User Tasks & Analysis:

Tasks, Overview and Detail; Brushing and Linking; Sensemaking

CS 4460 - Information Visualization

Spring 2019

Alex Endert

So far…

- **static** multi-variate visual representations
- Now you should understand
  - basic graphical primitives: what they are, why they work (perception)
  - how we place these primitives into visual metaphors to create basic techniques
-

today

- Moving on to **interactive** vis
- Today's focus:
  - low-level taxonomies on user tasks and analysis
  - details-on-demand
  - Also, high level description of how people think about data (e.g., sensemaking)

Why do we need task taxonomies?

- In order to build better visualizations, **we need to understand what people might use them for**
  - What tasks do they want to **accomplish**?
  - What are **models/diagrams** for those tasks?
  - What do we know about the **tasks that people perform**?
  - How do people want to make sense of their data?

Structural decomposition of concepts
Help us understand categories and their relationships
Low and high level taxonomies exist

Recall, two high-level tasks we already talked about

- **search** vs. **browsing**
- but these are pretty generic, let's be more specific

Some Thoughts

- Maybe infovis isn't about answering questions or solving problems… hmmm
- Maybe it's about:
  - asking **better/new questions**?
  - knowing what to search for?
  - getting the "gist" of the data, without knowing individual pieces of data?

Task Taxonomies

- Taxonomies are structural decompositions of concepts to help us understand categories and their relationships

- Number of different ones exist, important to understand what **process** they focus on
  - Creating understanding
  - Creating a “final vis” that presents something
  - …

- Low and High-level taxonomies exist
  - we will talk about both during the 2 days of this topic
  - low-level today

User Tasks

- Wehrend & Lewis created a **low-level, domain independent taxonomy of user tasks** in visualization environments

- 11 **basic** actions
  - identify, locate, distinguish, categorize, cluster, distribution, rank, compare within relations, compare between relations, associate, correlate

Wehrend & Lewis
Vis ‘90

Another Perspective

- Shneiderman proposed **task × data type taxonomy** to understand **what people do with visualization**

- Mantra: "*Overview first, zoom and filter, then details on demand*"
  - Design paradigm for infovis systems
  - more on this later today

Shneiderman
VL '96

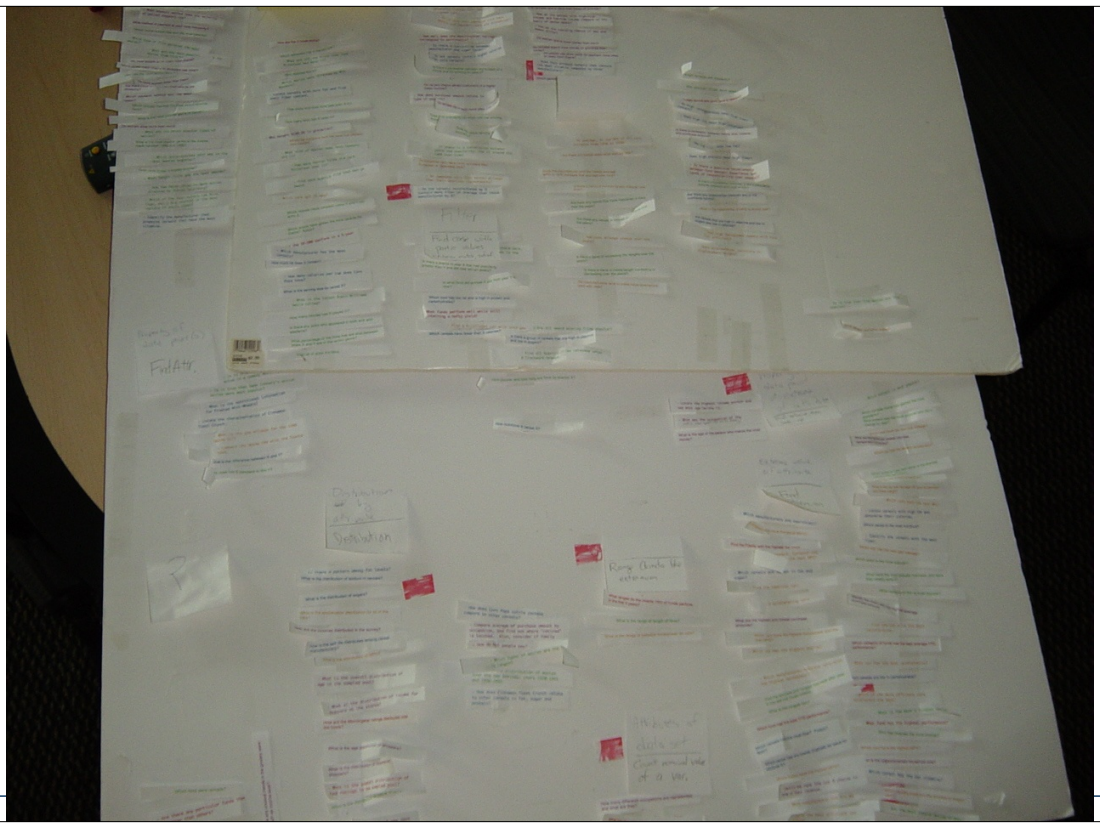
Taxonomy

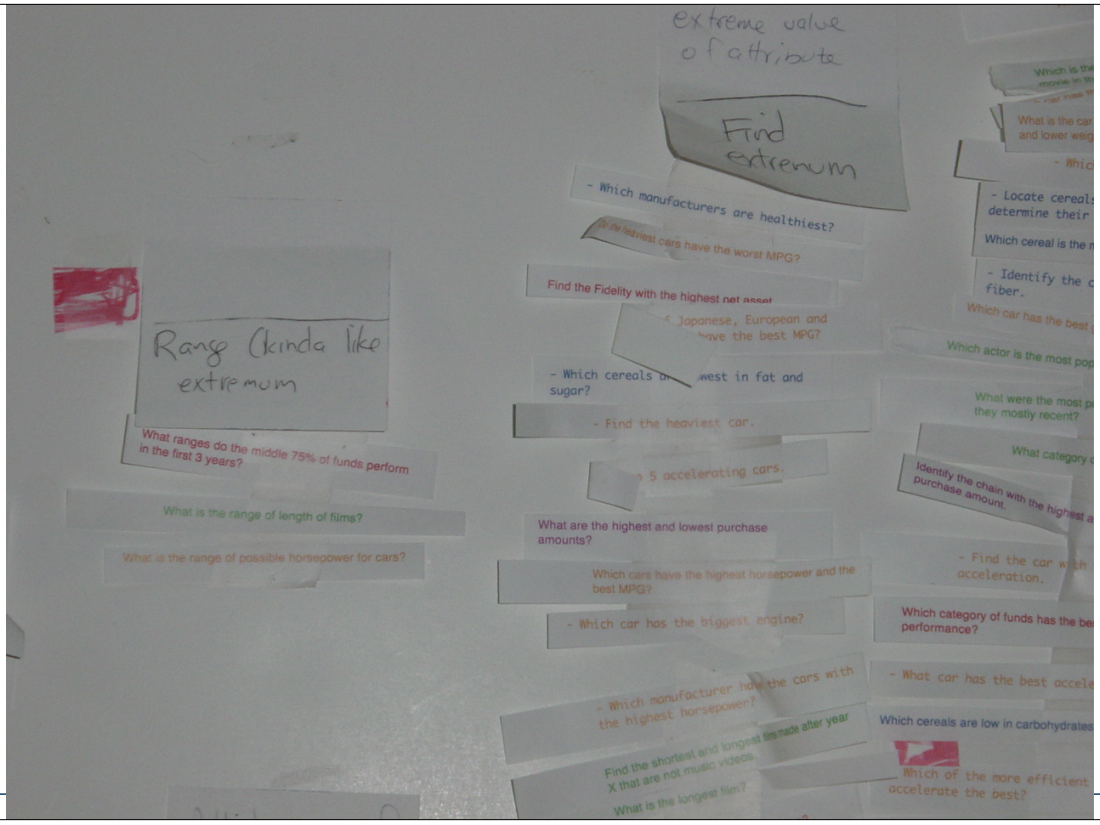
- Data Types
  - 1D
  - 2D
  - 3D
  - Temporal
  - ND
  - Tree
  - Network

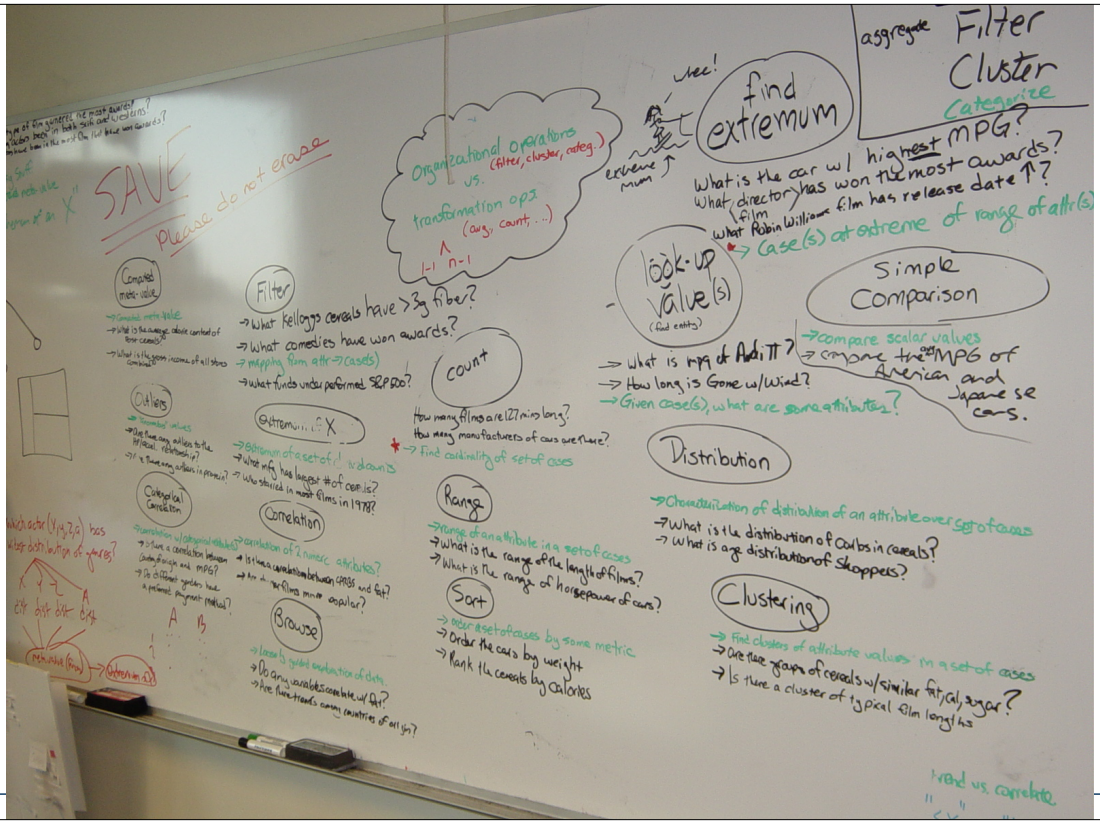
Tasks
- Overview
- Zoom
- Filter
- Details-on-demand
- Relate
- History
- Extract

Primary Task Taxonomy used in this course

- Amar, Eagan, & Stasko – InfoVis ’05
- Let’s first talk about how they came up with this




extreme value of attribute
Find extremum
Range (kinda like extremum
- Which manufacturers are healthiest?
Find the Fidelity with the highest net asset
What ranges do the middle 75% of funds perform in the first 3 years?
What is the range of length of films?
What is the range of possible horsepower for cars?
What are the highest and lowest purchase amounts?
Which cars have the highest horsepower and the best MPG?
- Which car has the biggest engine?
- Find the heaviest car.
5 accelerating cars.
- Which cereals are low in fat and sugar?
Which cereal is the
- Identify the
fiber.
Which car has the best
Which actor is the most
What were the most
they mostly recent?
What category
Identify the chain with the highest
purchase amount.
- Find the car
acceleration.
Which category of funds has the
performance?
- What car has the best
Which cereals are low in carbohydrates
Which of the more efficient
accelerate the best?
- Which manufacturer
the highest horsepower?
Find the shortest and longest
X that are not music videos.
What is the longest film?


SAVE
Please do not erase
Find extremum
aggregate
Filter
Cluster
Categorize
wee!
Organizational operations (filter, cluster, categ.)
vs.
transformation ops (avg, count, ...)
1-1 n-1
extreme num
What is the car w/ highest MPG?
What director has won the most awards?
What Robin Williams film has release date ↑?
Look-up value(s) (find entity)
-> case(s) at extreme of range of attr(s)
Simple Comparison
-> compare scalar values
-> compare the MPG of American and Japanese cars.
-> What is mpg of Audi TT?
-> How long is Gone w/Wind?
-> Given case(s), what are some attributes?
Computed derived value
Filter
-> What Kellogs cereals have >3g fiber?
-> What comedies have won awards?
-> mapping from attr -> case(s)
-> what funds underperformed S&P500?
count
How many films are 127 mins long?
How many manufacturers of cars are there?
-> Find cardinality of set of cases
Outliers
Extremum
Categorical correlation
Correlation
Range
-> range of an attribute in a set of cases
-> What is the range of the length of films?
-> What is the range of horsepower of cars?
Sort
-> order set of cases by some metric
-> Order the cars by weight
-> Rank the cereals by calories
Browse
Distribution
-> Characterization of distribution of an attribute over set of cases
-> What is the distribution of carbs in cereals?
-> What is age distribution of shoppers?
Clustering
-> Find clusters of attribute values in a set of cases
-> Are there groups of cereals w/similar fat/cal/sugar?
-> Is there a cluster of typical film lengths
read vs. correlate

Terminology

- **Data case** – An entity in the data set
  - a.k.a “data item”
- **Attribute** – A value measured for all data cases
  - a.k.a “dimension”, “variable”
- **Aggregation function** – A function that creates a numeric representation for a set of data cases (eg, average, count, sum)

1. Retrieve Value

General Description:
Given a set of specific cases, **find attributes of those cases**.

Examples:
- What is the mileage per gallon of the Audi TT?
- How long is the movie Gone with the Wind?

2. Filter

General Description:
Given some concrete conditions on attribute values,
find data cases satisfying those conditions (and hide others).

Examples:
- What Kellogg's cereals have high fiber?
- What comedies have won awards?
- Which funds underperformed the SP-500?

3. Compute Derived Value

General Description:
Given a set of data cases, **compute an aggregate numeric representation of those data cases**.

Examples:
- What is the gross income of all stores combined?
How many manufacturers of cars are there?
- What is the average calorie content of Post cereals?

4. Find Extremum

General Description:
Find data cases possessing an extreme value of an attribute over its range within the data set.

Examples:
- What is the car with the highest MPG?
- What director/film has won the most awards?
- What Robin Williams film has the most recent release date?

5. Sort

General Description:
Given a set of data cases, **rank** them according to some ordinal metric.

Examples:
- Order the cars by weight.
1. Rank the cereals by calories.
2. Sort students by age.

6. Determine Range

General Description:
Given a set of data cases and an attribute of interest,
find the span of values within the set.

Examples:
- What is the range of film lengths?
- What is the range of car horsepowers?

7. Characterize Distribution

General Description:
Given a set of data cases and a quantitative attribute of interest, **characterize the distribution of that attribute's values over the set**.

Examples:
- What is the distribution of carbohydrates in cereals?
- What is the age distribution of shoppers?

8. Find Anomalies

General Description:
Identify any anomalies within a given set of data cases with respect to a given relationship or expectation, e.g. statistical outliers.

Examples:
- Are there any outliers in protein?
- Are there exceptions to the relationship between horsepower and acceleration?

9. Cluster

General Description:
Given a set of data cases, **find clusters of data items with similar attribute values**.

Examples:
- Are there groups of cereals w/ similar fat/calories/sugar?
- Is there a cluster of typical film lengths?

10. Correlate

General Description:
Given a set of data cases and two attributes, **determine useful relationships between the values of those attributes**.

Examples:
- Is there a correlation between carbohydrates and fat?
- Is there a correlation between country of origin and MPG?
- Do different genders have a preferred payment method?
- Is there a trend of increasing film length over the years?

Discussion/Reflection

- Compound tasks
  - "Sort the cereal manufacturers by average fat content"
    Compute derived value; Sort
  - "Which actors have co-starred with Julia Roberts?"
    Filter; Retrieve value

Discussion/Reflection

- What questions were left out?
  - **Basic math**
    "Which cereal has more sugar, Cheerios or Special K?"
    "Compare the average MPG of American and Japanese cars."
  - **Uncertain criteria**
    "Does cereal (X, Y, Z…) sound tasty?"
    "What are the characteristics of the most valued customers?"
  - **Higher-level tasks** (we will talk about these next time)
    "How do mutual funds get rated?"
    "Are there car aspects that Toyota has concentrated on?"
  - **More qualitative comparison**
    "How does the Toyota RAV4 compare to the Honda CRV?"
    "What other cereals are most similar to Trix?"

Contributions

- Set of **grounded, low-level analysis tasks**
- Potential use of tasks as a language/vocabulary for **comparing** and **evaluating** infovis systems
  - we can now say what a vis is good for; what tasks it supports (or should support)
- Note: this taxonomy did not talk about **how** these tasks should be completed. The interaction design aspect comes later in the semester

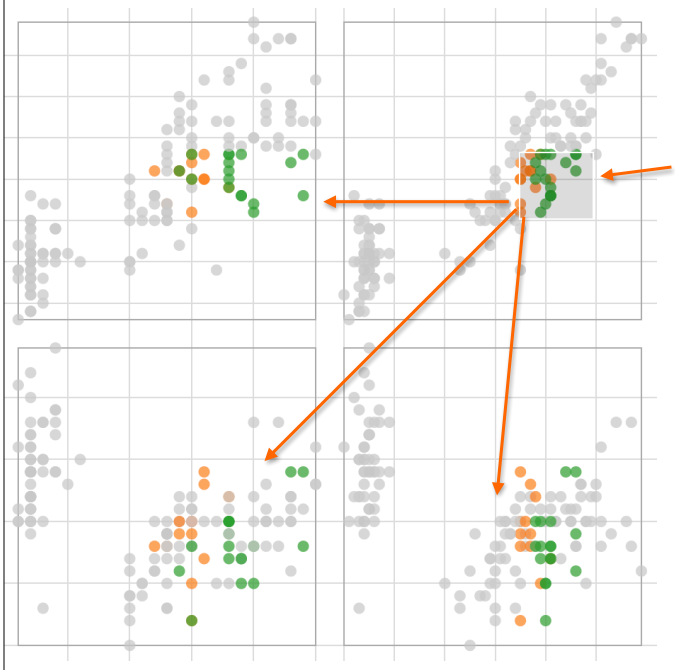
Tasks, in summary

1. Retrieve Value
2. Filter
3. Compute Derived Value
4. Find Extremum
5. Sort
6. Determine Range
7. Characterize Distribution
8. Find Anomalies
9. Cluster
10. Correlate

Brushing and Linking

- Let’s look an example that uses these tasks
- Popular mechanism to coordinate two visualizations.
- Selection in 1 view highlights corresponding data items in second view.

Brushing and linking example



So then, what's Interaction?

- User goals and tasks **carried out through interaction with visualization**
  - The interactive dialog helps people explore
- Interaction **transforms viewers into active participants** in the analysis process
- Interaction moves visualization **from one state to another**

- *Very important component of InfoVis!!*

Interactive Dynamics

- "taxonomy of interactive dynamics that contribute to successful analytic dialogues"
  - part interaction, part task —> There is an obvious connection!

Data and View Specification	**Visualize** data by choosing visual encodings.
	Filter out data to focus on relevant items.
	Sort items to expose patterns.
	Derive values or models from source data.
View Manipulation	**Select** items to highlight, filter, or manipulate them.
	Navigate to examine high-level patterns and low-level detail.
	Coordinate views for linked, multidimensional exploration.
	Organize multiple windows and workspaces.
Process and Provenance	**Record** analysis histories for revisitation, review, and sharing.
	Annotate patterns to document findings.
	Share views and annotations to enable collaboration.
	Guide users through analysis tasks or stories.

Heer & Shneiderman
CACM '12

much more on user interaction later

- For now, let's put taxonomies and user interaction together and expand on Shneiderman's Mantra

much more on user interaction later

- For now, let's put taxonomies and user interaction together and expand on the Shneiderman's Mantra

Information Visualization: Mantra

- Overview, zoom & filter, details-on-demand
- Overview, zoom & filter, details-on-demand
- Overview, zoom & filter, details-on-demand
- Overview, zoom & filter, details-on-demand
- Overview, zoom & filter, details-on-demand
- Overview, zoom & filter, details-on-demand
- Overview, zoom & filter, details-on-demand
- Overview, zoom & filter, details-on-demand
- Overview, zoom & filter, details-on-demand
- Overview, zoom & filter, details-on-demand

Fundamental Problem

- Scale - Many data sets are too large/complex to visualize on one screen
  - May simply be **too many cases** (too large)
  - May be **too many variables** (too complex)
  - May only be able to highlight particular cases or particular variables, but viewer's focus may change from time to time

One Solution :^)

You can just buy more pixels

there are actually other advantages other than just more space, we'll get to that

Perhaps another driving factor

Devices with even smaller screens are becoming more popular!

This means vis designers have less pixels to work with :(

Overview

- Providing an overview of the data set can be extremely valuable
  - Helps present overall patterns
  - Assists user with navigation and search
  - Orients activities
- Generally visualizations should start with an overview
  - Again, Shneiderman mantra

Details

- Viewers also will want to examine details, individual cases and variables
- How to allow user to **find and focus on details of interest**?
- Generally provide **details on demand**

Providing Both

- **Overview + detail** displays can be combined via either time or space
  - Time - Alternate between overview and details sequentially in same place // **interaction**
  - Space - **Use different portions of screen** to show overview and details
- Each has advantages and problems
- Hybrid approaches also exist

An Example

Overview and detail (from Civilization II game)

Survey of Techniques

- Application concern: viewing and editing large images
- Expanding the notion of the one dimensional scroll bar: zooming, diagonal panning, multiple detailed views
- List of visualization/interaction solutions

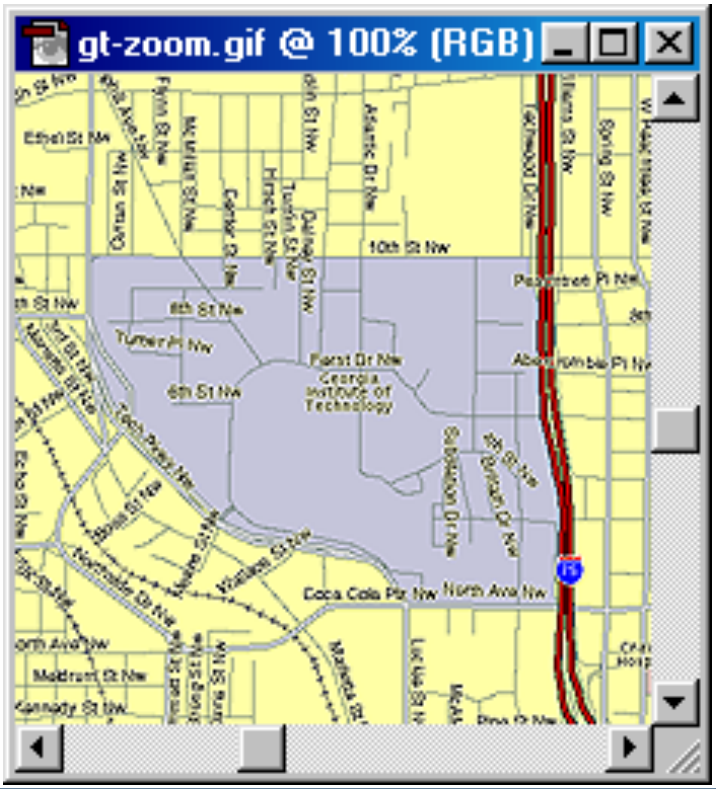
- **Let's take a look at some variations and examples!**

Plaisant et al
IEEE Software '95

1. Detail-only

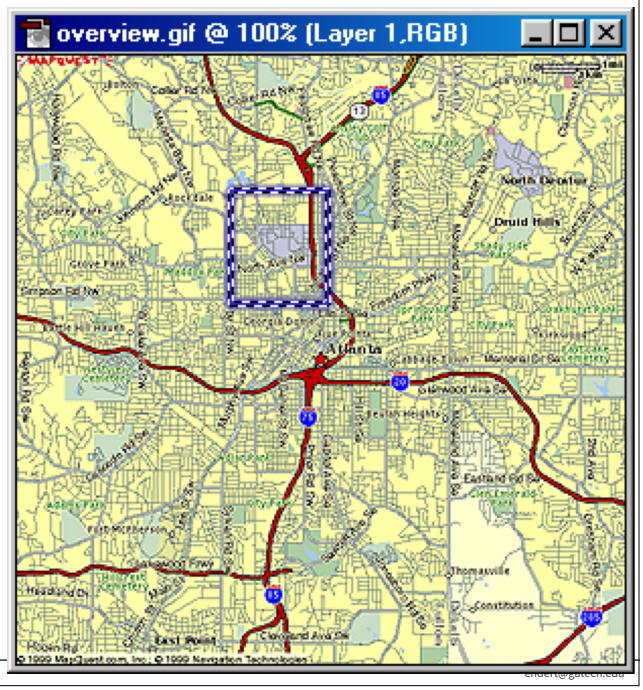
- Single window with horizontal and vertical panning

- Works only when zoom factor is relatively small



2. Single window with zoom and replace

- Global view with **selectable zoom area** which then becomes entire view
- Variations can let users pan and adjust zoomed area and adjust levels of magnification
- Context switch can be disorienting
- Example: CAD/CAM

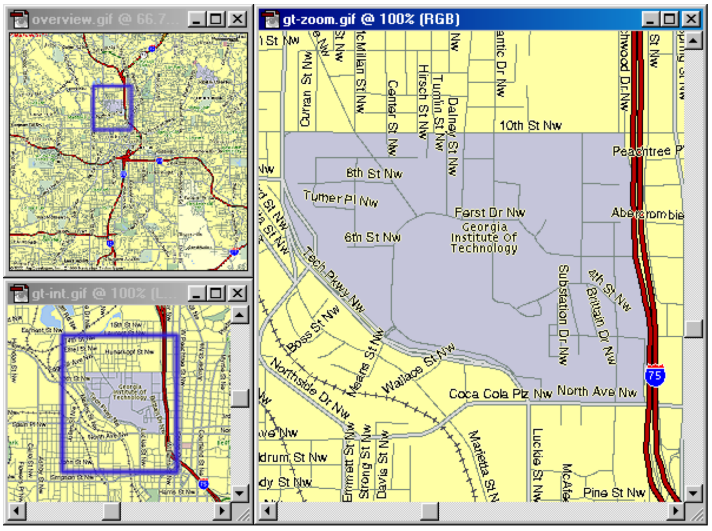


3. Single coordinated pair

- Combined display of the overview and local magnified view (separate views)
- Some implementations reserve large space for overview; others for detail
- Issue: How big are different views and where do they go (overlap?)?

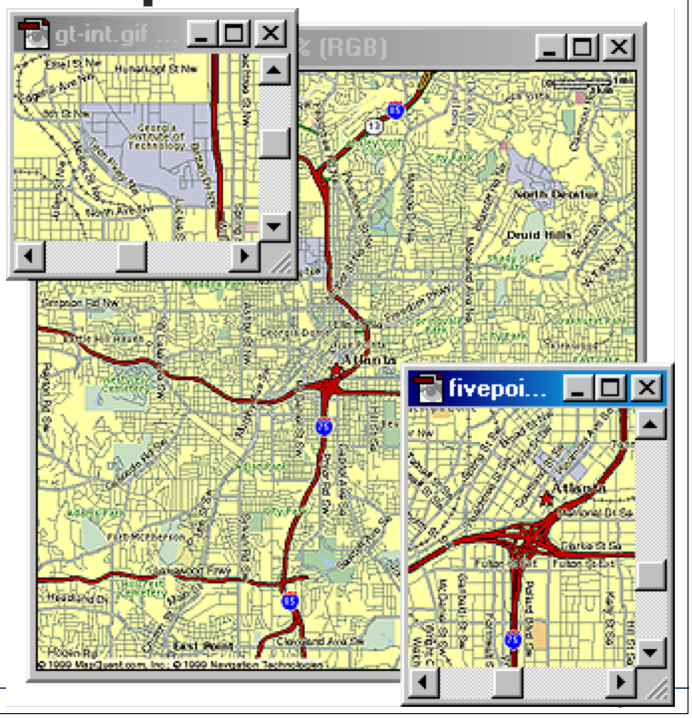
4. Tiled multilevel browser

- Combined global, intermediate, and detail views
- Views do not overlap
- Good implementations closely relate the views, allowing panning in one view to affect others



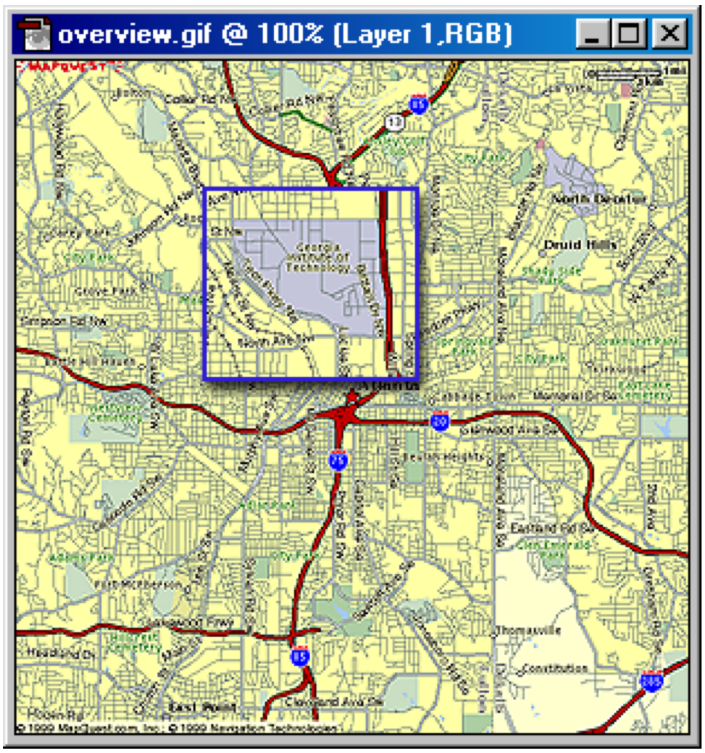
5. Free zoom and multiple overlap

- Overview presented first; user selects area to zoom and area in which to create detailed view
- Flexible layout, but users must perform manual window management



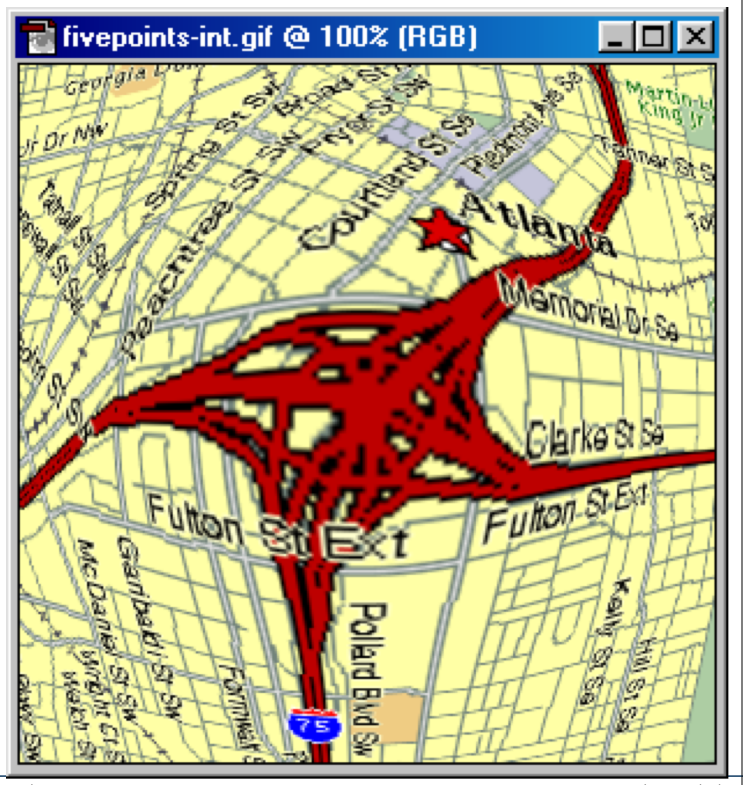
6. Bifocal magnified

- "Magnifying glass" zoomed image floats over overview image
- Neighboring objects are obscured by the zoomed window



7. Fish-eye view

- Magnified image is distorted so that focus is at high magnification, periphery at low
- All in one view
- Distortion can be disorienting
- More details coming...



Focus + Context Views

- Same idea as overview and detail, with one key difference:
  - Typically, the **overview and the detail are combined into a single display**
  - Mimics our natural vision systems more closely

Prototypical Example

- When people think about focus+context views, they typically think of the **Fisheye** View (distortion)
- Introduced by George Furnas in 1981 report, more famous article is 1986 SIGCHI paper

Definition

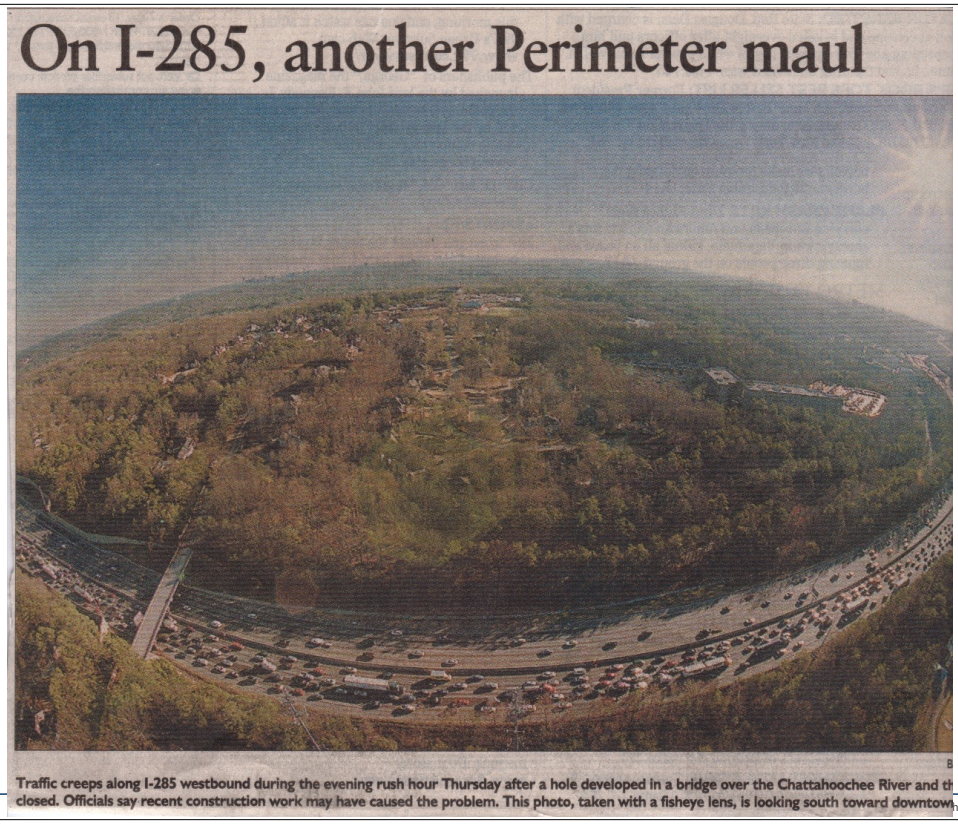
- Fisheye View -

"Provide[s] detailed views (focus) and overviews (context) without obscuring anything…The focus area (or areas) is magnified to show detail, while preserving the context, all in a single display."

-(Shneiderman, DTUI, 1998)

Real fisheye camera lens

Atlanta Journal



On I-285, another Perimeter maul

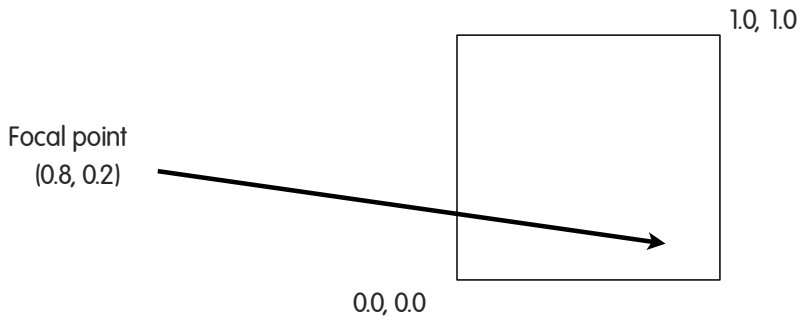
Traffic creeps along I-285 westbound during the evening rush hour Thursday after a hole developed in a bridge over the Chattahoochee River and th closed. Officials say recent construction work may have caused the problem. This photo, taken with a fisheye lens, is looking south toward downtown.

Fisheye Terminology

- Focal point
- Level of detail
- Distance from focus
- Degree of interest function

Focal Point

- Assume that viewers focus on some item, some coordinate, some position,…
- This position (or point) is called the focal point

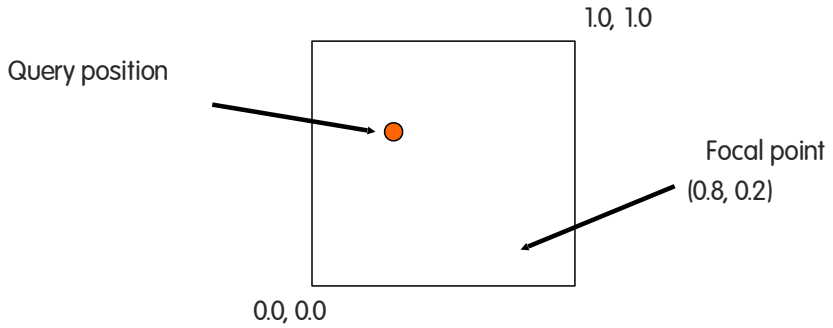


Level of Detail

- Some intrinsic value or quantity on each data element
- How important is it to you in a general sense?
- Simplest example is that all data items have same level of detail

Distance from Focus

- Calculation of how far each data item is from the focal point
  - What level of detail should the query position have, given the focal point location?

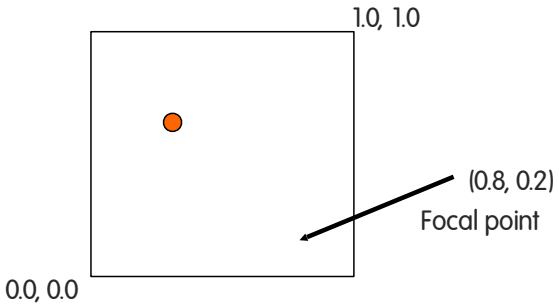


Degree of Interest Function

- Function that determines how items in display are rendered (what level of detail)
  - There are many variations of DOI functions. Each has tradeoffs

E.g., Degree of Interest (a,f) = Level of Detail (a) - Distance from Focus f

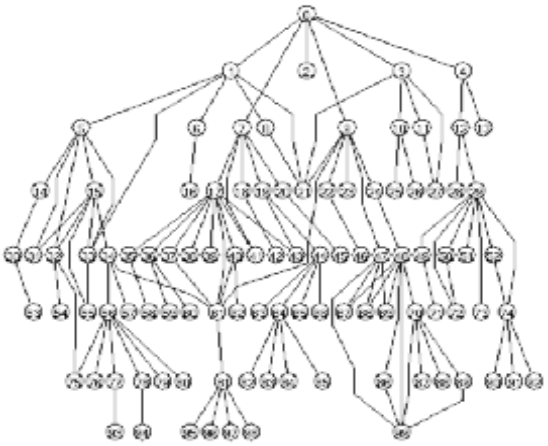
a = point in vis
f = focus point



Degree of Interest Function Types

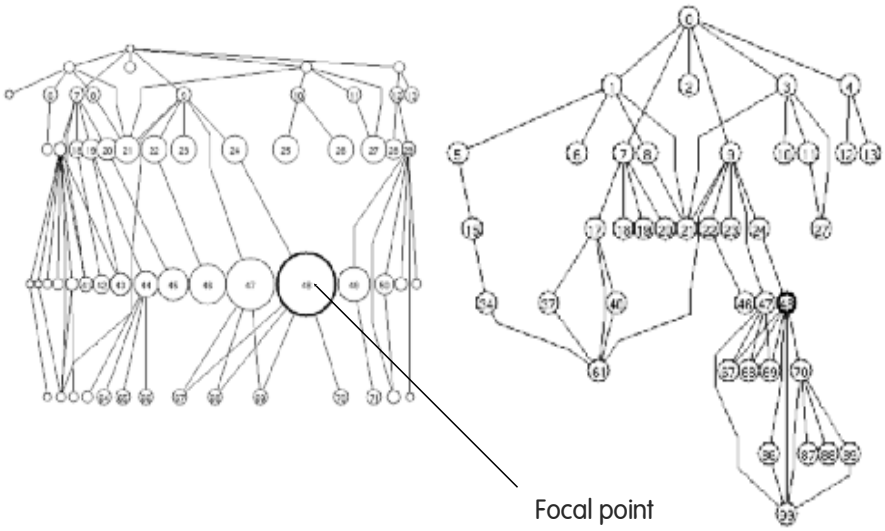
- Can take on various forms
  - **Continuous** - Smooth interpolation away from focus
  - **Filtering** - Past a certain point, objects disappear
  - **Step** - Levels or regions dictating rendering
    0<x<.3 all same, .3<x<.6 all same
  - **Semantic changes** - Objects change rendering at different levels

Example



Original

Example



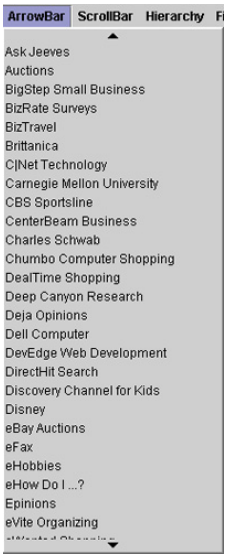
Focal point

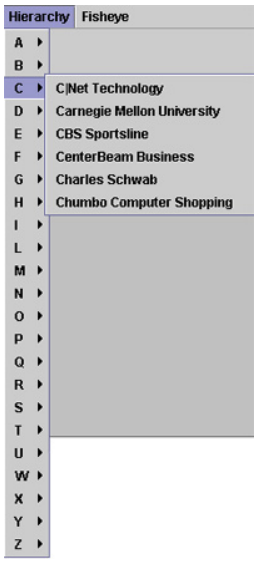
Fisheye Application

- Using Fisheye view in menus
  - Menus have too many items
  - Especially a menu of data items (fonts)
  - Scrolling arrows & bars
  - Hierarchical groups

Bederson
UIST '00

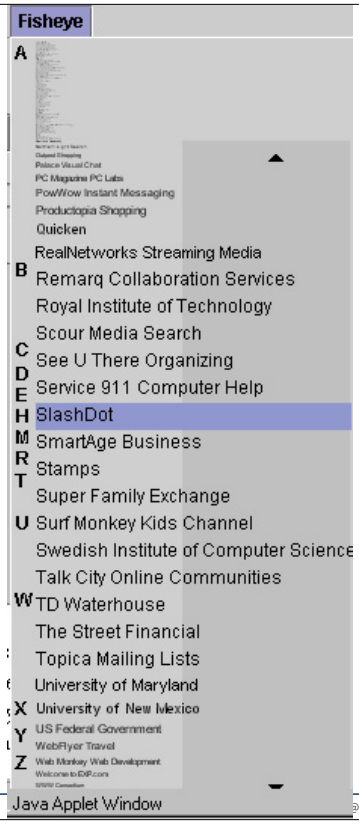
Existing Options





Fisheye Menus

- Dynamically change size of menu item & provide focus area around the pointer
- Items near cursor displayed at full size
- Items further away on either side are smaller
- Uses a distortion function so items will always fill menu



Disadvantages

- Distortion can be annoying
- Can be very difficult to implement
  - at times computationally expensive / slow
- Any change in focal point potentially requires recalculation of DOI for all objects and hence re-rendering of all objects -> Expensive!

now, let's move on from low-level

- higher-level understanding of user tasks and analysis
  - sensemaking
  - knowledge discovery
  - insight
- techniques to help people think (about data)

concepts we will cover

- Models describing the process of data analysis
  - mental models
  - sensemaking loop; *Pirolli and Card*
  - data-frame model; *Klein et al.*
  - *insight*

Mental models

- People have mental models of phenomena, situations, events, etc.
- We construct and edit these based on our experiences during our life (including our experiences with data visualizations)
  - Watching how people act in certain situations might build our mental model of social patterns
  - Watching water flow in a stream might teach us about physics
- Mental models are constantly created, adjusted
  - This process, broadly, is called sensemaking

Sensemaking Loop

some context

- Peter Perolli
  - Research Fellow, PARC
  - was at UC Berkley prior

- Stuart Card
  - big name in HCI
  - “pioneer” of considering human factors for computers
  - written lots of papers, books, etc. (h-index = 77)

The Gist of this Work

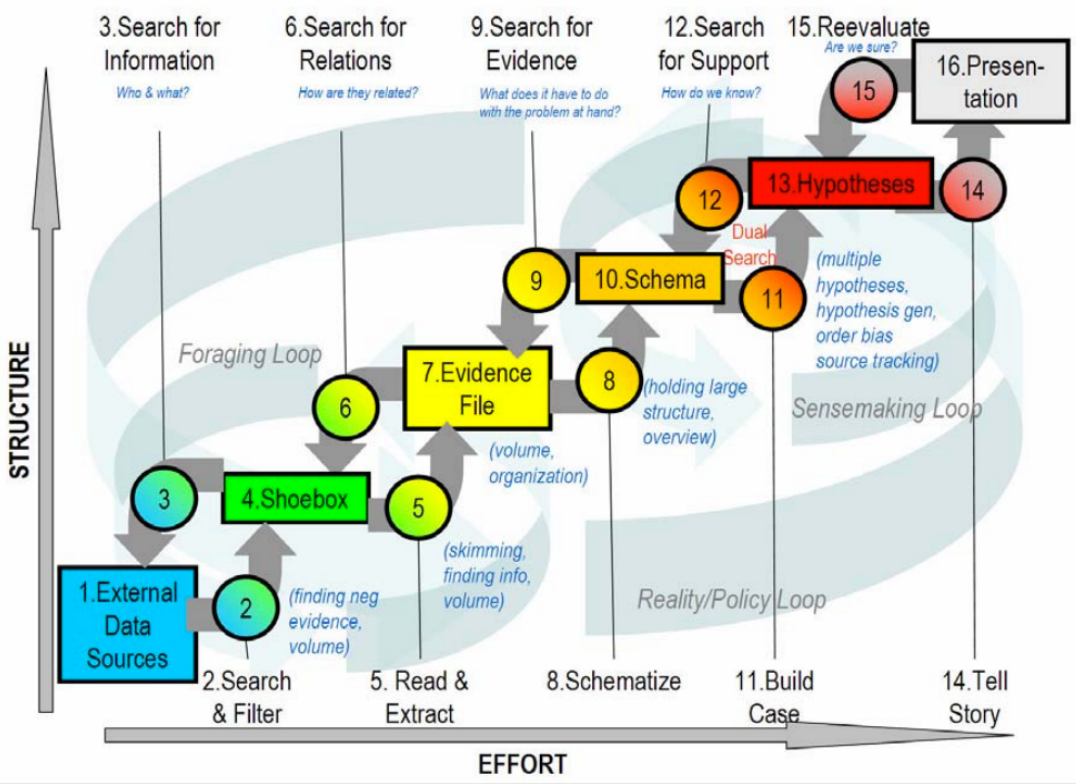
- provide empirical descriptions of the cognitive processes involved in **intelligence analysis**
- worked closely with a (small) number of analysts
  - observed how they worked
  - performed **cognitive task analysis**
- produced a commonly referenced and used figure (notional model) for sensemaking

Intelligence analysis

- specific type of analysis intended to gain “intelligence” (or knowledge, evidence, support)
- typically associated with the “intelligence community” (but doesn’t have to be)
- think: the TV show 24, but less flashy…

they observed schemas

- ordered lists
- organized piles
- linked maps
- hierarchical organization charts
- overall, these are **spatial structures** that hold tacit information
  - authors claim that the mental counterpart to these are schemas

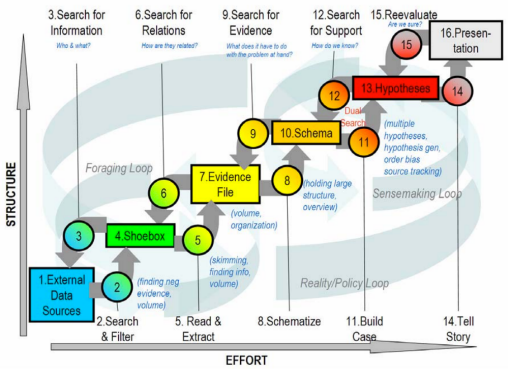


Bottom-up

- search and filter
- read and extract
- schematize
- build case
- tell story

Top-down

- re-evaluate
- search for support
- search for evidence
- search for relations
- search for information



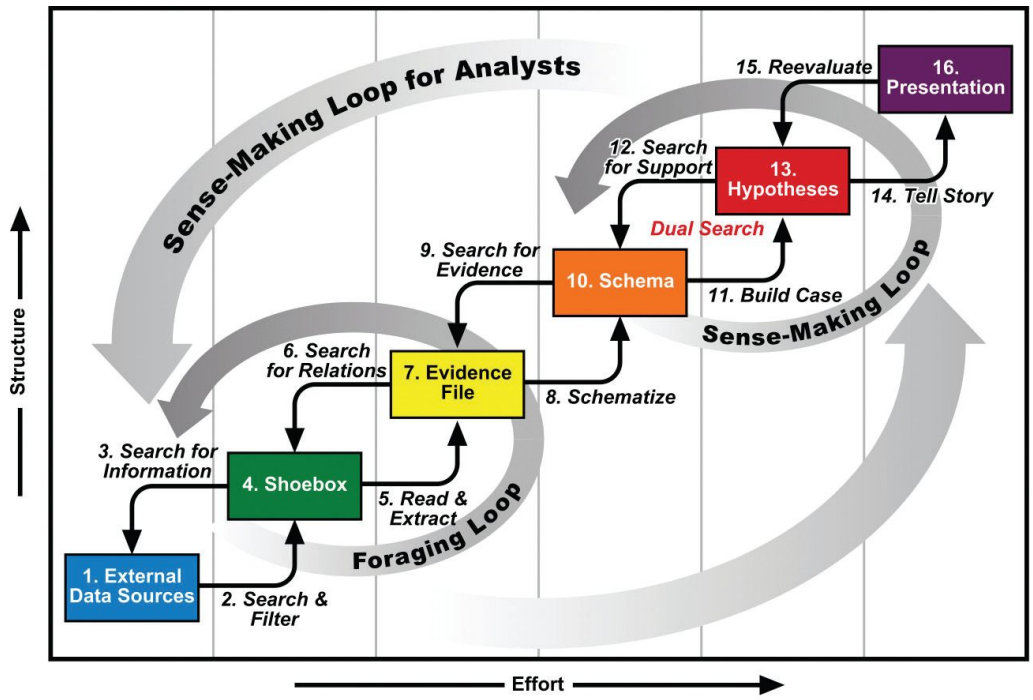
Information Foraging Tasks

- **Exploring**: like recall in IR, increasing the span of new information
- **Enriching**: narrowing to a smaller set of items
- **Exploiting**: generate inferences, notice pattern (generally, the more in-depth analysis of a smaller sets of data)

Cost structures for analysis

- authors describe cost structures associated with key points in analysis
- E.g., cost of shifting attentional control, cost of follow-up searches, etc.
  - some of these provide opportunities for tools to help (we'll talk about some of these tools this semester)

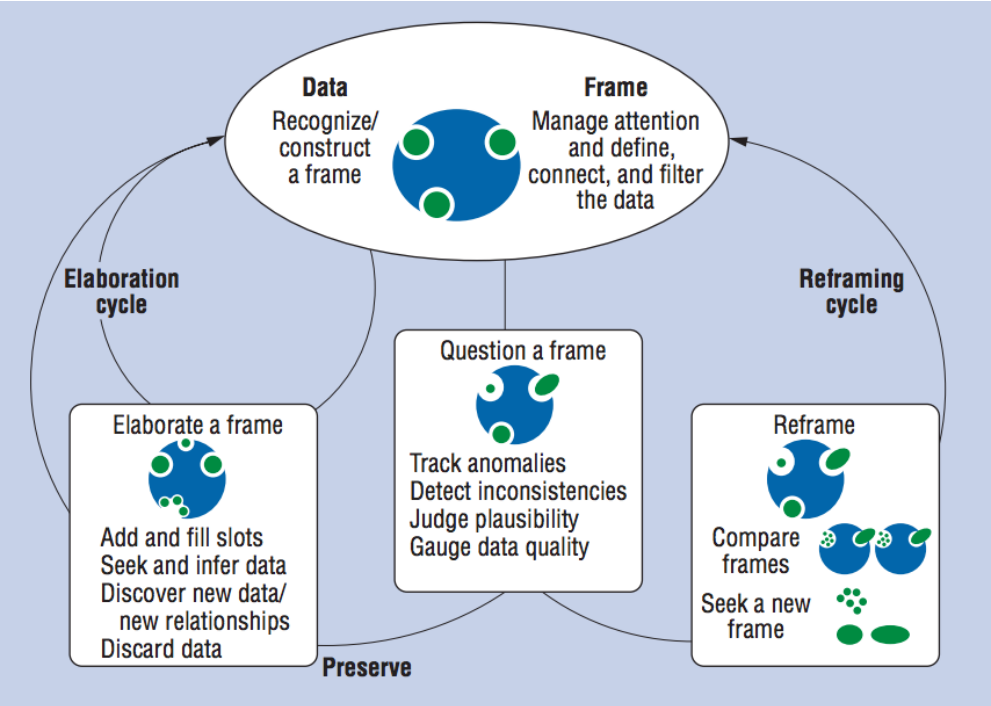
Sensemaking Loop



The Data-Frame Model

Data-Frame Model

Klein et al., 2006



Seeking a frame

- Establishing new “anchors” or relationships that allow you to explain a collection of data
  - a Frame is a way to describe sets of data points, and why they exist
  - E.g., be able to explain why your program crashed, given the error logs, may fit in your frame of how your system is supposed to run
-

Elaborate the frame

- Given an existing frame
  - seek data to fill it out; confirm it
  - extending the frame to “fill slots” with data that you expected
- Then, incorporate this back into your collection of frames for the process of sensemaking

Questioning the frame

- you just found data that violates an expectation that you had about a frame
- inconsistency with what you expect to find
-

Preserving the Frame

- explain the anomaly in the data that forced you to question to frame
- ability to understand error in data, therefore the frame holds

Comparing Frames

- The act of sharpening and refining the distinction between frames to clarify the boundaries/definitions of both
  - E.g., Two people may have a different “frame” for what a “nice day” is. Clarifying the distinction can help define both valid frames.
-

about the process

- While there are loops, many of these tasks happen quickly and in parallel
- The process can happen even without any interaction with a system
  - E.g., As you sit back and think about your data currently shown, you might be going through many iterations of this model, all without interaction or new data.

Ultimately, it's about gaining insight

Defining insight

- Many views on how to define it exist
  - Insight is the "critical breakthrough" that **suddenly sparks a new way of thinking about a problem**, or answering it (Saraiya et al.)
  - Observing old information in a new way so that you understand

- Defined as a "**product**"/result (North, 2006). Insight is
  - **Complex**: involves lots of data points, not simply individual values
  - **Deep**: builds up over time; generates further questions, thus further insight
  - **Qualitative**: multiple levels of resolution, not exact
  - **Unexpected**: serendipitous; unpredictable
  - **Relevant**: dependent on the domain (an insight in one domain may not be in another)

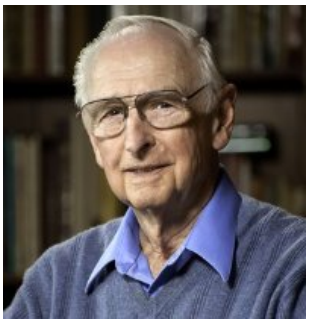
Defining insight

- Others have defined it as a **process**
  - Think about sensemaking
  - The process of forming frames may be insightful/insight
  - The frames may be your insight
  - …

- *insight and knowledge discovery is complex. We know how some of the components work, and how they fit together, but there's lots left to understand (and is likely beyond the scope of this course)*

let's talk about insight as a process

- Richards Heuer
- Former Central Intelligence Agency (CIA) employee
- extensive work in Structured Analytic Techniques
  - one of which is ACH, Analysis of Competing Hypotheses
-



process implies considering prior knowledge

- Focus on the "role of the observer in determining **what is observed** and **how it is interpreted**"
  - "perceiving the world through a lens", may distort/shape opinion?
- those who "know the most about a subject have the most to **unlearn**"
  - prior experiences can help, but can also hurt
  - analysis rarely starts with an "empty mind"
- mental models contain a lot of information
  - users need tools to help them organize it, map it into the data being analyzed

Thinking about Thinking

- How does one map what you know into/onto what is being analyzed?
- What are some **advantages of prior knowledge**?
- What are some **disadvantages of prior knowledge**?

some examples…

some examples…

Georgia Tech

some examples…

Impressions resist change.

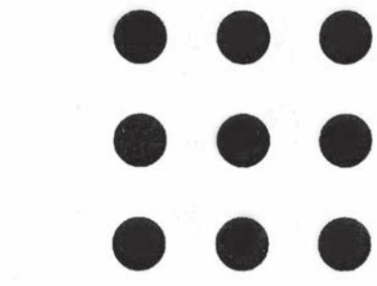
Georgia Tech

some examples…

It is difficult to look at the same information from different perspectives.

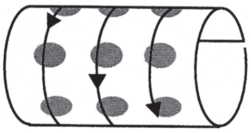
some examples…

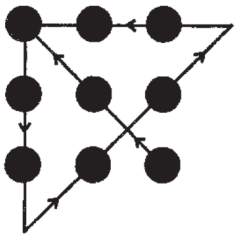
- Try it:
  - connect all dots
  - can only use 3 lines
  - cannot lift pencil

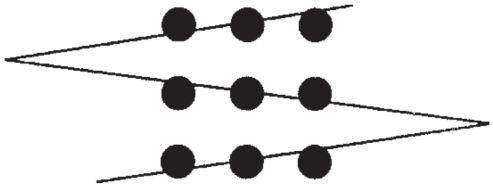


some examples…

- Try it:
  - connect all dots
  - can only use 3 lines
  - cannot lift pencil







Review from today

- 2 models for describing sensemaking
  - Sensemaking Loop
  - Data-frame model
- Defining insight
  - Product-oriented views
  - Process-oriented views
- Example of using visualization (i.e., spatial layout) as a method for externalizing sensemaking process
-