DATA-DRIVEN ORGANIZATION DESIGN

Sustaining the competitive edge through organizational analytics

RUPERT MORRISON

KoganPage

CONTENTS

PART ONE Introduction 1

1.1 Data-driven organization design 3

Effective organization design and management 5
Why people are struggling with organization design 6
How this book will help you 8
Structure of the book 10
Final thoughts 10
Notes 11

1.2 Challenges 13

Introduction 13
The reality of the challenge 14
Common challenges in organization design 17
Final thoughts 23
Notes 24

1.3 Foundations and core concepts 25

Introduction 25
Foundation 1: The organization is a system 26
Foundation 2: Organizational data is hierarchical 31
Foundation 3: Organizational data is messy 33
Applying the foundations in practice 35
Method 1: create hierarchical data structures 37
Method 2: connect the system 40
Method 3: visualize to analyse 46
Final thoughts: see it as a journey 56
Notes 57

PART TWO Macro design 59

2.1 Introduction 61

Macro design in practice 62
Structure and logic for Part Two 63

2.2 Strategy articulation and design criteria 65

Introduction 65
Define where you are going and why 66
The case for change 71
Constraints and risks 75
Setting the design criteria 76
Final thoughts 77
Notes 79

2.3 Structural options and business case 80

Introduction 80
Develop options 82
Analysing and deciding your best option 89
Develop and sign-off the business case 97
Final thoughts: Have a strong case for change 99
Notes 100

PART THREE Micro design 101

3.1 Introduction 103

Structure and logic for Part Three 105

3.2 The data goldmine 108

Introduction 108
Data blockers and myths 109
Building the baseline – get ready to wrangle 113
Performing analysis: common statistical traps 122
Final thoughts 130
Notes 131

3.3 Objectives management 133

Introduction 133
Why organizations struggle with corporate-wide performance management 134
Process for setting and managing objectives 138
Maintaining and improving objectives management 148
Final thoughts 150
Notes 152

3.4 Fixed process design 153

Introduction 153
Types of process map 154
Understanding the as-is 159
Analysing the as-is 163
Design the to-be process and structures 170
Link the to-be work to to-be positions and hierarchical structure 178
Final thoughts 183
Notes 184

3.5 Dynamic process design 185

Introduction 185
Dynamic process design methodologies 186
Detailed example: project-driven organizations 188
Mini-examples 196
Final thoughts 200
Notes 202

3.6 Competency management 203

Introduction 203
Objectives versus competency management 205
The business case for competency management 206
Developing and putting the framework to work 210
Final thoughts 218
Notes 219

3.7 Rightsizing 220

Introduction 220
Common currently used flawed methods 221
Four methods for rightsizing your organization 222
Method 1: Ratio analysis 223
Method 2: Activity analysis 230
Method 3: Driver analysis 236
Method 4: Mathematical modelling 243
Final thoughts 247
Notes 248

PART FOUR Making it Real 249

4.1 Introduction 251

What is Making it Real? 251
Realize the design is only a sketch 253
Structure and logic for Part Four 254
Notes 256

4.2 Common traps 257

Introduction 257
Knowing when and what things are going wrong 257
Final thoughts 262

4.3 HOWWIP 264

Introduction 264
HOWWIP outcomes 265
Exercises 268
Final thoughts 273
Note 274

4.4 Implementation 275

Introduction 275
Impact assessment 276
Communication 281
Transition management and consultation 285

Collateral 293
Final thoughts 294
Note 295

4.5 Workforce planning 296

Introduction 296
What we really mean by workforce planning 297
How micro design helps to drive workforce planning 301
Workforce planning in practice 302
Final thoughts 307
Notes 309

4.6 Talent management, succession development and succession planning 310

Introduction 310
Why all the hype? 311
Approaches to succession planning 313
Implementing talent and succession planning 319
Final thoughts 321
Notes 322

4.7 Organizational getting things done 323

Introduction 323
The power of the list 324
Getting things done in a team 326
Using a data-driven approach to get things done 330
Final thoughts: Use the right management tools 333
Notes 334

4.8 Sustaining the edge 335

Introduction 335
Tracking the system 335
Celebrating, reflecting and resting 340
Final thoughts 342
Notes 343

Appendix 345

The data goldmine

3.2

Introduction

There is currently a gap. A business opportunity if you will: to connect your people with business performance; to tie the crucial elements of the business together by tracking and managing drivers from across the organization, and understanding their impact. Data lie at the heart of achieving this – linking all the elements of the system together by bringing organizational and business data together.

We are living in the information age and the pace of growth in human knowledge is exponentially increasing. In 1965 Gordon Moore noted that the processing power of computer circuits had doubled every year since their invention and predicted, correctly, that this trend would continue into the future, except now the accepted time is 18 months (Moore's Law). As Buckminster Fuller predicted, new knowledge, which doubled every century until 1900, is now estimated to double every 18 months.[1]

This is a hard thing for the mind to grapple with. For example, take a piece of paper a tenth of a millimetre thick. Fold it ten times and it will be will be 102 millimetres thick; 20 times takes you to 105 metres, 30 into space at 107 kilometres, 42 to the moon, and after 51 folds you reach the sun. After 103 folds you will be outside the observable universe, which is estimated at 93 billion light years in diameter.[2] With exponential growth comes exponential value. For example, in the 1990s it cost billions to map the human genome. Today it costs less than $3,000 and that is predicted to drop even further.[3]

The most successful organizations of today are those that can harness these exponential gains of the information age. For example, when Google first indexed the internet in 1998 it found 26 million pages; it now indexed over a trillion unique URLs.[4] And Facebook gained more than a billion users in just 10 years.[5] Businesses are sitting on a goldmine of organizational data that, if used in the right way, could return exponential amounts of value by connecting the organizational system. However, many still struggle to simply track headcount.

A lot of the basic building blocks for getting the most out of organizational data is currently missing: ensuring data quality; having consistent role titles; counting FTEs by a common understanding of grade, function or geographical unit. Organizations need an end-to-end data-driven approach to process and maintain organizational data over time and connect the various elements of the system. Today we are at the beginning of this journey. A shift is emerging as organizations begin to appreciate the value of this data more. For example, Norwegian retailer Elkjøp, working with psychological testing company Cut-e, increased sales performance of new hires by 14 per cent. They identified the personality types most effective in sales roles and by screening candidates they were able to target recruitment. Then by connecting their recruitment profiling data with sales data they were able to compare the two and measure the value returned to the business.[6]

In this chapter, I set out a method for building a baseline of data and gaining an understanding of your organizational as-is. I begin by exploring four data myths, which have traditionally hindered those working with organizational data from making progress. I then provide an end-to-end seven-stage process for how to collect, understand and maintain a baseline of data. Finally, the last section investigates how to analyse your data effectively so that you are getting the answers you need and you can avoid some common statistical traps when analysing it. This chapter provides the platform for the rest of the book. Building a comprehensive and focused baseline is essential for designing and managing your organization, because if you don't know where you are, how can you know where you are going? By getting your data processes right you are setting yourself up for success throughout the rest of your design.

Data blockers and myths

A trend has emerged over the last few years that spending on technology is increasingly moving away from IT function and towards specific digital units.[7] One of the best examples of this is the rise of digital marketing. The ease with which marketers have been able to gather and analyse data has transformed the industry: from how consumers are targeted down to the very skills considered essential to become a marketing professional. As Marketo's chief marketing officer Sanjay Dholakia commented, 'Marketing's changed more in the past five years than in the past 500 years.' It is a 'whole new learning curve for marketers, who are starting to have to be more and more technical.'[8] Indeed, in 2012, Gartner analyst Laura McLellan declared that by 2017 the Chief Marketing Officer would spend more on IT than the Chief Information Officer.[9]

People and organizational effectiveness functions are increasingly seeing the need for investment into systems that make the most of their data. However, they are lagging far behind other business functions such as marketing. Why have other organizational functions got so far ahead in their utilization of data? For a start, in contrast to functions like marketing and finance, organizational analytics are less obviously connected to business outcomes, and so by their nature it is harder to demonstrate their value. More importantly, though, I believe it comes to an approach. Unlike marketing and finance, functions typically dealing with organizational data has been slower and less eager to embrace a data-centric approach. The result is they have found it harder to substantiate arguments and demonstrate value to the business, in turn making it hard for them to get the funding required to put people at the top of the business agenda. A range of data myths has emerged over the years which have obscured, blocked and been provided as excuses for not embracing a data-centric approach. In this section I outline four of these myths to help highlight how the systematic process of building a baseline (covered in the next section) can help you make the most of your data and overcome these traditional blockers.

- Myth One: 'Data and technology is scary.'
- Myth Two: 'Organizational data is too hard to process.'
- Myth Three: 'One system fits all.'
- Myth Four: 'More data means more insight.'

Myth One: Data and technology is scary

Two years ago I worked on a project with an HR manager who was extremely competent in many aspects of HR. We were looking at a spreadsheet containing employee pay, pay increase and bonus. I asked her to calculate the total pay increase (a simple calculation of pay increase + annual bonus) and it emerged she had never learned to use the sum function in Excel. Despite holding a master's degree in HR, and five years of work experience, she had not moved beyond the stage where Excel terrified her.

Too often when working with people who have traditionally dealt with organizational data, I hear 'I don't do numbers' or 'I don't do data.' Unlike the marketing function, organizational analytics have not been made a business priority. There is a lack of urgency to appreciate and understanding for the value data can give to enhance people functions and employee relations. For example, how a data-driven approach to areas such as objectives or competency management can help individuals improve performance and develop by identifying specific areas for improvement. Professionals working with organizational data is going to have to transition from a soft-skill focus and combine them with hard skills, if they are to keep up with skills required for the demands of data-driven organizations. Likewise, organizations that do not invest in this area will find themselves missing out on a huge area of potential business value. To help address this myth explicitly I have put together a set of definitions for technical terms when handling data and building a baseline (see Figure XXX in the Appendix).

Myth Two: Organizational data is too hard to process

As I highlighted in Chapter 1.3, organizational data is naturally complex and messy. It can feel impossible to process because it is stuck in numerous isolated Excel islands and systems. Data professionals have reported that cleaning and transforming their datasets is an extremely time consuming and boring part of their analytical workflows, often comprising of 80 per cent of the work.[10] When dealing with organizational data this represents a real challenge, and can often be used as an excuse to not begin to use data.

Data quality should never be used as an excuse. You have to start somewhere and it is only by using data that you will be able to understand what data is valuable, what is not, and where improvements are needed. There is no doubt that data is hard work, especially to begin with when building your baseline. This means employees need to have the benefits the data will bring to themselves and the organization as a whole communicated to them. Too often the employees handling data get little personal value from the outputs of their work. It is no wonder, therefore, that excuses are found to avoid working on what is perceived to be a thankless task.

Myth Three: One system fits all

Technology is achieving some amazing things, with a range of tools now on the market to help those working with organizational data (for a discussion on types of technology see XXX). And yet, Excel and PowerPoint are still the main tools of choice. Excel is great for many things but it has many flaws when dealing with organizational data: it isn't that visual; links between worksheets are easily broken and hard to track; it struggles to handle hierarchies; and it is typically used offline by one person at a time, making duplications very easy to make.

Static use of technology has meant organizations are struggling with their data, and professionals find it time consuming and uninspiring to process. As our understanding of data and their possibilities has grown, so has the refinement of the tools available to us. Depending on the data you are collecting and what you are trying to achieve with them, you may require a different tool. At this point in time, I think there are four broad options you can consider to aid your approach to your data: a data warehouse; an Enterprise Resource Planning (ERP) system; a data intermediary; and specific visualization software. A data warehouse is the standard choice, but as already mentioned in Chapter 1.3, it does not suit organizational data very well. Going down the data warehouse route is a hugely complex, long and painful journey, only solved by brute force and herculean commitment in order to collect, structure and query the data. ERPs are a vital organizational tool because of their ability to integrate and handle flows of information from across the organization. However, they often leave you short of flexible analytics and reporting capabilities, and the ability to model and run organizational scenarios so key in organization design. As the name suggests, data intermediaries are beginning to offer a middle, more niche ground. These are software and applications designed for specific organizational events rather than overarching operations; for example, specific organizational modelling software. Finally, out-and-out visualization software allows you to perform advanced analytics on top of your data, although it may lack some of the modelling capabilities of intermediary software and have very few operational capabilities. The trick is to work out exactly what you are trying to achieve and then choose the best tool for job. Having the right tool makes the data less scary and easier to process along every stage of the data journey. And given the potential value, the investment is certainly worth it.

Myth Four: More data means more insight

Just because organizational data represent a goldmine, much of which hasn't been exploited yet, it doesn't mean you need to collect all of it. Think of data as a value chain, starting at source systems and ending in actions and behaviours. Good Business Intelligence (BI) starts by thinking through the information needs for the highest priority actions and decisions first, and working backwards. Too often an IT team will be told 'Give me BI' without guidance about the end requirement or output. The team then starts by building a data warehouse with the information it has from various systems and never moves beyond that point. I have seen companies with a ten-year plan to build a warehouse, and other companies spend millions to achieve little more than the insight they started with.

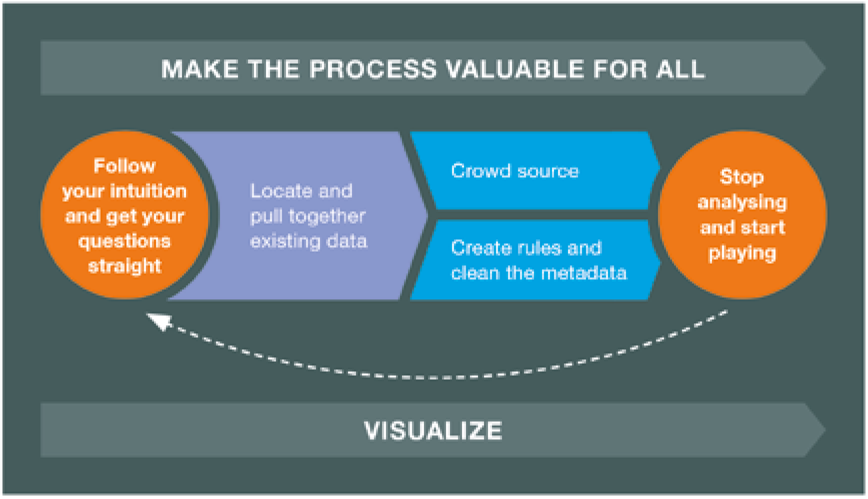
There is so much data around that often the challenge is actually to specify and focus on what you are trying to achieve and, therefore, which data is relevant. Too much information is no information. I've heard of consultants compiling lists of 800+ HR metrics in the belief that if you think of everything, then you will be able to answer any question. However, collecting too much data will overburden people and cripple your data collection, and analysis processes will be more difficult, running the risk of nothing getting delivered. Much better is to collect high-priority data and then add to your baseline as you go. Remember, there is nothing stopping you from collecting more and new data further down the line.

Most of these myths are due to a mind set towards data. By approaching data in the right way you can overcome many of the barriers that have traditionally 'stopped' organizational analytics being at the heart of business decisions. Much of the challenge is simply a question of discipline when building up and maintaining a reliable baseline of data to work from. The next section provides a systematic process for how to build a baseline, taking into account and tackling head on the myths addressed in this section.

Building the baseline – get ready to wrangle

As I have already highlighted, organizational data can be a nightmare to work with; incomplete, hidden in numerous systems and spreadsheets, and constantly moving. No set of organizational data will ever be absolutely perfect. So get ready to wrangle. Data wrangling is at the heart of building a baseline, and covers the steps of cleaning, converting and manipulating data into a usable, convenient format. I love the idea of wrangling with data. The word *wrangle* captures beautifully the essence of the battle with data to keep it up to date, uniform and useful. And it is a battle. Making the baseline process as simple and as accessible for employees as possible will immediately get you a great deal nearer to business-changing insights. In this section I go through seven iterative steps of how to go about building a baseline set of data as shown in Figure 3.2.1. These include the steps from deciding which data to collect, right through to the collection, visualization and analysing of that data.

FIGURE 3.2.1 Building a baseline process



1 Follow your intuition and get your questions straight

For many years, strategy consultants have used a hypothesis-led approach to focus investigations – and it works. Data and analytics are not all about facts and figures. Often the best insights come initially from intuition, investigation and creative thinking. Therefore, when building a baseline of data think beyond the simple organizational metrics such as headcount and absence. Go back to your design criteria to help direct your focus. For example, if a key part of the design criteria is to boost productivity through centralizing the business, come up with a series of questions which will help you understand productivity throughout the business: Are some departments more productive than others? Does training affect productivity? Where do our top performers come from? These types of question will help direct the metrics you need to collect to build your baseline.

When doing a major transformation there are certain sets of metrics that you will almost certainly need, regardless of the focus of the project. A sample list is shown in Figure 3.2.2. Remember though, your baseline is not a one off, but an ongoing set of data which should support the ongoing development of your organization. So when tackling any problem, refer back to your intuition and the fundamental questions to guide your approach.

FIGURE 3.2.2 Example baseline properties (fields)

Category

Type	Must have (20)	Should have (20)	Could have (28)	Calculated (17)
Business (8)		Actual sales Customer satisfaction Sales targets 3	Gross revenue Net revenue Revenue Process level cost Direct/indirect cost 5	
Core OD (16)	Number of hours worked (per week) FTE Function/department ID number Manager ID Title (role; post) Area (geographic) Vacancy Country Grade 10	Office location (City) Is temporary Subfunction 3		Is current employee Span of control Team potential rating 3
Demographic (12)	Birthday Gender Leave date Start date 4		Ethnicity Role start / appointment date Expected retirement age 3	Age Tenure Tenure in current role Tenure in previous role Years to retirement 5
Financial (22)	Current salary Currency Currency exchange rate 3	Share-based payments Cost centre 2	Cost of termination Current bonus Expenses Healthcare costs Other benefits Payroll taxes Pension Recruitment cost Relocation costs Total benefits Total payroll cost Functional cost 12	Currency converted bonus Currency converted employee costs Currency converted salary FTE salary Total cost 5
Personal (9)	Email address First name Surname 3	NI or tax number Other unique identifier Personal address Photo work Telephone number 5		Full name 1
Talent and performance (18)		Absence detail Employee engagement Employee potential rating How recruited Prior period performance ranking Successor information Training days 7	Absence days Absence instances Absence type Left voluntarily? Performance ranking Reason for leaving Training cost Target performance distribution 8	Bradford factor Team absences Performance rating category 3

Category: Must have · Should have · Could have · Calculated

2 Make the process valuable for all

As a general rule, people don't like to provide data. It costs them time and effort; it can cause embarrassment and show their shortcomings. Even if you have top-down sponsorship, your priority won't necessarily be the same as every employee. Not everyone is intimidated by authority, and those providing data can be highly creative in their delay tactics. So make the process worthwhile for them. At a high level, make sure you can answer the key questions of why you are collecting the data, how they are going to help, what the consequence of not getting the data will be, and if there is any upside for them. A good exercise is to tie the data back to the list of questions in step 1. The answers should be compelling enough to get buy-in.

At the lower data-entry levels, for those performing the day-to-day task of entering and cleaning data, you have to answer the data owner's concerns: 'If I'm going to input data, what's in it for me?' There are two methods here: first, ensure the data owner can see the value they are contributing or getting back. One way of doing this is ensuring they can see the end-to-end process. They can then understand the holistic benefits the data will lead to. Second, make the process into a game. For example, on one project, a network of hospitals wanted to clean their data. However, none of them was progressing. To incentivize them we put up a map of how each was progressing and who was top week by week. Just a week after the map went up, the progress made increased exponentially. Gamifying the process through visual feedback and competition made the process fun and part of a team effort. To get the most out of your data you have to be creative. Not just with how you analyse them, but getting them in the first place. After all, data sit with people, so you have to find ways to incentivize them and get them to buy into the process.

3 Locate and pull together existing data

It amazes me how often organizations jump straight to implementing new data collation and creation, overlooking their existing data. Often you already have in place a lot of the data you need in order to see the as-is picture; you just need to bring it all together. By trying to get insight out of your data early on in the process, you can:

- show those working hard to provide the data that they are being used and how you are using them;
- start to answer some of your hypothesis and make sure you are collecting the right data for your needs;

- gain a much better understanding of the data cleaning and further collection processes ahead.

The starting point is to be extremely clear about the data you need to collect. Figure 3.2.2 shows a sample list of properties (or fields in relationship database speak) that are frequently collected as part of an OD project. Note that I have categorized them by two dimensions. The first is the level of priority. Some properties are critical and you can't build your baseline or do an OD project without them. These nine properties fall within the critical dimensions with eight in core OD and one in financial types of data.

Once you know what data you need, the next step is to determine where to source them. In doing this, understand:

- What systems are used in the organization and where are they being used? For example, often different geographical locations will use different systems.
- How do the required properties map to the various systems and which properties currently don't exist in any system?
- Who owns which system and what is the process for getting the data extracted?

If you need to repeat the extraction process, ensure you document and test the process. Good sources to begin collecting data from include:

- Payroll – Typically this is the best source because it should contain all your people information. It is certainly a good place to start because, if you then match it up and compare it with your other sources, you can begin to see if you have a consistent view of your organization. You would be surprised how often payroll doesn't match other source systems.
- Performance management systems – Comparing this source system with payroll is a great way of testing the quality of your people data. If they are the same, then great, but if they are different you need to look at your data quality.
- Excel spreadsheets – I have not seen an organization yet that does not have Excel islands, with pockets of information siloed away. It is a problem we constantly have to tackle in my company and with our clients. However, bringing all this data together is a good chance to collect all that 'lost' data.
- HR Management Information Systems (HRIS).
- Enterprise Resource Planning (ERP) systems.

- Customer Relationship systems (CRM).
- Financial systems for customer revenue and margins, and P&L cost information.
- Applicant tracking systems.
- Learning management systems.
- Quandl for externally published sources, eg unemployment, demographic, or economic growth.

4 Visualize

Visualizing data is a key part of the data and analytics processes as highlighted in Chapter 1.3. When visualizing the data try to stay away from Excel, and where possible use specifically designed software that helps bring all your data together in an interactive and visual way. Visualizing the data as early as possible in the process will bring it to life and achieve buy-in from people working with it and affected by it. It may also start uncovering some basic insights to help reaffirm your hypotheses and ensure you are going in the right direction.

Think back to your original hypotheses. What are the best ways to look at the data and what stories are you trying to tell? For example, if you are interested in salary across the spans and layers of your organization, take your existing ERP data with the given reporting lines. This only requires five properties of data: Employee ID, Manager ID, function, salary and grade. Using this data you can answer dozens of questions about the organizational structure and consistency of salaries across the organization. To tell this story, think back to some of the visualizations in Chapter 1.3 such as the box grid visualizations (Figures 1.3.13 and 1.3.14). Cycle the data through these visualizations so you can see different perspectives. When looking for further insights think through the various visualization options:

- traditional statistical representations (bars, columns, lines, scatter plots, box plots, Gantt charts, tables, dashboards);
- additional analytical elements such as colouring, scaling, hierarchies;
- plotting a fourth dimension of time to highlight trends.

From this point onwards this process of visualizing data will add value throughout each stage of building your baseline. For example, visualizing data quality alone (detailed in stage 5) as demonstrated in Figure 3.2.3, helps identify where the gaps are in your data and where there may be inaccuracies. Equally, when crowdsourcing data, handing visuals back to data owners helps them identify outliers and inaccuracies in seconds (stage 6).

5 Create rules and clean the metadata centrally

Inevitably when you bring data together there will be data quality issues. So, as soon as you can, create clear rules on data structure. For example, on one project a client had a database of more than 30,000 items of IT hardware spend. The vast majority of the items were duplicates and the data could be reduced a couple of hundred items; that accounted for well over 95 per cent of the spend. The company had made its life much harder by not having clear rules about how data was to be collected and stored.

Cleaning data is not a one-off exercise. Given that people are collecting and cleaning data in different offices or even different countries, once your data is clean you need to make rules about data collection and storage. For example, often the lowest level of data is acceptable, but it is how the data is categorized or tagged that is important. A common and easy category to grasp is gender. The male sex may appear as Male, male, m, 0 or 1, MALE... and that is just in English. Consistency is key. You have to be clear about definitions; for example, what is an FTE, or total cost of utilization? If you decide this upfront, you will save serious amounts of time later on.

Data quality can be used as a reason for not generating information. In my experience, if data is being used and presented in an understandable way, they will drive behaviour and the data quality will improve naturally. Visualizing your data will help clean data naturally for two reasons. First, it is much easier to identify inaccuracies and outliers. For example, the visualization in Figure 3.2.3 provides an overview of all the metadata and the distribution of data for each property. Second, it highlights the completeness and consistency of particular data fields (dimensions and measures) and where there are orphans in the organizational hierarchy, so you can immediately identify problem areas and where attention is needed (see Chapter 1.3 for definitions of hierarchical data).

Visualizing data in this way makes it easy to obtain the data you need from across the organization. Alongside new and intuitive connections between data and visualizations, this means you no longer have rely on database administrators or IT to solve data issues. This is because traditional data visualization layers in software are becoming increasingly intertwined with database layers. Visualizations are no longer just a representation of data, they also link back to the data they represent. The consequence is that where previously you had to change data within a database yourself, or via a database engineer to alter the visualization, now you can alter data directly through the visualizations. You can drag and drop pieces of data to alter numbers of fields at once, making the process of cleaning data in line with

FIGURE 3.2.3 Example data types and patterns dashboard

your rules on data structure on an ongoing and ad hoc basis infinitely easier, saving time and effort in the cleaning process.

6 Crowd source

The data you need to further improve quality or for new properties will sit in different places of the organization. Often those who are in the best position to fix data is the ones who want the maximum insight from it but do not have the time to correct it. Therefore, make it easy for people to be able to provide and alter data. Technology now allows for fast feedback, such as surveying capability, which automatically updates a person's records once completed. For example, a problem in many organizations is a simple question of organizational hierarchy. Who reports to whom? One particularly useful exercise may be to send a two-question survey to all employees asking 'Who do you report to?' and 'Who reports to you?' Collate the data, draw an org chart and very quickly you will expose any confusion or inaccuracies. Big data is all very glamorous, but the most value is often in getting the basics right. Here are three tips to help you when crowdsourcing data:

1. **Make data easy to upload.** If the upload process is complex people will become disengaged and turn to other work instead. For example, sending out a webform that takes two minutes to complete will give you instant updates on your data.
2. **Report data back to owners**. Those who know the data can interact with the summary findings will see the value returning from their data and they will be able to easily identify further outliers that represent data quality issues.
3. **Make the process mobile.** Often the best time to spend a couple of minutes uploading or analysing data is when you are in transit or a little bored. If the data is available at all times and on all devices such as smartphones and tablets, it makes it easy to review charts and add insight.

To begin with, collecting data from different areas of the organization will come down to you and your team. However, the sign that you are getting this right is when the data is seen as a business issue and function leaders come to you to drive new insights and get further data.

7 Stop analysing and start playing with your data!

As the HR Information Systems Manager at one retail client once said: 'One of the most exciting things about bringing all types of organizational data together, and visualizing is that you can start to answer questions that people didn't realize they wanted the answer to.' Having built your baseline, just enjoy yourself. Start exploring, slicing and dicing the data. (The slice is the ability to split the data according to any dimension or measure. The dice is the ability to use any aggregator such as sum, standard deviation, percentile, skewness, and many more.)

When analysing your data for the first time you will be amazed what you can find. One consultant I worked with, after quickly combining two client datasets and analysing the training impact, was able to tell her client that it was wasting half of its graduate sales training budget because one group of the target audience learned nothing. By contrast, for the other group it was a great investment. The client just needed to know which graduates to target for what type of training. Simply by splitting the trainees by geographic location showed that particular office sales teams didn't increase their sales post teams, while others almost doubled theirs. The underlying reasons and the action taken as a result are a separate story, but what is powerful is the insight gained by combining different datasets, in this case sales and training data.

The combination of data is frequently referred to as mash-up data. Bring together your people data with data from across the organization and see what you can find. To make sense of all the data, break down areas of insight into topics by organization or theme. For example, functional areas: HR; finance; procurement; manufacturing; or organisational areas: people, process, systems; or strategic, operational, transactional. Using these categories you can start performing advanced analytics. For instance, can you see whether sales performance links to education or recruitment channels? Do those who attend training have lower attrition? Which actions post an employee engagement review had meaningful impacts? What human factors drive profit, productivity or retention?

We have now got to into the in-depth analysis. But when analysing and drawing your insights you have to be careful. What might seem obvious in the data may be wrong statistically speaking. It is too easy to oversimplify and make errors of judgement. The next section sets out three checks to ensure you analyse your data as effectively as possible.

Performing analysis: common statistical traps

All through these practical steps you will have been building your baseline understanding. However, I have included this section because data can often be misrepresentative and insights are only as good as those who draw them. The world is full of wrong and dangerous analysis and statistics. Numbers are aggregated and conclusions erroneously drawn. Linear correlations are calculated with high correlation coefficients (R) at the wrong level and bad science is peddled as robust analysis. There are many PhDs that can be written on bad analysis. Given the scope of this book, I am going to focus on the three logical statistical flaws that I see most frequently. This doesn't mean you should stop the learning here. More, I hope, that you gain an interest to want to learn more and avoid making recommendations that give the illusion of being robust but are not. The three I am going to cover include:

1. The ecological fallacy
2. Taking correlation for causation
3. Ignoring statistical significance

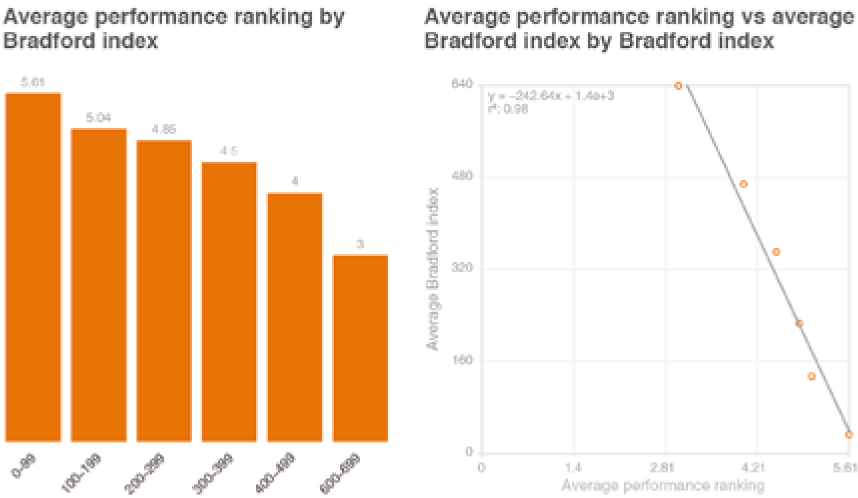
The ecological fallacy

In the 2004 US election George W Bush won the 15 poorest states and John Kerry won 9 of the 11 wealthiest. The conclusion: wealthy people vote Democrat and the poor Republican. And yet 62 per cent of voters with annual incomes over $200,000 voted for Bush, while only 36 per cent of voters with annual incomes of $15,000 or less voted for Bush.[11] So that conclusion is completely inaccurate.

This is a great example of the ecological fallacy in practice. 'An ecology fallacy is a logical fallacy in the interpretation of statistical data where inferences about the nature of individuals are deduced from inference for the group to which those individuals belong.'[12]

In the context of HR, an example of the ecological fallacy in practice is the likely hypothesis of a correlation between performance and absenteeism as measured through the Bradford index. The Bradford index weights the number of absence instances higher than the total duration of absence. The theory is that short, frequent and unplanned absences are more disruptive than longer absences. The formula is $B = S^2 * D$ where S is the number of spells (instances) and D is the duration. For example, if someone was sick once for 10 days, their Bradford score would be $1^2 * 10 = 10$. But if they were sick 10 times a day at a time, giving the same total of 10 days, their score would be $10^2 * 10 = 1,000$. For a given group of employees, the histogram (in Figure 3.2.4) clearly shows that those with a Bradford score below 100 have an average performance of 5.6 out of 10, while those between 600 and 700 have an average of only 3.

FIGURE 3.2.4 Histogram and aggregate scatter with linear regressions of average performance rankings by average Bradford index category

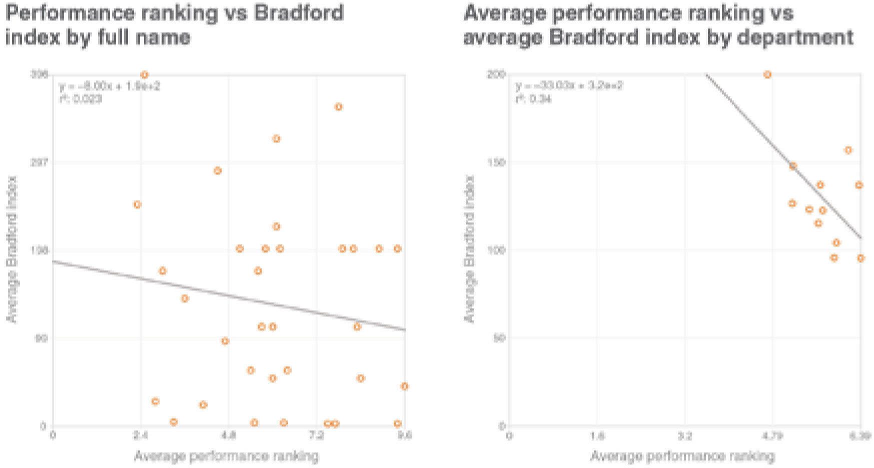


This is proven even more emphatically through the 'scientific-looking' correlation in Figure 3.2.4. The figure on the left shows the average performance for each Bradford index group while that on the right takes each of those six groups and draws a correlation between each group's performance and their average Bradford score. The same groups of employees (those with Bradford factors less than 100, between 100 and 200, and so on) are correlated with their performance ratings. Unsurprisingly, the correlation (r) is –0.97 or almost perfect (a score of r = 1 or –1 is perfectly correlated while a score of r = 0 means there is zero correlation). *Wow*, think the people reviewing these numbers. Conclusions are drawn: 'We need to sort out the poor-performing employees.' 'It's so logical that it's these people who aren't performing!' The analyst gets a pat on the back and now the management team thinks it has the proof and burning platform to take action against these low performers.

However, in this instance, they are all wrong. In fact, they are terribly wrong. They have just fallen for the ecological fallacy. In the above example, if you dig just a fraction deeper, the distribution of the 46 employees included within the analysis is shown in Figure 3.2.5.

FIGURE 3.2.5 Pivot table of performance rankings by Bradford index category

FIGURE 3.2.6 Departmental and individual scatter plots of performance rankings by Bradford index category



You can see, straightaway, that there are very few (four) data points of those with a Bradford score above 300. You can also see a large range of performance rankings of those with a Bradford score below 100. In fact, if you draw the scatter chart at the individual rather than group level, you can instantly see there is, in fact, no correlation (it is just 0.047).

There are only two employees with high Bradford scores and poor performance scores. This is a good example of what is officially called 'outlier analysis'. So no drastic or group action is necessary. Instead, a quick investigation into their historical performance to find out if they have always been poor performers would be advisable (if possible). Then a conversation can happen either to find out if something or someone is affecting their performance, or how the manager can help bring them back on track. This is obviously a tailored example, but it is an excellent reminder of the value of digging into statistics, questioning what you see and ensuring that you don't make sweeping conclusions which could have a negative impact.

Correlation versus causation

Did you know that it is dangerous to eat ice-cream and then go swimming? There is a correlation between ice-cream consumption and the likelihood of shark attacks. Sharks like people who have just eaten ice-cream, right? Clearly, no. So how can they be correlated? The real correlation is that the hotter it is, the more ice-creams people eat, the more they swim and the more likely it is that there are going to be attacked by sharks. This may be obvious, but every day people jump to overly simplistic and inaccurate conclusions, confusing correlation with causation.

A correlation is just a number. It happens to calculate the strength of a linear relationship between variables, but it does not carry any information about causation. Why? Because causation is far more complicated than the idea of how nicely it can fit a line to data. In general, it is very difficult to show causation. Why? Because we need to rule out all other factors that could influence the relationship. In general, showing causation is only possible in scenarios such as a laboratory environment where external factors can be controlled or at least limited and explored.

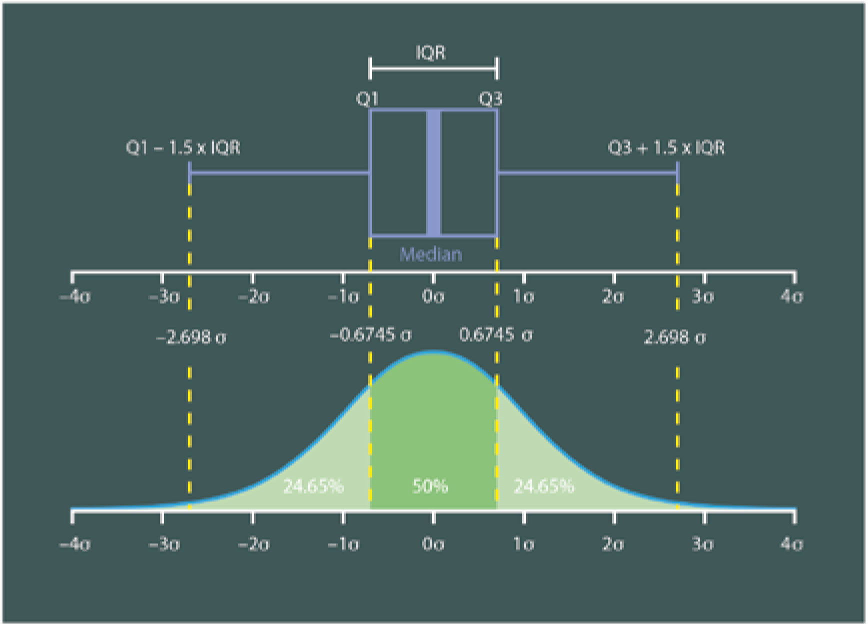
In a business context, we can show correlations are high and use these effectively for prediction. But it is important to ask why this correlation is occurring in order to examine causation. For instance, suppose that performance ratings were inversely correlated to number of absences. However, the driver of the relationship is unclear. That is, we cannot determine whether performance ratings were assigned as a penalty for absences or whether missed days actually contributed to lower performance. The causation is unclear but the correlation is still useful to drive conversation, investigate the potential drivers before acting and run relevant analysis for future conclusions to be drawn.

Statistical significance

If we are going to draw conclusions about a set of data, then we should have some understanding of how strongly we can draw those conclusions. I am going to use a simplified example to demonstrate my point. The example below shows that for a set of employees, males are higher performers than females. So does that mean that we should only hire men? The problem here is that the stats can get quite complex, so here are a couple of simple tools and thoughts. The first is to look at the variation within the dataset. The greater the standard deviation, the less likely you can draw the conclusion. Why? Standard deviation measures the average of how spread-out the data is. You can think about it this way: if female employees all have really similar performance measures, then it is possible to make better generalizations about that group. More similarity means small standard deviation. If, instead, some are performing really well, some are performing really poorly and some are doing okay, it is harder to tell a story about what is happening to the whole group. This is a situation with a large standard deviation – lots of differences, a more spread-out group.

A good way to visualize this is through the box plot diagram, which is described in Figure 3.2.7. The boxplot shows the distribution of the data.

FIGURE 3.2.7 Box plot explanation

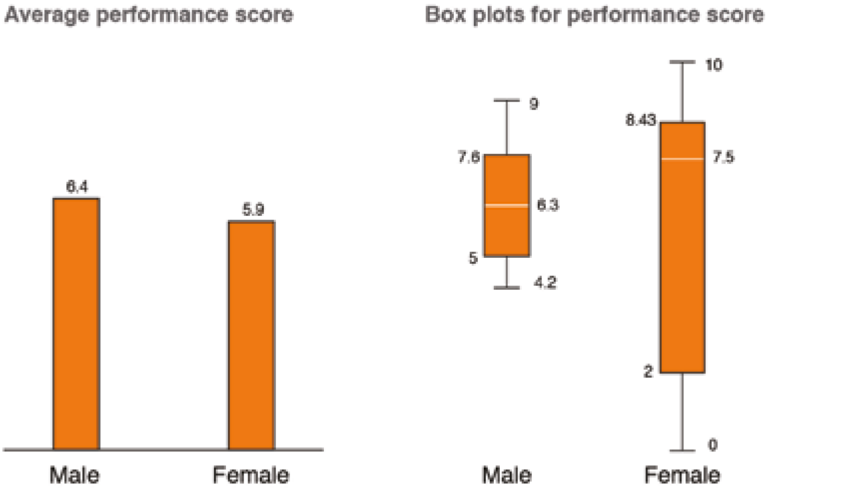


For the median, there is a thick line and a box is drawn to show the majority of the data points. This could be the upper and lower deciles (where 90 per cent of the points are above it and 10 per cent of the points below) or quartiles (the same, but 25 per cent and 75 per cent). A further set of lines shows the upper and lower limits of the data. This could be the max and min, or the 99 per cent and 1 per cent points (the percentiles). The box plot gives you a visual sense of distribution, which helps to indicate where there is more than just a simple story in the data.

The risk of using bar charts and the power of the box plot is highlighted in Figure 3.2.8. The left-hand bar chart shows the performance scores of males and females. At face value it appears males are performing far better, with an average of 6.4 out of 10 compared with the female average of 5.9. But when the same data is plotted using the box plot, another picture emerges. The median performance for females is actually 7.5. There is far more variability within the female population, with the lowest scores of '0' being 'achieved' and therefore taking the average down.

The next thing to look at when thinking about statistical significance is the sample size. The smaller the sample size, the greater the danger. This also makes sense. If you want to make a claim about the performance of women and you only measure the performance of 10 women, how confident are you

FIGURE 3.2.8 Average and box plot of gender performance levels



that they reflect *all women* the company would ever be interested in hiring? If you get performance measures from more women, you have a clearer idea of what is happening on average. How can we make a statistical decision about whether men perform better on average? We turn to the most common statistical significance test, which is called the *t*-test. We can plug all of these numbers into the formula for a *t*-test comparing two averages. The result from such a test is a single percentage called a *p*-value.[13] (It is not in this book's scope to outline the *t*-test. For further reading on this please see the suggestions below.) In our scenario, this value is 15 per cent. Great – so how do I get any information about my employees' performance based on this number? In this scenario, the *p*-value measures the likelihood that these differences in employee performance arise completely by chance rather than due to any real average performance differences. It is standard practice to conclude that a *p*-value less than 5 per cent indicates a 'significant' difference. Here, with our value of 15 per cent we cannot make this claim. Are men performing better than women on average? Yes. But are they performing *significantly* better on average? No. So should we hire only men to boost performance scores? Not according to the data.

If you are told that a value is 'statistically significant', you should ask: 'Statistically significant compared with what?' and 'According to which statistical test?' Statistical significance is used within a section of statistical analysis called hypothesis testing and is used to infer how likely it is that we observe data by chance. Instead of using the phrase 'statistical significance', you will save a lot of energy and confusion by just explaining in words the conclusion your data leads you to. For example, it isn't (statistically) possible to conclude that men are performing better than women. It is also good practice to include details about the statistical test used to reach your conclusion and the number of data points included in the calculation.

While each of these three statistical traps is relevant, the point to take away is that data analysis isn't as much of a magic wand as many people think. Too often, people work extremely hard to build a baseline of data and leave the 'insight phase' of the work to the last minute. They then jump to erroneous conclusions because they don't think hard enough and make too many predictable mistakes, such as the three outlined above. Unfortunately, from what I have observed from countless business presentations, most of the business world is full of terrible statistics. The bar chart is probably the most common stats tool used, yet as shown above it hides a whole range of flaws. The aim of this section isn't to make you an expert, but to improve your awareness of the dangers. If you are interested in learning more there are a couple of books I suggest you look at: *Essentials of Social Statistics for a Diverse Society*[14] and *Statistics for People Who (Think They) Hate Statistics*.[15]

Final thoughts

Data should be used to turn hindsight into insight and then foresight. While building a looking-back baseline as the logical first step in the micro stage of design, it is that dataset that you will come back to the most. In fact, it never stops being important, either through the design or afterwards. By maintaining the dataset and mashing it with other data, greater insight will be generated. You will use this data to generate scenarios and plans. With enough history, other projections will become possible and you move into helping to shape the future. The key is to make sure that you set yourself up for success by collecting, cleaning, analysing and reporting on your data consistently.

Organizational analytics have an opportunity to deal with the world as it really is: a complex, interconnected system. Consistent and focused data collection must be backed by a deep understanding of the multiple connections between people, roles, skills and activities. Too many change programmes have been overwhelmed by the sheer complexities of the organization. As you continue to build your baseline you can add more data from across the organization to really see which areas are performing well, need improvement or a downright overhaul. This is not just about a one-off design. In the future, those functions dealing with organizational data and analytics have to be able to resolve business issues before they happen by connecting the organizational system. They need to drive leaders from across the whole business to think through what their people do now, what they will do in the future and what the cost impact will be if activities change. With this, your data will become a goldmine: an invaluable resource with exponential and sometimes unpredictable benefits.

Remember this

1. To get the most from your data you need to recognize potential value and communicate that value to the business.
2. Be disciplined throughout the data process. Be clear about the data you need and who or where you need to get them.
3. Make the process of collecting data valuable to everyone involved, from business owners to the employees working on the ground with the data.
4. Use the right tool for the right job. Do not try to do everything in Excel, and where possible use specialist tools to help with particular tasks. It is worth it!
5. When analysing your data, be wary of jumping too quickly to conclusions. Use analysis as guidance for further questions. It is not always a final answer and be mindful of the many hidden statistical traps.

Notes

1 [!Reference to come!]

2 Diaz, J [accessed 19 February 2015] Sploid, [Online] http://sploid.gizmodo.com/if-you-fold-a-paper-in-half-103-times-it-will-be-as-thi-1607632639

3 Stratified Medicine in the NHS; An assessment of the current landscape and implementation challenges for non-cancer applications (2014), Association of the British Pharmaceutical Industry

4 Official Google Blog [accessed 19 February 2015] We knew the web was big, Google, [Online] http://googleblog.blogspot.co.uk/2008/07/we-knew-web-was-big.html

5 Sedghi, A [accessed 28 March 2015] Facebook: 10 years of social networking in numbers, Guardian [Online] http://www.theguardian.com/news/datablog/2014/feb/04/facebook-in-numbers-statistics

6 As presented by Evensen, K and Aaserød, J, presentation to HR Recruitment Days, HR Norge, March 2015

7 Press, G [accessed 15 January 2015] Gartner Predicts Top 2015 and Beyond Trends for Technology, IT Organizations, and Consumers, Forbes [Online] http://www.forbes.com/sites/gilpress/2014/10/09/gartner-predicts-top-trends-for-technology-it-organizations-and-consumers-for-2015-and-beyond/3/

8 Koetsier, J [accessed 21 November 2014] Marketo CMO: 'Marketing has changed more in 5 years than the past 500', Venture Beat, [Online] http://venturebeat.com/2014/03/01/marketo-cmo-marketing-has-changed-more-in-5-years-than-the-past-500-interview/

9 Gartner official website [accessed 19 February 2015] Gartner [Online] http://my.gartner.com/portal/server.pt?open=512&objID=202&mode=2&PageID=5553&resId=1871515&ref=Webinar-Calendar

10 Kristi, M (204) Support the Data Enthusiast: Challenges for Next-Generation Data-Analysis Systems, Proceedings of the VLDB Endowment, **7** (6) pp 453–56

11 Gelman, A (2008) *Red State, Blue State, Rich State, Poor State*, Princeton University Press

12 Robinson, W (1950) Ecological Correlations and the Behavior of Individuals, *American Sociological Review*, **15** (3), pp 351–57

13

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \qquad s = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

For our example:

Variable	Men	Women
Mean	$\bar{X}_1 = 5.43$	$\bar{X}_2 = 4.72$
Standard deviation	$s_1 = 2.47$	$s_2 = 1.66$
Sample size	$n_1 = 29$	$n_2 = 17$

14 [!Reference to come!]

15 [!Reference to come!]

This is an extract from the book Data-driven Organization Design
published by Kogan Page Limited in 2015

"This book puts into words what I've been trying to explain for years. Managers should read this book to become more data-driven; data-driven people should read it to understand the imperatives of management."
Stéphane Hamel, Digital Analytics Thought Leader, Immeria

"Morrison's dashing and enthusiastic style persuasively makes the case for using good data analytics as the basis for successful organization design. This book is now on my 'recommend' list."
Dr Naomi Stanford, organization design consultant

"Provides insightful connection between strategy and the enabling organizational design. It includes an inspiring collection of theory, tools and experience to drive change and transformation."
Pär Åström, Senior Vice President, Business Development, Husqvarna Group

"This book nails it. Honest and practical, it shows you how to deeply analyse and design an organization – and implement it."
Nathan Adams, HR Director, Aviva

Data is changing the nature of competition. Making sense of it is tough; taking advantage of it is tougher. There is an important business opportunity for organizations to use data and analytics to transform performance. Organizations are a constantly evolving system made up of objectives, processes and people, and all of it organized in a governance structure. It is dynamic, and constantly changing. Using data and analytics you can connect all the elements of the system to design an environment for people to perform in – one that has the right people, in the right place, doing the right things, at the right time. Only when everyone performs to their potential will your organization have a hope of sustaining a competitive edge. Data-driven Organization Design provides a practical framework for HR and organization design practitioners: build a baseline of data, set objectives, carry out fixed and dynamic process design, map competencies and rightsize the organization. It shows how to collect the right data, present it meaningfully and ask the right questions of it. Whether you're planning a long-term transformation, a large redesign or a one-off small project, Data-driven Organization Design will show you how to make the most of your organizational data and analytics to drive business performance.