


TEACHING DATA STRUCTURES AND RELATIONSHIPS USING THE GENERAL SOCIAL SURVEYS

Stephen Sweet
Ithaca College

Professional Enhancement Programs Conference
Macalester College
June 17, 2005

What Can a Unicorn and a Carpenter Tell Us About Math and Quantitative Literacy?



The Two Mathematics

Source: Madison, Bernard. 2004. "Two Mathematics: Ever the Twain Shall Meet?" *Peer Review* 6:9-12.

- "Real" Mathematics
 - Geometry
 - Algebra
 - Trigonometry
 - Calculus
- Focus is on math as a purpose unto itself. Principles to be studied, dismantled, and synthesized.
 - Absolute precision is expected
- Quantitative Literacy
 - The blending of mathematical tools with linguistic constructs
- Focus is on the application of mathematical reasoning to consider the workings of the natural and social worlds
 - Math-Lite?
 - Pragmatic orientation to imprecision
 - Greater acceptance mathematical blackboxes

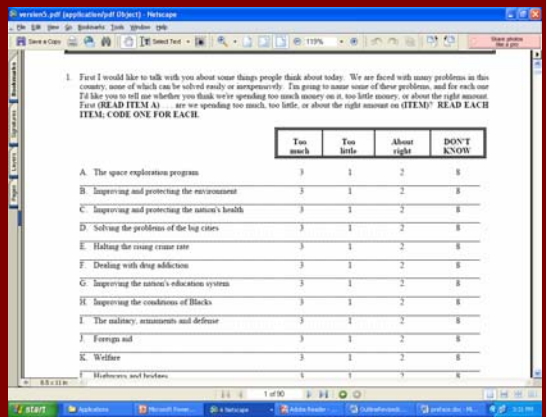
A Deweyian Approach to Quantitative Literacy

Dewey, J. (1916). *Democracy and education*. New York: Macmillan.

- Structure a lab situation that opens opportunities for questioning and exploring data.
- As part of the experience, students engage in critical reasoning through dialogue.
- Process
 - Introduce data
 - Form hypotheses
 - Test hypotheses with bivariate methods
 - Critically reflect on findings through dialogue and form new hypotheses
 - Test these relationships with bivariate and multivariate methods

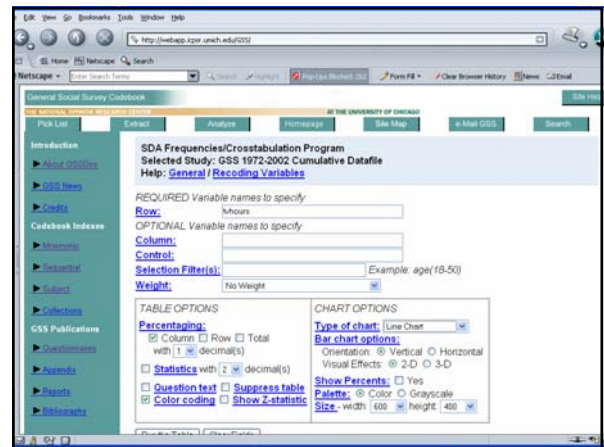
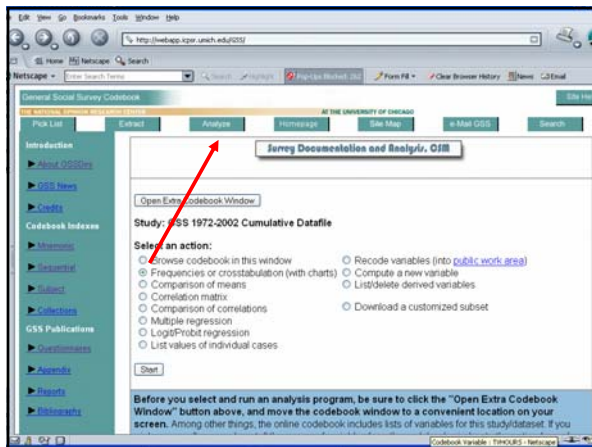
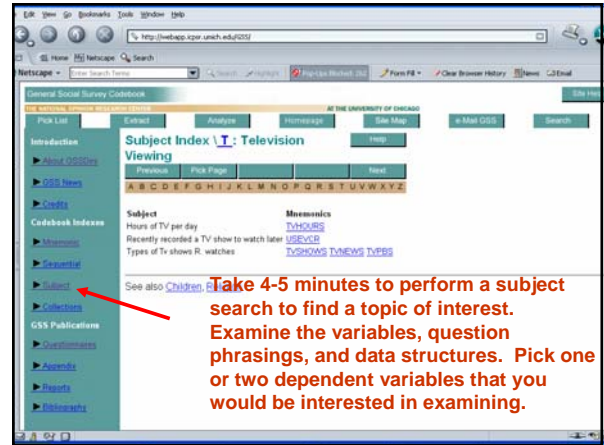
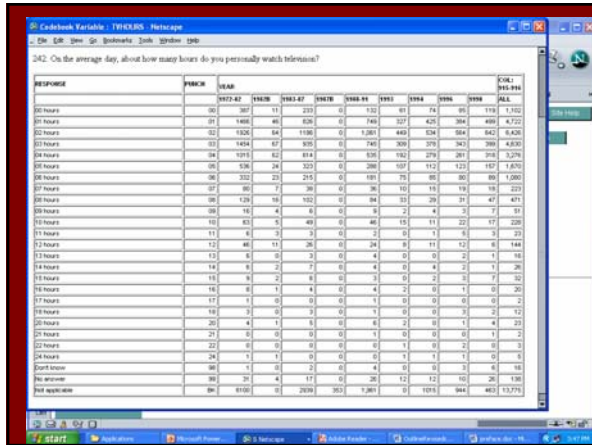
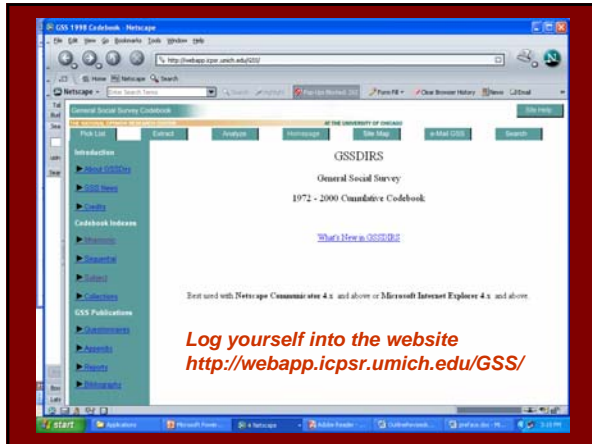
The General Social Surveys (GSS)

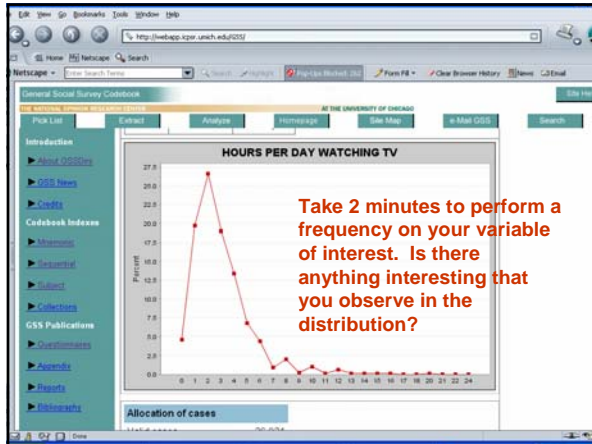
- Annual surveys conducted by the National Opinion Research Center
- The first survey took place in 1972 and most recent survey was in 2002.
- Some questions have been asked in all surveys, other questions have been introduced as "modules." Strong emphasis on reliability across years.
- Since its inception more than 38,000 respondents have answered over 3,260 different questions.
- Enables longitudinal analyses and the study of change over a thirty year span
- Enables comparisons of subgroups from a representative sample of American adults
- By combining data across years, the GSS enables the identification of subgroups in the population, even those that are usually with too few number to be studied in other surveys (e.g. retirees, homosexuals, members of minority religions, etc.)



1. First I would like to talk with you about some things people think about today. We are faced with many problems in this country, some of which can be solved easily or unimportantly. I'm going to name some of these problems, and for each one I'd like you to tell me whether you think we're spending too much money on it, too little money, or about the right amount. First (READ ITEM A) ... are we spending too much, too little, or about the right amount on (ITEM)? READ EACH ITEM. CODE ONE FOR EACH.

	Too much	Too little	About right	DO NOT KNOW
A. The space exploration program	3	1	2	8
B. Improving and protecting the environment	3	1	2	8
C. Improving and protecting the nation's health	3	1	2	8
D. Solving the problems of the big cities	3	1	2	8
E. Halting the rising crime rate	3	1	2	8
F. Dealing with drug addiction	3	1	2	8
G. Improving the nation's education system	3	1	2	8
H. Improving the conditions of Blacks	3	1	2	8
I. The military, armaments and defense	3	1	2	8
J. Foreign aid	3	1	2	8
K. Welfare	3	1	2	8
L. Healthcare and health	3	1	2	8





Survey Documentation and Analysis, OJIB

Open Extra Codebook Window

Study: GSS 1972-2002 Cumulative Datafile

Select an action:

- Browse codebook in this window
- Frequencies or crosstabulation (with charts)
- Comparison of means
- Correlation matrix
- Comparison of correlations
- Multiple regression
- Logit/Probit regression
- List values of individual cases
- Recode variables (into [public work area](#))
- Compute a new variable
- List/delete derived variables
- Download a customized subset

Start

Before you select and run an analysis program, be sure to click the "Open Extra Codebook Window" button above, and move the codebook window to a convenient location on your screen. Among other things, the online codebook includes lists of variables for this study/dataset. If you

Role	Name	Label	Range	MD	Dataset
Correlate	tvhours	HOURS PER DAY WATCHING TV	0-24	-1,98.99	1
Correlate	year	GSS YEAR FOR THIS RESPONDENT	1972-2002		1

Correlation Matrix		
	tvhours	year
tvhours	1.00	-.01
year	-.01	1.00

Missing data excluded: Listwise

Color coding: <0.45, <0.30, <0.15, <0.00, >0.00, >0.15, >0.30, >0.45

Allocation of cases

Valid cases	26,921
Cases with invalid codes on variables correlated	16,777
Total cases	43,698

- ### Why Is No Relationship Observed?
- Perhaps there is no relationship!
 - Perhaps we need to reconstruct the measure of TV viewing, not as a continuous measure, but rather as a concern of "heavy" and "light" watching.
 - Perhaps individual years do not give us sufficient sample sizes to detect trends.
 - Let us try recoding variables and changing data structures and see if that changes our conclusion

Survey Documentation and Analysis, OJIB

Open Extra Codebook Window

Study: GSS 1972-2002 Cumulative Datafile

Select an action:

- Browse codebook in this window
- Frequencies or crosstabulation (with charts)
- Comparison of means
- Correlation matrix
- Comparison of correlations
- Multiple regression
- Logit/Probit regression
- List values of individual cases
- Recode variables (into [public work area](#))
- Compute a new variable
- List/delete derived variables
- Download a customized subset

Start

Before you select and run an analysis program, be sure to click the "Open Extra Codebook Window" button above, and move the codebook window to a convenient location on your screen. Among other things, the online codebook includes lists of variables for this study/dataset. If you

SDA Recode Program

Selected Study: GSS 1972-2002 Cumulative Datafile

Help: [General / Recoding Rules](#)

NAMES of the variables

Name for the new variable to be created: tvhours2

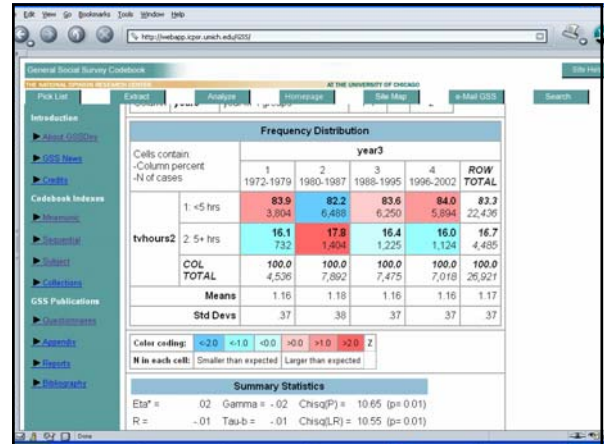
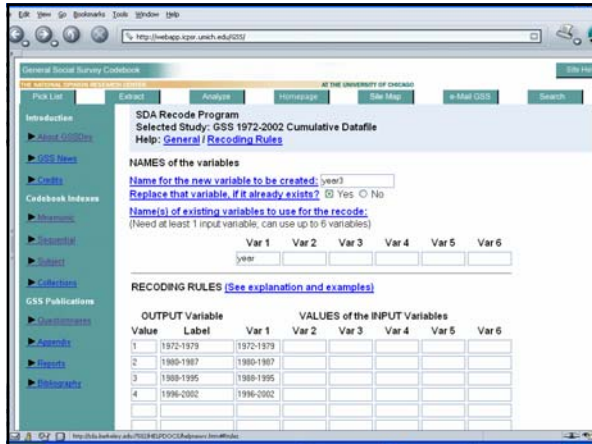
Replace that variable, if it already exists? Yes No

Name(s) of existing variables to use for the recode: (Need at least 1 input variable, can use up to 6 variables)

Var 1	Var 2	Var 3	Var 4	Var 5	Var 6
tvhours					

RECODING RULES (See explanation and examples)

OUTPUT Variable	VALUES of the INPUT Variables						
Value	Label	Var 1	Var 2	Var 3	Var 4	Var 5	Var 6
1	<5 hrs	0-4					
2	5+ hrs	5-24					

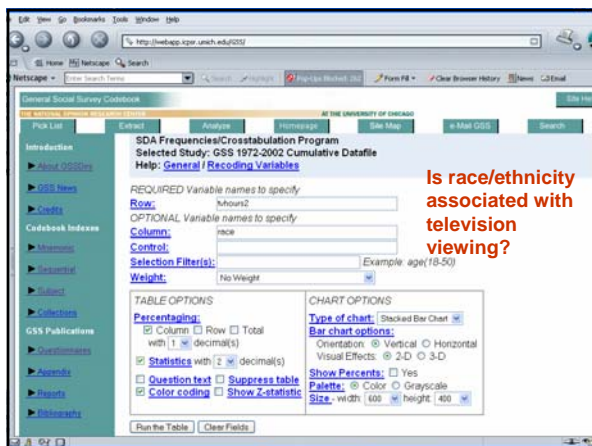


Examine Trends

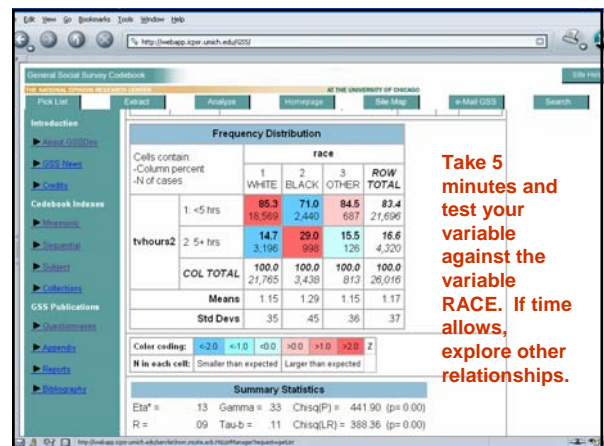
- Everyone take 10 minutes to see if your variable of interest shows any relationship the variable YEAR (the date in which the respondent took the survey). You may want to recode your data if it is continuous, and possibly restructure the variable YEAR. If you do not have multiple years represented, you may want to select another variable to explore. If time allows, explore longitudinal relationships with other variables in your subject area.

Comparing Subgroups in the Population

- Sex
- Race
- Income
- Age
- Political Orientation
- Religiosity
- Sexual orientation
- etc

Is race/ethnicity associated with viewing?



Take 5 minutes and test your variable against the variable RACE. If time allows, explore other relationships.

Frequency Distribution

	race			ROW TOTAL
	1 WHITE	2 BLACK	3 OTHER	
1 <=5 hrs	85.3 18,569	71.0 2,440	84.5 697	83.4 21,696
2 5+ hrs	14.7 3,196	29.0 965	15.5 126	16.6 4,320
COL TOTAL	100.0 21,765	100.0 3,438	100.0 813	100.0 26,016
Means	1.15	1.29	1.15	1.17
Std Devs	.35	.45	.36	.37

Cells contain: Column percent, N of cases

Color coding: <2.0 <1.0 <0.0 >0.0 >1.0 >2.0 Z

N in each cell: Smaller than expected Larger than expected

Summary Statistics

Eta² = .13 Gamma = .33 Chsq(P) = 441.90 (p= 0.00)
 R = .09 Tau-b = .11 Chsq(LR) = 388.36 (p= 0.00)

Why do African Americans watch more TV?
 Can we test this with multivariate methods?

SDA Frequencies/Crosstabulation Program

Selected Study: GSS 1972-2002 Cumulative Datafile
 Help: General / Recoding Variables

REQUIRED Variable names to specify
 Row: thours2
 OPTIONAL Variable names to specify
 Column: race
 Control: wkstat
 Selection Filter(s):
 Weight: No Weight

TABLE OPTIONS
 Percentaging: Column Row Total with 1 decimal(s)
 Statistics with 2 decimal(s)
 Question text Suppress table Color coding
 Show Z-statistic

CHART OPTIONS
 Type of chart: Stacked Bar Chart
 Bar chart options:
 Orientation: Vertical Horizontal
 Visual Effects: 2-D 3-D
 Show Percents Yes
 Palette: Color Grayscale
 Size: width 600 height 400

Frequency Distribution

	wrkstst						
	1 WORKING FULLTIME	2 WORKING PARTTIME	3 TEMP NOT WORKING	4 UNEMPL, LAID OFF	5 RETIRED	6 SCHOOL	7 KEEPING HOUSE
1 <=5 hrs	91.8 12,524	85.5 2,367	85.8 490	69.5 957	72.6 2,477	83.5 624	69.6 3,279
2 5+ hrs	8.2 1,106	14.5 401	14.2 81	30.5 245	27.4 936	16.5 123	30.4 1,433
COL TOTAL	100.0 13,430	100.0 2,768	100.0 571	100.0 802	100.0 3,413	100.0 747	100.0 4,712
Means	1.08	1.14	1.14	1.31	1.27	1.16	1.30
Std Devs	.27	.35	.35	.46	.45	.37	.46

Cells contain: Column percent, N of cases

Color coding: <2.0 <1.0 <0.0 >0.0 >1.0 >2.0 Z

N in each cell: Smaller than expected Larger than expected

Summary Statistics

Eta² = .26 Gamma = .46 Chsq(P) = 1,834.96 (p= 0.00)
 R = .25 Tau-b = .23 Chsq(LR) = 1,792.44 (p= 0.00)

SDA Frequencies/Crosstabulation Program

Selected Study: GSS 1972-2002 Cumulative Datafile
 Help: General / Recoding Variables

REQUIRED Variable names to specify
 Row: thours2
 OPTIONAL Variable names to specify
 Column: race
 Control: wkstat
 Selection Filter(s):
 Weight: No Weight

TABLE OPTIONS
 Percentaging: Column Row Total with 1 decimal(s)
 Statistics with 2 decimal(s)
 Question text Suppress table Color coding
 Show Z-statistic

CHART OPTIONS
 Type of chart: Stacked Bar Chart
 Bar chart options:
 Orientation: Vertical Horizontal
 Visual Effects: 2-D 3-D
 Show Percents Yes
 Palette: Color Grayscale
 Size: width 600 height 400

Statistics for wrkstst = 1 (WORKING FULLTIME)

	race			ROW TOTAL
	1 WHITE	2 BLACK	3 OTHER	
1 <=5 hrs	93.3 10,134	62.4 1,349	89.8 416	81.8 11,899
2 5+ hrs	6.7 729	17.6 289	10.2 47	8.2 1,064
COL TOTAL	100.0 10,863	100.0 1,637	100.0 463	100.0 12,963
Means	1.07	1.18	1.10	1.08
Std Devs	.25	.38	.30	.27

Cells contain: Column percent, N of cases

Color coding: <2.0 <1.0 <0.0 >0.0 >1.0 >2.0 Z

N in each cell: Smaller than expected Larger than expected

Summary Statistics for wrkstst = 1 (WORKING FULLTIME)

Eta² = .13 Gamma = .42 Chsq(P) = 226.02 (p= 0.00)
 R = .10 Tau-b = .12 Chsq(LR) = 194.61 (p= 0.00)
 Somers' d² = .09 Tau-c = .05 d² = .2

Take 5 minutes and introduce a control variable into your analysis, or test a relationship that someone else reported that you believe may be spurious.

SDA Logit/Probit Regression Program

Selected Study: GSS 1972-2002 Cumulative Datafile
 Help: General / Dummy vars / Product terms

Type of Regression: Logit Probit
 Dependent: thours2
 Independent: (You can tab from one input box to the next)
 1. wkstat (d1.2*work) 2. race (d2*black+1) 3. sex 4. year
 5. 6. 7. 8.
 9. 10. 11. 12.
 13. 14. 15. 16.

More independent variables
 Selection Filter(s):
 Weight: No Weight

Other statistics:
 T-tests Univariate stats
 Color coding Question text

Change number of decimal places to display:
 For coefficients and t-tests: 3

Variables					
Role	Name	Label	Range	MD	Dataset
Dependent	tvhours2(Recoded)	5 or more hours of tv per day	0-1		2
Independent	wrkstat(d: 1,2)	LABOR FRCE STATUS: working =1	0-1		1
Independent	race(d: 2)	RACE OF RESPONDENT (1972-2000): black=1	0-1		1
Independent	sex	RESPONDENTS SEX	1-2		1
Independent	year	GSS YEAR FOR THIS RESPONDENT	1972-2002		1

Logit Coefficients					Test That Each Coefficient = 0	
	B	SE(B)	T-statistic	Probability		
wrkstat	-1.304	.036	-35.819	.000		
race	.870	.044	19.832	.000		
sex	.119	.037	3.252	.001		
year	.001	.002	.255	.799		
Constant	-2.448	4.479	-.547	.585		

SDA Multiple Regression Program
 Selected Study: GSS 1972-2002 Cumulative Datafile
 Help: [General](#) / [Dummy vars](#) / [Product terms](#)

Dependent: tvhours

Independent: (You can tab from one input box to the next)

1 tvhours (d:1,2) recd	2 race (d:2) black=1	3 sex	4 year
5	6	7	8
9	10	11	12
13	14	15	16

More independent variables

Selection Filter(s): Example: age(18-50)

Weight: No Weight

Other statistics:
 T-tests Global F-test Univariate stats
 Correlation matrix Covariance matrix
 Color coding Question text

Run Regression Clear Fields


Change number of decimal places to display:
 For coefficients: 3

Variables					
Role	Name	Label	Range	MD	Dataset
Dependent	tvhours	HOURS PER DAY WATCHING TV	0-24	-1,98.99	1
Independent	wrkstat(d: 1,2)	LABOR FRCE STATUS: working =1	0-1		1
Independent	race(d: 2)	RACE OF RESPONDENT (1972-2000): black=1	0-1		1
Independent	sex	RESPONDENTS SEX	1-2		1
Independent	year	GSS YEAR FOR THIS RESPONDENT	1972-2002		1

Regression Coefficients						Test That Each Coefficient = 0	
	B	SE(B)	Beta	SE(Beta)	T-statistic	Probability	
wrkstat(d:1,2)	-1.253	.028	-.269	.006	-44.605	.000	
race(d:2)	1.027	.040	.153	.006	25.895	.000	
sex	.028	.028	.006	.006	.997	.319	
year	.003	.002	.009	.006	1.479	.140	
Constant	-1.583	3.462			-.457	.648	

Conclusion and Discussion

- A Deweyian approach to quantitative literacy
 - Engages students in the construction of questions
 - Challenges students to generate answers to their own questions
 - From this approach, students learn both skills and the value of quantitative literacy.



Discussion – What are the concerns of using such an approach in the context of mathematics courses, or as the QL standard for a bachelor's degree?