Urban Legends and SAS
Stanley Fogleman, Harvard Clinical Research Institute, Boston, MA

ABSTRACT
This paper presents a variety of well-known misconceptions about SAS®. Sometimes the most difficult part of implementing a new project is overcoming these preconceived notions. The paper will draw on submissions solicited from SAS-L and personal experience. The aim of the paper is to be fun and thought-provoking.

INTRODUCTION
Part of what you sign up for when you become a SAS programmer is to overcome people’s anxieties, fears and doubts about the language. The less informed and higher up in the organization hierarchy one is, the more difficult it is to overcome this. Note that it is impossible in a paper of this length to present and debunk every possible misconception about SAS, so I will limit my paper to what I perceive to be the most common.

SAS IS USED MAINLY FOR STATISTICS
This myth is the granddaddy of them all which has some historic basis. SAS was originally a series of programs(procedures) written (starting in approximately 1970) that could be used to perform common statistical calculations. “SAS then consisted of the basic data step and language, SAS datasets, and procedures. It is interesting that the basic design has held through from then until at least now. There were probably at most 8-10 procedures, including the ones that eventually became GLM and PLOT.”1 Of course, there are many other applications for SAS (such as data transformation, business intelligence, data mining, etc.) that do not rely exclusively on its statistical capabilities. I used SAS for over 10 years without touching a statistical proc (used FREQ and SUMMARY for counting and summarizing).

SAS IS A MAINFRAME PACKAGE2
Again, it might have started out that way in the seventies, when it was a combination of assembler, PL/I and FORTRAN, but since being rewritten in C for version 6, it has been ported to a variety of platforms too numerous to list here. Suffice to say that SAS runs on almost every OS except Apple’s OS/X. While we’re at it, there is a corollary that SAS code written on one platform has to be drastically rewritten to run on another. One of the beauties about SAS is that (with the exception of filename and libname statements), the code is pretty much platform independent, unless you spend an inordinate amount of your time doing machine-dependent things like bit-testing.

SAS CAN OVERCOME (FILL IN BLANK) SECURITY
I’ve heard this howler from DBA’s, network administration and helpdesk personnel. While it is certainly the case that you can enhance security by making use of the metadata library in version 9 and the more primitive approach of password-protecting data sets in versions before that, but in no instance have I been able to defeat operating system or database security solely by using SAS. It should go without saying that any respectable security system would focus on authentication and without those user and password couplets, you wouldn’t get very far. In the case of DBA’s it usually just involves a calm conversation about assigning a user and password with read-only capabilities to sensitive databases.

SAS IS TOO SLOW FOR WHAT WE’RE TRYING TO ACCOMPLISH
Funny, because some of the most time-sensitive applications I have worked on (updating a pricing system after market close, updating a call center system after a nightly database load) were exclusively in SAS and I don’t remember anyone suggesting that it be rewritten in some other language. Part of the appeal of the language to me is the ability to be concise and to only have to spend time on the problem you are working on and a minimal amount of time on housekeeping-type chores. It is certainly true that, like any other programming language, you can write inefficiently and solve programs in a roundabout manner which will produce less than optimal results. This is the fault of the programmer, not the language. Paul Dorfmann, in a posting on SAS-L ran SAS against COBOL in a parallel test. “As David has mentioned, I have conducted an experimentum crucis to actually measure SAS and Cobol performance, both I/O and CPU bound. Specifically, I made SAS 8.1 and IBM Cobol (with optimized compiler options) read and write a flat file with 10 million records. As I recall, SAS was about 10% faster at input, Cobol - about 5% faster on the output. In other words, about equal. Then I made both loop 1 billion...
times, each time incrementing an index. SAS, with its 'universal' double-float, was much faster than Cobol if the index was declared as full-word binary, and slightly faster with the half-word (S9(4) comp.).

SAS IS ON ITS' WAY OUT. NO ONE WILL BE USING IT IN 5 YEARS.
I remember some search representative advising me to stay away from SAS because it was a niche market. Worst advice I've never taken. One should seriously question the source of the reading materials of the person making this claim.

SAS IS GOOD IN DATA MINING, BUT THEY'RE NOT A BI COMPANY.
This was actually in an article quoting an Oracle executive. Evidently their subscription to sas.com had expired! One of the major focuses of SAS is to offer a variety of business intelligence solutions. Perhaps this statement might have made some sense 10 years ago, but certainly not today, but then business intelligence was a very poorly defined term 10 years ago, so I'm not certain what that would have meant.

SAS IS EXPENSIVE
This is very difficult to quantify and/or defend against. Case one: You use one copy of SAS at home and write three programs a year. (answer: True). Case two: Your employer has a site license and you use it for analysis, reporting, etc. (answer: Almost certainly false). Anecdotal evidence: At one company where I was consulting, there was a query from the FDA which, of course, needed to be answered immediately. My recollection is that three programmers were assigned to the task – the object being who could come up with the quickest verifiable answer. My recollection is that one programmer was using PL/SQL, I used SAS (of course!), and the third program wrote queries using Microsoft Access. Surprise: I came up with the answer comfortably ahead, and the rest of the time was spent trying to verify the result using one of the other platforms (what a headache!). Perhaps the most sensible approach to this is to say that the reduced development time can and should offset the cost of the software.

USING SAS TO AUTOMATICALLY GENERATE MONTHLY AND WEEKLY REPORTS IS COMPLICATED
Again, very difficult to quantify and/or defend against. I can only say that I have used SAS for reporting on and off for 15 years, and I don't understand why this would be an issue. Speculatively, I can only guess that someone was tempted to write a “super_program” that does extraction, transformation and reporting all at the same time, and I can only comment that the time taken to modularize the solution (write one program for each step) will pay handsome dividends when maintenance is required, as it almost certainly will be. Certainly if one has Data Warehousing available and one takes the time to develop a reporting database, and leave the actual reporting to Enterprise Guide (which can be turned into a Stored Process) and utilize the built-in scheduler provided with Data Integration Studio, you have a much better solution.

WE DON’T COMPETE WITH SAS, THEY’RE NOT BIG INTO DATA WAREHOUSING
Admittedly, this quote is a few years old, but that wouldn’t prevent someone from repeating it! Having just spent over a year installing, testing and validating Data Integration Studio(formerly ETL Studio), I can assure you that this is emphatically not true. I'm not sure of the source of this, other than the belief that a data warehouse consisting solely of SAS datasets wouldn’t be very useful. But since DIS can accept data from virtually any source and target any commercial database in addition to trusty SAS datasets, I don’t find that very useful.

YOU HAVE TO BE A PROGRAMMER TO USE SAS
I have had occasion to work with user groups who had limited computer expertise, and (even with the old program editor) had them using WHERE clauses in no time. Just using VIEWTABLE, they were able to write fairly sophisticated queries on existing SAS datasets in very little time. With the advent of Enterprise Guide, the process should be even easier. Of course, it certainly helps to be fluent in SAS to produce complex reports, but that
shouldn’t be a hindrance to end users running queries on prepared data. The more sophisticated the available Warehouse is, the better.

SAS DOESN’T DO VERY GOOD GRAPHS
In a word: stupefying. Again, someone spending 15 minutes or less with the on-line Enterprise Guide tutorial would discover that this claim is without merit. Certainly, becoming a SAS/GRAPH guru is its own area of expertise, and one could spend many hours mastering PROC ANNOTATE, but the quality of graphs is a non-issue. Of course, there are those people who insist that certain colors be used in certain ways for certain graphs, and those will always require manual tweaking.

CONCLUSIONS
It is useful when dealing with a non-SAS audience to be aware of these common prejudices and misconceptions. I am hopeful that in shedding some light on this subject that they can be more quickly overcome. I’m surprised by the way, in how often the views are held by other IT professionals who aren’t accustomed to SAS, and that they are often the most difficult to convince otherwise.

REFERENCES
1) From URL: http://groups.google.com/group/comp.soft-sys.sas. Post from William Viergiver on Oct 19, 1999
3) From URL: http://groups.google.com/group/comp.soft-sys.sas. Post from Paul Dorfmann on Nov 14, 2001
4) Manila Bulletin Online. May 01, 2006. Oracle stepping up assault to promote business intelligence. Taken from URL: http://www.mb.com.ph/issues/2006/05/01/INFO2006050162786.html

ACKNOWLEDGMENTS
SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.


Ted Conway for publicizing my post on “The Unofficial SAS Weblog” and making me a temporary hero in the eyes of my teenage son.

CONTACT INFORMATION
Your comments and questions are valued and encouraged. Contact the author at:

Stanley Fogleman
Harvard Clinical Research Institute
930 Commonwealth Ave West
Boston MA 02215
Web: www.hcri.edu

* * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *