SECOND EDITION

# VISUAL SIX SIGMA

## MAKING DATA ANALYSIS LEAN

IAN COX • MARIE A. GAUDARD • MIA L. STEPHENS

WILEY

# Visual Six Sigma

# Wiley & SAS Business Series

The Wiley & SAS Business Series presents books that help senior-level managers with their critical management decisions.

Titles in the Wiley & SAS Business Series include:

*Agile by Design: An Implementation Guide to Analytic Lifecycle Management* by Rachel Alt-Simmons

*Analytics in a Big Data World: The Essential Guide to Data Science and Its Applications* by Bart Baesens

*Bank Fraud: Using Technology to Combat Losses* by Revathi Subramanian

*Big Data, Big Innovation: Enabling Competitive Differentiation through Business Analytics* by Evan Stubbs

*Business Forecasting: Practical Problems and Solutions* edited by Michael Gilliland, Len Tashman, and Udo Sglavo

*Business Intelligence Applied: Implementing an Effective Information and Communications Technology Infrastructure* by Michael Gendron

*Business Intelligence and the Cloud: Strategic Implementation Guide* by Michael S. Gendron

*Business Transformation: A Roadmap for Maximizing Organizational Insights* by Aiman Zeid

*Data-Driven Healthcare: How Analytics and BI Are Transforming the Industry* by Laura Madsen

*Delivering Business Analytics: Practical Guidelines for Best Practice* by Evan Stubbs

*Demand-Driven Forecasting: A Structured Approach to Forecasting, Second Edition* by Charles Chase

*Demand-Driven Inventory Optimization and Replenishment: Creating a More Efficient Supply Chain* by Robert A. Davis

*Developing Human Capital: Using Analytics to Plan and Optimize Your Learning and Development Investments* by Gene Pease, Barbara Beresford, and Lew Walker

*Economic and Business Forecasting: Analyzing and Interpreting Econometric Results* by John Silvia, Azhar Iqbal, Kaylyn Swankoski, Sarah Watt, and Sam Bullard

*Financial Institution Advantage and the Optimization of Information Processing* by Sean C. Keenan

*Financial Risk Management: Applications in Market, Credit, Asset, and Liability Management and Firmwide Risk* by Jimmy Skoglund and Wei Chen

*Fraud Analytics Using Descriptive, Predictive, and Social Network Techniques: A Guide to Data Science for Fraud Detection* by Bart Baesens, Veronique Van Vlasselaer, and Wouter Verbeke

*Harness Oil and Gas Big Data with Analytics: Optimize Exploration and Production with Data Driven Models* by Keith Holdaway

*Health Analytics: Gaining the Insights to Transform Health Care* by Jason Burke

*Heuristics in Analytics: A Practical Perspective of What Influences Our Analytical World* by Carlos Andre, Reis Pinheiro, and Fiona McNeill

*Hotel Pricing in a Social World: Driving Value in the Digital Economy* by Kelly McGuire

*Implement, Improve and Expand Your Statewide Longitudinal Data System: Creating a Culture of Data in Education* by Jamie McQuiggan and Armistead Sapp

*Killer Analytics: Top 20 Metrics Missing from your Balance Sheet* by Mark Brown

*Mobile Learning: A Handbook for Developers, Educators, and Learners* by Scott McQuiggan, Lucy Kosturko, Jamie McQuiggan, and Jennifer Sabourin

*The Patient Revolution: How Big Data and Analytics Are Transforming the Healthcare Experience* by Krisa Tailor

*Predictive Analytics for Human Resources* by Jac Fitz-enz and John Mattox II

*Predictive Business Analytics: Forward-Looking Capabilities to Improve Business Performance* by Lawrence Maisel and Gary Cokins

*Statistical Thinking: Improving Business Performance, Second Edition* by Roger W. Hoerl and Ronald D. Snee

*Too Big to Ignore: The Business Case for Big Data* by Phil Simon

*Trade-Based Money Laundering: The Next Frontier in International Money Laundering Enforcement* by John Cassara

*The Visual Organization: Data Visualization, Big Data, and the Quest for Better Decisions* by Phil Simon

*Understanding the Predictive Analytics Lifecycle* by Al Cordoba

*Unleashing Your Inner Leader: An Executive Coach Tells All* by Vickie Bevenour

*Using Big Data Analytics: Turning Big Data into Big Money* by Jared Dean

*Visual Six Sigma, Second Edition* by Ian Cox, Marie Gaudard and Mia Stephens

For more information on any of the above titles, please visit www.wiley.com.

# Visual Six Sigma

*Making Data Analysis Lean*

**Ian Cox**
**Marie A. Gaudard**
**Mia L. Stephens**

Second Edition

WILEY

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.

Published simultaneously in Canada.



***Library of Congress Cataloging-in-Publication Data:***

Names: Cox, Ian, 1956–
Title: Visual six sigma : making data analysis lean / Ian Cox, Marie A. Gaudard, Mia L. Stephens.
Description: Second edition. | Hoboken : Wiley, 2016. | Series: Wiley and SAS business series | Revised edition of Visual six sigma, 2010. | Includes index.
Identifiers: LCCN 2016001878 (print) | LCCN 2016003459 (ebook) | ISBN 9781118905685 (hardback) | ISBN 9781119222262 (epdf) | ISBN 9781119222255 (epub)
Subjects: LCSH: Six sigma (Quality control standard) | Decision support systems. | Decision making—Statistical methods. | Organizational effectiveness. | BISAC: BUSINESS & ECONOMICS / Strategic Planning.
Classification: LCC HD30.213 .C69 2016 (print) | LCC HD30.213 (ebook) | DDC 658.4/013—dc23
LC record available at http://lccn.loc.gov/2016001878

Cover Design: Wiley
Cover Image: ©Studio-Pro/iStock.com

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

# Contents

# Preface to the Second Edition

The first edition of this book appeared in 2010, so we decided to produce an updated and expanded second edition. The purpose of the book remains unchanged—to show how, using the three principles of Visual Six Sigma, you can exploit data to make better decisions more quickly and easily than you would otherwise. And, as you might expect given their power and utility, these principles are also unchanged. However, production of this second edition allows us to take advantage of some interim developments that make the implementation of Visual Six Sigma even easier, further increasing the scope and efficacy of its application. It also allows us to improve and enhance the content and form of the first edition.

The staying power of Six Sigma as a methodology can be attributed to the fact that it can provide a common language for, and approach to, project-based improvement initiatives. Nonetheless, as we pointed out in the first edition, there is a clear need to evolve the mechanics of Six Sigma both to accommodate the greater availability of data and to address the fact that, historically, approaches to analyzing data were overly concerned with hypothesis testing, to the detriment of the hypothesis generation and discovery needed for improvement. We believe that Visual Six Sigma can foster this evolution, and this is part of our motivation for keeping this text current.

At the same time, the past five years have seen the explosion of "big data," at least as an identifiable area that software providers and implementation consultants make strenuous efforts to market to. In this language, the increased data availability mentioned above is measured using three dimensions: volume, variety, and velocity. Even though the precise definition of big data is not always clear, we think there is much for would-be data scientists to learn from the principles of Visual Six Sigma and their application. In addition, if a project-based approach is warranted, the language of Six Sigma may also be useful.

Although the principles of Visual Six Sigma are general, their effective and efficient adoption in practice is reliant on good enabling software. The first edition was tied to version 8.01 of JMP, Statistical Discovery software from SAS Institute®. This second edition has been revised to be consistent with the version current at the time of writing, JMP 12.2.0. Generally, JMP aims to exploit the synergy between visualization and analysis, and its continuing development has opened up new possibilities for Visual Six Sigma. In some cases, these are simply matters of detail and efficiency, but in others there are important new capabilities we can use.

A key feature of the book remains the six self-contained case studies. Given feedback from the first edition, we are even more convinced of the advantage of this format in showing how seemingly disparate techniques can be used in concert to accomplish something useful. We interweave the new capabilities of JMP where they usefully support or extend the case studies.

Consistent with the requirements of Visual Six Sigma in the new era of big data, we have introduced two new chapters:

- Chapter 4, "Managing Data and Data Quality," precedes the case studies and addresses the management of data and data quality. Data quality, at an organizational level, is a ubiquitous topic that is often seen as mainstream to the point of being boring. However, the importance of data quality for project teams and anyone making decisions with data cannot be overstated. As we shall see, the Visual Six Sigma context leads to some important and interesting nuances.
- Chapter 11, "Beyond 'Point and Click' with JMP," follows the case studies and shows how to go beyond the interactive usage of JMP for discovery and improvement. No matter how simple or complex, the performance of empirical models always degrades over time. Once improvements are made, there is always the need to monitor and adapt with an appropriate frequency. In turn, this means that analyses need to be repeated as new data arrive, and this is often best done with an element of automation.

The case studies appear in Part Two of the book. Chapter 4 is appended to Part One, making this section four chapters long. Given the nature of the content, Chapter 11 appears as a singleton chapter in Part Three.

Finally, we have tried to make the case studies easier to use by having clearer typographic separation between the narrative (consisting of the *why*, the *what*, and the findings of each technique as it is used in a specific context) and the "how to" steps required in JMP. As well as helping to keep things concise, this arrangement better accommodates users with different levels of prior familiarity with JMP, and may make it easier to use other software should this be required or mandated.

As in the first edition, we have used different fonts to help identify the names of data tables, of columns in data tables, and commands. Data table names are shown in **MeridienLTStd-Bold**, the names of columns (which are variable names) are shown in *italic Helvetica*, and the names of commands and other elements of the user interface are shown in **bold Helvetica**.

We are now living through a time of rapid change in the world of data analysis. We have tried to reflect this in our changes and additions. We hope that this second edition on Visual Six Sigma contains even more of interest for current or would-be Six Sigma practitioners, or more generally for anyone with

a stake in exploiting data for the purpose of gaining new understanding or of driving improvement.

## Supplemental Materials

We anticipate that you will follow along, using JMP, as you work through the case studies and Chapters 4 and 11. You can download a trial copy of JMP at www.jmp.com/try. Chapter 10 requires JMP Pro. You can request a trial version of JMP Pro at www.jmp.com/en_us/software/jmp-pro-eval.html. JMP instructions in this book are based on JMP 12.2.0. Although the menu structure may differ if you use a different version of JMP, all the functionality described in this book is available in JMP 12.2.0 or newer versions.

The data sets used in the book are available at http://support.sas.com/visualsixsigma. This folder contains a journal file, **Visual Six Sigma.jrn**, that contains links to the data tables, scripts, and add-ins discussed in this book. The color versions of the exhibits shown in the book are also available here. Exhibits showing JMP results were taken using JMP 12.2.0 running on Windows.

# Preface to the First Edition

The purpose of this book is to show how, using the principles of Visual Six Sigma, you can exploit data to make better decisions more quickly and easily than you would otherwise. We emphasize that your company does not need to have a Six Sigma initiative for this book to be useful. Clearly there are many data-driven decisions that, by necessity or by design, fall outside the scope of a Six Sigma effort, and in such cases we believe that Visual Six Sigma is ideal. We seek to show that Visual Six Sigma can be used by a lone associate, as well as a team, to address data-driven questions, with or without the support of a formal initiative like Six Sigma.

To this end, we present six case studies that show Visual Six Sigma in action. These case studies address complex problems and opportunities faced by individuals and teams in a variety of application areas. Each case study was addressed using the Visual Six Sigma Roadmap, described in Chapters 2 and 3. As these case studies illustrate, Visual Six Sigma is about exploration and discovery, which means that it is not, and never could be, an entirely prescriptive framework.

As well as using the case studies to convey the Visual Six Sigma Roadmap, we also want to use them to illustrate Visual Six Sigma techniques that you can reuse in your own setting. To meet this goal, sometimes we have deliberately compromised the lean nature of the Visual Six Sigma Roadmap in order to take the opportunity to show you extra techniques that may not be strictly necessary to reach the conclusion or business decision. Striking the balance this way means that you will see a wider repertoire of techniques from which to synthesize an approach to Visual Six Sigma that works for you.

Because of its visual emphasis, Visual Six Sigma opens the doors for non-statisticians to take active roles in data-driven decision making, empowering them to leverage their contextual knowledge to pose relevant questions, get good answers, and make sound decisions. You may find yourself working on a Six Sigma improvement project, a design project, a data mining inquiry, or a scientific study—all of which require decision making based on data. After working through this book, we hope that you will be able to make data-driven decisions in your specific situation quickly, easily, and with greater assurance.

## How This Book Is Organized

This book is organized in two parts. Part I contains an introductory chapter that presents the three Visual Six Sigma strategies, a chapter on Visual Six Sigma,

and a chapter introducing JMP statistical software (from SAS® Institute), which will be used throughout the case studies.

Case studies are presented in Part Two. These case studies follow challenging real-world projects from start to finish. Through these case studies, you will gain insight into how the three Visual Six Sigma strategies combine to expedite project execution in the real world. Each case study is given its own chapter, which can be read independently from the rest. A concise summary of the storyline opens each case study. Although these case studies are real, we use fictitious names for the companies and individuals to preserve confidentiality.

Within each case study, visualization methods and other statistical techniques are applied at various stages in the data analysis process in order to better understand what the data are telling us. For those not familiar with JMP, each case study also contains the relevant how-to steps so that you may follow along and see Visual Six Sigma in action.

The data sets used in the case studies are available at http://support.sas.com/visualsixsigma. Here you can also find the exhibits shown in the case studies, allowing you to see screen captures in color. Additional Visual Six Sigma resource materials will be made available on the website, as appropriate.

## A Word about Software

The ideas behind Visual Six Sigma are quite general, but active learning—in our view, the only kind of learning that works—requires that you step through the case studies and examples in this book to try things out for yourself. For more information about JMP, and to download a trial version of the software, visit www.jmp.com/demo.

JMP is available on Windows, Mac, and Linux platforms. The step-by-step instructions in this book assume that you are working in Windows. Mac and Linux users should refer to the JMP documentation for details on differences. This book is based on JMP version 8.0.1.

# Acknowledgments

Stating the obvious, this book would not exist without its first edition. Even though some have moved on, we remain deeply indebted to all those listed who made the first edition possible. Most importantly, we want to thank Leo Wright of SAS and Phil Ramsey of the North Haven Group, LLC, our co-authors on the first edition, who provided some of the original case studies and helped to make this book possible.

Both editions of the book were substantially improved by suggestions from Mark Bailey of SAS. We greatly appreciate his time, interest, valuable feedback, and insights. We want to thank Andy Liddle, now of Process Insight Consulting Limited, who assisted with the review of the original version of "Improving a Polymer Manufacturing Process" (now Chapter 9). We also want to thank Volker Kraft of SAS, who provided valuable feedback in connection with updates to this case study for the book's second edition.

This project was greatly facilitated by Stacey Hamilton and Stephenie Joyner of SAS Publishing. Their support, encouragement, and attention to detail at every step of this adventure were invaluable.

Finally, we would like to thank Jon Weisz and Curt Hinrichs of JMP for their support and encouragement in updating this book. And, as before, a special thank-you goes to John Sall, Bradley Jones, Chris Gotwalt, Xan Gregg, Brian Corcoran, and the JMP Development Team for their continuing work on a visionary product that makes Visual Six Sigma possible.

# About the Authors

**Ian Cox** currently works in the JMP Division of SAS. Before joining SAS in 1999, he worked for Digital Equipment Corporation, Motorola, and BBN Software Solutions Ltd. and has been a consultant for many companies on data analysis, process control, and experimental design. A Six Sigma Black Belt, he was a Visiting Fellow at Cranfield University and is a Fellow of the Royal Statistical Society. Cox holds a Ph.D. in theoretical physics.

**Marie A. Gaudard** is a consultant specializing in statistical training with the use of JMP. She is currently a statistical writer with the JMP documentation team. She earned her Ph.D. in statistics in 1977 and was a professor of statistics at the University of New Hampshire from 1977 until 2004. She has been heavily involved in statistical consulting since 1981. Gaudard has worked with a variety of clients in government agencies, medical areas, and manufacturing. She has extensive experience in consulting and training in the areas of Six Sigma, Design for Six Sigma, forecasting and demand planning, and data mining.

**Mia L. Stephens** is an academic ambassador with the JMP division of SAS. Prior to joining SAS, she was an adjunct professor at the University of New Hampshire and a partner in the North Haven Group, a statistical training and consulting company. Also a coauthor of *JMP Start Statistics: A Guide to Statistics and Data Analysis Using JMP, Fifth Edition* and *Building Better Models with JMP Pro*, she has developed courses and training materials, taught, and consulted within a variety of manufacturing and service industries. Stephens holds an M.S. in statistics from the University of New Hampshire.

# Visual Six Sigma

PART
**ONE**

# Background

CHAPTER 1

# Introduction

## WHAT IS VISUAL SIX SIGMA?

Visual Six Sigma is about leveraging interactive and dynamic graphical displays to help transform data into sound decisions. It is not an algorithm. It is a creative process that employs visual techniques in the discovery of new and useful knowledge, leading to quicker and better decisions than do the methods in general use today. It signals a new generation of Six Sigma techniques.

At the heart of Six Sigma is the concept of data-driven decision making, that is, of exploiting the data from measurements or simulations at various points in the life cycle of your product or service. Visual Six Sigma aims to produce better alignment between Six Sigma practice and the key idea of discovery, providing benefits for all those who have a stake in solving problems and in making improvements through data.

Visual Six Sigma consists of three main strategies:

1. Using dynamic visualization to literally *see* the sources of variation in your data.
2. Using exploratory data analysis techniques to *identify key drivers and models*, especially for situations involving many variables.
3. Using confirmatory statistical methods only when the conclusions are not obvious.

Six Sigma programs often use the so-called *DMAIC approach* for team-based process improvement or problem-solving efforts. The acronym *DMAIC* stands for the major phases in a team's project: *Define*, *Measure*, *Analyze*, *Improve*, and *Control*. DMAIC provides a structure for a team's efforts, just as an overall Six Sigma program provides a structure for a company's efforts. Each phase of DMAIC comes with a list of techniques that are considered appropriate in that phase; the team moves from one phase to another, using this sequence of techniques as a general guide. In a similar way, Six Sigma projects aimed at design follow various structures, such as *Define*, *Measure*, *Analyze*, *Design*, and *Validate* (DMADV) and *Identify*, *Design*, *Optimize*, and *Validate* (IDOV).

Visual Six Sigma is not a replacement for the DMAIC, DMADV, or IDOV frameworks. Rather, Visual Six Sigma supports these frameworks by simplifying and enhancing methods for data exploration and discovery whenever they are needed. In addition, when circumstances make a full-blown project-based or team-based approach undesirable or unworkable, Visual Six Sigma can still be used by individual contributors such as you. In a nutshell, Visual Six Sigma helps to make the DMAIC and design structures—and data analysis in general—lean.

### Moving beyond Traditional Six Sigma

It is our belief that the tools, techniques, and workflows in common use with Six Sigma efforts are typically not aligned with the key idea of discovery. In

the early days of Six Sigma, relevant data rarely existed, and a team was often challenged to collect data on its own. As part of the Measure phase, a team usually conducted a brainstorming session to identify which features of a process should be measured. In some sense, this brainstorming session was the team's only involvement in hypothesis generation. The data collected were precious, and hypothesis testing methods were critical in separating signals from noise.

Project teams struggling with a lack of useful data generally rely on an abundance of subjective input, and often require hypothesis testing to minimize the risk of bad decisions. This emphasis on hypothesis testing is reasonable in an environment where data are sparse. In contrast, today's Six Sigma teams often find warehouses of data that are relevant to their efforts. Their challenge is to wade through the data to discover prominent features, to separate the remarkable from the unremarkable.

These data-rich environments call for a shift in emphasis from confirmatory methods, such as hypothesis testing, to exploratory methods, with a major emphasis on the display of data to reveal prominent features that are hidden in the data. Since the human interpretation of the data context is a vital part of the discovery process, these exploratory techniques cannot be fully automated. Also, with large quantities of data, hypothesis testing itself becomes less useful—statistical significance comes easily and may have little to do with practical importance.

Of course, the simple abundance of data in a warehouse does not guarantee its relevance for improvement or problem solving. In fact, it is our experience that teams working in what they believe to be data-rich environments sometimes find that the available data are of poor quality or are largely irrelevant to their efforts. Visualization methods can be instrumental in helping teams quickly reach this conclusion. In these cases, teams need to revert to techniques such as brainstorming, cause-and-effect diagrams, and process maps, which drive efforts to collect the proper data. But, as we shall see, even in situations where only few relevant data are available, visualization techniques, supported as appropriate by confirmatory methods, prove invaluable in identifying telling features of the data.

## Making Data Analysis Lean

Discovery is largely supported by the generation of hypotheses—conjectures about relationships and causality. Today's Six Sigma teams, and data analysts in the business world in general, are often trained with a heavy emphasis on hypothesis testing, with comparatively little emphasis given to hypothesis generation and discovery. They are often hampered in their problem-solving and improvement efforts by the inability to exploit exploratory methods, which could enable them to make more rapid progress, often with less effort.

In recent times, we have seen incredible advances in visualization methods, supported by phenomenal increases in computing power. We strongly believe that the approaches now allowed by these methods are underutilized in current Six Sigma practice. It is this conviction that motivated us to write the first edition of this book and, following its success, to produce a second edition that takes advantage of recent software advances. We hope you find this book useful as you shape and build your own real-world Six Sigma experience.

## Requirements of the Reader

This leads to another important point, namely, that you are "part of the system." Discovery, whether practiced as an individual or as a team sport, involves both divergent and convergent thinking; both creativity and discipline are required at different times. You should bear this in mind when forming a team or when consulting with individuals, since each person will bring his or her own skill set, perspective, and strength to the discovery process.

Given the need to be data driven, we also need to recognize one of the basic rules of using data, which is that any kind of analysis that treats data simply as a list of numbers is doomed to failure. To say it differently: All data are contextual, and it is this context and the objectives set out for the project that must shape the analysis and produce useful recommendations for action. As a practitioner, your main responsibility should always be to understand what the numbers in the data actually mean in the real world. In fact, this is the only requirement for putting the ideas in this book into practice in your workplace.

CHAPTER 2

# Six Sigma and Visual Six Sigma

This chapter introduces the key ideas behind Six Sigma and Visual Six Sigma; our focus is on the latter. Six Sigma is a potentially huge topic, so we only have space to mention some of its essential ideas. There are already numerous well-written books and articles dealing with the many and diverse aspects of Six Sigma as commonly practiced.[1] We also note that today, digital tools (software, databases, visual media, etc.) are leveraged extensively in Six Sigma initiatives.[2]

Our goal in this chapter is to provide an overview of Six Sigma so that you start to see how Visual Six Sigma fits into this picture. However, it is worth pointing out in advance that you can only gain a proper appreciation of the power of visualization techniques by working with data that relate to real problems in the real world.

## BACKGROUND: MODELS, DATA, AND VARIATION

There is no doubt that science and technology have transformed the lives of many and will continue to do so. Like many fields of human endeavor, science proceeds by building pictures, or *models*, of what we think is happening. These models can provide a framework in which we attempt to influence or control inputs so as to provide better outputs. Unlike the models used in some other areas, the models used in science are usually constructed using data that arise from measurements made in the real world.

At the heart of the scientific approach is the explicit recognition that we may be *wrong* in our current world view. Saying this differently, we recognize that our models will always be imperfect, but by confronting them with data, we can strive to make them better and more useful. Echoing the words of George Box, one of the pioneers of industrial statistics, we can say, "Essentially, all models are wrong, but some are useful."[3]

## MODELS

The models of interest in this book can be conceptualized as shown in Exhibit 2.1. This picture demands a few words of explanation:

- In this book, and generally in Six Sigma, the outcomes of interest to us are denoted with a Y. For example, Y1 in Exhibit 2.1 could represent the event that someone will apply for a new credit card after receiving an offer from a credit card company.
- Causes that may influence a Y will be shown as an X. To continue the example, X1 may denote the age of the person receiving the credit card offer.
- Rather than using a lengthy expression such as "we expect the age of the recipient to influence the chance that he or she will apply for a credit