

AUTHORS

ZACH GEMIGNANI • CHRIS GEMIGNANI



FOUNDERS OF JUICE ANALYTICS

DATA FLUENCY

EMPOWERING YOUR ORGANIZATION
WITH EFFECTIVE DATA COMMUNICATION

Ez. 02



An abstract graphic in the background of the book cover features several thick, curved lines in shades of red, blue, and grey. These lines intersect and flow across the page, creating a sense of data movement and connectivity. The lines are set against a dark blue rectangular area at the top and a solid red rectangular area at the bottom.

WITH DR. RICHARD GALENTINO
& DR. PATRICK SCHUERMANN

FLOWINGDATA.COM

WILEY

Data Fluency

Data Fluency

**Empowering Your Organization with
Effective Data Communication**

Zach Gemignani

Chris Gemignani

Dr. Richard Galentino

Dr. Patrick Schuermann

WILEY

Data Fluency: Empowering Your Organization with Effective Data Communication

Published by
John Wiley & Sons, Inc.
10475 Crosspoint Boulevard
Indianapolis, IN 46256
www.wiley.com

Copyright © 2014 by John Wiley & Sons, Inc., Indianapolis, Indiana

Published simultaneously in Canada

ISBN: 978-1-118-85101-2
ISBN: 978-1-118-85089-3 (ebk)
ISBN: 978-1-118-85100-5 (ebk)

Manufactured in the United States of America

10 9 8 7 6 5 4 3 2 1

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except as permitted under Sections 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 646-8600. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, or online at <http://www.wiley.com/go/permissions>.

Limit of Liability/Disclaimer of Warranty: The publisher and the author make no representations or warranties with respect to the accuracy or completeness of the contents of this work and specifically disclaim all warranties, including without limitation warranties of fitness for a particular purpose. No warranty may be created or extended by sales or promotional materials. The advice and strategies contained herein may not be suitable for every situation. This work is sold with the understanding that the publisher is not engaged in rendering legal, accounting, or other professional services. If professional assistance is required, the services of a competent professional person should be sought. Neither the publisher nor the author shall be liable for damages arising herefrom. The fact that an organization or Web site is referred to in this work as a citation and/or a potential source of further information does not mean that the author or the publisher endorses the information the organization or website may provide or recommendations it may make. Further, readers should be aware that Internet websites listed in this work may have changed or disappeared between when this work was written and when it is read.

For general information on our other products and services please contact our Customer Care Department within the United States at (877) 762-2974, outside the United States at (317) 572-3993 or fax (317) 572-4002.

Wiley publishes in a variety of print and electronic formats and by print-on-demand. Some material included with standard print versions of this book may not be included in e-books or in print-on-demand. If this book refers to media such as a CD or DVD that is not included in the version you purchased, you may download this material at <http://booksupport.wiley.com>. For more information about Wiley products, visit www.wiley.com.

Library of Congress Control Number: 2014946679

Trademarks: Wiley and the Wiley logo are trademarks or registered trademarks of John Wiley & Sons, Inc. and/or its affiliates, in the United States and other countries, and may not be used without written permission. All other trademarks are the property of their respective owners. John Wiley & Sons, Inc. is not associated with any product or vendor mentioned in this book.

*To our parents, who shared a love of art and joy of teaching
that we try to pass on to those communicating with data.*

About the Authors

This book was a collaborative effort built upon years of experience helping companies make better use of data. Zach led the writing effort and defined the Data Fluency Framework that is the foundation of this book. Chris is responsible for many of the design and data visualization ideas and approaches that we share. Richard contributed from his experience in healthcare, education, and nonprofits, conceived of the Data Fluency Inventory, and took on the task of coordinating with our editors. Patrick worked with our research associate Tim to contribute content on organizational development, helping make this book a tool for leaders interested in transforming their organizations.

Zach Gemignani is co-founder of Juice Analytics and has helped build the company's reputation for designing engaging information experiences and delivering unique data visualization solutions. As CEO, he is responsible for the strategic direction, thought leadership, and business development of the company. Prior to Juice, Zach led reporting and analytics efforts at AOL and was a consultant with Diamond Technology Partners and Booz Allen, where he developed a reputation for creating exquisite slide presentations. He graduated from Haverford College with a Bachelor of Arts degree in Economics and received his MBA degree from The Darden School at the University of Virginia. Zach lives in Nashville, TN with his wife and three children.

Chris Gemignani is co-founder of Juice Analytics and the company's technology visionary. Chris earned his data chops in the credit card industry, taking on responsibility for risk modeling and analyzing cardholder behavior patterns. He combines this analytical experience with the ability to bring these insights to the screen with a hypercritical eye for user interface and interaction design. Chris graduated from Williams College with a Bachelor of Arts degree in Computer Science and Economics. He received a Masters in Economics degree from Washington University in St. Louis.

Dr. Richard Galentino. serves as CEO of Stratable, Inc., a strategic planning and organizational development consulting firm. Prior to launching Stratable, Richard led an international medical effort sending hundreds of doctors, nurses, and allied health professionals to more than 27 countries. Selected as a Harvard International Education Policy Fellow, Richard is a graduate of Harvard University (Administration, Planning, and Social Policy; Ed.M.) and Georgetown University's School of Foreign Service (Economics; B.S.F.S.). Richard earned his doctorate in education leadership and public policy from Vanderbilt University. Richard and his family reside in Nashville, TN.

Dr. Patrick Schuermann is a research professor at Vanderbilt University's Peabody College of Education. Having previously served as the director of policy and technical assistance for the national center on educator compensation reform for the U.S. Department of Education and the PI for numerous research projects in school leadership and education technology, Patrick currently serves as the director of the independent school leadership master's degree program and chair of the Peabody professional institutes. Patrick resides in Nashville with his talented singer-songwriter wife and their two dogs.

Credits

Executive Editor

Carol Long

Project Editor

Adaobi Obi Tulton

Technical Editor

Nathan Yau

Production Editor

Christine Mugnolo

Copy Editor

San Dee Phillips

**Manager of Content Development
and Assembly**

Mary Beth Wakefield

Director of Community Marketing

David Mayhew

Marketing Manager

Carrie Sherrill

Business Manager

Amy Kries

**Vice President and Executive Group
Publisher**

Richard Swadley

Associate Publisher

Jim Minatel

Project Coordinator, Cover

Patrick Redmond

Compositor

Maureen Forsy, Happenstance Type-O-Rama

Proofreader

Nancy Carrasco

Indexer

Robert Swanson

Cover Designer

Wiley

Cover Image

Courtesy of Zach Gemignani and
Chris Gemignani

Acknowledgments

A decade ago, Chris and I set out to create a company that bridged the gap between data and the people who might use it. This book draws from the many lessons we've learned from our discussions with colleagues and clients about presenting, visualizing, and sharing data.

I'm grateful and humbled by the energy and commitment demonstrated by all my colleagues at Juice. A special thanks to Ken Hilburn, James Lytle, and Michel Guillet for helping stretch our thinking along the way. Coming to work is always a joy when I get to collaborate with talented individuals including Djam Saidmuradov, Meghna Kukreja, Lindsay Conchar, Tim O'Guin, and Jennie Gemignani.

I'd like to thank Nathan Yau for inviting us into the Wiley family. His unflagging dedication to learning, sharing, and discussing all elements of data visualization is impressive and inspiring. As our project editor, Adaobi Obi Tulton has been a patient guide through the process.

I also appreciate the thoughtful efforts of Tim Drake, who contributed to the writing and research of the book. Tim is a Ph.D. student in Education Leadership and Policy at Vanderbilt University. Tim writes, researches, and teaches in the areas of quantitative research design and methods, data-driven decision making, and K-12 education leadership and policy.

Finally, a heartfelt thanks to my wife, Andrea, and kids—Owen, Maya, and Lila—who have been supportive and patient as I took on yet another responsibility.

Zach

Contents

<i>Foreword</i>	xix
<i>Introduction</i>	xxi
Chapter 1	
The Last Mile Problem	1
The Information Age: Driving the Need for Data Fluency	2
Data Fluency: Unlock the Potential Energy of Data in Your	
Organization	4
Big Data and Data Metaphors	5
Our Data Fluency Framework.....	7
Case Studies: A Window into the Framework for	
Data Fluency	8
<i>Data Consumers: Fantasy Football</i>	8
<i>Producers of Data Products: U.S. News</i>	11
<i>Organizational-Level Consumers: School District Woes</i> 13	
<i>Organizational-Level Producers: Insurance Company Bottom Lines</i>	15
Chapter 2	
The Data Fluency Framework	19
The Data Fluency Framework	21
Individuals and the Organization.....	22
Using Data versus Presenting Data	23
Element 1: Data Literate Consumers	23
Element 2: Data Fluent Producers	24
Element 3: The Data Fluent Culture.....	26
Element 4: The Data Product Ecosystem.....	27
Connective Tissue.....	28
Resources for More Depth	28
Benefits of the Data Fluent Organization	29
How to Use This Framework	30
How Organizations Struggle	31
Summary.....	32
Chapter 3	
How Organizations Struggle with Data Fluency	33
Pitfalls on the Path to Data Fluency	35
Report Proliferation	35
Balkanized Data	36

Data Elitism	38
The Supermodel	39
Searching for Understanding.....	40
Data Care	42
Metric Fixation	43
Finding Balance	45
Chapter 4 A Consumer’s Guide to Understanding Data	47
Data Products	49
Everyday Data Products	51
Barriers to Using Data Products	55
<i>Jargon</i>	55
Not Knowing Where to Start or What to Focus On.....	57
Inconsistency.....	58
Learning the Language of Data	60
Atomic Data.....	60
Summarized Data	61
Exploring Data Tables	63
<i>Rows Tell Stories</i>	63
<i>Columns Give the Bigger Picture</i>	64
Understanding Charts and Visualizations	64
Comprehensibility.....	65
Dissecting Data Products.....	66
<i>Where Does It Come From?</i>	67
<i>What Can I Learn from It?</i>	68
<i>What Can You Do with It?</i>	74
Wrapping Up	77
Chapter 5 Data Authors: Skilled Designers of Data Presentations	79
A Rare Skillset	80
What You’ll Learn.....	81
Guided Conversations	82
Finding Your Purpose and Message	84
<i>Let the Data Speak</i>	85
<i>Your Objectives</i>	85
<i>Your Audience</i>	86
Information Discrimination	88
Defining Meaningful and Actionable Metrics	92
Creating Structure and Flow to Your Data Products	95
<i>A Guided Path: Structure and Flow</i>	95
<i>Why Structure Matters?</i>	97
<i>Structure Options</i>	97

Designing Attractive, Easy-to-Understand Data Products	103
<i>Form</i>	103
<i>Visualizing Your Data</i>	104
<i>Color</i>	107
<i>Typography</i>	109
<i>Wrapping Data in Context</i>	111
<i>Language</i>	112
Creating Dialogue with Your Data Products	114
<i>Your Audience's Audience</i>	114
<i>Data Leading to Dialogue</i>	114
Design Principles	115
<i>Visualizations</i>	115
Design Principles for Data Products	121
<i>Compactness and Modularity</i>	121
<i>Gradual Reveal</i>	121
<i>Guide Attention</i>	122
<i>Support Casual Use</i>	122
<i>Lead to Action</i>	123
<i>Customizable</i>	124
<i>Explanation before Information</i>	124
Viva the Authors of Data Products	125
Chapter 6 The Data Fluent Culture	127
Leadership, Culture, and Communicating Priorities	129
<i>Set and Communicate Expectations</i>	130
<i>Articulate Specific, Measureable Indicators</i>	131
<i>Celebrate Effective Data Use and Products</i>	132
<i>Use Data to Inform Decisions and Actions</i>	133
Establishing Key Metrics to Rally Around	134
<i>What Makes a Good Metric?</i>	134
Using Metrics to Drive Organizational Improvement....	137
<i>Choose a Few Key Metrics at Any Given Level</i>	138
<i>Select Key Metrics That Align with the Mission and Vision</i>	138
Show Employees That Their Contributions Are Essential	138
<i>Reference Key Metrics and Data Analysis When Communicating Goals</i>	139
<i>Avoiding Metrics Pitfalls</i>	139
Shared Understandings.....	141
Common Vocabulary and Terminology Relating to Organization-Specific Data	143
<i>Clear Definitions of Measures</i>	143
<i>Standard Forms for Collecting Data</i>	143

Understanding and Appreciating Credible, Reliable Data Sources.....	145
<i>Understand the Strengths and Weaknesses of Data Sources</i>	145
Provide Transparency into How Data Is Manipulated and Modeled	147
Define a Shared Set of Key Metrics	148
Understanding the Purpose and Motivation for Data Products	149
Everyday Activities.....	151
Data Consumers.....	152
<i>Help Individuals Evaluate Data without Distraction..</i>	152
<i>Focus on the Message.....</i>	153
Data Products.....	153
<i>Establish Clear Guidelines for Quality</i> <i>Data Products.....</i>	153
<i>Develop a Feedback Mechanism for Data Products to Help Evolve and Improve Content.....</i>	155
<i>Celebrate Examples of Quality Data Products.....</i>	156
Data Usage.....	156
<i>Encourage Data-Driven Decision-Making</i>	157
<i>Evaluating Effective Data Use within the Organization.....</i>	157
Evolution of Data Fluent Cultures	158

Chapter 7

The Data Product Ecosystem	161
Data Products for Information Delivery.....	162
Necessary Conditions	163
Learning from the App Store.....	165
Demand.....	167
Top-Down Demand Map	167
Grassroots Needs.....	169
Where to Begin.....	169
Design	170
Objective.....	170
Start with a Style Guide	172
Develop	172
"It's a Poor Craftsman Who Blames His Tools"	174
Discover.....	175
Objective.....	176
Where to Begin: A Centralized Inventory of Data Products	176

Discuss	177
Objective.....	177
Where to Begin: Create a Place to Capture Insights	178
Distill.....	179
Learning from Wikipedia.....	179
Objective.....	180
What Can You Do Without?	180
“Only Connect”.....	181
Chapter 8 The Journey to Data Fluency	183
Why Data Fluency?	185
Data Consumers: Creating a Sophisticated Audience.....	187
Data Product Producers: The Skills to Enable Effective Data	
Communication	188
Data Fluent Culture: Building a Shared Understanding of Data.....	189
Data Product Ecosystem: Tools and Processes to Facilitate the	
Fluid Exchange of Information	189
Begin the Journey	190
Feature The Data Fluency Inventory	193
The Data Fluency Inventory Survey Questions	194
Component 1: Data Consumer Literacy.....	196
<i>Use of Data</i>	197
<i>Data Skills</i>	198
<i>Value Placed on Data</i>	199
Component 2: Data Product Author Skills	200
<i>Tools</i>	201
<i>Skills</i>	202
<i>Perceptions and Attitudes</i>	203
Component 3: Data Fluent Culture	204
<i>Leadership</i>	205
<i>Key Metrics</i>	205
<i>Shared Understanding and Everyday Data Use</i>	206
Component 4: Data Product Ecosystem	207
<i>Demand and Design</i>	208
<i>Develop</i>	208
<i>Discover</i>	209
<i>Discuss and Distill</i>	209
Summary of the DFI	210
DFI Scoring Guide (for Organizations)	210
Question Types and Point Values.....	211
Organization Scores.....	211
Component 1: Data Consumer Literacy.....	211
<i>Thoughts for an Organizational Leader</i>	212

Component 2: Data Product Authors	212
<i>Thoughts for an Organizational Leader</i>	213
Component 3: Data Fluent Culture	213
<i>Thoughts for an Organizational Leader</i>	214
Component 4: Data Product Ecosystem	214
<i>Thoughts for an Organizational Leader</i>	214
Organizational Scoring Summary.....	215
Scoring Guide (for Individuals)	215
What Can You Measure as an Individual?	216
Component 1: Data Consumer (Individual Only).....	216
Component 2: Data Authors (Individual Only)	217
Individual Scoring Guide Summary.....	217
DFI Supporting Materials.....	218
Survey Introduction E-mail	218
Data Literacy Quiz.....	219

Appendix A

Designing Data Products	223
A Checklist for Creating Data Products	224
Think Like a Designer	226
Designed to Be Used	228
Breaking Free of the One-Page Dashboard Rule	229
Dashboard Alerts Checklist	231
Context: Users Need to Understand How an Alert Is Defined and How It Fits into the Larger Picture.....	232
Cogency: An Alerting System Needs to Avoid Causing Unnecessary Alarm While Delivering Easy-to- Understand Information That Can Be Acted Upon... .	232
Communication: Alerts Must Be Designed to Effectively Capture Attention and Inform	233
Control: Advanced Alert System Should Give Users the Ability to Customize and Manage Alerts.....	233
8 Features of Successful Real-time Dashboards	234

Appendix B

Style Guide	237
Style Guide Sample 1: Fonts.....	239
Style Guide Sample 2: Colors.....	240
Style Guide Sample 3: Date/Number Formatting.....	241
Style Guide Sample 4: Bar Charts.....	242
Style Guide Sample 5: Trend Charts	243
Style Guide Sample 6: Tables.....	244
<i>Index</i>	245

Foreword

It's been more than a decade since I took my first statistics course in college. Unlike for many, my introduction to statistics brings back happy memories of an enthusiastic professor who jaunted up and down the stairs of the lecture hall. It's not easy to get excited about beginning concepts in distributions and hypothesis testing, but he pulled it off. I grew interested in working with and understanding data which eventually led to many years of graduate school. I had no clue back in college that statistics—or more generally, using data—would be so popular now. I just liked to play with data. And there's a lot of data to play with these days.

Every day I read or hear about companies and organizations that use data in some way. There's a wide array of applications: improving business, providing better service to customers, helping to make the lives of others easier, or communicating complex processes. There's an excitement. People want to gain insights from all this data they collected.

There's a gotcha though, and it's a big one. You can't just take a stream of data, plug it into the most expensive software you can find, and gather instant results—regardless of whether you're one person or a big organization. It's never that easy. Anyone who tells you otherwise either doesn't know what he is talking about or is trying to sell you something.

As someone focused on data visualization, I would love to build a dashboard or develop an interactive tool that enables people to understand their data in an instant. No background needed. However, you have to learn how to use the tool before anything worthwhile comes of it. You must know what data represents and how to analyze and interpret.

When you start to look at how an entire organization can grow more fluent in the language of data, you introduce other challenges. Those in management have different responsibilities than those working on the floor, but there must be a proper foundation for everyone to work together in an effective way.

Zach and Chris Gemignani, co-founders of Juice Analytics, help groups with these challenges every day, and now they educate others with *Data Fluency*. The two brothers and their team

have been consulting long before “big data” became a thing, before Google’s chief economist Hal Varian said that the job of a statistician is sexy, and before I started FlowingData. The Gemignanis’ experience shows in their articles online and in this book. Their advice is practical but general enough so that you can apply frameworks to your own situation.

When I first searched for “data visualization” years ago, the Juice Analytics site was one of the first ones I found and still subscribe to today. So I was excited when Zach and Chris agreed to write *Data Fluency*. However, this isn’t a book about visualization. It certainly covers the topic, but *Data Fluency* provides a wider view.

When you have visualization floating around in your organization—reports, talking slides, and data displays—does it actually matter if no one looks or gets anything out of it? It ends up in the recycle bin or as background noise. You can have the most efficiently designed charts in the world, but at the end of the day, you need people to pay attention. The goal is to bring data closer to the front so that everyone from management on down can make better informed decisions.

At the same time, there is no promise of a panacea or a new tool to make all data problems go away. It’s a realistic view that stems from the Gemignanis’ experience. They understand that often a lot of moving parts in groups might move slower than others or are difficult to change. I’m just a one-man show with FlowingData, but in my own consulting work, I understand the pains of bureaucracy all too well. The key is to work with the areas that do change and go from there. *Data Fluency* is an excellent guide to figuring out how you can do this.

Sitting here, thinking about what data might look like another decade from now, I can only imagine more of it, at a more detailed level. In the present day, the rate of collection far exceeds the rate at which we can understand. However, the growing rate at which people *want* to understand is a different story. So the more people who can learn the language of data now, the better we will be for it later.

Nathan Yau

Introduction

How do you change minds?

My brother and I huddled in my basement, putting the finishing touches on our analysis. The sun had set and our presentation was the following morning. We had spent the last month gathering data about student retention at online schools. We wanted to know what caused students to leave and what kinds of students tended to stay.

We had the slides to share with the executive team. The presentation summarized an attrition model, segmented the student population, and offered recommendations. Yet we felt something was missing.

How could we teach people to care?

Behind our numbers were individual students who chose the online school, took out student loans to pay for their education, spent hours with the online coursework, consulted with teachers, and tried to keep up with the schedule. Our analysis flattened the individual stories, successes, and struggles of these students. How could we bring real life back into our presentation?

As the clock ticked toward midnight, we got an idea: We'd create an animated movie. It would show how every student found their way into the school from different points of origin, how they progressed through their schoolwork, and how they eventually made the decision to stay or leave.

The movie-making was more quick and dirty than elegant. We constructed images showing where each student existed on their journey then joined them into a single image for each day of the school year (Figure 1). The students were positioned precisely and moved like stop-motion figures. Finally, Chris wrote a script to generate hundreds of single-day snapshots then weaved

them together into an animation. For a couple of tired data junkies, our couple minutes of movie magic felt like a Spielbergian masterpiece.

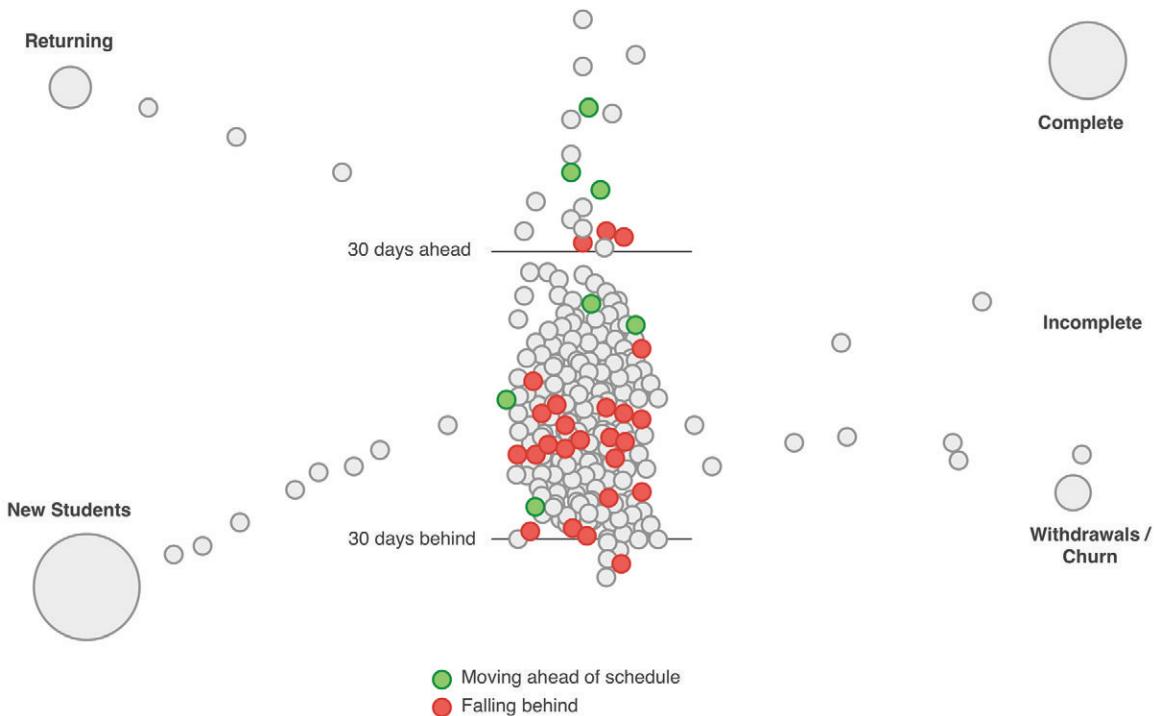


FIGURE 1 A point in time from our movie about student retention

The students marched across the screen on their way to joyful completion or disappointing withdrawal. The data had new life. And it sparked a conversation with our client.

That creative exercise ignited a passion. We had started Juice Analytics a few months prior, knowing that we wanted to help businesses gain insight and understanding from their data.

That night helped us turn from focusing solely on the numbers to how they are communicated. We realized we wanted to find better and more creative ways to help people understand data. Could we bridge the gap between data analysts and the people who can take action from their work?

For almost a decade we have pursued this goal. Juice Analytics has worked with over a hundred companies—from start-ups trying to deliver data to their customers to global brands looking for better ways to communicate data to executives. We’ve designed engaging interactive dashboards, reports, and analytical tools—all with the goal of helping real people make sense of and act on data.

Along the way, we’ve learned a few important lessons.

DATA IS THE NEW LANGUAGE OF BUSINESS

Data is a medium to communicate and convince. Its value has been recognized and elevated over the last few years. Media sites such as FiveThirtyEight (from ESPN) and Upshot (from *The New York Times*) are creating public discussions that combine data analysis and visualization with journalistic storytelling. These sites are a public expression of a trend that has been percolating within many smart organizations.

However, not everyone is comfortable communicating with data. Many of the audiences we design for—administrators, attorneys, marketers—are unfamiliar and inexperienced with getting value from data, even in small doses. One of the great challenges of data communication is building a dialogue. As much as a speaker must express himself through clear, accurate data presentation, the audience needs to be a willing and capable recipient. Presenters of data need to meet their audiences where they are, in ways that their audience can comfortably engage with the content.

How do you create common ground for more effective data communication? You can start by teaching the fundamental grammar of data visualization: metrics, dimensions, distributions, relationships, outliers, and variance. You can encourage good choices for how to express data by picking the right chart to emphasize the important elements in the data. You can learn from the expert data communicators to see how they fluently use the language of data.

More than ever, data will be a large part of how you convey messages. You need to ensure that everyone in your organization can participate in the discussion.

DATA COMMUNICATION IS A SOCIAL PROBLEM, NOT A TECHNOLOGY PROBLEM

For years, many organizations found it important to strive for data volume and invest in bigger databases and feature-rich platforms. Lost in the focus on size was the real prize—actionable insights in the hands of people who can do something about it. The first generation of business intelligence was about delivering complex, full-featured solutions designed for the IT team. Yet vast quantities of data collected by organizations remain disconnected from the people who might make use of it.

Making data useful is a problem that ultimately must be solved by people—people who understand the specific context of the data, people on the front-lines of decisions, and people who deeply understand the problems that data can illuminate. Data is useful when people use it to tell stories, craft compelling visualizations, and construct thoughtful analyses. People are the missing ingredient.

Unlocking the value of data takes more than individual efforts. It takes the interactions between people who communicate with data, discuss meanings, and debate what actions to take. There is a need for a data culture within organizations that embraces informed decision making.

Data is a cold, lonely medium on its own. Data needs to be humanized and human-sized. It needs to be made relevant to the audience by being clearly linked to relatable problems. It should be presented in intuitive, visual, and simple ways. And like any language, data should be about conveying meaning.

CONNECTING AND COLLABORATING

Much of the conversation on data occurs across the great divide between those who have a cultivated knowledge of data and those who have responsibilities that seldom involve digging into data. There are language barriers, biases, and misconceptions between these groups of people.

On the one hand, consider a data analyst who has created a complex spreadsheet that helps explain inefficiencies in operations or perhaps defines

marketing channel attribution. To do this they have learned the intricacies of APIs, how data is collected and defined, where it is gathered and how it can be joined to other data. To them, data is a flashlight illuminating a bit of truth in a chaotic world.

The analyst's boss comes from an entirely different world. She is equally smart and invested in finding ways to improve the organization's bottom line, but she has little attention or time for a detailed spreadsheet or black-box data algorithm. She's more likely to be moved by a compelling story than a table of numbers.

If these types of people can find a way to collaborate, the organization can benefit. The analyst's work can see the light of day and drive smarter decisions. His manager can help keep the focus on the pressing problems where better analysis can impact actions.

Our goal has been to create a productive dialogue among those who are data fluent and those who are just learning the language of data. If we can do this, we can connect people who can ask the best questions with those who can answer the questions.

TURNING DATA INTO ACTION

I was presenting to an audience about the untapped value of data when I saw a hand shoot up in the back of the room. It belonged to a serial entrepreneur who I had known for years. He commented to the crowd, "Data isn't valuable. In fact, it is costly. Think about all the money that goes into gathering, storage, management, and software. The insights that can be found in the data, isn't that what's valuable?" He offered a valuable distinction. But maybe we should go further. It is only through actions taken that true value that can be unleashed from data.

The data industrial complex—big business intelligence providers like IBM, MicroStrategy, and SAP—have plenty of incentive to make you believe that gathering more data is a source of value. More data needs more powerful tools with complex feature sets. Bigger data is better data.

Don't buy it.

A core principle of product design, according to Joshua Porter, Director of User Experience for HubSpot, is “Usefulness is job #1.” He goes on to say, “If your product is not *useful*, if humans do not find use for it, then the design has failed. Your product must help people do something valuable in their lives.”¹

Your data is made useful when it helps people do something better in their job. When data is communicated in ways that are easy for people to grasp, the information can drive dialogue and discussion. The insights and stories in data can get people talking in a productive way. And from these discussions come better decisions and actions. Without this journey of insight to action, all your data might as well be hot air.

VISUALIZATION IS ONLY A PIECE OF THE SOLUTION

The human brain has an incredible ability to absorb and process visual information. We may not remember someone’s phone number or the name of that person we just met, but our visual systems can put computers to shame. Those annoying CAPTCHA systems used to verify that you aren’t a spamming robot are evidence that our minds can process information and find patterns with ease. This is the foundation on which data visualization has been built.

Over the last decade, we have had a front-row seat to watch the explosive growth of data visualization. In 2005, we started blogging about data visualization and a handful of passionate practitioners and academics were at the core of a fledgling data visualization community. Today, there are dozens of conferences, established academic programs, and thousands of designers eager to visualize your data in an infographic.

Lost in this beehive of activity is a simple fact: Data visualization is but a means to an end. That end is to effectively communicate ideas and insights by transforming and representing numbers in ways that everyone can understand. The means can—and must—go beyond charts, infographics, and sophisticated interactive visualizations.

How do we reach beyond infographics? Helping an audience work with data requires creating a logical flow through the information—a flow that has a clear beginning and an end that can lead to actions. Good data communication means providing guidance about the meaning of data elements and timely

insertion of contextual information. Presenting data isn't just about how that data is visualized, but also about how the user can interact, explore, and have a great experience that works with what they already know.

If data is a new medium for communication, there is something to be learned from the many other forms of communication that have come before—like print, photography, and film. A movie director does much more than string together a series of images. He connects to the viewer by artfully combining elements like music, sound effects, editing, and cinematography.

The role of the data communicator is similarly complex. The goal should be to create “information experiences” that transform how audiences think about a subject and make better decisions.

Whether better decisions are made with data often depends on organizational dynamics. Organizational culture deeply influences whether “data products” lead to better decisions. Processes need to support effective data communication. And, in dynamic organizations, effective data communication constantly shapes better processes, systems, and decisions.

It is these lessons, and many more, that we share in *Data Fluency*. Our ambitions are broad. We hope to construct a roadmap that organizations large and small can use to improve how they work with data.

WHO IS THIS BOOK FOR?

The journey to data fluency is important for any organization that wants to have data inform decision-making. It takes a diverse set of people to build the skills and culture for data fluency. Whether you are a leader or an analyst, this book offers practical guidance to help you and your organization on this journey. Data fluency requires:

- **Organizational leaders** who want to create a culture of data communication and usage, and need to build a team of savvy data consumers and analysts.
- **Data novices** who are beginning to learn how to communicate with data.
- **Experienced analysts** who work deeply with data and need to effectively communicate their findings.

- **Data product authors** who design and create dashboards, reports, and data-rich applications, and want their work to engage their audiences.
- **Technology leaders** who are responsible for managing data, creating data solutions, and ensuring the organization extracts maximum value from its data assets.
- **Human Resource managers** who need to select and train their staff with the skills necessary to work with data.

The first chapter explains the Data Fluency Framework in detail. With an understanding of the Data Fluency Framework, Figure 2 identifies a few of the best chapters to focus on if you are interested in quickly getting to the content most relevant to your role.

Chapter	Organization Leader	Technology Leader	HR Manager	Analyst	Content Creator	Data Novice
3. How Organizations Struggle Diagnose the challenges your organization is facing	✓	✓				
4. Consumers Guide to Data The skills you need to have to become an effective consumer of data			✓			✓
5. Producers Guide to Data The skills to communicate with data		✓	✓	✓	✓	✓
6. Data Fluent Culture Creating conventions and standards for a data fluent culture	✓		✓	✓	✓	
7. Data Product Ecosystem Creating an environment for the fluid and efficient sharing of data presentations	✓	✓			✓	

FIGURE 2 Chapter Guide

NOTE

1. Porter, Joshua. "Principles of Product Design." *Bokardo.com*, n.d. Retrieved from <http://bokardo.com/principles-of-product-design/>

CHAPTER 1

The Last Mile Problem

*Water, water, everywhere, and all the boards did shrink;
Water, water, everywhere, but not a drop to drink.*

*Samuel Taylor Coleridge
The Rime of the Ancient Mariner*

The Texas Oil Boom at the turn of the 20th century saw populations in Dallas double and double again in a few short years. This tremendous growth was driven by a wholesale shift of the U.S. economy and infrastructure to automobiles. The first Ford Model T left the factory on September 27, 1908. Less than 20 years later, in 1927, Ford had produced 15 million cars as supply and demand caused a massive societal shift—the United States literally drove into a new society.

Similarly, the advent and widespread use of the personal computer at the end of the 20th century led to a new boom: the Information Age. Thirty years after the launch of the personal computer, we entered a new era of big data and data-driven decisions that has been compared to the oil boom. Certainly the hype is gushing and hopes are high that *data will deliver smart insights and a more intelligent enterprise*.

This is the promise, but here's the hitch: Although generations of teenagers have learned how to drive as a rite of passage into adulthood, and with it have found new degrees of freedom, we have not systematized the acquisition of skills essential to ensure data fluency. The language at the heart of the highways of commerce, now, and even more so in the future, is available to relatively few.

The 1920s and 1930s saw the advent of factories to efficiently mass produce automobiles, and generations of insight inform our patterns of *consumption* of this critical product. Almost 100 years later, a critical set of skills is now needed to ensure individuals and organizations are thoughtful *consumers* of data, and emerging skill sets are essential to *produce* data-based presentations and actionable data products.

It is to the development of this critical set of skills—those of being informed, capable consumers of data, and of being accomplished producers of data presentations and products—that this book is dedicated. Our goal is to help individuals and organizations understand and develop data fluency, as we contend it is the new language, the new highway, of commerce in the 21st century.

THE INFORMATION AGE: DRIVING THE NEED FOR DATA FLUENCY

Fantastic advances in data storage capacity have fundamentally changed the trade-offs we need to make regarding what to keep and what to delete. Rather than having to carefully decide what elements of our digital reality to capture and which to throw away, we can now keep everything. We can have it all—and we do. According to independent research organization SINTEF (The Foundation for Scientific and Industrial Research), 90 percent of all the data in the world has been generated over the last two years.¹ With more Instagram pictures, more tweets, more history of where customers go on the web, we are rapidly growing the amount of data we can sift through.

In a sense, the raw materials for informed decision-making have never been more plentiful. Yet the promise of data nirvana still seems far off. Students, scholars, employees, and executives are often still making crude decisions based on chance, gut, or whims of the crowd. In this era of data as the new oil boom, where's the payoff? Are we making better decisions and are we better able to understand our world? Are we driving cars or still riding horses along the digital highway?

On the ground, in the organizations we've worked with at Juice Analytics, people are often frustrated by their inability to effectively use data. They've built data warehouses, invested in expensive business intelligence solutions, and spent finite fiscal resources to hire data scientists. They've data-mined, analyzed, defined key metrics, and created dashboards. Despite these efforts, data is often under-used and misunderstood.

Few people, and fewer organizations, consistently engage with the data and use it to guide their thinking. Our vision is for everyone, from front-line customer service agents to senior executives, to leverage the mountain of data at their disposal. Forget the complex Wall Street trading models or IBM's Watson computer diagnosing disease—data in your organization can and should be used in simple, incremental ways to improve conversations, focus resources on priorities, and make small, everyday decisions with clarity.

Making use of data is a problem common to organizations large and small, public and private, and across market segments. According to a study conducted by the consulting firm Avanade, “more than 60 percent of respondents said their employees need to develop new skills to translate big data into insights and business value.”²

With all the promise that data holds, and the hope that data can help us make more informed decisions, the big question is: What is causing the gap between the vast opportunity of data and the reality of organizations struggling to act on this data? Here are a few theories:

1. Many people are data phobic and unwilling to engage with data to make decisions.

While at a leading Internet media company, we witnessed analysis teams dutifully churning out detailed reports about how online content was performing while the report recipients, content managers, dutifully ignored the information. Decisions had always been made based on gut and continued despite more and more detailed data about content usage and users.

2. Technology and personnel limitations constrain organizations' ability to work with their data sources.

Many organizations we encounter lament their spreadsheet-driven culture. Every department has its own mechanism for gathering, analyzing, and reporting on its unique data. No consistent “source of truth” exists and data analysts become indispensable because they are the only people in the organization who know how a financial model works, how to access and understand the data sources, and its strengths and weaknesses. People in these organizations wish for a technology solution that could bring all the information together and make it available to all decision makers in interactive, visual dashboards.

3. Organizational constraints inhibit the effective use of data.

In school districts around the country, superintendents often lament the lack of good data. Indeed, the Gates Foundation is currently putting significant resources into developing district level dashboards to inform decisions. Yet, critical organizational challenges remain with respect to collecting the data in a timely manner, linking data to competency and performance assessments, and engaging teachers in the process.

We believe that data-phobia, technology limitations, or organizational dysfunction are symptoms of something broader—not the root causes of the lack of payoff we are currently realizing from data. The root cause is something we call “the last mile” problem. Fundamentally, failing to use data isn’t a technological problem, but a social problem.

The last mile analogy comes from telecommunications where bridging the final few feet from the big pipes carrying gigabytes of Internet traffic throughout your city to each individual house is the most costly. With data, collecting and storing information is the easy part. The technologists have done their job. It is analytics, application, and adoption that pose the greatest challenge. Although data storage can be done *en masse*, the last mile is personal and often organization-specific. Revealing insights, influencing decisions, and taking action requires skill and motivation at a personal and organizational level. This is the missing link—the last mile—requiring individual and organizational data fluency.

This book is about how organizations can more effectively communicate with data—both internally and with external constituents. It is about people and the specific skills needed to be capable consumers and effective producers of data-based reports and presentations.

DATA FLUENCY: UNLOCK THE POTENTIAL ENERGY OF DATA IN YOUR ORGANIZATION

In many ways, data is like oil—and it is certainly so in the economic engine of your organization. Just like you can’t pull crude oil from the ground and pump it directly into your gas tank, or mold it into a plastic LEGO® brick, you can’t dump data into an organization and expect it to be useful. Creating

value from data is a complex puzzle; one that few organizations have solved. Although there isn't a simple answer (and thus why so many organizations struggle), the good news is that understanding the nature of the problem offers a starting point for our path forward. Data fluency is the path—the ability to use the language of data to fluidly exchange and explore ideas that are important to your organization.

In this book, the goal is to help you unlock the potential of data in your organization. Your data challenges have less to do with technologies and organizational constraints, and more to do with developing the capacity of data consumption and production within individuals and organizational teams.

Data fluency applies to individuals (everyone needs the skills to “read and write” and “listen and speak” using data) and also to organizations that must create an environment that rewards productive data conversations. There are many practical resources and books that address individual data skills. However, there are fewer resources for helping transform an organization to achieve data fluency. This book draws on foundational organizational development literature as well as best practices from current industry leaders.

The goal is to offer a framework that can help you understand the pieces required to construct a data fluent organization. At the same time, it provides practical guidance that you can act on. You don’t live in the theoretical, so the insights in this book won’t stop there, but instead are rooted in real-life examples intended to provide actionable guidance.

The advice will be technology-agnostic. There are many tools that can help you communicate with data more effectively. You will like some more than others, and circumstances often dictate what you have to work with. The guidance offered in this book can be applied across the full spectrum of particular products or platforms you utilize and can help you make the best of them.

BIG DATA AND DATA METAPHORS

Data is everywhere. Peta- and exa- and zettabytes of data. Amounts that require mind-blowing analogies, like the 40 zettabytes of digital data that will be generated by 2020—a figure estimated to be 57 times the amount of all the grains of sand on the beaches of the earth!¹³ Though the amount is staggering, data is also captured from more sources and at a faster pace than ever before. In lockstep with the increase in the amount and kinds of data has been the technological capacity to house, store, and analyze it.

Big data—that is, the *volume, velocity, and variety* of information currently being gathered and stored⁴—represents the next great frontier, not only for business, but also for other sectors, including government, education, and healthcare. Anecdotes about the sophisticated use of data can catch your imagination and excitement:

Billy Beane's use of statistics to help the Oakland A's acquire undervalued players, an act that led to a playoff berth for a team that spent one-third the amount on salary as the New York Yankees.

Google analyzing Internet search patterns to provide early warning signals for flu and dengue fever outbreaks.

The United States Federal Government funding initiatives across departments to "greatly improve the tools and techniques needed to access, organize, and glean discoveries from huge volumes of digital data."⁵

Technology professionals, business leaders and academics all view data as a means to forecasting, reduce costs, identify new business opportunities, improve research, and improve sales efficiency and effectiveness. As a result, big data startups have flooded the market. Venture capitalists are investing more than one billion dollars a year in the sector, funding hundreds of companies.⁶

The possibilities truly seem endless. Imagine, all your enterprise data, seamlessly integrated, available anywhere, anytime, just waiting to be translated into dollars. In this regard, data truly does seem like the new oil. In fact, there have been many metaphors used to describe our current era of big data. Tyler Bell, who is a big data industry analyst for O'Reilly, classified these according to the following typology:

- **Natural resources** ("the new oil," "goldrush" and of course "data mining"): Highlights the singular value inherent in data, tempered by the effort required to realize its potential.
- **Natural disasters** ("data tornado," "data deluge," "data tidal wave"): Frames data as a problem of near-biblical scale, with subtle undertones of assured disaster if proper and timely preparations are not considered.
- **Industrial devices** ("data exhaust," "firehose," "Industrial Revolution"): A convenient grab-bag of terminologies that usually portrays data as a mechanism created and controlled by us, but one that will prove harmful if used incorrectly.⁷

Of course, the metaphor that may resonate with you is largely determined by your organizational capacity. Are you like 73 percent of a recently surveyed group of U.S. business and IT executives who admit converting large data volumes into actionable intelligence remains a challenge?⁸ Or are you like more than one-half of a representative sample of business leaders in the world that say that they have more data than they know what to do with? Do you view data as crude oil, something to be extracted, processed, and refined, a precious resource that must be converted if it is to be transformed into something of value? Or like gold, something to be targeted and mined? Or maybe a fire hose, essential when there's a fire, but overpowering when you just want a drink?

Clearly, embedded in these metaphors are a range of attitudes toward the data boom—from a fever pitch of excitement to dire warnings about the use (and abuse) of big data, warnings seemingly substantiated by high-profile data leaks.

Falling somewhere in between these two extremes are the vast majority of people, individuals who view data as valuable but are frustrated by an inability to capitalize on its value. Even leading research company Gartner sees big data slipping into the *trough of disillusionment* as interest begins to wane and organizations become frustrated as implementations fail to deliver value.⁹

So what has caused the fall? Why are many beginning to lose faith in the possibility and promise of data?

Simply stated, there is a growing recognition among many that gathering data, developing sophisticated storage systems, and hiring data scientists will not automatically translate into a competitive advantage. The efforts of countless executives, healthcare companies, school districts, and other governmental and nongovernmental organizations are a testament to this. A 2013 survey of companies that have implemented business intelligence solutions, shows that just 28 percent feel that the investment has delivered significant business impact.¹⁰ The frustration with data, then, stems not merely from failed expectations, but from the inability of many end users to be thoughtful producers, consumers, and users of it. This book attempts to remedy this problem at both the individual and organizational level.

OUR DATA FLUENCY FRAMEWORK

With the goal of helping you unlock the potential of data for individual work, for collaborative working teams, and entire organizations, we have developed

a framework for data fluency. The framework, as shown in Figure 1-1, portrays the skill sets and competencies that you can develop.

The framework specifies two primary categories of skill required to develop data fluency, namely those required to be an expert *consumer* of data presentations as well as the expertise required to be a skilled *producer* of data presentations. The development and application of these skills occur at two levels: the *individual* and the *organizational*. The goal is to catalyze the development of data fluency associated with all four quadrants of the framework.

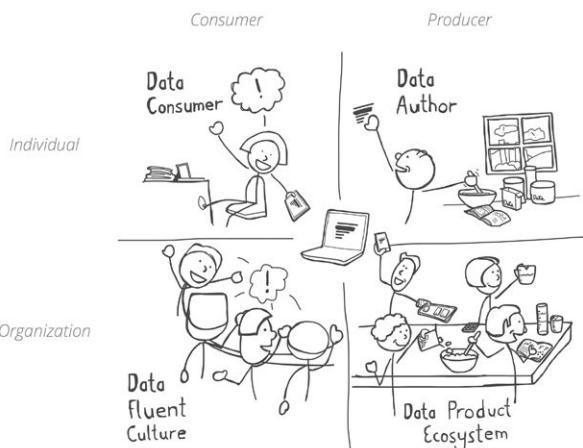


FIGURE 1-1 The Data Fluency Framework

CASE STUDIES: A WINDOW INTO THE FRAMEWORK FOR DATA FLUENCY

Subsequent chapters of this book are dedicated to helping individuals and organizations develop skill in each quadrant of the Framework for Data Fluency. This section provides four case studies to provide an initial window into the framework and illuminate the type of skills it addresses. For each struggle or promising practice outlined in the cases, subsequent chapters provide a set of actionable guidelines for individual and organizational growth.

Data Consumers: Fantasy Football

It may not be a stretch to say more Americans have learned about data and statistics through fantasy football than every college statistics course in the country. Each week, some 19 million NFL football fans spend their Sundays meticulously setting team line-ups based on statistical projections, historical patterns, and analysis of week-to-week variance. The couch potatoes who once relished on-field hits and in-game strategies now spend an average of more than eight hours a week diving into the data of the sport.

For the uninitiated, fantasy sports let fans play the role of team owners and managers by picking players for their own fantasy team and making weekly roster decisions. As the action plays out each week on the field, fantasy owners collect points against other competitors within their fantasy leagues. To

win, fantasy owners quickly realize that success often depends on studying player and team performance data closely.

Here are a few ways that NFL fantasy players incorporate data into their thinking:

Variation in Player Performance

The best fantasy owners understand the nature of week-to-week variance and its relationship to earning points. For example, touchdowns generally earn a fantasy owner six points; but touchdowns occur rarely and can fluctuate wildly. In contrast, the number of touches players receive may be a better indicator of how much the team is using them and their opportunity to provide the owner with points. Because *consistent* performance matters, successful owners often focus on players with more stable predictors of success (for example, touches) versus more sporadic events (for example, touchdowns).

Rankings Can Be Misleading

Fantasy football cheat-sheets offer rankings of players in every position. These rankings mask the differences and dispersion of expected performance. For instance, the top running back may be expected to perform 20 percent better than the second rated running back, who in turn is only expected to score five percent more points than the third through sixth rated running back. The data shows that players often cluster into tiers of performance. This statistical understanding was publicly explained by Boris Chen who stated that “players within a tier are largely equals. The amount of noise between the ranks within a tier and actual results is high enough that it is basically a dice roll in most situations.” This concept has been widely adopted by fantasy owners as a player drafting strategy.

The Only Constant Is Change

The worst fantasy football owners are stuck in the past and pick players and teams that they have relied on in the past to generate points. That is, they fail to update their assumptions about the best teams, players, and trends. Following the data closely reveals when certain players have gone past their prime and when teams that once had high-scoring offenses can no longer put up big points. Clinging to past success may be a formula for disaster because the only constant in fantasy football is change.

Context Fills Out the Picture

Data viewed in isolation can be deceiving. Say, for example, that your top wide receiver scored only one-half the number of points that he scored on

average in a season. Is this a new and troubling trend? Should you trade? A little research might reveal that he matched up against one of the league's top cornerbacks, or his quarterback was knocked out of the game, or perhaps he tends to perform poorly in cold weather, away games. These environmental factors make a difference with respect to outcomes. Performance data cannot be understood in isolation—context matters.

So how did fantasy football create legions of fans who have developed a specialized dialect of data fluency? It has been a combination of education, effective data presentation, common data conventions, and incentives. Fantasy football owners have been taught how to use data to their advantage through the efforts of the NFL, ESPN, Yahoo!, and a cloud of other websites dedicated to football analyses. Organizations like Football Outsiders built new media businesses around data modeling and projections of player performance.

Leading online fantasy football sites like ESPN and Yahoo! have been aggressive in pushing data and data visualizations to their users. These sites include trend charts for every player, drive charts, player comparison graphics, and predictive models for estimating game outcomes.

The educated fantasy football community is also highly engaged with the sport. The community loves football! The fantasy league has provided a whole new (and rewarding) dimension to its fandom. No longer is it tied down to rooting for a single team—instead, the whole league becomes fodder for its attention as it picks and chooses players from each of the 32 NFL teams. In addition, the fantasy football industry has coalesced around consistent formats for leagues, points, and key metrics (Figure 1-2). Terms like PPR, running back by committee, waiver wire, and flex are well understood, facilitating conversations among league owners. And with \$1.18 billion bet in fantasy football leagues annually and a passionate fan base, fantasy owners have huge incentives to make informed decisions. When money or bragging rights are on the line, individuals invest time and energy into developing the skills and abilities to become data fluent.



FIGURE 1-2 Yahoo! Sports fantasy football game status page

In short, these factors have brought data fluency to the masses. Millions of fans have learned how to read charts, grasp basic data concepts, and allow deeply embedded data to inform how they make decisions—all critical skills associated with quadrant one in our framework.

Producers of Data Products: U.S. News

The president of a prestigious nationally ranked university waited for the news. Ranked #20 in the previous year in *U.S. News & World Report's* Best College list, she felt enormous pressure from its large and influential alumni base to stay in the top 20. Though prepared to say that rankings didn't matter, she couldn't help but think about how she spent the past year touting its entry into the top 20 at every major fundraising and recruiting event. Not too far away from her office, high school seniors from the city's best school would use the rankings to guide their college selection.

The stakes are high. Economists James Monks and Ronald Ehrenberg found that a negative change in rankings would result in a decrease in applications, a decrease in applicant quality, and a drop in fundraising among select colleges and universities (Monks and Ehrenberg, 1999)¹¹. A recent study further found that moving to the front page of the *U.S. News* ranking provides a significant increase in enrollment across all institution types (Bowman and Bastedo, 2009)¹².

Yet these rankings are controversial. Critics argue that the rankings show little variation across years—a measure that favors those who are consistently on top, institutions like Harvard, Princeton, Stanford, and a handful of other prestigious schools. Proponents, however, contend that the rankings help standardized college reporting and open visibility to the public with respect to financial data, teacher quality, student satisfaction and a host of other variables. Deciding where to apply and attend from approximately 2,400 public and private 4-year colleges is no easy decision.

Despite the controversy over the rankings, popular interest in the list has grown. For example, every September when *U.S. News* releases its Best Colleges list, the USNews.com website experiences massive volumes of visitors. In 2013, 2.6 million unique visitors generated almost 19 million views in a single day around the release of the rankings.¹³ The online interest suggests people find the rankings to be an incredibly useful tool to sift through the vast amount of information available from colleges—information that is presented differently by each institution—to help guide college choice decisions.

So what can be learned from *U.S. News'* capability to bring data to a broad audience? How did it create such a successful formula? How did the *U.S. News* become the unofficial guide in the college attending process?

The *U.S. News Best Colleges* rankings list started as a reputational survey of college presidents. *People Magazine* had the “50 Most Beautiful People”; *Forbes* the “400 Wealthiest People.” Both were top sellers. *U.S. News* and its leadership team understood the power of rankings. Bob Morse, director of Chief Data Strategist for *U.S. News*, and colleagues did not just recognize the power of lists to sell—they were also skilled data producers. They recognized that lists were a powerful tool that could consolidate vast amounts of information into a single, easily digestible metric.

In creating and disseminating the best college rankings, Morse and his team demonstrate four important principles in becoming a skilled data producer.

Categorize Information

Morse sorted higher education institutions into categories based on the Carnegie Classification of Institutions of Higher Education. Over time, new categories have been created by *U.S. News* to meet the demands of consumers. U.S. colleges and universities are classified into national universities, national liberal arts colleges, regional universities, regional colleges, best-value schools, up-and-coming schools, top public schools, and schools with the most international students, for example. These categories provided college-goers with important information to begin their search.

Focus on Ranking and Comparison

Having established definable categories, other information could be used to rank and compare institutions within a given category. For instance, variables such as tuition and fees, total enrollment, acceptance rate, average freshman retention rate, 6-year graduation rate, class size, and average SAT scores could be used to sort and rank institutions. Ranking and comparison are easily accessible ways to present data.

Help Consumers to Understand What They're Seeing

Variables used to rank institutions were added over time. For example, the inaugural rankings in 1983 were based upon a simple survey to 1,308 college and university presidents asking them about the five best schools in their category. By 1988 the rankings included information on selectivity, faculty

strength, educational resources, and graduation rates.¹⁴ Over time, Morse and his team introduced new variables to the models that produced the rankings. After the public understood the basics, *U.S. News* added complexity to refine the measures.

Create a Data Experience for a Diverse Audience

To aid students and families in the college choice process, *U.S. News* created My Fit Engine, an interactive tool that allows them to enter their preferences and find a list of matching schools. Although rankings appeal to students and families working through the college choice process, they have just as much (if not more) appeal to higher education institutions. In polling these institutions, *U.S. News* found that institutions are often concerned with historical trends and patterns. Accordingly, *U.S. News* worked with Juice Analytics to create Academic Insights, an online, interactive tool to analyze and explore data over time.

The great innovation of the *U.S. News* list is that it organizes vast amounts of information into a uniform format that is easily understood and then distributes that information via readily accessible print and online formats. *U.S. News* created a system where information on hundreds of institutions could be processed and reduced to a set of lists. Instead of sorting through countless websites and glossy brochures, families now had a quick way to reduce the amount of time invested in the college search process. In doing so, *U.S. News* demonstrates the skills central to quadrant 2 of our framework—the skills to create effective data presentations.

Organizational-Level Consumers: School District Woes

The meeting ended badly. For the last five years, the district had worked tirelessly to implement a new teacher evaluation system—and a corresponding database and reporting system to support its use. The district even allocated a large percentage of its \$10 million grant from the federal government to hire “data coaches” to work with principals and teachers to develop familiarity with the data system and capitalize on its use. But the money ran out last year. The data coaches had to go, many to positions outside the district. The few that remained took central office jobs to work with the IT department on a wide variety of tasks, but with no unified purpose.

Now, one year later, the School Board called a special meeting to investigate the status and impact of this expensive endeavor. The central office team gave a dire report—unable to visit schools and without enough money to

provide ongoing professional development, it appeared that principals and teachers stopped using the system. The School Board members pressed the district about the functionality of the system by asking such questions as: *Is this a technological problem? Does your team not have the required staff to keep things running? Other than the data coaches, has anything else changed in the system since last year when most of the schools seemed to be using the data?* The district team responded that the data system not only functioned properly, but that it also had been improved to allow principals and teachers access to more reports and greater tools to help them explore and visualize the data.

"The system is way better now that we're working with the IT department," the leader of the department commented. "We're constantly updating what school personnel can do with it—the reports they can access, the tools they can use. We've made sure to send out weekly e-mails with all the changes and with reports that we believe are most useful. But we hear back from only a few who are really tech savvy and love data. We don't understand why the rest aren't using the system. It's clear that they don't like something about the system. They just never tell us what it is!"

As part of its presentation, the team provided information from a survey given to a random sample of schools in the district. The results were discouraging. Less than 10 percent mentioned using the system "regularly," and approximately 55 percent marked "never." Follow-up survey questions that explored the reasons for nonuse noted that a large majority of principals and teachers felt that the data systems and reports were "too complicated" to use. In addition, they felt that the district provided "too much information" and only 25 percent agreed that they provided "the right kind" of information.

After presenting the survey results, a group of teachers and principals that had been invited to the meeting were given time to discuss their experiences and answer questions. Their comments were revealing:

One principal said: "I don't have time to open every e-mail I receive from the data office. And when I do, it seems like they're always changing the format. Sometimes it's a PDF, other times I have to open them in Excel."

A teacher offered: "Half of the reports we get aren't even relevant to my school context and my student population."

Another teacher asked: "How am I supposed to use teacher value-added information when I don't even understand how they calculate it?"

A principal commented: "They're always changing how they do things. Some days I get on and find what I'm looking for, but then I go back a few days later

and it's not there! And things that I get from their team hardly look like other information I receive from HR or my area superintendent."

Another principal stated: "We're too busy to take the time to sit down as a leadership team and look at the data."

A final principal said: "We have a hard enough time as 'insiders' making sense of all the data . . . It's not like any parents are going to understand this stuff."

After the teachers and principals concluded their remarks, the School Board spent the rest of the meeting arguing over one major question: *Why did we invest so much money in a system that isn't being used . . . and where do we go from here?*

These challenges, which are explored in detail in Chapter 6, are all central to organizational efforts to create a foundation for a common interpretation of data presentations.

Organizational-Level Producers: Insurance Company Bottom Lines

Large companies spend millions of dollars a year on healthcare claims for their employees. Health insurance companies are responsible for explaining where all that money went, and why premiums go up annually, despite carefully designed programs to make a healthier workforce. Account executives at insurance companies have a common challenge: how to tell this story to their customers using data.

The goal is to show increased quality of health services being offered, better health outcomes for the employees, and lower employer costs. Telling a good story with actionable data would result in a contract renewal. Too often, however, these data presentations are met with boredom and frustration rather than enthusiasm. A lack of a good story line is often the result of disjointed data points pulled from a number of databases. Significant time is spent collecting data and then trying to compare it to national, regional, and industry benchmarks, and often little energy is put into telling the story. Or conversely, much effort is put into telling a good story—with little data to back it up. Either way, poor presentations are often followed up by requests for different metrics and more data, which may or may not be helpful. And for an account executive trying to close a deal, a poor presentation could mean no contract renewal this month.

Consider the challenge of collecting wellness data on 250,000 employees from a large company, and matching it with employee productivity, wellness,

and retention metrics at local and regional levels. First, data on the 250,000 employees is collected from many sources (employee, employer, and a variety of healthcare and wellness sources: hospitals, clinics, counseling centers, therapists, antitobacco programs, online programs, and even gyms). Some parts of the wellness program might be done in-house, whereas other campaigns like quitting tobacco and fighting depression were handled externally. Data collected on the 250,000 employees is then merged with that of 40 million other lives covered by this large insurance company—a large pool of data from which statistically significant comparisons could be made between those participating in a wellness program and those not. Data points came in daily—sometimes overnight via data sync between the employer’s employee database and the insurance company’s secure servers, and at other times via snail mail as in the case of gym reimbursement forms sent in by individual employees. Different departments collected different data points into 52 different databases. Extracting, analyzing, and presenting this data back to an employer in a useful and meaningful way is a challenge—demonstrating lowering healthcare costs, improving employee productivity and retention, and improving employee health and morale. Showing gains in these three areas would have to be contextualized for local, regional, and national trends.

Although this seems like an overwhelming task, large health insurance companies often have some of the most complete and comprehensive databases of employee claims and historical trends. Combined with top notch teams of statisticians and data scientists, it would seem pulling appropriate data for a large client presentation would be easy.

However, there are many breakdowns in the data ecosystem that prevent the systematic development and sharing of data presentations that can tell a compelling story. Perhaps these four challenges sound familiar:

- Data is stored in many separate, disjointed databases. Data is often collected from dozens of separate reports. Getting these reports may require a “ticket,” queuing it up for a database administrator (DBA). Data retrieved might spark interest in another thread of reasoning and require another “ticket.” And if reports are readily accessible without a DBA, significant time might still have to be spent logging into different data systems to collect the needed information.
- Consolidating all this data into a single PowerPoint presentation requires discretion. Synthesizing data and trying to make it look clean, professional, and visually engaging isn’t a simple feat. Different reports

will have different metrics, and making sure the different metrics are communicated clearly in a visually accessible manner takes skill.

- Telling a story is an act of creativity. Stringing these metrics into a story rather than a series of facts takes additional time and effort—and requires individuals who know both how to work technically and creatively.
- Balancing a client's desire for detail on disparate topics with time constraints is a trick. Attending to a client's conflicting desire for detailed data with a 90-minute meeting constraint means that only the most meaningful data should be shared—but done so in an authentic manner—superficiality will not carry the day.

Telling a convincing story with actionable data sets the stage for a different type of meeting between organizations and their clients—a meeting where the client isn't even thinking about contract renewal or renegotiation but rather how to solve problems informed by thoughtful analysis. In the end, many companies struggle to tell a compelling story, grounded in carefully crafted data reports, to nondata-savvy consumers. These quadrant 4 skills, which are targeted in Chapter 7, are critical to developing an ecosystem for building and sharing data.

NOTES

1. Dragland, Åse, "Big Data—for better or worse." SINTEF, retrieved on July 22, 2013 from <http://www.sintef.no/home/Press-Room/Research-News/Big-Data--for-better-or-worse/>
2. Avanade, "Global Survey: Is Big Data Producing Big Returns?," published June 2012, retrieved on July 22, 2014 from <http://www.avanade.com/Documents/Research%20and%20Insights/avanade-big-data-executive-summary-2012.pdf>
3. Lucas Mearian, "By 2020, there will be 5,200 GB of data for every person on the Earth," *Computer World*, December 11, 2012, http://www.computerworld.com/s/article/9234563/By_2020_there_will_be_5_200_GB_of_data_for_every_person_on_Earth?pageNumber=1
4. Doug Laney, "3D data management: Controlling data volume, velocity and variety," META Group Research Note 6 (2001), retrieved on July 22, 2014 from <http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>

5. Office of Science and Technology Policy, "Obama Administration unveils 'Big Data' Initiative: Announces \$200 Million in R&D Investments," retrieved on July 22, 2014 from http://www.whitehouse.gov/sites/default/files/microsites/ostp/big_data_press_release_final_2.pdf
6. Eric Sherman, "VC Bets \$1.4 Billion on Big Data in 2012," *Inc.*, February 28, 2013, retrieved on July 22, 2014 from [http://www.inc.com/erik-sherman/vcs-bet-\\$14-billion-on-big-data-in-2012.html](http://www.inc.com/erik-sherman/vcs-bet-$14-billion-on-big-data-in-2012.html)
7. Tyler Bell, "Big Data: An Opportunity in Search of a Metaphor," *Radar*, February 10, 2011, <http://radar.oreilly.com/2011/02/big-data-metaphor.html>
8. CompTIA, "2nd Annual Big Data Insights and Opportunities" study, Retrieved on July 22, 2014 from <https://www.comptia.org/resources/2nd-annual-big-data-insights-and-opportunities>
9. Svetlana Sicular, "Big Data Is Falling into the Trough of Disillusionment," published January 22, 2013, retrieved on July 22, 2014 from <http://blogs.gartner.com/svetlana-sicular/big-data-is-falling-into-the-trough-of-disillusionment/>
10. Cindi Howson, "BI Adoption Remains Flat," *BI Scorecard*, April 1, 2014, retrieved on July 22, 2014 from <http://www.biscorecard.com/bi-adoption-flat/>
11. Monks, J. and Ehrenberg, R. G. (1999). The impact of U.S. News & World Report college rankings on admissions outcomes and pricing policies at selective private institutions (CHERI Working Paper #1). Retrieved June 1, 2014, from Cornell University, ILR School site: <http://digitalcommons.ilr.cornell.edu/cheri/1>
12. Bowman, Nicholas & Bastedo, Michael. "Getting on the Front Page: Organizational Reputation, Status Signals, and the Impact of U.S. News and World Report on Student Decisions." *Research in Higher Education*. August 2009, Volume 50, Issue 5, pp 415-436. Retrieved on June 1, 2014 from: <http://www-personal.umich.edu/~bastedo/papers/BowmanBastedo.ResHE2009.pdf>
13. U.S. News Breaks Online Traffic Record, published September 12, 2013. <http://www.usnews.com/info/blogs/press-room/2013/09/12/us-news-breaks-online-traffic-record>.
14. Luke Myers and Jonathan Robe, "College Rankings: History, Criticism, and Reform," *Center for College Affordability and Productivity*, March 2009, retrieved on July 22, 2014 from *Productivity* http://www.centerforcollegeaffordability.org/uploads/College_Rankings_History.pdf

CHAPTER 2

The Data Fluency Framework

Motion pictures are among the most popular media for telling stories. They didn't become so overnight. Over their 100-year history, changes in the industry have led to disruptions and challenges that shape the way you experience movies today. This evolution, though, has brought a rich and distinctive system to deliver content and fluently communicate ideas in ways that engage audiences.

The *technologies* of moviemaking have changed dramatically over the years. From the initial transition from silent films to "talkies" to innovations in animation and 3-D, filmmakers have continuously learned new skills as technologies create new opportunities for storytelling. At the same time, the film industry has understood that establishing standards creates a healthy ecosystem. *Standards*—everything from film aspect ratios to cinema sound systems to DVDs—give producers a shared set of tools and expectations to work with. At the same time, MPAA ratings, well-understood movie genres, and industry awards help consumers know what to expect when they buy a film experience.

The people who make films have undergone their own evolution as well. Out of a period when directors were under the control of studio executives emerged *auteurs*—talented directors whose passion and personal influence are such that they are considered "authors" of the films they produce. The industry also

created film schools and apprenticeship systems to educate new generations of talent. Meanwhile, *audiences* have become more educated and refined in their understanding of the medium. The unique tastes of a niche audience can be efficiently matched to films that meet those interests. The films that touch a nerve with their audience are lauded while weak films are publicly panned.

Together, these foundational elements—innovative technologies, industry standards, creative producers, and educated consumers—have created the modern film industry. George Lucas and the beginning of the Star Wars saga is an illustrative example of innovative technologies, industry standards, creative producers and consumers coming together way back in 1977—resulting in the pop culture blockbuster (Figure 2-1).

A similar transformation is currently transpiring for the communication of data. As demonstrated in the film industry, a multitude of factors must come together to create an established environment with the necessary language, standards, and tools to facilitate communication between data producers and consumers. Data literacy for an individual is one thing; creating fluency across an organization and an industry is more complex. The community needs to define how data can and should be communicated. Individuals need to learn a new sophistication in engaging with data as a communication medium.

This book offers a path to a data fluent organization—achieved through a web of connected elements. Data fluency requires people who speak the language of data; skilled producers of data products; an organizational culture with the conditions to support data discussions; and the systems, tools, and ecosystem to create and share data products. These foundational elements—and the flow of information between them—represent the necessary conditions for data fluency in your organization. If that sounds complex, it is. It also helps to explain why so many organizations struggle in their pursuit of data fluency.



FIGURE 2-1 The Star Wars movies found their audience.

Nevertheless, it is a worthy journey—even if you make it only part way there. Data is a powerful, underutilized asset in most organizations. Creating organizational fluency with this data can enable productive conversations, smart decisions, and organizational priorities. It can influence the many small decisions that are made by your front-line staff every day. And it can guide big thinking about innovation and strategic direction.

When individuals and teams within an organization achieve data fluency, unproductive discussions about “where did that data come from?” and “what does that value mean?” are exchanged for deeper discussions about key organizational goals and metrics. Data fluency speeds the communication between people, letting them focus on the message and not the medium. Data fluency empowers people to more effectively do their job, informed by a shared data language and organizational understanding of that data—understanding that accelerates organizational growth.

By the end of this chapter you should understand:

- The components of the data fluency framework
- How a set of existing books and resources on data visualization skills and data-driven enterprises offers useful but targeted portions of the overall solution
- Various ways to use the framework to improve your organization

THE DATA FLUENCY FRAMEWORK

The data fluency framework, provided in Figure 2-2, is composed of four foundational elements. The elements are grouped along two dimensions. Along the horizontal dimension is a distinction between the capabilities required at an *individual level* (top) and at the *organizational level* (bottom). Along the vertical dimension, you can focus on the *audience*, or *consumers*, of data (left) separate from the *authors*, or *producers*, of data products (right). Below we clarify these distinctions and introduce how to develop data fluency in each quadrant—topics explained in detail in subsequent chapters of the book.

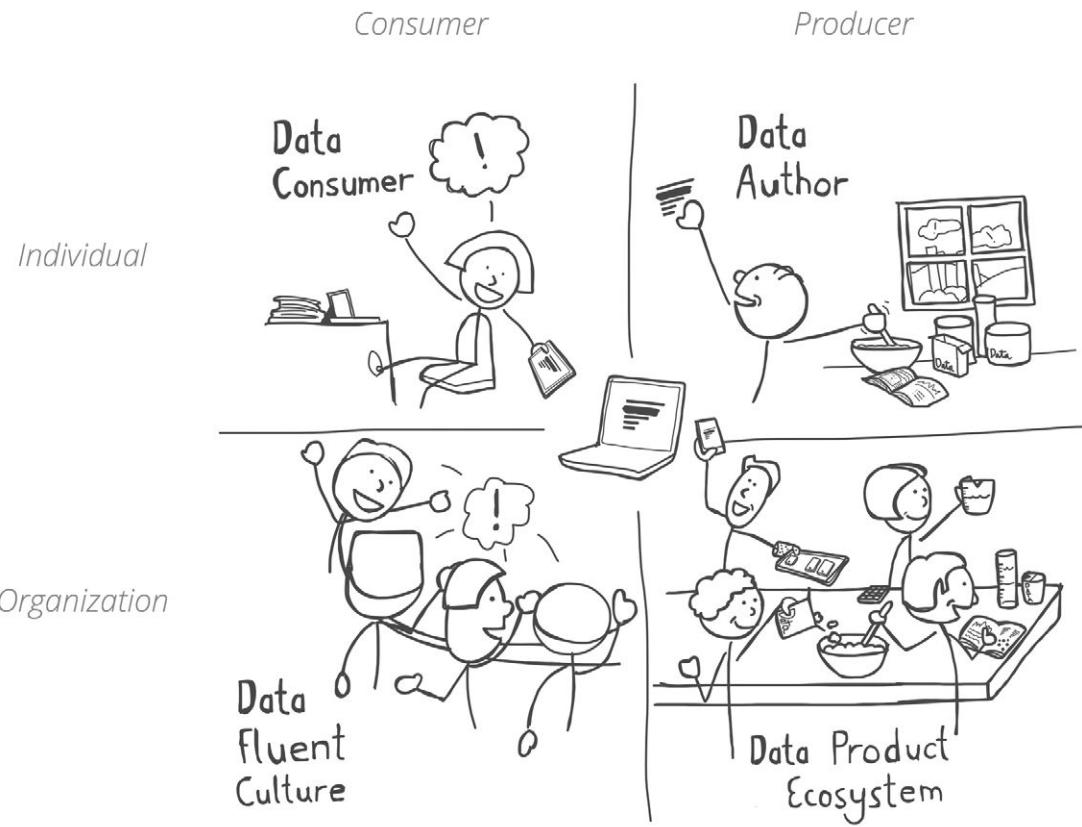


FIGURE 2-2 The four quadrants of the data fluency framework

INDIVIDUALS AND THE ORGANIZATION

Your organization has its own ways of doing things and often an intangible set of conventions and expected behaviors. Your staff both shapes and is shaped by these processes and culture. Developing data fluency skills at an individual level and creating an environment in which data fluency flourishes is important. Both are necessary to create a data fluent organization. Individuals trained in communicating with data will grow frustrated in an organization without the tools and expectations that align with their skills. Likewise, the best efforts of leaders to shape a data fluent culture will make little progress without individuals who have the ability to engage with data.

USING DATA VERSUS PRESENTING DATA

Supply and demand. Authors and readers. Producers and consumers. Data fluency is the exchange of ideas with data as a medium. Without both sides of an exchange, no message can be conveyed. However, separating people into consumers or producers of data content is a bit simplistic; the same person may play both roles at different times. Sales people may consume reports about organizational performance, product trends, and sales scoreboards. At the same time, they can be well served to communicate to the sales manager about their performance with data analyses rather than anecdotal complaints about poor quality leads. Within the framework, you can learn how to build a cohort of sophisticated data consumers and a culture that facilitates shared understanding. On the producer side, the framework offers guidance on teaching the skills for effective data communication and the systems required to make these producers successful. Now take a deeper look at the pieces of the data fluency puzzle.

ELEMENT 1: DATA LITERATE CONSUMERS



If *data fluency* is the fluid exchange of ideas using data, *data literacy* is the ability for individuals to understand and draw meaning from data. You should start here because reaching a willing and informed audience is perhaps the most fundamental condition for a data fluent organization. The famous philosophical thought experiment asks, "If a tree falls in a forest and no one is around to hear it, does it make a sound?" That is, does something exist if it is not perceived? Does your data matter without someone to read, understand, and interpret it?

Yet data is a relatively new language for most people. The goal of data literacy, like reading literacy, is to develop control over surface-level comprehension so the reader “can focus on understanding the deeper levels of meaning embedded in the text.”¹ It takes time to move from the grasp of individual vocabulary to the instantaneous internalization of meaning. It is a worthwhile journey because this new language of data is useful for interpreting your world. It can be argued that data will become a new lingua franca, that is, a language systematically used to facilitate communication possible between people not sharing a mother tongue.²

In Chapter 4, which focuses on element 1, you can see that it takes more than a solitary listener to give meaning to your data. It takes an educated audience that appreciates the nature of the data presented, knows core data concepts, can find insights in data presentations (both simple tables and advanced charts), and—like the movie critic—sets high expectations.

ELEMENT 2: DATA FLUENT PRODUCERS



Not always, but too often, authors of data communications forsake their role. You've certainly been witness to the results: raw data tables delivered as “reports”; long PowerPoint presentations with slide after slide of repetitive

charts; and dashboards densely packed with colorful figures. What the authors are actually saying with these data products is:

- I don't know what is important in the data.
- I don't respect the time and attention of my audience.
- I don't know how to visually communicate data.
- I don't know that I need to communicate meaning.

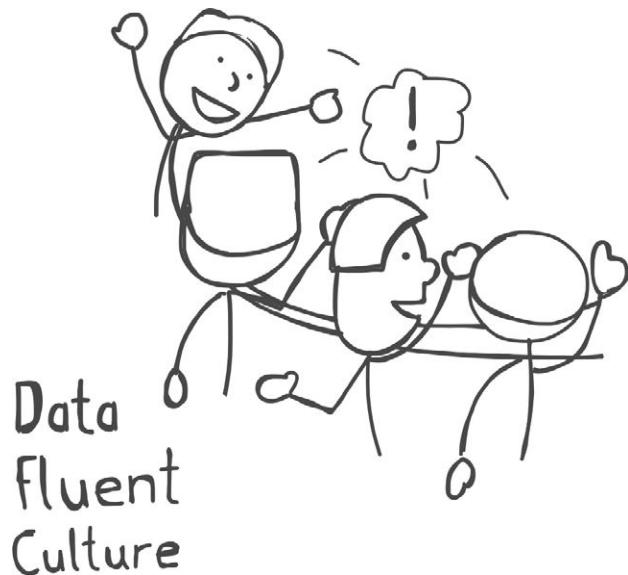
A common problem is awareness of one's role. Data producers are first and foremost teachers tasked with communicating information and actionable insights from the data. In this regard, it may be helpful to consider the Montessori educational system. In contrast to more traditional forms of teaching, Montessori teachers consider themselves guides for their students. Rather than dispensing information to students in a teacher-deliverer model, Montessori teachers create a fertile environment in which students can engage in their own learning, have the necessary tools needed for learning, and have the freedom to find what is of interest to them. Similarly, data authors need to consider the needs and actions of their audience and deliver content in ways that are easy to absorb.

Every member of your audience comes to the data with different priorities, needs, and perspectives. Successful communication of data builds on this variety as an asset—everyone in the discussion can add to the overall understanding of the group and also find their own insights.

Chapter 5 focuses on best practices for creating effective data products that tell stories while engaging and guiding audiences. You need to find your message and be discriminating in the content that you put on the page. Just as no book or movie can convey everything about a topic, don't try to say everything that can be said in your next report.

Chapter 5 also explains the nuts and bolts of presenting data, including choosing the right metric and the right chart to sharpen your message. And when your data product starts to become complex, it becomes important to consider how the information fits together. The structure and flow of data visualized together makes a huge difference in making the message and meaning accessible to your audience. Likewise, how the data product is delivered and how it looks matter, too. This chapter will be valuable for members in your organization that develop data products, while offering important insights for those who lead and manage data producers. You may not be an actual data product designer, but you can find guidelines and tips that can enable you to play the part on TV.

ELEMENT 3: THE DATA FLUENT CULTURE



If individual skills are a requirement for the exchange of ideas through data, the culture of the organization is the nutrient-rich soil that helps these individual skills grow. A data fluent culture is built upon organizational leadership, shared understanding and conventions, and everyday habits and processes that make data communication a priority.

People in any organization will look to their leaders to set expectations and model behaviors. First and foremost, data fluent leaders should communicate using data to support their arguments and priorities. By doing so, they set the standard for quality data products and demonstrate their data literacy in public forums.

The language of data has its own dialect from organization to organization. This dialect reflects the organization's industry, history, and circumstances, and results in unique meanings associated with its data. For example, a concept such as *customer satisfaction* can have wide-ranging meaning. The definition will tie to the organization's approach in collecting the data, expectations of what *satisfied* means, required time period for data gathering, and sample size. The goal of a data fluent culture, in part, is to ensure that everyone knows what is meant by a term like *customer satisfaction*. A data fluency culture breaks down when people spend more time debating terminology,

calculations, and validity of data sources rather than discussing what action to take based on the results.

A culture is also defined by acceptable and unacceptable behaviors. A data fluent culture brings well-understood key metrics into meetings; knows exactly how to measure the performance of a new project or product; and includes data fluency skills in the hiring and employee evaluation processes. This same culture does not accept willful data ignorance and rejects impassioned anecdotal evidence as the sole input into decisions. Chapter 6 outlines key strategies for developing a data fluent culture.

ELEMENT 4: THE DATA PRODUCT ECOSYSTEM



The previous three elements in the framework focus on people and behaviors. To enable the flow of information and the creation and sharing of data products, there must be processes, tools, and standards in place. Consider the ecosystem that Apple has created for iPhone and iPad applications. Apple created a system in which high-quality applications can be designed, developed, tested, marketed, and sold. There are mechanisms for user feedback, processes for checking application quality and conformity to Apple's standards, and a well-understood way for consumers to find the apps that suit their needs.

On a much smaller scale, successful data fluent organizations create an ecosystem that enables the creation and sharing of data products. The objectives are similar to Apple’s app store: Ensure quality data products; allow distribution of the right data to the right decisions makers at the right time; and encourage curation of data products so the most valuable ones are discovered and used.

The ecosystem requires three fundamental conditions: standards, tools, and processes. *Standards* are the design patterns and style guidelines that make it easier for data producers to effectively communicate with data. The best *tools* make it efficient to design and build data products and ensure they are discoverable for the target audience. *Processes* encourage sharing of insights, collaboration between producers and audiences, data hygiene, and quality control. A healthy ecosystem can help facilitate data fluency throughout the organization. Chapter 7 reviews these fundamental conditions and provides specific guidance for organizational leaders to create a data fluent ecosystem.

CONNECTIVE TISSUE

The elements of the framework are linked together by guidelines, tools, standards, and training. For example, the *data fluent organization* needs to define and create training resources to help individuals learn basic data comprehension skills. Likewise, the design standards defined in the *ecosystem* should result in a style guide that data product producers use to standardize their output.

RESOURCES FOR MORE DEPTH

The objective for this book is to explain all the elements that must come together to create data fluency for your organization. As a result, this book provides a comprehensive solution for each of the elements in the framework. Fortunately, most of the resources about data communication, data visualization, and data-driven enterprises take a narrower and deeper view. The following books and websites can fill out your knowledge of the elements in the framework.

- Data Literate Consumers
 - Michael Milton, *Head First Data Analysis* (O'Reilly Media, 2009)
 - Alistair Croll, Benjamin Yoskovitz, *Lean Analytics* (O'Reilly Media, 2013)

- Data Product Authors
 - Nathan Yau, *Visualize This* (Wiley, 2011)
 - Stephen Few, *Information Dashboard Design* (Analytics Press, 2013)
 - Alberto Cairo, *The Functional Art* (New Riders, 2012)
 - Nancy Duarte, *Resonate* (Wiley, 2010)
 - Edward Tufte, *Visual Display of Quantitative Information* (Graphics Press, 2001)
 - David McCandless, *Information is Beautiful* (Collins, 2012)
- Data Fluent Culture
 - Thomas H. Davenport, Jeanne G. Harris, Robert Morison, *Analytics at Work* (Harvard Business Review Press, 2010)
 - Avinash Kaushik, Occam's Razor blog, <http://www.kaushik.net/avinash/>
 - Thomas Redman, *Data Driven* (Harvard Business Review Press, 2008)
 - Peter Drucker, *The Essential Drucker* (HarperBusiness, 2008)
- Data Ecosystem
 - Cindi Howson, *Successful Business Intelligence* (McGraw-Hill, 2013)
 - Larissa Moss, *Extreme Scoping* (Technics Publications, 2013)

BENEFITS OF THE DATA FLUENT ORGANIZATION

When all the elements shown in the framework exist and work together, an organization can use data as a common language. Data fluency can have a number of benefits for your organization including:

- **Better decision-making**—A primary objective is to create a more systematic, objective approach to making smarter decisions and taking informed actions in your organization. A clear understanding of data and ability to transform it into meaningful analysis will move your organization away from historical biases, decision-by-anecdote, and pure gut-instinct.

- **Efficient, accurate communication**—Miscommunication and misunderstanding about the meaning of data is one of the hidden barriers to progress in many organizations. When everyone is on the same page as to the meaning of the data, the discussions are about the content and ideas, not deciphering the language.
- **Alignment and accountability**—Clear communication about key metrics gives everyone in an organization an understanding of priorities and how their actions influence those priorities. Furthermore, when things are measured in well-understood ways, it is much easier to hold people accountable to those measures. As management guru Peter Drucker said, “What’s measured improves.”
- **Learning culture**—Discussions around data frequently raise more questions than they answer. In an organization in which data communication works effectively, new, better questions are raised. These questions lead to deeper knowledge, more insights, and hopefully greater productivity.

HOW TO USE THIS FRAMEWORK

This book is for the executive and the business analyst, for the most data-savvy head of operations and the most data-phobic marketer. Creating an organization that can converse using a common data language is the responsibility of front-line staff as well as the organizational leaders. The introduction explained which chapters are the best places to start given your role. Now with the framework details in mind, you can understand some of the immediate benefits that can be achieved through specific chapters in this book.

- **Evaluate strengths and weaknesses**—“Data Fluency Inventory” defines a data fluency inventory for determining where your organization is in its journey to data fluency. This survey instrument points out the areas where you have room for improvement and points you to the parts of the book where there is specific guidance.
- **Define an organizational plan**—The chapters focused on organizational capabilities (the bottom row of the framework) give leaders a roadmap for getting in place the pieces necessary to support your staff. Chapter 6 is for the CEO or department head who wants to define

a better data-driven culture through expectations, guidelines, and behaviors. Chapter 7 speaks more directly to the technology leaders in your organization.

- **Define training and skills needed**—Understand the distinction between the skills required for data consumers (Chapter 4) and those required for data producers (Chapter 5). Both these chapters provide an outline of many of the important skills necessary to create and comprehend data products.
- **Data product inventory**—Throughout this book you read about data products—the reports, presentations, dashboards, and data analyses—that are the medium for communicating data within your organization. A good starting point is to build an inventory of the collection of data products currently in use, their purpose, and the audience they are intended for. Chapter 5 (Data author) can help you understand how these products succeed or fail in their purpose, whereas Chapter 7 (ecosystem) gives you insight into how to manage and share your collection of data products.
- **Set a technology roadmap**—The framework isolates much of technology discussion within the concept of building a data product ecosystem. In Chapter 7, you can gain an understanding of some of the core technology requirements as they relate to providing data access, capabilities to visualize and present data, and having a place where data products can be discovered.

HOW ORGANIZATIONS STRUGGLE

The aspirational framework describes the ideal conditions for a data fluent organization. In working with dozens of organizations in many different industries, the common denominator has been the struggle to establish these conditions. The next chapter describes real-world situations in which weaknesses in one or more elements of the framework undermine the most committed efforts to use data to create a high performance organization. This “learning from the struggles of others” provides an important springboard for growth and development—to which the remainder of the book is dedicated.

SUMMARY

Creating data fluency within your organization is a complex task. It requires individuals who speak the language of data, skilled producers of data products, an organizational culture with the conditions to support data discussions, and the systems, tools, and ecosystem to create and share data products. In this chapter, we outlined the core components of the data fluency framework, our roadmap for improving organizational data fluency. These core components include data literate consumers, skilled data producers, the data fluent culture and the data product ecosystem.

NOTES

1. For more on reading fluency, one might refer to the article by Timothy Rasinski: Timothy Rasinski, "Creating Fluent Readers," *Educational Leadership*, March 2004. Retrieved on July 27, 2014 from <http://www.ascd.org/publications/educational-leadership/mar04/vol61/num06/Creating-Fluent-Readers.aspx>
2. For a better understanding of the nuances of a lingua franca please refer to Wikipedia. Retrieved on July 27, 2014 from http://en.wikipedia.org/wiki/Lingua_franca

CHAPTER 3

How Organizations Struggle with Data Fluency

Few people need to be convinced about the value of using data to inform decisions. A 2012 survey conducted by the technology consulting firm Avanade found that 84 percent of senior leaders think using data has helped them make better business decisions. Furthermore, data use can help drive growth; as 73 percent of respondents claim to have already leveraged data to increase revenue.¹ Increasingly, the opportunities and benefits of data are well understood. What is generally not yet understood is how to actually use data and convert the potential energy stored within this information into kinetic energy that will propel growth and facilitate goal attainment.

When tracking the progress of data use, many organizations value quantity over quality. In 2014, Facebook shared that it was storing 300 petabytes of user data, an amount that had tripled since the previous year.² That's more than a CD's worth of data for each man, woman, and child in the United States. A study by Oracle indicates that the average data warehousing project costs \$1.1 million and takes 10 months to deliver.³ Tremendous volumes of data and

massive investments are common ways that organizations demonstrate that they are on the big data bandwagon. Measures of frantic reporting activity can be equally misguided. While we worked at AOL, the business intelligence group was proud of the hundreds of reports it generated every month and how this number was growing.

Investments in technology and massive quantities of raw data don't guarantee you can have thoughtful conversations using data. Data fluency means having quality discussions, a common use of measures, and shared appreciation for the meaning of your unique data. This chapter is dedicated to the development of these skills.

Your organizational strategy may be defined by a few big decisions by your leadership team. However, execution requires many small decisions.

Strategy communications should always be accompanied by metrics, which help frontline employees take ownership over their roles in the execution. The message should be two-fold: This is what we are trying to achieve and this is how we will measure if we are achieving it.

Amy Gallo, Harvard Business Review⁴

When an organization is data-fluent, individuals use data in everyday activities. They understand key metrics and how they relate to strategy. Useful, usable data products and tools influence how people think about their decisions.

According to IDG Enterprise's Big Data report for 2014, the average organization will spend approximately \$8 million on big data projects this year.⁵ Whether that number sounds like loose change or a king's ransom, you want to ensure your organization's data investments aren't squandered.

This chapter describes some of the pitfalls that organizations face when trying to make good use of their data. The problem is seldom a lack of commitment or investment. As you learned earlier in Chapter 2, "The Data Fluency Framework," data fluency requires a strong foundation in all four quadrants of the framework. Weak data culture, poor individual skills, or a fragile ecosystem can doom organizations to cycles of wasted effort.

By the end of this chapter, you will be able to:

- Identify common ways organizations struggle with data fluency
- Understand how these struggles are aligned with weaknesses in elements of the data fluency framework

PITFALLS ON THE PATH TO DATA FLUENCY

By itself, data isn't enough. A singular focus on investing in technology to manage and visualize your data won't get it done. Leaders committed to a data-driven enterprise are great . . . but still more is needed. Even a team of good people isn't enough. While all of these factors are necessary for data fluency, the pieces alone are not sufficient to transform your organization's use of data.

Below we discuss a few illustrative ways organizations are unable to align all the factors necessary to reach data fluency. Of course, there are many ways to fail. We will describe a few. In doing so, we hope you can draw comparisons to your own organization so that it becomes more clear how an area of weakness in your organization's culture, skills, or capabilities can act as a roadblock that stifles your best efforts at data informed decision-making.

In sharing the cases below, we also hope you will feel a little less alone as you and your organization seek greater levels of data fluency.

REPORT PROLIFERATION

Reports have a way of multiplying like rabbits. Start with a perfectly useful and important report: a monthly sales report with product enhancements and utilization metrics sent to strategic accounts to make them aware of improvements coming and past usage. Customers see the information and want to know more. The report grows. A missing metric is added along with a detailed breakout. The report expands to the point where it gets split into separate reports, each targeted as a separate audience. Then a new executive arrives with her own perspective on the best way to present the same data. New reports are spawned, but the old ones don't go away. Someday, somebody might still find them useful.

We've seen this report proliferation in all kinds of organizations—hospitals, schools, technology companies, insurance companies, and manufacturers. "If we report on everything, surely the right information will exist somewhere in a report." Perhaps they're right, but if no one can find what they need, everyone's left sorting rabbits.

Even sophisticated analytics companies can scarcely control the tendency toward report carpet-bombing. We consulted for a large healthcare company who was fighting against an institutional culture that always pushed for more

data, more reporting, and more metrics. An important part of its services is to deliver annual summary reports to its clients about the costs that affect health benefits. The company's analysts would spend thousands of hours preparing detailed PowerPoint presentations, chock full of hundreds of "key" metrics. When it came time to present to its clients, the discussions quickly dove deep into the details. Important recommendations were lost in a pile of minutia. The first step in recovery is to recognize you have a problem. Our client knew this volume of information wasn't good for the data product authors or the data consumers.

REPORT PROLIFERATION AND THE DATA FLUENCY FRAMEWORK



Report proliferation may mean you have a lot of people interested in data but are having trouble defining what's important. Conversely, you may have people less interested in data but who need reports nonetheless to validate their activities. In either case, useless reports distract attention and prioritization is essential.

BALKANIZED DATA

Organizations which design systems . . . are constrained to produce designs which are copies of the communication structures of these organizations.

Conway's Law

Departments in an organization can easily become *independent silos*, operating with their own set of norms, conventions, and terminology. This impacts what you can do with your data and what you can understand. Almost 50 years ago, computer programmer Melvin Conway observed a similar issue with software

development and coined Conway's Law. You've experienced this problem if you've ever been on a customer service call where you give all your personal information at the start of the call and then have to give it all again every time you're transferred.

Although Conway's Law was partly meant in jest, it's an accurate sociological observation that has been confirmed in studies. Communication is hard and people with different motivations and backgrounds will create a fractured design.

A similar challenge applies to data. Each organizational department may use different data systems and terminology, processes, and conventions in data conversations and products.

We recently experienced this in a large urban school district. The Office of Assessment and Accountability (OAA) is a data-lover's paradise, with the latest technologies, a team of skilled data scientists, and an effective data warehouse beloved by principals and teachers alike.

BALKANIZED DATA AND THE DATA FLUENCY FRAMEWORK



If you don't have a shared data culture and ecosystem, your data will be siloed. This limits what you can learn and the questions you can ask. The solution isn't purely technical, it's organizational. Chapter 6, "The Data Fluent Culture," and Chapter 7, "The Data Product Ecosystem," discuss how organizations can break down these silos to create a more data fluent organization.

However, policy changes were encouraging school leaders to think about teacher quality and performance alongside the more traditional measures of student achievement. With this in mind, members of the OAA began to

coordinate with the human resources (HR) department to connect information on teachers with their students. Unfortunately, HR had no systematic way of generating electronic records on teachers. In fact, most of the records were kept as hard copies in file drawers. When HR began entering the information by hand into spreadsheets, it found that teacher identification numbers changed annually and didn't match state identification numbers. Years of student data could not be connected back to teachers charged with these students.

DATA ELITISM

Working with data can require a lot of technical skill. And data can tell stories and reveal truths that an organization may not want to share broadly. Why not centralize your efforts and limit access to data to the highly trained few who can be trusted to bring order to chaos? This is the viewpoint of Tom Davenport in his book, *Competing on Analytics*. In his view, the best analytical organization is one that has centralized control:

Stage 5 organizations develop a robust information management environment that provides an enterprise wide set of systems, applications, and governance processes. They begin by eliminating legacy systems and old spaghetti code and press forward to eliminate silos of information like data marts and spreadsheet marts. They hunt for pockets of standalone analytic applications and either migrate them to centralized analytic applications or shut them down.

(Harvard Business Review Press, 2007)

Like an over-eager police force hunting down deviants, this IT-led vision of business intelligence focuses on control, consistency, and data management. An extreme approach, however, comes at the expense of the individuals who use the data. Distancing analysis from the people who must use it results in data producers and their products that are disconnected from the decision-making process. Data products aren't trusted and they often aren't useful. All the problems of a command and control economy emerge.

We've seen this happen in credit card organizations where data scientists built models to predict customer behaviors that were not in alignment with the regulatory environment. The knowledge of what decisions could be made given the regulatory environment was in the hands of the data end users. Getting the data scientists to understand the real business environment took extra time and effort and the initial set of models were useless.

DATA ELITISM AND THE DATA FLUENCY FRAMEWORK



Disconnecting analysis from the people that use it decreases employee engagement. The data elite should be building frameworks that empower others to do for themselves.

THE SUPERMODEL

It's easy to mistakenly believe that dashboards must look slick and flashy with eye-catching visualizations that sizzle with graphical luster. Unfortunately, dashboards of this type often say little, and what they do manage to communicate can be a distraction.

Banks are notorious for creating supermodel dashboards that conceal as much as they present. One large bank displays the retirement accounts of its clients with a snazzy dashboard. The landing page in the website gives you many options including the ability to compare yourself to others. Enter a few data points (age, marital status, savings, and salary), and you'll soon be comparing your retirement nest egg to that of "people like you." The presentation is capped by a colorful animated pie chart showing your funds by asset class (large cap growth, global/international, and small/mid/speciality).

If you want to substantively analyze your retirement account year-over-year, you'll have to break out your calculator and gather old PDF statements. If you want to see how much your fund grew due to contributions versus market performance, it's your responsibility. Good luck finding the impact of fees on your investment returns over time.

These dashboards distract customers from what's actually important and steer them toward options that work for the bank. They might look fancy but as soon as you start asking questions you hit a wall. Dashboards must display the right information with a clear message in a limited space.

THE SUPERMODEL AND THE DATA FLUENCY FRAMEWORK



It's easy to lose sight of one's goal with data products. Authoring tools should support consumer needs and data consumers need to be literate enough to demand the metrics that matter.

SEARCHING FOR UNDERSTANDING

Your visions will become clear only when you can look into your own heart. Who looks outside, dreams; who looks inside, awakes.

C.G. Jung

An organization's capability to make fluent decisions from data depends on how well the organization knows itself. Self-awareness helps you answer the difficult questions: What does success look like? Are we moving in the right direction? Who should we compare ourselves to?

For a new organization—especially one in an emerging market—it takes time to figure out what matters most. These organizations often lack focus in their data analysis, measurement, and communication while on the path of discovery. Even with the best intentions, organizations can struggle to make good use of their data as they search for the information and metrics that will align with their emerging strategy.

We experienced this with a client that was innovating in the health benefits market. This company had a new approach for helping families with healthcare issues using high-touch customer service. By giving every employee access to a *health concierge*, the company encouraged preventative care and supported employees with serious, long-term conditions like diabetes. Helping employees navigate the healthcare system led to lower benefits' costs for

employers. Despite its success, the company wasn't sure how to explain the connection between its model and the reduction in costs.

The company leadership wanted to use data to communicate with its customers as well as to measure internal performance. Without clear links between its actions and health outcomes, it struggled to make its case even with a growing volume of data.

Presentations to clients varied from telling anecdotes to deep-dive analyses, but the company never quite demonstrated how its services led to financial impact. Internally, the organization had not yet arrived at a common data language. Different areas used different metrics to track performance, often focusing on things that were easy to measure, like volume of calls or customer satisfaction surveys. A focus was on activities rather than outcomes that affected costs.

It was obvious to us that this company was on a journey. Without a well-understood market, it needed to learn to know itself, which then could be expressed through its reporting, dashboards, and communications with clients.

SEARCHING FOR UNDERSTANDING AND THE DATA FLUENCY FRAMEWORK



For organizations in the early learning stages, weaknesses in their data fluency foundation have little to do with the skills of individuals or a system to create effective data products. The newness of the business model itself leads to an immature culture around data. There isn't shared understanding of what data means and what matters most. Even with strong leadership committed to use of data, people need to build a common awareness about how the organization works and how the system can be measured.

DATA CARE

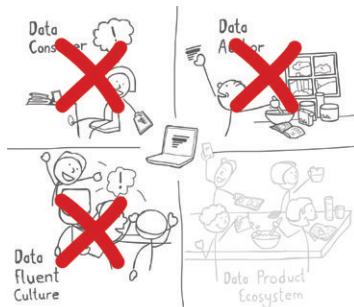
Sometimes data quality is a proxy for organizational discipline and morale. Do people throughout an organization really care or are they just collecting a paycheck? Team members who actually care about an organization and the organization's mission will go the extra mile to make sure data is correctly populated, erroneous data is fixed, and extra data needed for future reference is added. We have found issues of data care at large publicly traded companies, academic institutions, and nonprofits.

Working with the fundraising team of a large, yet struggling nonprofit organization, we found data care and organizational morale low. The fundraising team used *Raiser's Edge*, a popular fundraising software platform to capture the all-important donor information.

Gift officers were given reports with the names of individuals in their region that contributed more than \$5,000 in any of the last 3 calendar years. Armed with a regional list they were asked to make donor calls, visit donors personally, and raise a certain dollar amount (often five to seven times their salary). None of the gift officers reached their goals. Why? Poor data care.

When trained on *Raiser's Edge*, gift officers were asked to keep track of e-mails and visits with prospective and current donors. Although this information would be reported at a biweekly teleconference with the whole fundraising team, little effort was made to populate the information in the company's data product. As a result, each gift officer kept track of his donor knowledge independently, and institutional knowledge was not built. When gift officers left the organization, replacements started with little knowledge about donors in their region. Although you knew past gift history, you didn't know why they gave, what they liked, what was conveyed with the giving of the gift, or if they had a preferred contact number. Knowledge that would be needed for future reference was never recorded and, therefore, wasn't generated for the organization.

DATA CARE AND THE DATA FLUENCY FRAMEWORK



Fundraising is an illustrative case of the importance of data care. Failing to create a strong data product, failing to use data products available, and an organizational culture that doesn't hold employees and vendors accountable for data care leads to an inefficient team.

METRIC FIXATION

Leaders and organizations can become mesmerized by one metric at the expense of others. Examples range from the management of a customer service call center, the derailment of a university present, to an increase in hospital-acquired infections. Laser focus can be helpful in achieving a goal; however, leaders need to ensure that important things that are temporarily out-of-focus are also being monitored.

For example, customer service call centers can often become fixated on metrics to the detriment of both employees at these call centers as well as customers. Some common metrics at customer service call centers include average call wait time, average call length, whether or not the issue was resolved on the first call (and doesn't call back within 30 days), and retention of customers (keeping customers who are trying to cancel service, even if it means changing their subscription package to a lesser service). Incentives are often established for call center employees, typically low wage earners, against metrics like these. Let's take a closer look at one metric: customer retention. If customer service employees are fixated on customer retention and their livelihood depends on it, a whole host of tactics will be used to keep folks subscribed to their current service (phone, cable, internet, gym, magazine, etc.). Transferring calls to retention specialists, call dropping, and offering a different subscription service at a lower monthly cost are a few.

Adrianne Jefferies reports about the pressure on Comcast customer service employees with respect to retaining customers and how their monthly income could be greatly affected by the loss of services. Jefferies reports that “metrics-obsessed reps are . . . trying to reach a predetermined outcome in the call, and they’re trying to do it in under 11 minutes. Comcast has turned its customer service reps into sales reps.” Even when customers explicitly tried to cancel their service, they found difficulty.

In another setting, a president of a private liberal arts college was hyper-focused on increasing enrollment and lost sight of other important metrics. Increasing enrollment is important for a variety of reasons and had worked for him at a previous institution. In his mind, increasing enrollment meant increasing revenue, which would lead to new programs and pay for new buildings.

Everybody at the college was soon busily increasing the number of students applying, admitted, and enrolling. Tuition was significantly discounted in year one to attract students with competitive financial aid and merit scholarships. This hyper-focus led to more students in the short-run. Meanwhile, little attention was paid to retention.

Without the capacity to properly handle the new students, students became frustrated by the inability to get into the classes they wanted, overcrowded first year dormitories, and reduced financial aid in subsequent years. Consequently, the 6-year graduation rate dropped. Revenue gains by increasing the freshman class were soon lost by students transferring or dropping out. The board lost confidence in the president and a leadership change was made.

Hospital leadership can also become hyper-focused on a single metric to the detriment of other metrics. One hospital, for example, had an internal issue with hospital-acquired sepsis, a life-threatening infection. Concerned, hospital educators created competencies around hospital-acquired sepsis for all nurses. Continuing education focused on sepsis and the chief nursing officers held lunch-and-learn sessions. Awareness was raised and patient outcomes improved. However, although sepsis decreased, urinary tract infections increased. Attention paid to one infection type meant attention diverted from another. One chief nursing officer lamented, “You can have your team focus only on so many things.”

As a result of metric fixation, thought leaders have tried to introduce broad concepts that measure and score all underlying metrics. In healthcare, for example, *patient experience* is a new metric that healthcare providers are beginning to measure. Hospital Consumer Assessment of Healthcare Providers and

Systems (HCAHPS), a survey administered to a sample of patients leaving any hospital setting and reported to the government, attempts to measure *patient experience*. The verdict is still out as to whether this will result in better health outcomes. It is an attempt at moving organizations beyond metric fixation.

METRIC FIXATION AND THE DATA FLUENCY FRAMEWORK



When a leader and culture become hyper-focused on one metric and loses sight of the big picture, balance suffers. Data producers begin creating hyper-focused data products—reporting just on a single metric often from different vantage points. The conversations of data consumers begin reflecting the reports given to them. Just because everybody is marching to the same drummer doesn't mean you're headed in the right direction.

FINDING BALANCE

This chapter explored a few of the ways organizations struggle to make the most of their data. Whether it is an issue with too many reports or too few, a lack of commitment to using data or a fixation on a particular metric, many organizations have found it challenging to build data fluency.

The path toward data fluency requires balance. In particular, organizations should look for the following:

- Balance between the needs of data users and the interests of the data producers
- Balance between letting the data speak and allowing the data author to define the message

- Balance between the business users who make decisions using data and the technologists who manage and deliver the data
- Balance between focusing on a few key metrics and the flexibility to change measures of success as needs evolve
- Balance between the deep analysis of data scientists and the direct reporting that provides the most important data in a simple format
- Balance between data presentation that puts understanding first and data products that use aesthetics to capture an audience's imagination

A data fluency framework can help to provide guidance about how to make the right choices. The following chapters will explain the four quadrants of the framework and the organizational capabilities needed to avoid the struggles just described.

NOTES

1. Avande, "Global Survey: Is Big Data Producing Big Returns?" June 2012. Retrieved July 31, 2014 from: http://www.avanade.com/Documents/Research_and_Insights/avanade-big-data-executive-summary-2012.pdf
2. Pamela Vagata and Kevin Wilfong, "Scaling the Facebook data warehouse to 300 PB," 2014. Retrieved July 31, 2014 from: <https://code.facebook.com/posts/229861827208629/scaling-the-facebook-data-warehouse-to-300-pb/>
3. Oracle, "An Overview of Oracle Business Intelligence Applications," June 2009. Retrieved July 31, 2014 from: <http://www.oracle.com/us/solutions/business-intelligence/064103.pdf>
4. Amy Gallo, "Making Your Strategy Work on the Frontlines," Harvard Business Review Blog Network, June 24, 2010. Retrieved July 31, 2014 from: <http://blogs.hbr.org/2010/06/making-your-strategy-work-on-t/>
5. Alex Woodie, "Big Data Spending to Hit \$8M Per Organization, IDG Says," Datanami, January 16, 2014. Retrieved July 31, 2014 from: [http://www.datanami.com/datanami/2014-01-16/big_data_spending_to_hit_\\$8m_per_organization_idg_says.html](http://www.datanami.com/datanami/2014-01-16/big_data_spending_to_hit_$8m_per_organization_idg_says.html)

CHAPTER 4

A Consumer's Guide to Understanding Data

One muggy midsummer night in the 1960s, Andy Warhol pointed a camera at the top of the Empire State Building to record an experimental film he called *Empire*. For more than 6 hours, as dusk fell on New York's most famous building, the camera's eye captured the long transition from day to night and the mundane details of lights flickering on and off its exterior (Figure 4-1).

This film is the essence of data; its unique capability to be more patient and more omniscient than you or I can be in our observations of the world. Data captures actions and characteristics of the real world and transforms them into something that can be examined and explored after the fact. And like Warhol's film, when recording data, we must always consider perspective—where the lens is pointed and what lies outside the frame.

Today, unblinking lenses record where and when we shop, what we buy, when we travel, and what media we consume. Thousands of security cameras quietly film their own version of *Empire* every night. Our opinions are captured explicitly through surveys, and our interests are inferred through data that captures our behavior.



FIGURE 4-1 View of the Empire State Building at Night

We are capturing more, storing more, and interacting with more data than ever before. As data consumers, this unending flow of data promises to help answer questions of all sorts. It can illuminate the patterns of disease in a population, show the areas where individual students are struggling, and reveal the elements of an advertising campaign that grabbed the most attention.

As the ability to capture and store data has increased, you are presented with a new set of challenges—how do you interpret and act on the many different forms of data that are readily at your disposal? How do you make sense of it? How do you decide what you should focus on and what you should ignore? What skills do you need to easily consume data products and use them to improve your decision-making? Your data literacy skills (Figure 4-2) are the first frame in the data fluency framework.

Creating a data fluent organization doesn't just happen. It starts with people who love using data as a tool to improve their job performance—people who have learned to converse with others in the language of data. It needs people who expect and demand better, more useful data products from themselves and others. It starts with you.

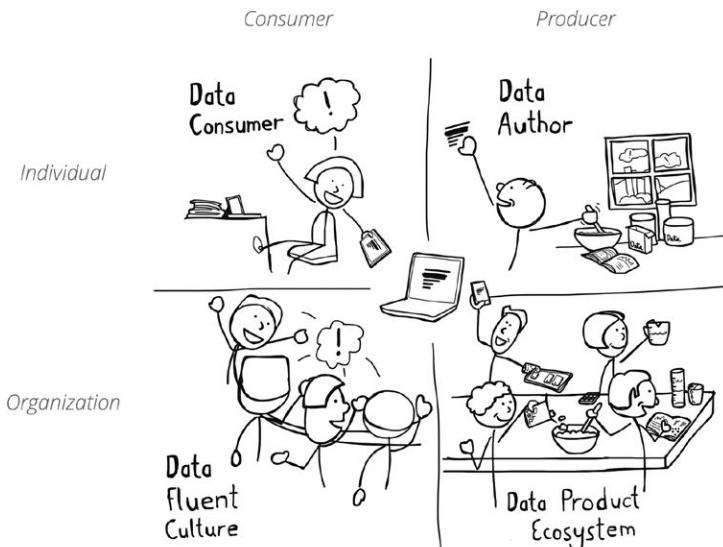


FIGURE 4-2 Building data literacy skills for yourself and within an organization is one of the four pillars of building a data fluent organization.

By the end of this chapter you should be able to:

- Become conscious of everyday data products.
- Develop strategies to overcome common barriers to using data products.
- Increase your data fluency.
- Dissect and find meaning in data products.

DATA PRODUCTS

In *Empire*, Andy Warhol captured data from a constant data source and created a unique data product. Twenty-four times per second, the camera captured data, an image of the New York skyline. The data source was the skyline from 8 p.m. to 2 a.m. The data product is the film.

Broadly defined, *data* means events that are captured and made available for analysis. A *data source* is a consistent record of these events. And a *data product* translates this record of events into something that can easily be understood.

Data products have changed rapidly with technology. Today, data products may be automated, like a website that shows you reports in real time, or they may be built by hand, like a PowerPoint presentation. They can be raw data extracts—perhaps a CSV file containing sales details, or they can be highly processed, like a complex, interactive data visualization.

Consider a large healthcare company, a client of ours. This client captures data on hospitals from around the country and often uses Center for Medicare and Medicaid Services (CMS) data to populate important hospital data fields including addresses, total Medicare expenditures per hospital, and the total number of Medicare patients served. CMS, a data source, provides a consistent record of these publicly available events. After the CMS data is uploaded into the healthcare company's Salesforce.com database, hospital names are matched to unique CMS identification numbers so that useful reports can be generated. The sales team prioritizes its prospects based on Medicare spending—and works from a Salesforce.com report, a data product, which lists hospitals accordingly. The following table gives examples of the data source, data, and data product for the preceding film and healthcare cases.

DATA SOURCE	DATA	DATA PRODUCT
Skyline	Photos	<i>Empire</i> , the film
CMS CSV file	Hospital address, Medicare expenditure, number of patients	Prioritized Salesforce.com Report (Hospitals by Medicare expenditure)

Data products can be organized and characterized by a series of continuums that describe the nature of the data and how it is presented (Figure 4-3). Data can range from being raw to processed, granular to summarized, and textual to quantitative. Data presentations might be static or dynamic, small or massive, or anywhere in between (Figure 4-3).



FIGURE 4-3 Characteristics of data products

Countless data products, varying on the different continuums presented in Figure 4-3, can be found in everyday situations.

EVERYDAY DATA PRODUCTS

We use data products every day to manage and understand our world, often without being conscious of their impact on our daily routines. For example, when you were getting ready to leave home this morning, you may have checked the weather to see what to wear (Figure 4-4).



FIGURE 4-4 Online weather forecast

Data reminds us to wear a jacket.

Later, on your way into work, you worried about traffic (Figure 4-5).

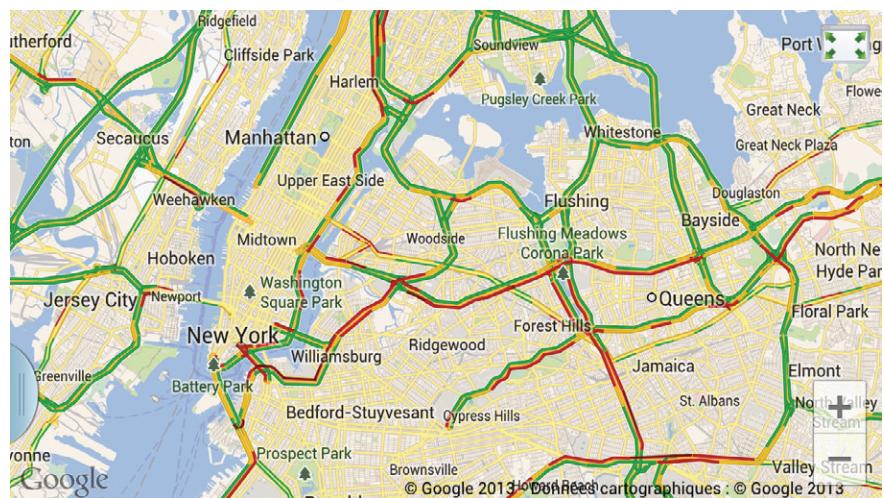


FIGURE 4-5 Online driving directions with real-time traffic updates

Data keeps us on time.

You stop to grab breakfast and read a menu that visually conveys the price list (Figure 4-6).

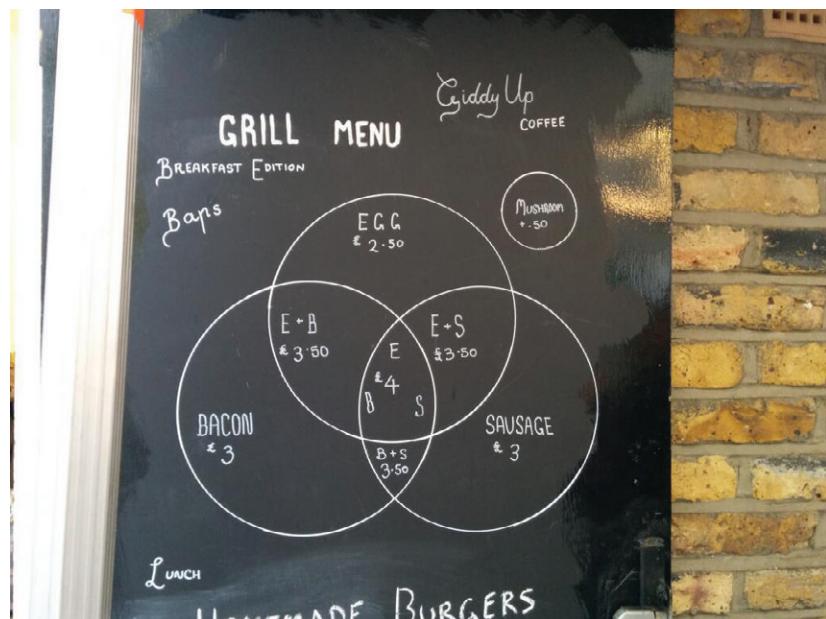


FIGURE 4-6 Breakfast menu

At work, you no doubt interacted with a number of data products, some more obvious than others (Figure 4-7).



FIGURE 4-7 Laptop battery strength indicator

This laptop's battery says "don't worry, you've got time."

Finally, after a long day, you sit down for a nice dinner and a bottle of wine (Figure 4-8).

A wine bottle label helps you find the right food.

These examples show a few ways in which you can interact with data products on a regular basis. Everyday objects, signs, and communication often come with subtle and simple data products attached. Like those represented here, these objects can be expertly designed to convey a clear message for the viewer.

As consumers, you are exposed to sophisticated data products every day. If designed well, like the work by professional information design teams at publications like the New York Times, The Guardian, or ProPublica, they are powerful tools for engaging curiosity, interactive storytelling, and understanding our world.



FIGURE 4-8 Descriptive wine label

One good example comes from Nathan Yau of FlowingData.com who published this beautiful graphic based on data security breaches. For a casual reader, it consolidates information about this increasingly common problem, and makes it easy to compare the breadth of impact and type of breach. The color helps us see that data lost through hacking (green) far outweighs breaches through stolen, lost, or fraudulent activities (Figure 4-9).

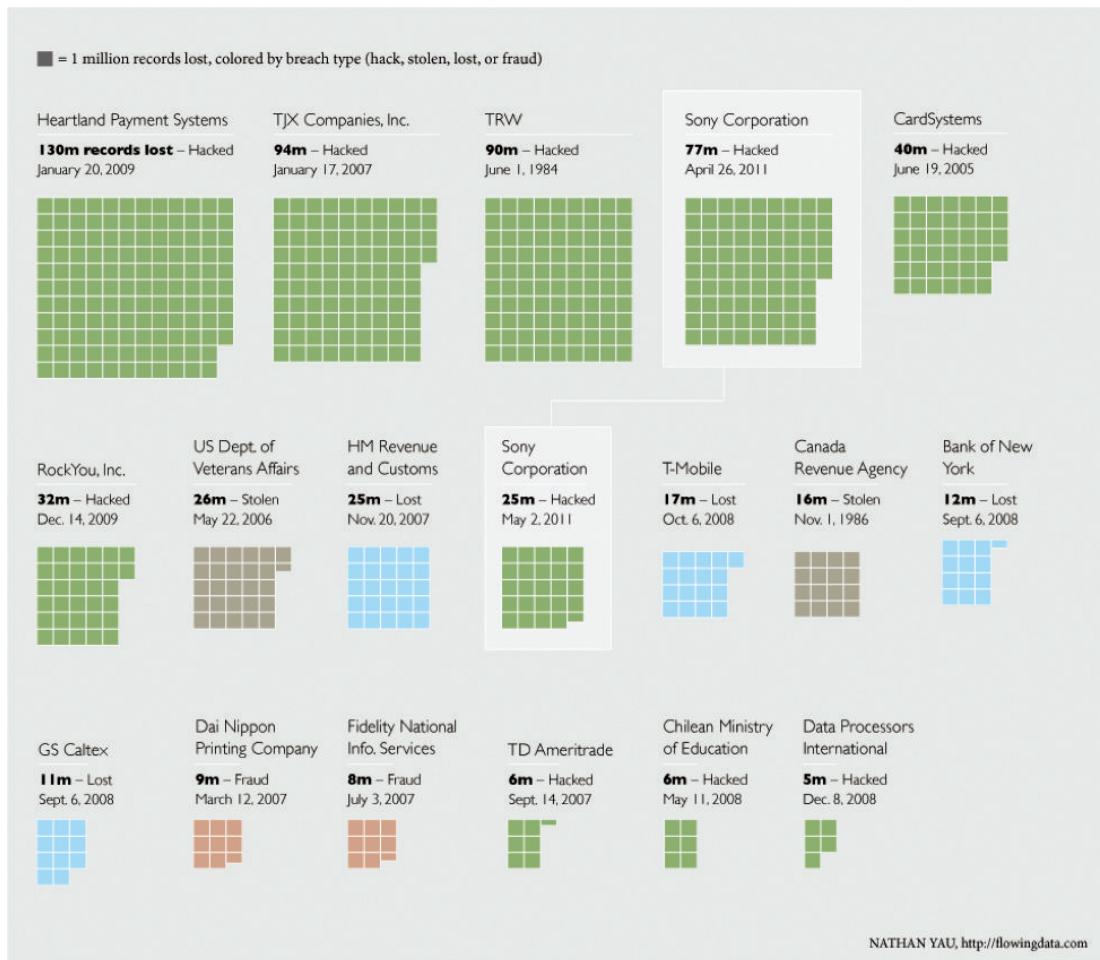


FIGURE 4-9 The largest data loss incidents on record

Source: <http://flowingdata.com/2011/06/13/largest-data-breaches-of-all-time/>

Sadly, this graphic is the exception. Most data products are created by people without the experience demonstrated here. (Hopefully Chapter 5 rectifies this problem.) More often, data products have arbitrary visualization choices, an unclear message, and data presentation without thoughtful design or concern about the users' priorities. They leave the burden of understanding to the audience. The remainder of this chapter is intended to give you guidance and tools to find your way through the data as it is shared in many organizations. You explore some barriers to using data products and then experience some practical challenges that help you increase your data fluency.

BARRIERS TO USING DATA PRODUCTS

Learning the language of a new industry can be difficult. Likewise, you often encounter barriers to engaging with data products. These barriers include jargon, not knowing where to start and what to focus on, and inconsistency.

Jargon

Unleveling. Fitnessgram testing. Scaffolding text-to-text connections. Dahlia Lithwick, who was keenly interested in learning what her second grader was doing in school so that she could provide support at home, described going to a parent-teacher conference at her child's public school only to encounter all these terms. The teachers didn't take time to explain the terms but assumed that all the parents knew what they meant. When she looked around, all the other parents were smiling and nodding in agreement, and she certainly didn't want to appear to be uneducated or uninvested in her child's success.

Jargon can make a speaker sound smart and listeners feel dumb. As a consumer of data, it is important to remember that most jargon represents simple and often familiar concepts.

When learning a new language, it can be intimidating to encounter a native speaker. *¿Dónde está la biblioteca?* may uncork a fountain of Spanish that you cannot begin to comprehend. It is the same in learning to speak data. The following table presents a common jargon list used by experienced data analysts, with an explanation of the relatively simple and familiar concepts they represent.

TERMINOLOGY	WHAT IT MEANS
Correlation and causation	<p>Even though there may be a relationship between two things (correlation), there isn't evidence that one thing directly causes the other thing to happen (causation).</p> <p>Suntan lotion sales and ice cream sales both go up at the same time. It doesn't mean that suntan lotion causes you to crave ice cream. They're both related to temperature, the common factor driving sales of both.</p>
Statistical significance	<p>You want to know if you can be certain that the results are not due to chance—that they really are a result of an intervention, action, or explanation. Statistical significance means that a finding or value is reliable. That is, the result isn't likely explained by random chance.</p>
Over-indexing, under-indexing	<p>It is often helpful to view values through the lens of an index, where the index may be an average, benchmark, or standard value. In some cases, the index is set to 100, so it is easy to compare an over-indexing value (greater than 100) or under-indexing values (less than 100).</p> <p>For example, a Pew survey showed that 26 percent of African-American Internet users surveyed said they used Twitter, compared to 14 percent of Caucasian users and 19 percent of Hispanics.¹ In this example, African-American users over-index on Twitter.</p>
Regression model	<p>A statistical approach to determine the relationship between variables. Typically many independent variables are included in the regression model in an effort to explain movements in a dependent variable. For example, many descriptive factors about customers may be used in a model to try to predict the likelihood of a customer leaving.</p>
Generalizable	<p>A determination of whether the results from the study/survey may be applied to a broader population than was included in the group who actually participated in the study/survey.</p>
Regression to the mean	<p>Are the results influenced by the tendency for things to move back toward the average? For example, initial results may be impacted by a unique performance instead of a person's average performance over time, and what we are now seeing is a person's typical performance.</p>
Confounding factors	<p>A situation in which a variable that helps explain a relationship has been left out of a statistical model. For example, a perceived relationship between two factors (suntan location sales and ice cream sales) may be misunderstood for lack of including a confounding factor (temperature).</p>

NOT KNOWING WHERE TO START OR WHAT TO FOCUS ON

Books, movies, and music all have a beginning. Data products often do not. When you see a dashboard littered with a dozen competing graphics, it's easy to disengage. Tables of raw data can be even worse. Data products are often designed to deliver everything and the kitchen sink.

Good data products start with a high-level summary and then let users progressively and logically drill into more complex details. They are also simple and uncluttered. They use white space and have a clear visual hierarchy (Figure 4-10).

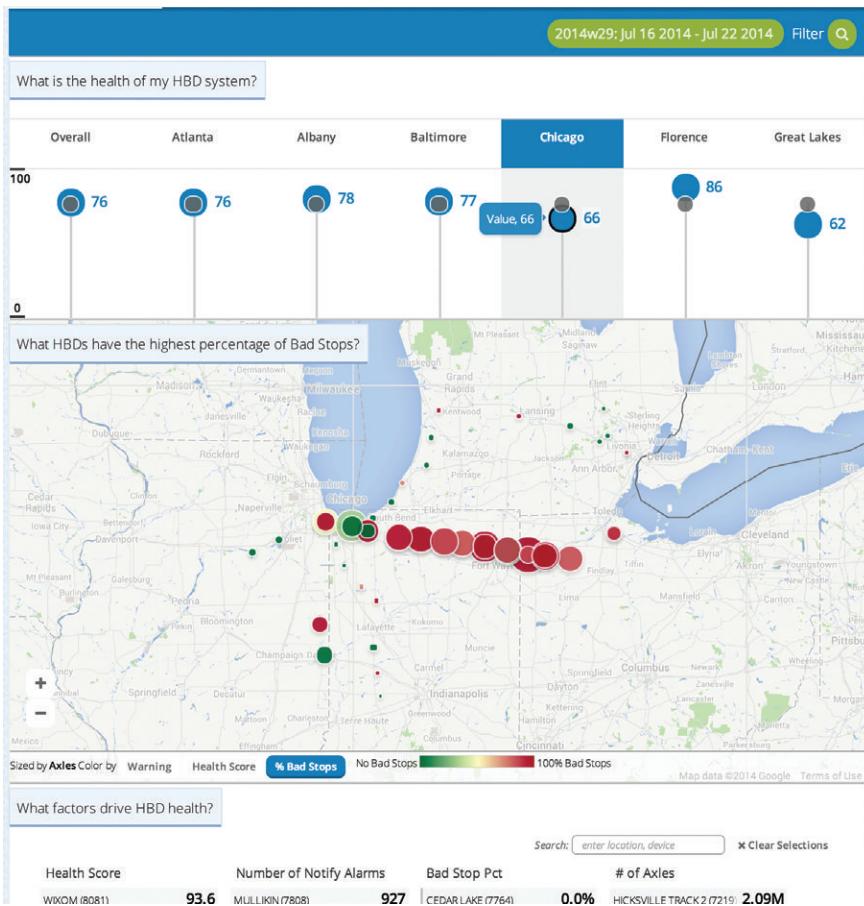


FIGURE 4-10 Railroad analytics dashboard

When data products lack a clear starting point, you as the reader must find your own path through the information. Later in this chapter, you will read some pointers on how to extract meaning from charts and tables, and ideas of where to start your investigation.

INCONSISTENCY

Often you'll get conflicting stories as you look at data products. Try not to get caught up in trying to stamp out these inconsistencies. Different data products were developed for different needs and have different histories.

Returning to the healthcare client that uses CMS data (referenced in the section titled "Data Products" in this chapter), you might consider the prioritized list provided to the sales team. The top hospitals by government Medicare expenditure isn't a bad list to use when trying to determine which institutions are the biggest and deserve the most attention for future sales calls. However, when a sales person compared this government list (via the SalesForce data product) to a popular website listing the top 50 healthcare institutions—there was inconsistency. Why? Because large hospital systems may have a different CMS identification number for each of their facilities. The SalesForce data product had facility level data. The popular website found by the sales person included health systems that were not broken down into their respective facilities. Regardless, either list would have been a good place to start from with respect to client prospecting! If you spend the majority of your time stamping out such inconsistencies, you might have perfectly synchronized lists; however, you'll never get around to making any sales call. Data needs to be good enough to help you make a better decision about something you face right now. Tomorrow there will be new (and hopefully better) sources of data.

Data products aren't synchronized swimming (Figure 4-11)—perfect harmony isn't required.

Water polo is a better analogy (Figure 4-12)—move the ball forward and stay in your position.



FIGURE 4-11 Synchronized swimming



FIGURE 4.12 Water polo

LEARNING THE LANGUAGE OF DATA

Understanding statistical concepts like those described above is an important skill. Clive Thompson emphasizes this point in a *Wired* magazine article entitled *Why We Should Learn the Language of Data*.² He frets that a collective inability to understand statistics leads to misguided conclusions. “Gamblers think their number is more likely to come up this time because it didn’t come up last time. Political polls are touted by the media even when their samples are laughably skewed.” He concludes: “We often say, rightly, that literacy is crucial to public life: If you can’t write, you can’t think. The same is now true in math. Statistics is the new grammar.”

As you engage with data products, you’ll find it equally important to understand the raw materials of the information you are looking at. By knowing the language of data and how to quickly read it, you’ll have a step up in interpreting what you see and where to find insights.

This section offers practical guidance on how to engage with data products and extract meaning. It tackles the problem in four parts:

- Learn the components of data to understand how data products are constructed.
- Explore data tables to start to get a sense of the shape and insights in the data.
- Look for the unexpected in data products.
- Find a starting point within a complex data product.

Before you begin to explore data tables and data products, look at the underlying structure of data. Understanding the building blocks of data can help you be more efficient in reading tables, charts, and other data products.

ATOMIC DATA

Each row in raw data represents an *element*, a person, place, or object doing something. The data that is captured includes *dimensions*, which describe characteristics of the elements involved, and *metrics* answer the question “how much?” or “how many?” Metrics can be added up, averaged, or otherwise manipulated with arithmetic or statistics. Now consider this nonsense sentence.

The quick brown fox jumped over the lazy dog.

If you were representing this as data, maybe for an important study on the jumping ability of animals, your observation of Mr. Fox would look like this:

Jump date: July 1, 2013, 10:13 a.m.

Jumper species: Vulpes vulpes

Jumper color: #826548

Jumper height at shoulder: 0.42 m

Jump height: 0.78 m

Number of jumps: 1

Target species: Canis lupus familiaris

Target color: #454244

Target height at shoulder: 0.71 m

For both the *jumper* and the *jumpee*, we've captured the *dimensions* of species and color. These are characteristics of the fox and dog.

We've also described with numerical *metrics* other characteristics of the actors as well as the action that was taken (one jump of 0.78 meters). We could create calculated metrics. How high can the fox jump relative to his height? In this case it's 0.78m / 0.42m. That's almost twice his height—bravo, Mr. Fox.

SUMMARIZED DATA

When you summarize data, you group data with the same dimensions together and do something sensible with the metrics. Now imagine you've gone to a zoo with the intent to observe animal bounciness. You've observed for days and captured the leap of the tiger and the spring of the gazelle. You want to summarize this data, so you group the animals by their dimensions and sum up the metrics.

Some metrics are additive, like the total number of jumps. For others it makes more sense to calculate averages or rates. It's more common to see data summarized than raw.

These three concepts (elements, dimensions, and metrics) begin to explain how data tables and charts are constructed and can be analyzed. A straightforward way to understand most data is to:

1. Filter elements by dimensions to create a scope that is relevant. For example, limit results for data collected in the last 30 days.
2. Summarize the metrics by adding or averaging all the values of the elements. For example, average jump height.
3. Break out the metrics by dimensions to see the characteristics where that value is unexpectedly high or low. For example, average jump height broken out by species.

Most analyses are that simple: Filter by relevant dimensions; summarize metrics; and breakout by dimensions. Web analytics tools such as Google Analytics may seem complex (Figure 4-13), but they are basically giving you ways to choose metrics and break them out by dimensions.

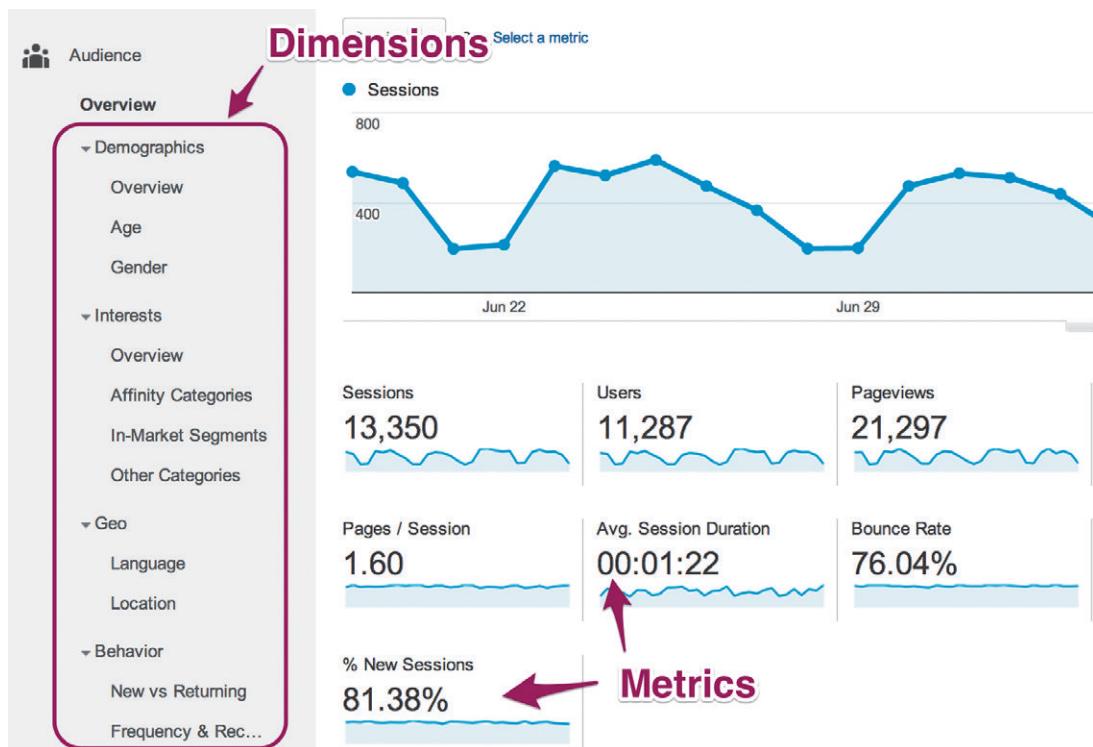


FIGURE 4-13 Dashboard from Google Analytics

EXPLORING DATA TABLES

In some cases, you may encounter data products that are raw or summarized data presented in tables. Even if you are not an experienced analyst, there is much you can do to gain an appreciation for the structure and insights to be found in the data. Rather than trying to absorb the full data table, start with understanding rows, and then move on to summarizing columns, as shown in Figure 4-14.

Rows Tell Stories

Each row in raw data represents an element—an individual, atomic thing from the real world. It may be an event, like a customer transaction at a specific location and time, or an advertisement of a specific size, position, and location (refer to Figure 4-14). A row could represent a person, place, or object.

<i>Student Name</i>	Facility	Department	Overall Score
Amelia Hughes	St. Marys	ICU	83.0
Annette Willis	Centennial	ICU	97.0
Arthur Zimmerman	Centennial	OR	77.0
Beverly Armstrong	Centennial	ER	84.0
Brandi Ballard	Centennial	Support Services	92.0
Cassandra Klein	Centennial	Support Services	82.0
Clark Larson	St. Marys	ER	78.0
Cory Boyd	St. Marys	OR	47.0

FIGURE 4-14 Understanding how to use rows and columns

To begin to absorb the data, read across the table to understand what this element is and how the columns describe it. A picture emerges of what the data is about. Take your one row and compare it to others with the same

dimension. Perhaps it is a student, who is performing well in some classes but struggling in others, or a customer service call that took too long and resulted in some unhappy comments. You'll find that as you read each row, a small character study emerges.

Columns Give the Bigger Picture

If the rows are the trees, the columns help you understand the forest. Start by removing columns that aren't helpful. Data tables straight from a data source are often littered with fields that are empty, redundant, or hold no useful information. Find those columns that you can relate to your experience in the real world. Separate the interesting columns into two buckets: dimensions (the who, what, where, and when of the elements) and metrics (how much).

To understand dimensions, ask these questions:

- How many unique values are there?
- What are the most common values?
- What proportion of all the elements are represented by the top few dimensional values?
- Analyzing the metrics in your data will focus more on summarizing values:
 - What is the average value?
 - What are the minimum and maximum values?
 - What causes the outliers?
 - What is the distribution of values?

Answering these types of questions can build a familiarity with the data and reveal insights about the real world behind that data.

UNDERSTANDING CHARTS AND VISUALIZATIONS

We have an inbuilt ability to manipulate visual metaphors in ways we cannot do with the things and concepts they stand for—to use them as malleable, conceptual Tetris blocks or modeling clay that we can more easily squeeze, stack, and reorder. And then—whammo!—a pattern emerges, and we've arrived someplace we would never have gotten by any other means.

David Byrne, from the introduction to The Best American Infographics3

Charts and visualizations let you use your eyes to find patterns. If they're designed well, they represent characteristics and quantities in your data in a consistent, clear way.

When we say *consistent*, we mean that if we're comparing values and one value is twice as big, then the bar, pie piece, or line that represents it should be twice as large onscreen. When you look at the consistent chart, you're getting an accurate picture of how numbers compare.

To be *clear*, charts should display data without distracting decoration. Unneeded details are what Edward Tufte—the godfather of modern data visualization—calls *chart junk*. They include shadings that reduce contrast, 3D views that make it hard to see where a bar starts and ends, and gradients and colors that are purely decorative. Clarity also requires using the right type of chart. The following table gives the benefits and limitations for common chart types.

CHART TYPE	GREAT FOR SHOWING ...	LIMITATIONS
Pie	How a whole breaks up into constituent parts	Becomes hard to read and label when there are more than a few slices in the pie
Line	Changes in values over time	Too many data series makes it hard to distinguish trends; lines often overlap and collide.
Column/bar (single series)	Comparing of values for a list of items	A bar chart is typically superior to columns because it affords more room for readable labels.
Histogram	Distribution of values across a continuous variable	Different “bucket” sizes can impact what results are visible.
Scatterplot/Bubble	Finding outliers across multiple dimensions; seeing patterns as items cluster	Difficult to label many items; outliers can make it difficult to see the rest of the data.

Choosing the right chart type given your data is covered in more detail in Chapter 5 as we explore the role of the author of data products.

COMPREHENSIBILITY

Do you understand what it is saying? Do you know enough about the information being shown to know what it is saying? There are many data products that simply fail to convey information in ways that serve their users. For the worst of these, we've used the phrase “in-chart encryption.” Pie charts with dozens

of different colored values, 3-D bar charts with illegible labels, and tables with microfonts can make it almost impossible to recover the original data. Figure 4-15 is just one of many confusing charts available for public consumption.

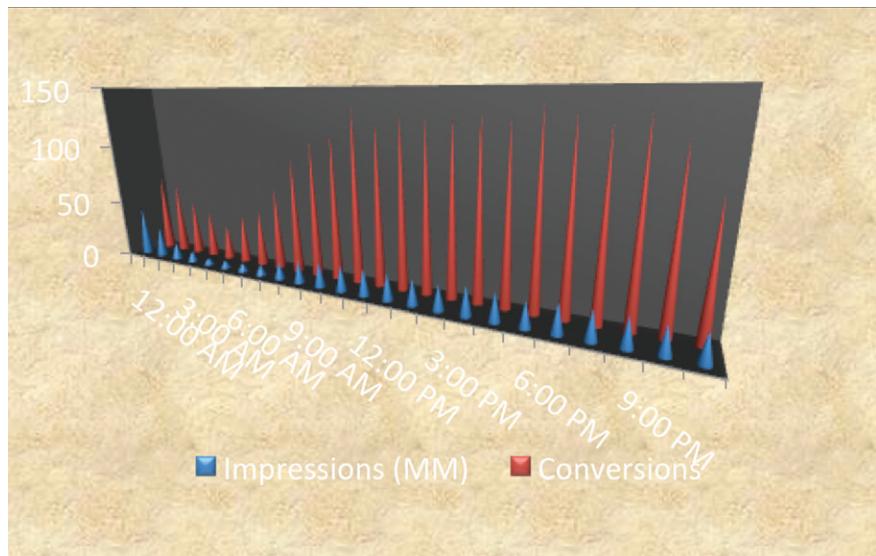


FIGURE 4-15 A confusing call center chart

This call center chart does a great job of blocking any data understanding. To avoid confusion, or to understand confusing data products, the following section suggests three guiding questions to dissect data products.

DISSECTING DATA PRODUCTS

You experience data products in everyday life—labels on food, traffic maps, and weather sites. The reports, presentations, and dashboards you see at work are seldom so expertly designed and easy to understand. To become a data fluent consumer, you'll need to build on your understanding of data and charts to explore the meaning in complex data products. This section provides a description of three key questions:

- Where does the data product come from?
- What can you learn from it?
- What can you do with it?

Where Does It Come From?

"Where are you from?" is one of the first questions we ask when meeting a new person. Historically, people were named after their birthplace (Joseph of Arimathea, Robin of Loxley, and so on) or by their profession, which went a long way toward understanding their background.

Data is no different. Understanding where data comes from means understanding how the data was collected and how it was processed before it came into your hands. It also means exploring the goals and motivations of the data product author. Try asking the following questions:

1. What real life behavior does it reflect? Data is a lens, a perspective on real life that is grounded in something tangible. Consider, for instance, the fantasy football example mentioned in the first chapter. Fantasy football scoring does not always directly correlate with actual game scoring, but it does tie to common events on the field. The more your players are involved in the action—catching passes, gaining yards—the more you can expect to score.

2. What are the strengths or weaknesses of the data sources? In examining any data product, a good place to start is to consider the source. Who recorded the information? Consider the motivation behind those providing the data product. Does the source have a financial incentive to provide accurate information?

A client of ours, for instance, discovered that the office locations of a large percentage of its national customers had been entered as the headquarters of its own company. When the company began to investigate, it found that those entering the data chose to set the address to the quickest thing that came to mind. There weren't any incentives to enter the data accurately. *The Data Journalism Handbook* notes:

The easiest way to show off with spectacular data is to fabricate it. It sounds obvious, but data as commonly commented upon as GDP figures can very well be phony. Former British ambassador Craig Murray reports in his book, Murder in Samarkand, that growth rates in Uzbekistan are subject to intense negotiations between the local government and international bodies. In other words, it has nothing to do with the local economy.⁴

3. What information is emphasized? Though data may be objective, subjective people author data products. There is often a message that has been baked into the data presentation. The choice of metrics often indicates what is important to the author, and the visual decisions can emphasize one conclusion

over another. For example, an executive dashboard that highlights quarterly sales and new customers is authored by someone who celebrates growth. Implicitly, the same dashboard minimizes the importance of customer retention and cost control. You might be growing revenue, but at what cost? Priorities will shine through.

4. What are the relationships between the different elements of data shown? Much of data analysis and visualization is aimed at revealing the connections and relationships in data. A trend chart shows the relationship between a metric and time. A scatterplot chart shows how two metrics relate to each other. Some of the relationships may already be obvious to you from your experience; others may surprise you with new insights into how the world works. View the data product as a tool for uncovering these relationships.

5. At what level of granularity is the data shown? There are different types of insights that can be gleaned from data depending on whether it is summarized or granular. Summarized data can show the big picture but glosses over important relationships and drivers of results. For example, the average height of a basketball team may be 6 foot, 4 inches, but this hides the fact that there is a 7-foot center and two guards under 6 feet. Granular data exposes individual correlations, outliers, and exceptions at the risk of believing a small pattern applies broadly.

6. What is the scope of the data? To draw accurate conclusions from a data product, you need to understand how the data has been filtered or narrowed. Time range is a good place to start because data for the last week may tell a different story from a year of history. A good data product often tries to cut out data that isn't relevant to the discussion—perhaps there are discontinued products, non-standard transactions, or categories of customers that would skew your results.

7. Do you have a shared understanding of the meaning of all the data fields? As you examine the data product, are there terms, calculations, or labels that require more explanation? Discussions about data are often derailed by different understandings of how a metric is calculated or what a dimension represents.

What Can I Learn from It?

To become a savvy data consumer, you need to transition from understanding "what it is" to "what it means." This section considers how to find insights

from raw data and examines some common charts to help you begin to find meaning. In doing so, you see the importance of recognizing the “unexpected” and contextualizing data through comparison.

Data visualization helps you see results and easily compare those results to your expectations. Data that matches your expectations can help you understand that they are on the right path and encourage you to do more of the same. Results that are unexpected, however, will spur people to action. You want to know: Does the world work the way you think it does? Or does the data show you something that challenges assumptions of expected results?

Data visualizations are designed to emphasize patterns and deviations in data. In fact, each specific chart type is well suited to highlighting particular forms of insight.

A skilled author of data products will choose the right visualization to emphasize a message. The data, chart, and supporting descriptions should work in harmony to point out what is interesting. The reader simply goes along for the ride.

Unfortunately, many data products end up a muddled mess of chart choices, obscure labeling, and arbitrary layout. The producer has passed responsibility to the audience to find the meaning. Chapter 5, “Data Authors: Skilled Designers of Data Presentations,” discusses the responsibilities of the data product author to build a “guided safari” for readers.

Here is a quick tour of common charts to show you how they help you spot the unexpected and use comparisons to find insights.

Unexpected Distributions

Pie charts are designed to show how something breaks into its constituent pieces. The slices add up to the whole, and the volume of each slice indicates its piece of the pie. (The obvious and tasty real-world analogy helps account for the popularity of pie charts in the face of withering criticism by data visualization experts.) The primary insight offered in a pie chart comes from slices that are smaller or larger than you would expect. Readers must imagine what they would expect the pie chart to look like and then find the differences.

In this pie chart (Figure 4-16), readers might be surprised to find that confects are nearly one-half their diet by volume. That’s not healthy eating.

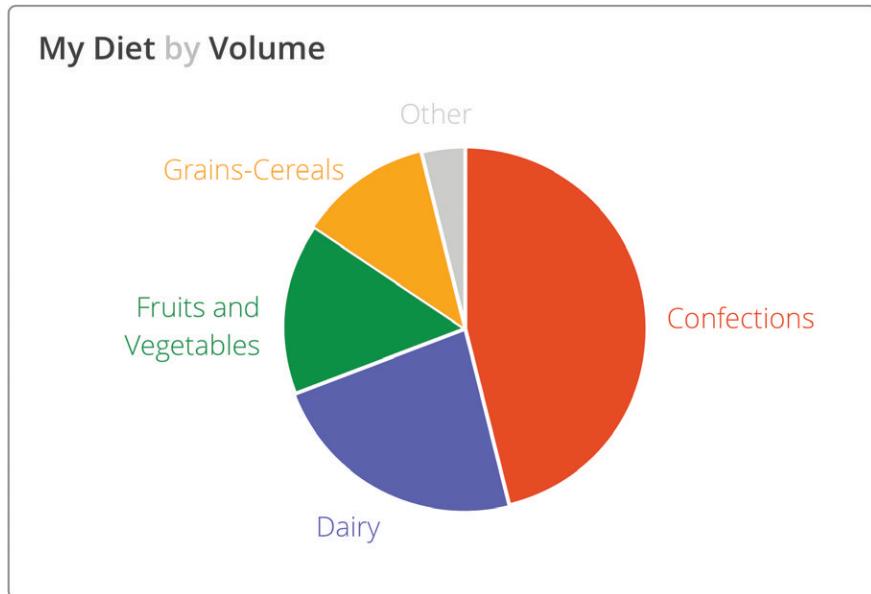


FIGURE 4-16 My diet by volume

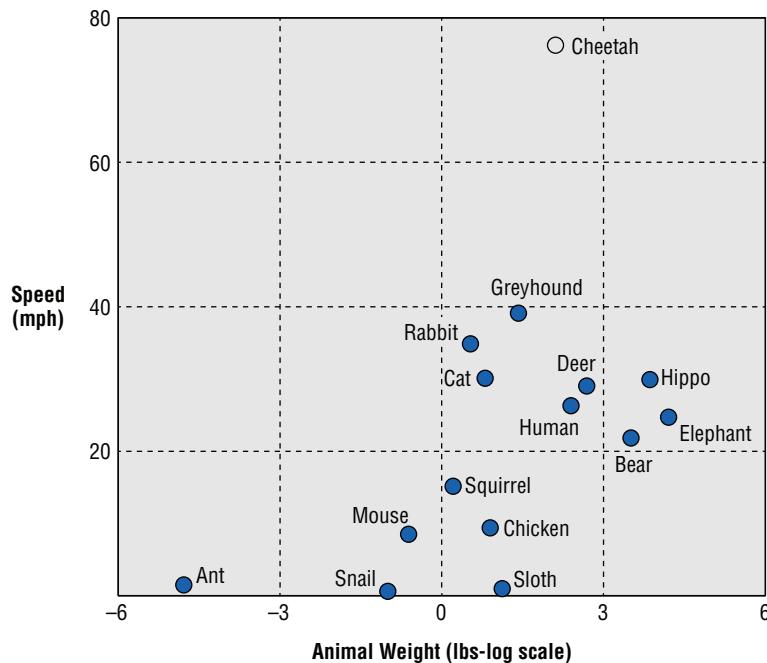
Unexpected Patterns or Relationships

Plotting data in a scatterplot or bubble chart is a way to show relationships between two or more variables. The pattern of the points may express a correlation that is either expected or surprising. Furthermore, outliers from this pattern are interesting because they break the mold.

This scatterplot (Figure 4-17) shows animal size versus weight. The data indicates a positive relationship between the size of the animal and its top speed. Bigger is faster, but with a lot of variation. The cheetah is an outlier with an unusually high ratio of speed to body mass. It's good to see that humans are faster than bears. Unfortunately, a detailed reading of the data behind this chart reveals that the human data point is Usain Bolt, world record holder in the 100 meter sprint.⁵

Unexpected Trends

Trends across time are another common place to look for insights. Line charts can make obvious deviations apparent when compared to expected patterns or trends. Like the pie chart, readers need to overlay their assumptions on the shape of the lines. Do you expect there to be an upward trend? Should the values remain steady over time, or is it normal to see substantial fluctuations?

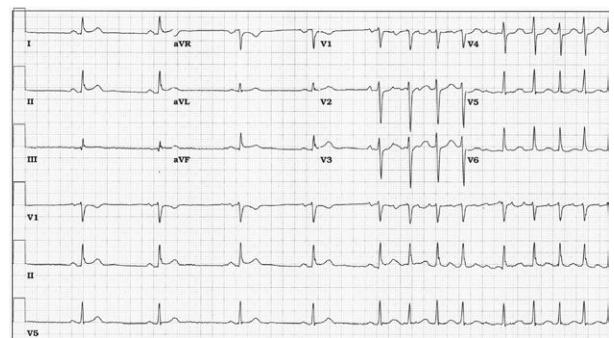
Animal Speed (mph) vs. Size (lbs-log)**FIGURE 4.17** Animal speed versus sizeSource: Data from *Natural History Magazine*, March 1974

An astute clinician could interpret this EKG tracing (Figure 4-18) to identify a patient at high risk for malignant heart arrhythmias.

Cardiologists use electrocardiograms (aka EKGs) to trace the electrical activity from the heart. A healthy heart demonstrates familiar patterns in the lines; changes to these patterns indicate problems. An experienced cardiologist can see abnormal heart rhythms, chamber enlargement, and signs of impaired blood flow through changes in the shape of the lines.

Comparisons

Data without context may offer little meaning. But adding a comparison value—whether an industry benchmark, an organizational goal, or a regulatory standard—brings values into focus. Comparisons across time periods can

**FIGURE 4-18** EKG rhythms and data

communicate improvement or regression. Direct comparisons can show how two or more entities rank compared to each other. Numerous specialized data visualizations have been designed to enable quick comparison, including bullet charts, “stop-lighting,” and leaderboards.

This dashboard (Figure 4-19) compares bank brands by a series of survey questions. Rankings and side-by-side comparison make it obvious who is performing better for each brand performance measure.

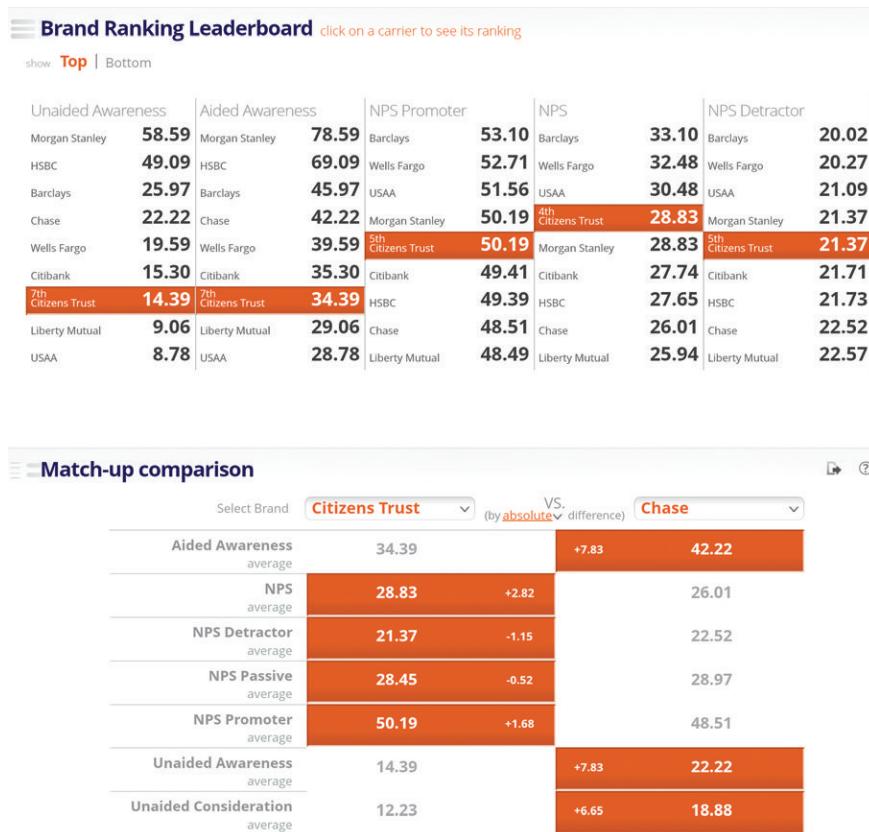


FIGURE 4-19 Brand ranking leaderboard

Find a Starting Point

A dashboard, report, or data visualization can feel like an ocean of information competing for your attention—like a *Where's Waldo* puzzle. Rather than trying to take in the whole picture at once, it's a good idea to focus your attention on a small piece of the picture. Focusing on a single element can help you grasp

the nature of the data, the dimensions and metrics displayed, and eventually how a small piece fits into the whole. Take the following data visualization comparing hospitals by patient experience as an example (Figure 4-20).

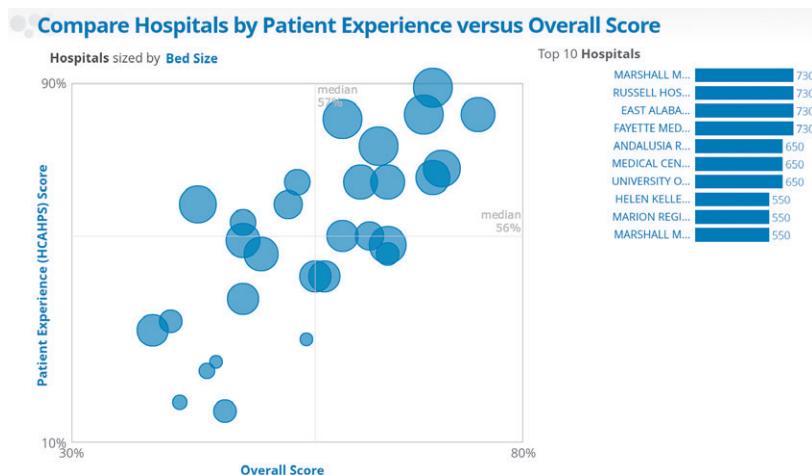


FIGURE 4-20 Hospitals and patient experience

There is a lot going on here. The bubble chart shows three separate metrics about each hospital. The meaning, size, and positioning of the bubbles requires new readers to carefully review the axes and legend to get their bearings. The connection between the bubbles and the bar chart on the right is not immediately obvious.

It's a lot easier to tell the story of a single bubble as illustrated in Figure 4-21.

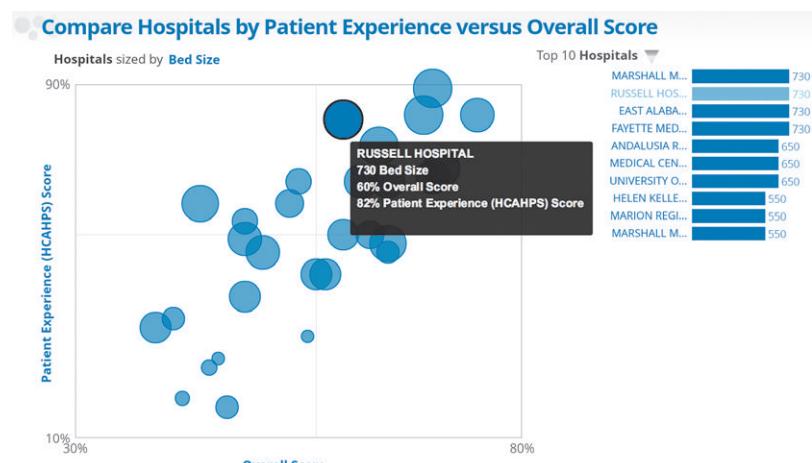


FIGURE 4-21 Hospitals and patient experience with emphasized bubble

The highlighted bubble represents a hospital, and the three metrics used to size and position the bubble are shown in the tooltip. The connection to the bar chart becomes obvious as Russell Hospital is identified as one of the largest hospitals by bed size. This particular data point may not be the most interesting or unexpected in this chart, yet now you have a much better sense of what the data product is trying to convey about hospitals.

It is often said that you know you have become fluent in a foreign language when you dream in the language. Novice language learners often translate their new language back into their mother's tongue word for word. Translating back to one's native tongue, while a crutch, is often needed for a time until one increases in fluency. Gaining data fluency is like developing language fluency. Without instant recognition of the meaning of a data visual, it can be useful to convert the information into a language in which you are familiar. In time, translating back isn't necessary.

Referring to Figure 4-21 to understand the message in the data, you might translate a highlighted data point into a descriptive sentence such as "Russell Hospital has 730 beds, tied with three other hospitals" or "Russell Hospital's patient experience score is near the top of all hospitals shown." This is a way of capturing and testing your understanding of the data depicted.

By breaking a complex data product into its smallest pieces and finding something comprehensible, you can start to understand both what the author is trying to show and how to read the content.

What Can You Do with It?

You are left with the job of sorting the good from the bad. Distinguishing between a data product that is useful and one that is distracting—one must understand whether it is reliable and/or valid. With reliability and validity determined, data products must be actionable.

Reliability and Validity

When considering whether you should use data, you first need to determine if it is reliable and valid. *Reliability* refers to the extent to which a measurement procedure yields consistent results; that is, if a particular technique, applied repeatedly to the same object or phenomenon, produces the same result each time.⁶ For example, you could think of using a scale to weigh yourself. If you stepped on the scale 10 times in a row and recorded your weight each time, a *reliable* scale would give you the same result. If, however, the recorded

weight fluctuated each time, and nothing changed about you between weighings (that is, you're not munching on Twinkies after each round!), you would conclude that the scale is *unreliable*.

Validity, however, refers to whether the measure accurately reflects the object or phenomenon that you are trying to measure.⁷ Your scale may tell you each time that you are 185 pounds; however, in reality you may be pushing 225. In this case, your scale would not be *valid*.

In Figure 4-22, Target A shows a reliable measure that is not valid—notice how the blue dots consistently arrive at a similar point, but they are not accurate and miss the bull's eye. Target B shows a valid measure in that the blue dots, on average, will arrive at a measure that is at or near the bull's eye; however, the accuracy is poor, making it unreliable. Target C, however, lacks accuracy and precision—the dots do not produce a valid result on average, nor do they consistently arrive at the same point. Finally, Target D shows a valid and reliable measure, one that arrives at a precise and consistent location, and one centered on the bull's eye.

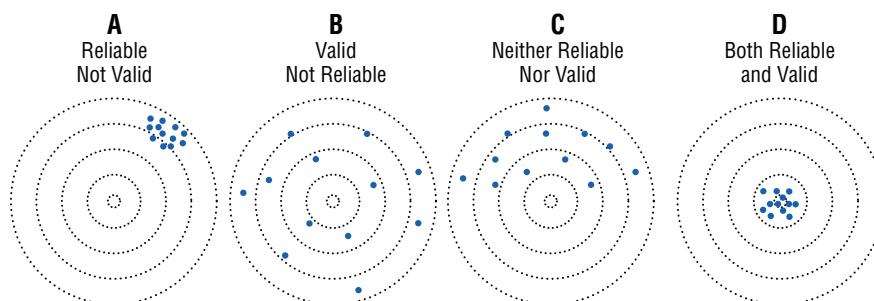


FIGURE 4-22 Understanding reliability and validity

Source: Juice

In assessing the reliability and validity of the data, it is important to consider a few threats to each. For reliability, think about the instrument of measurement. Is it a person? A computer? A tool or machine? Or is it a process that might include each of these? Although there are a number of formal statistical tests to measure reliability, these questions can be an important first step in questioning a data's reliability. In addition, processes that repeatedly measure the object or phenomenon of interest can also help determine the data's reliability. For example, you can never find out if your scale is unreliable based on a single measurement. Stepping onto the scale repeatedly in a short time will quickly help you determine if it produces a reliable measure of weight. The same is true for other forms of measurement, especially those that involve

multiple steps or components. A good place to start is learning more about the origin of the data and the process for collecting it.

When formal statistical evidence of reliability is not presented, we are left to use our own judgment and intuition about the reliability of the data and the process for gathering it.

There are also a number of ways to determine the data's validity. The most important and easiest of these is to ask whether the data has *face validity*. Do the numbers make sense? Are they consistent with past trends? If I weigh myself each day on the same reliable scale and it says 185, and then I weigh myself on a different scale and it says 225, does this mean I've suddenly gained 40 pounds? Of course not. Because the first scale is reliable, I know that the new scale is not valid. In education, if a student that consistently scores high on standardized exams designed to measure intelligence and academic ability fails the final test, I would think that the last test is not a valid measure of the student's ability. Perhaps she was sick that day or had just found out that her parents were getting a divorce.

Another consideration is to think about the data in relation to other measures. And although these relationships can be formally tested with statistics, it is important to do quick, informal checks for the validity of relationships between the data that you might expect to see. For instance, does an increase in the number of hours logged by a sales team also correspond to an increase in the number of sales? Does the retirement of a number of teachers in a school (without their replacement) lead to a higher student-teacher ratio? Does the multimillion dollar advertising campaign spotlighting a revamped website lead to increased search and visits to that site? Often these predictable relationships can help establish validity.

Finally, one must consider the selection process in determining the validity of data. Again, a number of questions are helpful:

- Who does the data represent?
- Is it a sample or does it include everyone?
- If it is a sample, how were they chosen? Why? Are the numbers "cherry-picked" to tell a specific story or sell a certain idea? Or was it random?
- Is the sample representative of the target population of interest? That is, if it is a sample of U.S. hospitals, for example, were only hospitals in major urban cities on the east coast chosen, or was the information drawn from a range of urban and rural hospitals from across the country?

With all these questions, the goal is to find out if there is any *bias* in the data, that is, is the information misrepresented in a particular direction. Bias can range from the accidental to the purposive, from poor procedures to corrupt processes. As with reliability, data fluent consumers are often left to use their own judgment about the validity of results. Face validity, examining the relationship between data elements, and considering bias are three important tools to help you determine what you can do with the data.

Actionable

If your data is reliable and valid, now what? What can you do with it? Does it offer information that informs decisions you can take?

Actionable data draws a direct and obvious line between the implications of the data and specific actions the reader can take. In our work with clients, we often help define key performance indicators, or KPIs. To test whether a KPI is worthwhile, we like to ask, “What would you do if that number changed?” Sometimes there is an obvious next step. A person can make an adjustment or react to the change. An activity is ramped up or a process is changed. But when the response is, “It sure would be interesting information to know,” then the KPI is probably more of a distraction than of value.

For a commonplace example, take goingtorain.com, the late, lamented website that was pure actionable simplicity. It asks the most important question many people want to know from their weather forecast—do I need to bring an umbrella to work today—and answers it in the most direct possible way (Figure 4-23).

In contrast, many weather sites are overloaded with news stories about dramatic weather in far off locations. These stories are more interesting than actionable. In our age of information overload, it is important to sift through information to the most relevant and actionable. In asking what you can do with data, it is critical that you begin to focus on using data that produces actions—shapes and changes the way you behave—because *that* data will be of most value.

it will be partly cloudy with a high of 96°f today in atlanta, ga

FIGURE 4-23 *Saying it directly*
Source: goingtorain.com

WRAPPING UP

It surrounds us and penetrates us; it binds the galaxy together.

Yoda

Data products surround you. They are stories, often summarized, sometimes raw. As a user of these data products, you want to be educated about what you are consuming. You need to answer questions like:

- Is the information trustworthy?
- What data is most important and relevant for you?
- How do you interpret and evaluate the message being delivered by the data product author?
- What information is actionable?

Data consumers are the foundation of data fluent organizations. They are on the front lines of decision-making—making the thousands of large and small decisions that will determine organizational success. Data consumers need more than the tools to make smart decisions; they need the skills to properly use these tools.

NOTES

1. Cotton Delo, "Pew Study: Blacks Over-Index on Twitter; Whites on Pinterest." *AdAge*. February 14, 2013. Retrieved on July 30, 2014 from <http://adage.com/article/digital/pew-study-blacks-index-twitter-whites-pinterest/239810/>
2. Clive Thompson, "Clive Thompson on Why We Should Learn the Language of Data." *Wired*. April, 19, 2010. Retrieved on July 30, 2014 from http://www.wired.com/2010/04/st_thompson_statistics/
3. Gareth Cook, *The Best American Infographics 2013* (Houghton Mifflin Harcourt, 2013).
4. Gray et al., *The Data Journalism Handbook: How Journalist Can Use Data to Improve the News* (O'Reilly Media, 2012). http://datajournalismhandbook.org/1.0/en/understanding_data_0.html#sthash.3tab55Y1.dpuf
5. Natural History Magazine, March 1974; The American Museum of Natural History; and James G. Doherty, general curator, The Wildlife Conservation Society via <http://hubpages.com/hub/Relative-Top-Speeds-of-Common-Land-Animals>
6. R. K. Schutt, *Investigating the Social World: The Process and Practice of Research*. (SAGE Publications, 2011).
7. E. R. Babbie, *The Basics of Social Research*. (Cengage Learning, 2012).

CHAPTER 5

Data Authors: Skilled Designers of Data Presentations

Data product authors (or data authors, as we'll use for shorthand) aren't just concerned with getting data on the page. That's like saying writers are just collecting enough words to fill a book . . . or movie directors are shooting purely to capture a few hours' worth of film. Rather, data authors, and the presentations they craft, must have a specific, targeted purpose. And the purpose must be directed at the audiences of the data products. Data can both help explain how things are, and how they could be, but data becomes truly powerful only when it informs, instructs, and leads to smart discussions, decisions, and actions.

Skilled data authors are an essential element in our data fluency framework (Figure 5-1). Your challenge as an author of data products is to build a bridge of understanding between your data and the intended audience. This bridge requires that your audience can interpret the data as presented, understand its relevance, and consider the implications and available actions.

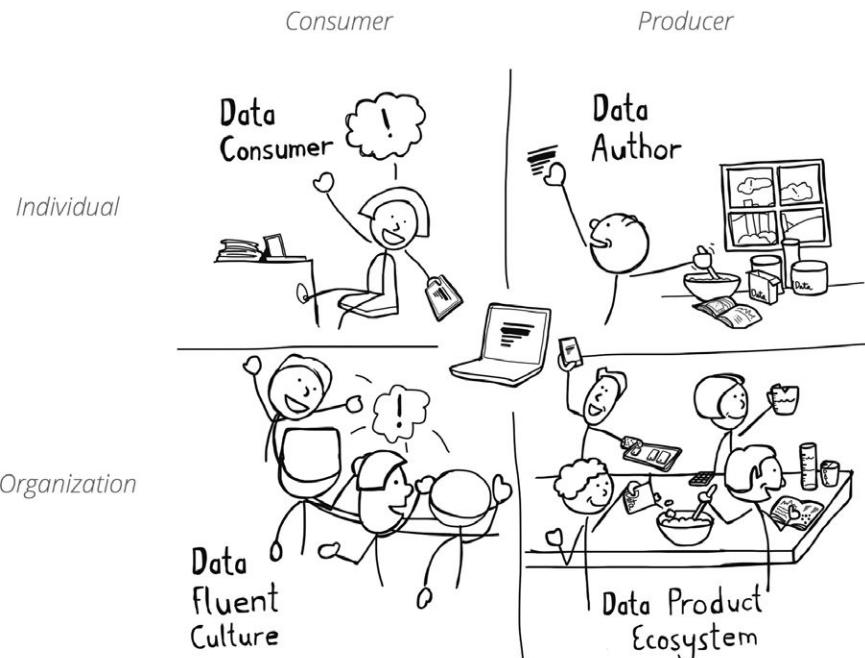


FIGURE 5-1: Data Product Authors within the Data Fluency Framework

A RARE SKILLSET

To bridge the gap between data and an audience, the data author has a complex job. The author must decide what data is most important to focus on to answer the questions at hand, and how to optimally harness and depict data to inform thinking and action. Effective communication with data is a rare skillset. Here are a few hats that data product authors must wear:

- **Salesman**—Data product authors must know their audience. To do so, they must consider how to best capture their audience’s attention, how they might perceive the data, and what it may take to gain the audience’s buy-in. An effective data product needs to be enticing, clear, and convincing to lead an audience to action.
- **Therapist**—Data product authors need empathy—the desire and ability to understand and share the feelings of others. By getting into the hearts and minds of the audience, they can find the questions that are

most influential. What will motivate an audience to action? What is the audience afraid of? How can the data address these concerns?

- **Connoisseur**—The best data authors are willing to make tough distinctions between data that is interesting and data that is important and actionable. It can go against our nature to put aside data that others might want to see. But the desire to deliver everything needs to be offset by an appreciation for an audience's limited attention span.
- **Data analyst**—Data authors can't create great art if they don't like working with their materials. Data authors need to be comfortable with core statistical concepts and comfortable with manipulating data. Getting involved with deep data analysis can reveal the important messages and accurate ways to convey them.
- **Ethnographer**—Data product authors should have a perspective on how the data product will fit into the way people work within their organization. How will the data product get incorporated into your audience's workflow? How does information travel throughout the organization? What do people care about and what do they ignore?

Unfortunately, many of the people who are asked to be data product authors aren't trained in these skills. Only now are leading academic institutions recognizing this market demand and educational gap. Although data analytics and data communications is becoming a part of leading academic public health and education programs, a commonly understood set of prerequisites doesn't exist.

Note We spoke to the director of Business Intelligence at a global computer manufacturer and asked who designs the company's dashboards. His response was, "nobody who is qualified." The user-interface designers in his organization weren't equipped to understand the data, and the data analysts weren't adept at presenting their results. In this gap, there needs to be a skilled data product author.

WHAT YOU'LL LEARN

This chapter is designed to help you build your skills for communicating—giving you the practical skills and mindset to become a better author of data products. Specifically, you will learn how to create effective data products by examining the following core concepts:

- How to know your audiences, develop purposeful products, and deliver a targeted message

- How to be discriminating with the data you choose to present
- How to define metrics that are meaningful and lead to action
- How to create a logical structure and narrative flow to your data products
- How to master basic design skills for making your data presentations attractive and easy to understand, including choosing the right form and language to present the data
- How to create data products that meet the needs of multiple constituent groups and start a dialogue about the data

Before diving into this set of skills, we first want to propose a helpful approach you can bring to designing data products.

GUIDED CONVERSATIONS

People working in the data analytics field believe in the importance of visual data presentation and the concept of *data storytelling*. There are core truths in the concept of data storytelling. First, data communication has an ability to both create and resolve tension. That is, data products often pose a problem and then deliver an insight that helps address the problem. Second, the author is an active participant in the exchange of information. The author sets a path that the audience goes down.

Nonetheless, we feel that the concept of data storytelling has gotten a little overloaded. For example, consider comments like the following:

All of this data visualization is, of course, really just a new way to tell stories (or create experiences).

Garrick Schmitt, Ad Age

The author of this visualization experimented with different views and arrived on the two that told the story most completely, most effortlessly.

From Data to Story: Dissecting a Well-Made Visualization

In fact, the analogy to storytelling is limited when applied to communicating with data. Data visualization has fundamental characteristics missing from traditional storytelling. For example, interactive data visualizations let audiences explore information to find insights that resonate with them. Visualizations

take shape based to a large extent on the underlying data. And as this data changes, the emphasis and message of the visualization is likely to change.

Therefore, communicating with data is less often about telling a specific story and more like starting a *guided conversation*. It is a dialogue with the audience rather than a monologue. While some data presentations may share the linear approach of a traditional story, other data products (analytical tools, in particular) give audiences the flexibility for exploration. In our experience, the best data products combine a little of both: a clear sense of direction defined by the author with the ability for audiences to focus on the information that is most relevant to them. The attributes of the traditional story approach combined with the self-exploration approach leads to the guided safari analogy (Figure 5-2).

Communicating with data is like creating a guided safari. On a guided safari your guide (the data product author) takes you down a path in which he thinks there will be a good chance to see wildlife. If you are on a guided safari in the Serengeti, you may be searching for the big five: elephants, lions, cape buffalo, black rhinos, and leopards. The tourists are led to places where the animals are likely to congregate, near food sources or watering holes. The safari guide knows the best places to look, but the story will change every day. The data product author's role is to create and lead an audience to a target-rich environment for finding data stories. It is up to audiences to recognize stories that are relevant to them.

Having considered your role as a data author in an environment with multiple influences, now turn to six core principles for creating an effective data product:

- 1. Find the purpose and message of your data products and know your audience.**
- 2. Be discriminating with what data you present.**
- 3. Define metrics that are meaningful and can lead to action.**
- 4. Create a logical structure and narrative flow to your data product.**
- 5. Master basic design skills for making your data presentation attractive and easy to understand, including choosing the right form and language to present the data.**
- 6. Create data products that serve a broader audience and start a dialogue.**

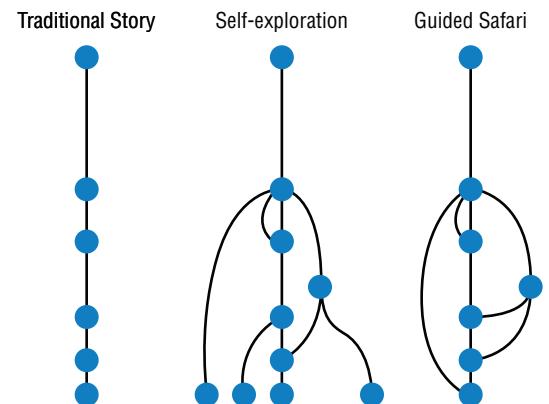


FIGURE 5-2: Analogies for storytelling and data visualization

The following sections provide details on each of these principles.

FINDING YOUR PURPOSE AND MESSAGE

Data products must strive for simplicity—a capability to communicate the most important data clearly, effectively, and directly. The message of a data product is likely to change as the data changes, but finding a simple way to present the data is a constant challenge for data product authors.

In the book *Simple* (Twelve, 2013), authors Alan Siegel and Irene Etzkorn propose a framework for simplicity that applies as much to data products as it does to their discussion of business writing, documents, and presentations. For them, simplicity requires the author to:

- **Empathize** by perceiving others' needs and expectations.
- **Clarify** by making the offering easier to understand, use, and benefit.
- **Distill** by boiling down and customizing what's being offered to meet needs.

Like Siegel and Etzkorn, you can put the audience at the forefront of your efforts to design data products. You need to know what information is most critical to your audiences' decision-making and the questions that they need answered to be more successful. Data products—data-rich presentations, reports, dashboards, and analytical tools—clarify by presenting data visually, using the right chart or graphic for the job, and by using clear language and thoughtful context to ensure understanding. Finally, data product authors need to be extremely discriminating in what information they choose to focus on. Only the right metrics and most actionable data should be featured to make the most of your audiences' attention.

In this section, you learn the following:

- How to create a balance between what the data is saying and the needs of your audience
- How to understand the perspective of your audiences and consider what data will be most important to them
- How to narrow your focus to deliver the most important information

The goal is to achieve a balance between the data, audience, and objective in your data product. Your data product has a purpose—whether it is as broad as

informing executives about overall organizational performance or as narrow as delivering a weekly e-mail highlighting sales activity compared to goals. With many competing demands, it isn't always easy to clearly define this purpose and therefore the scope and message of the data you need to communicate. The purpose is a balancing act, weighing the needs of the audience against your own objectives and considering the results and insights that percolate from the data itself (Figure 5-3). Now consider some thoughts on how to perform this balancing act, starting with the data.

Let the Data Speak

Whatever your personal objectives and the desires of your audience, your data product must be rooted in fair and honest analysis and summary of the data. Some data products, such as research reports and customer survey analyses, should rely first on the data to uncover a message. In these cases, there may not be a specific action that you are guiding your audience toward. The data can lead to many unexpected places and may raise more questions than it answers.

However, we caution against a blanket philosophy of "letting data tell its story." The data may not have a story. Or the story may vary for different audiences. Or the data may point to outcomes that are entirely outside of your organization's control. The cost of this focus on pure objectivity—and therefore discarding the inherent knowledge of the author or the needs of the audience—is often a directionless and sprawling document.

Your Objectives

Data products will, and should, bear the stamp of their author. They should reflect a goal that is important to you, as the person who is choosing what data to show and how it gets displayed. What outcome are you looking for? What do you hope to change in your organization by creating this report, dashboard, or analysis tool?

Data products can serve many purposes. The following (non-exhaustive) list includes many of the popular reasons for building data products:

- Help management define what is important.
- Educate people in the organization about the things that matter.
- Set goals and expectations for specific individuals or groups.

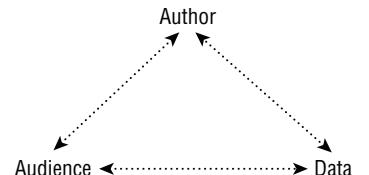


FIGURE 5-3: Finding balance between the data, audience, and author's objective

- Help organizational leaders sleep at night because they know what's going on.
- Encourage specific actions in a timely manner.
- Highlight exceptions and provide alerts when problems occur.
- Communicate progress and success.
- Provide a common interface for interacting with and analyzing important organizational data.

These are general motivations. You will have specific motivations for each data product you are involved with. Before letting the data speak or simply reflecting the will of your audience, consider what is important to you.

Your Audience

Data products need to start with the audiences in mind. Who are the consumers of the information? What do they already know? What are their experiences and prejudices? As you design the data product, understanding the audiences can help you craft a product that fits their needs and is something they love to use.

The needs and perspectives of different audiences can vary wildly. Consider the needs of a marketing analyst versus a chief marketing officer. A marketing analyst wants to know the details of a specific marketing campaign. He wants to know how results are calculated and the data sources that were used in a report. His job is to understand the data deeply, and he enjoys exploring data to find his own insights. Spoon-feeding him just the top-line results is likely to frustrate this marketing analyst, causing him to disengage from the information.

In contrast, the chief marketing officer has a much narrower attention span for diving into the data. The top-line results are exactly where she will want to start. If everything meets her expectations, she may not want to know anything more. However, the CMO's standards for the clarity and aesthetics of the data presentation are likely to be much higher than the marketing analyst. She spends her day looking at marketing campaigns that have a specific message and a polished display of information.

Often there will be multiple audiences for a data product. In fact, delivering the same report across different parts of the organization can offer the benefit of getting people on the same page. However, a diverse set of audiences complicates your efforts to communicate the data. The structure of a data

product can help. For example, separating the executive summary from an appendix allows the executive and the data analysts to both get what they want. Interactive data products allow users to drill into areas where they want to find out more.

The following table shows a few of the factors to consider about your audience and the implications for a data product design.

AUDIENCES'	QUESTIONS	IMPLICATION
Role	What decisions do they make? What questions do they need answered?	Structure the information to make it easy and direct to answer high-priority questions.
Workflow	In what context will they be reviewing the dashboard? What information are they using on a daily basis? How much time do they have to review the numbers?	The form and information display needs to fit into an existing workflow. For example, an on-the-road salesperson may need information delivered to her phone, not designed for an online wide-screen monitor.
Data comfort and skills	How sophisticated are they with using data? Do they enjoy digging into the numbers?	The dashboard's level of detail and analytical capabilities should match the audiences' comfort zone.
Industry and data expertise	How familiar are they with the key metrics? Do they understand where the data comes from? Are they familiar with internal company or industry terminology?	This determines the need for embedded explanations and use of natural language.

Asking the Right Questions

Albert Einstein said, "If I had an hour to solve a problem I'd spend 55 minutes thinking about the problem and 5 minutes thinking about solutions." The previous table is meant as a guide to help ask the right questions—spending 55 minutes thinking about the problem. Good questions, which are specific in nature, can help lead you to an understanding of the solution.

The website DoINeedanUmbrella.com is an example of asking the right question. It asks the first question many people have when they wonder about

the weather forecast. Instead of serving up a comprehensive 10-day detailed forecast, DoINeedanUmbrella.com simply provides a direct answer to a good question, as shown in Figure 5-4.

When looking at the data in our organization, isn't this really what we're trying to get to? Should we hire another developer? Are we going to meet our sales objectives? Will the product release be on time? Start with your message in mind and make sure, if you don't do anything else, that you deliver that message.

INFORMATION DISCRIMINATION

The most common mistake in ineffective data products is an inability to make difficult decisions about what information is most important. Amanda Cox of *The New York Times* graphics department summed up this problem perfectly, "Data isn't like your kids—you don't have to pretend to love them equally."



FIGURE 5-4: DoINeedanUmbrella.com gets right to the point.

Often information gets included in data products for reasons that are superfluous to the purpose, audience, and message—reasons that cater the product to someone influential or use information that has been included historically. The bar should be higher. Here are a few strategies to help narrow down to the information that matters:

- **Find the core problem**—Your data product should be more than a lot of data on a screen or page. It should have a core theme based on the essence of the problem. For example, a sales dashboard may be designed around the question, “How can we more effectively move leads through our pipeline?” Or a marketing dashboard may strive to answer, “How can we optimize our marketing investments?” Finding this core problem can give you the logic and argument for discarding extraneous information.
- **Ask a better question**—Data requirements can quickly turn into a laundry list of unrelated metrics, dimensions, and half-baked analyses. The root of this problem stems from asking, “What would you like to know?” The follow-up question that narrows down the list of requirements is, “What would you do if you knew this information?” This question separates the novel and whimsical desires from the important and actionable information. The next section talks more about actionable metrics and data products.
- **Push to the appendix**—Sometimes, it is impossible to ignore the requests for certain information to be included in a data product. Multiple audiences demanding a multitude of metrics. In these cases, it can be helpful to create an appendix report that includes some of these requests. Doing so can help keep the focus on the most critical data.
- **Separate reporting from exploration**—Tools designed for reporting need to be narrowly focused with a clear topic and address a limited set of questions. Exploration or data analysis tools serve a different purpose. There are many data products designed to give the users a broad palette to explore a variety of data. It helps to understand which of these goals you are setting out to serve.

The success of many data products is determined by an ability to distinguish between useful, productive information and interesting but ultimately extraneous information. In short, we echo the sentiment expressed by French author Antoine de Saint-Exupéry, “Perfection is achieved, not when there is nothing more to add, but when there is nothing left to take away.”

YOU CANNOT WRITE WELL WITHOUT DATA.¹

An author of data products faces many of the same creative challenges as writers. You struggle to find a message, speak to an audience, and develop an expressive style. Over the years, many accomplished authors have offered their wisdom for new writers. The lessons translate even if you work with data instead of words.

Find Inspiration

Find a subject you care about and which you in your heart feel others should care about. It is this genuine caring, not your games with language, which will be the most compelling and seductive element in your style.

—Kurt Vonnegut

If there is a magic in story writing, and I am convinced there is, no one has ever been able to reduce it to a recipe that can be passed from one person to another. The formula seems to lie solely in the aching urge of the writer to convey something he feels important to the reader. If the writer has that urge, he may sometimes, but by no means always, find the way to do it.²

—John Steinbeck

Develop a Direct, Clear Style

Readers want our pages to look very much like pages they have seen before. Why? This is because they themselves have a tough job to do, and they need all the help they can get from us.³

—Kurt Vonnegut

Try to leave out the part that readers tend to skip. A rule that came to mind in 1983. Think of what you skip reading a novel: thick paragraphs of prose you can see have too many words in them. What the writer is doing, he's writing, perpetrating hoopedoodle, perhaps taking another shot at the weather, or has gone into the character's head, and the reader either knows what the guy's thinking or doesn't care. I'll bet you don't skip dialogue.

—Elmore Leonard

Learn from the Best

No aspiring author should content himself with a mere acquisition of technical rules. ... All attempts at gaining literary polish must begin with judicious reading, and the learner must never cease to hold this phase uppermost. In many cases, the usage of good authors will be found a more effective guide than any amount of precept.

It is also important that cheaper types of reading, if hitherto followed, be dropped. Popular magazines inculcate a careless and deplorable style which is hard to unlearn, and which impedes the acquisition of a purer style."

—H.P. Lovecraft in *Literary Composition*

Serve your Audience

Forget your generalized audience. In the first place, the nameless, faceless audience will scare you to death and in the second place, unlike the theater, it doesn't exist. In writing, your audience is one single reader. I have found that sometimes it helps to pick out one person—a real person you know, or an imagined person and write to that one."4

—John Steinbeck, from an interview in the Fall 1975 *The Paris Review*

Write to please just one person. If you open a window and make love to the world, so to speak, your story will get pneumonia.5

—Kurt Vonnegut's 8 Tips on How to Write a Great Story

Discover the Core Message of Your Story

A short story must have single mood and every sentence must build towards it.

—Edgar Allan Poe

The solution to a problem—a story that you are unable to finish—is the problem. It isn't as if the problem is one thing and the solution something else. The problem, properly understood = the solution. Instead of trying to hide or efface what limits the story, capitalize on that very limitation. State it, rail against it.

—Susan Sontag, *As Consciousness Is Harnessed to Flesh: Journals and Notebooks, 1964–1980*

Give your readers as much information as possible as soon as possible. To hell with suspense. Readers should have such complete understanding of what is going on, where and why, that they could finish the story themselves, should cockroaches eat the last few pages.6

— Kurt Vonnegut's 8 Tips on How to Write a Great Story

Perhaps the most relevant guidelines for an author of data products comes from The Unpublished David Ogilvy,⁷ in a memo to his staff:

The better you write, the higher you go in Ogilvy & Mather. People who think well, write well.

Woolly minded people write woolly memos, woolly letters and woolly speeches.

Good writing is not a natural gift. You have to learn to write well. Here are 10 hints:

Read the Roman-Raphaelson book on writing. Read it three times.

Write the way you talk. Naturally.

Use short words, short sentences and short paragraphs.

Never use jargon words like reconceptualize, demassification, attitudinally, judgmentally. They are hallmarks of a pretentious ass.

Never write more than two pages on any subject.

Check your quotations.

Never send a letter or a memo on the day you write it. Read it aloud the next morning—and then edit it.

If it is something important, get a colleague to improve it.

Before you send your letter or your memo, make sure it is crystal clear what you want the recipient to do.

If you want ACTION, don't write. Go and tell the guy what you want.

DEFINING MEANINGFUL AND ACTIONABLE METRICS

Metrics, as you know from Chapter 4, "A Consumer's Guide to Data," are the values that you use to judge performance. They are also the focal point of any data product. They are the numerical reflection of the real-world behavior that your organization wants to improve, avoid, or shape. For example:

- In the credit card industry, "share of wallet" is a popular measure. The number represents the portion of a consumer's total spending that the specific card issuer has managed to capture. For the card issuer, it can't influence the total disposable income of its customers, but it can try to influence them to pull out their card more often at the cash register or in front of their computers.
- The No Child Left Behind legislation created an unprecedented focus on metrics in American schools. The primary measurement to receive federal funding under the Act is Annual Yearly Progress (AYP). To receive federal funding and avoid the "failing school" label, all students must meet a minimum level of proficiency on state standardized tests.

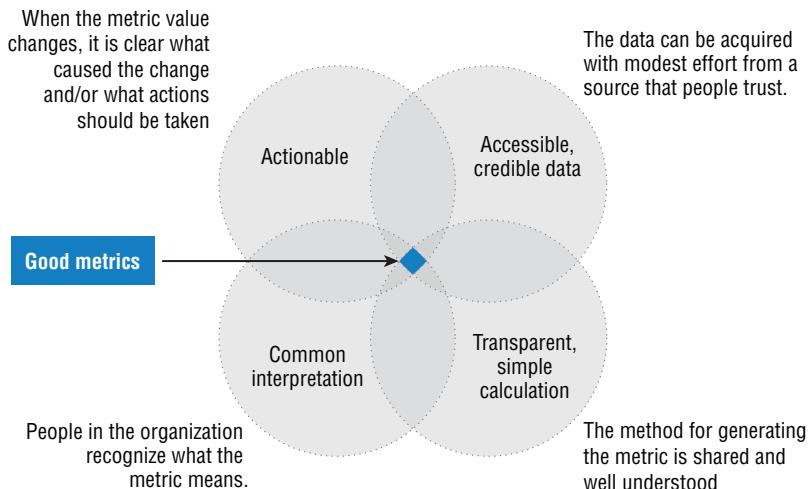
Although the requirements are waived in many states, districts and schools have dramatically reshaped their behavior to focus efforts on a new “bottom line” of education—state standardized test scores.

Metrics can create focus and alignment in an organization by providing clarity about what improvement looks like. Metrics can also create behaviors that are counter-intuitive or contrary to organizational goals. Beware of the call center that measures agent performance based on average call duration!

Despite the importance of choosing the right metrics, it is common to not think deeply about what to measure. Here are a few of the most common mistakes in choosing what metrics to focus on:

- **Historical conventions**—History isn’t always a good guide. Many organizations blindly follow the conventional wisdom in their industry or simply follow the path of history without thinking about the implications.
- **Simplistic metrics**—The path of least resistance may be to take what your data gives you. Time spent on a website is easily tracked but doesn’t tell you anything about whether the users were actively engaged or running off to make a cup of coffee.
- **Complex metrics**—The opposite of simplistic metrics are complex, contrived metrics that combine data from many data sources. If people don’t understand how a value is created, they may react skeptically to your analyses.
- **Too many metrics**—Many organizations fall into a trap of trying to have a comprehensive collection of metrics that track everything that managers care about. If you work with dozens of key metrics, those metrics may not deserve to be called “key.”
- **Vanity metrics**—Some metrics can make you or your organization look good but don’t tie to outcomes that are important. Be careful that you don’t select metrics simply because you know they’ll make you look good.

To avoid these common pitfalls and choose the best possible metric, consider the following framework, one that balances four factors, as shown in Figure 5-5.

**FIGURE 5-5:** Choosing the right metric

Source: Juice Analytics.

A lot of metrics fail in at least one of these dimensions. Following are a few examples:

- **Common interpretation**—We had a client who made a distinction between “leads” and “prospects” in its marketing organization. Prospects had theoretically expressed more interest in the service through their actions. Unfortunately, the line between leads and prospects was always hard to decipher and the definitions were hard to communicate.
- **Actionable**—Metrics are frequently too broad for the impact that a particular group can have. Customer satisfaction is a popular dashboard staple, but it is hard for most managers to see how they can have a significant impact on the number.
- **Accessible, credible data**—Sometimes, the most valuable and obvious metrics are frustratingly hard to track. In the web analytics world, unique visitors is important data to know, but user deletion of cookies makes it hard to trust the data.
- **Transparent, simple calculation**—Top NFL agent Leigh Steinberg says of the famous football quarterback ratings metric, “Other than one attorney in our office, I am unaware of a single human being who has the capacity to figure a quarterback rating.” I don’t know what

kind of art majors he hires, but all they need to do is use the simplified formula: $(83.33 * \text{Comp \%}) + (4.16667 * \text{Yds per att}) + (333.333 * \text{TD pct}) - (416.667 * \text{INT pct}) + 25 / 12$.

CREATING STRUCTURE AND FLOW TO YOUR DATA PRODUCTS

Your job as an author of data products is multilayered. The last section discussed choosing the right metrics. Your next challenge is to bring order and structure to the way your data is visualized. What is the general structure of the content you want to communicate? How does the content connect? How does one data or visualization element flow into the next? Like an author writing a book, there should be a starting point, themes, segues between chapters, and a conclusion.

In this section you gain ideas for the different ways that data products can be structured, guiding your audience along a path of discovery, and tips for laying out a page or screen with content.

A Guided Path: Structure and Flow

Some people describe this process as *data storytelling*. As mentioned in the introduction, storytelling might not be the right analogy. Stories are fixed and narrow in interpretation. Data is less so because it is always changing and can tell new stories. The creator of a data product is closer to a “guide”—laying out a path for the audience to travel along, but the findings and conclusions of the audience are more fluid. The landscape may always be changing.

Where do you want to lead your audience? As an author communicating with data, your objective is to help readers travel a path through the presented data in a way that they can move their understanding forward. Unfortunately, the standard practice for laying out most dashboards and reports has been to simply fit everything on the page in a grid or a jumbled mass of information. Figure 5-6 shows a dashboard with a grid overlay.

This rigid structure tells you nothing about how different charts relate to each other. It offers no clue as to where to begin understanding the data and nothing about what information is most important. Creating a layout that helps frame the content of your data product is one of the most undervalued areas of information design. The dashboard in Figure 5-7 starts with the few

high-level metrics (e.g., federal IT spend and change in spend) and then flows into the details about the agencies most responsible for these key metrics.



FIGURE 5-6: This common dashboard layout provides little guidance for users about how to explore the data.

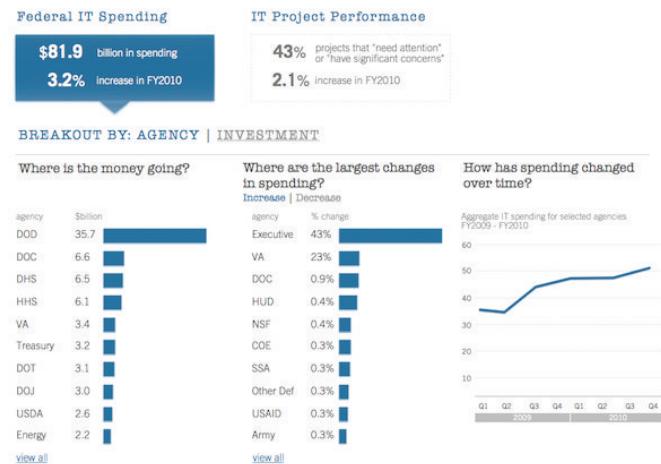


FIGURE 5-7: A redesigned version of the Federal IT dashboard makes it easy to see where to start and where to go next.

Source: Juice Analytics

Why Structure Matters?

Stephen Few, author of *Information Dashboard Design*, calls structure one of the greatest challenges of dashboard design:

Dashboard content must be organized in a way that reflects the nature of the information and that supports efficient and meaningful monitoring. Information cannot be placed just anywhere on the dashboard, nor can sections of the display be sized simply to fit the available space. Items that relate to one another should usually be positioned close to one another. Important items should often appear larger, thus more visually prominent, than less important items. Items that ought to be scanned in a particular order ought to be arranged in a manner that supports that sequence of visual attention.⁸

Whether it is a dashboard or a data-rich presentation, the structure of your data product is an opportunity to define the logical way to look at a problem or the business. How you choose to lay out the information shapes how your audience understands the big picture and how the smaller pieces fit together. At a more practical level, structure can serve as a navigational mechanism for the user. It shows where to start and where to go next.

Structure Options

A good data product layout requires a deep understanding of how the system you are measuring works. There are many ways to break something down into manageable parts. For example, the performance of an American football game can be deconstructed in many ways: 1) by offense, defense, and special teams; 2) by down and distance; 3) by time period; 4) by drive; and 5) by running versus passing plays. Data presented around each of these organizing principles would tell a different story.

Choosing the right way to structure your data for presentation is entirely specific to the situation. However, in our experience, there are three common categories for structuring your information on the page: flow, relationships, and grouping.

Flow

A flow-based structure emphasizes a sequence of events or actions across time (Figure 5-8). Systems that fit this model include leads moving through a sales pipeline, stages of customer support, and operational processes. Notice how the dashboard in Figure 5-9 is built around the flow of information from the summary metrics to submetrics to dimensional breakouts. The commitment to

the vertical flow structure makes it clear to users how to think about exploring the data.



FIGURE 5-8: Data products may be structured based on the flow of the underlying process.



FIGURE 5-9: Dashboard with a flow structure

Relationships

The structure of a data presentation can also emphasize the relationships between entities or measures (Figure 5-10). These relationships or connections may be mathematical, geographical, organizational, or functional. Figure 5-11 shows a visualization that displays scenarios for the valuation of a start-up. With the simple addition of a connective line and descriptive labels, the information in this tool helps users see the relationship between inputs on the left and projections on the right.

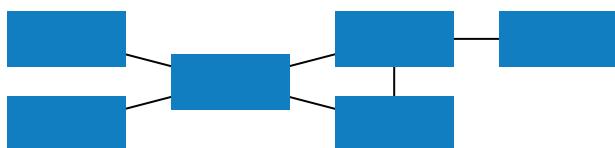


FIGURE 5-10: Data products may be structured to emphasize relationships.

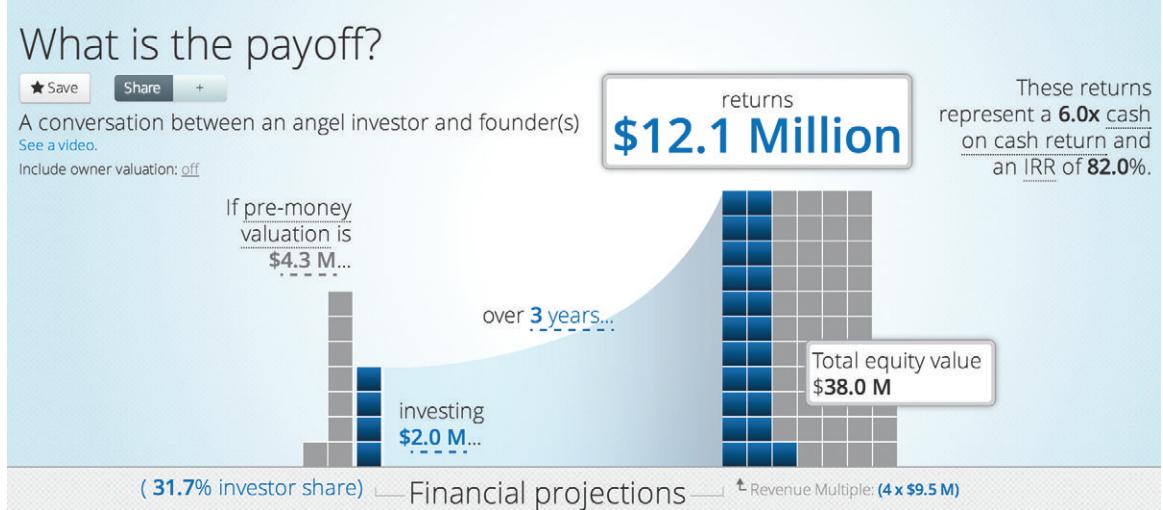


FIGURE 5-11: Company valuation tool shows the relationship between inputs and outputs (labs.juiceanalytics.com/valuation).
Source: Juice Analytics

Grouping

The structure of last resort is to group related information into categories or a hierarchy. The simple act of putting similar things together can bring some logic and accessibility to an otherwise haphazard collection of data (Figure 5-12). Figure 5-13 groups related metrics into four categories. Based on their needs, different users will know where to start in reviewing the performance data.

Organizing the Page

The little things in designing a data product can make a big difference—such as where you place the key metrics, where you position charts, and how much information you try to fit onto a page. For example, many dashboards can do an admirable job of displaying individual charts effectively, but how the charts are arranged on the screen can create confusion. Oftentimes, the information is compressed to fit on a page in such a way that readers cannot easily determine what the most critical information is. Here are a few things to keep in mind when laying out data on the page.

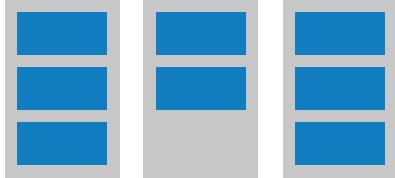


FIGURE 5-12: Data products may group elements to make the information easier to navigate.

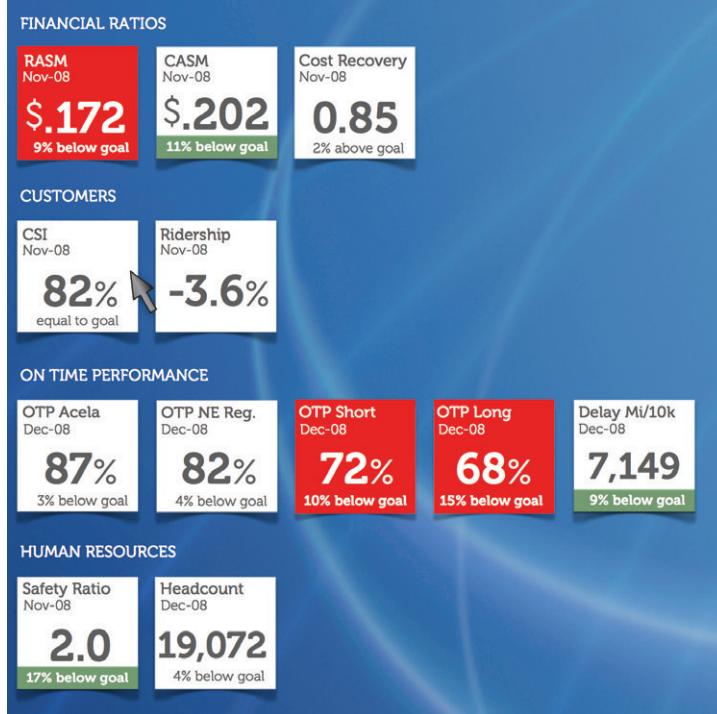


FIGURE 5-13: A dashboard with metrics grouped into categories
Source: Juice Analytics

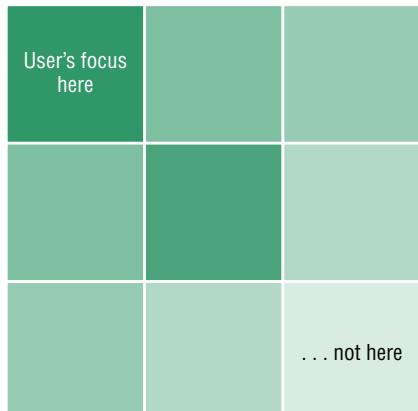


FIGURE 5-14: Where people focus their attention on a page. Source: Juice Analytics

Pay Attention to Attention

Position the most important information where people look first. Studies show that people tend to scan a page in a similar manner. Imagine that the image in Figure 5-14 is overlaid on your page of content. The research indicates that users look first for information on the top and left. Users also focus their attention down the left side. The center gets a fair bit of attention as well. But the bottom and right may not be noticed by your users at all. Therefore, as you choose how to present information, place the most critical information toward the upper left to capture your readers' attention.

Grids

Many web designers use a grid system—a series of columns and gutters of certain widths—to ensure that key lines in their

designs align. This approach brings a coherence and order to the page that puts users at ease.

We can overlay a grid on a site like TechMeme to see how this system gets applied in practice, as shown in Figure 5-15.

Not everything fits into one column, but it is obvious that this site has a pre-defined grid and sticks faithfully to it. The result is a clean, effective page layout.

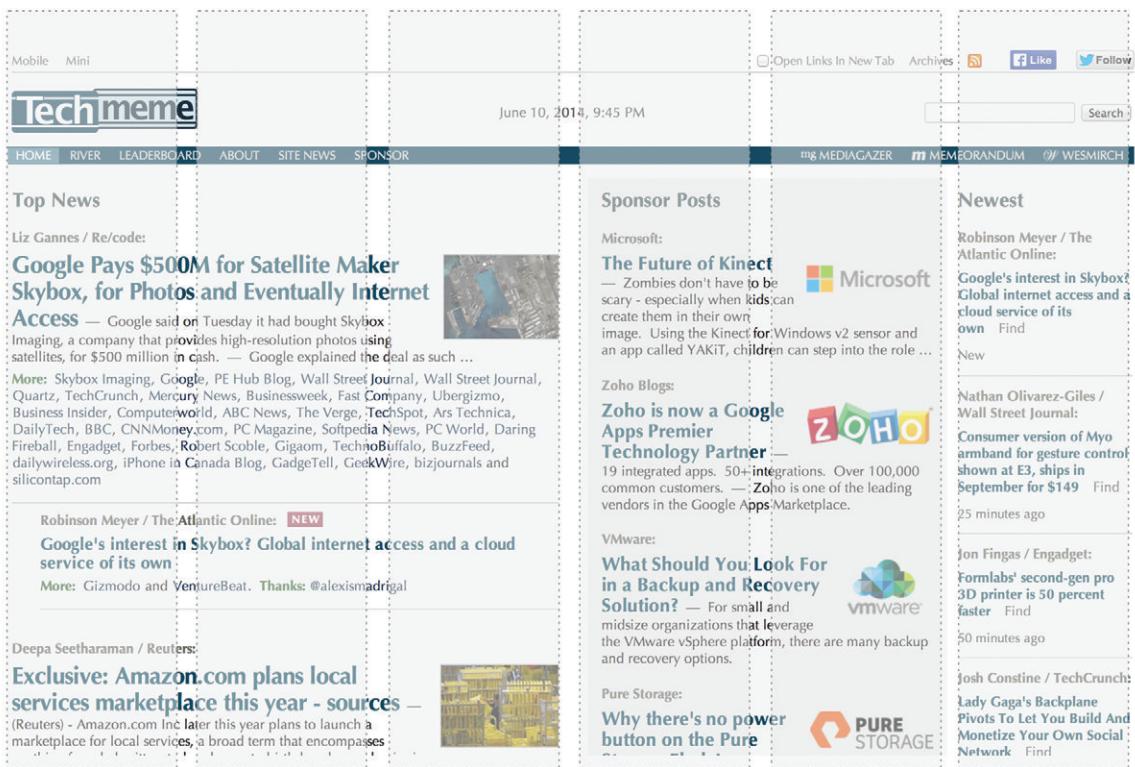


FIGURE 5-15: Many websites use a grid system to make the layout pleasing to the eye.

White Space

You don't just see objects themselves; you also see the space that is not there. White space in an interface design is important and too often overlooked. Maximizing real estate for data must also mean creating places for the eye to "rest" so that the nonwhite space has more impact. When you don't have sufficient spacing, everything runs together and you can't see what is most important, as shown in Figure 5-16.

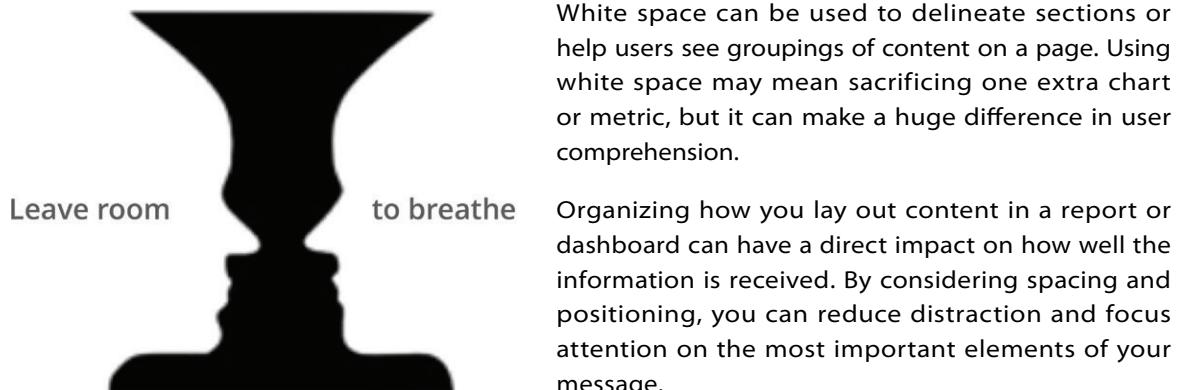


FIGURE 5-16: Use whitespace to aid comprehension.

PIXAR'S 22 RULES FOR STORYTELLING

Pixar's Rules for Storytelling⁹—as shared by Emma Coats,¹⁰ Pixar's Story Artist—are almost as relevant for communicating with data as they are to filmmaking. Whether you are a creator of data-rich presentations, infographics, or data visualizations, here's an abridged list of the most relevant Pixar principles for data storytelling:

- #2. You gotta keep in mind what's interesting to you as an audience, not what's fun to do as a writer. They can be very different.
- #3. Trying for theme is important, but you won't see what the story is actually about til you're at the end of it. Now rewrite.
- #5. Simplify. Focus. Combine characters. Hop over detours. You'll feel like you're losing valuable stuff but it sets you free.
- #8. Finish your story, let go even if it's not perfect. In an ideal world you have both, but move on. Do better next time.
- #11. Putting it on paper lets you start fixing it. If it stays in your head, a perfect idea, you'll never share it with anyone.
- #12. Discount the 1st thing that comes to mind. And the 2nd, 3rd, 4th, 5th—get the obvious out of the way. Surprise yourself.
- #13. Give your characters opinions. Passive/malleable might seem likable to you as you write, but it's poison to the audience.
- #14. Why must you tell THIS story? What's the belief burning within you that your story feeds off of? That's the heart of it.
- #17. No work is ever wasted. If it's not working, let go and move on—it'll come back around to be useful later.
- #22. What's the essence of your story? Most economical telling of it? If you know that, you can build out from there.

DESIGNING ATTRACTIVE, EASY-TO-UNDERSTAND DATA PRODUCTS

The last section discussed strategies for bringing order and structure to how data is displayed on a page. Your next challenge is to consider how the data looks, in what form it is delivered, and how words are best incorporated to facilitate understanding.

In this section you will learn:

- How the right form of data product can enhance presentation for your audience
- How functionality can enhance communication for your audience
- How attractive visual design can enhance presentation for your audience
- Core design principles for data products

Form

Data products can come in many forms. The volume of content will vary (from a single key metric to a comprehensive annual report) as will the medium used to transmit the information. A website, e-mail, animated slides, poster, glossy pamphlet, or 55-inch TV can all serve to communicate data. The challenge is in determining the right format for your audience.

The function of a data product is to communicate critical information to your audience in a way they can understand, delivered when and where they need the information. The form should capture its attention at the right time and move it to action.

Before you decide how you want to deliver your data product, consider a few factors that may influence the form:

- **Timeliness**—How frequently is the data updated?
- **Aesthetic value**—How important is it that the data presentation looks attractive, or can it be purely utilitarian?
- **Mobility**—Does the audience need to access the information through mobile devices?
- **Connectivity**—Does the dashboard need to connect to live data sources or can it be updated on a less frequent basis?

- **Data detail**—Will the data product offer a capability to drill down to see more context?
- **Data density**—How information-rich will views of the data be?
- **Interactivity**—Will the user benefit from interacting with the data?
- **Collaboration**—Is it important that your audience can easily share and collaborate with others about the data?

The following table assesses six forms of data products by their effectiveness in addressing the preceding factors. Which forms perform well for the factors your audience cares about?

FACTORS	HARDCOPY REPORT	EXCEL	WEB APP	E-MAIL REPORT	LARGE SCREEN DISPLAY
Timeliness	-	+	+	+	+
Aesthetic value	+		+		+
Mobility	+		+	+	-
Connectivity	-		+	+	+
Data detail	-	+	+	-	
Data density	+			-	
Interactivity	-		+	-	-
Collaboration				+	-

"+" means the form can work well for the specific factor; "−" means the form is not effective.

Visualizing Your Data

After considering the form of your data product, you should next consider the best ways of visualizing data to convey meaning. This section considers the steps of choosing the right chart, color, and typography and also discusses wrapping your data in context.

Choosing the Right Chart

Charts are the building blocks from which reports and dashboards are constructed. Choosing what building blocks to use is a seemingly simple decision that can determine whether your data product solidly communicates your message or misleads your audience into focusing on the wrong things.

What is the right chart for your data? This is the question we hear more than any other. Unfortunately, there isn't a simple answer or secret decoder ring.

However, if you can understand how different charts address different types of questions, you'll be well on your way to making good chart choices.

The first element in selecting charts is to understand that the purpose of a chart is to make it easy for readers to identify patterns and see how values compare to each other. The best charts maximize comprehension by playing to the strengths of visual perception. People can judge values accurately on a two-dimensional plane—the length of a line, the height of a bar, or the position of a data point. Bar charts and line charts play to these strengths. The ability to judge accurately declines when charts use other visual features to show values—width, area, color intensity, and radial distance. Therefore pie charts are generally less effective than bar charts because they require readers to compare radial distances, as shown in Figure 5-17.

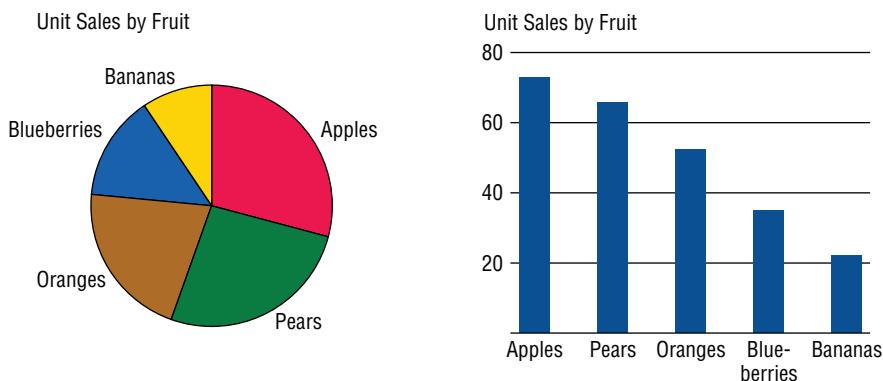


FIGURE 5-17: Bar charts make it easier to compare values.

The second fundamental element in choosing charts is to know what you want to show. There is no good answer to what is the *right* chart without knowing the question you want to answer. The beauty of defining this question is that it can reveal clues as to the chart you need to use. Take the following examples:

- How do my sales break down by customer type?
- What is the change in visitors over the last 3 months?
- Which students have the best test scores and have improved the most?

Each question tells you both what data elements are important and what you want to emphasize in your display of the data.

QUESTION	METRIC	DIMENSION	CHART CATEGORY
How do my sales break down by customer type?	Sales	Customer type	Composition
What is the change in visitors over the last 3 months?	Number of visitors	Time	Trend
Which students have the best test scores and have improved the most?	Test scores, change in test scores	Individual students	Comparison

Among the basic chart types that you find in most data analysis tools such as Excel, there are a limited number of categories of charts:

- **Distributions**—How do elements in the data break out across one or more dimensions? Examples: Bar chart and scatterplot.
- **Composition**—How does the whole of something break into its constituent pieces? Examples: pie chart, bar chart, and tree map.
- **Trends**—How are values changing across time? Examples: line chart and area chart.
- **Relationship or comparison**—How do elements in my data compare to each other? Examples: bubble chart and bullet chart.

Many years ago we gathered the charts associated with each of these categories into a free, simple online tool called Chart Chooser (www.chartchooser.com). Chart Chooser enables you to select the category of chart that you need, and then it displays the available chart choices, which you can download in an attractive Excel or PowerPoint template, as shown in Figure 5-18.

Numerous books, blogs, and websites discuss the design of charts for effective interpretation and deliver critiques of the many poorly designed charts and infographics. For more depth on this subject, consult the works of Stephen Few, a leader in this field with a passionate dedication to effective data communications.

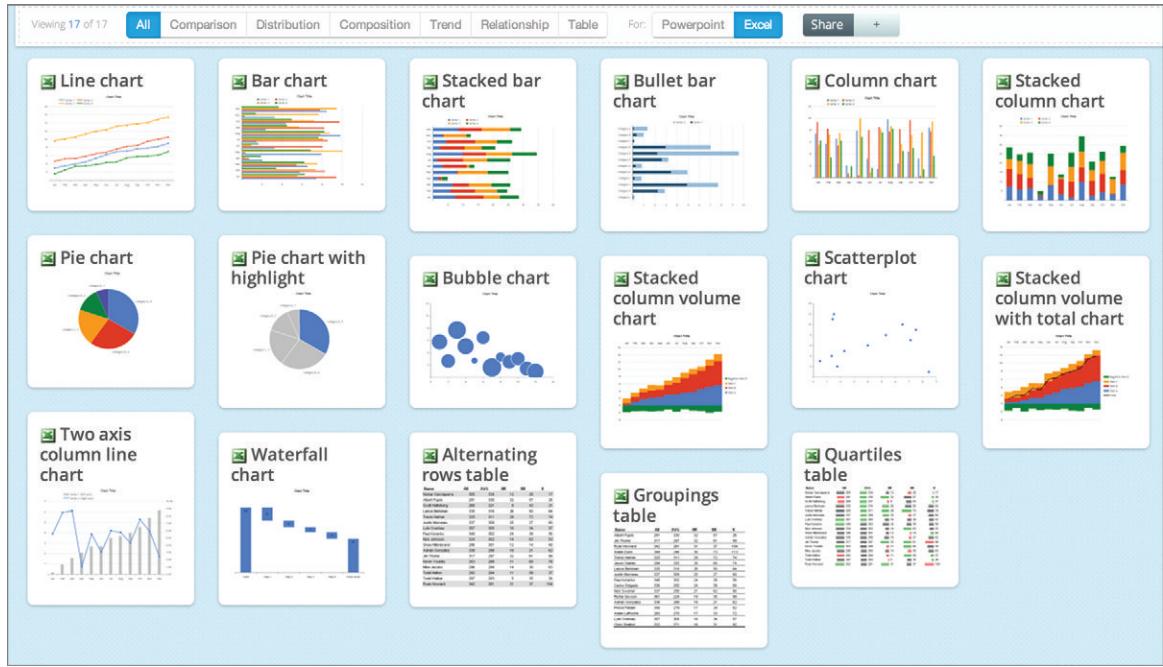


FIGURE 5-18: *Chart Chooser (www.chartchooser.com) lets you pick the right kind of chart for the job and download Excel or PowerPoint templates.*

Color

Quite often, data presentations get dressed-up with color like a festive holiday costume. Color is applied indiscriminately and adds little to the meaning of the data. Color should be used with restraint, adding more only when it brings appropriate emphasis or meaning to elements on the page.

Color Brings Meaning

Color can draw your eye to what is important and tie together similar things. For example, if you increase color brightness, it can attract attention and make a point seem important. Similarly, use of the same color hue can connect related things.

At a more subtle level, the color scheme you choose can evoke an emotion or feeling about the data. Colors can be broken into high-level dichotomies, such as earthtones versus unnatural colors. We perceive earthtones as calming (Edward Tufte has said that these are the kinds of colors you want to use if you just want to use color gently on your page). In contrast, unnatural colors jump out at your audience, making them ideal for showing an alert, as shown in Figure 5-19.

Earths tones

Gentle browns, blues. Calming, sinks into the page



Cool

Soothing, restful, calm



Unnatural colors

Alarming, unnerving, draws attention



Warm

Optimistic, active, vivid



Increasing color intensity

Increasing saturation and brightness draws the eye and means the point is more important.



FIGURE 5-19: The right colors can help re-enforce the meaning or message of your data.

Color to Display Data

When you use color in your graphs to represent data, consider three types of color schemes:

- **Sequential**—When you order values from low to high
- **Divergent**—When the values are ordered and there is a critical midpoint (for example, an average or zero)
- **Categorical**—When data falls into distinct groups (for example, countries) and therefore requires contrast between adjacent colors

Figure 5-20 provides an example of each of these color schemes.

Sequential

Colors can be ordered from low to high



Diverging

Two sequential schemes extended out from a critical midpoint value



Categorical

Lots of contrast between each adjacent color



FIGURE 5-20: Color schemes must fit their purpose in a chart

Typography

Typography is an obsession for some; it can include everything from choosing a typeface to picking the right point size, kerning, tracking, and leading. In the meantime, you have a data product to produce. This section provides a simple and useful framework for making decisions that ensure quality typography in your data displays.

With just few simple decisions, you can ensure that the text on the dashboard will both look good and communicate effectively. The majority of text on a page falls into four categories:

- **Body** text is clean, readable content.
- **Headers** separate and name major sections of your work.
- **Notes** describe additional things the reader should be aware of. For the most part these should fade into the background.
- **Emphasis** text is what you want your reader to pay particular attention to.

Figure 5-21 describes an approach for deciding how to display each of these text types. The highlighted areas indicate where you need to make decisions.

	Purpose	Size	Font	Color	Style
Body	Clean, readable text. 50–80% of your text will look like this.	10–16 pts	sans-serif: Helvetica, Arial, Tahoma, Verdana serif: Georgia, Times	Neutral	Normal No bold or italic 1.2 line spacing
Header	Separate and name major sections of your writing.	150–200% of body	Same as body or flip serif/sans-serif	Neutral	Normal, bold, or italic Whitespace above
Notes	Additional things a reader should be aware of, data sources, etc. Should “fade into the background.”	85% of body	Same as body	De-emphasize, lower contrast	Normal No bold or italic
Emphasis	Draw the eye to key points you need to make.	Same as body	Same as body	High impact color choose one or both	Bold or italic

your decisions

FIGURE 5-21: Our Simple Font Framework helps you use typography effectively with a few simple decisions.

You need to make only three basic decisions:

1. Choose size and font of the body text.
2. Decide if the header is going to flip to serif or sans-serif and whether it is going to have any style.
3. Decide what to do about emphasis, such as color or bold or italic.

A few things don't fit neatly into one of the four text categories listed, such as table headers and graph titles. You probably tend to use a combination of styles to handle these exceptions and will also want to emphasize text that is particularly important to your message.

Stick to this framework and your data product will have a more consistent, clean appearance.

Wrapping Data in Context

Context is the frame you put around your data to make it understandable to your audience. Context gives data meaning. Take the number 73, for example. This is a number, like many numbers, that means little by itself. Giving it a unit of measurement helps, such as 73 *home runs*. Here's a little more context: 73 is the record number of home runs hit in Major League Baseball in a season held by Barry Bonds. And if you are a baseball fan, the number means even more (or perhaps less) when put in the context of the steroid scandals of Bonds' era. By wrapping the number in the relevant information, it becomes meaningful and useful.

As an author of data products, you have many choices about context to help explain your data. The simplest place to start is with labeling. Any chart, data point, or visualization needs basic labeling to ensure the audience knows what it is looking at: title, axes, units of measurement, and source. Similarly, the scope of the data is critical baseline information. For example, does the key metric represent the last month or the last day of data?

These basics ensure the reader knows what you are showing. The data becomes more intriguing when values can be compared. Consider the following types of context to start to reveal insights:

- **Benchmarks**—Is there an industry value or average that reveals how your organization compares to others? Keep in mind that the most useful benchmarks are those that your readers consider “comparables” relative to factors such as size, industry, or region.
- **Goals**—Comparing a value to a goal goes a long way toward showing the audience how it might feel about the data it sees.
- **Changes**—How has a value changed over time? It is often more useful to consider data trends in values than the values themselves.

The key metrics shown in Figure 5-22 are a good example of providing context around some basic metrics. The values are surrounded by an indicator of change and shown relative to a quarterly goal and the quarter-to-date goal. The user of this dashboard is given much more than some key metrics; she can see whether those numbers are good or bad as compared to expectations.

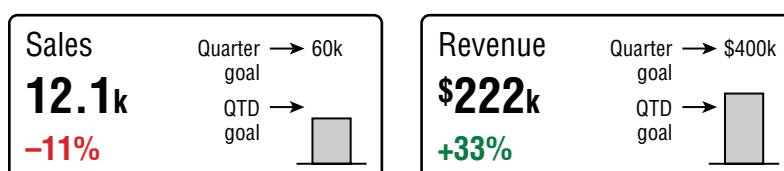


FIGURE 5-22: Key metrics shown with relevant context

A final layer of context to consider is what broad set of information should be shown on the same page or simultaneously. One-page dashboards, like the one in Figure 5-23, make an implicit statement that it is important to show all those metrics together to create an overall context. Effectively, the user needs to see all the metrics to have a good sense of whether any one metric is a particular concern and to get the big-picture understanding that would be obscured if each metric were shown separately. However, before you try to show everything at once in the name of context, consider the limitations of your audience's attention and ability to absorb the volume of content.

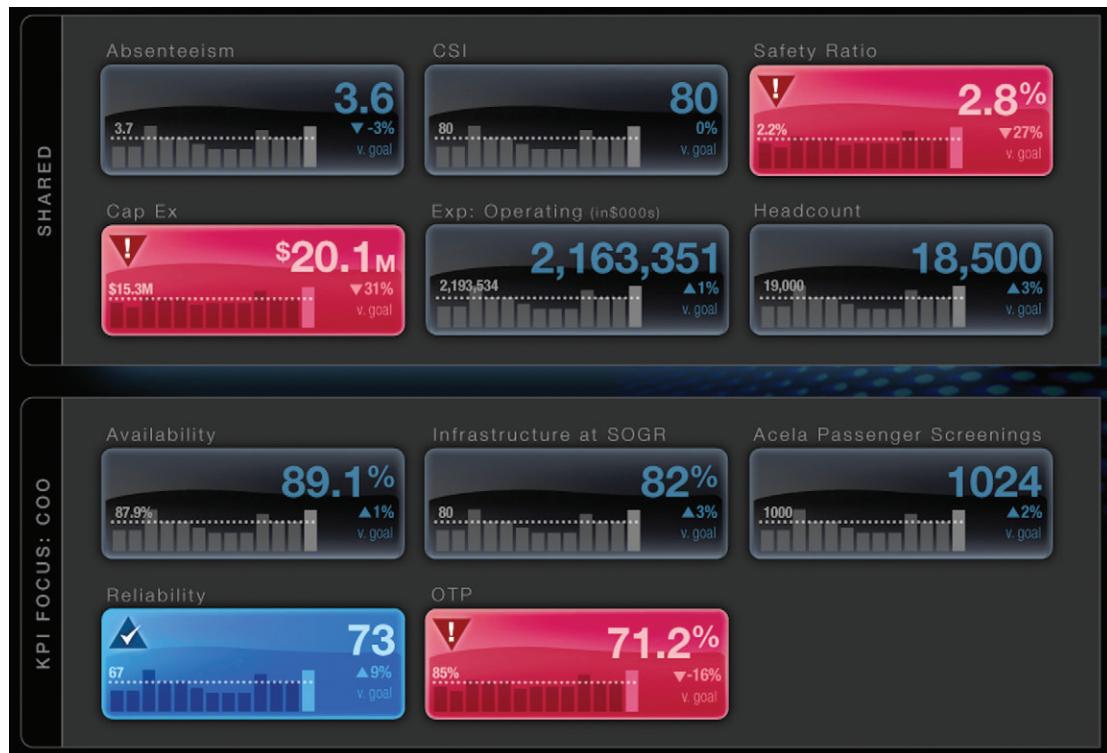


FIGURE 5-23: This dashboard implies that seeing the metrics together is necessary to understand the full context of the system status.

Language

Data products need more than data. The effective use of words can make the difference between understanding and confusion to your audience. When considering the language used to explain your data, the principle objective is to be understood by your target audience. In most cases, you can benefit from

choosing common, natural language over complex jargon, and using fewer, simpler words. For example, within the medical world it may be understood that *cephalalgia* is a medical term for headache, but if you tell the average person that you have a *cephalalgia*, he may wonder if that's a disease or some kind of growth. If the users of your data product need to look up terms or ask what an acronym means, you've distracted from the purpose of the data product.

As with everything else in authoring data products, audiences come first. Consider how language can be tailored to the way your audiences speak. What terms are they familiar with? Do they understand the meaning behind complex metrics? For internal audiences, commonly used shorthand may be perfectly accessible. But when crossing different functional areas, the familiarity with terminology may dissipate. It is better to err on the side of expressing concepts in simple and more explicit terms. CPC and PPC are well understood in marketing, but a finance executive is better served by referring to "cost-per-click" or "pay-per-click" advertising. The importance of language becomes even more pronounced when crossing organizational and international boundaries.

Many marketers, web designers, and consultants have established a "voice" in their writing that creates a balance between friendly and professional. The goal of business writing is described by Carter Daniel in his book *Reader-Friendly Reports* (McGraw-Hill, 2012):

Business writing isn't a different language with a separate set of words and phrases. In receipt of, as per your request, beg to acknowledge, and please be advised are relics of the past. They aren't used anymore, or shouldn't be anyway.

Business writing also isn't dull and stereotyped. Bad writing is dull; good writing is interesting. These statements are true for all writing, business or any other kind. If you are interested in the subject but find the report dull, something's wrong with the way the report is written.

Business writing is like any other form of writing except more efficient. Think for a minute about who your audience is, and you'll understand the reason for insisting on efficiency. Business people are intelligent, suspicious, and busy. So when you write for them you have to be factual, persuasive, and efficient: factual because they are too intelligent to be fooled by vagueness and bluffing; persuasive because you have to overcome the suspicion that always accompanies money matters; and efficient because you'll lose your reader if you waste time.

Therefore, when you are writing the descriptions, labels, or notes that accompany your data product, focus on how the text can be easy to read for your target audience and expressed in a way that is straightforward, concise, and engaging.

CREATING DIALOGUE WITH YOUR DATA PRODUCTS

Although data product authors must attend to a host of technical aspects when designing presentations, the products of your work serve a broad audience of stakeholders and are ultimately tools for creating dialogue and decision. In this section, you learn:

- The importance of serving your audiences' audience
- How data products gain value as your audience responds to the content
- Strategies for capturing feedback and refining your data products

Your Audience's Audience

Throughout this chapter, the importance of serving your audiences has been emphasized. This ability to understand and take action on the data is what ultimately matters. However, your data product seldom stops with your primary audiences. It will be passed on and shared from your primary audiences to their colleagues. In some cases, your audiences' audience may be more important in the impact the data product can have.

For example, imagine a marketing analyst who has prepared a quarter-end presentation to explain the progress the group has made in attracting new customers. This presentation may be presented to the chief marketing officer (CMO) and other members of the team to evaluate where to focus energies in the coming months. However, the CMO is likely to cherry-pick a few charts from the presentation to share with the organization's executive team. Meanwhile, another portion of the presentation may find its way to the finance group to help with budgeting and projections. The data product author may not have set out to communicate to people across the organization, but information will flow where it is needed.

Knowing that your data products or visualizations may reach a secondary audience, it is good to construct them with that in mind. You may build products that are modular in format with certain modules being deleted for certain audiences.

Data Leading to Dialogue

Data alone isn't valuable. In fact, it can be expensive in time and resources to manage and maintain. The analysis of this data is closer to something that is valuable. A clearly communicated analysis starts to transform a reflection of the world into knowledge in the minds of people.

Even so, knowledge alone does not make your organization better. It is the decisions and actions of people—based on this data-sourced knowledge—that is the goal. But these decisions are seldom made in a vacuum. In most organizations, decisions are a collaborative, social experience. People come together to discuss options, review their knowledge of the situation, and arrive at a path to go down. Herein is one of the great powers of effective data products: They can shape and guide these discussions. Conclusions are seldom clear-cut, even when there is data to support a direction.

Your job as an author of data products isn't necessarily to drive people directly to data-driven decisions. It is seldom about delivering carefully packaged answers. Your job is to encourage productive dialogue. You can help people focus on the right things, be informed, and ask relevant questions.

The outcome of a data fluent organization is informed discussion, better questions, and new insights about how your organization works, what your customers need, and how you can better serve them.

DESIGN PRINCIPLES

This final section includes a list of principles to consider when you design data visualizations and data products. Though every data set and situation is different, these principles offer guidelines that apply in many cases and can improve what you deliver to your audience.

VISUALIZATIONS

When creating charts and tables for presenting data, the following eight design principles are worth considering before you click Print or Send.

1. Reduce chart-junk and increase data-to-ink ratio. These are the first two recommendations from Edward Tufte, well-known information design expert. Reduce chart junk by removing elements that are decorative or ornamental. Three-dimensional chart effects, for example, add nothing of value to your chart. Increase data-to-ink ratio by making every pixel tell a story about your data. Watch as Figure 5-24 is gradually stripped of extraneous chart junk.

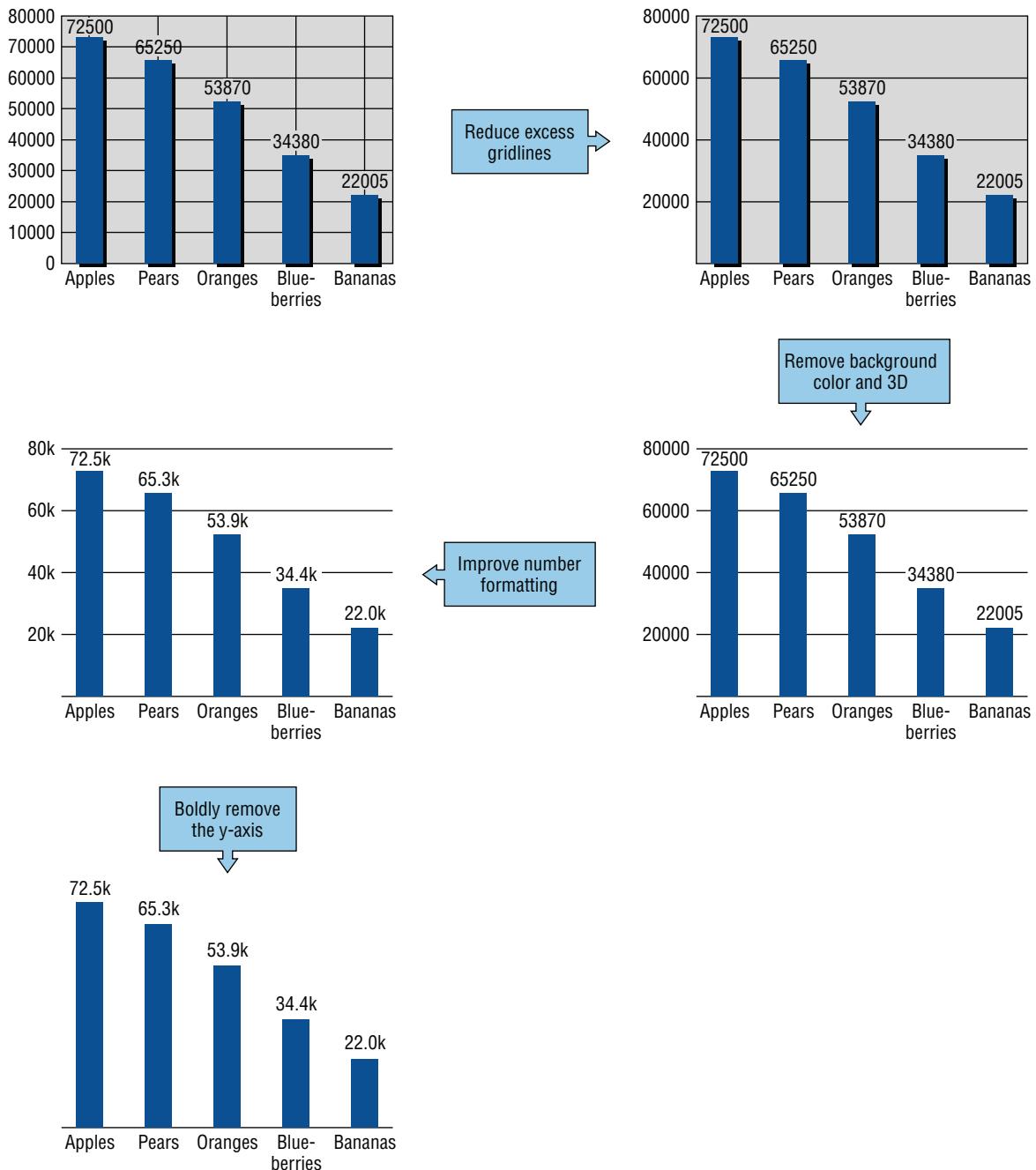


FIGURE 5-24: Reduce chart junk.

- 2. Use contrast to make the data visible.** If there is low contrast between the data and labels compared to surrounding elements, it can be difficult for readers to see the most important information on the page. For many years, the standard Excel default chart, for example, made it more difficult than necessary to distinguish the line from the background. A white background and de-emphasized gridlines can help, as shown in Figure 5-25.

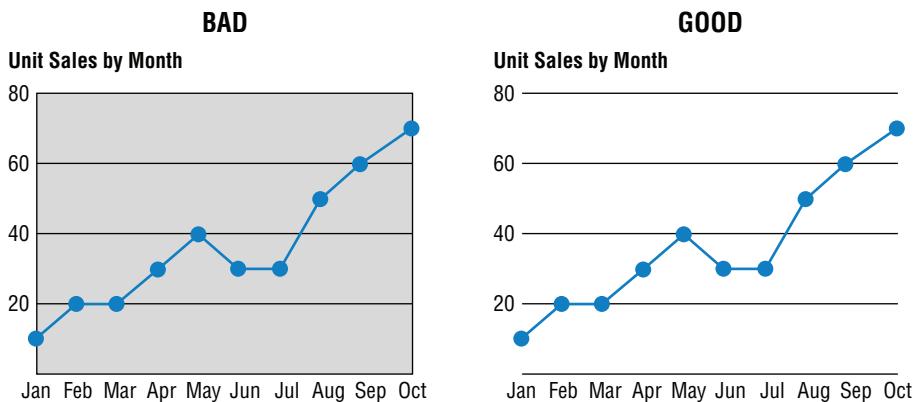


FIGURE 5-25: Improved contrast makes the data easier to read.

- 3. Readable labels.** Clear labeling is often overlooked. Whenever possible, avoid rotated labels or tiny font. When the labels are a struggle to read, they distract from focusing on the numbers, as shown in Figure 5-26.

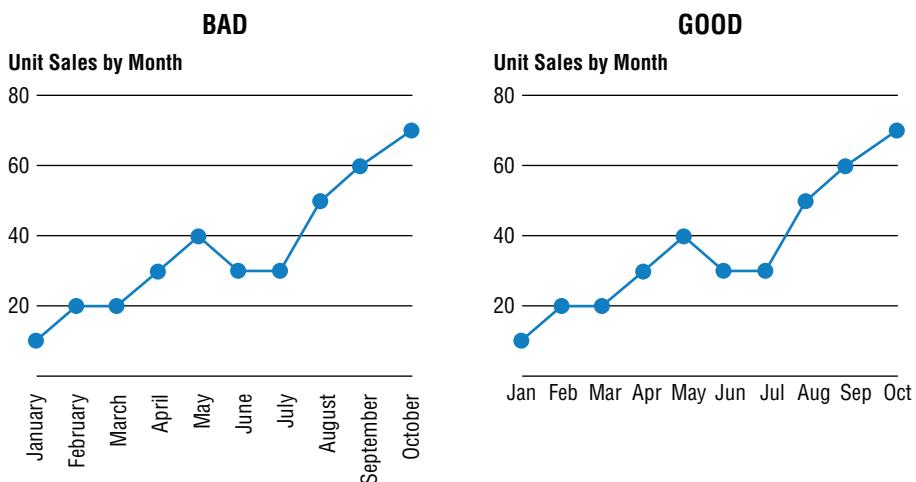


FIGURE 5-26: Don't make it hard on your audience to read labels.

- 4. Don't repeat yourself; repetition is bad.** It's not necessary to have both a legend and a title for single series graphs. Likewise, the title of a chart may sufficiently explain what the reader is looking at, as shown in Figure 5-27.

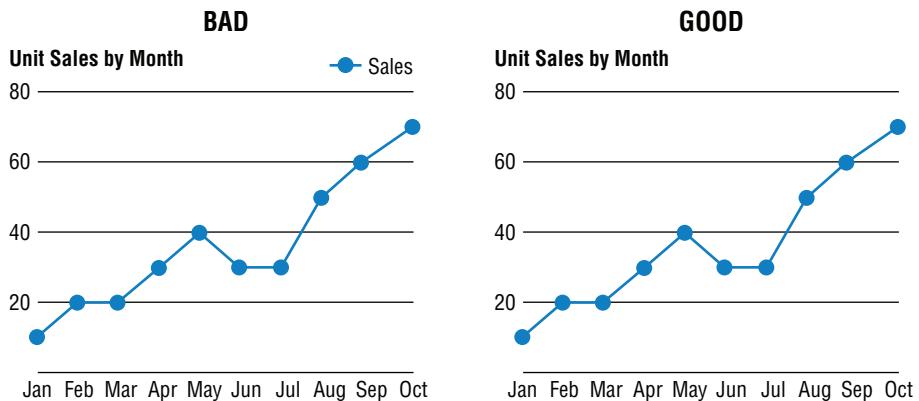


FIGURE 5-27: Remove repeated elements to clean up your charts.

A useful alternative to a legend is to label individual data series directly, as shown in Figure 5-28.

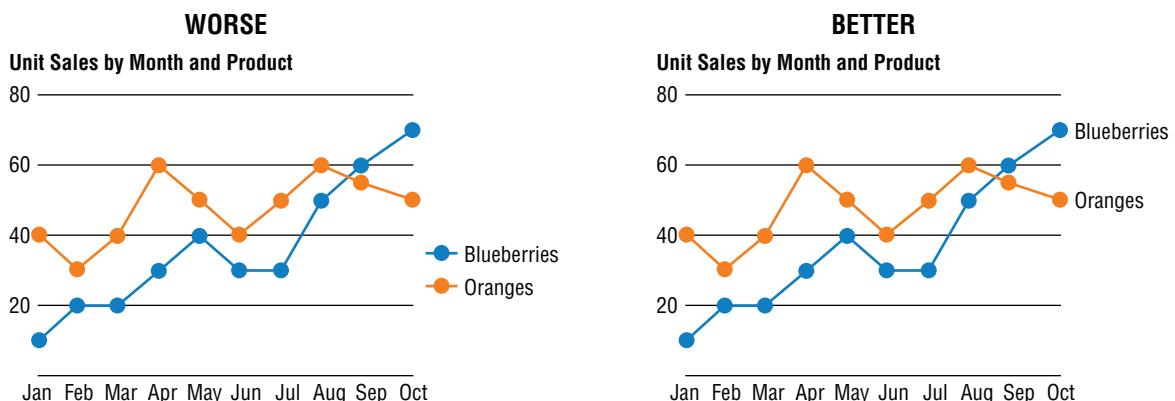


FIGURE 5-28: Direct labeling of data series

- 5. Avoid smoothing and 3-D.** A smoothing feature added to your line gives the impression of data points that are not actually there. A smoothing feature added to your line gives the impression of data points that are not actually there (Figure 5-29). Similarly, glossy 3-D effects are a visual that was impressive in 1999 but contributes no value to your chart.

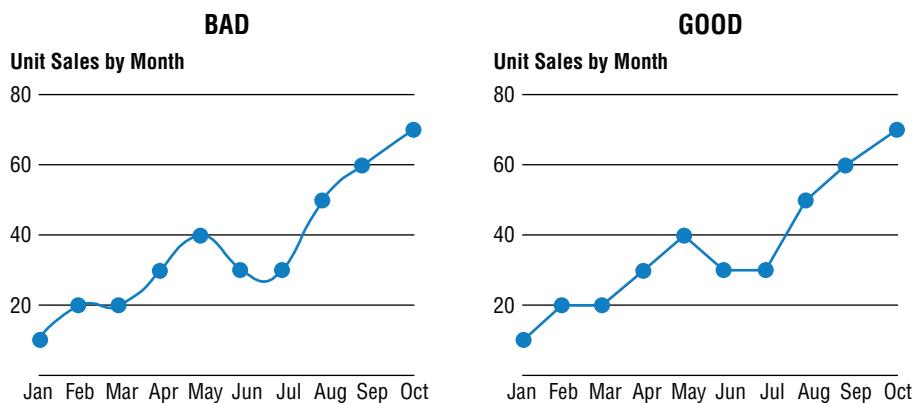


FIGURE 5-29: Smoothing of lines is seldom useful.

- 6. Careful use of gradients.** Use flat colors or a bare minimum gradient. Ensure that the bar endpoint is visible. When the gradient fades toward the endpoint, it reduces the ability to clearly see the length of the bar, as shown in Figure 5-30.

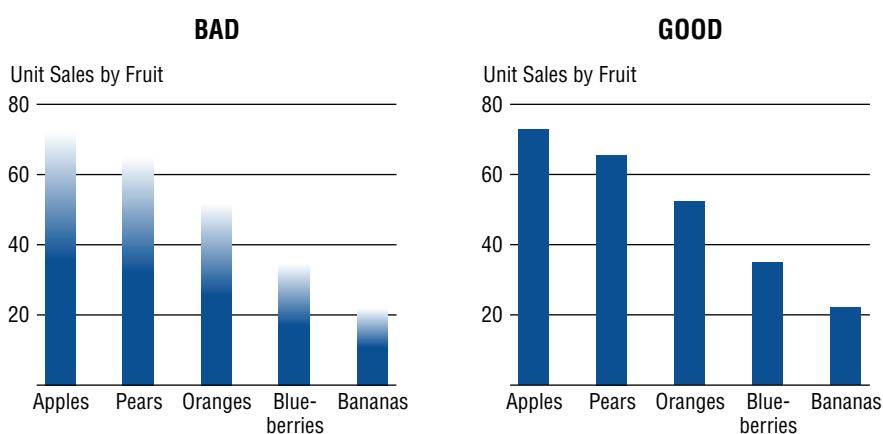


FIGURE 5-30: Bars should not fade out toward their endpoint.

- 7. Sort for comprehension.** Add structure and clarity to the chart by sorting by a metric of interest. Users want to see what values are largest or smallest. They shouldn't have to hunt through a chart to find this basic information, as shown in Figure 5-31.

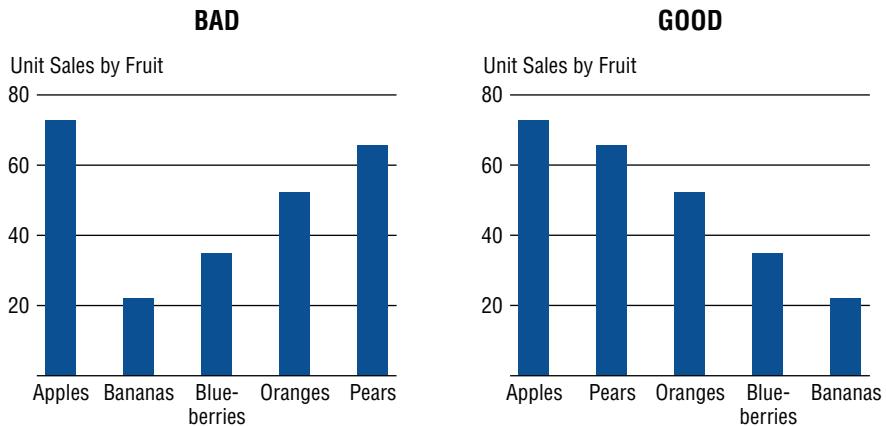


FIGURE 5-31: Sort columns to make it easy to find the interesting insights.

- 8. Use color variants.** If you are displaying multicolumn or stacked charts, use variants on a hue or gray to show different data series. This approach reduces the need for a wide variety of distracting colors, as shown in Figure 5-32.

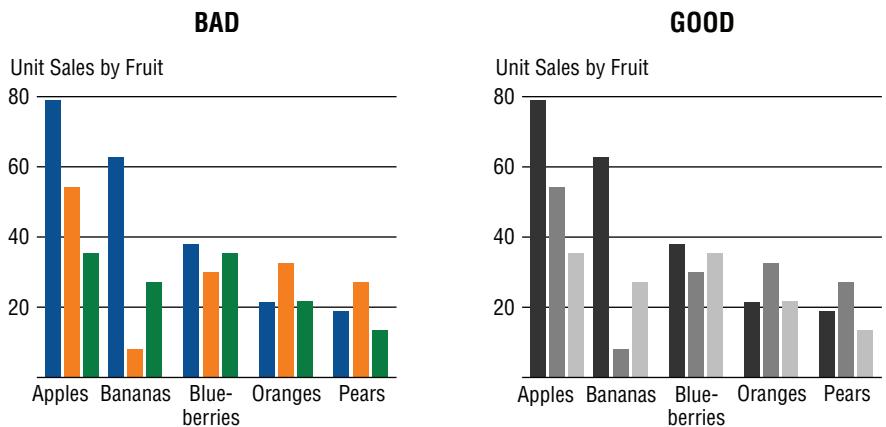


FIGURE 5-32: Fewer colors can be better.

Also, don't vary the colors by point like in the chart in Figure 5-33. This creates a lot of unnecessary visual noise and makes similar colors seem related.

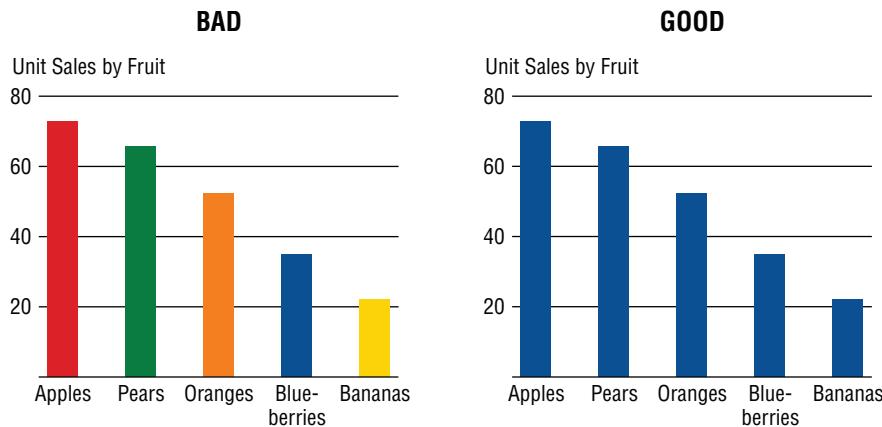


FIGURE 5-33: Separating colors by column adds no value to this chart.

DESIGN PRINCIPLES FOR DATA PRODUCTS

As you get closer to putting pencil on paper to design your data product, here are a few core design principles to use as reminders of what is important. By no means should you feel compelled to follow all these principles; in fact, it is better to pick one or two high-priority principles to help stay focused.

Compactness and Modularity

Some data products become large and unwieldy in an effort to create a single comprehensive view of an entire business or process. Eric Steven Raymond, writing about good software design, provides this guidance:

Compactness is the property that a design can fit inside a human being's head . . . Compact software tools have all the virtues of physical tools that fit well in the hand. They feel pleasant to use, they don't obtrude themselves between your mind and your work, [and] they make you more productive.¹¹

A data presentation should be broken into bite-sized pieces, each built around a key question, topic, or concept.

Gradual Reveal

Reveal information as the user expresses interest. In other words, don't bombard the user with all the information at once. We frequently use levels of increasing detail from (a) key metric to (b) context around the metric to (c)

full breakout detail for the metric. Figure 5-34 shows an interface for Datran Media's Aperture online advertising dashboard that demonstrates this model:



FIGURE 5-34: This interface reveals more detailed information as the user selects what they are interested in.

Guide Attention

It isn't good enough to make the information available; you need to use visual cues and functionality to draw the user to the things that matter most. A few mechanisms that can help are alerts, positioning on the page, and careful use of color and fonts. The visualization in Figure 5-35 emphasizes those metrics that need attention.

Support Casual Use

Minimize the barrier to entry for new users by avoiding feature overload, minimizing clicks for each task, and providing clear, concise descriptions of what things mean. The data interface we designed for network usage highlights only the most important metrics, uses a parallel structure to delineate between upstream and downstream data, and has a simple visual to show the largest values, as shown in Figure 5-36.

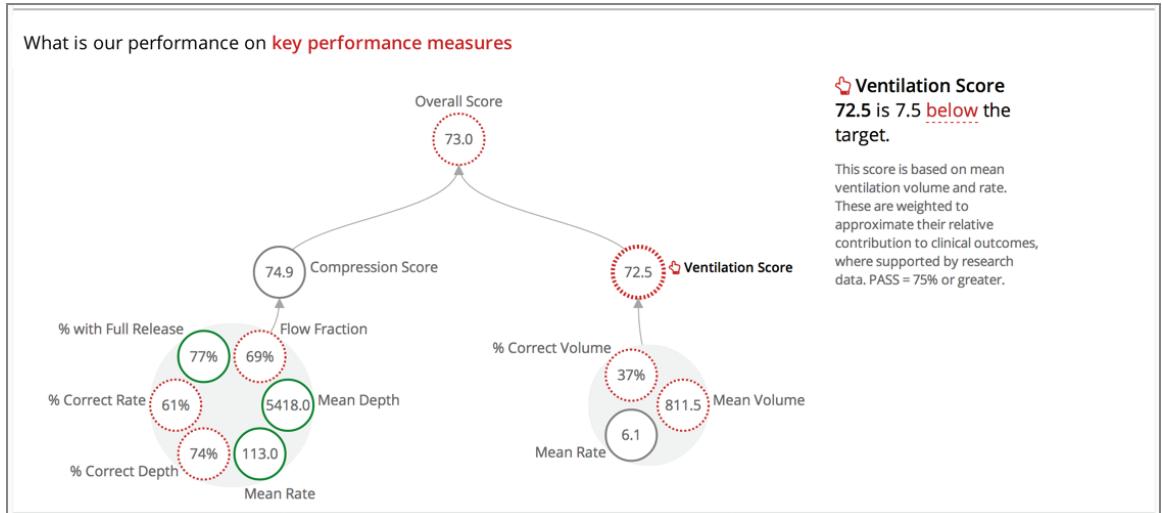


FIGURE 5-35: Colored alerts guide attention to the important information on the screen.



FIGURE 5-36: An executive-level interface to track network utilization

Lead to Action

Empower users to finish their task quickly and understand the action that should be taken based on the results. You can build in explicit guidance about what a change in a metric means or who to contact to address an issue that

is highlighted in the data product. In the interface in Figure 5-37, the primary user action is to build a list of students who need help with their online coursework. The tool helps narrow down the list of all students to the select few who require action.



FIGURE 5-37: This data product provides guidance to suggest specific actions based on the results.

Customizable

Build in flexibility to allow the data presentation to become relevant for different users. The most common way to allow users to customize the dashboard is by defining the scope of the data by using filters. Beyond filters, other types of functionality that enable customization include the ability to save a view of the data that they've configured and features to tag or highlight items that are important to the user.

Explanation before Information

Users need context and explanations to understand new and unfamiliar events. Providing data without these higher-level analyses is the difference between a chef presenting a fine dinner and a fish monger throwing a fish at

your head. For some situations when communicating data, there is sufficient time to interpret the results and provide a summary before presenting to your audience. Letting the data speak for itself can be a recipe for misinterpretation and confusion. The following quote about journalism offers a broader perspective on how to present complex information:

*The most “basic” [journalistic] acts are reporting today’s news and providing current information, as with prices, weather reports and ball scores. We think of “analysis,” “interpretation,” and also “explanation” as higher order acts. They come after the news has been reported, building upon a base of factual information laid down by prior reports [However, there are some stories] **where until I grasp the whole, I am unable to make sense of any part.** Not only am I not a customer for news reports prior to that moment, but the very frequency of the updates alienates me from the providers of those updates because the news stream is adding daily to my feeling of being ill-informed, overwhelmed, out of the loop.*

—Jay Rosen, NYU Journalism Institute¹²

These design principles are some of the fundamental guidelines that can shape your thinking about a data product and help you evaluate whether you've been successful.

VIVA THE AUTHORS OF DATA PRODUCTS

This chapter was about the people in organizations who do the hard work of creating reports, presentations, and dashboards that communicate data. It is a common activity, but a role that is seldom identified explicitly or described with a set of skills and objectives. This chapter explained what it is to be an author of data products, and why this role is essential to the success of a data fluent organization. It is not just that these producers facilitate the flow of information between people. They also add tremendous value to the data they are communicating by analyzing, summarizing, structuring, storytelling, visualizing, and contextualizing.

It takes many diverse skills to be good at designing data products. Although some of these skills were mentioned, they were not described in sufficient depth to make you an expert. More than teaching skills, you were provided with guidance as to what matters the most in the process by focusing on the following:

- Know what your audience wants to know and what they can act on.
- Strive to focus your message to present the right metrics and content.
- Build logic, structure, and flow into your data products so that audiences can more easily digest the information you are delivering.

- Develop your skills in the fundamentals of information display—from little things, such as good use of typography, to bigger elements, such as choosing the right chart to highlight the critical elements in your data.

People are the basis of your organization, and their skills are the basis of a data fluent organization. The past two chapters have focused on essential skills for data consumers and producers of data products. In the next chapter, we turn your attention to the structural and cultural factors that can help them flourish within the framework of your organization.

NOTES

1. George Higgins, American author.
2. Maria Popova, "Six Tips on Writing from John Steinbeck." *Brain Pickings*, n.d. Retrieved from <http://www.brainpickings.org/index.php/2012/03/12/john-steinbeck-six-tips-on-writing/>
3. Maria Popova, "How to Write with Style: Kurt Vonnegut's 8 Keys to the Power of the Written Word." *Brain Pickings*, n.d. Retrieved from <http://www.brainpickings.org/index.php/2013/01/14/how-to-write-with-style-kurt-vonnegut/>
4. See endnote 2.
5. Maria Popova, "Kurt Vonnegut's 8 Tips on How to Write a Great Story." *Brain Pickings*, n.d. Retrieved from <http://www.brainpickings.org/index.php/2012/04/03/kurt-vonnegut-on-writing-stories/>
6. ibid.
7. Maria Popova, "10 Tips on Writing from David Ogilvy." *Brain Pickings*, n.d. Retrieved from <http://www.brainpickings.org/index.php/2012/02/07/david-ogilvy-on-writing/>
8. Stephen Few, "Pervasive Hurdles to Effective Dashboard Design." *Visual Business Intelligence Newsletter*, January 2007. Retrieved from http://www.perceptualedge.com/articles/visual_business_intelligence/pervasive_hurdles_to_dd.pdf
9. Adapted from Emma Coats, "Pixar's 22 Rules of Storytelling." Retrieved from <http://aerogrammestudio.com/2013/03/07/pixars-22-rules-of-storytelling/>
10. <https://twitter.com/lawnrocket>
11. Eric Steven Raymond, "The Art of Unix Programming." Retrieved from <http://catb.org/~esr/writings/taoup/html/ch04s02.html>
12. Jay Rosen, "National Explainer: A Job for Journalists on the Demand Side of News." *Pressthink: Ghost of Democracy in the Media Machine*, August 13, 2008. Retrieved from journalism.nyu.edu/pubzone/weblogs/pressthink/2008/08/13/national_explain.html

CHAPTER 6

The Data Fluent Culture

The only thing of real importance that leaders do is to create and manage culture. If you do not manage culture, it manages you, and you may not even be aware of the extent to which this is happening.

*Edgar Schein
MIT Sloan School of Management*

The aim of this chapter is to guide organizational leaders in the formation of a data fluent culture. Although knowledgeable individuals are necessary for a meaningful exchange of ideas, the culture of an organization is the context that helps these individual efficiently support the mission of the organization. A culture of data engagement and fluency is built upon organizational leadership, shared understandings and conventions about data products, and everyday habits and processes that make data-informed communication a priority.

In this chapter, you learn about:

- How leaders can develop a data fluent organizational culture and communicate priorities
- How to establish key organizational metrics that guide individual and organizational efforts
- How to create a shared language and understanding of data in your organization
- How to incorporate data into everyday activities

It is natural that people in an organization look to their leaders to set expectations, model behaviors, and shape the culture. In fact, culture emerges from the modeled behaviors and reinforced expectations of the leaders. Initially, leadership actions drive culture. Over time, as patterns are established and the culture takes hold, the organizational substrate begins to shape and reinforce employee actions.

This chapter offers guidance to help organizational leaders purposefully shape and sustain a high-performing, data fluent culture. However, before diving into the how, start with the why. What are the benefits of data-focused organizational cultures? Why should leaders invest time in shaping the culture in the workplace?

In *Shaping School Culture: The Heart of Leadership*, Deal and Peterson identify research studies that highlight the powerful effects that a strong, positive, collaborative culture can have on an organization.

- **Culture can foster organizational effectiveness and productivity**—A culture focused on productivity, performance, and improvement helps employees overcome uncertainty in their work (Lorti, 1975) by providing focus, social motivation, and collegiality (Purkey and Smith, 1983; Levine and Lezotte, 1990; Newmann and Associates, 1996).
- **Culture can nurture collaborative activities that foster communication and problem-solving**—Collaborative cultures promote the exchange of ideas, the enhancement and spread of effective practices, and team-based problem solving (Little, 1982; Peterson and Brietkze, 1994).

- **Culture can promote successful change and ongoing improvement efforts**—Cultures that support inertia, apathy, and the status quo will be devoid of innovation, whereas a culture fostering experimentation, risk-taking, and progress will likely enjoy the benefits of continual improvement and advancement (Little, 1982; Louis and Miles, 1990; Deal and Peterson, 1990).
- **Culture can build institutional commitment**—People are motivated and feel committed to organizations that articulate their values and core purposes through a mission that is inspiring and valued (Shein, 1985).
- **Culture can increase the focus of employee efforts on what is important and valued by an organization**—Although job descriptions and policies shape employee efforts, the unstated expectations embedded in the culture have immediate and enduring influence on individual actions and organizational progress (Deal and Kennedy, 1982; Schein, 1985).

Although each of these concepts applies generally to organizational culture, they also offer a rationale and incentive for building a data fluent culture.

LEADERSHIP, CULTURE, AND COMMUNICATING PRIORITIES

Within every industry sector, a host of leadership skills are necessary to meet the specific needs of the organization. Yet, the essence of effective leadership has common practices regardless of the venue. Applying these practices toward the development of a data fluent organization can be summarized as follows:

1. **Set the stage.**
 - **Create a clear vision**—Leaders must create a clear vision for the importance of data and how it will be leveraged to increase productivity within their organization.
 - **Articulate clear expectations**— Leaders must set and clearly communicate high expectations for data engagement and fluency within their organization.

2. Train the team.

- **Train and support**—All employees must receive ongoing, targeted training to ensure they have the knowledge and skills to meet expectations and help the organization achieve the vision.
- **Hold people accountable**—Leaders hold all team members accountable for meeting goals and provide actionable feedback on their performance.

3. Promote growth.

- **Encourage curiosity**—Although there is need for accountability for results and a rationale for the use of resources, leaders must counterbalance these demands, both directly and indirectly, through established cultural norms with a healthy promotion of inquisitiveness and innovation.
- **Accelerate informed decisions**—Leaders encourage decisions informed by data and develop a culture that values the process of planning, testing, analyses, and action.

The ability of your organization to use data hinges on developing a team of people who share a common vocabulary and skillset to understand data. The steps above provide a broad roadmap for developing such a team. The rest of this chapter provides insights and examples about how organizational leaders go about these important tasks.

SET AND COMMUNICATE EXPECTATIONS

Data fluent leaders must lead by example. They should express expectations for quality data products and then use data products to support the organization’s mission. To build a data fluent culture, leaders must communicate using data to support their decisions and organizational priorities. By doing so, they set the standard for quality data products and demonstrate their data literacy in public forums, modeling the expected behaviors of their team.

A senior executive at a global beverage brand was asked to present to the Board of Directors to explain the impact of his function—customer service—on the bottom line. Faced with a room packed full of heavy-hitters, he chose an unusual but effective approach. Rather than stand in front of the group with a series of bland PowerPoint charts—like so many of the previous presenters—he

turned the tables by asking the board members to stand. He arranged them in a line and asked them to imagine they were ordering at a fast food restaurant. Then he brought the data to life by demonstrating the cost of wait times in line, the profitability of each individual fast food customer, and the impact of a poor product experience. When a customer left the line because of slow service, it might cost the company only a dollar in profit, but it wasn't hard for people in the room to extrapolate a single data point by the many thousands of fast food locations. The Board quickly came to understand the data behind the vague concept of customer service in a way it never would have through rote presentation of data in slides.

ARTICULATE SPECIFIC, MEASUREABLE INDICATORS

The language of data has nuanced dialects that vary from organization to organization and industry to industry. The dialect reflects the organization's geography, industry, history, and circumstances. This organization-specific data language results in unique meanings associated with data fields and metrics. A concept as simple as *customer satisfaction* can have a wide-ranging meaning. The definition is directly linked to the organization's approach to collecting the data, expectations of what *satisfied* means, required time period for data gathering, and sample size. A data fluent culture ensures everyone in the organization has a common understanding of exactly what is meant by a term like customer satisfaction.

Furthermore, the benefits of data diminish when people spend more time debating terminology, specifics of calculations, and validity of sources rather than discussing what action to take based on the results. This is not to suggest that validity and reliability and precise calculations are not important topics. Yet, when these definitional questions linger in an organization, productive discussions are derailed in meeting after meeting.

Amazon.com is well known for creating small teams that set out to solve specific business problems. Aside from the cross-functional, cross-divisional nature of these teams, the most innovative aspect of this model has been the use of a single fitness function to measure team performance. For Amazon.com, a *fitness function* is a performance metric that directly reflects the goals of the team and is approved by senior leadership. The choice of the fitness function—and how it is calculated—is critically important because it represents the driving objective for the team. It creates team alignment and accountability on a clear goal.

CELEBRATE EFFECTIVE DATA USE AND PRODUCTS

Everyone watches leaders in an organization. What is appreciated, recognized, and celebrated by leaders signals the key values of the organization. In addition, these celebrations of work products provide an opportunity to demonstrate that high standards and performance are achievable. Leaders can celebrate data products by using them as a centerpiece of communication with the organization or by putting resources behind the people and teams that create the most useful, effective data products.

We've worked with many executives who have come to understand that attractive presentation of data grabs attention and engages audiences. In some cases, a single visualization can serve as a focal point to understand the business or describe opportunities or issues. For a leading health insurance provider, we used a treemap visualization (Figure 6-1) as the mechanism to understand the distribution and drivers of costs. By summarizing the full picture of benefits claims in a single graphic, the leadership team had an easy way to see what was driving patient expense overall and what components were increasing over time. Moreover, the leadership team signaled to the company the value of synthesizing core business data into an easy-to-digest format.

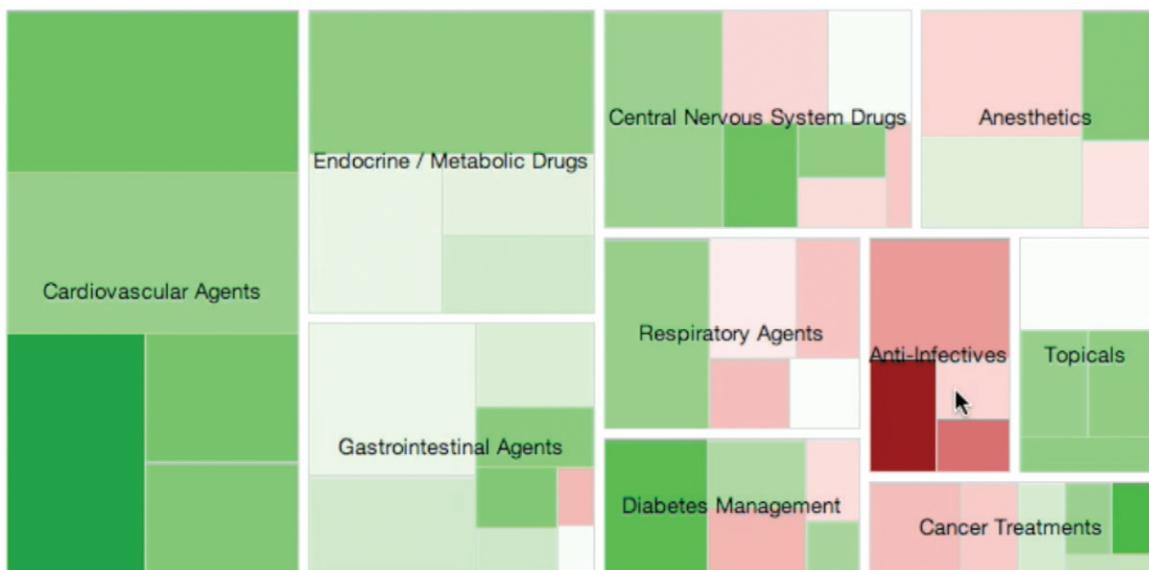


FIGURE 6-1 A treemap visualization allowed healthcare executives to focus on the drivers of costs.

Use Data to Inform Decisions and Actions

Leaders and employees in a data fluent culture bring well-understood key metrics into meetings, understand how to measure the performance of a new project or product, and include data fluency skills in the hiring and employee evaluation processes.

People are the heart, soul, and horsepower of an organization. If a leader fails to recruit, develop, and retain the right people for a company, the organization's future is in jeopardy. One of the core tenants of Jim Collins' widely cited framework in his book *Good to Great* is to start first with the *who* and then move to the *what*. High-performing organizations have the right people on the bus and in the right positions.

There are several ways that leaders seeking to develop a culture of data fluency can implement this goal. To begin, leaders must ensure that the organization uses data to inform the recruitment and hiring processes. Getting the right people on the bus means first knowing what types of expertise and skills the organization needs. This question can be answered through proper use and understanding of data during the selection process. Leaders can ensure that prospective employees have the aptitude and attitude aligned with the data fluency demands of the organization. In the forthcoming Feature, "The Data Fluency Inventory (DFI Supporting Materials"), we have included a sample data literacy quiz for this specific purpose. When hired, leaders must ensure employees receive the ongoing support and development needed to meet existing and emergent requirements for being skilled consumers and authors of data products. Finally, when allocating employees to specific teams or tasks, a data fluent leader can draw upon evidence to support placements and have clear metrics for assessing the quality of such decisions.

Google's innovative use of data has served as a case study for bringing analytics to the Human Resources function. Whereas traditionally HR has been built on relationships, Google's goal is to "bring the same level of rigor to people-decisions that we do to engineering decisions."¹ A dedicated people analytics group is tasked with gathering, analyzing, and informing decisions by identifying the measurable characteristics of outstanding leaders, improving employee retention by predicting at-risk staff, and creating data models for the factors in hiring that are most likely to lead to successful employees.

This type of advanced analytics may not be well suited to every organization, but it does offer ideas for more deeply incorporating data into the process of managing people. In addition to any direct benefits from using data to improve

HR decisions, by incorporating analytics into the HR process, leaders can also set a tone and expectation for data fluency, even in areas of the organization that do not typically use data.

When you have the right people and leadership priorities in place, the next challenge is to define the right metrics to focus the organization on.

ESTABLISHING KEY METRICS TO RALLY AROUND

Can you imagine a basketball game without a hoop or a tennis match without sidelines? In the absence of clear goals and a way to measure progress, organized sports would be . . . disorganized. The same is true for organizations of all types—in the absence of clearly communicated metrics of organizational success, employees have no way to know if they are on track, contributing in meaningful ways, or merely spinning their wheels. To create a culture of data fluency, leaders must establish key organizational metrics that guide individual and organization efforts. This section first discusses the characteristics of a good metric and then describes specific ways that leaders can leverage metrics to drive organizational improvement.

WHAT MAKES A GOOD METRIC?

Misaligned goals, distorted behaviors, and a misguided sense of success . . . no, we're not referring to college graduates or politicians. We're describing the problems caused by using poor metrics in your organization. One common example is the customer service department that tracks calls handled without regard for the quality of the caller experience. The first step in establishing what makes a good metric is to understand what you want to achieve.

Metrics can serve two purposes: identifying problems and measuring performance. When the goal is to identify problems and pinpoint areas of operational inefficiency and ineffectiveness, defining the right metric requires a bit of detective work. It requires you to uncover the data residue of a problem and to determine what evidence can be found and how exactly it shows up. When the goal is to measure performance, the right success metrics focus on measures that can be controlled and where improvement in the metric is an unambiguously good thing.

After you have clarity on the purpose for which you will use a particular metric, the next step is to choose a good metric. Good metrics include several specific dimensions that are shown in Figure 6-2.

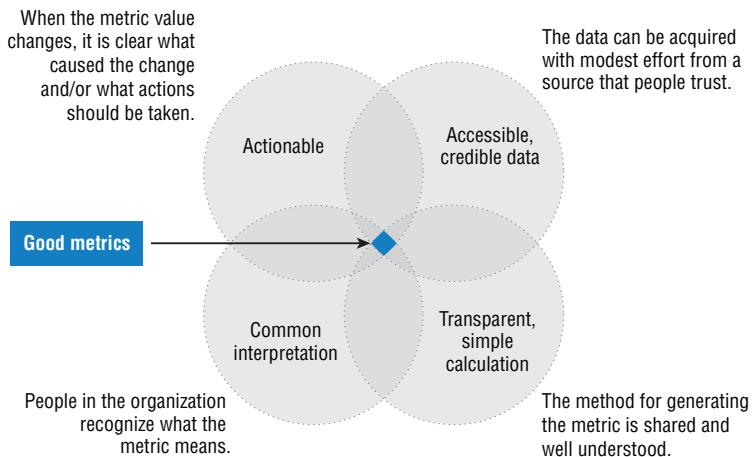


FIGURE 6-2 Effective metrics are actionable, credible, transparent, and have a shared interpretation across your organization.

- **Good metrics are created from accessible, credible data that is easy to gather from a trusted source.** Every metric takes time and effort to construct. Before choosing the metric, it is worth asking questions of the data source: Is the data collected in a way that accurately captures reality? Is the data difficult to gather or delayed in a way that will undermine its value? Will consumers find reason to question the data source? For the results to have validity and for a culture of data engagement to emerge within your organization, data producers must have access to credible, trusted data.
- **Good metrics are transparent and generated by understandable calculations.** For the metrics identified by an organization to shape individual and collective practice, they must be easily understood, affording transparency for how they were developed. Further, as Andrew Jaquith, author of *Security Metrics* points out, good metrics are best expressed as a number or percentage. And by number, he specifies a *cardinal number*—something that specifies how many of something there are—rather than an *ordinal number* that denotes

something's position relative to other things. He goes on to add that the numbers should also contain at least one unit of measure that characterizes the thing being counted. The units of measure enable data analysts to consistently express results of a measurement process. Using two measures can assist this process. For example, instead of just reporting "number of new sales accounts opened," a second unit of measure such as "per 1,000 calls" enables the constructed metric to be used for benchmarking.

- **Good metrics have a common interpretation and are consistently measured.** It is imperative in an organization that all people in the organization recognize what the metrics mean and have assurance that the metrics are consistently generated. To facilitate this, leaders should ensure that measurement processes are clearly documented and that employees share a common set of terminology to understand the meaning of data presentations. A common issue in sales organizations is the definition of a *lead*. What level of interest does a lead need to have demonstrated? How ready and qualified does the person need to be in order to be labeled as a lead? These definitional issues—if not dealt with consistently in data capture and discussions—can lead to a lot of confusion about the meaning of any key metrics based on data about leads.
- **Good metrics are contextually specific and actionable.** Good metrics hold meaning for members of an organization and signal specific actions that can be taken. They shed light on an underperforming part of an organization, chronicle the results of continuous improvement efforts, or demonstrate the value of specific people and processes within the organization (Jaquith, 2007).

The table below provides a set of performance metrics that are representative of the four dimensions of good metrics. They are all composed of data that is often accessible and easy to gather; the calculations are transparent and easily understood; they can all be expressed in numbers and units of measurement; and they offer contextual detail that can be actionable. Having described the characteristics of a good metric, we now turn to a discussion of the ways that leaders can leverage metrics to promote organizational performance and build data fluent cultures.

SECTOR	METRIC
Freight	<ul style="list-style-type: none"> ▶ Freight cost per mile (total spent on truck or train carriage divided by miles) ▶ Load factor (percentage utilization of total capacity) ▶ Empty miles (percentage of total miles that carried no inventory)
Warehousing	<ul style="list-style-type: none"> ▶ Cost per square foot (total warehouse operating costs divided by size) ▶ Inventory turns (total cost of goods sold annually divided by average inventory value for the period)
Education K-12	<ul style="list-style-type: none"> ▶ Change in test scores correlated with an increase in per teacher expenditure on professional development; number of graduating high school students matriculating in a 4-year college correlated with hours of guidance counselling
Marketing	<ul style="list-style-type: none"> ▶ Customer Acquisition Cost (total sales and marketing cost divided by the number of new customers within a specific time period) ▶ Return on Marketing Investment (New sales from a marketing program minus cost of the marketing program, divided by cost of the marketing program)
E-commerce	<ul style="list-style-type: none"> ▶ Website conversion rate (percentage of unique visitors to the website who make a purchase)
Cable	<ul style="list-style-type: none"> ▶ Subscription costs to acquire (cost of all marketing, subsidies, and discounts divided by the total number of acquired customers) ▶ Average revenue per user (total revenue divided by total number of users)

Source: Modification and expansion of figure from Janquith, 2007

USING METRICS TO DRIVE ORGANIZATIONAL IMPROVEMENT

Have you seen that movie scene where the weathered old cowboy tames the impossibly wild stallion? The wise frontiersman takes his time, careful not to frighten the horse, and gradually shows it what he expects. He uses a soft voice and doesn't ask too much at first. After he has established trust, it isn't long before he hops on the stallion and lets it do what it wants to do naturally: gallop into the sunset. There are both leadership and metric lessons in this scene.

Choose a Few Key Metrics at Any Given Level

Too many organizations end up with dozens of “key” metrics. In actuality, these organizations typically don’t know what to focus on. Like the wise cowboy, leaders must invest time and energy to socialize their organization to a small set of essential metrics. The challenge is to offer a clear definition of what is being measured and demonstrate its importance. It is not imperative to initiate discussions about goals at this point. Like the wild mustang, you will probably be dealing with a jittery cast of characters (no sudden movements or loud noises). After a set of three to five metrics have been established, incorporate the metrics in meetings and other venues. You want these metrics to become part of the organizational vernacular. Gradually, people will become accustomed to the metrics and their implications about the business.

Select Key Metrics That Align with the Mission and Vision

Like the stallion, a transformation can occur in organizations that learn what to focus on. With the awareness and understanding of a small set of key metrics, people begin to wonder what goals are associated with those metrics. It is a natural progression. In many cases, employees will ask for goals as if you as the leader had never considered the concept. Let them own the idea.

Goals associated with a few, well-understood key metrics is a powerful combination. For both internal and external stakeholders, there is a strong alignment between organization mission, vision, goals, and tracking of progress. The efforts of everyone can be directed at these measurable goals, and people will focus on the processes that can impact these metrics.

We worked with a software vendor that had a crystal clear understanding of what it took to grow its business. For the vendor, it was all about driving new leads into its efficiently designed sales machine. Therefore, measures related to product development, product usage, or customer satisfaction were all subservient to the cost per lead, volume of leads, and various conversion rate measures through their sales funnel. The company could spend on marketing campaigns and have a precise expectation for the revenues that would result.

SHOW EMPLOYEES THAT THEIR CONTRIBUTIONS ARE ESSENTIAL

Truly exceptional leaders not only articulate organizational goals and identify metrics for assessing operational progress, they also can help employees understand how their work contributions are an essential component of the

broader goals and measures of progress. Metrics should have a hierarchical connection that allows the front-line staff to see how their measures link up to the key metrics for the whole organization. In nature, form follows function, and we have countless examples of alignment between physical attributes and environmental conditions. Similarly, the more individuals have a clear line of sight between their daily contributions and the organization's priorities, the better able (and willing) they are to make smart decisions and commit their energy in ways that sustained overall organizational performance.

Reference Key Metrics and Data Analysis When Communicating Goals

Data fluent leaders lead by example. Accordingly, leaders must model reliance upon data analysis and reference key metrics when communicating goals and priorities to stakeholders. To build a data fluent culture, leaders must communicate in the language of their metrics, using data to support their decisions about organizational progress. By doing so, they reinforce the importance of the selected key metrics and demonstrate their data literacy in public forums, thus modeling wanted habits of mind.

Avoiding Metrics Pitfalls

Having spent a bit of time focused on how leaders can use metrics to benefit their organizations, this section concludes by providing a few warnings about potential pitfalls in the use of metrics.

Although selecting the right metric can foster success, the wrong metric can create unexpected, misguided, and counter-productive behaviors. The wrong metrics can create incentive structures for behaviors that actually work against company goals. For example, one school superintendent—under pressure to meet metric goals set by the No Child Left Behind Act—developed a scheme to keep low-performing students out of classrooms. By removing these students from the classroom, the schools achieved the target thresholds and received the funding and bonuses the school administrators craved.² You get what you measure.

And then there is the example of a former VP for marketing for an airline company as she was traveling on her own airline. While rushing to catch her flight, she arrived late to the gate and found the employees were intent on closing the door to the airplane 10 minutes in advance of the flight. At this particular airline, success is measured by the percentage of flights that push-back from the gate before the scheduled time. The employees, her own colleagues, were perfectly willing to slam the door in her face to meet their success targets.

Push-back time was probably considered something employees could control and reasonably correlated with on-time arrivals. In addition, push-back time is straightforward to measure—unlike the seething anger of a customer trying to get on a plane that is still sitting at the end of the jetway.

Most companies make the mistake of evaluating performance (establishing success metrics) along functional activities. This perpetuates disconnected, dysfunctional silos within an organization. Marketing becomes focused on driving revenue. Operations focus on driving down costs. Product development is a success if new features appear on time. The results of these incentives are predictable: Marketing, which is focused on new customer volume, brings in customers who have low life-time value; operations, seeking to minimize costs, squeezes customer service at the price of customer satisfaction; and product development, focused purely on the clock, doesn't bother to deliver the features that can help marketing bring in new customers.

One alternative is to put customer experiences or processes at the forefront by defining success metrics that cut across organizational structures. No matter what industry you work in, happy and high-value customers are at the center of your businesses success—but require multiple functions to work together. Customer-centric metrics ensure you are aligning your organizational processes to this wanted goal. A commonly accepted metric with strong leadership support and real accountability can be a catalyst to get these formerly unconnected groups working together.

Along these lines, consider the following when deciding on appropriate metrics:

- **Trending, distributions, and comparison**—A single, lonely number often doesn't reveal a true understanding of an organization's performance. It is necessary to put a concept around the value to allow people to gain a more complete understanding of the relative goodness or badness. By showing trending, distribution, and comparison, you improve understanding for the audience of the metric. Trending shows whether the metric moves closer or further from a goal. Comparisons to industry benchmarks show how your organization stacks up in a larger context. Finally, break out the value by different dimensions (for example, locations, products, and customer segments) to reveal whether one factor is propping up or dragging down the metric value.

- **Edge cases**—There will always be edge cases, or outliers, where a metric may not mean what you think it means. These situations are worth understanding, but you shouldn't allow the perfect to be the enemy of the good. As a leader, you need to weigh the benefits of choosing metrics that work for 90 percent, 95 percent, or 99 percent of cases with the costs of those incremental gains.
- **Accountability test**—Could you hold someone accountable for this metric without them offering a dozen reasons why it doesn't make sense? If not, you may need to reconsider the validity and value of the metric. This simple thought exercise is a decent test of the value of the metric.
- **Self-serving**—Be careful that you don't select metrics simply because you know they'll make you look good. These short-term victories have a way of incrementally turning into a losing long-term strategy for organizational competitiveness and success.
- **Letting go**—Putting a metric out to pasture, especially within the context of a large, multilayered, complex organization, is a hard thing to do. There are a few reasons this happens:
 - a) The metric was developed at great effort and high expense.
 - b) After the process of collecting the data for the metric stops, it can be restarted only at great effort and high expense.
 - c) People higher in the organization (who last paid attention to the metric when it was useful) might come looking for it if something goes awry.

SHARED UNDERSTANDINGS

A culture of data fluency needs to be built on a shared understanding of the data sources, data analysis, key metrics, and data products. It requires employees to be on the same page about how data is used and why it is important.

In general, the concept of shared understanding is a fundamental building block for organizational success. It allows everyone to be working toward the same goals, have a common set of principles, and be in alignment about

the organizational vision. Creating a shared understanding allows employees to act with autonomy and creativity because they can align their actions to the established mission of the organization's leaders, as shown in Figure 6-3.

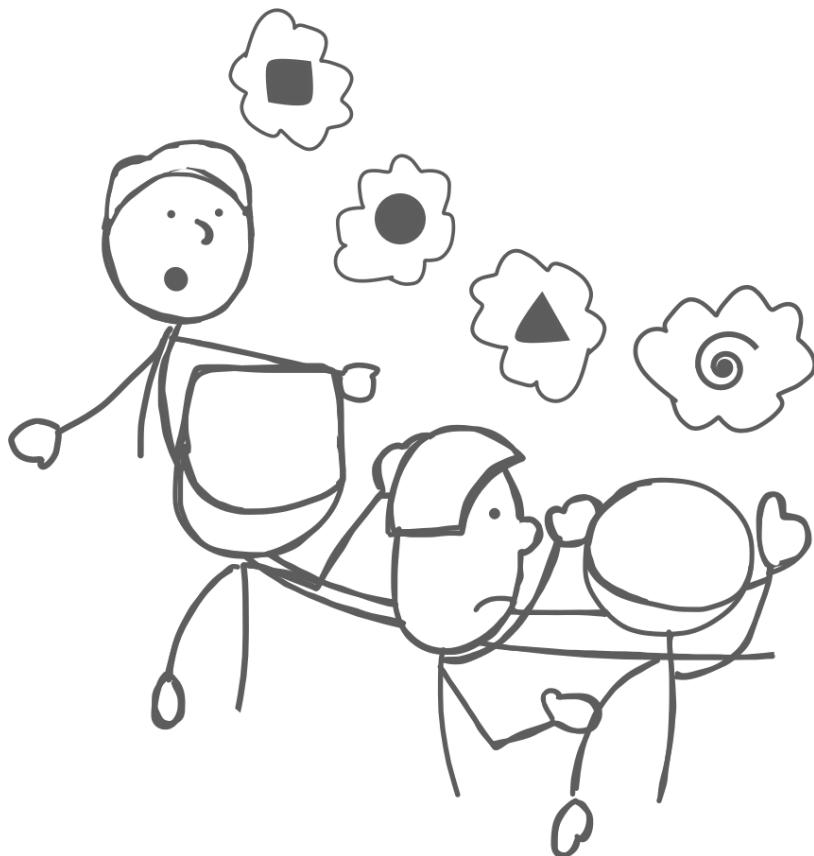


FIGURE 6-3 It can be difficult to get everyone on the same page.

Within the context of data fluency, shared understanding is based on people knowing what data means and how it is used. The previous section touched on the importance of clearly defining key metrics, the challenges that arise when people have different definitions for data, and how these differences can derail progress. Following are the primary areas in which shared understanding can help foster a data fluent culture:

- Using consistent vocabulary and terminology for organization-specific data
- Demonstrating the credibility and reliability of the data sources

- Creating transparency into how data is manipulated and analyzed
- Establishing a shared set of key metrics and explaining how these metrics link to organizational goals
- Building agreement on the purpose and motivation for data products

Each of these objectives represents a step toward a data fluent organization. Leaders play a critical role in implementing these steps in building shared understanding and data fluency organization-wide.

COMMON VOCABULARY AND TERMINOLOGY RELATING TO ORGANIZATION-SPECIFIC DATA

Every organization has data that is specific to its organization and industry. Describing, defining, and regularly using vocabulary and terminology over time can create a data jargon. Jargon has many purposes including legitimizing a profession, keeping nonexperts out, socializing new employees, and facilitating communication. A rich data jargon can facilitate communication. By defining the measures that decide whether success is achieved, leaders help an organization develop a common vocabulary and terminology.

Clear Definitions of Measures

As discussed in the previous section, organizations as a whole, as well as policies and procedures, should include a written, precise definition for data elements that are central to the organizational culture. One way to handle this is to craft a data dictionary that defines all measures and categories within measures. It is helpful to customize definitions to the particular circumstances of specific programs and to include examples of how to resolve ambiguities. This should result in clear definitions for all measures.

Standard Forms for Collecting Data

If a group of researchers interview different people, each researcher uses the same instrument. In this way the interview is conducted using the same vocabulary and terminology with each interviewee. Likewise, organizations should use standard forms for data collection that include all the data elements and categories that

Note A data dictionary is a great onboarding tool. New employees are often swimming in alphabet soup upon arrival. Take, for example, a new sales representative and junior accountant fresh out of college learning the difference between NOV (New Order Value), TOV (Total Order Value), and revenue. What these terms mean for a new sales representative and what these terms mean for a new accountant might be different. A sales person might assume that NOV resulting from a sale equals revenue for that year. Because the contract is for three years, however, revenue may be spread over three years. NOV, however, is probably recorded just in the year of the sale.

are referenced in the database system. Employees record information onto data forms following specified procedures and then key the information from these forms into the program database. Employees should not need nor be allowed to enter their own codes or variables because this causes data entry errors and hurts reliability and validity.

Develop a common understanding of data fields. Leaders should insist on a shared understanding of data fields. You can't rally around key metrics if people have different interpretations of data fields that are keeping track of those metrics.

THE CHALLENGE OF EMERGING NEW HEALTHCARE FACILITY TYPES— DEFINITIONAL DIFFERENCES AND DATA

One source of miscommunication at a large insurance company was the definitional difference between *inpatient* (IP) versus *outpatient* and how data was aggregated for these two categories. It seems simple: Inpatient care happens in a hospital and outpatient care happens outside of a hospital. With the goal of brainstorming how to decrease the number of IP ACUTE episodes by high-risk members (members with diabetes, cardiovascular disease, and chronic kidney disease), the meeting devolved quickly. Team members realized that IP ACUTE included long-term acute care hospitals (LTACHs) and skilled nursing facilities (SNFs). Having a member at an SNF or LTACH as compared to a traditional hospital, however, had different cost implications. A meeting objective of trying to figure out what member characteristics and factors would be predictive of an inpatient episode devolved into a definitional discussion of what was an inpatient versus what was an outpatient. As a result of the confusion, the team defined inpatient acute as a traditional hospital episode and outpatient as a term was discarded in favor of a breakout of non-IP ACUTE possibilities. These outcomes included LTACHs, SNFs, ALFs, and home health. Outpatient care as defined as ambulatory was left for another day.

Understanding the origin of a word or term can help provide deeper understanding into the everyday language of an organization. The sources of data

provide the etymology of data and create a deeper shared understanding among team members who understand the root words of data.

Definition Data is the plural form of the word datum, which in Latin means “something given.”

UNDERSTANDING AND APPRECIATING CREDIBLE, RELIABLE DATA SOURCES

"Consider the source," is a warning with much relevance for modern organizations. When making decisions or forming opinions, thoughtful people consider the source of information and the methods with which it was collected. Data sources carry a reputation, like people.

Outside data sources can have a reputation for credibility. The Dartmouth Atlas of Health Care is an academic source that has built a strong reputation for gathering geographically based information about the use of health resources. For measuring online audiences, comScore has emerged as the standard for advertisers and agencies when negotiating ad spend. In other sectors, government sources like the Bureau of Labor Statistics and Department of Education have established a role in delivering credible data. In part, these reputations are developed because these data sources have established protocols and processes to vet and verify sources.

Sources of internal organizational data also develop a reputation over time. Many times we've worked with organizations that have one database that is considered the trusted source of truth, whereas another data set seems to capture similar information but is known to be faulty. The difference between these data sources is based on how the data is collected, stored, accessed and perceived. Reputation of data sources is also influenced by employee, team, department, organization and industry biases, whether warranted or not. A data fluent organization needs to create broad awareness of these issues to guide analyses toward the data sources that are most credible.

Understand the Strengths and Weaknesses of Data Sources

Data, a piece of information in and of itself, has limited utility. First, data fluent members of an organization must understand the importance of reliable data, or what is referred to as *internal validity*. Next, to make inferences about the data to a larger population, we move into the terrain of generalizability, or *external validity*. A leader has to know whether the data collected will be generalizable to other situations. Do the processes or protocols used in one situation apply to another situation, or are the populations quite different? Leaders, over time, often develop the skill to determine if the data collected will be generalizable to other markets, regions, sectors, populations, or situations. Following are a set of conditions that impact the reliability and generalizability of data. For each condition, data fluent organizations ask themselves if there is any way that the condition can explain the story that we would otherwise think the data tells.

RELIABILITY	
Selection	Consider subject-related variables that may impact how people respond to interventions.
History	Consider anything during the time frame of the intervention that may impact results.
Maturation	Consider that people change and grow. What may look like a result of an intervention is purely participant maturation.
Selection-Maturation Interaction	Consider that different populations will mature at different rates.
Repeated testing	Consider that prior exposure to the intervention may affect results of subsequent measures. For example, people can become test-wise and are thus less affected by the intervention.
Instrumentation	Consider the reliability of the instrument used to gather the data or the precision of the researcher who takes the measurements. Inter-rater reliability helps with this.
Regression to the mean	Consider that extreme cases are likely to move in one direction over the course of the pre-post test sequence . . . always toward the average.
Experimenter Bias	Consider the reality that expectations and investment in a project will impact the outcome of a study.
GENERALIZABILITY	
Pretest	Consider that a pretest is artificial when moving outside to the broader population. When we take the “treatment” to the external population, there is no pretest. Here, the pretest is part of the treatment.
Subject selection	Consider the impact of this sampling issue. Even if a sample of people is randomly assigned within an experiment, if they don’t represent the larger community in terms of demographics, the experiment is not generalizable.
Reactive effects/the Hawthorne effect	Consider that the attention the subjects receive in the experiment is impacting the outcome—this won’t happen in the larger community—therefore an implementation dip will occur.

PROVIDE TRANSPARENCY INTO HOW DATA IS MANIPULATED AND MODELED

If you were fed the same stew every day, you'd want to know what the cook is putting in the pot. It's the same for a data product. Knowing how source data is analyzed, manipulated, and modeled helps users trust what they see. By contrast, performing complex manipulation in a black box leads to many of the unproductive debates previously mentioned. Ultimately, trust is a core element of shared understanding.

Any presentation of data, whether a simple calculated metric or a complex predictive model, is going to have a set of assumptions and choices that the producer has made to get to the output. The more that these can be made explicit, the more the audience of the data will be open to accepting the message offered by the presenter.

One example of this issue is what has happened with a complex metric that ESPN, "the worldwide leader in sports," developed to measure NFL quarterback performance. In 2011, ESPN created a measure called Total Quarterback Rating (QBR) that competed with the commonly accepted (though equally arcane) NFL Passer Rating. QBR represented an "improved" way to evaluate quarterbacks, in part because it attempted to evaluate how the quarterback contributes to scoring points and helping his team win. However, reception has been frosty. ESPN has been reluctant to share the precise approach for calculating QBR—opting instead to assert that it is extremely complex.³ The lack of transparency resulted in this measure not being adopted by fans. The problems faced by ESPN are explained by Michael David Smith of Profootballtalk.com:⁴

One of the aspects of Total QBR that could be both a strength and a drawback is that it considers data that the average fan doesn't have access to, like how far a pass travels in the air, and whether the quarterback was under pressure when he threw it . . . it means fans can't see for themselves exactly where Total QBR comes from—fans just have to trust that the distance the ball traveled was correctly measured, and how much pressure the quarterback felt on the play was correctly assessed . . . If ESPN is committed to this stat and is able to clearly and concisely explain it on the Worldwide Leader's NFL broadcasts, then fans will quickly become familiar with it and it will soon become a staple of how we talk about quarterbacks. On the other hand, if the stat comes across as too convoluted—or if it doesn't really seem like much of an improvement on the current passer rating—then this will all feel like a rather pointless exercise.

Closer to home, many organizations struggle with shared understanding when individual employees have proprietary spreadsheets for delivering important outputs. Think of the marketing analyst who has developed a complex spreadsheet for consolidating marketing campaigns or the financial analyst with the perfect product profitability model. These single-owner analyses are opaque for outsiders, and therefore hard for everyone to fully buy-in to the data that comes out. It is incumbent on the spreadsheet owner to explain the assumptions and communicate how the analytical model works to gain trust.

The same holds true for something as simple as the charts displayed in a report or dashboard. Presenting the data isn't enough. It is imperative that the context around the data is also included. What filters have been applied to the data? What values are shown and how are they defined? This emphasis on providing transparency into the underlying data—whether in a chart, complex data model, or calculated metric—will build credibility with the consumers of data and shared understanding of the data's meaning.

DEFINE A SHARED SET OF KEY METRICS

The importance of choosing a limited set of well-defined key metrics was previously discussed in this chapter. Ideally, everyone in an organization understands the critical measures that link directly to organizational success. With this kind of shared understanding, people have a common language to support productive conversations and a reference point for decision-making.

The counterpoint is the organization that has allowed key metrics to proliferate. This is a persistent problem in healthcare. There are many ways to evaluate healthcare performance—from operational to clinical to financial. In each category there are dozens of ways to measure performance across hundreds of conditions and health departments. The result is that individuals rarely have the time or capacity to wrap their minds around all the data. And the metrics themselves lose their value as decision-making tools.

WHAT CAN YOUR METRIC MEASURE? A SHIFT TO PROACTIVITY

Collecting customer service metrics is a challenge for many organizations. Consider a call center supporting hospitals across the country for education and training. With the stated organizational goal of becoming customer-centric, the call center has a short telephonic survey after every "ticket," representing an issue or challenge needing recovery. Collecting data from in-bound calls after each ticket is resolved seems like a great way to measure customer service. However, it doesn't tell the whole story. Achieving high recovery rates might show the skill level of the call center employees (the client representatives or success managers); however, if the vision is to create a client-centric company, the data collected might be a sample of convenience. With better engagement the incoming calls might have never come in!

Progressive call centers often implement a policy where success managers call out to clients to see how they are doing without an issue needing resolution. Another strategy employed by some organizations is the embedding of high-skilled engineers or other technical skilled persons in the call center so that more complex in-bound issues can be resolved in real time rather than being referred to another team for handling at a later time. Measuring "customer engagement" as the scores realized on recovery tickets as opposed to engagement of customers regardless of recovery tickets would result in very different conversations between your organization and perspective clients.

UNDERSTANDING THE PURPOSE AND MOTIVATION FOR DATA PRODUCTS

The endpoint for data sources is, of course, the data products presented to a target audience. Establishing a shared understanding for data products also requires building credibility and transparency for the consumer. Essentially, people want to know why the data is shown and what they are expected to learn from the data. Cynically, data consumers may wonder whether there is an ulterior motive to how the data is presented. They want to trust the asserted results. A more open-minded person might simply want to know what the organization hopes to achieve through the data product.

For example, we have worked with several services companies who need to deliver periodic reports to their large corporate clients. In many cases, these reports are unproductive because the presenter struggles to articulate the

purpose of the discussion. Without a clear sense of what the report should achieve, the client discussion becomes meandering, overbroad, and inevitably a download of disconnected facts and figures. The audience is left to wonder whether the report is simply intended to confuse them, distract from an issue, or deliver a reference document.

Similarly, you may have seen the ever-popular corporate dashboard that graces a large screen TV in the lobby of company headquarters. Numbers flash colorfully across the screen and a glossy pie chart shows revenue breakdown in the last quarter. Who is the intended audience of this dashboard and how are they supposed to be moved to action? It is unlikely the lobby dashboard has a purpose. And if it does, there is an even smaller chance the viewers will know the purpose.

It is difficult to get everyone in an organization on the same page about data. Data sources are complex, as are the analyses and data products that result. Nevertheless, leadership in a data fluent culture requires diligence to define a shared understanding about the meaning of data. Trusted data sources, transparent data analyses, and purposeful data products can lead to an organization that uses data together as a productive tool.

GUIDESTAR.ORG AND NONPROFIT TRANSPARENCY

Guidestar.org is an example of an organization that has online data products with a clear purpose and motivation. This organization's purpose is to shed light on nearly all nonprofit organizations registered in the United States. Whether interested in a foundation like the Bill & Melinda Gates Foundation or an organization like Habitat for Humanity, one can review different organizational metrics that are publically available. Raw data is also available in an organization's posted IRS 990 form. Want to know all the organizations that a foundation supported last year and the amount they gave to each, review the foundation's 990. Want to know how much money the executives made at a nonprofit, review its 990. Want to know the assets of a nonprofit, review its 990. Such raw data is important when writing grants, evaluating the health of your favorite charity, or deciding how much to negotiate for when applying for a job at a not-for-profit. Guidestar.org gives users the possibility of quick summarized data or deep raw data.

This next section of the chapter focuses on another critical element of a high-performing, data fluent culture—how leaders utilize a set of everyday activities to support data consumers and producers in effective data use in their organizations.

EVERDAY ACTIVITIES

In a recent conference presentation, Jason Fried described a typical workday:

People go to work, and they're basically trading in their work day for a series of "work moments." That's what happens at the office. . . . It's like the front door of the office is like a Cuisinart, and you walk in and your day is shredded to bits, because you have 15 minutes here and 30 minutes there, and then something else happens and you're pulled off your work and you've got to do something else, then you have 20 minutes, then it's lunch. Then you have something else to do. Then you've got 15 minutes, and someone pulls you aside and asks you this question, and before you know it, it's 5 pm, and you look back on the day, and you realize that you didn't get anything done.

TED: Why Work Doesn't Happen at Work

He later attributes these ineffective work days at the office to M&Ms—managers and meetings. Managers “whose job is to interrupt people” and meetings that breed more meetings, an incestuous cycle that has you spending your time “[talking] about things you’re supposed to be doing later,” rather than doing them. Added to the M&Ms are a never-ending stream of electronic distractions—e-mails, tweets, social media interruptions—all asking for your immediate attention.

As a leader, you are no stranger to Fried’s insights. In fact, 15 minutes here or 30 minutes there may seem like a luxury you no longer have. Nonetheless, you also know that meetings, e-mails, hallway conversations, and conference calls are everyday activities that drive organizational culture and productivity.

The last section of this chapter helps you shape these activities in the service of a data fluent organization. It describes how you can build a culture that supports effective data consumption, production, and use.

DATA CONSUMERS

A central tenet of a data fluent organization is that everyone in the organization is a thoughtful and informed consumer of data. To get there, leaders must attend to two interrelated activities. First, leaders must develop a culture in which individuals are trained to extract important insights from data products without being sidetracked by chance random occurrences or natural variation in the data. Second, leaders must ensure that meetings and discussions focus on the *content* of data products, rather than getting mired in discussions about data quality, sources, or calculations.

Help Individuals Evaluate Data without Distraction

R.A. Fisher, known as one of the fathers of modern statistical science, once argued that experimenters (invoking the classic image of white lab coats, thick-rimmed glasses, clipboards, randomized experimental assignment, treatment, and control groups) go to extreme lengths in the execution of their experiments to “decrease or eliminate all the causes of fortuitous variation that might obscure the evidence.”⁵ Though we realize that your data might seem far removed from these clinically based randomized experiments, as a leader it is vital that you help develop an organizational culture in which individuals recognize the difference between meaningful patterns and insights and those that are brought about by chance random occurrence. One important way of safeguarding against the trap of mistakenly identifying natural variation in the data as meaningful is to use evidence from multiple sources or time points. Often a single anomalous event can be identified by examining past trends or relying on other sources of information for triangulation, including intuition and personal experience.

For example, public companies are subject to the critical evaluation of industry analysts who attempt to predict the future prospects for the company. Many investors look for the consensus among analysts to try to determine the growth and risk prospects in buying shares. What happens when a lone analyst breaks from the pack to offer a contrarian—say negative—opinion? Should this outlier analyst be considered a forward-thinking “canary-in-the-coal-mine” or someone simply looking for attention? We know there will be variations in opinions, so the first reaction by the company and investors should not be to panic at an extreme viewpoint. Does this analyst have credibility as a data source? Is there a history with the analyst in providing prescient predictions, or a history of histrionics? The weight put on this analyst’s opinion should reflect this context.

Focus on the Message

Chapter 5, “Data Authors,” outlines a few of the critical elements of successful data products. An important principle highlighted was that the most effective data products tell a story. One of the reasons why stories are so effective is that they help the consumer focus on the message and content *without* getting distracted by other issues. Questions surrounding data sources, calculations, or interpretation can often distract from the central message of a data product.

As a leader, you have the opportunity to model this behavior, focusing your attention, questions, and insights on the meaningful information and *not* the ancillary details that derail meetings. More important, your success in keeping discussions on track will be largely dependent on individuals in your organization who produce the data. As such, now turn to issues surrounding data production and data products.

DATA PRODUCTS

The previous chapter reviewed principles for creating successful data products. As a leader, you can support the work of data producers by establishing clear guidelines for quality and effective communication between data producers and consumers. As part of data communication, you should encourage a feedback mechanism to help improve the content while celebrating examples of quality data products.

Establish Clear Guidelines for Quality Data Products

In your efforts to help producers focus on the message, it is important to establish a set of clear guidelines designed to produce quality data products. In creating these guidelines, it is helpful to draw an analogy with writing. On questions of vocabulary and style in written communication, there seems to be agreement about the pursuit of simplicity and ensuring that the message, not the words, takes center stage. Following are three common writing guidelines and what they mean for helping to guide the data communication choices your data producers will make.

- **Use smaller, simpler words**—C.S. Lewis once commented: “Don’t use words too big for the subject. Don’t say ‘infinitely’ when you mean ‘very’; otherwise you’ll have no word left when you want to talk about something really infinite.” The simplest visualization techniques, including bar charts, line charts, tables, and even individual numbers,

are effective for the majority of data communication. Like simple words, you can be assured that when your data products use these simple forms your audience will quickly understand the message. It is better to encourage your data producers to refrain from advanced visualizations, like animated bubble charts, except in those rare occasions when the breadth and nuance of your data requires more visual sophistication. Even so, there is actual value in data producers becoming familiar with advanced data visualization techniques. Just as a large vocabulary offers an author more fodder with which to clearly communicate her message, advanced visualization techniques can offer similar benefits to a data producer.

- **Remember that using too many words is a symptom of poor understanding**—Albert Einstein once said, “If you can’t explain something simply, you don’t understand it well.” The quickest way to lose a meeting is to realize that the data presenter is thinking out loud. In data presentation it seems especially tempting to think out loud by showing all the data in every conceivable way. Unfortunately, we’ve all been a recipient of the exhaustive 80-page PowerPoint report with seemly infinite variations on the same data. Make sure that the expectation is that authors of data products form an opinion and get to their point immediately.
- **Words are for communication, not show**—The great English poet Samuel Taylor Coleridge once wrote, “Words in prose ought to express the intended meaning; if they attract attention to themselves, it is a fault; in the very best styles you read page after page without noticing the medium.” Far too many data presentations contain images that are intended to draw attention to the visualization and not the message. It’s not about the data or the visualization; it’s about the message that the data presenter is trying to communicate. This is not simply a question of efficient communication; it is also a question of perception. When data products are unnecessarily complex or showy, they draw attention to the wrong things. Figure 6-4 is an example of a chart that is more about the chart than the message:

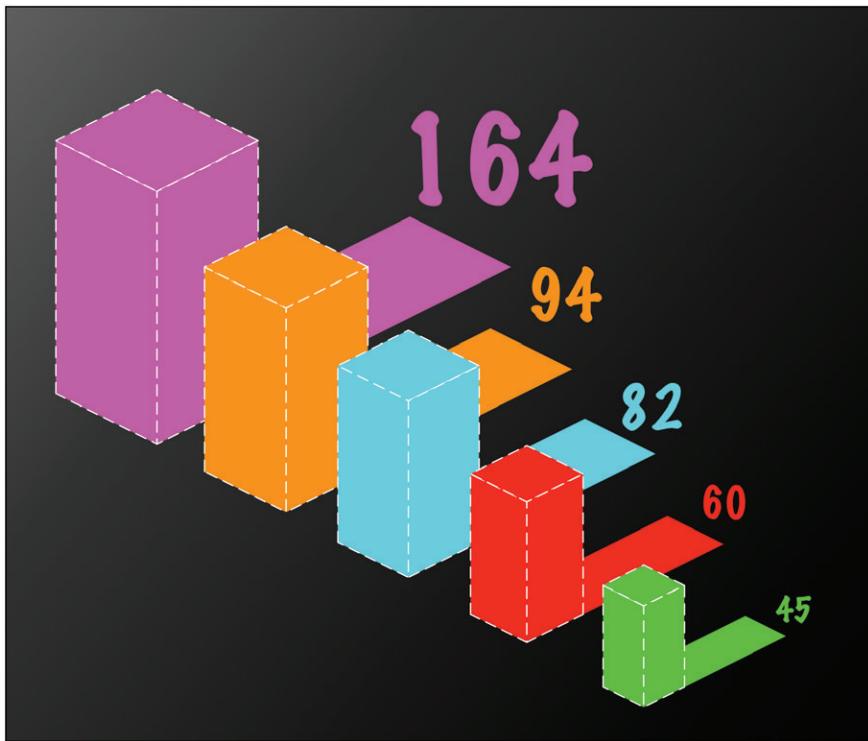


FIGURE 6-4 This chart loses the message in its efforts to dazzle.

Develop a Feedback Mechanism for Data Products to Help Evolve and Improve Content.

The key to a successful data product is to ensure that there are mechanisms in place to allow for feedback to improve and evolve content. Often, users need to see results before they can ask better, more detailed questions of the data. These feedback loops provide critical learning to improve data products.

In providing feedback, there will be times when you need to retire a certain data product. In this regard, the economists' refrain, "sunk costs are irrelevant," is worth remembering. Despite the past time, energy, attention, and resources sunk into generating a data presentation or product, products that are no longer useful must be excised from your data ecosystem. Retaining data products that are not used makes for a cluttered system and can create confusion for future employees unfamiliar with the product's history or former utility. From an employee perspective, it is a waste of time to have to

sort through or consider an outdated data product. Key considerations with respect to retiring a data product include:

- How often is it used and by whom?
- What product or products replace it?
- What other reports or data connect to it?

Although some data products need to be retired manually (that is, someone needs to make a decision and hit the Delete key), other data products can be retired by inactivity. The next chapter discusses how your data ecosystem can be designed in such a way to efficiently and effectively manage this process.

Celebrate Examples of Quality Data Products

Finally, leaders need to celebrate the best data products through public praise. In doing so, make it clear how your organization defines “quality.” Why is this product praised? Is it because of its utility? If so, how has it been useful? Has it contributed to decision-making? What were the results? Making explicit the reasons for celebrating a specific data product will go a long way toward developing a data fluent culture.

In addition, you can celebrate quality data products by encouraging their use. As a leader, however, there is often no better way to celebrate quality data products than by using them in your own daily activities. The next section elaborates on the topic of data use.

DATA USAGE

Developing a data fluent culture in which individuals focus on the message of quality data products is not enough: The insights gained must be both actionable *and* acted upon. Leaders must encourage individuals throughout all levels of the organization to incorporate data into their decision-making processes. Just as there should be a feedback loop in the production of quality data products, so too should there be an opportunity to reflect upon previous data-driven decisions. Furthermore, you should continually evaluate the progress your organization is making in effectively using data and making data-drive decisions.

Encourage Data-Driven Decision-Making

In the words of the Scottish writer Andrew Lang, data should not be used like a “drunk man uses lamp posts, for support rather than illumination.” Data-informed decision-making must be just that—*informed* by the data. Arriving at a conclusion and then searching for the data to support it is tempting. As a leader, however, you should encourage individuals throughout all levels of the organization to *first* examine data products with the intent of *then* finding actionable insights.

More important, you can create a culture in which organizational teams and individuals reflect upon previous data-driven decisions to inform future decision-making. One litmus test you can use is to ask leaders to have groups think about the last three decisions that were informed by data. What were the conditions? Were there any commonalities across decisions? Building a culture of reflection is an important mechanism for establishing a data-driven culture.

Evaluating Effective Data Use within the Organization

How do you know that data is used to drive decisions in your organization? It is a subtle shift to measure. However, there are a number of indicators that can give you a directional understanding of how well your organization has adopted use of data:

- **Evaluating new products or projects**—Has your organization created a well-understood and accepted model for evaluating new investments? Of course, each new endeavor will have unique aspects, but the performance tracking approach should start from a common template.
- **Data trumps anecdotes**—Many organizations suffer from the tyranny of anecdotes. A single data point, or long-ago experience, becomes the deciding factor in decisions. These anecdotes can be instructive and provide useful color to a discussion. They should not be accepted as data.
- **Data products for important processes**—For each of the priorities in the organization, someone should have created a data product that sheds light on the progress toward this goal. The data product may be a quarterly summary of results or a real-time dashboard of operations.

Is there an important element of what your organization does that is not transparent and could benefit from an effective data product?

- **Grass roots expectations**—As noted earlier, soon after key metrics are introduced, people will begin to wonder about the goals associated with these measures. This is a powerful indicator that the attention of the organization has turned to measuring performance. Likewise, employees should start to expect that meetings will be informed by data and that the executive team communicates priorities and results backed by data.

EVOLUTION OF DATA FLUENT CULTURES

Building a data fluent culture takes time. It is a journey. We've worked with many companies who are committed to the process of putting data at the center of their organizational conversations. But it is rare to find an organization that has had success in all the areas described in this chapter.

In one case, we engaged with an innovative healthcare company that helps reduce the cost of employee healthcare. Its customer service-based model is a new approach in the industry. From the CEO on down, it has a deep commitment to using data to improve its processes and prove its value to customers. However, its business model did not have an established measurement approach and key metrics were elusive. As a result, the company tracks a myriad of metrics and continues to search for a shared understanding of what is the right data to focus on. This organization doesn't lack for commitment in its pursuit of data fluency, but needs to continue to evolve.

In contrast, we worked with a leading Internet content provider who had the opposite problem. This organization had the capability to measure everything about its business and existed in an industry in which there were commonly accepted measures of success. Key metrics were tracked diligently by analysts and reporting groups throughout the organization. It even had established credible "sources of truth" in the data. The problem was with the leadership's commitment to using data to guide decisions. We could easily tell what content was engaging audiences and returning value on the investment. But these objective and accepted measures of success were seldom prioritized over the opinions of content owners.

It is a long-term challenge to put together all the pieces of a data fluent culture. In the end, it takes three core elements:

- **Leaders** who value data as a core element of their approach
- **Commitment** to work through the many complications and investments necessary to develop credible data sources and quality data products
- **Focus** on the right metrics, useful data products, and communicating with a clear message

NOTES

1. John Sullivan, "How Google is Using People Analytics to Completely Reinvent HR." *TLNT*, February 26, 2013. Retrieved from: <http://www.tlnt.com/2013/02/26/how-google-is-using-people-analytics-to-completely-reinvent-hr/>
2. Manny Fernandez, "El Paso Schools Confront Scandal of Students who 'Disappeared' at Test Time," *The New York Times*, October 13, 2012. Retrieved from http://www.nytimes.com/2012/10/14/education/el-paso-rattled-by-scandal-of-disappeared-students.html?nl=todaysheadlines&emc=edit_th_20121014&pagewanted=all&r=0
3. Mike Sando, "How to Identify NFL's Best Quarterbacks," *ESPN.com*, August 1, 2011. Retrieved from: http://espn.go.com/blog/nfcwest/post/_/id/41811/how-to-identify-nfls-best-quarterbacks
4. Michael David Smith, "ESPN Tries to Build a Better Quarterback Rating." August 1, 2011. Retrieved from: <http://profootballtalk.nbcspor ts.com/2011/08/01/espn-tries-to-build-a-better-quarterback-rating/>
5. *Statistical Methods and Scientific Inference*, 3rd edition (Edinburgh: Oliver and Boyd, 1973), p.97.

CHAPTER 7

The Data Product Ecosystem

The teacher's first duty is to watch over the environment, and this takes precedence over all the rest. Its influence is indirect, but unless it be well done there will be no effective and permanent results of any kind, physical, intellectual or spiritual.

—Maria Montessori

Our children attended Montessori schools for the first years of their education and we loved it for the unique educational approach. Maria Montessori believed the right conditions encourage the right behaviors.

Among the core principles she espoused was a “prepared environment,” or a learning environment that is carefully conceived to support student success. The classrooms are often quite beautiful. The space is clean, open, and well organized. Big learning concepts are broken into small building blocks for “work.” The work materials are of high quality and easy for children to manipulate. Students know where to find different types of work, and they have free reign to choose the pieces that are most interesting to them.

What are the conditions in your organization that will encourage and support the flow of information through data products? How can you create an

environment in which data product authors have the tools they need to be successful? What does it take to allow good data products to flourish?

In this chapter, you will learn:

- How data products can be different in their purposes but share common characteristics
- The ways in which the Apple app store provides a model for a data product ecosystem
- The six foundational elements that compose an effective data product ecosystem

DATA PRODUCTS FOR INFORMATION DELIVERY

This book describes data products as the medium for communication in your organization. Data products include the periodic reports that explain what's happening in a functional area, the dashboards that attempt to provide a high-level view of performance, the spreadsheet models that describe financial projections, and the presentations that tell a story with data. In a blog post about data products, Harlan Harris says, "Data products make it easy for regular people to get what they need without having to dive into a very complex set of data and a very complex set of algorithms."¹

We use the term "products" because they represent the transformation of data into something useful that can be delivered to the people who want to consume the information. The best data products are:

- **Results-focused**—Data products should be targeted to answer specific questions or solve particular problems.
- **Right-sized**—A data product that tries to do too much will be hard to explain to the audience, whereas one that tackles too little will seem trivial.
- **Context-sensitive**—Authors of data products should have domain expertise and use this knowledge in choosing what data to emphasize.

From this definition, there is a clear gap between data products and general-purpose tools for business intelligence, statistics, analysis, or visualization. These analytical tools may be used to create data products, but they are not themselves focused, purpose-built products.

Data products come in many varieties—both in form and function. In some cases, data products emphasize the ability for users to explore the data independently. Others lean toward guided storytelling, a vision driven by the author more than the audience. Along another dimension, there are a set of data products that offer varying degrees in which the author can enhance the data through analysis, introspection, and features. Figure 7-1 shows how different types of data products can be mapped by these characteristics.

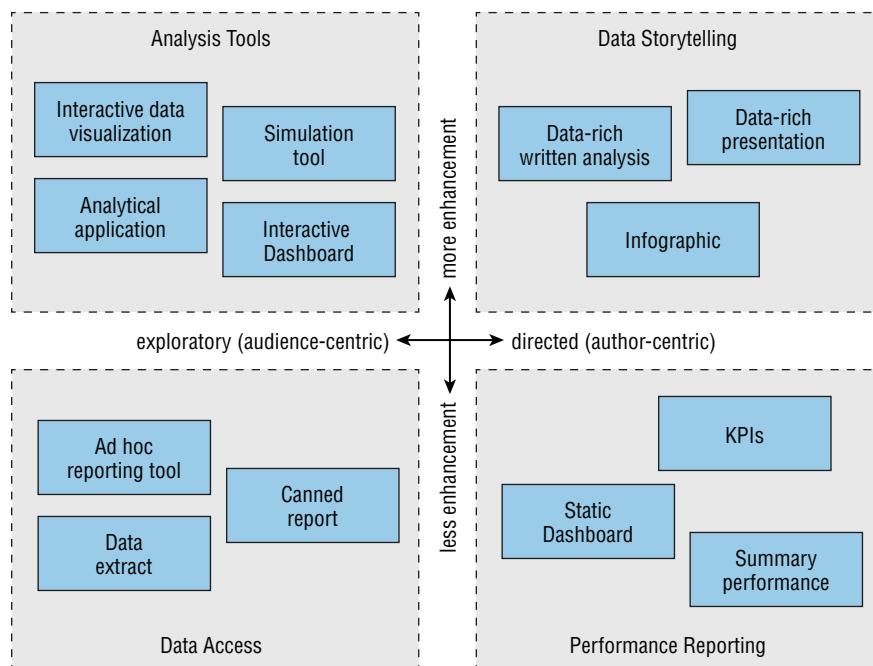


FIGURE 7-1 Data products classified by level of enhancement and audience vs. author control

NECESSARY CONDITIONS

Sun and water and soil make the flowers grow. What will make your data products flourish? This chapter is an exploration of the conditions that, taken

together, create a fertile environment for data to flow freely. Data products need an environment that:

- Encourages the right data products to be created and shared;
- Makes it easy for knowledge workers to find the right information for their job;
- Cleans out the old and irrelevant data products, like a gardener who diligently weeds.

Creating an ecosystem for data products requires processes and tools. Processes set standards and ensure that the right priorities and qualities are built into every data product. Technology tools gather data, visualize the results, and distribute data products to users. Here are the six conditions (“the Six Ds” shown in Figure 7-2) that we have found most critical to the data product ecosystem:

Demand—What are the most important areas that would benefit from the insights and guidance of better data?

Design—What processes and tools can help ensure effective design of data products?

Develop—What processes and tools support the efficient production of data products, including gathering multiple data sources, presenting this data, providing user customization, and delivering the information to data consumers?

Discover—How can you help people discover the many data products in your organization and find the right information for their individual needs?

Discuss—What capabilities encourage data consumers to take the insights they find in the data and share these insights with others?

Distill—How do you filter out the irrelevant content and provide feedback to enhance those data products that remain?

Together, these conditions provide the environment that can allow you to create useful, engaging solutions that drive informed actions. The next sections provide details to help you understand what you can do in your organization to create these conditions. Each section describes the conditions, why it is important to building your data product ecosystem, a vision of the ideal capabilities to strive for, and a pragmatic view of actions you can take today to move in the right direction.

Finally, a word of warning: In our experience most organizations are in their infancy with respect to data fluency. Organizations often have small pockets of individuals that have a conscious understanding of which data products they need and how to design and create good ones. These skilled data consumers and data producers, however, make up the minority and often find roadblocks to sharing their work and permeating the organizational culture with a level of data fluency. As a result, some strong data products that are powerful communication tools often exist side-by-side with outdated, unused, and misguided reports. The Six Ds of a healthy data ecosystem help create an environment for the minority of savvy data users in your organization to extend their influence and organizational performance in the process.

The conditions expressed here represent an ideal model. Strive toward them and appreciate the value of moving toward a better ecosystem.

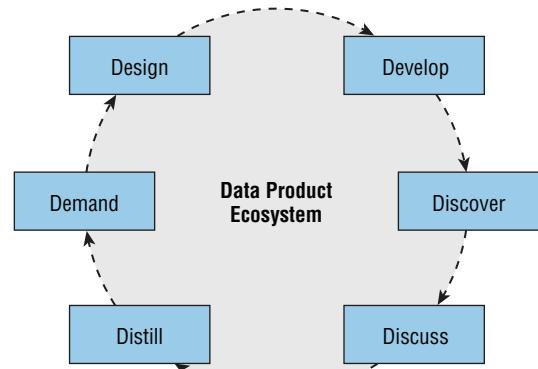


FIGURE 7-2 The six Ds are the necessary conditions for a healthy data product ecosystem.

LEARNING FROM THE APP STORE

Over a holiday weekend, 28-year-old Dong Nguyen made a mobile game. His idea was simple—tap your screen to fly a bird between a series of vertical green pipes. For each set of pipes passed, you get a point. He called the game *Flappy Bird*.

Flappy Bird launched in May 2013. Later that year, the game went viral. By the end of January 2014, it had been downloaded more than 50 million times, making Dong an estimated \$50,000 a day in in-game advertising. Ten days later, the game was gone—pulled by its creator.

"I was just making something fun to share with other people," Dong said. "*Flappy Bird* was designed to play in a few minutes when you are relaxed. But it happened to become an addictive product. I think it has become a problem. To solve that problem, it's best to take down *Flappy Bird*. It's gone forever."² Days after, smartphones with the game began to be auctioned on eBay.

Although the sales were never permitted, some bids ranged in the thousands of dollars—a testament to the addiction that Dong had feared.

But how did Dong launch this app? How did his idea for a simple game go from virtual anonymity to the devices of millions of mobile users across the world in little more than a month? The answer is in the incredible data ecosystem Apple, Google, and others created to distribute apps.

Let's use Apple's App Store as an example because it was the first. With the smartphone revolution came a huge increase in demand for mobile applications. For instance, since the launch of the App Store on July 10, 2008, the number of iPhones sold has grown beyond 500 million.³ With this many iPhones, the needs of users have become increasingly clear through top-selling lists of apps and increasingly clear categories on the App Store.

At the same time, Apple has been careful to provide app developers with a system that allows individuals and organizations to effectively design and develop mobile apps for use on the iPhone and for sale in the App Store. Apple has gone so far as to provide a guidebook for designing Apps called the *iOS Human Interface Guidelines*. This book provides design principles to (strongly) encourage application developers to create user interfaces that fit neatly into the Apple ecosystem. Specific guidance includes:⁴

"Don't take space away from the content people care about."

"Use a tone that's informal and friendly, but not too familiar."

"Focus on the needs of 80 percent of your users. When you do this, most people won't have to supply any settings because the app is already set up to behave the way they expect."

Most important, Apple created an App Store that allowed users to discover apps, discuss them, and distill feedback in a convenient and meaningful way to improve future iterations of the app.

Flappy Bird is a case study of the power of Apple's app ecosystem. To develop *Flappy Bird* for sale in the App Store, Dong first had to download and install Xcode, the application in which all iPhone apps are created. Next, he used a common programming language, Objective-C, to program the functions and overlay his graphics for the game. In Xcode, Dong could also test the game in an iOS simulator, try it on his own iPhone, debug it, and distribute a test version to others. With this process completed, Dong could then purchase a development account through Apple's App Store for \$99 a year. With this

license, he's ready to launch his app. In Xcode, he would use the Application uploader tool and wait for approval from Apple. When approved, the game goes live for download on the App Store.

Although Dong's story is clearly an outlier—one of maybe a dozen mobile games that have gone viral, and possibly the only game that's been pulled because of too much use—it illustrates the power of a successful ecosystem. Dong, a 20-something who was living with his parents in Hanoi, Vietnam, entered an environment that allowed him to easily design, develop, and distribute *Flappy Bird*, one of the most discussed games in mobile application history.

DEMAND

There is nothing so useless as doing efficiently that which should not be done at all.

—Peter Drucker

We begin with the end in mind. The consumers of data have needs. A healthy ecosystem will support those needs through the right data products. Discovering the information that will best serve the organization is the first step.

Understanding data consumer demand is not a one-time endeavor. It requires a process of continually mapping the important decisions made by the organization and evaluating whether and how data can improve those decisions.

Organizations without an understanding or responsiveness to data needs may suffer from gaps in areas in which data could inform decisions. At the same time, these organizations may be prone to extreme analytical fixation on particular areas at the expense of other areas of greater importance.

There are a couple of ways to build a complete awareness of your organizational data demands: a top-down map or listening for grassroots needs.

TOP-DOWN DEMAND MAP

The best opportunities for useful data products come at the intersection of important and frequent decisions and situations in which better data leads directly to better decisions. Your organization may decide to enter a new market, but if this decision is guided by executive experience and happens

infrequently, then there is little a data product can add. In contrast, how a sales team decides to discount for new customers is an area that may happen often and can be informed by an understanding of the long-term value of the new customer.

The following prioritization process helps identify where to start:

- 1.** Build a comprehensive list of processes, functions, and decision-points that happen in your organization. You may want to start from the first point you interact with a customer and trace all the interactions directly or indirectly related to the customer. Consider functional areas such as human resources, finance, supply chain, operations, and sales.
- 2.** For each element, rate it on a scale of one to five for importance to the success of your organizational goals. One means the element has little to no impact on organizational success; five means it is a core and critical decision.
- 3.** For each element, rate it on a scale of one to five for the value better data can bring to decisions. One means data brings little to no value to the decision process; five means data always makes the difference between good and bad decisions.
- 4.** Plot all the elements in a chart, as shown in Figure 7-3.

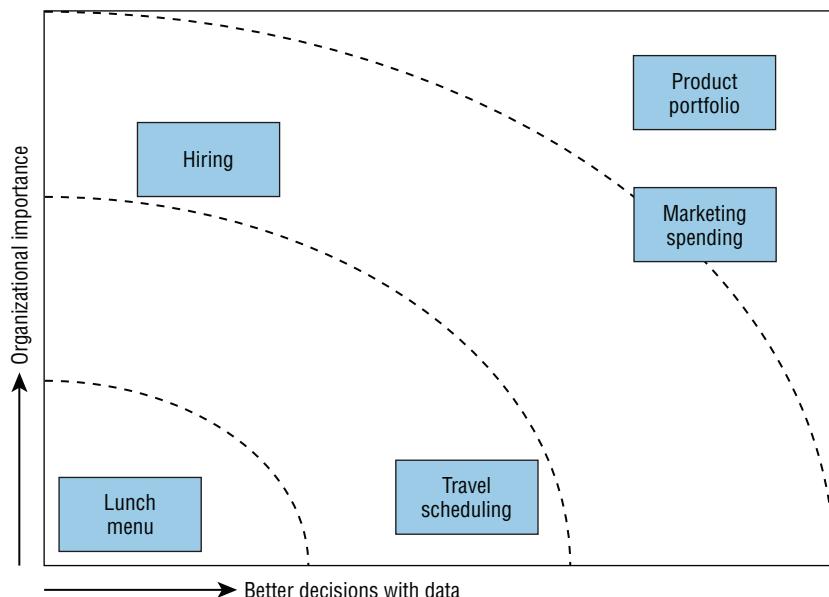


FIGURE 7-3 An approach for prioritizing data product needs

In this chart, the further an element is from the lower-left corner, the more value a data product will bring.

GRASSROOTS NEEDS

The alternative to a top-down mapping exercise is to let the demand for data needs surface from the people who want it most. As a leader, how do you hear what the “market” needs? Encourage people to ask difficult questions and create an environment where these questions are embraced.

Warren Berger, author of *A More Beautiful Question* (Bloomsbury, 2014), suggests that encouraging questioning in your organization can start with the leader but is a grassroots effort:

“A culture of inquiry starts at the top—with leaders who question.”

“Questioning should be rewarded (or at least, not punished).”

“Give people the time and space to question deeply.”

“Provide the tools to question well.”

The last point is of particular practical use. Approaches like “The Five Whys” provide a proven model for getting people to think more deeply about their circumstances. When they do, they are likely to wonder why decisions are made the way they are made and whether better information would improve those decisions.

WHERE TO BEGIN

Understanding where you need better visibility within your organization can be complex for a large university, a growing public company, or for a small start-up. The answer starts with a simple question: What goal—if achieved—would cause your organization to celebrate in the next 12 months?

With that answer in mind, you can define what you most need to measure to track progress toward this goal. Perhaps there are a number of factors that must come together for the goal. Data products focused on this goal will give you confidence that you are moving forward on the set of crucial factors—or where you need to make adjustments.

DESIGN

Less than 30 percent of the potential users of organizations' standard business intelligence (BI) tools use the technology today. This low take-up is due to the fact that longstanding tools and approaches to BI are often too difficult to use, slow to respond or deliver content of limited relevance.

—Gartner⁵

The three reasons cited by Gartner for this problem are:

- 1.** Ease of use ("is hard to work with")
- 2.** Performance ("users are frustrated by delays")
- 3.** Relevance ("does not express content in line with their frame of reference")

The first and last reasons link directly to issues of poor data product design.

In our role as dashboard and analytical application designers, this is an area that is close to home. We see it all the time: reports and dashboards that lack focus and a message that targets their audience, which is often undefined. We see poor choices in data visualization that distract from the important elements in the data and put the burden of deciphering meaning on the readers. We see data products that lack an obvious starting point and logical flow to conclusions.

Poor design is wasteful. It results in solutions that users don't want to use, as noted by Gartner. It wastes the audience's valuable time as it struggles to comprehend the data. And it wastes the development and distribution efforts necessary to deliver the data product.

Overcoming these design shortcomings requires systematic processes and capabilities within your organization. Chapter 5, "Data Authors," discusses the skills involved with designing effective data products. But how do you incorporate these skills into your processes and create repeatable ways to ensure quality design?

OBJECTIVE

Better information design in your organization will come from a combination of skilled people and processes that ensure design is a priority.

The Apple App Store has strict guidelines for applications that are made available to customers. Apple publishes design guidelines, enforces a rigorous review process, and provides tools that help app developers get their designs right. As a result, consumers of Apple applications get a consistent, high-quality experience. The design focus allows novice users to quickly learn a new app because they have come to understand common conventions and design practices.

You should consider four areas when building an environment that infuses good design into your data products:

- 1. Train** your data authors on the design skills for communicating data. Chapter 5 describes many of the areas of content that are foundational knowledge. The best information designers are discriminating in what data they choose to present and thoughtful about choosing metrics that encourage appropriate behaviors. They have the ability to create a logical flow through a series of attractive and clear data visualizations. There are also practical design skills to teach. For example, we always create wireframes—rough mock-ups of user interfaces—before building the actual data presentations. In this way, we can experiment with layout and content using rapid, lightweight iterations.
- 2. Incorporate** designers early in the development process. Many of the organizations we work with go from concept to development of data products without considering the best way to communicate the data. When an executive wants a dashboard with a set of required metrics, the data analyst gets right to work on putting all the content together. The analyst might meet the specific requirements, but not in a form that anyone wants to consume. One client referred to these types of reports and dashboards as “ransom notes” because of the great variation in styles on the page. Designers should be included early in the requirements gathering process—not at the last moment to try to polish graphics.
- 3. Document guidance** to ensure consistent designs. Many of the best practices and information design advice shared in Chapter 5 can be gathered into documents for designers. We have had success working with clients to create style guides for information design and communication. In addition, templates for reports, presentations, spreadsheet models, and dashboards can give designers a consistent starting point for their work. We also make sure to capture design patterns and principles as a way to establish models for approaching similar problems.

4. **Review process** to maintain quality control. Your organization should set expectations for the quality of data products that are used to communicate your valuable data and share insights. The costs—confusion, frustration, wasted time, and misleading information—are not worth tolerating poor design. The skilled designers, armed with the documented guidance, may form a review board to evaluate and provide feedback on the design of data products in your organization.

START WITH A STYLE GUIDE

Good design builds on consistent, fundamentally sound choices and learns from the quality work of others. This is where your organization can start its efforts to improve data presentation design.

In our experience, an effective first step is to create a style guide for the presentation of data. It can be modeled after the branding style guides that many large organizations create to ensure a consistent presentation of content in print and online. Your style guide should provide data product authors with direction on choice of colors, typography, chart selection, and number formatting. The document provides examples of well-designed data visualizations and sets the rules for legends and labeling. In Appendix B, we have included an outline for such a style guide along with examples of the content.

“Imitation is the sincerest form of flattery,” said Charles Caleb Colton. Your data product designers should have a standard style guide and set of illustrative data products they can reference to understand quality norms and the expectations of organizational leaders. To do so, select a few examples of dashboards or reports that set a high bar. Find opportunities to publicly present these data products—setting a standard for clarity and content.

DEVELOP

There are dozens of technology solutions for developing data products. Microsoft offers a variety of options: PowerPoint for data-rich presentations, Excel for models and ad hoc analysis, Power BI for self-service analytics, SQL Server Reporting Services for reporting, SharePoint with dashboards, and undoubtedly many more. The fact that one organization offers numerous

tools for communicating data says a lot about the complexity of choosing the right tool for your unique situation.

Ideally, you want to have a small set of data tools that support the variety of types of data products your organization needs. A single solution is unlikely to offer the breadth of capabilities necessary. In our experience, four to five tools for data presentation are usually sufficient for most organizations. The following table describes a set of tool categories and the specific solution we've seen work most successfully.

Types of Data Product Development Tools

TYPE	PURPOSE	OUR CHOICE
Data presentation	Data storytelling with text and images to provide context to the data visualization	PowerPoint*
Scenario tools and analyses	Predictive models or other calculations used to explain how something works	Excel
Data visualizations	Interactive data visualizations to explain a data analysis	Tableau
Dashboards	Summary of performance in a single interface	DataHero
Analytical applications	Interactive data presentation to augment existing software applications	Fruition
Do it yourself	Charting libraries to create your own dashboards and data visualizations	D3.js

*There is a natural cringe-reaction for many people when they think of PowerPoint or Excel as tools for communicating data. While these tools have their flaws (most notably the encouragement of poor chart choices), the critiques should focus more on how individual people have misused the tools rather than the general capabilities of the tools themselves.

There are many forms your data products may take. And for every form, there are many technology options. However, here are some common features that are worth evaluating in almost every case:

- **End-user customization**—Some presentations may target a single audience. This is the exception to the rule. Most often, a data product goes out into the world alone and is used by many people, each of whom comes from a unique perspective. Whether it is their department, region, or products, all audience members will want to see data

that is customized and scoped to their situation. Many interactive applications can support this ability to filter to the relevant data.

- **Sharing support**—Data should spur conversation. However, some solutions for data products create an isolating environment. The data product should make it easy to share, discuss, and capture insights—whether the discussion happens online, offline, on a desktop, or on mobile devices.
- **Quality visualization**—It matters how data is visualized. Clean, clear charts can make it easy for readers to quickly understand the data. The default settings for data visualizations should adhere to the fundamentals shared by well-known data visualization authors like Stephen Few and Edward Tufte.
- **Fit workflows**—Finally, we believe it is important that data products integrate into how people do their jobs. If the consumer of data is constantly on the run, bombarded by information of all types, an effective data product will deliver simple, narrow content to this person. If the consumer wants to drill deeply into the data to understand underlying assumptions, the functionality should exist to allow for this need.

"IT'S A POOR CRAFTSMAN WHO BLAMES HIS TOOLS"

The tools you have today probably are a suitable start for developing data products. Better to start with the investments you've already made and the technologies your team is comfortable with. Communicating data is far more about the people creating the data products than the technology tools they use.

Start by refining the skills of the individuals using these tools—even if they are simply Microsoft Office products like PowerPoint and Excel. The design skills and capabilities described earlier in this chapter and in Chapter 5 have little to do with the specific technical solution. Furthermore, many tools—Excel in particular—can be extremely capable with practice and creativity. For example, we built the visualization in Figure 7-4 using Excel, inspired by a hand-crafted visual shown on *The New York Times* website.

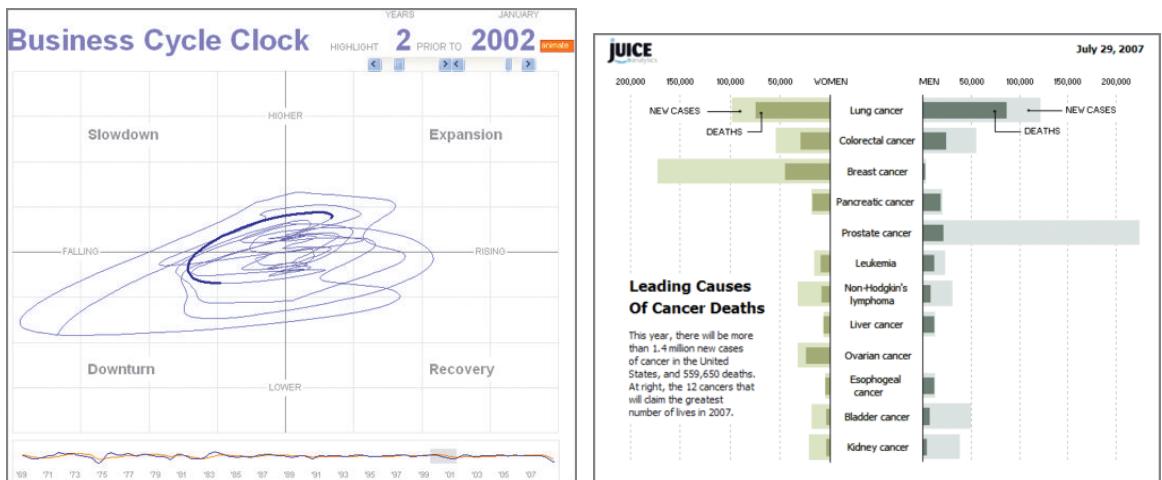


FIGURE 7-4 The flexibility of Excel

DISCOVER

Knowledge is of two kinds: we know a subject ourselves, or we know where we can find information upon it.

—Samuel Johnson, long before Google

We've never seen an organization that is short on data. Nor have we seen one without volumes of reports, analytical tools, data models, and dashboards. It isn't that the right data doesn't exist; it is that individual people may not know where to find it or how to access it.

This is the data discovery gap. The gap between supply and demand for data often is the result of, ironically, incomplete information. Just like imperfect information leads to price distortions in the economic marketplace, imperfect information within an organization leads to organizational data inefficiencies. The person who could most benefit from an understanding of customer lifetime value may not know that a well-researched analysis already exists. Better data discovery also has the added benefit of revealing duplicate, competing, or conflicting data products.

OBJECTIVE

Data product discovery should mirror the capabilities of online content subscription services. Podcasts, blogs, or Twitter, all have established features for ensuring an audience can find and access the latest content. These include:

- 1. Searching** of metadata about the content, including title, author, and description
- 2. Browsing** of content sorted into categories and ranked by popularity or ratings
- 3. Surfacing** of related content based on the consumer's expressed areas of interest
- 4. Subscribing** to allow consumers to sign up to receive updates to content
- 5. Automated pushing** that allows consumers to receive updated content automatically rather than having to remember to return to the source
- 6. Using user permissions** to control who has access to applications and content

Organizations are starting to recognize the need to provide a centralized place to find and access applications such as data products. Gartner expects that by 2017, 25 percent of enterprises will have an enterprise app store for managing corporate-sanctioned apps on PCs and mobile devices.

An app store can be a natural way to share new applications within the enterprise, recognize great applications, provide feedback to development teams, and even create a bit of competition between them—all to drive the development of better solutions.⁶

—Gartner

WHERE TO BEGIN: A CENTRALIZED INVENTORY OF DATA PRODUCTS

Before building an enterprise app store, there are easy, low-tech ways to deliver discoverability to your audiences. For example, a centralized inventory of all the data products in your organization is a start. The inventory should catalog the product author, update schedule, content area, source data, and directions for subscribing or accessing the data product. A shared drive and common naming conventions will go a long way to making it easier for data consumers to find what they need.

DISCUSS

In a vacuum, analysis and insight amounts to little. But when those results are discussed and debated, they have the potential to be acted upon. Data must get mixed with people.

The best data ecosystems don't simply assume discussions will occur. They encourage discussions through mechanisms for sharing, capturing, and saving information and insights. The discoveries found in the data are treated as precious assets—after all, they are the purpose of all the effort put into creating data products. Finally, the ecosystem should encourage people to take action when the discussion is complete.

Some organizations consider data products a one-way information broadcast. They implicitly assume that a dashboard is intended to deliver an information result, not drive an action.

Consider the evolution of television and its audiences as an analogy. Before the pervasiveness of social media, television was a powerful medium for telling a story, but there was little room for engaging the audience beyond the broadcast. "Water cooler" conversations sometimes occurred after a big event with spiked TV rankings—when J.R. was shot or after the last episode of "M.A.S.H." While there are some examples of early fan engagement via letter writing and call-in voting, generally the focus was always on the one-way delivery of a show to a broad audience.

Now things are different. TV shows have to build stronger connections with smaller audiences. To do so, the broadcast is only the beginning of the conversation. Loyal fans and TV critics write detailed reviews of every episode. Theories and predictions are hashed out on fan discussion boards. The analysis and collective synthesis of the TV show deepens the audience's understanding and engagement with the content. It took many years for television executives to appreciate and want to foment audience discussion. Now it is a crucial element in a TV show's success.

OBJECTIVE

An ecosystem that encourages discussion first needs the capability to capture insights found in data products. No user is an island; they need to take useful information they find and share it, with the context in which they found it. We call this concept *snapshots*. If an effective data product is a guided safari

through the data, then a snapshot is a picture of a trip highlight by the data tourist at a point in time. And like tourists, they will want to save these pictures, send them to relatives at home, and organize them into albums that can help recall the experience.

The technical mechanism for capturing snapshots will vary. It may be something as simple as taking screen shots of a dashboard or saving a graph from a report as a separate file. More important, these snapshots are best when accompanied by a note that describes what makes the information interesting.

Beyond capturing insights, a data fluent organization supports and encourages the discussion that is spurred by data. A monthly report on customer attrition may raise eyebrows when the trend line shifts up, but what happens next? What follow-on discussions occur to dig deeper and start to create a plan for action?

The content of the data products can help encourage their audience to discuss and take action. Often there may be direct and well-understood links between certain data or charts and suggested actions. In the case of increasing customer attrition, the organization may have learned that there are common underlying factors that contribute to this issue (for example, competitors are running promotional campaigns, recent service problems, or outages). This information should be integrated into the data product.

Information and insights are just the beginning of the journey. The DIKW Pyramid (Data-Information-Knowledge-Wisdom) in Figure 7-5 is a framework from the realm of information science. It emphasizes the hierarchy of understanding, the narrowing of volume, and the elevation in value as data is eventually transformed into wisdom.

The true objective for facilitating discussion should be to gradually synthesize the information extracted from data products into greater organizational knowledge—shared expertise and intuition—and eventually into wisdom—the ability to apply that knowledge effectively to bring more value to the organization.

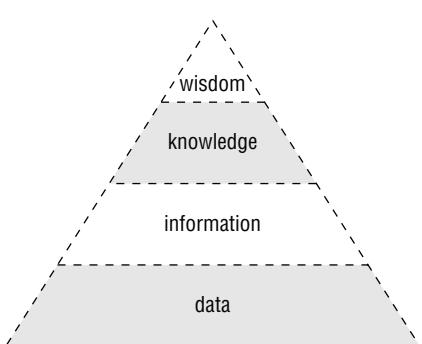


FIGURE 7-5 DIKW pyramid

WHERE TO BEGIN: CREATE A PLACE TO CAPTURE INSIGHTS

Discussions on data—like most of data fluency—are more a social and human problem than a technology problem. The

technology approaches may be simple. For each data product, create a document or folder for capturing insights. The document may simply be screen shots of the relevant part of the content along with an annotation explaining why it is interesting. As a historical artifact, this document will reveal patterns of common issues and best practice approaches for responding to those issues. More important than a complex technology solution is an organizational culture that encourages dialogue and action after the insights are first found.

DISTILL

The scourge of data in most organizations is the ever-growing collection of reports that get generated month after month. New reports are created but seldom killed. The mass of data products quickly becomes difficult to navigate and the right information is hard to find. Even for the data that has found a suitable audience, there is seldom a feedback loop. The direction and needs of the organization may change, yet the content does not change to fit evolving demands.

Data products should be living documents. They should improve over time or be removed if they are no longer relevant. It is a matter of survival of the fittest.

Data fluent organizations recognize that too much content—particularly data content—will clog up the channels of communication. The data products must be distilled to the essential information. You want to clean out distractions and emphasize the most useful remaining parts.

To filter and clean your data product ecosystem, you need processes in place to gather feedback from your users. The feedback needs to impact how data products are designed and produced.

LEARNING FROM WIKIPEDIA

Possibly the best example of an online living “document” is Wikipedia. Although many people benefit from the front end of Wikipedia, there is a whole ecosystem undergirding the popular website. The data that is viewed by all is created and managed by a network of administrators, bureaucrats, stewards, reviewers, and other specialty groups. The ecosystem governs the content

and makes public all edits on any page identifying the author. Each page and piece of data is vetted and improved by the power of collective knowledge. Wikipedia, a nonprofit organization, is fueled by voluntary effort. Yet, few large organizations have figured out how to harness the collective knowledge of their employees or associates.

OBJECTIVE

The goal for a data fluent organization should be to have multiple mechanisms to provide feedback on data products to ensure they evolve as needs evolve. Those data products that don't continually change to fit audience needs may slip into obscurity and be discontinued. Following are three basic approaches for distilling your best content:

- 1. Create a lightweight feedback mechanism.** This should be a simple process that allows users to rate data products. Did you find this data product valuable? What information was most useful? What decisions did it influence? The results from a few questions would provide the authors with valuable guidance on improving what they deliver.
- 2. Track usage.** For online products, you may find it straightforward to find out what content people are engaging with. Which reports are downloaded most often? How many people viewed the operational dashboard every day? Like content on a website, these usage metrics are a direct way to understand what is popular or not.
- 3. Conduct content reviews.** From time to time, gather the audience for a data product and have an open discussion about what they find valuable and what features or data they would like to see added. To avoid the inevitable slide toward more scope, it is useful to create an artificial constraint (for example, number of metrics or pages) to force users into hard choices about what they need versus what they want.

WHAT CAN YOU DO WITHOUT?

With an objective to bring greater focus to your organization, consider the following simple plan to narrow the volume of data products being produced. First, for each functional area build a list of all the data products currently being used. Second, create a simple survey based on this list. The survey can have a single question: "What three pieces of information would you recommend to a new employee in your role?" Third, determine what data products received

no votes. These are the data products that need to prove their value. Would anyone miss them if they were gone? In the final step, stop distributing these data products and see if anyone complains. If the data is being presented but not used, it is a good bet your resources are better focused elsewhere.

"ONLY CONNECT"

Because knowledges are so specialized we need also a methodology, a discipline, a process to turn this potential into performance. Otherwise, most of the available knowledge will not become productive; it will remain mere information. To make knowledge productive, we will have to learn to connect.

—Peter Drucker, *Pro-Capitalist Society*

Drucker is referencing E.M. Forster's belief in the imperative of human connection. This chapter is about connection and how to create the conditions that facilitate connections. The six Ds are pillars that ensure data gets turned into productive action, connecting people to the information that leads to smarter decisions:

- **Demand** is the connection of your organization's needs with the people who will produce data products.
- **Design** ensures quality outputs that will communicate to the audience.
- **Develop** is the mechanism to create data products, the connective glue that will deliver useful data to the people who can use it.
- **Discover** is connecting supply and demand after data products have been created.
- **Discuss** is the collaboration between people to understand and digest the meaning of the data.
- **Distill** is the feedback loop that ensures organizational needs are heard after a data product is created.

An ecosystem with these elements in place will allow data to flow freely, ensure quality communication, and improve data products over time.

The conditions for a data fluent ecosystem are far from trivial. In fact, in our decades of experience working with organizations that value data, we have

never seen an organization that has put all these elements together. To help on the journey, we have included suggestions for manageable places to start.

This chapter completes our tour of the four quadrants of the framework for data fluent organizations. Next this question is answered: Where are you on your journey to data fluency?

NOTES

1. Harlan Harris, "The Data Products Venn Diagram," *Data Community DC*, September 19, 2013. Retrieved from: <http://datacommunitydc.org/blog/2013/09/the-data-products-venn-diagram/>
2. Lan Anh Nguyen, "Exclusive: Flappy Bird Creator Dong Nguyen Says App 'Gone Forever' Because it was an 'Addictive Product,'" *Forbes*, February 11, 2014. Retrieved from <http://www.forbes.com/sites/lananhnguyen/2014/02/11/exclusive-flappy-bird-creator-dong-nguyen-says-app-gone-forever-because-it-was-an-addictive-product/>; See also David Kushner, "The Flight of the Birdman: Flappy Bird Creator Dong Nguyen Speaks Out," *Rolling Stone*, March 11, 2014. Retrieved from <http://www.rollingstone.com/culture/news/the-flight-of-the-birdman-flappy-bird-creator-dong-nguyen-speaks-out-20140311>
3. Kelly Hodgkins, "Apple Quietly Sells Its 500 Millionth iPhone," *Mac Rumors*, March 25, 2014. Retrieved from <http://www.macrumors.com/2014/03/25/apple-500-million-iphones-sold/>
4. Mark Wilson, "8 Tips From Apple's Official Guide To App Design," *Fast Company*, May 15, 2014. Retrieved from <http://www.fastcodesign.com/3030652/8-tips-from-apples-official-guide-to-app-design>
5. "Gartner Says Consumerization of BI Drives Greater Adoption," *Gartner Newsroom*, July 20, 2011. Retrieved from <http://www.gartner.com/newsroom/id/1748214>
6. "Gartner Says That by 2017, 25 Percent of Enterprises Will Have an Enterprise App Store," *Gartner Newsroom*, February 12, 2013. Retrieved from: <http://www.gartner.com/newsroom/id/2334015>

CHAPTER 8

The Journey to Data Fluency

Data is music to his ears. Stephen Brady was just out of college and had landed his dream job. He was working for a country music label in his hometown Nashville, Tennessee. Stephen wasn't an aspiring musician or music producer; his talents were in data analysis.

A few months after joining Black River Entertainment, Stephen could hardly believe the possibilities before him. He had access to broad and deep data about radio promotion and plays. The potential seemed enormous: He could see how different songs performed as they moved up or down the charts. He could track the viral nature of songs as they spread across geographic areas. And yet, it seemed as if nobody—within the industry, much less his organization—was making use of these data assets.

Stephen set to work building probabilistic models to forecast radio airplay trends. He looked for predictive variables that could help improve the efficiency of radio promotion. His analyses could help the music label promote its artists more effectively by understanding when and where to release singles. He set up a meeting to share his insights with the music label's CEO Gordon Kerr and the promotional team.

Eighty slides later and after seeing visualizations like the one in Figure 8-1, the promotional team shared a collective look of awe and confusion. What was it

supposed to do with his new probabilistic model? What did he mean when he shared a graphic showing "the cosine similarity between radio programming trends using a clustering algorithm to find clusters of stations that exhibit similar programming trends"?

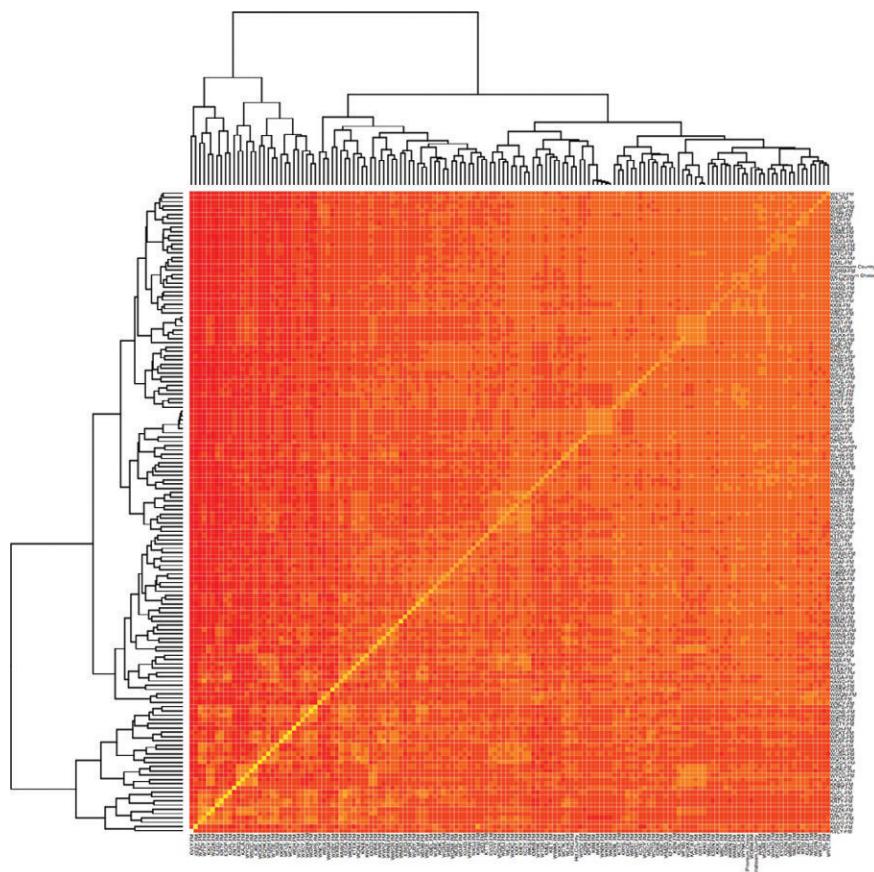


FIGURE 8-1: Radio station visualization

Fortunately for Stephen, Gordon saw the potential in his work. And from a CEO's vantage point, the analyses could provide a unique differentiation for the music label. He could talk to artists about Black River's capability to give them the best chance to have a hit single through targeted promotions.

Stephen's hard work had begun the journey toward smarter, more confident decision-making. He'd shown that there was value in the data—it just needed

to be unlocked. But despite these efforts, Stephen and Gordon still had to learn to communicate the data in ways that everyone in the organization could understand and act on.

WHY DATA FLUENCY?

It is the experience of Stephen, and all the other analysts toiling with data, that have animated our efforts over the years. We've seen the energy that gets poured into making sense of data—and all the waste as it gets obscured in reports, detailed analyses, and slide presentations that fall flat with their audiences. If the data were communicated differently, these analysts might move people to action, change minds, and shine light on important choices. If the value in data is to be unlocked, it takes better communication skills and an engaged audience.

In 2007, Professor Thomas Davenport wrote an influential book called *Competing on Analytics: The New Science of Winning*.¹ He stoked a smoldering ember into a flame by examining the power of analytics to improve organizations. The book catalyzed a generation of business leaders to look to find the value in their data.

Of course, we had quibbles when the book came out. In particular, Davenport advocates a centralized model for analytics, where the data is managed at an enterprise-level by a cadre of data scientists building complex models to drive decisions throughout the organization. He believed that the best organizational structure is a . . .

central analytics and data science organization, based in a Strategy function, with analysts assigned to and rotated among business units and functions: This is, I think, the optimal structure and home for analytics and data science. The central function allows for a critical mass of quants and for central coordination of their skill and career development. It should be a shared service, where anyone with the money and the high-priority requirements can get the help they need. The assignment and rotation allows for close relationships between analysts and decision-makers.²

We believe in another approach, one that emphasizes front-line decision-making in your organization. For us, management guru Peter Drucker sums it up best:

Most discussions of decision making assume that only senior executives make decisions or that only senior executives' decisions matter. This is a dangerous mistake. Decisions are made at every level of the organization, beginning with individual professional contributors and frontline supervisors. These apparently low-level decisions are extremely important in a knowledge-based organization.³

Senior leaders in your organization may make the so-called “big strategic” decisions, in effect choosing the path to travel down. But the speed with which you travel toward your goal and stay on course when distractions arise—these decisions are controlled by your front-line staff. We want to empower nurses, teachers, administrators, marketers, and salespeople (as well as data analysts) with the language of data.

This belief that data can inform better decisions throughout an organization is part of our motivation. There are other foundational beliefs that have fueled this book.

Data needs to be formed into targeted, purposeful solutions to be of use to most people. The common practice of delivering a general purpose analytical tool to end users and expecting something useful to happen with it, typically results in little added value. People are busy with their jobs. The last thing most information workers have time for is to learn how to use a new analysis tool, figure out what data might be relevant to them, and dive deep into a data analysis exercise. There is no scale or leverage to such an approach. Before adding this burden on people, the data product authors (Chapter 5, “Data Authors: Skilled Designers of Data Presentations”) should take time and thoughtfully find the right data, the right chart, and the right message to provide the end users with a helpful solution. It is the difference between throwing someone an anchor and throwing them a lifeline.

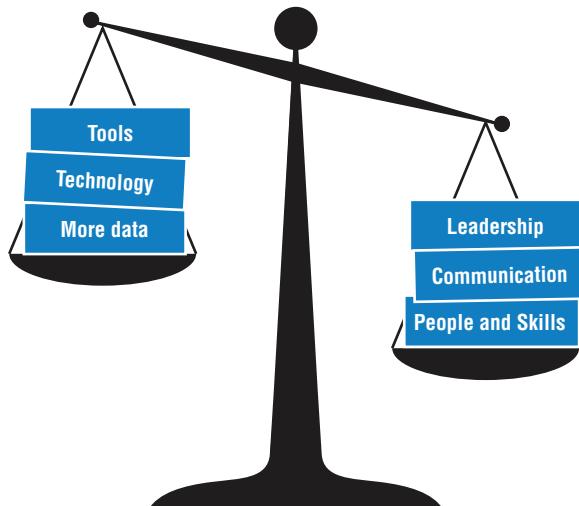


FIGURE 8-2: People over technology

We have also made clear our belief in people over technology (Figure 8-2). There are many suitable technologies for capturing, managing, manipulating, and presenting data. Better technology or tools is seldom the problem. Actually, many of the data challenges that required large

information technology investments a decade ago can be done quickly and economically today. The challenges are in the skills and collaboration of the people that use those technologies. Poor communication, misalignment of values, limited data communication skills, unfocused messages . . . these are the challenges that most organizations we work with face today. The good news is that these are all solvable by focusing on the skills of your people.

This book is about creating organizations that value and use data, and the individuals that can make this happen. The book started with a focus on individual skills by explaining the specific knowledge that can make people effective at working with data. We drew a distinction between authors of data products and consumers of data products. (Although, many readers may find themselves in both camps.) From this granular view, you looked at how organizations can encourage data fluency. An environment was described in which the data fluent individuals can flourish and high-quality data products reach their intended audiences.

The Data Fluency Framework comes in four parts, subdivided by the producer-consumer and individual-organization dichotomies. Each quadrant can be understood as a necessary and interdependent building block to a fully data fluent organization.

DATA CONSUMERS: CREATING A SOPHISTICATED AUDIENCE

*Soon the digital divide will not be between the haves and the have-nots.
It will be between the know-hows and the non-know-hows.*

—Howard Rheingold

The same could be said of data fluency. Rather than the haves and have-nots, there are those who engage with data and those who choose to avoid it. It is a path chosen based on comfort level, education, and experience. As data becomes more common as part of everyday communication—in the media, in entertainment, and in the workplace—this data-divide becomes a more substantial problem.

Chapter 4, “A Consumer’s Guide to Understanding Data,” described the skills and mindset required to become a sophisticated data consumer. It covered

the basic elements of data analysis and data visualization that help someone move beyond basic interpretation to a deeper appreciation of the message in the data. Educated data consumers are the foundation of a data fluent organization because they will appreciate, understand, and use the data products that are delivered to them. Without this, all the efforts to share data have little impact. A good data consumer conscientiously listens to and understands the language of data.

DATA PRODUCT PRODUCERS: THE SKILLS TO ENABLE EFFECTIVE DATA COMMUNICATION

Overload, clutter, and confusion are not attributes of information, they are failures of design. So if something is cluttered, fix your design, don't throw out information. If something is confusing, don't blame your victim—the audience—instead, fix the design.

—Edward Tufte

Chapter 5, “Data Authors: Skilled Designers of Data Presentations,” focused on what it takes to become an effective communicator of data. Knowing how to visualize data is at the core of this skill set. Good data product authors know how to choose the right chart and let the data shine. They also are concerned with the overall message and how different pieces of information build to support that message. They understand their audience and try to deliver data in a form that fits into their audience’s needs and workflow.

The authors of data products play a critical role in connecting vision to action. The many design decisions that they make—in choosing what metrics to show, what matters in a data analysis, and how to communicate it—can create alignment and purposeful action in an organization. Or, if done poorly, the results can lead to confusion and disengagement from important, useful data. A good data product author can effectively communicate in the language of data.

DATA FLUENT CULTURE: BUILDING A SHARED UNDERSTANDING OF DATA

People were also driven by the clear definition of the goals and an overarching purpose; they could see how their collaboration would benefit a cause larger than that of any individual. Collaboration is effective only if goals are evident: What problem are we trying to solve together? What can we do to solve this collectively?⁴

—Harvard Business Review

Organizations are social entities that come together to achieve shared goals—whether to solve a problem, make money, or deliver a service to a target population. In the everyday hustle and bustle, these common goals can become lost. Data can serve to illuminate the best course of progress amidst everyday activity just like a lighthouse's light guides a ship in a fog. By building a shared understanding of the important metrics that tie to organizational success, data can help align interests and actions.

Chapter 6, “The Data Fluent Culture,” discussed what it takes to build a culture that values and uses data in decision-making. It starts with leadership; people at the top must set the tone by incorporating data products into how they communicate priorities. For example, a shared executive dashboard provides a concrete view into what organizational leaders believe is most important.

The importance of building a common understanding of the meaning of your data was emphasized. Too often, wasted energy occurs because people debate and discuss what a data point means, where it came from, and whether it can be trusted. A data fluent culture strives to settle these issues so that people can move on to the business of understanding the message in the data.

DATA PRODUCT ECOSYSTEM: TOOLS AND PROCESSES TO FACILITATE THE FLUID EXCHANGE OF INFORMATION

The iOS and Mac App Stores alone, with more than \$10 billion in sales last year, would themselves slot in at number 270 [in the Fortune 500 list] ahead of credit firm Discover and casino giants MGM and Caesar's.⁵

—appleinsider

From July 2008, when it offered 500 apps, to 6 years later, when it offered more than 1.2 million apps, the Apple App Store has become a model for creating a fertile business ecosystem. Other companies have created similar ecosystems, most notably the Amazon Appstore and Google Play. The application stores—with their overabundance of options—are merely the visible manifestations of connected technologies and business systems. These systems are carefully constructed to deliver specific solutions to targeted consumer markets. Apple's App Store found success by providing developers with development tools and design guidelines; offering ready and plentiful demand from millions of iDevice wielding consumers; a process for approval and feedback to ensure quality; and a payment system that make purchasing easy.

Chapter 7, "The Data App Ecosystem," is not about building your own App Store. It is about learning from the App Store to understand the conditions that make for a fertile environment for delivering the right solution to the right audience. Specifically, how do you ensure the data products—your vehicles for communicating insights—are designed for and discovered by the many unique audiences in your organization?

Like growing a garden, your job doesn't end with planting the seeds. The ineffective data products must be weeded out, and the useful ones should be pruned so that they can continue to bear fruit.

BEGIN THE JOURNEY

Data fluency is the ability to use the language of data to fluidly exchange and explore ideas within your organization. There are many trends that are driving the need for data fluency:

- A desire to make use of the investments that have already been made in data gathering and infrastructure;
- Use of data in journalism and marketing to tell convincing stories, setting new expectations for communicating ideas with data;
- A decentralized workforce of knowledge workers with the tools and expectations to make independent decisions;
- New enterprise tools focused on helping employees, customers, and suppliers share and collaborate.

As you consider the path toward data fluency—whether for yourself or your organization—we want to emphasize once again that it is both challenging and worthwhile.

We have worked with dozens of organizations; every one has been deeply motivated to make use of their data. We often find pockets of skilled employees who were data fluent both as consumers and authors. Yet these people are usually frustrated by colleagues who were unable or unwilling to engage with the data. We have seen organizational cultures that have a clear and shared understanding of the metrics that matter. Yet it is more familiar for us to see organizations that are burying themselves in dozens or hundreds of “key metrics.” Most rare of all are organizations that have a stable process for creating quality data products, disseminating them to the right audiences, and improving their content and messages over time.

The journey to data fluency is a difficult path but the potential value is vast. Imagine if you could understand your customer’s behaviors, and deliver this insight into the eager hands of salespeople and customer service representatives. Imagine if everyone in your organization knew which metrics they should focus on, and where to find reports that explain opportunities to improve these metrics. Imagine if your meetings revolved around pointed analyses that attendees could understand and next steps were clearly defined. Imagine if follow-up was tracked. Imagine if everyday decisions were informed by just the right data, and everyday discussions balanced experience with analytical insight.

We hope that by reading this book you can begin to recognize the skills and organizational capabilities you already have—and find the gaps you need to fill. This book offers the landmarks that will help you track your progress and a guide for your journey to data fluency.

NOTES

1. Thomas H. Davenport and Jeanne G. Harris, *Competing on Analytics: The New Science of Winning* (Boston, Harvard Business Review Press, 2007).
2. Thomas Davenport, “Organizing Analytics and Big Data,” International Institute for Analytics. Published September 23, 2013. Retrieved on June 1, 2014 from <http://ianalytics.com/research/organizing-analytics-and-big-data>

3. Peter Drucker, "What Makes an Effective Executive," *Harvard Business Review*, June 2004.
4. Vineet Nayar, "A Shared Purpose Drives Collaboration," *Harvard Business Review*, April 2, 2014. Retrieved on August 16, 2014 from <http://blogs.hbr.org/2014/04/a-shared-purpose-drives-collaboration/>
5. Shane Cole, "App Store growth makes Apple's iTunes business alone more valuable than Xerox, CBS," *Appleinsider*, February 11, 2014. Retrieved on August 16, 2014 from <http://appleinsider.com/articles/14/02/11/app-store-growth-makes-apples-itunes-business-alone-more-valuable-than-xerox-cbs>

FEATURE

The Data Fluency Inventory

What gets measured gets done!

—Any management expert

The *Data Fluency Inventory (DFI)* is a tool that assesses your organization's ability to effectively communicating with data. In addition to measuring skills necessary for data fluency, it measures the value employees place on effective data communication. Ideally, your organization is measuring what you value most. However, this is not always the case. We designed the DFI to assist you in aligning your values, measurement, and actions. Specifically, the DFI will help you identify those areas and individuals within your organization that value and demonstrate data fluency, and those that need help.

At the *individual level*, the DFI measures the competency of *data consumers* and *data authors*. Do employees understand the data presented to them? Do they possess the skills necessary to create effective data presentations?

At the *organizational* level, aggregate scores across a team, a business unit, or an entire organization give leaders insight into where to direct resources and training to improve data literacy. What percentage of your organization

is data literate? Do you have enough data product authors in your organization? Does your organization's culture value data fluency?

Lastly, the inventory has items to assess the building blocks of a *data ecosystem*. The DFI measures whether policies and procedures are in place to foster effective communication with data. Are data consumers, data product authors, and organizational processes constantly evolving as better knowledge is generated? How well does your organization deliver the six Ds of a data ecosystem (Demand, Design, Develop, Discover, Discuss, and Distill)?

By benchmarking strengths and weaknesses of employees, teams, departments, and divisions within an organization, leaders can better understand areas of potential miscommunication or low performance. Such a process can also give organizational leaders an opportunity to recognize rising talent.

This chapter includes a series of tools to help you use the Data Fluency Inventory, including the survey questions, a scoring guide, a sample introductory e-mail for conducting the survey, and a data fluency quiz that can be used for evaluating new hires.

THE DATA FLUENCY INVENTORY SURVEY QUESTIONS

The Data Fluency Inventory is aligned with the Data Fluency Framework (Figure 1). There is a component of the DFI survey associated with each quadrant of the framework. For example, the first component of the DFI measures Data Consumer Literacy, shown in the upper left of the framework and detailed in Chapter 4.

You can use each of the four components of the DFI in a variety of ways. For example, you can use the inventory as a self-assessment to understand your own data fluency skill. Alternatively, you can use individual level data to inform talent management or human capital decisions, such as targeted training initiatives.

The DFI can raise important questions:

- Should you include data literacy in your orientation or training for new employees?
- Do data literacy training opportunities need to be created for existing employees?
- Do support structures align with goals and key metrics of success?

You can use the following inventory components as standalone data gathering elements or to provide a more holistic picture of individual and organizational data fluency.

Note Human resource teams may consider a simple data literacy quiz as part of the interview process to discern the data skills of new hires. At the end of this chapter, we include a sample data literacy quiz.

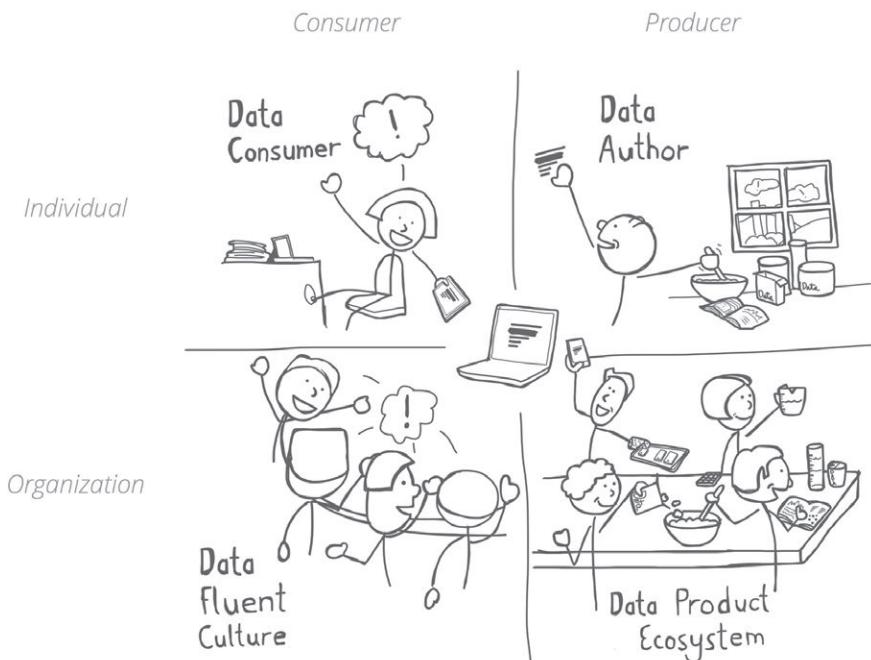


FIGURE 1 Data Fluency Framework

COMPONENT 1: DATA CONSUMER LITERACY

Component 1 of the DFI measures the extent to which employees have the abilities necessary to thoughtfully consume data. It captures how data literate your teams are and the value employees place on data fluency. Although different levels of data literacy may be needed by employees on different teams, some level of basic data literacy skills are needed by nearly all employees.



The following inventory items assess the abilities of data consumers to understand and use the data presented to them in everyday, work-related situations. Beyond data literacy skills, the inventory also measures individual attitudes toward using data.

Use of Data

PLEASE INDICATE YOUR RESPONSE AS YES OR NO.	YES	NO
IN MY ORGANIZATION		
1. I know the most important metrics for measuring the success of my organization.		
2. I like to have access to data that helps me make more informed decisions.		
3. I am able to translate data into actions that I can take in my role.		
4. I try to understand the motivations of the people who create our reports, dashboards, and presentations.		
5. I try to understand the source of the data and any issues with the data.		
MY COLLEAGUES		
1. My colleagues know the most important metrics for measuring the success of my organization.		
2. My colleagues like to have access to data that helps them make more informed decisions.		
3. My colleagues are able to translate data into actions that they can take in their roles.		
4. My colleagues try to understand the motivations of the people who create our reports, dashboards, and presentations.		
5. My colleagues try to understand the source of the data and any issues with the data.		

Data Skills

	NO SKILL	LOW SKILL	AVERAGE SKILL	STRONG SKILL	EXCEPTIONAL SKILL
HOW WOULD YOU ASSESS YOUR ABILITY TO					
1. Summarize data to find averages and totals.					
2. Understand distributions of values and identify outliers.					
3. Find interesting and useful trends in data.					
4. Read charts and graphics and understand the important highlights.					
5. Identify insights in the data that can be acted upon.					
6. See inconsistencies or weaknesses in the data.					
HOW WOULD YOU ASSESS YOUR COLLEAGUES' ABILITIES TO					
1. Summarize data to find averages and totals.					
2. Understand distributions of values and identify outliers.					
3. Find interesting and useful trends in data.					
4. Read charts and graphics and understand the important highlights.					
5. Identify insights in the data that can be acted upon.					
6. See inconsistencies or weaknesses in the data.					

Value Placed on Data

PLEASE INDICATE YOUR RESPONSE TO THE FOLLOWING STATEMENTS.			
FOR ME,	NEVER	SOMETIMES	ALWAYS
1. I feel it is important that our organization be adept at using data to guide decisions.			
2. I feel that data presentations should have a clear message.			
3. I feel that our organization's overall strategy should be linked to performance metrics.			
4. I value insights and conclusions that are based on data.			
5. I feel that it is important that our data sources be understood, reliable, and valid.			
MY COLLEAGUES,			
1. My colleagues feel it is important that our organization be adept at using data to guide decisions.			
2. My colleagues feel that data presentations should have a clear message.			
3. My colleagues feel that our organization's overall strategy should be linked to performance metrics.			
4. My colleagues value insights and conclusions that are based on data.			
5. My colleagues feel that it is important that our data sources be understood, reliable, and valid.			

COMPONENT 2: DATA PRODUCT AUTHOR SKILLS

You can use Component 2 of the DFI to measure the extent to which data product authors have the skills necessary to create data products that effectively communicate meaning. Chapter 5 discusses the important role of data product authors within an organization and explains the importance of providing data product authors with the tools and training necessary to improve data fluency.



The first set of survey questions in Component 2 focuses on the tools used by data product authors in your organization. The second set assesses the context within which data products are created and considers whether team members are using best practices in their day-to-day data authoring. The third set of questions is constructed to provide insight about perceptions and attitudes related to creating data products.

Tools

WHAT TYPE OF TOOLS DO YOU USE AS A DATA PRODUCT AUTHOR?	LIST THE NAME(S) OF THE TOOLS YOU USE	FREQUENCY WITH WHICH YOU USE THE TOOLS:	YOUR SKILL WITH THE TOOLS:
		0 = NEVER	0 = NOVICE/BEGINNER
		1 = INFREQUENTLY	1 = COMPETENT
		2 = SOMETIMES	2 = PROFICIENT
		3 = OFTEN	3 = EXPERT
1. Presentation tools such as PowerPoint, Keynote, or Prezi			
2. Spreadsheet tools such as Excel or Google Spreadsheets			
3. Statistical analysis packages such as R, SAS, or SPSS			
4. Visual analytics tools such as Tableau or Qlikview			
5. Software packages for tracking operational activities (e.g. Salesforce for CRM)			
6. Specialized tools for creating dashboards or visualizations			

Skills

PLEASE INDICATE YOUR RESPONSE TO THE FOLLOWING STATEMENTS.			
	NEVER	SOMETIMES	ALWAYS
AS A DATA PRODUCT AUTHOR			
1. I have the data tools I need.			
2. I have access to the data I need.			
3. I consider my audience as I decide what to communicate.			
4. I consider the needs of my audience's audience.			
5. My data products put non-essential or extraneous information in an appendix or equivalent.			
6. My data products help management or my audience define what is important.			
7. My data products encourage specific actions in a timely manner.			
8. My data products highlight exceptions and provide alerts when problems occur.			
9. My data products communicate progress and success.			
10. My data products clearly communicate data by reducing distracting elements (for example, chart-junk).			
MY COLLEAGUES AS DATA PRODUCT AUTHORS			
1. Consider their audience as they decide what to communicate.			
2. Consider the needs of their audience's audience.			
3. Their data products put non-essential or extraneous information in an appendix or equivalent.			
4. Their data products help management or their audience define what is important.			
5. Their data products encourage specific actions in a timely manner.			

PLEASE INDICATE YOUR RESPONSE TO THE FOLLOWING STATEMENTS.

	NEVER	SOMETIMES	ALWAYS
6. Their data products highlight exceptions and provide alerts when problems occur.			
7. Their data products communicate progress and success.			
8. Their data products clearly communicate data by reducing distracting elements (for example, chart-junk).			

Perceptions and Attitudes

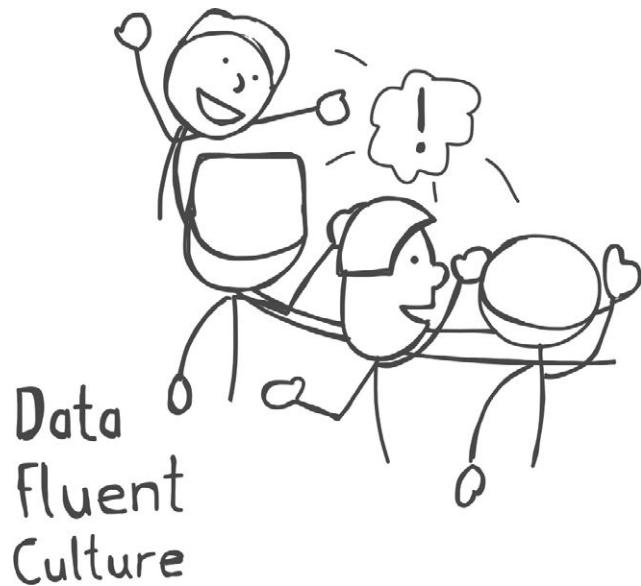
PLEASE INDICATE YOUR RESPONSE TO THE FOLLOWING STATEMENTS.

	NEVER	SOMETIMES	ALWAYS
I PLACE A HIGH VALUE ON			
1. Data products that inform decision making.			
2. Data products that can be manipulated to filter and focus on just the data that is important to the audience.			
3. Data products that link to the organization's mission and vision.			
4. Data products that have a reference guide or help for new users.			
MY COLLEAGUES PLACE A HIGH VALUE ON			
1. Data products that inform decision making.			
2. Data products that can be manipulated to filter and focus on just the data that is important to the audience.			
3. Data products that link to the organization's mission and vision.			
4. Data products that have a reference guide or help for new users.			

The first two components of the inventory help you assess individual and organizational data literacy and skills. In essence, *do you have the people in place to build a data fluent organization?* If your team is lacking the right skills and talents, you will need to train the team you have or recruit new team members.

The next two components of the inventory look to the broader organization to assess whether you have the right conditions in place to promote data fluency among data consumers and authors.

COMPONENT 3: DATA FLUENT CULTURE



Component 3 of the DFI can help you assess the extent to which your organization's leadership, culture, policies, and processes support effective data communication. Chapter 6 underscores the importance of leaders identifying and communicating a set of key metrics to rally around, and the importance of developing shared understandings for data usage. Your organization's performance metrics should be directly aligned with the overall strategy and mission, and team- and department-level measurement should be similarly aligned. Organizational policies should help create a shared understanding of the data being analyzed. Expectations of data fluency should be continually assessed and reinforced by leaders within an organization.

Leadership

PLEASE INDICATE YOUR RESPONSE TO THE FOLLOWING QUESTIONS.			
	NEVER	SOMETIMES	ALWAYS
1. Do organizational leaders' presentations use data to support the organization's mission and vision?			
2. Are specific, measurable indicators identified for each element of the organization's mission and vision?			
3. Do organizational leaders set expectations for high-quality data products?			
4. Do leaders ensure that prospective employees are data fluent during the employee recruitment process?			
5. Do leaders assess data fluency of employees when allocating them to teams, projects, and leadership positions?			
6. Do leaders provide ongoing support and development to ensure the organization has the skills to consume and use data effectively?			

Key Metrics

PLEASE INDICATE YOUR RESPONSE TO THE FOLLOWING QUESTIONS.			
	NEVER	SOMETIMES	ALWAYS
1. Does every employee understand the key metrics that are valued by the organization?			
2. Do the key metrics align to the organization mission and vision?			
3. Do team/division leaders align team/division metrics to overall organizational metrics?			

(continues)

(continued)

PLEASE INDICATE YOUR RESPONSE TO THE FOLLOWING QUESTIONS.

	NEVER	SOMETIMES	ALWAYS
4. Do all employees understand how their work contributions align with key metrics and organizational goals?			
5. Do organizational leaders reference key metrics and data analysis when they communicate goals and priorities?			

Shared Understanding and Everyday Data Use

PLEASE INDICATE YOUR RESPONSE TO THE FOLLOWING QUESTIONS.

	NEVER	SOMETIMES	ALWAYS
1. Do people who use data for analysis understand how the data is collected?			
2. Does your organization have ways to gather and share information about data sources (for example, definitions, known issues, and documentation)?			
3. Do users of data in your organization share a common vocabulary and terminology to understand the meaning of the data?			
4. Do people in your organization have easy access to raw data sources for data analysis?			
5. Can the organization evaluate data presentations and make decisions without being sidetracked by natural variations in data?			
6. Are the data reports received quickly to allow you to make timely decisions?			
7. Can you recall decisions in the last few months that were determined through data analysis?			

COMPONENT 4: DATA PRODUCT ECOSYSTEM

Component 4 of the DFI measures how well your organization has developed an ecosystem for effective data communication. Chapter 7 reviews the elements of an effective ecosystem, the six Ds: Demand, Design, Develop, Discover, Discuss, and Distill. Like Apple's App Store and supporting tools for building mobile applications, your organization can strive to make data products that are of consistent quality, respond to end-user needs, are easy to find and evaluate, and evolve based on feedback.

Demand and Design

PLEASE INDICATE YOUR RESPONSE TO THE FOLLOWING STATEMENTS.			
	NEVER	SOMETIMES	ALWAYS
1. Our organization has a process to identify the areas where better data is needed.			
2. I have influence over what data products are created and shared.			
3. Reports and other data products are designed with ease of use in mind.			
4. Reports and other data products come with relevant information to explain the context.			
5. Data product authors receive ample training about designing data products and presenting data.			
6. Data products are influenced by guidelines or rules for creating consistent product designs.			

Develop

PLEASE INDICATE YOUR RESPONSE TO THE FOLLOWING STATEMENTS.			
	NEVER	SOMETIMES	ALWAYS
1. We have the tools and processes necessary to support the efficient production of data products.			
2. Users receive reports and other data products in a timely manner.			
3. Reports and other data products used in our organization allow for easy customization.			
4. We incorporate end-user needs and designers early in the process of developing data products.			

Discover

PLEASE INDICATE YOUR RESPONSE TO THE FOLLOWING STATEMENTS.			
	NEVER	SOMETIMES	ALWAYS
1. Reports and data products are easily shared throughout our organization.			
2. It is easy to find the data, reports, or presentations that are applicable to my work.			
3. We have the capability to search and/or browse data products by title, author, description, or category.			
4. We have a centralized online environment in which we can search for data reports and see the demand for these reports.			

Discuss and Distill

PLEASE INDICATE YOUR RESPONSE TO THE FOLLOWING STATEMENTS.			
	NEVER	SOMETIMES	ALWAYS
1. Colleagues in my organization regularly discuss insights from data.			
2. Data analysts are encouraged to share their data insights with leadership.			
3. Data insights are valued and recognized.			
4. Our data reports and products filter our irrelevant content.			
5. Data products are continually being enhanced and improved based on feedback from users.			
6. Data products that are not used are retired or archived.			

SUMMARY OF THE DFI

The level of data fluency within an organization depends on the literacy of data consumers, the ability of data authors, organizational processes that foster effective communication, and a healthy ecosystem that supports the production and distribution of quality data products.

Assessing your organization is a good first step to understand where you are with respect to talent development and data fluency. An analysis of DFI survey data can lead to self-awareness at the individual and organizational level. You may conclude that your organization requires new hiring processes to ensure that new team members have a facility with data. The results may lead to new organizational processes for the use of data and the sources of data used.

The Data Fluency Inventory is a tool to help foster a conversation about data within your organization. The following section provides a simple scoring approach for each of the sections of the inventory.

DFI SCORING GUIDE (FOR ORGANIZATIONS)

The Data Fluency Inventory measures individual and organizational ability to work and communicate with data. The following section provides a recommended approach for scoring and analyzing the results of a DFI survey across your organization.

QUESTION TYPES AND POINT VALUES

All questions in the DFI can have a maximum value of 2 points. Below is a table of question types and the corresponding response point value.

QUESTION TYPE	RESPONSE VALUE
Yes/No	No = 0 points Yes = 2 points
Never/Sometimes/Always	Never = 0 points Sometimes = 1 points Always = 2 points
Skill Level	No Skill = 0 points Low Skill = 0.5 points Average Skill = 1 points Strong Skill = 1.5 points Exceptional Skill = 2 points

ORGANIZATION SCORES

An organization that takes the DFI will receive a score for each of the four components of the Data Fluency Framework. The score for each component is calculated as the percentage of total points earned out of the total available points. An organization might score 80 percent for Data Consumer Fluency, 75 percent on Data Product Authoring, 65 percent for Data Fluent Culture, and 65 percent for the Data Product Ecosystem.

COMPONENT 1: DATA CONSUMER LITERACY

The following questions evaluate the skills of an individual with respect to data consumption within the context of an organization. Do I have the skills needed today to be a conscientious data consumer within the context of my organization? Questions in this part of the DFI also ask employees to evaluate their colleagues with respect to data consumption skills.

PART OF SURVEY	HOW IT IS SCORED	AVAILABLE POINTS
Use of data	Yes = 2 No = 0	20
Data skills	Never = 0 Sometimes = 1 Always = 2	24
Value placed on data	No skill = 0 Low skill = 0.5 Average skill = 1 Strong skill = 1.5 Exceptional skill = 2	20

Thoughts for an Organizational Leader

The ability to be a conscientious data consumer is essential in today's knowledge economy. We recommend that organizations establish a baseline score and track progress year over year. Furthermore, individual scores can be returned to employees in a confidential manner in order for them to establish a performance improvement plan with regard to data consumer literacy. Individual scores versus colleague scores can be evaluated to understand if there are gaps between self-assessed and perceived skills.

COMPONENT 2: DATA PRODUCT AUTHORS

The first set of questions about data product authors takes an inventory of common data authoring tools by measuring their usage across employees. No points are awarded to these questions; however, important information may be revealed.

The skills of data product authors are evaluated in the remainder of this survey component. The individual employee skills are self-assessed, then respondents are asked to record their perceptions about colleagues' skills.

PART OF SURVEY	HOW IT IS SCORED	AVAILABLE POINTS
Skills	Never = 0	36
	Sometimes = 1	
	Always = 2	
Perception and Attitude	Never = 0	16
	Sometimes = 1	
	Always = 2	

Thoughts for an Organizational Leader

Data product authoring skills are an important aspect of knowledge economy jobs. Most professionals need to know how to create a PowerPoint deck or manipulate an Excel spreadsheet. Data product authoring, however, encompasses many more skills that may not be as common. We suggest benchmarking data product authoring skill levels year over year. Creating a performance index can help identify the missing skills and need for training. Typically, organizations score lower on data product authoring skills (Component 2) than data consumer literacy (Component 1).

COMPONENT 3: DATA FLUENT CULTURE

An organization can put policies and procedures in place as well as establish an organizational culture to guide and promote data fluency. In particular, we look at factors related to leadership, awareness of key metrics, shared understanding, and everyday data use.

PART OF SURVEY	HOW IT IS SCORED	AVAILABLE POINTS
Leadership	Never = 0	12
	Sometimes = 1	
	Always = 2	
Key Metrics	Never = 0	10
	Sometimes = 1	
	Always = 2	
Shared Understanding and Everyday Data Use	Never = 0	14
	Sometimes = 1	
	Always = 2	

Thoughts for an Organizational Leader

We recommend looking at the total organizational culture score and then looking at the three subsets to identify areas for specific improvement. More important than achieving a specific score in a given year is benchmarking your organization's organizational culture score and then tracking it over time. Small steady improvements in organizational culture are better than sweeping policy changes that may or may not take hold. As an organizational culture improves, the capacity for teams to embrace larger initiatives like building a data product ecosystem (Component 4) become more of a reality.

COMPONENT 4: DATA PRODUCT ECOSYSTEM

Having a strong data product ecosystem is aspirational for most organizations. Making it part of the DFI was intentional as it suggests a pathway to data fluency that organizations will be striving towards as a next step in organizational evolution. The six Ds of a data product ecosystem and the corresponding questions are listed below.

PART OF SURVEY	HOW IT IS SCORED	AVAILABLE POINTS
Demand and Design	Never = 0 Sometimes = 1 Always = 2	12
Develop	Never = 0 Sometimes = 1 Always = 2	8
Discover	Never = 0 Sometimes = 1 Always = 2	8
Discuss and Distill	Never = 0 Sometimes = 1 Always = 2	12

Thoughts for an Organizational Leader

Each of the six Ds highlighted will most likely be used to foster discussion about creating a healthy data product ecosystem and provide a benchmark for progress over time. Typically, organizations score lower on this component

of the Data Fluency Framework than the other three components. Building a healthy data product ecosystem is a commitment that takes buy-in from the grassroots as well as the board level. We recommend that organizations begin to address areas related to Component 1 through 3 in a staged process and then develop a roadmap for a data product ecosystem.

ORGANIZATIONAL SCORING SUMMARY

Becoming a data fluent organization is a long-term endeavor. Employees come and go. Policies change. Leaders change. Organizational culture can change over time for the better or for the worse. Obtaining a data fluency score for a team, department, division or organization can be a healthy way to begin a discussion about what matters most in an organization.

With a score for each of the four components of the Data Fluency Framework, an average of the four components produces an Organizational Data Fluency Score. Drawing an analogy to language acquisition, it takes time to become fluent or even bilingual. How does your organization stack up?

Organizational Data Fluency Benchmarks

ORGANIZATIONAL DATA FLUENCY MEAN SCORE (MEAN OF COMPONENTS 1-4)	LEVEL OF FLUENCY
0-25%	Basic Vocabulary
26%-50%	Working Knowledge
51%-75%	Partially Fluent
76%-90%	Fully Fluent
91%-100%	Bilingual

SCORING GUIDE (FOR INDIVIDUALS)

While the Data Fluency Inventory (DFI) was created as a tool to help organizations assess their level of data fluency, individuals interested in improving their skills with regard to data literacy and data product authoring can also use it. Below is a scoring guide to assist individuals in their self-assessment and self-improvement journey.

WHAT CAN YOU MEASURE AS AN INDIVIDUAL?

At the individual level, the DFI measures competency with respect to data consumer literacy (Chapter 4) and data product authoring (Chapter 5). The DFI can provide a valuable self-assessment tool for individuals who want to be better and more critical consumers of data in work environments where data is a language of choice. Most knowledge economy jobs now require individuals to re-package data for either internal or external audiences. The DFI can help you assess your data authoring skills and suggest areas for improvement. Lastly, as career mobility often depends on a favorable interview, becoming a savvy data consumer and data product author are skills that may need to be demonstrated in order to rise in any candidate pool.

The individual assessment taken as a standalone instrument doesn't ask questions about colleagues. Individual assessments can be returned confidentially to team members to encourage skill building with appropriate training opportunities.

Individual assessments (without an organizational context) are a combination of questions from Component 1 and Component 2 of the DFI.

COMPONENT 1: DATA CONSUMER (INDIVIDUAL ONLY)

This section focuses on basic understanding of how to read and interpret data when it is presented in reports, dashboards, or analytical tools.

PART OF SURVEY	HOW IT IS SCORED	AVAILABLE POINTS
Use of data	Yes = 2 No = 0	20
Data skills	Never = 0 Sometimes = 1 Always = 2	24
Value placed on data	No skill = 0 Low skill = 0.5 Average skill = 1 Strong skill = 1.5 Exceptional skill = 2	20

COMPONENT 2: DATA AUTHORS (INDIVIDUAL ONLY)

The skills for data authoring are explored in this second part of the survey.

PART OF SURVEY	HOW IT IS SCORED	AVAILABLE POINTS
Skills	Never = 0	36
	Sometimes = 1	
	Always = 2	
Perception and Attitude	Never = 0	16
	Sometimes = 1	
	Always = 2	

INDIVIDUAL SCORING GUIDE SUMMARY

An individual self-assessment is a combination of Component 1 and 2 of the Data Fluency Framework. Being a strong data consumer with growing data product authoring skills can help you distinguish yourself in the marketplace. Improving data product authoring skills is a way to separate yourself from the pack in an interview process or from your colleagues with respect to performance. Don't be discouraged if you only have a basic vocabulary or working knowledge in the language of data today. Work to become fully fluent or even bilingual in the preferred language of the knowledge economy.

Data Fluency Benchmarks for Individuals

ORGANIZATIONAL DATA FLUENCY MEAN SCORE (MEAN OF COMPONENTS 1-4)	LEVEL OF FLUENCY
0-25%	Basic Vocabulary
26%-50%	Working Knowledge
51%-75%	Partially Fluent
76%-90%	Fully Fluent
91%-100%	Bilingual

DFI SUPPORTING MATERIALS

Achieving a good response rate on a survey often depends on how it is introduced and by whom it is introduced. Before sending out the DFI survey to colleagues in your organization, consider who the recipients will see as the sender and when is a good time to gather an interested (or even captive) audience. The first few minutes of a town hall meeting or a weekly team meeting might be an ideal time to achieve a high response rate. Below is a sample introductory e-mail for the appropriate team member to use when sending out the DFI.

In this section we have also included a data literacy quiz to be used by your human resources team in the evaluation of job candidates. We have found that asking someone if they are data literate typically gets an affirmative response; the data literacy quiz can be used as part of the interview process to validate a candidate's general data skills.

SURVEY INTRODUCTION E-MAIL

The following is a sample e-mail for organizational or team leaders when sending out the DFI survey.

Dear team members,

We request your participation in a survey evaluating how we communicate with data as an organization. This survey can help our organization assess the four frames of the data fluency framework: data literacy, data product authoring, organizational culture, and data product ecosystem.

We believe that by having a team with strong data literacy skills and data authoring abilities, as well as an organizational culture supporting effective data communication, we can enhance our organizational performance. Effective data ecosystems make it easy for team members to find the data products and reports they need as well as provide the opportunity to build off the work of other team members.

Participation is voluntary and your answers are confidential. The link to the survey is here: [insert link]

Thank you,

(Name)

DATA LITERACY QUIZ

During an interview you can ask potential team members about their data fluency. Most will respond that they are comfortable working with data. A more effective interview might include a data fluency quiz. The following is a sample quiz for you to use or build upon:

- 1.** Ms. Smith received the following scores on a recent student formative assessment (scale: 1 to 20):

15, 14, 14, 18, 15, 17, 14, 15, 15, 14, 17, 16, 14, 15, 18, 19, 2

Which measure of central tendency should Ms. Smith use to summarize these grades?

- a.** Mean
- b.** Median
- c.** Mode
- d.** Standard deviation
- e.** None of the above

Answer: b. Median is not influenced by the outlier value of 2 like mean is; there are multiple modes. Standard deviation is not a measure of central tendency.

- 2.** A group of students are taking a new survey called *StudentView* about their teachers. The survey is administered three times during the school year. The students consistently rate their teachers with similar scores on the *StudentView* each time; however, these scores are different than scores the same teachers received on another student survey that is known to accurately describe their performance. You would most likely conclude that *StudentView* is which of the following?

- a.** Not a valid or reliable survey
- b.** A valid but not a reliable survey
- c.** A reliable but not a valid survey
- d.** A valid and reliable survey
- e.** None of the above

Answer: c. Reliability relates to the *consistency* of the results. In this case, *StudentView* yields a similar or consistent result with each administration. However, *StudentView* does not seem to be an *accurate* measure of teacher performance, as measured by the other student survey.

3. Student race/ethnicity is a good example of which of the following?

- a.** A nominal dimension
- b.** An ordinal dimension
- c.** A ratio
- d.** An interval
- e.** A continuous dimension

Answer: a. Nominal dimensions deal with qualitative data that is assigned a predetermined category that *cannot* be ranked. In this case, race/ethnicity cannot be ranked with a certain racial/ethnic group *above* another.

4. A histogram of teacher value-added scores would be *most* useful in examining which of the following?

- a.** The number of teachers
- b.** The mean value of teacher value-added
- c.** The median value of teacher value-added
- d.** The distribution of teacher value-added

Answer: d. Histograms are designed to examine the *distribution* of scores. And although the other information (that is, mean, median, and such) can be provided in a histogram, their primary function is to examine data distributions and spread.

5. Mr. Jones received his recent classroom observation score. He was told he scored in the *second quartile* in his grade. Assume that the following scores represent the values for the teachers in his grade:

7 5 3 10 15 6 12 9 14 16

Which score did Mr. Jones most likely receive?

- a.** 5
- b.** 10
- c.** 9
- d.** 3

Answer: c. 9

Rearranging the data in ascending order = 3 5 6 7 9 10 12 14 15 16:

The median value (50th percentile) = 9.5.

The median value of the lower half of the data (25th percentile) = 6.

Mr. Jones most likely scored between 6 and 9.5.

The best answer is 9.

6. For the previous scores, what is the *interquartile range*?

- a.** 8
- b.** 7
- c.** 5
- d.** 9

e. The interquartile range cannot be calculated with this data.

Answer: a. 8

The interquartile range measures the spread of the center half of the dataset, or $Q3 - Q1$.

$$14 - 6 = 8$$

7. A line chart of a student's test score performance would be an appropriate way to display which of the following?

- a.** The student's average test score
- b.** The student's overall grade
- c.** The student's growth over time
- d.** The student's abilities
- e.** You shouldn't use line charts for individual students.

Answer: c. Line graphs are best used to model growth or change over time.

8. To examine the relationship between student free/reduced lunch status and student test scores, you would most likely use which of the following?

- a.** A chi-square goodness of fit test
- b.** A one-way analysis of variance
- c.** A correlation coefficient
- d.** A Poisson probability distribution

Answer: c. Correlation coefficients measure both the strength and direction of a relationship, in this case, the relationship between free/reduced lunch status and test scores. The other statistical tests are appropriate for modeling differences between groups of students.

- 9.** Which graphical display is most appropriate for modeling the relationship described in question 8?
- a.** A box and whisker plot
 - b.** A stem and leaf plot
 - c.** A pie chart
 - d.** A scatterplot
 - e.** A stacked bar graph

Answer: d. Scatterplots are used to model the relationship between two variables.

- 10.** Student achievement on the state standardized exam approximates a normal distribution. Approximately what percentage of values lie within 1 standard deviation of the mean?
- a.** 50 percent
 - b.** 68 percent
 - c.** 75 percent
 - d.** 82 percent

Answer: b. The normal distribution is a bell-shaped, symmetrical curve. When this is the case, the *empirical rule* states that approximately 68 percent of the data values will be within one standard deviation.

APPENDIX A

Designing Data Products

The idea for this book grew out of many of the ideas we developed over the years writing our blog. In this appendix, we have edited together a collection of blog posts that discuss some important principles and concepts on designing data products.

Articles

A Checklist for Creating Data Products—An overview for product managers of data solutions to begin the process of transforming their organization's valuable data into customer-facing products. <http://bit.ly/data-product-checklist>

Think Like a Designer—A set of design principles for creating a data dashboard, including the use of unity/harmony, proximity/hierarchy, clear space, balance, contrast, proportion, and simplicity. <http://bit.ly/think-designer>

Designed to Be Used—A short list of design principles that we use to deliver applications that are valuable and productive for users. <http://bit.ly/designered-used>

Breaking Free of the One-Page Dashboard Rule—Many people believe a dashboard needs to be a one-page, single view interface. We offer a

different definition of data dashboards, focusing on clarity and conciseness.
<http://bit.ly/one-page-rule>

Dashboard Alerts Checklist—Alerts are a common feature in data products. This article summarizes our four C's for effective alerts: context, cogency, communication, and control. <http://bit.ly/dashboard-alerts>

8 Features of Successful Real-time Dashboards—Many operational dashboards require data that is updated in real time. We describe what it takes to make this type of dashboard work for users. <http://bit.ly/real-time-dashboards>

A CHECKLIST FOR CREATING DATA PRODUCTS

Data is the new oil.

—Everyone (first by Clive Humby)

Treat data like money.

—Jim Davis (SAS CMO) in *The Economist*

Are you sitting on a gold mine—if only you could transform your unique data into a valuable, monetizable data product?

Over the years, we've worked with dozens of clients to create applications that refine data and package the results in a form users will love. We often talk with product managers early in the conception phase to help define the target market and end-user needs, even before designing interfaces for presenting and visualizing the data.

In the process, we've learned a few lessons and gathered many useful resources. The checklist shown in Figure A-1 is divided into four sections:

- 1. Audience**—Understand the people who need your data.
- 2. Data**—Define and enhance the data for your solution.
- 3. Design**—Craft an application that solves problems.
- 4. Delivery**—Transition from application to profitable product.

	Actions	What we've learned	Resources to learn more	Completion
AUDIENCE	Target users vs. target buyers	Look beyond the analysts who use the data to find the people who see the value and have the budget to back up their interest.	7 Companies That Totally 'Get' Their Buyer Personas (HubSpot) bit.ly/Buyer_Personas	<input type="checkbox"/>
	Understand users' data fluency	You'll find a lot of variation in the users' ability to work with data. How can you thrill data junkies while supporting novices?	Data Is Useless Without the Skills to Analyze It (HBR) bit.ly/data-without-skills	<input type="checkbox"/>
	Fit users' workflow	An engaging solution fits into users' existing processes, delivers data in the right format, and is timely for decision-making.	Form and function in data delivery (Juice) bit.ly/data-form-function	<input type="checkbox"/>
	Inform user actions	Many solutions provide data that is interesting but doesn't connect to decisions. What information can people act on?	Choosing the Right Metric (Juice) bit.ly/30Days-Metric	<input type="checkbox"/>
	Balance flexibility vs. guided analysis	Too much flexibility puts the burden on users to find answers. Too little flexibility is frustrating and won't serve a broad audience.	How to Tell Stories with Data (E. Segel) bit.ly/stories-with-data	<input type="checkbox"/>
DATA	Get the right data	All data is not created equal. Find the best data with the most valuable insight.	More Meaningful Big Data bit.ly/meaningful-data	<input type="checkbox"/>
	Choose your metrics	Apply unique or specialized metrics that make the data highly valued in your industry.	Franken-measures (Juice) bit.ly/franken-measures	<input type="checkbox"/>
	Make your data special	Spend time with your data. Do the analysis and understand its value.	Big Data + Intimate Stories (Juice) bit.ly/data-intimate-stories	<input type="checkbox"/>
	Know what's fair game	Be aware of the laws and constraints around sharing your data internally and externally.	Privacy requirements (Teach Privacy) bit.ly/data-privacy-security	<input type="checkbox"/>
DESIGN	Define product architecture	Content modules, visualizations, and features are woven together into a structure to make it easy for users to accomplish their goals.	The Dribblisation of Design bit.ly/dribblisation	<input type="checkbox"/>
	Plan for data variation & outliers	The design needs to anticipate the worst-case scenarios for data. E.g. sparse data, null values, odd distributions, and long labels.	Random Data Generator bit.ly/random-data-gen	<input type="checkbox"/>
	Enable sharing	Users will want to share your data, collaborate with colleagues, and distribute their findings. Consider features for capturing and exporting user parameters, configurations, and annotations.	InVision's prototyping app has beautiful ways for sharing content bit.ly/invision-app	<input type="checkbox"/>
	Create logical flow	Provide a guided path for users to explore and find insights. Help your audience by giving them an obvious place to start and the visual indicators to deepen their analysis.	Guide to Dashboard Design, see Pt. 2 (Juice) bit.ly/guide-dash-design	<input type="checkbox"/>
	Write clear labels and description	Use plain, jargon-free language to explain the data. Craft titles, labels, and legends that explain the meaning of the data and descriptions of actions to be taken with the results.	8 Rules for Great Writing (K. Vonnegut) bit.ly/write-with-style	<input type="checkbox"/>
DELIVERY	Focus on customer development	Start small by thinking about discrete data applications that solve real customer problems and answer their questions.	Four Steps to the Epiphany (S. Blank) stanford.io/clo4lu	<input type="checkbox"/>
	Prioritize features	Novel and fun features can outpace their need. Make sure features line up with company's goals.	60 second business case (J. Brett) slidesha.re/S8YnvQ	<input type="checkbox"/>
	Find the right price	Ask users how they plan to use the product and assess the value they will derive.	KISS Metrics on Pricing bit.ly/kissmetrics-pricing	<input type="checkbox"/>
	Run experiments	When you're not sure (and you rarely will be), create an experiment, learn and move on.	Run Experiments to Tackle Risks (A. Maunya) bit.ly/run-experiment	<input type="checkbox"/>
	Reach customers	How do you get marketing budget without proven sales? Find early adopters, excel at service, and ask for referrals.	Why I Hate Funnels (B. Chestnet) bit.ly/hate-funnels	<input type="checkbox"/>

FIGURE A-1 Data product checklist

THINK LIKE A DESIGNER

By Ken Hilburn

Recently, Nancy Duarte participated in an interview with Jimmy Guterman of the MIT Sloan Management Review, which resulted in the article “How to Become a Better Manager by Thinking Like a Designer.” The quote that summarizes the article follows:

Often managers . . . rely heavily on data and information to tell the story and miss the opportunity to create context and meaning . . . leaving lots of room for interpretation, which can spawn multiple cycles and limit advancement.

It’s the same with information presentation. A focus on design at the beginning expands the ability to deliver context and meaning. But before you discount design as a concept for well, you know, “those artsy types,” keep in mind, as Nancy puts it:

Design is . . . crafting communications to answer audience needs in the most effective way.

What this means is that the more you focus on design, the more you’ll “speak” to your audience—which means you’ll be more effective with your data presentation. Remember, it’s about the audience, not you.

Here are some dashboard design principles that we use (with a few enhancements from Nancy’s interview) to make sure we become better information presenters by thinking like designers:

- **Unity/harmony**—A sense that everything in the application belongs together, resulting in a “whole” that is greater than the sum of the parts. All the elements complement, augment, and enhance, as opposed to distract and detract from each other.
 - **Takeaway**—Identify the problem you’re solving and make sure every element you place moves you closer to answering that question.
- **Proximity/hierarchy**—Things that are near each other are related. Hierarchy demonstrates relationships between items where appropriate. Proximity and hierarchy both provide tremendous contextual cues leading to better understanding.
 - **Takeaway**—Place related things near each other and separate unrelated things. Remember, dogs and cats don’t play well together.

- **Clear space**—White space in information display is important and too often overlooked. Maximizing dashboard real estate means creating places for the eye to “rest” so that the nonwhite space is more effective.
 - **Takeaway**—Use white space with proximity to help your viewers follow the story the information is telling.
- **Balance**—Dominant focal points either give the viewer a sense of comfort (balanced) or spur them to action (unbalanced). Nancy points out, “That does not mean all things must be in balance all the time. It is often effective to jar people and thereby effect a change in behavior or thought. Be aware, though, that once something has been thrown out of balance, it is the nature of the universe to find a new state of equilibrium.”
 - **Takeaway**—Make sure the primary focal points in your information presentation tell the viewer either “it’s ok, move on,” or “you need to do something.”
- **Contrast**—Contrast creates interest to focus attention or highlight differences. Again quoting from the article “The value of contrast lies neither in the black nor the white, but in the tension between them.”
 - **Takeaway**—Use Contrast to shift balance so the viewer focuses and acts more quickly.
- **Proportion**—More important elements deserve more real estate. It’s tempting to want to present an unbiased view of the data. However, as Amanda Cox of the NYT graphics department stated at the OEDC seminar “Innovative Approaches to Turn Statistics into Knowledge,” “Data isn’t like your kids, you don’t have to pretend to love them equally.”
 - **Takeaway**—Increase the size and emphasis of the values and decrease the size of labels and you’ll find dramatically better impact and speed of understanding.
- **Simplicity**—Stay focused on the specific fact on which you’re trying to shine light. This sometimes means showing *less* data and a simpler display. I think Garr Reynolds sums it up best: “Don’t confuse ‘simplicity,’ which is hard to achieve, with ‘simplistic,’ which is easy and usually lacking value.”
 - **Takeaway**—Help your viewers focus on what’s actually important by pointing them to the kernels and not the chaff.

DESIGNED TO BE USED

By Ken Hilburn

I have become curiously interested in a recent blog post that talks about how it's difficult to correctly write an application for the iPhone. The assertion is that writing software for the iPhone is harder than for a desktop, not because of the technology, but because "*everything counts so much — every design choice, every line of code, everything left in and everything left out.*"¹

Very eloquently and precisely put. If you've ever used any sort of mobile computing platform, not just the iPhone, you know how much proper design can make an application really useful—or totally useless.

But then again, isn't this the case with *any* application? Some applications seem to have their genesis in the charter "build an application that allows the user to perform all these actions," whereas others are built on the charge "build an application that helps the user solve this problem"—it's the battle of *functionality* versus *purpose*.

We put together a short list of some design principles that we use to keep the user productive:

- **Solve a problem**—Make sure the end product provides a specific solution to a specific problem so that users can easily understand how it helps them.
- **Enable casual use**—Minimize the "barrier to entry" for new users by avoiding feature overload, minimizing clicks for each task, and not letting polish become bling.
- **Tell a story**—Relate the data to the key questions, answering them in a logical order and revealing layers of detail as users express interest in knowing more, not before.
- **Lead to action**—Empower users to finish their task quickly. (The "task" is not "using software.")
- **Encourage exploration**—Use the experienced guide approach to give users enough context to understand the problem and then point them in the right direction to learn about new factors that can expand their insight.

BREAKING FREE OF THE ONE-PAGE DASHBOARD RULE

By Zach Gemignani

Conventional wisdom says that an executive dashboard *must* fit on a single page or screen. The argument hinges on a pair of assertions about this constraint: It provides necessary discipline to focus on only the most critical information, and it enables the audience to see results “at a glance.”

The “discipline” argument is made forcefully by Avinash Kaushik (among others).

If your dashboard does not fit on one page, you have a report, not a dashboard. . . . This rule is important because it encourages rigorous thought to be applied in selecting the golden dashboard metric.²

I buy wholeheartedly into the value of constraints. However, defining a useful constraint as a “rule” assumes there is only one viable means to achieve the desired ends. Confining visual real estate is just one way to focus your thinking. There are others: How about limiting yourself to five key measures? How about demanding that a dashboard can be understood in 3 minutes by a new user? How about presenting only exceptions?

The argument that a one-page dashboard necessarily provides a view of your business “at a glance” is more self-deceiving. Well-known information-ista Stephen Few uses this rationale in his definition of a dashboard:

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.³

I check my speedometer “at a glance.” I “glanced” at a Heads-up Display (HUD) on a video game showing how much energy my character has remaining. These displays communicate a single number that is already hovering on the corner of my consciousness. If we follow this advice literally, we’d show this (Figure A-2).

Assuming one page gives you quick, easy comprehension is like assuming all red cars are fast. That’s simply not true. More often, people follow the one-page dashboard rule off a cliff, cluttering



FIGURE A-2 A single dashboard metric

up the page with pie charts, line and bar graphs, and a series of gauges that display information from a variety of sources.

There are real problems with this definition of a dashboard, as shown in Figure A-3.

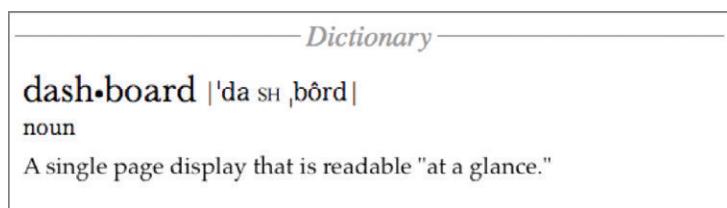


FIGURE A-3 The dictionary definition of "dashboard"

In reality, the one-page rule leads to jamming information into the available space. When everything must fit on a page, there isn't room to describe the connections between information or fashion a story from the data. A good dashboard raises more questions than it can answer. Sticking to a static piece of paper limits any ability to find or present explanations.

Don't get me wrong: A one-page dashboard is often an effective way to create "a visual display of the most important information needed to achieve one or more objectives." But with streaming video, interactive visualizations, podcasts, Kindles, smartphones, video projectors . . . is it necessary to limit ourselves to an 8.5" x 11" piece of paper? Or might we open ourselves up to some more creative solutions to sharing the numbers: a short movie, a few slides, a short text narrative, or 140 characters.

I'd like to use this definition instead, as shown in Figure A-4.

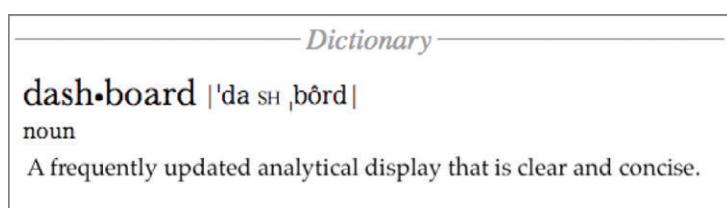


FIGURE A-4 A better dashboard definition

DASHBOARD ALERTS CHECKLIST

By Zach Gemignani

The tendency with reporting, and information dashboard design in particular, is to cram as much information on the page as possible. It is a problem that Avinash describes with typical candor:

This [is] one of the core reasons why most dashboards are “crappy,” i.e., they are data pukes that provide little in terms of context and even less in terms of actionable value.

In the past, we have offered tools to make data presentation as clear as possible (chart chooser and Excel chart cleaner). Sometimes, clean isn’t enough; a more dramatic approach is needed.

One alternative is to shift the focus from the full data to changes in the most critical data points. By pulling out the important exceptions, you can make it easier for your audience to digest what matters and take action.

Stephen Few says in his book *Information Dashboard Design* (O’Reilly Media, 2006):

The best way to condense a broad spectrum of information to fit onto a dashboard is in the form of summaries and exceptions . . . given the purpose of a dashboard to help people monitor what’s going on, much of the information it presents is necessary only when something unusual is happening; something that falls outside the realm of normality, into the realm of problems and opportunities. Why make someone wade through hundreds of values when only one or two require attention? We call these critical values exceptions.

Alerts are one mechanism to turn the focus to the exceptions, outliers, and data highlights. Whether embedded in the dashboard or presented separately, alerts can be the extra layer of abstraction that make a dashboard useful. Unfortunately, they are hard to get right. I’ve arrived at four C’s for effective alerts—context, cogency, communication, and control. Here’s a checklist to consider as you build alerts into a dashboard or report:

CONTEXT: USERS NEED TO UNDERSTAND HOW AN ALERT IS DEFINED AND HOW IT FITS INTO THE LARGER PICTURE

- **Are the parameters well defined?** An alert is commonly defined by the following factors: *metric* (for example, revenue), *dimension* (for example, time), *delta* (for example, month over month change), *scope* (for example, Northeast region, Peanut-product line), *threshold* (for example, increase or decrease of 10 percent).
- **Is the timing of the alerts actionable?** One client explained to us that fluctuations in many of its metrics make monthly alerts too frequent—it would unnecessarily alarm people when, from its perspective, no significant trend had been established.
- **Is the change statistically significant?** This is of particular importance when you are measuring deltas. A doubling of traffic from a referring site doesn't mean much when it is moving from one to two visitors.

COGENCY: AN ALERTING SYSTEM NEEDS TO AVOID CAUSING UNNECESSARY ALARM WHILE DELIVERING EASY-TO-UNDERSTAND INFORMATION THAT CAN BE ACTED UPON

- **Can the alerts be described in simple terms that even an executive can understand?** Alerts should have a real-world meaning that users are familiar with. If an alert is based on a complex metric, for example, users will be confused as to the implications.
- **Is the alert actionable?** In the best cases, alerts should point users to both the drivers of the alert and the actions that can address the situation.
- **Are the alerts so granular and/or frequently triggered that users will get alert fatigue?** Excessive use of alerts can undermine their credibility. We saw this happen at one client where an IT-designed system threw off alerts like they were going out of style. The application went out of style the next year when users decided it was more distracting than useful.

COMMUNICATION: ALERTS MUST BE DESIGNED TO EFFECTIVELY CAPTURE ATTENTION AND INFORM

- **Is the alert placed in context?** For example, Google Finance does a nice job of putting news alerts within the stock chart.
- **Is it clear what users should do next?** Give users a clear path to more information so that they can understand the full context of the alert.
- **Does the sophistication of your alerts match the sophistication of your audience?** I've found that it is better to start with some simple alerts so that your audience can begin to learn what it means and how to react. Over time, these alerts can become more refined and focused to capture complex situations.
- **Does the alert draw the eye without being visually overwhelming or annoying?** The visual noise in a dashboard can make it difficult to see what is important.
- **Is color used appropriately?** Red means bad. Yellow is sorta bad. Green means good. (But "good" things don't need to be alerts.) It isn't particularly fair for color blind folks, but these conventions are deeply rooted.
- **Have you found the best mechanism for presenting alerts?** Alerts can be sent through e-mail, as an SMS message, blasted over the office intercom system, or posted to the wall in the bathroom. What is the most convenient and appropriate medium?

CONTROL: ADVANCED ALERT SYSTEM SHOULD GIVE USERS THE ABILITY TO CUSTOMIZE AND MANAGE ALERTS

- **Can users identify the important alerts and avoid the others?** As hard as you may try in designing the dashboard or report, you aren't in the shoes of the users. They will learn what they want to pay attention to and what information is extraneous.
- **Can users adjust the parameters?** With more sophisticated dashboards, you want to give users the ability to adjust parameters to hone in on the exceptions that require action.
- **Can users analyze alert frequency and trends?** I've never seen a system that does this, but having the ability to view and analyze alert history seems critically important to getting a holistic view of performance.

8 FEATURES OF SUCCESSFUL REAL-TIME DASHBOARDS

By Zach Gemignani

Real-time dashboards—the types that show up on a big screen in a call center—are an entirely different beast than your standard management dashboard. Their job is to support immediate decision-making. As a result, the information must be easy to interpret, alert users to problems, and make the next action obvious. In addition to key success metrics, real-time dashboards may show detailed data about the action “on the ground.” Here are eight characteristics that can make a real-time dashboard effective:

- 1. A summary status that indicates how things stand overall.** Users need to tell at a glance whether they should worry. Here’s a great example from an iPhone app called Umbrella. The app provides the most important information you need from a weather report (Figure A-5).



FIGURE A-5 The direct approach to weather forecasts

- 2. Reflect a well-understood structure of the business.** By the time you design a real-time dashboard, you should have a strong theory for how the pieces of the business fit together (that is, the relationships between key measures, drivers, and available actions). For example, in the call center business, there are clearly defined success measures (for example, wait time), a mathematical relationship between these measures and their underlying drivers (for example, call volume), and known levers to address problems (for example, staffing levels).
- 3. Support quick diagnosis of problems.** The data presentation should point directly to the likely source of the problem. Real-time dashboards aren't the place for deep analysis or introspection into the drivers of the business.
- 4. Simple data presentation.** Real-time dashboards aren't the place for complex or advanced data visualizations. Imagine you were Napoleon and you had to use a half-completed version of Minard's graphic of the march to Russia to make a battlefield decision in the next 5 minutes.
- 5. Granular view of the “unit of action.”** Real-time dashboards are often about tracking activity. It may be useful to show the raw data around these events in the form of a ticker, scroll, or RSS feed. We use a real-time tracker for our website called Sitemeter. It does a nice job of tracking the basic unit of action—visitors (Figure A-6).

Juice Analytics Recent Visitors by Referrals	
Detail	Referring URL
[1]	unknown
[2]	http://www.google.com.au/search...excel&btnG=Google Search&meta=
[3]	http://www.portafolioblog.com/...antillas-de-powerpoint-gratis/
[4]	unknown
[5]	unknown
[6]	unknown
[7]	http://medicablogs.diariomedic...get/tag/plantillas-powerpoint/
[8]	unknown
[9]	http://www.google.com/reader/view/
[10]	http://www.google.com/search?h...fancy writing styles generator

FIGURE A-6 A granular view of traffic on our website

- 6. Appropriate time window.** Getting time right on an operational dashboard is critical. If the measures and trends represent too long a time period, users may not react to changes quickly enough. However, small time windows encourage frantic reactions to changes that may not represent real trends. Ideally, the dashboard should offer the ability to configure this time range and “freeze” a moment in time.
- 7. Prominent but balanced alerts.** Naturally, alerting users to problems is a central mission for real-time dashboards. The challenge (as always with alerts) is to balance between “the sky is falling” hysteria and “don’t worry, be happy” apathy. One item to emphasize is the need to show a sense of relative importance. Not all problems have the same impact on the business, and finding a way to communicate this relative importance is valuable.
- 8. Point to specific action.** If real-time dashboards are about identifying and responding to issues, the tool should point users to what they can do about a problem. This can be as simple as displaying the phone number of the right person to call.

Real-time dashboards can be ignorable, create mayhem, or drive great behavior in an organization. Thinking carefully about the design and functionality will make a huge difference.

NOTES

1. Brett Simmons, November 2008. See <http://inessential.com/2008/11/>
2. Avinash Kaushik, March 2007. See <http://www.kaushik.net/avinash/five-rules-for-high-impact-web-analytics-dashboards/>
3. Stephen Few, February 2006. See http://www.perceptualedge.com/articles/Whitepapers/Common_Pitfalls.pdf

APPENDIX B

Style Guide

A Data Presentation Style Guide is a shared document that helps everyone in an organization create high-quality, consistent data presentations. The document provides a starting point for anyone creating a report, dashboard, or analytical interface regardless of the tool being used. In fact, the style guide helps to smooth out the inevitable differences across tools. The following table of contents provides a sample of the content that may be covered in a style guide.

- **Chart selection**—Guidelines for selecting the right chart for the data and your message.
- **Layout**—Design principles for laying out charts and other content on the page, including using gridlines and whitespace.
- **Typography**—Font sizes and weights for headings, body, labels, and notes text.
- **Color**—Categorical, sequential, and divergent color schemes to be used in charts.
- **Number formatting**—Formats for dates, decimals, currencies, and large numbers.
- **Basic charts**—Bar, line, scatterplot, bubble, and pie.

- **Tables**—Formats for table text, borders, headers, text alignment, and emphasis.
- **Advanced charts and graphics**—Styling for maps, bullet charts, two-axis charts, sparklines, treemaps, and other advanced visuals. Guidance on when to use these graphics and which ones to avoid.
- **Common widgets**—Key metrics, selectors, sliders, input areas, and tooltips.
- **Legends and labels**—Rules and styling for chart legends and labels.

This appendix includes six sample pages based on style guides that we have produced for diverse organizations. These style guide examples are intended to jump start your process for creating your own style guide.

STYLE GUIDE SAMPLE 1: FONTS

TYPOGRAPHY

Use Open Sans if it available. Otherwise use Arial

OPEN SANS

Title text

Short description of what is shown in the chart.

Size: 18 pt, 80% bigger
than body text
Weight: Semibold

Color: #333333
RGB: 51, 51, 51
80% black

ARIAL

Size: 18 pt, 80% bigger
than body text
Weight: Regular

Color: #333333
RGB: 51, 51, 51
80% black

Subhead text

Used when there are multiple charts under one title and each one needs a smaller title/description

Size: 12 pt, 20% bigger
than body text
Weight: Semibold

Color: #666666
RGB: 102, 102, 102
60% black

Size: 12 pt, 20% bigger
than body text
Weight: Bold

Color: #666666
RGB: 102, 102, 102
60% black

Body text

Body text is all other written text, such as longer descriptions below the subhead to give additional information

Size: 10 pt
Weight: Regular

Color: #666666
RGB: 102, 102, 102
60% black

Size: 10 pt, 10% bigger
than body text
Weight: Regular

Color: #666666
RGB: 102, 102, 102
60% black

Source, notes, axis,
legends, and data labels

Size: 9 pt, 10%
smaller than body text
Weight: Light

Use notes to annotate key details about the data that aren't clearly displayed in the chart.

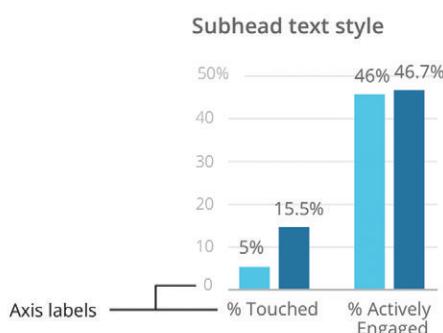
Color: #999999
RGB: 153, 153, 153
40% black

Size: 9 pt, 10% smaller
than body text
Weight: Regular

Color: #999999
RGB: 153 153, 153
40% black

Title text style

Body text style



Source/notes labels — Source: XXXX

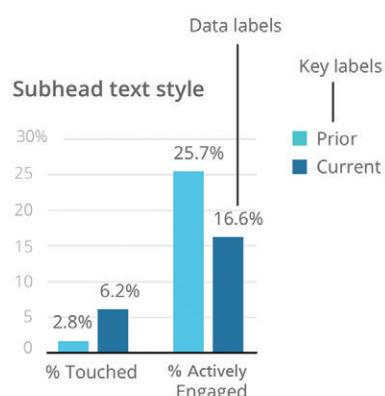


FIGURE B-1 By setting a few rules around choices of font families, sizes, and colors, the style guide encourages consistent, readable data products.

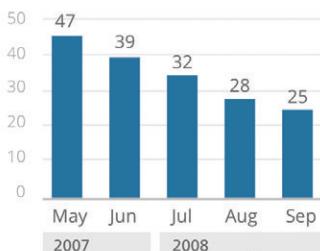
STYLE GUIDE SAMPLE 2: COLORS

COLOR

Basic colors

The main bar/line chart color:

█ RGB: 48, 116, 158



Year background color:

#E6E6E6
RGB: 230, 230, 230

10% black
14 pts tall

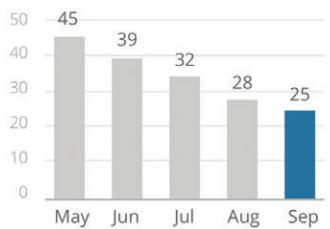
Highlight bars and/or to text draw attention to pertinent information:

Highlighted bar
(or text):

█ RGB: 48, 116, 158

Other bars:

#BDBDBD
RGB: 176, 176, 176
30% black



Color schemes

There are 3 types of color schemes that can be used.

Sequential Colors

Use this group of colors when you are ordering values from low to high.

█ RGB: 93, 48, 14

█ RGB: 186, 95, 27

█ RGB: 232, 119, 34

█ RGB: 241, 173, 122

█ RGB: 246, 201, 167

Divergent Colors

Use this group of colors when the values are ordered and there is a critical mid-point (e.g. an average or zero).

█ RGB: 93, 48, 14

█ RGB: 186, 95, 27

█ RGB: 232, 119, 34

█ RGB: 246, 201, 167

RGB: 230, 230, 230

█ RGB: 148, 203, 198

█ RGB: 79, 163, 156

█ RGB: 13, 119, 110

█ RGB: 9, 79, 73

Categorical Colors

Use the colors below when data falls in distinct groups and therefore requires contrast between adjacent colors.

█ RGB: 232, 119, 34

█ RGB: 238, 187, 51

█ RGB: 161, 165, 52

█ RGB: 79, 163, 156

█ RGB: 46, 98, 150

RGB: 176, 176, 176

█ RGB: 109, 68, 145

█ RGB: 237, 129, 154

FIGURE B-2 The style guide provides a core emphasis color as well as color schemes to support sequential, divergent, and categorical needs in charts.

STYLE GUIDE SAMPLE 3: DATE/NUMBER FORMATTING

NUMBER FORMATTING

Data type	Notes	Format options
Number in thousands	Use 'k' to shorten the number of digits	32,924 should be 32.9k.
Number in millions	Use 'M' to shorten the number of digits	10,456,209 should be 10.5M.
Number in billions	Use 'B' to shorten the number of digits	10,456,209,009 should be 10.5B
Using year as label	Write out the year, unless there isn't enough space. Then abbreviate.	Preferred: 2012 Limited space: '12
Using month and day and possibly year as label	Do not include the year unless time span is more than one year. Be sure to note the year somewhere else though, such as in the title. Always spell out the first 3 letters of the month unless space is limited.	Preferred: Jan 5 Jan 5, 2012 Limited space: 1/5 1/5/12
Using month as label	Spell out the first 3 letters of the month unless space is limited.	Preferred: Jan Limited space: J
Decimals	Round all numbers unless there is a specific reason to. Then, keep it to as few decimals as necessary to show differences between values.	32.9241 should be 32.9 if this provides enough precision
Using month and year as label	See below	

If it is a month and year range: Jan-Sep '11

If it is too crowded: 1/11-9/11

Reach rate

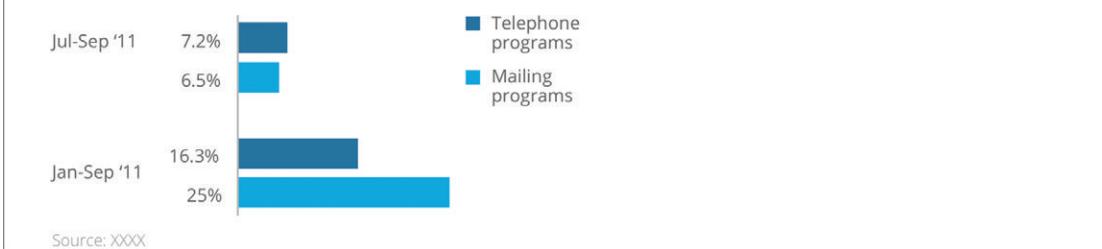


FIGURE B-3 Numbers' formats encourage "shortening" to make large numbers more readable.

STYLE GUIDE SAMPLE 4: BAR CHARTS

BAR CHARTS

Use a bar chart when comparing groups of data. There are many types:

1. **Standard bar chart** for comparing multiple groups of data.
2. **Stacked bar chart** for comparing breakdowns and totals for multiple groups of data. Also for percentages of a whole as an alternative to pie charts.
3. **Paired bar chart** for comparing 2 sets of data with the same time frame, or other grouping.
4. **Average comparison bar chart** for comparing current/prior data to a norm, or a set of data to an average.
5. **Bullet chart** for comparing data to a bad/good/great range and/or an optimal amount.

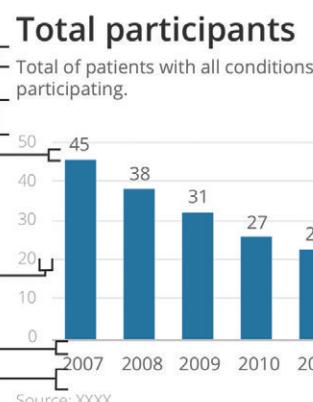
1. Standard bar chart

Vertical bar chart

There are 10 pts between the title and body text.

Total participants

There are 20 pts between the bottom of the body text and



There are 5 pts between the top of the bars and the bottom of the data label text

There are 8 pts between axis lines and the y axis labels, as well as between the baseline and the x axis labels.

There are 12 pts between the bottom of the x axis labels and the source text.

Bars*

 RGB: 48, 116, 158
17 pts wide
15 pts between vertical bars

Axis lines

 RGB: 230, 230, 230
10% black
1 pt stroke

Baseline

 RGB: 179, 179, 179
30% black
1 pt stroke

Horizontal bar chart

Use a horizontal bar chart when labels are too long to read clearly in a vertical bar chart.

If there is no body text between the title and chart, there are 15 pts between the bottom of the title and the

Total enrolled by Diagnosis

Breast (159) 32%

Prostate (58) 12%

Colon (28) 6%

Ovarian (26) 5%

Lung and Bronchus (23) 5%

There are 10 pts between horizontal bars.

Source: XXXX

* For information on highlighting a particular bar, refer to Basic Colors, page 3.

FIGURE B-4 The style guide describes the individual elements and layout of common chart types.

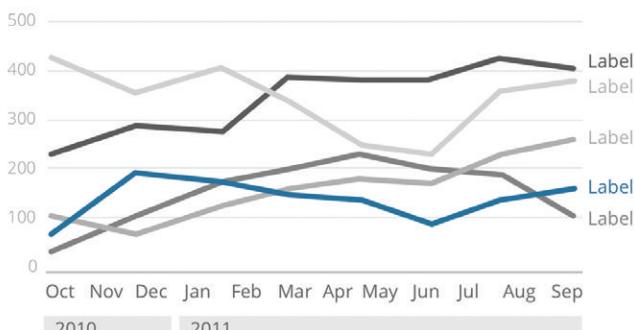
STYLE GUIDE SAMPLE 5: TREND CHARTS

TREND CHARTS

Line chart Use a line chart to show trends over time. For spacing information, refer to page 9, Bar Charts.

Chart title

Lead in here with more explanation if needed.



Source: XXXX

Trend line

■ RGB: 48, 116, 158
3 pt stroke

Axis lines

■ RGB: 230, 230, 230
10% black
1 pt stroke

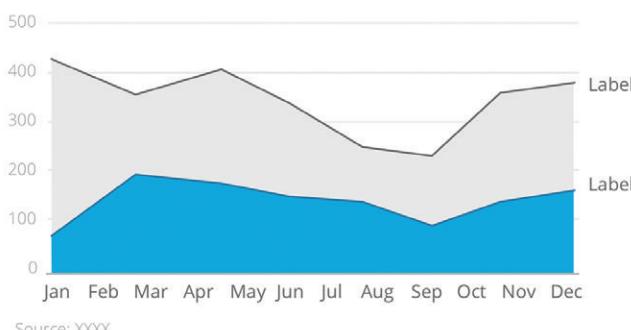
Baseline

■ RGB: 179, 179, 179
30% black
1 pt stroke

Area chart Use an area chart to show cumulated totals over time.

Chart title

Lead in here with more explanation if needed.



Grey trend line

■ RGB: 92, 92, 92
65% black
3 pt stroke

■ RGB: 230, 230, 230
10% black

Orange trend line

■ RGB: 48, 116, 158
3 pt stroke

■ RGB: 35, 95, 142

FIGURE B-5 Guidelines for trend charts cover how to color and label the lines.

STYLE GUIDE SAMPLE 6: TABLES

TABLES

Table title
Lead in here with more explanation if needed.

Jul. 11, 2011 - Sept. 11, 2011			Jan. 11, 2011 - Sept. 11, 2011			
Program	# of Touched Members	% of Eligible Members	Norm	# of Touched Members	% of Eligible Members	Norm
Item 1	Metric 1	Metric 1	Metric 1	Metric 1	Metric 1	Metric 1
Item 2	Metric 2	Metric 2	Metric 2	Metric 2	Metric 2	Metric 2
Item 3	Metric 3	Metric 3	Metric 3	Metric 3	Metric 3	Metric 3
Item 4	Metric 4	Metric 4	Metric 4	Metric 4	Metric 4	Metric 4
Item 5	Metric 5	Metric 5	Metric 5	Metric 5	Metric 5	Metric 5

Legend:

- Blue square: RGB: 48, 116, 158
- Text box: Size: 12 pt, 20% bigger than body text, Weight: 65 Bold, RGB: 77, 77, 77, 70% black
- Vertical line: Size: 12 pt, 20% bigger than body text, Weight: 55 Regular, Color: #666666, RGB: 102, 102, 102, 60% black
- Horizontal line: 40 pts of space between dividing lines
- Dividing line: Dividing line
- Gray square: RGB: 230, 230, 230, 10% black, 1 pt stroke

FIGURE B-6 Well-formatted tables make it possible for readers to understand the hierarchy of the data and help reduce visual clutter.

Index

A

Academic Insights, 13
accountability, 30
accountability test, 141
accuracy, 75, 219
actionable data, 74, 77
actionable metrics, 92–95, 136
Ad Age, 82
alerts
 cogency, 232
 color, 107, 233
 communications, 233
 context, 232
 control, 233
 dashboard, 231–233
 guiding attention, 122
 real-time dashboards, 236
alignment, 30
Amazon Appstore, 190
Amazon.com, 131, 190
anecdotes, 157
Annual Yearly Progress (AYP), 92
AOL, 34
Apple, 27, 166, 171, 189–190
Apple App Store, 165–167, 171, 189–190, 207
appleinsider, 189
As Consciousness Is Harnessed to Flesh
 (Sontag), 91
atomic data, 60–61
attention
 data product structure, 100
 guiding, 122, 123
audiences, 20
 data products, 86–87
 dialogue, 114
 needs and perspectives, 86
automated pushing, 176

B

balance, 227
Balkanized data, 36–38
bar charts, 65, 105, 242
baseball, 111
battery strength indicator, 53
Beane, Billy, 6
Bell, Tyler, 6
Berger, Warren, 169
The Best American Infographics, 64
Best College list, 11
bias, 146
big data, 1, 5–7, 34
Bill & Melinda Gates Foundation, 150
Black River Entertainment, 183–184
blogs, 176, 223
Bolt, Usain, 70
Bonds, Barry, 111
Brady, Stephen, 183–185
"Breaking Free of the One-Page Dashboard
 Rule" (Gemignani), 229–230
browsing, 176
bubble charts, 65, 70, 73–74
Bureau of Labor Statistics, 145
business intelligence solutions, 2, 7
 take-up rates, 170
Byrne, David, 64

C

cardinal number, 135
Carnegie Classification of Institutes of Higher
 Education, 12

- case studies
 - fantasy football, 8–11
 - insurance company bottom lines, 15–17
 - school district woes, 13–15
 - U.S. News & World Report*, 11–13
- casual use, 122, 228
- categorizing information, 12
- causation, 56
- Center for Medicare and Medicaid Services (CMS), 50
- Chart Chooser, 106, 107
- chart junk, 65, 115, 116
- charts, 64–65
 - bar, 65, 105, 242
 - bubble, 65, 70, 73–74
 - choosing, 104–106
 - column, 65
 - insights expressed, 69–74
 - labeling, 66, 111
 - line, 65, 221, 243
 - pie, 65–66, 69, 70, 105
 - scatterplot, 65, 70, 222
 - style guide, 237, 238, 242–243
 - trend, 243
 - types, 65
- Chen, Boris, 9
- clarity, 231
- clear space, 227
- CMS. *See* Center for Medicare and Medicaid Services
- Coats, Emma, 102
- cogency, 232
- Coleridge, Samuel Taylor, 1, 154
- Collins, Jim, 133
- color, 107–109
 - alerts, 107, 123, 233
 - guiding attention, 122, 123
 - style guide, 237, 240
 - variants, 120–121
- Colton, Charles Caleb, 172
- column charts, 65
- columns, 64
- Comcast, 44
- communication, 30
 - alerts, 233
 - goals, 131, 139
 - leadership, 129–134
 - metrics, 131, 139
- compactness, 121
- comparison, 12, 71–72
 - metrics, 140
- Competing on Analytics* (Davenport), 38, 185
- complex metrics, 93
- comprehensibility, 65–66
- comScore, 145
- confounding factors, 56
- constraints, 3–4, 229
- content reviews, 180
- context, 9–10
 - alerts, 232
 - data product design, 111–112
 - metrics, 136
- contrast, 117, 227
- control, 233
- Conway, Melvin, 36
- Conway's Law, 36–37
- correlation, 56
- correlation coefficients, 221
- Cox, Amanda, 88, 227
- CSV files, 50
- culture, 25–26. *See also* data fluent culture
 - effects, 128–129
 - leadership, 129–134, 214
 - learning, 30
 - report proliferation, 35–36
 - spreadsheet-driven, 3
- customer
 - centric metrics, 140
 - engagement, 149
 - retention, 45
 - satisfaction, 26, 131
- customer service metrics, 149

D

- D3.js, 173
- Daniel, Carter, 113

- Dartmouth Atlas of Health Care, 145
 "Dashboard Alerts Checklist" (Gemignani), 231–233
 dashboards, 2, 97
 alerts, 231–233, 234–236
 appropriate information, 39
 comparisons in, 72
 constraints, 229
 context, 111
 features, 234–236
 granularity, 235
 problem diagnosis, 235
 real-time, 234–236
 reflecting priorities, 68
 structure, 235
 supermodel, 38–39
 time window, 236
- data
 actionable, 74, 77
 atomic, 60–61
 Balkanized, 36–38
 celebrating effective use, 132
 constraints on using, 3–4
 dialects, 26
 evaluating effective use, 157–158
 granularity of, 68
 language of, 60–77
 last mile problem, 4
 problems using, 3–4
 rate of production, 2
 relationships in, 68
 reliability, 74–77
 scope of, 68
 selection, 76
 siloed, 37
 solutions from, 186
 standard forms for collecting, 143–144
 summarized, 61–62
 validity, 74–77
- data authors, 79, 80, 83, 125–126, 188
 goals, 67
 leadership, 213
 Ogilvy guidelines, 91–92
- skills inventory, 200–204
 skillset, 80–81
 training, 171
 data care, 42–43
 data coaches, 13
 data consumers, 152–153, 187–188
 data literacy, 23–24, 196–199, 212
 fantasy football, 8–11
 school districts, 13–15
- data dictionary, 143
 data discovery gap, 175
 data elitism, 38–39
 data experiences, 13
 data fields
 common understanding, 68, 144
 non-useful, 64
- data fluency, 4–5, 185–187
 data producers, 24–25
 foundational elements, 20, 21, 22
 consumers, 23–24
 culture, 26–27
 ecosystem, 27–28
 individuals and organization, 22
 producers, 24–25
 using and presenting, 23
- need for, 2–4, 190
 pitfalls, 35–45
 Balkanized data, 36–38
 data care, 42–43
 data elitism, 38–39
 metric fixation, 43–45
 report proliferation, 35–36
 searching for understanding, 40–41
 supermodel, 38–39
- data fluency framework, 7–17, 21–28
 Balkanized data, 37
 data authors, 80
 data care, 43
 data elitism, 38
 finding balance, 45–46
 metric fixation, 45
 report proliferation, 36
 resources, 28–29

- searching for understanding, 41
- supermodel, 39
- using, 30–31
- Data Fluency Inventory (DFI), 30, 193–194
 - data literacy quiz, 219–222
 - scoring guide
 - individuals, 215–217
 - organizations, 210–215
 - supporting materials, 218–222
 - survey introduction e-mail, 218
 - survey questions, 194–209
 - data consumer literacy, 196–199
 - data fluent culture, 204–206
 - data product author skills, 200–204
 - data product ecosystem, 207–209
- data fluent culture, 128–129, 189
 - data informed decisions, 133–134
 - developing, 129–130
 - everyday activities, 151–158
 - data consumers, 152–153
 - data products, 153–156
 - data usage, 156–158
 - evolution, 158–159
 - inventory, 204–206
 - key metrics, 134–141
 - communicating goals, 139
 - employee contributions, 138–141
 - goal communication, 139
 - identifying good, 134–141
 - organizational improvement, 137–138
 - pitfalls, 139–141
 - shared understandings, 148
 - leadership, 129–134, 214
 - expectations, 130–131
 - indicators, 131
 - shared understandings, 141–151
 - data product purpose and motivation, 149–151
 - data sources, 145–146
 - key metrics, 148
 - transparency, 147–148
 - vocabulary and terminology, 143–144
 - data fluent organizations, 28
 - benefits, 29–30
- discussion, 178
- distilling information, 179
- plans, 30–31
- The Data Journalism Handbook*, 67
- data literacy, 20, 23–24, 196–199
 - skills, 48, 49
- data literacy quiz, 219–222
- data loss incidents, 54
- data metaphors, 5–7
- data mining, 2
- data phobia, 3, 4
- Data Presentation Style Guide, 237–244
- data presentations, 50, 231, 235
- data producers
 - data fluency, 24–25
 - insurance companies, 15–17
 - roles of, 25
 - U.S. News & World Report*, 11–13
- data product authors. *See* data authors
- data products, 49–59
 - audience, 86–87
 - barriers to using, 55–58
 - inconsistency, 58
 - jargon, 55–56
 - not knowing where to start or focus, 57–58
 - celebrating effective, 132
 - celebrating quality, 156
 - characteristics, 50
 - checklist for creating, 224–225
 - classification, 163
 - core principles, 83
 - attractive and easy-to-understand design, 103–113
 - dialogue, 114–115
 - information discrimination, 88–92
 - meaningful and actionable metrics, 92–95
 - purpose and message, 84–88
 - structure and flow, 95–102
 - design
 - color, 107–109
 - context, 111–112
 - form, 103–104

- language, 112–113
- principles, 121–125
- typography, 109–110
- visualizations, 104–106
- dissecting, 66–77
 - actionable, 77
 - insights in, 68–74
 - sources, 67–68
- ecosystem, 189–190
 - centralized inventory, 176
 - demand, 167–169
 - design objectives, 170–172
 - develop, 172–174
 - discovery, 175–176
 - discussion, 177–179
 - distilling, 179
 - document guidance, 171
 - grassroots needs, 169
 - inventory, 207–209
 - necessary conditions, 163–165
 - prioritizing needs, 168
 - review process, 171
 - Six Ds, 164, 165, 181, 194, 207
 - style guide, 172
 - tools, 173
 - top-down demand map, 167–169
- establishing guidelines, 153–154
- everyday, 51–55
- feedback, 155–156
- information delivery, 162–163
- inventory, 31
- letting data speak, 85
- as living documents, 179
- objectives, 85–86
- purpose and motivation, 149–151
- qualities of good, 162
- workflow fit, 174
- data security breaches, 54
- data source, 50
 - complex metrics, 93
 - credible and reliable, 145
 - data fluent culture, 27
 - data tables from, 64
- defining, 49
- examining data products, 67–68
- good metrics, 135
- internal, 145
- outside, 145
- shared understandings, 141, 145
- trusted, 145, 150
- data storytelling, 82, 95
- data tables, 63–64
- data usage, 156–158
- data visualization, 69, 82, 104–106
- data warehouses, 2
- database administrator (DBA), 16
- data-driven decisions, 1, 156–157
- DataHero, 173
- Data-Information-Knowledge-Wisdom Pyramid (DIKW Pyramid), 178
- data-to-ink ratio, 115
- Datran Media, 122
- Davenport, Tom, 38, 185
- Davis, Jim, 224
- DBA. *See* database administrator
- decision making, 29, 33, 186
 - data informing, 133–134
 - data-driven, 1, 156–157
- demand, 167–169
- Department of Education, 145
- design
 - data product ecosystems, 170–172
 - document guidance, 171
 - incorporating in development, 171
 - poor, 170
 - review process, 172
 - style guide, 172
- design principles, 115–125
 - data products, 121–125
 - casual use, 122
 - compactness, 121
 - customization, 123, 173–174
 - explanation before information, 124–125
 - gradual reveal, 121–122
 - guiding attention, 122

lead to action, 123–124
 modularity, 121
 objectives, 170–172
 visualizations, 115–121
 chart junk, 115, 116
 color variants, 120–121
 contrast, 117
 data-to-ink ratio, 115
 gradients, 119
 labels, 117
 repetition, 118
 smoothing, 119
 sorting, 120
 3-D effects, 119
 "Designed to be Used" (Hilburn), 228
 development, 172–174
 DFI. *See* Data Fluency Inventory
 dialects, 26
 dialogue, 114–115
 DIKW Pyramid. *See* Data-Information-Knowledge-Wisdom Pyramid
 dimensions, 60
 discussion, 177–179
 distribution, 220
 distributions of metrics, 140
 document guidance, 171
 DolNeedanUmbrella.com, 87–88
 Dong Nguyen, 165
 driving directions, 52
 Drucker, Peter, 167, 181, 185
 Duarte, Nancy, 226

E

eBay, 165
The Economist, 224
 ecosystem, 27–28
 edge cases, 141
 Ehrenberg, Ronald, 11
 "8 Features of Successful Real-Time Dashboards" (Gemignani), 234–236
 Einstein, Albert, 87, 154

electrocardiograms (EKGs), 71
 element, 60
Empire (film), 47, 49
 Empire State Building, 47, 48
 empirical rule, 222
 end-user customization, 123, 173–174
 ESPN, 10, 147
 Etzkorn, Irene, 84
 everyday data products, 51–55
 Excel, 106, 117, 172, 173, 174
 expectations, 130–131
 experimenter bias, 146
 explanation, 124–125
 exploration, 228
 external validity, 145

F

face validity, 76
 Facebook, 33
 fantasy football, 8–11, 67
 feedback, 155–156, 180
 Few, Stephen, 97, 106, 174, 231
 Fisher, R.A., 152
 fitness function, 131
 The Five Whys, 169
Flappy Bird (game), 165–166
 flow, 95–102
 flow-based structure, 97–98
 FlowingData.com, 54
 focal points, 227
 fonts
 chart labels, 66, 117
 choosing, 110
 guiding attention, 122
 style guide, 239
 Football Outsiders, 10
 Forbes, 12
 Ford Model T, 1
 form, of data products, 103–104
 formatting numbers, 237, 241
 forms, standard, 143–144

Forster, E.M., 181
 The Foundation for Scientific and Industrial Research (SINTEF), 2
 four C's, 231–233
 Fried, Jason, 151
From Data to Story: Dissecting a Well-Made Visualization, 82
 Fruition, 173
 functionality *versus* purpose, 228

G

Gallo, Amy, 34
 Gartner, 7, 170, 176
 Gemignani, Zach, 229, 231, 234
 generalizability, 56, 145, 146
 goals
 context from, 111
 data fluency improving, 21
 of data product authors, 67
 data product focus, 169
 metric alignment, 134
 metric selection, 138
 metrics communicating, 131, 139
 metrics contrary to, 93
 trending metrics, 140
Good to Great (Collins), 133
 Google, 6, 133
 Google Analytics, 62
 Google Play, 190
 gradients, 119
 gradual reveal, 121–122
 grassroots expectations, 158
 grassroots needs, 169
 grids, 100–101
 grouping, 99
 The Guardian, 53
 guided conversations, 82–83
 guided storytelling, 163
 Guidestar.org, 150
 guiding attention, 122, 123
 Guterman, Jimmy, 226

H

Habitat for Humanity, 150
 harmony, 226
 Harris, Harlan, 162
Harvard Business Review, 34, 189
 Hawthorne effect, 146
 HCAHPS. See Hospital Consumer Assessment of Healthcare Providers and Systems
 Heads-up Display (HUD), 229
 health concierge, 40
 hierarchy, 226
 Hilburn, Ken, 226–227, 228
 histogram, 65, 220
 historical conventions, 93
 history, 146
 Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS), 44–45
 hospital-acquired sepsis, 44
 "How to Become a Better Manager by Thinking Like a Designer" (Guterman), 226
 HUD. See Heads-up Display
 Humby, Clive, 224

I

IBM, 3
 IDG Enterprise, 34
 inconsistency, 58
 indicators, 131
 information, 124–125
 categorizing, 12
 dashboards, 39
 delivery with data products, 162–163
 design, 170
 discrimination, 88–92
 distilling, 179
 Information Age, 1
 data fluency need, 2–4
Information Dashboard Design (Few), 97, 231
 information discrimination, 88–92

"Innovative Approaches to Turn Statistics into Knowledge" (seminar), 227
 inspiration, 90
 Instagram, 2
 instrumentation, 146
 internal validity, 145
 interquartile range, 221
iOS Human Interface Guidelines (Apple), 166
 iPhone, 166, 228

J

Jaquith, Andrew, 135
 jargon, 55–56, 143
 Jeffries, Adrienne, 44–45
 Johnson, Samuel, 175
 Juice Analytics, 2, 13
 Jung, C.G., 40

K

Kaushik, Avinash, 229, 231
 Kerr, Gordon, 183–185
 key metrics, 2, 111, 191
 data fluent culture, 134–141
 employee contributions, 138–141
 fantasy football, 10
 goal communication, 139
 identifying good, 134–136
 organizational improvement, 137–138
 pitfalls, 139–141
 shared set, 148
 key performance metrics (KPIs), 77

L

labeling, 111
 charts, 66, 117
 clear, 117
 style guide, 238
 Lang, Andrew, 157

language
 of data, 60–77
 design, 112–113
 fluency, 74
 word choice, 153–154
 last mile problem, 4
 layout, 237
 leadership
 celebrating effectiveness, 132
 communication, 129–134
 data consumer literacy, 212
 data fluent culture, 129–134, 214
 data informing decisions and actions, 133–134
 data product authors, 213
 data product ecosystem, 214–215
 expectations, 130–131
 indicators, 131

learning culture, 30

legends, 238

Leonard, Elmore, 90

Lewis, C.S., 153

line charts, 65, 221, 243

Lithwick, Dahlia, 55

living documents, 179–181

Lovecraft, H.P., 91

Lucas, George, 20

M

Major League Baseball, 111
 maturation, 146
 mean, regression to, 56, 146
 meaning

 color, 107

 context, 111–112

meaningful metrics, 92–95

measurement

 clear definitions, 143

 reliability and validity of, 75

measures of central tendency, 219

median, 219

- menus, 52
 - message, 84–88
 - metadata, 176
 - metric fixation, 43–45
 - metrics, 60–61
 - accountability test, 141
 - actionable, 92–95, 136
 - alerts, 232
 - choices, 67
 - common mistakes in choosing, 93
 - comparisons, 140
 - complex, 93
 - context, 136
 - customer service, 149
 - customer-centric, 140
 - distributions, 140
 - edge cases, 141
 - framework for choosing, 93–95
 - goals communicated with, 131, 139
 - key, 2, 111, 134–141, 191
 - KPIs, 77
 - letting go, 141
 - meaningful, 92–95
 - organizational improvement, 137–138
 - pitfalls, 139–141
 - self-serving, 141
 - simplistic, 93
 - trending, 140
 - vanity, 93
 - Microsoft, 172
 - Microsoft Office, 174
 - MIT Sloan Management Review, 226
 - modeling
 - analytics, 185
 - customer behavior, 38
 - misalignment with environment, 38
 - probabilistic, 183–184
 - regression, 56
 - transparency, 147–148
 - modularity, 121
 - Monks, James, 11
 - Montessori, Maria, 161
 - Montessori educational system, 25, 161
 - A More Beautiful Question* (Berger), 169
 - Morse, Bob, 12, 13
 - moviemaking, 19–20
 - My Fit Engine, 13
- N**
- necessary conditions, 163–165
 - The New York Times*, 53, 88, 174
 - NFL, 8–10
 - NFL Passer Rating, 147
 - No Child Left Behind, 92, 139
 - nominal dimensions, 220
 - normal distribution, 222
 - number formatting, 237, 241
- O**
- OAA. *See* Office of Assessment and Accountability
 - Oakland A's, 6
 - Objective-C, 166
 - objectives
 - data discovery, 176
 - data product design, 170–172
 - of data products, 85–86
 - discussion, 177–178
 - feedback, 180
 - Office of Assessment and Accountability (OAA), 37
 - Ogilvy, David, 91
 - ordinal numbers, 135
 - O'Reilly, 6
 - organizational constraints, 3–4
 - organizational dysfunction, 3, 4
 - organizational improvement
 - analytics for, 185
 - metrics driving, 137–138
 - organizational plan, 30–31
 - organizing pages, 99
 - over-indexing, 56

P

The Paris Review, 91
 patient experience, 44–45, 73
People Magazine, 12
 personnel constraints, 3, 4
 Pew, 56
 pie charts, 65–66, 69, 70, 105
Pixar Rules for Storytelling, 102
 podcasts, 176
 Poe, Edgar Allan, 91
 Power BI, 172
 PowerPoint, 16, 24, 50, 106, 130, 172–174
 precision, 75
 prepared environment, 161
 pretest, 146
 proactivity, 149
Pro-Capitalist Society (Drucker), 181
 processes, 28
 Profootballtalk.com, 147
 proportion, 227
 ProPublica, 53
 proximity, 226
 purpose, 84–88

Q

QBR. *See Total Quarterback Rating*

R

radio promotion, 183
Raiser's Edge, 42
 ranking, 9, 12
 reactive effects, 146
Reader-Friendly Reports (Daniel), 113
 real-time dashboards, 234–236
 regression model, 56
 regression to mean, 56, 146
 relationships
 in data, 68
 data product structure, 98
 patterns, 68
 unexpected, 70

reliability, 146, 219
 data, 74–77
 data sources, 145
 measurement, 75
 repeated testing, 146
 repetition, 118
 report proliferation, 35–36
 review process, 172
 Reynolds, Garr, 227
 Rheingold, Howard, 187
 Rosen, Jay, 125
 rows, 63–64

S

Saint-Exupéry, Antoine de, 89
 Salesforce.com, 50
 scatterplot charts, 65, 70, 222
 Schein, Edgar, 127
 Schmitt, Garrick, 82
 searching
 metadata, 176
 for understanding, 40–41
Security Metrics (Jaquith), 135
 selection
 data, 76
 metrics, 138
 subject, 146
 selection-maturation interaction, 146
 self-awareness, 40
 self-serving metrics, 141
Shaping School Culture: The Heart of Leadership, 128
 share of wallet, 92
 shared understandings, 141–151
 SharePoint, 172
 sharing, 174
 Siegel, Alan, 84
 siloed data, 37
Simple (Siegel and Etzkorn), 84
 simplicity, 227
 simplistic metrics, 93

SINTEF. *See* The Foundation for Scientific and Industrial Research

Six Ds, 164, 165, 181, 194, 207

skills, 31

Smith, Michael David, 147

smoothing, 119

snapshots, 177–178

Sontag, Susan, 91

sorting, 120

spreadsheet-driven cultures, 3

SQL Server Reporting Services, 172

standards, 19, 28

Star Wars, 20

statistical significance, 56, 232

Steinbeck, John, 90, 91

Steinberg, Leigh, 93

storytelling

- data, 82, 95
- guided, 163
- Pixar rules for, 102

structure, 95–102

- business, 235
- dashboards, 235
- guided path, 95–96
- importance, 97
- options, 97–102
 - attention, 100
 - flow, 97–98
 - grids, 100–101
 - grouping, 100
 - organizing page, 100
 - relationships, 98
 - white space, 101–102
- style guide, 237–244
 - charts, 237–238, 242–243
 - color, 237, 240
 - data product ecosystems, 172
 - fonts, 239
 - labeling, 238
 - typography, 237, 239
- subject selection, 146
- subscribing, 176
- summarized data, 61–62

summary status, 234

supermodel dashboards, 38–39

surfacing, 176

T

Tableau, 173

tables, 238, 244

TechMeme, 101

technological constraints, 3, 4

technology roadmap, 31

terminology, common, 143–144

Texas Oil Boom, 1

"Think Like a Designer" (Hilburn), 226–227

Thompson, Clive, 60

3-D effects, 119

time window, 236

tools, 28

top-down demand map, 167–169

Total Quarterback Rating (QBR), 147

training, 31

transparency, 147–148

trend charts, 243

trending metrics, 140

trough of disillusionment, 7

Tufte, Edward, 65, 107, 115, 174, 188

Twitter, 56, 176

typography, 109–110

- style guide, 237, 239

U

under-indexing, 56

unexpected distributions, 69

unexpected patterns or relationships, 70

unexpected trends, 70–71

United States Federal Government data initiatives, 6

unity, 226

The Unpublished David Ogilvy, 92

U.S. News & World Report, 11–13

user permissions, 176

V

validity, 219
 data, 74–77
 external, 145
face, 76
internal, 145
measurement, 75
vanity metrics, 93
variation, 9
visual real estate, 229
visualizations, 64–65, 69, 104–106
 design principles, 115–121
 chart junk, 115, 116
 color variants, 120–121
 contrast, 117
 data-to-ink ratio, 115
 gradients, 119
 labels, 117
 repetition, 118
 smoothing, 119
 sorting, 120
 3-D effects, 119
labeling, 111
quality, 174

vocabulary, 143–144
Vonnegut, Kurt, 90, 91

W

Warhol, Andy, 47, 49
Watson, 3
weather forecast, 51
wellness data, 15–17
white space, 101–102, 227
Why We Should Learn the Language of Data
 (Thompson), 60
widgets, 238
Wikipedia, 179–181
wine labels, 53
Wired magazine, 60
workflows, 174

X

Xcode, 166–167

Y

Yahoo!, 10
Yau, Nathan, 54

