Data and Experimental Design

First Semester
Senior Seminar

Typical Analytical Data Sequence

(1) Conceptualization of problem
  What data do we need and why?

(2) Data collection
  What methods to use?

(3) Processing and analyzing the data
  What statistics and methods to use?

(4) Presentation of results
  What are the results?

(5) Discussion of results
  What do the results mean?

Data - record of system observations

Qualitative

Quantitative

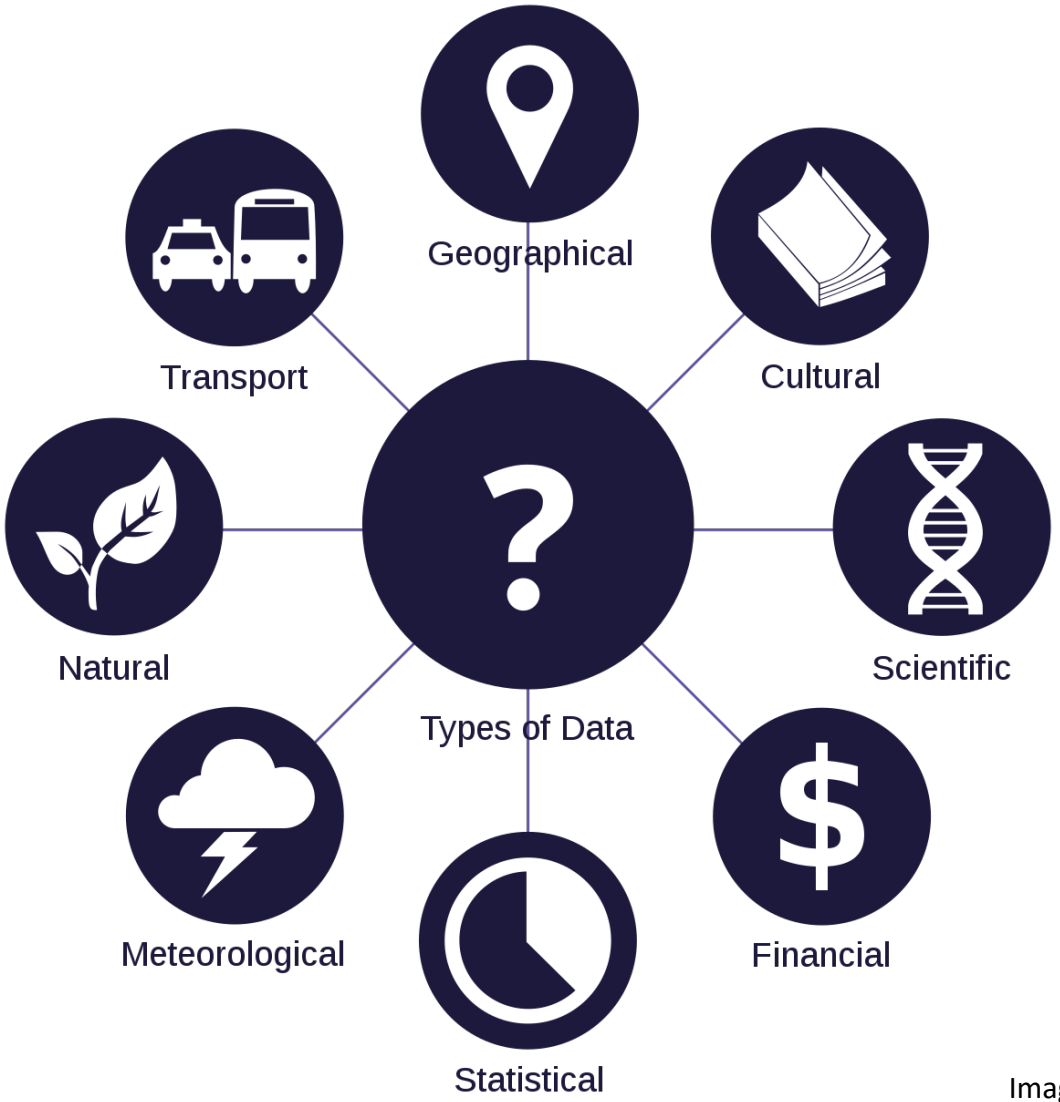


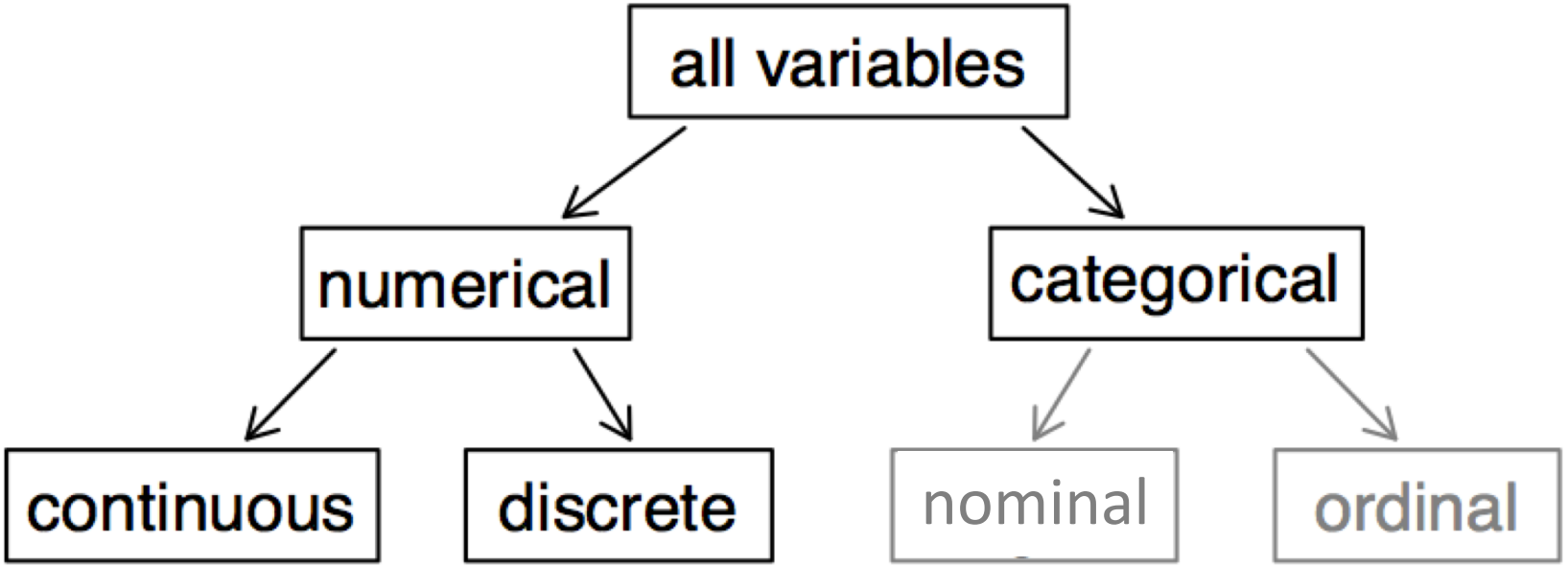
Image: Wikipedia

Data - record of system observations

How will you structure your observations?
How many observations are enough?
What will I do with them?

Image: Wikipedia

Types of Data Measurements



No order or ranking (e.g., male, female, non-binary; control & treatment)

Ranked (e.g., short, tall, very tall)

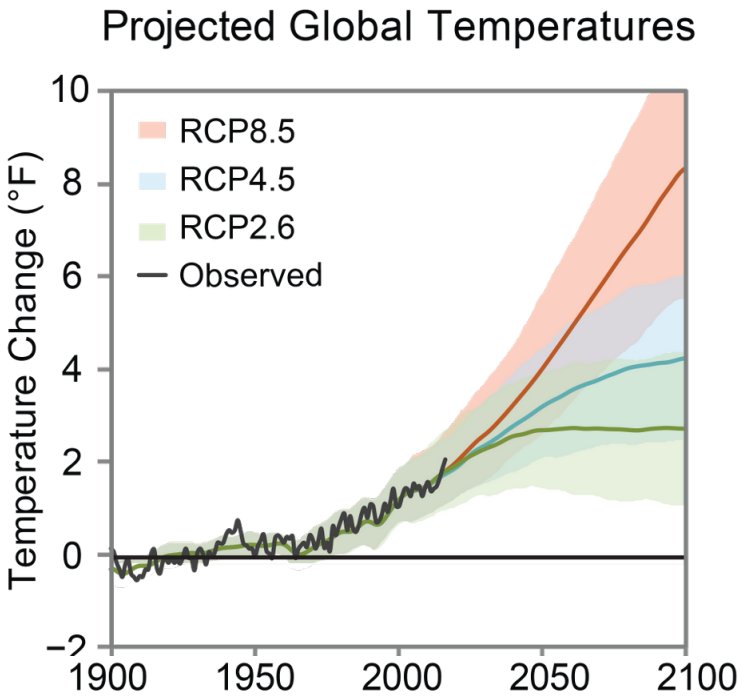
Types of Data

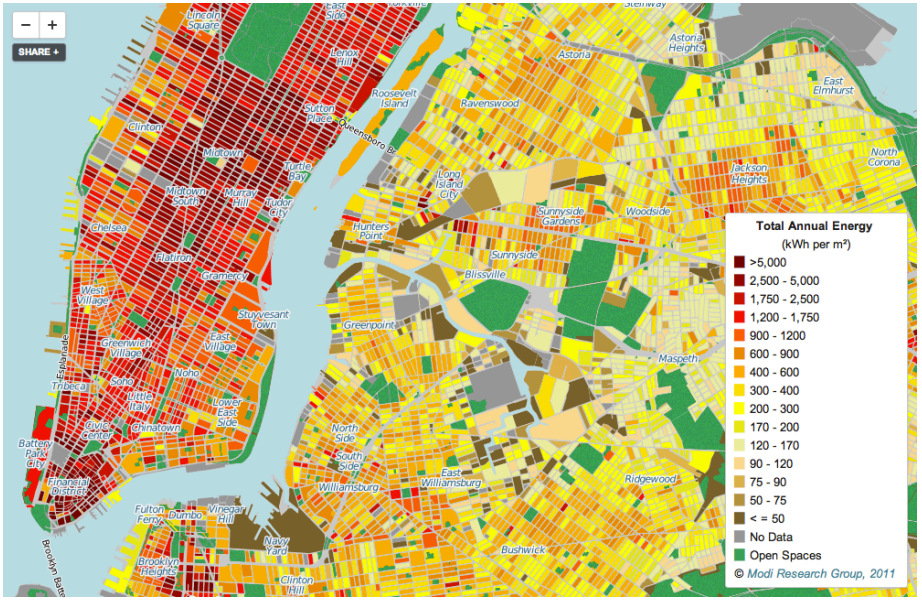
Time series

- Change with time e.g. temperature, unemployment rate

Spatial data:

- Data with locations / geographic coordinates
- Most environmental data have a spatial component





Types of Variables

Dependent:
the variables you are interested in explaining or predicting (what you measure)

Independent / Predictor:
the variables you believe affect your dependent variable (what you manipulate)

Recording Data

- ✓ Notebooks
- ✓ Digital data loggers
- ✓ Online apps – digital recorders
- ✓ Data sheets
- ✓ Surveys – online / paper
- ✓ Voice recorders
- ✓ Methods, metadata

Data Management

Data Management

- ✓Enter data as soon as possible
- ✓Record metadata and methods
- ✓Use informative file names
- ✓Proof read and calculate summary stats
- ✓Exploratory data analyses
- ✓Note your preliminary findings / observations

Data Management

- ✓Use a log book to track your progress in field work / data collection, data entry
- ✓Separate files with informative names
- ✓BACK UP YOUR DATA

Analyzing data

- ✓ Summary statistics (mean, median, range)
- ✓ Variation (standard deviation, error, confidence intervals)
- ✓ Correlation and regression
- ✓ T-tests and ANOVAs
- ✓ Contingency tables and Chi-Square tests
- ✓ Logistic regression
- ✓ Content analyses with code books

Data reality…it is complicated!

Realism: Do your data make sense?
(Range of values, nature)

Confound factors
Randomization
Controls
Consistency
Co-variates
Independent observations

"There's a flaw in your experimental design.
All the mice are scorpios."

Where to go for help

Your mentor and Senior Seminar advisor

The library
- Statistics books
- Methods sections of relevant papers

Statistical Consulting – Columbia Research Data Services
Drop in hours
https://library.columbia.edu/services/research-data-services/schedule.html

Empirical Reasoning Center – Drop in hours
https://erc.barnard.edu/visit-us