Dashboards for Excel

Deliver Critical Information and Insight at the Speed of a Click

Jordan Goldmeier
Purnachandra Duggirala

Apress®

Dashboards for Excel

Jordan Goldmeier
Purnachandra Duggirala

Apress®

Dashboards for Excel

ISBN-13 (pbk): 978-1-4302-4944-3

ISBN-13 (electronic): 978-1-4302-4945-0

Managing Director: Welmoed Spahr
Lead Editor: James DeWolf
Development Editor: Chris Nelson
Technical Reviewer: Fabio Claudio Ferracchiati
Editorial Board: Steve Anglin, Mark Beckner, Gary Cornell, Louise Corrigan, Jim DeWolf, Jonathan Gennick, Jonathan Hassell, Robert Hutchinson, Michelle Lowman, James Markham, Susan McDermott, Matthew Moodie, Jeffrey Pepper, Douglas Pundick, Ben Renow-Clarke, Gwenan Spearing, Matt Wade, Steve Weiss
Coordinating Editor: Melissa Maldonado
Copy Editor: Kim Wimpsett
Compositor: SPi Global
Indexer: SPi Global
Artist: SPi Global

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

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

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

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

For Katherine.

—Jordan

To Jyosthna, of course.

—Purnachandra

Contents at a Glance

About the Authors xxi

About the Technical Reviewer xxiii

Acknowledgments xxv

Introduction xxvii

■ Part I: Dashboards and Data Visualization 1

■ Chapter 1: Introduction to Dashboard and Decision Support Development 3

■ Chapter 2: A Critical View of Information Visualization 19

■ Chapter 3: The Principles of Data Visualization in Microsoft Excel 35

■ Chapter 4: The Excel Data Presentation Library 55

■ Part II: Excel Dashboard Design Tools and Concepts 79

■ Chapter 5: Getting Started: Thinking Outside the Cell 81

■ Chapter 6: Visual Basic for Applications for Excel, a Refresher 111

■ Chapter 7: Avoiding Common Pitfalls in Development and Design 131

■ Chapter 8: The Elements of Good Excel Dashboards and Decision Support Systems 155

■ Part III: Formulas, Controls, and Charts 177

■ Chapter 9: Introducing Formula Concepts 179

■ Chapter 10: Advanced Formula Concepts 199

■ Chapter 11: Metrics: Performance and Context 217

■ Chapter 12: Charts with Heart (or, How to Avoid a Chart Attack) 227

■ Chapter 13: Creating an Interactive Gantt Chart Dashboard 253

■ Chapter 14: An Interactive Gantt Chart Dashboard, Data Visualization............. 273

■ Chapter 15: An Interactive Gantt Chart Dashboard, Data Details on Demand......... 295

■ Part IV: From User Interface to Presentation................................. 317

■ Chapter 16: Working with Form Controls 319

■ Chapter 17: Getting Input from Users 349

■ Chapter 18: Storage Patterns for User Input 369

■ Chapter 19: Building for Sensitivity Analysis 393

■ Chapter 20: Perfecting the Presentation 423

■ Part V: Data Models, PowerPivot, and Power Query 449

■ Chapter 21: Data Model Capabilities of Excel 2013... 451

■ Chapter 22: Advanced Modeling with Slicers, Filters, and Pivot Tables............ 465

■ Chapter 23: Introduction to Power Query 483

■ Chapter 24: Introduction to PowerPivot 515

Index 535

Contents

About the Authors xxi

About the Technical Reviewer xxiii

Acknowledgments xxv

Introduction xxvii

Part I: Dashboards and Data Visualization 1

Chapter 1: Introduction to Dashboard and Decision Support Development 3

The Data Problem 3

Enter Excel: The Most Dangerous Program in the World 4

Not Realizing How Far Spreadsheets Have Come 4

Garbage In, Gospel Out 5

How Excel Fits In 5

What Excel Is Good For 6

A Commercial-Off-the-Shelf Solution 6

Flexible and Customizable 6

Familiarity and Ubiquity 6

Inexpensive-ish 7

Quick Turnaround Time 7

A Good Return on Investment 7

What Excel Isn't Good For 7

A Full-Fledged Database 8

Enterprise-Level Reporting 8

A Full Software Package 8

Predicting the Future 8

Buzzword Bingo: Dashboards, Reports, Data Visualization, and Others 9
Dashboards...9
Decision Support Systems...10
The Excel Development Trifecta 10
Good Visualization Practices...11
Good Development Practices..12
Thinking Outside the Cell...15
Available Resources 16
The Last Word.. 17
■ **Chapter 2: A Critical View of Information Visualization 19**
Understanding the Problem.. 19
Of Pilots and Metaphors 20
A Metaphor Too Far: Driving Down the Information Superhighway............................... 22
A Brief History of Dashboards and Information Visualization.. 23
A Quick Summary Before Taking a Critical Look 24
Dashboards by Example: U.S. Patent and Trademark Office 24
Radial Gauges..26
So Many Metrics, So Little Working Memory..27
Is the Logo Necessary? 28
Visualizations That Look Cool but Just Don't Work 29
Data Journalism 31
Why These Examples Are Important 33
The Last Word.. 33
■ **Chapter 3: The Principles of Data Visualization in Microsoft Excel 35**
What Is Visual Perception and How Does It Work?.. 35
Perception and the Visual World 36
Our Bias Toward Forms: Perception and Gestalt Psychology...............................36

The Preattentive Attributes of Perception 46
Color Attributes 47
High-Precision Judging 49
Lower Precision, but Still Useful 51
The Last Word 53
Chapter 4: The Excel Data Presentation Library 55
Tables 55
Line and Bar Charts 56
Scatter Charts 60
Scatter Charts vs. Line Charts 61
Correlation Analysis 64
Correlation Fit and Coefficient 65
Linear Relationship and Using R^2 Correctly 66
Bullet Graphs 67
Small Multiples 68
Charts Never to Use 69
Cylinders, Cones, and Pyramid Charts 69
Pie Charts 71
Doughnut Charts 72
Charts in the Third Dimension 72
Surface Charts 74
Stacked Columns and Area Charts 75
Radar Charts 77
The Last Word 78

■ **Part II: Excel Dashboard Design Tools and Concepts........................ 79**

■ **Chapter 5: Getting Started: Thinking Outside the Cell 81**

House Hunters: Excel Edition.. 82

The Purely VBA Method 83

The Semi-code Method 86

The No-Code Method... 95

Sorting.. 101

The Rollover Method 105

Rollover Method Basics 107

Implementing the Rollover Method... 107

The Last Word.. 109

■ **Chapter 6: Visual Basic for Applications for Excel, a Refresher 111**

Making the Most of Your Coding Experience 111

Tell Excel: Stop Annoying Me!... 112

Make Loud Comments.. 113

Pick a Readable Font... 115

Start Using the Immediate Window, Immediately.................................. 115

Opt for Option Explicit... 116

Naming Conventions 117

Hungarian Notation... 117

"Loose" CamelCase Notation 118

Named Ranges 119

Sheet Objects 119

Referencing.. 121

Shorthand References... 121

Worksheet Object Names 122

Procedures and Macros.. 122

Development Styles and Principles.. 123

Strive to Store Your Commonly Used Procedures in Relevant Worksheet Tabs 123

No More Using the ActiveSheet, ActiveCell, ActiveWorkbook, and Selection Objects........................127
Render Unto Excel the Things That Are Excel's and Unto VBA the Things That Require VBA..............128
Encapsulating Your Work...129
The Last Word... 129
Chapter 7: Avoiding Common Pitfalls in Development and Design 131
Calculation Pitfalls.. 132
Volatile Functions and Actions..132
Understanding Different Formula Speeds 141
Spreadsheet Errors..144
Code Pitfalls 147
Copy/Paste Iterations...148
Testing Properties Before Setting Them 148
Bad Names 151
The Last Word... 153
Chapter 8: The Elements of Good Excel Dashboards and Decision Support Systems 155
Types of Dashboards 155
Strategic 155
Operational 156
Analytical...156
Decision Support Systems 156
Simplified Layout... 157
Information-Transformation-Presentation.. 159
Common Dashboard Problems.. 162
Too Much Formatting and Embellishment..162
Too Many Tabs 165
Bad Layout...167
Needless Protection..171
Instructions and Documentation..174
The Last Word... 175

Part III: Formulas, Controls, and Charts 177

Chapter 9: Introducing Formula Concepts 179

Formula Help 179

F2 to See the Formula of a Select Cell 179

F9 for On-Demand and Piecewise Calculation 179

Evaluate Formula Button 180

Excel Formula Concepts 181

Operators, in Depth 181

The Range Operator (:) 182

The Union Operator (,) 184

The Intersection Operator ( ) 185

When to Use Conditional Expressions 187

Deceptively Simple Nested IF Statements 187

CHOOSE Wisely 189

Why This Discussion Is Important 190

Introduction to Boolean Concepts 190

Condensing Your Work 193

The Legend of XOR()-oh 193

Do We Really Need IF? 195

The Last Word 197

Chapter 10: Advanced Formula Concepts 199

Filtering and Highlighting 199

Filtering with Formulas 199

Conditional Highlighting Using Formulas 203

Selecting 205

Aggregating 211

Using SUMPRODUCT for Aggregation 211

You're About To Be FOILed! 214

Reusable Components 215

The Last Word 216

Chapter 11: Metrics: Performance and Context 217
Telling the Whole Story Like a Reporter: An Introduction to Analytics 217
Who and Where 218
When 220
Why, How, and What 220
What If? 222
Metrics, Metrics, Metrics 223
Working Criteria for Choosing Metrics 224
The Last Word 226
Chapter 12: Charts with Heart (or, How to Avoid a Chart Attack) 227
Challenging the Defaults 227
Data-Ink Density 227
Charts: More Art Than Science, in Some Ways 228
Chart Formats for Simple Comparison 230
Column and Bar Charts 230
Layout Constraints 232
Banding 234
Bullet Charts 241
The Last Word 251
Chapter 13: Creating an Interactive Gantt Chart Dashboard 253
Features of the Gantt Chart Dashboard 254
Pop-up Information Box 254
Banded Region Chart 255
Dynamic Chart Legend 255
Behind the Scenes 256
The Data Worksheet Tab 257
The Calculation Worksheet Tab 258
The Dashboard Worksheet Tab 263
The Last Word 271

Chapter 14: An Interactive Gantt Chart Dashboard, Data Visualization............ 273

The Banded Region Chart.. 274

Creating the Chart 275

Placing the Chart onto the Dashboard..282

Creating the Banded Chart Legend...284

The Dynamic Legend.. 287

Creating the Endpoints of the Legend 288

Interpolating Between the Endpoints 290

Applying the Visual Effects 291

The Last Word... 293

Chapter 15: An Interactive Gantt Chart Dashboard, Data Details on Demand.........295

The Rollover Method 296

A Quick Review of the Rollover Method...296

A Conceptual Model for Rollover Method Implementation 296

Implementing a Hover Table 298

Implementing Conditional Formatting 300

Gantt Chart Rollovers 304

Rollover Implementation..304

Getting the Information to Create the Details-on-Demand Pop-up 307

Designing the Pop-up 311

Making the Pop-up Follow Your Mouse 311

The Last Word... 316

Part IV: From User Interface to Presentation................................ 317

Chapter 16: Working with Form Controls 319

Welcome to the Control Room... 319

Form Control Fundamentals... 321

The ComboBox Control 323

The ListBox Control..325

The Scroll Bar Control..327

The Spinner Control..329

The CheckBox Control 329

The Least Favorites: Button, Label, Option Button, and GroupBox Controls 330

Creating Scrollable Tables 333
Highlighting Data Points on Charts 341
The Dynamic Legend 345
The Last Word 348
Chapter 17: Getting Input from Users 349
Of Input Forms and Excel 349
A Simple Input Form 351
Custom Formats for Input Validation 352
Creating a Spreadsheet-Based Wizard 355
Layout Patterns for the Spreadsheet-Based Wizard 356
The Helper Tab 357
Moving Between Views 358
Views That Require Additional Instruction 360
Components That Provide Information 366
The Last Word 368
Chapter 18: Storage Patterns for User Input 369
The World Health Organization: An Applied Example 369
Design of Your Spreadsheet File 370
The Input Wizard 371
Setting Focus to the First Input Cell 373
The Database 376
Input Entry Table 376
Database Information Table 380
The Backend Database Table 381
Menu Screen Functionality 382
Inserting a New Record 383
Editing an Existing Record 384
Deleting a Selected Record 387
Linking the Column of Country Names to the Form Control ListBox 388
Wizard Summary Buttons 389
The Last Word 391

Chapter 19: Building for Sensitivity Analysis 393

Weighted Average Models 394

Sensitivity Analysis on a Weighted Average Model 396

One-Way Sensitivity Analysis 397

Creating a Linked Values Table 399

Linking to the Database 402

Building the Tool 404

Getting to the Backend, the Intermediate Table 405

Scrolling Capability 406

Adjusting the Scroll Bar 411

Formula-based Sorting Data for Analysis 411

The Sort Column, Your New Best Friend 413

The Match Index Column, the Sort Column's Buddy 417

You Have a "Unique" Problem 419

Seeing It Work Altogether 420

The Last Word 421

Chapter 20: Perfecting the Presentation 423

Implementation and Design of the Weight Adjustment System 423

Displaying Data from the Intermediate Table 426

Results Information Label 426

The Current Rank of Each Country 427

Country Name 428

Total Scores for Each Country 428

In-cell Bar Charts for All Metrics 429

Best Possible Comparisons 431

Weight Box Progress Meters 431

"Sort By" Dropdown and Sort Labels 432

Dropdown Metric Selection 432

Using Boolean Formulas to Define Which Metric Has Been Selected 433

Connecting Everything with Conditional Format Highlighting 435

The Presentation Display Buttons 440

Going Back to the Menu 440

Resetting the Weights 440

Data Display and Aesthetics 441

Weighted vs. Not-Weighted Metrics 442

Color Choices 444

Data Spacing 447

The Last Word 448

■ **Part V: Data Models, PowerPivot, and Power Query 449**

■ **Chapter 21: Data Model Capabilities of Excel 2013 .... 451**

Relationship Advice 451

Preparing Data for Relationships 452

Setting Up the Relationships 454

Working with Relationships in the Excel Data Model 456

Exposing Relationships with Pivot Tables 456

Your First Multitable Pivot 458

Slicing and Dicing with Relationships 459

Rules for Happy Relationships 462

The Last Word 463

■ **Chapter 22: Advanced Modeling with Slicers, Filters, and Pivot Tables 465**

Creating an Interactive Excel Dashboard 465

Let's Meet the Data 466

Steps for Creating the Dashboard 467

Problems with Pivot Tables 480

Tips for Using Pivot Tables, Slicers, and the Data Model Effectively in Your Dashboards 480

The Last Word 481

Chapter 23: Introduction to Power Query 483

What Is Power Query, and How Can It Help You? 484

Bringing Data to Excel 484

Transforming Data 484

Loading Data to Where You Want It 484

Five Data Problems and How Power Query Can Solve Them 484

Problem 1: Unclean Data 485

Loading the Data in to Power Query 486

Removing Extra Spaces from the Name Column 487

Removing the Extra Periods 489

Fixing the E-mail Addresses 490

Bringing the Data Back to Excel 490

Problem 2: Partial Data Sets 492

Loading the Text File Data into Power Query 493

Saving and Loading the Query 497

Merging Customer and Phone Number Data 498

Giving a Proper Name to the Phone Number Column 501

Loading the Merged Data in to Excel 501

Problem 3: Data Consolidation 502

Taking a Look at the Old Data 502

Setting Up a New Connection 503

Appending Data 504

Problem 4: Duplicates in Your Data 507

Problem 5: Inconsistent Data Formats 509

The Last Word 514

Chapter 24: Introduction to PowerPivot 515

What Is PowerPivot? 516

A Note About How to Get PowerPivot 516

What to Expect from This Chapter 517

Getting Started with PowerPivot 517

Loading Sample Data into PowerPivot........ 517

What Is the PowerPivot Data Model?........ 519

Let's Enter the World of PowerPivot 520

Creating Your First PowerPivot Table........ 521

The Real Power of PowerPivot: DAX Formulas........ 523

Let's Create Your First-Ever DAX Formula Measure in PowerPivot........ 524

Let's Create a Few More DAX Measures........ 527

Example PowerPivot Report: Top Five Products Based on Average Quantity per Customer........ 529

PowerPivot and Excel Dashboards........ 532

The Last Word........ 533

Index........ 535

About the Authors

Jordan Goldmeier is an internationally recognized analytics professional and data visualization expert, author, speaker, and CEO. He is the owner of Cambia Factor, a data consulting agency, and a talent analyst for Ernst & Young. He is the author of *Advanced Excel Essentials*, a book on developing advanced analytics with Excel. He has consulted with and provided training for NATO Training Mission, the Pentagon, Air Force and Navy, Financial Times, University of Cincinnati, and others. His work has been cited by and quoted in the Associated Press, Bloomberg BusinessWeek, Dice News, and American Express OPEN Forum. He has held the prestigious MVP award from Microsoft since 2013 and is an owner and producer of Excel.TV, an online community devoted to sharing the stories, setbacks, and lessons from experts and practitioners in Excel, business intelligence, big data, and analytics.

Purnachandra "Chandoo" Duggirala is an author, speaker, blogger, teacher, and CEO. He is the creator of one of the most popular Excel and data analytics blogs, Chandoo.org, where he helps people become awesome in Excel. He is the author of *The VLOOKUP Book*, a book on various lookup formulas in Excel. He has worked as a consultant and trainer for many reputed organizations around the world, including Microsoft, KPMG, CapGemini, Novartis, and so on. He has held the prestigious MVP award from Microsoft since 2009. He lives in India and likes to travel, bicycle, and play with LEGOs in his free time.

About the Technical Reviewer

Fabio Claudio Ferracchiati is a senior consultant and a senior analyst/developer using Microsoft technologies. He works at BluArancio SpA (`www.bluarancio.com`) as a senior analyst/developer and Microsoft Dynamics CRM specialist. He is a Microsoft Certified Solution Developer for .NET, a Microsoft Certified Application Developer for .NET, a Microsoft Certified Professional, and a prolific author and technical reviewer. Over the past ten years, he's written articles for Italian and international magazines and coauthored more than ten books on a variety of computer topics.

Acknowledgments

This book would not exist if it weren't for the incredible Apress team. In particular, I have to thank Jim DeWolfe, whose patience and tireless efforts behind the scenes brought this book to completion. I don't even know how many times he stuck his neck out for me, but I imagine it was a lot. I also have to thank Chris Nelson, my development editor, whose intervention and wise guidance helped make this book into something great. Apress believed in the vision that is this book, and you wouldn't be reading this without them.

To the Excel bloggers, enthusiasts, analysts, accountants, and hobbyists—this book was inspired by your untiring effort to push Excel beyond its limits. This book owes so much to your imagination and creativity. Thank you, and please, never stop.

—Jordan

I must thank Jordan for inviting me to collaborate on this book. Special thanks to Chris Nelson for his patience and amazingly detailed reviews that made this book even more awesome. Thanks to Microsoft for creating Excel and helping millions of users around the world make sense of their data and for helping me to make a living from teaching people how to use Excel better. Thanks to all my readers, listeners, and students at Chandoo.org for supporting me and helping me indirectly in co-writing this book.

—Purnachandra

Introduction

This book will challenge you to think differently about Excel. *It will challenge you to think outside the cell.*

Most of us are not used to thinking differently about Excel. We're used to viewing Excel as a one-dimensional platform with a fixed set of capabilities. This myopic view has caused us to build dashboards and models, a not-so-small-number of which are perceived as being slow and clunky. Some have even accepted "slow and clunky" as the compromise to building dashboards in Excel. But it doesn't have to be this way. You can create intuitive, interactive data and analytics applications in Excel without sacrificing speed and utility, if you know how.

This book will teach you how. But it won't happen through a series of clicks or rote memorization. Indeed, memorizing everything in this book will likely not help you. To master the techniques in this book, you must learn to think creatively and differently. You must be willing and perhaps even courageous to challenge Excel's perceived limitations. You must be able to extend focused examples into your own work, adapting them to new scenarios and different problems. Luckily, as an analyst and professional, you likely already have the creative gene to get the job done. This book will show you how to apply it to Excel dashboards and development.

What to Expect

Here's what you should know. This is an intermediate to advanced-level book. We assume you are not a novice to Excel formulas and have some degree of experience with Visual Basic for Applications programming, even if it's just the macro recorder. If you are open to learning new things, this book will teach you. Remember, the most important skill is being able to think about problems differently. Chances are you already do.

This book will not help you pick the correct metrics for your dashboards, but it will help you present them in the best way possible. This is not a book dedicated to pivot tables, VBA, or formulas but rather to how you can use them appropriately, and sometimes counterintuitively, to create engaging and meaningful analytical tools.

Much of this book will focus on reverse engineering complete designs. There are many reasons we chose to present examples in this way. For one, we want to establish the use of reusable components. We'll go into reusable components later in the book, but for now, you can understand them as collections of formulas, charts, form controls, and other Excel features that allow you to repeat and implement similar design patterns across different scenarios and applications. By reverse engineering completed work, you'll see that most applications in Excel share very similar constructions that are easy to build and ready to be reused in different products.

More important, however, is that real life rarely allows you to start from scratch. More often than not, you look to the work of others for both guidance and inspiration. Perhaps you are taking over a dashboard designed by your predecessor. Or, maybe you're adapting the solution you found on an Excel forum after an Internet search. Whatever the reasons, being a good developer requires that you know how to read, understand, apply, and adapt the work of others. How to take a small concept and apply its lessons across a range of other scenarios is a skill worth developing in its own right, and to the extent possible, this book will challenge you to do just that. It's not enough to be good at Excel development; you must learn to think like a developer. This is how Excel can transcend its unfounded reputation.

How This Book Is Laid Out

This book is split into five parts.

Part I, “Dashboards and Data Visualization”

Part II, “Excel Dashboard Design Tools and Concepts”

Part III, “Formulas, Controls, and Charts”

Part IV, “From User Interface to Presentation”

Part V, “Data Models, PowerPivot, and Power Query”

Part I introduces what it takes to make an awesome Excel application: good development practices, good data visualization principles, and thinking outside the cell. “Thinking outside the cell” refers to when and how you can use Excel differently. We also review dashboards, explaining how some are successful and why many are not. Finally, we cover data visualization principles so you can understand the best ways to present your data on your dashboards and decision support systems.

Part II explains what’s required for to build and design a dashboard in Excel. Specifically, we’ll focus on what to do and what not to do, while applying the data visualization principles outlined in Part I. Moreover, we’ll walk you through how this book uses VBA and formulas differently than you might be used to. Part II is filled with new ways to use Excel features you might already be familiar with, such as form controls and shapes.

Part III will investigate a Gantt chart dashboard used for workforce planning. This part will apply much of what was presented in Part II to implement a fully functional dashboard system with visual and interactive elements.

Part IV will investigate how to build a decision support system used to investigate different healthcare systems across different countries. The chapters in this part will go through taking in user input, storing the input in a back-end database, and displaying the final results to the user. This part also introduces some easily implemented analytics techniques from the field of management science.

Part V was written by Purnachandra, a.k.a. Chandoo, and he shows how many of Excel’s new features can help build dashboards similar to those presented in the first four parts. These features include data models, slicers, and PowerPivot and Power Query. These new features make up the new Microsoft Power Business Intelligence platform.

Above all, this book will focus on thinking creatively. It will focus on showing you how to hone your skills as an analyst to use Excel to create work that will astound your boss and colleagues. Most of all, this book is about helping you take your Excel skills to the next level. This book is a journey, and you’re just at the beginning. The road ahead is long, and the challenges are formidable but well within your grasp if you keep at it.

Good luck on your journey. I know it will be a good one.

—Jordan Goldmeier

PART I

Dashboards and Data Visualization

Chapter 1. Introduction to Dashboard and Decision Support Development

Chapter 2. A Critical View of Information Visualization

Chapter 3. The Principles of Data Visualization in Microsoft Excel

Chapter 4. The Excel Data Presentation Library

CHAPTER 1

Introduction to Dashboard and Decision Support Development

Microsoft Excel has proved itself to be a powerful development platform, and the need for material on advanced Excel development has never been greater. In recent years, however, few books have afforded developers the instruction and depth required to create advanced Excel applications. Two titles in particular, *Professional Excel Development* by Bullen, Bovey, and Green and *Excel Dashboard and Reports for Dummies* by Michael Alexander, have given serious focus to these subjects; yet a dearth of quality material still remains. Where printed material has failed to fill this gap, thriving communities and blogs have spawned on the Internet over the last few years devoted to Excel development. Naturally, the last few years have also seen growing interest in data visualization, business intelligence, and data analytics.

The Data Problem

We are so immersed in data that it's hard to make sense of it all. So far, our solution has been to rely on products and people who purportedly transform this data into something we can digest; vendors, smartphones, computers, and even the news media slice and dice this information to make it palatable for us to consume. But even these devices and people can, and often do, fail to deliver; the institutions we've entrusted for guidance through this mess have not always had our interests in mind. Where we should have met their work with skepticism, they have imbued in us a false sense of security. Intentional or not, they have led us astray.

Thus, we exist at an important and exciting time for professional Excel developers. With knowledge of Microsoft Excel spreadsheets, formulas and Visual Basic for Applications, and the desire to learn, developers can create powerful data and visualization platforms that rival the work performed by vendors. In other words, with the right tools—and the right mind-set—we can take control of our data. We can take the power back from the vendors.

However, the desire for more data visualization and analytics and the ability to create these products in Excel has not always translated into better products, especially in the last few years. Data visualization is often misunderstood by the vendors that sell it. This misunderstanding has crept its way into Excel. Some vendors and even Excel blogs sell products or instruct users on how to create fancy but useless widgets that sparkle with metallic finishes in three dimensions. But all that glitters is not gold. Just because we can add these things to our Excel spreadsheets does not mean we should.

At its core, the problem stems from a lack of clarity. At every turn we are confounded by information. When products and research fail to make us smarter, they add to our confusion. If we are to see clearly through the data storm, we must stick to what we know—what the research tells us is true. This book will not teach you how to make the newest flashy gadgets; instead, it will teach you established principles based on proven research. We will continue down the path so rarely taken by vendors: the path of research and best practices. Think of it as an adventure.

Enter Excel: The Most Dangerous Program in the World

No, really, it's possible that Microsoft's Excel is the most dangerous software on the planet.

—Tim Worstall in *Forbes Magazine*

Worstall is referring to JPMorgan's use of Microsoft Excel and its potential contribution to the global recession that began in 2008. The story is a familiar one. JPMorgan, one of the big banks, required a new way for their chief investment officer to model the risk associated with their portfolio of collateralized debt obligations (CDOs) . CDOs, as you might recall, are those complex financial instruments that are comprised of a bundle of mortgages used to speculate and manage the risk for default. In the end, many homeowners eventually defaulted on their mortgages. Yet, JPMorgan's model had not captured this at all. According to an internal report by JPMorgan, their model contained an error that "likely had the effect of muting volatility by a factor of two and of lowering the VaR [value at risk]."

What had happened? How could one of the biggest banks employ a model that was so absent from reality? According to the report, their model was "operated through a series of Excel spreadsheets, which had to be completed manually, by a process of copying and pasting data from one spreadsheet to another." The error was generated when "After subtracting the old rate from the new rate, the spreadsheet divided by their sum instead of their average, as the modeler had intended."

Worstall, like many others, blames Excel. It's too insecure, and there's no audit trail, he contends. Its flexibility is its greatest strength and Achilles heel. His argument isn't a new one, of course. Without proper internal governance and procedures regarding archiving, getting lost in a nightmare of error-ridden spreadsheets and mixed-up versions is a real possibility.

But is that really what happened in the case of JPMorgan? Are more audit controls needed? I'm not so sure.

Not Realizing How Far Spreadsheets Have Come

Many organizations still operate with the mind-set that Excel is an incredibly hard tool to use. Early versions of Excel may have required manual updates, but these days Excel provides a lot of efficient and easy ways to pull in information. As well, what reason was there to maintain many different spreadsheets? Why did they require someone to copy and paste the information with every update to back-end data? Macros easily streamline these rote activities.

It's troubling then to think a large financial institutional such as JPMorgan didn't choose to use all the capabilities available in Excel. Instead, they tended to use it in a very antiquated way. But more than that, I submit that when you don't take advantage of the features that exist to address your needs, not only are you using Excel inefficiently, you're using it incorrectly—and dangerously.

One of the main reasons I still see organizations operate with the mind-set that they need to maintain many different spreadsheets with manual updates is because they want a historical imprint of *every piece of information for every step in the process*. These files become the historical records, or in other words, the audit trail of changes. According to their internal report, JPMorgan knew that operating with several spreadsheets was foolhardy. But for many organizations—and quite possibly for JPMorgan—operating with several spreadsheets can give the appearance of having a trail of changes. JPMorgan knew they had to automate the model's process but never did. Where was the urgency?

Are more audit controls to track every change really the solution? It would have certainly helped. But I don't think it addresses the problem in full.

Garbage In, Gospel Out

Another major contributing factor was that nobody seemed interested in validating the numbers the model was spitting out. Even when their numbers seemed different from that of the other banks, JPMorgan failed to properly validate the model to verify its calculations were correct. Why? One reason might be because JPMorgan liked the numbers coming out of the model.

This is a common problem in my experience. We're less likely to double-check results that seem viscerally desirable. Often these results take the form of these three cases:

- When the models provide information that appears consistent with everyone else's numbers, we may meet the numbers with less skepticism.
- When our models provide information that is counterintuitive, we may again be less skeptical because we think we might have found a counterintuitive response.
- When our models seem to average information together, giving us a grand bargain between one or more competing ideas, we may think we have results that everyone should like.

Part of the problem is how we view models. If a lot of money and research goes into developing a model, we're less likely to turn our backs on all of the work we've performed even when we suspect there are issues. We might think we can save it, if only we had more time and resources. However, some work is just not usable when we realize it's wrong. We can try to ignore its results, but it's hard to ignore results even when we know they're bad. Think of it as an incorrect entry on a crossword puzzle written in ink: even when we want to ignore it, we often forget that the entry is wrong. Bad models affect our ability to judge in the same way.

When our model delivers results that are incorrect, however seemingly truthful, the model can be described as being "garbage in, garbage out." When management believes the results no matter what, we might describe management's behavior as "garbage in, gospel out."

■ **Remember** Think outside the cell. Excel is a great tool. I love it. But it can't do everything—and we shouldn't force it to do what we want to verify our own biases.

How Excel Fits In

The unfortunate reality is that Excel cannot make decisions for us. We cannot use it like a crystal ball to gaze for the truth into the future. We cannot pretend that disparate and conflicting information might form a real fluent idea as if we were reading tea leaves in high resolution. In fact, no program, no mechanism, and no promise however earnest can do this for us.

We should start by knowing what Excel can and cannot do for us. If we use Excel correctly, we can bring important decision-making insight to our management. If we try to turn it into something it's not, we might mistakenly misunderstand or even waste its potential. At best, if we misuse our tools, we make a mistake that is well understood and correct it immediately; at worse, we follow down a path of bad results unaware that they are bad.

What Excel Is Good For

There are several good reasons why Excel might be useful for our project. Among other reasons, Excel is

- A no-nonsense commercial-off-the-shelf (COTS) solution
- Flexible and customizable for our needs
- Inexpensive-ish
- Familiar to many in the industry
- A good return on investment
- Mostly backward compatibility

A Commercial-Off-the-Shelf Solution

Commercial-off-the-shelf (COTS) software is a common term among consumers of software and information systems. In particular, the U.S. federal government holds a special designation for COTS software packages. The reason COTS solutions are preferred is their name: they're "off the shelf." By that, we mean that COTS products are essentially ready for us to use as if we purchased them off the shelf at a retailer. Most COTS products have some assembly required, but they represent a piece of software that doesn't require a significant undertaking and investment into infrastructure to implement, train users, and maintain operations.

And this is where Excel really brings the advantage. There are many data visualization packages out there, and many of them require a certain amount of data architecture or "business intelligence" to implement. Excel doesn't require any of these technologies to begin building dashboards and powerful data visualizations.

Flexible and Customizable

Even though we can characterize Excel as an off-the-shelf solution, it still brings a strong amount of flexibility and customizability. By flexible, I mean that Excel gives you many options in which you present and visualize your data and analysis. By customizable, I mean that you are able to do things to your spreadsheet to make it indiscernible from a regular piece of software. For example, you can protect certain cells from user input. You can even remove the tab banner at the top of the screen when you send out your dashboard. These things allow you to essentially modify the user experience from looking at an Excel spreadsheet to looking at a dashboard.

Familiarity and Ubiquity

Excel is ubiquitous. Currently, Microsoft Excel is the most widely used spreadsheet platform across the world. With familiarity comes less training. Because Excel is so widely used in industry, chances are your customers don't need intense training to just pick up your work and use it right away. As well, your Excel solutions *should* work on any computer that also has Excel because the solutions are hardware independent (to a certain extent—there are notable limitations, and we'll talk about how to overcome some of them later in the book). Finally, because many users are often behind—that is, they are using versions of Excel older than the most recent one available—Microsoft, as a general rule, makes worksheets fairly backward compatible. Even if you develop in a new version, chances are most if not all of the functionality will work in the most recent previous version.

Inexpensive-ish

I say "ish" because Excel isn't incredibly cheap, especially as you move into the more expensive versions of Office, such as Professional and Professional Plus. As of this writing, the latest version of Excel, Excel 2013, was selling for $499 for the Professional Plus edition. That said, $499 is still a cheaper license than many of the full-blown data visualization packages. But chances are, you don't have to worry about buying Excel new at your local office store. You probably purchased a computer that came with a Microsoft Office Professional license—or, your company already has some license agreement with Microsoft. Whether it came with your equipment, it was purchased on your behalf, or you had to purchase it for yourself (or for your business), the purchase represents a small fixed cost to you and essentially no additional charge to carry over to your customer.

Quick Turnaround Time

In my experience, Excel solutions just don't take as long as other software solutions. On the front end, you can add and modify the format with ease and without code. As well, the coding language behind the scenes, Visual Basic for Applications (VBA), provides a high-level ability to modify the spreadsheet on the fly without tons of coding.

SPREADSHEETS AS PROGRAMMING

Many of us don't think of developing a dashboard in Excel as developing software or programming, even when we employ VBA. However, recent research suggests that a spreadsheet is a form of software programming, albeit in a less rigid environment. In *Quality Control in Spreadsheets: A Software Engineering-Based Approach to Spreadsheet Development*, the authors write the following:

The process of formula construction in spreadsheets is basically a form of computer programming.[1]

A Good Return on Investment

This should be a no-brainer at this point. When we consider all of the advantages discussed earlier, it's not hard to see why Excel could provide a much better return on investment than a big, expensive data visualization package. But the proof is in the pudding: managers will see the good work you do and see how much they've paid for it.

What Excel Isn't Good For

There are lots of things Excel cannot do (and should not be expected to do). Excel should not be used in the following ways:

- As a full-fledged database
- For enterprise-level reporting
- For replacing a full software package
- For predicting the future

In the next sections, I'll review cases in which Excel is not the best tool for the job.

[1]Rajalingham, Kamalasen, David Chadwick, Brian Knight, and Dilwyn Edwards. "Quality control in spreadsheets: a software engineering-based approach to spreadsheet development." In *System Sciences*, 2000. Proceedings of the 33rd Annual Hawaii International Conference, pp. 10. IEEE, 2000.