Syllabus for EPSY 887, Spring 2011

Instructor: Robert Pruzek, Professor (rmpruzek@yahoo.com, fax & phone: 518-442-5066)
Most communications with students to be within the **wiki** prepared for the course: moderngraphics11.pbworks.com (first page open for all to see; many other pages require registration).

Teaching Assistant: none

Prerequisites: EPSY 630 or equivalent (basically two intro statistics courses; and some background with R software is of special value, although not a prerequisite).

Schedule: Each week, classes meet on Thursdays from 4:15 until 7:05 pm in B-13 of the SOE, i.e., the Mac Lab. (This facilitates computer-based analyses, instruction, and interactions where each student has control of the software used for data analyses and graphics.) The wiki will serve as a basis for all material that is covered. Assignments are given in class, and spelled out in detail on website. Two or three relatively brief exams as well as a final are scheduled; the topics for these exams (which may be given as homework to be handed in) will be spelled out in detail using class reviews. No general paper is assigned for this course, but presentations will be invited, and a typical class will incorporate at least one 10-15 minute presentation (often by teams of students).

Textbook: No textbook, as such, is used. But several books and articles (see below) will be cited as having a central role in the class. Numerous documents, mostly pdfs, and various URL links, are used to support student learning. These range from wiki files prepared (mostly) by the instructor to cover the topics on which we focus, to documents identified at prescribed URLs, as well as applets for demonstrations and commentaries. It is expected that students will regularly access and take advantage of these and other resources.

***The course is Graded*: A-E**. Grades determined principally by student performances on two or three brief exams plus a final (adding all points). Participation, in the form of class presentations, and project work, submitted exercises (about one per week), and discussion is also required. But while all homework and presentations are assessed, and to the extent possible feedback is provided to each student, these assessments do not contribute to student grades. Interactions with students are in general (except in special situations) shared with all students online. Guidelines for homework are provided in class as well as on the course wiki. (As a graduate class most grades are either A's or B's; consistent and effective participation usually ensures at least a B in this class.)

***Catalog Description*:** *EPSY 887 Modern graphics for social science research:*
Graphical and related methods for applications: modern fixed and dynamic graphics and related methods; tools for analysis and interpretations of data. Major goals will be to tailor the course to needs of students for analysis and presentation of data, to help ensure operational skills as well as sound understanding of modern methods for both analysis and graphics to facilitate communication. Software applications, especially R, will be emphasized, taking advantage of new vehicles (youtube, web videos, cloud computing, etc.) to facilitate analysis and display. Prerequisites: EPSY 630 or equivalent. (3 credits)

Course Overview: Recent years have seen huge increases in the numbers of powerful personal computers and access to the high speed internet. Of special note, recent years have seen enormous growth in software capabilities also, software that in the case of statistics and related methodology, offers great promise for profound changes in statistical computing and especially in graphics. Much of this software is free, readily downloaded from the net. In particular, R

software (r-project.org) has become the software of choice for professional statisticians and data analysts around the world; and it has become central not only to research and production but also instruction. Indeed, the advent of these technologies have already afforded progress on too many fronts to be understood by any one person. For example, R has more than 2500 packages with numerous functions in each, most of which are being continuously updated by their authors. The existence of such software, free, widely available, powerful and growing, represents a class of riches that cannot be ignored by any modern university, especially those with strong orientations toward research and development. This course aims to introduce students to a selected sample of modern graphics that holds promise for improving their understanding of, and abilities to use, many new forms graphical methodology in core areas of measurement, design, analysis methods and ways to interpret data. The key idea of the course is to build on what students have experienced in textbooks, articles and research projects, to help envision new and especially meaningful ways of conducting effective, and perhaps more meaningful research, quite possibly to be more efficient and effective than has been possible with older methods and technology. To the extent possible, the course will be tailored to the needs of students, with an emphasis on methods that appear to hold special promise for them. The role of students is also central: to learn to use relevant software, to construct graphics that they can interpret, to ask key questions, and to help one another to engage in this methodology.

What follows is a listing of *principal topics for the course*, as well as representative questions on which students may profitably focus as each topic is covered. Many of these topics are closely related to one another as will be made clear when they are discussed. All methods will be implemented using R functions; code for R will be supplied, and invited. Students are expected to learn R from one another as well as from materials that have been provided.

◆ Methods for displaying and comparing groups, from the simple to (moderately) complex:

- Q's: How can one visualize simple data sets, such as for comparing two independent groups? What different ways might be effective in different contexts? How effectively can summary differences, such as effect sizes, t-statistics, confidence intervals be seen and interpreted in different contexts? (In all cases illustrations will be most desirable.)

◆ Graphics for specialized (descriptive) comparisons:

Q's: What novel methods are available, and promising, for particular needs? E.g., for identifying individuals or subgroups; for comparing multiple groups by pairs; for being effective when groups are either very small or very large; or for highlighting special features? What are q-q plots?

◆ Graphics to facilitate statistical inferences when comparing groups:

Q's: How to facilitate ANOVA inferences, for independent and dependent samples? How to display confidence intervals (and interpret them)? How to use and display bootstrapping results when traditional (normal) theory assumptions are unrealistic? How do independent and dependent sample problems differ, and how can corresponding data effectively be visualized?

◆ Graphics for basic and intermediate regression (prediction) problems:

Q's: How can one improve on linear regression (e.g., loess and more), and show results graphically? What are some key issues that arise when moving away from linear regression methods, and how might some of them be displayed graphically? How can graphics be used to show when transformations or re-expressions of data are likely to be useful? How might two predictors be used in predicting a quantitative outcome using non-linear (e.g., loess) regression?

- Dynamic graphics, and how they can improve understanding of data:
  Q's: What are some of the key forms of dynamic graphics? What are their primary uses? What are major limitations currently for these methods?

- Graphics for propensity score analysis (to be introduced and discussed in detail):
  Q's: What are some of the key graphical methods for the following? Describing and comparing distributions of derived propensity scores; examining balance for covariate distributions (across treatment groups); comparing treatment groups on outcome variables?

- Lattice graphics to afford conditional examinations of data:
  Q's: Can you name and discuss at least four different kinds of lattice graphics? E.g., for comparing subgroups, perhaps using histograms, boxplots or density plots? Also, for comparing groups on bivariate distributions across subgroups? For profile plots, possibly for observed and fitted profiles (e.g. in longitudinal data analysis)? Etc.

- Dendograms and related graphics for use in cluster analysis:
  Q's: What are the key virtues and limitations of dendograms? How do dendograms relate to derived trees? What (cluster) methods generally provide dendogram displays? What other methods can be used effectively for cluster displays?

- Graphics for displaying categorical data:
  Q's: What are association plots? What are mosaic plots (2d and 3d)? How might different variations on such plots be useful? What other methods or graphics deserve mention for categorical data display?

- Graphics for longitudinal (and time series) data:
  Q's: What are some of the standard approaches for displaying data across time? What are some of the main problems that interfere with interpretations of corresponding graphics? How (say w/ lattice methods) might some of these problems be ameliorated? How might individual differences effectively be displayed or compared?

The following ***books***, and major pdfs, will be central resources for the preceding.

Cleveland, W.S. *Visualizing Data* (1993);

Sarkar, D. *Lattice: Multivariate Data Visualization with R* (2008) (A major pdf related to this book, w/ R code, is here: http://lmdvr.r-forge.r-project.org/figures/figures.html);

Adler, J. *R in a Nutshell* (2010) (pdf available)

… other books to be added… by you, and by me…

--- see the course wiki for pdfs, URLs, and more ---

Reasonable Accommodations: Reasonable accommodations are provided for students with documented physical, sensory, systemic, cognitive, learning and psychiatric disabilities. If you believe you have a disability requiring accommodation in this class, please notify the Director of Disabled Student Services (Campus Center 137, 442-5490). That office will provide the course instructor with verification of your disability, and will recommend appropriate accommodations. For more information, visit the website of the UAlbany Office for Disabled Student Services: http://www.albany.edu/studentlife/DSS/guidelines/accommodation.html

Academic Integrity (A comparable statement from Universities' Admissions & Graduate Requirements appears on the moodle website, and is referenced at Exam times)

Whatever you produce for this course should be your own work and created specifically for this course. You cannot present work produced by others, nor offer any work that you presented or will present to another course. If you borrow text or media from another source or paraphrase substantial ideas from someone else, you must provide a reference to your source.

The university policy on academic dishonesty is clearly outlined in the Student Bulletin, and includes, but is not limited to plagiarism, cheating on examinations, multiple submissions, forgery, unauthorized collaboration, and falsification. These are serious infractions of University regulations and could result in a failing grade for the work in question, a failing grade in the course, or dismissal from the University.