
36-303: Sampling, Surveys and Society

Data Collection Methods
Brian Junker
132E Baker Hall
brian@stat.cmu.edu

Handouts

- Food Frequency Survey
- Lecture Notes
- HW02
- Handout on Probability for Surveys
(for next week)
- Remember to turn in on Blackboard today:
 - HW01
 - Project assignment I.1

Outline

- Office Hours
- Assignments and Readings
- Project Assignments
- Methods of Data Collection (Groves Ch 5)
 - Evolution of methods
 - Interviewer effects
 - Response Rates
 - Sampling Frame and Coverage

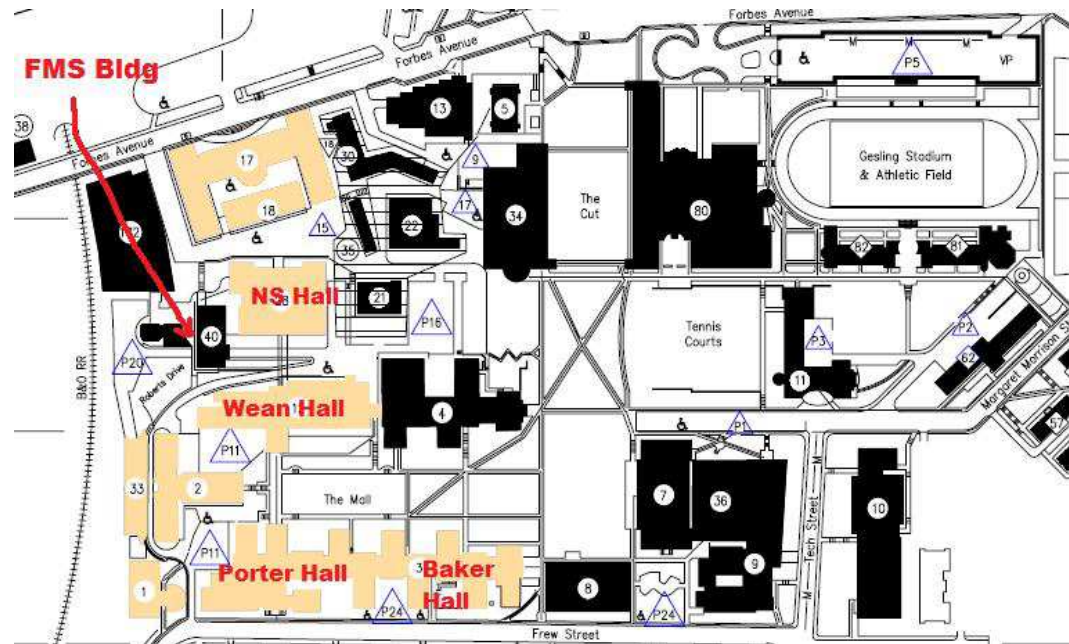
Office Hours

- Based on the in-class poll, office hours for Michael will be:

M 4-5pm

FMS 320

- Reminder: BJ is avail right after class or by appt.



Assignments & Readings

- Things to read:
 - Done already: Groves Ch 1-3
 - This week:
 - Groves, Ch 5
 - Next week:
 - Groves, Ch 11 (sections 1-6)
 - Groves, Ch 4 (sections 1-3)
 - Handout on probability for surveys
- Due Next Tues (Feb 7):
 - HW02
 - Team Assignment I.2

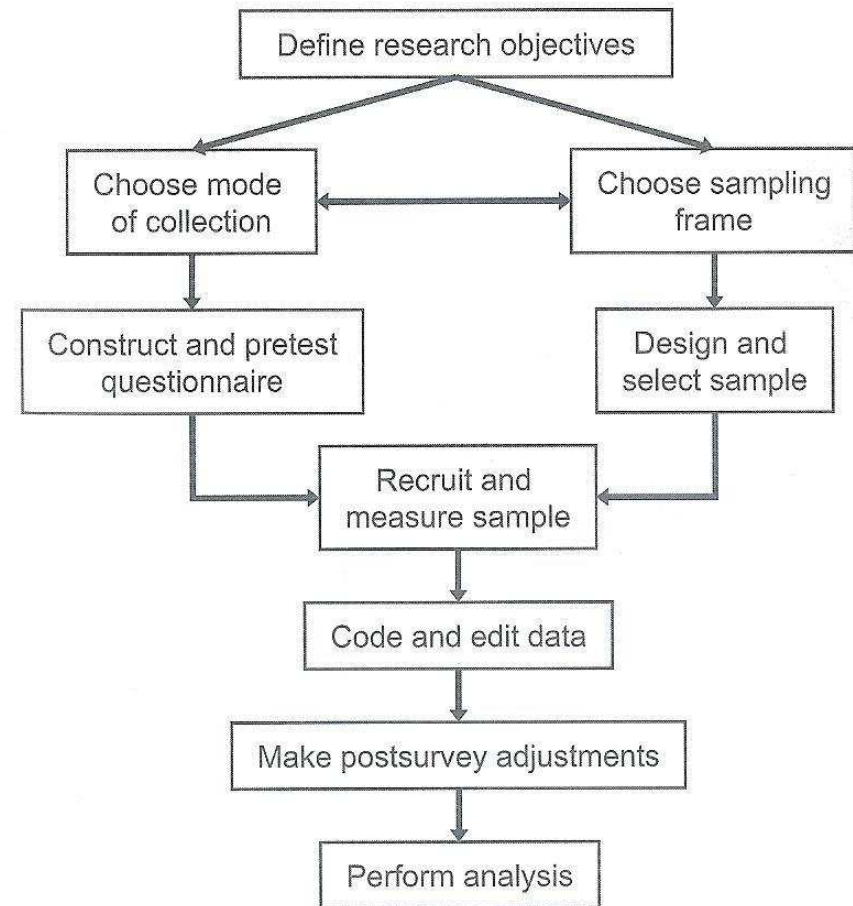
Project Assignments

- Due Today – Complete answers (for both proposed topics):
 - I.1 A, B, C: Interesting Topic -- Research Questions -- Past Research
 - I will email feedback to each group

 - Due next Tues – Complete answers (for both proposed topics):
 - I.2 D, E, F, G: Sampling Frame -- Target Population -- Mode of Data Collection -- Variables to Measure
 - Don't just turn in D, E, F, G, but rather revise A, B, C and add D, E, F, G, in a single document.
 - Do this for both of your project proposals!
-

Methods of Data Collection

- **Mode** or **method** of data collection: how to get data from each respondent?
- Must answer two questions:
 - ❑ *What method is most appropriate for your research question(s)?*
 - ❑ *How does the method impact errors and costs?*



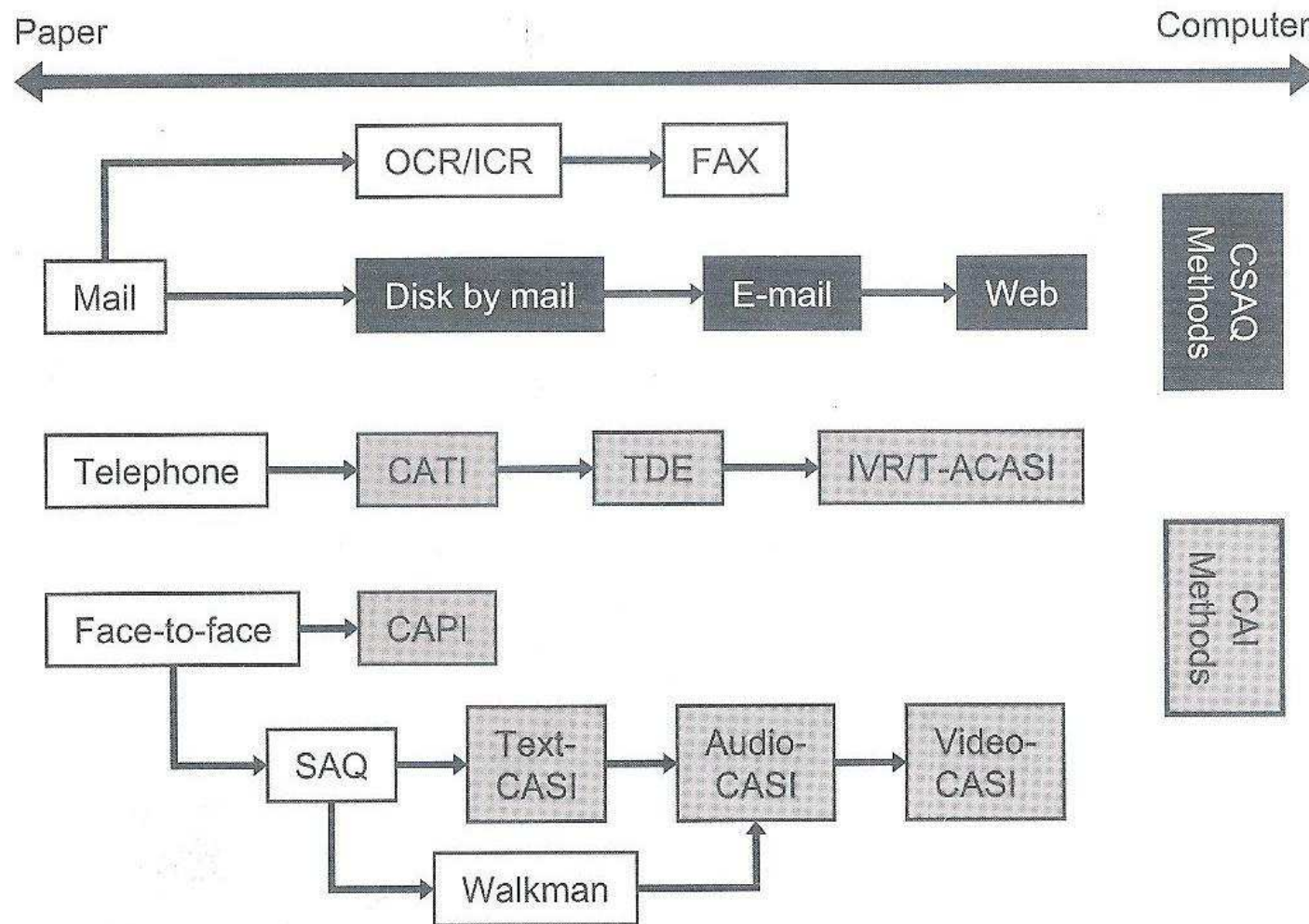
Some Typical Methods of Data Collection

- 1930's-1950's:
 - Face to Face Interviews
 - Mail Surveys
- 1960's added:
 - Telephone Surveys (vs face-to-face)
- 1980's-1990's added:
 - Disk by mail (vs paper & pencil mail surveys)
- Today, added:
 - Computer assisted methods
 - Email and Web surveys

Technology-Assisted Methods

- Computer-assisted personal interview (CAPI)
 - Interviewer reads questions aloud from laptop, enters respondent answers on laptop
- Computer-assisted telephone interview (CATI)
 - CAPI with a phone
- Computer-assisted self-interview (CASI)
 - Respondent operates the laptop
- Audio questions, keypress or voice-recog response
 - ACASI: Audio questions CASI (text/video too)
 - T-ACASI or IVR (Interactive Voice Response)
- Self-Administered Questionnaires (SAQ)
 - P&P mail surveys, email surveys
 - Computerized SAQ (CSAQ), Web surveys

Evolution of Survey Technology



Mode of Data Collection vs. Contact Method

- Some modes specifically imply the method of contacting respondents
 - email, phone
 - Other modes can be combined with different contact methods
 - CAPI, CASI, (C)SAQ, P&P
 - Go to respondent (home, business, etc.) or wait for respondent (man-on-the-street)
 - The research question can sometimes determine contact method or location
 - Observing traffic at a specific intersection
 - Observing & interviewing smokers outside a building or buildings
 - Observing gas prices posted at gas stations
-

Degree of Interviewer Involvement

- Can be deep or non-existent
 - Interviewer reads questions, records responses
 - Mail-in questionnaire (no interviewer at all)
- Can blend approaches
 - SAQ as part of an interview
 - (A)CASI for sensitive questions (NSDUH drug use questions)
 - Look at this video and then I will ask you questions about your reaction...

Interviewer Benefits and Costs

■ Benefits

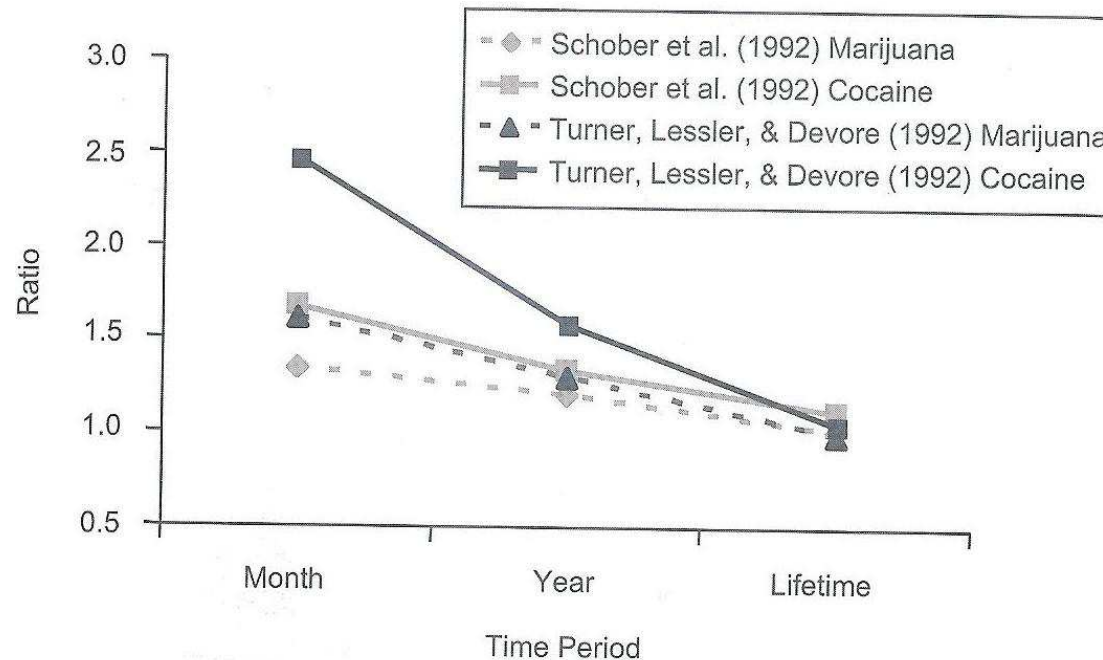
- Interviewers can be effective recruiters – higher response rates
- High interaction with respondent → more raw data per respondent → more control over measurement process
- Answering “meta-questions” about the survey can increase respondent cooperation
- Assist in clarifying, probing and motivating respondents (reduces item non-response -- higher proportion of answered questions/respondent)
- Even when no “interview” is involved, trained observers can solve data collection problems in the field
 - Gas prices survey used trained observers “at the pump”, not telephone reports
 - NAEP periodically collects academic transcripts and other administrative information from schools
 - Seek personal permission from farmer to do crop measurements

Interviewer Benefits and Costs

■ Costs

- ❑ Some sensitive questions show large interviewer effects
 - Race/gender/personal affect of interviewer
 - Personal questions, social desirability questions
 - Privacy issues
 - ❑ Interviewer presence or proximity
 - ❑ Proximity of other respondents
 - ❑ Randomized response methods
- ❑ Accessibility/comfort level of locations, interviewer fraud, ...
- ❑ Requires a corps of trained, equipped, motivated interviewers
 - Interviewers require supervision, monitoring and support
 - Large portion of survey cost and survey organizational effort
- ❑ Costs of transcribing and coding complex interviewer interactions

Interviewer Effects for Sensitive Questions



- Ratio of Proportion of respondents reporting illicit drug use in self-administered questionnaire vs interviewer-administered questionnaire (Groves Ch 5)

As We Remove the Interviewer from the Process

- Channels of Communication
 - Figures and Text
 - Audio
 - Video
 - Social Presence
 - Is respondent “socially focused” on the survey, or multitasking?
 - Order effects
 - Primacy: the first choice is preferred
 - Recency: the last choice is preferred
 - First try at an answer
 - Respondent’s first try may be “socially desirable” or not thought out
 - Giving respondents the ability to backtrack/review/revise answers reduces “first try” effects
-

Response Rates

- Hager et al. (2003, *Nonprofit and Voluntary Sector Quarterly*, pp 252-267):

What constitutes an acceptable return rate? Research methods textbooks differ in their guidelines, although most argue strongly for securing a high rate of return as a means of minimizing nonresponse bias. Babbie (1990) contended that a return of 50% is adequate, although Bailey (1987) set the adequacy bar at 75%. More recently, Schutt (1999) instructed students that “a response rate below 60 percent is a disaster” and concluded that “*it is hard to justify the representativeness of the sample if more than a third failed to respond*” (pp. 254-255).

- The key is to be able to justify representativeness. This is much easier if the response rate is higher.
- *Nonresponse Bias*: How different are the responders from the nonresponders?
 - *Big issue for Web surveys and other self-selected samples!*

Response Rates: FTF, Telephone, Mail

- Generally speaking

- Face to Face Interviews have the highest response rates
- Telephone Interviews have the next highest response rates

- O'Niel et al (1979)

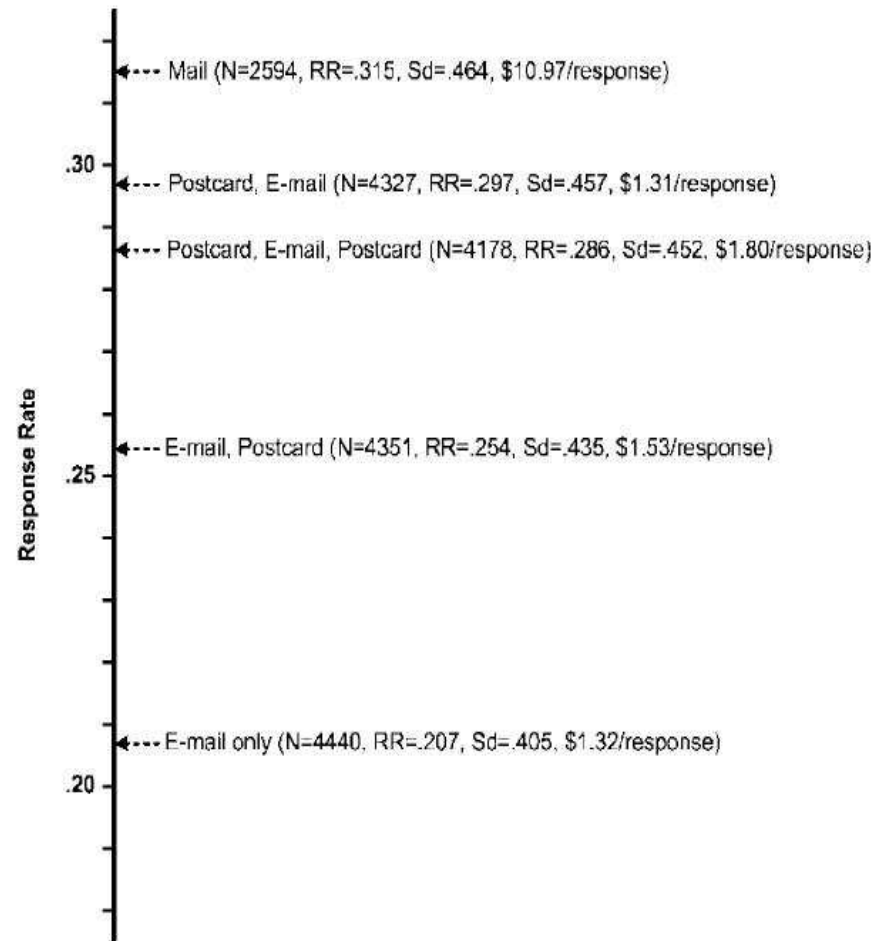
- Response rates ranging from 50% to 70% (seems high)
 - Pew (2008 Religious Landscape Survey) reports 24%
- Most refusals in the first few minutes, so careful how you start!
- Mail-in surveys generally lower response rates than FTF or telephone

- Key issues [for mail, FTF, telephone, and email]

- Motivation of respondents
- Advance notification, number of contact attempts, followup of nonresponders

Response Rates: Email vs Snail Mail

- Kaplowitz et al (2004, *Public Opinion Quarterly*, pp 94-101)
- Compared mail-in vs email surveys for Michigan State undergraduates with excellent email access
- This is a “best case” scenario!



Some Interesting Response Rate Data

■ **Office Hours Poll:**

- ❑ Last year I did a doodle poll:
14/42 = 33% responded
- ❑ This year I did an in-class census:
30/34 = 88% responded

■ **Past 303 groups:**

- ❑ Email to invite students to take a poll at surveymonkey.com: 20-25% of those emailed actually go to surveymonkey.com
- ❑ Telephone survey, roughly 10% response rate.

Low response rates...

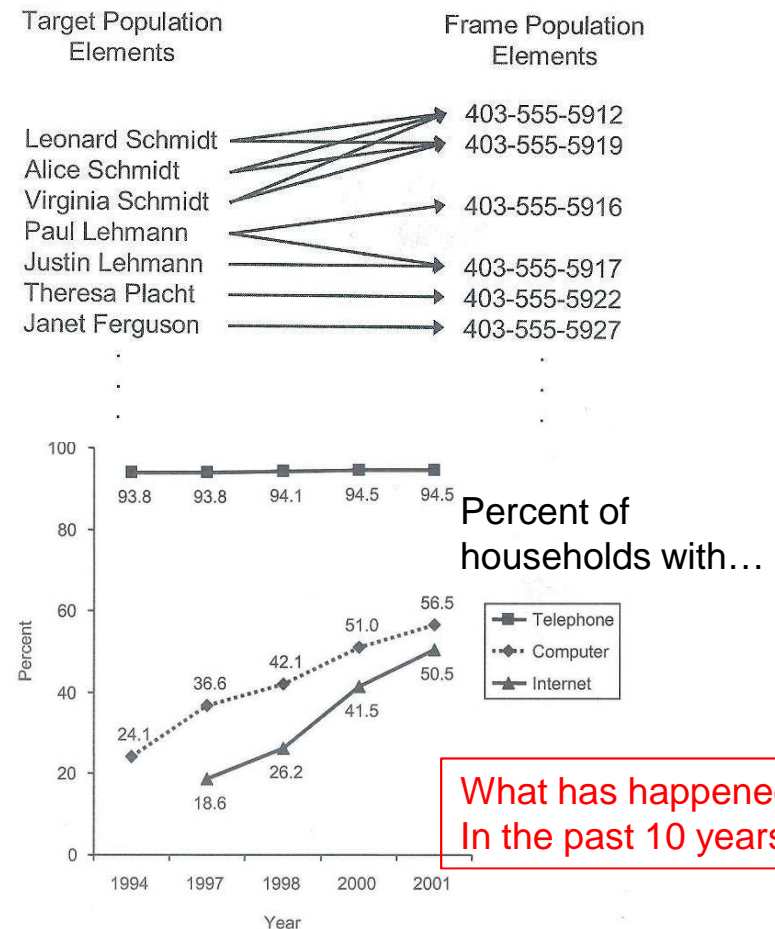
- Lower response rates (10-30%) can be OK if the respondents are like a random sample
 - We will see that a small fraction of total population is OK

- For samples that are not randomly chosen, different guidelines (Sue & Ritter, 2007, Conducting On-line Surveys, p. 34):
 - Useful sample sizes are typically 30-500
 - Within that range, sample roughly 10% of total population
 - Sample should be roughly 10 times larger than number of variables being studied
 - Choose the largest sample you can possibly afford

Sampling Frame and Coverage Issues

■ Mode affects frame and coverage

- Face to face – need accessible locations
- Telephone – how does phone book affect coverage?
- Land line vs cellphone
- Email
 - Availability, self-selection
- Web Survey
 - Availability, self-selection



Choosing the Mode

- Face to Face interviews
 - ❑ Expensive, seen most often in large scale federally funded surveys
 - ❑ Can be appropriate in small surveys if you really need to control or problem-solve the measurement process
- Comparing modes (FTF, Phone, Mail, Email, Web, ...)
 - ❑ Cost
 - ❑ Respondent Attention
 - ❑ Sensitive Questions
 - ❑ Response Rate
 - ❑ Nonresponse bias & Coverage bias

Institutional Review Board (Ethics)

- Historically IRB has been more focused on medical research than social research
- In recent years, liability concerns (risk/benefit, confidentiality, etc.) have spread IRB review to most social and survey style research
 - Studies conducted for research must undergo IRB review
 - Studies for commercial clients, for consulting, or for class credit, often do not require IRB approval
- In this class:
 - You must take & pass the CITI training (part of HW02).
 - You must complete an IRB application for your project, which I will review (team project schedule I.3 & I.6).
- I will talk more about Ethics next week

Review

- Turn In First Team Assignment on Blackboard today: *Two Possible Projects*
 - I will email you feedback this week
 - Next Tuesday, Team Assignment 1.2 is due: *Complete items A,B,C,D,E,F,G from project outline*
- HW01 due today
- HW02 assigned today
- Methods of Data Collection (Groves Ch 5)