

Appendix B

Software Resources for EDA

The purpose of this appendix is to provide information on internet resources for EDA. Most of these were discussed in the body of the text, but some were not. Also, most of them are for MATLAB code, but some of them are stand-alone packages.

B.1 MATLAB Programs

In this section of Appendix B, we provide websites and references to MATLAB code that can be used for EDA. Some of the code is included with the EDA Toolbox (see [Appendix E](#)). However, we recommend that users periodically look for the most recent versions.

Bagplots (included with the EDA Toolbox)

The code for the bagplot can be found at

www.agoras.ua.ac.be/Public99.htm

The authors include software for FORTRAN, S-Plus, and MATLAB. Papers are also available for download from this same website.

Computational Statistics Toolbox

This toolbox was written for the book *Computational Statistics Handbook with MATLAB*. It contains many useful functions, some of which were used in this book. It is available for download from

www.stat.unipg.it/stat/statlib/

Some of the functions from this toolbox are included with the EDA Toolbox (see [Appendix E](#)).

Data Visualization Toolbox

The authors of this MATLAB toolbox provide functions that implement the graphical techniques described in *Visualizing Data* by William Cleveland. Software, data sets, documentation, and a tutorial are available from

www.datatool.com/Dataviz_home.htm

Some of these functions are included with the EDA Toolbox (see [Appendix E](#)).

Generative Topographic Mapping

The GTM Toolbox is available for download from

www.ncrg.aston.ac.uk/GTM/

This website includes links to documentation, papers on the GTM, and data sets. This software is available for noncommercial use under the GNU license. This toolbox is included when you download the EDA Toolbox.

Hessian Eigenmaps

The home page for Hessian eigenmaps or HLLE is

<http://basis.stanford.edu/WWW/HLLE/frontdoc.htm>

This website also has code that runs the Swiss roll demo. The HLLE code is included with the EDA Toolbox.

ISOMAP

The main website for ISOMAP is found at

<http://isomap.stanford.edu/>

This also has links to related papers, data sets, convergence proofs, and supplemental figures. The ISOMAP functions are in the EDA Toolbox.

Locally Linear Embedding – LLE

The main website for LLE is

www.cs.toronto.edu/~roweis/lle/

From here you can download the MATLAB code, papers, and data sets. The LLE function is part of the EDA Toolbox.

MATLAB Central

You can find lots of user-contributed code at this website:

www.mathworks.com/matlabcentral/

Always look here first before you write your own functions. What you need might be here already.

Microarray Analysis - MatArray Toolbox

The following link takes you to a webpage where you can download a toolbox that implements techniques for the analysis of microarray data. This toolbox includes functions for k -means, hierarchical clustering, and others.

www.ulb.ac.be/medecine/iribhm/microarray/toolbox/

Model-Based Clustering Toolbox

The code from the Model-Based Clustering Toolbox is included with the EDA Toolbox. However, for those who are interested in more information on MBC, please see

www.stat.washington.edu/mclust/

This website also contains links for model-based clustering functions that will work in S-Plus and R.

Robust Analysis

A toolbox for robust statistical analysis is available for download at

<http://www.wis.kuleuven.ac.be/stat/robust/LIBRA.html>

The toolbox has functions for univariate location and scale, multivariate location and covariance, regression, PCA, principal component regression, partial least squares and robust classification. Graphical tools are also included for model checking and outlier detection.

Self-Organizing Map

The website for the SOM Toolbox is

www.cis.hut.fi/projects/somtoolbox/links/

This website also contains links to documentation, theory, research, and related information. The software may be used for noncommercial purposes and is governed by the terms of the GNU General Public License. Some functions in the SOM Toolbox are contributed by users and might be governed by their own copyright. Some of the functions in the SOM Toolbox are included in the EDA Toolbox.

SiZer

This software allows one to explore data through smoothing. It contains various functions and GUIs for MATLAB. It helps answer questions about

what features are really there or what might be just noise. In particular, it provides information about smooths at various levels of the smoothing parameter. For code, see:

www.stat.unc.edu/faculty/marron/marron_software.html

For notes on a short course in SiZer and how it can be used in EDA, go to:

www.galaxy.gmu.edu/interface/I02/I2002Proceedings/...MarronSteve/MarronSiZerShortCourse.pdf

Statistics Toolboxes - No Cost

The following are links to statistics toolboxes:

www.statsci.org/matlab/statbox.html

www.maths.lth.se/matstat/stibox/

Here is a link to a toolbox on statistical pattern recognition:

cmp.felk.cvut.cz/~xfrancv/stprtool/index.html

Text to Matrix Generator (TMG) Toolbox

The TMG Toolbox creates new term-document matrices from documents. It also updates existing term-document matrices. It includes various term-weighting methods, normalization, and stemming. The website to download this toolbox is

scgroup.hpclab.ceid.upatras.gr/scgroup/Projects/TMG/

B.2 Other Programs for EDA

The following non-MATLAB programs are available for download (at no charge).

Crystal Vision

CrystalVision, copyright (c) 2000 by Crystal Data Technologies (Qiang Luo, Edward J. Wegman, and Xiaodong Fu), is a Windows 95/98/NT package for Wintel computers. A demonstration version of CrystalVision is available at

<ftp://www.galaxy.gmu.edu/pub/software/CrystalVisionDemo.exe>

This software includes parallel coordinates, brushing, linking, plot matrices, and others.

Finite Mixtures - EMMIX

A FORTRAN program for fitting finite mixtures with normal and t-distribution components can be downloaded from

<http://www.maths.uq.edu.au/~gjm/emmix/emmix.html>

GGobi

GGobi is a data visualization system for viewing high-dimensional data, and it includes brushing, linking, plot matrices, multi-dimensional scaling and many other capabilities. The home page for GGobi is

www.ggobi.org/

Versions are available for Windows and Linux.

MANET

For Macintosh users, one can use MANET (missings are now equally treated) found at:

www1.math.uni-augsburg.de/MANET/

MANET provides various graphical tools that are designed for studying multivariate features.

Model-Based Clustering

Software for model-based clustering can be downloaded at

<http://www.stat.washington.edu/mclust/>

Versions are available for S-Plus and R. One can also obtain some of the papers and technical reports at this website, in addition to joining a mailing list.

R for Statistical Computing

R is a language and environment for statistical computing. It is distributed as free software under the GNU license. It operates under Windows, UNIX systems (such as Linux), and the Macintosh. One of the benefits of the R language is the availability of many user-contributed packages. Please see the following website for more information.

www.r-project.org/

B.3 EDA Toolbox

The Exploratory Data Analysis Toolbox is available for download from two sites:

<http://lib.stat.cmu.edu>

and

<http://www.infinityassociates.com>

Please review the **readme** file for installation instructions and information on any changes.

The functions in the EDA Toolbox are available via a graphical user interface called **eda**, as well as the command line. The EDA Toolbox will be available when this book is published, but the **eda** GUI will not be available in the toolbox until February, 2005.

Most of the functions and capabilities in the toolbox have been discussed in this text, but several of them have not. Please see [Appendix E](#) for a list of functions that are included in the EDA Toolbox, along with a brief description of what they do. To view a *current* list of available functions (i.e., including those written after this book was published), type **help eda** at the command line. For complete functionality, you must have the Statistics Toolbox, version 4 or higher. For more information on the use of the EDA Toolbox, please see the accompanying documentation.