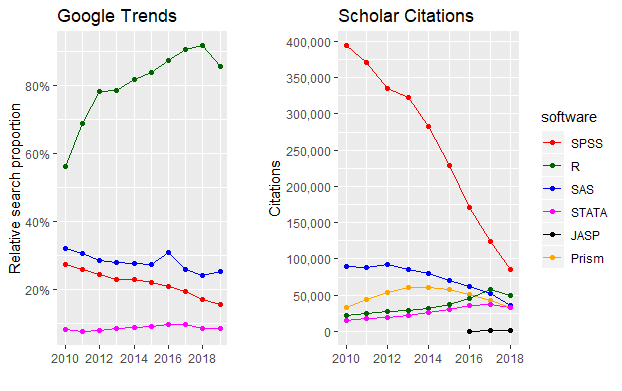
SPSS is dying. R takes the lead in 2020.

SPSS is dying. R is surging and I predict that R overtakes SPSS in yearly citations by 2020. The implications are clear: (1) if you use SPSS in your business or research, move to R now rather than later. (2) Do not ask for SPSS competences in job postings. You will not attract the good candidates. (3) We are doing students a disservice by teaching SPSS. Switch to [JASP](http://jasp-stats.org) and R. [Rstudio Desktop](https://www.rstudio.com/products/rstudio/download/) is a highly recommended interface to R.

# The numbers

The numbers have been clear for a number of years now that SPSS was on the decline. It was very clearly exposed by Robert A. Muenchen in a [comprehensive 2016-analysis of the use of data science software](http://r4stats.com/articles/popularity/) where he looked at everything from job postings to online queries to academic citations. I have updated two of these analyses to include data from 2017 and 2018: Google Search Trends and citations in the academic literature.



Here, we need to look at the trends rather than the absolute values for reasons I explain in the end of this post. Although R took a small dip in 2018, it is clear that it is getting traction along with STATA. It is a good guess that R and SPSS will par in citations in 2019 and that R will have overtaken SPSS by 2020. Let’s discuss what this means.

# Why SPSS is dying

A few years ago, [I wrote a blog](https://jasp-stats.org/2017/11/01/jasp-vs-spss/) about how a new GUI program, [JASP](http://jasp-stats.org), gets most things right, and how that exposing SPSS’s many shortcomings in the process ([SPSS now plans](http://r4stats.com/articles/how-to-search-for-data-science-articles/) to go in a more JASP-like direction for the GUI). SPSS simply feels *old and unmaintained*. Users have been screaming for simple statistics like Cohen’s d, confidence intervals on correlation coefficients, meta-analysis, etc., which has been a mandatory part of many major publication guidelines since 2000. This is not just some science formalia - these statistics are highly informative for industry as well. Despite repeated requests, SPSS has not implemented these, or many other standard statistical methods.

In addition, both industry and science now require greater reproducibility, transparency, and interaction with data. If you have ever tried using SPSS, you will know that it is fundamentally not fit for these.

# Why R is surging

R saves you time. First, it is free, saving you (and your collaborators) time handling licensing and asking for budget approvals.

As a statistician, most of the time is spent pre-processing data before doing the statistics. Since the advent of [tidyr](https://tidyr.tidyverse.org/) in 2014, this has become incredibly easy to do. Perhaps more importantly, it has become much easier to read the code, which facilitates seamless collaboration and empowers you to learn much quicker from examples online. Pre-processing is often a non-linear process where you go back and forth. R is like editing a recipe in a text editor, and SPSS is like having to dictate the whole thing on a tape every single time (point-and-click).

When researchers develop novel analysis methods, they will often publish them in user-friendly R-packages even before they publish the accompanying academic paper. For the most part, if you can think if it, it exists and is only one “install.package()” away. Not stumbling into road blocks saves you time.

Perhaps counterintuitively, it turns out that students like programming as it helps them better grasp what they are doing to the data than a point-and-click interface. In R, you can filter data and do a mixed model in just four relatively brief and self-explanatory lines of code, while it would take 20+ clicks in SPSS. If you want to do it for multiple datasets, you have to go through all that SPSS-clicking again. It is *really* easy to miss a click and unknowingly get wrong results. R saves debugging time.

After all, statistics is about the interaction and processing of variables, so programming often requires *less* abstraction than graphical user interfaces.

Programming is, of course, overkill for routine one-off analyses with little pre-processing. [JASP](http://jasp-stats.org), and its sibling [Jamovi](https://www.jamovi.org/), are free graphical user interfaces to R that fills in this space.

# Implications for industry and science

Consider SPSS liability. Either weakly through taking more person-hours to use. Or strongly, through the increased risk of errors. R may require a

Ask for R competences in job postings. If you ask for SPSS competences, you will select for applicants who are not up to date and filter out those who are, because they will want to avoid SPSS.

We should also stop teaching SPSS. Students spend a disproportionally large amount of learning the interface rather than learning the statistics. When they graduate, the cost of SPSS will incentivize them to avoid stats. JASP may be a good start for undergraduates because of the very shallow learning curve and the sensible defaults. Then switch to R in the next semester or for graduate students.

# Notes about the graph

The Google Trend values for R are probably inflated by the fact that it is a more technical audience which probably uses the internet more while SPSS users do less advanced analyses based on books. As a reflection of actual usage, we should probably just look at the trends.

The absolute citation numbers in the graph is a bit misleading since it goes down while we know that the annual publication volume is increasing. It seems that we simply cite the analysis software less frequently than we used to. However, the *relative* popularity of software packages is still valid.

To collect these data, I wrote a Google Scholar Scraper here (in R, of course) and I have posted the dataset here . I used Robert Muenchen’s search terms which I found valid. As a side effect, you can use this scraper to collect time-trends in all Google Scholar searches. I just happened to search for statistical packages.

Note that you need to update the full dataset to compare citations by year. For example, Robert found ~300.000 SPSS citations in 2011 where I found ~375.000 using the same search string. Google Scholar improves and more publications are retrospectively put online.