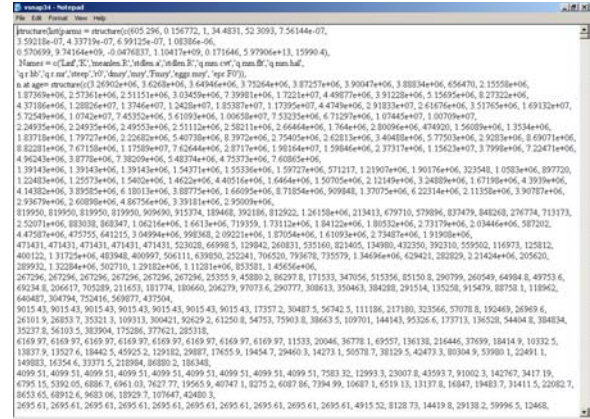


Figure3. “Vsnap34.rdat” data

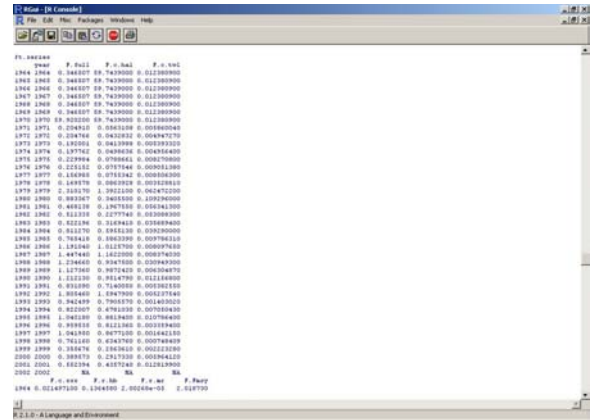


Figure4. R Console

Due to the overwhelming amount of data, the user would have to extract certain data sets to be analyzed each time the command is run. R is capable of storing data sets as variable to make them easily access or applied to the console for evaluation. The command “dget” retrieves the data set and allows the user to input the variable to represent the data We named the program “Matrix.plot.r” to help the user associate the

nature of the graphical data. The source command makes the Matrix.plot.r available to the R console (source("Matrix.plot.r")).

For testing purposes we used actual data collected by NOAA personnel to experiment with the model.

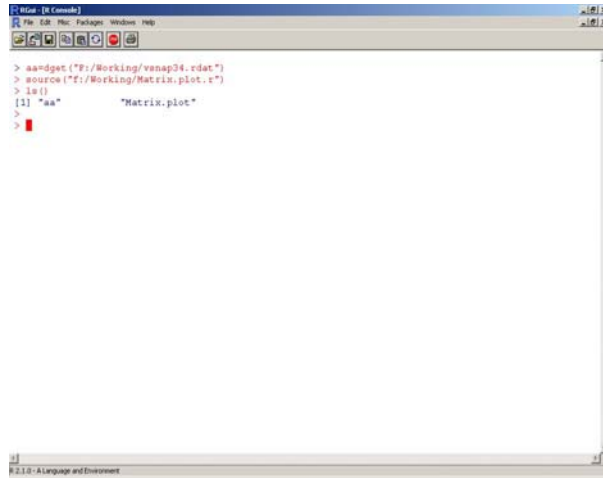


Figure 5. R console with Vsnap34.rdat and Matrix.plot imported.

Each segment has its own name to be able to differentiate from data that cannot be analyzed by a scatterplot matrix. Each data set is separated by a dollar sign, and the name of the data (\$filename). To expedite the process, the code includes a search function (grep) and it also searches for the \$.serres information (time series) and then locates all of the "F.*.*" data.

Now that the only essential data is displayed, the year column has to be removed and the F.Fmsy data. This step is very important because it requires modification without disturbing other data. Regardless of the position of the year and F.Fmsy the position as be located in the vector data and the number be extracted from the set.

The graph must have a title and it is generated within the code. In this scenario the title for the "vsnap34.rdat" data is stock assessment. The year data was removed because instead of the labels a variation of colors can be used to depict a relationship

between entries. In order to display elaborate colors the heat.colors function was used to indicate earlier years with dark colors and light colors are latter. Within consideration the program calculated the amount of data to generate the correct amount of color shades. There are numerous combinations and user definable colors such as rainbow, topo, terrain, and cm, but most importantly is whether they complement the data.

Upon execution of the graph, the user can specify whether the graph should be exported to a pdf file by adding "true" after the data expression. The fuction dev.print produces a .jpg image of the graph.

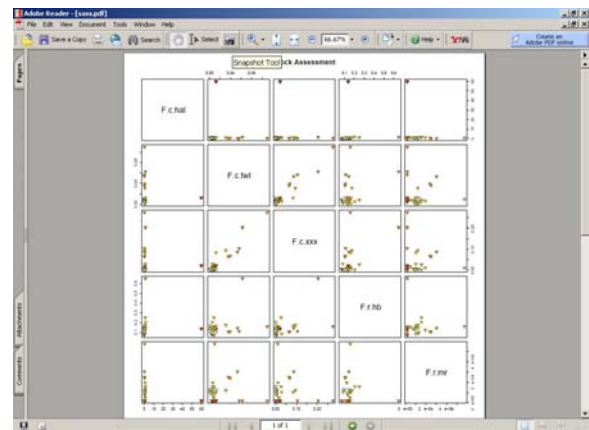
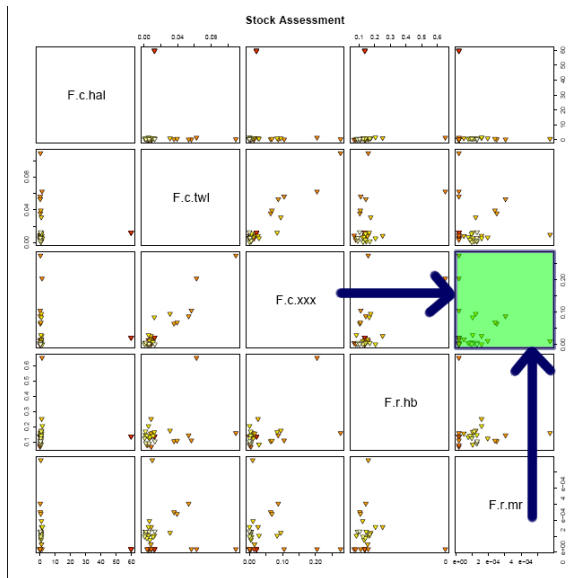


Figure . Stock Assessment Scatterplot Matrix in .PDF format.

III. Results

The scatter plot matrix is to be read by selecting two columns and the point of intersection displays the relationship between the data. The program that we have written allows any user to import the data in a matter of seconds produce professional graphs. The x-axis and y-axis scale can be found on the outmost edge of the graph.



[4]. “The Comprehensive R Archive Network”
<http://cran.r-project.org/>

[5]. “Tinn-R”
<http://www.sciviews.org/Tinn-R/>

IV. Future Research

Due to the complex learning curve we were unable to create an elaborate program that includes a great deal of flexibility. Each of the graphs are independent of each other, it would be beneficial if we were able to add a feature to allow the user to refer back to previous graphs and add overlay capabilities. To truly test the program it would also be beneficial to use the program to develop a research opportunity which relied primarily on the graph to formulate a thesis.

V. References

[1]. Maindonald, John. Braun, John. Data Analysis and Graphics Using R an Example Based Approach. 2003.

[2]. Ripley, BD. Venables, WN. Statistics and Computing “Modern Applied Statistics with S 4th Edition. 1999.

[3]. Ripley, BD. Venables, WN. Statistics and Computing Introducing Statistics with R. 2002