Arc in a Minute

Sanford Weisberg, University of Minnesota

August 24, 2005

1 Getting *Arc*

You can download *Arc* from the website www.stat.umn.edu/arc, for Windows, Linux/Unix and Macintosh. On Macintosh, *Arc* runs on System 9 or as a classic program under System X. On Windows, *Arc* includes an Uninstaller you can use to remove *Arc*; on Macintosh, simply move the *Arc* folder to the trash.

2 Data files

The data for the article by Vasilescu and Wainer can also be downloaded from www.stat.umn.edu/arc. If you plan to use these data with *Arc*, download the file vwdata.lsp, which is a text file that is in a special format for use with *Arc*.

Arc can also read a plain text file. The simplest format has: (1) *variable names* in the first row separated by blanks or tabs; (2) *data values* in succeeding rows with values separated by blanks or tabs. Data can be numbers or text. If you use text variables, avoid imbedded blanks (so Las Vegas could be LasVegas or Las-Vegas, for example); (3) *missing values* must be indicated by a missing value indicator, which is a question mark. The first 6 lines of a data file might look like this:

```
Country Crowdedness Fertility GDP
ARUBA 0.7 ? 20100
AUSTRIA 0.7 1.28 31187
AZERBAIJAN 2.1 2.1 853
BAHAMAS 1.3 2.29 14462
BELGIUM 0.6 1.66 29257
BERMUDA 0.6 1.67 51991
```

The value of Fertility is missing for Aruba, and is indicated by the ? in place of a value. Files like this one are easy to create with most statistics packages, including SAS, SPSS, Excel and others. *Arc* can also create data files from Excel files; see the "Arc add-ons" link on the *Arc* website.

You can give the file any legal name you like, such as vwdata.txt. You will find it convenient to put your data file in the *Arc* directory (c:\Arc on Windows), but you can put it anywhere.

3 Starting *Arc* and reading a file

Double-click on the *Arc* icon. The opening screen looks like this:
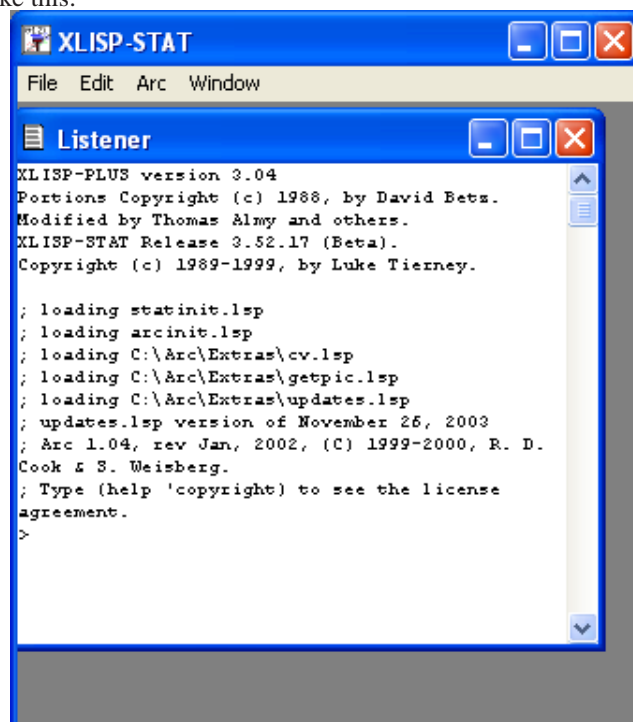
The *Listener* window shows text output. It can also be used for text input, but we won't cover that in this brief introduction. The *menubar* at the top of the screen is your entree to the program. The only non-standard menu at first is the Arc menu, but more will be created as you use the program.

To read a data file, select Arc → Load, and browse to the data file you created. If you are using Windows, you may need to select Files of type: → All files (*.*) to see the data file you saved. If you load the data file vwdata.lsp, the screen will look something like this:


The variable names and number of cases for each variable are printed in the Listener. Also, two menus are created. The first is called the Dataset menu, but its actual name depends on the name of the dataset. This name is either specified in the data file or else you will be given an opportunity to choose a name using a dialog. In the example, the name is UNdata . The second menu is always called Graph&Fit . Interactions with the program are through these two menus.

4 The Dataset and Graph&Fit menus

The Dataset menu is used to modify the data set through transformations, defining factors, and the like. For most of the items in this menu, you select the item and a dialog

is created. You fill in the dialog, and then an action is done. Most dialogs have a "Help" button, and in any case they are reasonably self-explanatory. You should be able to figure out much of the functionality of the program by simply trying the menu items.

The Graph&Fit menu is used to draw graphs and to fit models. To draw a scatterplot matrix shown in Figure 1 in the discussion of the paper, for example, select Graph&Fit → Scatterplot matrix of. . . to get the following dialog:
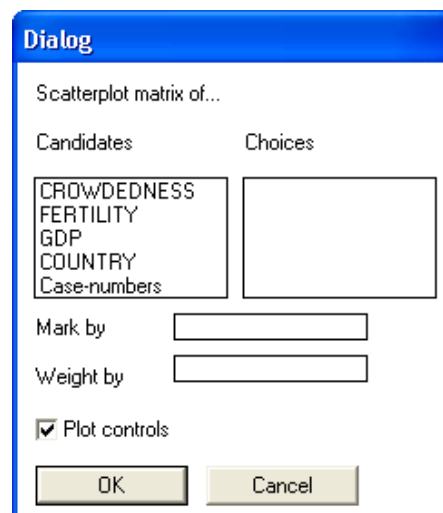
The variable names that you put in the "Choices" box will appear in the scatterplot matrix. You move items between the larger boxes by double-clicking. You move an item between any two boxes by clicking once to highlight the item you want to move, and clicking again on the destination. This pattern works in all dialogs in *Arc*. You can specify weights for use with plot controls that make use of weights, and you can specify a marking variable, which is a variable with just a few distinct values, to specify colors for the plotted points. For the example, simply double-click on the first three variable names and then on OK to get the scatterplot matrix:

2
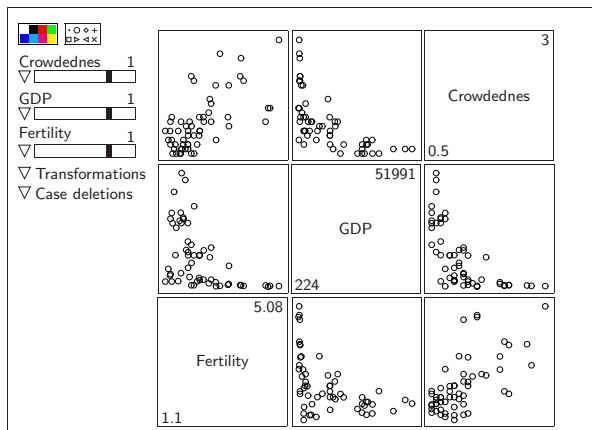
If you:

- Click on a point, it is highlighted in every frame of the scatterplot.

- Click on a point and then on a color or symbol, the color or symbol of the point is changed accordingly.

- Click or click and drag in one of the slide bars, the values for the corresponding variable are transformed. For example, moving the slider for *GDP* to 0.5 will replace GDP^1 by $GDP^{0.5}$. The value of zero corresponds to natural logarithms.

- Click and hold on one of the little down-pointing triangles to get a menu of more options. Try them out to figure out what they do! In particular the Transformations popup menu provides an automatic procedure for choosing transformations, and for saving the variables in the data set that correspond to the current transformations visible in the scatterplot matrix.

- Even more items are available in the Scatmat menu.

- Click on a graph while holding down the Control key (Option key on Macintosh) and see what happens.

- In addition, all the standard actions for a window will work, so you can grow, shrink, move or remove the window in the usual ways.

---

[1]The transformation is actually $(GDP^{05} - 1)/0.5$. This scaled transformation avoids a change in sign of correlations when the transformation parameter changes from positive to negative.

Try it!

After you find some transformations that seem to make all the frames in the scatterplot matrix approximately straight lines, you can use the Transformations popup menu on the scatterplot matrix to save the transformed variables, and then use Graph&Fit → Fit linear LS... to fit the regression model. This will give you (1) "usual" printed output; (2) a menu of options for working with the regression model; and (3) if you select Graph&Fit → Plot of..., standard summaries like residuals and Cook's distances can be graphed. *Arc* includes histograms, box plots, 2D and 3D scatterplots, and scatterplot matrices.

5 Saving and printing

You will want another program for saving printed output and graphs. The easiest method is to cut and paste from *Arc* windows to Microsoft Word, and then use Word for printing (printed output should use a fixed-width font like `Courier New`). Other ways of printing and saving are described in the reference below and on the website.

6 Quitting

Select File → Exit on Windows, or File → Quit on Macintosh.

7 More information

The best reference for *Arc* is the book by Cook and Weisberg (1999). The website `www.stat.umn.edu/arc` contains more information as well; browse and see for yourself. The remaining references below (apart from Vasilescu and Wainer, 2005) discuss the structure of *Arc* and why it is different from other statistical packages.

8 References

Cook, R. D. and Weisberg, S. (1999). *Applied Regression Including Computing and Graphics*. New York: Wiley. *This book provides the manual for* Arc.

Vasilescu, D. and Wainer, H. (2005). Old Mother Hub-
bard and the United Nations: An adventure in ex-
ploratory data analysis. *Chance*, 18(3), 38–47.

Weisberg, S. (1996). A graphical paradigm for regression.
In Fox, J. and Stine, R. (eds.), *Statistical Computing
Environments for Social Research*. Thousand Oaks:
Sage. *Discusses the design philosophy used in* Arc.

Weisberg, S. (2005). Lost opportunities: Why we
need a variety of statistical languages. *Journal of
Statistical Software*, Vol 13, Number 1, available
at http://www.jstatsoft.org. *Discusses some of the
unique features of* Arc.