The origins of SageMath – creating a viable open source alternative to Magma, Maple, Mathematica, and Matlab

William Stein

SageMath, Inc., and University of Washington

June 11, 2016

Thank You!

To the organizers, the audience, and the BP job.

What it is like to be a BP...

William Stein

Harvard University

August 22, 2003 for Microsoft Research

$$Ш(A_f/\mathbf{Q}) = \ker\left(H^1(\mathbf{Q}, A_f) \to \bigoplus_{\text{all } v} H^1(\mathbf{Q}_v, A_f)\right)$$

$$Ш(A_f/\mathbf{Q})$$

SageMath Survey

Survey

- Who has ever heard of Sage (open source math software)?
- Who has used Sage?
- Who has contributed code to Sage?

What this talk is about

This Talk

Why and how I started the Sage math software project at Harvard as a BP, and what has happened since.



Yau's 2004 birthday which was when I started Sage.

1997–2004: my background in mathematical software

- **1991-93:** Computer Science undergraduate
- **1997-99: Hecke** and interpreter in C++;
  for modular forms research with Ribet, Buzzard, and Mazur.
- **1998: Kohel:** Introduced me to both "open source" and Magma, and said "too bad you have to write an interpreter"...
- **1999-2004:** I wrote a lot of Magma code (3 Sydney visits), and tried to convert everyone I met to using Magma.
- **2004: Problems**: Magma is closed source, closed development model, expensive; authorship issues, no user-defined objects; hard to save/load data – not a **mainstream** programming language. But algorithms in Magma are way, way, **way** ahead of open source.

MAGMA
COMPUTER • ALGEBRA

2004: open source?

In 2004 I looked at my laptops and my rack of servers (that Will Hearst donated to Harvard) and the only closed source program on them was Magma. For everything I did, **except the one thing I cared the most about**—mathematics research—open source was a viable option.

My laptops in my office (5th floor of Science Center)

MECCAH **M**athematics **E**xtreme **C**omputation **C**luster at **H**arvard

2004: Magma conference in Paris at IHP

Manjul Bhargava gave a talk about his work on quadratic forms; much of it involved frustration with shortcomings in Magma that he (and Jon Hanke) couldn't address because Magma was closed source.



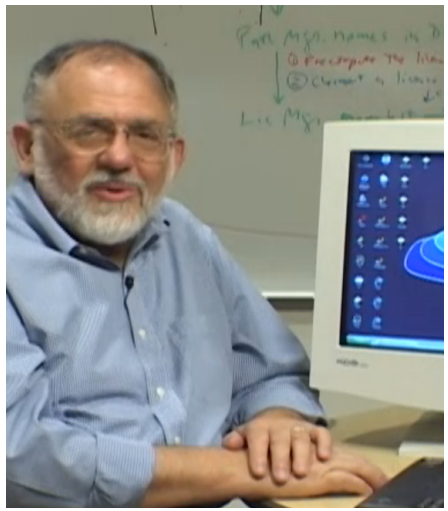
What happened?: Jon Hanke
tried over the next few years to provide an open-source foundation for research on quadratic forms; the result was failing to get tenure at University of Georgia, so he was **forced to leave academia**... Manjul followed a different path, and his academic career was more successful.

2004: Magma, Maple, Mathematica, and Matlab Dominate

Allan Steel (Magma)

Michael Monagan (Maple)

Stephen Wolfram (Mathematica)



Cleve Moler (Matlab)

Don't you worry your pretty little head...

reference.wolfram.com/language/tutorial/WhyYouDoNotUsuallyNeedToKnowAboutInternals.html

WOLFRAM Products & Services Technologies Solutions

Why You Do Not Usually Need to Know about Internals

"You should realize at the outset that while knowing about the internals of Mathematica may be of intellectual interest, it is usually much less important in practice than you might at first suppose. Indeed, in almost all practical uses of Mathematica, issues about how Mathematica works inside turn out to be largely irrelevant. Particularly in more advanced applications of Mathematica, it may sometimes seem worthwhile to try to analyze internal algorithms in order to predict which way of doing a given computation will be the most efficient. But most often the analyses will not be worthwhile. For the internals of Mathematica are quite complicated."

`http://reference.wolfram.com/mathematica/tutorial/WhyYouDoNotUsuallyNeedToKnowAboutInternals.html`

2004: I got a job offer at UCSD with **tenure**!

Todo:
- really learn maple "easy to use"
- learn mathematica

So I can make Monin more familiar, explain it, etc. to users of those systems.

Need better name than "MANIN"
- pymath
- Sage ← "System for Arithmetic Geometry experimentation"
- —

→ but also a sacred Navajo herb

System for Arithmetic Geometry Experimentation

2005: I Launched Sage

What I want

I really, really, **really** want open source software for research mathematics that isn't way behind Magma.

We are willing to spend years on a single mathematics research problem. I just got tenure, so surely I can spend years to create such software.

I launch the first version of Sage at Harvard in Feb 2005 after a year of investigating options and building prototypes. David Joyner becomes the first ever users/developer.

SAGE: System for Arithmetic Geometry Experimentation

http://modular.fas.harvard.edu/SAGE/

William Stein

Asst. Professor of Mathematics, Harvard University

March 24, 2005, PYCON, Washington, D.C.

Arithmetic Geometry

Arithmetic geometry is about geometric questions that have an arithmetic flavor. Sample famous problems:

- **Fermat's Last Theorem:** $x^n + y^n = z^n$ has no solutions with $n \geq 3$ and x, y, z all nonzero integers. Andrew Wiles proved this in 1995 using elliptic curves and modular forms.
- **The Birch and Swinnerton-Dyer Conjecture:** Discovered from computations in the 1960s. Simple criterion for whether or not for given a, b the elliptic curve $y^2 = x^3 + ax + b$ has infinitely many rational solutions. (Clay \$$10^6$ dollar problem.)
- **The Riemann Hypothesis:** Nontrivial zeros of Riemann Zeta function are on line $\mathrm{Re}(s) = 1/2$. Solution gives deep understanding of distribution of the prime numbers $2, 3, 5, 7, 11, 13, \ldots$. (Clay million dollar problem.)
- **Cryptography:** Factor integers quickly. E.g., Hendrik Lenstra, used elliptic curves to give a new algorithms for this. The number field sieve is a sophisticated algorithm for factoring and uses computation in number fields. Also, cryptosystems come from elliptic curves over finite fields.

The Problem

Create a system for doing computations with the mathematical objects mentioned above. Main Goals:

- **Efficient:** Be very fast – comparable to or faster than anything else available. This is very difficult, since most systems are closed source, algorithms are often not published, and finding fast algorithms is often extremely difficult (years of work, Ph.D. theses, luck, etc.)
- **Open Source:** The source code must be available and readable, so users can understand what the system is really doing and trust the results more.
- **Comprehensive:** Implements enough different things to be really useful.
- **Well documented:** Reference manual, API reference with examples for every function, and at least one published book. Make documentation and source a peer reviewed package, so get academic credit like a journal publication.
- **Extensible:** Be able to define new data types or derive from builtin types, or make code written in your favorite language part of the system.
- **Free:** Must be sufficiently free (at least GPL).

Existing Mature Systems

- **Mathematica, Maple, and MATLAB:** Arithmetic geometers are not their target audience. Mathematica does well at special functions, and both do calculus very well, which is almost never useful in arithmetic geometry. These systems are *closed source, very expensive, for profit.*
- **MAGMA:** *By far* the best software for arithmetic geometry. It's very efficient, comprehensive, and well documented. Great design and class hierarchy. BUT: It's closed source (but non-profit), expensive, and not easily extensible (no user defined types or C/C++-extension code). I've contributed substantially to MAGMA.
- **PARI:** Efficient, open source, extendible and free. But the documentation is not good enough and the memory management is not robust enough. Also, PARI does not do nearly as much as what is needed.
- **Maxima, Octave, etc.:** Open source, but not for arithmetic geometry.

(There's always something else that I don't know about.)

All these system use their own custom programming language.

2005: Backlash

"This
is to formally advise you that your permission to run a general-purpose calculator based on Magma expires on Dec 31, 2005. This was originally set up at your request so students in your courses at Harvard could have easy access to Magma."
– John Cannon

2005: "Sage is essentially doomed"

I've talked with founders of Maple, Magma, Mathematica, and Maxima... who have all told me that Sage is doomed without being commercial.

> *"By avoiding applications (say, to engineering design, finance, education, scientific visualization, etc etc)* **Sage is essentially doomed**. *Why? Government funding for people or projects will be a small percentage of the funding for pure mathematics. That's not much.* **And the future is pretty grim.**"
>
> *–Richard Fateman, UC Berkeley computer scientist, 2005.*

I will prove them wrong. The pure mathematics community is amazing!!

2006: UC San Diego – The First Sage Days

Sage Days 1

Sage Days have been very successful: there have been over 80 of them. Next few this year are in Vancouver, Hawaii, France and Israel.

2006: I created a mission statement to clarify direction

Sage Mission Statement

Create a viable free open source alternative to Magma, Maple, Mathematica and Matlab.

Our goal isn't to be better than other open source projects, or publish papers, or solve a particular problem. It is to use and improve the entire open source world in order to **give you and your students** more options.

2006: Built a badass team at Univ of Washington!

2008: There may be a flaw in my master plan

Everybody graduated and I couldn't hire any of them. Instead, they got amazing jobs at Google, etc.

Brilliant devs would show up and write incredible code, but then vanish since I couldn't pay them. **(Everybody assumed I had tons of grants!)**

Everybody left.

I failed to get funding to hire even a single person fulltime to focus on Sage. (In 2016 in Europe, the first ever fulltime Sage person – Erik Bray – was finally hired with a new EU grant, and he's awesome!)

Heartbleed – "the worst security bug ever"

A massive bug in OpenSSL (discovered by Google) made it so a huge number of the world's webservers can be hacked. OpenSSL is used by hundreds of thousands of sites and many multibillion dollar companies.

> *"There should be at least a half dozen full time OpenSSL team members, not just one, able to concentrate on the care and feeding of OpenSSL without having to hustle commercial work. If you're a corporate or government decision maker in a position to do something about it, give it some thought.* **Please. I'm getting old and weary...**" – *Steve Marquess*

2012: Simons Foundation Roundtable - Finally!

I was incredibly excited in 2012 when Eisenbud invited me to a meeting at the Simons Foundation headquarters with the following goal:

"The purpose
of this round table is to investigate what sorts of
support would facilitate the development, deployment
and maintenance of open-source software used
for fundamental research in mathematics and [...]
The scale of foundation support could be substantial,
perhaps up to several million dollars per year."

Simons

Wow, these guys are serious, and really understand pure mathematics. Finally, there is hope! This was the moment I had spent the years preparing for.

The Decision: *fund Magma for all North American institutes.* WTF!? Many people worked on or used Sage because they couldn't afford Magma. Holy sh!t; this went far worse than I could have ever possibly imagined.

(Simons Foundation follow up)

People *frequently* tell me *"Jim Simons is a true academic at heart [...] Approach his fund. I'm 100% sure he'll give you a grant on the spot."* OK! So in 2015, I wrote to Yuri Schinkel, current director of the Simons Foundation:

Dear William,

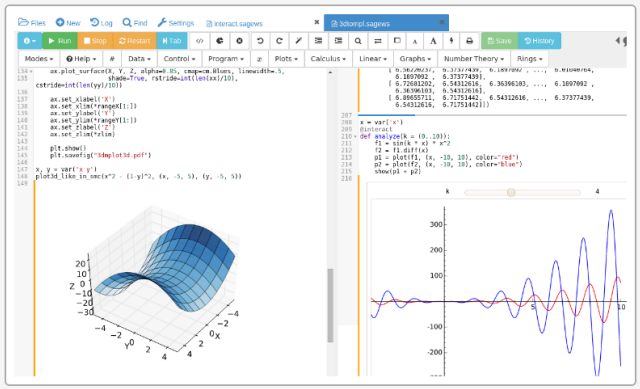
Before I joined the foundation, there was a meeting conducted by David Eisenbud to discuss possible projects in this area, including Sage.

After that meeting it was decided that the foundation would support Magma.

Please keep me in the loop regarding developments at Sage, but I regret that we will not fund Sage at this time.

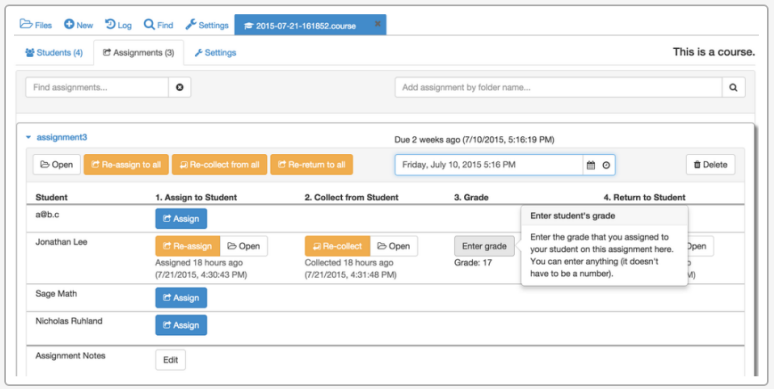
2012: Started building SageMathCloud...

Interactive Worksheets



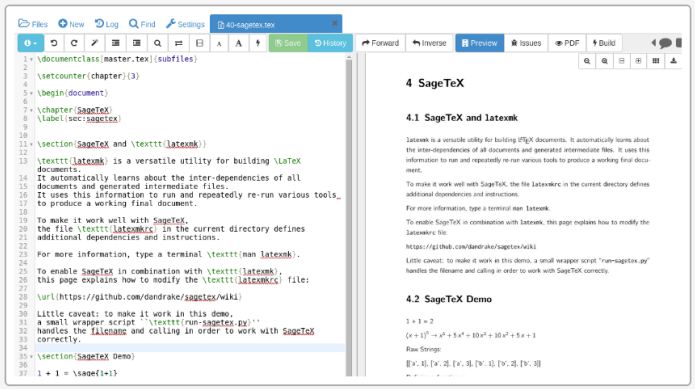
Interactively explore mathematics, science and statistics. **Collaborate with others in real time**. You can see their cursors moving around while they type — this works for Sage Worksheets and even Jupyter Notebooks!

Course Management



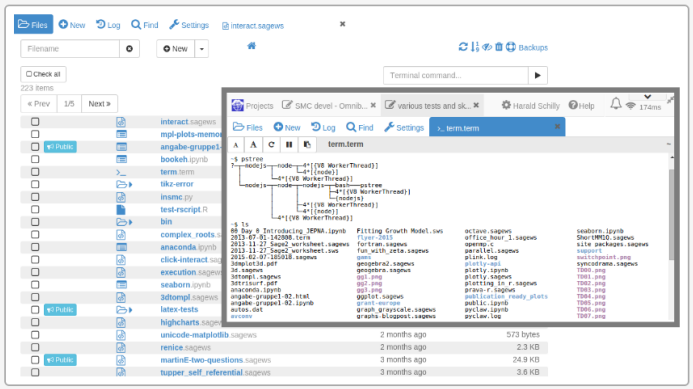
SageMathCloud helps to you to **conveniently organize a course**: add students, create their projects, see their progress, understand their problems by dropping right into their files from wherever you are. Conveniently handout assignments, collect them, grade them, and finally return them. (SMC used for Teaching and learn more about courses).

LaTeX Editor



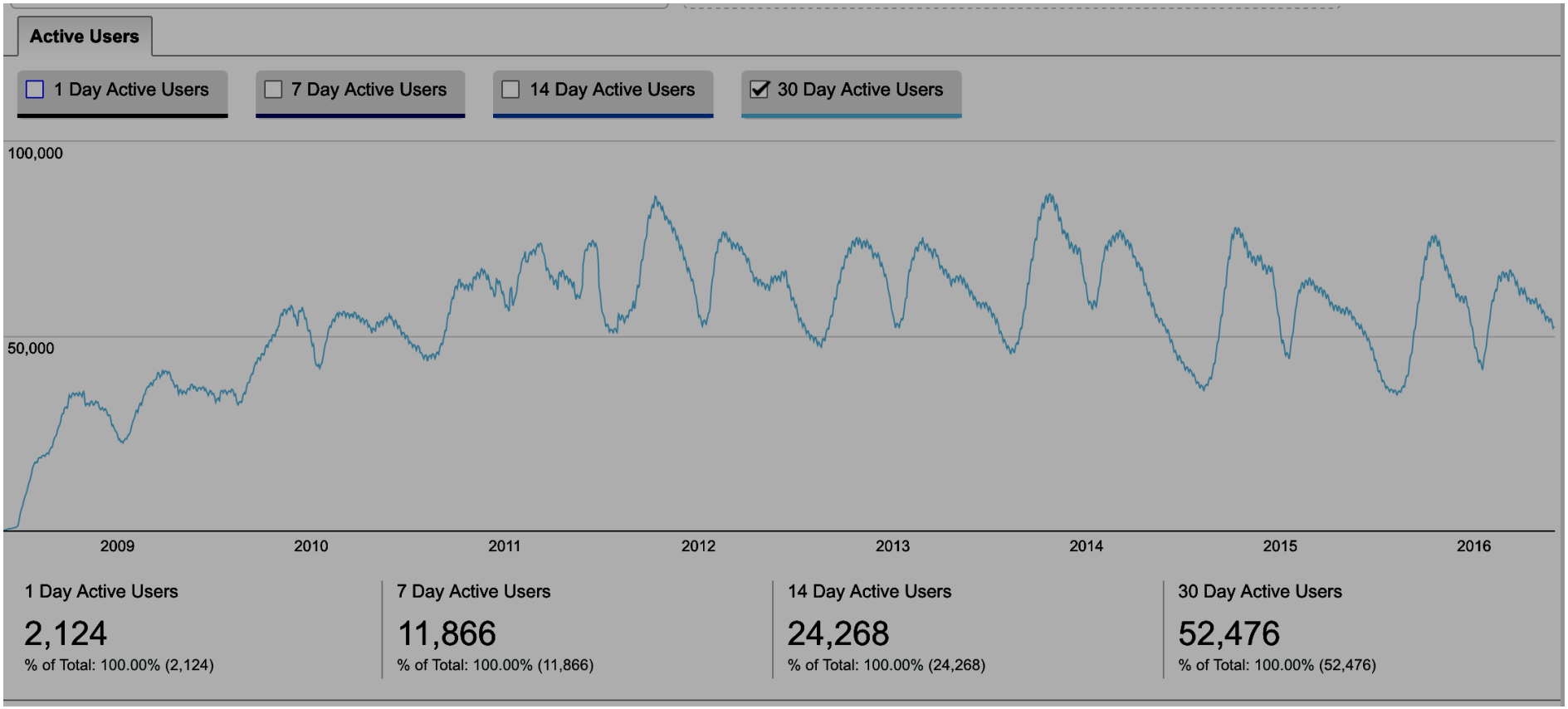
SageMathCloud supports authoring documents written in LaTeX, Markdown or HTML. The **preview** helps you understanding what's going on. The LaTeX editor also supports **forward and inverse search** to avoid getting lost in large

The Sky is the Limit



SageMathCloud does not arbitrarily restrict you. **Upload** your own files, **generate** data and results online, then download or **publish** your results. Besides Sage Worksheets and Jupyter Notebooks, you can work with a **full Linux terminal** and

2016 Status: Growth of Sage completely stopped in 2011



sagemath.org: 50K Monthly Active Users

Sage dev is very active with **hundreds of contributors**. However, **clearly something is missing** to Sage being a viable alternative to the Ma's! They have millions of users.

Status: Sage is still significantly behind Magma

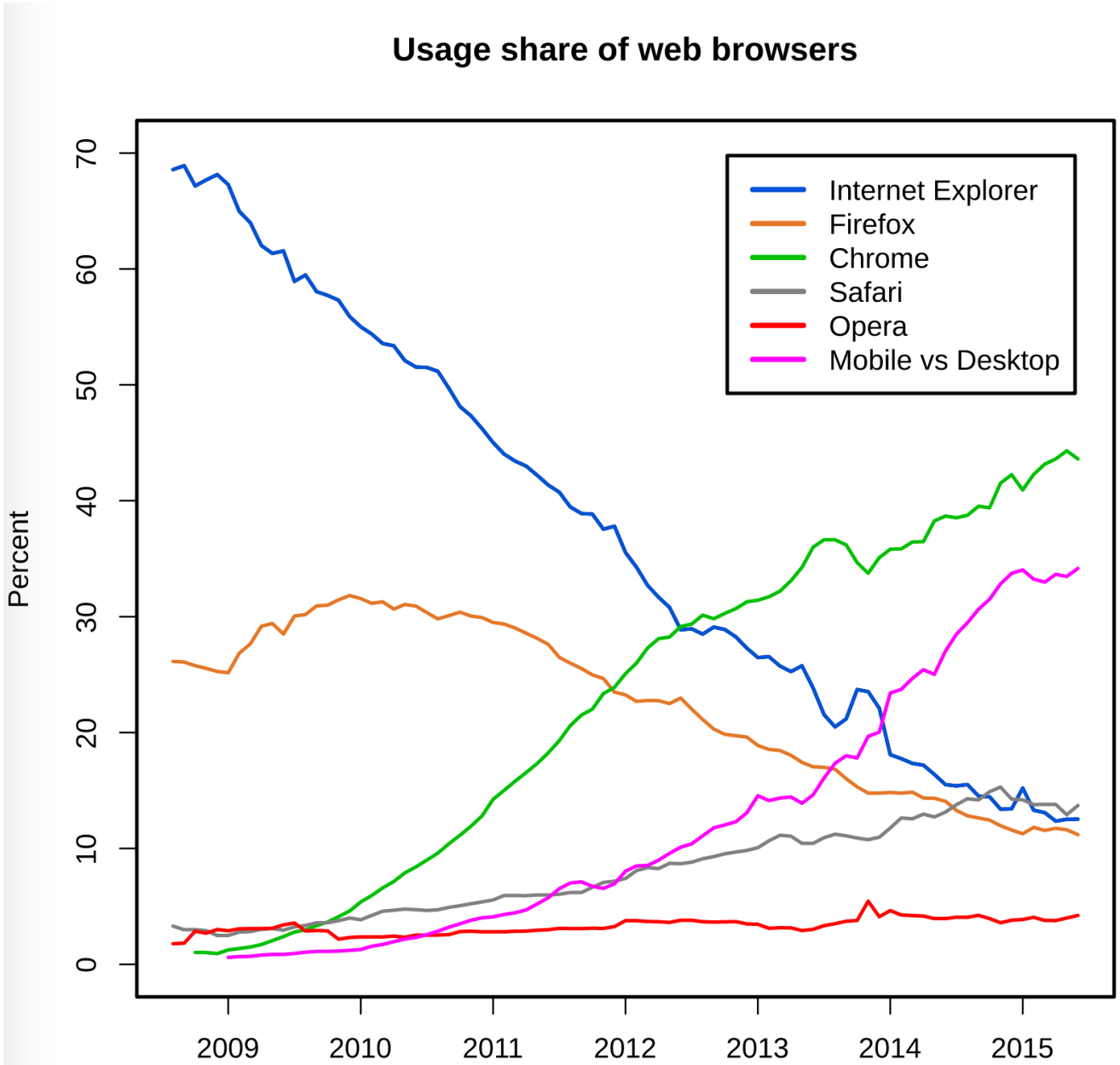
David Kohel knows both Magma and Sage intimately:

David Kohel (this morning)

"I should think how I can contribute to development in arithmetic geometry, in which **Sage is not on par with Magma.**"

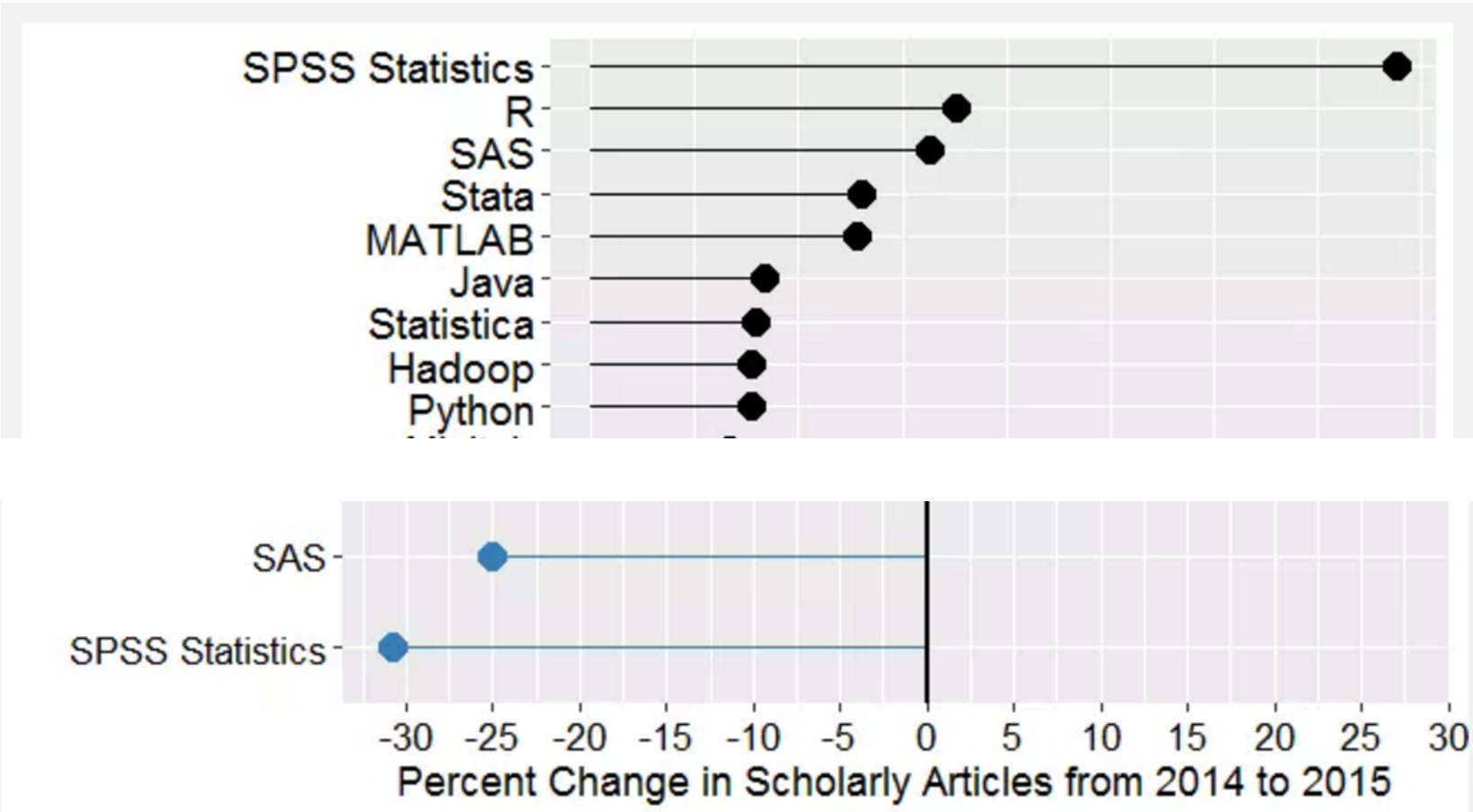
And there is also Mathematica, etc...

But Open Source Can Win: Web Browsers



But Open Source Can Win: Statistics Software

R is free, open source, and the defacto standard in statistics.



Chrome and R have very strong company support

Stephen Wolfram (1993): "The [mathematical community] sort of hate one aspect of what I have done, which is to take intellectual developments and **make a company** out of them and sell things to people.

My own view of that, which has hardened over the years, is, my god, that's the right thing to do. If you look at what's happened with TeX, for example, which went in the other direction...well, Mathematica could not have been brought to where it is today if it had not been done **as a commercial effort**. The amount of money that has to be spent to do all the details of development, **you just can't support that in any other way than this unique American idea of the entrepreneurial company.**"

I used to think Wolfram was wrong. Now I am not so sure.

2016: If you were a new faculty member again...

Dear Prof. Stein,

I'm a (relatively new) assistant professor in math with a heavy software bent and can't help but note your recent blog posts about your frustration with grants for Sage and, more generally, academic respect for software libraries. To put it mildly, I find this concerning, as my software output is (by far) my biggest effort and output.

If you were a new faculty member again, would you start something like SageMathCloud sooner or simply leave for industry?

Sincerely,

2016: Deep mathematical software is not appreciated...

Hi William,

I am sitting on an offer from Google and am increasingly frustrated by continual evidence that it is more valuable to publish a litany of computational papers with no source code than to do the thankless task of developing a niche open source library. **Deep mathematical software is not appreciated by either the mathematicians or the public.**

I had been on the fence about accepting the offer, and this conversation led to me making the difficult decision.

– Jack Poulson, Stanford

A leader of a major open source project in academia told me yesterday...

Issues with software dev in academia

- Hard money for software development is virtually nonexistent: I can't think of anyone I know who got tenured based on his or her software.
- Researchers on soft money are systematically discriminated against in favor of tenure-track and tenured faculty.
- Researchers are increasingly evaluated solely on bibliometric counts rather than an informed assessment of their overall portfolio of papers, code, software, industry engagement, or student supervision.

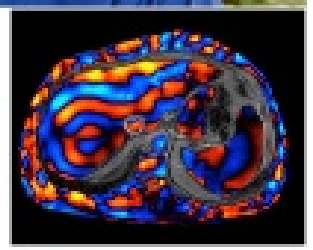
"The hits particularly close to home... I am sort of the walking example of that painful trifecta." – Fernando Perez, founder of IPython and Jupyter, another major open source project.

"Every great open source math library is built on the ashes of someone's academic career."

For example...

Travis Oliphant - CEO

- PhD 2001 from Mayo Clinic in Biomedical Engineering
- MS/BS degrees in Elec. Comp. Engineering
- Creator of **SciPy** (1999-2009)
- Professor at BYU (2001-2007)
- Author of **NumPy** (2005-2012)
- Started **Numba** (2012)
- Founding Chair of **Numfocus** / **PyData**
- Previous PSF Director



NumPy SciPy

Num FOCUS Open Code, Better Science

python

I am leaving academia to build a company

Academia has been **good to me personally**, with grants, winning the Jenks prize, and getting hired with tenure right out of my BP.

But I can't figure out how to create Sage in academia. The money isn't there. The mathematical community doesn't care enough.

The only option left is for me to build a company.

SageMath, Inc.

We launched a company—SageMath, Inc.—and I will start working fulltime for it a week from now.