

DCSUG Newsletter

Washington, DC SAS Users Group

Fourth Quarter Meeting

When: Tuesday, December 2, 2003
Time: 8:30a.m. to noon
Place: Bureau of Labor Statistics
Postal Square Building
Room G440
First Street North East
Washington DC



Agenda:	8:30 - 9:00	Registration and refreshments
	9:00 - 9:15	Announcements
	9:15 - 10:15	Robert N. Rodriguez, SAS Institute Statistical Graphics Using ODS in SAS 9.1
	10:15 - 10:30	Break
	10:30 - 11:30	Andrew Karp, Sierra Information Services, Inc. A Peek at PROC DOCUMENT
	11:30 - 11:45	User-to User Questions and Answers Moderators - DCSUG officers
	11:45 - Noon	Giveaway Drawings

***Presentation times subject to change**

Visitors must use the visitor's entrance on First Street NE. Directions for those taking Metro: Take the Red Line to Union Station. Exit the station via the Amtrak exit, but do not go up to the train level. Instead, after going up to and out through the farecard gates, turn to the left where there is an exit to the street (First Street NE). Directly across from the exit are the employee and visitor's entrances to the building. After checking in with security, take the main elevators down to the G level. **BRING PICTURE ID!**



If you have special needs and are planning to attend this meeting, please contact any DCSUG officer no later than 11/17/03

SAS® is a registered trademark of the SAS Institute, Cary, NC

NOTE: If you are not currently a paid member of the DCSUG and are planning to attend the meeting, please email Paul Gorrell at pgorrell@s-3.com or call 301-628-3237 by Nov 17th and leave your name. Security at the Bureau of Labor Statistics is very tight and having your name on a list of possible attendees will make your entrance to the building easier.

President's Message

Peter Knapp

As many of you know, this year's NESUG was held in Arlington, VA. DCSUG's very own Larry Altmayer and Ralph Leighton of The Hartford chaired a great conference. I'd like to thank them and all the other volunteers who worked on the conference and insured it was a success! NESUG, DCSUG, and all the other SAS user's groups rely on many people volunteering their time and energy. It brings me great personal satisfaction to be part of a community that works together to create and sustain vibrant user groups like DCSUG and NESUG.

I hope you will be able to attend the December 2 DCSUG meeting. The SAS Institute is lending us Bob Rodriguez to talk about ODS Stat Graphics. Andrew Karp will also be joining us to talk about Proc Document. I look forward to learning from both of them. DCSUG thrives in part because people are willing to share their experiences with SAS and because so many of you turn out for our quarterly meetings. Keep up the good work.

Speaker Abstracts

Statistical Graphics Using ODS in SAS 9.1

Robert N. Rodriguez, SAS Institute

Effective graphical displays are essential in modern data analysis, statistical modeling, and data mining.

Traditionally, SAS users have produced statistical graphics by modifying output data sets from statistical procedures and using generic graphics procedures. This forces the user to reconstruct analytical context that is available within the statistical procedure.

In SAS 9.1, over two dozen SAS/STAT and SAS/ETS procedures have been modified to use an experimental extension to the Output Delivery System (ODS) that enables them to create statistical graphics as automatically as tables. This extension, referred to as ODS for Statistical Graphics (ODS Graphics for short), provides commonly used displays, including scatter plots, histograms, box-and-whisker plots, and contour plots, in ODS output. Many ODS features for tables, such as destination statements, apply equally to graphics. ODS styles control the appearance and consistency of all graphs, whereas ODS templates control the layout and details of individual graphs.

This presentation introduces statistical users and other SAS programmers to ODS Graphics. A series of examples will be used to explain the basic functionality, which requires minimal syntax, and common tasks for graph management. Some familiarity with ODS for tables will be assumed.

Bob Rodriguez joined SAS in 1983 and is currently a director in SAS R&D with responsibility for development of SAS/STAT and SAS/QC. He received his Ph.D. in statistics in 1977 from the University of North Carolina. He is editor of the ASA-SIAM Book Series, chair of the ASA Section on Physical and Engineering Sciences, and serves on the editorial board of The American Statistician.

A Peek at PROC DOCUMENT

Andrew Karp, Sierra Information Services, Inc

A major enhancement to the SAS Output Delivery System (ODS) in SAS 9 Software is implementation of the DOCUMENT Destination, which allows SAS users to store ODS-generated objects in a reuseable file structure called a "document." ODS DOCUMENTS can be accessed, re-played, and manipulated using the new DOCUMENT Window or PROC DOCUMENT, an interactive BASE SAS Procedure which is also new to the SAS System in Version 9. Learning to use the DOCUMENT Destination, DOCUMENT Window and DOCUMENT Procedure will give you enhanced control and flexibility when using ODS to generate output.

Andrew Karp is President of Sierra Information Services, Inc., a SAS Software consulting and training firm located in California wine country. A 20-year SAS Software user, Andrew is a SAS Certified Professional for Version 8 of the SAS System and has given dozens of papers at SAS users group meetings around the world. A graduate of The George Washington University and former Dupont Circle area resident, Andrew regularly returns to DC and has spoken at several prior DCSUG meetings. (See page 7 for other presentations by Andrew Karp in the DC area)

DCSUG 2003 Steering Committee Officers

CHAIR -- Peter Knapp

(202) 482-1359 (202) 482-1388(fax)
peter_knapp@ita.doc.gov

SECRETARY -- Paul Gorrell

(301) 628-3237 (301) 628-3201 (fax)
pgorrell@s-3.com

TREASURER -- Arlene Siller

(301) 458-4498 (301) 458-4032 (fax)
asiller@cdc.gov

SAS LIAISON -- Mike Rhoads

(301) 251-4308 (301) 294-2040 (fax)
rhoadsm1@westat.com

OTHER COMMITTEE MEMBERS

Larry Altmayer

(301) 763-2569 (301) 457-4437 (fax)
Lawrence.W.Altmayer@census.gov

Linda Atkinson

(202) 694-5046 (202) 694-5718 (fax)
Atkinson@ers.usda.gov

Karen Dennis

kcdennis@comcast.net

Frank Fry

(202) 452-2666 (202) 452-6433 (fax)
frank.fry@frb.gov

Roger Goodwin

(301) 763-4774 (301) 457-2744 (fax)
roger.l.goodwin@census.gov

Howard Schreier

(202) 482-4180 (202) 482-4614 (fax)
Howard_Schreier@ita.doc.gov

Christopher Zogby

(703) 671-6330 (703) 671-6360 (fax)
czogby@erols.com

DESKTOP SPECIAL INTEREST GROUP

CO-CHAIR -- Joe Mirabal

(703) 265-4620
jmmirabal@aol.com

CO-CHAIR -- Yesvy Gustasp

(301) 625-9576 (202) 307-5888 (fax)
Yesvy_Gustasp@yahoo.com

SAS Talk

by Ian Whitlock

A colleague passed around an interesting question from his database class. Given a file with NAME (of student) and CLASS (taken by student), make a dataset with all pairs of names, say WHOTOMATCH and FOUND, where FOUND is taking all the classes taken by WHOTOMATCH. For example, let's use numbers for names and for classes, so that data can be easily generated.

NAME	Class
10	1
10	2
20	2
20	3
30	1
30	2
30	3

The result is

WHOTOMATCH	FOUND
10	30
20	30

For this data 30 is taking all of 10's classes and 30 is also taking all of 20's classes. No other names appear in the lists since there are no students taking all the classes that 30 is taking.

The job was to be done with SQL. I found that restriction good, because I thought it the best language for expressing the solution. At first I wanted to use a reflexive left join on matching classes, but found this way difficult to turn into code. I then turned to counting, and said that 30 can be known to taking all 10's classes because 10 and 30 have 2 classes in common and 10 is only taking 2 classes. Thus I needed a count of classes for each student and a count of common classes for each pair of students. Here is a solution based on this idea.

```
proc sql ;
  create table q2 as
  select distinct a.name as whotomatch , b.name as found
  from ( select * , count(*) as count
        from people
        group by name ) as a
  , people as b
  where a.name ^= b.name
  and a.class = b.class
  group by a.name , b.name
  having count (*) = a.count
  order by a.name , b.name
  ;
quit ;
```

The subquery calls for remerging the count back on to the original file and will produce a note on the log to that effect. Many SQL's make this illegal, but fortunately SAS does not because it something we often want in SAS.

Another colleague came up with the following solution.

```
proc sql;
  create table q1 as
  select distinct p1.name as whotomatch
    , p2.name as found
  from people as p1 left join people as p2
  on p1.name ne p2.name and p1.class = p2.class
  where not exists ((select class
                    from people
                    where p1.name = name)
  except
  (select class
   from people
   where p2.name = name))
  order by p1.name , p2.name
  ;
quit;
```

This solution seems more intuitive in that it is not based directly on counting. The subquery takes all of the classes for the first student and removes any class that is also taken by the second student. Now the WHERE clause says we want the pair of students when the remaining set of first student's classes is empty. This solution shows the use of set theoretic constructions EXCEPT and EXISTS in SQL. The use of sets in SQL is often misunderstood and therefore avoided.

However, the second solution is based on two correlated subqueries since P1.NAME and P2.NAME refer to names in the outer query, i.e. the inner name must be correlated with an outer name. This means that the subqueries depend on which pair of names one is looking at. Although the subqueries do not appear to be nested they act like nested DO-loops since both outer names must be determined before the subqueries can be executed. Consequently the subqueries must be performed many times. In contrast the first solution contains a subquery which does not depend on the out query. Hence it can be done once before the outer query.

In general, in SAS, correlated subqueries can be very time consuming. To illustrate the point let's generate 1000 students taking up to 5 of 100 classes.

```
data people ;
  do name = 1 to 1000 ;
    do i = 1 to ceil ( 5 * ranuni ( 560957652 ) ) ;
      class = ceil ( 100 * ranuni ( 560957652 ) ) ;
      output ;
    end ;
  end ;
run ;
```

Both solutions produce the same final dataset, however the first took under 2 seconds while the correlated subquery solution took over 4 minutes. Why? With a thousand students there are a million pairs (999,000 when the same name pairs are eliminated). This means that the first solution did one subquery while the second did the two subqueries almost a million times. Moreover, the number of times the subqueries must be executed and consequently the amount of execution time increases like the square of the number of students.

Calendar of Meetings and Events

This feature is meant to give users an idea of some of the activities of area users groups and special interest groups, as well as regional and international groups. If your group would like to be regularly included in this feature, please call Mike Rhoads or one of the DCSUG officers. There is a DCSUG directory with phone, fax, and E-mail contact information elsewhere in this newsletter.

November

12 **DCSUG** Desktop SIG Meeting
Time: 12:30 - 1:30 pm
Site: Room S3031
1800 M St. NW
Program: dfPower DBMS/Copy as a SAS
Workalike
Contact: Charlie Hallahan at (202) 694-5051
hallahan@ers.usda.gov

Other Groups

CENSUG Census SAS Users Group
Contact: Andrea Potter at (301) 763-1687

MDSUG Maryland SAS Users Group
Contact: Mr. R.H. Miller, (410) 740-4229
millerrh@juno.com
www.umbc.edu/chpdm/mdsug/

December

2 **DCSUG** DC SAS Users Group, General Meeting
Time: 8:30 am - 12:00 noon
Site: Bureau of Labor Statistics
2 Massachusetts Ave. NE
Program: See front page announcement
Contacts: Mike Rhoads at (301) 251-4308
RhoadsM1@Westat.com
Peter Knapp at (202) 482-1359
peter_knapp@ita.doc.gov

NCHSSUG National Center for Health Statistics SAS Users Group
Contacts: Arlene Siller at (301) 458-4498
asiller@cdc.gov
Linda Tompkins at (301) 458-4533

NIHSUG National Institutes of Health SAS Users Group
Contact: Ray Danner at (301) 496-6037

NOVASUG Northern Virginia SAS Users Group
Contact: Gary McQuown at (703) 628-5681
gmcquown@dasconsultants.com
www.novasug.org

March 2004

2 **DCSUG** DC SAS Users Group, General Meeting
Time: 8:30 am - 12:00 noon
Site: TBA
Program: TBA

SIGSTAT Special Interest Group in Statistics, CPCUG and WINFORMS
Contact: Charlie Hallahan, 202-694-5051
hallahan@ers.usda.gov

May 2004

9-12 **SUGI 29** SAS Users Group International Conference
Site: Montréal (Palais des congrès de Montréal)
Chair: Duke Owen
Further Info: support.sas.com/usergroups/sugi/sugi29/

VASUG Virginia SAS Users Group (Richmond area)
Contact: Carrie Mariner at
president@vasug.org
www.vasug.org

DCSUG Thanks Its Sponsors

Westat

An Employee-Owned Research Corporation

Social & Scientific Systems, Inc.

An Employee-Owned Company

The Bureau of Labor Statistics

These organizations make it possible for DCSUG to host quarterly meetings featuring informative presentations about SAS software while keeping dues affordable.

Thanks

Election of 2004 DCSUG Officers



Peter Knapp -- Candidate for Chair

Peter Knapp is a Computer Specialist at the International Trade Administration at the Department of Commerce. For thirteen years Mr. Knapp has supported an office of 70 trade analysts who use SAS to administer international trade law. He learned SAS while working as an intern at the Bureau of Labor Statistics where he worked on the Consumer Price Index. Mr Knapp has presented papers at both NESUG conferences and DCSUG meetings. He is the current president of the DC SAS Users Group.

Arlene Siller -- Candidate for Treasurer

Arlene Siller is a Computer Specialist with the National Center for Health Statistics and has used SAS for 24 years. Her responsibilities include programming and analysis using SAS/Base and SAS/Stat for National health surveys. She is the SAS Site Representative for her agency and co-chair of the in-house NCHS Users Group (NCHSSUG). She is very involved in providing support for NCHS SAS users, and

enjoys this personal interaction part of her responsibilities the most. She has programmed using the TSO, Wylbur, VMS, DOS, Windows and UNIX operating systems.

Paul Gorrell - - Candidate for Secretary

Paul Gorrell is a Programming Manager at Social & Scientific Systems, Inc. He has over twenty years of SAS experience working with different types of data on a variety of platforms. These days he works mostly with health care survey data. He has been a presenter and section co-chair at NESUG and, for the past year, has served as DCSUG Secretary.

Roger Goodwin -- Candidate for SAS Liaison

Roger Goodwin started working at the Census Bureau in 1999. At the Bureau, Mr. Goodwin currently works in the Decennial Directorate as a Mathematical Statistician. He received his BS in Computer Science in 1988 from Old Dominion University and an MS in Applied Statistics in 1998 from Old Dominion University. Mr. Goodwin has been a SAS user since grad school, and is currently an unofficial steering committee member of DC SAS User's Group. He has taught computer software courses at Tidewater Tech in Chesapeake VA, and has taught a math course at Prince George's Community College in MD.

DCSUG 2004 BALLOT

Place an "X" In the Box beside the candidate of your choice. You may write in the name of someone who has not been nominated if you wish.

DCSUG Officers

Chairperson

Description: Lead DCSUG, delegate assignments, chair group meetings, preside over nomination and election of officers.

Peter Knapp

Treasurer

Description: Manage the group finances: collect dues and fees, make disbursements, maintain record of income and disbursements. The Treasurer will also prepare a report of income and disbursements for distribution at the fourth quarter general meeting.

Arlene Siller

Secretary

Description: Record minutes of group meetings, maintain membership list, prepare meeting notices.

Paul Gorrell

SAS Liaison

Description: Represent and increase the influence of DCSUG to the SAS Institute

Roger Goodwin

DCSUG Dues

Dues are set annually by the steering committee and voted on by the membership at the fourth quarter meeting. Dues will be payable by the first quarter general meetings. Dues will be used to finance mailings, telephone charges, newsletter and other expenses as approved by the group of officers.

Keep proposed (current) rate (\$10.00 per year individual, \$50.00 per year group)
 Increase Decrease

Please mail ballot to:
Washington DC SAS Users Group
P.O. Box 44670

Washington, DC 20026-4670

SAS(r) Training Seminars in the DC Metro Area: Jan & Feb 2004

Sierra Information Services, Inc. is presenting a series of SAS and statistical seminars in the Washington, DC area during Jan and Feb 2004. The seminars address a wide range of SAS Software and data analytic topics, and are given by a team of well-known experts. They will be held at the UVA/VA Tech Northern VA Center, a short walk from the West Falls Church Metrorail station. For more details, please visit www.SierraInformation.com

Date	Instructor	Seminar
Tuesday, January 06, 2004	Andrew Karp	SAS 101: An Intro. to the SAS System
Wednesday, January 07, 2004	Paul D. Allison, Ph.D.	Treatment of Missing Data (Day 1)
Thursday, January 08, 2004	Paul D. Allison, Ph.D.	Treatment of Missing Data (Day 2)
Friday, January 09, 2004	Andrew Karp	ODS for Data Analysts and Statisticians
Monday, January 12, 2004	David Cassell	Sampling with the SAS System (Day 1)
Monday, January 12, 2004	Ben Cochran	Intro to SAS/SQL Programming
Tuesday, January 13, 2004	David Cassell	Sampling with the SAS System (Day 2)
Tuesday, January 13, 2004	Ben Cochran	Advanced SAS/SQL Programming
Wednesday, January 14, 2004	Ben Cochran	Manipulating Data with SAS Functions & Arrays
Wednesday, February 04, 2004	Andrew Karp	SAS 101: An Intro. to the SAS System
Thursday, February 05, 2004	Andrew Karp	Summarizing and Reporting Data Using the SAS System
Friday, February 06, 2004	Andrew Karp	Advanced Reporting with the SAS Output Delivery System
Monday, February 09, 2004	Art Carpenter	From %MACRO to %MEND: An Intro. to the SAS Macro Language
Monday, February 09, 2004	Andrew Karp	Building and Applying Predictive Models Using the SAS System
Tuesday, February 10, 2004	Art Carpenter	Advanced Topics in the SAS Macro language
Wednesday, February 11, 2004	Andrew Karp	SAS System Programming Efficiencies: Tips & Techniques
Wednesday, February 11, 2004	Art Carpenter	Building Dynamic Applications with the SAS Macro Language
Thursday, February 12, 2004	Art Carpenter	Intro to SAS/STAT
Thursday, February 12, 2004	Andrew Karp	ODS for Data Analysts & Statisticians

DCSUG -Desktop SIG meeting - jointly with SIGSTAT

Wednesday, November 12th from 12:30 to 1:30 pm

1800 M St., NW.

Room S3031 (Enter the South Tower & take the elevator to the 3rd floor to check in at the guard's desk)

Contact Charlie Hallahan at: HALLAHAN@ers.usda.gov to ensure quick entry.

For directions: <http://www.cpcug.org/user/sigstat/directns.htm>

Code Portability: SAS Data step in DBMS/Copy - Using it as a SAS Work-Alike

by Yesvy Gustasp

Many SAS(r) users are familiar with Conceptual Software's DBMS/COPY, a widely used utility for transferring data between software packages. Conceptual's product lines have been acquired by DataFlux [a subsidiary of SAS Institute]. DBMS/Copy includes DBMS/Analyst, a database manipulation and programming system. The modules of DBMS/Analyst provide functionality resembling that of base SAS software (including the DATA step). The presentation will look at the feasibility of deploying DBMS/Analyst as a complement to SAS, including interoperability issues and the porting of applications. The emphasis will be on code (SAS DATA step programs as well as statements invoking the SAS SORT, MEANS, TABULATE and other summarization procedures).

Users can submit code they would like to be translated [along with a miniature version of the data set ; provided it is good for illustration in the presentation, to Yesvy.Gustasp@Yahoo.com



The drive to succeed. It never ends.

Broad and dynamic challenges that develop new skills. Working with a company that does more than evolve...it sets standards. These are your career goals. ICON Clinical Research is your company.

At ICON Clinical Research, we see and do things differently. Dedicated to providing the pharmaceutical and biotechnology industries with exceptional clinical research and biometrics services, we're the first CRO to be awarded ISO 9002 quality certification across all functions. Join us and build a remarkable career... one that will set a new criterion for success.

To continue our legacy of excellence at ICON, GloboMax, a dynamic and innovative Product Development Organization committed to helping pharmaceutical and biotech companies bring products through development from early stage to regulatory approval, is now the Strategic Pharmaceutical Development Division of ICON.

Senior SAS Programmer

Hanover, MD

GloboMax - Strategic Pharmaceutical Development Division of ICON.

The successful candidate will lead the programming effort on various projects. Write programs in support of analyses to be performed by pharmacokineticists. Write programs to generate listing and summary tables for inclusion in clinical study reports, as described in the applicable Analysis Plan. Provide all necessary documentation of all programs. Draft, finalize, and implement programming guidelines and SOPs. Provide technical resource for SAS (e.g. SAS upgrades, licenses, etc.). Work under the direction and supervision of Manager, Clinical Data Management. Provide SAS training to GloboMax personnel as necessary. Balance concurrent tasks/priorities within defined timelines.

Minimum of a Bachelor's degree with at least 7 years' SAS experience and at least 7 years' pharmaceutical, biotechnology, or health-related industry experience. Strong analytical, organizational, and communication skills is a must.

TO APPLY and for more information on our opportunities and locations, please visit us at www.iconclinical.com No agencies, please.

Come to ICON and be an important part of a well-respected and growing company. You'll enjoy highly competitive benefits in addition to relocation assistance, tuition reimbursement, credit union, health club privileges, and many other exceptional advantages.

ICON. We are the right solution.

ICON Clinical Research is an Equal Opportunity Employer committed to strength in diversity.

www.iconclinical.com

• Consulting for Healthcare, Banking & Financial Services, Insurance, Power & Energy, Manufacturing, Retail, •

Certified Consultants • GSA Schedule • SAS® Solutions: Supplier Relationship Management, Strategic Performance Management, Financial Management

Pharmaceutical and Government • HealthView • WebView Services • Member of the American Association of Health Plans • SAS Alliance Gold Member

Results, not Promises

Qualex
Consulting Services, Inc.

Aligning with SAS® Intelligence Solutions



www.qlx.com for more information
877-887-4SAS (4727)

• Training • Free proof of concept • Free SAS tools: Cataloger, Exporter, Find2000, Zipcoder, Datepick, Helpman •

2004 Meetings

Remember to Join DCSUG!!!

Plan ahead and mark your calendars now. The DCSUG meetings in 2004 are TENTATIVELY scheduled for March 2, June 8, Sept 14, and Dec 7. Meeting times and locations to be announced. If at any time you have a question about an upcoming meeting or other DCSUG information, call one of the members of the Steering Committee.

~~You're~~
~~Invited~~



To visit the DCSUG home page

The URL for the DCSUG home page on the Web:

<http://dc-sug.org>

Don't forget to renew your DCSUG membership for 2004. Dues are used to defray the costs of producing and mailing the newsletter and provide refreshments at our meetings. You'll make sure that you keep receiving the newsletter and keep DCSUG going. If you have not joined DCSUG, now is a great time to do so. Joining DCSUG is as easy as completing the membership form included in this newsletter. Individual memberships are only \$10; group memberships are \$50



Membership Application/Renewal

Date: _____

Name: _____

Company Name: _____

Mailing Address: _____

City, _____

State _____ Zip _____

Phone # _____

Fax # _____

E-Mail: _____

Preferred mode of contact (check one):

Fax

E-mail

SAS Version(s) _____

SAS Product(s): _____

Operating System(s) _____

Check one: _____

Individual membership

for 1 year is \$10.00

Group membership

for 1 year is \$50.00

Please indicate a contact person for Group Membership

Check one:

New membership

Renewal membership

Mail to:

Washington DC SAS Users Group
P.O. Box 44670
Washington, DC 20026-4670