

# Introduction

John M. Abowd  
January 2005

# Outline

- Introductions and preliminaries
- Protocols
- A little background
- Course objectives
- Course requirements

# Introductions

- Information technologies coordinator: Lars Vilhuber
- Teaching assistant: Ian Schmutte
- Producer: Jason Kenyon
- INFO 447 (undergraduate version) instructor: Warren Brown

# Protocols

- You will always see me in the distance learning frame
- You should see the lecture presentation in the VNC session active in the classroom
- If you want to ask a question, hold up the sign “QUESTION: Site” in your camera window. When I can take a question, I’ll call on you by site.

# Tools

- Course web page:  
<http://instruct1.cit.cornell.edu/courses/info747>
- Virtual RDC:
  - You will need software for this.

# The Information Technologies Research Grant from NSF

- A program that encourages innovative, high-payoff IT research and education
- Our grant proposal cited the many research studies and data products created by previous NSF support for the Research Data Center network and the Longitudinal Employer-Household Dynamics Program

# What Is It?

- \$2.9 million 3-year grant to the RDC network (Cornell is the coordinating institution)
- To provide core support for scientific activities at the RDCs
- To develop public use, analytically valid synthetic data from many of the RDC-accessible data sets
- To facilitate collaboration with RDC projects that help design and test these products

# Public Use Data Products Are the Lifeblood of Statistical Agencies

- RDC-based teams understand the public use data products produced at Census and how they relate to the underlying confidential data products
- In the demographic area there are many public use micro data products
  - But, their confidentiality protection is increasingly challenging
- In the economic area there are very few public use micro-data products
  - But all the data are used for public use aggregate products

# Public Use Data Products Are the Lifeblood of Statistical Agencies

- Integrated data products, like LEHD, produce public use summary data (QWIs) but no micro data products
  - But, synthetic data offers the possibility of releasing customized micro data products

# Principal Investigators

- Ron Jarmin, Center for Economics Studies
- Trivellore Raghunathan, University of Michigan
- Stephen Roehrig, Carnegie Mellon University
- Matthew Shapiro, University of Michigan
- I am the coordinating PI

# Project Teams and Coordinators

- Neil Bennett, CUNY Baruch
- Gail Boyd, Argonne National Laboratories
- Marjorie McElroy, Duke University
- Wayne Gray, Clark University
- John Haltiwanger, University of Maryland
- Andrew Hildreth, UC Berkeley
- Margaret Levenstein, University of Michigan
- Jerome Reiter, Duke University
- Jeremy Wu, LEHD Census
- Ray Bair, Argonne National Laboratories
- Lars Vilhuber, Cornell University

# Team Locations

- At Census, in the Center for Economic Studies and the Longitudinal Employer-Household Dynamics Program
- At the RDCs in Washington Plaza, Boston (NBER, Cambridge), California (UCLA and Berkeley), Chicago (Consortium administered by Northwestern), Ann Arbor (University of Michigan), Research Triangle (Consortium administered by Duke), New York (Cornell and Baruch)

# This Course

- Was sponsored by this NSF grant to develop a new generation of researchers who can use and produce these tools.

# Course Objectives

- Teach researchers to use the confidential microdata that underlie public use products
- Cover the legal, ethical, statistical and computational issues
- Develop the tools for editing, linking, analyzing those data
- Integrate research on the confidential data with improvement of the public use data

# Course Requirements

- Lectures
- Labs
- Familiarity with Census RDCs
- Virtual RDC use
- For a grade: either take-home final or submit a project proposal to the RDC system