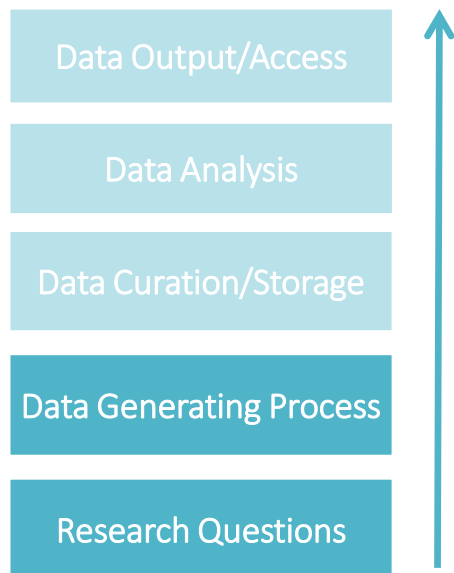
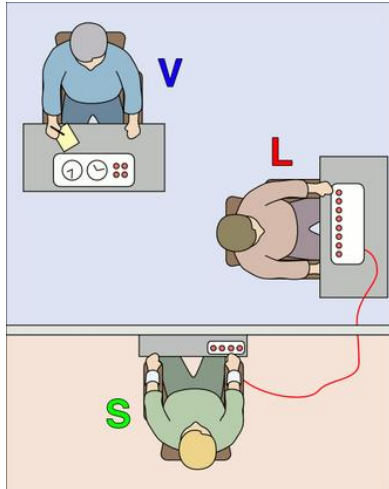


© FRAUKE KREUTER @ UMD, UMA, IAB, UMICH



© FRAUKE KREUTER @ UMD, UMA, IAB, UMICH

## Experiments (laboratory)



© FRAUKE KREUTER @ UMD, UMA, IAB, UMICH

- Systematic variation of treatment between one or more groups
- Ideally units are randomized to groups with the goal of equal distribution of characteristics across groups (no confounders)
- Often smaller in scale
- Designed for a specific research purpose
- Often difficult to balance the strong internal validity and the often weaker external validity

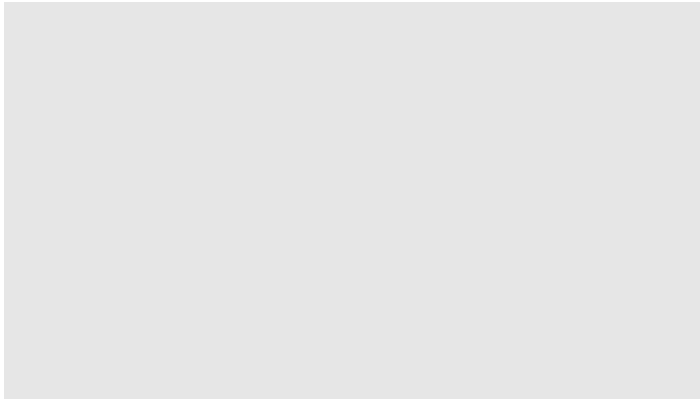
## Survey Interview Data



© FRAUKE KREUTER @ UMD, UMA, IAB, UMICH

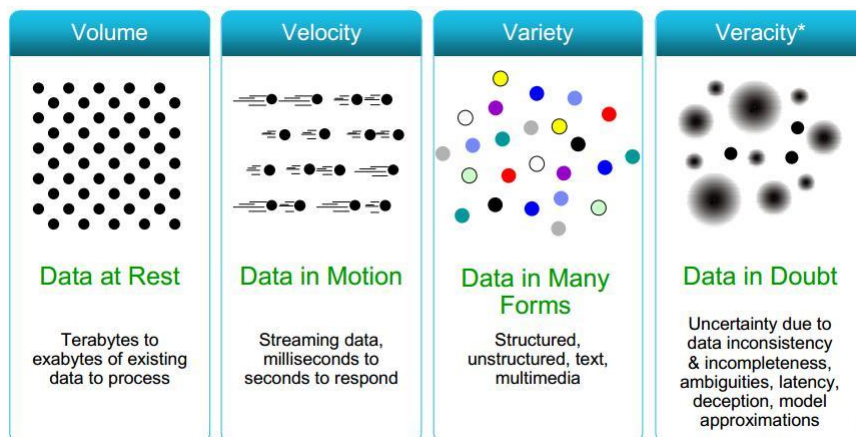
- Data collected with specific research purpose in mind
- Examples: political polling, national crime, and victimization surveys, health surveys, consumer satisfaction...
- Large variation in size and quality
- Timeliness depends in part on mode of data collection
- Often cross-section (on-off) sometimes panel studies (repeated data collection on the same units)
- Usually a sample not a Census

## “Found Data” -- Boston Street Bump



© FRAUKE KREUTER @ UMD, UMA, IAB, UMICH

## Has All of the Big Data Characteristics



<http://www.rosebt.com/blog/data-veracity>

© FRAUKE KREUTER @ UMD, UMA, IAB, UMICH

## Administrative Data

20130030201

2

3 **Steuernummer** [redacted]

4 eTIN IL Lohnsteuerbescheinigung(en), sofern vorhanden [redacted] eTIN IL weiterer Lohnsteuerbescheinigung(en) [redacted]

**Einkünfte aus nichtselbständiger Arbeit**

5 **Angaben zum Arbeitslohn** Lohnsteuerbescheinigung(en) Steuerklasse 1 - 5

6 **Steuerklasse** 168 EUR Ct

7 **Bruttoarbeitslohn** 110 [redacted]

8 **Lohnsteuer** 140 [redacted]

9 **Solidaritätszuschlag** 150 [redacted]

- Data generated as part of an administrative or production process
- Examples: insurance plans, welfare benefits, social security systems, tax forms, reimbursements for services ...
- Generated with a purpose, but primary purpose is not research, thus quality often best for program relevant (mandatory) variables (fields)
- Delay in reporting and dissemination (i.e. slow)
- Includes (most) units in program/process