Finding Data for your Class Paper, Thesis, or Dissertation

Heidi Lyons
October 19, 2009
You want to study what!?!?

• Research Question
  – Ask A Professor
    • Go bounce ideas off of them.
    • I am interested in and you are interested in…….
    • Know what different professors study.
      – Research Projects
      – Research Affiliates
Other ways to get topics

• Class Syllabus or Class Discussion
• Books Chapters
  – What is missing in the literature?
  – justification cite
Next Step…

• Is it possible with secondary data?
You are a slave to your data…Kinda

• One of the biggest mistakes people make is they pick a research question that they cannot answer with a quality dataset.
  – Hate groups are interesting but….
• Go talk to your teacher/advisor before you get too far into the project.
You have a topic and it is approved...Now What?

- Go talk to someone that researches that topic.
- Look at papers that have been published on that topic.
  - What datasets are they using?
  - If there are only obscure datasets being used on the topic it might be a warning sign.
What is a good dataset anyway?

• Has it been used before?
• Is it nationally representative? If not is there a nationally representative dataset you could use? If there is you will need to justify why you are using that dataset.
• Does it have a good reputation?
Once you pick a dataset

- Go talk to someone who has used that dataset to find out what is wrong with it.
- What are some of the datasets you have used? What are the pros/cons?
In House Datasets

• TARS
• OLS
• Marriage Matters

• YOU NEED PERMISSION TO USE THESE DATASETS!
ICPSR

• Need to setup an account
• Description
• Downloads Available
• Related Literature-who has published using this data (can be old but a good start)
• ICPSR
Commonly Used Data

• Add Health
• Fragile Families
• NSFG
• NLSY79/97
The National Longitudinal Study of Adolescent Health (Add Health)

- Three Waves of Data
  - Wave I: 1995- nationally rep of 7th-12th grade. Parents were interviewed too (about 20,745)
  - Wave II: 1996-almost 15,000 interviewed
  - Wave III: 2001 they are 18-26 years old N=15,197 plus 1,507 romantic partners
Add Health

• Physical and Mental Health Measures
• Personal traits
• Families
• Friendships
• Romantic relationships
• Peer groups and schools
• Neighborhoods, and communities.
Public vs. Restricted

- In-School Questionnaire
- Wave I In-Home Interview
- AHPVT
- Wave I Parent Questionnaire
- Wave II In-Home Interview
- Contextual data
- In-school network data

- Wave I Adolescent In-Home Interview with AHPVT, Parent In-Home Questionnaire, and Adolescent In-School Questionnaire data attached
- Adolescent In-School Questionnaire
- Wave I, II School Administrator Questionnaire
- Wave II In-Home Interview
- School Information Data
- Wave III In-Home Interview, with field interviewer characteristics and STD
CFDR can help you get access

- Email the CFDR to get access to the restricted data.
- Sign consent form.
- **Only use computers with secure server.**
  - You cannot take these data with you.
- **Add Health**
Fragile Families

• 1. What are the conditions and capabilities of unmarried parents, especially fathers?
• 2. What is the nature of the relationships between unmarried parents?
• 3. How do children born into these families fare?
• 4. How do policies and environmental conditions affect families and children?
Fragile Families

• Cohort of 5,000 children born between 1998 and 2000 with over-sample of non-marital births in 20 large US cities.
• Interviews with mother and fathers at birth, 1, 3, and 5.
• Parent interviews: attitudes, relationships, parenting behavior, demographics, health, economic and employment status, neighborhood qualities, and program participation.
Fragile Families

- 4,898 families
- About 3,700 are unmarried when the child was born
Fragile Families

• Public data is available by completing a brief application and a 25-word abstract about your research project.

• **Fragile Families Home Page**

• http://www.fragilefamilies.princeton.edu/documentation/FF_NCMR_ICPSR.pdf
NSFG 6

- 2002 is nationally representative of 12,571 respondents ages 15-44.
- Includes both men and women.
- Provide reliable national estimates on marriage, divorce, contraception, parenthood, infertility, and the health of men, women and infants in the United States.
- Very health behavior oriented.
NSGF 6

- NSFG Cycle 6 Webdocs
NLSY79

- 12,686 young men and young women who where 14 to 22 years of age in 1979.
- The NLSY79 public-use file includes data from rounds 1-20 (1979-2002).
- Questions about schooling, moved out of their parental homes, labor market, served in the military, and married and having children.
- It details the life course representative of all American men and women born in the late 1950s and early 1960s.
NLSY 79

• Codebooks are available on CD-rom obtainable from the CFDR office. The NLSY79 errata is also available at http://www.nlsinfo.org/nlsy79/nlsy79_errata.php3.

• Similarly there is a NLSY 1997 – 12 to 16 years old as of December 31, 1996
Still Lost?

• CFDR can help
• CFDR Data Holdings
Even More Lost?

• Come See Us. We love talking about Data!
Thank You

• Please visit http://www.bgsu.edu/organizations/cfdr/
• Questions?
• Next Workshop:
  – ARRAYS and MACROS
    Monday Nov 2, 2009