**EXPLORATORY and GRAPHICAL DATA ANALYSIS, PSY 319**

**Spring 2014**

**Time and Place: Wed 9:10-12:00, Admin Bldg 104 Instructor: Joe Rodgers (Hobbs 202)**

**Textbooks: Tukeys Exploratory Data Analysis, Clevelands Visualizing Data, readings**

Wed, January 8 First class meeting, hand out personal EDA, 1&2; Roberts, 2001

data collection exercise

Introduction to Course, The philosophy of EDA

Wed, January 15 Permissible EDA/CDA combinations EDA, 3&4

Ethics and Statistical Analysis

Research & Science -- the big picture Wilkinson, Am Psych, 1999

Portraying numbers Tukey

Wed, January 22 Graphical data analysis Vdat, 1

Wainer & Velleman, Ann Rev of Psych, 2001

Wed, January 29 Graphical masterpieces Vdat, 2

Beniger & Brown, Am Stat,

Feb 1978

Kosslyn, JASA, Sept. 1985

EDA Laboratory – Introduction Cleveland & McGill, JASA, Sept. 1984

Wed, February 5 Graphing Scatterplots & Correlations Rodgers & Nicewander,

Am Stat, Feb 1988

Thissen et al, Psych Bull,

1981; Vdat 3

Wed, February 12 Scatterplots & Correlations (cont.)

Wilkenson’s The Grammar of Graphics

EDA Laboratory team meetings

Wed, February 19 Quiz 1

Re-expression and data transformations EDA, 5; Vdat, 4-6

Multiway GDA EDA, 10

Wed, February 26 Two-way analyses & median smoothing EDA, 11

EDA Laboratory team meetings

Wed, March 12 Robustness and Resistance

EDA Laboratory

Wed, March 19 Straightening, Flattening, Hanning, & Smoothing EDA, 6&7

EDA Laboratory team meetings

Wed, March 26 Quiz 2

Book Review, Robust Estimate of Location

Data mining and CART McArdle chapter, 2010

Wed, April 2 Team presentations

Wed, April 9 Individual Presentations

Wed, April 16 Individual Presentations

EDA A retrospective Overview Rodgers, Amer Psych,

Turn in Individual Project Portfolios 2010

**Grading:** You will be formally evaluated in four different ways

1. You will turn in your individual data collection/data analysis projects on Wednesday,

January 29. These projects are described in a separate handout. This project

will be worth 10% of your grade.

2) There will be two 30-minute quizzes over lecture material and readings

(on Feb 19 and March 26). These will be objective style quizzes. Make-ups

will only be given under extraordinary circumstances and by prior arrangement. Each quiz = 10% of your grade.

3) You will give two presentations. Your group will give a 15-20 minute presentation of

Its data analysis problem on April 2. In addition, you will give a 10-minute

individual presentation on April 9 or 16. These presentations each

account for 10% of your grade. Each student will get an automatic 9.0 (i.e., a

neutral A) if you give your presentations in a timely and well-prepared fashion

(i.e., no counting off for nervousness, for presentation quality, etc. this is a

chance to practice a formal presentation in a relatively pressure-free setting).

4) The major semester projects (described on the separate hand-out) is a set of

EDA Projects, which you will summarize in an EDA Portfolio to be handed in on Wednesday April 16. Project = 50% of your grade.

In addition, the following is expected of you in this class:

1) Attend class every day. (If you cant attend for some reason, please let me know that before

class; call me, e-mail, leave voice mail, leave a note, or

otherwise inform me that you will be absent; telepathy doesnt count!)

2) Discuss your projects with your instructor and other class members. EDA is a process

that is facilitated by discussion. Feel free (expected) to do so.

3) Be enthusiastic about EDA (pretend if necessary)!!!!!!

**Disabilities:** Any student who has a disability that might interfere with normal participation in the class should contact the instructor. All possible effort will be made to support and accommodate students with disabilities.

**Academic Misconduct:** No form of cheating, plagiarism, or other forms of academic misconduct will be tolerated. If students are unsure of the boundaries, they should contact the instructor.

EXPLORATORY DATA ANALYSIS

PSYCHOLOGY 319, Spring, 2014

**Required EDA Projects**

EDA is like basketball. We can watch either being done, and appreciate the art and skill involved in high-level performance. In the hands of Lebron James or Michael Jordan, a basketball is a highly-tuned artistic instrument; in the hands of John Tukey, a graph sings the praises of data in melodies both harmonious and discordant, reflecting model, data, and mood. Part of this course will be devoted to **Watching** and **Studying** the master at his work.

But basketball is played by thousands of bodies with less than NBA training and ability. Some novice basketball players are just learning their craft, and others will evolve into future LJ’s and MJs; others have lower aspirations, yet still enjoy participating. So also should EDA be played. A second part of this course will involve **learning to do EDA** by **Doing It**.

Each of you will be expected to do **several** EDA projects. These projects will be done during EDA Labs in class, as well as during out-of-class effort. The nature of most of the particular projects will be entirely up to you. You will report to your instructor during EDA Labs on what you have been doing and what you plan to do. You will give a 15-minute individual presentation to the class at the end of the course on what you did in one or two of your major projects. There are three requirements. Your projects must include some **hand plotting**. Your projects must include some **computer work**. And your projects must include one **group project** with two other people (your team projects). Results of this group project will also be presented toward the end of the course.

Each project will require some data, to which EDA techniques will be applied. You are welcomed (in fact, strongly encouraged) to use data with which you are currently involved; dissertation or thesis data, a research project, the almanac, data from an article, data from EDA or VDat, data you collect from your family or friends, or data provided to you by your instructor are possible sources.

Examples of appropriate EDA projects are listed below. I hope some or all of these will be worked on by members of the class. You should develop and work on your projects individually (except for the group project), but discussion with the instructor and others class members is encouraged and in fact expected.

**Possible EDA Projects**

1) Draw plots by hand of some data that are of interest to you, and transform the variables in

several different ways. Interpret your results.

2) Choose some data from EDA or VDat; table or plot them in a way that Tukey/Cleveland didnt.

3) Find some population data of interest to you (e.g., Tennessee, Davidson County, Afghanistan,

your ant farm, etc.) and do several hand plots like those in Chapter 5 of EDA. Interpret results.

4) Find some data in the World Almanac and plot and/or table them.

5) Use some two-way data, and repeatedly extract the medians like Tukey does in Ch. 10 & 11.

6) Find some time series data, and smooth them in several different ways (see EDA, ch. 7).

Data with seasonal patterns are especially interesting (see VDat, pp. 152-172).

7) Write an R, SAS-Graph, SPSS, BASIC, FORTRAN, C, JAVA, or other program to portray influence-enhanced scatterplots. Produce scatterplots of several relationships.

8) Write a BASIC, FORTRAN, C, SAS, SPSS, JAVA, or other program to portray scatter plots on

a computer. Give the user the option to plot X and/or Y as either raw data, logs,

squares, cubes, reciprocals, roots, etc.

9) Write an R, SAS-Graph, SPSS, BASIC, FORTRAN, C, JAVA or other program to produce some exotic version of stem-and-leaf diagrams.

10) Write a an R, SAS-Graph, SPSS, BASIC, FORTRAN, C, JAVA or other program to plot in

three-dimensions with time as one of the dimensions (i.e., a kinostatistical plot).

11) Use R or SAS-Graph or some other dedicated graphical package to plot some interesting data (preferably in color, possibly in 3D, maybe even in higher than 3D).

12) Write an R/SAS routine to do median smoothing by three, and use it on some data.

13) Write an R program, or SAS MACRO or SAS PROC or SAS program to produce some EDA

output (but dont duplicate what PROC UNIVARIATE already does).

14) Find an R program in the R library that does interesting EDA; apply it to some interesting data.

15) Produce a correlation matrix between many variables, and develop a scatterplot matrix from it.

16) Read the literature on graphical data analysis and develop some new graphical techniques. Program your techniques. Apply them to real data.

17) Invent a new EDA graphical application, and apply it to real data.

You should keep a log describing all EDA projects you undertake (including the group project). At the end of the course, three things will happen. **First**, you will give a 15- to 20-minute presentation of your Group EDA Project (in teams of ≈ three). The members of each group must contribute approximately equally to the presentation. **Second**, you will give a 12- to 15-minute in-class presentation in which you choose one of your EDA projects to describe to the class. Your description should include the goal of the project, the data you used, and a demonstration (PowerPoint, handout, holdup, computer demo, etc.) of the product. **Third**, you will turn in an **EDA Portfolio** which consists of two components:

1) A report describing all your projects. There must be at least two projects --

a hand-plotting project and a computer project (your group project may or may

not account for one of these). Or you may do 8 or 10 or even more projects. The

total number depends on the scope and difficulty of each project (a range of 4-6 is

typical). There may be projects that you dont finish. Thats fine; EDA projects

are hardly ever completely finished; write them up anyway. The projects

should be numbered consecutively (i.e., in the order in which you began them),

and should include for each project a description of the goal, the product

(computer program, hand graph, computer graph, etc.), the data, and some

interpretation.

Reports must be Word Processed and of high quality in terms of writing,

grammar, presentation, etc. You will be evaluated on writing, presentation

quality, and quality of the data analysis activities.

2) A prototypical example of the product of each project (e.g., a graph, computer

code, etc.). You may wish to put computer output into binders or appendices,

graphs into report folders, etc.

Portfolios will not be returned; if you wish to have a copy, make one before you turn it in. Portfolios are due on Wednesday of the last week of class, April 16, 2012. Project reports **will not be accepted late**. Please, no exceptions!!!