

SESUG 2018 Abstracts - by Track/Section Application/Macro Development

| SESUG Paper # | Title | Primary Author | Abstract |
|----------------|---|---------------------|--|
| SESUG 116-2018 | Saving and Restoring Startup (Initialized) SAS® System Options | Lafler, Kirk Paul | Processing requirements sometimes require the saving (and restoration) of SAS® System options at strategic points during a program's execution cycle. This paper and presentation illustrates the process of using the OPTIONS, OPTSAVE, and OPTLOAD procedures to perform the following operations: - Display portable and host-specific SAS System options; - Display restricted SAS System options; - Display SAS System options that can be restricted; - Display information about SAS System option groups; - Display a list of SAS System options that belong to a specific group; - Display a list of SAS System options that can be saved; - Save startup SAS system options; - Restore startup SAS System options, when needed. |
| SESUG 144-2018 | Assigning agents to districts under multiple constraints using PROC CLP | Sloan, Stephen | The Challenge: assigning outbound calling agents in a telemarketing campaign to geographic districts. The districts have a variable number of leads and each agent needs to be assigned entire districts with the total number of leads being as close as possible to a specified number for each of the agents (usually, but not always, an equal number). In addition, there are constraints concerning the distribution of assigned districts across time zones, in order to maximize productivity and availability. Our Solution: uses the SAS/OR procedure PROC CLP to formulate the challenge as a constraint satisfaction problem (CSP), since the objective is not necessarily to minimize a cost function, but rather to find a feasible solution to the constraint set. The input consists of the number of agents, the number of districts, the number of leads in each district, the desired number of leads per agent, the amount by which the actual number of leads can differ from the desired number, and the time zone for each district. |
| SESUG 158-2018 | Using SAS Macro Functions to Manipulate Data | Cochran, Ben | The SAS DATA step has the reputation for being one of the best data manipulators in the IT world. While the author of this paper agrees with this statement, it is possible to go beyond the capabilities of the DATA step by using SAS Macro functions. It would be difficult to show the full power of these Macro Functions in an hour presentation, so, this paper will look at a few commonly used Macro Functions and compare and contrast them to DATA step functions. These functions can be used not only to manipulate data, but to manipulate entire programs as well. |
| SESUG 159-2018 | Create a SAS Program that Can Read the Contents of a Directory | Cochran, Ben | On occasions, a SAS user might find themselves in the position where they need to write a SAS® program that can read and process files in a specific directory. In this case, the contents are all excel spreadsheet files. All these files need to be read and converted into SAS datasets. This paper illustrates how to do this in a step by step process using the DATA step. |
| SESUG 175-2018 | Using PROC FORMAT to Automate Data correction process. | Dalvi, Shreyas | SAS has exceptional analytics capabilities, but to process data we often need to extract, transform, validate and correct the data that we get from various sources to make best use of its capabilities. Suppose in an application, we process customer data where we get information from data entry monthly, with multiple records having data entry errors. Periodically, we need to identify and correct those entries in the final SAS dataset as part of the data validation and correction process. It is time consuming to manually update each record monthly, therefore the need for an automated process arises in order to produce a final corrected data set. This paper demonstrates how we can update only the incorrect values in a SAS data set by using the external file which provides only the corrected values (finder file). This process does not make any data merges or SQL joins for the data correction. The process will use PROC format and will create the customized formats using CNTLIN for the finder file. Proc format will create the variable to be corrected and a unique master key having several variables concatenated to avoid errors in the correction process. Using this format, code will correct the invalid values in the variable and all remaining variables will remain the same. This paper is intended for intermediate level SAS Developers who want to build data validation and data correction programs using SAS. |
| SESUG 177-2018 | A User Defined SDTM Data Quality Checking and Tracking System | Zeng, Zemin | The quality of Study Data Tabulation Model (SDTM) datasets plays a crucial role in clinical studies and regulatory submissions. Inspired by Pinnacle 21 report, a similar SDTM quality checking system with additional built-in features to effectively monitor database cleaning and SDTM mapping quality progress has been developed. The checking system consists of collecting checking rules in Excel, writing small SAS macros to check SDTM datasets, and using SAS data steps and ExcelXP tagset to generate issue summary report in Excel file. This user defined checking system can be easily maintained and expanded through SAS programs creation and modification, and can be shared from one study to another. The paper describes the development steps of the SDTM quality checking and tracking system in details. |
| SESUG 184-2018 | Knowing What You've Got | Thompson, Stephanie | Have you ever needed to use data from a large collection of files that you know nothing about? If so, this paper is for you. PROC FREQ is a handy way to look at the contents of your data but sometimes you just want to generate a summary of a bunch of tables all at one time. This paper uses a case study of over 700 tables from several databases and one simple, easy, and reusable method is presented. PROC DATASETS, PROC SQL, the MACRO facility, and even PROC IML are used. A dataset of table information and summaries of the populated variables in each table are the end result. |

SESUG 2018 Abstracts - by Track/Section Application/Macro Development

| SESUG Paper # | Title | Primary Author | Abstract |
|----------------|---|--------------------------|---|
| SESUG 197-2018 | A Sysparm Companion, Passing Values to a Program from the Command Line | Fehd, Ronald | SAS(R) software has sections in its global symbol table for options and macro variables. Sysparm is both an option and a macro variable. As an option, it can be assigned during startup on the command line; in programs, values can be assigned with the options statement; values are stored and referenced as a macro variable. The purpose of this paper is to provide a general-purpose program, parse-sysparm.sas, which provides a method of deconstructing a list of comma-separated values (csv) into separate macro variables. This is a useful method of passing a set of parameter values from one program to another. |
| SESUG 236-2018 | Web Scraping in SAS: A Macro-Based Approach | Duggins, Jonathan | Web scraping has become a staple of data collection due to its ease of implementation and its ability to provide access to wide variety of free data. This paper presents a case study that retrieves data from a web site and stores it in a SAS data set. PROC HTTP is discussed and the presented technique for scraping a single page is also then automated using a SAS macro. The result is a macro that can be customized to access data from a series of pages and store the results in a single data set for further analysis. The macro is designed with a built-in delay to help prevent server overload when requesting large amounts of information from a single site. The macro is designed for both academic and industry use. |
| SESUG 240-2018 | Efficient Use of Disk Space in SAS® Application Programs | Billings, Thomas | A high-level overview of managing disk space for SAS® data sets and files created by or for the SAS system. Basic housekeeping is covered: keep files that are in-use and backup or discard files that are not in use. Backup methods are discussed, including the important question whether the operating system that your SAS site runs on might change in the future, necessitating use of the special transport format for backup files. SAS procedures that are commonly used for disk file management are described: PROC DELETE, DATASETS, and CATALOG. SQL DELETE and SAS DATA step functions for file management are also discussed. File compression is a very important tool for saving disk space, and the SAS features for this are described. Logical deletion of rows in a data set can waste disk space; prototype SAS code to detect files with this condition is supplied in an appendix. Multiple SAS programming techniques that promote efficient use of disk space are described, as well as suggestions for managing the SAS WORK library. |
| SESUG 251-2018 | Life in the Fast Lane: SAS Macro Language with Parallel Processing | Rabb, Merry | The SAS Macro language is widely used for implementing programs that are reusable, flexible and easily repeatable. When moving to a SAS Grid you gain the ability to run independent tasks within your SAS program in parallel in order to shorten run time. If those programs are built around SAS Macros used as code modules and SAS Macro variables used as parameters to control execution, some changes to or restructuring of your code may be needed to duplicate the program logic in a distributed environment. When using Macro programming in conjunction with RSUBMIT blocks for parallel processing, the code inside a RSUBMIT block is executed in a remote session, but macro code can be compiled in the main SAS program task. If a block of code that you plan to submit to a remote session contains macro calls, macro programming logic or even just macro variable references, some thought needs to be given to how and when the macro processing should occur. This paper looks at methods for incorporating SAS Macro programming logic when using parallel processing. |
| SESUG 253-2018 | Case Study: Using Base SAS to Automate Quality Checks of Excel Workbooks that have Multiple Worksheets | Mendez, Lisa | This case study provides a real world example of how Base SAS was used to read in over 185 Excel workbooks to check the structure of over 10,000 worksheets – and to repeat the process quarterly. It will illustrate how components such as the LIBNAME XLSX Engine, PROC SQL (to create macro variables), SAS Dictionary Tables, and SAS Macros were used together to create exception reports exported to MS Excel workbooks. The structure of the worksheets, such as worksheet names and variable names, were checked against pre-loaded templates. Values within the worksheets were also checked for missing and invalid data, percent differences of numeric data, and ensuring specific observations were included in specific worksheets. This case study describes the process from its inception to the ongoing enhancements and modifications. Follow along and see how each challenge of the process was undertaken and how other SAS User Group conference proceeding papers contributed to this quality check process. |
| SESUG 270-2018 | Synchronized tracking of dataset versions with programs and logs using Proc PRINTTO and other SAS tricks. | Ganapathi, Laxminarayana | It is common to generate several versions of output data in a production environment. Quite often, the most important information, the program and its version that generated the data is documented manually. We describe an automated process that tracks each data set with a specific version of the code as well as the log file generated in the specific run. We will also demonstrate other uses of proc PRINTTO in tracking the versions of the production runs to make meaningful documentation of code and data in a production environment. Use of automatic macro variables generated by the system increases the reliability of versioning and audit trails. The process is suited for tracking minor changes implemented during a production run. It implicitly takes care of documenting any changes in the inputs to the program. Hence changes affected by running the same code in a changed environment can be dynamically documented. We will demonstrate the use of the process in a Linux as well as windows environment. |

SESUG 2018 Abstracts - by Track/Section Application/Macro Development

| SESUG Paper # | Title | Primary Author | Abstract |
|----------------|--|---------------------|---|
| SESUG 282-2018 | Quick 'n Dirty - Small, Useful Utility Macros | Droogendyk, Harry | Macros are useful to define “canned” code that may be made available to other users in your organization. But, there’s also simple, less sophisticated macros that are useful in your day to day coding efforts, if only to reduce keystrokes. This presentation will demonstrate a few of these macros, even ones that generate only partial statements: %dups generate code snippet to identify duplicate observations %fiscal return formatted SAS dates offset by fiscal year %cleanup cleanup work datasets and/or global macro variables %single resolve macro variables within single quotes |
| SESUG 283-2018 | Building Neural Network model in BASE SAS ® (From Scratch) | Mandalapu, Soujanya | Artificial Neural Networks (ANNs) are extremely popular in deep learning applications such as image recognition and natural language processing. ANNs are also being implemented in finance, marketing, and insurance domains. Most neural network models are implemented in Python, Java, C++, or Scala. Although Base SAS is a preferred language in regulated environments such as finance and clinical trials, it cannot be used to implement ANN models. This brings difficulties for financial modelers who want to use ANNs to improve their models to gain efficiency. This paper aims at those modelers who would like to implement machine learning models using only Base SAS and SAS macros. A standard three-layer (one input, one hidden and one output) feed forward backward propagation algorithm with three separate macros (forward propagation, backward propagation and to control the number of iterations) for each repetitive step was implemented in this paper. This algorithm is scalable to increase as many features and hidden nodes. |
| SESUG 288-2018 | Sorting Arrays Using the Hash Object | Dorfman, Paul | Before the advent of the SAS Hash Object, two options had been available for sorting SAS arrays: (1) using the SORTN or SORTC call routines or (2) implementing a sorting algorithm, such as the quick sort, heap sort, etc. The first option is limited by its dependence on the data type and inability to handle duplicate keys and parallel arrays. The second option requires rather sophisticated custom programming. The SAS hash object, with its built-in ability to sort its hash items internally, is devoid of the above deficiencies. In this paper, we show how it can be used to sort SAS arrays simply and efficiently. |

SESUG 2018 Abstracts - by Track/Section Building Blocks

| SESUG Paper # | Title | Primary Author | Abstract |
|----------------|--|-------------------|--|
| SESUG 107-2018 | Mining Bitcoins: A Step-by-Data Step Simulation | Hoffman, Seth | Have you ever thought about investing in a crypto-currency, but first wanted to understand more about how they work? The best way to understand a computer program is to read its software code and see it run. This paper walks through a Base SAS(R) simulation of a Bitcoin miner to demonstrate all the parts needed to implement the Bitcoin payment system. |
| SESUG 110-2018 | Introduction to Data-driven Programming Techniques Using SAS® | Lafler, Kirk Paul | Data-driven programming, or data oriented programming (DOP), is a specific programming paradigm where the data, or data structures, itself controls the flow of a program and not the program logic. Often, data-driven programming approaches are applied in organizations with structured data for filtering, aggregating, transforming and calling other programs. Topics include how SAS® users can access metadata content to capture valuable information about the librefs that are currently assigned, the names of the tables available in a libref, whether a data set is empty, how many observations are in a data set, how many character versus numeric variables are in a data set, a variable's attributes, the names of variables associated with simple and composite indexes, access the content of read-only SAS metadata data sets called DICTIONARY tables or their counterparts, SASHELP views, and how SAS metadata can be dynamically created using data-driven programming techniques. |
| SESUG 117-2018 | A Beginner's Babbelfish: Basic Skills for Translation Between R and SAS® | Woodruff, Sarah | 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 |
| SESUG 131-2018 | A Quick Bite on SAS(r) Studio Custom Tasks | Kola, Michael | Custom Tasks in SAS Studio help users generate reports through a point and click user interface (UI). As a developer, there will be a need to create custom tasks for non-SAS programmers as well as novice SAS users or programmers. After creating a custom task, it can be shared with stakeholders. Custom Tasks are saved as CTM (Custom Task Model) files or CTK (Custom Task Kit) files. CTK file is acting as front-end GUI application, and CTM file are back-end files. Both CTM and/or CTK files could be shared through email or by just placing it on a network shared area. The user simply has to run the file and it generates the User Interface, as the user navigates through the UI, the SAS code is generated simultaneously, which will help the user review the SAS code related to the UI. The task framework is flexible. All tasks use the same common task model and the Velocity Template Language. The task consists of task elements and they are: Registration, Metadata, UI, Dependencies, Requirements and Code Template. We will use all these elements to build our example custom task. |
| SESUG 178-2018 | Twenty ways to run your SAS program faster and use less space | Sloan, Stephen | 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. |
| SESUG 179-2018 | Table Taming – Using ODS to Simplify Data Presentation | Perry, Emily | PROC REPORT is a powerful tool, but may be much more than you need. It contains twelve statements and pages of options, making it difficult to display exactly what you want in exactly the way you want it. Consider an alternative method of implementing ODS TRACE and ODS OUTPUT. These SAS statements can be used to output the results of almost any PROC step into a dataset that can then be manipulated to your heart's content. You have more control over your analysis with a straightforward and intuitive method. In this article, we will discuss using these ODS methods in conjunction with typical PROC steps, such as PROC MEANS and PROC FREQ, along with data manipulation to create tables that you can be proud of. |

SESUG 2018 Abstracts - by Track/Section Building Blocks

| SESUG Paper # | Title | Primary Author | Abstract |
|----------------|--|------------------|---|
| SESUG 192-2018 | Emulating FIRST. and LAST. SAS® DATA Step Processing in SQL? Concepts and Review | Billings, Thomas | The questions of whether and how FIRST. and LAST. BY variable SAS® DATA step processing can be emulated in SQL qualify as FAQs: frequently asked questions. We begin with an overview of terminology and explain the different environments involved. Then we cover the common approaches to this issue. #1: if the data rows are uniquely ordered by existing variable(s), those can be used to support emulation. #2: in SAS PROC SQL, DATA step views can be used to inject the values of FIRST. and LAST. BY variables; alternately the undocumented SAS function monotonic() can be used to inject a row number into the data, which can be used in approach #1. However as the order of pull can change from run to run in SQL, the results with monotonic() are not reproducible, hence problematic for data validation and audits. #3: most other SQL dialects have non-standard SQL extensions that will supply a row number; again the order of pull can change from run to run. Then we discuss whether this type of emulation makes sense; that is why the article title includes a question mark (?). We discuss some non-SQL alternatives, and specify the very narrow contexts where the emulation is reasonable. Finally, we note that DS2 supports BY, FIRST., LAST. processing and it can run in-database, making it a possible alternative to SQL and the DATA step. |
| SESUG 196-2018 | An Autoexec Companion, Allocating Location Names during Startup | Fehd, Ronald | Like other computer language SAS(R) software provides a method to automatically execute statements during startup of a program or session. This paper examines the names of locations chosen in the filename and libname statements and the placement of those names in options that enable all programs in a project to have standardized access to format and macro catalogs, data sets of function definitions and folders containing reusable programs and macros. It also shows the use of the global symbol table to provide variables for document design. The purpose of this paper is to examine the default values of options, suggest naming conventions where missing, and provide both an example autoexec and a program to test it. |
| SESUG 202-2018 | Dating for SAS Programmers | Horstman, Josh | 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? |
| SESUG 221-2018 | When Reliable Programs Fail: Designing for Timely, Efficient, Push-Button Recovery | Hughes, Troy | Software quality comprises a combination of both functional and performance requirements that together specify not only what software should accomplish, but also how well it should accomplish it. Recoverability—a common performance objective—represents the timeliness and efficiency with which software or a system can resume functioning following a catastrophic failure. Thus, requirements for high availability software often specify the recovery time objective (RTO), or the maximum amount of time that software may be down following an unplanned failure or a planned outage. While systems demanding high or near perfect availability will require redundant hardware, network, and additional infrastructure, software too must facilitate rapid recovery. And, in environments in which system or hardware redundancy is infeasible, recoverability only can be improved through effective software development practices. Because even the most robust code can fail under duress or due to unavoidable or unpredictable circumstances, software reliability must incorporate recoverability principles and methods. This text introduces the TEACH mnemonic that describes guiding principles that software recovery should be timely, efficient, autonomous, constant, and harmless. Moreover, the text introduces the SPICIER mnemonic that describes discrete phases in the recovery period, each of which can benefit from and be optimized with TEACH principles. Software failure is inevitable but negative impacts can be minimized through SAS® development best practices. |
| SESUG 255-2018 | Tips and Tricks for Ensuring Accurate Routine Data Processing when Logs are Redirected to Text Files | Moyses, Davia | SAS users often develop and maintain a set of codes that process the same expected inputs through to expected outputs and that are run on a routine basis as datasets are updated with new observations. A common efficiency implemented in this data management scenario is to redirect SAS log contents to a permanent text file, rather than display in the terminal log window. This efficiency does provide processing time savings, but may lead to oversight of processing errors when inputs or outputs stray from their expected formats. This paper will discuss techniques that may be implemented to aid a SAS user in reviewing redirected processing logs and the success of processing, especially when no errors or warnings present themselves. Topics will include automatically opening and programmatically reviewing the redirected log text file and checking the intended structure of input, intermittent, and output datasets. Focus will be given to directing some messages back to the terminal log window in the middle of processing, as well as ensuring that all observations are processed through conditional IF-ELSE IF processing. These techniques utilize PROC CONTENTS, PROC SQL, CALL SYMPUT, %SYSEXEC, and %SYSFUNC. |

SESUG 2018 Abstracts - by Track/Section Building Blocks

| SESUG Paper # | Title | Primary Author | Abstract |
|----------------|---|--------------------|--|
| SESUG 266-2018 | Proc SQL in Twenty Minutes | Cohen, John | Proc SQL can be a valuable addition to your SAS tool set. While the syntax will seem foreign to experienced Data Step programmers, the Structured Query Language can be quite elegant in its own right. Further, unlike a requirement for us to “computer engineer” our Data Step for efficiency, our friends in Cary have built-in intelligence to allow us to submit generic SQL and optimization under the covers does at least some of the rest. The SQL Procedure makes certain tasks easier (such as Cartesian Products), is a requirement for pulling data from many external databases (to link via “proc SQL pass-through” to link to Oracle, Teradata, MS/Access, and the like), and for many is a preferred tool for common tasks (such as capturing record counts, accessing SAS Dictionary Tables, creating views and indexes, and capturing summary statistics for loading into SAS macro variables). Finally, in today’s complex IT environments, having some familiarity with SQL will allow one to better engage in this broader environment. We will not turn you into an overnight expert, but we will lay a foundation for you to continue to explore and learn on your own. |
| SESUG 284-2018 | Hash Beyond Lookups - Take Another Look | Axelrod, Elizabeth | Using the hash object in SAS – beyond table lookup and data retrieval – enables you to do some pretty cool stuff, solving complicated problems with code that runs efficiently and flows intuitively. And... it’s easy to maintain and modify. What’s not to like? This is an updated and expanded version of my original paper “Hash Beyond Lookups - Your Code Will Never Be the Same”. It provides a further exploration of some useful HASH techniques, and offers more situations that can be elegantly solved using Hash. The main message? Don’t be afraid of Hash! |
| SESUG 285-2018 | The Categorical Might of PROC FREQ | erinjeri, jinson | PROC FREQ is essential to anybody using Base SAS® for analyzing categorical data. This paper presents the various utilities of the FREQ procedure that enable in effective data analysis. The cases presented include a range of utilities such as finding counts, percentages, unique levels or records, Pearson chi-square test, Fisher’s test, McNemar test, Cochran-Armitage trend test, binomial proportions test, relative risk, and odds ratio. In addition, this paper will show the ODS features available to effectively display your results. All the cases presented in this paper will prove the categorical might of PROC FREQ beyond doubt. |

SESUG 2018 Abstracts - by Track/Section Coder's Corner

| SESUG Paper # | Title | Primary Author | Abstract |
|----------------|---|-----------------------|--|
| SESUG 114-2018 | A Visual Step-by-step Approach to Converting an RTF File to an Excel File | Laffler, Kirk Paul | Rich Text Format (RTF) files incorporate basic typographical styling and word processing features in a standardized document that many programs and applications are able to read. In today's high-tech arena sometimes the contents of an RTF file needs to be viewed as, and even converted to, an Excel file. You would think that since both RTF and Excel are Microsoft standards that this would be a simple process to achieve, but you may be surprised to find out that it is not. This paper and presentation highlights a visual step-by-step approach to converting an RTF file to an Excel File using the SAS® software. |
| SESUG 122-2018 | When SAS Doesn't Behave As Expected | McNeill, Megan | SAS does not always behave in a way a programmer would expect. Sometimes this is due to errors in programming, while other times this is due to how SAS processes data and performs computations. This paper will provide two examples of instances where SAS does not provide the result a programmer was expecting, and give tips on how to program defensively in order to avoid such issues. One example will highlight a case where the result computed from SAS's internal process differs from the mathematically correct result. The other example will touch on a broader issue of SAS processing date variables differently between a data step and a macro. This paper will cover different processing methods and the importance of knowing how SAS behaves in these situations. |
| SESUG 138-2018 | Easily Updating Programs with the SAS® Macro Language: an EDFacts Example | Brown, Aaron | Many times, a programmer may have to generate the same report several times, such as weekly, monthly, or annually. If these reports have minor changes, making the necessary updates can be time-consuming and cause bugs or errors due to typos. This paper describes examples of how to easily update programs by using the SAS® macro language, using the example of the annual EDFacts data submissions (which state Departments of Education must are required to send to the federal Department of Education). This paper utilizes both macro variables (using the %let statement) and macros. |
| SESUG 140-2018 | Matching SAS Data Sets with Hash Objects: If at First You Don't Succeed, Match, Match Again | Go, Imelda | Two data sets can be matched by using a number of techniques available in SAS. If the goal is to match as many records as possible between two data sets, then the matching process may have to be repeated several times between the two data sets by using different sets of matching criteria. Although programmers new to hash objects will need to learn a few things before they can feel confident about using the technique, the resulting ease of use/maintenance of the code is well worth the effort. {This is the hash-object version of two prior papers on the same topic: the 2009 SESUG paper focused on PROC SQL and the 2004 SESUG paper focused on the DATA step.} |
| SESUG 148-2018 | Using a Picture Format to Create Visit Windows | Watson, Richann | 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. |
| SESUG 154-2018 | Improving Query performance by saying "NO" to Heterogeneous Joins | Pothuraju, Suryakiran | Heterogeneous joins between SAS dataset and DBMS table is a common requirement most SAS programmers often come across. SAS/ACCESS support for DBMS temporary tables has the ability to retain DBMS temporary tables across multiple SAS steps. If you can make the process happen in the DBMS in some certain situations, the overall performance will enhance immensely. In this paper we will discuss how to access small subset of data from extremely large Database table, depending on the key variables from a SAS Table. |
| SESUG 155-2018 | You're Doing It Wrong! Volume 001 | Rosanbalm, Shane | 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 001 will be on the appropriate naming of macro parameters. |

**SESUG 2018 Abstracts - by Track/Section
Coder's Corner**

| SESUG Paper # | Title | Primary Author | Abstract |
|----------------|--|---------------------|---|
| SESUG 168-2018 | "Use SAS® PROC Tabulate to create a PDF with Bookmarks" | Chien, Chuchun | Recently our client requested a report of cross frequencies containing many variables to compare and review their data. Our original thought was to run a cross frequency however this created a very long output. While this initial output met the client's request, the information was difficult to review and did not contain a defined structure. It was determined the final report needed to provide a structure to the end user which would allow easy access to the comparisons within the report and have a finished appearance. We discovered that you can do so much more to improve the look and readability of the report by using ODS and SAS® Proc tabulate. This paper will explain our progression from a plain SAS® output of a simple cross frequency to a PDF and then to a more user friendly PDF with meaningful bookmarks using macros, SAS® ODS and SAS® Proc Tabulate. We will show you how to take your simple Cross Frequencies beyond simple output to this next level of a finished report. |
| SESUG 189-2018 | Automated Comparison Tables using PROC REPORT | Kwiatkowski, Evan | Comparison tables showing rows of summary statistics for continuous and categorical variables (e.g. age and gender) across columns of a categorical variable (e.g. treatment vs. placebo) are ubiquitous in clinical research. Many SAS procedures are used to generate these data summaries and associated tests of statistical significance (e.g. PROC FREQ, PROC TTEST). Other pertinent information, such as formatted variable names and number of non-missing observations per variable, are often included in the final report. This paper presents a one-step macro for producing reports that contain summary statistics, tests for significance, and other relevant information in a comparison table in RTF format. |
| SESUG 193-2018 | Password Protection of SAS® Enterprise Guide® Projects | Billings, Thomas | SAS® Enterprise Guide® lets a user assign a password to a project, after which any/all users must enter the correct password to be able to open and run the project. In this paper we test the security of password protected projects against 2 hacks. We first present a sample project that has multiple types of Tasks that will be used for testing. Before assigning a password to the project, we demonstrate a simple hack that facilitates the examination of some of the internal metadata of a SAS Enterprise Guide project, i.e., a .egp file. Then we assign a password to the sample project and repeat the hack on the password-protected project, with the result that the hack does not work. Next, we try another simple hack - viewing the .egp file in a text editor - and find that this does not work either. The tests here provide independent confirmation that password-protected .egp files are secure against the 2 hacks demonstrated. We end with a brief discussion on .egp file metadata and its role in the SAS ecosystem. SAS products: SAS Enterprise Guide. User-level: beginner. |
| SESUG 194-2018 | Do the Loop De Loop ! | Lindquist, Jennifer | Like the break neck speed of a rollercoaster careening through loops, barreling through programs using SAS® DO loops and arrays instead of repetitive statements can shorten your programming time and lines of code. This introduction to arrays and loop processing covers the basics - starting with the syntax of each and preceding to examples. By combining the two concepts you can construct a simple iterative DO loop to traverse an array with ease. Make going around in circles work for you! |
| SESUG 213-2018 | Accessing Password Protected Microsoft® Excel Files in A SAS® Grid Environment | Welch, Brandon | Microsoft® Excel continues as a popular choice for data storage and manipulation. Extracting data from Excel files is often challenging. If a file is password-protected, the challenge intensifies. DDE (Dynamic Data Exchange) is the classic approach to reading password-protected Excel files. The DDE approach fails in a SAS grid environment. This paper illustrates an alternative using a Visual Basic Scripting (VBScript) file. The script is built within a SAS program and executed with a X command, then the text file is accessed via PROC IMPORT. The methods presented in this paper highlight SAS techniques that will educate programmers at all levels. In addition, the reader will learn how to build a simple VBScript file. |

**SESUG 2018 Abstracts - by Track/Section
Coder's Corner**

| SESUG Paper # | Title | Primary Author | Abstract |
|----------------|---|------------------|--|
| SESUG 218-2018 | Using PROC SQL to Generate Shift Tables More Efficiently | Cody, Jenna | Shift tables display the change in the frequency of subjects across specified categories from baseline to post-baseline time points. They are commonly used in clinical data to display the shift in the values of laboratory parameters, ECG interpretations, or other categorical variables of interest across visits. The "into:" statement in PROC SQL can be used to create macro variables for the denominators used in these tables. These macro variables can be accessed throughout the program, allowing for easy computation of percentages and the ability to call the same macro variable to display the subject count value in the header. This paper outlines the steps for creating a shift table using an example with dummy data. It describes the process of creating macro variables in PROC SQL using the "into:" step, creating shift table shells using the DATA step, conducting frequency tabulations using PROC FREQ, calling the macro variables to calculate and present the count and percent, and using the macro variables for the subject count value in the headers. It then discusses the efficiency of the use of PROC SQL to create macro variable denominators over other methods of calculating denominators, such as in the PROC FREQ step. Code examples are provided to compare shift table generation techniques. |
| SESUG 224-2018 | Abstracting and Automating Hierarchical Data Models: Leveraging the SAS® FORMAT Procedure CNTLIN Option To Build Dynamic Formats That Clean, Convert, and Categorize Data | Hughes, Troy | The SAS® FORMAT procedure "creates user-specified formats and informats for variables." In other words, FORMAT defines data models that transform (and sometimes bin) prescribed values (or value ranges, in the case of numeric data) into new values. SAS formats facilitate multiple objectives of data governance, including data cleaning, the identification of outliers or new values, entity resolution, and data visualization, and can even be used to query or join lookup tables. SAS formats are often hardcoded into SAS software, but where data models are fluid, formats are best defined within control files outside of software. This modularity—the separation of data models from the programs that utilize them—allows SAS developers to build and maintain SAS software independently while domain subject matter experts (SMEs) separately build and maintain the underlying data models. Independent data models also facilitate master data management (MDM) and software interoperability, allowing a data model to be maintained as a single instance, albeit implemented not only with SAS but also Python, R, or other languages or applications. The CNTLIN option (within the SAS FORMAT procedure) facilitates this modularity by creating SAS formats from data sets. This text introduces the BUILD_FORMAT macro that greatly expands the utility of CNTLIN, allowing it to build formats not only from one-to-one and many-to-one format mappings but also from multitiered, hierarchical data models that are built and maintained externally in XML files. The numerous advantages of BUILD_FORMAT are demonstrated through successive SAS code examples that rely on the taxonomy of the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5). |
| SESUG 233-2018 | Recursive Programming Applications in Base SAS® | erinjeri, jinson | Programmers employ recursive programming when faced with tasks which are structured hierarchically. Recursive programming involves the call of the same program or function within itself to solve such tasks. For the recursive program to execute effectively, it requires an entry condition, a recursive sequence and an exit condition. Recursive programming can be employed in base SAS® and this paper presents some of the cases where recursive programming can be applied. In addition, the cases presented in the paper can help instill a programmer of how and when to take the recursive approach while programming. |
| SESUG 235-2018 | Investigation of Impact on Tobacco Use due to Vaporizing Nicotine in High Schools | Williams, Justin | Approximately 431,000 deaths each year are attributed to nicotine use in the United States. This is directly responsible for more than US \$100 billion in healthcare and indirect costs. Nicotine is very addictive. However, nicotine has evolved to be consumed in many forms, one of which is in the form of vapor (vaping). Nicotine vapor consumption has become somewhat of a social trend amongst high school students, due to the assumption that consuming nicotine through vaporization is less hazardous to their health than smoking cigarettes. This paper will demonstrate the use of the Chi-Square Test, T-Test, and Regression Analysis in SAS to examine the phenomenon of vaping amongst the teenage population. |

SESUG 2018 Abstracts - by Track/Section Coder's Corner

| SESUG Paper # | Title | Primary Author | Abstract |
|----------------|--|---------------------|--|
| SESUG 238-2018 | A Macro to Add SDTM Supplemental Domain to Standard Domain | Kunwar, Pratap | Many pharmaceutical and biotechnology industries are now preferring to setup Study Data Tabulation Model (SDTM) mapping in the beginning of the study rather at the end, and use SDTM datasets to streamline the flow of data from collection through submission. With having SDTM datasets in disposal, it is a logical choice to use them for any clinical reports. Getting information from supplemental (SUPP) domain back to parent domain is a regular step that programmers can't avoid. But, this step can be very tricky when either (1) SUPP domain contains multiple types of identifying variables, or (2) SUPP domain empty or does not exist. In this presentation, I will present an easily understandable macro that will produce correct results in every possible scenario. |
| SESUG 239-2018 | SAS Generated Dashboard - Useful Tool for A SAS Programmer | Kunwar, Pratap | Being savvy in SAS technical skills is a key to become a top SAS programmer, but at the same time it is important to possess other supporting tools that can provide related project and other information instantly during programming. A SAS generated HTML based dashboard page with relevant information and instant navigation to such as project folders, people, other documents and any related html links, is highly desirable, resulting in saving time during the work day. In this presentation, first I will share my dashboard that I have generated using SAS. Second, because of its popularity among my colleagues and non-programmers, I will show how I have setup a SAS program for non-programmer so dashboard can be generated from Excel with SAS interface, not necessarily even opening SAS in interactive mode. |
| SESUG 242-2018 | Order, Order! Four Ways to Reorder Your Variables, Ranked by Elegance and Efficiency | Hadden, Louise | SAS(r) practitioners are frequently required to present variables in an output data set in a particular order, or standards may require variables in a production data set to be in a particular order. This paper and presentation offer several methods for reordering variables in a data set, encompassing both data step and procedural methods. Relative efficiency and elegance of the solutions will be discussed. |
| SESUG 244-2018 | Divide and Conquer: A Macro to Split Data Based on Duplicate Values | Tayshetye, Meredith | Have you ever had a dataset that needed to be split into smaller datasets so that one variable did not have duplicate values? An email program the team was using did not allow more than one email to an address within the same send. So, if one email address was listed multiple times for multiple records, the program would only send an email to one record, excluding the others. As a workaround, team members would manually copy duplicates into different files and treat each duplicate file as a separate send. While we could not solve for the limitations of the email program, we could make it easier to split the files. Using SAS Enterprise Guide 7.13, we have created a macro that solves this problem. The macro will determine the maximum number of times any particular value appears and will create that many files. Though an email address can be linked to multiple records, you may not initially know the maximum number of times an address appears. For example, if a file has one email address linked to ten different records, there will be ten output files. The macro will automatically determine the maximum number of files needed. Then, it will write out the necessary number of files to a specified directory where they can be used for other purposes such as an input to the email program. This macro is customizable to a variety of data and industries. It is easy to use and can save companies valuable time and manual work. |
| SESUG 256-2018 | Quote the SASLOG | Kuligowski, Andrew | "For every action, there is an equal and opposite reaction." Sir Isaac Newton was talking about physics, but the concept also applies to other areas of life - including quotation marks in SAS code. Unfortunately, SAS coders do not always properly provide balanced quotation marks while coding. SAS will detect possible occurrences of this problem, signaling its concern with a SASLOG message: WARNING: The quoted string currently being processed has become more than 262 characters long. You might have unbalanced quotation marks. This presentation contains a few coding suggestions to identify unbalanced quotation marks in SAS code. |

SESUG 2018 Abstracts - by Track/Section Coder's Corner

| SESUG Paper # | Title | Primary Author | Abstract |
|----------------|--|-------------------|--|
| SESUG 261-2018 | Tips for Pulling Data from Oracle® Using PROC SQL® Pass-Through | Cohen, John | <p>For many of us a substantial portion of our data reside outside of SAS®. Often these are in DBMS (Data Base Management Systems) such as Oracle, DB2, or MYSQL. The great news is that the data will be available to us in an already-structured format, with likely a minimum of reformatting optimized database effort required. Secondly, these DBMS' come with an array of manipulation tools of which we can take advantage. The not so good news is that the syntax required for pulling these data may be somewhat unfamiliar to us. We will offer several tips for making this process smoother for you, including how to leverage a number of the DBMS tools. We will take advantage of the robust DBMS engines to do a lot of the preliminary work for us, thereby reducing memory/work space/sort space, data storage, and CPU cycles required of the SAS server – which is usually optimized for analytical work while being relatively weaker (than the DBMS) at the heavy lifting required in an increasingly Big Data environment for initial data selection and manipulation. Finally, we will make our SAS Administrators happy by reducing some of the load in that environment.</p> |
| SESUG 267-2018 | PROC SQL - GET "INTO:" IT! | Schlessman, Kelly | <p>The "INTO:" clause within PROC SQL is an extremely useful tool, but may be a mystery to those unfamiliar with SQL. A SELECT statement paired with the INTO clause in PROC SQL provides a simple method of storing data in a macro variable - or many macro variables! What can go into these macro variables? Values of selected variables from a dataset, summary statistics, and delimited lists of values, to name a few. A task that may require multiple steps in traditional DATA STEP programming can be achieved in a single line of PROC SQL code. All that is required to harness the power of the INTO clause is basic knowledge of SELECT statements in PROC SQL. This paper will provide an introduction to the INTO clause for novice PROC SQL users, and demonstrate some useful ways in which it can be put to work in your code.</p> |
| SESUG 268-2018 | A beginner's take on conditional logic using SAS Enterprise Guide 7.15 and SAS Macros. | Walker, Mykhael | <p>Learning any new system can be a daunting task, but learning a system that has multiple ways to solve the same problem can be refreshing and frustrating at the same time. When I was introduced to SAS Enterprise Guide 7.15, I was overwhelmed by all the features that were available. Once the initial shock subsided, it was time to put the tools to work. A task that commonly comes up when trying to solve a problem is the scenario: If A happens, then do B...conditional logic. SAS Enterprise Guide 7.15 provides a way for the user to manipulate conditional logic in a visual way and SAS code provides multiple ways to manipulate conditional logic via code. In this paper I will discuss how I used the conditional logic capability built in SAS Enterprise Guide 7.15 to solve an issue. I will also discuss how I used macros in SAS code to solve the same issue, but with a little more flexibility.</p> |
| SESUG 273-2018 | Automation Methods: Using SAS to Write the Code for Repetitive Quality Control Checks | Go, Imelda | <p>This paper illustrates how you can use a SAS data set, which we will call a quality control specifications (QCS) data set, to store the information on how each variable in the data set will be scrutinized via QC methods. For example, we will need to run PROC FREQ for variables A to Z and run a PROC MEANS for variables X1-X100. Let us suppose we have an existing data set we need to run quality control checks on. We first apply PROC CONTENTS to this data set to get the complete list of variables in the data set. We then focus on adding to this data set the variables that identify which QC methods apply to each variable. Once this information has been added to the data set, this becomes our QCS data set. We then use SAS to process the information in it so that SAS creates the programming statements required to execute the QC checks. This simplifies the task tremendously for the programmer because the programmer can focus on making sure that each variable has been marked with all the required QC checks and be assured that SAS will create the code for you.</p> |

**SESUG 2018 Abstracts - by Track/Section
Coder's Corner**

| SESUG Paper # | Title | Primary Author | Abstract |
|--------------------------|--|---------------------------|---|
| SESUG 274-2018 | Automation Methods: Using SAS to Write PROC SQL Joins for You | Go, Imelda | When the number of variables involved is large and there are a number of aliases, it can actually take some time to type the join code even though the task is inherently simple. We first apply PROC CONTENTS to the two data sets to be joined in order to get a full list of variables for each data set. We then focus on adding to each data set indicator variables that show us which variables are to be included in the join. We can also specify aliases, identify which variables will be used for the join, and in general, specify other information that could be useful for the join. We then use SAS to process the information in these data sets so that SAS creates the programming code required for the join. This method allows the programmer to focus on specifying the features of the join instead of focusing on typing the SAS code flawlessly. |
| SESUG 301-2018 | Panel Discussion: Tip Top Top Tier SAS® Tips: What makes a great SAS tip | Hadden, Louise | Tip Top Top Tier SAS® Tips: What makes a great SAS tip Four panelists and a moderator |

SESUG 2018 Abstracts - by Track/Section Data Management/Big Data

| SESUG Paper # | Title | Primary Author | Abstract |
|----------------|--|-------------------|---|
| SESUG 118-2018 | Reducing the Length of Character Variables in a SAS® Data Set | Gilsen, Bruce | In the big data era, reducing the defined length of character variables to their actual maximum length in any observation is one way to reduce disk storage use and improve processing time with no loss of information. A simple way to do this is with the FIXLENG macro, which determines the maximum length for each character variable, and if any character variables have a maximum length smaller than their defined length does the following: 1. Prints a summary report listing the defined length and maximum length for each character variable. 2. Based on the value of a user-specified parameter either (1) reduces the size of all character variables to their maximum length, or (2) generates and displays SAS® code that would reduce the size of all character variables to their maximum length. FIXLENG provides a simple alternative to compression, which is commonly used to reduce disk storage. One advantage of the FIXLENG macro is that it is only executed once, whereas a data set must be compressed every time it is written and uncompressed every time it is read. |
| SESUG 139-2018 | Data Integration Best Practices | Droogendyk, Harry | The creation of and adherence to best practices and standards can be of great advantage in the development, maintenance, and monitoring of data integration processes and jobs. Developer creativity is always valued, but it is often helpful to channel good ideas through process templates to maintain standards and enhancement productivity. Standard control tables are used to drive and record data integration activity. SAS® Data Integration Studio (or Base SAS®) and the judicious use of auto-call utility macros facilitate data integration best practices and standards. This paper walks you through those best practices and standards. |
| SESUG 142-2018 | Reducing the space requirements of SAS data sets without sacrificing any variables or observations | Sloan, Stephen | 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. |
| SESUG 143-2018 | Fuzzy Matching Programming Techniques Using SAS Software | Sloan, Stephen | Data comes in all forms, shapes, sizes and complexities. Stored in files and data sets, SAS users across industries know all too well that data can be, and often is, problematic and plagued with a variety of issues. When unique and reliable identifiers are available, users routinely are able to match records from two or more data sets using merge, join, and/or hash programming techniques without problem. But, what happens when a unique identifier, referred to as the key, is not reliable or does not exist. These types of problems are common and are found in files containing a subscriber name, mailing address, and/or misspelled email address, where one or more characters are transposed, or are partially and/or incorrectly recorded? This presentation introduces what fuzzy matching is, a sampling of data issues users have to deal with, popular data cleaning and user-defined validation techniques, the application of the CAT functions, the SOUNDEX (for phonetic matching) algorithm, SPEDIS, COMPLEV, and COMPGED functions, and an assortment of programming techniques to resolve key identifier issues and to successfully merge, join and match less than perfect or messy data. |
| SESUG 152-2018 | A Shorter Path to Moving Data Profiles from SAS® Data Management Studio to SAS® Data Management Server | hileman, wayne | Typically, one uses SAS® Data Management Studio to develop and run a data profile on a much smaller subset of data than the final target data set. Once the profile is ready for prime time or you need to profile a large data set, the data profile is run on the SAS® Data Management Server through a batch job using SAS® Data Integration Studio. This presentation covers the steps to move a data profile from the user's machine to the SAS® Data Management Server and run through a DataFlux® Batch Job, in SAS® Data Integration Studio. Also covered are recommendations to follow that are helpful to know in advance, rather than hindsight, in getting my first data profile to run successfully. Before scaring you away, the intended audience is anyone who has access to SAS® Data Management Studio, SAS® Data Management Server, SAS® Data Integration Studio, and a simple data set that you would like to profile. Experience with either application is minimal. |
| SESUG 153-2018 | Obtain better data accuracy using reference tables | venna, Kiran | Data accuracy can be improved tremendously by using reference tables, especially when data is loaded from external files into target tables. Metadata and data quality checks can be easily performed with help of these reference tables. Data accuracy for many files can be performed by using a single reference SAS table with help of macro. In this paper we will discuss step by step approach how to build this reference tables and way to automatize the data accuracy checks. |

SESUG 2018 Abstracts - by Track/Section Data Management/Big Data

| SESUG Paper # | Title | Primary Author | Abstract |
|----------------|--|----------------|--|
| SESUG 171-2018 | Taking XML's Measure: Using SAS® to Read In and Create XML for Analytic Use and Websites | Hadden, Louise | XML has become a standard over the years for populating websites and transferring information. This presentation demonstrates how to parse mystery XML files, read in XML files that you can't right-click on and open with Microsoft Excel, read into Microsoft Excel using SAS®, and how to use maps and schemas to input and output various XML representations. Constructing and outputting "measure code" data sets from input data to maximize the flexibility of XML data representation and usage is discussed. Metrics comparing using the hierarchical measure code structure to standard, more rectangular XML structures are presented. |
| SESUG 222-2018 | Parallel Processing Your Way to Faster Software and a Big Fat Bonus: Demonstrations in Base SAS® | Hughes, Troy | SAS® software and especially extract-transform-load (ETL) systems commonly include components that must be serialized due to real process dependencies. For example, a transform module often cannot begin until the data extraction completes, and a corresponding load module cannot begin until the data transformation completes. While process dependencies such as these cannot be avoided in many cases and necessitate serialized software design, in other cases, programs or data can be distributed across two or more SAS sessions to be processed in parallel, facilitating significantly faster software. This text introduces the concept of false dependencies, in which software is serialized by design rather than necessity, thus needlessly increasing execution time and deprecating performance. Three types of false dependencies are demonstrated as well as distributed software solutions that eliminate false dependencies through parallel processing, arming SAS practitioners to accelerate both their software and salaries. |
| SESUG 281-2018 | How To Dupe a Dedup | Dorfman, Paul | Many ETL transformation phases begin with cleansing the extracted data of duplicate data. For example, transactions with the same key but different dates may be deemed duplicate, and the ETL needs to select the latest transaction. Usually, this is done by sorting the extract by the key and date and then choosing the most recent record. However, this technique is quite resource-costly, particularly when the non-key variables are numerous and/or long and the result needs to be resorted back into the original order. In this paper, we show how the same goal can be achieved via a principally different algorithm based on modifying the extract file by marking the duplicate records for deletion and thus obviating the need to sort the entire extract even once. For large extract files and relatively sparse duplicate records, this approach may result in cutting the processing time by an order of magnitude or more. |
| SESUG 305-2018 | Speed up your SAS processing automatically with no change to code | Rafiee, Dana | Long running SAS jobs have been a big problem for many SAS users and organizations. With absolutely no change to SAS code, Destiny will discuss business use cases and Vexata, a disruptive technology certified by SAS Institute that takes SAS data step, procedures, and model processing from many days and hours down to minutes, making job execution faster and simpler than ever before. |

SESUG 2018 Abstracts - by Track/Section E-Posters

| SESUG Paper # | Title | Primary Author | Abstract |
|----------------|---|------------------|--|
| SESUG 105-2018 | Using SAS® to test the Psychometric Properties of Computer-Based Prostate Cancer Screening Decision Aid Acceptance Scale among African-American Men | Tavakoli, Abbas | Prostate cancer incidence and mortality rates among African-American men are 1.5 and 2.5 times greater than Whites, respectively, in the southeastern United States. Prostate cancer screening could result in early detection and potentially reduce cancer health disparities. This study tested the psychometric properties of the Computer-Based Prostate Cancer Screening Decision Aid and Acceptance Scale (PCDAA). We analyzed baseline data from a purposive sample of 352 African-American men aged 40 years and older who resided in South Carolina. Exploratory factor analysis (EFA) was conducted using maximum likelihood, squared multiple correlation, and Promax rotation. Internal consistency reliability was assessed using Cronbach's alpha. Pearson's correlation assessed the association between factors and subscale items. A Scree plot and item statistics assessed factor loadings. EFA revealed the optimal factor structure of the PCDAA among African American men was a 24-item, 3-factor model. Factor loadings ranged from 0.32 to 0.94 with 11 items loading on Factor 1, 8 items on Factor 2, and 5 items on Factor 3. A negative weak correlation between Factors 1 and 3, a strong positive correlation between Factors 1 and 2, and no correlation between Factors 2 and 3. A positive correlation between factors and the total subscale ranging from .32 to .85. Scale reliability in this study was examined by estimating the Cronbach's alpha. The reliability of each subscale ranged from 0.87 to 0.94 with a total scale reliability of 0.87. The PCDAA is a reliable measure of prostate knowledge, decision self-efficacy, and intention to make an informed decision among African American men. |
| SESUG 145-2018 | Let's Get FREQy with our Statistics: Data-Driven Approach to Determining Appropriate Test Statistic | Watson, Richann | 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. |
| SESUG 160-2018 | The Art of Accurate Reports | Garcia, Victoria | Many times, we find ourselves with an overwhelming amount of data at our fingertips. The goal of this presentation is to aid the user in creating accurate reports in SAS Enterprise Guide in a simple top-down approach using four steps; Envisioning, Planning, Creating, then Testing. For this paper, we will be working data from the Florida Fish and Wildlife Conservation Commission. |
| SESUG 163-2018 | Creating and Implementing a SAS Visual Analytics Dashboard Style Guide | Barulich, Danae | The University of Central Florida's Institutional Knowledge Management (IKM) office is the official data source for the entire university. The office provides a variety of data reports ranging from simple ad hoc projects to complex multi-page interactive SAS Visual Analytics (VA) dashboards. As the office has been tasked with creating a larger volume of interactive dashboards in last few years, it became apparent that a style guide was needed. The style guide serves two main purposes: to standardize the look and feel of IKM dashboards and to save time on dashboard development. The style guide contains detailed information in the following areas: layout, color and font selection (including suggested color choices for frequently used data points), chart styling, and general guidance on dashboard organization. The IKM VA Style Guide is currently being used for all new dashboard development. |
| SESUG 166-2018 | Tips on Developing SDTM Datasets for Complex Long-Term Safety Studies | Ling, Yunzhi | There is an increase tendency in clinical research practice to combine multiple studies' extension phases into one long-term safety (LTS) study, as the cost of conducting one LTS study is much lower than maintaining individual LTS study for each core study separately, from clinical operation point of view. However, such combination significantly increases the complexity of LTS study's Clinical Study Report (CSR) programming handling, especially in Study Data Tabulation Model (SDTM) datasets development. This paper illustrates the key areas that need to pay special attention, comparing with a typical study's SDTM datasets creation, when developing SDTM datasets in order to effectively facilitate LTS study's Analysis Data Model (ADaM) datasets creation and potential datasets pooling for integrated analyses. |

**SESUG 2018 Abstracts - by Track/Section
E-Posters**

| SESUG Paper # | Title | Primary Author | Abstract |
|----------------|---|-----------------|---|
| SESUG 198-2018 | ODS Excel Tips for an Output Makeover | Slater, Rita | The interaction between SAS® ODS Excel statement and SAS PROC REPORT is amazing. With minimal manipulation, it is relatively easy to create visually informative and easy-to-read spreadsheets. The techniques we present show how a user can perform data error detection with color coding, formatting, and shading of cells. We will also present some additional items including filtering a column, adding super or subscripts, and adding hyperlinks to a footnote. |
| SESUG 208-2018 | Consumer Emotional Loyalty – Measuring What Motivates Your Customers | Queller, Sarah | In this world of price transparency and competition, why should customers choose you? They can easily pick up their phones and find the lowest price for sirloin steak, a car, or a vacation. Yet, some customers will buy a higher priced Tiffany diamond over a Costco diamond, even if the size and quality are exactly the same. Researchers and marketers have posited that this kind of behavior is motivated by Emotional Loyalty, which can produce ‘stickier’ consumer behavior, often with less cost than competing on price. Previous academic research posits that Status, Habit and Reciprocity contribute to consumer Emotional Loyalty (e.g., Henderson, Beck & Palmatier, 2011). We make these claims actionable by creating an Emotional Loyalty Assessment. To create this scoring tool, we analyzed responses of 1200 loyalty program members on 60+ agree/disagree items intended to assess consumers’ emotional loyalty motivations. Factor analysis (SAS proc factor) validated a 3 factor orthogonal solution with each customer having one score for Status, one for Habit, and one for Reciprocity. Importantly, the Emotional Loyalty Scores were significantly related to emotional attachment to companies, as well as to intentions to purchase and to promote products or services. (Analysis was multivariate regression using SAS proc reg with Type III SS, co-varying age and gender.) We further validated the Emotional Loyalty Scores using an oversample of 200 actual hotel loyalty program members and found, as expected, that higher status scores were associated with increased Marriott/Star Preferred Guest loyalty program membership and higher Habit scores were associated with increased Choice loyalty program membership. (Analysis was logistic regression using SAS proc logistic, co-varying age and gender with Marriott/SPG as the target group; Status odds ratio = 1.88, p=.003; Habit odds ratio = 0.61, p=.015). Emotional Loyalty Scores can be used to increase customer engagement and emotional loyalty to a company by ensuring that the messaging, benefits and offers customers see are relevant to what motivates them. When emotional loyalty motivations are related to behaviors such as churn, this can suggest important changes to marketing and messaging strategies to match what motivates your customers. |
| SESUG 250-2018 | Comparison of SAS®, SAS University Edition®, and Microsoft Excel® in Collaborative Data Analysis of Physical Properties by Scientist and Statistical Programmer | Schreiber, John | Calibration data for novel instrumentation and laboratory data were analyzed using common regression techniques presented in physics literature and a pooled regression respecting the underlying physical constraints of mechanical analysis. This paper compares and contrasts implementation of these techniques in SAS®, SAS University Edition®, and Microsoft Excel®. |
| SESUG 257-2018 | Sentiment Analysis of Users’ Review to Improve 5 Star Rating Method | Arya, Surabhi | Recommendation Systems plays a significant role in e commerce industry for making personalized recommendations to users. Users with the help of different rating methods rate items and these ratings are utilized by algorithm to make recommendations to other users. Five-star rating method is a popular platform to record users rating, but it also has some disadvantages like incapability to capture important features about a product or service, or bi modal distribution of ratings i.e. users' ratings tends to cluster at extreme points which in our case is 1 and 5, whose overall interpretation can be ambiguous. This paper attempts to solve above said problems by leveraging sentiment analysis capability of SAS Enterprise Miner. The solution provided in this paper can be utilized by e commerce websites to perform sophisticated feature addition for customer interaction and optimize their credibility towards website. |

SESUG 2018 Abstracts - by Track/Section Education/Institutional Research

| SESUG Paper # | Title | Primary Author | Abstract |
|----------------|--|---------------------|---|
| SESUG 106-2018 | Enrollment Management Utilities | Mamrick, Marla | With the increased demand for data-driven decision-making in higher education, institutional researchers are required to provide data analyses that extend well beyond dashboards and descriptive statistics. Limited resources preclude IR professionals from writing research reports for all campus constituents. Enrollment Management Utilities (EMU) was designed to address these issues by allowing select users to access institutional data via a point and click environment to perform typical data manipulation and analyses used to create IR reports. A manual and online help system are available to explain how to perform each function and how to interpret the results. The system is free of charge to any institution of higher education that has a single user SAS license for SAS Base, SAS Stat, and SAS Graph. This session will present an overview of the program as well as a simple example of the use of the software. |
| SESUG 167-2018 | Opportunities and Challenges of Visual Business Intelligence Course for MBA Students | Aggarwal, Anil | As social network diffuses so does the data generated through them. This data contains text, videos, graphs, pictures and many other varieties. Data is generated at the speed of light and organizations must take advantage of it. Organizations are demanding graduates who can understand, manage and make sense of such data. Universities are obligated to provide such skills to their graduates. Visualization is becoming an important tool for exploring big data. Visualization typically has 2 parts. Exploration and Explanation. First part deals with data analysis and second part deals with data modeling and validation. These parts are not mutually exclusive. Both parts should be included in a course. This creates challenges for professors who need to either offer a new course in visualization or at a minimum include it in an existing course. It is important that our students understand this new phenomenon in addition to existing systems. There are opportunities given that software like SAS, JMP and Tableau are making visualization a reality and challenges due to business background of students. The course needs to be data-driven with visualization to be useful to business students. This paper is an attempt in developing such new visualization course. |
| SESUG 185-2018 | Developing In-House Student Success Models | Thompson, Stephanie | The number of vendors providing solutions for predicting student success keeps increasing. Each claim to be able to use your data and predict the likelihood of retention or graduation for you. This all comes at both a financial cost as well as the time and effort needed to prepare and update data to feed their systems. Is creating an in-house system a better solution? This paper walks through one way to make that determination. If your institution has gone through changes impacting the continuity of data or has made other significant structural changes, finding an off the shelf solution may be more challenging. Should you include additional variables that the commercial products do not consider? A side benefit of creating the models in-house is validating multiple models on past students as well as working with those who will be using the model output to generate buy in. Rolling out a commercial solution is of no benefit if those using it do not trust the results. SAS® has the tools to do the modeling whether you use Base SAS®, SAS Enterprise Miner®, or SAS Visual Analytics®. Why not leverage what you have to support your student's success? |
| SESUG 201-2018 | The "Why" and "How" of Teaching SAS to High School Students | Richards, Jennifer | Learning SAS for the first time as a budding epidemiologist in graduate school, I remember being amazed at how powerful and customizable SAS was in making sense of huge amounts of data. I also remember being very intimidated by all the moving parts involved in learning such a program! My study and use of SAS has continued since that time, specifically now as a Doctor of Public Health student. I have often wondered how much further along my research would be at this point, had I learned SAS in high school. When one of my professors recently told me that resources are available for teaching SAS at that level, it prompted me to ask a local high school teacher what she thought about teaching SAS to her students. When she replied that she thought it was a great idea but did not know how she could do this, I realized this was a gap that needed to be filled. The aim of this presentation is to raise the awareness of educators about the importance of, and resources available for, teaching SAS to high school students. |
| SESUG 212-2018 | Redesigning Project Management Within Institutional Research in a SAS Environment | Watts, Andre | The University of Central Florida is a large metropolitan university located in Central Florida. Due to its size and scope, Institutional Knowledge Management (IKM) within the university processes a considerable number of requests by numerous constituents for data and analysis throughout the year. Like many institutional research offices, IKM is faced with an increasing demand for complex data analysis and the creation of interactive dashboards. Part of the department's challenge is managing and delivering these large-scale projects in a timely fashion when faced while also dealing with other ad-hoc and smaller projects. As such, over the past year, the office is in the process of overhauling its project management processes to improve delivery and meet university demands. This presentation will cover how the UCF IKM office has integrated new project management processes to meet these demands and tighter deadlines within the current SAS suite of tools. We will show how these project management processes work with and alongside SAS products and have evolved due to the changing landscape of analytics at the institution. Attendees will hopefully come away with some ideas and approaches they can utilize in their organizations or institutions. |

SESUG 2018 Abstracts - by Track/Section Education/Institutional Research

| SESUG Paper # | Title | Primary Author | Abstract |
|----------------|---|-----------------|--|
| SESUG 215-2018 | Going from Zero to Report Ready with PROC TABULATE | Straney, Rachel | The TABULATE procedure in SAS® can be used to summarize your data into organized tables. This procedure can calculate many of the descriptive statistics that the MEANS, FREQ, and REPORT procedures do, but with the flexibility to display them in a customized tabulated format. At first, the syntax may seem difficult and overwhelming but with practice and some basic examples, you can go from zero to report ready in no time. This paper will discuss the benefits of using PROC TABULATE and identify the kinds of reports that this procedure is best suited. An example in a higher education setting will be used to illustrate the syntax and statements needed to generate a complex table. The table will include multiple classification variables as well as more than one numeric variable for various computed statistics. Readers will learn the functions of the CLASS, VAR and TABLE statements and how to include subtotals and totals with the keyword ALL. To make the finished table 'report ready,' examples of how to apply formats, labels, and styles will also be shared. |
| SESUG 229-2018 | From CSV to SAS®: Dealing with Empty Fields and Repeated Observations | Smith, Kelly | Sometimes, higher education students need to register for classes before official transcripts from prior education institutions have arrived at their new institution. If a student needs to enroll in a class with prerequisites and official transcripts have not been evaluated, the student can obtain an override from an advisor, using unofficial documentation, in order to complete the registration process. At Central Piedmont Community College (CPCC) in North Carolina, information about student overrides is pulled from the institutional database as a CSV file. In a recent effort to standardize and improve student placement data maintained by Institutional Research, it was decided to transform the override CSV files into SAS data sets. This paper summarizes the multi-step process involved in the data set creation. Producing the data set required the modification of student identification and date information into standard formats, the appropriate addition of student identification and date information to observations with missing values, and the creation of a condensed variable from multiple observations. Particular attention is paid to the use of IF THEN processing, the use of RETAIN and DROP in a DATA step, and the use of ARRAY to collapse a multiple-value variable. In addition, the use of internal check mechanisms within the SAS code is presented as a method for data validation. The creation of a SAS data set from institutional data is a beneficial step that maximizes the information's potential for later use. |
| SESUG 230-2018 | Student Placement: Using SAS® to Combine and Prioritize Information | Smith, Kelly | When a student applies to a North Carolina community college in order to take curriculum classes requiring college-level proficiency, it is still common for placement tests to be administered so the college-readiness of the student can be assessed. However, placement tests are not the only mechanism used to determine a student's final placement. Other considerations include state initiatives such as Multiple Measures, transfer credit from other institutions of higher education, and overrides provided by community college advisors. In a recent effort to standardize and improve student placement data, the Office of Institutional Research at Central Piedmont Community College (CPCC) revised the SAS code used to finalize student placement. This paper summarizes the refinement of the student placement program. Final student placement is constructed from four data sets: (a) the testing file, which is based on information from the student test data files; (b) the curriculum students by term file, which contains all students who participated in the registration process for a particular academic term; (c) the transfer credit file, which contains all transfer credits awarded to students for a particular academic year; and (d) the overrides file, which contains all overrides provided to students who participated in the academic term. The correct determination of student placement requires the precise selection of appropriate information from the transfer and override data sets with WHERE, the careful use of MERGE procedures to combine data sets, the creation of flag variables to track merged information, and the meticulous application of IF THEN processing. The refinement of the SAS code for determining student placement illustrates the logic analysis required to construct SAS programs that generate accurate outcomes. |
| SESUG 241-2018 | Undergraduate Student Retention. Can we catch them before they fall? | Bhandari, Lohit | Educational institutions are very concerned about the retention of students in their undergraduate programs. Until few years ago, the institutions focused only on storing and maintaining student data without utilizing it for making data informed decisions. With the availability of data mining techniques and tools, we can try to derive useful insights from such datasets. The datasets available for this study include demographic, admissions, financial, student employee and academic information. These datasets are available through the Institutional Research and Information Management division at Oklahoma State University. The study would focus on the factors that affect retention of first year freshman students. The factors or reasons for not being able to retain the students from their first semester into their second semester will be part of this study. The study would like to focus on how financial support affects the student's decision to continue their studies. The financial support includes the amount of financial aid provided to the student, and the availability of the on-campus employment opportunities helps those without the financial aid. The study will also include the demographic and other significant factors as part of the model. The technical approach planned is to cleanse, transform and consolidate the datasets in Base SAS. The final output dataset will be imported into SAS Enterprise miner. The predictive modeling techniques available in SAS Enterprise Miner will be used to assess if the above-mentioned financial indicators are significant towards student retention. This would help in improving the quality of decisions that affect the retention of students in higher education. Students, faculty, and program administrators all stand to benefit from such useful insights. |

SESUG 2018 Abstracts - by Track/Section Education/Institutional Research

| SESUG Paper # | Title | Primary Author | Abstract |
|----------------|---|-------------------------|--|
| SESUG 243-2018 | Predicting the Risk of Attrition for Undergraduate Students using SAS® Enterprise Miner | Kandula, Sivateja Reddy | A larger number of undergraduate students drop out of college prior to their graduation despite the efforts of college management to improve the retention rates. In order to improve these retention rates, it is important to identify the significant factors that contribute towards students' attrition. In this paper, I will concentrate on various aspects that play an important role in a student's decision to drop out of college. According to surveys on student attrition, it is evident that students' incoming credentials like high school grades, demographic factors like gender, race, and distance from the college, financial factors, socio-economic factors, academic performance, and campus involvement of students play a pivotal role in students' decision to attrite from college. In this paper, I will mainly concentrate on students' attrition in the third semester meaning how likely a student drop out or transfer from a college before reaching the third semester. For this project the data is acquired from a large mid-west university for Fall 2016 - Fall 2017 and the model will be validated on data from Spring 2016 - Spring 2017. Variables mentioned above will be used along with some calculated fields to predict college attrition/retention. This paper focuses on analyzing student applicant data and their campus involvement within the first two semesters using SAS Enterprise Guide 7.1, SAS Enterprise Miner 14.1, and SAS Visual Studio. This project will determine the probability of attrition of each student. Results from this study will help university officials provide services to those students who may be at risk for drop out. |
| SESUG 246-2018 | Text Mining to Predict College Admission Trends | Chebrolu, Shashikant | Any institution receives numerous applications for college admission. Out of those, a few lucky ones get an offer of admission to the college. But from the pool of students who are given out admits, there are many those who do not accept the offer. Due to this, a student who's deserving and willing to join the college losses out an opportunity to do so. This project aims at finding the link between the students who accept an admission offer and the interactions they have with college over emails to determine if there is a pattern that can be discovered to predict an acceptance with more accuracy. The data used has been provided by OSU's Institutional Research & Information Management which contains the data of all the communications between the student being offered the admit and the university for the past three years. SAS Enterprise Miner 14.2 has been used for the text analysis of this project. The ultimate goal of this project is to incorporate the text analytics results into a predictive model to detect if a student accepts the admit or not so more accurate decision making can be performed during the admission process. |
| SESUG 252-2018 | ANALYSIS OF FACTORS INFLUENCING DROPOUTS IN SOPHOMORE ENROLLMENT | Chandrasekaran, Apoorva | Bill Gates called U.S colleges' staggering dropout rates as 'tragic' in his blog post about 'Putting students first'. The United States is leading in terms of the number of people who start college but is lagging far behind in terms of the number of people who actually finish college. As a matter of fact, only about 50% of these students really leave the college with a diploma. This situation can be remedied. Most of the students that dropout, usually do so before entering the sophomore year. This analysis is based on data from Oklahoma State University, Institutional Research and Information Management (IRIM) Department. In this analysis, we leverage the power of predictive analytics and SAS® 9.4 to predict whether or not a student will enroll for their sophomore year. It will take into consideration various influential factors such as financial background, scholarships, education history, family background, university services use while in college, student employment data and athletic status among others to determine students who are at-risk. Data preparation and predictive models will be handled by Base SAS® 9.4 and SAS® Enterprise Miner™ 14.2. Preliminary findings show that variables such as enrollment in the previous two semesters and college of enrollment heavily influence whether a student will enroll in their sophomore year. This study hopes to provide an opportunity for institutions to detect and help at-risk students at an early stage, by determining the course of action they can take in order to avoid dropping out of college. |
| SESUG 260-2018 | Using SAS® to create HTML codebooks and more! | Ritchie, Ethan | With some basic knowledge of SAS® and HTML, you can create custom documents using your data and associated documentation by using a file statement and put statements to write to an HTML file. This paper provides an example of creating an organized and formatted HTML codebook based on a national education survey and the resulting data. The codebook is complete with variable labels, survey question wording, notes, value labels, and summary statistics. The process is dynamic (i.e., variable names do not need to be specified and the code does not need editing when documentation, variables, or values change). This allows the code to be ported between projects with minimal modification. |

SESUG 2018 Abstracts - by Track/Section Hands on Workshops

| SESUG Paper # | Title | Primary Author | Abstract |
|----------------|--|------------------------|--|
| SESUG 113-2018 | Hands-on SAS® Macro Programming Essentials for New Users | Lafler, Kirk Paul | The SAS® Macro Language is a powerful tool for extending the capabilities of the SAS System. This hands-on workshop teaches essential macro coding concepts, techniques, tips and tricks to help beginning users learn the basics of how the Macro language works. Using a collection of proven Macro Language coding techniques, attendees learn how to write and process macro statements and parameters; replace text strings with macro (symbolic) variables; generate SAS code using macro techniques; manipulate macro variable values with macro functions; create and use global and local macro variables; construct simple arithmetic and logical expressions; interface the macro language with the SQL procedure; store and reuse macros; troubleshoot and debug macros; and develop efficient and portable macro language code. |
| SESUG 135-2018 | Introduction to Monte Carlo Simulation | Brinkley, Jason | 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. |
| SESUG 147-2018 | Animate Your Data! | Watson, Richann | 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 hands-on workshop 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! |
| SESUG 157-2018 | Using a Handful of SAS Functions to Clean Dirty Data | Cochran, Ben | This paper shows how to use SAS Functions to clean dirty data, as well as to make data more consistent. There are a series of examples that take the audience through a step by step process to make data clean and consistent. |
| SESUG 164-2018 | Working with the SAS® ODS EXCEL Destination to Send Graphs, and Use Cascading Style Sheets When Writing to EXCEL Workbooks | Benjamin Jr, William E | This Hands-On-Workshop will explore the new SAS® ODS EXCEL destination and focus on how to write Excel Worksheets with output from SAS Graph procedures and spice it up using Cascading Style Sheet features available on modern computer systems. Note that the ODS EXCEL destination is a BASE SAS product, which makes it available on all platforms. The workshop will be limited to the Windows platform, but it should be simple to port the code to other operating systems. The code will be on the computers and you will get a chance to see how it handles. |
| SESUG 165-2018 | The ODS Excel Destination: Assorted Tips and Techniques to Simplify Writing SAS(R) Data to Excel Workbooks. | Benjamin Jr, William E | This Hands on Work shop will introduce many of the Actions, Options and Sub-Options of the New ODS EXCEL Destination, and give the students experience visualizing the results of the new ODS EXCEL Destination outputs by executing code and viewing the results. This class will cover the some of the topics in the following subject areas: <ul style="list-style-type: none"> • Differences between SAS Tagsets and SAS ODS Destinations • SAS ODS EXCEL Destination Actions • How to Set Excel Property Values • Modifying Output Features Using Anchor, Style, and CSS • Options that Apply to Titles, Footnotes, and Sheet processing • Ways you can adjust Print Features like Headers and Margins • Options that Impact Columns, Rows, and Cells |
| SESUG 204-2018 | Getting Started with the SGPLOT Procedure | Horstman, Josh | Do you want to create highly-customizable, publication-ready graphics in just minutes using SAS? This workshop will introduce the SGPLOT procedure, which is part of the ODS Statistical Graphics package included in Base SAS. Starting with the basic building blocks, you'll be constructing basic plots and charts in no time. We'll work through several different plot types and learn some simple ways to customize each one. |
| SESUG 205-2018 | Doing More with the SGPLOT Procedure | Horstman, Josh | 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. |

SESUG 2018 Abstracts - by Track/Section Hands on Workshops

| SESUG Paper # | Title | Primary Author | Abstract |
|----------------|---|-------------------|---|
| SESUG 258-2018 | From Raw, Messy Data to a Clean Analytic Dataset: Common Data Management Techniques using SAS ® | Smith, Raymond B. | Despite the amount of quantitative research that exists in the social and behavioral sciences, many graduate programs do not offer classes focused on the multitude of steps necessary to manage quantitative datasets. Instead, this skill is often learned through trial and error with the beginning SAS user having to use multiple resources, including, but not limited to the plethora of proceedings papers from SAS Global Forum as well as regional users' groups as well as other SAS friendly resources such as UCLA's Institute for Research and Digital Education and SAS publications. Although these resources are incredibly useful when a SAS user knows what procedure he/she needs, they are less useful for the novice analyst who does not know where to begin. The focus of this paper is to help guide the novice user through common data management techniques to transform raw, messy data into an analytic dataset. This paper contains information on data management processes as basic as getting an external data set into SAS to more advanced techniques such as using a macro to examine missing data mechanisms as well as using PROC SURVEYSELECT to split the data into an exploratory sample and a holdout sample. We illustrate the various processes using Wave 1 public use data from the National Longitudinal Study of Adolescent to Adult Health (Add Health). |
| SESUG 262-2018 | Working in SGPLOT: Understanding the General Logic of Attributes | Blum, Jim | ODS Graphics use common style elements for distinct graphical entities—text, symbols, fills and lines—through attribute options. Most graphical entities produced, either by default or requested via a specific option, can have their styles modified with an ATTRS-type option. The syntax of such options is built to be consistent across instances of these graphical entities for most plotting statements that generate them. Knowing what graphical entities are generated by various statements and options, along with understanding how SAS categorizes graphical entities and their style elements, can help make style modification much easier to understand. Examples will focus on the commonalities present across a wide variety of plotting statements and options within them. Some non-standard style elements will be considered. |

SESUG 2018 Abstracts - by Track/Section Life Sciences/Healthcare/Insurance

| SESUG Paper # | Title | Primary Author | Abstract |
|----------------|---|---------------------|--|
| SESUG 102-2018 | Patient-Level Longitudinal Analysis Plots Showing Adverse Event Timelines and Dose Titration Levels | Gerlach, John R | The analyses in a clinical trial often include subject-specific longitudinal plots that depict, for example, a lab measurement taken at various times during a study period. Because the study involves dose titration levels, the plots might include vertical lines indicating dose levels, which can increase and decrease during the study period. Perhaps the analysis includes another component: the timelines of adverse events, if any. Both the vertical reference lines and horizontal time lines are superimposed on the longitudinal plot. Obviously, the task becomes more intricate since each plot is subject-specific, having different measurements, dose titration levels, and adverse events. This paper explains how to produce this intricate graphical analysis. |
| SESUG 123-2018 | Hardcoding In Clinical Trials: A (Sometimes) Necessary Evil | Weller, Gregory | Hardcoding is generally considered to be a dirty word in the clinical trials programming world, especially in statistical programming. However, sometimes it is the only way to accurately report data that is known to be incorrect. This paper will explore two examples from clinical trials where it was determined hardcoding post database lock was necessary to fix incorrect raw data. Every step of the hardcoding process will be explained, including identification of the problem, figuring out the appropriate solution, and the documentation involved. This paper will also explore an example of a programmer unintentionally hardcoding while trying to handle problematic raw data values. While unintentional hardcoding is not done with malice, it is nonetheless a serious problem because: 1) it bypasses the steps required to properly validate and approve hardcoding, and 2) the code resulting from an unintentional hardcoding is unlikely to be useful beyond the specific version of raw data it was created for. |
| SESUG 199-2018 | SDTM - Just a Walk in the (Theme) Park: Exploring SDTM in the Most Magical Place on Earth | McNichol, Christine | Most in the pharmaceutical industry know the basic concepts of SDTM. Let us now infuse some magic and take a virtual vacation to look at SDTM from a perspective inspired by an exciting vacation. Some of the basic as well as more complicated aspects of SDTM can be shown in examples of things encountered during a day at a theme park in Central Florida. With a little imagination, our virtual vacation can be equated to a clinical trial. Come along as the vacationers (subjects) take part in a vacation (study) where they will experience various rides and attractions (treatments). The manner in which these experiences could be mapped to SDTM will be examined, from describing the plan of the virtual vacation in the TDM, through capturing the events of the day in SDTM domains. Within this trip, some more common but important domain examples such as DM, AE, DS, VS and more will be discussed. Additionally, some more advanced concepts in SDTM in areas within the vacation example, such as the usage of EX vs EC, basic trial design strategies to use in TDM creation, EDC and external data to support core study data, as well as a few applicable device domains will be detailed. These concepts will be examined in a new and creative way as SDTM with a magical twist is explored. |
| SESUG 220-2018 | Dynamic Program Tracker | Khole, Tanmay | A program tracking document for statistical programming in clinical trial is an essential document which gives us the progress report of the project. Maintaining this document is one of the prime responsibilities of a statistical programmer. By making the tracking document dynamic, we can check the development and validation status of programs which will in turn save time in assessing the status of the study with respect to statistical programming deliverables. A document which can dynamically track the process and alert the user of any process deviations can prove very helpful in statistical programming for clinical trial projects. Dynamic Program Tracker also greatly reduces the time and effort required to check if all programs are following the Program Development Life Cycle, as well as check for any issues with programs. This paper will describe the design, features and comprehensive checks done by Dynamic Program Tracker. The code for Dynamic Program Tracker is developed in Microsoft VBA (Visual Basic for Applications) which is available within all Microsoft Office products - with no additional costs for new software. The purpose of this tracker is to check compliance and serve as a project management tool. |

**SESUG 2018 Abstracts - by Track/Section
Life Sciences/Healthcare/Insurance**

| SESUG Paper # | Title | Primary Author | Abstract |
|----------------|--|--------------------|---|
| SESUG 245-2018 | Predictive Modeling for Healthcare Professionals: The use of time-series analysis for health-related data and the application of ARIMA modeling techniques in SAS for Public Health Practice | Comeaux, Camillia | The use of time-series analysis in public health practice is an under-utilized tool that can aid in effective activities such as: program and health planning; appropriate health service and provider delivery; improved emergency preparedness action; and much more. The ARIMA technique is a type of trend recognition tool in time-series analysis, that can sort through large amounts of data and create a statistical model for forecasting. Time-series models, such as ARIMA, were historically used in financial industries to assess risks and market changes overtime to predict future economic outcomes. The ARIMA technique is a process in which the stages of model identification, parameter estimation, and diagnostic checking are repeated to find the most appropriate fitting model for prediction (Chen, 2008). This modeling technique is operational when data is assumed to have stationarity, or without a trend, and uses longitudinal data with at least forty-five data points to increase the accuracy of forecasting (Chen, 2008). Engle (2001) suggests that the utility of modeling techniques in time-series analysis, is their ability to factor in major shocks and volatility shifts over time; therefore, in the public health field, the social, the economic, and the ecological factors correlated to these shocks over time can be analyzed for health outcome forecasting and future population health planning. This product gives a generalized overview of time-series analysis and its application to public health practice. Using the SAS statistical package, health professionals will be empowered to use time-series analysis, specifically ARIMA modeling techniques, to analyze and interpret large health data sets to predict or forecast factors that impact health outcomes on populations. |
| SESUG 269-2018 | Utilization of SAS® Input Statements for the Barell Injury Diagnosis Matrix to Categorize Senior Sports Injury by Body Region and Nature of Injury | McGlonn, Kassandra | The Barell Injury Diagnosis Matrix is used by a wide variety of health practitioners and researchers worldwide to identify, analyze, and communicate injury information. In this paper, we will present a thorough investigation of Florida statewide emergency department (ED) data for senior sports injuries using SAS® input statements that categorized injuries by body region and nature of injury per the Barell Injury Diagnosis Matrix for ICD-9-CM codes. This paper will detail practical use, benefits, limitations, and ease of use for Barell Matrix SAS® input statements. |
| SESUG 287-2018 | Analyzing Hospital Medicare Cost Report Data Using SAS | Andrews, Kimberly | 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 files; describing the characteristics of the data, and providing basic SAS program code which can be used to analyze the data. |
| SESUG 290-2018 | SAS/STAT® 14.3 Round-Up: Modern Methods for the Modern Statistician | Stokes, Maura | The latest release of SAS/STAT® software has something for everyone. The new CAUSALMED procedure performs causal mediation analysis for observational data, enabling you to obtain unbiased estimates of the direct causal effect. You can now fit the compartment models of pharmacokinetic analysis with the NLMIXED and MCMC procedures. In addition, variance estimation by the bootstrap method is available in the survey data analysis procedures, and the PHREG procedure provides cause-specific proportional hazards analysis for competing-risk data. Several other procedures have been enhanced as well. Learn about the latest methods available in SAS/STAT software that can modernize your statistical practice. |

SESUG 2018 Abstracts - by Track/Section Planning/Administration

| SESUG Paper # | Title | Primary Author | Abstract |
|----------------|--|---------------------|---|
| SESUG 101-2018 | What's the Difference? – Comparing SAS® Artifacts After a Migration or Upgrade | Varney, Brian | There are times when it is convenient to be able to compare two file system areas that are supposed to be equivalent. This could be during a migration activity involving changing servers and/or SAS® versions. This paper intends to cover techniques that can help facilitate the comparison of programs, logs, excel files, and SAS data sets as well as techniques for digesting the results. |
| SESUG 162-2018 | Learnin' about Churnin' | Garcia, Victoria | What is churn? According to sources it's; "The percentage of customers who cut ties with your company". Churn rate tell a story. If you listen, you'll learn a lot more about your customers and company than you think. In this presentation, we will be taking a look at our agencies churn rate; how we got it, what we learned from it, and what we are doing to improve it. For this presentation, we will be working data from the Florida Fish and Wildlife Conservation Commission. |
| SESUG 169-2018 | Get Smart! Eliminate Kaos and Stay in Control – Creating a Complex Directory Structure with the DLCREATEDIR Statement, SAS® Macro Language, and Control Tables | Hadden, Louise | An organized directory structure is an essential cornerstone of data analytic development. Those programmers who are involved in repetitive processing of any sort control their software and data quality with directory structures that can be easily replicated for different time periods, different drug trials, etc. Practitioners (including the author) often use folder and subfolder templates or shells to create identical complex folder structures for new date spans of data or projects, or use manual processing or external code submitted from within a SAS® process to run a series of MKDIR and CHDIR commands from a command prompt to create logical folders. Desired changes have to be made manually, offering opportunities for human error. Since the advent of the DLCREATEDIR system option in SAS version 9.3, practitioners can create single folders if they do not exist from within a SAS process. Troy Hughesi describes a process using SAS macro language, the DLCREATEDIR option, and control tables to facilitate and document the logical folder creation process. This paper describes a technique wrapping another layer of macro processing which isolates and expands the recursive logical folder assignment process to create a complex, hierarchical folder structure used by the author for a project requiring monthly data intake, processing, quality control and delivery of thousands of files. Analysis of the prior month's folder structure to inform development of control tables and build executable code is discussed. |
| SESUG 200-2018 | Are You Ready for It? Preparing for Your Next Technical Interview | McNichol, Christine | Whether you are interviewing for a position as an intern, programmer, statistician, standards expert, or technical manager, you are likely to come across... the technical interview. Even if you have experience and know how to do your job inside and out, do you really need to spend time preparing for an interview? Yes! Though you may feel it is just going to be questions about what you do every day, there are some points about preparation and presentation that should be considered. Whether you are about to graduate college and go on your first interview, or you have not interviewed since before smart phones, or even if you interview every few years, it does not hurt to be a little bit more prepared for an interview that may determine how and where you spend your days Monday through Friday. We will cover topics and considerations from preparing for an interview through participating in and following up after a technical interview. Though each interviewer will have different styles, preferences and opinions, we will cover topics that will get you in the interview mindset and get you thinking about ways to present yourself and your skillset to an interviewer. |
| SESUG 203-2018 | So You Want To Be An Independent Consultant | Horstman, Josh | While many statisticians and programmers are content with a traditional employment setting, others yearn for the freedom and flexibility that come with being an independent consultant. While this can be a tremendous benefit, there are many details to consider. This paper will provide an overview of consulting as a statistician or programmer. We'll discuss the advantages and disadvantages of consulting, getting started, finding work, operating your business, and various legal, financial, and logistical issues. |

SESUG 2018 Abstracts - by Track/Section Planning/Administration

| SESUG Paper # | Title | Primary Author | Abstract |
|----------------|--|----------------|---|
| SESUG 207-2018 | Planning for Migration from SAS® 9.4 to SAS® Viya® | Hayes, Spencer | SAS® Viya® is the new Analytics architecture based on the Cloud Analytic Services (CAS) In-Memory Engine delivered by the SAS Institute Inc. SAS Viya brings a change in the fundamental methodology of installing SAS by moving away from SAS Software Depots and towards industry-standard software deployment methods. This paper will compare and contrast SAS 9.4 with SAS Viya in several key areas for Administrators including Pre-Install Requirements, Installation processes, Administration tools and methods, and data source connectivity including library definitions. The paper will also discuss Upgrade and Migration Planning. |
| SESUG 226-2018 | Accredited, Bona Fide, Certified, Diploma'ed, and Edumacated: The ABCDEs of Automating the Validation and Monitoring of Professional Requirements for Employees and Job Candidates Through Dynamic, Data | Hughes, Troy | Job postings typically have stated requirements such as education, training, certifications, and other criteria. Some requirements must be met before a candidate is hired while others can be fulfilled after employment during a “grace period” of a specified duration. Because many certifications require periodic refresh or renewal, validation of professional requirements often continues after candidates are hired and throughout their careers. While validation and monitoring professional requirements for a small team might be accomplished in minutes, the effort increases with the number of employees, the diversity of employee roles, and the diversity of requirements that must be tracked. This text demonstrates a flexible, scalable, reusable macro (COMPLIANCE) that validates and monitors education, training, and professional certifications for job candidates or employees. The data-driven SAS® solution relies on an external data model (operationalized in an Excel spreadsheet) that specifies the requirements, desired achievements, and applicable grace periods. This software modularity—the separation of the data model from the underlying SAS software—ensures that the solution can be adapted to any industry, environment, or organization by modifying only the spreadsheet. COMPLIANCE creates dynamic, color-coded HTML reports that delineate personnel who meet requirements, who are within a grace period, who do not currently meet requirements but have met requirements in the past, and who have never met requirements, enabling immediate visual identification of top performers or non-compliant employees. |
| SESUG 277-2018 | How to HOW: Hands-on-Workshops Made Easy | Kincaid, Chuck | 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 see if they people can 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. |
| SESUG 292-2018 | Comparing SAS® Viya® and SAS® 9.4 Capabilities: A Tale of Two SAS Platform Engines | Peters, Amy | SAS(r) Viya(r) extends the SAS(r) Platform in a number of ways and has opened the door for new SAS(r) software to take advantage of its capabilities. SAS(r) 9.4 continues to be a foundational component of the SAS Platform, not only providing the backbone for a product suite that has matured over the last forty years, but also delivering direct interoperability with the next generation analytics engine of SAS Viya. Learn about the core capabilities shared between SAS Viya and SAS 9.4, and about where they are unique. See how the capabilities complement each other in a common environment, and understand when it makes sense to choose between the two and when it makes sense to go with both. In addition to these core capabilities, see how the various SAS software product lines stack up in both, including analytics, visualization, and data management. Some products, like SAS(r) Visual Analytics, have one version aligned with SAS Viya and a different version with SAS 9.4. Other products, like SAS(r) Econometrics, leverage the in-memory, distributed processing of SAS Viya, while at the same time including SAS 9.4 functionality like Base SAS(r) and SAS/ETS(r) software. Still other products target one engine or the other. Learn which products are available on each, and see functional comparisons between the two. In general, gain a better understanding of the similarities and differences between these two engines behind the SAS Platform, and the ways in which products leverage them. |

SESUG 2018 Abstracts - by Track/Section Planning/Administration

| SESUG Paper # | Title | Primary Author | Abstract |
|----------------|---|-------------------|---|
| SESUG 293-2018 | Achieving Your Cloud-First Mandate with SAS® | Peters, Amy | Public cloud offers organizations of all shapes and sizes the opportunity for cost reduction and flexibility, ubiquitous access, and ease of maintenance. This session highlights the SAS strategy for cloud-first initiatives. We also cover common patterns and best practices for organizations seeking to pursue their own cloud-first mandates, including transitioning existing workloads to public cloud infrastructure and highlighting the possibilities for building entirely new, cloud-based workloads. |
| SESUG 299-2018 | Panel Discussion: Building Your Data Professional Career | Lafler, Kirk Paul | This session features a moderator and panelists that are experienced industry leaders. We will discuss how to build your data career and answer questions about what moves careers forward and what causes them to stall out, should you try an independent path, and what to focus on at different parts of your career. |
| SESUG 300-2018 | Panel Discussion: Building An Organization's Analytics Maturity Roadmap | Kincaid, Chuck | This session features a moderator and panelists that are leaders in companies implementing and executing on an analytics maturity roadmap. We will discuss how organizations build their analytics maturity roadmaps with respect to technology, process, culture, governance, and hiring & training |

SESUG 2018 Abstracts - by Track/Section Reporting/Visualization/JMP

| SESUG Paper # | Title | Primary Author | Abstract |
|----------------|--|-------------------|---|
| SESUG 103-2018 | Interrelationship Digraphs and Sociograms Using SAS® | Alexander, Melvin | This presentation is about two important visualizations: one used in quality management and planning (MP), the other for studying patterns of interactions between people in groups. The two tools are the Interrelationship Digraphs and Sociograms. The Interrelationship Digraph (ID) or Relations Diagram is designed to clarify intertwined, causal relationships among a group of items, issues, problems, or opportunities. This tool helps analysts gain insights into potential complex relationships of root causes that may underlie recurring problems despite efforts to resolve them. The Sociogram is a graphic representation of the connections or links that a person has with other persons in a social network. Sociograms are useful in analyzing interrelationships between people such as: determining which individuals direct most of their comments to others or the group, who interrupts others, which person is the center of attention, how members of the group are connected, and other observable, interpersonal relationships. IDs and Sociograms are specialized applications of Graph Theory, and Social Network Analysis where connections and interactions between objects and systems are made. My plan is to give some background about IDs; describe how to construct sociograms and the two standard forms of IDs using base SAS®, SAS/IML®, the %DS2CONST macro; and conclude with final notes and thoughts. |
| SESUG 104-2018 | Does Anybody Really Know What Time It Is? Mapping Time Zones with SAS | Okerson, Barbara | In today's workplace, project teams are often virtual and spread across the country (or even the world). It is never a pleasant surprise to receive a meeting invitation two hours outside one's working hours. When planning for meetings and events, it is important for the organizers to know where and in what time zone the attendees are located. Along with many other mapping tools, SAS provides time zone information by zip code that can be used to map time zones and identify time for employee locations. This presentation will provide several examples of mapping time zones in SAS. |
| SESUG 156-2018 | Backsplash patterns for your world: A look at SAS OpenStreetMap (OSM) tile servers | Okerson, Barbara | 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. |
| SESUG 172-2018 | Wow! You Did That Map With SAS®?! Round II | Hadden, Louise | This paper explores the creation of complex maps with SAS® software, incorporating the wide range of possibilities provided by SAS/GRAPH and polygon plots in the SG procedures, as well as replays, overlays in both SAS/GRAPH and SG procedures, and annotations including Zip Code level processing. The more recent GfK maps now provided by SAS that underlie newer SAS products such as Visual Analytics as well as traditional SAS products will be discussed. The pre-production SGMAP procedure released with Version 9.4 Maintenance release 5 will be introduced and some exciting examples displayed. |
| SESUG 180-2018 | Using PROC TABULATE and ODS Style Options to Make Really Great Tables | Wright, Wendi | 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. |

SESUG 2018 Abstracts - by Track/Section Reporting/Visualization/JMP

| SESUG Paper # | Title | Primary Author | Abstract |
|----------------|---|------------------|---|
| SESUG 191-2018 | A Macro for Creating Time Trend Plots with Four Y Axes | Beal, Dennis | Time trend plots are used often in statistical analysis to see how variables change over time. SAS® can easily produce time trend plots with one or two Y axes, but displaying more than two Y axes can be challenging. If there are three or more Y variables with very different scales that are to be compared together across time on the same graph, placing these Y variables simultaneously on a two-dimensional graph can be done. This paper shows SAS code in a macro that uses annotation with PROC GPLOT to produce time trend graphs with four Y axes. This paper is for intermediate SAS users of Base SAS® and SAS/GRAPH. |
| SESUG 214-2018 | Five Crazy Good Visualizations and How to Plot Them | Mintz, David | We all have our favorite visualizations. The best ones deliver a clear message to the intended audience. Over the years, there are a few that have won my affection. I would like to share my top five with you, along with the code and a few anecdotes about why they make the list. Some of these examples are static; others are interactive. This paper will cover SAS/GRAPH® and ODS Graphics procedures. It will also touch on a few basic elements of good graphical design. |
| SESUG 225-2018 | The Doctor Ordered a Prescription...Not a Description: Driving Dynamic Data Governance Through Prescriptive Data Dictionaries That Automate Quality Control and Exception Reporting | Hughes, Troy | Data quality is a critical component of data governance and describes the accuracy, validity, completeness, and consistency of data. Data accuracy can be difficult to assess, as it requires a comparison of data to the real-world constructs being abstracted. But other characteristics of data quality can be readily assessed when provided a clear expectation of data elements, records, fields, tables, and their respective relationships. Data dictionaries represent a common method to enumerate these expectations and help answer the question What should my data look like? Too often, however, data dictionaries are conceptualized as static artifacts that only describe data. This text introduces dynamic data dictionaries that instead prescribe business rules against which SAS® data sets are automatically assessed, and from which dynamic, data-driven, color-coded exception reports are automatically generated. Dynamic data dictionaries—operationalized within Excel workbooks—allow data stewards to set and modify data standards without having to alter the underlying software that interprets and applies business rules. Moreover, this modularity—the extraction of the data model and business rules from the underlying code—flexibly facilitates reuse of this SAS macro-based solution to support endless data quality objectives. |
| SESUG 271-2018 | CONSORT Diagrams with SG Procedures: Adding Efficiencies | Rosanbalm, Shane | The PharmaSUG 2018 paper "CONSORT Diagrams with SG Procedures" was a major step forward in simplifying the process of creating CONSORT diagrams in SAS. The SGPLOT-based approach superseded the previous RTF approach, which was adequate but tedious and quirky. Unfortunately, the SGPLOT approach requires the user to specify a lot of metadata. In order to create a diagram with 15 rectangles the user has to specify 404 numeric values over 223 lines of code. In this paper we explore options for reducing this specification burden. The proposed modifications result in the specification of 142 numeric values over 123 line of code. |
| SESUG 279-2018 | Active Students Filing an Intent to Graduate at the University of Central Florida (UCF) | Piemonti, Carlos | SAS® software is about improving ourselves as SAS developers, and as a consequence, is about always improving user experience. At Institutional Knowledge Management (IKM), official source of UCF institutional data, we strive in providing accurate, insightful, and actionable information, timely, to the university and the public. This paper will present you with a project converted from SAS® Enterprise Guide® to SAS® Stored Processes allowing users to select specific criteria via cascading prompts. These SAS Stored Processes are available through our SAS® Information Delivery Portal, providing a daily snapshot of the data on demand, instead of manually generating these reports involving many steps in the process. Challenges regarding the application of the new source of data being used are briefly addressed. |
| SESUG 286-2018 | Graph Visualization for PROC OPTGRAPH | Henshaw, Andrew | PROC OPTGRAPH is an extensive set of tools for graph and network analysis, but the lack of visualization capability is limiting. We've developed a simple, yet powerful, online visualization that can be directly accessed using the SAS macro language. Our tool creates and renders various graph layouts using a tightly-linked external website and the Graphviz programming toolkit. The external web service is programmed in Python and is easily integrated into an existing capability. A macro wrapper around PROC HTTP is used to push the output of the PROC OPTGRAPH procedure to the visualization web service. Node and edge-specific attributes (such as edge weight or node color) may be directly manipulated using added table columns. Whenever new data is pushed, the graph layout and the visualization are both updated automatically. |

SESUG 2018 Abstracts - by Track/Section Reporting/Visualization/JMP

| SESUG Paper # | Title | Primary Author | Abstract |
|----------------|--|------------------|--|
| SESUG 294-2018 | Tips and Techniques for Designing the Perfect Layout with SAS® Visual Analytics | Norris, Ryan | Do you want to create better reports but find it challenging to design the best layout to get your idea across to your consumers? Building the perfect layout does not have to be a rocky experience. SAS(r) Visual Analytics provides a rich set of containers, layout types, size enhancements, and options that enable you to quickly and easily build beautiful reports. Furthermore, you can design reports that work across different device sizes or that are specific to a particular device size. This presentation explores how to use the layout system, describes the feedback you will encounter, and finally demonstrates building several sample reports. |
| SESUG 295-2018 | Insights from a SAS Technical Support Guy: A Deep Dive into the SAS® ODS Excel Destination | Parker, Chevell | SAS is a world leader in the area of data analytics, while Microsoft Excel, with over 30 million active users, is a leader when it comes to spreadsheet packages. Excel spreadsheets are heavily used for calculations, information organization, statistical analysis, and graphics. SAS(r) can leverage the power of its world-class analytics and reporting capabilities to produce stylistic and highly functional Excel spreadsheets by using the Output Delivery System (ODS) Excel destination. This paper, relevant to anyone who uses Microsoft Excel, offers insights into the depths of the ODS Excel destination by illustrating how you can customize styles in Microsoft Excel worksheets and discusses common layout and reporting questions (including limitations). In addition, the discussion covers useful applications for automating and executing Excel worksheets. After diving deep into this discussion about the ODS Excel destination, you should understand the behavior and capabilities of the destination so that you can create aesthetic and effective Excel worksheets. |
| SESUG 296-2018 | Transformation of Patient Reported Outcome Survey Responses Made Simple with JMP | Shapiro, Mira | JMP has a robust set of functions that facilitate data cleaning and transformation. This discussion will focus on use of the MATCH function to transform survey responses for use in subgroup analysis. Using data from the MyLymeData patient registry, JMP will be used to create a Likert Scale from patient survey responses and create visual representations of patient subgroups. |
| SESUG 302-2018 | Panel Discussion: Plotting in SAS - How to Pick the Best Method | Okerson, Barbara | Panel discussion on the ways to plot in SAS - GPLOT, SGPLOT, DSGI, JMP etc with panelists that regularly use one of the methods - sort of a back and forth debate with audience participation. |
| SESUG 303-2018 | Great Time to Learn GTL: a Step by Step Approach at Creating the Impossible | Watson, Richann | 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 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. |
| SESUG 304-2018 | A Map is Just a Graph Without Axes | Wooding, Nat | SAS' ® PROC GMAP can produce a variety of maps of varying complexity but to go beyond the basic capabilities of GMAP, it is necessary to use the ANNOTATE Facility in order to add additional information such as symbols in specific places. Furthermore, there are times that the desired map is simply a sketch of geographically related measurements that need to be displayed in a simulated, not to scale, map. A map is simply a collection of coordinates that are plotted but for which no X/Y axis system is typically shown (although items such as road atlases or military maps or charts may have a grid and axes to help locate specific points of reference). By remembering this, one can sometimes create an embellished map using PROC GPLOT without having to create an ANNOTATE data set. Furthermore, by using GPLOT with the axes, one can locate invalid map coordinates in user created map files. Finally, an example of creating a plotted outline map with dots showing environmental variables using ANNOTATE and PROC GPLOT is offered. ANNOTATE is used in the latter case since it was necessary to dynamically scale the dots that represent the location and magnitude of the plotted values. |

SESUG 2018 Abstracts - by Track/Section Statistics/Data Analysis

| SESUG Paper # | Title | Primary Author | Abstract |
|----------------|---|-----------------|---|
| SESUG 119-2018 | Using the R interface in SAS ® to Call R Functions and Transfer Data | Gilsen, Bruce | Starting in SAS ® 9.3, the R interface allows 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 both SAS/IML users and all 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 wish to take advantage of the R interface. |
| SESUG 133-2018 | Forecasting: Something Old, Something New | Dickey, Dave | Abstract: ARIMA (AutoRegressive Integrated Moving Average) models for data taken over time were popularized in the 1970s by Box and Jenkins in their famous book. SASTM software procedures PROC ESM (Exponential Smoothing Models) and PROC UCM (Unobserved Components Models which are a simple subset of statespace models – see PROC SSM) have become available much more recently than PROC ARIMA. Not surprisingly, since ARIMA models are universal approximators for most reasonable time series, the models fit by these newer procedures are very closely related to ARIMA models. In this talk, some of these relationships are shown and several examples of the techniques are given. At the end, the listener will find that there is something quite familiar about these seemingly new innovations in forecasting and will have more insights into how these methods work in practice. The talk is meant to introduce the topics to anyone with some basic knowledge of ARIMA models and the examples should be of interest to anyone planning to analyze data taken over time. |
| SESUG 134-2018 | A Case Study of Mining Social Media Data for Disaster Relief: Hurricane Irma | Gadidov, Bogdan | In the wake of two recent hurricanes, Harvey and Irma, local, state, and federal governments are trying to provide relief to the millions of affected people. With projected property damage in the hundreds of billions of dollars, these recent natural disasters will have long-lasting effects on their respective areas where recovery could take years. This paper aims to use social media data, specifically Twitter, to analyze how people in the affected areas reacted to these natural disasters in the days leading up to the storm, during the storm, and after the storm. The goal is to see if there are any trends detected in the responses of affected citizens which can be used to help relief efforts in future natural disasters. For the most recent hurricane, Irma, we collected tweets in South Florida and analyze the discussed topics among civilians. Data was collected from Thursday (9/7/2017) to Wednesday (9/13/2017) (with the hurricane making landfall on Sunday morning). We use SAS® Enterprise Miner™ for the analysis of the tweets. Techniques such as stemming and lemmatization of words are used in the pre-processing of the text data. Topic modeling, text clustering, and time series are combined to better understand peoples' reactions throughout a storm event. This analysis is performed at the hourly level. |
| SESUG 136-2018 | Regression Models for Count Data | Brinkley, Jason | 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 talk 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 this presentation 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. |
| SESUG 149-2018 | Advanced Project Management beyond Microsoft Project, Using PROC CPM, PROC GANTT, and Advanced Graphics | Sloan, Stephen | The Challenge: Instead of managing a single project, we had to craft a solution that would manage hundreds of higher- and lower-priority projects, taking place in different locations and different parts of a large organization, all competing for common pools of resources. Our Solution: Develop a Project Optimizer tool using the CPM procedure to schedule the projects, and using the GANTT procedure to display the resulting schedule. The Project Optimizer harnesses the power of the delay analysis feature of PROC CPM and its coordination with PROC GANTT to resolve resource conflicts, improve throughput, clearly illustrate results and improvements, and more efficiently take advantage of available people and equipment. |

SESUG 2018 Abstracts - by Track/Section Statistics/Data Analysis

| SESUG Paper # | Title | Primary Author | Abstract |
|----------------|---|---------------------------|---|
| SESUG 228-2018 | Fight Human Trafficking with Text Analytics | Mehrotra, Stuti | Human trafficking is an international challenge and it affects every country. The United Nations defines it as the recruitment, transportation, transfer, harboring, or receipt of persons by improper means (such as force, abduction, fraud, or coercion) for an improper purpose including forced labor or sexual exploitation. It is a grave violation of human rights. The problem has escalated as human traffickers have now tapped into the power of online advertising. Online advertising is cost effective and has an exhaustive reach which allows human traffickers to operate on a global level and reach their targeted customers. Hundreds of online ads are posted on websites such as backpage.com where sex traffickers advertise their services. In an attempt to tackle the problem this paper uses the power of text analytics to build a robust model for identification and categorization of potential human and sex trafficking ads. The ads were scraped from Backpage.com. They were categorized by human volunteers and Text Mining nodes provided by SAS® Enterprise Miner™ 14.2. The categorization by volunteers was then used to validate the output from Text Mining nodes provided by SAS® Enterprise Miner™ 14.2. While Backpage.com was closed down by the federal government, identification of online ads will help in identifying other malicious websites that promote such ads. Making it riskier for traffickers to reach their customers using online advertising thus affecting their demands. |
| SESUG 234-2018 | Data Driven Approach in the NBA Pace and Space Era | Ferrara, Tom | Whether you're an NBA executive or Fantasy Basketball owner or a casual fan, you can't help but begin the conversation of who is a top tier player? Currently who are the best players in the NBA? How do you compare a nuts and glue defensive player to a high volume scorer? The answer to all these questions lies within segmenting basketball performance data. A k-means cluster is a common used guided machine learning approach to grouping data. I will apply this method to human performance. This case study will focus on NBA basketball individual performance data. The goal at the end of this case study will be to apply a k-means cluster to identify similar players to use in team construction. SAS Procedures utilized in this approach are proc standard, proc fastclus, proc gplot and proc anova. |
| SESUG 247-2018 | Logistic and Linear Regression Assumptions: Violation Recognition and Control | Schreiber-Gregory, Deanna | Regression analyses are one of the first steps (aside from data cleaning, preparation, and descriptive analyses) in any analytic plan, regardless of plan complexity. Therefore, it is worth acknowledging that the choice and implementation of the wrong type of regression model, or the violation of its assumptions, can have detrimental effects to the results and future directions of any analysis. Considering this, it is important to understand the assumptions of these models and be aware of the processes that can be utilized to test whether these assumptions are being violated. Given that logistic and linear regression techniques are two of the most popular types of regression models utilized today, these are the ones that will be covered in this paper. Some Logistic regression assumptions that will be reviewed include: dependent variable structure, observation independence, absence of multicollinearity, linearity of independent variables and log odds, and large sample size. For Linear regression, the assumptions that will be reviewed include: linearity, multivariate normality, absence of multicollinearity and auto-correlation, homoscedasticity, and measurement level. This paper is intended for any level of SAS® user. This paper is also written to an audience with a background in theoretical and applied statistics, though the information within will be presented in such a way that any level of statistics/mathematical knowledge will be able to understand the content. |
| SESUG 248-2018 | Regularization Techniques for Multicollinearity: Lasso, Ridge, and Elastic Nets | Schreiber-Gregory, Deanna | Multicollinearity can be briefly described as the phenomenon in which two or more identified predictor variables are linearly related, or codependent. The presence of this phenomenon can have a negative impact on an analysis as a whole and can severely limit the conclusions of a research study. In this paper, we will briefly review how to detect multicollinearity, and once it is detected, which regularization techniques would be the most appropriate to combat it. The nuances and assumptions of R1 (Lasso), R2 (Ridge Regression), and Elastic Nets will be covered in order to provide adequate background for appropriate analytic implementation. This paper is intended for any level of SAS® user. This paper is also written to an audience with a background in theoretical and applied statistics, though the information within will be presented in such a way that any level of statistics/mathematical knowledge will be able to understand the content. |

SESUG 2018 Abstracts - by Track/Section Statistics/Data Analysis

| SESUG Paper # | Title | Primary Author | Abstract |
|----------------|--|------------------------------|---|
| SESUG 249-2018 | Worker safety in energy production in America: A comparative analysis of fuel sources and accompanying occupational risk | Venn, Alyssa | <p>The world runs on energy. From light bulbs to cell phones to cars, energy is essential. In our increasingly technology-based society, its importance too is increasing. As our reliance on energy increases, the need to find safe and sustainable energy production sources only becomes more urgent. In the United States as of 2016, coal energy makes up about 30% of electricity generated, while hydroelectric provides about 7%. In 2007, the rate of fatal injuries for coal miners was almost six times the rate of fatal injuries in private industry, and it seems that every time you check the news, there are stories detailing occupational hazards in the US energy industry. In 2006, an explosion trapped thirteen miners underground in West Virginia, killing all but one. It is not just coal mining that comes with occupational risk: hydroelectric power carries the risk of the dams collapsing, possibly leading to catastrophic flooding, injury and death. In 1976, the Teton Dam in Idaho collapsed, flooding the surrounding area and resulting in fourteen deaths. This project gathers domestic data on injuries (fatal and non-fatal) from the U.S. Bureau of Labor Statistics' website, and data on energy produced from the U.S. Energy Information Administration. The goal of this project is to compare and contrast occupational risks (neither health nor environmental) in energy production fields through the years. We also standardize the data relative to total energy produced, number of worksites, catastrophic events vs. day-to-day hazards, etc. For example, coal provides more energy than hydroelectric, and so it is natural to expect more injuries from coal mining. However, when standardized by energy produced, hydroelectric power is occasionally more dangerous, as occurred in the United States in 2005 and 2006, aligning with the passing of the Energy Policy Act of 2005. Since the act provided tax benefits to entities that avoid greenhouse gasses, it is possible that there was a surge in activity in hydroelectric plants, leading to workers being put in more danger. This research will look for explanations to other spikes and dips in the danger of each field.</p> |
| SESUG 264-2018 | Understanding Crime Pattern in United States by Time Series Analysis using SAS Tools | Kar Choudhury, Soumya Ranjan | <p>Crime, be it property or violent, personal or public, has always been a social evil and a drawback to the inclusive development of the society. United States specifically has had a significant reduction in both property and violent crime over the past quarter century. Uniform Crime Reporting, an organization within the FBI, has a data repository constituting different kinds of crimes for 20 years spanning 1995 through 2014 at a state level, collected by several city agencies across the country. Using this dataset, it is possible to forecast crime levels using Time Series analysis in order to understand the severity and provide a ballpark for crimes in the upcoming 3 years (from 2015 to 2017). This forecast can then be compared with the actual crime numbers to realize the accuracy of the forecast. SAS Enterprise guide is used for running the basic forecasting model by state and crime type. SAS Studio and SAS 9.4 are then used to build moving average auto-regressive models by state and crime type. As an end result, all these models are all packaged into one macro in order to facilitate easy viewing of all the models at one place. Several SAS tools are leveraged all through the project to showcase how a combination of SAS products can help build better systems and analytics. This paper illustrates the simplicity of explaining time series forecasting in SAS tools and since UCR doesn't forecast crimes on its website, the above process can be used to provide an estimation for future crime rates in all states. This will help FBI understand the trend of various types of crimes in all the states and thus necessitate preventive measures in accordance with the insights.</p> |
| SESUG 276-2018 | Efficient DATA Step Random Sampling Out Of Thin Air | Dorfman, Paul | <p>Many ETL and data analysis tasks require to generate a random sample of unique K out of N available integers such that $K \ll N$. While it can be done using the SURVEYSELECT procedure, it needs an input data set with N observations. Unfortunately, when N is large enough, it is not a practical option. For example, at $N=1E14$, merely looping from 1 to N would take close to six days. However, the required sample can be created in a DATA step "out of thin air" using algorithms that require looping only from 1 to K regardless of how large N is. Moreover, their bookkeeping memory storage can be kept strictly bounded by K. In this paper, we will discuss a number of such algorithms and their DATA step implementations based both on arrays and the SAS hash object.</p> |

SESUG 2018 Abstracts - by Track/Section Statistics/Data Analysis

| SESUG Paper # | Title | Primary Author | Abstract |
|----------------|---|----------------|---|
| SESUG 289-2018 | A Flexible Approach to Computing Bayes Factors with PROC MCMC | Hicks, Tyler | Whereas frequentist tests focus on p-values, Bayesian tests center on Bayes' factors. The Bayes' factor quantifies the extent to which data support one hypothesis over another. Each test has advantages and disadvantages. For example, the p-value is simpler to calculate but trickier to interpret relative to the Bayes' Factor. The Bayes' factor is trickier to calculate but simpler to interpret relative to the p-value. Although it is complicated to implement the Bayes factor procedure, Markov Chain Monte Carlo (MCMC) methods enables empirical derivation of the Bayes' Factor as opposed to analytic. After considering an example where the Bayes' Factor procedure is preferable to the p-value procedure, this paper describes calculating the Bayes' Factor with the SAS/STAT® MCMC procedure. Although such calculation is not currently a default option in PROC MCMC, this paper shows that built-in functions enable its users to calculate the Bayes' Factor to evaluate models of varying complexity. |
| SESUG 291-2018 | Modeling Longitudinal Categorical Response Data | Stokes, Maura | Longitudinal data occur for responses that represent binary and multinomial outcomes as well as counts. These data are commonly correlated and often include missing values, so any analysis needs to take both of these factors into consideration. This tutorial focuses on using generalized estimating equations for analyzing longitudinal categorical response data, but it also discusses the generalized linear mixed models approach. Strategies such as weighted generalized estimating equations for managing missing data are also discussed, along with the assumptions for these methods. Techniques are illustrated with real-world applications using SAS procedures such as GENMOD, GLIMMIX, and the GEE procedure. *Experience with logistic regression is required for this tutorial. |
| SESUG 298-2018 | Computing Predicted Values and Residuals from Data Sets containing Classification Variables with Larger Numbers of Levels | Littell, Ramon | Linear models frequently include terms to reduce bias in parameter estimation and extraneous variation. Large data sets typically contain observations from multiple sources, such as different locations, time periods or product types. If one or more of the model variables is a classification (CLASS) variable, then the required computations might overwhelm memory capacity and disable computation of model coefficients. This problem can be partially solved by using the ABSORB statement in the GLM procedure, which enables estimation of coefficients of non-classification variables. But it does not permit computation of linear combinations involving coefficients of CLASS variables, such as predicted values or comparisons of levels of among CLASS variables. This paper demonstrates a computational method that can be carried out in data steps that accomplishes the same objectives as the ABSORB statement, but enables computation of predicted values and residuals. An illustrative data set contains prices of machine products that were sold to multitudes of customers. The data were collected to estimate the effects of product cost, demand and a possible change in economic environment on price, adjusted for product effects. (Data for the example were simulated to represent a real data.) The HPmixed procedure has capabilities that overcome the shortcomings of the GLM ABSORB statement, but uses more computational resources, as is briefly shown. |