

## RESEARCH DESIGN

RESEARCH QUESTION

start here

LEVEL/UNIT OF ANALYSIS

WHAT DO I WANT TO MEASURE?

DESCRIPTIVE AND PREDICTIVE ANALYSIS?

## SOURCE

### DOCUMENTS

Average text length:

Document Format (pdf, html, json, ...):

Language(s):

Structure and tidiness level:

Which preprocessing steps are absolutely necessary?

### DOCUMENT METADATA

Meta-data I want to collect:

### AVAILABILITY

Data gaps:

How to deal with it?

## MEASUREMENT STRATEGY

HOW DO I CONCEPTUALIZE IT?

### LEVEL OF MEASUREMENT

- Full document, paragraph, sentence, (span of) word(s)
- individual text
- pairs of text

### WHAT'S THE MEASUREMENT TASK?

- **classification**: texts => categories
- **scoring**: texts => scale/spectrum
- **comparison**: [text, text] => score

### INDUCTIVE OR DEDUCTIVE?

**classification**: you know the categories to which texts can belong? ◦ yes ◦ no  
**scoring/comparison**: you know the underlying dimension? ◦ yes ◦ no