Users, tasks, data

Thomas Torsney-Weir

Overview

Last time

- factors for effective visualization
  - data
  - tasks
  - users
- People have been using visualization for centuries
- Human visual perception is extremely powerful

Designing an effective visualization

- Today: scenario
  - Users
  - Tasks
  - Data
- Next time: rendering
  - Visualization pipeline
  - Visual elements

Why not just data?

Users : Visual literacy, etc

Tasks : Hypotheses, interactions, etc

Data : Information we want to examine

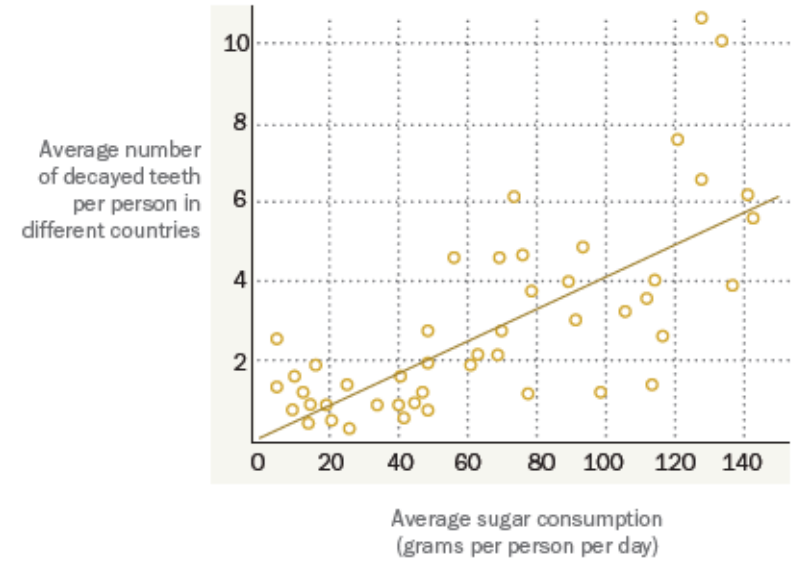
Users

User considerations

- Visualization literacy
- Color blindness
- Patience
  - Development
  - Evaluation
- Training considerations

63% of American Adults Can Correctly Read This Chart

Which of the following statements best describes the data in the graph below?



A. In recent years, the rate of cavities has increased in many countries

B. In some countries, people brush their teeth more frequently than in other countries

C. The more sugar people eat, the more likely they are to get cavities (CORRECT)

D. In recent years, the consumption of sugar has increased in many countries

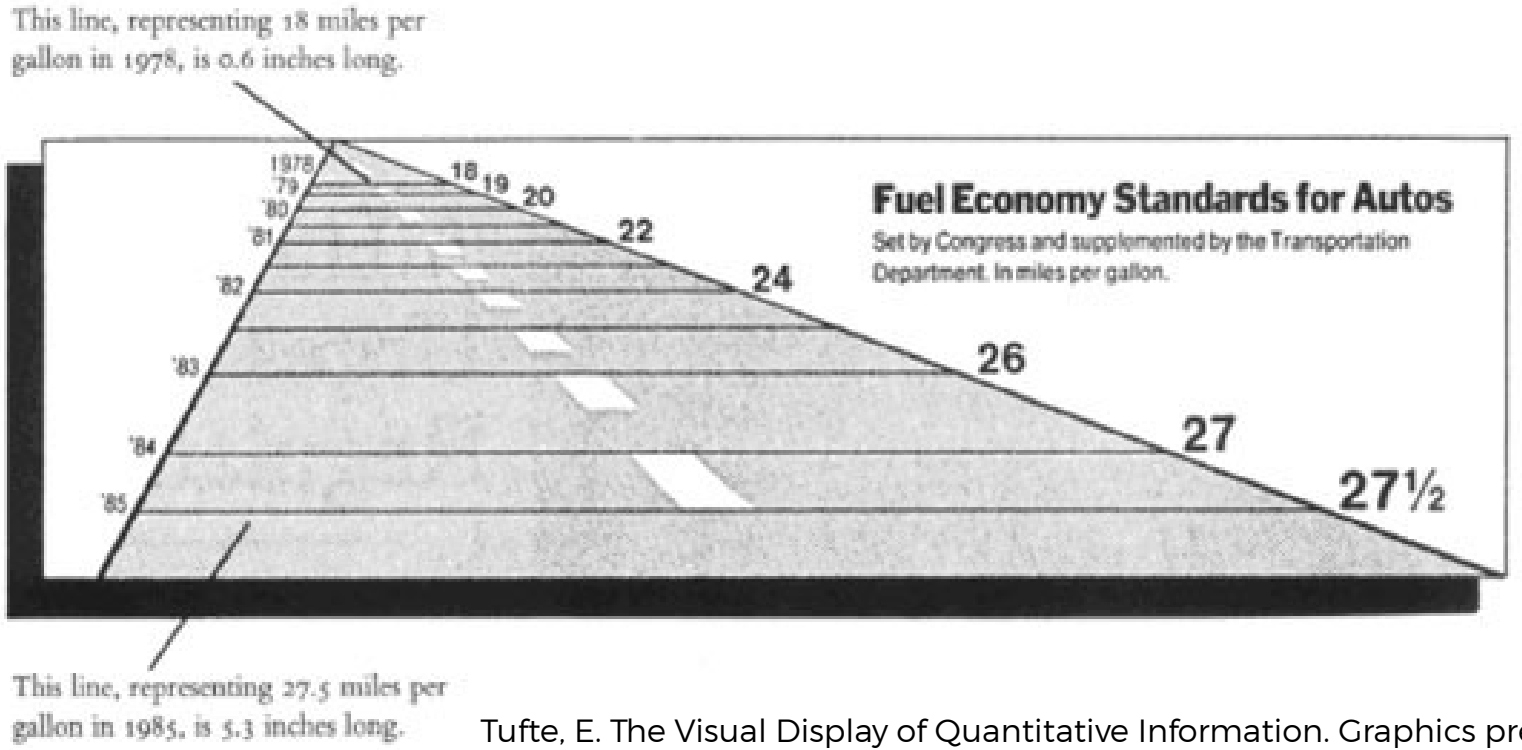
Source: American Trends Panel (wave 6). Survey of U.S. adults conducted Aug. 11-Sept. 3, 2014.

PEW RESEARCH CENTER

https://www.pewresearch.org/fact-tank/2015/09/16/the-art-and-science-of-the-scatterplot/

The lie factor

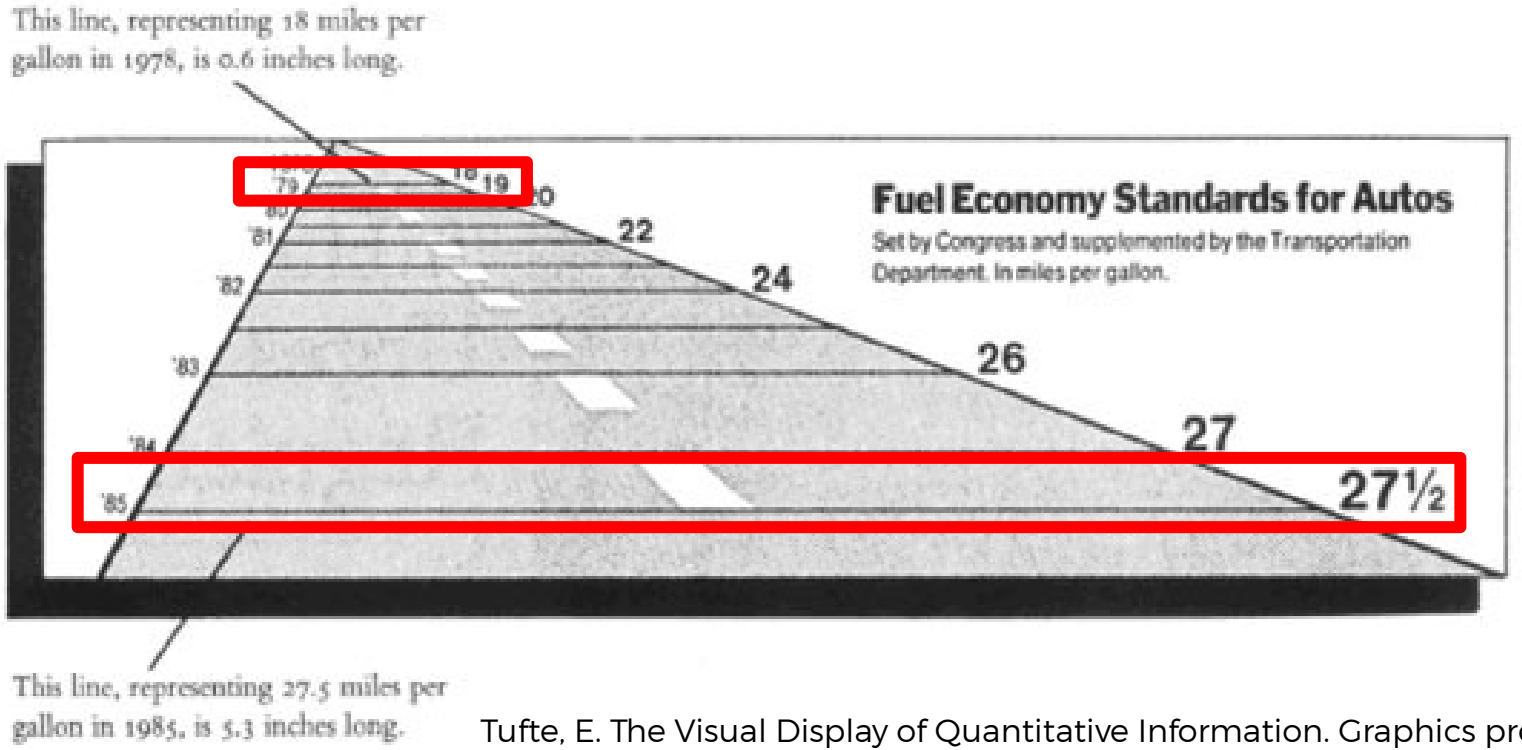
(size of effect in graphic)/(size of effect in data)



Tufte, E. The Visual Display of Quantitative Information. Graphics press, 1983.

The lie factor

$$\frac{\frac{5.3-0.6}{0.6}}{\frac{27.5-18}{18}}$$

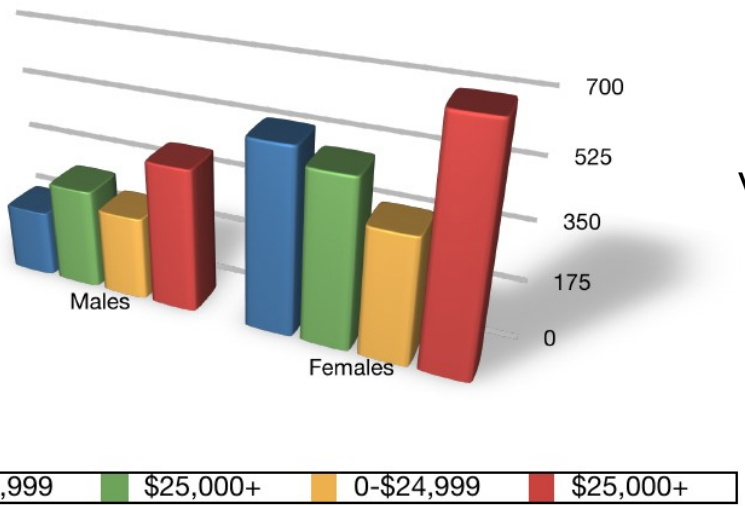


Tufte, E. The Visual Display of Quantitative Information. Graphics press, 1983.

Data-ink ratio

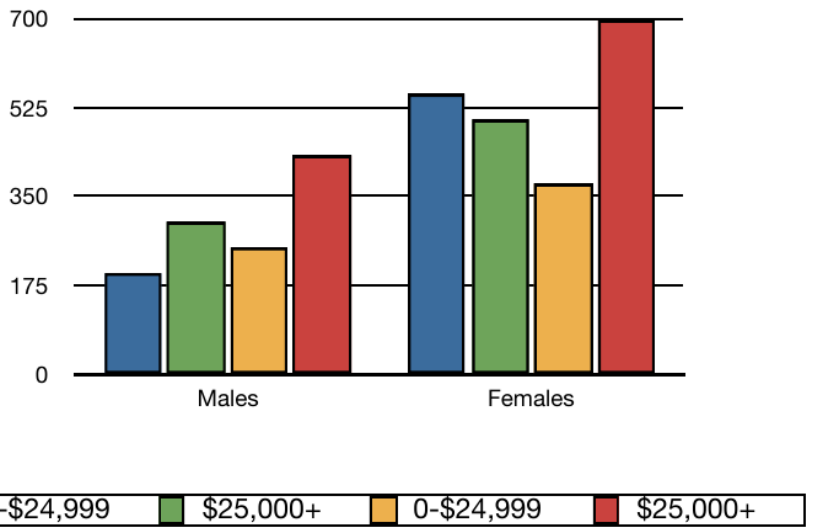
Data-ink : the ink used to show data
Data-ink ratio : data-ink / total ink used

bad



vs

good

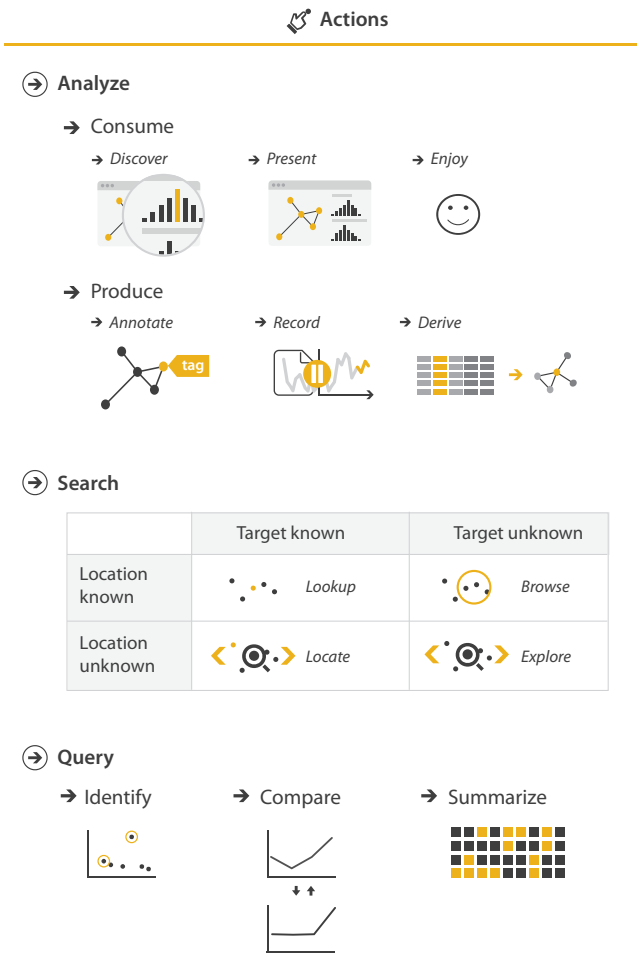


Tasks

What are tasks?

- For what purpose is a particular visualization effective?
- Often described abstractly
- Abstracting tasks allows us to compare across domains
  - "constrast patients in the ICU after one month vs first week"
  - "see if the treated tissues differ from untreated"
  - compare values between 2 groups

Types of tasks



High level : Cure cancer

Mid-level : Identify most influential factors

Low level : Click on line, band select key points

Munzner, T. Visualization Analysis & Design. CRC press, 2014.

Shneiderman's mantra

Overview first, zoom and filter, details on demand

There are many visual design guidelines but the basic principle might be summarized as the Visual Information Seeking Mantra:

Overview first, zoom and filter, then details-on-demand
Overview first, zoom and filter, then details-on-demand
Overview first, zoom and filter, then details-on-demand
Overview first, zoom and filter, then details-on-demand
Overview first, zoom and filter, then details-on-demand
Overview first, zoom and filter, then details-on-demand
Overview first, zoom and filter, then details-on-demand
Overview first, zoom and filter, then details-on-demand
Overview first, zoom and filter, then details-on-demand
Overview first, zoom and filter, then details-on-demand

Shneiderman, B. "The eyes have it: A task by data type taxonomy for information visualizations," Proceedings 1996 IEEE Symposium on Visual Languages. 1996.

Data

What is data?

Data is just information

Semantics : real-world meaning of data

Type : Storage type in computer

e.g. height can be `int`, `double`, or a `Tuple{int,int}`

Attributes

Visualizations are based on data attributes

Quantitative : Continuous numbers (10cm, 75kg, etc)

Ordinal : Categories with order (small, medium, large, eggs in a box)

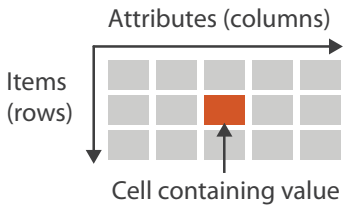
Nominal : Categories with no order (male, female, apples, oranges, etc)

Dataset

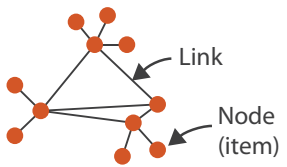
Collection of attributes

➔ **Dataset Types**

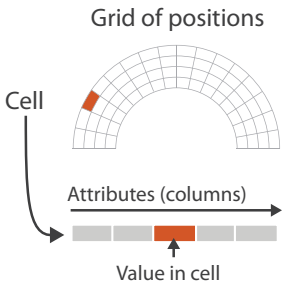
➔ Tables



➔ Networks

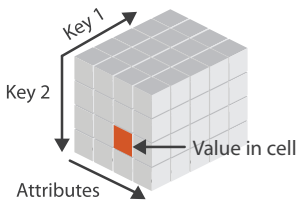


➔ Fields (Continuous)

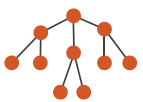


➔ Geometry (Spatial)

➔ *Multidimensional Table*

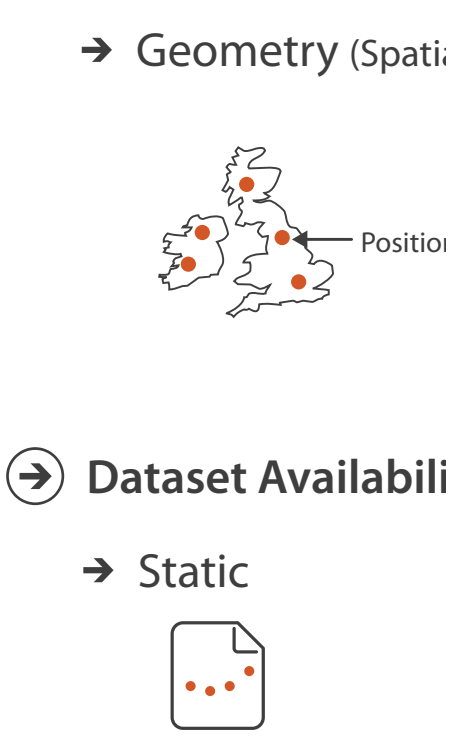


➔ *Trees*



Munzner, T. Visualization Analysis & Design. CRC press, 2014.

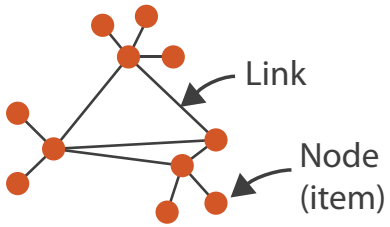
Tabular data



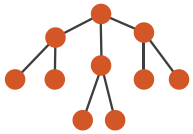
- Likely most common
- Attributes in columns
- Data “item” is a row

Networks

➔ Networks

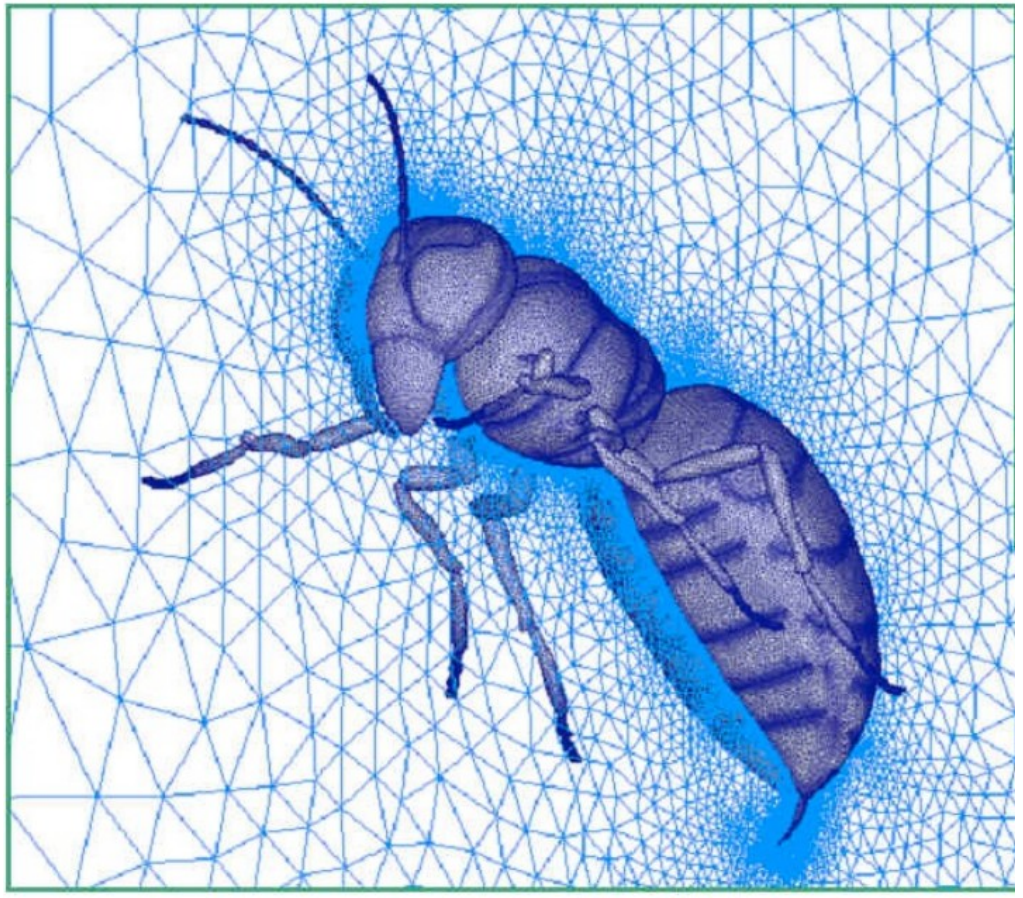


➔ *Trees*

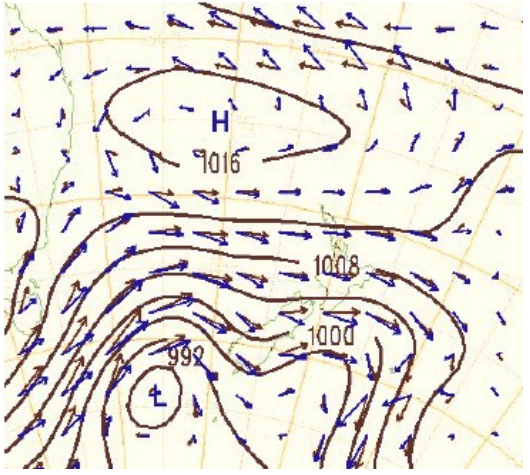


- Nodes have attributes
- Edges have attributes
- Often represented as 2 tables

Fields (continuous)

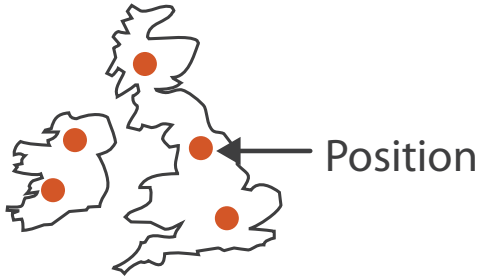


- Common in volume & flow visualization
- Grids can be many types

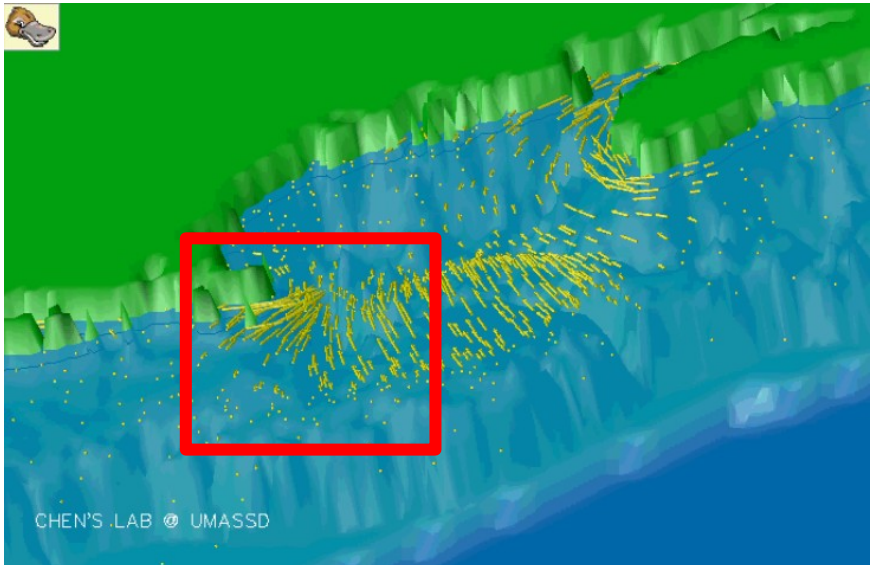


Geometry

➔ Geometry (Spatial)



- Location
- Distance
- Areas of effect

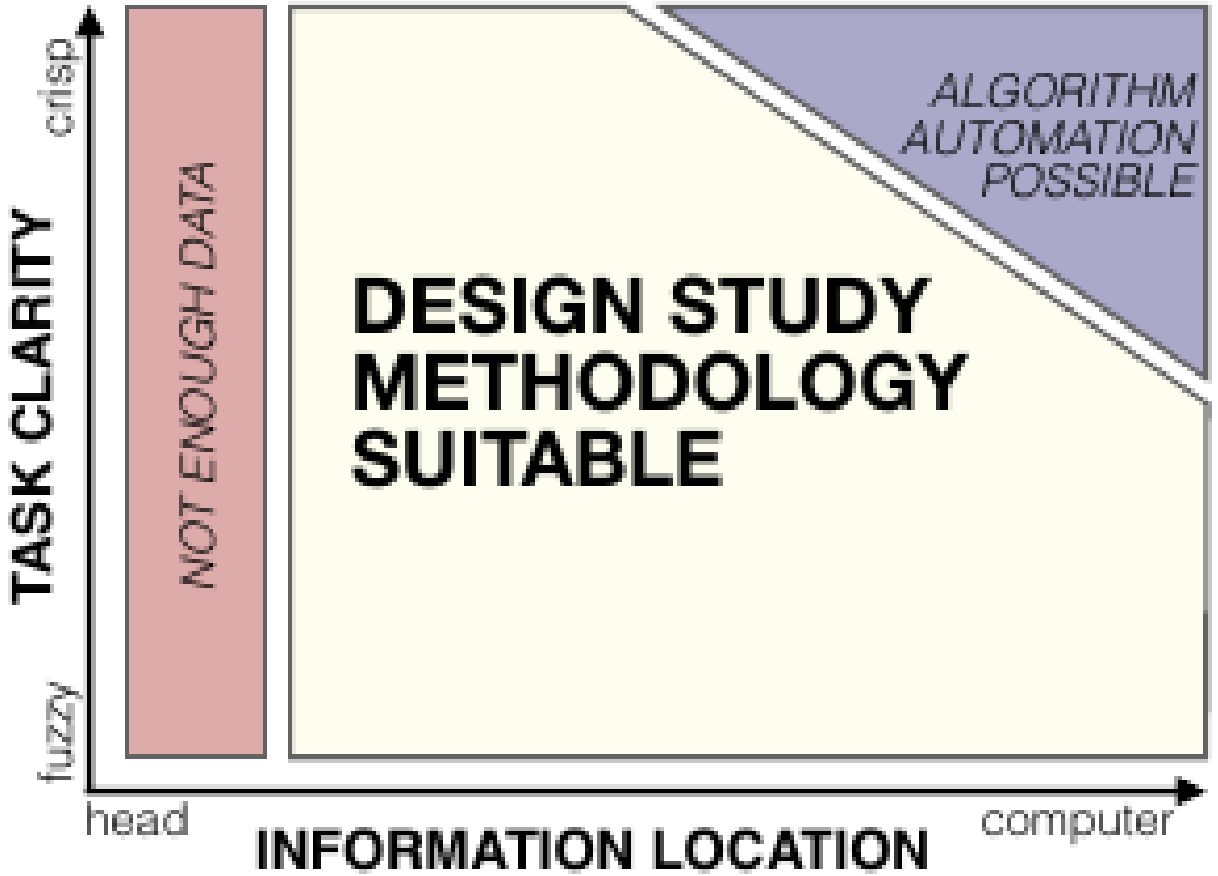


What about text?

- Text is unstructured data
- Examples of encoding into attributes
  - Word counts (aka bag of words)
  - N-grams
  - word2vec

Conclusion

When to use visualization



Sedlmair, Michael, Miriah Meyer, and Tamara Munzner. "Design study methodology: Reflections from the trenches and the stacks," IEEE Transactions on Visualization and Computer Graphics. 2012.

Summary

- Shneiderman's mantra is a good starting point for design
- Humans are diverse like snowflakes
- Task abstractions allow us to compare across domains
- Data has both semantics and type
- Identifying attributes is key to understanding data