Paper #	Title	Primary Author	Abstract
APP-15	Generating Reliable Population Rates Using SAS® Software	Shoemaker, Jack	The business of health insurance has always been to manage medical costs so that they don't exceed premium revenue. Monitoring and knowing about these patient populations will mean the difference between success and financial ruin. At the core of this monitoring are population rates like per member per month costs and utilization per thousand. This paper describes techniques using SAS® software that will generate these population rates for an arbitrary set of population dimensions. Keeping the denominators in sync with the numerators is key for implementing trustworthy drill-down applications involving population rates.
APP-36	A macro of building predictive model in Proc Logistic with AIC-optimal variable selection embedded in cross-validation	Yang, Hongmei	Logistic regression leveraging stepwise selection has been widely utilized for variable selection in health care predictive modeling. However, due to the drawbacks of stepwise selection, new ideas of variable selection are emerging, including Akaike Information Criterion (AIC)-optimal stepwise selection which utilizes AIC as the criterion for variable importance and builds a model based on a combination of stepwise logistic regression and information criteria. As predictive factors selected over a single sample may over fit the sample and have poor prediction capability on independent test data, embedding variable selection in resampling techniques, such as cross-validation, is recommended to appropriately estimate expected prediction error, especially with a limited sample size. When processing the AIC-optimal selection through cross-validation, different lists of influential variables may be selected over the iterations. Simply averaging the coefficients would yield a final model with many more predictors than necessary, and therefore reduced predictive accuracy. This paper proposes additional steps to address this issue. Variables selected in the AIC-optimal stepwise process are ranked by their frequency appearing in the AIC-optimal lists obtained from cross-validation iterations. A final model is obtained by sequentially adding the variables with the same frequency until an optimal averaged area under the Receiver Operating Characteristic curve (AUC) is achieved. We present the algorithm and the macro used to achieve the selection in the context of cross-validation. Intended audience: SAS users of all levels who work with SAS/STAT and PROC LOGISTIC in particular.
APP-38	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, the SPEDIS, COMPGED, and COMPLEV 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.

Paper #	Title	Primary Author	Abstract
APP-72	Advanced project management, beyond Microsoft Project, using PROC CPM and 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 allows different individuals to input resource requirements and resource availability and to set priorities. 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.
APP-87	Detecting Outlying Data in Proficiency Studies with SAS	Slone, Stacey	Regardless of the industry, one of the first steps in analyzing data from a collaborative or a proficiency study is to detect any outliers in the data. Regulations published by the International Organization for Standardization, ISO, enumerate several tests to consider (ISO 5725-2, 1994; ISO/FDIS 13528, 2015). To identify outlying means in the laboratories participating in proficiency study, the ISO regulations suggest the single Grubb's test and the Mandel's h test. Similarly, the ISO regulations propose using the Cochran's test and the Mandel's k test to identify outlying standard deviations among participating laboratories. ISO 5725-2 provides a partial table of significant values for these four outlier tests (ISO 5725-2, 1994). However, since these collaborative studies typically analyze multiple parameters, a programmatic solution to finding not only the scores, but the critical values, leads to savings in time and effort. In his 2013 paper, Wilrich (Wilrich, 2013) derived the formulas for the critical values for each of the fore mentioned tests. The included SAS® macro can perform any of these tests on outliers, compare the result with the correct critical values and print a summary or graph as appropriate based on the tests chosen. Options for printing include significant results for the Grubbs or Cochran test, graphs for the Mandel's h and k, with reference lines at the 95% and 99% critical values, and a summary table of results.
APP-99	Ron Fehd, SAS-L's Macro Maven, Answers Your Macro Questions	Fehd, Ronald	SAS(R) software consists of two languages, SAS and its macro language. The purpose of this talk is to provide overview and perspective of how SAS works and how the macro language can work both within and before SAS program statements and steps.
APP-101	Macros for creating a custom report of figures	Williams, Laura	Often, with clinical studies, a report that includes both tables and figures can be preferred. PROC REPORT can be used to place multiple images in a document, such as an RTF file, in combination with summary tables or other text. First, a set of macros have been developed to control the appearance of figures within our organization, using a single template with GTL (Graph Template Language). Plus, the macros also control the file type, size, and resolution of the ODS output. Second, another set of macros have been developed to join these figures into an RTF file and populate the appropriate titles and footnotes. This second set of macros is a joint set of macros that can also produce tables. Therefore, we are able to create reports in a certain style that have a consistent appearance regardless of who programs the analysis output.
APP-109	A SAS macro replacement for Dynamic Data Exchange (DDE) for use with SAS grid computing	Kinney, Saki	The ability to fill in Excel table shells by writing to specific cells of workbooks is an important feature of DDE that users lose when migrating to grid computing. SAS Add-In for Microsoft Office has this functionality but requires programming in Visual Basic and complicates quality control. To replace DDE functionality we have developed a set of SAS macros which use PROC IML to integrate the functionality of R packages XLConnect and openxlsx for working with Excel workbooks into SAS programs. This paper describes the macros, their performance and implementation with SAS grid computing, and capabilities and limitations with respect to DDE.

Paper #	Title	Primary Author	Abstract
APP-123	The %FRED Macro – Finding Risk in Education Data	Edora, Fred	There has been increasing attention given to the analysis of the risk of students who are disproportionally represented in special education settings, such as the overidentification of students in specific disabilities and/or racial/ethnic groups who receive special education services around the country. With recently passed regulations related to the Individuals with Disabilities in Education Act (IDEA), states and local school districts must determine if new standards in disproportionality and risk will affect how they provide services to students with disabilities. The often complicated risk ratio formulas require many repetitive calculations for multiple demographic categories. This makes the issue a perfect candidate for a macro! The %FRED macro ("Finding Risk in Education Data") uses the SAS® macro facility along with array processing and the SAS® Prompt Manager to help stakeholders quickly and more accurately determine if their state or local school district is at risk for disproportionality based on recommended formulas for states to use. This will allow analysts in SAS® Enterprise Guide to run a robust program without having to worry about programming the calculations themselves. This paper will outline the steps from conception to completion.
APP-127	A Macro for Generating the Adverse Events Summary for ClinicalTrials.gov	Moseby, Andrew	In the clinical trials industry, the website ClinicalTrials.gov serves as a publicly accessible outlet of information on trial outcomes. Once a federally sponsored trial concludes, the study team is responsible for publishing adverse events data to ClinicalTrials.gov. These data must be delivered in a very particular structure, uniform across all studies. This uniformity allows for generalization via a SAS macro. The purpose of this paper is to introduce such a macro. Macro variables offer flexibility to account for study-specific characteristics such as number of treatment groups and variable naming conventions. Input datasets may be modified in the work directory in cases where the macro variables do not accommodate a particularly non-standard input dataset.
APP-136	SAS Macros for Time-Dependent Effects and Risk Factors on Survival: Quantile and Landmark Analysis	Zhang, Chao	SAS Macros for Time-Dependent Effects and Risk Factors on Survival: Quantile and Landmark Analysis Chao Zhang, Manali Rupji, and Jeanne Kowalski* Emory University, Atlanta, GA 30329 Abstract: In traditional Kaplan-Meier or Cox regression analysis, typically a risk factor measured at baseline is examined for it's association with survival thereafter. During follow-up however, things may change, such that either the effect of a fixed baseline risk factor may vary over time, resulting in a weakening or strengthening of associations over time, or the risk factor itself may vary over time. In the former case, such as effect is often seen in what appears to be significant differences in survival not necessarily overall, among all survival times, but early on or at later survival times. We address such time-dependent effects on survival by creating two SAS macros, one for landmark and another for quantile survival regression. As opposed to the currently available SAS procedure for quantile survival regression, our quantile survival analysis macro includes a test for significant differences among the levels of a categorical covariate within quantiles of survival time by constructing a plot of the estimated difference in survival time between levels of a categorical covariate that are shown along with 95% confidence intervals (CI). Using these plots, the significant quantiles are readily identified based on whether or not the 95% CI includes zero or not and the corresponding survival times are also displayed. For addressing a baseline risk factor known to vary at a single fixed a priori specified time (a.k.a., landmark time), we developed a SAS macro for landmark survival analysis that constructs Kaplan–Meier plots in two parts: pre- and post-landmark, along with log-rank p-values.

SESUG 2017 Abstracts - by Track/Section
Application/Macro Development

Paper #	Title	Primary Author	Abstract
APP-175	Leads and Lags: Static and Dynamic Queues in the SAS® DATA STEP, 2nd ed.	Keintz, Mark	From stock price histories to hospital stay records, analysis of time series data often requires use of lagged (and occasionally lead) values of one or more analysis variable. For the SAS® user, the central operational task is typically getting lagged (lead) values for each time point in the data set. While SAS has long provided a LAG function, it has no analogous "lead" function – an especially significant problem in the case of large data series. This paper will (1) review the lag function, in particular the powerful, but non-intuitive implications of its queue-oriented basis, (2) demonstrate efficient ways to generate leads with the same flexibility as the lag function, but without the common and expensive recourse to data re-sorting, and (3) show how to dynamically generate leads and lags through use of the hash object.
APP-214	A Second Look at the ODS Destination for PowerPoint	Elsinger, Jane	This paper demonstrates how to use the ODS destination for PowerPoint to create attractive presentations from your SAS® output. Packed with examples, this paper gives you a behind-the-scenes tour of how ODS creates Microsoft PowerPoint presentations. You get an in-depth look at how to customize the ODS PowerPoint style templates that control the appearance of your presentation. With this information you can quickly turn your SAS output into an engaging and informative presentation. This paper is a follow-on to the SAS® Global Forum 2013 paper "A First Look at the ODS Destination for PowerPoint."

Paper #	Title	Primary Author	Abstract
BB-14	Tales from the Help Desk: Solutions to Common DATA Step Tasks	Gilsen, Bruce	In 32 years as a SAS ® consultant at the Federal Reserve Board, I have seen some questions about common SAS tasks surface again and again. This paper collects the most common questions related to basic DATA step processing from my previous "Tales from the Help Desk" papers, and provides code to explain and resolve them. The following tasks are reviewed: 1. Using the LAG function with conditional statements. 2. Avoiding character variable truncation. 3. Surrounding a macro variable with quotes in SAS code. 4. Handling missing values (arithmetic calculations versus functions). 5. Incrementing a SAS date value with the INTNX function. 6. Converting a variable from character to numeric or vice versa and keeping the same name. 7. Converting character or numeric values to SAS date values. 8. Using an array definition in multiple DATA steps. 9. Using values of a variable in a data set throughout a DATA step by copying the values into a temporary array. 10. Writing data to multiple external files in a DATA step, determining file names dynamically from data values. In the context of discussing these tasks, the paper provides details about SAS processing that can help users employ SAS more effectively. See the references for seven previous papers that contain additional common tasks.
BB-24	Quality Control in SAS®: Checking Input, Work, and Output	Brown, Aaron	Part of being a responsible SAS® programmer is checking that one's programs work as intended. Errors and bugs can come from several causes. This paper goes through some tools for checking input data, one's work as one works on the data, and output data. It is unlikely that all of these tools will be applicable for every program, but the author hopes that this paper will give tools that other programmers can use as quality control resources. The author also hopes this paper will emphasize the need for quality control and checking assumptions. These tools utilize the COMPARE, CONTENTS, FREQ, MEANS, PRINT, and SQL procedures.
BB-25	Migrating from PC-SAS to SAS-Grid on Linux	Myers, Susan	SAS Grid is available for Windows or Unix, and it offers features and benefits not available under PC SAS. What challenges do programmers face when an organization makes the move from Windows-based PC SAS to Linux-based SAS Grid? To adapt to the new system, one must address differences between Windows and Linux as well as the differences between PC SAS and SAS Grid, and some batch operations must coordinate programs on both platforms as part of the same overall process. The authors present their experience and lessons learned regarding Windows versus Linux, the use of PuTTY versus MobaXTerm terminal emulators, data management, program conversion and the use of Windows command files that call Linux shell scripts. The resulting systems work well, but the learning curve can be steep.
BB-51	Data Analysis and Storytelling – Communicating Analytical Results with Clarity, Precision and Efficiency	Lafler, Kirk Paul	The data analysis process involves the gathering and collection, cleansing, transforming, and modeling of data from various sources. The purpose is to discover, evaluate, understand and derive useful information from the data to support decision-making. Unfortunately, and all too often, data analysts omit a very crucial step – the development of a narrative, or story, of the data analysis process and outcome. This omission not only fails to bring context, insight and interpretation of the data analysis results to stakeholders, it neglects to bring meaning, relevance and interest to the "key" points of the data analysis results. This presentation describes the importance, considerations and steps needed in developing a compelling narrative, along with the necessary visual analytics, to communicate a convincing point-of-view to help persuade others to understand the complexities associated with the data analysis results.

Paper #	Title	Primary Author	Abstract
BB-67	Old Age and Treachery vs. Youth and Skill: An Analysis of the Mean Age of World Series Teams	DeMaio, Joe	Every October, baseball fans discuss and debate the ages of World Series opponents. However, this comparison is only for the two teams competing for the Commissioner's Trophy. Does the older of the two teams win more often than not? If so, how does the age of that team compare to the rest of the league? Perhaps the older of the two teams won the World Series but was still younger than the league average. Furthermore, there are two distinct types of players in baseball. Batters are needed to score runs. Pitchers are needed to prevent the other team from scoring runs. In 2016, the mean age of the World Series Champion Chicago Cubs' pitchers was 31.1 years and older than the mean age of the Cleveland Indians' pitchers at 25.3 years. The reverse was true for batters. The mean age of Cubs' batters was 27.2 years verses the Cleveland Indians' batters at 29 years. In this talk we examine how player's ages have changed over time, if there exist differences in ages between pitchers and batters and how ages of World Series teams compare against league averages.
BB-68	Hash Beyond Lookups: Your Code Will Never be the Same!	Axelrod, Elizabeth	If you've ever used the Hash object for table lookups, you're probably already a fan. Now it's time to branch out and see what else Hash can do for you. This paper shows how to use Hash to build tables and aggregate data. Why would you ever want to do this? How about achieving a complex process that would have taken multiple sorts and many passes through your data – all in a single data step that flows intuitively and isn't hard to write! I learned this technique from a presentation by Hash masters Paul Dorfman and Don Henderson, and now my code – and how I think about solving problems – will never be the same.
BB-75	Using SAS to Employ Propensity Score Matching in an Institutional Research Office to Create Matched Groups for Outcomes Analyses	Frye, Bobbie	It is common to encounter student unit record data in the community college and to analyze the impact of educational interventions using two groups of students, those exposed to the intervention and those not exposed. Yet results are limited in that the students are not typically randomly selected into experimental and control groups. Non-random selection implies that the two groups of students may be very different on key factors that affect the results of analyses through self-selection bias and other differences. Propensity matching is a technique designed to simulate an experimental design, controlling for selection bias and creating almost equivalent experimental and comparison groups on key indicators. Propensity score matching using key characteristics such as diagnostic/placement test scores, Pell status, age, gender and race/ethnicity will be used to select the experimental and comparison groups. Comparisons of student outcomes using propensity matching has been used to yield less biased results than are derived using simple comparisons (Rojewski et al., 2010).
BB-81	Let SAS® Do Your DIRty Work	Watson, Richann	Making sure that you have saved all the necessary information to replicate a deliverable can be a cumbersome task. You want to make sure that all the raw data sets and all the derived data sets, whether they are Study Data Tabulation Model (SDTM) data sets or Analysis Data Model (ADaM) data sets, are saved. You prefer that the date/time stamps are preserved. Not only do you need the data sets, you also need to keep a copy of all programs that were used to produce the deliverable, as well as the corresponding logs from when the programs were executed. Any other information that was needed to produce the necessary outputs also needs to be saved. You must do all of this for each deliverable, and it can be easy to overlook a step or some key information. Most people do this process manually. It can be a time-consuming process, so why not let SAS® do the work for you?

Paper #	Title	Primary Author	Abstract
BB-96	SAS and the Voluntary Framework of Accountability: A Prime Example of the Use of SAS in Education	Smith, Kelly	Community colleges in the United States provide accessible learning opportunities for more than 10 million students each year. The student population served by community colleges often differs from the student population served by four year colleges. Consequently, accountability metrics developed for four year colleges are not always appropriate for two year community colleges. Metrics designed to better evaluate the effectiveness of community colleges were recently established as part of the Voluntary Framework of Accountability (VFA). The metrics used in the VFA were developed by the American Association of Community Colleges, the Association of Community College Trustees and the College Board Advocacy and Policy Center. VFA metrics utilize data that community colleges are already collecting for internal and external reporting (state and federal). College data is provided to VFA through the production and submission of CSV (comma delimited) files. The creation of these data files requires SAS users to utilize a wide array of techniques in extracting, recoding, and merging data from multiple data sources at the college. This paper summarizes the development process and testing of SAS code to create the VFA datasets at Central Piedmont Community College (CPCC) in North Carolina. Particular attention is paid to the sub-setting, merge, and export procedures that were repeatedly utilized in the dataset creation process. The process of creating SAS code for the VFA datasets provides an excellent example of the use of SAS in the field of education.
BB-103	A Tutorial on the SAS® Macro Language	Cohen, John	The SAS Macro language is another language that rests on top of regular SAS code. If used properly, it can make programming easier and more fun. However, not every program is improved by using macros. Further-more, it is another language syntax to learn, and can create problems in debugging programs that are even more entertaining than those offered by regular SAS. We will discuss using macros as code generators, saving repetitive and tedious effort, for passing parameters through a program to avoid hard coding values, and to pass code fragments, thereby making certain tasks easi-er than using regular SAS alone. Macros facilitate conditional execution and can be used to create program modules that can be standardized and re-used throughout your organization. Finally, macros can help us create interactive systems in the absence of SAS/AF. When we are done, you will know the difference between a macro, a macro variable, a macro statement, and a macro function. We will introduce interaction between macros and regular SAS language, offer tips on debugging macros, and discuss SAS macro options.
BB-117	Joining Data in SAS - SQL or MERGE?	Droogendyk, Harry	This paper explores the joining of datasets / tables using both the data step MERGE and PROC SQL. Similarities between the two methods are identified and occasions when one method might be preferred over the other are discussed. Specific issues relating to INNER, OUTER and FULL joins are covered, as are the vagaries of the SQL ON vs. WHERE clauses. An array of examples will illustrate exactly how joins are accomplished in both data step and SQL environments. Included in the presentation is a method of displaying SQL's inner workings providing hints for query optimization.

Paper #	Title	Primary Author	Abstract
BB-119	SAS® Survey Report Macro for Creating User-Friendly Descriptive Summaries	Dickenson, Tammiee	The use of online platforms such as Survey Monkey and Qualtrics allow for ease of survey administration. Survey items are entered into the system and the data are collected directly from respondents electronically. While these online platforms provide summaries of results, the reports are often not in a user-friendly format that is ready to share with stakeholders. The purpose of our survey report macro is to prepare descriptive summaries of item-level results from surveys collected from online systems in a user-friendly format. Features of the macro code include data steps using arrays and do loops to organize the data for analysis, PROC TABULATE to produce accessible tables of results, and use of the Output Delivery System (ODS) to create reports in Word or Excel format. The survey report macro produces survey results for Likert-type items from a single survey administration. In our work as program evaluators, our projects typically include multiple schools that are implementing a common program. We often wish to summarize survey results for all schools within the project to inform the program overall and for individual schools to share with school personnel for use in program planning at the school-level. The survey report macro allows flexibility to run analyses for different subsets of participants for whom the user wishes to report results. There are also user-specified choices for reporting options built into the macro code.
BB-122	Table Lookups: Getting Started With Proc Format	Cohen, John	Table lookups are among the coolest tricks you can add to your SAS® toolkit. Unfortunately, these techniques can be intimidating both conceptually and in terms of the programming. We will introduce one of the simplest of these techniques, employing Proc Format and the CNTLIN option as part of our construct. With any luck, this will prove both easy-enough to program and more efficient to run.
BB-132	Tired of CALL EXECUTE? Try DOSUBL	Fan, Jueru	DOSUBL was first introduced as a function in SAS® V9.3. It enables the immediate execution of SAS® code after a text string is passed. Macro variables that are created or updated during the execution of the submitted code are exported back to the calling environment. With this feature it can be treated as a powerful alternative to traditional CALL EXECUTE when creating macros. This paper demonstrates the principal difference between DOSUBL and CALL EXECUTE with a real-world application. With a thorough understanding of the discussion in this paper a reader should be able to apply the technique more generally.
BB-133	Automatic Verification of Combined Datasets	Maher, Marcus	One common task for a SAS programmer is to combine datasets from different sources, perhaps from multiple rounds of a study or from different panels in a survey. When combining these datasets, it is important to ensure that the final dataset(s) have the correct combination of rows, both that they have all of the expected rows from each source dataset and that the rows from the source datasets are not in conflict. We present a solution for automatically verifying a combined dataset. We verify that every row in the source datasets has a matching row in the combined dataset, that every row in the combined dataset comes from a unique source, that the ID variable in the final dataset is unique, and provide several examples of possible logic checks that can be added to the application depending on needs, such as verifying that the study round matches the dataset source. The intended audience for this presentation is an intermediate programmer. We use hash tables in this application, but present them in sufficient detail for a programmer with no background in hash tables to understand their purpose and use.

Paper #	Title	Primary Author	Abstract
BB-135	Analysis of Tweets on Demonetization in India using SAS Enterprise Miner	Patel, Jital	Author(s): Jital Patel, Narmada Panneerselvam University: Oklahoma State University, Stillwater, Oklahoma, USA Location of Research: Stillwater, Oklahoma, USA Mentor(s): Narmada Panneerselvam, Dr. Goutam Chakraborty and Dr. Miriam McGaugh The citizens of a country often face the benefits or the brunt of various policies adopted by the government. Social platforms then become sounding boards for them to express their displeasures or concerns for the matter at hand. One such policy that sent Twitter into a frenzy was the demonetization order that was announced rather abruptly by the Indian Prime Minister on November 8th 2016 without any form of prior intimation to public. In this paper, we have analyzed the tweets that helped us recognize whether demonetization was perceived positively or negatively by the citizens. About 15 days after the demonetization decision was announced, a dataset of 8,000 tweets spanning two days, was collected from a publicly available data source. Using the commonly used terms in the tweets and studying the strength of their relations using concept links, in general, show a positive feedback to the demonetization implementation policy. Using text clustering and text topic to group people with similar thoughts based on the tweets, reveal that demonetization was positively supported by a large number of people. Thus, our overall analysis showed that a vast majority of Indians accepted the demonetization policy positively while some of them expressed their displeasure over it.
BB-139	Building a Sequential Programs for a Routine Task with Five SAS Techniques	Yu, Gongmei	When a task needs to be implemented on a regular basis and has a tight timeline, the SAS programs should be designed with minimum required updates to the code and maximum automation in each processing step. This paper intends to illustrate the process of developing such programs with five SAS techniques: (1) Macro variable; (2) Macro program; (3) Conditional and iterative statement; (4) %Include statement; (5) DDE (Dynamic Data Exchange) outputting. This paper assumes basic knowledge of SAS procedures and macro language and the use of program logic. The emphases of the paper is to illustrate how the five techniques can be applied to the specific task of calculating healthcare payment weights. The code may include other techniques not addressed in this paper.
BB-144	Beyond IF THEN ELSE: Techniques for Conditional Execution of SAS Code	Horstman, Josh	Nearly every SAS® program includes logic that causes certain code to be executed only when specific conditions are met. This is commonly done using the IF…THEN…ELSE syntax. In this paper, we will explore various ways to construct conditional SAS logic, including some that may provide advantages over the IF statement. Topics will include the SELECT statement, the IFC and IFN functions, the CHOOSE and WHICH families of functions, as well as some more esoteric methods. We'll also make sure we understand the difference between a regular IF and the %IF macro statement.
BB-145	Merge with Caution: How to Avoid Common Problems when Combining SAS Datasets	Horstman, Josh	Although merging is one of the most frequently performed operations when manipulating SAS datasets, there are many problems which can occur, some of which can be rather subtle. This paper examines several common issues, provides examples to illustrate what can go wrong and why, and discusses best practices to avoid unintended consequences when merging.
BB-158	Using SAS® for Application Programming Interface Requests	Jadoo, Mike	Application Programming Interface (API) is a method of requesting data and has been around for some time. Their primary uses are for front-end webs developers who wish to use these respective data sets for charts, tables, and maps. However, this method of data request can also be useful for data processing and analysis as fewer steps in the data products process can be achieved. In the paper we will discover how an API request can remove steps in the data production process and how to make a request for data from a major statistical accounts producer. Moreover, some useful tips will be relieved when using this method.

Paper #	Title	Primary Author	Abstract
BB-170	Using the Force of Python and SAS Viya on Star Wars Fan Posts	Heyne, Grace	The wealth of information available on the Internet includes useful and interesting data, and it is easy to gather using the connection between SAS Viya™ and Python. Since the announcement at SAS Global Forum 2016 that SAS Viya™ supports running CAS actions within Python, SAS programmers have been exploring the many ways that SAS works well with Python. With easy transitions between the Python language and SAS, this connection allows programmers to use the best of both languages to build more powerful tools for data analysis. This paper includes specific examples of ways to use both SAS and Python together in Jupyter Notebooks to leverage Python's data collection, cleaning, and management tools, along with ways to gather data from the web and perform necessary transformations to prepare and analyze the data.
BB-177	Unleash the Power of PROC REPORT With the ODS Excel Destination	Sekar, Devi	A new ODS destination for creating Microsoft Excel workbooks is available with SAS® 9.4M3. This destination is an extremely easy and handy tool for producing ad-hoc as well as production Excel reports. The ODS EXCEL destination has several advantages over ODS ExcelXP tagset. With the ODS EXCEL destination, you can bring all those powerful features available with the REPORT procedure such as predefined styles, traffic-lighting, custom formatting, and compute block flexibility straight into your Excel reports. Once you start using the ODS EXCEL destination, you will quickly realize that the PRINT procedure is not sufficient to meet all the formatting demands for your Excel Reports. This paper covers various techniques that you can use with PROC REPORT and the ODS EXCEL destination, to make your Excel reports pretty and publication-ready!
BB-178	One-to-one, One-to-many, and Many-to-many Joins Using PROC SQL	Shankar, Charu	A powerful and essential PROC SQL programming technique that all SAS users should understand, and be comfortable performing, is the process of joining (or combining) two or more tables of data. This presentation describes and illustrates the join process, including what a join is, exploration of one-to-one, one-to-many, and many-to-many data relationships, identifying a primary key (or unique identifier), special preparation requirements for each table being specified in a join, and popular join techniques available to SAS users. To demonstrate the power of the join process, examples of conventional (symmetrical matching) and unconventional (asymmetrical left, right and full matching) using PROC SQL join programming techniques are illustrated.
BB-180	PROC IMPORT and more. Or: when PROC IMPORT just doesn't do the job	horvath, david	PROC IMPORT comes in handy when quickly trying to load a CSV or similar file. But it does have limitations. Unfortunately, I've run into those limitations and had to work around them. This session will discuss the original CSV specification (early 1980's), how Microsoft Excel violates that specification, how SAS PROC IMPORT does not follow that specification, and the issues that can result. Simple UNIX tools will be described that can be used to ensure that data hilarities do not occur due to CSV issues. Recommendations will be made to get around some of PROC IMPORT limitations (like field naming, data type determination, limitation in number of fields, separator in data). CSV, TAB, and DLM types will be discussed.
BB-184	Zen and the Art of Problem Solving	horvath, david	Although software development is taught as a STEM out of Science or Engineering schools, it is as much an art or craft a creative process as a science. This presentation focuses on innovative problem solving techniques the tools and techniques to use when your normal process just doesn't seem to get you to a solution. Much of the information in this talk is based on Robert Pirsigs Zen and the Art of Motorcycle Maintenance, which, although it is focusing on Motorcycles, applies to all kinds of problem spaces (and Pirsig was a tech writer for IBM). These techniques have served me well over the years. The difference in art versus science approaches is actually supported by the way the brain works. * Traditional Problem Solving Methods, where they fall down * Eastern techniques, advantages, disadvantages * Tips * Brain versus Mind

Paper #	Title	Primary Author	Abstract
BB-186	UCF Visual Analytics - Dataset Development for the Undergraduate Attrition Study	Milbuta, Scott	This presentation will focus on the Development cycle of the underlying data in order to provide the nessesary elements required for implementation in a comprehensive SAS Visual Analytics envirionment. Topics will include customer requirements, development strategy, data collection and manipulation, development challenges and images of the final Report visualizations.
BB-193	Setting the Percentage in PROC TABULATE	Franklin, David	PROC TABULATE is a very powerful procedure which can do statistics and frequency counts very efficiently, but it also it has the capability of calculating percentages on many levels for a category. This paper looks at the automatic percentage calculations that are provided, and then delves into how a user can specify the denominator for your custom percentage.
BB-210	SAS® Spontaneous Combustion: Securing Software Portability through Self-Extracting Code	Hughes, Troy	Spontaneous combustion describes combustion that occurs without an external ignition source. With the right combination of fire tetrahedron components—including fuel, oxidizer, heat, and chemical reaction—it can be a deadly yet awe-inspiring phenomenon, and differs from traditional combustion that requires a fire source, such as a match, flint, or spark plugs (in the case of combustion engines). SAS® code as well often requires a "spark" the first time it is run or run within a new environment. For example, SAS programs may operate correctly in an organization's original development environment, but may fail in its production environment until necessary folders are created, SAS libraries are assigned, control tables are constructed, or configuration files are built or modified. And, if software portability is problematic within a single organization, imagine the complexities that exist when SAS code is imported from a blog, white paper, textbook, or other external source into a new environment. The lack of software portability and the complexities of initializing new software often compel development teams to build code from scratch rather than attempting to reuse or rehabilitate existent code. One solution is to develop SAS code that flexibly builds and validates its environment and required components during execution. To that end, this text describes techniques that increase the portability, reusability, and maintainability of SAS code and demonstrates self-extracting, spontaneously combustible code that requires no spark.
BB-216	SAS® and Hadoop: The 6th Annual State of the Union	Kent, Paul	The fourth maintenance release for SAS® 9.4 and the new SAS® Viya platform bring even more progress with respect to the interoperability between SAS® and Hadoop (the industry standard fro big data). This talk brings you up-to-date with where we are: more distributions, more data types, more options. And then there is the cloud. Come and learn about the exciting new developments for blending your SAS processing with your shared Hadoop cluster.

Paper #	Title	Primary Author	Abstract
BB-223	Challenges and Solutions for Handling Re-screened Subjects in SDTM	Quick, Charity	A common problem for SDTM is tabulating data for subjects who enroll multiple times in a single trial. Currently, the FDA advises that as long as there is one USUBJID for each specific subject, a different subject identifier (SUBJID) value can be used for each screening attempt. It can be challenging to handle this data while conforming to both the SDTM Implementation Guide and FDA-published validation rules. Here, I suggest retaining one record in the Demographics domain (DM) for a trial subject using the latest SUBJID value and saving the previous subject identifiers in SUPPDM. From there, the ability to link observations with previous ID values in Supplemental data and adequate usage of VISIT and EPOCH values will facilitate mapping of most trials involving re-screened subjects. Specific considerations will include: 1. Identifying the subset of domains which will contain data from multiple enrollments. 2. Managing possible duplicate records when the same events or treatments can be reported at multiple screening visits, such as for the Adverse Events (AE), Concomitant Medications (CM), or Medical History (MH) domains. 3. Populating Findings domains as well as Subject visits (SV) and Disposition (DS) domains for multiple enrolled subject records.
BB-224	The Future of the SAS Platform	Author TBD, SAS	SAS has delivered integrated capabilities that organizations use to access, explore, transform, analyze, and govern data, delivering trusted insights on time and at scale. Over the last few years, recent trends in business and technology such as cloud, open APIs, microservices and new and emerging use cases have driven a need for the SAS platform to undergo an evolutionary transformation. This presentation will describe that transformation, how it came about and where the SAS platform is headed in the future.
BB-226	Finding the Gold in Your Data: An Overview of Data Mining	Dickey, Dave	The term "data mining" has appeared often recently in analytic literature and even in popular literature, so what exactly is data mining and what does SAS* provide in terms of data mining capabilities? The answer is that data mining is a collection of tools designed to discover useful structure in large data sets. With an emphasis on examples, this talk gives an overview of methods available in SAS Enterprise Miner and should be accessible to a general audience. Topics include predictive modeling, decision trees, association analysis, incorporation of profits and neural networks. We'll see that some of the basic ideas underlying these techniques are closely related to standard statistical techniques that have been around for some time but now have new more appealing names than their statistical ancestors and have been automated to become more user friendly.

Paper #	Title	Primary Author	Abstract
CC-26	SAS Does Not Store Dates	Wooding, Nat	Posts frequently appear on discussion boards with a statement like: 'I have a date stored in SAS(R) in ... format". This paper is intended for newer programmers who do not fully understand how dates can be read into SAS data sets and why this statement is not meaningful or at best confusing. The paper will also touch on why date values read from external sources will often not appear to be processed by SAS date handling functions.
CC-35	A practical application of Vtable	Subramaniam, Sumathi	You may have read about "Dictionary.Tables", which is one of the available sources of information available for developers to take advantage of Meta data based on different requirements but this paper is talks about a simple but powerful application with an example. Several articles have been published in the past many years but this paper provides a practical example of how it is used in some financial transaction processing in some leading banks. We have some processes built on a requirement that looks for transactions recorded in the past 30 days or past 10 days depending on the business requirement. They can be based on files that may have been on an irregular schedule. This calls for a way to look up the creation date of the file within SAS. Fortunately, SAS provides session tracking variables that are available through Vtable which is available through SASHELP library. This not only stores information about SAS files that are part of the standard SAS libraries, but also stores information from a users current SAS session and libraries defined within the program and file names defined within the program. The columns those are available for reading provides useful information that allows a user to operate with much flexibility. In this paper, I show example code of how the files are created and the corresponding Meta data is stored in the SASHELP.Vtable Meta data dataset and how it can be retrieved to dynamically build code that can further be used in subsequent processing. This is just a tip of the iceberg but it is up to the user to take advantage of the available information to the best use as deemed appropriate.
CC-41	Dictionary.columns is your friend while appending or moving data	venna, Kiran	Dictionary.columns is a dictionary table which gives metadata information of columns in various tables. Dictionary.columns is very helpful in various data related issues. Data integrity issues can be understood while appending new data from external files to SAS Datasets with help of Dictionary.columns. Appropriate appending of external files to SAS Datasets can be done with help of SAS Macros. Data movement from Hive to Teradata by using SAS will cause an issue when variable name length is greater than 32. This longer variable name issues can be evulated with the aid of Dictionary.columns. This variable names length issue can be circumvented by Proc FedSQL. While inserting data from SAS to RDBMS different column lengths can cause issues and this issues can be analyzed with help of dictionary.columns
CC-60	Sending SAS® Data Sets and Output to Microsoft Excel®	Go, Imelda	For many of us, using SAS and Microsoft Excel together is inevitable. This paper is useful for a beginner who wants to know how the two can be used together and how to send SAS data and/or output to Excel. A number of SAS features facilitate the sharing of data between the two. There are tools in SAS Enterprise Guide (EG) that help send output data into Excel. With SAS/ACCESS, PROC EXPORT can convert a SAS data set into an Excel file. The Output Delivery System (ODS) offers a number of destinations, such as the Excel destination that facilitates the transfer of SAS output into Excel. Using Dynamic Data Exchange (DDE) is a way for a user to populate a preexisting or preformatted Excel file with SAS data. SAS traffic lighting (color coding of SAS output elements) in a report can also be applied to SAS output such that the traffic lighting will appear in Excel.

Paper #	Title	Primary Author	Abstract
CC-98	Advanced Programming Concepts: History of the List Processing and Cardinality Ratio Memes	Fehd, Ronald	A statement in a natural language contains three parts: subject, verb and object. A statement in a computer language contains only two parts: verb and an object; the subject, or actor, is the computer's operating system. The predecessor of the meme list processing is the computer language LISP, in which every statement is a function call and the object is a list. The list processing paradigm of programming contains these steps: 1. identify an object, an \textit{item} 2. write a function, process, or procedure for an item 3. prepare a list, a \textit{list} of items 4. use a loop on the list, to process each item . This paper reviews the author's development of the concept of list processing and its implementation in SAS(R) software. The purpose of this exposition is to highlight the author's papers published on these topics and to provide a critique of earlier ideas. This is accomplished by a review of the development of processes for calculating the cardinality ratios of the variables in a data set.
CC-112	Extracting Clinical Research Data from OnCore® and Preparing it for Statistical Analysis in SAS®	Xu, Chang	This paper outlines a simple and straightforward method to extract data from the OnCore® data base system and to prepare it for analysis using SAS®. With this method, the user only needs to have basic online permission in OnCore® along with basic knowledge of SAS® and Excel. Example code and instructions are provided.
CC-115	Automate Secure Transfers with SAS and PSFTP	Thompson, Kyle	PSFTP, or PuTTY SFTP offers a way to securely send files between computers. SAS can be used to automate this secure transfer of files with the versatile and encrypted SSH-2 scheme via the free, open-source PuTTY application. After reviewing the use of the command line interface and associated code in SAS, we describe PuTTY commands that can be used for single-line execution or coupled together into batch submission within a SAS program running in Windows. From there, we offer examples of SAS code which can be used in whole or in part to integrate and automate PuTTY's use within a SAS program.
CC-116	Identifying and Removing Subjects with Duplicates in Long Data Format: A Simple Approach Using PROC SORT and PROC SQL	Chao, Szu-Fu	This paper will illustrate an efficient two-step method to identify and remove all examinees with duplicate observations in long format testing data, rather than simply the redundant records. This approach applies in the former case for instances where each row in the dataset contains only one item record per subject, and is generally implemented only when the sample size is sufficiently large and the number of participants who have duplicate records only account for a small proportion (e.g., < 3%) of the overall data set. The issue with duplicates is resolved by implementing the combination of the NODUPKEY and DUPOUT options in PROC SORT and one of two queries in PROC SQL. Although the option NODUPKEY from PROC SORT can indeed provide a good way to handle duplicates (i.e., removing redundancies and only keeping intended observations with appropriate key variables specified), there seems to be no explicit function in SAS to remove both the primary and duplicate records for these subjects. However, the goal can be achieved easily by applying one additional step in PROC SQL with the information extracted from the data generated by the option DUPOUT in PROC SORT. Variations can be implemented to either create a new data set or overwrite the original data directly. This paper provides a quick and efficient way of producing duplicate-free long format data with problematic subjects excluded. A complete example of the method will be demonstrated using data from a computer-based assessment field test with 32 items and 198 examinees that initially producing 5960 rows in total. People from different professional backgrounds who may be dealing with similar issues are likely benefit from the techniques demonstrated in this paper with relatively limited effort.

Paper #	Title	Primary Author	Abstract
CC-125	An Introduction to Visit Window Challenges and Solutions	Ngo, Mai	In clinical trial studies, statistical programmers often face the challenge of subjects' visits not occurring on the exact scheduled visit dates. As a result, visit windowing is often needed for analysis purposes. This paper aims to provide a general introduction to visit window programming. I present some different scenarios where visit windowing is needed and sample SAS® codes for each of these cases. I also discuss strategies for applying the visit window rules to multiple programs with different visit frequencies and window lengths and cover validation strategies in these cases.
CC-130	Identifying Semantically Equivalent Questions Using Singular Value Decomposition	Reddy Akkaloori, Varsha	Quora is a Q&A social network centered on the idea of promoting ideas and knowledge. With over 100 million monthly visitors, it's not surprising that many people ask similarly worded questions causing site visitors to spend more time discovering the best response to their question. This also frustrates authors because they feel they need to answer multiple versions of the same question. This paper aims at solving a challenge released by Quora to improve the experience of its authors and site visitors by grouping queries with similar intent using SAS. To ensure that different words are processed equivalently as the same representative parent term, Pydictionary module in Python, based on thesaurus.com, was used for extracting synonyms for the most frequently occurring terms in the term-by-document matrix. With the help of SAS Enterprise Miner, singular value decomposition (SVD) was implemented to reduce the dimensions of the term-by-document frequency matrix. Euclidean distance was used to determine distance between sentences that have been projected into the SVD space. In addition, Inverse Document Frequency weight was chosen so that frequently occurring terms will have been down-weighted and the rarer but concentrated terms have the greatest influence on similarity. The accuracy of the classification of question pairs was 62.4%. Further research would be continued to make a utility which would predict if a question is duplicate based on the prior knowledge imbibed into it thereby acting as a recommender system for Quora.
CC-134	Format-o-matic: Using Formats To Merge Data From Multiple Sources	Maher, Marcus	User-defined formats are often the best way to merge a single value onto a larger dataset, but they can seem overly complicated for the novice programmer to use, particularly when remembering all of the specific variable names and details like the 'OTHER' row. We first explain the use case for user-defined formats and some of the considerations to keep in mind when using them. Then, we present a single macro that creates a user-defined format from an already existing dataset, with parameters for all of the commonly used options and some of the less common ones, written for maximum flexibility. The macro is intended to be able to be used by novice programmers without complete knowledge of the workings of the process, but the advanced options make it appropriate for any level of programmer. This presentation does not require any understanding of user-defined formats, or even SAS formats at all. The intended audience is novice and intermediate level programmers, as well as anyone interested in an off-the-shelf user-defined format macro.

Paper #	Title	Primary Author	Abstract
CC-140	Backward Variable Selection for Logistic Regression based on Percentage Change in Odds Ratio	Kwiatkowski, Evan	Variable selection is a fundamental component of statistical modeling. A common variable selection method used in health sciences is backward variable selection, which iteratively removes variables based on their relevance to the model. Often, automated backward variable selection procedures determine variable relevance based on overall statistical significance. However, many epidemiologists, including formative thinkers Greenland and Robins, favor a "change-in-estimate" approach to variable selection rather than an overall significance approach. We developed a macro to implement a backward variable selection procedure for logistic regression using the "change-in-estimate" method. Our macro implements backwards variable selection in the logistic regression model in the situation where there is a single independent variable (IV) and single dependent variable (DV) of interest, with additional covariates that are eligible for removal based on their relevance to the model. This relevance is based on the percentage change in odds ratio between the IV and DV in a full model including all additional covariates and a reduced model which removes a single covariate at a time. This macro provides epidemiologists and other health science professionals with a theoretically sound option for automated backward variable selection in logistic regression, and is an extension of backward variable selection options provided in the LOGISTIC procedure. The macro is easily implemented in any dataset by having the user specify the IV, DV, additional covariates, and threshold of difference in odds ratio which is used for removal of additional covariates.
CC-148	RETAINing YOUR SANITY: 5 IDEAS TO MANIPULATE DATA USING THE RETAIN STATEMENT	Taylor, Matthew	SAS® PROCs are generally structured to work with long, vertically oriented datasets. The challenge with datasets that are created this way is that it is very easy to work with data across observations but it is cumbersome to try to work with data vertically. Thus enters the RETAIN statement, a way to hold values across observations and assist in those comparisons. This paper will review five ways to use the RETAIN statement to solve common dataset related problems.
CC-153	Applying IFN and IFC Functions	erinjeri, jinson	The SAS® IFC and IFN functions are helpful in compact as well as in elegant coding. Both IFC and IFN functions can be employed in Base SAS and PROC SQL as an alternative to the traditional IF THEN ELSE, SELECT END and SELECT CASE statements. The paper presents the applications of IFN and IFC functions in various scenarios along with a peek at efficiency comparisons of some of the traditional approaches. Also, presented are pitfalls using these functions and the ways around the pit for a successful run.
CC-159	ODS Basics	Jadoo, Mike	Using the output delivery system (ODS) can have many advantages for SAS® users. This paper will go over some of the basics on how to use ODS in your code. Moreover, different options will be presented using this method.
CC-194	Everyone can use a little Currency – when dependent data set updates silently make your analysis data set out of date.	Worrell, Scott	Nearly every analysis data set has dependencies on multiple other data sets. The dependent data sets may be raw data sets or other analysis data sets. At times a dependent data set is updated, but the person in charge of updating the analysis data set is not notified of the modification to the dependent data set. This makes the analysis data set with the data dependency invalid. In turn, this condition produces data integrity issues in every Table, Listing or Figure (TLFs) that uses the analysis data set. Before running TLFs, it is a good idea to verify that all dependent data sets are older or the same age as the analysis data sets. The Data Currency Utility detects data set currency issues and reports them so that they can be corrected through an update run of the analysis data set creation program. This presentation is appropriate for persons with all levels of SAS experience, but assumes basic understanding of SAS's Dictionary Tables. The utility was written and tested using SAS 9.4 on the Linux operating system. * Paper was presented at PharmaSUG 2017 in Baltimore

Paper #	Title	Primary Author	Abstract
CC-197	Creating a DOS Batch File to Run SAS® Programs	Franklin, David	We often have many SAS programs to run in a directory. While it is possible to run each individually, it is better if a DOS Batch file be created with the list of programs being run and the order in which they are run. This paper looks at a SAS macro that will take the list of SAS programs in a directory, as specified by the user, and create a DOS Batch file to which can then be run to run all the SAS programs. Also presented will we a small SAS program that you can run at the end of the program to send you an email saying when the programs had finished, and whether there are any issues in the SAS LOG to review!
CC-204	Dynamically Assigning Length to Transposed Variables	Ritchie, Ethan	When using PROC TRANSPOSE to transform "narrow" data (a single subject variable stored in many rows) into "wide" data (one row with subject values stored as distinct variables), the length of the original subject variable is used as the length for each new variable. As a result, variable lengths may be much larger than necessary. This paper will demonstrate a method to automatically assign the smallest length necessary to each variable once the "narrow" data is transposed into "wide" data. The method is currently used on large sample survey data that is updated daily and accounts for the possibility that the optimal length of a variable may change from day to day.
CC-220	SAS ODS EXCEL Destination: Using the ID Option to Write Different Excel Workbooks at the same time.	Benjamin Jr, William E	One of the features of the new ODS EXCEL destination is the ID option. This feature allows you to define different output Excel Workbooks that use the same SAS Code but use different output options when writing an Excel Workbook. I will show you how to create three workbooks at once using this ODS option.
CC-221	SAS ODS EXCEL Destination: Using the STYLE Option to spruce up your Excel output workbook.	Benjamin Jr, William E	The SAS environment maintains many different output styles to use to enhance the visual display of your output data. The ODS EXCEL destination can take advantage of these SAS maintained styles to apply formatting and color schemes to your EXCEL output workbooks. I will show you how to use the ODS Excel destination STYLE option to enhance your output workbooks.
CC-222	SAS ODS EXCEL Destination: Using the START_AT sub-option to place your data where you want it on the Excel Worksheet.	Benjamin Jr, William E	The SAS ODS Excel destination statement option called "OPTIONS" has many sub-options. One of those sub-options is START_AT which allows you to place your output anywhere on the open output Excel worksheet. I will show you how the put your starting output data element into any row and column in the open worksheet.

Paper #	Title	Primary Author	Abstract
DM-57	Data movement issues: Explicit SQL Pass-Through can do the trick	venna, Kiran	This paper presents few case studies to enumerate benefits of Explicit SQL Pass-Through in ETL and SAS Macros. Data movement between Teradata and SAS will have huge impact on run time of SAS Job. Usage of Explicit SQL Pass-Through will reduce Data movement and also improves query performance. Proper ETL planning by using Explicit SQL Pass-Through whenever possible will be very helpful in attaining required results by using optimal resources. Explicit SQL Pass-Through can also enhance execution of SAS macros, when a Teradata Table is used to create a SAS data set. This paper also explores how to implement popular SAS character, numeric and date functions along with few SAS statements in Teradata, when using Explicit SQL-pass through.
DM-74	Identifying Duplicate Variables in a SAS ® Data Set	Gilsen, Bruce	In the big data era, removing duplicate data from a data set can reduce disk storage use and improve processing time. Many papers have discussed removal of duplicate observations, but it is also useful to identify duplicate variables for possible removal. One way to identify duplicate variables is with PROC COMPARE, which is commonly used to compare two data sets, but can also compare variables in the same data set. It can accept a list of variable pairs to compare and determine which variable pairs are identical. This paper shows how to obtain a summary report of identical variables for all numeric or character variables in a data set, using the following steps: 1. Dynamically build a list of all possible numeric or character variable pairs for PROC COMPARE to analyze. 2. Convert PROC COMPARE pairwise results (e.g., N1 is identical to N3, N3 is identical to N5, N1 is identical to N5, etc.) into a summary report that groups all identical variables (e.g., N1, N3, and N5 are identical). For very large data sets, the paper shows how to substantially improve performance by first executing PROC COMPARE one or more times on a small number of observations to reduce the number of extraneous comparisons.
DM-77	A Software Toolkit for Data Management	Chantala, Kim	We have developed a suite of data tools to help keep data preparation on schedule and within budget. This toolkit provides a remarkably easy, low-cost way to create codebooks, master lists of SAS data sets for a project, reports of variables needing special investigation, and data crosswalks showing the relationship of variables across datasets. Traditionally, these documents are produced at the end of a project with a great deal of programming or are manually produced, but our tools allow the programmer to seamlessly create these documents at any time during the data preparation task. We have found that producing these documents early in data collection improves data quality and communication between the data collection team and client. This toolkit provides a comprehensive way to document and review not only datasets you create, but also datasets that you receive, especially if they do not have good documentation. Our paper illustrates how to run these macros and provides tips to add embellishments or tailor the documents to your project needs.
DM-85	Parallel Processing in a SAS Grid	Rabb, Merry	In 2016 RTI International migrated over 400 SAS users from SAS on PCs and several stand-alone Linux servers to a SAS Linux Grid. In addition to other improvements new grid provided an opportunity to improve the performance of some of our long running jobs. SAS jobs often contain independent tasks that can be split up and distributed across multiple nodes of a SAS Grid. Running independent multiple tasks in parallel will usually result in a reduction in total elapsed time for the job. This paper will discuss practical considerations for introducing parallel processing into a SAS grid job, including identifying types of concurrency and how best to divide the program into subtasks. We will cover the basics of how to add the required code, including how to pass information from one task to another. This paper will provide a sample program to illustrate code modifications to take advantage of parallel processing on the grid, and cover some real life examples of programs that have been move to the grid using these techniques.

Paper #	Title	Primary Author	Abstract
DM-108	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.
DM-171	Statistician's secret weapon: 20 ways of detecting raw data issues	Liu, Lixiang	Unclean clinical raw data is always statistician and statistical programmer's nightmare for all the downstream SDTM, ADaM and TFLs development work. Raw data issues could mess up the programming logic, create OpenCDISC reject, error, warning messages, and worst of all, if incorrect data is analyzed, study team could draw wrong or inaccurate conclusions regarding drug's safety and efficacy, which could put patients' safety in jeopardy and have significant impact on company's financial status. This paper will review 20 effective ways of detecting raw data issues. Since they are applied to the drug dispense, labs, and safety related CRF data (Including adverse event, medical history, concomitant therapy, drug exposure etc.), which are common for all clinical trials, these methods and their associated SAS programs could be easily used for clinical trial studies across different therapeutic areas.
DM-173	Data Quality Control: Preventing Information Loss Through High Performance Binning Procedures	Schreiber-Gregory, Deanna	It is a well-known fact that the structure of real-world data is rarely complete and straight-forward. Keeping this in mind, we must also note that the quality, assumptions, and base state of the data we are working with has a very strong influence on the selection and structure of the statistical model chosen for analysis and/or data maintenance. If the structure and assumptions of the raw data are altered too much, then the integrity of the results as a whole are grossly compromised. The purpose of this paper is to provide programmers with a simple technique which will allow the aggregation of data without losing information. This technique also has the capability to check for the quality of binned categories in order to improve the performance of statistical modeling techniques. The SAS® high performance analytics procedure, HPBIN, gives us a basic idea of syntax as well as various methods, tips, and details on how to bin variables into comprehensible categories. Through this paper, you will learn about the theory behind binning, how to employ binning procedures, and how to check whether the produced categories are reliable and realistic by reviewing the WOE (Weight of Evidence), and IV (Information Value) for the binned variables. This paper is intended for any level of SAS User interested in quality control and/or SAS high performance analytics procedures.

Paper #	Title	Primary Author	Abstract
DM-185	Guide to ETL Best Practices in SAS® Data Integration Studio	Potluri, Sai	This Paper converses about the ETL best practices, tips and techniques that can be followed by developers when using SAS® Data Integration Studio for building and maintaining data warehouses/data marts. This Paper discusses the following topics. • Best ETL Design Practices. • Helpful coding insights in SAS DI studio. • Techniques and implementation using the Key transformations in SAS DI studio.
DM-188	ETL Load performance bench marking using different load transformations in SAS® Data Integration Studio.	Potluri, Sai	This paper is primarily intended to provide some helpful insights to the developers in loading large volumes of data into an Oracle Database by carrying out a performance benchmarking between different load transformations available in SAS® Data Integration Studio. The performance benchmarking for bulk loads has been performed on the following Load Transformations using different load styles available with them. • Oracle Bulk loader • Table loader The performance comparison for update operation has been performed between the following transformations • Merge • Table Loader
DM-202	From Words to Actions: Using Text Analytics to Drive Business Decisions	Baughman, Reid	Companies from a variety of industries collect free-form text data that can be used to identify new patterns and relationships. Because unstructured text data does not fit the standard row and column format, it can be more difficult to analyze and utilize. This growing array of data has the potential of yielding new insights for companies seeking to better understand customers and gain an edge relative to competitors. By transforming free-form text data into a structure that can be analyzed and visualized, analysts can use supervised and unsupervised data mining techniques to shed new light on old business problems or develop fresh insights. Call center transcripts, medical records, practitioner's notes, survey responses, or any free-form response fields can all hold valuable insights for any organization This paper will outline an example shows of how to manage and transform free-form text data, and apply advanced analytical methods to extract useful patterns, and develop actionable insights.

Paper #	Title	Primary Author	Abstract
DM-207	From FREQing Slow to FREQing Fast: Facilitating a Four-Times-Faster FREQ with Divide-and-Conquer Parallel Processing	Hughes, Troy	With great fanfare, the release of SAS® 9 delivered multithreaded processing to a single-threaded SAS world. Procedures such as SORT, SQL, and MEANS could now run faster by taking advantage more fully of system resources through parallel processing paradigms. Multithreading commonly implements divide-and-conquer methodologies in which data sets or data streams are decomposed into subsets and processed in parallel rather than in series. Multithreaded solutions are faster (but typically not more efficient) than their single-threaded counterparts because execution time (but not system resource utilization) is decreased. As the costs of memory and processing power have continued to decrease, however, there remains no excuse for not implementing multithreaded processing wherever possible. To this end, and because SAS unfortunately abandoned some hapless procedures in single-threaded Sheol, this text aims to reunite the single-threaded FREQ procedure with its multithreaded bedfellows. The FREQFAST macro is introduced and espouses divide-and-conquer parallel processing that performs a frequency analysis more than four times faster than the out-of-the-box FREQ procedure. Non-environmental factors affecting FREQ performance (e.g., number of observations, number of unique observations, file size) are elucidated and modeled to demonstrate and predict performance improvement delivered through FREQFAST.
DM-217	I spy PII: Detect, Protect and Monitor Personal Data with SAS® Data Management	Hoffritz, Cecily	The clock is ticking! Is your company ready for 25 May 2018 when the General Data Protection Regulation that has to do with data privacy laws across Europe comes into force? Failing to comply, companies incur very large fines and may lose customer trust if sensitive information is compromised. With data streaming in from multiple channels in different formats, sizes and wavering quality, it is increasingly difficult to keep track of personal data so that you can protect it. SAS® Data Management helps companies on their journey towards governance and compliance involving tasks such as detection, quality assurance and protection of personal data. This paper focuses on SAS® Federation Server and SAS® Data Management Studio from the SAS® Data Management suite to surface and manage that hard to find personal data. SAS® Federation Server provides you with a universal way to access Hadoop, Teradata, SQL Server, Oracle, SAP HANA and other data without data movement during processing. SAS® Federation Server's advanced data masking and encryption capabilities can be utilized when virtualizing data for users. Purpose-built data quality functions are used to perform identification analysis, parsing, matching and extraction of personal data in real time. We will also provide insight on how SAS® Data Management Studio's exploratory data analysis enables you to scan through your investigation hub to identify and categorize personal data.

Paper #	Title	Primary Author	Abstract
HOW-33	Improving a Graph Using PROC GPLOT® and the GOPTIONS Statement	Wright, Wendi	Starting with a SAS PLOT program, we will transfer this plot into PROC GPLOT and review the many ways you can improve the look of the plot using SAS GRAPH statements. We will make the plot really shine by customizing titles, footnotes, symbols, legends, axes, and even the reference line. At each step, a hands-on example will be presented where the user will choose their own features such as symbol colors and placement of the legend. In the end, you will have built your own personalized graph using the Title, Footnote, Symbol, Legend, and Axis statements.
HOW-43	Getting Started with SAS® Prompts	Varney, Brian	Allowing SAS® users to leverage SAS prompts when running programs is a very powerful tool. Using SAS prompts makes it easier for SAS users to submit parameter driven programs and for developers to create robust, data driven programs. This workshop will demonstrate how to create SAS prompts while in SAS Enterprise Guide and leverage them from SAS programs and SAS Stored Processes.
HOW-52	A Hands-on Introduction to SAS® DATA Step Hash Programming Techniques	Lafler, Kirk Paul	SAS software supports a DATA step programming technique known as hash that enables faster table lookup, search, merge/join, and sort operations. This hands-on workshop introduces what a hash object is, how it works, and the syntax required. Attendees learn essential programming techniques to define a simple key, sort data, search memory-resident data using a simple key, match-merge (or join) two data sets, handle and resolve collision scenarios where two distinct pieces of data have the same hash value, as well as more complex programming techniques that use a composite key to search for multiple values.
HOW-110	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.
HOW-113	SAS In The Classroom: Exploratory Data Analysis with SAS Studio	Duggins, Jonathan	With the advent of SAS Studio, and SAS University Edition in particular, there are new opportunities for exposing students at all levels to SAS. Here we focus on an activity for Exploratory Data Analysis (EDA) in a data set that is larger than those traditionally used in classroom settings. Following a brief introduction to SAS Studio it will be used to demonstrate both its interactive nature and its ability to transition users to adapting and authoring SAS code. The activity is designed to support classrooms from pre-AP statistics up through college courses by aligning with the newly adapted 2016 GAISE College Report.
HOW-168	Hands-On with an Excel-Based Code Playground for Creating and Sharing SAS ODS Graphics	conway, ted	You've heard that SAS ODS Graphics provide a powerful and detailed syntax for creating custom graphs, but for whatever reason still haven't added it to your bag of SAS tricks. Let's change that! Workshop participants will quickly gain experience creating a variety of charts by using an Excel-based code "playground" to submit SAS code examples and view the results directly from Excel. More experienced users will also find the code playground useful for compiling SAS ODS Graphics code snippets for themselves and to share with colleagues, as well as for creating Excel-hosted dashboards containing precisely sized and placed SAS graphics.
HOW-190	Hello World! - Getting Started with the SAS DS2 Language	Aanderud, Tricia	DS2 is a object-oriented programming language that is used for advanced data manipulation and data modeling applications. DS2 features extend the DATA step capabilities by adding variable scoping, user-defined methods, ANSI SQL data types and user-defined packages. In this workshop, you'll write some DS2 programs to familiarize yourself with the language. You'll have some ideas of ways to upgrade or extend your existing code when you return to the office.
HOW-227	Introduction to ODS Graphics	Kincaid, Chuck	This presentation teaches the audience how to use ODS Graphics. Now part of Base SAS®, ODS Graphics are a great way to easily create clear graphics that enable any user to tell their story well. SGPLOT and SGPANEL are two of the procedures that can be used to produce powerful graphics that used to require a lot of work. The core of the procedures is explained, as well as some of the many options available. Furthermore, we explore the ways to combine the individual statements to make more complex graphics that tell the story better. Any user of Base SAS on any platform will find great value in the SAS ODS Graphics procedures.

Paper #	Title	Primary Author	Abstract
LS-27	One Project, Two Teams: The Unblind leading the Blind	Harrington, Kristen	In the pharmaceutical world, there are instances where multiple independent programming teams exist to ensure blinded treatments are maintained by the appropriate parties. For the purposes of this discussion, blinded corresponds to fake data and unblinded corresponds to actual data. Within these projects, blinded programmers use temporary fake data to create programs which produce both blinded and unblinded results. To ensure the blind is maintained and ethics are upheld, only the blinded programming team produces and modifies the programs. While robust programming is key, another major contributing factor for success is communication. This discussion will explore the process of initializing a project supported by blinded and unblinded teams, successful communication techniques when real data comes into play, and ways to effectively troubleshoot validation issues without providing unblinding information.
LS-34	A Data-Driven Approach for Generating Define.xml v2.0 using ADaM Specifications and ADaM Datasets	Gill, Jaskaran	Case Report Tabulation Data Definition Specifications (CRT-DDS) commonly known as Define-xml is an integral part of FDA's electronic submissions. Define.xml facilitates the review process by providing information about the structure and content of standard clinical data (SDTM and ADaM) in a machine readable format. There are two versions of define.xml v1.0 and v2.0 available. In March 2016, FDA announced to end support for Version 1.0 of define.xml and mandate the use of version 2.0 for all the electronic submissions. This paper presents a data-driven approach for generating define.xml v2.0 using ADaM specifications and ADaM datasets. A Define ready ADaM specification plus ADaM datasets are passed into a macro that automatically generates define.xml v2.0. This macro also provides an error report for any inconsistencies identified between ADaM datasets and ADaM specifications. This automation process considerably reduces the time and effort in generating a CDISC compliant define.xml for FDA submissions.
LS-78	De-Identifying Data When Sharing Clinical Data	Saranadasa, Pushpa	In a particular clinical study a great deal of information is collected that may be of use in other research investigations. The sharing of knowledge may include the patient-level detailed data used in clinical studies. For many reasons, not the least of which is patient privacy, any shared data must first be de-identified. The process of de-identification and creating substitute identifiers involves masking the original identifiers in a way that preserves the utility of the data while at the same time minimizing the risk of re-identification of the patients and any of their unique data that might have been intrinsic to the study data base. This paper will explicitly discuss: 1) what is meant by de-identification, 2) how robust is the de-identification process and 3) the challenge of ensuring study subject anonymity while preserving data base utility.
LS-79	Automated Validation of Complex Clinical Trials Made Easy	Watson, Richann	Validation of analysis datasets and statistical outputs (tables, listings, and figures) for clinical trials is frequently performed by double programming. Part of the validation process involves comparing the results of the two programming efforts. COMPARE procedure output must be carefully reviewed for various problems, some of which can be fairly subtle. In addition, the program logs must be scanned for various errors, warnings, notes, and other information that might render the results suspect. All of this must be performed repeatedly each time the data is refreshed or a specification is changed. In this paper, we describe a complete, end-to-end, automated approach to the entire process that can improve both efficiency and effectiveness.

Paper #	Title	Primary Author	Abstract
LS-105	Streamlining ADaM Dataset Specifications	Vega, Hunter	In a large group of statistical programmers, you often find that people have a varied style for how they might create an analysis dataset specification. It could either be too detailed so that it compromises the independent programming concept or lacking in key information, e.g. how flags are to be derived. In an effort to consolidate some core practices for analysis dataset programming, we put together a training for our statistical programming and biostatistics teams to go over basic ADaM dataset concepts and the process to be followed at CROS NT. The purpose of the training was to create guidelines for how to create an analysis dataset specification, how much information one should be expected to include and when the biostatisticians should review. This collaboration of our biostatistics and statistical programming groups has improved the quality of our analysis dataset specifications as well as improved the efficiency of our programming of these datasets.
LS-157	Building an Out-of-Pocket Cost (OOPC) Calculator for Medical Expenses using the Medical Expenditure Panel Survey (MEPS)	Merriman, Barry	Building an Out-of-Pocket Cost (OOPC) Calculator for Medical Expenses using the Medical Expenditure Panel Survey (MEPS) Background: It can be difficult for consumers buying health insurance on the health insurance marketplace to compare plans. Aside from different premiums plans can have different cost-sharing characteristics: copayment, coinsurance, deductible, or maximum out-of-pocket payment. In addition, copayments and coinsurance can vary across different types of services and even drugs. Having a way to estimate out-of-pocket costs (OOPC) would help consumers compare plans and make it easier to choose a suitable plan. There is an OOPC calculator on healthcare.gov, the federally-facilitated marketplace. However, it is based on data from a commercial claims database and is expensive. I sought to find a suitable data source that is publicly available and is free. Any resulting calculator could be used in any marketplace including state-based. Methods: In order to estimate OOP costs you first need to estimate utilization and associated costs. Then you apply the plan cost-sharing characteristics to each of the healthcare events to estimate patient responsibility or the OOPC. I used MEPS utilization data from 2013 for the under 65 population to create healthcare domains, for example, inpatient facility charges, primary care visits, or outpatient surgery. Patients were grouped by age and sex and further divided into low, medium, and high groups according to the cost of their healthcare. Plan cost-sharing characteristics are applied to the costs associated with typical utilization within each domain for each age/sex/low/medium/high group to give an OOPC estimate for each group. Results: If the tool is used it will be in the background of a web interface where customers shop for health insurance. Also in the background will be a dataset with plan cost-sharing characteristics. The customer will input age/sex for all family members to be insured. The estimated OOPC is calculated for any plan the customer selects. The customer can combine the OOPC with plan premiums to compare total costs across plans. Conclusion: The OOPC calculator based on MEPS data yields comparable results to a calculator based on commercial claims and is therefore a viable alternative that is free.
LS-174	Timing Variables In Clinical Trials: Avoiding Common Mistakes And Dealing With Unforeseen Issues	Weller, Gregory	Timing variables are an often overlooked part of clinical trials data collection and mapping, particularly in the SDTM and ADaM standards. However, failure to properly plan timing variable collection and mapping before study data is collected can have major consequences later on. This paper will review good practices for collecting and mapping timing information in clinical trials, from clinical database setup through SDTM and ADaM. Examples from real studies will be used to highlight the challenges that can happen when either: a) insufficient timing information is collected, or b) unforeseen issues arise in the collection of timing variables.

Paper #	Title	Primary Author	Abstract
LS-200	Discovering CMS Data for Healthcare Research	Ganapathi, Laxminarayana	Center for Medicare and Medicaid (CMS) hosts a plethora of health care data and at the outset aims to disseminate the benefits of this vast information to the public, without compromising the individual's privacy. It is estimated that over 98% of adults age 65 and over are enrolled in Medicare and over 99% of the deaths in this group are accounted for in the Medicare program, thus making it a vital repository of data for healthcare research about the senior population of about 50 million. While the size of the data could be intimidating, researchers have taken up the challenge of working with this ocean of data to evaluate, model, hypothesize and help CMS in improving its services. It is welcome news for SAS users that most of the data is available as SAS data sets. This work aims to introduce the newcomers to the structure of the CMS databases and point to the need and advantage of a deeper understanding of the data for the success of a study.
LS-213	Enrollment Simulation in Clinical Trials	DePuy, Venita	Enrollment predictions for clinical trials are often as simple as estimating A subjects per site per month x B sites x C months = D target enrollment. This prediction does not model the uncertainty of the estimated enrollment duration. The commonly used method provides an exact number, but in reality, while we may average a certain number of subjects per month, we are likely to see some variability in enrollment. In addition, this method does not incorporate possible enrollment delays, such as delaying enrollment between each subject due to dose toxicity concerns. We propose a simulation method to allow for the creation of estimated enrollment durations with associated confidence intervals, while also allowing for optional delays in recruitment. This method simulates exponentially distributed times to subject enrollment. In addition, this approach also allows us to take into consideration variables such as start-up times or enrollment holds due to data monitoring committee meetings.

Paper #	Title	Primary Author	Abstract
PSA-59	You as a Brand: Tips for Managing Your Data Career	Aanderud, Tricia	Yes! You are a brand. The branding you advertise to the target market (your manager and colleagues) is very similar and just as important as the branding that goes into a product or company. All brands require management to be successful. During this talk, we will review how you establish your brand, using social media to promote your brand, and simple ways you can maintain your brand. You will walk away with some strategies you can implement immediately.
PSA-73	Document and Enhance Your SAS® Code, Data Sets, and Catalogs with SAS Functions, Macros, and SAS Metadata	Hadden, Louise	Discover how to document your SAS® programs, data sets, and catalogs with a few lines of code that include SAS functions, macro code, and SAS metadata. Do you start every project with the best of intentions to document all of your work, and then fall short of that aspiration when deadlines loom? Learn how SAS system macro variables can provide valuable information embedded in your programs, logs, lists, catalogs, data sets and ODS output; how your programs can automatically update a processing log; and how to generate two different types of codebooks.
PSA-76	Can A SAS Programmer Stay Cool	Hu, Jiangtang	The traffic of SAS mailing list is keeping decreasing. There are less active SAS bloggers, including myself. It seems everyone is talking about the coolness of R, Python, machine learning or deep learning. The other side is, SAS as a software and a programming language is getting better. As a long time and continuous SAS heavy user, I'm asking: can a SAS programmer stay cool? Furthermore, can a SAS programmer like me cool? It's not a silly question. Lots of SAS programmers feel the crunch. In this paper, I will talk about the challenges a SAS programmer can face. The challenges can come from external competitors like R and Python, technology trend like machine learning and deep learning. I will talk about in which SAS is good at, while in which not. In the end, I will explore some how-to for a SAS programmer to stay alert, marketable, and even more, cool. I will also encourage SAS programmers to keep update with SAS software; Lua, Python, machine learning and other cool stuff are already imbedded in SAS system! All codes and supporting materials will be available in my Github page: https://github.com/Jiangtang/SESUG/tree/master/2017

Paper #	Title	Primary Author	Abstract
PSA-86	Getting on SAS Grid, Our Journey	Wilson, David	RTI International made the move from PC and server-based SAS to SAS Grid this past year. The decision to move to the Grid was based on cost over time, scalability - allowing room for growth, centralization of SAS administration, improved performance, and a highly available environment. This paper will focus on how the overall transition occurred from the decision to move to SAS Grid to getting 400+ SAS users successfully transitioned within a 12-month period on both the Federal Information Processing Standard (FIPS) Low (limited adverse effect) and FIPS Moderate (serious adverse effect) environments. The process included getting sponsorship from the business units and IT, identifying an implementation partner, forming stake-holder teams, gathering requirements and designing the system, establish a plan and timeline for communication, implementation, training, and roll-out. Along the way, we addressed challenges in our environment: which platform to use, moving from Windows to Linux, FIPS mod requirements that impacted the architecture of the system, user response to the transition, regional users and data housed at regional offices, as well as legacy code that required modification due to the move to SAS Grid. Some unexpected positive items to come out of the transition included more collaboration between SAS programmers across the business units which continue after the transition was completed and overall knowledge of what SAS programmers throughout the company utilize in their project specific code. As we transition from implementation to operational mode as a company that utilizes SAS Grid, our previous project technical advisory team has become a Grid advisory committee, a cross-functional team that continues to emphasize the business needs of the analytical community as well as improve our relationship with SAS and leverage the latest features SAS offers. We hope sharing our experience and lesson-learned will be valuable to others considering the transition to SAS Grid.
PSA-88	Wait, I don't want to be the Linux Administrator for SAS Visual Analytics!	Boase, Jonathan	Whether you are a new SAS administrator or switching to Linux an environment, you have a complex mission. This job becomes even more formidable when you are working with a system like SAS Visual Analytics that requires multiple users loading data daily. Eventually a user will have data issues or create a disruption that causes the system to malfunction. When that happens, what do you do next? In this paper, we will go through the basics of a SAS Visual Analytics Linux environment and how to troubleshoot the system when issues arise.
PSA-100	Job Upward Mobility: Getting Better When You're Already Good	Hall, Molly	You're a SAS® Programmer, or a Statistical Analyst, or a Data Scientist, or a Manager of SAS Programmers. You make a healthy six figures, or close to it. Perhaps you drive a Lexus, or other luxury car. You're reasonably happy in your career. Still, sometimes you can't help feeling a little stuck. You're curious about exploring opportunities, but you want to do it inconspicuously, and you don't want to jeopardize your current position. We will talk about the SAS job market. We will explore networking channels that work. Can using Social Media boost your career? Is it appropriate? Let's discuss. Some people say that resumes are dead – is this true? What about the cover letter – is it necessary? We will also examine interview tips and techniques. Even if you are happy at your current company, or in your current job, this presentation is for everyone, and it is appropriate for people at all levels. It's time to stop procrastinating and, perhaps, level up.
PSA-126	Parallelizing Windows Operating System Services Job Flows	Kratz, David	SAS® Job flows created by Windows operating system services have a problem: At present they can only execute jobs in series (one at a time). This can slow job processing down, and limits the utility of these flows. This paper shows how one can alter Windows operating system services flows after they have been generated to enable jobs to run in parallel (side by side). A high level overview of a SAS PROC GROOVY script which automates these changes is provided, as well as a summary of the positives and negatives of running jobs in parallel.

Paper #	Title	Primary Author	Abstract
PSA-147	Super Happy Fun Times: Diagnosing and Resolving an Intermittent Failure with the SAS 9.4 Grid Workspace Server	Hayes, Rebecca	As any SAS user or Administrator knows – intermittent problems are often the most difficult to diagnose. This paper will explore a real world scenario in which the client experienced intermittent failures when attempting to assign a logical application server in Enterprise Guide or when attempting to validate the Workspace Server in SAS Management Console. These attempts would succeed approximately 1 out of 3 times. The issue stemmed from changes in how the Object Spawner functions in SAS 9.4 versus previous releases of SAS and how this can manifest itself in a SAS Grid environment. In this paper we will discuss and review those changes and the temporary and permanent solutions.
PSA-151	Understanding the influence of the day of week in the reviews written	Alugubelli, Sujal Reddy	In recent times, manufacturers give importance to reviews written by users on social networking and e-commerce websites. Understanding customers' needs is a critical aspect of every business which helps in satisfying their customers. With the huge number of reviews available it is an arduous job to narrow down and find significant comments or understand what external factors may influence those comments. One external factor that could be a potential influence on the customer reviews is the effect of day of the week on the way a user writes a review. This factor is analyzed and examined in this paper. According to published research, human beings react differently in different times. They tend to overreact when they are under pressures and are more accepting when they are relaxed. As most individuals are stressed on weekdays because of the workload, they are likely more stressed during weekdays and are calmer on weekends. To investigate this, in this paper analysis is done separately for weekday and weekend reviews. This may help a company to understand the attenuating effect of an external factor on the nature of the review. Reviews written on Amazon.com for two products from electronics product category are explored in this paper. Each product has more than 4000 reviews. SAS Enterprise Miner and SAS Sentiment Analysis Studio is used to analyze reviews written on weekdays and weekends.

Paper #	Title	Primary Author	Abstract
PSA-209	Pinching Off Your SAS® Log: Adapting from Loquacious to Laconic Logs To Facilitate Near-Real Time Log Parsing, Performance Analysis, and Dynamic, Data-Driven Design and Optimization	Hughes, Troy	Too often in SAS® literature, the role of the program log is narrowly conceptualized as a static, post hoc quality control review that validates program success through the detection of program failure or the lack thereof. Especially when software development occurs outside of a formalized software development life cycle (SDLC), as is often the case with non-production software and within end-user development environments, SAS practitioners must painfully parse logs in search of notes, warnings, runtime errors, and other often elusive indications of functional failure. A substantial body of SAS literature advances antediluvian manual log review through the automation of log parsing and analysis and subsequent communication of program success (or failure) to stakeholders. To a lesser extent, SAS literature depicts how log parsing can be utilized to extract, analyze, and ultimately improve software performance metrics. After a cursory review of SAS automated log parsing literature, this text elucidates and expands this second objective of automated log parsing, demonstrating how performance metrics can be analyzed in near-real time to drive program flow dynamically. By pinching off shorter logs and saving these as temporary text files, SAS programs can analyze performance metrics for individual procedures or processes, enabling software to detect anomalous or undesirable CPU, input/output (I/O), or memory consumption and to respond dynamically to optimize execution.
PSA-219	Tips for Effective SAS Platform and User Administration	Sadof, Michael	Effective administration and support of a SAS Enterprise BI environment is difficult due to the many and varied type of users and SAS products. Presented here will be some tips and techniques of supporting your environment and users. Some topics to be discussed are how to organize the folder structure on the platform, monitoring the environment, training users, and providing them with support, standards, and resources for self-help. To keep things running smoothly certainly involves tuning of the environment but it is essential to optimize user programs as well. Enclosed within are some ideas to gently turn your users into SAS rock stars. When the SAS environment is running efficiently and programs are delivering timely and accurate results to the users you don't have to worry and you can be happy.

Paper #	Title	Primary Author	Abstract
RIV-18	How to Data Science: Visualization of Spatial Data - Beyond the Standard Proc Gmap	Hoffman, Seth	The bulk of a data scientist's job is data preparation and cleaning. However, presenting the findings in a compelling manner is what generates speaking engagements. "A Map!" is the obvious answer for visualizing location data. Though the definition is quite broad, most people picture a general reference map, or maybe a heat map if it is the middle of summer. This paper uses Proc Gmap as the drawing tool to explore some less traditional metaphors and ideas for displaying analysis results which contain a spatial component.
RIV-19	Applying JMP®'s Imaging Analytic Tools to Target Cancer Tumors that Guides External Beam Radiation Therapy	Alexander, Melvin	The motivation of this paper came from questions I received on my SESUG 2016 presentation. There, I showed how John Ponte's JMP Image Column Analyzer add-in could select frames of MRI prostate cancer images (saved in an Expressions column). The selected frames showed special markers near the prostate. These markers identified the location where to aim radiation beams that will kill the cancer cells without harming healthy cells. Selected frames with the markers were output to JMP data tables representing the image data. I appended the marker-frame data tables into a larger JMP data table. In this presentation, I compute kernel density estimates (non-parametric measures) of the marker's intensity values from the Bivariate Analysis > Nonpar Density platform. Contour lines with the largest density of marker's intensity values provide useful metrics for further data analysis and exploration.
RIV-22	Parallel Coordinates Plot Made Easy	Rosanbalm, Shane	A parallel coordinates plot is useful for visualizing multivariate data. Unfortunately, there isn't a PARALLEL statement in SGPLOT. In this paper we present a macro called %parallel. Using a minimum of parameters (data=, var=, group=) the macro will produce a parallel coordinates plot via SGPLOT.
RIV-31	Using PROC REPORT® and ODS STYLE Options to Make Really Great Tables	Wright, Wendi	Questar Assessment, Inc. had a need to automate many tables in a large public report created every year. This paper describes the techniques used to automate the tables in SAS. The tables can now be created with little or no modification before being posted, making the creation of the report much more efficient. This intermediate-level paper shows examples of how to use the ODS RTF STYLE= option in PROC REPORT to customize the borders in table headers and table cells. It will also cover the many other ways the STYLE option can be used to set widths, heights, and justification and to change fonts and colors in the table's foreground and background. Also covered will be how to add subscript, superscript, and carriage returns to cell contents.
RIV-39	Student Development and Enrollment Services Dashboard at UCF	Piemonti, Carlos	At the University of Central Florida (UCF), Student Development and Enrollment Services (SDES) combined efforts with Institutional Knowledge Management (IKM), which is the official source of data at UCF, to venture in a partnership to bring to life an electronic version of the SDES Dashboard at UCF. Previously, SDES invested over two months in a manual process to create a booklet with graphs and data that was not vetted by IKM; upon review, IKM detected many data errors plus inconsistencies in the figures that had been manually collected by multiple staff members over the years. The objective was to redesign this booklet using SAS® Web Report Studio. The result was a collection of five major reports. IKM reports use SAS® Business Intelligence (BI) tools to surface the official UCF data, which is provided to the State of Florida. Now it takes less than an hour to refresh these reports for the next academic year cycle. Challenges in the design, implementation, usage, and performance are presented.

Paper #	Title	Primary Author	Abstract
RIV-42	Methods for Creating Sparklines using SAS®	Andrews, Rick	A sparkline is a very small line chart, typically drawn without axes or coordinates, to represent the general shape of the variation (typically over time) of some measurement, such as temperature or stock market price, in a simple and highly condensed way (Spark, 2017). This paper will discuss how SAS® can be used to mimic sparklines created in Microsoft® Excel. A number of methods exist for generating sparklines in SAS, two of which will be discussed. The first is a series of images that can be added to the REPORT procedure employing an SGPLOT and the second utilizes a highly customized process using the annotate facility available with SAS.
RIV-58	SAS® Visual Analytics Tricks We Learned from Reading Hundreds of SAS® Community Posts	Aanderud, Tricia	After you know the basics of SAS Visual Analytics, you realize there are some situations that require unique strategies. Sometimes tables are not structured right or become too large for the environment. Maybe creating the right custom calculation for a dashboard can be confusing. Geospatial data is hard to work with if you haven't ever used it before. We looked through 100s of SAS Communities posts for the most common questions. These solutions (and a few extras) were extracted from the newly released Introduction to SAS Visual Analytics book.
RIV-66	Visualization of Patient Electronic Records to Support Exploratory Analysis and Variable Derivation of Categorical Data	Thomas, Steven	Background Electronic medical records (EMR) have become a standard data source for epidemiological, outcomes, and health services research. However, there are challenges caused by the size and complexity of EMR data. Data are collected continuously across multiple systems and are stored in a variety of structures. Systems and structures can include free text, long or wide forms, and complex temporal information. These complexities make EMR data similar to an evolving ecosystem rather than a static source found in most studies. In a natural ecosystem, data sources are assessed to ensure that information is consistent with expectations. Institutions should approach EMR data in a similar manner to provide insight and build confidence among team members with diverse backgrounds. New tools and processes need to be developed that support assessment of analytic decisions and are available to all members of the team. Methods This paper proposes three visual tools to utilize in exploratory analyses before variable derivation. These tools are designed to promote discussion and build consensus between team members using EMR data. They allow examination of individual patient records and trends across time so common operational considerations (e.g., defining variables via multiple features, selection of time-windows) are addressed using both the data and therapeutic expertise. This paper will present SAS graphic language templates for patient profiles, cumulative heat maps, and Sankey diagrams with example discussions and decisions that each visual is designed to support. Results/Conclusions Studies seeking to maximize use of EMR data involve multiple stakeholders that need to understand nuances in the data. Visualizations can facilitate team discussions and improve the process of feature extraction, variable construction, and project planning. These visuals can be particularly useful for sequential analysis, treatment patterns, and defining episodes of care, but all studies using EMR can benefit from the use of visuals.

Paper #	Title	Primary Author	Abstract
RIV-92	SAS/GRAPH® and GfK Maps: a Subject Matter Expert Winning Combination	Hadden, Louise	SAS® has an amazing arsenal of tools to use and display geographic information that is relatively unknown and underutilized. High quality GfK Geocoding maps have been provided by SAS since SAS 9.3 M2, as sources for inexpensive map data dried up. SAS has been including both GfK and "traditional" SAS map data sets with licenses for SAS/GRAPH for some time, recognizing there will need to be an extended transitional period. However, for those of us who have been putting off converting our SAS/GRAPH mapping programs to use the new GfK maps, the time has come, as the "traditional" SAS map data sets are no longer being updated. If you visit SAS MapsOnline, you will find only GfK maps in current maps. The GfK maps are updated once a year. This presentation will walk through the conversion of a long-standing SAS program to produce multiple US maps for a data compendium to take advantage of GfK maps. Products used are Base SAS® and SAS/GRAPH®. SAS programmers of any skill level will benefit from this presentation.
RIV-97	Exploring JMP® Modeling Functionality Using Consumer Expenditure Data	Klick, Josh	The Consumer Expenditure (CE) Surveys collect information on demographic, housing, income, and expenditure characteristics of households in the US. The purpose of this presentation is to explore JMP® modeling techniques that explain consumer unit (CU) expenditures of the CE Public Use Micro Data from the CE Interview survey. An initial linear model is constructed based on an exploratory analysis of the data. The second and third models utilize stepwise regression based on the Bayesian Information Criterion and the Max K-Fold RSquare. The fourth model implements Bayesian model averaging via an interface to the R Bayesian Adaptive Sampling (BAS) package. Last, a fifth model partitions data based on the relationship between predictors and response variables. This paper concludes with a summary of model performance.
RIV-107	Clinical Data Visualization using TIBCO Spotfire® and SAS®	Gupta, Ajay	In Pharmaceuticals/CRO industries, you may receive requests from stakeholders for real-time access to clinical data to explore the data interactively and to gain a deeper understanding. TIBCO Spotfire 7.6 is an analytics and business intelligence platform, which enables data visualization in an interactive mode. Users can further integrate TIBCO® Spotfire with SAS® (used for data programming) and create visualizations with powerful functionality e.g. data filters, data flags. These visualizations can help the user to self-review the data in multiple ways and will save a significant amount of time. This paper will demonstrate some basic visualizations created using TIBCO Spotfire and SAS using raw and SDTM datasets. This paper will also discuss the possibility of creating quick visualizations to review third party vendor (TPV) data in formats like EXCEL® and Comma Separated File (CSV).
RIV-114	Mapping Roanoke Island Revisited: An OpenStreetMap (OSM) Solution	Okerson, Barbara	In a previous presentation, SAS® was used to illustrate the difficulty and solutions for mapping small pieces of coastal land that are often removed from map boundary files to smooth boundaries. Roanoke Island, one of the first areas of the current United States to be mapped (1585), was used as an example since it is smoothed out of many current maps. While these examples isolated Roanoke Island, they didn't provide detail beyond city names on the map. Originally limited to SAS Visual Analytics, SAS now makes available background maps with street and other detail information for SAS/GRAPH using open source map data from OpenStreetMap (OSM). This presentation will review the previous solutions and then look at how to map Roanoke Island using SAS/GRAPH and OSM.

Paper #	Title	Primary Author	Abstract
RIV-128	Building a Nested Range Bar Chart with SAS® Graph Template Language	Pan, Fan	A bar chart is a graph with rectangular bars with lengths that are proportional to the values that they represent, which is probably one of the most useful of all graphs. A bar chart is an alternative to line plots to display the mean or midpoint response when the responses are categorized by discrete values. Bar charts are usually used to compare a response value among specific categories, and widely employed in many fields, including clinical research, agriculture, and education. There are many situations where a simple bar chart does not tell the whole story, or can even be visually misleading. In this way, stacked or subgroup bar charts were showing the distribution of different subgroups among the cluster groups is an excellent visualization tool. This paper demonstrates how to create different kinds of stacked bar charts including: vertically and horizontally stacked bar charts, 100% Stacked Bar Chart and 3D stacked bar charts, and how to control their appearance and layout. Stacked bar chart can be created in GCHART using the VBAR, SUBGROUP, and GROUP options. SGPLOT also quickly produce the bar charts with one or two classification groups, whereas three levels of categorization can be done easily using the SGPANEL procedure. Besides making sure that the graph is not misleading, the paper also demonstrates ways to improve the good looking and layout of these stacked bar charts. For example, how to specify properties of graphs, such as format, color, name, height, axis and width, how to send bar graphs to different ODS destinations and how to apply ODS styles to graphs. Other features such as midpoints and statistic information, text label, inside and outside test annotations, and legend will also be combined to generate an advanced stacked bar chart. Based on the information, a researcher who is not quite familiar with SAS programming could be still able to create stacked bar charts with the appropriate layout.
RIV-138	Building Heat Maps for Data Cleaning and Beyond	Much, Kristen	A heat map is a graphical representation of data that translates data values into colors within a matrix. This type of data visualization summarizes a vast amount of data within a single snapshot which helps to quickly communicate relationships between data values. Using SAS, a heat map can be generated easily with the HEATMAPPARM statement in either the SGPLOT/SGPANEL procedures or Graph Template Language (GTL). The focus of this paper is to explore the syntax of the HEATMAPPARM statement within both SGPLOT/SGPANEL and GTL as well as to walk through a direct application of a heat map to be used in preparation for database lock. The example demonstrates how to develop a heat map that indicates the status of all case report forms (CRFs) within a database across all visits for all subjects. When used effectively, this tool can greatly enhance understanding within a clinical team regarding the status of requirements for a successful database lock.
RIV-156	An Obvious Yet Helpful Guide to Developing Recurring Reports in SAS®	Straney, Rachel	Analysts, in particular SAS programmers, are often tasked with compiling data and generating reports at a moment's notice. It is paramount to make report generation efficient for ease of replication. The method of preparing and compiling recurring reports can be made easier when the process of reporting is well-defined. The objective for this paper is to outline a checklist that can guide the creation of a reporting process for your data. SAS capabilities like labels and formats as well as the SAS Macro Facility play an important role in automating a reporting process. Identifying SAS procedures ahead of time, like PROC PRINT or PROC TABULATE, and developing a program that is flexible and adaptable can save time. General steps to successfully develop a reporting process will be described and detailed examples to show how to achieve this will be provided. The overall benefits of planning a reporting process are to reduce reporting errors, reduce programming time and have a well-designed report that is meaningful.

Paper #	Title	Primary Author	Abstract
RIV-169	Removing the Mask of Average Treatment Effects in Chronic Lyme Disease Research Using Big Data and Sub-Group Analysis	Shapiro, Mira	There is considerable controversy in the medical community surrounding chronic Lyme disease (CLD). CLD sufferers have persistent Lyme symptoms resulting from being untreated, under-treated or, having a lack of response to their antibiotic treatment protocol. Some of the past Lyme disease studies that reported on average treatment effects were unable to identify treatment successes. Using patient- reported outcome data collected by LymeDisease.org via their online registry, MyLymeData, we will show that using sub-group analysis techniques can unmask valuable information about treatment effectiveness.
RIV-192	Mapping Participants to the Closest Medical Center	Franklin, David	"How far are patients from Clinics?" That was the question which was asked on a wintry afternoon. Recorded in the database was the ZIP codes of each participant, but how could we use that data to find the closest distance from the nearest clinic? "And would it not be nice if we could map it?" This paper looks at calculating the distance from a participant ZIP code to a Medical Center using the GEODIST function, finding the smallest distance for each participant to attend, and finally color-coding the plots depending on the distance calculated using the GMAP procedure. Also helpful is producing reports, one of which will be the those participants who are farthest away and the distance.

Paper #	Title	Primary Author	Abstract
SD-17	PMML & SAS: An Introduction to PMML and PROC PSCORE	Gannon, Andrew	The purpose of this paper is an introduction to the Predictive Model Markup Language (PMML). It goes through a general structure of PMML and an example of a general regression model. The second half of the paper is an introduction to the PROC PSCORE procedure in SAS. It goes through assigning the PMML and the output of the procedure. It ends with an example of using the PMML and PROC PSCORE to generate a simple regression with output. This paper is accompanied by several appendixes with the code used throughout and the SAS code that is created during the PSCORE procedure.
SD-29	Using Categorical Variables in Regression Analysis	Bilenas, Jonas	In this tutorial we will review how to deal with categorical variables in regression models. Techniques will show how categorical variables can test for changes in intercept or slope parameters in regression models. We will review how to use the CLASS statement in many regression PROCS and the default settings for various PROCS. We will also review how to select the reference level for more than 2 categories.
SD-61	Decomposing the R-squared of a Regression Using the Shapley Value in SAS®	Coleman, Charles	After creating a linear regression model, quantifying the importance of each variable is often desired, either for academic purposes or to design the most effective interventions. This normally decomposing the variance, or, equivalently, R^2. Using the Shapley Value, known in this literature as the LMV, it is possible to do such a decomposition. This decomposition can be applied to models with interactions and is robust against multicollinearity. It has the interpretation that it is the marginal contribution of each variable to R^2. This is implemented in macro %DECOMPOSE_R2 using PROCs GLM, IML, macro string functions and data step-like macro processing.
SD-62	Tips and Tricks for Raking Survey Data with Advanced Weight Trimming	Battaglia, Michael	Raking to population control totals is often the final step in developing survey weights. Raking is an iterative procedure that brings the weighted sample into agreement on socio-demographic variables that are available for the sample and the population. It is primarily used to reduce unit nonresponse bias. Raking can lead to some observations ending up with extreme weights, in other words, weights that are very large or very small compared to the mean weight, resulting in inflated standard errors. In 2009, we developed a SAS raking macro that implements weight trimming during the raking iterations, ensuring that the weighted sample will agree with the population. We recently enhanced the macro to add several options related to weight trimming. There are four trimming values that can be specified: two "high end" trimming values – the individual high cap value and the global high cap value; and two "low end" trimming values -- the individual low cap value and the global low cap value. The macro also includes two trimming options. In the "OR" method an observation has its weight reduced if the value of the weight is greater than the individual high cap value or the global high cap value. An observation has its weight increased if the value of the weight is less than the individual low cap value or the global low cap value. In the "AND" method an observation has its weight reduced if the value of the weight is greater than the individual high cap value and the global high cap value. An observation has its weight increased if the value of the weight is less than the individual low cap value and the global low cap value. Using survey data, we illustrate how the new weight trimming features are used: whether to trim weight or not, by which method to trim, which end to trim, and how much to trim. We also demonstrate how to set different convergence criteria for a subset of the raking variables when raking convergence problems arise. These enhancements have proven to be of interest to the survey research community.

Paper #	Title	Primary Author	Abstract
SD-69	Binning of Predictors for the Cumulative Logit Model	Lund, Bruce	Binning of a predictor is widely used when preparing a predictor for use in binary logistic regression. In this paper the practice of binning is extended to the cumulative logit model. Here the target has more than 2 levels. A binning algorithm and SAS® code is presented. The user may choose to bin a predictor X so that the ordering of X is maintained. In this case, for each k for 2 ⬚ k ⬚ L (where L= number of levels of X), the algorithm can find the k-bin solution which is optimal with respect to log-likelihood or to a generalized information value (IV). The optimal solution is found by complete enumeration of all solutions. Additionally, the algorithm can find if there is an optimal k-bin solution which is monotonic versus the empirical cumulative logits of the target. Alternatively, the user may choose to bin X without regard to the ordering of X. In this case the algorithm maximizes log-likelihood or generalized IV at each step in the binning process but an optimal k-bin solution is not guaranteed. A generalization of weight of evidence (WOE) is defined and SAS code for WOE coding is generated. Examples are presented.
SD-71	MOPP the SOP - Makeover of Programming Process Steps in the Standard Operating Procedure	Harkins, Kathy	Our goal was to redesign our standard operating procedures and the corresponding tools and templates for defining, developing and validating programs to support analysis and reporting (A&R) performed by Biostatistics and Research Decision Sciences (BARDS). Current processes were evaluated and comprehensive documentation, templates and training materials were developed for each of the system development lifecycle stages – Define, Develop, Validate and Operations. This paper will provide insights into the planning, implementation and execution of a major procedure rewrite in order to streamline processes and align with industry best practices.
SD-106	Missing Data and Complex Sample Surveys Using SAS®: The Impact of Listwise Deletion vs. Multiple Imputation Methods on Point and Interval Estimates when Data are MCAR, MAR, and MNAR	Kellermann, Anh	Social scientists from many fields use secondary data analysis of complex sample surveys to answer research questions and test hypotheses. Despite great care taken to obtain the data needed, missing data are frequently found in such samples. Even though missing data is a ubiquitous problem, the methodological literature has provided little guidance to inform the appropriate treatment for such missingness. This Monte Carlo study used SAS to investigate the impact of missing data treatment (hot deck-based multiple imputations versus regression-based multiple imputation versus listwise deletion) when data are MCAR, MAR, and MNAR. By using 10% to 70% of missing data (along with complete sample conditions as a reference point for interpretation of results), the research focused on the parameter estimates in multiple regression analysis in complex sample data. Results are presented in terms of statistical bias in the parameter estimates and both confidence interval coverage and width.
SD-120	Tornado Inflicted Damages Pattern	Sharma, Vasudev	On average, about a thousand tornadoes hit the United States every year. Three out of every four tornadoes in the world occur in the United States. They damage life and property in their path and they often hit with very little, sometimes no, warning. Tornadoes cause approximately 70 fatalities and 1,500 injuries in US every year. Once a tornado destroyed an entire motel in Oklahoma and the motel's sign was later recovered in Arkansas. Tornadoes most frequently hit "Tornado Alley" which is mainly made up of Nebraska, South Dakota, Oklahoma, Texas, and Kansas. A tornado extends from a thunderstorm to the ground and appears as a funnel shaped cloud rotating with winds, which can reach 300 miles per hour and can exceed a one-mile radius. Tornadoes can travel very long distances making them very devastating. Since the ability to detect the intensity and direction of tornadoes prior to formation is limited, predicting the likelihood a tornado will form with accuracy can save many lives, as well as property. The purpose of the study is to find a pattern in the fatalities, injuries and property loss. The tools used are Base SAS, SAS Enterprise Miner, R, and Tableau. The results include statistical analysis, descriptive analysis, predictive analysis and visualizations from these tools.

Paper #	Title	Primary Author	Abstract
SD-124	Two Sample Parametric and Nonparametric Statistical Tests for a Large Number of Comparisons Made Easy	Beal, Dennis	Parametric and nonparametric two sample statistical tests have been in SAS® for a long time. However, for environmental data there often are many constituents to compare their mean or median concentrations to see if they are statistically the same or not. Typically there can be hundreds of chemical or radionuclide constituents that need to be statistically tested between two sampling locations. This paper provides SAS code that efficiently calculates the summary statistics and output from two sample statistical tests by constituent including p-values and a final determination whether they are significantly different at a given significance level. Parametric t-tests are performed when the data are normally or lognormally distributed in both samples. The nonparametric two sample Wilcoxon Rank Sum test is performed when the data are neither normally nor lognormally distributed. ODS output is utilized to export p-values from statistical tests into a data set so conclusions can be made easily whether two sample locations have significantly different levels of constituents. This paper is for intermediate SAS users of SAS/BASE.
SD-137	Applying Logistics Regression to Forecast Annual Organizational Retirements	Dunham, Alan	This paper briefly discusses the labor economics reasoning underlying employee decisions affecting retirements from corporations and government organizations, and then shows how that reasoning can be used for practical forecasting by applying logistics regression. A test dataset is generated and stochastically perturbed for the forecasting exercise. The SAS® LOGISTICS procedure is used to forecast the probabilities of annual individual retirements. Given the binomial nature of the retirement decision, the probabilities can be summed to obtain annual total retirement forecasts. Theoretical underpinnings, test data, code, and results for the test dataset are provided.
SD-150	Analysis of Unstructured Data: Topic Mining & Predictive Modeling using Text	Allaparthi, Ravi Teja	Eighty percent of the data generated in digital space is unstructured. While the amount of textual data is increasing rapidly, the ability to summarize and make sense of such data for making better business decisions remains challenging. This paper provides insights into how to analyze textual survey data for extracting public opinion from a huge collection of feedback forms and also to come up with rules in segregating the outcome. The dataset to be analyzed is collected based on the Toronto Casino Feedback Form which contains 17,000 records with information about open ended questions such as why do you not support the establishment of a casino as well as closed ended questions such as age group and gender. The primary objective is to understand and predict the opinion of a user towards the establishment of a casino by considering the survey filled by the user using unstructured data. Topics are extracted using the Text Topic node in SAS Enterprise Miner in identifying the public opinion. Text Rule Builder node is used to build rules that can differentiate the outcome. From the analysis, we are able to identify that the majority of the public worried about gambling leading to addiction, increase in crime rate, traffic congestion and relationships. People also opined positively about new jobs being created, tourism, additional revenue generated through taxes and tourism and increase in entertainment options. In addition to understanding public opinion the rules formed from mining the textual data can be easily understood to predict the user's perception on the outcome.

Paper #	Title	Primary Author	Abstract
SD-152	Intracompany Social Media Support for a SAS GRID Migration	Gordek, Harper	In 2016 RTI International migrated over 400 SAS users from SAS on PCs and several stand-alone Linux servers to a SAS Linux Grid. Despite best efforts, pre-migration company-wide training could not cover every potential topic. In addition, post-migration support from IT was necessarily focused on issues of base functionality. In order for specific programming issues to be addressed with expediency it became necessary then for new methods of communication and collaboration to arise. But how to best do this when these 400 users spanned several departments with limited historical communication between them? In response, staff began to use to social media applications like Slack and Yammer and created an internal migration support community. In this presentation, we'll cover how the limitations of company-wide approaches led to innovative solutions that gathered together a diverse internal community to better problem solve and share knowledge. We will also discuss how those solutions evolved over time and both their strengths and limitations.
SD-160	Multicollinearity: What Is It, Why Should We Care, and How Can It Be Controlled?	Schreiber-Gregory, Deanna	Multicollinearity can be briefly described as the phenomenon in which two or more identified predictor variables in a multiple regression model are highly correlated. The presence of this phenomenon can have a negative impact on the analysis as a whole and can severely limit the conclusions of the research study. This paper reviews and provides examples of the different ways in which multicollinearity can affect a research project, and tells how to detect multicollinearity and how to reduce it once it is found. In order to demonstrate the effects of multicollinearity and how to combat it, this paper explores the proposed techniques by using the Youth Risk Behavior Surveillance System data set. This paper is intended for any level of SAS® user. This paper is also written to an audience with a background in behavioral science or statistics.
SD-191	Churn the Data Around. A Machine Learning Approach to Understanding Why Customers Leave.	Ankenbruck, Sean	Customer attrition is a common problem that affects companies across a variety of industries. Companies value profitable customers that are consistent and engage in long-term interactions that contribute to a high customer lifetime value. Identifying and understanding the factors that influence customer churn allows companies to gain a competitive advantage. This paper discusses a data-driven approach to identify important factors that influence churn and classify individuals based on their predicted likelihood to churn. It will also provide a visual analysis of churn rates across various customer groups. The entire process is accomplished using SAS tools such as SAS Visual Data Mining and Machine Learning.
SD-212	Meta-Analysis and Matrix Multiplication: Adapting an IML-Based Macro for DerSimonian and Laird's error calculation to use Base SAS.	DePuy, Venita	Proc IML is a wonderful tool; but not all companies opt to purchase the SAS license for that product. This paper summarizes my journey to adapt Hertzmark & Spieglman's %METAANAL macro, which produces DerSimonian-Laird estimators for meta-analysis, to use Base SAS. The macro itself is also provided.

Paper #	Title	Primary Author	Abstract
SD-215	Propensity Score Methods for Causal Inference with the PSMATCH Procedure	Yung, Yiu-Fai	In a randomized study, subjects are randomly assigned to either a treated group or a control group. Random assignment ensures that the distribution of the covariates is the same in both groups and that the treatment effect can be estimated by directly comparing the outcomes for the subjects in the two groups. In contrast, subjects in an observational study are not randomly assigned. In order to establish causal interpretations of the treatment effects in observational studies, special statistical approaches that adjust for the covariate confounding are required to obtain unbiased estimation of causal treatment effects. One strategy for correctly estimating the treatment effect is based on the propensity score, which is the conditional probability of the treatment assignment given the observed covariates. Prior to the analysis, you use propensity scores to adjust the data by weighting observations, stratifying subjects that have similar propensity scores, or matching treated subjects to control subjects. This paper reviews propensity score methods for causal inference and introduces the PSMATCH procedure, which is new in SAS/STAT® 14.2. The procedure provides methods of weighting, stratification, and matching. Matching methods include greedy matching, matching with replacement, and optimal matching. The procedure assesses covariate balance by comparing distributions between the adjusted treated and control groups.
SD-218	Bayesian Hierarchical Modeling using PROC MCMC	Chen, Fang	This tutorial describes how to use the MCMC procedure to fit Bayesian mixed models and compares the Bayesian approach to fitting the classical models with familiar mixed modeling procedures (MIXED, GLIMMIX, and NLMIXED) The tutorial describes fundamental concepts in making Bayesian inferences (such as the selection of noninformative prior distributions) and covers key elements in translating mixed models into PROC MCMC (such as coding CLASS variables and fitting models with various covariance types). A series of increasingly complex examples illustrate applications with PROC MCMC, from linear models to generalized linear models to repeated measurements. Additional details, such as custom hypothesis testing and predictions, are also presented. The tutorial concludes with a brief discussion on fitting classes of hierarchical models that are beyond the capabilities of the mixed modeling procedures.

Paper #	Title	Primary Author	Abstract
EPO-20	Using SAS® to examine the relationship between nursing attitudes towards suicidal patients in the Emergency Department (ED)	Tavakoli, Abbas	Mental illness visits in Emergency Department (ED) are annually 12 million in United States. There is need to prepare staff with mental issues. The purpose of this paper is using SAS® to examine the impact of an educational initiative on nurses' attitudes about patients expressing suicidal ideation in the emergency departments 2 rural hospitals. Two rural communities ED were used for this study. Deliberate Self- Harm Questionnaire (ADSHQ) used to assess attitudes included confidence, empathy, dealing effectively, and coping. The outcomes were attitude of nurses towards suicidal patients were measured at baseline and after intervention. The intervention for this study required that specialized training be provided to nursing staff working. Aggregated data was collected from health stream when the study was completed. Each Item was analyzed and reported in a percentage or rate per item. Proc Mean, Freq, ttest, and univariate used to analyze this data. T-Test and Wilcoxon signed ranked test used to examine the difference of means for attitude of nurse pre and post of intervention. The results showed there was improvement in attitude nurses towards suicidal patients after intervention. However, the results were not statistically significant.
EPO-30	How to use UNIX commands in SAS code to read SAS logs	Willis, James	SAS batch programs running in UNIX can be chained together using include statements. When a long chain of programs is executed and each program in the chain has its log printed separately, reviewing all the logs from the process, can be difficult but is definitely time consuming. Reading the logs using UNIX commands and SAS logic simplifies and speeds up the log verification process. "ERROR:", "WARNING:", "CONVERSION", "MISSING", "NOT FOUND" and all other messages can be located, ranked, written to a spreadsheet, and finally, the spreadsheet is attached to an email that is sent to you.
EPO-63	Behind the Scenes: from Data to Customized Swimmer Plots Using SAS® Graphical Template Language (GTL)	Tsang, Rita	A Swimmer plot is a useful visual tool to assess treatment efficacy for individual subjects in oncology clinical trials. These efficacy endpoints may include tumor response evaluation over time as measured by Response Evaluation Criteria in Solid Tumors (RECIST), or percent change from baseline in target lesions for measurable disease as defined by the presence of at least one measurable lesion. These endpoints are always juxtaposed with other valuable information, such as treatment doses, duration of treatment, primary tumor type, and treatment status to create a compelling visual presentation. There are many industry publications describing how to create swimmer plots in SAS. They provide a useful framework for using either SGPLOT or GTL to generate swimmer plots. This paper will build on the flexibility of SAS GTL to go further behind the scenes to illustrate how to prepare the data structure and variables for a customized swimmer plot. Focus will be placed on using dynamic programming for addressing missing data. The SGRENDER procedure will be used to create the graph based on the prepared data set and the graph template created in PROC TEMPLATE.
EPO-82	When ANY Function Will Just NOT Do	Watson, Richann	Have you ever been working on a task and wondered whether there might be a SAS® function that could save you some time? Let alone, one that might be able to do the work for you? Data review and validation tasks can be time-consuming efforts. Any gain in efficiency is highly beneficial, especially if you can achieve a standard level where the data itself can drive parts of the process. The ANY and NOT functions can help alleviate some of the manual work in many tasks such as data review of variable values, data compliance, data formats, and derivation or validation of a variable's data type. The list goes on. In this poster, we cover the functions and their details and use them in an example of handling date and time data and mapping it to ISO 8601 date and time formats.
EPO-91	Purrfectly Fabulous Feline Functions	Hadden, Louise	Explore the fabulous feline functions and calls available in SAS® 9.1 and later. Using CAT functions and CAT CALLs gives you an easier way to streamline your SAS code and facilitate concatenation of character strings. So, leave verbose coding, myriad functions, and the vertical bar concatenation operators behind! SAS® 9.2 and later enhancements will also be demonstrated.

Paper #	Title	Primary Author	Abstract
EPO-95	Using SAS® to proactively monitor drug alerts: Implementing the URAC "Drug-Drug Interactions" measure for a specialty pharmacy	Wallace, Karen	To satisfy annual URAC accreditation requirements, one of the mandatory reported "Safe Care" domain measures is "Drug-Drug Interactions" (DM2012-13). This measure relies upon pharmacy and enrollment data to assess the percentage of patients who: 1) received a prescription for a target medication during the measurement period, and 2) were dispensed a concurrent prescription for a precipitant medication. The resultant data are stratified by line of business (URAC, 2016). Using SAS® Base programming, in conjunction with technical and clinical review, the preliminary methodology: 1) imports and joins several files using the data elements of patient identifier as well as medication and insurance plan, 2) cleanses the mixed-case, punctuated MediSpan-generated drug alert text, 3) categorizes the text into discrete flags based upon predetermined criteria, 4) sums the number of flags per observation, and finally, 5) segments data by insurance type. The reporting showcases de-identified, summarized descriptive statistics that may be used to better monitor drug alerts and, more importantly, intervene where there is perceived harm in the patient population. This paper is intended for the SAS® novice who may be challenged on ways to perform on-the-fly textual analysis with a mindset towards implementing standardized, automated reporting after receiving requisite oversight and sign-off.
EPO-111	Beginner Beware: Hidden Hazards in SAS® Coding	Wise, Alissa	New SAS® programmers rely on errors, warnings, and notes to discover coding issues. However, it is important to note that some coding issues may be hiding in plain sight. Herein are a few examples of these issues including incomplete comparisons and inadvertently truncating variables with the IMPORT procedure. The explanations provided are meant to assist new SAS® programmers navigate these hazards so that results are clean and programs run more efficiently.
EPO-131	Using SAS® Enterprise Miner for Categorization of Customer Comments to Improve Services at the U.S. Postal Service	Olatunji, Olayemi	Delivering high-quality service and providing excellent customer experiences are performance outcome goals the U.S. Postal Service has established to measure corporate strategy success and continuous improvement efforts. Social media has opened the door for customer engagement and decision making. With the help of Twitter, Facebook, and Yelp, government agencies are more informed about how customers feel about their service and experience. Using Yelp data we will text mine comments about U.S. Postal Service customer service, retail service, mail delivery, and facility services using SAS® Text Miner of SAS® Enterprise Miner 7.1. The aim of this paper is to provide ways to categorize consumer comments regarding U.S. Postal Service services to improve the customer experiences at stations.
EPO-189	Statistical Analyses of Public Health Surveys Using SAS® Survey Package	Ha, Trung	We will discuss the key SAS® survey package (SAS 9.4 or later) procedures that one can use to perform statistical analyses of complex survey data with replicate weights. We will address the topic of replicate weights and balanced repeated replications for variance estimation. The key methods for obtaining summary statistics and performing testing for quantitative and qualitative outcomes will be discussed. In addition, model building approaches for logistic regressions with fixed effects will be addressed. Examples will illustrate the concepts and procedures for smoking-related behaviors using the Tobacco Use Supplement (TUS) to the Current Population Survey (CPS) data set with 160 replicate weights. The survey is widely used to estimate the prevalence of smoking, as well as rates of smoking-related events and behaviors, e.g., quitting behaviors among smokers and attitudes towards smoking laws.
EPO-199	How High will my Magikarp Jump?	Maher, Marcus	The Pokémon Company's new mobile game, Magikarp Jump, has simple game play where the goal is to train your Magikarp to jump as high as possible. The distance jumped (in meters) is solely determined by the jump points possessed by the Magikarp, which can be earned by eating berries and by training your Magikarp. Measurements for jump points are imprecise and only recorded with three significant digits, and the relationship between the two is not defined by the game. What is the relationship between height jumped and jump points, and how does one deal with potential bias from imprecise measurement? Using data obtained by the presenters and their children through several dozen playthroughs, this presentation explores techniques used in estimation and the importance of statistical graphs for exploring the data and assessing quality of fit.

Paper #	Title	Primary Author	Abstract
EPO-203	Mysteries of Posting Results on reddit	Wilcox, Adeline	After reading a December 2016 USA TODAY story revealing data leaked from the U.S. Department of Veterans Affairs (VA), this retired VA employee doubted the quality of the leaked data. After I completed my analysis of the data published by USA TODAY, I wanted to share my results, but not by corresponding with USA TODAY. Naively, but successfully, I created a reddit account and posted my analysis on the subreddit named dataisdepressing. My post remained at the top of the front page of the dataisdepressing subreddit for nearly three months, despite receiving no comments and few upvotes. For analyses of general interest that produce cheerless results, the dataisdepressing subreddit can be a means for sharing results or visualizations that can be displayed on a single screen. Reddit users submitting content to the dataisdepressing subreddit should not be disappointed if their post receives no more than a few hundred views.
EPO-225	YAGI: a tool to help Newcomers to SAS efficiently learn Code and Syntax	Louis, Steph-Yves	SAS is one of the most widely used software packages for statistical calculations and data manipulation purposes across many fields. While books and online resources already exist to help with syntax, researching these materials can be tedious and time-consuming. For a new student, the learning can be overwhelming. Even for an individual familiar with the program, attempts to retrieve code from other users' codes may take longer than expected. Considering the above issues, YAGI, a software designed to facilitate the learning for novice coders and to represent a personal dictionary of codes for any returning user, may present a solution.