

Using the Internet to Teach Statistics and the Statistical Environment R

Vieira, Vilson

Santa Catarina State University, Department of Computer Science

Campus Prof. Avelino Marcante s/n – Bom Retiro

Joinville 89223-100, Brazil

E-mail: vilson@void.cc

Henning, Elisa

Santa Catarina State University, Department of Mathematics

Campus Prof. Avelino Marcante s/n – Bom Retiro

Joinville 89223-100, Brazil

E-mail: elisa@joinville.udesc.br

Introduction

Statistics, like any Mathematical science, can make an erroneous hard feeling for students. Specially if these students come from different areas and levels of knowledgement, or even come from different courses. It is the case of the Statistics course at the Santa Catarina State University¹, from Brazil. Students of this course come from three different areas of study: Engineering, Information Technology and Computer Science. For these areas, the mathematics and computing knowledge level vary. Engineering students have a high level of Mathematics knowledge, but just a few understanding of computer programming languages and no previous usage of a statistical computing environment. At the other hand, Information Technology and Computer Science students have a mid level of Mathematics but a high capacity to use statistical computing environment. But just the Computer Science students knows about the use of computer programming languages. Other necessity exists too, like the few numbers of computer laboratories, no resources to buy expensive software licenses from commercial statistical environments and the necessity of more documentation about the statistical computing environment chosen in portuguese language.

How to solve these necessities ? How to teaches Statistics in a practical way, using a statistical computing software, with just a few resources ? The answer is Internet. A project have been created to make documentation about Statistics and the chosen statistical computing software. And all the documentation is on-line, at a public Web site² that every student around Brazil – with a vary level of knowledge of Computation and Mathematics – can use.

This poster wants to describe the actual situation of the project called “Divulgando o Ambiente R” (or Divulgating the R Environment), the used technology and education methodology. The future spectatives will be discussed, aiming to set this project as an example of success case on using Internet to propagate educational content.

Technology Helps

A statistical computing environment is a core requisite to Statistics, specially to its education. Actually there are popular proprietary softwares (SAS/STAT, STATA, MINITAB, 2007), but with expensive cost licenses, difficulting their usage at public education institutions. The alternative is to use a free software. The choice of the project was to use the most powerful one, the R (R PROJECT, 2007).

¹<http://www.joinville.udesc.br>

²<http://r.void.cc>

R is a free software environment for statistical computing and a complete and powerful computer programming language. As a statistical software R can do complex calculations, data handling, storage and analysis, linear and non-linear modelling, classification, statistical tests, and can generate high quality graphics as results of its computations. R is not just a statistical environment, it is a true computer programming language, giving to the user the possibility of extends R using itself. Because of this, R has a large number of packages that extends its functionality, creating from extra statistical techniques, to uncommon features to a statistical software, like the possibility to interface with external programming languages – like C or Fortran – or to create complete software applications even with a Graphical User Interface or database access.

These characteristics of R made it a complex software that need some documentation. The R official web site has the most documentation needed and it is completed by external ones. The problem is the need of documentation in portuguese, specially advanced documentation about the use of R as a programming language, a feature that motivates and gets the interest of students, mainly the Information Technology and Computer Science ones. As discussed, a project to write this kind of documentation was needed. And to publish the content was chosen a collaborative web site management: a Wiki.

The use of a Wiki makes the writing of content easy. The philosophy behind a Wiki is “let the users write and collaborate together” (CUNNINGHAM & LEUF, 2001), so the content was written fast and in a right way, with minor errors, because the users of Wiki can edit the content of everyone at everywhere and on everytime. Again, a free software Wiki system was used, called PmWiki³, and one more time the project was save of software licenses.

The free of costs and powerful technologies used in the project motivates also the students to use it on their own academics and professional initiatives. And now the project can have the help of these same students or any other person because the use of Internet and a collaborative content management system turns it possible.

Education Methodology Changes

The adoption of powerful technologies and Internet changes even the way to teach Statistics. R makes possible to show students how to use the Statistics Theory in a practical way. With the high quality graphics, students can see the results of their statistical analysis. And using R as a programming language, they can even create complex computation routines to do the work. For non Computer Science students, R can be also an introduction to computer programming, motivating the students to learn more about computation and applied Statistics as well.

Because R is a free software, all the student can download and install it on their own computer, without paying software licenses per copy. Using Internet to create and publish the documentation about R and Statistics, students can learn to use the software even at their houses.

The Results

By now the project is accomplishing its objectives. The content of the web site is at constantly production. Because the small number of references about R in portuguese language, the project is turning on a good source of information, specially in how to use R as a programming language. The use of Internet to publish the Statistics course and R content turn possible to eliminate the need to use of computer laboratories. Now students can do their research everywhere, even at home.

The use of R on Statistics course motivates the students to learn it and to apply Statistics Theory in their own projects. With R the students feels that Statistics can be applied at any area of

³<http://www.pmwiki.org>

study, from Engineering to Computer Science or Information Technology.

The Future

The project has some spectatives. One of these is to create a brazilian community of R users and developers. The objective of this community will be to translate R and R packages, extends R developing new packages or fixing the current ones, support brazilian projects that use R, and by other things, create a communication channel for the brazilian people that uses and wants to help the development of R.

At the Statistical course of the institution, the use of Rpad – a R package that makes possible to share a R session of every computer on a network (RPAD, 2007) – to create on-line classes is at study. If it can make possible, the practical classes will be made also at Internet, with teacher communicating with students by a chat room, and showing to them the R commands and its results using Rpad.

The use of R and the collaborative opportunity of the Internet extends the domain of a simple Statistical course to a national source of information, and maybe to a community of users and a e-learning initiative.

REFERENCES

R PROJECT. The R Project for Statistical Computing. <http://www.r-project.org> (04/09/2007).

SAS/STAT. SAS/STAT Software. <http://www.sas.com/technologies/analytics/statistics/stat/index.html> (04/09/2007).

STATA. Statistical Software for Professionals. <http://www.stata.com> (04/09/2007).

MINITAB. Minitab Brasil Statistical Software. <http://www.minitabbrasil.com.br> (04/09/2007).

CUNNINGHAM, Ward and LEUF, Bo. The Wiki Way. Quick Collaboration on the Web. Addison-Wesley, 2001.

RPAD. Rpad Interactive, Web-based Analysis Program. <http://www.rpad.org> (04/09/2007)

ABSTRACT

R is a free software environment for statistical computing. It is also a powerful programming language and produces professional made statistical graphics. A software with these features – a powerful professional tool with a free license – is demonstrating some advantages to the education of undergraduate students of a engineering and computer science university from Brazil. Because the quality of this software, the students being motivated to use it on their own academic and business works. This kind of interest created a demand for more information about R and Statistics itself, specially written in brazilian portuguese idiom.

To help this necessity, an academic extension project has been created and called “Divulgando o Ambiente R” (or Divulging the R Environment). Today, the project has a Web site⁴ with many content about R and the use of R for statistical applications. The content is from the basic R installation tutorial to the advanced documents about the use of R as a programming language, and its extension using other programming languages as C. The purpose of this poster is to show the education methodology used with the creation of this project and to analyze its results at this institution and out of that, with discussions about the used technology as well.

⁴<http://r.void.cc>