Integrating Data into the Curriculum
Qualitative Data Experience

With this assignment, you will experience the thrill of "doing" sociology rather than just reading sociological findings. We will do most of this project together, in class, but you each will be responsible for completing one interview on your own, and sharing your typed transcripts in file form with the entire class.

Day 1: We will break into four groups (each group will have five students). We have been reading Scott Coltrane's *Gender and Family* book and he has presented you with findings from many different research projects about the enduring presence of gender stratification in family life. Some of you have been convinced by his arguments and others have not. Here is your chance to identify a research area that interests you about families and inequality (gender or another form of inequality) and find out for yourselves. Each group will have 10 minutes to come up with two areas of general interest that they would like to explore in interview data. We will then re-convene and share our interests with one another. Before the end of the class we will vote on the area to pursue. We will also identify a research sample, that is, who we want to study. Each student will be required to formulate a hypothesis about the general topic (with an independent, dependent and control variable) and a set of questions that could be used to test the hypothesis. That assignment will be do at the next class.

Day 2: We will list all the hypotheses on the board. We will decide which of the hypotheses to investigate and the instructors will use the appropriate questions to create an *interview schedule* (that is what sociologists call the list of questions we use during the interview).

Day 3: The instructors will return to class with the *interview schedule* that we can all use and we will practice interviewing one another, and critiquing our techniques. We will work on learning to be comfortable, and when and how to probe.

You will have two weeks to do an interview, type the transcript and email to the rest of the students in your group. Each of you is responsible for reading each transcript (all 20!). Your group will need to have all 20 interviews printed for your work in class. Organize inside your group so that each of you print out five (different) interviews. Each student should be assigned to print out five of them and bring them to class on Day 4 of this exercise.

Day 4: Each group will be assigned ONE hypothesis to test. In class, you will meet and talk about your findings. Each group will be responsible for producing a one or two page double-spaced paper about their findings. You will have one week.

Day 5: Each group will present their findings to the class, and submit their collective paper about the findings.

GRADING:
This assignment is worth 15 points. Transcripts are worth 10 points, group projects 5 points.

Sociology of the Family
Summer 2004

Learning to Read Tables: "Predictors of Family Structure"

This assignment will require you to locate three tables on-line (detailed directions are provided), print them out, interpret the data in the tables, and answer several questions on what you have learned from the tables. You may work together to create the tables. You may NOT work together to interpret the tables and answer the questions. Most of your answers to the questions will simply be percentages. However, question 7 is a short answer question and question 13 is a discussion question. Please limit your answer to question 13 to under a page.

Your answers must be typed, 12-point font, and double-spaced. Attach the copies of your three tables to the back of your assignment.

This assignment will help you learn how to:
- ❑ Test hypotheses using quantitative data
- ❑ Produce a bivariate table (bivariate means having two variables)
- ❑ Evaluate the implied hypothesis (is the hypothesis supported by the data?)
- ❑ Provide bivariate tables with a control variable
- ❑ Evaluate the hypothesis under the conditions of the control variable

We will be analyzing 1990 census data. It is easy to find using these simple directions! If it is not working for you, check your spelling.

1. Load WebChip. The URL is http://www.ssdan.net/ida_resources.shtml (don't miss the "s" before "html"!)
2. Click on "WebChip Launcher".
3. Using this screen, you will activate a data set.
4. Using the MANUAL launcher, on the line that says "datasets/pc", type "centrend/fmin5090.dat" so that it reads: "datasets/pc/centrend/fmin5090.dat". We will be using three variables from these data sets: family type ("FamType"), race ("RaceLat"), and family income ("FmIncome").

Conceptualizing the variables: Definitions used by the Census Bureau

Family Income ("FamIncome") is the combined earnings of adults in the household expressed as an annual amount.

- ▪ <15K- less than $15,000/year
- ▪ 15-25K- between $15,000 and $25,000/year
- ▪ 25-35K- between $25,000 and $35,000/year
- ▪ 35-50K- between $35,000 and $50,000/year
- ▪ 50K+ more than $50,000/year

Family Type ("FamType") indicates the type of household defined by gender and marital status.

- MrrdCpl- married couple household
- MaleFam- male headed household, no wife/mother present (i.e. single parent family)
- FemlFam- female headed household, no husband/father present

Race (Race) individual's self-identification as:

- Black- all persons who indicated their race as black
- NonBlack- all persons who indicated their race is not black

Creating Table 1: Family Type by Race

1. Make sure to click on "%Down" at the far left of the screen. This means you will read down the columns on your tables. Notice how if you add up all the percentages in each column, you will get 100%. (Helpful hint: Columns are vertical, rows are horizontal)

2. You want your independent variable as your column variable and your dependent variable as your row variable. Race is your independent variable and family type is your dependent variable (remember, variables such as race, age, and biological sex will not be dependent variables).

3. Click on "CrossTab". Voila!

Now that you have your data, answer the following questions:

1) What percent of NonBlacks live in the family type, married couple (MrrdCpl)?
2) What percent of Blacks live in the family type, married couple?
3) In what family type are NonBlacks <u>most</u> likely to live?
4) In what family type are Blacks <u>least</u> likely to live?
5) In what family type are Blacks <u>most</u> likely to live?
6) What family type is most rare for both NonBlacks and Blacks?
7) Does race predict family type? That is, does the data support the hypothesis:
   Race → Family Type

Creating Tables 2 and 3: Family Type by Race, Controlling for Family Income

What is a control variable? A control variable is a conditional variable. It allows us to look at the relationship between two variables under different conditions.

Example:
We could look at the relationship between gender and income controlling for marital status. In this case, we could look at the effect of gender on income for married people in one table and non-married people in another table. We would be able to see if gender had the same effect on income for married and non-married people. Marital status, then, is the condition under which we are the testing the relationship between gender and income.

How to create a table with a control variable:

1. Select your row and column variables (same as before).
2. Click on your control variable found in the box next to "Control by". Since we want to control by family income, click on "FmIncome".
3. After you click on your control variable, click on "Control by" and then "CrossTab". (You must click in this order: "FmIncome", "Control by", "CrossTab", or the program will not run).

Present your findings for the poorest family type (<15K) in Table 2

Present your findings for the richest family type (50K+) in Table 3

Using this data, answer the following questions:

8) When family income is 15K or less, what percent of NonBlacks live in the family type, married couple (MrrdCpl)?
9) When family income is 15K or less, what percent of Black live in the family type, married couple?
10) What percent of Blacks live in the family type, married couple, when family income is 50K or more?
11) What percent of NonBlacks live in the family type, married couple, when family income is 50K or more?
12) Evaluate the following hypothesis. Does your conclusion hold for the poorest group and the richest group?
Race → Family Type

Question 13 (length should be about a page):

a. Other than the fact that many groups are lumped together under "Black" and "NonBlack", what is your best explanation for these results?
b. What other questions do these results raise? (While these results definitely raise value questions, e.g. is this "right" or "wrong", try to formulate sociologically testable questions).
c. What other hypotheses might you test to answer those questions?

Creating Univariate and Bivariate Tables

We will be analyzing Census data in order to:
> Evaluate hypotheses using both quantitative and qualitative data.
> Use simple statistics to analyze social data.
> Develop skills in using evidence.

I. WebChip
For this assignment, our first goal is to gain familiarity with WebChip. WebChip is an internet based system which includes data access and data analysis. You may enter WebChip under Internet Explorer or Netscape.

1. Load WebChip. The URL is http://www.ssdan.net/ida_resources.shtml.
   (See Attachment 1A for what the page will look like.)
2. Click on WebChip Launcher.
3. Using the next screen, you will activate a data set. On the line that says "datasets/pc" select "cen1990",
4. Using the next screen to select the data we will use, select "earn9.dat" by hitting the Select Target button.
5. This will produce a page like attachment 1B. This data set has 4 variables: income (Earning), race (RaceLat), gender (Gender) and age of worker (WkAge)

II. Variables
Although there are several ways in which these variables may be conceptualized, defined and measured, these are the definitions used by the Census Bureau.

Income (Earning) – the money a person makes from working, as wages, salary, or a form of self-employment, expressed as an annual amount.

Race (RaceLat) – individual's self-identification as:
- **Non-Latino White (NLWhite)** – all persons who indicated their race as white and not of Latino origin.
- **Black** – all persons who indicated their race as black.
- **Latino (Hispanic origin)** – persons of white or "other" races who identified themselves as Mexican, Puerto Rican, Cuban, or Other Spanish/Hispanic.
- **Asian (or Pacific Islander)** – includes all persons who indicated their race or ethnicity as Chinese, Filipino, Japanese, Asian Indian, Korean, Vietnamese, Cambodian, Hmong, Laotian, Thai, or other Asian as well as Hawaiian, Samoan, Guamanian or other Pacific Islander.
- **American Indian (AmIndian)** – all persons who classified themselves as American Indian, Eskimo or Aleut.
- **Other (NLOther)** – includes persons who indicated their race as "other" and are not of Latino origin.

Gender (Gender) - individual's self-identification as either male or female.

III. Frequencies

To get a listing of the variables and their frequencies, go to the left top of the screen and "click on *"Marginals,"* you should get the following output on your screen:

RaceLat

NLWhit	Black	Latino	Asian	AmIndi	NLOthe	Total
80.4	9.7	6.5	2.8	.6	.1	= 100.0%

Gender

Male	Female	Total
61.4	38.6	= 100.0%

WkAge

16-24	25-34	35-44	45-54	55-64	65+	Total
9.	31.	28.6	18.9	10.4	2.1	= 100.0%

Earning

<15K	15-25K	25-35K	35-50K	50K+	Total
22.2	30.4	21.1	15.1	11.1	= 100.0%

Print this table and keep it in your notes.

This is a frequency table (also called a univariate table because we are looking at one variable at a time) for all full-time workers in 1990. According to the table, 61.4% of all full-time workers are male as opposed to only 38.6 who are female. 22.2% of full-time workers earn $15,000 a year or less while 11.1% earn $50,000 or more. 80.4% of full-time workers are non-Latino white, 9.7% are black, 6.5% are Latino and 2.8% are Asian.

The frequency table allows one to get an overall sense of the distribution or a particular variable or set of variables which is an important place to start. However, what we are interested in exploring further is the impact of gender and race on earnings. That is, do men typically earn more than women? What racial group typically has the highest earnings? Which group has the lowest earnings? Do Blacks earn more than Latinos? In order to address this question, we will need to cross tabulate the variables of interest.

IV. CrossTab

1. In order to do the crosstabulation of two variables, you want to be sure that you know how the variables are associated. It makes sense to say that one's gender may influence his or her earnings, but it does not make sense to say that one's earnings influences whether one is male or female. The variable that influences or affects another variable is known as the independent variable (x) and the variable that is influenced or affected by another variable is called the dependent (y) variable. You

can write this as: x → y. In this case, we are interested in how gender influences earnings. Gender is the independent variable (x) and earning is the dependent variable (y). Gender → Earning. In other words, earnings to some extent, depends on gender.

2. To create a crosstabulation table in WebCHIP, you need to tell the program which is the independent and which is the dependent variable. First make sure that the circle by "%Down" at the far left of the top of the page is marked. (See Attachment 1C) Choose the variable that is to be the independent variable by clicking on the arrow under "Column Variable." Click on "Gender". Choose the dependent variable by clicking on the arrow under "Row Variable" and choosing "Earning." Now click on "Crosstab", on the top, right of Attachment 1C. You should get the following table:

	Male	Female	All
<15K	15.9%	32.3%	22.2%
15-25K	25.6%	38.0%	30.4%
25-35K	23.0%	18.1%	21.1%
35-50K	19.4%	8.3%	15.1%
50K+	16.0%	3.3%	11.1%
100%=	45,544,808	28,646,328	N = 74,191,128

In this case, we can see that 16.0% of male full-time workers make $50,000 or more a year as opposed to only 3.3% of females. On the other hand, 32.32% of female full-time workers earn under $15,000 a year as opposed to only 15.9% of male full-time workers.

Now examine how race influences earnings, following the same process as step 2 above. (Start by deciding which are your independent and dependent variables.) You should be able to complete the table in order to answer the following questions.

Earnings	NLWhite	Black	Latino	Asian	AmIndian	NLOther	All
<15K							
15-25K							
25-35K							
35-50K							
50K +							
100%=	59660684	7174617	4842169	2056312	414470	42881	74191128

1. What racial group has the highest earnings?
2. What racial group has the lowest earnings?
3. Since these are all full-time workers, what factors might contribute to the observed

racial differences in earnings?
    4. What, if anything, do you find most surprising about these findings?

(Adapted from Timothy Thornton's "Census Data Assignment")