



PhilaSUG newsletter



Spring 2018 Meeting Announcement

PhilaSUG Spring 2018 Meeting Monday, June 18, 2018

The Philadelphia Area SAS Users Group Spring Meeting will be on Monday, June 18, 2018 at 1:00 PM, and will be hosted by the School of Business Administration at Widener University, Latham Hall, E 13th Street and Potter Street, Chester, PA 19013. A map, detailed driving, and parking instructions are available later in the newsletter.

Registration will begin at 12:15 PM. The meeting will commence at 1 PM. Membership for the year is \$30. There are no other fees for attending PhilaSUG meetings. This year we will be accepting cash, check (made out to "PhilaSUG"), or credit/debit card (this charge will show up on your credit/debit card statement as paid to "Michael Davis / PhilaSUG"). If you are a student of Widener, fees for this meeting will be waived, but please complete the meeting sign-up form so that we can get a better food count. Receipts will be available at registration time.

Abstracts and bios can be found later in this newsletter.



Important – In order to obtain an accurate food count we are asking all attendees to please complete the required [Meeting Attendee Sign-up Form](#) by **June 15th**.

Agenda

We thank our host for providing lunch and break refreshments

12:15-1:00	Registration and lunch
1:00-1:10	Opening Remarks by Jayati Ghosh, Ph.D, Dean of the School of Business Administration
1:10-2:00	Don Henderson - The SAS Supervisor
2:00 -2:30	Richard Goeke - What Teaching SAS Has Taught Me: Lessons Learned from Academia and the Business World
2:30-3:00	Paul Dorfman - Efficient Selective Unduplication Using the MODIFY Statement
3:00-3:30	Break, posters, and refreshments
3:30-4:20	Paul Dorfman & Don Henderson - Divide and Conquer Using a Hash Function
4:20-4:30	Open Forum, Raffle Prizes and closing Remarks

Poster

Weifeng Xu - Lung Cancer Diagnosis

The presenters and the PhilaSUG Executive Committee will adjourn for dinner at UNO Pizzeria & Grill, 1516 Providence Ave, Chester, PA 19013 when the meeting concludes. Google indicates the restaurant is 0.2 miles away.

About Our Host



Widener University

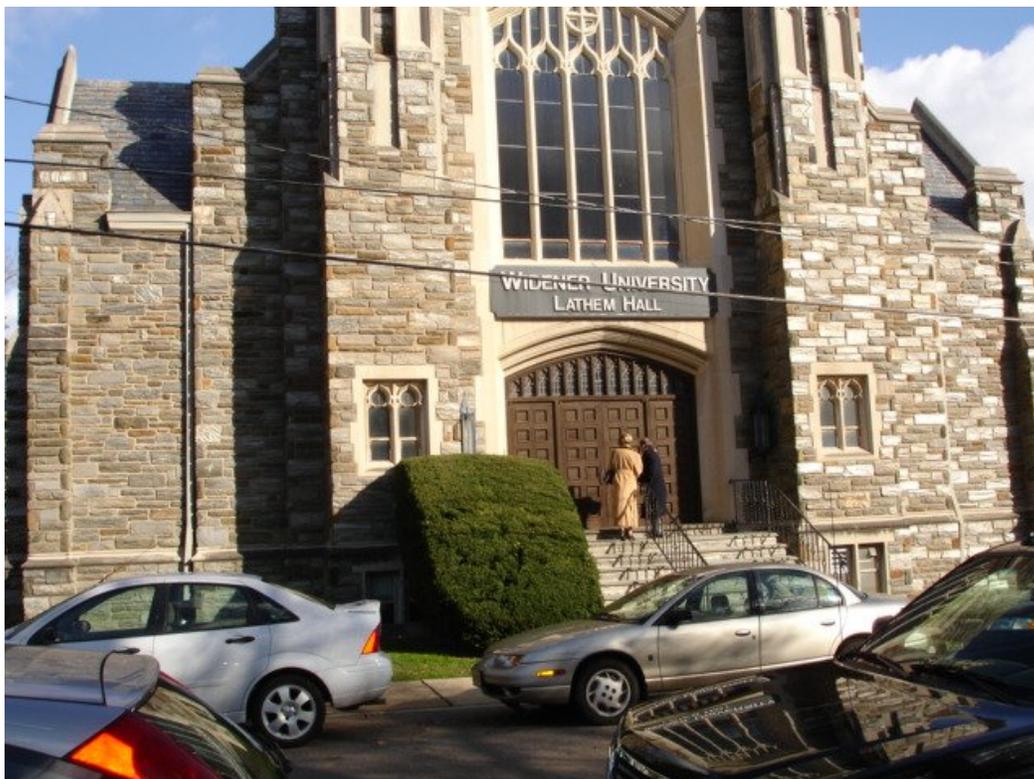
Widener University is a private, metropolitan university that connects curricula to social issues through civic engagement. Dynamic teaching, active scholarship, personal attention, leadership development and experiential learning are key components of the Widener experience. A comprehensive doctorate-granting university, Widener comprises eight schools and colleges that offer liberal arts and sciences, professional and pre-professional curricula leading to associate, baccalaureate, master's and doctoral degrees. The university's campuses are in Chester and Harrisburg, Pa., and Wilmington, Del. Visit the university website, www.widener.edu.

The Widener University School of Business Administration prepares undergraduate and graduate students for thriving careers in business and hospitality management. By pairing cutting-edge industry knowledge and hands-on learning experiences with a focus on leadership development, we provide a foundation for success.

Widener University's School of Business Administration now offers a joint certificate program with the SAS Global Academic Program in Business Intelligence and Analytics. This is the first joint certificate program of its kind offered at the undergraduate level in the greater Philadelphia region.

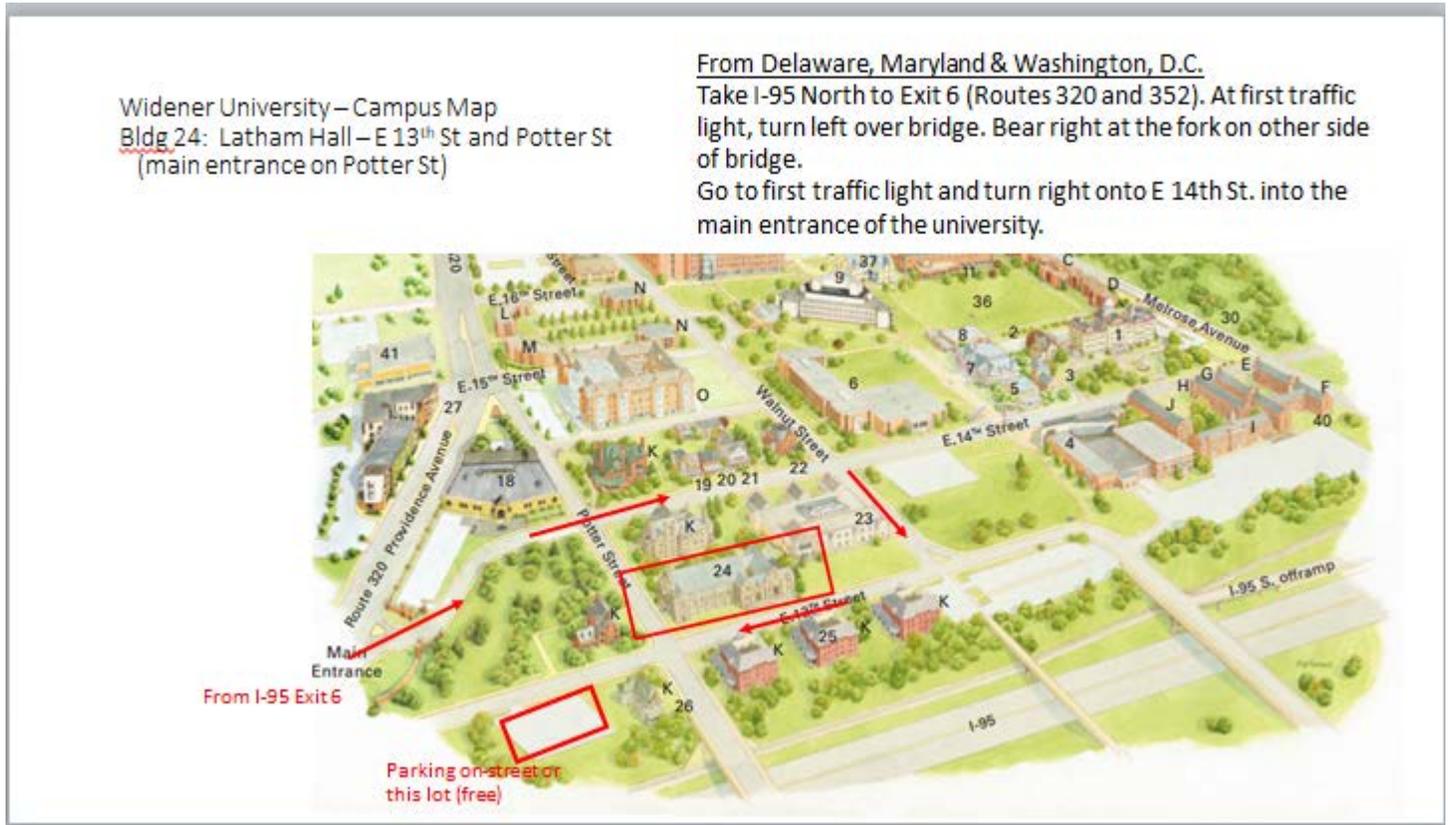
The SAS Global Academic Program created the SAS Joint Certificate Program to better assist universities in preparing students to work in a data-rich business environment. The joint certificate documents students' coursework using SAS' award-winning software to solve real-world business problems.

"This certificate provides additional assurance for employers that Widener students are graduating with extensive training on SAS software and have a variety of technical skills that allow them to handle big data, make sense of that data and communicate findings in writing and verbally," said Dr. Richard Goeke, associate professor.



Directions to Our Host Site

Our meeting will be held at Widener University, School of Business Administration, Latham Hall (Building #24 on the Campus Map) E 13th St and Potter St. (main entrance on Potter St), Chester, PA 19013. Widener's Main Campus, located in Chester, Pennsylvania (14 miles south of Philadelphia, Pennsylvania), serves all daytime undergraduate students, as well as graduate students and adult learners and business professionals who attend University College evening and weekend classes.



- **From New York, Northern New Jersey, and New England**

Take the New Jersey Turnpike South to Exit 2. Proceed on Route 322 West across the Commodore Barry Bridge. Coming off the bridge, follow signs for I-95 North and Philadelphia. Take I-95 North to Exit 6 (Routes 320 and 352). At first traffic light, turn left over bridge. Bear right at the fork on other side of bridge.

Go to first traffic light and turn right onto E 14th St. into the main entrance of the university.

- **From Philadelphia and Philadelphia Airport**

Take I-95 South to Exit 7 (I-476 N toward Plymouth Meeting). Take the **MacDade Blvd.** exit, **EXIT 1**. Keep **left** to take the **MacDade Blvd. West** ramp. Merge onto **MacDade Blvd.**, which becomes **E 22nd St**. Turn **left** onto **Providence Ave/PA-320**.

Turn **left** onto **E 14th St**.

- **From Western & Northern Pennsylvania**

Follow the Pennsylvania Turnpike to Exit 333, I-476. Take I-476 South to Exit 1, **MacDade Blvd.** exit. Keep **right** to take the **MacDade Blvd. West** ramp. Merge onto **MacDade Blvd.**, which becomes **E 22nd St.** Turn **left** onto **Providence Ave/PA-320.**

Turn **left** onto **E 14th St.**

- **From Delaware, Maryland & Washington, D.C.**

Take I-95 North to Exit 6 (Routes 320 and 352). At first traffic light, turn left over bridge. Bear right at the fork on other side of bridge.

Go to first traffic light and turn **right** onto **E 14th St.** into the main entrance of the university.



Taylor Arboretum at Widener University is a 30 acre reserve of plantings and natural lands

PhilaSUG Executive Committee

Michael Davis, President
John Cohen, Membership Coordinator
Diane Foose, Treasurer
Robert Schechter, Newsletter Editor, and Web Master
Jonas V. Bilenas
Barry Cohen
Mark Keintz
Haibin Shu
Max Cherny
David Horvath
Steve Rhoades

The PhilaSUG EC team appreciates the efforts of all contributors, without whom this newsletter would not be possible. We are always looking for volunteers to contribute content to the quarterly newsletter. If you have ideas or suggestions for the newsletter, please share them with us. We would love to hear them!

Host Sites Wanted

We continuously seek host sites for future PhilaSUG meetings. There is not a lot of work involved, and it is a great way to put your company on the local SAS map. We need your help with this. If your company would like to host a meeting, within reasonable geographic proximity to Philadelphia, PhilaSUG would be grateful if you would contact Michael Davis at michael.davis@alumni.duke.edu.

E-mail Announcements



PhilaSUG-L is a low volume, announcement-only e-mail notification service provided free of charge to all members who wish to subscribe. In order to sign up for this service, you need only send a blank e-mail message to PhilaSUG-L-subscribe@onelist.com. Note that you can subscribe as many times with as many different e-mail addresses as you wish to have the e-mail sent to; e.g., home and office.

PhilaSUG Web Site

Our site on the World Wide Web always contains the latest information concerning upcoming meetings, SAS training and seminars, links to SAS related hot topics, and local SAS job opportunities.



Visit us regularly at: <http://www.PhilaSUG.org>

Presenters Wanted

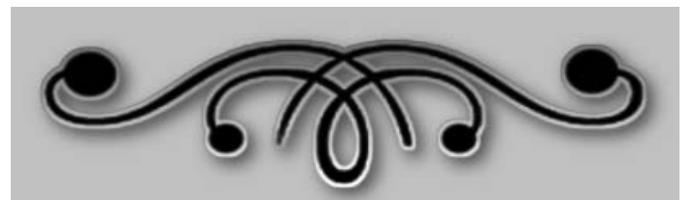


The PhilaSUG Executive Committee invites presentation abstracts. If you have thought presented before, PhilaSUG is an excellent venue for first-timers.our web site. Presentations can range from ten minute coders corner tips to 50 minutes talks. Both new and previously presented topics are welcome. If this is of interest to you, PhilaSUG invites you to submit electronic abstracts and bio for either a poster or oral presentation by using our online submission form. Check our home page for the Call for Papers.

A Thank You



The PhilaSUG Executive Committee wishes to thank Dr. Tunde Odusami, Ph.D. CFA, Associate Professor of Finance at Widener University, School of Business Administration for arranging the hosting of this meeting. In addition, we wish to thank two members of our Executive Committee, Mark Keintz for speaker/program coordination and Haibin Shu for site coordination.



Paper Abstracts

Divide and Conquer Using a Hash Function

Paul Dorfman, Proprietor, Dorfman Consulting
Don Henderson, Principal, Henderson Consulting Services, LLC

Aggregating or joining large data volumes can be hindered by system constraints and fail. One approach to address the problem is to divide the input into a number of segments, process them independently and combine the results. For such divide-and-conquer tactic to work, two seemingly contradictory criteria must be met. First, to aggregate or combine the data correctly, no segment can share its key-values with the rest. Second, the segments must be more or less equal in size. We demonstrate how a hash function can be used to satisfy both criteria for arbitrary input with no a priori knowledge of the input key-values. Effectively, the method renders any task of aggregating or combining data of any size principally doable by splitting its input into a large enough number of nearly equal segments. The trade-off is the need to partially re-read the data. However, it is a rather small price to pay to make a failing or endlessly running task finish in a reasonable time.



Paul Dorfman is an Independent Consultant. He specializes in developing SAS software solutions from ad hoc programming to building complete data management systems in a range of industries, such as telecom,

banking, pharmaceutical, and retail. A native of Ukraine, Paul started using SAS while pursuing his degree in physics in the late 1980's. In 1998, he pioneered using hash algorithms in SAS programming by designing a set of hash routines based on SAS arrays. With the advent of the SAS hash object, Paul was first to use it practically and to author a SUGI white paper on the subject. In the process, he introduced hash object techniques for metadata-based parameter type matching, sorting, unduplication, filtering, data aggregation, dynamic file splitting, and memory usage optimization. Paul has presented papers at global, regional, and local SAS conferences and meetings annually since 1998.



Don Henderson is the Owner and Principal of Henderson Consulting Services, a SAS Affiliate Partner. Don has used SAS software since 1975, designing and developing business applications with a focus on data warehouse, business

intelligence, and analytic applications. Don was one of the primary architects in the initial development and release of SAS/IntrNet software in 1996, and he was one of the original developers for the SAS/IntrNet Application Dispatcher. Don is the author of SAS Server Pages: Generating Dynamic Content, Building Web Applications with SAS/IntrNet: A Guide to the Application Dispatcher, and Data Management Solutions Using SAS Hash Table Operations: A Business Intelligence Case Study. Don has presented numerous papers at SUGI and regional SAS user group meetings and continues to be a great supporter of SAS and its products.

Coming Soon:

Data Management Solutions Using SAS® Hash Table Operations: A Business Intelligence Case Study

By Paul Dorfman and Don Henderson

Anticipated publication date: Third quarter 2018

Hash tables can do a lot more than you might think! This book concentrates on solving your challenging data management and analysis problems via the power of the SAS hash object, whose environment and tools make it possible to create complete dynamic solutions. To this end, this book provides an in-depth overview of the hash table as an in-memory database with the CRUD (Create, Retrieve, Update, Delete) cycle rendered by the hash object tools. By using this concept and focusing on real-world problems exemplified by

sports data sets and statistics, the book seeks to help you take advantage of the hash object productively, in particular, but not limited to:

- select proper hash tools to perform the hash table operations
- use proper hash table operations to support specific data management tasks
- utilize the dynamic, run-time nature of hash object programming
- understand the algorithmic principles behind hash table data look-up, retrieval, and aggregation
- recognize why the hash object is exceptionally well suited for data aggregation and learn how to do it
- manage hash table memory footprint, especially when processing big data
- use hash object techniques for other data processing tasks, such as filtering, combining, splitting, sorting, and unduplicating.

With hash tables, you can answer your toughest questions in the most efficient way possible!

The SAS Supervisor

Don Henderson, Principal, Henderson Consulting Services, LLC

How SAS processes jobs is the responsibility of the SAS Supervisor and an understanding of its function is important. While the details of how it works have changed over time, some of the basics of the SAS Supervisor have been reasonably consistent over time.

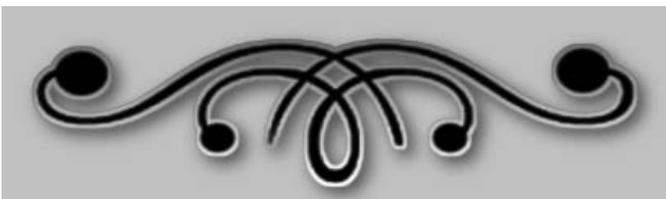
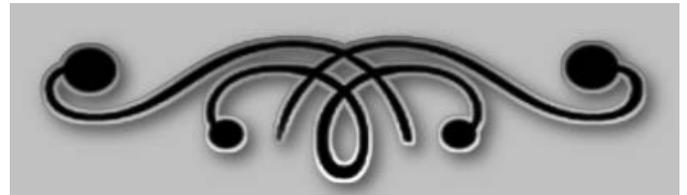
This paper was first presented at SUGI (now known as SGF) in 1983 and was a standard talk at many conferences back in the Jurassic period of SAS. It discusses the functions of the SAS Supervisor during the execution of a SAS DATA Step program.

The functions of the SAS Supervisor can be categorized as follows:

- Compiling SAS Source Code, and
- Executing Resultant Machine Code

The actions of the Supervisor during both the compile and execution phases of a SAS job will be illustrated.

When a SAS DATA Step program is written, the DATA Step "module" must be integrated within the structure of the SAS System. This integration is done by the SAS Supervisor.



Efficient Selective Unduplication Using the MODIFY Statement

Paul Dorfman, Proprietor, Dorfman Consulting

Many an ETL transformation phase starts with cleaning the extracted data from "duplicate" data. For example, transactions with the same key but different dates may be deemed "duplicate", and the ETL needs to choose the latest transaction. Usually, this is done by sorting the extract file by the key and the date and then selecting the most recent record. However, this approach may exact a very high processing cost if the non-key (aka the satellite) variables are numerous and long and especially if the result needs to be resorted to bring the extract into the original order. This paper shows that this kind of data cleansing can be accomplished by using a principally different algorithm based on merely marking the duplicate records for deletion in the original extract file. In real-life scenarios, the proposed approach, particularly when the "duplicates" are relatively rare, may result in cutting the processing time by more than an order of magnitude.

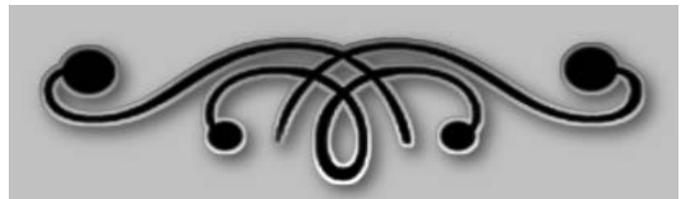
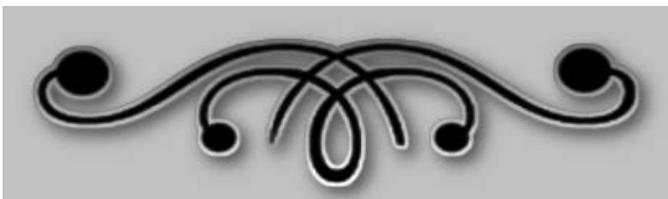
What Teaching SAS Has Taught Me: Lessons Learned from Academia and the Business World

Richard Goeke, Widener School of Business Administration

Whether in industry or academia, learning and using SAS can range from rewarding to exasperating. Gaining SAS expertise depends heavily on many individual factors, including intelligence, problem-solving skills, motivation, systems-thinking, resilience, and humility. In addition, a new generation of learners and technologies offers additional challenges and opportunities. This presentation will shed some light on what has (and has not) worked when it comes to teaching and learning SAS.



Richard Goeke received his PhD in Management Systems from Kent State University, and is an Associate Professor in the Widener University's Management Department, where he teaches undergraduate and graduate courses in business analytics, information systems, database management, and statistics. His research has appeared in scholarly and practitioner journals including Communications of the ACM, Information Resources Management Journal, Quality Management Journal, Business Process Management Journal, and Journal of Global Mobility. Before his doctorate, Goeke worked 17+ years at The Goodyear Tire and Rubber Company, where he used SAS to develop systems and applied analytics in accounting, finance, treasury, marketing, and operations.



Paper Abstracts

Lung Cancer Diagnosis Ascertainment from the Population-based Administrative Databases

Weifeng Xu, BARDS, Merck and Co., Inc.

Population-based administrative databases are increasingly used for real-world evidence generation. However, its application towards cancer research is limited. One main reason is that current diagnosis classification systems used for reimbursement purposes, either ICD9 or ICD10, do not specify key cancer characteristics (e.g., histological types, stages). One potential solution is to fully utilize all the information in claims and construct sophisticated coding algorithms for ascertainment of cancer cases of interest. Recently, successful work on machine and statistical learning on classifying lung cancer severity using the Surveillance, Epidemiology and End Results (SEER) cancer registry program and Medicare claims data in conjunction with classifier ensemble (superlearner) has been reported in the literature. Given the large size of the administrative data, the superlearner algorithm can be used to its full potential. We carried out a simulation study to better understand the behavior and performance of the superlearner. In our presentation we will show how different first level learners effectively adapt to linear and non-linear data structure. We will also discuss how the non-negative least squares combiner inherently addresses the collinearity of the first level learners' outputs. Finally, we will present preliminary work on lung cancer diagnosis ascertainment using the SEER database.

Future Meetings and SAS Events



PhUSE US Connect2018

Where: Raleigh, NC

When: 3rd - 6th June 2018

Meeting Facility: Raleigh Convention Center



Next PhilaSUG Meeting

To Be Announced

PhilaSUG Fall 2018 Meeting

Host: TBA

**When: TBA,
but targeting Mid October
Through Mid November**

