Forecast Update: Will 2014 be the Beginning of the End for SAS and SPSS?

May 14, 2013

By Bob Muenchen

I recently updated my plots of the data analysis tools used in academia in my ongoing article, The Popularity of Data Analysis Software. I repeat those here and update my previous forecast of data analysis software usage.

Learning to use a data analysis tool well takes significant effort, so people tend to continue using the tool they learned in college for much of their careers. As a result, the software used by professors and their students is likely to predict what the next generation of analysts will use for years to come. As you can see in Fig. 1, the use of most analytic software is growing rapidly in academia. The only one growing slowly, very slowly, is Statistica.
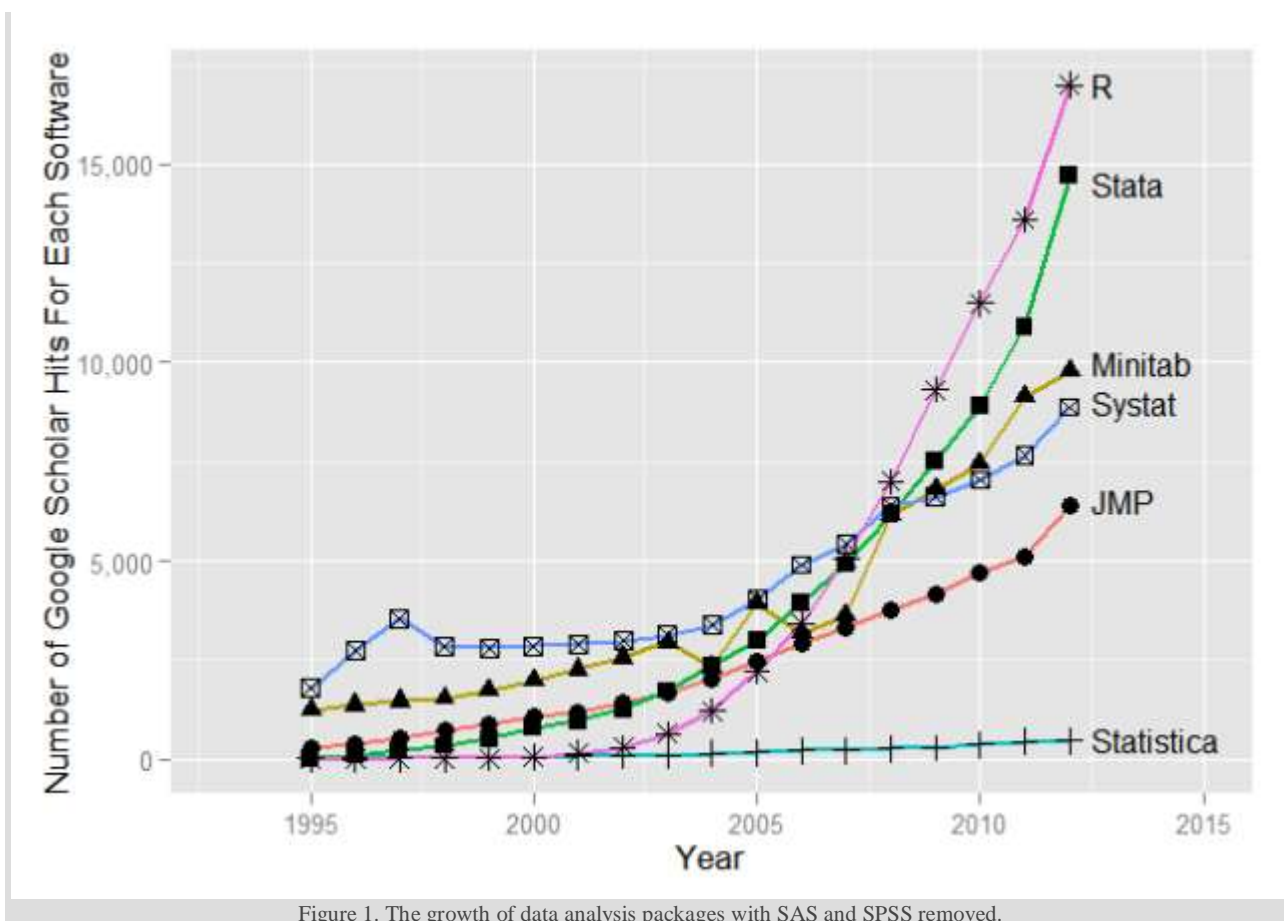

Figure 1. The growth of data analysis packages with SAS and SPSS removed.

While they remain dominant, the use of SAS and SPSS has been declining rapidly in recent years. Figure 2 plots the same data, adding SAS and SPSS and dropping JMP and Statistica (and changing all colors and symbols!)
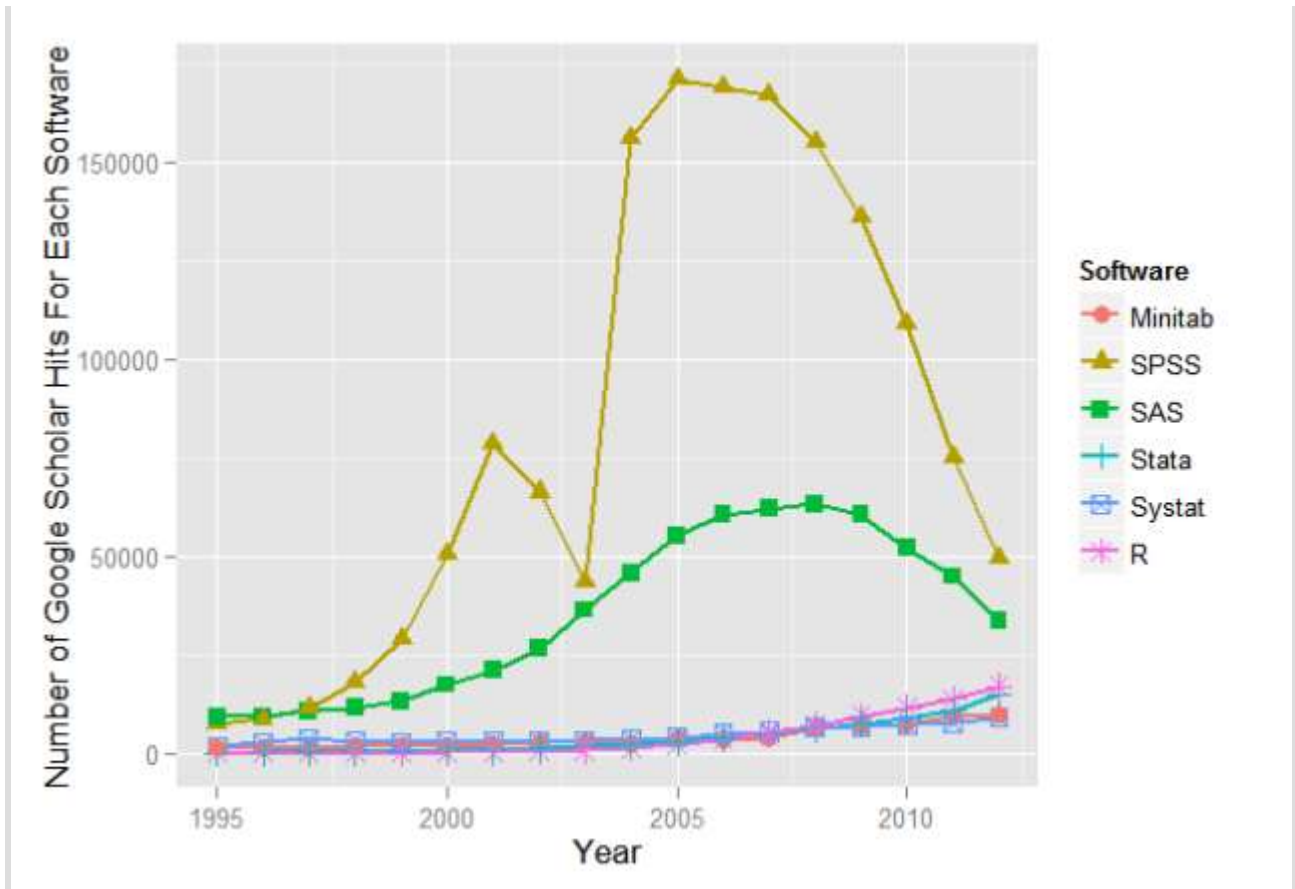

Figure 2. Scholarly use of data analysis software with SAS and SPSS added, JMP and Statistica removed.

Since Google changes its search algorithm, I recollect all the data every year. Last year's plot (below, Fig. 3) ended with the data from 2011 and contained some notable differences. For SPSS, the 2003 data value is quite a bit lower than the value collected in the current year. If the data were not collected by a computer program, I would suspect a data entry error. In addition, the old 2011 data value in Fig. 3 for SPSS showed a marked slowing in the rate of usage decline. In the 2012 plot (above, Fig. 2), not only does the decline *not* slow in 2011, but both the 2011 and 2012 points continue the sharp decline of the previous few years.
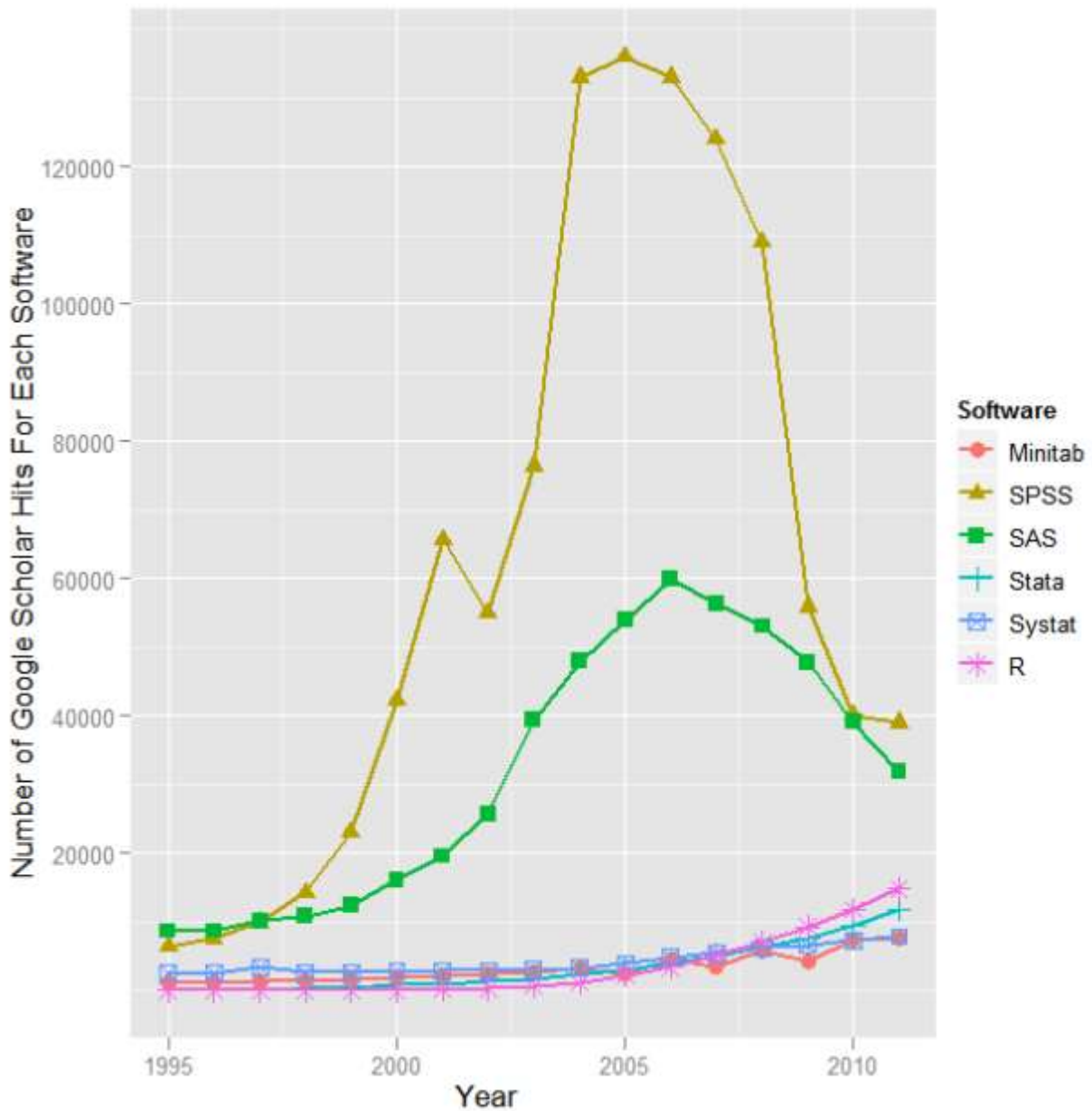

Figure 3. Scholarly use of data analysis software, collected in 2011. Note how different the SPSS value for 2011 is compared to that in Fig. 2.

Let's take a more detailed look at what the future may hold for R, SAS and SPSS Statistics.

Here is the data from Google Scholar:

```
         R    SAS  SPSS    Stata
1995     7   9120  7310       24
1996     4   9130  8560       92
1997     9  10600 11400      214
1998    16  11400 17900      333
1999    25  13100 29000      512
2000    51  17300 50500      785
2001   155  20900 78300      969
2002   286  26400 66200     1260
2003   639  36300 43500     1720
2004  1220  45700 156000    2350
2005  2210  55100 171000    2980
2006  3420  60400 169000    3940
2007  5070  61900 167000    4900
2008  7000  63100 155000    6150
2009  9320  60400 136000    7530
2010 11500  52000 109000    8890
2011 13600  44800  74900   10900
2012 17000  33500  49400   14700
```

ARIMA Forecasting

I forecast the use of R, SAS, SPSS and Stata five years into the future using Rob Hyndman's forecast package and the default settings of its auto.arima function. The dip in SPSS use in 2002-2003 drove the function a bit crazy as it tried to see a repetitive up-down cycle, so I modeled the SPSS data only from its 2005 peak onward. Figure 4 shows the resulting predictions.
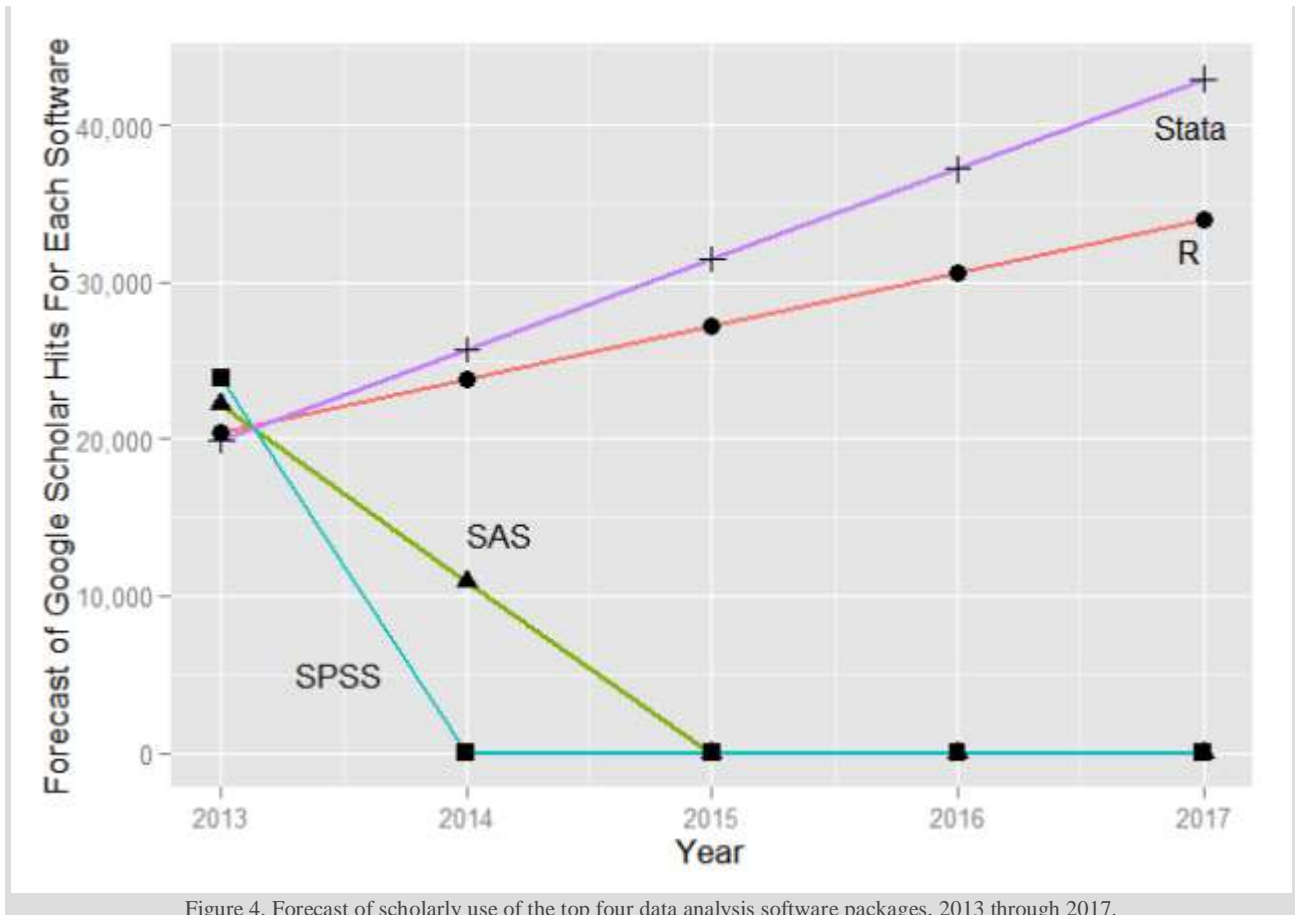

Figure 4. Forecast of scholarly use of the top four data analysis software packages, 2013 through 2017.

The forecast shows R and Stata surpassing SPSS and SAS this year (2013), with Stata coming out on top. It also shows all scholarly use of SPSS and SAS stopping in 2014 and 2015, respectively. Any forecasting book will warn you of the dangers of looking too far beyond the data and above forecast does just that.

Guestimate Forecasting

So what will happen? Each reader probably has his or her own opinion, here's mine. The growth in R's use in scholarly work will continue for three more years at which point it will level off at around 25,000 articles in 2015. This growth will be driven by:

- The continued rapid growth in add-on packages
- The attraction of R's powerful language
- The near monopoly R has on the latest analytic methods
- Its free price
- The freedom to teach with real-world examples from outside organizations, which is forbidden to academics by SAS and SPSS licenses (IBM is loosening up on this a bit)

What will slow R's growth is its lack of a graphical user interface that:

- Is powerful
- Is easy to use
- Provides direct cut/paste access to journal style output in word processor format
- Is standard, i.e. widely accepted as *The One to Use*
- Is open source

While programming has important advantages over GUI use, many people will not take the time needed to learn to program. Therefore they rarely come to fully understand those advantages. Conversely, programmers seldom take the time to fully master a GUI and so often underestimate its full range of capabilities and its speed of use. Regardless of which is best, GUI users far outnumber programmers and, until resolved, this will limit R's long term growth. There are GUIs for R, but with so many to choose from that none becomes the clear leader (Deducer, R Commander, Rattle, at least two from commercial companies and still more here.) If from this "GUI chaos" a clear leader were to emerge, then R could continue its rapid growth and end up as the most used software.

The use of SAS for scholarly work will continue to decline until it matches R at the 25,000 level. This is caused by competition from R and other packages (notably Stata) but also by SAS Instute's self-inflicted GUI chaos. For years they have offered too many GUIs such as SAS/Assist, SAS/Insight, IML/Studio, the Analyst application, Enterprise Guide, Enterprise Miner and even JMP (which runs SAS nicely in recent versions). Professors looking to meet student demand for greater ease of use are not sure which GUI to teach, so they continue teaching SAS as a programming language. Even now that Enterprise Guide has evolved into a respectable GUI, many SAS users do not know what it is. If SAS Institute were to completely replace their default Display Manager System with Enterprise Guide, they could bend the curve and end up at a higher level of perhaps 27,000.

The use of SPSS for scholarly work will decline less sharply in 2013 and will level off in in 2015 at around 27,000 articles because:

- Many of the people who needed advanced methods and were not happy calling R functions from within SPSS have already switched to R or Stata
- Many of the people who like to program and want a more flexible language than SPSS offers have already switched to R or Stata
- Many of the people who needed more interactive visualization have already switched to JMP

The GUI users will stick with SPSS until a GUI as good (or close to as good) comes to R and becomes widely accepted. At The University of Tennessee where I work, that's the great majority of SPSS users.

Although Stata is currently the fastest growing package, it's growth will slow in 2013 and level off by 2015 at around 23,000 articles, leaving it in fourth place. The main cause of this will be inertia of users of the established leaders, SPSS and SAS, as well as the competition from all the other packages, most notably R. R and Stata share many strengths and with one being free, I doubt Stata will be able to beat R in the long run.

The other packages shown in Fig. 1 will also level off around 2015, roughly maintaining their current place in the rankings. A possible exception is JMP, whose interface is radically superior to the others for exploratory analysis. Its use could continue to grow, perhaps even replacing Stata for fourth place.

The future of SAS Enterprise Miner and IBM SPSS Modeler are tied to the success of each company's more mainstream products, SAS and SPSS Statistics respectively. Use of those products is generally limited to one university class in data mining, while the other software discussed here is widely used in many classes. Both companies could significantly shift their future by combining their two main GUIs. Imagine a menu & dialog-box system that draws a simple flowchart as you do things. It would be easy to learn and users would quickly get the idea that you could manipulate the flowchart directly, increasing its window size to make more room. The flowchart GUI lets you see the big picture at a glance and lets you re-use the analysis without switching from GUI to programming, as all other GUI methods require. Such a merger could give SAS and SPSS a game-changing edge in this competitive marketplace.

So there you have it: the future of analytics revealed. No doubt each reader has found a wide range of things to disagree with, so I encourage you to do your own forecasts and add links to them in the comment section below. You can use my data or follow the detailed blog at Librestats to collect your own. One thing is certain: the coming decade in the field of analytics will be interesting indeed!