SESUG Speaker Sharing Program

Presentation Titles, Abstracts and Author Bios

Program Goals:

- Speaker sharing is one of the many ways SESUG supports local and in-house SAS user groups in the SouthEast region. When available, SAS will also provide a speaker to the NorthEast region
- SESUG provides SAS user groups a yearly presentation selected from our list of topics and delivered virtually by our SESUG EC featured speakers.
- SESUG speakers in the local area are also available for in-person presentations. SAS User Groups can also choose to fund an out of the area presenter for an in-person presentation.

Program Specifics:

- Local and in-house SAS user groups in the SESUG region (Alabama, Florida, Georgia, Kentucky, Maryland, Mississippi, North Carolina, South Carolina, Tennessee, Virginia, West Virginia, Washington DC, Puerto Rico) are eligible to participate.
- For organizations with locations in multiple regions, SESUG also supports entities with a presence (SAS users) in this region.
- User groups in the former NESUG region are also eligible to apply.
- Sponsorship is limited to one virtual speaker each calendar year. Selected speakers can be expected to provide more than one presentation at a single user group meeting if requested.
- SESUG fulfills local and in-house SAS user group requests for speakers based upon speaker availability.
- In person speakers are available if in the local area or at the expense of the SAS User Group.
- In order to accommodate multiple user group requests for a speaker, SESUG may request use of an alternate speaker.

Requesting a Speaker:

- Please plan as far in advance as possible so that SESUG can accommodate your request.
- After reviewing the available topics, speakers, and speaker bios **BELOW**, select your "top three choices" of speakers, and email your request to Abbas Tavakoli. Please include:
    - Desired **date** / **time** / **location** of meeting
    - User Group name and contact information
    - Virtual connection information
    - **Three** preferred speakers, and the presentation titles of interest.
- SESUG will review your request and, if fulfilled, SESUG will provide confirmation of:

- Speaker / title(s)
- Date

NOTE: Requests are fulfilled based on speaker availability.

Alphabetic List of Presentation Titles and Presenting Authors

A Beginner's Babblefish: Basic Skills for Translation Between R and SAS®; Sarah Woodruff

A Macro to Unravel Macros; Sarah Woodruff

An Introduction to SAS® Arrays; Andrew T. Kuligowski

Animate your Data!; Richann Watson

Array, Hurray, Array; Consolidate or Expand Your Input Data Stream Using Arrays; William E. Benjamin Jr.

Automated Validation of Complex Clinical Trials Made Easy; Richann Watson

Avoiding Sinkholes: Common Mistakes During ADaM Data Set Implementation; Richann Watson

Backsplash Patterns for Your World: A Look at SAS® OpenStreetMap (OSM) Tile Servers; Barbara B. Okerson

Check Please: An Automated Approach to Log Checking; Richann Watson

Compare MIXED and GLMMIX to Analyze Breast Cancer Longitudinal Study**;** Abbas S. Tavakoli

Considerations in ADaM Occurrence Data: Handling Crossover Records for Non-Typical Analysis; Richann Watson

Creating Zip Code Level Maps With SAS®; Barbara B. Okerson

Data in the Doughnut Hole: Using SAS® to Report on What is NOT There; Sarah Woodruff

Data Management Plans: Stages, Components, and Activities; Abbas S. Tavakoli

Developing an Analytics Center of Excellence; Chuck Kincaid

Easier than You Think: Creating Maps with SAS® Enterprise Guide®; Stephanie Thompson

Easy Ways to get SAS® Enterprise Guide® to Send Tables and Graphs to Microsoft Office Suite, Without Writing Code; William E. Benjamin Jr.

Evaluating Sample Code for an Interview; Stephanie Thompson

Examining Mediator and indirect effects of Loneliness in Social Support on Social Well-Being Using Baron and Kenny and a Bootstrapping Method Data Management Plans: Stages, Components, and Activities Compare MIXED and GLMMIX to Analyze Breast Cancer Longitudinal Study; Abbas S. Tavakoli,

Exploring HASH Tables vs. SORT/DATA Step vs. PROC SQL; Richann Watson

Extend the Power of SAS® to Use Callable VBS and VBA Code Files Stored in External Libraries to Control Excel Formatting Routines; William E. Benjamin Jr.

Flat Pack Data: Converting and ZIPping SAS® Data for Delivery; Sarah Woodruff

Four Thousand Reports Three Ways; Stephanie Thompson

FREQ Out – Exploring Your Data the Old School Way; Stephanie Thompson

Generating Custom Shape Files for Data Visualization - note this is a Quick Tip and can be done with another choice; Stephanie Thompson

Getting the Most Out of Your SAS® Conference; Andrew T. Kuligowski

Getting Your Data into SAS®; Stephanie Thompson

Give the Power of SAS® to Excel Users Without Making Them Write SAS Code as a "Take-out"; William E. Benjamin Jr.

Great Time to Learn; Richann Watson

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

Guidelines for Selecting the Covariance Structure in Mixed Model Analysis; Chuck Kincaid

How to Succeed in Consulting Without Really Trying; Chuck Kincaid

How to be a Data Scientist with SAS®; Chuck Kincaid

How to HOW: Hands-on-Workshops Made Easy; Chuck Kincaid

I'm a SAS® Programmer. Why Should I JMP®?; Barbara B. Okerson

In Search of the LOST CARD; Andrew T. Kuligowski

Intermediate SAS® Macro Programming; Chuck Kincaid

Interactions between the DATA Step and External Files – INFILE / INPUT and more; Andrew T. Kuligowski

Introduction to Data Simulation; Jason Brinkley

Introduction to Hierarchical Linear Modeling (HLM); Jason Brinkley

Intro to Proc Report; Barbara B. Okerson

Intro to Proc Report with PDF Output; Barbara B. Okerson

Keeping Up Appearances: Turning Specifications into SAS® Format Libraries and Statements; Sarah Woodruff

Leave Your Bad Code Behind: 50 Ways to Make Your SAS® Code Execute More Efficiently; William E. Benjamin Jr.

Let SAS® Do Your DIRty Work; Richann Watson

Let's Get FREQy with our Statistics: Data-Driven Approach to Determining Appropriate Test Statistic; Richann Watson; Lynn Mullins

Looking Beneath the Surface of Sorting; Andrew T. Kuligowski

Managing Your SAS® Assets; Stephanie Thompson

Messy Data in SAS®; Jason Brinkley

More than Models: The Data Mining Team; Stephanie Thompson

Navigating Workforce Trends in the Human Age; Chuck Kincaid

New to SAS® and New to Programming? What You Need to Do Before Typing Code; Stephanie Thompson

ODS Graphics; Chuck Kincaid

Parsing Useful Data Out of Unusual Formats Using SAS®; Andrew T. Kuligowski

Perusing, Choosing, and Not Mis-using: Non-parametric vs Parametric Tests in SAS®; Venita DePuy

PROC SURVEY… Says!: Selecting and Analyzing Stratified Samples; Darryl Putnam

Predicting Analytics: Where will Analytics be in 20 years?; Chuck Kincaid

Pruning the SASLOG – Digging into the Roots of NOTEs, WARNINGs, and ERRORs; Andrew T. Kuligowski

Regression Models for Count Data; Jason Brinkley

SAS® Users Can Command Microsoft Excel to Automatically Create Graphs from SAS ExcelXP Output; William E. Benjamin Jr.

Smoke and Mirrors!!! Come See How the _INFILE_ Automatic Variable and SHAREBUFFERS Infile Option Can Speed Up Your Flat File Text-Processing Throughput Speed; William E. Benjamin Jr.

So, You've Got Data Enterprise Wide (SAS®, ACCESS, EXCEL, MySQL, Oracle, and Others); Well, Let SAS Enterprise Guide Software Point-n-Click Your Way to Using It; William E. Benjamin Jr.

Standardized, Customized or Both? Defining and Implementing (MedDRA) Queries in ADaM Data Sets; Richann Watson

Talking Past Each Other? How to Communicate with Medical Writers When Preparing Clinical Research Manuscripts for Journal Submission; Stephanie Thompson

Telling the Story Better: Advanced ODS Graphics; Chuck Kincaid

The Building Blocks of SAS® Datasets: S-M-U (Set, Merge, and Update); Andrew T. Kuligowski

The Care and Feeding of Magical Creatures: Managing Statisticians and Data Scientists in an Analytically Enthused World; Chuck Kincaid

The Little Engine That Could: Using LIBNAME Engine Options to Enhance Data Transfers Between SAS® and Microsoft Excel Files; William E. Benjamin Jr.

Time Series Mapping with SAS®: Visualizing Geographic Change over Time in the Health Insurance Industry; Barbara B. Okerson

Tips and Tricks for Introductory Workshops in SAS® for Health Professionals; Jason Brinkley

Tricks and Tips for Using the Bootstrap in JMP® Pro;Jason Brinkley

Using a Picture Format to Create Visit Windows; Richann Watson

Using Big Data to Visualize People Movement Using SAS® Basics; Stephanie Thompson

Using GENMOD to Analyze Correlated Data on Military Health System Beneficiaries Receiving Inpatient Behavioral Health Care in South Carolina Health Care Systems; Abbas S. Tavakoli

Using GLIMMIX and GENMOD Procedures to Analyze Longitudinal Data from a Department of Veterans Affairs Multisite Randomized Controlled Trial; Abbas S. Tavakoli

Using INFILE and INPUT Statements to Introduce External Data into the SAS® System; Andrew T. Kuligowski

Using JMP® to apply Decision Trees and Random Forests as Screening Tools for Limiting Candidate Predictors in Regression Models; Jason Brinkley

Using SAS® to Examine Mediator, Direct and Indirect Effects of Isolation and Fear on Social Support Using Baron& Kenny Combined with Bootstrapping Methods; Abbas S. Tavakoli

'V' for … Variable Information Functions to the Rescue; Richann Watson

When ANY Function Will Just NOT Do; Richann Watson

Where Should I Dig? What to do Before Mining Your Data; Stephanie Thompson

Windows and Unix Computers Now Have Multiple CPU's; Why Not Control Two or Three or More Parallel Executing SAS® Batch Jobs from One Master Job!; William E. Benjamin Jr.

Working with Character Data; Andrew T. Kuligowski

Author Bios

William E. Benjamin Jr.
William E. Benjamin, Jr. his expertise includes Base SAS® Software, and SAS Macros. William has a BS degree in computer science from Arizona State University and an MBA from Western International University. He has been a SAS software user since 1983 and a computer programmer since 1973. His programming experience spans from vacuum tube mainframes, to current PC computers. William currently owns a consulting company called OWL Computer Consultancy, LLC in Phoenix AZ. His new SAS Press book titled "Exchanging Data between SAS and Microsoft Excel: Tips and Techniques to Transfer and Manage Data More Efficiently" is now available.

Jason Brinkley
Dr. Jason Brinkley is a Senior Researcher at the American Institutes for Research health division out of Chapel Hill, NC. He maintains an Adjunct Professor of Public Health and is an affiliate of the NC Agromedicine Institute at East Carolina University where he advises on data collected on agriculture worker health (including migrant populations). Brinkley spent almost 8 years on the faculty of the department of Biostatistics at East Carolina and has worked with a variety of individuals in the medicine, public health, business, education, and the natural and social sciences. He has expertise in a wide range of statistical and biostatistical methods along with basic psychometric methods and customized data visualizations. He currently serves as NC Chapter Representative for the American Statistical Association and on the executive board for the Southeast SAS Users Group. In his role at AIR, Dr. Brinkley leads and participates in research on a variety of health services research projects including evaluation, quality improvement, access, quality of care, performance management, and health care disparities.

Venita DePuy
Venita is an experienced statistician and SAS programmer, and owner of Bowden Analytics. She specializes in providing statistical and programing services to pharmaceutical companies, CROs, and academic organizations, with an emphasis on clinical trials and CDISC. Her therapeutic experience includes cardiology, oncology, depression, veterinary medicine. She earned her Bachelor of Science in Statistics from the University of Alaska Fairbanks, then continued on to earn her Master of Statistics at North Carolina State University and PhD in Biomathematics with a Statistics minor at NCSU as well.

Chuck Kincaid
Chuck is a Principal Data Scientist at Experis with the Business Intelligence and Analytics Practice. His Computer Science and Statistics degrees are from Kansas State University and PhD work in Statistics at the University of Florida. Previously, he was the Senior Engagement Director for the Business Intelligence and Analytics Practice. Since 2002 he managed an eclectic analytics team in their Kalamazoo Center of Excellence, providing BI and Analytics services to almost every industry since the early 90’s. Chuck often speaks at events like this one, including the International Workshop on “Advanced Analytics and Data Science” in Warsaw, Poland, as well as giving the keynote presentations for SESUG 2014 and 2016. All were awesome experiences for which he is humbly grateful. His interests are in analytics talent, data science, analytics infrastructure, education and visualization.

Andrew T. Kuligowski
Andrew T. Kuligowski has been a SAS user for well over 25 years, and is currently a Manager for HSN in St Petersburg, Florida. He has augmented his professional coding experiences in the retail, media, insurance, and petrochemical fields by speaking at various SAS conferences and user group meetings. Andrew was conference chair for SAS Global Forum 2012 in Orlando, Florida, and co-chair for SESUG'97 in Jacksonville, Florida and Tennessee SAS Users Day in Knoxville, Tennessee. In his spare time, Andrew can also be found volunteering at the Florida Aquarium in Tampa.

Barbara B. Okerson
Barbara B. Okerson, Ph.D. is an independent SAS contractor, especially for the health services industry and for historical data applications. She has worked in the healthcare industry for over 20 years and has been a SAS user for over 25 years. She serves in leadership roles at SESUG and has been active in both in-house and local SAS user groups... She is a SAS Certified Professional, a Certified Professional in Healthcare Quality, and a Fellow, Academy for Healthcare Management.

Darryl Putnam
Darryl Putnam has 15 years of experience designing and implementing statistical and data analyses systems to solve business problems in a wide variety of industries. Mr. Putnam has specialized in designing and implementing the SAS software solution suite to build information decision support systems for his clients. These solutions range from statistical quality control systems, geo-demographic mapping systems, behavioral modeling systems, and enterprise wide production systems. He has a B.S. and M.A. in Economics from Virginia Commonwealth University. He is always looking for the next problem to solve.

Abbas Tavakoli
Dr. Tavakoli received a Bachelor of Science in Animal Husbandry from Tehran University, Tehran, Iran 1985, a Master of Public Health in Biostatistics at the University of South Carolina in 1989, a Doctor of Public Health in Biostatistics at the University of South Carolina in 1998, and a Master of Engineering in Computer Engineering at the University of South Carolina in 2003. Dr. Tavakoli teaches statistics courses for USC College of Nursing DNP and PhD students. Dr. Tavakoli has been involved with many research projects and manuscripts during his tenure at the College. His research interests are health related topics and he has considerable experience in data management, linear modeling, logistic regression, and experimental design.

Stephanie Thompson
Over twenty years of experience in applying statistical and modeling techniques to solve business problems in various commercial and academic environments. Strong understanding of data structures, a variety of analytical tools, and operating environments. Views problems (opportunities?) in a broad context by examining the interrelations between issues and the local and broader operational framework. Demonstrated skill at effectively communicating and working across multiple functional areas and at all organizational levels. Stephanie has made dozens of presentations at local, regional, and international meetings and conferences to technical and non- technical audiences.

Richann Watson
Richann Watson is an independent statistical programmer and CDISC consultant. She has been using SAS since 1996. She is also a member of the CDISC ADaM team and various sub-teams. In addition, she is the chairperson for the local SAS user group in her area and is actively involved with SAS Global Forum, PharmaSUG and other SAS User Groups.

Sarah Woodruff
Sarah Woodruff has been programming in SAS professionally since 2006. She works on reporting and analysis for Westat's Health, Education and Transportation divisions. She served as the Operations Chair for SESUG 2012 and has served on the SESUG executive council as well as section chaired for several years. Her undergraduate work includes a BS in mathematics and statistics from Georgia State University and a BS in microbiology from University of Maryland. Sarah also holds an MS in bioinformatics.

Presentation Abstracts (in alphabetic order by presentation title)

A Beginner's Babblefish: Basic Skills for Translation Between R and SAS®
Sarah Woodruff

SAS professionals invest time and energy in improving their fluency with the broad range of capabilities SAS software has to offer. However, the computer programming field is not limited to SAS alone and it behooves the professional to be well rounded in his or her skill sets. One of the most interesting contenders in the field of analytics is the open source R software. Due to its range of applications and the fact that it is free, more organizations are considering how to incorporate it into their operations and many people are already seeing its use incorporated into project requirements. As such, it is now common to need to move code between R and SAS, a process which is not inherently seamless.

This paper serves as a basic tutorial on some of the most critical functions in R and shows their parallel in SAS to aid in the translation process between the two software packages. A brief history of R is covered followed by information on the basic structure and syntax of the language. This is followed by the foundational skill involved in importing data and establishing R data sets. Next, some common reporting and graphing strategies are explored with additional coverage on creating data sets that can be saved, as well as how to export files in various formats. By having the R and SAS code together in the same place, this tutorial serves as a reference that a beginner can follow to gain confidence and familiarity when moving between the two.

A Macro to Unravel Macros
Sarah Woodruff

Do you find it tedious to pinpoint problems in your macro code? Are you annoyed when the log points to the line number of the macro call rather than the code in the macro that generated the error? Would you like a little help? You want to produce macros that are utilitarian black boxes, but you need to understand what the code is actually doing as part of the development process, and that can be frustrating.

This paper presents a simple wrapper macro that is used to enclose existing code, which can call any number of other macros. When you use this technique along with the mPrint and mFile options, the output file produced bymPrint/mFile contains a completely "de-macroed" version of the program. You can then examine this output as a working program, thus making it easier to both understand and edit.

An Introduction to SAS® Arrays
Andrew T. Kuligowski

So, you've heard about SAS® arrays, but are not sure when – or why – you would use them. This presentation will provide the attendee / reader with a background in SAS arrays, from an explanation as to what occurs during compile time through to their programmatic use, and will include a discussion regarding how DO-loops and macro variables can enhance array usability. Specific examples, including Fahrenheit to Celsius temperature conversion, salary adjustments, and data transposition / counting will assist the user with effective use of SAS arrays in their own work, and provide a few caveats as to their usage, as well.

Animate your Data!
Kriss Harris; Richann Watson

When reporting your safety data, do you ever feel sorry for the person who has to read all the laboratory listings and summaries? Or have you ever wondered if there is a better way to visualize safety data? Let's use animation to help the reviewer and to reveal patterns in your safety data, or in any data!

This presentation demonstrates how you can use animation in SAS® 9.4 to report your safety data, using techniques such as visualizing a patient's laboratory results, vital sign results, and electrocardiogram results and seeing how those safety results change over time. In addition, you will learn how to animate adverse events over time, and how to show the relationships between adverse events and laboratory results using animation. You will also learn how to use the EXPAND procedure to ensure that your animations are smooth. Animating your data will bring your data to life and help improve lives!

Array, Hurray, Array; Consolidate or Expand Your Input Data Stream Using Arrays

William E. Benjamin Jr.

You have an input file with one record per month, but need an output file with one record per year. But you cannot use PROC TRANSPOSE because other fields need to be retained or the input file is sparsely populated. The techniques shown in this paper will enable you to be able to either consolidate or expand your output stream of data by using arrays. Sorted files of data records can be processed as a unit using "BY Variable" groups and building an array of records to process. This technique allows access to all of the data records for a "BY Variable" group and gives the programmer access to the first, last and all records in between at the same time. This will allow the selection of any data value for the final output record.

Automated Validation of Complex Clinical Trials Made Easy

Richann Watson; Josh Horstman

Validation of analysis datasets and statistical outputs (tables, listings, and figures) for clinical trials is frequently performed by double programming. Part of the validation process involves comparing the results of the two programming efforts. COMPARE procedure output must be carefully reviewed for various problems, some of which can be fairly subtle. In addition, the program logs must be scanned for various errors, warnings, notes, and other information that might render the results suspect. All of this must be performed repeatedly each time the data is refreshed or a specification is changed. In this paper, we describe a complete, end-to-end, automated approach to the entire process that can improve both efficiency and effectiveness.

Avoiding Sinkholes: Common Mistakes During ADaM Data Set Implementation

Richann Watson; Karl Miller

The ADaM Implementation Guide was created in order to help maintain a consistency for the development of analysis data sets in the pharmaceutical industry. However, since its inception we have seen issues with guideline non-conformance which can impede this development process and carry impacts that are felt down-stream in subsequent processes. When working with ADaM data sets, non-compliance and other related issues are likely the number one source for numerous hours of re-work; not only creating unnecessary additional work for the data sets themselves, but also for reports, compliance checks, the Analysis Data Reviewers Guide (ADRG), etc. all the way down to the ISS/ISE processes. Considering this breadth of impact, one can see how devastating these sinkholes can be. Like any sinkhole, there is a way out of it but it is a long, tedious process that will consume a lot of resources and it is always better to avoid the sinkhole entirely. This paper will assist you in creating compliant ADaM data sets, provide the reasoning on why you should avoid these sinkholes, all of which will help minimize rework and likely eliminate the need for additional work.

Backsplash Patterns for Your World: A Look at SAS® OpenStreetMap (OSM) Tile Servers

Barbara B. Okerson

Originally limited to SAS Visual Analytics, SAS now provides the ability to create background maps with street and other detail information in SAS/GRAPH® using open source map data from OpenStreetMap (OSM). OSM provides this information using background tile sets available from various tile servers, many available at no cost. This paper provides a step-by-step guide for using the SAS OSM Annotate Generator (the SAS tool that allows use of OSM data in SAS). Examples include the default OpenStreetMap tile server for streets and landmarks, as well as how to use other free tile sets that provide backgrounds ranging from terrain mapping to bicycle path mapping. Dare County, North Carolina is used as the base geographic area for this presentation.

Check Please: An Automated Approach to Log Checking

Richann Watson

In the pharmaceutical industry, we find ourselves having to re-run our programs repeatedly for each deliverable. These programs can be run individually in an interactive SAS® session, which enables us to review the logs as we execute the programs. We could run the individual programs in batch and open each individual log to review for unwanted log messages, such as ERROR, WARNING, uninitialized, have been converted to, and so on. Both of these approaches are fine if there are only a handful of programs to execute. But what do you do if you have hundreds of programs that need to be re-run? Do you want to open every single one of the programs and search for unwanted messages? This manual approach could take hours and is prone to accidental oversight. This paper discusses a macro that searches a specified directory and checks either all the logs in the directory, only logs with a specific naming convention, or only the files listed. The macro then produces a report that lists all the files checked and indicates whether issues were found.

Compare MIXED and GLMMIX to Analyze Breast Cancer Longitudinal Study

Abbas S. Tavakoli

The importance of choosing the type of statistical program that we use to analyze data in a longitudinal study are growing as a specialty due to the fact that computerized data analyses have become the standard for scientific research. There are many procedures in SAS ® that can be used to analyze longitudinal data. The purpose of this paper is to compare MIXED and GLIMMIX procedure in SAS to analyze data from a e longitudinal study. A randomized trial design was used in which 185 participants were assigned to a therapeutic group by teleconference (n=92) in which participants interacted in real time with each other or the control group (n=93) who received usual psychosocial care (any support used by the patient in the course of cancer treatment). The randomization was stratified by treatment type. Data were collected at baseline, the end of the intervention, and 16 weeks from baseline. A mixed-effects repeated measures model was used to assess outcome variable of social well-being (social connection) by group over time. The effect of group, time and interaction effect of group by time were examined after controlling for several confounding factors. SAS is the most powerful statistical program for data analyses in a r longitudinal study.

Considerations in ADaM Occurrence Data: Handling Crossover Records for Non-Typical Analysis

Karl Miller; Richann Watson

With the release of the new ADaM Occurrence Data Model for public comment in the first quarter of 2014, the new model is clearly established to encompass adverse events as well as concomitant medications, along with other data into this standard occurrence analysis structure. Commonly used analysis for this type of occurrence structure data can be found utilizing subject counts by category, based on certain criteria (e.g. treatment, cohort, or study period). In most cases, the majority of the analysis data will be in a one-to-one relationship with the source SDTM record.

In this paper, the authors will discuss the creation of ADaM occurrence data for specific cases outside the common or typical analysis where analysis requires a record in SDTM data, which spans across multiple study treatments, periods or phases, to be replicated for inclusion in non-typical analysis with a record being analyzed under multiple study treatments, periods or phases. From the assignment and imputation of key timing variables (i.e. APERIOD, APHASE, ASTDT), through the appropriate derivation of indicator variables and occurrence flags (i.e. ANLzzFL, TRTEMFL, ONTRTFL and AOCCRFL, AOCCPFL, etc.) the authors guide you through the non-typical process in order to maintain efficiency along with ensuring the traceability in the generation of this analysis-ready data structure.

Creating Zip Code Level Maps with SAS®

Barbara B. Okerson

SAS® provides programmers with the tools to create professional and colorful maps. Provided with SAS/GRAPH are boundary files for U.S. states and territories, as well as internal boundaries at the county level. While much data and results can be displayed at this level, often a higher degree of granularity is needed. The U.S. Census Bureau provides ZIP code boundary files in ESRI shapefile format (.shp) by state for free download and import into SAS using SAS PROC DATAIMPORT. This paper illustrates the use of these ZIP code tabulation area (ZCTA) files with SAS to map data at a ZIP code level. Example maps include choropleth, distance, and heat maps. Examples included in this paper use Base SAS, SAS/STAT and SAS/GRAPH and require a SAS/GRAPH license but are not platform-specific and can be adapted by beginning through advanced SAS users.

Data in the Doughnut Hole: Using SAS® to Report on What is NOT There

Sarah Woodruff

Typically the request for a table, graph or other report concerns data that currently exists and needs to be explored. However, there is often a need to examine data that is expected to be present but currently is not. Yet how can we report on something that is not there? This paper describes and explores ways to create simulations of what is expected and then match those with data that is actually present. Making comparisons between what has been anticipated and what actually exists then opens up the ability to report on what is not there. This paper also explores how to test the differences between the groups based on a variety of conditions and how to determine what is legitimately absent. Reporting on the absent data is then described, including how to check that everything which is expected is actually present and be able to report that as well. DATA step techniques are combined with PROC SQL and PROC REPORT to create a step-wise process which can be easily modified to fit most any set of specifications.

Data Management Plans: Stages, Components, and Activities

Abbas S. Tavakoli

Data management strategies have become increasingly important as new computer technologies allow for larger and more complex data sets to be analyzed easily. As a consequence, data management has become a specialty requiring specific skills and knowledge. Many new investigators have no formal training in management of data sets. This paper describes common basic strategies critical to the management of data as applied to a data set from a longitudinal study. The stages of data management are identified. Moreover, key components and strategies, at each stage are described.

Developing an Analytics Center of Excellence

Chuck Kincaid

Analytics is becoming a mission-critical venture for companies who want to be ahead of their competition. But what does it take for companies to really succeed in analytics? We have a talent shortage of analytics professionals, a wealth of tools and technology, tons of buzz words, and organizations that know they want it, but don't know what "it" is. Based on our experience and expertise, this presentation will give the reader insight into some of the challenges of building an Analytics Center of Excellence and guidance in

ways to overcome them. The presentation will be accessible and valuable to any attendee in any industry. It will be of particular interest to anyone with the desire to promote analytics within their organization.

Easier than You Think: Creating Maps with SAS® Enterprise Guide®

Stephanie Thompson

Have you ever wanted to display data on a map to add more punch to your analysis? Sometimes seeing things geographically can put things into a completely new perspective. Letting your data tell a story on a map is easier than you might think. A two-step approach with SAS® Enterprise Guide® is all you need. From using supplied shape files right through customizing your graph, this paper walks you through the process. Some best practices are also included to make sure your customers are impressed. They will never know how easy it really was.

Easy Ways to get SAS® Enterprise Guide® to Send Tables and Graphs to Microsoft Office Suite, Without Writing Code

William E. Benjamin Jr.

This paper will demonstrate several ways to output tables, graphs, databases, and files that are in formats that Microsoft products are able to process. All of these methods require no programming, and just use little known or hidden techniques that are available directly from the SAS® Enterprise Guide user interface. The focus will be directly on how a new user to SAS® Enterprise Guide can prepare and deliver data to the Microsoft Office (MS Office) suite of tools.

Evaluating Sample Code for an Interview

Stephanie Thompson

Requesting sample SAS® code from job candidates before an interview is a good way to gauge their level of experience, ability, and style. Once you have reviewed the code, prepare your questions for the candidate. You want to make sure that they 1) wrote the code themselves, 2) understand what they wrote, and 3) are familiar enough with it to explain why things were done the way they were. This paper will provide you with some guidelines on how to get answers to these questions in an interview. Code samples that I have received over the last few years and the questions that arose will be used as examples. This is also a good time to ask if the candidate can think of another way to accomplish the same task using a different approach. Evaluating work samples can be a way to gain invaluable insight into your candidate's skills and this paper will help you get the most out of it that you can.

Examining Mediator and indirect effects of Loneliness in Social Support on Social Well-Being Using Baron and Kenny and a Bootstrapping Method Data Management Plans: Stages, Components, and Activities Compare MIXED and GLMMIX to Analyze Breast Cancer Longitudinal Study

Abbas S. Tavakoli

This study presentation examines mediator effect and indirect effect of loneliness in social support on social well-being by using two methods: Baron & Kenny, and Bootstrapping. A cross-sectional data were used here from the longitudinal study randomized trial design in which 185 participants were assigned to the therapeutic group (n=93) who received by teleconference with participants interacting in real time with each other and control group (n=92) who received usual psychosocial care (any support used by the patient in the course of cancer treatment. Baron & Kenny (1986) steps and Hayes (2004) we reused to examine mediation effect. Results of Baron indicated that the relationship between social support and social well-being was significant (c =.634 (total effect) (p=.0001)) and that there was significant relationship between mediator and predictor variable (α =-2.11 (p=.0001)). Also, previously found

significant relationship between social support and social well-being w (c' =.595 (direct effect) (p=.0001) when both social support and loneliness were in the model. The indirect effect was .039 and Sobel test was significant (P=.028). Therefore, there was no mediator effect for loneliness in the relationship between social support and social well-being. The results of bootstrapping methods indicated the direct effect wares .591 (95% CI: .589-.593 for normal theory and .481- .690 for percentile) and indirect effect was .040 (95% CI: .039-.040 for normal theory and .006-.087for percentile). The result showed both methods had significant indirect effect.

Exploring HASH Tables vs. SORT/DATA Step vs. PROC SQL

Richann Watson; Lynn Mullins

There are often times when programmers need to merge multiple SAS® data sets to combine data into one single source data set. Like many other processes, there are various techniques to accomplish this using SAS software. The most efficient method to use based on varying assumptions will be explored in this paper. We will describe the differences, advantages and disadvantages, and display benchmarks of using HASH tables, the SORT and DATA step procedures, and the SQL procedure.

Extend the Power of SAS® to Use Callable VBS and VBA Code Files Stored in External Libraries to Control Excel Formatting Routines

William E. Benjamin Jr.

Did you ever wish you could use the power of SAS® to take control of 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 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 VBA macro code and control EXCEL. This level of control allows you to make EXCEL do what you want, without leaving any traces of a macro behind. This is Power.

Flat Pack Data: Converting and ZIPping SAS® Data for Delivery

Sarah Woodruff

Clients or collaborators often need SAS data converted to a different format. Delivery or even storage of individual data sets can become cumbersome, especially as the number of observations or variables grows. The process of converting SAS data sets into other forms of data and saving files into compressed ZIP storage has become not only more efficient, but easier to integrate into new or existing programs. This paper describes and explores various methods to convert SAS data sets to other formats as well as effective strategies to ZIP data sets along with any other files that might need to accompany them.

PROC IMPORT and PROC EXPORT have been long standing components of the SAS toolbox, so much so that they them in code being run in batch or to include them in programs that may be run interactively but "hands free". The syntax of each is described with a particular focus on moving between SAS, STATA and SPSS, though some attention is also given to Excel. Once data sets and their attendant files are ready for delivery or need to be put into storage, compressing them into ZIP files becomes helpful. The process of using ODS PACKAGE to create such ZIP files is laid out and can be connected programmatically to the creation of the data sets or documents in the first place.

Four Thousand Reports Three Ways

Stephanie Thompson

How do you go about generating over four thousand PDF reports in up to three different versions? When a large, southern research university decided to add up to five optional questions per class section and up to five more questions at the prefix level to their core set of fifteen questions on the student

evaluation of faculty survey, it seemed like a project that would never be completed. If the additional questions weren't enough, the reports were being revamped at the same time to improve their appearance for delivery on the web. Each report had a tabular section and two customized box and whisker plots. Thanks to ODBC / SAS Access, PROC SQL, macro, DATA Step programming, PROC GPLOT, goptions, and ODS it all came together. This paper summarizes how each SAS® component was used and contributed to the completion of the project.

FREQ Out – Exploring Your Data the Old School Way

Stephanie Thompson

The tried-and-true FREQ procedure just doesn't get the attention it deserves. But, as they say, it is an oldie but a goodie. Sometimes you just need a quick look at your data and a few simple statistics. PROC FREQ is a great way to get an overview of your data with a limited amount of code. This hands-on workshop explores everything from the basic framework of the procedure to how to customize the output. It also presents an overview of some of the options that are available.

Generating Custom Shape Files for Data Visualization - note this is a Quick Tip and can be done with another choice

Stephanie Thompson

There are times you may want to represent your data graphically but the existing options do not work. Using a custom shape file or even a file with coordinates you create yourself can get you what you want. This quick presentation will show you how.

Getting the Most Out of Your SAS Conference

Andrew T. Kuligowski

Your company has approved your attendance at the conference. They've agreed to an outlay of money to cover your registration, transportation, food, and lodging. Equally important, they've made a commitment of your time, allowing you to be at the event instead of sitting at your computer terminal. They fully anticipate that you will return their investment through the knowledge you bring back. How do you ensure that you make the best use of your time – to get the most out of your conference experience?

I've been a speaker at many SAS User Group conferences and events over the years. In that time, I think I've learned some tricks to best utilize my time at user group events. This presentation will be a short, informal look into how I make the best use of my time; hopefully, it will assist you in getting the most out of your conference.

Getting Your Data into SAS®

Stephanie Thompson

Where is your data stored?

- Oracle tables
- SQL Server tables
- Microsoft Access
- Microsoft Excel
- Text file
- All over the place

Over the years there has been a proliferation of ERP systems and other ways to collect and store data. Many times you need data from different systems to complete a single analysis. Sometimes getting data out can seem like quite a challenge. Luckily, SAS has the capability to access many different types of

data but also different ways to do so. This workshop will cover how to use SAS to access data from a variety of sources through both presentation and live demonstration. Some practical tips on which methods work best or are fastest will also be covered.

The following methods to access data will be covered in the workshop:

- SAS libname engine
- PROC SQL
- ODBC
- PROC IMPORT (including the Import Wizard)

Give the Power of SAS® to Excel Users Without Making Them Write SAS Code as a Take-out"

William E. Benjamin Jr.

Merging the ability to use SAS® and Microsoft Excel can be challenging. However, with the advent of SAS® Enterprise Guide®, SAS® Integration Technologies, SAS® BI Server software, JMP® software, and SAS® Add-In for Microsoft Office; this process is less cumbersome. Excel has the advantages of being cheap, available, easy to learn, and flexible. On the surface, SAS and Excel seem widely separated without these additional SAS products. But wait, BOTH SAS AND EXCEL CAN INTERFACE WITH THE OPERATING SYSTEM. SAS can run Excel using the command and Excel can run SAS as an "APPLICATION." This is NOT DDE; each system works independently of the other. This paper gives an example of Excel controlling a SAS process and returning data to Excel.

Great Time to Learn

Kriss Harris; Richann Watson

It's a Great Time to Learn GTL! Do you want to be more confident when producing GTL graphs? Do you want to know how to layer your graphs using the OVERLAY layout and build upon your graphs using multiple LAYOUT statement? This paper guides you through the GTL fundamentals!

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 rely on pre-defined templates built with GTL. 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 **G**reat **T**ime to **L**earn 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.

Guidelines for Selecting the Covariance Structure in Mixed Model Analysis

Chuck Kincaid

Mixed Models is rapidly becoming a very useful tool for statisticians. As a general paradigm it can be used to handle almost every situation, especially if you extend the Linear Mixed Model to the Generalized Linear Mixed Model case or the Nonlinear Mixed Model case. It's also an area in which a lot of research is being done, because the questions are far from being answered. Advanced computing power is giving us the capability to answer those questions. One important question which, unfortunately, still has no good answer is how to select the covariance structure. This presentation is an attempt to survey the information available for answering the question.

How to Succeed in Consulting Without Really Trying

Chuck Kincaid

Maybe you are an awesome programmer working on a company's internal consulting team, but you have a hard time getting work done by the deadline. Maybe you are a strong SAS developer who does independent consulting, but clients get upset when changes they ask for cost them more money. Just because you're good in the technical skills doesn't mean that you can succeed as a consultant. This presentation will give you tips on how to do just that. With a combination of project management and consulting skills you can go much farther, whether it's doing internal consulting, independent consulting or working for a consulting company. This presentation will be good for people who want to do better at managing the world outside of their code.

How to be a Data Scientist with SAS

Chuck Kincaid

The role of the Data Scientist is the viral job description of the decade. And like LOLcats, there are many types of Data Scientists. What is this new role? Who's hiring them? What do they do? What skills are required to do their job? What does this mean for the SAS programmer and the statistician? Are they obsolete? And finally, if I'm a SAS user, how can I become a Data Scientist? Come learn about this "job of the future" and what you can do to be part of it.

How to HOW: Hands-on-Workshops Made Easy

Chuck Kincaid

Have you ever attended a Hands-on-Workshop and found it useful? Many people do! Being able to actually try out the things that you're learning is a wonderful way to learn. It's also a great way to teach. You can actually see the people apply what they're learning.

Have you ever thought that it would be fun to teach other people in a hands on format? Maybe you weren't sure what it takes or how to approach the course. This presentation will help you with those questions and struggles. What to teach? How much to teach? How should I teach it? How is a Hands-on Workshop different than lecture style? How much to put into PowerPoints? What if they ask me something I don't know? What if they have a computer problem? All those questions that you have will be answered in this presentation.

I'm a SAS® Programmer. Why Should I JMP®?

Barbara B. Okerson

JMP® software, when used as a stand-alone package, provides a variety of ways of understanding, visualizing and communicating what your data is telling you. Adding JMP software and functionality to a SAS® programming environment can result in the best of both worlds. Topics covered include previewing SAS data, running SAS procedures from JMP, using JMP for further exploration of SAS results, and using SAS geographic data with JMP. Through these features and others in JMP's point-and-click environment, JMP extends the power and functionality of SAS. The techniques represented are not platform-specific and can be adapted by beginning through advanced SAS and JMP users.

In Search of the `LOST CARD`

Andrew T. Kuligowski

"Everyone who's not here, raise your hand." It's an old joke, but it points out the difficulty of identifying persons or things that are not present. The SAS® System has its own version of this chestnut, the SASLOG message indicating that there are one or more gaps in one's input data:

```
NOTE: LOST CARD.
```

This presentation will focus on the creation and use of ad hocs to explore input data, in order to locate the positions where the input data might be incomplete. The goal will be to identify where the missing data should be, so that you can code around the limitations of your data.

Interactions between the DATA Step and External Files – INFILE / INPUT and more

Andrew T. Kuligowski

Chances are, your raw data was not created within the SAS® System. There is a good likelihood that your data may also need to be packaged and passed along to another non-SAS package. This presentation will provide basic answers to two questions common to new SAS users:

- How do I get my data into SAS for analysis?
- How do I get my data out of SAS?

The focus for this presentation will be on two pairs of DATA step statements: INFILE / INPUT and FILE / PUT. We will discuss syntax and usage, citing various types of files as examples.

Intermediate SAS® Macro Programming

Chuck Kincaid

The SAS Macro Language powerfully enhances a programmer's capabilities by providing an advanced level of flexibility and robustness to otherwise static programs. Programmers can use the Macro language to create dynamic programs that can be re-used in new situations, driven by the data, dependent upon the operating environment, and made to handle repetition without extensive coding. This Hands-On Workshop will explore some of the more advanced capabilities with the SAS Macro Language that may not be familiar to those just starting in the language. Some of the features explored include understanding variable scope, interacting with data steps, evaluating expressions, quoting variables, macro looping, macro debugging, and creating data driven programs. Basic experience with macro programming is assumed.

Introduction to Data Simulation

Jason Brinkley

Creating synthetic data via simulation can often be a powerful tool for a wide variety of analyses. The purpose of this workshop is to provide a basic overview of simulating data for a variety of purposes. Examples will include power calculations, sensitivity analysis, and exploring nonstandard analyses. The workshop is designed for the mid-level analyst who has basic knowledge of data management, visualizations and basic statistical analyses such as correlations and t-tests.

Introduction to Hierarchical Linear Modeling (HLM)

Jason Brinkley

HLM, also known as multi-level analysis, is a more advanced form of simple linear regression and multiple linear regression. Multilevel analysis allows variance in outcome variables to be analyzed at multiple hierarchical levels, whereas in simple linear and multiple linear regression all effects are modeled to occur at a single level. Thus, HLM is appropriate for use with nested data.

Intro to Proc Report

Barbara B. Okerson

Proc Report is a SAS tool that combines and enhances the features of the print, means, and tabulate procedures to summarize and display data. This presentation looks at the basic syntax and commonly used customization, both using standard SAS output templates and writing output to excel spreadsheets.

Intro to Proc Report with PDF Output

Barbara B. Okerson

Proc Report is a SAS tool that combines and enhances the features of the print, means, and tabulate procedures to summarize and display data. The flexibility of the report procedure allows the design of reports as per requested specifications with every aspect fully customizable. Examples create client-ready reports in PDF format without post-processing. Customizations illustrated include: splitting column labels, creating spanning headers, grouping and ordering results, currency formatting, and cell-width modifications.

Keeping Up Appearances: Turning Specifications into SAS® Format Libraries and Statements

Sarah Woodruff

Specifications documents concerning the desired appearance of SAS data are often provided in an Excel spreadsheet format. While such an arrangement provides ease of use to the person creating it, typically the client to whom the final data delivery will eventually be returned, having format information set up this way is not directly conducive to its use in SAS. This paper describes a process by which formats presented in this way can be easily converted into both SAS format libraries and format statements. The process centers around the use of standard reports from Oracle Clinical to provide the basic information, but the methods can be applied to any specifications document in Excel. Along with the final format products, this encompasses evaluating the appropriateness of format names and making broadly applicable changes as needed, ensuring unique variable names, particularly if data is coming from multiple sources, and building in appropriate quality control steps along the way. This conversion requires both DATA step work and several common procedures (IMPORT, CONTENTS, FREQ), but is accomplished entirely in Base SAS and does not need any special products. Flexibility remains in the process to compile or subdivide both the format catalog and statements as needed by the user or as applicable based on the requirements of a particular deliverable.

Leave Your Bad Code Behind: 50 Ways to Make Your SAS Code Execute More Efficiently

William E. Benjamin Jr.

This laundry list of tips shows 50 ways to help SAS programmers make their code run faster and more efficiently. Topics include maximizing each DATA step, making simple tasks take less code, using macro variables to simplify maintenance, using built-in features, optimizing code to save disk space, using sorts for more than just sorting, and ways to make the program code just read better.

Let SAS® Do Your DIRty Work

Richann Watson

Making sure that you have saved all the necessary information to replicate a deliverable can be a cumbersome task. You want to make sure that all the raw data sets and all the derived data sets, whether they are Study Data Tabulation Model (SDTM) data sets or Analysis Data Model (ADaM) data sets, are saved. You prefer that the date/time stamps are preserved. Not only do you need the data sets, you also need to keep a copy of all programs that were used to produce the deliverable, as well as the corresponding logs from when the programs were executed. Any other information that was needed to produce the necessary outputs also needs to be saved. You must do all of this for each deliverable, and it can be easy to overlook a step or some key information. Most people do this process manually. It can be a time-consuming process, so why not let SAS® do the work for you?

Let's Get FREQy with our Statistics: Data-Driven Approach to Determining Appropriate Test Statistic

Richann Watson; Lynn Mullins

As programmers, we are often asked to program statistical analysis procedures to run against the data. Sometimes the specifications we are given by the statisticians outline which statistical procedures to run. But other times, the statistical procedures to use need to be data dependent. To run these procedures based on the results of previous procedures' output requires a little more preplanning and programming. We present a macro that dynamically determines which statistical procedure to run based on previous procedure output. The user can specify parameters (for example, fshchi, plttwo, catrnd, bimain, and bicomp), and the macro returns counts, percents, and the appropriate p-value for Chi-Square versus Fisher Exact, and the p-value for Trend and Binomial CI, if applicable.

Looking Beneath the Surface of Sorting

Andrew T. Kuligowski

Many things that appear to be simple turn out to be a mask for various complexities. For example, as we all learned early in school, a simple drop of pond water reveals a complete and complex ecosystem when viewed under a microscope. A single snowflake contains a delicate crystalline pattern. Similarly, the decision to use data in a sorted order can conceal an unexpectedly involved series of processing and decisions. This presentation will examine multiple facets of the process of sorting data, starting with the most basic use of PROC SORT and progressing into options that can be used to extend its flexibility. It will progress to look at some potential uses of sorted data, and contrast them with alternatives that do not require sorted data. For example, we will compare the use of the BY statement vs. the CLASS statement in certain PROCs, as well as investigate alternatives to the MERGE statement to combine multiple datasets together.

Managing Your SAS® Assets

Stephanie Thompson

Hiring talented SAS® professionals can be an expensive proposition. While it is great to get another position for your department, replacing someone can be a real downer. Why did they leave? Was it the work? Did we make a poor hiring decision? Was it just a better opportunity? Could they have been happy here? Maybe they just did not like me. All of these questions are valid, and you can keep guessing and never really know. As they say, an ounce of prevention is worth a pound of cure. What can you do as a manager to keep your best talent happy and challenged? Keeping those in your department engaged is a big piece of the puzzle. Not a manager? Well, you're not off the hook either. There are things you can do to help your manager manage you better. I've been on both sides of this fence, and this paper will talk about some of the things that will help keep everyone happy. Think of your company's SAS talent as SAS assets!

Messy Data in SAS

Jason Brinkley

Transitions between character and numeric data, working with dates and time, summarizing large data into smaller datasets by counts or groupings.

More than Models: The Data Mining Team

Stephanie Thompson

Gathering data from various sources, preparing it for modeling, imputing, partitioning, testing various models, choosing the best, presenting it to your boss, then failure? Data mining is about more than just variables and models. Developing an understanding of each variable may take more than just running some summary statistics and deciding if it is ordinal or nominal. The need for comprehension is even more critical when the data you use are from different areas of the organization. Putting together a group of subject matter experts in the early phases of a data mining project can make a big difference in the outcome of your project. They can help you eliminate extraneous or duplicative variables and put others in context to help you better understand and interpret results. This paper will discuss how subject matter experts can aid in data mining using examples from several actual projects. Learn how to leverage the knowledge to derive a better conclusion and avoid costly errors.

Navigating Workforce Trends In The Human Age

Chuck Kincaid

The job market today is much different from a decade ago, and employers and job candidates need to be aware of the workforce trends impacting society and the world of work. This is particularly true for statisticians because we see a building demand for individuals with the skills needed to help the business succeed by effectively analyzing their data. Despite high unemployment in the world today, there is a pervasive talent mismatch where employers cannot find the people with the right skills for mission-critical positions. This puts the issue of talent at center stage as we emerge from the current recession and enter what we are calling the Human Age; an era where people with the right talents and skills will be critical to business success.". Understanding how to navigate workforce trends becomes a key consideration for employers and job seekers alike.

New to SAS® and New to Programming? What You Need to Do Before Typing Code

Stephanie Thompson

People from all kinds of backgrounds are using SAS® software to meet their business needs. Some come with previous programming experience and some have never typed a line of code in their career. Have you been told what each proc does, that each DATA step must end in a run command, and then been told to write a SAS program? Have you been given code to launch and told to "just change the date and run it every week"? Beware. There is a lot more to SAS programming than knowing some syntax and launching a job. Being a successful SAS programmer means bringing yourself up to speed with some basic programming skills. Before typing even the first line of code into the Enhanced Editor, some preparation needs to be done. Doing this will help you more than you can imagine. This paper provides the basics of programming that can help you write good code and ensure you are getting the answer you intended to get.

ODS Graphics

Chuck Kincaid

This presentation will teach the audience how to use SAS® ODS Graphics. Now part of Base SAS 9.3, ODS Graphics are a great way to easily create clear graphics that allow any user to tell their story well. SGPLOT and SGPANEL are two of the procedures that can be used to produce powerful graphics that used to require a lot of work. The core of the procedures are explained, as well as the options available. Furthermore, we explore the ways to combine the individual statements to make more complex graphics that tell the story better. Any user of Base SAS on any platform will find great value from the SAS ODS Graphics procedures.

Parsing Useful Data Out of Unusual Formats Using SAS®
Andrew T. Kuligowski

Most “Introduction to Programming” courses will include a section on reading external data; the first assumption they make will be that the data are stored in some sort of documented and consistent format. Fortunately, in the “real world”, a lot of the data we deal with has the same basic assumption of occurring in a documented, consistent format – a lot of it, but not all of it. This presentation will address some techniques that can be used when we are not dealing with cleanly formatted data, when the data we want is in a less-than-ideal format, perhaps intermingled or seemingly buried with unnecessary clutter. It will discuss the principles of using SAS® to parse a file to extract useful data from a normally unusable source. This will be accomplished by citing examples of unusual data sources and the SAS Code used to parse it.

Perusing, Choosing, and Not Mis-using: Non-parametric vs Parametric Tests in SAS (R)
Venita DePuy

Most commonly used statistical procedures, such as the ttest, are based on the assumption of normality. The field of non-parametric statistics provides equivalent procedures that do not require normality, but often require assumptions such as equal variances. Parametric tests (which assume normality) are often used on non-normal data; even non-parametric tests are used when their assumptions are violated. This paper will provide an overview of parametric tests and their non-parametric equivalents; what assumptions are required for each test; how to perform the tests in SAS®; and guidelines for when both sets of assumptions are violated. Procedures covered will include PROCs ANOVA, CORR, NPAR1WAY, TTEST and UNIVARIATE. Discussion will include assumptions and assumption violations, robustness, and exact versus approximate tests.

Predicting Analytics: Where will Analytics be in 20 years?
Chuck Kincaid

Analytics has gotten very popular in the last 10 years. Technological innovations, executive backing, new programs and even movies about analytics have built that excitement. What does the next 20 years bring for the field? What impact will the trends in analytics talent and virtual reality have upon the way we do analytics? In this presentation we will explore the potential futures of these types of areas to predict what field of analytics will be like in 2036.

PROC SURVEY… Says!: Selecting and Analyzing Stratified Samples
Darryl Putnam

Statisticians and analysts need to design stratified survey plans and analyze the results of those surveys. Gone are the days when the analyst can ignore survey design tools when drawing inferences from the surveys. By forgoing the SAS© survey analysis procedures, estimates of the mean and standard error will be incorrect. By combining DATA STEP processing with the SAS© survey analysis procedures of PROC SURVEYSELECT and PROC SURVEYMEANS, we can determine the sample size, allocate the sample size across strata, and then draw correct inferences. This paper will demonstrate how to use these survey design and analysis tools with a stratified sample of an inventory audit.

Pruning the SASLOG – Digging into the Roots of NOTEs, WARNINGs, and ERRORs®
Andrew T. Kuligowski

You've sat through constant design meetings. You've endured countless requests for "just one more little change". You even managed to find a creative solution to that nagging technical problem. But, you

persevered, and despite all of the obstacles, you've managed to eliminate the final syntax error in your newest SAS routine. Time to sit back and relax -- uh, not quite. The primary focus of this presentation will be on techniques to ensure comprehension of your input data. We will look at several messages that are often found in the SASLOG, such as:

```
NOTE: MERGE statement has more than one data set with repeats of BY values.
```

that imply that there may be gaps in your knowledge of your data! Special emphasis will be placed on the use of ad-hoc queries to assist in finding data anomalies that can cause problems with your SAS code. It is assumed that the reader has a basic understanding of the SASLOG, including its composition, format, and the SAS system options which control its content.

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 workshop is to illustrate how count regression models can outperform traditional methods while utilizing the data in a more appropriate manner. Most of the first half of the workshop will focus on Poisson Regression techniques but some time will be devoted to using Negative Binomial Regression when the data are overdispersed and using Zero-Inflated techniques for data with many more zeroes than is expected under traditional count regression models. Time will be spent applying such models to example data so that the participant can see when and how to utilize these methods. A strong grasp of traditional linear regression methods is suggested.

SAS Users Can Command Microsoft Excel to Automatically Create Graphs from SAS ExcelXP Output

William E. Benjamin Jr.

The SAS ODS Tagset ExcelXP creates *.xml output, and *.xml output cannot contain graphs. So how can SAS programmers get graphs into your Excel workbooks? One way is to build them in Excel yourself. This paper shows you how to create data using SAS, and then command Microsoft Excel to read the data, create a graph or fully reformat a worksheet, without putting an Excel macro into the output Excel Workbook. And the program will do it all while you watch, including for multiple sheets in a workbook. The SAS code, and Excel code, shown is a fully integrated system to create and format macro-free Excel workbooks, using SAS® 9 and Excel 97 and above.

Smoke and Mirrors!!! Come See How the _INFILE_ Automatic Variable and SHAREBUFFERS Infile Option Can Speed Up Your Flat File Text-Processing Throughput Speed

William E. Benjamin Jr.

Processing speed for small files is never really a problem; one eye blink looks and feels much the same as two or three blinks of the eye. With today's computer speeds any machine will do a lot of work in three blinks of the eye. However, when processing hundreds of thousands, or millions of records blinks can become pauses, yawns, naps, or even an over-night sleepover. When a Base SAS® data step reads in text files and writes out nearly the same file the work proceeds as follows; (a) read input file data into input buffer, (b) move data values to data step variables (c) process data variables, (d) move data to output buffer, (e) write output file from output buffer. SAS reads your source code every time you run the code

and builds the “current” processing steps (unless the code is compiled separately first) then executes them. By packing more work into each instruction less time is wasted between instructions, this means more work in less real time. The SHAREBUFFERS option eliminates moving data from the input to the output buffer (a BIG time saver). And the _infile_ automatic variable can be used in function calls on either side of the “=” sign. This eliminates some or all of the data step variable processing like this; (a) read input file data into a common buffer, (b) process data values, (c) write data output text file from common buffer. The resulting time saved can result in faster execution of the program.

So, You've Got Data Enterprise Wide (SAS, ACCESS, EXCEL, MySQL, Oracle, and Others); Well, Let SAS Enterprise Guide Software Point-n-Click Your Way to Using It

William E. Benjamin Jr.

This paper will walk a SAS user through the process of opening and using files from different (non-SAS) generated sources and processing the data with SAS Enterprise Guide (EG). The focus of this paper will be to get access to the data, since most functions available to EG are drop down menus and wizards which help teach a new user how to process data. For completeness a simple chart will be shown with instructions about how to each type of data shown. This paper will not show how to get the data into non-SAS files; it will be assumed that the user’s enterprise will guide them through that process.

Standardized, Customized or Both? Defining and Implementing (MedDRA) Queries in ADaM Data Sets

Richann Watson; Karl Miller

Investigation of drug safety issues for clinical development will consistently revolve around the experience and impact of important medical occurrences throughout the conduct of a clinical trial. As a first step in the data analysis process, Standardized MedDRA Queries (SMQs), a unique feature of MedDRA, provide a consistent and efficient structure to support safety analysis, reporting, and also address important topics for regulatory and industry users. A variance in working with SMQs is the ability to limit the scope for the analysis need (e.g., “Broad” or “Narrow”) but there is also the ability outside of the specific SMQs in allowing the ability to develop Customized Queries (CQs). With the introduction of the ADaM Occurrence Data Structure (OCCDS) standard structure, the incorporation of these SMQs, along with potential CQs, solidified the need for consistent implementation, not only across studies, but across drug compounds and even within a company itself. Working with SMQs one may have numerous questions: What differentiates the SMQ from a CQ and which one should be used? Are there any other considerations in implementation of the OCCDS standards? Where does one begin? Right here…

Talking Past Each Other? How to Communicate with Medical Writers When Preparing Clinical Research Manuscripts for Journal Submission

Stephanie Thompson

Clinical research manuscripts are often a blend of qualitative prose and quantitative data. And, the people who help the author(s) prepare each section — introduction, methods and materials, results, and discussion — are typically from very different parts of the research organization and have different backgrounds. Speaking a common language makes the process easier. This paper will present some tips on how to improve communication between medical writers and SAS® programmers, such as biostatisticians. Presented from both points of view, what does and doesn’t work will be discussed in a point, counter-point fashion.

Telling the Story Better: Advanced ODS Graphics

Chuck Kincaid

The SAS® Statistical Graphics (SG) procedures: SGPLOT, SGPANEL, SGSCATTER, and SGRENDER are exciting additions first put into production in SAS 9.2 that give easy access to some of the power of the Graphics Template Language (GTL). As good as these procedures are in helping you tell your analytical story, and they are good, there are techniques that can add extra value to your graphs. This presentation will discuss four concepts introduced by William S. Cleveland in his book Visualizing Data – level ordering, slicing, banking, and stacking. For these techniques, we will explain the concepts, provide examples and outline the basic algorithms that can be found in the associated paper. The audience for this presentation is the statistician or business analyst who wants to tell their story even better.

The Building Blocks of SAS® Datasets – S-M-U (Set, Merge, and Update)

Andrew T. Kuligowski

S-M-U. Some people will see these three letters and immediately think of the abbreviation for a private university and associated football team in Texas. Others might treat them as a three-letter word, and recall a whimsical cartoon character created by Al Capp many years ago. However, in the world of the SAS® user, these three letters represent the building blocks for processing SAS datasets through the SAS DATA step. S, M, and U are first letters in the words SET, MERGE, and UPDATE – the 3 commands used to introduce SAS data into a DATA step. This presentation will discuss the syntax for the SET, MERGE, and UPDATE commands. It will compare and contrast these 3 commands. Finally, it will provide appropriate uses for each command, along with basic examples that will illustrate the main points of the presentation.

The Care and Feeding of Magical Creatures: Managing Statisticians and Data Scientists in an Analytically Enthused World

Chuck Kincaid

The stereotypical statistician used to be the geeky guy in the back office working by himself for many months on problems that others may or may not care about. It's exciting for our profession that those days are over. The response we get from others is no longer "I had statistics in college and I hated it." It's now "How can I learn more?"

This enthusiasm of the masses is exciting for us, but, at the same time, it does present new challenges for analytical teams in hiring, training, equipping, managing and retaining the magical creatures that are statisticians and data scientists. How do we build teams when the talent gap is so large? How do we enable the teams to truly impact the business? How do we develop and challenge our team so that they are with us for a long time? We will discuss these questions in more detail and give suggestions that you can take back to your own organization.

The Little Engine That Could: Using LIBNAME Engine Options to Enhance Data Transfers Between SAS® and Microsoft Excel Files

William E. Benjamin Jr.

Many people are not aware that the SAS ® Access for PC Files product will allow SAS Programmers to access an Excel spreadsheet in much the same way as any other SAS file. There are of course some restrictions, but there are also a lot of options that help remove some of the bumps in the road. The LIBNAME statement allows the user to define an Excel file in SAS terms and gives the programmer access to LIBNAME and data set options to control how the Excel file is defined, accessed, and yes even how the data will be formatted. This paper will describe some of those options.

Time Series Mapping with SAS®: Visualizing Geographic Change over Time in the Health Insurance Industry

Barbara B. Okerson

Changes in health insurance and other industries often have a spatial component. Maps can be used to convey this type of information to the user more quickly than tabular reports and other non-graphical formats. SAS® provides programmers and analysts with the tools to not only create professional and colorful maps, but also the ability to display spatial data on these maps in a meaningful manner that aids in the understanding of the changes that have transpired. This paper illustrates the creation of a number of different maps for displaying change over time with examples from the health insurance arena.

Tips and Tricks for Introductory Workshops in SAS® for Health Professionals

Jason Brinkley

It can sometimes be the case that general health professionals need some basic SAS training in order to effectively create simple reports and manipulate incoming data. The presenter will share his experiences in leading SAS Workshops Series in a university setting across the course of several years. Heading a team of university faculty members, the presenter has designed, implemented, and refined short term SAS overview training for general health professionals. While multiple topics have been discussed in these workshops, some have fared better with a general health professional audience than others. Topics will include tips on introducing code based work to individuals with no previous experience, workshop format, good practices on instruction and delivery, and introducing SAS macros in an example based manner.

Tricks and Tips for Using the Bootstrap in JMP Pro

Jason Brinkley

The bootstrap has become a very popular technique for assessing the variability of many different or unusual estimators. Starting in JMP Pro 10 the bootstrap feature was added to a wide variety of output options; however, there has not been much development as to the possible uses of this somewhat hidden feature. This paper will discuss a handful of uses that can be added to routine analyses. Examples include confidence interval estimates of the 5% trimmed mean, validation of covariates in regression analysis, comparing the differences in Spearman correlation estimates across two groups, and eigenvalues in principal components analysis. The examples will show the extra depth that can be easily added to routine analyses.

Using a Picture Format to Create Visit Windows

Richann Watson

Creating visit windows is sometimes required for analysis of data. We need to make sure that we get the visit/day in the proper window so that the data can be analyzed properly. However, defining these visit windows can be quite cumbersome especially if they have to be defined in numerous programs. This task can be made easier by applying a picture format, which can save a lot of time and coding. A format is easier to maintain than a bunch of individual programs. If a change to the algorithm is required, the format can be updated instead of updating all of the individual programs containing the visit definition code.

Using Big Data to Visualize People Movement Using SAS® Basics

Stephanie Thompson

Visualizing the movement of people over time in an animation can provide insights that tables and static graphs cannot. There are many options but what if you want to base the visualization on large amounts of data from several sources? SAS is a great tool for this type of project. This paper will summarize how visualizing movement was accomplished using several datasets, large and small, and the various SAS

PROCS to pull it together. The use of a custom shape file will also be highlighted. The end result is a gif which can be shared that provides insights not available with other methods.

Using INFILE and INPUT Statements to Introduce External Data into the SAS® System

Andrew T. Kuligowski

The SAS® System has numerous capabilities to store, analyze, report, and present data. However, those features are useless unless that data is stored in, or can be accessed by, the SAS System. This presentation is designed to review the INFILE and INPUT statements. It has been set up as a series of examples, each building on the other, rather than a mere recitation of the options as documented in the manual. These examples will include various data sources, including DATALINES, sequential files, and CSV files.

Using GENMOD to Analyze Correlated Data on Military Health System Beneficiaries Receiving Inpatient Behavioral Health Care in South Carolina Health Care System

Abbas S. Tavakoli

Many SAS® procedures can be used to analyze large datasets with correlated data. This study was a secondary analysis of data obtained from the South Carolina Revenue and Fiscal Affairs Office (RFA), which includes medical claims from all health care systems in South Carolina (SC). We used the SAS procedure, GENMOD, to analyze a large dataset with correlated data from Military Health Care (MHS) system beneficiaries who received inpatient behavioral health care in South Carolina Health Care Systems from 2005 to 2014. Behavioral health (BH) was defined by Major Diagnostic Code (MDC) 19 (mental disorders and diseases) and 20 (alcohol and drug use). MDCs are formed by dividing all possible principal diagnoses from the International Classification Diagnostic (ICD-9) codes into 25 mutually exclusive diagnostic categories. The sample included a total of 6,783 BH visits and 4,827 unique military service members, veterans, and their adult and child dependents who have MHS insurance coverage. PROC GENMOD included a multivariate GEE model with type of BH hospitalization (mental health or substance abuse) as the dependent variable; and gender, race, age group, and discharge year as predictors. Hospital ID was used in the repeated statement with different correlation structures. Gender was significant for both independent correlation ($p = .0001$) and exchangeable structure ($p = .0003$). However, age group was significant using the independent correlation ($p = .0160$), but non-significant using the exchangeable correlation structure ($p = .0584$). SAS is a powerful statistical program for analyzing large, correlated datasets with categorical outcomes.

Using GLIMMIX and GENMOD Procedures to Analyze Longitudinal Data from a Department of Veterans Affairs Multisite Randomized Controlled Trial

Abbas S. Tavakoli

Many SAS® procedures can be used to analyze longitudinal data. This study employed a multisite randomized controlled trial design to demonstrate the effectiveness of two SAS procedures, GLIMMIX and GENMOD, to analyze longitudinal data from five Department of Veteran Affairs Medical Centers (VAMCs). Older male veterans ($n = 1222$) seen in VAMC primary care clinics were randomly assigned to two behavioral health models, integrated ($n = 605$) and enhanced referral ($n = 617$). Data were collected at baseline, and 3, 6, and 12 month follow-up. A mixed-effects repeated measures model was used to examine the dependent variable, problem drinking, which was defined as count and dichotomous from baseline to 12 month follow-up. Sociodemographics and depressive symptoms were included as covariates. First, bivariate analyses included general linear model and chi-square tests to examine covariates by group and group by problem drinking outcomes. All significant covariates were included in the GLIMMIX and GENMOD models. Then, multivariate analysis included mixed models with Generalized Estimation Equations (GEEs). The effect of group, time, and the interaction effect of group by time were examined after controlling for covariates. Multivariate results were inconsistent for GLIMMIX and GENMOD using Lognormal, Gaussian, Weibull, and Gamma distributions. SAS is a powerful statistical program in data analyses for longitudinal study.

Using JMP® to apply Decision Trees and Random Forests as Screening Tools for Limiting Candidate Predictors in Regression Models

Jason Brinkley

There are many techniques for evaluating candidate predictors in regression models. Some, such as stepwise regression, have been well studied and with known limitations. Others, such as shrinkage techniques (i.e. LASSO), have become increasingly popular and are showing true potential in helping to provide quality regression models for estimating effects in a multiple variable environment. Many of these techniques become difficult in the world of big data, especially if that data is long (many columns or variables) and short (fewer numbers of observations or rows). Tree based methods are a good alternative as a framework for data exploration and identification of variables to be used in building a quality regression model. Indeed tree methods provide an entirely different framework for model building that can oftentimes provide better predictions. However, when the main goal is still effect estimation, tree methods can be a very useful screening tool. This work examines the practice based on several examples using options very commonly found in both JMP® and JMP Pro software. The focus is on using EITHER single classification trees as well as so-called 'Random' or 'Bootstrap' Forests.

Using SAS to Examine Mediator, Direct and Indirect Effects of Isolation and Fear on Social Support Using Baron& Kenny Combined with Bootstrapping Methods

Abbas S. Tavakoli

This study presentation examines mediator, direct and indirect effects of isolation and fear on social support by using two methods: Baron & Kenny and Bootstrapping. This paper used a cross-sectional data from the longitudinal study randomized trial design in which 185 participants were assigned to the therapeutic group (n=93) who received by teleconference with participants interacting in real time with each other and control group (n=92) who received usual psychosocial care (any support used by the patient in the course of cancer treatment. Baron and Kenny (1986) steps and Hayes (2004) were used to examine for direct and indirect effects. Results of Baron indicated that the relationship between fear and social support was significant (c =-1.151 (total effect) (p=.0001)) and that there was significant relationship between isolation and fear (α =1.22 (p=.0001)). Also, previously significant relationship between fear and social support was not significant (c' =-.40 (direct effect) (p=.1876) when both fear and isolation were in the model. The indirect effect was -1.11 and Sobel test was significant (P=.0001). The results of bootstrapping methods indicated the direct effect wares -.41 (95% CI: -.42, -.40 for normal theory and -.41 (95% CI: -.99, .14 for percentile) and indirect effect was -1.06 (95% CI: -1.09, -1.08 for normal theory and -1.09, -1.55 for percentile). The result showed both methods had significant indirect effect.

'V' for … Variable Information Functions to the Rescue

Richann Watson; Karl Miller

There are times when we need to use the attributes of a variable within a data set. Normally, this can be done with a simple CONTENTS procedure. The information can be viewed prior to programming and then hardcoded within the program or it can be saved to a data set that can be joined back to the main data set. If the attributes are hardcoded then what happens if the data set changes structure, then the program would need to be updated accordingly. If the information from PROC CONTENTS is saved and then joined with the main data set, then this would need to be done for all data sets that need to be processed. This is where knowing your 'V' functions can come in handy. The 'V' functions can be used to return the label, format, length, name, type and/or value of a variable or a string within the data step. These functions can come quite in handy when you need to create summary statistics and if you need to perform an algorithm on a variable with a specific naming convention.

When ANY Function Will Just NOT Do

Richann Watson; Karl Miller

Have you ever been working on a task and wondered whether there might be a SAS® function that could save you some time? Let alone, one that might be able to do the work for you? Data review and validation tasks can be time-consuming efforts. Any gain in efficiency is highly beneficial, especially if you can achieve a standard level where the data itself can drive parts of the process. The ANY and NOT functions can help alleviate some of the manual work in many tasks such as data review of variable values, data compliance, data formats, and derivation or validation of a variable's data type. The list goes on. In this poster, we cover the functions and their details and use them in an example of handling date and time data and mapping it to ISO 8601 date and time formats.

Where Should I Dig? What to do Before Mining Your Data

Stephanie Thompson

Data mining involves large amounts of data from many sources. In order to successfully extract knowledge from data, you need to do a bit of work before running models. This paper covers selecting your target and data preparation. You want to make sure you find golden nuggets and not pyrite. The work done up front will make sure your panning yields results and is not just a trip down an empty shaft.

Windows and Unix Computers Now Have Multiple CPU's; Why Not Control Two or Three or More Parallel Executing SAS Batch Jobs from One Master Job!

William E. Benjamin Jr.

This paper will show the processes required to start multiple, parallel batch copies of SAS on the SAME Microsoft Windows or UNIX based computer. It may be easy to click the SAS Short-cut button or type SAS in a command prompt window a few times, but those jobs are independent of each other, and use resources to update the screen images for the operator (you). The process proposed here allows one SAS job to start others as child batch processes, with no screen update requirements, with their own parameters and monitor their completion. This paper exposes some of the tricks required to get two (or more) batch copies of SAS running on a Windows or Unix computer.

Working with Character Data

Andrew T. Kuligowski

The DATA Step allows one to read, write, and manipulate many types of data. As data evolves to a more free-form state, SAS's ability to handle character data becomes increasingly important. This presentation will address character data from multiple vantage points. For example, what is the "default" length of a character string, and why does it appear to change under different circumstances? What sort of formatting is available to us for character data? How can we examine and manipulate character data? This paper will aim at a beginning to intermediate audience, and will stress an introduction to the numerous character functions available in SAS®, including the basic LENGTH, and SUBSTR, plus many others.