Data Analytics for Organisational Development

Unleashing the Potential of Your Data

UWE H KAUFMANN • AMY BC TAN

WILEY

Preface

Do you like statistics?

If you feel like affirming this question, you belong to a minority, in our opinion. In this case, you will hopefully recognise many familiar tools and methods in this book. You may discover some new insights into their application and our point of view from the perspective of organisational development.

If you feel like answering our question in the negative, you may not want to put away our book too fast. We know how you feel, because we had been in a similar situation like you, before we had to learn to deal with data as part of our work. Our job was then and still is organisational development in very different organisations with manifold cultural background.

Here is more about that:

After my study of engineering and passing all exams with nice results, I had put away statistics as a part of my life that was over and done with. Even though I always feel quite comfortable when it comes to dealing with numbers, the importance and usefulness of statistics has not been too clear to me during my study. While I could easily make use of Gauss' normal distribution in control cards for the benefit of customers of a production line in a modern German engineering plant, only after joining General Electric (GE), I was taught, explained, and coached how an ANOVA could add value to a typical banking environment.

At GE, data analytics has been used for improving the organisation, better serving customers and employees, and hence increasing efficiency and profitability. General statements like "With this new method we will be able to increase our profitability," were no longer accepted. "How much increase of productivity can we safely assume? What is the risk for this investment going to waste?" were questions that could only be replied with data-based, statistically proven answers.

During my hard and application-oriented education at GE, I learned that the use of data without the support of the appropriate statistical tools is usually inadequate.

With the case studies shown and explained in this book, I wish to pass on this perspective.

Uwe H Kaufmann

I have been in Human Resources the longest time of my professional life. At first as the director HR for an insurance company, then as director people matters of a ministry and lastly in the same position with the Organising Committee of the first Youth Olympic Games. I grew up with the understanding that data analytics for people in HR means calculating averages and percentages.

Only in 2002, when I was tasked by my CEO to take up the role of a Black Belt for some insurance-related and some HR projects, I got in touch with serious data analytics. To be honest, at the beginning I was not really excited to take up this role. Only after a very tough training by Uwe H Kaufmann and his persistent coaching, I was able to appreciate that the new methods and tools would not only benefit my HR function but also improve processes in many other functions of our organisation.

And of course, an HR director who can roll up her sleeves to study activities and increase efficiency in any part of the organisation does really gain acceptance throughout the firm.

Our organisation development activities have been and still are data-driven and supported with statistical methods ever since.

Therefore, I have really benefitted from the training to become a Black Belt - a Lean Six Sigma Black Belt is a first-class data scientist - and changed my perspective on and approach to my work within and outside HR.

Organisational development without data analytics will not work.

Amy BC Tan

Together we have written this book because of our career and its close connection between organisational development and data science. This comprehensive perspective is the key to your success in your organisation, as it was and still is for us.

You as leader or manager or supervisor in any function have the non-delegable side-job - often even the main job - as organisational developer, because you have responsibility for business results, for resource utilisation, for customer satisfaction and for engaging your staff. Nowadays, you have more data accessible at your fingertips about all activities and matters of daily business life than ever before. And you have enormously powerful hard- and software available for acquiring, converting, cleaning, and analysing this data to transform it into information beneficial for your organisation.

Use this opportunity!

If you acquire the necessary knowledge and skills, it will not only give your organisation an edge in the fast-changing market but also yourself. With our cases it should be possible for you to follow through typical applications of data analytics in many different organisational development situations and to translate the learned steps into your own environment.

We wish you all the best on the journey to unleash the potential in your data and to be a step ahead of the competition.

Amy BC Tan Uwe H Kaufmann

Introduction

1 Why Data Analytics is Important

"In God we trust, all others bring data."
W Edwards Deming

Everyone remembers the troublesome process of raising a request with IT or the vendor for a small data analysis task and waiting days or even weeks before getting the result. More often than not, the result was not presented in the most useful way or has just raised a follow-up question which required new data to answer, new requests for IT or the vendor. For decades, managers have relied on this kind of process because they had no choice. This process has a fundamental flaw: if you want to make timely decisions, lagging and outdated information cannot be used. Hence, timely decisions had to be done without having the foundation of real-time data and sometimes based on gut feeling.

Time has changed. The amount of available data in all functions of any organisation is growing daily. And the access to this data gets easier and easier. Nearly everyone can acquire the data necessary to run his own analyses. And almost everyone has a formidable computer with powerful analytics tools right on his desk. The question now is, how to turn the data into business relevant information for making the right decisions when needed.

Hence, it is about time to ensure that the right data is collected in an appropriate way, screened, transformed and analysed using valid methods in a manner that delivers business relevant information which is turned into intelligence preparing appropriate decisions towards business success.

Data analytics**{** XE "Data analytics" **}** is the process of collecting, processing and analysing data with the objective of discovering useful information, suggesting conclusions, and supporting problem solving as well as decision making (Wikipedia, Wikipedia, n.d.).

"Data Analytics is a business practice
every Manager should be familiar with."

Data analytics encompasses the main components **Descriptive analytics** (post-mortem analysis), **Predictive analytics** and **Prescriptive analytics**.

Big Data describes sets of data that are so voluminous and complex that traditional data processing applications cannot handle them. Big Data is defined by its three Vs, Volume, Velocity, Variety (Russom, 2011). At the beginning of the 2000s, big data posed a serious problem to many organisations. On the one hand, the volume of available data went up exponentially. On the other hand, CPU speed and storage capacity could not keep up with the data amount at hand. At that time, handling big data was reserved to a few companies and organisations who were relying on the analysis of data to stay in business.

Nowadays however, computers with huge data storage and handling capacity are available to nearly any organisation, be it by installing hardware and software inhouse

or be it by renting external capacity. Two trends seem to be the result of this change in the IT environment. Firstly, more and more organisations have the means and see the need to collect data about their operational environment. Secondly, these organisations widen the scope of their data analytics activities.

Whereas some researchers used to suggest that data analytics is mainly describing the handling of user data that is produced by CRM and similar systems and turned into customer intelligence, the scope of data analytics opens up to include all functions of an organisation.

Not only is there a move from so-called Big Data analytics to analytics of any kind of data, there is also a healthy trend towards involving all levels of management and even staff into this not so new field of information management. Progressive managers are familiar with data available and with trends, shifts or other patterns in their data and use them for decision making.

The former specialty data analytics is gaining popularity amongst all managers of an organisation. Hence, it is about time to ensure that the right data is collected in an appropriate way, screened, transformed and analysed using valid methods in a manner that delivers business relevant information which is turned into intelligence that prepares appropriate decisions towards business success.

"The ability to take data - to be able to understand it, to process it, to extract value from it, to visualize it, to communicate it - that is going to be a hugely important skill in the next decades."
Google's Chief Economist Dr Hal R Varian in 2009

2 Why This Book Has Been Written

To be very honest, I did not like statistics very much. After I covered statistics during my engineering study, I made sure I pass all tests and exams and phew… Never again.

Only after joining General Electric (GE) in the nineties, I had to relearn the statistics. At GE, acquiring knowledge in statistics was purpose-driven, i.e. it was real business problems at hand I had to solve using numbers. Consequentially, I started to grow some degree of interest for the math and the stats.

"Having wrong data is worse than having no data at all."

As the following examples testify, having numbers is good but not enough. In addition, we need to ensure the data is properly collected, cleaned and analysed before making any decision.

> Some years ago, the director of a blood bank came back from a meeting with other blood bank heads. She was not happy because she had the chance to compare certain blood bank performance indicators with others and recognised that her own blood bank obviously was wasting significantly more blood products than some other blood banks. She was talking about bags with platelets that were taken from blood donors, tested and then disposed

> because they did not meet the quality standards. By her criteria, this kind of situation was not easily accepted.
>
> A team was set up to investigate the root causes for the wastage in the most precious blood products. After data collection and some basic analysis, it became clear that the blood products were not of lower quality than in other countries. The root cause was in the process of evaluating the quality of blood bags – the data collection.
>
> This case example is explained later in this book.

This case alone generated some life-long learnings we wish to turn into some recommendations:

Firstly, **do not trust numbers blindly**. Even numbers that are spat out by a computer can be wrong, biased or otherwise made useless. Check how these numbers get into the computer in the first place.

Secondly, before you perform any data analytics, ensure that the **data is collected following proper procedure**. Therefore, this book does not start at data analysis. But it starts where the collection of the data is thought about and designed.

Thirdly, like the head of the blood bank had her very **powerful business case**, confirm that your data analytics serves a purpose, a need the people whom you work for with your data analytics know, understand and share. Only with this need, this business case, your data analytics case is more than playing with numbers.

In the following chapters, we are going to elaborate the use of Data Analytics to solve business problems, to make critical decisions and to drive the organisational strategy. And, we will show some typical pitfalls and remedies on the way to data analytics.

At the moment, there are numerous Data Analytics courses available in the market. Interestingly, many of them are titled around Data Analytics for HR Professionals or for Customer Relationship Management (CRM). This book takes a wider scope and shows the application of data analytics in any organisational situation, where the proper use of data is critical. Therefore, we call it "Unleashing the Potential of Data – Using Data Analytics for Organisational Development".

This book is written with the intention to close a well-known gap mentioned again by Amy Gallo (Gallo, 2018). Every manager should know four powerful analytics concepts in order to be informed about his organisation and to make data-based decisions. These concepts are in no way new. However, they gain more importance with the increased amount of data available and the apparent need – and the chance – to turn this data into business relevant information. This is supported by the availability of a multitude of easy to handle tools for data analysis and data visualisation.

These tools can only be used by managers if these managers understand the basics of data analytics from data acquisition to data analysis. Therefore, so Gallo, managers need to know the most basic concepts.

These concepts are randomised controlled experiments, hypothesis testing, regression analysis and statistical significance.

Randomised controlled experiments include data collection techniques such as any kind of surveys, pilot studies, field experiments and lab research. Instead of outsourcing such services to specialists and relying on them analysing the result and developing recommendations, it could be beneficial to understand data and analysis process. This knowledge would certainly help to draw customised conclusions for the organisation; conclusions an outsider cannot easily draw. Experiments also comprise testing new routines or products on their performance. Experimenting with processes is a powerful way of improving the output whilst watching other important indicators at the same time in a controlled way.

Hypothesis testing{ XE "hypothesis testing" **}** contains statistical tools that compare stratified business relevant data and answer the question for the "better one" including the inherent risk of this decision to be wrong. Hypothesis tests find their application in all units of any organisation. Analysing survey results uses hypothesis tests to answer questions like "Is there a difference between last year's and this year's rating?" or "Did department A perform better than Department B". The result of a hypothesis test can be much more than just "Yes" or "No" to such questions. Hypothesis tests always give a risk that comes with making a decision; a risk for making a wrong decision. Many hypothesis tests even give an indication on what the minimum difference or minimum improvement is, which leads to much better decisions on the impact of a change or improvement. "What is the minimum improvement if we buy our supplies from Supplier B compared to Supplier A?" can be answered with hypothesis tests.

Regression analysis comprises statistical tools that are used for similar tasks as hypothesis tests. Whilst hypothesis testing usually answers questions about the relationship between two variables, regression models may include a large number of variables at the same time. With this, the interaction between multiple drivers (independent variables) for the same result (dependent variable) can be analysed which is less effective with hypothesis tests. Hence, regression models help explaining complex relationships between many variables at the same time. Additionally, these tools are often applied in predictive statistics, i.e. to use existing data for forecasting the behaviour of machines, devices, organisational units and even workforce.

The aforementioned groups of methods are based on one important concept: **statistical significance**. This often-misunderstood concept is the backbone of all statistics, the backbone of all data analytics. Statistical significance informs about the risk one is to take when making a business decision based on data analytics.

In statistics, "never" and "always" do not exist. "0% probability" and "100% probability" are usually not the results of randomised, controlled experiments, hypothesis test or regression. Most likely, the result of an analysis lies somewhere in between. Then, it is up to the manager to make a smart, informed and data-based choice. Understanding the concept of significance is key on the way to a quality decision.

3 How This Book Is Structured

As this book is about application of data analytics for organisational design, the cases discussed later cover different data analytics situations in any domain of the value chain of an organisation (Figure 3-1).

Under the **Customer** domain, we cover collection, processing and analysis of customer related data. This includes survey data from different customer environments and data measuring the "moment of truth", the moment when the customer experiences the product or service offered.

The **Operations** domain includes gathering data from many different operational environments and turning that into critical information for decision making.

The **Workforce** domain offers ideas for handling HR related data that are used to draw conclusions about different workforce related aspects, be it recruitment, turn-over, engagement or workforce planning amongst others.

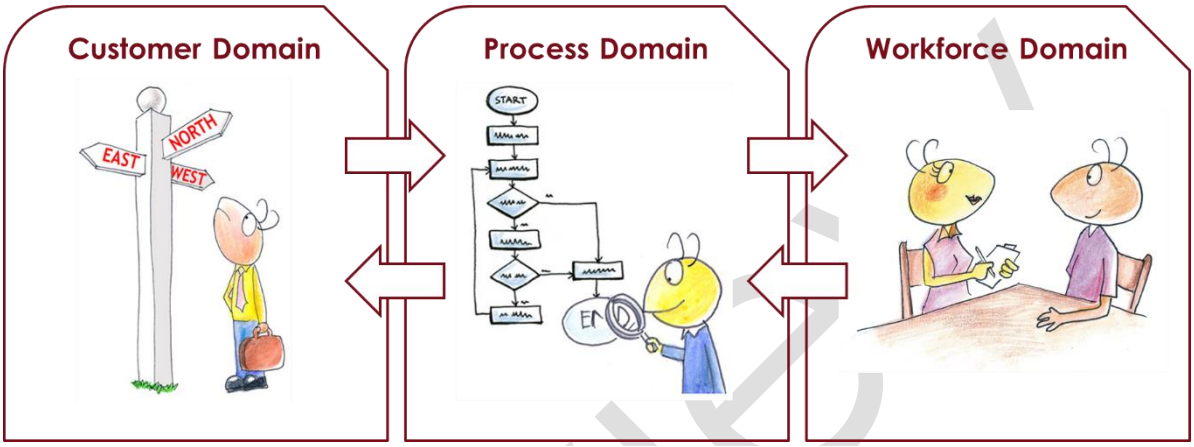


Figure 3-1: Domains of an Organisations Value Chain

For each case, we follow through all steps from the questions, hypotheses or business cases over all stages of data analytics to get to the right decision. The steps mentioned in the following chapters are (Figure 3-2) Formulating a Business Question, Performing Data Acquisition, Conducting Data Preparation, Executing Data Analysis and Drawing a Business Decision.

Figure 3-2: Steps of a Data Analytics Case

Business Question

In this first step, this business-related issue has to be clearly identified. And, it has to be translated into an indicator, a KPI that makes the issue measurable. Better yet, this indicator is on the scorecard or dashboard of management members, i.e. it is important to someone.

Data Acquisition

There is a multitude of ways to collect data to answer the Business Question. It is usually necessary to validate the method of data collection to ensure useful data for analysis, i.e. data that is representative, reproducible and accurate enough to provide sufficient information for answering the business question. There are statistical tools that help identifying potential problems within the data collection process.

Data Preparation

Even if the method of data collection is proven and the instrument is statistically accepted, it can still be that data is not useful.

In surveys, for example, some survey participants may not give useful input. Part of the reason might be that they were either forced or incentivised to participate in the survey. In general, we can assume they were not interested in it. Hence, they may have provided valid input to a well-established survey questionnaire, but the input may not be useful. Or worse yet, the input could spoil the following analysis steps. Such input could be random rating numbers or the same rating numbers for all questions or statements.

Therefore, data preparation is necessary to find and omit such input to feed only data into the analysis that is really value-added.

Data preparation also includes formatting the data so that it can be used by the preferred analysis software. More often than not, downloaded data from a system is not in the right format to be fed into the analysis software, Excel for example. However, in most cases data can be reorganised, reformatted or transformed so that the software can handle it.

Not always will the analysis software stop working because of the wrongly formatted data. In worst cases, it might just work and spit out wrong results.

Data Analysis

Generally, data analysis is done in a graphical and in a statistical way. Usually, both are necessary to ensure proper conclusions. Additionally, graphical analysis may be needed for visualising data and storytelling.

However, graphical analysis without statistical support may lead to wrong decisions. And similarly so for running statistical analysis without graphical support.

Hence, all data analyses should be done in a two-step approach. Firstly, one or more graphs should be plotted to visualise the data. This visualisation alone may have the power to drive the decision. Secondly, however, it is a good practice to support the graphical analysis with statistics.

Table 3-1: Graphical Analysis Tools for Various Data Type Situations

		X	
		Discrete	**Continuous**
Y	**Discrete**	Bar Charts, Pie Charts,	Probability Plots Reversed Stratified Frequency Plots
	Continuous	Stratified Frequency Plots	Scatter Plots
		Hypothesis Tests	**Regression**

For the analysis of data, a variety of tools is available. The selection of the appropriate tool depends on the business question that needs to be answered, the type of data

collected and their characteristics. The stratifying factor that drives a decision is usually called "X". The resulting outcome is usually called "Y".

When, for example, the rejection rate of a product is compared between suppliers A and B, then month signifies the independent variable X whereas rejection rate denotes the dependent variable Y. Almost all data analytics tasks fit in this structure. The application of tools depends on the data type found in X and Y.

Supplier, for example, is a discrete X and rejection rate is a discrete Y that is generated by counting satisfied customers and non-satisfied customers. Hence, the upper left field in Table 3-2 is applied. Since we have only two categories, Supplier A and Supplier B in X, the appropriate statistical tool would be a 2-Proportion-Test.

This table will be referred to in the following cases to select the applicable graphical and statistical tools.

Table 3-2: Statistical Analysis Tools for Various Data Type Situations

	X	
	Discrete	**Continuous**
Y	Proportion-Tests	Logistic Regression
	Parametric and Non-Parametric Tests for Central Tendency and Variance	Linear and Non-linear Regression
	Hypothesis Tests	**Regression**

Whilst the application of appropriate graphical (Table 3-1) and statistical tools (Table 3-2) will be demonstrated, the tools will not be explained in detail.

Business Decision

Very often, data analysis produces results that are hard to understand by staff who are not trained in data science. A "p-value", for example, is a key output of many statistical tools but is not easily understood by the majority.

However, the translation of an analysis output like "p-value = 0.03" into a result like "The risk of wasting our money by buying from the more expensive Supplier A is only 3%" changes the conversation about data science.

It is no longer the case of relying on the Data Analyst or Data Scientist to make this translation. Management should understand the basics of data science in order to turn data into information and drawing appropriate conclusions.

Every case is based on a real client case. However, to protect our clients, we have taken out names and have amended all data.

4 What Tools Are Used

Our intention was to provide a reference book, learners can follow along step by step. In order to do this, popular software is used. In our work with our clients we realised their requirements towards the software they are exposed to:

1. **Software must be easily available**. Nearly everyone in the world has a version of Microsoft Office on the computer. Integral part of this is MS Excel. MS Excel includes many functions that help conducting most of the data acquisition, data preparation and data analysis tasks described in this book. Some users do not know the add-in "**Analysis ToolPak**" that inserts even more tools into the MS Excel environment.

   MS Power BI extends the MS Office environment with potent and interactive visualisation and business intelligence facilities. MS Power BI offers data warehouse, data preparation and data discovery capabilities for building dynamic collaborative dashboards.

   R is a programming language for statistical computing and graphics. **R Studio** offers a user interface and development environment for R. Both software packages are available for free.

2. **Software must be easy to use**. At least MS Excel is a software nearly everyone has used before. This means analysts can work with a familiar environment, just adding some new tools. Nearly the same applies to MS Power BI. It might be new to many analysts, but then again it has Microsoft's user interface and many functions that are adapted from MS Excel. The learning curve for MS Power BI should be very steep and short.

   R is a programming language and free software environment for statistical computing and graphics. The R language is widely used among statisticians and data miners for developing software for wrangling and analysis of data (Wikipedia, R (programming language), 2020). Learning R (via R Studio) is easier for people with a light programming background. The learning curve for R might be longer for many. But the benefits are excellent. R has a sheer endless collection of readymade functions that grows every day. R can even be integrated to MS Power BI so that special functions and graphs can be produced in R and displayed in the familiar and more presentable MS environment.

3. **Software must be compatible** with other commonly used software. The integration of MS Excel tables and graphs into any MS PowerPoint presentation is as seamless as it can be. There is even the possibility of dynamically linking MS Excel or MS Power BI with the data source on any server or any website and inserting the analysis output into MS PowerPoint. This enables the analyst to have the usual impressive PowerPoint pitch or Power BI dashboard with the actual data, whenever PowerPoint is refreshed.

Each Time we introduced other software like Minitab, SigmaXL, SAS, or SPSS, the limited availability of these software packages was an obstacle to implementing the newly learned tools into the organisation.

Therefore, if you want to be successful in your change effort, ensure you consider the above-mentioned points. If we were business owners or managers responsible for the profit and loss, we would consider carefully, whether we need to buy a number of licenses with an annual license fee of a new software if MS Excel and R or Python can do the job and are available for free.

Hence, the cases in this book are showing analyses done using MS Excel, MS Power BI and R Studio. In order to follow through, you need to activate an MS Excel Add-In and install the other software. Here are the step-by-step instructions:

Activating and Using MS Excel's Analysis ToolPak

Many MS Excel users are not aware of the tools loaded into this awfully familiar office package. Not only has MS Excel functions for nearly every possible data manipulation and analysis task build in. It also comes with an Analysis tool pack that is hardly used.

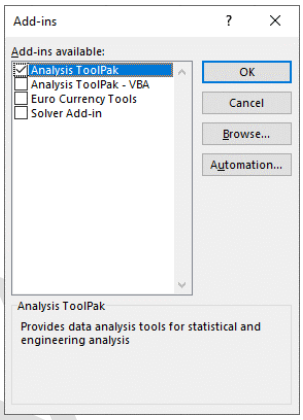


Figure 4-1: Activating MS Excel's Analysis ToolPak

And, it just needs to be activated to make it show up as a collection of macros that have the potential of making your analysis work much easier.

After loading MS Excel, press File - Options - Add-ins - Go and check the Analysis ToolPak checkbox (Figure 4-1). This is all you need to do in order to add a collection of commonly needed analysis tools to your Excel (Table 4-1). These tools can be found under Data - Data Analysis (Figure 4-2).

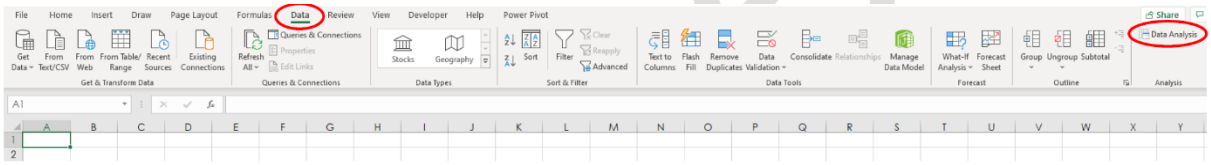


Figure 4-2: Analysis ToolPak in MS Excel

After activating the Analysis ToolPak, the following functions are available:

Table 4-1: Tools Available in MS Excel Analysis ToolPak

- Anova
- Correlation
- Covariance
- Descriptive Statistics
- Random Number Generation
- Fourier Analysis
- Histogram
- Moving Average
- Exponential Smoothing
- F-Test Two-Sample for Variances
- Rank and Percentile
- Regression
- Sampling
- t-Test
- z-Test

Let us get more familiar with the newly discovered set of tools.

Task 4-1: Generate random data

1. Open a new Excel Sheet.
2. Name column A Group and column B Data.
3. Select Data - Data Analysis - Random Number Generation.

4. Select 1 for Number of Variables, 1000 for Number of Random Numbers, Normal for Distribution, 100 for Mean and 5 for Standard Deviation and place the cursor in B2 after selecting Output Range (Figure 4-3).

5. Select Data - Data Analysis - Random Number Generation.

6. Select 1 for Number of Variables, 1000 for Number of Random Numbers, Patterned for Distribution, From 1 to 2 in Steps of 1, repeating each number 5 times, repeating the sequence 100 times and place the cursor in A2 after selecting Output Range (Figure 4-3).

7. Select Column Group, Home - Find & Select - Replace. Find what: 1, Replace with: Group 1, Replace All. Find what: 2, Replace with: Group 2, Replace All.

After you generate the data (Figure 4-4), you will have a different result on your worksheet.

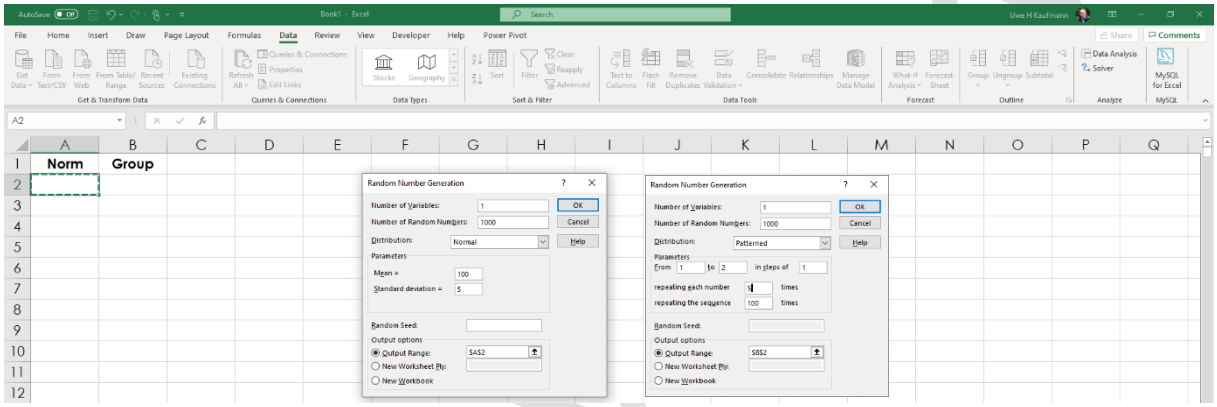


Figure 4-3: Generating 2 Columns with Random Data

Task 4-2: Analyse Descriptive Statistics of Norm

1. Select Data - Data Analysis - Descriptive Statistics.

2. Select $B$1:$B$1001 for Input Range. Or, just select B1 with your cursor and then select Control + Shift + ⇩ to mark the whole data range.

3. Select New Worksheet Ply and check Summary Statistics and Confidence Level for Mean at 95% (Figure 4-4).

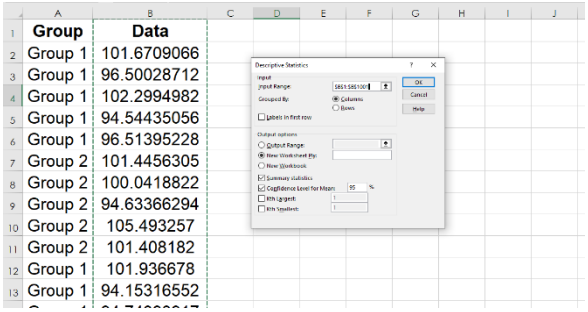


Figure 4-4: Determining Descriptive Statistics for Norm

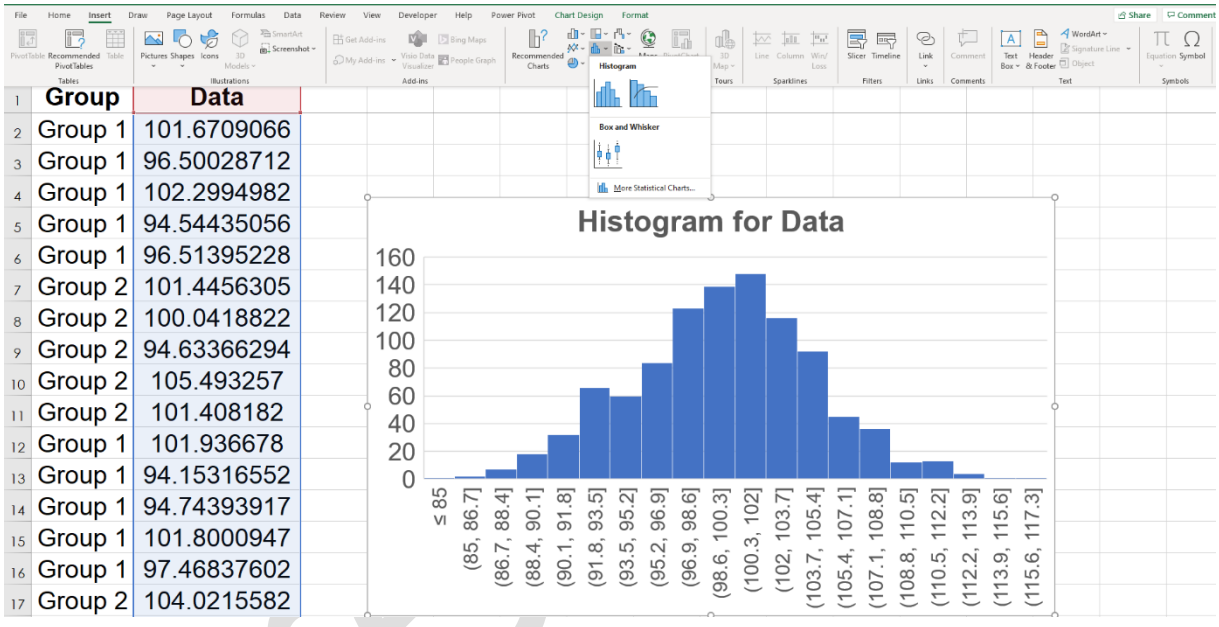
As a result, the descriptive statistics with a list of most basic indicators for your data Norm (Figure 4-5, your descriptive statistics will be different). The descriptive statistics will be explained later in this book.

Norm	
Mean	99.6988
Standard Error	0.1570
Median	99.8703
Mode	94.1532
Standard Deviation	4.9662
Sample Variance	24.6635
Kurtosis	- 0.0133
Skewness	0.0037
Range	31.0480
Minimum	84.9213
Maximum	115.9693
Sum	99,698.8065
Count	1000

Figure 4-5: Descriptive Statistics for Norm

Task 4-3: Plot Histogram for Norm

1. Select Data on Sheet1 (Mark B1 and select Control + Shift + ⇩).
2. Select Insert - Chart - Histogram.
3. Save your work with the name Norm.xlsx on the desktop or a folder of your choice.



	Group	Data
2	Group 1	101.6709066
3	Group 1	96.50028712
4	Group 1	102.2994982
5	Group 1	94.54435056
6	Group 1	96.51395228
7	Group 2	101.4456305
8	Group 2	100.0418822
9	Group 2	94.63366294
10	Group 2	105.493257
11	Group 2	101.408182
12	Group 1	101.936678
13	Group 1	94.15316552
14	Group 1	94.74393917
15	Group 1	101.8000947
16	Group 1	97.46837602
17	Group 2	104.0215582

Figure 4-6: Histogram for Norm

The histogram for Norm will be created (Figure 4-6). This histogram has been beautified by amending the X axis in Bin Width and Tick Marks and by adding a Chart Title. This histogram would be ready for being inserted into a PowerPoint presentation, Word Document or even Email or Chat.

Task 4-4: Show Boxplot for Norm by Group

1. Select Data and Group on Sheet1 (Mark A1:B1 and select Control + Shift + ⇩).
2. Select Insert - Chart - Box and Whisker.

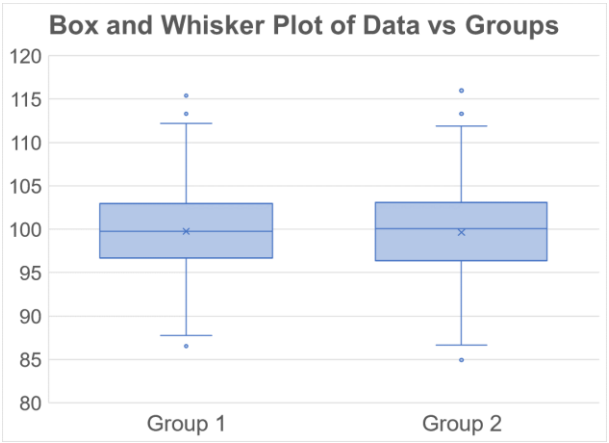


Figure 4-7: Box and Whisker Plot of Norm by Group

This command results in the box plot seen in Figure 4-7.

Downloading and Using MS Power BI

MS Power BI is available for download from Microsoft Store if MS Office is running on your computer. Microsoft Power BI puts visualisations at your fingertips. If Power BI web services are available to you, you may even share your visualisations, i.e. your dashboards with your team via Power BI server.

Task 4-5: Load the Data from Norm.xlsx into Power BI

1. Open MS Power BI Desktop.
2. Select Get Data - Excel - Connect.
3. Select Open Norm.xlsx in the location where you have saved it.
4. Check the box in front of Sheet1 - Load (Sheet1 carries the data table).

Unlike Excel, data will not appear as table on the screen. Data tables and columns are listed under Fields on the right.

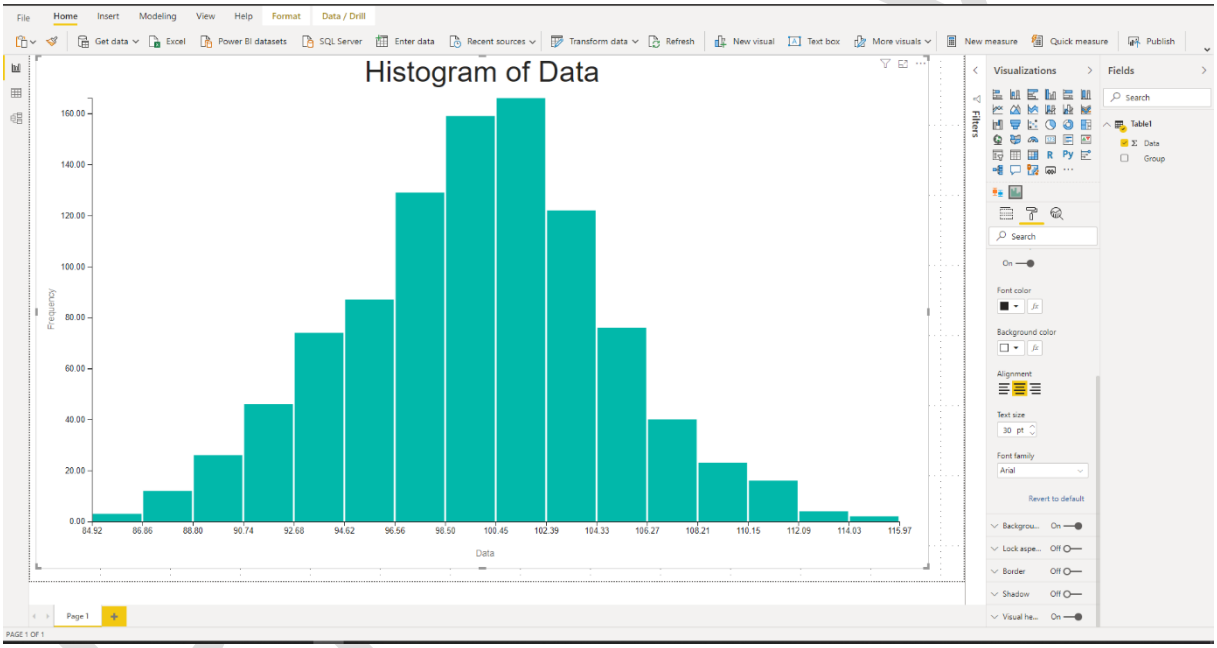


Figure 4-8: Power BI Histogram of Norm data

Task 4-6: Load a visual for a histogram from the Power BI repository

1. Select "…" under Visualizations - Get more visuals
2. Search for histogram
3. Add Histogram Chart (or any other histogram of your choice)
4. Select the newly imported histogram icon
5. Increase the size of the plot area to your liking
6. Ensure the histogram plot area is selected.
7. Pull the field Norm from the right into the Values box under Visualizations.

8. Amend the format of Title, Axes, etc.

Power BI exists for desktop and mobile devices. All versions are available from the Microsoft Store.

Downloading and Using R and R Studio

Firstly, install R and R Studio from any server provided on the websites:

- Download R from https://www.r-project.org/.
- Download R Studio from https://www.rstudio.com/.
- Start R Studio.

R programming offers a set of large and ever-growing set of libraries that help programming routines, analyse data and plot visualisations with minimal code and great flexibility. Most of these libraries need to be loaded before functions can be used. We will elaborate this in detail during our case applications in this book.

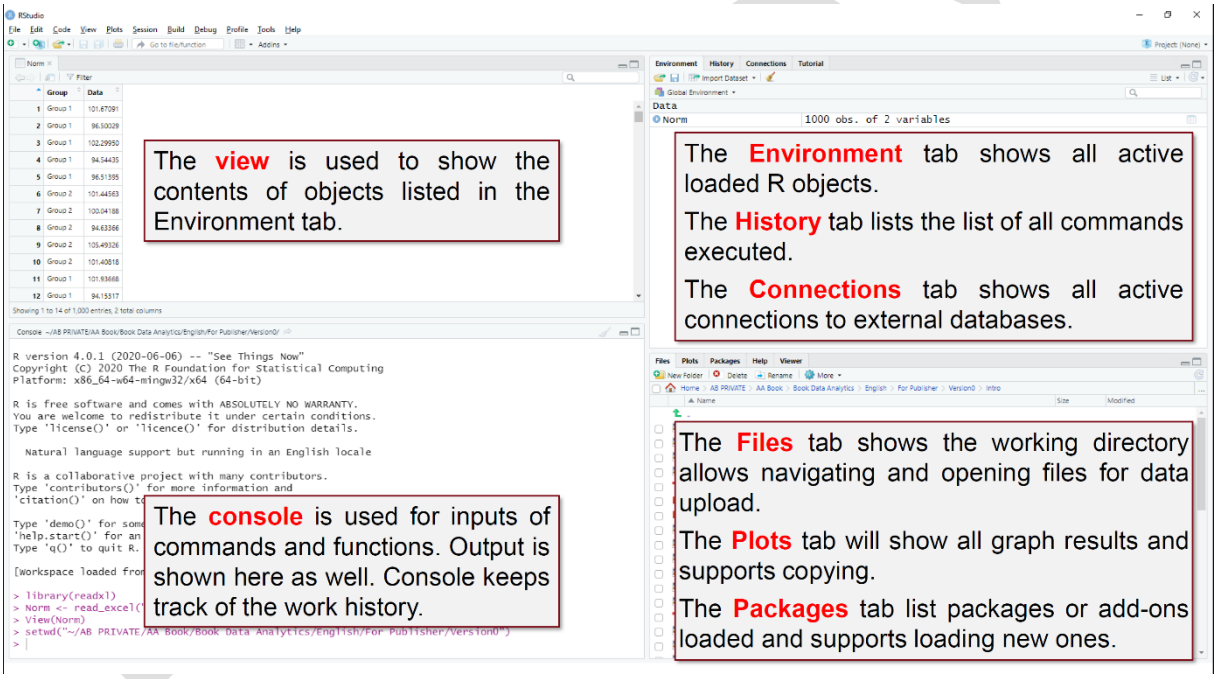


Figure 4-9: R-Studio User Interface

Task 4-7: Load the Data from Norm.xlsx into R Studio and Load basic analysis and graphic pack "pastecs" (Figure 4-9).

1. Open R Studio.
2. Select File - Import Dataset - From Excel.
3. Select Browse - Norm.xlsx from the location where you have saved it.
4. Select Sheet: Sheet1 - Import.
5. Data table Norm is shown in Tab Norm and in Environment.
6. In Tab Packages, Search for "pastecs", if found, check it.

7. If not found, select Install.
8. Install Packages - Packages: pastecs - Install.

Task 4-8: Analyse Descriptive Statistics of Norm.

Type the following in the Console:

```
# See all packages installed and available in Tab Packages
library()     # If package pastecs is part of the list, do
install.packages("pastecs")
# Show packages currently loaded
search()      # If package pastecs is not loaded, do
library(pastecs)

# Show descriptive statistics for all columns at data frame (table) Norm
stat.desc(Norm)
# Show descriptive statistics for column Norm at data frame (table) Norm
stat.desc(Norm$Data)
```

Output:

1. All columns at table Norm are included in descriptive statistics (Figure 4-10).
2. Only column Norm at table Norm is included in descriptive statistics (Figure 4-10).

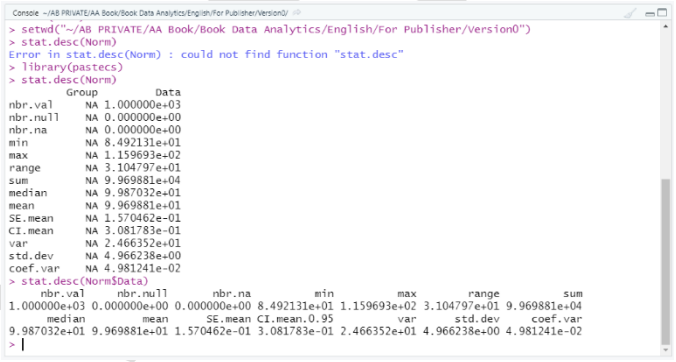


Figure 4-10: Descriptive Statistics for Norm

The output of the descriptive statistics gives an overview of parameters that describe the dataset. It is the most basic analysis for drawing conclusions about central tendency, variation and shape of the distribution of the dataset.

Task 4-9: Perform Normality Test (Shapiro-Wilk test) for Norm

```
# Performing test on normality of column Norm at data frame Norm
shapiro.test(Norm$Data)
```

Output:

The normality indicators for Norm are shown indicating that Norm follows normal distribution (Figure 4-11), i.e. p-value > 0.05.

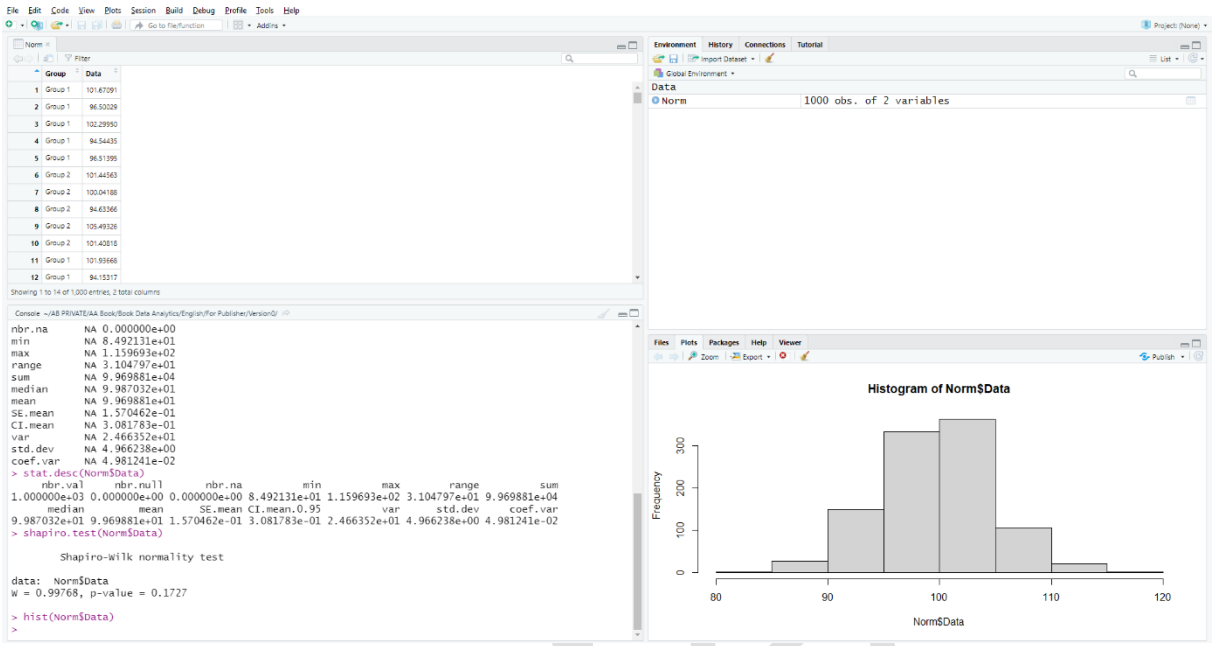


Figure 4-11: Descriptive Statistics, Normality Test and Histogram for Norm

Task 4-10: Plot Simple Histogram for Norm

```
# Plotting histogram for Norm
hist(Norm$Norm)
```

Output:

A basic histogram for Norm is displayed in tab Plots (Figure 4-11).

Since we have just confirmed that our data in column Norm is indeed normally distributed, we may also add the bell shape to the histogram.

Task 4-11: Plot Simple Histogram for Norm with Bell Shape

```
# Moving column Norm into variable Data
data <- Norm$Data
# Calculating mean and standard deviation of data
m <- mean(data)
std <- sqrt(var(data))
# Plotting histogram for Norm
hist(data, density=20, breaks=20, xlab="Norm", main="Normal Curve over Histogram",
cex.main=2.00, col="lightblue", cex.lab=1.50, cex.axis=1.50, prob=TRUE,)
# Set margin
par(mar = c(5, 5, 5, 5))
# Plot bell shape
curve(dnorm(x, mean=m, sd=std), lwd=2, add=TRUE, yaxt="n", col="darkblue")
```

Output:

A basic histogram for Norm including bell shape representing normality is displayed in tab Plots (Figure 4-12).

Select Plots - Export - Copy to Clipboard makes the histogram available for use in other programs. Figure 4-12 has been inserted this way.

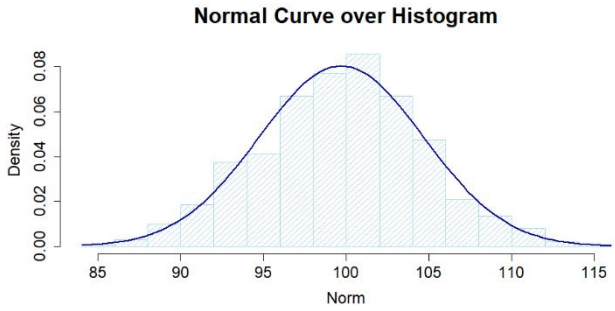


Figure 4-12: Histogram and Bell Shape for Norm

These very basic examples show how R Studio can help doing analyses and visualisations very fast and with minimal input. The Shapiro-Wilk test on normality does not exist in MS Excel and therefore is an excellent addition to your collection of tools.

Especially complex data analyses like exploratory factor analysis (EFA), confirmatory factor analysis (CFA), structural equation modelling (SEM) and many more rather complicated procedures can be run in R Studio with minimal effort. It might be, that additional packages need to be loaded to make other tools available. All these packages are accessible for download and installation.

The tab Help offers info to available functions and the packages they are in.

The abovementioned functions are not available in MS Excel and would be very hard to be programmed in Visual Basic. R Studio fills this gap and enriches your toolbox enormously.

5 What Is Provided

In addition to this book, a comprehensive set of aids is provided to the reader. These aids comprise of:

- All data used in case studies in this book are provided in the format Microsoft Excel.
- All data analyses have been prepared using Microsoft Excel, partially with the built-in macro collections "Data Analysis Tools" and R Studio.
- All graphical analyses have been conducted using either Microsoft Excel, Microsoft Power BI or R Studio.

The respective data sets as well as graphical and analytical results are provided, too.

There are a multitude of analytics software packages. And most of the cases in the book can be analysed using any of these. However, we have not met too many people who have fully mastered Microsoft Excel and Microsoft Power BI, which are most likely available for all readers without any investment into new software and the task to learn a new user interface.

We would like to encourage our readers make full use of Microsoft Excel and Power BI for data analysis and presentation. There is no need to rush for other software packages if MS Excel and Power BI meet the requirements.

With this, it should be possible for you, the reader, to follow through typical data analytics cases step by step as described in the book, and to customise these steps for your own data analytics cases.

6 Which Cases Should I Study?

Case studies introduced and analysed in this book make use of different tools for data science tasks. If you wish to start your journey through data analytics with the well-known and powerful pack of **Microsoft Excel** functions, the following cases might be of interest to you:

1. Which supplier has the better product quality?
2. Why Does Finance Pay Our Vendors Late?
3. Do we have enough people to run our organisation?

In the above-mentioned chapters, only MS Excel tools are employed for data preparation and analysis tasks. Of course, we make use of Excel's Analysis ToolPak.

If you wish to add **Microsoft Power BI** for displaying your data and creating informative and interesting dashboards to your tools backpack, we recommend studying the following cases:

4. What influences our staff's innovative work behaviour?
5. How to create a patient satisfaction dashboard?

After you have mastered the basic steps of data analytics with above mentioned tools, you may want to extend your toolset by a very powerful instrument for data wrangling, data visualisation and data analysis. The **programming language R** has

become an easy to use and extraordinarily powerful software environment for all tasks data scientists are faced with.

Do not worry. There is no need to learn a programming language. We have prepared all commands in our case studies that will open the magic of R for you. As you have already seen in the previous chapter, it needs one line of function to perform a normality test, one line to draw a histogram and one line to give a comprehensive set of descriptive statistics. Whilst learning R, we had to conclude that we are usually much faster in R than doing a similar task in other software. The following cases will help you diving into R:

6. Great, we have improved … Or not?
7. What drives our patient satisfaction?
8. How to create a patient satisfaction dashboard? (with R in Power BI)
9. Why are we wasting blood?
10. Making Better Decisions - Knowing the Risk of Being Wrong
11. What does our engagement survey result mean?
12. What drives our staff out?

Especially the last case is filled to the brim with R code to support a rather complex logistic regression analysis task. If you wish to start with R, you may want to begin your journey with the first examples on the list above.

We wish you success and fun for studying and following through our cases. After that it should not be too difficult to apply these tools on your own cases for organisational development.