BIG DATA & DEVELOPMENT EVALUATION
SESSION 1

INTRODUCTION

Kerry Bruce
GROUP INTRO
Little to No Experience with Big Data

Some Experience with Big Data
First Time at MERL Tech

Been to MERL Tech before
TODAY’S OBJECTIVES

Learn about the tools and techniques of big data science

Learn about the benefits and challenges with using big data

Deepened understanding of data analytics

Guidelines for integrating big data into evaluation
WHAT IS “BIG” DATA?

Big Data

Large Data

Small Data
Huge in volume and generated very fast

Too big, usually, to process with one computer

Much of the data were collected for a different purpose than the evaluation

Typically covers the whole population (not a sample)

integrated data bases that create common fields so that many different kinds of data can be synthesized and analyzed

Data are more granular and permit more detailed disaggregation
BIG DATA IN DEVELOPMENT

**What are the sources of “big data?”**

*Often comprised of many “small datasets” pulled together*

- Monitoring data
- Baselines / Midterms / Ending / Follow-Up surveys
- Qualitative datasets
- Other secondary data
  - DHS / Health Facility Data / Surveillance data
  - Census
  - Labor statistics
- Education data (test scores, enrollment, attendance)
- Social Media data
- Transactional data (sales, telephone calls, banking transactions)
THE DATA CONTINUUM

Data Sources

- Big Data
- Large Data
- Small Data

Types of Data Analysis

- Big Data Analytics
  - Multiple sources of small and large data create integrated data bases requiring big data analytics
- Computer-based statistical analysis
- Small data analysis using quantitative and qualitative methods
  - Complementing big data with quantitative (small data) analysis
WHAT HAVE EVALUATORS USED?

Big Data

Large Data

Small Data
WHAT CAN EVALUATORS USE?

Big Data

Large Data

Small Data
LET’S ADD BIGGER DATA TO OUR TOOLBOX
DEVELOPMENT EVALUATION IN THE AGE OF BIG DATA

The dramatic expansion of digital technologies

Defining Big Data for development

The Big Data ecosystem
THE DRAMATIC EXPANSION OF DIGITAL TECHNOLOGIES

The world is more inter-connected

Every aspect of life is now captured digitally

Multiple sources of data in real-time.

Data can capture complex inter-relationships

We can now quantify and monitor ourselves
These rich sources of data are becoming increasingly accessible to individuals, researchers, businesses and DEVELOPMENT AGENCIES through all types of technologies.
THE THREE “V'S” OF BIG DATA
UN GLOBAL PULSE’S DEFINITION OF DATA FOR DEVELOPMENT

- Digitally generated
- Passively produced
- Automatically collected
- Geographically or temporally trackable
- Continuously analyzed
DATA ECOSYSTEM
Producers, analysts, users and regulators of Big Data

DATA GENERATION
Generation of new sources of data

DATA ANALYTICS
Organization, integration, analysis and dissemination of big data

EVALUATION OFFICES

AFFECTED POPULATIONS
Individuals and groups affected by how big data about them is used

THE BIG DATA ECOSYSTEM
WEAPONS OF MATH DESTRUCTION

How Big Data Increases Inequality and Threatens Democracy
THE PURPOSES OF BIG DATA

Planning, implementation and evaluation
Commercial applications
Urban and regional planning and social engineering
Research
Evaluation
Political control and social manipulation
THE ANALYTIC ACTORS
EVALUATORS

Social science methods
Research study design
Outcome measurement
Statistical analysis
Qualitative analysis
Cost-benefit analysis
Reporting

Evidence-based decision-making
DATA SCIENTISTS

- R, Python, SQL
- Basic statistics
- Machine learning
- Data transformation
- Data visualization & communication
- Software engineering

**Probability-based decision-making**
DATA ANALYTICS

- Detection/Identification
- Description/Exploration
- Prediction/Correlation
- Evaluation/Prescription