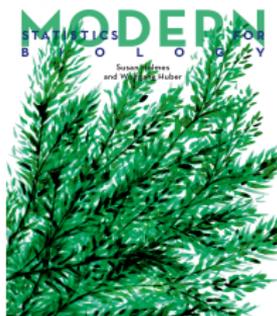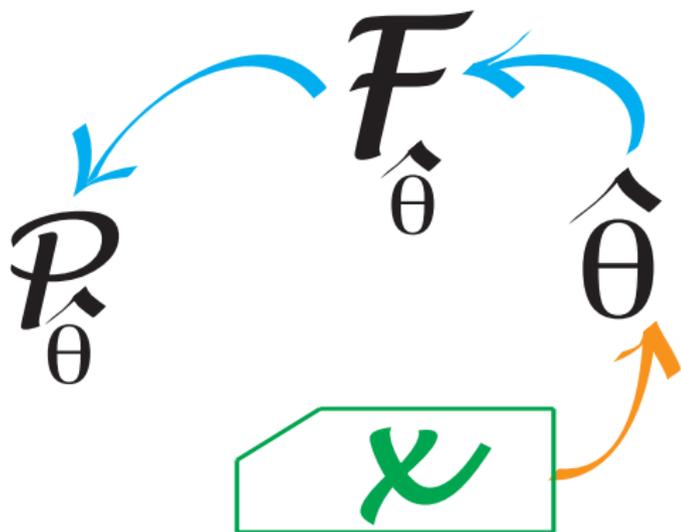# Data Types for Heterogeneous Data

Susan Holmes,@SherlockpHolmes

ASM Workshop, San Francisco, June 2019

Statistics: separate the model from the data

See a complete book:
`http://bios221.stanford.edu/book/`

Heterogeneity of Data

- Status : response/ explanatory.
- Hidden (latent)/measured.
- Types :
  - Continuous
  - Binary, categorical
  - Graphs/ Trees
  - Images
  - Spatial Information
  - Rankings
- Amounts of dependency: independent/time series/spatial.
- Different technologies used (Illumina, MassSpec, NMR, RNA-seq).

Glossary

vector $v[1]$ $v[2]$ $v[3]$ .....$v[n]$

matrix The dimension of A here is 4 by 3.

$$
\begin{array}{ccc}
A[1,1] & A[1,2] & A[1,3] \\
A[2,1] & A[2,2] & A[2,3] \\
A[3,1] & A[3,2] & A[3,3] \\
A[4,1] & A[4,2] & A[4,3]
\end{array}
$$

factor A categorical variable with levels: "O", "AB", "B" etc. See here `http://web.stanford.edu/class/bios221/book/Chap-Generative.html`

list A container for different objects, usually of different type and dimensions.

data.frame A list whose components are of the variables, usually of different types.

status response/explanatory: the "formula" $y \sim x$.

Resources for dealing with different data types and structures in R

- If you've never used R:
  (a lecture 0 link here: `Lecture0-DirectoryFiles.html`).
- About data types: `Lecture4-DataTypes.html`
- Hadley's chapter on data structures:
  `http://adv-r.had.co.nz/Data-structures.html`