

# Programming Languages and Statistics: New Possibilities for Statistical Computing

## 1. Introduction

The development of statistical software is an exciting area. Many statisticians around the world are using statistical packages to reach at conclusions everyday. Developing useful and modern statistical packages is thus crucial. Programming languages play a key role in developing statistical applications. In next section, several commonly used programming languages for statistical software will be briefly overviewed. Section 3 introduces a user-friendly software which incorporates certain modern computer technologies. Some advanced facilities for statistical computing, such as mobile development and web service deployment are discussed in Section 4, while a concluding discussion is given in Section 5.

## 2. Programming Languages for Statistical Computing

Fortran, C, Lisp and Java are commonly used programming languages for developing statistical applications. Fortran and C are imperative programming languages, Lisp is a functional language and Java is an object-oriented language (see Tucker and Noonan (2007)). Fortran (Formula Translator), the first scientific programming language, might be used most widely. There are a variety of numerical libraries written in Fortran. That is one main reason for its popularity. Fortran plays a key role in the developing process of S system. As indicated by Becker (1994), the inventors of S system initially wanted an interactive interface to the algorithms in the SCS (Statistical Computing Subroutines) library, which is a large, well-documented Fortran library. Fortran is still evolving and Fortran 04 is the very new version.

Many operating system components are written in C, which is highly efficient and the programmers can use to access to the machine language level. R, the widely used statistical software, was written in ANSI standard C. Although S and R have very similar syntax, R usually provides better performance in term of timing (see Ihaka and Gentleman (1996)).

Functional programming, motivated by the researchers in artificial intelligence, symbolic computation, theorem proving, rule-based systems, and natural language processing, emerged in the early 1960s (see Tucker and Noonan (2007), chapter 14). Lisp is one of functional programming languages. XLISP-STAT, based on a family of LISP implementations XLISP, is an environment for dynamic graphics and statistics with objects. Although XLISP-STAT has not been as frequently used as before, the dynamic graphics facilities in R are not as seamless as in XLISP-STAT (see Narasimhan (2005)). However, many statisticians who used XLISP-STAT before now choose R/S systems as teaching or research tools.

Object-oriented programming is an old paradigm, at least dated back to Simula in the 1960s. Nevertheless, Java definitely plays a crucial role in making object-oriented programming very popular to the programmers. Java provides a great deal of advanced facility. Furthermore, modern software technologies, such as frameworks and design patterns, have been extensively used in Java APIs. That is one reason why Java is so structured and succinct. A few Java applications for statistics, including Jasp, JStatCom and JStatSoft have been implemented. Jasp, a powerful statistical system, utilizes a GUI and a CUI (Character User Interface) seamlessly (Yamamoto et al. (2002)) and provides a procedural and object-oriented statistical language based on Pnnts (Kobayashi et al. (2002)). JStatCom (Kratz (2007)), based on software engineering strategies, provides a framework for the development of rich desktop clients for data analysis in a rather general way. JavaStatSoft (see Wei and Chen (2008)), a user-friendly software for data analysis, has been implemented based on several software reuse techniques, including design patterns and frameworks.

## 3. JavaStatSoft Using Design Patterns

As developing a modern software with numerous codes, software reuse is a commonly used software engineering strategy in the design process. One of the techniques to support software reuse is design patterns, which was popularized by GoF (the Gang of Four, see Gamma et al. (1995)) and is now a common discussion topic in software development teams around the world. As indicated by GoF (1995, p. 360): "A design pattern systematically names, motivates, and explains a general design that addresses a recurring design problem in object-oriented systems. It describes the problem, the solution, when to apply the solution, and its consequences. It also gives implementation hints and examples. The solution is a general arrangement of objects and classes that solve the problem. The solution is customized and implemented to solve the problem in a particular context". A good design pattern not only helps the software developers organize or polish the codes but also provides the other programmers a blueprint for examining codes, analyzing application development, and communicating software design. Several design patterns can be used by a framework, which is another technique for software reuse. Based on a well-designed framework, the developers can build their software with similar structures efficiently and effectively.

Wen Hsiang Wei

Department of Statistics, Tung Hai University, Taiwan

E-mail: [wenwei@thu.edu.tw](mailto:wenwei@thu.edu.tw)

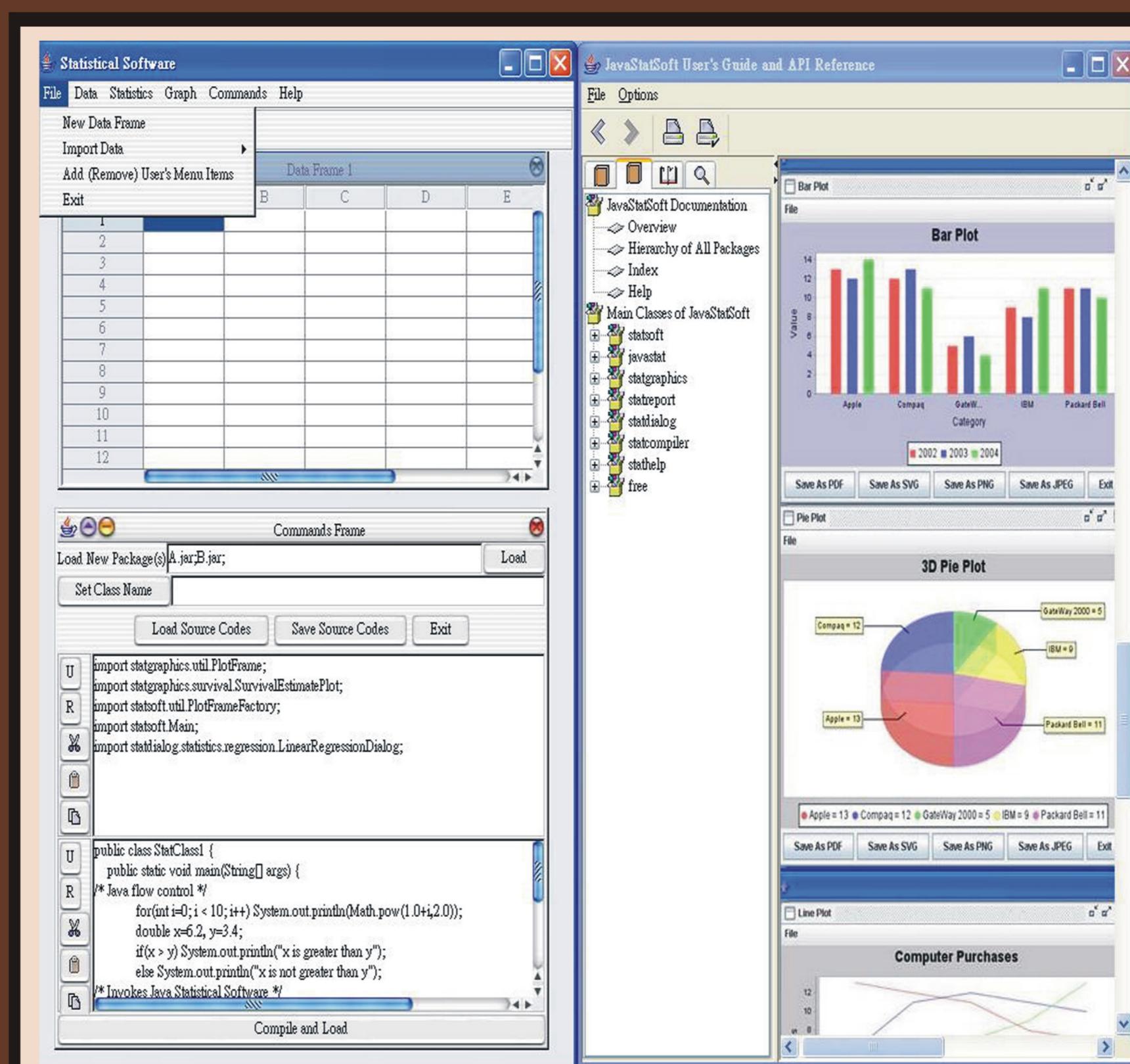


Figure 1:JavaStatSoft: Look and Feel

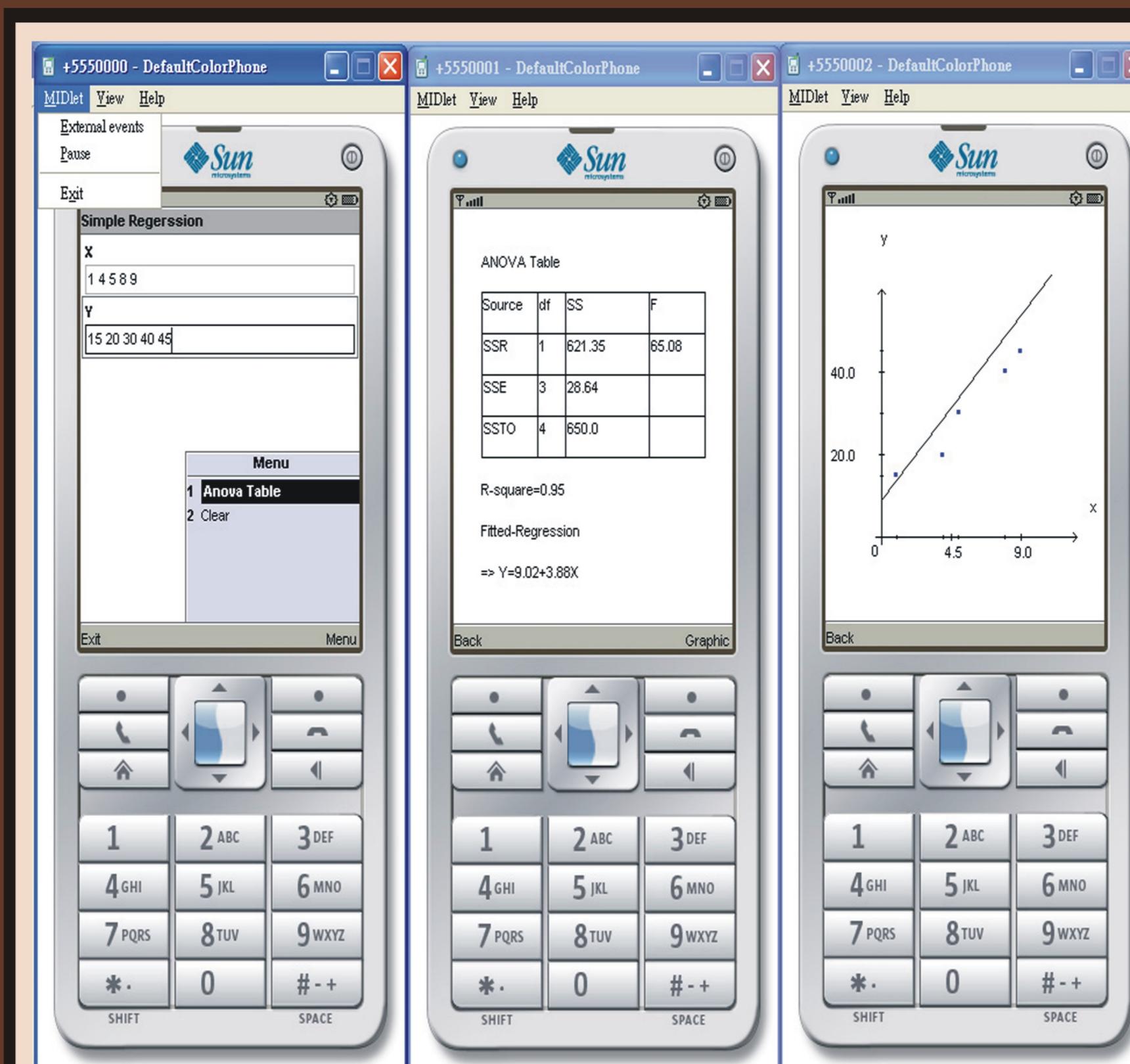


Figure 2:Running A Simple Regression Application

JavaStatSoft, implemented using Java and designed based on a proposed framework (see Wei and Chen (2008)), referred to as Data Analysis Module, is a software with a GUI (Graphical User Interface). The aim of the software is to offer a freely accessible alternative to commercial statistical software. For students or people used to spreadsheet-like environment, JavaStatSoft allows them to carry out statistical analysis via menu selections. For programmers, JavaStatSoft can compile Java code created by them and also incorporates with BeanShell (<http://www.beanshell.org/>), a free and embeddable Java source interpreter with object scripting language features, for running scripting language. In addition, the users can customize the GUI of JavaStatSoft and extend the functionalities of JavaStatSoft. The software and its user's guide can be downloaded from the web site

<http://www2.thu.edu.tw/~wenwei/>

The software can be invoked by clicking on the Windows executable jar (Java Archive) file `javastatsoft_beta1.2.jar`. The look and feel of the software is given in Figure 1.

## 4. Mobile Development and Web Service Deployment

Cell phone, used booming worldwide, currently supports many additional services and accessories. However, there is no statistical application for cell phones. Since the third generation is coming and the multi-function cell phones are the trend in the future, developing a user-friendly statistical application is thus crucial.

Java is also a powerful programming language to develop statistical applications for cell phone. The Java ME (Java Micro Edition) is the platform that the programmers can use to develop a variety of applications for mobile devices. The aim of the Java ME is to develop the intelligent wireless devices and small computing devices that can provide products to incorporate cross-platform functionality. These devices that have limited processing power, storage capabilities, and intermittent or fairly low-bandwidth network connections, include cell phones, pagers, wireless devices and set-top boxes among others.

Some simple applications based on Java ME have been developed by Liang(2008). For example, the users can input the required data for a simple regression analysis, and then the ANOVA table and fitted regression line can be generated, as presented in Figure 2.

Web services, a new distributed computing paradigm, can perform almost any kind of task. The applications all over the world can communicate with each other and share data via web services. Web services based on different software programs can exchange information via common internet protocols and XML (Extensible Markup Language). Little work has been done for statistical web services. NISS (National Institute of Statistical Sciences) has developed a web service for data swapping (see Sanil et al. (2002)). Wang (2008) has developed web services using Java for statistical inference on the population mean and proportion.

## 5. Concluding Discussions

Java is very powerful and has a great deal of functionality. From mobile phone, desktop, to distributed computing environment, Java can help the statisticians develop a variety of statistical applications. Further, the other advanced facilities provided by Java, such as multithreading support, networking facilities, database connectivity, internationalization, the control of input and output of multi-media, and Enterprise JavaBeans development, might provide a very strong foundation on which a powerful statistical software might be built to meet the needs for the statisticians, the students and the people of other disciplines.

## References

Becker, R. A. (1994) A brief history of S. Technical report.

Gamma, E., Helm, R., Johnson, R., Vlissides, J. (1995). *Design patterns: elements of reusable object-oriented software*. Reading, MA: Addison Wesley.

Kobayashi, D., Fujiwara, T., Nakano, J., and Yamamoto, Y. (2002). A procedural and object-oriented statistical scripting language. *Computational Statistics*, 17, 395-410.

Kratz, M. (2007). A software framework for data analysis. *Computational Statistics & Data Analysis*, 52, 618--634.

Ihaka, R. and Gentleman, R. (1996). R: a language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, 5, 299--314.

Liang, F. C. (2008). *Java micro edition technology for statistical and graphical analysis*. Master thesis, Department of Statistics, Tung Hai University.

Narasimhan, B. (2004). Lisp-Stat to Java to R. *Journal of Statistical Software*, <http://www.jstatsoft.org/v13/i04>.

Sanil A., Gomatam, S. and Karr, A. F. (2003). NISS WebSwap: a web service for data swapping. *Journal of Statistical Software*, <http://www.jstatsoft.org/v08/i07>.

Tucker, A. B. and Noonan, R. E. (2007). *Programming languages: principles and paradigms*. McGraw Hill.

Wang, S. H. (2008). *Web services for statistical analysis*. Master thesis, Department of Statistics, Tung Hai University.

Wei, W. H. and Chen, G. J. (2008). JavaStatSoft: design patterns and features. *Computational Statistics*, 52(2): 235-251.

Yamamoto, Y., Nakano, J., Fujiwara, T., and Kobayashi, D. (2002). A mixed user interface for a statistical system. *Computational Statistics*, 17, 379--393.