CSE 664
Visualization & Visual Analytics

The Views of Edward Tufte (and Some Others)

Klaus Mueller

Computer Science Department
Stony Brook University
<table>
<thead>
<tr>
<th>Lecture</th>
<th>Topic</th>
<th>Projects</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Intro, schedule, and logistics</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Applications of visual analytics, basic tasks, data types</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Introduction to D3, basic vis techniques for non-spatial data</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Data assimilation and preparation</td>
<td>Project #1 out</td>
</tr>
<tr>
<td>5</td>
<td>Data assimilation and preparation</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Bias in visualization</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Data reduction and dimension reduction</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Visual perception</td>
<td>Project #2(a) out</td>
</tr>
<tr>
<td>9</td>
<td>Visual cognition</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Visual design and aesthetics</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Cluster analysis: numerical data</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Cluster analysis: categorical data</td>
<td>Project #2(b) out</td>
</tr>
<tr>
<td>13</td>
<td>High-dimensional data visualization</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Dimensionality reduction and embedding methods</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Principles of interaction</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Midterm #1</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>Visual analytics</td>
<td>Final project proposal call out</td>
</tr>
<tr>
<td>18</td>
<td>The visual sense making process</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>Maps</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>Visualization of hierarchies</td>
<td>Final project proposal due</td>
</tr>
<tr>
<td>21</td>
<td>Visualization of time-varying and time-series data</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>Foundations of scientific and medical visualization</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>Volume rendering</td>
<td>Project 3 out</td>
</tr>
<tr>
<td>24</td>
<td>Scientific and medical visualization</td>
<td>Final Project preliminary report due</td>
</tr>
<tr>
<td>25</td>
<td>Visual analytics system design and evaluation</td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>Memorable visualization and embellishments</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>Infographics design</td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>Midterm #2</td>
<td></td>
</tr>
</tbody>
</table>
Seminal Books by Edward Tufte

Standard literature for every visualization enthusiast

Edward Tufte

Well recognized for his writings on information design

- a pioneer in the field of data visualization
- taught information design at Princeton University
- now a professor at Yale University

Popularized concept of “small multiples”

- aka trellis chart or panel chart
- similar charts of same scale + axes
- allows them to be easily compared
- use multiple views to show different partitions of a dataset
E. Muybridge’s Horses in Motion (1886)

- proofed for the first time that horses CAN have all 4 legs in the air
- work was also foundational to the development of the motion picture
FA Walker’s census charts (1870)

- population is broken down by state and then occupation, including a count of those attending school
- also has tree maps!
Cumulative projected deaths in each state

- Deaths
- Projected Deaths
- Previous Estimates (Mean Projection)

- Alabama: High: 670, Low: 600
- Alaska: 43
- Arizona: 2k
- Arkansas: 360
- California: 2k, 1k
- Colorado: 1k
- Connecticut: 6k
- D.C.: 530
- Delaware: 330
- Florida: 3k
- Georgia: 6k
- Hawaii: 53
- Idaho: 110
- Illinois: 3k
- Indiana: 2k
- Iowa: 1k

Bloomberg page
# Small Multiples

## Tracking Covid-19

<table>
<thead>
<tr>
<th>State</th>
<th>Cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maryland</td>
<td>2k</td>
</tr>
<tr>
<td>Massachusetts</td>
<td>9k</td>
</tr>
<tr>
<td>Michigan</td>
<td>5k</td>
</tr>
<tr>
<td>Minnesota</td>
<td>900</td>
</tr>
<tr>
<td>Mississippi</td>
<td>1k</td>
</tr>
<tr>
<td>Missouri</td>
<td>920</td>
</tr>
<tr>
<td>Montana</td>
<td>41</td>
</tr>
<tr>
<td>Nebraska</td>
<td>1k</td>
</tr>
<tr>
<td>Nevada</td>
<td>470</td>
</tr>
<tr>
<td>New Hampshire</td>
<td>160</td>
</tr>
<tr>
<td>New Jersey</td>
<td>12k</td>
</tr>
<tr>
<td>New Mexico</td>
<td>300</td>
</tr>
<tr>
<td>New York</td>
<td>32k</td>
</tr>
<tr>
<td>North Carolina</td>
<td>610</td>
</tr>
<tr>
<td>North Dakota</td>
<td>1k</td>
</tr>
<tr>
<td>Ohio</td>
<td>1k</td>
</tr>
<tr>
<td>Oklahoma</td>
<td>770</td>
</tr>
<tr>
<td>Oregon</td>
<td>240</td>
</tr>
<tr>
<td>Pennsylvania</td>
<td>5k</td>
</tr>
<tr>
<td>Puerto Rico</td>
<td>110</td>
</tr>
</tbody>
</table>
Edward Tufte

Also popularized “sparklines”
  • small integrative visualizations

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Bid</th>
<th>Ask</th>
<th>Last</th>
<th>Change</th>
<th>High</th>
<th>Low</th>
<th>Value Change</th>
<th>Value</th>
<th>Gain</th>
</tr>
</thead>
<tbody>
<tr>
<td>DCLL</td>
<td>09.54</td>
<td>09.31</td>
<td>09.34</td>
<td>+1.14</td>
<td>10,310,100</td>
<td>90.10</td>
<td>89.92</td>
<td>17,950</td>
<td>13,147</td>
</tr>
<tr>
<td>CPG</td>
<td>25.87</td>
<td>25.86</td>
<td>25.79</td>
<td>-0.18</td>
<td>25,827,700</td>
<td>51.14</td>
<td>50.91</td>
<td>4,844</td>
<td>1,831</td>
</tr>
<tr>
<td>SOTI</td>
<td>26.14</td>
<td>25.36</td>
<td>25.38</td>
<td>+0.12</td>
<td>504,600</td>
<td>27.36</td>
<td>27.30</td>
<td>13,188</td>
<td>17,115</td>
</tr>
<tr>
<td>COMS</td>
<td>46.12</td>
<td>46.16</td>
<td>45.94</td>
<td>-0.24</td>
<td>3,131,100</td>
<td>47.15</td>
<td>47.00</td>
<td>6,503</td>
<td>2,793</td>
</tr>
<tr>
<td>LU</td>
<td>111.50</td>
<td>111.11</td>
<td>111.91</td>
<td>+1.51</td>
<td>5,104,600</td>
<td>112.56</td>
<td>112.00</td>
<td>5,570</td>
<td>2,273.9</td>
</tr>
<tr>
<td>YHOO</td>
<td>368.17</td>
<td>368.17</td>
<td>368.17</td>
<td>+0.17</td>
<td>3,757,800</td>
<td>381.15</td>
<td>381.10</td>
<td>9,213</td>
<td>-0.41</td>
</tr>
<tr>
<td>AOK</td>
<td>162.38</td>
<td>162.16</td>
<td>163</td>
<td>+8</td>
<td>10,039,800</td>
<td>164</td>
<td>163.16</td>
<td>5,705</td>
<td>73.08</td>
</tr>
<tr>
<td>CMGI</td>
<td>57.36</td>
<td>57.12</td>
<td>57.12</td>
<td>+5.78</td>
<td>1,333,800</td>
<td>59.12</td>
<td>59.00</td>
<td>9,100</td>
<td>87.60</td>
</tr>
<tr>
<td>SPLN</td>
<td>333.33</td>
<td>333.15</td>
<td>333.16</td>
<td>+7.16</td>
<td>300,200</td>
<td>34</td>
<td>34.00</td>
<td>6,763</td>
<td>64.60</td>
</tr>
<tr>
<td>BSAS</td>
<td>13.12</td>
<td>13.58</td>
<td>13.58</td>
<td>-0.16</td>
<td>359,200</td>
<td>14</td>
<td>13.16</td>
<td>1,363</td>
<td>-0.17</td>
</tr>
<tr>
<td>GNET</td>
<td>103.36</td>
<td>103.36</td>
<td>103.36</td>
<td>+5.18</td>
<td>307,600</td>
<td>97</td>
<td>96.43</td>
<td>10,131</td>
<td>130.26</td>
</tr>
<tr>
<td>NMK</td>
<td>67.67</td>
<td>67.67</td>
<td>67.67</td>
<td>+2.34</td>
<td>1,233,900</td>
<td>69</td>
<td>69.15</td>
<td>6,700</td>
<td>78.87</td>
</tr>
<tr>
<td>MSFT</td>
<td>173.17</td>
<td>173.17</td>
<td>173.56</td>
<td>+1.34</td>
<td>13,284,500</td>
<td>174.76</td>
<td>174.69</td>
<td>17,331</td>
<td>54.74</td>
</tr>
<tr>
<td>INTC</td>
<td>133.34</td>
<td>133.13</td>
<td>133.13</td>
<td>-0.14</td>
<td>6,994,300</td>
<td>137.16</td>
<td>137.12</td>
<td>20,763</td>
<td>65.20</td>
</tr>
<tr>
<td>TOTAL</td>
<td>205,302</td>
<td>80,993</td>
<td>80,993</td>
<td>2,293</td>
<td>143,280</td>
<td>79.41</td>
<td>79.41</td>
<td>83,377</td>
<td></td>
</tr>
</tbody>
</table>

Sparklines inspired “word size visualizations”
  • charts or graphs tightly integrated into text or even computer code

Although Tufte is said to have invented **sparklines**, in actuality he invented only the name and popularized it as technique.[15] Sparklines are a condensed way to present trends and variation, associated with a measurement such as average temperature or stock market activity, often embedded directly in the text; for example: The Dow Jones index for February 7, 2006 ... [16][17] These are often used as elements of a **small multiple** with several lines used together. Tufte explains the sparkline as a kind of “word” that conveys rich information without breaking the flow of a sentence or paragraph made of other ”words” both visual and conventional. To date, the earliest known implementation of sparklines was done by interaction designer Peter Zelchenko and programmer Mike Medved in early 1998.[18]
According to Tufte (pg. 51):

- Graphical excellence is the well-designed presentation of interesting data
  - a matter of **substance**, **statistics**, and **design**

- Graphical excellence consists of complex ideas communicated with:
  - **clarity**, **precision**, and **efficiency**

- Graphical excellence is that what gives the viewer:
  - the **greatest number of ideas**
  - in the **shortest time**
  - with the **least ink**
  - in the **smallest space**

- Graphical excellence is nearly always multivariate

- Graphical excellence requires telling the truth about the data

(Nevertheless, visualizations should be visually pleasing and may very well have an artistic touch)
Tufte’s views on

- visual embellishments → “chart junk”
- abuse of physically-motivated distortions → “lie factor”
Avoid Misleading Embellishments = Chart Junk
Nigel Holmes’ Famous Chart

MONSTROUS COSTS
Total House and Senate campaign expenditures, in millions

1972 '74 '76 '80 '82 est.
Avoid Misleading Use of Graphics Effects

This line, representing 18 miles per gallon in 1978, is 0.6 inches long.

Fuel Economy Standards for Autos
Set by Congress and supplemented by the Transportation Department. In miles per gallon.

This line, representing 27.5 miles per gallon in 1985, is 5.3 inches long.

real effect: \( \frac{27.5 - 18}{18} = 53\% \)

graphical effect: \( \frac{5.3'' - 0.6''}{0.6''} = 783\% \) \( \Rightarrow \) lie factor: \( \frac{783}{53} = 14.8 \)
Tell the Truth About the Data

REQUIRED FUEL ECONOMY STANDARDS:
NEW CARS BUILT FROM 1978 TO 1985

19.1 mpg, expected average for all cars on road, 1985

13.7 mpg, average for all cars on road, 1978

If You Must Embellish…

REQUIRED FUEL ECONOMY STANDARDS:
NEW CARS BUILT FROM 1978 TO 1985

19.1 mpg, expected average for all cars on road, 1985

13.7 mpg, average for all cars on road, 1978

Avoid Suggestive Distortions

In the Barrel...

Price per bbl. of light crude, leaving Saudi Arabia on Jan. 1

April 1
$14.55

1979
1978
1977
76
75
74
$0.95
$2.41
$10.46
$11.51
$12.09
$12.70
$13.34

Lie factor: 9.4 (2D), 59.4 (3D)
Show the Data in Their Proper Context
Avoid Display of Out-of-Context Data

**SOLAR RADIATION AND STOCK PRICES**

Graphical Integrity

- Quoting data out of context and/or too sparse (recall: graphics allows high data density)
  - example: Connecticut traffic deaths (pg. 74/75)

Principles that ensure graphical integrity:

- The representation of numbers should be directly proportional to the numerical quantities represented (see the growing barrels)
- Clear and detailed labeling should be used to defeat graphical distortion and ambiguity
- Show data variations and not design variations (see the fuel economy graph)
- In time-series displays of money, show deflated and standardized units
- The number of information carrying dimensions should not exceed the data dimensions (see the growing barrels, the shrinking doctor)
- Graphics must not quote data out of context (see the Connecticut traffic deaths)
- Convincing graphics must demonstrate cause and effect (see Challenger disaster)
Do these bare graphs engage a human audience?
  • are they memorable?

A recent (research) trend
  • will embellishment help memorability, engagement?
  • do we need what Tufte calls “chart junk”
Experiment set up as a game on Amazon Mechanical Turk

• workers were presented with a sequence of images (about 120)
• presented for 1 second, with a 1.4 second gap between consecutive images
• workers had to press a key if they saw an image for the second time in the sequence (spacing 1-7 images with “filler” images in between)
Memorability Experiment by Borkin et al.

most memorable

most memorable
after removing
human recognizable
cartoons

least memorable

Borkin et al. IEEE TVCG 2014
What Do People Remember?

**Experiment Design**

**Labeled Visualization Database**
- 393 visualizations
- Visualizations are taken from [8], and the label taxonomy described in Table 1 is applied.

**Encoding**
- 100 “target” visualizations
- 10 seconds / image
- **Output:** Eye-tracking fixation locations and durations.

**Recognition**
- Same 100 targets + 100 “fillers”
- 2 seconds / image
- **Output:** Eye-tracking fixation locations and durations, and whether visualization is recognized.

**Recall**
- Correctly recognized blurred targets
- 20 min - as many images as participant can complete
- **Output:** Text descriptions of what participant recalls about the visualization.

---

Borkin et al. IEEE TVCG 2016
Fig. 7. Examples of the most and least recognizable visualizations from [8]. **TOP:** Eye-tracking fixation heat maps (i.e., average of all participants’ fixation locations) from the *encoding* phase of the experiment in which each visualization was presented for 10 seconds. The fixation patterns demonstrate visual exploration of the visualization. **BOTTOM:** Eye-tracking fixation heat maps from the *recognition* phase of the experiment in which each visualization was presented for 2 seconds or until response. The most recognizable visualizations all have a single focus in the center indicating quick recognition of the visualization, whereas the least recognizable visualizations have fixation patterns similar to the encoding fixations indicative of visual exploration (e.g., title, text, etc.) for recognition.

Borkin et al. IEEE TVCG 2016
Takeaways:

• 393 visualizations and eye movements of 33 participants and 1,000s of participant-generated text descriptions of the visualizations
• titles and supporting text should convey the message of a visualization
• if used appropriately, pictograms do not interfere with understanding and can improve recognition
• redundancy helps effectively communicate the message
• visualizations that are memorable “at-a-glance” are also capable of effectively conveying the message of the visualization

→ thus, a memorable visualization is often also an effective one
Important are:
- attributes like color
- inclusion of a human recognizable object

However, link to human engagement not explicitly established
- “just” memorability

Our own studies show that embellishments can get humans interested in studying an image
- but prefer conventional charts for problem solving
Infographic

Graphic visual representations of information, data or knowledge intended to present information quickly and clearly

Evolved in recent years to be for mass communication
- designed with fewer assumptions about the readers knowledge base than other types of visualizations
- but can be misleading and express the opinion of the author
Using Icons as Bar Graphs

Wang et al. CHI 18
Data-Driven Design Guides

Kim et al. TVCG 17
video and more is here