Types, Tasks, Marks and Channels

- **Data Types:** *What* people visualize
  - Terminology
  - Tables, Networks, Trees
  - 1-, 2-, 3- and multi-dimensional data
  - Spatial & Temporal data

- **Tasks:** *Why* people use visualizations
  - The Information Seeking Mantra
  - Produce / Consume / Search / Query

- **How** people visualize data
  - Introducing Marks and Channels
  - Accuracy, Discriminability, Separability
  - Graphical Integrity
Data types: Terminology

• **Item** (sometimes also called case): entity described by one or several attributes
  - Examples: car, person, HTTP access event, IP packet

• **Attribute**: specific property whose value can be measured
  - Examples: price, name, URL, source IP

• **Value** of an attribute can be nominal, ordinal, interval, ratio, …
  - should be known from statistics 101 ?!??

• **Link**: relationship between items, typically in a network
  - Examples: friendship between people, route between gateways, can be directed

• **Position**: location in 2D or 3D space, given as coordinates
Value Types with Examples

• Ordered
  - Ordinal: A, B, C, D, ...
  - Quantitative
    • Interval: °C
      • „5 degrees warmer“
    • Ratio: °K
      • zero really means zero
      • „twice as warm“

• Ordering direction:
  - Sequential: °K
  - Diverging: °C
  - Cyclic: (e.g., time)

• Categorical
  - Nominal: D, CH, A, ...
  - Categories: Male, Female
**Dataset types: Terminology**

- **Table**: 2D (x/y) arrangement of data in rows and columns, where
  - each *row* y contains an *item*
  - each *column* x contains an *attribute*
  - each *cell* x/y contains the *value* of attribute x of item y

- **Network** (also called *graph*): set of items and links, where
  - each pair of items (called *nodes*) can be connected by…
  - one or more links (sometimes called *edges*) between them
  - edges can be *directed*
  - both nodes and edges can have *attributes*

- **Tree**: network with a hierarchical structure, i.e., no cycles

- **Field**: array of measurements from a continuous domain (e.g., height)

- **Geometry**: given spatial arrangement, e.g., countries, 3D models, …
1-dimensional data

- Linear list of items with $\geq 1$ attribute each
- Organized in a linear manner (by order of sequence)
- Examples:
  - Arbitrary text
  - List of names
  - Program code
    - Seesoft (Eick et al. 1992)
- NB: 1-dimensional data does not equal 1-dim. display!!
Special case of 1-dimensional: temporal data

- Special **meaning** and **structure** of the ordering dimension
  - Meaning: we cannot move arbitrarily in time
  - Structure: Cyclic nature of time at different levels
  - Example:

[Diagram showing total KW consumption over time]
Example: E.J. Marey's Train Schedule

Discussion: What’s lost in this recent copy?

http://engineering.org.cn/EN/10.1016/J.ENG.2016.03.019
2-dimensional data

- Array of items with $\geq 2$ attributes
- 2 of the attributes can be used for indexing
  - Ideal mapping to the 2 dimensions of screen
- Examples:
  - brightness values at position $(x, y)$ on a light sensor (i.e. digital photo)
  - altitude of the terrain at geographical position $(\text{lat}, \text{lon})$ (i.e. topogr. map)
3-dimensional data

- Array of items with $\geq$ 3 attributes, 3 of which can be used for indexing
- Often represents 3-dimensional objects or spatial data from the real world
- Examples:
  - 3D medical imaging, such as CT scans or MRI
  - Results of measurements or physical simulations
  - 3D models in computer graphics

https://commons.wikimedia.org/wiki/File:MRI.png
http://www.petr-lorenz.com/emgine/
Example from the Visible Human Project

https://www.youtube.com/watch?v=dPPjUtGh4Ys
Multidimensional (>3 = high-dimensional) data

- Can be indexed and accessed along more than 3 dimensions
- Simple Example: Music collection

- Complex example:
  - index of all web pages, such as the google index
  - each word in a (collected) lexicon is a dimension
  - occurrence or frequency of word is value along this dimension
Types, Tasks, Marks and Channels

- **Data Types:** *What* people visualize
  - Terminology
  - Tables, Networks, Trees
  - 1-, 2-, 3- and multi-dimensional data
  - Spatial & Temporal data

- **Tasks:** *Why* people use visualizations
  - The Information Seeking Mantra
  - Produce / Consume / Search / Query

- **How** people visualize data
  - Introducing Marks and Channels
  - Accuracy, Discriminability, Separability
  - Graphical Integrity
Ben Shneiderman’s Information seeking Mantra

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

- **Overview**: Gain an overview of the entire collection.

- **Zoom**: Zoom in on items of interest

- **Filter**: Filter out uninteresting items.

- **Details-on-demand**: Select an item or group and get details when needed.

- **Relate**: View relationships among items.

- **History**: Keep a history of actions to support undo, replay, and progressive refinement.

- **Extract**: Allow extraction of sub-collections and of the query parameters.

User Tasks according to Tamara Munzner (I)

• Produce
  - **annotate** items by adding new information to them
  - **record** the results of (an interaction with) a visualization
  - **derive** new data elements from existing ones, e.g., derived attributes

• Consume
  - **present** the data to a target audience
    • not just static (e.g. interactive graphics in online newspapers)
  - **discover** new information in the data
    • e.g., generate / verify a hypothesis
  - **enjoy** data, for example in a “casual” visualization
    • Example: Name Voyager (see next slide)
Example: Name Voyager

http://www.babynamewizard.com/voyager
User Tasks according to Tamara Munzner (II)

- **Search**

<table>
<thead>
<tr>
<th></th>
<th>target known</th>
<th>target unknown</th>
</tr>
</thead>
<tbody>
<tr>
<td>location known</td>
<td>lookup</td>
<td>browse</td>
</tr>
<tr>
<td>location unknown</td>
<td>locate</td>
<td>explore</td>
</tr>
</tbody>
</table>

- **Query**
  - **identify** a single target
  - **compare** two or multiple targets
  - **summarize** a set of selected targets
Types, Tasks, Marks and Channels

• Data Types: **What** people visualize
  - Terminology
  - Tables, Networks, Trees
  - 1-, 2-, 3- and multi-dimensional data
  - Spatial & Temporal data

• Tasks: **Why** people use visualizations
  - The Information Seeking Mantra
  - Produce / Consume / Search / Query

• **How** people visualize data
  - Introducing Marks and Channels
  - Accuracy, Discriminability, Separability
  - Graphical Integrity
Marks & Channels: Terminology

- **Mark**: basic geometric element that conveys an item or a link

- 0D mark = point
- 1D mark = line
- 2D mark = area
- 3D mark = volume

- **Channel**: controls the appearance of a mark
Visual Channels

- Position
  - horizontal
  - vertical
  - both
- Color
- Shape
- Tilt
- Size
  - length
  - area
  - volume
Visual Language is a Sign System

- Image perceived as a set of signs
- Sender encodes information in signs
- Receiver decodes information from signs

- Jacques Bertin
  - French cartographer [1918-2010]
  - Semiology of Graphics [1967]

- Theoretical principles for visual encodings
According to Bertin ...

**Channels**

- position
- size
- (grey)value
- texture
- color
- orientation
- shape

**Marks**

- points
- lines
- areas

**LES VARIABLES DE L'IMAGE**

<table>
<thead>
<tr>
<th>XY</th>
<th>dimensions du plan</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**LES VARIABLES DE SÉPARATION**

- Grain
- Couleur
- Orientation
- Forme

<table>
<thead>
<tr>
<th>Points</th>
<th>Lignes</th>
<th>Zones</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Describing Examples in terms of Marks & Channels

- Which marks are used?
- Which channels are used?
- Which data types do they encode?
  - good/bad mappings?
- Can all channels be combined?
  - remember preattentive perception?
Pre-Attentively Perceived Visual Channels

Grey value
Elongation
Curvature
Added surround box
Shape
Added surround color
Filled
Sharpness
Cast shadow
Convex and concave
Sharp vertex
Joined lines
Misalignment
Blinking
Direction of motion
Phase of motion

http://visualizedd.wordpress.com/process-paragraph/
Channel Types: Magnitude vs. Identity

**Channels: Expressiveness Types and Effectiveness Ranks**

**Magnitude Channels: Ordered Attributes**
- Position on common scale
- Position on unaligned scale
- Length (1D size)
- Tilt/angle
- Area (2D size)
- Depth (3D position)
- Color luminance
- Color saturation
- Curvature
- Volume (3D size)

**Identity Channels: Categorical Attributes**
- Spatial region
- Color hue
- Motion
- Shape

https://www.safaribooksonline.com/library/view/visualization-analysis-and/9781466508910/image/fig5-1.png
Choosing Color Scales

• Match the attribute type!
  - Categorical
    • how many categories?
  - Ordered
    • Sequential
    • Diverging
    • Cyclic?

• Avoid rainbows if possible ;-)
• Tons of resources on the web

https://betterfigures.org/2015/06/23/picking-a-colour-scale-for-scientific-graphics/
Source of Inspiration: Color Brewer

http://colorbrewer2.org/
Marks and Channels for Grouping

- containment
- connection

- proximity
  - same spatial region
- similarity
  - same values as other categorical channels

**Identity Channels: Categorical Attributes**

- Spatial region
- Color hue
- Motion
- Shape

**Marks as Links**

- Containment
- Connection
Connection & Hierarchy

- marks as links (vs. nodes)
  - common case in network drawing
- 1D case: connection
  - ex: all node-link diagrams
  - emphasizes topology, path tracing
  - networks and trees
- 2D case: containment
  - ex: all treemap variants
  - emphasizes attribute values at leaves (size coding)
  - only trees
Expressiveness and Effectiveness of channels

• **Expressiveness** Principle: A visual representation…
  - should express all of the underlying attributes
  - should not express any attributes not in the data

• **Effectiveness** Principle:
  - Importance of attribute should match salience of channel
  - I.e., the most important attributes should be encoded in the most effective channels
Expressiveness: match channel and data characteristics

**Magnitude Channels: Ordered Attributes**
- Position on common scale
- Position on unaligned scale
- Length (1D size)
- Tilt/angle
- Area (2D size)
- Depth (3D position)
- Color luminance
- Color saturation
- Curvature
- Volume (3D size)

**Identity Channels: Categorical Attributes**
- Spatial region
- Color hue
- Motion
- Shape
Effectiveness: encode most important attributes with highest ranked channels

- **Magnitude Channels: Ordered Attributes**
  - Position on common scale
  - Position on unaligned scale
  - Length (1D size)
  - Tilt/angle
  - Area (2D size)
  - Depth (3D position)
  - Color luminance
  - Color saturation
  - Curvature
  - Volume (3D size)

- **Identity Channels: Categorical Attributes**
  - Spatial region
  - Color hue
  - Motion
  - Shape

Spatial position ranks high for both
Accuracy of a channel

- How well can we judge the encoded value?
- Many perceptual processes follow a power law $S=I^N$
  - $N = 1$ means linear (good!)
  - $N < 1$: scale is compressed
  - $N > 1$: scale is magnified
Discriminability of a channel

- Are the differences between values noticeable in the channel?
- How many different values can be encoded?
  - group values into bins
  - channels vary largely!

Image from Munzner (2014), page 107

Separability of channels

- Do different channels interfere with each other?

- Different levels of interference exist

- Match the characteristics to the data encoded
Zoomout: The Visualization Pipeline (one version)

- **Data Analysis**: prepare data for visualization (e.g., by applying a smoothing filter, interpolating missing values, or correcting erroneous measurements)

- **Filtering**: selection of data portions to be visualized

- **Mapping**: focus data are mapped to marks and channels; most critical step for achieving Expressiveness and Effectiveness.

- **Rendering**: geometric data are transformed to image data.

  - **Goal**: create a fitting **mental model** of the data with your visualization

  - **Brief discussion**: Who does what? Where does the end user interact?
My mother commutes into Munich and loves to knit. For 2018, she has knit a „train delay scarf“: 2 rows per day, grey for <5min., pink for 5-30min. and red for >30min. delay on one or both trips that day.