



Swansea University
Prifysgol Abertawe



Cronfa - Swansea University Open Access Repository

This is an author produced version of a paper published in:
Computer Graphics Forum

Cronfa URL for this paper:
<http://cronfa.swan.ac.uk/Record/cronfa48247>

Paper:

Rees, D. & Laramee, R. (2019). A Survey of Information Visualization Books. *Computer Graphics Forum*
<http://dx.doi.org/10.1111/cgf.13595>

This item is brought to you by Swansea University. Any person downloading material is agreeing to abide by the terms of the repository licence. Copies of full text items may be used or reproduced in any format or medium, without prior permission for personal research or study, educational or non-commercial purposes only. The copyright for any work remains with the original author unless otherwise specified. The full-text must not be sold in any format or medium without the formal permission of the copyright holder.

Permission for multiple reproductions should be obtained from the original author.

Authors are personally responsible for adhering to copyright and publisher restrictions when uploading content to the repository.

<http://www.swansea.ac.uk/library/researchsupport/ris-support/>

A Survey of Information Visualization Books

D. Rees and R. S. Laramée

Swansea University, UK

Abstract

Information visualization is a rapidly evolving field with a growing volume of scientific literature and texts continually published. To keep abreast of the latest developments in the domain, survey papers and state-of-the-art reviews provide valuable tools for managing the large quantity of scientific literature. Recently a survey of survey papers (SoS) was published to keep track of the quantity of refereed survey papers in information visualization conferences and journals. However no such resources exist to inform readers of the large volume of books being published on the subject, leaving the possibility of valuable knowledge being overlooked. We present the first literature survey of information visualization books that addresses this challenge by surveying the large volume of books on the topic of information visualization and visual analytics. This unique survey addresses some special challenges associated with collections of books (as opposed to research papers) including searching, browsing and cost. This paper features a novel two-level classification based on both books and chapter topics examined in each book, enabling the reader to quickly identify to what depth a topic of interest is covered within a particular book. Readers can use this survey to identify the most relevant book for their needs amongst a quickly expanding collection. In indexing the landscape of information visualization books, this survey provides a valuable resource to both experienced researchers and newcomers in the data visualization discipline.

CCS Concepts

•General and reference → Surveys and overviews; General literature; •Human-centered computing → Information visualization;

"The book you don't read won't help." – Jim Rohn - entrepreneur, author, motivational speaker.

"You cannot open a book without learning something." – Confucius - teacher, politician, philosopher.

1. Introduction and Motivation

Books provide a very valuable resource for education, research and to further understanding in any field. Books cater to many different levels of information visualizers, from beginners to specialists, however, knowing what books are relevant for which topics and competency levels can be very challenging due to the rapidly expanding landscape of information visualization books.

One of the first books to discuss the use of information visualization is Jacques Bertin's *Sémiologie Graphique* [BBS67] published in 1967. Few books were published in the following 40 years, however in the last decade there has been a dramatic increase of published books, as evident in Figure 1. In fact there are so many, we cannot survey all of them. They are being published so rapidly that it's very challenging to keep abreast of the volume (similar to research papers). We focus on what we believe is the most important

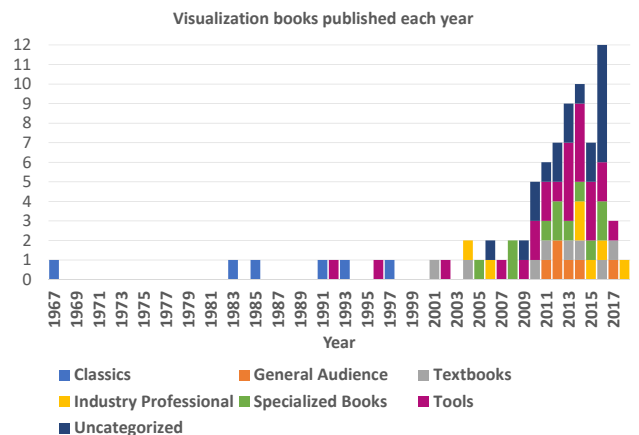


Figure 1: Number of identified information visualization and visual analytics book publications by year.

core subset. See section 1.3 for more details on the scope. Keeping up with the book volume is a difficult task both in time and money, with over 23,000 pages represented in this survey.

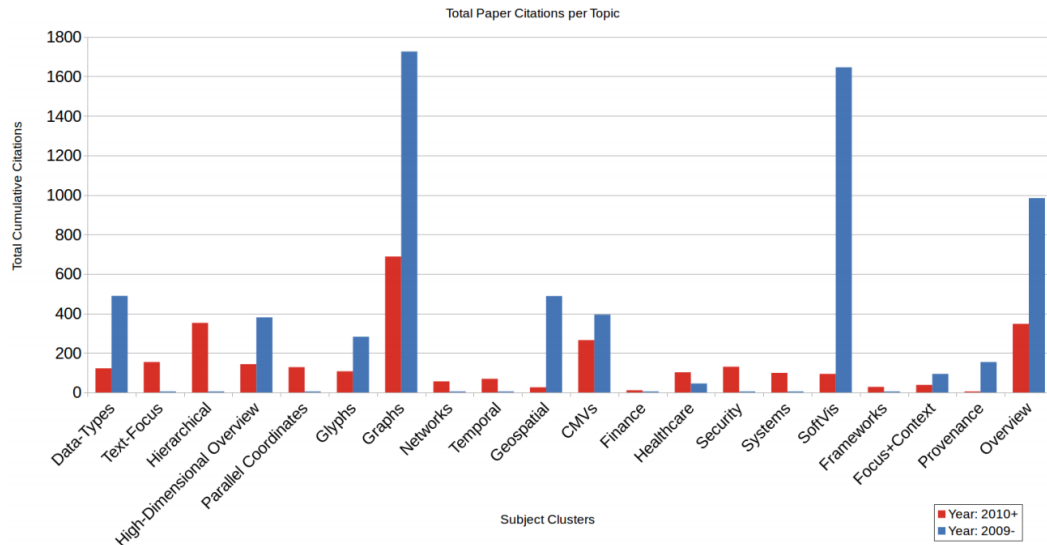


Figure 2: Citations in survey papers, ordered by topic, before and after 2010, courtesy of McNabb and Laramée [ML17], highlighting that the majority of research paper citations predate 2010.

Books are predominantly overlooked in traditional literature surveys, leading to missed opportunities to pass on valuable knowledge which, as scientists, is one of our primary goals. For example, Eduard Imhof published color use guidelines for geo-spatial visualization which are rarely (if ever) used by data visualization researchers [Imh82]. This is unfortunate because the authors of these books invest several years in their writing efforts. As an example, Inselberg’s book on Parallel Coordinates [Ins09], required five years to write (based on personal communication) and consists of 554 pages and over 230 color figures. Inselberg’s book contains much mathematical formulation on parallel coordinates not featured elsewhere in the information visualization conferences and journal papers. Another example is Samet’s book on Foundations of Multidimensional and Metric Data Structures [Sam06] which required over ten years of dedication to complete (based on personal communication). Ten years is not unusual when we consider multiple editions of a book.

Books offer some advantages over research papers in that they feature a longer format, therefore enabling greater detail to be provided. Greater use of figures, along with the use of more descriptive case studies, can also be used to aid understanding of a subject. Books also offer advantages to beginners and non-specialists, the large volume of published literature, with the looser length restrictions, enabling simplifications and background to facilitate understanding. Research papers, in contrast, often assume advanced understanding. Books also tend to be more accessible to people outside the subject who are looking for a starting point or overview of visualization.

Books, however, are at a disadvantage from research papers in that they do not necessarily represent the state of the art within the subject due to the publication time required. However, even this is mitigated with the advent of e-books and short books such as the SpringerBriefs series in computer science. Analysis performed by McNabb and Laramée [ML17] shows that the majority of citations

in survey papers are a number of years old already. See Figure 2. Books also do not undergo a rigorous peer review process, however, they are scrutinized by customer reviews.

This non-traditional literature survey aims to provide a novel overview of the valuable landscape of the most influential information visualization and visual analytics books and to categorize the common topics within the book chapters. The contributions of this unique survey are to provide:

- the first overview of the books on the topic of information visualization and visual analytics,
- a novel two-level categorization of both books and the topics covered within each book,
- valuable meta-data from each book to help the reader compare and navigate this vast and information-rich literature space,
- recommendations for choosing a suitable visualization book.

We believe that a literature survey of this type is very useful for readers both inside and outside the visualization community.

1.1. Survey Challenges

There are a number of challenges involved in performing a literary survey of this nature.

Search: Although online book stores, such as Amazon.com [Ama18], enable fast searching of millions of books, sometimes appropriate books can be lost in irrelevant material when performing a generic search. For example, searching for the term “visualization” on Amazon.com returns 16 books on the first page, with only five dedicated to data visualization. Whereas a visualization research paper in the domain has a limited number of sources, conferences and journals, that can be searched [IHK*17], there are a myriad of publishers capable of publishing appropriate books. Once a book has been identified it can also be difficult to ascertain whether it is appropriate for the reader as summaries are not always avail-

able. The only authoritative method of verification is to obtain a full copy.

Cost: Acquiring the literature for a traditional survey is normally an inexpensive task for an individual researcher, with institutional membership often providing access to online digital libraries. This is, however, not the case with books. Although libraries offer access to a range of books, there is a limit to the number of books that they archive. And, in our experience, due to the specialized nature of visualization books, they are often unavailable. Even if a university library could store all of these books, a library member may only check-out a very limited number simultaneously. Prices for the books contained within this survey range from \$20 to \$102 with the average book priced at \$43, as can be seen in Figure 3. The total cost of the literature within this survey comes to a total of approximately \$3,600 based on prices from Amazon.com.

Volume: Over 35 focus books are included within this survey, representing a vast amount of reading. To make the survey manageable, we include both focus and non-focus books. Further explanation is provided in section 1.3 on scope.

1.2. Book search methodology

This survey is partly inspired by the growing number of books on our own shelves, therefore we began with these. Further references found within books are analyzed and relevant books added to the survey. The online book stores Amazon.com and Amazon.co.uk [Ama18] are searched, using both their search function as well as the "customers who bought this item also bought" recommendations. Publisher websites are also consulted such as Springer Science+Business Media [Spr18] along with use of their search functions. We searched the following sources for books:

Online book stores

- Amazon.com
- Amazon.co.uk
- John Smith
- Waterstones
- Blackwells
- Wordery.com

Publisher websites

- CRC Press
- Springer Science+Business Media
- John Wiley & Sons
- O'Reilly Media

Other resources

- Our local university Library

The following search terms were used:

- Visualization
- Data visualization
- Information Visualization
- Visual analytics
- Visual analysis

We also scanned the references of each book for further related literature.

1.3. Book survey scope

Due to the large volume of published data visualization books, restrictions are applied to constrict the scope of this survey for manageability. The first condition concentrates the survey on information visualization and visual analytics books. This excludes books that primarily focus on scientific visualization. We define scientific visualization as visual design based on data with an intrinsic spatial position such as volume and flow data. Searching for the term 'scientific visualization' in Amazon.com's search returns over 300 books, although some books are repeated due to multiple editions and some results are not suitable. A notable book that falls into the scientific visualization category is Telea's Data Visualization: Practice and Principles [Tel14], and is therefore out of scope.

Another restriction to the scope is infographics, which are excluded due to the large number of books. Infographics are a manually generated graphic representation of information that presents behavior quickly and clearly. These are primarily used in journalistic articles and instruction cards. An Amazon.com search for 'infographics' returns over 600 book results. Example infographic books are *Cool Infographics: Effective Communication with Data Visualization and Design* by Krum [Kru13], *The Wall Street Journal Guide to Information Graphics: The Dos and Don'ts of Presenting Data, Facts, and Figures* by Wong [Won10] and books by McCandless [McC09, McC14]. A particularly excellent example is *Data Flow: Visualising Information in Graphic Design* by Klanten [Kla08] which presents over 200 pages of glossy designs.

Edited volumes are also excluded due to the large volume within this category, with the publisher Springer offering over 50 edited volumes on the subject. Edited volumes are predominantly comprised of a collection of scientific papers which are covered in traditional surveys. This excludes books such as *Beautiful Visualization: Looking at Data Through the Eyes of Experts* compiled by Steel and Iliinsky [SI10] and *Mastering the Information Age: Solving Problems with Visual Analytics* by Keim *et al.* [KKEM10].

Books specializing in geographic information systems (GIS) are also excluded due to the extensive volume of books. Searching for 'Geographic Information Systems' on Amazon.com returns over 4,000 book results, although not all books are relevant and there are some duplicates. Many of these relate to GIS tools such as ArcGIS [esr] and QGIS [QGI]. Cartographic books are considered as GIS and are therefore out of scope. An example cartographic book is the *Springer Handbook of Geographic Information* by Kresse and Danko [KD12]. This particular book has a cost of \$241, highlighting the financial challenge associated with a survey of this kind.

Books dedicated to visualization tools are partially included. There are a large number of books dedicated to some individual tools, for example D3.js, a JavaScript visualization library, has over 20 books dedicated to it. As a compromise a single representative book on each individual tool is reviewed with references to the most recent related books. There are many tools that can be used to create data visualizations, therefore a scope boundary is established between propriety tool and open-source tools, with propriety tools excluded and open-source tools included. A book dedicated to Microsoft's business intelligence tools by Aspin [Asp14] is out of scope because it describes proprietary technology, while a book by

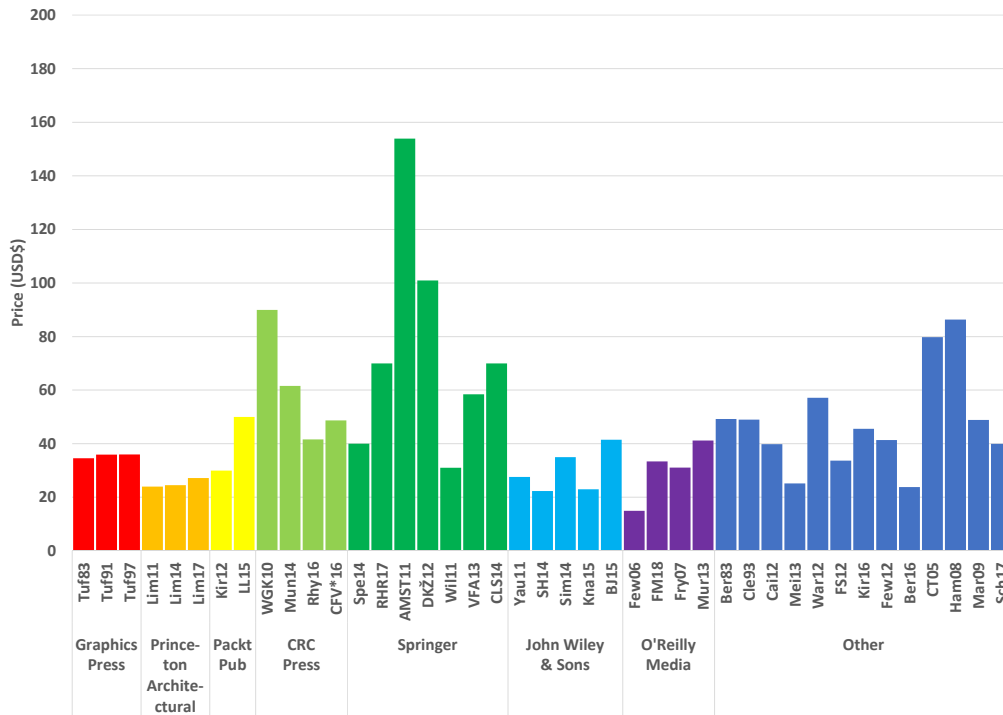


Figure 3: Approximate cost price of each book in USD from Amazon.com. Bars are color coded according to book publisher.

Murray on D3.js [Mur13] is included, because it is open source and free.

1.4. Two-level classification

We propose a novel two-level classification for the books reviewed. The first level is based on the target audience of each book while the second level classifies the topics discussed within each book chapter.

1.4.1. Primary book classification

We classify books into categories based on the target reader and purpose of each book. We examine both the target reading audience described in the foreword of each book and the book's content in order to categorize each. Using this methodology we define six categories for the classification:

- **Classic:** Books that were originally published over 20 years ago and that are considered as outstanding in the field. For Example Bertin's pioneering *Sémiology of Graphics* [BBS67].
- **General Audience:** Books with a broader audience that concentrate more on visual impact rather than the visualization process and casual reading books. We cite Lima's *The Book of Trees* as an example [Lim14].
- **Academic and Textbooks:** These are books that have a more educational stance written for students of data visualization, such as *Interactive Data Visualization: Foundations, Techniques, and Applications* [WGK15]. These are normally targeted at university level students.
- **Industry Professional:** Where the intended reading audience

of the book is a professional, from any industry, seeking to learn broad visualization tips. These books differ from academic books as they provide less background detail. An example in our survey is *The Visual Organization: Data Visualization, Big Data, and the Quest for Better Decisions* [Sim14].

- **Specialized Books:** We consider the books within this category to be for a very specific niche purpose and audience, such as Aigner *et al.* visualizing only time oriented-data [AMST11].
- **Tools:** Books that are dedicated to providing instructions on a particular visualization tool or programming language for developing visualizations. For example *Interactive Data Visualization for the Web* [Mur13].

Books are classified chronologically by the year of the first edition within each category.

1.4.2. Second-level book chapter classification

Topics discussed in each book chapter are used as our second-level classification axis. These chapters are then classified into broader topics. These broad topics, based on the information visualization pipeline [CMS99], consist of introductory topics, data analysis, data mapping, visualization techniques, rendering, and interaction and perception. Table 2 shows the overall classification of the books. The values indicate an approximate number of pages allocated to each of the corresponding topics. The introductory topics are subdivided into two categories, visualization history and the visualization pipeline. Books often introduce the topic of data visualization using a historic aspect, citing the work of John Snow [Sno55] and Minard [Min69] amongst others. The visual pipeline by Card *et al.* [CMS99], or modifications of it, are also often used

to introduce the subject and to convey its processes. The next broad topic of data analysis includes books that discuss techniques that manipulate the data in some aspect. This is subdivided into six topics: data pre-processing and enhancement, dimension reduction techniques, dealing with uncertainty and error margins within the data, visual analytics, different forms of data, and information theory.

Mapping the data to visual information channels is the third broad topic. This includes the visual design, the use of color and color theory, and broader discussion within the topic. Visualization techniques is the fourth broad topic which is subdivided into the techniques used for various data types and visual design styles. Those subtopics are:

- Spatial data
- Geospatial data
- Time-oriented data
- Multivariate data
- Glyph-based visualizations
- Hierarchical data
- Graphs and networks
- Text and document visualization
- Scientific visualization
- Applications

Topics associated with rendering are the next categories listed. These include discussions on presentation, animation, multiple views and the evaluation and comparison of visualization techniques.

The final subject includes content based around interaction and perception. The full list of subtopics are: interaction, human perception and cognition, aesthetics, profiles of prominent people within the subject, and research directions. Our chapter topic classification mirrors the information visualization pipeline from Card's classic book [CMS99] and the topic classification used in a Survey of Surveys [ML17].

1.5. Online resource

We collect books referenced in this survey in an online resource using SurVis [BKW16] for ease of referencing and book curation. This can be found at the following URL: <http://visbooks.swansea.ac.uk/>. The books in this collection can be curated and expanded over time.

1.6. Related work

To our knowledge, this is the first paper to review and classify books within data visualization and visual analytics. There are, however, a number of critical reviews of data visualization books available as articles and blog posts on various websites such as Bremer on her visual cinnamon website [Bre] and a blog post by Cotgreave on the Tableau website [Cot13]. These kinds of reviews have a tendency to be subjective, do not classify the topics covered within each book and rarely review more than ten books. Reviews for each individual book can also be found on online book stores such as Amazon.com [Ama18], however these are scattered, often short, subjective and do not classify topics.

McNabb and Laramee present a Survey of Surveys (SoS)

[ML17] which classifies over 80 survey papers within the information visualization and visual analytics domains. Alharbi *et al.* [AAM*17] also present a survey of surveys of molecular visualization within the computational biology field. These meta-surveys present a broad overview of data visualization landscape, however these surveys only review journal and conference papers and do not consider books.

2. Survey

The surveyed books are organized according to the classification outlined in section 1.4.1, with books placed into one of six categories according to the intended reading audience. Within these categories books are ordered chronologically. The topics discussed within each book and the number of pages dedicated to each topic are indicated in Table 2. Topics are outlined in section 1.4.2.

The following section provides summaries of the books within this survey. Each book is presented in its classified category. We attempt to describe each chapter in one to two sentences. We also try to group every set of two chapters into a paragraph. Chapters one and two are in the first paragraph, chapters three and four in the second paragraph etc. This way readers know where in the book they are whilst reading the summaries. It also enables skimming through the context because new paragraphs correspond to odd numbered chapters.

At the end of each classification category, we provide a comparison and recommendation section which compares each book and highlights strengths within the category and provides our book recommendations.

In addition to the meta-data presented in Section 1, graphical meta-data for this survey is presented and discussed throughout the survey to facilitate comparison of the titles (see Table 1).

Description	Section	Description Page N ^o	Figure Page N ^o
Comparison 1 uses the citation count and Amazon.com sales rank to compare book influence	2.1	9	10
Comparison 2 compares the number of references, pages, and figures in each book	2.2	13	14
Supplementary material is described in Comparison 3	2.3	17	18
Amazon.com reviews are discussed in Comparison 4	2.4	22	23
Comparison 5 compares the topics discussed in each of the categories	2.5	27	27

Table 1: Tables and figures that include meta-data are distributed throughout this survey.

2.1. Classics

The classics section contains books that were initially published over 20 years ago and are noteworthy within the data visualization community. Six books are found in this category, the pioneering *Sémiology of Graphics* by Bertin [BBS67], three books by Edward Tufte [Tuf83, Tuf90, Tuf97], and two by Cleveland [Cle85, Cle93], see Figure 4.

Table 2: Book classification. Books are classified into rows according to their audience and date of first publication. Columns indicate the indication of how many pages are dedicated to each topic indicated by numbers and the color intensity.

		Intro.		Data Analysis					Mapping			Visualization Techniques										Presentation
		Visualization History	Visualization Foundations/Pipelines	Data Pre-processing & Enhancement	Dimension Reduction	Visual Analytics	Data Types	Information Theory	Visualization Design	Color	Data to Visual Primitive Mapping	Spatial Data	Geospatial Data	Time-Oriented Data	Multivariate Data	Glyph-based Visualization	Hierarchical Data	Graphs and Networks	Text and Document Visualization	Sci-Vis	Applications	
Classics	[Ber83]			20					190	8	20		128	76	6	10						
	[Tuf83]	20							110	2			10	20	5							
	[Tuf90]	30							126	14			21									
	[Cle93]					323	7							253						41		
	[Tuf97]	50							150	2												
General Audience	[Lim11]	47														30	172					
	[Cai12]							92							1							
	[Kir12]		10	5				200				14										
	[Mei13]	23					2			2		80	57			29	38	20				
	[Lim14]	80														200						
[Lim17]	60							255				25										
Textbooks	[Spe14]	7	13						10			27		24	3	2	10	15	1			
	[War12]									26						5						
	[WGK15]	16	44	14	3		2		26	4		38	32	32	34	6	10	12	24			
	[Yau11]			27			5		16	10			56	58	106	13						
	[FS12]	20	63	16	1		10		88	6											9	
	[Mun14]					21		20	51	17	25	20			31	2						
	[Kir16]		40	6			34		179	30	7											
	[RHR17]								152													
Industry Pro.	[Few12]	4	13				24		216	5	49				15							
	[Few06]	4					5		110	5					180							
	[Sim14]	1	26	10					5													
	[SH14]			30					10						60							
	[Kna15]								160	3	10				50							
	[Ber16]	15							155	8		5										
	[FM18]		45		1		10		31													
Specialized Books	[CT05]	14		30	2	186	16														26	12
	[Ham08]								12									176				
	[Mar09]		49	21			45		45	5	8	1	22				8				206	
	[AMST11]	30	15		20		24		33	8		13	250									
	[Wil11]	20	2	6			10		44	1	45	2	256	50					4			
	[DKZ12]					200																
	[VFA13]			91	74													17				
	[CLS14]											69										
	[BJ15]	12		32					61	11	31			23			21	156				29
	[Rhy16]	10								170												40
	[CFV*16]		2					194														86
[Sch17]	3				38																	
Tools	[Fry07]			67					18	5			60	40	51		38	44				
	[Mur13]			18					9		10		15				7					
	[LL15]										94									135		



Figure 4: Book covers for classic classified books [BBS67, Tuf83, Tuf90, Cle93, Tuf97].



Sémiologie of Graphics is an English translation of the 1973 book *Sémiologie graphique* written by Jacques Bertin and translated by William J. Berg [Ber83].

Bertin begins by explaining the difference between the invariant and components of an information graphic, how to identify them, and details the importance of declaring these in titles and legends. Identifying the number of components within a graphic is examined along with the number of categories within the components. The second chapter discusses the properties of graphic marks or glyphs. The three classes of marks are discussed; the point, the line, and the area whilst four groups of glyphs are identified; diagrams, networks, maps, and symbols. A detailed analysis of possible graphical variables is then compiled which are able to encode data, these being variations in two-dimensional location, size, value, texture, color, orientation, and shape. These graphical variables, first outlined here, form the basis of graphical data visualization and are discussed in many other texts.

The next chapter begins with a dataset describing the amount and percentage of people working in each of three sectors for each region within France in 1954. One hundred different visualizations are presented of this dataset to demonstrate the effectiveness and variability of visual examples. The efficiency of figures is considered in terms of how easily a user is able to extract information and three functions of graphic representations are identified: to record, to communicate, and to process information. Rules for the construction of different groups of glyphs are then given for their best utilization. Diagrams are the subject of the fourth chapter and are defined as visualizations with one common variable between all data. Multiple examples of diagrams are given for different scenarios such as the number of variables, and different types of variables such as continuous or discrete data.

The fifth chapter's subject is networks - visualizations that have multiple correspondence between variables such as trees and flow charts. Different network arrangements are examined such as rec-

tilinear or irregular, and their construction considered. The final chapter of the book is dedicated to geographic maps. Identifying the scale and the location represented is examined, along with representing the globe as a flat representation and representing geographic features.



The *Visual Display of Quantitative Information* Second Edition by Tufte [Tuf83] begins with a chapter that provides guidelines on what makes an excellent visualization, the ability to simply and clearly show the data and gain insight, along with multiple examples of high quality visual designs. The visual designs also provide a historical perspective of data visualization, often being the first example of that type of visualization. The second chapter discusses visual designs that represent data in a poor manner, sometimes deliberately, so that the data can be misinterpreted. Many examples of visual designs are given and explanations of the problems within them, such as miss-aligned axis scales.

The third chapter queries the reasons for poor graphical integrity, concluding that illustrators lack the necessary skill to communicate data effectively. Chapter four introduces the Data-Ink ratio, a ratio of the amount of ink used to represent data divided by the total amount of ink used in the visualization. Tufte concludes that by maximizing the data-ink ratio a better visualization would be produced, although in some cases, non data-ink is useful for interpretation.

Chartjunk is the focus of the fifth chapter. Three types of chartjunk are identified: moiré patterns, grids and graphical decoration, all of which should be removed from a visual design. A redesign of box plots, bar charts and scatterplots is presented in the next chapter that maximize data-ink.

Chapter seven gives example visual designs that use multifunctional graphics, such as markers that also give data values and encoded glyphs on a scatterplot. In chapter eight, data density is defined as the size of a visualization in relation to the amount of data displayed. Tufte notes that the data density in many published information visualizations is low and that most could easily be shrunk to half their size without the loss of legibility or information. Small multiples are also introduced – shrunk down visualizations that are repeated with one variable changing, highlighting the change in the given variable.

The final chapter provides guidelines for creating aesthetically pleasing information visualizations. These guidelines define when to use tables, the use of words, line weighting, and the shape of the graphic.



In the first chapter of *Envisioning Information* [Tuf90], Tufte outlines design strategies that map many dimensions of data onto the 2D medium of paper or a screen and that maximize the amount of data for a given area. This is achieved with the use of multiple examples that include a timetable for a Java railroad from 1937, and methods for representing sunspots over time. The second chapter concerns the representation of large volumes of data in a concise way. Examples of a Tokyo railway timetable and a choropleth of the population distribution around Tokyo are used to demonstrate the principles discussed.

Using layers to differentiate information is discussed in chapter three. This is primarily achieved with the use of color to differentiate different levels of data, for example, annotations on a diagram. Using lighter colors for the use of guiding marks within figures,

such as grid lines, and the consideration of negative space are also examined. Small multiples is the focus of the fourth chapter, the use of multiple small charts to create a single visual design that allows for the comparison of multiple attributes. The use of small multiples demonstrates depicting sequences and comparisons of river lengths.

Chapter five discusses the use of color and describes the rules set out by Eduard Imhof, a Swiss cartographer, for the use of color [Imh82]. The final chapter studies the depiction of motion over time using examples that include representations for the movement of the moons of Jupiter, various transport timetables and dance instructions.



Cleveland begins *Visualizing Data* [Cle93] by outlining different data types by which the book is organized: uni-variate, bi-variate, tri-variate, hyper-variate, and multi-way. For each data type an example chart is provided. The role of visualization is also outlined; to provide a framework for studying assumptions. Uni-variate data is studied in the second chapter where the visualization of data using quantile-quantile (q-q) plots is explained and exemplified, along with box plots. Using logarithmic and factor scales for dealing with non-normal data distributions is also demonstrated and validated with the use of visualization.

Chapter three is dedicated to bi-variate data. Validating line fitting using loess (local regression) to scatter plot data is demonstrated by visualizing error residuals. Interaction by brushing data for label information is also briefly discussed. Using techniques presented in the previous chapters, chapter four demonstrates the discovery of variable dependence for tri-variate data. Interacting with scatter plot matrices using brushing in multiple views is also discussed along with contour plots and 3D surface plots.

Chapter five extends the techniques from the previous chapters to further variables. The addition of data cropping is used to find dependencies. The scatter plot matrix is used to find outliers and dependencies. The final chapter discusses multiway data, where there is only one quantitative variable and multiple categorical variables. Techniques are used to highlight discrepancies in the data. The book concludes with an example of a previously analyzed data-set where previous statistical analysis failed to indicate an error, easily identifiable using visualization.



Cleveland has also authored another book titled *The Elements of Graphing Data* [Cle85]. This book focuses on the basics of visualization designs such as the proper use of legends, labels, and scales whereas *Visualizing Data* focuses on visual analytics.



Visual Explanations: Images and Quantities, Evidence and Narrative by Tufte [Tuf97] begins with an analysis of visual techniques used to depict quantities using historic images. The complications of projecting a 3D visualization onto a 2D display medium, the use of scales to aid perception, and misrepresentation of scales is also discussed. In the second chapter, Tufte explores the statistical and graphical analysis of two events as a case study of the use of visualization. The first event is the cholera epidemic of London in 1854 where John Snow's analysis proved that cholera was a waterborne disease. The second event was the Challenger space shuttle disaster, where engineers' concerns about the launch temperature failed to prevent the launch.

Chapter three discusses the conveying of instructions demonstrated by pictorial descriptions of magic tricks, as well as the distraction techniques used by the magicians themselves to deceive the audience. The fourth chapter illustrates how minor changes to a visualization can improve its effectiveness. An example is given of an annotated pictorial diagram of an ear, which is then improved by thinning the annotation lines and another example of a topological map with a suitable and unsuitable color-map.

The fifth chapter discusses the use of repetition within visualizations. Examples explored include the calendar used on the Russian Salyut 6 space station during the longest space mission of the time, and the repetition of letters on the Trajan Inscription. Chapter six discusses multiples, repetitions of a figure to show different states. Examples given include a depiction of the different views of Saturn as seen from Earth from 1659 and a representation of medical records.

Confectioned figures are the topic of the final chapter. Composite imagined scenes and compartmentalized figures are exemplified and explored.

2.1.1. Comparison of Classics and Recommendations

All of the classic books make excellent use of diagrams to demonstrate principles discussed in the text. The diagrams and figures used in the Tufte texts tend to be of a higher quality than the diagrams in *Sémiology of Graphics* [Ber83] due to the incorporation of color. The Tufte books also make use of more color-mapped diagrams in comparison. Computing technology is an underrepresented topic within all books classified as classics due to the immaturity of computer hardware at the time of publication.

Both Bertin and Tufte provide recommendations for producing effective and aesthetically pleasing visualizations, and to avoid perception issues such as moiré effects. Tufte recommends increasing the data ink ratio, where as Bertin describes the optimal use of ink as 5% to 10% for best "retinal legibility".

Sémiology of Graphics [Ber83] provides a comprehensive description of the principles and fundamental rules for the use of graphics to represent data. Bertin incorporates many diagrams to outline principles discussed in the text and presents many visualization designs, with cartographic designs being particularly well represented. Rules and guidelines for the construction of a graphical diagram and for legibility of a diagram are presented and are particularly useful. The book does however suffer with minimal use of color and, due to the age of the literature, digitally created images are not discussed. Bertin's *Sémiology of Graphics* is suitable for beginners in data visualization to understand good graphics principles, or as a reference book for best practice. We recommend the Bertin book for readers that are interested in seeing a historical view of the field, especially on how the field has evolved since 1967. It holds a very special place since it was first, and much of the content is still relevant today.

The Visual Display of Quantitative Information [Tuf83] also makes good use of figures to demonstrate principles outlined in the text. Particularly are figures showing best practice and poor design choices. A good demonstration and description of misleading graphics is provided. Rules for legibility of a diagram are a valuable resource to anyone looking to produce good visual designs.

We recommend Tufte's book on *The Visual Display of Quantitative Information* for readers coming from any background. Tufte's book offers a very engaging and informative exposition accessible to a very wide audience. It is both educational and entertaining.

Tufte's second book *Envisioning information* [Tuf90] continues in a similar vein to his previous book, with many historical figures used to demonstrate themes from the text. Themes discussed in this book include representing multiple dimensions, using layers in visual designs, and the use of color.

Visualizing Data [Cle93] is more technical than the other books within this category and expects a knowledge of basic statistics. This book provides a statistician's perspective of visualization. It demonstrates the step-by-step process for the use of visualization to analyze and validate data. The book is well supported by numerous figures to demonstrate the techniques discussed. We recommend this book for those who are interested in visual analytics as it provides detailed instructions on the subject.

Visual Explanations [Tuf97], as the title suggests, focuses on graphics and diagrams that explain, e.g. demonstrates the use of graphics as evidence for decisions. This is the only book surveyed to discuss the use of pictorial descriptions. The chapter titled "*The Smallest Effective Difference*" also offers pictorial examples of how small changes in visual designs can make significantly influence the impact of a diagram, something not offered in other books. The book does however forego information on how to map data to graphics that other books in this category discuss. *Visual Explanations*, like the other Tufte books, makes use of a large number of high quality images and is written in an entertaining style. The book complements Tufte's other books and is suitable to anyone interested in data visualization. We recommend Tufte's other books to those readers who usually enjoy his classic best seller and are interested in more and similar material including more depth on a range of topics.

Comparison 1. Comparison of Literature Influence Books within the 'Classics' category are all highly cited, this can be seen in Figure 5 which charts the influence of each focus book described in the survey. The number of citations are on the vertical axis and the Amazon.com sales rank [Ama18] on the horizontal axis (using a logarithmic scale). Tufte's book *The Visual Display of Quantitative Information* is notable as being the book with the most citations according to Google Scholar [Goo18], with over 10,500 citations, almost double the citations of the next book, Tufte's *Envisioning Information* [Tuf90]. Tufte's other book *Visual Explanations* [Tuf97] is the seventh most cited book out of the books reviewed in this survey, highlighting the influence of Tufte on the data visualization field. The other book classified as classic, Bertin's *Sémiologie of Graphics*, has also amassed over 3,000 citations according to Google Scholar. It should be noted however, that the books in this category have had a longer time to collect citations.

Knaflig's *Storytelling with Data* [Kna15] is ranked as the highest book in Amazon's sales rank, followed by two of Tufte's books.

2.2. General Audience Books

Books in this section are written for a wider audience than just the data visualization community. We include six books in this category, the covers for each can be seen in Figure 6.



Visual Complexity: Mapping Patterns of Information by Lima [Lim11] begins by investigating historic tree models, citing them as early precursors to network diagrams. Lima notes that early tree imagery is often due to the cultural importance of trees, and over time these developed into descriptions of scientific knowledge. Dürer's 1509 engraving "*The fall of Man*" [Dur04] and the table of contents from Chambers' *Cyclopaedia* [Cha38] are examples of the images used in the chapter. The chapter concludes with a note that tree diagrams are still widely used today, although their designs today are non figurative. The second chapter discusses the need for visualizing structures more complex than trees. The need for network visualization is identified in complex problems, such as currency stabilization, where multiple factors can influence an outcome. Other network problems identified are in urban development, neurological research, the internet, information classification and natural ordering.

The idea of network science is introduced in the third chapter, with the first documented network analysis, credited to Euler [Eul41], with further developments by Moreno and Northway in the field of psychology [NRM55]. The similarity between cartography and network visualizations is highlighted and principles for creating a network visualization are outlined, such as defining the purpose of the visualization and encoding the data to visual primitives. The fourth chapter gives multiple examples of network designs organized by the type of data which they convey. Examples of the figures shown are a chord diagram showing email communication by Baker [Bak07], and an arc diagram showing the differences and similarities between the holy books of different religions by Steinweber and Koller [SK07].

Chapter five features 90 images and their descriptions, organized into design type. Figures include a flow chart visualizing Last.fm listeners by country made by Adjei and Holland-Cunz [AHC08], and an elliptical implosion depicting word frequency and associations in Carroll's "*Alice in Wonderland*" by Paley [Pal09]. The aesthetics of network visualizations is explored in the sixth chapter. Comparisons are made between modern abstract works and network visualization.

The final chapter features four essays by separate invited authors, which discuss trends in information visualization and future possibilities such as visualizing complex ecological relationships.



The Functional Art: An Introduction to Information Graphics and Visualization by Cairo [Cai12] begins with an example of fertility rates by country and uses data from the UN to create a visualization to support a hypothesis. The explosion of data is mentioned and the need to turn the data into knowledge and wisdom. The second chapter uses examples of infographics published in the media analyzing them critically and suggesting alternative visualizations that better represent the underlying phenomena, highlighting the need of having a variety of visualization techniques for the data.

The aesthetic of information visualizations is the subject of the third chapter, with discussion of finding balance between complex

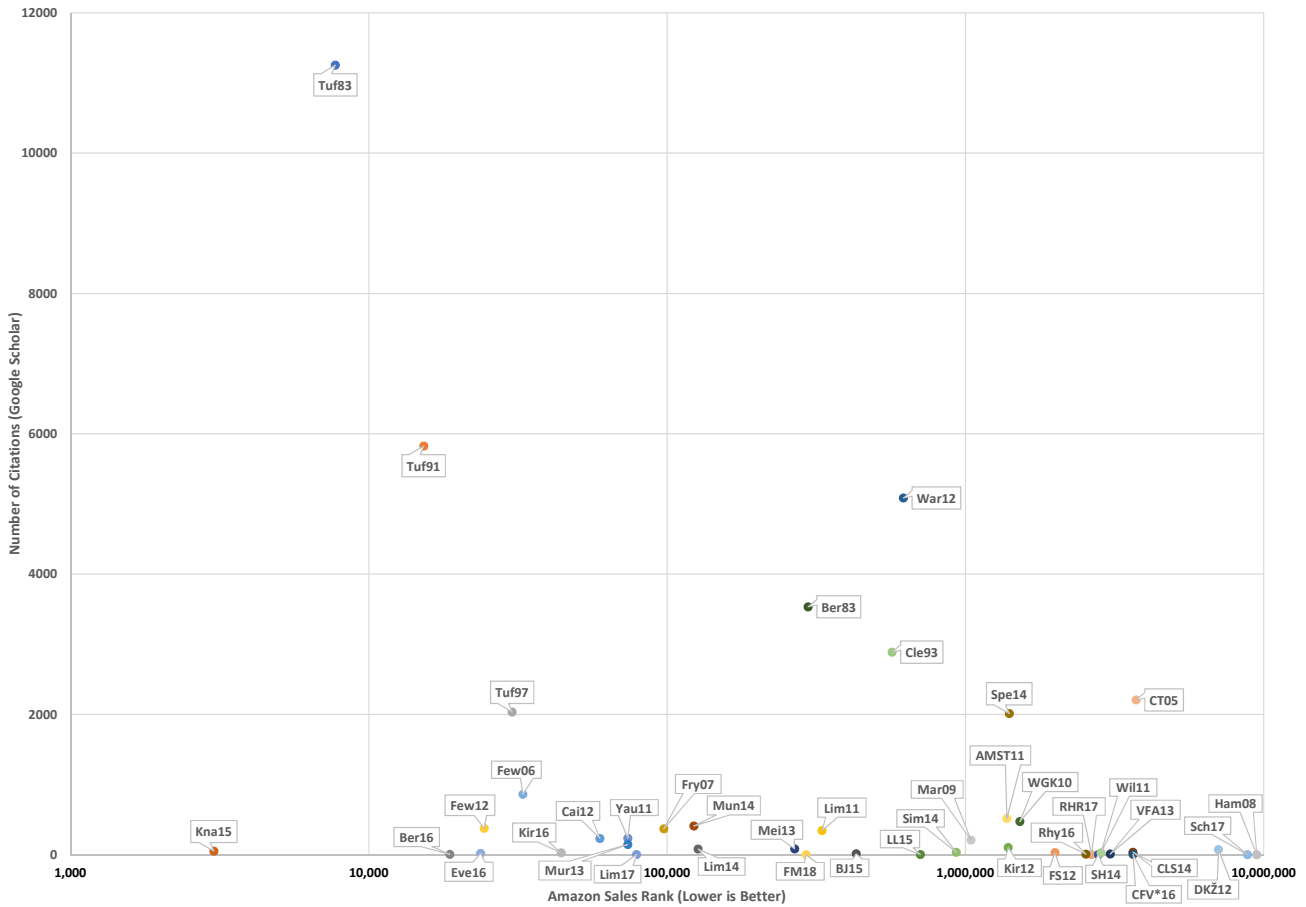


Figure 5: Scatter plot of influence. A visualization of the number of citations (according to Google Scholar) vs Amazon sales rank on a logarithmic x-axis. Analysis of this Figure is discussed in Comparison 1.

and deep visuals, and intelligible and shallower visual designs. The complexity and volume of information displayed is discussed in the context of the audience of the visualization. Minimalist visualization versus highly decorated visualizations are also examined. The themes from chapter three continue into chapter four, where further published information visualizations are critically analyzed for the information they show and their effectiveness.

The fifth chapter describes the physiological structure of the human eye and how we build a perception of our surroundings by eye movement. The difference between what we see and what we perceive is also highlighted with some examples of optical illusions. Applying what is known of human perception to visualizations, is the subject of chapter six. The use of shade is shown to be easier to differentiate compared to the use of shape, and space between objects help us create natural groupings. Different methods of encoding data, as outlined by Bertin [BBS67], are also ranked by their ability to allow for more accurate judgments.

Chapter seven describes the process of how humans identify ob-

jects, such as faces, by comparing them to our long-term memory. It explains that by only having outlines of objects enables us to recognize objects faster than when they have complex detail. The eighth chapter gives examples of published information visualizations along with some preliminary sketches of those visualizations to show how they developed.

In chapter nine four considerations are highlighted when designing interactive graphics: Visibility - ensure that interactive objects such as buttons are visible, feedback - to offer a conformation of the user input, constraints - the limitations of the interaction are clear, and consistency - fonts, colors, positioning and interaction techniques should be consistent throughout a visualization. Examples of interactive visual designs are provided. The final chapter presents a profile of ten different information graphic designers, including an interview with each and examples of their work.



Cairo published a follow up book in 2016, "The Truthful Art", [Cai16] which expands on the principles from The Functional Art [Cai12].



Figure 6: Book covers for general audience information visualization books [Lim11, Cai12, Kir12, Mei13, Lim14, Lim17] described in section 2.2.



Data Visualization: A Successful Design Process by Kirk [Kir12] highlights the importance of data visualization in the modern era to help make sense of data, and describes some design objectives that enable assessments of visual designs to be made. The importance of assessing the goal of a visualization and the management of its creation is the focus of the second chapter. Defining the audience and purpose of the visualization whether to explain, explore or exhibit are considered to ascertain the design.

Chapter three concentrates on the editorial focus of a visualization and the process of data acquisition, sorting and analysis. Analysis of data to find stories is discussed and demonstrated with an example exploring the number of medals won by countries over a number of Olympic events. The design options are discussed in the fourth chapter. Choosing the visualization method and the accuracy of different encoding channels are considered along with the use of color and annotations. Interactivity, focus and context and layout are also studied, with all design choices dependant on data properties, data relationships, and audience.

The fifth chapter provides a taxonomy of many different visual designs such as flow charts, bar charts, and different forms of cartographic charts. The final chapter introduces tools that can be used to analyze and represent data such as Tableau, and programming environments such as D3.js. Finally guidance is given on evaluating a finished project and utilizing feedback for future work.



Design for Information: An Introduction to the Histories, Theories, and Best Practices Behind Effective Information Visualizations by Meirelles [Mei13] begins by describing the philosophy of the book, to analyze existing visual imagery for design and information content.

The first chapter addresses hierarchies. The importance of spatial layout of a visualization from a cognitive standpoint and the way humans detect patterns is discussed. Example images of hierarchical visualizations are given including a graphic by Ramon Lull

from 1515 and computerized visualizations from the 1990s. A case study of TreeMaps is made, in particular the ‘Map of the Market’ visualization by SmartMoney [Wat98], to describe in detail the intricacies of the design. In chapter two the concept of a network as a set of nodes with links is explained, along with extensions such as link direction, weighted links and network path traversal. Node-link diagram examples are given to highlight potential problems in their creation such as occlusion and label placement and to show the many different layout examples such as circular, force directed geographical, radial, and Sankey diagrams.

Visualizing time-based data is the subject of the third chapter and is explored through a number of visual designs such as Jacques Barbeu-Dubourg’s history since creation from 1753 [BD53] and “Fifty Years of Space Exploration” by McNaughton and Velasco in 2010 [MV10]. A brief history of maps is given in chapter four, accompanied by an explanation of map projections, a discussion on map scale and encoding data onto a map. Multiple examples of maps with different coding of data attributes are given, for example, choropleth and isometric maps.

The fifth chapter examines visual designs that have variables in both time and space. Examples are used such as Minard’s 1869 “Napoleon March to and from Russia” [Min69] and mobility patterns showing traffic in Milan. Visual layouts that represent text data are the subject of the sixth chapter. Wordle [Fei14] and IBM ManyEyes [VWVH*07] are used as case studies to show examples of text visualization.



The Book of Trees by Lima [Lim14] starts with a historical look at the use of tree representations throughout history, from tree impressions on Bronze Age, Mesopotamian cylinder seals to treemap representations of folder structures on a PC’s hard drive. Significant characters who used depictions of trees throughout history are also discussed, from Aristotle, through René Descartes to Shneiderman.

The first chapter, shows multiple tree images depicted as realistic trees are presented in a chronological order. Chapter two focuses on vertical hierarchical structures without dendrological depictions. Sixteen figures are given of various vertical trees such as “X-Men Family Tree” by Joe Stone in 2011 [Sto11].

The third chapter discusses horizontal trees. Multiple figures are given of seminal trees including the table of contents from 1728 Cyclopaedia by Ephraim Chambers [Cha38]. Chapter four discusses multidirectional trees, trees with a hierarchical structure not fixed to a horizontal or vertical axis.

Trees with a radial layout is the subject of the fifth chapter. Examples are given such as a complete phylogenetic tree for all mammal species by Bininda-Emonds et al [BECJ*07]. The sixth chapter explores hyperbolic trees, a variation of radial trees which allows the focus of the visualization to be placed on a single node using layout algorithms. This enables the depiction of large datasets with focus and context interaction.

Chapter seven examines rectangular treemaps, a hierarchical visualization using nested rectangles. Examples of rectangular treemaps given are a 1874 chart showing the population of states of the USA by Walker [Wal74], Wattenberg’s “Map of the Market” [Wat98] and a treemap of a PC folder structure by Johnson and Shneiderman [JS91]. Voronoi treemaps, where the space is filled with mathematically derived space rather than rectangles, is

the subject of the eighth chapter.

The next chapter discusses another variation of treemaps, the circular treemap. Lima notes that these aren't as popular as other treemaps due to their ineffective use of space. Example images include a circular treemap of the market value of 2,000 public companies featured in the previous chapter. Sunbursts visualizations or radial treemaps are the subject of the penultimate chapter. These are similar to treemaps, but use a radial layout with the root of the hierarchy at the center of the visualization.

The final chapter discusses icicle trees, these are a non-nested method of showing a hierarchy that can be used in either a horizontal or vertical arrangement.



The *Book of Circles* by Lima [Lim17] commences with an exploration of circular objects in nature and the use of circles in human history. Circular cities and artifacts, such as Stonehenge in the United Kingdom, circular symbols, the use of circles as metaphors and human preference for curvilinear lines are examined. The second chapter classifies different forms of circular designs into seven families upon which each of the next chapters will be based. The families consist of rings and spirals, wheels and pies, grids and graticules, ebbs and flows, shapes and boundaries, maps and blueprints, and nodes and links.

Over 30 images of ring and spiral representations are presented in the third chapter such as a NASA compiled photo of Jupiter focused on its North Pole. Wheels and pies are presented in the next chapter. These are circular diagrams that have radiating lines from the center of the circle. Some examples presented in the book are multiple pie charts depicting the religious makeup of each US state and Florence Nightingale's diagram depicting causes of mortality during the Crimean war.

The fifth chapter features images combining circular depictions with grid structures. Circular depictions which ebb and flow are the subject of the sixth chapter. Figures such as radar charts, circular bar charts, radial area charts and radial line charts fall into this category. Visual designs that captivate images of circles within a circle, circular treemaps, and Voronoi patterns are presented in the following chapter. An example of an image depicted is a Voronoi diagram representing the various sizes of countries' economies.

The penultimate chapter looks at circular depictions of geospatial information. Example figures include architectural blueprints, historic maps and other geographical depictions. The final chapter explores different circular node link diagrams such as a depiction of co-authorship between physicians publishing on hepatitis C.

2.2.1. Comparison and Recommendations of General Audience Books

These general audience books by their nature feature an accessible reading style and have a high number of large images.

Visual Complexity by Lima [Lim11] features almost three hundred full-colored images with brief captions on the data they represent. A detailed description of the visual history of trees and networks is provided. Other than historical aspects, the book does not provide explicit educational function in order to study visualization techniques. However multiple designs are presented and due to the large number of examples, the book may be a good source of inspiration. We recommend this book to a wide audience, however the

book is not recommended to those looking for pedagogical guidance in data visualization. This book is very image-oriented. Lima does an outstanding job of collecting a wide range of a visual designs and examples inspired by trees and networks. He pays special attention to aesthetics.

The Functional Art by Cairo [Cai12] covers a wide range of data visualization topics, however the focus of the book is towards infographics. A thorough description of visual perception and cognition is provided, as well as tips for creating interactive graphics. Profiles of infographic designers are also given along with examples of their work, a unique feature within all books surveyed. Images are used to support topics discussed in the text, however *The Functional Art* doesn't have as many visualizations as are presented in the Lima books. We recommend this book to beginners in information visualization, particularly those with an interest in infographics.

The process of creating a visualization is discussed by Kirk in *Data Visualization: A Successful Design Process* [Kir12]. Different perspectives on data visualization are discussed dependant on the purpose of the visualization and the intended audience. A useful taxonomy of visualizations is presented, showing a number of visualization designs, along with a useful range of digital tools to aid visualization. The book is written in an informal style, however it does not feature as many interesting images as the other books within this category. We recommend this book to beginners in data visualization with the book providing a basic guide and a useful list of digital resources. This book is more pedagogical than the others in this category. It's intended for those that are really interested in applying the guidelines and principles presented. However it's not in the textbook category because it's for a general audience (as opposed to a special target student audience).

Meirelles' *Design for Information* [Mei13] is divided into six chapters, each of which discusses the visualization of particular data type; hierarchical, relational, temporal, spatial, spatio-temporal, and textual. For each of these data types, a history of visualization is presented along with a number of visualization designs to demonstrate different methods to encode the data. The book provides a number of high quality images to compliment the text which provides an entertaining read. This book, along with others in this category do not present information with regard to computational technology and creating digital representations. It's a concise book which can be read quickly, so we recommend this book to readers interested in a quick and casual introduction to information visualization. We also recommend this book to those with interest in visualization designs, such as design students. The author's style follows a principle-by-example style.

Lima's second book within this category, *The Book of Trees* [Lim14], is the successor to visual complexity with a stronger focus on tree visualizations. Almost two hundred examples are presented in color, along with a detailed history of tree visualizations. The book is arranged by visualization design such as different forms of treemaps and sunbursts. Similar to Lima's *Visual Complexity*, educational content is limited to a historical aspect, but many visualization designs are presented. As with Lima's other books we recommend this book to a wide audience or for those interested in hierarchical visualization and who also liked his first book. Again Lima's style focuses on images and examples. He has gone to great

length to collect interesting examples of tree-like visual layouts that are very engaging. This is definitely a strength of each of his books.

The Book of Circles [Lim17] is Lima's third book in this survey and follows a similar recipe to the others. Imagery with circles is the focus of this book with the examples themselves taking centre stage. A historical aspect is presented along with three hundred graphic examples. Visualization fundamentals and other educational aspects are not presented. We recommend this book to a wide audience interested in nice visuals or for those interested in circular designs and readers who enjoyed his two previous books. We strongly recommend Lima's books for those readers interested in an example-image-based approach to the topic. His collection of examples are impressive and fascinating.

Comparison 2. Comparison of Book Quantities Figure 7 shows the number of references, pages and figures contained in each book described in the survey. The most notable book in the chart is Bertin's *Sémiology of Graphics* [BBS67] which has the greatest number of figures and references no other work. Books from the Textbooks and Academic category tend to have a higher number of pages and cite more references in comparison to other books within the survey. Ware's *Information Visualization: Perception for Design* [War12] contains the most references of the surveyed books and *Interactive Data Visualization by Ward et al.* [WGK15] has the highest number of pages.

2.3. Textbooks and Academic

This section contains eight books that are targeted predominantly at data visualization students at the university level. Book covers can be seen in Figure 8.



Information Visualization: An Introduction, Third Edition by Spence [Spe14] begins by showing different examples of visual designs from history and explains why information visualization is important – to gain insight.

In chapter two an example dataset is used to construct a visual design and to show the benefits of using the appropriate visualization method, highlighting challenges along the way.

The third chapter introduces visual layout techniques such as star plots and parallel coordinates. These are then critically analyzed with human perception factors in mind. Interaction techniques are briefly discussed and data relationship visualizations are assessed. Display media other than paper or a PC screen are also discussed. Chapter four explores the use of techniques such as focus+context and zooming as methods to make the most of display space. A discussion of using temporal space is also presented and the considerations of the use of both spaces to enable effective visualization.

Norman's action cycle is introduced as a model to develop interaction techniques and an example of its use is given in the fifth chapter. The considerations of intuitive design and expected results from interactions are discussed. The author identifies eight explicit steps for creating an interactive information visualization in chapter six. These steps range from receiving the commission, through idea generation and sketching, to presenting the final product.

Case studies of designing visualizations for various problems are given and critically analyzed in the final chapter. A description of

videos to compliment topics raised in the book are given in the appendix.



Information Visualization: Perception for Design, Third Edition by Ware [War12] considers the importance of human perception when designing visual layouts along with the importance of reducing cognitive load. Perceiving real-world environmental factors such as surface textures and the behavior of light on computer monitors are considered in the second chapter. Other digital visualization techniques such as augmented reality, virtual reality and display walls are also discussed with a focus on how much the human visual system can comprehend.

Chapter three examines contrast, luminance, and brightness and the way that they are perceived. A detailed description of how color is perceived and accurately reproduced is given in chapter four along with guidance on the best use of color.

The fifth chapter explores visual primitives such as color, size, shape and orientation. To create distinctive glyphs to stand out among others, at least one of those visual primitives must be vastly different. The perception of textures, two dimensional shapes, proximity and connectedness are evaluated in chapter six. The visualization of multi-variate data is explored, overlapping data, parallel coordinates, treemaps, and dynamic visual designs are all examined with a view to find patterns in the data.

Chapter seven explores the visualization of data in 3D space. Depth cues such as shading and textures in standard and stereoscopic displays are considered. The perception of surfaces and the positions of data points in space are also examined. Theories of object recognition and pictorial representations are discussed in chapter eight.

The use of images, text, verbal dialogue and animations in data visualization is scrutinized in the ninth chapter, with the aim of a narrative in mind. The importance of the use of gestures, such as the use of arrows to highlight, is also discussed. The tenth chapter provides a detailed explanation of how data interaction techniques should behave, e.g. providing visual feedback, and the theory of why is discussed.

The thinking process of the user is considered in chapter 11 with emphasis on cognitive load and memory. Example algorithms are given that describe the user interaction with differing visual designs, with the aim of reducing cognitive load. The appendix gives information on CIE color standards, goals that would evaluate a visualization, and summarizes all of the guidance points highlighted throughout the book.



Ware has published an additional book [War10] in which the perception concepts discussed in *Information Visualization: Perception for Design* are expanded on to provide advice for designers. In the interests of scope, managing volume and overlapping subject material this book is not given a full review within this survey.



Interactive Data Visualization: Foundations, Techniques, and Applications, Second Edition by Ward et al. [WGK15] starts with an introduction to information visualization and why it is important, for information understanding, along with a brief history of information visualization. The visualization pipeline is introduced and human perception discussed. Data characteristics are described in the second chapter along with

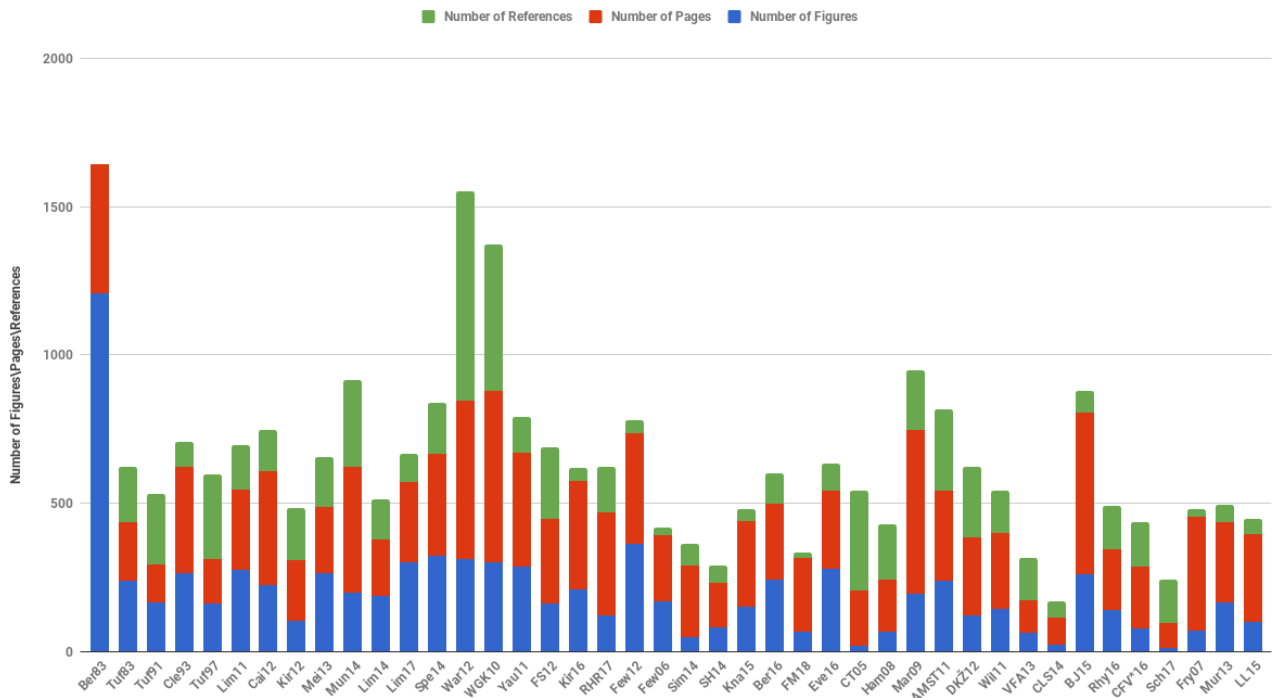


Figure 7: Stacked bar chart comparing the number of references, pages and figures in each book, discussed in Comparison 2.



Figure 8: Sleeve covers for books classified as textbooks [Spe14, War12, WGK15, Yau11, FS12, Mun14, Kir16, RHR17] described in section 2.3.

some data preprocessing techniques such as normalization and sub-sampling.

Chapter three deals with human perception of graphics and images specifically in information visualizations. A physiological description of the eye is given along with many figures to highlight perception oddities. Card *et al.*'s visualization pipeline [CMS99] is discussed in detail in chapter four, along with the use of visual primitives. Data visualization taxonomies, such as Shneider-

man's [Shn96], are also discussed.

Chapter five presents different techniques used for scientific data, 1D, 2D and 3D spatial coordinates along with dynamic flow. The sixth chapter provides a basic introduction to geospatial data with specific real-world coordinates. Map projections, representing the globe as flat are followed by techniques used to visualize area, line and point data on maps.

Chapter seven focuses on different categorizations of time oriented-data and gives an example of each. The categorization is based on abstract or spatial data, univariate or multivariable data, linear or cyclic time, instantaneous or interval data, static or dynamic visual designs, and 2D or 3D visualizations. Chapter eight discusses different techniques used to show multivariate data categorized into point, line and region based, including combinations of techniques.

Visualization techniques for trees, graphs and networks for data interrelationships are presented in the ninth chapter, such as treemaps, sunburst displays and node-link diagrams for both hierarchical and arbitrary data. Chapter ten introduces text and document visualization techniques and some example approaches such as word clouds and arc diagrams are given. Document collection visualizations are also presented.

Chapter 11 introduces different classes of interaction techniques: navigation, selection, filtering, reconfiguring, encoding, connecting, abstracting, as well as combinations of those classes. Interaction spaces are also identified as screen, data value, data structure, attribute, object or visualization structure. Chapter 12 discusses and exemplifies algorithms used for each of the interaction spaces iden-

tified in the previous chapter. Algorithms for animating the interaction and interaction control principles are also discussed.

In chapter 13 guidance is given for designing visualizations such as data mapping, selecting views, information density, labelling, use of color, and the importance of aesthetics. Potential problems that may occur when producing visual layouts are also identified. The importance of identifying the purpose and the audience of the visualization are highlighted in chapter 14, along with data characteristics and image characteristics in order to benchmark different visualizations.

The penultimate chapter presents software tools for producing different forms of data visualizations such as GGobi [SLBC03], Graphviz [EGK*01] and Tableau [Tab]. Active research fields in data visualization are highlighted in the final chapter. A history of computer graphics and visualization is given in the appendix along with some sample datasets and some example program code.



Visualize This: The FlowingData Guide to Design, Visualization, and Statistics by Yau [Yau11] begins by explaining that visualization is useful to present numerical data in an engaging manner, and to show patterns and relationships. Basic principles are outlined for creating visual designs such as labeling axes and data correctly, citing data sources, and to consider the target audience. The second chapter discusses the data used to create the visual layout, where to find data sources and how to format the data. Example Python code is given to extract data from an online resource and some applications are discussed that will reformat data.

Example tools used to visualize data are presented in the next chapter. The applications are split into ‘Out of the box’ tools such as Microsoft Excel and Tableau, ‘programming’ tools such as Python and R, ‘illustration’ tools such as Adobe Illustrator and Inkscape, and geographical ‘mapping’ tools such as ArcGIS and Polymaps. Chapter four concerns the visualization time-dependant data. Bar graphs, scatter plots and line graphs are used to display the time data along with detailed description of their creation using R and refined with Adobe Illustrator.

The fifth chapter concerns graphics that use proportional representations, such as pie charts, stacked bar graphs, treemaps and stacked area graphs. Some R code examples are provided for their creation. Visual designs for finding data relationship are discussed in chapter six. R code examples are again given for scatterplots, scatterplot matrices, bubble charts and histograms.

Chapter seven is concerned with visualizing high dimensional data. Heatmaps, Chernoff faces, star charts and parallel coordinates are used as examples along with R code for their creation. Dimensionality reduction is also briefly addressed. The penultimate chapter is based on geographic data. Python and R visualization script examples are given that link to various APIs such as Google Maps to create images. Interactive flash based animations are also used along with coding details.



Yau also has a related book called *"Data Points: Visualization That Means Something"* which focuses more on the graphical side of data visualization. [Yau13].



Interactive Visualization: Insight Through Inquiry by Bill Ferster [FS12] commences with different genres of the

topic and describes the historic advancements with the introduction of computers and the internet. Some theoretical models are discussed such as the cognitive load theory and, in particular, the ASSERT model. The ASSERT model (Ask, Search, Structure, Envision, Represent, Tell) provides a framework for creating interactive visual designs and also provides the structure for the next chapters. Chapter two is based on the first step of the ASSERT model and is titled *"Ask a Question"*. The chapter discusses how to create a solid question to drive the development of a visual design. Considerations are made to the intended audience and their expertise, and for creating a focused question. Techniques for generating a good question are also discussed.

The third chapter addresses the *"Search for Information"* section of the ASSERT model, which considers sources of data. The difference between primary, secondary and tertiary sources of data are explained, data quality is discussed and example sources of data are provided. Structuring the information is the next step of the ASSERT model. Quantitative versus qualitative information is explained and data structures and comparison with other datasets are discussed. Linked Data and Semantic Web projects that aim to structure data on the internet are also mentioned.

Chapter five concerns the Envision section of the ASSERT model which examines strategies for analysing and representing the data to answer the question. Analysis techniques are discussed for quantitative data such as relationships, and qualitative data such as frequency analysis. Creating a visual design is the topic of the sixth chapter which is the Represent section of the ASSERT model. Many aspects to be considered when creating imagery are discussed such as perception and cognition, usability, aesthetics, use of color, and spatial arrangement. Different display strategies are discussed and a number of tools that can be used to create visual depictions are presented such as Wordle [Fei14].

The final part of the ASSERT model is tell a story. The use of a storytelling component in a visualization is delineated and examples of storytelling visualizations such as Minard’s map of Napoleon’s march to Moscow are given before a discussion of misleading representations. Chapter eight provides a case study of depicting the books available at the University of Virginia’s first library when it opened, to show the application of the ASSERT model.

The ninth chapter discusses the technology behind the internet and how it is accessed including the use of Adobe’s Flash and vector graphics. Statistical techniques for data processing are presented in the tenth chapter, for example standard deviation and regression.

Chapter 11 provides guidance on using spreadsheet programs such as Microsoft Excel and Google Docs. Databases are briefly explained in chapter 12 along with XML.

The 13th chapter discusses methods to improve the accessibility of visual designs created for example taking into consideration the colorblind and providing clear alternative text. The final chapter discusses VisualEyes [Uni], a Flash-based authoring tool to create dynamic visual designs.



Visualization Analysis and Design by Munzner [Mun14] begins with a discussion of why computer graphics for data is needed, for the discovery of knowledge in data, and gives a broad overview of considerations, such as resources available, for creating visual designs. Data types, their format and

their sources are categorized and explained in the second chapter. Example data types discussed are hierarchical, geometric and time dependant data.

The third chapter explores the importance of considering the reasons for creating visual representations, who the user is and what goals the imagery is trying to achieve. This is to ensure that the appropriate design choices are made when creating a visual design. Examples are given of visual designs created using derived data. Chapter four describes four distinct levels for creating a visual design; identifying the requirements of the visual design, abstracting the data, designing the visual layout and interaction, and encoding the visual design computationally. The importance of validating each of these stages is highlighted.

Methods for displaying data are discussed in chapter five, including different ways of encoding data into graphical primitives or 'channels' such as color and size. A detailed analysis is also made of the effectiveness of the different attributes with a consideration of human perception. Chapter six presents eight rules of thumb for consideration when designing visual layouts along with detailed justification of each rule. Example rules of thumb include the use of Shneiderman's information seeking mantra [Shn96].

Visualization techniques for data found in tabular form i.e. information data, are discussed in the seventh chapter, along with examples such as scatter plots, stream graphs, bar charts and parallel coordinates. Spatially ordinated data visualizations are considered with examples given in chapter eight. These include geographical data as well as scalar and vector fields. The techniques used to visualize them are also discussed here.

The ninth chapter examines network and tree diagrams, various node-link diagram examples are given and their benefits weighed against matrix views. Hierarchical visual designs examples are also provided including treemaps. The use of color to map data attributes is studied in detail in chapter ten, in particular the perception of luminance, hue, and saturation used in different scenarios. The use of other attributes such as shape, size and texture are also mentioned.

Interaction techniques such as highlighting elements and changing the viewpoint by zooming and scaling are described in chapter 11 along with cutting and slicing of 3D scenes to produce 2D visualizations. Considerations of dissecting data to create multiple views are discussed in chapter 12. Adjacent views and layered views are presented with multiple programs using the techniques described.

Chapter 13 presents options for reducing data by filtering. Three broad methods are described; reducing the number of items, reducing the number of attributes, and grouping data. Examples of applications that use these techniques are given. Focus and context techniques are split into three categories in chapter 14; superimposing layers, geometry distortion, filtration and aggregation. Example applications for each of the categories are given.

The final chapter provides case studies of six different applications, namely Graph-Theoretic Scagnostics [WAG05], VisDB [KK94], the Hierarchical Clustering Explorer [SS02], PivotGraph [Wat06], InterRing [YWR02], and Constellation [MGR99].



Data Visualisation: A Handbook for Data Driven Design by Kirk [Kir16] begins by defining data visualization as "*The representation and presentation of data to facilitate*

understanding", and presents three principles for good visualization design, to be trustworthy, accessible and elegant. The second chapter outlines a workflow and describes a mindset for creating an information visualization design. The workflow forms a basis for the layout of the rest of the book with the first step of formulating the purpose of the design, followed by data handling, the focus of the visual design, and designing the visualization.

Chapter three discusses the initial stage of creating a design, formulating the design brief. Considerations of the audience, constraints such as time, available technology and required deliverables are discussed. A discussion is made of different data types, textual, nominal, ordinal, interval and ratio in chapter four. Sources of data and preprocessing the data into a useable format are also discussed.

Chapter five presents the design approach to visualization, the editorial angle, what data is included and what the focus of the visualization is. Examples of visual designs are studied to demonstrate the editorial aspect such as a series of charts showing National Football League touchdown passes. The sixth chapter discusses visual encoding methods and describes 49 different types of charts and their properties, ordered by the type of data that they convey.

Interaction options are split into two features, data adjustment and presentation adjustment in chapter seven. Examples of data adjustment include filtering, animating, and data exploration, whilst examples of presentation adjustment are visual emphasis, annotations and orientation. The use of annotations is discussed in chapter eight. The use of headings, introductions and footnotes, as well as chart marks, labels, and legends is examined.

An overview of color theory is given in chapter nine before a discussion of the most appropriate use of color for data legibility and emphasis. The penultimate chapter discusses the composition and layout of a visualization. Details such as the labeling and ordering of scales, and chart size and orientation are examined.

In the final chapter a critical analysis of a visual design is provided, highlighting the techniques made throughout the book. The multidisciplinary nature of being a data visualizer is also explored.



Five Design-Sheets: Creative Design and Sketching for Computing and Visualisation by Roberts *et al.* [RHR17] begins by detailing the book's purpose, as a method to help create a software tool, in particular a data visualization tool. The layout of the book consists of three parts matching the process of developing a tool: think, prepare, and sketch. The second chapter provides an overview of the Five Design-Sheets method for planning and designing software interfaces by sketching. As the name suggests, the method involves five large sides of paper, the first for sketching ideas, the second, third and fourth for different design solutions and the fifth for a sketch of the final design. The structure of the sheets is described and the use of the Five Design Sheets method in different scenarios is detailed.

Chapter three discusses different forms of problems, characterizing them and various ways of thinking to address a variety of problems. Social and ethical considerations are studied in the fourth chapter. Issues such as evaluating the need for a particular software tool, whether an existing tool exists, if the resources are available to create the tool and if the authority exists to create the tool are examined along with ethics of software development.

The fifth chapter looks at sketching skills and techniques. The use of sketching in planning is highlighted and uses of color and line thickness to enhance sketches is explained. Tips to improve sketching skills are also provided with a recommendation to create a sketching kit. Chapter six explores the Gestalt principles and how humans interpret graphical marks on a page. Bertin's 'Representational Styles' presented in Semiology of Graphics [BBS67], which classify whole graphical images, are also explored.

Methods to help creative thinking and generate ideas are presented in the seventh chapter such as taking inspiration from nature and being well rested. The intricate details of the Five Design-Sheets method are laid out in the next three chapters. Chapter eight details the first sheet which is to frame the problem, gather ideas, refine them and to define three potential ideas to move onto the next stage.

Sheets 2, 3 and 4 of the method are presented in chapter nine, which describes each of the three design solutions. Each design should be carefully considered and detailed with pros and cons for each design as well as how the user will interact with the design. The final sheet, detailed in chapter ten, features one of the previous three designs taken forward as a final solution. The sheet should be detailed enough that the product can be created from it.

The final chapter of the book provides two examples and runs through the use of the Five Design-Sheets method with each. The first is a tool for capturing and displaying prehistoric monuments and the second is a tool for visualizing computer algorithms.

2.3.1. Comparison and Recommendations of Academic Books

Spence's book *Information Visualization: An Introduction* is aimed at university students from any discipline who need to visualize data. The book focuses on creating a visual design whilst presenting data visualization principles. Exercises are also given at the end of each chapter. Some unique topics discussed in this book are eye tracking and gaze heat maps as well as alternative canvases. Spence's book is the first academic text book on information visualization and thus features early research work in this field. We recommend this book for readers that would like a concise academic introduction to the field as it is not as comprehensive as some of the other books.

Information Visualization: Perception for Design by Ware provides a comprehensive analysis of human perception and vision with an emphasis on data visualization. Although other books within this survey discuss perception, none are as detailed as Ware's work, and often reference Ware's work themselves. Other topics within information visualization are not discussed in much detail within this book due to its strong focus on visualization perception. We recommend this book as the principal book for vision and perception topics because the author has a deep knowledge and extensive background in the field of vision psychology.

Interactive Data Visualization: Foundations, Techniques, and Applications by Ward *et al.* is more formal than other books in this survey with references to scientific papers and discussions of algorithms. Like some other books in this category, exercises are presented for university students to complete. A topic unique to this book in this category is research directions, where future research topics are explored. We recommend for those with some

prior knowledge of computer science. This book is more technical than other books in this section, with the exception of the focus of Ware's book. This is actually the textbook we recommend as a starting point for university (computer science) students in data visualization. This is because the book is up-to-date and has the most comprehensive view of the field. It is a good starting point.

Yau's book *Visualize This: The FlowingData Guide to Design, Visualization, and Statistics* provides step-by-step instructions on how to create visual layouts, providing coding examples that a reader can follow along with for multiple tools. The book is written in a casual and informal style. We recommend this book for beginners in visualization who are looking to quickly create imagery, with the book focusing on producing visual designs over educational content. We also recommend this for university students that are not necessarily computer science students. It provides more step-by-step guidelines for readers than many of the other books in the category.

Interactive Visualization: Insight Through Inquiry by Ferster provides a framework for creating a visualization design, based on the data visualization principles. A list of digital visualization resources is provided for creating a visual design.

Visualization Analysis and Design by Munzner provides an easy to read and educational introduction to data visualization, with many predominant topics of the subject covered. See Table 2. Of particular use are eight rules of thumb that Munzner presents for the creation of a visualization. In contrast to Ward *et al.*'s book, that works starting with algorithms as its base and moving upwards, Munzner's book starts from the conceptual level and works its way down and stops one level above algorithms. We recommend this book for university students more interested in the conceptual level of visual design.

Kirk's *Data Visualisation: A Handbook for Data Driven Design* provides another perspective for creating a visualization design from concept to realization, similar to Spencer's, Ferster's and Yau's books. An easy and informal style is used. A taxonomy of 49 different visual designs is a useful resource available in this book, however fewer images are used overall. Exercises are available with the book, but only from an online resource. We recommend Kirk's book to students that are less interested in the research aspects of data visualization and are more interested in the contemporary aspects and culture. This is because it draws on many contemporary and popular sources for examples and inspirations such as many web pages, news web sites, and YouTube.

Five Design-Sheets: Creative Design and Sketching for Computing and Visualisation by Roberts *et al.* presents a method for creating ideas, designing, and developing a software tool, with a focus on an information visualization tool. This book differs from the others in this category by not focusing on visualization design principles. We recommend this book to readers who are interested in deriving and brainstorming visual designs at the conceptual level. This book goes into the most depth on how to come up with a range of visual designs centered around addressing a specific challenge. It is very practical in nature and provides detailed guidance on how to brainstorm and even recommends specific materials to use.

Comparison 3. Comparison of Supplementary Material A

Book title	Supplementary material available	Supplementary material description
Visual Explanations [Tuf97]	www.tufte.com	Author website
Visual Complexity [Lim11]	visualcomplexity.com	Showcase of visualization projects
The Functional Art [Cai12]	DVD	Three video lessons
Information Visualization: An Introduction [Spe14]	http://extras.springer.com/2014/978-3-319-07340-8	Supplementary video clips
Interactive Data Visualization [WGK15]	www.idvbook.com	Data sets, images, videos and software
Visualize This [Yau11]	http://www.wiley.com/WileyCDA/WileyTitle/productCd-0470944889.html http://book.flowingdata.com/	Supplementary code
Visualization Analysis and Design [Mun14]	https://www.cs.ubc.ca/~tmm/vadbook/	Supplementary slides, figures, references and author interviews
Data Visualisation: A Handbook for Data Driven Design [Kir16]	book.visualisingdata.com	Exercises, additional reading and case studies on website
The Visual Organization [Sim14]	www.philsimon.com	Author website
Storytelling with Data [Kna15]	http://www.storytellingwithdata.com/	Author podcast, blog and videos
Making Data Visual [FM18]	http://resources.oreilly.com/examples/0636920041320	Supplementary code and video demonstrations
Illuminating the Path [CT05]	http://nvac.pnl.gov	Site expired
Applied Security Visualization [Mar09]	CD	Live CD containing tools
Visualization of Time-Oriented Data [AMST11]	http://www.timeviz.net/	Showcase of visualization projects
Multidimensional Data Visualization [DKŽ12]	Datasets	Dataset details are provided in the appendix
Graph-Based Clustering and Data Visualization Algorithms [VFA13]	www.abonyilab.com	Research group website
Visualizing the Data City [CLS14]	Over 60 URLs provided in appendix	Links to projects and data sources
Graph Analysis and Visualization [BJ15]	www.wiley.com/go/GraphAnalysisVisualization	Data files, sample code
Visualizing Data [Fry07]	http://benfry.com/writing/	Author website
Interactive Data Visualization for the Web [Mur13]	http://alignedleft.com/tutorials/d3	Link to code repository
OpenGL Data Visualization Cookbook [LL15]	https://www.packtpub.com/books/content/support/19795	Code available at website

Table 3: The supplementary material available from reviewed books, where available. Books are ordered in classification order, as can be seen in Table 2.

number of books within this survey complement the printed volume with additional supplementary material, this is especially true for textbooks where five of the eight books provide additional material. Table 3 highlights supplementary material with each publication within this survey. The supplementary material, where available, varies from details of datasets referenced in the books to supporting videos. Supplementary material is usually provided on a dedicated website allowing for corrections to errata as they become apparent and the addition of new material. However websites can sometimes be removed with the loss of the supplementary material as with the web address referenced in "*Illuminating the Path*" [CT05].

2.4. Industry Professional

Seven books are identified in this category, books aimed at professionals from any industry, their covers can be seen in Figure 9.

These differ from textbooks as they do not provide as much background important in a pedagogical context.



Show Me the Numbers: Designing Tables and Graphs to Enlighten by Few [Few12] commences with the purpose of the book which is to enable the reader to produce effective means of displaying quantitative information. The second chapter examines different types of data. The difference between quantitative and categorical data is explored along with different types of relationships, for example nominal, hierarchical and correlation. Statistical methods such as mean, median and standard deviation are also explained. Guidelines for the use of currency data are also given, numbers should be adjusted for inflation when comparing over time and differing currencies should be converted to be uniform.

Chapter three discusses the use of tables and graphs for displaying data and when to use each. Tables are used when precise, indi-



Figure 9: Book covers for industry professional books [Few12, Few06, Sim14, SH14, Kna15, Ber16, FM18] described in section 2.4.

vidual values are required or when more than one unit of measure is used while graphs are used to display the shape of the data or to reveal relationships. A brief history of the use of graphs is also given with references to the work of Descartes [Des37] and Playfair [Pla86]. The use of tables is also discussed in the fourth chapter. Table layouts and designs are explained for different purposes and data.

The fifth chapter discusses human perception of shapes and marks. A description is given of how the eye and brain processes images and how this influences the perception of length, width, size, shape, color and position as a means of displaying quantitative information. The Gestalt principles of visual perception are also discussed. Different methods of encoding data such as points, lines and color are introduced along with different graph relationships such as time series, distributions, and geospatial in chapter six. Graph designs for each type of relationship are presented. An additional section appears at the end of the chapter with six scenarios of different types of data, where the reader is tasked with identifying the most suitable visualization method.

Chapter seven outlines design features of a visualization such as the layout of a figure, the use of titles, legends and labels, and the use of highlighting and emphasis. The design features specific to tables are discussed in the eighth chapter. The use of grids, white space, and fill color to delineate columns and rows, the arrangement of the data, formatting of text and the use of summarizing values are examined. Exercises asking the reader to identify limitations of tables presented are given at the end of the chapter.

Chapter nine meanwhile features general design features of graphs. The importance of quantitative scales that start at zero and of consistent scale spacing is highlighted, along with the interpretation challenges with using 3D layouts. The tenth chapter continues with a description of the format of more specific design features of graphs such as graph marks, points, lines and bars. The format and use of indicator components such as tick marks, annotations and reference lines is also explored.

Displaying multiple variables in one figure is the topic of chapter 11. This is achieved with the use of small multiples. The tech-

nicalities of the layout, sequencing, and general use of the small multiples are studied. Chapter 12 highlights graphs that should be avoided due to their inability to communicate data effectively. These include doughnut charts, radar charts and circle charts along with the reasoning behind their dismissal. The chapter is appended with exercises asking the reader to identify shortcomings with some depicted charts.

Storytelling is the focus of chapter 13. Few presents characteristics that effective stories have in common, for example, they are simple, informative and contextualized. The final chapter concludes that the reader should construct their own design standards that take inspiration from this book.

The appendix contains summaries of when to use which type of graph, recommended reading, answers to exercises from other chapters, how to calculate adjusting for inflation and instructions on how to create a table lens display and box plots in Excel.



Information Dashboard Design: The Effective Visual Communication of Data by Few [Few06] begins with a brief history of Business Intelligence (BI) data visualization and defines an information dashboard: "A dashboard is a visual display of the most important information needed to achieve one or more objectives; consolidated and arranged on a single screen so the information can be monitored at a glance." Few explains that most information dashboards used in business fall short of their potential and provides a number of examples of business dashboards. In chapter two Few identifies three roles for dashboards: Strategic - that provide a quick overview of a business and focus on high-level data; Analytical - providing more detailed data that allow interaction; Operational - for monitoring activities and events that provide instant notifications. Different types of information that are typically used in dashboards are also explored.

Common mistakes in dashboard design is the focus of the third chapter. Examples from vendor websites are used to highlight design mistakes such as, choosing inappropriate graphics, visual designs cluttering displays with decoration, and fragmenting data into separate screens. Human perception is the topic of the next chapter. Visual memory is discussed along with data encoding and the Gestalt principles of visual perception.

The fifth chapter discusses characteristics of well designed dashboards and details techniques to producing effective visual designs, such as reducing graphics that do not contribute to the information to be displayed. Using the best type of visual layout for the information to be displayed is the focus of the sixth chapter. A number of visual layout types are given such as bullet graphs, bar charts, scatter plots, and treemaps. The use of icons, spatial maps, tables, and small multiples are also exemplified.

Chapter seven discusses designing dashboards for usability. Effective organization of the information is considered along with making the dashboard aesthetically pleasing and maintaining consistency of colors and interaction. User testing of the design is recommended. The final chapter provides examples of dashboards in different scenarios. Examples dashboards are given for telesales, marketing, sales, and for use by a chief information officer. Eight examples of sales dashboards are also given for critique.



The Visual Organization: Data Visualization, Big Data, and the Quest for Better Decisions by Simon [Sim14] introduces the benefits of data visualization in the modern

context of big data and emphasizes that most businesses do not utilize visualization.

The first chapter discusses the changing landscape of IT. Changes in the way people interact with IT and how technology companies are adapting by creating new tools and collection data. Data visualization tools for businesses are explored in the second chapter. Dedicated visualization tools such as Tableau, Microsoft Excel, and visualization tools incorporated within business intelligence software are examined. The ability of visualization tools to integrate with databases and statistical software and their ease of use are discussed. Open-source tools such as D3 and R are also discussed.

The third chapter studies the TV and film streaming site Netflix and its use of data visualization. The extent to which Netflix utilizes the data it collects to provide insights into its customers and to improve the service it offers is examined. The online polling company Wedgies is the subject of the fourth chapter. The company's use of Google Analytics, D3, and other open-source software is highlighted.

Chapter five meanwhile features the University of Texas System and how they use SAS to visualize their data. The sixth chapter introduces four levels of organizations categorized by the data that they visualize. The first level organization visualizes static, small scale data, the second level utilizes interactive visualizations on small datasets. The third and fourth levels utilize big data sets using static and interactive techniques respectively. Progression of organizations up this scale is discussed.

Chapter seven provides an example of human resource visual design at Autodesk. A visualization shows the personnel changes over time, produced using Java, Processing and Graph Viz. The broad outline to utilize visualizations within business is presented in the eighth chapter. Obtaining the data, processing it and utilizing available tools such as Hadoop or Amazon Web Services is the first step. Considering what kind of visual layout is required for individual scenarios is an important factor along with visualization design, and user experience. Creating a culture of data analysis and visualization within an organization is also important.

The penultimate chapter looks at common pitfalls for organizations that may try to utilize visualization. Common visualization mistakes, the importance of continuous development and tailoring to the visualization audience are explored.



Presenting Data: How to communicate your message effectively by Swires-Hennessy [SH14] outlines some principles for using numbers when presenting numeric information, such as consistently using the same symbol for decimal separators, right justifying numbers in columns, and using a typeface that has a constant width for all digits. Details are also given on how to effectively round numbers. Presenting numeric data in tables is the subject of the second chapter. A series of principles are described for the effective use of tables such as putting totals at the bottom of columns and to the right of rows, and putting data for comparison into columns.

Chapter three discusses the use of charts. Principles are outlined for the correct use of bar charts, histograms, pie charts, and graphs. Multiple examples of charts are given along with how they could be improved, for example by ensuring axes start at zero. Numbers in text is the subject of the fourth chapter, such as in headlines or

statements. Principles are outlined for their use such as ensuring the correctness of the data, using plain language, and ordering messages in terms of importance.

The final chapter discusses using the internet as a form of sharing data. Some history of national statistics data on the internet is given along with some examples of current designs such as InfoBase Cymru [Dat].



Storytelling with Data: A Data Visualization Guide for Business Professionals by Nussbaumer Knaflic [Kna15] begins with examples of inadequate visual designs. The first chapter, discusses the importance of understanding the purpose of a visual design and to consider the audience before its creation. Chapter two describes different forms of graphic visualizations such as bar graphs, scatterplots and line graphs. Examples of charts to avoid are also given such as pie charts and 3D graphs.

The third chapter discusses the issue of clutter, defined as visual elements that don't contribute to an increased understanding of data. Gestalt principles of visual perception are explored and used in an example where clutter is removed from a scatter graph. Chapter four begins by discussing human perception and memory and how these can be used to draw the attention of an audience to a particular element within a visual design.

The fifth chapter discusses the use of design principles for creating a visual layout. Key considerations are reducing clutter, highlighting important elements, avoiding overcomplexity, textual annotations, and aesthetics. Five examples of model visualizations are given in chapter six that are analysed for the design choices and that highlight the principles outlined so far in the book.

Chapter seven discusses how to build an effective story, using a beginning, a middle, and an ending. The order of the narrative of the story is also discussed. The eighth chapter emphasises principles from the previous chapters and gives an example of a series of line graphs that tell a story of how prices have changed over time.

Chapter nine addresses five situations that have not been addressed so far in the book, visualizations with a dark background, producing a static figure of an animated visualization, ordering in categorical data, avoiding over complex line graphs, and alternatives to pie charts. The final chapter summarizes the principles from the book and provides some extra tips such as learning the visualization tool, producing draft iterations of visual designs and use of feedback, and to look at other examples of layouts.



Good Charts by Berinato [Ber16] begins with an introduction to data visualization and how it is becoming an increasingly important skill to provide insightful and relevant designs. The first chapter details some of the history of data visualizations, from works published by Playfair [Pla86], through the work of Bertin [BBS67] and Tufte [Tuf83, Tuf90, Tuf97] to the latest research by Rensink and Baldrige [RB10] and Harrison *et al.* [HYFC14]. Berinato discusses what features tend to stand out in a visualization, such as a peak on a line graph, and how too much information makes it difficult to find individual data points. The second chapter also discusses the importance of using common conventions, such as red is hot, blue is cold and north is up.

Four types of visualizations are identified in the third chapter: conceptual, declarative visualizations for idea illustration; conceptual, exploratory visualizations for idea generation; data-driven, exploratory visualizations for visual discovery; and data-driven

declarative for presentation. The skills and tools required to produce the visual design can be determined by first identifying the type of visual design. The fourth chapter describes an hour long process for developing a visual design; the first five minutes are for preparation, the next 15 are to discuss - to establish the purpose of a visual design and what to highlight, 20 minutes are for sketching ideas, and the final 20 minutes for creating a prototype.

Refining visual layouts to make them visually appealing is the topic of the fifth chapter. Using a fixed structure, aligning elements, removing unnecessary elements, and correct use of colors are all discussed. The next chapter discusses how to emphasize particular data within a visual design to make a statement. Emphasizing data using color, labels, and markers to draw attention, removing elements that distract, adding elements that support and shifting reference points are examined.

Chapter seven features visualization techniques that create misleading visual designs. These are the use of a truncated axes, using multiple axis on the same plane, and representing quantitative data on geographical maps. The eighth chapter examines verbal presentation of visualizations and storytelling. Guidance for presenting are given such as not to read the picture, using reference charts, use of simple charts, use of storytelling, and to avoid talking when first presenting a chart.

The final chapter encourages the reader to find published visual designs and preform a critical review of them. Three examples of critiques are provided.



Making Data Visual by Fisher and Meyer [FM18], a new book, commence with an example data set to outline why visualization is useful, to facilitate discovery. The second chapter focuses on developing a concrete visualization task from a broad question about a data set, exemplified with the question "Who are the best movie directors". The question is broken down using some visually derived assumptions to actionable tasks.

Chapter three details how to interview and communicate with domain experts and stakeholders to establish what questions and assumptions can be made to create insightful visualizations of a data set. Data types are the focus of the fourth chapter with three principle types identified: continuous data, ordinal data, and categorical data. Three additional specialized data types are also identified: temporal data, geographical data, and relational data. Techniques for processing the data, such as transforming between data types and dimension reduction, are also briefly discussed.

The fifth chapter explores different chart types and explains what charts are suitable for what task and what data type. These charts include scatterplots, histograms, bar plots, node-link views, treemaps, and word clouds. Multiple linked views is the subject of the sixth chapter, which explores visualization techniques such as small multiples, scatterplot matrices, and dashboards. Some considerations for multiple views are outlined such as consistency of scales, and ensuring interaction, such as filtering, is mirrored across views.

A case study is presented in chapter seven using a real world example that one of the authors personally worked on. The author recounts the iterative process of creating a business intelligence dashboard based on software telemetry. A second case study is outlined in the penultimate chapter focusing on visualization of fruit fly genetics. The case study is based on another author's experience,

demonstrating the approach outlined within the book and making particular use of multiple linked views.



We cite another book that fits in this category, **Creating More Effective Graphs** by Robbins [Rob12]. The book details how to choose and create appropriate graphs for many different situations. We don't include a full summary due to space limitations.

2.4.1. Comparison and Recommendations of Industry Professional Books

Show Me the Numbers by Few [Few12] provides detailed descriptions of the principles of data visualization for a beginner to get started. Unlike books classified as textbooks, the book is targeted towards professionals required to present data in industry settings. Because of the professional focus of the book, it includes presenting data using tables as well as more advanced visual designs. We recommend this book for business professionals seeking best practice for presenting data in an industry setting. The hardcover printing of this book is quite big compared to the others. The visual designs featured are not generally interactive and are more at the beginner level.

Few's second book in this survey *Information Dashboard Design* [Few06] focuses on the design on visual dashboards. This is a unique topic within this survey, however some common subjects are also included such as perception. We would recommend this book for people developing visual interfaces who already have basic knowledge of data visualization. Again the visual layouts are not generally very interactive and the visual designs are at the beginner's level. The book also contains a lot of white space.

The Visual Organization by Simon [Sim14] provides a discussion of the use of data visualization in industry and how proper implementation can lead to success. The book is written in an entertaining and engaging manner. This book is suitable for a business manager or leader as a demonstration of how effective the use of data visualization can be. This book does not contain as many images as the other books and also features some infographics themes. The examples used are at the beginner's level.

Presenting Data by Swires-Hennessy [SH14] provides formal instructions for best practice in the presentation of data. A strong focus is placed on the use of tables and spreadsheets however a section is dedicated to charts. Rules are provided for producing simple, truthful, charts. We recommend this book for business professionals who would like a very compact and concise introduction to visualization. The visual designs are basic and generally not interactive.

Storytelling with Data by Knaflic [Kna15] provides an informal guide for creating engaging visualizations. The book is more concise than Few's *Show Me the Numbers* and provides a more creative viewpoint. Presenting visualization aspects incrementally to give a narrative is also discussed. We recommend this book for professionals who are required to produce persuasive visualizations in their work. The example visual designs are basic and not interactive. Presentation is the focus. It's easy to read and is targeted for all industry professional backgrounds.

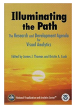
Good Charts by Berinato [Ber16] focuses on developing and improving visualizations skills. This book features the most images, examples, and figures of those in this category. Figures are used to demonstrate good practice and guidance is given on how to go about creating visualization ideas. The book provides an easy to read style and uses casual language, but also has a very awkward landscape orientation. We recommend this book to professionals seeking a wide variety of sample visual designs in a concise browsing fashion.

Making Data Visual by Fisher and Meyer [FM18] outlines a method for establishing how to develop visual designs for the visualization of a particular data set. The book details of how to interrogate stakeholders to establish how the visual design will be used and what insights are required. The iterative process of developing a visualization design to solve a particular problem is demonstrated with two case studies. The book is written in an easy to read, informal style and only covers topics discussed in a brief fashion without providing any complex topics.

Comparison 4. Comparison of Online Review Scores Figure 10 presents the Amazon.com review score of the summarized books within this survey. Only books with at least one review are included in the figure, this excludes nine books from the visualization. Notable from the image is that no books have a review score of less than three. One book, by Vathy-Fogarassy and Abonyi [VFA13], receives the maximum score of five, although it should be noted that this is from only one review. Tufte's *The Visual Display of Quantitative Information* [Tuf83] has the most reviews, with 270, with an average score of 4.4. Another popular book is *Storytelling with Data* by Knaflic [Kna15] with 226 reviews and an average score of 4.5. This is despite having a relatively recent publishing date in 2015.

2.5. Special Focus

Twelve books have been identified for this category which features books focused on more specific directions within data visualization. See Figure 11. Due to the specialized focus of the books in this category, books generally have a lower Amazon.com sales rank, although the precise way in which the rank is calculated is unknown. The eight books with the lowest sales rank are in this category. The books within this category also tend to be the most expensive.



Illuminating the Path: The Research and Development Agenda for Visual Analytics by Thomas and Cook [CT05] is published as a research agenda for visual analytics as a response for the need for better data analysis tools to help counter terrorism. The book is sponsored by the US department of homeland security.

The first chapter explores potential terrorism threats and the challenges of analyzing vast amounts of data to counter the threat, while still allowing for personal freedom. The need for advancements in visual analytics to address these challenges is highlighted. Chapter two discusses the use of visual analytics as a tool to support analytical reasoning. The process of analytical reasoning is explored, and the inclusion of visual analytics to facilitate perception and understanding of data and to enable rapid decision making.

Visual representation of data and interaction is discussed in the third chapter. Scientific principles for data visualization are explored along with interaction methods. The use of the appropriate visual design for specific datasets and queries are considered along with their scalability with increasing data. The fourth chapter discusses data and data processing from raw format to representations that may provide more insight. Different forms of data are discussed such as video, image, text, and geospatial data along with their characteristics.

Chapter five highlights the importance of accurate communication of visual analytic results. Existing tools are reviewed and the automation of effective visualization generation is discussed. A framework for moving visual analytics research into practical use is presented in the sixth chapter. Four areas are identified as requiring attention: evaluation – to address if a new tool or technique is effective; privacy and security – to support approaches such as anonymization to protect privacy; system interoperability – to avoid incompatibility with other tools; and technology insertion – the widespread adoption of the tool or technique.

The final chapter discusses the requirements and need for meeting the agenda set out in the book.



Interactive Visualization of Large Networks: Technical challenges and practical applications by Van Ham [Ham08] begins with a chapter outlining the objective of the literature, how computer supported visualizations can be used to help users gain insight into the content and structure of large networks. The second chapter defines a simple graph, a series of nodes connected by edges, and different forms of graphs. Considerations are made from the aim of a user when visualizing a graph to what constraints may exist to the creation of a visualization.

Previous work to design graphs is outlined in the third chapter. This is subdivided into three topics, graph drawing, clustering and visual techniques such as treemaps and interaction. Chapter four acts as an introduction to the next four chapters, outlining four different graph problems for which visual designs are created.

The fifth chapter introduces Beamtrees [VHvW03], a visualization based on treemaps that uses perceptual cues to clarify its structure. Details of how to construct a Beamtree and how to calculate the size of each node are presented, along with a user evaluation of the method. A method for visualising state transition graphs is presented in the sixth chapter. The technique relies on clustering states to enable the visualization of large numbers of nodes that provides a global view of the structure of the graph.

A visual layout that enables interactive visualization of a large graph structure with hierarchy is presented in the seventh chapter. An example of the exploration of a software engineering project is used as an example throughout the chapter. The eighth chapter presents a method for conveying small world graphs. The layout utilizes clustering to minimize clutter on a high abstraction level, while focusing on a particular cluster allows for detail to be seen.

The penultimate chapter reflects on each of the four projects outlined in the previous chapters, comparing their flexibility and scalability. Common aspects between the four visualizations are discussed and insights gained from the projects are presented.

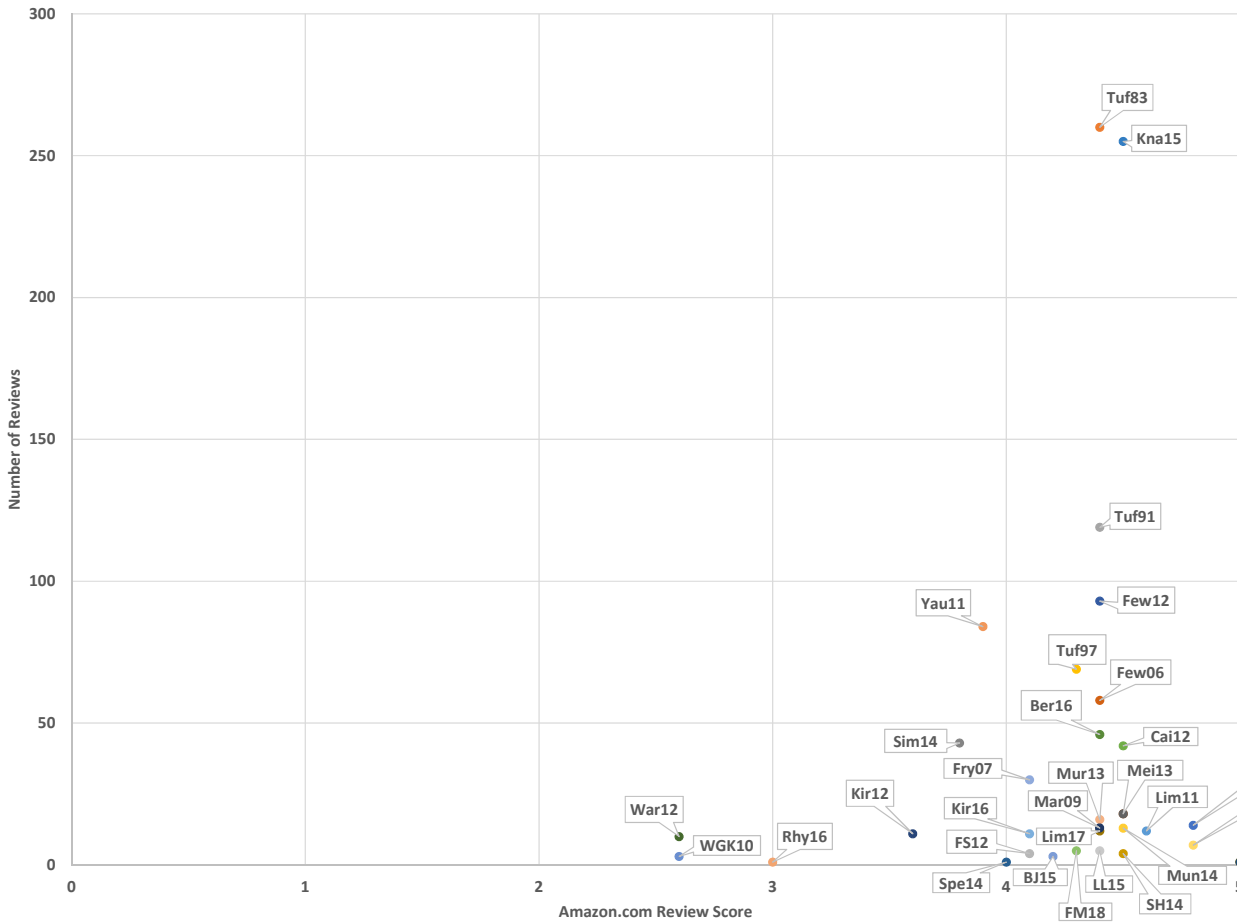


Figure 10: Average review score against the number of reviews on Amazon.com, for all summarized books within the survey with a review, discussed in Comparison 4.



Applied Security Visualization by Marty [Mar09] introduces data visualization and its benefits in particular within the computer security sector. Perception is briefly discussed along with some graph design principles such as annotating data and showing comparisons. The second chapter discusses security data sources, predominantly log files, the parsing of the data and potential issues such as missing log entries. Specific security data sources are discussed and how to retrieve them such as network traffic, firewall logs and operating system logs.

Methods of visually representing the data are discussed in the third chapter including the use of color, shape and size. A number of visual designs are described, for example bar charts, line charts, treemaps, node link diagrams and parallel coordinates, and the kind of data each can display. The use of 3D charts is also discussed along with pitfalls such as occlusion and the use of interaction and animation to overcome these. Chapter four introduces the visualization pipeline as six steps. The six steps are: define the problem, assess the data, process the data, visual transformation, view trans-

formation and refine and iterate. An example of web traffic analysis is used to demonstrate the process. A brief description of data processing tools is also presented.

The fifth chapter applies the principles from the previous chapters to analyse security data. Three broad categories of analysis are identified, reporting, historical analysis and real-time monitoring. Time-based layouts, interaction and forensic analysis are discussed along with examples meanwhile discussion about real-time monitoring is predominantly centered around data dashboards. Examples analyzing threats to network perimeters are given in chapter six. Examples include traffic flow analysis using parallel coordinates and vulnerability visual designs using treemaps.

The seventh chapter discusses the use of visualization for compliance such as with auditing, compliance monitoring and visualizing risk. The penultimate chapter discusses the use of visualization for detecting threats from inside a network, in particular for showing user activity.

The final chapter presents a range of visualization tools.



Figure 11: Book covers for books classified as special focus [CT05, Mar09, Ham08, AMST11, Wil11, DKŽ12, VFA13, CLS14, BJ15, Rhy16, CFV*16] described in section 2.5.

Visualization of Time-Oriented Data by Aigner *et al.* [AMST11] begins with an introduction to time-oriented data along with the special associated challenges such as recognising periodic cycles. A formal introduction to visualization is also given with academic references describing the process of creating an interactive visualization. An application example is given of VisuExplore, an interactive medical history visualization tool [RMA*10]. Historical depictions of time-oriented data are explored in the second chapter such as William Playfair’s work [Pla86]. Time in storytelling is also discussed such as the use of comic strips and movies as well as music and paintings.

Chapter three characterises time by its scale, scope and arrangement and also looks at the types of data that can be represented. The considerations of creating a visualization from data are discussed in the fourth chapter by asking what, why and how data is presented. The visual representation of time is explored along with representation of data, both spatial and abstract, and how to encode both.

Different forms of interaction are first identified such as selecting and filtering before some interaction principles are outlined in chapter five. Methods of interacting are discussed such as direct manipulation and brushing, as well as automatic methods such as event-based visualization. Analytical techniques for data mining are discussed in the sixth chapter. Analysis tasks for time-oriented data are for classification, clustering, search & retrieval, pattern discovery and prediction. Data abstraction is discussed along with principle component analysis.

Chapter seven presents a survey of over 100 visualization tech-

niques, categorized by data characteristics. Techniques range from simple point plots to visualization techniques developed for a specialist field. The final chapter presents future research challenges for time oriented data.

Visualizing Time: Designing Graphical Representations for Statistical Data by Wills [Wil11] begins by emphasising the importance of time and of creating visual designs for it, citing Stonehenge as an example. William Playfair’s work [Pla86] and Minard’s figure of Napoleon’s march on Russia [Min69] are other historical works of time-oriented layouts that are studied, along with comic book style depictions. The second chapter introduces visualization concepts, briefly describing visual primitives, data processing steps, aesthetic considerations, mapping data, and interaction concepts.

The process taken to design a visualization is the subject of the third chapter. Wills introduces the Goal Question Metric [VSB-CR02] approach to visualization design; first the user goal needs to be established, questions are then formulated, the answers to which will achieve the goals, finally the data is mapped to the visual primitives. Chapter four discusses different formats of date and time, time intervals and different scales of measurement.

Mapping time to an axis is the topic of the fifth chapter. As time is an independent variable it is usually placed on the horizontal axis. Techniques for plotting events and sequenced data are discussed. The sixth chapter discusses using non-Cartesian coordinate systems, labelling time to an axis, and using small multiples.

Mapping time to another attribute such as color or size is the subject of the seventh chapter. An example dataset of medieval soldier activity is used to demonstrate the different techniques. A further example is used to demonstrate ordering within text-based data. Chapter eight discusses the distortion of the time axis. This enables for better representation of some time-based data enabling emphasis of a particular element. Techniques discussed include changing from a linear time scale to an ordered time scale and using frequency analysis.

Ways of interacting are discussed in chapter nine. Modifying different parameters of such as axes, mapping primitives, coordinates, and scales are explored along with brushing techniques. Chapter ten exemplifies problems encountered with specific datasets. Methods for dealing with large datasets such as aggregation are examined along with techniques for visualizing timelines and linked events.

The final chapter discusses the complexity of visualizations and a formula is introduced to derive and quantify complexity based on the number of variants and primitives mapped. All figures used in the book are then ranked according to the formula.

Multidimensional Data Visualization: Methods and Applications by Dzemyda *et al.* [DKŽ12] defines what data is considered multidimensional, i.e. a dataset with multiple features. Two method groups of multidimensional data visualization are identified, direct visualization methods – the plotting of each feature; and projection – where dimensional reduction techniques are used. Chapter two describes in further detail the visualization methods mentioned in the previous chapter. Direct visualization methods include geometric visualizations such as scatter-plots, multiline graphs, parallel coordinates and RadViz; glyph display methods such as Chernoff faces and star plots; and hierarchical displays methods of dimensional stacking and trellis

display. A number of dimension reduction methods are also presented including principle component analysis, linear discriminant analysis, multidimensional scaling, manifold based visualization, and local linear embedding.

The third chapter explores multidimensional scaling in further detail. Multidimensional scaling describes a range of techniques that reduce the dimensionality of a dataset with the aim of retaining the dissimilarity between any two objects. Different methods of optimizing the dissimilarity measure are investigated. The concept of artificial neural networks is explained and the use of a self organizing map, neural gas and SAMANN networks with multidimensional scaling is discussed in chapter four.

The final chapter provides examples of the application of the techniques discussed in the book across many different fields.



Graph-Based Clustering and Data Visualization Algorithms by Vathy-Fogarassy and Abonyi [VFA13] commences with an examination of vector quantisation algorithms that can be used to convert complex data into connected graphs. Techniques explained include growing neural gas vector quantisation, dynamic topology representing network and the weighted incremental neural network. The second chapter discusses two graph-based clustering algorithms and introduces novel techniques to improve their performance. The first clustering algorithm discussed is the minimal spanning tree method. Vathy-Fogarassy and Abonyi introduce a modification by removing inconsistent edges and utilize the Gath-Geva algorithm [GG89] to cluster the results. The resultant algorithm is titled the Hybrid minimal spanning tree-Gath-Geva algorithm and utilizes a fuzzy hyper volume validity measure.

The final chapter provides an overview of algorithms that unfold and represent the topology of data, in particular the Isomap, Isotop and Curvilinear Distance Analysis methods. An algorithm that combines the topology representing methods and multidimensional scaling is studied and analysed. The appendix provides details of the various algorithms described throughout the book.



Visualizing the Data City: Social Media as a Source of Knowledge for Urban Planning and Management by Ciuccarelli *et al.* [CLS14] researches and analyses geo-located social media data mining from an urban development standpoint. The second chapter emphasises the possibilities of using geo-located social media to characterise urban uses and issues. Some existing projects are reviewed, and an aim of producing a platform that can visualize the data is proposed.

Chapter three presents a taxonomy of state-of-the-art geo-located data visualization techniques that deal with large datasets and with real-time interaction. Some 56 projects are analysed and categorized according to the data topic and by project goals. The fourth chapter describes how the authors created a framework for investigating urban interest topics using geo-located social media data. A discussion of the evaluation of the work is also given.

Telltale and Urban Sensing, two projects that visualize geo-located social media data, are presented in the penultimate chapter. Examples of visual designs and data analysis performed using the two projects are presented, such as using twitter to track individual movements, and mapping cultural distribution through the language used in social media. The final chapter presents a discus-

sion on the potential of using geo-located social media analysis and its shortcomings.



Graph Analysis and Visualization: Discovering Business Opportunity in Linked Data by Brath and Jonker [BJ15] first describes the importance of data visualization with particular emphasis of the use of graph-based visual designs. Multiple historic examples are given of the use of graph visualizations and the benefits these visual designs provided, such as managing network supply chains and mapping social hierarchies. The second chapter identifies different forms of graphs and their uses in representing relationships, hierarchies, communities, flows and spatial networks.

Sources of data to be visualized are identified, such as flight stats or Tweets, and methods of obtaining the data are discussed in chapter three. Processing the data so that only relevant information is included and file storage methods and formats are discussed such that information can be easily imported into graphing software. Chapter four discusses meta-data of graphs as the number of nodes and edges, and centrality as an indication of the types of relationship in the graph. Based on this understanding and the objective, different graph layouts are also outlined.

The fifth chapter discusses mapping different visual attributes to a graph. The use of color, labels, font, line weight and size is discussed along with the merits of the use of each. Interaction techniques such as zooming and filtering are explained and the use of legends, annotations and sequences to help communicate insights are studied in chapter six.

Chapter seven provides an overview of the capabilities of five different graph creating tools, namely Excel, NodeXL, Gephi, Cytoscape and yEd. Meanwhile chapter eight describes how graph layouts can be constructed using the Python environment and the JavaScript environment using the D3 libraries. See section 2.7.2.

The ninth chapter discusses the relationship connections within a graph, when they should be kept separate and their aggregation. Applications of changing the connections are provided such as actor relations across films. Hierarchies are discussed in the tenth chapter, represented as trees and treemaps.

Techniques for identifying and analysing clusters within graphs are discussed in chapter 11. An example of a Facebook discussion about the Toronto Raptors NBA team is used to demonstrate some of the techniques such as identifying cliques. Chapter 12 discusses flow based diagrams such as Sankey diagrams and chord diagrams and gives detailed instructions on how to create an example of each using JavaScript and the D3 library.

The 13th chapter explores spatial networks, where where physical characteristics need to be preserved. The London Underground network map is given as an example. Rose link diagrams and grouping of small worlds are also explored. Chapter 14 discusses issues encountered when dealing with big data. The use of database and query languages for exploration of structures within the data is discussed and detailed examples of subtracting data for visualization in Gephi is given. The use of tiled approaches for visualizing scalable large datasets is also discussed.

Visualizing graphs that change with time is discussed in chapter 15. Different techniques for achieving this are discussed including the use of animation, small multiples, focus+context and an additional spatial dimension. The final chapter studies design choices

for a graphs such as the shape and size of nodes and links, and the use of iconography. The use of an appropriate color palette is also examined.



Applying Color Theory to Digital Media and Visualization by Rhyne [Rhy16] introduces the three color models, the RGB model of light and display, the CMYK model for printing, and the RYB model for paint. A history of the development of color theory and an example of color theory application is also given. The second chapter details how human vision perceives color and luminosity, and discusses various color vision deficiencies and how to apply color theory to compensate for this. Problems with using a rainbow color map are also highlighted.

The color gamut and the color spaces that are produced from different displays and media are discussed in the third chapter. Various color spaces are explained such as the sRGB, and the HSV color spaces. Color harmony, defined as the colors that work well together in the composition of an image, is the topic of the fourth chapter. Color wheels for the color models are reviewed and color harmonies between hues on the wheels are examined.

Chapter five reviews eight different online and mobile tools for producing color schemes. These tools include ColorBrewer 2.0 [HB03], Adobe Color CC [Ado18], and PANTONE Studio [Pan18]. The sixth chapter provides three case studies on the use of color theory to produce color schemes for visual designs. The case studies consist of a biological dataset, a geospatial dataset, and a hurricane animation.



Information Theory Tools for Visualization by Chen *et al.* [CFV*16] begins by describing basic concepts of information theory, such as entropy, conditional entropy, and mutual information. The second chapter details how information theory can be applied to data visualization, as the communication of information from data, through visual representation, to perception.

The application of information theory in determining the best viewpoints for scientific visualizations is demonstrated in the third chapter. Further examples of the use of information theory within volume visualizations are given in chapter four. Example problems include transfer function design in direct volume rendering and iso-surface extraction.

Chapter five provides examples of the use of information theory in flow visualization. The concept of an entropy field is introduced and applications of its use for streamline analysis and seeding. The final chapter discusses the use of information theory in information visualization. Among the applications of information theory are the definition of a quality metric of visualizations and measures to analyse the relationships of variables in multivariate datasets.



User-Centered Evaluation of Visual Analytics by Jean Scholtz [Sch17] commences with an outline of the objectives of the book, to provide assistance to researchers in visual analytics on conducting formal user-centered evaluations. The second chapter provides a brief history of intelligence analysis in the United States of America as well as a discussion on critical thinking.

Analytical methods are discussed in chapter three, in particular

the activities involved with sensemaking. Chapter four describes why visual analytics is needed, citing Thomas and Cook's Illuminating the Path [CT05]. Some visual analytics tools such as Tableau are also discussed.

The evaluation of the ease of use of software is discussed in the fifth chapter, and how it can be used with visual analytics tools. Different Human-Computer Interaction (HCI) evaluation methods are also presented such as the use of guidelines, usability studies and A/B studies. The sixth chapter discusses the HCI evaluation needs specific to the visual analytics field. Modifications to the HCI methods presented in the previous chapter are presented, that make them more applicable to visual analytics. The use of metrics to measure effectiveness of tools is also discussed.

Chapter seven examines the current use of evaluation in visual analytics systems, and cites work by Isenberg *et al.* [IIC*13] to determine the use of human participants in evaluations of published conference proceedings. Proposals by Sedlmair *et al.* [SMM12] of a methodology of developing visualization software along with evaluation of its use, are also cited. The penultimate chapter identifies trends in the visual analytics field in particular in collaborative work and in the use of streaming data.

This book is the most recent book in the Synthesis Lectures on Visualization series, another seven books are available in this series which we only reference in the interest of content management [FGKR16, End16, SP16, Hur15, Tom15, Mac11, LM10].

2.6. Comparison and Recommendations of Special Focus Books

Due to the varying subject matter of the books within this category, comparisons and recommendations are not suitable across very different topics. However three books categorized in this section discuss network visualization, and another two books discuss visualization of time therefore it is logical to compare these.

Van Ham's *Interactive Visualization of Large Networks* [Ham08] is a published PhD thesis therefore is recommended for experienced data visualizationists and researchers. Similarly Vathy-Fogarassy and Abonyi's *Graph Based Clustering and Data Visualization Algorithms* [VFA13] features published research work and is therefore recommended for researchers. This book is part of the Springer Briefs in computer science series and is thus printed in black and white. Both books discuss different subject matters within network visualization, with Van Ham's work concentrating on visualizing network structure and Vathy-Fogarassy and Abonyi's publication focusing on clustering algorithms.

Graph Analysis and Visualization by Brath provides a beginners guide to using graphs to analyze relationships between data. The book is aimed at analysts and data scientists, with many examples taken from a business environment. Network visualization tools are explored including some coding examples using Python and D3. We would recommend this book to beginners in data visualization with linked data to visualize and also those that would like a comprehensive introduction to graph analysis and visualization. It is also more applied in nature than the other two graph books.

Both Wills' *Visualizing Time* and Aigner *et al.*'s *Visualization of Time-Oriented Data* examine the visualization of time with many

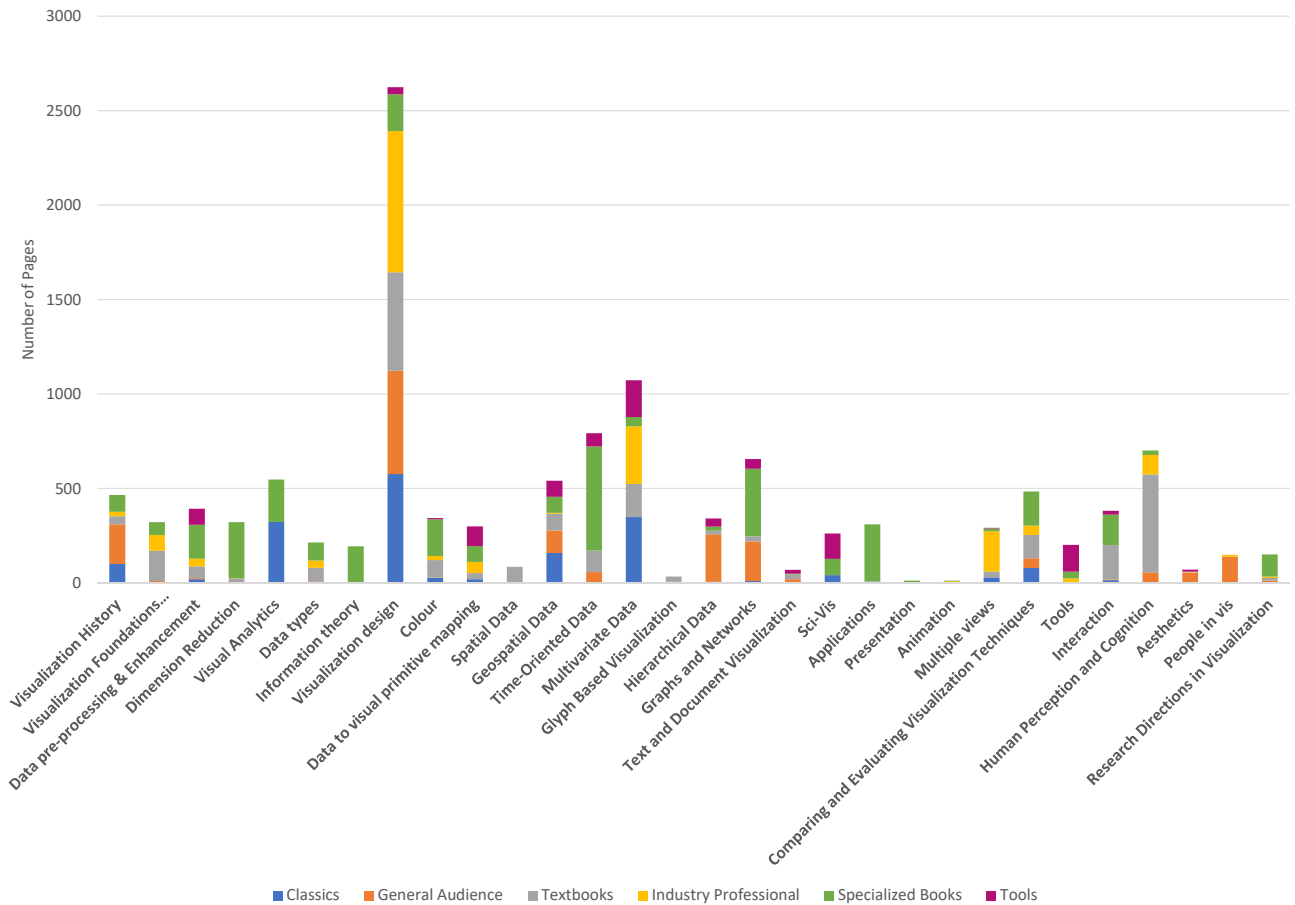


Figure 12: Total number of pages, from surveyed books, dedicated to each topic, colored by book classification. Analysis of this Figure can be found in Comparison 5.

overlapping topics between the two books such as interaction. Wills’ book however has greater emphasis on creating a visual design with data mapping being a strong subject. Meanwhile Aigner *et al.* provide greater prominence to the use of visualization for data mining and analysis. Aigner *et al.* also present over 100 examples of time visualizations categorized by data characteristics, useful for inspiration. We recommend Wills’ book for beginners requiring to visualize time. For people requiring visualizations for data analytics purposes, we recommend the book by Aigner *et al.*, and for individuals creating visualizations for presentation and storytelling we recommend the book by Wills. Aigner *et al.*’s book is more appropriate for researchers undertaking time-oriented visualization and provides a survey of the research landscape.

Comparison 5. Comparison of Topics Across Books Figure 12 indicates the number of pages dedicated to each topic from all of the focus books within this survey. Notable is the number of pages dedicated to the visual design of a visualization, which is spread over most classifications of books apart from tools. The topics with the fewest dedicated pages are glyph based visualization, presentation and animation. Books classified for general audiences and

classics predominantly discuss the same topics, where as textbooks and special focus books focus on specialist topics.

2.7. Information Visualization Tools

Books in this category are instructional books for specific tools. Due to the large number of instructional books for software tools, only one recent book for each tool is reviewed whilst other related books are referenced. The books are summarized for this category cover Processing, D3, and OpenGL, see Figure 13.

2.7.1. Processing

The Processing language is the first visualization tool that we survey. In order to keep the survey succinct, we only survey Fry’s Visualizing Data: Exploring and Explaining Data with the Processing Environment [Fry07]. Many other books that present information on Processing environment are available, we list five recent ones in Table 4 along with meta-data.



Figure 13: Book covers for books classified as tools [Fry07, Mur13, LL15] and described in section 2.7.



Visualizing Data: Exploring and Explaining Data with the Processing Environment by Fry [Fry07] identifies seven stages for visualizing data; these are Acquire, Parse, Filter, Mine, Represent, Refine and Interact. An example dataset of zip code data is used to demonstrate the seven stages. The second chapter introduces Processing, a programming language for creating visual designs. Some basic commands for using Processing are given such as drawing primitives, loading data and mouse interactions.

Mapping data to visual space is the topic of chapter three. Detailed Processing code of an example of placing data on a geographical map is given with data attributes mapped to size and color. The next chapter focuses on visualizing data that changes over time. An example data set is used to demonstrate importing data into Processing and creating a labelled scatter graph, line graph and bar chart.

Chapter five uses baseball data from MLB.com to create a visual design. Detailed Processing code is given on how to read and process the data and to create a connection to the data. A Perl script is also given to enable for the image to be shown in a web browser. An example of showing US zip codes on a geographical map is used in chapter six. The seven stages identified in chapter one are followed and Processing code is given for the various stages to create interactive visual layouts.

The seventh chapter discusses using recursion programming in Processing to create hierarchical visualizations such as treemaps. An example is given with Processing code that creates a Treemap of the file structure of the PC. The eighth chapter provides Processing code to create graphs. The integration of Processing with the Eclipse IDE is also demonstrated to create an interactive graph diagram of website traffic.

Chapter nine presents various techniques of acquiring and pre-processing data, including storing data in databases. Parsing Data is the subject of the tenth and penultimate chapter. Different formats of input files are discussed along with techniques of extracting the most relevant data. The final chapter serves as a Process reference for integration with the Java programming language.

2.7.2. D3

D3 [BOH11] is a JavaScript library for producing web based data visualizations. We found 23 books that introduce and describe how to use the D3 library, five recent examples can be found in Table 5. We survey one recent example for this paper [Mur13].

Ref	Number of pages	Amazon Sales Rank	Google Scholar Citations	Price (USD)
[RF07]	672	98,445	502	59.66
[Shi09]	564	51,476	93	41.06
[BGLL12]	472	181,654	44	50.63
[RF15]	238	39,897	104	16.62
[NN17]	576	1,813,622	0	59.24

Table 4: Meta-data for five recent Processing books



Interactive Data Visualization for the Web by Scott Murray [Mur13] begins with a compilation of tutorials for beginners in D3. In the second chapter Murray introduces the JavaScript library D3, its purpose, its benefits, and its limitations. Alternative tools are also described.

Chapter three briefly describes the technologies behind web hosting and offers a beginners guide to HTML, CSS, JavaScript and SVG. The fourth chapter describes how to setup D3 on a PC or Mac in order to begin coding.

The process of importing data into D3 is described in chapter five. Details are also given on how to check that the data has been properly read, and how to load the data into functions. Chapter six provides detailed code and explains how to draw basic bar graphs, bubble plots, and scatterplots using D3 and SVG. The inclusion of labels is also described.

The details of how to scale data is described in the seventh chapter. The use of the `d3.min()` and `d3.max()` functions are described in order to establish the range of the data and therefore establishing the scale required. Alternative methods of scaling, such as logarithmic are briefly described. Chapter eight describes how to add axes to a scatter plot.

The ninth chapter describes how to update visual designs already created with updated data. Details are also given on how to effect the transition of the visualization from one dataset to another, and how to add additional data and remove data from a visual layout. Interactivity, the tenth chapter guides the reader on how to highlight elements when hovering over with a mouse and how to sort elements. The application of tooltips is also described.

Chapter 11 examines different layouts in D3. Pie, stack, and force layouts are described along with their application. The 12th chapter looks at creating geospatial visualizations using D3. Importing GeoJSON files and mapping data to geographical elements are explained.

The final chapter explores the options for exporting results. Three methods are described, using print screen, printing to a .pdf file, and saving code as a .svg file.

2.7.3. OpenGL

OpenGL is a cross-platform API for rendering computer graphics, first released in 1992, and is used extensively across multiple disciplines within computer science. Due to the maturity and the broad adoption of the technology, many books exist on the tool. An Amazon.com search for 'OpenGL' returns over 500 results. We summarize a recent book by Lo and Lo [LL15] for this survey, which

Ref	Number of pages	Amazon Sales Rank	Google Scholar Citations	Price (USD)
[Zhu13]	338	1,029,905	27	44.99
[Nel14]	316	2,812,251	1	25.69
[Kin14]	288	888,699	5	27.64
[Mee15]	352	169,709	5	31.13
[RT16]	272	1,843,691	0	39.97

Table 5: Meta-data for five recent, related D3 books

has a focus on data visualization. We cite another five examples of OpenGL books in Table 6 with accompanying meta-data.



The OpenGL Data Visualization Cookbook by Lo and Lo [LL15] opens with an introduction to OpenGL and provides details on how to set up an OpenGL programming environment on Windows, OS X and Linux. The coding and compiling of a basic program is also detailed. The second chapter demonstrates drawing simple primitives of points, lines and triangles. An example of drawing an electrocardiogram trace is used and a 3D Gaussian function example with height matched to color is also demonstrated.

Rendering 3D images is the focus of the third chapter. The camera perspective is discussed and how to set it up, and an example of visualizing a 3D Gaussian is given. Details on how to interact with the rendering are also given, with code on how to move the perspective on the model. An example of code for volume rendering is given to finish the chapter. Visualizing images and videos are introduced in the fourth chapter. To do this, programming using shaders is introduced, along with textures and the OpenCV library [Bra00] for videos.

Chapter five builds on what has been introduced so far to render 3D depth fields captured using the Microsoft Kinect [Mic13]. The sixth chapter discusses rendering stereoscopic 3D models and using the OpenAsset Import Library (Assimp) [SGK*12] to import 3D models.

Using OpenGL for Embedded Systems (OpenGL ES) to render on Android devices is explained in the seventh chapter. Detailed instructions on how to setup the framework required, to create the code and to run the code on an Android phone are given. Chapter eight continues from the previous chapter with the addition of real-time rendering of information recorded using the phone sensors to create an interactive demo.

The final chapter details how to create an augmented reality based visual design for mobile and wearable platforms. This builds on the platform developed in the previous two chapters with the addition of the OpenCV library to enable live video capture, and the use of textures to render the video.

Other Tools Microsoft Excel is proprietary software therefore is not included in the survey. However its ubiquitousness merits some inclusion, therefore an Excel book review is included in the supplementary material.

We also note that there are a large number of books on R [Wic16], a useful visualization tool. It is not covered here due to space limitations. We are generally interested in interactive tools.

Ref	Number of pages	Amazon Sales Rank	Google Scholar Citations	Price (USD)
[Mov13]	326	894,921	9	49.99
[KSS16]	976	362,336	1070	47.29
[Wol11]	394	722,147	43	49.59
[WJHSL15]	880	330,601	482	45.40
[AS11]	768	210,517	3	55.99

Table 6: Meta-data of five recent, related OpenGL books

We present the tools mentioned in each book in Table 7. Tools where explicit instruction guides or code is given are marked in bold, enabling the quick discovery of guidance on a particular software technology.

3. Future Work and Discussion

By analyzing Table 2 and Figure 12 we can begin to see some patterns in the contents of the books with the classification we have applied. The books we have classified as ‘classics’ have limited pages dedicated to data analysis topics, and to perception and interaction topics. Books classified as ‘general audience’, ‘industry professional’ and ‘tools’ also have few pages dedicated to data analysis topics. ‘Industry professional’ classified books also do not, in general, discuss topics classified as visualization techniques.

The most popular topic from all of the books surveyed is the ‘visualization design’ category with most books having pages related to the topic. Other topics such as ‘information theory’ only have one book, which is wholly dedicated to the topic. Other popular topics are multivariate data, time-oriented data, and human perception and cognition.

Topics such as presentation, animation and sci-vis have a low number of pages dedicated to them, however these are covered in books from outside the data visualization field with presentation covered in infographics books, animation covered in graphics and sci-vis in its own topic.

It is worth noting that there is no scientific method of analyzing the topics discussed in each book and that the categorization used is a subjective matter.

Books continue to be published rapidly therefore a natural extension of this work would be to include newly published books as they become available [MP18, Ara18, Gab18, Ren18, Hea18, Ben19]. This survey predominantly covers information visualization books however many other data visualization books are available in different areas of the field such as scientific visualization, infographics, and GIS. In future we would like to extend our survey to cover these other sub-topics of data visualization. One area that is experiencing very rapid growth currently is the digital humanities. A survey of literature including books would be very valuable due to some overlap with text analytics and text visualization. An example of such a book is Introduction to Text Visualization by Cao and Cui [CC16].

We had difficulty in procuring some books that would otherwise be included in this survey, such as Tufte’s Beautiful Evidence

Title	Microsoft Excel	Tableau	Google Tools	D3	Processing	Adobe Illustrator	Other Tools
Visual Complexity [Lim11]	X		X				Apple Numbers
The Functional Art [Cai12]	X	X					
Data Visualization [Kir12]	X	X	X	X	X	X	Adobe After Effects, Adobe Flash, Arc GIS, CartoDB, closr.it, DataWrapper, Geocommons, Gephi, Google Chart Tools, Google Fusion Tables, Google Visualization API, Grapheur, InDesign, Indiemapper, Inkscape, Instant Atlas, Kartograph, KendoUI, KeyLines, Keynoe, Leaflet, Maya 3D, Microsoft Excel, Microsoft PowerPoint, Miso, Nodebox, Mondrian, OpenStreetMap, Panopticon, Paper.js, Polychart, Polymaps, QlikView, Quadrigram, R, Raphael, Tableau, TIBCO Spotfire, TileMil, WebGL, Wolfram Mathematica, Wordle, zoom.it
Design for Information [Mei13]							ColorBrewer, GLEAMviz.org, ManyEyes, Wordle
Information Visualization: An Introduction [Spe14]							Analyst's Notebook, IBM i2, MIND, Spotfire, WireVis
Information Visualization: Perception for Design [War12]							ColorBrewer
Interactive Data Visualization [WKG15]		X		X	12		ArcGIS, AVS, Circos, ERADS IMAGINE, GGobi, GraphViz, GRASS, InfoScope, Jigsaw, OpenDX, Prefuse, QGIS, SAS, SCIRun, Spotfire, SPSS, Tom Sawyer Software, VTK, Weave, XmdvTool
Visualize This [Yau11]	X	X	X		X	17	Python (24), R (77), Protovis (11), Adobe Flash Builder (23) ArcGIS, ColorBrewer, Inkspace, Many Eyes, Modest Maps, Mr. Data Converter, PHP, Polymaps, R
Interactive Visualization [FS12]	19	X	X		X	X	Adobe Flash, Adobe Flex, Apple iWork, ArcGIS, DataWrangler, DBpedia, FOAF, Gephi, Java, JavaScript, Juxta, ManyEyes, Microsoft Word, NodeXL, OpenOffice, OWL, PHP, Prefuse, Protovis, Python, Ruby, SPARQL, VIDI, VisualEyes, WinPure, Wordle
Visualization Analysis and Design [Mun14]		X					ColorBrewer, Constellation, Graph-Theoretic Scagnostics, InterRing, PivotGraph, the Hierarchical Clustering Explorer, VisDB
Data Visualisation [Kir16]	X	X	X	X		X	Microsoft Word
Show Me the Numbers [Few12]	7						
The Visual Organization [Sim14]	X	X	X	X	X		Amazon Redshift, BrioQuery, Business Objects, Cognos, Crystal Reports, DataHero, Gephi, Graph Viz, Hadoop, Hive, Impromptu, Jaspersoft, Lawson Business Intelligence, ManyEyes, MapReduce, Microsoft SQL Reporting Services, MicroStrategy, Pentaho, PIG, PowerPlay, QlikView, R, Reporter-Smith, SAP, SAS, SPSS, Teradata, Visually
Presenting Data [SH14]	X						
Storytelling with Data [Kna15]	X	X	X	X	X	X	Microsoft PowerPoint, Python, R
Good Charts [Ber16]	X	X	X	X		X	Apple Numbers, Charbuilder, Datawrapper, Highcharts, Inogram, Plot.ly, Qlik Sense, Quadrigram, Raw, Silk, Vizable
Making Data Visual [FM18]	X	X	X	X	X	X	Python(S), Vega(S), VegaLite(S) ggplot, Microsoft Virtual Earth, MizBee, PowerBI, R
Illuminating the Path [CT05]							AVS software, Comand Post of the Future (CPOF), GeoTime, ILOG library, IN-SPIRE, InfoVis-Toolkit, Microsoft PowerPoint, Oculus Info Sandbox, Office, OpenDx, POLESTAR, Rivet, SeeSoft, Spotfire, ThemeView, TreeJuxtaposer, Visual Insights ADVIZOR, Word, XmdvTool
Interactive Visualization of Large Networks [Ham08]							Dot, GEM, MG, Narcissus, Neato, NicheWorks, SemNet
Applied Security Visualization [Mar09]	X	X	X				Argus(4), awk(1), NetFlow(1), NFDump(1), Perl(2), sed(1) Advizor, AfterGlow, C#, C++, Crystal Reports, Cytoscape, Dotty and Ineato, EtherApe, ggobi, gITail, gnuplot, GraphViz, grep, GUESS, InetVis, Java, Large Graph Layout, Many Eyes, Mathematica, MatLab, Mondrian, MRTG/RRD, Notepad, NVisionIP, Oculus, OpenOffice, Parvis, Ploticus, R, Real Time 3D Graph Visualizer, Rumint, Shoki, SpotFire, StarLight, Swivel, TimeSearcher, TNV, Treemap, Tulip, Walrus
Visualization of Time-Oriented Data [AMST11]							Microsoft Word, Polaris, Visage, VisuExplore, Xmdv-Tool
Multidimensional Data Visualization [DKŽ12]							Matlab, Xmdv
Visualizing Time [Wil11]	X						Adobe Photoshop
Graph Analysis and Visualization [BJ15]	5			29		X	Cytoscape(10), Gephi(7), Influent(6), NodeXL(8), Python(18), yEd(4) Aperature Tiles, Geotime, Hadoop, Microsoft PowerPoint, Mondrian, SPARQL, Titan
Applying Color Theory to Digital Media and Visualization [Rhy16]		X					Adobe Color CC, Adobe Photoshop, Coblis, Color Companion, Color Gamut Mask, Color picker tool, Colorbrewer, COLOURlover's Color Palette Software, Paletton, PANTONE Studio app, Vischeck
User-Centered Evaluation of Visual Analytics [Sch17]		X					Gotham, Metropolis, Tableau, TRIST
Visualizing Data [Fry07]					326		MySQL(4), OpenOffice(2), Peal(4)
Interactive Data Visualization for the Web [Mur13]		X	X	183	X		HTML(43) Arbor.js, Crossfilter, Cubism, Dashku, DataWrapper, dc.js, Flot, ggplot2, gRaphael, Highcharts JS, JavaScript InfoVis Toolkit, jqPlot, jQuery Sparklines, Kartograph, Leaflet, Modest Maps, NVD3, Paper.js, Peity, PhiloGL, Polychart.js, Polymaps, Raphael, Rickshaw, Sigma.js, Three.js, Timeline.js, Tributary, YUI Charts
OpenGL Data Visualization Cookbook [LL15]							Android SDK(4), Java(11), OpenCV + C++(11), OpenGL + C++(168), OpenGL ES + C++(51)

Table 7: Table showing software tools mentioned in books. The numbers indicate the number of dedicated pages to explicit operating instructions or code and the colour intensity an indication of the percentage of the book dedicated to the tool. The 'X' mark and non bold text indicates a tool that is mentioned without explicit instruction. **Bold** text indicates tools where explicit operating instructions or code is given along with the number of pages where the tool is described. 'S' indicates code given in supplementary material.

[Tuf06]. In future we would like to add full reviews of these books into the survey as they are highly cited literature. While conducting the search for suitable books, it was noticeable that there are a number of books dedicated to specific data visualization tools, for

example we found 23 books on D3 [BOH11], a JavaScript data visualization library, over a dozen books on the Processing language, and nearly 4,000 books on Excel. This indicates that the majority of information visualization books are related to visualization tools,

and therefore each tool may be suitable for its own book survey. It also indicates that these areas are saturated and that as an author they are areas to best avoid.

References

- [AAM*17] ALHARBI N., ALHARBI M., MARTINEZ X., KRONE M., ROSE A. S., BAADEN M., LARAMEE R. S., CHAVENT M.: Molecular Visualization of Computational Biology Data: A Survey of Surveys. In *EuroVis 2017 - Short Papers* (2017), Kozlikova B., Schreck T., Wischgoll T., (Eds.), The Eurographics Association, pp. 133–137.
- [Ado18] ADOBE SYSTEMS INCORPORATED: Adobe color cc, 2018. URL Accessed: 4 January 2019. URL: <https://color.adobe.com/>.
- [AHC08] ADJEI C., HOLLAND-CUNZ N.: Monitoring and visualizing last.fm. *Diplom thesis, Fachhochschule Mainz* (2008). URL Accessed: 4 January 2019. URL: http://visualizinglastfm.de/projects/data_visuals/visualizing_lastfm_01/.
- [Ama18] AMAZON.COM, INC.: Amazon.com, 2018. URL Accessed: 4 January 2019. URL: <https://www.amazon.com>.
- [AMST11] AIGNER W., MIKSCH S., SCHUMANN H., TOMINSKI C.: *Visualization of Time-Oriented Data*. Springer Science & Business Media, Berlin, Germany, 2011.
- [Ara18] ARAGUES A.: *Visualizing Streaming Data: Interactive Analysis Beyond Static Limits*. O'Reilly Media, Sebastopol, CA, 2018.
- [AS11] ANGLE E., SHREINER D.: *Interactive Computer Graphics: A Topdown Approach with Shader-Based OpenGL*. Addison-Wesley, Boston, MA, 2011.
- [Asp14] ASPIN A.: *High Impact Data Visualization with Power View, Power Map, and Power BI*. Apress, Berkeley, CA, 2014.
- [Bak07] BAKER C. P.: Email map, 2007. URL Accessed: 4 January 2019. URL: <http://christopherbaker.net/projects/mymap/>.
- [BBS67] BERTIN J., BARBUT M., SERGE B.: *Sémiologie graphique: les diagrammes, les réseaux, les cartes*. Mouton/Gauthier-Villars, Paris, 1967.
- [BD53] BARBEU-DUBOURG J.: *Chronographie universelle, 1753*. URL Accessed: 4 January 2019. URL: <http://dataphys.org/list/barbeu-dubourgs-machine-chronologique/>.
- [BEC*07] BININDA-EMONDS O. R., CARDILLO M., JONES K. E., MACPHEE R. D., BECK R. M., GRENYER R., PRICE S. A., VOS R. A., GITTLEMAN J. L., PURVIS A.: The delayed rise of present-day mammals. *Nature* 446, 7135 (2007), 507.
- [Ben19] BENOIT G.: *Introduction to Information Visualization: Transforming Data into Meaningful Information*. Rowman & Littlefield Publishers, Lanham, 2019.
- [Ber83] BERTIN J.: *Semiology of graphics: Diagrams, networks, maps (WJ Berg, Trans.)*. Madison, WI, 1983.
- [Ber16] BERINATO S.: *Good Charts*. Harvard Business Review Press, Boston, MA, 2016.
- [BGLL12] BOHNACKER H., GROSS B., LAUB J., LAZZERONI C.: *Generative design: visualize, program, and create with processing*. Princeton Architectural Press, New York, NY, 2012.
- [BJ15] BRATH R., JONKER D.: *Graph analysis and visualization: discovering business opportunity in linked data*. John Wiley & Sons, Indianapolis, IN, 2015.
- [BKW16] BECK F., KOCH S., WEISKOPF D.: Visual analysis and dissemination of scientific literature collections with survis. *IEEE transactions on visualization and computer graphics* 22, 1 (2016), 180–189.
- [BOH11] BOSTOCK M., OGIEVETSKY V., HEER J.: D³ data-driven documents. *IEEE Transactions on Visualization and Computer Graphics* 17, 12 (2011), 2301–2309.
- [Bra00] BRADSKI G.: The opencv library. *Dr. Dobbs' Journal: Software Tools for the Professional Programmer* 25, 11 (2000), 120–123.
- [Bre] BREMER N.: Reviews of data visualization books - visual cinnamon. URL Accessed: 4 January 2019. URL: <https://www.visualcinnamon.com/resources/learning-data-visualization/books>.
- [Cai12] CAIRO A.: *The Functional Art: An introduction to information graphics and visualization*. New Riders, Berkeley, 2012.
- [Cai16] CAIRO A.: *The truthful art: Data, charts, and maps for communication*. New Riders, Berkeley, 2016.
- [CC16] CAO N., CUI W.: *Introduction to text visualization*. Springer, London, 2016.
- [CFV*16] CHEN M., FEIXAS M., VIOLA I., BARDERA A., SHEN H.-W., SBERT M.: *Information Theory Tools for Visualization*. CRC Press, Boca Raton, FL, 2016.
- [Cha38] CHAMBERS E.: *Cyclopaedia: Or an Universal Dictionary of Arts and Sciences*. Midwinter, London, 1738.
- [Cle85] CLEVELAND W. S.: *The Elements of Graphing Data*. Wadsworth Advanced Books and Software, Monterey, CA, 1985.
- [Cle93] CLEVELAND W. S.: *Visualizing data*. Hobart Press, Summit New Jersey, 1993.
- [CLS14] CIUCCARELLI P., LUPI G., SIMEONE L.: *Visualizing the Data City: Social Media as a Source of Knowledge for Urban Planning and Management*. Springer, Cham, 2014.
- [CMS99] CARD S. K., MACKINLAY J. D., SHNEIDERMAN B.: *Readings in information visualization: using vision to think*. Morgan Kaufmann, San Francisco, CA, 1999.
- [Cot13] COTGREAVE A.: 8 great books about data visualization, 2013. URL Accessed: 4 January 2019. URL: <https://www.tableau.com/about/blog/2013/7/list-books-about-data-visualisation-24182>.
- [CT05] COOK K. A., THOMAS J. J.: *Illuminating the path: The research and development agenda for visual analytics*. Tech. rep., Pacific Northwest National Laboratory (PNNL), Richland, WA, 2005.
- [Dat] DATA CYMRU: Infobasecymru. URL Accessed: 4 January 2019. URL: www.infobasecymru.net.
- [Des37] DESCARTES R.: *La géométrie*. Jan Maire, Leiden, 1637.
- [DKŽ12] DZEMYDA G., KURASOVA O., ŽILINSKAS J.: *Multidimensional Data Visualization: Methods and Applications*, vol. 75. Springer Science & Business Media, New York/London, 2012.
- [Dur04] DURER A.: The fall of man, 1504.
- [EGK*01] ELLSON J., GANSNER E., KOUTSOFIOS L., NORTH S. C., WOODHULL G.: Graphviz—open source graph drawing tools. In *International Symposium on Graph Drawing* (2001), Springer, pp. 483–484.
- [End16] ENDERT A.: Semantic interaction for visual analytics: Inferring analytical reasoning for model steering. *Synthesis Lectures on Visualization* 4, 2 (2016), 1–99.
- [esr] ESRI: ArcGIS. URL Accessed: 4 January 2019. URL: <https://www.arcgis.com/>.
- [Eul41] EULER L.: Solutio problematis ad geometriam situs pertinentis. *Commentarii academiae scientiarum Petropolitanae* 8 (1741), 128–140.
- [Fei14] FEINBERG J.: Wordle, 2014. URL Accessed: 4 January 2019. URL: <http://www.wordle.net/>.
- [Few06] FEW S.: *Information Dashboard Design*. O'Reilly, Sebastopol, CA, 2006.
- [Few12] FEW S.: *Show Me the Numbers: Designing Tables and Graphs to Enlighten*. Analytics Press, Burlingame, CA, 2012.
- [FGKR16] FALK M., GROTTTEL S., KRONE M., REINA G.: Interactive gpu-based visualization of large dynamic particle data. *Synthesis Lectures on Visualization* 4, 3 (2016), 1–121.

- [FM18] FISCHER D., MEYER M.: *Making Data Visual*. O'Reilly, Sebastopol, CA, 2018.
- [Fry07] FRY B.: *Visualizing data: Exploring and explaining data with the processing environment*. O'Reilly Media, Inc., Sebastopol, CA, 2007.
- [FS12] FERSTER B., SHNEIDERMAN B.: *Interactive visualization: Insight through inquiry*. MIT Press, Cambridge, MA, 2012.
- [Gab18] GABRIELLE B. R.: *Storytelling with Graphs: A New Approach Beyond Data Visualization*. Insights Publishing, Kirkland, WA, 2018.
- [GG89] GATH I., GEVA A. B.: Unsupervised optimal fuzzy clustering. *IEEE Transactions on pattern analysis and machine intelligence* 11, 7 (1989), 773–780.
- [Goo18] GOOGLE LLC: Google scholar, 2018. URL Accessed: 4 January 2019. URL: <https://scholar.google.co.uk/>.
- [Ham08] HAM F. v.: *Interactive Visualization of Large Networks: Technical Challenges and Practical applications*. VDM Verlag, Saarbrücken, 2008.
- [HB03] HARROWER M., BREWER C. A.: Colorbrewer.org: an online tool for selecting colour schemes for maps. *The Cartographic Journal* 40, 1 (2003), 27–37.
- [Hea18] HEALY K.: *Data Visualization: A Practical Introduction*. Princeton University Press, 2018.
- [Hur15] HURTER C.: Image-based visualization: Interactive multidimensional data exploration. *Synthesis Lectures on Visualization* 3, 2 (2015), 1–127.
- [HYFC14] HARRISON L., YANG F., FRANCONERI S., CHANG R.: Ranking visualizations of correlation using weber's law. *IEEE Transactions on Visualization and Computer Graphics* 20, 12 (2014), 1943–1952.
- [IHK*17] ISENBERG P., HEIMERL F., KOCH S., ISENBERG T., XU P., STOLPER C. D., SEDLMIR M. M., CHEN J., MÖLLER T., STASKO J.: vispubdata.org: A Metadata Collection about IEEE Visualization (VIS) Publications. *IEEE Transactions on Visualization and Computer Graphics* 23 (2017).
- [IIC*13] ISENBERG T., ISENBERG P., CHEN J., SEDLMIR M., MÖLLER T.: A systematic review on the practice of evaluating visualization. *IEEE Transactions on Visualization and Computer Graphics* 19, 12 (2013), 2818–2827.
- [Imh82] IMHOF E.: *Cartographic relief presentation*. Walter de Gruyter GmbH & Co KG, Berlin/Boston, 1982.
- [Ins09] INSELBERG A.: *Parallel Coordinates: Visual Multidimensional Geometry and its Applications*. Springer, New York, NY, 2009.
- [JS91] JOHNSON B., SHNEIDERMAN B.: Tree-maps: A space-filling approach to the visualization of hierarchical information structures. In *Proceedings of the 2nd conference on Visualization '91* (1991), IEEE Computer Society Press, pp. 284–291.
- [KD12] KRESSE W., DANKO D. M.: *Springer Handbook of Geographic Information*. Springer, Dordrecht, 2012.
- [Kin14] KING R. S.: *Visual Storytelling with D3: An Introduction to Data Visualization in JavaScript*. Addison-Wesley Professional, Upper Saddle River, NJ, 2014.
- [Kir12] KIRK A.: *Data Visualization: a successful design process*. Packt Publishing Ltd, Los Angeles, CA, 2012.
- [Kir16] KIRK A.: *Data visualisation: a handbook for data driven design*. Sage, Los Angeles, CA, 2016.
- [KK94] KEIM D. A., KRIEGLER H.-P.: Visdb: Database exploration using multidimensional visualization. *IEEE Computer Graphics and Applications* 14, 5 (1994), 40–49.
- [KKEM10] KEIM D., KOHLHAMMER J., ELLIS G., MANSMANN F.: Mastering the information age-solving problems with visual analytics. ISBN-13 9783905673777 (2010).
- [Kla08] KLANTEN R.: *Data Flow: Visualising Information in Graphic Design*. Die Gestalten Verlag, Berlin, 2008.
- [Kna15] KNAFLIC C. N.: *Storytelling with data: a data visualization guide for business professionals*. John Wiley & Sons, Hoboken, NJ, 2015.
- [Kru13] KRUM R.: *Cool infographics: Effective communication with data visualization and design*. John Wiley & Sons, Indianapolis, IN, 2013.
- [KSS16] KESSENICH J., SELLERS G. M., SHREINER D.: *OpenGL Programming Guide: The Official Guide to Learning OpenGL*. Addison-Wesley Professional, Boston, 2016.
- [Lim11] LIMA M.: *Visual Complexity. Mapping Patterns of Information*. Princeton: Princeton Architectural Press, Princeton, NJ, 2011.
- [Lim14] LIMA M.: *The Book of Trees: Visualizing Branches of Knowledge*. Princeton Architectural Press, Princeton, NJ, 2014.
- [Lim17] LIMA M.: *The Book of Circles: Visualizing Spheres of Knowledge*. Princeton Architectural Press, Princeton, NJ, 2017.
- [LL15] LO R. C., LO W. C.: *OpenGL Data Visualization Cookbook*. Packt Publishing Ltd, Birmingham, UK, 2015.
- [LM10] LAM H., MUNZNER T.: A guide to visual multi-level interface design from synthesis of empirical study evidence. *Synthesis Lectures on Visualization* 1, 1 (2010), 1–117.
- [Mac11] MACIEJEWSKI R.: Data representations, transformations, and statistics for visual reasoning. *Synthesis Lectures on Visualization* 2, 1 (2011), 1–85.
- [Mar09] MARTY R.: *Applied security visualization*. Addison-Wesley, Upper Saddle River, NJ, 2009.
- [McC09] MCCANDLESS D.: *Information is Beautiful*. Collins, London, 2009.
- [McC14] MCCANDLESS D.: *Knowledge is Beautiful*. HarperCollins, London, 2014.
- [Mee15] MEEKS E.: *D3.js in Action*. Manning Publ., Shelter Island, 2015.
- [Mei13] MEIRELLES I.: *Design for Information: An Introduction to the Histories, Theories, and Best Practices behind Effective Information Visualizations*. Rockport Publishers, Beverly, MA, 2013.
- [MGR99] MUNZNER T., GUIMBRETIERE F., ROBERTSON G.: Constellation: A visualization tool for linguistic queries from mindnet. In *Information Visualization, 1999.(Info Vis' 99) Proceedings. 1999 IEEE Symposium on* (1999), IEEE, pp. 132–135.
- [Mic13] MICROSOFT CORPORATION: Kinect - Windows app development, 2013. URL Accessed: 4 January 2019. URL: <https://developer.microsoft.com/en-us/windows/kinect>.
- [Min69] MINARD C. J.: *Carte Figurative des pertes successives en hommes de l'armée qu'Annibal conduisit d'Espagne en Italie en traversant les Gaules (selon Polybe)*. Regnier et Dourdet, Paris, 1869.
- [ML17] MCNABB L., LARAMEE R. S.: Survey of Surveys (SoS) - Mapping The Landscape of Survey Papers in Information Visualization. *Computer Graphics Forum* 36, 3 (2017), 589–617.
- [Mov13] MOVANIA M. M.: *OpenGL Development Cookbook*. Packt Publishing Ltd, Birmingham, UK, 2013.
- [MP18] MILLER-PRICE J.: *GREAT STORYTELLING IN BUSINESS: How to present your data as a compelling story*. Independently published, 2018.
- [Mun14] MUNZNER T.: *Visualization Analysis and Design*. CRC Press, Boca Raton, FL, 2014.
- [Mur13] MURRAY S.: *Interactive Data Visualization for the Web*. O'Reilly Media, Inc., Sebastopol, CA, 2013.
- [MV10] MCNAUGHTON S., VELASCO S.: Fifty years of space exploration. *National Geographic* (2010).
- [Nel14] NELLI F.: *Create Web Charts with D3*. Apress, Berkeley, CA, 2014.

- [NN17] NYHOFF J., NYHOFF L.: *Processing: An Introduction to Programming*. Chapman and Hall, Boca Raton, FL, 2017.
- [NRM55] NORTHWAY M. L., ROOKS M. M., MORENO J.: Creativity and sociometric status in children. *Sociometry* 18, 4 (1955), 194–201.
- [Pal09] PALEY W.: Textarc: Alice's adventures in wonderland, 2009. URL Accessed: 4 January 2019. URL: <http://www.textarc.org/Alice.html>.
- [Pan18] PANTONE INC.: Pantone studio, 2018. URL Accessed: 4 January 2019. URL: <https://www.pantone.com/studio>.
- [Pla86] PLAYFAIR W.: *The commercial and political atlas and statistical breviary*. J. Debrett, Piccadilly, 1786.
- [QGI] QGIS DEVELOPMENT TEAM: QGIS. URL Accessed: 4 January 2019. URL: <https://www.qgis.org/>.
- [RB10] RENSINK R. A., BALDRIDGE G.: The perception of correlation in scatterplots. In *Computer Graphics Forum* (2010), vol. 29, Wiley Online Library, pp. 1203–1210.
- [Ren18] RENDGEN S.: *The Minard System: The Complete Statistical Graphics of Charles-Joseph Minard*. Princeton Architectural Press, New York, NY, 2018.
- [RF07] REAS C., FRY B.: *Processing: a programming handbook for visual designers and artists*. Mit Press, Cambridge, MA, 2007.
- [RF15] REAS C., FRY B.: *Getting Started with Processing: A Hands-On Introduction to Making Interactive Graphics*. Maker Media, Inc., San Francisco, CA, 2015.
- [RHR17] ROBERTS J. C., HEADLEAND C. J., RITSOS P. D.: *Five Design-Sheets: Creative Design and Sketching for Computing and Visualisation*. Springer, Cham, 2017.
- [Rhy16] RHYNE T.-M.: *Applying color theory to digital media and visualization*. CRC Press, Boca Raton, FL, 2016.
- [RMA*10] RIND A., MIKSCH S., AIGNER W., TURIC T., POHL M.: Visuexplore: gaining new medical insights from visual exploration. In *Proc. Int. Workshop on Interactive Systems in Healthcare (WISH@CHI2010)* (2010), pp. 149–152.
- [Rob12] ROBBINS N. B.: *Creating More Effective Graphs*. Wiley, Hoboken, NJ, 2012.
- [RT16] RININSLAND Æ., TELLER S.: *Learning D3.js Data Visualization*. Packt Publishing Ltd, Birmingham, UK, 2016.
- [Sam06] SAMET H.: *Foundations of multidimensional and metric data structures*. Morgan Kaufmann, San Francisco, CA, 2006.
- [Sch17] SCHOLTZ J.: User-centered evaluation of visual analytics. *Synthesis Lectures on Visualization* 5, 1 (2017), i–71.
- [SGK*12] SCHULZE T., GESSLER A., KULLING K., NADLINGER D., KLEIN J., SIBLY M., GUBISCH M.: Open asset import library (assimp). *Computer Software*, URL: <https://github.com/assimp/assimp> (2012).
- [SH14] SWIRES-HENNESSY E.: *Presenting Data: How to Communicate Your Message Effectively*. John Wiley & Sons, Chichester, 2014.
- [Shi09] SHIFFMAN D.: *Learning Processing: A Beginner's Guide to Programming Images, Animation, and Interaction*. Morgan Kaufmann, Amsterdam, 2009.
- [Shn96] SHNEIDERMAN B.: The eyes have it: A task by data type taxonomy for information visualizations. In *Visual Languages, 1996. Proceedings., IEEE Symposium on* (1996), IEEE, pp. 336–343.
- [SI10] STEELE J., ILIINSKY N.: *Beautiful Visualization: Looking at Data through the Eyes of Experts*. O'Reilly Media, Inc., Beijing, 2010.
- [Sim14] SIMON P.: *The visual organization: data visualization, Big Data, and the quest for better decisions*. John Wiley & Sons, Hoboken, NJ, 2014.
- [SK07] STEINWEBER P., KOLLER A.: Similar diversity, 2007.
- [SLBC03] SWAYNE D. F., LANG D. T., BUJA A., COOK D.: Ggobi: evolving from xgobi into an extensible framework for interactive data visualization. *Computational Statistics & Data Analysis* 43, 4 (2003), 423–444.
- [SMM12] SEDLMAIR M., MEYER M., MUNZNER T.: Design study methodology: Reflections from the trenches and the stacks. *IEEE transactions on visualization and computer graphics* 18, 12 (2012), 2431–2440.
- [Sno55] SNOW J.: *On the mode of communication of cholera*. John Churchill, London, 1855.
- [SP16] SEDIG K., PARSONS P.: Design of visualizations for human-information interaction: A pattern-based framework. *Synthesis Lectures on Visualization* 4, 1 (2016), 1–185.
- [Spe14] SPENCE R.: *Information visualization: An introduction*. Springer, Cham, 2014.
- [Spr18] SPRINGER, SCIENCE+BUSINESS MEDIA: Springer - international publisher science, technology, medicine, 2018. URL Accessed: 4 January 2019. URL: <http://www.springer.com/>.
- [SS02] SEO J., SHNEIDERMAN B.: Interactively exploring hierarchical clustering results [gene identification]. *Computer* 35, 7 (2002), 80–86.
- [Sto11] STONE J.: X-men family tree, 2011. URL Accessed: 4 January 2019. URL: <https://www.joe-stone.co.uk/familytrees>.
- [Tab] TABLEAU SOFTWARE: Tableau. URL Accessed: 4 January 2019. URL: <https://www.tableau.com/>.
- [Tel14] TELEA A. C.: *Data visualization: principles and practice*. CRC Press, Boca Raton, FL, 2014.
- [Tom15] TOMINSKI C.: Interaction for visualization. *Synthesis Lectures on Visualization* 3, 1 (2015), 1–107.
- [Tuf83] TUFTE E. R.: *The Visual Display of Quantitative Information*. Graphics Press, Cheshire, CT, 1983.
- [Tuf90] TUFTE E. R.: *Envisioning Information*. Graphics Press, Cheshire, CT, 1990.
- [Tuf97] TUFTE E. R.: *Visual Explanations*. Graphics Press, Cheshire, CT, 1997.
- [Tuf06] TUFTE E. R.: *Beautiful Evidence*. Graphics Press, Cheshire, CT, 2006.
- [Uni] UNIVERSITY OF VIRGINIA: Visualeyes. URL Accessed: 4 January 2019. URL: <http://www.viseyes.org/viseyes.htm>.
- [VFA13] VATHY-FOGARASSY Á., ABONYI J.: *Graph-based clustering and data visualization algorithms*. Springer, London, 2013.
- [VHVW03] VAN HAM F., VAN WIJK J. J.: Beamtrees: Compact visualization of large hierarchies. *Information Visualization* 2, 1 (2003), 31–39.
- [VSBOR02] VAN SOLINGEN R., BASILI V., CALDIERA G., ROMBACH H. D.: Goal question metric (GQM) approach. *Encyclopedia of software engineering* (2002).
- [VWVH*07] VIEGAS F. B., WATTENBERG M., VAN HAM F., KRIS J., MCKEON M.: Manyeyes: a site for visualization at internet scale. *IEEE Transactions on Visualization and Computer Graphics* 13, 6 (2007).
- [WAG05] WILKINSON L., ANAND A., GROSSMAN R.: Graph-theoretic scagnostics. In *Proceedings of the Proceedings of the 2005 IEEE Symposium on Information Visualization* (Washington, DC, USA, 2005), INFOVIS '05, IEEE Computer Society, pp. 21–.
- [Wal74] WALKER F. A.: Chart showing the principal constituent elements of the population of each state. In *Statistical atlas of the United States based on the results of the ninth census 1870 with contributions from many eminent men of science and several departments of the government* (New York, 1874), J. Bien.
- [War10] WARE C.: *Visual Thinking: For design*. Morgan Kaufmann, Amsterdam, 2010.
- [War12] WARE C.: *Information visualization: perception for design*. Elsevier, Amsterdam, 2012.
- [Wat98] WATTENBERG M.: Map of the market, 1998. URL Accessed: 4 January 2019. URL: <http://www.bewitched.com/marketmap.html>.

- [Wat06] WATTENBERG M.: Visual exploration of multivariate graphs. In *Proceedings of the SIGCHI conference on Human Factors in computing systems* (2006), ACM, pp. 811–819.
- [WGK15] WARD M. O., GRINSTEIN G., KEIM D.: *Interactive data visualization: foundations, techniques, and applications (2nd Edition)*. CRC Press, Boca Raton, FL, 2015.
- [Wic16] WICKHAM H.: *ggplot2: Elegant Graphics for Data Analysis*. Springer, New York, NY, 2016.
- [Wil11] WILLS G.: *Visualizing time: Designing graphical representations for statistical data*. Springer Science & Business Media, New York, 2011.
- [WJHSL15] WRIGHT JR R. S., HAEMEL N., SELLERS G. M., LIPCHAK B.: *OpenGL SuperBible: Comprehensive Tutorial and Reference*. Addison-Wesley Professional, Boston, MA, 2015.
- [Wol11] WOLFF D.: *OpenGL 4.0 Shading Language Cookbook*. Packt Publishing Ltd, Birmingham, 2011.
- [Won10] WONG D. M.: *The Wall Street Journal Guide to Information Graphics: The Dos and Don'ts of Presenting Data, Facts, and Figures*. WW Norton, New York, NY, 2010.
- [Yau11] YAU N.: *Visualize this: the FlowingData guide to design, visualization, and statistics*. Wiley, Indianapolis, IN, 2011.
- [Yau13] YAU N.: *Data points: Visualization That Means Something*. John Wiley & Sons, Indianapolis, IN, 2013.
- [YWR02] YANG J., WARD M. O., RUNDENSTEINER E. A.: Interring: An interactive tool for visually navigating and manipulating hierarchical structures. In *Information Visualization, 2002. INFOVIS 2002. IEEE Symposium on* (2002), IEEE, pp. 77–84.
- [Zhu13] ZHU N. Q.: *Data visualization with D3.js cookbook*. Packt Publishing Ltd, Birmingham, 2013.