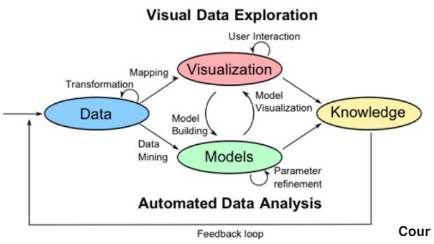
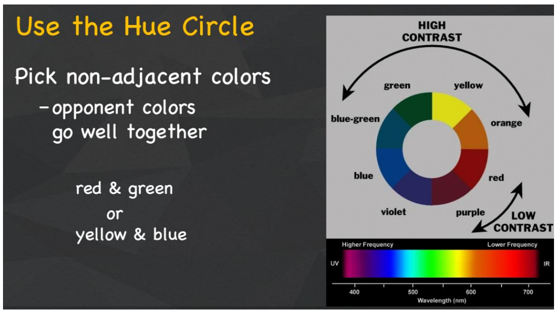
Lecture1 1.Represent information 2.Analyze data 3.Communicate data

1. Helps us think 2.Uses perception to offload cognition 3.Serves as an external aid to augment working memory 4.Boosts our cognitive abilities



- **Data transformation:** converts raw data into a more suitable intermediate representation (e.g., resampling, interpolation, coordinate transformation)
- **Mapping:** converts intermediate data into a number of graphical entities called visualization icons (e.g., isosurfaces, vector arrows, LIC textures, streamlines)
- **Rendering:** Displays the graphical entities on screen
- **Visual perception:** recognition of visual attributes such as color, shape, length, contrast and texture
- **Cognition:** association of low and high level visual information with meaning (e.g., yellow means high temperature, vectors show the direction and speed of a water flow)

Lecture2 **Definition of Visualization:** Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively. **Why Having Human-in-the-loop?** 1.Visualization is suitable when there is a need to augment human capabilities rather than replace people with computational decision-making methods 2.Beyond human patience 3.Scale to large datasets 4.Support interactivity **Benefits of Visualization for Humans** 1.Long-term use for end users (e.g., exploratory analysis of scientific data) 2.Presentation of known results 3.Stepping stone to better understanding of requirements before developing models 4.Helps developers of automatic solution refine/debug, determine parameters 5.Helps end users of automatic solutions verify and build trust



Lecture3 **Visual Perception:** about the nature of the signals coming in; what you see. Perception is the organization, identification and interpretation of sensory information in order to represent and understand the environment. **Cognition** :about how you understand and interpret what you see. In science, cognition is a group of mental processes that includes attention, memory, producing and understanding language, solving problems, and making decisions. **Gestalt Theory** 1.Figure and Ground 2.Closure 3.Similarity 4.Continuity 5.Proximity

Lecture4 **Types of data visualization methods in EDA:** 1.Plotting of raw data 2. Plotting of statistical values 3. Multiple coordinated views (Dashboard)

Principles of Graphical Integrity

- **Minimize the Lie Factor**
  - Tell the truth about the data – above all else show the data
- **Use Consistent Scale**
  - Show data variation, not design variation
  - Make large data sets coherent
- **Present Data in Context**
  - Reveal the data at several levels of detail from broad overview to the fine structure

$$Lie\ Factor = \frac{\text{size of effect shown in graphic}}{\text{size of effect in data}}$$

$$size\ of\ effect = \frac{|\text{second value - first value}|}{\text{first value}}$$

Tufte's Rule

Five Key Principles of Data Visualization

- **Graphical Integrity:** Visual representations of data must tell the truth
- **Data-Ink Ratio:** Data Ink is the ink on a graph that represents data. Good graphical representations maximize data-ink and erase as much non-data-ink as possible
- **Avoid Chart Junk:** Chart Junk is the excessive and unnecessary use of graphical effects in graphs
- **Data Density:** Proportion of total size of the graph that is dedicated to displaying data. Maximize data density and data matrix within reason
- **Small Multiples:** Series of the same small graph repeated in one visual. Small multiples are a great tool to visualize large quantities of data and with a high number of dimensions

Lecture5 **Transformation** 1.Normalization 2.Smoothness 3.Sampling 4.Binning/Discretization 5.Dimensionality Reduction (1) Principal Components Analysis (PCA) (2) Multidimensional Scaling (MDS) (3) Self-Organizing Map (SOM) 6.Clustering

Evaluation Method:

- Inspection or Principled Rationale
  - Apply design heuristics, perceptual principles
  - Expert interviews
- Formal User Study
  - Controlled experiments
  - Interviews
  - Surveys
  - Usability study
- Field Deployment or Case Studies
  - Have people use visualization, observe and interview
  - Document effects on work practices
- Theoretical Analysis and Benchmarking
  - Time and space complexity
  - Performance

Evaluation: when and what?

- Earlier stages:
  - Observe and interview target users (needs assessment)
  - Design data abstraction/operation (data types, transformation, operations)
  - Justify encoding/interaction design (design heuristics, perception research)
  - Informal analysis/qualitative analysis of prototypes (task-based)
  - Algorithm complexity analysis/evaluation
- Mid- and later stages:
  - Qualitative analysis of system (task-based)
  - Algorithm performance analysis
  - Lab or crowdsourced user study
  - Field study of the deployed system

Levels in Text Information

- Based on understanding levels:
  - Lexical Level: basic unit of text.
  - Syntactic Level: structure information.
  - Semantic Level: meaning of text contents.

Research Direction

- Type of text:
  - Single document.
  - Document collection.
  - Temporal documents.
- Features of text:
  - Content.
  - Structure.
  - Multi-level information.

Process: Text Analysis Mining

- Preprocessing:
  - Text wrangling.
- Feature Extraction:
  - Keywords.
  - Word frequency.
  - Topics.
- Measurement:
  - Document similarity
  - Clustering

Process: Visualization

- Visual Design:
  - Chernoff Face.
  - Flow metaphor.
  - …
- Layout:
  - Force-directed.
  - Treemap
  - Scatterplot
  - …

Process: Interaction

- Linking & Brushing:
  - Brushing: focusing.
  - Linking: highlighting.
- Zooming
- Focus + Content
- Filtering

八 Text Analysis Methods : 1.Tokenization 2.Vector-space Model 3.Topic retrieving **Visualizing Document Content** : 1.Word cloud 2.themeriver3.ConceptVector, Senten Tree

Visualizing Document Content

- Keyword-based Visualization
- Temporal Document Visualization
- Feature Distribution Visualization
- Opinion Analysis
- Query Visualization
- Software Visualization

十 **Properties of Network** 1.Complex relations(Directions, weights)2.Network centrality (Degree,Closeness, Betweenness , Eigenvector) **Graph Representation** 1.Node-link diagram (Sugiyama,Force-directed layout, MDS layout) 2.Adjacency matrix 3.Attribute-based representation

Nodes Position Computation*

- From random or initial **configuration**
- Loop:
  - Compute **the repulsion and attraction** force for every pair of nodes.
  - Accumulate the **force** (vector) for every node.
  - Update nodes position **step by step** according to their forces.
- Loop stops when the layout is "good enough"

Gmap Procedures 1.Layout the nodes on the graph 2. Clustering 3. Constructing voronoi diagram for every cluster 4. Coloring

Visual Requirements:

- Key idea: to present the relationship among nodes clearly and efficiently.
  - Order of nodes follow their hierarchies.
  - Less edge crossing.
  - Shorten edge length.
  - Be aware of the aspect ratio.

Hierarchical Data Visualization

- Node-link diagram (structure-clarity).
  - Node-link tree.
  - Hyperbolic tree.
  - Cone tree.
- Space-filling methods (space-efficiency).
  - Treemap.
  - Voronoi treemap.
- Hybrids.

Properties of Streaming Data 1.Unlimited in data volume 2.Uncontrollable with volume or order of the coming data 3.Online analytical processing (OLAP) is needed 4.There exists data error and data loss

Types of Interactions

Dix and Ellis (AVI 1998)	Daniel Keim (TVCG 2002)	Yi et al. (TVCG 2007)
Highlighting and focus	Projection	Select
Accessing extra info – drill down and hyperlinks	Filtering	Explore
Overview and context – zooming and fisheyes	Zooming	Reconfigure
Same representation, changing parameters	Distortion	Encode
Linking representations – temporal fusion	Brushing & linking	Abstract/Elaborate
		Filter
		Connect

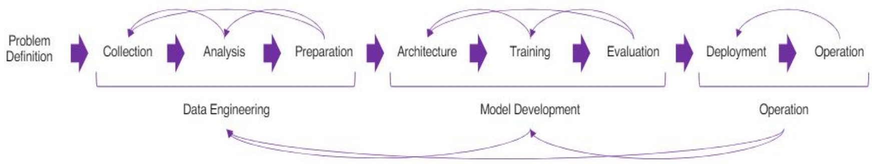
- Classifiers that are commonly recognized as understandable, and hence need little effort to explain them

$$x \longrightarrow \boxed{f_\theta} \longrightarrow y$$

Categories		Related Papers	Remarks
Interpretable Classifiers	Interpretable Architecture	Decision Trees [7], Rule Lists [27, 59], Rule Sets [60]	rule-based
		Linear Models [6]	linear
		kNNs [12, 22]	instance-based
	Learning Sparse Models	Decision Trees [43], Sparse SVMs [11], Sparse CNNs [29]	simplification
		Sparsity by Bayesian [56], Integer Models [55, 58]	direct-sparsity

Not as explainable as they seemed to be!

The Life Cycle of a Classifier:

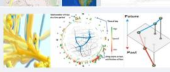
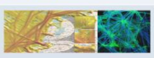
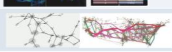


AI4VIS research primarily focuses on 7 problems: 1.**Transformation** converts visualizations from one format to another 2. **Assessment** measures the absolute or relative quality of a visualization in terms of scores or ranking 3.**Comparison** estimates the similarity or other metrics between two visualizations 4.**Querying** finds the target visualization relevant with a user query within visualization collections 5.**Reasoning** challenges AI to interpret visualizations to derive high-level information like insights and summaries 6.**Recommendation** automates the creation of visualizations by suggesting data and/or visual encodings 7.**Mining** discovers insights from visualization databases

- Good visualizations are accumulating rapidly
  - Diverse, creative, and professional
  - Clear patterns: alignment, color, font, etc.
- Understand and reuse existing designs
  - Design ideas: How to present information effectively
  - Design details: How to present information aesthetically
- Challenges
  - Understand: Extract semantic models from examples
  - Reuse: Find appropriate examples to reuse

- Basic Insights:
  - Read a value
  - Identify extrema
  - Characterize distribution
  - Describe correlation

十五 **Visual Analysis of Trajectories** 1.motivation:Wireless Exploration 2.Challenges:Uncertainty,Noise,Scalability 3.Applications:Disaster Forecasting,Traffic Analysis and Management,Analysis of Movements of People,Location-based Service

	Techniques	Visualizations	Features
Spatial	Points Curves Proximately		1. Simple 2. Visual clutter
Temporal	Animation Color Shape 3D		1. Clear overview 2. Instant display vs durative display
Other	Color Height + Illumination		Clear visual patterns
User Interaction	Focus + Overview Brushing		User involved in the loop
Clutter Reduction	Force-directed Model Edge Bundling		1. Obscured semantics 2. Loss of information
Visual Analysis System	Interactive Progressive		Visual comparison and refinement

- Nested Qualitative Methods
  - Experimenter observations
  - Think-aloud protocol
  - Collecting participant opinions
- Inspection evaluation methods
  - Usability Heuristics
  - Collaboration Heuristics
- Information Visualization Heuristics
  - In situ observational studies
  - Participatory observation
  - Laboratory observational studies
  - Contextual interviews