BASUG.org
BOSTON AREA SAS® USERS GROUP

BASUG Meeting Announcement

We are thrilled with your response to the Call-for-Papers for our Q4 Coders' Corner meeting. This meeting will feature many presenters – some first-timers, some seasoned – covering a variety of topics. Please join us to learn lots of tips 'n techniques from your fellow BASUG colleagues.

After the meeting, we will provide an informal light buffet lunch for all attendees. We hope you can stay for this opportunity to network and socialize with your fellow SAS users.

Topic	**Coders' Corner 2015**
When	Thursday, December 3, 2015 8:30AM - Noon
Where	Microsoft New England Research & Development Center (NERD) [1] One Memorial Drive Cambridge MA 857-453-6000
Directions	Please visit the meeting site directions page (http://microsoftnewengland.com/About)
How to Register	Individual, on-line registration required. Sorry, NO WALK-INs. Register Now! (http://www.basug.org/index.php?p=eventreg)
Payment	$10 -- if paid on-line by Monday, November 23, 2015 $15 -- if paid on-line by NOON on Wednesday, December 2, 2015 $20 -- at the door (checks only)
Contact	If you have questions about the meeting, please contact the meeting organizers (mailto:basugwm@basug.org)

Agenda[*]

Time	Activity
8:15AM	Sign-in and Refreshments
8:45AM	Announcements

Time	Activity
9:00AM-Noon	**PRESENTATIONS:**
9:00AM	**"Calculating the Percentage Using PROC TABULATE"**, by David Franklin, Quintiles
9:12AM	**"SAS$^®$ Macro to divide long text between words for CDISC compliance"**, by Maksim Kazanski, PAREXEL
9:24AM	**"Proc Summary vs. Retain By"**, by Neal Jawadekar, Predilytics
9:36AM	**"Where did those formats go?"**, by Jeffrey Lavenberg, Harvard T.H. Chan School of Public Health
9:48AM	Break
10:08AM	**"Custom Graphs Created by Proc Template"**, by Michelle Bissell Peak, Siemens Healthcare
10:20AM	**"Chaining Stored Processes With SAS$^®$ Sessions"**, by Christopher Lampron, Siemens Healthcare
10:32AM	**"%Assert() your way to sleep-filled nights: A one line data validation macro"**, by Quentin McMullen, Siemens Healthcare
10:44AM	**"SAS$^®$ Tips and Techniques: SAS Function to Validate NPI, How to eliminate data error notes from the SAS log"**, by Ali Sabouri, Blue Cross Blue Shield of Massachusetts
10:56AM	Break
11:16AM	**"File Management Using Pipes and 'X' Commands in SAS$^®$"**, by Emily K.Q. Sisson, Boston University School of Public Health
11:28AM	**"Global Health Applications: Creating Multilingual Reports with the SAS$^®$ Unicode Server"**, by Sharon Coleman, Boston University School of Public Health
11:40AM	**"WAPTWAP, but remember TMTOWTDI"**, by Jack Shoemaker, Self-employed
Noon	Informal buffet lunch (provided by BASUG)

[*] Note: Times (and sequence) are approximate and subject to change. Please re-visit the BASUG website (http://www.basug.org/index.php) for updated information.

Abstracts and Speaker Biographies

"Calculating the Percentage Using PROC TABULATE", by David Franklin

PROC TABULATE is a very powerful procedure which can do statistics and frequency counts very efficiently, but it also it has the capability of calculating percentages on many levels for a category.This presentation looks at the automatic percentage calculations that are provided, and then delves into how you as a user can specify the denominator for your custom percentage.

David started programming in SAS in 1985 in the land known now as "Middle Earth". After finding the way to surface he worked in Europe and later found his way to New England where he now calls home. Since 2004 David has been editor of TheProgrammersCabin.com which is dedicated to the SAS programmer, providing many tips to help learn new ideas. Until recently David was a consultant working with many Blue Chip companies in the US and Europe but in 2014 he put his hat up and travel bag away and accepted a position at Quintiles Real World Late Phase division in Cambridge, MA. He currently is a Manager of Statistical Programming and Editor for a monthly SAS Users Newsletter within the division.

"SAS® Macro to divide long text between words for CDISC compliance", by Maksim Kazanski

CDISC requires splitting text strings longer than 200 characters into values up to 200 long. The text should be split between words to improve readability as per CDISC SDTM Implementation Guide (Version 3.2) section 4.1.5.3.2. In the paper, the author will show how create and use the code to achieve this desired result in SAS data step.

Maksim is a Principal Data Analyst at PAREXEL Informatics. He has been serving PAREXEL since 2007 and working with SAS since then, utilizing his previous data and programming experience for ORACLE-based systems. In 2009 he was invited to work locally in the Boston area and moved. As a truly believer in the power and overall benefit of Standardization, he has developed a set of functions, macros, code templates, and best practices for the variety of clinical data export tasks in CDISC and non-CDISC formats. Now all this is employed leading to the better short and long-term results of Data Analytics team's developments and cross-supports. Maksim came from Minsk, Belarus. He enjoys working with data.

"Proc Summary vs. Retain By", by Neal Jawadekar

"Proc summary" and "retain by" are two methods which can be used in SAS to summarize data. Both have their advantages and disadvantages. Proc summary's syntax is more succinct, and can be used to output simple statistics, whether it be summarizing data by certain variables, calculating the means, and more. Meanwhile, "retain by" can be advantageous when trying to maintain the meat and bones of a dataset, while also manipulating some variables along the way. The purpose of this presentation will be to dive deeper into the various functionalities of proc summary and retain by.

Neal Jawadekar is a Big Data Analyst at Predilytics, a Welltok company. He uses SAS to aggregate and analyze healthcare data, predicting a variety of outcomes (from hospital readmission to medication adherence to program engagement). Neal is excited about Welltok's mission, which is to provide data-driven incentives for individuals to optimize their health. Neal received his Master of Public Health degree in Epidemiology and Biostatistics from Tufts University in 2014. In his free time, he enjoys playing tennis and the guitar.

"Where did those formats go?", by Jeffrey Lavenberg

Three macros, %LocFMT, %MissFMT, and %MultiChkFMT, will make working with formats much easier. %LocFMT outputs which format catalog the format definition is being applied from. The path to the format catalog is also provided. Given a library reference, the %MissFMT macro searches datasets within that location and outputs which formats are not found in the available format catalog(s). The %MultiChkFMT macro scans the defined format catalog(s) for overlapping format definitions and highlights which one will be used. These macros are useful whether you are receiving datasets, creating new ones, or distributing them.

Jeff Lavenberg is the Assistant Head of the CBAR Programming Core at Harvard T.H. Chan School of Public Health. He has using SAS software for data management, analysis, and reporting with clinical trials since 2009.

"Custom Graphs Created by Proc Template", by Michelle Bissell Peak

Proc Template can be used to construct custom graphs that create a visual representation of an analysis. Two examples to be detailed in this paper are a scatter plot with a broken or split axis and a regression plot paired with box plots of categorical variables. Both of these examples will use lattice functionality within Proc Template. By using tools such as Proc Template to create graphics beyond the standard output in SAS a thorough and concise visual summary can be produced to complement an analysis.

Michelle is a Senior Biostatistician at Siemens Healthcare. The primary focus of her role is pertaining to technical operations and manufacturing support in relation to Immunoassay technology. Michelle has chemistry and statistics background including a Master's degree in Biostatistics from Grand Valley State University. Currently, she is obtaining a second Master's degree in Biomedical Diagnostics from Arizona State University. In her free time Michelle enjoys hiking, random outdoor adventures and playing with the most adorable Miniature Schnauzer, Bosco.

"Chaining Stored Processes with SAS$^{®}$ Sessions", by Christopher Lampron

Uploading a file from a personal machine to the SAS server to be used within multiple SAS stored processes is not as simple as a PROC IMPORT. The easiest way to import the file from the local machine to the server is through HTML forms. A single stored process can be dedicated to uploading a file, while subsequent stored process can be called to access the dataset and use within SAS programs. This can be done by chaining stored processes together. Chaining stored processes consists of linking multiple stored processes together and passing variables from one process to the next. The dataset can be passed from one process to the next through SAS sessions. A SAS session is the data that is saved from one SAS stored process that is carried over to the next process via macro variable and library members. The session is designed so that all users have independent sessions

This paper demonstrates how to upload an Excel worksheet (*.xlsx) into a SAS dataset and access the dataset in multiple chained stored processes via SAS sessions. In the paper, the author will show how to create the file upload capability using an HTML input statements within a form and how to chain several stored processes together. The author will also show how to create a session in the first stored process which can be used to temporarily save the uploaded dataset and access it in subsequent stored processes. This paper can serve as a guide on setting up basic stored process that utilizes HTML code and SAS sessions.

Chris is a Biostatistician for the Siemens Healthcare company. Having been with Siemens since 2005, Chris has worked in multiple business units within the company. Chris started he career in a customer service roll, then moving in the laboratory in Molecular Quality Control. After several years in quality, Chris moved into a Technical Operations group within Molecular to identify the root cause of failures in product and implement changes to resolve the issues. After seven years in the Molecular business, Chris joined the Biostatistics group in Siemens. Along with providing statistical support for the business, Chris has been developing web applications in SAS that have been used to streamline processes and reduce downtime. In his spare time, Chris enjoys regular visits to the local comic book store purchasing comic books to add to his collection.

"%Assert() your way to sleep-filled nights: A one line data validation macro", by Quentin McMullen

Data are messy, and should never be trusted. This paper presents a simple SAS macro, %Assert, which allows the user to state a Boolean expression that is expected to be true, e.g.: %assert(Age>0). If the expression is false, an error message is printed to the log. The macro may be used to verify expectations about data at any point in a DATA step. Use of the %Assert macro provides two important benefits: it provides a method for automated error detection, enhancing confidence that results are correct; and it allows the programmer to explicitly state expectations regarding the data being processed, enhancing the readability of code. Assertions are a critical tool for both defensive coding and test-driven development. Principles of macro design encountered during the development of the macro are addressed. [The full NESUG 2012 paper is available here (http://www.lexjansen.com/nesug/nesug12/cc/cc31.pdf).]

Quentin has been programming in SAS for more than 15 years. Currently his SAS focus is on macros, Stored Processes, and Business Intelligence web apps. His personal focus is on his family, which includes a new dog.

"SAS® Function to Validate NPI", by Ali Sabouri

An NPI is a unique 10-digit health care identifier mandated by the administrative simplification provisions of HIPAA legislation. The federal regulation requires all providers who submit standard HIPAA transactions to have an NPI. NPIs eliminate the need for providers to use different identifiers numbers for different health plans. NPIs are used for electronic claims, eligibility and responses, claim status inquiries and responses, referrals, and remittance advices.

Healthcare data analysts often receive a long list of NPI's to perform related provider tasks. Having a valid list of NPI's would save time and effort to do just that. you can use the Luhn algorithm to distinguish a good provider NPI from a bad one (where we're qualifying the ID, not the actual provider)

"How to eliminate data error notes from the SAS® log", by Ali Sabouri

SAS programmers who create and run rather large codes often would have several hundred lines of logs returned after the execution. This simple code would allow you to direct all potential error notes to "error-only" file in which would be easier to look into and quickly fix it.

Ali Sabouri is a Senior Network Technical Consultant at Blue Cross Blue Shield of Massachusetts in Boston where he does data analyses, programming and reporting. Ali is SAS Certified Programmer and has been using SAS since 1993 mostly within the healthcare industry.

"File Management Using Pipes and 'X' Commands in SAS®", by Emily K.Q. Sisson

SAS for Windows can be an extremely powerful piece of software, not only for analyzing data, but for storing, organizing, and maintaining output and permanent datasets. By employing pipes and 'x' commands within a SAS session, you can easily and effectively manage files of all types stored on your local network. Specific examples using pipes and 'x' commands to archive datasets and SAS output will be presented.

Emily K.Q. Sisson, M.A., graduated from the Boston University Graduate School of Arts & Sciences in 2005 with a concentration in Mathematics & Statistics. She is a Statistical Manager at the Data Coordinating Center at BUSPH and has collaborated on numerous public health projects during her 5-year tenure. Ms. Sisson has worked

as an analyst on fifteen different research projects since 2005 and has published over 20 journal articles. The Data Coordinating Center has been a data management resource center since 1984 with responsibilities that include design and collection of data collection protocols, subject and data tracking systems, site monitoring, implementation of data entry systems, statistical analyses, and study closeout.

"Global Health Applications: Creating Multilingual Reports with the SAS® Unicode Server", by Sharon Coleman

Global public health initiatives are critical approaches to improving the health and well-being of populations worldwide. Research that incorporates data from various countries with different languages and scripts can be challenging. The implementation of the SAS Unicode server, a SAS session that uses the Unicode Transformation Format (UTF-8), allows the encoding of characters from multiple languages and bridges the language gap. Installation of the Unicode server for Windows and the creation of multilingual reports, including the use of Cyrillic script, will be discussed.

Sharon Coleman, M.S., MPH, graduated from the Boston University School of Public Health (BUSPH) in 2004 with a concentration in Epidemiology and Biostatistics. She is a Research Manager at the Data Coordinating Center at BUSPH and has collaborated on numerous public health projects, including studies in Russia and South Africa. Sharon has worked as a research manager and analyst on ten different research projects since 2004 and has co-published over 40 journal articles. She has presented study esults at several national meetings including the American Public Health Association annual meetings.

"WAPTWAP, but remember TMTOWTDI", by Jack Shoemaker

Writing a program that writes a program (WAPtWAP) is a technique that allows one to solve computing problems in a flexible and dynamic fashion. Of course, there is more than one way to do it (TMTOWTDI). This presentation will explore three WAPtWAP methods available in the SAS system.

Jack is currently building out a business intelligence environment at an upstate health plan in Greenville, SC. Prior to that, Jack was the Vice President of the Healthcare Practice at d-Wise, a technology consultancy based in RTP which serves the Life Sciences and Healthcare industries. Prior to d-Wise, Jack had a long run at WellCare Health Plans where he held several leadership roles in the information technology and operations departments using SAS technologies to build a robust business intelligence platform and support various merger, acquisition, and integration initiatives. Jack is a graduate of the Massachusetts Institute of Technology with a BS in Economics and of Harvard University with a MS in Health Policy and Management.

BASUG Contacts

Mailing Address:

> BASUG
> PO Box 170253
> Boston MA 02117

Email Our Webmaster (mailto:basugwm@basug.org)

[1] *The Microsoft New England Research & Development Center (NERD) is a research and software innovation campus located in the heart of Cambridge, Massachusetts. The NERD vertical campus spans two buildings with its primary presence and conference center located at One Memorial Drive and a recently renovated and expanded space located at One Cambridge Center. NERD is home to some of Microsoft's most strategic teams including Microsoft Research New England, Microsoft Application Virtualization (App-V), SharePoint Workspace, Microsoft Technical Computing, Microsoft Advertising, Microsoft Lync, Microsoft Office 365 and more. NERD has become a hub of activity for the local tech community and has hosted more than 500 events and welcomed more than 40,000 visitors during the past two years.*