SUSS
SINGAPORE UNIVERSITY OF SOCIAL SCIENCES

ANL201 Data Visualisation for Business

STUDY GUIDE

School of Business

Course Development Team

Head of Programme	:	Assoc Prof James Tan
Course Developer(s)	:	Eric Tay
Revised By	:	Dr. Priyanka Gupta
Technical Writer	:	Diane Quek, ETP
Video Production	:	Danny Chin, ETP

© 2023 Singapore University of Social Sciences. All rights reserved.

No part of this material may be reproduced in any form or by any means without permission in writing from the Educational Technology & Production, Singapore University of Social Sciences.

ISBN 978-981-4685-32-0

Educational Technology & Production
Singapore University of Social Sciences
463 Clementi Road
Singapore 599494

How to cite this Study Guide (APA):

Tay, E. (2023). *ANL201 Data visualisation for business*. Singapore: Singapore University of Social Sciences.

Release V1.8.0

Build S1.0.15, T1.7.0

Table of Contents

Course Guide

Study Unit 1: Business Performance Measurement

Study Unit 2: Science and Art of Data Visualisation

Study Unit 3: Four Stages of Data Visualisation

Study Unit 4: Data Visualisation for Discrete and Continuous data

Study Unit 5: Business Performance Dashboard

Study Unit 6: Advanced Data Visualisation Techniques

List of Tables

List of Figures

List of Lesson Recordings

Course Guide

Data Visualisation for Business

1. Welcome

Presenter: Dr. Priyanka Gupta

This streaming video requires Internet connection. Access it via Wi-Fi to avoid incurring data charges on your personal mobile plan.

Click here to watch the video. [i]

Welcome to the course ***ANL201 Data Visualisation for Business***, a 5 credit unit (CU) course.

This Study Guide will be your personal learning resource to take you through the course learning journey. The guide is divided into two main sections – the Course Guide and Study Units.

The Course Guide describes the structure for the entire course and provides you with an overview of the Study Units. It serves as a roadmap of the different learning components within the course. This Course Guide contains important information regarding the course learning outcomes, learning materials and resources, assessment breakdown and additional course information.

[i] https://suss.ap.panopto.com/Panopto/Pages/Viewer.aspx?id=a98eb45f-4c69-45a4-a173-af4600652f31

2. Course Description and Aims

ANL201 Data Visualisation for Business is designed to equip students with the skills and knowledge to use Data Visualisation in measuring and monitoring business performance. Students will acquire knowledge of performance metrics and measurement framework, components, stages and types of data visualisation and how to create an effective business performance dashboard. This course introduces business performance measurement in a strategy map framework and data visualisation in a problem-solving framework. It goes through the visualisation stages in a systematic process, and uses examples to demonstrate data preparation and visualisation. The course also examines the applications of Tableau for data visualisation. Students will work along the examples and exercises in a "consulting" mode to reproduce various frameworks and visualisations and draw insights and conclusions from them. At the end of this course, students will be competent in using Data Visualization software to create Business Performance Dashboard.

Course Structure

This course is a 5-credit unit course presented over 6 weeks.

There are six Study Units in this course. The following provides an overview of each Study Unit.

Study Unit 1 – Business Performance Measurement

This unit introduces the meaning, purpose and foundation of performance management. Through this process, an organisation establishes criteria to determine quality of their activities, efficiency, and effectiveness in conducting their business operations. Business performance metrics is used to measure different aspects of business and categorised under different business functions using the balanced scorecard framework. Strategy map serves as the system that provides linkages between vision, mission, strategic objectives, performance measures, targets and strategic initiatives of an organisation.

Further, this unit will discuss the internalization of business performance measurement in an organisation.

Study Unit 2 – Science & Art of Data Visualisation

This unit introduces the concept, benefits and applications of data visualisation. There is a shift on the word visualisation's meaning from being an internal object in our mind to becoming an external object to support the decision-making process. It demonstrates the benefits and applications of data visualisation through various examples and shows the information processing model of human visual perception. Data visualisation is built upon data and the four components, which are visual cues, coordinate systems, scales, and context. These components work together and each of them affects each other.

Study Unit 3 – Four Stages of Data Visualisation

This unit introduces the four stages of data visualisation, together with three feedback loops. In addition, we will also discuss the types of data sets, namely, record data, graph-based data, and ordered data. This unit also discusses the first two stages of data visualisation in detail. Data collection and storage is the methodological process of gathering information about a specific subject and appropriately storing it in a secure. Data preparation is the process of identifying data quality issues and treating these issues so that data can be used for business intelligence (BI), analytics and data visualisation applications. As part of data preparation process, this unit also discusses different data types, dataset architecture and exploratory analysis techniques.

Study Unit 4 – Data Visualisation for Discrete and Continuous Data

This unit introduces several techniques for building data visualisation for categorical and time series data. For categorical data, we discuss the appropriateness of bar charts, stacked bar charts, side-by-side bar charts, pie charts, heat maps and tree maps vis-à-vis the visualisation task at hand. Time series data can be categorised as either discrete (bucketed) or continuous (unbroken). One consideration when presenting time series data is how we treat the discrete and continuous time series data. For the time series data,

depending on the visualisation problem, we study the suitability of each of the line charts (with trend and reference lines), stacked area charts and gantt charts. We learn to create all these charts using a problem solving framework in data visualisation software.

Study Unit 5 – Business Performance Dashboard

This unit introduces the concept of business performance dashboard- a visual display of the most important information needed to achieve one or more objectives, consolidated and arranged on a single screen so that the information can be monitored at a glance. In addition, it also discusses the three types of dashboards - strategic dashboard, tactical dashboard, and operational dashboard. We also demonstrate how to create an impactful business performance dashboard using the data visualisation software and following the effective dashboard design principles in a problem-solving framework.

Study Unit 6 – Advanced Data Visualisation Techniques

This unit introduces visualisation for spatial data. Spatial data are related to the location of the subject matter. There is a natural hierarchy to spatial data that allows, and often requires, us to explore at different granularities. We study maps, spider maps and path maps for representing spatial data. This unit also discusses visualisations to examine relationships among two or more variables in the data- scatter plot and bubble chart and visualisations to benchmark performance against target or predicted value for a variable in the data- bullet chart. It also demonstrates advanced visualisations such as pareto chart and forecasting using tableau.

3. Learning Outcomes

Knowledge & Understanding (Theory Component)

By the end of this course, you should be able to:

- Identify key metrics for business performance measurement.
- Explain the end-to-end process of creating business performance dashboard from data.

Key Skills (Practical Component)

By the end of this course, you should be able to:

- Use strategy map for business performance measurement.
- Inspect and prepare data for visualisation.
- Determine appropriate data visualisation techniques based on given data.
- Develop data visualisation dashboard using a software package.

4. Learning Material

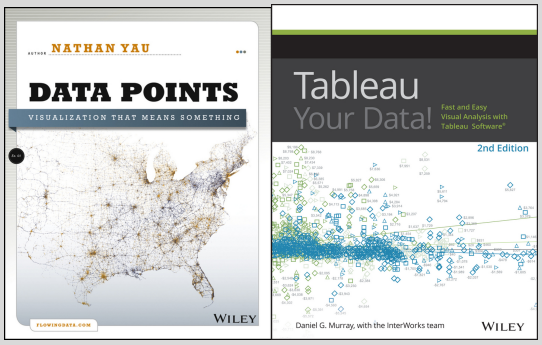
The following is a list of the required learning materials to complete this course.

Required Textbook(s)

Yau, N. (2013). *Data points: Visualization that means something*. John Wiley & Sons.

Murray, D. G. (2016). *Tableau your data*. John Wiley & Sons.

If you are enrolled into this course, you will be able to access the eTextbooks here:



To launch eTextbook, you need a VitalSource account which can be created via Canvas (iBookStore), using your SUSS email address. Access to adopted eTextbook is restricted by enrolment to this course.

Recommended software

1	Software	Microsoft Office software, and Data Visualisation software, like Tableau Desktop.

5. Assessment Overview

The overall assessment weighting for Day & Evening Time course is as follows:

Assessment	Description	Weight Allocation
Overall Continuous Assessment Score	Pre-Course Quiz 01	2%
	Pre-Class Quiz 01	2%
	Pre-Class Quiz 02	2%
	Tutor-Marked Assignment	18%
	Group-Based Assignment	20%
	Participation during Seminar	6%
Overall Examinable Score	End-Of-Course Assessment	50%
TOTAL		100%

The overall assessment weighting for the full online classes for this course is as follows:

Assessment	Description	Weight Allocation
Assignments	Pre-Course Quiz	2%
	Pre-Class Quiz 1	2%
	Pre-Class Quiz 2	2%
	Discussion	10%
	Participation	6%
	Tutor-marked Assignment	18%
	Group-based Assignment	10%
Examination	End-Of-Course Assessment	50%
TOTAL		100%

The following section provides important information regarding Assessments.

Continuous Assessment:

There will be continuous assessment in the form of assignments, test and participation. In total, continuous assessment will constitute 50 percent of the overall course grade. The assignments are compulsory and are non-substitutable. These assignments will test conceptual understanding of both the fundamental and more advanced concepts and applications. It is imperative that you read through your assignment questions and submission instructions before embarking on your assignment.

Examination:

The ECA will constitute the other 50 percent of overall student assessment and will test the ability to apply business analytics concepts, models and applications to solve particular business problems commonly faced by decision makers.

Passing Mark:

To successfully pass the course, you must obtain a minimum passing mark of 40 percent for both the Continuous Assessment and the End-of-Course Assessment separately. For detailed information on the course grading policy, please refer to the Student Handbook (Award of Grades section under Assessment and Examination Regulations). The Student Handbook is available on the Student Portal.

Non-graded Learning Activities:

Each study unit consists of activities for self-directed learning. These learning activities are meant to help you assess your own understanding and achievement of the learning outcomes. The activities can be in the form of Formative Assessments, Quizzes, Review Questions, or Application-Based Questions. You are expected to complete the suggested activities either independently or collaboratively.

6. Course Schedule

To help monitor your study progress, you should pay special attention to your Course Schedule. It contains study unit related activities including Assignments, Self-assessments, and Examinations. Please refer to the Course Timetable in the Student Portal for the updated Course Schedule.

Note: You should always make it a point to check the Student Portal for any announcements and latest updates.

7. Learning Mode

The learning process for this course is structured along the following lines of learning:

a. Self-study guided by the study guide units. Independent study will require at least 3 hours per week.
b. Working on assignments, either individually or in groups.
c. Classroom Seminar sessions (3 hours each session, 6 sessions in total).

iStudyGuide

You may be viewing the interactive StudyGuide (iStudyGuide), which is the mobile-friendly version of the Study Guide. The iStudyGuide is developed to enhance your learning experience with interactive learning activities and engaging multimedia. You will be able to personalise your learning with digital bookmarking, note-taking, and highlighting of texts if your reader supports these features.

Interaction with Instructor and Fellow Students

Flexible learning—learning at your own pace, space, and time—is a hallmark at SUSS, and we strongly encourage you to engage your instructor and fellow students in online discussion forums. Sharing of ideas through meaningful debates will help broaden your perspective and crystallise your thinking.

Academic Integrity

As a student of SUSS, you are expected to adhere to the academic standards stipulated in the Student Handbook, which contains important information regarding academic policies, academic integrity, and course administration. It is your responsibility to read and understand the information outlined in the Student Handbook prior to embarking on the course.

Study Unit 1

Business Performance Measurement

Learning Outcomes

At the end of this unit, you are expected to:

- Identify key metrics for business performance measurement
- Use strategy map for business performance measurement

Overview

This unit introduces the meaning, purpose and foundation of performance management. Through this process, an organisation establishes criteria to determine quality of their activities, efficiency, and effectiveness in conducting their business operations. Business performance metrics is used to measure different aspects of business and categorised under different business functions using the balanced scorecard framework. Strategy map serves as the system that provides linkages between vision, mission, strategic objectives, performance measures, targets and strategic initiatives of an organisation. Further, this unit will discuss the internalization of business performance measurement in an organisation.

Chapter 1: Business Performance Measurement Metrics

Lesson Recording

Business Performance Measurement Metrics

Business performance measurement is the process whereby an organisation establishes parameters within which resources, programmes, investments and acquisitions are to achieve the desired business results. Through this process, an organisation establishes criteria to determine quality of their activities, efficiency, and effectiveness in conducting their business operations.

1.1 What are Business Performance Measurement Metrics?

The ability to understand, manage and improve business performance is largely dependent on the ability to successfully measure business performance. By measuring their business performance, organisations will be able to derive meaningful information from such measures and present them in a report or graphical presentation for review or evaluation.

Business metrics are quantifiable measures used to track business processes to judge the performance level of your business. There are hundreds of these metrics because there are so many different kinds of businesses, with many different processes.

Generally, individual divisions or departments within a company, such as manufacturing, marketing and sales, are responsible for monitoring the metrics that track the performance of their respective parts of the business. Senior executives track more general metrics. CFOs, for instance, track earnings before interest, taxes, depreciation and amortisation (EBITDA), a universal measure of profitability, and the metrics that feed into it, such as net sales, operating expenses and operating profit.

The COO of a manufacturing company, meanwhile, might want to track the perfect order rate, a key performance indicator (KPI) to measure the performance of warehouse operations. CEOs are likely to closely monitor just a handful of summary metrics drawn from the dashboards of each of their direct reports.

Key business performance measurement metrics

1.1.1 Sales Metrics

Sales metrics measure and evaluate the sales-related performance and activities of an individual, team or company over a given period of time (for example, weekly, quarterly or annually). Analysing sales metrics helps identify what is and isn't working and provides insights into actions to take to improve sales performance. Table 1.1 lists a few key metrics to track in sales.

Table 1.1 Key sales metrics

Description	Calculation
Net sales revenue	
Represents a company's revenue after the associated deductions have been taken out	Net sales = Gross sales – Discounts – Returns – Costs associated with discounts and returns
Quota attainment	
Can show you where to concentrate your efforts. You can monitor the sales targets by measuring their performance against the quota.	Quota attainment = Amount of sales achieved by a particular representative or region / Goal for that rep or region E.g., if the goal is $10 million and the rep achieved $9 million in sales, they're at 90% of quota.
Growth rate	

Description	Calculation
Year-over-year growth is an important overall indicator of the health of your business. When compared to industry benchmarks, it tells you how well or how poorly your sales team is performing compared to the competition.	Sales growth rate = (Current year's revenue – Previous year's revenue) / Previous year revenue × 100 If your sales were $12 million this year and $11 million last year, your growth rate is ($12 – $11) / $11 × 100, or 9.09%.
Churn rate	
Churn rate is the percentage of customers who cancel or don't renew their contracts or subscriptions for a company's services or products. Rising churn rates could indicate a problem with a company's offerings or customer service approach, or it could mean the company is losing business to competitors.	Churn rate = Number of customers lost during period / Starting number of customers at beginning of period × 100 For example, if a company begins Q3 with 5,000 customers and ends Q3 with 4,000 customers, then the difference in the number of customers (1,000) indicates a 20% churn rate.
Lead response	
Besides quota attainment, you might want to look into how long it takes reps to contact a new lead. Lead response time can be immensely important in certain industries because the quicker a salesperson responds to a person's inquiry, the more engaged that person is likely to be and the greater the chance of a sale.	Lead response time = Sum of time between lead contact to sales rep response for all contacts / Total number of leads For example, if a sales rep is given 9 leads and responds to 5 leads within 1 day, 3 leads within 2 days and 1 lead in three days, the lead response time is (5 × 1) + (3 × 2) + (1 × 3) / 9 = 1.56 days.

Description	Calculation
Average revenue per user (ARPU)	
ARPU measures the amount of money that we can make from each customer.	ARPU = Total revenue / Total number of users
Average order value (AOV)	
AOV measures the amount of money that we can make from each customer order. A customer can place multiple orders.	AOV= Total order value/Number of orders

1.1.2 Marketing Metrics

There are so many ways for businesses to market and advertise their product or service — direct mail, email, websites, social media — that it's essential to know what mix works best. Adopting key marketing metrics helps your marketing team determine how effective its methods and channels are in supporting the success of your business. Table 1.2 lists some of the key marketing metrics.

Table 1.2 Key marketing metrics

Description	Calculation
Return on marketing investment (ROMI)	
ROMI focuses on the profits of incremental sales that can be attributed to marketing activity — or more simply, profit generated by the marketing department. ROMI can provide insights into the value of marketing activities in general or differentiate the relative	Return on marketing investment = (Sales growth – Marketing cost) / Marketing Investment × 100 For example, imagine you invest $10,000 in an email marketing campaign, which generates $60,000 in sales at a 20% margin, thus contributing $12,000 to company

Description	Calculation
performance of different marketing channels and campaigns.	profit. Your ROMI for this effort is (60,000 × .20 – 10,000) / 10,000) × 100 = 20%
Customer acquisition cost (CAC)	
How much does it cost to turn a prospect into a customer? CAC should take into account all marketing and sales costs, from salaries and benefits of the staff to the media spend to turn a prospect into a customer.	Customer acquisition cost = Total marketing and sales spend / Number of new customers If you invest $1 million in marketing and sales and get 500 new customers, your CAC is $1,000,000 / 500 = $2,000 per customer.
Customer lifetime value (CLV)	
CLV is the profit earned from a customer over the entire time they remain a customer.	Customer lifetime value = (Average transaction value × Average number of transactions in a year × Average customer retention in years) × Profit margin Suppose a company with an overall 20% profit margin retains customers for five years on average. The company has an average transaction value of $100 and each customer makes 10 purchases per year. Its CLV = (100 × 10 × 5) × .20, or $1,000
Customer retention	
Customer retention is the percentage of existing customers that stay during a	Customer retention = (Number of customers at end of a period –

Description	Calculation
specific period of time. Knowing how costly it is to acquire new customers demonstrates how important it is to retain the customers you already have.	Customers added during period) / Number of customers at beginning of period For example, if a company had 500 customers at the start of a year, added 50 customers during the year and ended with 500 total customers, it would have a customer retention rate of (500 – 50) / 500, or 90%.
Customer engagement score	
Customer engagement scores can help you understand how much and how often your customers engage with your SaaS (Software as a Service) solution, such as how often they log in, how often they use specific tools and features, what they use the software for and more.	There's no one formula to calculate a customer engagement score, so a business must create its own model and system to do so. Create a list of inputs or actions that predict customer engagement, perhaps based on habits of long-term customers. Then score each input or action based on how critical it is to customer retention and add up each customer's engagement score. Continually evaluate your rating system to ensure you're appropriately picking the right features that predict retention and churn.
Net promoter score (NPS)	
NPS estimates the likelihood users will recommend your service to others. NPS	NPS = Percentage of promoters – Percentage of detractors

Description	Calculation
is usually measured through a one-question survey to customers: "How likely are you to recommend us to a friend or colleague?" with a 0-to-10 scale (where zero means they won't recommend your product and 10 means they definitely would). Respondents are then bucketed into the following categories: Detractors, or respondents who answer with a 0 to 6. Passives, or respondents who answer with a 7 or 8. Promoters, or respondents who answer with a 9 or 10.	For example, out of 100 survey respondents, 20 are detractors, 50 are promoters and 30 are passives. Your NPS would be 30 (50% – 20%).
Cost per lead (CPL)	
It is the cost to identify, attract, qualify and retain a customer. Determining how much each lead costs will help you allocate your budget appropriately. But just because a particular channel incurs a higher CPL, doesn't mean you should drop it— those customers might actually convert at a higher rate or spend more than customers gained through a lower-CPL channel.	Cost per lead = Total marketing spend / Number of new leads

Description	Calculation
Cost per click (CPC)	
It is dependent on the number of people clicking through to your website from paid ads. This includes traffic from linked Google ads, and other paid advertisements. We can track this through campaign-tagged URLs where the medium is defined as 'cpc' or 'paid'.	CPC = cost of an advertising campaign / Number of clicks Example: If we paid $500 for an ad and we get 200 clicks, then the CPC = $2.5
Website traffic-to-lead ratio	
A sales qualified lead (SQL) from your website is someone that is not only aware of the company but interested enough to enter information about themselves on the website in order to, for example, get past a filter or to get your newsletter.	Website traffic-to-lead ratio = Number of leads / Number of unique website visitors A business whose website is visited by 500,000 individuals in a month, 5,000 of whom convert to leads, has a traffic-to-lead ratio of 1%.
Conversion rate	
Conversion rate is a way to measure the percentage of users or customer prospects who complete a desired action, such as making a purchase, registering an account or starting a free trial. Tracking this metric can help you get a feel for how well your marketing strategy is working.	Conversion rate = (Conversions / Total unique visitors) × 100 For example, suppose a subscription business offers a free trial to 1,000 potential customers in total, and 200 of them take advantage of it. The conversion rate is (200 / 1,000) × 100, or 20%.

Description	Calculation
Website bounce rate	
Like conversion rate, this metric can help you track how effective your marketing strategy is. Bounce rate tracks how well a website landing page generates visitor interest by calculating the percentage of visitors who enter the site and leave before viewing other pages within the same site.	Bounce rate = (Number of site visits that access only one page / Total number of site visits) × 100 If a site has 100,000 visitors, and 50,000 of them view only one page, its bounce rate is (50,000 / 100,000) × 100, or 50%. The higher the bounce rate, the less likely the site engages customer interest. A low bounce rate is ideal.
Cart abandonment rate	
It is the percentage of customers who added at least an item to carts and then abandon the purchase.	Example: 200 customers added items to the cart, but only 120 of them proceeded to the checkout page. The cart abandonment rate is therefore (200-120)/200 = 40%
Checkout abandonment rate	
This is the percentage of customers who initiated checkouts and then abandoned purchases.	Example: If on a website, 120 customers proceeded to checkout, but only 60 of them completed the purchases, then the checkout abandonment rate = (120-60)/120 = 50%
Lead response	

Description	Calculation
How long does it take us to respond to a new lead? Faster response to a person's inquiry often result in better satisfaction and greater chance of a sale.	Lead response time = Sum of response time durations / Total number of leads Example: If we have a total of 9 leads, and if we respond to 5 leads within 1 day, 3 leads within 2 days and 1 lead in three days, the lead response time is (5 × 1 + 3 × 2 + 1 × 3) / 9 = 1.56 days
Traffic source	
Improve marketing and advertising efforts by knowing which source is working well. Sources include paid, organic search, referrals, direct, and others.	
Traffic volume	
Traffic volume reflects level of activity on the business's website.	This is measured in visitors (users) and times visiting the site (sessions).

1.1.3 Financial Metrics

For finance teams, the metrics that matter most are the ones that reflect the financial health of the business. After all, a company's survival hinges on its financial health. Thus, most financial metrics concern factors like revenue, cash flow, accounts receivables and assets and liabilities, as illustrated in Table 1.3.

Table 1.3 Key financial metrics

Description	Calculation
Net income	
Also known as the bottom line, **net income** is an important starting point for calculating other key metrics, like net profit margin and earnings per share. Net income can help assess whether revenue exceeds business expenses and, if so, by how much.	Net income = Total revenue – Cost of goods sold – Operating expenses – Other expenses – Interest – Taxes – Depreciation and Amortisation Net income is different from gross income, which only subtracts the cost of goods or services sold from revenue.
Net profit margin	
Measures how much actual profit is netted for each dollar of revenue made. This is important because revenue increases may not always translate into increased profitability.	Net profit margin = (Net income / Total revenue) × 100
Gross profit margin	
Unlike net profit margin, gross profit margin shows a company's profits before subtracting interest, taxes and operating expenses like rent, utilities and wages.	Gross profit margin = (Revenue – Cost of goods or services sold) / Revenue
Current ratio	
Measures a company's ability to pay off financial obligations that are due within a year. Current assets are those expected to convert to cash within a year (such as accounts receivable), while current	Current ratio = Current assets / Current liabilities Generally, a current ratio above 1.0 is considered healthy. A ratio of 2.0, for

Description	Calculation
liabilities are obligations due within a year (such as accounts payable).	example, suggests the business has two times more current assets than current liabilities. However, a current ratio above 3.0 could indicate the business isn't efficiently handling working capital. Note that current ratio is only a quick, short-term snapshot of solvency and must be calculated regularly.
Working capital	
Keeping a close eye on working capital can help you figure out ways to free up cash, use funds more effectively or learn to reduce dependence on outside funding, while getting a clear sense of the business's liquidity.	Working capital = Current assets – Current liabilities
Accounts receivable turnover ratio	
Measures how effectively the accounts receivable department collects debt owed by clients. The higher the ratio, the better the company is at collecting payments, which makes it more likely to have cash on hand to make its own payments or invest in growth. A lower turnover ratio can indicate illiquid customers, slow-to-pay customers or an inefficient debt collection process — potentially stunting a business's growth.	Accounts receivable turnover ratio = Net credit sales in a given period / Average accounts receivable of period

Description	Calculation
Percentage of accounts payable overdue	
Indicates cash flow problems — the more overdue payments, the more likely the business is having trouble paying suppliers, indicating a need for funding or a new business strategy. The lower the percentage, the better a company is at paying its debts on time.	Accounts payable overdue rate = (Accounts payable overdue / Total accounts payable) × 100

1.1.4 Human Resources Metrics

Human resources metrics can help indicate employee satisfaction and performance. These metrics generally track data related to employee turnover, development and engagement, company culture and training costs — all of which can help you spot workforce trends and dynamics and proactively solve potential issues, like burnout or ineffective training programmes. Table 1.4 lists down key human resources metrics.

Table 1.4 Key human resources metrics

Description	Calculation
Employee turnover rate	
High turnover rates can reflect talent management issues, unhappy workers or a pattern of hiring employees unfit for their positions.	Turnover rate = (Number of separations in a given period / Average number of employees in period) × 100
Revenue per employee (R/e)	
The more revenue per employee, the more productive a business is and	Revenue per employee = Total revenue / Current number of employees

Description	Calculation
the more likely it's efficiently using resources — both of which can directly relate to greater profits.	
Employee net promoter score (eNPS)	
Like the traditional NPS, eNPS offers a standardised approach to understanding how employees feel about the company, using a scale from 0 to 10.	eNPS = Percentage of promoters – Percentage of detractors
Training spend	
Companies should track training expenses to see whether they're getting a return on their investment. Tracking training expenses alongside employee productivity and profitability can help a business determine whether training strategies are effective.	Training spend per employee = Total training expenses / Total number of employees
Career path ratio	
This metric helps track the ratio of vertical promotions to lateral transfers. Tracking career path ratio can help a company measure employee mobility.	Career path ratio = Total promotions / (Total promotions + Total transfers) Values above 0.7 indicate more vertical promotions, meaning the organisation may be getting too "top heavy" and should look to start expanding roles laterally. Values under 0.2 indicate more lateral transfers, suggesting not

Description	Calculation
	enough employees are being primed for promotion.

1.1.5 Other Business Metrics

To provide a sense of the sheer volume of business metrics available, Table 1.5 illustrates a few additional metrics that can be useful for the C-suite, inventory teams, manufacturing companies and other business departments and industries.

Table 1.5 Additional metrics

Description	Calculation
Revenue vs. forecast	
This metric can help executives see whether company performance is matching expectations or coming up short. If actual revenue is falling short of expectations, executives must act to find out what is causing the disconnect.	Variance percentage = ((Actual – forecast) / Actual revenue) × 100
Inventory turnover rate	
A crucial financial metric for manufacturing and retail companies, inventory turnover rate tracks how many times a company sells and replaces its inventory over a given period. The higher the number the better, because it means the company is holding less inventory — and holding inventory is expensive. However, if a high inventory	Inventory turnover rate = (Cost of goods sold / Average inventory) × 100

Description	Calculation
turnover rate is accompanied by lost sales due to item unavailability, the company is probably running too lean.	
Scrap	
Measuring scrap, or waste, is a key metric for manufacturing teams. Specifically, scrap measures the amount of rejected or unusable manufactured items, usually due to manufacturing defects.	Scrap = (Total scrap units, volume or weight / Total product run in units, volume or weight) × 100 The less scrap, the more effective the manufacturing process.
Return on assets (ROA)	
ROA calculates the per-dollar profit a company makes on its assets and is used to assess profitability. This is a particularly important metric for the banking industry.	Return on assets = Net income / Total assets The higher the ROA, the more efficient a company is with its assets. However, it is important to note that bank ROAs generally hover around 1% — which is still considered a healthy number for the industry. This is because banks often have more debt than equity.
Average support ticket resolution time	
A key metric for customer service departments, average support ticket resolution time tracks how long it generally takes to resolve support tickets. Although response time is also	

Description	Calculation
important, resolution time is a better metric because a short resolution time typically will indicate quick responses as well. Support teams should strive to have fast resolution times in order to keep customers satisfied.	
Customer satisfaction	
Customer satisfaction is critical to a successful business. One way to measure customer satisfaction is through a CSAT, or customer satisfaction, scale. These are typically simple questions that follow up on a single customer experience: "On a scale of 1 to 10, how satisfied were you with X experience?" with 1 being extremely unhappy and 10 being extremely happy. Businesses then add up all the scores and divide that figure by number of respondents to arrive at a customer satisfaction value. The higher the number, the more satisfied customers are with the experience.	

1.2 Why It Is Important to Track Business Metrics

Tracking the metrics that are most important to your business — and managing operations based on the results — maximises the business's chances of success. It is that simple and yet that hard. The key word here is "important". For instance, executive search professionals might track how many candidates they bring to a client. But what matters

to the client is the speed with which the position is filled and the quality of the candidates; those are the important metrics to measure and track.

Performance improvement: Tracking the right business metrics tells you how well or poorly the business is doing and provides direction for how to improve operations.

Comparative analysis: Tracking business metrics reveals whether the business is over- or underperforming on key industry benchmarks.

Alignment: Business metrics can be used to ensure the entire company is working toward shared organisational goals.

Compliance: Mandates to track certain business metrics from governmental and other regulatory agencies require companies to monitor them to stay in compliance.

Communication: Reporting business metrics is a vital communications tool for customers, shareholders, employees or society at large.

Identifying problems: Analysing business metrics can help identify emerging problems in time to correct them before they become major pain points.

1.3 What Business Metrics Should You Use?

This is a vital question because there are so many business metrics to choose from. In the absence of clarity around business goals, some organisations can go "metrics-crazy" and try to monitor too many things. While every business is different and, therefore, the metrics that matter are different for each of them, these three questions can be powerful tools to identify what matters most to any business:

1. Is the metric directly relevant to the performance of the business?
2. Does it help predict future performance in a useful way?
3. Can it be reasonably measured?

Figure 1.1 Examples of key business metrics for different businesses

Chapter 2: Balanced Scorecard

Lesson Recording

Balanced Scorecard

There is agreement in the management and performance evaluation literature that performance is a multi-dimensional construct (Kaplan & Norton, 2005; Moore, 1995; Wongrassamee, Simmons & Gardiner, 2003) and performance measurement has been recognised as a key component of good governance. A robust model of business performance measurement requires measures to be established in all the areas so as to provide a holistic picture of organisation business performance now and in the future.

The development of an accounting system is synonymous with the industrial age where the only means of looking at performance was by looking at the financial results of organisations. For many years, business performance was judged based on financial results, usually at the end of one financial year. By reviewing business performance based on financial results, it is simply looking at events that are already over. This approach is considered as reactive instead of proactive.

In 1998, David Norton and Robert Kaplan (Harvard Business School) described how many world-class organisations have adopted business performance measurement systems way beyond looking only at financial objectives alone. Kaplan and Norton correctly pointed out that measuring financial results alone was flawed in that they were documented after the event. Kaplan and Norton described it "like driving a car by looking at the white line in the rear view mirror."

In their quest for success, the above world-class organisations placed significant emphasis on non-financial measures. To provide a more holistic view of an organisation's overall performance in meeting the competition, and strategic objectives, Kaplan and Norton

sought a better model of performance measurement. They came up with the concept of a balanced scorecard through their research.

Without a balanced scorecard, a business tends to be judged only by short-term financial results, and this may hide serious problems—for example, the reduction of short-term costs by deferring all maintenance expenditure. If the BSC was to include a question that probes the "percentage of maintenance completed on schedule," this safety issue would surface.

The BSC enables organisations to clarify their visions and strategies, and puts them into action. It provides feedback around both the internal business process and external objectives in order to continuously improve on strategic performance and results. When fully deployed, the BSC transforms strategic planning from an academic exercise into the nerve centre of enterprise.

The BSC suggests that we view an organisation from four perspectives or dimensions, and we develop measures, collect data and analyse the data relative to each of the following perspectives:

- The learning and growth perspective, which is sometimes referred to as the innovation and development perspective
- The internal process perspective
- The customer perspective
- The financial perspective

2.1 The Four Balanced Scorecard Perspectives

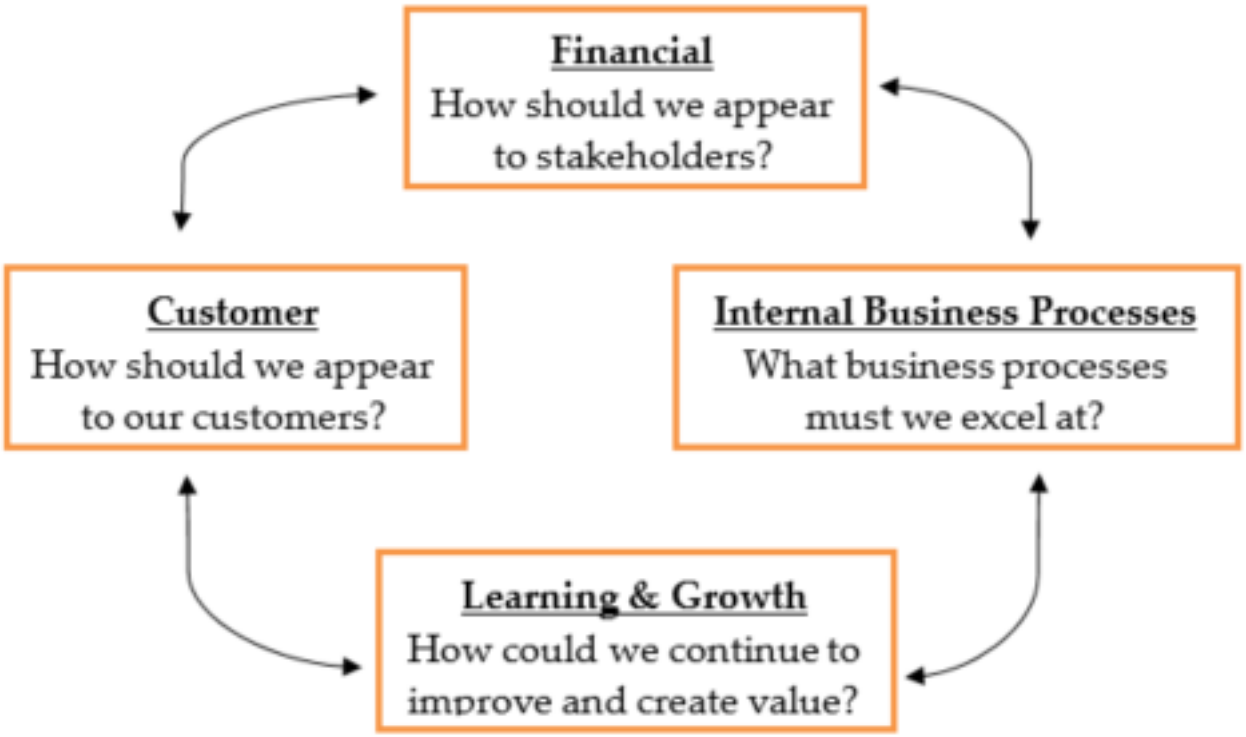


Figure 1.2 The four perspectives of the balanced scorecard

The BSC advocates a set of measures that relate to the four major perspectives, which are financial, customer, internal processes, and learning and growth, to determine the overall performance of an organisation. Customer, internal processes, and learning and growth perspectives can be considered as operational measures that ultimately drive financial performance. The management will then have a complete view of an organisation's performance when viewing the overall performance in the current period compared with the previous period.

2.1.1 The Customer Perspective

Management philosophy in recent times has shown an increase in realisation of the importance of being customer focused (i.e. customer satisfaction) in any business. This is a leading indicator because if customers are not satisfied, they will eventually find other suppliers that will meet their needs. Poor performance from this perspective is thus a leading indicator of future decline in an organisation's performance, even though the current financial outlook may look good.

Under the customer perspective, an organisation needs to determine what the customers require and think are important in terms of the products and services that the organisation provides them. Although customers' needs are varied and different from customer to customer, the needs also depend on the types of product and services offered. There are some basic customer requirements that are usually common across industries. These basic requirements relate to time, quality, service, functional performance, and value. Organisations have to establish goals for each key requirement. Measures must then be selected to gauge the organisation's success in achieving these goals.

Examples of measures:

- Net promoter score
- Customer satisfaction
- Customer engagement
- Product returns
- Customer retention

2.1.2 The Internal Business Processes Perspective

Measures based upon internal process perspective allow managers to know how well their business is running, and whether its products and services meet customer requirements. Under this perspective, the organisation must identify key business processes that need to be excelled in terms of organisation's operations. Measures must then be established to monitor the performance of these key business processes.

For example, if one of the key business processes is new product introduction, then a possible measure can be the new product lead time, which is the time between the conception of new product idea and the actual introduction of the new product in the market.

As every organisation is unique in its composition, the measures for internal process perspective have to be carefully designed by those who know these processes most intimately.

Examples of measures:

- Innovation lead time
- New products introduced
- Inventory turnover
- Product adoption rate
- Grievance resolution lead time

2.1.3 The Learning and Growth Perspective

This perspective includes employee training and corporate cultural attitudes related to both the individual and corporate self-improvement. In a knowledge worker organisation, people – the only repository of knowledge – are the main resource. In the current climate of rapid technological change, it is becoming necessary for knowledge workers to be in a continuous learning mode.

Innovation usually refers to an organisation's ability to introduce new products and services, while *development* usually refers to an organisation's ability to continually make improvements to its key processes and improve productivity. The organisation will then have to establish goals that relate to innovation, growth and development. Measures must be selected so that they can gauge the organisation's success in achieving the goals. For example, an organisation might set the following goal related to innovation — to increase the number of patents filed per year. A measure for this goal might be the number of patents filed annually per research and development (R&D) employee.

Kaplan and Norton emphasised that learning is more than just training. It includes people like mentors and tutors within the organisation, as well as that ease of communication amongst workers that allows them to readily receive help when needed. Learning also includes acquisition of technological tools or high performance work systems.

At times, an organisation may find it difficult to hire new talents, at the same time there is a decline in training of existing employees. This is a leading indicator of "brain drain" that must be reversed. Measures can be put in place to guide managers in focusing training funds where they can help the most. Learning and growth constitute the essential foundation for the success of any knowledge-worker organisation.

Examples of measures:

- Employee turnover
- Employee attrition
- Training hours
- Absenteeism
- Job satisfaction

2.1.4 The Financial Perspective

Kaplan and Norton do not disregard the traditional need for financial data. Timely and accurate funding data will always be a priority, and managers will do whatever necessary to provide such data. In fact, often, there is more than enough handling and processing of financial data. The important point to take note of is that over emphasis on financial information can lead to an "unbalanced" view of performance. Emphasising non-financial related information such as risk assessment and cost-benefit data are equally important.

The measures for the other three perspectives that relate to customer, internal business process and people must be appropriately selected in order (for financial performance) to yield results. If the measures in the financial perspective are not showing gains while measures in the other three perspectives are, it could mean that the organisation has set the wrong goals.

For any profit-oriented organisation, the bottom-line profitability is important. Although the concept of the BSC is to take performance in many non-financial areas into consideration, the measures in the financial perspective must eventually show gains in order for the organisation to conclude that its overall performance has indeed improved. Typical financial objectives are growth, profitability and shareholder value. For instance, take growth as a goal, an organisation may set a goal that relates to sales growth. A measure for sales growth could then be the increase in sales dollar.

The BSC provides substantial focus, motivation and accountability in both profit and non-profit organisations. It provides the rationale for their existence (serving customers), and communicates to all stakeholders the outcome and performance drivers by which the organisation achieves its mission and strategic objectives.

Examples of measures:

- Revenue
- Net profit
- Sales growth
- Return on assets
- Return on investment
- Total expenditure

Chapter 3: Strategy Map

Lesson Recording

Strategy Map

Business performance measurement in itself serves limited purpose unless it provides linkages between vision, mission, strategic objectives, performance measures, targets and strategic initiatives in order for it to be considered an invaluable asset to the organisation. The term, "strategy map," can be used to describe such a system.

The *"strategy map"* lays out the process of executing strategies through planned activities that are intended to achieve certain strategic objectives. Measurable business performance indicators are then established to track advancement towards strategic objectives, and to drive necessary improvement actions that will bring the organisation towards its intended strategic objectives. A strategy map framework is shown in Figure 1.3.

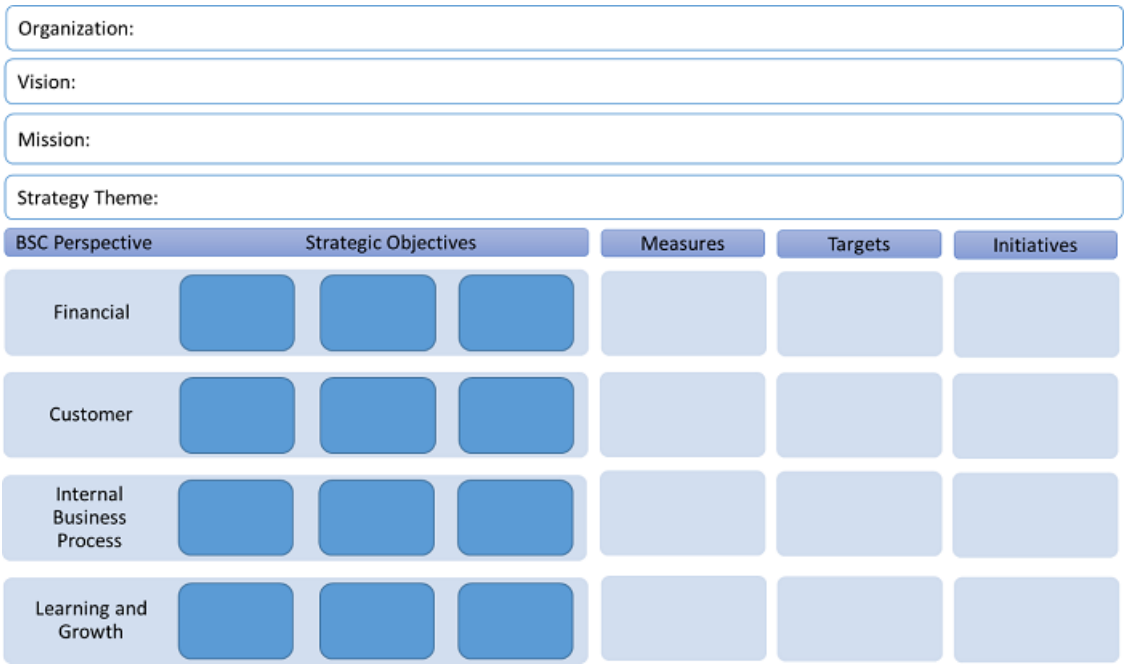


Figure 1.3 Strategy map framework

3.1 Strategy Map

A *strategy map* is a diagram used to communicate the strategic objectives organisations would like to achieve. When we use a strategy map for the balanced scorecard approach, individual strategic objectives are positioned systematically to demonstrate the relationships between them, and reveal the logic under each strategic theme. It also provides the management with a framework for describing and managing strategies in a knowledge-worker and strategy focused organisation.

Each strategic objective is related and linked to other strategic objectives on the strategy map using lines that represent cause-and-effect relationships. Such linkages logically reflect the natural value chain and culture. The cause-and-effect relationships are critical components of a strategy map. They are represented by lines with arrows, linking various items in the strategic map such as strategies, strategic themes, strategic objectives and business performance measures.

Figure 1.4 illustrates the architecture of a strategy map for a balanced scorecard approach for a retail company specialising in men's apparels. The cause-and-effect logic of this

design constitutes the hypotheses of the strategy. The financial perspective consists of two themes: *growth* and *productivity*, for improving shareholder value (i.e. profit). The themes in the customer perspective emphasise the importance of customer shopping experience, brand image, and product fit.

There are four strategic themes in the internal business process perspective— brand dominance, fashion excellence, sourcing and distribution networks, and customer shopping experience— to build operational excellence, increase customer values and build franchise. Finally, the four strategic themes in the learning and growth perspective are strategic awareness, goal alignment, staff competence, and technology infrastructure. They support the four internal business processes.

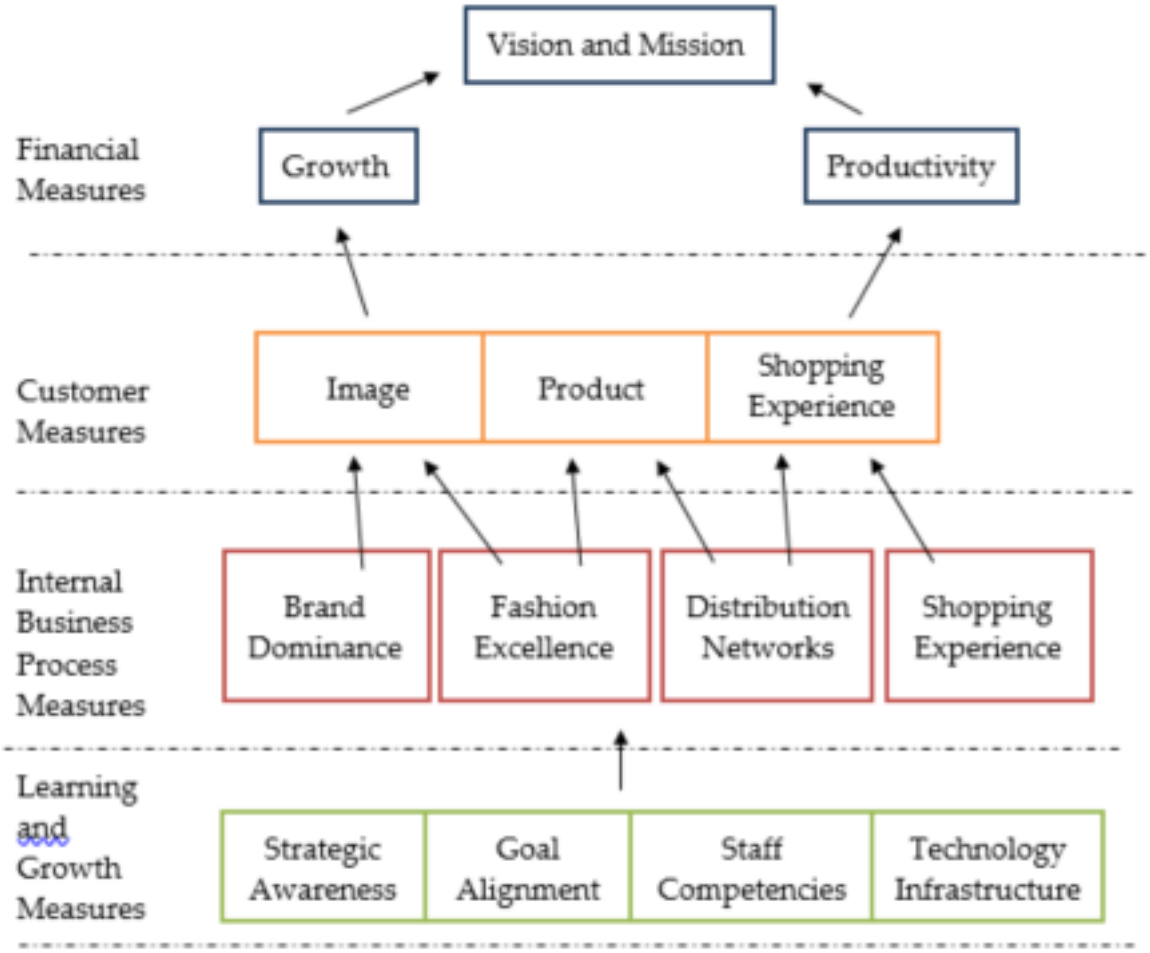


Figure 1.4 A fashion retailer's balanced scorecard strategy map

3.1.1 Vision, Mission and Strategy

Although vision and mission statements provide a sense of direction, many people often could not draw a distinction between the mission statement and the vision statement. By definition, these statements can be defined in simple terms:

- A vision statement sets out the desired future state of the organisation.
- A mission statement is the reason for the organisation to exist.

Further understanding can be achieved by looking at the period which these accomplishments are made as shown in Figure 1.5.

Figure 1.5 Vision and mission statements

Strategy formulation and strategic planning are important tasks carried out by the management to steer the organisation to gain higher performance and greater profitability. Strategies formulated by organisations are usually relevant for three to five years, depending on the competitive landscape and technological changes.

A strategy is not formulated in a vacuum. It is usually guided by the organisation's mission, vision, and core values. Although the form of strategy being established is a product of competition and operating environment, the ideals of the strategy always revolve around the mission, the vision and the core values of an organisation.

While strategy planning is a key task, of equal importance is its implementation, many organisations fail not because they do not have good strategies, but because they do not execute their strategies effectively. A business performance measurement framework serves as an important role in supporting the execution of strategy within an organisation. Once a strategy is formulated by the CEO, it must be translated into strategic objectives and business performance measures. With the establishment of business performance measures, performance of various business activities within the organisation can be monitored towards achieving its strategic objectives.

Here are some examples of strategy established by different organisations:

- Increase growth of earnings to 15% in three years and achieve a return on equity of 20%, so that the organisation will become the leader in its new products and attain lower costs than its competitors.
- Growth (through diversification) by merging with or acquiring firms in different, but related industries.
- Satisfy customers by providing quality products, reducing time taken to offer new products in the market, improving efficiency of all plants and processes, and building teamwork amongst employees, suppliers and dealers.
- Become the market leader by offering best-of-class products to customers.
- Protect and improve the firm's position as the number one brand globally. Build a strong momentum in growing the female consumer market. Continue to drive for increased margins through efficient inventory management.
- Global product strategy: same product and same marketing approach everywhere.
- Return to profitability: stabilise operation by downsizing, retrenchment and revitalising organisational resources and capabilities.

A few observations can be made by reviewing the strategies above. Most of the strategies have elements of improving the firm's financial performance, competitiveness and market position. To the top management, the broad interpretation of the strategies is quite obvious. However, to the lower management level, this understanding may be unclear or lacking.

3.1.2 Strategic Themes

Strategic themes or *strategic thrusts* are the main high-level business strategies that an organisation must excel in in order to achieve its vision. Strategic themes knit together independent activities and focuses on effort and resources of functional groups that are significantly important. For instance, a budget airline's strategic theme can be "Now Everyone Can Fly", while a national airline's strategic themes can be "Your Best Travelling Experience".

Strategic themes can be regarded as an organisation's pillar of excellence when they apply it to every part of the organisation, and they define what major areas the organisation will focus on to achieve its vision. This is illustrated in Figure 1.6.

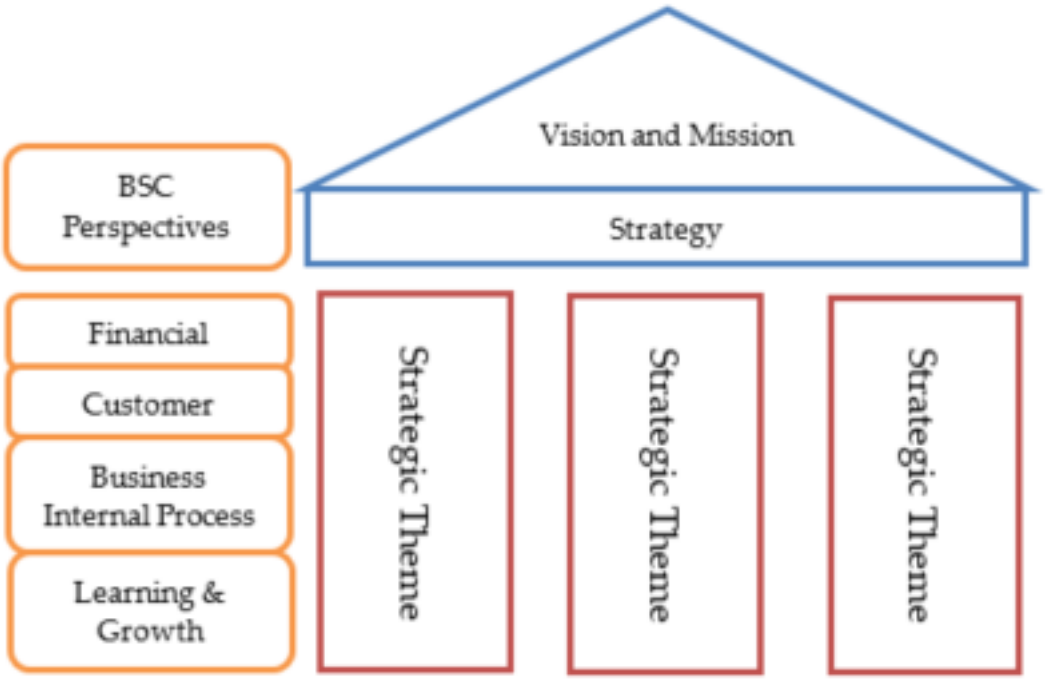


Figure 1.6 Strategic themes as pillars supporting an organisation's vision and mission

Strategic themes are intended to create shared understanding of what strategic objectives are, and actions to focus on (usually over a period of three to five years planning horizon). They should be designed to be measurable since they are considered to be something similar to major or top objectives. Furthermore, strategic themes can be used to separate an organisational strategy into several focused themes, and these themes can affect the four BSC perspectives.

Examples of a set of strategic themes for a leading fast-food-chain company include improving operating margin by re-engineering supply chain processes and reducing cost of capital employed.

3.1.3 Strategic Objectives

Strategic objectives are briefly defined objectives on the strategy map, which require further supporting information, initiatives and plans on how they can be achieved. Ideally, initial short-form wording of objectives should be meaningful, tangible, and written in everyday

language that allows middle and lower level management personnel to understand the purpose of performing certain activities.

Using the BSC as a platform, strategic objectives can be established for any broad strategy devised by an organisation. In the BSC model, each perspective has a defining question. For example, from the financial perspective, the defining question is: How should we appear to stakeholders?

In Figure 1.7, under the financial perspective, there are two strategic themes: financial growth and productive use of financial resources. These themes are translated into two strategic objectives, such as to increase revenue or to lower operations cost.

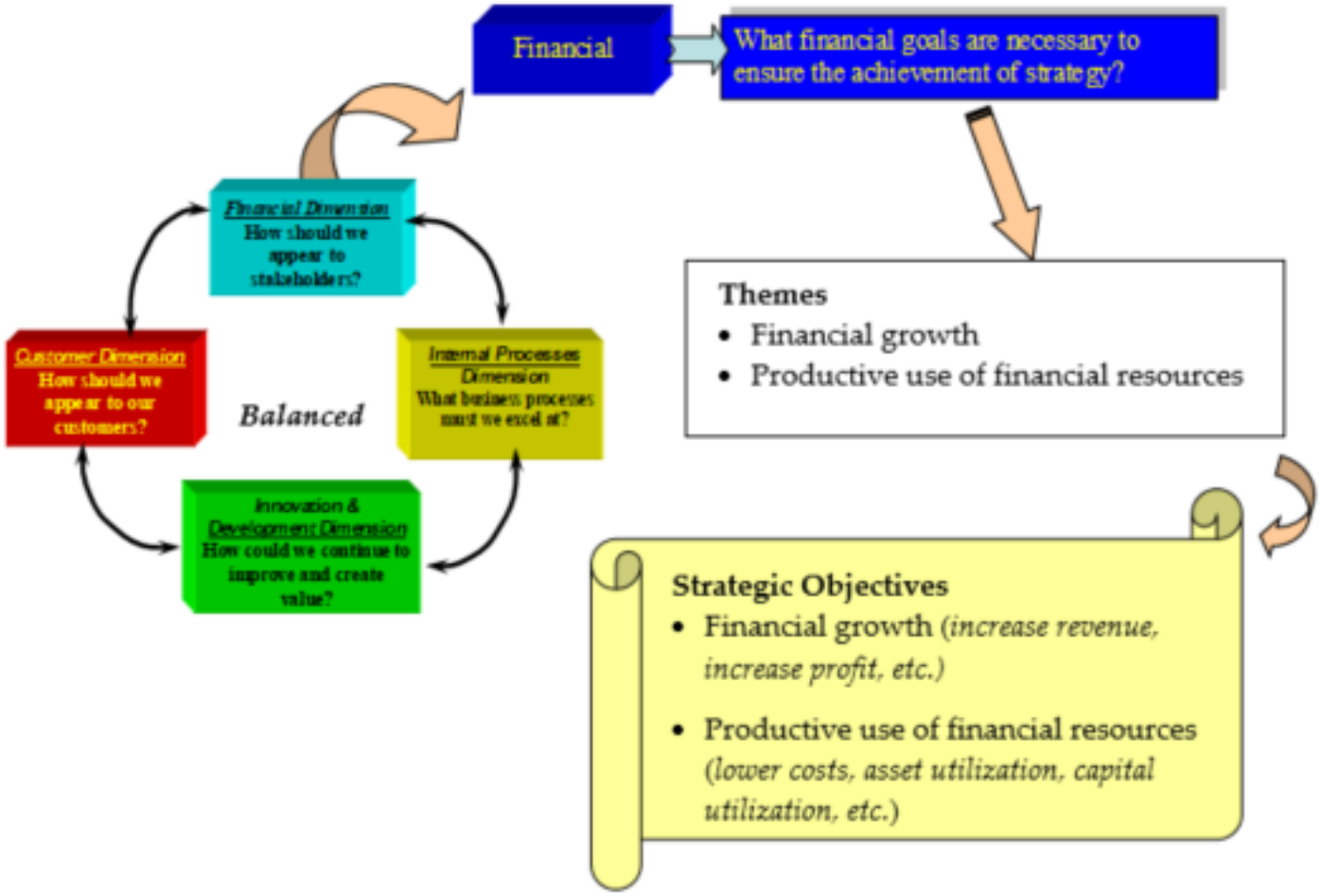


Figure 1.7 Strategic themes to strategic objectives under the BSC financial perspective

3.1.4 Measures

In order to monitor if the strategic objectives can be accomplished, the right measures must first be defined. Business performance measures are intended to provide objective information on the effectiveness and efficiency of a particular process or activity. This will lead to the accomplishment of goals. Linking a performance measure to a strategic objective provides relevance for the measurement. It also helps drive the organisation's activities towards achieving the organisation's strategy.

To select the appropriate measure, the considerations that need to be made are:

1. **Is it measurable?**

   This relates to the objectivity of the measure and the ability of current technology to perform the measurement. Not all situations are directly measurable (e.g. customer or employee level of satisfaction). If no direct measure can be used, proxy measures are used in place. It is important that sufficiently accurate proxy measures are used to represent the direct measurement sought.

2. **Are the data currently available or easily available?**

   Although a good measure can be selected after careful consideration, it may turn out that trying to collect data relating to the measure may be beyond the capability of the organisation at the current state of operations. Thus, data availability has to be factored into the selection of measures. Most of the time, business performance data are available, thus it is a matter of standardising the procedures to formally collect and collate data. However, if business performance data are not readily available, the question is then whether they can be made easily available. In such situations, new instruments may have to be used to generate the data. The instruments can be in the form of new sensors, new IT systems, or just new survey questionnaires. It should be noted that when data are not easily available, considerable resources will have to be used to generate the data.

3. **What should the frequency be?**

   In determining the measurement frequency, it is important to select an appropriate time period for every measure. Measures associated with strategic performance tend to be measured on a quarterly or yearly basis, whereas measurements associated with operational performance are more likely to be taken monthly, weekly, or even daily.

Once measures have been identified, processes are then designed to collect data for these relevant measures with the objective of improving collection efficiency, storage and presentation.

The measure attributes to be defined and documented are

- the name of measure,
- relation of measure to a specific strategic goal,
- a specified target,
- formulae (if any) to compute values for the measure,
- frequency of data collection, and
- responsibility and ownership.

At times, it may be difficult to create a business performance measure for business performance that is not easy to quantify, such as customer relations improvement. Although such measure is common and popular, measuring it is not always straightforward or possible. But we can measure components that contribute to customer relations improvement, such as the number of customer complaints satisfactorily handled during the month, or the average time required to resolve a customer's problem.

Likewise, we may encounter a situation where there may not be a single measure that properly or accurately represents the intent of a strategic objective. In such a circumstance, a special type of measure called a *proxy* (or *surrogate*) *measure* can be used. A proxy measure is an alternative choice of measurement used when a direct measure is unavailable. For example, a hospital assesses the effectiveness of medical facilities with a proxy measure like patients' average recovery time. Another example is the number of female members of a chamber of commerce as a proxy measure for estimating the percentage of female business owners.

A proxy measure is useful in providing a close approximate indication of actual business performance. Although it may not be perfect or conceptually ideal, it, nonetheless, can intuitively represent business performance until a better measure is available. A proxy measure is usually adopted when the data of the ideal measure is unavailable, or too expensive to collect, or when technically better measures are not intuitive to stakeholders. Proxy measures are not usually used as long-term substitutes for ideal

measures. However, an organisation will be better off using proxy measures than not measuring at all, until a better measure comes along.

The measures under the balanced scorecard approach should contain a mix of leading and lagging indicators (measures) of business performance. If there are little or no leading indicators (i.e. leading indicator is essentially historical representation of business performance), we will end up knowing little about the "how" when it comes to operations. Conversely, a preoccupation with leading indicators will not reveal whether improvements are leading to improved process and results. Table 1.6 summarises the differences between leading and lagging indicators.

Table 1.6 Differences between lagging and leading indicators (measures)

	Lagging	**Leading**
Definition	Measures focus on results at the end of a time period Based on historical performance	Measures that drive or lead to the performance of lag measures Measures intermediate process and activities
Examples	Revenue Employee satisfaction	Punctuality Absenteeism
Advantages	Often easily identified and captured	Predictive in nature, and allows an organisation to make adjustments based on results
Issues	Historical in nature and do not reflect current activities Lack predictive power	May prove to be difficult to identify and capture. Often, new measures have no history
The balanced scorecard should contain a mix of leading and lagging indicators.		

The BSC approach strategy maps are designed to help an organisation navigate its journey and the changing tides. As such, they must be updated to ensure that they remain relevant and effective. In today's environment, minor updates or modifications can be expected within 1-2 year's timeframe, while major changes occur when there is a change in business focus or strategy, resulting from a change of CEO or leadership.

3.1.5 Targets

Once measures are selected, the next stage is to set targets that have to be achieved by an organisation. *Business performance targets* are objective values that share the same unit of measurement as the measures per se. *Targets* represent a performance level that the organisation wants to achieve in order to fulfil its stated strategy either partially or fully. With the identified targets, the gap between current and intended performance level can be ascertained. Business performance targets give a sense of direction and purpose to establish strategic initiatives that will drive improvement activities and breakthroughs in the processes, and those activities that are being measured. Targets drive change by influencing behaviours, so people try to do the right thing the right way.

Business performance targets can be short-term (achieve within one year), mid-term (achieve within three years), or long term (achieve within five years). They usually correspond to an organisation's strategic objectives in terms of being incremental objectives, stretch objectives and visionary objectives respectively. A systematic approach towards setting up targets can start by specifying the target for each strategic objective, and subsequently to each measure. Such an approach enhances the cause-and-effect linkages of the entire business performance measurement system.

Targets should still be set even if managers are unsure about what they should be. Having an indicative target would still be preferred over a blank target, as it would drive certain business performance. In such situations, it is important to note and communicate appropriately that targets for such measures are merely indicated and subject to future refinement. Good commitment can also be obtained by getting relevant employees to participate in setting the future target.

Targets set must be reviewed on a regular basis to ensure the ongoing validity of the targets, and to ensure that employees are motivated towards achieving them. Targets that are too easily attained or too difficult to achieve will affect motivation of employees. Over time, the use of targets can form a basis for establishing internal benchmarks or best practices for improving internal processes.

The characteristics of a well-chosen target are:

1. easily understood and communicated,
2. clear in establishing expectations, and
3. encouragements given to stretch performance.

It is common to see organisations become too involved and mechanical in setting targets. They fail to realise that targets will shape organisation behaviours. Targets should be established to motivate people to perform in a certain way. Poorly chosen targets can cause confusion, undermine strategy, cause people to pursue the wrong goals, unnecessarily consume resources, and de-motivate employees. As targets have cause-and-effect links to strategic objectives, it would certainly affect the overall strategy of the organisation. At times, interim targets encourage an organisation to perform at a preferred pace. Progress towards a strategic objective is easily measured against multiple steps of the interim targets. Interim targets also offer a sense of progress and motivation.

3.1.6 Initiatives

Initiatives are specific projects, programmes or planned activities directed at key processes for the purpose of enhancing their output performance, and to meet or exceed established targets. Such initiatives are directed at closing the gaps identified during routine business performance measurement. A gap occurs when the measured business performance falls short of the target. Therefore, in order to bring the level of performance up to a desirable level, corrective actions or improvements must be made to the processes. These improvement activities are incorporated as strategic initiatives.

3.2 Success Factors in Implementing and Sustaining Business Performance Measurement Systems

The business performance measurement holds tremendous promise when implemented with passion and commitment. Resistance to change should be replaced with pursuit for improvement and a desire to advance the organisation's vision and mission.

An often overlooked method that can enhance the implementation of a business performance measurement system is advertising and celebrating success. The successful completion of a business performance measurement system implementation stage or the attainment of an organisational objective is an ideal time to recognise it, and use such success as a "role model" for other business units, especially sceptics with negative views. Such effort can pique the interest of staff that have not yet been part of the process. Success often breeds success, and a success story can influence another business unit to follow the footsteps leading to actual success.

For example, an organisation can advertise a successful business performance measurement system implementation of one department, division or business unit in an internal newsletter. Apart from advertising and promoting the initiative, such a newsletter can also educate staff under business units that have not yet been involved and encourage future involvement. For organisations to implement and, more importantly, to sustain the operation of a business performance measurement system, the following success factors are important considerations:

3.2.1 Top Management Support

In order to provide a supportive and conducive environment for moving a strategy forward, the organisation must obtain buy-in for the project from both the top-down and bottom-up levels of the organisation.

Top management buy-in is important to the success of a business performance measurement system implementation. Such buy-in often drives acceptance of the system by employees at all levels of the organisation. Financial support and human resources will

likely be diverted to a project like this only if top management is convinced of its benefits. If the top management is not convinced, how will others in the organisation be convinced? Political battles can be minimised or avoided when the top management endorses and sponsors a business performance management initiative, and publicises its benefits and importance to the organisation.

Bottom-up buy-in is also important because employees who are occupied with their daily work and responsibilities may not place much time and emphasis on a business performance management project, which they regard as non-essential in fulfilling their job responsibilities according to their job descriptions.

However, having good buy-in does not mean that the business performance management initiative will be spared from failure or termination. At times, a change in the top management can mean that the "pet projects" of the outgoing management team be eliminated. To ensure the business performance management initiative does not derail, it has to be integrated into the culture of the organisation. Using the results from business performance management to set the agenda for routine management meetings or as key input to management decision process is essential to ensure survival and sustainability. Generally, management tools are not primary targets for termination when there is a change in management, but projects are, especially when there is a budget cut.

3.2.2 Project Champion

As with any other long running projects that involve the entire organisation, a project champion must be elected to drive the activities associated with business performance measurement. The project champion should be someone from the senior management so as to be able to communicate effectively with all levels of the management.

3.2.3 Adequate Resources

Implementing a business performance measurement system and then sustaining it over a long period of time requires a substantial amount of resources. Other than competent staff to track and collate data, the organisation has to deploy supporting IT systems and

infrastructure to manage the collated data from business performance measurements. These IT systems include automated data collection systems, data repositories, data warehouse, scalable server systems; and data analysis tools, such as business intelligence and analytics software. Compatibility issues must also be ironed out in order to facilitate the smooth transfer of data between systems.

3.2.4 Employee Participation

To reduce resistance to change and obtain proactive participation, buy-in from the lower level management is required. In many situations, the measurement of business performance is undertaken at the source level, which usually involves a business process managed by employees within a particular department of an organisation. The responsibility of taking up business performance measurements at regular periods will have to be left to the employees managing the process. Thus, it is imperative that employees accept this responsibility and undertake the business performance measurement task conscientiously.

3.2.5 Concise Reporting

Measurements collected have to be presented in a concise and efficient manner to managers, so that effective decision making can be done. Good reporting format and practices, like using performance dashboards, must be established.

3.3 Linkage Between Strategy, Strategic Themes, Strategic Objectives, Business Performance Measures, Business Performance Targets, and Strategic Initiatives

For a business performance measurement system to be successfully implemented and useful, its business performance measures must link to the organisation's strategies. Only then will the results from the measures be used to drive the organisation towards greater competitiveness and profitability. Cause-and-effect linkage is one good method to link strategy to performance measures.

Developing a comprehensive business performance measurement system is not cheap, simple or quick for the following reasons:

- Business performance is a multi-dimensional concept, which includes efficiency, effectiveness, quality, equity, fiscal stability, and conformance with government policy and standards.
- Data collection is not cheap, and managers may not be able to capture all aspects of business performance.
- Those who will use business performance information may dictate which business performance dimensions to focus on.
- Although strategic objective plays a vital role in the designing of business performance measurement systems, the multiplicity of strategic objectives may exist and may be inconsistent at times.
- The lack of control of processes and time constraint may make measuring business performance difficult.
- Organisation barriers such as lack of accountability, insufficient analytical skills, lack of analytics staff, perverse reward system, inadequate performance evaluation and requirement of large investment upfront are possible hindrances to the implementation of business performance management systems.
- Personal barriers such as attitudes, traits, behaviours, risk avoidances, conceptual confusions, inadequate control, manager alibis, and perceived threats to job security are common barriers in implementing business performance measurement systems.

Summary

Most organisations today operate through decentralised business units and teams that are much closer to the customer than large corporate management staff. These organisations recognise that competitive advantage comes more from intangible knowledge, capabilities and relationships created by employees than from investments in physical assets and access to capital. Strategy implementation, therefore, requires all business units, support units and employees to be aligned and linked to the strategy. With rapid changes in technologies, competition, and regulations, the formulation and implementation of strategies must become a continual and participative process. The ability to understand, manage and improve business performance is largely dependent on the ability to successfully measure business performance. Business metrics are quantifiable measures used to track business processes to judge the performance level of your business. Generally, individual divisions or departments within a company, such as manufacturing, marketing and sales, are responsible for monitoring the metrics that track the performance of their parts of the business. Senior executives track more general metrics. There are hundreds of these metrics because there are so many different kinds of businesses, with many different processes. Tracking the metrics that are most important to your business —and managing operations based on the results—maximises the business's chances of success.

Balanced scorecard approach allows organisations to execute their strategies and offers many advantages to an organisation. First, it allows organisations to have a balanced view of its performance in all critical areas of the business. Second, it requires managers to first establish strategic objectives before the business performance measures are selected. This ensures relevancy of the measures. Third, the business performance measures channel improvement actions by the employees and the managers towards achieving the strategic objectives. Fourth, with balanced perspectives, selected measures tend to have a mixture of leading and lagging indicators. This ensures that organisations plan for the future, and at the same time understand the past and the current business performance.

The strategy map offers a big picture in which strategic objectives are represented visually and are linked to one another through cause-and-effect relationships. Below is a summary of the process of linking strategies to business performance measures.

1. Define vision and mission: For example, to be the market leader by offering the best possible product to customers.
2. Define strategic themes: For example, financial growth and productive use of financial resources.
3. Define strategic objectives: For example, under the financial growth strategic theme, one of the strategic objectives can be identified as obtaining a 50% increase in net profit in the next three years.
4. Define business performance measures: For example, the percentage increment in net profit per year.
5. Conduct qualifying checks on the defined business performance measures using the questions below:
    a. Does each measure have impact on the strategic objectives, and is it measurable?
    b. Are the data to derive the measures easily available to and accessible by the organisation?
    c. What should the measurement frequency be, and can the organisation update the data based on the preferred measurement frequency?
6. Establish a target for the measures: For example, the target percentage increment in net profit is equal to 20% per year.

A gap occurs when the measured business performance falls short of the target. In order to bring the level of performance up to a desirable level, corrective actions or improvements must be made to the processes. Such improvement activities are incorporated as strategic initiatives.

Major obstacles that may affect the effective implementation of business performance measurement systems are discussed in the last section of this unit. The significance of this unit is to emphasise that corporate strategy, business performance measurement, and

continuous improvements are intertwined; all three are essential for an organisation to stay competitive. The top management buy-in is a critical step in attempting to obtain support and acceptance of business performance management by employees.

Activity 1.1

Prepare your answers for the following questions, as your instructor will ask you to present your answers in front of the class:

1. Why is business performance management an important activity in an organisation?
2. List down five benefits of having business performance management.
3. Explain the differences between the past and the present business performance measurement systems.
4. Describe the four essential perspectives in the balanced score card (BSC) business performance management approach.
5. Describe the key components in the strategic map framework.
6. Explain why it is important that an organisation's broad strategy be translated into strategic objectives.

   You are a senior manager in a major company producing athletic shoes. Your organisation has just established the following two strategies:

   - Build a strong momentum to grow the female consumer market.
   - Drive increased margins through efficient inventory management.
7. Using the BSC approach, suggest appropriate strategic objectives and business performance measures for your company.
8. Explain why choosing the right business performance measure is important.
9. What should one do if a business performance measure is not directly measurable? How to you measure customer experience and network efficiency?

10. Explain what the leading and lagging indicators (measures) are.
11. Why is a business performance target important?
12. What do you think are the important success factors in implementing a sustainable business performance measurement system?

Formative Assessment

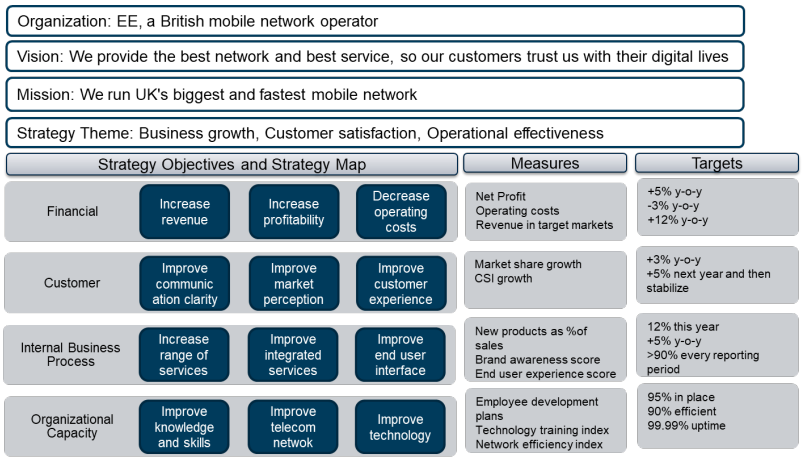
1. Please pick up an organisation of your choice and construct a strategy map.

Solutions or Suggested Answers

Formative Assessment

1. Please pick up an organisation of your choice and construct a strategy map.

I have illustrated the construction of a strategy map by choosing a British mobile network operator, EE.



Click here for the enlarged image

Tableau Demo

1. Download and install *Tableau Desktop* on your laptop as per the instructions on Canvas.
2. Study the quick user guide online on how to connect the Tableau Desktop to Data

According to the Tableau Online Help, to create charts in the Tableau Desktop, we need to first open a workspace page. There are many ways we can open a workspace page. For example, we can drag and drop any data source icon (i.e. excel file icon) onto the Tableau icon. It will open the Tableau worksheet page for the selected data source. We can open as many connections as we want in Tableau by going to the start page or data connection page to select a new connection. Figure 1.8 shows the Tableau worksheet that is connected to a sample superstore sales excel data set used to create scatter plots.

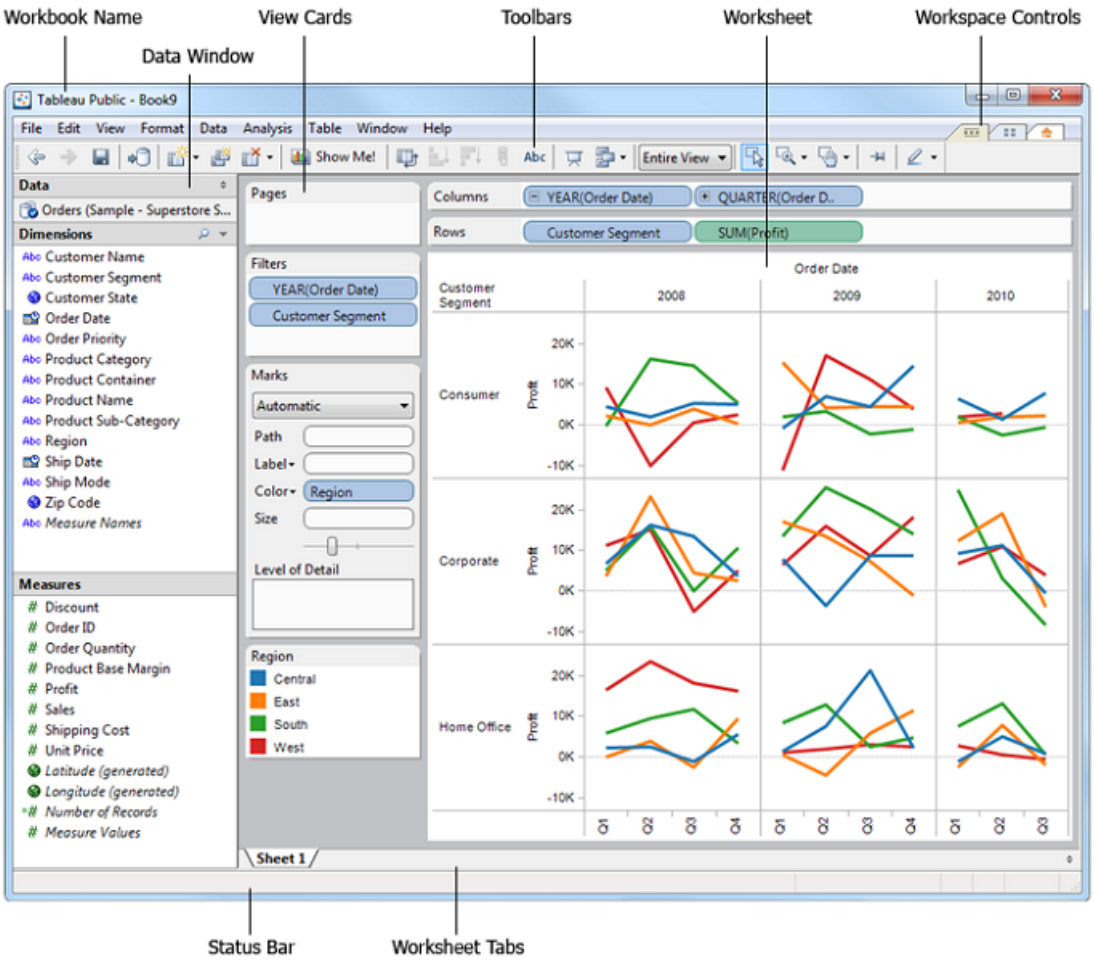


Figure 1.8 Tableau's Workspace (Tableau Online Help, 2014)

Data sources appear on the left side of the workspace in the Data window, as shown in Figure 1.9. We can hide and show the Data window by selecting **View -> Data Window**. Alternatively, we can also click the minimise button in the upper right corner of the Data window.

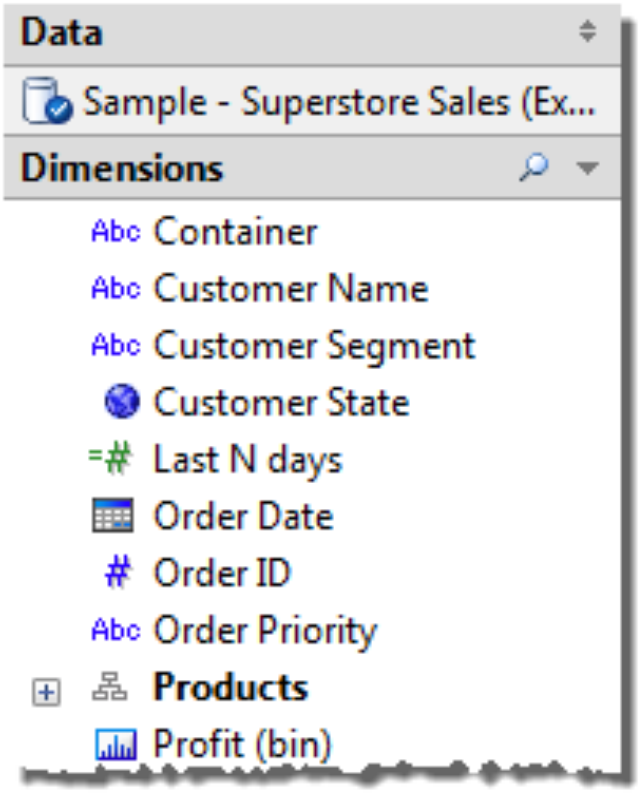


Figure 1.9 Tableau's Data Window (Tableau Online Help, 2014)

We can search the fields in the Data window by clicking the magnifying glass icon and typing in the text box. To access commands related to each field, we can right click the field in the data window.

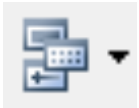
The Tableau's toolbar contains commands such as Connect to Data and Save, as shown in Figure 1.10. In addition, the toolbar contains analysis and navigation tools such as Sort, Group, and Zoom. We can unlock the toolbar by grabbing its left edge and then dragging it to a new location. We can hide or display the toolbar by selecting **View -> Toolbar**.

Figure 1.10 Tableau's Toolbar

Table 1.7 shows the functions of each toolbar button:

Table 1.7 Tableau's Toolbar Buttons

Toolbar Button	Description
	Undo: undoes the last task you completed.
	Redo: repeats the last task you cancelled with the Undo button.
	Save: saves the changes made to the workbook.
	Connect to Data: opens a dialogue box where you can create a new connection or select one from your repository.
	New Sheet: creates a new, blank worksheet.
	Duplicate Sheet: duplicates the existing content onto the new worksheet.
	Clear: clears the current worksheet. Use the drop-down list to clear specific parts of the view such as filters, formatting, and sizing.
	Automatic Updates: controls whether Tableau automatically updates the view when changes are made. Use the drop-down list to automatically update the entire sheet or just quick filters.
	Run Update: runs a manual query of the data to update the view with changes when automatic updates is turned off. Use the drop-down list to update the entire sheet or just quick filters.

Toolbar Button	Description
	Show Me!: displays alternative views of the data, in addition to the best view according to best practices. The options available depend on the selected data fields when you click this button.
	Swap: moves the fields from the Rows shelf to the Columns shelf and vice versa. The Hide Empty Rows and Hide Empty Columns settings are always swapped using this button.
	Sort Ascending: applies a manual sort in ascending order of a selected field based on the measures in the view.
	Sort Descending: applies a manual sort in descending order of a selected field based on the measures in the view.
	Group Members: creates a group by combining selected values.
	Show Mark Labels: toggles between showing and hiding mark labels for the current sheet.
	Presentation Mode: toggles between showing and hiding everything but the view.
	View Cards: shows and hides the specified cards in a worksheet. Select the cards you want to hide or show from the drop-down list.
	Fit Selector: specifies how the view should be sized within the application window. Select either a Normal Fit, Fit Width, Fit Height, or Entire View.

Toolbar Button	Description
	Fix Axes: toggles between locking the axes to a specific range and showing all of the data in the view.
	Highlight: turns on the highlighting function for the selected sheet. Use the options on the drop-down list to define how values will be highlighted.

The Status Bar is located at the bottom of the Tableau Workbook. It displays descriptions of menu items, as well as information about the current view. For example, the status bar in Figure 1.11 shows that the view has 72 marks shown in one row and one column. It also shows that the Sum(Sales) for all the marks is $51.3M.

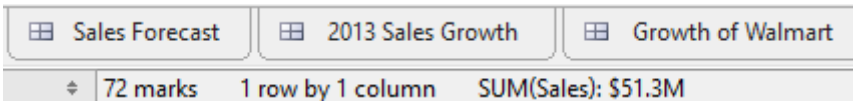


Figure 1.11 Tableau's Status Bar

We can hide the Status Bar by selecting **View -> Status Bar**.

Occasionally, the Tableau will display warning icons at the bottom right corner of the status bar to indicate errors that have or may occur. Table 1.8 below shows the possible warning icons and what they mean.

Table 1.8 Tableau's Status Bar possible warning icons

Warning Icon	Description
	Cancel Query Indicator: when you cancel multiple queries, an indicator will appear to show you how many queries are still running on the database and using the resources. For more information about this warning, refer to Abandoned Queries.

Warning Icon	Description
	Precision Warning: some fields are more precise in the database than Tableau can model. When you add these fields to the view, a precision warning will be displayed at the status bar. For more information about this warning, refer to Precision Warnings. Geocoding Warning: if Tableau cannot geocode some of your location values, this warning will show up. Geocoding warnings may happen if you have unknown location names or names that exist in multiple countries and states.

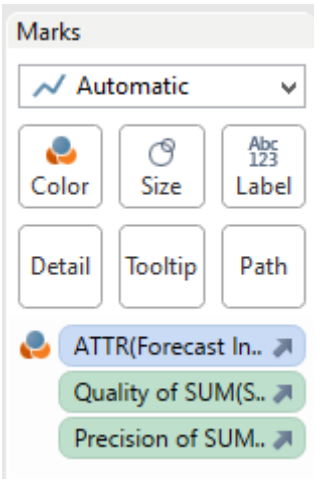


Figure 1.12 Tableau's Mark Card

Mark Cards can be shown or hidden, as well as rearranged around the worksheet. The following list describes each card and its contents.

- Columns Shelf is the shelf where we can drag fields to add columns to the view.
- Rows Shelf is the shelf where we can drag fields to add columns to the view.
- Pages Shelf is the shelf where we can create several different pages, with respect to the members in a dimension or the values in a measure.
- Filters Shelf is the shelf where we can specify the values to include in the view.
- Measure Names/Values Shelf is the shelf where we can use multiple measures along a single axis.

- Colour Legend contains the legend for colour encodings in the view. It is only available when there is a field on the colour shelf.
- Shape Legend contains the legend for shape encodings in the view. It is only available when there is a field on the shape shelf.
- Size Legend contains legend for the size encodings in the view. It is only available when there is a field on the size shelf.
- Map Legend contains the legend for symbols and patterns on a map. The map legend is not available to all map providers.
- Quick Filters is available for every field in the view. Use these cards to easily include and exclude values from the view without having to open the Filter dialogue box.
- Marks contains a mark selector where we can specify the mark type as well as the Path, Shape, Text, Colour, Size, Angle, and Level of Detail shelves. The availability of these shelves are dependent on the fields in the view.
- Title contains the title of the view. Double-click this card to edit the title.
- Caption contains a caption that describes the view. Double-click this card to edit the caption.
- Summary contains the summary of each of the measures in the view, including the Min, Max, Sum, and Average.
- Map Options allows us to modify the various labels and boundaries shown on the online maps. We can also use this card to overlay metro statistical area information.
- Current Page contains playback controls for the pages shelf and indicates the current page that is displayed. This card is only available when there is a field on the pages shelf.

Each card has a menu that contains common controls that apply to the contents of the card. For example, we can use the card menu to show and hide the card. Access the card menu by clicking on the arrow at the upper right corner of the card.

References

Ariyachandra, T. R., & Frolick, M. N. (2008). Critical success factors in business performance management – striving for success. *Information Systems Management*, *25*, 113-120.

Bacal, R. (1999). *Performance management.* New York: McGraw Hill.

Harvard Business Review (1998). *Measuring corporate performance.* Boston: Harvard Business School Press.

Harvard Business School Press (2006). *Performance management.* Boston: Harvard Business School Press.

Hobbs, L. (2008). *Strategic DNA.* Chicago: Agate Publishing.

Kaplan, R.S., & Norton, D.P. (2006). Avoiding the pitfalls. *Harvard Business Review, 2006*, 1-14.

Kaplan, R.S., & Norton, D.P. (2008). Mastering the management system. *Harvard Business Review, January 2008*, 1-7.

Kaplan, R.S. & Norton, D.P. (2005). The balanced scorecard – measures that drive performance. *Harvard Business Review, July-August 2005*, 1-8.

Kaplan, R.S. (1996). *Mobil USM&R: Linking the balanced scorecard.* Boston: Harvard Business School Press.

Lawson, R., Desroches, D., & Hatch, T. (2008). *Scorecard best practices.* West Sussex: John Wiley & Sons.

Moore, M. H. (1995). *Creating public value: Strategic management in government*. Harvard university press.

Neely, A. (Ed.). (2002). *Business performance measurement – theory and practice.* Cambridge: Cambridge University Press.

Neely, A., Bourne, M., Mills, J., Platts, K., & Richards, H. (2002). *Strategy and performance: Getting the measure of your business.* Cambridge: University Press.

Olve, N., Roy, J., & Wetter, M. (1997). *Performance drivers*: *A practical guide to using the balanced scorecard.* West Sussex: John Wiley & Sons.

Spitzer, D.R (2007). *Transforming performance measurement.* New York: AMACOM (American Management Association).

Tyagi, R. K., & Gupta, P. (2008). *A complete and balanced service scorecard.* New Jersey: FT Press.

Wongrassamee, S., Simmons, J. E., & Gardiner, P. D. (2003). Performance measurement tools: the Balanced Scorecard and the EFQM Excellence Model. *Measuring business excellence*.

Study Unit 2

Science and Art of Data Visualisation

Learning Outcomes

At the end of this unit, you are expected to:

- Determine appropriate data visualisation techniques based on given data.

Overview

This unit introduces the concept, benefits and applications of data visualisation. There is a shift on the word visualisation's meaning from being an internal object in our mind to becoming an external object to support the decision-making process. It demonstrates the benefits and applications of data visualisation through various examples and shows the information processing model of human visual perception. Data visualisation is built upon data and the four components, which are visual cues, coordinate systems, scales, and context. These components work together and each of them affects each other.

Chapter 1: Data Visualisation

Lesson Recording

Data Visualisation

Data by definition means facts or information that are used to find out things or to make decisions (Oxford English Dictionary). Until recently, the term *visualisation* meant the act of forming a picture of something or somebody in our mind (Oxford English Dictionary). However, *visualisation* now means something more akin to a graphical representation of data or concepts. There is a shift on the word visualisation's meaning from being an internal object in our mind to becoming an external object to support the decision-making process. Therefore, *data visualisation* means a process to create a visual representation of data.

1.1 Benefits of Data Visualisation

In this section, we will use an example to illustrate the benefits of data visualisation. Figure 2.1 shows a data visualisation of the Passamaquoddy Bay between Maine, in the United States, and New Brunswick, in Canada, where the tides are the highest in the world. Data come from approximately one million measurements and are collected by the University of New Brunswick's Ocean Mapping Group, using EM-1,000 multi-beam sonar system onboard the CHS vessel F.G. Creed.

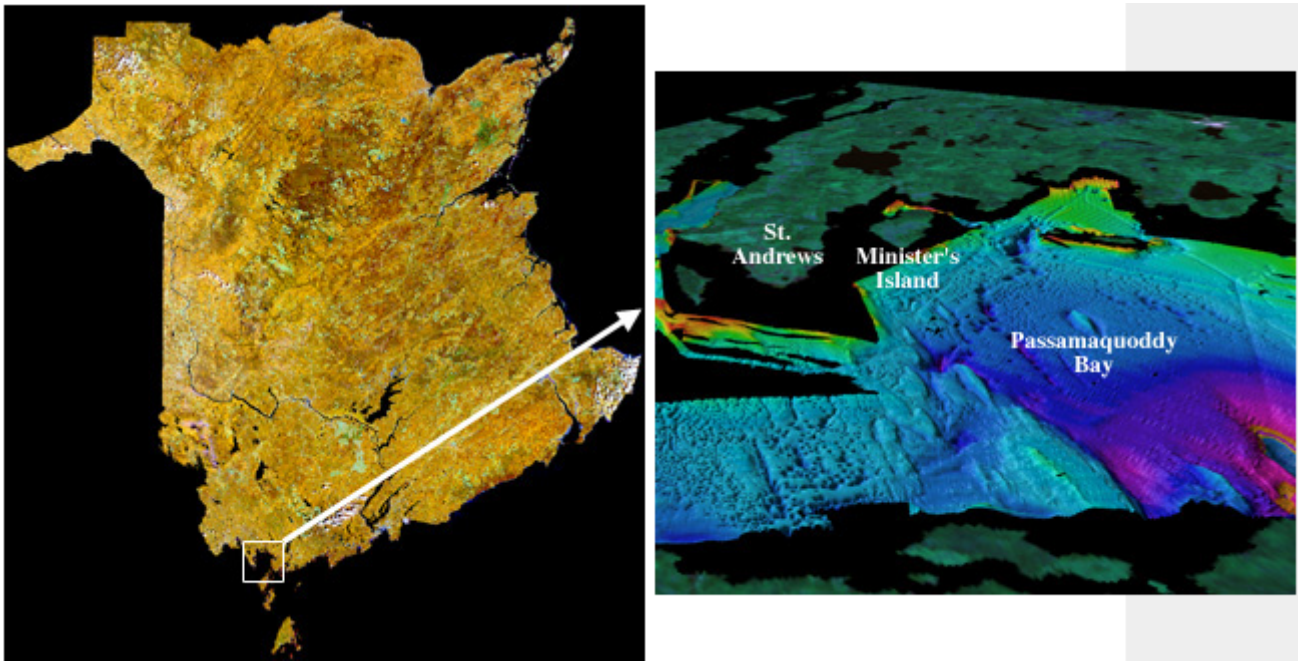


Figure 2.1 3D data visualisation of colour-coded bathymetry data from Passamaquoddy Bay (University of New Brunswick, 2014)

If the above one million measurements data were presented using numbers rather than a 3D colour-coded map, it will be challenging for the research team from the University of New Brunswick to gain insight into the data. This example highlights a number of benefits of data visualisation:

- Data visualisation provides us the "ability" to comprehend huge amounts of data. Thus, using data visualisation, researchers were able to interpret important information from more than a million measurements.
- Data visualisation allows the perception of emergent properties that are not anticipated. The areas that have high tides (i.e. colour-coded in brighter colours) appearing within the same cluster is immediately evident.
- Data visualisation often enable problems with data to become immediately apparent. It is common for data visualisation to reveal things not only about the data, but also about the way the data are collected. With appropriate data visualisation, missing data or errors in the data will surface. Thus, data visualisation can be invaluable in quality control.

- Data visualisation facilitates the understanding of both large- and small-scale features of the data.
- Data visualisation facilitates hypothesis formation. For example, the data visualisation in Figure 2.1 helps the team from the University of New Brunswick in their research.

1.2 Data Visualisation in Everyday Life

Data visualisation has found its way into our everyday life. Almost all online contents are stored in databases and file systems. As people become more comfortable with computers and mobile devices, analysts and data visualisation software developers can create applications with interfaces that display the sheer quantity of data all at once.

1.2.1 Google News and Newsmap

Some of us perhaps like to use Google News to search for news. As shown in Figure 2.2, the Google News application shows a list of headlines, complemented with a thumbnail. Some of the top stories are listed at the top, while the most popular stories (i.e. based on the number of readers) are listed on the right sidebar. Other stories are categorised and shown below the top stories.

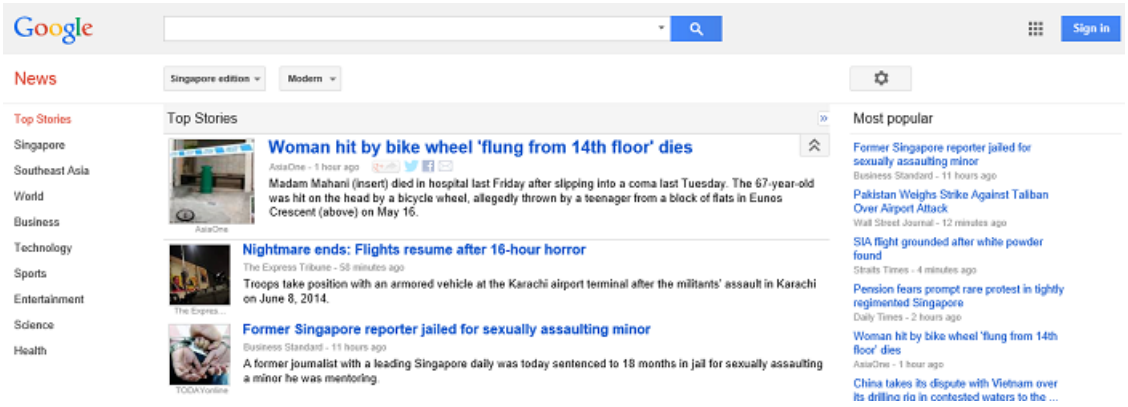


Figure 2.2 Google News (Google News, 2014)

If we were ever overwhelmed by the amount of news on Google News, the Newsmap data visualisation application that was launched in 2004 by design engineer Marcos Weskamp

offers us an option to view the news. The Newsmap application collects data from Google News and displays the headlines sized by popularity (i.e. based on the number of readers), as shown in Figure 2.3.

Figure 2.3 Newsmap data visualisation application (Weskamp , 2014)

Each rectangle represents a "clickable story" and is coloured according to topics, such as world, national or business, so that users can get a sense of what is going on in the world at a glance. In addition, there is a variety of options, such as country of interest and time frame, for topics that users want to include and exclude as their profile preference.

1.2.2 Google Maps

Geographical maps are heavily used as a navigation tool where people can look up directions from point A to point B. The Google Maps application, as shown in Figure 2.4, was created to offer us navigation features with additional layers of information and in the context of areas, such as nearby restaurants and other businesses.

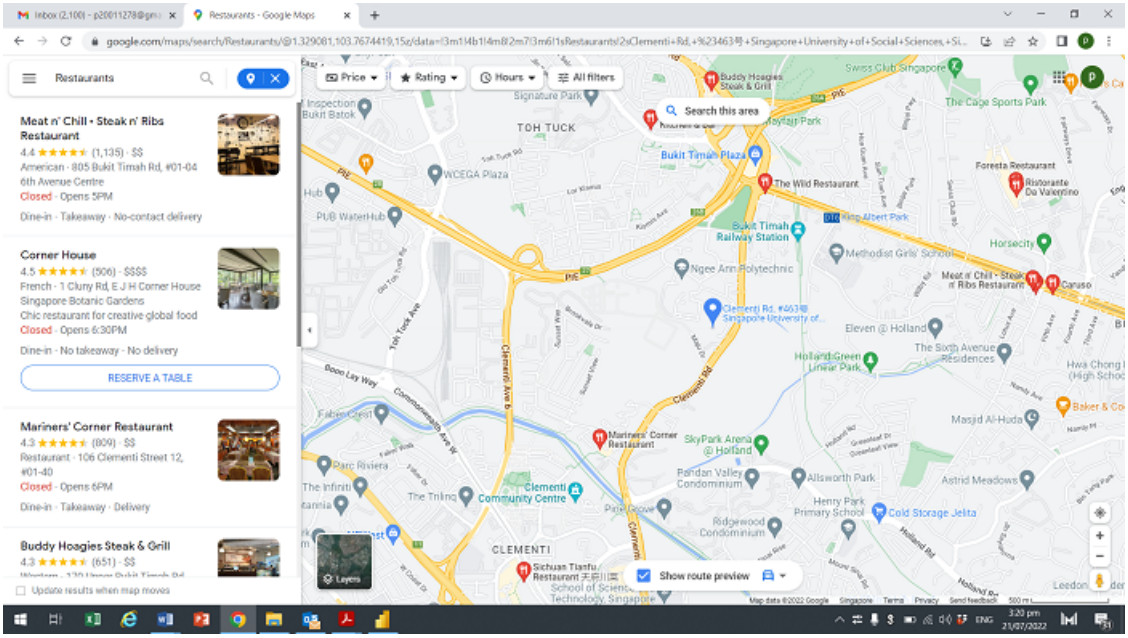


Figure 2.4 Google Maps

1.2.3 Planetary

Some data visualisation may completely change the way we interact and relate to data. The Planetary is an iPad application that places iTunes music library into the context of a solar system, as shown in Figure 2.5.

Figure 2.5 The Planetary

This application uses stars to represent artists. Planets that orbits the stars represent albums, and moons that orbits the planets represent the music tracks. Instead of creating a visualisation of opening a music library and choosing a specific song inside the music library, this application transforms music libraries into a landscape that we can explore and rediscover. Furthermore, as we use this application via the iPad touch interface, the data almost feels tangible to us.

1.3 Perceptual Processing Model

Figure 2.6 shows the information processing model of human visual perception according to Ware (2013).

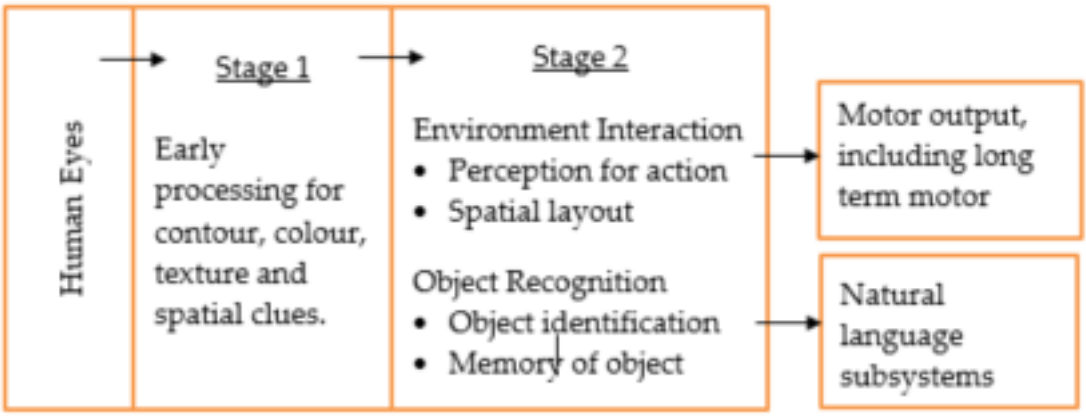


Figure 2.6 Overview of a two-stage model of human visual information processing (Ware, 2013)

In the first stage, information is processed in parallel to extract basic features of the environment. In the second stage, visual attention plays a much more active role, and items in the environment tend to be examined sequentially.

Stage 1: Parallel processing to extract low-level properties of the visual scene

Visual information is first processed by large arrays of neurons in the eye, and in the primary visual cortex at the back of the brain. Individual neuron is selectively tuned to certain kinds of information, such as the orientation of edges or the colour of a patch of light. In each subarea, large arrays of neurons work in parallel, extracting particular features of the environment. In the early stages, this rapid parallel processing proceeds whether we like it or not. It is largely independent of what we choose to attend to, not where we look at. If we want people to understand information quickly, we should present it in such a way that it can easily be detected by these large, fast computational systems of the brain.

Stage 2: Sequential goal-directed processing

The second stage is divided into two separate parts – a subsystem specialised in object recognition, and a subsystem specialised in interacting with the environment. To identify an object, people must match the object's visual characteristics with properties of the object stored in their memories. In addition, the task that the observer is performing will also

influence what is perceived. One of the primary mechanisms relating what is perceived to the task is visual attention. Some aspects of the second stage processing occur sequentially where one visual object is processed at one time.

There is increasing evidence that tasks involving eye-hand coordination and locomotion may be processed in pathways distinct from those involved in object recognition. Milner and Goodale (1995) suggested two-visual-system hypotheses: one system for locomotion and action, another for symbolic object manipulation.

A number of theorists suggests that there is an intermediate representation of the world at the boundary between Stage 1 and Stage 2 processing. Marr (1982) called this the $2\frac{1}{2}$ sketch. Triesman (1980) called it a *feature map*. But much is still unknown about how visual objects are constructed from features, memories and the locus of visual attention.

Beyond the visual processing stages shown in Figure 2.6 are interfaces of other subsystems. A visual object recognition process interfaces with the verbal linguistic subsystems of the brain, so that words can be connected to images. The environment interaction subsystem interfaces with the motor systems that control muscle movements.

Chapter 2: Four components of Data visualisation

Lesson Recording

Four Components of Data Visualisation

According to Yau (2013), data visualisation is built upon data and the four components, which are visual cues, coordinate systems, scales, and context (refer to Figure 2.7). Sometimes, these four components are explicitly displayed, but other times they may form an invisible framework. These components work together and each of them affects each other.

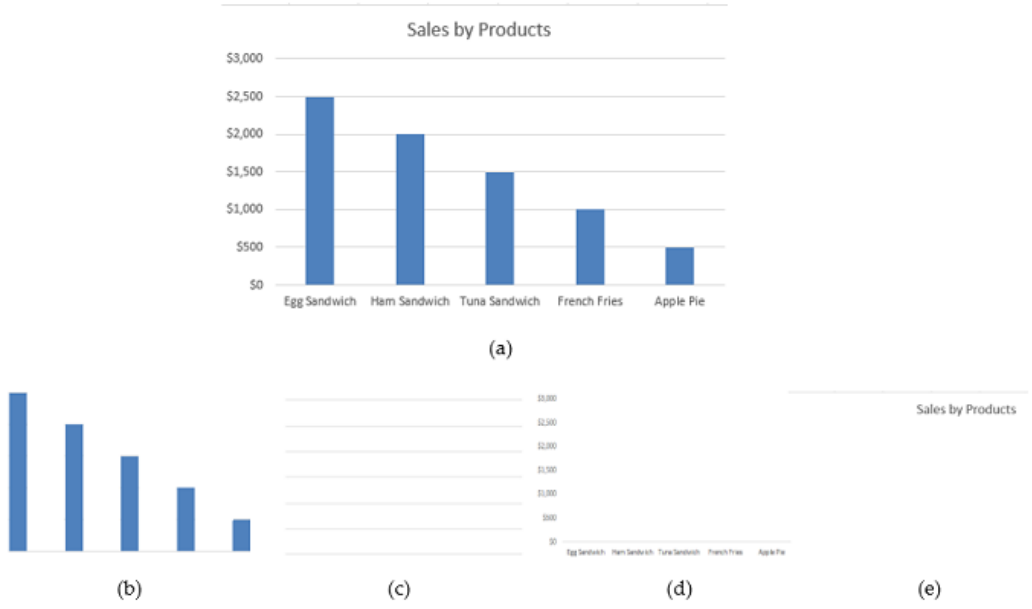


Figure 2.7 (a)Data visualisation of sales by products; (b)Visual cue of the visualisation; (c) Coordinate system of the visualisation; (d)Scale of the visualisation; and (e) Context of the visualisation

2.1 Visual Cues

In its most basic form, data visualisation is simply mapping data onto geometry and colour. It works because the human brain is wired to find patterns and it can switch back and forth between the visual and the number it represents.

We must make sure that the essence of data is not lost in the back and forth between visual and the number it represents. If we cannot map back to the data, the data visualisation is just a bunch of shapes. Which visual cues we choose depends on data and business objectives. We must choose the right visual cue, which changes by purpose, and we must use it correctly. That depends on how we perceive the varied types of visual cues, like position, length, angle, etc.

2.1.1 Position

Position defines where in the space or coordinate system the data is. When we use position as the visual cue, we compare a value based on where the other values are placed in a given space or coordinate system. For example, in Figure 2.8, we use a scatter chart to compare a data point based on its x- and y-coordinates relative to other data points.

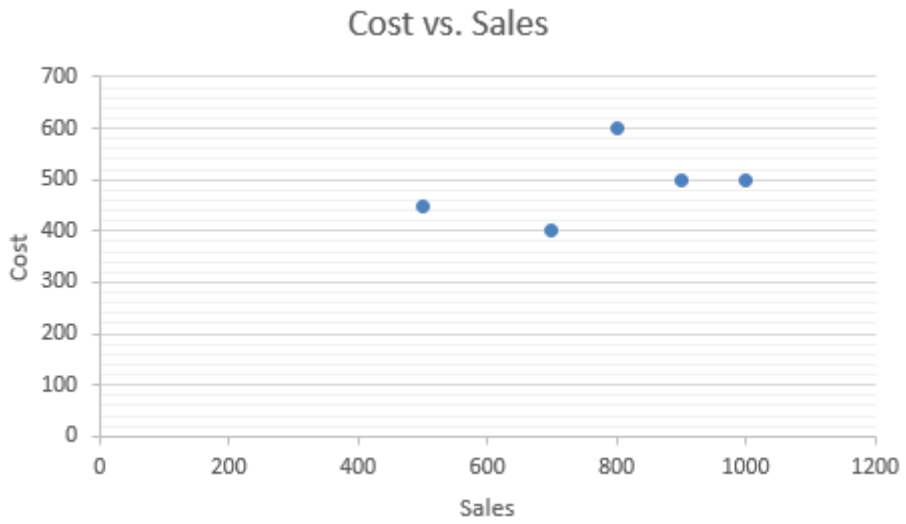


Figure 2.8 Cost versus sales shown using a scatter chart, where position is used as the visual cue

The above scatter chart uses position as a visual cue so that we can compare the cost of the product when the sales equals to $500, and the cost of product when the sales equals to other values, like $700, $800, etc. Position visual cues tend to take up less space than other types of visual cues, because we can draw all the data points within the x and y planes and represent each data point using a symbol like a dot, cross, square or diamond.

Unlike other types of visual cue that take up space to compare values, all data points in position-based data visualisation take up the same amount of space (i.e. symbol size). By using the position visual cue, we can spot trends, clusters and outliers by plotting a lot of data points within the x and y planes. However, it will be challenging to identify what each data point represents. Even in the interactive scatter plot, we still need to mouse over or select a data point to get more information. The overlapping of two or more data points can cause more challenges in identifying what the data points represent.

2.1.2 Length

Length is the most commonly used visual cue to compare data values in several chart types, like bar chart, bullet chart, combo chart, etc. The longer a bar or a bullet point, the greater the absolute value. It can work in both horizontal and vertical directions.

To judge the length visually, we measure the distance from one end of a shape to the other end. In this case, to compare data values based on lengths, we must see both ends of the bars or bullet points. Otherwise, we may end up having a skewed view of maximums, minimums, and everything in between. For example, as shown in Figure 2.9, we use two bar charts with length visual cues to display the same sales amount of two products of a company.

Figure 2.9 Sales amount of products A and B shown using a length visual cue

If both bar charts are shown quickly, the difference in sales amounts between Product A and Product B in Chart (b) will be perceived to be greater than that of Chart (a). This is because on Chart (a), the axis starts at zero and its major unit is labelled every $700, whereas in Chart (b), the axis starts at $900 and its major unit is labelled every $100.

2.1.3 Angle

Angle, as a visual cue, ranges from zero to 360° on a circle. A straight line is 180°, an obtuse angle is an angle between 90° and 180°, and an acute angle is an angle of less than 90°.

For each angle in between zero and 360°, there is an implied opposite angle that completes the rotation. The opposite angle is called the conjugate of an angle. This is why angles are commonly used to represent parts of a whole, as shown in Figure 2.10.

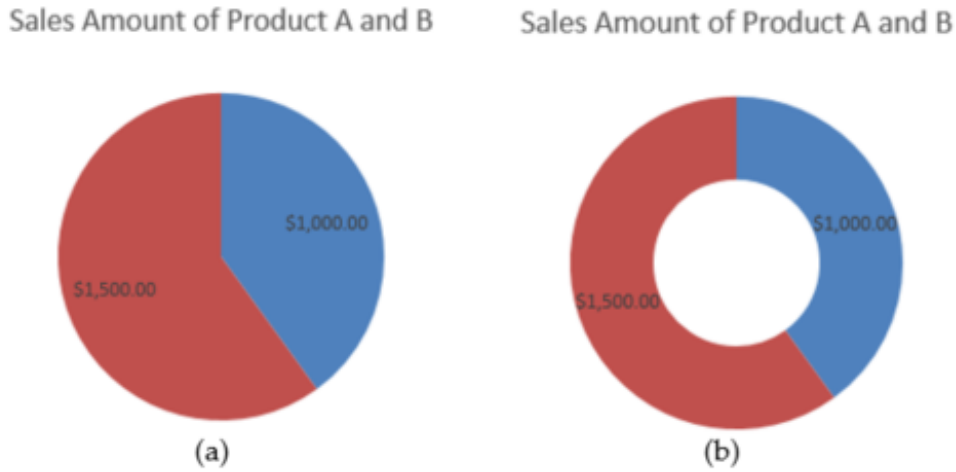


Figure 2.10 Sales amount of products A and B shown using the angle visual cue

In Figure 2.10, Chart (a) is a pie chart that shows the proportion of sales amounts of Product A and Product B, where the relative degrees of $1,000.00 (i.e. Product A's sales amount) and $1,500.00 (i.e. Product B's sales amount) in the circle is the angle visual cue. Chart (b) is a doughnut chart that shows the proportion of sales amounts of Product A and Product B. The length of the arc is the angle visual cue because the centre of the circle, which indicates angles, is removed.

2.1.4 Direction

Direction as a visual cue is similar to angle. Instead of relying on two vectors joined at a point, direction relies on a single vector's orientation in a coordinate system to see which way is up, down, left, right, and everything in between. Direction helps us determine the slope to see increases, decreases and fluctuations, as shown in Figure 2.11.

Figure 2.11 Sales trend of Product A shown using line chart, where direction is used as the visual cue

The amount of perceived changes depends a lot on the scale, as shown in Figure 2.12. For example, we can make a small change in percentage look like a lot by stretching out the scale. Likewise, we can make a big change look like a little by compressing the scale.

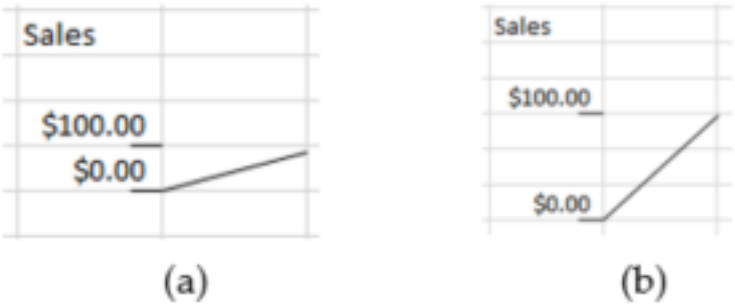


Figure 2.12 The same amount of change shown using different scales

The rule of thumb is to scale the visualisation so that the direction fluctuates mostly around 45 degrees, but this rule is not cast in stone. We can start with this rule and adjust accordingly based on the context. If a small change is significant, then it may be appropriate to stretch the scale so that we can see the shift. In contrast, if a small change is insignificant, we should not stretch out the scale, so that the change will not look major.

2.1.5 Shape

A *shape* or symbol is commonly used as a visual cue to differentiate categories and objects. When we use shapes on a map, they can represent things in the real world. For example, we may use trees as a symbol to represent forests, or use houses as a symbol to represent residential areas. In a chart, we may use triangles and squares in a scatter chart to differentiate different categories of data, as shown in Figure 2.13. Nevertheless, various shapes can provide a context that data points alone cannot.

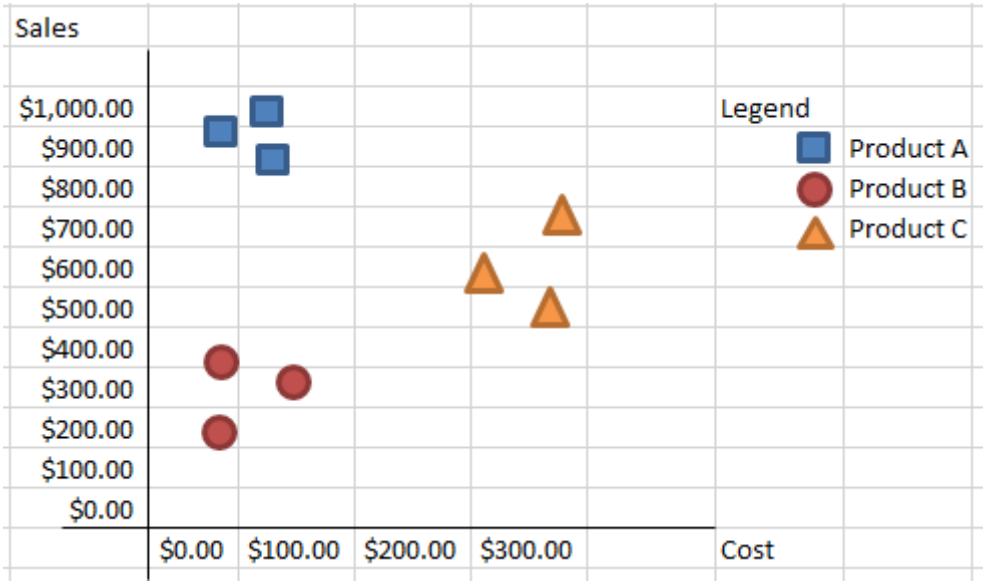


Figure 2.13 Different shapes representing different products on a scatter chart

2.1.6 Area and Volume

When we use *area* and *volume* as visual cues, bigger objects represent greater values. For instance, length, area, and volume can be used to represent data with sizes (i.e. amount of space) in either two or three dimensional space. When we use circles (i.e. two dimensional spaces) or spheres (i.e. three dimensional space) as shape visual cues, we can also use different sizes of circles or spheres to represent different data values.

The most common mistake is to scope a two or three dimensional object using one of the dimensions, such as height, without keeping the proportion of all other dimensions. This

will result in shapes that are either too big or too small, which makes it impossible to fairly compare the data values.

For example, when we use a square shape with two dimensions (i.e. width and height) as a visual cue to represent our data, and we apply the area visual cue, we know that the greater the value, the greater the area of a square. If one value is 50% greater than another, we want the area of the square to be 50% greater than the other. However, if we increase the width and height of the square by 50%, instead of keeping the proportion of width and height, the larger square area will be increased by 125%, as shown in Figure 2.14.

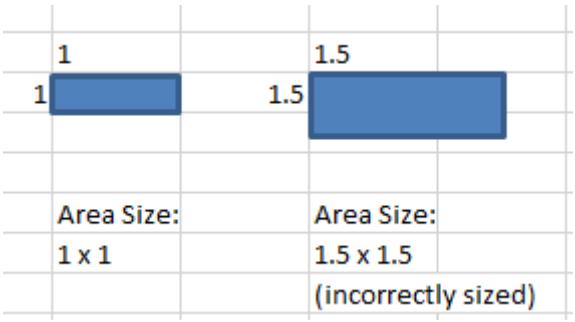


Figure 2.14 Squares sized by different dimension values

We will run into the same problem with three dimensional objects, but the mistake will be more pronounced. For example, when we increase the width, height and depth of a cube by 50%, the volume of the larger cube will be increased by approximately 238%.

2.1.7 Colour

Colour, as a visual cue, helps us break camouflage because things differ visually from their surroundings only by their colours. For example, if we have colour vision, we can easily see that there is a black bear in Figure 2.15. If we do not have colour vision, it will be much harder for us to spot the black bear. Clearly, this can be a life-or-death decision for the hunters or for a person who walks in the jungle.

Figure 2.15 The importance of having colour as a visual cue (Sears, 2014)

The role that colour plays ecologically suggests ways that it can be used in data visualisation. Colour is good for labelling and categorisation, but poor for displaying shape, detail or space.

2.2 Coordinate Systems

According to Yau (2013), when we create data visualisation, we need to place the objects somewhere. The *coordinate system* is a data visualisation component that determines the position of the objects (e.g. data points, shapes or symbols). There are three most basic coordinate systems: *cartesian*, *polar* and *geographic* coordinate systems.

2.2.1 Cartesian

The *cartesian coordinate system*, as shown in Figure 2.16, is a coordinate system that specifies each data point on a plane by a pair of numerical coordinates. The numerical coordinates are the signed distances from the data point to the two fixed perpendicular reference lines, which are measured in the same unit of length. These reference lines are called x- axis and y-axis. Both axes meet at a point, called the *origin*, which is usually represented by ordered pair (0, 0). The numerical coordinates can also be expressed as a signed distance from the origin.

We can also use the same concept to define the position of any data point in n-dimensional space by n cartesian coordinates. The coordinates are the signed distance from the data

point to the *n* fixed perpendicular reference lines which are measured in the same unit of length.

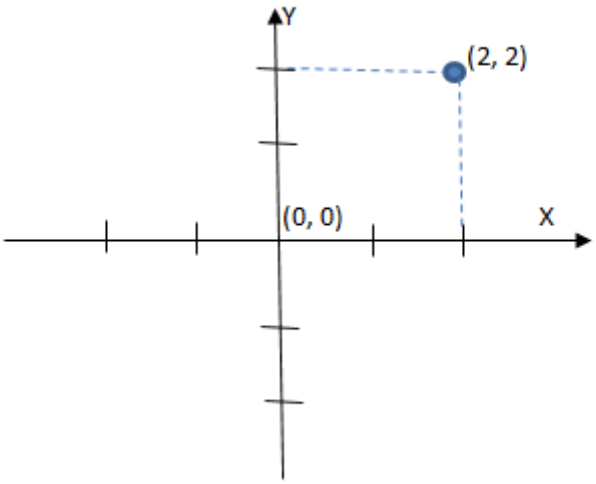


Figure 2.16 Cartesian coordinate system

2.2.2 Polar

The *polar coordinate system*, as shown in Figure 2.17, is a two-dimensional coordinate system. Each data point is determined by the distance between a fixed point and an angle from a fixed direction. The fixed point, which is analogous to the origin in the cartesian coordinate system, is called the *pole*. The ray or half-line from the pole in the fixed direction is called the *polar axis*. The distance from the pole is called the *radial coordinate* or *radius*, and the angle is called the *angular coordinate*, *polar angle* or *azimuth*. This coordinate system is less used than the cartesian coordinate system, but it is useful in cases where the angle or direction is important.

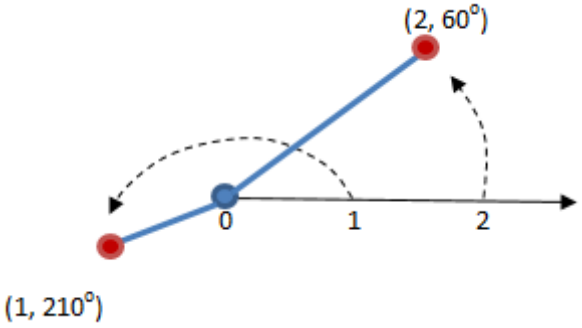


Figure 2.17 Polar coordinate system

Polar coordinates are often used in navigation. An aircraft uses a slightly modified version. Polar coordinates make calculation easier for physicists and engineers— they are used in representing electrical fields, magnetic fields and temperature fields.

2.2.3 Geographic

The *geographic coordinate system*, as shown in Figure 2.18, is a coordinate system that enables every location on the earth to be specified by a set of numbers or letters. To represent the location, the coordinate system commonly uses latitude and longitude —- angles relative to the equator and the prime meridian, respectively. Sometime the coordinate system may also use elevation.

Latitude lines run east and west, which indicates north and south positions on the globe. *Longitude lines* run north and south, which indicates east and west positions. *Elevation* can be thought of as a third dimension. Analogous to the cartesian coordinate system, the latitude is like the horizontal axis, and the longitude is like the vertical axis.

The surface of the earth is wrapped around a spherical mass, but we usually want to display a location on earth on a two-dimensional surface, like on a piece of paper or a computer screen. Therefore, there are a variety of ways to map the surface of the Earth on a two-dimensional surface, which are called *projections*.

Figure 2.18 Geographic coordinate system (Krygier, 2014)

2.3 Scales

Scales is a data visualisation component that defines where the shapes are placed and how they are shaded. There are several types of scales as listed below.

2.3.1 Linear Scale

In a *linear scale*, the visual spacing between each of the data points is the same regardless of where the data points are on the axis.

For example, a customer relationship manager of an apparel company would like to visualise customer satisfaction with the company's products. He can use the linear Scale (i.e. 0 representing very disappointed and 5 representing very satisfied), as illustrated in Figure 2.19, to visualise the data. If we were to measure the distance between two data points on the lower end of the linear scale, it would be the same as the distance between two data points on the higher end of the linear scale.

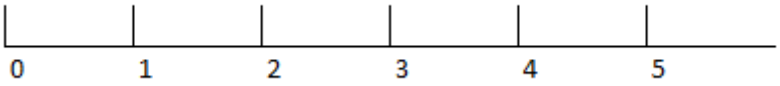


Figure 2.19 Linear scale

2.3.2 Logarithmic Scale

The *logarithmic scale* condenses the distance between each of the data points when the value of the data points increase. This scale is useful if we are interested to visualise the percentage difference or the rate of change amongst each of the data points, especially when the data has a wide range. For example, the sales amount of a food and beverages company like Burger King in year 2000 was S$ 1,000 and doubled each year. The following charts show the sales amount on both linear and logarithmic scales:

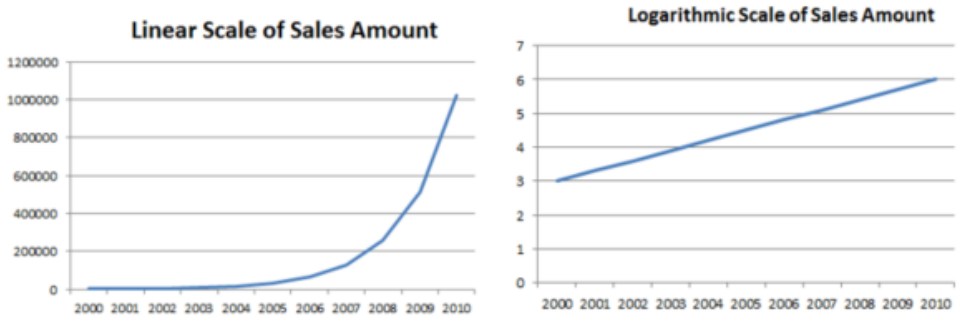


Figure 2.20 Linear scale versus logarithmic scale

The linear scale shows the absolute number of sales over time, while the logarithmic scale shows the rate of change of the sales amount over time. By visualising the data using the logarithmic scale, it is much clearer that the rate of change is constant. Although the logarithmic scale is extremely useful, it is not easily understood by all. Thus, the data visualisation designer must know their target audience.

2.3.3 Percent Scale

Percent scale is usually linear, but when it is used to represent part of the whole data, its maximum is 100%. In addition, the sum of percentages should not exceed 100%.

For example, using the food and beverages company scenario, the company sales in year 2000 was S$ 1,000 in the Singapore market, S$ 500 in the Malaysia market, and S$ 500 in the Thailand market. The following chart show the sales amount in a percent scale:

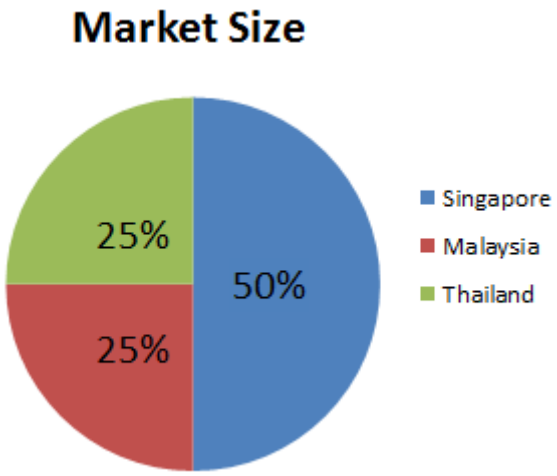


Figure 2.21 Percent scale

2.3.4 Categorical Scale and Ordinal Scale

We use *categorical scales* when we want to provide visual separation of categorical data, such as country of residence or gender. The categorical scale often works with a linear scale or a logarithmic scale.

For example, a sales manager of an apparel company would like to visualise product sales in year 2000. The sales manager can use a bar chart as shown in Figure 2.22, where a categorical scale is used on the horizontal axis, and a linear scale is used on the vertical axis to show the sales amount for different years. The spacing between each category (i.e. each bar) is arbitrary because it does not depend on the numeric value; it is usually adjusted to provide clarity.

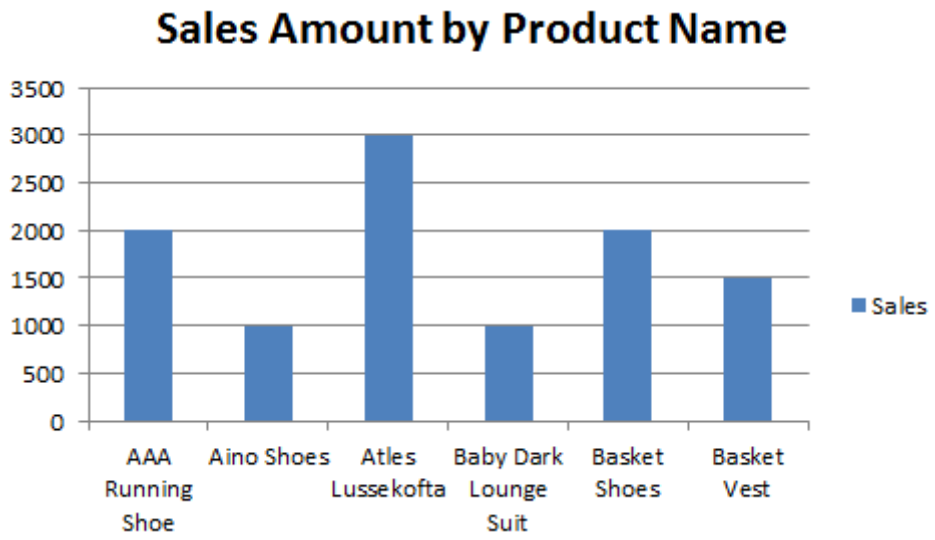


Figure 2.22 Categorical scale works with linear scale

When the order of the categorical data matters, we should order the categorical scale in the context of the data from year 2000 to the latest year, which is year 2010 in Figure 2.23. This type of scale is known as an *ordinal scale*.

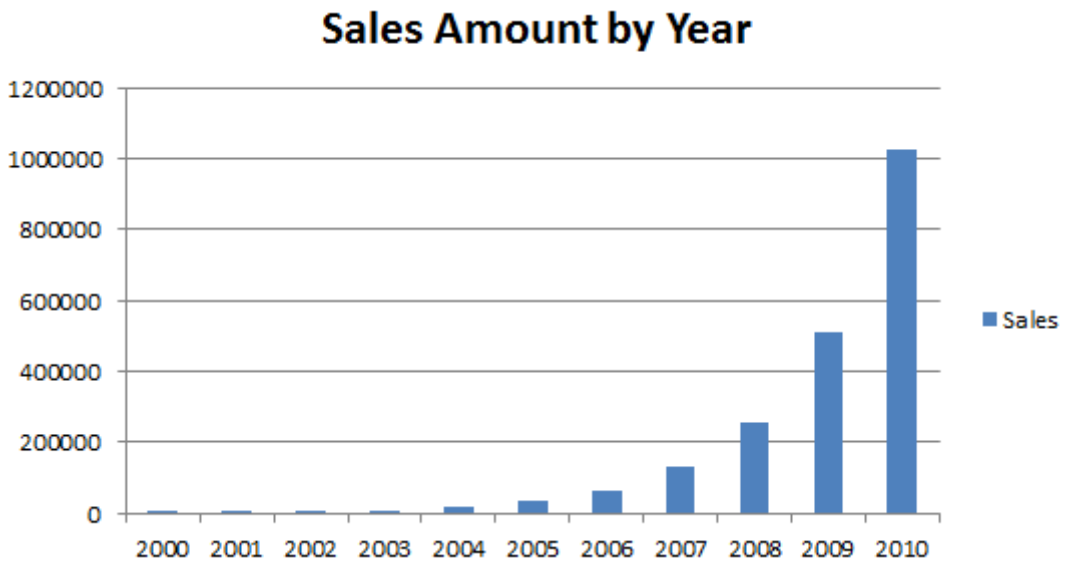


Figure 2.23 Ordinal scale works with linear scale

2.3.5 Time Scale

We use the *time scale* when we want to plot temporal data on a linear scale, or to divide the temporal data into a categorical scale, such as by year, month or day. Figure 2.23 is an example of using a time scale (i.e. year) in a bar chart. When communicating data to the audience, the time scale gives us the advantage of lending the reader connection, because time is a part of everyday life.

2.4 Context

Context is a data visualisation component that lends to a better understanding of the who, what, when, where and why of the data. Context can make the data clearer for interpretations.

When we would like to enable data visualisation viewers to see the data visualisation object of primary interest in full detail, and at the same time get an overview within the context (i.e. surrounding information) available, this is what we call a *focus-context problem* in data visualisation. Card et al. (1999) explained that a focus-context problem starts from three premises: First, the viewer needs both context (i.e. overview of the information), and focus (i.e. detail of the information) simultaneously. Second, the information needed in the overview may be different from that needed in detail. Third, these two types of information need to be combined within a single interactive data visualisation.

A focus-context data visualisation therefore allows viewers to have the information of interest in the foreground and all the remaining information in the background which are simultaneously visible (i.e. seeing the trees without missing the forest). The focus-context problem is not always a spatial related problem. There are also structural and temporal related problems.

Spatial related problems are common to all data visualisation that use maps. For example, a market researcher may wish to understand the behaviour of the individual customer of a shop in Ang Mo Kio. This information is understood in the context of the distance between the shop and the other shops, and the distance between the shop and the customers. Structural related problem arises when we try to visualise data that have structural

components at many levels. Suppose a company sells their products in the U.S. and internationally. We wish to visualise the sales performance of their stores. We may need to understand the company's structure from country, state or province level to city or town level of details.

Temporal related problem involves understanding the timing of data at very different scales. For example, to visualise the performance of Apple shares in the U.S. share market, it can be useful to know the overall performance of the Apple share on a daily basis. However, it can also be useful to analyse the share performance on hourly, weekly, monthly or yearly basis.

We are fortunate that the human brain is able to integrate detailed information from successive fixation of the fovea, with the less detailed information that is available at the periphery of vision. This is combined with the information that comes from the prior sequence of fixations. For each new fixation, our brain will somehow match the key objects in the previous view with those same objects that are moved to new locations. Differing level of details is supported in normal perception because objects are seen at much lower resolution at the periphery of vision than in the fovea. In addition, we have no difficulty in recognising objects at different distances, which means that scale-invariance operations are supported in normal perception. Therefore, solutions to the focus context problem are found by taking advantage of these perceptual capabilities.

There are four different visualisation techniques to solve the focus context problem: distortion, rapid zooming, elision, and multiple windows.

2.4.1 Distortion

The *distortion* technique spatially distorts a data presentation to give more room to the designated points of interest, and to decrease the space given to regions away from those points. What is of specific interest is spatially expanded at the expense of what is not, thus providing both focus and context. Figure 2.24 shows one of such methods called *hyperbolic tree browser* (Lamping et al., 1995), where parts of the graph are dynamically repositioned

and resized based on selected points of interest. The selected node is expanded to show its content.

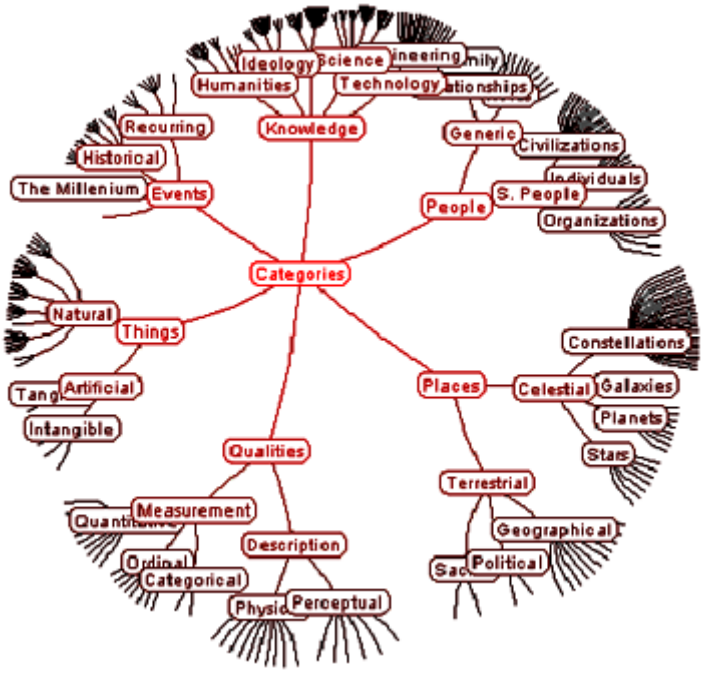


Figure 2.24 The hyperbolic tree browser (Heer, 2004)

2.4.2 Rapid Zooming

The *rapid zooming* technique allows viewers to zoom rapidly in and out of points of interest. This technique provides a large information landscape, even though only a part of them is visible in the viewing window at any one time. This means that the focus and context are not simultaneously available, but viewers can move rapidly and smoothly from focus to context and vice versa. Figure 2.25 illustrates the Pad and Pad++ systems (Bederson & Hollan, 1994), which are based on this principle.

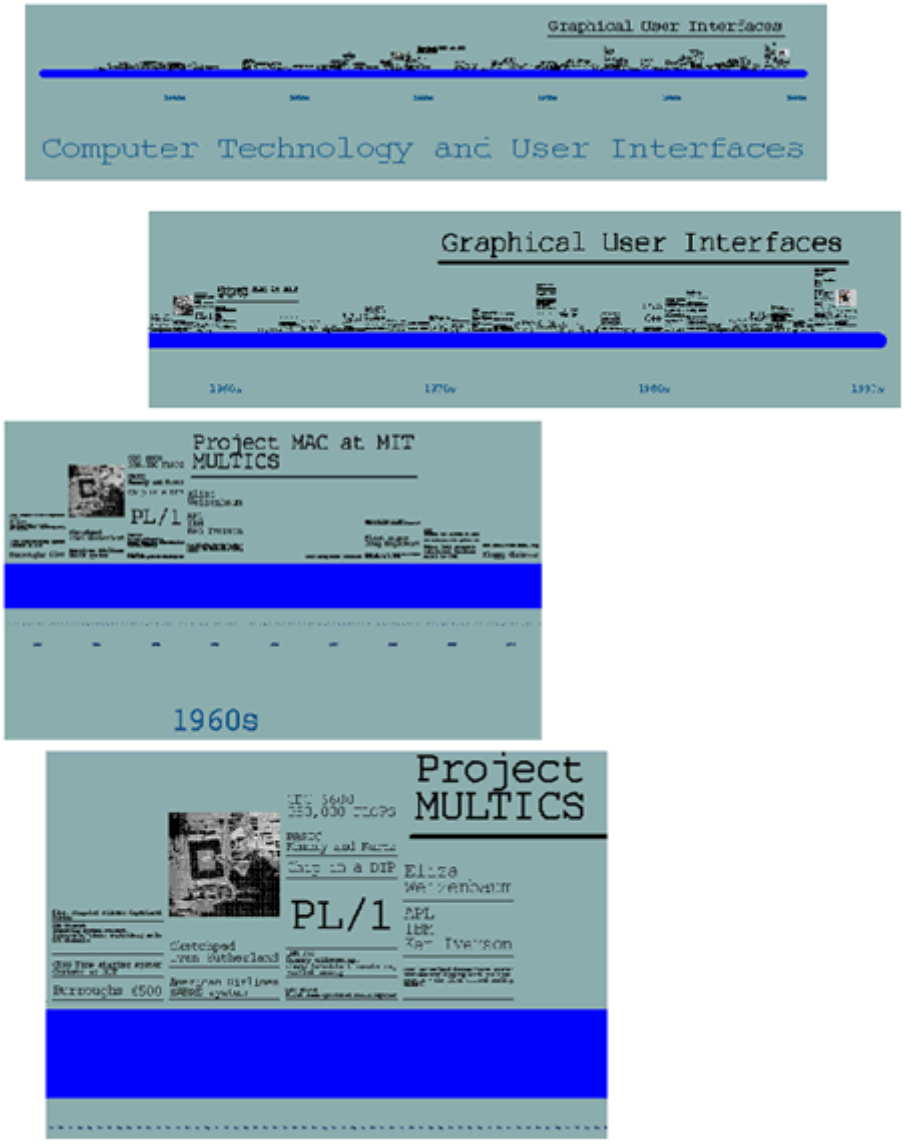


Figure 2.25 Sequence views of Pad++ system (The University of New Mexico, 1996)

2.4.3 Elision

The *elision* technique hides parts of a structure from viewers until they are needed. Typically, this can be achieved by collapsing a large graphical structure into a single graphical object. This technique can be applied to text as well as to graphics. Figure 2.26 illustrates the generalised, fish-eye technique of viewing data (Furnas, 1986), where lesser and lesser detail is shown as the distance from the focus of interest increases.

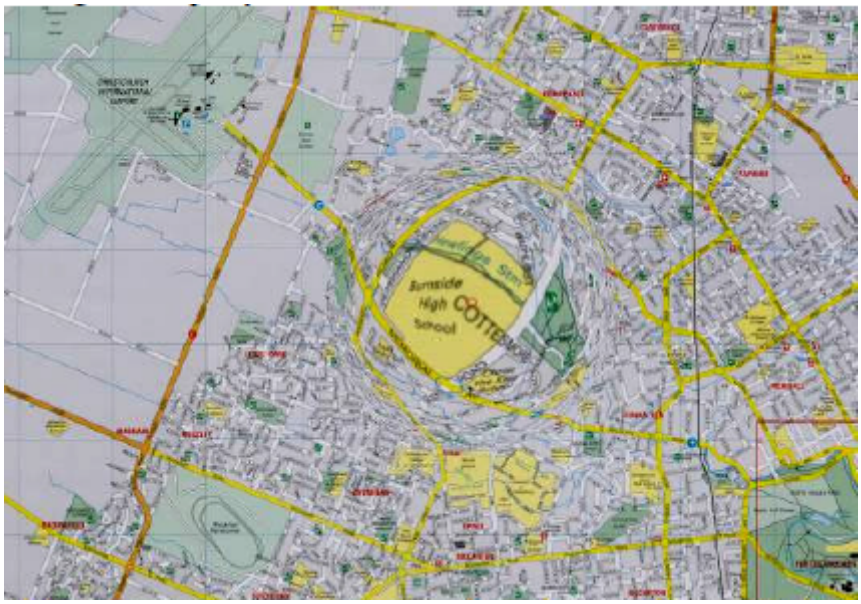


Figure 2.26 The generalised fish- eye technique for viewing data (Cockburn et al. , 2007)

2.4.4 Multiple Windows

The *multiple windows* technique allows viewers to have one window that shows an overview of the data, and several other windows that show the expanded details. The major perceptual problem with the multiple windows technique is that detailed information in one window is disconnected from the overview (i.e. the context) as shown in another window. A solution for this problem is to use lines or transparent overlay to connect the boundaries of the zoomed window to the source window.

Figure 2.27 illustrates the spiral calendar (Card, et al., 1999), where information in one calendar (i.e. zoom window or focus) is linked to the source window (i.e. the context) by a connecting transparent overlay.

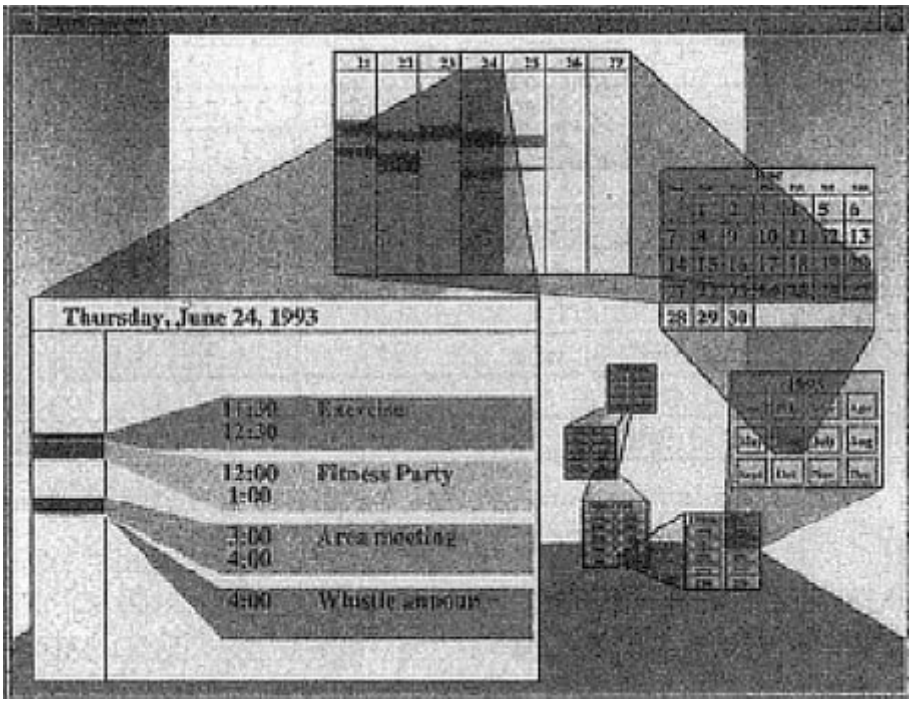


Figure 2.27 The spiral calendar (Zhai et al. , 1996)

The advantage of the multiple windows technique over the others listed previously is that this technique will not distort the object. It is able to show the focus and the context simultaneously.

Summary

Data visualisation applies visual research to practical problems of data analysis in much the same way as engineering applies physics to practical problems of building manufacturing plants. Just as engineering has influenced physicists to become more concerned in areas such as semiconductor technology, we may hope that the development of an applied discipline of data visualisation can encourage visual researchers to intensify their efforts in addressing problems, such as task-oriented perception and user interfaces in 3D space.

Visual cues, coordinate systems, scales and context are the components of data visualisation. A visual cue is the main component that a data visualisation viewer will see, while coordinate systems and scales provide structure and a sense of space to the data visualisation. Last but not least, contexts make data visualisation understandable and relevant. For example, when we use length as the visual cue, the cartesian coordinate system and the categorical scale on the horizontal axis (with the linear scale on the vertical axis in the context of understanding a company's sales amount by financial year) will result in a bar chart of sales amount by financial year.

In summary, we need to know how these components of data visualisation work. Once we have completed our design, we can try to ask other people to look at our design and observe whether they can understand what they see.

Activity 2.1

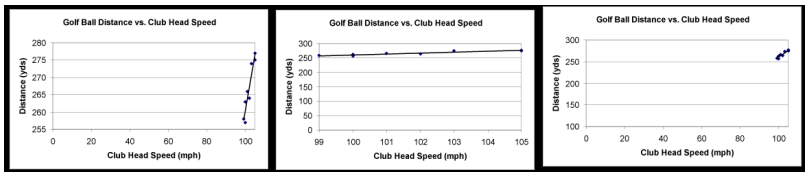
Prepare your answers to the following questions, as your instructor will ask you to present your answers in front of the class:

1. Explain the definition of data visualisation.
2. Describe the benefits of data visualisation.
3. Illustrate how data visualisation can be used in our everyday lives.

4. Illustrate the concepts of the perceptual processing model.
5. Describe the four components of data visualisation.
6. Compare the differences between the seven components of visual cues.
7. Compare the differences between the cartesian, polar, and geographic coordinate systems.
8. Compare the differences between the linear, logarithmic, categorical, ordinal, percent, and time scales.
9. Explain what the focus-context problem is.
10. Illustrate the applications of visual cues, coordinate systems, scales and context in visualising business related data.

Formative Assessment

1. Describe why the plots/charts are bad visualisation examples and suggest ways to improve the plots/charts.

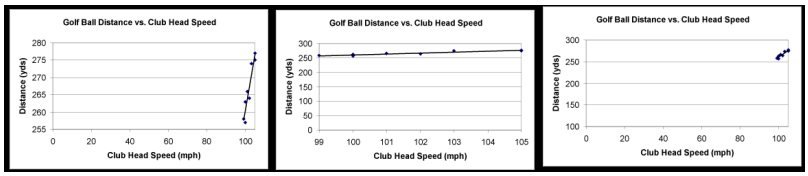


Click here for the enlarged image

Solutions or Suggested Answers

Formative Assessment

1. Describe why the plots/charts are bad visualisation examples and suggest ways to improve the plots/charts.



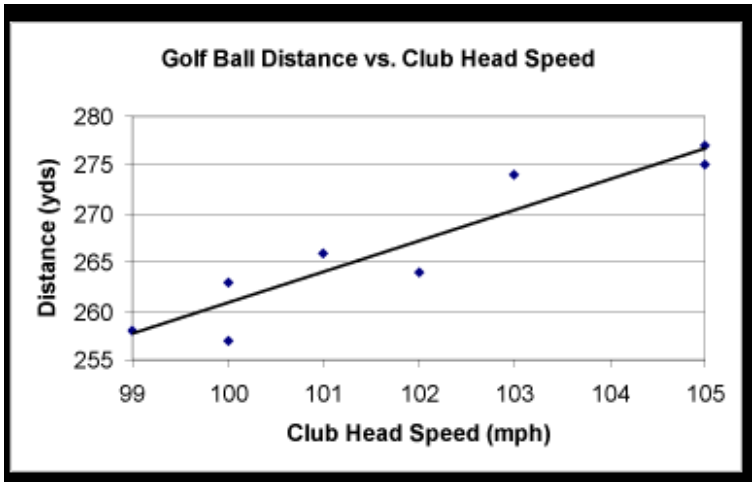
Click here for the enlarged image

In the first graph, the x-axis scale chosen is from 0-100 whereas the club speed ranges from 99-105 and hence all the data points are clustered along a vertical line.

In the second graph, the y-axis scale chosen is from 0-300 whereas distance ranges from 255-280, hence all the data points seem to be clustered in the top most band of the graph.

In the third graph, both the x-axis and y-axis scale chosen are incorrect and hence all data points are clustered in the upper right-hand corner.

The above issue can be rectified by choosing the x-axis scale between 99-105 and y-axis scale between 255-280 as shown below:


Golf Ball Distance vs. Club Head Speed
Distance (yds)
280
275
270
265
260
255
99
100
101
102
103
104
105
Club Head Speed (mph)

References

Barthes, R. (1983). *Empire of signs*. New York, NY, USA: Hill and Wang.

Bederson, B., & Hollan, J. (1994). Pad++: A zooming graphical interface for exploring alternate interface physics. *Proceedings of UIST '94, ACM*, 17–36.

Bertin, J. (1977). *La graphique et le traitement graphique de l'information*. Flammarion. Paris, France.

Bloom. (2014). *Planetary [Electronic Illustration]*. Retrieved 24 May 2014 from https://apps.apple.com/us/app/planetary-remastered/id1473561807

Card, S., Mackinlay, J., and Shneiderman, B. (1999). Readings in Information Visualization: Using Vision to Think. Morgan-Kaufmann.

Cockburn, A., Karlson, A., & Bederson, B.B. (2007). *A review of overview+detail, zooming, and focus+context interfaces*. Retrieved 28 June 2014 from https://www.cs.ubc.ca/~tmm/courses/533/readings/cockburn-computingSurveys09.pdf.

Few, S. (2015). *Signal: Understanding what matters in a world of noise.* Burlingame, CA, USA: Analytics Press.

Friendly, M. (2009). *Gallery of data visualisation*. Retrieved June 21, 2014, from http://euclid.psych.yorku.ca/SCS/Gallery/images/tlens01-s.gif.

Furnas, G. W. (1986). Generalized fisheye views. *Proceedings of CHI '86, ACM*, 17–26.

Furuti, C.A. (2013). *Polyconic projections*. Retrieved June 21, 2014, from http://www.progonos.com/furuti/MapProj/Normal/ProjCon/ProjPolycon/projPolycon.html.

Gaba, E. (2008). *File: world map sinusoidal proj-odeg centered.svg*. Retrieved June 21, 2014, from http://commons.wikimedia.org/wiki/File:World_map_sinusoidal_proj-0deg_centered.svg.

Glen. (2014). CIE chromaticity diagram. Retrieved June 14, 2014, from http://www.ntsc-tv.com/ntsc-index-05.htm.

Google Map [Electronic Illustration]. Retrieved May 17, 2014, from https://www.google.com.sg/maps/place/SIM+University.

Google News [Electronic Illustration]. Retrieved May 17, 2014, from http://news.google.com.

Heer, J. (2004). Position paper for the workshop on information visualisation software infrastructures at IEEE 2004 visualisation. Retrieved June 21, 2014.

Helmholtz, H. L. F. (1866). *Handbuch der Physiologischen Optik*, 1st ed. (3rd ed. translated by J. P. C. Southall). Opt. Soc. Am., 1924-5).

Kard, S. K., Mackinlay, J. D., & Shneiderman, B. (1999). *Readings in information visualization: Using vision to think.* San Francisco, CA, USA: Morgan Kaufmann Publishers.

Knippers, R. (2009). *Map projections*. Retrieved June 21, 2014.

Krygier, J.mB. (2014). *The power of maps and GIS*. Retrieved 21 June 2014 from http://go.owu.edu/~jbkrygie/krygier_html/geog_222/geog_222_lo/geog_222_lo01.html.

Lacan, J. (1964). *The four fundamental concepts of psychoanalysis*. New York, NY, USA: W.W. Norton and Company.

Lamping, J., Rao, R., & Pirolli, P. (1995). A focus + content technique based on hyperbolic geometry for viewing large hierarchies. *Proceedings CHI '95, ACM*, 401–408.

Lévi-Strauss, C. (1963a). *Structural anthropology*. New York, NY, USA: Basic Books.

MacEvoy, B. (2005). *Do "primary" colors exist?* Retrieved June 14, 2014, from http://www.handprint.com/HP/WCL/color6.html.

Mapof.net. (2014). *World map winkel tripel proj – the world maps*. Retrieved June 21, 2014, from http://mapsof.net/map/world-map-winkel-tripel-proj#.VAM6q_zlrIV.

Marr, D. (1982). *Vision*. New York, NY, USA: W. H. Freeman and Company.

Michell, J. (1997). Quantitative science and the definition of measurement in psychology. *British Journal of Psychology, 88*, 355–383.

Michell, J. (1999). *Measurement in psychology – A critical history of a methodological concept*. Cambridge, UK: Cambridge University Press.

Milner, A. D., & Goodale, M. A. (1995). The visual brain in action. *Oxford Psychology Series 27*. Oxford, UK: Oxford University Press.

Murray, G. D. (2016). *Tableau your data!: Fast and easy visual analysis with Tableau software.* (2nd ed.). Indianapolis, IN, USA: John Wiley & Sons Inc.

Rao, R., & Card, S. K. (1994). The table lens: Merging graphical and symbolic representations in an interactive focus + context visualization for tabular information. *Proceedings of CHI '94, ACM*, 318–322.

Rogowitz, B., & Treinish, L. (1996). *Isomorphic color map*.

Sears, T. (2014). *Black bear population boom*. Retrieved June 14, 2014, from http://thejunglestore.blogspot.sg/2013/07/black-bear-population-boom.html.

Stam, J. (2007). *Simulating diffraction*.

Stockman, A. (2003). *Colorimetry*. Institute of Ophthalmology. London, UK.

The University of New Mexico. (1996). *Zoomable graphical sketchpad*. Retrieved June 21, 2014, from http://www.cs.umd.edu/hcil/pad++/papers/jvlc-96-pad/zoom-5.html.

Triesman, A. (1980). Preattentive processing in vision. *Computer Visions, Graphics and Image Processing, 31*, 156-177.

TUBS. (2011). *File: world location map (equirectangular 180).svg*. Retrieved June 21, 2014, from http://commons.wikimedia.org/wiki/File:World_location_map_(equirectangular_180).svg.

Tweedie, L. (1997). Characterizing interactive externalizations. *CHI'97, ACM*, Atlanta Proceedings, 375-382.

University of New Brunswick. (2014). *3-D data visualisation of colour-coded bathymetry data from Passamaquoddy Bay* [Electronic Illustration]. Retrieved May 17, 2014, from http://www2.unb.ca/gge/News/1996/Aquaculture/PassamaquoddyBay.html

Ware, C. (2013). *Information visualization: Perception for design (Interactive technologies)* (3rd ed.). Waltham, MA, USA: Elsevier Inc.

Waterman, S. (1999). *The world of maps*. Retrieved June 21, 2014, from http://www.watermanpolyhedron.com/worldmap.html .

Weskamp, M. (2014). *Newsmap* [Electronic Illustration]. Retrieved May 17, 2014, from http://newsmap.jp.

Worthey, J. A. (2010). *Color matching with amplitude not left out*. Retrieved June 14, 2014, from http://www.jimworthey.com/jimtalk2004nov.html.

Yau, N. (2013). *Data points: visualization that means something*. Indianapolis, IN, USA: John Wiley & Sons Inc.

Young, T. (1802). On the theory of light and colours. *Phil. Trans. R. Soc. Lond. 92*, 20-71.

Zhai, S., Buxton, W., & Milgram, P. (1996). The partial-occlusion effect: Utilizing semitransparency in 3D human-computer interaction. *ACM Transactions on Computer-Human Interaction* (TOCHI), 3(3), 254-284.

Study Unit 3

Four Stages of Data Visualisation

Learning Outcomes

At the end of this unit, you are expected to:

- Explain the end-to-end process of creating a business performance dashboard from data through four stages of the data visualisation process
- Inspect and prepare data for visualisation

Overview

This unit introduces the four stages of data visualisation, together with three feedback loops. In addition, we will also discuss the types of data sets, namely, record data, graph-based data, and ordered data. This unit also discusses the first two stages of data visualisation in detail. Data collection and storage is the methodological process of gathering information about a specific subject and appropriately storing it in a secure. Data preparation is the process of identifying data quality issues and treating these issues so that data can be used for business intelligence (BI), analytics and data visualisation applications. As part of data preparation process, this unit also discusses different data types, dataset architecture and exploratory analysis techniques.

Chapter 1: Data Visualisation Stages

Lesson Recording

Data Visualisation Stages

According to Ware (2013), there are four basic stages in the data visualisation process, together with three feedback loops as illustrated in Figure 3.1. The four basic stages are:

- Data collection and storage: the collection and storage of data
- Data pre-processing: the pre-processing of data to transform them into something one can understand
- Graphics engine: the display hardware and the graphics algorithms to produce data visualisation on screen
- Human visual and cognitive processing: human perceptual and cognitive systems that are involved in interpreting and visualising data

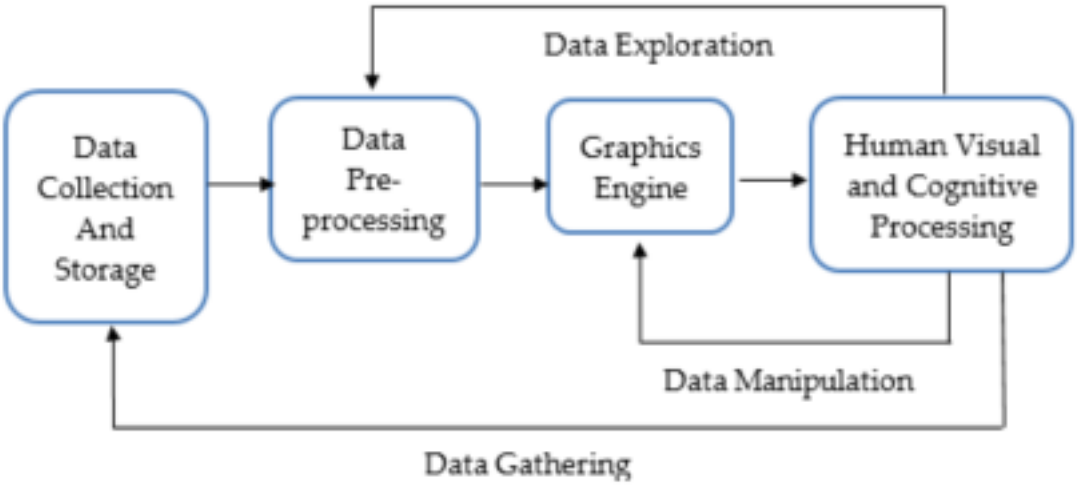


Figure 3.1 The four stages of data visualisation Process (Ware, 2013)

The longest feedback loop is the data gathering loop where the data seeker, such as a business analyst or a market analyst, can choose to gather more data to follow up on an interesting insight provided by data visualisation. Both the physical and the social environments are involved in the data gathering loop. The physical environment is the physical data sources, while the social environment determines in subtle and complex ways what is collected and how the data and data visualisation are interpreted.

The data exploration loop controls the computational preprocess that takes place prior to data visualisation to transform the data before passing the data to the graphic engine. Finally, in the data manipulation loop, the data visualisation process may be highly interactive. For example, in a data visualisation, a market analyst may choose a different vantage point to better understand the market, where the analyst can use a laptop mouse to select parameter ranges that are most interesting.

1.1 Metadata

Metadata are also known as *derived data* in the database modelling community because they are data about data. Metadata are structured information that explain, describe or locate the original (i.e. also known as primary data), and otherwise make the using of original data more efficient. The comparison between primary data and their metadata are illustrated in Figure 3.2 below:

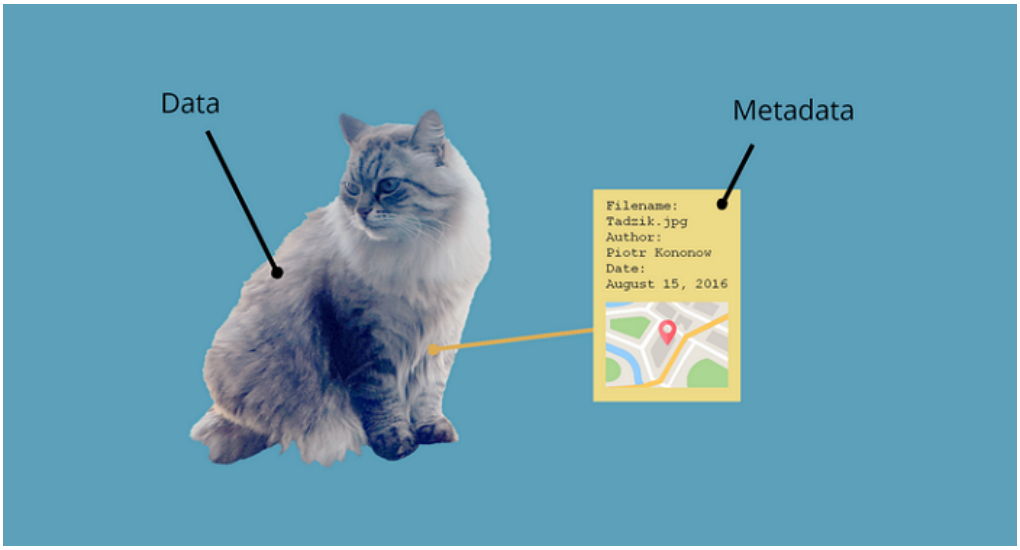


Figure 3.2 Example of data versus metadata (Kononow, 2018)

1.2 Data Dictionary

A data dictionary is a collection of descriptions of the data objects or items in a data model for the benefit of people who need to refer to them.

No.	Name	Description	Measurement	Data Type	Frequency
1	USER_ID	ID of the customer	Nominal	ID	Monthly
2	AGE	Age of the customer	Integer	Numeric	Monthly
3	GENDER	Gender of the customer	Nominal	Character	Monthly
4	SECTOR	The sector that the customer is working in	Nominal	Character	Monthly
7	TOTAL_WEEKLY_SALES	Total purchases made by the customer in a week	Integer	Numeric	Weekly
8	SALES_RANK_MTH	Ranking of customer based on the total purchases made by the customer in a month	Ordinal	Numeric	Monthly

Figure 3.3 Example of data dictionary

1.3 Types of Datasets

Coming on the types of data sets, we define them into three categories— namely, record data, graph-based data, and ordered data.

1.3.1 Record Data

Majority of data mining work assumes that data is a collection of records (data objects). The most basic form of record data has no explicit relationship among records or data fields, and every record (object) has the same set of attributes. Record data is usually stored either in flat files or in relational databases.

Tid	Refund	Marital Status	Taxable Income	Defaulted Borrower
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

(a) Record data.

TID	ITEMS
1	Bread, Soda, Milk
2	Beer, Bread
3	Beer, Soda, Diaper, Milk
4	Beer, Bread, Diaper, Milk
5	Soda, Diaper, Milk

(b) Transaction data.

Projection of x Load	Projection of y Load	Distance	Load	Thickness
10.23	5.27	15.22	27	1.2
12.65	6.25	16.22	22	1.1
13.54	7.23	17.34	23	1.2
14.27	8.43	18.45	25	0.9

(c) Data matrix.

	team	coach	play	ball	score	game	win	lost	timeout	season
Document 1	3	0	5	0	2	6	0	2	0	2
Document 2	0	7	0	2	1	0	0	3	0	0
Document 3	0	1	0	0	1	2	2	0	3	0

(d) Document-term matrix.

Figure 3.4 Types of record data (Gupta, 2019)

There are a few variations of record data, which have some characteristic properties (as shown in Figure 3.4).

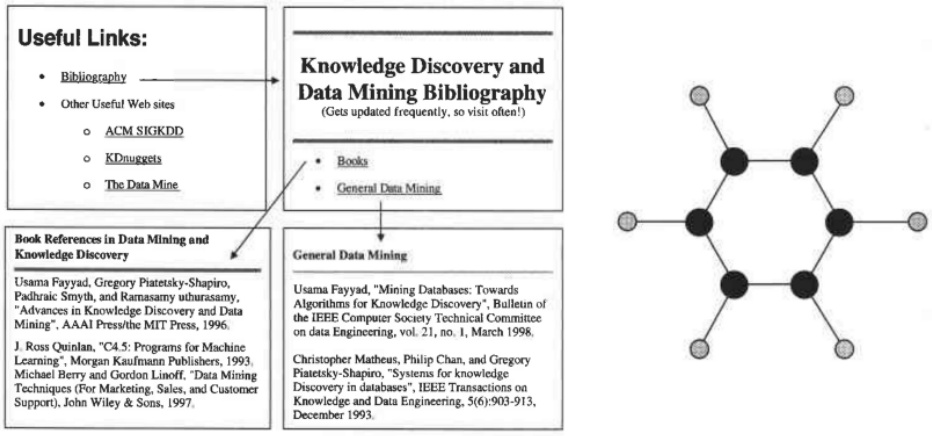
Transaction or market basket data: It is a special type of record data, in which each record contains a set of items. For example, shopping in a supermarket or a grocery store. For any particular customer, a record will contain a set of items purchased by the customer in that respective visit to the supermarket or the grocery store. This type of data is called

market basket data. Transaction data is a collection of sets of items, but it can be viewed as a set of records whose fields are asymmetric attributes. Most often, the attributes are binary, indicating whether or not an item was purchased or not.

The data matrix: If the data objects in a collection of data all have the same fixed set of numeric attributes, then the data objects can be thought of as points (vectors) in a multidimensional space, where each dimension represents a distinct attribute describing the object. A set of such data objects can be interpreted as an m × n matrix, where there are n rows, one for each object, and n columns, one for each attribute. Standard matrix operation can be applied to transform and manipulate the data. Therefore, the data matrix is the standard data format for most statistical data.

The sparse data matrix: A sparse data matrix (sometimes also called document-data matrix) is a special case of a data matrix in which the attributes are of the same type and are asymmetric; i.e., only non-zero values are important.

1.3.2 Graph-based Data



(a) Linked Web pages. (b) Benzene molecule.

Figure 3.5 Types of graph-based data (Gupta, 2019)

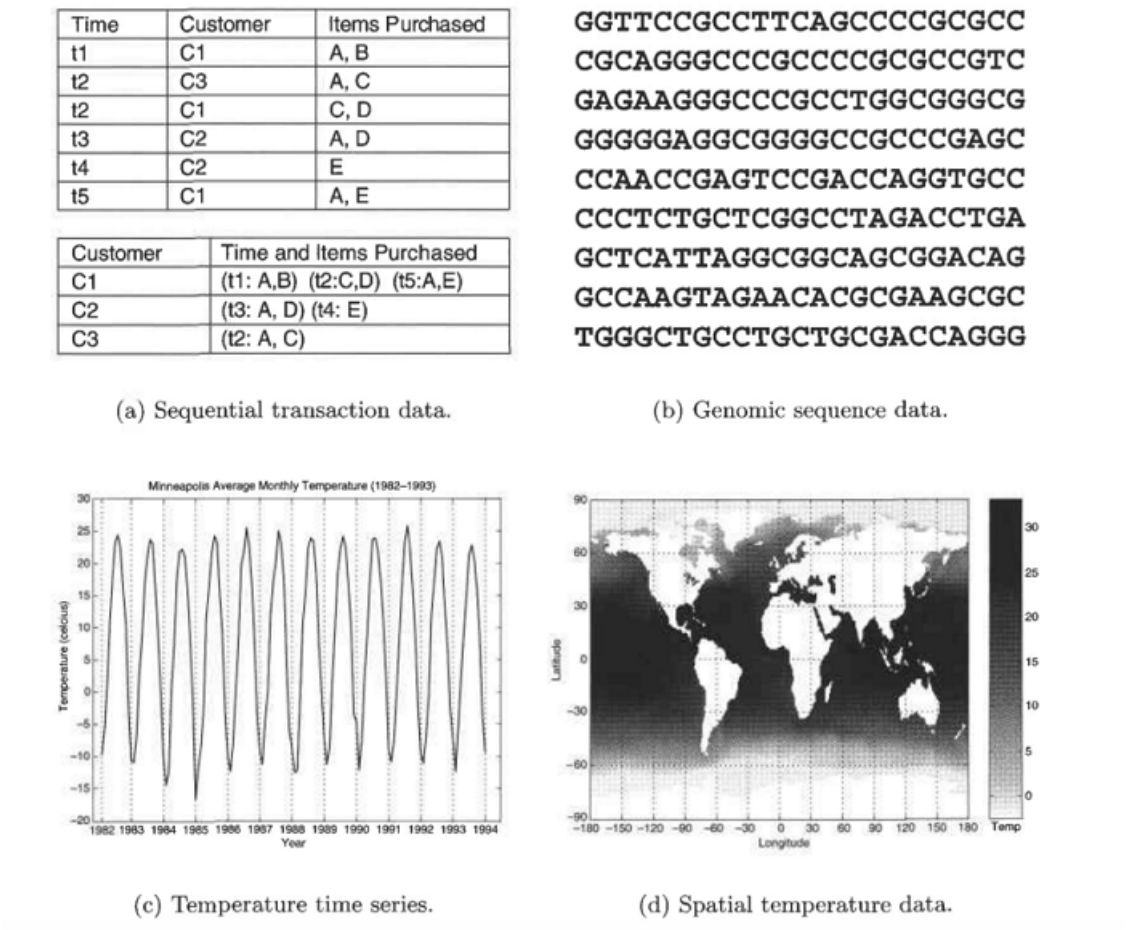
This can be further divided into types (as shown in Figure 3.5):

Data with relationships among objects: The data objects are mapped to nodes of the graph, while the relationships among objects are captured by the links between objects and link

properties, such as direction and weight. Consider web pages on the World Wide Web, which contain both text and links to other pages. In order to process search queries, web search engines collect and process web pages to extract their contents.

Data with objects that are graphs: If objects have structure, that is, the objects contain sub objects that have relationships, then such objects are frequently represented as graphs. For example, the structure of chemical compounds can be represented by a graph, where the nodes are atoms and the links between nodes are chemical bonds.

1.3.3 Ordered Data



Time	Customer	Items Purchased
t1	C1	A, B
t2	C3	A, C
t2	C1	C, D
t3	C2	A, D
t4	C2	E
t5	C1	A, E

Customer	Time and Items Purchased
C1	(t1: A,B) (t2:C,D) (t5:A,E)
C2	(t3: A, D) (t4: E)
C3	(t2: A, C)

(a) Sequential transaction data.

```
GGTTCCGCCTTCAGCCCCGCGCC
CGCAGGGCCCGCCCCGCGCCGTC
GAGAAGGGCCCGCCTGGCGGGCG
GGGGGAGGCGGGGCCGCCCGAGC
CCAACCGAGTCCGACCAGGTGCC
CCCTCTGCTCGGCCTAGACCTGA
GCTCATTAGGCGGCAGCGGACAG
GCCAAGTAGAACACGCGAAGCGC
TGGGCTGCCTGCTGCGACCAGGG
```

(b) Genomic sequence data.

(c) Temperature time series.

(d) Spatial temperature data.

Figure 3.6 Types of ordered data (Gupta, 2019)

For some types of data, the attributes have relationships that involve order in time or space. As you can see in the picture above, they can be segregated into four types (as shown in Figure 3.6):

Sequential data: Also referred to as temporal data, they can be thought of as an extension of record data, where each record has a time associated with it. Consider a retail transaction data set that also stores the time at which the transaction took place.

Sequence data: Sequence data consists of a data set that is a sequence of individual entities, such as a sequence of words or letters. It is quite similar to sequential data, except that there are no time stamps; instead, there are positions in an ordered sequence. For example, the genetic information of plants and animals can be represented in the form of sequences of nucleotides that are known as genes.

Time series data: Time series data are a special type of sequential data in which each record is a time series, i.e., a series of measurements taken over time. For example, a financial data set might contain objects that are time series of the daily prices of various stocks.

Spatial data: Some objects have spatial attributes, such as positions or areas, as well as other types of attributes. An example of spatial data is weather data (precipitation, temperature, pressure) that is collected from a variety of geographical locations.

Chapter 2: Data Collection and Storage

Lesson Recording

Data Collection and Storage

Data collection is the methodological process of gathering information about a specific subject. It is crucial to ensure your data is complete during the collection phase and that it is collected legally and ethically. If not, your analysis won't be accurate and could have far-reaching consequences.

2.1 Data Collection Methods

2.1.1 Surveys

Surveys are physical or digital questionnaires that gather both qualitative and quantitative data from subjects. One situation in which you might conduct a survey is gathering attendee feedback after an event. This can provide a sense of what attendees enjoyed, what they wish was different, and areas you can improve or save money on during your next event for a similar audience. Because they can be sent out physically or digitally, surveys present the opportunity for distribution at scale. They can also be inexpensive; running a survey can cost nothing if you use a free tool. If you wish to target a specific group of people, partnering with a market research firm to get the survey in the hands of that demographic may be worth the money. Something to watch out for when crafting and running surveys is the effect of the following biases:

Collection bias: It can be easy to accidentally write survey questions with a biased lean. Watch out for this when creating questions to ensure your subjects answer honestly and are not swayed by your wording.

Subject bias: Because your subjects know their responses will be read by you, their answers may be biased toward what seems socially acceptable. For this reason, consider pairing survey data with behavioural data from other collection methods to get the full picture.

2.1.2 Transactional Tracking

Each time your customers make a purchase, tracking that data can allow you to make decisions about targeted marketing efforts and understand your customer base better.

Often, e-commerce and point-of-sale platforms allow you to store data as soon as it's generated, making this a seamless data collection method that can pay off in the form of customer insights.

2.1.3 Interviews and Focus Groups

Interviews and focus groups consist of talking to subjects face-to-face about a specific topic or issue. Interviews tend to be one-on-one, and focus groups are typically made up of several people. You can use both to gather qualitative and quantitative data. Through interviews and focus groups, you can gather feedback from people in your target audience about new product features. See them interact with your product in real-time and record their reactions and responses to questions that can provide valuable data about which product features to pursue.

As is the case with surveys, these collection methods allow you to ask subjects anything you want about their opinions, motivations, and feelings regarding your product or brand. It also introduces the potential for bias. Aim to craft questions that don't lead them in one particular direction. One downside of interviewing and conducting focus groups is they can be time-consuming and expensive. If you plan to conduct them yourself, it can be a lengthy process. To avoid this, you can hire a market research facilitator to organise and conduct interviews on your behalf.

2.1.4 Observation

Observing people interacting with your website or product can be useful for data collection because of the candour it offers. If your user experience is confusing or difficult, you can witness it in real-time. Setting up observation sessions can be difficult. You can use a third-party tool to record users' journeys through your site or observe a user's interaction with a beta version of your site or product.

While less accessible than other data collection methods, observations enable you to see first-hand how users interact with your product or site. You can leverage the qualitative and quantitative data gleaned from this to make improvements and double down on points of success.

2.1.5 Online Tracking

To gather behavioural data, you can implement pixels and cookies. These are both tools that track users' online behaviour across websites and provide insight into what content they are interested in and typically engage with. You can also track users' behaviour on your company's website, including which parts are of the highest interest, whether users are confused when using it, and how long they spend on product pages. This can enable you to improve the website's design and help users navigate to their destination.

Inserting a pixel is often free and relatively easy to set up. Implementing cookies may come with a fee but could be worth it for the quality of data you will receive. Once pixels and cookies are set, they gather data on their own and do not need much maintenance, if any. It is important to note: tracking online behaviour can have legal and ethical privacy implications. Before tracking users' online behaviour, ensure you are in compliance with local and industry data privacy standards.

2.1.6 Forms

Online forms are beneficial for gathering qualitative data about users, specifically demographic data or contact information. They are relatively inexpensive and simple to set up, and you can use them to gate content or registrations, such as webinars and email

newsletters. You can then use this data to contact people who may be interested in your product, build out demographic profiles of existing customers, and in remarketing efforts, such as email workflows and content recommendations.

2.1.7 Social Media Monitoring

Monitoring your company's social media channels for follower engagement is an accessible way to track data about your audience's interests and motivations. Many social media platforms have analytics built in, but there are also third-party social platforms that give more detailed, organised insights pulled from multiple channels.

You can use data collected from social media to determine which issues are most important to your followers. For instance, you may notice that the number of engagements dramatically increases when your company posts about its sustainability efforts.

2.2 Data Collection Challenges and Improvements

This section provides information on some of the most common challenges encountered in the data collection process and improving data collection practices in general. It provides advice on improving data quality for analysis and reporting purposes, and information on privacy and security requirements.

Table 3.1 Data collection challenges and improvements

Challenges	Improvements
Inconsistent data collection standards Currently, there are many national and state-wide data standards which are used for collecting administrative data. These standards are not always broadly applied, and may themselves be inconsistent, and this can impact the comparability of data collections.	**Data items have pre-defined responses** Where appropriate, it is recommended that data items have a pre-defined set of response options at the point of entry into a data management system. This reduces the potential for typographical errors and

Challenges	Improvements
For example, many specialist and non-disability specific support services collect data on disability drawing on definitions used by the National Disability Insurance Scheme (NDIS), the National Disability Agreement and state governing bodies. Different types of services may apply different standards depending on what is most relevant for their service provision. As a result, the scope and detail of information collection may not be consistent across services, making it complex to compare data between services, or to population level data sets.	enables more efficient data collection and subsequent analysis.
Data collection is not core to business function Administrative data are typically collected as a by-product of operational requirements or to meet an internal business need and may only include core information needed to perform a service, such as a client's contact details. In such cases, information on an individual's sexual orientation, cultural background or disability may not be seen as an operational requirement for organisations that do not offer specialised services. As a result, organisations may only	**Using data-related key performance indicators (KPIs)** Setting KPIs linked to data and evidence can be a motivating factor for organisations to ensure improvement in their data collection practices. KPIs can target many aspects of data quality including completeness (how many records have a recorded value), and precision (how many records have a meaningful or valid value). For example, an organisation finds that only 50% of the clients contained within their record management system have a

Challenges	Improvements
collect a narrow range of data items, which lack sufficient detail needed for broader secondary use purposes, such as conducting state-wide service analysis, monitoring or research.	recorded gender. The organisation sets a KPI for 100% of clients to have a recorded gender, and they monitor this goal over the course of a six-month period to ensure that improvements made are effectively moving towards this goal.
Lack of training in data collection As the primary role of front-line service staff will generally not be data collection, they may not receive training in this area. If staff do not receive training or understand why they need to collect particular data, they may feel less confident to ask the associated questions, or ask them in a different way.	**Staff training** in the benefit of collecting these data items, and in sensitive or culturally appropriate ways to do so, can build staff understanding of the value of these types of data, and assist in building data quality and consistency. Training should emphasise why it is important to collect data and highlight the benefits of data for operations, planning, research and evaluation. If staff understand the rationale for collecting certain information, they will feel more confident to ask for these data items and to explain why it is important.
Lack of quality assurance processes There may be limited opportunities to confirm information with a person who has been in contact with a service, meaning that the data initially collected cannot be verified. Additionally, the sophistication of record keeping systems	**Creating mandatory data fields** Where appropriate, it is advised that service providers and agencies update their data collection infrastructure to utilise mandatory data fields (or at minimum, prompts, on all non-optional data items). Therefore, the person

Challenges	Improvements
can vary and data quality is often reliant on the person entering the data correctly. Depending on the resourcing of an organisation, time may not permit staff to review information for completeness and obtain missing data.	inputting the data cannot move to the next screen without entering a response in mandatory data fields. For example, to improve their collection of gender information, an organisation updates their data entry system so that a response for gender must be recorded when entering details about a new client before the new entry can be completed.
Economic and IT restrictions Some organisations may not have the capacity or infrastructure to prioritise improvements to data collection systems and processes. This may be due to a backlog of paper-based records to be digitised, a small workforce to input and maintain data, and lack of budget to upgrade records management systems. In some cases, these IT systems may have limited capacity to include multiple response values or dynamic questioning, which supports sophisticated data collection. The introduction of multiple response options may also present problems for exporting and analysing data.	**Commitment from all levels of an organisation** Improving data collection and the quality of data holdings requires a concerted effort from an entire organisation and should begin with a top-down commitment for change. This includes identifying priority areas for improvement and barriers to improvement, adopting best practice procedures for collecting quality data, ensuring IT infrastructure is kept up to date and allows for efficient and effective data collection, and providing training where needed to those collecting data to ensure confidence and consistency in data collection practices.

2.3 Data Storage Challenges and Solutions

These are some of the most important potential data storage issues you will need to consider:

Table 3.2 Data storage challenges and solutions

Challenges	Solutions
Infrastructure Data needs a place to rest, the same way objects need a shelf or container; data must occupy space. If you plan on storing vast amounts of data, you will need the infrastructure necessary to store it, which often means investing in high-tech servers that will occupy significant space in your office or building.	**Cloud hosting** One of the easiest workarounds is to use cloud hosting and cloud storage, which take advantage of another company's infrastructure to save you that space and the trouble of setting things up yourself.
Cost Running your own data centre is an expensive operation. You will need to spend money on initial setup, ongoing maintenance, and the costs associated with the people responsible for maintaining it.	**Outsource the work** Again, the best solution here is to outsource the work; you will probably have to pay a monthly fee, but it will save you money in the long run.
Security Security is a major issue to overcome. Hypothetically, if your data is stored somewhere, it's possible for a third party to obtain it. There are many layers	**Run a tight operation** You will need to run a tight operation, choosing the best partners and keeping

Challenges	Solutions
of security that can help you prevent this unauthorised access, including encryption and reliance on third-party providers, but there's a limit to how well these can protect you—even the FBI has trouble maintaining the security of its data when its own best practices aren't followed.	your own team adhering to best practices at all times.
Corruption Practically every form of data storage has the potential to be corrupted. Stray particles can interfere with most forms of data storage, and anything relying on magnetic strips or electric storage can be corrupted by electromagnetic interference. Even if there is not an outside source directly interfering with it, data can naturally degrade over time.	**Use multiple backups** Your best bet for protection here is utilising multiple backups.
Scale You might be able to find a storage solution that serves your current needs adequately, but what happens if those needs change suddenly? How will you account for your needs as they stand in five years? Your data storage solution needs some capacity to scale.	**Explore options** Here, it pays to give yourself as many options as possible, since you will not be sure exactly how your needs will change in the future.
UI and accessibility	**Use a system with good UI**

Challenges	Solutions
Your data will not be much good to you if it is hard to access; after all, data storage is just a temporary measure so you can later analyse the data and put it to good use.	Accordingly, you will need some kind of system with an intuitive, accessible user interface (UI), and clean accessibility for whatever functionality you want.
Compatibility If you plan on using multiple systems or applications with your data, you will need to ensure they are compatible.	**Open API** For that, you will need to find a data storage partner with an open API and a clean system of transition.

Chapter 3: Data Preparation

Lesson Recording

Data Preparation

Data preparation is the process of combining, structuring and organising data so it can be used in business intelligence (BI), analytics and data visualisation applications. The components of data preparation include profiling, cleansing, and transformation.

3.1 What Are the Benefits of Data Preparation?

One of the primary benefits of data preparation is to ensure that raw data being readied for processing and analysis is accurate and consistent so the results of BI and analytics applications will be valid. Data is commonly created with missing values, inaccuracies or other errors, and separate data sets often have different formats that need to be reconciled when they are combined. Correcting data errors, validating data quality and consolidating data sets are big parts of data preparation projects.

Data scientists often complain that they spend most of their time gathering, cleansing and structuring data instead of analysing it. A big benefit of an effective data preparation process is that they and other end users can focus more on data mining and data analysis — the parts of their job that generate business value. For example, data preparation can be done more quickly, and prepared data can automatically be fed to users for recurring analytics applications.

- Done properly, data preparation also helps an organisation do the following:
- ensure the data used in analytics applications produces reliable results;
- identify and fix data issues that otherwise might not be detected;

- enable more informed decision-making by business executives and operational workers;
- reduce data management and analytics costs;
- avoid duplication of effort in preparing data for use in multiple applications; and
- get a higher ROI from BI and analytics initiatives.

Effective data preparation is particularly beneficial in big data environments that store a combination of structured, semi-structured and unstructured data, often in raw form until it is needed for specific analytics uses. Those uses include predictive analytics, machine learning (ML) and other forms of advanced analytics that typically involve large amounts of data to prepare. For example, in an article on preparing data for machine learning, Felix Wick, corporate vice president of data science at supply chain software vendor Blue Yonder, is quoted as saying that data preparation "is at the heart of ML."

3.2 Data Preparation Principles

1. Understand the data consumer – who is going to use the data and what questions do they need answered.
2. Understand the data – where it is coming from and how it was generated.
3. Save the raw data. If the data engineer has the raw data, then all the data transformations can be recreated. Additionally, do not move or delete the raw data once it is saved.
4. Ensure that transforms are reproducible and deterministic. Each transform must produce the same results each time it is executed given the same input data set, without harmful effects.
5. Future proof your data pipeline. Version not only the data and the code that performs the analysis, but also the transforms that have been applied to the data.

3.3 Steps in the Data Preparation Process

The data preparation process comprises of four steps:

1. Understand the data: data types

2. Understanding the data: dataset architecture
3. Exploratory data analysis (EDA) to identify the following:
    - Properties of data
    - Data quality issues
4. Apply appropriate treatment methods

3.3.1 Understanding Data: Data Types

1. **Nominal**

    Nominal measurement measures items based on their labels or categories or other qualitative classification the items belong to with no implied order. Two types of nominal assignments are Type A where the "numbering" of product represents the identification of the product, and Type B where the "numbering" for each member of a same product type or product category is assigned the same numeral. Since the purpose is served when any two designating numerals are interchanged, this measurement form remains invariant under the general substitution or permutation group (sometimes called the *symmetric group of transformations*).

    The only statistic relevant to nominal measurements of Type A is the number of cases. For example, the count of product identification numbers in the company's sales order system. For Type B, we can determine the most common item (the mode), and under certain conditions we can test, through contingency methods, hypotheses regarding the distribution of cases among the classes, for example the most frequent product categories bought by customers or the count of products in each product categories.

2. **Ordinal**

    The ordinal measurement arises from the operation of rank ordering. An example of an ordinal measurement is the measurement of customer satisfaction rating. Other instances are found among measurements of examination grade and quality of product, or measurements used by psychologists.

"In the strictest propriety, the ordinary statistics involving means and standard deviations ought not to be used as the ordinal measurements, because these statistics imply a knowledge of something more than the relative rank-order of data. Means and standard deviations computed on the relative rank-order data are in error to the extent that the successive intervals on the measurements are unequal in size. When only the rank-order of data is known, we should proceed cautiously with our statistics, especially with the conclusions we draw from them.

Even in applying those statistics that are normally appropriate for the rank-order data, we sometimes find rigour compromised. Thus, although it is indicated in Table 3.3 that the percentile measures may be applied to the rank-ordered data, it should be pointed out that the customary procedure of assigning a value to a percentile by interpolating linearly within a class interval is, in all strictness, wholly out of bounds. Likewise, it is not strictly proper to determine the mid-point of a class interval by linear interpolation, because the linearity of an ordinal measurement is precisely the property which is open to questions." (Stevens, 1946)

3. **Interval**

The interval measurement allows us to measure the degree of difference between items, but not the ratio between them. We can use almost all the usual statistical measurements here, but we cannot use the interval measurement to measure items that have a zero point. An example of the interval measurement is the measurement of temperature with the centigrade or Celsius scale. The Celsius scale defines the freezing and boiling points of water at specific conditions, then separate both points into 100 intervals. 10°C cannot be said to be thrice as hot as 30°C. However, one temperature difference can be said to be thrice that of another.

Another example is the measurement of time. When we measure dates from an arbitrary epoch like AD, the measurement can be transformed to those on

another by using the $y=ax+b$ equation, but it will be meaningless to say that one value in an arbitrary epoch is twice or some other proportion greater than value in another arbitrary epoch. Periods of time, however, can be measured using ratio measurements, because one period may be correctly defined as double of another.

4. **Ratio**

Ratio measurements estimate the ratio between a magnitude of a continuous quantity and a unit magnitude of the same kind (Michell, 1997, 1999). Under ratio measurement, we can use the operations for determining all four relations: equality, rank-order, equality of intervals, and equality of ratios. An absolute zero is always implied, even though the zero value on some scales (e.g. absolute temperature) may never be produced. All types of statistical measures are applicable to ratio measurements.

Table 3.3 Overview of the four levels of measurement (Stevens, 1946)

Level of Measurement	Basic Empirical Operations	Mathematical Group Structure	Permissible Statistics
Nominal	Determination of equality	Permutation group $x' = f(x)$ f(x) means any one-to-one substitution	• Number of cases • Mode • Contingency correlation
Ordinal	Determination of greater or less	Isotonic group $x' = f(x)$ f(x) means any monotonic increasing function	• Median • Percentiles

Level of Measurement	Basic Empirical Operations	Mathematical Group Structure	Permissible Statistics
Interval	Determination of equality of intervals or differences	General linear group $x' = ax + b$	• Mean • Standard deviation • Rank-order correlation • Product-moment correlation
Ratio	Determination of equality of ratios	Similarity group $x' = ax$	• Coefficient of variation

Discrete and continuous data

Discrete data is a numerical type of data that includes whole, concrete numbers with specific and fixed data values determined by counting. Discrete data refers to individual and countable items (discrete variables). Some examples of discrete data one might gather — number of customers who bought different items, number of computers in each department, number of items you buy at the grocery store each week, etc.

Continuous data includes complex numbers and varying data values that are measured over a specific time interval. Continuous data is a type of numerical data that refers to the unspecified number of possible measurements between two realistic points. Some examples of continuous data include— weight of newborn babies, daily wind speed, temperature of a freezer, etc.

Table 3.4 Comparison of discrete vs continuous data characteristics

Discrete data	Continuous data

Takes specific countable values	Takes any measured value within a specific range
Some common examples of discrete data are the number of students, the number of children, the shoe size, and so on	Some common examples of continuous data are height, weight, length, time, temperature, age, and so on
Ordinal data values and integer values represent discrete data	Decimal numbers and fractions represent continuous data
Easily counted on something as simple as a number line	Requires more in-depth measurement tools and methods like curves and skews
Discrete data remains constant over a specific time interval	Continuous data varies over time and can have separate values at any given point

3.3.2 Understanding Data: Dataset Architecture

1. **Variable hierarchy**

   A hierarchical dimension is a means of organising and structuring parent-child (one-to-many) data within a single dimension and using self-relations to organise the values of the hierarchical dimension into groups. A hierarchy exists when values within a dimension are arranged in levels, with each level representing the aggregated total of the data from the level below. Some dimensions have multiple hierarchies based on them.

   For example, rather than defining two separate dimensions, one for city and the other for region, you could define a hierarchical dimension named GEOGRAPHY that contains both city and region values. Similarly, as shown in Figure 3.7 below, the vehicle rating comprises of several dimensions that are shown as second level: energy efficiency, safety features, style and design, comfort, ease of obtaining spare parts and maintenance costs.

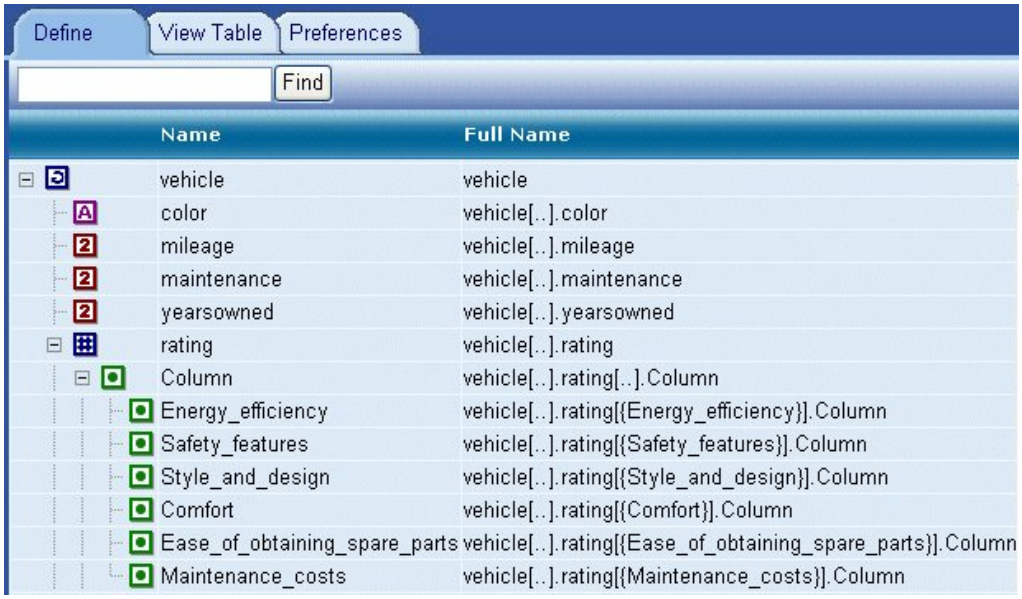


Figure 3.7 Variable hierarchy example (UNICOM Intelligence, n.d.)

2. **Data granularity**

Data granularity is a measure of the level of detail in a data structure. In time-series data, for example, the granularity of measurement might be based on intervals of years, months, weeks, days, or hours. For ordering transactions, granularity might be at the purchase order level, or line item level, or detailed configuration level for customised parts. The name field could represent the full name or have separate entries for first name, middle name, and last name. The level of data granularity determines what analysis can be performed on the data, and whether results from that analysis lead to appropriate conclusions. A good example of data granularity is how a name field is subdivided, if it is contained in a single field or subdivided into its constituents such as first name, middle name and last name. As the data becomes more subdivided and specific, it is also considered more granular.

3. **Data aggregation**

Data aggregation may be done manually or through specialised software called automated data aggregation. For example, new data can be aggregated over a given period to provide statistics such as sum, count, and average, minimum,

maximum. After the data is aggregated and written to a view or report, you can analyse the aggregated data to gain useful insights about particular resources or resource groups.

There are two types of data aggregation.

Time aggregation: It is data points for a single resource over a specified period.

Spatial aggregation: It is data points for a group of entities over a specified period. The group of entities could be people in a team, SKU's of a product, products in a basket, places in a region/country etc.

Companies often collect data from their online customers and website visitors. Google collects data in the form of cookies to show targeted advertisements to its users. Facebook is doing the same thing by collecting and analysing the information and showing ads to its users. The aggregate data would combine statistics on customer demographics and behaviour metrics, such as average age or number of transactions. The consumer uses a single master personal identification number (PIN) to give them access to their various accounts (such as those for financial organisations, airlines, book and music clubs, and so on). Performing that type of data aggregation is sometimes referred to as "screen scraping."

The aggregated data can be handled by the marketing team to personalise messaging, offers, and more in the user's digital experience with the brand. It overall improves the user experience. It can also be used by the product team to learn which products are successful and which are struggling. Moreover, the data can also be used by company executives and finance teams to help them decide how to allocate budget towards marketing or product development strategies.

4. **Slicing and dicing**

Refers to a way of segmenting, viewing and comprehending data in a database. Large blocks of data is cut into smaller segments and the process is repeated until the correct level of detail is achieved for proper analysis. Therefore, slicing and

dicing presents the data in new and diverse perspectives and provides a closer view of it for analysis. For example, a report is showing annual performance of a particular product. If we want to view the quarterly performance, we can use a slicing and dicing strategy to drill down to the quarterly level. It is mainly done using the filter actions in the software.

Chapter 4: Exploratory Analysis and Treatment Methods

Data sets pulled together from different source systems are highly likely to have numerous data quality, accuracy and consistency issues to resolve. It is a time-consuming process: The 80/20 rule is often applied to analytics applications, with about 80% of the work said to be devoted to collecting and preparing data and only 20% to analysing it.

Table 3.5 outlines the exploratory data analysis methodologies, measures within each methodology that help identify the issues with the data and then points to the appropriate treatment methods that can be deployed to resolve the issues.

Table 3.5 Exploratory data analysis methodologies

S No.	Method	Measure	Issue	Treatment
1	Visual examination		Inconsistent format in data fields	Format standardisation examples— name, address and date
2	Summary statistics	Number of observations Mean, median and mode	Missing values	Drop the observations Estimate missing value Ignore missing values
		Minimum, maximum, range	Inconsistent data values	Replace with additional information or null value
3	Data distribution	Histogram box and whisker plot	Outliers	Drop or truncate Quantile based capping Median imputation

4.1 Visual Examination

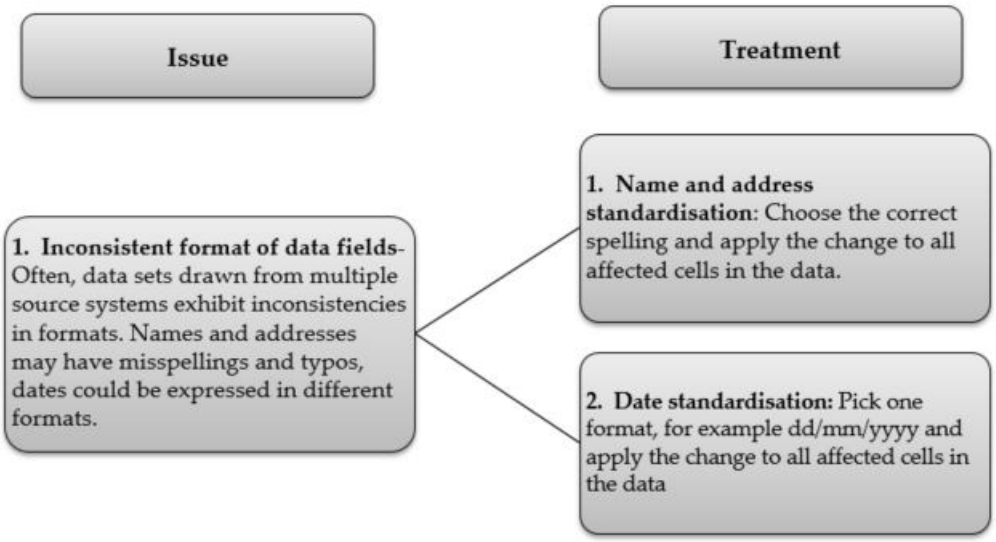


Figure 3.8 Data quality issues highlighted by visual examination and corresponding treatment methods

Click here for the enlarged image

4.2 Summary Statistics

Table 3.6 Summary statistics

Measure	Definition	Issues
Number of observations	Number of non-null or missing values	Missing values
Mean	Average of all the observations present in the table. Sum of observations / Total number of elements in data set	
Median	Middle value of the collection of data when arranged in	

Measure	Definition	Issues
	ascending order and descending order	
Mode	Variable or number or value which is repeated maximum number of times in the set	
Minimum	Lowest value of the variable	Inconsistent data values
Maximum	Highest value of the variable	
Range	Evaluate the spread among the data members Maximum value – Minimum value	

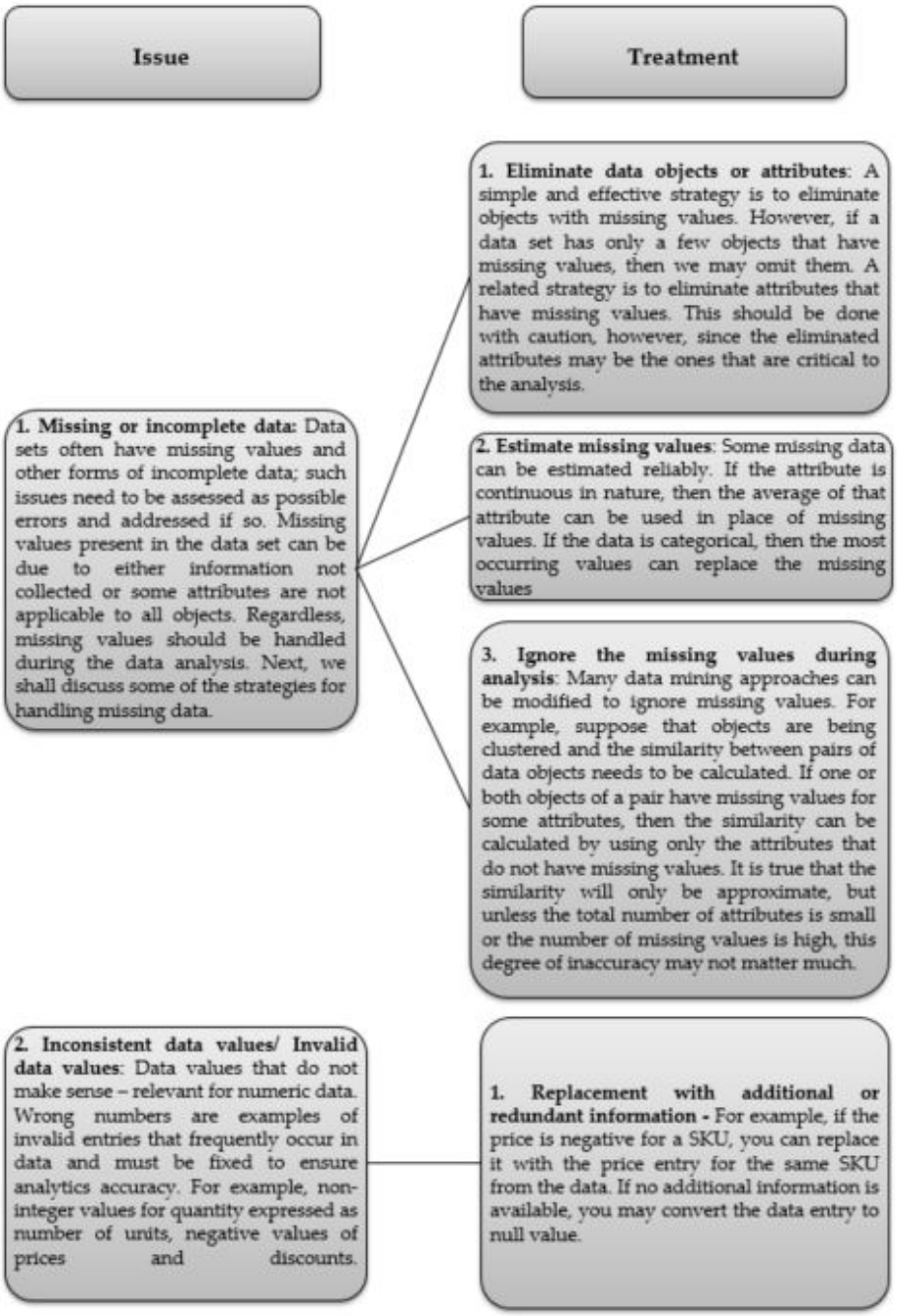


Figure 3.9 Data quality issues highlighted by summary statistics and corresponding treatment methods

Click here for the enlarged image

4.3 Data distribution

In addition to middle or average value of our data, we should also know more about the distributions of our data. We can plot the distribution of our data with histogram and box and whisker plots. For distribution of one variable, using a histogram will enable us to see where the data is clustered, i.e. whether it is positively skewed, negatively skewed, or normally distributed. Box and whisker plots give additional information about outliers, if any, on the vertical axis.

4.3.1 Histogram

A histogram turns continuous variables into discretely bucketed bins of variables. It uses bar length to display the value of the variable. Using the e-commerce company scenario, when the vice president of sales of the company would like to visualise the distribution of the sales in the year 2013, he can use the histogram in Tableau as illustrated in Figure 3.10. Below are the steps to creating a histogram in Tableau:

1. Create a new worksheet.
2. Right click on one measure (i.e. Sales 2013) in the "Data" window and select "Create Bins".
3. On the "Create Bins" interface, key in "Sales Bin" under "New Field Name", and "1000" under "Size of Bins". Click the "Load" button and it will generate the Min, Max and Diff values. Finally, click on "OK" to close the "Create Bins" interface.
4. Drag one measure (i.e. Sales 2013) into the worksheet's rows.
5. Drag the "Sales Bin" into the worksheet's columns.

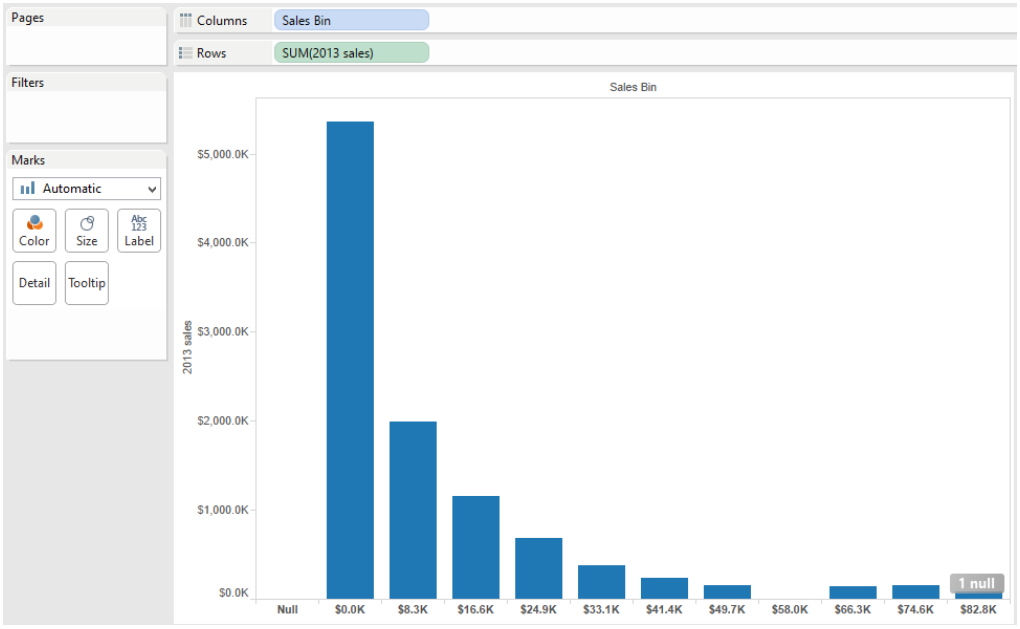


Figure 3.10 Histogram Chart

From Figure 3.10, it can be seen that the distribution is skewed on the right, which means that most of the sales values fall in the range of $0-$25k. There could be a few values which are extremely high, known as outliers, and can interfere in data analysis. Box and whisker plots further highlight such values in the data.

Read

Here are some optional readings to serve as additional references for students who want to learn more about the topics just covered:

1. Yau, N. (2013), pp. 149, 197-198.
2. Tableau Online Help – Histogram at Tableau Online Help – Bubble Charts at http://kb.tableausoftware.com/articles/knowledgebase/bulletgraph.

4.3.2 Box and Whisker plots

Box plots are intended to show a distribution of data, and that can be difficult when data is aggregated. You can configure lines, called whiskers, to display all points at the maximum extent of the data. Boxes indicate the middle 50% of the data. Below are the steps to creating box and whisker plots in Tableau:

1. Create a new worksheet.
2. Drag one measure (i.e. Discount) into the worksheet's rows.
3. Drag the "Segment" and "Region" into the worksheet's columns.
4. From the drop down options on the right hand side, select box and whisker plot.

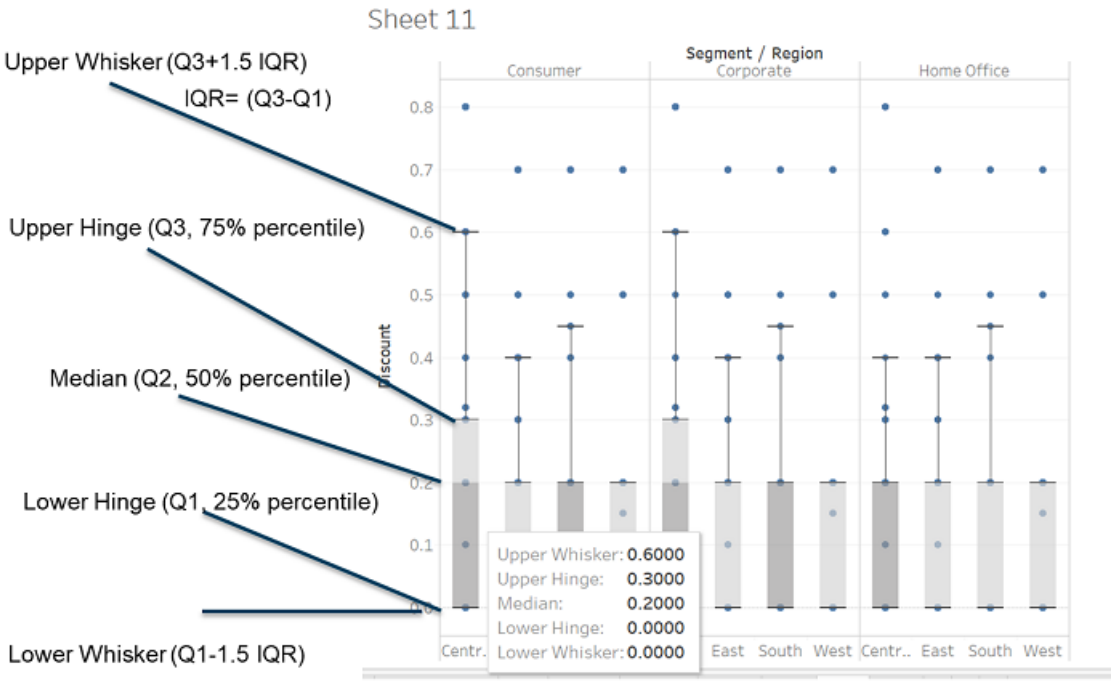


Figure 3.11 Box and whisker plot

"Percentile" is in everyday use, but there is no universal definition for it. The most common definition of a percentile is a number where **a certain percentage of scores fall below that number.** You might know that you scored 67 out of 90 on a test. But that figure has no real meaning unless you know what percentile you fall into. If you know that your score is in the 90th percentile, that means you scored better than 90% of people who took the test.

The 25th percentile is also called the first quartile. The 50th percentile is generally the median. The 75th percentile is also called the third quartile. The difference between the third and first quartiles is the interquartile range. The default boxplot created has whiskers that is 1.5 times of IQR with upper whisker = Q3 + (Q3-Q1)*1.5 and lower whisker = Q1 - (Q3-Q1)*1.5. The whiskers can be set to the min and max value of profit as well.

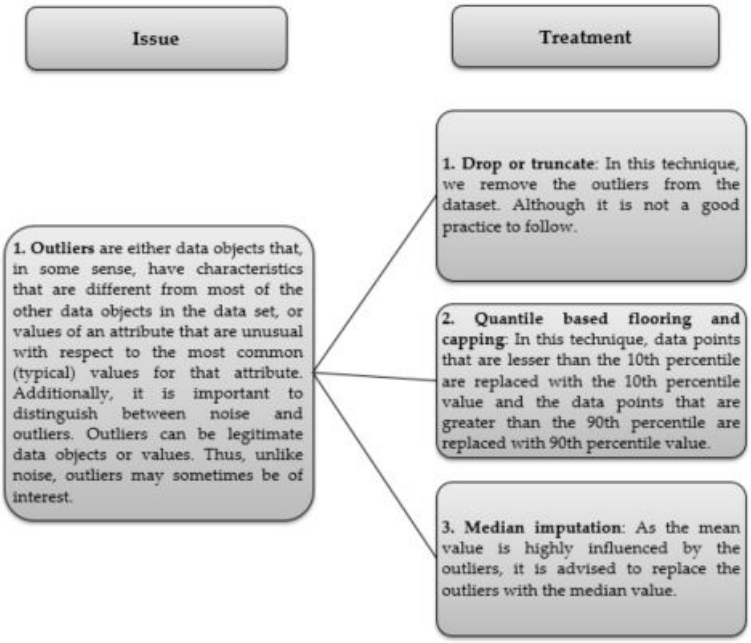


Figure 3.12 Data quality issues highlighted by summary statistics and corresponding treatment methods

Click here for the enlarged image

Summary

There are four basic stages in the data visualisation process, together with three feedback loops— data collection and storage, data pre-processing, graphics engine, human visual and cognitive processing. Metadata are structured information that explain, describe or locate the original (i.e. also known as primary) data, otherwise make the using of original data more efficient. A data dictionary is a collection of descriptions of the data objects or items in a data model for the benefit of people who need to refer to them.

Record data is a collection of records (data objects). Record data is usually stored either in flat files or in relational databases. Transaction or market basket data is a special type of record data, in which each record contains a set of items. The data matrix is a collection of data that all have the same fixed set of numeric attributes. A sparse data matrix (sometimes also called document-data matrix) is a special case of a data matrix in which the attributes are of the same type and are asymmetric. Sequential data can be thought of as an extension of record data, where each record has a time associated with it. Sequence data consists of a data set that is a sequence of individual entities, such as a sequence of words or letters. Time series data is a special type of sequential data in which each record is a time series.

Data collection is the methodological process of gathering information about a specific subject. Common data collection methods are surveys, transactional tracking, interviews and focus groups, observation, online tracking, forms and social media monitoring. It is important to consider the inherent strength and weaknesses of each of the data collection methods and use the one that is most suitable to the purpose and budget. Some of the common challenges of data collection methods can be addressed by having pre-defined responses for the data items, using data related KPI's, staff training, creating mandatory data fields and ensuring commitment from all levels of the organisation.

Data preparation is the process of combining, structuring and organising data so it can be used in business intelligence (BI), analytics and data visualisation applications. The components of data preparation include profiling, cleansing, and transformation. Data

sets pulled together from different source systems are highly likely to have numerous data quality, accuracy and consistency issues to resolve. It is a time-consuming process: The 80/20 rule is often applied to analytics applications, with about 80% of the work said to be devoted to collecting and preparing data and only 20% to analysing it.

Effective data preparation is particularly beneficial in big data environments that store a combination of structured, semi-structured and unstructured data, often in raw form, until it is needed for specific analytics uses. Variables can be classified into nominal, ordinal, interval and ratio depending on the measurement type. Nominal measurement measures items based on their labels or categories or other qualitative classification the items belong to with no implied order. The ordinal measurement arises from the operation of rank ordering. The interval measurement allows us to measure the degree of difference between items, but not the ratio between them. Ratio measurements estimate the ratio between a magnitude of a continuous quantity and a unit magnitude of the same kind. Data can also be classified into discrete versus continuous. Discrete data are a numerical type of data that include whole, concrete numbers with specific and fixed data values determined by counting. Continuous data include complex numbers and varying data values that are measured over a specific time interval.

Exploratory data analysis methods such as visual examination, summary statistics and data distribution plots can be used to identify the data anomalies and a combination of knowledge and experience can be used to deploy the most appropriate treatment method depending on the situation at hand.

Activity 3.1

Prepare your answers to the following questions, as your instructor will ask you to present your answers in front of the class:

1. Explain the four basic stages of the data visualisation process.
2. Select any dataset from among the class demonstration datasets in Canvas. For the dataset selected, answer the following questions:

   a. Identify the dataset type with explanation— record data, graph data or sequential data?
   b. Briefly explain the variable hierarchy, data granularity and aggregation as seen in raw data.
3. Run exploratory data analysis on the selected dataset.
   a. Identify the issues with the data.
   b. Apply appropriate treatment methods for data preparation.

Formative Assessment

1. Taking the example of “Car”, suggest the characteristics of “Car” that will belong to each of the variable types.

Solutions or Suggested Answers

Formative Assessment

1. Taking the example of "Car", suggest the characteristics of "Car" that will belong to each of the variable types.
    1. **Nominal - classifies entities based on their labels or categories – colour**
    2. **Ordinal - orders entities based on rank – quality rating**
    3. **Interval - measures the degree of difference between entities- year of manufacture**
    4. **Ratio – measures the equality of ratios for continuous variables – engine horsepower**

References

Alton, L. (n.d.). The 7 Biggest Problems in Data Storage – and How to Overcome Them. *SmartData Collective*. https://www.smartdatacollective.com/7-biggest-problems-data-storage-overcome.

Bhanot, P. (2021, July 28). Six Essential Data Preparation Steps for Analytics. *Actian*. https://www.actian.com/blog/data-integration/the-six-steps-essential-for-data-preparation-and-analysis.

C3 AI Suite. (n.d.). *Data Granularity*. https://c3.ai/glossary/features/data-granularity.

Cote, C. (2021, December 02). 7 Data Collection Methods in Business Analytics. *Business Insights*. https://online.hbs.edu/blog/post/data-collection-methods.

Gupta, T. (2019, August 10). Types of Data Sets in Data Science, Data Mining & Machine Learning. *Towards Data Science*. https://towardsdatascience.com/types-of-data-sets-in-data-science-data-mining-machine-learning-eb47c80af7a.

Kononow, P. (2018, September 16). What is Metadata (with examples). *Dataedo*. https://dataedo.com/kb/data-glossary/what-is-metadata.

Krunal. (2020, January 24). What is Data Aggregation and How it is Useful. *AppDividend*. https://appdividend.com/2020/01/24/what-is-data-aggregation-examples-of-data-aggregation.

Michell, J. (1997). Quantitative science and the definition of measurement in psychology. *British journal of Psychology, 88*(3), 355-383.

Michell, J. (1999). *Measurement in psychology: A critical history of a methodological concept* (Vol. 53). Cambridge University Press.

Stevens, S. S. (1946). On the theory of scales of measurement. *Science, 103*, (2684), 677-680.

UNICOM Intelligence (n.d.). *Working with Hierarchical Data and understanding grid tables*. https://support.unicomsi.com/manuals/intelligence/75/index.html#page/Server%20User%20Guides/WorkWithHierarchicalData.20.17.html.

VIC Government. (n.d.). *Data Collection Challenges and Improvements*.https://www.vic.gov.au/victorian-family-violence-data-collection-framework/data-collection-challenges-and-improvements.

Ware, C. (2013). *Information visualization: Perception for design (Interactive technologies)* (3rd ed.). Waltham, MA, USA: Elsevier Inc.

Yau, N. (2013). *Data points: visualization that means something*. Indianapolis, IN, USA: John Wiley & Sons Inc.

Zangre, A. (2019, July 10). Discrete vs Continuous Data – What's the Difference?. G2. https://www.g2.com/articles/discrete-vs-continuous-data

Study Unit 4

Data Visualisation for Discrete and Continuous data

Learning Outcomes

At the end of this unit, you are expected to:

- Determine appropriate data visualisation techniques based on given data

Overview

This unit introduces several techniques for building data visualisation for categorical and time series data. For categorical data, we discuss the appropriateness of bar charts, stacked bar charts, side-by-side bar charts, pie charts, heat maps and tree maps vis-à-vis the visualisation task at hand. Time series data can be categorised as either discrete (bucketed) or continuous (unbroken). One consideration when presenting time series data is how we treat the discrete and continuous time series data. For the time series data, depending on the visualisation problem, we study the suitability of each of the line charts (with trend and reference lines), stacked area charts and gantt charts. We learn to create all these charts using a problem solving framework in data visualisation software.

Chapter 1: Visualisation for Categorical Data

Lesson Recording

Data Visualisation for Categorical and Time Series Data

We have learnt in Study Unit 2 that designing good data visualisation is surely an art, but just as surely, it is one that ought to be supplemented by science. Over the last decade, there has been a modest revolution in the analysis of categorical data. Graphical methods and techniques of data visualisation are so commonly used for quantitative data that the development for such techniques to be applied to frequency data and discrete data has begun.

1.1 Best Practices to Visualise Categorical Data

One-way frequency tables may be conveniently displayed in a variety of ways. As typical as bar charts, which are often ordered by frequency rather than by bar-label, dot charts (Cleveland, 1993) or pie charts can be used when the percentage of the chart expression's total is important.

For example, a sales manager at an apparel company like Charles and Keith in Singapore may have shoe sales data as shown in Figure 4.1 below. The figure shows the shoe sales data for the various categories of shoes for the past one year.

Shoe Category	Sales($'000)
Ballerinas	73
Boots	48
Flats	23
Heels	70
Loafers	35
Mary Janes	86
Mules	82
Pumps	36
Sneakers	18
Wedges	56

Figure 4.1 Shoe sales data

It will be challenging to imagine how the sales trend looks like if the sales manager uses only the data presentation above. Thus, the sales manager can use a bar chart to visualise the data, as shown in Figure 4.2. Each rectangle represents a category. The longer the rectangle is, the greater the value it represents. Whether a higher value means better or worse depends on the dataset and the analysis of the point of view.

In this example, the higher the value is the better the sales performance is. We can see that the bar graphs are restricted in a way that each bar must start at the zero-axis and must extend straight across or upward to the corresponding value. Do take note that we should avoid using a line chart to depict categorical data.

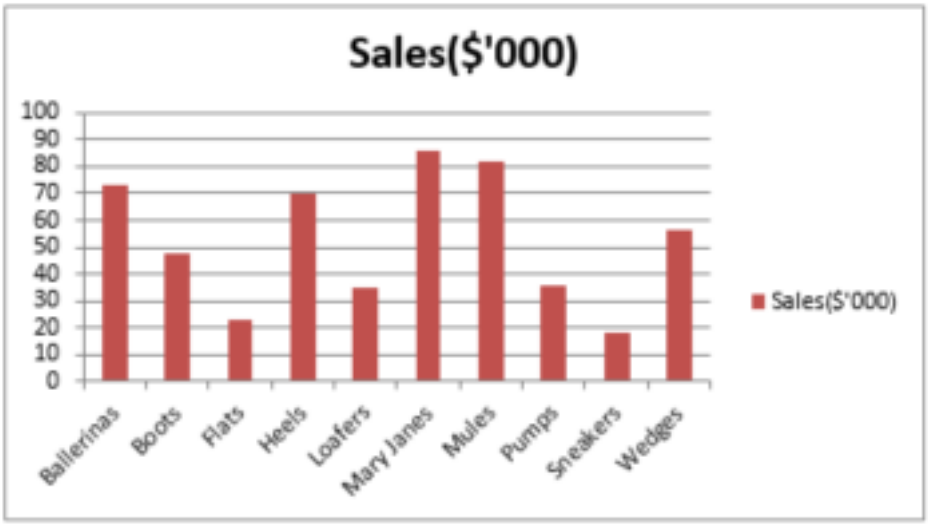


Figure 4.2 Using a bar chart to visualise the shoe sales data

When we put categorical data together, the sum of the parts can be equal to the whole. Using the same business scenario, a sales manager would like to view the sales data for each shoe category versus the aggregate sales amount. Visualising categories as a single unit can be beneficial if the sales manager wants to see distributions or the spread across a single population. This is where the pie chart, as illustrated in Figure 4.3, comes into the picture.

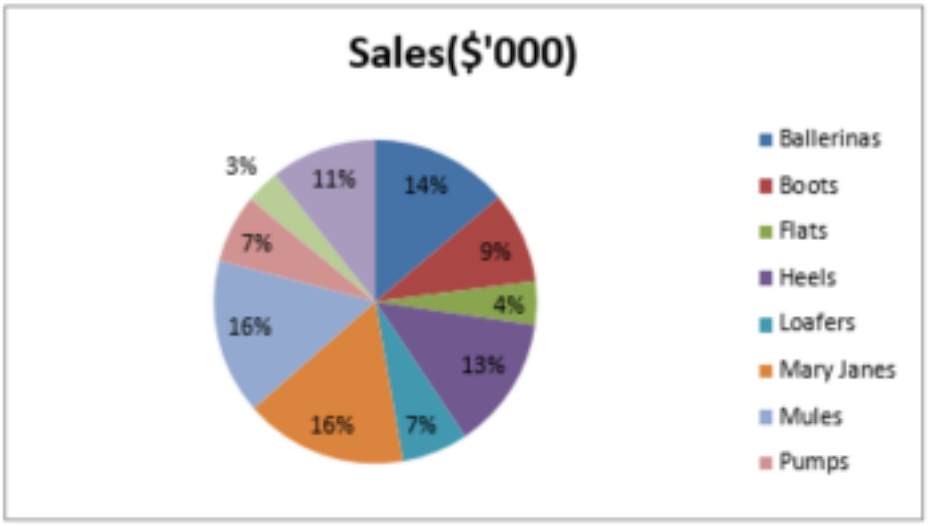


Figure 4.3 Using a pie chart to visualise ten-year sales data

A full circle represents 100%, and each wedge is a portion of that 100%. The sum of all wedges equals 100%, and the angle is the visual cue. Pie charts tend to clutter quickly when we have a lot of categories, simply because there is only so much space in the circle and small values end up as slivers.

Sometimes, we may encounter categorical data within categories. This type of categorical data is called *sub-categorical data.* They are often more revealing than main categorical data. As we drill down from main category to sub-category, there may be higher variability and more interesting insights to visualise.

Using the same business scenario, a sales manager would like to visualise the sales performance on per financial year basis, and he would like to further drill down to the product that has the highest sales amount in a particular financial year. He can use a tree-map chart, as illustrated in Figure 4.4, to visualise the sub-categorical data within the categorical data.

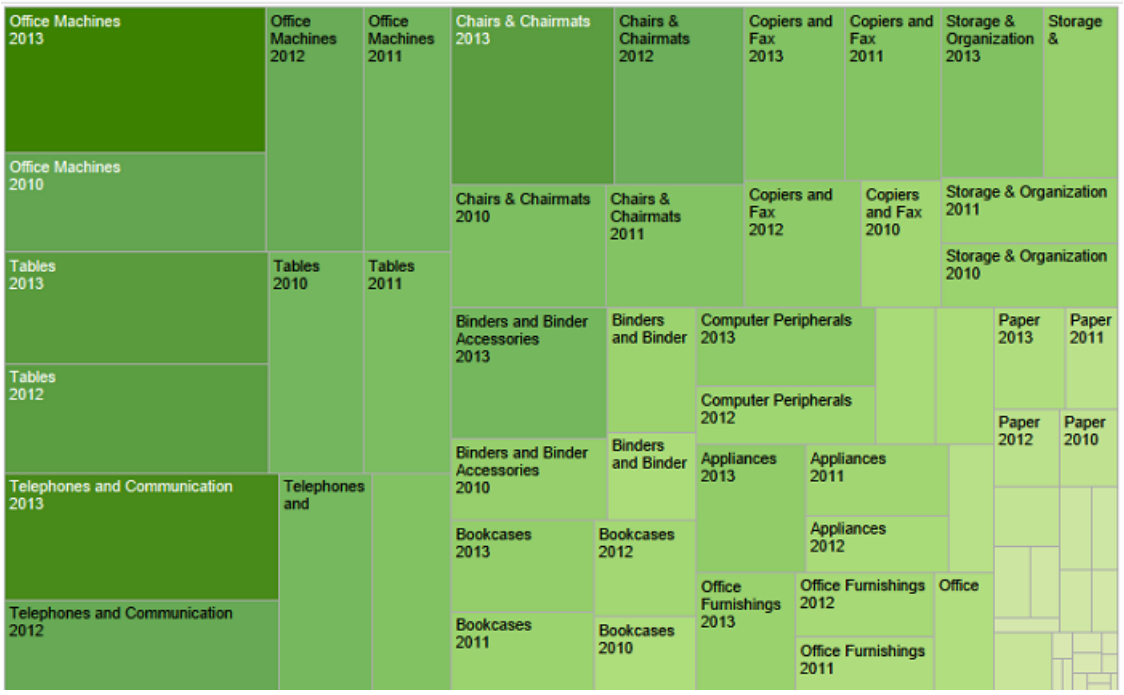


Figure 4.4 Using a tree-map chart to visualise sales data by year category and product sub-category

For each category sales value, we can look for both maximum and minimum values because they give us a sense of the data range, and they can be easily found by sorting the values in ascending or descending order. The table below summarises the most suitable chart type for each of the visualisation tasks listed in the first column. We will read about each one of them in detail in the following sections.

Table 4.1 List of chart type matched to the visualisation task for categorical data

Visualisation task	Chart type
Compare continuous variable (sales, profit, discount, quantity, cost, etc.) across different values of categorical data (product categories, brands, regions, markets, outlets, sales managers, etc.)	Bar chart
Compare continuous variable (sales, profit, discount, quantity, cost, etc.) across different values of categorical	Side-by-side / Stacked bar chart

Visualisation task	Chart type
data with next level of detail (product sub-categories, cities in regions, outlets within markets, salesman under sales managers, etc.)	
Contribution of each part (levels of categorical variable) to the total volume/value of continuous variable	Pie chart
See how the magnitude of continuous variable (sales, profit, discount, quantity, cost, etc.) varies (using colour intensity) over two dimensions (category and segment, category and sub-category, category and region, etc.)	Heat map
Display large amounts of hierarchically structured (tree-structured) data. (Total into categories into sub-categories into SKU, country into region into cities, sales head into area managers into salesman, cities into markets into outlets, etc.)	Tree map

1.1.1 Bar Chart, Side-by-Side Bar Chart, and Stacked Bar Chart

These charts facilitate one-to-many comparisons. The bar chart is the most effective way to compare values across dimensions due to their linear nature that makes precise comparisons easy. The side-by-side bar chart provides a way to compare measure across dimensions on a single axis. The stacked bar chart is an efficient way to show the share of each sub-category in the category total. In cases where there are too many sub-categories, too many colours will make the bar look very cluttered and difficult to interpret. As a rough rule of thumb, a stacked bar-chart is ideal for sub-categories below 7 or 8.

Let us say the sales manager of a company, like global superstore, would like to visualise how much sales happened for each product category sold. He/she can use a horizontal bar chart or a vertical bar chart in Tableau, as illustrated in Figures 4.5 and 4.6. Below are the steps to creating a bar chart in Tableau:

1. Create a new worksheet.
2. For horizontal bar charts, drag "Category" dimension into the worksheet's rows. For vertical bar charts, drag "Category" dimension into the worksheet's columns.
3. For horizontal bar charts, drag one or more measures ("Sales" in this case) into the worksheet's columns. For vertical bar charts, drag one or more measures ("Sales" in this case) into the worksheet's rows.
4. Drag and drop "Category" dimension on the "Color" marks card shown on the left in Figures 4.5 and 4.6.
5. Drag and drop "Sales" measure on the "Label" marks card.

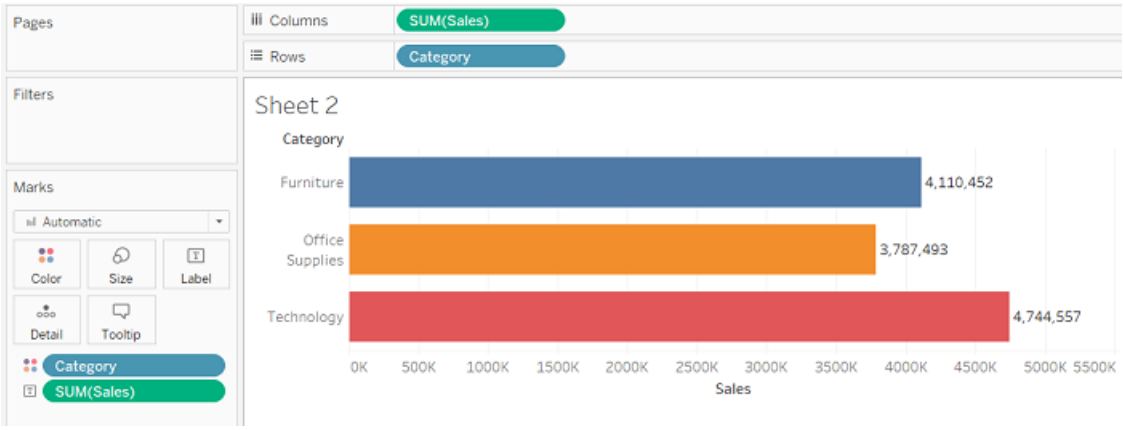


Figure 4.5 Generating a horizontal bar chart using Tableau

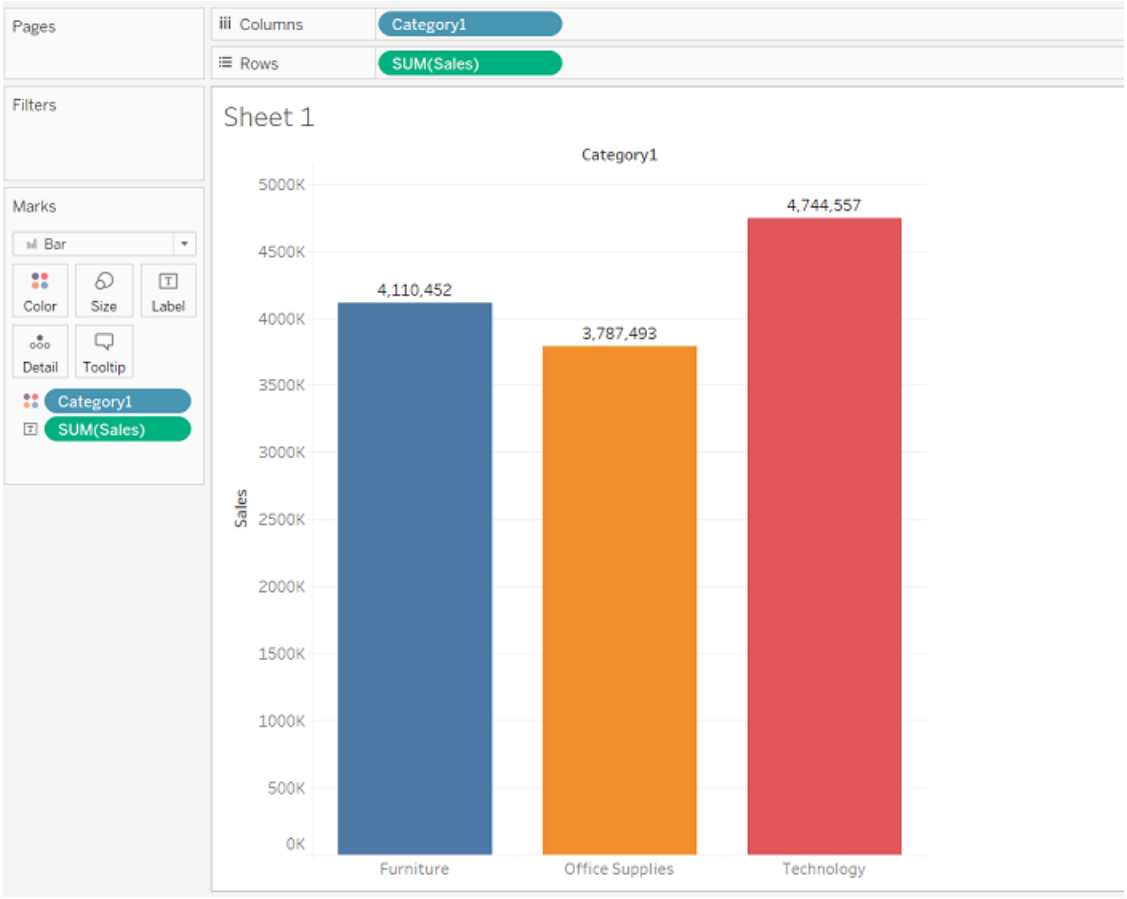


Figure 4.6 Generating a vertical bar chart using Tableau

Read

Here are some optional readings to serve as additional references for students who want to learn more about the topics just covered:

1. Murray. D. G. (2016), pp. 93-154.
2. Yau, N. (2013), pp. 97, 143.
3. Tableau Online Help - Bar Charts at https://help.tableau.com/current/pro/desktop/en-us/buildexamples_bar.htm

The same sales manager would now like to visualise the sales for each sub-category within the product category. He/she can use the horizontal side-by-side bar chart or the vertical side-by-side bar chart in Tableau, as illustrated in Figures 4.7 and 4.8. Below are the steps to creating a side-by-side bar chart in Tableau:

1. Create a new worksheet.
2. Drag the "Sales" measure into either the worksheet's columns (for side-by-side horizontal bar chart) or rows (for side-by-side vertical bar charts).
3. For horizontal side-by-side bar charts, drag "Category" and "Sub-Category" dimensions into the worksheet's rows. For vertical side-by-side bar charts, drag the "Category" and "Sub-Category" dimensions into the worksheet's columns.
4. Drag and drop the "Category" dimension on the "Color" marks card shown on the left in Figures 4.7 and 4.8.
5. Drag and drop "Sales" measure on the "Label" marks card.

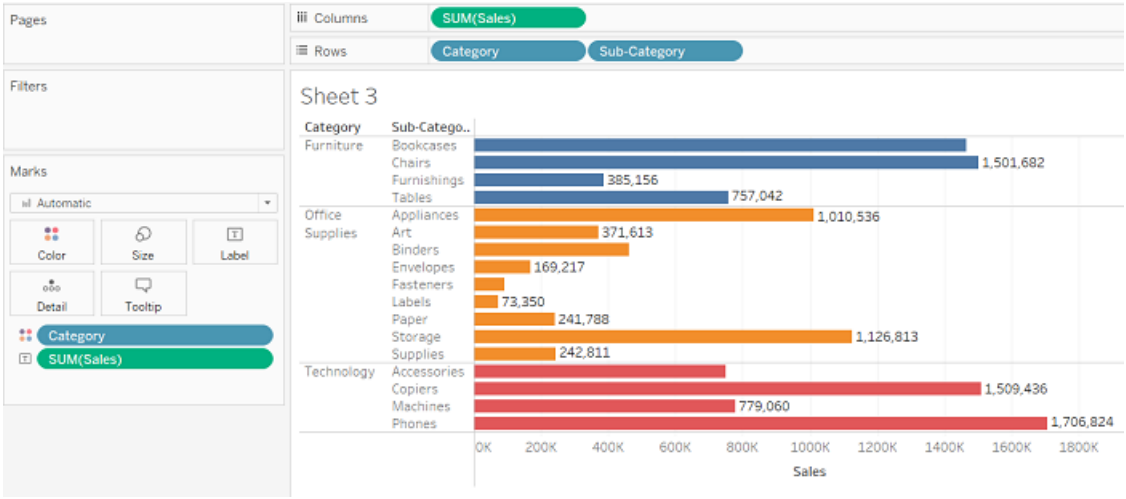


Figure 4.7 Generating a horizontal side-by-side bar chart in Tableau

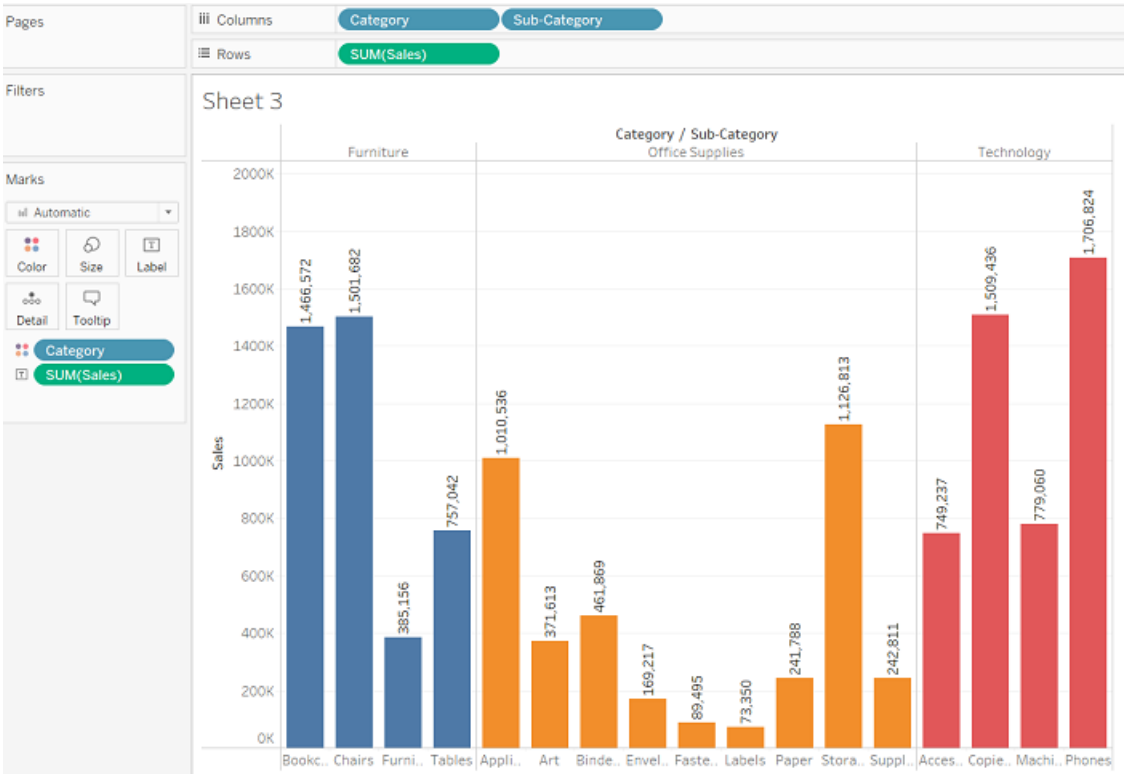


Figure 4.8 Generating a vertical side-by-side bar chart in Tableau

Read

You should now read:

1. Tableau Online Help – Side-by-Side Bar Charts athttps://kb.tableau.com/articles/howto/creation-of-a-grouped-bar-chart

The same sales manager would like to visualise how much each sub-category contributed to the total sales of the product category. He/she can use horizontal stacked bar charts or vertical stacked bar charts in Tableau, as illustrated in Figure 4.9. Below are the steps to creating a vertical stacked bar chart in Tableau:

1. Create a new worksheet.

2. For a vertical stacked bar chart, drag the ”Category” dimension into the worksheet’s columns.
3. For a vertical stacked bar chart, drag the ”Sales” measure into worksheet’s rows.
4. Drag and drop “Sub-Category” dimension on the “Color” marks card as shown on left in Figure 4.9.
5. Drag and drop “Sales” measure on the “Label” marks card.

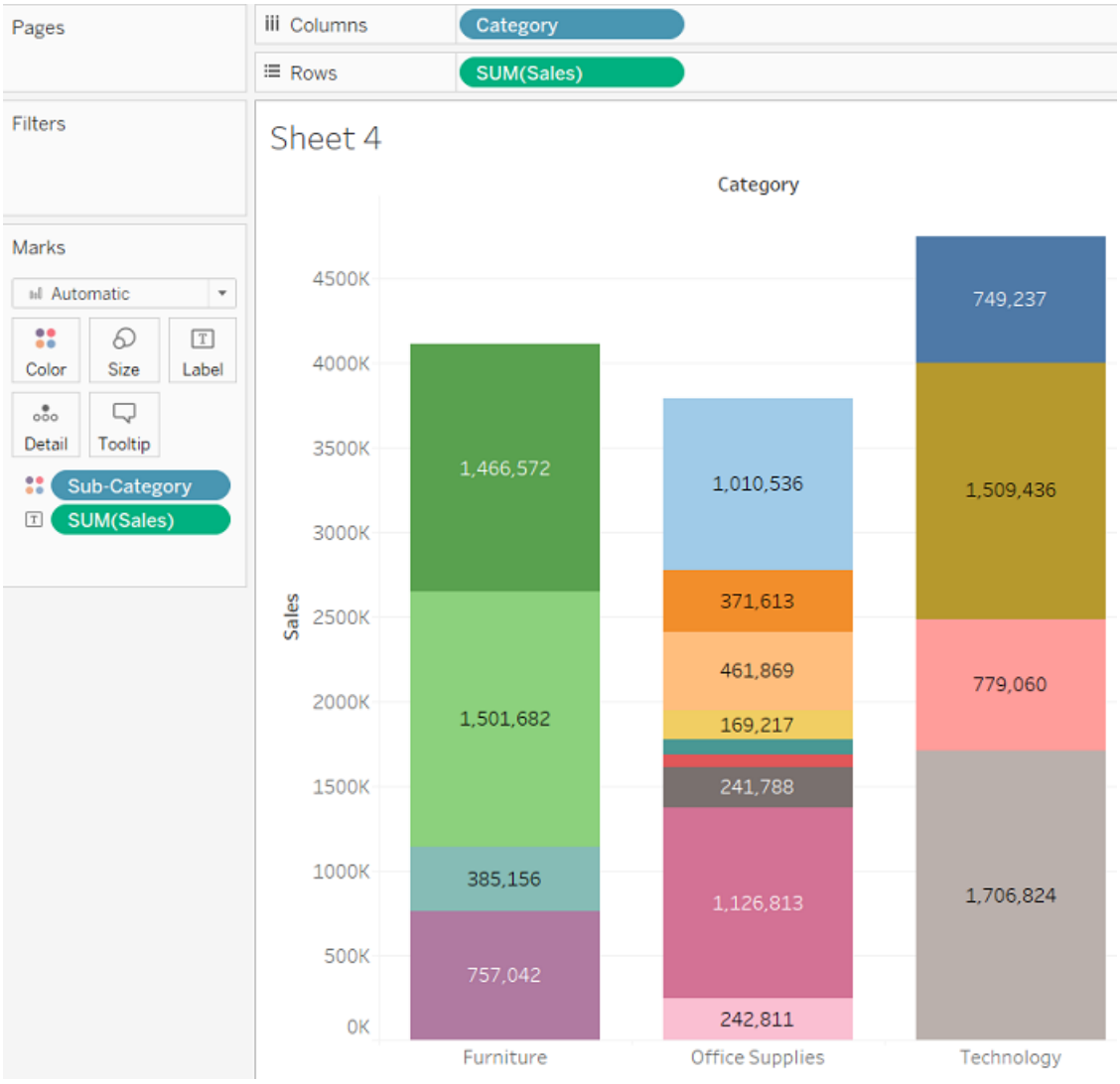


Figure 4.9 Generating a vertical stacked bar chart in Tableau

You should now read:

1. Yau, N. (2013), pp. 146.

2. Tableau Online Help – Stacked Bar Charts at http://kb.tableausoftware.com/articles/howto/stacked-bar-chart-multiple-measures.

1.1.2 Pie Chart

Pie Charts should be used when you want to visualise the contribution/share of each part to the total. It should be used to get a general sense of magnitude, but not for precise comparisons. Pie Chart is a static chart, meaning it gives the share of the part in the total only for a specific time frame. If you want to visualise how this share changes across years, one may have to plot multiple pie charts - one for each year. For such purposes, one may also consider stacked bar charts placed side by side for each year or one of the other visualisations discussed in the section ahead for time-series data.

Using the same example when the sales manager would like to visualise the contribution of each product category sales to total sales of the company. He/she can use a pie chart in Tableau as illustrated in Figure 4.10. Below are the steps to creating a pie chart in Tableau:

1. Create a new worksheet.
2. Drag and drop"Category" dimension on the "Color" marks card.
3. Drag and drop "Sales" measure on the "Size" marks card.
4. Choose "Pie Chart" as the chart type.
5. Drag and drop "Sales" measure on the "Angle" marks card.
6. Drag and drop "Category" dimension on "Label" marks card.
7. Drag and drop "Sales" measure on the "Label" marks card. Click the drop down arrow on the "Sales" tab added to "Label" marks card in above step and select "Quick table calculation" -> "Percent of Total" from the drop down menu.

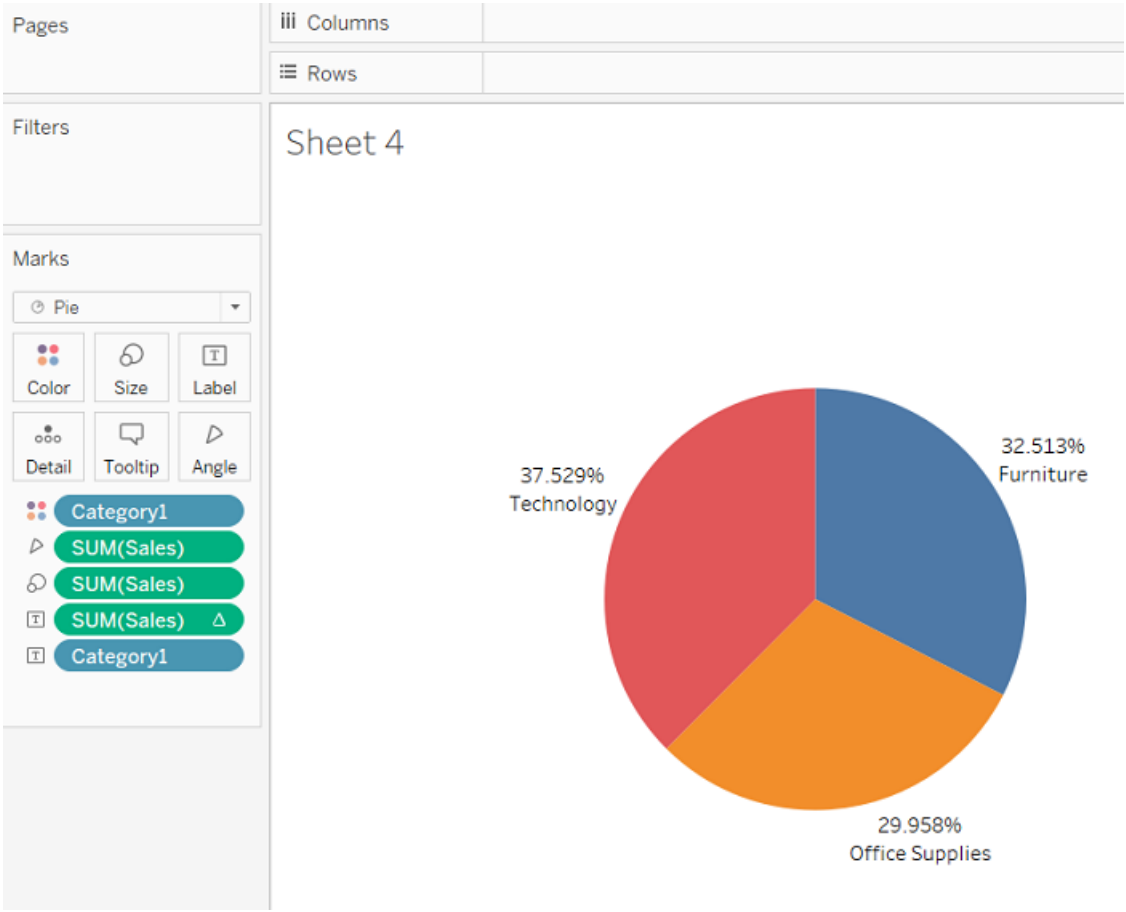


Figure 4.10 Pie chart settings in Tableau

Read

You should now read:

1. Yau, N. (2013), pp. 98.
2. Tableau Online Help – Pie Charts at https://help.tableau.com/current/pro/desktop/en-us/buildexamples_pie.htm

1.1.3 Heat Map and Treemap

Comparing granular combinations of dimensions and measures can be done effectively with heat maps and treemaps. Heat maps use colours and sizes to compare up to two measures. Treemaps effectively display larger dimension sets using colours and sizes to display one or more dimensions, and up to two measures.

A heat map is a data visualisation technique that shows the magnitude of a phenomenon as a colour in two dimensions. The variation in colour may be by hue or intensity, giving obvious visual cues to the reader about how the phenomenon is clustered or varies over space.

When the sales manager would like to visualise the distribution of each product category sales across consumer segments using colours and sizes, he can use a heat map in Tableau, as illustrated in Figure 4.11. Below are the steps to creating a heat map in Tableau:

1. Create a new worksheet.
2. Drag ”Category” dimensions into worksheet’s rows.
3. Drag ”Segment” into worksheet’s columns.
4. Drag and drop “Sales” measure on the “Color” marks card.

From the heatmap, we can see that the office supplies category sales lagged behind furniture and technology sales across consumer segments. But this could be due to the category size itself— hence for more conclusive evidence, it is recommended that one looks at the profitability of each of the categories across segments rather than absolute sales numbers.

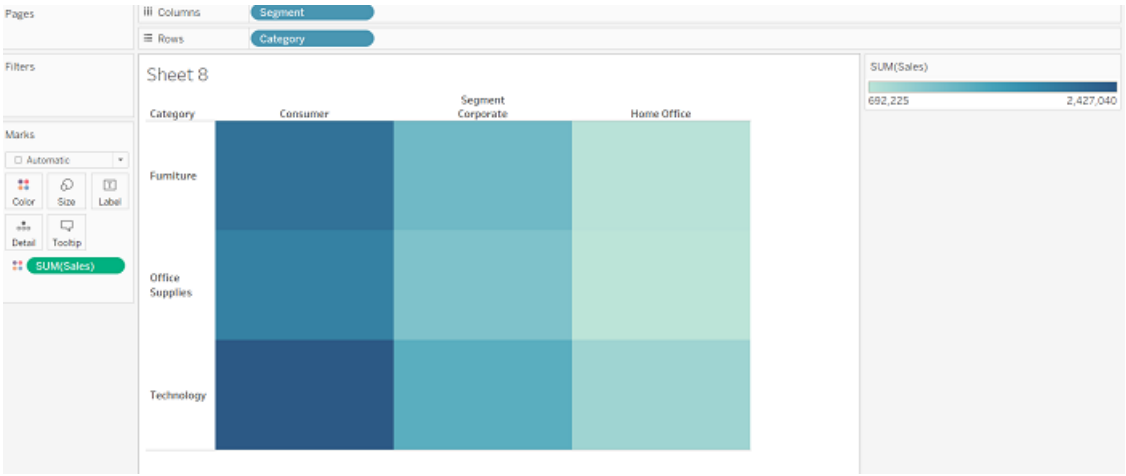


Figure 4.11 Creating a heat map in Tableau

Read

You should now read:

1. Yau, N. (2013), pp. 164, 182-184, 195.
2. Tableau Online Help – Heat Maps Chart at https://help.tableau.com/current/pro/desktop/en-us/buildexamples_highlight.htm

Treemaps are ideal for displaying large amounts of hierarchically structured (tree-structured) data. The space in the visualisation is split up into rectangles that are sized and ordered by a quantitative variable.

When the sales manager of the company would like to visualise the contribution of each product category sales to total sales and also each sub-category sales to category sales using colours and sizes, he/she can use a treemap in Tableau as illustrated in Figure 4.12. Below are the steps to creating a treemap in Tableau:

1. Create a new worksheet.
2. Drag "Category" and "Sub-Category" dimensions into the worksheet's rows.

3. Drag "Sales" measure into the worksheet's columns.
4. Choose Treemaps as the chart type.
5. Drag and drop "Category" dimension" on the "Color" marks card.
6. Drag and drop "Sales" measure on the "Label" marks card.

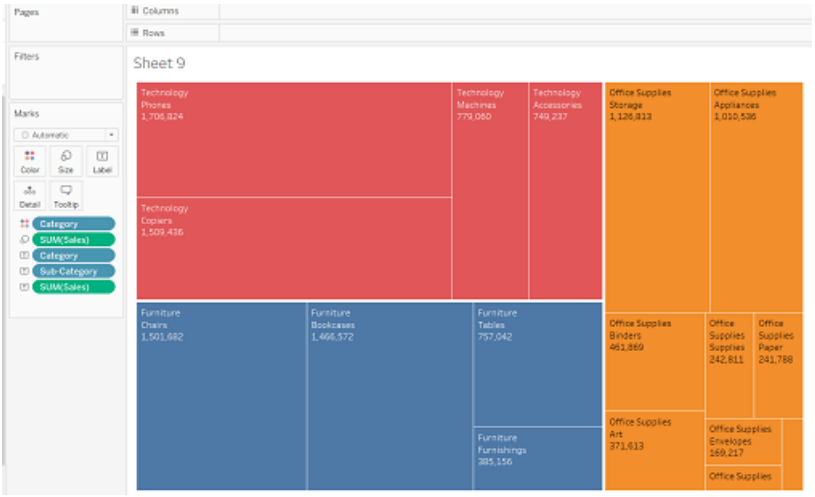


Figure 4.12 Treemap settings in Tableau

Read

You should now read:

1. Yau, N. (2013), pp. 53, 57, 146, 150, 281.
2. Tableau Online Help – Treemaps Chart at https://help.tableau.com/current/pro/desktop/en-us/buildexamples_treemap.htm

Chapter 2: Visualisation for Time Series Data

We look at time every day. The most common thing we look for in time series or temporal data is trends, e.g. is something increasing or decreasing, is this a seasonal cycle, etc. To find the patterns, we have to look beyond individual data points to get the whole picture. It is easy to pick out a single value from a point in time, but when we look at what came before and what comes after, we can have a better understanding of what that single value means. Furthermore, the more we know about our data, the better the insight that we can gain.

2.1 Best Practices to Visualise Time Series Data

Time series data can be categorised as either discrete (bucketed) or continuous (unbroken). One consideration when presenting time series data is how we treat the discrete and continuous time series data.

2.1.1 Discrete Time Series Data

For discrete time series data, values are from specific points or blocks of time, and there is a finite number of possible values. For example, the number of product sold last year is discrete. Last year's sales figure will not change afterward. Something like the currency exchange rate, however, is continuous, because it can be measured at any time of the day during any interval, and the rate is constantly changing.

A line chart is the most effective way to display time series data. A line chart for discrete time series data places breaks between time units like year, quarter, month and day.

Bar and stacked bar charts can also be used to display time series data. The x-axis of the bar and stacked bar chart (or the time axis) provides a place for points in time that are ordered chronologically. The y-axis of the bar and stacked bar chart (or the value axis) indicates the scale of the graph. Always start the value axis at zero, otherwise the bar or stacked bar chart could display incorrect relationship.

Sometimes, points can also be used to display time series data instead of bars because points use less space and there are no bins. Points can also allow dashboard viewers to feel that there is flow from one point to another. This type of point chart is commonly known as a scatter plot chart. It can be used to display non-time series data. A scatter plot chart uses position as the visual cue, so that we can compare each point in time to other points in time based on where they are placed in the x- and y-coordinates. Because of this, the value axis of a scatter plot chart does not always have to start at zero, even though it is usually a good practice.

2.1.2 Continuous Time Series Data

Visualising continuous time series data is similar to visualising discrete time series data. We still have a discrete number of data points, even if the dataset is continuous. The difference between visualising continuous time series data from visualising discrete time series data is in what they represent in the physical world. Continuous time series data represents a constantly changing phenomena.

The line chart for continuous time series data is presented as unbroken lines. We have the points on x- and y-coordinates, and the line will connect the points to help us see the trend in our data. It is usually a best practice to start the value axis at zero so that it will not affect the scale. How far we can stretch the x-axis can also affect the appearance of the trend. If we stretch it too far out, we may not see the pattern. If we squish the axis too much, the increase from point to point may look more than it is.

One of the drawbacks of a standard line chart is that it implies steady change from one point to other points. That is about right with a measure like a country's population, but it will not be right for a case where a measure stays at a value for a long time and all of a sudden it declines or inclines. For example, the interbank interest rate may stay the same for months and suddenly drop within a day. For such continuous time series data types, we need to use a step chart. Instead of directly connecting one point to another, the line stays at the same value until there is a change, at which point it jumps up or down to the next value.

When we have a lot of data, or the data we have is *noisy*, it can be hard to spot trends and patterns. Thus, it will be easier if we can estimate the trend line. To estimate the trend line, we can draw a line that goes through where there is the most points, and minimise the summarised distance from these points to the fitted line. The most straightforward method is to create a straight fitted line using the basic slope intercept equation below:

$$Y = mx + b$$

where "m" refers to the slope of the fitted line, and "b" refers to the intercept of the fitted line to the y-axis. When the fitted line is not linear, we can use a statistical method created by William Cleveland and Susan Devlin (1988) called LOESS, or locally weighted scatterplot smoothing. This method will enable us to fit a curve to our data, because LOESS starts at the beginning of the data and takes small slices. At each slice, it estimates a low-degree polynomial for just the data in the slice. LOESS moves along the data, fits a bunch of tiny curves, and together the tiny curves form a single curve.

2.2 Creating Charts to Visualise Time Series Data

To create charts in Tableau, we first need to open a workspace page, and select the data connection. We shall use the global superstore sales data to illustrate the steps for building charts in this section. The table below summarises the most suitable chart type for each of the visualisation tasks listed in the first column. We will read about each one of them in detail in the following sections.

Table 4.2 List of chart types matched to the visualisation task for time-series data

Visualisation task	Chart type
Look at trends in continuous variable (sales, profit, cost, quantity, etc.) across time (treated as continuous)	Line chart
Look at trends in continuous variable (sales, profit, cost, quantity, etc.) across time (treated as continuous)	Stacked area chart

Visualisation task	Chart type
by categories/levels (brands, tiers, locations, etc)	
Compare/plot timing and duration of activities or events, usually for project planning, logistics, scheduling	Gantt chart
See patterns in data that can provide predictive value, by drawing a line that best fits the values in the visualisation.	Line chart with trend line
Compare actual data presented in the visualisation against targets, or identify outliers that may require our attention or additional analysis.	Line chart with reference line

2.2.1 Line Chart

The line chart is the most effective way to display time series data. Discrete time series data are presented with breaks between time units, while continuous time series data are presented in unbroken lines.

Let us say the sales vice president of the global superstore would like to visualise monthly sales trends between the years 2011 and 2015. He can use the line chart in Tableau, as illustrated in Figure 4.13. Below are the steps to creating a line chart in Tableau:

1. Create a new worksheet.
2. Drag ”Order Date” dimension into the worksheet’s columns.
3. Expand the x-axis into month and year by clicking on the “Order Date” dimension in the columns and selecting “Month year” option from the drop down menu.
4. Drag “Sales”measure into the worksheet’s rows.

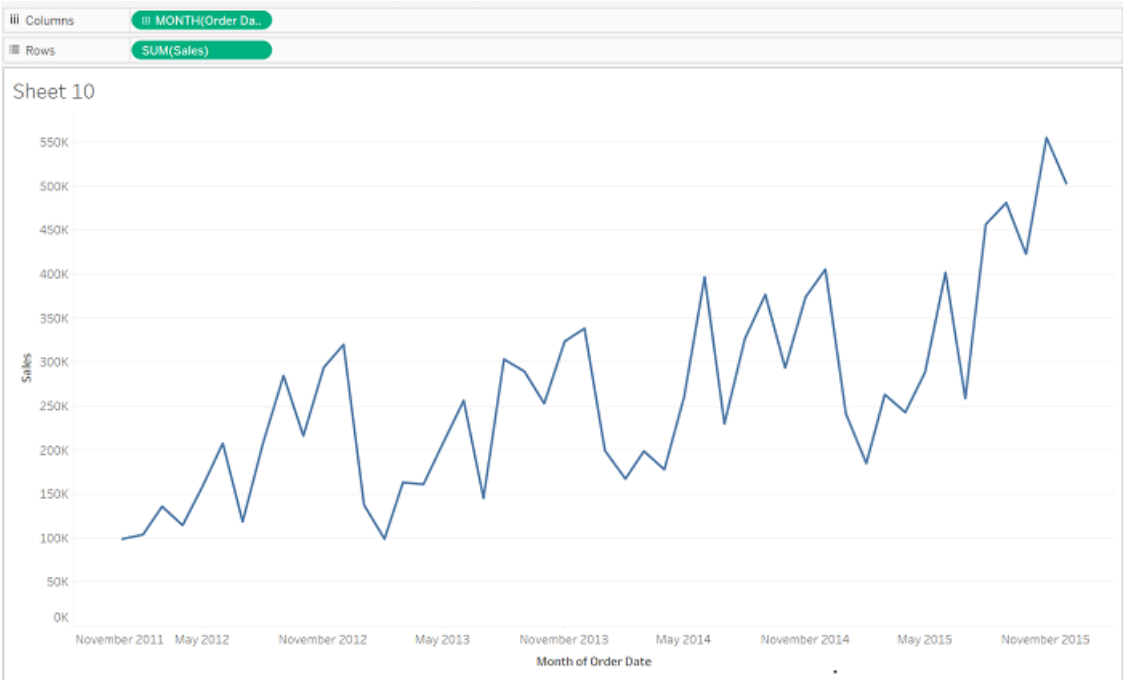


Figure 4.13 Creating a line chart in Tableau

Read

You should now read:

1. Yau, N. (2013), pp. 155, 159.
2. Tableau Online Help – Line Charts at https://help.tableau.com/current/pro/desktop/en-us/buildexamples_line.htm

To take a closer look at the sales trends, the sales vice president of the same global superstore would like to visualise monthly sales trends between the years 2011 and 2015 for each of the product category. This will also give him/her a fair idea of each category's performance. One of the common ways to visualise this is by drawing a different coloured line for each product category as shown below:

1. Create a new worksheet.

2. Drag "Order Date" dimension into the worksheet's columns.
3. Expand the x-axis into month and year by clicking on the "Order Date" dimension in the columns and selecting "Month year" option from the drop down menu.
4. Drag "Sales"measure into the worksheet's rows.
5. Drag and drop "Category" dimension on the "Color" marks card on left as shown in Figure 4.14.

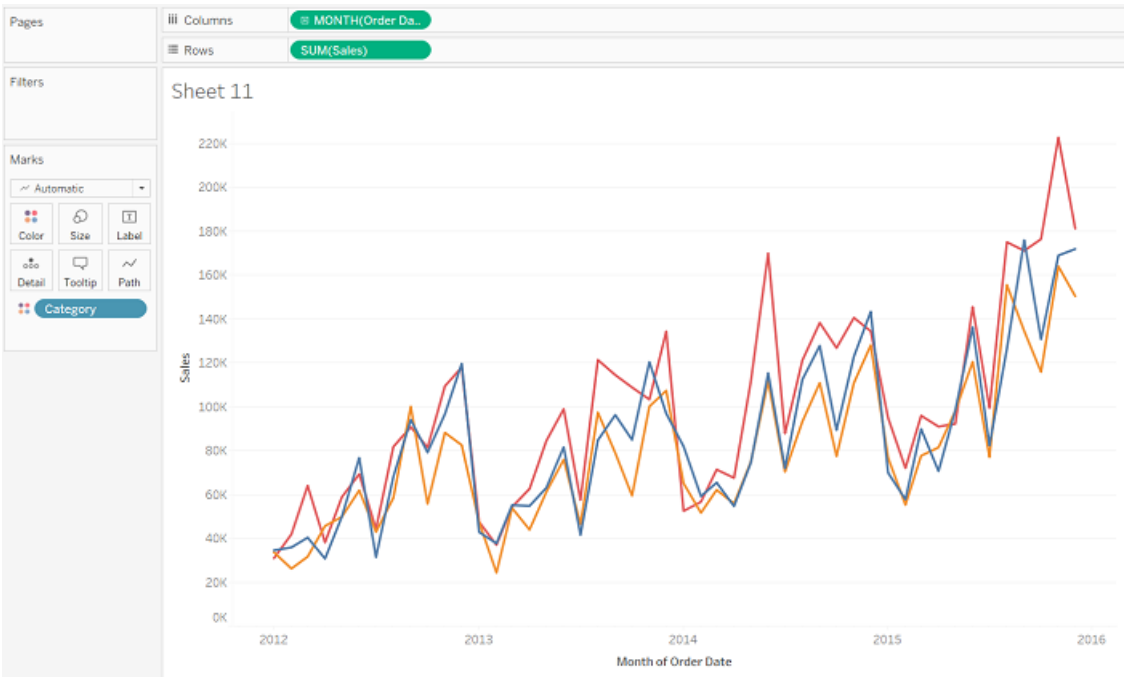


Figure 4.14 Creating multiple line charts in Tableau

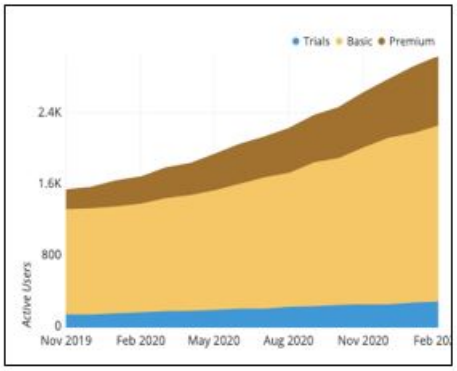
From Figure 4.14, it is difficult to contrast or compare the sales trend of each of the product category across months as the lines are overlapping and it looks visually cluttered. To overcome this drawback, another option is to use a stacked line/area chart. The stacked chart plots cumulative values of the measure on the vertical axis, taking one of the categories as the baseline and plotting every other category one above the next. This prevents overlapping of lines or data points and makes the chart visually easier to interpret.

2.2.2 Stacked Area Chart

Stacked area charts plot values as bands, thus it is easy to misinterpret the top band as being the largest value in a data set. The stacked area chart is a variation of the multi-line chart in that the lines do not cross each other and each coloured area is well demarcated from the other.

How to interpret stacked area charts?

Take an example of a service that has three categories of users based on subscription type — trials, basic and premium. The area chart depicts active number of users per month for each category. The boundary of the uppermost tier denotes the cumulative number of users on the vertical axis. Take the example of the month of November, 2019, where the blue area ends at 154, which is the number of users in the "Trials" category. The yellow area starts from 154 onwards and ends at 1334 on the vertical axis, which is the cumulative number of active users for the "Trials" and the "Basic" category. Hence the active users for the "Basic" category will be 1334-154=1180. A similar interpretation holds true for the "Premium" category active users in the month of November 2019, as seen from the stacked area graph shown in Figure 4.15.



MONTH	TRIALS	BASIC	PREMIUM
2019-11	154	1180	201
2019-12	157	1186	219
2020-01	170	1195	270
2020-02	180	1213	285

MONTH	TRIALS	+ BASIC	+ PREMIUM
2019-11	154	1334	1535
2019-12	157	1343	1562
2020-01	170	1365	1635
2020-02	180	1393	1678

Figure 4.15 Number of active users per month basis subscription type

Hence, to visualise the sales trend product category wise, a stacked area chart can be created as illustrated in Figure 4.16. Below are the steps to create a stacked area chart in Tableau:

1. Create a new worksheet.
2. Drag "Order Date" dimension into the worksheet's columns.
3. Expand the x-axis into month and year by clicking on the "Order Date" dimension in the columns and selecting "Month year" option from the drop down menu.
4. Drag "Sales"measure into the worksheet's rows.
5. Drag and drop "Category" dimension on the "Color" marks card on the left as shown in Figure 4.16.
6. Choose "Area Chart" as chart type.

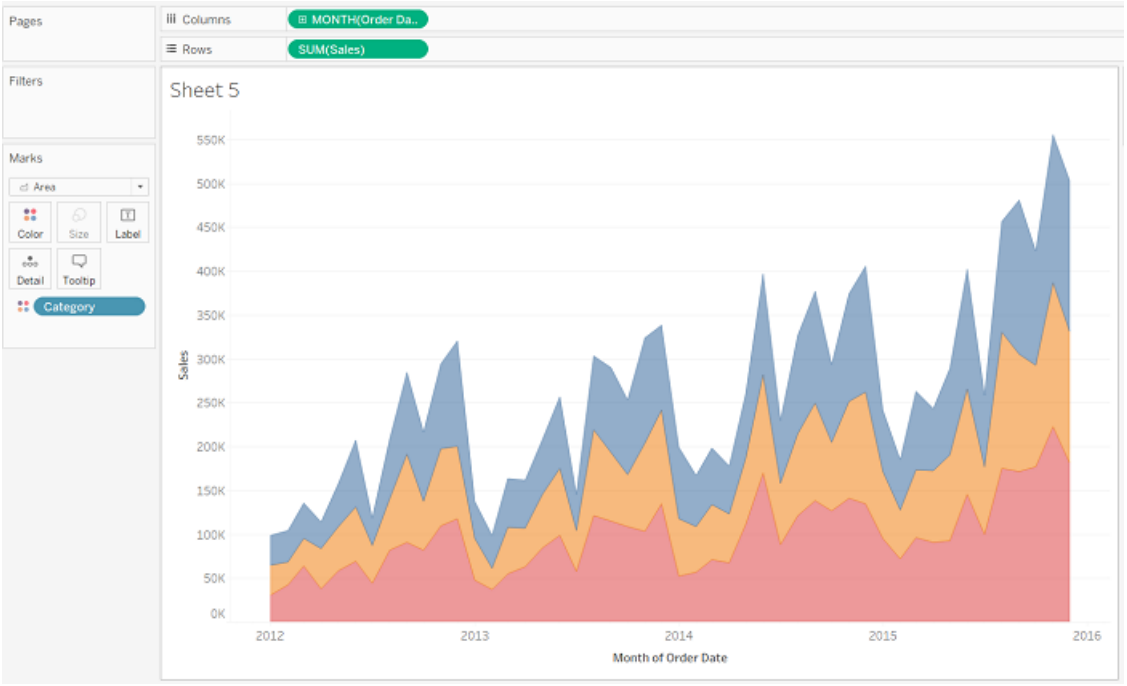


Figure 4.16 Creating a stacked area chart in Tableau

Read

You should now read:

1. Tableau Online Help – Area Charts at https://help.tableau.com/current/pro/desktop/en-us/qs_area_charts.htm

2.2.3 Gantt Chart

A Gantt chart is commonly being used for project planning. The activity or event start time is visualised by the bar's horizontal position, and the duration of each activity or event is visualised by the individual bar length. Thus, this chart is useful to visualise the timing and duration of activities or events.

Let us say that the supply chain manager of the global superstores would like to visualise the shipping lead time (time between order and shipping date) for each order in the month of April. He/she can use a Gantt chart in Tableau, as illustrated in Figure 4.17. Below are the steps to creating a Gantt chart in Tableau:

1. Create a new worksheet.
2. Select "Analysis", then "Create Calculated Field".
3. In the "Calculated Field" dialogue box, type "Duration in Days" in the name textbox and type "DATEDIFF" ('second', [Order Date], [Ship Date])/86400.
4. Drag the "Order Date" field into the worksheet's columns.
5. Right click on the "Order Date" field and select "Exact Date".
6. Drag the "Order ID" field into the worksheet's rows and into the "Colour Mark".
7. On the Mark Card's drop down menu, select "Gantt Bar".
8. Drag the "Duration in Days" field to "Size".
9. Right click on the "Order Date" field and select "Edit Axis".

10. In the "Edit Axis" dialogue box, under "Range", select "Fixed". Change the date range, and click "Apply".
11. On the "Tick Marks" tab, fix the major and minor tick marks to every one day. Change the origin date to April 1, 2014. When finished, click "OK".
12. Drag "Order Date" field into Filters Cards to filter the "Order Date" from April 1, 2014 to April 30, 2014.

From the Gantt chart, we can take a closer look at the lead time to shipping an order from the date the order was placed in the month of April. This helps us to understand if there was a supply chain disruption in the month of April, how much delay did it cause and which were the worst affected orders.

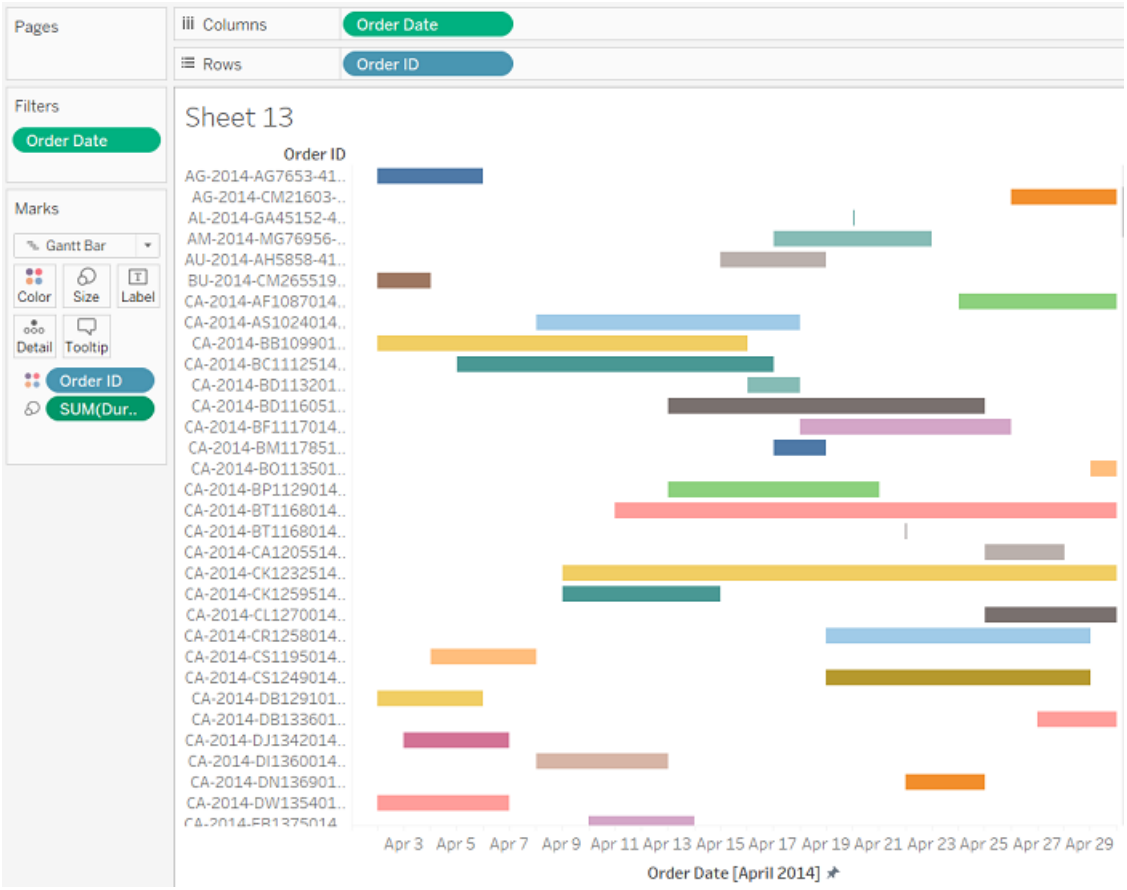


Figure 4.17 Creating a Gantt chart in Tableau

Read

You should now read:

1. Tableau Online Help – Gantt charts at

   https://help.tableau.com/current/pro/desktop/en-us/buildexamples_gantt.htm

2.2.4 Trend and Reference Lines

When we try to visualise granular data, sometimes it may result in random looking data visualisation. Trend lines can help us see patterns that can provide predictive value, by drawing a line that best fits the values in the visualisation. Reference lines can help us compare the actual data presented in the visualisation against targets, create statistical analyses of the deviation contained in the visualisation, or create the range of values based on fixed or calculated numbers. By looking at reference lines, we will be able to identify outliers that may require our attention or additional analysis.

Using the global superstore scenario, when the VP of sales of the company would like to see the year-on-year sales trend for each product category across consumer segments, he/she can use trend lines together with a line chart in Tableau, as illustrated in Figures 4.18. Below are the steps to creating trend lines in Tableau:

1. Create a new worksheet.
2. Drag “Category” and ”Order Date” dimension into the worksheet’s columns. From the drop down menu for “Order Date” dimension select “Year”.
3. Drag “Segment” dimension and “Sales” measure into the worksheet’s rows.
4. Right click on the chart and select “Show Trend Lines”.

5. For more the "Trend Lines" option, point at the "Trend Line", right click, and select "Edit Trend Lines". This will expose the Trend Line Menu. Select "Show Confidence Bands".

From the graph below, we can see that the trend line has a steepest slope for the "Consumer" segment as compared to the "Corporate" segment, followed by the "Home office" segment, for all three product categories. We can say that the rise in year-on-year sales for the "Consumer" segment is the highest, followed by the "Corporate" segment, and sales remained at the same level or showed marginal increment for the "Home office" segment.

Figure 4.18 Using trend lines in Tableau

Using the same scenario as above, when the VP of sales of the company would like to compare the actual sales against average sales in between the years 2012 and 2015, he/she can use reference lines together with a line chart in Tableau, as illustrated in Figure 4.19. Below are the steps to creating reference lines in Tableau:

1. Create a chart, for example, a line chart, by following the steps that was described in the line chart section previously.
2. Right click on the axis on which we want to apply the reference line, and select "Add Reference Line", "Band", or "Box".
3. Explore the line, band, and distribution buttons in conjunction with the computation value's dropdown menu to see all the available options for reference line types.

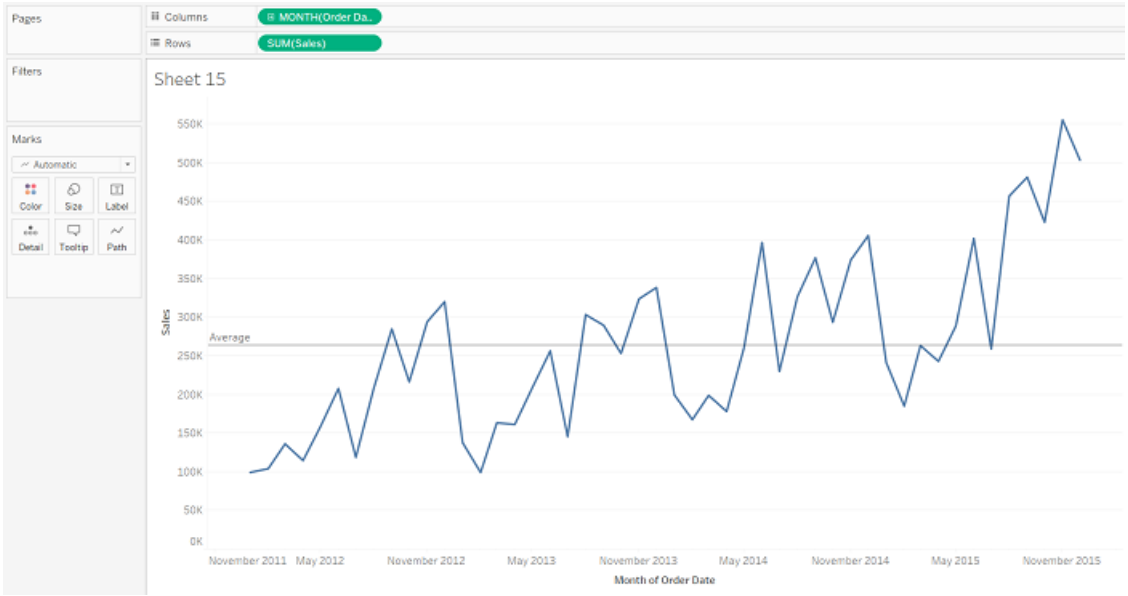


Figure 4.19 Creating reference lines in Tableau

Read

You should now read:

1. Tableau Online Help – Reference Lines at https://help.tableau.com/current/pro/desktop/en-us/reference_lines.htm
2. Tableau Online Help – Trend Lines at https://help.tableau.com/current/pro/desktop/en-us/trendlines_add.htm

2.3 Disaggregation and Aggregation of Data

Disaggregation or aggregation of data helps us to visualise data at different levels of detail depending on the information that is contained in the dataset. For example, in time-series data, time could be measured in years, months, weeks, days, or hours. For ordering transactions, it might be at the purchase order level, or line item level, or detailed configuration level for customised parts. Depending on the visualisation task, the x-axis can represent years, month or weeks and the line chart or bar graph can be broken down into line item level or detailed configuration level.

Tableau supports disaggregation and aggregation in the data through a click and select menu. For example, if the date is in MM/YYYY format in the data, upon selecting the time variable in Tableau, one may select either year or month as the unit on the x-axis from the drop down menu. Similarly, for displaying information at various sub-category levels, one may select the appropriate dimension from the menu on the left hand side column in Tableau.

Summary

More often than not in our daily lives, , we may like to group our business related data based on customer, region and other dimensions, and store these data as categorical data. One of the most common charts to visualise categorical data is the bar chart, including the side-by-side bar chart and the stacked bar chart.

We can use a pie chart to visualise data distributions across a single population when we only have one measure or use a stacked bar chart when we have more than one measure. Sometimes, we may encounter categorical data within categories. We can use a treemap chart to visualise the sub-categorical data within the categorical data.

Time is embedded in our day-to-day life; thus visualising time series data can be quite intuitive. By visualising time series data, we can understand how things change and evolve, and the challenge is to find out how much the change or evolution is, and to know what to look for in the visualisation.

In general, it is easy to spot the trend on a time series data visualisation, and to say that something is good or bad, as that is what data visualisation is meant for. Nevertheless, we should take this further by drilling down on sections of time and questioning ourselves on why there were drops or spikes on some days, but nowhere else.

After we understand our data well, we need to explain and highlight our insights so that data visualisation viewers will know which part of the visualisation they should pay more attention to. Data visualisation can be nice for its creator, but without context, the visualisation may be meaningless to everyone else.

Activity 4.1

1. Using the dataset of your choice, create the following:
    a. Bar chart, side-by-side bar chart, and stacked bar chart
    b. Pie chart
    c. Heat-map and treemap
    d. Line chart, stacked area chart, Gantt chart, reference line, and trend line

Formative Assessment

Select the most suitable chart option for each of the examples given below:

1. You need to compare sales for different brands. Which is the most suitable chart?
    a. Bar chart
    b. Tree map
    c. Line chart
    d. Area chart

2. Which is the most suitable chart to evaluate the share of each category in total sales?
    a. Bar chart
    b. Pie chart
    c. Heat map
    d. Area chart

3. You are given the task of preparing the project schedule indicating the number of days each stage will take from start to completion. Which chart will you pick to showcase the project schedule?
    a. Heat map
    b. Stack bar chart
    c. Line chart
    d. Gantt chart

4. There has been a dip in sales last month. A sales review is called, and you are supposed to start the presentation by showing the sales movement across previous quarters till date. Which chart will you pick?
    a. Tree map
    b. Side-by-side bar chart
    c. Line chart

d. Horizontal bar chart

5. In addition to sales for previous quarters, your manager also wants to compare it against the average sales for last year to understand how good or bad was the sales as compared to last year. Which is the most suitable chart for this?
    a. Pie chart
    b. Grantt chart
    c. Line chart with trend line
    d. Line chart with reference line

Solutions or Suggested Answers

Formative Assessment

Select the most suitable chart option for each of the examples given below:

1. You need to compare sales for different brands. Which is the most suitable chart?

    a. Bar chart

    Correct. Refer to Table 4.1 for explanation.

    b. Tree map

    Incorrect.

    c. Line chart

    Incorrect.

    d. Area chart

    Incorrect.

2. Which is the most suitable chart to evaluate the share of each category in total sales?

    a. Bar chart

    Incorrect.

    b. Pie chart

    Correct. Refer to Table 4.1 for explanation.

    c. Heat map

    Incorrect.

    d. Area chart

    Incorrect.

3. You are given the task of preparing the project schedule indicating the number of days each stage will take from start to completion. Which chart will you pick to showcase the project schedule?

    a. Heat map

    Incorrect.

    b. Stack bar chart

    Incorrect.

    c. Line chart

    Incorrect.

    d. Gantt chart

    Correct. Refer to Table 4.2 for explanation.

4. There has been a dip in sales last month. A sales review is called, and you are supposed to start the presentation by showing the sales movement across previous quarters till date. Which chart will you pick?

    a. Tree map

    Incorrect.

    b. Side-by-side bar chart

    Incorrect.

    c. Line chart

    Correct. Refer to Table 4.2 for explanation.

    d. Horizontal bar chart

    Incorrect.

5. In addition to sales for previous quarters, your manager also wants to compare it against the average sales for last year to understand how good or bad was the sales as compared to last year. Which is the most suitable chart for this?

a. Pie chart

Incorrect.

b. Grantt chart

Incorrect.

c. Line chart with trend line

Incorrect.

d. Line chart with reference line

Correct. Refer to Table 4.2 for explanation.

References

Cleveland, W. S. (1993). *Visualizing data*. Summit, NJ, USA: Hobart Press.

Cleveland, W. S., & Devlin, S. J. (1988). Locally weighted regression: an approach to regression analysis by local fitting. *Journal of the American statistical association, 83*(403), 596-610.

Few, S. (2015). *Signal: Understanding what matters in a world of noise.* Burlingame, CA, USA: Analytics Press.

Murray, D. G. (2016). *Tableau your data!: Fast and easy visual analysis with Tableau software.* (2nd ed.).Indianapolis, IN, USA: John Wiley & Sons Inc.

Tableau Online Help. (2014). Retrieved July 1, 2014, from http://onlinehelp.tableausoftware.com/current/pro/online/http://onlinehelp.tableausoftware.com/v6.1/public/online/

Yau, N. (2013). *Data points: Visualisation that means something*. Indianapolis, IN, USA: John Wiley & Sons.

Study Unit 5

Business Performance Dashboard

Learning Outcomes

At the end of this unit, you are expected to:

- Explain the end-to-end process of creating business performance dashboard from data
- Develop data visualisation dashboard using a software package

Overview

This unit introduces the concept of business performance dashboard- a visual display of the most important information needed to achieve one or more objectives, consolidated and arranged on a single screen so that the information can be monitored at a glance. In addition, it also discusses the three types of dashboards - strategic dashboard, tactical dashboard, and operational dashboard. We also demonstrate how to create an impactful business performance dashboard using the data visualisation software and following the effective dashboard design principles in a problem-solving framework.

Chapter 1: Introduction to the Business Performance Dashboard

Lesson Recording

Business Performance Dashboard

A business performance dashboard can be defined as a visual display of the most important information needed to achieve one or more objectives, consolidated and arranged on a single screen so that the information can be monitored at a glance (Few, 2004). The original idea of the business performance dashboard is taken from a vehicle dashboard, which provides all the measures that are critical to the operation of the vehicle, such as temperature and speed. Similarly, a business performance dashboard serves to provide all the pertinent measures needed to drive the performance of an organisation.

A *business performance dashboard* provides useful information that enables managers to mount improvement activities on areas that show performance below expectation. It serves as a critical performance reporting function in the strategic performance management model. It acts like a magnifying glass to focus an organisation's attention on deficiencies that impede the achievement of the organisation's overall strategy.

Eckerson (2006) found strong interest in business performance dashboards in a survey among IT professionals. At least 43% of the 689 respondents have implemented or are in the process of implementing business performance dashboards. According to the profile of the respondents, this interest is not only limited to those in the manufacturing industry, but also cuts across all industries, including the government sector. The basic functionalities of a business performance dashboard according to Eckerson are to:

- monitor critical business processes and activities, using the relevant measures that track business performance; and

- manage people and processes to improve decision making, optimise performance, and steer the organisation in the right direction.

Figure 5.1 illustrates an award-winning business performance dashboard of an airline company, used to track the company's relevant performance (Few, 2006).

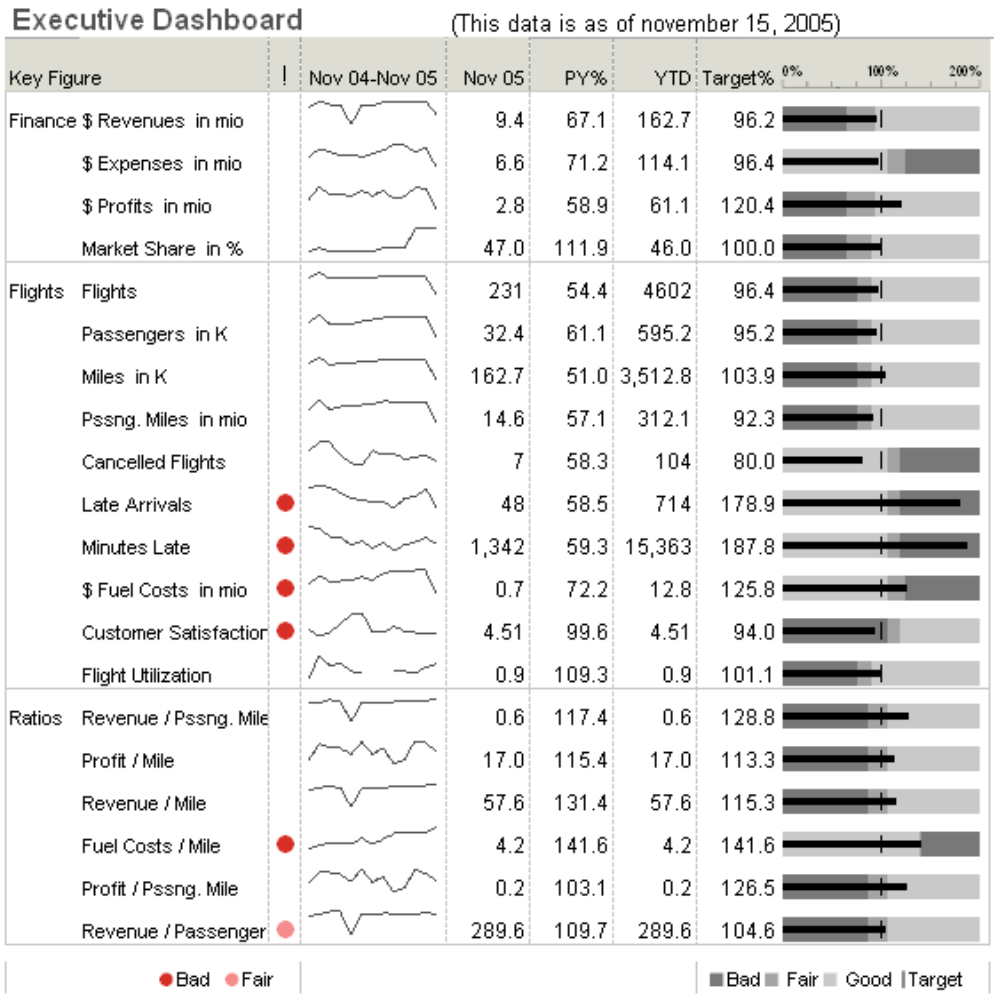


Figure 5.1 Business performance dashboard for an airline company

Business performance dashboards can generally be classified into three main types. Each type of dashboard caters to a different group of users. The three types of dashboards are strategic dashboard, tactical dashboard, and operational dashboard.

1.1 Strategic Dashboard

The *strategic dashboard* is used by executives and managers to monitor the execution of strategic objectives. These strategic objectives can be the results of using a balanced scorecard approach. The key features of the strategic dashboard are simplicity and high visual impact. An intuitive strategic dashboard is something that allows the top management, who often have tight schedules, to have a quick and fast look at the company's performance. If required, top management will be able to carry out further analysis by drilling down to the details of the respective object or to the underlying tactical dashboard.

Figure 5.2 illustrates an example of a strategic dashboard.

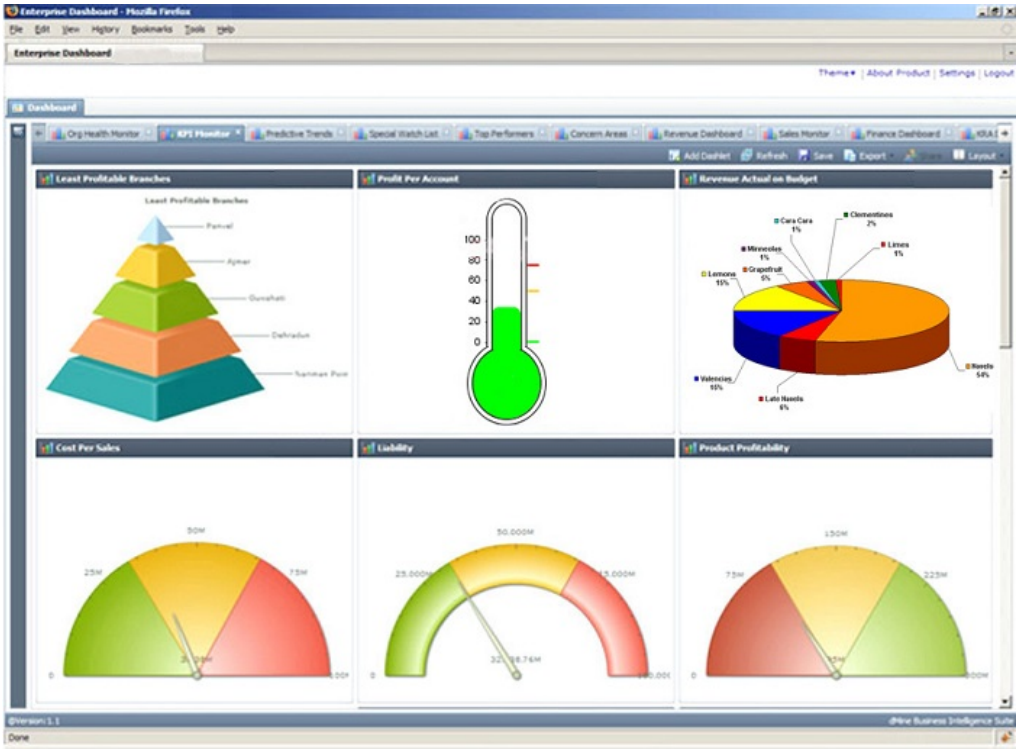


Figure 5.2 Strategic dashboard

1.2 Tactical Dashboard

A *tactical dashboard* is used by managers and analysts to track the progress of departmental processes, and projects against budget plans, forecasts or last period's achievement. A financial dashboard is an example of a tactical dashboard used by the finance department to track revenue and expenses versus the budgeted numbers on a monthly or quarterly basis. A typical user group of the tactical dashboard is business and financial analysts, where they use multi-dimensional charts, slice-and-dice of online analytical processing cube, etc.

Figure 5.3 illustrates an example of a tactical dashboard.

Figure 5.3 Tactical dashboard

1.3 Operational Dashboard

The *operational dashboard* is used by operational staff and their supervisors to monitor operational processes. In a sales department, the operational dashboard is used to monitor the performance of individual sales representatives in terms of sales achievement for

a specific period of time, which is usually monthly or quarterly. In the manufacturing industry, the operational dashboard is used to monitor the performance of manufacturing engineers and assistants in terms of units of output for a day. Such dashboards are very specific to a particular operation or process.

Figure 5.4 illustrates an example of the operational dashboard.

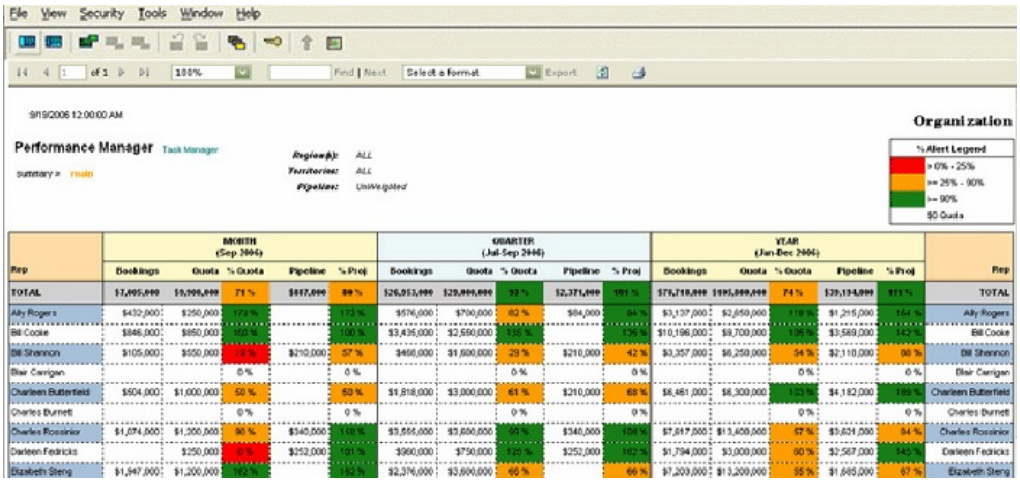


Figure 5.4 Operational dashboard

1.4 Benefits of Business Performance Dashboard

Organisations implement business performance dashboards because of the many benefits that can be reaped by everyone in an organisation. These benefits, according to Eckerson (2006), are as follows:

1.4.1 Communicate Strategy

Business performance dashboards facilitate the translation of corporate strategy into measures with their corresponding targets. Anyone who has access to the business performance dashboard will be aware of the organisation's strategic objectives and what needs to be done in his/her area to achieve these objectives.

1.4.2 Increase Insight

A business performance dashboard provides the management team with greater insights into business performance in a timely manner. This minimises being caught by surprise by unforeseen problems that can affect corporate bottom-line or results. Corporate strategy can then be promptly fine-tuned with minor corrections, when required, instead of veering off-course.

1.4.3 Increase Motivation

With the measures and corresponding targets publicly displayed in business performance dashboards, the motivation to excel in the measured areas is increased. Such motivation compels people to work harder out of pride, and the desire to achieve higher pay; especially when the pay moves in tandem with the measures. Knowing "what gets measured, gets done," transparency in performance measurement and results increases motivation to excel in the measured areas.

1.4.4 Increase Coordination

Objectivity, openness and transparency in the business performance dashboard help promote effective coordination, and encourage different departments to work more closely. Employees will be encouraged to engage in meaningful dialogue on performance results and forecasts. This allows managers to conduct more frequent and constructive performance reviews.

1.4.5 Consistent View of Business

The business performance dashboard consolidates and integrates business performance information using a common platform, definition, measures, and rules. A single version of such information minimises conflicts and disagreements.

1.4.6 Reduce Cost and Redundancy

Standardising information and reporting based on a business performance dashboard can eliminate redundant silos of information and duplication of reports. A well designed business performance dashboard is capable of replacing independent reporting systems, spreadsheet marts, data marts, and warehouses. This also helps streamline reporting and simplifies the organisation's information system.

1.4.7 User Sufficiency and Empowerment

A business performance dashboard allows users to be self-sufficient in creating, organising, and presenting information on business performance. Users are able to create the necessary materials promptly and efficiently without relying on the IT department or report writing specialist. With timely information, the management is able to resolve business issues promptly without wasting excessive time looking for the right information or report.

Chapter 2: Creating the Business Performance Dashboard

A business performance dashboard that relies on grids or overly complex charts does not communicate insights effectively to the viewers. It may take a longer time to visualise the data than the one that does not have complex charts and grids.

For example, a sales report displaying 12 months of history for 30 products (i.e. 12 × 30 = 360 data points) in a table does not help dashboard viewers see the trends and outliers as easily as a time-series chart of the same information. Also, it will take a longer time to load the table than a time-series chart. Worst yet, if viewers need to use the web browser to view the dashboard and the Internet connection is slow, the interactivity of the table will be slower than the interactivity of the time-series chart.

The dashboard shown in Figure 5.5 displays another example of the wrong ways of building a dashboard. The pie chart on the left has too many slices, making it difficult for dashboard viewers to perform precise comparisons of each product (i.e. A, B, C, etc.). The bar chart on the left represents the sales margin and displays it on y-axis ranges from zero to 90%, even though every data point is above 80%. It will be better to change the range of the y-axis values to give a more zoomed-in effect.

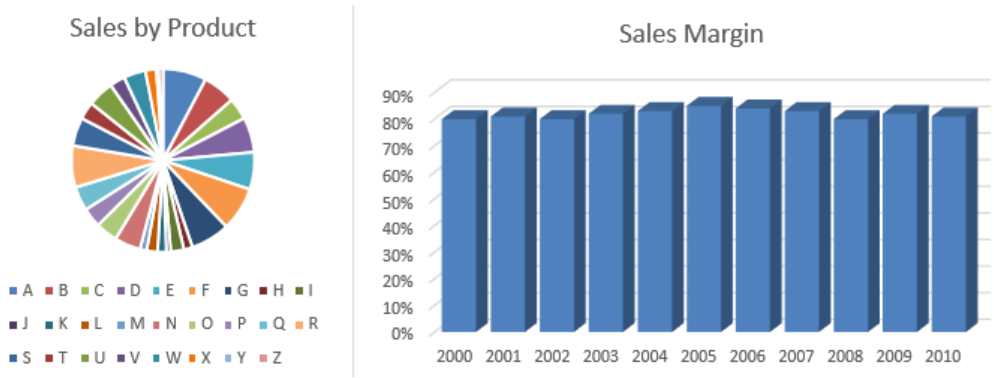


Figure 5.5 Two wrong ways of building a dashboard

2.1 Choose the Display Option

One of the first things we should consider when assembling worksheets in a dashboard is the available space that our viewers have to view them— for example, whether the dashboard will be viewed on an overhead projector with limited resolution, or the dashboard will be displayed on a personal computer or smart devices.

The top-left hand corner of the dashboard worksheet in Tableau displays all of the worksheets contained in the workbook. The bottom-left hand corner of the dashboard worksheet provides access to other object controls for adding text, images, blank space or live web pages into the dashboard worksheet. It also has control over specifying the size of the dashboard, and a checkbox for adding a dashboard title.

We can define the sizing of the entire dashboard and the individual objects included in the dashboard workspace. Before any worksheet is added into the dashboard workspace, we need to click the "size shelf", as shown in Figure 5.6, to define the size of the dashboard to accommodate the worst-case scenario in which the dashboard will be viewed. The automatic option expands the dashboard to fill up the available screen space. The exact option allows us to lock the dashboard width and height. The range option enables us to define minimum and maximum limits of the size of the dashboard. Once the size has been defined, we are ready to add individual worksheet objects into the dashboard.

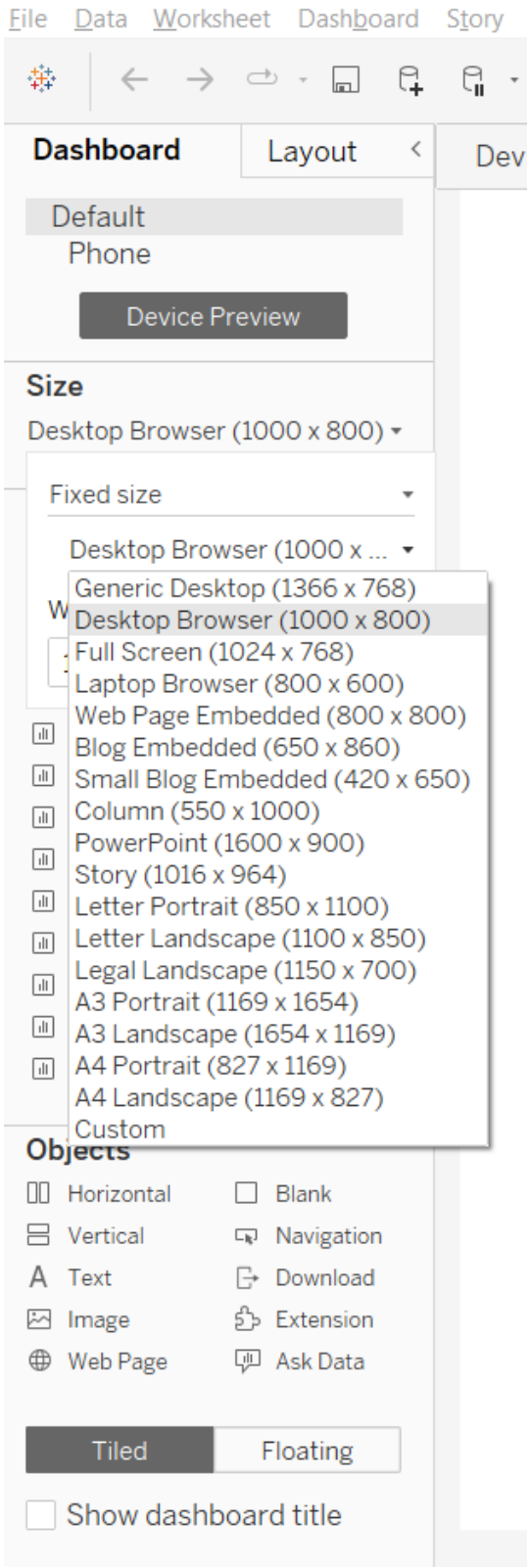


Figure 5.6 Tableau's dashboard layout size definition

2.2 Define the Visualisation Task

The important questions to ask here are – who will be looking at the dashboard and at what frequency? The answers to these two questions are key to defining the visualisation task which will help us ascertain the type and amount of information that will go in the dashboard. Let us take the following scenario: the sales manager for a global superstore has to do monthly reporting of his KPIs— number of customers, sales and profit achieved per month. In such a scenario, the sales dashboard should include monthly breakdown of total number of transactions, sales and profit for the year until the current month. In addition, the dashboard should also include certain performance tracker, for example, average revenue and profit per transaction broken down by product category for better analysis.

The dashboard should start with three bar graphs— one each for number of transactions, sales and profit per month. The next layer should be a line graph that compares sales and profit per transaction across months for each product category.

2.3 Position Objects in the Dashboard Workspace

In Tableau, after creating multiple worksheets, you can combine the worksheets to give an integrated view. Figure 5.7 shows an empty dashboard worksheet.

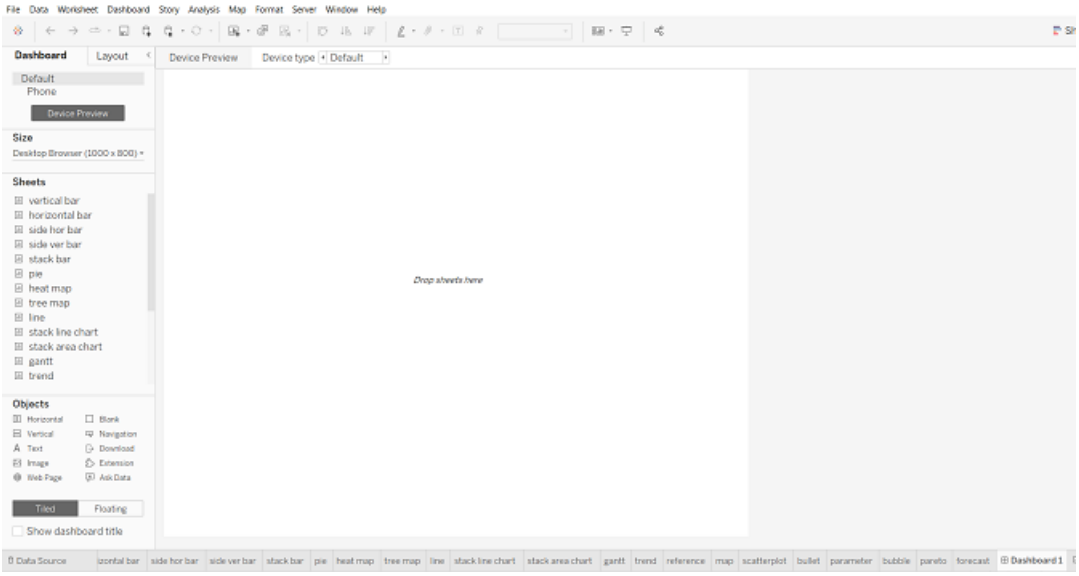


Figure 5.7 Tableau's dashboard worksheet

There are two ways to add objects into the dashboard workspace. We can either drag the selected object into the "Drop sheet here" area, or double click the worksheet objects on the top-left hand corner. By double clicking on each worksheet object, Tableau will place that object onto the dashboard workspace automatically. To control the placement of an individual object more precisely, drag the object into the view. Once we lift up the left-click button, Tableau will provide a preview of the area that the object will occupy by shading it in grey.

Following up on the visualisation task defined in Section 2.2, it makes sense for the bar charts showing the number of transactions and sales and profit to be displayed in the first row, followed by category-wise split of average revenue and profit per transaction, as illustrated in Figure 5.8.

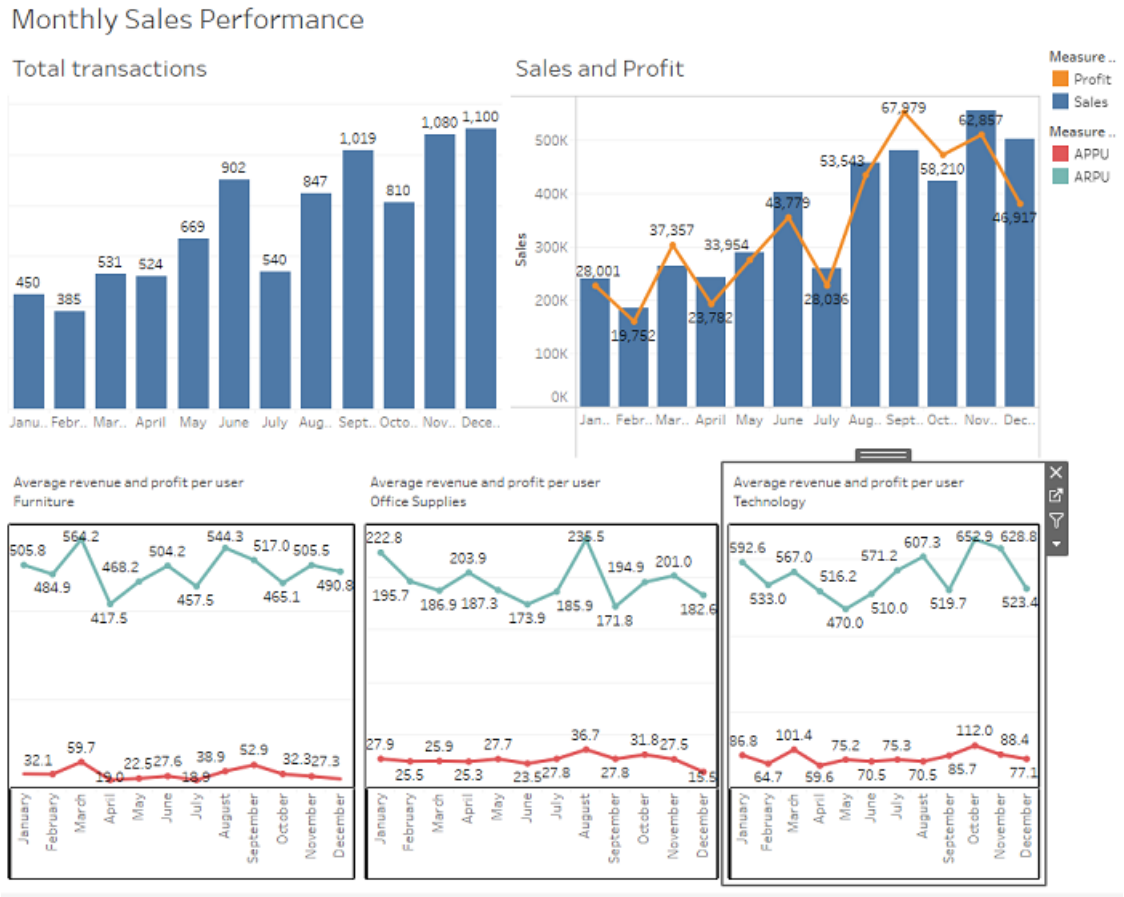


Figure 5.8 dashboard for tracking monthly sales KPI's performance

Add a title to the dashboard by selecting the "Show Title" option on the bottom left of the dashboard shelves. The default title will be the name of the dashboard worksheet. Edit the title text by double clicking on the default name and type in the new title. You can also edit the font size, colour, type and position of the title.

2.4 Dashboard Design Principles

To make a dashboard communicate the information more effectively to its viewers, we need to follow the basic dashboard design principles taken from Alexander (2008) and summarised as follows:

2.4.1 Keep It Simple

The well-known *KISS (Keep It Simple, Stupid)* principle of having simplicity as a key goal applies to the business performance dashboard's design. It is important not to have too many measures and visual distractions so that attention is focused on the essentials.

For example, we can hide the axis headers and mark labels to provide more space and make the dashboard look less cluttered. The axis headers can be hidden in each chart by pointing at the axis header area, right-clicking and unchecking the "show header" option. If the dashboard is going to be printed and consumed on paper, it is not a good idea to remove the axis headers. Mark labels can always be displayed, but space would be better utilised if mark labels are displayed only when the dashboard viewers want to see the labels, as illustrated in Figure 5.8.

2.4.2 Don't Display Everything

Fight the tendency to display everything, including raw data and all summaries of the data that one can think of. Display only what is required. This means that we clearly define the visualisation task and choose the appropriate charts as illustrated in section 2.2 of the chapter.

2.4.3 Keep to a Single Page

Eliminate the need to scroll back and forth, up and down, to look at the measures. Since the dashboard should provide all relevant measures at a glance, segregate raw data from the actual dashboard whenever possible. The dashboard in Figure 5.8 follows this principle as it puts everything on a single page.

2.4.4 Avoid Fancy Formatting

Avoid using colours and other fancy formatting like pattern fills, gradients, shadows and glows on the dashboards. In fact, plain formatting is more effective in drawing attention to the measures, graphics, and the traffic lights.

2.4.5 Use Layout and Placement

Even after selecting the critical performance measures or components that need to be displayed on the dashboard, there may be differing levels of importance among them. Alexander (2008) referred to studies done where the upper left and middle-left (i.e. dark blue) portion of a dashboard or document is called the priority zone, because it draws the most attention from dashboard viewers, as shown in Figure 5.9. Hence, place the more important measures or components in the priority region of the dashboard.

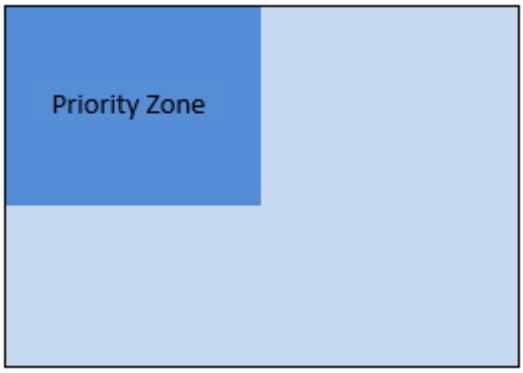


Figure 5.9 Priority zone in a dashboard

Following the above principle, in Figure 5.8, we placed the bar charts in the upper half to draw greater attention to the total transactions and sales and profit. The other three charts that showed secondary level of detail— average revenue and profit per user —were placed side by side below the bar graphs for the same reason.

2.4.6 Format Numbers Effectively

Format numbers to display commas and dollar symbols when the measures are monetary values. Limit the number of decimal places (i.e. usually to two for currency values) and display in denominations of thousands or millions instead of the actual amount where appropriate (e.g. display 1.4 mil instead of 1,437,387).

2.4.7 Use Titles and Labels Effectively

Always remember to have descriptive titles, column headers, and footnotes, with minimum use of short forms and acronyms where possible. Adding more descriptive object titles will make it easier for viewers to interpret it (refer to Figure 5.8).

2.5 Advanced Dashboard Navigation— Using Actions

Actions facilitate insight discovery by altering the context of the dashboard based on selections made by the dashboard viewers. We can build actions that filter and highlight the main dashboard. To create advanced navigation options— filter, highlight, go to url and others— we need to access the dashboard's menu option, then select the actions menu to expose the actions dialogue box, as shown in Figure 5.10.

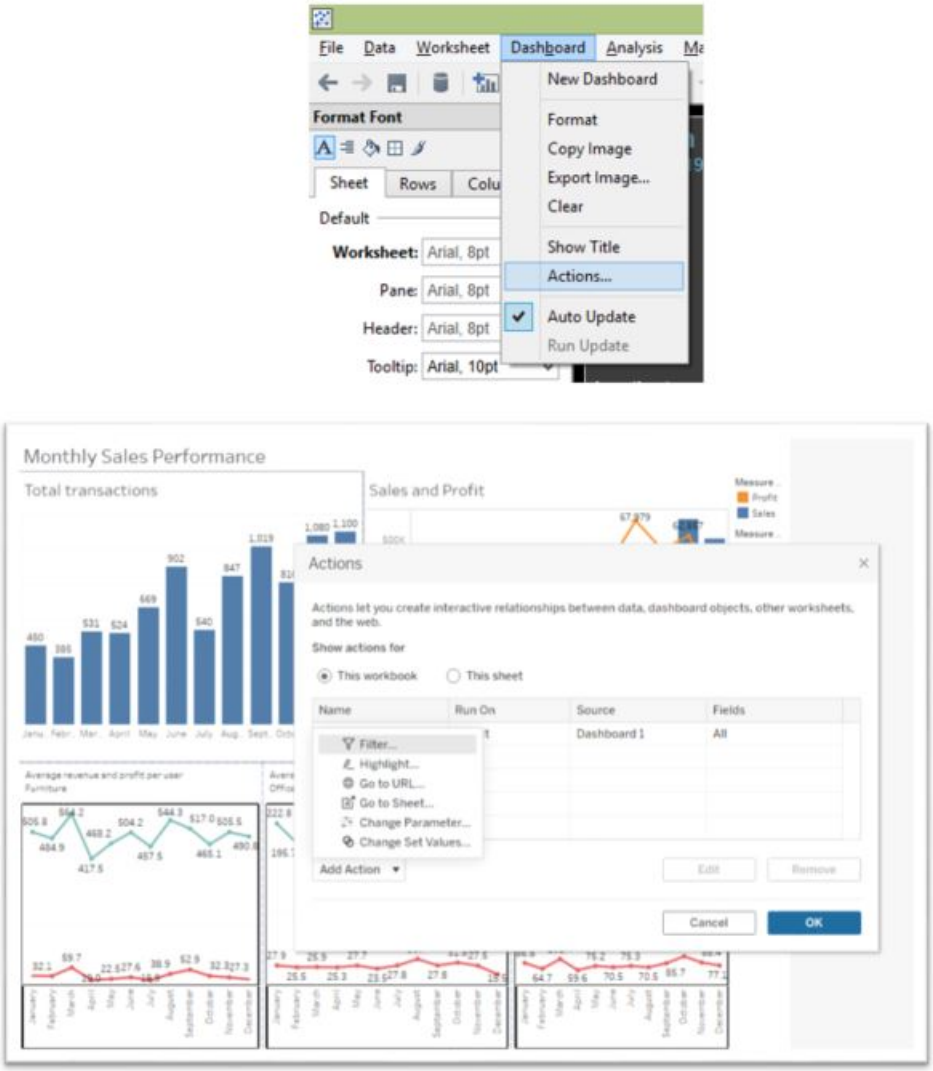


Figure 5.10 "Actions" menu in Tableau

2.5.1 Filter Action

Filter action filters the display area in target sheet based on the selection made in the source sheet. With reference to the dashboard in Figure 5.8, we want to filter the states only in one particular region corresponding to the region clicked on the pie chart, so we use the "Filter action" to achieve this, as illustrated in Figures 5.11a & b.

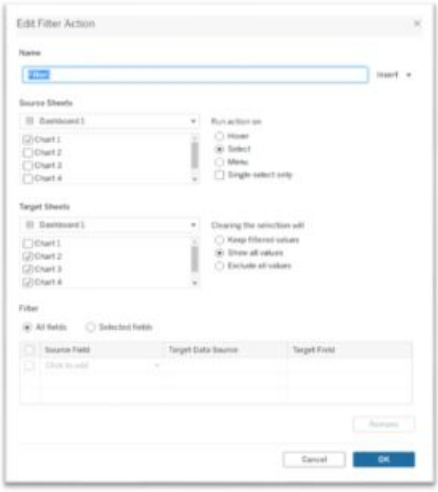


Figure 5.11a: Editing filter action in Tableau

Figure 5.11b: Dashboard with monthly filter action

Figure 5.11 a: Editing filter action in Tableau, b: Dashboard with monthly filter action

2.5.2 Highlight Action

Highlighting helps dashboard viewers to see related information more easily. For example, in the dashboard created (refer to Figure 5.8), we intend to highlight any one particular month based on the selection made in the bar chart. This can be achieved using the "Highlight Action" as illustrated in Figures 5.12a & b.

Figure 5.12a: Editing Highlight Action in Tableau

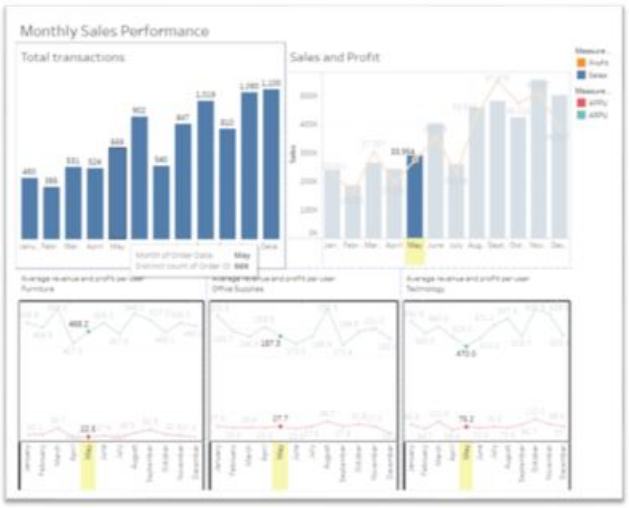


Figure 5.12 a: Editing Highlight Action in Tableau, b: Dashboard with monthly Highlight action

2.5.3 Go to URL Action

The "Go to URL" action is used when we want to access additional information/references with respect to a field selected on one of the charts in the dashboard. For example, if we want to read up on any of the measures— for example, ARPU and APPU— in the charts, we can add a "Go to URL" action for users to select the hyperlink from the menu that directs them to the URL/Wikipedia link. The action is illustrated in Figures 5.13a & b.

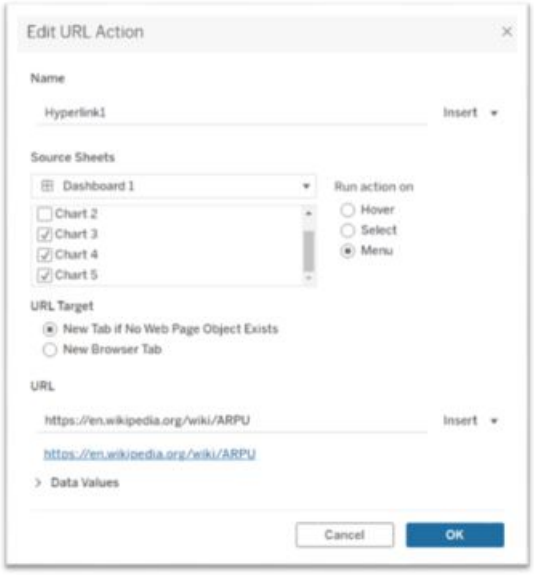


Figure 5.13a: Editing the Go to URL action

Figure 5.13b: Dashboard with Go to URL action

Figure 5.13 a: Editing the Go to URL action, b: Dashboard with Go to URL action

2.6 The Bigger Picture

The dashboard discussion so far has been based on historical data— data based on past performance or results. This data is collected, stored and retrieved as needed from the relevant databases. This is the most common type of data that is available and prevalent in organisations for performance analysis and visualisation. Figure 5.14 represents the process of visualisation and interpretation from historical data that has been covered in the current course. However, there is another type of data that is collected at a higher frequency, which could be daily or hourly, and is relevant at a more operational level. This is referred to as real-time data.

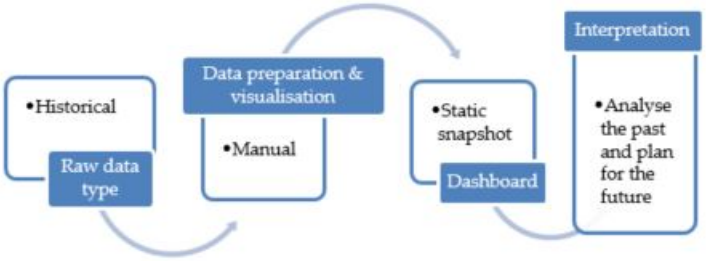


Figure 5.14 Dashboard interpretation with historical data

Figure 5.15 represents the process of visualisation and interpretation with real-time data. Data preparation and visualisation with real time data requires automated or programmed loop that runs at the same frequency at which the data is collected. This is often executed with the aid of programming languages (MATLAB, Python or R) and advanced statistical packages (SPSS, STATA), which is outside the purview of this course. Interpretation in this case is more about real time monitoring for any deviation from the allowed range of values or target values and facilitating immediate corrective action.

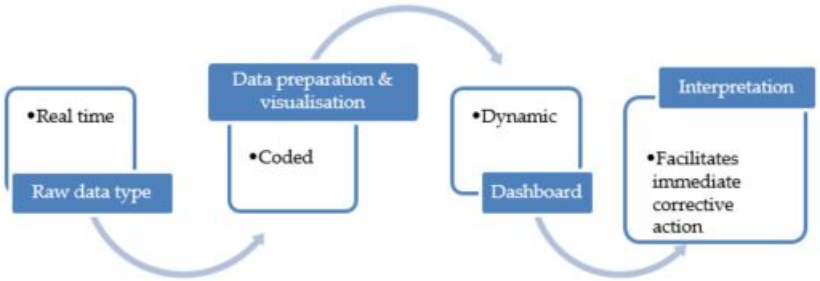


Figure 5.15 Dashboard interpretation with historical dataDashboard interpretation with real time data

Summary

A business performance dashboard can be defined as a visual display of the most important information needed to achieve one or more objectives, consolidated and arranged on a single screen so that the information can be monitored at a glance. It provides useful information that enables managers to perform improvement activities on areas that show performance that is below expectation. Business performance dashboards can generally be classified into three main types. Each type of dashboard caters to a different group of users. The three types of dashboards are the strategic dashboard, tactical dashboard, and operational dashboard.

The *strategic dashboard* is used by executives and managers to monitor the execution of strategic objectives. These strategic objectives can be the results of using a balanced scorecard approach. A *tactical dashboard* is used by managers and analysts to track the progress of departmental processes and projects against budget plans, forecasts or the last period's achievement. The *operational dashboard* is used by operational staff and their supervisors to monitor operational processes. Such dashboards are very specific to a particular operation or process.

Organisations implement business performance dashboards because of the many benefits that can be reaped by everyone in an organisation— such as strategy communication, increased insight, motivation and coordination, cost reduction and empowerment. For effective communication of information displayed on the dashboard, it is important to follow key design principles, such as keeping it simple, not trying to put everything on the dashboard, and using descriptive titles and labels.

Whilst an organisation can achieve significant benefits from the initial implementation of a spreadsheet-based scorecard system that is manual, they will eventually feel the need for a more sophisticated system that uses either off-the-shelf software packages or in-house developed software. The integration of the performance dashboard with

existing enterprise software systems can bring many benefits when the entire performance measurement system is unified.

Activity 5.1

1. Based on the worksheets that you have created in week 4 and week 5's class activities, create a dashboard using the Tableau Desktop.
2. Apply the dashboard design principles to ensure that your dashboard effectively communicates the right information to its viewers.
3. Use filter and highlight actions to help dashboard users and viewers see the related information in a dashboard more easily.

Formative Assessment

1. Which statement listed below is not a benefit of a business performance dashboard?
    a. Business performance dashboards facilitate the translation of corporate strategy into measures with their corresponding targets.
    b. A well-designed business performance dashboard is capable of replacing independent reporting systems, spreadsheet marts, data marts, and warehouses.
    c. Objectivity, openness and transparency in a business performance dashboard help promote effective coordination and encourage different departments to work towards their individual targets.
    d. A business performance dashboard provides the management team with greater insights into business performance in a timely manner.

2. The basic dashboard design principles are:
    i. Avoid fancy formatting
    ii. Keep it simple
    iii. Display everything
    iv. Format numbers effectively
    a. i, ii, and iii
    b. i, iii, and iv
    c. i, ii, and iv
    d. All of the above.

3. Which portion of the dashboard is the priority zone?
    a. A large portion in the dashboard
    b. The upper-left and middle-left portion
    c. The central portion

d. The bottom portion

4. Business performance dashboards can generally be classified into three main types. Which one is incorrect?
    a. Strategic dashboard
    b. Information dashboard
    c. Tactical dashboard
    d. Operational dashboard

5. Which among the below is not a dashboard actions option in Tableau?
    a. Filter
    b. Highlight
    c. Go to URL
    d. Copy-paste

Solutions or Suggested Answers

Formative Assessment

1. Which statement listed below is not a benefit of a business performance dashboard?

   a. Business performance dashboards facilitate the translation of corporate strategy into measures with their corresponding targets.
   Incorrect.

   b. A well-designed business performance dashboard is capable of replacing independent reporting systems, spreadsheet marts, data marts, and warehouses.
   Incorrect.

   c. Objectivity, openness and transparency in a business performance dashboard help promote effective coordination and encourage different departments to work towards their individual targets.
   Correct. Objectivity, openness and transparency in business performance dashboards help promote effective coordination and encourage different departments to work more closely together and not towards their individual targets.

   d. A business performance dashboard provides the management team with greater insights into business performance in a timely manner.
   Incorrect.

2. The basic dashboard design principles are:

   i. Avoid fancy formatting
   ii. Keep it simple
   iii. Display everything
   iv. Format numbers effectively

a. i, ii, and iii
Incorrect.

b. i, iii, and iv
Incorrect.

c. i, ii, and iv
Correct. A dashboard should avoid fancy formatting, is simple, and displays only what is required with numbers that are formatted effectively.

d. All of the above.
Incorrect.

3. Which portion of the dashboard is the priority zone?

a. A large portion in the dashboard
Incorrect.

b. The upper-left and middle-left portion
Correct. The upper-left and middle-left portion of a dashboard or document is called the priority zone because it draws the most attention from dashboard viewers.

c. The central portion
Incorrect.

d. The bottom portion
Incorrect.

4. Business performance dashboards can generally be classified into three main types. Which one is incorrect?

a. Strategic dashboard
Incorrect.

b. Information dashboard

Correct. The three types of dashboards are the strategic dashboard, tactical dashboard, and operational dashboard.

c. Tactical dashboard

Incorrect.

d. Operational dashboard

Incorrect.

5. Which among the below is not a dashboard actions option in Tableau?

a. Filter

Incorrect.

b. Highlight

Incorrect.

c. Go to URL

Incorrect.

d. Copy-paste

Correct. Copy-paste is not an option when you click on the dashboard actions tab in Tableau.

References

Alexander, M. (2008). *Excel 2007 dashboards & reports for dummies*. USA: Wiley Publishing.

Eckerson, W. W. (2006). *Performance dashboards: Measuring, monitoring and managing your business*. New York: John Wiley & Sons.

Eckerson, W.W. (2006, July). *Deploying dashboards and scorecards*. TDWI best practices report.

Few, S. (2007). *Pervasive hurdles to effective dashboard design*. Visual Business Intelligence Newsletter. http://www.perceptualedge.com/articles/visual_business_intelligence/pervasive_hurdles_to_dd.pdf.

Few, S. (2006, November). *Dashboard design for rich and rapid monitoring*. Visual Business Intelligence Newsletter.

Few, S. (2004, March). *Dashboard confusion*. http://www.perceptualedge.com/articles/ie/dashboard_confusion.pdf.

Murray, G. D. (2016). *Tableau your data: Fast and easy visual analysis with Tableau software* (2nd ed.). Indianapolis, IN, USA: John Wiley & Sons Inc.

Study Unit 6

Advanced Data Visualisation Techniques

Learning Outcomes

At the end of this unit, you are expected to:

- Determine appropriate data visualisation techniques based on given data

Overview

This unit introduces visualisation for spatial data. Spatial data are related to the location of the subject matter. There is a natural hierarchy to spatial data that allows, and often requires, us to explore at different granularities. We study maps, spider maps and path maps for representing spatial data. This unit also discusses visualisations to examine relationships among two or more variables in the data- scatter plot and bubble chart and visualisations to benchmark performance against target or predicted value for a variable in the data- bullet chart. It also demonstrates advanced visualisations such as pareto chart and forecasting using tableau.

Chapter 1: Visualisation for Spatial Data

Lesson Recording

Visualisation Techniques for Spatial and Multi Variable Data

Spatial data are related to the location of the subject matter, such as customer country, supplier address, etc. According to Yau (2013), there is a natural hierarchy to spatial data that allows, and often requires, us to explore at different granularities. For example, from the corporate level, the senior management may be interested to analyse the data based on country level. However, in order to know what happened in the business units, tactical and operational managers may want to zoom in on the data into provinces or states, counties, districts, cities, towns, neighbourhoods, all the way down to the individual household. Table 6.1 summarises the chart type that is most suitable for the visualisation task defined in the left column of the table.

Table 6.1 List of chart type matched to the visualisation task for spatial and multi variable data

Visualisation task	Chart type
Compare continuous variable (sales, profit, discount, quantity, cost, etc.) across locations (markets, cities, states, countries, regions)	Map
Plot the path from starting point to destination	Spider maps, path maps
Observe relationships between two numeric variables—for example, association and correlation	Scatter plot

Visualisation task	Chart type
Compare the performance of a primary measure to one or more other measures One example— did the actual sales meet target sales or not?	Bullet chart
Display one-to-many comparison by using size and colour without the need for precise visual ranking between measures	Bubble chart
Show the percentage of total sales/profit that comes from the top products, and thus help the business manager identify the products that are most important for the business's success	Pareto chart

1.1 Best Practices to Visualise Spatial Data

The most common way to visualise spatial data is with maps that place values within a geographic coordinate. However, a map is not always the most informative way to visualise spatial data. For example, we can treat "Region" as a category and use a bar chart to visualise the data.

We can visualise the geographic coordinate of a location by mapping the latitude and longitude coordinates to a two-dimensional space, and draw a point on the space. This seems quite straightforward, but it can pose challenges when there are a lot of locations to be visualised.

When the density of individual locations across a region is more informative than the overlapping points on a map, we may want to colour code the region based on the density scale, or use lines to show data continuously over geography. We can also size the regions by the data and ignore the physical area, so that the regions with high density data will appear bigger than the regions with low density data.

Rather than separating locations, we may sometimes want to explore connections between two locations. For example, when we analyse social media data, we would like to know

the relationship between one person and the others. In this scenario, we can plot each individual on a map, and draw lines to connect one individual with the others.

1.1.1 Map

To create maps in the Tableau, we need to assign a "Geographic Role" to the field that contains geographic data, such as country, region, state, zip code, and so on. Fields with a geographic role will automatically generate longitude and latitude coordinates to be visualised as a map. Geographic roles are arranged in a hierarchy where the field with the lowest level of detail is what is shown on the map.

The sales vice president of global superstores would like to see how each state performed as compared to others in terms of profit generated in the U.S. One way to visualise this is using a bar chart, but the categorical axis will get too long as there are almost 50 states in the U.S.. Another chart that could be used here is the pie-chart, but again, it will become too cluttered showing almost 50 slices in a single pie. The best way to visualise this information so that it can be interpreted easily is to show the profit generated by each state on the map of the U.S. and use some sort of colour gradation to show how each state performed as compared to others on a spectrum of minimum to maximum profit; the steps to creating this visualisation in Tableau using a map is illustrated in Figure 6.1 below:

1. Create a new worksheet.
2. If necessary, assign the geographic role to fields that are not automatically recognised by Tableau as geographic data.
3. Drag and drop "Longitude" measure to columns.
4. Drag and drop "Latitude" measure to rows.
5. Drag and drop "Country" dimension on the "Detail".
6. Add "Country" dimension to filters and filter "United States" from the countries list.
7. Drag and drop "State" dimension on the "label" marks card.
8. Drag and drop "Profit" measure on the "Color" marks card.
9. Drag and drop "Profit" measure on the "Label" marks card.

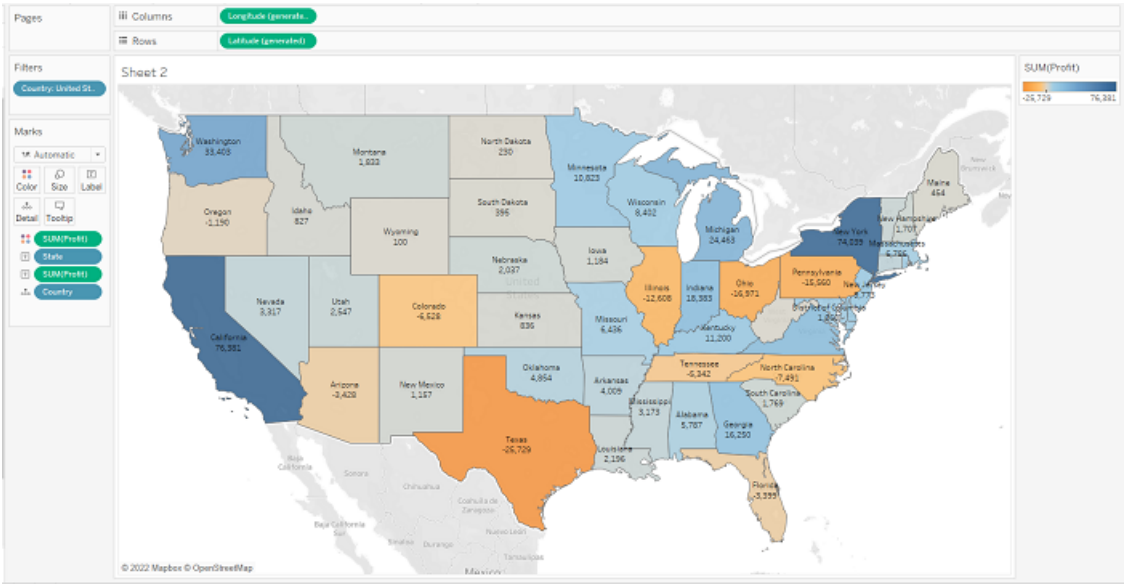


Figure 6.1 A map in Tableau

Read

Here are some optional readings to serve as additional references for students who want to learn more about the topics just covered:

1. Murray. D. G. (2016), pp. 191-220.
2. Yau, N. (2013), pp. 165-176.
3. Tableau Online Help

   https://help.tableau.com/current/pro/desktop/en-us/maps.htm

1.1.2 Mapping Point-to-Point Details on Map

The most often used point-to-point maps in data visualisation are either path maps or spider maps— also known as origin destination maps. Path maps show point-to-point detail or the path between two locations, as shown in the airlines flight destination map in Figure 6.2.

Figure 6.2 Airlines flight destination map

Spider maps show paths between an origin and one or more destination locations as shown in the SMRT map in Figure 6.3.

Figure 6.3 SMRT map

(Source: LTA.GOV)

Mapping point-to-point details on maps require that the data support plotting and linking each point. To create this type of map in Tableau, data should include the following information, as illustrated in Figure 6.4:

- A path ID for every unique path
- Numbers to define the drawing order of each data point
- Latitude and longitude coordinates for every location

Line (suburban metro line number)	Line Group (Path ID)	Order of Points	Station (Paris, France)	Latitude	Longitude	Traffic
1	1	1	La Défense (Grande Arche)	48.891934	2.237883	14,275,382
1	1	2	Esplanade de la Défense	48.887843	2.250442	9,843,051
1	1	3	Pont de Neuilly	48.884509	2.259892	6,902,931
10 BOUCLE	10	1	Boulogne-Jean-Jaurès	48.842222	2.238836	3,847,782
10 BOUCLE	10	2	Porte d'Auteuil	48.848074	2.258648	687,237
10 BOUCLE	10	3	Michel-Ange-Auteuil	48.847740	2.264297	2,222,709

Figure 6.4 Dataset for point-to-point mapping

Below are the steps to creating a point-to-point map in Tableau (illustrated in Figure 6.5) using the dataset <point to point map dataset.xls>:

1. Create a new worksheet.
2. In the new worksheet, from Measures, drag **Longitude** to the **Columns** shelf, and **Latitude** to the **Rows** shelf.
3. From Dimensions, drag **Line Group (Path ID)** to **Detail** on the Marks card.
4. On the Marks card, click the Mark Type drop-down and select **Line.**
5. From Measures, drag **Point Order** to **Path** on the Marks card.
6. Point Order is aggregated as a sum.
7. On the Marks card, right-click the **SUM(Point Order)** field and select **Dimension.**
8. From Dimensions, drag **Line Group (Path ID)** to **Color** on the Marks card.
9. Each line now has its own colour associated with it, and a colour legend is added to the view.
10. From Measures, drag **Longitude** to the **Columns** shelf and place it to the right of the first longitude field.
11. On the Marks card, under the bottom AVG (longitude) tab, click the Mark type drop-down and select **Automatic.**

12. On the Columns shelf, right-click the second **AVG (Longitude)** field (on the right), and select **Dual Axis**. Your map views are now layered on top of one another.
13. From Measures, drag **Traffic** to **Size**, on the bottom AVG (longitude) Marks card. The size of the data points update to show the amount of traffic per station.
14. On the Marks card, click **Size** and move the slider to the right.
15. On the Marks card, click **Color**, and then, under Effects, click the **Border** drop-down and select a colour.
16. The view is now complete. You can quickly find the stations on each metro line with the most traffic.

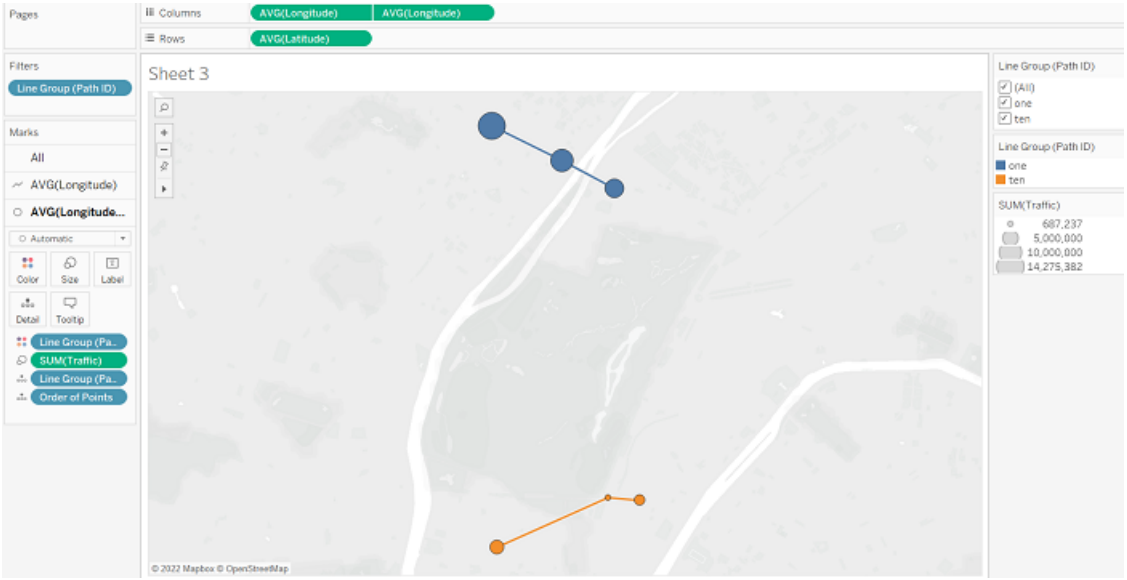


Figure 6.5 Mapping point-to-point detail

Read

Here is an optional reading to serve as an additional reference for students who want to learn more about the topics just covered:

1. Tableau Online Help

   https://help.tableau.com/current/pro/desktop/en-us/maps.htm

Chapter 2: Visualisation for Multi Variable Data

According to Yau (2013), data often come in a table format with multiple columns. Each column represents a variable. These data may come from survey questionnaires, results of an experiment that measure multiple aspects of a system, or demographic data on countries that include multiple information of each country.

2.1 Best Practices to Visualise Multi Variable Data

To visualise multi-variable data, we can fit all the data onto a screen and display the relationships amongst variables or trends in each variable. However, sometimes, the relationships amongst variables or trends are not so straightforward. In such cases, we need to use multiple views through more straightforward charts.

We can use scatter plots to visualise the relationships amongst variables. We can plot one variable on the horizontal axis, and the other variable on the vertical axis. The statistical relationship amongst variables is called correlation. Correlation is a statistical measure that measures the degree of how one variable's value tends to change in a certain way as the other variable's value changes. For example, taxi fares and the price of gasoline are positively correlated as both have been increasing over the years.

As we analyse the relationships amongst variables, we should not confuse correlation with causation. Causation is the extent of how one variable will impact another. To differentiate between correlation and causation, using the taxi fare and gasoline prices as an example, we may need to ask whether taxi fares go up by default if we increase the price of gasoline. Next, if taxi fare does go up, is it because of the increase in the price of gasoline or is it due to external factors such as taxi rental. It is difficult to account for external factors, thus it is challenging to prove the causation. It is however easy to find correlation. Correlation can help us predict one variable's value by knowing the other variable's value.

Read

Here is an optional reading to serve as an additional reference for students who want to learn more about the topics just covered:

1. Yau, N. (2013), pp. 176-189, 189-199

2.2 Creating Charts to Visualise Multi Variable Data

2.2.1 Scatter Plot

Scatter plots' primary uses are to observe and show relationships between two numeric variables. The dots in a scatter plot not only report the values of individual data points, but also patterns when the data are taken as a whole. Identification of correlational relationships are common with scatter plots. In these cases, we want to know if we were given a particular horizontal value, what a good prediction would be for the vertical value. You will often see the variable on the horizontal axis denoted as an independent variable, and the variable on the vertical axis as the dependent variable. Relationships between variables can be described in many ways: positive or negative, strong or weak, linear or nonlinear.

A scatter plot can also be useful for identifying other patterns in data. We can divide data points into groups based on how closely sets of points cluster together. Scatter plots can also show if there are any unexpected gaps in the data and if there are any outlier points. This can be useful if we want to segment the data into different parts, like in the development of customer profiles.

When we have lots of data points to plot, this can develop into the issue of overplotting. Overplotting is the case where data points overlap to a degree where we have difficulty seeing relationships between points and variables. It can be difficult to tell how densely-packed data points are when many of them are in a small area. There are a few common

ways to alleviate this issue. One alternative is to sample only a subset of data points: a random selection of points should still give the general idea of the patterns in the full data. We can also change the form of the dots, adding transparency to allow for overlaps to be visible, or reducing point size so that fewer overlaps occur. As a third option, we might even choose a different chart type like the heatmap, where colour indicates the number of points in each bin.

Scatter plots also use colour, shape and size to express more aspects of the variables. Using the global superstore scenario, when the sales vice president would like to compare the profit and sales across categories and regions for the U.S., he can use a scatter plot in Tableau, as illustrated in Figure 6.6. Below are the steps to creating a scatter plot in Tableau:

1. Create a new worksheet.
2. Drag "Sales" measure into the worksheet's columns. The measure is automatically aggregated as a summation.
3. Drag "Profit" measure into the worksheet's rows. The measure is automatically aggregated as a summation.
4. Drag and drop "Region" dimension on the "Color" marks card.
5. Drag and drop "Category" dimension on the "Shape" marks card.
6. Drag and drop "Country" dimension in "Filters" and select "United States" in the country name drop down list.
7. Click the "Analytics" tab next to "Data" pane. From the drop down menu, double click on "Reference Line" under "Custom".

Figure 6.6 Creating a scatter plot in Tableau

The above scatter plot can be used to make the following observations:

Both office supplies and technology categories generated above average sales and profit in the east and west of the U.S. They can be loosely called the star performers in these two regions. On the other hand, all three categories performed below average in terms of both sales and profit in the South. This could be a reason for concern and also indicates some problem that needs further investigation. Also, except the technology category, the other two categories did below average in central US. This could be another area that should be studied to understand the reasons for poor performance.

On the other hand, the technology category generated above average profit despite below average sales in central US. This means that this category has higher profit margin and could be the driver for sales growth in future. Juxtapose it with the performance of the furniture category in east and west US, which generated above average sales but below average profits. This means that though the volume business is higher for these categories, profit margins are lower. These categories must be addressing a mass market with low

product differentiation and these could be loosely called the "cash cows" for the company. Such categories generate sales with lesser marketing investment as long as you get the pricing right.

Read

Here are some optional readings to serve as additional references for students who want to learn more about the topics just covered:

1. Yau, N. (2013), pp. 176-189.
2. Tableau Online Help

   https://help.tableau.com/current/pro/desktop/en-us/buildexamples_scatter.htm

2.2.2 Bullet Chart

A bullet chart is a bar chart that includes reference lines and reference distributions for each cell in the chart (1 dimension and 2 measures). A bullet graph is useful for comparing the performance of a primary measure to one or more other measures. The bullet chart is usually created to compare actual value with target value of the measure.

When the sales manager of the company would like to visualise the sales in the years 2012 and 2013 in each region together with target sales, he can use a bullet chart in Tableau, as illustrated in Figure 6.7. Below are the steps to creating a bullet chart in Tableau:

1. Create a new worksheet.
2. Drag "Sub-Category" dimension into the worksheet's rows.
3. Drag "Sales" and "Profit" measures into the worksheet's columns.
4. From the drop down menu on the right hand side, choose "Bullet Chart" as the chart type.

5. Right click on x-axis and select "Swap Reference Line Fields" option from the drop down menu to change the "Sales" and "Profit" positions in the graph.

For this dataset, we will compare sales with profit.

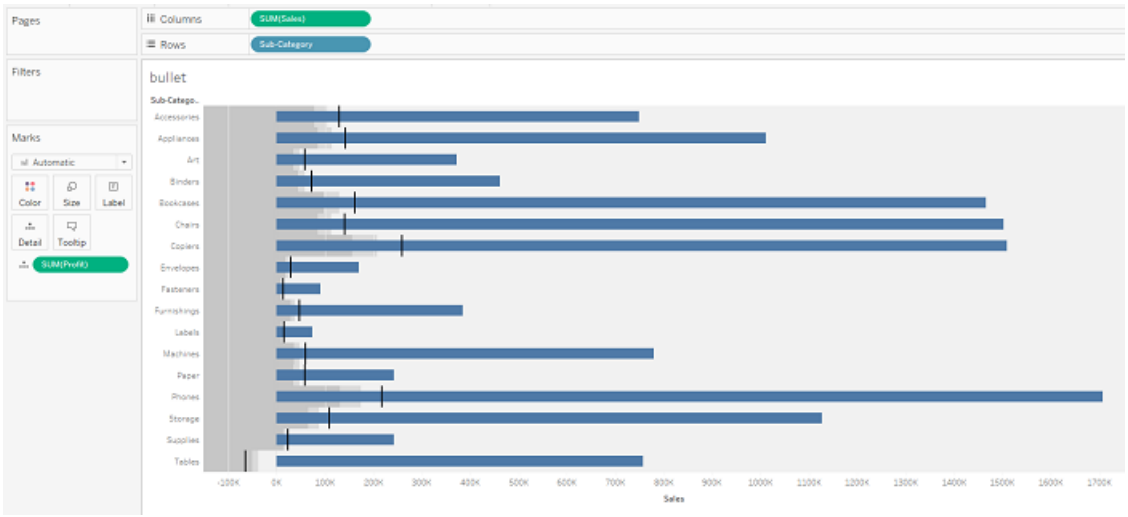


Figure 6.7 Creating a bullet chart in Tableau

This graph shows actual sales compared to estimated sales (profit, in this case). The centre blue bar represents actual value. The black vertical line represents a target value (profit, in this case). The grey coloured bands represent ranges, such as poor, average, and good.

Read

Here are some optional readings to serve as additional references for students who want to learn more about the topics just covered:

1. Yau, N. (2013), pp. 149, 197-198.
2. Tableau Online Help – Bullet Charts at

   https://help.tableau.com/current/pro/desktop/en-us/qs_bullet_graphs.htm

2.2.3 Building Parameters

Using parameters, data visualisation users can change normally static values into dynamic entities that facilitate ad-hoc analysis without the need to change the design of the data visualisation. A parameter is a workbook variable such as a number, date, or string that can replace a constant value in a calculation, filter, or reference line. For example, you may create a calculated field that returns "True" if sales is greater than $100,000 and returns "False" otherwise. You can replace the constant value of "100000" in the formula with a parameter. Then, using the parameter control, you can dynamically change the threshold in your calculation. The steps to creating parameters in Tableau is illustrated using a horizontal bar graph as below:

1. Create a new worksheet.
2. Set "Sales" measure as column and "Sub-Category" dimension as rows. Sort the bars from highest to lowest sales.

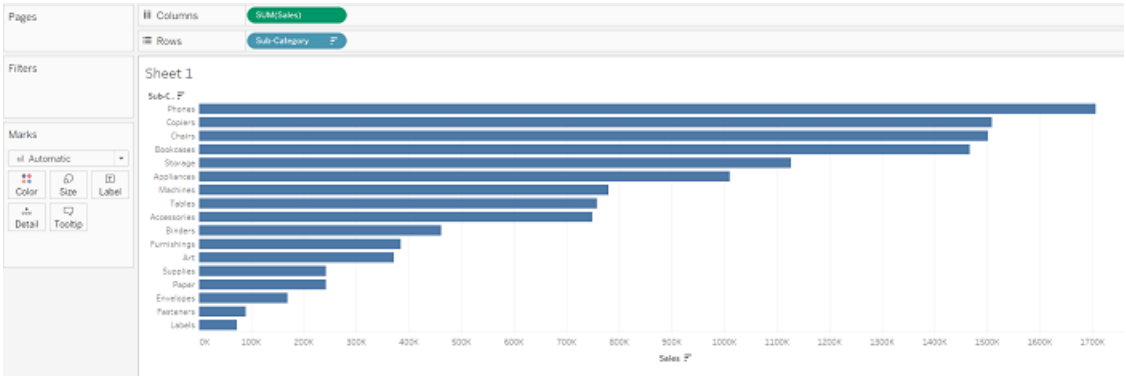


Figure 6.8 A horizontal bar graph in Tableau

3. Click the drop down menu as shown in circle below and select "Create Parameter".

Figure 6.9a: Opening the "Create Parameter" window

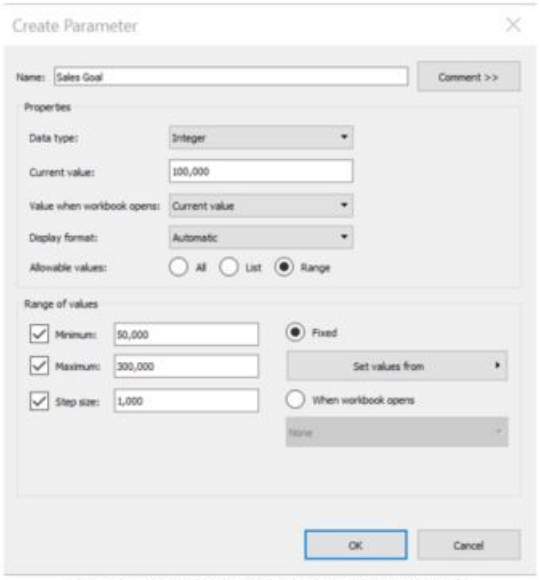


Figure 6.9b: Settings on the "Create Parameter" window

Figure 6.9 a: Opening the "Create Parameter" window,
b: Settings on the "Create Parameter" window

4. Set the data type to be "Integer" and format to "Currency (standard)".
5. Specify current value as 100,000.
6. Select Range option in Allowable values option.
7. Specify the minimum as 50,000, maximum as 300,000 and step size as 1000.
8. Open the "Create Calculated Field" window, as shown in Figure 6.10a.

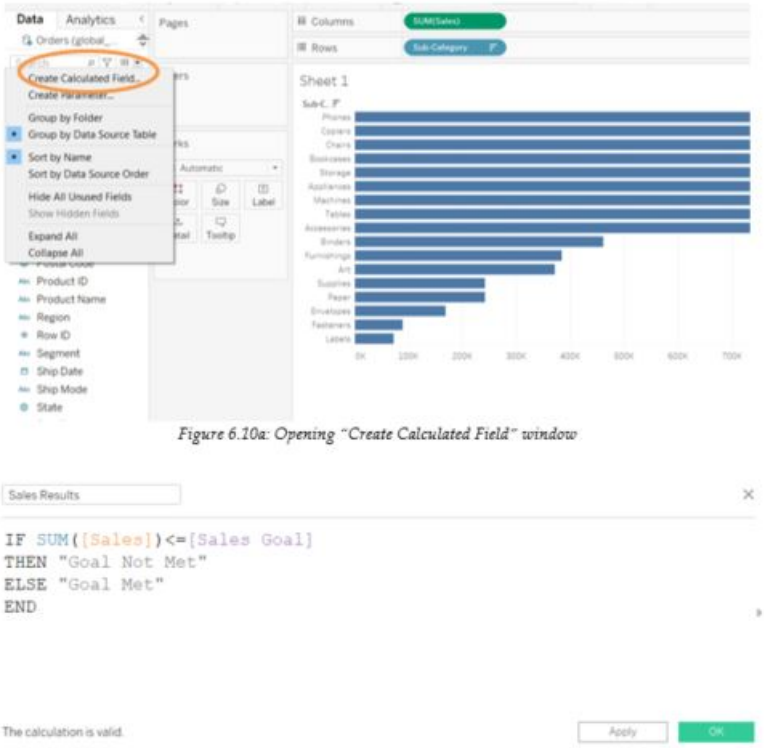


Figure 6.10a: Opening "Create Calculated Field" window

Figure 6.10 a: Opening "Create Calculated Field" window b: "Create Calculated Field" window

9. Create a calculated field called "Sales Results", as illustrated in Figure 6.10b.
10. Set sales results as the colour and sales goal as the detail.
11. Right click on Sales Goal under Markers and Select "Show Parameter"— see how the colour changes accordingly when you move the sales goal slider as illustrated in Figure 6.11.

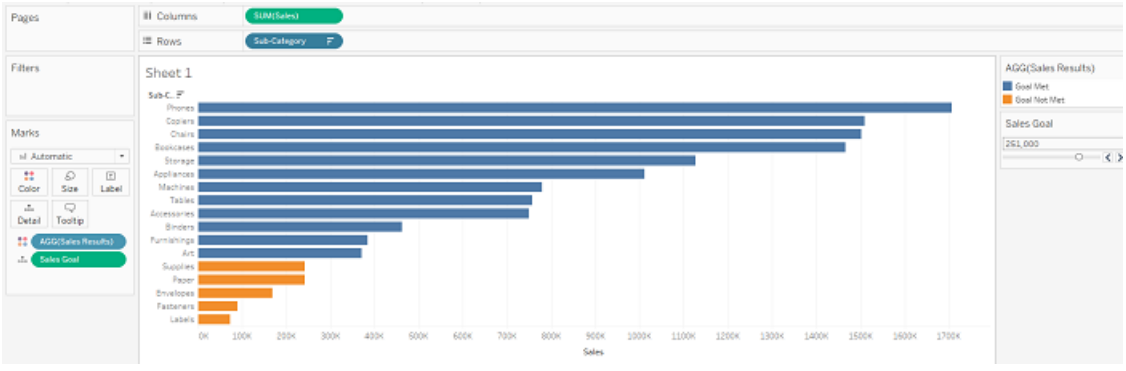


Figure 6.11 Horizontal bar graph with a created parameter

Read

Here is an optional reading to serve as an additional reference for students who want to learn more about the topics just covered:

1. Tableau Online Help

   https://help.tableau.com/current/pro/desktop/en-us/parameters_create.htm

2.2.4 Bubble Chart

A bubble chart displays one-to-many comparisons by using size and colour. However, it does not allow precise comparison between different bubbles. Therefore, we should use a bubble chart only when we do not require precise visual ranking of the bubble. Use packed bubble charts to display data in a cluster of circles. Dimensions define the individual bubbles, and measures define the size and colour of the individual circles.

When the sales manager of the company would like to do a rough comparison of the profit and the sales across product categories/sub-categories/regions, he/she can use a bubble chart in Tableau, as illustrated in Figure 6.12. Below are the steps to creating a bubble chart in Tableau:

1. Create a new worksheet.
2. Drag the "Category" dimension into the worksheet's columns.
3. Drag the "Profit" measures into the worksheet's rows.
4. Choose "Packed Bubbles" as the chart type.
5. Drag the "Sales" measure into "Colour" on the Marks Card to colour code the bubbles based on the "Sales" value.

Figure 6.12 Creating a bubble chart in Tableau

From the bubble chart above, we can roughly conclude that the technology category has both the highest sales and profit whereas the furniture category has the lowest profit despite medium sales. On the other hand, office supplies makes good profit despite having the lowest sales among the three categories, which could also indicate that the category has higher profit margins and with increased sales with the right marketing stimuli, it can drive the bottom-line of the company.

Read

Here is an optional reading to serve as an additional reference for students who want to learn more about the topics just covered:

1. Tableau Online Help – Bubble Charts at https://help.tableau.com/current/pro/desktop/en-us/buildexamples_bubbles.htm

2.2.5 Pareto Chart

The Pareto chart visualises the 80-20 rule, as developed by Vilfredo Pareto in 1960. It was originally developed to describe the unequal distribution of wealth in his country. In general, the 80-20 rule states that 20% of the inputs account for 80% of the outputs. Pareto made the observation that 80% of land was typically owned by 20% of the population. Pareto extended his principle by observing that 20% of the peapods in his garden contained 80% of the peas. Eventually, the principle was further expanded by others to propose that for many events, roughly 80% of the effects come from 20% of the causes. In business, for example, 80% of profits not infrequently derive from 20% of the available products.

A Pareto chart is a type of chart that contains both bars and a line graph, where individual values are represented in descending order by bars, and the ascending cumulative total is represented by the line. It shows the percentage of total sales that come from the top products, and thus helps the business manager identify the key segments of the customer base that are most important for the business's success. Below are the steps to creating a Pareto chart in Tableau, as illustrated in Figure 6.13.

1. Create a new worksheet.
2. Drag and drop “Sub-Category” dimension under columns in worksheet.
3. Drag and drop “Sales” measure under rows in worksheet and arrange the bars in descending order of sales.
4. Drag and drop “Sales” measure next to the first “Sales” and select chart type as “Line”. Two separate charts are created.
5. Now click on the second “Sales” in the rows and select dual axis option from the drop down menu. The line chart is superimposed on the bar graph.
6. For the “Sales” corresponding to the line chart, click add table calculation and select “running total” under primary calculation type and select add secondary calculation option.
7. Select “Percent of total” under secondary calculation type and click ok.

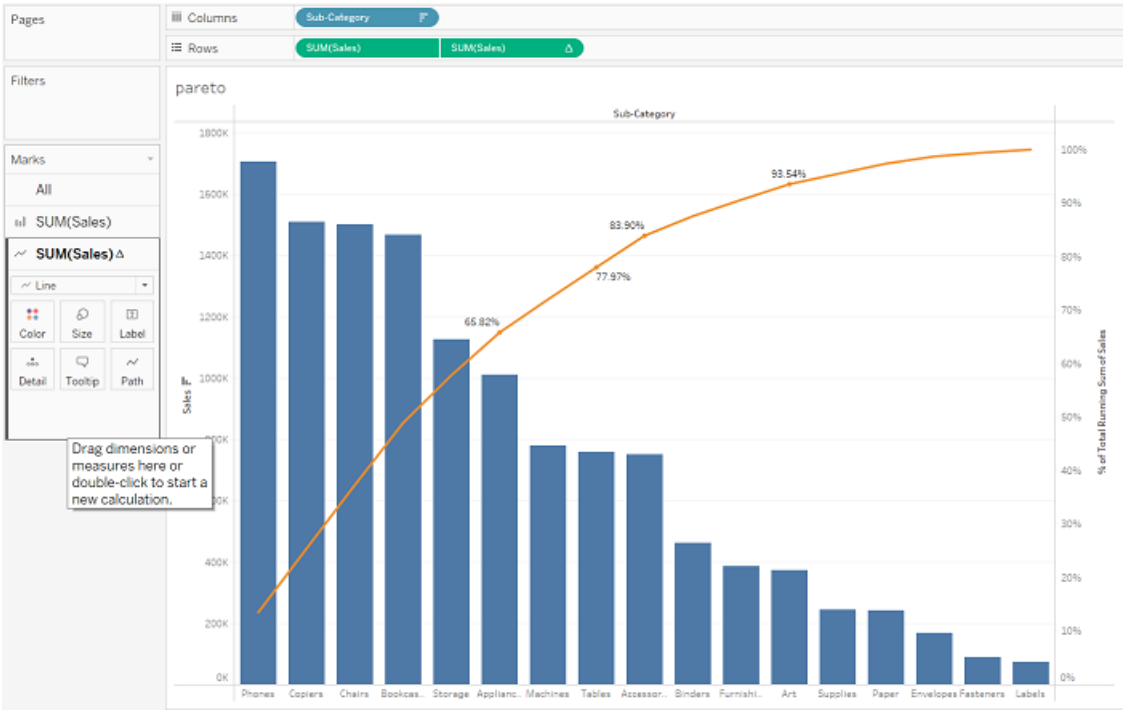


Figure 6.13 Creating a Pareto chart in Tableau

The line in the Pareto chart now corresponds to the cumulative sales value as a percentage of total sales for the categories up to that point. In the above chart, the label values on the line graph indicate that the top eight sub-categories (from "Phones" to "Tables" on the x-axis) account for ~78% of sales and if we include "Accessories", the top nine categories account for ~84% of sales. Thus, out of the 17 sub-categories, 84% of the sales comes from 9 or ~53% of the products. This can be useful in determining the allocation of resources to respective products.

Read

Here are some optional readings to serve as additional references for students who want to learn more about the topics just covered:

1. Tableau Online Help

   https://help.tableau.com/current/pro/desktop/en-us/pareto.htm

Chapter 3: Forecasting

Predicting future values is by nature imprecise. However, using extensive data from the past will give us an idea of what to expect in the future. Forecasting is the act of predicting future values based on historical values.

3.1 Creating Forecast

In Tableau, we can forecast data in different ways. Tableau will recommend what it deems is the best method to forecast our data. If we do not want to accept the default method, we can always edit the forecast model by right-clicking on the worksheet and selecting the forecast option menu.

Tableau provides the following forecast trend models:

- Trend and season
- Trend only: five data points
- Season only: two seasons or one season plus five periods
- No trend or season

Depending on the amount and granularity of data, each of the above forecast trend models generates different results. Trend and season model will generate the most volatile forecast data. The number of periods that Tableau forecasts is dependent on the data range and on the data aggregation level presented in the data visualisation. Tableau requires at least five data points in the time series (dates) to estimate a trend, and enough data points for at least two seasons or one season plus five periods to estimate seasonality. For example, at least nine data points are required to estimate a model with a four quarter seasonal cycle (4 + 5), and at least 24 to estimate a model with a twelve month seasonal cycle (2 * 12). When there is not enough data in the visualisation, Tableau automatically tries to forecast at a finer temporal granularity, and then aggregates the forecast back to the granularity of the visualisation.

Forecasting in Tableau uses a technique known as exponential smoothing. Forecast algorithms try to find a regular pattern in measures that can be continued into the future. All forecast algorithms are simple models of a real-world data generating process (DGP). For a high quality forecast, a simple pattern in the DGP must match the pattern described by the model reasonably well. Quality metrics measure how well the model matches the DGP. Tableau automatically selects the best of up to eight models, the best being the one that generates the highest quality forecast.

The steps to create a forecast chart in Tableau, as illustrated in Figure 6.14, are as below:

1. Set "month (order date)" to be column and "sum (sales)" as rows.
2. Click on "Analysis" tab and select "Forecast" from the drop down menu. By default, the forecast period is set at 13 and the interval is months. You may modify this by selecting "Forecast" -> "Forecast Options" from the drop down menu.

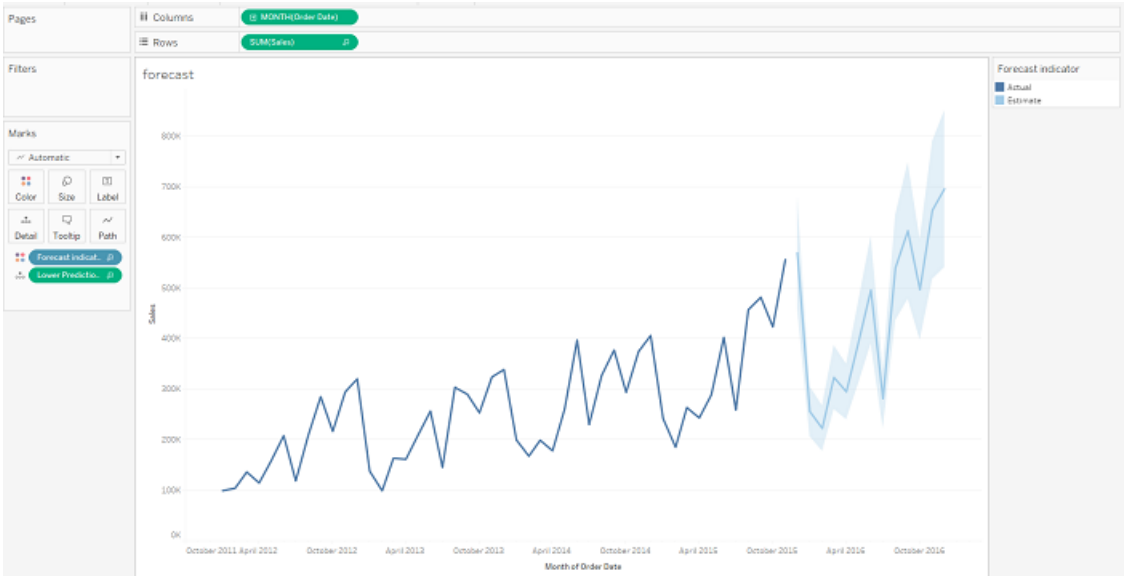


Figure 6.14 Creating a forecast chart in Tableau

For each measure that is forecasted, a summary table can be displayed by clicking on the "Analysis" tab and selecting "Forecast" -> "Describe Forecast" from the drop down menu. The summary table describes the forecast, as illustrated in Figure 6.15.

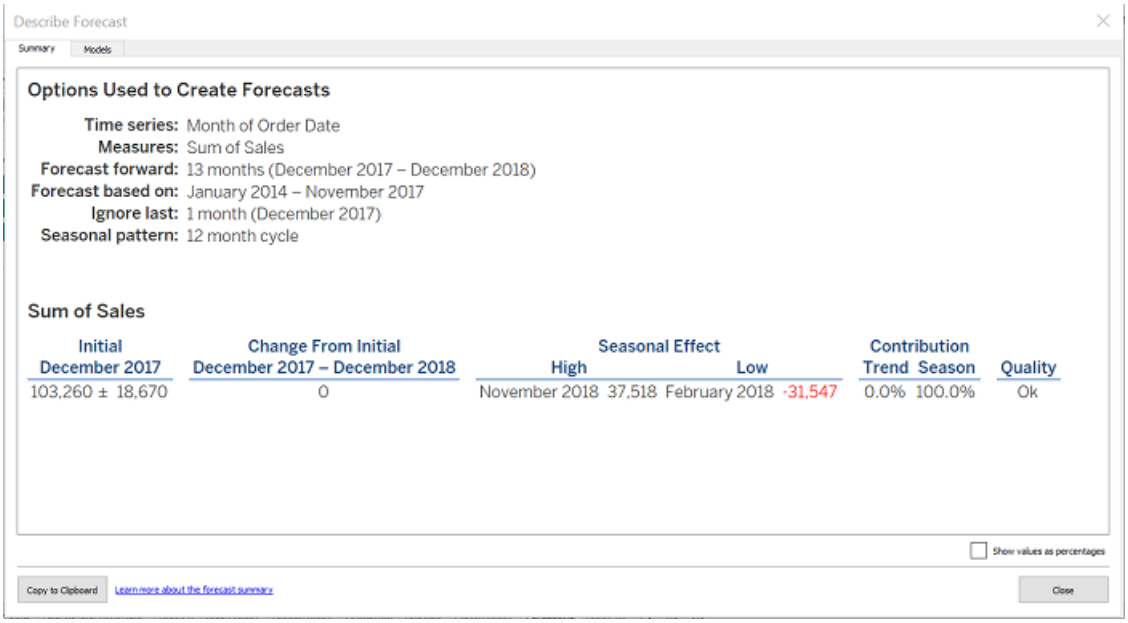


Figure 6.15 The forecast summary table in Tableau

The fields in the forecast summary table are:

Initial—the value and prediction interval of the first forecast period.

Change from initial—the difference between the first and the last forecast estimate points. The interval for those two points is shown in the column header. When values are shown as percentages, this field shows the percentage change from the first forecast period.

Seasonal effect—these fields are displayed for models identified as having seasonality—that is, a repeating pattern of variation over time. They show the high and low value of the seasonal component of the last full seasonal cycle in the combined time series of actual and forecast values. The seasonal component expresses the deviation from the trend and so varies around zero and sums to zero over the course of a season.

Contribution—the extent to which trend and seasonality contribute to the forecast. These values are always expressed as percentages and add up to 100%.

Quality—indicates how well the forecast fits the actual data. Possible values are GOOD, OK, and POOR. A naïve forecast is defined as a forecast that estimates that the value of the next period will be identical to the value of the current period. Quality is expressed

relative to a naïve forecast, such that OK means the forecast is likely to have less error than a naïve forecast, GOOD means that the forecast has less than half as much error, and POOR means that the forecast has more error.

Read

Here are some optional readings to serve as additional references for students who want to learn more about the topics just covered:

1. Murray. D. G. (2016), pp. 221-242.
2. Tableau Online Help https://help.tableau.com/current/pro/desktop/en-us/forecast_create.htm

Summary

In visualising spatial data, we map data to colour and geometry on a map. It seems easy because we just need to draw and colour the geometry. However, it is challenging to figure out what shapes and colours work best, where to put them, and how to size them.

Analysing relationships in our data can be challenging and it requires more critical thinking than blindly graphing measures and dimensions. But it can also be the most rewarding and informative method for us, because we will be able to gain an actionable insight by understanding the relationships in our data. It is how our data relate and interact among themselves. By knowing the relationships in our data, we will be able to tell the best stories behind the data.

Tableau provides a best fit forecast function that data visualisation users can modify by selecting from a menu of available forecast options, including automatic, automatic with seasonality, trend and season, trend only, season only, or no trend or season.

Activity 6.1

1. Using the sample superstore dataset, create:
    a. Point-to-point map
    b. Scatter plot
    c. Bullet chart, parameter and bubble chart
    d. Pareto chart
    e. Forecasting

Formative Assessment

Select the most suitable chart option for each of the examples given below:

1. You need to compare sales across locations on a map. Which is the most suitable chart?
    a. Bar chart
    b. Pie chart
    c. Map
    d. Area chart

2. A logistics company wants to plot the route of its cargo vehicles from starting point to destination in Singapore. Which is the most suitable chart to plot the routes?
    a. Point-to-point map
    b. Line chart
    c. Map
    d. Side-by-side bar chart

3. In a GMAT applicant pool, the selectors wants to check if the GMAT score has a positive relationship with the work experience of the applicant. Which chart should they pick to showcase the relationship between the GMAT score and work experience?
    a. Spider map
    b. Scatter plot
    c. Bubble chart
    d. Grantt chart

4. During the year-end review of the sales team, the manager wanted to see how the sales team performed in terms of actual sales when compared to projected sales at

the beginning of the year (target set at the start of the year). Which chart should the sales manager choose?

a. Line chart
b. Horizontal bar graph
c. Line chart with a reference line
d. Bullet chart

5. During a strategy meeting, the CEO of the firm wanted to see which top 5 products among the 20 products in the firm's portfolio are in terms of profitability and how much they contribute to overall profits. Which chart should the CEO look at?

a. Pie chart
b. Bubble chart
c. Pareto chart
d. Line chart with reference line

Solutions or Suggested Answers

Formative Assessment

Select the most suitable chart option for each of the examples given below:

1. You need to compare sales across locations on a map. Which is the most suitable chart?

    a. Bar chart

    Incorrect.

    b. Pie chart

    Incorrect.

    c. Map

    Correct. Refer to Table 5.1 for explanation.

    d. Area chart

    Incorrect.

2. A logistics company wants to plot the route of its cargo vehicles from starting point to destination in Singapore. Which is the most suitable chart to plot the routes?

    a. Point-to-point map

    Correct. Refer to Table 5.1 for explanation.

    b. Line chart

    Incorrect.

    c. Map

    Incorrect.

    d. Side-by-side bar chart

Incorrect.

3. In a GMAT applicant pool, the selectors wants to check if the GMAT score has a positive relationship with the work experience of the applicant. Which chart should they pick to showcase the relationship between the GMAT score and work experience?

   a. Spider map
      Incorrect.

   b. Scatter plot
      Correct. Refer to Table 5.1 for explanation.

   c. Bubble chart
      Incorrect.

   d. Grantt chart
      Incorrect.

4. During the year-end review of the sales team, the manager wanted to see how the sales team performed in terms of actual sales when compared to projected sales at the beginning of the year (target set at the start of the year). Which chart should the sales manager choose?

   a. Line chart
      Incorrect.

   b. Horizontal bar graph
      Incorrect.

   c. Line chart with a reference line
      Incorrect.

   d. Bullet chart

Correct. Refer to Table 5.1 for explanation.

5. During a strategy meeting, the CEO of the firm wanted to see which top 5 products among the 20 products in the firm's portfolio are in terms of profitability and how much they contribute to overall profits. Which chart should the CEO look at?

   a. Pie chart

      Incorrect.

   b. Bubble chart

      Incorrect.

   c. Pareto chart

      Correct. Refer to Table 5.1 for explanation.

   d. Line chart with reference line

      Incorrect.

References

LTA, SMRT map https://www.lta.gov.sg/content/ltagov/en/getting_around/public_transport/rail_network.html

Murray, D. G. (2016). *Tableau your data: Fast and easy visual analysis with Tableau software. (2nd Edition)* Indianapolis, IN, USA: John Wiley & Sons Inc.

Pareto, V. (1960). Elites, Force, and Governments. *Images of Man: The Classic Tradition in Sociological Thinking*, 262-91.

Tableau Online Help. (2014). Retrieved July 1, 2014, from http://onlinehelp.tableausoftware.com/current/pro/online/mac/en-us/help

Yau, N. (2013). *Data points: Visualisation that means something*. Indianapolis, IN, USA: John Wiley & Sons.