IAT 814
Visualization

Representation: Design Idioms 1

Lyn Bartram

These slides borrow heavily from T. Munzner and S. Few, and may be incompletely attributed. Work in progress.
Recall: Data Abstractions

- **Tables**
  - Data item (row) with attributes (columns): row = key, cells = values

- **Networks**
  - Item (node) with attributes (features) and relations (links)
  - Trees (hierarchy)
  - Node = key, node-node, link = key, cell = value

- **Text/Logs**
  - Grammar
  - Bag of words
  - Derived values

- **Image**
  - 2d location = key, pixel value expresses single attribute or combo of attributes according to coding (RGB)
Recall: Why
Vis tasks and targets [Munzner]

**Actions**
- Analyze
  - Consume
    - Discover
    - Present
    - Enjoy
  - Produce
    - Annotate
    - Record
    - Derive

**Targets**
- All Data
  - Trends
  - Outliers
  - Features
- Attributes
  - One
    - Distribution
  - Many
    - Dependency
    - Correlation
    - Similarity
- Network Data
  - Topology
  - Networks
- Spatial Data
  - Shape

**Search**

<table>
<thead>
<tr>
<th>Target known</th>
<th>Target unknown</th>
</tr>
</thead>
<tbody>
<tr>
<td>Location known</td>
<td>&lt;_lookup&gt;</td>
</tr>
<tr>
<td>Location unknown</td>
<td>&lt;locate&gt;</td>
</tr>
</tbody>
</table>
User Tasks (why)

• Amar & Stasko created a taxonomy of user tasks in visualization environments
• 10 basic actions
• Retrieve Value, Filter, Compute Derived Value, Find Extremum, Sort, Determine Range, Characterize Distribution, Find Anomalies, Cluster, Correlate
1. Retrieve Value

• General Description:
  • Given a set of specific items, find attributes-related values of those items.

• 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.

• **Examples:**
  • What Kellogg's cereals have high fiber?
  • What comedies have won awards?
  • Which funds underperformed the S&P-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 Extreme

• 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?
  • Which national hockey team has won the most Olympic Gold medals since 1912?
5. Sort/Order

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

• **Examples:**
  • Order the cars by weight.
  • Rank the cereals by calories.
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?
  • What actresses are in the data set?
7. Characterize Distribution

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

• **Examples:**
  - What is the distribution of carbohydrates in cereals?
  - What is the age distribution of shoppers?
7a. Part-whole

• General Description:
  • Given a set of data items that differ in values in one category, reveal the portion that each value represents to some whole.

• Examples:
  • How many Conservatives are women?
  • How do our tax dollars get spent?
  • What is the recommended daily caloric intake across food types for pregnant women?
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 cereals that have high calories but low sugar?
  • Are there exceptions to the relationship between horsepower and acceleration?
9. Cluster

• General Description:
  • Given a set of data cases, find clusters of similar attribute values.
  • Note: you need to determine what “similar” is

• Examples:
  • Are there groups of cereals w/ similar fat/calories/sugar?
  • Are all comedies the same length?
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: 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

- “What’s the variation in cereal sugar content”?  
  - Compute derived value; determine range; characterise distribution
What was left out? Aha analytics!

- **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**
  - “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?”
Fundamental principles: how

Expressiveness:
• the visual encoding should express all of, and only, the information in the dataset attributes

Effectiveness:
• the importance of the attribute should match the salience of the channel.
• Use the strongest and most accurate channels for the most important interpretation tasks (data)
Recall How version 1: Channel Expressiveness and Effectiveness

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

Credit: T. Munzner, 2014
Roadmap so far

Part 1: principles
• Data
• Perception
• Visual encoding
• (interaction to come later)

Part 2: Methods (How v2!)
• What defines the design space?
• Taxonomy of design considerations
• How many views?
• How to reduce
A Framework for Analysis (Munzner)
• A visualisation **idiom** is a distinct approach to creating and manipulating visual representations.

• **Data**: the types and hierarchical salience of the information to represent
  • **Design**: the visual encoding and organisation choices
  • **Interaction**: the methods to man
<table>
<thead>
<tr>
<th>ID</th>
<th>Sepal Length</th>
<th>Sepal Width</th>
<th>Petal Length</th>
<th>Petal Width</th>
<th>Species</th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
<td>4.3</td>
<td>3</td>
<td>1.1</td>
<td>0.1</td>
<td>setosa</td>
</tr>
<tr>
<td>39</td>
<td>4.4</td>
<td>3</td>
<td>1.3</td>
<td>0.2</td>
<td>setosa</td>
</tr>
<tr>
<td>43</td>
<td>4.4</td>
<td>3.2</td>
<td>1.3</td>
<td>0.2</td>
<td>setosa</td>
</tr>
<tr>
<td>9</td>
<td>4.4</td>
<td>2.9</td>
<td>1.4</td>
<td>0.2</td>
<td>setosa</td>
</tr>
<tr>
<td>42</td>
<td>4.5</td>
<td>2.3</td>
<td>1.3</td>
<td>0.3</td>
<td>setosa</td>
</tr>
<tr>
<td>23</td>
<td>4.6</td>
<td>3.6</td>
<td>1</td>
<td>0.2</td>
<td>setosa</td>
</tr>
<tr>
<td>48</td>
<td>4.6</td>
<td>3.2</td>
<td>1.4</td>
<td>0.2</td>
<td>setosa</td>
</tr>
<tr>
<td>4</td>
<td>4.6</td>
<td>3.1</td>
<td>1.5</td>
<td>0.2</td>
<td>setosa</td>
</tr>
<tr>
<td>7</td>
<td>4.6</td>
<td>3.4</td>
<td>1.4</td>
<td>0.3</td>
<td>setosa</td>
</tr>
<tr>
<td>3</td>
<td>4.7</td>
<td>3.2</td>
<td>1.3</td>
<td>0.2</td>
<td>setosa</td>
</tr>
<tr>
<td>30</td>
<td>4.7</td>
<td>3.2</td>
<td>1.6</td>
<td>0.2</td>
<td>setosa</td>
</tr>
<tr>
<td>13</td>
<td>4.8</td>
<td>3</td>
<td>1.4</td>
<td>0.1</td>
<td>setosa</td>
</tr>
<tr>
<td>12</td>
<td>4.8</td>
<td>3.4</td>
<td>1.6</td>
<td>0.2</td>
<td>setosa</td>
</tr>
<tr>
<td>31</td>
<td>4.8</td>
<td>3.1</td>
<td>1.6</td>
<td>0.2</td>
<td>setosa</td>
</tr>
<tr>
<td>25</td>
<td>4.8</td>
<td>3.4</td>
<td>1.9</td>
<td>0.2</td>
<td>setosa</td>
</tr>
<tr>
<td>46</td>
<td>4.8</td>
<td>3</td>
<td>1.4</td>
<td>0.3</td>
<td>setosa</td>
</tr>
<tr>
<td>38</td>
<td>4.9</td>
<td>3.6</td>
<td>1.4</td>
<td>0.1</td>
<td>setosa</td>
</tr>
<tr>
<td>10</td>
<td>4.9</td>
<td>3.1</td>
<td>1.5</td>
<td>0.1</td>
<td>setosa</td>
</tr>
<tr>
<td>2</td>
<td>4.9</td>
<td>3</td>
<td>1.4</td>
<td>0.2</td>
<td>setosa</td>
</tr>
<tr>
<td>35</td>
<td>4.9</td>
<td>3.1</td>
<td>1.5</td>
<td>0.2</td>
<td>setosa</td>
</tr>
<tr>
<td>58</td>
<td>4.9</td>
<td>2.4</td>
<td>3.3</td>
<td>1</td>
<td>versicolor</td>
</tr>
<tr>
<td>107</td>
<td>4.9</td>
<td>2.5</td>
<td>4.5</td>
<td>1.7</td>
<td>virginica</td>
</tr>
<tr>
<td>36</td>
<td>5</td>
<td>3.2</td>
<td>1.2</td>
<td>0.2</td>
<td>setosa</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>3.6</td>
<td>1.4</td>
<td>0.2</td>
<td>setosa</td>
</tr>
<tr>
<td>50</td>
<td>5</td>
<td>3.3</td>
<td>1.4</td>
<td>0.2</td>
<td>setosa</td>
</tr>
</tbody>
</table>

Sepia and petal length for three species of iris [Fisher 1936]
Possible views – scatter plot – why?

Distribution and correlations between variables
Possible views – scatter plot – why?

Easy with 2/3 dimensions
Harder with more!!
We can remodel the data

<table>
<thead>
<tr>
<th>ID</th>
<th>Species No</th>
<th>Organ</th>
<th>Length</th>
<th>Width</th>
<th>Species Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.1</td>
<td>sepal</td>
<td>5.1</td>
<td>3.5</td>
<td>setosa</td>
</tr>
<tr>
<td>2</td>
<td>1.1</td>
<td>sepal</td>
<td>4.9</td>
<td>3</td>
<td>setosa</td>
</tr>
<tr>
<td>3</td>
<td>1.1</td>
<td>sepal</td>
<td>4.7</td>
<td>3.2</td>
<td>setosa</td>
</tr>
<tr>
<td>4</td>
<td>1.1</td>
<td>sepal</td>
<td>4.6</td>
<td>3.1</td>
<td>setosa</td>
</tr>
<tr>
<td>5</td>
<td>1.1</td>
<td>sepal</td>
<td>5</td>
<td>3.6</td>
<td>setosa</td>
</tr>
<tr>
<td>6</td>
<td>1.1</td>
<td>sepal</td>
<td>5.4</td>
<td>3.9</td>
<td>setosa</td>
</tr>
<tr>
<td>7</td>
<td>1.1</td>
<td>sepal</td>
<td>4.6</td>
<td>3.4</td>
<td>setosa</td>
</tr>
<tr>
<td>8</td>
<td>1.1</td>
<td>sepal</td>
<td>5</td>
<td>3.4</td>
<td>setosa</td>
</tr>
<tr>
<td>9</td>
<td>1.1</td>
<td>sepal</td>
<td>4.4</td>
<td>2.9</td>
<td>setosa</td>
</tr>
<tr>
<td>10</td>
<td>1.1</td>
<td>sepal</td>
<td>4.9</td>
<td>3.1</td>
<td>setosa</td>
</tr>
<tr>
<td>11</td>
<td>1.1</td>
<td>sepal</td>
<td>5.4</td>
<td>3.7</td>
<td>setosa</td>
</tr>
<tr>
<td>12</td>
<td>1.1</td>
<td>sepal</td>
<td>4.8</td>
<td>3.4</td>
<td>setosa</td>
</tr>
<tr>
<td>13</td>
<td>1.1</td>
<td>sepal</td>
<td>4.8</td>
<td>3</td>
<td>setosa</td>
</tr>
<tr>
<td>14</td>
<td>1.1</td>
<td>sepal</td>
<td>4.3</td>
<td>3</td>
<td>setosa</td>
</tr>
<tr>
<td>15</td>
<td>1.1</td>
<td>sepal</td>
<td>5.8</td>
<td>4</td>
<td>setosa</td>
</tr>
<tr>
<td>16</td>
<td>1.1</td>
<td>sepal</td>
<td>5.7</td>
<td>4.4</td>
<td>setosa</td>
</tr>
<tr>
<td>17</td>
<td>1.1</td>
<td>sepal</td>
<td>5.4</td>
<td>3.9</td>
<td>setosa</td>
</tr>
<tr>
<td>18</td>
<td>1.1</td>
<td>sepal</td>
<td>5.1</td>
<td>3.5</td>
<td>setosa</td>
</tr>
<tr>
<td>19</td>
<td>1.1</td>
<td>sepal</td>
<td>5.7</td>
<td>3.8</td>
<td>setosa</td>
</tr>
<tr>
<td>20</td>
<td>1.1</td>
<td>sepal</td>
<td>5.1</td>
<td>3.8</td>
<td>setosa</td>
</tr>
<tr>
<td>21</td>
<td>1.1</td>
<td>sepal</td>
<td>5.4</td>
<td>3.4</td>
<td>setosa</td>
</tr>
<tr>
<td>22</td>
<td>1.1</td>
<td>sepal</td>
<td>5.1</td>
<td>3.7</td>
<td>setosa</td>
</tr>
<tr>
<td>23</td>
<td>1.1</td>
<td>sepal</td>
<td>5.1</td>
<td>3.6</td>
<td>setosa</td>
</tr>
<tr>
<td>24</td>
<td>1.1</td>
<td>sepal</td>
<td>5.1</td>
<td>3.3</td>
<td>setosa</td>
</tr>
<tr>
<td>25</td>
<td>1.1</td>
<td>sepal</td>
<td>4.8</td>
<td>3.4</td>
<td>setosa</td>
</tr>
</tbody>
</table>

• Add abstraction
Scatter plots show correlations

Perfect positive : $r=1$

Strong positive : $r=0.97$

Positive : $r=0.8$

Strong negative : $r=10.98$

No correlation: $r = 0.16$

Nonlinear correlation
However …

- Scatter plots can be difficult to understand
- What alternatives are there?
- More generally, what kinds of techniques are best for what kinds of problems?
Common 2D design idioms
Design choices for tabular data

Tabular data are Key→Value vectors.

- **Key/Attribute**: property of the data that can be used to index into (sort by/look up) the set. (independent variable)
  - \( N, O \)

- **Value**: the actual value of an individual item
  - \( N,O,Q, \)
If all you want is a single precise value ....

<table>
<thead>
<tr>
<th>Year</th>
<th>Jan</th>
<th>Feb</th>
<th>Mar</th>
<th>Apr</th>
<th>May</th>
<th>Jun</th>
<th>Jul</th>
<th>Aug</th>
<th>Sep</th>
<th>Oct</th>
<th>Nov</th>
<th>Dec</th>
<th>Annual</th>
</tr>
</thead>
<tbody>
<tr>
<td>1990</td>
<td>127.4</td>
<td>128.0</td>
<td>128.7</td>
<td>128.9</td>
<td>129.2</td>
<td>129.9</td>
<td>130.4</td>
<td>131.6</td>
<td>132.7</td>
<td>133.5</td>
<td>133.8</td>
<td>133.8</td>
<td>130.7</td>
</tr>
<tr>
<td>1991</td>
<td>134.6</td>
<td>134.8</td>
<td>135.0</td>
<td>135.2</td>
<td>135.6</td>
<td>136.0</td>
<td>136.2</td>
<td>136.6</td>
<td>137.2</td>
<td>137.4</td>
<td>137.8</td>
<td>137.9</td>
<td>136.2</td>
</tr>
<tr>
<td>1992</td>
<td>138.1</td>
<td>138.6</td>
<td>139.3</td>
<td>139.5</td>
<td>139.7</td>
<td>140.2</td>
<td>140.5</td>
<td>140.9</td>
<td>141.3</td>
<td>141.8</td>
<td>142.0</td>
<td>141.9</td>
<td>140.3</td>
</tr>
<tr>
<td>1993</td>
<td>142.6</td>
<td>143.1</td>
<td>143.6</td>
<td>144.0</td>
<td>144.2</td>
<td>144.4</td>
<td>144.4</td>
<td>144.8</td>
<td>145.1</td>
<td>145.7</td>
<td>145.8</td>
<td>145.8</td>
<td>144.5</td>
</tr>
<tr>
<td>1994</td>
<td>146.2</td>
<td>146.7</td>
<td>147.2</td>
<td>147.4</td>
<td>147.5</td>
<td>148.0</td>
<td>148.4</td>
<td>149.0</td>
<td>149.4</td>
<td>149.5</td>
<td>149.7</td>
<td>149.7</td>
<td>148.2</td>
</tr>
<tr>
<td>1995</td>
<td>150.3</td>
<td>150.9</td>
<td>151.4</td>
<td>151.9</td>
<td>152.2</td>
<td>152.5</td>
<td>152.5</td>
<td>152.9</td>
<td>153.2</td>
<td>153.7</td>
<td>153.6</td>
<td>153.5</td>
<td>152.4</td>
</tr>
<tr>
<td>1996</td>
<td>154.4</td>
<td>154.9</td>
<td>155.7</td>
<td>156.3</td>
<td>156.6</td>
<td>156.7</td>
<td>157.0</td>
<td>157.3</td>
<td>157.8</td>
<td>158.3</td>
<td>158.6</td>
<td>158.6</td>
<td>156.9</td>
</tr>
<tr>
<td>1997</td>
<td>159.1</td>
<td>159.6</td>
<td>160.0</td>
<td>160.2</td>
<td>160.1</td>
<td>160.3</td>
<td>160.5</td>
<td>160.8</td>
<td>161.2</td>
<td>161.6</td>
<td>161.5</td>
<td>161.3</td>
<td>160.5</td>
</tr>
<tr>
<td>1998</td>
<td>161.6</td>
<td>161.9</td>
<td>162.2</td>
<td>162.5</td>
<td>162.8</td>
<td>163.0</td>
<td>163.2</td>
<td>163.4</td>
<td>163.6</td>
<td>164.0</td>
<td>164.0</td>
<td>163.9</td>
<td>163.0</td>
</tr>
<tr>
<td>1999</td>
<td>164.3</td>
<td>164.5</td>
<td>165.0</td>
<td>166.2</td>
<td>166.2</td>
<td>166.2</td>
<td>166.7</td>
<td>167.1</td>
<td>167.9</td>
<td>168.2</td>
<td>168.3</td>
<td>168.3</td>
<td>166.6</td>
</tr>
<tr>
<td>2000</td>
<td>168.8</td>
<td>169.8</td>
<td>171.2</td>
<td>171.3</td>
<td>171.5</td>
<td>172.4</td>
<td>172.8</td>
<td>172.8</td>
<td>173.7</td>
<td>174.0</td>
<td>174.1</td>
<td>174.0</td>
<td>172.2</td>
</tr>
<tr>
<td>2001</td>
<td>175.1</td>
<td>175.8</td>
<td>176.2</td>
<td>176.9</td>
<td>177.7</td>
<td>178.0</td>
<td>177.5</td>
<td>177.5</td>
<td>178.3</td>
<td>177.7</td>
<td>177.4</td>
<td>176.7</td>
<td>177.1</td>
</tr>
<tr>
<td>2002</td>
<td>177.1</td>
<td>177.8</td>
<td>178.8</td>
<td>179.8</td>
<td>179.8</td>
<td>179.9</td>
<td>180.1</td>
<td>180.7</td>
<td>181.0</td>
<td>181.3</td>
<td>181.3</td>
<td>180.9</td>
<td>179.9</td>
</tr>
</tbody>
</table>
The first question: Table or graph?

• Will the data be used to look up and compare individual values, or will the data need to be precise? If so, you should display it in a table.

• Is the message contained in the shape of the data—in trends, patterns, exceptions, or comparisons that involve more than a few values? If so, you should display it in a graph.

• NOTE: (You can use both. Next time and beyond).
Key-value question defines idiom choice (1) [Munzner]

- **2 values**

- **1 Key and 1 value**

- **2 Keys and 1 value**

![Graph showing relationship between circumference and height.](image)
7 interesting relationships
1. Time Series
   • Quantitative across equal intervals (of time)
2. Ranking
   • Sequenced by size of attribute value
3. Part-Whole
   • Portion that each value represents to some whole,
4. Deviation
   • Differ from reference (baseline)
5. Distribution
6. Correlation
   • How one value affects another
7. Nominal
   • Simple categorical
The Power of Space

Categorical

What/where
Planar position
Hue
Shape
Stipple/texture

Relational/Same category

Grouping
Containment (2D)
Connection
Similarity (other channels)
Proximity (position)

Ordered/Quantitative

How Much
Position common scale
Position unaligned scale
Length
Tilt/angle
Area
Curvature
Lightness
Saturation
Texture
density
The Power of Space

Categorical

- What/where
  - Planar position
  - Hue
  - Shape
  - Stipple/texture

Relational/Same category

- Grouping
  - Containment (2D)
  - Connection
  - Similarity (other channels)
  - Proximity (position)

Ordered/Quantitative

- How Much
  - Position common scale
  - Position unaligned scale
  - Length
  - Tilt/angle
  - Area
  - Curvature
  - Lightness
  - Saturation
  - Texture
  - Density
We encode data spatially to

- **Express** (show) values
- **Arrange** data groupings
  - **Separate** /distinguish regions by categorical **key**
  - **Order** groups by **ordinal key**
  - **Align** for visual comparison along **reference value**
Space

Spatial Channels
- Values
  - Express
- Regions
- Separate
- order
- Align
  - 1D
  - 2D
- Given Use

Geographic Fields
- scalar

Spatial Layouts
- Parallel
- Rectilinear
- Radial
- Spacefilling
- Dense
Design choices [Munzner]

- **Arrange Tables**
  - Express Values

- **Separate, Order, Align Regions**
  - Separate
  - Order
  - Align
    - 1 Key List
    - 2 Keys Matrix
    - 3 Keys Volume
    - Many Keys Recursive Subdivision

- **Axis Orientation**
  - Rectilinear
  - Parallel
  - Radial

- **Layout Density**
  - Dense
  - Space-Filling
Single view methods

- All information integrated in one view
- basic visual encodings
  - spatial position
  - color
  - other channels
  - pixel-oriented techniques
- visual layering
  - global compositing
  - item-level stacking
- glyphs
Expressing values

- Scatterplots
- Axes encode 2D values
- Color for key/category
- Additional value in size
Scatterplot idiom

<table>
<thead>
<tr>
<th>Idiom</th>
<th>Scatterplots</th>
</tr>
</thead>
<tbody>
<tr>
<td>What: Data</td>
<td>Table: two quantitative value attributes.</td>
</tr>
<tr>
<td>How: Encode</td>
<td>Express values with horizontal and vertical spatial position and point marks.</td>
</tr>
<tr>
<td>Why: Task</td>
<td>Find trends, outliers, distribution, correlation; locate clusters.</td>
</tr>
<tr>
<td>Scale</td>
<td>Items: hundreds.</td>
</tr>
</tbody>
</table>
Data transformations can enhance value

Log transformations show strong correlation between size and price
Organising data

- List alignment
- Can be ordered along the list axis or by value
- Nominal and ranking (if axis ordered)
- Emphasises individual values
Bar chart idiom

- Categorical attributes match well with spatial **regions**
- **Separate, order, align**
- Can be hard to find patterns in the data “shape”

<table>
<thead>
<tr>
<th>Idiom</th>
<th>Bar Charts</th>
</tr>
</thead>
<tbody>
<tr>
<td>What: Data</td>
<td>Table: one quantitative value attribute, one categorical key attribute.</td>
</tr>
<tr>
<td>How: Encode</td>
<td>Line marks, express value attribute with aligned vertical position, separate key attribute with horizontal position.</td>
</tr>
<tr>
<td>Why: Task</td>
<td>Lookup and compare values.</td>
</tr>
<tr>
<td>Scale</td>
<td>Key attribute: dozens to hundreds of levels.</td>
</tr>
</tbody>
</table>

Credit: T. Munzner, 2014
Stacked bars

- Multidimensional tables with 2 keys/attributes
- Typically use colour or texture for 2nd

<table>
<thead>
<tr>
<th>Idiom</th>
<th>Stacked Bar Charts</th>
</tr>
</thead>
<tbody>
<tr>
<td>What: Data</td>
<td>Multidimensional table: one quantitative value attribute, two categorical key attributes.</td>
</tr>
<tr>
<td>How: Encode</td>
<td>Bar glyph with length-coded subcomponents of value attribute for each category of secondary key attribute. Separate bars by category of primary key attribute.</td>
</tr>
<tr>
<td>Why: Task</td>
<td>Part-to-whole relationship, lookup values, find trends.</td>
</tr>
<tr>
<td>Scale</td>
<td>Key attribute (main axis): dozens to hundreds of levels. Key attribute (stacked glyph axis): several to one dozen</td>
</tr>
</tbody>
</table>
Few’s correlation bar graph

- Inches

- U.S. Dollars

- Height

- Salary
Paired Bar graph with trend lines (Few)

*Stephen Few, Show Me the Numbers (Oakland, California: Analytics Press, 2004), pg. 86.*
Streamgraphs

- **Stacked time series**

<table>
<thead>
<tr>
<th>Idiom</th>
<th>Stacked Bar Charts</th>
</tr>
</thead>
<tbody>
<tr>
<td>What: Data</td>
<td>Multidimensional table: one quantitative value attribute, two categorical key attributes.</td>
</tr>
<tr>
<td>How: Encode</td>
<td>Bar glyph with length-coded subcomponents of value attribute for each category of secondary key attribute. Separate bars by category of primary key attribute.</td>
</tr>
<tr>
<td>Why: Task</td>
<td>Part-to-whole relationship, lookup values, find trends.</td>
</tr>
<tr>
<td>Scale</td>
<td>Key attribute (main axis): dozens to hundreds of levels. Key attribute (stacked glyph axis): several to one dozen</td>
</tr>
</tbody>
</table>

- Show shape of the data and part-whole relationships
- De-emphasise individual values

Figure 7.6. Streamgraph of music listening history. From [Byron and Wattenberg 08, Figure 0].
Line charts and dotplots

• Position to express value according to key

• Line charts use angle/shape to show trends

• Frequently time
### Line Chart idiom

- Line charts, dotplots
- Good for ordered data

<table>
<thead>
<tr>
<th>Idiom</th>
<th>Dotplot</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>What:</strong> Data</td>
<td>table:</td>
</tr>
<tr>
<td></td>
<td>1 quant value attrib,</td>
</tr>
<tr>
<td></td>
<td>1 ordered key attrib</td>
</tr>
<tr>
<td><strong>How:</strong> Encode</td>
<td>point marks</td>
</tr>
<tr>
<td></td>
<td>aligned vertical position to express value attrib,</td>
</tr>
<tr>
<td></td>
<td>separate/order by key attrib into horiz regions</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Idiom</th>
<th>line chart</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>What:</strong> Data</td>
<td>table:</td>
</tr>
<tr>
<td></td>
<td>1 quant value attrib,</td>
</tr>
<tr>
<td></td>
<td>1 ordered key attrib</td>
</tr>
<tr>
<td><strong>How:</strong> Encode</td>
<td>point marks, connection marks,</td>
</tr>
<tr>
<td></td>
<td>aligned vert position to express value attrib,</td>
</tr>
<tr>
<td></td>
<td>separate/order by key attrib into horiz regions</td>
</tr>
</tbody>
</table>
Mind the Gap - An Economic Chart Remake

Percentage of Employed Who Are Senior Managers, by Gender, 2008
Percentage of Employed Who are Senior Managers, by Gender, 2008

<table>
<thead>
<tr>
<th>Country</th>
<th>Women</th>
<th>Men</th>
</tr>
</thead>
<tbody>
<tr>
<td>United States</td>
<td></td>
<td></td>
</tr>
<tr>
<td>New Zealand</td>
<td></td>
<td></td>
</tr>
<tr>
<td>United Kingdom</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ireland</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Australia</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Estonia</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Belgium</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Greece</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Canada</td>
<td></td>
<td></td>
</tr>
<tr>
<td>France</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Iceland</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Finland</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OECD average</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hungary</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spain</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Israel</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Slovenia</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poland</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Czech Republic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Switzerland</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Austria</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Portugal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Norway</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Slovak Republic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Germany</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sweden</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Luxembourg</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Turkey</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Denmark</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mexico</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Korea</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Percentages for each country are shown in the chart for both women and men.
Percentage of Employed Who are Senior Managers, by Gender, 2008
(Percent) 

Mind the Gap - An Economic Chart Remake
What’s wrong?
The semantics of mark types

- Bar charts, line charts and dotplots all encode a quantitative value against a key attribute in a rectilinear layout.
- Often use additional encoding for other categories
- Lines also use connection marks to show inter-item relations
  - Only use for ordered data!
Which to use when

- Bars and bubbles emphasise comparison and association of individual values

- Lines (explicit and implied) emphasise trends
Lines and bars

Lines imply connections
• “the more male some the taller he is”

Use when there is some progression between the on the x-axis
• “12 year olds are taller 10 year olds”

Tufte’s Sparklines

- Give a hint of the trend, but don’t show the actual axes and scales.

- Good for dashboards and small spaces

peer2patent.org
Lines: Aspect ratio matters!

- Our ability to judge angles is more accurate at exact diagonals than at arbitrary direction.
  - We can judge distances “off” 45 or 90 degrees (43°) but cannot see the difference between 20 and 22 degrees.
- Multiscale banking to 45 degrees – algorithm to compute informative aspect ratios to maximise line segments close to the diagonal.

![Graph showing line segments close to the diagonal with x-axis values from 1700 to 1950 and y-axis values from -150 to 0.](image-url)
Matrix alignment

- Heatmap
- 2 keys, 1 value
- Good for dense encoding
- Re-ordering for clusters
Parallel layouts

- Parallel coordinates
- Many key attributes
- Different correlations
- Value vector is a line
Parallel coordinates

<table>
<thead>
<tr>
<th>Idiom</th>
<th>Parallel Coordinates</th>
</tr>
</thead>
<tbody>
<tr>
<td>What: Data</td>
<td>Table: many value attributes.</td>
</tr>
<tr>
<td>How: Encode</td>
<td>Parallel layout: horizontal spatial position used to separate axes, vertical spatial position used to express value along each aligned axis with connection line marks as segments between them.</td>
</tr>
<tr>
<td>Why: Tasks</td>
<td>Find trends, outliers, extremes, correlation.</td>
</tr>
<tr>
<td>Scale</td>
<td>Attributes: dozens along secondary axis. Items: hundreds.</td>
</tr>
</tbody>
</table>

13 items, 7 keys

(a)

16K items, 5 keys

(b)
What about Pies?
Radial layouts

Use polar coordinates
• 1 categorical key, 1 quantitative value
## radial idioms

<table>
<thead>
<tr>
<th>Idiom</th>
<th>Star plot</th>
</tr>
</thead>
<tbody>
<tr>
<td>What: data</td>
<td>Table: 1 quant value, 1 categorical attribute</td>
</tr>
<tr>
<td>How: Encode</td>
<td>length coding along point marks at 1D spatial position along axis + 1D spatial position for aligned axes</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Idiom</th>
<th>Pie chart</th>
</tr>
</thead>
<tbody>
<tr>
<td>What: data</td>
<td>Table: 1 quant value, 1 categorical attribute</td>
</tr>
<tr>
<td>How: Encode</td>
<td>area and angle</td>
</tr>
</tbody>
</table>
Percent Blue relative to Red?
Percent Blue relative to Red?
Few’s criteria for an effective visualization

• Clearly indicate the nature of the relationship
• Represent the quantities accurately
• Makes it easy to compare the quantities
• Makes it easy to see the ranked order of values
• Makes obvious how people should use the information
Total Deaths in American by Cause in 2007

- Heart disease
- Cancer
- Stroke (cerebrovascular diseases)
- Chronic lower respiratory diseases
- Accidents (unintentional injuries)
- Alzheimer’s disease
- Diabetes
- Influenza and Pneumonia
- Nephritis, nephrotic syndrome, and nephrosis
- Septicemia
- All other causes
Clearly indicate the nature of the relationship?

Total Deaths in American by Cause in 2007

- Heart disease
- Cancer
- Stroke (cerebrovascular diseases)
- Chronic lower respiratory diseases
- Accidents (unintentional injuries)
- Alzheimer’s disease
- Diabetes
- Influenza and Pneumonia
- Nephritis, nephrotic syndrome, and nephrosis
- Septicemia
- All other causes
Represents quantities accurately?

Total Deaths in American by Cause in 2007

- Heart disease
- Cancer
- Stroke (cerebrovascular diseases)
- Chronic lower respiratory diseases
- Accidents (unintentional injuries)
- Alzheimer’s disease
- Diabetes
- Influenza and Pneumonia
- Nephritis, nephrotic syndrome, and nephrosis
- Septicemia
- All other causes
Makes it easy to compare quantities?

Total Deaths in American by Cause in 2007

- Heart disease
- Cancer
- Stroke (cerebrovascular diseases)
- Chronic lower respiratory diseases
- Accidents (unintentional injuries)
- Alzheimer’s disease
- Diabetes
- Influenza and Pneumonia
- Nephritis, nephrotic syndrome, and nephrosis
- Septicemia
- All other causes
Makes it easy to see ranked values?

Total Deaths in America by Cause in 2007

- Heart disease
- Cancer
- Stroke (cerebrovascular disease)
- Chronic lower respiratory diseases
- Accidents (unintentional injuries)
- Alzheimer’s disease
- Diabetes
- Influenza and Pneumonia
- Nephritis, nephrotic syndrome, and nephrosis
- Septicemia
- All other causes
Makes it easy to see how people should use information?
A better way

- Heart disease: 25.42%
- Cancer: 23.22%
- Stroke (cerebrovascular diseases): 5.61%
- Chronic lower respiratory diseases: 5.28%
- Accidents (unintentional injuries): 5.10%
- Alzheimer's disease: 3.08%
- Diabetes: 2.95%
- Influenza and Pneumonia: 2.18%
- Nephritis, nephrotic syndrome, and nephrosis: 1.92%
- Septicemia: 1.44%
- All other causes: 23.81%
Percent Water

IAT 814 | Design Choices 1
Percent Water

Body

Brain

Blood
Bad
Better

The youth vote and everybody else

Republican vote share vs Age

- 2000
- 2004
- 2008
Even Better*

![Graph showing the percentage of youth voting Republican compared to 25, 45, and 65 year olds over the years 2000, 2004, and 2008. The graph indicates a decrease in the percentage of youth voting Republican over time.]
Too Little

Too Much

About Right

National Spending to Deal with Drug Addiction
Too Little

About Right

Too Much

Male
Female

National Spending to Deal with Drug Addiction
National Spending to Deal with Drug Addiction

- Too Little
- About Right
- Too Much

Female | Male
---|---
Too Little
About Right
Too Much

IAT 814 | Design Choices 1
National Spending to Deal with Drug Addiction

- Too Little
- About Right
- Too Much

Female

Male
1. Is the independent variable quantitative?
   - No: Qualitative
   - Yes: Quantitative

2a. Does it have one independent variable?
   - No: Composite Bar Graph, Grouped Bar Graph
   - Yes: Bar Graph

2b. Is the independent variable ordinal?
   - No: Scalar
   - Yes: Ordinal

3. Does it have one independent variable?
   - No: Multi-line Graph
   - Yes: Line Graph, Bar Graph

4. Do you want to see local trends?
   - No: 3D Scatter Plot, Contour Plot, Area Rendering
   - Yes: Multi-line Graph, Surface Plot

5. Do you want to see individual data points?
   - No: Histogram
   - Yes: Scatter Plot

http://chartchooser.juiceanalytics.com/