Chapter 2

What is Vis?
Why do it?

The Big Picture

- **Visualization provide visual representations of datasets designed to help people carry out tasks more effectively**
  - **Vis is suitable when there is a need to augment human capabilities rather than replace people with automatic methods.**
- **About visualization**
  - **Design space is huge**
    - **Includes the considerations of both how to create and how to interact with visual representations**

The Big Picture

- Vis design is full of trade-offs, and most possibilities are ineffective for a particular task
  - So validating the effectiveness of a design is necessary and difficult.
- Must take into account three very different kinds of resource limitations
  - Those of computers
  - Those of humans
  - Those of displays

Rationale behind aspects of vis definition

- Why have a human in the decision-making loop?
- Why have a computer in the loop?
- Why use an external representation?
- Why depend on vision?
- Why show the data in detail?
- Why use interactivity?
- Why is the vis idiom design space huge?
- Why focus on tasks?
- Why are most designs ineffective?
- Why care about effectiveness?
- Why is validation difficult?
- Why are there resource limitations?
- Why analyze vis?

Why have a human in the decision-making loop?

- **Vis allows people to analyze data when they don’t know exactly what questions they need to ask in advance**
  - **When people have well-defined questions to ask about data, they can use purely automatic methods from statistics and machine learning**
- **Many analysis problems are ill specified: people don't know how to approach the problem**
  - **There are many questions to ask, and people don’t know which are the right ones in advance**
  - **The best path forward is an analysis process with a human in the loop**
    - **People can exploit the powerful pattern detection properties of human visual system in the design.**

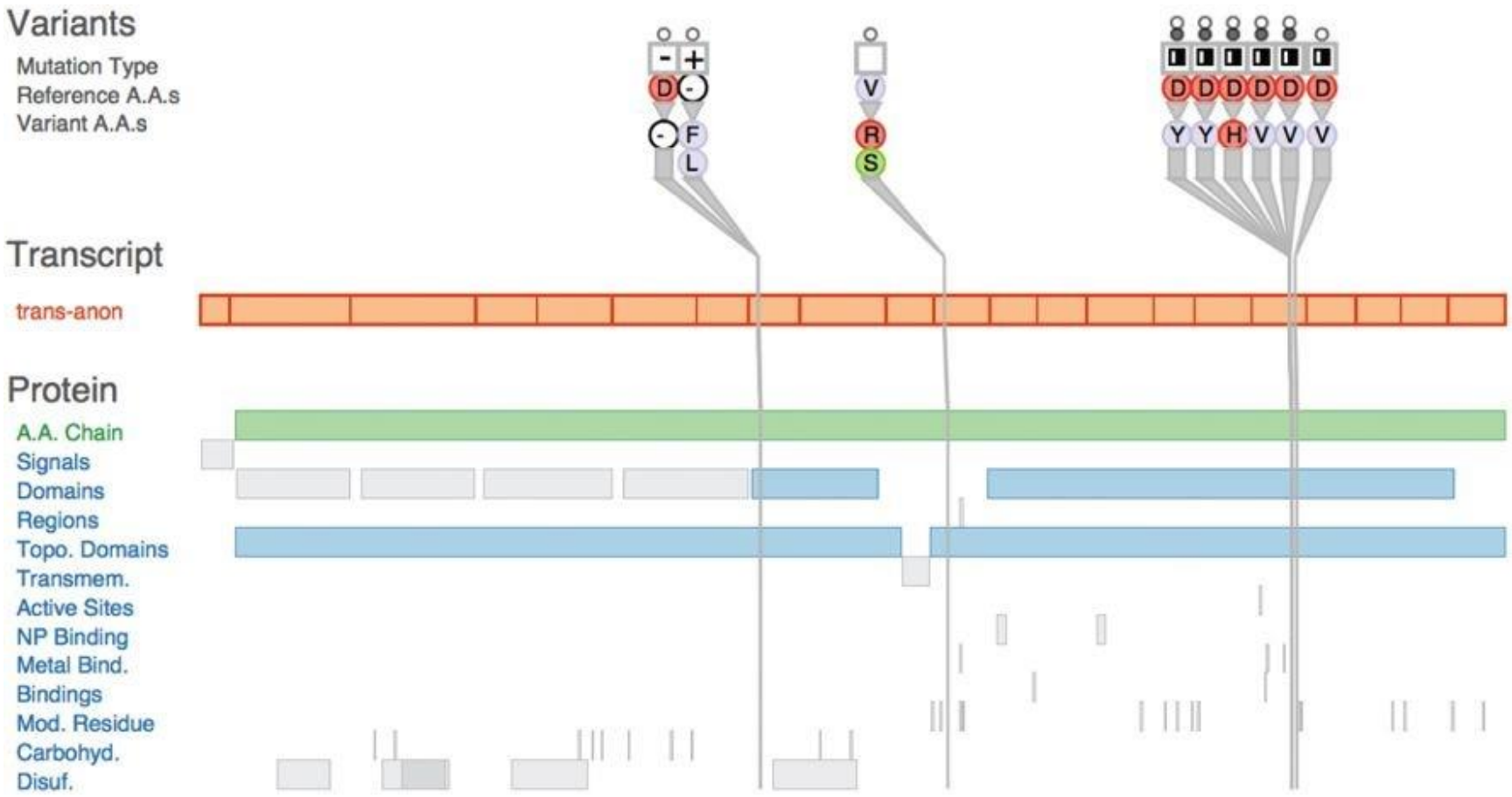
Why have a human in the decision-making loop?

- **Design vis tools for many kinds of uses**
  - **Make a tool intended for transitional use by helping the designers of future solutions that are purely computational.**
    - **Use vis tool to explore data to find questions for automatic methods**
  - **Make a tool for long-term use, in a situation where there is no intention of replacing the human any time soon.**
    - **Use vis to explore data, find questions, and answer the questions.**
  - **Build a tool aimed at the designers of a purely computational solution, to help them refine, debug, or extend that system's algorithms or understand how the algorithms are affected by the changes of parameters**

Why have a human in the decision-making loop?

- **Design vis tools for many kinds of uses**
  - **Design a tool for end users in conjunction with other automatic methods to judge if automatic system is doing right things**
    - **Use vis tool to display the result of automatic methods**
    - **Vis tools are abandoned after the decision is made, or continue to be in play with long-term use to monitor the system**
  - **Design tools for presentation**

Why have a human in the decision-making loop?

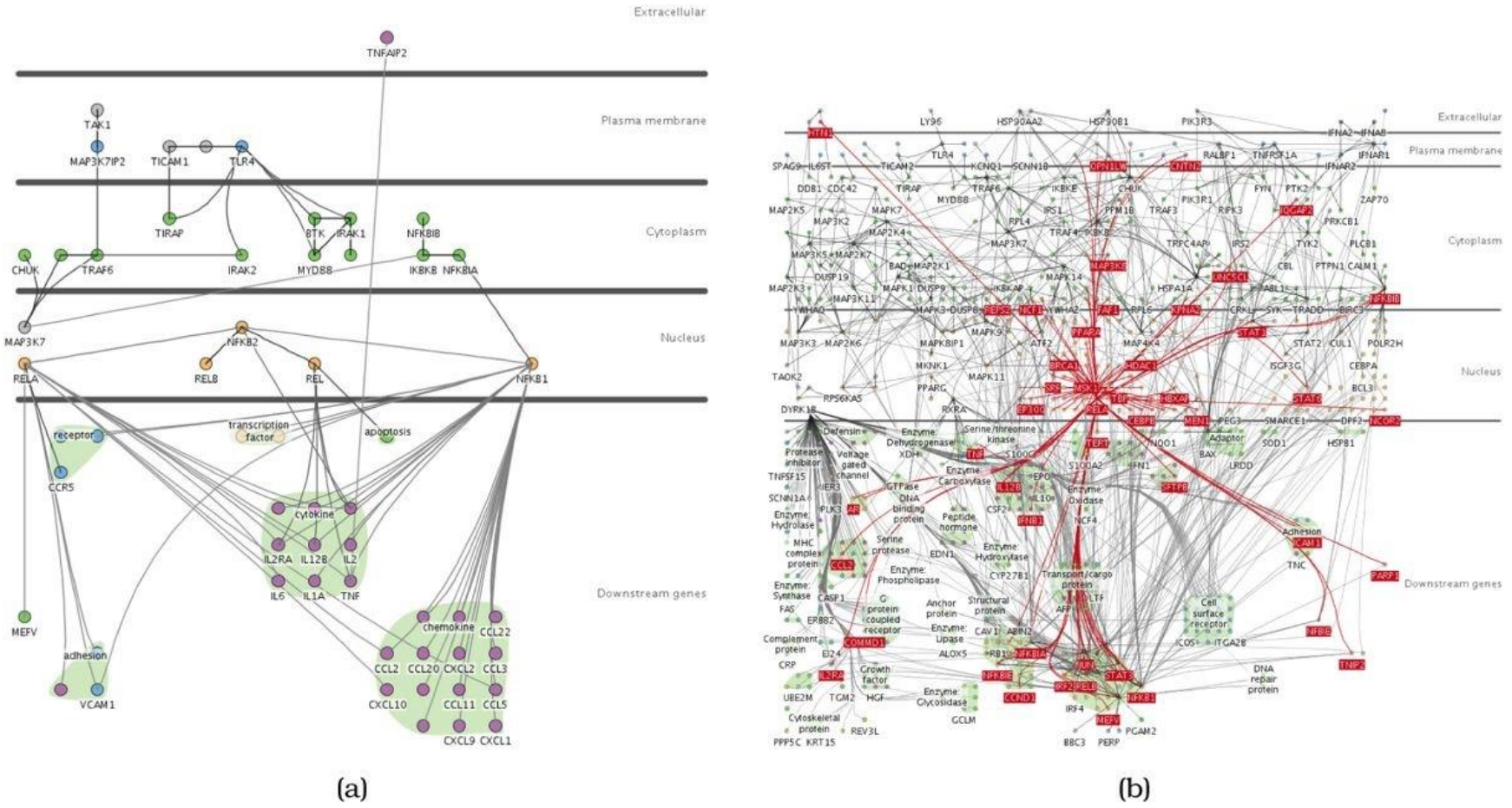


The Variant View vis tool support biologists in assessing the impact of genetic variants by speeding up The exploratory analysis process. [Ferstay et al. 13]

Why have a computer in the loop?

- With computer, we can build tools that allow people to explore or present large datasets that would be infeasible to draw by hand
  - Data set is large
  - Data set changes dynamically over time
- Think about what aspects of hand-drawn diagrams are important to automatically create drawings that retain the hand-drawn spirit

Why have a computer in the loop?



Hand-drawn diagrams in textbooks with vertical layers corresponding to places within a cell where interactions between genes occur. (a) 57 nodes/74 edges network by hand.
(b) 760 nodes/1269 edges + interactive exploration – automatic layout [Barsky et al. 07]

Why use an external representation?

- **External representations augmented human capacity by allowing us surpass the limitations of our own internal cognition and memory.**
- **Vis allows people to offload internal cognition and memory usage to the perceptual system, using carefully designed images as a form of external representations.**
  - **Can take many forms**
    - **Touchable physical objects like an abacus or a knotted string**
    - **2D display**

Why depend on vision?

- **Visualization is based on exploiting the human visual system as communication**
  - **Visual system provides a very high-bandwidth channel to brains**
  - **Visual information processing occurs in parallel**
- **Why don't use other senses?**
  - **Sound is poorly providing overviews of large information spaces compared with vision**
  - **Taste and smell don't have viable recording and reproduction technology at all**

Why show the data in detail?

- **Vis help om situations where seeing the dataset structure in detail is better than seeing only a brief summary of it**
  - **Occurs when exploring the data to find patterns, both to confirm expected ones and find unexpected ones**
  - **Occurs when assessing the validity of a statistical model, to judge whether the model in fact fits the data**

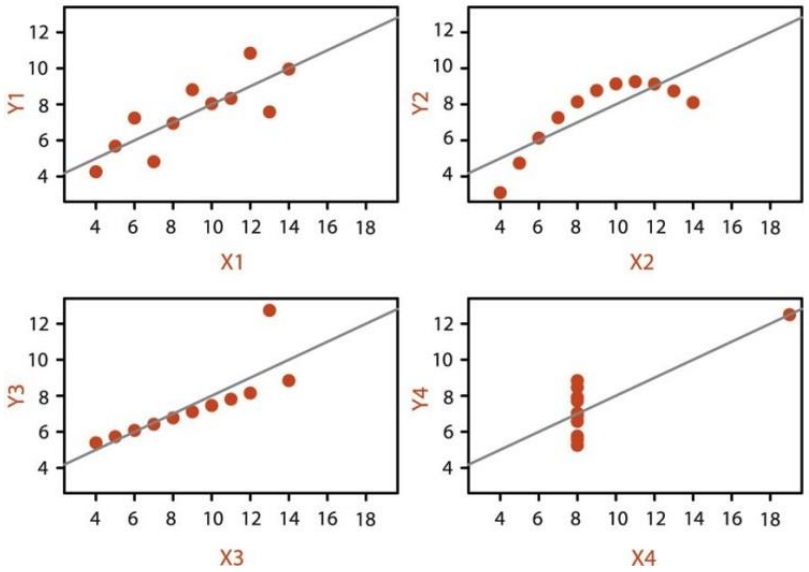
Why show the data in detail?

- **Statistical characterization of datasets is a powerful approach, but it may lose information through summarization.**
  - **datasets that have identical descriptive statistics, such as mean, variance, correlation, and linear regression can have very different structures**
  - **A single summary is often an oversimplification that hides the true structure of the dataset, applies even more to large and complex datasets**

Why show the data in detail?

Anscombe's Quartet: Raw Data

	1		2		3		4	
	X	Y	X	Y	X	Y	X	Y
	10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
	8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
	13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
	9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
	11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
	14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
	6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
	4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
	12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
	7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
	5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89
Mean	9.0	7.5	9.0	7.5	9.0	7.5	9.0	7.5
Variance	10.0	3.75	10.0	3.75	10.0	3.75	10.0	3.75
Correlation	0.816		0.816		0.816		0.816	



Dataset 1: ok
Dataset 2: nonlinear
Dataset 3: a single outlier
Dataset 4:

Why use interactivity?

- **Interactivity is crucial for building vis tools that handle complexity.**
  - **When datasets are large, limitations of people and display preclude just showing everything at once; interaction where user actions cause the view to change is the way forward**
    - **Support investigation at multiple levels of detail, ranging from a very high-level overview down through multiple levels of summarization to a fully detailed view.**
  - **Multiple views**
    - **A single static view can show only one aspect of the dataset**
    - **Multiple views support many possible queries**

Why is the vis idiom design space huge?

- **A vis idiom is a distinct approach to creating and manipulating visual representations**
  - **There are many ways to create a visual encoding of data as a single picture**
  - **The design space gets even bigger when people consider how to manipulate one or more of these pictures with interaction**
- **In this course, we will focus on the framework for thinking about the vis idiom design space systematically by considering a set of design choices**

Why focus on tasks?

- **A tool that serves well for one task can be poorly suited for another, for exactly the same dataset**

- **Refraining the task from domain form into abstract form allows us to consider the similarities and differences what people need across many real-world usage contexts.**

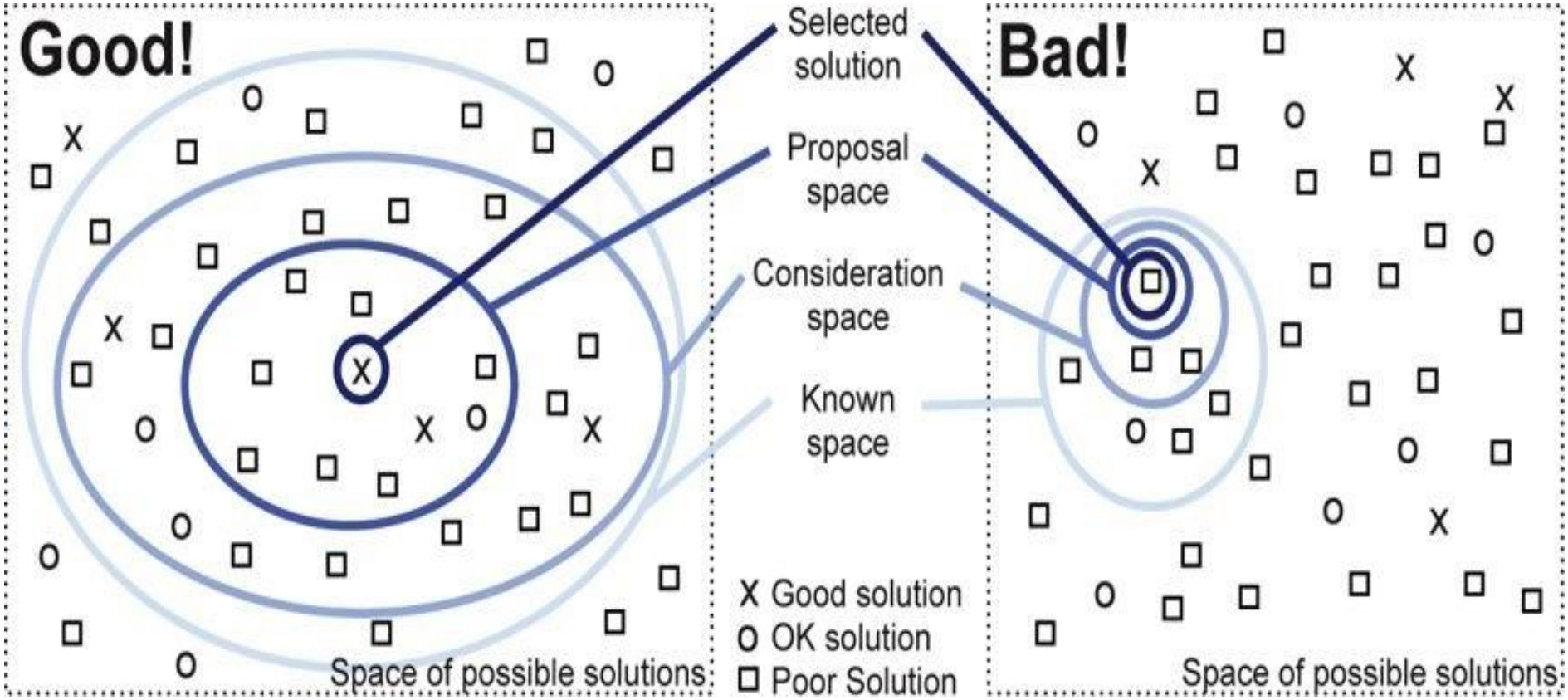
Why care about effectiveness?

- **The focus on effectiveness is a corollary of defining vis to have the goal of supporting user tasks.**
  - **This goal leads to concerns about correctness, accuracy, and truth playing a very central role in vis.**
- **It is not just about making pretty pictures**
  - **The goal is not met if the result is beautiful but not effective**

Why are most designs ineffective?

- **Vis design is difficult**
  - **The vast majority of the possibilities in the design space will be ineffective for any specific usage context**
  - **Only a very small number of possibilities are reasonable choices**
  - **Only an even smaller fraction are excellent choices**
- **Finding the very best solution from the design space is not a useful goal**
- **Finding one of the many possible good solutions is more appropriate**

Why are most designs ineffective?



Small consideration space will have higher probability of only considering Ok or poor solutions and missing a good one.

Why is validation difficult?

- **There are so many questions that you could ask when considering whether a vis tool has met your design goals**
  - **How do you know if it works?**
  - **How do you argue that one is better or worse than another for the intended users?**
  - **What does better mean?**
    - **Users get something done faster?**
    - **They have more fun doing it?**
    - **They work more effectively?**
  - **What does effectively mean?**
  - **What is the user?**
  - **What is faster mean?**

Why are there resource limitations?

- When designing or analyzing a vis system, we must consider at least three different kinds of limitations.
  - computational capacity (time and memory)
  - human perceptual and cognitive capacity
  - display capacity
- Vis systems are inevitably used for larger datasets than those they were designed for
  - Scalability is a central concern
    - Design system to handle large amount of data gracefully

Why are there resource limitations?

- **Computer time and memory are limited resources.**
  - **Soft constraints**
    - **Vis system shares with other programs running on the computer**
  - **Hard constraints**
    - **Even if vis system use nearly all available memory, dataset size can easily outstrip that finite capacity**
  - **Need to design systems that gracefully handle larger datasets that do not fit into core memory**
  - **Major concern**
    - **Algorithms for dataset preprocessing, transformation, layout, and rendering**

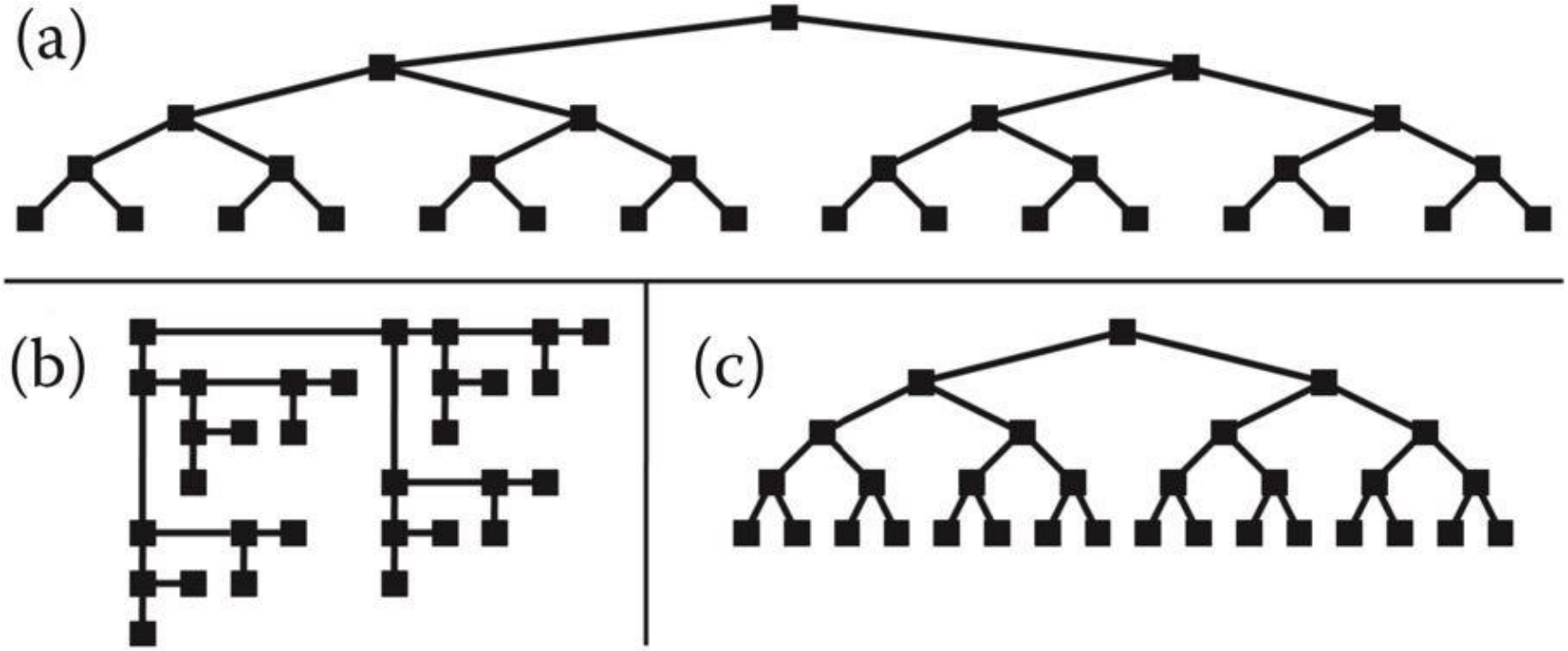
Why are there resource limitations?

- Visual preattentive mechanisms that carry out massively parallel processing
- Human memory for things that are not directly visible is limited
  - Limit for long-term recall
  - Limit for shorter-term working memory
    - Change blindness
      - The phenomenon where even very large change are not noticed if we are attending to something else in our view

Why are there resource limitations?

- **Display capacity is a third kind of limitation to consider**
  - **Visual designer often run out of pixels**
  - **The resolution of screen is not enough to show all desired information simultaneously**
- **The information density of a single image is a measure of the amount of information encoded versus the amount of unused space**
  - **Information density vs. visual encoding**

Why are there resource limitations?



(a) Low information density, and level is clear
(b) High information density, and level is not clear
(c) High information density, and level is clear

Why are there resource limitations?

- **Trade-off between**
  - **Showing as much as possible**
    - **Can minimize the need for navigation and exploration**
  - **The costs of showing too much at once**
    - **User is overwhelmed by visual clutter**
- **The goal of idiom design choices is to find an appropriate balance**

Why analyze?

- **Analyzing existing systems is a good stepping stone to designing new ones.**
  - **Many existing idioms and tools -> a big collection of different possibility, that is good!**
  - **So many possible combinations of data, tasks, and idioms -> hard for designers to find exactly what needed to know by reading papers**
- **We need an analysis framework that imposes a structure on this enormous design space**
  - **Help you think about design choice systematically**

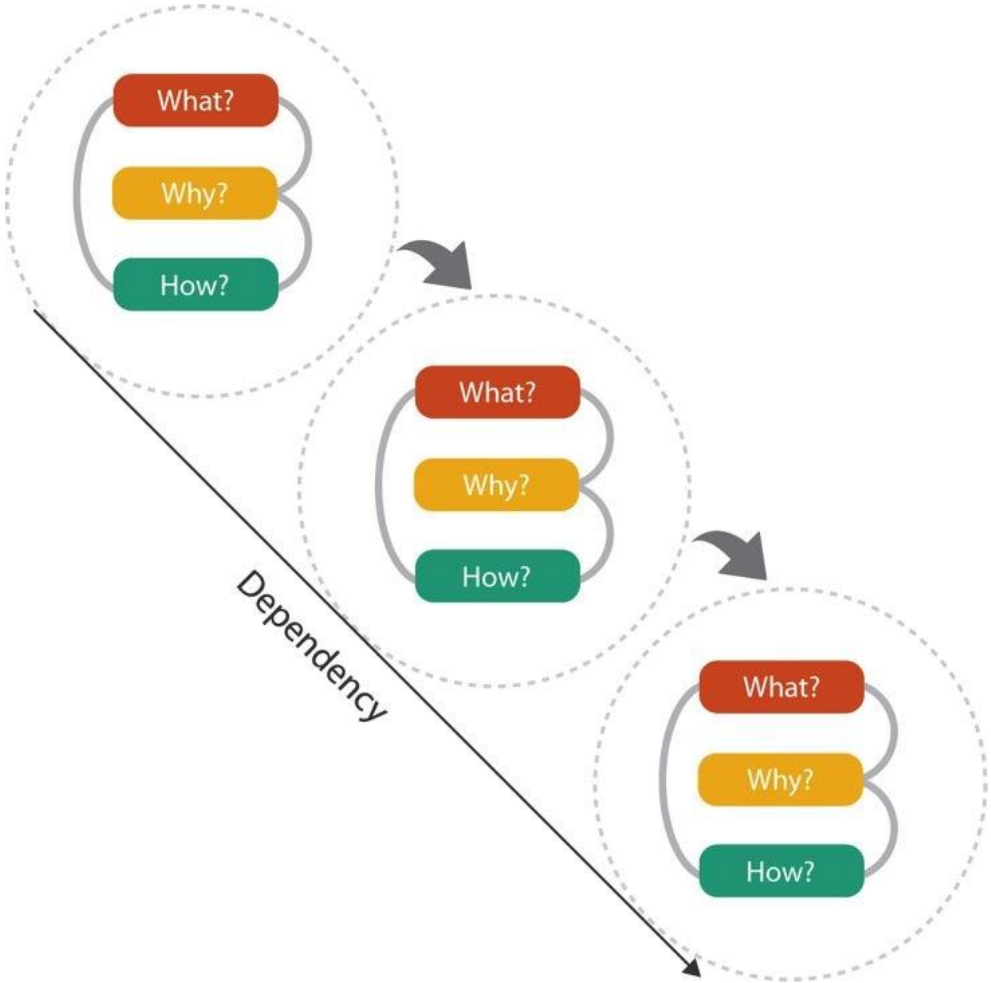
Why analyze?

- **High-level analysis framework**
  - **What: What data users sees**
  - **Why: Why the user intends to use a vis tool**
  - **How: How the visual encoding and interaction idioms are constructed**

Why analyze?

- Each what–why–how question has a corresponding data-task-idiom answer trio
  - One of these analysis trios is called an instance
- Simple vis tools can be fully described as an isolated analysis instance
- Complex vis tool often requires a sequence of instances
  - Express dependence

Why analyze vis?



Analyzing vis usage as chained sequence of instances, where the output of one instance is the input to another