

2. WORKING WITH DATA

EXPOSITION

1. THE DATA VALUE CHAIN

Key to understanding how to work with data is the data flow or the **data value chain**. It's generally seen as sequential. Some people prefer to use a lifecycle visualization, this has some benefits. In either case, you should adopt a holistic view: even as you move step by step along the flow, you always need a clear idea of where you want to end before you start.

2. DATA COLLECTION: SOURCES OF DATA

Data can come from internal or external sources.

What are the benefits of each source of data?

- Internal data sources, where data is created by the company, are typically easier to collect and can be more relevant for the company's own purposes and insights.¹ They're free of cost to the company. can launch an array of big data initiatives without ever looking beyond their own walls
- External data is that acquired from outside the company.
 - Third party data – can be specific to your market or customers, but difficult to attain and/or expensive.
 - Open data – ie government statistics, data related to social impact. Can be of great value but often requires more investment in analysis to provide insights of relevance.
 - Collectively-generated data. Many more business are participating in meetups and hackathons.

Case study – is external data worth the cost?

- 7 small companies in Northern Ireland selling processed food products gained access to a large supermarket chain's data on areas such as consumer life stage and lifestyle, market-basket analysis, and best-performing stores for the small firms' products.
- Owners and managers attended workshops to learn how to retrieve and analyse the most relevant data from the database, in order to answer questions such as "How is my category performing?" "What is the most popular flavor of bread?" "What type of consumer buys a product similar to mine?"
- A yogurt maker learned older adults were a key market, so when they visited supermarkets for in-store tastings, they no longer tried to entice younger shoppers and instead focused on older people. The tactic improved the events' productivity.
- Other companies were able to envision long-range innovations, rather than simply looking at other products on the market and trying to imitate them or following the guidelines laid down by the big retail buyers.
- The data amplified the firms' inherent entrepreneurial nature. Workplaces became more collegial: Most of the owner-managers shared the card information with their firms and encouraged employees to get involved and offer new ideas.²

¹ <http://www.aanalytics.com/understanding-analytics-part-1-top-internal-sources-of-big-data/>

² <https://www.conference-board.org/blog/post.cfm?post=2526>

Small business owners face obstacles to investing in external data. Barriers include cost, their lack of expertise, or belief in their own market knowledge.

Once they were given access to the data, most of the businesses were quick to formalize their approach to marketing planning.

In this example the small businesses were able to access the external data for free.

How much could they have paid for it to still be worthwhile?

DATA COLLECTION: TYPES OF DATA

Data is presented as a DATA SET. A **data set** is a grouping of information that's related to each other.

- Data sets can be Quantitative or Qualitative. The difference seems stark but there can be lots of ways to have them interact. Qualitative data may be difficult to precisely measure and analyze. The data may be in the form of descriptive words that can be examined for patterns or meaning, sometimes through the use of coding. Coding allows the researcher to categorize qualitative data to identify themes that correspond with the research questions and to perform quantitative analysis.
- Data can be Discrete or Continuous.
- Data can be:
 - **Nominal:** Nominal data is qualitative; there is no inherent scale or value attached to the data.
 - **Ordinal:** Data that has a natural order, like a Likert scale where 1 means "hate it!" and 5 means "love it!" Just remember that the difference between ordinals (e.g., between 1 and 2 or between 3 and 4) is not necessarily equal. Often the distance between extreme like and dislike is greater than between feeling neutral or slightly positive or negative.
 - **Interval:** Data that has a natural order and the distance between each value is equal, like temperature.
 - **Ratio:** Data that has a natural order, equal distance between values and a natural zero point.³

3. DATA STORAGE

It's vital to think ahead when planning your data collection. Think about how you will use the data you collect and what decisions you want to make based on it. Do you want to get information about a real-time situation? Do you want to understand causality? Do you want to compare two variables and select the best one?

Onsite storage v remote storage

- Traditionally data was stored onsite, but increasing volume of data and decreasing cost of data storage service providers make remote storage in the cloud more attractive.
- The cloud-based storage market is dominated by Amazon Web Services, Google and Microsoft Azure.
- All storage clouds are not created equal. Some clouds are optimized to handle archiving, others have the performance and stability to act as the back end for a primary data storage system, and still others aren't worth the risk for any purpose. The lowest price cloud may end up being much more expensive in the long run if data is lost or inaccessible

³ <https://communitymedicine4asses.com/2013/01/13/scales-of-measurement-nominal-ordinal-interval-ratio/>

- Many solutions provide the flexibility to choose where data is stored: on premise and/or in the cloud.

4

How to choose how to store your data?

When it comes to storing data, there is no 'one-size-fits-all' solution.

Businesses must understand the amount and type of data they have along with the motivation behind storing the information.

WHO needs access to which type of information? How often and how fast do we need to access the data?

- When setting up processes, identify the organization's most important data and prioritize storage management resources appropriately. email may be a company's top priority, but storing and archiving email data for one particular group, say the executives, may be more critical than other groups
- "Save money by only using your fastest storage, like SSD, for data that you actively use, and utilize less expensive platforms, like the cloud, to store your archival or backup data,"

WHAT type of data do we need to store?

- Might we need to combine different data sets in the future?

HOW MUCH data do we have and by how much will it grow in the future?

- Don't store redundant data. Many companies capture and store redundant data. This used to be a less of a problem when disk was expensive, but with capacity being more reasonably priced, the tendency is to store everything.

HOW LONG do we need to store the data?

- A data retention policy is a necessity for both internal data governance and legal compliance. Some of your data must be retained for many years, while other data may only be needed for days."

HOW SECURE does the data need to be? What regulatory requirements need to be adhered to?

- Have a disaster recovery plan. Whichever method or methods of backup you use, be sure to test them and be sure you can recover your data.

3 DATA ANALYSIS

To answer your strategic guiding questions and gain a deeper understanding of your business you will need to take your raw data and transform it using a combination of basic arithmetic and descriptive statistics to reveal patterns—both those that you expect to see and those that may be surprising.⁵

What to look for: facts, stats and trends are all useful but for different reasons.

Note: when talking about data visualization it can be seen as PART OF the analytic process and separately, as a means to present results. They're intrinsically linked, and one can usually support the other.

Examples sales and profitability chart. This is often a typical outcome of data analysis. (Source: Forbes)

⁴ <https://techcloudlink.com/14-things-you-need-to-know-about-data-storage-management/>

⁵ <https://www.schusterman.org/playbooks/data/making-meaning/find-patterns-stats/>

The below Sales & Profitability dashboard by Ann Jackson is a visually compelling, cleanly designed summary of the data that shows the changes over time, the geographical differences, losses for certain product categories and summarizes key performance indicators as numbers. Would Ann hand this dashboard over to her stakeholders and call it done? No, because now the real discussions start. Ann can sit with her audience and drill into further details to explore why certain results have come about and identify opportunities for improving business performance. Further investigation of the data, exploring it with your audience, that is the value you add as an analyst, beyond producing dashboards. They give you an excellent basis for these discussions but shouldn't be the end point.⁶

4. DATA INTERPRETATION AND USE

Having data is not enough. Analyzing data is not enough. It is vital to translate the insights gained from data into actions that support business growth.

As with any good strategy, knowing when and how to apply it, when to change it, and how to measure its success are all critical for organizations to move from being ones that does data, to being truly data-driven.

The cycle of using data is ongoing – it doesn't just stop when the insights are presented but requires indepth discussion. Analysing the data is a starting point, not an end point.

In which chart is there a bigger difference?

Data use arises from science, but it is also an art. Data is objective but it can be presented subjectively. Given that data visualization is so important for making sense of large data and for driving decisions, it is important to understand how to go about it.

Slide 1: In the first chart it looks like there is a big difference.

Slide 2: In this chart, we see both are a difference between 35.5% and 41.5% - the same difference. This shows that

- charts can be deceptive
- it is important to communicate data in formats that are accurate and to reasonable scale.
- When information is added on scale, it becomes much clearer.

Which way of displaying the data is most effective?

You have to think about how you communicate data for your purposes AND how to communicate it clearly so your message is understood.

In this example, the coloured chart is the easiest to understand. What changed?

The original blue chart was “decluttered” – the axis and labels were removed and the blank space reduced. Then salient details were “emphasized” – the colours were added to reflect a continuum of responses and the numbers were added to show relative values.

Conclusion:

Any time you use data to convey something, you are essentially telling a story. It might be in a quarterly presentation for your board, in a formal program evaluation report, in an annual report to constituents.

⁶ <https://www.forbes.com/sites/evamurray/2019/03/29/what-is-the-difference-between-data-analysis-and-data-visualization/>

Storytelling with data is powerful, and it comes with an ethical imperative. Data can be easily manipulated to tell a story that is not there, or to minimize a story that is.⁷ It is important to start with the data and let it tell you the story, instead of crafting your story ahead of time and only selecting the data that supports it. For example, if you present a summary of a program evaluation that only highlights the successes and aspects of your service that people liked, and does not include the components that came up short and revealed areas for improvement, you are not telling the full story.

This project has been funded with support from the European Commission. This publication [communication] reflects the views only of the author, and the Commission cannot be held responsible for any use, which may be made of the information contained therein.

⁷ <http://www.bdbanalytics.ir/media/1123/storytelling-with-data-cole-nussbaumer-knaflic.pdf>