Day One – Intro to Data Science

Data comes from:

* Functional Area Support Systems
* Corporate Databases
* Government Websites
* Commercial Providers
* Academic/Research Institutions
* Myriad Electronics Devices (IoT)
* DIY

Quantitative vs. Qualitative

* Quantitative – Measured Objectively
  + Discrete – Limited to certain values ie. # of students in a class
    - Counts and Proportions
    - Nominal Variables – Unordered: Gender, Location
    - Ordinal Variables – Ordered: Grade Level, Income Brackets
    - Continuous Variables – Grouped into Small # of Categories: (interval, bins) Blood Pressure Levels; normal, high-normal, high
  + Continuous – Unfixed number of possible measurements between 2 realistic points
    - Interval Data – Ordered units with same difference: -10, -5, 0, 5, 10
    - Ratio Values – Ordered units with same difference but with a “true” zero
  + Measurement Hierarchy
    - Nominal < Ordinal < Interval
    - Methods for nominal data can be used for interval but not the other way around
  + Contingency Tables – Used to summarize discrete data
* Qualitative – Categorize data, not numerical in nature
  + Obtaining qualitative data
    - Interviews
    - Focus Groups
    - Observation
    - Longitudinal Studies
    - Case Studies
  + Deductive vs. Inductive Reasoning
    - Deductive – Based on predetermined structures to analyze data
      * Generalized Principles > Specific Cases
    - Inductive- Not based on predetermined structures
      * Specific Cases > Generalized Principles
  + Advantages of Qualitative Data: Helps with in-depth analysis, Rich Data
  + Disadvantages of Qualitative Data: Time consuming, hard to generalize, Skill-dependent

Biases

* Selection bias
* Non-response bias
* Social desirability bias