What I heard so far ...

We are already using administrative data.

Where to start?

We don’t have data on everybody.

We need to do something now!

The data are filled with measurement error and we have no control.

Think of change management!
1. You need to know your (inferential) goal

2. Dare to combine (imperfect) data

3. Empower your team
1. You need to know your inferential goal
Inferential Goals: Products

- Description

- Causation

- Prediction
Inferential Goals: Units

- Individuals

- Blocks

- Communities ... all geocoded
1. You need to know your inferential goal

2. Dare to combine (imperfect) data
Imperfect data

- Surveys: filled with measurement error

- Administrative data: filled with error

- Found data: filled with error

... we are used to that
1. You need to know your inferential goal

2. Dare to combine (imperfect) data

3. Empower your team
Big Data for Federal Agencies
Learn how to collect data yourself, and how data are generated through administrative and other processes. Learn how to link data. APIs

Learn how to curate and manage data. Understand the advantage of databases.

Learn a variety of analysis methods suited for different data types: Network/Text

Learn how to communicate results and distribute your data. Know its limitations.

Be very clear about your research goal and investigate which data are best suited.
Designed (survey and admin) and organic data (transaction and aspirational), linkage, matching

Practical training in data base management, SQL, editing, coding, imputation, etc.

Statistical methods, machine learning, Bayesian, hierarchical, small area estimation

Visualization, disclosure control, ethics, privacy

Economics, public policy, criminology, journalism, public health, sociology, etc.
Executive training in the works: NYU – U of Chicago - UMD

- Teams: across disciplines and across academia/gov/private sector
- Learning around data + topic: ACS could be one
- Workshops on modules combined with data work in-between
- Problem + output oriented
Thank you