Paper #	Title	Authors	Abstract
Development and Support			
DEV121	Application of Fuzzy Matching Techniques Using SAS® Software - A Panel Discussion	Kirk Paul Lafler and Stephen Sloan	Data comes in all forms, shapes, and sizes. When consistent and reliable identifiers (or keys) between two or more files exist, SAS® users are typically able to search and match observations without a problem. But, when a unique identifier (or key) is not consistent or reliable then the matching process can become compromised. This panel discussion will explore topics related to what fuzzy matching is, common data issues, popular data cleaning and validation techniques, the application of the five CAT functions, the application of the SOUNDEX (for phonetic matching) algorithm, the SPEDIS, COMPLEV, and COMPGED functions, and an assortment of SAS programming techniques to resolve key identifier issues and to successfully search, merge/join and sort less than perfect or messy data.
DEV123	Orchestrating and triggering SAS process execution from custom process control tables	Angye Rivero and Osmel Brito Bigott	When we have different processes in charge of feeding a Data Warehouse, commonly in the design of the process and the pipeline a process control table is included, from which it must be read in order to generate a trigger that starts the execution of one or more several programs that carry a logical sequence at the time of loading information, thus being able to integrate the execution of SAS processes to the execution of processes from different sources or technologies.

This paper will expose a way to generate an orchestration and trigger scheme from the data loaded in a control table with a simple and easy design, as well as the implementation so that it can be customized according to your needs.We will also show you a practical example of how you can use a control table, your SAS programs and schedule these executions to improve the automation of the execution of your processes. |
| DEV135 | Names From Template | KANNAN DEIVASIGAMANI | SAS programmers at times have these datasets with plenty of variables that are similar or cloned from another dataset but contain additional variables due to some intermediate computations. At the end of the day, we are tasked with retaining the original variables and drop any additional ones. We end up copying a list of variables manually or typing them in the keep statement. What if there is a way to automatically have the list available and the copy to be made using the template dataset that has the variables we need. Well, you've got it. This macro does that exactly that. It keeps the variables from the template dataset and makes a new copy or overwrites the existing one depending on your requirement. The macro accepts 3 parameters as input from the user. The first parameter is the input dataset that has all the variables that needs to be cleaned up. The second parameter is the variable-template which is a SAS dataset that has the variables of interest. The final parameter is the name of the output dataset that is the cleaned up version. This could either be the original name or a new name depending on the user's need. No more typing the variable names from another dataset or copying manually from elsewhere. This macro is a solution to the challenge. This increases productivity and makes a programmer more powerful. |
| DEV139 | Twenty Ways to Run Your SAS® Program Faster and Use Less Space | Stephen Sloan | When we run SAS® programs that use large amounts of data or have complicated algorithms, we often are frustrated by the amount of time it takes for the programs to run and by the large amount of space required for the program to run to completion. Even experienced SAS programmers sometimes run into this situation, perhaps through the need to produce results quickly, through a change in the data source, through inheriting someone else's programs, or for some other reason. This paper outlines twenty techniques that can reduce the time and space required for a program without requiring an extended period of time for the modifications. The twenty techniques are a mixture of space-saving and time-saving techniques, and many are a combination of the two approaches. They do not require advanced knowledge of SAS, only a reasonable familiarity with Base SAS® and a willingness to delve into the details of the programs. By applying some or all of these techniques, people can gain significant reductions in the space used by their programs and the time it takes them to run. The two concerns are often linked, as programs that require large amounts of space often require more paging to use the available space, and that increases the run time for these programs. |
| DEV140 | Reducing the space requirements of SAS® data sets without sacrificing any variables or observations | Stephen Sloan | The efficient use of space can be very important when working with large SAS® data sets, many of which have millions of observations and hundreds of variables. We often need to fit the data sets into a fixed amount of space. Many SAS data sets are created by importing Excel or Oracle data sets or delimited text files 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 we can do to make the SAS data sets more compact and 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 shrink the space of the entire data set. There are also tests we can run to shrink the length of different variables and evaluate whether they are more efficiently stored as numeric or 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.
DEV162	Pol-y-mor-phism in SAS, Or, Good Programmers are Lazy	david horvath	This session reviews techniques for implementing simple polymorphism in SAS programs. As part of an effort to productionalize a large number of models with multiple variables with long time series, a datamart was created in SAS datasets to minimize impact on the corporate data warehouse. Rather than coding the inputs multiple times in the build process, a set of macros were created to define the DDL for input to each model. By including different files that define those macros, different functions were implemented: Prime, shift months, define keep string.
DEV163	Data Wrangling in a Highly Regulated Industry	david horvath	Data is rarely clean – it needs to be wrangled into a form that your analysts can actually use it. Wrangling that data becomes more complicated in a highly regulated industry such as Pharma or Banking/Finance. This session will look into those areas and provide suggestions on how to deal with the related issues.
DEV165	Utility Macros to Check for Changes in Macro Variables, Options, or Formats in SAS®	Aaron Brown	Imagine a scenario where you create a particular macro variable, set your options as needed, or create formats you need early in a program. Later, you accidentally change one of these settings, which breaks assumptions further down in the program. Such a scenario can be difficult to troubleshoot. This paper introduces three sets of macros that use the metadata SAS® creates during a SAS session in order to check if macro variables, options, or formats changed. It utilizes the COMPARE, FORMAT, SQL, and OPTSAVE procedures.
DEV174	Calling for Backup When Your One-Alarm Becomes a Two-Alarm Fire: Developing SAS® Data-Driven Concurrent Processing Models through Control Tables and Dynamic Fuzzy Logic	Troy Hughes	In the fire and rescue service, a box alarm (or, simply, "alarm") describes the severity of a fire. As an alarm is elevated from a one-alarm fire to a multi-alarm fire, additional, predetermined resources (e.g., personnel and apparatuses) are summoned to combat the blaze more aggressively. Thus, a five-alarm fire—or its equivalent "five-alarm" Cincinnati chili—represents something extremely hot and dangerous. After extinguishment, and as smoke and embers recede and overhaul begins, fire and rescue resources are released "back into service" so they can be utilized elsewhere if necessary. Essential to managing complex fireground operations, this load balancing paradigm is also common in grid and cloud computing environments in which additional computational resources can be shifted temporarily to an application or process to maximize its performance and throughput. This text instead demonstrates a programmatic approach in which SAS® extract-transform-load (ETL) operations are decomposed and modularized and subsequently directed (for execution) by control tables. If increased throughput is required, additional instances of the ETL program can be invoked concurrently, with each software instance performing various operations on different data sets, thus decreasing overall runtime. A shared control table provides the communications backbone for all SAS sessions by tracking incomplete, in-progress, and completed operations for all data sets. A configuration file allows end users to specify prerequisite processes (that must be completed before an operation can commence), thus facilitating the dynamic fuzzy logic that autonomously selects the specific ETL operations to be executed. This data-driven design approach ensures that the execution of operations can be prioritized, optimized, and, to the extent possible, run in parallel to maximize performance and throughput.
DEV210	SaviApp - An Analysis Toolkit for SAS Environments	ALAN CHURCHILL	SAS language and usage analysis is critical for determining SAS product usage within the programs. Additionally, SAS Enterprise Guide projects are very complex and need to be deciphered across a large pool of projects. SaviApp is a free tool available to the SAS community that enables large scale discovery and analysis of SAS programs, EG, datasets, logs, et al. It uses parsing technology to discover the components of a SAS program even if embedded in an EG project.
DEV211	A Comparison of Typical Programming Processes Using SAS® and R	Brian Varney	This hands-on-workshop will walk participants through examples of typical programming processes in SAS and R. This will enable R users interesting in learning SAS or SAS users interested in learning R to add new methods to their tool box. Programming processes will include but not be limited to: * Reading/Writing Data * Combining Data * Group by Processing * Transposing/Pivoting Data * Handling Duplicates * Summarizing Data After this workshop, participants will have usable code snippets which they can apply to their daily work.
DEV224	N-gram Fuzzy Matching Revisited	Floyd Nevseta	N-gram fuzzy matching logic offers another useful method for measuring the similarity between strings, particularly for strings with word order differences such as street addresses and company names. Unfortunately, Base SAS® software does not provide a built-in

			function that encapsulates the method. Some past conference papers thoroughly discussed the fuzzy matching technique with code that implemented the logic for performing the comparison. From the code in those papers, the author created a function with the FCMP procedure to make it as easy to use as SPEDIS, COMPLEV and the other string comparison functions. However, while testing the function flaws in the algorithm surfaced. This paper will explain the flaws in the original algorithm and the improvements to correct them.
DEV231	Using SAS® to Optimize Your Data Dictionary and Improve Documentation	Muhammad Salaam	When working with large datasets, it is important to accurately document the datasets and their contents. This makes the user knowledgeable of all of the nuances in the data, while also providing a guide for teach others about the data. The most commonly used method to data documentation is done by developing a data dictionary. While there are numerous ways to create a data dictionary, data dictionaries are often created and are underutilized or are never updated over time. This presentation will explore commonly used data documentation methods and how SAS can be used to improve these methods to develop a comprehensive data dictionary.
DEV233	Long to Wide Format	Isaiah Omerhi	I mostly work with HIV data, and the reports generated from the various medical record sources are in the long format with multiple rows for a single ID variable. To perform the desired analysis on the data all unique ID must be in one row. To do that I use SAS to reshape the long format into wide format using proc TRANSPOSE. However, the data comes with multiple repeating categorical variables that need to be transposed and this introduces a special challenge to the Transpose procedure.

Industry Applications

IND127	Standardized, Customized or Both? Defining and Implementing (MedDRA) Queries in ADaM Data Sets	Richann Watson and Karl Miller	Investigation of drug safety issues for clinical development will consistently revolve around the experience and impact of important medical occurrences throughout the conduct of a clinical trial. As a first step in the data analysis process, Standardized MedDRA Queries (SMQs), a unique feature of MedDRA, provide a consistent and efficient structure to support safety analysis, reporting, and also address important topics for regulatory and industry users. A variance in working with SMQs is the ability to limit the scope for the analysis need (e.g., "Broad" or "Narrow") but there is also the ability outside of the specific SMQs in allowing the ability to develop Customized Queries (CQs). With the introduction of the ADaM Occurrence Data Structure (OCCDS) standard structure, the incorporation of these SMQs, along with potential CQs, solidified the need for consistent implementation, not only across studies, but across drug compounds and even within a company itself. Working with SMQs one may have numerous questions: What differentiates the SMQ from a CQ and which one should be used? Are there any other considerations in implementation of the OCCDS standards? Where does one begin? Right here…
LRN131	SAS® PROC GEOCODE and PROC SGMAP: The Perfect Pairing for COVID-19 Analyses	Louise Hadden	The new SAS® mapping procedure PROC SGMAP is adding capability with every release. PROC SGMAP was introduced in SAS 9.4M5 as an extension of (ODS) Graphics techniques to render maps and then overlay plots such as text, scatter, or bubble plots. It has contributed a lot of functionality which used to be reserved for SAS/GRAPH users to BASE SAS – including PROC GEOCODE. PROC GEOCODE has been available in SAS/GRAPH since Version 8.2, and recently became available in BASE SAS with a number of other tools in Version 9.4 Maintenance Release M5. SAS provides a link to files required for street level geocoding and more on SAS MAPSONLINE. The ongoing COVID-19 pandemic has produced massive amounts of epidemiological and surveillance data, much of which can be linked to geography as countries, including the United States, grapple with how to address the constantly transmuting contagion. The combination of PROC SGMAP and PROC GEOCODE is well positioned to help researchers address and visualize COVID-19 data. This paper and presentation will walk through the graphic representation of publicly available COVID data sources using PROC SGMAP and PROC GEOCODE.
IND166	Using SAS® to Prepare Postsecondary Data Partnership (PDP) Data Submission Files	Bobbie Frye	Transforming the way a higher education institution measures and reports student progress so that every student can thrive can be accomplished with joining the Postsecondary Data Partnership (PDP) and submitting comprehensive student data to the National Student Clearinghouse. The PDP empowers institutions with more comprehensive data, easier analysis, and better visual representations to help institutions understand, improve, and communicate student momentum, outcomes, and equity.

In 2021, Achieving the Dream launched a PDP Onboarding Course designed to support the initial data submission process. The PDP Onboarding Course is supported by The Bill and Melinda Gates Foundation and is designed to support both two-year and four-year institutions in successfully onboarding to the PDP.

To facilitate the preparation of data files, the National Student Clearinghouse (NSC) provides documentation for file formatting and field length requirements. PDP Technical Briefs have been authored by Achieving the Dream SAS® users to provide support to those onboarding to the PDP. The purpose of this paper is to showcase the power of SAS® software in facilitating the file creation process. Specifically, the use of macro variables in the SAS® environment adds efficiency to the file creation process, and macro variables utilized throughout file creation aid in accomplishing file and field formatting requirements. The flexibility of macros equipped the analyst with a file creation process that is replicated across multiple years or terms. Additionally, to harness the power of SAS® macros in file creation processing, macros were combined with LIBNAME statements to point to file locations, used to add additional required fields, and used in file naming conventions.

In this paper, several pieces of SAS® code were extracted and isolated for illustration purposes. The initial data collection is the most time consuming, including up to five years of data, but the process outlined in this document can be replicated for future data collections. The SAS® code can be utilized by beginners or intermediate programmers to prepare files for the PDP. The PDP equips institutions with accessible reports and visualizations and SAS® software provides a flexible roadmap to successful data submissions. |
| IND177 | One Click Excel Cleanup - Generate High Quality | Suresh Acharya | The focus of this paper is to generate a high-quality Excel specification document in one click with the help of a Visual Basic for Applications (VBA) macro. In the pharmaceutical SAS programming industry, documents such as SDTM and ADaM specifications are required to create SDTM and ADaM datasets, and most times these are in Excel format. Statisticians and SAS programmers typically update these documents multiple times in-life as per the study requirements and analysis needs. These updates are frequently tracked |

	SDTM/ADAM Specification		by using various types of markings such as different colored text, highlighted cells, and overstrikes. These updates are often not formatted consistently leaving highly inconsistent text formats and fonts in the document. Removal of formatting is a manual process which is tedious and time consuming. It takes 20-30 hours per protocol deliverable with increased likelihood of human error and inconsistencies. As the number of worksheets increases, the time spent to clean up the markings also increases significantly. The VBA tool presented in this paper generates a high-quality Excel specification document that is consistently formatted, as per the formatting standards, and saves 20-30 hours during each delivery cycle. Final output is clean of any strikeouts, empty rows, empty columns, out of context borders and colors. In addition, it also updates font style, size and color to a consistent standard format while removing any additional cell fill color. With an option to clean a single sheet or multiple sheets, this tool assists in creating high quality excel SDTM or ADAM specifications.
IND184	REDCap: Your SAS Friend for EHR Manual Abstraction	Brooke Ellen Delgoffe and Steffani Roush	This presentation will demonstrate the utility of REDCap in the context of abstracting electronic health records (EHR) data in a way that talks well with SAS. REDCap allows SAS to query against it using an API for easy data exchange. The associated paper will present an advanced export macro that also automatically downloads the data dictionary and creates a labeled SAS data set, complete with labels and a format library based on the REDCap data dictionary. It will also cover special data concerns when using REDCap for housing PHI and methods for interfacing with an EHR.
IND188	A two-staged local regression based binning method for weight of evidence transformation in credit scoring models	hui wang, Shirong Huang, Emma Zhou and Erin Martin	Weight of Evidence (WOE) variable transformation method is widely and commonly used in credit risk analysis. This work provides a two-staged, local regression based binning method to estimate WOE. Using the banking industry dataset as an example, this paper shows that when a sufficient number of bins are selected and appropriate smooth factor is chosen, the loss of information and prediction accuracy could be minimized. The proposed method performs well on imbalanced dataset. It can also handle either monotonic or U-shaped relationships between the transformed WOE and original variable by delivering results that have business soundness. The model created with this approach can enable the users to experience more smooth credit score migration when a financial ratio shifts from one bin to another. Considering the widely acknowledged advantages of using WOE method, such as the ease of handling missing values, and the good interpretability of the model after transformation, this approach is considered to have good performance compared to the existing variable transformation approaches and can meet the business needs specifically for credit risk analysis.
IND192	SAS Heatmaps in the Analysis Real Estate Markets	Harshita Budumuru	The pandemic has introduced a new dynamic of societal uncertainty for stakeholders and home buyers fueled by rapid inflation and soaring house prices. The federal reserve has stepped in this past month by increasing interest rates to limit credit flows. Some analysts are predicting another housing bubble while others say logistic growth due to stagnant supply chains. Will this volatile housing market crash or are we in a perpetual state of exponential growth? The question of where and when to invest has become a challenging discussion, but is there a solution? Scientific literature on the use of heat maps for statistical analysis increased in almost every industry and field. Of late, heat maps have become increasingly popular in real estate, especially as a useful analytical tool for making informed decisions in the housing market. This paper aims to illustrate how SAS can be used to generate these heat maps which present complex data as simple visual aids. SAS can aid in the analysis of these hot housing markets to be readily available to entrepreneurs, mortgage institutions, and policy makers. In this article, real estate data will be used to identify key parameters such as employment growth, population growth, first time home buyers, listing prices, listing quantities, rental prices, occupancy rates, and much more. The future of the housing market landscape can be demonstrated and the differences between the spiraling recession and the booming pandemic can be addressed.
IND193	To Lend, or Not to Lend, That Is the Question! Evaluating Loan Application with Machine Learning	Rafae Abdullah, Sucharitha Vallabhaneni and Jaideep Muley	Have you ever lent money to someone? If yes, you must know how much headache lenders go through when they do not get their money back within the expected time. This is true for any financial institution. The goal of this paper is to apply various machine learning models by using SAS® Enterprise Miner in the loan lending process of a peer-to-peer lending company and result in the best approach that can be used by peer-to-peer lending companies to automate the loan acceptance process and by other financial institutions to build a more robust model for them since their amount of lend and amount of information on borrowers is huge. Algorithms such as Decision Tree, Random Forest, Neural Network, and so on were used to find the best alternative to ease the loan lending process. The data set used contains plenty of information about the accounts of the clients and the loans funded in the period from 2007 to 2018. We started our analysis with looking at statistical summary and visualizations of the variables, data cleaning and pre-processing, missing value imputation, and finally model building and evaluation. From our analysis, we have found that the Random Forest model outperforms the others. Based on the outcome, we hope that financial institutions can apply Machine Learning to automate their loan approval procedure. Also, to handle, transform, and keep track of around 100 inputs, a tool like SAS® Enterprise Miner is very much useful.

IND209	A unique and innovative end-to-end demand planning and forecasting process using a collection of SAS products	Stephen Sloan, Kevin K. Gillette, Sushil B. Kumar and Puspita Kumari Padhi	Forecasting demand can be a very tricky process. Questions arise about which statistical algorithms to use when forecasting based on past sales, how to incorporate business knowledge into the forecast, planning for unforeseen events, and planning for unique events that would not be predictable from sales history. As an example, COVID caused many forecasts to be wrong about quantities when purchasing switched from services to goods, from in-store sales to remote purchases, and from work in the office to remote work. Even when the total quantities of a product were forecasted incorrectly, the percentage distribution of the sales for some of the subcategories within the larger categories was often accurate, and vice versa. An example of this is the switch from eating and drinking in restaurants to take-out, where the total quantities of the items might have been the same, but the packaging changed (fewer kegs, more six-packs). The solution, then, is to leverage the useful information from the statistical forecasts while allowing the people who know the business to make individual or mass updates. All of this can be accomplished using existing SAS products: using SAS EG and SAS DI to read in, manipulate, and output data; using the High-Performing Forecasting (HPF) SAS PROCs, which underpin SAS Enterprise Miner, to create statistical forecasts; and using SAS Financial Management (SAS FM), which incorporates Excel features while remaining within SAS, to allow users to make individual and mass changes to SAS data sets.
IND215	Using SAS to Geocode Injury-related Deaths in North Carolina	Bruce Nawrocki, Scott Proescholdbell, Shana Geary and Mike Dolan Fliss	BACKGROUND: Person, time and place are core epidemiological concepts. In North Carolina (NC), historical units of analyses have been state and county, yet other geographic units might enable a better understanding of the patterns of injury deaths and a more robust response. This GIS project's purpose was meant to explore other various geographic units available in NC that had not been utilized for injury control before, such as zip codes, census tracts and political boundaries (US Congressional Districts, NC Senate and NC House districts), and develop a production process to handle these in a standardized way. METHODS: We used NC death certificate resident addresses starting with 2014 and standard ICD-10 coding to select specific injury deaths in NC (unintentional drug overdose and violent deaths). We wrote new SAS code to geocode these injury-related deaths. We used SAS Enterprise Guide (EG) to incorporate all of our individual data preparation and processing tasks into a single process flow that first used SAS geocode datasets and procedures to assign Lat/Long values to addresses, then assigned each geocoded death record into specific geographic units such as zip, census tract and US Congressional Districts. We then analyzed and mapped violent deaths and overdoses using these various geographic units.
IND217	Equity Data and SAS®: Modeling Intersectionality	Kelly Smith	Multiple aspects of students' background and social/economic environment can impact academic success; therefore, equity data encompasses a wide range of variables including but not limited to demographic and academic metrics. By using an intersectional lens, researchers can obtain a deeper understanding of educational equity and potential hinge points for improving student success. A strong equity model should simultaneously consider multiple demographic data points and their relationship with academic metrics such as retention, persistence, and academic achievement. The presentation, appropriate for intermediate and advanced SAS users, will examine the use of Base SAS and Enterprise Guide to analyze equity data through an intersectional lens. Several approaches will be presented; the pros and cons of each approach will be included.
INDSAS2	SAS User Groups	Kate Ulveling	SAS User Groups.
INDPAN	How Industries Use SAS: A Panel Presentation		Discussion of SAS use in various industries.

Leadership, Careers and Planning

LCP109	Exploring the Skills Needed by the Data Scientist	Kirk Paul Lafler	As 2.5 quintillion bytes (1 with 18 zeros) of new data are created each and every day, the age of big data has taken on new meaning. More and more organizations across industries are embracing Data Science / Computer Research Scientist skills resulting in an emerging demand for qualified and experienced talent. According to the Bureau of Labor Statistics (BLS) the number of data science jobs is expected to grow 19 percent over the next two decades – nearly three times as fast as the average growth rate for all jobs. Energized by this employment outlook, students and professionals across job functions are preparing for tomorrow's growing data science / analytic demands by acquiring a comprehensive skill set. To prepare for this growing demand, many colleges, junior colleges, Universities, and vocational training organizations offer comprehensive degrees and certificate programs to fulfill the increasing demand for analytical skills. This presentation explores the skills needed by the Data Scientist / Analytics professional including non-technical skills such as critical thinking; business acumen and verbal/written communications; and technical skills such as data access; data wrangling; statistics; use of statistical programming languages like SAS®, Python, and R; Structured Query Language (SQL); Microsoft Excel; and data visualization.
LCP143	Developing and running an in-house SAS Users Group	Stephen Sloan	Starting an in-house SAS ® Users Group can pose a daunting challenge in a large worldwide organization. However, once formed, the SAS Users Group can also provide great value to the enterprise. SAS users (and those interested in becoming SAS users) are often scattered and unaware of the reservoirs of talent and innovation within their own organization. Sometimes they are Subject Matter Experts (SMEs); other times they are new to SAS but provide the only available expertise for a specific project in a specific location. In addition, there is a steady stream of new products and upgrades coming from SAS Institute and the users may be unaware of them or not have the time to explore and implement them, even when the products and upgrades have been thoroughly vetted and are already in use in other parts of the organization. There are often local artifacts like macros and dashboards that have been developed in corners of the enterprise that could be very useful to others so that they don't have to "reinvent the wheel".
LCP147	Acing Technical Questions in SAS Programmer Interviews	Lauren Rackley	Part of the interview process for SAS programming positions often involves technical assessments or technical questions. Important topics to know are SQL joins, ways to create macro variables, by group processing, and correct use of the RETAIN statement. In an effort to help candidates getting started as a SAS programmer, this paper will provide guidance on how to tackle technical questions and technical assessments given during interviews.
LCP171	Data for Good: Statistical Volunteering with SAS and How to Get Involved	David Corliss	This presentation brings together experiences from SAS professionals working as volunteers for organizations, charities, and in academic research. Pro bono work, much like that done by physicians, attorneys and professionals in other areas is rapidly growing in statistical practice as an important part of a statistical career, offering the opportunity to utilize your skills in a places where they are so needed but cannot be supported in a for-pay position. Statistical volunteers also gain important learning experiences, mentoring, networking, and other opportunities for professional development. The presenter will share experiences from volunteering for local charities, NGO's and other organizations and causes, both in the US and around the world. The mission, methods and focus of some organizations are presented, including DataKind, Statistics Without Borders, Peace-Work and others.
LCP202	Adventures in Independent Consulting: Perspectives from Two Veteran Consultants Living the Dream	Josh Horstman and Richann Watson	While many statisticians and programmers are content in a traditional employment setting, others yearn for the freedom and flexibility that come with being an independent consultant. In this paper, two seasoned consultants share their experiences going independent. Topics include the advantages and disadvantages of independent consulting, getting started, finding work, operating your business, and what it takes to succeed. Whether you're thinking of declaring your own independence or just interested in hearing stories from the trenches, you're sure to gain a new perspective on this exciting adventure.
LCP212	Organizational Considerations When Replacing or Adding a New Software Language	Brian Varney	Replacing or adding a new software language for your organization's users can add a lot of value and productivity to your team(s). There are, however, many factors and considerations to take into account to avoid chaos and hidden costs.
LCP216	SAS, SAS Coders, and IDEA (Inclusion, Diversity, Equity, Accessibility)	Kelly Smith	Fifty years ago, SAS coding was in development at North Carolina State University; the establishment of SAS as a business organization was still in the future. At that time, the American population was over 80% White and less than half of eligible women (43%) participated in the workforce. Fast forward to 2022; the American population is now approximately 60% White and over half of eligible women (56%) are in the workforce.

			America's population and workforce have changed significantly over the past 50 years. Recently, I started wondering … Where do SAS coders come from? What do they look like? How can SAS and SAS coders support the concepts of inclusion, diversity, equity, and accessibility through their work? Join me on a journey into understanding where SAS is now and where it's headed.
LCP218	Leading Change: Using Data and Feedback to Launch a Strategic Plan or New Initiative	Anne Oxenreider	Traditionally, new strategic plans begin with reviewing the past plan and scanning the environment. Establishing trust between planners and stakeholders is a crucial first step to managing and leading change. Creating a campus-wide feedback loop with data and idea-sharing provides the necessary groundwork to launch transformative change. This presentation will describe a pre-kickoff communication campaign that shared illustrative data points and gathered information from faculty and staff close to practice. The campaign used a Qualtrics survey to perform a gap analysis of core college functions and yielded valuable qualitative data. With current practitioner knowledge in hand, a kick-off event was launched that captured and maximized stakeholders' attention.
LCP219	A Brief History of SAS® Software, from a User's Perspective	David Maddox	This paper shares the experiences of the author as a SAS® user, administrator, and advocate beginning in the 1970's. The purpose of the paper is to convey to current users how SAS developed and evolved from a users perspective. While the story of SAS, the company, is quite interesting and informative, the focus here is on the software product itself and some of the applications that were developed over the years. In that context, maybe users of today can better understand how to position SAS in their careers.
LCPSAS1	The "For-You Page" of SAS: SAS Tips and News Continuously Delivered	Chris Hemedinger	SAS and social media.

Learning SAS I

ID	Title	Authors	Description
BTB106	Working with large decimals: how precise is my data with SAS ?	Osmel Brito Bigott, Yenireth Gil and María Victoria Daboín	SAS is a powerful tool for statistical analysis and calculation that makes easier when it comes to manipulating and reporting our data. However, what they have never told us about is what inconveniences can arise when it comes to having to be very precise in the quality of the numerical value of a calculation. In this sense, it is important to understand the ideal way to store numerical data when having an information repository in other storage systems (databases or files) other than SAS. In a project where we used Teradata to store results of historical statistical analysis in SAS we had faced with the question: How precise were numerical values in SAS to store decimals? And how should we upload them to a Tersdata Data Warehouse? It turns out that this answer depends a lot on the operating system where SAS is installed (Windows, Linux). In this paper we want to show the findings we made about the precision of numerical values and how we were able to solve the problem faced.
BTB107	Functions (and More!) on Call!	Richann Watson and Louise Hadden	SAS® Functions have deservedly been the focus of many excellent SAS papers. SAS call routines, which rely on and collaborate with functions, are less well known, although many SAS programmers use these routines frequently. This paper and presentation will look at numerous SAS functions and call routines, as well as explaining how both functions and call routines work in practice. There are many areas that SAS call routines cover including CAS (Cloud Analytic Services) specific functions, character functions, character string matching, combinatorial functions, date and time functions, external routines, macro functions, mathematical functions, sort functions, random numbers, special functions, variable control, and variable information. While there are a number of call routines and functions, we plan to drill down on character function call routines including string matching; macro, external and special routines; sort routines; random number generation routines; and variable control and information routines. We could go on and on about SAS Call Routines, but we are going to limit the call routines discussed in this paper, excluding any environment specific call routines such as those designated for use with CAS and TSO as well as other redundant examples. We hope to demystify SAS call routines generally, and will demonstrate real world applications of specific call routines, bringing some amazing capabilities to light.
BTB112	Demystifying PROC SQL Join Algorithms	Kirk Paul Lafler	When it comes to performing PROC SQL joins, users supply the names of the tables for joining along with the join conditions, and the PROC SQL optimizer determines which of the available join algorithms to use for performing the join operation. Attendees learn about the different types of join algorithms and explore nested loop (brute-force), sort-merge, index, and hash join methods along with selected options to control processing.
LRN116	The Battle of the Titans (Part II): PROC REPORT versus PROC TABULATE	Kirk Paul Lafler and Josh Horstman	Should I use PROC REPORT or PROC TABULATE to produce that report? Which one will give me the control and flexibility to produce the report exactly the way I want it to look? Which one is easier to use? Which one is more powerful? WHICH ONE IS BETTER? If you have these and other questions about the pros and cons of the REPORT and TABULATE procedures, this presentation is for you. We will discuss, using real-life report scenarios, the strengths (and even a few weaknesses) of the two most powerful reporting procedures in SAS® (as we see it). We will provide you with the wisdom you need to make that sometimes difficult decision about which procedure to use to get the report you really want and need.
BTB136	Have your cake and eat it too: Automated, sequential SAS batch jobs conditional on programmed log review.	Isaiah Gerber	While it is useful to break processing pipelines across multiple programs, at scale it can also make re-running analyses extraordinarily time-consuming and prone to human error. Prior work has tackled this problem by both laying out how to automatically batch numerous SAS programs in sequence (using tools like SAS Enterprise) as well as by developing macros to automate mass log review. This paper explores an approach which allows users to combine these two useful categories of utilities by sequentially batching any number of SAS programs while requiring each program in that sequence to pass a basic log check before batching the next program. The author reviews a simple SAS interface which calls a PowerShell function developed by the author to both 1) automate SAS programs in sequence and to 2) check each resulting log automatically for errors and warnings. This approach allows SAS users to have both the efficiency of automated sequential batching as well as the efficacy of automated log review contained within the same processing pipeline.
BTB148	Finding Duplicate Names and Addresses in Your Consumer Database	Joe DeShon	Despite the best efforts of all, duplicates are certain to be created in any consumer database. Because of variations in spelling and formatting of names and addresses, simply processing the database with PROC SORT NODUPKEY is not enough. Many companies offer services to help de-dupe such databases, but the process can be very inflexible and expensive. This paper describes a system written in Base SAS which identifies the most important parts of each name and address and thus produces a de-duped database with results comparable to most outsourced solutions. This is appropriate for any company with a large consumer database and limited resources needing to detect and respond to duplicate names and addresses in a consumer database.

BTB152	Using SQL Dictionaries to Research the Global Symbol Table	Ronald Fehd	The sql procedure in SAS software provides a number of dictionaries that can be used to research entries in the global symbol table. These dictionaries include lists of dataset and variable names, option values, and catalog entries for format values and macro definitions. Purpose: This paper provides example programs to research values in the global symbol table assigned by the global statement options, procedure output from the format procedure, and macro definitions.
BTB153	Have a Date with ISO®? Using PROC FCMP to Convert Dates to ISO 8601	Richann Watson	We have all had to deal with dates and sometimes determining whether a date is in a day-month or month-day format can leave us confounded. Because of this confusion, CDISC has implemented the use of ISO® 8601 format for datetimes in SDTM domains. However, converting these datetimes from the raw data source to ISO 8601 format is no picnic. While SAS® has many different functions and CALL routines there is no magic function to take raw datetimes and convert them to ISO 8601. Fortunately, SAS allows us to create our own custom functions and subroutines. This paper illustrates a custom function with custom subroutines that takes raw datetimes in various states of completeness and converts them to the proper ISO 8601 format.
BTB154	What's Your Favorite Color? Controlling the Appearance of a Graph	Richann Watson	The appearance of a graph produced by the Graph Template Language (GTL) is controlled by Output Delivery System (ODS) style elements. These elements include fonts and line and marker properties as well as colors. A number of procedures, including the Statistical Graphics (SG) procedures, produce graphics using a specific ODS style template. This paper provides a very basic background of the different style templates and the elements associated with the style templates. However, sometimes the default style associated with a particular destination does not produce the desired appearance. Instead of using the default style, you can control which style is used by indicating the desired style on the ODS destination statement. However, sometimes not a single one of the 50-plus styles provided by SAS® achieves the desired look. Luckily, you can modify an ODS style template to meet your own needs. One such style modification is to control which colors are used in the graph. Different approaches to modifying a style template to specify colors used are discussed in depth in this paper.
LRN157	Leveraging the CONTENTS PROCEDURE for easier SET statements	Tamar Roomian	When setting together multiple data sets in the data step, two common problems can arise. First, the data sets can share variable names in common that are of different datatypes, resulting in error messages in the log. The error messages list which variables are both numeric and character, but fail to mention for which data sets, making it difficult for the user to resolve the issue using the log messages alone. Second, the data sets can share character variables of different lengths, which results in truncation of data when set together. The SAS Institute has published a macro program, %union, to combine data sets that have variables in common of different lengths, however it can only be used for two data sets at a time, and it does not account for differences in case or format length. When joining more than two data sets that contain hundreds of variables in common, resolving these discrepancies becomes time consuming and tedious. This paper will demonstrate two macro programs that both take advantage of the OUT statement of the CONTENTS procedure. The first generates an easy-to-read table of all variables in common across the data sets that have mismatched datatypes and lists which type is in which data set. The second takes the maximum length and maximum format length, and generates code saved in a macro variable to be used in the data step. Together, these two programs make setting multiple data sets in the data step faster, easier, and with less programming.
LRN164	A Gentle Introduction to Creating SAS Graphs	Dane Korver	Creating graphs using SAS can be intimidating even for an experienced SAS programmer. This paper will provide you with a very gentle introduction to creating graphs in SAS so that it is not an overwhelming experience when you are tasked with creating one.
LRN179	REST API for the Weary Beginner	jinson erinjeri	API stands for Application Programming Interface which is a mode of communication between programs in order to transfer data. REST is an acronym for REpresentational State Transfer which is nothing but a standard that guides the design and development of processes that enable effective communication between programs. Therefore, in the World Wide Web environment, REST API is used to interact with data stored on web servers. The objective of this paper is to present the basics of REST API's for a novice learner in simple terms coupled with examples of publicly available API. In addition, this paper will present features available in SAS as well as Python for communicating with REST API web services using the same set of examples.
BTB187	Offensive Programming: A Threesome of Error-Throwing Macros	Quentin McMullen	SAS® programming can feel fraught with danger. Each time a program is executed, there is a risk that an error in the source data or the code will lead to undetected erroneous results. Offensive programming techniques are designed to decrease that risk by adding code to detect and report errors as a program runs. This paper presents three offensive utility macros for runtime validation of SAS jobs: %DupCk detects duplicate key values in a data set; %Assert detects values in a data set that are invalid or unexpected; and %CheckRecordCounts detects the accidental deletion of records.

			By increasing the detection rates of common errors, these utilities increase the programmer's confidence in their results. Principles of offensive programming are discussed, as are principles of macro design encountered during the development of the macros.
BTB191	Array Searching Algorithms and Techniques	Paul Dorfman	As SAS data structures, arrays have a variety of uses, all based on quick direct accessibility of each array element by its own unique index. However, the latter is most important for using arrays for key lookup and data retrieval: Since array indices can be programmatically manipulated in every conceivable way, any known search algorithm can be implemented using SAS arrays. This paper is a classified compendium of array searching algorithms and their practical implementations in the SAS language.
LRN195	Take the Train to Transpose	Alissa Wise	To train the data to transpose to your style, the data decide. What are the basic elements in need? Do you need all sent? Do you need ten when nine are collapsible? Then have we got a poster for you.
LRN197	Learning Fun(damental) Character String Cleaning and Parsing Methods in SAS!	William Smith	Cleaning and parsing character strings is an important step in data processing activities to allow the data to be as analytically useful as possible. This often requires character strings to be summarized into discrete numeric categories (coded) or requires using character strings to fuzzy match records between datasets. In order to do these activities, a number of cleaning and parsing techniques are needed. This paper will examine different SAS functions and custom macros to clean character string data, remove or replace unwanted characters and symbols, and parse single character strings into separate strings based on user-defined characters and delimiters. The functions and macros explored within are easy to understand and fundamental for all SAS users at any skill level.
BTB203	Getting Started with DATA Step Hash Objects	Josh Horstman	The hash object provides a powerful and efficient way to store and retrieve data from memory within the context of a DATA step. This presentation will introduce the hash object, cover its basic syntax and usage, and walk through several examples that demonstrate how it can offer new and innovative solutions to complex coding problems. This presentation is intended for SAS® users who are already proficient with basic DATA step programming.
LRN220	Using a Hash Table to Add Diagnosis Related Information to a Health Claims File	Roberta Glass	PROC SORT is one of the most "expensive" routines in SAS in terms of resources required. When working with health claim files with millions of records, the last thing you want to do is sort the data multiple times. This paper presents a method of adding information from a file of diagnosis (dx) codes and their estimated probability of being an avoidable emergency room visit onto a claims file which does not require sorting the claims by diagnosis code. A hash table in SAS® base is used to create a look-up table of dx codes and their associated probability estimates. When processing the claims file, the probability of the dx code on the claim is located in the hash table and saved to the claims file. In addition to eliminating the need to sort, a hash table allows for the look-up of multiple diagnosis codes on one claim. This technique is intended to serve as an example for the SAS® user who needs to efficiently add information to a large data set.
BTB221	The PRXMATCH Function: A Perl of Great Price	Lauren Rackley and Josh Horstman	The PRXMATCH function is one of many hidden pearls in the SAS language. By harnessing the power of Perl regular expressions, PRXMATCH provides rich pattern-matching functionality useful for a variety of tasks involving text data. This paper provides an overview of Perl regular expressions and the PRXMATCH function as well as several examples demonstrating their use. If you work with text data at all, this is a tool you definitely want to have in your toolbox!
BTB223	Where Form(at) Meets Function: More Than Just a Pretty Face	Louise Hadden and Nancy McGarry	SAS® practitioners are accustomed to using PROC FORMAT to transform variables into an enhanced or aggregated version of the original schemas, via an assigned format used in a SAS procedure, through the use of put statements, and other intriguing uses, but there is so much more to this venerable SAS procedure. This paper will explore and demonstrate some lesser known techniques to enhance your SAS data sets and processing, including user defined functions to extend the power of PROC FORMAT, creation of complex format libraries, and comparison of format libraries to aid in quality assurance activities.
BTB229	GIT with GIT. A Stubborn SAS User's perspective	Zeke Torres	GIT is not just copying from github. And procrastinate too much and you will see the issues later when your resume lacks the right team coding skills. It's time to GIT with GIT. Let's take a folder with your project code and GIT started with GIT demonstrating how this saves us time and grief. Even if it's our own solo code or especially if it's a team project code folder. This presentation will cover: - Myth vs facts about what GIT is or is not. - Resources for what to download, where useful resources are. - Including "best practices" and "do's and don'ts" to be mindful of. Even if your team and work today doesn't include GIT - it likely will soon or in your next career chapter. GIT and version control are not something you can afford to ignore.

BTBSAS3	SAS 9.4 - 5 top migration updates for programmers	Charu Shankar	
LRNSAS4	Know Thy Data – techniques for data exploration	Charu Shankar	
LRNSAS5	Getting Started with ODS	Chevell Parker	
LRNSAS9	Generating Simple Statistics with Base SAS	Jane Eslinger	
LRNSAS10	A Brief Introduction to SAS® DS2	Mark Jordan	

Learning SAS II (HOW)			
HOW108	Enhancing Your Skillset with SAS® OnDemand for Academics (SODA) Software	Kirk Paul Lafler	The free cloud-based SAS OnDemand for Academics (SODA) software is an exciting development for SAS users and learners around the world! The software includes Base SAS, SAS Studio, SAS/STAT, SAS/GRAPH, SAS/ETS, SAS/OR, SAS/IML, SAS/QC, SAS/CONNECT, SAS Enterprise Miner, and SAS/ACCESS to PC Files. SODA offers users with extensive learning opportunities to enhance skills for career development and advancement using data access, data manipulation, data management, programming techniques, analytics, data visualization, and statistical analysis capabilities. Topics include an introduction and overview of SAS OnDemand for Academics (SODA) software, demonstration of SAS Studio features, and programming examples to showcase this exciting software suite.
HOW110	Essential Programming Techniques Every SAS® User Should Learn	Kirk Paul Lafler	SAS® software boasts countless functions, algorithms, procedures, options, methods, code constructs, and other features to help users automate and deploy solutions for specific tasks and problems, as well as to access, transform, analyze, and manage data. This paper identifies and shares essential SAS programming techniques that the pragmatic user and programmer should learn. Topics include determining the number of by-group levels that exist within classification variables; data manipulation with the family of CAT functions; sorting data; merging / joining multiple tables of data; performing table lookup operations with user-defined formats; creating single-value and value-list macro variables with PROC SQL; examining and processing the contents of value-list macro variables; determining FIRST, LAST, and Between by-group rows; and using fuzzy matching techniques.
HOW158	Processing Complex Data Types in SAS and Python	david horvath	SAS provides versatile tools for ingesting complex data types like XML, JSON, and CSV. Python also provides tools to read these data types. This session will review methods of doing so under SAS and Python via the Jupyter notebook. The basic structures of these files will be reviewed. With available time, additional techniques of reviewing the raw data and understanding that structure will be covered.
HOW167	PARSING: Using SAS® When the Data Are Hiding in a Non-Standard Format	Andrew Kuligowski	Sequential files? Spreadsheets? Databases? There are numerous tutorials that instruct the SAS® user in techniques to extract data from standard sources. Sometimes, however, the desired data is hidden inside a non-standard source; information may be found within the flow of a text document, for example.

This presentation will address some techniques that can be used when not dealing with cleanly formatted data, through use of an example where data are found within a free-form text file. It will deal with identifying what can be considered useful data and what can be discarded, then tackle techniques to extract the data for further analysis, reporting, or whatever is the desired end result. |
HOW205	Map It Out: Using SG Attribute Maps for Precise Control of PROC SGPLOT Output	Josh Horstman	The SGPLOT procedure, part of the ODS Statistical Graphics package, allows for extensive customization of nearly all aspects of plot output. These capabilities are commonly used to distinguish between groups or categories being compared through the use of distinct plot attributes, such as symbols and colors. However, there are times when it is advantageous to be able to associate specific plot attributes with specific data values. SG attribute maps provide functionality that does exactly that. This hands-on workshop will provide an introduction to the use of SG attribute maps in conjunction with PROC SGPLOT. A series of examples will demonstrate how attribute maps are used and why they are useful as a programming tool. Both discrete and range attribute maps will be used to modify a variety of plot attributes, such as plot marker symbols and colors, line styles and fill patterns.
HOW208	Introduction to Data Simulation	Jason Brinkley	Creating synthetic data via simulation can often be a powerful tool for a wide variety of analyses. The purpose of this workshop is to provide a basic overview of simulating data for a variety of purposes. Examples will include power calculations, sensitivity analysis, and exploring nonstandard analyses. The workshop is designed for the mid-level analyst who has basic knowledge of data management, visualizations and basic statistical analyses such as correlations and t-tests.
HOW225	Data Driven Programming for Beginners	Brian Varney	Data Driven Programming for Beginners is intended for SAS programmers that have been using SAS but want to learn more about how to use The SAS Dictionary Tables, SAS Macro, call execute, and more to make their programs more robust. Allowing the data itself to drive a program reduces hardcoding and creates a program that is easier to maintain.

Showcasing SAS

SHO102	Simplified Linux SAS ® Log Comparison: Filtering Differences Into "Useful" and "Useless" Files	Bruce Gilsen	SAS ® log files from two executions of the same program are frequently compared. For example, we might compare today's results to a baseline run we know is correct to test if a change to the SAS release, operating system, external data, our code, or anything else has altered the results. Log file comparisons are often done programmatically, but it's also common to generate a line-by-line comparison with the Linux diff command and eyeball the results. This "eyeball analysis" can be difficult when programs have many steps because the diff output is often cluttered with many records such as step summaries with real and/or CPU time differences and page headings with date/time differences that are useless to most people. The DIFFSPLIT macro runs diff for two SAS log files, parses the output with Perl Regular Expressions in a DATA step, and splits the diff output into two files: one with useful file differences, and the other with useless differences, allowing us to focus on just the differences of interest. The macro in this paper compares files in Linux and treats step summaries with real and/or CPU time differences and page headings with date/time differences as useless. As shown in the paper, it can easily be extended to run on other platforms such as Windows or split the text using different criteria.
SHO124	Q&A with the macro maven: is sql our lingua franca?	Ronald Fehd	SAS software provides an implementation of Structured Query Language (sql). Usage of the sql procedure creates an opportunity for the user to learn database terminology and consider how to address their programming problems and reporting as a matter of the design of the data structure. The purpose of this Q&A dialogue is to review the suitability of sql as a common language for users and programmers. Its benefits include a knowledge of design of data structure for ease of reporting as well as the ability to research the SAS global symbol table. The dialogue is an exposition of how we learn %a language,both natural and artificial languages, and how we use them to communicate.
SHO129	Form(at) or Function? A Celebratory Exploration of Encoding and Symbology	Louise Hadden	The concept of encoding is built into SAS® software in a number of forms, including PROC FORMAT, which transforms values in variables for reporting and to create new variables. Similarly, symbol tables are built into SAS software, so as to communicate with different platforms and systems. This quick tip demonstrates how to PROC FORMAT, and by extension PROC FCMP, to create a system to convert user provided text into Morse Code, and then convert that Morse Code word into sounds, all using SAS. This fun exploration is highly informational about the sort sequence used by SAS software on different platforms, as well as demonstrating the use of PROC FORMAT, PROC FCMP, and sound generation in SAS.
SHO149	PROC IMPORT: Controlling SAS Data Types and Character Variable Lengths When Reading CSV Files	Imelda Go and Abbas Tavakoli	The task at hand is we have a SAS data set and the client needs a CSV file containing all the data in that SAS data set. Validate the CSV file to make sure that it contains exactly what is in the SAS data set. Using PROC IMPORT to read CSV files is convenient. However, relying solely on PROC IMPORT features does not always produce data sets with the desired properties, which can make combining two or more data sets difficult. This paper goes through an example of how you can control or adjust variable types in a data set generated by PROC IMPORT by using GUESSINGROWS=NO, reading as the first record a character string that defines the length of variables, and then reading the CSV file. The technique can be useful when we do not have the source SAS data sets and are given CSV files, which we have to read and then combine into a single SAS data set.
SHO156	Introduction To SCL Functions For Macro Programmers	Ronald Fehd	Another dialect of SAS software is SAS Component Language (scl). This paper shows scl methods for common tasks in jobs and macro definitions as statements and as functions using macro function %sysfunc.
SHO168	Quote the SASLOG(r)	Andrew Kuligowski	I received a "cold call" email from a SAS user in Europe that asked my assistance in tracking down unbalanced quotes in her SAS code. This interactive, audience participation (hopefully) presentation looks at multiple possible solutions to the problem.
SHO189	Data-Driven Interactive Fiction: Designing the DDORK Interpreter and "Risky Ribbons," the Inaugural Base SAS Text Adventure for the SAS OnDemand for	Troy Hughes	Come play an interactive fiction (i.e., text adventure) game that installs you as a SAS conference attendee! Navigate the conference environment and intrigue, attend informational lectures, and collect ribbons for your badge as you hobnob with other attendees, authors! The game can be played in the SAS® OnDemand for Academics (SODA) platform (https://www.sas.com/en_us/software/on-demand-for-academics.html), and requires download and installation, which can take 24 hours. In addition to live, hands-on game play, this session will discuss the data-driven design principles and methods that facilitated creation of the first-ever text adventure game developed using the Base SAS and the SAS macro language!

	Academics (SODA) Platform		
SHO190	Picking Scabs and Digging Scarabs: Refactoring User-Defined Decision Table Interpretation Using the SAS® Hash Object To Maximize Efficiency and Minimize Metaprogramming	Troy Hughes and Louise Hadden	Decision tables allow users to express business rules and other decision rules within tables rather than coded statically as conditional logic statements. In the first author's 2019 book, SAS® Data-Driven Development, he describes how decision tables embody the data independence that data-driven programming requires, and demonstrates a reusable solution that enables decision tables to be interpreted and operationalized through the SAS macro language. In their 2019 white paper Should I wear pants?, the authors demonstrate the reusability of this solution by utilizing the same data structure and underlying code to interpret unrelated business rules. In the current paper, the authors refactor this code by replacing metaprogramming techniques and macro statements with a dynamic hash object that performs the decision table lookup. In prioritizing backward compatibility (of the user-defined data structure), the new interpreter is able to operationalize historic decision tables. This "scab picking"—anecdotally, the refactoring of software to improve performance—yields a more flexible interpreter that is not unnecessarily limited by metaprogramming constraints. Finally, the selection of archaeological dig sites toward the objective of finding treasured scarab artifacts is demonstrated using a decision table.
SHO199	Don't be so One-Dimensional: How to Engineer Multi-Dimensional, High Cardinality Categorical Inputs for Machine Learning	Aleksandar Nikolic	Knowing how to process categorical inputs is a key skill that every data scientist must know. While a wide array of techniques is available to prepare categorical data for machine learning, they do not address all the problems that a data scientist could face. One such problem arises from working with transactional datasets where there could be multiple observations per level of interest. It is possible for a categorical input to not only have a high number of unique values across the entire dataset, but also multiple unique values within the level of interest. Since traditional categorical feature engineering techniques do not address the multi-dimensional aspect of this problem, this paper will outline a novel approach that does. This paper demonstrates the solution using base SAS® code, and SAS® Cloud Analytic Services (CAS) procedures found in the SAS® Viya™ 3.5 platform for machine learning model comparisons.
SHO213	Why Write Base SAS Code When the Macro Processor Can Do It for You	Andrew Walker	As SAS programmers, we need to infuse resiliency in our programs while reducing maintenance time and the likelihood of errors caused by faulty logic or typos. The macro processor is another tool in our toolbox to achieve such goals. This paper demonstrates how to use the macro processor to write dynamic Base SAS® code with two examples: 1) write a dynamic if/else if statement and 2) bin creation when completing a by-group analysis. This paper concludes with details on how to document resolved macros in the log and export the macro to an output file for documentation.
SHO214	Q&A with the macro maven: Do we need Macros? An Essay on the Theory of Application Development	Ronald Fehd	This paper examines the theoretical steps of applications development (ApDev) of routines and subroutines in SAS software. It compares and contrasts the benefits of using the %include statement versus macros. It examines the methods of calling subroutines, e.g., sql, call execute and macro loops.

The purpose of this paper is to highlight the benefits of using macros to support unit and integration testing, and searching for and finding issues during maintenance. |
SHO228	Student SAS Showcase: SAS in the University	Barbara Okerson, Ivan Flores Martinez, Vaibhav Kumar Chaubey, Hong Li, Di Zhu, Joshua Cook, Rafae Abdullah and Sucharitha Vallabhaneni	University and College students that use SAS in their classwork and research will share how SAS is being used in academic institutions throughout the region and how they are personally using SAS.
SHOSAS6	Fun with the SAS Filename Statement	Mark Jordan	
SHOSAS7	Working with APIs at ReST in SAS® 9.4	Mark Jordan	

Statistics, Analytics and Reporting

RPT105	Use of JMP®'s Prediction Profiler to Predict Outcomes of Acute Spinal Cord Injury	Melvin Alexander	For 30 years JMP®'s Prediction Profiler (JPP) has been the cutting-edge, "secret-sauce" tool for finding optimal settings of predictor, regression factors that impact fitted responses from designed experiments and correlated observational data (Jones, 2021). Aarabi, Alexander et al. (2011) used stepwise regression to identify the key input variables from trauma admission, demographic, injury severity measures from Magnetic Resonance Imaging (MRI), and clinical indicators that impacted the response variables of the American Spinal Injury Association (ASIA) motor scores, Functional Independence Measure (FIM), manual dexterity, and dysesthetic pain (e.g., burning, prickling, or aching) experienced after surgery for trauma patients with acute traumatic central cord syndrome (ATCCS) due to spinal stenosis (spinal-cord narrowing that presses on the nerves passing through the spine). This paper uses JPP's Desirability functions and Simulator to gain additional the insights between 10 independent variables on four output response variables that were not considered in the original study. JPP's Desirability functions and Montel Carlo simulation helps to find the most robust data settings that achieves desirable goals for the dependent response variables, especially when some responses have conflicting objectives.
STA134	Using Parallel Analysis to Determine the Dimensionality of Students' Perspectives of Online Caring	Abbas Tavakoli, Kathrine Jones and Phyllis Raynor	There is a significant body of literature about online learning especially during COVID-19 pandemic. Caring is essential in nursing and in online nursing education. There are several nurse researchers who have studied online faculty behaviors that signal caring to nursing students. The instrument Student Perspectives of Online Caring (SPCO) was created by several leading nurse researchers whose research focuses on faculty caring behaviors. This paper used parallel analysis to determine the dimensionality of Perspectives of Faculty Caring among College of Nursing graduate students. Parallel analysis using a SAS macro and option parallel in Proc Factor, was used to determine the dimensionality of the Perspectives of Faculty Caring. Exploratory factor analysis was conducted using maximum likelihood, squared multiple correlations, and Promax rotation. Parallel analysis using a SAS macro and option parallel in Proc Factor, was used to determine the dimensionality of the Perspectives of online Caring. Internal consistency reliability was assessed using Cronbach's alpha. Pearson's correlation assessed the association between factors and subscale items. Parallel analysis showed there are four or five meaningful underlying factors. All 24 items loaded positively on four different factors at 0.30 or above with factor loadings ranging from 0.34 to 0.84, except PERC5 which is not loaded high in any four factors. In addition, PERC23 is loaded in two factors. All 24 items loaded positively on five different factors at 0.30 or above with factor loadings ranging from 0.39 to 0.91. Parallel analysis is a valuable method for determining the dimensionality of the SPCO Scale.
STA144	Assigning agents to districts under multiple constraints using PROC CLP	Stephen Sloan and Kevin Gillette	The Challenge: assigning outbound calling agents in a telemarketing campaign to geographic districts. The districts have a variable number of leads and each agent needs to be assigned entire districts with the total number of leads being as close as possible to a specified number for each of the agents (usually, but not always, an equal number). In addition, there are constraints concerning the distribution of assigned districts across time zones, in order to maximize productivity and availability. Our Solution: uses the SAS/OR ® procedure PROC CLP to formulate the challenge as a constraint satisfaction problem (CSP), since the objective is not necessarily to minimize a cost function, but rather to find a feasible solution to the constraint set. The input consists of the number of agents, the number of districts, the number of leads in each district, the desired number of leads per agent, the amount
RPT146	Using SAS® Data Integration Studio as an effective Data Virtualization Tool	Raj Bhosale	Data Virtualization is a process of merging data from diverse sources/systems to create one or more meaningful datasets that reside in a common location. These datasets can be used towards creating dashboards, which can serve as a critical tool for making data-informed decisions. There are several challenges in this process including, but not limited to, data type differences, frequency of data refresh, handling unstructured data, denormalization and performance overhead. While developing a project that built interactive dashboards in an academia setting, I was able to overcome several of these challenges using SAS® Data Integration Studio. Transformations such as Extract, Join, Append, and LASR Analytic Loader along with SAS® expressions like PUT, INPUT, SUBSTR and CASE helped me in creating clean datasets which were then fed into a SAS® Visual Analytics Dashboard. This paper is geared towards all levels of SAS users (Beginners to Advanced) and will demonstrate how SAS® Data Integration Studio can be used as a powerful Data Virtualization tool and how it can drive SAS® Visual Analytics Dashboards to provide useful information to the end user.
STA169	Embracing Cross-Loading to Improve	Kevin Gittner, Niloofar	A major criticism of estimating traditional latent variable measurement models is the unrealistic assumption of fixed zero cross-loadings with non-primary factors. Latent variable modeling has evolved from these traditional zero cross-loadings to allowing

	Latent Variable Models Fit: A comparison of available options in SAS, Mplus, and R	Ramezani and Katherine Mobley	estimation of non-zero cross-loadings while validating latent constructs of interest. Even though these non-zero cross-loadings are restricted to minimal cross-loadings, they enable researchers to achieve better model fit while allowing a more realistic relationship between items and factors. Both Exploratory Structural Equation Modeling (ESEM) and Bayesian Structural Equation Modeling (BSEM) estimation methods go beyond the traditional model approaches and allow for relaxing restrictive cross-loading assumptions in measurement models. However, different software packages have been slow in adopting these recent developments in latent variable analysis. An extensive literature review of ESEM and BSEM has found estimation availability of both methods only in Mplus and R, with SAS adoption of ESEM, and not BSEM, to date (Muthén & Muthén, 2017; R Core Team, 2022; SAS Institute Inc., 2018). A previous study, which compared the traditional independent cluster model of confirmatory factor analysis (ICM-CFA) approach to ESEM and BSEM, emphasized the improved model estimation with the new flexible ESEM and BSEM approaches in Mplus (Gucciardi & Zyphur, 2016). Using previously collected mental ability test scores data (Holzinger & Swineford, 1939), this study compares the available CFA and ESEM estimation algorithms in SAS to the CFA, ESEM and BSEM estimations algorithms in other software. Results will provide researchers with appropriate application of these methods while evaluating consistency of parameter estimates and model global fit across software packages. Conclusions from this methodological study will highlight how software choice may lead practitioners to making inconsistent decisions about latent variable measurement models. The authors highlight potential areas of growth for SAS PROC CALIS to adopt more recent latent variable modeling options and become more competitive with other software. This could greatly benefit applied researchers and practitioners who use SAS and are interested in adopting more flexible estimation techniques within latent variable modeling and its applications in different fields, especially social science.
STA170	Cutting Edge Regression Methods: Ridge, LASSO, LOESS, and GAM	David Corliss	This session offers a brief introduction to recent advances in regression methods. Techniques demonstrated include ridge regression, LASSO, local polynomial regression (LOESS), and generalized additive models (GAM). The presentation is practical and example driven, emphasizing which procedures to consider and how to apply them in real situations. A quick introduction to each method followed by two worked examples, with discussion of use cases, and options for SAS procedures and producing graphical output.
STA172	Time Series Analysis of School Shootings as Stochastic Terrorism	David Corliss	This study presents a historical analysis of school shooting attacks in the United States. Only attacks using firearms are included: raw data from the US federal CHDS school shooting database was analyzed to exclude suicides, accidents, and threatening actions not resulting in injury. Exploratory data analysis indicates two historical time periods with distinct behaviors. Time series cluster analysis indicates a change in attack behavior about the year 1991: prior to that time, attacks and fatalities were relative few, normally distributed, and show little variation from year to year. Beginning about 1991, a second type of attack pattern is found superimposed on the historical pattern, with both continuing to the present. This second pattern is found to be consistent with stochastic terrorism and is characterized by a highly variable annual attack rate, marked increases in deaths and lethality (deaths per incident), and a skewed distribution with a risk of a high number of fatalities driven by a small number of extreme events. A forecast for the second half of the 2021-2022 school year is presented and analyzed.
STA178	A SAS Macro to Calculate Blinding Index in Clinical Trials: %blinding_index , an application of PROC IML	Jingwei Wu and Di Zhu	Randomized clinical trials are often considered as gold standard for clinical research, owing the fact of its rigorous study design and implementation. Success of blinding, i.e., study participants and key layperson are unaware of the treatment assignment or therapeutic allocation they received, is key component to minimize post-randomization bias. There are two commonly used methods that can quantify blinding success in a double-blinded randomized control trial based on administered post-randomization questionnaire data, namely, James Blinding Index (James_BI) and Bang Blinding Index (Bang_BI). James_BI is a scaled number between 0 and 1, increases as the success of blinding increases. Bang_BI is calculated per treatment arm, is a scaled number between -1 to 1, with 0 as the most desirable situation under successful blinding. There is no current SAS macro program that can calculate the two blinding indexes. To fit this gap, we have written a SAS macro: %blinding_index to provide researchers a computational tool to calculate these blinding indexes. We used PROC IML to calculate the indexes, along with the corresponding 95% confidence intervals for statistical inferences. Researchers may thus use it to assess successfulness of blinding in clinical trials. This presentation will review the two methods for the assessment of blinding, and demonstrate the developed macro with a trial application.
STA182	Cliometrics: An Underused Quantitative Approach to History	Barbara Okerson	The study of history emphasizing statistical or quantitative methods is known as cliometrics. Originally applied to the study of economic history, cliometrics now includes the full range of quantitative historic studies. These quantitative studies are used when identifying and studying varied phenomena of the past including wartime patterns, literacy levels, agriculture, and consumerism. Comparing across time and space, we can demonstrate correlations, patterns, and trends that provide insight into cause and

			consequence. This paper will begin with a review of cliometrics as a field of study and discuss and illustrate some of its applications and relevance.
STA185	A SAS Macro for Implementing a Nonparametric Control Charting Scheme for Individual Observations	Austin Brown	In process monitoring applications, a commonly encountered scenario is one in which several presumably identical processes are desired to be monitored simultaneously. Such a process is referred to as a "Multiple Stream Process (MSP)." Several parametric control charts have been developed for monitoring MSPs with the majority being derived from the classical Shewhart X-bar chart. These control charts perform best when the assumption of normality is met, but their performance can deteriorate substantially when the underlying data are non-normal. This deterioration is magnified when the sample size is n=1. While some work has been done to address the issue of non-normality through the development of a nonparametric MSP chart, a review of the literature shows an apparent gap for nonparametric MSP charts where the sample size is one. Thus, this paper introduces a novel method to address this gap, which is based on the classical nonparametric test, Cochran's Q Test and utilizes the EWMA framework, as well as a SAS macro for implementing the new technique. Derivation of the charting statistic and control limits will be given as well as recommendations for selection of chart parameters for a desired average run length performance in addition to an applied example of usage.
STA186	A SAS Macro That Automates Model Fitting of Group-Based Trajectory Modeling Using Proc TRAJ	Donald Warden and Yu Jiang	The third-party SAS procedure TRAJ developed by Jones et.al gives users a convenient tool for group-based trajectory modeling. The procedure allows the user to fit any number of classes across five polynomial orders that each class follows. To fit these models, users must assess model fit by comparing the Bayesian Information Criteria (BIC) of the models with different polynomial orders. Moreover, the parameter estimate for each class polynomial must be statistically significant – generally presumed as a T test alpha less than 0.05. With each increase in the number of classes, the number of models to check increases exponentially. Traditional methodology employs bootstrapping to generate valid, parsimonious models without checking every model. However, this methodology likely does not find the best fitting model. In this paper, we develop a SAS macro, autoTRAJ, which assesses all possible permutations of polynomial order and outputs a list with the best fitting model based on BIC up to eight classes. Further, the macro checks for common statistical abnormalities such as false or singular convergences. The macro utilizes latent output files to determine the statistical significance of the polynomial T statistics and only shows users statistically significant models. By automating the traditional methodology, the programmer will no longer exclude better fitting models due to time constraints. This allows for more precise trajectory modeling with no uncertainty that the selected model fits best. The intended audience for this presentation is intermediate SAS users with knowledge of Base and IML SAS and the TRAJ procedure.
RPT194	Crime Rates in Atlanta, Georgia relative to Covid-19	Benoit Bernadel and Joe DeMaio	Atlanta offers many extraordinary cultural experiences along with museums, dining, shopping, and entertainment. Unfortunately, criminal activity follows along as a companion to any city with so much to offer. And so, Atlanta is no different in this respect to other world class cities. In 2021, the violent crime rate in Atlanta was 697 per 100,000 people. The Atlanta Police Department makes available raw reported crime data from 2009. Said data is broken down across six zones within the city boundaries of Atlanta. Using SAS Base programming we determine which times of day and which days it is safer to visit different zones in Atlanta. How does this data compare to pre-Covid 19 rates? March 2020 saw Atlanta and the rest of the world shut down due to Covid-19. June 12, 2020, saw the death of Rayshard Brooks in the parking lot of a south Atlanta Wendy's restaurant. Using data from January 2009 to February 2020, we construct a model to predict frequency of reported crime in Atlanta. Using this predictive model, we consider if crime rates changed in March 2020 and/or on June 12, 2020, in Atlanta. The content of this paper is accessible to all levels of SAS users.
STA198	Invoking Survey Weights to Calculate Bayes' Factor with PROC MCMC to Generate More Generalizable Inferences	Tyler Hicks and Graham Rifenbark	In many instances, such as the one considered in this paper, analysts use survey weights to enhance the representativeness of samples. As an example, pollsters use survey weights to better forecast elections with cheaper samples. Naturally, survey weights are relevant when generalizing statistical tests to populations. To navigate between contradictory models, Bayesian statisticians turn to Bayes' Factors. The Bayes Factor (BF) quantifies the extent to which data support one data better than another. As an example, BF=5 indicates data support the alternate give times better than the null. Although adjusting the BF with sampling weights is not small feat, Markov Chain Monte Carlo (MCMC) algorithms enables their successful empirical derivation. Using an example-based approach, this pedagogical paper shows how even SAS users without extensive technical expertise in Bayesian methods and survey weights can fit a two-stage hurdle model to recover BFs from data that adjust for survey weights using in-built functions in the SAS/STAT® MCMC

			procedure. This paper thus helps SAS users take the extra step of marshaling the full power of survey weights in Bayesian testing with PROC MCMC.
RPT204	Using the Output Delivery System to Create and Customize Excel Workbooks	Josh Horstman	In years past, SAS® output was limited to the text-based SAS listing. However, the Output Delivery System (ODS) greatly enhanced the capabilities of the SAS system by allowing users to create highly-customizable output in a variety of document formats, including Microsoft Excel® workbooks. This paper provides a brief overview of how to use the ODS EXCEL destination to create excel workbooks and how to customize the various visual attributes of the output such as fonts, colors, styles, and much more.
STA222	Application of Feature Selection and Dimension Reduction Techniques on Large-Scale CT Dataset for Lung Cancer Diagnosis Based on Radiomics	Mostafa Zahed and Maryam Skafyan	A German physicist Wilhelm Konrad Rontgen started to work on medical imaging in 1895. Later, in 1978, Hounsfield developed computed tomography (CT) technology to study medical images to screen and detect tumors in the fastest way. Because of the importance of searching medical images, Lambin 2012 proposed an approach to analyzing medical images, which is called radiomics. This approach has involved several steps from the beginning to the end: segmentation, feature extraction, feature selection, and statistical modeling. Extracted features can be categorized in the description of tumor gray histograms, shape, texture features, and the tumor location and surrounding tissue. Because of the massive number of features from radiological images, machine learning plays an important role in analyzing the big data obtained from tumors. In other words, the dimension of extracted features needs to be reduced to describe the tumor better. Many linear and nonlinear dimension reduction techniques, including Principal Component Analysis (PCA), Linear Discriminant Analysis (LDS), Local Feature Analysis (LFA), and manifold learning have been developed. In this paper, a large-scale CT dataset for Lung cancer diagnosis (Lung- PET-CT-Dx) which was collected by Huiping Han, Funing Yang, and Rui Wang of Harbin from Medical University in Harbin in China is used to illustrate the dimension reduction techniques, which is a main part of radiomics process, via SAS. This dataset consists of CT DICOM images of lung cancer subjects with XML Annotation files that indicate tumor location with bounding boxes. Pyradiomics through 3D Slicer medical software was used to extract features for 74 patients out of 130 as the provided annotation file did not work for all patients through Python. This study has been shown how to apply the dimension reduction methods available in SAS to the lung cancer dataset, including Principal Component Analysis (PCA) and Cluster Analysis through PROC FACTOR, PROC PRINCOMP, and TEXTMINE Procedure. Both approaches suggested if the data were categorized into six subcategories, the reduced data would adhere to the dominant variation in the original data.
RPT227	Custom Panel Graphs Using PROC TEMPLATE	Jim Blum	PROC SGPANEL gives a variety of styles of panel graphs, but they have significant limitations in the scope of layouts available. Layouts have equal cell sizes, structured in some form of a grid. They also use the same plotting statement(s) for each of the cells, distinguished only by the variables used in the PANELBY statement. PROC TEMPLATE can construct a wide variety of panel graphs, including those with graphs of different sizes, types, and variable sets, This paper gives a brief overview of the PROC TEMPLATE tools that mimic the PROC SGPANEL structures, and then focuses on structures that cannot be built in PROC SGPANEL. Example output and code for graphs built from common SASHELP data sets are provided.
STA230	Using SAS Macro and ODS Output to efficiently examine the descriptive and analytic statistics in epidemiology studies	Yue Pan	The key feature of descriptive and analytic epidemiology is a comparison group, i.e. the exposure, and how it is associated with the study interest, i.e. the outcome. Researchers and investigators are required to present the descriptive and analytic statistics of both exposure and outcomes and test their relationship to describe their study sample. However, if there are many exposures and outcomes to examine, the syntax usually become repetitive and hard to navigate and identify useful results. This paper will present how to use SAS macro and ODS output to efficiently examine and generate descriptive and analytic statistics for epidemiology studies.

This presentation is aimed at beginner to intermediate SAS programmers and healthcare analysts who already have a basic understanding of SAS Macro and ODS Output and are looking to efficiently examine their data. |
| RPT232 | Excel with ODS Excel Destination and Proc Report! | Devi Sekar | ODS EXCEL destination for creating Microsoft Excel workbooks is available starting 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. This paper covers various techniques that you can use with PROC REPORT and the ODS EXCEL destination, to bring excellence to your work! |
| STASAS8 | Bayesian Mixed Models | Danny Modlin | |

	Super Demos		
SDEM1	Converting Character, Numeric and Date Variables	Jane Eslinger	We all convert character to numeric and numeric to character, and then there are always those pesky date variables. Let's look at conversion examples.
SDEM2	Creating Excel Pivot Tables with SAS	Chevell Parker	Generate styled Pivot tables directly from SAS capable of generating new summarized views with a spin of the data
SDEM3	Using the SAS Extension for Visual Studio Code	Chris Hemedinger	The SAS programming language has arrived in VS Code, the new favorite coding environment for programmers around the world. Learn how to get started!
SDEM4	Bayesian Generalized Linear Mixed Models	Danny Modlin	This presentation will feature the BGLIMM procedure available in SAS/STAT. This will allow the participant to model non-normal responses and include random effects within their Bayesian approach.
SDEM5	The SAS sorting hat & Serpentine Sort	Charu Sankar	Sorting in SAS is an expensive process in terms of both time and resources consumed. In this session, prepare to explore some of the common and lesser known sorts that SAS provides. Become like the sorting hat in Harry Potter!
SDEM6	Introducing SAS Analytics Pro on SAS Viya	Chris Hemedinger	SAS Viya on your desktop – in just a few minutes! Learn about this new way to use SAS from a Docker container. You can write and run code, run batch jobs, connect to databases, use Python and more!
SDEM7	Jedi SAS Tricks for SAS® DS2 Programmers	Mark Jordan	This presentation demonstrates some unique coding benefits of DS2, including: deferred array dimensioning, formatting all array elements at once, using FIRST. / LAST. without pre-sorting, Using matrix math to resolve complex problems, and writing output to CSV files using ODS.