

**Computing in Sociological Research**  
**Sociology 365, Fall, 2007**  
**11:00-12:15 TR, 3218 Sewell Social Science Bldg.**

**Prof. John Allen Logan**

Email: logan@ssc.wisc.edu  
Phone: 262-0995  
Office hours: Monday 10-12  
Office: 4438 Social Science

**Scheduled Guest instructors:**

**Bill Buckingham**, Applied Population Laboratory, Rural Sociology

**Charlie Fiss**, Social Science Research Services

**Ryan Horrisberger**, Social Science Research Services

**Ann Lewis**, Social Science Research Services

**Jack Solock**, Social Science Research Services

**Meeting Locations:** The timetable's designated classroom is 6116 Sewell. Most (or all) sessions will be held in 3218 Sewell instead. I'll announce any sessions to be held in 6116.

**Subject Matter and Objectives:** This course teaches computer skills appropriate for professional social science researchers. Students completing the course should be able to use computers at all stages of research from recording, cleaning, transforming, and analyzing data to preparing reports and presentations.

**Course Theme and Intended Audience:** This course teaches what should lie in between research *data* on the one hand, and professional *publications* on the other. This involves not only the exercise of various computer skills and statistical expertise, but also developing ways of permanently recording the steps between data and paper. The course is for students who intend to do graduate-level research or to work as data analysts either professionally or as research assistants for others.

**Course Software:** Though effective use of computers in social research involves more than knowledge of particular programming environments, we will focus on the **SAS** statistical analysis system, the **Emacs** text editor, and the **Linux** (or **Unix**) operating system. SAS skills are highly valued in the marketplace for researchers, while Emacs and Unix are particularly well suited to scientific research, and are quite widely used and valued. Students who have mastered the fundamentals of these programs are ready to quickly master other environments, and should be well qualified for research positions. We will also cover the SQL database query language and the ArcGIS geographic information system.

**Prerequisite:** An introductory statistics course, e.g., Sociology 360. Basic familiarity with the Windows operating system and use of Powerpoint is helpful.

**Final exam period:** Thursday, December 20, 10:05 a.m. -12:05 p.m.

**Required Textbooks** (available at University Book Store):

Delwiche, Lori D., and Susan J. Slaughter. 2003. *The Little SAS Book: A Primer, Third Edition*. Cary, N.C.: SAS Institute.

Cameron, Debra, et al. 2005. *Learning GNU Emacs, Third Edition*. Sebastopol, CA: O'Reilly Media.

Ray, Deborah S., and Eric J. Ray. 2007. *Unix: Visual Quickstart Guide, Third Edition*. Berkeley, CA: Peachpit Press.

**Additional Readings:**

Bennett, Steve, Mark Myatt, Damien Jolley, and Andrzej Radalowicz. 2001. **Data Management for Surveys and Trials; A Practical Primer Using Epidata**. 98 pp. Download from <http://www.epidata.dk/php/downloadc.php?file=dmeidata.pdf> .

Juul, Svend. November, 2004. **Take Good Care of Your Data**. 57 pp. Download from <http://www.folkesundhed.au.dk/uddannelse/software/takecare.pdf> .

**Social Science Computing Coop publications** (various numbers), online at <http://www.ssc.wisc.edu/sscc/pubs/allpubs.htm> .

Readings are assigned in the syllabus. Some other publications, especially from SSCC, may be added as assignments during the semester.

**Arriving for class:** Most class sessions will require you to log in to a lab computer. Please try to arrive early enough to do this before the schedule start of class.

**Open Class Sessions:** A number of class sessions are designated in the syllabus as open to pre-registered students from the broader research community, via SSCC's training seminar series. This arrangement is part of what enables us to draw on such a broad range of expert lecturers.

Please be helpful and courteous to visitors in the open class sessions. *Registered students in Sociology 365 will have precedence for computer access and seating over visitors, so make sure to let me or another instructor know if you have any difficulties getting a seat or a properly working computer before the session starts.*

**SSCC Linux Computers:** Most of the course work will be done by connecting the classroom computers remotely to Linux/Unix computers operated by the Social Science Computing Coop (SSCC). The SSCC system is usually off-limits to undergrads because of restrictions in the research contracts that support it. Though we will use the SSCC's computers, and will get expert lectures from SSCC staff, it is important that students do not request consulting help from SSCC staff or attempt to use any of its facilities except as explicitly authorized in class.

Please be especially careful to follow the rules for printing on the SSCC printers, as will be presented in class. No printing of materials unrelated to the class is permitted.

*Classroom computers are to be used only for our class's activities. No reading your other email, browsing the web, instant messaging, etc. Grade penalties will be imposed for violations.*

Students should plan to spend a good amount of time outside class doing computer assignments. This can be done in the SSCC's computer labs (3218 and 4218 Social Science Building) or, depending on one's initiative and resources, at other campus labs or on a personally-owned computer by connecting to the SSCC computer system. Students will be provided with accounts that allow for remote access to the computer system as well as physical access to the lab.

**Assignments and Grading:** There will be two in-class exams, a group project and presentation, and an individually-written paper based on the group project. Homework assignments will be given once or twice a week. Grading percentage allocation: homework 30%, first exam 15%, second exam 15%, project paper and supporting documentation 20%, project presentation 10%, participation in lecture discussions 10%.

Homework will be graded using this scheme: zero, passable, good, or outstanding. Late assignments will have a maximum grade of passable.

**Term Project, Presentation, and Paper:** Groups of three or four undergrad students each will develop a term project that answers a research question using publicly available data. Group members will collaborate on defining the question, locating and reading suitable data, creating a working data set, analyzing the data and producing a presentation that includes high-resolution tables and/or graphs. Each student will also write an individual report as a term paper. Resources for all these steps will be identified in lectures.

Undergrads' term papers should be about five pages, plus tables and graphs. Required formats are paper (the substance made out of fiber) *and* pdf; both must be submitted.

**Supporting Documentation:** You must also submit a complete set of data and supporting documentation allowing all the analyses, tables, and graphs in the paper to be regenerated from the original data set without any further modifications or intervening steps. These files can be submitted by freezing a final directory or directories on the SSCC system, *making sure your permissions are set so that I can access them*, and sending me an email stating when that was done.

The paper and all supporting documentation are due at the scheduled time of the final exam, but can be submitted earlier.

**Grad Student Project, Presentation, and Paper.** Graduate students must work individually on all parts of the project, presentation, and paper, rather than in groups. Higher quality is expected, and the page target is relaxed to about 10 pages.

**Collaboration.** It is okay if you talk with each other about how to do your homework. A little collaboration on getting the computer programs to do what you want is encouraged. But you are required to write up your own answers to the homework, independently of what other students have written. *You are also required to produce and attach your own computer printout*, whenever printout is needed. No photocopies or second printings of other students' printouts are allowed. No cutting and pasting, or other copying, from others' programs is allowed in creating your own programs. *These rules also apply for the term papers.*

The final presentation, on the other hand, should be a completely collaborative project with all group members sharing the work equally. Individual members of each group will be asked to submit notes indicating their assessments of the work efforts of their colleagues.

[SEE SCHEDULE OF TOPICS AND READINGS ON NEXT PAGE.]

### Schedule of Topics and Assigned Readings

Readings are indicated like this: "U 1-3" means all of chapters 1 through 3 of U. "U p1-6" means pages 1-6. "U 4.1-4.6" means chapter 4, sections 1-6 of U. "SSCC # 7-28" means all of SSCC publication number 7-28. And so on. See code definitions below the table.

Session, Day, Date and Topics	Open*	Readings/Co-Instructor
1. T Sep 4. Intro, w/ SSCC intro,	no	SSCC Handbook/Ann Lewis
2. R Sep 6. Research Data Resources	yes	--- / Jack Solock, Charlie Fiss
3. T Sep 11. Epidata Data Entry Program	no	Bennett, et al., pp. 7-63.
4. R Sep 13. Research Data Management	no	Juul, pp. 1-37
5. T Sep 18. Introduction to Linux	yes	U 1-3; SSCC # 7-28 / Ryan H.
6. R Sep 20. Linux Permissions; File Manip.	no	U 5-6
7. T Sep 25. Introduction to Emacs	yes	LGE 1-3
8. R Sep 27. Emacs Buffers & Work Environment	no	LGE 4,5
9. T Oct 2. Review/Catch-up	no	---
10. R Oct 4. Exam 1	no	---
11. T Oct 9. Intro to SAS; Running SAS in Emacs	no	LSB 1 (skim PC parts)
12. R Oct 11. Reading/Saving/Converting Data	no	LSB 2.4-16, 2.19-22, 9.1-5, 9.7
13. T Oct 16. SAS Data Step I	yes	SSCC # 4-18 / Russell D.
14. R Oct 18. SAS Data Step II	yes	SSCC # 4-18 / Russell D.
15. T Oct 23. In-Class Data Exercises	no	LSB 4, 6
16. R Oct 25. SAS Statistical Procs	no	LSB 8
17. T Oct 30. Numerical Issues in Computing	no	t.b.a.
18. R Nov 1. SAS High Resolution Tables/Graphs	no	t.b.a. / Doug Hemken
19. T Nov 6. Review/Catch-up	no	---
20. R Nov 8. Exam 2	no	---
21. T Nov 13. SAS Array Processing	yes	LSB 3.10-11 / Doug Hemken
22. R Nov 15. SAS Macros	yes	LSB 7 / Doug Hemken
23. T Nov 20. Emacs Macros and Registers	no	LGE 6; p194-201
(Thanksgiving)		
24. T Nov 27. Structured Query Language (SQL)	yes	t.b.a. / Ryan Horrisberger
25. R Nov 29. GIS overview.	yes	t.b.a. / Bill Buckingham
26. T Dec 4. ArcGIS	yes	t.b.a. / Bill Buckingham
27. R Dec 6. In-class project workshop	no	t.b.a.
28. T Dec 11. Project reports I	no	---
29. R Dec 13. Project reports II	no	---

#### NOTES:

\*Open to pre-registered visitors through SSCC training series.

Reading codes:

U = Unix; LGE = Learning Gnu Emacs; LSB = Little SAS Book;

Juul and Bennett, et al., are listed in "Additional Readings," above.

SSC = SSCC Publication, from [www.ssc.wisc.edu/sscc/pubs/allpubs.htm](http://www.ssc.wisc.edu/sscc/pubs/allpubs.htm)