

LA&S 792, Managing Research Data in the Social Sciences

1 to 3 credit hour course - Summer 2018

**June 5 – June 15** (if taken for 1 Credit),  
**June 5 – June 25** (if taken for 2 Credits),  
**June 5 – June 29** (if taken for 3 Credits)

1:30-3:30pm 413 Summerfield

Line Number: 81108  
Instructor: Larry Hoyle



**Objectives:**

This course is designed to teach the skills necessary to read, reorganize, transform, display, clean, archive, and export simple or complex data. The primary tool used in the class is the SAS system. Structured Query Language (SQL) in both SAS and Microsoft Access receives a strong emphasis.

Documentation of the research process and outputs using structured metadata also receives significant attention.

**Topics:**

The SAS Environment – using SAS interactively

Reading Data

Reading More Complex Data - e.g. reading a Web page

Transforming Data

Reorganizing Data

Hierarchical Files –e.g. Census Public Use Microdata Samples (PUMS)

Joining and Concatenating Files SQL and Data steps

Designing your data structures

Custom output, e.g. Web pages, XML

Parameterizing Programs / Macros – Iterating procedures

More SAS Data Step Techniques, Perl regular expressions

Data Interchange – moving data among software packages, operating systems, XML

Metadata and the data lifecycle

Data Cleaning

A brief tour of some other topics

Archiving Data

Survey design

Multidimensional databases

Online Analytical Processing (OLAP servers), Pivot Tables

Visualizing data, Geographic Information Systems – GIS

Data Warehousing

**Course structure:** approximately 50% lecture/demonstration and 50% hands-on lab.

**Auditing Policy:** Faculty and permanent staff are welcome to sit in as space permits.

Please contact Larry – [LarryHoyle@ku.edu](mailto:LarryHoyle@ku.edu)

Monday, June 4, 2018	Tues., June 5, 2018	Wed., June 6,2018	Thurs., June 7, 2018	Friday, June 8, 2018
Read LSB <sup>1</sup> Chapters 1, 8 and 10 sometime during the first week	Overview & SAS Environment, Documenting Process	Reading Data  <i>LSB</i> chap 2	Reading Data  <i>LSB</i> chap 2	Transforming Data  <i>LSB</i> chap 3
Monday, June 11, 2018	Tues., June 12, 2018	Wed., June 13, 2018	Thurs., June 14, 2018	Friday, June 15, 2018
Restructuring Data  <i>LSB</i> chap 6	XML and Data Interchange, DDI (ICPSR) Metadata <i>LSB</i> chap 9 <i>Palmer</i> SUGI paper 036-29 <i>Hoyle</i> SGF 030-2010	Text - Fuzzy matches Perl Regular expressions text mining  <i>Cody</i> SUGI paper 265-29.pdf <i>Pless</i> SUGI paper 043-29.pdf	SQL SELECT ODBC  SQL chaps 1,2 pp1-37. SQL chap 2 pp38-52 SQL chap 3 pp 54-68	SQL Joining and Merging Data Indexing files Pums example  SQL chap 3 pp 69-81 SQL chap 4 pp84-103  <b>Project – Reading Data Due</b>
Monday, June 18, 2018	Tues., June 19, 2018	Wed., June 20, 2018	Thurs., June 21, 2018	Friday, June 22, 2018
Custom Output Popups and drill down  <i>LSB</i> chap 4 and 5 <i>Zdeb</i> SUGI papers 090-29.pdf, 251-29.pdf	SQL Set operations – Union, etc  <i>SQL</i> chap 5 pp 106-125	SQL create, update, alter, integrity constraints	The data lifecycle, data and metadata. DDI	The data lifecycle continued  <i>LSB</i> chap 6  <b>SECOND ASSIGNMENT DUE</b>
Monday, June 25, 2018	Tues., June 26, 2018	Wed., June 27, 2018	Thurs., June 28, 2018	Friday, June 29, 2018
Parameterizing your programs  <i>LSB</i> chap 7	Designing entering and storing Your Data Checking and Cleaning Your Data  <i>Cody</i>	R And Python	Visualizing data	Text Mining  <b>LAST ASSIGNMENT DUE</b>

**Note: all of the reference material needed for this course is available online. You may find the recommended books below to be useful.**

*LSB* – Delwiche, Lora D. and Slaughter, Susan J., ***The Little SAS Book a primer, Fifth Edition*** ISBN 978-1-61290-343-9. Fourth Edition would be ok as well (Recommended)

Slaughter, Susan J and Delwiche, Lora D. ***The Little SAS Book for Enterprise Guide® 4.2*** ISBN 978-1-59994-726-6 (Recommended)

*STEP* – SAS Publishing ***Step-by-Step Programming with Base SAS Software*** ISBN 1-58025-791-7 (Recommended)

*Cody* – Cody, Ron, ***Cody's Data Cleaning Techniques Using SAS Software, Second Edition*** ISBN 978-1-59994-659-7 (Recommended)

SQL – SAS Institute Inc., **SAS SQL Procedure User's Guide, Version 9** available free online at [http://support.sas.com/documentation/onlinedoc/91pdf/sasdoc\\_91/base\\_sqlproc\\_6992.pdf](http://support.sas.com/documentation/onlinedoc/91pdf/sasdoc_91/base_sqlproc_6992.pdf)  
(Available online)

SAS User Group International (SUGI) & SAS Global Forum (2007 on) conference papers can be found online for free via <http://www.lexjansen.com/sugii/> Recommended links at the bottom of the page for special topics : ARRAYS, MACROS, ODS, PROC FORMAT, SQL, XML

Example papers:

Carpenter, Arthur L. Using PROC FCMP to the Fullest: Getting Started and Doing More, SGF2013 paper 139-2013  
<http://support.sas.com/resources/papers/proceedings13/139-2013.pdf>

Cody, Ron SUGI paper 265-29 *An Introduction to Perl Regular Expressions in SAS 9*  
<http://www2.sas.com/proceedings/sugi29/265-29.pdf>

Dunn, Toby; Lafler, Kirk Paul SAS Global Forum 2012 Paper 226-2012 *The Good, The Bad, and The Ugly*  
<http://support.sas.com/resources/papers/proceedings12/226-2012.pdf>

Hoyle, Larry. *Using XML Mapper and Enterprise Guide to Read Data and Metadata from an XML File*. SAS Global Forum 2010 paper 030-2010, Seattle, Washington 2010  
<http://support.sas.com/resources/papers/proceedings10/030-2010.pdf>

Huntley, Scott and Woody Middleton *A Different Point of View with ODS PDF in SAS® 9.3* SAS Global Forum 2012 paper 260-2012  
<http://support.sas.com/resources/papers/proceedings12/260-2012.pdf>

Langston, Rick SAS Global Forum 2007 Paper 027-2007 *Handling Large Stream Files with the '@string' Feature*.  
<http://www2.sas.com/proceedings/forum2007/027-2007.pdf>

Langston, Rick SAS Global Forum Paper 052-2009 *Creating SAS® Data Sets from HTML Table Definitions*  
<http://support.sas.com/resources/papers/proceedings09/052-2009.pdf>

Langston, Rick SAS Global Forum Paper 245-2012 *Using the New Features in PROC FORMAT*  
<http://support.sas.com/resources/papers/proceedings12/245-2012.pdf>

Li, Arthur X.. SAS Global Forum Paper 3181-2015 *How Best to Effectively Teach the SAS® Language*  
<http://support.sas.com/resources/papers/proceedings15/3181-2015.pdf>

Matange, Sanjay *Quick Results with SAS® ODS Graphics Designer* SAS Global Forum 2012 paper 153-2012  
<http://support.sas.com/resources/papers/proceedings12/153-2012.pdf>

McKinney, William *Python for Data Analysis Data Wrangling with Pandas, NumPy, and IPython 2<sup>nd</sup> edition* O'Reilly Media 2017 ISBN-13: 978-1491957660

Nelson, Rob *ODS LAYOUT to Create Publication-Quality PDF Reports of STD Surveillance Data* SAS Global forum paper 216-2010  
<http://support.sas.com/resources/papers/proceedings10/216-2010.pdf>

O'Connor, Daniel and Scott Huntley *Breaking New Ground with SAS® 9.2 ODS Layout Enhancements* SAS Global forum 2009 paper 043-2009  
<http://support.sas.com/resources/papers/proceedings09/043-2009.pdf>

Palmer, Michael paper 036-29 *XML in the DATA Step*  
<http://www2.sas.com/proceedings/sugi29/036-29.pdf>

Pless, Richard SUGI paper 043-29 *An Introduction to Regular Expressions with Examples from Clinical Data*  
<http://www2.sas.com/proceedings/sugi29/043-29.pdf>

Slaughter, Susan and Delwiche, Lora D. *Writing Code in SAS® Enterprise Guide®*  
[http://www.lexjansen.com/wuss/2013/83\\_Paper.pdf](http://www.lexjansen.com/wuss/2013/83_Paper.pdf)

Sweigart, Al *Automate the Boring Stuff with Python* <https://automatetheboringstuff.com/>

Zdeb, Mike SUGI paper 090-29 *Pop-Ups, Drill-Downs, and Animation*  
<http://www2.sas.com/proceedings/sugi29/090-29.pdf>

Zdeb, Mike SUGI paper 251-29 The Basics of Map Creation with SAS/GRAPH®  
<http://www2.sas.com/proceedings/sugi29/251-29.pdf>

Other Resources:

Delwiche, Lora D. and Slaughter, Susan J *Coming to SAS® from SPSS A Programming Approach*  
[http://support.sas.com/publishing/bbu/companion\\_site/62272.pdf](http://support.sas.com/publishing/bbu/companion_site/62272.pdf)

Other Course Notes

Course materials prepared by the instructor, together with the content of all lectures and review sessions presented by the instructor are the property of the instructor. Video and audio recording of lectures and review sessions without the consent of the instructor is prohibited. On request, the instructor may grant permission for students to audio tape lectures, on the condition that these audio tapes are only used as a study aid by the individual making the recording. Unless explicit permission is obtained from the instructor, recordings of lectures and review sessions may not be modified and must not be transferred or transmitted to any other person, whether or not that individual is enrolled in the course

Unless explicit permission is obtained from Larry Hoyle, materials on the course CD or DVD, server or web site may not be modified and must not be transferred or transmitted to any other person, whether or not that individual is enrolled in the course.

Students have found that reading the suggested material in advance is very helpful.

Because this is a "short" course, the following dates apply:

<b>Dept</b>	LA&S
<b>Course</b>	792
<b>Class Number</b>	81169
<b>Start Date</b>	6/5/2018
<b>End Date</b>	6/29/2018
<b>Last Day to Add with Approval *</b>	6/11/2018
<b>Last Day for CR\No CR **</b>	6/15/2018
<b>Last Day to Cancel</b>	6/8/2018
<b>Last Day to Withdraw</b>	6/25/2018
<b>Last Day for 100% Refund</b>	6/7/2018
<b>Last Day for 50% Refund</b>	6/10/2018
<b>Location</b>	LAWRENCE