11-7 – 11-19 – The role of data

* Creating vs. validating
  + Data can be used to create a model.
  + Data needs to be evaluated with respect to different data
    - You cannot use data to build something and evaluate it with the same data
  + We can sample from the same dataset to separate them into testing and building
* It doesn’t just go one way
  + A model may be used to take actions in the real-world, thus impacting what data is produced.
* We need data
  + Most (i.e. in volume) of the data “out there” is made of text, sounds, images, movies.
  + We need data we can deal with: spreadsheets and networks
* Importing and exporting data in NetLogo
  + Import: from files to NetLogo
  + Export: from NetLogo to files
  + Spaghetti code – types all the data into the code, refers to states by color
  + Good code – refers to states as numbers/categories, map to colors for visualization
    - Interacts with generic data files (No hardcoded data), more abstraction, more generic
  + Template to import data from a file
    - Ensure there is a file to read data from (**file-exists?**)
    - Open the file (**file-open**)
    - Read the data (**file-read**)
    - Close the file (**file-close**)
  + Read the data
    - Create global variables
    - Read them from the file, one at a time
    - Trigger the load procedure
    - Use them in the code
    - File-read will get the next element of the file, separated by a space.
      * Types must match
* Making sense of data files
  + Meta-data
  + Data dictionaries are not only painful and bulky, but they really mess up data types
  + Configuration file tells us in plain English what the numbers represent
    - INI files
      * Brackets for heading, semicolon for comments, only reads numbers
    - XML files
      * Kind of like folders
    - JSON files (JavaScript)
      * “VarName” : Value,
  + Or we can tell NetLogo to read it and do nothing with it
    - **Let variable file-read-line**
    - Then, **set variable file-read-line** for rest of times
* Exporting Data
  + Why should we export data?
    - Need a purpose for saving
    - Analyze data outside of NetLogo and answer a question
    - Randomness requires multiple runs
  + How to export
    - File -> Export -> Export World
      * Everything in your model
      * Not meant for analysis, to restore a model where you left it at
    - If a program exports everything, the file is called a dump
    - When we export data for analysis, we need to be more selective than dumping everything
  + Data Cleaning
    - Attributes are columns, entries are rows
    - Numerical columns should have consistent units, and just the numbers