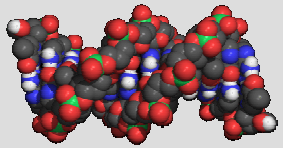


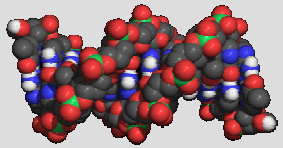
2.0 Definition and Goals

The Data:= “data generated from mathematical models or computations and from human and machine collection”

Goals of this chapter:

- relate data to reality (the world)
- discuss processing of data
- characterize data through data models
- give examples of (complex) data sets

UNIVERSITY OF PADERBORN
Computer Graphics, Visualization and Image Processing
Prof. Dr. Gitta Domik



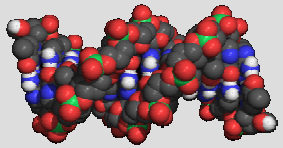
2.1 Data and the world being modeled*

Establish valid and reliable relationship between Data and World

Scientific methods and concepts: Rationale

- Domain specialists try to find a model that matches the “real world”
- Computer Science students have little or no experience in scientific data acquisition or model to real world mapping
- Important for CS students to understand „mindset“ of domain specialists

* Chapter 2.1 from Tutorial by Scott Owen

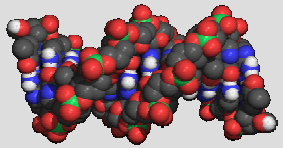


2.1 Data and the world being modeled

Scientific Objectives and Method

- Objectives
  - Attempt to explain the real world (e.g. ball-and-stick)
  - Understanding
  - Prediction
- Method
  - Create a model of the world
  - Acquire data to verify or refine the model

UNIVERSITY OF PADERBORN
Computer Graphics, Visualization and Image Processing
Prof. Dr. Gitta Domik

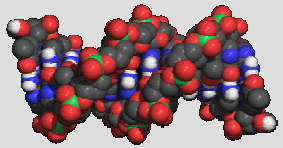


2.1 Data and the world being modeled

Scientific Concepts

- What is a model?
- Relationship between model and empirical data
- Approximations in model
- Features of empirical data
- Mathematical models to represent reality
  - e.g. linear / non-linear / differential equations

UNIVERSITY OF PADERBORN
Computer Graphics, Visualization and Image Processing
Prof. Dr. Gitta Domik

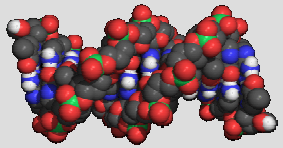


2.1 Data and the world being modeled

Relationship between model and empirical data

- Model guides data acquisition and investigation
- Data may change parameters in model
- Data may cause model to be changed
- Data may be wrong (error in experiment)

UNIVERSITY OF PADERBORN
Computer Graphics, Visualization and Image Processing
Prof. Dr. Gitta Domik

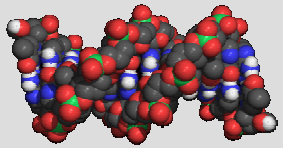


2.1 Data and the world being modeled

Approximation in Scientific Modeling

- Rigorously derive a model
- Make approximations until computationally tractable
- Make more realistic approximations when have:
  - Faster machines
  - Better algorithms

UNIVERSITY OF PADERBORN
Computer Graphics, Visualization and Image Processing
Prof. Dr. Gitta Domik

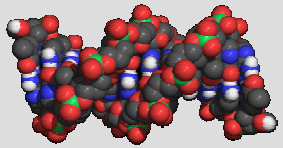


2.1 Data and the world being modeled

Examples of Changing Approximations

- Computer Graphics
  - Ambient light (Phong model)
  - Radiosity
- Quantum Mechanics (Molecular Orbital Calculations)
  - Huckel
  - Semi-Empirical
  - Ab Initio

UNIVERSITY OF PADERBORN
Computer Graphics, Visualization and Image Processing
Prof. Dr. Gitta Domik



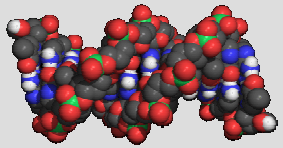
2.2 Processing the Data

Enhance information content of data

Example: Features of empirical data

- „Real world is a fuzzy place“*
- Data usually has noise (random errors)
  - Data may be smoothed
- Data is point sampled from an analog domain
  - Potential aliasing artifacts
- normalizing
  - make data comparable
- data cleansing
  - remove undesirable influences
- filtering: “Goal is to massage the data and not mutilate it”*

* useful comment by Scott Owen



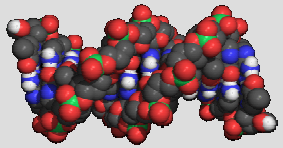
2.3 Data Models

data model = conceptual view of data

Characterize data by e.g.

- geometry
- topology
- value

UNIVERSITY OF PADERBORN
Computer Graphics, Visualization and Image Processing
Prof. Dr. Gitta Domik

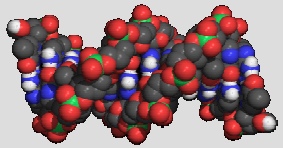


2.3.1 Advantage of Data Models

- discipline independent view on data
- choose expressive visualization technique
- avoid “mental road blocks”

[BRO92], [GAL94]

UNIVERSITY OF PADERBORN
Computer Graphics, Visualization and Image Processing
Prof. Dr. Gitta Domik

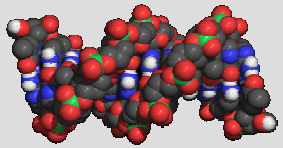


2.3.2 Overview of selected data characteristics

(Non-orthogonal characteristics)

- nominal, ordinal, quantitative
- point, scalar, vector
- “continuous” data
- topology/structure for non-continuous data
- data reliability
- valid range of data
- time descriptors

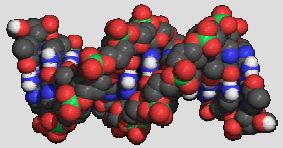
UNIVERSITY OF PADERBORN
Computer Graphics, Visualization and Image Processing
Prof. Dr. Gitta Domik



2.3.3 Nominal, ordinal, quantitative

- Nominal data - members of certain class, e.g. [Georgia, Florida, North Carolina, Delaware], or [Maple, Birch, Oak]
  - effective visual attributes: color (hue!), symbol
- Ordinal data - related by order, e.g. [low, medium, high], or [tiny, small, medium, large]
  - effective visual attributes: brightness, size, (color – hue, if meaningful to observer)
- Quantitative data - carry precise numerical value, e.g. [2.3, 4.56, 0.8, 2.5E-35]
  - effective visual attributes: position, length

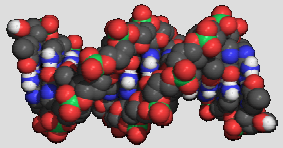
UNIVERSITY OF PADERBORN
Computer Graphics, Visualization and Image Processing
Prof. Dr. Gitta Domik



2.3.3 Nominal, ordinal, quantitative

Priorities of Visual Attribute for Various Data Types (Excerpt) [MAC96]

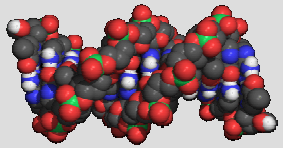
Quantitative	**Ordinal**	**Nominal**
Position	Position	Position
Length	Density	Hue
Angle	Saturation	Density
Slope	Hue	Saturation
Area	Length	Shape
Density	Angle	Length
Saturation	Slope	Angle
Hue	Area	Slope
Shape	Area	Area



2.3.4 Point, Scalar, Vector

Syntactical categories, additionally characterized by dimensions

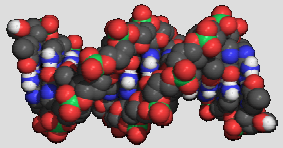
- **Point**
  - each data element is considered a position in n-dimensional space.
  - example: measurements of leaves: [length, width, tree type, age], e.g. [2.3, 1.2, B, 1], [4.3, 2.2, B, 3], [1.5, 1.5, M, 1], [3.0, 2.9, M, 3], ...
  - expressive visualizations: scatter plots, glyphs
- **Scalars**
  - each data element has a numeric expression
  - example: topography of terrain, expressed as 2-d field containing elevations
- **Scalar arrays** - often “discrete samples of continuous functions”
  - usually 1 (linear), 2 (image) , or 3 (volumetric) dimensional data sets; samples in equidistant or non-equidistant steps.
  - expressive visualizations: line graph, shaded surface, volume viewing



2.3.4 Point, Scalar, Vector

- **Vectors**
  - each data element is considered a straight directed line with a certain length (magnitude) in n-dimensional space.
  - example: Direction of particle flow in channel
  - expressive visualizations: arrows, stream lines, particle tracks

UNIVERSITY OF PADERBORN
Computer Graphics, Visualization and Image Processing
Prof. Dr. Gitta Domik

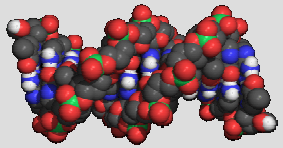


2.3.5 “Continuous” Data

“Continuous” data can be represented by (samples of) function:
yi = fi (X), where X = (x1 , x2 , x3 , ..., xn); i=[1,....,m]
x .... independent variables; e.g space, time, spectral (“dimensions”)
y .... dependent variables (“parameters”)

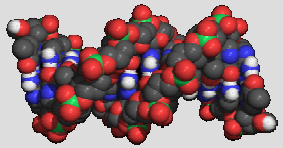
comes in regular and irregular formats

Expressive visualizations of functions: similar to scalar, quantitative, ordinal
Interpolation methods: must be meaningful in problem space
Computation time for visualization techniques faster on regular grids



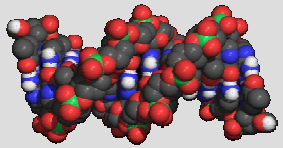
2.3.6 Topology/structure of non-continuous data

- Types of topology/structure, e.g.
  - sequential (text)
  - hierarchical
  - relational
  - single points and connectors
- Examples and corresponding expressive visualizations
  - molecules (e.g. ball-and-stick model)
  - data bases (cone tree; perspective wall)



2.3.7 Other data characteristics in a data model

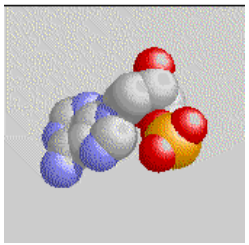
- Data reliability
  - Missing data or unreliable data
    - expressive visualizations: error bars; indicate borders between real/missing data
    - careful with interpolation
- Valid range of data
  - min / max / mean / median
- Time descriptors
  - Various meanings of time: simulation time, simulated/actual time frame, computation time, recording and playback time, user's time frame
  - “time models” to support time conversions necessary to synchronize

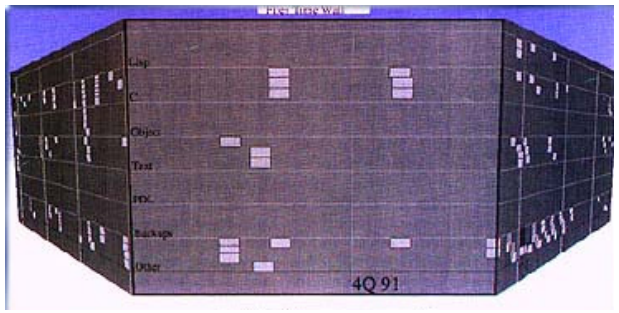


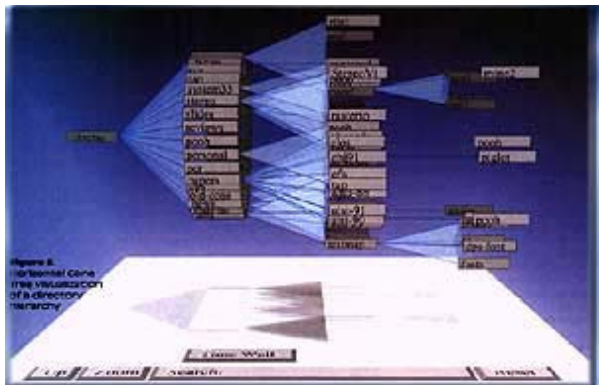
2.4 Examples

Complex data sets and their visual counterparts, e.g

- scientific visualization
- proteins
- software
- web pages







Perspective Wall and Cone Tree: from CACM April 1993, Information Visualizer by Robertson, Stuart and Mackinlay.