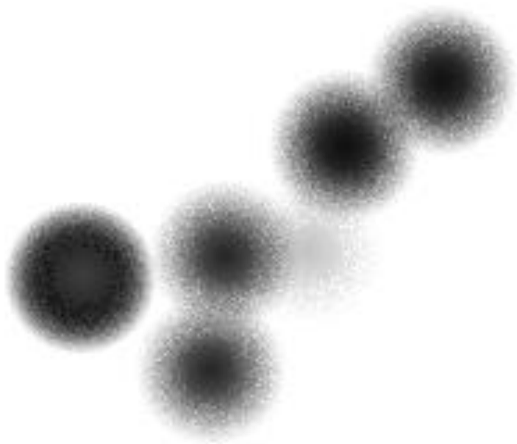
Dynamic Pattern Synthesis

Using Microsoft® Excel

2nd Edition

Philip Haynes

David Alemna



White Horse Books

Chichester, UK

www.whb.co.uk

© 2023 Philip Haynes and David Alemna

The rights of Philip Haynes and David Alemna to be identified as the authors of this work has been asserted by them in accordance with sections 77 and 78 of the Copyright, Designs and Patents Act 1988.

This digital work is licensed under the Creative Commons Attribution – Non Commercial – No Derivatives 4.0 International License.

To view a copy of the license, visit https://creativecommons.org/licenses/by-nc-nd/4.0/ .

Printing and photocopying of the entire work, or sections of the work, for the purposes of educational learning, teaching, and training is permitted, provided that the work is properly cited and attributed (for example by inclusion of the title and preliminary pages in any distribution). This permission includes making a digital copy of the complete work available on an educational, or similar, website. The purpose or aim of any distribution must be for the learning and education of others. No attempt should be made to alter the representation of the authors or the first publisher (White Horse Books).

Under the terms of the Creative Common Licence, adaptations of the work and/or the use, or reprint, or photocopying, or electronic digital distribution of the work, or part of the work, for the purpose of making a commercial profit, are not permitted.

To cite this work in additional and new - but related - materials, the following text is suggested:

Haynes, P. and Alemna, D (2023) *Dynamic Pattern Synthesis* Chichester: White Horse Books

Product or corporate names may be trademarks or registered trademarks and are used only for identification and explanation without intent to infringe.

British Library Cataloguing-in- Publication Data

A catalogue record for this book is available from the British Library

E-book

ISBN: 978-0-9534065-3-1

Contents

List of Figures 5

List of Tables 6

Introduction and preface to the 2nd Edition 8

Chapter one: What is Dynamic Pattern Synthesis (DPS) 10

- Exploring complexity 11
- Previous influences 12
- Sampling with DPS 13
- Replication 13
- Inferential statistics 13
- Effect 14
- Choosing cases and variables 14
- The training dataset 15

Chapter two: Case Patterns as clusters 17

- Preparing the data for cluster analysis 17
- Interpreting the table of Squared Euclidean Distance 19
- Analysing the possible cluster structure 21
- Interpreting the Euclidean distances 22
- Identifying unique pairs 22
- Validating the proposed cluster structure and number of clusters 23
- Compute descriptive statistics 24

Chapter three: Cluster pattern synthesis: clusters and their variable patterns 26

- Presenting the data as tables of cluster configurations 26
- Near misses and fuzzy relationships 31
- Cluster boundaries and outliers 32
- Reconstituting the table to include an outcome variable 32

Chapter four: Repeating the DPS with longitudinal data 35

- Data synthesis for 2016 35
- Data synthesis for 2017 42

Chapter five: Concluding the DPS: longitudinal change 52

- Changes of variable scores over time 52
- Cluster changes 53
- Combined case and variable changes 54

Checking the consistency of case based variable averages over time 62

Converting the scale data to crisp set categories 62

The qualitative DPS summary 68
Conclusion.. 69
The datasets 71
Acknowledgements 72

List of Figures

Figure 3.1 Excel® custom sort, by cluster 26

Figure 3.2 Custom sort, drop down menu 27

Figure 3.3 Custom sort sub menu 27

Figure 3.4 Final set up of the custom sort menu 27

Figure 3.5 Conditional formatting with colour scales 28

List of Tables

Table 1.1 The complexity domain in research 12

Table 2.1 Standardising a matrix of case and variable scores, 2015 data 18

Table 2.2 Squared Euclidean Distance Matrix for Cases, 2015 data 21

Table 2.3 Summary of the proposed cluster structure, 2015 data 23

Table 2.4 Cluster pattern synthesis – adding clusters and descriptive statistics, 2015 25

Table 3.1 Cluster Pattern Synthesis – table sorted by clusters and variables, 2015 28

Table 3.2 Cluster patterns, as demonstrated with colour scales, 2015 29

Table 3.3 Cluster Pattern Synthesis – using monochrome shading to identify patterns, 2015 30

Table 3.4 Clusters resorted against a single outcome variable, 2015, debtors 32

Table 4.1 Cluster construction data for 2016, applying the Squared Euclidean differences 36

Table 4.2 Cluster pattern synthesis – clusters and descriptive statistics, 2016 38

Table 4.3 Cluster pattern synthesis – using variable colour coding, 2016 39

Table 4.4 Cluster pattern synthesis – using monochrome shading to summarise clusters, 2016 40

Table 4.5 Clusters resorted to show a single outcome variable, debtors, 2016 41

Table 4.6 Cluster construction data for 2017, applying the Squared Euclidean differences 42

Table 4.7 Cluster pattern synthesis – adding clusters and descriptive statistics, 2017 45

Table 4.8 Cluster pattern synthesis – using variable colour coding, 2017 46

Table 4.9 Cluster pattern synthesis – using monochrome shading to summarise clusters, 2017 47

Table 4.10 Cluster 3 - Squared Euclidean Distance Matrix 48

Table 4.11 Cluster pattern synthesis – examining cluster 3 validation in more detail, 2017 49

Table 4.12 Clusters resorted to show a single outcome variable, debtors, 2017 50

Table 5.1 Variable mean changes: 2015-17 53

Table 5.2 Longitudinal cluster patterns 53

Table 5.3 Excel® calculation example for longitudinal average score for JB Alpha 54

*Table 5.*4 Longitudinal averages (2015-17) 55

Table 5.5 Raw data and simple cluster calculations for longitudinal case comparisons 56

Table 5.6 Longitudinal averages (2015-2017) sorted by clusters 58

Table 5.7 Longitudinal averages (2015-2017) showing cluster colour map 59

Table 5.8 Longitudinal averages (2013-2015) using monochrome shading to identify patterns 60

Table 5.9 Longitudinal averages (2013-2015) sorted by debtor outcomes 61

Table 5.10 An example of scale data converted to binary crisp set scores (2015) 63

*Table 5.*11 Setup for a longitudinal table with consistent above and below average scores 65

Table 5.12 Cases with consistently above or below median average scores from 2015-2017, sorted by variable patterns 66

Table 5.13 Cases with consistently above or below median average scores from 2015-2017, sorted by debtors outcome 67

Introduction and preface to the 2nd Edition

The primary purpose of this second edition is to update the research method for the applied social sciences, known as *Dynamic Pattern Synthesis.*

Dynamic Pattern Synthesis (DPS) is a multimethod for the social sciences that seeks to better represent social complexity in data and over time.

By multimethod, we mean a combination of techniques, to strengthen the overall approach, rather than relying on one technical approach.

The development of DPS was never intended to follow a single, rigid convention of the same repeated algorithms. The methodological aim was to find a pragmatic and relatively simple method using applied techniques to study social science data in the context and challenge of complexity.

The first edition used an example of Hierarchical Cluster Analysis (HCA) being combined with elements of Qualitative Comparative Analysis (QCA) to understand the construction of different case clusters and to attempt to explore and explain clusters relationship with outcomes. Also, to see how any emerging patterns of similarity and difference between cases evolved over time. It required both access to IBM SPSS and Microsoft® Excel. The methods outlined in this latest edition can all be performed with Microsoft® Excel. This will hopefully make the method more accessible to a wider audience.

One of the criticisms that followed the previous approach to DPS was its lack of attention to the growing number of QCA conventions. For example, DPS was criticised for practices like using too many variables with a small number of cases, and not adequately explaining contradictory outcomes (shared data conditions that have different outcomes) and logical remainders (data patterns that do not relate to any outcome). DPS in its earliest forms in the first edition was, at best, a working example of finding complex solutions in QCA and, at worst, failed to offer discrete QCA parsimonious and intermediate solutions. As QCA become increasingly disciplined and rigorous, in this respect, DPS looked more like a periphery and adulterated form of QCA, perhaps better separated in the future from these conventions and constraints to become a recognised configurational case-based method in its own right.

This separation from the idea of QCA is exactly what we have done with this second edition. Dynamic Pattern Synthesis is still influenced by the growing popularity and achievements of QCA, and its original case configurational philosophy, but nevertheless DPS now stands on its own merits as a member of a growing number of configurational cased based methods, each with its own strengths and weaknesses.

In this second edition of the DPS manual, DPS remains closer to the presentation and understanding of case diversity as evidenced in scale data, with much less use of binary crisp set approaches. This also allows DPS to be used, where appropriate, with other descriptive, inferential and effect based approaches to statistics (although we do not explore this in detail in this book, it is an option for those using DPS). Dynamic Pattern Synthesis remains closer to an explorative approach, than an explanatory approach, but nevertheless is committed to using explorative evidence to begin to explain possible causal evidence through qualitative interpretation. If inferential statistics are added to DPS in the future (this is an option, not a requirement), DPS can also have an increased ability for quantitative explanation, this in addition to the qualitative interpretation of the patterns discovered.

Like the first edition, this second edition book assumes the basic ability to use Microsoft® Excel and for the reader to have access to this software. It is certainly possible to compute, analyse and teach a DPS exercise with this software.

In order to work through the example in the book, there is access via the publishers' website to download the datasets. This is explained at the end of the book.

Dynamic Pattern Synthesis (DPS) seeks to model social science data over time. It does this with small samples of data. This enables it to get a sense of realism about the diversity of case experiences, while observing both similarity and difference. It borrows from both principles of quantitative and qualitative research methods.

The origins of the DPS method were forged from three different areas of research: complexity theory, cluster analysis, and Qualitative Comparative Analysis (QCA) (See Haynes, 2017). In addition, DPS has been influenced by critical realism and realistic evaluation (Pawson & Tilley, 1997). It accepts a need to search for partial causal mechanisms, while always placing these in a dynamic social context. These influences are discussed in chapter one.

Chapter two explains the contribution of cluster analysis to DPS. In this edition, the approach to cluster analysis is deliberately minimalist and exploratory, allowing the reader to use basic Excel® formula and algorithms, while also checking with data tables whether clusters derived from multivariate analysis have validity and make sense in practice. Do the mathematical patterns produced by cluster analysis have useful 'real word' meaning?

Chapter three progresses to include the validation and theorisation of clusters with configurative tables, as originally influenced by QCA in the first edition. This second edition takes further the idea of considering in more detail how variables influence cluster memberships in diverse ways. As in the first edition, this includes consideration of how to promote one variable to an outcome status, for purposes of evaluation and explanatory research.

Chapter four introduces the longitudinal element of the method, where the pattern models for each time point are linked and compared over time.

Chapter five concludes on the qualitative interpretation of what a full DPS model reveals, including the consideration of the longitudinal results on a single outcome variable.

Chapter one: What is Dynamic Pattern Synthesis (DPS)

Dynamic Pattern Synthesis (DPS) is designed to examine complex patterns in longitudinal datasets. The method has evolved to advance configurative case-based methods. It adds a sensitivity to exploring change over time, in order to better understand dynamic social and economic change and trends.

Dynamic Pattern Synthesis was first designed to be used with relatively small samples of cases, but can be scaled up to be used with larger samples. This book uses a small sample, as an example.

The method was first presented by one of the authors (Haynes) in 2014 at an international research seminar held at the University of Warwick, UK. The seminar was part of the UK Economic and Social Research (ESRC) Council seminar series on *Complexity and Methods in the Social Sciences*. Before that, he had been using separate examples of cluster analysis and qualitative analysis and had begun to experiment with ways of combining them (Haynes & Haynes, 2016; Haynes, 2014, 2012).

This second edition, like the first, is designed specifically to be a practical 'how to do it guide'. It shows the novice exactly how to compute and calculate a DPS model. For those who want more theoretical background about how complexity theory influences the development of DPS, it is suggested that you read Haynes' previous monograph: (2017) *Social Synthesis: Finding Dynamic Patterns in Complex Social Systems*, Oxon: Routledge. There are a growing number of working examples of DPS published in research journals. Examples are: trajectories of local government finance in England (Taylor, Haynes, & Darking, 2021), a comparison of West African and South American economies (Alemna et al, 2021; Alemna, 2022) and configurations of COVID-19 country fatalities (Haynes & Alemna, 2022).

Dynamic Pattern Synthesis has five stages:

Stage 1 The exploration of case patterns in clusters

Stage 2 The exploration of variable patterns to validate the best case-based cluster patterns

Stage 3 Reformulating stage 2 to focus on a specific outcome variable

Stage 4 The exploration of longitudinal patterns and theorising about dynamic patterns over time

Stage 5 Concluding on any longitudinal explanatory patterns

Exploring complexity

Dynamic Pattern Synthesis is developed from the world view offered by complexity theory (Boulton, Allen, & Bowman, 2015). Complexity theory illuminates that in many areas of science and social science any causal effects discovered are often contingent on the context. For example, casual mechanisms might vary according to the historical time point or the spatial location in which they are situated. In this sense, DPS has some similarities with the approaches of *critical realism* and the idea of *realistic evaluation* (Pawson & Tilley, 1997)

Table 1.1 illustrates the scientific issues when researching the complexity domain. In a simple domain, there is stability in existing cause and effect relationships. An example is the gravitational effect that creates tides. These are predictable to very precise times and can be published in standard tide tables. A complicated domain has strong elements of prediction, but the possibility that occasional phenomena and events might disrupt the predictability. Examples are flying an aircraft. Although computerised aircraft are highly predictable and safe in their mechanical predictability, there is still the very small chance that they can experience major disruption caused by human error and interference, mechanical failure, or an exceptional external weather event.

In the complex domain (table 1.1), disruptions to cause and effect are much more likely. This includes disruptions when trying to replicate a known cause and effect in a different time and place. This is exactly the domain where DPS is designed to operate. Research in this domain needs to search for patterns and to examine how consistent and replicable those patterns are over time and across physical space. An example in science research is studying the behaviour of a group of animals, like a herd of elephants or flock of birds. An example in social science is the use of psychological therapy, where a specific therapy may work in some situations, but it is difficult to generalise to multiple places and over time (as society and its cultures and resulting behaviour changes). Rather than demonstrating and replicating cause and effect it may be possible to identify patterns and probabilities about when the therapy is more likely to work or not work, but these patterns will be subject to numerous disruptions and cannot be expected to be highly reliable.

Finally, arguably the most difficult task for researchers is to research the domain of chaos. Here instability is the norm and any pattern replication will be very short term. While pattern analysis may still be relevant, the focus might have to be on single or very small numbers of cases. Weather forecasting is an example in scientific research where there is much instability. Rather than trying to make clear statements of prediction, such as it will rain today, it may be better to say there is a 70% probability of rain in the next few hours. Similarly, when unexpected events like floods and disease disrupt an economy, it is extremely difficult to predict the impact with any precision. In social science research, an example is trying to predict the employment attendance of those who persistently consume excessive amounts of alcohol.

Table 1.1 The complexity domain in research

	Simple	*Complicated*	*Complex*	*Chaos*
Scientific prediction	Predictable	Bounded prediction	Temporary forecasting	Short term probabilities
Dynamic	Stable	Stable with occasional disruptions	Mix of stability and instability	Unstable
Research evidence	Cause and effect	Linear trends and statistical controls	Pattern analysis	Individual case studies Event probabilities
Science example	Tide tables	Flying an aircraft	Human behaviour	Weather
Social science example	National controls on the price of alcohol influence national levels of consumption	Local level of alcohol consumption and local emergency admissions	The use of a psychological therapy to manage depression	Employment attendance by excessive consumers of alcohol

Adapted from Snowden & Boone (2007)

Previous influences

Dynamic Pattern Synthesis is developed from previous research methods that are well tested and widely used. In this sense, it is an incremental development of previous practice. The two major historical influences on DPS are: cluster analysis and Qualitative Comparative Analysis (QCA). These are both examples of configurational case-based research (Ragin, 2014). Configurational case-based research looks for evidence that groups of cases are similar, at least, for a period of time. Cases are recognisable social entities like people, organisations, businesses, or even regions and nation states (Byrne, 1998). We do not expect cases to be identical, but assume some will share important similarities.

Cluster Analysis (CA) was first developed to categorise animal and plant species into similar types (Everitt, 1993). Multiple measurements can be combined in calculations of case-based similarity and difference, to give evidence for possible configurations about which cases are most likely to be similar. A strength of cluster analysis is its ability to compute using multiple variables with a range of scale measures. A weakness is that some small differences between cases and clusters can be found in mathematical patterns which have little substantive usefulness or meaning in real life.

Qualitative Comparative Analysis (QCA) was developed to theorise about causality in comparative sociology and political science, where the number of countries being compared was often small and where different configurations of political and social influence might still lead to the same result (Ragin, 1987). For example, countries could become stable democracies after following different historical paths with only some elements of shared similarity and other patterns of diverse influences (Rihoux & Ragin, 2009). In social science theory this is referred to as 'equifinality'.

For some of the most up to date information on researching complex configurations, see the UK government public administration research advice (Bicket, et al., 2020).

Sampling with DPS

The fundamental sampling principle in DPS is to choose a group of cases that can reasonably be compared together. Therefore, the sample needs a degree of expected similarity. These will be cases that share some common attributes, but where it is also reasonable to expect them to have differences that are interesting and important to understand. For example, if interested in macro political economy, one would start by comparing a group of countries based on shared continental geography (Western Africa), or with broadly similar economies (European countries sharing the Euro currency).

The sampling strategy with DPS is, therefore, purposeful, and not inferential. Inferential sampling is where the researcher takes a random sample from a large population and then uses probabilistic inference to predict whether sample results can be generalised to the larger population.

DPS sampling is purposeful because there is a deliberate attempt to compare a small group of cases. In this sense, the sampling method is more similar to the approach taken in qualitative social science research, rather than in quantitative research.

Replication

Having found a dynamic pattern in a first, small sample, it is then reasonable to compare the results with another group of cases. As with any principle of research replication, the next sample group should be chosen so it has some similarities to the previous sample (so that patterns can still be managed and observed) and with the purpose of finding an interesting and logical further comparison. For example, if having studied all the members of the Euro currency, one might then conduct a DPS on all the other countries in the European Union that are not in the single currency. Or if researching a large database of human participants, one can start with taking a small subgroup, like all those of a very similar age. Having found a patten in this group, one can then compare with a slightly older, or younger age group, and progress to understand larger patterns in this way.

Inferential statistics

In general, the authors do not prioritise the use of inferential statistics with DPS, because DPS is not designed to generalise from a sample to a larger population. Nevertheless, there may be situations where inferential statistics can be usefully applied to the results of DPS. For example, where a small sub sample has been drawn from a previous inferential sample of a national population and the researcher wants to calculate if results found in the small sample could reasonably be inferred back to the original population. In this situation, having decided that clusters were robust and valid, standard statistical operations like ANOVA could be used to see if the differences in mean scores between clusters are statistically significant or not (as when handling the groups as independent samples). Similarly, there are occasions when repeat measure inferential statistics might be used to examine whether cluster changes over time are chance effects, or statistically significant.

Effect

The consideration of the effect of variables on clusters is one key element of DPS. This analysis takes place in the second stage of the method. Tables of variable patterns are used to diagnose the effect of variables on cluster membership. These configurational tables allow for a more complex and appropriate understanding of how variable effects can be experienced differently by diverse configurations of cases and clusters.

Choosing cases and variables

The choice of cases is determined by purposeful sampling, as referred to above. The choice of variables is based on finding as reliable a set of indicators as possible. Ideally, this will be secondary longitudinal dataset from a reputable database and source. Aim to have at least three time points, using the same variable measurements, in order to examine changing patterns over time.

DPS is flexible enough for additional categorical variables to be added at stage 2. Variables with binary categories with no ordinal or scale differences, can be also considered in the tables. This gives flexibility in understanding the influence of variables on cluster membership. At stage 2, additional binary categories can also be added to the model, alongside the scale variables entered at stage 1. These practices are not demonstrated in this short guide to DPS, but are entirely feasible.

The training dataset

The training dataset used in this book is available at the supporting publisher's website (https://whb.co.uk). It is a fictional dataset that compares 12 research businesses. The dataset is fictional and for the purpose for training and teaching the method.

The 12 cases are:

JB Alpha

Cosign Research

Mini Max

System Synthesis

Open Thinking

LKS Data

Strategy Statistics

Visual Research

Ashton Algorithms

Linear Logics

Sun Focus

New Perspectives

(The cases are labelled in a categorial variable called: Business Name.)

The scale variables (each with a measurement for each of the three years) are:

Capital Expenditure 2015, as a percentage of income

Capital Expenditure 2016, as a percentage of income

Capital Expenditure 2017, as a percentage of income

Annual income growth 2015, percentage change from previous year

Annual income growth 2016, percentage change from previous year

Annual income growth 2017, percentage change from previous year

Postgraduate level qualifications 2015, percentage of the workforce

Postgraduate level qualifications 2016, percentage of the workforce

Postgraduate level qualifications 2017, percentage of the workforce

Gender pay gap 2015, percentage of gross income

Gender pay gap 2015, percentage of gross income

Gender pay gap 2015, percentage of gross income

Marketing Expenditure as a percentage of income 2015

Marketing Expenditure as a percentage of income 2016

Marketing Expenditure as a percentage of income 2017

Number of staff per line manager 2015, ratio

Number of staff per line manager 2016, ratio

Number of staff per line manager 2017, ratio

Overseas business 2015, percentage of customers

Overseas business 2016, percentage of customers

Overseas business 2017, percentage of customers

Customers retained 2015, percentage

Customers retained 2016, percentage

Customers retained 2017, percentage

Late payment invoices over one year, 2015, percentage of customers

Late payment invoices over one year, 2016, percentage of customers

Late payment invoices over one year, 2017, percentage of customers

Staff turnover 2015, percentage of staff

Staff turnover 2016, percentage of staff

Staff turnover 2017, percentage of staff

Employee absence with illness 2015, average days absent

Employee absence with illness 2016, average days absent

Employee absence with illness 2017, average days absent

Chapter two: Case Patterns as clusters

The first stage of DPS involves using cluster analysis. Cluster analysis examines case scores across all variables and compares cases for similarity.

Cluster Analysis has been used for many decades and became more sophisticated after the wide availability of computers. This also led to some criticism of the method, as it was soon discovered that using slightly different mathematical techniques to measure the similarity and differences between cases could lead to different conclusions about where the boundaries between clusters should lie, and which cases should sit in which clusters.

In short, these challenges reveal the 'fuzzy' nature of case-based clusters and that cases are similar and different to each other in a multitude of ways, depending on which variables and variable interactions the researcher and their algorithms focus on.

Given the fuzzy nature of cluster boundaries, it is sometimes difficult to place a case with others in a very conclusive way, and it is not uncommon for a model to find the existence of a case that could potentially be argued to sit with two different cluster groups, or perhaps would be better left as a single outlier.

Our approach in this book is to encourage the researcher to make final decisions about the location of cases in clusters with the best possible insight into case-variable relationships, this rather than making the decision solely on the basis of a single mathematical algorithm.

The approach to clustering taken in this book is to keep the mathematics relatively simple, and to give you – as the researcher – the maximum ability to see how variables influence cluster memberships. The researcher can then make an informed decision about where they want to argue the boundaries exist between clusters in the model they are developing. The mathematics informs the research decision rather than dictating it.

The reader should be aware that there are much more sophisticated mathematical approaches to measuring and modelling cluster memberships using more advanced Excel® techniques and alternative software. (For example, the first edition of this book, suggested using the IBM Statistical Package for the Social Sciences).

Preparing the data for cluster analysis

Unless all the variable data in your dataset is constructed with the same scales, i.e., percentage scores, and with the similar distributional characteristics, we recommend standardising data before performing a cluster analysis.

The simplest way to do this in Excel®, is to use z scores. Here, the data for all variables is standardised to a common scale where the mean is 0. Excel® uses the mean, and standard deviation of a variable to make this standardisation.

The Excel® formula to achieve this is:

=STANDARDIZE(x,mean,standard_dev)

Where is x is a variable score for a single case, mean is the mean average calculation for the variable with all cases, and standard_dev is the standard deviation calculation for the variable with all cases.

Table 2.1 Standardising a matrix of case and variable scores, 2015 data

	A	B	C	D	E	F	G	H	I	J	K	L
		Capexpend2015	AnIncomeGrow2015	PGT2015	Genderpaygap2015	Marketing2015	Managers2015	Overseas2015	Continuecustomers2015	Debtors2015	Staffturnover2015	Sicknessdays2015
3	JB Alpha	12.3	2.9	72.0	2.0	5.0	0.10	0.0	90.0	2.0	30.0	6.0
4	Cosign Research	11.1	3.0	54.0	3.0	4.3	0.03	6.0	84.0	2.0	15.0	4.0
5	Mini Max	4.5	4.0	32.0	3.0	5.2	0.02	0.0	86.0	3.0	16.0	7.0
6	System Synthesis	9.2	13.7	34.0	7.0	8.1	0.01	12.0	82.0	3.0	13.0	6.0
7	Open Thinking	8.7	15.6	67.0	1.0	4.2	0.05	6.0	100.0	0.5	16.0	5.0
8	LKS Data	3.1	8.9	76.0	1.0	4.0	0.05	5.0	98.0	1.0	8.0	4.0
9	Strategy Statistics	2.1	6.9	90.0	1.0	4.6	0.04	3.0	89.0	1.0	21.0	9.0
10	Visual Research	9.8	20.3	43.0	3.0	5.7	0.05	8.0	84.0	3.0	2.0	7.0
11	Ashton Algorithms	7.1	2.8	56.0	1.0	7.2	0.03	4.0	77.0	3.5	14.0	6.0
12	Linear Logics	7.4	2.3	42.0	8.0	6.1	0.05	23.0	76.0	3.0	9.0	3.0
13	Sun Focus	5.7	7.1	56.0	2.0	3.7	0.04	4.0	69.0	5.0	7.0	4.0
14	New Perspectives	4.7	7.3	45.0	4.0	2.3	0.0	11.0	80.0	3.0	11.0	6.0
15	*Mean*	*7.1*	*7.9*	*55.6*	*3.0*	*5.0*	*0.04*	*6.8*	*84.6*	*2.5*	*13.5*	*5.6*
16	*Standard Deviation*	*3.1*	*5.6*	*17.0*	*2.2*	*1.5*	*0.02*	*6.0*	*8.5*	*1.2*	*6.9*	*1.6*
17		Standardized scores										
18	JB Alpha	1.68	-0.90	0.96	-0.45	-0.02	2.70	-1.13	0.64	-0.41	2.38	0.26
19	Cosign Research	1.29	-0.88	-0.09	0.00	-0.48	-0.59	-0.14	-0.07	-0.41	0.22	-0.99
20	Mini Max	-0.86	-0.70	-1.38	0.00	0.11	-1.06	-1.13	0.17	0.41	0.36	0.88
21	System Synthesis	0.67	1.04	-1.27	1.79	2.03	-1.53	0.86	-0.30	0.41	-0.07	0.26
22	Open Thinking	0.51	1.38	0.67	-0.89	-0.55	0.35	-0.14	1.81	-1.66	0.36	-0.36
23	LKS Data	-1.32	0.18	1.20	-0.89	-0.68	0.35	-0.30	1.57	-1.24	-0.79	-0.99
24	Strategy Statistics	-1.64	-0.18	2.02	-0.89	-0.29	-0.12	-0.64	0.52	-1.24	1.08	2.13
25	Visual Research	0.87	2.23	-0.74	0.00	0.44	0.35	0.19	-0.07	0.41	-1.66	0.88
26	Ashton Algorithms	-0.01	-0.92	0.02	-0.89	1.43	-0.59	-0.47	-0.89	0.83	0.07	0.26
27	Linear Logics	0.08	-1.01	-0.80	2.24	0.71	0.35	2.68	-1.01	0.41	-0.65	-1.61
28	Sun Focus	-0.5	-0.14	0.02	-0.45	-0.88	-0.12	-0.47	-1.83	2.07	-0.94	-0.99
29	New Perspectives	-0.8	-0.1	-0.6	0.447	-1.81	-0.1	0.69	-0.54	0.414	-0.361	0.26

For example, in Table 2.1, the formula is used to calculate standardized scores. As an example, JB Alpha’s standardised scores for Capital expenditure (Capexpend2015) is calculated as:

=STANDARDIZE(B3,B$15,B$16)

Where the raw data used is: B3 = 12.3, B$15 is 7.1 and B$16 is 3.1

For the above Excel® formula, the dollar sign $ is used to make sure that the formula activates using the same, fixed, row number. Therefore, given that the mean and standard deviation data is fixed in rows 15 and 16, the $ sign comes before the row number to ensure that when the formula is copied and pasted, Excel® still computes from the correct rows.

One of the best-known algorithms for assessing similarity and difference between cases is the Squared Euclidean Distance.

$$=\sum (x_i-y_i)^2$$

Where x and y are two comparable arrays of case scores.

In Excel® the Squared Euclidean Distance uses the following formula:

=SUMXMY2(array_x,array_y)

The algorithm is comparing two arrays of case scores rather than two arrays of variable scores.

Interpreting the table of Squared Euclidean Distance

Table 2.2 shows the result of computing the Excel® formula to calculate the Squared Euclidean Distance for each case's row of results. This is done, pair by pair, with each case compared with the others. First the case names must be cut and pasted into a row above the new table being formulated, as shown in table 2.2. A matrix is formed.

The formula therefore takes its two lines of array scores from two rows of data in table 2.1, not columns.

The first cell in the case-by-case matrix is the computation of JB Alpha with itself

=SUMXMY2(B$18:L$18,B18:L18).

This gives the result of 0.00, as there cannot be any difference between identical scores

The $ symbol in this formula tells Excel® to keep row 18 (JB Alpha's array of scores) constant in the calculation. The second array has no $symbols because it needs to change, as it is cut and pasted down through column B, so that it picks up a different case row, for each of the case comparisons.

For example, to compare the first case JB Alpha, with Cosign Research, the arrays entered into the formula are the first two rows of the data of standardised scores in table 2.1

=SUMXMY2(B$18:L$18,B19:L19)

This formula can be cut and pasted to calculate the rest of column B in table 2.2, but for each new column the formula needs to be altered manually.

The first array must be updated to include the correct column reference: **B$19:L$19** This ensures it picks up the second row of standardised scores.

The second array must be updated to the constant B18:L18 at the top of each column of the matrix.

As an example, for all the case comparisons for the second column (Cosign Research), the column reference for the first array becomes **B$19:L$19**.

This formula can now be cut and pasted to calculate all the column comparisons for Cosign Research.

Continue this process, updating the formula for the first array in each column, before copying and pasting it down the column.

The results for the whole matrix are shown in table 2.2

The matrix automatically repeats the calculations across the diagonal, and the diagonal is represented in table 2.2 by the 0.00 scores where a case is compared with itself and there is no dissimilarity.

Table 2.2 Squared Euclidean Distance Matrix for Cases, 2015 data

	A	B	C	D	E	F	G	H	I	J	K	L	M
		JB Alpha	Cosign Research	Mini Max	System Synthesis	Open Thinking	LKS Data	Strategy Statistics	Visual Research	Ashton Algorithms	Linear Logics	Sun Focus	New Perspectives
32	JB Alpha		20.26	31.78	48.45	21.08	30.30	27.06	38.99	26.57	49.58	40.07	34.16
33	Cosign Research	**20.26**		12.15	19.23	13.49	15.76	26.37	19.83	10.10	19.94	14.95	10.88
34	Mini Max	31.78	12.15		17.26	23.21	21.32	19.51	20.00	**7.72**	31.86	16.87	10.44
35	System Synthesis	48.45	19.23	17.26		31.60	38.57	42.69	**14.39**	16.72	**17.91**	28.97	22.56
36	Open Thinking	21.08	13.49	23.21	31.60		**7.06**	18.01	18.18	24.85	42.19	33.56	20.59
37	LKS Data	30.30	15.76	21.32	38.57	**7.06**		**15.75**	24.73	22.34	37.83	25.27	16.89
38	Strategy Statistics	27.06	26.37	19.51	42.69	18.01	15.75		34.14	21.18	55.69	36.22	23.02
39	Visual Research	38.99	19.83	20.00	**14.39**	18.18	24.73	34.14		18.61	30.47	20.47	16.26
40	Ashton Algorithms	26.57	**10.10**	**7.72**	16.72	24.85	22.34	21.18	18.61		26.05	11.58	16.04
41	Linear Logics	49.58	19.94	31.86	17.91	42.19	37.83	55.69	30.47	26.05		25.50	19.11
42	Sun Focus	40.07	14.95	16.87	28.97	33.56	25.27	36.22	20.47	11.58	25.50		**9.83**
43	New Perspectives	34.16	10.88	10.44	22.56	20.59	16.89	23.02	16.26	16.04	19.11	**9.83**	
44													
45	*Lowest distance*	20.26	10.10	7.72	14.39	7.06	7.06	15.75	14.39	7.72	17.91	9.83	9.83

Analysing the possible cluster structure

Row 45 in table 2.2 calculates the lowest distance pair in each column. This is where the lowest score is the least distance apart between the pairs, and hence the pair of cases are the most similar in that column.

Before performing the lowest distance calculations, you should delete all the zero scores along the central diagonal where a case is compared with itself. The diagonal is then represented by blank cells as in table 2.2.

As an example, the formula in Row 45, for the first column, finds the most similar pair with JB Alpha using:

=MIN(B32:B43)

The most similar pair of cases in each column in the matrix in table 2.2 can then be identified and indicated in bold text, or similar. This is Open Thinking and LKS Data (7.06).

You can now begin the process of developing simple cluster analysis.

Before doing so, it is important to keep in mind the limitations of these simple paired distance measures. These measures are an aggregate of all the variable scores for each case compared with another aggregate for another case. An aggregation of scores in this way does not guarantee the best understanding of how variable scores pattern with cases. For example, there might be just three variables from the eleven in total where the cases are identical, but eight where their scores are very different and at opposite points of the variable distributions. This will not give a similarity score that indicates they are similar! Even though the pattern is potentially one of interest.

It is for this reason that the calculation of the aggregate score of the Euclidean distance is only a starting point in the simple cluster analysis and not the final word on the matter. As we will see it is important to check the patterns suggested by the Euclidean distances with other detailed information about how each variable relates to pairs and groups of cases.

Interpreting the Euclidean distances

The approach below to exploring possible clusters from the Euclidean distances is based on the premise that we are seeking to find small groups of cases (clusters) that are significantly different to each other. In other words, we are not seeking to find a single larger group of cases that are similar. This dictates the approach we use below where we systematically identify 'unique pairs', rather than looking for one single group of closely related cases.

The Squared Euclidean Distance table 2.2 shows you the aggregate similarity and differences when all cases are compared with each other. Therefore, the process of identifying unique closely related pairs that maximise cluster differences across the matrix is important for finding clusters which are defined by their own unique variable patterns. In other words, we want to find the best set of unique clusters, where the clusters are different to each other, not a single, aggregate pattern of similarity. If we only work from the lowest pair score, joining cases to this one pair using a hierarchy of ranked lowest scores across the whole matrix, we will find a single aggregate cluster and not maximise our ability to identify the maximum diversity between several clusters. This is why we focus on identifying unique pairs, not a simple hierarchy of the lowest scoring pairs across the whole matrix.

Identifying unique pairs

Start by identifying the most similar pair of cases. This is the pair with the lowest score in the whole matrix in table 2.2. Then continue to search for new, unique pairs to start a new cluster.

The lowest score is for Open Thinking and LKS Data (7.06). This forms the start of the first cluster.

Next is Mini Max and Ashton Algorithms (7.72). As a unique pair, this forms the start of the second cluster.

Next is New Perspectives and Sun Focus (9.83). As a unique pair, this forms the start of the third cluster.

The next most similar pair is Cosign Research and Ashton Algorithms (10.10) Ashton Algorithms is already located in the second cluster, so this is not a unique pair.

The next pair is Visual Research and System Synthesis (14.39). This is a unique a pair, so they start a fourth cluster.

Next is Strategy Statistics with LKS Data (15.75) This is not a unique pair. LKS Data is already assigned to the first cluster.

Next is Linear Logics and System Synthesis (17.91), System Synthesis is already located, and this is not a unique pair.

The final lowest column score is for JB Alpha with Cosign Research (20.26). This is a unique pair. It can be noted at this point that JB Alpha has the highest Squared Euclidean scores with other cases in table 2.2, and so in the next stage of DPS, when variable and case scores are examined in more detail, it will be important to consider whether it should be considered as an outlier.

The unique pairs allocated into clusters are:

1. Cluster 1: Open Thinking and LKS Data
2. Cluster 2: Mini Max and Ashton Algorithms
3. Cluster 3: New Perspectives and Sun Focus
4. Cluster 4: Visual Research and System Synthesis
5. Cluster 5: JB Alpha and Cosign Research

The remaining cases are: Strategy Statistics and Linear Logics.

The remaining cases can be located with their 'best' pairing. So, Strategy Statistics joins cluster 1, and Linear Logics joins cluster 4 (table 2.3)

Table 2.3 Summary of the proposed cluster structure, 2015 data

Cluster 1	Open Thinking, LKS Data, Strategy Statistics
Cluster 2	Mini Max, Ashton Algorithms,
Cluster 3	New Perspectives, Sun Focus
Cluster 4	Visual Research, System Synthesis, Linear Logics
Cluster 5	Cosign Research, JB Alpha

Validating the proposed cluster structure and number of clusters

The next stage is to form a cluster pattern table. This allows the detailed aspects of similarity and difference between the cases to be scrutinised more closely. This allows you, as the researcher, to make a final decision about how many clusters and outliers to validate and include in your model. It is reasonable to use this table validation process to make changes to the clusters proposed in summary table 2.3, providing of course you have evidence to do so.

To do this validation, return to your raw data table for the cases (see the top of table 2.1, before the standardisation process), and sort the data to represent the clusters in table 2.3.

To carry out this sorting and visual analysis, you will need to create a cluster membership variable in a new column (table 2.4), and then use the Excel® custom sort menu to sort by clusters (figure 3.1).

Open your Excel® datafile.

Add a new cluster membership variable in a new column. You can do this by simply creating a new column to the right of the existing variable columns.

After ensuring you have a cluster membership variable in your data table, you can add descriptive statistics in a row immediately beneath the table for all the variables used in the cluster analysis. This will assist you in making easier visual comparisons between clusters and their variable patterns in relation to the central tendency and distribution of the whole dataset.

Compute descriptive statistics

Table 2.4 shows three rows of descriptive statistics added below the data table.

The Excel® formulas used are:

Mean average	=AVERAGE(B2:B13)
Median average	=MEDIAN(B2:B13)
Standard Dev	=STDEV.P(B2:B13)

Note: Standard Dev of the Population(P) is used, as this dataset is not seeking to represent a larger population via a sample.

The results of the descriptive statistics are shown in the bottom three rows of table 2.4

In the next chapter, you will see that by using an Excel® function to 'sort by cluster', you can see the four clusters in alignment. You can also then sort against a variable of choice, to begin to get a sense of which variables are most influencing cluster memberships. The important point with cluster analysis, is that different clusters will be influenced by different variable patterns, rather than there being one variable pattern that predominates across the whole data matrix. Cases are also 'fuzzy' in their relationships with clusters. You will find some cases that sit close to more than one cluster and share some variable similarities with a cluster while being different to it for other variable comparisons.

Table 2.4 Cluster pattern synthesis – adding cluster and descriptive statistics, 2015

	Capexpend2015	AnIncomeGrow2015	PGT2015	Genderpaygap2015	Marketing2015	Managers2015	Overseas2015	continuecustomers2015	debtors2015	staffturnover2015	sicknessdays2015	cluster
JB Alpha	12.3	2.9	72.0	2.0	5.0	0.10	0.0	90.0	2.0	30.0	6.0	5
Cosign Research	11.1	3.0	54.0	3.0	4.3	0.03	6.0	84.0	2.0	15.0	4.0	5
Mini Max	4.5	4.0	32.0	3.0	5.2	0.02	0.0	86.0	3.0	16.0	7.0	2
System Synthesis	9.2	13.7	34.0	7.0	8.1	0.01	12.0	82.0	3.0	13.0	6.0	4
Open Thinking	8.7	15.6	67.0	1.0	4.2	0.05	6.0	100.0	0.5	16.0	5.0	1
LKS Data	3.1	8.9	76.0	1.0	4.0	0.05	5.0	98.0	1.0	8.0	4.0	1
Strategy Statistics	2.1	6.9	90.0	1.0	4.6	0.04	3.0	89.0	1.0	21.0	9.0	1
Visual Research	9.8	20.3	43.0	3.0	5.7	0.05	8.0	84.0	3.0	2.0	7.0	4
Ashton Algorithms	7.1	2.8	56.0	1.0	7.2	0.03	4.0	77.0	3.5	14.0	6.0	2
Linear Logics	7.4	2.3	42.0	8.0	6.1	0.05	23.0	76.0	3.0	9.0	3.0	4
Sun Focus	5.7	7.1	56.0	2.0	3.7	0.04	4.0	69.0	5.0	7.0	4.0	3
New Perspectives	4.7	7.3	45.0	4.0	2.3	0.04	11.0	80.0	3.0	11.0	6.0	3
Mean	*7.1*	*7.9*	*55.6*	*3.0*	*5.0*	*0.04*	*6.8*	*84.6*	*2.5*	*13.5*	*5.6*	
Median	*7.3*	*7.0*	*55.0*	*2.5*	*4.8*	*0.04*	*5.5*	*84.0*	*3.0*	*13.5*	*6.0*	
Standard Deviation	*3.1*	*5.6*	*17.0*	*2.2*	*1.5*	*0.02*	*6.0*	*8.5*	*1.2*	*6.9*	*1.6*	

In the next chapter, we examine how to sort the clusters to better examine the variable patterns for each cluster in relation to the descriptive statistics.

Chapter three: Cluster pattern synthesis: clusters and their variable patterns

In the first edition of this book, we recommended transforming the scale data results from the cluster analysis into binary categories, similar to the process used in crisp set QCA (Ragin, 1987), so as to be able to see how different variable patterns are associated with the membership of different clusters.

In this second edition, we prefer a method that allows the researcher to continue to see the diversity of scale scores while comparing the clusters. For this purpose, we propose the use of pattern synthesis with configurational tables. The scale data is sorted simply into different configurations, to allow the researcher to search for the data patterns that best represent cluster membership.

If you prefer to use the previous binary, categorical approach to cluster comparison, this is available in the first edition of the book, in chapter two.

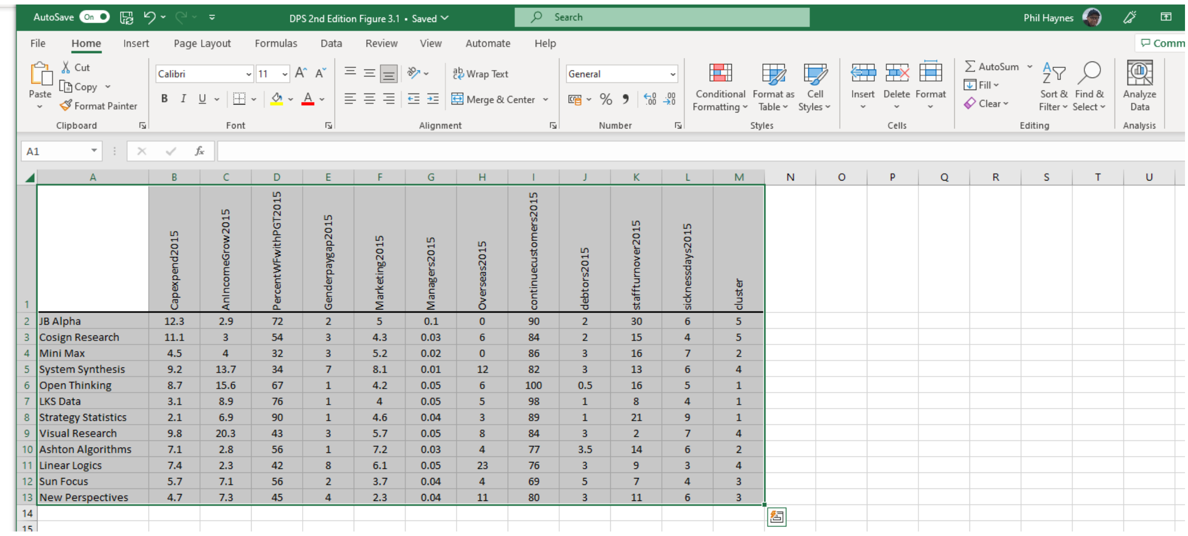
Presenting the data as tables of cluster configurations

Next, we begin to identify the different variable configurations that form each cluster.

For this purpose, we need to sort the Excel® table into cluster groupings.

The Excel® sort menu is used to manipulate the table, to assist the comparative analysis.

First select the data, including the case labels in column A, and the variable names in Row 1, as shown below (figure 3.1).



	Capexpend2015	AnIncomeGrow2015	PercentWFwithPGT2015	Genderpaygap2015	Marketing2015	Managers2015	Overseas2015	continuecustomers2015	debtors2015	staffturnover2015	sicknessdays2015	cluster
JB Alpha	12.3	2.9	72	2	5	0.1	0	90	2	30	6	5
Cosign Research	11.1	3	54	3	4.3	0.03	6	84	2	15	4	5
Mini Max	4.5	4	32	3	5.2	0.02	0	86	3	16	7	2
System Synthesis	9.2	13.7	34	7	8.1	0.01	12	82	3	13	6	4
Open Thinking	8.7	15.6	67	1	4.2	0.05	6	100	0.5	16	5	1
LKS Data	3.1	8.9	76	1	4	0.05	5	98	1	8	4	1
Strategy Statistics	2.1	6.9	90	1	4.6	0.04	3	89	1	21	9	1
Visual Research	9.8	20.3	43	3	5.7	0.05	8	84	3	2	7	4
Ashton Algorithms	7.1	2.8	56	1	7.2	0.03	4	77	3.5	14	6	2
Linear Logics	7.4	2.3	42	8	6.1	0.05	23	76	3	9	3	4
Sun Focus	5.7	7.1	56	2	3.7	0.04	4	69	5	7	4	3
New Perspectives	4.7	7.3	45	4	2.3	0.04	11	80	3	11	6	3

Figure 3.1 Excel® custom sort, by cluster

Then select **Sort & Filter** from the **Home** menu at the top of the screen. When you see the drop-down menu, select **Custom Sort** (figure 3.2).

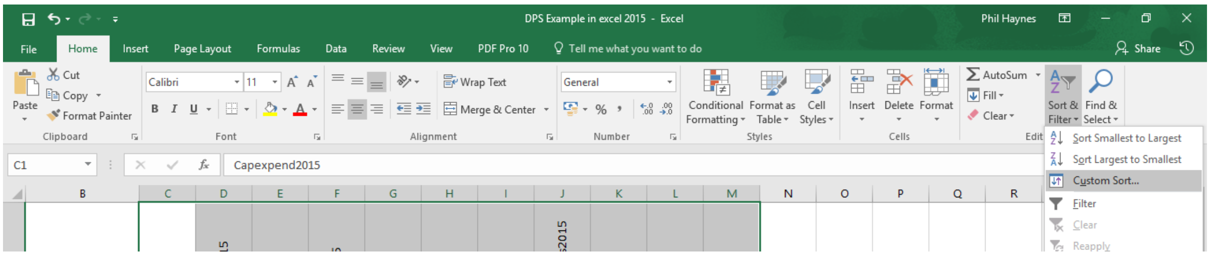


Figure 3.2 Custom sort, drop down menu

The **Custom Sort** sub menu is revealed, as shown below (figure 3.3).

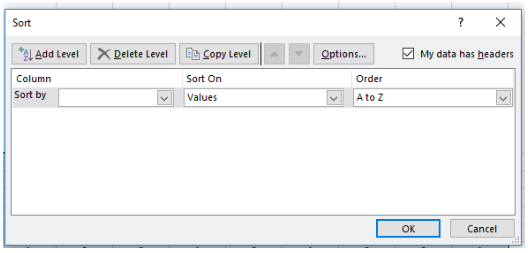


Figure 3.3 Custom sort sub menu

Check that the option **My data has headers** is selected and ticked.

The **Custom Sort** menu allows you to sort by Cluster. The drop-down **Sort by** menu will reveal a list of all the possible variables, including the cluster that the case belongs to.

Use the **Add Level** to Sort by cluster first. Cluster then becomes the first level in your sort.

Keep the default option of **Sort On**: **Values**.

Ensure that the **Order** is **smallest to largest**, so that cluster 1 will be shown at the top of the resulting sort table.

The sort menu is shown below (figure 3.4).

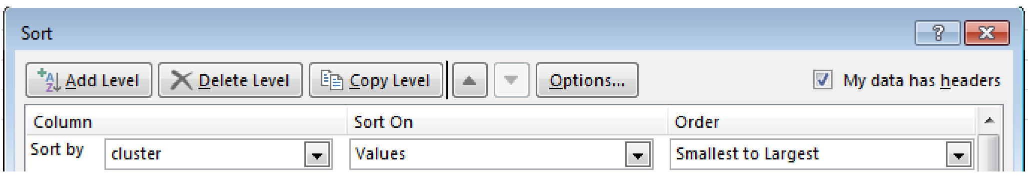


Figure 3.4 Final set up of the custom sort menu

Now click on **OK** and run the sort. The resulting table is shown below (table 3.1).

Table 3.1 Cluster Pattern Synthesis – table sorted by clusters and variables, 2015

	Capexpend2015	AnIncomeGrow2015	PGT2015	Genderpaygap2015	Marketing2015	Managers2015	Overseas2015	continuecustomers2015	debtors2015	staffturnover2015	sicknessdays2015	cluster
Open Thinking	8.7	15.6	67.0	1.0	4.2	0.05	6.0	100.0	0.5	16.0	5.0	1
LKS Data	3.1	8.9	76.0	1.0	4.0	0.05	5.0	98.0	1.0	8.0	4.0	1
Strategy Statistics	8.7	15.6	67.0	1.0	4.2	0.05	6.0	100.0	0.5	16.0	5.0	1
Mini Max	4.5	4.0	32.0	3.0	5.2	0.02	0.0	86.0	3.0	16.0	7.0	2
Ashton Algorithms	7.1	2.8	56.0	1.0	7.2	0.03	4.0	77.0	3.5	14.0	6.0	2
Sun Focus	5.7	7.1	56.0	2.0	3.7	0.04	4.0	69.0	5.0	7.0	4.0	3
New Perspectives	4.7	7.3	45.0	4.0	2.3	0.04	11.0	80.0	3.0	11.0	6.0	3
System Synthesis	9.2	13.7	34.0	7.0	8.1	0.01	12.0	82.0	3.0	13.0	6.0	4
Visual Research	9.8	20.3	43.0	3.0	5.7	0.05	8.0	84.0	3.0	2.0	7.0	4
Linear Logics	7.4	2.3	42.0	8.0	6.1	0.05	23.0	76.0	3.0	9.0	3.0	4
JB Alpha	12.3	2.9	72.0	2.0	5.0	0.10	0.0	90.0	2.0	30.0	6.0	5
Cosign Research	11.1	3.0	54.0	3.0	4.3	0.03	6.0	84.0	2.0	15.0	4.0	5
Mean	*7.1*	*7.9*	*55.6*	*3.0*	*5.0*	*0.0*	*6.8*	*84.6*	*2.5*	*13.5*	*5.6*	
Median	*7.3*	*7.0*	*55.0*	*2.5*	*4.8*	*0.0*	*5.5*	*84.0*	*3.0*	*13.5*	*6.0*	
Standard Deviation	*3.1*	*5.6*	*17.0*	*2.2*	*1.5*	*0.0*	*6.0*	*8.5*	*1.2*	*6.9*	*1.6*	

In order to see clearly the resulting cluster data patterns, it is necessary to identify clusters that consistently share above and below average variable scores, or other similar distributional aspects of variables.

The easiest method for comparing visually cluster variable patterns is to use:

Conditional Formating/ Color Scales

As shown in figure 3.5

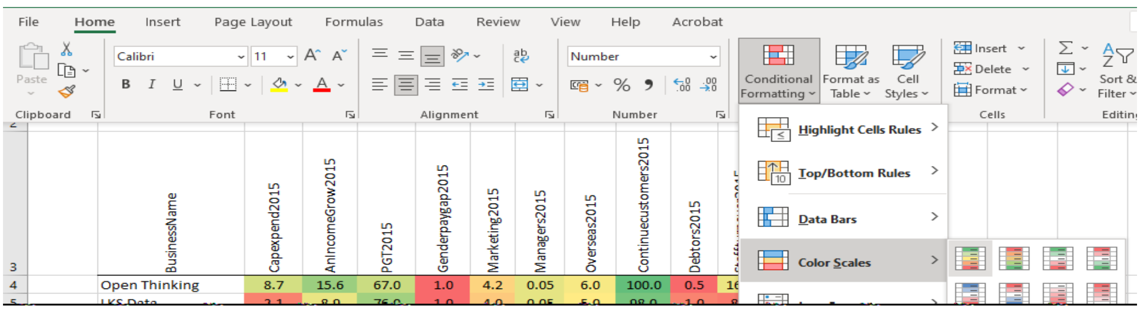


Figure 3.5 Conditional formatting with colour scales

Select a single variable column and apply the colour scaling. The above example, in figure 3.5, uses the first drop down option, where green represents higher scores and red represents lower. Scores around the central tendency (close to the mean/median) are shown in yellow.

Work your way along each column, applying the scales to each column in turn. The results are shown in table 3.2.

Do not attempt to copy and paste the colour scale rule to the whole data array in a single operation, as this will lead to a comparison of the whole array, not each variable. Hence, you need to work systematically in individual columns.

Table 3.2 Cluster patterns, as demonstrated with colour scales, 2015

	Capexpend2015	AnIncomeGrow2015	PGT2015	Genderpaygap2015	Marketing2015	Managers2015	Overseas2015	Continuecustomers2015	Debtors2015	Staffturnover2015	Sicknessdays2015	Cluster
Open Thinking	8.7	15.6	67.0	1.0	4.2	0.05	6.0	100.0	0.5	16.0	5.0	1
LKS Data	3.1	8.9	76.0	1.0	4.0	0.05	5.0	98.0	1.0	8.0	4.0	1
Strategy Statistics	2.1	6.9	90.0	1.0	4.6	0.04	3.0	89.0	1.0	21.0	9.0	1
Mini Max	4.5	4.0	32.0	3.0	5.2	0.02	0.0	86.0	3.0	16.0	7.0	2
Ashton Algorithms	7.1	2.8	56.0	1.0	7.2	0.03	4.0	77.0	3.5	14.0	6.0	2
Sun Focus	5.7	7.1	56.0	2.0	3.7	0.04	4.0	69.0	5.0	7.0	4.0	3
New Perspectives	4.7	7.3	45.0	4.0	2.3	0.04	11.0	80.0	3.0	11.0	6.0	3
System Synthesis	9.2	13.7	34.0	7.0	8.1	0.01	12.0	82.0	3.0	13.0	6.0	4
Visual Research	9.8	20.3	43.0	3.0	5.7	0.05	8.0	84.0	3.0	2.0	7.0	4
Linear Logics	7.4	2.3	42.0	8.0	6.1	0.05	23.0	76.0	3.0	9.0	3.0	4
JB Alpha	12.3	2.9	72.0	2.0	5.0	0.10	0.0	90.0	2.0	30.0	6.0	5
Cosign Research	11.1	3.0	54.0	3.0	4.3	0.03	6.0	84.0	2.0	15.0	4.0	5
Median	*7.3*	*7.0*	*55.0*	*2.5*	*4.8*	*0.04*	*5.5*	*84.0*	*3.0*	*13.5*	*6.0*	
Mean	*7.1*	*7.9*	*55.6*	*3.0*	*5.0*	*0.04*	*6.8*	*84.6*	*2.5*	*13.5*	*5.6*	
Standard Deviation	*3.1*	*5.6*	*17.0*	*2.2*	*1.5*	*0.02*	*6.0*	*8.5*	*1.2*	*6.9*	*1.6*	

Table 3.2 reveals evidence of variable patterns that do match cluster memberships.

For example, in cluster 1, there are shared green (higher) scores for PGT and continuing customers. There are shared red scores (lower scores) for gender pay gap and debtors.

There are other patterns of similarity around the central tendency. Marketing is close to, but marginally below, the central tendency. Managers is close to, but marginally above, the central tendency.

If you want to use a similar method of pattern analysis, that uses monochrome printing instead of colour, a method to achieve this is to shade the cluster scores that share similar scores using the Excel® Home/Format/Format Cells - options. (This is often best done using the right mouse button options.)

Experiment with these options and you will find you can change the cell colour and text colour of a group of cells. This allows you to create patterns that are monochrome rather than colour based.

You can use the Excel® cell formatting menu to devise your own shading presentation.

For example, in table 3.3, clusters that share above mean average scores are shared in black with white text and clusters that share below mean average scores are shared in light grey with black text. This does not show the graded detail permitted in the colour table, but provides a nice summary of the key variable relationships.

Table 3.3 shows the results of a monochrome analysis.

Table 3.3 Cluster Pattern Synthesis – using monochrome shading to identify patterns, 2015

	Capexpend2015	AnIncomeGrow2015	PGT2015	Genderpaygap2015	Marketing2015	Managers2015	Overseas2015	continuecustomers2015	debtors2015	staffturnover2015	sicknessdays2015	cluster
Open Thinking	8.7	15.6	67.0	1.0	4.2	0.05	6.0	100.0	0.5	16.0	5.0	1
LKS Data	3.1	8.9	76.0	1.0	4.0	0.05	5.0	98.0	1.0	8.0	4.0	1
Strategy Statistics	2.1	6.9	90.0	1.0	4.6	0.04	3.0	89.0	1.0	21.0	9.0	1
Mini Max	4.5	4.0	32.0	3.0	5.2	0.02	0.0	86.0	3.0	16.0	7.0	2
Ashton Algorithms	7.1	2.8	56.0	1.0	7.2	0.03	4.0	77.0	3.5	14.0	6.0	2
Sun Focus	5.7	7.1	56.0	2.0	3.7	0.04	4.0	69.0	5.0	7.0	4.0	3
New Perspectives	4.7	7.3	45.0	4.0	2.3	0.04	11.0	80.0	3.0	11.0	6.0	3
System Synthesis	9.2	13.7	34.0	7.0	8.1	0.01	12.0	82.0	3.0	13.0	6.0	4
Visual Research	9.8	20.3	43.0	3.0	5.7	0.05	8.0	84.0	3.0	2.0	7.0	4
Linear Logics	7.4	2.3	42.0	8.0	6.1	0.05	23.0	76.0	3.0	9.0	3.0	4
JB Alpha	12.3	2.9	72.0	2.0	5.0	0.10	0.0	90.0	2.0	30.0	6.0	5
Cosign Research	11.1	3.0	54.0	3.0	4.3	0.03	6.0	84.0	2.0	15.0	4.0	5
Mean	*7.1*	*7.9*	*55.6*	*3.0*	*5.0*	*0.04*	*6.8*	*84.6*	*2.5*	*13.5*	*5.6*	
Median	*7.3*	*7.0*	*55.0*	*2.5*	*4.8*	*0.04*	*5.5*	*84.0*	*3.0*	*13.5*	*6.0*	
Standard Deviation	*3.1*	*5.6*	*17.0*	*2.2*	*1.5*	*0.02*	*6.0*	*8.5*	*1.2*	*6.9*	*1.6*	

Tables 3.2 and 3.3 allow us to see the homogeneity and heterogeneity of the clusters. The five cluster structure provides some homogeneity for each of the individual clusters.

Smaller clusters (pairs) are likely to be more homogeneous in their variable patterns.

Near misses and fuzzy relationships

It is important to appreciate the 'qualitative interpretation' of these pattern synthesis tables when using the monochrome summary. There are often mathematical 'near misses', that still indicate a degree of similarity between cases. If you are using a more sophisticated clustering computer program, it will offer a wide variety of algorithms to choose from, to pick the cluster patterns. Different algorithms will produce different patterns, not the same patterns. Of course, where cases are strongly related with each other, they are likely to remain in the same cluster even when you change the computer algorithm, but in many situations the relationships between cases are more 'fuzzy' and therefore cases may move clusters depending on the algorithm used. This reflects the reality of the real-world relationship between complex cases. Cases can be similar in some respects and different in others. Nevertheless, the advantage of the simple mathematical approach to developing cluster analysis taught in this book, with the use of exploration through tables, is that you as the researcher remain in control of the process of exploration and you can clearly see where the complexity of case relationships is at its most fuzzy.

Firstly, the researcher needs to decide how to pattern scores at the mean. For example, scores identical to the mean average in table 3.3 can potentially be interpreted as similar to other cases at both above and below mean scores. A specific example is the variable gender pay gap, for cluster 4. Here, one score for a case in the cluster is at the mean (3.0). The other two scores are above the mean. The researcher has to decide whether to shade the cell area for the cluster as representing similar scores that are in general terms higher representations of the variable.

Likewise, some larger clusters may have one variable observation that is dissimilar to the other cases in their cluster, but where all other cases share similar scores. In the first edition of this book, these situations were defined as 'near misses' and worthy of close examination. Such decisions might be based on how marginal the difference of the one score is. For example, if a score is below the mean but above the median, it is close to the central tendency and it seems reasonable to argue it is a near miss. The assessment of near misses is likely to be especially important for larger clusters, where the cluster is relatively large compared to the total size of the case population.

Finally, the approach to near misses should be dependent on the modelling requirement of the researcher and the degree of tolerance for fuzzy scores etc. The important point is that the researcher should be consistent and transparent about how they develop such models and include reflections about their cluster decisions in their methods write up.

Cluster boundaries and outliers

The use of cluster pattern synthesis tables enables the researcher to make a final and informed decision about what separation of clustering to focus on in their data analysis. When using tables to explore the relationships in detail, the researcher is avoiding depending on cluster algorithms alone.

For example, when first constructing the clusters, it was noted that JB Alpha was a potential outlier with a series of relatively high Squared Euclidean difference scores that suggested it might be rather separate from other cases. The pattern synthesis tables 3.2 and 3.3 allows this to be examined with more scrutiny.

There does not seem to be a conclusive argument for removing JB Alpha from the cluster 5 pairing on the basis of the overall evidence. It shows some variable similarities with its pair: Cosign Research. This validation illustrates why the use of detailed tables is important to understand the precise relationship of variables with clusters, rather than relying solely on the Squared Euclidean difference aggregations.

Similarly, if you are interested in less homogeneous clusters, that are linked to only on a few variables, you might also check at this point in the DPS whether you want to argue to combine any clusters because they have aspects of cluster similarity. For example, in tables 3.2 and 3.3, cluster 1 and 2 share some variable score patterns of similarity, and likewise with clusters 3 and 4. Nevertheless, on balance, there are also some key differences that make all the clusters different from each other.

In this situation, when examining the 2015 data, the evidence is that the five cluster structure is valid.

Having confirmed the cluster structure, it is also possible to explore causality, with regard to any single outcome variable. For example, if we are interested in why some businesses have more debtors than others, we can make debtor an outcome variable (table 3.4).

Reconstituting the table to include an outcome variable

The default approach with DPS is to use the number of clusters as the outcome variable in the pattern synthesis tables. This enables the variable patterns to be checked against case clusters, to ensure a logical and realistic number of clusters is chosen. The clusters are validated by the observation of shared variable scoring patterns.

Once the researcher has decided on the optimum number of clusters that is useful, it is then possible to reorganise the table so that you can consider a different outcome variable, as an alternative to focusing on cluster validation alone. First, reopen the **Custom Sort** menu and ensure that the chosen outcome variable (debtors) is the first level of the new sort. Make cluster the second sort.

You can also add and order the sort of the other variables, if you wish. For example, one logic that can be used is to put the other variables that are most likely to influence the outcome, higher up the sorting hierarchy.

Having performed the sort, table 3.4 shows the results. This table also requires the outcome variable, debtors2015, to be cut and pasted into the last column of the table, for ease of presentation.

Table 3.4 Clusters resorted against a single outcome variable, 2015, debtors

	Capexpend2015	AnIncomeGrow2015	PGT2015	Genderpaygap2015	Marketing2015	Managers2015	Overseas2015	continuecustomers2015	staffturnover2015	sicknessdays2015	cluster	debtors2015
Open Thinking	8.7	15.6	67.0	1.0	4.2	0.05	6.0	100.0	16.0	5.0	**1**	0.5
Strategy Statistics	2.1	6.9	90.0	1.0	4.6	0.04	3.0	89.0	21.0	9.0	**1**	1.0
LKS Data	3.1	8.9	76.0	1.0	4.0	0.05	5.0	98.0	8.0	4.0	**1**	1.0
JB Alpha	12.3	2.9	72.0	2.0	5.0	0.10	0.0	90.0	30.0	6.0	**5**	2.0
Cosign Research	11.1	3.0	54.0	3.0	4.3	0.03	6.0	84.0	15.0	4.0	**5**	2.0
Mini Max	4.5	4.0	32.0	3.0	5.2	0.02	0.0	86.0	16.0	7.0	2	3.0
New Perspectives	4.7	7.3	45.0	4.0	2.3	0.04	11.0	80.0	11.0	6.0	3	3.0
System Synthesis	9.2	13.7	34.0	7.0	8.1	0.01	12.0	82.0	13.0	6.0	**4**	3.0
Linear Logics	7.4	2.3	42.0	8.0	6.1	0.05	23.0	76.0	9.0	3.0	**4**	3.0
Visual Research	9.8	20.3	43.0	3.0	5.7	0.05	8.0	84.0	2.0	7.0	**4**	3.0
Ashton Algorithms	7.1	2.8	56.0	1.0	7.2	0.03	4.0	77.0	14.0	6.0	2	3.5
Sun Focus	5.7	7.1	56.0	2.0	3.7	0.04	4.0	69.0	7.0	4.0	3	5.0
Mean	*7.1*	*7.9*	*55.6*	*3.0*	*5.0*	*0.04*	*6.8*	*84.6*	*13.5*	*5.6*		*2.5*
Median	*7.3*	*7.0*	*55.0*	*2.5*	*4.8*	*0.04*	*5.5*	*84.0*	*13.5*	*6.0*		*3.0*
Standard Deviation	*3.1*	*5.6*	*17.0*	*2.2*	*1.5*	*0.02*	*6.0*	*8.5*	*6.9*	*1.6*		*1.2*

We can now view patterns in relation to the specified outcome. We can see which clusters are most likely to have variable patterns that are partially associated with the outcome, and clusters that do not. The configurations are fuzzy and some clusters are dispersed against the outcome (clusters 2 and 3). This suggests little association with the outcome for those clusters. The configurations between the outcome, clusters and variables are diverse, non-symmetrical, and need a qualitative judgement from the researcher regarding their interpretation rather than the simple application of a mathematical rule.

For cases with a lower number of debtors, table 3.4 shows some evidence that clusters 1 and 5 match below mean scores for debtor levels. Cluster 1 shares above average PGT qualifications and continuing customers. It has shared below mean scores for gender pay gap and marketing. Cluster 5 has above average capital expenditure and staff turnover, and below average scores for annual income grow, gender pay gap and overseas customers.

In terms of 'near misses', cluster 5 scores for PGT qualifications and continuing customers are close to the mean and above, indicating a degree of similarity with cluster 1.

Cluster 4 shares slightly above average scores for debtors. It shares above average scores for capital expenditure, gender pay gap, marketing, and overseas orders. The level of PGT qualifications, continuing customers and staff turnover, is consistently below average.

Three of the variables across clusters 1, 5 and 4 show a partial symmetry with debtors: PGT qualifications, gender pay gap, overseas, and marketing.

This approach to considering one variable as an outcome illustrates the sensitivity of the method to each individual case.

There is the potential for examining just one case in detail compared to all others, and groups of others, if the focus of your research is from the perspective of one organisation. This flexibility and sensitivity to complex differences is the major advantage of configurational methods. Likewise, some researchers are interested in why one or two cases stand out as being exceptional and rather different to others, and this exceptionalism might become the legitimate focus of the research project. All these approaches are likely to benefit from 'mixed methods', where this type of quantitative approach is followed up with other qualitative research.

Chapter four: Repeating the DPS with longitudinal data

Having developed a pattern synthesis for 2015, the next step is to repeat the pattern analysis with 2016, and then 2017, data.

This chapter shows the key results for 2016 and 2017, but does not repeat the detail about how to use Excel® to carry out the analysis. You may need to refer back to chapter three to recall the detail of how to make the calculations with Excel®.

Data synthesis for 2016

You can begin the process of a simple cluster analysis for the 2016 data using table 4.1.

Start by identifying the most similar pair of cases. This is the pair with the lowest score in the whole matrix. Then continue to search for new, unique pairs to start a new cluster.

The lowest score is for New Perspectives and Mini Max (10.01). This forms the start of the first cluster.

Next is Mini Max and Cosign Research (10.95). It is not a unique pair, as Mini Max is already located with New Perspectives. So, this pair is held for later consideration.

Next is Cosign Research and LKS Data (11.38). This is a unique pair, and forms the start of the second cluster.

The next most similar pair is Cosign Research and Open Thinking (12.30) Cosign Research has already been located in the second cluster, so this is not a unique pair.

The next pair is Linear Logics and New Perspectives (12.80). New Perspectives has already been allocated to cluster 1, so this is not a unique pair

Next is Sun Focus and Linear Logics (12.84) This is a unique pair as neither have yet been allocated to a cluster. This forms cluster 3

Next is Visual Research and Mini Max(13.52), Mini Max is already located and this is not a unique pair.

Next is Strategy Statistics and LKS Data (14.09) LKS Data is already allocated to a cluster and so this is not a unique pair.

Next is Ashton Algorithms and Visual Research (15.33). This is a unique pair, as neither have been previously allocated to a cluster. This is cluster 4.

Next is Strategy Statistics and JB Alpha (19.23) this is a unique pair, as neither have been previously allocated to a cluster/ This forms cluster 5.

Table 4.1 Cluster construction data for 2016, applying the Squared Euclidean differences

	Capexpend2016	AnIncomeGrow2016	PGT2016	Genderpaygap2016	Marketing2016	Managers2016	Overseas2016	Continuecustomers2016	Debtors2016	Staffturnover2016	Sicknessdays2016
						Raw scores					
JB Alpha	14.6	3.1	70.0	1.0	4.1	0.09	0.0	92.0	2.0	15.0	5.0
Cosign Research	11.0	4.2	55.0	3.0	5.2	0.03	4.0	90.0	2.0	7.0	3.0
Mini Max	5.5	6.7	49.0	2.0	4.9	0.03	2.0	87.0	3.0	3.0	6.0
System Synthesis	8.4	15.0	40.0	6.0	7.2	0.02	10.0	80.0	4.0	7.0	6.0
Open Thinking	8.3	3.0	65.0	2.0	6.4	0.06	5.0	98.0	1.0	4.0	6.0
LKS Data	4.1	4.5	75.0	2.0	5.2	0.04	2.0	97.0	1.0	7.0	2.0
Strategy Statistics	4.5	5.0	85.0	0.0	5.2	0.05	3.0	92.0	2.0	17.0	5.0
Visual Research	10.2	-5.6	45.0	4.0	5.5	0.04	9.0	87.0	3.0	6.0	7.0
Ashton Algorithms	8.1	-1.0	58.0	2.0	9.0	0.03	4.0	85.0	4.0	9.0	5.0
Linear Logics	8.2	0.6	51.0	3.0	7.1	0.06	20.0	78.0	2.0	5.0	4.0
Sun Focus	6.3	-2.3	59.0	2.0	5.1	0.03	8.0	72.0	2.5	2.0	3.0
New Perspectives	4.6	7.1	52.0	3.0	4.1	0.05	14.0	84.0	2.0	9.0	6.0
Mean	*7.8*	*3.4*	*58.7*	*2.5*	*5.8*	*0.04*	*6.8*	*86.8*	*2.4*	*7.6*	*4.8*
Standard Deviation	*3.0*	*5.0*	*12.5*	*1.4*	*1.4*	*0.02*	*5.5*	*7.3*	*0.9*	*4.3*	*1.5*
						Standardized scores					
JB Alpha	2.29	-0.05	0.91	-1.04	-1.21	2.43	-1.22	0.70	-0.40	1.72	0.11
Cosign Research	1.07	0.17	-0.29	0.35	-0.40	-0.72	-0.50	0.43	-0.40	-0.14	-1.25
Mini Max	-0.78	0.66	-0.77	-0.35	-0.62	-0.72	-0.86	0.02	0.67	-1.06	0.80
System Synthesis	0.20	2.31	-1.49	2.42	1.06	-1.51	0.59	-0.93	1.73	-0.14	0.80
Open Thinking	0.16	-0.07	0.51	-0.35	0.48	0.85	-0.32	1.52	-1.47	-0.83	0.80
LKS Data	-1.25	0.23	1.31	-0.35	-0.40	-0.20	-0.86	1.38	-1.47	-0.14	-1.94
Strategy Statistics	-1.12	0.33	2.11	-1.73	-0.40	0.33	-0.68	0.70	-0.40	2.18	0.11
Visual Research	0.80	-1.78	-1.09	1.04	-0.18	-0.20	0.41	0.02	0.67	-0.37	1.48
Ashton Algorithms	0.10	-0.87	-0.05	-0.35	2.38	-0.72	-0.50	-0.25	1.73	0.33	0.11
Linear Logics	0.13	-0.55	-0.61	0.35	0.99	0.85	2.39	-1.20	-0.40	-0.60	-0.57
Sun Focus	-0.51	-1.12	0.03	-0.35	-0.48	-0.72	0.23	-2.02	0.13	-1.30	-1.25
New Perspectives	-1.08	0.74	-0.53	0.35	-1.20	0.39	1.31	-0.39	-0.40	0.33	0.80

Squared Euclidean Distance

	JB Alpha	Cosign Research	Mini Max	System Synthesis	Open Thinking	LKS Data	Strategy Statistics	Visual Research	Ashton Algorithms	Linear Logics	Sun Focus	New Perspectives
JB Alpha		21.36	33.44	62.73	20.06	30.18	19.23	31.91	37.52	38.94	41.44	30.39
Cosign Research	21.36		10.95	25.66	**12.30**	**11.38**	23.38	14.95	17.36	17.61	12.92	15.15
Mini Max	33.44	**10.95**		**20.31**	13.88	19.80	24.24	**13.52**	16.35	23.15	13.81	**10.01**
System Synthesis	62.73	25.66	20.31		40.80	50.97	56.08	25.04	24.52	28.82	34.69	23.43
Open Thinking	20.06	12.30	13.88	40.80		12.88	18.82	17.33	22.24	20.03	24.97	15.64
LKS Data	30.18	11.38	19.80	50.97	12.88		**14.09**	35.72	30.30	30.18	21.42	21.74
Strategy Statistics	**19.23**	23.38	24.24	56.08	18.82	14.09		37.53	27.22	37.51	32.32	21.15
Visual Research	31.91	14.95	13.52	25.04	17.33	35.72	37.53		**15.53**	15.99	18.52	15.11
Ashton Algorithms	37.52	17.36	16.35	24.52	22.24	30.30	27.22	15.53		20.43	19.28	26.90
Linear Logics	38.94	17.61	23.15	28.82	20.03	30.18	37.51	15.99	20.43		**12.84**	12.80
Sun Focus	41.44	12.92	13.81	34.69	24.97	21.42	32.32	18.52	19.28	12.84		17.21
New Perspectives	30.39	15.15	**10.01**	23.43	15.64	21.74	21.15	15.11	26.90	**12.80**	17.21	
Minimum distance	19.23	10.95	10.01	20.31	12.30	11.38	14.09	13.52	15.53	12.80	12.84	10.01

The unique pairs allocated into clusters are:

1. New Perspectives and Mini Max
2. Cosign Research and LKS Data
3. Linear Logics and New Perspectives
4. Ashton Algorithms and Visual Research
5. Strategy Statistics and JB Alpha

The remaining cases are: System Synthesis and Open Thinking.

These can be located with their best pairing. So, System Synthesis joins cluster 1, and Open Thinking joins cluster 2 (table 4.2).

Table 4.2 shows the initial allocation of clusters for 2016 data based on the identification of unique pairs from the Squared Euclidean distance data. Descriptive statistics have also been added.

Table 4.2 Cluster pattern synthesis – clusters and descriptive statistics, 2016

	Capexpend2016	AnIncomeGrow2016	PGT2016	Genderpaygap2016	Marketing2016	Managers2016	Overseas2016	Continuecustomers2016	Debtors2016	Staffturnover2016	Sicknessdays2016	Cluster
Mini Max	5.5	6.7	49.0	2.0	4.9	0.03	2.0	87.0	3.0	3.0	6.0	1
System Synthesis	8.4	15.0	40.0	6.0	7.2	0.02	10.0	80.0	4.0	7.0	6.0	1
New Perspectives	4.6	7.1	52.0	3.0	4.1	0.05	14.0	84.0	2.0	9.0	6.0	1
Cosign Research	11.0	4.2	55.0	3.0	5.2	0.03	4.0	90.0	2.0	7.0	3.0	2
Open Thinking	8.3	3.0	65.0	2.0	6.4	0.06	5.0	98.0	1.0	4.0	6.0	2
LKS Data	4.1	4.5	75.0	2.0	5.2	0.04	2.0	97.0	1.0	7.0	2.0	2
Linear Logics	8.2	0.6	51.0	3.0	7.1	0.06	20.0	78.0	2.0	5.0	4.0	3
Sun Focus	6.3	-2.3	59.0	2.0	5.1	0.03	8.0	72.0	2.5	2.0	3.0	3
Visual Research	10.2	-5.6	45.0	4.0	5.5	0.04	9.0	87.0	3.0	6.0	7.0	4
Ashton Algorithms	8.1	-1.0	58.0	2.0	9.0	0.03	4.0	85.0	4.0	9.0	5.0	4
JB Alpha	14.6	3.1	70.0	1.0	4.1	0.09	0.0	92.0	2.0	15.0	5.0	5
Strategy Statistics	4.5	5.0	85.0	0.0	5.2	0.05	3.0	92.0	2.0	17.0	5.0	5
Mean	*7.8*	*3.4*	*58.7*	*2.5*	*5.8*	*0.04*	*6.8*	*86.8*	*2.4*	*7.6*	*4.8*	
Median	*8.2*	*3.7*	*56.5*	*2.0*	*5.2*	*0.04*	*4.5*	*87.0*	*2.0*	*7.0*	*5.0*	
Standard Deviation	*3.0*	*5.0*	*12.5*	*1.4*	*1.4*	*0.02*	*5.5*	*7.3*	*0.9*	*4.3*	*1.5*	

Table 4.3 validates the clusters by exploring patterns of variable similarity. The heat map shows some evidence that the clusters do relate to specific variable patterns and with different variable patterns validating the five clusters.

Table 4.4 simplifies the pattern synthesis by using the monochrome presentation introduced in the previous chapter. The monochrome shading shows clusters that share above and below mean scores for specific variables, following the same method taught in chapter three.

Cluster 1 has above mean scores for annual income growth and sickness, and below mean scores for postgraduate qualifications.

Clusters 2 and 3 show different patterns, despite some overlap, like shared below mean scores for annual income growth and staff turnover. There are other variable patterns that make the two clusters different from each other and support retaining them as two separate clusters (in particular, the contrasting pattern for continuing customers).

Clusters 4 and 5 share above mean average sickness, but they have different characteristics on lots of other variables. In summary, there is amble evidence that the five cluster structure is valid and demonstrates discrete clusters.

Table 4.3 Cluster pattern synthesis – using variable colour coding, 2016

	Capexpend2016	AnIncomeGrow2016	PGT2016	Genderpaygap2016	Marketing2016	Managers2016	Overseas2016	Continuecustomers2016	Debtors2016	Staffturnover2016	Sicknessdays2016	Cluster
Mini Max	5.5	6.7	49	2	4.9	0.03	2	87	3	3	6	1
System Synthesis	8.4	15	40	6	7.2	0.02	10	80	4	7	6	1
New Perspectives	4.6	7.1	52	3	4.1	0.05	14	84	2	9	6	1
Cosign Research	11	4.2	55	3	5.2	0.03	4	90	2	7	3	2
Open Thinking	8.3	3	65	2	6.4	0.06	5	98	1	4	6	2
LKS Data	4.1	4.5	75	2	5.2	0.04	2	97	1	7	2	2
Linear Logics	8.2	0.6	51	3	7.1	0.06	20	78	2	5	4	3
Sun Focus	6.3	-2.3	59	2	5.1	0.03	8	72	2.5	2	3	3
Visual Research	10.2	-5.6	45	4	5.5	0.04	9	87	3	6	7	4
Ashton Algorithms	8.1	-1	58	2	9	0.03	4	85	4	9	5	4
JB Alpha	14.6	3.1	70	1	4.1	0.09	0	92	2	15	5	5
Strategy Statistics	4.5	5	85	0	5.2	0.05	3	92	2	17	5	5
Mean	*7.8*	*3.4*	*58.7*	*2.5*	*5.8*	*0.04*	*6.8*	*86.8*	*2.4*	*7.6*	*4.8*	
Median	*8.2*	*3.7*	*56.5*	*2*	*5.2*	*0.04*	*4.5*	*87*	*2*	*7*	*5*	
Standard Deviation	*3*	*5*	*12.5*	*1.4*	*1.4*	*0.02*	*5.5*	*7.3*	*0.9*	*4.3*	*1.5*	

Table 4.4 Cluster pattern synthesis – using monochrome shading to summarise clusters, 2016

	Capexpend2016	AnIncomeGrow2016	PGT2016	Genderpaygap2016	Marketing2016	Managers2016	Overseas2016	Continuecustomers2016	Debtors2016	Staffturnover2016	Sicknessdays2016	Cluster
Mini Max	5.5	6.7	49.0	2.0	4.9	0.03	2.0	87.0	3.0	3.0	6.0	1
New Perspectives	4.6	7.1	52.0	3.0	4.1	0.05	14.0	84.0	2.0	9.0	6.0	1
System Synthesis	8.4	15.0	40.0	6.0	7.2	0.02	10.0	80.0	4.0	7.0	6.0	1
Cosign Research	11.0	4.2	55.0	3.0	5.2	0.03	4.0	90.0	2.0	7.0	3.0	2
Open Thinking	8.3	3.0	65.0	2.0	6.4	0.06	5.0	98.0	1.0	4.0	6.0	2
LKS Data	4.1	4.5	75.0	2.0	5.2	0.04	2.0	97.0	1.0	7.0	2.0	2
Linear Logics	8.2	0.6	51.0	3.0	7.1	0.06	20.0	78.0	2.0	5.0	4.0	3
Sun Focus	6.3	-2.3	59.0	2.0	5.1	0.03	8.0	72.0	2.5	2.0	3.0	3
Visual Research	10.2	-5.6	45.0	4.0	5.5	0.04	9.0	87.0	3.0	6.0	7.0	4
Ashton Algorithms	8.1	-1.0	58.0	2.0	9.0	0.03	4.0	85.0	4.0	9.0	5.0	4
JB Alpha	14.6	3.1	70.0	1.0	4.1	0.09	0.0	92.0	2.0	15.0	5.0	5
Strategy Statistics	4.5	5.0	85.0	0.0	5.2	0.05	3.0	92.0	2.0	17.0	5.0	5
Mean	*7.8*	*3.4*	*58.7*	*2.5*	*5.8*	*0.04*	*6.8*	*86.8*	*2.4*	*7.6*	*4.8*	
Median	*8.2*	*3.7*	*56.5*	*2.0*	*5.2*	*0.04*	*4.5*	*87.0*	*2.0*	*7.0*	*5.0*	
Standard Deviation	*3.0*	*5.0*	*12.5*	*1.4*	*1.4*	*0.02*	*5.5*	*7.3*	*0.9*	*4.3*	*1.5*	

Table 4.5 Clusters resorted to show a single outcome variable, debtors, 2016

	Capexpend2016	AnIncomeGrow2016	PGT2016	Genderpaygap2016	Marketing2016	Managers2016	Overseas2016	Continuecustomers2016	Staffturnover2016	Sicknessdays2016	Cluster	Debtors2016
LKS Data	4.1	4.5	75.0	2.0	5.2	0.04	2.0	97.0	7.0	2.0	2	1.0
Open Thinking	8.3	3.0	65.0	2.0	6.4	0.06	5.0	98.0	4.0	6.0	2	1.0
Cosign Research	11.0	4.2	55.0	3.0	5.2	0.03	4.0	90.0	7.0	3.0	2	2.0
JB Alpha	14.6	3.1	70.0	1.0	4.1	0.09	0.0	92.0	15.0	5.0	5	2.0
Strategy Statistics	4.5	5.0	85.0	0.0	5.2	0.05	3.0	92.0	17.0	5.0	5	2.0
New Perspectives	4.6		52.0	3.0	4.1	0.05	14.0	84.0	9.0		1	2.0
Linear Logics	8.2	0.6	51.0	3.0	7.1	0.06	20.0	78.0	5.0	4.0	3	2.0
Sun Focus	6.3	-2.3	59.0	2.0	5.1	0.03	8.0	72.0	2.0	3.0	3	2.5
Mini Max	5.5		49.0	2.0	4.9	0.03	2.0	87.0	3.0		1	3.0
Visual Research	10.2	-5.6	45.0	4.0	5.5	0.04	9.0	87.0	6.0	7.0	4	3.0
Ashton Algorithms	8.1	-1.0	58.0	2.0	9.0	0.03	4.0	85.0	9.0	5.0	4	4.0
System Synthesis	8.4	15.0	40.0	6.0	7.2	0.02	10.0	80.0	7.0	6.0	1	4.0
Mean	*7.8*	*3.4*	*58.7*	*2.5*	*5.8*	*0.04*	*6.8*	*86.8*	*7.6*	*4.8*		*2.4*
Median	*8.2*	*3.7*	*56.5*	*2.0*	*5.2*	*0.04*	*4.5*	*87.0*	*7.0*	*5.0*		*2.0*
Standard Deviation	*3.0*	*5.0*	*12.5*	*1.4*	*1.4*	*0.02*	*5.5*	*7.3*	*4.3*	*1.5*		*0.9*

Table 4.5 shows a new format for the cluster data and is a re-sort of the 2016 data to allow a visual focus on the outcome of levels of debtors. The debtors variable is now in the final column, as with the similar 2015 example in chapter three.

There is a fair degree of congruence between the clusters and this outcome variable, with only cluster 1 being fragmented and dispersed when debtors is prioritised as the first variable in the sort.

For clusters 2 and 5, the lower levels of debtors are associated with lower overseas orders, and higher continuing customers. Cluster 5 also shares other variable similarities (as previously identified in this chapter when analysing table 4.3 and 4.4). Clusters 2 and 5 have differing scores for staff turnover when compared with their low level of debtors.

Cluster 3 has debtor levels that reflect the central tendency of the distribution. It does not share the other variable patterns shared by clusters 2 and 5.

Cluster 4 has the highest debtor scores in the range. In contrast to the shared cluster variable patterns mentioned above it has higher capital expenditure, and lower scores for annual income growth, postgraduate qualifications, and the ratio of managers. Like cluster 5, it has higher sickness scores.

Data synthesis for 2017

Next the same approach to exploring and explaining cluster patterns is applied to 2017 data.

Table 4.6 Cluster construction data for 2017, applying the Squared Euclidean differences

	Capexpend2015	AnIncomeGrow2015	PGT2015	Genderpaygap2015	Marketing2015	Managers2015	Overseas2015	Continuecustomers2015	Debtors2015	Staffturnover2015	Sicknessdays2015
JB Alpha	15.1	4.3	65.0	1.0	5.3	0.08	0.0	94.0	0.5	16.0	6.0
Cosign Research	9.9	6.1	59.0	1.0	5.1	0.03	5.0	92.0	1.5	5.0	4.0
Mini Max	7.1	5.3	48.0	2.0	4.7	0.04	0.0	87.0	3.5	4.0	6.0
System Synthesis	7.1	5.0	43.0	7.0	6.2	0.03	12.0	85.0	4.0	8.0	6.0
Open Thinking	8.5	7.0	65.0	3.0	6.3	0.05	8.0	96.0	1.0	9.0	4.0
LKS Data	2.3	5.3	73.0	1.0	5.3	0.04	6.0	97.0	1.0	8.0	3.0
Strategy Statistics	6.3	14.0	84.0	1.0	5.2	0.05	4.0	90.0	5.0	13.0	6.0
Visual Research	12.5	4.0	55.0	6.0	4.3	0.04	6.0	83.0	4.0	8.0	9.0
Ashton Algorithms	8.2	2.3	61.0	4.0	8.2	0.03	6.0	89.0	4.5	7.0	6.0
Linear Logics	6.9	4.7	58.0	5.0	6.3	0.05	14.0	90.0	3.0	8.0	5.0
Sun Focus	6.4	5.6	61.0	4.0	5.9	0.04	10.0	79.0	5.0	7.0	7.0
New Perspectives	5.7	5.4	55.0	5.0	5.0	0.04	12.0	86.0	4.0	13.0	8.0
Mean	*8.0*	*5.8*	*60.6*	*3.3*	*5.7*	*0.04*	*6.9*	*89.0*	*3.1*	*8.8*	*5.8*
Standard Deviation	*3.2*	*2.7*	*10.3*	*2.1*	*1.0*	*0.01*	*4.3*	*5.1*	*1.6*	*3.3*	*1.6*
					Standardized scores						
JB Alpha	2.24	-0.53	0.43	-1.14	-0.36	2.79	-1.60	0.98	-1.64	2.15	0.10
Cosign Research	0.60	0.13	-0.15	-1.14	-0.56	-1.02	-0.44	0.59	-1.00	-1.15	-1.13
Mini Max	-0.28	-0.17	-1.22	-0.65	-0.97	-0.25	-1.60	-0.39	0.26	-1.45	0.10
System Synthesis	-0.28	-0.28	-1.70	1.78	0.56	-1.02	1.18	-0.78	0.58	-0.25	0.10
Open Thinking	0.16	0.46	0.43	-0.16	0.66	0.51	0.25	1.37	-1.32	0.05	-1.13
LKS Data	-1.80	-0.17	1.20	-1.14	-0.36	-0.25	-0.21	1.56	-1.32	-0.25	-1.74
Strategy Statistics	-0.54	3.03	2.27	-1.14	-0.46	0.51	-0.68	0.20	1.21	1.25	0.10
Visual Research	1.42	-0.64	-0.54	1.30	-1.37	-0.25	-0.21	-1.17	0.58	-0.25	1.95
Ashton Algorithms	0.06	-1.27	0.04	0.32	2.59	-1.02	-0.21	0.00	0.90	-0.55	0.10
Linear Logics	-0.35	-0.39	-0.25	0.81	0.66	0.51	1.64	0.20	-0.05	-0.25	-0.51
Sun Focus	-0.5	-0.06	0.04	0.32	0.25	-0.25	0.72	-1.95	1.21	-0.55	0.72
New Perspectives	-0.73	-0.1	-0.5	0.81	-0.7	-0.3	1.18	-0.6	0.58	1.25	1.33

Squared Euclidean Distance

	JB Alpha	Cosign Research	Mini Max	System Synthesis	Open Thinking	LKS Data	Strategy Statistics	Visual Research	Ashton Algorithms	Linear Logics	Sun Focus	New Perspectives
JB Alpha		32.30	37.52	56.35	22.14	37.86	39.41	38.50	47.35	37.01	49.46	40.50
Cosign Research	32.30		**8.49**	22.46	8.03	10.59	30.27	24.52	20.22	15.37	21.15	24.68
Mini Max	37.52	8.49		18.51	19.56	21.87	32.86	15.04	20.41	19.16	14.00	19.56
System Synthesis	56.35	22.46	18.51		22.09	35.14	45.66	14.43	13.14	**7.36**	8.40	8.38
Open Thinking	**22.14**	**8.03**	19.56	22.09		**8.10**	**24.27**	30.44	18.30	7.76	23.28	21.13
LKS Data	37.86	10.59	21.87	35.14	8.10		27.68	45.31	28.29	18.10	31.35	29.96
Strategy Statistics	39.41	30.27	32.86	45.66	24.27	27.68		40.66	41.11	32.68	27.86	28.27
Visual Research	38.50	24.52	15.04	14.43	30.44	45.31	40.66		24.79	20.00	11.46	10.51
Ashton Algorithms	47.35	20.22	20.41	13.14	18.30	28.29	41.11	24.79		12.17	13.00	20.79
Linear Logics	37.01	15.37	19.16	**7.36**	**7.76**	18.10	32.68	20.00	**12.17**		9.90	9.50
Sun Focus	49.46	21.15	14.00	8.40	23.28	31.35	27.86	11.46	13.00	9.90		**7.57**
New Perspectives	40.50	24.68	19.56	8.38	21.13	29.96	28.27	**10.51**	20.79	9.50	**7.57**	
Minimum distance	22.14	8.03	8.49	7.36	7.76	8.10	24.27	10.51	12.17	7.36	7.57	7.57

The next task is to search for unique pairs, as the basis for clusters, from the Squared Euclidean Distance Matrix (table 4.6).

The lowest score is Linear Logics and System Synthesis (7.36). This forms the start of the first cluster.

Next is Sun Focus and new Perspectives (7.57). This forms the start of the second cluster.

The next pair is Open Thinking and Linear Logics (7.76). this is not a unique pair because Linear logics has already been located in cluster 1 with System Synthesis.

Next is Cosign Research and Open Thinking (8.03). This is a unique pair and forms the start of cluster three.

The next most similar pair is LKS Data and Open Thinking (8.10). This is not a unique pair as Open Thinking is already located with Cosign Research.

Next is Mini Max and Cosign Research (8.49). As we have just seen, this is not a unique pair, because Cosign Research is already allocated to a cluster.

Next is Visual Research and New Perspectives (10.51). this is not a unique pair, as New Perspectives is already in cluster two.

The next most similar pair is Ashton Algorithms and Linear Logics (12.17). This is not a unique pair because Linear Logics is already allocated to cluster one.

Next is JB Alpha and Open Thinking (22.14). This is not unique, because Open thinking is already in cluster three.

The final pair is Strategy Statistics and Open Thinking (24.27). This is not unique, because Open Thinking is already in cluster three.

The result of this first stage before examining the table evidence is that there are three clusters. Clusters not in unique pairs are allocated to clusters on the basis of their first pairing and the cluster allocation given to their pair. On this basis the clusters are as follows.

Cluster 1 Linear Logics, System Synthesis, Ashton Algorithms

Cluster 2 Sun Focus, New Perspectives, Visual Research

Cluster 3 Open Thinking, Cosign Research, LKS Data, Mini Max, JB Alpha, Strategy Statistics

Given the relatively small number of unique pairs and the size of clusters, exploring the cluster pattern tables become particularly important for checking the validity of the cluster structure.

The resulting clusters are shown in table 4.7

Table 4.7 Cluster pattern synthesis – adding clusters and descriptive statistics, 2017

	Capexpend2017	AnIncomeGrow2017	PGT2017	Genderpaygap2017	Marketing2017	Managers2017	Overseas2017	Continuecustomers2017	Debtors2017	Staffturnover2017	Sicknessdays2017	Clusters
System Synthesis	7.1	5.0	43.0	7.0	6.2	0.03	12.0	85.0	4.0	8.0	6.0	1
Linear Logics	6.9	4.7	58.0	5.0	6.3	0.05	14.0	90.0	3.0	8.0	5.0	1
Ashton Algorithms	8.2	2.3	61.0	4.0	8.2	0.03	6.0	89.0	4.5	7.0	6.0	1
Visual Research	12.5	4.0	55.0	6.0	4.3	0.04	6.0	83.0	4.0	8.0	9.0	2
New Perspectives	5.7	5.4	55.0	5.0	5.0	0.04	12.0	86.0	4.0	13.0	8.0	2
Sun Focus	6.4	5.6	61.0	4.0	5.9	0.04	10.0	79.0	5.0	7.0	7.0	2
Mini Max	7.1	5.3	48.0	2.0	4.7	0.04	0.0	87.0	3.5	4.0	6.0	3
Cosign Research	9.9	6.1	59.0	1.0	5.1	0.03	5.0	92.0	1.5	5.0	4.0	3
LKS Data	2.3	5.3	73.0	1.0	5.3	0.04	6.0	97.0	1.0	8.0	3.0	3
Open Thinking	8.5	7.0	65.0	3.0	6.3	0.05	8.0	96.0	1.0	9.0	4.0	3
JB Alpha	15.1	4.3	65.0	1.0	5.3	0.08	0.0	94.0	0.5	16.0	6.0	3
Strategy Statistics	6.3	14.0	84.0	1.0	5.2	0.05	4.0	90.0	5.0	13.0	6.0	3
Mean	*8.0*	*5.8*	*60.6*	*3.3*	*5.7*	*0.04*	*6.9*	*89.0*	*3.1*	*8.8*	*5.8*	
Median	*7.1*	*5.3*	*60.0*	*3.5*	*5.3*	*0.04*	*6.0*	*89.5*	*3.8*	*8.0*	*6.0*	
Standard Deviation	*3.2*	*2.7*	*10.3*	*2.1*	*1.0*	*0.01*	*4.3*	*5.1*	*1.6*	*3.3*	*1.6*	

The larger cluster (3) is important to consider for its degree of homogeniety.

Table 4.8 shows the colour gradients for case scores. Table 4.9 applies the monochrome shading principle to the cluster pattern synthesis. This exposes more clearly the variable patterns that contribute the most to cluster definitions.

Cluster 1 has a shared below mean scores for annual income growth and staff turnover, and a shared above mean scores for gender pay gap and marketing. In addition, capital expenditure is relatively low, with Ashton Algorithms only marginally above the mean of 8.0 with a near miss score of 8.2. Likewise, postgraduate qualifications are relatively low with one near miss. (Ashton Algorithms scores marginally above the mean of 60.6 at 61.0.) Overseas is a near miss for shared above average scores, with only Ashton Algorithms marginally below the mean of 6.9 with a score at the median point of 6.0. Debtors is a near miss for above the mean (3.1), with only Linear Logics scoring close to the mean at 3.0. Staff sickness is a near miss for above average scores with only Linear Logics (5.0) marginally below the mean (5.8) This cluster is a good example of how important it can be to consider near misses. The volume of near misses provides evidence that cluster 1 is a homogeneous cluster.

Table 4.8 Cluster pattern synthesis – using variable colour coding, 2017

	Capexpend2017	AnIncomeGrow2017	PGT2017	Genderpaygap2017	Marketing2017	Managers2017	Overseas2017	Continuecustomers2017	Debtors2017	Staffturnover2017	Sicknessdays2017	Cluster
System Synthesis	7.1	5.0	43.0	7.0	6.2	0.03	12.0	85.0	4.0	8.0	6.0	1
Ashton Algorithms	8.2	2.3	61.0	4.0	8.2	0.03	6.0	89.0	4.5	7.0	6.0	1
Linear Logics	6.9	4.7	58.0	5.0	6.3	0.05	14.0	90.0	3.0	8.0	5.0	1
Visual Research	12.5	4.0	55.0	6.0	4.3	0.04	6.0	83.0	4.0	8.0	9.0	2
Sun Focus	6.4	5.6	61.0	4.0	5.9	0.04	10.0	79.0	5.0	7.0	7.0	2
New Perspectives	5.7	5.4	55.0	5.0	5.0	0.04	12.0	86.0	4.0	13.0	8.0	2
Open Thinking	8.5	7.0	65.0	3.0	6.3	0.05	8.0	96.0	1.0	9.0	4.0	3
Mini Max	7.1	5.3	48.0	2.0	4.7	0.04	0.0	87.0	3.5	4.0	6.0	3
LKS Data	2.3	5.3	73.0	1.0	5.3	0.04	6.0	97.0	1.0	8.0	3.0	3
JB Alpha	15.1	4.3	65.0	1.0	5.3	0.08	0.0	94.0	0.5	16.0	6.0	3
Cosign Research	9.9	6.1	59.0	1.0	5.1	0.03	5.0	92.0	1.5	5.0	4.0	3
Strategy Statistics	6.3	14.0	84.0	1.0	5.2	0.05	4.0	90.0	5.0	13.0	6.0	3
Mean	*8.0*	*5.8*	*60.6*	*3.3*	*5.7*	*0.04*	*6.9*	*89.0*	*3.1*	*8.8*	*5.8*	
Median	*7.1*	*5.3*	*60.0*	*3.5*	*5.3*	*0.04*	*6.0*	*89.5*	*3.8*	*8.0*	*6.0*	
Standard Deviation	*3.2*	*2.7*	*10.3*	*2.1*	*1.0*	*0.01*	*4.3*	*5.1*	*1.6*	*3.3*	*1.6*	

In table 4.9, Cluster 2 shares two below mean scores and three above mean scores, indicating a homogeneous pair. In addition, the managers score is uniform at the mean (0.04).

In table 4.9, Cluster 3 is a large cluster with six case members. It shares one below mean score for gender pay, but has two important near misses. These misses are indicated in table 4.9 with shading because of this being a larger cluster. Open Thinking is the only near miss (below the mean) for marketing and Mini Max is the only near miss for continuing customers (above the mean).

We can further consider this complex, larger cluster with reference to the exploration of clusters resulting from the Squared Euclidean Distance matrix in table 4.6. In fact, we can recreate a simpler version of the table that just displays the matrix for cluster 3 (see table 4.10).

Table 4.9 Cluster pattern synthesis – using monochrome shading to summarise clusters, 2017

	Capexpend2017	AnIncomeGrow2017	PGT2017	Genderpaygap2017	Marketing2017	Managers2017	Overseas2017	Continuecustomers2017	Debtors2017	Staffturnover2017	Sicknessdays2017	Clusters
Linear Logics	6.9	4.7	58.0	5.0	6.3	0.05	14.0	90.0	3.0	8.0	5.0	1
System Synthesis	7.1	5.0	43.0	7.0	6.2	0.03	12.0	85.0	4.0	8.0	6.0	1
Ashton Algorithms	8.2	2.3	61.0	4.0	8.2	0.03	6.0	89.0	4.5	7.0	6.0	1
New Perspectives	5.7	5.4	55.0	5.0	5.0	0.04	12.0	86.0	4.0	13.0	8.0	2
Sun Focus	6.4	5.6	61.0	4.0	5.9	0.04	10.0	79.0	5.0	7.0	7.0	2
Visual Research	12.5	4.0	55.0	6.0	4.3	0.04	6.0	83.0	4.0	8.0	9.0	2
Open Thinking	8.5	7.0	65.0	3.0	6.3	0.05	8.0	96.0	1.0	9.0	4.0	3
Mini Max	7.1	5.3	48.0	2.0	4.7	0.04	0.0	87.0	3.5	4.0	6.0	3
LKS Data	2.3	5.3	73.0	1.0	5.3	0.04	6.0	97.0	1.0	8.0	3.0	3
JB Alpha	15.1	4.3	65.0	1.0	5.3	0.08	0.0	94.0	0.5	16.0	6.0	3
Cosign Research	9.9	6.1	59.0	1.0	5.1	0.03	5.0	92.0	1.5	5.0	4.0	3
Strategy Statistics	6.3	14.0	84.0	1.0	5.2	0.05	4.0	90.0	5.0	13.0	6.0	3
Mean	*8.0*	*5.8*	*60.6*	*3.3*	*5.7*	*0.04*	*6.9*	*89.0*	*3.1*	*8.8*	*5.8*	
Median	*7.1*	*5.3*	*60.0*	*3.5*	*5.3*	*0.04*	*6.0*	*89.5*	*3.8*	*8.0*	*6.0*	
Standard Deviation	*3.2*	*2.7*	*10.3*	*2.1*	*1.0*	*0.01*	*4.3*	*5.1*	*1.6*	*3.3*	*1.6*	

Table 4.10 Cluster 3 - Squared Euclidean Distance Matrix

	JB Alpha	Cosign Research	Mini Max	Open Thinking	LKS Data	Strategy Statistics
JB Alpha		32.30	37.52	22.14	37.86	39.41
Cosign Research	32.30		***8.49***	**8.03**	***10.59***	30.27
Mini Max	37.52	***8.49***		**19.56**	***21.87***	32.86
Open Thinking	22.14	***8.03***	***19.56***		***8.10***	24.27
LKS Data	37.86	***10.59***	***21.87***	**8.10**		27.68
Strategy Statistics	39.41	30.27	32.86	24.27	27.68	
Minimum distance	22.14	8.03	8.49	8.03	8.10	24.27

Table 4.10 shows four cases within the cluster that have substantially lower scores between their pairs than the other two cases.

The four cases are: Cosign Research, Open Thinking, Mini Max and LKS

Table 4.11 Cluster pattern synthesis – examining cluster 3 validation in more detail, 2017

	Capexpend2017	AnIncomeGrow2017	PGT2017	Genderpaygap2017	Marketing2017	Managers2017	Overseas2017	Continuecustomers2017	Debtors2017	Staffturnover2017	Sicknessdays2017	Clusters
Mini Max	7.1	5.3	48.0	2.0	4.7	0.04	0.0	87.0	3.5	4.0	6.0	3
Cosign Research	9.9	6.1	59.0	1.0	5.1	0.03	5.0	92.0	1.5	5.0	4.0	3
LKS Data	2.3	5.3	73.0	1.0	5.3	0.04	6.0	97.0	1.0	8.0	3.0	3
Open Thinking	8.5	7.0	65.0	3.0	6.3	0.05	8.0	96.0	1.0	9.0	4.0	3
JB Alpha	15.1	4.3	65.0	1.0	5.3	**0.08**	**0.0**	**94.0**	0.5	16.0	6.0	3
Strategy Statistics	6.3	14.0	84.0	1.0	5.2	**0.05**	**4.0**	**90.0**	5.0	13.0	6.0	3
Mean	*8.0*	*5.8*	*60.6*	*3.3*	*5.7*	*0.04*	*6.9*	*89.0*	*3.1*	*8.8*	*5.8*	
Median	*7.1*	*5.3*	*60.0*	*3.5*	*5.3*	*0.04*	*6.0*	*89.5*	*3.8*	*8.0*	*6.0*	
Standard Deviation	*3.2*	*2.7*	*10.3*	*2.1*	*1.0*	*0.01*	*4.3*	*5.1*	*1.6*	*3.3*	*1.6*	

This can be examined in more detail with variable scores for the cluster using a monochrome table (Table 4.11)

Table 4.11 indicates the resulting similarity of the pairing of the potential cluster outliers, JB Alpha and Strategy Statistics. This illustrates the risk of relying entirely on the calculation of the Squared Euclidean distance between pairs. While Strategy Statistics and JB Alpha have a high score for their Euclidean distance (39.41) an examination of their individual variables scores in table 4.11 shows that they share similar scores for seven variables, but the other four variables show opposite scores in the distribution range for these two cases. JB Alpha and Strategy Statistics have some variable scores that are similar, but others where they are very different. This illustrates the problem with relying on the aggregate Squared Euclidean Distance score alone for case comparisons, without considering the individual variables. The Euclidean distance scores aggregate the experience of the two cases. They do not reveal the variable complexity, but tables do demonstrate this.

We can also see the fuzzy and imperfect nature of cluster definitions, especially with larger clusters. In this example, cluster 3 is retained because of its overall sharing of similarity with three variables, despite the more homogeneous pairing of JB Alpha and Strategy Statistics. An alternative analysis and model is to place JB Alpha and Strategy Statistics into a new cluster 4.

Table 4.12 Clusters resorted to show a single outcome variable, debtors, 2017

	Capexpend2017	AnIncomeGrow2017	PGT2017	Genderpaygap2017	Marketing2017	Managers2017	Overseas2017	Continuecustomers2017	Staffturnover2017	Sicknessdays2017	Clusters	Debtors2017
JB Alpha	15.1	4.3	65.0	1.0	5.3	0.08	0.0	94.0	16.0	6.0	3	0.5
LKS Data	2.3	5.3	73.0	1.0	5.3	0.04	6.0	97.0	8.0	3.0	3	1.0
Open Thinking	8.5	7.0	65.0	3.0	6.3	0.05	8.0	96.0	9.0	4.0	3	1.0
Cosign Research	9.9	6.1	59.0	1.0	5.1	0.03	5.0	92.0	5.0	4.0	3	1.5
Linear Logics	6.9	4.7	58.0	5.0	6.3	0.05	14.0	90.0	8.0	5.0	1	3.0
Mini Max	7.1	5.3	48.0	2.0	4.7	0.04	0.0	87.0	4.0	6.0	3	3.5
System Synthesis	7.1	5.0	43.0	7.0	6.2	0.03	12.0	85.0	8.0	6.0	1	4.0
New Perspectives	5.7	5.4	55.0	5.0	5.0	0.04	12.0	86.0	13.0	8.0	1	4.0
Visual Research	12.5	4.0	55.0	6.0	4.3	0.04	6.0	83.0	8.0	9.0	2	4.0
Ashton Algorithms	8.2	2.3	61.0	4.0	8.2	0.03	6.0	89.0	7.0	6.0	1	4.5
Sun Focus	6.4	5.6	61.0	4.0	5.9	0.04	10.0	79.0	7.0	7.0	1	5.0
Strategy Statistics	6.3	14.0	84.0	1.0	5.2	0.05	4.0	90.0	13.0	6.0	3	5.0
Mean	*8.0*	*5.8*	*60.6*	*3.3*	*5.7*	*0.04*	*6.9*	*89.0*	*8.8*	*5.8*		*3.1*
Median	*7.1*	*5.3*	*60.0*	*3.5*	*5.3*	*0.04*	*6.0*	*89.5*	*8.0*	*6.0*		*3.8*
Standard Deviation	*3.2*	*2.7*	*10.3*	*2.1*	*1.0*	*0.01*	*4.3*	*5.1*	*3.3*	*1.6*		*1.6*

Table 4.12 shows the further sorting of the table to demonstrate debtors as an outcome variable.

Cluster 3 remains as a relatively homogenous explanation of below mean debtor scores where scores are tending to the lowest third of the range for debtors comparing all 12 cases. There is an association with the highest quartile of above mean scores for continuing customers. There is an important exception. Strategy Statistics, in cluster 3, has the highest score for debtors.

When considering above average debtor scores, if Linear Logics is treated as a near miss with its debtors score of 3.0 (0.1 below the mean) clusters 1 and 2 evidence higher comparable debtor scores and this is associated with below average annual income growth. There is also a near miss for above average sickness where only Linear Logics (5.0) is below the mean (5.8). Further near misses are: postgraduate qualifications, where only Ashton Algorithms and Sun Focus (61.0) are marginally above the mean (60.6); gender pay gap (mean is 3.3), where only Mini Max is below the mean (2.0); and staff turnover is all below the mean (8.8), apart from New Perspectives (13.0).

It can be argued that there is some degree of evidence of symmetry in the negative association between higher continuing customers and lower debt. Clusters 1 and 2, if combined in table 4.8, show continuing customer scores that are all at the mean (89.) or below, apart from Linear Logics (90.0) that has the highest score for continuing customers in this group. Linear Logics also has the lowest relative score for debtors in this combined cluster with a debtor score at 3.0 (mean is 3.1).

A conclusion is that there is evidence of lower debtors being associated with higher continuing customers, a lower gender pay gap and lower sickness (LKS Data, Open Thinking and Cosign Research in table 4.12).

Chapter five: Concluding the DPS: longitudinal change

The final stage of DPS is to consider longitudinal change. This requires comparing the three years of data.

When concluding the results of DPS it is important to conclude for:

(1) the changes of variable scores over time;

(2) case similarities and differences over time;

(3) the dynamic interactions between variable changes with case pattern changes.

Changes of variable scores over time

A good starting point is to plot the mean and/or median variable averages for each year and to conclude on the trend characteristic for each variable. Table 5.1 shows the mean variable changes, and the final row indicates a qualitative text-based conclusion about the trend for that variable. Standard deviations could also be added to this table, but are not currently shown.

Where overall change over the three years is relatively small, it is appropriate to conclude that the variable trend change is stable, although this may also depend on the range of scale measurement used by that variable.

Noticeable variable trends from the example data are a decline in annual income growth (although some evidence of a recovery in 2017), an increasing percentage of employees with postgraduate qualifications, an increase in continuing customers, and a decline in staff turnover.

Table 5.1 Variable trend mean changes: 2015-17

	Capexpend	AnIncomeGrow	PGT	Genderpaygap	Marketing	Managers	Overseas	continuecustomers	debtors	staffturnover	sicknessdays
2015	7.1	7.9	55.6	3.0	5.0	0.04	6.8	84.6	2.5	13.5	5.6
2016	7.8	3.4	58.7	2.5	5.8	0.04	6.8	86.8	2.4	7.6	4.8
2017	8.0	5.8	60.6	3.3	5.7	0.04	6.9	89.0	3.1	8.8	5.8
	stable	v	^	stable	stable	stable	stable	^	stable	v	stable

Cluster changes

The best method for considering cluster stability and case changes between clusters is to first compare them visually in a table (table 5.2). Cases that remain in similar across the three years are highlighted in bold.

Table 5.2 Longitudinal cluster patterns

<table>
<tr><th>2015</th><th>2016</th><th>2017</th></tr>
<tr><td>Cluster 1
Open Thinking
LKS Data
Strategy Statistics</td><td>Cluster 2
LKS Data
Open Thinking
Cosign Research</td><td>Cluster 3
LKS Data
Open Thinking
Cosign Research
Mini Max
JB Alpha
Strategy Statistics</td></tr>
<tr><td>Custer 5
JB Alpha
Cosign Research</td><td>Cluster 5
JB Alpha
Strategy Statistics</td><td></td></tr>
<tr><td>Cluster 2
Mini Max
Ashston Algorithms</td><td>Cluster 1
Mini Max
New Perspectives
System Synthesis</td><td rowspan="2">Cluster 2
New Perspectives
Sun Focus
Visual Research</td></tr>
<tr><td>Cluster 3
Sun Focus
New Perspectives</td><td>Cluster 3
Linear Logics
Sun Focus</td></tr>
<tr><td>Cluster 4
Visual Research
System Synthesis
Linear Logics</td><td>Cluster 4
Visual Research
Ashton Algorithms</td><td>Cluster 1
Linear Logics
System Synthesis
Ashton Algorithms</td></tr>
</table>

Open Thinking and LKS Data are linked as a pair across all three years, in addition, Cosign Research, LKS data and Open Thinking are also linked between 2016 and 2017.

New Perspectives and Sun Focus return to the same cluster in 2017, having been first identified together in 2015, but are separated in the 2016 model. The same is true for the relationship of System Synthesis and Linear Logics.

JB Alpha and Strategy Statistics are linked in the 2016 and 2017 models.

Combined case and variable changes

Another method for considering the overall characteristics over time of the DPS is to create a longitudinal pattern table. This device uses an Excel® formula to identify cases that have similar trajectories of average variable change across the three-year period. It is therefore a method for identifying change patterns over the period.

Table 5.3 Excel® calculation example for longitudinal average score for JB Alpha

	t1 Capexpend2015	t2 Capexpend2016	t3 Capexpend2017	
JB Alpha	12.3	14.6	15.1	
Formula In Excel®				(t1+t2+t3)/3 =mean(B17…C17)
JB Alpha				**14.0**

Table 5.3 shows the Excel® process and method for computing the longitudinal average where the case variable scores for 2015,2016, 2017 are in cells B2, C2 and D2. Therefore, the mean average longitudinal score for capital expenditure for JB Alpha is 14.0

Table 5.4 shows a summary version with the average change for each case and variable across the three-year period.

Table 5.4 Longitudinal averages (2015 - 2017)

	capital expenditure	Annual income growth	Postgraduate staff	Gender pay gap	Marketing	Managers	Overseas	Continuing customers	Debtors	Staff turn over	Sickness
JB Alpha	14.0	3.4	69.0	1.3	4.8	0.09	0.0	92.0	1.5	20.3	5.7
Cosign Research	10.7	4.4	56.0	2.3	4.9	0.03	5.0	88.7	1.8	9.0	3.7
Mini Max	5.7	5.3	43.0	2.3	4.9	0.03	0.7	86.7	3.2	7.7	6.3
System Synthesis	8.2	11.2	39.0	6.7	7.2	0.02	11.3	82.3	3.7	9.3	6.0
Open Thinking	8.5	8.5	65.7	2.0	5.6	0.05	6.3	98.0	0.8	9.7	5.0
LKS Data	3.2	6.2	74.7	1.3	4.8	0.04	4.3	97.3	1.0	7.7	3.0
Strategy Statistics	4.3	8.6	86.3	0.7	5.0	0.05	3.3	90.3	2.7	17.0	6.7
Visual Research	10.8	6.2	47.7	4.3	5.2	0.04	7.7	84.7	3.3	5.3	7.7
Ashton Algorithms	7.8	1.4	58.3	2.3	8.1	0.03	4.7	83.7	4.0	10.0	5.7
Linear Logics	7.5	2.5	50.3	5.3	6.5	0.05	19.0	81.3	2.7	7.3	4.0
Sun Focus	6.1	3.5	58.7	2.7	4.9	0.04	7.3	73.3	4.2	5.3	4.7
New Perspectives	5.0	6.6	50.7	4.0	3.8	0.04	12.3	83.3	3.0	11.0	6.7
Mean	*7.7*	*5.7*	*58.3*	*2.9*	*5.5*	*0.04*	*6.8*	*86.8*	*2.7*	*10.0*	*5.4*
Median	*7.7*	*5.8*	*57.2*	*2.3*	*5.0*	*0.04*	*5.7*	*85.7*	*2.8*	*9.2*	*5.7*
Standard Dev	*3.0*	*2.7*	*13.1*	*1.7*	*1.2*	*0.02*	*5.1*	*6.7*	*1.1*	*4.3*	*1.3*

Using table 5.5, we can apply the same approach to an exploration of the Euclidean differences of longitudinal averages that we have used previously for each separate year in chapters 3 and 4.

After searching for unique pairs using the approach used in chapters 3 and 4, the following cluster structure results for validation with cluster pattern tables.

Open Thinking, LKS Data, System Synthesis and Mini Max and JB Alpha are in cluster one. The remainder of the cases are in cluster two.

Table 5.5 Raw data and simple cluster calculations for longitudinal case comparisons

	Capexpend	AnIncomeGrow	PGT20	Genderpaygap	Marketing	Managers	Overseas	Continuecustomer	Debtors	Staffturnover	Sicknessdays
						Raw scores					
JB Alpha	14	3.4	69	1.3	4.8	0.09	0	92	1.5	20.3	5.7
Cosign Research	10.7	4.4	56	2.3	4.9	0.03	5	88.7	1.8	9	3.7
Mini Max	5.7	5.3	43	2.3	4.9	0.03	0.7	86.7	3.2	7.7	6.3
System Synthesis	8.2	11.2	39	6.7	7.2	0.02	11.3	82.3	3.7	9.3	6
Open Thinking	8.5	8.5	65.7	2	5.6	0.05	6.3	98	0.8	9.7	5
LKS Data	3.2	6.2	74.7	1.3	4.8	0.04	4.3	97.3	1	7.7	3
Strategy Statistics	4.3	8.6	86.3	0.7	5	0.05	3.3	90.3	2.7	17	6.7
Visual Research	10.8	6.2	47.7	4.3	5.2	0.04	7.7	84.7	3.3	5.3	7.7
Ashton Algorithms	7.8	1.4	58.3	2.3	8.1	0.03	4.7	83.7	4	10	5.7
Linear Logics	7.5	2.5	50.3	5.3	6.5	0.05	19	81.3	2.7	7.3	4
Sun Focus	6.1	3.5	58.7	2.7	4.9	0.04	7.3	73.3	4.2	5.3	4.7
New Perspectives	5	6.6	50.7	4	3.8	0.04	12.3	83.3	3	11	6.7
Mean	*7.7*	*5.7*	*58.3*	*2.9*	*5.5*	*0.04*	*6.8*	*86.8*	*2.7*	*10.0*	*5.4*
Standard Deviation	*3.0*	*2.7*	*13.1*	*1.7*	*1.2*	*0.02*	*5.1*	*6.7*	*1.1*	*4.3*	*1.3*
						Standardized scores					
JB Alpha	2.15	-0.83	0.82	-0.95	-0.59	2.81	-1.34	0.78	-1.06	2.41	0.20
Cosign Research	1.03	-0.46	-0.17	-0.37	-0.50	-0.74	-0.36	0.28	-0.78	-0.23	-1.30
Mini Max	-0.66	-0.13	-1.16	-0.37	-0.50	-0.74	-1.20	-0.01	0.50	-0.53	0.65
System Synthesis	0.19	2.04	-1.47	2.20	1.50	-1.33	0.88	-0.67	0.95	-0.16	0.43
Open Thinking	0.29	1.05	0.56	-0.54	0.11	0.44	-0.10	1.67	-1.70	-0.06	-0.33
LKS Data	-1.51	0.20	1.25	-0.95	-0.59	-0.15	-0.50	1.57	-1.52	-0.53	-1.83
Strategy Statistics	-1.13	1.09	2.13	-1.30	-0.41	0.44	-0.69	0.52	0.04	1.64	0.95
Visual Research	1.07	0.20	-0.81	0.80	-0.24	-0.15	0.17	-0.31	0.59	-1.09	1.70
Ashton Algorithms	0.05	-1.57	0.00	-0.37	2.28	-0.74	-0.42	-0.46	1.23	0.01	0.20
Linear Logics	-0.05	-1.16	-0.61	1.38	0.89	0.44	2.39	-0.82	0.04	-0.62	-1.08
Sun Focus	-0.524	-0.79	0.03	-0.14	-0.50	-0.15	0.09	-2.02	1.41	-1.09	-0.55
New Perspectives	-0.896	0.35	-0.58	0.622	-1.453	-0.15	1.0768	-0.52	0.312	0.241	0.95149

Squared Euclidean Distance

	JB Alpha	Cosign Research	Mini Max	System Synthesis	Open Thinking	LKS Data	Strategy Statistics	Visual Research	Ashton Algorithms	Linear Logics	Sun Focus	New Perspectives
JB Alpha		25.84	37.19	66.49	22.47	37.69	24.78	37.55	40.16	48.78	46.00	39.30
Cosign Research	25.84		10.30	28.04	8.99	12.18	24.11	14.91	16.84	17.93	14.60	15.99
Mini Max	37.19	10.30		21.63	17.12	20.55	20.43	8.41	**13.48**	24.94	**10.58**	9.06
System Synthesis	66.49	28.04	21.63		31.80	47.53	42.94	14.24	24.64	20.88	26.09	18.32
Open Thinking	**22.47**	**8.99**	17.12	31.80		**8.08**	**14.56**	19.89	26.84	27.05	29.64	19.33
LKS Data	37.69	12.18	20.55	47.53	8.08		18.20	35.21	32.05	32.46	27.81	25.48
Strategy Statistics	24.78	24.11	20.43	42.94	14.56	18.20		28.84	28.19	43.23	28.72	19.36
Visual Research	37.55	14.91	**8.41**	**14.24**	19.89	35.21	28.84		17.09	18.55	13.82	**8.71**
Ashton Algorithms	40.16	16.84	13.48	24.64	26.84	32.05	28.19	17.09		18.42	13.52	23.84
Linear Logics	48.78	17.93	24.94	20.88	27.05	32.46	43.23	18.55	18.42		14.45	16.17
Sun Focus	46.00	14.60	10.58	26.09	29.64	27.81	28.72	13.82	13.52	**14.45**		11.73
New Perspectives	39.30	15.99	9.06	18.32	19.33	25.48	19.36	8.71	23.84	16.17	11.73	
Minimum distance	22.47	8.99	8.41	14.24	8.08	8.08	14.56	8.41	13.48	14.45	10.58	8.71

Using the clustering method employed in this book in the previous chapter, the minimum distances are used to allocate the cases into clusters. Two clusters result.

Table 5.6 shows the longitudinal averages with descriptive statistics and their allocated cluster groups.

Table 5.6 Longitudinal averages (2015-2017) sorted by clusters

	Capital expenditure	Annual income growth	Postgraduate staff	Gender pay gap	Marketing	Managers	Overseas	Continuing customers	Debtors	Staff turn over	Sickness	Cluster
Cosign Research	10.7	4.4	56.0	2.3	4.9	0.03	5.0	88.7	1.8	9.0	3.7	1
Open Thinking	8.5	8.5	65.7	2.0	5.6	0.05	6.3	98.0	0.8	9.7	5.0	1
LKS Data	3.2	6.2	74.7	1.3	4.8	0.04	4.3	97.3	1.0	7.7	3.0	1
Strategy Statistics	4.3	8.6	86.3	0.7	5.0	0.05	3.3	90.3	2.7	17.0	6.7	1
JB Alpha	14.0	3.4	69.0	1.3	4.8	0.09	0.0	92.0	1.5	20.3	5.7	1
Mini Max	5.7	5.3	43.0	2.3	4.9	0.03	0.7	86.7	3.2	7.7	6.3	2
System Synthesis	8.2	11.2	39.0	6.7	7.2	0.02	11.3	82.3	3.7	9.3	6.0	2
Visual Research	10.8	6.2	47.7	4.3	5.2	0.04	7.7	84.7	3.3	5.3	7.7	2
New Perspectives	5.0	6.6	50.7	4.0	3.8	0.04	12.3	83.3	3.0	11.0	6.7	2
Ashton Algorithms	7.8	1.4	58.3	2.3	8.1	0.03	4.7	83.7	4.0	10.0	5.7	2
Linear Logics	7.5	2.5	50.3	5.3	6.5	0.05	19.0	81.3	2.7	7.3	4.0	2
Sun Focus	6.1	3.5	58.7	2.7	4.9	0.04	7.3	73.3	4.2	5.3	4.7	2
Mean	*7.7*	*5.7*	*58.3*	*2.9*	*5.5*	*0.04*	*6.8*	*86.8*	*2.7*	*10.0*	*5.4*	
Median	*7.7*	*5.8*	*57.2*	*2.3*	*5.0*	*0.04*	*5.7*	*85.7*	*2.8*	*9.2*	*5.7*	
Standard Dev	*3.0*	*2.7*	*13.1*	*1.7*	*1.2*	*0.02*	*5.1*	*6.7*	*1.1*	*4.3*	*1.3*	

Table 5.7 Longitudinal averages (2015-2017) showing cluster colour map

	capital expenditure	Annual income growth	Postgraduate staff	Gender pay gap	Marketing	Managers	Overseas	Continuing customers	Debtors	Staff turn over	Sickness	Cluster
JB Alpha	14	3.4	69	1.3	4.8	0.09	0	92	1.5	20.3	5.7	1
Cosign Research	10.7	4.4	56	2.3	4.9	0.03	5	88.7	1.8	9	3.7	1
Open Thinking	8.5	8.5	65.7	2	5.6	0.05	6.3	98	0.8	9.7	5	1
LKS Data	3.2	6.2	74.7	1.3	4.8	0.04	4.3	97.3	1	7.7	3	1
Strategy Statistics	4.3	8.6	86.3	0.7	5	0.05	3.3	90.3	2.7	17	6.7	1
Mini Max	5.7	5.3	43	2.3	4.9	0.03	0.7	86.7	3.2	7.7	6.3	2
System Synthesis	8.2	11.2	39	6.7	7.2	0.02	11.3	82.3	3.7	9.3	6	2
Visual Research	10.8	6.2	47.7	4.3	5.2	0.04	7.7	84.7	3.3	5.3	7.7	2
Ashton Algorithms	7.8	1.4	58.3	2.3	8.1	0.03	4.7	83.7	4	10	5.7	2
Linear Logics	7.5	2.5	50.3	5.3	6.5	0.05	19	81.3	2.7	7.3	4	2
Sun Focus	6.1	3.5	58.7	2.7	4.9	0.04	7.3	73.3	4.2	5.3	4.7	2
New Perspectives	5	6.6	50.7	4	3.8	0.04	12.3	83.3	3	11	6.7	2
Mean	*7.7*	*5.7*	*58.3*	*2.9*	*5.5*	*0.04*	*6.8*	*86.8*	*2.7*	*10.0*	*5.4*	
Median	*7.7*	*5.8*	*57.2*	*2.3*	*5.0*	*0.04*	*5.7*	*85.7*	*2.8*	*9.2*	*5.7*	
Standard Dev	*3.0*	*2.7*	*13.1*	*1.7*	*1.2*	*0.02*	*5.1*	*6.7*	*1.1*	*4.3*	*1.3*	

Table 5.7 shows a heat map with colour gradients for the longitudinal clusters. These provide evidence that the following variables appear to be distinguishing the allocation of cases to the two clusters. These are: postgraduate qualifications, gender pay gap, overseas customers, continuing customers, debtors and staff turnover.

Table 5.8 examines this by using monochrome shading to indicate above and below mean average variable scores. Cluster 1, with less cases as members, shows more homogeneity.

Cluster 1 has above mean scores for postgraduate qualifications except for the near miss for Cosign Research that has a score (56.0) close to the central tendency, but marginally below the mean (58.3) and median (57.2). The cluster also has above mean scores for continuing customers. There are below mean scores for gender pay gap and overseas customers. Also, marketing is below mean, with one near miss exception for Open Thinking (5.6) where the mean is 5.5. And there are below mean scores for debtors, for all cases except Strategy Statistics which has a score at the mean (2.7).

Cluster 2 has noticeable above mean scores for debtors and sickness, but with one near miss for debtors (Linear Logics with a score at the mean, 2.7) and sickness having two near misses below the mean for Linear Logics and Sun Focus (scores of 4.0 and 4.7, where the mean is 5.4). There are below mean average scores for continuing customers and postgraduates. For postgraduates, there is a near miss for Sun Focus at 58.7 that is marginally above the mean of 58.3. Ashton Algorithms exhibits this mean score.

Table 5.9 resorts the data in table 5.8 to show the chosen outcome variable that is now in the final column, debtors.

This table confirms that there is some evidence of a symmetrical causal relationship between above average continuing customers and below average debtors, and the opposite relationship. Similarly, there is a suggested negative association between postgraduate staff and debtors.

There are several other variables of interest where the influence on debtors is not symmetrical. Below average scores in gender pay gap and overseas customers appear to be associated with below average scores for debtors. Higher than average scores for sickness have a tendency to be associated with above average scores for debtors.

Table 5.8 Longitudinal averages (2013-2015) using monochrome shading to identify patterns

	capital expenditure	Annual income growth	Postgraduate staff	Gender pay gap	Marketing	Managers	Overseas	Continuing customers	Debtors	Staff turn over	Sickness	Cluster
Cosign Research	10.7	4.4	56.0	2.3	4.9	0.03	5.0	88.7	1.8	9.0	3.7	1
Open Thinking	8.5	8.5	65.7	2.0	5.6	0.05	6.3	98.0	0.8	9.7	5.0	1
LKS Data	3.2	6.2	74.7	1.3	4.8	0.04	4.3	97.3	1.0	7.7	3.0	1
Strategy Statistics	4.3	8.6	86.3	0.7	5.0	0.05	3.3	90.3	2.7	17.0	6.7	1
JB Alpha	14.0	3.4	69.0	1.3	4.8	0.09	0.0	92.0	1.5	20.3	5.7	1
Mini Max	5.7	5.3	43.0	2.3	4.9	0.03	0.7	86.7	3.2	7.7	6.3	2
System Synthesis	8.2	11.2	39.0	6.7	7.2	0.02	11.3	82.3	3.7	9.3	6.0	2
Visual Research	10.8	6.2	47.7	4.3	5.2	0.04	7.7	84.7	3.3	5.3	7.7	2
New Perspectives	5.0	6.6	50.7	4.0	3.8	0.04	12.3	83.3	3.0	11.0	6.7	2
Ashton Algorithms	7.8	1.4	58.3	2.3	8.1	0.03	4.7	83.7	4.0	10.0	5.7	2
Linear Logics	7.5	2.5	50.3	5.3	6.5	0.05	19.0	81.3	2.7	7.3	4.0	2
Sun Focus	6.1	3.5	58.7	2.7	4.9	0.04	7.3	73.3	4.2	5.3	4.7	2
Mean	*7.7*	*5.7*	*58.3*	*2.9*	*5.5*	*0.04*	*6.8*	*86.8*	*2.7*	*10.0*	*5.4*	
Median	*7.7*	*5.8*	*57.2*	*2.3*	*5.0*	*0.04*	*5.7*	*85.7*	*2.8*	*9.2*	*5.7*	
Standard Dev	*3.0*	*2.7*	*13.1*	*1.7*	*1.2*	*0.02*	*5.1*	*6.7*	*1.1*	*4.3*	*1.3*	

Table 5.9 Longitudinal averages (2013-2015) sorted by debtor outcomes

	Capital expenditure	Annual income growth	Postgraduate staff	Gender pay gap	Marketing	Managers	Overseas	Continuing customers	Staff turn over	Sickness	Cluster	Debtors
Open Thinking	8.5	8.5	65.7	2.0	5.6	0.05	6.3	98.0	9.7	5.0	1	0.8
LKS Data	3.2	6.2	74.7	1.3	4.8	0.04	4.3	97.3	7.7	3.0	1	1.0
JB Alpha	14.0	3.4	69.0	1.3	4.8	0.09	0.0	92.0	20.3	5.7	1	1.5
Cosign Research	10.7	4.4	56.0	2.3	4.9	0.03	5.0	88.7	9.0	3.7	1	1.8
Strategy Statistics	4.3	8.6	86.3	0.7	5.0	0.05	3.3	90.3	17.0	6.7	1	2.7
Linear Logics	7.5	2.5	50.3	5.3	6.5	0.05	19.0	81.3	7.3	4.0	2	2.7
New Perspectives	5.0	6.6	50.7	4.0	3.8	0.04	12.3	83.3	11.0	6.7	2	3.0
Mini Max	5.7	5.3	43.0	2.3	4.9	0.03	0.7	86.7	7.7	6.3	2	3.2
Visual Research	10.8	6.2	47.7	4.3	5.2	0.04	7.7	84.7	5.3	7.7	2	3.3
System Synthesis	8.2	11.2	39.0	6.7	7.2	0.02	11.3	82.3	9.3	6.0	2	3.7
Ashton Algorithms	7.8	1.4	58.3	2.3	8.1	0.03	4.7	83.7	10.0	5.7	2	4.0
Sun Focus	6.1	3.5	58.7	2.7	4.9	0.04	7.3	73.3	5.3	4.7	2	4.2
Mean	*7.7*	*5.7*	*58.3*	*2.9*	*5.5*	*0.04*	*6.8*	*86.8*	*10.0*	*5.4*		*2.7*
Median	*7.7*	*5.8*	*57.2*	*2.3*	*5.0*	*0.04*	*5.7*	*85.7*	*9.2*	*5.7*		*2.8*
Standard Dev	*3.0*	*2.7*	*13.1*	*1.7*	*1.2*	*0.02*	*5.1*	*6.7*	*4.3*	*1.3*		*1.1*

The results of this resorting are a structure of case similarity and difference that is not always directly related to variable similarities over time as, indicated in the earlier comparison of years in table 5.1. Both postgraduate qualifications and continuing customers are upward trends across the sample (table 5.1) and these are important goals for organisations who want to reduce the money owed in debts. The sample trend with debtors is relatively stable, with only small annual fluctuations. Together, all evidence suggests a degree of instability in case-based similarity and difference, with it being relatively likely that cases will change over time in relationship to each other (table 5.2), despite more relative stability in the underlying variables that define the characteristics of these cases. This is not surprising given we are examining meso business cases in a dynamic market environment. Macro studies with DPS tend to illustrate more case-based stability over time, with countries being relatively less likely to change their similarity and difference to each other over time, despite any variable instability (see Haynes, 2017). Nevertheless, there are some case patterns that stand out in this research example, in terms of individual organisations that share characteristics that seem to help them to reduce the impact of debtors.

Checking the consistency of case based variable averages over time

There is a further technique for examining case based variable averages over time that was used in the first edition of this book.

This requires converting the annual datasets into binary scores, sometimes known as crisp sets, where each variable has a value of 0 or 1 based on a threshold point in the scale. For example, if the median is the threshold point used, 0 will define scores below the median and 1 will define scores above the median.

The first stage for using this technique is convert the scale data into crisp set, binary, categories. This approach has some similarities to the formulation of so called 'truth tables' (Ragin , 1987).

Converting the scale data to crisp set categories

A fixed algorithm is used to set the threshold in an Excel® workbooks. This threshold is used to automatically convert the scale data into the two binary categories (0,1).

Excel® Formula for converting binary data to crisp sets

Place this formula in the cell where you want the crisp set score (0 or 1) to appear.

=IF(B2>B$15,1,0)

In the example above, the original scale value for single case score is in B2 and the request is for the crisp set conversion to calculated from the algorithm that calculated the variable median, in cell B15.

This produced the results in table 5.10, for the first year of data, 2015.

Table 5.10 An example of scale data converted to binary crisp set scores (2015)

	Capexpend2015	AnIncomeGrow2015	PGT2015	Genderpaygap2015	Marketing2015	Managers2015	Overseas2015	continuecustomers2015	debtors2015	staffturnover2015	sicknessdays2015
JB Alpha	12.3	2.9	72.0	2.0	5.0	0.10	0.0	90.0	2.0	30.0	6.0
Cosign Research	11.1	3.0	54.0	3.0	4.3	0.03	6.0	84.0	2.0	15.0	4.0
Mini Max	4.5	4.0	32.0	3.0	5.2	0.02	0.0	86.0	3.0	16.0	7.0
System Synthesis	9.2	13.7	34.0	7.0	8.1	0.01	12.0	82.0	3.0	13.0	6.0
Open Thinking	8.7	15.6	67.0	1.0	4.2	0.05	6.0	100.0	0.5	16.0	5.0
LKS Data	3.1	8.9	76.0	1.0	4.0	0.05	5.0	98.0	1.0	8.0	4.0
Strategy Statistics	2.1	6.9	90.0	1.0	4.6	0.04	3.0	89.0	1.0	21.0	9.0
Visual Research	9.8	20.3	43.0	3.0	5.7	0.05	8.0	84.0	3.0	2.0	7.0
Ashton Algorithms	7.1	2.8	56.0	1.0	7.2	0.03	4.0	77.0	3.5	14.0	6.0
Linear Logics	7.4	2.3	42.0	8.0	6.1	0.05	23.0	76.0	3.0	9.0	3.0
Sun Focus	5.7	7.1	56.0	2.0	3.7	0.04	4.0	69.0	5.0	7.0	4.0
New Perspectives	4.7	7.3	45.0	4.0	2.3	0.04	11.0	80.0	3.0	11.0	6.0
Mean	*7.1*	*7.9*	*55.6*	*3.0*	*5.0*	*0.04*	*6.8*	*84.6*	*2.5*	*13.5*	*5.6*
Median	*7.3*	*7.0*	*55.0*	*2.5*	*4.8*	*0.04*	*5.5*	*84.0*	*3.0*	*13.5*	*6.0*
Standard Deviation	*3.1*	*5.6*	*17.0*	*2.2*	*1.5*	*0.02*	*6.0*	*8.5*	*1.2*	*6.9*	*1.6*
JB Alpha	1	0	1	0	1	1	0	1	0	1	1
Cosign Research	1	0	0	1	0	0	1	1	0	1	0
Mini Max	0	0	0	1	1	0	0	1	1	1	1
System Synthesis	1	1	0	1	1	0	1	0	1	0	1
Open Thinking	1	1	1	0	0	1	1	1	0	1	0
LKS Data	0	1	1	0	0	1	0	1	0	0	0
Strategy Statistics	0	0	1	0	0	1	0	1	0	1	1
Visual Research	1	1	0	1	1	1	1	1	1	0	1
Ashton Algorithms	0	0	1	0	1	0	0	0	1	1	1
Linear Logics	1	0	0	1	1	1	1	0	1	0	0
Sun Focus	0	1	1	0	0	1	0	0	1	0	0
New Perspectives	0	1	0	1	0	1	1	0	1	0	1

This conversion method can also be applied to the other data years, 2016, 2017.

Finally, an Excel® formula is created and used to identify case and variable patterns that are consistently above or below threshold (median) for the entire three-year period. It is therefore an additional method for identifying stable patterns over the time period. However, this method does not analyse cluster groupings before comparing cases and variables.

Table 5.11 shows the Excel® process and method for computing this longitudinal pattern of cases that have variable scores consistently above, or below average for all three years/

The formula used is:

=IF(OR(AND(B2=1,B15=1,B28=1),),"ABOVE",IF(AND(B2=0,B15=0,B28=0),"BELOW"," "))

In the example above, the formula is for the first comparison and the cells being compared are B2, B15 and B26. That is B2 is the data for 2015, B15 is the data for 2016, and B28 is the data for 2017.

The result of the Excel® formula computations is in the area at the bottom of table 5.11 sub headed: 'OVERALL threshold stability'. Cells scoring ABOVE, are where there is a consistent above threshold score between the case and variable for all three years, and those scoring BELOW, are where these is consistent below threshold score between the case and variable for all three years. If the cell is empty, with no text present, this indicates there is no consistent score over time for that case and variable relationship.

Tables 5.12 and 5.13 show a resorting of the overall threshold stability data, to find the consistent case pattern similarities, as defined by variables, over time. No cluster calculations or allocations are used in these tables. The sort in table 5.12 is set up to prioritise the order of cases according to the amount of consistent longitudinal scores in each variable. For example, the first variable in the sort sub menu for table 5.12, is set as the proportion of employees who have a postgraduate qualification, as this variable is consistently either above or below threshold for all cases across all three years (there are no empty cells in the column).

Table 5.13 resorts the same results to show debtors as the outcome variable in the last column. As this method does not include a cluster algorithm, but considers the cases separately, the focus of the results is a little different to table 5.9. In table 5.13, the relationship between above average continuing customers and below average debtors stands out for four cases, Open Thinking, Cosign Research, JB Alpha and LKS Data. This relationship is consistent across all three years. Compared to table 5.9 that highlighted a symmetrical negative relationship between postgraduate staff and debtors, in table 5.13 there is less evidence for this relationship and more emphasis on a partial association between three cases with above average postgraduate qualifications and below average debt (Open Thinking, JB Alpha, and LKS Data).

In general, the development of DPS in this second edition is to prioritise the visualisation and consideration of scale data where possible, rather than to oversimplify the complexity of the data via the use of binary categorical tables (as in the first edition).

Table 5.11 Setup for a longitudinal table with consistent above and below average scores

2015	Capexpend	AnIncomeGrow	PGT	Genderpaygap	Marketing	Managers	Overseas	Continuecustomers	Debtors	Staffturnover	Sicknessdays
JB Alpha	1	0	1	0	1	1	0	1	0	1	1
Cosign Research	1	0	0	1	0	0	1	1	0	1	0
Mini Max	0	0	0	1	1	0	0	1	1	1	1
System Synthesis	1	1	0	1	1	0	1	0	1	0	1
Open Thinking	1	1	1	0	0	1	1	1	0	1	0
LKS Data	0	1	1	0	0	1	0	1	0	0	0
Strategy Statistics	0	0	1	0	0	1	0	1	0	1	1
Visual Research	1	1	0	1	1	1	1	1	1	0	1
Ashton Algorithms	0	0	1	0	1	0	0	0	1	1	1
Linear Logics	1	0	0	1	1	1	1	0	1	0	0
Sun Focus	0	1	1	0	0	1	0	0	1	0	0
New Perspectives	0	1	0	1	0	1	1	0	1	0	1
2016											
JB Alpha	1	0	1	0	0	1	0	1	0	1	0
Cosign Research	1	1	0	1	0	0	0	1	0	0	0
Mini Max	0	1	0	0	0	0	0	0	1	0	1
System Synthesis	1	1	0	1	1	0	1	0	1	0	1
Open Thinking	1	0	1	0	1	1	1	1	0	0	1
LKS Data	0	1	1	0	0	0	0	1	0	0	0
Strategy Statistics	0	1	1	0	0	1	0	1	0	1	0
Visual Research	1	0	0	1	1	0	1	0	1	0	1
Ashton Algorithms	0	0	1	0	1	0	0	0	1	1	0
Linear Logics	1	0	0	1	1	1	1	0	0	0	0
Sun Focus	0	0	1	0	0	0	1	0	1	0	0
New Perspectives	0	1	0	1	0	1	1	0	0	1	1
2017											
JB Alpha	1	0	1	0	0	1	0	1	0	1	0
Cosign Research	1	1	0	0	0	0	0	1	0	0	0
Mini Max	0	0	0	0	0	0	0	0	0	0	0
System Synthesis	0	0	0	1	1	0	1	0	1	0	0
Open Thinking	1	1	1	0	1	1	1	1	0	1	0
LKS Data	0	0	1	0	0	0	0	1	0	0	0
Strategy Statistics	0	1	1	0	0	1	0	1	1	1	0
Visual Research	1	0	0	1	0	0	0	0	1	0	1
Ashton Algorithms	1	0	1	1	1	0	0	0	1	0	0
Linear Logics	0	0	0	1	1	1	1	1	0	0	0
Sun Focus	0	1	1	1	1	0	1	0	1	0	1
New Perspectives	0	1	0	1	0	0	1	0	1	1	1
OVERALL threshold stability – 2015,2016,2017											
JB Alpha	**ABOVE**	BELOW	**ABOVE**	BELOW		**ABOVE**	BELOW	**ABOVE**	BELOW	**ABOVE**	
Cosign Research	**ABOVE**		BELOW		BELOW	BELOW		**ABOVE**	BELOW		BELOW
Mini Max	BELOW		BELOW			BELOW	BELOW				
System Synthesis			BELOW	**ABOVE**	**ABOVE**	BELOW	**ABOVE**	BELOW	**ABOVE**	BELOW	
Open Thinking	**ABOVE**		**ABOVE**	BELOW		**ABOVE**	**ABOVE**	**ABOVE**	BELOW		
LKS Data	BELOW		**ABOVE**	BELOW	BELOW		BELOW	**ABOVE**	BELOW	BELOW	BELOW
Strategy Statistics	BELOW		**ABOVE**	BELOW	BELOW	**ABOVE**	BELOW	**ABOVE**		**ABOVE**	
Visual Research	**ABOVE**		BELOW	**ABOVE**					**ABOVE**	BELOW	**ABOVE**
Ashton Algorithms		BELOW	**ABOVE**		**ABOVE**	BELOW	BELOW	BELOW	**ABOVE**		
Linear Logics		BELOW	BELOW	**ABOVE**	**ABOVE**	**ABOVE**	**ABOVE**			BELOW	BELOW
Sun Focus	BELOW		**ABOVE**					BELOW	ABOVE	BELOW	
New Perspectives	BELOW	**ABOVE**	BELOW	**ABOVE**	BELOW		**ABOVE**	BELOW			**ABOVE**

Table 5.12 Cases with consistently above or below median average score from 2015 - 2017 , sorted by variable patterns

	Capexpend	AnIncomeGrow	PGT	Genderpaygap	Marketing	Managers	Overseas	Continuecustomers	Debtors	Staffturnover	Sicknessdays
Ashton Algorithms		BELOW	**ABOVE**		ABOVE	BELOW	BELOW	BELOW	ABOVE		
Open Thinking	**ABOVE**		**ABOVE**	**BELOW**		**ABOVE**	ABOVE	**ABOVE**	**BELOW**		
JB Alpha	**ABOVE**	BELOW	**ABOVE**	**BELOW**		**ABOVE**	BELOW	**ABOVE**	**BELOW**	ABOVE	
Sun Focus	**BELOW**		**ABOVE**					BELOW	ABOVE	BELOW	
LKS Data	**BELOW**		**ABOVE**	**BELOW**	**BELOW**		**BELOW**	ABOVE	BELOW	BELOW	BELOW
Strategy Statistics	**BELOW**		**ABOVE**	**BELOW**	**BELOW**	ABOVE	**BELOW**	ABOVE		ABOVE	
Linear Logics		BELOW	**BELOW**	**ABOVE**	ABOVE	ABOVE	ABOVE			**BELOW**	BELOW
System Synthesis			**BELOW**	**ABOVE**	ABOVE	BELOW	ABOVE	BELOW	ABOVE	**BELOW**	
Visual Research	ABOVE		**BELOW**	**ABOVE**					ABOVE	**BELOW**	ABOVE
Cosign Research	ABOVE		**BELOW**		BELOW	BELOW		ABOVE	BELOW		BELOW
New Perspectives	BELOW	ABOVE	**BELOW**	ABOVE	BELOW		ABOVE	BELOW			ABOVE
Mini Max	BELOW		**BELOW**			BELOW	BELOW				

*Table 5.*13 Cases with consistently above or below median average scores from 2015 - 2017, sorted by debtors outcome

	Capexpend	AnIncomeGrow	PGT	Genderpaygap	Marketing	Managers	Overseas	Continuecustomers	Staffturnover	Sicknessdays	**Debtors**
Linear Logics		BELOW	BELOW	**ABOVE**	**ABOVE**	**ABOVE**	**ABOVE**		BELOW	BELOW	
Strategy Statistics	BELOW		**ABOVE**	BELOW	BELOW	**ABOVE**	BELOW	**ABOVE**	**ABOVE**		
Mini Max	BELOW		BELOW			BELOW	BELOW				
New Perspectives	BELOW	**ABOVE**	BELOW	**ABOVE**	BELOW		**ABOVE**	BELOW		**ABOVE**	
System Synthesis			BELOW	**ABOVE**	**ABOVE**	BELOW	**ABOVE**	BELOW	BELOW		**ABOVE**
Ashton Algorithms		BELOW	**ABOVE**		**ABOVE**	BELOW	BELOW	BELOW			**ABOVE**
Visual Research	**ABOVE**		BELOW	**ABOVE**					BELOW	**ABOVE**	**ABOVE**
Sun Focus	BELOW		**ABOVE**					BELOW	BELOW		**ABOVE**
Open Thinking	**ABOVE**		**ABOVE**	BELOW		**ABOVE**	**ABOVE**	**ABOVE**			BELOW
Cosign Research	**ABOVE**		BELOW		BELOW	BELOW		**ABOVE**		BELOW	BELOW
JB Alpha	**ABOVE**	BELOW	**ABOVE**	BELOW		**ABOVE**	BELOW	**ABOVE**	**ABOVE**		BELOW
LKS Data	BELOW		**ABOVE**	BELOW	BELOW		BELOW	**ABOVE**	BELOW	BELOW	BELOW

The qualitative DPS summary

The outcome focus in this book has been on the percentage of debtors as defined by percentage of customers with late payment invoices over one year. Nevertheless, DPS allows for the researcher to resort and restructure the pattern matrix using a variety of table structures to consider any available variable as an outcome that might be affected by other variables. The method can also be used to focus on a single case of interest and how it compares to others.

The longitudinal table 5.13 shows four businesses that consistently have above average debtors and four others that have consistently below average debtors. These two groups do fall into different clusters, when cluster calculations and allocations are used in table 5.9. Examples of those with below average debtors are: Open Thinking, LKS Data, JB Alpha, and Cosign Research. Table 5.2 that looked at the consistency of cases in clusters over time identified some important evidence of consistency with Open Thinking and LKS Data in particular, less so with JB Alpha. Strategy statistics is also a case on the periphery of this relationship in table 5.2 and it is interesting that it does not have a consistently low debtor pattern over time (table 5.12). This is explained by its higher-than-average score for debtors in 2017 (table 4.12).

In addition, analysis provides evidence of some businesses that conversely have consistent challenges over time with the level of debtors and which also have scored consistently below average for retaining continuing customers.

Using the scale data and cluster approach in table 5.9 gives the clearest visual view of the degree of consistency between variable scores and any outcome of interest. The case simplification approach, without the allocation of clusters added in table 5.13, allows a more nuanced view of certain aspects of individual cases that might not be necessarily apparent in the clusters and table 5.9. Such an approach might be important if doing a detailed analysis of a single business case where the research focus is the priority of one single organisation and its circumstances.

For example, two of these cases, *System Synthesis* and *Ashton Algorithms*, in table 5.13 with above average debtors also have an ongoing pattern over time with above average expenditure on marketing (perhaps indicating a pressure to secure more new and reliable business) and a below average ratio of managers (suggesting that this needs exploring in terms of manager's ability to prioritise staff to chase and secure income owed). If undergoing a forensic analysis of one case only, these details might be more important than focusing on overall patterns and relationships. Table 5.2 shows that there is some important heterogeneity between *System Synthesis* and *Ashton Algorithms*, as they do not share consistent proximity in the cluster structures over time.

Similarly, an analysis can focus on subgroups of cases. For example, in table 5.13, the businesses that have consistently avoided having higher numbers of debtors across all three years are also above average in consistently retaining customers in this period. Three of these four are consistent in higher-than-average capital expenditure. Three cases are also consistent over time in employing a higher proportion of staff with postgraduate qualifications and having a lower gender pay gap. These features look to be part of a qualitative explanation about why these businesses can avoid delay in securing payments.

Conclusion

Dynamic Pattern Synthesis is influenced by the ontology of qualitative approaches but uses quantitative data evidence to explore case homogeneity and heterogeneity. Rather than building a model based on just aggregate or average scores, where real cases are considered according to their closeness to a typical or 'ideal' model of cases, DPS demonstrates dynamic differences between cases and the limits to their degrees of similarity. Similarity is balanced with considerations of difference also. This leads to a better sense of judgement about associations and causality, where such relationships are likely to be contextual, especially with regard to their persistence over time and place. These considerations are important in circumstances of economic and social complexity.

References

Alemna, D., Artaraz, K., Haynes, P., & Mwale, S. (2021). The complexity and instability of policy conditionality and transfer: IMF interventions in the political economy of South America. *Complexity Governance and Networks*, *6*(1), 14-31.

Alemna, D. (2022). Monetary Convergence Across the Economic Community of West African States: Lessons for the Envisioned West African Monetary Union. *Complexity, Governance & Networks*, *7*(1), 50-78.

Boulton, J., Allen, P., & Bowman, C. (2015). *Embracing Complexity: Strategic Perspectives for an Age of Turbulence.* Oxford: Oxford University Press.

Everitt, B. (1993). *Cluster Analysis.* London: Arnold.

Haynes, P. (2012). *Public policy beyond the financial crisis.* Oxon: Routledge.

Haynes, P. (2014). Combining the Strengths of qualitative comparative analysis and cluster analysis. *International Journal of Public Policy, 37*(9), 581-590.

Haynes, P. (2015). *Managing Complexity in the Public Services* (2nd ed.). Oxford: Routledge.

Haynes, P. (2015). The European Single Currency Project and the Concept of Convergence for European Welfare States - The Ideal and the Reality. *Social Policy and Administration, 44*(4), 466-489.

Haynes, P. (2017). *Social Synthesis: understanding patterns in complex social systems.* Oxford: Routledge.

Haynes, P., & Haynes, J. (2016). Convergence and Heterogeneity in Euro Based Economies: Stability and Dynamics. *Economies*, 1-16.

Haynes, P., & Alemna, D. (2022) From aggregations to multimethod case configurations. Case diversity in quantitative analysis when explaining COVID-19 fatalities. *International Journal of Social Research Methodology*, DOI: 10.1080/13645579.2022.2122224

Pawson, R., & Tilley, N. (1997). *Realistic Evaluation.* London: Sage.

Ragin, C. (1987). *The Comparative Method: Moving beyound Qualitative and Quantitative Strategies.* Berkeley, CA: University of California Press.

Rihoux, B., & Ragin, C. (2009). *Configurational Comparative Methods: Qualitative Comparative Methods (QCA) and Related Techniques.* London: Sage.

Snowden, D. J., & Boone, M. E. (2007). A leader's framework for decision making. *Harvard business review*, *85*(11), 68.

Taylor, L., Haynes, P., & Darking, M. (2021). English local government finance in transition: towards the ‘marketization of income’. *Public Management Review*, *23*(7), 1081-1106.

The datasets

Access to the dataset and example spreadsheets are via the publisher's websites at https://whb.co.uk/

It is possible to download the supporting files, for use in training and education.

Acknowledgements

My thanks to the following organisations and individuals who have helped to support the development of Dynamic Pattern Synthesis.

Brian Castellani

David Byrne

ESRC CECAN (Centre for the Evaluation of Complexity Across the Nexus)

ESRC Complex Methods Seminar programme

Lasse Gerrits

Jane Haynes

Jonathan Haynes

School of Humanities and Social Science, University of Brighton

University of Brighton Research Office and Impact Awards

Emma Uprichard

About the Authors

Philip Haynes is Professor of Public Policy at the University of Brighton where he teaches applied statistics, research methods, and social and public policy. He has published extensively in related areas including *Social Synthesis; Finding Dynamic Patterns in Complex Social Systems* (2017) *Managing Complexity in Public Policy*, 2nd Edition, (2015) and *Public Policy Beyond the Financial Crisis* (2012). He undertakes research and consultancy using applied statistics to solve policy and management related problems.

David Alemna is a lecturer in Politics and International Relations at the University of Portsmouth. Between 2021-22 he was an ESRC Post-Doctoral Research Fellow with the South Coast Doctoral Training Partnership and based at the University of Portsmouth (Ref: ES/W005743/1). He also works as an Analyst with the Inner City Fund (ICF - London) and completed his doctorate using DPS at the University of Brighton.

Dynamic Pattern Synthesis is one of a new range of case-based methods that seeks to balance the search for overall quantitative patterns based on aggregations of case information, with the complexity and diversity of individual cases. These methods try to avoid making assumptions about case similarity when using evidence of relationships between variables that seem to show cases are similar. Orthodox methods often ignore important aspects of case diversity and oversimplify the aggregate impact of some variables. Human and organisational diversity is vital to consider with greater attention, if applied social science research is to have a more useful impact on policy and practice in the future.

White Horse

whb.co.uk

2023

ISBN 978-0-9534065-3-1

9 780953 406531 >