

# Teaching Research and Analysis for Public Policy and Management: What Big Data Does and Does *Not* Change

Dahlia Remler

School of Public Affairs  
Baruch College, City University of New York

April 12, 2014  
APPAM Spring 2014 Conference

# Plan of Talk

- Research & Analysis for MPA/MPP
  - Courses, students/challenges, learning goals
    - Main focus on required R&A courses
- Changes in *what* we teach from Big Data?
  - New Stuff
  - Old Stuff Now More Important
  - Old Stuff Now Less Important
- Raising questions more than answering them (often)

# Research & Analysis for MPA/MPP

- All require some form of R&A:
  - Statistics, research methods for PA, quantitative methods, data analysis...
  - Variation in content, emphasis
    - Sometimes: just statistics, much or little qualitative, other quant methods besides stats, data collection
    - More variation than optimal?
  - How many courses?
    - Always 1, Usually 2, Rarely 3
    - Right amount?
- Capstone, More advanced R&A

# MPA Student Challenges

- In considering required content:
  - *Very* diverse quantitative backgrounds!
    - → Exacerbated by sequential nature of much R&A content
  - Very diverse IT backgrounds?
  - Diverse motivations (or lack) for R&A
  - Career stage variation
  - Limited time
  - Most MPAs do not become researchers or analysts!
    - Will almost everyone be analyst?

# MPA/MPP Learning Goals

- Critically consume research & analysis
  - Spot weak or invalid conclusions
  - Extract & apply relevant, valid conclusions
  - Quantitative literacy
- Commission research
  - Pose question/purpose effectively, realistically
- Perform R&A in policy/practice capacities
  - Quick, on the fly, analysis
  - Formal, traditional research
- Does Big Data change the mix?

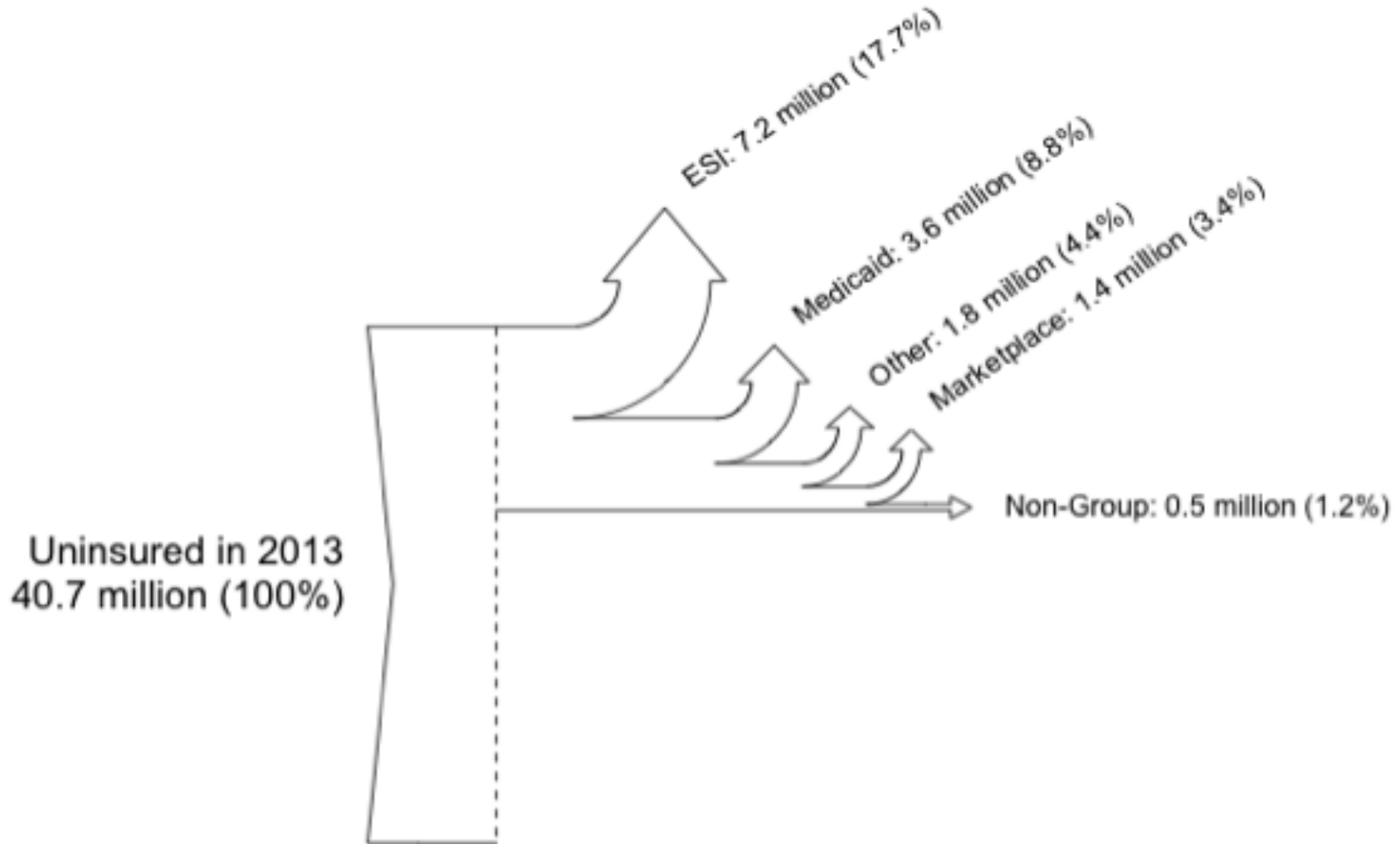
# Priorities!

- Time constraints, Student learning realities → Can't teach everything, particularly not effectively
- Can't only suggesting *adding*
- What do we prioritize?

# New Stuff from “Big Data”

- Qualitative data (existing)
  - Text...audio, video, images
  - Coding (convert to categorical or other quant data)
  - Analyzing
    - Not traditional qualitative data analysis
- Caution about “Inference” for census (not sample) data
- Data Visualization
  - Faculty need to catch up

# Image of Insurance Changes (John Graves)





# New Stuff from “Big Data”

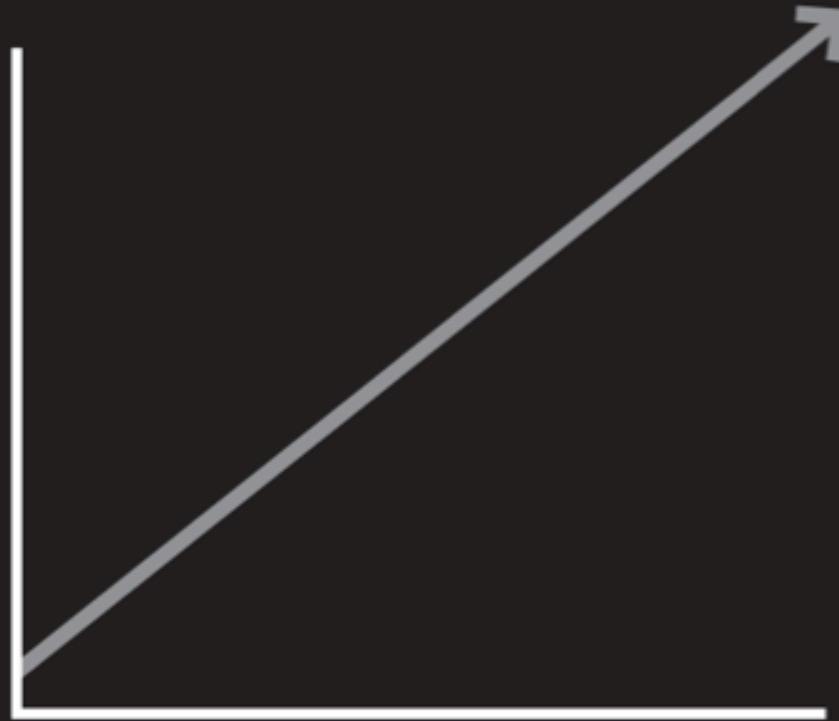
- Knowledge about new forms of data
  - “Raw” forms (what we leave on Internet)
  - How extracted
    - Ex: Scraping data off websites, pinging
  - How cleaned, merged...
  - An updated version of old message to know everything about your data
    - But much harder, more technical now
- New forms of analysis???
  - Machine learning, CS algorithms
  - Everyone will be *consuming* new analysis

# Old Stuff Now Even More Important

- Correlation vs. causation
  - Distinguish: Description, prediction, causal effects, explanation
    - Time trend correlation problem!
- Theory, mechanisms
- Measurement
  - What do you want to measure? Purpose?
  - Validity (concept vs. actual measure)
    - Google flu trends
  - Desired population vs. “sample”

# Correlations of time trends

HOW MUCH SOME PEOPLE  
THINK A CERTAIN THING



HOW OFTEN SOME PEOPLE  
DO A DIFFERENT THING

# Old Stuff Now Even More Important

- Data management
  - Data structures
- Data analysis
  - Vs. inference
  - Practice with actual data
- *Intuition* on
  - Over-fitting, Statistical significance, Multiple comparison problem
- “Everyday” analysis
  - Constantly monitoring data patterns

# Old Stuff Now Less Important

- Inferential statistics
  - But not basic intuition
  - Technical aspects, specific tests
  - Not a long list, but a lot of time and effort
- What else?
  - Either find stuff to cut or shorten... or add more courses?
  - What belongs in required courses? What in more advanced electives?

# Conclusions

- More data analysis, less inferential statistics
  - Data visualization
- Old skill “knowing your data” got a lot harder
  - Same measurement stuff in principle, a lot more IT knowledge in practice
- Correlation vs. causation, theory still very important, despite hype
- Faculty will need to learn!
- Hype may have been over-done and wrong in detail but Big Data is a big deal
- Many questions remain:
  - What to cut from required R&A? Or do we add courses?