



Journal of Statistical Software

July 2004, Volume 11, Software Review 1.

<http://www.jstatsoft.org/>

Reviewer: Antony Unwin
University of Augsburg

Aabel 1.5.7

Gigawiz Ltd. Co., Tulsa, OK. \$ 349 (standard), \$249 (academic).
<http://www.gigawiz.com/>

Aabel looks like a package that has been written to meet geologists' needs. It has some useful spatial tools that you would be hard put to find in other statistics packages (map projection transformations and Voronoi plots) and it has some specialist graphics that are likely to be useful for geological applications (e.g., Contour maps, Rose diagrams, Ternary plots, Harker diagrams, Vector displays, see Figure 1). The main features of the package are its data visualisation displays. Otherwise it has a few simple statistics and some limited multivariate analysis tools, which are implemented in a rather rudimentary way.

Aabel is a Macintosh-only package for OS X, but its interface reminds you more of Windows than Mac in some places. It tends to fussiness with lots of dialogs to work through, where users might hope for a smoother workflow. For example, it is quite a complicated matter to get a simple one-dimensional plot of a variable. You have to start a visualisation pipeline, select the chart-type you want and then select the variable. If you want the same plot of another variable from the same data set while keeping the first display available, you have to open another visualisation pipeline (or duplicate the current one). The idea is to allow you to combine data from many different sources, but for analyses of single data sets it is clumsy. Another example of a surprising lack of smoothness is in changing the size of a display. You don't drag it, as is common in most windows' systems, you change the scale in a control which can initially be hidden behind the list of variables available for that display. Since you change the scale of the display but not the window it is drawn in, this means that you have to then drag the window to an appropriate size to see the rescaled display. To change the axes you can move the cursor across one of them so that the cursor symbol changes. A mouse click then brings up an enormous dialog box with more options than you could ever want to use at once (Figure 2). Less is more in good interfaces and anyone hoping to make a quick change with such a system will surely be disappointed

In **Aabel** everything is done through pipelines: Data Processing, Visualization, Statistics, Multivariate Analysis. This is to enable the software to work with several different datasets at once without having to merge them. However, this is not as general as it sounds. You cannot combine data sets with the same cases, but different variables (i.e. additional information). You can only combine data sets with different cases (though they may have additional variables). You have to create a pipeline to choose the data sets you want to use for practically

everything you do, which can be very irritating, especially when only one data set is available and so there is no choice. Linking between graphics works fine and if you select points in a scatterplot or bars in a histogram, these cases are highlighted in all other displays including the data spreadsheet. Strangely, it does not work in the opposite direction. You cannot select cases in the spreadsheet and see them highlighted in the displays.

The import function offers several options (and can directly read Excel files), but stumbled when I tried to read in a file it could not read. Instead of issuing an error message, it just did nothing. To test the limits of **Aabel** I tried loading a data set with around 85000 cases and 400 variables. After several minutes the cursor was still spinning and I force-quit the programme. It seems to have been the number of variables rather than the number of cases as I was able to load a data set with a million rows with two variables and work with it relatively easily. Ideally developers should provide estimates of how much data their software can handle.

Although the interface tends to be fussy (too many options at once, too many helpful error messages), there are some interesting shortcuts. Clicking the button above a variable's column produces summary statistics and a histogram if the variable is continuous (Fig 3). That is a good idea, but is spoiled by providing too much information. Interactive systems can afford to be selective because the user can easily ask for more if they want more. There are too many statistics (how often do I want the sum and the skewness?) and superimposing both an estimated normal density and a cumulative distribution curve just confuses the display. The estimated normal density is even drawn for 0-1 variables! The histogram display can be varied interactively, which is helpful, but it is more appropriate to vary the binwidth than the number of bins. What might you expect to get by clicking the variable information button for a text variable? A barchart, a frequency table or what? For no obvious reason the software provides statistics on the length of the text entries (Fig 4)! Given that a barchart would be very helpful here, it's disappointing that it is not offered. Barcharts of text variables can be drawn in a visualisation pipeline, but they are called category histograms not barcharts.

Graphics on their own are not enough and it is good to see statistical tools included in the package, but the implementation could be better. Whereas the graphics look modern, include linking and offer great flexibility, the statistical outputs could not easily be distinguished from those of 1970s software (Fig 5 shows the output for an unpaired t test). **Aabel** is not alone in this, you can find it in some Data Mining packages as well. The fact that you can copy and paste the outputs into editing windows and then pretty them up suggests that **Aabel**'s designers recognise the problem and are presumably working to fix it. For statistical analyses you need an overview report of results and access to objects which contain details, not huge printouts in many sections with no proper summary.

The collection of methods in the Multivariate Analysis pipeline includes factor analysis, multiple regression and PLS regression. Single variable regression and some Analysis of Variance models are in the Statistics pipeline. A single cluster analysis method is available in the Graphics pipeline. Again, this lack of structure may be found in other packages, but there it is probably for legacy reasons.

Aabel offers many different kinds of graphics and many ways of adjusting and modifying them. It would be nice to be able to state that the default graphics are all excellent and that the handbook is full of instructive examples. However, that's not the case. When a visualisation pipeline is opened and a graph type chosen, the package attempts to plot the first variables in the data set. This can mean that the message No Valid Data appears (hardly encouraging) or that something silly is done, like plotting case numbers. Clustering a moderately sized

data set produced a dendrogram with totally unreadable labels (even after zooming). In the handbook there are plenty of displays illustrating what can be done, but the data sets are never explained, there is no discussion of why those graphics were chosen or of what they reveal. Many are flawed or unattractive, which is a pity, given that much better should be possible with the tools which are available in the package. (To be fair to **Aabel**, there are plenty of other handbooks with similarly disappointing graphics.)

Although **Aabel** can calculate Voronoi diagrams for sets of points and colour the resulting polygons according to another variable, it does not seem to be able to read in polygon data sets. This limits its use for geographers (and indeed statisticians). The software is able to display layered charts and one useful application would be to display background map images. Curiously, the examples in the handbook concentrate more on displaying decorative pictures and there is little advice given on how to align map images and data points.

The accompanying handbook is a hyperlinked pdf file, which is very easy to navigate. I wish all software would use this kind of approach for their documentation. The statistical references in the handbook are very old (Kendall & Stuart Vol 1 from 1969 is one of them), mostly out of date and are not referenced in the handbook text. The authors would be well advised to look at some more modern texts and other statistical software if they want to improve the statistical features in **Aabel**.

Overall I would not recommend this package for data exploration (**Data Desk** and **JMP** on the commercial side or **MANET** and **Mondrian** amongst freeware packages are much better geared to that), but **Aabel** has definite advantages in working with geological data and certainly has potential. I look forward to seeing future versions.

Reviewer:

Antony Unwin

University of Augsburg

Department of Computer Oriented Statistics and Data Analysis

Augsburg, Germany

E-mail: Antony.Unwin@Math.Uni-Augsburg.DE