Social Sciences and Ethics in CTD&S

# T: Theory

1. Identify general question to be addressed
2. What is the related theory of interest; identify categories of theory
3. Do we prove theories, substantiate them, or examine evidence?

## Ethics

1. Privacy

# G: Ways to gather data

## Conceptualization and Operationalization:

1. Move from general question or concept to dimension, variable, attribute
2. Why collected? Who collected? How collected? When collected?

## Data types

1. Data values: nominal, ordinal, interval, ratio
   1. Ratios: fractions, percentages, rates, odds
   2. Indices and scales
2. Unit of analysis
3. Variable and attributes of the variable

## Mechanisms for gathering data:

1. Target population, access frame, sample
2. Protocols: Collecting values through self-report, observe/measure, artifact
3. Methods
   1. Survey sample
   2. Experiment
   3. Administrative data
   4. Natural Experiment
   5. Observational

## Ethics

1. Subject treatment: consent, harm, inactive participation
2. Responsible parties: custodial, regularization, commoditization

# Q: Ways to assess data quality

## Bias and Validity

1. Construct valid (capture parameter of interest);
2. Face validity (concept well-represented by measurement) - measurement bias
3. Predictive validity: coverage bias, selection bias
4. Probability: means to achieve validity and reliability (in part)

## Reliability

1. Small variance (probability helps calculate variance)
2. Repeatable across time, instruments, recorders, units

## Ethics

1. Fairness and justice
2. Equality and equity

# U: Ways to understand data

## Univariate distribution:

1. Shape and meaning
2. Measure of center: choice
3. Measure of spread: choice and meaning

## Bivariate distribution:

1. Relationship shape and meaning
2. Disaggregation via side-by-side plots and super-posed plots

## Method of comparison

1. Benchmarks
2. Variation
3. Disaggregation
4. Rate: choice of denominator

## Ethics/Guidelines for informative/interpretable/meaningful plots

1. Histogram vs barplot
2. Side-by-side vs stacked bar plots and line plots
3. Choice of scale
4. Overplotting and smoothing
5. Choice of color

# Organization

## Case studies

1. Stand-alone
2. Motivating question + background
3. Cover G, Q, and U (gather, quality, understand)
4. Code included
5. Use snippets of other studies or reference other case studies or use this one case study to create assignments (possibly alternatives)

## Coverage

1. One or two chapters present Overarching / Review / Introduction of G, Q, U Concepts
2. Matrix of topics in G, Q, U covered
3. Matrix of topics in coding needed (not covered)
4. Example Sequences of case studies matched to the coding topics