Practical D3.js

Tarek Amr
Rayna Stamboliyska

Apress®

Practical D3.js

Tarek Amr
Amsterdam
The Netherlands

Rayna Stamboliyska
Issy-les-Moulineaux
Paris, France

ISBN-13 (pbk): 978-1-4842-1927-0
ISBN-13 (electronic): 978-1-4842-1928-7
DOI 10.1007/978-1-4842-1928-7

Library of Congress Control Number: 2016944319

Copyright © 2016 by Tarek Amr and Rayna Stamboliyska

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.

Trademarked names, logos, and images may appear in this book. Rather than use a trademark symbol with every occurrence of a trademarked name, logo, or image we use the names, logos, and images only in an editorial fashion and to the benefit of the trademark owner, with no intention of infringement of the trademark.

The use in this publication of trade names, trademarks, service marks, and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Managing Director: Welmoed Spahr
Acquisitions Editor: Louise Corrigan
Development Editor: Michael Koch
Technical Reviewer: Kent Russell
Editorial Board: Steve Anglin, Pramila Balen, Louise Corrigan, Jonathan Gennick, Robert Hutchinson, Celestin Suresh John, Nikhil Karkal, James Markham, Susan McDermott, Matthew Moodie, Ben Renow-Clarke, Gwenan Spearing
Coordinating Editor: Nancy Chen
Copy Editor: Laura Lawrie
Compositor: SPi Global
Indexer: SPi Global

Distributed to the book trade worldwide by Springer Science+Business Media New York, 233 Spring Street, 6th Floor, New York, NY 10013. Phone 1-800-SPRINGER, fax (201) 348-4505, e-mail orders-ny@springer-sbm.com, or visit www.springer.com. Apress Media, LLC is a California LLC and the sole member (owner) is Springer Science + Business Media Finance Inc (SSBM Finance Inc). SSBM Finance Inc is a Delaware corporation.

For information on translations, please e-mail rights@apress.com, or visit www.apress.com.

Apress and friends of ED books may be purchased in bulk for academic, corporate, or promotional use. eBook versions and licenses are also available for most titles. For more information, reference our Special Bulk Sales–eBook Licensing web page at www.apress.com/bulk-sales.

Any source code or other supplementary materials referenced by the author in this text is available to readers at www.apress.com. For detailed information about how to locate your book's source code, go to www.apress.com/source-code/.

Printed on acid-free paper

Contents at a Glance

About the Authors xi

About the Technical Reviewer xiii

Acknowledgments xv

Introduction xvii

■ Part 1: Understanding Data Visualization 1

■ Chapter 1: Understanding Data Visualization 3

■ Chapter 2: Structuring and Designing Data Visualizations 27

■ Chapter 3: Getting the Facts Right 39

■ Chapter 4: Sourcing Data 61

■ Part 2: Using D3.js for Practical Data Visualization 73

■ Chapter 5: Getting Started with D3 75

■ Chapter 6: Creating Complex Shapes 91

■ Chapter 7: Transforming Data with Layouts 109

■ Chapter 8: Using Advanced Layouts 127

■ Chapter 9: Working with Data 149

Index 163

Contents

About the Authors xi

About the Technical Reviewer xiii

Acknowledgments xv

Introduction xvii

Part 1: Understanding Data Visualization 1

Chapter 1: Understanding Data Visualization 3

Visualising Raw Data 4

Having a Good Eye for Data 6

Choosing the Right Form 9

Designing Interactive Data Visualizations 12

Structuring Interactive Data Visualizations 16

Techniques for Exploration and Interaction 18

Types of Interactive Experience 22

Why Use D3.js? 25

Summary 26

Chapter 2: Structuring and Designing Data Visualizations 27

First Things First 27

Organising Your Data Visualization 28

It's All About Gestalt 30

Summary 37

Chapter 3: Getting the Facts Right 39

What Is Your Story? 39

Differences between Data Points 45

Transformation, Not Representation 46

The Importance of Telling a Story 47

It Is D3, Not 3D 52

Things That Can Go Wrong 53

D3 in a Nutshell 53

Maps 54

Hierarchies 55

Spider and Chord Charts 57

Summary 59

Chapter 4: Sourcing Data 61

Finding and Acquiring Data 61

Open Data Portals 61

Reports by Institutions and Private Business Actors 62

User-Generated Data 64

(Big) Data Professionals 65

Data Scraping 65

Understanding D3 Selections 66

Reviewing Data Formats 67

CSV, TSV, and Other Data Files 68

JSON 69

Cocooning the Data 70

Problems with Data that You Can Solve 70

Problems with Data that the Data Provider Can Solve 71

More Bad Data (Science) 72

Summary 72

Part 2: Using D3.js for Practical Data Visualization 73

Chapter 5: Getting Started with D3 75

A Note on SVG 75

Time to Plot Something 75

Your First Bar Chart 76

When to Use D3 78

Basic Shapes with D3 79

Your First D3 Chart 81

D3 Scales with Your Needs 82

Web Development Tools 85

Your First Animated D3 Chart 86

Moving and Rotating Objects 88

Summary 90

Chapter 6: Creating Complex Shapes 91

Forget Rectangles, Go Freehand 91

A Simple Freehand Chart in D3 92

Introducing D3 Path Generators 93

Using Area Charts 95

Adding Text 97

JavaScript's this keyword 99

Adding an Axis 100

Connecting the Dots with SVG Diagonal 102

Using Arcs 106

Summary 108

Chapter 7: Transforming Data with Layouts 109
Pie Chart Layout 109
Treemap Layout 114
Pack Layout 117
Partition Layout 118
Stacking Stuff in Layers 121
Stacking Stuff in Circles 124
Summary 126
Chapter 8: Using Advanced Layouts 127
Using the Force Layout (to Move Objects) 127
Which Add-ons Do News Outlets Use? 131
A Closer Look at Gravity 134
Time to Defy Gravity 136
Creating Your Own Layouts 140
Method Chaining 140
Nested Functions and Closure 142
Array Map and Reduce Methods 142
"Meatballs" Layout 143
Mixing Layouts 146
Summary 148
Chapter 9: Working with Data 149
Using Third-Party Data 149
Charting the Global Gender Gap 149
Charting the Top Music Tracks 155

Combining Data from Multiple Sources 159
Open Data Around the World 159
Summary 162

Index 163

About the Authors

Tarek Amr achieved his postgraduate degree in data mining and machine learning from the University of East Anglia. He currently works as a data scientist in Amsterdam. He has more than 10 years' experience in software development. Tarek participates in training data journalists and he works on promoting open data. He can be reached at http://tarekamr.com.

Rayna Stamboliyska is a trained scientist whose professional journey has shifted into data-driven innovation and strategy. She consults for international organizations, businesses, media and nonprofits, and is a polylingual bookworm. She conducts risk and crisis management assessments using OSINT on a daily basis and frequently works in conflict and postconflict zones in MENA, Eastern Europe, and Africa. She is the founder of the first-ever action research and service consultancy around open knowledge in the MENA region, and the curator of Data Colada, the only French-speaking weekly resource on data. Crunching data is an integral part of her existence, and so is data visualization. She frequently trains curious individuals about open data and data journalism. She can be reached at http://about.me/raynast or @MaliciaRogue on Twitter.

About the Technical Reviewer

Kent Russell manages money on a professional basis and believes that D3 and R are essential tools for visual exploration of data and decision making. He resides in Birmingham, Alabama.

Acknowledgments

I'd like to dedicate this book to my late mother, Sherifa. I am also grateful to many people who taught me a lot. Some of these people I know in person, some I haven't met yet, but without them all this book wouldn't have been what it is now: my postgraduate supervisor, Dr. Beatriz De La Iglesia; Alberto Cairo, author of the book The Functional Art; Stoyan Stefanov, author of the book JavaScript Patterns; and last but not least, my friend and data journalist, Ramzi Jaber.

—Tarek Amr

Writing those chapters consumed a fair amount of dedication and work. Still, achievement would not have been possible if I did not have a support of many individuals: friends, family and like-minded nerds. Therefore, I would like to extend my sincere gratitude to all of them: my parents for their pride and trust in my abilities, from day 0; Benoît, for answering silly questions while troubleshooting Ansible scripts and bearing with me throughout the process; Pierre, for embarking on the joyful folie we call Data Colada, started out between two paragraphs of this book; Gabriel, for being the most sensitive and funkiest mentor in the world; Sylvain and the #DataGueule team for inspiration; and everyone who provided ideas and expressed keen interest (and kicked me to go to the gym, too).

Last but not least, special thanks go to the Apress team who provided us with this opportunity. Thanks, Louise and Nancy, for your support, and thanks Kent for the insightful and cool editorial suggestions.

—Rayna Stamboliyska

Introduction

Data visualization plays an essential role in communicating ideas and concepts on today's data-driven world. Not only your personal computer, but maybe your mobile phone, smart watch, and other sensors in your home are all producing enormous amount of data. Last month when they came to install Tarek's washing machine, they told him it needs access to the Internet, so he can track his washing habits. If someone told him such a thing few years ago, he'd have laughed at them. Away from them being buzz words, terms such as the Internet of Things (IoT) and Big Data reflect how our world is very much into producing, storing, and analyzing data. That's why data visualization is not limited to few data nerds any more; journalists, marketing people, researchers, teachers, and many others also need data visualization to be able to understand the data at their fingertips so that they can base their decisions on it, or be able to communicate the insights they find there to others.

D3 is one of the most powerful and flexible data visualization frameworks out there. If you have been paying attention to the data team in the *New York Times* or the *Guardian* you sure would have noticed how it enables them to create visualizations we never thought possible before. Nevertheless, such flexibility comes at a cost; its learning curve is some how steep. Furthermore, you aren't normally given some data and asked to visualize it. Most of the time, you will need to know where to find your data, how to clean it, explore it, and pick the best visualization to represent it.

In Practical D3 we will try to walk the whole journey together, and show you how to pose your story hypothesis, and then find data to prove or disprove this hypothesis. We will also show you how to prepare your data as well as how to use D3 to be able to present your data story in a visual manner.

Book Organization

In the first four chapters, Rayna will be your guide in your journey with data visualization. Then, Tarek will be your guide in the next five chapters. As you will see in a moment, although the chapters build on each other, they are also self-contained. In other words, you are free to read this book from cover to cover or you treat this book as a hop-on hop-off bus:

Chapter 1 answers the question "what is data visualization?" It also explains how good visualizations are meant to help you explore the data. Rather than dumbing down facts, we will discuss the design heuristics, the different chart types, and how to choose the best visual structure to represent your data.

Chapter 2 explains some of the basics of HTML, such as the DOM. The chapter also introduces you to vector graphics and their underlying mathematical concepts.

Chapter 3 should be data journalists' favorite. It starts with explaining what is a journalistic story, how to find it, and how to pose your story hypothesis. You will also learn how to explore and analyze the data you have. Some basic mathematics is always nice for data journalists to learn.

Chapter 4 will show you where to look for your data. We look at what kind of data formats are out there. Along the way, you will learn some concepts, such as data scraping, data cleaning, and user-generated data.

In Chapter 5 you learn to create some basic shapes and to plot your first chart using SVG, so you understand the underlying technology D3 uses. The chapter also details how to recreate your chart using D3 and introduces some D3 concepts, such as scales, event listeners, and animations in addition to shapes scaling and transformations.

Chapter 6 shows you how to create more complex shapes using paths and path generators. It also shows you how to create line and area charts and add axis to your charts.

Chapters 7 and 8 will focus on D3 layouts. Layouts are how D3 transforms data and help you create complex data visualization. Chapter 7 starts with a layout for creating something as basic as a pie chart, and then shows you how to play with your pie and turn it into a sunburst diagram. You also learn how to create visualizations using layouts such as treemaps, packs, partition, and stack. Chapter 8 covers the force layout and shows you how to create your own layouts.

Chapter 9 builds on Chapter 4 and explains how D3 can help you load external data, regardless of whether it is in CSV or JSON format. You will also learn how to combine data from different sources, and how to use D3 and JavaScript to sort, filter, and manipulate your data before plotting it.

At the time of writing, the stable version of D3 is 3.5.x. Thus, the examples and code here are all based on that version.

Preboarding Notes

This book assumes that you have basic knowledge of JavaScript, CSS, and related web technologies. In cases in which the required knowledge is beyond the basic knowledge, we are going to explain things in more detail and get you prepared. We try to keep the book useful to users who have minimal JavaScript experience as well as to those who are JavaScript experts but are not that experienced in D3.