Information Visualisation

Course Notes

Version of 16 Mar 2022

Ao.Univ.-Prof. Dr. Keith Andrews

ISDS
Graz University of Technology
Inffeldgasse 16c
A-8010 Graz
kandrews@iicm.edu

https://courses.isds.tugraz.at/ivis/ivis.pdf

Copyright 2022 by Keith Andrews, except as otherwise noted.
# Contents

**Contents**

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>List of Figures</td>
<td>iii</td>
</tr>
<tr>
<td>List of Tables</td>
<td>vii</td>
</tr>
<tr>
<td>Preface</td>
<td>ix</td>
</tr>
<tr>
<td>Credits</td>
<td>xi</td>
</tr>
</tbody>
</table>

## 1 Introduction

1.1 References ................................................. 2
1.2 Why Visualise? ............................................. 7
1.3 Visual Lies ................................................ 11
1.4 General Concepts of Information Visualisation ............. 14
1.5 Types of Information ................................. 18

## 2 Visual Perception

2.1 Human Vision ............................................. 22
2.2 Visual Input Takes Priority .............................. 23
2.3 Preattentive Processing ................................. 23
2.4 Visual Encoding ......................................... 31

## 3 History of Information Visualisation

3.1 Diderot and d’Alembert ................................ 33
3.2 Joseph Priestley .......................................... 34
3.3 William Playfair ......................................... 35
3.4 Edmund Cooper and John Snow .......................... 37
3.5 Florence Nightingale .................................... 38
3.6 Charles Minard ........................................... 41
3.7 Fletcher Hewes and Henry Gannett ........................ 42
3.8 Marey and Iby (1878) ................................... 43
3.9 Paul Otlet .................................................. 44
3.10 Jaques Bertin ............................................ 44

## 4 Visualising Linear Structures

4.1 Bifocal Display .......................................... 47
4.2 Perspective Wall .......................................... 48
4.3 Seesoft ..................................................... 49
4.4 Lifestreams .............................................................. 50
4.5 Spiral of Ranked Items ............................................. 51

5 Visualising Hierarchies .............................................. 53
  5.1 Outliners ............................................................. 54
  5.2 Layered Node-Link Tree Browsers .............................. 56
  5.3 Radial Node-Link Tree Browsers ................................. 59
  5.4 Layered Space-Filling Tree Browsers ......................... 65
  5.5 Radial Space-Filling Tree Browsers ......................... 66
  5.6 Inclusive Space-Filling Tree Browsers ....................... 68
  5.7 Overlapping Space-Filling Tree Browsers ................. 74
  5.8 Single-Level Tree Browsers .................................. 76

6 Visualising Networks and Graphs ................................ 79
  6.1 Adjacency Matrix .................................................. 79
  6.2 Predetermined Position ......................................... 80
  6.3 Layered Graph Drawing .......................................... 84
  6.4 Force-Based Layouts ............................................. 85

7 Visualising Multidimensional Metadata ......................... 87
  7.1 Interactive Tables ................................................ 87
  7.2 Interactive Scatterplots ......................................... 88
  7.3 Multidimensional Glyphs ....................................... 90
  7.4 Parallel Coordinates ............................................ 92
  7.5 Star Plot ........................................................... 94
  7.6 Interactive Attraction ........................................... 95
  7.7 Interactive Histograms .......................................... 95
  7.8 Circular Histograms ............................................. 97

8 Visualising Text and Object Collections (Feature Spaces) .... 99
  8.1 Distance-Based Projection ...................................... 99
  8.2 Force-Directed Placement (FDP) ............................. 102
  8.3 Example Systems ................................................ 102
  8.4 Self-Organizing Maps (SOM) ................................. 106

9 Other Kinds of Visualisation .................................... 109
  9.1 Visualising Query Spaces ...................................... 109
  9.2 Internal Document Visualisation (Content Analysis) .... 111

Bibliography .......................................................... 113
## List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Line Chart of Sales Data</td>
<td>8</td>
</tr>
<tr>
<td>1.2</td>
<td>Attentive Processing</td>
<td>8</td>
</tr>
<tr>
<td>1.3</td>
<td>Pre-Attentive Processing</td>
<td>9</td>
</tr>
<tr>
<td>1.4</td>
<td>Anscombe’s Quartet Plots</td>
<td>10</td>
</tr>
<tr>
<td>1.5</td>
<td>Steve Jobs’ 3D Pie Chart</td>
<td>12</td>
</tr>
<tr>
<td>1.6</td>
<td>Reconstruction of Steve Jobs’ 3D Pie Chart</td>
<td>13</td>
</tr>
<tr>
<td>1.7</td>
<td>2D Pie Chart of Steve Jobs’ Data</td>
<td>13</td>
</tr>
<tr>
<td>1.8</td>
<td>Bar Chart of Steve Jobs’ Data</td>
<td>14</td>
</tr>
<tr>
<td>1.9</td>
<td>Visualisation Pipeline</td>
<td>15</td>
</tr>
<tr>
<td>1.10</td>
<td>Straight-Line Connections</td>
<td>16</td>
</tr>
<tr>
<td>1.11</td>
<td>Edge Bundling</td>
<td>17</td>
</tr>
<tr>
<td>1.12</td>
<td>Excentric Labels in JExplorer</td>
<td>17</td>
</tr>
<tr>
<td>1.13</td>
<td>Scatterplot Matrix</td>
<td>18</td>
</tr>
<tr>
<td>2.1</td>
<td>Colour Hue</td>
<td>24</td>
</tr>
<tr>
<td>2.2</td>
<td>Colour Intensity (Grey)</td>
<td>24</td>
</tr>
<tr>
<td>2.3</td>
<td>Colour Intensity</td>
<td>25</td>
</tr>
<tr>
<td>2.4</td>
<td>Line Length</td>
<td>25</td>
</tr>
<tr>
<td>2.5</td>
<td>Line Width</td>
<td>25</td>
</tr>
<tr>
<td>2.6</td>
<td>2D Spatial Position</td>
<td>26</td>
</tr>
<tr>
<td>2.7</td>
<td>Orientation</td>
<td>26</td>
</tr>
<tr>
<td>2.8</td>
<td>Size</td>
<td>26</td>
</tr>
<tr>
<td>2.9</td>
<td>Shape</td>
<td>27</td>
</tr>
<tr>
<td>2.10</td>
<td>Added Mark</td>
<td>27</td>
</tr>
<tr>
<td>2.11</td>
<td>Enclosure</td>
<td>27</td>
</tr>
<tr>
<td>2.12</td>
<td>Lack of Enclosure</td>
<td>28</td>
</tr>
<tr>
<td>2.13</td>
<td>Curvature</td>
<td>28</td>
</tr>
<tr>
<td>2.14</td>
<td>Focus and Blur</td>
<td>28</td>
</tr>
<tr>
<td>2.15</td>
<td>3D Depth</td>
<td>29</td>
</tr>
<tr>
<td>2.18</td>
<td>Direction of Motion</td>
<td>30</td>
</tr>
<tr>
<td>2.19</td>
<td>Conjunction of Shape and Depth</td>
<td>30</td>
</tr>
<tr>
<td>2.20</td>
<td>Conjunction of Blur and Colour Hue</td>
<td>31</td>
</tr>
<tr>
<td>3.1</td>
<td>Tree of Diderot and d’Alembert</td>
<td>34</td>
</tr>
<tr>
<td>3.2</td>
<td>Priestley’s New Chart of History (1769)</td>
<td>35</td>
</tr>
<tr>
<td>3.3</td>
<td>Playfair’s Area Chart</td>
<td>36</td>
</tr>
<tr>
<td>3.4</td>
<td>Playfair’s Pie Chart</td>
<td>36</td>
</tr>
<tr>
<td>Section</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
<td>------</td>
</tr>
<tr>
<td>3.5</td>
<td>Snow’s London Cholera Map</td>
<td>38</td>
</tr>
<tr>
<td>3.6</td>
<td>Nightingale’s Bat’s Wing Diagram</td>
<td>39</td>
</tr>
<tr>
<td>3.7</td>
<td>Modern Bat’s Wing Diagram</td>
<td>39</td>
</tr>
<tr>
<td>3.8</td>
<td>Nightingale’s Wedge Diagram</td>
<td>40</td>
</tr>
<tr>
<td>3.9</td>
<td>Modern Wedge Diagram</td>
<td>41</td>
</tr>
<tr>
<td>3.10</td>
<td>Minard’s Map</td>
<td>42</td>
</tr>
<tr>
<td>3.11</td>
<td>Hewes and Gannett Parallel Coordinates (1883)</td>
<td>43</td>
</tr>
<tr>
<td>3.12</td>
<td>Marey and Iby Train Schedule</td>
<td>44</td>
</tr>
<tr>
<td>3.13</td>
<td>Bertin’s Reorderable Matrix</td>
<td>45</td>
</tr>
<tr>
<td>4.1</td>
<td>Bifocal Calendar</td>
<td>48</td>
</tr>
<tr>
<td>4.2</td>
<td>Perspective Wall</td>
<td>49</td>
</tr>
<tr>
<td>4.3</td>
<td>Seesoft source code visualisation</td>
<td>50</td>
</tr>
<tr>
<td>4.4</td>
<td>Lifestreams orders streams of item chronologically</td>
<td>51</td>
</tr>
<tr>
<td>4.5</td>
<td>GopherVR Spiral of Search Results</td>
<td>52</td>
</tr>
<tr>
<td>4.6</td>
<td>JUCS Spiral of Authors</td>
<td>52</td>
</tr>
<tr>
<td>5.1</td>
<td>The Java JTree tree view</td>
<td>54</td>
</tr>
<tr>
<td>5.2</td>
<td>MagTree</td>
<td>55</td>
</tr>
<tr>
<td>5.3</td>
<td>WebTOC</td>
<td>56</td>
</tr>
<tr>
<td>5.4</td>
<td>The Walker Tree Browser</td>
<td>57</td>
</tr>
<tr>
<td>5.5</td>
<td>The FSN file system navigator</td>
<td>57</td>
</tr>
<tr>
<td>5.6</td>
<td>The Harmony Information Landscape</td>
<td>58</td>
</tr>
<tr>
<td>5.7</td>
<td>Radial Tree</td>
<td>59</td>
</tr>
<tr>
<td>5.8</td>
<td>Original Hyperbolic Browser from Xerox</td>
<td>60</td>
</tr>
<tr>
<td>5.9</td>
<td>D3 Hypertree for Product Taxonomy</td>
<td>61</td>
</tr>
<tr>
<td>5.10</td>
<td>Hypertree of Life</td>
<td>61</td>
</tr>
<tr>
<td>5.11</td>
<td>H3</td>
<td>62</td>
</tr>
<tr>
<td>5.12</td>
<td>Walrus</td>
<td>63</td>
</tr>
<tr>
<td>5.13</td>
<td>SInVis</td>
<td>63</td>
</tr>
<tr>
<td>5.14</td>
<td>Cone Tree</td>
<td>64</td>
</tr>
<tr>
<td>5.15</td>
<td>Botanical visualisation</td>
<td>65</td>
</tr>
<tr>
<td>5.16</td>
<td>Xdu visualisation of JDK 1.2 distribution</td>
<td>66</td>
</tr>
<tr>
<td>5.17</td>
<td>Information slices</td>
<td>67</td>
</tr>
<tr>
<td>5.18</td>
<td>SunBurst</td>
<td>68</td>
</tr>
<tr>
<td>5.19</td>
<td>A tree map of the Dewey Decimal classification hierarchy</td>
<td>69</td>
</tr>
<tr>
<td>5.20</td>
<td>Market map of stock market</td>
<td>70</td>
</tr>
<tr>
<td>5.21</td>
<td>SequoiaView</td>
<td>70</td>
</tr>
<tr>
<td>5.22</td>
<td>Information pyramids</td>
<td>71</td>
</tr>
<tr>
<td>5.23</td>
<td>Information pyramids with treeview</td>
<td>72</td>
</tr>
<tr>
<td>5.24</td>
<td>InfoSky Cobweb</td>
<td>72</td>
</tr>
<tr>
<td>5.25</td>
<td>InfoSky Cobweb Zoomed</td>
<td>73</td>
</tr>
<tr>
<td>5.26</td>
<td>Voronoi Treemap</td>
<td>74</td>
</tr>
<tr>
<td>5.27</td>
<td>Cheops</td>
<td>75</td>
</tr>
<tr>
<td>5.28</td>
<td>BeamTree</td>
<td>75</td>
</tr>
<tr>
<td>5.29</td>
<td>BeamTree3D</td>
<td>76</td>
</tr>
<tr>
<td>5.30</td>
<td>GopherVR visualising one level of a Gopher tree</td>
<td>77</td>
</tr>
<tr>
<td>Section</td>
<td>Title</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>6.1</td>
<td>Example Adjacency Matrix</td>
<td>80</td>
</tr>
<tr>
<td>6.2</td>
<td>ThreadVis for Message Threads</td>
<td>80</td>
</tr>
<tr>
<td>6.3</td>
<td>Circular Plot of Migration Data</td>
<td>81</td>
</tr>
<tr>
<td>6.4</td>
<td>Flare Dependency Graph</td>
<td>82</td>
</tr>
<tr>
<td>6.5</td>
<td>Intermedia Global Map</td>
<td>83</td>
</tr>
<tr>
<td>6.6</td>
<td>Migration to California Flow Map</td>
<td>83</td>
</tr>
<tr>
<td>6.7</td>
<td>Local map of link connections around “grep” manual page</td>
<td>84</td>
</tr>
<tr>
<td>6.8</td>
<td>The Harmony Local Map 3D</td>
<td>85</td>
</tr>
<tr>
<td>6.9</td>
<td>SemNet</td>
<td>86</td>
</tr>
<tr>
<td>6.10</td>
<td>HyperSpace</td>
<td>86</td>
</tr>
<tr>
<td>7.1</td>
<td>Table Lens</td>
<td>88</td>
</tr>
<tr>
<td>7.2</td>
<td>The FilmFinder starfield display</td>
<td>89</td>
</tr>
<tr>
<td>7.3</td>
<td>Envision maps document attributes along two axes</td>
<td>89</td>
</tr>
<tr>
<td>7.4</td>
<td>Search Result Explorer</td>
<td>90</td>
</tr>
<tr>
<td>7.5</td>
<td>Chernoff Faces for Cars</td>
<td>92</td>
</tr>
<tr>
<td>7.6</td>
<td>Parallel Coordinates</td>
<td>93</td>
</tr>
<tr>
<td>7.7</td>
<td>InfoScope</td>
<td>93</td>
</tr>
<tr>
<td>7.8</td>
<td>Star Plot of Cereals</td>
<td>94</td>
</tr>
<tr>
<td>7.9</td>
<td>Dust and Magnet</td>
<td>95</td>
</tr>
<tr>
<td>7.10</td>
<td>The Attribute Explorer</td>
<td>96</td>
</tr>
<tr>
<td>7.11</td>
<td>JavaScript Attribute Explorer</td>
<td>97</td>
</tr>
<tr>
<td>7.12</td>
<td>Daisy</td>
<td>98</td>
</tr>
<tr>
<td>8.1</td>
<td>Distance-Based Projection Pipeline</td>
<td>100</td>
</tr>
<tr>
<td>8.2</td>
<td>Bead</td>
<td>103</td>
</tr>
<tr>
<td>8.3</td>
<td>SPIRE Galaxy and Themescape</td>
<td>103</td>
</tr>
<tr>
<td>8.4</td>
<td>VxInsight</td>
<td>104</td>
</tr>
<tr>
<td>8.5</td>
<td>VisIslands</td>
<td>105</td>
</tr>
<tr>
<td>8.6</td>
<td>InfoSky</td>
<td>106</td>
</tr>
<tr>
<td>8.7</td>
<td>SOMLib</td>
<td>107</td>
</tr>
<tr>
<td>8.8</td>
<td>WEBSOM</td>
<td>108</td>
</tr>
<tr>
<td>9.1</td>
<td>InfoCrystal Idea</td>
<td>110</td>
</tr>
<tr>
<td>9.2</td>
<td>InfoCrystal with Four Query Terms</td>
<td>110</td>
</tr>
<tr>
<td>9.3</td>
<td>MetaCrystal</td>
<td>111</td>
</tr>
<tr>
<td>9.4</td>
<td>Tile Bars</td>
<td>112</td>
</tr>
</tbody>
</table>
List of Tables

1.1 Table of Sales Data ......................................................... 7
1.2 Anscombe’s Quartet in Tabular Form ........................................ 9
1.3 Steve Jobs’ Pie Chart Data in Tabular Form ............................... 12

7.1 mtcars Dataset ............................................................... 91
7.2 Mapping to Facial Characteristics ....................................... 91
Preface


These lecture notes have evolved over many years of giving talks and teaching short courses on various aspects of information visualisation at conferences and for industry. I taught the first dedicated course on Information Visualisation at Graz University of Technology in summer semester 2005 and have also taught a version of the course at FH Joanneum in Graz, and at various conferences.

I would like to thank my students and colleagues past and present for their many suggestions and corrections, which have helped to massage these notes into their current form.

Thanks and happy reading.

Keith Andrews
Graz, Austria, 16 Mar 2022
Credits

I would like to thank the following individuals and organisations for permission to use their material:

• Figure 4.3 is used with kind permission of Steve Eick, Visual Insights.
• Figure 4.5 is used with kind permission of Tom Erickson.
• Figure 5.12 is used with kind permission of Young Hyun, CAIDA.
• Figure 5.18 is used with kind permission of John Stasko, Georgia Tech.
• Figure 5.13 is used with kind permission of Matthias Kreuseler, University of Rostock.
• Figure 5.15 is used with kind permission of Jack van Wijk, Eindhoven University of Technology.
• Figure 6.9 was scanned from an original photograph with kind permission of Kim Fairchild.
• Figure 7.2 is used under the terms of the Copyright Notice of the University of Maryland (see page xi).
• Figure 8.3 is used with kind permission of Pacific Northwest Naional Laboratory.

The following figures are used under the terms of the ACM Copyright Notice (see page xi):

• Figure 4.2, extracted from CHI ’91. Human factors in computing systems conference proceedings on Reaching through technology, pages 173-176.
• Figure 4.4, extracted from CHI 96 Electronic Proceedings, Videos.
• Figure 7.10, extracted from CHI 94 Electronic Proceedings, Videos.
• Figure 7.3, extracted from CHI 97 Electronic Proceedings, Demos.
• Figure 8.2, from the SIGIR ’92 Proceedings.

University of Maryland Copyright Notice

Copyright © University of Maryland 1984-1994
All works herein are Copyright (c) University of Maryland 1984-1994, all rights reserved. We allow fair use of our information provided any and all copyright marks, trade marks, and author attribution are retained.

ACM Copyright Notice

Copyright © by the Association for Computing Machinery, Inc.
Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit
is permitted. To copy otherwise, to republish, to post on servers, or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Publications Dept, ACM Inc., fax +1 (212) 869-0481, or permissions@acm.org.

For further information, see the ACM Copyright Policy.
Chapter 1

Introduction

“Let my dataset change your mindset.”

[ Hans Rosling, title of talk at TED@State, 03 Jun 2009 [Rosling 2009]. ]

Information visualisation (InfoVis) is the visual presentation of abstract information spaces and structures, together with accompanying interactions, so as to facilitate their rapid assimilation and understanding.

Visualisation

The broader field of visualisation has three main sub-fields:

• SciVis: Scientific Visualisation (SciVis) typically involves concrete (3d) objects, for example a medical scan of part of the body, or a simulation of air flow around an aircraft wing. SciVis visualisations often depict flows, volumes, and surfaces in (3d) space.

• GeoVis: Geographic Visualisation (GeoVis) is map-based. The data typically has inherent 2d or 3d spatial coordinates, and is generally shown in relation to a map.

• InfoVis: Information Visualisation (InfoVis) deals with abstract information structures, such as hierarchies, networks, or multidimensional spaces.

Data Visualisation (DataVis) = InfoVis + GeoVis.

Visual Analytics = DataVis (frontend) + Analytics (backend).

Information Visualisation

InfoVis ≠ SciVis or GeoVis:

• In SciVis, the visual representation (geometry) is generally given, suggested by objects in the data. [concrete objects, geometry is given]

• In GeoVis, the visual representation (geometry) is generally given, determined by spatial coordinates within the data. [spatial coordinates, geometry is given]

• In InfoVis, an appropriate visual representation must be (carefully) designed or “invented”. [abstract structures, geometry is chosen]
Visual Representation + Interaction

- The visual representation is only half the story.
- Interaction (navigational and manipulation) facilities are the other half.

Two Main Purposes

Interactive data visualisations are used for two main purposes:

- **Analysis**: Exploratory visualisations help researchers to explore and analyse unfamiliar datasets.
- **Presentation**: Explanatory visualisations present results and insights to a wider audience.

For numerous examples of each kind, see my talk at TEDxGraz 2015 [Andrews 2015].

1.1 References

Books (InfoVis)

++ Tamara Munzner; *Visualization Analysis and Design*; CRC Press, 26 Nov 2014. ISBN 1466508914 (com, uk) [Munzner 2014] [4 physical copies and online full text via TUB.]
++ Tominski and Schuhmann; *Interactive Visual Data Analysis*; CRC Press, 30 Apr 2020. ISBN 0367898756 (com, uk) [Tominski and Schuhmann 2020]
++ Colin Ware; *Information Visualization: Perception for Design*; 4th Edition, Morgan Kaufmann, 11 Mar 2020. ISBN 0128128755 (com, uk) [Ware 2020] [5 physical copies in TUB.]
+ Alberto Cairo; *The Functional Art*; New Riders, 2012. ISBN 0321834739 (com, uk) [Cairo 2012] [Online full text via TUB.]
+ Alberto Cairo; *The Truthful Art*; New Riders, 2016. ISBN 0321934075 (com, uk) [Cairo 2016] [Online full text via TUB.]

Books (Practical Guides)


+ Cole Nussbaumer Knaflic; *Storytelling with Data*; Wiley, 02 Nov 2015. ISBN 1119002257 (com, uk) [Nussbaumer Knaflic 2015]


**Books (Applications)**

+ Julie Steele and Noah Iliinsky; *Beautiful Visualization: Looking at Data Through the Eyes of Experts*; O’Reilly, 21 Jun 2010. ISBN 1449379869 (com, uk) [Steele and Iliinsky 2010]


+ Martin Dodge and Rob Kitchin; *Mapping Cyberspace*; Routledge, 2000. ISBN 0415198844 (com, uk) [Dodge and Kitchin 2000]


+ Chaomei Chen; *Information Visualisation and Virtual Environments*; Springer, 1999. ISBN 1852331364 (com, uk) [Chen 1999]


+ Fayyad et al; *Information Visualization in Data Mining and Knowledge Discovery*; Morgan Kaufmann, 2001. ISBN 1558606890 (com, uk) [Usama Fayyad 2001]


+ Alberto del Bimbo; *Visual Information Retrieval*; Morgan Kaufmann, 1999. ISBN 1558606246 (com, uk) [del Bimbo 1999]
CHAPTER 1. INTRODUCTION


**Podcasts**

++ Enrico Bertini and Moritz Stefaner; *Data Stories*; Podcast about data visualisation. datastori.es
+ Cole Nussbaumer Knaflic; *Storytelling with Data Podcast*; https://storytellingwithdata.com/podcast
+ Jonathan Schwabish; *PolicyViz Podcast*; https://policyviz.com/podcast/

**Online Resources**

++ *Multiple Views: Visualization Research Explained*; https://medium.com/multiple-views-visualization-research-explained
++ Robert Kosara; *EagerEyes*; eagereyes.org
++ Reddit; Data Is Ugly; https://reddit.com/r/dataisugly/
++ Reddit; Data Is Beautiful; https://reddit.com/r/dataisbeautiful/
+ Andy Kirk; *Visualising Data*; visualisingdata.com
+ David McCandless; *Information is Beautiful*; informationisbeautiful.net
+ Nadieh Bremer; *Visual Cinnamon*; visualcinnamon.com
- Cole Nussbaumer Knaflic; *Storytelling with Data*; storytellingwithdata.com
- Nathan Yau; *Flowing Data*; flowingdata.com
+ Makeover Monday; makeovermonday.co.uk
- Jonathan Schwabish; *HelpMeViz*; https://policyviz.com/helpmeviz/
- Lisa Charlotte Rost; *Chartable: A Blog by Datawrapper*; blog.datawrapper.de
- Severino Ribeccar; *The Data Visualisation Catalogue*; Overview of chart types. datavizcatalogue.com
- Alberto Cairo; *The Functional Art*; thefunctionalart.com
- Manuel Lima; *Visual Complexity*; visualcomplexity.com
- Michael Friendly; *Gallery of Data Visualization*; https://datavis.ca/gallery/
- *InfoVis Wiki*; infovis-wiki.net
- *InfoVis.org Mailing List*; https://infovis.org/pipermail/infovis/
Journals

++ Information Visualization; Sage; ISSN 1473-8716 https://journals.sagepub.com/home/ivi
++ IEEE Computer Graphics and Applications; ISSN 0272-1716 https://publications.computer.org/cga/
++ IEEE Transactions on Visualization and Computer Graphics; ISSN 1077-2626 https://publications.computer.org/tvcg/
++ Visual Informatics; ISSN 2468-502X https://sciencedirect.com/journal/visual-informatics
++ Nightingale; The Journal of the Data Visualization Society nightingaledvs.com [More magazine-style.]

Conferences

++ IEEE Vis. Formerly three separate co-located conferences: InfoVis, VAST, and SciVis. The first IEEE Symposium on Information Visualization (InfoVis) was in 1995. The main conference in the field, low acceptance rate (23% in 2017), very focussed, high quality papers. ieeewis.org
[The Vis 2020 conference talks are here: https://youtube.com/c/IEEEVisualizationConference/]
[The EuroVis 2021 and 2020 conference talks are on YouTube and are linked from the programme: https://eurovis21.ifi.uzh.ch/index.php/for-attendees/program/ and https://conferences.eg.org/egev20/program/]
++ Outlier. Annual Conference of the Data Visualization Society. outlierconf.com
[The Outlier 2021 conference talks are here: https://youtube.com/playlist?list=PLAm5TIX-yz7lK0eStK_vl&hl=en&gl=US]
++ IV (International Conference on Information Visualisation). Since 1997, usually in London. Broad in scope, fairly high acceptance rate, papers of mixed quality, multi-track. graphicslink.co.uk
++ PacificVis (IEEE Pacific Visualization Symposium). pvis.org
[The PacificVis 2021 conference talks are linked from the programme: http://vis.tju.edu.cn/pvis2021/program.html]
++ OpenVis. Annual conference, open-source tools for datavis. Usually in Boston, USA. Held upto 2018. openvisconf.com
++ Information+ Conference; informationplusconference.com
[The Information+ 2021 conference talks are linked from the schedule: https://informationplusconference.com/2021/]
++ See Conference, Germany. Largely German-speaking. see-conference.org
++ IVAPP (International Conference on Information Visualisation Theory and Applications). ivapp.scitevents.org
++ Some papers at CHI, AVI, UIST.

Professional Organisations

++ Data Visualization Society; datavisualizationsociety.com
[11,000+ members and very active Slack channels (datavizsociety.slack.com after approval)]
Local Organisations

+ On Data and Design Vienna; https://meetup.com/ondataanddesign-vienna/

InfoVis Companies

Suppliers of infovis toolkits and components:

• Tableau; tableau.com [Tableau was acquired by Salesforce on 01 Aug 2019 for $15.7 billion.]
• Qlik; qlik.com
• Microsoft Power BI; powerbi.microsoft.com
• Spotfire; https://tibco.com/products/tibco-spotfire [Spotfire was acquired by Tibco in May 2007 for $195 million.]
• Datawrapper; datawrapper.de
• Flourish; flourish.studio
• The Hive Group; hivegroup.com [The Hive Group was merged into Visual Action in Jan 2015.]
• Panopticon; panopticon.com
• macrofocus; macrofocus.com
• Maya Viz; mayaviz.com [Maya Viz was acquired by General Dynamics in Apr 2005.]
• OmniViz; omniviz.com [OmniViz was acquired by BioWisdom in Feb 2007. BioWisdom was acquired by Instem in Mar 2011.]
• AVS; avs.com
• NComVA; ncomva.com [NComVA was acquired by Qlik in 2013.]
• Advizor Solutions; advizorsolutions.com [Formerly Visual Insights, was renamed Advizor Solutions in 2003.]
• magnaview; magnaview.com
• Uncharted Software; uncharted.software [Formerly Oculus Info, was renamed Uncharted Software in 2015.]
• Tom Sawyer Software; tomsawyer.com
• yWorks; yworks.com
• ILOG; ilog.com [ILOG was acquired by IBM in Jan 2009.]
• Periscopic; periscopic.com
• Stamen Design; stamen.com
• datavisyn; in Linz; datavisyn.io [Located in Linz, Austria. Co-founded by Marc Streit and Alex Lex.]
### Table 1.1: Sales for 2012 in € by salesperson (ficticious sales data). Compare the sales figures of the two salespeople. Can you spot a trend?

<table>
<thead>
<tr>
<th>Month</th>
<th>Salesperson A</th>
<th>Salesperson B</th>
</tr>
</thead>
<tbody>
<tr>
<td>2012-01</td>
<td>28366</td>
<td>23274</td>
</tr>
<tr>
<td>2012-02</td>
<td>27050</td>
<td>21732</td>
</tr>
<tr>
<td>2012-03</td>
<td>29463</td>
<td>23845</td>
</tr>
<tr>
<td>2012-04</td>
<td>32561</td>
<td>28732</td>
</tr>
<tr>
<td>2012-05</td>
<td>28050</td>
<td>24023</td>
</tr>
<tr>
<td>2012-06</td>
<td>30100</td>
<td>26089</td>
</tr>
<tr>
<td>2012-07</td>
<td>22343</td>
<td>19026</td>
</tr>
<tr>
<td>2012-08</td>
<td>21506</td>
<td>17903</td>
</tr>
<tr>
<td>2012-09</td>
<td>24664</td>
<td>19387</td>
</tr>
<tr>
<td>2012-10</td>
<td>28842</td>
<td>23490</td>
</tr>
<tr>
<td>2012-11</td>
<td>30621</td>
<td>25873</td>
</tr>
<tr>
<td>2012-12</td>
<td>36254</td>
<td>28490</td>
</tr>
</tbody>
</table>

### 1.2 Why Visualise?

#### Table vs Line Chart

Compare the table of numbers in Table 1.1 with the visual representation (a line chart) of the same data in Figure 1.1.

- It is much easier to see trends and patterns in the visual representation.
- It is easier to make comparisons in the visual representation.
- It is easier to read off exact data values in the tabular representation (although you could, for example, display exact values upon mouseover in an interactive version of the line chart).

#### Attentive vs Pre-Attentive Processing

View Figure 1.2 and count the number of 3s. Then do the same with Figure 1.3.
Figure 1.1: Sales for 2012 in € by salesperson. Line chart of the same sales data. It is much easier to see the trends and compare the data, when it is presented visually.

175496490872545628327267094621
635280462905702676727325929055
561548569586711934907152874596
596289748716229184490082538851
180265490932887579802909278921
872634890928895000283058985889
927756990049828005987761883115

Figure 1.2: Count the number of 3s. Attentive processing requires conscious effort and proceeds serially.

- Text and numbers have to be processed attentively, which requires cognitive effort and proceeds in serial (slow).
- Certain visual attributes can be processed pre-attentively, which happens without conscious effort and in parallel (fast).

**Anscombe’s Quartet**

- Table 1.2 shows Francis Anscombe’s 1973 dataset [Anscombe 1973; Wikipedia 2013; Kosara 2011] with four variables in x and y.
- Standard summary statistics such as mean and standard deviation are almost identical for all four variables.
- However, when plotted graphically, as shown in Figure 1.4, the four variables are revealed to be very different.
1.2. WHY VISUALISE?

Figure 1.3: Count the number of 3s. Colour is a pre-attentive attribute. By encoding the target 3s in red, they can be rapidly processed by the human visual system pre-attentively. Pre-attentive processing occurs without conscious effort and in parallel.

<table>
<thead>
<tr>
<th></th>
<th>v1</th>
<th>v2</th>
<th>v3</th>
<th>v4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>x1</td>
<td>y1</td>
<td>x2</td>
<td>y2</td>
</tr>
<tr>
<td>10.00</td>
<td>8.04</td>
<td>10.00</td>
<td>9.14</td>
<td>10.00</td>
</tr>
<tr>
<td>8.00</td>
<td>6.95</td>
<td>8.00</td>
<td>8.14</td>
<td>8.00</td>
</tr>
<tr>
<td>13.00</td>
<td>7.58</td>
<td>13.00</td>
<td>8.74</td>
<td>13.00</td>
</tr>
<tr>
<td>9.00</td>
<td>8.81</td>
<td>9.00</td>
<td>8.77</td>
<td>9.00</td>
</tr>
<tr>
<td>11.00</td>
<td>8.33</td>
<td>11.00</td>
<td>9.26</td>
<td>11.00</td>
</tr>
<tr>
<td>14.00</td>
<td>9.96</td>
<td>14.00</td>
<td>8.10</td>
<td>14.00</td>
</tr>
<tr>
<td>6.00</td>
<td>7.24</td>
<td>6.00</td>
<td>6.13</td>
<td>6.00</td>
</tr>
<tr>
<td>4.00</td>
<td>4.26</td>
<td>4.00</td>
<td>3.10</td>
<td>4.00</td>
</tr>
<tr>
<td>12.00</td>
<td>10.84</td>
<td>12.00</td>
<td>9.13</td>
<td>12.00</td>
</tr>
<tr>
<td>7.00</td>
<td>4.82</td>
<td>7.00</td>
<td>7.26</td>
<td>7.00</td>
</tr>
<tr>
<td>5.00</td>
<td>5.68</td>
<td>5.00</td>
<td>4.74</td>
<td>5.00</td>
</tr>
<tr>
<td>mean</td>
<td>9.00</td>
<td>7.50</td>
<td>9.00</td>
<td>7.50</td>
</tr>
<tr>
<td>sd</td>
<td>3.3166</td>
<td>2.0316</td>
<td>3.3166</td>
<td>2.0317</td>
</tr>
</tbody>
</table>

Table 1.2: The four variables known as Anscombe’s Quartet look almost identical if only standard summary statistics such as mean and standard deviation (sd) are considered.
Figure 1.4: The four variables in Anscombe’s Quartet have almost identical summary statistics (mean, standard deviation, etc.). However, when plotted graphically, they are revealed to be rather different in nature.
1.3 Visual Lies

Sometimes, visualisations can be (deliberately) deceptive or misleading.

Resources

++ Alberto Cairo; *How Charts Lie*; W. W. Norton, 15 Oct 2019. ISBN 1324001569 (com, uk) [Cairo 2019]

+ Alberto Cairo; *Graphics Lies, Misleading Visuals*; Chapter 5 of New Challenges for Data Design, 27 Dec 2014. [Cairo 2014]

+ Nathan Yau; *How to Spot Visualization Lies*; 09 Feb 2017. [Yau 2017]


+ *WTF Visualizations*; viz.wtf [WTF 2017]

+ Reddit; *Data is Ugly*; https://reddit.com/r/dataisugly/

- David Shere and Hannah Groch-Begley; *A History Of Dishonest Fox Charts*; http://mediamatters.org/research/2012/10/01/a-history-of-dishonest-fox-charts/190225

- Mushon Zer-Aviv; *Disinformation Visualization: How to Lie with DataVis*; 31 Jan 2014. [Zer-Aviv 2014]


Misuse of 3D Perspective

- Surely, no-one would ever misuse 3d perspective to give a false impression?

- In his Macworld 2008 keynote speech [Jobs 2008, 00:08:24], Steve Jobs showed a 3d pie chart of US smartphone market share data by brand. Figure 1.5 shows the original slide and Figure 1.6 is a reconstruction of Jobs’ 3d pie chart. Table 1.3 shows the raw data in a table. [Video: https://youtu.be/i9XU6wk_mWY?t=08m24s]

  - The “Apple” slice of 19.5% is at the bottom of the pie chart and is tilted towards the viewer, making it appear much larger than the “Other” slice of 21.2% at the top.

  - A fairer representation would be to use a standard 2d pie chart (see Figure 1.7), or even better a bar chart (see Figure 1.8) of the same data.

- As Jack Schofield of the Guardian points out [Schofield 2008], the graphic is cleverly deceptive in at least two other ways as well:
  - Splitting market share by brand emphasises Apple. A more honest appraisal would be to split by operating system (OS), since certain OSes are used by several brands.
  - At the time, Symbian dominated the world smartphone market share, but was weak in the US, so showing only US data places Apple in the best light.
Figure 1.5: Steve Jobs’ 3d pie chart [Jobs 2008, 00:09:42].

<table>
<thead>
<tr>
<th>Manufacturer</th>
<th>Market Share (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RIM</td>
<td>39.0</td>
</tr>
<tr>
<td>Apple</td>
<td>19.5</td>
</tr>
<tr>
<td>Palm</td>
<td>9.8</td>
</tr>
<tr>
<td>Motorola</td>
<td>7.4</td>
</tr>
<tr>
<td>Nokia</td>
<td>3.1</td>
</tr>
<tr>
<td>Other</td>
<td>21.2</td>
</tr>
</tbody>
</table>

Table 1.3: Steve Jobs’ pie chart data showing U.S. smartphone market share for Q3 2007 as given by Gartner.

Video: Noah Iliinsky

- Noah Iliinsky; Data Visualizations Done Wrong; ORDcamp 2012 Ignite Talk, 5-minute video [Iliinsky 2012] Clips from [Video: https://youtu.be/1b7jaP1eEE4].
Figure 1.6: A reconstruction of Steve Jobs’ 3d pie chart, with 3d perspective making the lower segment appear larger than it should. [Chart created by Keith Andrews with LiquidDiagrams [Andrews and Lessacher 2010].]
1.4 General Concepts of Information Visualisation

General concepts which appear over and again in information visualisation.

Visualisation Pipeline

In Chapter 1 of their Readings volume, Card et al. [1999, page 17] describe a reference model for visualisation, shown in Figure 1.9:

• It describes the mapping of data to visual to support human interaction.
• It has become known as the visualisation pipeline.

The Information Visualisation Mantra

Ben Shneiderman’s information visualisation mantra:

“Overview, zoom and filter, details on demand”

Repeated ten times, once for each project where this principle was re-discovered...

From [Shneiderman 1996].

Utilising Human Visual Perception

Humans have remarkable perceptual abilities:
• to scan, recognise, and recall images rapidly.

• to rapidly and automatically detect patterns, outliers, and changes in size, colour, shape, movement, or texture.

Text-based interfaces require cognitive effort to understand their information content.

Information visualisation seeks to present information visually, in essence to offload cognitive work to the human visual perception system.

Visualisation + Interaction

• Interaction: being able to explore and manipulate the view is just as important as the underlying visual representation.

• Object Constancy: Smoothly animate transitions over about 1 sec. Eliminates the need for re-assimilation of the scene [Robertson et al. 1991a].

Focus-plus-Context

Focus on one or more areas of interest, while maintaining the surrounding context (but not in as much detail). Specific examples of focus-plus-context techniques include:

• Overview-plus-Detail: Two separate, but synchronised, windows or panes: an overview window (context) and a detail window (focus).

• 3d Perspective: Focus items are in the foreground, with the context in the background.

• Fisheye View: A geometric distortion (like a magnifying glass) is applied over the area of interest [Furnas 1981; Furnas 1986]. The focus in the centre is enlarged, while the surrounding context is made smaller.

• Focus-and-Blur: Focus items are (optically) in focus, while context items are blurred.

Guaranteed Visibility

Landmarks in the visualisation which are important to the user’s understanding remain visible at all times.
Edge Bundling

In node-link visualisations, there are sometimes so many edges that visual clutter results:

- Edge bundling [Holten 2006; Lhuillier et al. 2017] ties edges with similar paths together visually to reduce clutter.

- It is like applying cable ties to a bunch of computer network cables which follow the same path.

- In Figure 1.10, naive straight-line connections clutter the diagram.

- In Figure 1.11, edge bundling ties edges with similar paths together.

Excentric Labels

Sometimes, there is not enough room inside a visualisation to properly label individual elements:

- To avoid clutter, excentric labels [Fekete and Plaisant 1999; Welz 1999, pages 57 and 81] are drawn outside of the main visualisation and connected to the corresponding elements by lines.

See Figure 1.12.

Small Multiples

A small multiple is a table (lattice or grid) of charts or plots, using the same axes and scale for comparison. For example:

- The term was coined by Tufte [2001, page 170] and defined as “...a series of graphics, showing the same combination of variables indexed by changes in another variable.”
1.4. GENERAL CONCEPTS OF INFORMATION VISUALISATION

**Figure 1.11:** Edge bundling ties edges with similar paths together. [Figure made by Keith Andrews using sample R code from Holtz [2017]].

**Figure 1.12:** Excentric labels are drawn outside of the main visualisation and connected to the corresponding elements by lines. In the JExplorer tree browser, excentric labels are used to label the selected folder, its parent, and (some) siblings. [Screenshot made by Keith Andrews.]
Figure 1.13: The scatterplot matrix on the left shows all pairwise combinations of 13 performance variables about players from 16 clubs in five top European leagues in the 2017/18 season. The scatterplot on the right shows number of appearances plotted against total clearances. Players are represented by red triangles. [Screenshot made by Keith Andrews using mVis [Chegini et al. 2019].]

- For example, a scatterplot matrix is a table of scatterplots showing all pairwise combinations of variables, as shown in Figure 1.13.

1.5 Types of Information

- **Linear**: Tables, program source code, alphabetical lists, chronologically ordered items, etc.
- **Hierarchies**: Tree structures.
- **Networks**: General graph structures, such as hypermedia node-link graphs, semantic networks, webs, etc.
- **Multidimensional**: Metadata attributes such as type, size, author, modification date, etc. Items with n attributes become points in n-dimensional space.
- **Feature Spaces**: From information retrieval (IR), a feature vector represents each object in a collection. For collections of text documents, word frequencies are used to construct a term vectors. The feature space is projected to two or three display dimensions.
- **Query Spaces**: objects laid out according to their affinity with particular (combinations of) query terms.
- **[Spatial]**: Inherently 2d or 3d data such as floor plans, maps, CAD models, etc. with an obvious, natural geometric visual representation. [Not the focus of this course.]

**Video: David McCandless**

- David McCandless: *The Beauty of Data Visualization*, 18-minute video [McCandless 2010] [00:00–07:28, 17:09–17:54] [Video: https://ted.com/talks/david_mccandless_the_beauty_of_data...]
• More examples of his work are at informationisbeautiful.net.

**Video: Tamara Munzner**


• The nine prerecorded segments are available separately in this YouTube playlist: https://youtube.com/playlist?list=PLT4XLHmqHJBFfudbh19w_duD7mprvNf4Q

• Other talks and materials are available here: https://cs.ubc.ca/~tmm/vadbook/
Chapter 2

Visual Perception

“We see more with our mind than our eyes.”


References


+ Christopher Chabris and Daniel Simons; The Invisible Gorilla: And Other Ways Our Intuition Deceives Us; Harper Collins, Mar 2011. ISBN 000731731X (com, uk) [Chabris and Simons 2011b]


Online Resources

++ Christopher Healey; Perception in Visualization; https://csc.ncsu.edu/faculty/healey/PP/ [C. Healey 2009]

2.1 Human Vision

The eyes sample the environment 3–4 times per second, building up a picture of what is there in our mind:

- **Fixation**: a brief stationary period when visual information is sampled.
- **Saccade**: a period of rapid eye movement to a new location.

We actually see with our mind:

- What we see depends not only on what is actually there, but also on what we expect to see and where our attention is directed.
- “What you see is what you need”

### Selective Attention

We perceive only what the mind attends to:

- **Selective Attention Test (basketball passes)** by Dan Simons and Chris Chabris [Simons and Chabris 1999a; Chabris and Simons 2011b; Chabris and Simons 2011a].
- Based on work originally done by Ulric Neisser at Cornell University in the 1970s [Bazerman and Tenbrunsel 2016; Neisser 1979].
- **Original video** [Simons and Chabris 1999b] [Video: https://youtu.be/vJG698U2Mvo].
- **Multi-Colour Card Trick** by Richard Wiseman [Wiseman 2008] [Video: https://youtu.be/5HkG5rY189Y].

### Colour Perception

Colour perception is influenced, among other things, by:
2.2. VISUAL INPUT TAKES PRIORITY

- (Assumed) presence of shadows. [Video: https://youtu.be/XGG01xTYSoo?t=1s] [Video: https://youtu.be/z9Sen1HTu5o?t=24s]

Depth Perception

- Perspective, chair.
- Perspective, room.
- Perspective, rail lines. [Video: https://dailymotion.com/video/xq1rfl?start=736]

2.2 Visual Input Takes Priority

- Sight overrides hearing (McGurk effect). [Video: https://youtu.be/G-lN8vWm3m0?t=26s]
- Sight can override touch (rubber hand). [Video: https://youtu.be/Qsmkgi7FgEo?t=01m40s]

Visual Working Memory

The capacity of visual working memory is limited to 3–4 simple shapes [Vogel et al. 2001].

2.3 Preattentive Processing

Certain visual attributes can be processed preattentively (in parallel rather than serially) by the human visual system..

They “pop out” of the screen at you (without conscious effort).

When a visual target is encoded using a preattentive attribute, the time to locate it remains constant, regardless of the number of distractors.

Preattentive Attributes (Pop-Out Features)

The following attributes are pre-attentive:

- Colour hue (see Figure 2.1).
- Colour intensity (luminance) (see Figures 2.2 and 2.3).
- Line length (see Figure 2.4).
- Line width (see Figure 2.5).
- 2d spatial position (see Figure 2.6).
- Orientation (see Figure 2.7).
- Size (see Figure 2.8).
CHAPTER 2. VISUAL PERCEPTION

Figure 2.1: Colour hue is preattentive.

Figure 2.2: Colour intensity is preattentive.

- Shape (see Figure 2.9).
- Added mark (see Figure 2.10).
- Enclosure (see Figure 2.11).
- Lack of enclosure (see Figure 2.12).
- Curvature (see Figure 2.13).
- Focus and blur (see Figure 2.14).
- 3d depth (see Figure 2.15).
- Flicker (blinking) (see Figure 2.16).
- Velocity of motion (see Figure 2.17).
- Direction of motion (see Figure 2.18).

Some preattentive attributes are perceived more readily (are perceptually stronger) than others.
2.3. PREATTENTIVE PROCESSING

Figure 2.3: Colour intensity is preattentive.

Figure 2.4: Line length is preattentive.

Figure 2.5: Line width is pre-attentive.
Figure 2.6: 2D spatial position is preattentive.

Figure 2.7: Orientation is preattentive.

Figure 2.8: Size is preattentive.
2.3. PREATTENTIVE PROCESSING

Figure 2.9: Shape is preattentive.

Figure 2.10: An added mark is preattentive.

Figure 2.11: Enclosure is preattentive.
**Figure 2.12:** Lack of enclosure is preattentive.

**Figure 2.13:** Curvature is preattentive.

**Figure 2.14:** Focus and blur is preattentive.
2.3. PREAMTENTIVE PROCESSING

Figure 2.15: 3d depth is preattentive.

Figure 2.16: Flicker (blinking) is preattentive. [Figure is interactive, only works in HTML slides.]

Figure 2.17: Velocity of motion is preattentive. [Figure is interactive, only works in HTML slides.]
**Figure 2.18:** Direction of motion is preattentive. [Figure is interactive, only works in HTML slides.]

**Figure 2.19:** The conjunction of shape and depth remains preattentive.

**Conjunctions**

Beware when using multiple encodings in one display:

- Most preattentive attributes cannot be combined while still remaining pre-attentive.
- Conjunctions of motion, depth, colour hue, and orientation remain pre-attentive.

See Figures 2.19 and 2.20.


2.4 Visual Encoding

References


+ Peter Aldhous; *Data Visualization: Basic Principles*; https://paldhous.github.io/ucb/2019/dataviz/week2.html

Encoding Quantitative Values

Only two of the preattentive attributes can be perceived quantitatively with any accuracy:

• Line length.
• 2d spatial position.

Only these two can be used to accurately encode continuous quantitative numerical values, for example:

• location on a line graph (2d spatial position).
• location on a scatter plot (2d spatial position).
• length of a bar in a bar graph (line length).

Perceiving Continuous Differences

Differences in colour intensity and size can be perceived to a limited degree:

• one shade of grey is lighter than another
• one circle is larger than another

but not by how much.

It is hard to compare them accurately on a continuous scale:
• This shade of grey is 10% lighter than that one.
• This slice of a pie chart is 5% smaller than that one.

In practice, about 5 distinct shades of gray (or intensities of any single colour hue) can be distinguished.

**Encoding Categories**

Some preattentive attributes are more naturally non-continuous.

The following preattentive attributes are particularly useful for encoding discrete categories:

• Colour hue (palette).
• Shape (fixed number of distinct glyphs).

**Perceptually Distinguishable Colour Palettes**

Maureen Stone has proposed a selection of 8 distinguishable colours for colour-coding bar charts or line charts.


Chapter 3

History of Information Visualisation

References
++ Sandra Rendgen; History of Information Graphics; Taschen, 08 May 2019. ISBN 3836567679 (com, uk) [Rendgen 2019]
++ Michael Friendly and Howard Wainer; A History of Data Visualization and Graphic Communication; Harvard University Press, 08 Jun 2021. ISBN 0674975235 (com, uk) [Friendly and Wainer 2021]
++ Isabel Meirelles; Design for Information; Rockport, 01 Oct 2013. ISBN 1592538061 (com, uk) [Meirelles 2013]
+ Theodore W. Pietsch; Trees of Life; Johns Hopkins University Press, 02 May 2013. ISBN 1421411857 (com, uk) [Pietsch 2013]
+ Manuel Lima; The Book of Trees; Princeton Architectural Press, 08 Apr 2014. ISBN 1616892188 (com, uk) [Lima 2014]
+ Kruja et al; A Short Note on the History of Graph Drawing; Proc. Graph Drawing (GD 2001). [Kruja et al. 2002]
+ Jacques Bertin; Semiology of Graphics; ESRI Press, 2010. ISBN 1589482611 (com, uk) [Bertin 2010] [Reprint]
+ Jacques Bertin; Sémiologie graphique; Editions de l’Ecole des Hautes Etudes en Sciences, 1999. ISBN 2713212774 (com, uk) [Bertin 1999] [In French]

Online Resources
++ Michael Friendly and Daniel J. Denis; Milestones in the History of Thematic Cartography Statistical Graphics, and Data Visualization; 2018. https://datavis.ca/milestones/

3.1 Diderot and d’Alembert
• Denis Diderot and Jean le Rond d’Alembert, 1751.
A taxonomy of human knowledge, called the “Figurative System of Human Knowledge”.

Inspired by Francis Bacon’s taxonomy from 1620 [Bacon 1620].

Also known as “The tree of Diderot and d’Alembert” [Diderot and le Rond d’Alembert 1751b].

The top three branches are Memory, Reason, and Imagination.

**Figurative System of Human Knowledge (1751)**

Produced as the table of contents for the “Encyclopedia, or a systematic dictionary of the sciences, arts, and crafts” published in 1751 [Diderot and le Rond d’Alembert 1751a].

Shown in Figure 3.1.

### 3.2 Joseph Priestley

Joseph Priestley was an English scientist and theologian:

- He produced some of the first timeline charts [Sheps 1999].

- In 1765, he produced his Chart of Biography, a historical timeline of around 2000 people, divided into six horizontal bands (categories) such as “Poets and Artists”, “Mathematicians and Physicians”, and “Statesmen and Warriors”.

- In 1769, he produced the New Chart of History, shown in Figure 3.2. It spans 3000 years of history from 1200 BC to 1800 AD with 16 horizontal bands corresponding to geographical regions.
3.3 William Playfair

William Playfair was a Scottish engineer and economist:

- He is considered the founder of statistical graphics, inventing among other things the line chart, area chart, pie chart, and bar chart.

- An example of an area chart can be seen in Figure 3.3, taken from the 3rd edition of his Commercial and Political Atlas [Playfair 1801a].

- Two examples of a pie chart can be seen in Figure 3.4, taken from his Statistical Breviary [Playfair 1801b].
Figure 3.3: One of William Playfair’s area charts, this example from the 3rd edition of his Commercial and Political Atlas. [Schoenberg Center for Electronic Text and Image (SCETI). Used under the terms of Creative Commons CC BY 2.5.]

Figure 3.4: Two examples of a pie chart within this graphic from William Playfair’s Statistical Breviary. [Schoenberg Center for Electronic Text and Image (SCETI). Used under the terms of Creative Commons CC BY 2.5.]
3.4 Edmund Cooper and John Snow

London Cholera Map (1854)

- In Sep 1854, a cholera epidemic hit the Golden Square area of London around Broad Street (St. James’ parish).

- The first spot map of cholera deaths is usually attributed to Dr. John Snow, 1854, [Frerichs 2009], but in fact an earlier spot map was drawn by Edmund Cooper and presented on 26 Sept 1854, some 3 months earlier [Brody et al. 2000].

- Edmund Cooper was an engineer at the Metropolitan Commission of Sewers. Public complaints had linked the sewers to the cholera outbreak.

- Cooper plotted cholera deaths at addresses on a spot map, and used the map as an analytical tool to deduce that addresses near sewer holes did not exhibit higher numbers of deaths.

- Upto this time, cholera was thought to be an airborne disease, although Snow himself had long postulated a waterborne link.

- In popular retellings, such as Tufte [1997, pages 27–37] and Henig [1996, pages 1–2], Snow first plotted the deaths on a spot map (see Figure 3.5), and then used the map to discover a higher clustering of deaths around the Broad Street water pump. The handle on the Broad Street pump was removed, and the epidemic subsided.

- In fact, as recounted in Brody et al. [2000], Snow first presented a spot map on 04 Dec 1854, almost 3 months after the outbreak.

- Snow’s map was not used as an analytical tool, but rather as an illustration after the event, to illustrate his finding that cholera was a waterborne disease.

- Workers in the nearby brewery, which had its own water (and beer) supply, were largely unaffected.

- Not really InfoVis, more GeoVis, since it is based on an underlying map.
3.5 Florence Nightingale

- Famous British nurse and statistician.
- She used a number of statistical graphics to illustrate her data and support her argument for health care reform [Small 1998].

**Bat’s Wing Diagram (1858)**

- Shows changes over time for a small number of variables.
- In this case, the number of deaths from three different causes over a 12-month period, as shown in Figure 3.6.
- The time points are shown as 12 equally spaced radial lines (one for each month, winter at the bottom).
- The number of deaths is proportional to the length of the radial line (length-proportional).
- However, the coloured shading in of the areas (incorrectly) suggests to the viewer that the diagram is area-proportional.

**Wedge Diagram (1858)**

- Florence Nightingale, 1858.
- After criticism that her bat’s wing diagram could be misleading, Nightingale redesigned it as a wedge diagram, shown in Figure 3.8.
Figure 3.6: Florence Nightingale’s original bat’s wing diagram. [Image from the Wellcome Library, London. Used under the terms of Creative Commons CC BY-NC 2.0 UK.]

Figure 3.7: Modern bat’s wing diagram.
• Now, the number of deaths is actually proportional to the coloured shaded areas.

• However, there are some problems with occlusion, depending on the drawing order of the wedges.

• I use the name wedge diagram, following Small [1998]. It is sometimes also called a polar area diagram or a rose diagram. [It is also sometimes incorrectly referred to as a coxcomb diagram [Small 1998].]
3.6 Charles Minard

Diagram of Napoleon's Russian Campaign (1861)

• Charles Joseph Minard, 1861.
3.7 Fletcher Hewes and Henry Gannett

Fletcher Hewes and Henry Gannett produced the 1883 edition of Scribner’s Statistical Atlas of the United States [Hewes and Gannett 1883]:

- Plate 151 of the Atlas is shown in Figure 3.11.
- It shows a ranking of the then 47 states, according to ten dimensions.
- Each state appears once in each ranking and its instances are connected by a line, an early form of parallel coordinates plot.
- The darker blue boxes indicate the position of the United States as a whole.
3.8 Marey and Ibry (1878)

In his 1878 book [Marey 1878], Entienne Jules Marey published a graphic train schedule:

- The schedule is shown in Figure 3.12, which is taken from the 2nd Edition of the book [Marey 1885, page 20].

- He attributes the graphic to a Mr. Ibry, who is reported to have been deputy chief operating officer of the railway from Paris to Rouen [Wainer et al. 2013, page 56].

- Ibry had apparently been using this kind of display since at least 1846.

- As Wainer et al. [2013] describe, similar graphic train schedules had also been published in Russia in 1854.
Figure 3.12: Marey’s graphic train schedule, which he attributes to Ibry. [Image from the Internet Archive, in the public domain.]

3.9 Paul Otlet

Radial tree visualisation(s).

3.10 Jaques Bertin

Reorderable Matrix

- Jaques Bertin created a physical device for reordering matrices called Domino, see Figure 3.13.
- It is described in more detail by Henry [2008, page 78].
- Review paper and web app version at https://aviz.fr/bertifier.
Figure 3.13: Bertin’s reorderable matrix. [Photograph used with kind permission of Jean-Daniel Fekete.]
Chapter 4

Visualising Linear Structures

Linearly structured information:

• alphabetical lists
• program source files
• chronological lists
• ranked search results

4.1 Bifocal Display

• Invented by Robert Spence and Mark Apperley in 1980 [Spence and Apperley 2018; Apperley et al. 1982]

• Separate focus and context areas. See Figure 4.1.

• Video at http://www.ee.ic.ac.uk/r.spence/BobsVideos_files/V11BiFocal%20Office%20copy.mov (save first and then play).
4.2 Perspective Wall

- 3d perspective technique for linear information.
- Focus on the front panel and context in the side panels. See Figure 4.2.
- CHI’91 paper [Mackinlay et al. 1991] and video [Robertson et al. 1991b].  
  https://youtu.be/hYUZbrWtCZg
- US Patent 5339390 [Robertson et al. 1994b].
4.3 Seesoft

- AT&T Bell Labs, 1992.
- Focus + context technique for large amounts of source code.
- Lines of code are compressed down to rows of pixels. See Figure 4.3.
- Like hanging program listings on a wall several metres away.
- Zooming window supports several levels of zoom, from overview to lines of code.
- Articles [Eick et al. 1992; Ball and Eick 1996] and InterCHI’93 video [Steffen and Eick 1993].
Figure 4.3: Seesoft visualising software consisting of 38 files comprising 12037 lines of code. The newest lines are shown in red, the oldest in blue, with a rainbow colour scale in between. [Image used with kind permission of Steve Eick, Visual Insights.]

4.4 Lifestreams

- Yale University, 1995.
- Streams of chronologically ordered items.
- AAAI 1995 paper [Freeman and Fertig 1995], CHI’96 video [Fertig et al. 1996], Wired article [Steinberg 1997].
4.5 **Spiral of Ranked Items**

- A spiral is a compact way of displaying a list of ranked items.
- GopherVR used a spiral to display search results. See Figure 4.5.
- 1994 draft version of paper for ED-MEDIA 95 [McCahill and Erickson 1994].

**JUCS Spiral of Authors**

- High-profile authors in a sub-field of computer science are visualised in a spiral (in decreasing order of number of publications in that sub-field). See Figure 4.6.
- FIT 2009 paper [Afzal et al. 2009].
Figure 4.5: GopherVR spiral of search results. [Image used with kind permission of Tom Erickson]

Figure 4.6: JUCS spiral of authors.
Chapter 5

Visualising Hierarchies

“The organization of ideas and objects into categories and subcategories is fundamental to human experience. We classify to understand. Tree structures lie at the roots of our consciousness.”

[Peter Morville, Ambient Findability, page 127, Sept. 2005.]

Hierarchies are extremely common:

- file systems
- library classification systems
- family trees

Many general graphs (networks) can also be transformed to a tree plus backlinks.

References


+ Manuel Lima; The Book of Trees; Princeton Architectural Press, 08 Apr 2014. ISBN 1616892188 (com, uk) [Lima 2014]

+ Jürgensmann and Schulz; A Visual Survey of Tree Visualization; Best poster at IEEE InfoVis 2010. [Jürgensmann and Schulz 2010]

Online Resources

- Hans-Jörg Schulz; A Visual Bibliography of Tree Visualization; treevis.net
5.1 Outliners

5.1.1 Tree Views

- Tree view on left shows structure, list view on right shows items (files, documents) at a particular level.
- Windows Explorer.
- Java Swing JTree component (see Figure 5.1).

5.1.2 MagTree

- Andy Clark and Dave Smith, IBM, Jan 2001.
- File Magnitude Viewer (MagTree).
- Part of a tutorial on Java tree views [Smith and Clark 2001].
- Extends traditional tree view with preview of size statistics.
- Coloured pie indicates proportion of bytes in each subdirectory.
- Sizes are either relative to the root or relative to the parent.
5.1.3 WebTOC

- David Nation, Department of Defense and HCIL, 1997.
- Generates tree view of web site.
- Extends traditional tree view by overlaying supplementary statistical information.
- Coloured bars indicate proportion of various media types, shadows indicate number of files.
Figure 5.3: A WebTOC table of contents for the University of Maryland’s HCI Lab web site.

5.2 Layered Node-Link Tree Browsers

5.2.1 Walker Layout

- Walker [1990], improved by Buchheim et al. [2002].
- Drawing trees of unbounded degree in linear time.
- The root is at the top, children on successive layers beneath.
- The layout progresses bottom-up, allocating the same amount of space to each leaf node. See Figure 5.4.

5.2.2 File System Navigator (FSN)

- 3d landscape visualisation of file system.
- Files sit atop pedestals, subdirectories recede into the background. See Figure 5.5.
- Featured in Jurassic Park.
- Used in MineSet product to visualise decision trees.
- Software (binaries) available online [Tesler and Strasnick 1992].
- Patented under [Strasnick and Tesler 1996a; Strasnick and Tesler 1996b].
Figure 5.4: The Walker tree browser.

Figure 5.5: FSN landscape visualisation of a file system hierarchy. Files sit atop pedestals, subdirectories recede into the background.
5.2.3 Harmony Information Landscape

- 3d landscape visualisation of Hyperwave collection structures. See Figure 5.6.
- Similar to FSN, documents sit atop pedestals, subcollections recede into the background.
- Paper at IEEE InfoVis’95 (reprinted in [Card et al. 1999]), [Andrews 1996].
5.3 Radial Node-Link Tree Browsers

5.3.1 Radial Tree

- Classic tree layout technique.
- The root is in the centre of a circle, surrounded by its descendents on concentric circles based on their depth.
- In fixed-wedge layouts, each branch of the tree is constrained within a wedge of the circle (“annulus wedge”) [di Battista et al. 1999, pages 52–53].
- In flexible-wedge layouts, children are allowed to make use of free space in neighbouring wedges.
- Figure 5.7 shows the radial tree implemented in HVS [Andrews et al. 2007; Fuchs 2015].

5.3.2 Hyperbolic Browser

- Focus+context technique, always displaying the entire hierarchy. See Figure 5.8.
- Layout on a hyperbolic plane, which is then mapped to the unit disc. Each child places its children in a wedge of space.
- Developed as a software component by Inxight, acquired by Business Objects, now owned by SAP.

Figure 5.7: A radial tree from HVS.
**Figure 5.8:** The original hyperbolic browser from Xerox. The hyperbolic browser always displays the entire hierarchy. However, subtrees around the edge of the disk become so small they are invisible. [Image used with kind permission of Ramana Rao, Xerox PARC].

- Won the CHI’97 Great Browse Off!

### 5.3.3 D3 Hypertree

- Open-source implementation of hyperbolic browser in TypeScript using D3.
- Examples:
  - Product Taxonomy; https://keithandrews.com/demos/ivis/hypertree/hypertree-products/. See Figure 5.9.
  - Hyperbolic Tree of Life; hyperbolic-tree-of-life.github.io See Figure 5.10.
  - Several examples on Observable; https://observablehq.com/@glouwa.

### 5.3.4 3D Hyperbolic Browser

- Tamara Munzner, University of Minnesota and Stanford University.
- 3D hyperbolic browser. See Figure 5.11.
Figure 5.9: D3 Hypertree displaying Google’s product taxonomy.

Figure 5.10: The Hypertree of Life is a practical application of D3 Hypertree.
• For web sites, spanning tree is generated and laid out, cross-links are displayed on request.
• Paper at VRML’95 [Munzner and Burchard 1995] and InfoVis ’97 [Munzner 1997].

5.3.5 Walrus

• Young Hyun, CAIDA, 2002.
• 3D hyperbolic browser, homegrown implementation of Tamara Munzner’s algorithms. See Figure 5.12.
• Paper in BMC Bioinformatics Journal 2004 [Hughes et al. 2004], web site [Hyun 2005].

5.3.6 SInVis Magic Eye View

• The hierarchy is first laid out in 2d space according to the classic Reingold [Reingold and Tilford 1981] or Walker [Buchheim et al. 2002] algorithm.
• It is then mapped geometrically onto the surface of a hemisphere. See Figure 5.13.
• Smooth animated transitions are possible.
• The effect is similar to a hyperbolic browser, but hyperbolic geometry is not used.
Figure 5.12: The Walrus 3d hyperbolic browser displaying a directory tree. [Image used with kind permission of Young Hyun, CAIDA.]

Figure 5.13: The SLNvis Magic Eye View. [Image used with kind permission of Matthias Kreuseler, University of Rostock.]
5.3.7 Cone Tree

- 3d conical representation of tree. See Figure 5.14.
- A horizontal layout (cam tree) allows better labeling of nodes.
- Patented under [Robertson et al. 1994a].

5.3.8 Botanical Visualisation

- An abstract tree is converted into a geometric model of branches and leaves and then rendered.
- For better aesthetics, continuing branches are emphasised, long branches are contracted, and leaves are shown as fruit.
- Paper at InfoVis 2001 [Kleiberg et al. 2001]. [Video: https://youtu.be/mTipHD_MBZk?t=8m35s]
5.4 Layered Space-Filling Tree Browsers

5.4.1 Xdu

- Phil Dykstra, Army Research Laboratory, 1991.
- Utility for the X window system which displays a graphical disk usage for Unix file systems.
- Rectangles are stacked from left to right as the directory tree is descended.
- The vertical space allocated is proportional to size of each subdirectory.
- Software (source) available online [Dykstra 1991].
5.5 Radial Space-Filling Tree Browsers

5.5.1 Information Slices

- The hierarchy is fanned out across one or more semi-circular discs. See Figure 5.17.
- The number of levels displayed on each disc can be changed interactively, 4 or 5 works well.
- The area of each segment is proportional to the total size of its contents.
- Clicking on a directory in the left disc fans out its contents in the right disc, allowing rapid exploration of large hierarchies.
- For very deep hierarchies, clicking on a directory in the right disc causes the left disc to be miniaturised and slide off to the left (to join a stack of miniature discs), and a fresh disc is opened to the right.

5.5.2 SunBurst

- Much more advanced version of InfoSlices. See Figure 5.18.
- Uses full disc and implements fan-out of subtrees.
5.5. **RADIAL SPACE-FILLING TREE BROWSERS**

**Figure 5.17:** An Information Slices visualisation of the JDK 1.2 tree. For deeper hierarchies discs are stacked up in the left margin.

- Video at InfoVis 2000 [Stasko and E. Zhang 2000b]. [Video: https://youtu.be/YaMmDU6GIpo]
5.6 Inclusive Space-Filling Tree Browsers

5.6.1 Tree Maps

- Screen-filling visualisation by alternate vertical and horizontal slicing of available space ("Slice and Dice" treemap layout), as shown in Figure 5.19.
- The size of each rectangle is proportional to its weight, typically the total number or size of items within it.
- Child rectangles can be ordered (say alphabetically) within their parent rectangle, but rectangles can degenerate to very narrow strips.
- Visualization’91 paper [Johnson and Shneiderman 1991] and CHI’94 video [Turo 1994].
- Software at https://cs.umd.edu/hcil/treemap/.

5.6.2 Market Map

- Extension of tree map, avoiding the excessively narrow strips produced by Slice and Dice. See Figure 5.20.
- Uses a heuristic to slice up each rectangle into more evenly proportioned sub-rectangles ("Squarified" treemap layout) by placing the largest child first.
Figure 5.19: A tree map of the Dewey Decimal classification hierarchy widely used in libraries. Copyright ©University of Maryland 1984-1994, all rights reserved.

- Squarified tree maps look better and sub-rectangles are more easily compared in size, but at the cost of no ordering of child rectangles within the parent rectangle.


5.6.3 Cushion Treemaps

- Cushion treemaps: Three-dimensional shading is used to indicate the borders between treemap regions. See Figure 5.21.

- Explicit borders are no longer necessary, freeing up pixels for the data.

- Software package called SequoiaView produces disk usage maps using squarified cushion treemaps [SequoiaView 2005].

- Many clones of SequoiaView have been built for various platforms: KDirStat for Unix/X11 [Hundhammer 2010], WinDirStat for Windows [Seifert and Schneider 2018], and Disc Inventory X for Mac [Derlien 2010].


5.6.4 Information Pyramids

Figure 5.20: A market map of US stocks generated on 17th September 1999.

Figure 5.21: SequoiaView produces disk usage maps using squarified cushion treemaps.
A plateau represents the root of the tree. Other, smaller plateaux arranged on top of it represent its subtrees. See Figure 5.22.

The size of each block is, by default, proportional to the total size of its contents.

Separate icons are used to represent non-subtree members of a node such as files or documents.

The overall impression is that of pyramids growing upwards as the hierarchy is descended.

A redesigned version combines a pyramids display with a Java tree viewer. See Figure 5.23.

Figure 5.22: An Information Pyramids visualisation of the JDK 1.2 tree. The view from above gives a graphical disk usage. The dashboard provides user navigation. This version uses OpenGL for 3d output.


5.6.5 InfoSky Cobweb Browser

First example of recursive Voronoi subdivision for hierarchy visualisation.

Each parent polygon is subdivided into Voronoi polygons, one for each child, whose area is proportional to its size. See Figures 5.24 and 5.25.


5.6.6 Voronoi Treemap

Three years after InfoSky Cobweb.
Figure 5.23: The JExplorer combines a Java tree viewer with a synchronised Information Pyramids display.

Figure 5.24: The InfoSky Cobweb hierarchy browser.
Figure 5.25: The InfoSky Cobweb hierarchy browser, zoomed in for a closer look.

- Also uses space-filling, recursive voronoi subdivision to allocate available space to child polygons. See Figure 5.26.
- Video at [Balzer and Deussen 2005b]. [Video: https://youtu.be/3Qd56UBDKQE]
5.7 Overlapping Space-Filling Tree Browsers

5.7.1 Cheops

• Centre du recherche Informatique de Montréal, 1996.

• Compact 2d representation of a hierarchy by overlaying (squashing together) children to save on screen space. See Figure 5.27.

• Paper at Visualization'96 [Beaudoin et al. 1996].

• Software (Java classes) at CRIM [2002].

5.7.2 BeamTree

• Directories are in blue, files in other colours.

• The root beam is in the background, other beams are laid on top.

• See Figures 5.28 and 5.29.

• InfoVis 2002 paper [van Ham and van Wijk 2002].
Figure 5.27: Cheops uses stacked triangles to compactly display a hierarchy.

Figure 5.28: BeamTrees are a variation on treemaps using overlapping beams. Directories are coloured blue, files are other colours. The root directory is at the back, other directories are overlaid upon it.
5.8 Single-Level Tree Browsers

5.8.1 GopherVR

- University of Minnesota, 1995.

- 3d landscape visualisation of individual levels of a Gopher hierarchy. Members of a collection are arranged in a stonehenge-like circle.

- Spiral visualisation of Gopher search result sets, spiraling out from centre with decreasing relevance.

- Possibility to hand-place items, for example grouping related items.

- Cute physical simulation of shaking when landing, can jump up and look around.

Figure 5.30: GopherVR visualises one level of a Gopher hierarchy at a time. The central pyramid bears the name of the current level, clicking on it returns the user to the next higher level.
Chapter 6

Visualising Networks and Graphs

6.1 Adjacency Matrix

An adjacency matrix explicitly tabulates links between nodes.

- Also sometimes called a connectivity matrix and a design structure matrix (DSM).
- For a graph with $N$ nodes, an $N \times N$ matrix is used to indicate where edges occur.
- Figure 6.1 shows an example for a graph with five nodes and six edges.
6.2 Predetermined Position

The nodes of a graph are laid out in predetermined positions (linear, circular, grid, geographical) and the edges are routed and encoded in various ways.

6.2.1 Linear

Thread Arcs

- Thread Arcs were developed to visualise threads of conversation between email or newsgroup messages [Kerr 2003].
- ThreadVis is an implementation of Thread Arcs for the Thunderbird news and email client [Hubmann-Haidvogel 2008]. See Figure 6.2. [http://threadvis.mozdev.org/]

6.2.2 Circular

Also called ring-based layouts or chord diagrams.
Figure 6.3: A circular plot of migration flows between the world’s top 25 sending and receiving countries between 2005 and 2010. A large gap at the outer segment indicates destination, a small gap indicates origin. The tick marks show the numbers of migrants in millions [Image created by Keith Andrews using data and R code kindly provided by G. Abel [2014]].

Migration Flows

- Figure 6.3 shows a circular plot of migration flows between the world’s top 25 sending and receiving countries between 2005 and 2010.

Flare Dependency Graph

- The Flare Dependency Graph is a circular layout showing dependencies among classes within the Flare library [Heer 2010].
- Each class is placed around the edge of a circle. The exact radius indicates the depth of the class in the package structure tree.
- A link indicates that a class imports another.
- Edges are “bundled” together for greater clarity.
- See Figure 6.4.
6.2.3 Grid-Based

Intermedia Global Map

- Intermedia was a network hypermedia system developed for Apple computers in the 1980s. [Haan et al. 1992; Yankelovich et al. 1988].
- Hypermedia nodes (documents) are placed on a regular grid and links are drawn between them.
- See Figure 6.5.

6.2.4 Geography-Based

Flow Maps

- Historical flow maps created by Charles Minard.
- Layout algorithm described by Doantam Phan et al in InfoVis 2005 paper [Phan et al. 2005].
- See Figure 6.6 shows migration to California using US Census data from 2000.

**Figure 6.4**: The Flare Dependency Graph is a circular layout showing the dependencies between classes in the Flare library. Here the class Data has been selected.
Figure 6.5: The Intermedia Global Map. Hypermedia nodes (documents) are placed on a regular grid and links are drawn between them. [Image extracted from Conklin [1987]. Copyright ©1987 IEEE. Used with permission.]

Figure 6.6: A flow map showing migration to California from other US states, using data from the US Census 2000. [Image created by Keith Andrews using the software available from Phan et al. [2006]].
6.3 Layered Graph Drawing

Layered Graph Drawing, also called Sugiyama layout. User for (acyclic) directed graphs.

Three main steps:
1. Layering
2. Crossing reduction
3. X-coordinate assignment

Ideal for directed graphs: directionality is reflected in the layering (flow from top to bottom, or left to right).

6.3.1 Harmony Local Map

- Graph layout for nodes and links of a hypermedia network.
- Modified version of Eades and Sugiyama’s [Eades and Sugiyama 1990] graph layout algorithm [di Battista et al. 1999].
- Description in Chapter 8 of [Andrews 1996].

Harmony Local Map 3D

- Links in vertical plane superimposed atop information landscape. See Figure 6.8.
- Description in Chapter 8 of [Andrews 1996].
6.4 Force-Based Layouts

6.4.1 SemNet

- The first 3d information visualisation.
- 3d spatial layout of a semantic network. See Figure 6.9.
- Article [K. M. Fairchild et al. 1988].
- Video at CHI ’87 [K. Fairchild 1987]. [Video: https://youtu.be/wAu75r77Qko?t=03m00s]
- Patented under [Wexelblat and K. M. Fairchild 1991].

6.4.2 HyperSpace (Narcissus)

- University of Birmingham, 1995.
- Self-organising structure based forces and springs. See Figure 6.10.
- The number of links between documents provides the attractive force.
- Narcissus [Hendley et al. 1995], later renamed HyperSpace [Wood et al. 1995].
Figure 6.9: SemNet visualised a semantic network in 3d. [Image used with kind permission of Kim Fairchild.]

Figure 6.10: HyperSpace showing the structure of a website. White spheres represent web pages. Green lines represent links between pages. The size of a sphere is proportional to the total number of incoming and outgoing links. [Image used with kind permission of Andy Wood.]
Chapter 7

Visualising Multidimensional Metadata

"Getting information from a table is like extracting sunlight from a cucumber."

[Arthur and Henry Farquhar, Economic and Industrial Delusions, Putnam, New York, 1891.]

7.1 Interactive Tables

7.1.1 Table Lens

• Xerox PARC, 1994.

• Focus + context technique for large tables.

• Rows and columns are squeezed down to pixel and subpixel sizes. See Figure 7.1.


CHAPTER 7. VISUALISING MULTIDIMENSIONAL METADATA

Figure 7.1: The Table Lens represents rows of a table as rows of pixels. The user can focus and stretch out rows or columns to see the data, whilst maintaining surrounding context. [Screenshot made by Keith Andrews using the Inxight Eureka 30-day free trial.]

7.2 Interactive Scatterplots

7.2.1 FilmFinder

- Sliders and controls directly manipulate an on-screen scatterplot.
- The scatterplot is called a “starfield display”.
- Commercialised as part of the Spotfire toolkit [Spotfire 2000].

7.2.2 Envision

- Direct manipulation of search result sets by mapping document attributes along two axes.
- SIGIR’96 paper [Nowell et al. 1996] and CHI’97 online abstracts [Nowell et al. 1997].
7.2. **INTERACTIVE SCATTERPLOTS**

---

**Figure 7.2:** The FilmFinder, a starfield display combined with dynamic queries for rapid filtering. [Copyright © University of Maryland 1984-1994.]

---

**Figure 7.3:** Envision visualises a set of search results, by mapping document attributes along two axes. Where too many documents would occupy a cell, an ellipse is used as a container object. Another problem is where to place documents matching multiple categories. [Copyright © by the Association for Computing Machinery, Inc.]
7.2.3 Search Result Explorer

- IICM, 1999.
- Similar to Envision, Java implementation for the xFIND search engine.

7.3 Multidimensional Glyphs

Chernoff Faces

- Invented by Herman Chernoff in 1973 [Chernoff 1973].
- Dimensions (variables) are mapped to characteristics of cartoon-like human faces, such as the curvature of the mouth and the slant of the eyes.
- Figure 7.5 shows Chernoff faces for R’s built-in mtcars dataset.
<table>
<thead>
<tr>
<th>Name</th>
<th>mpg</th>
<th>cyl</th>
<th>disp</th>
<th>hp</th>
<th>drat</th>
<th>wt</th>
<th>qsec</th>
<th>vs</th>
<th>am</th>
<th>gear</th>
<th>carb</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mazda RX4</td>
<td>21.0</td>
<td>6</td>
<td>160.0</td>
<td>110</td>
<td>3.90</td>
<td>2.620</td>
<td>16.46</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Mazda RX4 Wag</td>
<td>21.0</td>
<td>6</td>
<td>160.0</td>
<td>110</td>
<td>3.90</td>
<td>2.875</td>
<td>17.02</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Datsun 710</td>
<td>22.8</td>
<td>4</td>
<td>108.0</td>
<td>93</td>
<td>3.85</td>
<td>2.320</td>
<td>18.61</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Hornet 4 Drive</td>
<td>21.4</td>
<td>6</td>
<td>258.0</td>
<td>110</td>
<td>3.08</td>
<td>3.215</td>
<td>19.44</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Hornet Sportabout</td>
<td>18.7</td>
<td>8</td>
<td>360.0</td>
<td>175</td>
<td>3.15</td>
<td>3.440</td>
<td>17.02</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Valiant</td>
<td>18.1</td>
<td>6</td>
<td>225.0</td>
<td>105</td>
<td>2.76</td>
<td>3.460</td>
<td>20.22</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Duster 360</td>
<td>14.3</td>
<td>8</td>
<td>360.0</td>
<td>245</td>
<td>3.21</td>
<td>3.570</td>
<td>15.84</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Merc 240D</td>
<td>24.4</td>
<td>4</td>
<td>146.7</td>
<td>62</td>
<td>3.69</td>
<td>3.190</td>
<td>20.00</td>
<td>1</td>
<td>0</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Merc 230</td>
<td>22.8</td>
<td>4</td>
<td>140.8</td>
<td>95</td>
<td>3.92</td>
<td>3.150</td>
<td>22.90</td>
<td>1</td>
<td>0</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Merc 280</td>
<td>19.2</td>
<td>6</td>
<td>167.6</td>
<td>123</td>
<td>3.92</td>
<td>3.440</td>
<td>18.30</td>
<td>1</td>
<td>0</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Merc 280C</td>
<td>17.8</td>
<td>6</td>
<td>167.6</td>
<td>123</td>
<td>3.92</td>
<td>3.440</td>
<td>18.90</td>
<td>1</td>
<td>0</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Merc 450SE</td>
<td>16.4</td>
<td>8</td>
<td>275.8</td>
<td>180</td>
<td>3.07</td>
<td>4.070</td>
<td>17.40</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Merc 450SL</td>
<td>17.3</td>
<td>8</td>
<td>275.8</td>
<td>180</td>
<td>3.07</td>
<td>3.730</td>
<td>17.60</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Merc 450SLC</td>
<td>15.2</td>
<td>8</td>
<td>275.8</td>
<td>180</td>
<td>3.07</td>
<td>3.780</td>
<td>18.00</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Cadillac Fleetwood</td>
<td>10.4</td>
<td>8</td>
<td>472.0</td>
<td>205</td>
<td>2.93</td>
<td>5.250</td>
<td>17.98</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Lincoln Continental</td>
<td>10.4</td>
<td>8</td>
<td>460.0</td>
<td>215</td>
<td>3.00</td>
<td>5.424</td>
<td>17.82</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Chrysler Imperial</td>
<td>14.7</td>
<td>8</td>
<td>440.0</td>
<td>230</td>
<td>3.23</td>
<td>5.345</td>
<td>17.42</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Fiat 128</td>
<td>32.4</td>
<td>4</td>
<td>78.7</td>
<td>66</td>
<td>4.08</td>
<td>2.200</td>
<td>19.47</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Honda Civic</td>
<td>30.4</td>
<td>4</td>
<td>75.7</td>
<td>52</td>
<td>4.93</td>
<td>1.615</td>
<td>18.52</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Toyota Corolla</td>
<td>33.9</td>
<td>4</td>
<td>71.1</td>
<td>65</td>
<td>4.22</td>
<td>1.835</td>
<td>19.90</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Toyota Corona</td>
<td>21.5</td>
<td>4</td>
<td>120.1</td>
<td>97</td>
<td>3.70</td>
<td>2.465</td>
<td>20.01</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Dodge Challenger</td>
<td>15.5</td>
<td>8</td>
<td>318.0</td>
<td>150</td>
<td>2.76</td>
<td>3.520</td>
<td>16.87</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>AMC Javelin</td>
<td>15.2</td>
<td>8</td>
<td>304.0</td>
<td>150</td>
<td>3.15</td>
<td>3.435</td>
<td>17.30</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Camaro Z28</td>
<td>13.3</td>
<td>8</td>
<td>350.0</td>
<td>245</td>
<td>3.73</td>
<td>3.840</td>
<td>15.41</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Pontiac Firebird</td>
<td>19.2</td>
<td>8</td>
<td>400.0</td>
<td>175</td>
<td>3.08</td>
<td>3.845</td>
<td>17.05</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Fiat X1-9</td>
<td>27.3</td>
<td>4</td>
<td>79.0</td>
<td>66</td>
<td>4.08</td>
<td>1.935</td>
<td>18.90</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Porsche 914-2</td>
<td>26.0</td>
<td>4</td>
<td>120.3</td>
<td>91</td>
<td>4.43</td>
<td>2.140</td>
<td>16.70</td>
<td>0</td>
<td>1</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>Lotus Europa</td>
<td>30.4</td>
<td>4</td>
<td>95.1</td>
<td>113</td>
<td>3.77</td>
<td>1.513</td>
<td>16.90</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>Ford Pantera L</td>
<td>15.8</td>
<td>8</td>
<td>351.0</td>
<td>264</td>
<td>4.22</td>
<td>3.170</td>
<td>14.50</td>
<td>0</td>
<td>1</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Ferrari Dino</td>
<td>19.7</td>
<td>6</td>
<td>145.0</td>
<td>175</td>
<td>3.62</td>
<td>2.770</td>
<td>15.50</td>
<td>0</td>
<td>1</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>Maserati Bora</td>
<td>15.0</td>
<td>8</td>
<td>301.0</td>
<td>335</td>
<td>3.54</td>
<td>3.570</td>
<td>14.60</td>
<td>0</td>
<td>1</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>Volvo 142E</td>
<td>21.4</td>
<td>4</td>
<td>121.0</td>
<td>109</td>
<td>4.11</td>
<td>2.780</td>
<td>18.60</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 7.1: R’s mtcars dataset.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Characteristic</th>
</tr>
</thead>
<tbody>
<tr>
<td>mpg</td>
<td>height of face</td>
</tr>
<tr>
<td>cyl</td>
<td>width of face</td>
</tr>
<tr>
<td>disp</td>
<td>structure of face</td>
</tr>
<tr>
<td>hp</td>
<td>height of mouth</td>
</tr>
<tr>
<td>drat</td>
<td>width of mouth</td>
</tr>
<tr>
<td>wt</td>
<td>smiling</td>
</tr>
<tr>
<td>qsec</td>
<td>height of eyes</td>
</tr>
<tr>
<td>vs</td>
<td>width of eyes</td>
</tr>
<tr>
<td>am</td>
<td>height of hair</td>
</tr>
<tr>
<td>gear</td>
<td>width of hair</td>
</tr>
<tr>
<td>carb</td>
<td>style of hair</td>
</tr>
<tr>
<td>mpg</td>
<td>height of nose</td>
</tr>
<tr>
<td>cyl</td>
<td>width of nose</td>
</tr>
<tr>
<td>disp</td>
<td>width of ear</td>
</tr>
<tr>
<td>hp</td>
<td>height of ear</td>
</tr>
</tbody>
</table>

Table 7.2: The mapping of variables in the mtcars dataset to facial characteristics.
7.4 Parallel Coordinates

Parallel Coordinates Plot

• Equidistant parallel vertical lines represent the axes of a multidimensional space.

• One vertical line for each dimension.

• Each object is plotted as a polyline defined by values along each dimension.

• Objects which are very similar will have polylines which follow each other.

• Figure 7.6 shows plot of 11 data points (students) on six dimensions (FirstName, Quiz1, Quiz2, Homework1, Homework2, Final).

• Figure 7.7 shows InfoScope [Brodbeck and Girardin 2003; Girardin and Brodbeck 2001; MacroFocus 2015], which combines a parallel coordinates view with a geographic map and a similarity map for exploration of multidimensional data about 73 cities. [Video: https://youtu.be/fnyKj8r0CM4?t=93m27s]

History of Parallel Coordinates

• The first known use was by Henry Gannett in 1883 [Hewes and Gannett 1883, Plate 151] used to visualise 10 attributes for each the (then 47) states of the USA. [It is shown in Figure 3.11.]

• Maurice d’Ocagne discussed the theoretical foundations of parallel coordinates in his book in 1885 [d’Ocagne 1885].

• The technique was independently re-invented by Alfred Inselberg in 1959 [Inselberg 2004]
Figure 7.6: Parallel Coordinates. The six vertical lines represent (from left to right) the name and marks of students in five exams. The eleven polylines represent the data from eleven students. Polylines which mirror one another closely from vertical lines 2 to 6 indicate students achieving very similar marks.

Figure 7.7: InfoScope [Brodbeck and Girardin 2003] allows users to explore multidimensional data about 73 cities. The parallel coordinates view at the bottom is complemented with geographic map and a similarity map. [Screenshot made by Keith Andrews with the InfoScope software [MacroFocus 2015].]
Inselberg published a comprehensive overview paper in the journal The Visual Computer in 1985 [Inselberg 1985] and has published widely on parallel coordinates since.

Inselberg published his definitive book on the subject in 2009 [Inselberg 2009].

[Alfred Inselberg passed away on 30 Dec 2019.]

References


++ Fletcher W. Hewes and Henry Gannett; Scriber’s Statistical Atlas of the United States; Charles Scriber’s Sons, Broadway, New York, 1883. [Hewes and Gannett 1883]

++ Maurice d’Ocagne; Coordonnées Parallèles et Axiales; Gauthier-Villars, Paris, 1885. [d’Ocagne 1885]

7.5 Star Plot

• Also known as star chart, radar chart, spider plot, wheel chart, and polar chart.

• Dimensions are laid out radially.

• Each record is represented as an irregular polygon.

• The visual image is highly dependent on the chosen ordering of dimensions.

• Sometimes lines are drawn on top of one another (occlusion).

Figure 7.8: A star plot of the classic cereals dataset. [Image created by Keith Andrews using Liquid Diagrams.]
7.6 Interactive Attraction

7.6.1 Dust and Magnet

- Metaphor of dust particles attracted to magnets, as shown in Figure 7.9.
- Magnets can be added and removed, and their strength increased or decreased.
- Paper in IV journal 2005 [Yi et al. 2005b].
- More recently re-implemented for touch interface. [Video: https://www.cc.gatech.edu/gvu/ii/dnm/dustnmagnet-16Hub.mp4]

7.7 Interactive Histograms

7.7.1 Attribute Explorer

- Imperial College, 1993.
- Direct manipulation of coupled views of histograms, as shown in Figure 7.10.
- Each attribute is assigned to a scale with histograms showing the population spread.
Figure 7.10: The Attribute Explorer. Terraced houses have been selected (red) together with a garden length of 120–140 feet (green). Houses fulfilling both criteria are shown in black. [Copyright © by the Association for Computing Machinery, Inc.]

- If there are few enough items, individual icons are shown.
- Sliders are provided for continuous attributes (e.g. price) and buttons for discrete attributes (e.g. type of house).
- The effect of one attribute on the others can be explored by selecting values of interest on one scale and viewing where those items appear on the other attribute scales.
- Figure 7.11 shows a more recent implementation of an Attribute Explorer in JavaScript.
Figure 7.11: A more recent implementation in JavaScript of an Attribute Explorer showing the classic cars dataset. Cars in the Miles per Gallon range of 35–40 have been selected. It is easy to see how these cars are distributed across the other dimensions. Clicking a single car icon highlights its position in the other dimensions. [Screenshot made by Keith Andrews.]

### 7.8 Circular Histograms

#### Daisy Chart

- Daisy Analysis, UK, 2003 [Daisy 2003].
- Attributes and (ranges of) their values are arranged around the perimeter of a circle. A polygon of connecting lines represents an individual item. See Figure 7.12.
Figure 7.12: The Daisy Chart maps attributes and ranges of their values to positions on the circumference of a circle. Here, four attributes (Day of the Week, Hour of the Day, Call Duration, and Exchange) are given space on the perimeter. Each item (phone call) is represented by a polygon touching each attribute once. In each mini-histogram, green indicates the number of calls, and cyan the total call time.
Chapter 8

Visualising Text and Object Collections (Feature Spaces)

References


8.1 Distance-Based Projection

Distance Calculation

Calculate the similarity (or dissimilarity) between every pair of objects in nD.

Techniques include:

• Vector space model and distance metric (such as scalar product).

• Normalised compression distance based on Kolmogorov complexity (NCD) [Telles et al. 2007].

• Distances are often normalised to values between 0 and 1.

• Results in a symmetric matrix of distance values.

Multidimensional Projection

• Each object is represented by a vector in nD space.

• Objects are mapped directly to positions in 1D, 2D, or 3D space.

• Preserving (as far as possible) the distance relationships from the high-dimensional space in the target space.

• Typically by minimising a stress function.
Figure 8.1: The visualisation pipeline for distance-based projection. Here, the vector space model has been chosen as the distance calculation algorithm and force-directed placement has been chosen as the projection algorithm.
8.1. DISTANCE-BASED PROJECTION

8.1.1 Linear Projection

- Input is nD vector space.
- Can be directly calculated (no need for iterative process).
- Each embedding axis is a linear combination of the original axes.
- Creates meaningful axes which can be interpreted (given a “name”).
- Straightforward mapping of new objects.
- Low computational complexity.

Linear Projection Techniques

- Principal Component Analysis (PCA):
  - Covariance matrix is decomposed into $m$ eigenvectors, the first $p$ with largest eigenvalues are chosen.
  - The first principal component accounts for as much of the variability in the data as possible.
  - Each succeeding component accounts for as much of the remaining variability as possible.
  - For a mapping to 2D, choose the first 2 principle components.

8.1.2 Non-Linear Projection

- Input is set (triangular matrix) of pairwise similarities (or dissimilarities).
- Similarity matrix can, of course, be calculated from an nD vector space.
- Usually needs an iterative process (cannot be directly calculated).
- Optimise a cost (stress) function.
- Change in objects means need to run a few more iterations (incremental layout).
- Can handle non-linear structures.
- New axes cannot be interpreted (given a “name”).

Non-Linear Projection Techniques

- Multi-Dimensional Scaling (MDS)
  - Majorisation: iterative nonlinear optimisation based on steepest descent towards a (local) minimum.
- Force-directed placement (FDP): Iterative solution of a force model.
- Fastmap [Faloutsos and Lin 1995].
- Nearest Neighbor Projection (NNP) [Tejada et al. 2003].
- Least Squares Projection (LSP) [Paulovich et al. 2008].
- SNE Stochastic Neighbor Embedding.
• t-SNE [van der Maaten and Hinton 2008; Wattenberg et al. 2016]. Stochastic Neighbor Embedding with a t-distribution.

8.2 Force-Directed Placement (FDP)

• Invented in 1984 by Peter Eades [Eades 1984]: spring model, forces move the system to a minimal energy state. Brute force, $O(n^3)$.

• Improved and named force-directed placement in 1991 by Fruchtermann and Reingold [Fruchtermann and Reingold 1991]. Forces are computed only to nearby objects (within a certain radius). Attempts to achieve uniform edge length.

• Series of improvements by Chalmers: 1992 $O(n^{2\sqrt{n}})$ [Chalmers and Chitson 1992], 1996 $O(n^2)$ [Chalmers 1996a], 2003 $O(n^{3/4})$ [Morrison et al. 2003].

• Jourdan and Melancon, MultiscaleMDS, in 2004 $O(n \log n)$ [Jourdan and Melancon 2004].

• Brandes and Pich; PivotMDS, in 2006 $O(n)$ through sampling and approximation [Brandes and Pich 2006].

• Ingram, Munzner, and Olano; Glimmer, in 2009 theoretically $O(n^2)$, but massively parallel [Ingram et al. 2009].

8.3 Example Systems

8.3.1 Bead

• Matthew Chalmers (EuroPARC, Ubilab), 1992 [Chalmers and Chitson 1992; Chalmers 1993; Chalmers 1996b].

• Vector space model and force-directed placement.

• Produces a 3d point cloud.

8.3.2 SPIRE


• Build vector space model from text (or other document) corpus.

• Anchored Least Stress (ALS): first project small subset of objects using PCA (first two principle components), then interpolate remaining objects.

• Results in 2d Galaxy View.

• From Galaxy View aggregate of keywords in height dimension to form Themescape. See Figure 8.3.

• Paper in ISKO [Hetzler et al. 1998], technical details paper in JASIS [Wise et al. 1995], good overview at I-Know ’01 [Thomas et al. 2001].

• Video at InfoVis 2004 Contest http://www.cs.umd.edu/hcil/InfovisRepository/contest-2004/

• Tutorial video https://youtu.be/YONTBZaxz8g.
Figure 8.2: Part of the Bead visualisation of 301 entries from the HCI Bibliography. The objects represent articles from the field of HCI. Documents containing similar keywords are placed near each other in the 3d point cloud. A search has been done on the keywords “information retrieval” and the results are highlighted. [Copyright © by the Association for Computing Machinery, Inc.]

Figure 8.3: SPIRE showing the Galaxy View (below), Themescape (upper right) and Probe Tool. [Image used with kind permission of Pacific Northwest National Laboratory.]
**Figure 8.4:** VxInsight showing some of 1,231 bibliographic records extracted from the physical sciences portion of the Science Citation Index Expanded. The layout uses similarities based on direct and cocitation links between articles. [Thanks to Brian Wylie, Sandia National Laboratories, for providing a demo version of VxInsight.]

### 8.3.3 VxInsight

- Sandia National Labs, 1998 [Davidson et al. 1998].
- VxOrd: force-directed placement.
- Accepts list of pre-computed similarities.
- Nodes are moved to minimise an energy function.

### 8.3.4 VisIslands

- IICM, 2001 [Andrews et al. 2001].
- First (as far as we know) interactive FDP (a few seconds).
- Build vector space from objects in search result set.
- Initially cluster objects using fast algorithm.
- Position cluster centroids using FDP.
- Place other cluster members around centroid, then run a few iterations of FDP.
8.3.5 InfoSky

- IICM + Know-Center + Hyperwave, 2002 [Andrews et al. 2002].
- InfoSky assumes objects are pre-placed within a hierarchy.
- Force-directed placement is not done globally, but recursively at each level of the hierarchy (only for the nodes at that level).
- First system to use recursive Voronoi subdivision of a hierarchy (now known as Voronoi Treemaps).
8.4 Self-Organizing Maps (SOM)

- Self-organizing map (SOM) invented by Kohonen [2000].
- Based on neural networks.
- The map consists of a regular grid of cells (“neurons”).
- The cells may be rectangular (like a shelf of post boxes), hexagonal (like a wine rack), or other regular shapes.
- A feature vector (descriptor) describes each cell.
- Each object is represented by a feature vector.
- Cell descriptors are usually initialised using a training set.
- An object is assigned to its closest cell. The feature vectors of that cell and neighbouring cells are then updated to reflect the new object.

8.4.1 SOMLib

- Based on a variant of the SOM algorithm [Rauber and Merkl 1999].

8.4.2 WEBSOM

- Self-organizing map (SOM) algorithm [Kohonen 2000].
Figure 8.7: The SOMLib system with the libViewer interface. Documents are assigned to sections of a bookshelf (or post boxes) according to their content. [Image used with kind permission of Andreas Rauber.]

- Papers in IEEE Transactions on Neural Networks [Kohonen et al. 2000] and Information Sciences [Lagus et al. 2004]
Figure 8.8: WEBSOM.
Chapter 9

Other Kinds of Visualisation

9.1 Visualising Query Spaces

9.1.1 InfoCrystal

• Anselm Spoerri, MIT, 1993.

• Builds upon the idea of a Venn diagram to indicate which sets of items match each permutation of a boolean query, as shown in Figure 9.1.

• $n$ boolean query terms are placed at the corners of an $n$-sided regular polygon.

• Icons representing items matching each permutation of a boolean query are positioned within the polygon.

• Figure 9.2 shows an InfoCrystal with four query terms.

• Papers at Vis’93 [Spoerri 1993a], CIKM’93 [Spoerri 1993b], and VL’93 [Spoerri 1993c], as well as PhD thesis [Spoerri 1995].

9.1.2 MetaCrystal

• Application of InfoCrystal idea to search results from multiple search engines [Spoerri 2004a; Spoerri 2004b].

• See Figure 9.3.

9.1.3 LyberWorld

• Matthias Hemmje, GMD-IPSI, 1993.

• LyberWorld’s Navigation Cone helps an analyst to progressively narrow down a result set of matching documents.

• Terms and documents are selected at alternate levels within a cone tree. [Video: https://youtu.be/uGBW7uPKyCM?t=02m03s]
CHAPTER 9. OTHER KINDS OF VISUALISATION

Figure 9.1: InfoCrystal builds upon the idea of a Venn diagram to indicate which sets of items match each permutation of a boolean query. [Image used with kind permission of Anselm Spoerri.]

Figure 9.2: InfoCrystal with four query terms. The numbers indicate how many document match the corresponding boolean query. [Image used with kind permission of Anselm Spoerri.]
Figure 9.3: MetaCrystal combines the results from multiple search engines in response to the same query. [Image used with kind permission of Anselm Spoerri.]

- LyberWorld’s Relevance Sphere allows the analyst to place search terms around the outside of a sphere and documents inside the sphere are attracted to them (similar to Dust and Magnets). [Video: https://youtu.be/uGBW7uPKyCM?t=05m30s]


9.2 Internal Document Visualisation (Content Analysis)

9.2.1 TileBars

- Visualisation of distribution of search terms within matching documents in a search result list.
- The structure of longer text documents is analysed, and the document is broken down into topical units.
- Each topical unit is a contiguous block of say a few paragraphs discussing the same themes.
- For each search term, a row of tiles indicates how frequently that term occurs in each topical unit (dark = very frequent). See Figure 9.4.

- Paper at CHI’95 [Hearst 1995], video at CHI’96 [Hearst and Pedersen 1996].
Figure 9.4: TileBars visualisation of the distribution of three search terms “osteoporosis”, “prevention”, and “research” within matching documents. The top-ranked document has some discussion of all three terms in the middle of the document. The fifth, sixth, and seventh matching documents all discuss research but have no mention of osteoporosis. [Image used with kind permission of Marti Hearst.]
Bibliography


BIBLIOGRAPHY


Paulovich, Fernando V., Luis G. Nonato, Rosane Minghim, and Haim Levkowitz [2008]. *Least Square Projection: A Fast High-Precision Multidimensional Projection Technique and Its Application to*


