

SESUG 2019 Conference Program



SESUG 2019
WILLIAMSBURG, VA
OCTOBER 20-22

Rachel Straney
Academic Chair

Chuck Kincaid
Operations Chair

www.sesug.org

LEARN, CONNECT AND DISCOVER

with other SAS® enthusiasts at the largest gathering of analytics professionals.

SAS®
**GLOBAL
FORUM**
2020

MARCH 29 - APRIL 1 | WASHINGTON, DC
WALTER E. WASHINGTON CONVENTION CENTER



#SASGF



SEE ANALYTICS APPLIED
to solve real-world challenges.



TAKE A DEEP DIVE
into sessions on leading-edge technologies.



BUILD CONNECTIONS
with a global community.



LEAVE INSPIRED
and ready to make an impact that people around you see and feel.

TABLE OF CONTENTS

Conference planning Team	4
SESUG Executive Council	6
Special Thanks	6
Conference Charity	7
Ribbons	8
Conference Information and Special Events	9
Academic Grant Winners	11
Professional Development Grants	12
Educational Opportunities	13
SAS Super Demos at SESUG	15
Conference Courtesies.....	17
Section Descriptions.....	20
Abstracts by paper number.....	27
Analytics Leadership	27
Coder's Corner	33
Data Management/Big Data.....	40
e-Posters	44
Education/Institutional Research.....	50
Hands on Workshops.....	54
Government.....	58
Healthcare/Pharmaceuticals	59
Know Your SAS: Foundations	62
Know Your SAS: Advanced Techniques.....	70
Open Analytics	81
Planning/Administration	86
Reporting/Visualization	89
Statistics/Data Analysis	98
SESUG Sponsors.....	113
Meeting Room Layout	120
Savannah 2020.....	122

CONFERENCE PLANNING TEAM

Conference Chairs

OPERATIONS

Chuck Kincaid

OperationsChair2019@sesug.org



ACADEMIC

Rachel Straney

AcademicChair2019@sesug.org



Operations Team

A/V Coordinator

Abbas Tavakoli
Venita DePuy

Graphics & Emails

Chuck Kincaid
Rachel Straney

Registration

Sarah Woodruff
Kristen Harrington

Sponsorships

Meenal Sinha
Chuck Kincaid
Venita DePuy

Food & Beverage

Chuck Kincaid
Venita DePuy
Kristen Harrington

Marketing

Chuck Kincaid
Rachel Straney

Signage

Richann Watson
Meenal Sinha

Volunteer Coordinators

Lesla Caves
Matthew Bell

Web Application

Brian Varney
Richann Watson

Grants

Barbara Okerson
Charlotte Baker

Publications

Lesla Caves
Linda Sullivan

Social Media

Jason Brinkley
Rachel Straney

Webmaster

Brian Varney
Abbas Tavakoli

Academic Section Chairs

Analytics Leadership

Linda Sullivan
Merry Rabb

e-Posters

Abbas Tavakoli
Brian Varney

Hands On Workshops

Mike Sadof
Andrea Lewton

Know Your SAS®: Advanced Techniques

Richann Watson
Harry Droogendyk

Pre/Post-Conference Workshops

Venita DePuy
Linda Sullivan

Coder's Corner

Matthew Bell
Venita DePuy

Education/Institutional Research

Meenal Sinha
Mel Alexander

Healthcare/ Pharmaceuticals

Meenal Sinha
Mel Alexander

Open Analytics

Jason Brinkley
Brian Varney

Reporting and Visualization

Barbara Okerson
Louise Hadden

Data Management/Big Data

Lesia Caves
Ethan Ritchie

Government

Meenal Sinha
Mel Alexander

Know Your SAS®: Foundations

Charlotte Baker
Kelly Smith

Planning and Administration

Josh Horstman
Carlos Piemonti

Statistics and Data Analysis

Kristen Harrington
Nat Wooding

SESUG EXECUTIVE COUNCIL

Kristen Harrington, *President* Sarah Woodruff, *Vice President* Venita DePuy, *Treasurer*

Barbara Okerson, *Secretary*

Charlotte Baker

Jason Brinkley

Lesia Caves

Chuck Kincaid

Meenal Sinha

Rachel Straney

Linda Sullivan

Abbas Tavakoli

Brian Varney

Richann Watson

SPECIAL THANKS

SAS Liaison

Denise Beech

Patsy Harbour

Academic Grants

Marilou Chanrasmi

CONFERENCE CHARITY

Williamsburg
**COMMUNITY
FOUNDATION**

Celebrating 20 Years

In twenty years, the Foundation has awarded grants and scholarships totaling \$6.8 million to more than 200 nonprofit organizations and over 650 students. We have 6 priority areas: Arts & Culture, Health & Community Wellness, Senior Services, Children & Young Adults, Environment & Conservation, and Scholarship.

MAKE THE WORLD A BETTER PLACE

GIVE CHUCK AND RACHEL A PIE IN THE FACE

If donations for the Williamsburg Community Foundation reach \$500, Chuck will get a pie in the face. If they reach \$750, Rachel will get pied, too!



DONATE AT REGISTRATION

If you want your donation to count for Pie-In-The-Face totals.

Otherwise, use this link or QR-code

<https://tinyurl.com/wcf-sesug>



RIBBONS

Some attendees have ribbons attached to their name badges. These indicate their type of participation in the conference. The ribbon colors and their meanings are:

Ribbon Color	Meaning
Jewel Blue	Conference Chair
Maroon	SESUG Executive Council Past, present and future conference chairs
Black	Section Chairs Organizers of facilities and presentations
Red	Speakers Presentation and Poster authors
Violet	Registration Registrars and Volunteers
Dark Blue	SAS Institute Participants Presenters and Exhibit Hall support staff
Gold	Session Coordinators
Peach	Sponsors
Navy Blue	Academic Grant Student/Faculty
Caramel	Grant Recipients Professional Development
White	Guests

The Conference Chairs and SESUG Executive Council members have been intimately involved with the planning of the conference. If you have any questions or comments about SESUG 2019, look for the people with the maroon ribbons and talk with them!

Remember: your input is essential to the continued success of the conference!

Post Conference: Downloadable zip file of conference papers available at www.sesug.org/SESUG2019

CONFERENCE INFORMATION AND SPECIAL EVENTS

Registration and Information

Location: Virginia Foyer
Time: Sunday 12:00 PM – 7:00 PM
Monday 7:30 AM – 5:00 PM
Tuesday 9:00 AM – 12:00 PM

Registration staff will be there to greet you, provide you with all your conference materials, and answer any questions you might have about the conference.

First Timer Meeting

Location: Virginia A
Time: Sunday 3:00 PM – 3:30 PM

There are three days of papers, over 135 papers, and the SAS Exhibit Hall too! At this session, experienced SESUG attendees will give you some helpful hints to help you get the most out of your SESUG experience! You don't have to be a newcomer to attend! There is information and fun for SESUG veterans as well.

Grant Recipient Meet and Greet

Location: Virginia A
Time: Sunday 3:30 PM – 4:00 PM

Section Chair Meeting

Location: Virginia E/F
Time: Sunday 3:30 PM – 4:00 PM

Speakers/Volunteer Meeting

Location: Virginia E/F
Time: Sunday 4:00 PM – 5:00 PM

Opening Session

Location: Virginia E/F
Time: Sunday 7:00 PM – 9:00 PM

Breakfast

Location: Virginia E/F
Time: Monday 7:30 AM – 9:00 AM
Tuesday 7:00 AM – 9:00 AM

Keynote

Location: Virginia E/F
Time: Monday 8:00 AM – 9:00 PM

Grant Recipient Lunch (Invitation Required)

Location: Liberty Room A/B
Time: Monday 11:30 AM – 1:00 PM

Lunch – on your own

Time: Monday 11:30 AM – 1:00 PM
Tuesday 11:30 AM – 1:00 PM

Evening Reception

Location: Virginia Foyer
Time: Monday 5:00 PM – 6:00 PM

Colonial 5K Fun Run/Walk

Location: Starting at Josiah Chowning's Tavern
Time: Monday 6:30 PM – 7:30 PM

We are meeting in the Virginia Foyer and going out from there. We'll end up by Josiah Chowning's Tavern.

Pub Crawl – on your own

Time: Monday 7:30 PM – ?

Ghost Tours – on your own

Time: Monday 8:00 PM – ?

If you like spooky fun, Williamsburg has Ghost Tours that will excite and titillate you. Here's a list of tours on TripAdvisor's site's. We've pulled out the main companies and listed them below.

Colonial Ghosts: <https://colonialghosts.com/>

Spooks and Legends: <http://www.spooksandlegends.com/>

The Original Ghost Tour: <http://www.theghosttour.com/original-ghost-tour/>

Ghostographer Tours: <https://www.ghostographer.com/>

Williamsburg Ghost Tours: <https://williamsburgghosttour.com/>

Morning Yoga with Charu

Location: Liberty Room
Time: Tuesday 7:00 AM – 7:45 AM

User Group Lunch (Ticket Required)

Location: Virginia F
Time: Tuesday 11:30 AM – 1:00 PM

Closing Session

Location: Virginia E/F
Time: Tuesday 4:30 PM – 5:00 PM

Join us while we say thank you to everyone for making this a great conference. There will be a preview of next year's conference.

ACADEMIC GRANT WINNERS

In early spring, SESUG accepted applications from college students and faculty for grants to attend the conference. Students and faculty using SAS were encouraged to apply and submit a paper for presentation at the conference. With support from SAS, the SESUG grant include a waived registration, limited funding to assist with hotel accommodation expenses, and a special luncheon.

From the list of well qualified applicants, we selected the following people to receive the scholarships. Check out the recipients papers at the conference. Look for the Grant Winner icon next to their names in the Presentations section of the program. Also, visit the Poster Area to view the scholarship winners' profiles.

Student Grantees

Arami Anwell	Virginia Commonwealth University
Yida Bao	Auburn University
Anunay Bhattacharya	Georgia Southern University
Liz Cervantes	Averett University
Zheng Dai	West Virginia University
Geoffrey Dean	Green Hope High School

Faculty Grantees

Michael King	George Mason University
Niloofer Ramezani	George Mason University
Tih Koon Tan	University of the District of Columbia
Xiaomin Wu	Virginia Commonwealth University

PROFESSIONAL DEVELOPMENT GRANTS

SESUG, with support from SAS, is providing professionals with grants to attend this conference. SAS professionals, who have been using SAS in their jobs for less than 3 years, were encouraged to apply. Paper presentations were encouraged, but not required. From the list of well-qualified applicants, the following professionals were selected to receive these grants.

Name	Company
Martin Ahlijah	United Health
John Bassler	University of Alabama at Birmingham (UAB),
Chris Barzola	Westat
Alfonza Brown	Army Public Health Center
Christi Copes	Central Carolina Community College
Wen Jiang	University of Central Florida
Pamela Kelman	Virginia Department of Health
Jerry Landry	Central Piedmont Community College
Sarah Morris	Datalys Center for Sports Injury Research and Prevention

EDUCATIONAL OPPORTUNITIES

Academic Presentations

Location: Allegheny A/B/C, Piedmont A/B/C, Virginia A/C
Time: Monday 9:00 AM – 11:30 AM
Monday 1:00 PM – 5:00 PM
Tuesday 8:30 AM – 11:30 AM
Tuesday 1:00 PM – 4:30 PM

SESUG has always been about education, and the main focus of our conferences is the academic presentations. Presentations are 10, 20, and 50 minutes followed by a few minutes to transition to the next speaker. Presentations are grouped into Academic Sections and each section is assigned to a specific room each day. Feel free to switch rooms as needed; no advance sign-up is required. Most papers can be found in the Conference Proceedings that can be downloaded at www.sesug.org/SESUG2019.

Hands-On Workshops

Location: Allegheny B/C
Time: Monday 9:00 AM – 11:30 AM
Monday 1:30 PM – 5:00 PM
Tuesday 8:30 AM – 11:30 AM
Tuesday 1:00 PM – 4:30 PM

Hands On Workshops (HOW) provide an engaging forum to share a variety of SAS skills in an interactive setting. These sessions are 75 minutes in length and instructors will demonstrate SAS code and procedures in real time with a specific task or goal in mind.

This year attendees will bring their own device and connect to SAS using Amazon Web Services (AWS). Attendees will need a laptop, web browser, and Remote Desktop client. A big thank you to one of our sponsors, Pinnacle Solutions, Inc., for making this possible!

e-Posters

Location: SESUG Exhibit Hall
Time: Monday 9:00 AM – 12:00 PM
Monday 1:00 PM – 5:00 PM
Tuesday 8:30 AM – 12:00 PM

Posters are available on monitors in the SESUG Exhibit Hall. Please stop by during Meet the Authors on Monday and Tuesday to learn more about these topics from the experts that created the posters! See the times to meet them in the Schedule at a Glance (SAAG).

Code Doctors

Location: Exhibit hall/Virginia Foyer (Tuesday afternoon only)

Time: Monday 9:00 AM – 11:00 AM
Monday 1:00 PM – 2:00 PM
Tuesday 9:00 AM – 11:00 AM
Tuesday 1:00 PM – 2:00 PM

SESUG Exhibit Hall

Location: Virginia D

Time: Monday 9:00 AM – 12:00 PM
Monday 1:00 PM – 5:00 PM
Tuesday 8:00 AM – 12:00 PM

Our sponsors would love to connect with you while you attend SESUG! Be sure to meet with our sponsors, including SAS, and learn about the latest software, services, training and books.

- Meet the sponsors that help make SESUG happen and learn what they can do to improve your SAS experience!
- Tap in to SAS expertise and resources as well as connect through communities.
- Talk one-on-one with SAS experts.
- Learn how to enhance your skills with SAS Training & Certification and SAS Books.

SAS SUPER DEMOS AT SESUG

All Super Demos will take place in Virginia D.

Date	Time	Presenter	Topic	Abstract
10/21/2019	9 - 9:30 AM 9:30 - 10 AM (encore)	Jim Kuell 	Important Performance Considerations When Moving SAS® to a Public Cloud	Any hardware infrastructure that is chosen by a SAS customer to run their SAS applications requires the following: a good understanding of all layers and components of the SAS infrastructure an administrator to configure and manage the infrastructure the ability to meet SAS requirements, not to just run the software but to also allow it to perform as optimally as possible This paper will talk about important performance considerations for SAS 9 (both SAS Foundation and SAS Grid Manager) and SAS Viya when hosted in any of the available public clouds (Amazon AWS, Microsoft Azure, and Google Cloud, to name a few). It will also give guidance on how to configure the cloud infrastructure to get the best performance with SAS.
10/21/2019	10 - 10:30 AM	Terry Woodfield 	Using SAS Visual Text Analytics to Explore the Aviation Safety Reporting System Data	The Aviation Safety Reporting System (ASRS) is a NASA funded operation that promotes aviation safety by providing a mechanism to voluntarily report potential aviation hazards that might affect public safety. SAS Visual Text Analytics can process thousands of safety reports to identify key concepts and themes related to specific safety concerns. In addition, the software can extract specific pieces of information from within documents, such as airport identifiers, to help pinpoint specific hazard conditions. This super demo will focus on two areas of concern: (1) operation noncompliance (e.g., failure to obey FAA regulations or Air Traffic Control requests), and (2) incursion hazards (e.g., runway or taxiway incursions that create a collision hazard).
10/21/2019	10:30 - 11 AM 11 - 11:30 AM (encore)	Charu Shankar 	Express Yourself with Python in SAS	With the entry of several open source languages, SAS users feel the need to learn them and understand the differences and commonalities between them. Come learn how to express your data needs by writing Python panda. Learn how Python stacks up with SAS code and do a compare and contrast.

Date	Time	Presenter	Topic	Abstract
10/21/2019	1:30 - 2 PM	Charu Shankar 	Sandwich your SAS Dataset to Excel Pivot Tables	Excel is universally loved. SAS has a way to bring excel into SAS so that you can analyze your data. Users now ask "Great, I can analyze my data in SAS, but my end users don't have SAS on their desktops. How can I give them SAS data in excel form". We'll go even further, instead of taking SAS into a standard Excel workbook, what if you could take SAS to an excel pivot table? Now you can. In this demo watch how quickly you can take a SAS dataset to excel pivot tables. See how in minutes, the Excel table shapes and forms right under your own eyes.
10/21/2019	2:30 - 3 PM	Yiu-Fai Yung 	What's New in SAS/STAT 15.1	Hear about the highlights of this important new release.
10/21/2019	3 - 3:30 PM	Yiu-Fai Yung 	Estimating Causal Effects from Observational Data with PROC PSMATCH	Learn how to use the PSMATCH procedure to do propensity score analysis and causal inference.
10/22/2019	9 - 9:30 AM	Yiu-Fai Yung 	Highlights of Model-Based Clustering	Learn about model-based clustering, its advantages and disadvantages, and how to use PROC MBC to handle data of this nature.
10/22/2019	10 - 10:30 AM 10:30 - 11 AM (encore)	Robert Allison 	Moving from SAS/Graph to ODS Graphics	Do you have years of legacy SAS/Graph code that you're still using, and not sure if it's possible to convert the code from SAS/Graph to ODS Graphics? I will provide tips on what's similar & what's different, and then present several non-trivial examples I have converted from SAS/Graph to ODS Graphics to show that it can be done!

CONFERENCE COURTESIES

Recording, taping or photographing any portion of any presentation is not allowed without the express permission of the presenter.

Turn off the ringers of cell phones, beepers and watch alarms, and keep conversation low and to a minimum during presentations.

Please help us make the 2019 SESUG Conference green! Please use your reusable water container and recycle any unwanted conference materials.

Finally, when you leave the room, please take your glass, cup or plate with you and place it in an appropriate location.

Thank you for your help!

SESUG POLICIES

The annual SouthEast SAS® Users Group (SESUG), Inc., Conference is primarily an educational gathering for the benefit of its attendees. SESUG recognizes that the majority of attendees are present as representatives of their employers for this purpose. Accordingly, SESUG does not condone, endorse, or encourage activities that may conflict with the educational nature of the conference. All attendees and sponsors are expected to abide by the Policies and Procedures set forth in this document.

Paper Content: Users are urged to present papers describing real-world applications using SAS software. SESUG also accepts a limited number of theoretical and general overview papers. Acceptance of all presentations is at the discretion of the Conference Co-Chairs. Oral presentations and written papers describing projects or services of a commercial nature may only be presented at the conference if they describe how the product relates to the use of SAS Institute software; and they do NOT include price lists, support commitments, or other material of a promotional or sales nature.

Right of Withdrawal: The SESUG Executive Council and the Conference Co-Chairs reserve the right to determine if any activity is in violation of these guidelines. They may, at their option, direct the withdrawal of a presentation or demonstration or the dismissal of a SESUG attendee from the conference.

Marketing and Recruiting: Any person or entity wishing to market their products or services or whose presence is primarily to recruit attendees at the annual SESUG conference must register as a sponsor. Registered sponsors are expected to conduct themselves with professionalism. The SESUG Executive Council reserves the right to refuse any or all sponsor registrations. In addition to, or in lieu of, a physical presence at the annual SESUG conference, sponsors may choose to have a virtual presence through means of advertising. Planned activities beyond interaction at a sponsor booth need to be approved in advance by the Conference Co-Chairs.

The Conference Program may include printed sponsor advertisements. Sales literature and promotional items may only be distributed to conference attendees in an approved manner, usually in the form of a conference bag insert or distributed from the sponsor's booth. Program advertisements and items for distribution must be shipped to a designee of the Conference Chairs and are subject to prior approval of form and content. Fees associated with advertising are included in various sponsor package options posted on the website.

Sponsors will be recognized in accordance with the sponsorship guidelines and package options posted on the website. Specific requirements (e.g., content, deadlines, and costs) for sponsor promotional opportunities are included in the sponsor program documents on the website or will be provided in a timely fashion.

SESUG reserves the right to approve any sponsor related activities involving attendees such as hospitality suites, recruitment or other similar activities. In the event any questions of interpretation arise, the decision of the Conference Chairs will apply.

Unless explicitly invited by SESUG, non-registered companies, their agents or individuals may not engage in any direct marketing or sales effort at the conference.

SECTION DESCRIPTIONS

Analytics Leadership

Co-Chairs: Linda Sullivan
Merry Rabb

Companies benefit even more from analytics when leadership takes a strong role in setting that direction. Leaders need information and tools to effectively implement and integrate analytics into their operations. This new track is being offered for leaders and champions of analytics to share experiences and highlight effective approaches to promoting analytics in an organization. Whether you are a senior leader with a mature analytics unit or just starting up an analytics effort, we want to hear from you!

- Presentation topics can include, but are not limited to,
- Case studies for new analytics initiatives
- Monitoring model performance
- Measuring analytical maturity and capabilities
- Leveraging data and analytics to drive transformation
- Managing and growing data science teams
- Moving from Business Intelligence to Analytics
- Making analytics value propositions

Coder's Corner

Co-Chairs: Matthew Bell
Venita DePuy

Every SAS® programmer, from the beginner to the expert, has found new or unusual ways to solve problems with SAS. Coder's Corner is the place to share tips and tricks, useful nuggets of programming, or techniques that make jobs easier. Presentations are 10 minutes in length and can come from a broad range of topics. Come and share your simple tricks to unlock some of SAS's mysteries!

Data Management/Big Data

Co-Chairs: Lesa Caves
Ethan Ritchie

Data manipulation and integration have been mainstays of SAS® software since its inception. Over time SAS has grown to include a full suite of data management capabilities like Data Quality, Data Flux, Data Governance, Master Data Management, and Data Federation. This section intends to highlight the capabilities of traditional Base SAS and SAS Data Management as well as the various ways SAS leverages Big Data. These presentations will cover topics like data integration, data federation, data quality, data access, and data governance while providing helpful insights and lessons learned along the way.

e-Posters

Co-Chairs: Abbas Tavakoli
Brian Varney

The E-Poster Section covers any area including: SAS® fundamentals; statistics; business intelligence; medical research; data mining; survey/panel results; social networking; and industry applications for the pharmaceutical, finance, education, environmental and entertainment industries. E-Posters will be displayed electronically on a wide screen monitor. In addition, a corresponding paper based upon the poster will be published in the conference proceedings. There will be allotted time to meet authors to discuss their E-Posters with conference attendees (Meet the Presenter session). Attendees will have the opportunity to examine E-Posters at their own pace and revisit displays several times during the conference.

Education/Institutional Research

Co-Chairs: Mel Alexander
Meenal Sinha

Papers in the Education/Institutional Research section focus on using the SAS System to find solutions for reporting and analysis in the education community. This section will present techniques, best practices, and solutions for data needs in primary, secondary, and postsecondary education.

- Topics may include, but are not limited to:
- School systems evaluation
- Integrated Postsecondary Education Data System (IPEDS) reporting
- Admissions data management
- Forecasting enrollment, retention, and graduation
- Standardized testing metrics

Hands On Workshops

Co-Chairs: Mike Sadof
Andrea Lewton

Hands On Workshops (HOW) provide an engaging forum to share a variety of SAS skills in an interactive setting. These sessions are typically 1.5 to 2 hours in length. Presenters will demonstrate SAS code and procedures in real time with a specific task or goal in mind. Any SAS topics that are better explained in an interactive setting are ideal for Hands On Workshops. These sessions are not reserved for complicated or elaborate SAS techniques only. A power point presentation, as well as sample code for the demonstration, is needed for the workshop. The datasets and code should be made available for download prior to the session, enabling attendees to run the code on their personal computers.

Workshops this year will be set up such that attendees bring their own device.

We will be using Amazon Web Services (AWS) to make presentation materials and SAS software available to attendees. Attendees will need a laptop, web browser, and Remote Desktop client. Please make sure you have an appropriate Remote Desktop client for your operating system.

- * Windows users: the default Remote Desktop client is fine
- * Mac OS X users: download the Microsoft RDP v10 app [here](#).

Attendees do not need to have SAS licenses on their own computers. However, if attendees decide they want to use their own computers and SAS licenses during the presentations, class materials will be available for download.

No laptop? No problem! Attendees are welcome to attend HOW sessions and watch as the instructor works through the exercises on the screen.

Government

Co-Chairs: Mel Alexander
Meenal Sinha

Papers in the Government section will focus on using the SAS System to help local, municipal, state, and federal government agencies apply analytics in delivering quality services in the face of budgetary and human resource constraints.

Topics may include, but are not limited to:

- Taxpayer funding, budgeting, and finance management
- Defense contracting
- Tax revenue and waste reduction oversight
- Public protection and safety
- Internal and external regulatory and policy compliance

Healthcare/Pharmaceuticals

Co-Chairs: Mel Alexander
Meenal Sinha

Papers in the Healthcare/Pharmaceuticals section will focus on using the SAS System to provide quality healthcare outcomes - saving patient lives; hospital readmission rates; drug/device discovery, injury and disease prevention; patient care and satisfaction; healthcare costs and delivery.

- Topics may include, but are not limited to:
- Patient safety and personal healthcare
- Medical and pharmaceutical intervention effectiveness
- Clinical trial evaluation
- Epidemiological, environmental, socioeconomic, and genomic studies
- Insurance premiums

Know Your SAS: Foundations

Co-Chairs: Charlotte Baker
Kelly Smith

This section is primarily intended for programmers new to SAS, and includes topics essential for building the knowledge and capacity of beginner SAS users. Appropriate topics include SAS programming fundamentals, SAS program design, basic SAS procedures, and SAS debugging techniques. The presentations will provide beginning programmers with a greater understanding of SAS coding procedures, rules, and troubleshooting. Programming topics across all fields and industries are encouraged and welcome.

Know Your SAS: Advanced Techniques

Co-Chairs: Harry Droogendyk
Richann Watson

Know your SAS: Advanced Techniques provides instructional content for intermediate and advanced SAS users. These presentations will help advanced programmers implement enhanced techniques to build on the power and flexibility afforded by SAS software. Programming topics across all fields and industries are encouraged and welcome.

- Subject matter will include advanced SAS programming topics including:
- ODS
- SAS Macro Facility
- Sophisticated and efficient PROC and DATA step programming solutions
- SAS Enterprise Guide functionality
- Integration with third-party software and databases
- Innovative application development and data integration
- Reporting and analytics solutions

Open Analytics

Co-Chairs: Jason Brinkley
Brian Varney

Open Source software has become popular for data visualization, analytics, and data manipulation and management. This year, SESUG will pilot a new section featuring open source software that can be useful to the SAS user community. As SAS states, "As we think about the entire analytics life cycle, it's important to consider data preparation, deployment, performance, scalability and governance, in addition to algorithms. Within that cycle, there's a role for open source and commercial analytics.*".

- The Open Analytics section will offer a variety of papers on topics leveraging open source applications in conjunction with or separately from SAS applications. Presentation topics that demonstrate clear benefit to the SAS user community may include, but are not limited to:
- Integrating open source software with SAS
- Scaling open source models to the enterprise
- Accomplishing specific tasks using open source software
- Developing and solving smaller scale projects with open source software
- Balancing user needs in an open analytics ecosystem
- Providing opportunities for open source users to learn SAS
- Managing diverse analytics assets and development tools

This section welcomes topics which include but are not limited to the following technologies:

- SAS Viya
- SASPy
- R
- Python
- OpenStreetMap
- Lua
- Groovy
- Google Analytics
- Google Maps
- RESTful web services
- Apache/Hadoop
- HDFS
- Spark
- Jupyter as an IDE

Planning/Administration

Co-Chairs: Josh Horstman
Carlos Piemonti

If all or part of your SAS® time includes supporting users, whether through systems architecture and administration or through consulting, training, and hiring, this section is the place for you to share your experiences with other members of the SAS community. This section will include guidelines, best practices, techniques, and resources for working efficiently and effectively in the SAS support community.

Possible topics include:

- SAS Systems architecture and administration
- Installation, deployment, and migration
- Virtualization
- Performance monitoring and tuning
- Other SAS systems support
- Recruiting, hiring and maintaining qualified staff
- Training and skill development
- SAS help desk support
- Project planning and management

Reporting and Visualization

Co-Chairs: Louise Hadden
Barbara Okerson

The Reporting and Visualization section invites presentations that demonstrate unique and innovative ways to visualize data and output. SAS provides many tools for visualizing and reporting data and results including Visual Analytics, SAS/Graph, SAS SG procedures, DSGI, and JMP®.

Presentation topics include but are not limited to:

- SAS graphics procedures, styles, templates, Output Delivery System (ODS) and Graphics Template Language (GTL)
- Customized reports, dashboards, scorecards, graphs, and maps
- SAS Visual Analytics
- JMP applications
- SAS and/or JMP integration with Microsoft Windows, Mac OS, R, Python, MATLAB, Tableau and TIBCO Spotfire

Statistics/Data Analysis

Co-Chairs: Kristen Harrington
Nat Wooding

Presentations in the Statistics and Data Analysis section address ways of transforming raw data into useful information that uncover important relationships and patterns in data to help gain insights for effective, data-driven decision-making. Papers do not need to present new statistical methods, although such topics are always welcome. Presentations are sought that involve the application of methods that many users of SAS statistics may not commonly see, such as methods for categorical, longitudinal, or censored data. Methods to facilitate analysis of very large data arrays, such as those that result from genetic studies or national surveys, are also sought for this section. Topics in this section should be of interest to a broad spectrum of SAS practitioners including analysts, developers, statisticians, data scientists, and DATA step programmers.

ABSTRACTS BY PAPER NUMBER

(by section then paper number)

Analytics Leadership

AL-138 ***Exploring the Skills Needed by the Data Science / Analytics Professional***
Kirk Paul Lafler

As 2.5 quintillion bytes (1 with 18 zeros) of new data are created each and every day, the age of big data has taken on new meaning with a renewed sense of urgency to prepare students, young professionals, and other workers across job functions for today's and tomorrow's analytics-roles along with the necessary analytical skills to tackle growing data demands. With the number of organizations embracing Data Science / Analytics skills and tools, organizations like LinkedIn, a leading professional networking and employment-oriented website and app, found that Data Scientists saw a 56% increase in the US job market in 2018. To keep up with the huge demand for analytics talent in 2019 and beyond, many colleges, Universities, and training organizations offer comprehensive Data Science / Analytics degrees and certificate programs to fulfill the increasing demand for analytical skills. This presentation explores the skills needed by the Data Science / Analytics professional including critical thinking; statistical programming languages such as SAS®, R or Python; Structured Query Language (SQL); Microsoft Excel; and data visualization.

(KEYWORDS: CRITICAL THINKING, SAS, R, PYTHON, SQL, EXCEL, DATA VISUALIZATION)

Kirk Paul Lafler is entrepreneur, consultant and founder of Software Intelligence Corporation, and has worked with SAS software since 1979. As a SAS consultant, application developer, programmer, data analyst, mentor, infrastructure specialist, educator and author at Software Intelligence Corporation, and an advisor and SAS programming adjunct professor at the University of California San Diego Extension, Kirk has taught SAS courses, seminars, webinars and workshops to thousands of users around the world. Kirk has also authored or co-authored several books including PROC SQL: Beyond the Basics Using SAS, Third Edition (SAS Press. 2019) and Google® Search Complete (Odyssey Press. 2014); hundreds of papers and articles on a variety of SAS topics; served as an Invited speaker, educator, keynote and section leader at SAS user group conferences and meetings worldwide; and is the recipient of 25 "Best" contributed paper, hands-on workshop (HOW), and poster awards.

AL-215

Self-service Analytics Platforms for Real-Time Insights into Clinical Data

Binoy Varghese

The clinical research industry is undergoing a dramatic shift in the way it looks at data. For decades, end users have been limited to reviewing static outputs that inhibited their ability to explore data without depending on the programming team. With the onset of industry-wide data standards and innovations in analytics technologies, more and more organizations are building self service capabilities that enable users to aggregate, drilldown and visualize data from the early stages of study conduct through study closeout. Furthermore, as the data collection landscape evolves these platforms are expected to assimilate huge volumes of data generated by wearable devices. Harnessing data via such high-performance systems empowers users with unparalleled efficiency that promotes exploratory enquiry and data driven insights while at the same time freeing up the programming team. This presentation explores the "Shiny Server" framework as a viable solution for an in-house analytics platform. Coupled with a live demo, we will present our findings with a focus on interactivity, responsiveness and ease of use along with 'behind the scenes' aspects such as scalability, upgrades and maintenance.

(KEYWORDS: ANALYTICS, SHINY SERVER)

Binoy Varghese presently serves as the Director of Biometrics Data Operations at AstraZeneca, where he is engaged in leading various enterprise wide initiatives on automation, analytics, infrastructure, partnerships, standards and submissions. He specializes in Statistical Programming and has worked with industry leading organizations over the past 18 years on phase 1 through post marketing studies across multiple therapeutic areas. He is a published author-presenter and has a M.S. in Computer Science.

AL-223

Need to develop your employees' SAS skills? A step by step framework.

Kelly Smith

SAS professionals come from a variety of educational and work backgrounds, requiring supervisors to craft personalized development plans for their employees. In addition, SAS runs on multiple platforms and offers multiple options to pull in data, to create and modify data, to perform data analysis, and to output and present the data. To top it all off, SAS offers numerous options for learning. The complexity of providing guidance to new and seasoned SAS professionals can seem overwhelming, even for experienced supervisors. In this presentation, a framework for crafting individualized development plans for SAS professionals is offered. The framework is grounded on research based principles and incorporates the concepts of andragogy, heutagogy, and formal/informal learning. The diverse learning options available for SAS are explored and suggestions for matching learner goals and learning options are provided. To conclude, a dozen development tips and a list of useful websites are presented.

(KEYWORDS: PROFESSIONAL DEVELOPMENT, TRAINING, SKILLS)

Before joining the SAS Universe, Kelly was a full-time educator. To improve her chances of getting hired as an institutional researcher, Kelly took advantage of the free learning resources offered by SAS. Her goal with this presentation is to provide managers with a workable plan to help their employees get the most out of their professional development options.

AL-236

Lessons from the Changing Analytic Landscape over the Past 40 Years.

Alan Dunham

Experiences inside over ten analytic organizations over the past forty-plus years show significant changes in preferred expertise, but also there are some constants related to human and organizational leadership behavior. The author discusses his perceptions of how the typical data analytics and statistical analysis jobs have shifted emphasis. Also there are several tips about how to quickly recognize the character and dynamics of different organizational leadership situations.

(KEYWORDS: ORGANIZATIONAL LEADERSHIP, CHANGING SKILL REQUIREMENTS, WORK EXPERIENCE)

AL-254

An Approach to Obtaining Investment funding and managing an Artificial Intelligence-focused initiative

Karen Davis

Organizations can develop significant inertia related to continuing to use known tools, techniques and processes related to how work is done. Adopting newer technologies, and moving into new areas, and applying new tools can run into significant obstacles, not the least of which is having enough resources to make a significant change happen. This presentation will cover RTI's approach to developing, adopting and implementing new approaches to data and analytics, including the application for funding, addressing change management, and ensuring executive support.

(KEYWORDS: ANALYTICS LEADERSHIP, CHANGE MANAGEMENT, ARTIFICIAL INTELLIGENCE)

Karen M. Davis, Vice President of RTI's Research Computing Division (RCD), was hired by RTI in 1999 to set a vision and strategy for RTI-wide business systems and establish an Information Systems (IS) organization to execute the plan. Since 2005, Ms. Davis has led RCD within the Social, Statistical and Environmental Sciences (SSES) research group; the division employs over 150 technology professionals working on projects for RTI's clients. In 2018, Ms. Davis became responsible for RTI's investment in artificial intelligence and machine learning. Ms. Davis has been a project director on multimillion-dollar projects as well as having taken a lead in systems security, and working with clients on requirements for systems compliance with FISMA-moderate. Ms. Davis' prior roles in computing and engineering have been with Data General Corporation, Kendall Corporation, Ford Motor Company, and Bethlehem Steel. Ms. Davis' education includes bachelor's degrees in Mathematics and English, and a Master's degree in Management, with a concentration in Information Systems.

Ideation and Innovation – Leveraging science and human behavior learnings to evolve your processes and teams**Steve Walker**

Organizations are being challenged with thinking differently and creating new approaches and products. In today's Digital Transformation era, teams are required to work and think in different ways to accelerate their overall digital transformation and agile maturity. Scientific research in human behavior and learning has created new concepts that can be applied to our business agility. In this session, we will discuss scientific principles of human behavior and learning and how these apply to today's workforce and our agile processes. In addition, we will see these behavior in action in the context of an innovation framework process that can applied to projects and products.

(KEYWORDS: LEARNING, HUMAN BEHAVIOR, TEAM DEVELOPMENT, AGILE, INNOVATION)

Steve Walker is a business and technology leader who focuses on bringing change and innovation to business. Steve has built his career on two core principles: You must try new things to improve and professionalism and ethics comes first. As the Experis Digital platform director and leader of the Experis Solution's Innovation Center, Steve leads Digital Transformation initiatives. He is responsible for identifying and creating new products and solutions for today's digital business. Having played many roles ranging from practice management to software engineer throughout his 25 year career in the solutions and system integration space, he has vast experience applying technology to new business problems.

Steve currently focuses on Digital Transformation enablement. Specifically, this is the application of new and emerging business models and technology to optimize and create new digital experiences and value. Across his career, he has helped organizations evolve by delivering complex solutions such as delivering new digital experiences, developing new business models, optimizing content production process, implementing complex software solutions, and improving training programs.

Steve delivers industry and client presentations across topics that are relevant to today's changing Digital Transformation business climate. Throughout his career he has worked with many successful organizations across industries and verticals including Microsoft, Apple, Dell, GE Lighting, General Dynamics IT, ChildFund, AllState Insurance, International Monetary Fund, Desert Schools Federal Credit Union, CFA Institute. Steve graduated from Virginia Polytechnic Institute and State University in 1992 with a double major in Finance and Economics and a minor in Computer Science.

Specialties: Emerging technologies and trends, Digital Experience Management, Web Strategy, Robotic Process Automation (RPA), Enterprise Content Management, Information Architecture, Usability, Marketing Technologies, Business Transformation, Enterprise Application Development, Analytics

Improving your organization's Digital Transformation agility**Steve Walker**

Organizations face a series of challenges to evolve in today's digital first environment. Digital no longer just means simply delivering a good customer experience but requires a series of organization investments and activities to adequately transform. Digital Transformation is a complex topic that covers various People, Process, and Technology challenges. How quickly your organization evolves can be directly attributed to series of tangible activities that must exist. We will discuss a set of criteria that an organization should evaluate to adjust their Digital Transformation Agility and show real-world examples of a digital leader who is making these adjustments.

(KEYWORDS: DIGITAL, DIGITAL TRANSFORMATION, AGILE, CUSTOMER EXPERIENCE)

Steve Walker is a business and technology leader who focuses on bringing change and innovation to business. Steve has built his career on two core principles: You must try new things to improve and professionalism and ethics comes first. As the Experis Digital platform director and leader of the Experis Solution's Innovation Center, Steve leads Digital Transformation initiatives. He is responsible for identifying and creating new products and solutions for today's digital business. Having played many roles ranging from practice management to software engineer throughout his 25 year career in the solutions and system integration space, he has vast experience applying technology to new business problems.

Steve currently focuses on Digital Transformation enablement. Specifically, this is the application of new and emerging business models and technology to optimize and create new digital experiences and value. Across his career, he has helped organizations evolve by delivering complex solutions such as delivering new digital experiences, developing new business models, optimizing content production process, implementing complex software solutions, and improving training programs.

Steve delivers industry and client presentations across topics that are relevant to today's changing Digital Transformation business climate. Throughout his career he has worked with many successful organizations across industries and verticals including Microsoft, Apple, Dell, GE Lighting, General Dynamics IT, ChildFund, AllState Insurance, International Monetary Fund, Desert Schools Federal Credit Union, CFA Institute. Steve graduated from Virginia Polytechnic Institute and State University in 1992 with a double major in Finance and Economics and a minor in Computer Science.

Specialties: Emerging technologies and trends, Digital Experience Management, Web Strategy, Robotic Process Automation (RPA), Enterprise Content Management, Information Architecture, Usability, Marketing Technologies, Business Transformation, Enterprise Application Development, Analytics

Current Issues in Analytics Leadership: A Panel Discussion**Kelly Smith, Karen Davis and Jared Dean (moderated by Linda Sullivan and Merry Rabb)**

This panel discussion and audience Q&A will provide engaging discussion on current analytics leadership issues including: managing analytic teams and ensuring they have the necessary skill sets, how to create and sustain support for analytics initiatives at your organization, and how to stay abreast of evolving and emerging analytics technologies.

(KEYWORDS: PANEL DISCUSSION, ANALYTICS LEADERSHIP)

A leaders' Look at Creating an Effective Analytical Ecosystem**Jared Dean**

As a leader of analytical talent, you face pressure to effectively maximize both human capital and technology investments. There are a number of competing interests and constraints as you make these decisions. This talk aims to provide a perspective on the trade-offs of different strategies and their consequences (both positive and negative). We will discuss key considerations with respect to tools, infrastructure, and processes. Included in this discussion will be SAS technologies along with open source projects including R, Python, Docker, Kubernetes, and Git.

(KEYWORDS: ANALYTICAL TALENT, TECHNOLOGY INVESTMENTS, TRADE-OFFS, SAS TECHNOLOGIES, OPEN SOURCE TECHNOLOGIES)

Jared is a Principal Data Scientist and Business Knowledge Series instructor at SAS. He has developed leading edge analytics for banking, entertainment, and mobile data. He has developed and maintains projects in Python and R including the SAS kernel for Jupyter and SASPy. Outside SAS, he is an adjunct professor in the MBA program for NCSU, Duke, and Elon as well as frequent speaker and presenter. Jared holds several patents in data mining. Previously, Jared was a Mathematical Statistician for the US Census Bureau and CTO of Reveal Mobile, a mobile marketing analytics startup. He was also a Senior Director of Research and Development for SAS Enterprise Miner. He holds an MS degree in computational statistics from George Mason University and served as an advisory board member to their statistics department. Jared is the author of Big Data, Data Mining, and Machine Learning: Value Creation for Business Leaders and Practitioners (John Wiley & Sons, Inc., May 2014). His book was listed as one of five must reads in big data for the summer. When not working, he spends time with his wife and four children. Their favorite family activities include travel, sports, enjoying good food, cheering for the Denver Broncos.

Coder's Corner

CC-108 ***I'm Warning You - Common SAS Programming Mistakes***

Christine Whitaker

This paper covers common programming mistakes, including those that don't generate error messages, but produce wrong results.

(KEYWORDS: COMMON SAS PROGRAMMING MISTAKES)

I am a 1982 graduate of WSSU (for all you Aggies in the house...deuces) and a 20 year employee. My title is Business and Technology Applications Analyst in IAR. I am a member of the NCAIR Executive board (Member-at-Large – Public Senior Institutions). I say 10+ years with SAS but it's been longer. I started out in 1990 (I believe) using SAS in batch mode where you had to write the code and the DCL (executable JCL - job control language to run the code. In August 2005, my director at that time looked over my shoulder one day and immediately introduced me to the Interactive Point and Click SAS. I have been using it since that day.

CC-117 ***Macroception: Maximizing the Capacity of Your Macro Variables***

Connor Cosenza

SAS® imposes a maximum character limit of 65,534 on macro variables. Running up against this limit can be frustrating and time-consuming. The author has run into this problem when using macro variables to search long lists of values or to dynamically generate code. This paper provides a program which solves this problem by partitioning the values to be encoded in a given macro variable, assigning those values to sub-macro variables, and encoding a parent macro variable with the names of the sub-macro variables. When the parent macro variable is called, all raw data values are decoded. This significantly increases the maximum number of characters which can be encoded into a single macro – specifically, this theoretical upper-limit increases from 65,534 to 623,883,680 characters.

(KEYWORDS: MACRO, VARIABLE, MAXIMUM, LENGTH)

Connor Cosenza is a Senior Associate at Ankura Consulting Group, LLC. Connor uses SAS every day to analyze litigation data and forecast aggregate liability for mass tort settlements. He has been using SAS for over four year.

CC-119 ***Validating Medicare Beneficiary Identifier Submissions***

Richard Pickett

Centers for Medicare and Medicaid Services (CMS) is removing Social Security Numbers (SSNs) from Medicare cards to fight medical identity theft for people with Medicare. A new Medicare Beneficiary Identifier (MBI) is replacing the SSN-based Health Insurance Claim Number (HICN). This presented a challenge to healthcare plans submitting data for the Medicare Health Outcomes Survey (HOS). Submissions included HICNs, Railroad Board Numbers (RRB) and MBIs mixed under one field which in turn presented a challenge to RTI in drawing additional information from the Enrollment Database using the identifier. This paper presents SAS macro code to identify invalid MBIs and display them for review and communication with the healthcare plan for correction.

(KEYWORDS: MEDICARE BENEFICIARY IDENTIFIER, DATA INTEGRITY, SAS MACRO)

Richard Pickett has 19 years of programming experience in constructing, maintaining, and analyzing large and complex survey and medical insurance claims data. He has excelled in performing analyses, intensive data cleaning, complex file construction, and verification, and testing of census bureau sampling programs and documentation; he has provided detailed and carefully written documentation for programs. Mr. Pickett also has extensive analytic experience from a 20 year career in wholesale grocery procurement and retail grocery operations.

CC-146 ***How to Write a Macro that Executes Other Macros***

Kelly Fenn

The ability to create SAS macro programs can make your life easier when you need to write a chunk of code that will be executed multiple times within a SAS program. Once you have mastered writing macro programs, however, you may find yourself in a situation where you need to call the macro several times, with a number of different parameters. This paper will introduce you to a technique for writing a macro program that will call another macro program repeatedly over a list of parameters, allowing you to avoid manually calling the macro for each unique parameter.

(KEYWORDS: MACROS, MACRO PROGRAMMING)

Kelly Fenn is a Marketing and Customer Insights Analyst at Fulton Financial Corporation in Lancaster, Pennsylvania. She uses SAS to extract insights about Fulton's customer base, and has enjoyed constantly improving her SAS skills -- often via frantic Googling. She lives in Lancaster with her husband, three kids, and four cats.

How am I different?**Kannan Deivasigamani**

Programmers often seek to use the procedure from SAS with the hope of finding differences between 2 datasets. One step further, the interest focuses on the variables that are different between the two datasets in question. Fortunately, "PROC COMPARE" provides information about the number of variables that are common between the 2 datasets and also prints the number of variables in the BASE dataset not in the COMPARE dataset and vice versa. However, there is some level of scrolling and manual tabulation to do if you consolidate the output of options such as LISTVAR, LISTBASEVAR, or LISTCOMPVAR. In many occasions, it turns out to be a manual process of eyeballing the fields or transferring the metadata to an Excel spreadsheet and using a VLOOKUP or some other function and then compiling the variables. As an alternative, what if there is a tool available where you can just plug-in the dataset names and it provides a neat print of the common columns, differences in the 2 datasets by listing out columns side-by-side in BASE but not in COMPARE dataset and vice versa in a presentable table format? Wouldn't that be nice? It would save several minutes of one's time. If this is done on multiple occasions, by multiple programmers, the benefit multiplies. That is exactly what this tool is intended to provide. The tool will accept the names of the two datasets to compare and will provide an output that will print common variables, variables in dataset-1 but not in dataset-2, and variables in dataset-2 but not in dataset-1 in 3 columns side by side. This can be presented to the individual who can address or justify the differences. The tool is a SAS macro that will accept two inputs and will handle different scenarios presented in the examples. This will take advantage of the SAS meta data and with minimal effort, will provide the desired output to the user.

(KEYWORDS: TABULATE DIFFERENCES, COMPARE, DIFFERENCES, VARIABLE NAMES, DICTIONARY.COLUMNS)

Kannan is currently contributing towards making Wellcare a better organization one day at a time. He has a MBA from Seton Hall & Doctorate degree from University of Phoenix. He is also a Remote Pilot, Photographer, and Ham Licensee. Love to watch movies, play outdoors and the sunny beaches of Florida and NASA.

You ought to be in pictures! Dressing up output with the picture format**Jennifer Lindquist**

SAS-provided and user-defined formats improve the appearance, interpretability, and readability of data. Using formats can eliminate the creation of additional 'parallel' text variables and inadvertent discrepancies between those 'parallel' character strings and the original numeric values. An under-utilized format option is the picture format. Picture formats allow output to be stylized by instructing SAS to present numeric values in a specified pattern. A picture format allows the user to add units and symbols (e.g. presenting a gain of 9.5 percent as +9.5%). Rescaling by a multiplication factor is easy to do (e.g. displaying 15,000 as 15K). Also rounding and truncation are a snap with picture formats. All can be accomplished without creating additional variables or calculations! Familiar with Excel formats? Almost all can be created with the SAS picture format. Come explore how to dress up your output and get into pictures!

(KEYWORDS: FORMAT, PICTURE FORMAT, OUTPUT STYLE)

Jennifer Lindquist is a statistician/data manager with the Durham Center Innovation to Accelerate Discovery and Practice Transformation. She has been in Health Services Research at the VA for 22 years and a SAS user for 28 years. Prior to her time at the VA, she was first a high school and then college math teacher.

CC-180 ***Computing Classical Item Statistics in SAS Using the Long vs. Wide Form of the Data***
Imelda Go

The wide form of data is intuitive to people because it resembles data as they are presented in a spreadsheet with each column representing a variable. There are advantages to storing and processing data in their long form versus in their wide form. This paper illustrates how to simplify and shorten the coded needed to compute classical item statistics when the data are in their long form.

(KEYWORDS: CLASSICAL ITEM STATISTICS)

Imelda "Mel" Go, Ph.D. is a Data Quality SAS Programmer with Questar Assessment's Psychometric and Research group. Prior to that she was with the South Carolina Department of Education's Office of Assessment for 15 years and with two SC public school districts for 7 years. She uses SAS to analyze student test data and to perform high-stakes calculations.

CC-190 ***Conditionally Executing Data Steps and Statements Based on the Presence of Variables in a SAS® Dataset***
Charles Coleman

Sometimes, a data step should only be executed if one or more variables are present. The same is true of statements within data steps. This paper provides several macros to detect the presence of variables and counts of the present variables. Examples show how these macros can be used to accomplish conditional execution for many scenarios. These macros have the advantage of only using macro statements, thus avoiding the costs of opening datasets.

(KEYWORDS: CONDITIONAL EXECUTION, MACROS, DATASETS)

Chuck Coleman is a Mathematical Statistician at the U.S. Census Bureau. He has been using SAS® since he began working there since 1998. Currently, his main responsibility is to develop specifications for estimates from the Survey of Construction in SAS.

CC-192 ***Three Ways to Transform Your Code into PROC SQL***
Charity Wilson

Does learning PROC SQL seem overwhelming? Or maybe you feel that it is not necessary? But did you know that PROC SQL is a powerful tool used to manipulate data in a dataset all in one procedure? This paper will introduce PROC SQL and demonstrate how to easily transform common DATA and PROC Steps you already use into PROC SQL code. Why would you want to use PROC SQL instead of DATA and PROC Steps you may ask? You can accomplish multiple tasks in one SELECT statement, easily join and append datasets, automatically print results without the use of a PROC print and utilize a time saving secret weapon – no presorting your dataset. Topics covered are great for novice SAS users to legacy DATA Step programmers.

(KEYWORDS: PROC SQL, CASE, SORT, APPEND)

Charity Wilson is a corporate analyst at Cobb EMC where she has been since 2015. Charity uses SAS® in her daily responsibilities of electric rate design, load forecasting, budgeting, and ad hoc rate analysis. She has used SAS for 4 years.

When in Doubt, Shell It Out**Timothy Egan**

Sometimes a table shell appears simple at first glance, but can quickly grow complicated when it comes time to program. Inconsistent blank lines that are needed within and between sections of the table, unpopulated categories that will not be produced with a PROC FREQ but still need to appear in the final table, future visits that still need to be presented, and inconsistent indentations are all subtle details that can challenge a programmer when they sit down to knock out a table. While the data step and PROC REPORT are perfectly valid approaches to create these tables, there are simpler methods in SAS that will allow a programmer to easily build these obscure or repetitive table structures. A key advantage to using these methods is that if the table structure needs to be updated at a later time, the program is already in a ready-to-update state for either the original programmer or someone else who has inherited the program. This presentation will show how to utilize the data step, DATALINES, or PROC SQL to create a complicated or repetitive table structure with a simple and easy to read program. It will also examine strengths and weakness of the different methods which will enable programmers to select the more efficient choice on a case-by-case basis.

(KEYWORDS: TABLE SHELL, TABLE CREATION, DATA STEP, DATALINES, PROC SQL)

Timothy Egan graduated from UNC-Chapel Hill in 2016 with a double major in Biostatistics and Mathematics. He gained his first exposure to the language through 2 SAS classes offered by the school of public health, as well as through undergraduate research. Upon graduating, Timothy immediately began working as a statistical programmer in the clinical trial where he has been programming with SAS for almost 3 and a half years. Timothy entered the field with a strong foundation in SAS, but has since gained much more experience and knowledge. He greatly enjoys discovering new programming techniques written by his coworkers.

User-Defined Functions that Concatenate Statistics for Standard Reporting**Martha Wetzel**

Statisticians often need to report results such as means/confidence intervals, medians/quartiles, and counts/percentages in a format that differs substantially from the data sets produced by SAS. Academic journals typically expect that tables display related results (e.g., the confidence interval associated with a mean) in a single cell with parentheses around select statistics. However, SAS procedures output data sets containing each statistic in a separate column. As a result, the statistician must either copy and paste results, which introduces an opportunity for error, or spend time writing lengthy concatenation statements. Without proper formatting, concatenation results in dropped trailing zeros, requiring additional manual changes. In order to increase efficiency, custom functions can be used to combine statistics into a character string following a standard pattern. To that end, four custom functions built with SAS Macro language are provided here:

1. MEANCI: Returns a concatenated variable of mean and confidence interval in the format "mean (lower confidence level, upper confidence level)"
2. MEANSTD: Returns a concatenated variable of mean and standard deviation in the format "mean (standard deviation)"
3. MEDQ: Returns a concatenated variable of medians and quartiles in the format "median (Q1, Q3)"
4. NPCT: Returns a concatenated variable of counts and percentages in the format "count (percent %)"

(KEYWORDS: REPORTING, USER-DEFINED FUNCTIONS, TRAILING ZEROES, JOURNAL-FORMATTED RESULTS)

Martha Wetzel is a biostatistician in the Department of Pediatrics at Emory University. Her research areas of interest include opioid policy and pain management. She holds an MSPH in health services research and health policy from Emory University and previously worked in the health care quality improvement field.

You're Doing It Wrong! Volume 002**Shane Rosanbalm**

You might think that you're a good programmer. But you're not. It's not just that you're doing it differently than I would do it. It's that you're actually doing it in a way that is unquestionably, incontrovertibly wrong! But, take heart. I am here to set you on the righteous path. Listen to me, and you will be adored by your coworkers, accepted by SUG section chairs, and solicited by recruiters. The focus of volume 002 will be the virtuosity of vertical code.

(KEYWORDS: VERTICAL)

Shane Rosanbalm is a Senior Biostatistician at Rho. He became a SAS user while studying biostatistics at the University of North Carolina at Chapel Hill in a time before there were such things as hybrid cars, GPS, or texting.

How to Keep Multiple Formats in One Variable after Transpose**Mindy Wang**

In many industries and research fields, proc transpose are used very often. When many variables with their individual formats are transposed into one variable, we lose the formats. We can do a series of if then statements to put the formats back in. However, when the variables involved are too many, the above method can be very tedious. This paper illustrates how to extract formats from dictionary.columns or sashelp.vcolumn, and then use PUTN function to assign the formats at run time and make the task much easier. In addition, it is much easier to apply the same method to other projects without a lot of hard coding in the SAS program. Efficiency is largely increased with this method.

(KEYWORDS: TRANSPOSE, FORMATS, PUTN, DICTIONARY.COLUMNS, SASHELP.VCOLUMN)

Mindy Wang is an Independent Consultant for SAS programming based in North Potomac, Maryland with over 20 years' experience. Mindy strives to make repetitive tasks easier using macro and other advanced date-driven techniques. She loves training junior staff and hosting webinars on SAS techniques. A SAS certified advanced programmer and Clinical Trials Programmer, Mindy is a frequent presenter at various regional and local SAS users group meetings.

Data Management/Big Data

DM-130

Reducing the space requirements of SAS® data sets without sacrificing any variables or observations

Stephen Sloan

The efficient use of space can be very important when working with large SAS data sets, many of which have millions of observations and hundreds of variables. We are often constrained to fit the data sets into a fixed amount of available space. Many SAS data sets are created by importing Excel or Oracle data sets or delimited text files into SAS and the default length of the variables in the SAS data sets can be much larger than necessary. When the data sets don't fit into the available space, we sometimes need to make choices about which variables and observations to keep, which files to zip, and which data sets to delete and recreate later. There are things that we can do to make the SAS data sets more compact and thus use our space more efficiently. These things can be done in a way that allows us to keep all the desired data sets without sacrificing any variables or observations. SAS has compression algorithms that can be used to shrink the space of the entire data set. In addition, there are tests that we can run that allow us to shrink the length of different variables and evaluate whether they are more efficiently stored as numeric or as character variables. These techniques often save a significant amount of space; sometimes as much as 90% of the original space is recouped. We can use macros so that data sets with large numbers of variables can have their space reduced by applying the above tests to all the variables in an automated fashion.

(KEYWORDS: SAS, EFFICIENCY, STORAGE, PROGRAMMING TECHNIQUES, ACCURACY)

Stephen Sloan has worked at Accenture in the Services, Consulting, and Digital groups and is currently a Data Science Senior Principal in the SAS Analytics area. He has worked in a variety of functional areas in Project Management, Data Management, and Statistical Analysis. He has had the good fortune to have worked with many talented people at SAS Institute. Stephen has presented at 20 SAS Global Forums and SAS Users Group events and has been published in professional journals.

Stephen has a B.A. cum laude with Honor in Mathematics from Brandeis University, M.S. degrees in Mathematics and Computer Science from Northern Illinois University, and an MBA (1st in class) from Stern Business School at New York University. He also has a graduate certificate in Financial Analytics from Stevens Institute.

DM-230 ***Exploring Efficiency in Data Manipulation with SAS: How to Get the Most Out of My Software and Hardware***

Jonathan Duggins and Jim Blum

An exploration of efficiency in combining data sets (merging/joining and interleaving) is considered under a variety of conditions. Methods are employed in each of the DATA step, PROC SQL, and PROC DS2 for a variety of data sets. Data sets considered include those with large numbers of observations, large numbers of variables, or both. Non-indexed and indexed versions of data sets are considered, with time required to construct the index included in the efficiency calculation. Multiple operating systems and hardware platforms are considered along with options available to set the resources available in the SAS session(s).

(KEYWORDS: EFFICIENCY, DATA STEP, SQL, DS2, MERGE, JOIN, INDEX, BIG DATA)

Jim Blum is a Professor of Statistics at the University of North Carolina Wilmington where he has developed original courses in SAS programming for the university over the past 15 years. These courses cover topics in base SAS, SAS/GRAPH—recently transitioned to template-based graphics, SAS/SQL, SAS/STAT, and macro language. Jim will be a primary instructor in UNC Wilmington’s Data Science program debuting in fall 2017.

DM-245 ***Should I Wear Pants? And Where Should I Travel in the Portuguese Expanse? Automating Business Rules and Decision Rules Through Reusable Decision Table Data Structures that Leverage SAS Arrays***

Troy Hughes and Louise Hadden

This paper introduces a data-driven approach to creating dynamic decision tables. The tables can have any number of variables, any number of columns, and any number of choices under each column. One example is provided herein, although we will be providing a second to demonstrate the insane level of flexibility that this solution offers.

(KEYWORDS: DECISION TABLES, DECISION RULES, DYNAMIC, DATA-DRIVEN, CONDITIONS, CONSTRAINTS, DATA RULES, ARRAYS)

Troy is a two-time SAS author who consults with the Department of Defense. In his spare time, he enjoys long, rum-induced walks on Monterey beaches.

Louise has been using and loving SAS since the days of computers the size of not-so-tiny houses! She is also the girl with the SAS tattoo!

Better To Be Mocked Than Half-Cocked: Data Mocking Methods To Support Functional and Performance Testing of SAS Software**Troy Hughes**

Data mocking refers to the practice of manufacturing data that can be used in software functional and performance testing, including both load testing and stress testing. Mocked data are not production or “real” data, in that they do not abstract some real-world construct, but are considered to be sufficiently similar (to production data) to demonstrate how software would function and perform in a production environment. Data mocking is commonly employed during software development and testing phases and is especially useful where production data may be sensitive or where it may be infeasible to import production data into a non-production environment. This text introduces the MOCKDATA SAS® macro, which creates mock data sets and/or text files for which SAS practitioners can vary (through parameterization) the number of observations, number of unique observations, randomization of observation order, number of character variables, length of character variables, number of numeric variables, highest numeric value, percentage of variables that have data, and whether character and/or numeric index variables (which cannot be missing) exist. An example implements MOCKDATA to compare the input/output (I/O) processing performance of SAS data sets and flat files, demonstrating the clear performance advantages of processing SAS data sets in lieu of text files.

(KEYWORDS: MOCKING DATA, PERFORMANCE TESTING, LOAD TESTING, STRESS TESTING, EFFICIENCY, SYSTEM RESOURCES)

Troy has been a SAS practitioner for more than 20 years, has managed SAS projects in support of federal, state, and local government initiatives, and is a SAS Certified Base, Advanced, and Clinical Trials Programmer. Since 2013, he has given more than 80 presentations, hands-on workshops, and trainings at SAS Global Forum, SAS Analytics Experience, WUSS, MWSUG, SCSUG, SESUG, and PharmaSUG, and has authored two groundbreaking books: - SAS Data Analytic Development: Dimensions of Software Quality (2016) - SAS Data-Driven Development: From Abstract Design to Dynamic Functionality (2018) Troy has an MBA in Information Systems Management and additional certifications, including: Project Management Professional (PMP), Risk Management Professional (PMI-RMP), Professional in Business Analysis (PMI-PBA), Agile Certified Professional (PMI-ACP), Certified Information Systems Security Professional (CISSP), Certified Secure Software Lifecycle Professional (CSSLP), ITIL Foundation, Certified ScrumMaster (CSM), Certified Scrum Developer (CSD), Certified Scrum Product Owner (CSPO), and Certified Scrum Professional (CSP). He is a US Navy veteran with two tours of duty in Afghanistan.

Data Governance: Harder, Better, Faster, Stronger**Peter Baquero**

All the major trends call for advanced control and accountability toward the use of data. From the migration to cloud applications and storage, to the deployment of big data environments, the democratization of analytics and artificial intelligence (AI), and the increasing requirements for data privacy and data protection—data governance has changed from something that is nice to have to being a must-have, with an ever-expanding scope to address. Gone are the days of marketing databases, some ERP processes, or specific regulations such as the Solvency 2 Directive or BCBS 239 being the limit. Most organizations came through strong challenges, aligning people and processes, and trying to sustain the governance effort; progressively this dream of enterprise data governance is fading. Organizations are now looking at more surgical initiatives to take control of their data lakes, and to ensure that their analytical processes are fed with reliable information and that their data privacy policies are enforced. They want results immediately. In this session, we look at how data governance can be smarter, how it can be automated, and how it can be fun by relying on analytics and AI.

(KEYWORDS: DATA GOVERNANCE, DATA PRIVACY, AUTOMATION WITH ANALYTICS AND AI)

Peter Baquero is a risk and data governance expert with over a decade of experience in development and implementation of enterprise analytic solutions. As a CPA Peter has been heavily involved in regulatory financial reporting for the financial services industry with a focus on solution design and implementation for IFRS 9/CECL and DFAST/CCAR. In addition to regulatory reporting, Peter also has extensive experience implementing high performance analytics in Big Data environments. Peter is a certified Base SAS programmer and has programmed in SAS for over 10 years. Before joining SAS Peter was a Risk Analytics senior consultant with Deloitte.

e-Posters

EP-105

Examining Imputation effect by Using FACTOR and MI Procedures in SAS® for a Bullying Scale for LGBTQ Youth in SC

Abbas Tavakoli, Laura Hein and Mary Cox

Missing data presents a challenge to researchers. This study used LGBTQ youth data from SC to test the effect of an imputation using FACTOR and MI procedures in SAS. There are many reasons for missing data. Factor analyses were run to develop a Bullying Scale for LGBTQ youth. These runs included no imputation, single imputation, and multiple imputation (1000 times) for missing data. Two factors emerged - hearing bullying and experiencing bullying. The inter-factor correlation was .46 for hearing bullying and experiencing bullying. Similar results for factor extraction for no imputation, single imputation, and multiple imputation were found. Testing revealed all reliability coefficients exceeded .80 with no imputation and with imputation using the SAS®6 PROC FACTOR and STANDARD, and MI procedures for data analysis.

(KEYWORDS: SAS, FACTOR, MI, BULLYING, LGBTQ)

Abbas Tavakoli currently work as Clinical Associate Professor with college of Nursing at the University of South Carolina. He has worked with office of research College of Nursing since 1992. He worked with Health Statistics at Raleigh (NC) from 1990 to 1992. His job entails teaching, involves with research, and using many statistical procedures. He has taught Statistics courses for doctoral nursing students since 2004. Dr Tavakoli taught Bios700 (introduction to Biostatistics) and Bios710 (Effective Data management in Public Health) for graduate students in the School of Public Health. He has served as a data manager, biostatistician, and research team member for seven previous NIH-funded R01 grants and many smaller grants that have required data management, display and analysis plans. He has assisted principal investigators to collect, manage, analyze, and present high quality data. He has involved many research project and manuscript.

Using SAS® to examine peripheral intravenous access using ultrasound guided
Abbas Tavakoli, Courtney Prince and Stephanie Burgess

Patients with difficult access usually undergo a central line or peripheral inserted central catheter (PICC) placement. Central venous access is more invasive, time consuming and prone to serious complications. Preventing complications related to central lines is an ongoing goal for healthcare providers, insurers, regulators and patient advocates. Establishing peripheral intravenous (PIV) access is a pivotal step in providing care for patients in hospital settings. This study analyzed data from a convenient sample of 70 of adult patient being treated in one hospital in South Carolina. These patients were identified as having difficult venous access by one of the nurses. After two failed attempts by nurses a consult requesting the vascular access team for this study. Five nurses are participating in the traditional insertion group and USGPIV group. Nurses are collecting randomized data and information using traditional coin flip-selections during a 31-day trial. Nurses participating in the project will complete online training modules, followed by didactic and hands-on training. Data is generated for the quality improvement project via nurses completing questionnaires designed to capture USGPIV and traditional PIV success rates, number of attempts required for successful peripheral access, and time used to place PIV's. Proc Mean and Freq used to describe the data. Proc Ttest, Npar1way, and Corr used examine peripheral intravenous access using ultrasound guided. The result indicated that the means of minutes to obtain IV, number of attempts, cost in salary, and cost for equipment were higher for traditional group as compare to USGPIV. The Chi-square and fisher exact test showed there was statistically significant between success rate and group (P value <.0001). All data analyses were performed using SAS/STAT® statistical software, version 9.4

(KEYWORDS: SAS, PERIPHERAL INTRAVENOUS, ULTRASOUND)

Abbas Tavakoli currently work as Clinical Associate Professor with college of Nursing at the University of South Carolina. He has worked with office of research College of Nursing since 1992. He worked with Health Statistics at Raleigh (NC) from 1990 to 1992. His job entails teaching, involves with research, and using many statistical procedures. He has taught Statistics courses for doctoral nursing students since 2004. Dr Tavakoli taught Bios700 (introduction to Biostatistics) and Bios710 (Effective Data management in Public Health) for graduate students in the School of Public Health. He has served as a data manager, biostatistician, and research team member for seven previous NIH-funded R01 grants and many smaller grants that have required data management, display and analysis plans. He has assisted principal investigators to collect, manage, analyze, and present high quality data. He has involved many research project and manuscript.

Using SAS Hash and Hiter Objects to Compute the Ability Levels that Correspond to the Rasch Model's Response Probabilities

Imelda Go

This paper illustrates how to use hash and hiter objects to determine the ability level that corresponds to an item's particular response probability (RP) under the Rasch Model. As an example, the Rasch model's RP50 (i.e., RP of 0.50) is identical to the item's difficulty parameter.

(KEYWORDS: HASH-HITER OBJECTS, RESPONSE PROBABILITIES, RASCH MODEL)

Imelda "Mel" Go, Ph.D. is a Data Quality SAS Programmer with Questar Assessment's Psychometric and Research group. Prior to that she was with the South Carolina Department of Education's Office of Assessment for 15 years and with two SC public school districts for 7 years. She uses SAS to analyze student test data and to perform high-stakes calculations.

EP-182 ***SAS PROC IMPORT Troubleshooting Guide***

Imelda Go and Wendi Wright

Although the task often appears simple, creating data sets from text files with PROC IMPORT can be tricky. This paper takes you through a progression of considerations and complications due to the way PROC IMPORT creates data sets. The discussion focuses on potential problems caused by different factors and, in part, by relying on SAS to define variable attributes; and how to avoid or solve such problems.

(KEYWORDS: PROC IMPORT)

Imelda "Mel" Go, Ph.D. is a Data Quality SAS Programmer with Questar Assessment's Psychometric and Research group. Prior to that she was with the South Carolina Department of Education's Office of Assessment for 15 years and with two SC public school districts for 7 years. She uses SAS to analyze student test data and to perform high-stakes calculations.

Wendi Wright has extensive experience in writing custom analysis and reporting programs in the educational testing industry. She has enjoyed presenting in many SAS conferences and enjoys the teaching aspect of both the conferences and her job.

EP-196 ***Ms. Independence (from the SAS® Format Library)***

Nancy McGarry and Louise Hadden

SAS® practitioners are frequently called up to format variables in SAS datasets they have received or created for various use cases. Analysts and other end users desire the convenient categorization, transformative nature, and attractive appearance that SAS formats can lend to variables for reports and further analytic and data set construction purposes. SAS formats can be created in SAS work space, and can be stored permanently in SAS catalogs, a specially purposed container for SAS files. SAS formats created in work space are ephemeral, and only exist for the duration of a SAS session. SAS formats stored in a SAS catalog are notoriously difficult to transfer across platforms, SAS versions and "bit" versions (32 bit vs 64 bit). Recipients of SAS data sets with "embedded formats" and/or SAS catalogs originating from incompatible systems find themselves in a quandary – SAS reports errors when it can't find a compatible catalog (IF a catalog accompanies a data set) for formats permanently associated with variables in a SAS data set. SAS catalogs are also very difficult to update, document and manipulate. This paper and presentation / poster propose straightforward SAS solutions for the creation, transfer and use of SAS formats.

(KEYWORDS: PROC FORMAT, CNTLIN, %INCLUDE)

Louise Hadden presented at her first SAS conference in 1996 and has never looked back, presenting at multiple conferences across the continent over the years. She supports analytic processing at Abt Associates Inc., a social science research company, and specializes in reporting and data visualization in the division of health and environment.

EP-202

Smoothing 3D drug overdose death data and displaying patterns with SAS/JMP

Nancy Han and Rong Wei

Many data presentation methods used to examine the trending behavior of drug overdose death rates by age over time rely on two dimensional line graphs with age defined as a categorical variable. With the use of SAS/JMP procedures, visual graphic methods can be used to explore patterns and relationships between age and year of death on death rates with three dimensional (3D) plots. The drug overdose death rates were calculated using mortality data from the National Vital Statistics System public use multiple cause-of-death files from 2007 to 2017. The SAS/JMP 3D plot software displays the changing death rates by age over the trending period when age is treated as a continuous variable rather than a grouped variable. Analysis is further enhanced, by the capability of the software to view the graphics from different angles. In addition to 3D data plots, locally smoothed 3D plots can be produced by SAS procedures prior to JMP plotting. SAS/TSPINE is used to produce the examples.

(KEYWORDS: U.S. TRENDING OF DRUG OVERDOSE DEATH RATES BY AGE, 2007-2017, 3D VISUAL GRAPHICS, SAS/TSPINE , JMP)

Nancy Han is an IT Specialist at Division of Analysis and Epidemiology, National Center for Health Statistics where she has been since 2011. Nancy uses SAS® in her daily responsibilities of data analyses. She has used SAS for more than 10 years.

EP-228

Utilizing SAS Functions to Generate Accurate Adherence Notifications for Clinical Trials

Lishu Zhang, Jiu Zhao and Wenjun He

For clinical trial studies, it is critical to keep close contact with participants after main study procedures are administered or during follow-up phases. Furthermore, it could be challenging to monitor participants and reinforce their adherence to medication taking, laboratory sample collection, or periodic doctor's visits outside clinical settings. Currently mobile technology has emerged to play an essential role by sending out automated simple reminders and notifications to maintain communication among participants, study coordinators and principal investigators. Accurate notifications can significantly improve study compliance and treatment efficacy. This poster will present a real case of how SAS® 9.4 can be utilized as a powerful and flexible tool to create a notification schedule that can be used to remind subjects to take dry blood samples at home for a clinical trial study. Notification dates and time were chosen based on the individual participants' given preferences. The poster will demonstrate step-by-step on how to apply LAG and INTNX functions to generate this type of notification using diagrams and actual programming code. It will also provide examples of how to logically implement changes of participant preferences during the course of study without jeopardizing the scientific rationale and study schedule.

(KEYWORDS: SAS LAG FUNCTION INTNX FUNCTION NOTIFICATION CLINICAL TRIAL)

EP-232

A SAS Journey From Tables in the Database to Graphics in your Inbox

Raghav Adimulam

The journeys that data can take from its origins in databases to its final destination in the form of reports and graphics are often diverse. This poster shows one such journey of data residing in SQL Server that ultimately becomes meaningful graphics and reports. Along the path, complex Macros are used to loop through a SAS dataset using Call Execute routine in a data step. The macro calls are automatically generated with macro parameters coming from dataset variable values. The various intricacies present in the Format, Report and SGPlot procedures are used to generate fixed structure reports and plots wrapped in the ODS destination sandwiches of Excel and PDF. The convenience of emailing the reports to users is accomplished by the Task Scheduler running on a SAS server. In this way, a complex set of data is converted to expressive graphics with the help of SAS software.

(KEYWORDS: GRAPHICS ODS MACROS EXCEL REPORT)

Raghav Adimulam is a Lead SAS Programmer for many projects at Westat and has also been the In-house Instructor of SAS classes for Westat Staff.

He has over 15 years of experience working in SAS Software in the fields of Financial, Pharmaceutical and Government domains.

He holds the Professional certifications of 'SAS Certified Advanced Programmer for SAS 9' and 'SAS Certified Base Programmer for SAS 9' credentials. Previously, he presented Papers at the SAS Global Forum and SESUG conferences.

EP-266

Sample size calculation using the win ratio approach for hierarchical composite of three outcomes

Matheos Yosef, Shokoufeh Khalatbari and Scott Hummel

Clinical trials and observational studies often involve multiple outcomes with hierarchy in terms of clinical importance. As a result, a composite outcome is commonly chosen as the primary endpoint for testing the treatment effect as well as determining the required sample size. Finkelstein & Schoenfeld (1999) proposed a non-parametric hierarchical testing of treatment effect for a composite endpoint of two outcomes. Pocock introduced the 'win ratio' estimator, a new approach to the analysis of hierarchical composite outcomes that accounts for clinical priorities of multiple outcomes. In this poster, we present our work on extending Pocock's SAS macros to compute the Finkelstein & Schoenfeld (FS) test and win ratio for a hierarchy of three outcomes, and our SAS program to calculate sample size and power for a clinical trial involving a composite of three outcomes.

(KEYWORDS: CLINICAL TRIALS, WIN RATIO, FINKELSTEIN-SCHOENFELD SCORE, HIERARCHICAL COMPOSITE OUTCOMES)

Cutting Out the Middle Man – Bypassing ADaM to Calculate Tumor Response Inputs using RECIST 1.1 Guidelines and Collected Data in SDTM Format**Amber Frazier**

Confirming vendor results? Providing recon listings to the client? No time or need for ADaM? Cut out the middle man! A high-level overview on how to use CRF data in SDTM format to reconcile tumor response assessment using Response Evaluation Criteria in Solid Tumor (RECIST 1.1) guidelines, from collecting the basic required data points to deriving the Best Overall Response. Tumor response categorization is key in monitoring patient disease progression and best overall response is frequently an analysis endpoint in Oncology studies. The RECIST 1.1 Guidelines offer a standard process for data collection and calculation for target, non-target, and new lesion progression based on assessments at timed intervals, generally 6-12 weeks apart. Efficient data analysis starts with effective data collection using simple case report forms (CRF) and ensuring all required data items are accounted for. Ideally, the CRF data is structured to easily fit into SDTM format using Tumor Identification, Tumor Results and Tumor Response domains (TU, TR, and RS). Using identification fields such as –LINKID and –LNKGRP, the domains can be merged to obtain nadir, diameter sums, and reduction percentages to derive response values per visit and overall.

(KEYWORDS: CRF DATA, SDTM FORMAT, ONCOLOGY STUDIES, RECIST 1.1 GUIDELINES)

Amber Frazier is an NC State graduate in Statistics, born and raised in Raleigh, NC. She is currently employed with PPD, Inc. as a statistical programmer analyst. She enjoys programming challenges, learning new programming techniques, and attending these conferences. She hopes to one day retire in Sarasota, Florida.

Education/Institutional Research

ED-159

Complete your IPEDS Completions in 60 minutes or less (using SAS and Colleague)

Yolanda Ingram

This workshop is designed for Institutional Research professionals who are responsible for completing the federal IPEDS Completions Survey. In this session, you will learn how to complete your IPEDS Completions report in 60 minutes or less by using a SAS, Colleague and Notepad++. The outcome will be an electronic file that can be uploaded by the key holder to minimize the need for manual data entry and to improve efficiency and accurate reporting. Learning Outcomes Implement analytical tools to Improve efficiencies in data reporting. Plan, design, implement and modify on an annual basis an automated method for IPEDS reporting.

(KEYWORDS: IPEDS, COMPLETIONS, INSTITUTIONAL RESEARCH)

ED-163

UCF's SDES Dashboard A Collection of Maintenance-Free SAS® Web Report Studio Reports

Carlos Piemonti

At the University of Central Florida (UCF), Student Development and Enrollment Services (SDES) combined efforts with Institutional Knowledge Management (IKM), which is the official source of data at UCF, to venture in a partnership to bring to life an electronic version of the SDES Dashboard at UCF. A collection of eleven reports was created using SAS® Web Report Studio. It just takes a few minutes to refresh, for an academic year cycle, 67 sections with more than 400 indicators in those eleven reports. Challenges in the design, implementation, usage, and performance are presented.

(KEYWORDS: UCF, DASHBOARD, WRS, MAINTENANCE-FREE)

ED-179

Prediction of College Admission Trend Using Predictive Analytics

Yash Prakash And Miriam Mcgaugh

The number of applications received by the universities and higher educational institutions is increasing every year. Out of these huge number of applications, limited number of applicants are provided with offer of admission and out of it, only few applicants accept the offer. Therefore, it is very important for the universities and institutions to offer the admission only to the prospective applicants who are more likely to join their universities and institutions. A prospective student can be identified by finding interactions of applicants with the college over phone, direct mails, e-mails and other communication channels.

This paper uses the power of predictive analytics to identify the factors influencing admission offer acceptance by the applicants and build operationalized models to predict the acceptance rate based on students' communication with the university via e-mail. The data used in this research has been provided by Marketing and Student Communication Center, Oklahoma State University. The dataset contains all the email communications date wise between the university and the students of the past three years starting from application submission to confirmed admission. The data set also contains the applicants' demographic information such as Active City, Region, Postal code, Gender, Race, First generation, etc. Also, the email communication between the student and the University are classified into categories such as Prospects, Inquirers, Applicants, Admitted, and Confirmed. The data is prepared and analyzed using different SAS tools like SAS Enterprise Guide®, SAS Viya® and SAS Enterprise Miner®. Variables such as Gender, Race, Inquiry Date, Application Date, Percent Clicked, Percent Opened and Application Submission Date are found to be the most important factors affecting the admission decision of the candidate.

(KEYWORDS: PREDICTIVE ANALYTICS, UNIVERSITY ENROLLMENT, ACCEPTANCE RATE)

ED-197

Making Life Easier on the SAS End: Best Practices for Collecting Survey Data

Keli Sorrentino and Julie Plano

An analysis-ready SAS dataset is the deliverable, the one and only thing the analysts are waiting for. It is the goal of any research project. Much effort goes into the collecting and cleaning of data before it is ready to be analyzed. Our research team has spent decades learning to design good data collection instruments, resulting in purposeful survey questions with practical response values. Whether the questionnaire is administered in person or collected via online software, it is important to begin with a clear and concise plan. Here we share our expertise and best practices to help research teams design a plan to efficiently get from data collection to an analysis ready dataset. Examples include using Qualtrics Research Suite to design surveys, gather data, and export tables. SAS 9.4 is used to create a functional dataset for further investigation and analysis.

(KEYWORDS: BASE SAS, QUALTRICS, SURVEY RESEARCH, DATA MANAGEMENT)

Improving Student Application Reporting Using a Slowly Changing Dimension and SAS Data Integration Studio**Lauren Schoenheit and Alexander Fantroy**

For a selective university shaping the incoming freshman class requires current operational data on the student applications that have been received, the admissions decisions that have been made, and the students that have committed in response, throughout the admissions cycle. A point in time comparison with historical data serves as a benchmark to better understand and anticipate the final makeup of the incoming students. To meet these demands one such institution shifted student application reporting from a base SAS program run on an ad hoc basis, that created a SAS dataset of the application information for that day saved to a drive; to a scheduled Data Integration Studio job that created a slowly changing dimension of year over year daily application information stored as an oracle table in the university's enterprise database. Changing the way data was stored reduced storage space, as only new applications and change records are added to the student application table instead of accumulating datasets with a repeated record for every day the application was in the system. Scheduling a Data Integration Studio job removed the task of running the program from a person's workload as well as standardized the collection of student application data. Together these changes facilitated more frequent snapshots of student application data as well as increased processing efficiency. A summary is loaded to SAS Visual Analytics nightly as the basis for a daily year over year analysis of student application decisions. This presentation will cover the business reasons for making this change, the technical details of creating the Data Integration job and the slowly changing dimensions, as well as screen caps of the Visual Analytics student application report.

(KEYWORDS: ADMISSIONS)

Clarifying the Assessment Picture: Defined Metrics and Programming Choices in SAS Coding**Jerry Landry and Kelly Smith**

Healthy Programs Reporting at Central Piedmont Community College was developed as an opportunity to allow programs to conduct a program review using metrics common to all programs at the institution along with program-specific or industry-specific outcomes. Ensuring that programs were able to assess their performance beyond standard institutional measures was seen as key in keeping the process focused on student success. The program review method was developed in a partnership between the College's Learning Unit and the Office of Planning and Research. Some of the metrics in the review require data provided by Planning & Research, which utilizes Base SAS and Enterprise Guide for reporting. In order to facilitate the writing of the SAS code, clear metric definitions had to be agreed upon by all project stakeholders to ensure that curriculum programs understood what data was being provided, that common parameters were established for analyzing applicable data points across programs, and that the data was meaningful for the programs. In the development process, data also had to be gathered from programs about their student intake process and timeline in order to establish the time parameters of the data report that would be provided. Shared in this presentation will be best practices from the development of Healthy Programs Reporting and the utilization of SAS in the process, with a particular focus on portions of SAS code which demonstrate programming choices to improve efficiency and ease of future adaptation.

(KEYWORDS: PROGRAM REVIEW, PROGRAM EVALUATION, OUTCOMES ASSESSMENT, PERFORMANCE MEASURES, DATA ANALYSIS, CODING)

In his role at Central Piedmont, Jerry Landry has been involved in various strategic planning, program assessment, and process improvement efforts. He strives to make required practices both efficient and meaningful for programs with the end goal of student success always in focus.

Kelly D. Smith has worked with SAS for over three years, using Base SAS and Enterprise Guide on a daily basis as senior institutional research analyst with Central Piedmont Community College. She is a firm believer in using SAS to improve the accuracy and efficiency of the reporting process, and hopes this presentation will provide timely and useful information to other SAS coders.

Hands on Workshops

HOW-127

Leveraging "UNIX Tools" (GNU) for Data Analysis

David Horvath

Life would be so much easier if everything was in a database or pulled via API. But that is not the case. All too often we get data files (or have to send them) in various formats. This session discusses some of the tools available to help you figure out what the file looks like so you can pull it apart using those tools or your tool-of-preference. While the GNU version of these tools will be the focus, the skills learned apply to many different platforms (Cygwin under MS Windows, MAC OSX, the Linux core of Android, commercial Linux — like Red Hat Enterprise, and commercial UNIX — like IBM's AIX or Sun/Oracle's Solaris). Of particular interest are 'head', 'tail', 'wc', 'awk', 'dd conv', and shells. A few of the differences between UNIX/Linux and Windows will also be discussed in case you ever have to deal with those environments in our heterogeneous environments. This knowledge also comes in handy if you need to migrate code from an existing UNIX/Linux-based application. There are times that the available tools make life so much easier!

(KEYWORDS: ADVANCED TOOLS, UNIX TOOLS, LINUX, UNIX, SHELL, GNU, WINDOWS, DATA ANALYSIS)

David is an IT Professional who has worked with SAS, off and on, since the late 1980's using it as a data processing (4GL/ETL) and analysis tool. He has presented at PhilaSUG previously and for other user groups and organizations (workshops and seminars) in Australia, France, the US, and Canada. His Masters is in Organizational Dynamics from UPENN, has consulted with CHERP at the VA hospital, and currently works for a regional bank in Risk Analytics Infrastructure at the Wilmington DE location. He has several books to his credit (none SAS-related) and is an Adjunct Instructor covering IT topics.

HOW-148

Interactive Graphs

Richann Watson and Kriss Harris

This paper demonstrates how you can use interactive graphics in SAS® 9.4 to assess and report your safety data. The interactive visualizations that you will be shown include the adverse event and laboratory results. In addition, you will be shown how to display "details-on-demand" when you hover over a point. Adding interactivity to your graphs will bring your data to life and help improve lives!

(KEYWORDS: ODS GRAPHICS, GRAPH TEMPLATE LANGUAGE, DRILL DOWN, INTERACTIVE GRAPHS, LAYOUT)

Richann Watson is an independent statistical programmer and CDISC consultant based in Ohio. She has been using SAS since 1996 with most of her experience being in the life sciences industry. She specializes in analyzing clinical trial data and implementing CDISC standards. Additionally, she is a member of the CDISC ADaM team and various sub-teams.

HOW-199

Using PROC TABULATE and ODS Style Options to Make Really Great Tables

Wendi Wright

We start with an introduction to PROC TABULATE, looking at the basic syntax, and then building on this syntax by using examples. Examples will show how to produce one-, two-, and three-dimensional tables using the TABLE statement. More examples cover how to choose statistics for the table, labeling variables and statistics, how to add totals and subtotals, working with percents and missing data, and how to clean up the table using options. A look at the three ways you can use the PRELOADFMT option is also covered. Next the ways to use the ODS STYLE= option in PROC TABULATE are covered. This option helps to customize the tables and improve their attractiveness. This option is very versatile and, depending on where the option is used, can justify cells or row and column headings, change colors for both the foreground and background of the table, modify lines and borders, controlling cell width, add a flyover text box in ODS HTML, or add GIF figures to the row or column headings.

(KEYWORDS: TABULATE, ODS, STYLES)

Wendi Wright has extensive experience in writing custom analysis and reporting programs in the educational testing industry. She has worked at several educational testing companies, and is currently at Data Recognition Corp.

HOW-208

Doing More with the SGPLOT Procedure

Josh Horstman

Once you've mastered the fundamentals of using the SGPLOT procedure to generate high-quality graphics, you'll certainly want to delve in to the extensive array of customizations available. This workshop will move beyond the basic techniques covered in the introductory workshop. We'll go through more complex examples such as combining multiple plots, modifying various plot attributes, customizing legends, and adding axis tables.

(KEYWORDS: SAS PROGRAMMING, ODS GRAPHICS, SGPLOT)

Josh Horstman is an independent statistical programming consultant and trainer based in Indianapolis with over 20 years' experience using SAS in the life sciences industry. He specializes in analyzing clinical trial data, and his clients have included major pharmaceutical corporations, biotech companies, and research organizations. Josh is a SAS Certified Advanced Programmer who loves coding and is a frequent presenter at SAS Global Forum and other SAS User Group meetings.

HOW-213

Introduction to Propensity Score Modeling and Treatment Effect Estimation

Jason Brinkley

Measuring the impact of treatments and interventions is an important aspect to all areas of evaluation and research. While a randomized experiment or clinical trial may be the gold standard for providing causal information regarding effectiveness, it is often the case that we must rely on observational data and secondary sources for measuring effectiveness. Confounding plays a huge role in measuring and evaluating effectiveness and statistical adjustments for confounding are a focal point of modern analytics. In many cases it is not enough to perform simple adjustments via outcome regression models to control for confounding, especially in cases where such confounding plays a role in the very selection of who gets which intervention of interest. Propensity score models have become a popular route for creating balance in observational data by implementing a pseudo or quasi-experimental set of conditions on the data of interest. This workshop is designed to give a soft introduction and background to propensity score modeling and associated treatment effect estimation. We will discuss the motivation for counterfactual data analysis and apply real world data (from the medical literature) to explore how to use the PSMATCH and CAUSALTRT procedures in SAS. Attendees need only have a working knowledge of multiple regression analyses.

(KEYWORDS: CAUSAL INFERENCE, PROPENSITY SCORE, REGRESSION MODELING)

Jason S. Brinkley, PhD, MS, MA is a Senior Researcher and Biostatistician at Abt Associates Inc. where he works on a wide variety of data for health services, policy, and disparities research. He maintains a research affiliation with the North Carolina Agromedicine Institute and serves on the executive committee for the NC Chapter of the American Statistical Association and the Southeast SAS Users Group. Follow him on Twitter, @DrJasonBrinkley.

HOW-231

Getting Started with PROC DS2

Jim Blum and Jonathan Duggins

This workshop is designed to give DATA step programmers foundational information to develop programs in PROC DS2. Starting with several common tasks given as DATA step program examples, the workshop goes through transitioning the code examples to PROC DS2 code step-by-step. As part of the process, various similarities and differences between the two steps are noted, and pros and cons of using each are discussed. Suggested topics for study for building on the PROC DS2 concepts presented are also provided, along with reference material to aid in further study.

(KEYWORDS: DATA STEP, DS2, ARRAYS, FUNCTIONS, METHODS)

Jim Blum is a Professor of Statistics at the University of North Carolina Wilmington where he has developed original courses in SAS programming for the university over the past 15 years. These courses cover topics in base SAS, SAS/GRAPH—recently transitioned to template-based graphics, SAS/SQL, SAS/STAT, and macro language. Jim will be a primary instructor in UNC Wilmington’s Data Science program debuting in fall 2017.

HOW-255

Easier and Better with Enterprise Guide

Laura Oliver

Enterprise Guide is a great Integrated Development Environment that helps people do the full advanced analytics lifecycle through a point-and-click or hard-core programming interface. Many times, though, the user hasn't had time to investigate all of the features that can be done and ways to make their life easier. For example, the GUI environment has good ways to accomplish tasks that had to be done in a program in the past. This Workshop is valuable for anyone who wants to use Enterprise Guide better, whether you are a user who is new to Enterprise Guide, or one who is experienced. It will cover such topics as Data Step Debugger, formatting code, ways to view your system options and macro variable options. It will also cover Data Explorer and ways to use EG Tasks to make the rest of your life easier.

(KEYWORDS: ENTERPRISE GUIDE, TIPS, SYSTEM OPTIONS, MACRO VARIABLES, DATA STEP DEBUGGER, DATA EXPLORER, TASKS)

HOW-256

Dashboards in Visual Analytics

Laura Oliver

Visual Analytics is a great tool to use to visualize and analyze your data and create dashboard for others to review data. This Workshop is designed for users that are new to report creating in Visual Analytics. It will give you hands on experience creating a powerful and interactive dashboard. We will build one with multiple reports and features to allow filtering of data.

(KEYWORDS: DASHBOARDS, VISUAL ANALYTICS, FILTERS, VISUALIZE ANALYZE, INTERACTIVE REPORTS)

Government

GT-134

Identify physical and mobile DMV Site Locations in North Carolina

Majed Al-Ghandour, Burcu Adivar, Rajeshwar Dutt, Puneet Kumar and Matthew Rogers

Growing population and changing demographics in North Carolina result in increased demand for DMV services, specifically for Real ID issuance. Considering the geographic distribution and the spatial characteristics of the demand, decision makers need to open new locations and/or reallocate limited resources among existing DMV locations to improve the operational efficiency and customer experience. The objective of this study is to provide an integrated approach for selecting the optimal DMV locations using expert knowledge, data mining, Analytic Hierarchy Process (AHP), Geographical Information System (GIS) and SAS Software. The proposed approach identifies sixteen location criteria through experts' input as part of the AHP process, yielding demographic attributes, flexibility, efficiency, cost and access to public facilities. Following the weight assessment for all criteria and sub-criteria, normalized weights are used for location suitability analysis in ArcGIS. Based on our projections for the demand and related geospatial data, alternative DMV locations are determined and visualized through ArcGIS. Finally, the alternative locations are evaluated by AHP weights and the multi-criteria location selection problem is optimized to maximize the coverage across the state.

(KEYWORDS: NCDOT-DMV ,SAS, NC REAL ID, MCLP,AHP)

GT-234

Scoring Classification and Recidivism in Corrections Through Outcome Analytics

Alan Mann

A data-driven design for inmate classification in corrections is proposed as a fact-based, analytic modeled alternative to the possibility of confirmation bias in risk assessment of prisoners entering incarceration, using retrospective assessment of recidivism risk measures of inmate data from New York State and Florida respectively. KPIs of Class Severity and predicted Custody Risk groupings based upon Offense and severity of New York State Offense Classification sentencing guidelines were a main result in creating an objective analysis rather than a subjective one. While initial findings from New York indicated a random and highly subjective pattern of classification, with class A misdemeanor offenders placed in Maximum custody, and class A-I felony offenders in Minimum, subsequent identification of objective markers in the New York, and Broward County, Florida data through Decision Tree and K-means models proved a more accurate, stronger scored, categorical and correlative relationship.

(KEYWORDS: LOGISTIC DECISION TREE K-MEANS CORRECTIONS RECIDIVISM CLASSIFICATION CONFIRMATION BIAS SUBJECTIVE)

Alan Mann has worked as a SAS resource and consultant since 1989, beginning as a programmer for Dupont Epidemiology, through now as a data scientist/statistician on assignment for the U.S. Treasury. Always alert in seeking new solutions with SAS and other analytic tools, he has made a career of delivering innovation to his clientele. He lives in Harpers Ferry, WV with his wife Louise, and two cats. He holds a BA from University of Delaware, a Postgraduate Diploma in Information Technology from the University of Liverpool, and a MS in Analytics from Capella University.

Healthcare/Pharmaceuticals

HP-120 ***Pooled Database Loading (PDL) – From Documentation to Code to Deliverable***
David Hartman and Pranjali Dafe

The combining of 2 or more studies into a single database is very difficult and labor intensive. While it is virtually impossible to eliminate human intervention, a process has been created where documentation containing everything from what studies are to be pooled to the mapping rules that are to be used to convert data from multiple studies into a common structure. This information (in the form of spreadsheets) is then executed via a tool called PDL (Pooled Database Loader) to produce the pooled database. This process is completely data independent and can pool data from multiple studies together regardless of the structure of the “original” data. The closer the study level data is to the CDISC standard structure, the more the mapping of a study will be created at the push of the button. The purpose of this paper is to illustrate how a user of PDL goes from individual study data to a final pooled dataset. Included in this paper will be examples of the information contained in the spreadsheets as well as the various calls to PDL that will create the pooled datasets.

(KEYWORDS: POOL, DATABASE, STUDIES, DOCUMENTATION, POOLED DATABASE)

HP-185 ***Data-driven Programming Techniques using SAS® Macros to Semi-automate Generation of Descriptive Tables in Healthcare Research***
Katie Mercaldi

Nearly all healthcare studies include one or more tables with descriptive statistics summarizing characteristics of the sample population. When the relevant variables in the SAS analytic data set are properly formatted and labeled, the process of producing publication-quality descriptive tables can be streamlined so that little more is needed than the names of the variables to use as table columns and rows. This process is implemented with four SAS macros -- %COLCNTL, %ROWCNTL, %MKTABLE, and %MKREPORT. The SAS macros %COLCNTL and %ROWCNTL create “control files” in the form of SAS data sets that act as sets of instructions for designing the table’s columns and rows respectively. The macros extract relevant metadata, including variable attributes such as labels and formats, about the column and row variables found in the specified analysis data set. Using this metadata, these macros then determine what type of summary statistics to produce for each row variable, such as means with standard deviations for continuous variables and counts with percentages for categorical measures. The SAS macros %MKTABLE and %MKREPORT generate the final output table. The %MKTABLE macro creates a version of the table as a SAS data set by converting the metadata to lists of macro variables using the INTO clause in PROC SQL or, alternately, by constructing macro calls within a data step with CALL EXECUTE. Finally, the %MKREPORT macro generates PROC REPORT syntax that creates the final deliverable-quality table in an Excel workbook.

(KEYWORDS: HEALTHCARE, DESCRIPTIVE STATISTICS, MACRO, METADATA, EXCEL)

Katie Mercaldi, MPH, is a Senior Data Analyst in Evidera’s Real-World Evidence group based in Waltham, Massachusetts. Ms. Mercaldi has served as both an analyst and epidemiologist on a variety of research studies in her work at Evidera, focusing mainly retrospective observational studies of insurance claims and other large, relational healthcare databases. She has nearly 10 years of SAS programming experience, with scientific and analytic works published in journals including Chest, Stroke, Diabetes Care, and American Journal of Clinical Nutrition.

HP-239 ***Analyzing Hospital Medicare Cost Report Data Using SAS® - Updated with Output***
Kimberly Andrews

Medicare-certified institutional health care providers are required to submit annual cost reports, which are maintained by the Centers for Medicare and Medicaid Services (CMS) in the Healthcare Cost Reporting Information System (HCRIS). Medicare Cost Reports (MCR) contain provider information such as facility characteristics, utilization data, total and Medicare costs, inpatient and outpatient charges, Medicare payment data, and financial statement data. HCRIS includes the following subsystems: Hospital, Skilled Nursing Facility (SNF), Home Health Agency (HHA), End-Stage Renal Disease (ESRD) Facility, Hospice, Community Mental Health Center (CMHC), and Rural Health Clinic (RHC)/Federally Qualified Health Center (FQHC). Our discussion focuses on the MCR for Hospitals (the most complex of the cost reports) explaining how to access the SAS datasets, available for 2010 through 2017; describing the characteristics of the data; providing basic SAS program code which can be used to analyze the data; and displaying the resulting output.

(KEYWORDS: HOSPITAL MEDICARE COST REPORTS)

HP-260 ***Who, When, and Where – A Step-by-Step Approach to Creating a Color-Coded Tracker for Clinical Trial Subjects***
Venita DePuy

In a perfect world, all subjects participating in a clinical trial would attend each and every scheduled visit on the planned date. Realistically, the project management team may spend a significant amount of time trying to keep track of who is where, and when. Busy medical offices may not get patients scheduled at the right times, participants may be too ill to attend visits, or any number of other issues may arise that lead to missed or out-of-window visits during the course of the trial. As subject participation in the trial draws to a close, there is increased pressure to schedule monitoring visits and site close-out visits as soon as possible, to allow final data cleaning prior to a timely database lock and the final study tables, listings, and figures. We will present a step-by-step approach to creating a traffic-lighted subject tracker, intended for use in clinical trials (but applicable to other fields as well). Based on each subject's initial treatment date, all future study visits are populated in the spreadsheet, using cell color and font characteristics to indicate a predicted future date. Overdue visits (when the tracker is run and the predicted date is prior to the program run date) are then highlighted to indicate a potential late visit. Completed visits are populated with the actual visit date (with out-of-window indicators if necessary) and color-coded to reflect a past visit. Subjects who have completed the trial are automatically moved to a separate tab of the sheet, with all information retained. This program can be set to run automatically each morning, after a daily data refresh, to have a handy, visual reference for subject visits and associated data entry.

(KEYWORDS: CLINICAL TRIAL)

Venita is a biostatistician and SAS programmer who has worked with multiple clinical research organizations, pharmaceutical companies, and academic research organizations. She owns Bowden Analytics, based in Apex, NC.

She received both her Master of Statistics and her PhD from North Carolina State University. She is also a SESUG 2015 conference co-chair, and looks forward to seeing you all in Savannah next year.

HP-294 ***Don't Let Complex Survey Data Get the Best of You! SAS® Survey PROCs for Categorical Data Analysis***

Charlotte Baker

Data from US federal health surveys frequently use complex survey structures, rendering traditional procedures not useful for analysis. The SAS survey procedures exist but have not yet become a regularly used asset in analysis. Instead, users frequently choose to use other programs or add-ons for even the most basic of analyses. This paper demonstrates why the survey procedures such as SURVEYFREQ and SURVEYLOGISTIC should be in everyone's toolbox when using complex survey data in research or practice.

(KEYWORDS: FEDERAL HEALTH SURVEYS, PROC SURVEYFREQ, SURVEYLOGISTIC, AND SURVEYREG, 2017 BRFS DATA FROM CDC)

Dr. Charlotte Baker is an assistant professor of Epidemiology at Virginia Tech, Her research interests include sports injury, physical activity, health disparities, and community based participatory research. Dr. Baker enjoys learning and pushing the boundaries of SAS and translating that knowledge into learning experiences for graduate students.

HP-297 ***Introduction to Proc Report***

Barbara Okerson

SAS PROC REPORT is a procedure that combines the capabilities of the PRINT, MEANS, and TABULATE procedures to produce distribution ready reports. This presentation introduces Proc REPORT syntax and reporting features in the context of healthcare reporting, including creating Microsoft Excel output. All examples were created from real production reports used by a major American health insurer.

(KEYWORDS: PROC REPORT, HEALTH INSURANCE DATA)

Barbara B. Okerson, Ph.D. is an independent contractor, biostatistician and SAS programmer. She has worked as a statistician/programmer in the healthcare industry for over 15 years and has been a SAS user for over 25 years. She is a member of the executive council at SESUG and a past SESUG conference chair. She is a SAS Certified Professional, a Certified Professional in Healthcare Quality, and a Fellow, Academy for Healthcare Management.

Know Your SAS: Foundations

FD-116 **Getting the Right DATE (With SAS)**

Marje Fecht

Do you struggle with dates in your programs? Do you have issues getting dates into the right format for database queries, or for reports and dashboards? Do you manually provide dates as input to your processes? This presentation will help you find the right date, and then generalize the coding to avoid manual input, repetitive and messy coding, and frustration. Examples emphasize the easy manipulation of dates, and focus on generalization to support flexible coding, including: >> Dynamically identifying date ranges, such as reporting and analytics periods (current calendar year; most recent 6 months; past 90 days; current fiscal year; year over year) >> Dynamically generating field names that represent date values or ranges >> Controlling the appearance of date values in reports >> Generating date-time stamps for file names, without special symbols.

(KEYWORDS: SAS DATES, GENERALIZING SAS PROGRAMS)

Marje Fecht is a Senior Partner with Prowerk Consulting, and has enjoyed using and teaching SAS software since 1979. Her current consulting work focuses on delivering business insights and analytics to support actionable decisions. She is passionate about developing efficient, reusable methodologies so that businesses can focus on insight instead of processes.

Marje enjoys sharing her SAS and business knowledge and she enjoyed working with the SAS community during her role as SAS Global Forum 2014 Conference Chair.

FD-124 **PROC IMPORT and more. Or: when PROC IMPORT just doesn't do the job.**

David Horvath

PROC IMPORT comes in handy when quickly trying to load a CSV or similar file. But it does have limitations. Unfortunately, I've run into those limitations and had to work around them. This session will discuss the original CSV specification (early 1980's), how Microsoft Excel violates that specification, how SAS PROC IMPORT does not follow that specification, and the issues that can result. Simple UNIX tools will be described that can be used to ensure that data hilarities do not occur due to CSV issues. Recommendations will be made to get around some of PROC IMPORT limitations (like field naming, data type determination, limitation in number of fields, separator in data). CSV, TAB, and DLM types will be discussed.

(KEYWORDS: PROC IMPORT, DELIMITED FILES)

David is an IT Professional who has worked with SAS, off and on, since the late 1980's using it as a data processing (4GL/ETL) and analysis tool. He has presented at PhilaSUG previously and for other user groups and organizations (workshops and seminars) in Australia, France, the US, and Canada. His Masters is in Organizational Dynamics from UPENN, has consulted with CHERP at the VA hospital, and currently works for a regional bank in Risk Analytics Infrastructure at the Wilmington DE location. He has several books to his credit (none SAS-related) and is an Adjunct Instructor covering IT topics.

FD-128

NOBS for Noobs

David Horvath

This mini-session will be a short discussion of the NOBS (number of observations) option on the SET statement. This includes one "gotcha" that I've run into with where clauses: NOBS is set before WHERE processing. If you have a reason to know the number of observations after the WHERE clause, another DATA step is needed.

(KEYWORDS: SET, NOBS=, OBSERVATIONS)

David is an IT Professional who has worked with SAS, off and on, since the late 1980's using it as a data processing (4GL/ETL) and analysis tool. He has presented at PhilaSUG previously and for other user groups and organizations (workshops and seminars) in Australia, France, the US, and Canada. His Masters is in Organizational Dynamics from UPENN, has consulted with CHERP at the VA hospital, and currently works for a regional bank in Risk Analytics Infrastructure at the Wilmington DE location. He has several books to his credit (none SAS-related) and is an Adjunct Instructor covering IT topics.

FD-131

Twenty ways to run your SAS program faster and use less space

Stephen Sloan

When running SAS® programs that use large amounts of data or have complicated algorithms we often are frustrated by the amount of time it takes for the programs to run and by the large amount of space required for the program to run to completion. Even experienced SAS programmers sometimes run into this situation, perhaps through the need to produce results quickly, through a change in the data source, through inheriting someone else's programs, or for some other reason. This paper outlines twenty techniques that can reduce the time and space required for a program without requiring an extended period of time for the modifications.

(KEYWORDS: SAS, PROGRAMMING TECHNIQUES, EFFICIENCY, PROGRAMMING SPEED, BASIC KNOWLEDGE)

Stephen Sloan has worked at Accenture in the Services, Consulting, and Digital groups and is currently a Data Science Senior Principal in the SAS Analytics area. He has worked in a variety of functional areas in Project Management, Data Management, and Statistical Analysis. He has had the good fortune to have worked with many talented people at SAS Institute. Stephen has presented at 20 SAS Global Forums and SAS Users Group events and has been published in professional journals.

Stephen has a B.A. cum laude with Honor in Mathematics from Brandeis University, M.S. degrees in Mathematics and Computer Science from Northern Illinois University, and an MBA (1st in class) from Stern Business School at New York University. He also has a graduate certificate in Financial Analytics from Stevens Institute.

Running Parts of a SAS Program while Preserving the Entire Program**Stephen Sloan**

The Challenge: We have long programs that accomplish a number of different objectives. We often only want to run parts of the programs while preserving the entire programs for documentation or future use. Some of the reasons for selectively running parts of a program are:

- Part of it has run already and the program timed out or encountered an unexpected error. It takes a long time to run so we don't want to re-run the parts that ran successfully.
- We don't want to recreate data sets that were already created. This can take a considerable amount of time and resources, and can also occupy additional space while the data sets are being created.
- We only need some of the results from the program currently, but we want to preserve the entire program.
- We want to test new scenarios that only require subsets of the program.

(KEYWORDS: SAS, TIPS AND TECHNIQUES, EFFICIENCY, MODULAR PROGRAMMING)

Stephen Sloan has worked at Accenture in the Services, Consulting, and Digital groups and is currently a Data Science Senior Principal in the SAS Analytics area. He has worked in a variety of functional areas in Project Management, Data Management, and Statistical Analysis. He has had the good fortune to have worked with many talented people at SAS Institute. Stephen has presented at 20 SAS Global Forums and SAS Users Group events and has been published in professional journals.

Stephen has a B.A. cum laude with Honor in Mathematics from Brandeis University, M.S. degrees in Mathematics and Computer Science from Northern Illinois University, and an MBA (1st in class) from Stern Business School at New York University. He also has a graduate certificate in Financial Analytics from Stevens Institute.

The Battle of the Titans (Part II): PROC TABULATE versus PROC REPORT**Kirk Paul Lafler, Ben Cochran and Josh Horstman**

Should I use PROC REPORT or PROC TABULATE to produce that report? Which one will give me the control and flexibility to produce the report exactly the way I want it to look? Which one is easier to use? Which one is more powerful? WHICH ONE IS BETTER? If you have these and other questions about the pros and cons of the REPORT and TABULATE procedures, this presentation is for you. We will discuss, using real-life report scenarios, the strengths (and even a few weaknesses) of the two most powerful reporting procedures in SAS® (as we see it). We will provide you with the wisdom you need to make that sometimes difficult decision about which procedure to use to get the report you really want and need.

(KEYWORDS: DETAIL OUTPUT, SUMMARY OUTPUT, TEMPLATE STYLES)

Kirk Paul Lafler is entrepreneur, consultant and founder of Software Intelligence Corporation, and has worked with SAS software since 1979. As a SAS consultant, application developer, programmer, data analyst, mentor, infrastructure specialist, educator and author at Software Intelligence Corporation, and an advisor and SAS programming adjunct professor at the University of California San Diego Extension, Kirk has taught SAS courses, seminars, webinars and workshops to thousands of users around the world. Kirk has also authored or co-authored several books including PROC SQL: Beyond the Basics Using SAS, Third Edition (SAS Press, 2019) and Google® Search Complete (Odyssey Press, 2014); hundreds of papers and articles on a variety of SAS topics; served as an Invited speaker, educator, keynote and section leader at SAS user group conferences and meetings worldwide; and is the recipient of 25 "Best" contributed paper, hands-on workshop (HOW), and poster awards.

Josh Horstman is an independent statistical programming consultant and trainer based in Indianapolis with over 20 years' experience using SAS in the life sciences industry. He specializes in analyzing clinical trial data, and his clients have included major pharmaceutical corporations, biotech companies, and research organizations. Josh is a SAS Certified Advanced Programmer who loves coding and is a frequent presenter at SAS Global Forum and other SAS User Group meetings.

Old But Not Obsolete: Undocumented Procedures for SAS® University Edition**Barbara Okerson**

Proc Spell? Proc Neighbor? Proc Browse? These and other SAS procedures have disappeared from the manuals, mainly because their functionality was picked up by newer, more robust, procedures and features of the SAS system. These procedures have remained available in standard SAS installations. But did you know they are also available for the University Edition? Since sometimes, simpler can be better, this paper will use the SAS University Edition to look at these and other available SAS procedures not in the current documentation and address situations where they can be useful today.

(KEYWORDS: SAS UNIVERSITY EDITION)

Barbara B. Okerson, Ph.D. is an independent contractor, biostatistician and SAS programmer. She has worked as a statistician/programmer in the healthcare industry for over 15 years and has been a SAS user for over 25 years. She is a member of the executive council at SESUG and a past SESUG conference chair. She is a SAS Certified Professional, a Certified Professional in Healthcare Quality, and a Fellow, Academy for Healthcare Management.

FD-167 ***Like, Learn to Love SAS® Like***

Louise Hadden

How do I like SAS®? Let me count the ways.... There are numerous instances where LIKE or LIKE operators can be used in SAS - and all of them are useful. This paper will walk through such uses of LIKE as: using the LIKE condition to perform pattern-matching; searches and joins with that smooth LIKE operator (and the NOT LIKE operator); the SOUNDS LIKE operator; and PROC SQL CREATE TABLE LIKE.

(KEYWORDS: LIKE, SAS FUNCTIONS, SAS STATEMENTS, PROC SQL)

Louise Hadden presented at her first SAS conference in 1996 and has never looked back, presenting at multiple conferences across the continent over the years. She supports analytic processing at Abt Associates Inc., a social science research company, and specializes in reporting and data visualization in the division of health and environment.

FD-204 ***Fifteen Functions to Supercharge Your SAS® Code***

Josh Horstman

The number of functions included in SAS® software has exploded in recent versions, but many of the most amazing and useful functions remain relatively unknown. This paper will discuss such functions and provide examples of their use. Both new and experienced SAS programmers should find something new to add to their toolboxes.

(KEYWORDS: SAS PROGRAMMING, FUNCTIONS, DATA STEP)

Josh Horstman is an independent statistical programming consultant and trainer based in Indianapolis with over 20 years' experience using SAS in the life sciences industry. He specializes in analyzing clinical trial data, and his clients have included major pharmaceutical corporations, biotech companies, and research organizations. Josh is a SAS Certified Advanced Programmer who loves coding and is a frequent presenter at SAS Global Forum and other SAS User Group meetings.

FD-219 ***PROC SQL for PROC SUMMARY Stalwarts***

Christianna Williams

One of the endlessly fascinating features of SAS is that the software often provides multiple ways to accomplish the same task. A perfect example of this is the aggregation and summarization of data across multiple rows “BY groups” of interest. These groupings can be study participants, time periods, geographical areas, or really just about any type of discrete classification that one desires. While many SAS programmers may be accustomed to accomplishing these aggregation tasks with PROC SUMMARY (or equivalently, PROC MEANS), PROC SQL can also do a bang-up job of aggregation – often with less code and fewer steps. The purpose of this step-by-step paper is to explain how to use PROC SQL for a variety of summarization and aggregation tasks, and will use a series of concrete, task-oriented examples to do so. For each example, both the PROC SUMMARY method and the PROC SQL method will be presented, along with discussion of pros and cons of each approach. Thus, the reader familiar with either technique can learn a new strategy that may have benefits in certain circumstances. The presentation style will be similar to that used in the author’s previous paper, “PROC SQL for DATA Step Die-Hards”.

(KEYWORDS: PROC SQL; PROC MEANS; PROC SUMMARY; DATA AGGREGATION)

Christianna Williams, PhD is an independent consultant based in Chapel Hill, North Carolina, focusing on study design and statistical analyses and reporting in epidemiology and health services research. Christianna started using SAS as a graduate student in population biology in the days when there was a single SAS manual, and she is still learning! She has been a frequent presenter at local and regional user group conferences as well as SAS Global Forum for more than 15 years. She also devotes as much time as possible to her other passions: running, vegetarian cooking and reading novels.

FD-220 ***SAS Formats Top 10***

Christianna Williams

SAS® FORMATS can be used in so many different ways! Even the most basic FORMAT use of modifying the way a SAS data value is displayed (without changing the underlying data value) holds a variety of nifty tricks, such as nesting formats, formats that affect various style attributes (such as alignment, font etc.), and conditional formatting. Add in PICTURE formats, multi-label FORMATS, using FORMATS for data cleaning, and FORMATS for joins and table look-ups, and we have quite a bag of tricks for the humble SAS FORMAT and the FORMAT procedure used to generate them. The purpose of this paper is to describe a few handfuls of very useful programming techniques that employ SAS FORMATS. While this paper will be appropriate for the newest SAS user, it will also focus on some of the lesser-known features of FORMATS and PROC FORMAT and so should be useful for even quite experienced users of SAS.

(KEYWORDS: FORMAT; UNICODE, NESTING FORMATS, PICTURE FORMATS, FORMAT LIBRARY)

Christianna Williams, PhD is an independent consultant based in Chapel Hill, North Carolina, focusing on study design and statistical analyses and reporting in epidemiology and health services research. Christianna started using SAS as a graduate student in population biology in the days when there was a single SAS manual, and she is still learning! She has been a frequent presenter at local and regional user group conferences as well as SAS Global Forum for more than 15 years. She also devotes as much time as possible to her other passions: running, vegetarian cooking and reading novels.

DO loops :INTO efficient programming: DATA STEP do loop + PROC SQL in boilerplate programming**Christopher Hargett**

Boilerplate SAS programs are real time savers, but frequently input datasets have different variables and values not accounted for in the boilerplate. Going through a whole program to manually update RENAME or SUM statements is laborious. This paper will introduce a technique to dynamically modify boilerplate programs based on input dataset. Combining a DO loop in a DATA STEP and the :INTO in PROC SQL results in a flexible method of creating macro variables that can be used in a wide array of applications. Using the COUNT function with a DO loop in the DATA STEP to create a SAS data set consisting of programming language, and the INTO clause in PROC SQL allows the programming language to be dynamically generated and applied efficiently in multiple scenarios in a longer program. This paper builds on concepts covered by Kelly Schlessman, "PROC SQL – GET "INTO:" IT!," SESUG Paper 267-2018.

(KEYWORDS: DO LOOP, DATA STEP, PROC SQL, :INTO, COUNT, MACRO VARIABLES)

Christopher Hargett is a SAS Support Specialist for the U.S. Department of Commerce, Washington, D.C. He started using SAS in 2004 as an International Trade Analyst and became a support specialist in 2013. His team writes and tailors boilerplate SAS programs for use by 120 analysts calculating Antidumping Duties for imports into the U.S. from around the world. The biggest challenges in his daily work life are translating legal prose into SAS code and explaining it to managers, and getting users to read instructions.

20 in 20: Quick Tips for SAS Enterprise Guide Users**Kelly Gray**

There are many time-saving and headache-saving tips and tricks you can use to make working in SAS® Enterprise Guide® a breeze. Did you know that you can change your layout so that you can see your code and your results at the same time? You will learn 20 tips and tricks for working in SAS Enterprise Guide in 20 minutes. One tip per minute, and out of the twenty you are guaranteed to find at least one nugget that will make your life easier.

(KEYWORDS: SAS ENTERPRISE GUIDE, TIPS AND TRICKS, TIME-SAVING)

Kelly Gray is a 20 year SAS user who loves manipulating data. She has a degree in Journalism from the University of North Carolina at Chapel Hill, and then went on to get her Certificate in Computer Programming from North Carolina State University. She has worked in Technical Support, Education, Sales, and Marketing and loves sharing her knowledge of programming and reporting using SAS.

Data Step versus Everybody: Approaching Problems as a Beginning Coder**Brian Varney**

SAS has had the Data Step and Base SAS procedures since its inception. PROC SQL was added in the late 1980s and introduced an additional powerful tool for problem solving. If you are relatively new to SAS, it can be confusing to choose an approach. This paper intends to guide a beginning SAS programmer on what to use depending on the ETL and/or macro variable creation needed for your programming problem.

(KEYWORDS: BASE SAS DATA STEP, PROC SQL, EFFICIENT PROGRAMMING, PROBLEM SOLVING)

Brian Varney has been a SAS consultant, trainer, and senior technical manager for over 28 years with the Experis Business Analytics Group, a SAS Consulting Alliance Partner. Located in Kalamazoo, Michigan, he keeps busy with providing SAS training, consulting support and business development (plus a few hours a week for soccer).

Know Your SAS: Advanced Techniques

AT-110 *The Knight's Tour in 3-Dimensional Chess*

John R Gerlach and Scott M Gerlach

Three dimensional chess typically uses three chess boards such that a chess piece can traverse the several boards according to the rules for that piece. For example, the knight can remain on the board where it resides or move to another successive board, then move in a perpendicular fashion. In three-dimensional chess, the Knight's Tour is a sequence of moves on multiple 8x8 chess boards such that the knight visits each square only once. Thus, for three boards, there would be 192 squares visited only once. The paper, *The Knight's Tour in Chess – Implementing a Heuristic Solution* (Gerlach 2015), explains a SAS® solution for finding such tours on a single chess board, starting from any square. This paper discusses several scenarios and SAS solutions for generating the Knight's Tour using multiple chess boards.

(KEYWORDS: CHESS, KNIGHT'S TOUR, 3-D CHESS, HEURISTIC ALGORITHM)

John has been using the SAS System for over 30 years specializing in the health and finances industries. John has written 50 papers including two award-winning papers. John has a BA in Italian Literature and studies classical piano.

AT-111 *Tales from the Help Desk: Solutions to Common Macro and Macro Variable Issues*

Bruce Gilson

In 35 years as a SAS® consultant at the Federal Reserve Board, I have seen some issues related to common SAS tasks surface again and again. This paper collects the most common simple issues related to macros and macro variables from my previous "Tales from the Help Desk" papers, and provides code to explain and resolve them. The following issues are reviewed:

1. Using an array definition in multiple DATA steps.
2. Using a DATALINES statement in a macro.
3. Surrounding a macro variable with single quotes.
4. Using comments in a macro.
5. Creating a macro variable and using it in the same DATA step.
6. Having an overlapping macro variable in a main macro and a called macro.
7. Using %SYSFUNC to execute DATA step functions in a macro.
8. Using the macro IN operator to check in a macro if a value equals one of the values in a list.

In the context of discussing these issues, the paper provides details about SAS processing that can help users employ SAS more effectively. See the references for seven previous papers that contain additional common issues.

(KEYWORDS: MACROS, MACRO VARIABLES, ARRAYS, DATALINES STATEMENT, MACRO COMMENTS, %SYSFUNC, MACRO IN OPERATOR)

Bruce Gilson has worked at the Federal Reserve Board for over 36 years. He spends most of his time as an in-house SAS consultant. His Bruce Force fantasy baseball team battled for its 6th league title in 31 years in 2019.

AT-122 ***UNIX X Command Tips and Tricks***

David Horvath

SAS provides the ability to execute operating system level commands from within your SAS code ? generically known as the ?X Command?. This session explores the various commands, the advantages and disadvantages of each, and their alternatives. The focus is on UNIX/Linux but much of the same applies to Windows as well. Under SAS EG, any issued commands execute on the SAS engine, not necessarily on the PC. X %sysexec Call system Systask command Filename pipe &SYSRC Waitfor Alternatives will also be addressed ? how to handle when NOXCMD is the default for your installation, saving results, and error checking.

(KEYWORDS: X, SYSTASK, %SYSEXEC, &SYSRC, FILENAME PIPE, WAITFOR)

David is an IT Professional who has worked with SAS, off and on, since the late 1980's using it as a data processing (4GL/ETL) and analysis tool. He has presented at PhilaSUG previously and for other user groups and organizations (workshops and seminars) in Australia, France, the US, and Canada. His Masters is in Organizational Dynamics from UPENN, has consulted with CHERP at the VA hospital, and currently works for a regional bank in Risk Analytics Infrastructure at the Wilmington DE location. He has several books to his credit (none SAS-related) and is an Adjunct Instructor covering IT topics.

AT-149 ***Quick, Call the "FUZZ": Using Fuzzy Logic***

Richann Watson and Louise Hadden

SAS® practitioners are frequently called upon to do a comparison of data between two different data sets and find that the values in synonymous fields do not line up exactly. A second quandary occurs when there is one data source to search for particular values, but those values are contained in character fields in which the values can be represented in myriad different ways. This paper discusses robust, if not warm and fuzzy, techniques for comparing data between and selecting data in SAS data sets in not so ideal conditions.

(KEYWORDS: FUZZY MATCH, COMPFUZZ, FLOOR, FUZZ, COMPARE, COMPGED, INDEXW, FIND, PRXMATCH, PROC FCMP)

Richann Watson is an independent statistical programmer and CDISC consultant based in Ohio. She has been using SAS since 1996 with most of her experience being in the life sciences industry. She specializes in analyzing clinical trial data and implementing CDISC standards. Additionally, she is a member of the CDISC ADaM team and various sub-teams.

Louise Hadden presented at her first SAS conference in 1996 and has never looked back, presenting at multiple conferences across the continent over the years. She supports analytic processing at Abt Associates Inc., a social science research company, and specializes in reporting and data visualization in the division of health and environment.

Hexadecimal Encoding Can Mitigate Some SAS® Macro Quoting Issues**Thomas Billings**

Passing special characters in a SAS® macro variable can be challenging; see Rosenbloom and Carpenter (2013). As messy as the situation is for fixed (constant) strings, the possibility that a macro variable is dynamically loaded from an input data source is even more challenging, as the values that come in via inputs may throw an error in production. Here we describe methods that mitigate some of the macro quoting issues surrounding special characters. We begin by clarifying what it means for a macro variable to contain nearly anything, i.e., the issue of trailing blanks and how they are handled by select input methods. Then we give a simple method (no encoding required) that works if the macro variable is used in the current session. Next we show how hex encoding can mitigate macro quoting issues when the objective is to use the target macro variable string(s) in a compiled macro to be used in other, downstream sessions. We end by discussing the constraints of this method.

(KEYWORDS: MACRO, QUOTING, MASKING, SPECIAL CHARACTERS, HEXADECIMAL, LUA, CALL EXECUTE, QUOTE FUNCTION)

Tom has used the SAS system intermittently since the mid-1970's, in a variety of different industries and application areas. He is currently working in Banking, using SAS Enterprise Guide and other tools to build and analyze databases.

Vetting Differences Between Relational Database Definitions and Actual Data with SAS®

Michael Raithe

SAS programmers are sometimes tasked with loading SAS data sets into relational databases management systems (RDBMS) such as Oracle, SQL Server, or SYBASE. Loading the data can be a painful, stop-and-go process when the relational database table definitions do not agree with the SAS data sets that are to be uploaded. Differences between SAS and the RDBMS in the number of variables, variable names, variable types, and variable lengths can cause issues with the loading or with subsequent processing. Therefore, it is prudent to vet the differences between the SAS data sets' metadata and the RDBMS tables' metadata before attempting to load SAS data sets into a database. This paper presents a method of comparing the metadata of SAS data sets against an RDBMS's table definitions. It produces four reports: variables found in SAS but not in the RDBMS; variables found in the RDBMS but not SAS; variable type differences; and variable length differences. Working with these reports and with the RDBMS's database administrator, SAS programmers can help shape the SAS data sets and table definitions that lead to an error-free loading of their relational databases.

(KEYWORDS: RELATIONAL DATABASE, RDBMS, ORACLE, SQL SERVER, SYBASE, TABLE DEFINITION, METADATA)

Michael A. Raithe is a Senior Systems Analyst for Westat, a leading research firm headquartered in Rockville, Maryland. He has been using SAS software for over 24 years and is an internationally recognized expert on using SAS in mainframe and UNIX environments. Michael has authored over twenty-five SAS technical papers and is a popular lecturer at SAS Global Forum, regional SAS conferences, and at local SAS User Group meetings. He has taught SAS classes at Westat, at American University, at SAS Global Forum meetings, and at SAS headquarters in Marlow, England.

*Michael has been a section chair at SUGI, SESUG, and NESUG, and he co-chaired NESUG in 1995. He is a frequent poster to the SAS-L listserv and a contributor to SASCommunity.org. Michael has written five books on SAS software; the most current one being *Did You Know That? Essential Hacks for Clever SAS Programmers*. A copy of his first SAS book, *Tuning SAS Applications in the MVS Environment*, resides in the Smithsonian Institution of American History's Permanent Research Collection of Information Technology.*

The Power of PROC SQL's SELECT DISTINCT INTO

HoaiNam Tran and Mai Anh Ngo

Many SAS programmers are familiar with PROC SQL procedures that join database tables or SAS datasets. However, they may not be familiar with using PROC SQL combined with SAS data steps and procedures to improve programming efficiency. Specifically, using the DISTINCT and SEPARATED BY arguments and SELECT and INTO clauses in PROC SQL as well as the macro SCAN function, DO loop, and array in SAS data steps and PROC REPORT procedure can provide programming flexibility, simplify SAS code, and minimize typographical errors. In this paper, we provide examples that illustrate how to use these specific PROC SQL arguments and clauses and SAS procedures to create automated data-driven programs that can be readily tailored to different applications.

(KEYWORDS: SELECT DISTINCT, SELECT INTO, MACRO SCAN, ARRAY, DO LOOP, PROC SQL, DATA DRIVEN-PROGRAM)

HoaiNam Tran is a Mathematical Statistician at the National Agriculture Statistics Service where she has been since 2005. She has been using SAS for about 10 years.

Integrating SAS IntraNet, SAS Macro facility, JavaScript, HTML, and .NET to build a dynamic web application to present NSSE data**Carlos Martinez, Wen Jiang and Uday Nair**

This presentation is intended for staff in the institutional research, survey research, or assessment office in the education industry who conduct and report survey data. The NSSE institute provides summary and item-level reports comparing a select institution to various other groups of institutions. However, internal university analysis by student characteristics is more insightful and actionable.

A dataset was created that combines NSSE data with the typical institutional research type data. The SAS ODS HTML destination within the SAS/IntraNet ® platform, HTML, and JavaScript were combined to create a dynamic web application consisting of cascading input query pages, custom reporting and easy-to-navigate reporting interface. PROC TABULATE was heavily used to build standard aesthetically-pleasing table templates. The SAS MACRO language was leveraged to improve efficiency of data processing, reporting, and back-end code maintenance. The web application was nestled inside a .NET controlled framework to ensure data access to select users.

This paper does not serve as instructions on how to build such a complex web application from the ground up. Instead, the goal is to highlight some of the more challenging tasks of integration and provide explanation as to how these were achieved. Knowledge of the programming languages SAS, JavaScript, HTML, and VB.NET or C# will be helpful in understanding the concepts covered in this paper. The presentation will include typical code examples. This paper is organized by certain features in the web application that were achieved through integration of several platforms and languages:

- Creating uniform layout and page sizing [cross-domain integration]
- Dynamically generated cascading dropdown menus
- Printing capabilities from the web application [cross-domain integration]
- Tracking web application user behavior [using Google Analytics]

(KEYWORDS: INTEGRATION, JAVASCRIPT, SAS INTRNET, HTML, NSSE)

Carlos Martinez has over twenty years of experience in application development in various programming languages, server administration, networking, and various other aspects of information technology. Currently he is employed as an Applications Programmer III at the University of Central Florida. Previously Carlos has been employed at Nortel Networks, Carlson Technology Solutions, MicroAge, and Digital Equipment Corporation. He has a Bachelor of Science degree in Information Systems Technology from UCF. He loves golf and UCF football. Go Knights, Charge on!!! Wen Jiang has a M.S. in Statistics and has been a SAS user for about 3 years. Uday Nair is working towards his Education Ph.D.. He has a M.S. in Industrial Engineering and an MBA. He has been a SAS user for 12+ years.

AT-210 ***Dating for SAS® Programmers***

Josh Horstman

Every SAS programmer needs to know how to get a date... no, not that kind of date. This paper will cover the fundamentals of working with SAS date values, time values, and date/time values. Topics will include constructing date and time values from their individual pieces, extracting their constituent elements, and converting between various types of dates. We'll also explore the extensive library of built-in SAS functions, formats, and informats for working with dates and times using in-depth examples. Finally, you'll learn how to answer that age-old question... when is Easter next year?

(KEYWORDS: SAS PROGRAMMING, FUNCTIONS, DATES AND TIMES, DATA STEP)

Josh Horstman is an independent statistical programming consultant and trainer based in Indianapolis with over 20 years' experience using SAS in the life sciences industry. He specializes in analyzing clinical trial data, and his clients have included major pharmaceutical corporations, biotech companies, and research organizations. Josh is a SAS Certified Advanced Programmer who loves coding and is a frequent presenter at SAS Global Forum and other SAS User Group meetings.

AT-214 ***When fuzzy matching doesn't work: using the CONTAINS and JOIN functions through SAS EG to find foreign words in text strings***

Arthur Laciak

I have found that fuzzy matching functions, such as SOUNDEX and SPEDIS, are not effective in matching foreign words, especially Chinese words. When transliterated into the Latin alphabet, too many syllables are similarly spelled that fuzzy matching is ineffective, even with strict parameters. Many Chinese words are unintentionally matched, resulting in more complications. This paper explores a work-around approach, using the CONTAINS and JOIN functions in SAS EG through PROC SQL to search for keywords from a reference table in text strings of a given data set. In this procedure, each text string is compared against keywords and matched keywords are displayed in a new column, resulting in clean data for further analysis. The example given in this paper matches Chinese addresses written in a text string to Chinese cities and provinces.

(KEYWORDS: DATA CLEANUP, FUZZY MATCHING, TEXT STRINGS, SAS EG, PROC SQL, CONTAINS, JOIN, FOREIGN WORDS, CHINESE)

Arthur Laciak is an Operations Research Analyst at the U.S. Consumer Product Safety Commission, an independent federal agency tasked with protecting the public from unreasonable risks of injury or death associated with the use of the thousands of types of consumer products. In his role in the Office of Imports Surveillance, Arthur primarily uses SAS Enterprise Guide and JMP to perform data quality of, visualize, and risk-assess importation data.

Using PROC FCMP to Implement Anomaly Detection Method**Tsung-hsun Tsai and Yung-chen Hsu**

The essential purpose of PROC FCMP, the SAS® Function Compiler procedure, is to allow creating user defined functions or CALL routines for DATA steps or SAS procedures. Those functions can also be shared with other SAS programmers to facilitate code reuse and increase productivity. For advanced users, it can serve as a wrapper function for PROC PROTO to enable incorporating C language structures, types, and functions into SAS®. In addition, for users who do not have access to SAS/IML®, PROC FCMP also provides CALL routines for performing typical matrix operations, such as multiplication, transpose, and inverse. In this paper, we use a simple anomaly detection method for screening field test items as an example to illustrate the use of PROC FCMP and its capability of performing matrix operations. In testing service organizations, evaluating field test items in the posterior human review is an elaborate inspection process during the test development. This work demonstrates a method to automatically flag potential flawed items by using PROC FCMP to conduct matrix and low-level array operations.

(KEYWORDS: PROC FCMP, ANOMALY DETECTION, MATRIX, ARRAY)

Yung-chen Hsu is a Senior Psychometrician at GED Testing Service since 2006 and has been using SAS for about 15 years.

Which report is appropriate? Let SAS figure it out!**Kelly Smith**

Central Piedmont Community College uses a series of progression and completion metrics to track multiple cohorts of new part-time and new full-time students through their first three years at the institution. Progression and completion metrics from matching cohorts in sequential years are also compared to determine if interventions are having a positive impact on student success. As cohorts progress through the three year cycle, data from new metrics becomes available and a fuller picture of the cohort emerges. The frequency and complexity of the reporting process encouraged us to search for ways to minimize errors that can occur through the updating of numerous macrovariables and to find a better method to determine which report was appropriate for each cohort at specific points in time. To minimize data entry errors, a series of macro variables have been defined using macro functions so that only an initial macro variable is updated by the user. To maximize efficiency and ease of use, macro functions and conditional processing within a macro identify which metrics and related output are appropriate based on a comparison of the current date to the starting point as defined by the user in the initial macro variable. The discussion includes macro and system functions such as %SYSFUNC, %EVAL, and %SUBSTR, in addition to the conditional processing functions %IF, %THEN DO, %ELSE %IF, and %END.

(KEYWORDS: MACRO FUNCTIONS, CONDITIONAL PROCESSING)

Kelly D. Smith has worked with SAS for over three years, using Base SAS and Enterprise Guide on a daily basis as senior institutional research analyst with Central Piedmont Community College. She is a firm believer in using SAS to improve the accuracy and efficiency of the reporting process, and hopes this presentation will provide timely and useful information to other SAS coders.

RegExing in SAS for Pattern Matching and Replacement**Pratap Kunwar**

SAS has numerous character functions which are very useful in manipulating character fields, but only knowing Perl Regular Expressions (RegEx) will help anyone to solve complex pattern matching and search-and-replace operations. Moreover, this skill can be easily portable to other popular languages such as Perl, Python, JavaScript, PHP and more. This presentation will cover basic of character classes and metacharacters and then use them to build regular expressions in simple examples, ranging from finding simple literals to finding complex string patterns and replacing them to show regular expressions are powerful, convenient and so much anyone can do, which makes it worth learning.

(KEYWORDS: REGEX, PERL REGULAR EXPRESSIONS, PATTERN MATCHING AND REPLACEMENT)

Pratap Kunwar is a Senior Principal SAS Programmer at The EMMES Company, LLC where he has been since 2012. Pratap uses SAS® in his daily responsibilities and has used SAS for more than 10 years.

User-Defined Multithreading with the SAS® DS2 Procedure: Performance Testing DS2 Against Functionally Equivalent DATA Steps

Troy Hughes

The Data Step 2 (DS2) procedure affords the first opportunity for developers to build custom, multithreaded processes in Base SAS®. Multithreaded processing debuted in SAS 9, when built-in procedures such as SORT, SQL, and MEANS were threaded to reduce runtime. Despite this advancement, and in contrast with languages such as Java and Python, SAS 9 still did not provide developers the ability to create custom, multithreaded processes. This limitation was overcome in SAS 9.4 with the introduction of the DS2 procedure—a threaded, object-oriented version of the DATA step. However, because DS2 relies on methods and packages (neither of which have been previously available in Base SAS), both DS2 instruction and literature has predominantly fixated on these object-oriented programming (OOP) aspects of the language rather than DS2 multithreading. To complicate the adoption of DS2 multithreading, one of the most ubiquitous examples of “multithreading” promulgated throughout SAS documentation and literature unfortunately fails to show any performance advantages in using DS2 over previous single-threaded methods—so many would-be DS2 developers may have cautiously dipped their toes in the multithreaded waters and, horrified with their own performance testing results from these published examples, quickly retreated back to the safer DATA step land. This text explores DS2 multithreading and demonstrates performance testing between DS2 procedures and functionally equivalent DATA steps and SAS procedures.

(KEYWORDS: DS2, MULTITHREADING, PARALLEL PROCESSING, DISTRIBUTED PROCESSING, MULTITHREADED, PERFORMANCE TESTING)

Troy has been a SAS practitioner for more than 20 years, has managed SAS projects in support of federal, state, and local government initiatives, and is a SAS Certified Base, Advanced, and Clinical Trials Programmer. Since 2013, he has given more than 80 presentations, hands-on workshops, and trainings at SAS Global Forum, SAS Analytics Experience, WUSS, MWSUG, SCSUG, SESUG, and PharmaSUG, and has authored two groundbreaking books: - SAS Data Analytic Development: Dimensions of Software Quality (2016) - SAS Data-Driven Development: From Abstract Design to Dynamic Functionality (2018) Troy has an MBA in Information Systems Management and additional certifications, including: Project Management Professional (PMP), Risk Management Professional (PMI-RMP), Professional in Business Analysis (PMI-PBA), Agile Certified Professional (PMI-ACP), Certified Information Systems Security Professional (CISSP), Certified Secure Software Lifecycle Professional (CSSLP), ITIL Foundation, Certified ScrumMaster (CSM), Certified Scrum Developer (CSD), Certified Scrum Product Owner (CSPO), and Certified Scrum Professional (CSP). He is a US Navy veteran with two tours of duty in Afghanistan.

AT-273 ***Introduction to Machine Learning In SAS***

Terry Woodfield

This presentation answers the questions of What is Machine Learning? And What does SAS offer for Machine Learning? Examples of specific machine learning techniques such as Random Forest, Gradient Boosting, Support Vector Machines, Neural Networks and K-means are covered.

(KEYWORDS: MACHINE LEARNING, RANDOM FOREST, GRADIENT BOOSTING, SVM, NEURAL NETWORKS)

A well-traveled data scientist with over 40 years experience using SAS software, Terry Woodfield has worked in the SAS Education Division for almost 20 years. He came to SAS in 1999 from the analytical consulting world, where he developed predictive modeling systems to streamline the process of data preparation, feature engineering, and predictive modeling in application areas like fraud and abuse detection, risk modeling, medical cost containment, physician profiling, and insurance claim-level loss reserving. Terry develops courses for machine learning and analytics applied to prediction problems. He teaches most of the courses in the SAS Enterprise Miner and SAS Viya curriculum related to prediction. He also teaches courses on statistical forecasting using SAS/ETS, SAS Forecast Server, and SAS Visual Forecasting. Terry also helped develop the current training materials for SAS Text Miner and SAS Visual Text Analytics. He holds a Ph.D. degree in Statistics from Texas A&M University.

AT-279 ***Discovering the Power of SAS Metadata: An Introduction to Dictionary Tables and Views***

Frank DiIorio

All SAS programs, regardless of size or complexity, create and populate dozens of metadata files, commonly known as Dictionary Tables. These tables are filled with information that is often difficult, and sometimes impossible, to obtain through other means. Any programmer who develops even simple general-purpose programs should be familiar with the tables' organization, content, and potential uses. This paper describes dictionary tables and their associated SASHELP library views. It:

- presents scenarios that show how they can be used
- gives high-level descriptions of some of the more important (a relative term, to be sure) tables
- identifies features of SQL and the macro language that are commonly used when writing programs that effectively use the tables
- shows examples of the tables' use, emphasizing the use of SQL and the macro language interface

The reader should come away from the discussion with an understanding of the tables as well as with a checklist of SQL skills that are required to use the tables most effectively.

(KEYWORDS: SAS METADATA, DICTIONARY TABLES, VIEWS)

A SAS programmer since 1975, Frank DiIorio is President of CodeCrafters, Inc. and the author of "SAS Applications Programming: A Gentle Introduction" and "Quick Start to Data Analysis with SAS." A frequent presenter at local and regional SAS user groups, he is past President of the SouthEast SAS Users Group, and co-chaired its 1994 and 1996 conferences. He is also active in several local SAS user groups and was a co-founder of the Research Triangle CDISC Users Group.

Use Callable VBS and VBA Code Files to Extend the Power of SAS® to Format Microsoft Excel Worksheets**William Benjamin**

Did you ever wish you could use the power of SAS® to take control of Microsoft Excel and make Excel do what you wanted when you wanted? Well, one letter is the key to doing just that, the letter "X", as in the SAS "X" command that opens the door to all operating system commands from SAS. The Microsoft Windows operating system comes with a facility to write a series of commands called scripts. These scripts have the ability to open and reach into the internals of Excel. Scripts can load, execute, and remove Visual Basic for Applications (VBA) macro code and control Excel. This level of control enables you to make Excel do what you want, without leaving any traces of a macro behind. This is Power.

(KEYWORDS: CALLABLE VBS, VBA SCRIPTS, FORMAT EXCEL WORKSHEETS WITH SAS)

William's expertise includes Base SAS® Software, and SAS Macros. He has a BSCS degree from Arizona State University and an MBA from Western International University. He has used SAS software since 1983 and programmed computers since 1973. His experience spans from vacuum tube mainframes, to current PC computers. His consulting company is called OWL Computer Consultancy, LLC. His first SAS Press book "Exchanging Data between SAS and Microsoft Excel: Tips and Techniques to Transfer and Manage Data More Efficiently" was released in April 2015. and his new book "Exchanging Data From SAS to Excel: The ODS Excel Destination" was published Aug 30, 2017.

Open Analytics

OA-104 ***%SUBMIT R: A SAS(R) Macro to Interface SAS and R***

Ross Bettinger

The purpose of the %SUBMIT R macro is to facilitate communication between SAS® and R under Windows. %SUBMIT R uses SAS' unnamed pipe device type to invoke the R executable. SAS datasets may be converted into R data frames and vice versa in a manner similar to using the SAS/IML ExportDataSetToR and ImportDataSetFromR functions. R graphics are also supported, and are displayed in the SAS results viewer. Graphs may be saved in user-specified locations as image files. R scripts may be created using a SAS data null step to write a file containing an R script, read from a user-specified .R input file, or by using the %R macro. Output of R execution may be directed to the SAS log file for inspection or to a user-specified .Rout file for later use.

(KEYWORDS: SAS MACRO, R SCRIPT, SAS PROGRAM FILE, UNNAMED PIPE, ODS, HTML)

Ross has been a professional SAS user for more than 20 years, contributing numerous SAS macros for the benefit of the SAS community, publishing technical articles, and lecturing on data mining algorithms and techniques. Ross' diverse academic background is one of his strengths. He has master's degrees in systems engineering from UCLA, business statistics from the University of Wisconsin, Madison, and electrical engineering from Northeastern University. Ross' expertise in statistics, computer science, and his creativity in algorithmic thinking contribute to his success in solving complex business problems for his clients.

OA-113 ***Comparison Word Clouds Using the %PROC_R macro and Base SAS® Interface***

Melvin Alexander

This presentation gives an example of calling and integrating R code from the Base SAS® environment. SAS makes it possible to run R code via SAS/IML®, SAS/IML Studio®, or SAS/Viya® as described by Gilson (2018). Interfacing R with these other SAS modules requires separate and installations that incur additional costs. R has a rich set of machine learning, text mining packages, and advanced graphic capabilities and complements SAS. I will demonstrate the R and Base SAS integration to construct comparison word clouds using a modified version of the %PROC_R macro of Wei (2012).

(KEYWORDS: COMPARISON WORD CLOUD, PROC_R, BASE SAS AND R INTEGRATION)

Melvin Alexander is an Analytician from Baltimore, MD and Greensboro, NC. He specializes in predictive and prescriptive analytics, statistical modeling and programming.

He has presented numerous papers at JMP® and SAS® Users Group meetings. He co-chairs the Mid-Atlantic JMP® Users Group (MAJUG). He earned a master's degree in Biostatistics from the University of North Carolina. He is an American Statistical Association member, an American Society for Quality (ASQ) Fellow and Certified Quality Engineer.

OA-121 ***Using the R interface in SAS® to Call R Functions and Transfer Data***

Bruce Gilson

Starting in SAS® 9.3, the R interface enables SAS users on Windows and Linux who license SAS/IML® software to call R functions and transfer data between SAS and R from within SAS. Potential users include SAS/IML users and other SAS users who can use PROC IML just as a wrapper to transfer data between SAS and R and call R functions. This paper provides a basic introduction and some simple examples. The focus is on SAS users who are not PROC IML users, but who want to take advantage of the R interface.

(KEYWORDS: R, R INTERFACE WITH SAS, SAS INTERFACE TO R)

Bruce Gilson has worked at the Federal Reserve Board for over 36 years. He spends most of his time as an in-house SAS consultant. His Bruce Force fantasy baseball team battled for its 6th league title in 31 years in 2019.

OA-152 ***Integrate Python with SAS using SASPy for a simpler, more effective script***

John Vickery

Why write two scripts in two different languages when you can get it done in one? By using the SASPy module, you can easily move data between SAS and Python sessions giving you the best of both worlds. At the NC State University Libraries, we need to manage a dynamic, million plus e-book collection with on-demand purchasing and auto-upgrade options. It is common for a publisher to release an e-book across multiple platforms each having differing access rights. In order to prevent duplicate auto-purchases we regularly combine data from SAS data sets and out catalog web services. In this case study, we show how Python handles the web service and SASPy allows us quick access to large data sets on disk. This paper will do a deep dive into the script and will show how effective an open analytics combination can be. In addition to SASPy the we also show examples of a few common Python modules such as Pandas, Requests and ElementTree.

(KEYWORDS: SASPY, PYTHON, PANDAS, WEB SERVICES)

John Vickery is analytics coordinator and collection manager for social sciences at NC State University Libraries. He's been with the NCSU Libraries since 2005 and programming in SAS since 2009. He has interests in applying analytical methods to library data for better organization in matters such as collections and service optimization.

OA-158

A Random Forest Example of the Boston Housing Data using the Base SAS® and the PROC_R macro in SAS® Enterprise Guide

Melvin Alexander

This presentation used the Boston Housing data to call and execute R code from the Base SAS® environment to create a Random Forest. SAS makes it possible to run R code via SAS/IML®, SAS/IML Studio®, or SAS/Viya® as described by Gilson (2018). Interfacing R with these other SAS modules requires separate installations that incur additional costs. R has a rich set of machine learning, text mining packages, and advanced graphic capabilities and complements SAS. I will demonstrate the R and Base SAS integration to create a Random Forest using the %PROC_R macro of Wei (2012).

(KEYWORDS: RANDOM FORESTS, PROC_R MACRO, BASE SAS AND R INTEGRATION)

Melvin Alexander is an Analytician from Baltimore, MD and Greensboro, NC. He specializes in predictive and prescriptive analytics, statistical modeling and programming.

He has presented numerous papers at JMP® and SAS® Users Group meetings. He co-chairs the Mid-Atlantic JMP® Users Group (MAJUG). He earned a master's degree in Biostatistics from the University of North Carolina. He is an American Statistical Association member, an American Society for Quality (ASQ) Fellow and Certified Quality Engineer.

OA-237

Face Recognition using SAS Viya: Guess who the person is!

Pratyush Dash and Anvesh Reddy Minukuri

Humans can take a look at an image and instantly recognize what the object is in the image, identify the person in the image or the location of the photo. The human cognitive system is fast and reliable, allowing people to perform very complex tasks like driving or operating a machine with little conscious thought. Performing these tasks for a computer would be very tough. Using fast, accurate and reliable algorithms could make computers to drive cars with sensors, enable them to recognize humans, operate different machines and even perform surgeries. The digital universe is expected to reach 44 zettabytes by 2020 because of the growth of Internet of Things (IoT). This shows us the massive opportunity we have in terms of digital content analytics. Facial recognition and classification algorithms like deep learning and neural networks can extract information from photos or videos and classify them almost instantaneously after it is posted online. There are many other applications of such algorithms as in security screening, medical image processing, and insurance claims. It is very challenging to perform this task as it requires extensive data preparation and lot of levels are needed to classify an image. In this project we have used the machine learning capabilities available in SAS® Viya® for image processing. In this paper we have used a dataset which consists of images of different people. The data was then divided into training and validation sets. For training 9 out of 10 images for each person were taken into consideration. The rest one image was used for validation. Several models will be designed using deep learning techniques like deep fully-connected neural networks (DNN), convolutional neural networks (CNN), and recurrent neural networks (RNN). The entire work has been done by using SAS DLPy which is by pulling in the Jupyter notebook into SAS® Viya®. This work proposes to analyze images and classify them as different people using the deep learning techniques and measuring accuracy. Through this project the objective of automatically detecting who the celebrity is achieved and it can be further used to segregate them into different folders.

(KEYWORDS: FACE RECOGNITION, SAS VIYA, DEEP LEARNING, COMPUTER VISION)

Chasing Master Data Interoperability: Facilitating Master Data Management (MDM) Objectives Through CSV Control Tables that Contain Data Rules that Support SAS® and Python Data-Driven Software Design

Troy Hughes

Control tables are the tabular data structures that contain control data—the data that direct software execution and which can prescribe dynamic software functionality. Control tables offer a preferred alternative to hardcoded conditional logic statements, which require code customization to modify. Thus, control tables can dramatically improve software maintainability and configurability by empowering developers and, in some cases, nontechnical end users to alter software functionality without modifying code. Moreover, when control tables are maintained within canonical data structures such as comma-separated values (CSV) files, they furthermore facilitate master data interoperability by enabling one control table to drive not only SAS software but also non-SAS applications. This text introduces a reusable method that preloads CSV control tables into SAS temporary arrays to facilitate the evaluation of business rules and other data rules within SAS data sets. To demonstrate the interoperability of canonical data structures, including CSV control tables, functionally equivalent Python programs also ingest these control tables. Master data management (MDM) objectives are facilitated because only one instance of the master data—the control table, and single source of the truth—is maintained, yet it can drive limitless processes across varied applications and software languages. Finally, when data rules must be modified, the control data within the control table must be changed only once to effect corresponding changes in all derivative uses of those master data.

(KEYWORDS: CSV FILES, CONTROL TABLES, PYTHON, SAS ARRAYS, DYNAMIC PROGRAMMING)

Troy has been a SAS practitioner for more than 20 years, has managed SAS projects in support of federal, state, and local government initiatives, and is a SAS Certified Base, Advanced, and Clinical Trials Programmer. Since 2013, he has given more than 80 presentations, hands-on workshops, and trainings at SAS Global Forum, SAS Analytics Experience, WUSS, MWSUG, SCSUG, SESUG, and PharmaSUG, and has authored two groundbreaking books: - SAS Data Analytic Development: Dimensions of Software Quality (2016) - SAS Data-Driven Development: From Abstract Design to Dynamic Functionality (2018) Troy has an MBA in Information Systems Management and additional certifications, including: Project Management Professional (PMP), Risk Management Professional (PMI-RMP), Professional in Business Analysis (PMI-PBA), Agile Certified Professional (PMI-ACP), Certified Information Systems Security Professional (CISSP), Certified Secure Software Lifecycle Professional (CSSLP), ITIL Foundation, Certified ScrumMaster (CSM), Certified Scrum Developer (CSD), Certified Scrum Product Owner (CSPO), and Certified Scrum Professional (CSP). He is a US Navy veteran with two tours of duty in Afghanistan.

OA-270 ***Make your data shine with R Shiny***

Pavan Vemuri

Data visualization and analysis is vital to the Pharma/CRO industry. Real time data access can enhance and streamline the ease of information transfer. This Paper aims to provide an example and serve as a template for other applications by making Patient profiles data reactive. R and R-shiny are used to build the profiles data as a web service instead of static reports giving the ability to access/visualize data in real time.

(KEYWORDS: R, SHINY, PATIENT PROFILES, SAS)

I have been programming for 11 years in both CRO and Pharma. I Like ADaM programming over TLFs or adhoc tasks that require deep thinking. Hobbies include wood working and eating (literally! I love eating).

OA-274 ***Deploying Models Using SAS and Open Source***

Jared Dean

In the excitement and hype around machine learning (ML) and artificial intelligence (AI), most of the time is spent on the model building. Much less energy is expended on how to take the insights from models and deploy them efficiently to create value and improve business outcomes. This paper shows a complete example using DevOps principals for building models and deploying them using SAS® in conjunction with open source projects including Docker, Flask, Jenkins, Jupyter, and Python. The reference application is a recommendation engine on a web property with a global user base. This use case forces us to confront security, latency, scalability, and repeatability. The paper outlines the final solution but also includes some of the problems encountered along the way that informed the final solution.

(KEYWORDS: OPEN ANALYTICS, DOCKER, FLASK, PYTHON, INTEGRATING SAS, MODELING)

Jared is a Principal Data Scientist and Business Knowledge Series instructor at SAS. He has developed leading edge analytics for banking, entertainment, and mobile data. He has developed and maintains projects in Python and R including the SAS kernel for Jupyter and SASPy. Outside SAS, he is an adjunct professor in the MBA program for NCSU, Duke, and Elon as well as frequent speaker and presenter. Jared holds several patents in data mining. Previously, Jared was a Mathematical Statistician for the US Census Bureau and CTO of Reveal Mobile, a mobile marketing analytics startup. He was also a Senior Director of Research and Development for SAS Enterprise Miner. He holds an MS degree in computational statistics from George Mason University and served as an advisory board member to their statistics department. Jared is the author of Big Data, Data Mining, and Machine Learning: Value Creation for Business Leaders and Practitioners (John Wiley & Sons, Inc., May 2014). His book was listed as one of five must reads in big data for the summer. When not working, he spends time with his wife and four children. Their favorite family activities include travel, sports, enjoying good food, cheering for the Denver Broncos.

Planning/Administration

PA-154 ***Developing and running an in-house SAS Users Group***

Stephen Sloan

Starting an in-house SAS Users Group can pose a daunting challenge in a large worldwide organization. However, once formed, the SAS Users Group can also provide great value to the enterprise. SAS users (and those interested in becoming SAS users) are often scattered and unaware of the reservoirs of talent and innovation within their own organization. Sometimes they are Subject Matter Experts (SMEs); other times they are new to SAS but provide the only available expertise for a specific project in a specific location. In addition, there is a steady stream of new products and upgrades coming from SAS Institute and the users may be unaware of them or not have the time to explore and implement them, even when the products and upgrades have been thoroughly vetted and are already in use in other parts of the organization. There are often local artifacts like macros and dashboards that have been developed in corners of the enterprise that could be very useful to others so that they don't have to "reinvent the wheel".

(KEYWORDS: SAS, USER, USER GROUP, USERS GROUP, COMMUNICATION)

Stephen Sloan has worked at Accenture in the Services, Consulting, and Digital groups and is currently a Data Science Senior Principal in the SAS Analytics area. He has worked in a variety of functional areas in Project Management, Data Management, and Statistical Analysis. He has had the good fortune to have worked with many talented people at SAS Institute. Stephen has presented at 20 SAS Global Forums and SAS Users Group events and has been published in professional journals.

Stephen has a B.A. cum laude with Honor in Mathematics from Brandeis University, M.S. degrees in Mathematics and Computer Science from Northern Illinois University, and an MBA (1st in class) from Stern Business School at New York University. He also has a graduate certificate in Financial Analytics from Stevens Institute.

PA-157 ***CONNECT TO vs. CONNECT USING for Security in SAS® PROC SQL***

Thomas Billings

We begin with an overview of the main SQL dialects in the SAS® system, and then describe the 2 modes of operation of PROC SQL: explicit pass-through and implicit pass-through. We focus on explicit pass-through, where PROC SQL is a wrapper to pass user-written, native RDBMS SQL code to a remote system. CONNECT TO syntax is illustrated, including the common usage of database passwords in source code. Methods to mitigate this issue are discussed – PROC PWENCODE, blotting, and encrypted, compiled macros that are user-specific. Next we illustrate an alternative, CONNECT USING paired with authentication domain LIBNAMEs. Here the password and userid are hidden in metadata (using SAS Personal Login Manager), providing a higher security alternative. We end with brief comments on explicit vs. implicit pass-through, efficiency vs. portability and maintainability.

(KEYWORDS: SECURITY, PROC SQL, CONNECT, EXPLICIT PASS-THROUGH, AUTHENTICATION DOMAIN, LIBNAME, PROC PWENCODE)

Tom has used the SAS system intermittently since the mid-1970's, in a variety of different industries and application areas. He is currently working in Banking, using SAS Enterprise Guide and other tools to build and analyze databases.

PA-227

Aligning Analytics Strategy with Business Strategy

Murali Sastry

As more and more organizations are adopting analytics, organizational strategic planning activities need to keep pace with the evolving business needs to include analytics function and data analytics in its strategic planning design. Strategic planning approaches (for example, Hoshin Kanri or Balanced Score Card) adopted need to exploit analytics capabilities of the organization for successful strategy implementation. Hoshin Kanri is a well-known strategic planning methodology implemented by Toyota® Motor Corporation, Bridgestone® Corporation, Hewlett-Packard Enterprise (HPE or HP®) and many others. Analytics function is critical in any organization specifically during strategic plan deployment. It is imperative to align analytics strategy with business strategy to ensure an organization achieves its tactical and strategic objectives while fulfilling the needs of its stakeholders. The intent of this paper is to familiarize Hoshin Kanri Strategic Planning Methodology to analytics professionals to provide them with the tools, techniques, and know-how to use in establishing a sound analytics strategy that is integrated with the organizational business strategy.

(KEYWORDS: STRATEGIC PLANNING, HOSHIN KANRI, BALANCED SCORE CARD, ANALYTICS, BUSINESS STRATEGY)

Murali Sastry received his MS in Analytics from Capella University in 2017. As a Data Reporting Analyst of Analytics Team in Finance at LGFCU in Raleigh, North Carolina, he assists the analytics team across departments and applies analytics methods using SAS products. He has been working for LGFCU since Jan 2018 and he has been using SAS since 2015. He has been designing, implementing, and monitoring Strategic Planning activity since early 2000s for several organizations as a member of corporate staff in executive leadership team for manufacturing and pharmaceutical companies. He has used Hoshin Kanri, Balanced Score Card and other strategic planning approaches during his career.

PA-257

Automating SAS® Viya® Admin tasks using CLI and Chatbot

Sandeep Grande

SAS® Viya® platform comes with a new command line interface to interact with microservices. This paper is an attempt to embrace the openness of Viya Platform by creating a Chatbot which helps SAS Administrator in performing his/her day to day tasks. While there are many ways to automate the Admin tasks, this paper explores the latest cloud services such as AWS Lex chatbot service, AWS Lambda which is a serverless computing platform to create user Interactive chat bot with Slack application chatbot being the front end. This chatbot can be easily customized to work at our voice commands. The lambda function uses python runtime environment and we also explore the way we can interact with microservices using python.

(KEYWORDS: ADMINISTRATION, CHATBOT)

Solving SAS Performance Problems: Our Methodology**Jim Kuell**

Diagnosing performance issues can be a lengthy and complicated process. For many, the most difficult step is figuring out where to begin. This typically leads to a track being opened with SAS Technical Support. The SAS Performance Lab has developed a standard methodology for diagnosing performance issues based on years of experience doing so both internally and at customer sites. This process is regularly applied when assisting with performance issues in SAS Technical Support tracks. This presentation goes through the methodology used by the SAS Performance Lab to diagnose performance issues and discusses resolutions to the most common problems.

(KEYWORDS: PERFORMANCE ISSUES, SAS PERFORMANCE LAB METHODOLOGY, RESOLVING COMMON ISSUES)

Jim is a Software Performance Engineer in the SAS® Platform R&D Performance Lab. He works closely with Technical Support, SAS customers, and SAS hardware partners on a daily basis. The goal of his work is to maximize performance at customer sites, both in on-premise and cloud-based environments.

Reporting/Visualization

RV-109 ***JMP®'s Visualization Analysis of SESUG Conference Attendance from 2008-2018***
Melvin Alexander

The SESUG 2018 Information and Visualization Competition had contestants who used creative, clever SAS® Visualization tools describing attendees of past conference years from 2008-2017. Although the competition was open to JMP® users, there were no JMP® submissions. This presentation applies JMP®'s visualization tools to answer some of the key questions raised in the competition. The tools featured include:

- Displaying SESUG attendance by Industry for each Conference year with image logos from JMP®'s Expressions columns;
- Animating how state and industry representation of attendees changed over time using JMP®'s Local Data Filter;
- Using Distribution Analysis of the Industries SESUG conference attendees came from so that committee planners can identify topics of interest for attendees.

The data used in this presentation covered the same past conference years, including 2018. JMP®'s local data filters, dynamic and interactive visualization functionality allows analysts to show off data in interesting and exciting ways.

(KEYWORDS: EXPRESSIONS COLUMNS, INTERACTIVE VISUALIZATION, LOCAL DATA FILTER ANIMATION)

Melvin Alexander is an Analytician from Baltimore, MD and Greensboro, NC. He specializes in predictive and prescriptive analytics, statistical modeling and programming.

He has presented numerous papers at JMP® and SAS® Users Group meetings. He co-chairs the Mid-Atlantic JMP® Users Group (MAJUG). He earned a master's degree in Biostatistics from the University of North Carolina. He is an American Statistical Association member, an American Society for Quality (ASQ) Fellow and Certified Quality Engineer.

RV-115 ***Pie is delicious but not nutritious: Graphics for univariate data.***
Peter Flom

When you have univariate data, that is, a single measure on a variety of units, the most common statistical graphic is a pie chart. But pie charts should not be used. Ever. When there are a lot of units, pie charts are unreadable. When there are only a few units, pie charts waste space. And research \cite{Cleveland1993,Cleveland1994} shows that, even with a moderate number of units, pie charts can distort the data (for example, using different colors leads to different estimates of the size of the wedges). Fortunately, there are better methods. SAS. No operating or version dependencies. Any level of user.

(KEYWORDS: PIE CHARTS, GRAPHICS, GTL)

Peter Flom is an independent statistical consultant working with researchers and graduate students in the social, medical and behavioral sciences. He has been using SAS for over 20 years and has presented at SGF, SUGI, NESUG, SESUG, MWSUG and local groups.

Making Your SAS® Results More Meaningful with Color**Kirk Paul Lafler**

Color can help make your SAS® results more meaningful. Instead of producing boring and ineffective results, users are able to enhance the appearance of their output, documents, reports, tables, charts, statistics, and spreadsheets to highlight and draw attention to important data elements, details, and issues, including using color in headings, subheadings, footers, minimum and maximum values, ranges, outliers, special conditions, and other elements. Color can be added to text, foreground, background, rows, columns, cells, summaries, totals, and traffic lighting scenarios. Topics include using color to results, documents, reports, tables, charts and spreadsheets can be enhanced with color, effectively add color to PDF, RTF, HTML, and Excel spreadsheet results using PROC PRINT, PROC FREQ, PROC REPORT, PROC TABULATE, and PROC SGPLOT and Output Delivery System (ODS) with style.

(KEYWORDS: COLOR, VISUALIZATION, SGPLOT, SGPANEL, ATTRIBUTE MAP, BACKGROUND COLOR, FOREGROUND COLOR)

Kirk Paul Lafler is entrepreneur, consultant and founder of Software Intelligence Corporation, and has worked with SAS software since 1979. As a SAS consultant, application developer, programmer, data analyst, mentor, infrastructure specialist, educator and author at Software Intelligence Corporation, and an advisor and SAS programming adjunct professor at the University of California San Diego Extension, Kirk has taught SAS courses, seminars, webinars and workshops to thousands of users around the world. Kirk has also authored or co-authored several books including PROC SQL: Beyond the Basics Using SAS, Third Edition (SAS Press, 2019) and Google® Search Complete (Odyssey Press, 2014); hundreds of papers and articles on a variety of SAS topics; served as an Invited speaker, educator, keynote and section leader at SAS user group conferences and meetings worldwide; and is the recipient of 25 "Best" contributed paper, hands-on workshop (HOW), and poster awards.

Building a Better Dashboard Using Base-SAS® Software**Kirk Paul Lafler and Joshua M. Horstman**

Organizations around the world develop business intelligence dashboards, sometimes referred to as enterprise dashboards, to display the current status of “point-in-time” metrics and key performance indicators. Effectively designed dashboards extract real-time data from multiple sources for the purpose of highlighting important information, numbers, tables, statistics, metrics, performance scorecards and other essential content on a single screen. Topics include the basic rules for “good” dashboard design, the metrics frequently used in dashboards, and the use of best practice programming techniques in the design of highly interactive, filterable, and drill-down dashboards using SAS® Base software. Attendees learn how to create a real-world static and interactive dashboard using SAS® Base programming techniques including the use of the DATA step, PROC FORMAT, PROC PRINT, PROC MEANS, PROC SQL, ODS, ODS Statistical Graphics, PROC SGRENDER, PROC SGPLOT, PROC SGSCATTER, PROC SGPANEL, and PROC TEMPLATE.

(KEYWORDS: BASE-SAS TECHNIQUES, DATA STEP, PROC FORMAT, SGPLOT, ODS, PROC TEMPLATE)

Kirk Paul Lafler is entrepreneur, consultant and founder of Software Intelligence Corporation, and has worked with SAS software since 1979. As a SAS consultant, application developer, programmer, data analyst, mentor, infrastructure specialist, educator and author at Software Intelligence Corporation, and an advisor and SAS programming adjunct professor at the University of California San Diego Extension, Kirk has taught SAS courses, seminars, webinars and workshops to thousands of users around the world. Kirk has also authored or co-authored several books including PROC SQL: Beyond the Basics Using SAS, Third Edition (SAS Press, 2019) and Google® Search Complete (Odyssey Press, 2014); hundreds of papers and articles on a variety of SAS topics; served as an Invited speaker, educator, keynote and section leader at SAS user group conferences and meetings worldwide; and is the recipient of 25 "Best" contributed paper, hands-on workshop (HOW), and poster awards. Josh Horstman is an independent statistical programming consultant and trainer based in Indianapolis with over 20 years' experience using SAS in the life sciences industry. He specializes in analyzing clinical trial data, and his clients have included major pharmaceutical corporations, biotech companies, and research organizations. Josh is a SAS Certified Advanced Programmer who loves coding and is a frequent presenter at SAS Global Forum and other SAS User Group meetings.

Josh Horstman is an independent statistical programming consultant and trainer based in Indianapolis with over 20 years' experience using SAS in the life sciences industry. He specializes in analyzing clinical trial data, and his clients have included major pharmaceutical corporations, biotech companies, and research organizations. Josh is a SAS Certified Advanced Programmer who loves coding and is a frequent presenter at SAS Global Forum and other SAS User Group meetings.

RV-147

Great Time to Learn GTL: A Step-by-Step Approach to Creating the Impossible

Richann Watson

Output Delivery System (ODS) graphics, produced by SAS® procedures, are the backbone of the Graph Template Language (GTL). Procedures such as the Statistical Graphics (SG) procedures dynamically generate GTL templates based on the plot requests made through the procedure syntax. For this paper, these templates will be referenced as procedure-driven templates. GTL generates graphs using a template definition that provides extensive control over output formats and appearance. Would you like to learn how to build your own template and make customized graphs and how to create that one highly desired, unique graph that at first glance seems impossible? Then it's a Great Time to Learn GTL! This paper guides you through the GTL fundamentals while walking you through creating a graph that at first glance appears too complex but is truly simple once you understand how to build your own template.

(KEYWORDS: ODS GRAPHICS, GTL, GRAPH TEMPLATE LANGUAGE, LAYOUT, OVERLAY, GRIDDED)

Richann Watson is an independent statistical programmer and CDISC consultant based in Ohio. She has been using SAS since 1996 with most of her experience being in the life sciences industry. She specializes in analyzing clinical trial data and implementing CDISC standards. Additionally, she is a member of the CDISC ADaM team and various sub-teams.

RV-165

Dressing Up your PROC SGPLOT and SGMAP Output with Attributes and Annotation

Louise Hadden

The output from PROC SGPLOT and SGMAP is often "camera-ready" without any intervention, but occasionally there is a need for customization. There are tools such as GTL to facilitate custom output, but SAS has also added a number of "ATTRS" which modify various elements of graphics output, and additionally permits annotation (for example, FILLATTRS, TITLEATTRS and MARKERATTRS). This paper will walk through these tools, focusing primarily on maps created with PROC SGPLOT's polygon statement and PROC SGMAP.

(KEYWORDS: POLYGON STATEMENT, ANNOTATION, ATTRIBUTES, SGMAP, SGPLOT, STYLE)

Louise Hadden presented at her first SAS conference in 1996 and has never looked back, presenting at multiple conferences across the continent over the years. She supports analytic processing at Abt Associates Inc., a social science research company, and specializes in reporting and data visualization in the division of health and environment.

RV-172 ***Probability Plots for Exploratory Data Analysis***

Dennis Beal

Probability plots are used in statistical analysis to check distributional assumptions, visually check for potential outliers and see the range, median and variability of a data set. Probability plots are an important statistical tool to use for exploratory data analysis. This paper shows SAS® code that generates normal and lognormal probability plots using the Output Delivery System (ODS) on a real environmental data set using PROC UNIVARIATE and interprets the results. This paper is for beginning or intermediate SAS users of Base SAS® and SAS/GRAPH®.

(KEYWORDS: NORMAL PROBABILITY PLOTS, LOGNORMAL PROBABILITY PLOTS, PROC UNIVARIATE)

The presenter is a senior statistician and risk scientist at Leidos in Oak Ridge, Tennessee. He supports a wide variety of projects for the U.S. Department of Energy, primarily in environmental cleanup of legacy wastes in Oak Ridge. He earned his Ph.D. in management science with a dissertation on statistical applications at the University of Tennessee in Knoxville. He also is an adjunct professor of statistics for several universities in the Knoxville area. He has used SAS for 30 years and has presented papers at SESUG every year since 2004.

RV-176 ***Using Heat Map or Gtile: Does Size Matter in Your Graphics?***

Keith Brown and Devi Sekar

With SAS 9.2 and beyond, ODS Graphics brings in a new way of generating high quality graphs. Many users still find themselves at the crossroads, trying to decide what path to follow – the traditional SAS/GRAPH or ODS Graphics. Both can produce most of the common types of graphs, such as scatter plots, regression and box plots, line graphs, bar charts, and histograms. In this paper we will share examples to generate a heat map using SGPLOT in ODS Graphics and a GTILE graph with SAS/GRAPH and discuss the advantages of each. With heat maps, you can display patterns in the data for a chosen response variable for one-dimensional data (in a map) or two-dimensional data (in a table or graph). Heat maps make tables and graphs easier to interpret, by shading the background color based on the frequency of observations in each cell of the graph or table. With GTILE, the response values are shown by both color gradient and area. We will also show how to customize a heat map by making use of the SAS Graph Template Language (GTL) and SGRENDER procedure. Users will appreciate how quick and easy it is to generate sophisticated graphs with ODS graphics.

(KEYWORDS: SAS/GRAPH, SGPLOT, HEAT MAP, ODS GRAPHICS, GTILE)

A Table 1 Macro that Produces Publication-Ready Results: %Table1nDone**Martha Wetzel**

Academic papers in many fields include a table summarizing the demographic characteristics of the sample and/or treatment groups—the ubiquitous Table 1. These tables often require an overall summary, as well as a between-group comparison. Furthermore, these tables summarize mixed data types, including both continuous and categorical data, requiring different statistical tests and SAS procedures. Without automation, analysts can spend hours per project calculating and arranging the results into the correct format, often having to redo the entire table when an investigator realizes there is a data error. The %Table1nDone macro was created to reduce analyst time spent on Table 1s. This paper presents a new Table 1 macro that calculates summary statistics overall and by group, performs the corresponding statistical testing as required, and produces an RTF file containing the final summary (i.e., the Table 1). The %Table1nDone macro expands on existing Table 1 macros by 1) streamlining variable input via an Excel file, 2) creating a table with both overall and by-group summary data, 3) producing an RTF table in the format expected by many journals, and 4) saving permanent data sets of key information for analyst review. Logic is built into the macro to select the appropriate statistical test based on user-supplied factors such as variable type (e.g., categorical, continuous) and data factors (e.g., number of comparison groups, cell-size counts, distributional assumptions). Publication-ready output contains results formatted as “N (%)”, “mean (standard deviation)”, or “median (quartile 1, quartile 3)”, depending on the type of data. In addition, the macro produces a report for the analyst to review for unexpected values in the data. This paper provides an overview of the macro’s capabilities, a description of the use and required parameters, an explanation of the statistical tests included, examples of output, and links to the macro code.

(KEYWORDS: TABLE 1, DESCRIPTIVES TABLE, DESCRIPTIVES MACRO, BIVARIATE ANALYSIS)

Martha Wetzel is a biostatistician in the Department of Pediatrics at Emory University. Her research areas of interest include opioid policy and pain management. She holds an MSPH in health services research and health policy from Emory University and previously worked in the health care quality improvement field.

The Gender Gap in the Education System**Geoffrey Dean and Jared Dean**

Despite our society's best efforts to close the education gap in gender, the data indicates that this has not yet been achieved. This paper will investigate the potential causes for the phenomena we see in our education system. This analysis of the gender gap in education will use SAS in conjunction with open source software to analyze data from a variety of sources. These sources include standardized tests such as the College Board's Advanced Placement exams and results from the ACT, school district reports from the North Carolina Department of Public Instruction, among others. These various data sources will be analyzed, compared, and visualized. The visualizations will aim to confirm or refute conventional wisdom, reveal underlying patterns, and increase awareness of both past and current conditions in the educational system.

(KEYWORDS: EDUCATION, OPEN SOURCE, VISUALIZATION, STATISTICS)

Geoffrey is a Senior at Green Hope High School. He writes for the school news site and is the Chief Data Editor. Geoffrey has experience creating graphics using SAS, JMP, Python, and D3.

*Jared is a Principal Data Scientist and Business Knowledge Series instructor at SAS. He has developed leading edge analytics for banking, entertainment, and mobile data. He has developed and maintains projects in Python and R including the SAS kernel for Jupyter and SASPy. Outside SAS, he is an adjunct professor in the MBA program for NCSU, Duke, and Elon as well as frequent speaker and presenter. Jared holds several patents in data mining. Previously, Jared was a Mathematical Statistician for the US Census Bureau and CTO of Reveal Mobile, a mobile marketing analytics startup. He was also a Senior Director of Research and Development for SAS Enterprise Miner. He holds an MS degree in computational statistics from George Mason University and served as an advisory board member to their statistics department. Jared is the author of *Big Data, Data Mining, and Machine Learning: Value Creation for Business Leaders and Practitioners* (John Wiley & Sons, Inc., May 2014). His book was listed as one of five must reads in big data for the summer. When not working, he spends time with his wife and four children. Their favorite family activities include travel, sports, enjoying good food, cheering for the Denver Broncos.*

Data Management Challenge: Select All That Apply: JMP® to the Rescue**Mira Shapiro**

Online surveys often provide questions with a variety of choices and instruct the user to "Select All That Apply". Depending upon the backend of the system, the desired analytics and output, and available tools, these sort of responses can sometimes create data management challenges. Hidden under JMP Col?Utilities are several tools that allow the user to recode the individual responses and then combine them into one column characterizing the user's complete response to a particular question. In this discussion we will show step by step how to use these techniques and demonstrate how the results can be used for analytics

(KEYWORDS: JMP, DATA MANAGEMENT, RECODE, SURVEY)

One Click to Analysis Results Metadata**Srivathsa Ravikiran and Priscilla Gathoni**

Analysis Results Metadata (ARM) is a plugin to ADaM Define.xml which provides information on Analysis data used to generate efficacy, safety and clinical study results. It establishes one to one relationship between the analysis result, ADaM data and variables used to perform the analysis. ARM plug-in benefits the reviewer by providing clear outline of the critical analysis displays, details on the analysis performed and methods used. Due to these added benefits of the ARM, in recent days there have been increased interest from regulatory agencies to submit ARM along with the ADaM define.xml. Creation of ARM will soon become a necessity for the regulatory submissions. Pinnacle21 Community version is a popular tool to generate the define.xml, however it does not have the ARM creation feature in the current version (v2.2.0). Due to this limitation and increasing interest to include the ARM in ADaM define.xml by the regulatory authorities, there is an unmet need for an alternative approach to generate the ARM metadata for submissions. This paper will present an approach using SAS to generate the ARM which can be integrated to an existing ADaM define.xml. The Resulting Integrated define.xml (ADaM define.xml + ARM) passes all compliance checks using the Pinnacle21 validator tool and is compliant for all regulatory submissions.

(KEYWORDS: ANALYSIS RESULTS METADATA, ARM, ADAM DEFINE.XML)

Srivathsa Ravikiran is a Senior Statistical Programmer at AstraZeneca Pharmaceuticals in Gaithersburg, Maryland. Srivathsa holds a Master's degree in Drug Regulatory Affairs from Northeastern University, Massachusetts. Srivathsa is a SAS certified programmer who specializes in clinical data analysis and reporting. Srivathsa holds a Private & Instrument Pilot License from the FAA and enjoys his free time flying. This is Srivathsa's first appearance in SESUG and a first-time author!

Priscilla Gathoni is a Statistical Programming Director at AstraZeneca Pharmaceuticals in Gaithersburg, Maryland. Priscilla holds an MS in Statistics and an MBA in Management from Colorado State University. Priscilla has authored several papers and is an active presenter in PharmaSUG, SESUG, and CDISC conferences. Priscilla is a respected leader in the pharmaceutical industry and has passion in leadership, people management, coaching-mentoring, sustainability, standards management, and loves to motivate, arouse enthusiasm, and inspire people around her.

Salesmanship with Graphics – A Panel Discussion**Nat Wooding, Robert Allison, Philip Gregory, and Martha Pohl (moderated by Barbara Okerson and Louise Hadden)**

Today we are a visual society that identifies products and brands by their logos and graphic advertisements. However, graphics are more than tools for 'at a glance' corporate and product recognition. They are also used as an art of persuasion to sell ideas, products, or philosophies. When used as intended, graphics can be very compelling for marketing. The ability to manipulate thoughts and ideas with graphics is also effective and has been practiced for a long time, e.g. an 1864 newspaper photo purported to be of General Grant on a horse in front of his troops (demonstrating the leadership the nation needed to see) when the body on the horse was actually Major General Cook. This panel will discuss:

- Effective marketing graphics,
- Effective graphics platforms, and
- Ethics when 'selling with graphics'.

Examples of each will be provided and audience participation is encouraged.

(KEYWORDS: PANEL DISCUSSION, REPORTING AND VISUALIZATION)

What's in the pipeline for the SGMAP Procedure?!?**Robert Allison**

SGMAP is a fairly new procedure in BASE SAS, and is undergoing rapid development and enhancements. It currently has rich capabilities for plotting data on tile-based maps, such as OpenStreetMaps® (as of release 9.4 maintenance 6), but only basic capabilities for creating choropleth (polygon-based) maps. In this paper I describe several new features under development that will help round out the choropleth mapping capabilities. These new features are "all but done" and should be available in a SAS release in the near future.

(KEYWORDS: SGMAP PROCEDURE, TILE-BASED MAPS, CHOROPLETH MAPS, FUTURE SAS FEATURES)

Robert Allison has worked at SAS for over a quarter century, and his specialty is creating custom graphs and maps.

Statistics/Data Analysis

SD-102 **Glass Box Neural Networks**

Ross Bettinger

Neural network models are typically described as “black boxes” because their inner workings are not easy to understand. We propose that, since a neural network model that accurately predicts its target variable is a good representation of the training data, the output of the model may be recast as a target variable and subjected to standard regression algorithms to “explain” it as a response variable. Thus, the “black box” of the internal mechanism is transformed into a “glass box” that facilitates understanding of the underlying model. Deriving a regression model from a set of training data analogous to a neural network is an effective means to understand a neural network model because regression algorithms are commonly-used tools and the interpretation of a regression model is straightforward and well understood.

(KEYWORDS: LOGISTIC REGRESSION, NEURAL NETWORKS, VARIABLE SELECTION, BLACK BOX, GLASS BOX)

Ross has been a professional SAS user for more than 20 years, contributing numerous SAS macros for the benefit of the SAS community, publishing technical articles, and lecturing on data mining algorithms and techniques. Ross' diverse academic background is one of his strengths. He has master's degrees in systems engineering from UCLA, business statistics from the University of Wisconsin, Madison, and electrical engineering from Northeastern University. Ross' expertise in statistics, computer science, and his creativity in algorithmic thinking contribute to his success in solving complex business problems for his clients.

SD-103 **Hybrid Rare Event Sampling Technique**

Ross Bettinger

We discuss a hybrid sampling methodology for building and validating machine learning models for classification for which the target event is a small proportion of the total population. This methodology may be useful for fraud detection or similar applications for which the response variable is typically scarce. We call this variation of cross-validation methodology the HYbrid Rare Event Sampling (HYRES) technique because it is meant to detect rare events with high resolution through sampling.

(KEYWORDS: CLASSIFICATION, IMBALANCED CLASSES, MODEL BUILDING, MODEL VALIDATION, RARE EVENT, SAMPLING PROTOCOL)

Ross has been a professional SAS user for more than 20 years, contributing numerous SAS macros for the benefit of the SAS community, publishing technical articles, and lecturing on data mining algorithms and techniques. Ross' diverse academic background is one of his strengths. He has master's degrees in systems engineering from UCLA, business statistics from the University of Wisconsin, Madison, and electrical engineering from Northeastern University. Ross' expertise in statistics, computer science, and his creativity in algorithmic thinking contribute to his success in solving complex business problems for his clients.

SD-114 ***Scatterplots: Basics, enhancements, problems and solutions.***

Peter Flom

The scatter plot is a basic tool for presenting information on two continuous variables. While the basic plot is good in many situations, enhancements can increase its utility. I also go over tools to deal with the problem of overplotting. SAS, any operating system or version, appropriate for all levels.

(KEYWORDS: SCATTERPLOT, GRAPHICS, GTL)

Peter Flom is an independent statistical consultant working primarily with researchers and graduate students in the social, behavioral and medical sciences. He has been using SAS for over 20 years and has presented at SGF, SUGI, NESUG, MWSUG, SESUG and local groups.

SD-118 ***The Best of Both Worlds: Forecasting Using Time Series with Inputs***

Dave Dickey

Regression and time series are both good tools for forecasting data taken over time. It stands to reason that combining the good features of both would give a better forecast. than either one alone. The AUTOREG procedure in SAS accomplishes this in a user friendly way. There is a nice degree of automation including an automatic identification of the autocorrelation structure within the broad class of stationary autoregressive structures. This is often a good first step or possibly even an acceptable final model. While the autoregressive class is large, there are more structures in the broader ARIMA class available in the ARIMA procedure. After a brief introduction to the ideas underlying the procedures, the talk will focus on examples. The two procedures will be compared and advantages of each will be illustrated. Anyone familiar with the REG, GLM or MIXED procedure will find it easy to transition to PROC AUTOREG. A theme will be the problems induced by using standard ordinary least squares methods when the residuals are correlated and how to overcome them through the use of the AUTOREG or ARIMA procedure.

(KEYWORDS: AUTOCORRELATION, PREDICTION, GENERALIZED LEAST SQUARES, IDENTIFICATION, ARIMA)

Dr. David Dickey has been a contract instructor for SAS since 1981, is a books by users author and a frequent presenter at SGF and SESUG. Dave is Emeritus Distinguished Professor of Statistics at NC State University and a Fellow of the American Statistical Association. He is co-developer of the Dickey-Fuller test for time series that appears in many time series computer packages including SAS.

Intervention at the Library: Assessing the Effect of an Event**Dave Dickey and John Vickery**

A highlight of any visit to the NC State University campus is a tour of the modern Hunt Library. Books are delivered for checkout by a robot called the BookBot and some interest was expressed by library personnel in determining the effect of the Hunt Library's opening on book circulation. Circulation numbers were available before and after the new library opening for two classes of books: those that stayed at the old library and those that were moved to the new and much more modern library. Such an event in time series is referred to as an intervention or sometimes as an interrupted time series. A simple before and after t-test showed some interesting deficiencies that led to a more thoughtful and ultimately revealing analysis. This talk shows the process of going from a simple initial analysis through more and more sophisticated approaches as the layers of complexity are peeled back. The lessons learned here are applicable to any situation in which the effect of a known event is to be investigated. Care in checking assumptions is emphasized.

(KEYWORDS: REGRESSION, INTERRUPTED TIME SERIES, INTERVENTION, INSTITUTIONAL RESEARCH)

Dr. David Dickey has been a contract instructor for SAS since 1981, is a books by users author and a frequent presenter at SGF and SESUG. Dave is Emeritus Distinguished Professor of Statistics at NC State University and a Fellow of the American Statistical Association. He is co-developer of the Dickey-Fuller test for time series that appears in many time series computer packages including SAS.

John Vickery is analytics coordinator and collection manager for social sciences at NC State University Libraries. He's been with the NCSU Libraries since 2005 and programming in SAS since 2009. He has interests in applying analytical methods to library data for better organization in matters such as collections and service optimization.

Generalized Linear Mixed Model Approach to Time-to-Event Data with Censored Observations**Kathleen Yeater, George Yocum, Kendra Greenlee, Julia Bowsher, Arun Rajamohan and Joseph Rinehart**

The time-to-event response is commonly thought of as survival analysis, and typically concerns statistical modeling of expected life span. In the example presented here, alfalfa leafcutting bees, *Megachile rotundata*, were randomly exposed to one of eight experimental thermoprofiles or two control thermoprofiles, for one to eight weeks. The incorporation of these fluctuating thermoprofiles in the management of the bees increases survival and blocks the development of sub-lethal effects, such as delayed emergence. The data collected here investigates the question of whether any experimental thermoprofile provides better overall survival, with a reduction and delay of sub-lethal effects. The study design incorporates typical aspects of agricultural research; random blocking effects. All *M. rotundata* prepupae brood cells were randomly placed in individual wells of 24-well culture plates. Plates were randomly assigned to thermoprofile and exposure duration, with three plate replicates per thermoprofile x exposure time. Bees were observed for emergence for 40 days. All bees that were not yet emerged prior to fixed end of study were considered to be censored observations. We fit a generalized linear mixed model (GLMM), using the SAS® GLIMMIX Procedure to the censored data and obtained time-to-emergence function estimates. As opposed to a typical survival analysis approach, such as Kaplan-Meier curve, in the GLMM we were able to include the random model effects from the study design. This is an important inclusion in the model, such that correct standard error and test statistics are generated for mixed models with non-Gaussian data.

(KEYWORDS: GLIMMIX PROCEDURE, CENSORED OBSERVATIONS, TIME-TO-EVENT RESPONSE, RANDOM EFFECTS)

Kathleen Yeater is an area statistician for the USDA-Agricultural Research Service-Plains Area Office of the Director. Dr. Yeater received her Ph.D in Biometry & Statistics, from the University of Illinois at Urbana-Champaign. She specializes in applied statistical methods for agricultural research studies, including multivariate methods and experimental design, with special attention to phenomic-genomic relationships.

PROC MCMC Application-Time Series Forecasting**Murali Sastry and Tyler Hicks**

Forecasting variables of interest in time series analysis can be quite intriguing. It can also be challenging to identify various factors that predict time series data patterns. For example, time series forecasting has been effectively applied in scheduling nurses in hospital emergency rooms, where the number of patients routinely fluctuates throughout the day based on the time of month and time of year. Similarly, time series forecasting has been helpful in forecasting credit card use by cardholders based on their past card usage history. The tricky part in modeling time series data is to combine variables such as seasonality, trend, and regressor components in one model to develop an accurate forecast. Whereas SAS/STAT® PROC ARIMA (Autoregressive Integrated Moving Average Procedure) offers SAS users a classical approach to estimating models for time series data with autoregressive, differencing, and moving average structures, it is also possible for SAS users to fit these same models to time series data using Bayesian approach with SAS/STAT® PROC MCMC (Markov Chain Monte Carlo procedure). The intent of this paper is to provide SAS users with example code and demonstrations on how to use PROC MCMC to estimate models suitable in time series analysis.

(KEYWORDS: PROC MCMC, BAYESIAN STRUCTURAL MODEL, PROC ARIMA, TIME SERIES FORECASTING)

Murali Sastry received his MS in Analytics from Capella University in 2017. As a Data Reporting Analyst of Analytics Team in Finance at Local Government Federal Credit Union (LGFCU) in Raleigh, North Carolina, he assists the analytics team across departments and applies analytics methods using SAS products. He has been working for LGFCU since Jan 2018 and he has been using SAS since 2015.

Effect of Manpower Policy Changes on Retention and Promotions in the U.S. Marine Corps

Jon Swanson

This analysis was conducted to advise and assist executive decision makers on possible effects on policy changes that would effect enlisted promotions and retention in the U.S. Marine Corps. On an annual basis the Marine Corps retains less than 30% of Marines on their initial contract. In some cases this can create gaps in manpower and readiness when Marines in certain occupational specialties are promoted the rank of E5 prior to the end of their initial contract. In fiscal year 18 approximately 30% of Marines who exited after their initial contract (48 months) reached the grade of E5 prior to exiting active service. The proposed policy change would not allow promotion to E5 prior to reaching 48 months of service in order to prevent this potential gap in manpower in specific occupational fields. What this led to in relation to this study was an in depth look into different occupational specialties and all enlisted ranks in regards to future manpower planning decisions. We used various SAS procedures and techniques along with other statistical analysis practices in the context of the Marine Corps to see where we could be affected by this policy change. Also, an in depth look into promotion rates which can effect the ability for Marine to complete required training prior to be considered for future advancement.

(KEYWORDS: MILITARY, FORECASTING, MANPOWER PLANNING)

Jonathan Swanson was born in Greensboro, North Carolina 26 March, 1982. He graduated from Page High School in June of 2000. Jonathan Swanson enlisted in the Marine Corps in 2000 and attended recruit training at Marine Corps Recruit Depot Parris Island. Upon completion of recruit training in October of 2000, Private First Class Swanson attended Bravo Company, Marine Corps Communications Electronics School, 29 Palms, California. He was assigned the military occupational specialty 2531, Field Radio Operator. Upon completion of Field Radio Operators Course, Lance Corporal Swanson was assigned to Communications Company Headquarters and Service Battalion, Fourth Force Service Support Group, Greensboro, NC. In January of 2003, Lance Corporal Swanson was mobilized in support of Operation Enduring Freedom where he served as a Radio Operator in Kuwait and Iraq. In October of 2003 he was promoted to the rank of Corporal. In February 2004, Corporal Swanson deployed to Central and South America in support of multi-national training exercise UNITAS, where he served as a Field Radio Operator. In April of 2005 he was promoted to the rank of Sergeant and deployed to Fallujah, Iraq in support of Operation Iraqi Freedom where he served on the Personal Security Detail for the Multi National Forces West Commander. In June 2007, Sergeant Swanson was selected to become an 0651 in the Active Reserve program and was assigned as a Data Network Specialist with Fourth Supply Battalion, Fourth Marine Logistics Group, Newport News, VA where he served as the Information Systems Coordinator for the battalion. In June of 2010, he was promoted to the rank of Staff Sergeant and assigned to Fourth Light Armored Reconnaissance Battalion, Fourth Marine Division, Camp Pendleton, CA where he served as the battalion's Communications Chief. In July of 2011 Staff Sergeant Swanson graduated from Liberty University with a Bachelor's of Science in Business Administration. In April of 2013, Staff Sergeant Swanson was assigned to Communications Company, Fourth Marine Logistics Group, Greensboro, NC where he serves as the Cyber Systems Chief and Assistant Operations Chief. In December 2014, he was promoted to the rank of Gunnery Sergeant. In June of 2016, Gunnery Sergeant Swanson was transferred to his current assignment at HQMC, Reserve Affairs to serve as Inventory and Requirements Chief. In December of 2016, he completed all requirements for a Master's of Science in Information Systems Management from Liberty University.

Master Sergeant Swanson's personal awards include the Navy Commendation Medal with two gold stars in lieu of a third award, the Navy/Marine Corps Achievement Medal with one gold star in lieu of second award, and the Combat Action Ribbon.

SD-174 ***ePRO: A View from a Statistical Programmer***

Vijetha Kode

Assessing patients experiences and perspectives into their clinical care is an important mechanism for evaluating the Quality of Life (QoL) of patients participating in cancer clinical trials. Patient-Reported Outcomes (PROs) commonly would capture patient perspective systematically and could assist in the development of new cancer therapies. European Organization for Research and Treatment of Cancer (EORTC) created and developed an integrated, modular approach for evaluating the QoL of patients participating in cancer clinical trials. This led to the development of the EORTC QLQ-C30, a quality of life instrument for cancer patients. Further more, EORTC developed various types of Questionnaires within various types of cancers. This paper primarily focuses on statistical programming aspects of PRO analysis for questionnaires (QLQ-C30, QLQ-LC13 and EQ-5D-5L) collected in Lung Cancer Indication trials. Details on the process of mapping collected data to Study Data Tabulation Model (SDTM), creation of Analysis Data Model (ADaM) datasets and various types of analysis reports typically included in a Clinical Study Report (CSR) will be discussed in this paper.

(KEYWORDS: EPRO, SDTM, ADAM, CDISC, QUESTIONNAIRES, ONCOLOGY, LUNG, CANCER, QOL, QLQC30, QLQLC13, EQ5D5L, EORTC)

SD-191 ***Bootstrapping Regression Models using PROC SURVEYSELECT***

Bryce Whitehead and Austin Brown

When constructing regression models, it is commonplace for a researcher to be interested in assessing the relationship between categorical predictor variables and some response variable. It is common for a categorical predictor to be dichotomous and coded as "0" or "1" in the dataset. However, when a large proportion of the observations fall into either category (i.e., greater than 80% or 90%) parameter estimation can become unreliable as the standard error of the estimator may become either inflated or deflated. Such a data situation may occur in observational types of analyses. One way of addressing this concern could be through taking a random sample from the larger group to match the sample size of the smaller group and then fitting the desired model. To efficiently use the total sample, this procedure could be performed multiple times using a bootstrapping technique. Here, several models are fit and the means of parameter estimates along with their standard errors are taken to be the final estimators. Using an example dataset containing final letter grades of domestic and international introductory statistics students over the course of several semesters, the aforementioned bootstrapping procedure will be demonstrated for a logistic regression using PROC SURVEYSELECT in SAS. Sampling techniques and assessing model fit will also be discussed.

(KEYWORDS: BOOTSTRAPPING REGRESSION FIT)

Bryce Whitehead is a graduate student at the University of Northern Colorado, in the Department of Applied Statistics and Research Methods, pursuing his PhD in Applied Statistics. Bryce has research interests in statistical process control and non-parametric statistics.

Dr. Austin Brown is an assistant professor at Kennesaw State University in the Department of Statistics and Analytical Sciences. Dr. Brown has research interests in statistical process control for non-normal processes, statistics in sports, statistical computing, and academic persistence in higher education. In his spare time, he enjoys hiking, spending time with family and friends, and watching sports.

SD-200

Summarizing some conventional methods to classify a binary target

Yida Bao and Philippe Gaillard

An average of about thousands of sports articles are published online every day. However, the quality of article varies, and a good article is always easy to neglected. To elicit readers' interest and give the user a better experience, World Wide Web hire experienced editors to manually classify articles as "Subjective" or "Objective". This classification procedure is tardiness, which may substantially hamper the efficiency of the website. We propose an automated way of classifying the sports articles, using several conventional methods to classify the sports articles and compare the misclassification error rate of each method. Each article has its own syntactic or semantic features, including parts of speech-grams, word level sentiment, and phrase-level sentiment. During our first step of classification, we use the SAS procedure PROC HPCLUS to explore 1000 sports article's cluster information based on these features. Later, PROC DISCRIM implements K-Nearest Neighbors and Discriminate Analysis. Also, we use SAS® Enterprise Miner to apply several machine learning methods into this case, such as Logistic regression, random forest, decision tree, and neural network.

(KEYWORDS: LABELS, CLASSIFICATION, QDA, RANDOM FOREST, KNN, NEURAL NETWORK, DECISION TREE, LOGISTIC, CLUSTER)

Yida Bao, he is a Ph.D. student from Auburn University, and he is also certified as a SAS Advanced Programmer.

Philippe Gaillard, he's Associate Professor and Director of Statistical Consulting Center from Auburn University.

SD-203

Data Visualizations using Census Shapefiles, PROC GMAP, SAS/GRAPH Animation, and BISG

Martha Pohl

The methodology to compute a proxy for missing race and ethnicity released in the summer of 2014 by the Consumer Financial Protection Bureau (CFPB) uses the surname of SMITH and the state of California as an example of this computation. This methodology uses Bayes' Theorem, the Census Surname List By Race Ethnicity from the Decennial Census, and the racial/ethnic composition of the population by geography from the Census Summary File 1 (100% sample). It is called the Bayesian Improved Surname Geocoding (BISG) proxy method. This presentation will look at the Surname of SMITH (the most frequently occurring surname in the United States) as well as other surnames at the County and Zip Code (ZCTA) levels using the data visualization technique of SAS/GRAPH animation to examine these changes by geographic location and concentration. Racial/ethnic composition of the population by county over time using the American Community Survey (ACS) yearly Estimates will be visualized nationwide. This methodology called BISG is used to compute a proxy for race/ethnicity when it is not available using the surname and geographic location. SAS/GRAPH Templates for SAS PROC GMAP using Shapefiles from the Census Bureau and SAS PROC GSLIDE for SAS animation will be provided. Disclaimer: Any opinions expressed in this paper are those of the author and do not constitute policy or opinion of the U.S. Department of Justice or any of its subcomponents.

(KEYWORDS: BISG, ANIMATION, PROC GMAP, SHAPEFILES)

Martha Dusenberry Pohl has a B.S. degree in a double major: Mathematics and Statistics from the American University in Washington, DC (AU). She also has done graduate work in Statistics at AU and is currently enrolled in the Masters of Science degree program in Data Science also at AU. She has been a Mathematical Statistician for several Federal agencies, including about 30 years at the U.S. Department of Justice. She has been a SAS user for almost all of her career.

A user-friendly and robust macro that produces a publication-ready Table 1**Daniel Brinton and Marshall Chew**

The most common first deliverable of a statistical analysis project is colloquially known as the Table 1—which compares the baseline demographics and characteristics of two or more groups of subjects. Completing a table 1 by hand can be time consuming and introduces opportunities for data entry error in an otherwise flawless analysis. Engineering out human error is a common goal of programmers, but auto-generating a Table 1 using portable code has proven to be a complex task. Typically, prior endeavors to create a table 1 macro required lots of modification by the end user to work correctly. This table 1 macro fills in the gaps by producing a publication-ready table 1 on any SAS dataset, output as a Microsoft Excel spreadsheet, along with a SAS table of the contents therein. Reported variables may be categorical or continuous. The macro reports the total sample size in each group. Categorical variables are reported by frequency and percent. Continuous variables are reported as mean \pm SD, as well as median [IQR] when not normal. Moreover, testing for normality is accomplished, with results and normality test utilized reported in a comments column. Finally, comparison of differences between groups are statistically tested using appropriate statistical tests—with p-values reported, and the name of the statistical test used reported in the comments column.

(KEYWORDS: TABLE 1, DEMOGRAPHIC ANALYSIS, MACROS)

Dr. Daniel Brinton is an Assistant Professor at the Medical University of South Carolina (MUSC) in the Department of Healthcare Leadership and Management. There he teaches Quality Improvement, Statistics, and Decision Analytics. He received his PhD in Health and Rehabilitation Science with a focus on Health Services Research in 2018 from MUSC. He is a methodologist and active health services researcher—engaged in many grant-funded research projects using large healthcare datasets. His research interests include comparative effectiveness research topics as well as the application of missing data techniques in big datasets to reduce bias.

Analyzing Airbnb reviews using SAS® Text Miner and Predicting the factors contributing for higher ratings

Ashlesha Sampathi

Airbnb, Inc. is a privately held global company headquartered in San Francisco that operates an online marketplace and hospitality service that is accessible via its website and mobile apps. Members can use the service to arrange or offer lodging, primarily homestays, or tourism experiences. It is the world's largest home sharing company and has over 4 million listings in more than 81,000 cities and 191 countries. Airbnb projects the prospect of making money by renting out our home with the platform. But homeowners, especially those renting out their homes for the first time, may have many questions: What price should I set my home at? Can I trust my home to guests? How can I ensure I get a good rating? Customer reviews and ratings play an important role in boosting a customer loyalty towards a brand. In fact, they can make or break a business. Therefore, it is very important for businesses to analyze the factors that are driving higher ratings and it is also equally important to have an overview of public opinion on the product. In this paper, I have predicted the main factors that are driving higher ratings and analyzed the reviews from customers using SAS® text miner and compared them with numerical ratings to analyze the correlation between written reviews and numerical ratings. In addition, I've also performed descriptive analysis to explore some key points which would be very helpful for business such as:

1. What are the most popular times of the year for Airbnb rentals in Seattle?
2. Which locations in Seattle are most valued according to Airbnb customers?

An open dataset from inside Airbnb website was used for the research purpose. The dataset provides information on home features, review scores, comments and the availability of 8,460 listings in Seattle for the year 2019. SAS® Viya was used to conduct visual analytics on the Airbnb data and SAS® Studio to perform linear regression to predict the factors driving higher ratings. In addition, SAS® Text Miner was used for text mining of customer reviews.

(KEYWORDS: MACHINE LEARNING,PREDICTIVE MODELLING,SENTIMENTAL ANALYSIS, STATISTICS AND DATA ANALYSIS)

Ashlesha is a graduate student in Business Analytics from Oklahoma State University, Stillwater. She has recently completed her Summer internship as a Data science and Advanced Analytics Intern at Chesapeake Energy. She also has 2 years of work experience as a BI Analyst at Accenture Solutions pvt. Ltd., India. Ashlesha has extensive hands-on experience with analyzing large datasets and building predictive models by implementing various machine learning algorithms. She has been actively using SAS for coursework, research and project related activities.

The Thorn in My Side!! Logistic Regression Continuous Variables that Violate the Assumption of Linearity on the Log-odds (Logit) scale: How to Identify and What to Do?

Janet Grubber and Cynthia Coffman

I find myself crossing my fingers when I check continuous variables that are to be included in logistic regression models for linearity on the log-odds scale“Please, please...let them be linear”! Recently, my pleading did not work for a variable that was of great interest to the study team. Though using the Box-Tidwell method for checking for linearity seems to have fallen out of favor, I used it anyway. It did identify that the variable was not linear on the log-odds scale – but the best next steps were unclear; and even the question of whether Box-Tidwell correctly identified the lack of linearity was unclear. Furthermore, even if the Box-Tidwell result was correct – what good was this knowledge when I needed to understand the shape of the non-linearity to appropriately use the variable in the model and understand its association with the dependent variable. This talk works through the sequence of steps that were ultimately used via SAS software to understand the nature of the non-linear relationship of the independent variable of interest with the dependent variable in a logistic regression model:

1. How to identify lack of linearity on the log-odds scale (Box-Tidwell, use of %PSPLINET macro (Frank Harrell) to plot the association, use of restricted cubic splines in PROC LOGISTIC procedure)
2. How to handle the continuous variable in the logistic regression model once the lack of linearity is detected (creation of multiple “dummy-like” continuous variables to represent the independent variable of interest, use of restricted cubic splines in PROC LOGISTIC procedure)

(KEYWORDS: LOGISTIC REGRESSION, LOG-LINEARITY, ASSUMPTION, CONTINUOUS VARIABLES, PLOTTING, CUBIC SPLINES)

Janet is an epidemiologist, working as a statistician, at the Durham Veterans Affairs Medical Center in Durham, North Carolina. She has been using SAS for 27 years. Her route to this career has not been straightforward! Janet worked as a personnel management specialist for the Navy prior to extracting DNA from approximately 3,000 mouse tails while working as a lab technician at what is now the National Cancer Institute’s Mouse Cancer Genetics Program. After the post-doc she was working with advised her that molecular epidemiology would be a good career for somebody (like her!) looking to combine her interests in biology, genetics, and people, she went to University of South Carolina, leaving with a Master’s of Science in Public Health in cancer epidemiology. Janet then worked as a breast and cervical cancer epidemiologist at the state health department in Delaware, followed by work as a statistician at the Center for Human Genetics and Cancer Center at the Duke University Medical Center for 12 years. She’s been at the Durham VA for 13 years, puzzling all the while about how to deal with independent continuous variables that are not linearly associated with the logit of binary dependent variables. She’s looking forward to sharing with you some of her painful experience in hopes of making yours a little easier!

Niloofar Ramezani

Machine learning, big data, and high dimensional data are very important topics of recent years which some consider as the wave of the future. Therefore, it is crucial to learn about appropriate statistical techniques to be applied within these relatively new topics. Since statistical approaches have been established for many years and their efficiency have already been evaluated, they can benefit newer data-related fields. Machine learning is an important topic these days as it involves a set of many different methods and algorithms that are suited to answer diverse research questions. Thus, choosing a proper algorithm is a critical step in the machine learning process to ensure it truly fits the solution proposed in answering a problem at hand (Segal, 2004). To better understand and select machine learning algorithms when dealing with real data, it is helpful to understand them within the framework of statistics and divide them into two main groups based on the nature of data and type of outcome variables. Depending on the nature of variables, classification and regression machine learning methods are discussed here and appropriate statistical techniques for modeling them are presented in this study. Within this paper, these two types are differentiated and related algorithms and statistical techniques, as well as the SAS procedures, are discussed in order to answer real world problems including high dimensional data scenarios. These procedures include, but are not limited to, PROC HPFOREST, PROC FACTOR, PROC CLUSTER, PROC LOGISTIC, PROC GENMOD, PROC NL MIXED, PROC CATMOD and PROC STANDARD.

(KEYWORDS: MACHINE LEARNING, STATISTICAL MODELS, BIG DATA, RANDOM FOREST, LOGISTIC MODELS)

Niloofar Ramezani, Ph.D., is an Assistant Professor of Statistics at George Mason University. She is an experience statistician and biostatistician who has served as a co-investigator for NIH-funded studies focused on teaching, motivation and addiction behavior, and mental health interventions. She has held multiple professional positions as a biostatistician, senior statistician, data scientist, methodologist, and senior statistical consultant working with multiple private and state agencies for over 10 years. She teaches graduate and undergraduate classes in statistics and data analytics and enjoys using SAS and R in her teaching and research. She has been a two-time recipient of SAS ambassadorship award as well as junior faculty scholarship through SASGF 2018. She also has been the seminar coordinator of all conference seminars at the Western Users of SAS Software conference 2017 held in Long Beach, California.

SD-243

Improving Employee Satisfaction Through Text Analytics and Predictive Modeling
Luis Martinez and Miriam McGaugh

The success of any business is heavily influenced by the productivity of its employees, making employees one of the most important assets that a company holds. One factor that is linked with employee productivity is work satisfaction. In other words, how well an employee works is partly determined by how happy he or she is with their job. Therefore, knowing the positive and negative feelings that employees have toward their employer can be useful for making better human resource decisions. One source of data on employee's opinions about their employer and work environment is online reviews. In this project, we take a collection of online reviews from a popular job search website and analyze it to gain insights into how employees feel about their company. The data set includes a pros and cons comment sections as well as a score from 1-5 on the overall opinion of the company. Thousands of reviews were collected on six popular technology companies. Online reviews have been analyzed descriptively in other research to determine what employees most commonly mention when reviewing their employer. Our own preliminary descriptive analysis showed a statistically significant difference among the six technology companies in relation to overall scores. This research paper, however, takes the analysis of these scores further by creating a predictive rule-based model that uses the text portions of the review as variables to classify comments into positive or negative. By interpreting the model, we can determine what factors may be most significant for causing high or low employee satisfaction. These factors can then be used by companies in taking appropriate steps to improve their employee sentiment and in turn, their overall productivity. The tools used to perform this analysis include SAS® Enterprise Miner™ and SAS® Studio. This research can also be easily expanded to include other companies and improved upon by adding more training data from other job sites.

(KEYWORDS: TEXT ANALYTICS, HUMAN RESOURCES, ONLINE REVIEW, PREDICTIVE MODELING)

SD-251

Validating a Probability of Default model using Enterprise Miner Scorecard
George Rezek

Despite the prevalence of Enterprise Miner (EM) with the scorecard Add-In, Probability of Default (PD) models are more often developed and validated using products outside of EM. While there is no shortage of demos pointing out features of using EM, this presentation focuses on stumbling blocks that are obvious to those familiar with the product, but often cause potential users to give up on EM citing time considerations. The Interactive Node is often mentioned as reason enough to use EM, because of the ease it brings to the binning process. Relevant statistical output, graphs and data are automatically generated—some are easier to find than others; code nodes allow interaction with base SAS and entire EM diagrams can be run as simulations.

(KEYWORDS: ENTERPRISE MINER SCORECARD MODEL VALIDATION PROBABILITY OF DEFAULT)

Market Basket Analysis on Instacart**Aravind Dhanabal and Balamurugan Mohan**

In the process of online shopping, you have probably seen a section called “suggestions for you” or “customers who bought this item also bought” in which Market Basket Analysis plays an important role. Market Basket Analysis is a technique used by retailers to understand customer behavior while purchasing from their stores. Instacart is an e-commerce website that allows users to shop for groceries from a local grocery store online, and then sends an Instacart personal shopper to pick up and deliver the orders made by users the same day. These processes allow retailers to conduct analysis on purchase iterations by users but understanding the customer purchasing patterns and behaviors can become tedious and challenging. SAS Viya, Tableau, SAS Enterprise Miner and Python were used for the initial analysis and model creation. The datasets provided by Instacart had complete information of over 3 million grocery orders from more than 200,000 Instacart users. Both product data and customer data from Instacart includes 50,000 unique products, week and the time of purchase, different product aisle and departments. Understanding the data, dairy products, fruits and vegetables were purchased the most across all the departments and people tends to purchase and reorders 60% of their previous orders mostly on Sunday and Monday. In this paper, my goal is to examine two major sections. The first section focuses on descriptive analysis on the customer purchase patterns, items purchased together and units that are highly purchased from the store to facilitate reordering and maintaining adequate product stock. In the second section, my goal is to identify the clusters and subgroups of customers possessing similar purchase behavior and to visualize the data to provide productive recommendations which focus on improving the revenue and customer experience through segmentation and prediction models. This paper will enable Instacart to enhance the user experience by suggesting the next likely product to purchase to the customer during the order process. Further, this paper will outline a marketing strategy for Instacart and similar retailers including sending personalized communications to customers reminding them to order again, by highlighting the predicted products in those communications.

(KEYWORDS: MARKET BASKET ANALYSIS, SAS MINER, PYTHON, TABLEAU, CLUSTER ANALYSIS, PREDICTION MODELS)

SD-271

An Introduction to Causal Effect Estimation with Examples Using SAS Software

Yiu-Fai Yung

How can you estimate a causal effect from nonrandomized data? As statisticians and data scientists are increasingly tasked with analyzing data that come from observational studies rather than randomized experiments, this is a question of increasing importance. This tutorial provides an overview of methods for estimating causal effects for dichotomous treatments. In particular, it illustrates causal effect estimation by propensity-score-based matching, inverse probability weighting, and doubly robust methods by using examples relevant to the biological and life sciences. The analyses are performed using the PSMATCH and CAUSALTRT procedures in SAS/ STAT® software. Also briefly discussed are approaches for constructing and evaluating the underlying models, comparisons of the estimation methods, and the assumptions required for identifying and estimating treatment effects.

(KEYWORDS: CAUSAL EFFECT, OBSERVATIONAL STUDIES, MODELING, SAS/STAT, PROPENSITY SCORES)

Yiu-Fai Yung is a senior manager in Advanced Analytics R&D at SAS Institute Inc. and has been developing statistical software for causal analysis and structural equation modeling. He obtained his PhD in psychology and MA degree in math from UCLA. He has led several structural equation modeling workshops at conferences such as SAS Users' Group meetings and Joint Statistical Meetings. Before joining SAS, he taught psychological statistics at the University of North Carolina at Chapel Hill. He has published research articles in Psychometrika, the British Journal of Mathematical and Statistical Psychology, and other Journals.

SD-296

Regression Models for Count Data

Jason Brinkley

Outcomes in the form of counts are becoming an increasingly popular metric in a wide variety of fields. For example, studying the number of hospital, emergency room, or in-patient doctor's office visits has been a major focal point for many recent health studies. Many investigators want to know the impact of many different variables on these counts and help describe ways in which interventions or therapies might bring those numbers down. Traditional least squares regression was the primary mechanism for studying this type of data for decades. However, alternative methods were developed some time ago that are far superior for dealing with this type of data. The focus of this paper is to illustrate how count regression models can outperform traditional methods while utilizing the data in a more appropriate manner. Poisson Regression and Negative Binomial Regression are popular techniques when the data are overdispersed and using Zero-Inflated techniques for data with many more zeroes than is expected under traditional count regression models. These examples are applied to studies with real data.

(KEYWORDS: POISSON REGRESSION, NEGATIVE BINOMIAL REGRESSION, MODELING COUNTS)

Jason S. Brinkley, PhD, MS, MA is a Senior Researcher and Biostatistician at Abt Associates Inc. where he works on a wide variety of data for health services, policy, and disparities research. He maintains a research affiliation with the North Carolina Agromedicine Institute and serves on the executive committee for the NC Chapter of the American Statistical Association and the Southeast SAS Users Group. Follow him on Twitter, @DrJasonBrinkley.



Want a résumé that outshines the rest?

Add analytics skills. Learn SAS® today.

Learn more
support.sas.com/learn





HELPING TO KEEP YOUR SAS® PLATFORM SECURE

We help you manage, review, troubleshoot, document, test and audit your SAS® platform installation with confidence.



@metacoda #SESUG2019 #SASadmin



MS IN BUSINESS ANALYTICS AND DATA SCIENCE*

A PROGRAM TO PREPARE STUDENTS TO MEET INDUSTRY NEEDS

The Spears School of Business developed one of the first analytics and data mining programs in the nation more than ten years ago, with two distinct tracks built on our strengths and diversity of talents. Over time, we have developed strong relationships with major analytics providers such as SAS, Teradata, IBM, and others.

The Oklahoma State University **Master in Business Analytics and Data Science (MS BANDS)*** employs multi platform experiential training in analytics leveraging our partnership with analytics vendors along with an integrated core set of data science experiences. In addition, students can specialize in one of four option areas or craft their own unique option from a wide variety of elective courses.

- **37 total hours** (22 hours core; 15 hours electives)
- Options include **Marketing Analytics, Advanced Data Science, Health Analytics, Cybersecurity Analytics***
- **STEM-designated** program
- Ability to earn **multiple SAS certificates**

"Business Analytics program at OKState is perfect for someone who wants to get into the field of Analytics & Data Science. I worked as a System Engineer at IBM before joining MSBAN in 2015. This program offered a wide range of electives and prepared me for a career in the industry of my choice. An amalgamation of course work in Statistics, Predictive Modeling, Marketing and Finance provided me with a stepping stone to launch my career as a Sr. Consultant at EY."

– Anirban Chakraborty, Class of 2015, Sr. Consultant EY



"While working full-time and earning my MS in Business Analytics degree, I was able to see connections between the technical, advanced analytics content from the MSBAN program and the problems I was working to solve at my job. Not only was the variety of content interesting to me personally, I learned the newest techniques and strategies in the advanced analytics realm. This program has equipped me with a competitive advantage and I will take the concepts I've learned from this degree throughout my entire career journey."

– Holly Hopkin, Distance Class of 2018, Business Analyst - Founders Brewing Company

FOR MORE INFORMATION PLEASE CONTACT:



Goutam Chakraborty, Ph.D.
SAS Professorship in Marketing
Analytics, School of Marketing and
International Business

goutam.chakraborty@okstate.edu
405-744-7644



Sarah Chabinak, M.S. Ed.
Senior Academic Counselor II
Watson Graduate School of
Management

sarah.chabinak@okstate.edu
405-744-2951



**SPEARS SCHOOL
OF BUSINESS**

*Pending approval from the OSU/A&M Regents and Oklahoma State Regents for Higher Education.



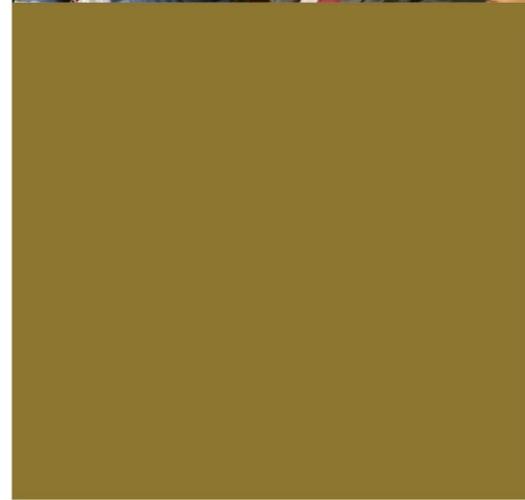
NOW ACCEPTING FALL 2020 APPLICATIONS

THE COMPETITIVE ADVANTAGE DEGREE

[HTTPS://MAYS.TAMU.EDU/MS-ANALYTICS/](https://mays.tamu.edu/ms-analytics/)
AVAILABLE FACE-TO-FACE AT HOUSTON CITYCENTR AND LIVE VIDEO STREAM



MS Analytics
MAYS BUSINESS SCHOOL





Pinnacle Solutions, Inc. ("Pinnacle Solutions") has been providing data analytics services to clients in the education, financial, health services and manufacturing sectors since 1996. While Indianapolis based, we service clients all over the US including Hawaii and have recently expanded into Canada. Our primary objective is to help clients makes sense of the data available to them utilizing data management, data mining, statistical analysis, forecasting, and ultimately visualizations. We are familiar with accessing and processing data in it various states from flat files, to databases, to big data solutions and to streaming data from sensors (IOT). Having access to both SAS and AWS analytics options allows us to meet the client where they are at in their analytics journey and move them forward. Our clients range from Fortune-level global enterprises to small businesses.



Operational Excellence and Assessment Support

UNIVERSITY OF CENTRAL FLORIDA

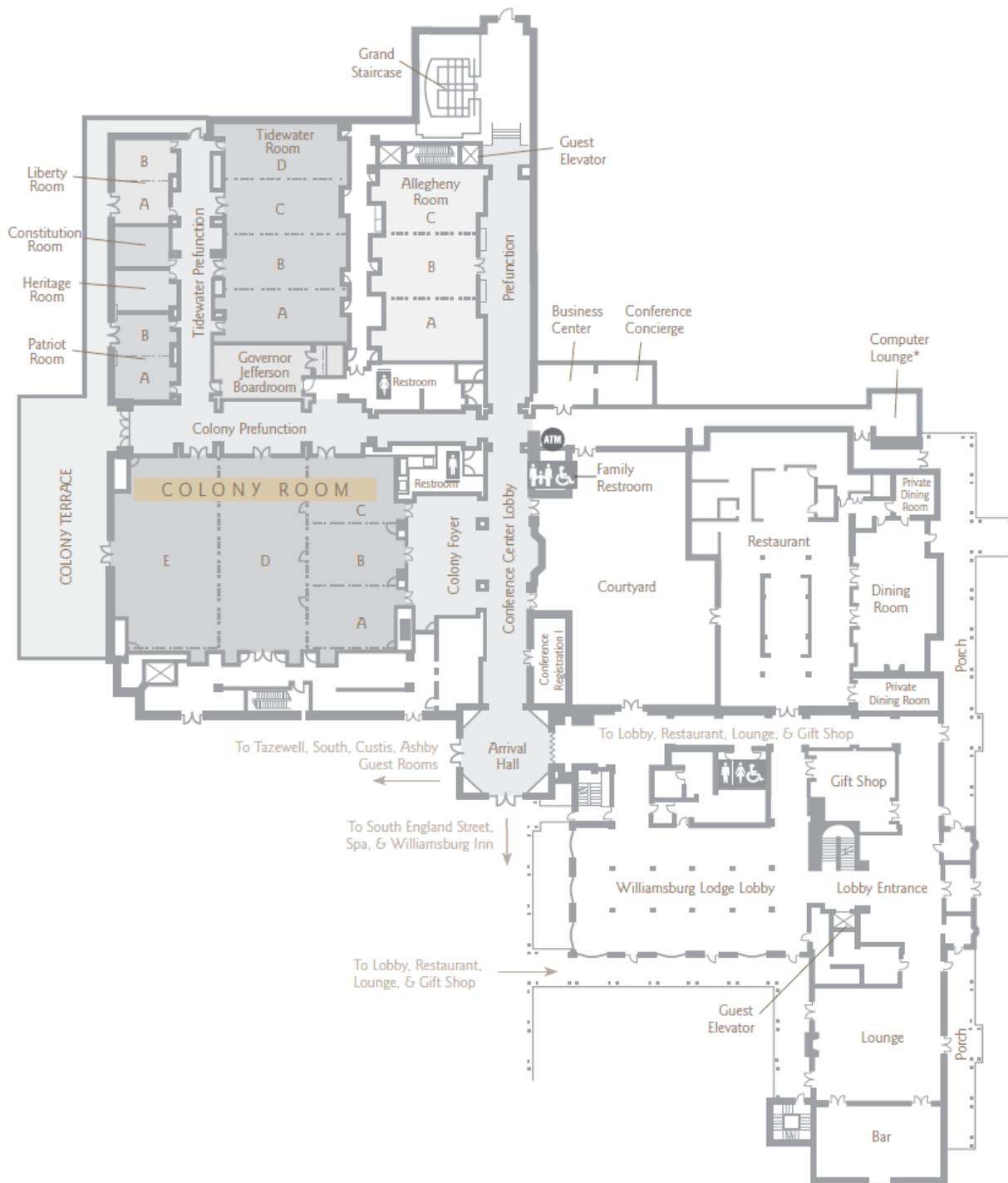
The mission of the office of Operational Excellence and Assessment Support (OEAS) at the University of Central Florida is to support efforts to improve the quality of student learning outcomes and the effectiveness and efficiency of university operations through assessment and analytics. OEAS will accomplish this by providing support to all academic programs and administrative units through integrated processes that include continuous quality improvement, analytical and survey studies, technology integration and guidance in assessment.



Experis Solutions delivers professional services that drive innovation and growth. We serve clients, including Fortune 500 and Global 1000 companies, from offices worldwide, including our delivery centers in the U.S. and abroad. Experis Solutions is a division of ManpowerGroup (NYSE: MAN).

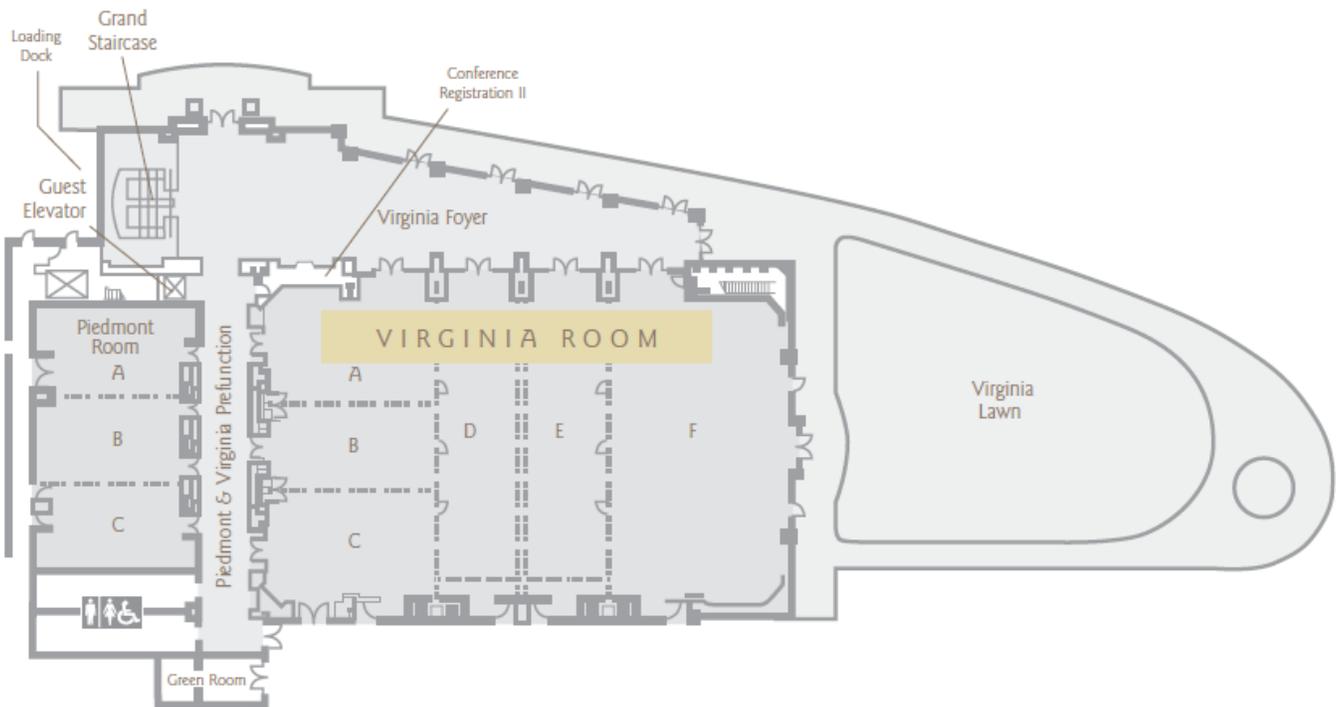
MEETING ROOM LAYOUT

WILLIAMSBURG LODGE CONFERENCE CENTER MAIN LEVEL



Post Conference: Downloadable zip file of conference papers available at www.sesug.org/SESUG2019

WILLIAMSBURG LODGE CONFERENCE CENTER LOWER LEVEL





SESUG 2020

Savannah, GA - October 25-27

Don't be afraid to save the date!!

Let your knowledge flow like the famous fountains of Savannah ... Share your expertise!



Branch out and extend your roots within the SAS community

Let seasoned experts guide you through the turbulent waters of modern analytics



Take time to enjoy the beauty and historic charm of Savannah

www.sesug.org/SESUG2020

Richann Watson
Academic Chair

AcademicChair2020@sesug.org



Lesla Caves
Operations Chair

OperationsChair2020@sesug.org