

Can SAS® Programmers Stay Cool?

Jiangtang Hu, d-Wise

ABSTRACT

Yes.

INTRODUCTION

The traffic of SAS® mailing list is keeping decreasing. There are less active SAS bloggers, including myself. It seems everyone is talking about the coolness of R, Python, machine learning or deep learning. The other side is, SAS as a software and a programming language is getting better. As a long time and continuous SAS heavy user, I'm asking: can a SAS programmer stay cool? Furthermore, can a SAS programmer like me cool? It's not a silly question. Lots of SAS programmers feel the crunch.

In this paper, I will talk about the challenges a SAS programmer can face. The challenges can come from external competitors like R and Python, technology trend like machine learning and deep learning. I will talk about in which SAS is good at, while in which not. In the end, I will explore some how-to for a SAS programmer to stay alert, marketable, and even more, cool. I will also encourage SAS programmers to keep update with SAS software; Lua, Python, machine learning and other cool stuff are already imbedded in SAS system!

All codes and supporting materials will be available in my Github page:

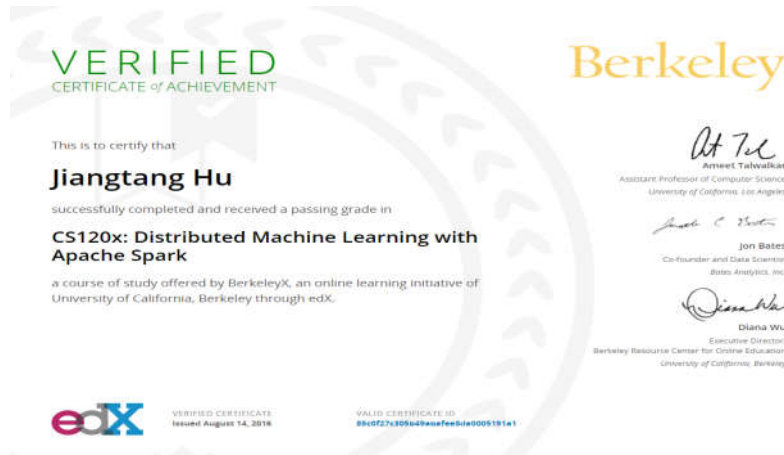
<https://github.com/Jiangtang/SESUG/tree/master/2017>

A. WHERE IS SAS IN THE BUBBLE WORLD OF BIG DATA, DATA SCIENCE?

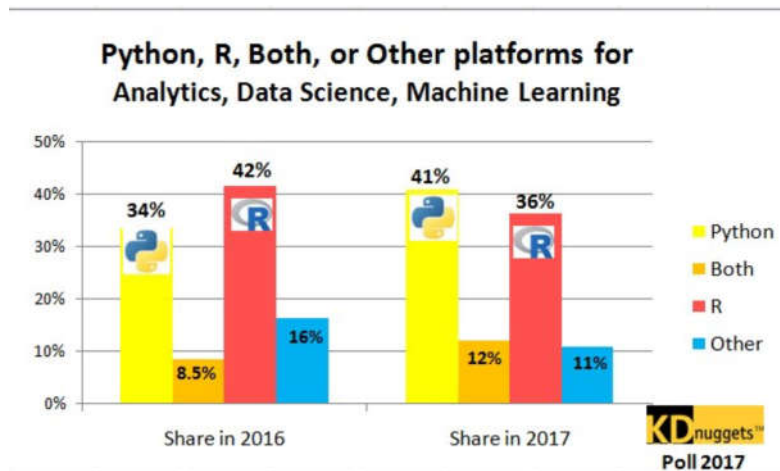
2016 was a year of struggle for me as a SAS programmer. I earn my bread mostly using SAS, I watch SAS blogs, mailing list and all things SAS frequently. But is a SAS programmer like me cool? So, I basically left SAS blogosphere for one year to explore the world beyond SAS.



In the past year, I learned some Python, machine learning, deep learning, Spark:



My impression was: SAS was absent in all these cool things cool kids talk about. In a latest Kdnuggets poll [1], SAS is even categorized in “other”:



B. SAS RECONSIDERED

A.1 SAS TABLES AND DATA STEPS

This is the gem of SAS programming language. One of the reasons that Python got its popularity in data analysis was the advent of Pandas, a DataFrame based library [2] in 2008. A DataFrame is basically a 2-dimensional table, and SAS programmers work with such kind of table since 1970s.

Pandas offers some advanced features over 2-dimensional table in which SAS lacks, but SAS is still cool compared. Suppose we need to create a new column by adding two columns up, in Pandas:

```
df["NewCol"] = df["Col1"] + df["Col2"]
```

while a typical SAS programmer will proudly do:

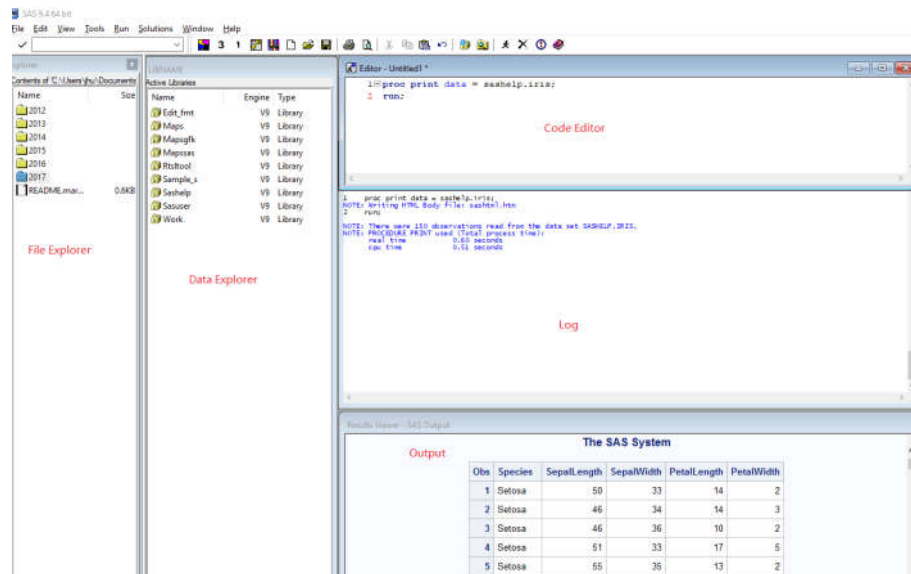
```
NewCol = Col1 + Col2
```

SAS programmers enjoy such advanced (yes it is) operations for generations. Before Pandas, this kind of simple manipulation need to take lines of loops.

A.2 SAS IDE

The traditional SAS IDE, namely SAS Display Manager (DM) faces lots of criticism among years. SAS company even discourage its usage in favor of SAS Enterprise Guide and/or SAS Studio. But considering a IDE in which

1. A code editor for typing your codes
2. A log window to check logs
3. A output window to check outputs
4. A data explorer to check datasets
5. A file explorer to check system files



This kind of IDE is what A python data scientist/analyst's dreaming IDE; they would call it the Data Science IDE and they finally got one in 2016 [3]. The interactive environment of SAS DM accelerates SAS programmer's productivity for long time. The irony is, when other programming languages try to catch up, SAS DM itself has a gloomy fate.

A.3 SAS MACROS

I'm a heavy macro user. Because of lack of advanced data structures, SAS macro is the only way to boost data driven or dynamic programming. It fits well on its purpose. When looking at Python, I'd say SAS macro is very isolated and outdated programing technique.

A.4 SAS SQL

SAS is not A language. It's rather a set of languages. I regard porting SQL to SAS was one of the best decisions made to enhance SAS. Investing of Proc SQL is definitely an asset.

A.5 SAS STAT

SAS/Stat software offers comprehensive statistical procedures with an incredible user guide (more than). I might not check SAS/Stat as frequently as I do for SAS programming techniques, but I do have an impression that SAS even under advertise its statistical and data mining capacities. Most people don't even know a SAS program can perform Support Vector Machine algorithm by simply firing Proc SVM or Proc HPSVM.

A.6 SAS SGPLOT

SAS statistical graphics procedures came out since SAS 9.2. It's a game changer for SAS programmers to do graphics. SAS SGPLOT is based on grammar of graphics, on which all other popular python or R graphics packages are based. The syntax of SAS SGPLOT is other story; the most important thing is to know the elements of graphics.

A.7 SAS PERL REGULAR EXPRESSION

Undisputedly, this is a positive asset. Perl itself as a programming language might be dying, long lives its regular expression.

A.8 SAS PROC LUA

Lua (pronounced as 'lu-aa') is scripting language like Python. It is very popular in gaming industry. In recent years it's also moved to machine learning world with a widely used library *Torch 7*, which is one of the core libraries used in the AlphaGo system. SAS introduced Lua to its Base module since version 9.4 which will definitely extend SAS programming functionality [4].

A.9 SAS PROC GROOVY

Groovy is a light version of JAVA, and you can do cool thing using SAS Proc Groovy [5].

A.10 SAS PROC IML

I think SAS IML is somehow awkward in SAS software family. A typical "minimum set of SAS system" in most SAS sessions includes the Base SAS, SAS/Stat and SAS/Graph, but no SAS/IML. This prevents SAS programmers to try to use a matrix language in their daily life.

Processing different rows in variable(s) is not an easy task in SAS data steps. Mostly SAS programmers utilize arrays (plus some tricky functions like LAG and DIF) and Dow-loop after its advent. There are advanced programming techniques, but in matrix with index, such manipulation is easy and straightforward (and saving your loops) [6].

Learning Proc IML is also a good bridge when a SAS programmer to learn R and Python. Since SAS/IML 14.2, it even introduces a new data structure that SAS programmers were dreaming of: List [7].

A.11 SAS VIYA

SAS is involving. It's 2017 and you might not realize that there are 2 parallel SAS offerings¹: the oldie and goodie SAS we were talking all around and the new comer, SAS Viya:

RESOURCES

- [Products & Solutions](#)
- [System Requirements](#)
- [Samples](#)
- [Install Center](#)
- [Third-Party Software Reference](#)
- [Documentation](#)
 - › [What's New in SAS](#)
 - › [Product Index A-Z](#)
 - › [SAS Viya](#)
 - › [SAS 9.4](#)

SAS Viya is a new computing platform in which you can use your preferred programming languages (SAS, Python, R, Java, Lua) to perform data manipulation, statistical analysis and data mining [8]. When I first saw this product, I think, the language battle among SAS, Python and R was dead now.

C. SAS PROGRAMMERS STAY COOL

The lesson I learned so far is, a SAS programmer is not a programming solely on one single programming language. Keep this in mind and you will figure out how to stay cool.

A.1 EXPLORE SAS PROGRAMMING CAPACITIES, ONCE MORE

In the previous sessions, I discussed several universal profitable and marketable SAS programming techniques:

- SAS data steps as a bridge to data frame like packages in Python and R
- SAS Proc SQL
- SAS embedded scripting languages like Proc Lua and Proc Groovy
- SAS IML for matrix manipulations
- R and Python in SAS through SAS Viya

A.2 YOU SHOULD GET WET IN MACHINE LEARNING

SAS/Stat offers most comprehensive documentation on statistical procedures in which covers lots of core algorithms in machine learning.

A.3 IT'S ALL ABOUT INCLUDNESS!

SAS programmer is only one label attached to the one who uses SAS. As discussed in the paper, SAS is not **A** language; it's rather a set of languages. A proficient SAS programmer is expected to be a good SQL programmer, data frame programmer; and if we add more skills to this pool, the future SAS programmer is compatible somehow as a Python programmer, R programmer, Lua programmer and such. In this sense, I claim that a good SAS programmer is a good programmer which I always consider cool.

CONCLUSION

It's not fair to compare SAS to Cobol or Perl. SAS as programming language(s), is evolving. A SAS programmer is expected to update their skill set accordingly. I didn't see any barriers to block a SAS programmer to be a good and cool programmer.

REFERENCES

- [1] Gregory Piatetsky, *Python overtakes R, becomes the leader in Data Science, Machine Learning platforms*, <http://www.kdnuggets.com/2017/08/python-overtakes-r-leader-analytics-data-science.html>
- [2] <http://pandas.pydata.org/>
- [3] <http://go.continuum.io/registration-jupyterlab-next-generation-ds-ide/>
- [4] Jiangtang Hu, *New Game in Town: SAS® Proc Lua with Applications*, http://analytics.ncsu.edu/sesug/2016/AD-133_Final_PDF.pdf
- [5] Falko Schulz, *How to import Twitter tweets in SAS DATA Step using OAuth 2 authentication style*, <https://blogs.sas.com/content/sascom/2013/12/12/how-to-import-twitter-tweets-in-sas-data-step-using-oauth-2-authentication-style/>
- [6] Jiangtang Hu, *Incorporate SAS/IML to Base SAS?* <http://jiangtanghu.com/blog/2012/10/16/incorporate-sas/iml-to-base-sas/>
- [7] Rick Wicklin, *Lists: Nonmatrix data structures in SAS/IML*, <https://blogs.sas.com/content/iml/2017/03/29/lists-sasiml.html>
- [8] <http://support.sas.com/documentation/onlinedoc/viya/index.html>

CONTACT INFORMATION

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

Jiangtang('JT') Hu
Life Sciences Consultant
d-Wise Technologies, Inc.
1500 Perimeter Park Dr., Suite 150, Morrisville, NC, 27560
www.d-Wise.com

(O) 919-334-6096
(C) 919-801-9659
(F): 888-563-0931
(O) Jiangtang.Hu@d-Wise.com
(P) jiangtanghu@gmail.com
<http://jiangtanghu.com/>

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.

Other brand and product names are trademarks of their respective companies.