

Introduction to Information Visualization

Riccardo Mazza

Introduction to Information Visualization

 Springer

Riccardo Mazza
University of Lugano
Switzerland

ISBN: 978-1-84800-218-0 e-ISBN: 978-1-84800-219-7
DOI: 10.1007/978-1-84800-219-7

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Control Number: 2008942431

© Springer-Verlag London Limited 2009

Apart from any fair dealing for the purposes of research or private study, or criticism or review, as permitted under the Copyright, Designs and Patents Act 1988, this publication may only be reproduced, stored or transmitted, in any form or by any means, with the prior permission in writing of the publishers, or in the case of reprographic reproduction in accordance with the terms of licences issued by the Copyright Licensing Agency. Enquiries concerning reproduction outside those terms should be sent to the publishers.

The use of registered names, trademarks, etc., in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant laws and regulations and therefore free for general use.

The publisher makes no representation, express or implied, with regard to the accuracy of the information contained in this book and cannot accept any legal responsibility or liability for any errors or omissions that may be made.

Printed on acid-free paper

Springer Science+Business Media
springer.com

To Vincenzo and Giulia

Preface

Imagine having to make a car journey. Perhaps you're going to a holiday resort that you're not familiar with. Your wife has been really eager to go. Spending the holidays in an appealing destination will certainly make her happy and it would go a long way toward rewarding her for the daily struggle of keeping the children under control, and even they could do with a change of scenery every now and again. Luckily, you have your trusty laptop at hand. You connect to the Internet (and who doesn't have an Internet connection these days) and log on to a well-known website that will show you exactly which route you should drive to reach your longed-for destination. The site is extremely efficient: Not only does it show, in meticulous detail, the course to follow (even the crossroads are accurately indicated), but it even provides you with a detailed map highlighting the advised route.

And you? Will you have to surrender your hobby for two weeks: golf? Never. Consulting the map provided by the website, you discover that, on the way to the highly anticipated getaway location, you will pass just 20 kilometers from an area that boasts one of the finest golf courses. But that's not all—you zoom into the area and find out that there are some interesting tourist spots that hadn't come up before due to the overly large scale of the map. You learn that there is a 19th-century castle nearby that offers relaxing weekends with beauty and spa treatments and provides a babysitting service to take care of your children all day, entertaining them with games, songs, and a variety of amusements. It's done: You've planned your journey and, thanks to the opportunity offered by the website, you have had the chance to please your family, without having to forgo your hobby.

All of this thanks to a website? Of course not. All the site offered you was a detailed map that featured, in addition to the route leading to your destination, a selection of information on the location, attractions, and places of interest in the area. These places of interest aren't necessarily found precisely along the route indicated by the ultra-efficient website: They could indeed be a short distance from your path but capture your interest nonetheless, as seen in the case presented here. But at what distance? And what sort of location might I be interested in? The situation just described illustrates a typical case in which the use of a graphical representation has taken on a decisive role in the discovery of new information. How can I go

about representing a series of various types of information that is complex by nature, linked together by a relationship of “proximity” that may be very vague and imprecise? (When are two locations deemed near? Are two locations near when they’re 20 kilometers from each other? And if they were 60 kilometers from each other, would they still be considered close?). How can I communicate to the users of my website which of the “nearby” locations might be of some artistic or architectural interest? And what if they are just interested in finding out which route to take?

In these circumstances, the use of graphical representations mediated by the computer can help in the analysis of and search for “imprecise” information. Imprecision is not in the nature of data (which in itself is precise) but rather by the type of search and interest that a generic user may not have made very clear beforehand. The “let’s look at the graphics and then see what to do” situation is one of the *modus operandi* in which the graphical representation of information is at its best, as portrayed in the previous case. We don’t know what to look for; therefore, we try to represent everything we possibly can, to then examine the information and come to a decision (“Along the way you pass close to a golf course? Then we can spend the weekend there!”), demonstrate a hypothesis (“The alternative path the website suggested indeed shorter in terms of kilometers, but requires crossing a mountain pass at an altitude of more than 1,000 meters. Therefore, the traveling time required is greater”), or even communicate an idea effectively (“See? If we stop here, we’ll be at the halfway mark. We’ll let the kids rest for a day and then set off again much more relaxed”).

This book illustrates such concepts in a simple and thorough manner. It aims to build a reference for the situations in which the graphical representation of information, generated and assisted by computer, can be helpful in carrying out explorative analysis on the data, effectively communicating ideas, data, or concepts, and helping to demonstrate or disprove a hypothesis on data.

Created as a support text for a university course, this book is also suitable for a wide and heterogeneous reading audience. It contains suggestions for setting communication systems based on or availing of graphical representations. The text will, above all, illustrate cases, situations, tools, and methods that can help make the graphical representation of information effective and efficient.

Lugano, October 2008

Riccardo Mazza

Acknowledgments

I would like to thank all the people who have contributed to this book in the various phases of its development, in particular, Enrico Bertini, Giuseppe Santucci, Daniele Galiffa, and Francesca Morganti for their helpful comments and suggestions. Special thanks to Lynsey Dreaper for helping with the translation from the first version in Italian. I would also like to thank all the authors and publishers of the copyrighted materials included in this book for their kind permission to reproduce the images.

Finally, I would like to thank my wife, Beatrice, for her support, encouragement, and understanding during the many late nights spent working on this book.

Contents

1	Introduction to Visual Representations	1
1.1	Presentation	4
1.2	Explorative Analysis	5
1.3	Confirmative Analysis	6
1.4	Information Visualization	7
1.5	From Data to Wisdom	8
1.6	Mental Models	10
1.7	Scientific Visualization	11
1.8	Criteria for Good Visual Representations	12
1.8.1	Graphical Excellence	13
1.8.2	Graphical Integrity	13
1.8.3	Maximize the Data–Ink Ratio	14
1.8.4	Aesthetics	14
1.9	Conclusion	15
2	Creating Visual Representations	17
2.1	A Reference Model	17
2.1.1	Preprocessing and Data Transformations	18
2.1.2	Visual Mapping	20
2.1.3	Views	23
2.2	Designing a Visual Application	24
2.3	Visual Representation of Linear Data	26
2.4	2D vs. 3D	30
2.5	Conclusion	31
3	Perception	33
3.1	Memory	33
3.2	Preattentive Properties	35
3.2.1	Color	35
3.2.2	Form	35
3.2.3	Spatial Position	35

- 3.2.4 Movement 38
- 3.3 Mapping Data to Preattentive Attributes 38
- 3.4 Postattentive Processing 39
- 3.5 Gestalt Principles 41
 - 3.5.1 Figure and Ground 42
 - 3.5.2 Proximity 42
 - 3.5.3 Similarity 42
 - 3.5.4 Closure 43
 - 3.5.5 Continuity 43
 - 3.5.6 Other Principles 44
- 3.6 Conclusion 44
- 4 Multivariate Analysis 45**
 - 4.1 The Problem of Multivariate Visualization 46
 - 4.2 Geometric Techniques 47
 - 4.2.1 Parallel Coordinates 47
 - 4.2.2 Scatterplot Matrix 50
 - 4.2.3 TableLens 52
 - 4.2.4 Parallel Sets 52
 - 4.3 Icon Techniques 55
 - 4.3.1 Star Plots 55
 - 4.3.2 Chernoff Faces 57
 - 4.4 Pixel-Oriented Techniques 59
 - 4.5 Conclusion 62
- 5 Networks and Hierarchies 63**
 - 5.1 Network Data 64
 - 5.1.1 Concept Maps and Mind Maps 65
 - 5.1.2 Complex Network Data 66
 - 5.1.3 Geographic Representations 72
 - 5.1.4 Transport Networks 73
 - 5.1.5 3D Graphs 74
 - 5.2 Hierarchical Data 75
 - 5.2.1 File System 77
 - 5.2.2 Representing Evolutionary Data with Trees 79
 - 5.2.3 Cone Tree 81
 - 5.2.4 Botanical Tree 81
 - 5.2.5 Treemap 83
 - 5.3 Conclusion 88
- 6 World Wide Web 91**
 - 6.1 Website Maps 91
 - 6.2 Website Log Data 93
 - 6.3 Visual Representation of Search Engine Results 98
 - 6.3.1 Clustering 100

- 6.4 Analysis of Interactions in Blogs 102
- 6.5 Conclusion 103
- 7 Interactions 105**
 - 7.1 The Problem of Information Overload 105
 - 7.2 Types of Interactive Visual Representations 106
 - 7.3 Manipulable Representations 107
 - 7.3.1 Scrolling 107
 - 7.3.2 Overview + Details 107
 - 7.3.3 Focus + Context 110
 - 7.4 Transformable Representations 116
 - 7.4.1 Filtering Input Data 116
 - 7.4.2 Data Reordering 117
 - 7.4.3 Dynamic Queries 117
 - 7.4.4 Magic Lens 119
 - 7.4.5 Attribute Explorer 121
 - 7.5 Conclusion 122
- 8 Evaluations 125**
 - 8.1 Human–Computer Interaction 125
 - 8.2 Evaluation Criteria 126
 - 8.3 Evaluating Visual Representations 127
 - 8.3.1 Analytic Methods 128
 - 8.3.2 Empirical Methods 129
 - 8.4 Conclusion 132
- References 133**
- Index 137**

List of Figures

1.1	Directions for the Lugano–Pisa route, provided in textual and in visual versions.	2
1.2	Mapping numerical values to the lengths of bars.	3
1.3	Visual representation of the march of Napoleon’s army in the Russian campaign of 1812, produced by Charles J. Minard.	5
1.4	A map of the United States showing the number of cancer-related deaths in the male population from 1970 to 1994, subdivided into counties.	6
1.5	A picture that compares the Italian MIBTEL stock market share index to the U.S. Dow Jones shares index.	7
1.6	Example of a visual representation that explains how to calculate the cathetus of a right triangle.	8
1.7	The continuum of understanding, according to Nathan Shedroff.	9
1.8	Two examples of scientific visualization.	12
2.1	The process of generating a graphical representation.	18
2.2	Examples of graphical elements.	21
2.3	Examples of graphical properties.	21
2.4	Accuracy in the perception of quantitative values for some graphical and spatial elements.	22
2.5	Use of primary colors to define categorical attributes and of color scales to define ordinal attributes.	23
2.6	A possible visual representation of the data collected from a discussion forum.	24
2.7	The GNP of various nations visually represented by a single-axis scatterplot and a bar chart.	27
2.8	Two-dimensional scatterplot that compares import and export values.	28
2.9	A three-dimensional scatterplot.	29
2.10	A two-dimensional scatterplot with GNP mapped to the area of the circle and to a color scale.	30

2.11	Visual representation of the topology of the Internet network MBone on a terrestrial scale.	31
3.1	Example showing how the hue and intensity of colors are processed preattentively	36
3.2	Preattentive attributes of form.	37
3.3	Preattentive attributes of concavity/convexity.	37
3.4	Combinations of more preattentive attributes.	39
3.5	Gestalt composition of elements.	41
3.6	Gestalt proximity example.	42
3.7	Gestalt similarity example	43
3.8	Gestalt closure example.	43
3.9	Gestalt continuity example.	44
4.1	Scatterplot representing four-variate attributes.	47
4.2	Parallel coordinates representation.	48
4.3	Parallel coordinates for 730 elements with 7 variant attributes.	49
4.4	Brushing of the values on the coordinate parallel axes.	50
4.5	Scatterplot matrix.	51
4.6	TableLens representation.	53
4.7	TableLens representation with sorted values.	53
4.8	Representation of the datasets of the Titanic disaster using parallel sets.	54
4.9	Star plot.	56
4.10	Star plot of the annual climatic data of some cities.	57
4.11	Chernoff face.	58
4.12	Climatic data of some cities represented by Chernoff faces.	58
4.13	Example of an application that uses the pixel as a basic unit.	60
4.14	Market Radar.	61
5.1	Examples of graphs.	64
5.2	Example of a concept map that describes concepts regarding a website.	65
5.3	Example of mind map for the planning of a university course.	66
5.4	An example of a complex graph.	67
5.5	An example of a graph that uses a force-directed algorithm to represent a social network.	69
5.6	Example of a reduced graph using the pathfinder technique.	70
5.7	Main NicheWorks layouts.	70
5.8	NicheWorks shows the graph of the <i>Chicago Tribune's</i> website structure.	71
5.9	An image that shows the topology of the NSFNET T1 backbone network in September 1991 in the United States.	73
5.10	Seenet, 3D representation of the Internet traffic across the NSFNET backbone network in 1993.	74

5.11	A map of the Madrid Metro system.	75
5.12	Examples of graphs with three-dimensional layouts generated by Tulip software.	76
5.13	A tree representing the classification of the wind instruments according to Curt Sachs and Erich von Hornbostel.	77
5.14	Interface of Finder tool in the Mac OS X system.	78
5.15	Xdiskusage.	79
5.16	Fsv in the TreeV version visualizes the organization of the file system.	80
5.17	Visual representation of a <i>stemma codicum</i> of a text.	80
5.18	Representation of hierarchical structure by a cone tree.	82
5.19	Botanical tree.	83
5.20	Example of a Treemap construction.	84
5.21	Treemap.	85
5.22	Newsmap uses a treemap algorithm to represent news from Google News U.S.	86
5.23	Smartmoney.com Map of the Market.	87
5.24	Map of the Market shows the details for the companies that deal with computer hardware.	87
5.25	Squarified cushion treemap in SequoiaView showing the whole content of the file system.	89
6.1	Map of the structure of the Nature Neuroscience website realized by Dynamic Diagrams.	93
6.2	Site map of the Dynamic Diagrams website.	94
6.3	Content map of an online course.	94
6.4	Google Analytics.	95
6.5	Website statistics through an overlay representation.	96
6.6	Representation of the navigation of a website through a graph.	97
6.7	INSYDER graphical interface.	99
6.8	Results of the web search using KartOO.	101
6.9	Selection of the www.lugano.ch website in the KartOO search interface.	101
6.10	Blogviz graphical interface.	102
7.1	Example of the application of the Overview + Details technique.	108
7.2	Map of the characters used by Microsoft Windows to assist the user in the choice of a special character.	108
7.3	Tourist map of Florence.	109
7.4	PhotoMesa	110
7.5	Bifocal display.	111
7.6	Perspective wall.	112
7.7	Fisheye distorted treemap.	113
7.8	Example of a fisheye view menu.	114
7.9	Example of a hyperbolic browser.	115
7.10	SpaceTree.	116

7.11 A collective view of all of the available films in FilmFinder. 119

7.12 Films listed after the filtering operation in FilmFinder. 120

7.13 Details of the selected film in FilmFinder. 120

7.14 Magic lens. 121

7.15 Attribute explorer. 123