# A Practical Guide to Data Visualisation

Julian Ellacott

Paul Teggin

Presented to The Staple Inn Actuarial Society

Staple Inn, London

4 February 2014

## About the authors

**Julian Ellacott**

Julian is Head of Actuarial Development at Just Retirement, having previously worked for Watson Wyatt.

julian.ellacott@gmail.com

**Paul Teggin**

Paul started his actuarial career at Royal & SunAlliance before moving to Prudential. He is now an independent consulting actuary.

paul.teggin@tegginconsulting.co.uk

## Acknowledgements

The authors would like to thank Matthew Edwards for his helpful comments and for chairing the SIAS meeting at which this paper was presented. The authors would also like to thank Jack Zhu for his kind assistance in producing the results used in the stochastic case study, Phil Joubert for suggesting the idea of word clouds to visualise technical or regulatory documents, and Allan Martin for drawing our attention to the presentation tool Prezi.

The authors very much welcome any feedback on the content of this paper, or on the subject more generally. We would also like to gauge demand for an online resource for the benefit of members of the profession (possibly on the IFoA website), collecting information on data visualisation, based on the links at the end of this paper. If there is sufficient interest, it might be worthwhile establishing a Member Interest Group or Working Party to take some of these ideas further.

## Disclaimer

The views expressed in this paper are those of the authors and do not necessarily represent the views of our employers or clients. The responsibility for any errors or omissions in this document rests with the authors; responsibility for referenced books or websites remains with their owners.

The authors are happy for reasonable use to be made of any of the content of this paper, with appropriate acknowledgement.

## Contents

# 1 Introduction

## 1.1 Overview

**What do we mean by data visualisation?** In response to this question we give three quotations from authors active in the field, which show that visualisation is a many-layered concept:

> *The main goal of data visualization is to* ***communicate information clearly and effectively*** *through graphical means.* (Friedman [1])
>
> *Important stories live in our data and data visualization is a powerful means to* ***discover and understand these stories****, and then to present them to others.* (Few [2])
>
> *Visualization gives you* ***answers to questions you didn't know you had****.*
> (Shneiderman [3])
>
> [All **emphases** added.]

The quotations suggest we should be thinking about **communication, graphics, narrative and novelty**. All these elements are important, but if we had to pick just one, it would be communication, rather than graphics. In the authors' view, the key to good visualisations is **good communication**: if the message of a visualisation is obscure, no amount of graphical skill will rescue it.

Perhaps unsurprisingly, then, the principles of good communication apply to data visualisation, albeit with adaptations where necessary. We shall see several examples of communication principles being applied to data visualisation throughout this paper.

**Graphics** are what distinguish data visualisation from other forms of communication. As with many other facets of life, there are people who are naturally gifted at graphic design, and then there are the rest of us. However, the basic principles of graphic design can be learned by anyone sufficiently motivated, and so we don't regard the need to engage with graphic design as being a barrier to actuaries using data visualisation techniques. This gives us the first of our principles of data visualisation:

> **Data visualisation techniques are not magical, they can be studied and learned.**

In fact, modern software packages and extensive online resources provide even the novice with a wide array of **tools** to help them communicate through visualisation - arguably the choice of tools is *too* wide, which can lead to a different type of problem, namely graphical overkill.

For obvious reasons, we restrict ourselves to **static** graphics in this paper – however **interactive** graphics can be very powerful in settings where they can reasonably be used (such as presentations or demonstrations).

**Narrative** can be seen as an extension of communication. The human brain appears to be "wired for narrative"[1] and tapping into this ability in a positive way is very useful.

The **novelty** factor is what makes it all worthwhile: the "lightbulb moment" when people see something that they hadn't realised previously, and really start to engage with some data. While there are people who can find things by staring at tables of numbers, the authors believe that graphs and diagrams are much more powerful at finding new insights.

One point of terminology: in this paper we use the word "data" in a very general sense of "**quantitative information**", rather than more specific actuarial usages (for example relating to attributes of customers of a financial institution, or observations from which assumptions are derived).

We believe that visualisation techniques can also fruitfully be applied to **qualitative** actuarial information (such as the structure of models or data-flows within them, or process diagrams), and we give a couple of examples in section 8, although in the interest of space we don't explore this topic in any depth.

### 1.2 Arrangement of Paper

Section 2 gives two examples to motivate the power of data visualisation. The remainder of the paper is focused on developing core principles to guide data visualisation, and illustrating these principles with short examples and longer case studies.

In section 3, we develop some high-level principles, inspired by analogies with the principles of good written communication. Section 4 presents our first case study, on presenting a correlation matrix for governance approval. In the first case study we also start to draw out some points about design and implementation, which we cover in more detail in sections 5 and 6. A second case study in section 7 on visualising stochastic model results pulls all the quantitative strands together, while section 8 touches briefly on qualitative visualisations.

Appendix A contains the visualisation references used in the main text, and also lots of links to web-based resources, with some commentary. Some of the websites are themselves collections of links to other resources. Finally in Appendix B we collect together the key principles developed throughout in the paper.

## 2 Motivating Examples

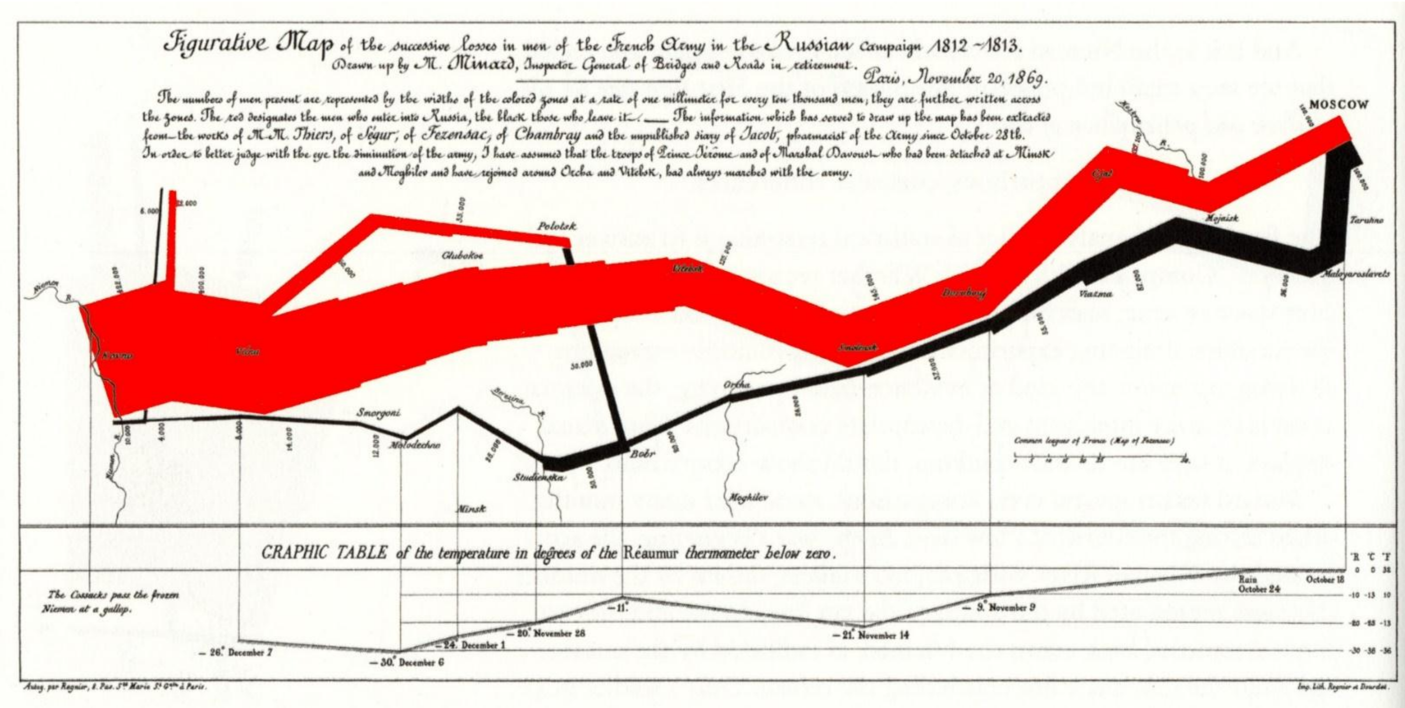
Before we start to describe the principles and practices of data visualisation, we give two visualisations which are generally considered to be excellent examples, illustrating a number of the key concepts which can be applied to our own visualisations.

[1] See for example www.newscientist.com/blogs/culturelab/2010/11/storytelling-20-when-new-narratives-meet-old-brains.html

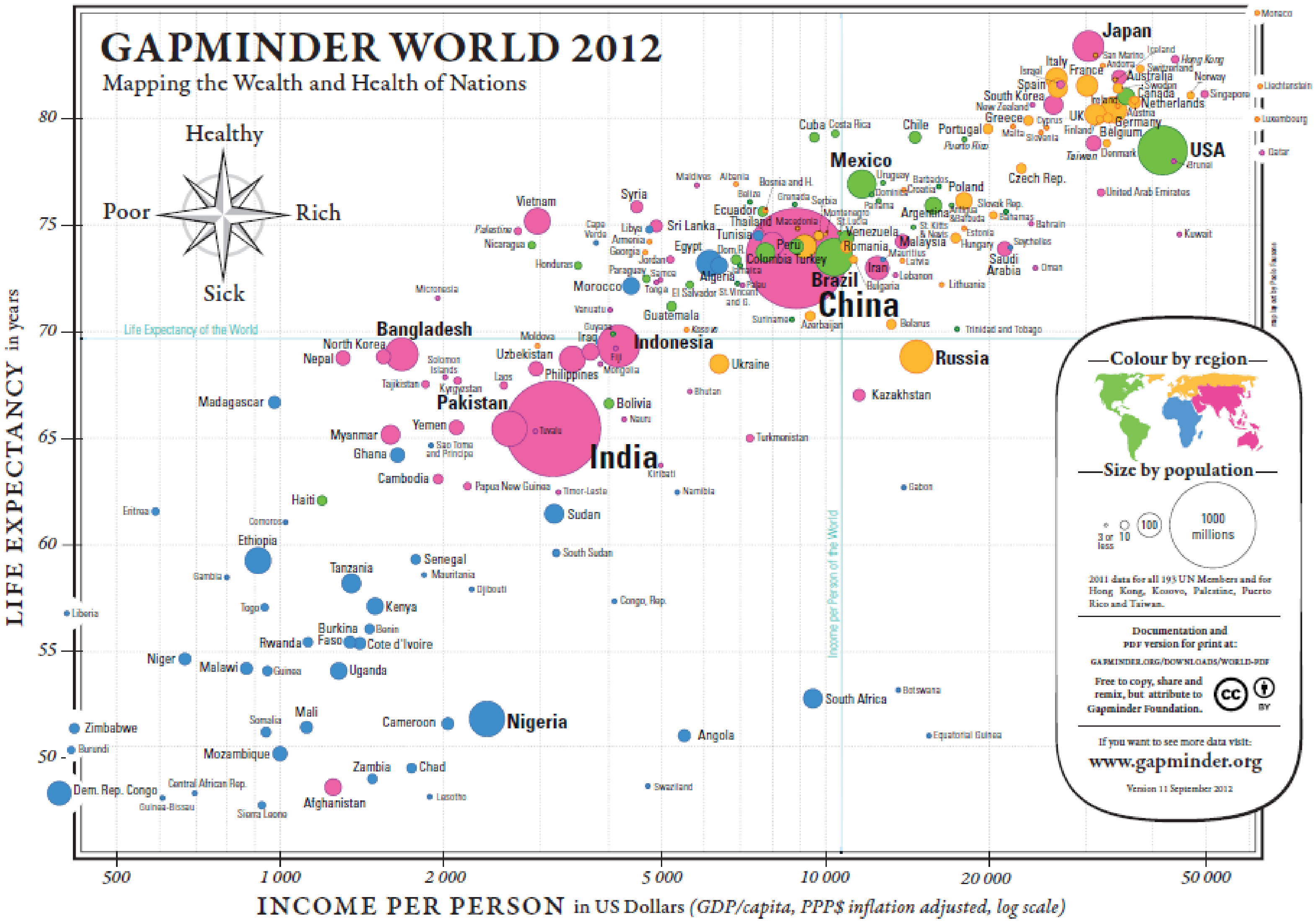
## 2.1 Napoleon's March to Moscow

The French version of this chart was produced by Charles Minard in 1869. (The one below has been translated to English and re-coloured, but is otherwise entirely in the spirit of the original.) The main purpose of the chart is to illustrate the losses incurred by Napoleon's army during its disastrous Russian campaign of 1812. Key devices used include:

- Use of the thickness of the red (attack) and black (retreat) lines to show the diminishing numbers in Napoleon's army, caused by the impact of successive battles, disease and cold. The actual numbers are shown periodically alongside the line.
- Limited use of place names and rivers, to give some spatial context but without turning it fully into a map, which would have cluttered the page and detracted from the key narrative.
- An illustration of the temperatures experienced at different points of the return march, shown along the bottom of the graphic.



## 2.2 World Health Chart



The main purpose of this chart is to illustrate the apparent relationship between key measures of “health” and “money”, for almost all countries in the world. This is an especially good demonstration of how more than 2 dimensions of data can be shown in a 2 dimensional form; key devices used include:

- Use of logarithmic scale to emphasise the strong positive correlation between the two variables.
- Use of colour to show the behaviour of subsets of the data, for each continent – for example Africa (highly clustered at the bottom left), Europe (broadly clustered at the top right), and Asia (very spread out).
- Use of “bubbles” to indicate the population of each country (and therefore, subconsciously, the weight to be placed on each data point in the context of a global average).
- Use of italics to show where data points are based on estimates rather than hard data – a very subtle way of making a relatively minor point.

# 3 High-Level Data Visualisation Principles

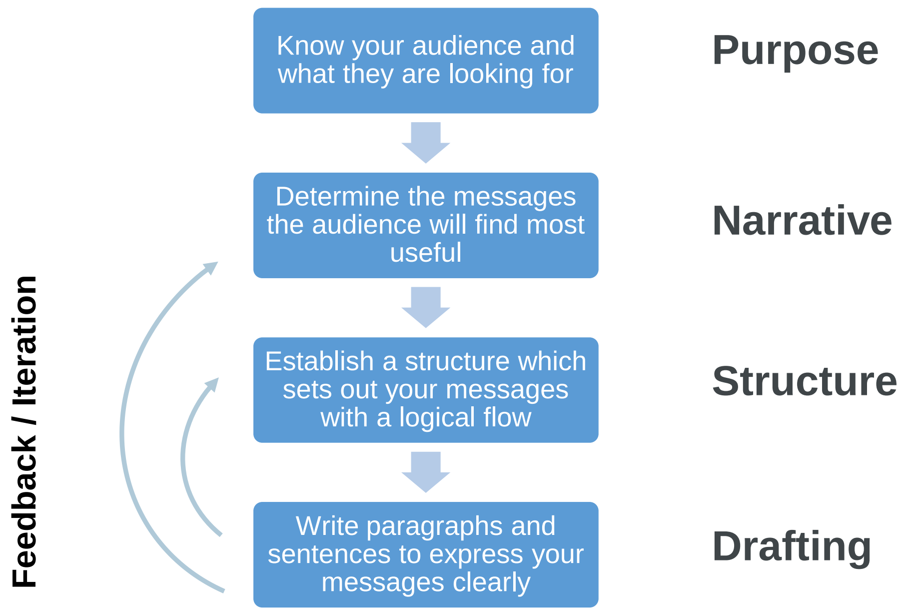
This paper is structured around a series of principles for data visualisation. In keeping with our emphasis on data visualisation as a form of communication, we start by applying well-established techniques of written communication to data visualisation. In fact, this is important enough to be a principle in its own right:

**Data visualisation is a form of communication, and so the techniques of good communication are highly applicable.**

Data visualisation is not purely about communication, and so we will develop more specific principles later on, but as we will see, there is a lot of mileage to be had from applying good communication techniques to data visualisation.

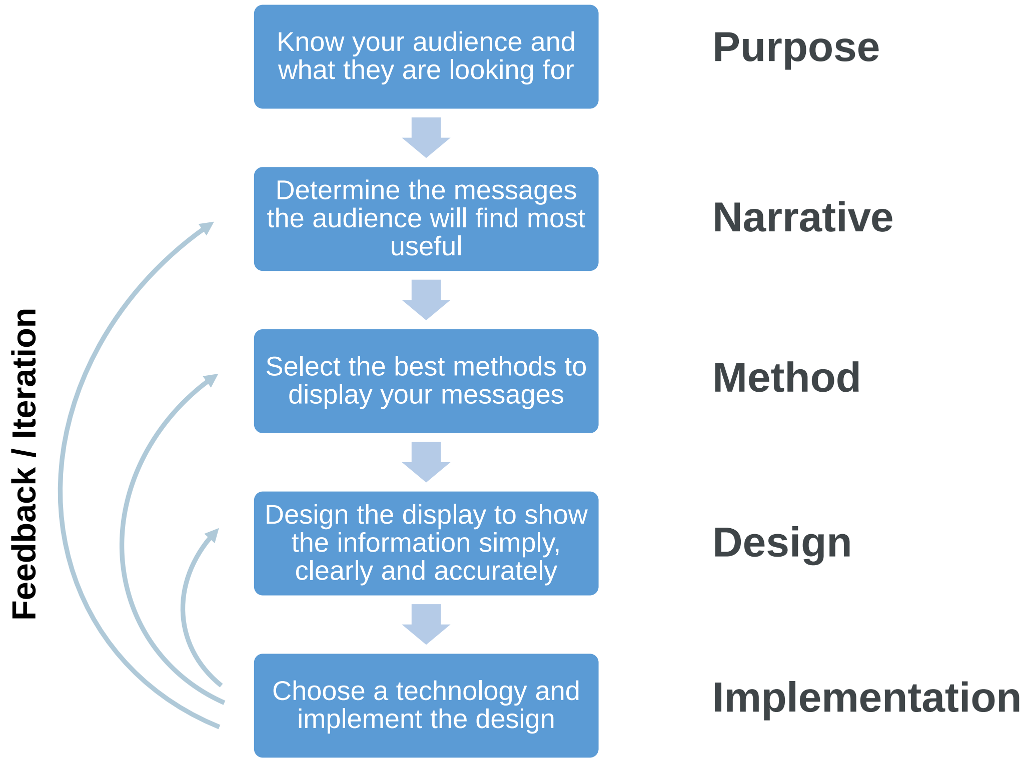
## 3.1 Structured Process

At a "macro" level, a structured process to written communication can be very useful, for example:




At the most basic level, the idea here is to know who you're addressing, what they're looking for, and what you're trying to say, and to get your thoughts into a logical progression prior to putting pen to paper (or finger to keyboard). We have probably all read reports where it wasn't entirely clear what the author was trying to say, or where the level of detail was unhelpful for the audience. It's unlikely that the first draft will be perfect, either structurally or at the sentence level, and so one or more incremental refinements are usually necessary.

We believe that a broadly similar structured process (below) can be very useful for data visualisation. The data visualisation version has an extra stage, for Implementation, and the stages of Structure and Drafting have become Method and Design:




Here purpose and narrative are common to both written communication and data visualisation: if you don't take account of your audience's needs, and don't know what you're trying to convey, then your visualisation is unlikely to be successful:

**A top-down structured process to data visualisation, starting with the needs of its users, is likely to be more successful than diving straight in.**

Structure and drafting can be regarded as high- and low-level aspects of constructing a document. Likewise, choosing a method is a high-level aspect, and corresponds broadly to laying out a structure, although there is more to it with data visualisation, as there are many more methods that can be applied than reasonable structures for a document. Design corresponds to paragraph and sentence construction, insofar as it relates to lower-level detail within a method / structure. There is much advice for authors on paragraph and sentence construction, and likewise we shall give and discuss some principles for designing good visualisation later.

Implementation is specific to data visualisation: the “implementation” of a written text is probably going to involve either a word processor or a slide presentation package, without many further considerations, and so was omitted from the earlier diagram. By contrast, the tools available to implement data visualisations can often constrain the method and design, or permit unhelpful methods and designs, so an understanding of the range of implementations is useful.

Both written communication and data visualisation benefit from an iterative approach, where attempts are made, critiqued, and improved:

**Trial and error will nearly always lead to improvements.**

In practical terms, the trial and error process can operate on a number of levels, for example:

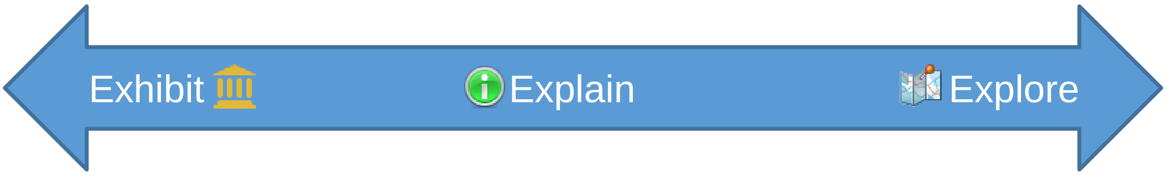
- Self-criticism by the author of the visualisation
- Discussion and challenge between the author and a colleague
- Explicit request by the author for feedback from the audience to which the visualisation is presented – asking what works, and what doesn’t work for them
- Presentation of the same data to the same audience using a number of different visualisations, asking which they prefer

In both written communication and data visualisation, a key skill to develop is a sense of what does and doesn’t work, and how to improve it. This applies at several levels in the overall process, and the amount of work will depend on the level. For example, changing the visualisation method is usually more work than changing the colour scheme, in the same way that changing a document’s structure can be more disruptive than re-drafting a paragraph. This emphasises a key benefit of the top-down approach, in which sound methods (or structures) are established earlier rather than later, before lots of detailed work has been carried out.

The process above is meant to provide guidance in a positive way: if it works in a particular situation, well and good; if not, please do not regard it as a straightjacket. We will illustrate this structure in case studies later in the paper.

## 3.2 Purpose

We classify the purpose of a visualisation in 3 different ways[2], illustrated below:




**What is the data?**

- Precision
- Clarity
- Simplicity

**What does the data mean?**

- Context
- Patterns
- Trends

**What else can I discover?**

- Relationships
- Connections
- Associations

At one end, we are simply trying to set out (or **exhibit**) the data, indicated with a museum icon in the diagram above. Exhibition of data is a very common actuarial requirement: think of all the papers setting out precise numerical assumptions or results.

Taking exhibition a little further, we often need to **explain** data (information icon) by setting it in context. This could just be showing the prior set of assumptions or results alongside for comparison, or taking it a stage further, a more formal analysis of change / surplus. There is often a strong **narrative** element to these analyses (such as "surplus has increased through strong investment management performance, offset slightly by a strengthening of reserves").

At the other end, a precise exhibition of the data may be less important; rather it may be useful to **explore** the data (map icon), to assist users to discover relationships, connections and other insights. Exploration is perhaps where the greatest **novelty** is to be found.

It is unusual for the same visualisation to score highly for all purposes; precision and simplicity often conflict with the methods used to explore relationships. As such, hybrids of different visualisation methods can be useful if the users' needs span a range of purposes.

**Choosing a data visualisation method requires an understanding of the purpose of the visualisation, as different methods are suited to different purposes.**

[2] Several authors of books on visualisation take a similar approach, although they don't all use the same terminology.

## 3.3 Method

The primary question when selecting a method to present some data is whether to set it out in a **table**, or a **chart** of some kind, or a **combination**.

We regard tables as a form of data visualisation, and do not focus exclusively on charts and other graphics. This is because much actuarial work involves presenting numerical assumptions or results, and a table is an excellent way of **exhibiting** such data, particularly if some straightforward principles are followed. Further, with a little care, tables can also set data in context and help to **explain** it.

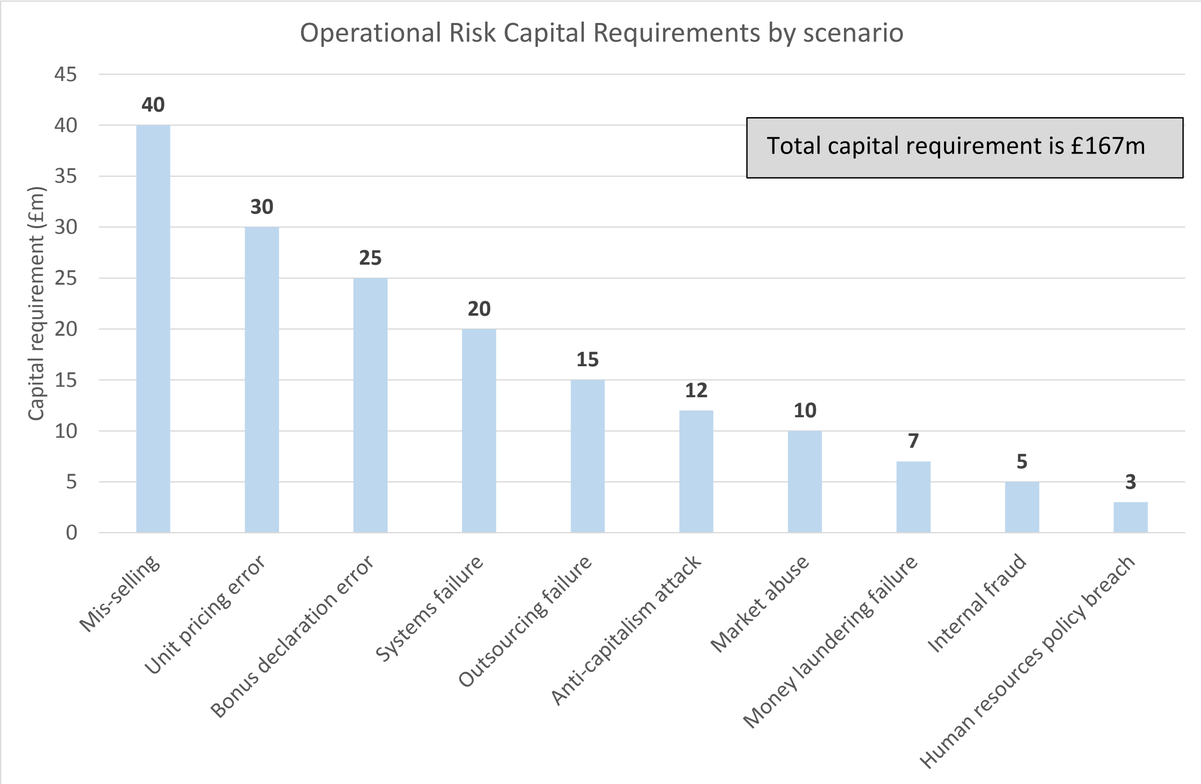
A fundamental use of a table is to allow **individual values** to be **looked up**, using row and/or column headings as guides. Here we give an operational risk example, in which we show capital requirements for a range of scenarios[3]:

| Scenario | Capital Requirement |
|---|---:|
| Anti-capitalism attack | 12 |
| Bonus declaration error | 25 |
| Human resources policy breach | 3 |
| Internal fraud | 5 |
| Market abuse | 10 |
| Mis-selling | 40 |
| Money laundering failure | 7 |
| Outsourcing failure | 15 |
| Systems failure | 20 |
| Unit pricing error | 30 |
| **Total (undiversified)** | **167** |

One use of this table is to say what the capital requirement is for a particular scenario: we read down the left-hand side until we find the scenario we want, and then look across to get the capital requirement. The scenarios have been sorted in alphabetical order, which promotes this kind of use. Arguably a more useful arrangement would be to sort in descending order of capital requirements, so that the largest scenarios were at the top, although the look-up would then operate in the other direction (ie "what's the second-largest scenario?" rather than "what's the capital requirement for Unit pricing error?").

[3] Scenario analysis is a reasonably common way to model operational risk. Each scenario represents a real-life situation which could occur in practice, and business experts are asked to assess the frequency of loss events for that scenario, and the severity of losses if a loss event occurred. Scenario analysis is usually augmented by data on historical loss events, either statistically or through having regard to past losses when assessing scenario frequency and severity. The operational risk examples in this section are somewhat contrived to allow us to make points about tables and charts.

In general, charts are more useful for **explaining** and **exploring** data than for exhibition, although the use of numerical labels on charts can also make them useful for **exhibiting** data, too. Indeed, in some cases, a chart with data labels can take the place of a table:



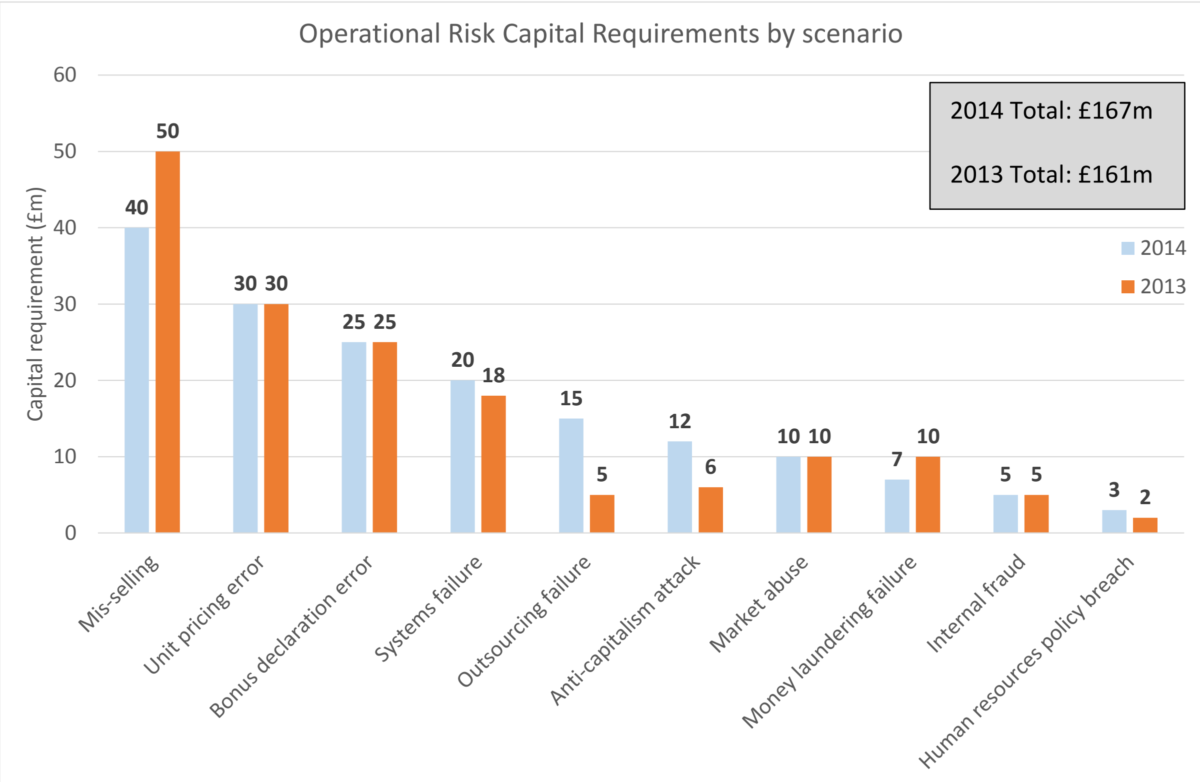

The only thing in the table of scenarios above that we didn't want to show directly in the chart was the total capital requirement, and so we included it as an annotation. (Including totals in bar charts is generally a bad idea as the total dominates the other elements and makes it harder to see patterns and trends.)

As well as **exhibiting** the data, this chart allows us to see **patterns**, for example that there is a fairly steady progression from highest to lowest with no jumps. Might this indicate that a degree of "smoothing" of capital requirements has gone on? This could also be seen from the table if the values were sorted, but the whole pattern can be seen at once in the chart, whereas in the table we would have to look at all the numbers.

Tables also permit **comparison of values**, by looking up values for different row and/or column headings. So we could compare the operational risk capital requirements with the previous year's values:

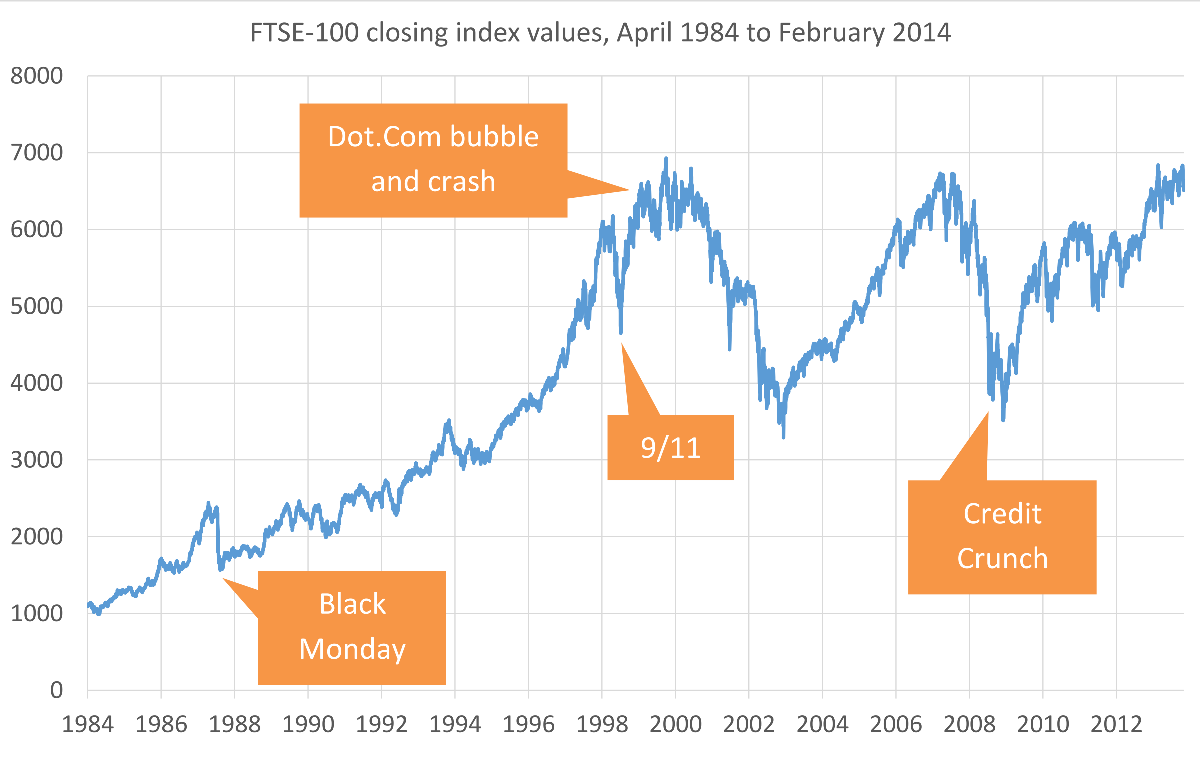
| **Scenario** | **2013** | **2014** | **YoY change** |
|---|---|---|---|
| Mis-selling | 50 | 40 | -10 |
| Unit pricing error | 30 | 30 | 0 |
| Bonus declaration error | 25 | 25 | 0 |
| Systems failure | 18 | 20 | 2 |
| Outsourcing failure | 5 | 15 | 10 |
| Anti-capitalism attack | 6 | 12 | 6 |
| Market abuse | 10 | 10 | 0 |
| Money laundering failure | 10 | 7 | -3 |
| Internal fraud | 5 | 5 | 0 |
| Human resources policy breach | 2 | 3 | 1 |
| **Total (undiversified)** | **161** | **167** | **6** |

We can also show this in a chart with data labels:




Both the table and chart work reasonably well in comparing year-on-year values.

The row and column headings can form an overall series (by date, for example), but if there are lots of rows and/or columns, a chart may well be preferable:




Here we show the FTSE-100 index between April 1984 and February 2014. We would rarely show the data underlying this kind of chart in a table, because the **trends** in the sequence of values are far **more important than** the **precise values** of the index[4]. The ordering of the data is also important here; the natural sequence of the dates means we don't have the freedom to sort that we had with categorical data like the operational risk scenarios.

Both tables and charts are often enhanced by **narrative** commentary. Embedding focused narrative messages in an additional table column or through text boxes placed on a chart (as in the FTSE example above) is often helpful, although it can clutter the presentation if overdone. Bullet points below the table or chart can be used if cluttering is an issue or to make secondary points.

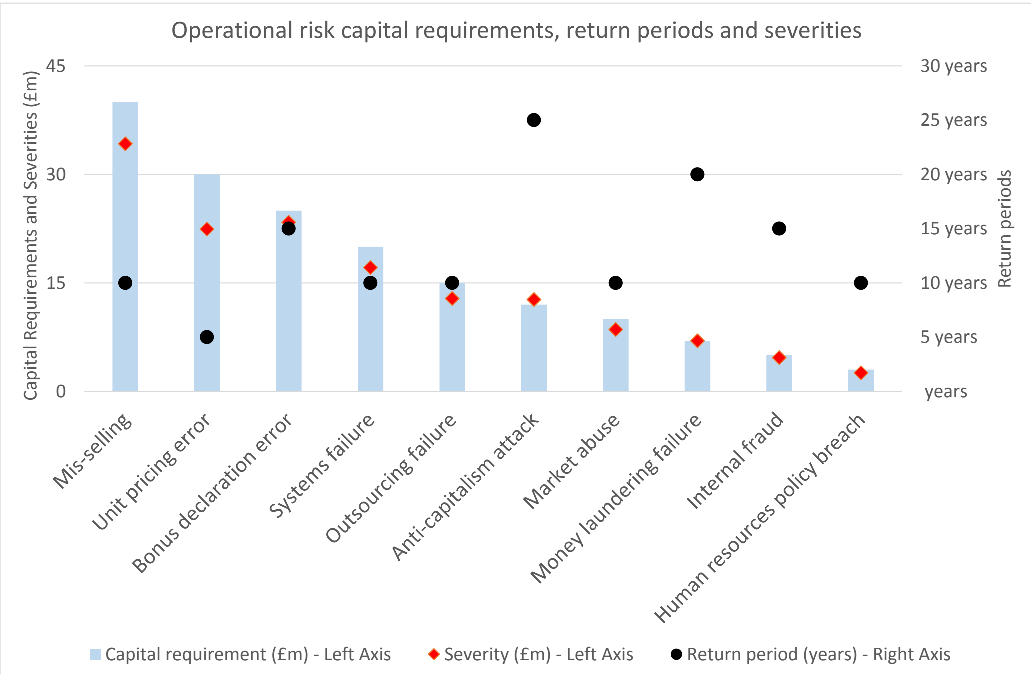
The FTSE example above highlights that charts are to be preferred when the overall **pattern** or **trend** of results is important. Charts are also useful to highlight **outliers** and other **exceptions**, and where comparisons are between whole series of data rather than individual values. (Section 5.3 discusses how best to display multiple equity indexes on the same chart.)

[4] A table of all the data would be very large, too, although we could just show annual or monthly values.

A table is usually preferable when there are fundamentally different kinds of number to be shown next to each other. We would have no qualms showing operational risk frequency and severity assumptions alongside capital requirements in a single table, for example:

| Scenario | Return period (years) | Severity (£m) | Capital requirement (£m) |
|---|---|---|---|
| Mis-selling | 10 | 34 | 40 |
| Unit pricing error | 5 | 22 | 30 |
| Bonus declaration error | 15 | 23 | 25 |
| Systems failure | 10 | 17 | 20 |
| Outsourcing failure | 10 | 13 | 15 |
| Anti-capitalism attack | 25 | 13 | 12 |
| Market abuse | 10 | 9 | 10 |
| Money laundering failure | 20 | 7 | 7 |
| Internal fraud | 15 | 5 | 5 |
| Human resources policy breach | 10 | 3 | 3 |
| Total (undiversified) | | | 167 |

Showing different kinds of number in a chart can be done, but requires a fair amount of care, and even then can be complicated, perhaps overly so. In the table immediately above, there are three kinds of value, two of which are monetary and broadly comparable with each other, so we can show them on the same y-axis. The frequency assumptions are quite different, so we need a secondary y-axis:



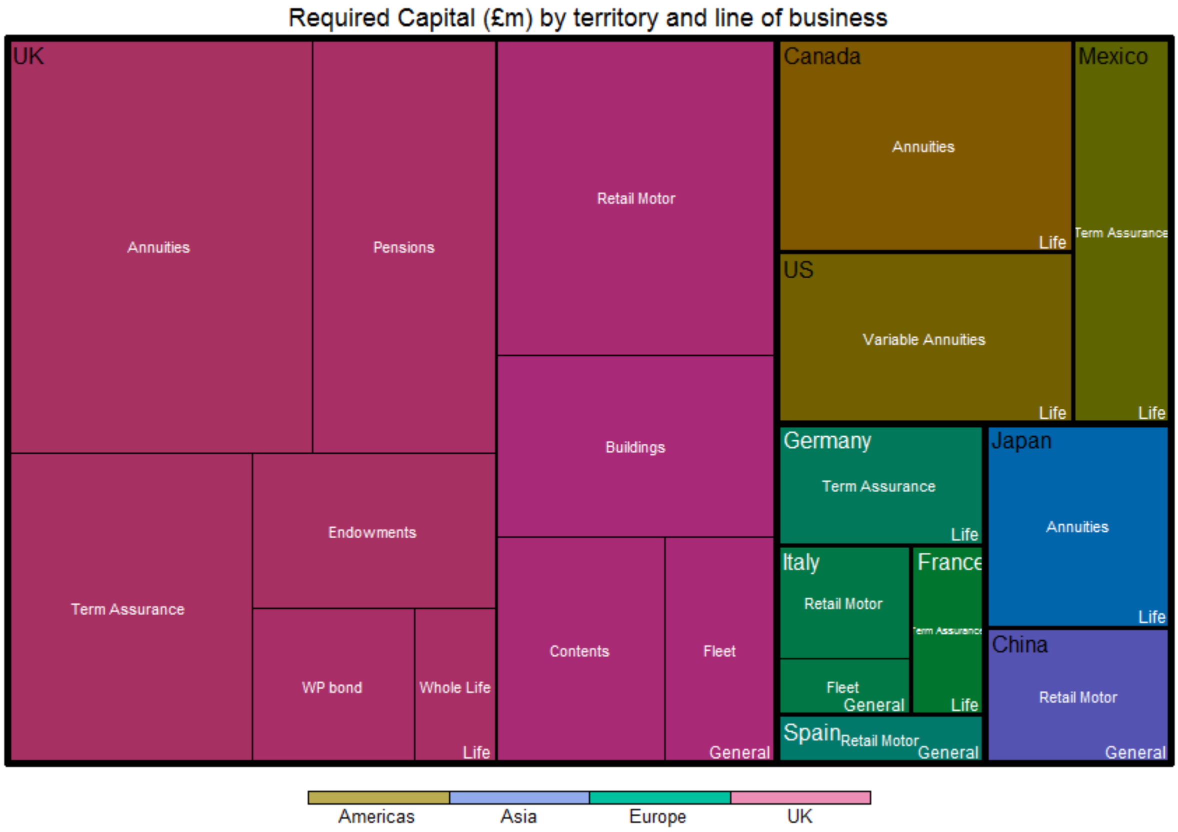
Despite the complication, this chart does allow us to see that capital requirements (blue bars) are not too dissimilar to the severity assumptions (red diamonds), and that where two scenarios have similar severities, the scenario with the shorter return period has the higher capital requirement – for example, compare Anti-capitalism attack with Market abuse. We would really struggle to show a table with 3 fundamentally different kinds of values in a chart, however, as we can't have 3 vertical axes (and even 2 vertical axes is pushing it).

Tables are particularly useful where there is a need to show data at different levels of detail, because we can group the row and/or column headings, and use sub-totals. In the example below we show (contrived) capital requirements for a composite multinational insurance group, broken down by country and line of business. As we noted earlier, this wouldn't work as well with a bar or column chart, because totals and sub-totals tend to be large in relation to individual elements.

| **Required Capital (£m)** | **Americas** | | | **Asia** | | **Europe** | | | | **UK** | **Total** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | **CAN** | **MEX** | **USA** | **CHN** | **JPN** | **FRA** | **GER** | **ITL** | **SPN** | | |
| **General** | | | | **10** | | | | **9** | **4** | **80** | **103** |
| Business | | | | | | | | | | 20 | 20 |
| Contents | | | | | | | | | | 15 | 15 |
| Fleet | | | | | | | | 3 | | 10 | 13 |
| Retail Motor | | | | 10 | | | | 6 | 4 | 35 | 55 |
| | | | | | | | | | | | |
| **Life** | **25** | **15** | **20** | | **15** | **5** | **10** | | | **140** | **230** |
| Annuities | 25 | | | | 15 | | | | | 50 | 90 |
| Endowments | | | | | | | | | | 15 | 15 |
| Pensions | | | | | | | | | | 30 | 30 |
| Term Assurance | | 15 | | | | 5 | 10 | | | 30 | 60 |
| Var. Annuities | | | 20 | | | | | | | | 20 |
| Whole Life | | | | | | | | | | 5 | 5 |
| WP bond | | | | | | | | | | 10 | 10 |
| | | | | | | | | | | | |
| **Total** | **25** | **15** | **20** | **10** | **15** | **5** | **10** | **9** | **4** | **220** | **333** |

"Pivot tables" available in spreadsheets and database packages are a very structured way of arranging data using different combinations and groupings of row and column headings. A pivot table can permit rapid experimentation with different layouts and levels of detail. Having established a good layout, the data can be formatted for presentation in a report. (Pivot tables are typically weaker at formatting.) Excel also has a notion of "pivot charts" which can similarly allow rapid prototyping of different charts and levels of detail, although typically even more effort needs to be put into formatting the results.

A possibly less familiar kind of chart called a **treemap**[5] can be used to visualise hierarchical data such as the capital requirements above, provided all the elements are positive. Elements are shown as rectangles, which are subdivided as the level of detail increases. In the example treemap below, which shows the same data as in the table of required capital above, there are four overall rectangles for the territories of UK, Europe, Americas, and Asia, within which are rectangles for country, then Life and General, and then for specific lines of business. The size of each rectangle is proportional to the capital requirement contained within it (subtotalled and totalled as the level of detail reduces).




*Produced using the 'treemap' library in the statistical package R.*

We can easily tell from the treemap that UK Life requires somewhat more capital than Americas, Europe and Japan combined, and that Asia and Europe require about the same amount of capital. We could do the same with the table, but it would require more subtotals, and be harder to fit on the page. That said, the table does exhibit precise values.

In this example, the colours are used to reinforce the location within the group, but it would also be possible to use colour to show another dimension, such as realistic reserves or capital intensity (required capital as a proportion of reserves).

[5] The treemap was invented by Ben Shneiderman, who wanted an easy way to locate the large files consuming his hard disk space - a file system is even more hierarchical than an insurance group's balance sheet. See www.cs.umd.edu/hcil/treemap-history/ for a discussion of the development of treemaps.

To summarise the discussion of tables and charts:

| Use tables when | Use charts when |
| --- | --- |
| 1. Individual values will be looked up or compared<br>2. Only individual values will be compared, rather than whole series of values<br>3. Precise values are required<br>4. Values involve more than one unit of measure<br>5. Values must be presented at various levels of aggregation (i.e. summary and detail) | 1. The message is contained in patterns, trends, and exceptions<br>2. Entire series of values must be seen in a whole and/or compared<br>3. To explore relationships, connections and associations<br><br>*Source: Adapted from 'Show me the numbers' (Stephen Few 2012)* |

Exhibit ⟷ Explain ⟷ Explore

**It will often make sense to use both tables and charts to display the same information**

Finally, when choosing a method, be aware that a visualisation is a **model** of the data, and many of the general considerations about models apply to visualisations too. We highlight two points:

- First, within reason models/visualisations should be selective, rather than showing every last detail – the aphorism "the map is not the territory"[6] is often employed to make this point.
- Second, all models have limitations, and when developing a visualisation one should be comfortable that its limitations are acceptable in the context of its use.

**A visualisation is in effect a model of the data being visualised, and standard advice about models applies. In particular, be selective where necessary ("the map is not the territory"), and be aware of limitations of the model / visualisation.**

For a more detailed discussion of modelling issues, please see "The Philosophy of Modelling" by Matthew Edwards and Zaid Hoosain, SIAS 2012, available from www.sias.org.uk/data/meetings/SIASMeetingJune2012/attachment/at_download.

[6] Other than in Blackadder Goes Forth.

## 3.4 Context of presentation

Every visualisation is presented within a context, namely how it will be presented and to whom, and therefore the effectiveness of a visualisation depends on how well it suits this context, just as much as its inherent or technical quality. Aspects of the context include:

- **The audience** – for example the level of knowledge or familiarity with the subject in hand. A stunning visualisation from the perspective of subject matter experts may be way over the heads of an audience of novices (or may need a number of intermediate visualisations as a "build up" for the benefit of the latter group).
- **The medium of presentation** – for example the size and density of text used will depend materially on whether the visualisation is to be presented in hard copy form or shown on a projector screen, likewise the potential to use animation or other forms of interaction; the types and range of shading that can be used will depend on whether the material will be shown in colour or black and white.
- **The method of presentation** – principally, whether the visualisation will be accompanied by a verbal description, text description, or whether it needs to stand alone.

In many cases allowance needs to be made for the same visualisation to be appropriate to multiple contexts – for example a presentation which is initially accompanied by verbal description may be used on a standalone basis thereafter; a chart originally printed in colour may end up being photocopied in black and white.

Whilst attempts to completely "future proof" a visualisation could take up inordinate amounts of time and detract from the primary purpose, some thought should be given to these issues, and if there are any straightforward changes available to make the visualisation work in other contexts, these should be made.

| **Ensure that a visualisation is suited to the context within which it will be presented.** |
| --- |

# 4 Case Study 1: Correlation Matrix

## 4.1 Introduction

Suppose you work in the risk management area of a large composite insurance group[8], and have been given the task of presenting a correlation matrix[9] for governance approval.

The Group writes life and non-life business in the UK, and has 25 different risk factors[10], categorised as follows:

| | |
|---|---|
| Market risk | Equity returns |
| | Credit spreads |
| | Interest rates – parallel shifts |
| | Interest rates – twists |
| | Interest rates – curvature |
| | Property returns |
| Credit risk | Corporate bond default / migration |
| | Other counterparty |
| Life insurance risk | Mortality/longevity |
| | Morbidity |
| | Lapse |
| Non-life insurance risk | Motor |
| | Home |
| | Legal |
| | Travel |
| | Liability |
| Expense | Current expenses |
| | Expense inflation |
| Operational | Internal fraud |
| | External fraud |
| | Employment Practices |
| | Business Practice |
| | Physical Asset Damage |
| | System Failures |
| | Execution Management |

[8] While we use the concrete example of a composite insurance group, very similar issues would apply to pension schemes, asset managers, and many other kinds of financial institution.

[9] For those unfamiliar with correlation matrices, in very broad terms a correlation matrix sets out the strength of the relationship between each pair of risks, with values ranging from +1 (risks are perfectly dependent on each other) to 0 (independence) to -1 (perfect negative dependence). Correlation matrices are used in economic capital models and govern the amount of diversification benefit available. All correlation matrices are symmetrical (the correlation between risks A and B is the same as between B and A), and all the entries on the leading diagonal are 1 (a risk is perfectly correlated with itself). Thus there are (625 –25) / 2 = 300 separate numerical assumptions to be set in our 25x25 matrix, all of which need to be precisely set out for governance approval, although some of them have a more material impact on economic capital than others. Be thankful you only have to present the assumptions, not perform the analyses required to set them all!

[10] These are not intended to be entirely realistic, but to provide a large enough set of risk factors to make the challenges of dealing with correlation matrices apparent. In practice, much larger correlation matrices are used in some organisations.

Correlation matrices have to be "internally consistent" and not contain combinations of correlations which in some sense contradict each other[11]. The Group's matrix is somewhat inconsistent and needs to be repaired – this is likely to lead to many of the values changing to a degree.

We work through the challenge of presenting the matrix for approval, following the principles we have already established, and drawing out some further principles.

## 4.2 Structured top-down approach

Mapping this situation onto our overall structured approach gives the following:

**Purpose: Know your audience and what they are looking for**

Our audience is a governance body (let's call it a Technical Committee). To satisfy the Group's governance requirements, they will have to see all the numbers, before and after adjustments for internal consistency. Not all members of the Committee fully understand the finer points of the statistics behind correlation matrices, and it is unlikely that the Committee will relish the opportunity to wade through 300 separate assumptions. So we're going to have to prioritise and ensure that focus is given to the most important aspects, and convince them that there are no other important aspects that should also have been considered.

**Narrative: Determine the messages the audience will find most useful**

We think that the following messages will be useful:

- Which assumptions deserve most attention (either through materiality or contentiousness)
- The internal-consistency adjustments are relatively small and do not have a material financial effect.
- There are no "silly mistakes" in the assumptions (given the tedious and complex nature of setting correlation assumptions, it is entirely possible for mistakes to creep in).

[11] If risks A and B are perfectly correlated, and so are risks B and C, then A and C must also be perfectly correlated, and if not there is an inconsistency; more subtle examples exist. The mathematical expression of internal consistency is that the matrix must be positive semi-definite; this is discussed in many textbooks, for example "Quantitative Risk Management" (McNeil, Embrechts and Frey, 2005).

**Method, Design, Implementation: Select the best methods to display your messages, design the display, and choose a technology**

We can't start to design without understanding what the method looks like, so let's simply put the matrix down on paper as a table:

| | | Market risk | | | | | | Credit risk | | Life insurance risk | | | Non-life insurance risk | | | | | Expense risk | | Operational risk | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Equity returns | Credit spreads | Interest rates – parallel shifts | Interest rates – twists | Interest rates – curvature | Property returns | Corporate bond default / migration | Other counterparty | Mortality/longevity | Morbidity | Lapse | Motor | Home | Legal | Travel | Liability | Current expenses | Expense inflation | Internal fraud | External fraud | Employment Practices | Business Practice | Physical Asset Damage | System Failures | Execution Management |
| Market risk | Equity returns | 1.00 | -0.75 | -0.25 | 0.25 | 0.00 | 0.50 | -0.50 | -0.25 | 0.00 | 0.00 | -0.50 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.25 | 0.25 | -0.25 | -0.25 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| | Credit spreads | -0.75 | 1.00 | 0.50 | -0.25 | 0.00 | -0.50 | 0.50 | 0.25 | 0.00 | 0.00 | 0.50 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | -0.25 | -0.25 | 0.25 | 0.25 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| | Interest rates – parallel shifts | -0.25 | 0.50 | 1.00 | 0.10 | 0.10 | -0.25 | 0.50 | 0.50 | 0.00 | 0.00 | -0.25 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| | Interest rates – twists | 0.25 | -0.25 | 0.10 | 1.00 | 0.10 | 0.25 | 0.25 | 0.25 | 0.00 | 0.00 | 0.10 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| | Interest rates – curvature | 0.00 | 0.00 | 0.10 | 0.10 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| | Property returns | 0.50 | -0.50 | -0.25 | 0.25 | 0.00 | 1.00 | -0.50 | -0.25 | 0.00 | 0.00 | -0.50 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Credit risk | Corporate bond default / migration | -0.50 | 0.50 | 0.50 | 0.25 | 0.00 | -0.50 | 1.00 | 0.25 | 0.00 | 0.00 | -0.50 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.25 | 0.25 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| | Other counterparty | -0.25 | 0.25 | 0.50 | 0.25 | 0.00 | -0.25 | 0.25 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.25 | 0.25 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Life insurance risk | Mortality/longevity | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| | Morbidity | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| | Lapse | -0.50 | 0.50 | -0.25 | 0.10 | 0.00 | -0.50 | -0.50 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.25 | 0.25 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Non-life insurance risk | Motor | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.25 | 0.25 | 0.25 | 0.25 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| | Home | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.25 | 1.00 | 0.25 | 0.25 | 0.25 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| | Legal | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.25 | 0.25 | 1.00 | 0.25 | 0.25 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| | Travel | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.25 | 0.25 | 0.25 | 1.00 | 0.25 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| | Liability | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.25 | 0.25 | 0.25 | 0.25 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Expense risk | Current expenses | 0.25 | -0.25 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.25 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.25 | 0.25 | 0.25 | 0.25 | 0.25 | 0.25 | 0.25 | 0.25 |
| | Expense inflation | 0.25 | -0.25 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.25 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.25 | 1.00 | 0.25 | 0.25 | 0.25 | 0.25 | 0.25 | 0.25 | 0.25 |
| Operational risk | Internal fraud | -0.25 | 0.25 | 0.00 | 0.00 | 0.00 | 0.00 | 0.25 | 0.25 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.25 | 0.25 | 1.00 | 0.50 | 0.50 | 0.50 | 0.50 | 0.50 | 0.50 |
| | External fraud | -0.25 | 0.25 | 0.00 | 0.00 | 0.00 | 0.00 | 0.25 | 0.25 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.25 | 0.25 | 0.50 | 1.00 | 0.50 | 0.25 | 0.50 | 0.50 | 0.50 |
| | Employment Practices | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.25 | 0.25 | 0.50 | 0.50 | 1.00 | 0.50 | 0.50 | 0.25 | 0.50 |
| | Business Practice | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.25 | 0.25 | 0.50 | 0.25 | 0.50 | 1.00 | 0.50 | 0.50 | 0.25 |
| | Physical Asset Damage | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.25 | 0.25 | 0.50 | 0.50 | 0.50 | 0.50 | 1.00 | 0.50 | 0.50 |
| | System Failures | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.25 | 0.25 | 0.50 | 0.50 | 0.25 | 0.50 | 0.50 | 1.00 | 0.50 |
| | Execution Management | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.25 | 0.25 | 0.50 | 0.50 | 0.50 | 0.25 | 0.50 | 0.50 | 1.00 |

Probably the only good thing we can say about the table above is that it does succeed in exhibiting all the assumptions, so as a method it has achieved something. However, in terms of design, it isn't at all clear and simple, in fact it looks very complicated and off-putting, with the only concession to guiding the reader being the use of slightly heavier lines to group the assumptions into categories (market risk, credit risk, etc). Some further criticisms, and proposals to address them, are in the table below:

| **Criticism** | **Possible remedy** |
|---|---|
| All values are formatted to 2 decimal places, which makes them all look very similar to each other. | We need 2 decimal places for the 0.25 and 0.75 values. However, we could show the zeros as 0 rather than 0.00, and retain 2 decimal places for the non-zero values. Using 1 decimal place for 0.5 is probably confusing, because 0.5 will appear visually to be somewhere between 0 and 0.25 whereas numerically it isn't. So we'll use 2 decimal places for non-zero numbers, and no decimal places for zeros. This should make the zeros and non-zero values stand out from each other. |
| It's hard to see which values are negative, because minus signs don't use much ink/many pixels, and so don't stand out. | Use () rather than minus signs to present negative values. |
| The numerically large correlations (0.5/0.75 and negatives) don't really stand out. | Use bold formatting to highlight any correlations whose absolute values are equal to or greater than 0.5. |
| The symmetry of the matrix leads to repetition, in that the upper-right triangle has the same values as the lower-left triangle. | Remove the upper-right triangle. |
| The values on the leading diagonal are always going to be 1.0, and don't carry any information as such, but look important. | Remove them. |
| If we look at a correlation towards the middle of the matrix, it's hard to see what risks it relates to, ie to look up the row and column headings. | Shade alternate rows and columns. |
| The heavier lines grouping assumptions into categories are reasonably useful, but it isn't particularly easy to identify which correlations are within a category, and which are between categories | Use coloured borders to identify blocks of correlations within a category. |
| There is no indication of the financial significance of the risks. | Show pre-diversification capital requirements beside row/column headings. |

An updated presentation addressing these criticisms is below:

| | | | Equity returns | Credit spreads | Interest rates – parallel shifts | Interest rates – twists | Interest rates – curvature | Property returns | Corporate bond default / migration | Other counterparty | Mortality/longevity | Morbidity | Lapse | Motor | Home | Legal | Travel | Liability | Current expenses | Expense inflation | Internal fraud | External fraud | Employment Practices | Business Practice | Physical Asset Damage | System Failures | Execution Management |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | 100 | 50 | 60 | 10 | 5 | 20 | 10 | 5 | 30 | 5 | 40 | 20 | 15 | 10 | 10 | 5 | 5 | 10 | 10 | 5 | 8 | 7 | 4 | 6 | 3 |
| Market risk | Equity returns | 100 | | | | | | | | | | | | | | | | | | | | | | | | | |
| | Credit spreads | 50 | **(0.75)** | | | | | | | | | | | | | | | | | | | | | | | | |
| | Interest rates – parallel shifts | 60 | (0.25) | **0.50** | | | | | | | | | | | | | | | | | | | | | | | |
| | Interest rates – twists | 10 | 0.25 | (0.25) | 0.10 | | | | | | | | | | | | | | | | | | | | | | |
| | Interest rates – curvature | 5 | 0 | 0 | 0.10 | 0.10 | | | | | | | | | | | | | | | | | | | | | |
| | Property returns | 20 | **0.50** | **(0.50)** | (0.25) | 0.25 | 0 | | | | | | | | | | | | | | | | | | | | |
| Credit risk | Corporate bond default / migration | 10 | **(0.50)** | **0.50** | **0.50** | 0.25 | 0 | **(0.50)** | | | | | | | | | | | | | | | | | | | |
| | Other counterparty | 5 | (0.25) | 0.25 | **0.50** | 0.25 | 0 | (0.25) | 0.25 | | | | | | | | | | | | | | | | | | |
| Life insurance risk | Mortality/longevity | 30 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | | | | | | | | | | | | | | | | | |
| | Morbidity | 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | | | | | | | | | | | | | | | | |
| | Lapse | 40 | **(0.50)** | **0.50** | (0.25) | 0.10 | 0 | **(0.50)** | **(0.50)** | 0 | 0 | 0 | | | | | | | | | | | | | | | |
| Non-life insurance risk | Motor | 20 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | | | | | | | | | | | | | | |
| | Home | 15 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.25 | | | | | | | | | | | | | |
| | Legal | 10 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.25 | 0.25 | | | | | | | | | | | | |
| | Travel | 10 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.25 | 0.25 | 0.25 | | | | | | | | | | | |
| | Liability | 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.25 | 0.25 | 0.25 | 0.25 | | | | | | | | | | |
| Expense risk | Current expenses | 5 | 0.25 | (0.25) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.25 | 0 | 0 | 0 | 0 | 0 | | | | | | | | | |
| | Expense inflation | 10 | 0.25 | (0.25) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.25 | 0 | 0 | 0 | 0 | 0 | 0.25 | | | | | | | | |
| Operational risk | Internal fraud | 10 | (0.25) | 0.25 | 0 | 0 | 0 | 0 | 0.25 | 0.25 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.25 | 0.25 | | | | | | | |
| | External fraud | 5 | (0.25) | 0.25 | 0 | 0 | 0 | 0 | 0.25 | 0.25 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.25 | 0.25 | **0.50** | | | | | | |
| | Employment Practices | 8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.25 | 0.25 | **0.50** | **0.50** | | | | | |
| | Business Practice | 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.25 | 0.25 | **0.50** | 0.25 | **0.50** | | | | |
| | Physical Asset Damage | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.25 | 0.25 | **0.50** | **0.50** | **0.50** | **0.50** | | | |
| | System Failures | 6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.25 | 0.25 | **0.50** | **0.50** | 0.25 | **0.50** | **0.50** | | |
| | Execution Management | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.25 | 0.25 | **0.50** | **0.50** | **0.50** | 0.25 | **0.50** | **0.50** | |

We've shown the undiversified capital requirements both numerically and graphically. The bars in each row are produced using Excel conditional formatting data bars, and the bars in each column use Excel sparklines. (The capital requirements are entirely contrived for illustrative purposes, and are not intended to be realistic.) Now we're starting to get somewhere: we can see which correlations relate to financially significant risks, and we can also start to see some patterns and relationships, for example:

- There are lots of zero correlations (which are significant in their own right) and we can very easily distinguish where correlations are zero or non-zero purely by the typography.
- The non-zero correlations seem to have been set primarily within the coloured blocks, with comparatively few non-zero correlations between risks in different blocks.
- Non-life risk is moderately correlated in itself, but life risk isn't.
- Risks within the operational risk category are fairly heavily correlated with each other (lots of bold), but operational risk isn't terribly financially significant.

All of these points may be quite reasonable for this particular Group, but the presentation of the matrix brings them out as areas of potential challenge by the Technical Committee.

This improved presentation may be enough, although we could continue to adjust it:

- We could retain the blocks and sort the risks within each block in descending order of financial significance. Or we could order all the risks by financial significance, so that the top-left corner involved the most significant risks. (On balance, we prefer keeping the block structure, because the way correlations are set in practice often involves different teams who might work on separate blocks.)
- We could apply the colours to a greater extent within each block, for example colouring the headings, the numerical correlations, or the shaded row/column guides. (We haven't done this here in case the result looked too colourful: colouring the bars/columns and the borders seems to us to be a sufficient visual cue, although ultimately this is a matter of taste.)
- We could present and discuss the matrix block-by-block, starting with the "intra-category" correlations with the coloured borders, and moving on to "cross-category blocks". In practice, we'd probably do this as well as showing the whole matrix. However, block-by-block presentation doesn't allow an overall sense of the whole matrix to be formed.

The changes we've made have mostly been about **design**; our overall **method** remains to exhibit the numbers, albeit with more **context** now, by showing the capital requirements. However, we can develop the method by using the blank space in the upper right corner to visualise the correlations in a stronger way than the decimal places and bold formatting:

| | | | Equity returns | Credit spreads | Interest rates – parallel shifts | Interest rates – twists | Interest rates – curvature | Property returns | Corporate bond default / migration | Other counterparty | Mortality/longevity | Morbidity | Lapse | Motor | Home | Legal | Travel | Liability | Current expenses | Expense inflation | Internal fraud | External fraud | Employment Practices | Business Practice | Physical Asset Damage | System Failures | Execution Management |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | 100 | 50 | 60 | 10 | 5 | 20 | 10 | 5 | 30 | 5 | 40 | 20 | 15 | 10 | 10 | 5 | 5 | 10 | 10 | 5 | 8 | 7 | 4 | 6 | 3 |
| Market risk | Equity returns | 100 | | | | | | | | | | | | | | | | | | | | | | | | | |
| | Credit spreads | 50 | **(0.75)** | | | | | | | | | | | | | | | | | | | | | | | | |
| | Interest rates – parallel shifts | 60 | (0.25) | **0.50** | | | | | | | | | | | | | | | | | | | | | | | |
| | Interest rates – twists | 10 | 0.25 | (0.25) | 0.10 | | | | | | | | | | | | | | | | | | | | | | |
| | Interest rates – curvature | 5 | 0 | 0 | 0.10 | 0.10 | | | | | | | | | | | | | | | | | | | | | |
| | Property returns | 20 | **0.50** | **(0.50)** | (0.25) | 0.25 | 0 | | | | | | | | | | | | | | | | | | | | |
| Credit risk | Corporate bond default / migration | 10 | **(0.50)** | **0.50** | **0.50** | 0.25 | 0 | **(0.50)** | | | | | | | | | | | | | | | | | | | |
| | Other counterparty | 5 | (0.25) | 0.25 | **0.50** | 0.25 | 0 | (0.25) | 0.25 | | | | | | | | | | | | | | | | | | |
| Life insurance risk | Mortality/longevity | 30 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | | | | | | | | | | | | | | | | | |
| | Morbidity | 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | | | | | | | | | | | | | | | | |
| | Lapse | 40 | **(0.50)** | **0.50** | (0.25) | 0.10 | 0 | **(0.50)** | **(0.50)** | 0 | 0 | 0 | | | | | | | | | | | | | | | |
| Non-life insurance risk | Motor | 20 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | | | | | | | | | | | | | | |
| | Home | 15 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.25 | | | | | | | | | | | | | |
| | Legal | 10 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.25 | 0.25 | | | | | | | | | | | | |
| | Travel | 10 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.25 | 0.25 | 0.25 | | | | | | | | | | | |
| | Liability | 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.25 | 0.25 | 0.25 | 0.25 | | | | | | | | | | |
| Expense risk | Current expenses | 5 | 0.25 | (0.25) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.25 | 0 | 0 | 0 | 0 | 0 | | | | | | | | | |
| | Expense inflation | 10 | 0.25 | (0.25) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.25 | 0 | 0 | 0 | 0 | 0 | 0.25 | | | | | | | | |
| Operational risk | Internal fraud | 10 | (0.25) | 0.25 | 0 | 0 | 0 | 0 | 0.25 | 0.25 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.25 | 0.25 | | | | | | | |
| | External fraud | 5 | (0.25) | 0.25 | 0 | 0 | 0 | 0 | 0.25 | 0.25 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.25 | 0.25 | **0.50** | | | | | | |
| | Employment Practices | 8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.25 | 0.25 | **0.50** | **0.50** | | | | | |
| | Business Practice | 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.25 | 0.25 | **0.50** | 0.25 | **0.50** | | | | |
| | Physical Asset Damage | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.25 | 0.25 | **0.50** | **0.50** | **0.50** | **0.50** | | | |
| | System Failures | 6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.25 | 0.25 | **0.50** | **0.50** | 0.25 | **0.50** | **0.50** | | |
| | Execution Management | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.25 | 0.25 | **0.50** | **0.50** | **0.50** | 0.25 | **0.50** | **0.50** | |

This is a reasonably standard statistical technique for visualising correlations. The convention is that zeros are represented (appropriately enough) by white circles; positive correlations point from bottom-left to top-right with the shape of the ellipse becoming increasingly narrow as the correlation tends to 1, reaching a diagonal line (/) at +1.0. Conversely, negative correlations point from top-left to bottom-right, with -1.0 being the opposite diagonal line (\).

To reinforce the directional convention, we have coloured positive ellipses in shades of red and negative ellipses in shades of blue (red represents "hot" and blue "cold")[12].

This has been **implemented** using VBA, and shows that with a degree of perseverance it is possible to extend the built-in charting capabilities of Excel. A similar effect can also be achieved using the statistical package R, and in fact we worked with the R implementation before developing the VBA.

The statistical motivation behind the use of ellipses is that each ellipse is the equi-density contour of a Gaussian copula parameterised with the given linear correlation. Regardless of technical explanations, one can grasp the overall convention rapidly, and to start **explore** the data, seeing **patterns** which go beyond the zero/non-zero distinction. For example:

- The top two rows have ellipses which alternate in colour and direction, reflecting the opposing sign conventions for equity returns and credit spreads. (Increases in equity returns are good, increases in credit spreads are bad.)
- The variety of correlations used within market risk stands out, as does the overall uniformity and intensity of the operational risk correlations. Having seen this visualisation, might our Technical Committee start to question the consistency of the process for constructing the correlation matrix?

[12] This is preferable to using red and green because a small but significant proportion of the population suffers from red-green colour blindness – something which should be borne in mind with all data visualisations.

As we noted earlier, making a correlation matrix internally consistent changes many of the values:

| | | | Market risk | | | | | | Credit risk | | Life insurance risk | | | Non-life insurance risk | | | | | Expense risk | | Operational risk | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Equity returns | Credit spreads | Interest rates – parallel shifts | Interest rates – twists | Interest rates – curvature | Property returns | Corporate bond default / migration | Other counterparty | Mortality/longevity | Morbidity | Lapse | Motor | Home | Legal | Travel | Liability | Current expenses | Expense inflation | Internal fraud | External fraud | Employment Practices | Business Practice | Physical Asset Damage | System Failures | Execution Management |
| | | | 100 | 50 | 60 | 10 | 5 | 20 | 10 | 5 | 30 | 5 | 40 | 20 | 15 | 10 | 10 | 5 | 5 | 10 | 10 | 5 | 8 | 7 | 4 | 6 | 3 |
| Market risk | Equity returns | 100 | | | | | | | | | | | | | | | | | | | | | | | | | |
| | Credit spreads | 50 | **(0.80)** | | | | | | | | | | | | | | | | | | | | | | | | |
| | Interest rates – parallel shifts | 60 | (0.23) | 0.47 | | | | | | | | | | | | | | | | | | | | | | | |
| | Interest rates – twists | 10 | 0.21 | (0.20) | 0.08 | | | | | | | | | | | | | | | | | | | | | | |
| | Interest rates – curvature | 5 | 0.00 | (0.00) | 0.10 | 0.10 | | | | | | | | | | | | | | | | | | | | | |
| | Property returns | 20 | **0.54** | **(0.54)** | (0.24) | 0.21 | 0.00 | | | | | | | | | | | | | | | | | | | | |
| Credit risk | Corporate bond default / migration | 10 | (0.43) | 0.42 | **0.53** | 0.17 | 0.00 | (0.43) | | | | | | | | | | | | | | | | | | | |
| | Other counterparty | 5 | (0.24) | 0.25 | 0.50 | 0.24 | 0.00 | (0.24) | 0.27 | | | | | | | | | | | | | | | | | | |
| Life insurance risk | Mortality/longevity | 30 | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | 0.00 | (0.00) | | | | | | | | | | | | | | | | | |
| | Morbidity | 5 | 0.00 | (0.00) | (0.00) | 0.00 | (0.00) | (0.00) | 0.00 | (0.00) | (0.00) | | | | | | | | | | | | | | | | |
| | Lapse | 40 | (0.41) | 0.39 | (0.21) | 0.01 | 0.00 | (0.42) | (0.34) | 0.02 | (0.00) | (0.00) | | | | | | | | | | | | | | | |
| Non-life insurance risk | Motor | 20 | (0.00) | 0.00 | (0.00) | 0.00 | 0.00 | (0.00) | 0.00 | (0.00) | 0.00 | 0.00 | (0.00) | | | | | | | | | | | | | | |
| | Home | 15 | (0.00) | 0.00 | 0.00 | 0.00 | 0.00 | (0.00) | 0.00 | (0.00) | 0.00 | (0.00) | 0.00 | 0.25 | | | | | | | | | | | | | |
| | Legal | 10 | (0.00) | 0.00 | (0.00) | 0.00 | 0.00 | (0.00) | 0.00 | (0.00) | 0.00 | (0.00) | (0.00) | 0.25 | 0.25 | | | | | | | | | | | | |
| | Travel | 10 | 0.00 | 0.00 | (0.00) | 0.00 | 0.00 | (0.00) | 0.00 | (0.00) | 0.00 | (0.00) | (0.00) | 0.25 | 0.25 | 0.25 | | | | | | | | | | | |
| | Liability | 5 | (0.00) | 0.00 | (0.00) | 0.00 | 0.00 | (0.00) | 0.00 | (0.00) | 0.00 | 0.00 | (0.00) | 0.25 | 0.25 | 0.25 | 0.25 | | | | | | | | | | |
| Expense risk | Current expenses | 5 | 0.22 | (0.21) | (0.02) | 0.03 | (0.00) | (0.02) | (0.05) | (0.00) | 0.00 | (0.00) | 0.19 | 0.00 | (0.00) | 0.00 | (0.00) | 0.00 | | | | | | | | | |
| | Expense inflation | 10 | 0.22 | (0.21) | (0.02) | 0.03 | (0.00) | (0.02) | (0.05) | (0.00) | (0.00) | 0.00 | 0.19 | (0.00) | 0.00 | (0.00) | 0.00 | (0.00) | 0.27 | | | | | | | | |
| Operational risk | Internal fraud | 10 | (0.25) | 0.24 | 0.01 | (0.00) | (0.00) | (0.00) | 0.25 | 0.25 | 0.00 | (0.00) | 0.01 | 0.00 | 0.00 | (0.00) | 0.00 | 0.00 | 0.24 | 0.24 | | | | | | | |
| | External fraud | 5 | (0.25) | 0.24 | 0.01 | (0.00) | (0.00) | (0.00) | 0.25 | 0.24 | (0.00) | 0.00 | 0.01 | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | 0.24 | 0.24 | **0.51** | | | | | | |
| | Employment Practices | 8 | 0.00 | 0.00 | (0.00) | (0.00) | 0.00 | 0.01 | 0.01 | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 | (0.00) | 0.00 | (0.00) | 0.25 | 0.25 | 0.50 | 0.50 | | | | | |
| | Business Practice | 7 | 0.00 | (0.01) | 0.00 | (0.00) | 0.00 | 0.00 | 0.01 | 0.00 | (0.00) | 0.00 | 0.01 | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | 0.25 | 0.25 | **0.50** | 0.25 | 0.50 | | | | |
| | Physical Asset Damage | 4 | 0.00 | (0.00) | (0.00) | (0.00) | 0.00 | 0.00 | 0.01 | 0.00 | (0.00) | (0.00) | 0.00 | 0.00 | 0.00 | (0.00) | (0.00) | 0.00 | 0.25 | 0.25 | 0.50 | 0.50 | **0.50** | **0.50** | | | |
| | System Failures | 6 | 0.00 | 0.00 | (0.00) | (0.00) | 0.00 | 0.01 | 0.01 | 0.00 | 0.00 | (0.00) | 0.01 | (0.00) | (0.00) | 0.00 | (0.00) | 0.00 | 0.25 | 0.25 | 0.50 | 0.50 | 0.25 | 0.50 | **0.50** | | |
| | Execution Management | 3 | 0.00 | (0.00) | (0.00) | (0.00) | 0.00 | 0.00 | 0.01 | 0.00 | (0.00) | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.25 | 0.25 | 0.50 | 0.50 | **0.50** | 0.25 | **0.50** | **0.50** | |

After making the matrix internally consistent, we've lost our nice zero/non-zero distinction, because many of the zero values have now become numbers close to but not exactly zero. These are shown as 0.00 to 2 decimal places, although we could pretend they were exactly zero and shown then as 0 in the interest of picking patterns out. However, there are plenty of small values such as 0.01 that used to be zero, so that doesn't get us too far.

The ellipses are still useful for seeing the overall pattern, though, as a small change in the value of the correlation shows up as a small change in the shape of the ellipses. (The same can't be said for the rounding convention, which is brittle in the face of small changes.)

We can also visualise the changes to correlations that have resulted from making the matrix internally consistent, by subtracting the internally-consistent values from the original values:

| | | | Market risk | | | | | | Credit risk | | Life insurance risk | | | Non-life insurance risk | | | | | Expense risk | | Operational risk | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Equity returns | Credit spreads | Interest rates – parallel shifts | Interest rates – twists | Interest rates – curvature | Property returns | Corporate bond default / migration | Other counterparty | Mortality/longevity | Morbidity | Lapse | Motor | Home | Legal | Travel | Liability | Current expenses | Expense inflation | Internal fraud | External fraud | Employment Practices | Business Practice | Physical Asset Damage | System Failures | Execution Management |
| | | | 100 | 50 | 60 | 10 | 5 | 20 | 10 | 5 | 30 | 5 | 40 | 20 | 15 | 10 | 10 | 5 | 5 | 10 | 10 | 5 | 8 | 7 | 4 | 6 | 3 |
| Market risk | Equity returns | 100 | | | | | | | | | | | | | | | | | | | | | | | | | |
| | Credit spreads | 50 | 0.05 | | | | | | | | | | | | | | | | | | | | | | | | |
| | Interest rates – parallel shifts | 60 | (0.02) | 0.03 | | | | | | | | | | | | | | | | | | | | | | | |
| | Interest rates – twists | 10 | 0.04 | (0.05) | 0.02 | | | | | | | | | | | | | | | | | | | | | | |
| | Interest rates – curvature | 5 | (0.00) | 0.00 | 0.00 | 0.00 | | | | | | | | | | | | | | | | | | | | | |
| | Property returns | 20 | (0.04) | 0.04 | (0.01) | 0.04 | (0.00) | | | | | | | | | | | | | | | | | | | | |
| Credit risk | Corporate bond default / migration | 10 | (0.07) | 0.08 | (0.03) | 0.08 | (0.00) | (0.07) | | | | | | | | | | | | | | | | | | | |
| | Other counterparty | 5 | (0.01) | 0.00 | 0.00 | 0.01 | (0.00) | (0.01) | (0.02) | | | | | | | | | | | | | | | | | | |
| Life insurance risk | Mortality/longevity | 30 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | (0.00) | 0.00 | | | | | | | | | | | | | | | | | |
| | Morbidity | 5 | (0.00) | 0.00 | 0.00 | (0.00) | 0.00 | 0.00 | (0.00) | 0.00 | 0.00 | | | | | | | | | | | | | | | | |
| | Lapse | 40 | (0.09) | 0.11 | (0.04) | 0.09 | (0.00) | (0.08) | (0.16) | (0.02) | 0.00 | 0.00 | | | | | | | | | | | | | | | |
| Non-life insurance risk | Motor | 20 | 0.00 | (0.00) | 0.00 | (0.00) | (0.00) | 0.00 | (0.00) | 0.00 | (0.00) | (0.00) | 0.00 | | | | | | | | | | | | | | |
| | Home | 15 | 0.00 | (0.00) | (0.00) | (0.00) | (0.00) | 0.00 | (0.00) | 0.00 | (0.00) | 0.00 | (0.00) | 0 | | | | | | | | | | | | | |
| | Legal | 10 | 0.00 | (0.00) | 0.00 | (0.00) | (0.00) | 0.00 | (0.00) | 0.00 | (0.00) | 0.00 | 0.00 | (0.00) | 0 | | | | | | | | | | | | |
| | Travel | 10 | (0.00) | (0.00) | 0.00 | (0.00) | (0.00) | 0.00 | (0.00) | 0.00 | (0.00) | 0.00 | 0.00 | (0.00) | (0.00) | 0 | | | | | | | | | | | |
| | Liability | 5 | 0.00 | (0.00) | 0.00 | (0.00) | (0.00) | 0.00 | (0.00) | 0.00 | (0.00) | (0.00) | 0.00 | (0.00) | 0.00 | (0.00) | (0.00) | | | | | | | | | | |
| Expense risk | Current expenses | 5 | 0.03 | (0.04) | 0.02 | (0.03) | 0.00 | 0.02 | 0.05 | 0.00 | (0.00) | 0.00 | 0.06 | (0.00) | 0.00 | (0.00) | 0.00 | (0.00) | | | | | | | | | |
| | Expense inflation | 10 | 0.03 | (0.04) | 0.02 | (0.03) | 0.00 | 0.02 | 0.05 | 0.00 | 0.00 | (0.00) | 0.06 | 0.00 | (0.00) | 0.00 | (0.00) | 0.00 | (0.02) | | | | | | | | |
| Operational risk | Internal fraud | 10 | (0.00) | 0.01 | (0.01) | 0.00 | 0.00 | 0.00 | (0.00) | 0.00 | (0.00) | 0.00 | (0.01) | (0.00) | (0.00) | 0.00 | (0.00) | (0.00) | 0.01 | 0.01 | | | | | | | |
| | External fraud | 5 | (0.00) | 0.01 | (0.01) | 0.00 | 0.00 | 0.00 | (0.00) | 0.01 | 0.00 | (0.00) | (0.01) | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.01 | (0.01) | | | | | | |
| | Employment Practices | 8 | (0.00) | (0.00) | 0.00 | 0.00 | (0.00) | (0.01) | (0.01) | (0.00) | (0.00) | (0.00) | (0.01) | (0.00) | (0.00) | 0.00 | (0.00) | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | | | | | |
| | Business Practice | 7 | (0.00) | 0.01 | (0.00) | 0.00 | (0.00) | (0.00) | (0.01) | (0.00) | 0.00 | (0.00) | (0.01) | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | (0.00) | (0.00) | 0.00 | | | | |
| | Physical Asset Damage | 4 | (0.00) | 0.00 | 0.00 | 0.00 | (0.00) | (0.00) | (0.01) | (0.00) | 0.00 | 0.00 | (0.00) | (0.00) | (0.00) | 0.00 | 0.00 | (0.00) | 0.00 | 0.00 | 0.00 | 0.00 | (0.00) | (0.00) | | | |
| | System Failures | 6 | (0.00) | (0.00) | 0.00 | 0.00 | (0.00) | (0.01) | (0.01) | (0.00) | (0.00) | 0.00 | (0.01) | 0.00 | 0.00 | (0.00) | 0.00 | (0.00) | 0.00 | 0.00 | 0.00 | 0.00 | (0.00) | 0.00 | (0.00) | | |
| | Execution Management | 3 | (0.00) | 0.00 | 0.00 | 0.00 | (0.00) | (0.00) | (0.01) | (0.00) | 0.00 | (0.00) | (0.01) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | 0.00 | 0.00 | 0.00 | 0.00 | (0.00) | (0.00) | (0.00) | (0.00) | |

Here the ellipses are the right **shape** for the differences (all of them are nearly circular, indicating that the differences are all quite small), but the **colours** are based on four times the differences in the bottom-left triangle, to make the larger differences stand out visually. We can see that the main differences are with lapse risk, where some correlations have moved by around 10% as a result of eliminating internal inconsistencies, and to a lesser extent with default risk. We're probably not too worried about default risk given that the capital requirements are fairly small, but maybe lapse risk could be reviewed to see if we could justify different assumptions that led to a lower impact of making the matrix internally-consistent – in the interest of space we don't pursue this analysis further.

## 4.3 Conclusions

Let’s take stock of what we’ve learned from this case study. The initial presentation was very weak because we had a massive block of numbers with no easy way to see what was going on. Guided by the principle of trying to bring out things which are important to the recipients, we improved on the initial presentation both by **adding elements** (brackets for negative numbers; bold text for numerically large correlations; financial significance elements; colouring) and, perhaps more importantly, by **removing elements** (upper-right triangle, 1s on the leading diagonal, decimal places for zeros).

In short, we sought to make the important aspects of the data stand out visually, both by emphasising the important and by de-emphasising or removing the unimportant. This process was described eloquently by Edward Tufte in “The Visual Display of Quantitative Information” as follows:

> *‘Just as a good editor of prose ruthlessly prunes out unnecessary words, so a designer of statistical graphics should prune out ink that fails to present fresh data-information.’*

As well as emphasising the connection with written communication, this quotation draws attention to the quantity of ink (or pixels) used in relation to the data or information being presented, which we can assess using what Tufte calls the **‘data : ink ratio’**. Higher values of the data : ink ratio are good either because we have increased the amount of data being shown for a given amount of ink, or reduced the amount of ink required to show a given amount of data. (In more arithmetical terms, if we want to increase a ratio we can either increase the numerator or reduce the denominator.) So we might be tempted to introduce a principle “maximise the data : ink ratio”. However, it’s possible to over-do this, and remove so many supporting elements that a graphic becomes overly minimal and hard to interpret. Thus we add the caveat “within reason”, to arrive at the following very important principle:

> **Maximise the data : ink ratio, within reason.**

In our experience, when combined with trial-and-error, this principle leads to all sorts of improvements; many of the changes made to the initial presentation of the correlation matrix were made directly as a result of applying these two principles.

Expanding on the data : ink principle:

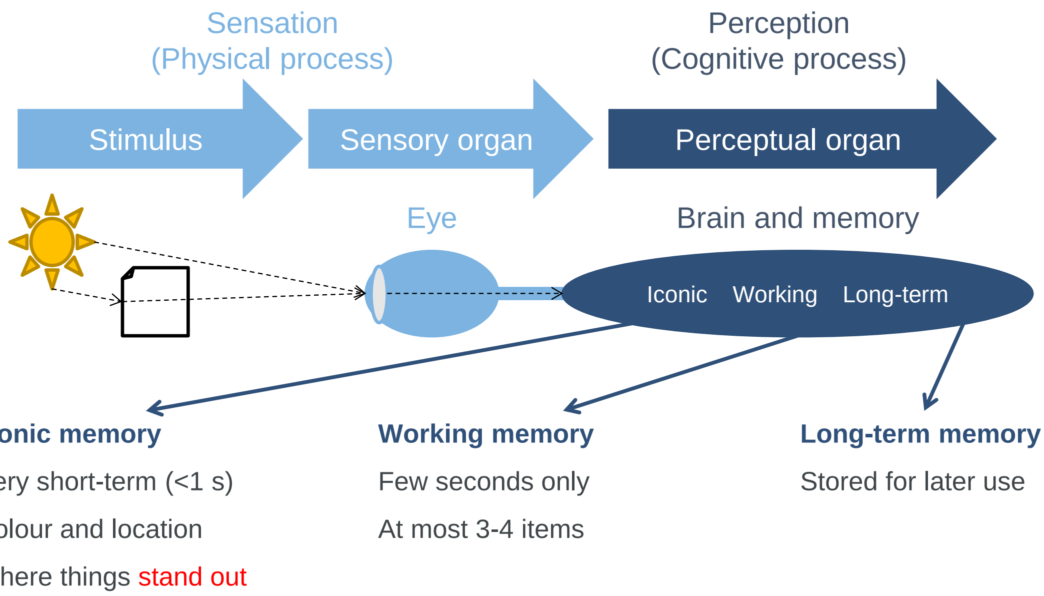
> **Remove everything that isn’t necessary. Identify what is data and what isn’t. Make the data prominent and clear, and make the non-data as unobtrusive as is sensible.**

Non-data are things like axes and grid-lines, which can sometimes be too “busy” – if so, consider removing them, or use a light grey colour so they are still present but less intrusive.

# 5 Design

## 5.1 Perception and Cognition

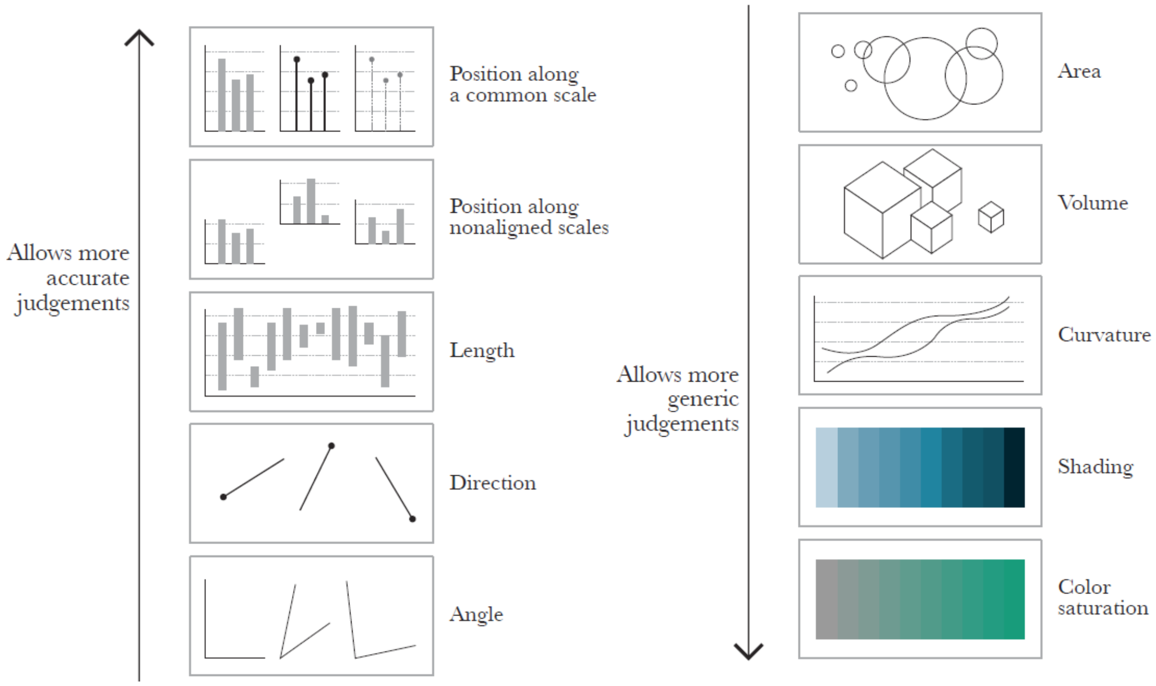
The human brain is very good at seeing patterns, and when designing charts and other visualisations it's useful to have some high-level understanding of how humans perceive and process visual images. The diagram below illustrates some key points:




- The impact of visual stimuli on the eye means nothing until these stimuli are processed by the brain. Put differently, you perceive with your brain, not your eyes.
- Images are initially processed by the brain in 'iconic' memory: this is very short term, and focused on the detection of basic features such as colour, location and movement. The red text stand out probably caught your eye immediately on turning to this page: that's iconic memory in operation.
- Some images then progress to working memory, which lasts for a lot longer than iconic memory, but still only for a few seconds. Working memory performs more detailed analysis and identification. There is a limit to how much can be stored in working memory, typically 3-4 items, perhaps slightly more. One practical implication for data visualisation is that complex charts with legends containing many entries will simply be too much for users to grasp, and they will keep having to go back to the legend to work out which series is which.
- Finally, some images bed down into long-term memory, and can be recalled consciously. This requires a combination of repetition and connections with other ideas already in long-term memory. So new kinds of visualisation can be learned, but it's easier if they relate to things already understood. This means that, for example, a complex new chart in a monthly Board pack will take a few months to become fully understood.

## 5.2 Judgement of Quantity

Some visual aspects are more easily perceived than others. The charts below put a range of visual aspects in rank order by how easily they allow accurate judgements of quantity to be made:




Adapted from 'The Functional Art: An introduction to information graphics and visualisation' by Alberto Cairo.

For example, charts involving lengths (bar and column charts) will allow more accurate judgements than charts involving angles and areas (pie charts) – we will see some examples shortly that bear this out.

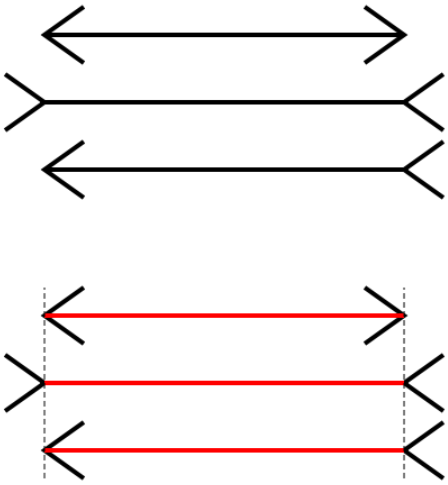
The rankings above should help to choose **methods** as well as to improve **designs** – we've included it here because of the connection to perception.

## 5.3 Misleading charts

Now we are starting to develop an understanding of how the brain perceives images, a word of warning: it is possible to exploit this to mislead, as well as to enlighten:

| **Ensure that a visualisation enlightens rather than misleads.** |
|---|

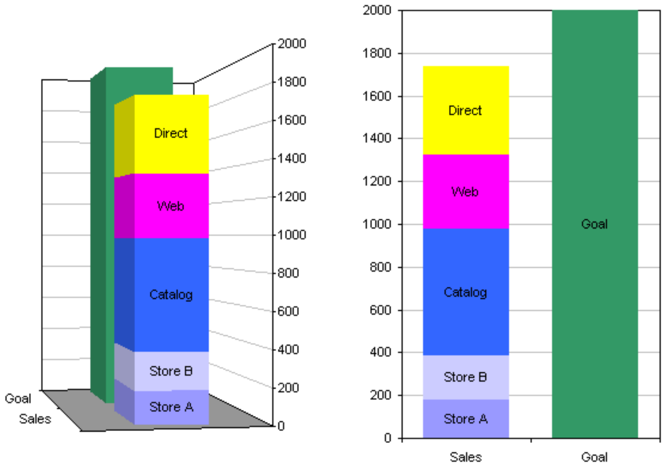
Everyone is familiar with the notion of an optical illusion. Optical illusions arise because the brain actively maps our surroundings and uses rules and tricks to do it – these rules and tricks can all too easily be subverted. One example is the Müller-Lyer illusion, in which the lines between the arrows are all the same length – this is much easier to see in the bottom diagram than the top:



Source: Wikipedia – attribution at [en.wikipedia.org/wiki/File:M%C3%BCller-Lyer_illusion.svg](en.wikipedia.org/wiki/File:M%C3%BCller-Lyer_illusion.svg)


Similarly illusory effects can easily be created by misusing certain kinds of chart, or presenting them badly. We give some examples below.

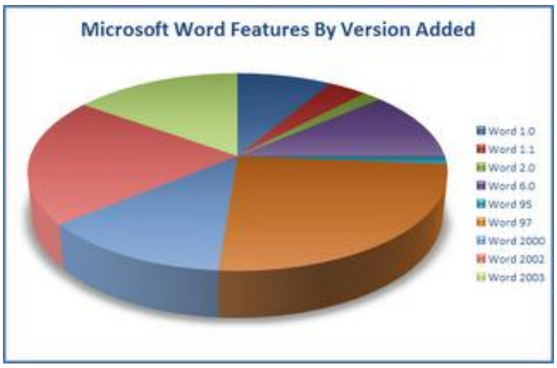
## Spurious 3d charts




Source: PeltierTech: [peltiertech.com/WordPress/the-perils-of-being-in-3d](peltiertech.com/WordPress/the-perils-of-being-in-3d)

Unless the situation you want to visualise is genuinely 3-dimensional, 3-d effects are generally misleading and confusing. A fairly benign example is given above, which shows sales coming from different channels (direct, web, etc). On the left-hand side, we can't easily see the absolute values of sales from each channel, because the 3-d effect obscures the relationship with the vertical axis. However, at least the relative sizes of sales in each channel are clear. On the right-hand side, we can see both the absolute and relative sizes easily, and it's clear that the 3-d effect added nothing of substance, and removed a key part of what the chart sought to convey.

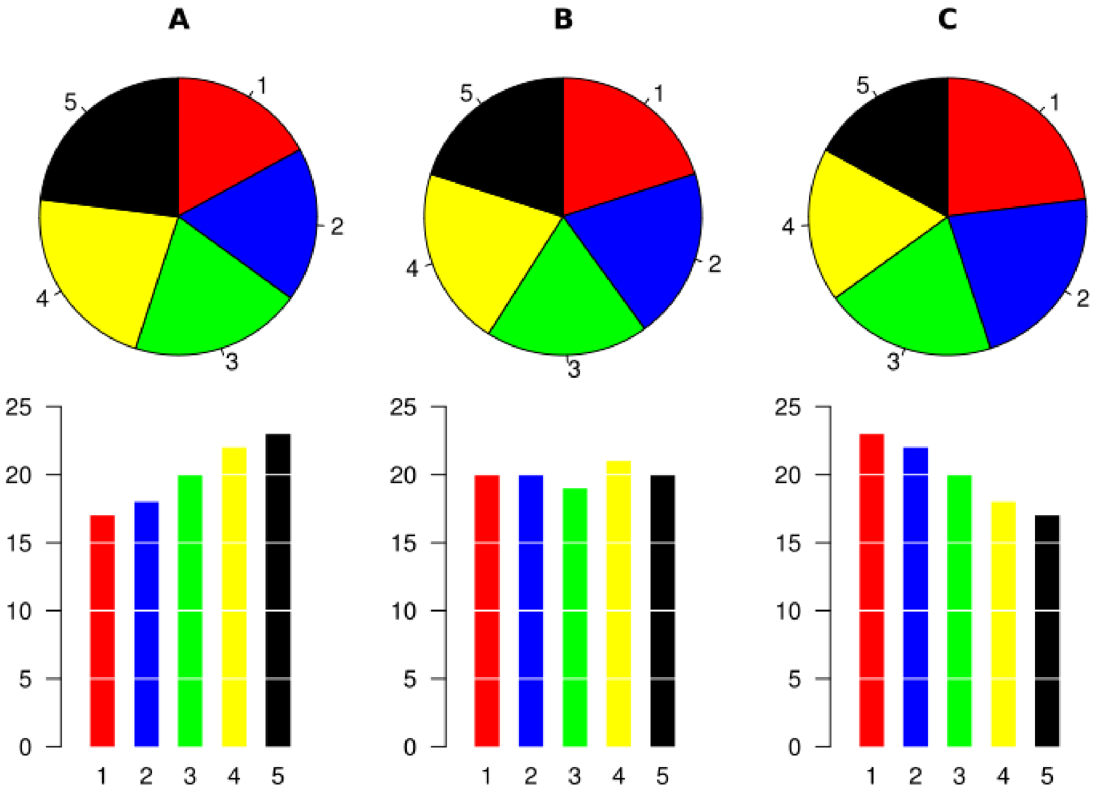
A more pernicious example is the following chart from Microsoft setting out the progression of features in different versions of Word:




This is weak on a number of levels: it gives no indication whatsoever of the absolute number of features added (surely more relevant than the relativity of features across versions). Even worse, the 3-d presentation seriously obscures the relationship between versions, because the perspective used makes slices at the back appear less significant than they really are. Finally, the use of slightly different shades of red, blue and green is a masterstroke of obsfuscation.

**Pie charts**

Even without spurious 3-d effects, pie charts can be misleading, and it is almost always possible to improve on them using another kind of chart. The following example from Wikipedia makes this very clear:




*Source:* [commons.wikimedia.org/wiki/File:Piecharts.svg](commons.wikimedia.org/wiki/File:Piecharts.svg) *where author information can be found.*
*Licensed under the Creative Commons Attribution 1.0 Generic licence.*

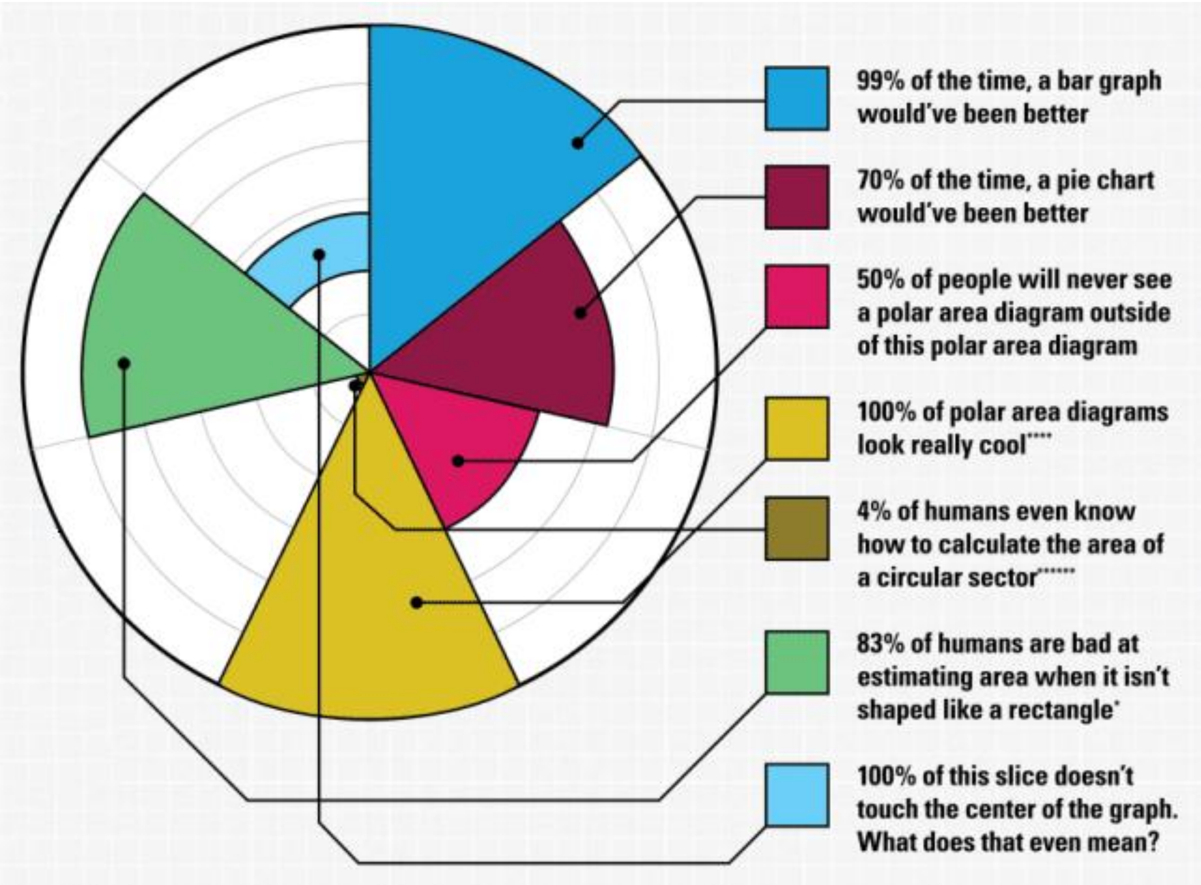
Looking at the pie charts at the top, we perceive that the segments are roughly the same size for all 3 charts (A, B and C), but we can't really tell much more than this, and we have no indication of the value of each segment. By contrast, at the bottom we can see both the absolute value and a clear progression across categories in charts A and C, which was obscured by the rotational presentation of the pie charts.

In short, the bar chart gave us more information more clearly. We could also plot multiple sets of data on the same bar chart if we wanted, to facilitate further comparisons, whereas we can only place pie charts side-by-side.

The main reason the bar chart works better is that as we saw in section 5.2, the brain is better at making judgements based on length than area, and so small changes in length stand out more than small changes in area. If the overall circular shape is important, then a doughnut chart (with a hole in the middle) can be used, as this is more length-like than area-like.

**Radar charts**

Radar charts (also called "polar area" charts) suffer many of the same shortcomings as pie charts, and the same data can usually be shown using a clearer method. The following illustration largely speaks (in a tongue-in-cheek manner!) for itself:



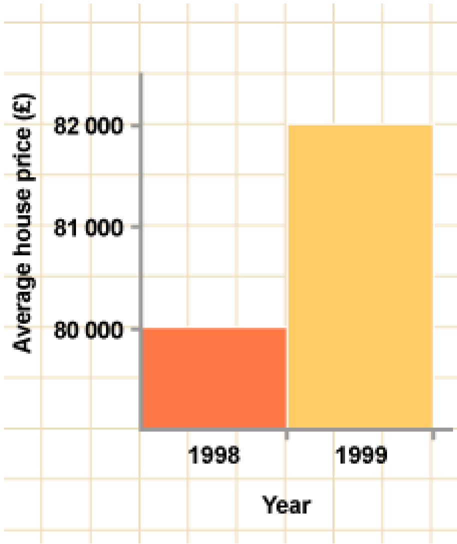

Source: SMBC: [www.smbc-comics.com/index.php?db=comics&id=3167](www.smbc-comics.com/index.php?db=comics&id=3167)

However, this particular example does illustrate one obvious advantage of well-presented charts – namely that they are eye catching, probably more so than many more conventional and technically "better" charts. In certain contexts this advantage might outweigh other shortcomings – a useful principle to bear in mind:

**A technically imperfect chart that stimulates discussion and insight is better than a perfect one which is ignored.**

**Bar charts without a zero point**

Even a bar chart can be used to mislead:




Source: BBC GCSE Bitesize: www.bbc.co.uk/schools/gcsebitesize/maths/statistics/representingdata2rev5.shtml

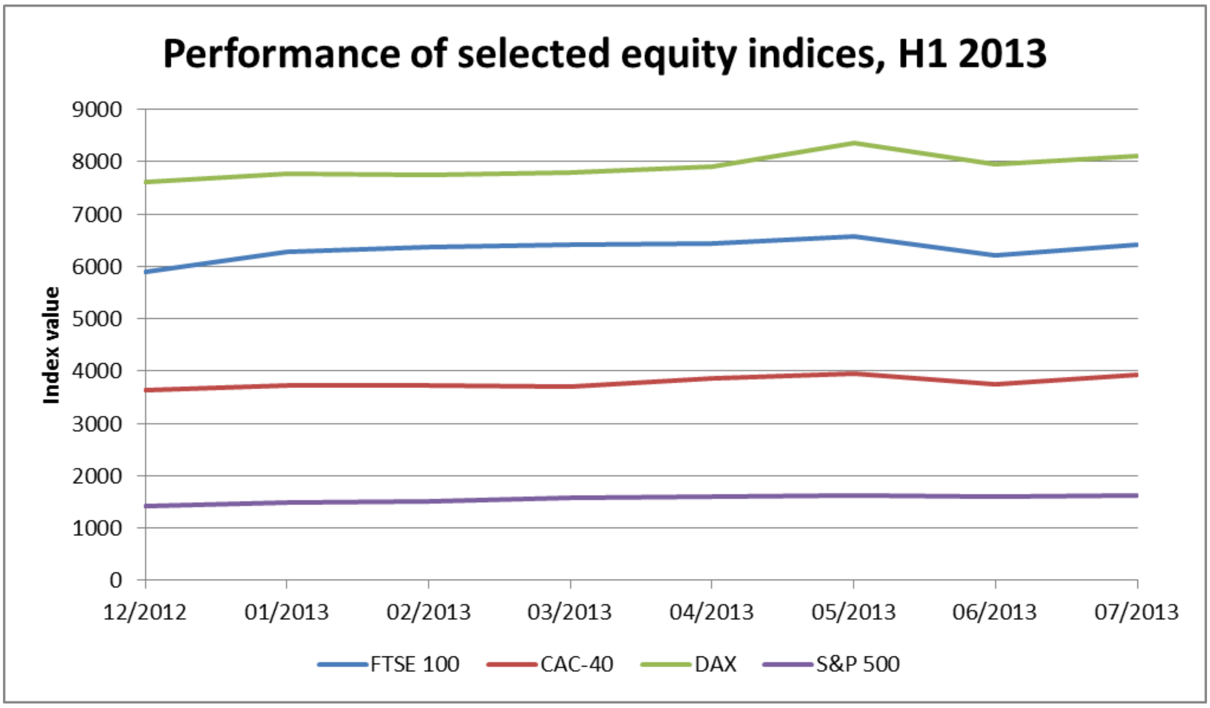
Starting the scale at 79,000 (which is not even labelled) makes 1999 look 3 times better than 1998, even though the values are almost identical (80,000 vs 82,000). The pattern seen by the brain is just not there at all in reality, but it's very hard to shift the idea that 1999 was much better than 1998 once it's been formed.

A similar effect can be created by using a log-scale on the y-axis. This is not to say that log-scales can't be useful, only that they tend to be "unexpected", and their presence should be highlighted upfront to avoid misleading the user.

**Correct but unhelpful charts**

Even if obviously misleading elements are absent, it is still very possible for charts to be unhelpful, missing opportunities to give the reader clarity of the message being conveyed, and (from a practical sense) causing the reader to spend more time viewing a chart than ideally necessary, and hence delaying them.

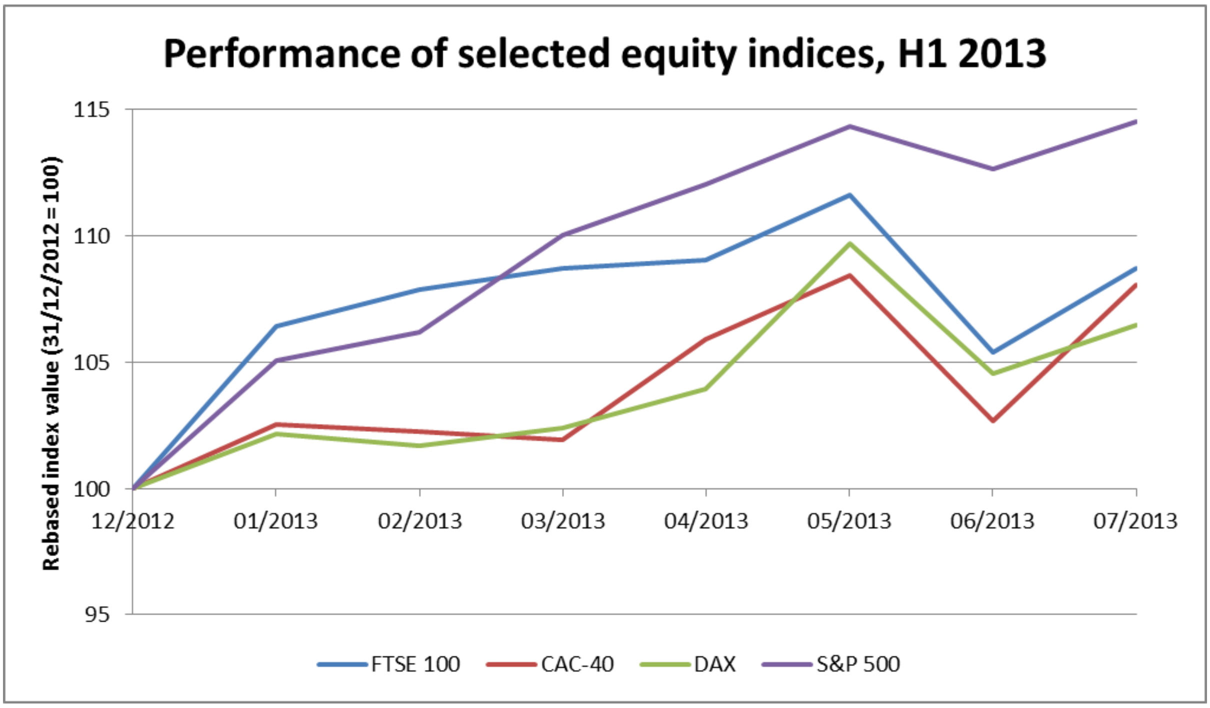
For example, evaluate how useful the following chart is (we have recreated it without attribution from a recent article in a financial newsletter, to save any embarrassment):




From this chart, a number of conclusions could validly be drawn:

- The different indices have widely different values
- None of the indices appeared to grow significantly over the period; the S&P probably least of all
- The DAX and FTSE were more volatile than the CAC and S&P

Now consider the following version, which uses exactly the same data, this time applying the simple device of rebasing all values to their respective starting points:



For the sake of a very simple transformation of the underlying figures, the chart has become significantly more useful. Growth rates and volatilities are far easier to deduce; in particular it is evident that the S&P was actually the best performer over the period, whereas in the first chart it appeared to be the worst. Information about the absolute level of each index has been lost, but since it should be widely known that the absolute values of indices are arbitrary anyway, this is no detriment to the reader[13].

## 5.4 Miscellaneous Advice

We collect together some largely self-explanatory principles that should help to achieve successful designs:

**Keep type and colour under control. Limiting the amount of colours and different fonts in your graphics will help you create a sense of unity. Stick to just two or three colours and play with their shades. Do the same with fonts and sizes.**

**If something is important, highlight it in such a way that readers can sense its relevance and how it operates.**

**Entities of similar nature should look alike and conversely things which are significantly different should look different.**

Adapted from 'The Functional Art: An introduction to information graphics and visualisation' by Alberto Cairo.

[13] We would be interested to hear from any reader who takes a contrary view, ie that the first depiction has any merit whatsoever relative to the second.

# 6 Implementation

Excel is the obvious tool for data visualisation, familiar to and used by most actuarial staff. More recent versions of Excel have more visualisation features:

- Lots of chart types – but avoid spurious 3-d charts
- Sparklines (small charts embedded in cells, used in the correlation matrix case study)
- Advanced forms of conditional formatting such as icon sets or data bars (also used in the correlation matrix case study, to complement sparklines, and also in the upcoming stochastic model case study).
- Pivot Charts – useful to decide quickly on an arrangement of categories and series, subject to later refinement by setting up a similar chart from scratch.

Since Excel is so familiar, it's easy to allow visualisations to be constrained by what it can and can't support. This is a shame, because there are many alternative tools, and in any case, Excel can support more methods than might first be apparent. peltiertech.com/Excel/Charts/ChartIndex.html is an excellent resource for producing additional types of chart in Excel itself, or customising the built-in charts in ways that might not be immediately apparent.

Excel can also lay out all sorts of shapes on screen, and at a pinch this can be automated using VBA, as we did in the correlations case study, although we'd recommend getting comfortable with a visualisation in another way before spending time developing any VBA.

If you have an idea for a visualisation that it seems Excel just can't do, even with VBA, one alternative that might be available within your organisation is the statistical package R. This has very powerful and flexible charting capabilities but has a steep learning curve. In our experience, people start with R using 'incantations' (scripts which work but you don't really understand why) and gradually start to understand what's really going on.

www.statmethods.net/graphs/index.html has some useful tips on constructing charts in R and there are many books and other resources online, some of which given in the references in Appendix A.

There are also bespoke packages (either standalone or add-ins for Excel or R) for some kinds of visualisation. For example RGGobi is an add-in to R which can be used to visualise high-dimensional data sets. And some visualisations can be produced online without needing to install bespoke software (see section 8 for an example).

Our overall advice is:

**Try not to let your visualisation ideas be constrained by technology – there will often be a way to do it with the tools you already have, and if not then there is a huge number of resources to help.**

# 7 Case Study 2: Stochastic Model Output

## 7.1 Introduction

In this case study, the situation at hand is the presentation of results from a stochastic simulation of the Solvency 2 balance sheet, in this case of an annuity company[14]. Stochastic simulation outputs, being high dimensional in nature, present an opportunity to add real value through using appropriate visualisation techniques. By "high dimensional", we refer to the fact that such outputs usually consists of combinations of:

- Different risk factors (eg longevity, credit, interest rates)
- High numbers of simulations/scenarios (anywhere from low numbers of thousands up to millions)
- Different runs based on sensitivities to the calibration of the underlying models
- Different runs with different "as at" dates

Presenting these results in an intuitive and cogent manner can therefore be a challenge. In this case study we suggest a few different methods which may add value; as with Case Study 1 we do not suggest that there exists a single "best" method of presenting the results, and that the value of any particular visualisation will be specific to the organisation and audience concerned.

In a similar way as for Case Study 1, we work through some iterations of potential ways of visualising the data, drawing out some conclusions.

## 7.2 Structured top-down approach

Mapping this situation onto our overall structured approach gives the following:

**Purpose: Know your audience and what they are looking for**

Our audience includes a range of people, from the reviewer of the runs, through the Chief Actuary, potentially to the regulator who may wish to review these results as part of an ICAS process. In all cases the focus is likely to be on achieving high level comfort that the results look reasonable, and that the risk profile accords with their existing understanding.

**Narrative: Determine the messages the audience will find most useful**

We think that the following messages will be useful:

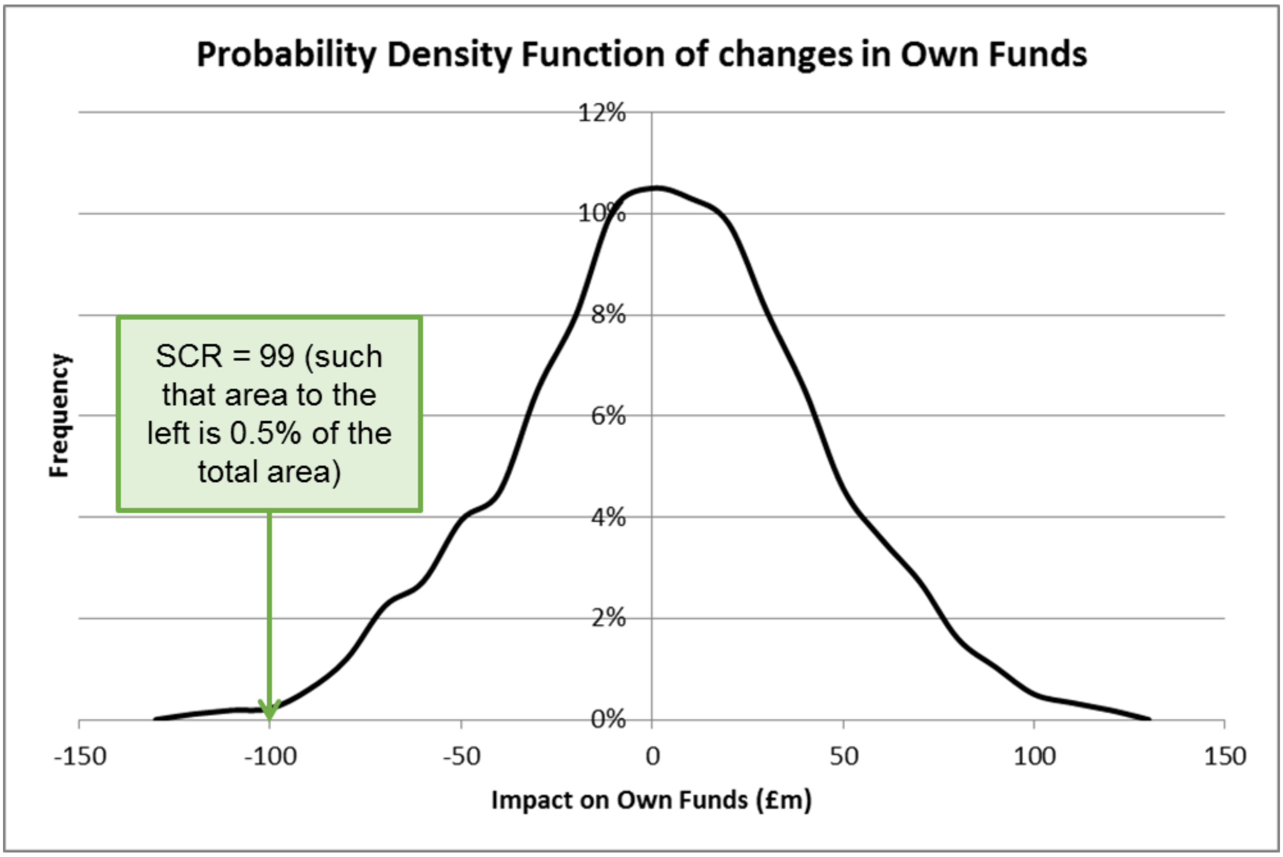
- What is the value of the SCR?

[14] Although we use an annuity company holding corporate bonds as an example, the principles are equally applicable to other types of insurance company undertaking a similar set of simulations, or a pension fund or investment fund.

- What is the shape of the Probability Distribution Forecast ("PDF"[15]), from which the SCR is derived?
- What are the balance sheet losses at other key percentiles of the distribution?
- What are the principal risks (and interactions between those risks) that drive the shape of the lower tail of the distribution, and hence the level of the SCR?

**Method, Design, Implementation: Select the best methods to display your messages, design the display, and choose a technology**

The starting point[16] is to display a fairly traditional view of the PDF, showing the SCR:



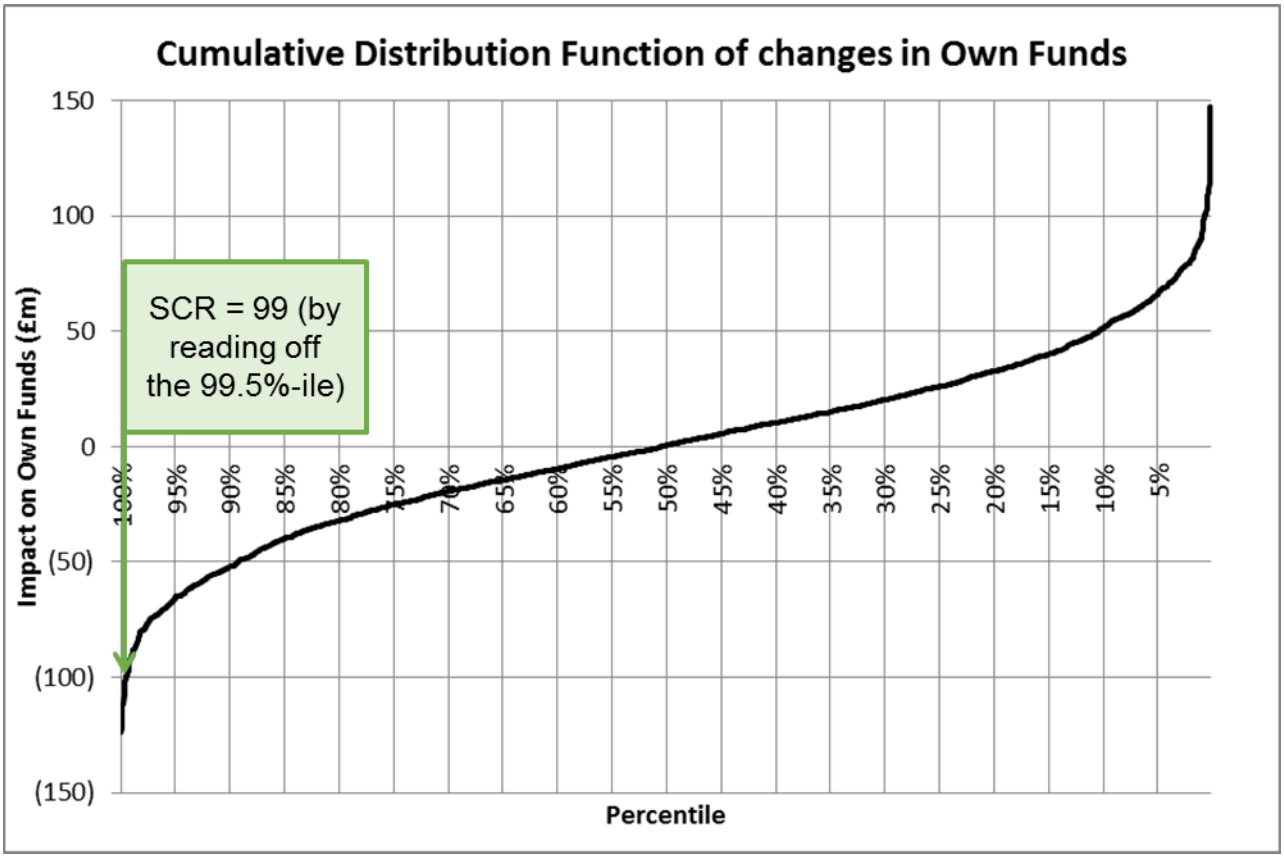

We evaluate the advantages and disadvantages of the chart as follows:

| Advantages | Disadvantages |
|---|---|
| Frequency plots are a familiar presentation | Values on the y-axis are not very meaningful |
| It gives a good indication of the overall symmetry of the PDF, especially if multiple distributions are shown on the same axis | It is not easy to "read off" values at specific percentiles (in practice you need to derive these from the underlying data first then draw them onto the chart). This is an issue if percentile values are of interest to the user |

[15] PDF could also stand for "Probability Density Function", which is a similar concept – for our purposes we can use the definition interchangeably.
[16] For the sake of brevity we have started with a reasonably well presented chart, avoiding some common pitfalls such as poor/non-existent labelling and superfluous decimal places on the axes.

| Advantages | Disadvantages |
|---|---|
| | (as they usually are). |

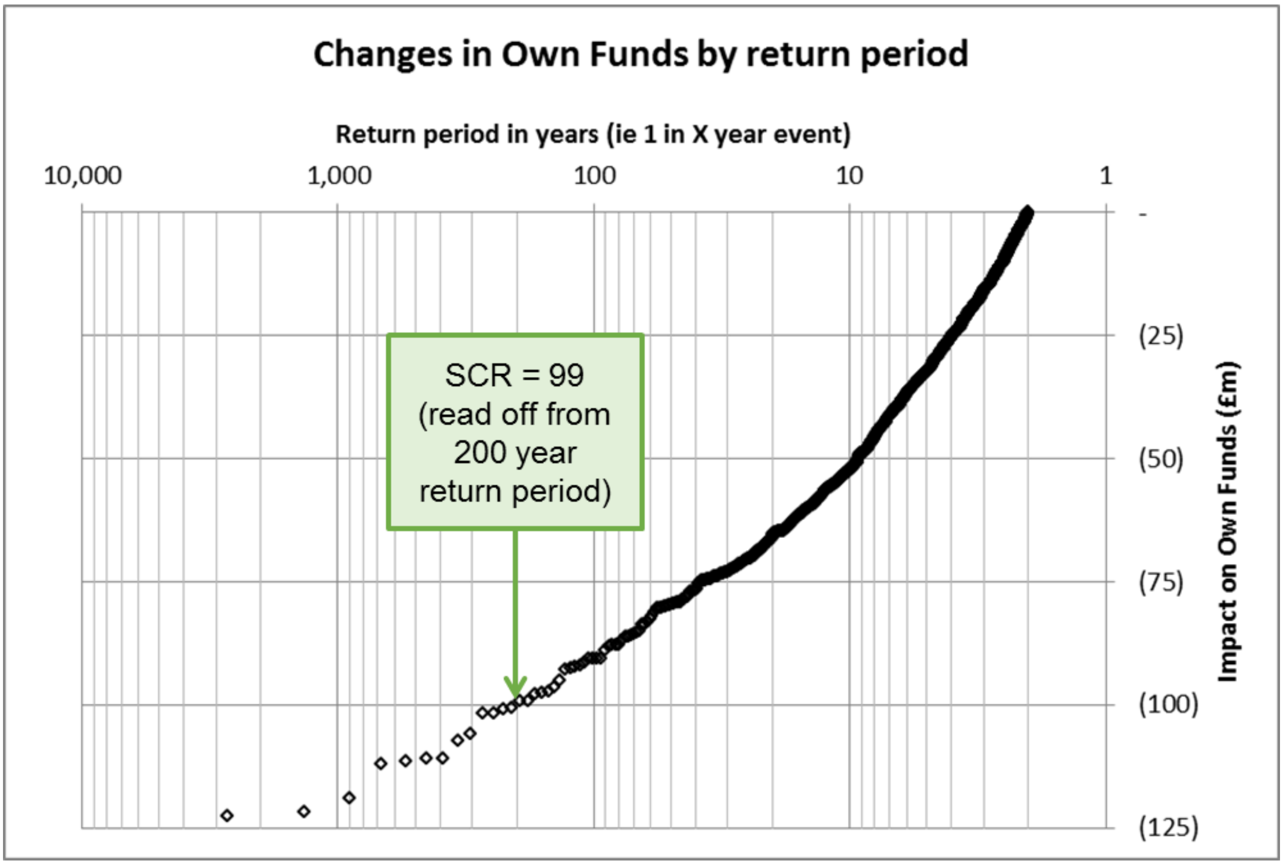
Therefore we try an alternative presentation of the same data below – this time, in the form of a Cumulative Distribution Function ("CDF"):




Evaluating the advantages and disadvantages:

| Advantages | Disadvantages |
|---|---|
| Still a reasonably familiar presentation, albeit not as widely used as a PDF. | Uses more "ink" than the PDF (eg gridlines, but these facilitate the reading of values). |
| Most percentiles can easily be read off, especially the median. | Percentiles in the tail, which are usually of most interest, are still not easy to read off. |

Mindful of the importance in this context of the extreme lower tail, we try a further iteration using a log-scale, which deliberately gives more ink and space to this area of interest:

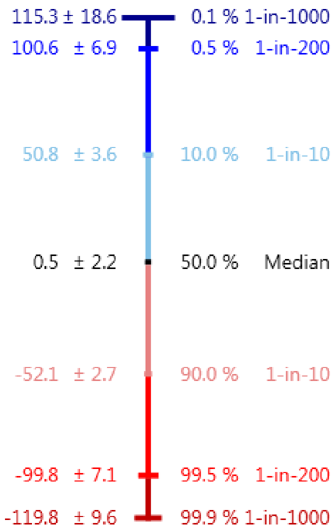



Evaluating the advantages and disadvantages:

| Advantages | Disadvantages |
|---|---|
| For the first time, the SCR (and similarly extreme percentiles) can actually be read off the graph directly. | The log scale is not immediately intuitive. |
| Plotting actual data points[17] rather than a line gives a clear indication of the granularity (or lack of it) in the extreme tail. | The upper tail of the distribution is lost. |

[17] We have also taken the opportunity here to demonstrate a useful tip for scatter plots – namely to use "hollow" markers instead of solid ones. Because hollow markers look like solid markers only when they are tightly clustered, they give the reader an indication of how tightly clustered the data points are – whereas solid markers form a solid "blob" regardless of whether there are many data points clustered or only a few. Another technique sometimes used is to apply "jitter" (small random perturbations) to the markers, as this can improve the visual representation of marker density.

To round off this series of iterations, we present the same data but in an even more abstract form – a "PDF summary plot" – which shares characteristics with box/whisker and candlestick plots:




The numbers to the left of the candlestick represent the change in Own Funds, with an assessment of simulation error[18]. On the right we show the corresponding points on the distribution, expressed as percentiles and return periods. The positions of the markers are scaled to the Own Funds figures; the length of the horizontal bars to the simulation error. It would also be possible to include assessments of parameter error or model error in addition to simulation error, although this needs to be balanced against over-complicating the chart.

Even though, relative to the previous presentations, much of the data (and hence the ink) has been removed, we can still tell a lot from this:

- The distribution is fairly close to symmetrical
- 1-in-200 events have roughly twice the impact of 1-in-10, like the Normal distribution
- 1-in-1000 events do not have significantly more impact than 1-in-200

Summarising the advantages and disadvantages:

| Advantages | Disadvantages |
|---|---|
| High data : ink ratio. | Data is only shown for 7 points on the distribution; it would need to be redrawn to include other points. |
| Addition of another dimension, ie simulation error. | |

[18] Simulation error arises because Monte Carlo simulation converges as the number of samples tends to infinity, and only a finite number of samples can be used in practice. A degree of simulation error is therefore unavoidable, and the question is really what level of simulation error is acceptable in a particular context.

For the final part of this Case Study we use the same underlying data, but present individual scenarios from the extreme tail, in tabular form. This is likely to be a real life use of such data, as it enables actuaries to “drill into” the specific drivers of the capital requirement.

The starting point is a direct copy of the model output, for the most extreme percentile (ie above the 99%-ile):

| Percentile | Scenario | All risks change in OF (£m) | Interest rate risk change in OF (£m) | Credit risk change in OF (£m) | Longevity risk change in OF (£m) | Operational risk change in OF (£m) | Implied non-linearity/non-separability (£m) |
|---|---|---|---|---|---|---|---|
| 100.0% | 89 | -123.74 | -0.10 | -81.26 | -42.23 | -0.24 | 0.10 |
| 100.0% | 2336 | -122.47 | 3.33 | -92.75 | -29.48 | -0.29 | -3.28 |
| 99.9% | 238 | -121.60 | -23.18 | -73.54 | -47.37 | -0.05 | 22.55 |
| 99.9% | 1540 | -119.01 | -3.11 | -78.76 | -40.12 | 0.00 | 2.97 |
| 99.9% | 1105 | -111.91 | -0.10 | -89.79 | -21.94 | -0.17 | 0.09 |
| 99.8% | 2258 | -111.39 | -5.18 | -78.71 | -31.75 | -0.32 | 4.56 |
| 99.8% | 2716 | -111.00 | 2.46 | -63.67 | -47.55 | -0.06 | -2.19 |
| 99.7% | 389 | -110.81 | -2.98 | -82.69 | -27.16 | -0.59 | 2.61 |
| 99.7% | 510 | -107.30 | -3.41 | -90.41 | -16.25 | -0.12 | 2.89 |
| 99.7% | 235 | -105.98 | 0.07 | -86.58 | -18.98 | -0.43 | -0.06 |
| 99.6% | 557 | -101.64 | 3.21 | -60.59 | -41.42 | -0.26 | -2.58 |
| 99.6% | 440 | -101.57 | 5.12 | -86.57 | -15.70 | -0.29 | -4.13 |
| 99.6% | 1365 | -100.99 | 2.38 | -71.90 | -29.39 | -0.18 | -1.90 |
| 99.6% | 2285 | -100.58 | -2.82 | -70.24 | -29.07 | -0.66 | 2.20 |
| 99.5% | 689 | -99.19 | -1.86 | -73.04 | -25.65 | -0.08 | 1.44 |
| 99.4% | 1202 | -99.18 | -5.21 | -100.09 | 2.19 | -0.08 | 4.00 |
| 99.4% | 2761 | -97.72 | 2.18 | -59.24 | -38.97 | -0.01 | -1.68 |
| 99.4% | 2274 | -97.50 | 4.76 | -50.22 | -48.22 | -0.13 | -3.68 |
| 99.3% | 640 | -97.27 | 2.28 | -52.02 | 23.52 | -70.70 | -0.35 |
| 99.3% | 541 | -96.54 | 4.53 | -75.73 | -21.77 | -0.10 | -3.47 |
| 99.3% | 2879 | -94.96 | -0.25 | -95.03 | 0.81 | -0.68 | 0.19 |
| 99.2% | 373 | -92.90 | 1.84 | -82.36 | -10.86 | -0.18 | -1.34 |
| 99.2% | 1948 | -92.60 | 3.91 | -79.71 | -13.85 | -0.09 | -2.86 |
| 99.2% | 1038 | -92.23 | 2.49 | -74.69 | -18.08 | -0.15 | -1.80 |
| 99.1% | 2752 | -91.98 | -1.16 | -70.20 | -21.29 | -0.16 | 0.83 |
| 99.1% | 1083 | -91.33 | -2.88 | -51.47 | -39.00 | 0.00 | 2.03 |
| 99.0% | 2355 | -90.69 | -2.16 | -61.66 | -27.93 | -0.45 | 1.51 |
| 99.0% | 1100 | -90.63 | 3.57 | -36.60 | -54.82 | -0.24 | -2.54 |
| 99.0% | 1750 | -90.50 | -2.94 | -66.98 | -22.55 | -0.08 | 2.05 |

Summarising the advantages and disadvantages:

| **Advantages** | **Disadvantages** |
|---|---|
| All relevant data is presented and easy to look up. | Wasted ink, eg gridlines, repetitious headings, unnecessarily precise values. |
| | Difficult to spot patterns or relationships. |

Using some simple methods to address the shortcomings yields the following iteration:

| Euler allocation by risk | | (99) | 1 | (75) | (21) | (4) | (0) |
|---|---|---|---|---|---|---|---|
| | | Changes in Own Funds arising from variations in... (£m) | | | | | |
| Percentile | Scenario | All risks | Interest rate risk | Credit risk | Longevity risk | Operational risk | Implied non-linearity & non-separability |
| 100.0% | 89 | (124) | (0) | (81) | (42) | (0) | 0 |
| 100.0% | 2336 | (122) | 3 | (93) | (29) | (0) | (3) |
| 99.9% | 238 | (122) | (23) | (74) | (47) | (0) | 23 |
| 99.9% | 1540 | (119) | (3) | (79) | (40) | (0) | 3 |
| 99.9% | 1105 | (112) | (0) | (90) | (22) | (0) | 0 |
| 99.8% | 2258 | (111) | (5) | (79) | (32) | (0) | 5 |
| 99.8% | 2716 | (111) | 2 | (64) | (48) | (0) | (2) |
| 99.7% | 389 | (111) | (3) | (83) | (27) | (1) | 3 |
| 99.7% | 510 | (107) | (3) | (90) | (16) | (0) | 3 |
| 99.7% | 235 | (106) | 0 | (87) | (19) | (0) | (0) |
| 99.6% | 557 | (102) | 3 | (61) | (41) | (0) | (3) |
| 99.6% | 440 | (102) | 5 | (87) | (16) | (0) | (4) |
| 99.6% | 1365 | (101) | 2 | (72) | (29) | (0) | (2) |
| 99.6% | 2285 | (101) | (3) | (70) | (29) | (1) | 2 |
| 99.5% | 689 | (99) | (2) | (73) | (26) | (0) | 1 |
| 99.4% | 1202 | (99) | (5) | (100) | 2 | (0) | 4 |
| 99.4% | 2761 | (98) | 2 | (59) | (39) | (0) | (2) |
| 99.4% | 2274 | (97) | 5 | (50) | (48) | (0) | (4) |
| 99.3% | 640 | (97) | 2 | (52) | 24 | (71) | (0) |
| 99.3% | 541 | (97) | 5 | (76) | (22) | (0) | (3) |
| 99.3% | 2879 | (95) | (0) | (95) | 1 | (1) | 0 |
| 99.2% | 373 | (93) | 2 | (82) | (11) | (0) | (1) |
| 99.2% | 1948 | (93) | 4 | (80) | (14) | (0) | (3) |
| 99.2% | 1038 | (92) | 2 | (75) | (18) | (0) | (2) |
| 99.1% | 2752 | (92) | (1) | (70) | (21) | (0) | 1 |
| 99.1% | 1083 | (91) | (3) | (51) | (39) | (0) | 2 |
| 99.0% | 2355 | (91) | (2) | (62) | (28) | (0) | 2 |
| 99.0% | 1100 | (91) | 4 | (37) | (55) | (0) | (3) |
| 99.0% | 1750 | (91) | (3) | (67) | (23) | (0) | 2 |

In particular, the use of the conditional formatting bars (within Excel) immediately gives a much more immediate and clearer impression of which risks dominate:

- Interest rate risk is very low – suggesting that interest rate hedging has succeeded
- Credit risk is the dominant risk, contributing the majority of the overall change in Own Funds in each scenario
- Longevity risk is the secondary contributor to the overall change in Own Funds
- The impact of operational risk, and non-linearity/non-separability phenomena, is low

Summarising the advantages and disadvantages:

| **Advantages** | **Disadvantages** |
|---|---|
| Use of conditional formatting makes patterns much more visible. | |
| Reduction in ink from removal of unnecessary gridlines and decimal places improves clarity (compensating for the extra ink from the conditional formatting). | |

Numerous variations or iterations of this presentation of the data are possible – as ever, there is no “best” solution, as it will depend on the precise purpose and audience.

## 7.3 Conclusions

Much of this case study arose from a phone call in which the authors challenged each other to improve the initial iteration, and so this case study is a real example of applying the trial and error principle. Some of the improvements arose from changing the ‘units of measurement’ (here from density to distribution), or the scale (linear to log), all guided by trying to bring out the important features of the data.

One way in which we assessed whether the visualisation was working or not was to see how clearly it told us things we already knew about the situation being visualised. For example, we expected interest rate risk to be low, since the example portfolio had been constructed with a close match between asset and liability cash-flows – including the conditional formatting bars in the table of numbers brought this out clearly.

# 8 Qualitative Visualisations

Although we have deliberately focussed in this paper on visualisations for quantitative data, there are several types of qualitative visualisation (not involving numbers) which are useful to actuaries, and we give some examples below:

- Diagrams of model structure, showing inputs and outputs to a formula, or the overall flow of data within a model. (Several popular actuarial modelling packages include basic versions of this kind of diagram.)
- Data flow diagrams showing the movement of data between systems and models.
- Diagrams of the hierarchy of legal entities and funds within a Group, possibly with some colouring convention to indicate the types of business written or the major risks each entity or fund is exposed to.
- Cognitive maps, as advocated in the 2011/2012 sessional paper 'A review of the use of complex systems applied to risk appetite and emerging risks in ERM practice' (Cantle et al., www.actuaries.org.uk/sites/all/files/documents/pdf/review-use-complex-systemsprintversion.pdf)
- Gantt charts for project management.

Many of the principles articulated earlier in this paper can be applied to the production of qualitative visualisations, with adaptations as necessary.

One possibly less familiar type of qualitative visualisation is the "word cloud", which shows the most frequent words used in a text, with larger font sizes being used for the highest frequencies. Two examples are overleaf, produced using www.wordle.net, from which we leave readers to draw their own conclusions.

**June 2013 IFRS Exposure draft on insurance contracts**


**The Pensions Regulator Code of Practice 03 on Funding Defined Benefits**

## Appendix A: References and Resources

Links were valid at the time of writing, but may not remain so indefinitely. Our commentary is included as a guide – readers may very well have different views.

### Books

| | |
|---|---|
| **The Visual Display of Quantitative Information**<br>(Edward Tufte, 2001) | *A beautifully-written book from one of the gurus in the field. Everyone interested in data visualisation should read this book, even if parts of it are a little dogmatic.* |
| **Show Me the Numbers: Designing Tables and Graphs to Enlighten**<br>(Stephen Few, 2012) | *Very clear and thorough exposition – devotes much attention to tables as well as charts.*<br>*Stephen Few blogs at www.perceptualedge.com* |
| **Data Visualization: a successful design process**<br>(Andy Kirk, 2012) | *Covers the end-to-end process – full of practical advice, but weaker on the theory than some of the other books.*<br>www.visualisingdata.com *is the accompanying website.*<br>www.visualisingdata.com/index.php/resources<br>*is a very useful collection of onward links to further resources.* |
| **The Functional Art: An introduction to information graphics and visualisation**<br>(Alberto Cairo, 2013) | *Covers infographics (more qualitative) as well as data visualisation (more quantitative) – many of the same considerations apply to both areas.* |
| **Information is Beautiful**<br>(David McCandless, 2009) | *Doesn't cover the theory of data visualisation, but useful for idea generation. Accompanying website* www.informationisbeautiful.net<br>*is also useful.* |
| **Graphic Methods for Presenting Facts**<br>(Willard C. Brinton 1914) | *A very early textbook (from 1914!) on what we would now call data visualisation. Facsimile available at archive.org/details/graphicmethodsfo00brinrich* |

## Websites

| | |
|---|---|
| **Gapminder**<br><br>www.gapminder.org | *Probably the single most influential resource for the data visualisation community, the site of the influential Hans Rosling, source of the health vs wealth graphic shown in section 2.2, plus much, much more.* |
| **Economist Graphic Detail blog**<br><br>www.economist.com/blogs/graphicdetail | *Regularly updated charts and graphics covering current affairs, from arguably the world's foremost newspaper, with a strong emphasis on quality and housestyle.*<br><br>www.economist.com/blogs/graphicdetail/2013/02/property-prices *is an example showing the probability of selling a house at a loss based on moving-in / moving-out years.*<br><br>www.economist.com/news/books-and-arts/21580446-revolution-taking-place-how-visualise-information-winds-change *is a review of some fairly recent books on data visualisation and infographics.* |
| **Visual.ly**<br><br>visual.ly | *Portal for a wide range of data visualisations and infographics.* |
| **Spatial.ly**<br><br>spatial.ly | *Focuses on spatial analysis and visualisation.*<br><br>spatial.ly/2013/07/interactive-crime-map-london *is an interactive example showing crime across London. Links through to the FT website (subscription may be required).* |
| **Flowingdata**<br><br>flowingdata.com | *Regularly updated website/blog by Dr Nathan Yau, showing a variety of interesting visualisations (for example this one showing a map of vernacular across the USA).* |
| **Eager Eyes**<br><br>eagereyes.org | *Blog on visualisation and visual communication from Robert Kosara.*<br><br>eagereyes.org/techniques/parallel-coordinates *is a decent explanation of how to interpret parallel coordinate charts.* |
| **Visual Literacy Periodic Table**<br><br>www.visual-literacy.org/periodic_table/periodic_table.html | *A "periodic table" for data visualisation, with a very comprehensive list of examples laid out in categories, along the lines of the periodic table of chemical elements. A useful source of inspiration.* |

| | |
|---|---|
| **Peltier Technical Services**<br><br>peltiertech.com | *Useful tips for working with Excel charts and discussions of data visualisation techniques. For example* peltiertech.com/WordPress/tornado-charts-and-dot-plots *discusses how to present "population pyramids".* |
| **Codex 99**<br><br>www.codex99.com | *Collection of fairly long and considered posts about the history of graphic design and the visual arts.* |
| **Finviz**<br><br>www.finviz.com | *A website using visualisations to show (live) activity in financial markets – treemaps in particular.* |
| **Mapping London**<br><br>mappinglondon.co.uk | *Varied collection of maps about London.*<br><br>mappinglondon.co.uk/2013/fire-insurance-maps-of-london *is an historical example showing fire insurance coverage.* |
| **Miscellaneous** | |
| www.msu.edu/~howardp/softdrinks.html | *Illustrates brand proliferation in the soft drinks industry.* |
| www.energyliteracy.com/wp-content/uploads/2009/10/Spaghetti-20091.gif | *Illustrates the flow of energy from source to usage. The technique used is called a Sankey diagram.* |
| olihawkins.com/2013/09/2 | *Uses animation to show the impact of confidence intervals on central estimates.* |

## Tools

| | |
|---|---|
| **Google Motion Chart**<br><br>code.google.com/apis/visualization/documentation/gallery/motionchart.html | *Online tool allowing you to create motion charts, particularly useful for displaying time dependent data.* |
| **Datawrapper**<br><br>datawrapper.de | *Tool for creating embedded graphics for webpages.* |
| **Prezi**<br><br>prezi.com | *Online and offline tool for creating "non-linear" presentations, useful for moving between big-picture and detail, as distinct from a linear sequence of PowerPoint slides.* |
| **R**<br><br>www.r-project.org | *Free software environment for statistical computing and graphics. Extremely powerful but steep learning curve. www.statmethods.net/graphs/index.html has some useful tips on constructing charts in R. www.rstudio.com is a much nicer development/execution environment that the standard R console.* |
| **Infogr.am**<br><br>infogr.am | *App for creating infographics for embedding into websites.* |
| **Wordle**<br><br>www.wordle.net | *Website for producing word clouds.* |
| **Slate for Excel**<br><br>www.slateforexcel.com | *Tool for visualising the structure of and data flows within Excel spreadsheets.* |

# Appendix B: Summary of Key Principles

## High-level Principles

**Data visualisation techniques are not magical, they can be studied and learned.**

**Data visualisation is a form of communication, and so the techniques of good communication are highly applicable.**

**A top-down structured process to data visualisation, starting with the needs of its users, is likely to be more successful than diving straight in.**

**Trial and error will almost always lead to improvements.**

## Principles for Methods

**Choosing a data visualisation method requires an understanding of the purpose of the visualisation, as different methods are suited to different purposes.**

### Use tables when

1. Individual values will be looked up or compared
2. Only individual values will be compared, rather than whole series of values
3. Precise values are required
4. Values involve more than one unit of measure
5. Values must be presented at various levels of aggregation (i.e. summary and detail)

### Use charts when

1. The message is contained in patterns, trends, and exceptions
2. Entire series of values must be seen in a whole and/or compared
3. To explore relationships, connections and associations

*Source: Adapted from 'Show me the numbers' (Stephen Few 2012)*


**It will often make sense to use both tables and charts to display the same information**

**A visualisation is in effect a model of the data being visualised, and standard advice about models applies. In particular, be selective where necessary ("the map is not the territory"), and be aware of limitations of the model / visualisation.**

**Ensure that a visualisation is suited to the context within which it will be presented.**

## Principles for Design

**Maximise the data : ink ratio, within reason.**

**Remove everything that isn't necessary. Identify what is data and what isn't. Make the data prominent and clear, and make the non-data as unobtrusive as is sensible.**

**Ensure that a visualisation enlightens rather than misleads.**

**A technically imperfect chart that stimulates discussion and insight is better than a perfect one which is ignored.**

**Keep type and colour under control. Limiting the amount of colours and different fonts in your graphics will help you create a sense of unity. Stick to just two or three colours and play with their shades. Do the same with fonts and sizes.**

**If something is important, highlight it in such a way that readers can sense its relevance and how it operates.**

**Entities of similar nature should look alike and conversely things which are significantly different should look different.**

**It is unlikely that there will be a single 'best' design, although there should be broad agreement on whether a given design works or not.**

## Principles for Implementation

**Try not to let your visualisation ideas be constrained by technology – there will often be a way to do it with the tools you already have, and if not then there is a huge number of resources to help.**