

Who's Bringing That Big Data Energy?

A 47-Year Longitudinal Analysis of 30,000 Presentations in the SAS® User Community To Elucidate Top Contributors and Rising Stars

Troy Martin Hughes

ABSTRACT

This analysis examines presentations at SAS® user group conferences between 1976 and 2023. It includes presentations referenced on www.LexJansen.com (aka “the LEX”) during this timeframe, which are drawn from multiple conferences, including: SAS User Group International (SUGI, *may she rest in peace*), SAS Global Forum (SGF, *may she be revived*), SAS Explore, Western Users of SAS Software (WUSS), Midwest SAS Users Group (MWSUG), South Central SAS Users Group (SCSUG), Southeast SAS Users Group (SESUG), Northeast SAS Users Group (NESUG), Pacific Northwest SAS Users Group (PNWSUG), and Pharmaceutical Software Users Group (PharmaSUG). This analysis identifies top contributors, including authors who have presented most abundantly at specific conferences, as well as across all conferences. For example, the SAS superstars and most prolific presenters of all time are ranked and recognized for bringing that big data energy (BDE)—Kirk Paul Lafler, Arthur L. Carpenter, Louise Hadden, Charlie Shipp, and Ronald J. Fehd! Rising stars, who may be new to the conference scene, yet are contributing significantly, are also identified. In addition to quantifying and extolling the contributions of these authors, this analysis aims to assist the leaders of future conferences in identifying key speakers to invite. Threats to data quality are also discussed—the reality of third-party data, as conferences often feed the LEX with raw presentation metadata that either are unstandardized or have patently wrong information (like incorrect author names or missing coauthors). Finally, and perhaps with some irony, Python 3.11.5 (and Pandas 2.1.0) was used exclusively to extract, clean, transform, and analyze all data.

INTRODUCTION

SAS user groups have been a staple of the SAS community for decades, and offer SAS users of all ilk the opportunity to meet, socialize with, and strategize with other users, and with SAS technical support, SAS developers, SAS authors, and industry thought leaders throughout the SAS analytic community. Global conference events include those hosted by SAS, such as SUGI, SGF, and SAS Explore. Regional user groups (RUGs) are hosted by non-profit organizations, with active conferences including WUSS and SESUG, inactive conferences (not having revived since the COVID-19 pandemic) including MWSUG and SCSUG, and defunct conferences including NESUG and PNWSUG. PharmaSUG—originally branded as the Pharmaceutical SAS Users Group, and now more broadly as the Pharmaceutical *Software* Users Group—remains the solitary industry-specific user group.

Central to all SAS user group conferences are the user- and SAS-supplied content—presentations, poster presentations, hands-on trainings, live demos, and panel discussions that convey best practices and lessons learned about all facets of SAS software, and which demonstrate solutions users have programmed in the SAS language. And for years, the single repository to which SAS practitioners have turned to view conference content is the www.LexJansen.com website (aka, “the LEX”), which is run by the former SAS employee—and fellow tall guy—the incomparable Lex Jansen! That is, long before the SAS Support Community site (<https://communities.sas.com/>) was built (in part) to host solutions to SAS technical challenges, and long before Stack Overflow (<https://stackoverflow.com/>) had hosted its first SAS-related question, SAS practitioners—the world over—had turned to the LEX for SAS programming support.

This analysis evaluates 30,002 presentations from the foregoing conferences whose metadata are maintained on the LEX, comprising approximately 14,918 unique presenters. Of these presentations, 96.5 percent (N=28,966) have an accompanying “white paper” that is maintained as a PDF. However, this analysis focuses on presentations as opposed to papers, and evaluates presentation metadata, such as conference, title, presentation type, and author(s). Thus, no attempt to analyze paper content is made. Appendix A lists the 150 top contributors (i.e., most prolific presenters) across all time and all conferences.

And as the title suggests, this analysis focuses primarily on the presenters who deliver content at conferences—preparing for weeks, oftentimes building a PowerPoint or other slide deck, practicing their presentation, sweating through their shirts and skirts, and fielding questions from thirsty SAS users. Highlighted herein are the indisputable top contributors. Finally, a gracious “thank you” (along with some Zacapa XO rum) is owed to Lex for his friendship, his review of this manuscript, and his tireless efforts in maintaining his eponymous trove of SAS knowledge.

INTRODUCING THE LEX

Also known for its Rock and Roll repository, the LEX primarily showcases SAS presentations and papers that span the ages. Early papers represent digital scans of the conference proceedings—the enormous tomes that SAS users once conveyed home from each conference, and which now occasionally dot the eBay landscape; I own a few and these boys are girthy! Later papers, rather, were created digitally, and thus are more legible. Through the catalog, one can follow the birth and growth of the SAS language while exploring the ins and outs of Base SAS, the SAS macro language, and countless SAS tools and modules.








[in](#)
[t](#)
[v](#)
[i](#)
[d](#)

SAS (and R) Conference Proceedings (1976 – present) ... and more

This searches **37037** conference papers from SAS Global Forum, SUGI, PharmaSUG, PhUSE, NESUG, SESUG, WUSS, MWSUG, PNWSUG, SCSUG, SEUGI, ...

ENHANCED BY Google

Perform a [search](#) for papers based on title, author or keywords.

[38 Papers](#) written by [Lex Jansen](#).
[Contact me](#)

This website is a personal and not-for-profit project maintained by [Lex Jansen](#). It is not affiliated with, or representing the views of [SAS](#), [CDISC](#) or any other organization.

578 PharmaSUG China papers (2012-2023) PharmaSUG China 2023 August 4-5 - Nanjing, China	181 RPharma presentations (2018-2021) RPharma 2023 September 17-20 - Chicago, IL	337 PHUSE/FDA CSS posters (2012-2023) PHUSE CSS 2024 June 3-5
81 SAS Explore presentations (2022) SAS Explore 2023 September 11-14 - Las Vegas, NV	2371 PHUSE EU Connect papers (2005-2022) PHUSE EU Connect 2023 November 5-8 - Birmingham, UK	4030 PharmaSUG papers (1997-2023) PharmaSUG 2024 May 19-22 - Baltimore, MD
3207 SESUG papers (1993-2022) SESUG 2023 October 22-24 - Charlotte, NC	2914 WUSS papers (1993-2022) WUSS 2023 Oct 31 - Nov 2 - San Diego, CA	1016 PHUSE US Connect papers (2018-2023) PHUSE US Connect 2024 February 25-28 - Bethesda, MD
12850 SUGI / SAS Global Forum papers (1976-2021)	2225 CDISC related papers and posters (2001-2023)	2111 MWSUG papers (1990-2019)
1402 SCSUG papers (1991-2019)	3123 NESUG papers (1988-2013)	286 PNWSUG papers (1996-2009)
115 VIEWS papers (2001-2004)	2435 SEUGI papers (1983-2003)	

Users can select a specific conference to view only its data. For example, clicking on “WUSS” yields a listing of each WUSS conference year, dates, and location.

← → ↺

lexjansen.com/wuss/

Home

Contact

SAS Paper Search

SAS:MWSUG papers

SAS:NESUG papers

R/R/Pharma presentations

SAS:PharmaSUG papers

SAS:PharmaSUG China papers

SAS:PHUSE EU papers

SAS:PHUSE US papers

FDA/PHUSE CSS posters

SAS:PNWSUG papers

SAS:SCSUG papers

SAS:SESUG papers

SAS:SEUGI papers

SAS:SASGF/SUGI papers

SAS:SAS Explore presentations

SAS:VIEWS papers

SAS:WUSS papers

SAS:CDISC papers

Fortune records

Marsh 1001

WUSS

Western Users of SAS® Software ([WUSS](#)) is a non-profit organization that exists to provide [SAS®](#) software users a forum for the exchange of ideas and to promote education, training, networking, and professional development opportunities for SAS® users. WUSS primarily does these through their annual conference, which brings together 500 SAS® users from the WUSS region and beyond. The WUSS region includes Alaska, Arizona, California, Colorado, Hawaii, Idaho, Montana, Nevada, Oregon, Utah, Washington and Wyoming. Western Users of SAS® Software is the officially recognized regional SAS® users group of the Western United States.



Conferences

[WUSS 2022](#) September 14-16, 2022, San Francisco, California

[WUSS 2019](#) September 4-6, 2019, Seattle, Washington

[WUSS 2018](#) September 5-7, 2018, Sacramento, California

[WUSS 2017](#) September 20-22, 2017, Long Beach, California

[WUSS 2016](#) September 7-9, 2016, San Francisco, California

[WUSS 2015](#) September 9-11, 2015, San Diego, California

[WUSS 2014](#) September 3-5, 2014, San Jose, California

[WUSS 2013](#) November 13-15, 2013, Las Vegas, Nevada

[WUSS 2012](#) September 5-7, 2012, Long Beach, California

[WUSS 2011](#) October 12-14, 2011, San Francisco, California

[WUSS 2010](#) November 3-5, 2010, San Diego, California

[WUSS 2009](#) September 1-3, 2009, San Jose, California

[WUSS 2008](#) November 5-7, 2008, Universal City, California

[WUSS 2007](#) October 17-19, 2007, San Francisco, California

[WUSS 2006](#) September 27-29, 2006, Irvine, California

[WUSS 2005](#) September 21-23, 2005, San Jose, California

[WUSS 2004](#) October 13-15, 2004, Pasadena, California

[WUSS 2003](#) November 5-7, 2003, San Francisco, California

[WUSS 2002](#) September 4-6, 2002, San Diego, California

[WUSS 2001](#) September 5-7, 2001, San Francisco, California

[WUSS 2000](#) September 17-19, 2000, Scottsdale, Arizona

[WUSS 1999](#) October 20-22, 1999, Los Angeles, California

[WUSS 1998](#) October 14-16, 1998, Oakland, California

[WUSS 1997](#) October 22-24, 1997, Universal City, California

[WUSS 1996](#) September 4-6, 1996, San Francisco, California

[WUSS 1995](#) September 27-29, 1995, Long Beach, California

[WUSS 1994](#) October 26-28, 1994, San Francisco, California

[WUSS 1993](#) October 20-22, 1993, Santa Monica, California

[←](#) [→](#) [↺](#) [http://laxjansen.com/cgi-bin/saspapers_query.php](#)

Home	
Contact	
SAS Paper Search	
SAS:MWSUG papers	
SAS:NESUG papers	
R/R/Pharma presentations	
SAS:PharmaSUG papers	
SAS:PharmaSUG China papers	
SAS:PHUSE EU papers	
SAS:PHUSE US papers	
FDA/PHUSE CSS posters	

Search proceedings from MWSUG, NESUG, PharmaSUG, PHUSE, PNWSUG, SAS Global Forum, SUGI, SEUGI, SCSUG, SESUG, VIEWS, and WUSS. This search is based on title and author. You need to specify at least 3 characters in one field.

Searches are case insensitive and can be done in one of 3 ways:

- A literal string within double quotes. Examples: "dynamic data exchange" or "dde "
- Multiple words separated by a plus (+). The results must contain all words. Example: clinical + metadata
- Multiple words separated by blanks. The results must contain at least one of the words. Example: report tabulate

TITLE contains

AND AUTHOR contains

Search Sort by: ☒ Date ☐ Title

Maximum number of results returned: 800

A word cloud of names, where the size of each name corresponds to its frequency. The names are arranged in a dense, overlapping pattern. The most prominent names, shown in the largest font, include Carpenter, Chen, Wang, Smith, and Brown. Other large names include Ross, Langston, Hadden, and Bessler. The names are in various colors, including shades of blue, green, yellow, and red, and are oriented in different directions, creating a dynamic and visually complex composition.

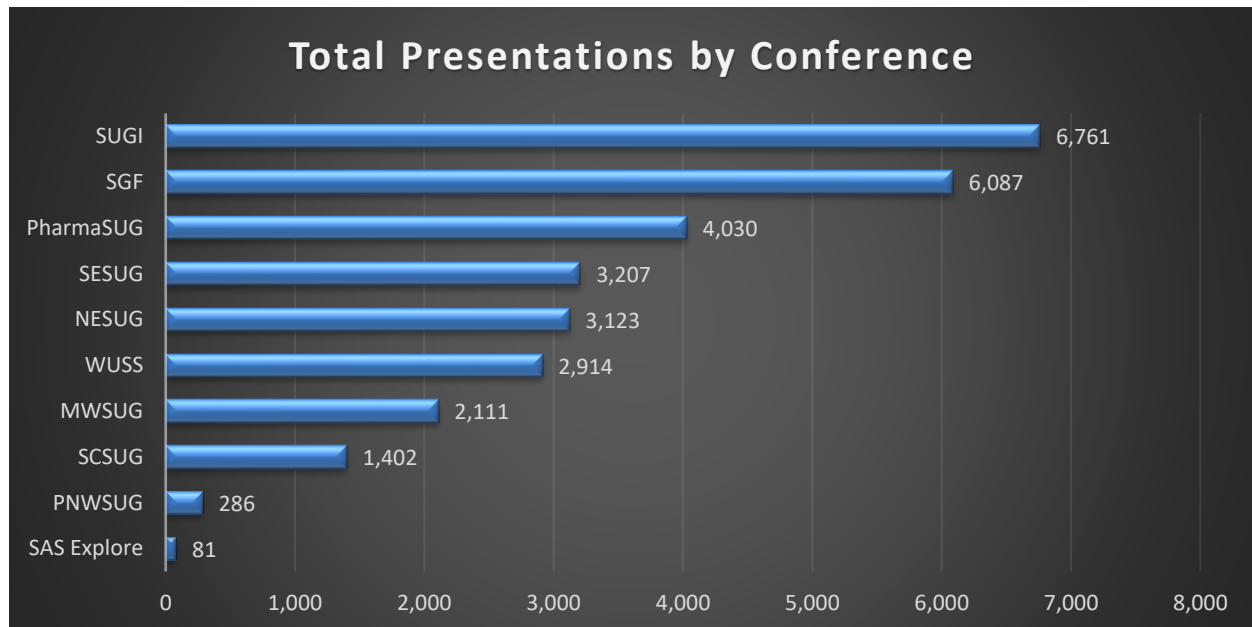
The following conferences are included in the analysis, and the respective index page (on the LEX) for each is provided:

- SAS Users Group International (SUGI) 1976 to 2006 – <https://www.lexjansen.com/sugi/>
- SAS Global Forum (SGF) 2007 to 2021 – <https://www.lexjansen.com/sugi/>
- SAS Explore 2022 – <https://www.lexjansen.com/sasexplore/>
- Midwest SAS Users Group (MWSUG) 1990 to 2019 – <https://www.lexjansen.com/mwsug/>
- Northeast SAS Users Group (NESUG) 1998 to 2013 – <https://www.lexjansen.com/nesug/>
- Pharmaceutical Software Users Group (PharmaSUG) 1997 to 2023 – <https://www.lexjansen.com/pharmasug/>
- Pacific Northwest SAS Users Group (PNWSUG) 1996 to 2009 – <https://www.lexjansen.com/pnwsug/>
- South Central SAS Users Group (SCSUG) 1991 to 2019 – <https://www.lexjansen.com/scsug/>

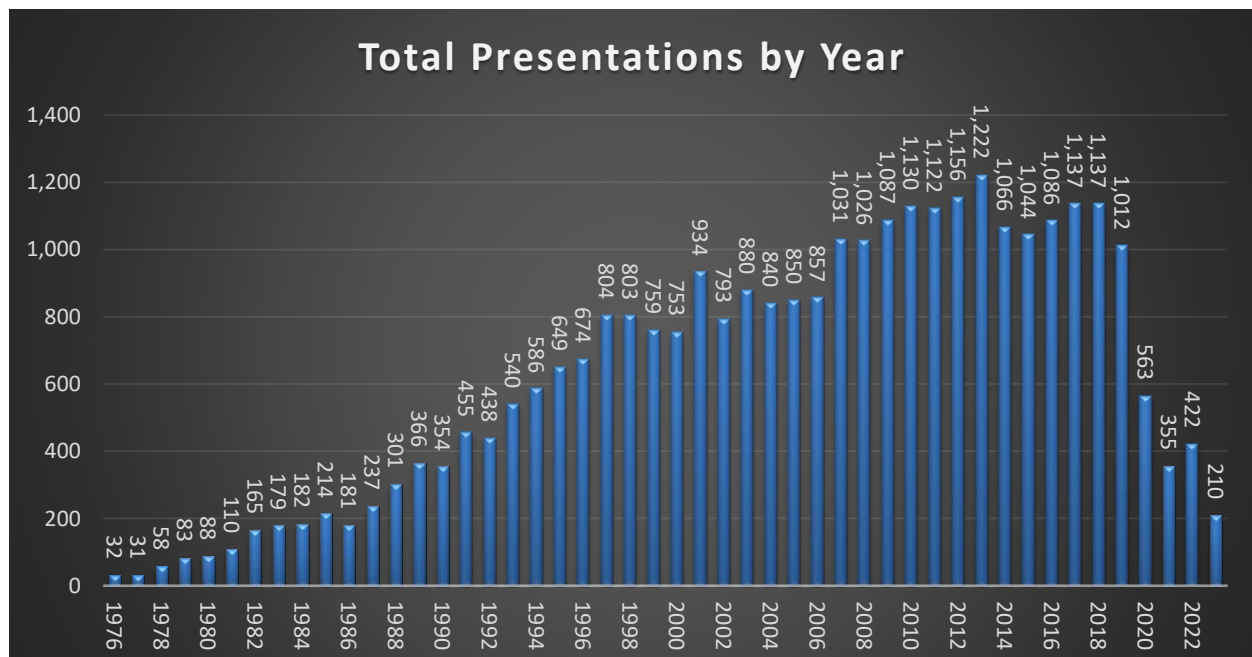
- Southeast SAS Users Group (SESUG) 1993 to 2022 – <https://www.lexjansen.com/sesug/>
- Western Users of SAS Software (WUSS) 1993 to 2022 – <https://www.lexjansen.com/wuss/>

Note that as this analysis was performed during the summer of 2023, the fall user conferences—including SAS Explore 2023, WUSS 2023, and SESUG 2023—are not included. Also note that the contents of SAS Analytics Experience (held by SAS Institute in 2016, 2017, and 2018) are not included; these presentations are neither maintained on the LEX nor on the SAS website, and a request to SAS Institute for the content was denied, with the heartbreaking explanation that the data have not been archived.

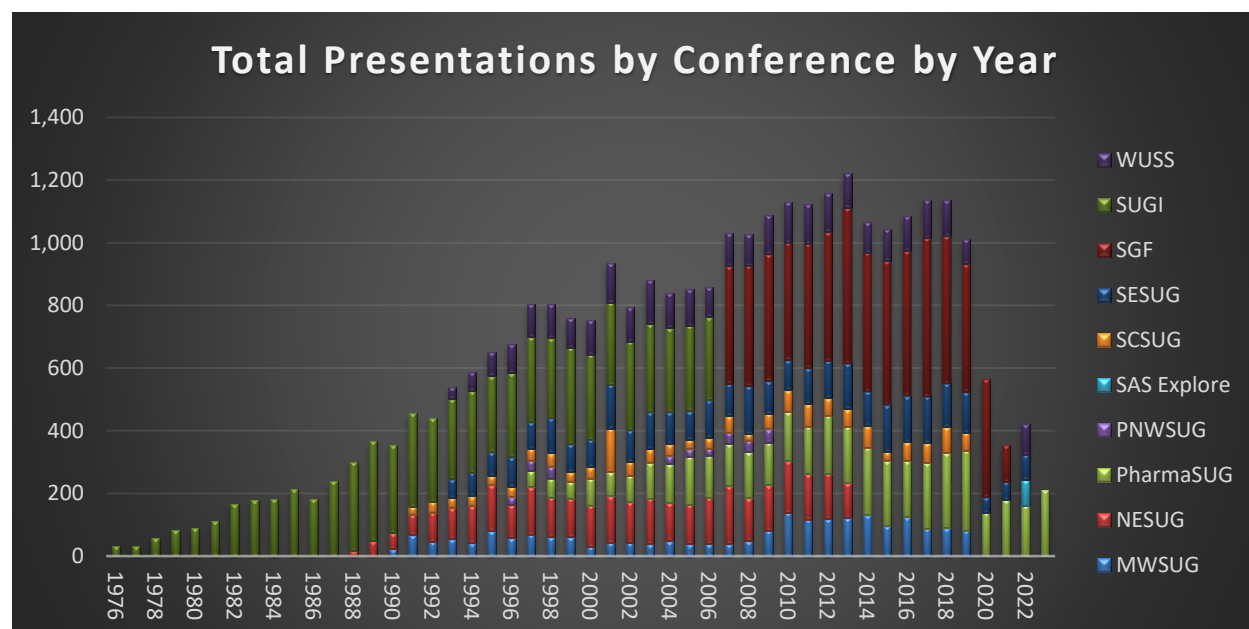
SUGI (N=6,761) and SGF (N=6,087) have had the largest number of presentations over the course of their lifetimes, with SUGI terminating in 2006 and being rebranded as SGF in 2007.



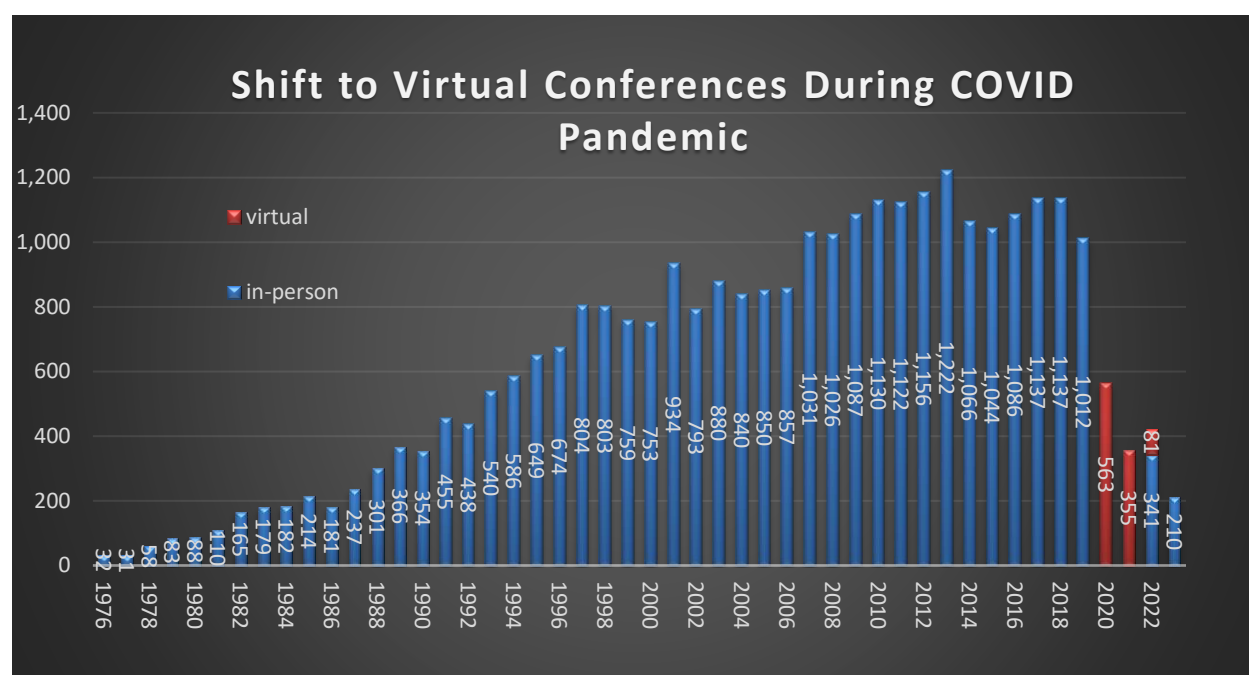
The combined SAS conferences offered the most presentations (N=1,222) in 2013. Thus, I joined the SAS conference scene at its apex, having attended my first conference (SESUG) in 2013 in St. Pete's, Florida.



Unfortunately, 2013 was also the last year of NESUG, which announced at its user conference that it was shuttering due, allegedly, to a catastrophic decrease in funding from SAS Institute. Thus, in 2014, and with the cessation of NESUG, total conference presentations dropped to 1,066.



The seemingly precipitous crash in 2020 (from 1,012 to 563 presentations) resulted from the COVID pandemic, and prompted a two-year shift from in-person to virtual events. For example, SGF—historically a spring event—was canceled only one month before its scheduled opening, with some content posted as papers and some content filmed by users and posted to YouTube. PharmaSUG, also a spring conference, was canceled the following month, and papers were posted online without presentations.



Again, note that the paucity of presentations (N=210) in 2023 results, in part, from only the PharmaSUG 2023 conference being included, as SAS Explore, WUSS, and SESUG fall events had not yet occurred. Unfortunately, however, two of the RUGs (MWSUG and SCSUG) have not returned since 2019, although MWSUG is planning a conference in 2024, and talks are underway regarding a possible return of SCSUG.

METHODOLOGY

The first step in this analysis was the identification of all conference URLs on which presentations are listed (and white papers, where available, could be downloaded). The Python BeautifulSoup package was utilized to scrape the following conference URLs (maintained as a Python list), and to determine during which years each conference was held:

```
conf_url_list = ['https://www.lexjansen.com/mwsug/',
                 'https://www.lexjansen.com/nesug/',
                 'https://www.lexjansen.com/pharmasug/',
                 'https://www.lexjansen.com/pnwsug/',
                 'https://www.lexjansen.com/scsug/',
                 'https://www.lexjansen.com/sesug/',
                 'https://www.lexjansen.com/wuss/',
                 'https://www.lexjansen.com/sugi/',
                 'https://www.lexjansen.com/sasexplore/']
```

Thereafter, a Pandas DataFrame was created to maintain the conference name, conference year, conference location (city and state), and conference URL (on the LEX). In total, the proceedings of 226 separate conferences were included in this analysis, 36 of which I have attended.

Next, the DataFrame was iterated to identify each conference URL, and BeautifulSoup again was leveraged to scrape presentation-specific metadata from each page, including:

- **Section** – In some cases, a section such as “e-Posters” or “Hands-on Workshops” denotes not the content but rather format of the presentation. In other cases, a section such as “Advanced Techniques” or “Statistics” is intended to reflect the content of those presentations.
- **Presentation Title** – This usually, but not always, matches an accompanying white paper or PowerPoint presentation.
- **Presenter(s)** – One or more presenters are listed.
- **Affiliation(s)** – At some conferences, and in some years, the affiliation (e.g., school, employer) of each presenter is listed.
- **Presentation ID** – Each presentation has a unique ID referenced as the “ID” tag in the accompanying HTML markup.
- **Presentation URL** – As mentioned, approximately 96.5% of all presentations have an accompanying white paper that can be downloaded as a PDF.

At this point, separate DataFrames were produced, including one having one row for each of the 30,002 presentations, and one having one row for each of the 14,918 unique authors. Thus, the author-level DataFrame decouples authors and coauthors, and facilitates author-specific analysis.

OVERCOMING DATA QUALITY ISSUES

The LEX is an incredible amalgam of data—ahem, rather *third-party* data, as in dirty data dumped on Lex’s doorstep by conference organizers more focused on the relief that their conference has concluded than on cleaning conference metadata. Owing to this dumpage, some presenter names are misspelled, coauthors are missing, titles are incorrect, or HTML markup codes glaringly appear in names, affiliations, and titles. This doesn’t render the LEX any less incredible, although it does illustrate that more effort could be made by conference organizers to supply *correct* metadata to ensure that authors are appropriately credited and information is conveyed.

Invalid data aside, the formatting of presentation metadata can also differ substantially among conferences. For example, a comma may delimit authors where multiple presenters exist. However, commas can also be used to separate a presenter’s name from his degree, affiliation, or location. Moreover, over time,

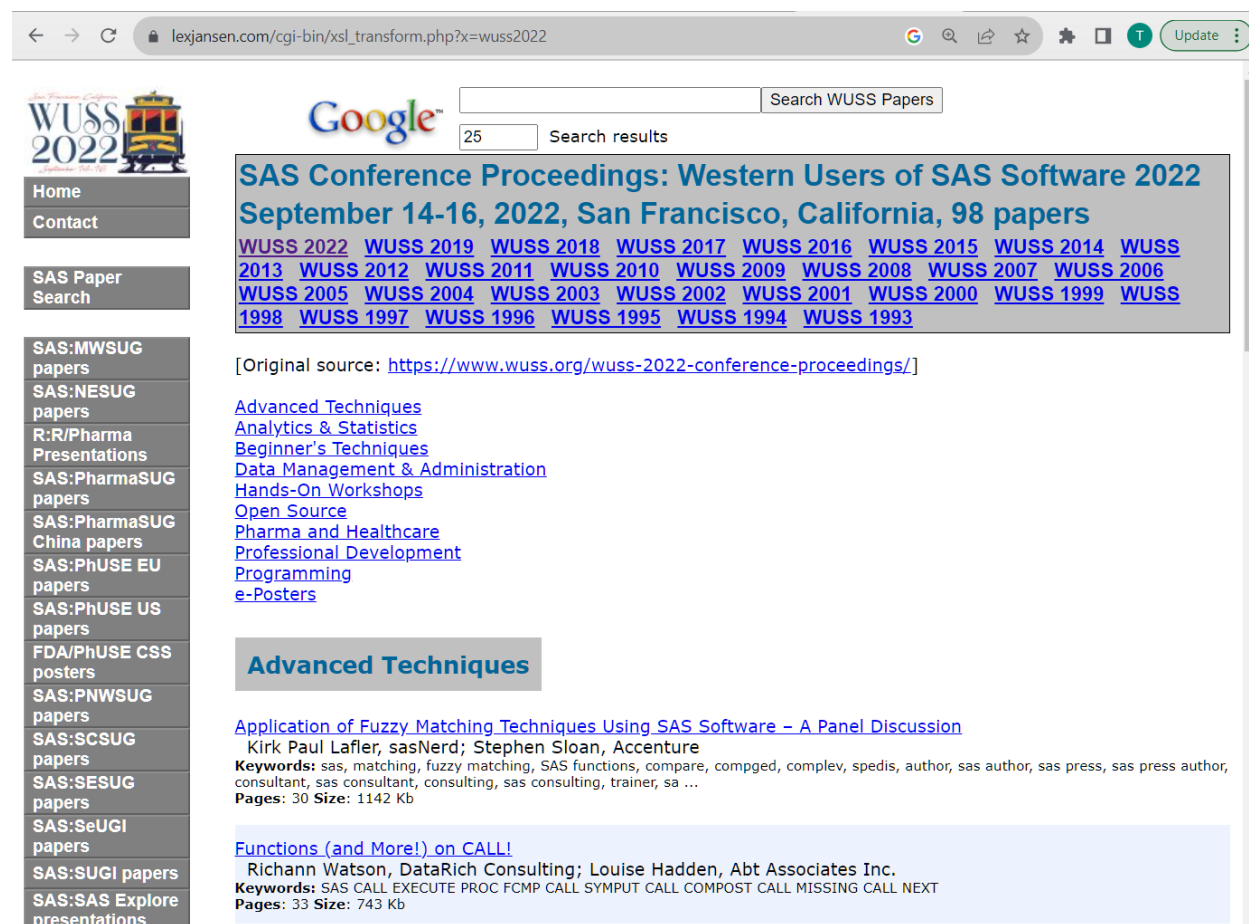
conferences have changed how they record presentation data, and thus what data they transmit to the LEX. This is all to say that a substantial amount of code (not shown) was required to identify metadata formatting, and to read metadata into the correct fields.

Finally, a third data quality issue involves the standardization of presenter names. Again, because this analysis aims to identify and credit those most prolific presenters, it was essential to entity resolve individuals so that their contributions could be appropriately quantified. A review of all authors for all papers identified more than 2,000 presenter names that required transformation to a standardized version of their name. For example, all instances of “Troy Hughes” were standardized to “Troy Martin Hughes.”

These three aspects of data quality—including metadata identification, metadata correction, and metadata standardization (i.e., entity resolution)—are discussed in the next three subsections.

DATA QUALITY – 1. METADATA IDENTIFICATION

In a perfect world, data would be presented with standardized formatting, including both what data are displayed and how those data are identified or delimited. And, in most cases, each conference page of the LEX—that is, a single conference for a single year, like WUSS 2022—is standardized. For example, the header for the WUSS 2022 conference, as well as the first two papers, appears as the following:



The screenshot shows a web browser displaying the WUSS 2022 conference page. The browser's address bar shows the URL: `lexjansen.com/cgi-bin/xsl_transform.php?x=wuss2022`. The page features a Google search bar at the top with the text "Search WUSS Papers". Below the search bar, the main heading reads "SAS Conference Proceedings: Western Users of SAS Software 2022 September 14-16, 2022, San Francisco, California, 98 papers". A list of links for various WUSS years (2022, 2019, 2018, 2017, 2016, 2015, 2014, 2013, 2012, 2011, 2010, 2009, 2008, 2007, 2006, 2005, 2004, 2003, 2002, 2001, 2000, 1999, 1998, 1997, 1996, 1995, 1994, 1993) is provided. On the left side, there is a sidebar with a "Home" and "Contact" link, and a "SAS Paper Search" section. Below this, a list of paper categories is shown: SAS:MWSUG papers, SAS:NESUG papers, R:R/Pharma Presentations, SAS:PharmaSUG papers, SAS:PharmaSUG China papers, SAS:PhUSE EU papers, SAS:PhUSE US papers, FDA/PhUSE CSS posters, SAS:PNWSUG papers, SAS:SCSUG papers, SAS:SESUG papers, SAS:SeUGI papers, SAS:SUGI papers, and SAS:SAS Explore presentations. The main content area displays the "Advanced Techniques" section, which includes a link to "Application of Fuzzy Matching Techniques Using SAS Software – A Panel Discussion" by Kirk Paul Lafler, sasNerd; Stephen Sloan, Accenture. The keywords for this paper are: sas, matching, fuzzy matching, SAS functions, compare, compged, complev, spedis, author, sas author, sas press, sas press author, consultant, sas consultant, consulting, sas consulting, trainer, sa ... The page size is 30 and the size is 1142 Kb. Below this, the "Functions (and More!) on CALL!" section is shown, which includes a link to "Functions (and More!) on CALL!" by Richann Watson, DataRich Consulting; Louise Hadden, Abt Associates Inc. The keywords for this paper are: SAS CALL EXECUTE PROC FCMP CALL SYMPUT CALL COMPOST CALL MISSING CALL NEXT. The page size is 33 and the size is 743 Kb.

Digging under the covers, the HTML markup for the first paper appears as the following:

```
<div xmlns:gcse="uri:dummy-google-ns" class="paper wh"><a
id="wuss2022.038.at"></a><a class="nofloatbox" target="_blank"
href="https://www.lexjansen.com/wuss/2022/WUSS-2022-Paper-38.pdf">Application of
Fuzzy Matching Techniques Using SAS Software – A Panel Discussion</a><br />Kirk
Paul Lafler, sasNerd; Stephen Sloan, Accenture<br /><span
```

```

class="key"><b>Keywords:</b> sas, matching, fuzzy matching, SAS functions, compare,
compged, complev, spedis, author, sas author, sas press, sas press author,
consultant, sas consultant, consulting, sas consulting, trainer, sa ...</span><br
/><span class="size"><b>Pages</b>: 30 </span><span class="size"><b>Size</b>: 1142
Kb </span></div>

```

Note that presenters are delimited by a semicolon, and that each presenter and his affiliation are delimited by a comma. However, commas can occur in either a presenter's name or an affiliation.

For example, Cleester Heath, IV, has a comma in his name:

```

<div xmlns:gcse="uri:dummy-google-ns" class="paper wh"><a id="wuss2010.3034_2_dpr-
heath1.dp"></a><a class="nofloatbox" target="_blank"
href="https://www.lexjansen.com/wuss/2010/DataPresentation/3034_2_DPR-
Heath1.pdf">Creating Presentation-Quality ODS Graphics Output</a><br /> Cleester
Heath, IV<br /><span class="size"><b>Pages</b>: 13 </span><span
class="size"><b>Size</b>: 1068 Kb </span></div>

```

This also occurs when you're a junior:

```

<div xmlns:gcse="uri:dummy-google-ns" class="paper wh"><a id="wuss2009.app-
benjamin.ad"></a><a class="nofloatbox" target="_blank"
href="https://www.lexjansen.com/wuss/2009/app/APP-Benjamin.pdf">The Little Engine
That Could: Using EXCEL LIBNAME Engine Options to Enhance Data Transfers between
SAS® and Microsoft® Excel Files</a><br />William E Benjamin, Jr., Owl Computer
Consultancy, LLC<br /><span class="size"><b>Pages</b>: 10 </span><span
class="size"><b>Size</b>: 873 Kb </span></div>

```

And some presenters list their degree after their name, delimited by a comma:

```

<div xmlns:gcse="uri:dummy-google-ns" class="paper bl"><a
id="wuss2004.c_das_empirical_investigatio.st"></a><a class="nofloatbox"
target="_blank"
href="https://www.lexjansen.com/wuss/2004/data_analysis/c_das_empirical_investigati
o.pdf">Empirical Investigation of Time Series Frequency Conversion with PROC
EXPAND</a><br />Keiko I. Powers, Ph.D., Power Information Network, Westlake
Village, CA<br /><span class="key"><b>Keywords:</b> PROC EXPAND ARIMA time series
analysis </span><br /><span class="size"><b>Pages</b>: 9 </span><span
class="size"><b>Size</b>: 170 Kb </span></div>

```

The affiliation can also contain a comma:

```

<div xmlns:gcse="uri:dummy-google-ns" class="paper wh"><a
id="wuss2022.103.hw"></a>ODS Document &amp; Item Stores: A New Beginning<br />Bill
Coar, Axio, a Cytel Company<br /></div>

```

Here, the coauthor was listed parenthetically after the primary author:

```

<div xmlns:gcse="uri:dummy-google-ns" class="paper wh"><a
id="wuss2011.papers_ottesen_r_74966.rs"></a><span class="hilite">Best Contributed
Paper</span><br /><a class="nofloatbox" target="_blank"
href="https://www.lexjansen.com/wuss/2011/res/Papers_Ottesen_R_74966.pdf">This is
how we do it: teaching SAS® in the Cal Poly Statistics Department</a><br />Rebecca

```

Ottesen (Jimmy Doi), City of Hope / Cal Poly SLO
 Pages: 12
 Size: 60 Kb

Once upon a time, some conferences delimited presenters using commas:

```
<div xmlns:gcse="uri:dummy-google-ns" class="paper wh"><a
id="scsug2019.rs09"></a>Using INFILE and INPUT Statements to Introduce External
Data into the SAS® System<br />Jimmy DeFoor, Andrew Kuligowski<br /></div>
```

In some cases, a semicolon delimits the author and the affiliation:

```
<div xmlns:gcse="uri:dummy-google-ns" class="paper wh"><a
id="wuss2005.ess_iob_security_using.se"></a><a class="nofloatbox" target="_blank"
href="https://www.lexjansen.com/wuss/2005/sas_essentials/ess_iob_security_using.pdf"
">Job Security: Using the SAS® Macro Language to Full Advantage</a><br />Arthur L.
Carpenter; California Occidental Consultants, Oceanside, California<br /><span
class="size"><b>Pages</b>: 8 </span><span class="size"><b>Size</b>: 66 Kb
</span></div>
```

Starting with SUGI 21 (i.e., 1996), presenter names are linked to presenter pages that show all of their presentations:

```
<div class="paper wh"><a id="sugi21.026-21"></a><a class="nofloatbox"
target="_blank" href="https://www.lexjansen.com/sugi/sugi21/at/026-21.pdf">The
Secret Life of the DATA Step</a><br /><a
href="/cgi-bin/xsl_transform.php?x=am&c=SUGI#tommiron">Thomas Miron</a>
<br /><span class="size"><b>Pages</b>: 7 </span><span
xmlns:gcse="uri:dummy-google-ns" class="size"><b>Size</b>: 244 Kb </span></div>
```

Starting with SGF 2014, paper titles are preceded by paper IDs:

```
<div class="paper wh"><a id="sgf2014.1558-2014"></a><span class="code">1558:
</span><a class="nofloatbox" target="_blank"
href="https://support.sas.com/resources/papers/proceedings14/1558-
2014.pdf">%COVTEST: A SAS® Macro for Hypothesis Testing in Linear Mixed Effects
Models via Parametric Bootstrap</a><br /><a
href="/cgi-bin/xsl_transform.php?x=ao&c=SUGI#peterott">Peter Ott</a>
<br /><span class="key"><b>Keywords:</b> <a href="/cgi-
bin/xsl_transform.php?x=tmac&s=SUGI_t&c=SUGI">Macro</a> </span><br /><span
class="size"><b>Pages</b>: 9 </span><span xmlns:gcse="uri:dummy-google-ns"
class="size"><b>Size</b>: 328 Kb </span></div>
```

Sometimes, the author is preceded by "Presented by":

```
<div xmlns:gcse="uri:dummy-google-ns" class="paper wh"><a
id="nesug2006.ap11"></a><a class="nofloatbox" target="_blank"
href="https://www.lexjansen.com/nesug/nesug06/ap/ap11.pdf">How Do I Love Hash
Tables? Let Me Count The Ways!</a><br />Judy Loren, Independent Consultant;
Presented by Alan Dickson, Independent Consultant<br /><span
```

```
class="size"><b>Pages</b>: 14 </span><span class="size"><b>Size</b>: 141 Kb
</span></div>
```

Sometimes the author is listed only as an organization, such as “SAS Institute Staff” in the following:

```
<div xmlns:gcse="uri:dummy-google-ns" class="paper wh"><a
id="wuss2000.061.ii"></a><a class="nofloatbox" target="_blank"
href="https://www.lexjansen.com/wuss/2000/WUSS00061.pdf">Patterns for Producing
Web-Enabled Applications</a><br /><b>SAS Institute Staff</b>, SAS Institute Inc.<br
/><span class="size"><b>Pages</b>: 10 </span><span class="size"><b>Size</b>: 713 Kb
</span></div>
```

Thus, the first step in this analysis was the identification of metadata so that specific metadata could be attributed to the correct variables. This included building various templates, as metadata structure, formatting, and content differed across conferences and years.

DATA QUALITY – 2. METADATA CORRECTION

After accounting for all the variations across conferences and across the years (some of which are shown in the previous subsection), there were still clear errors in the representation of some metadata. In some cases, malformed HTML resulted in escape codes like “&” being displayed in lieu of an ampersand symbol. In other cases, delimiters were missing, incorrect, or duplicated. Finally, many errors or ambiguities were discovered only through manual comparison of the header of the accompanying white paper to its metadata representation on the LEX. For example, the LEX might list a single author for a presentation, despite the accompanying white paper listing four authors.

Here, the semicolon is misplaced (between the first author and her affiliation), and the second author is concatenated incorrectly:

```
<div xmlns:gcse="uri:dummy-google-ns" class="paper bl"><a
id="wuss2014.052.qt"></a><a class="nofloatbox" target="_blank"
href="https://www.lexjansen.com/wuss/2014/52_Final_Paper_PDF.pdf">It's Simpler than
You Think: An example of a Patient Summary Report generated using SAS and ODS
PDF</a><br /><b>Julie Kilburn; City of HopeLeanne Goldstein, City of Hope</b><br /><span
class="size"><b>Pages</b>: 8 </span><span class="size"><b>Size</b>: 697 Kb
</span></div>
```

Sometimes an ampersand in the title is shown as an escape code:

```
<div xmlns:gcse="uri:dummy-google-ns" class="paper wh"><a
id="wuss2022.103.hw"></a><b>ODS Document & Item Stores: A New Beginning</b><br />Bill
Coar, Axio, a Cytel Company<br /></div>
```

Sometimes an extra semicolon is included:

```
<div class="paper wh" xmlns:gcse="uri:dummy-google-ns"><a
id="pharmasug2017.ib03"></a><span class="code">IB03: </span><a class="nofloatbox"
href="https://www.lexjansen.com/pharmasug/2017/IB/PharmaSUG-2017-IB03.pdf"
target="_blank">How to define Treatment Emergent Adverse Event (TEAE) in crossover
clinical trials</a><br /><b>Mengya Yin, Ultragenyx Pharmaceutical;; Wen Tan, Ultragenyx
Pharmaceutical</b><br /><span class="key"><b>Keywords:</b> 'Treatment Emergent
Adverse Event' TEAE Crossover
```

```

'Washout Periods'      'Severity Change'      'Partial Dates'
</span><br/><span class="size"><b>Pages</b>: 6 </span><span
class="size"><b>Size</b>: 421 Kb </span></div>

```

Similarly, an ampersand in the affiliation might be shown as an escape code:

```

<div xmlns:gcse="uri:dummy-google-ns" class="paper wh"><a
id="wuss2015.014.ad"></a><a class="nofloatbox" target="_blank"
href="https://www.lexjansen.com/wuss/2015/14_Final_Paper_PDF.pdf">A Risk Factor
Coding Custom Report Created With ODS Report Writing Interface With Recent
Literature and Peer Review</a><br />Robert Springborn, Office of Statewide Health
Planning &amp; Development<br /><span class="size"><b>Pages</b>: 13 </span><span
class="size"><b>Size</b>: 469 Kb </span></div>

```

Sometimes the “&” escape code was malformed (e.g., into “&amp;”):

```

<div xmlns:gcse="uri:dummy-google-ns" class="paper wh"><a
id="wuss2011.papers_liu_x_75059.cc"></a><a class="nofloatbox" target="_blank"
href="https://www.lexjansen.com/wuss/2011/coders/Papers_Liu_X_75059.pdf">Producing
Cumulative Distribution Frequency Figures with Minimal Important Difference
Reference Lines to Assess Quality of Life Treatment Effects</a><br />Xiao Liu, ICON
Late Phase &amp;amp; Outcomes Research<br /><span class="size"><b>Pages</b>: 7
</span><span class="size"><b>Size</b>: 156 Kb </span></div>

```

Sometimes names clearly had been concatenated such as “JohnM. [sic] Castelloe”:

```

<div xmlns:gcse="uri:dummy-google-ns" class="paper wh"><a
id="pharmasug2001.sp16"></a><span class="code">SP16: </span><a class="nofloatbox"
target="_blank"
href="https://www.lexjansen.com/pharmasug/2001/Proceed/Stats/sp16_castelloe.pdf">Po
wer and Sample Size Determination for Linear Models</a><br />
JohnM.Castelloe, SASInstitute Inc., Cary,NC; Ralph G. O'Brien, Cleveland
Clinic Foundation, Cleveland, OH<br /><span class="key"><b>Keywords:</b>
power analysis anova regression mixed models statistics
</span><br /><span class="size"><b>Pages</b>: 14 </span><span
class="size"><b>Size</b>: 577 Kb </span></div>

```

In some cases, a white paper that had been digitally scanned from a hard copy was interpreted incorrectly, such as “Kattamuri” being interpreted in two instances as “Kattamurl”—that is, an “l” being parsed as an “L”:

```

<div xmlns:gcse="uri:dummy-google-ns" class="paper wh"><a
id="wuss2001.037.st"></a><a class="nofloatbox" target="_blank"
href="https://www.lexjansen.com/wuss/2001/WUSS01037.pdf">Using SAS Enterprise Miner
for Forecasting Response and Risk</a><br />
Kattamurl S. Sarma, Ecostat Research Corp.<br /><span
class="size"><b>Pages</b>: 10 </span><span class="size"><b>Size</b>: 1103 Kb
</span></div>

```

In some cases, a degree like “Ph.D.” had been substituted for a name:

```
<div class="paper wh"><a id="sgf2009.416-2009"></a><a class="nofloatbox"
target="_blank" href="https://support.sas.com/resources/papers/proceedings09/416-
2009.pdf">Using Integrated Data Systems to Maximize Public Resources</a><br /><a href="/cgi-
bin/xsl_transform.php?x=am&c=SUGI#phdmreno">Ph.D. Moreno</a>
<br /><span class="size"><b>Pages</b>: 1 </span><span
xmlns:gcse="uri:dummy-google-ns" class="size"><b>Size</b>: 17 Kb </span>
```

In all of these cases, where errors are visibly salient on the LEX itself, errors were communicated to Lex Jansen, and he graciously modified dozens of entries for which he had been supplied invalid metadata from conferences.

In other cases, however, only through inspection of an associated white paper could a discrepancy or other error be detected. The most common example occurs when only a single presenter is listed in the metadata, whereas the accompanying white paper has two or more authors. In some cases, this discrepancy was likely intentional—for example, when one individual had authored a paper, but a different individual (perhaps a coworker) presented the paper.

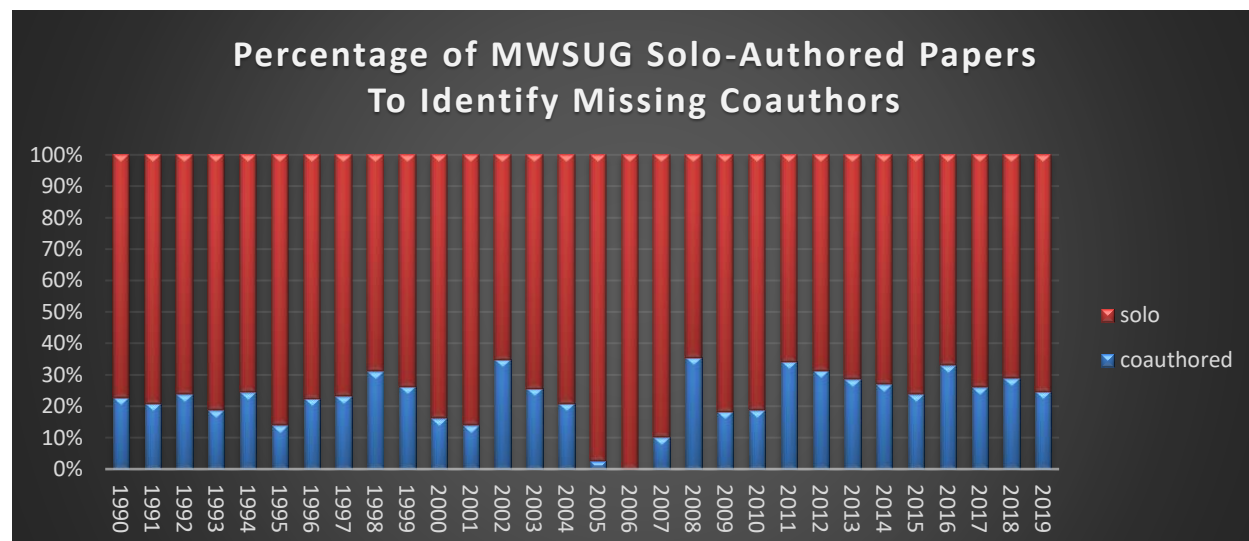
Consider the following entry for a paper authored only by Toby Dunn, yet presented only by Chang Chung:

```
<div xmlns:gcse="uri:dummy-google-ns" class="paper bl"><a
id="nesug2006.cc21"></a><a class="nofloatbox" target="_blank"
href="https://www.lexjansen.com/nesug/nesug06/cc/cc21.pdf">Handling Dates in the
Macro Facility</a><br />
Toby Dunn, Eagle Group Int.; Presented by Chang Chung, Princeton University<br
/><span class="key"><b>Keywords:</b>
<a href="/cgi-bin/xsl_transform.php?x=tmac&s=NESUG_t&c=NESUG">Macro</a>
</span><br /><span class="size"><b>Pages</b>: 4 </span><span
class="size"><b>Size</b>: 24 Kb </span></div>
```

When these instances were detected, the author name(s) were utilized in lieu of the presenter(s). This methodology ensures that authors receive credit, as writing a fifteen page paper is inarguably more taxing and time consuming than speaking about the paper for fifteen minutes!

Thus, the most commonly detected metadata issue appeared to be the truncation of coauthors, in which only the first author was listed in the LEX metadata. Although this occurred sporadically, some conference years were more prone to this issue, and these outliers could be identified statistically and visually.

For example, the following graph demonstrates that in 2005, MWSUG had only one coauthored paper (out of 38 total papers), and no coauthored papers in 2006 (out of 39 total presentations).

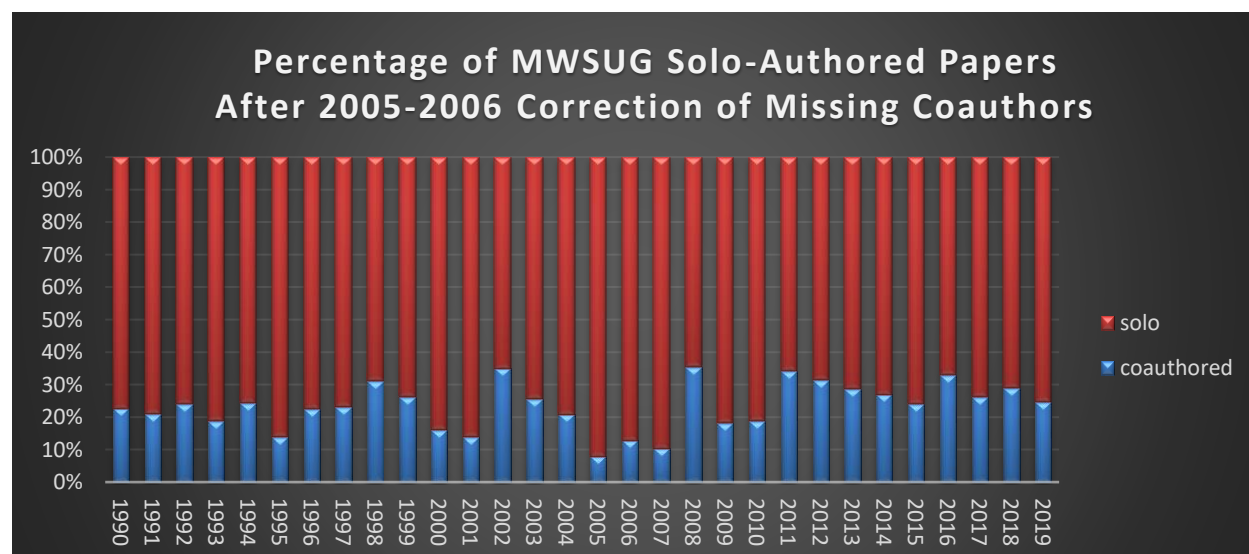


After manual review of all MWSUG 2005 and 2006 papers, an additional two coauthored papers were discovered in 2005, and five coauthored papers in 2006:

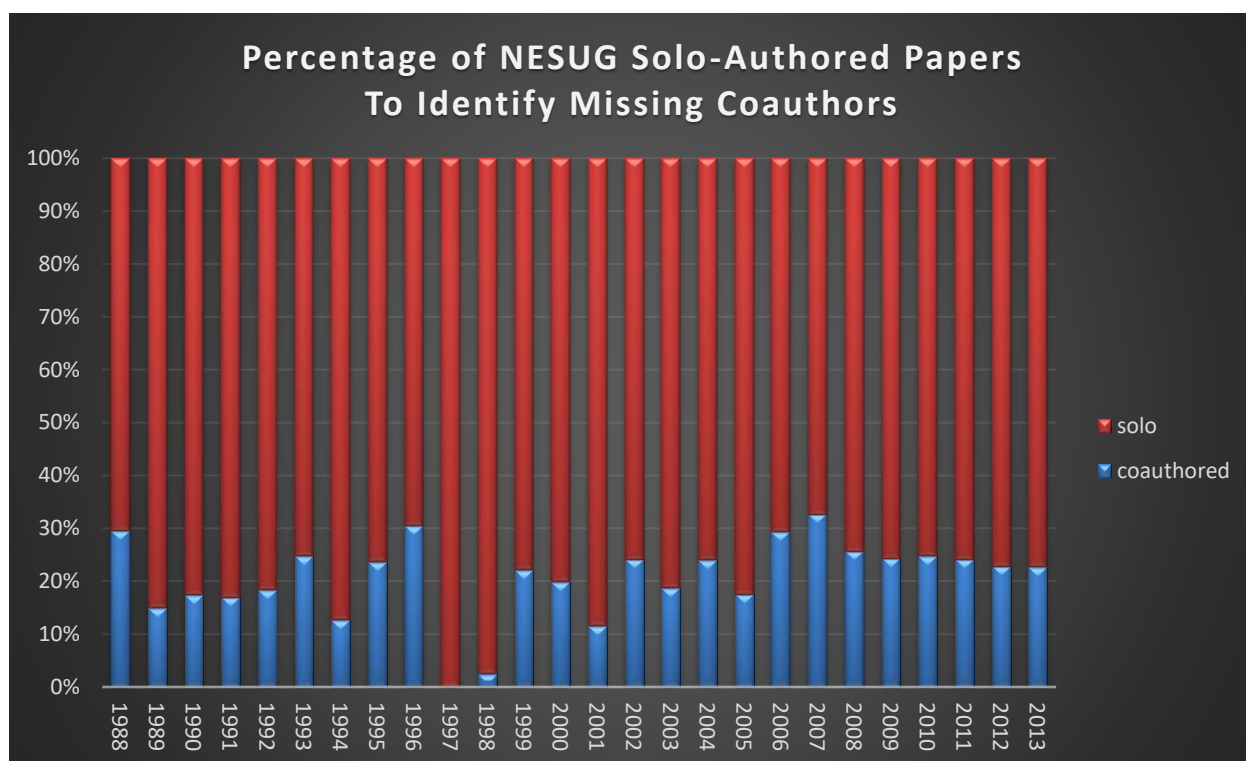
paper_id	raw_author_affil	paper_authors
mwsug2005.dm200	Michael Leonard, SAS Institute	['Michael Leonard','Brenda Wolfe']
mwsug2005.ph600	Padmini Sekar, University of Cincinnati	['Padmini Sekar','Matthew Flaherty','Daniel Woo','Ralph Buncher']
mwsug2006.ad08	Joe Murphy, RTI International	['Joe Murphy','Adam Safir','Susan Myers','Christy Hottinger']
mwsug2006.cc04	Cindy Liu, Wells Fargo	['Cindy Liu','Dongmei Lan']
mwsug2006.cc05	Hangtao Xu, Eli Lilly and Company	['Hangtao Xu','Mario Widel','Jyoti Rayamajhi']
mwsug2006.sd05	Matt P Shevrin, Veterans Affairs HSR&D	['Matthew P. Shevrin','Maria J. Silveira','Anamaria Segnini Kazanis']
mwsug2006.sd10	Qixuan Chen, The Department of Biostatistics	['Qixuan Chen','Brenda Gillespie']

To be clear, not all conferences and not all years were reviewed, but where noticeable drops like this were identified, a manual review of all papers for that conference occurred.

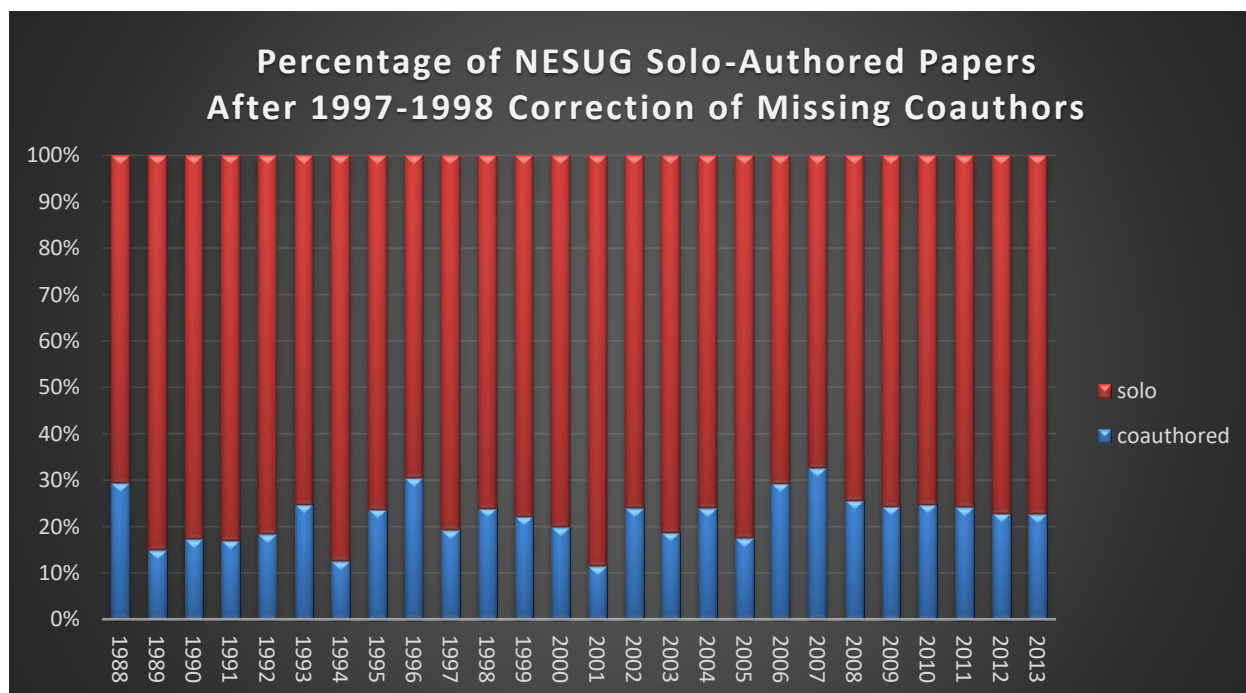
After the addition of coauthors, the updated MWSUG graph is shown:



A graph of NESUG demonstrates the LEX is missing coauthors in 1997 and 1998:



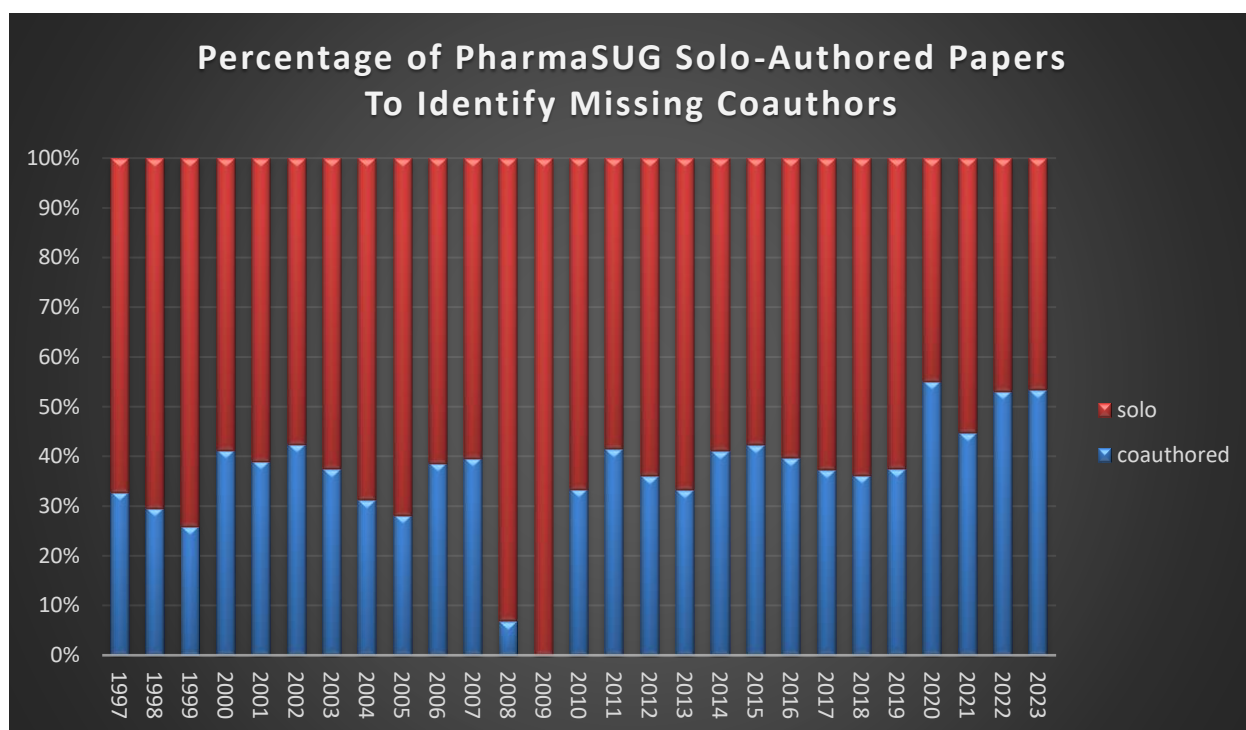
Upon inspection, 56 NESUG articles required revision to add missing coauthors. After these revisions, the updated NESUG graph lacks these outliers:



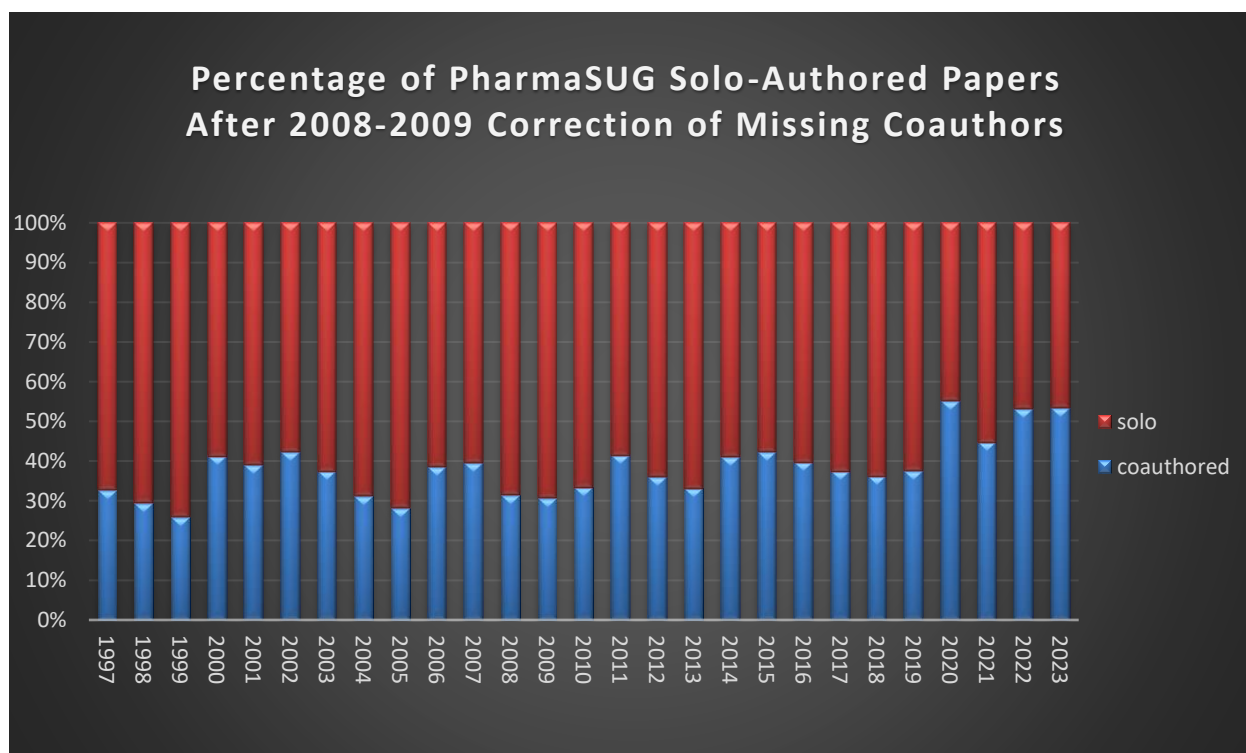
The 56 NESUG papers that were missing one or more coauthors include:

paper_id	raw_author_affil	paper_authors
nesug1997.at_gerlach	John Gerlach	['John R. Gerlach','Naoko S. Stearns']
nesug1997.at_gober	John Charles Gober	['Diana Zhang Wobus','John Charles Gober']
nesug1997.at_loren	Judy Loren	['Judy Loren','Gregory S. Barnes Nelson']
nesug1997.at_ma	Meimei Ma	['Juliana Mei-Mei Ma','Andrew H. Karp']
nesug1997.at_nelsen	Linda Nelsen	['Linda Nelsen','Margaret M. Coughlin']
nesug1997.at_wilson	Alan Wilson	['Alan C. Wilson','Marge Scerbo']
nesug1997.ad_elpas	Krista Elspas	['Krista A. Elspas','Julie Shadoan']
nesug1997.ad_roper	Christopher Roper	['Christopher A. Roper','Michael Gilman']
nesug1997.bt_dickson1	Alan Dickson	['Alan Dickson','Ray Pass']
nesug1997.cc_asam	Ellen Asam	['Ellen L. Asam','Donna Usavage']
nesug1997.db_burrowes	Philip Burrowes	['Philip Burrowes','Jason Oliveira']
nesug1997.db_kallukar	Paul Kallukaran	['Paul Kallukaran','Jerry Kagan']
nesug1997.db_lafler	Kirk Paul Lafler	['Kirk Paul Lafler','Sunil Kumar Gupta','Charles Edwin Shipp']
nesug1997.db_rogers	Greg Rogers	['Greg Rogers','Ellen Joyner']
nesug1997.iv_elkin	Steven Elkin	['Steven E. Elkin','William Mietlowski','Kevin McCague','Andrea Kay']
nesug1997.ii_lefebvre	Eric Lefebvre	['Eric J. Lefebvre','Thomas Kunselman']
nesug1997.po_brusilov	Pavel Brusilovskiy	['Pavel Brusilovskiy','Robert Hernandez']
nesug1997.po_ching	Eileen Ching	['Eileen Ching','Rosemary Oakes']
nesug1997.po_goetz	Margaret Ann Goetz	['Margaret Ann Goetz','Grace Chang']
nesug1997.po_rothwell	Sandra Rothwell	['Sandra T. Rothwell','Mary Ann Bush','Ilene Gottfried','Dawn M. Scott']
nesug1997.po_sun	James Sun	['James Q. Sun','Connie X. Li']
nesug1997.so_stuelpner	Janet Stuelpner	['Janet E. Stuelpner','Elizabeth Kaptanov']
nesug1997.st_blanchard	Suzette Blanchard	['M.S. Blanchard','Peter P. Gaccione']
nesug1997.st_daponte	John DaPonte	['John S. DaPonte','Joseph N. Vitale','George Tselioudis','William Rossow']
nesug1997.st_garavaglia	Susan Garavaglia	['Susan B. Garavaglia','Asha Sharma']
nesug1997.st_kotler	Mitchell Kotler	['Mitchell Kotler','Larry L. Laster']
nesug1997.st_shtatland	Ernest Shtatland	['Ernest S. Shtatland','Mary B. Barton']
nesug1997.st_stokes	Maura Stokes	['Gordon Johnston','Maura E. Stokes']
nesug1997.st_suchower	Lisa Suchower	['Lisa J. Suchower','Margaret D. Copenhaver']
nesug1998.p133	Patrick Roohan	['Patrick J. Roohan','John R. Piddock','Peter B. Lannon']
nesug1998.p033	Bruce Gilsen	['Bruce F. Gilsen','Peter Sorock']
nesug1998.p051	Gina Thomas	['Gina M. Thomas','Christopher A. Roper','Michael Gilman']
nesug1998.p077	Adam Terr	['Michael L. Davis','Adam Terr']
nesug1998.p117	Chapman Gleason	['Chapman P. Gleason','Michael Yu','Myles E. Powers','Jay Jacob Wind']
nesug1998.p148	Peter Prause	['Peter N. Prause','Charles S. Patridge']
nesug1998.p100	Ray Pass	['Ray Pass','Daphne E. Ewing']
nesug1998.p106	Daphne Ewing	['Daphne E. Ewing','Ray Pass']
nesug1998.p194	Donna Torrence	['Donna Torrence','Julianne Staub Perry']
nesug1998.p008	Janet Stuelpner	['Janet E. Stuelpner','Boris Krol']
nesug1998.p004	Danbo Yi	['Danbo Yi','Lei Zhang']
nesug1998.p005	Lei Zhang	['Lei Zhang','Danbo Yi']
nesug1998.p006	Daniel Sargent	['Daniel Sargent','Clinton S. Rickards']
nesug1998.p036	Carolyn Kuntz	['Carolyn M. Kuntz','Scott T. LaPorte','Carolyn S. Maass']
nesug1998.p114	Cornelius Chase	['Cornelius T. Chase','Susan K. Gutschmidt']
nesug1998.p191	Cathy Phipps	['Catherine A. Phipps','Jim Davis']
nesug1998.p043	Robert Workman	['Robert S. Workman','Richard F. Pless']
nesug1998.p023	Helen-Jean Talbott	['Helen-Jean Talbott','Earl R. Westerlund']
nesug1998.p135	Lev Sverdlov	['Lev Sverdlov','John F. Noble']
nesug1998.p137	Patricia Houck	['Patricia R. Houck','Sati Mazumdar','Charles F. Reynolds, III']
nesug1998.p172	Aldona Rajecki	['Eric E. Kahle','Aldona A. Rajecki']
nesug1998.p185	Louise Hadden	['Louise Hadden','Jim McIntosh']
nesug1998.p190	Ernest S. Shtatland	['Ernest S. Shtatland','Mary B. Barton']
nesug1998.p048	Christopher Roper	['Christopher A. Roper','Gina M. Thomas']
nesug1998.p070	David Izrael	['David Izrael','David Russo']
nesug1998.p046	Susan Garavaglia	['Susan B. Garavaglia','Asha Sharma']
nesug1998.p188	Chris Ollinger	['Christopher R. Ollinger','Randall D. Tobias']

PharmaSUG is missing coauthors in 2008 and 2009:



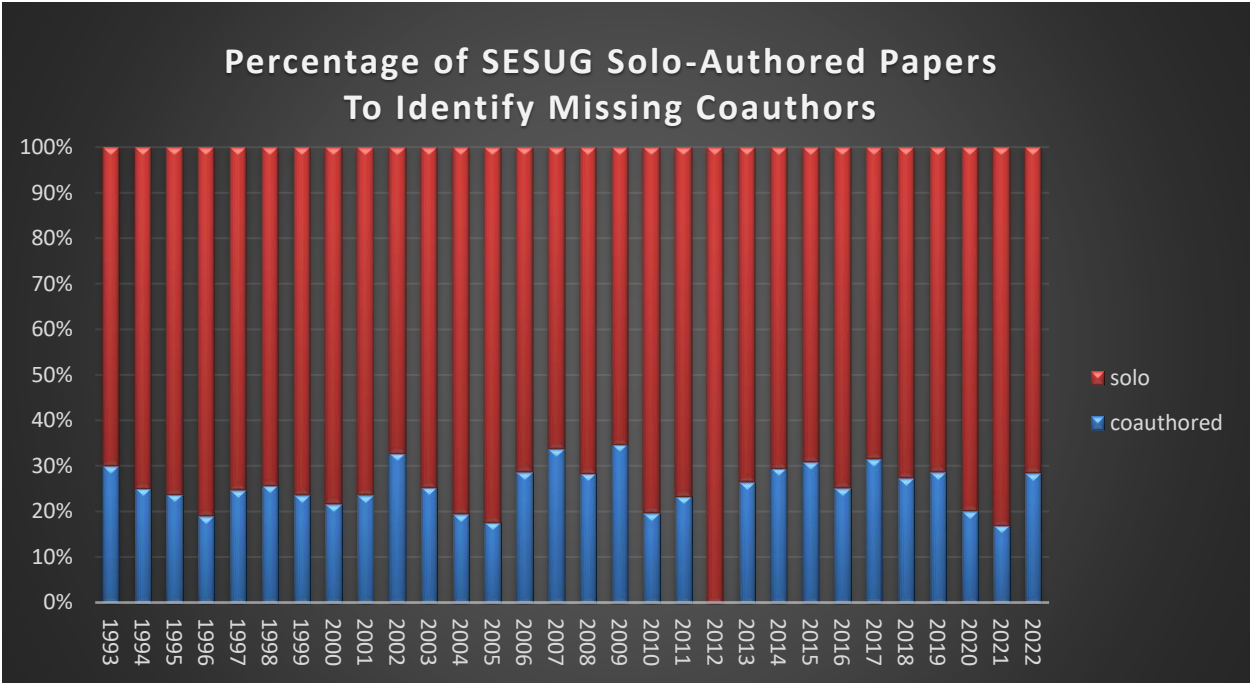
These outliers represent a substantial deviation, as on average, roughly 40 percent of PharmaSUG presentations have one or more coauthors. Thus, 78 presentations were identified as missing one or more coauthors, and once revised, the data no longer demonstrate these outliers:



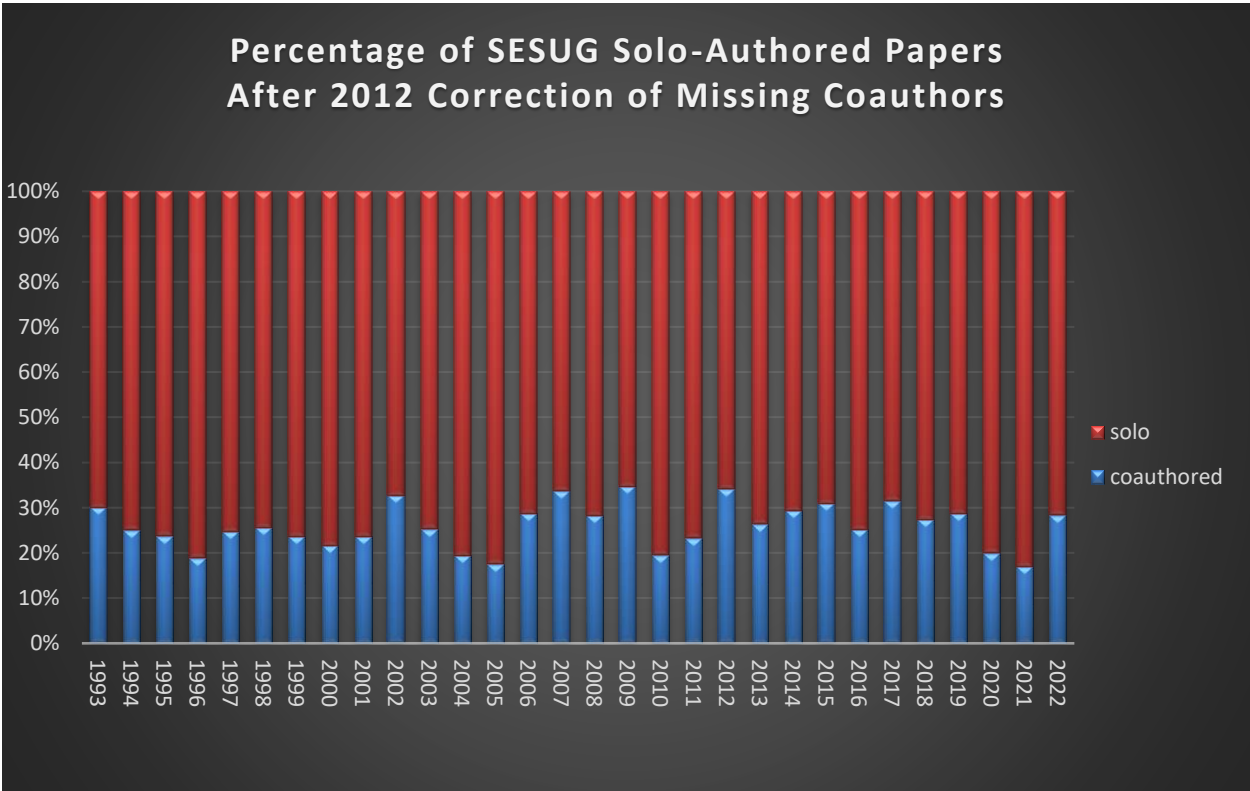
The 78 papers that required the addition of one or more coauthors include:

paper_id	raw_author_affil	paper_authors
pharmasug2008.ad03	Dongmin Shen, Schering-Plou	['Dongmin Shen', 'Hsiwei (Michael) Yu']
pharmasug2008.ad13	Mirjana Stojanovic, Duke Uni	['Mirjana Stojanovic', 'Donna Niedzwiecki']
pharmasug2008.ad16	Xiaopeng Cai, Amylin Pharma	['Jiangang Jameson Cai', 'Jianlin (Jay) Zhou']
pharmasug2008.ad17	David Clawson, Amylin Pharm	['David Clawson', 'Jiangang Jameson Cai']
pharmasug2008.ad19	Brian Fairfield-Carter, ICON Cl	['Brian P. Fairfield-Carter', 'Stephen W. Hunt']
pharmasug2008.cc05	Wenjie Song, Johns Hopkins	['Wenjie Song', 'Ke Yang']
pharmasug2008.cc06	Kalyani Kothapalli, MDS Phar	['Wenjie Wang', 'Jade Huang']
pharmasug2008.cc13	Simon Lin, Merck Co.	['Simon Lin', 'Huei-Ling Chen']
pharmasug2008.cc22	Yanmei Lu, Datameans	['Yanmei Lu', 'Russell M. Lavery']
pharmasug2008.cc26	David Escobar, Seattle Childre	['David M. Escobar', 'Ann Fowler', 'Barbara Mathewson']
pharmasug2008.cc28	Shiqun Li, MiniMax Informati	['Shiqun (Stan) Li', 'Wei Zhou']
pharmasug2008.dm04	Jim Fang, sanofi-aventis	['Amos Shu', 'Jiming Fang', 'Reddy Ayyalapu', 'Patrick Dunleavy']
pharmasug2008.dm07	Alejandro Jaramillo, Data Mei	['Alejandro Jaramillo', 'Russell M. Lavery']
pharmasug2008.dm08	Greg Nelson, ThotWave	['Emilio J. Power', 'Gregory S. Barnes Nelson']
pharmasug2008.ma01	Jackie Lane, ICON Clinical Res	['Jackie Lane', 'Robert Thomas Stempler']
pharmasug2008.po07	Jian (Daniel) Huang, Forest Re	['Daniel Huang', 'Lois Lynn']
pharmasug2008.po18	Aiming Yang, Merck Co.	['Aiming Yang', 'Huei-Ling Chen']
pharmasug2008.po21	Mei Dey, Merck Co.	['Mei Dey', 'Lisa Pyle']
pharmasug2008.pr04	Claire Chiang, i3DrugSafety	['Claire Chiang', 'Sue Su']
pharmasug2008.rs09	Frank Dilorio, CodeCrafters, Ir	['Jeffrey M. Abolafia', 'Frank C. Dilorio']
pharmasug2008.rs11	Robert Diseker, PPDI	['Robby Diseker', 'Lindsay Dean']
pharmasug2008.sa01	Teresa Arthur, SAS Institute I	['Teresa Arthur', 'Mary Jafri']
pharmasug2008.sa02	David Duling, SAS Institute In	['David R. Duling', 'Howard Plemmons, Jr.', 'Nancy Rausch']
pharmasug2008.sa12	Eric Brinsfield, SAS Institute Ir	['Eric C. Brinsfield', 'David Olaleye']
pharmasug2008.sa13	Richard Foley, SAS Institute Ir	['Richard Foley', 'Anthony L. Friebe']
pharmasug2008.sp03	Junliang Chen, Talecris Biothe	['Junliang Chen', 'Shannon Stock', 'Chunqin Deng']
pharmasug2008.sp04	William McCarthy, MMRI	['William F. McCarthy', 'Nan Guo']
pharmasug2008.sp07	Suhas Sanjee, MaxisIT Inc	['Suhas Ramesh Sanjee', 'Sheng Zhang']
pharmasug2008.tt01	Brian Fairfield-Carter, ICON Cl	['Brian P. Fairfield-Carter', 'David W. Carr']
pharmasug2008.tt04	Yanmei Lu, Datameans	['Yanmei Lu', 'Russell M. Lavery']
pharmasug2008.tt13	Robert Diseker, PPDI	['Anthony Harris', 'Robby Diseker']
pharmasug2008.tt14	Jasmin Fredette, ICON Clinica	['Jasmin Fredette', 'Brian P. Fairfield-Carter']
pharmasug2008.tu02	Paul Dorfman, Independent S	['Paul M. Dorfman', 'Lessia S. Shajenko']
pharmasug2008.tu07	Sandeep Gaudana, Ingenix	['Judy Loren', 'Sandeep Gaudana']
pharmasug2008.tu09	Lawrence Rasouliyan, ICON Cl	['Lawrence Rasouliyan', 'Dave P. Miller']
pharmasug2008.tu10	Laura Archer, Cancer Center,	['Laura E. Archer', 'Donna Hollis']
pharmasug2009.ad04	Huei-Ling Chen	['Huei-Ling Chen', 'Qian Dong']
pharmasug2009.ad05	Suhas Ramesh Sanjee	['Suhas Ramesh Sanjee', 'Christopher Tong']
pharmasug2009.ad07	Xiangchen (Bob) Cui	['Xiangchen (Bob) Cui', 'Shan Chen', 'Mei Wu']
pharmasug2009.ad10	Sathish Sundaram	['Sathish Sundaram', 'Ganesh Munuswamy', 'Mominul Islam']
pharmasug2009.ad11	Sandeep Juneja	['Sandeep Juneja', 'Vivek Mohan']
pharmasug2009.ad15	Cheng Jun Tian	['Cheng Jun Tian', 'JauRuey Huang']
pharmasug2009.ad17	Yining Wang	['Cheng Jun Tian', 'Yining Wang', 'Lisa Zhou']
pharmasug2009.ad18	Mila Chiflikyan	['Mila Chiflikyan', 'Mai Nguyen', 'Renee Karlsen']
pharmasug2009.ad21	Art Carpenter	['Arthur L. Carpenter', 'Richard O. Smith']
pharmasug2009.cc03	Sheng Zhang	['Sheng Zhang', 'Xingshu Zhu', 'Shuping Zhang', 'Weifeng Xu', 'Jane X. Liao', 'Amy E.
pharmasug2009.cc05	Wenjie Wang	['Wenjie Wang', 'Jade Huang']
pharmasug2009.cc06	Daniel Li	['Daniel Li', 'Suwen Li', 'Stephanie Sproule']
pharmasug2009.cc11	Fikret Karahoda	['Fikret Karahoda', 'Eric Qi']
pharmasug2009.cc12	Shiqun (Stan) Li	['Shiqun (Stan) Li', 'Wei Zhou']
pharmasug2009.cc14	David Gray	['David Gray', 'Zhao Chen']
pharmasug2009.cc17	Lin Yan	['Lin Yan', 'Aiming Yang']
pharmasug2009.cc20	Sandeep Sawant	['Sandeep Sawant', 'Gary Allen']
pharmasug2009.cc22	Ruben Chiflikyan	['Ruben Chiflikyan', 'Mila Chiflikyan', 'Donna Medeiros']
pharmasug2009.dm03	Chris Decker	['Chris Decker', 'Stephen Baker']
pharmasug2009.dm07	Zhuo Chen	['Zhuo Chen', 'David Gray']
pharmasug2009.dm08	Maruf Chowdhury	['Maruf Chowdhury', 'Barry Brolley']
pharmasug2009.ma04	Jonathan Botha	['Jonathan Botha', 'Sandra Vermeulen']
pharmasug2009.po02	Suwen Li	['Suwen Li', 'Daniel Li', 'Stephanie Sproule', 'David Zhu']
pharmasug2009.po06	Paul OldenKamp	['Paul OldenKamp', 'Paul D. Hamilton', 'Dante DiTommaso', 'Ann Marie Martin']
pharmasug2009.po08	Tracy Lin	['Tracy Lin', 'Jie Huang']
pharmasug2009.po16	Dante Di Tommaso	['Dante DiTommaso', 'Ann Marie Martin', 'Paul OldenKamp', 'Paul D. Hamilton']
pharmasug2009.po17	Cheng Jun Tian	['Cheng Jun Tian', 'Denis Michel']
pharmasug2009.po19	Cheng Jun Tian	['Cheng Jun Tian', 'Qin Li', 'Jiangfan Li']
pharmasug2009.po23	Bradford Danner	['Bradford J. Danner', 'Steven R. Kirby']
pharmasug2009.pr03	Mark Farman	['Mark Farman', 'William Koh', 'Sandra L. Close']
pharmasug2009.pr05	Jing Wang	['Jing Wang', 'Marla J. Husnik']
pharmasug2009.rs01	Gregory Steffens	['Gregory C. Steffens', 'Ian Fleming']
pharmasug2009.rs05	Beilei Xu	['Beilei Xu', 'Changhong Shi']
pharmasug2009.rs08	Steve Wong	['Christine Connolly', 'Kevin King', 'Amanda Tweed', 'Steve Wong']
pharmasug2009.sp01	Aiming Yang	['Aiming Yang', 'Nancy Liu', 'Olga M. Kuznetsova']
pharmasug2009.sp02	William McCarthy	['William F. McCarthy', 'Nan Guo']
pharmasug2009.sp03	Robert Downer	['Robert G. Downer', 'Patrick J. Richardson']
pharmasug2009.sp04	Daniel DiPrimeo	['Daniel DiPrimeo', 'Bruce Johnston', 'Lisa Price', 'Junming Yang']
pharmasug2009.sp10	Xiaomin He	['Xiaomin He', 'Shwu-Jen Wu']
pharmasug2009.sp11	Jing Su	['Jing Su', 'Peng Sun', 'Xiaodong Li', 'Alan H. Hartford']
pharmasug2009.tt08	John King	['John Henry King', 'Michael S. Zdeb']
pharmasug2009.tu09	Jeffrey Abolafia	['Jeffrey M. Abolafia', 'Susan J. Kenny', 'Stacy Patchell', 'Gregory C. Steffens', 'Willie

SESUG was shown to be missing all coauthors in 2012:



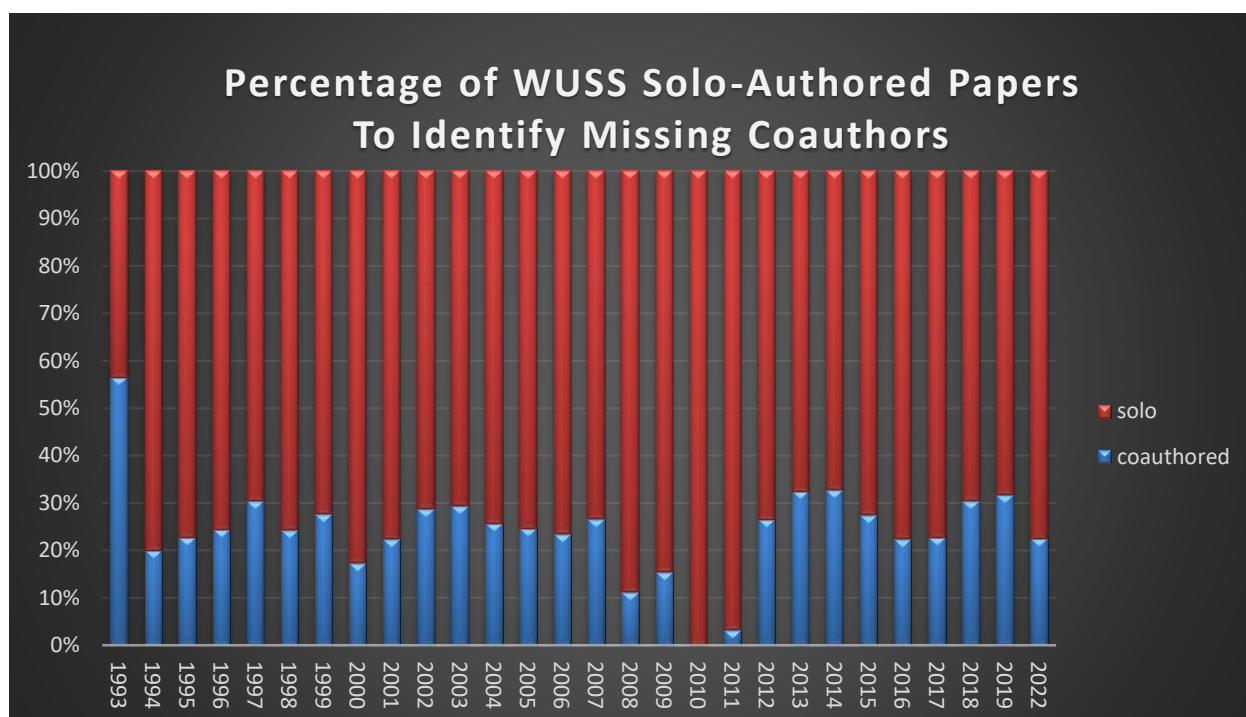
After correcting the 41 presentations that were missing a coauthor, the outlier year is no longer an issue:



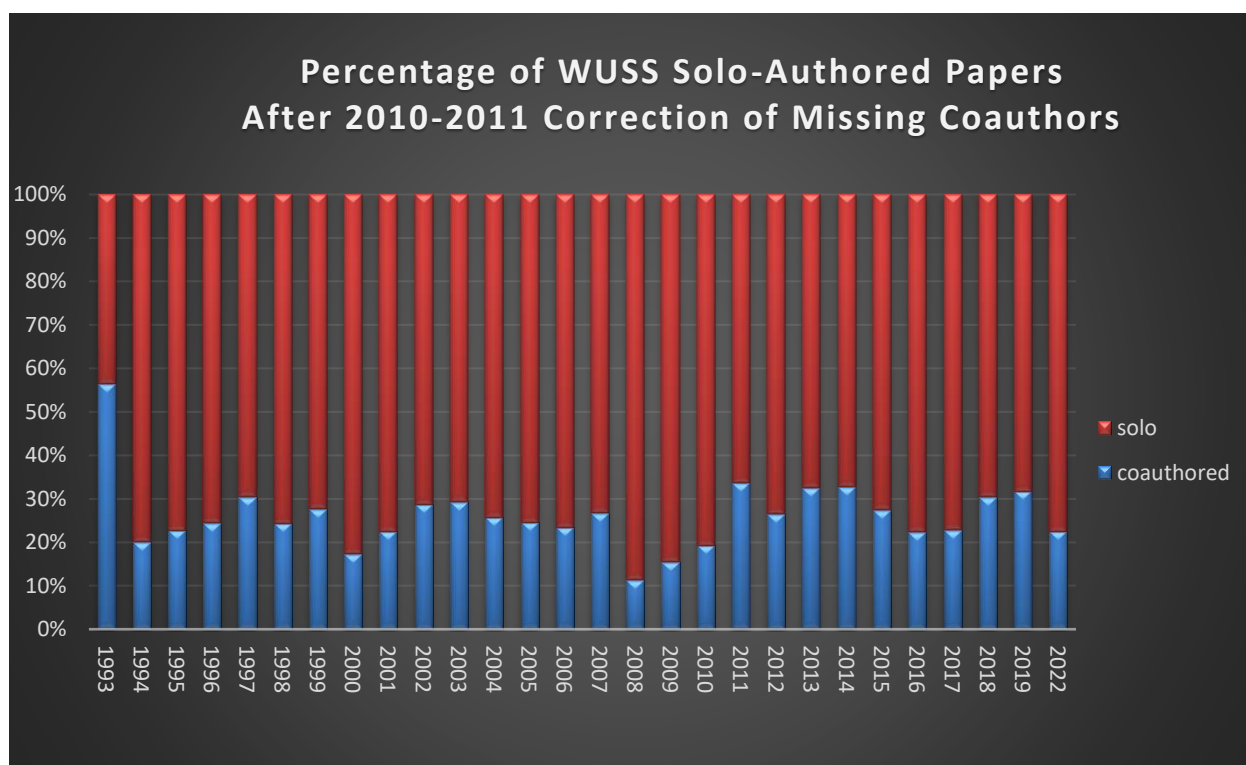
These 41 SESUG corrections include:

paper_id	raw_author_affil	paper_authors
sesug2012.bb-02	Patricia Aanderud	['Tricia Aanderud','Angela M. Hall']
sesug2012.bb-08	David Abbott	['David H. Abbott','Rebecca B. McNeil']
sesug2012.bb-11	Stuart Long	['Stuart D. Long','Edward E. Heaton','Daniel R. Scharf']
sesug2012.ct-07	Phil d'Almada	['Philip J. d'Almada','Daniel Wojdyla']
sesug2012.ct-11	Elizabeth Heath	['Elizabeth Heath','Priya Suresh']
sesug2012.ct-13	Ravi Kankipati	['Ravi Kankipati','Abhilash Chimbirithy']
sesug2012.ct-20	Meghal Parikh	['Meghal Parikh','Elayne Reiss']
sesug2012.ct-30	Yusheng Zhai	['Yusheng Zhai','Ryan Diver','Xia Lin']
sesug2012.dj-02	Annie Dudley Zangi	['Diane K. Michelson','Annie Dudley Zangi']
sesug2012.hw-01	Peter Eberhardt	['Peter W. Eberhardt','Louanna Kong']
sesug2012.it-01	Mira Shapiro	['Mira Shapiro','Kirk Paul Lafler']
sesug2012.it-05	Kevin Russell	['Charles E. Mullin','Kevin Russell']
sesug2012.it-06	Lesa Caves	['Lesa Caves','Nicole Williams']
sesug2012.ph-02	Hisham Madi	['Hisham Madi','Anna Maeser','Michael Zichy']
sesug2012.ph-06	Mei Dey	['Mei Dey','Lisa Pyle']
sesug2012.ph-07	Richard Zink	['Richard C. Zink','Russell D. Wolfinger']
sesug2012.ph-11	John R Gerlach	['John R. Gerlach','Glenn O'Brien']
sesug2012.pa-07	Margaret Crevar	['Margaret A. Crevar','Tony Brown']
sesug2012.po-02	Rebecca Horney	['Rebecca Horney','Karen Jones','Annette Wiseman']
sesug2012.po-05	Abbas Tavakoli	['Abbas S. Tavakoli','Richard Walker']
sesug2012.po-07	Kunal Agnihotri	['Kunal Agnihotri','Kenneth W. Borowiak']
sesug2012.po-10	Jeffrey Kromrey	['Diep T. Nguyen','Patricia Rodríguez de Gil','Eun Sook Kim','Aarti P. Bellara','Ann F
sesug2012.po-11	Daniel Hertz	['Daniel L. Hertz','Daniel J. Crona','Jasmine Talameh','Scott J. Brantley','Luke Rood
sesug2012.po-12	Neeta Shenvi	['Neeta Shenvi','Amita Manatunga','Andrew Taylor']
sesug2012.po-13	Patricia Rodriguez de Gil	['Patricia Rodriguez de Gil','John M. Ferron']
sesug2012.po-14	So Young Park	['So Young Park','Marcela Alfaro-Cordoba']
sesug2012.po-19	Anil Kumar Pantangi	['Anil Kumar Pantangi','Musthan Kader Meeran Mohideen','Goutam Chakraborty
sesug2012.po-20	Mihaela Ene	['Mihaela Ene','Whitney Smiley','Bethany A. Bell']
sesug2012.ri-01	Patricia Aanderud	['Tricia Aanderud','Angela M. Hall']
sesug2012.ri-03	Patrick Hall	['Patricia H. Hall','Ryan Snyder']
sesug2012.ri-09	Ben Robbins	['Ben Robbins','Michael Drutar']
sesug2012.ri-10	Neeta Shenvi	['Neeta Shenvi','Xin Zhang','Azhar Nizam']
sesug2012.ri-12	Helen Smith	['Helen Smith','Susan Myers']
sesug2012.ri-14	Shih-Ching Wu	['Shih-Ching Wu','Shane McLaughlin']
sesug2012.ri-15	Ed Summers	['Ed Summers','Juliana Langston','Robert Allison','Jennifer Cowley']
sesug2012.ri-17	Darrell Massengill	['Darrell Massengill','Alexander Kolovos']
sesug2012.sd-01	Abbas Tavakoli	['Abbas S. Tavakoli','Sue Heiney']
sesug2012.sd-06	Rajendra Kadel	['Rajendra P. Kadel','Kevin E. Kip']
sesug2012.sd-08	John Chantis	['John Chantis','Kandasamy Selvavel']
sesug2012.sd-11	Leonard Gordon	['Leonard Gordon','Brian Pavilonis']

WUSS was shown to be missing coauthors for 2010 and 2011:



After correcting the 64 presentations that were found to be missing one or more coauthors, 2010 and 2011 were no longer outliers::



The 64 WUSS corrections for 2010 and 2011 include:

paper_id	raw_author_affil	paper_authors
wuss2010.3067_3_anl-carp	Christopher Ake	['Christopher F. Ake', 'Arthur L. Carpenter']
wuss2010.2963_6_anl-derr	AnnMaria De Mars	['AnnMaria De Mars', 'Revati Kadu']
wuss2010.2954_6_anl-dert	Nate Derby	['Nate Derby', 'Laura Vo']
wuss2010.3024_2_anl-rodr	Robert Rodriguez	['Maura E. Stokes', 'Robert N. Rodriguez', 'Robert A. Cohen']
wuss2010.3038_4_app-gicc	Marc Desgrouilliers	['Marc Desgrouilliers', 'Romain Miralles']
wuss2010.3039_5_app-gicc	Ale Gicqueau	['Ale Gicqueau', 'Marc Desgrouilliers']
wuss2010.2916_4_app-sha	Mira Shapiro	['Mira Shapiro', 'Kirk Paul Lafler']
wuss2010.2932_3_cod-bee	Faith Beery	['Faith S. Beery', 'Dave P. Miller']
wuss2010.2964_4_cod-den	AnnMaria De Mars	['AnnMaria De Mars', 'Ernesto Flores']
wuss2010.2998_2_cod-kilb	Julie Kilburn	['Julie A. Kilburn', 'Rebecca Ottesen']
wuss2010.2962_3_cod-truc	Sy Truong	['Sy J. Truong', 'Carey G. Smoak']
wuss2010.2993_2_cod-woc	Jefferson Wood	['Jefferson L. Wood', 'Dennis G. Fisher', 'Grace L. Reynolds', 'Lucy E. Napper']
wuss2010.2999_4_cod-xia	Lan Xiao	['Lan Xiao', 'Jun Ma']
wuss2010.2996_4_ddi-hule	Denis Hulett	['Denis Hulett', 'Glenn Wright', 'Jaycee Karl']
wuss2010.2997_2_ddi-wrig	Glenn Wright	['Glenn Wright', 'Denis Hulett']
wuss2010.2994_3_how-del	Lora Delwiche	['Lora D. Delwiche', 'Susan J. Slaughter']
wuss2010.2940_2_hor-mill	Stefanie Millar	['Stefanie J. Millar', 'David J. Pasta']
wuss2010.2930_2_hor-sho	Sarah Short	['Sarah A. Short', 'David J. Pasta']
wuss2010.2995_2_hor-terv	Anna ter Veer	['Anna ter Veer', 'Julie A. Kilburn', 'Ann Vanderplas', 'Rizvan Mamet', 'Rebecca Ottesen']
wuss2010.3008_4_pos-dar	Joffrey Daras	['Joffrey Daras', 'Marc Desgrouilliers']
wuss2010.2921_4_pos-fan	Frank Fan	['Frank Fan', 'Ale Gicqueau']
wuss2010.2985_5_pos-fer	Ryan Fernandez	['Ramya R. Fernandez', 'Mary Wilson']
wuss2010.3009_5_pos-hu	Bing Hu	['Bing Hu', 'Marc Desgrouilliers']
wuss2010.3007_3_pos-le	Rebecca Le	['Rebecca Le', 'Daniel R. Jeske']
wuss2010.2968_2_pos-smi	Denise Smith	['Denise J. Smith', 'Daniel Schulz', 'Wei Cheng']
wuss2011.papers_lai_g_74	Ginny Lai, ICON Clinical Resea	['Ginny P. Lai', 'David R. Mink', 'David J. Pasta']
wuss2011.papers_langholz	Bryan Langholz, University of	['Bryan Langholz', 'David B. Richardson']
wuss2011.papers_li_y_749	Yunfeng Li, Cerexa, Inc.	['Yunfeng Li', 'Lily Llorens', 'Ganesh Tamarana']
wuss2011.papers_lisha_n	Nadra Lisha, University of Sou	['Nadra E. Lisha', 'Rachel Grana', 'Ping Sun', 'Louise A. Rohrbach', 'Donna Spruijt-Met']
wuss2011.papers_stokes_r	Maura Stokes, SAS Institute	['Maura E. Stokes', 'Fang K. Chen', 'Ying C. So']
wuss2011.papers_carpent	Art Carpenter, CA Occidental	['Eric T. Sun', 'Arthur L. Carpenter']
wuss2011.papers_carpent	Art Carpenter, CA Occidental	['Arthur L. Carpenter', 'Dennis G. Fisher']
wuss2011.papers_garla_s	Satish Garla, Oklahoma State	['Satish Garla', 'Goutam Chakraborty']
wuss2011.papers_gicquea	Ale Gicqueau, Clinovo	['Ale Gicqueau', 'Miki Huang', 'Stephen Chan']
wuss2011.papers_thornto	Patrick Thornton, SRI Internat	['Patrick Thornton', 'Iuliana Barbalau']
wuss2011.papers_garla_s	Satish Garla, Oklahoma State	['Satish Garla', 'Goutam Chakraborty', 'Gary Gaeth']
wuss2011.papers_liu_x_75	Xiao Liu, ICON Late Phase & O	['Xiao Liu', 'Stefanie J. Millar']
wuss2011.papers_melguiz	Maria Melguizo Castro, Unive	['Maria S. Melguizo Castro', 'Jerry R. Stalnaker', 'Christopher J. Swearingen']
wuss2011.papers_rosenbl	Mary Rosenbloom, Edwards Li	['Mary F. O. Rosenbloom', 'Kirk Paul Lafler']
wuss2011.papers_rosenbl	Mary Rosenbloom, Edwards Li	['Mary F. O. Rosenbloom', 'Kirk Paul Lafler']
wuss2011.papers_rosenbl	Mary Rosenbloom, Edwards Li	['Mary F. O. Rosenbloom', 'Arthur L. Carpenter']
wuss2011.papers_short_s	Sarah Short, ICON Clinical Res	['Sarah A. Short', 'Dave P. Miller']
wuss2011.papers_stein_a	Aleksandra Stein, Celerion/Ur	['Aleksandra Stein', 'Steven Kirby']
wuss2011.papers_budda	Aruna Buddana, TiVo Inc	['Aruna Buddana', 'Raghunathan Chakravarthy']
wuss2011.papers_derby_n	Nate Derby, Stakana Analytics	['Nate Derby', 'Laura Vo', 'Perry Watts']
wuss2011.papers_derby_n	Nate Derby, Stakana Analytics	['Perry Watts', 'Nate Derby']
wuss2011.papers_truong_s	Sy Truong, Meta-Xceed, Inc.	['Sy J. Truong', 'Carey G. Smoak']
wuss2011.papers_cheng_w	Wei Cheng, Isis Pharmaceutical	['Wei Cheng', 'Sanjay N. Matange']
wuss2011.papers_slaught	Susan Slaughter, Avocet Solut	['Susan J. Slaughter', 'Lora D. Delwiche']
wuss2011.papers_chilukuri	Kalyani Chilukuri, clinovo	['Kalyani Chilukuri', 'Ale Gicqueau']
wuss2011.papers_chu_l_74	Hao Chu, Kaiser Permanent Sc	['Li-Hao Chu', 'Aniket Kawatkar', 'Annie Gu']
wuss2011.papers_chung_n	Nadia Chung, City of Hope	['Nadia Chung', 'Huiyan Ma', 'Jenna Voutsinas', 'Leslie Bernstein', 'Sophia Wang', 'Jam']
wuss2011.papers_masias_a	Andrea Masias, University of t	['Andrea Masias', 'Vijaya Vemulakonda']
wuss2011.papers_miller_l	Lori Miller, University of Wash	['Lori L. Miller', 'Fred Wilds']
wuss2011.papers_zhao_b	Beinan Zhao, Palo Alto Medici	['Beinan Zhao', 'Eric C. Wong', 'Latha Palaniappan']
wuss2011.papers_zhao_w	Wuchen Zhao, University of Sc	['Wuchen Zhao', 'Arthur Xuejun Li']
wuss2011.papers_shaik_z	Zubair Shaik, Oklahoma State	['Zubair Shaik', 'Satish Garla', 'Goutam Chakraborty']
wuss2011.papers_lafler_k	Kirk Paul Lafler, Software Intel	['Kirk Paul Lafler', 'Charles Edwin Shipp']
wuss2011.papers_lafler_k	Kirk Paul Lafler, Software Intel	['Charles Edwin Shipp', 'Kirk Paul Lafler']
wuss2011.papers_carey_h	Helen Carey, Carey Consulting	['Helen D. Carey', 'Ginger Carey']
wuss2011.papers_slaught	Susan Slaughter, Avocet Solut	['Susan J. Slaughter', 'Lora D. Delwiche']
wuss2011.papers_schneid	Connie Robison, SAS Institute	['Mark L. Schneider', 'Donna Bennett', 'Connie Robison']
wuss2011.papers_stokes_r	Maura Stokes, SAS Institute	['Maura E. Stokes', 'Gary G. Koch']
wuss2008.tut06.tt	Jimmy Efird, Cincinnati Childr	['Jimmy T. Efird', 'Mi-Kyung Hong']

DATA QUALITY – 3. METADATA STANDARDIZATION

With author names correctly identified in (and extracted from) the LEX metadata, and hundreds of (albeit, admittedly, not all) missing coauthors added, the final step toward data quality required the standardization (i.e., entity resolution) of author names. This step was required to ensure that, to the extent possible, only one instance of each author would occur throughout time and across all conferences.

For example, William Benjamin, Jr., a popular SAS author, is represented eight different ways:

1. William Benjamin
2. William E. Benjamin
3. William Benjamin, Jr
4. William E Benjamin, Jr.
5. William E Benjamin, Jr
6. William E Benjamin, Jr.
7. William E. Benjamin, Jr
8. William E. Benjamin, Jr.

And Greg Nelson is represented nine different ways:

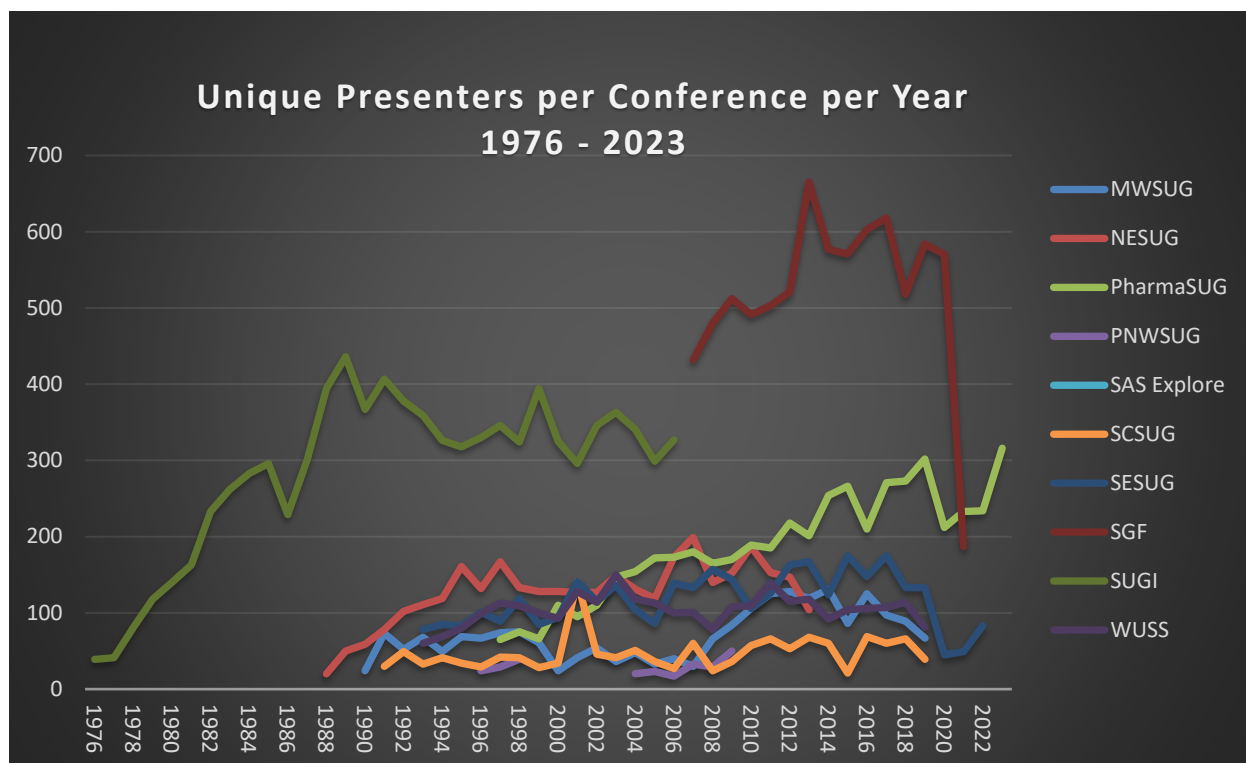
1. Greg Nelson
2. Greg S Nelson
3. Greg S. Nelson
4. Gregory Nelson
5. Gregory S. Nelson
6. Gregory B. Nelson
7. Greg Barnes Nelson
8. Greg S. Barnes Nelson
9. Gregory Barnes Nelson

This lengthy transformation process is not fully described herein; suffice it to say that thousands of author names required transformation. Notwithstanding, standardization is an inexact science that is thwarted by cases for which an entity cannot be verified or even established. For example, as some participants choose to use initials for their first and/or middle names, or as some participants choose to anglicize a non-English name, or as some women change their surnames after marriage, or as any user changes his name for any reason whatsoever—each of these shifts can cause the data to reflect that two or more users exist, when, in fact, only one does. When these transformations were complete, approximately 14,918 unique presenters were found to have presented—or authored—presentations across all referenced SAS conferences.

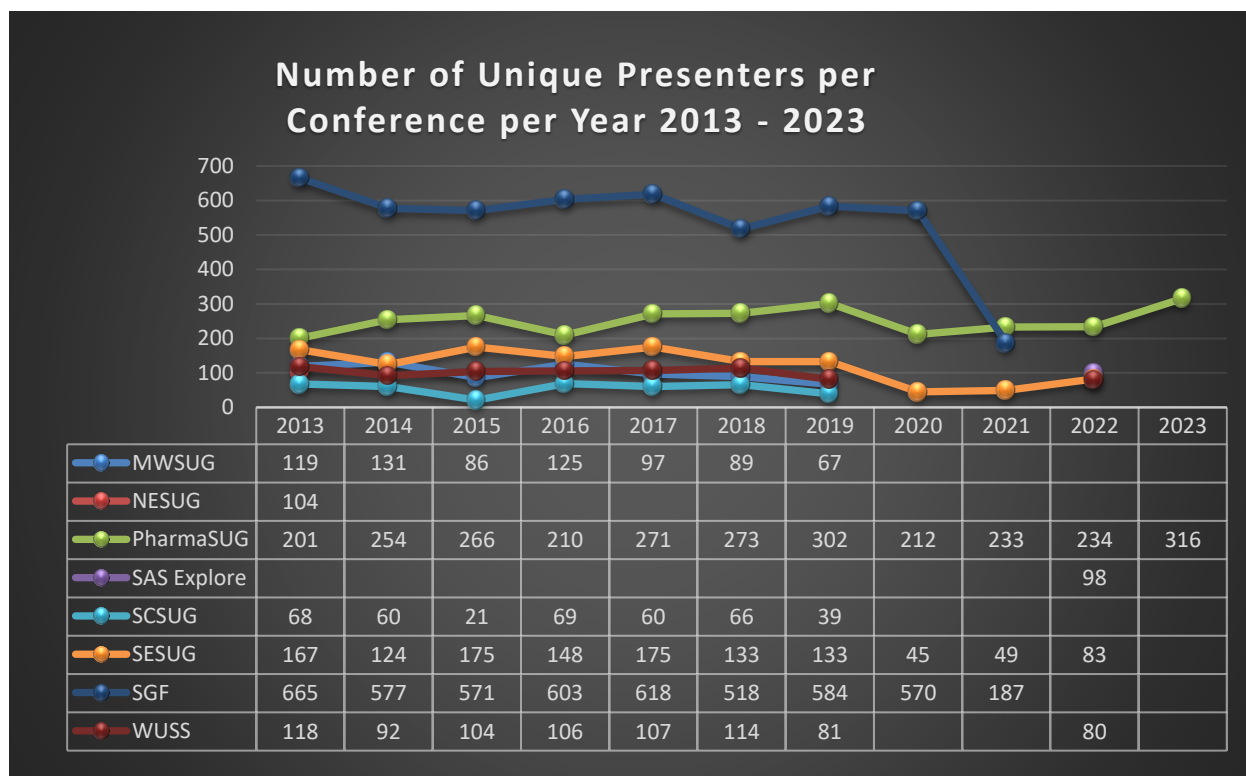
Future analysis could provide additional transformation and/or verification of names by programmatically evaluating the accompanying white papers, where they exist, to extract and compare contact information (such as name, company/educational affiliation, or email address) that typically appears on the first page of each paper, as well as in the “Contact” section the concludes each paper. Notwithstanding, presenters can change affiliation and/or email address over time, so these, too, are inexact identity indicators.

UNIQUE PRESENTERS

Just as the total number of presentations per conference has decreased significantly since COVID, so, too, has the number of unique presenters decreased for most conferences. Only PharmaSUG stands alone in its ability to have rebounded in 2022 to feature 316 unique presenters, surpassing its pre-COVID high of 302 presenters in 2019. However, WUSS and SESUG still have a chance to rally in 2023.



Over the past ten years, the most precipitous drop in unique presenters has been in SAS Institute conferences, with SGF dropping from 570 presenters in 2020 to only 187 presenters in 2021, and to only 98 presenters in 2022 (rebranded as the SAS Explore conference).



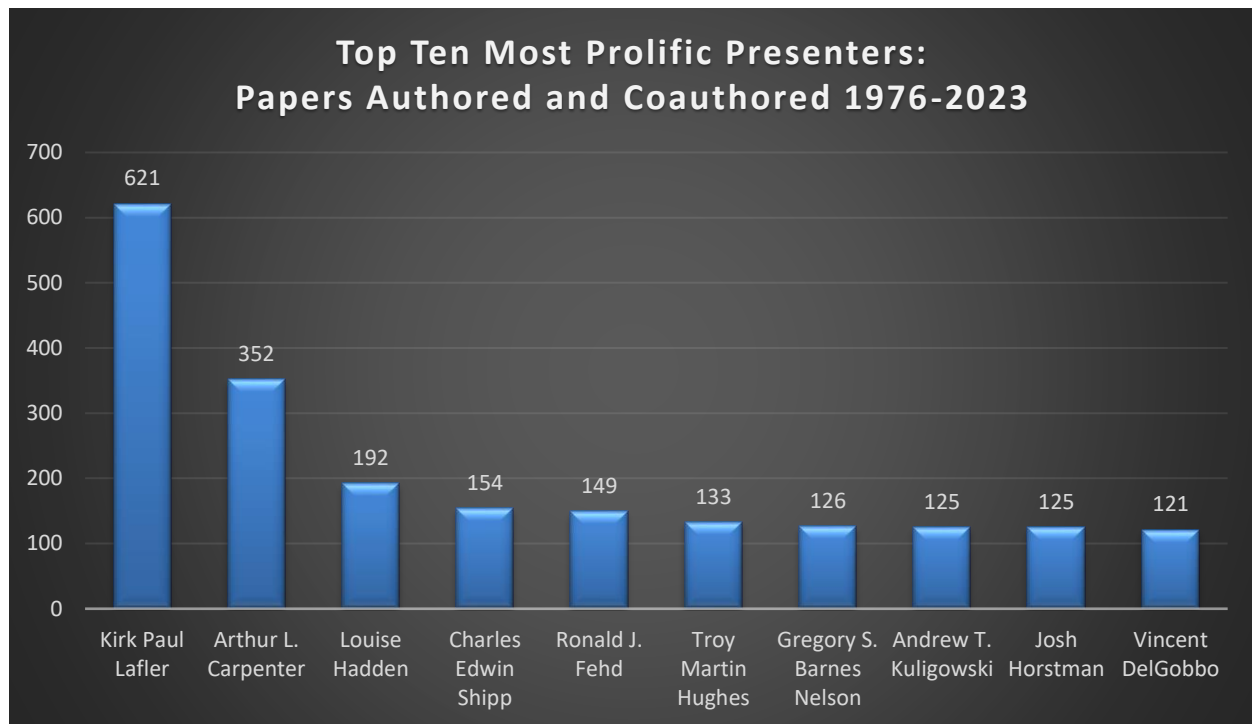
WHO'S BRINGING THAT BDE?

Talk about burying the lead—more than 20 pages in, and big data energy (BDE)—and who's bringing it—hasn't even been discussed! The remainder of this text highlights those most prolific presenters and our significant contributions to the SAS user community.

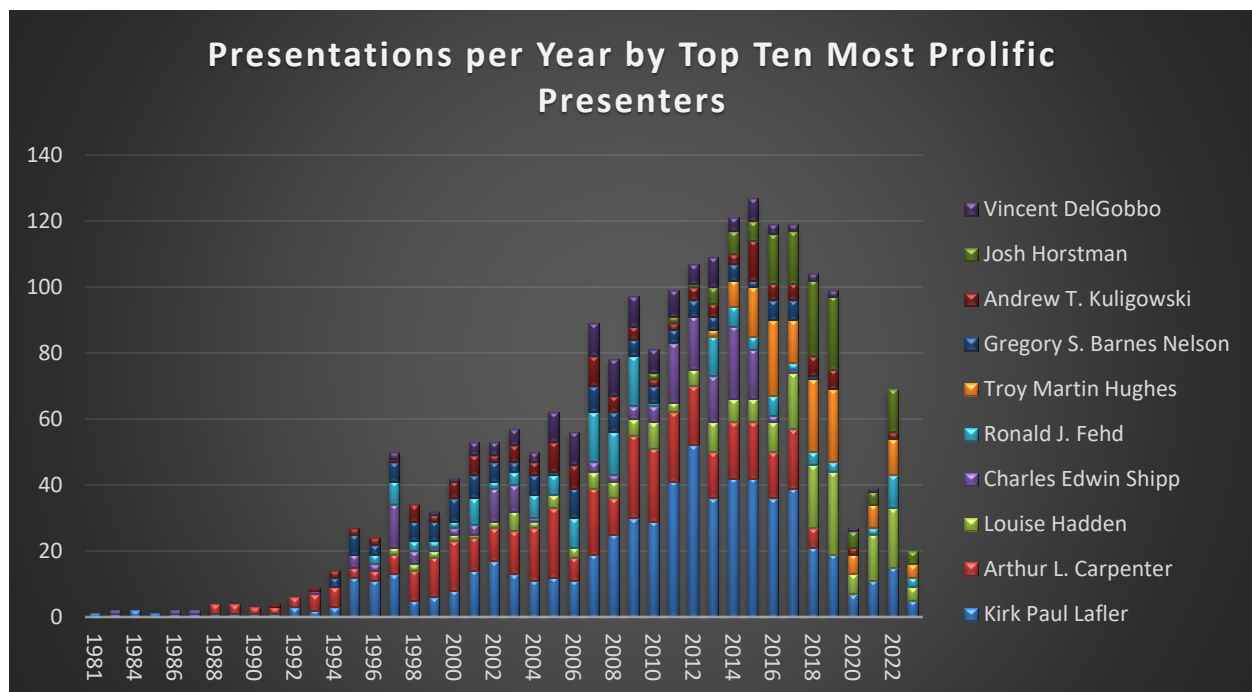
Ranked in order, the following 25 authors have most prolifically presented at SAS user conferences:

Rank	Author	Number of Presentations
1	Kirk Paul Lafler	621
2	Arthur L. Carpenter	352
3	Louise Hadden	192
4	Charles Edwin Shipp	154
5	Ronald J. Fehd	149
6	Troy Martin Hughes	133
7	Gregory S. Barnes Nelson	126
8	Andrew T. Kuligowski	125
8	Josh Horstman	125
10	Vincent DelGobbo	121
11	Peter W. Eberhardt	117
12	Andrew H. Karp	111
13	Ray Pass	110
14	Richann Watson	104
15	LeRoy Bessler	100
16	Ian Whitlock	97
17	Paul M. Dorfman	90
18	Bruce F. Gilson	88
19	Ron Cody	87
20	Maura E. Stokes	86
21	Sunil Kumar Gupta	85
21	Goutam Chakraborty	85
23	Benjamin Cochran	84
24	Frank C. Dilorio	83
24	Richard D. Langston	83

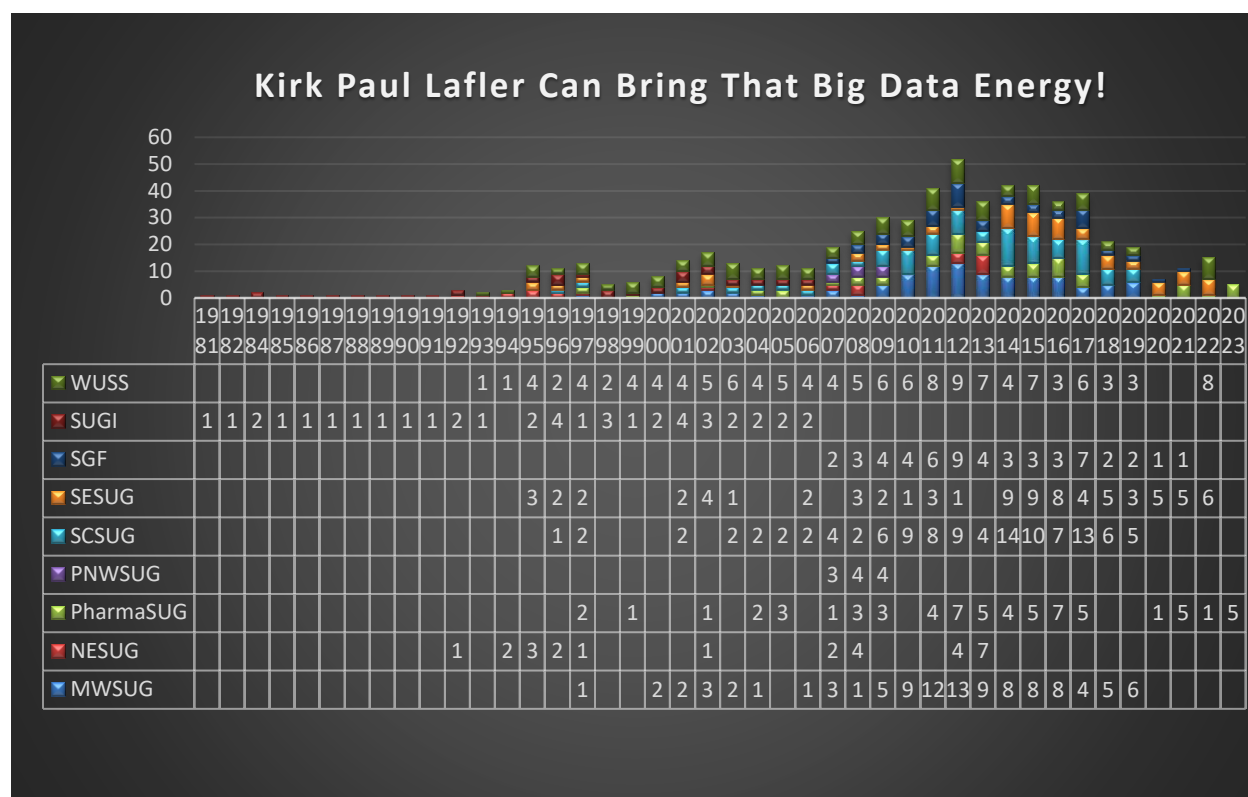
And it would be an understatement to say that Kirk Paul Lafler has eclipsed the rest of this crew with his 621 presentations! And hats off to Louise Hadden for being the solitary female in the top ten to be bringing that BDE!



The following graph demonstrates the activity of these ten presenters through the years across all SAS user group conferences, with the enormity of Kirk's BDE shown on the bottom of each stack.



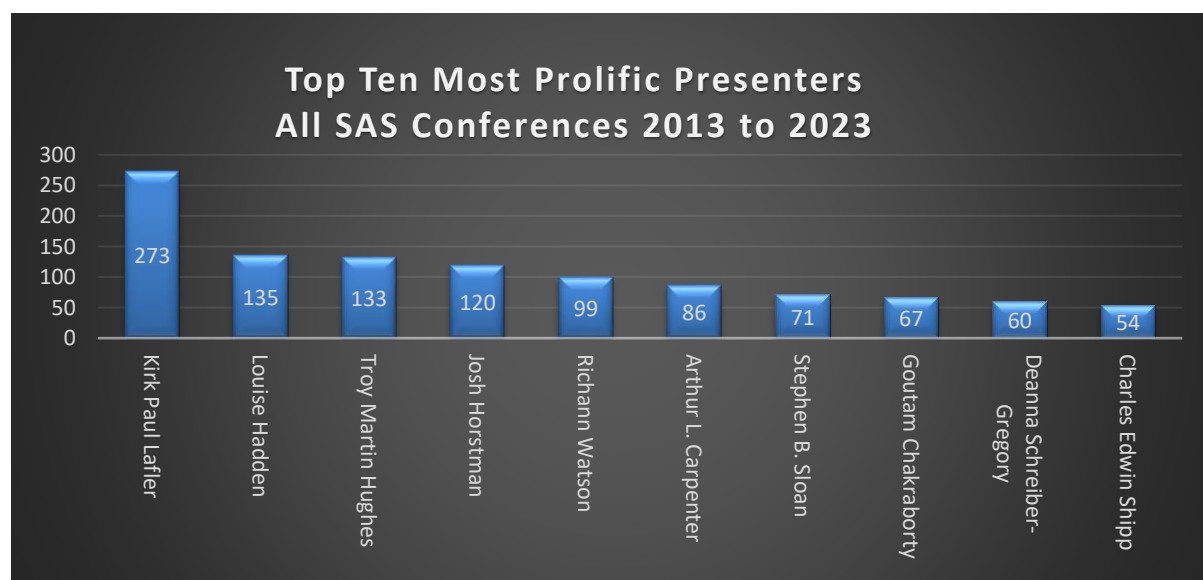
Kirk Paul Lafler began presenting in 1981 and has been bringing that BDE ever since!



It becomes apparent that as some current presenters had not yet been born—or were still in diapers—when Kirk had already begun presenting papers, recognition also should be afforded to the rising stars, including those presenters who are most recently contributing most significantly. This analysis is presented next.

RIISING STARS: WHO'S BRINGING THAT BDE FROM 2013 TO 2023

The following ten presenters have most prolifically presented from 2013 to 2023 across all SAS user group conferences:



Thus, by slicing and subsetting conference metadata to analyze only the most recent years, conference chairs and other leaders can best identify those users who should be invited to participate in future conferences. For example, an attendee might never have presented at the WUSS conference, but might be invited based on his significant contributions to SESUG or another conference.

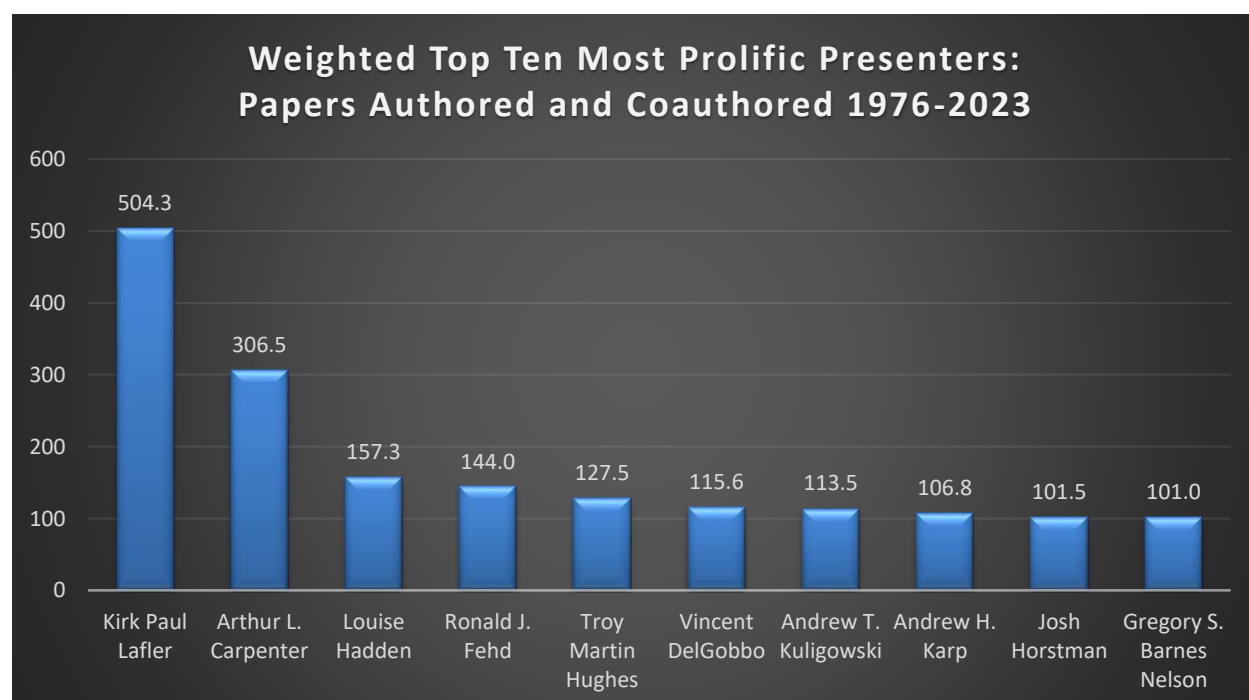
AUTHORING VS. COAUTHORING: WEIGHTED CONTRIBUTIONS OF PRESENTERS

As demonstrated in the “Data Quality – 2. Metadata Correction” section, coauthors and copresenters are commonplace among SAS user group conferences. And the argument could be made that coauthoring reduces the workload of each author. For example, where four coauthors present a twenty-minute talk (and write a ten-page paper), each of their individual contributions (and effort) are inarguably reduced in this zero-sum game. Thus, “weighted” contributions of presenters can be analyzed by scoring coauthors based on the number of coauthors, and one model would award only one-fourth of a presentation point where four coauthors presented, or only one-half of a presentation point where two coauthors presented.

When presentations are weighted, the following presenters are the all-time most prolific presenters:

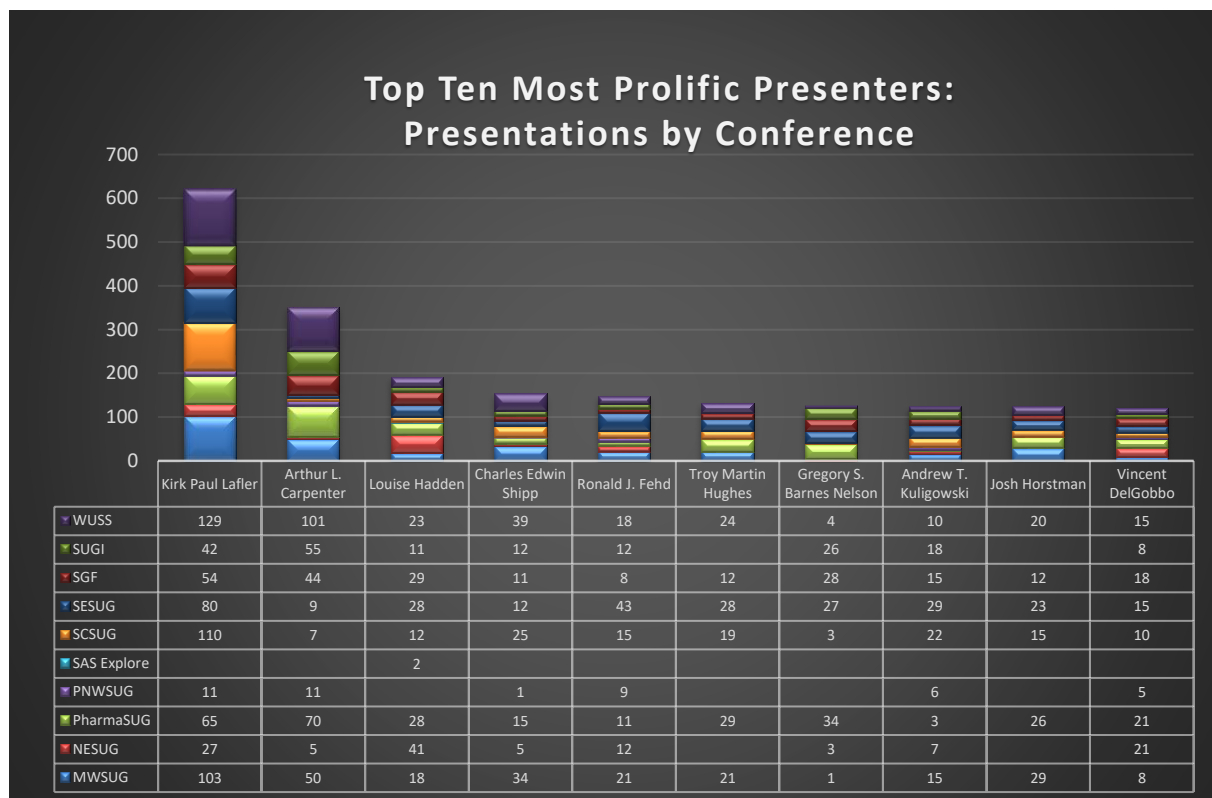
1. Kirk Paul Lafler
2. Arthur L. Carpenter
3. Louise Hadden
4. Ronald J. Fehd
5. Troy Martin Hughes
6. Vincent DelGobbo
7. Andrew T. Kuligowski
8. Andrew H. Karp
9. Josh Horstman
10. Gregory S. Barnes Nelson

The list closely follows the list of unweighted presenters, with the exception that Charlie Shipp moves off the top-ten list, and Andrew Karp moves onto the list.



CONFERENCE-LEVEL ANALYSIS

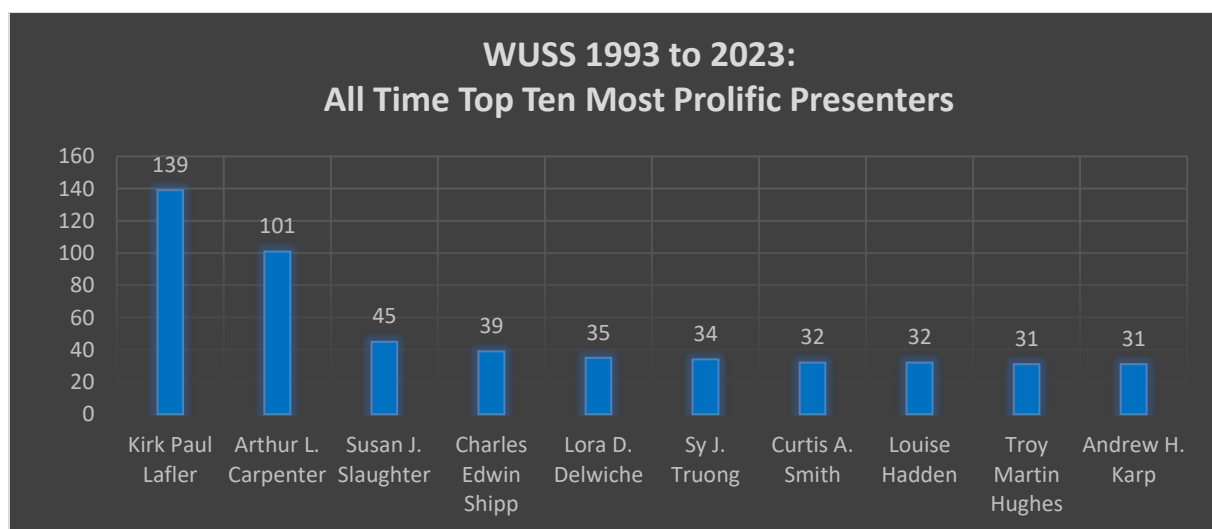
When evaluating the most prolific presenters, it becomes apparent that we all have presented across numerous conferences.



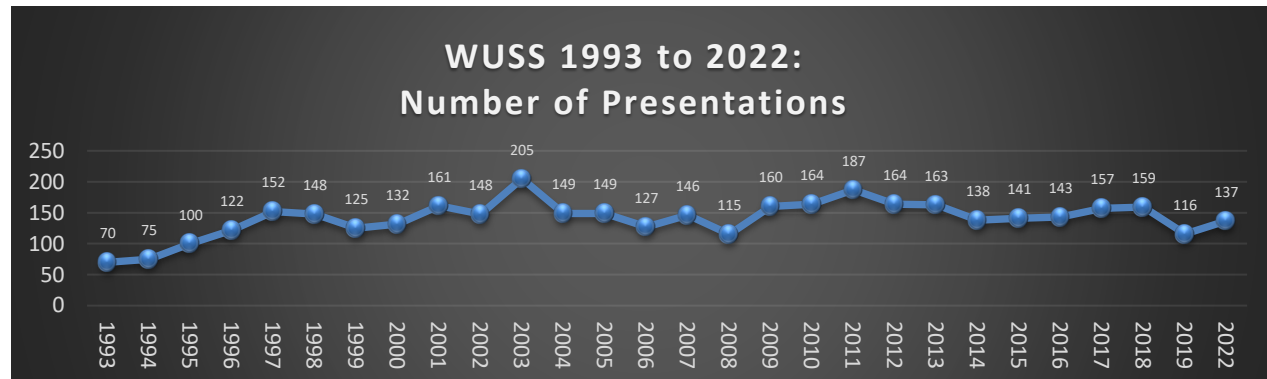
Not all presenters, however, have presented at multiple conferences, and some have been rock stars only at a couple or even one conference—and this is still commendable! Thus, the remaining subsections demonstrate conference-level analysis for the current conferences—WUSS, SESUG, and PharmaSUG.

WHO'S BRINGING THAT BDE TO WESTERN USERS OF SAS SOFTWARE (WUSS)?

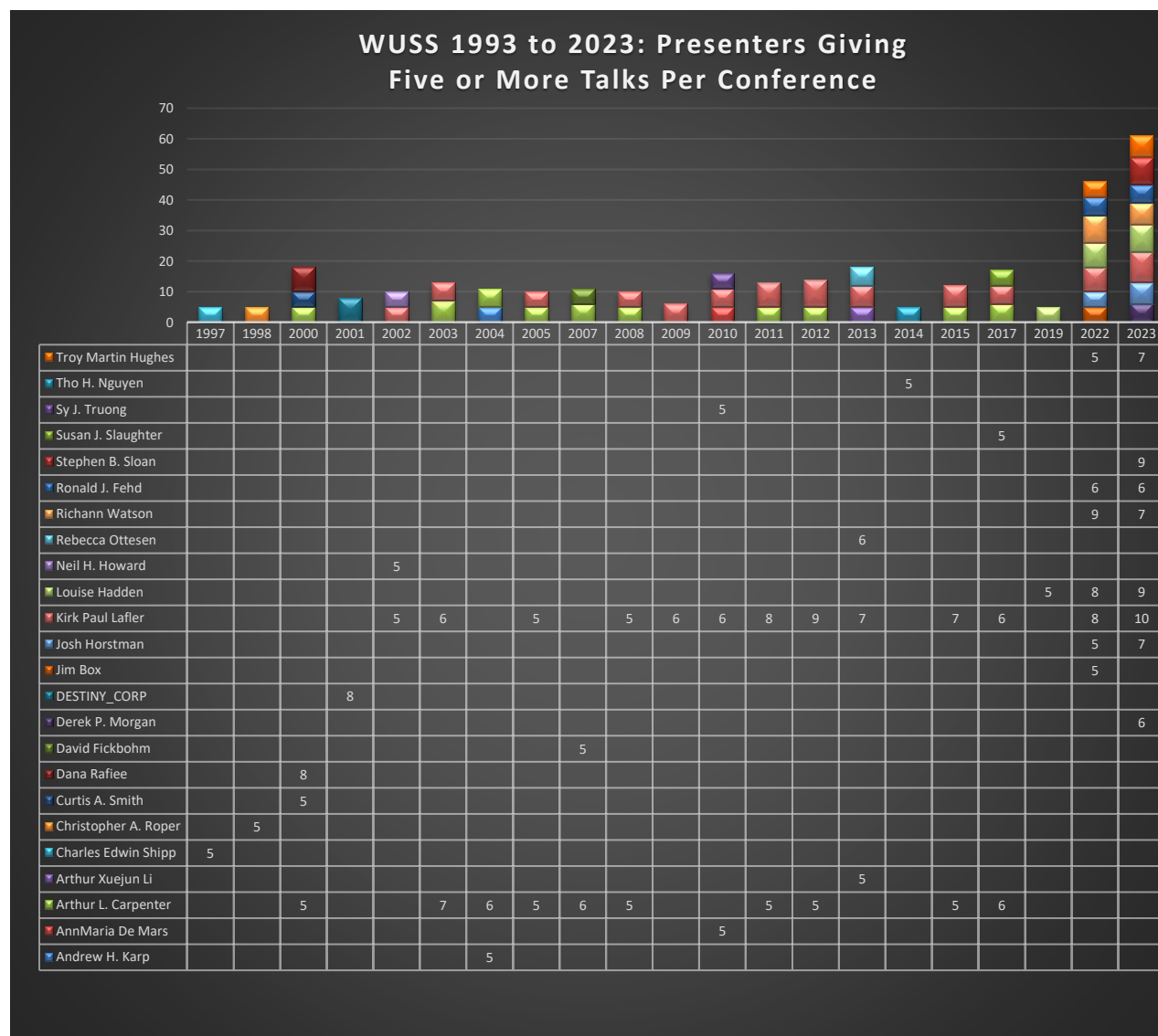
The following ten authors have most prolifically presented at WUSS throughout its history—note that 2023 data have been included in this analysis only:



Looking at WUSS presentations over time, most concerning is 2019, in which the total number of presentations dipped to 116—a huge drop that occurred even before the COVID pandemic.



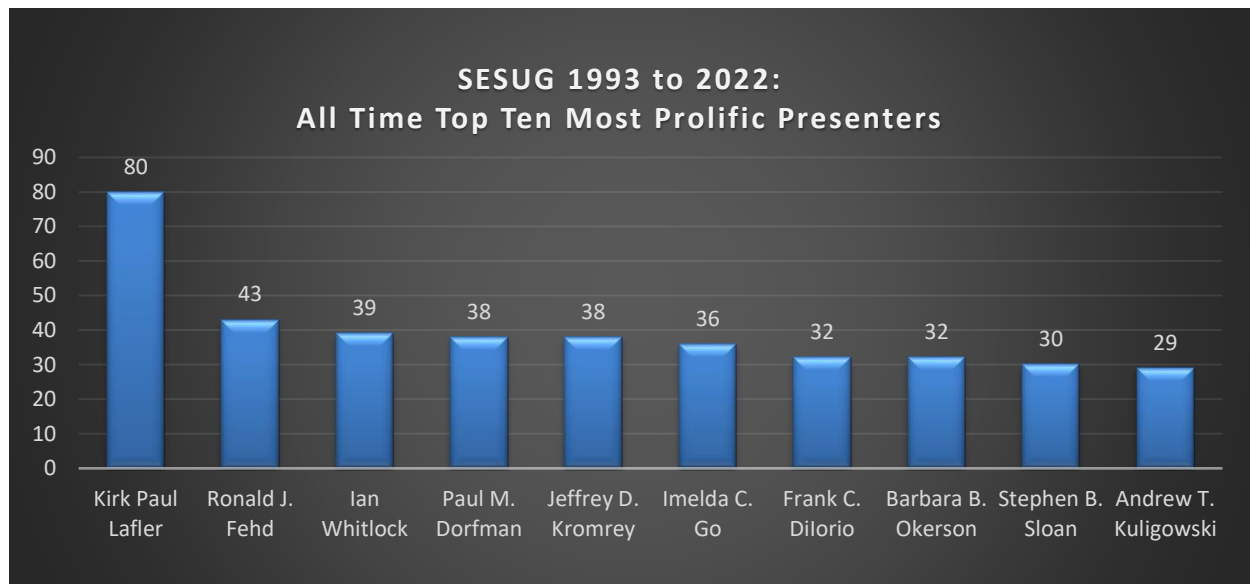
Another concerning metric for WUSS is the prevalence of authors who are now presenting five or more papers per conference, with a huge uptick occurring in 2022 and 2023, with seven authors presenting five or more talks in 2022, and eight authors presenting five or more talks in 2023.



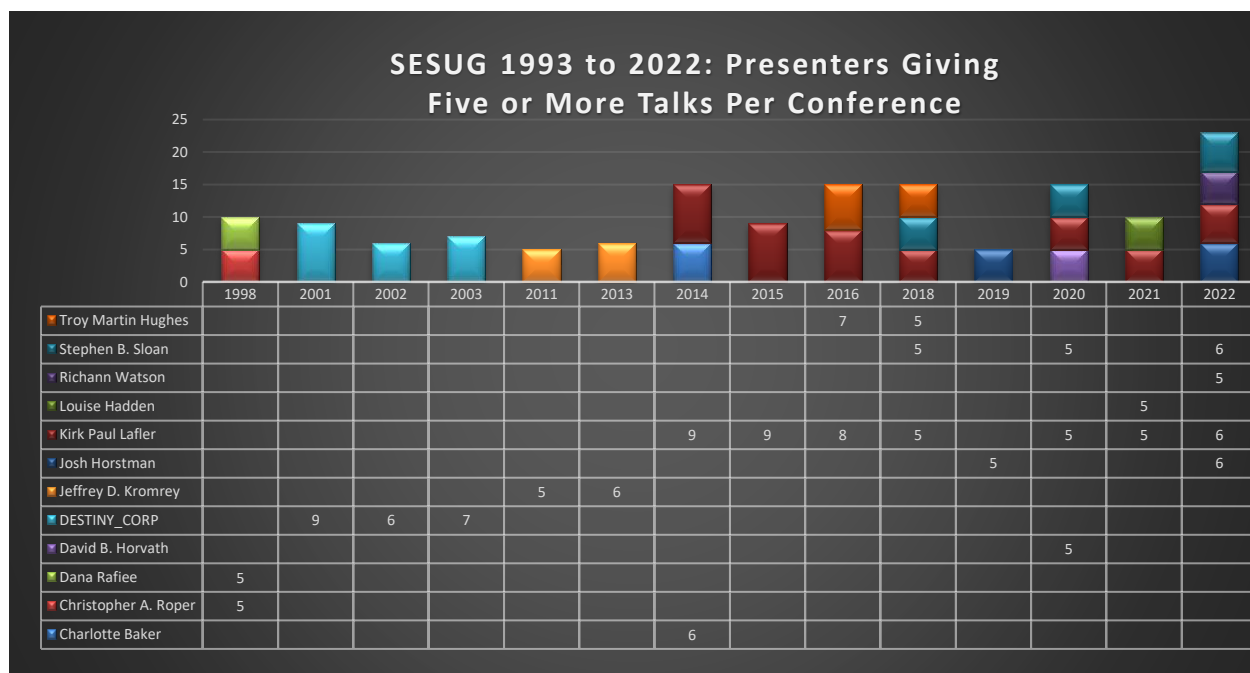
Thus, the concern becomes that a very few number of authors will be presenting the majority of the content for a conference, which might provide an indication that interest is waning by users who once presented, or that new presenters are not being attracted to the conference. That is, when only seven or eight users are delivering more than half of a conference's content, has the conference become a self-licking ice cream cone?!

WHO'S BRINGING THAT BDE TO SOUTHEAST SAS USERS GROUP (SESUG)?

The following ten authors have most prolifically presented at SESUG throughout its history—note that 2023 data are not included:

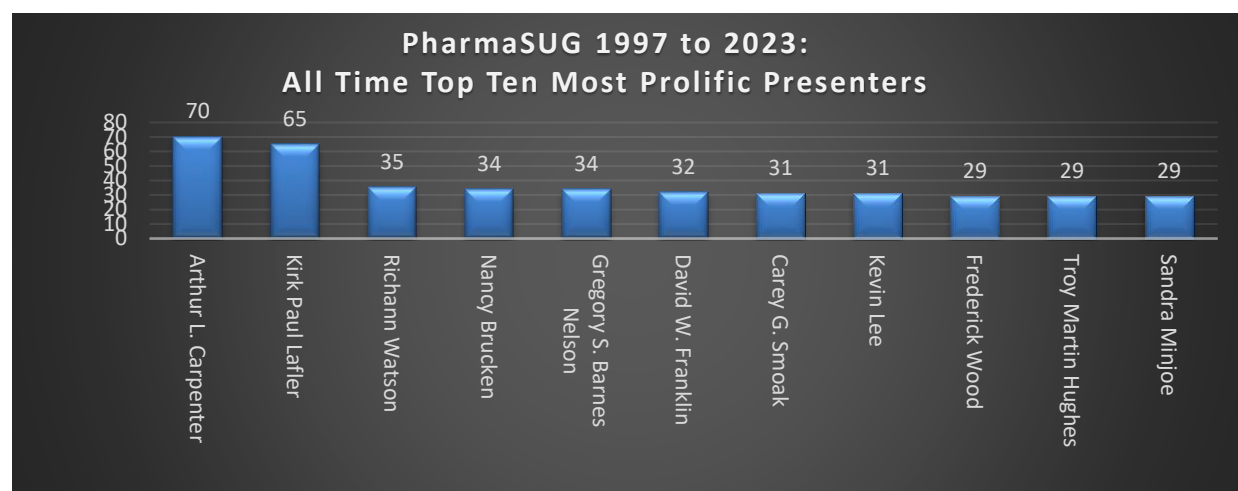


Although SESUG similarly had its highest number of authors giving five or more talks in 2022, this increase was not nearly as substantial as the uptick experienced by WUSS. Thus, in 2022, four presenters—Stephen Sloan, Richann Watson, Josh Horstman, and Kirk Paul Lafler—each gave five or more talks. However, in both 2018 and 2020, three presenters had given five talks a piece at both SESUG conferences.



WHO'S BRINGING THAT BDE TO PHARMASUG?

The following ten authors have most prolifically presented at PharmaSUG throughout its history—note that 2023 data are included:



NEXT STEPS

This initial evaluation of the LEX focused primarily on data quality and the improvement thereof, the longitudinal analysis of the number of presentations and presenters per year and per conference, and the identification of the most prolific presenters by conference and across all conferences. Thus, the metadata analyzed very narrowly comprised only the conference, year, and author(s).

A logical next step would analyze paper content longitudinally to assess whether and which presentations have been being repeated by a given author. This analysis could differentiate, for example, between authors that are presenting new material versus those repeating the same talks at the same or at different conferences. However, to identify repeated talks, it would be insufficient to analyze presentation title alone, as subtle spelling or word variations are common when a paper is presented over time, and as some presenters retitle presentations each time they are given. In other cases, a presenter may give the same talk multiple times under the same title, yet continue to change or expand the content. Thus, given these convolutions, even differentiating new presentations from old ones would become a lengthy analysis.

CONCLUSION

The LEX provides an invaluable resource to SAS practitioners in search of programming best practices, and is readily navigable by conference, year, author, and keywords. This analysis explored presentation metadata, including conference, year, and author, and provided the first-ever longitudinal evaluation of these data. Through the identification of those SAS authors who are bringing the big data energy (BDE), conference chairs and leadership for future SAS conferences can identify and invite those highest-contributing users. And by attracting additional and/or new high performers, conference leadership can mitigate against the recent trend (evidenced especially at WUSS) of too few presenters providing too much content. As the LEX draws its data from third-party sources, challenges to data quality are anticipated, and several threats to data quality were both identified and mitigated. And with these data quality improvements, and the substrate of this analysis, the groundwork is laid for future LEX evaluation. Finally, Appendix A lists the top 150 (i.e., most prolific) presenters across all time and all SAS conferences.

CONTACT INFORMATION

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

Name: Troy Martin Hughes

E-mail: troymartinhughes@gmail.com

APPENDIX A. TOP 150 MOST PROLIFIC SAS PRESENTERS

The following table alphabetically lists the 150 most prolific SAS presenters, and includes their total paper count and ranking.

Author	Count	Rank
Tricia Aanderud	36	104
Jeffrey M. Abolafia	35	107
Melvin T. Alexander	27	147
Elizabeth N. Axelrod	34	112
Caroline C. Bahler	30	131
William E. Benjamin, Jr.	81	27
John E. Bentley	33	117
LeRoy Bessler	100	15
Jonas V. Bilenas	36	106
Thomas E. Billings	46	72
Kenneth W. Borowiak	28	139
Eric C. Brinsfield	28	142
Tony Brown	32	121
Nancy Brucken	50	65
Dan Bruns	41	85
Helen D. Carey	27	152
Arthur L. Carpenter	352	2
David L. Cassell	53	59
Deb Cassidy	50	67
Patricia B. Cerrito	31	125
Goutam Chakraborty	85	22
Wei Cheng	31	126
Bill Coar	34	115
Benjamin Cochran	84	23
Ron Cody	87	19
John J. Cohen	54	57
Barry R. Cohen	37	98
Ted Conway	27	153
David J. Corliss	41	81
Margaret A. Crevar	46	73
Xiangchen (Bob) Cui	29	137
Michael L. Davis	42	80
Vincent DelGobbo	121	10
Lora D. Delwiche	67	35
Nate Derby	67	34
Richard A. DeVenezia	34	114
David A. Dickey	47	71
Frank C. Dilorio	83	24

Paul M. Dorfman	90	17
Harry J. Droogendyk	48	69
Toby A. Dunn	35	108
Joseph E. Earley	28	140
Peter W. Eberhardt	117	11
Daphne E. Ewing	55	56
Brian P. Fairfield-Carter	36	102
Marje Fecht	66	37
Ronald J. Fehd	149	5
George C. J. Fernandez	41	83
Steven J. First	48	68
Peter L. Flom	29	136
Malachy J. Foley	39	89
David W. Franklin	61	44
Eric A. Gebhart	28	141
John R. Gerlach	61	43
Bruce F. Gilson	88	18
Imelda C. Go	38	94
Sunil Kumar Gupta	85	21
Louise Hadden	192	3
Angela M. Hall	31	124
Jack F. Hamilton	27	149
Kriss Harris	37	96
Diane Hatcher	39	90
Lauren E. Haworth	56	54
Edward E. Heaton	30	129
Chris J. Hemedinger	33	118
Donald J. Henderson	58	49
John Henry King	27	148
Sigurd W. Hermansen	27	151
Josh Horstman	125	9
Neil H. Howard	64	38
Troy Martin Hughes	133	6
Jay Iyengar	33	116
Lex Jansen	28	143
Andrew H. Karp	111	12
Mark Keintz	36	105
Chuck Kincaid	52	63
Jeffrey D. Kromrey	62	41
Warren F. Kuhfeld	36	103
Andrew T. Kuligowski	125	8
John M. LaBore	37	101
Kirk Paul Lafler	621	1

Richard D. Langston	83	25
Russell M. Lavery	75	28
Kimberly Stinson LeBouton	28	145
Kevin Lee	41	87
Scott P. Leslie	32	122
Arthur Xuejun Li	59	47
Judy Loren	35	111
Peter J. Lund	57	53
Bruce S. Lund	31	128
Darrell Massengill	27	150
Sanjay N. Matange	59	48
Miriam McGaugh	30	134
Sandy McNeill	27	146
Gary J. Mehler	37	99
Sandra Minjoe	42	78
Michael J. Molter	29	135
Derek P. Morgan	68	33
Roger D. Muller	41	82
William C. Murphy	31	127
Gregory S. Barnes Nelson	126	7
Tho H. Nguyen	27	154
Barbara B. Okerson	60	46
Christopher R. Olinger	32	119
Rebecca Ottesen	30	130
Chevell T. Parker	38	95
Ray Pass	110	13
David J. Pasta	37	100
Douglas Paul Kent	46	74
Jeffrey A. Polzin	30	132
Frederick E. Pratter	41	84
Dana Rafiee	66	36
Michael A. Raithel	57	52
Nancy Rausch	35	109
Dianne Louise Rhodes	32	120
S. David Riba	62	40
Kari Richardson	45	76
Robert N. Rodriguez	72	30
Christopher A. Roper	51	64
Eric Rossland	57	51
John P. Sall	28	144
Marge Scerbo	57	50
Christopher W. Schacherer	29	138
Deanna Schreiber-Gregory	63	39

Jason Secosky	32	123
Charu Shankar	38	93
Mira Shapiro	34	113
Charles Edwin Shipp	154	4
Jack N. Shoemaker	43	77
Susan J. Slaughter	82	26
Stephen B. Sloan	74	29
Curtis A. Smith	52	60
Carey G. Smoak	45	75
Maura E. Stokes	86	20
Janet E. Stuelpner	56	55
Diana D. Suhr	38	92
Abbas S. Tavakoli	39	88
Stephanie R. Thompson	41	86
Patrick Thornton	35	110
Randall D. Tobias	48	70
Sy J. Truong	68	32
Brian K. Varney	52	62
Robert Virgile	50	66
Richann Watson	104	14
Perry Watts	62	42
Ian Whitlock	97	16
Christianna S. Williams	54	58
Steven A. Wilson	37	97
Thomas J. Winn, Jr.	42	79
Frederick Wood	30	133
Wendi L. Wright	38	91
Shi-Tao Yeh	60	45
Michael S. Zdeb	52	61
Cynthia L. Zender	70	31