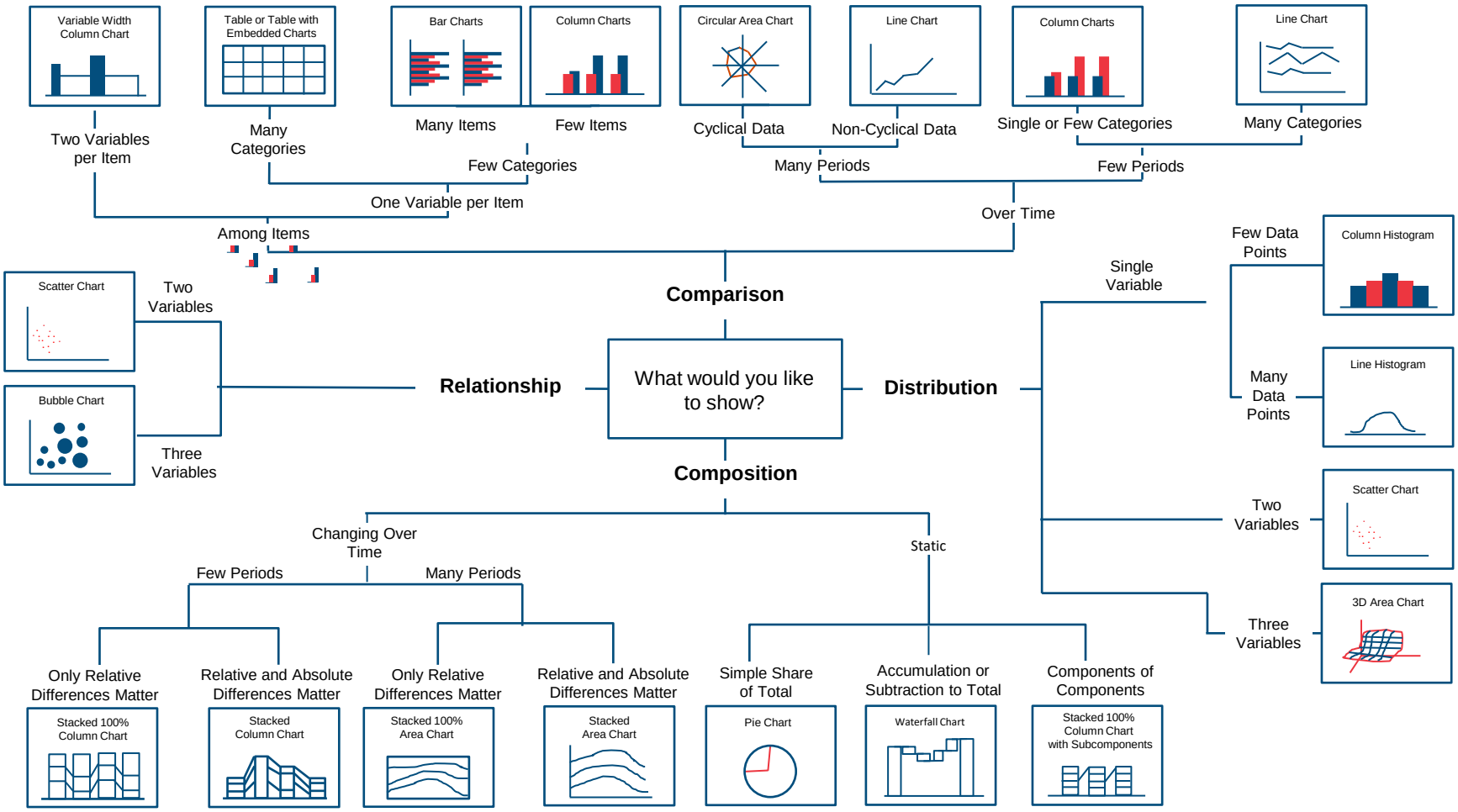
Data Visualization For Expert

Course Objectives

- Improve use of visualization to enhance your decision making
- Understand advanced concepts in data visualization and Dashboarding and corresponding best practices.
- Develop practically applicable skills through group discussion about how to design an effective data visualization solution.
- Appreciate different popular tools on the market: Excel PivotCharts, Qlik, PowerBI, Tableau.
- Appreciate real-world scenarios in end to end implementation process of extracting insights effectively from dashboards.

Leading practices in data visualization

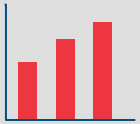
Leading practices: Chart Selection

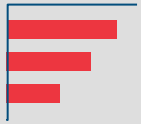


Guideline: Based on the Form of Data Analysis

Comparison of Data

- Identify the highs and lows of data
- Use case: Who is the best sales person?


Column Chart


Bar Chart

Transition of Data

- Understand the trend of change from time-based data
- Use case: Website traffic over the past 30 days

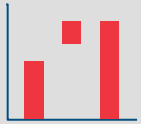

Line Chart


Area Chart

Composition of Data

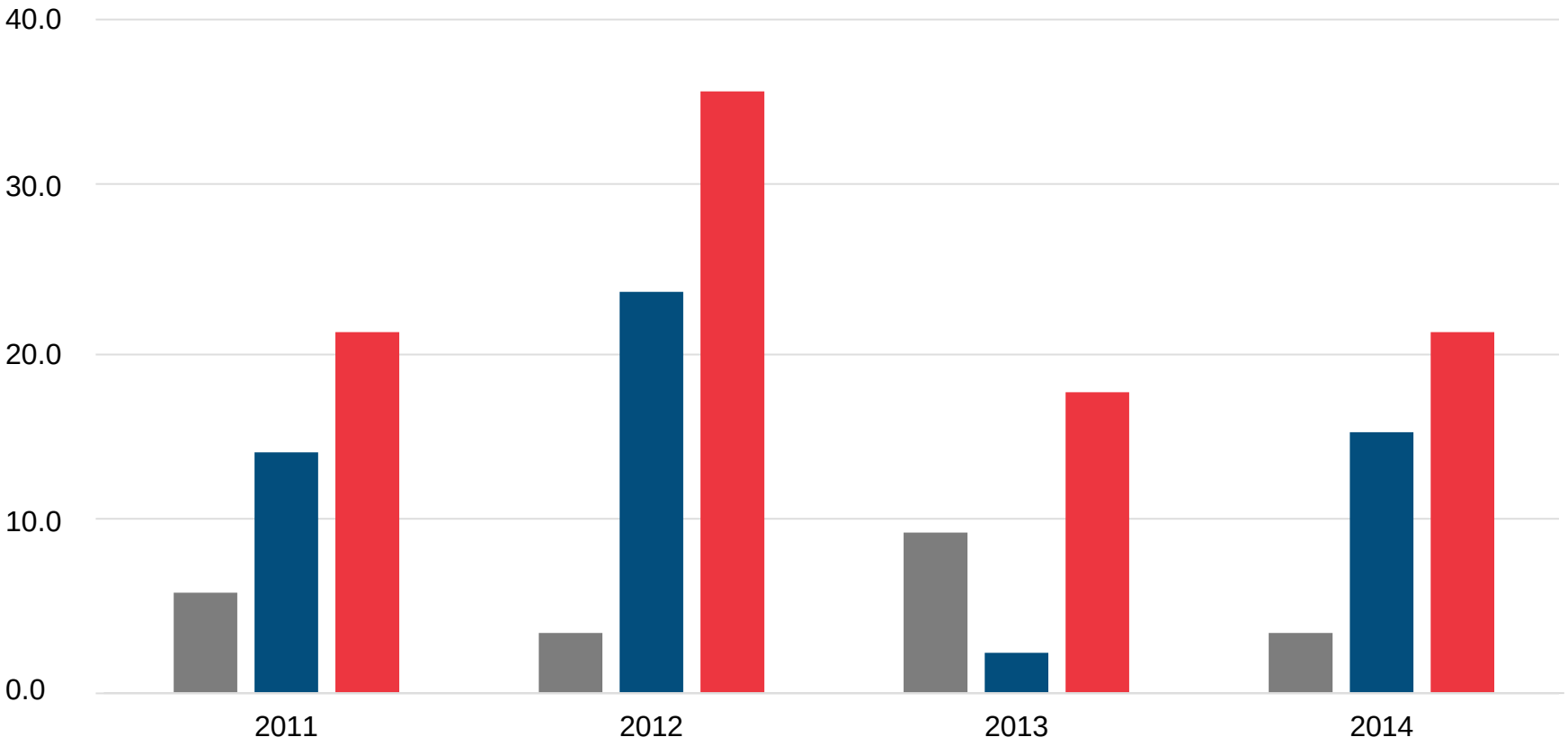
- Understand the breaks down of a data value into its constituents
- Use case: Breakdown of the website traffic


Pie Chart

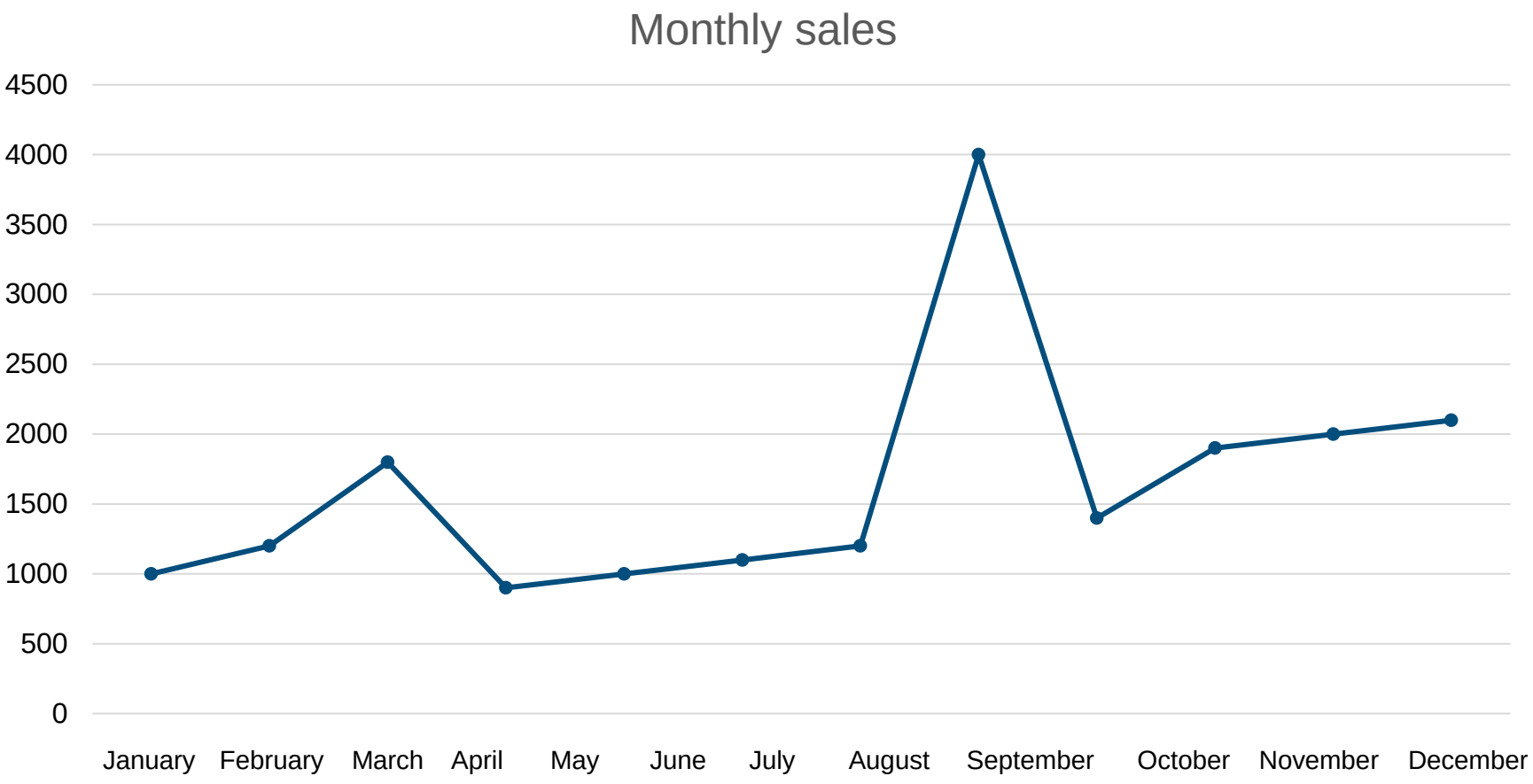

Waterfall Chart

Usage of Different Chart

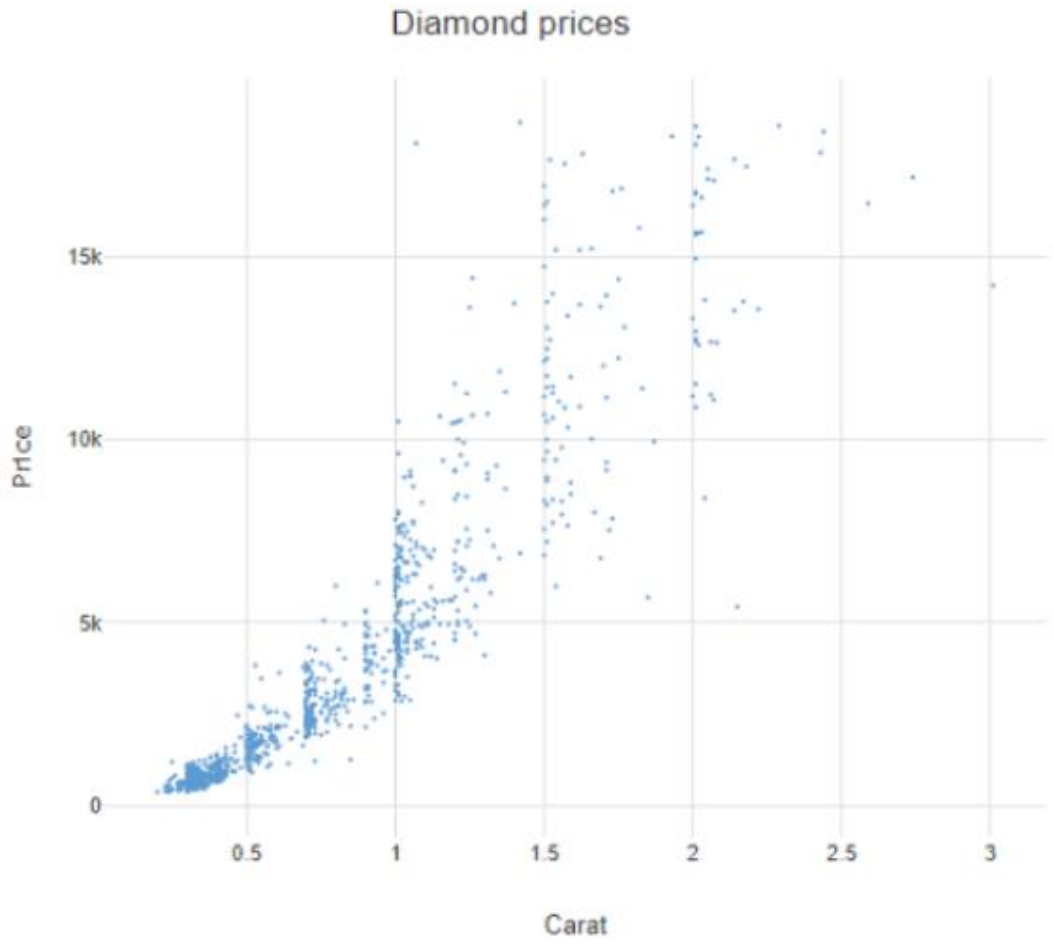
Bar chart



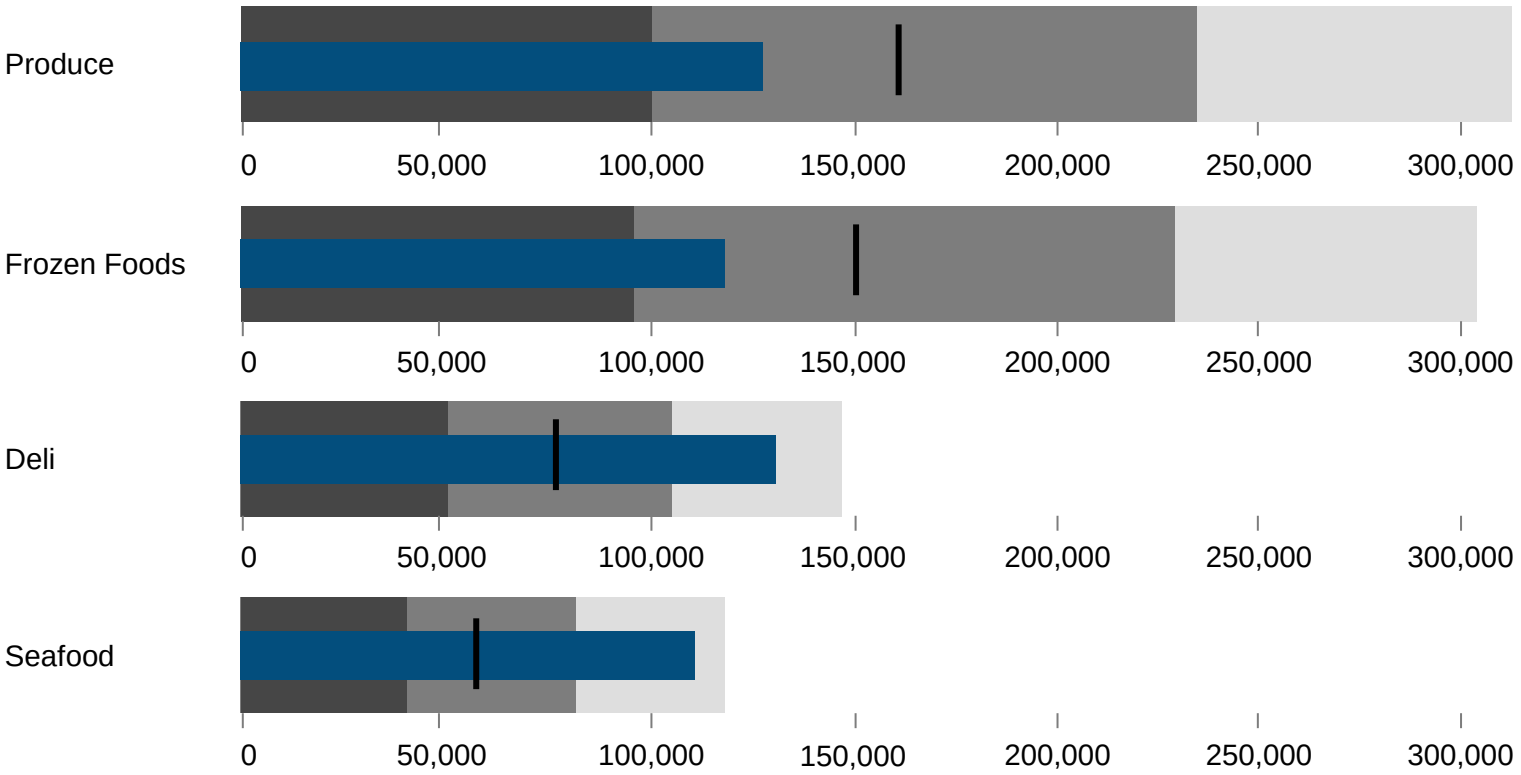
Line chart


Monthly sales
4500
4000
3500
3000
2500
2000
1500
1000
500
0
January
February
March
April
May
June
July
August
September
October
November
December

Scatter plot



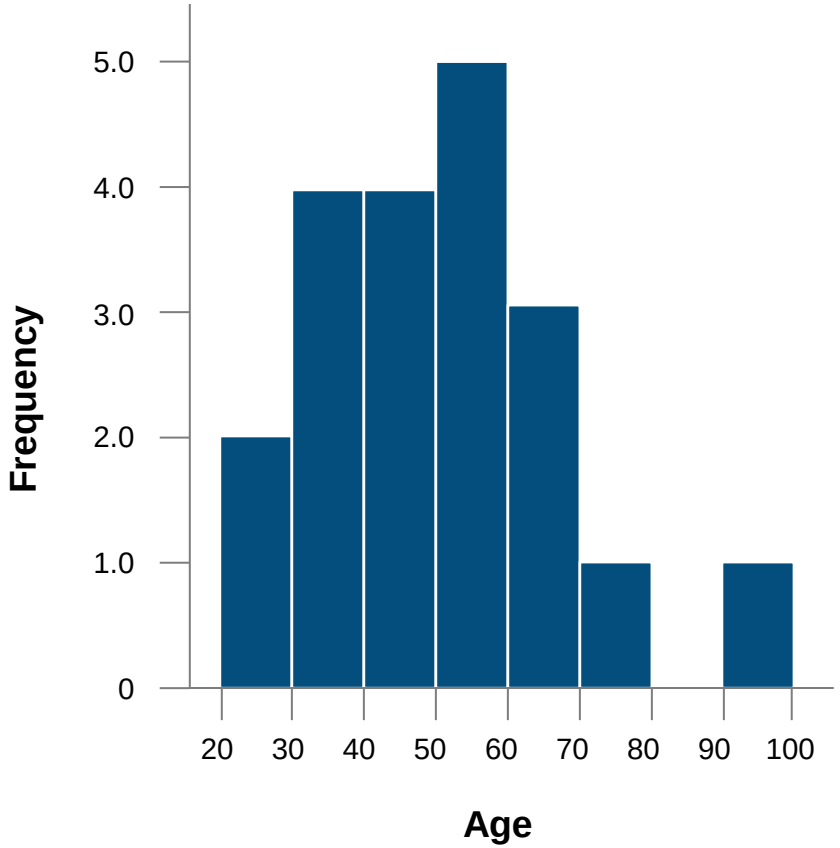
Bullet Chart



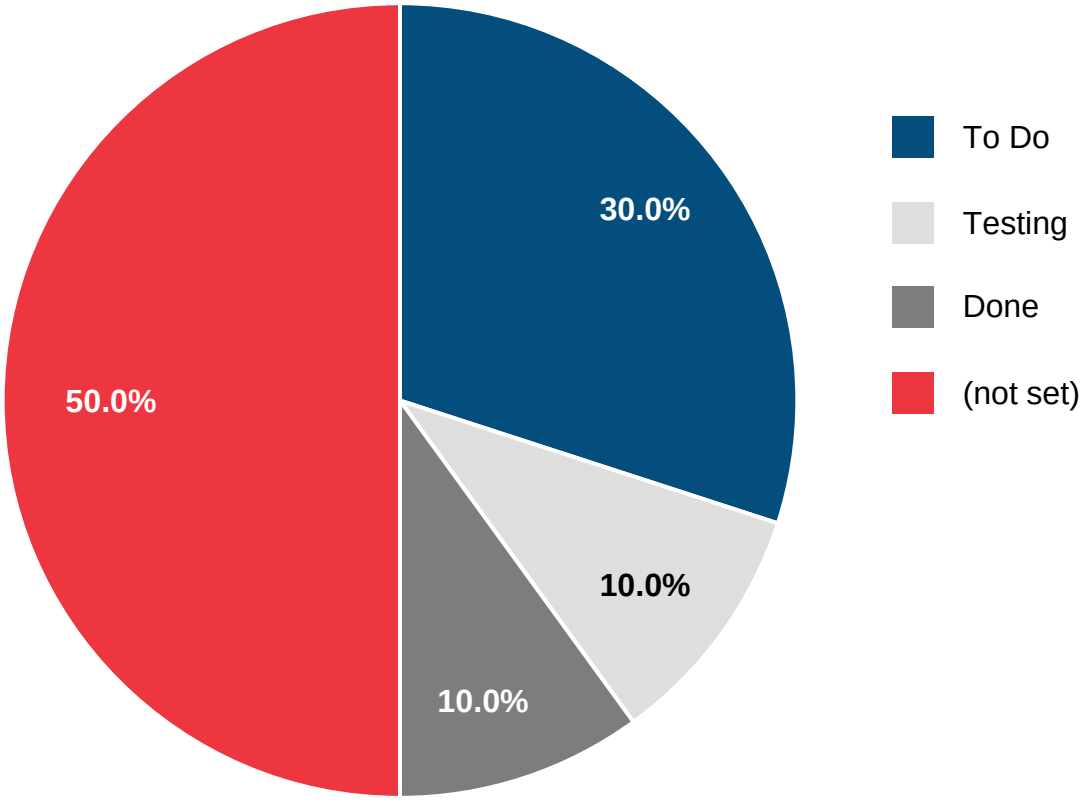
Highlight Tables

Order Date				
Container	2010	2011	2012	2013
Jumbo Box	($11,512)	($13,540)	$1,874	($20,775)
Jumbo Drum	$60,518	$65,472	$89,956	$96,210
Large Box	$5,971	$2,892	$47,094	$9,533
Medium Box	$7,195	$26,747	$28,581	$17,941
Small Box	$138,709	$195,976	$167,860	$311,776
Small Pack	$8,361	$15,881	$13,980	$28,940
Wrap Bag	$4,082	$4,421	$4,728	$3,572

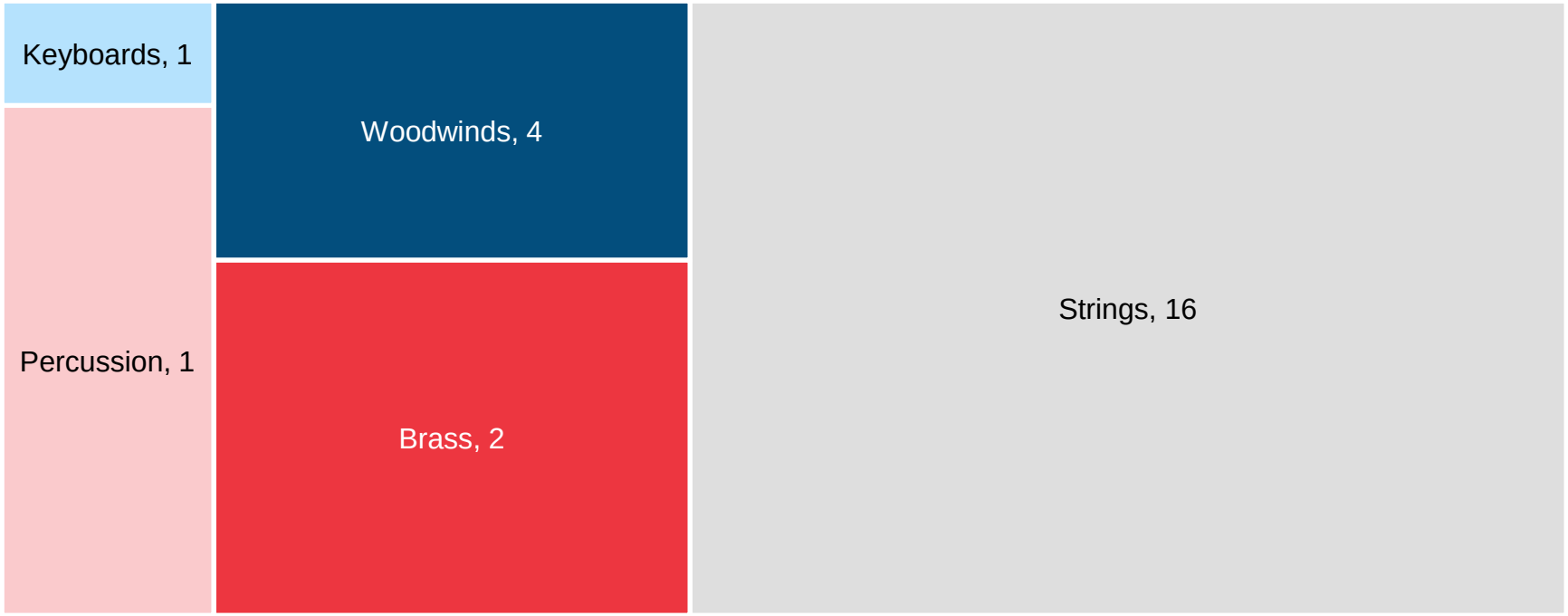
Histogram



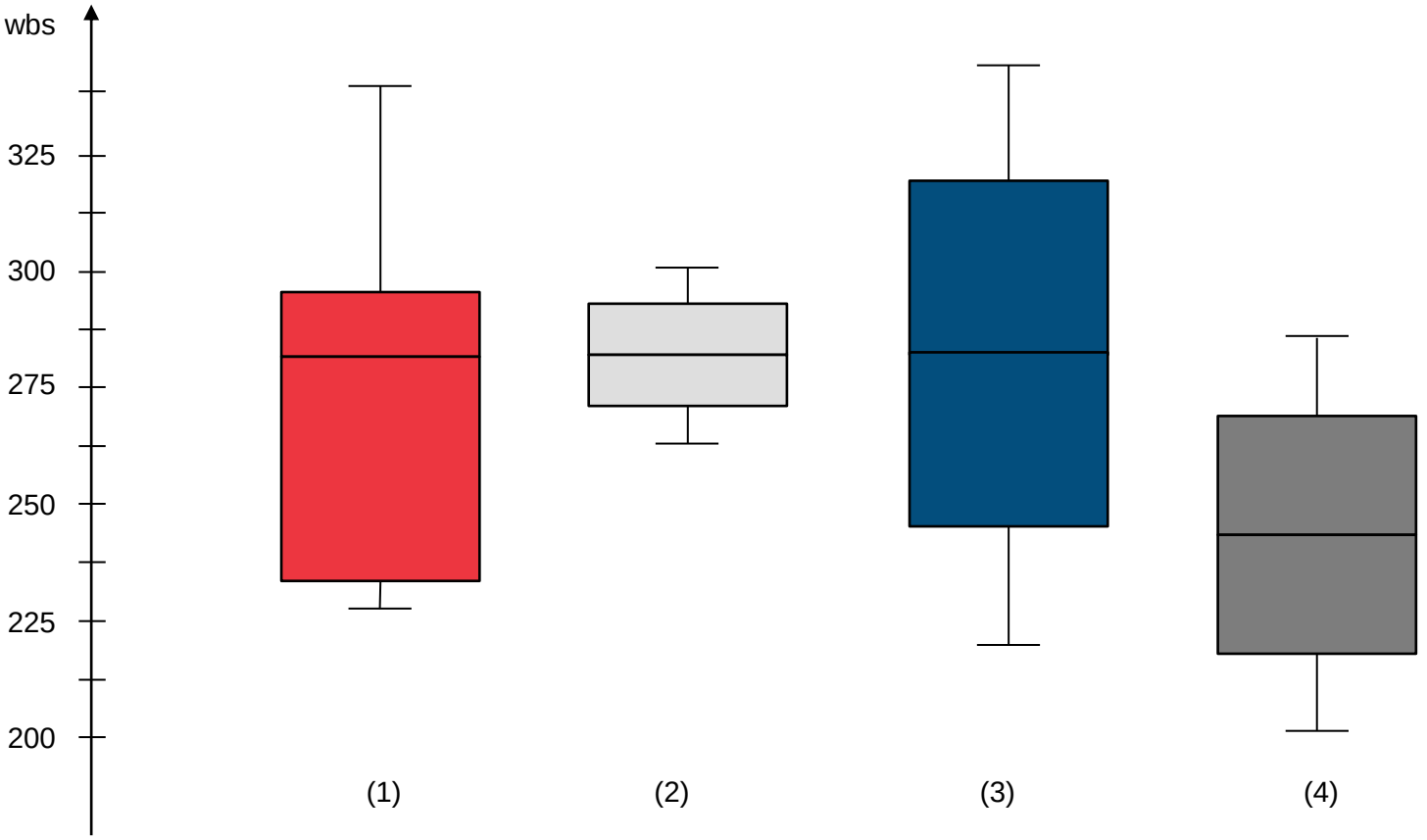
Pie Chart



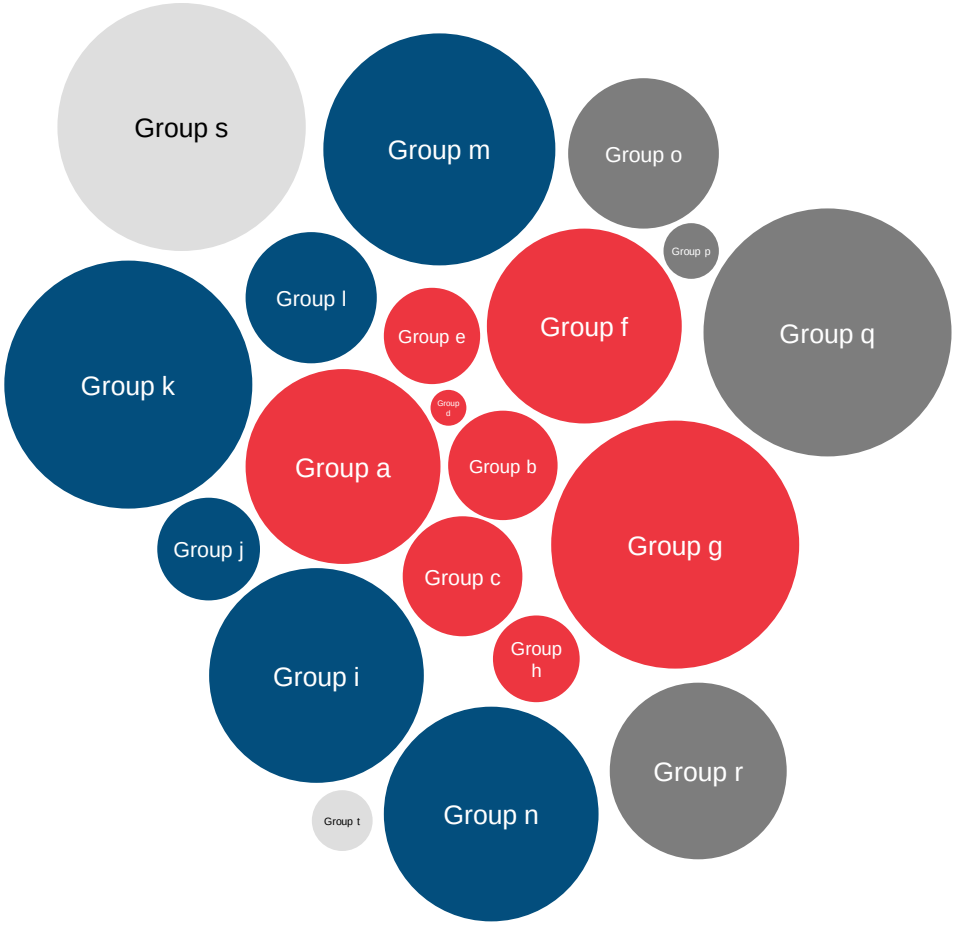
Treemap



Box Plot



Bubble Chart



Exercise – Selecting the right chart

Scenario	Type of charts
Sales trending over time	
Key Performance Indicator to measure proportion of orders being dispatched on time	
Project budget remaining at a point in time	
Percentage of staff passing and failing exams	
Number of males and females being promoted	

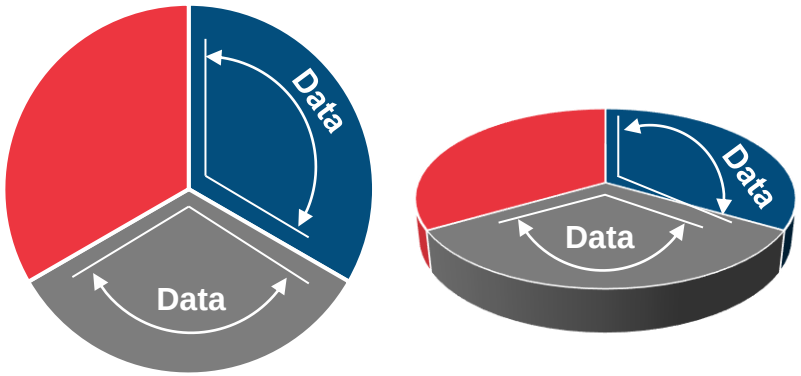
Exercise – Selecting the right chart (continued)

Scenario	Type of charts
Sales trending over time	Line chart
Key Performance Indicator to measure proportion of orders being dispatched on time	Gauge chart
Project budget remaining at a point in time	Funnel chart
Percentage of staff passing and failing exams	Pie chart
Number of males and females being promoted	Bar chart

Leading practices: Chart Designing

- Avoid the use of visual representations that don't accurately represent the data set, like pie charts in 3D.
- SAME but look DIFFERENT!

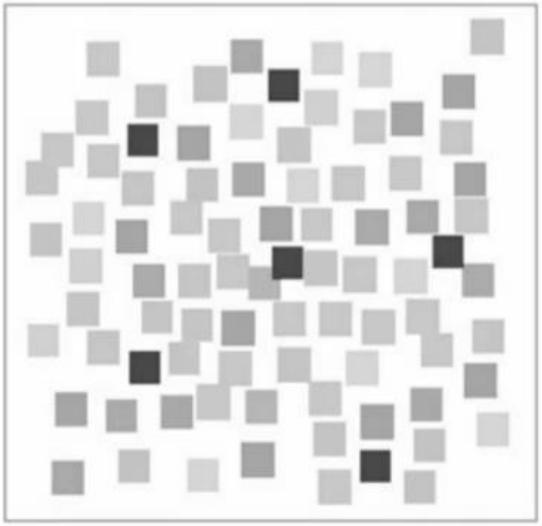
Angle

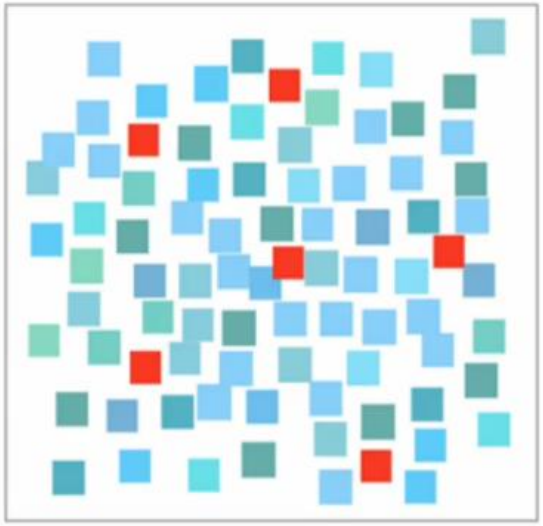


- Using sufficiently large font sizes and adequate contrast between type and the background are also helpful
- Contrast separates

Leading practices: Chart Color

- Do not use multiple colours to represent the same type of data
- Complement the use a different shade, or occasionally a different colour to convey different types of information
- Beware of how colours are perceived to people who are Colour Blind and in printing
- Use colors that have high contrast
- Use text or icons to label elements





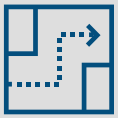
Leading practices: Dashboard designing (continued)

Avoid clutter

Remove as much as possible while ensuring the end user gets the right insight from the dashboard

Trick:

1) Does it have too much on it?
2) Is there anything you can remove or rearrange to add clarity?

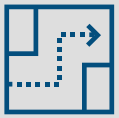


Use a grid layout

A grid helps provide a reading order for the dashboard, allowing the users to guide themselves through the dashboard in a predictable and logical way

Trick:

Use a columnar- or row-based flow to create a narrative that leads users from overview to detail



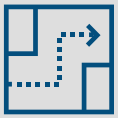
Leading practices: Dashboard designing (continued)

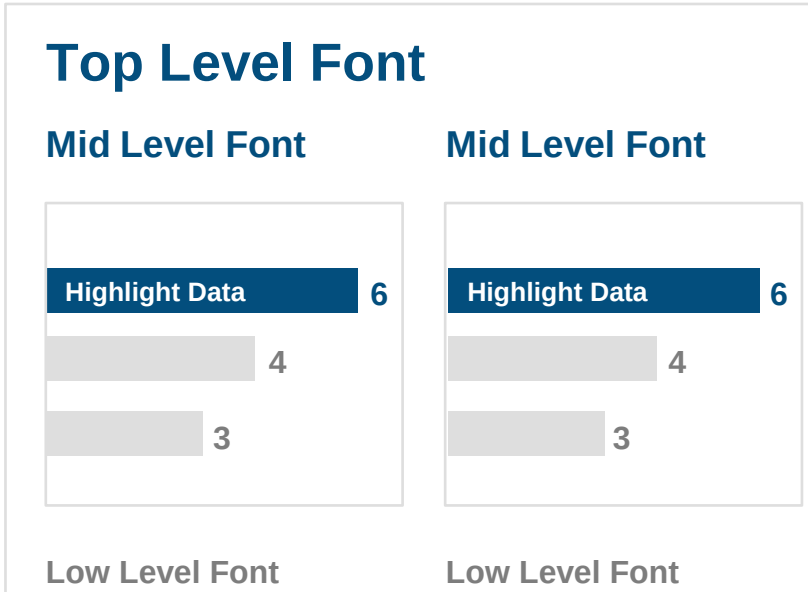
Use the right fonts

Remove as much as possible while ensuring the end user gets the right insight from the dashboard

Trick:

1) Do not use too many font types and sizes in a dashboard
2) Define a clear hierarchy for your typography





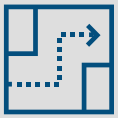
Leading practices: Dashboard designing (continued)

Simplify the use of color

Too many colors can create visual overload for your users and impede analysis

Trick:

1) Choose palettes that work well universally, avoiding reds and greens
2) Use neutral colors with a single, bright color to highlight what you want your viewers to pay attention to

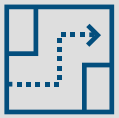


Show it big

Hits users with the most vital information immediately, and charts provide further context.

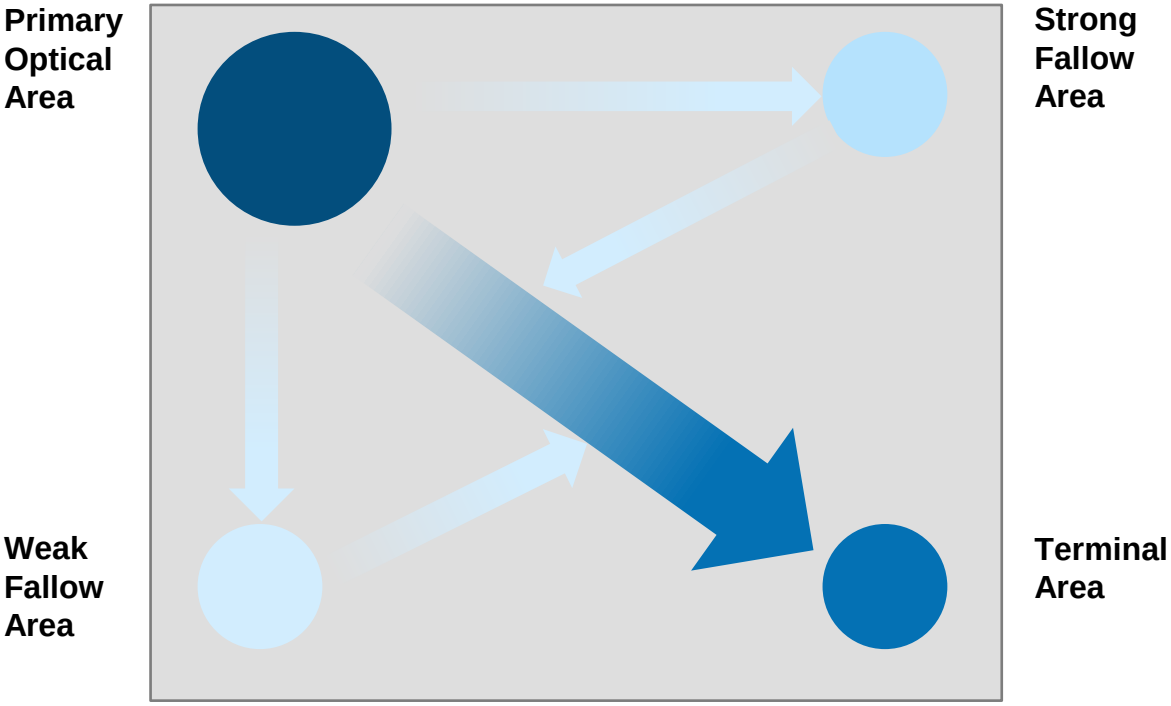
Trick:

Show the most important fact, i.e. KPI as big number in the dashboard



Leading practices: Dashboard designing (continued)

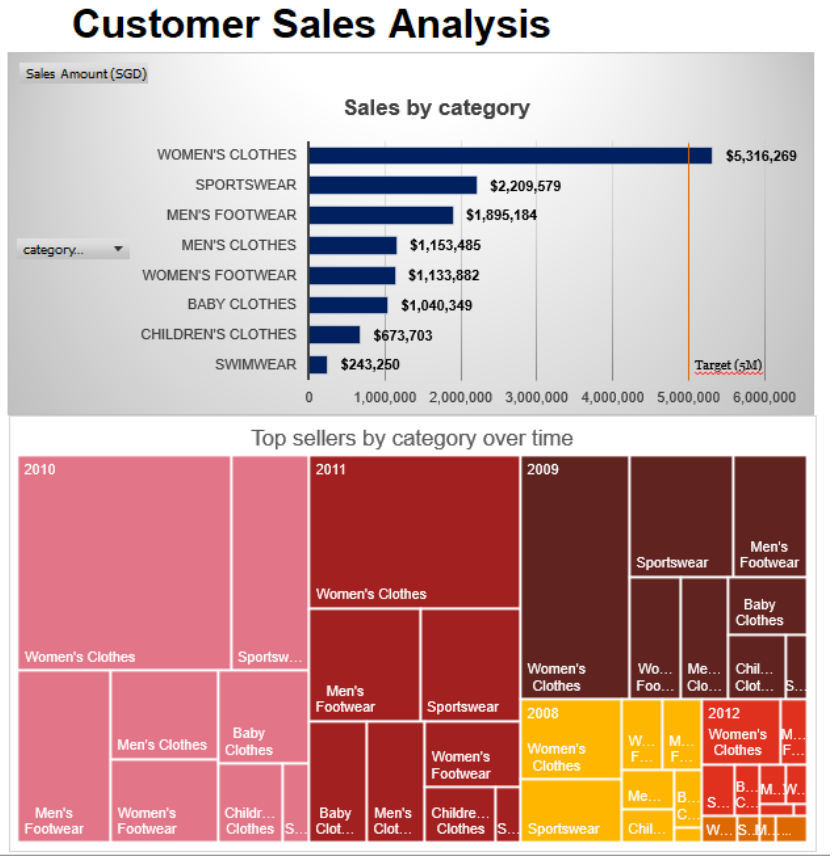
Reading Gravity



Demo

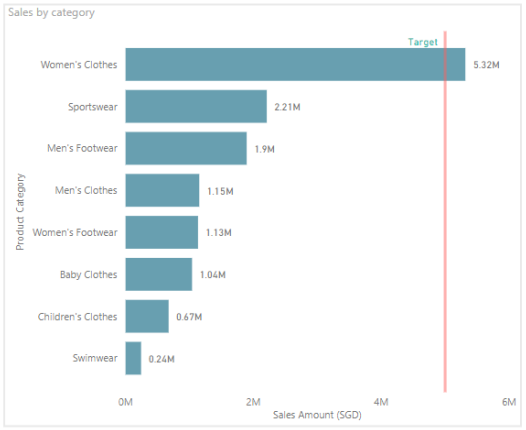
Demo of Microsoft Excel

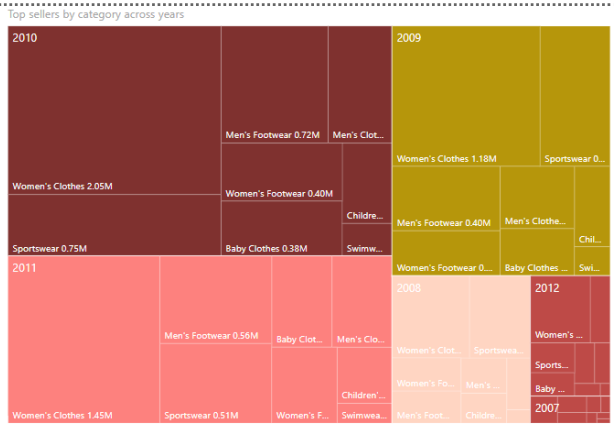
Microsoft Excel is developed by Microsoft for insightful spreadsheets. It has features such as calculation, graphing tools, pivot tables, and a macro programming language called Visual Basic for Applications.

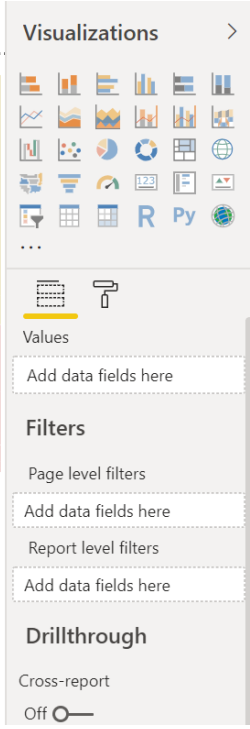


Demo of Power BI

Power BI is a business analytics solution that lets you visualize your data and share insights across your organization, or embed them in your app or website. Connect to hundreds of data sources and bring your data to life with live dashboards and reports.

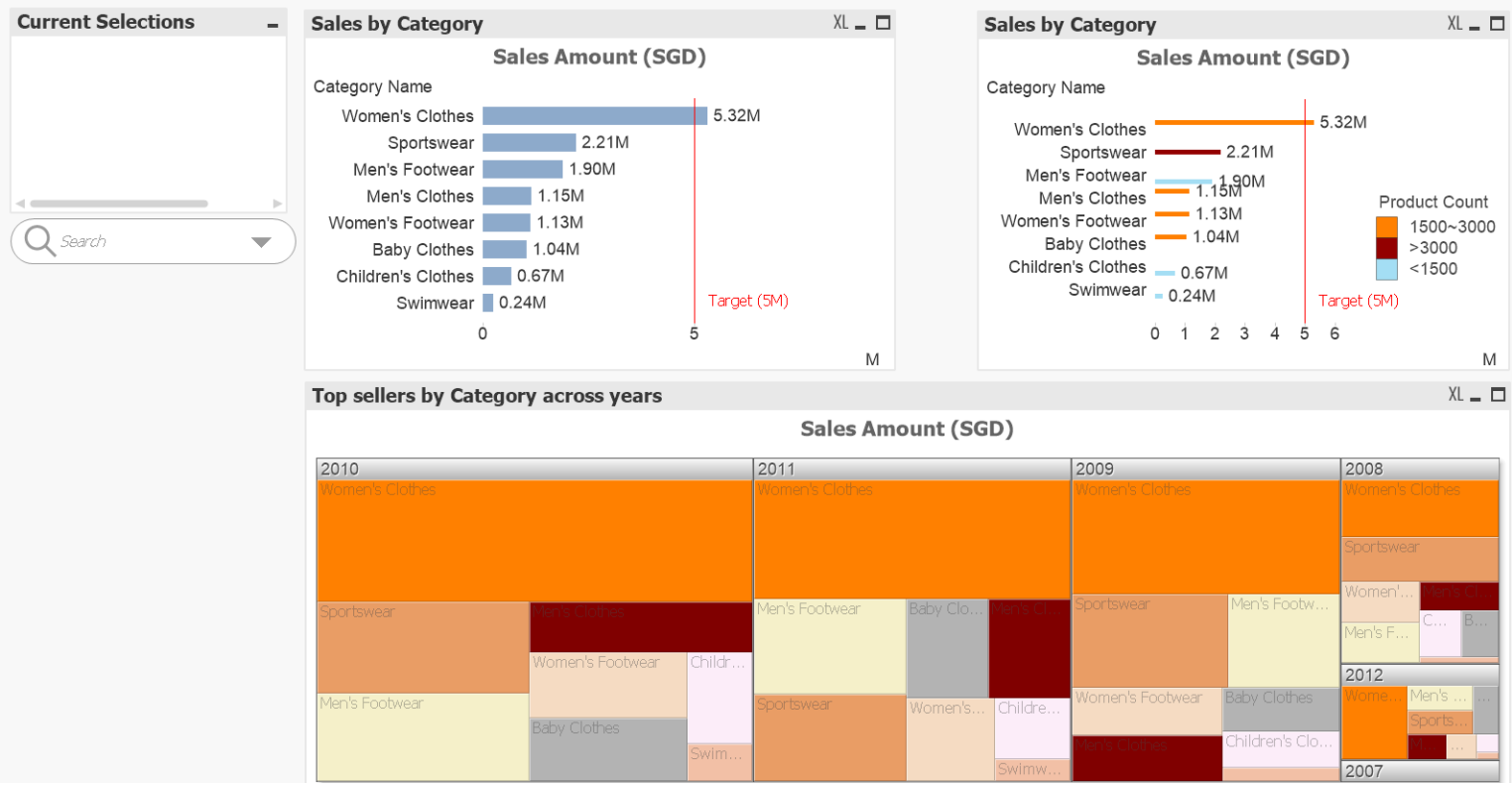






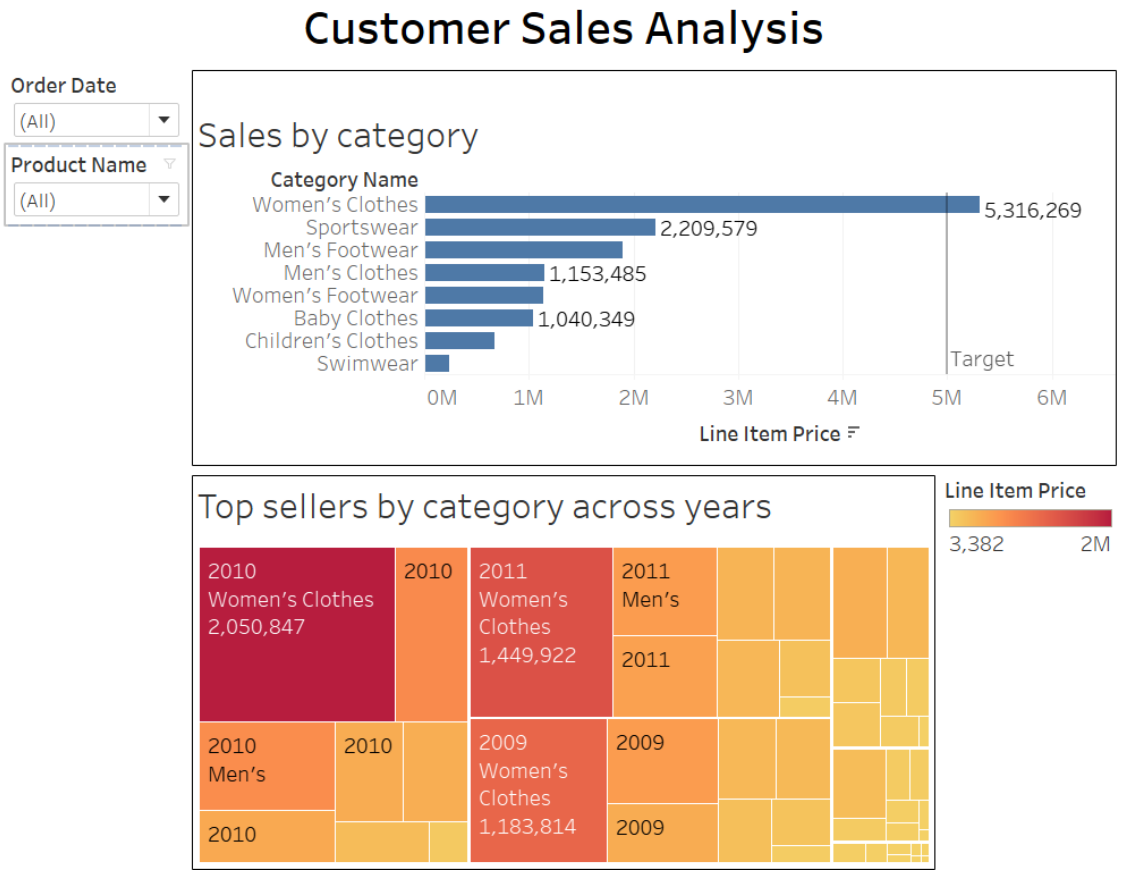
Demo of QlikView

QlikView is a business discovery platform that provides self-service BI for all business users in organizations.



Demo of Tableau

Tableau is business intelligence software that helps people see and understand their data.



Detailed visualization case study including hands-on session

Getting started

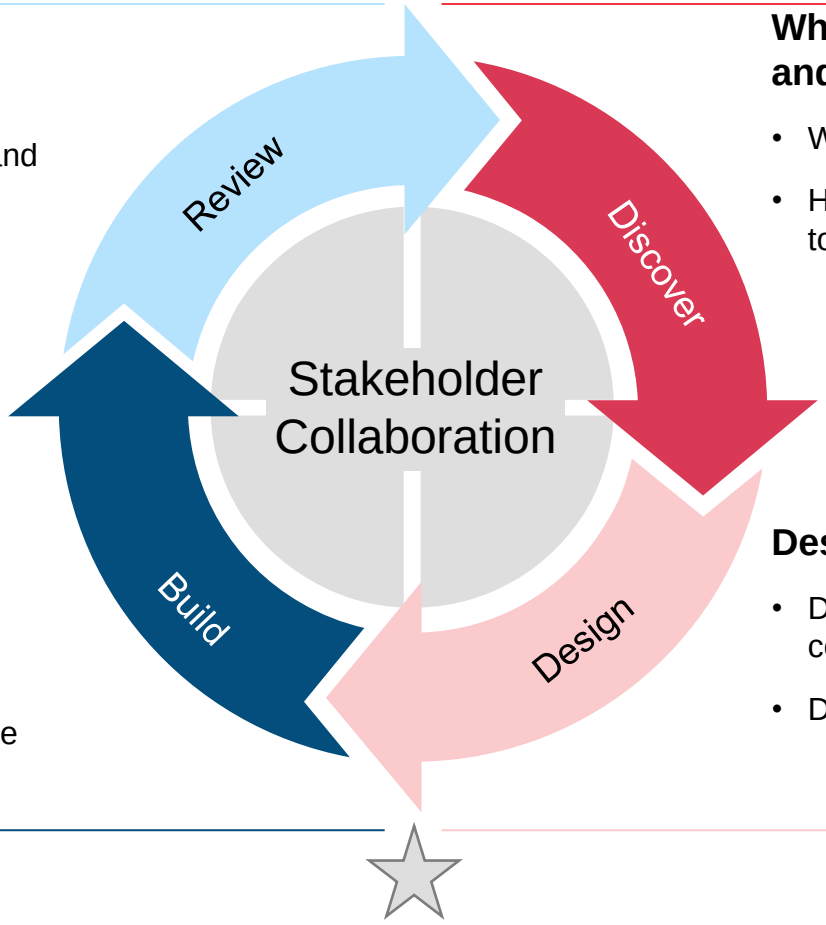
Data visualization project lifecycle

Review

- Perform testing
- Check labels, fonts, colours and alignments
- Review with client and refine requirements

What is the client challenge and visualization objective?

- What data is needed?
- How do data sources link together?



Build data model

- Develop template
- Build front end visualizations
- Enter comment code and write documentation

Design

- Data model and naming conventions
- Design front end visualizations

Discover: What data is needed?

Business problem

- A client is finding their current static sales reporting to be inflexible. They require an interactive solution to allow them to view different trends happening in the sales volume over a certain time by the different products, geographic areas, payment types and customer activity.

Client's requirement

- Sales geographic analysis
- Trend analysis covering a minimum of two years of detailed data.

What data is needed?

- Employee data
- Customer data
- Past 3 years' orders details

Design: What is the data model?

A data model refers to the logical inter-relationships and data flow between different data elements involved.

- Documents the way data is stored and retrieved
- Facilitate communication business and technical development
  - Accurately representing the requirements of the information system
  - Designing the responses needed for those requirements

Data keys

Keys are optimised when they are an integer type because integers are stored in a much more compact format than character formats and processed faster. This is because most primary keys will be foreign keys in another table.

Primary Key

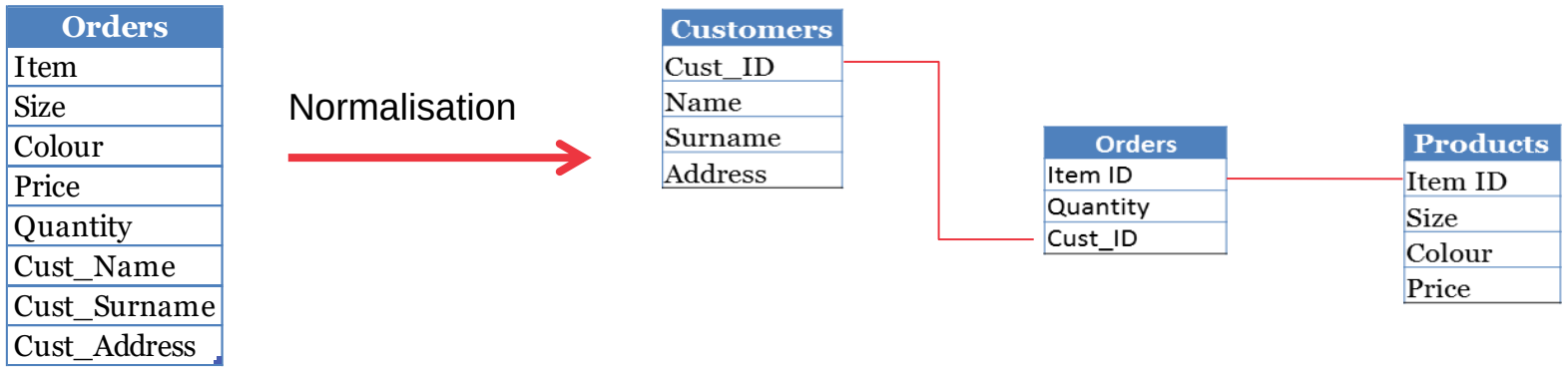
- Unique
- No duplicates
- No blanks/null

Foreign Key

- Link a record of one table to a primary key field of another table
- Define the relationship of records in different tables.

Normalisation and denormalisation

- Normalisation

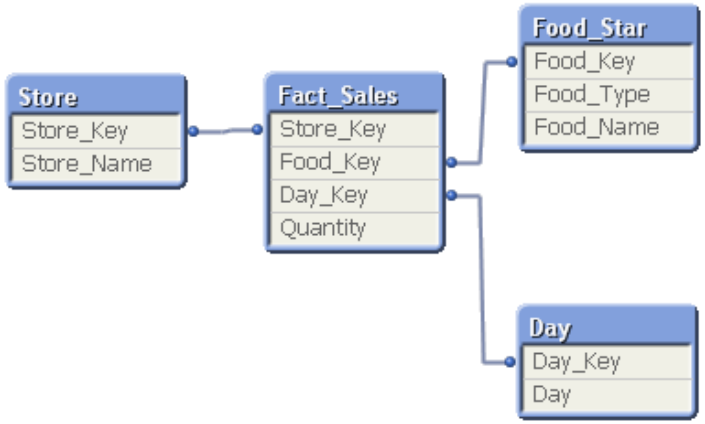


- Denormalisation
  - Can improve performance
  - Minimize the need of joins
  - Reduce the number of tables

Types of denormalised data models

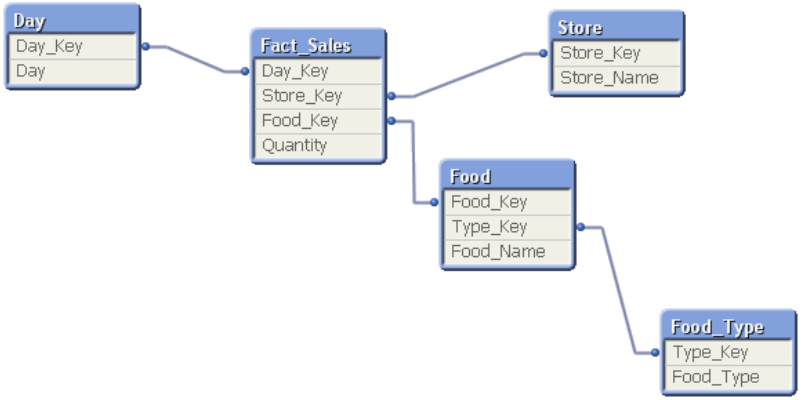
Star Schema

- One join between fact and dimension tables
- Ideal for quicker information retrieval as it has a denormalised structure.



Snowflake Schema

- More than one join between fact and dimension tables
- Sometimes partially denormalised structure can help performance



Data model challenges

Creates data associations between the fields you have imported

If two fields have the same name, they will be automatically associated between the datasets

This can be advantageous as it allows keys to be identified, however sometimes the fields should not be associated (for example ID in Products should not be associated to ID in Customers)

Association can be prevented by renaming fields or qualifying the tables

Data model challenges
Circular reference

A circular reference occurs when there are two or more routes to one table

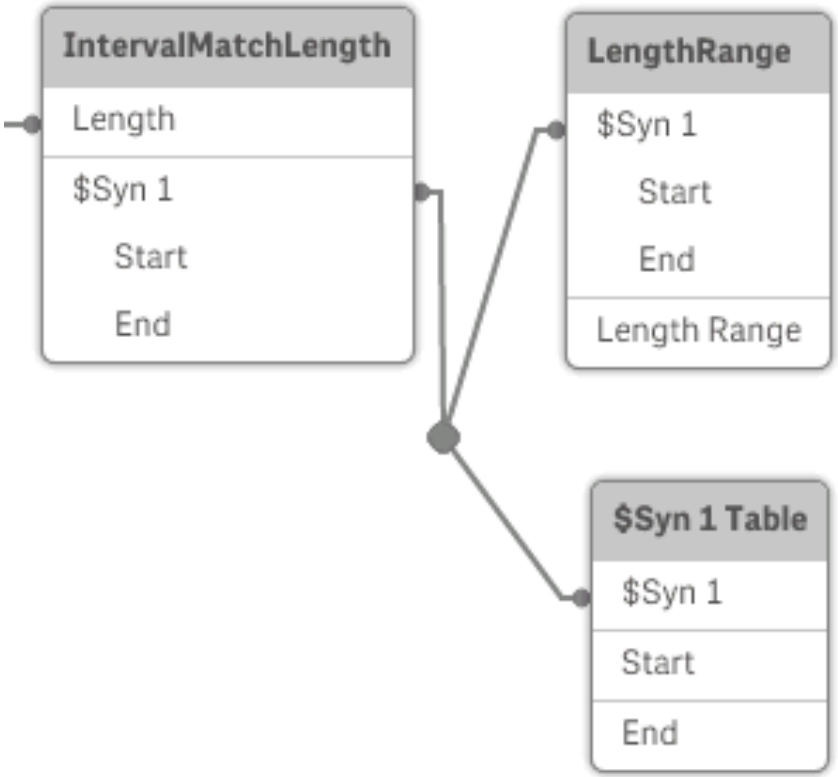
They cause data inconsistency and poor performance

Analyze the source data and talk to business users – what is the correct relationship between the tables?

Here, Gender in Employees table is not the same as Gender in Customers so one of these fields should be renamed.

Data model challenges
Synthetic key

- When two or more data tables have two or more fields in common, this suggests a composite key relationship. Qlik Sense handles this by creating synthetic keys automatically.
- They cause poor performance.
- Multiple synthetic keys are often a symptom of an incorrect data model, but not necessarily.
- Fields can be concatenated in order to remove the synthetic key.



Data model challenges
High cardinality fields

- These are fields with which have high level of uniqueness, for example time.
- Qlik Sense compresses data in fields by removing duplicate values. It performs better when there are fewer distinct values.
- An example, a timestamp field with date and time in one column has over 31M distinct possibilities.
- Evaluate your data model: is there data in one column that could be split across two?
- Separate date and time into their own fields during the data load. There are now 86k distinct possibilities.

Data model challenges
Multiple fact tables

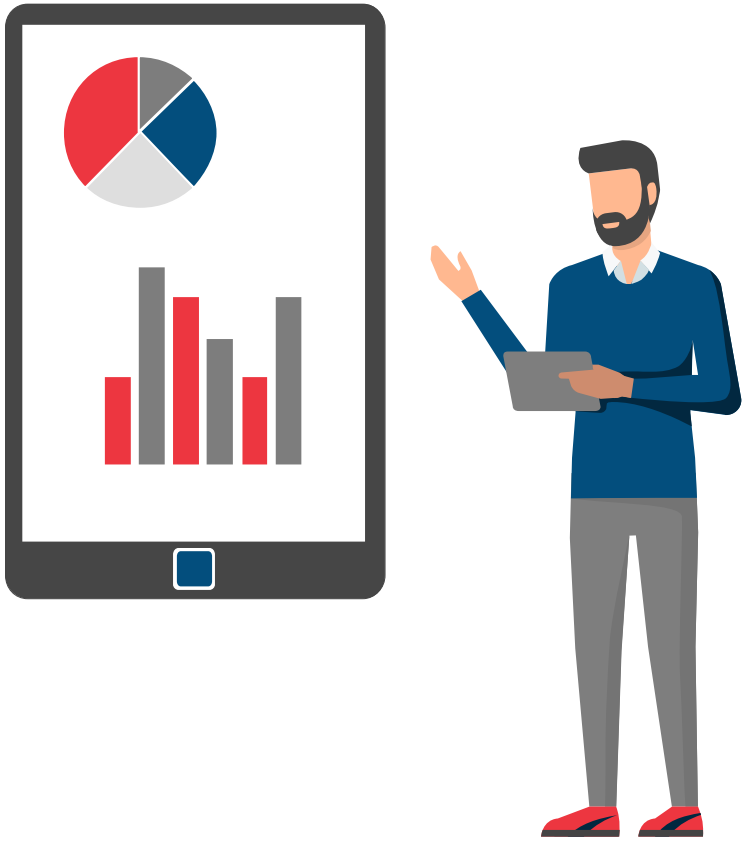
Often we see similar data stored in different tables, for example prior year financial data is stored in a separate data file or budgeted and actual data is stored in separate tables

These need to be concatenated into one Qlik Sense table

Qlik Sense will automatically concatenate tables with the same fields, unless you specify the No Concatenate option

What does a good data model look like?

- Converts real life data flows to dimensional models based on business processes
- Has a consistent and transparent naming convention
- Removes unused fields
- Uses number-based key fields for faster performance
- Eliminates all synthetic keys
- Eliminates any circular references
- Splits high cardinality fields
- Concatenates common fact tables
- Uses link tables to join multiple dimensional models
- Is flexible to future changes



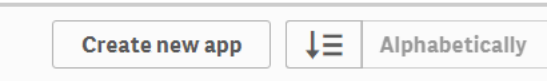
Build: How to build the data model?

Exercise 2
Data Import

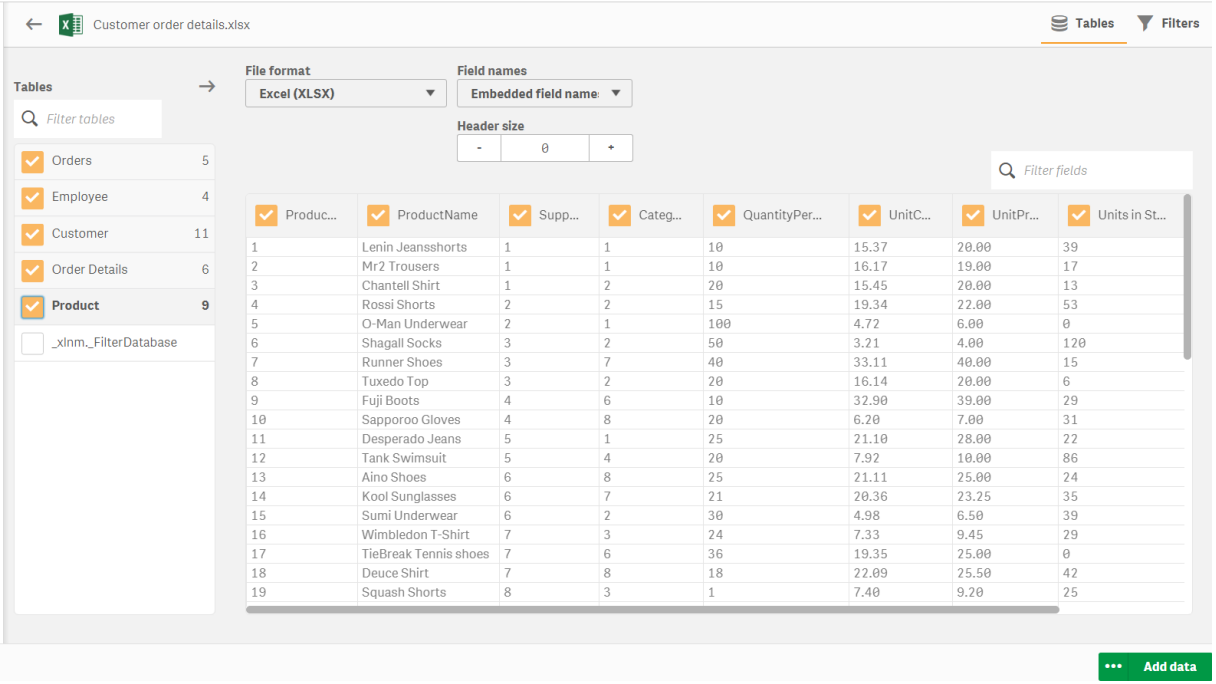
1. **Create new application** called ‘Customer Order Analysis’ in Qlik Sense.
2. **Add data to your app**: Click ‘add data from files and other sources’. Import *Orders, Employee, Customer, Order Details* and *Product* tab from Customer order details.xlsx.
3. **Add association to your data:** establish associations between the imported tables
4. **Load data**

Exercise 2 Solution
Data Import

1. **Create new application** called ‘Customer Order Analysis’ in Qlik Sense.



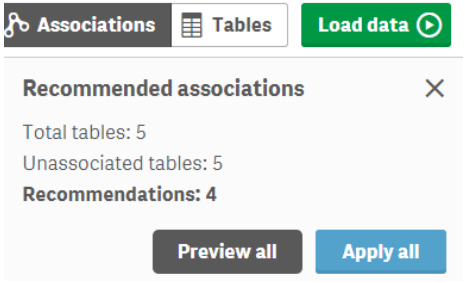
2. **Add data to your app**: Click ‘add data from files and other sources’. Import *Orders, Employee, Customer, Order Details* and *Product* tab from Customer order details.xlsx.

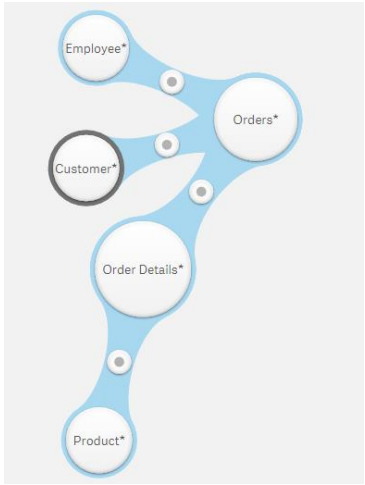


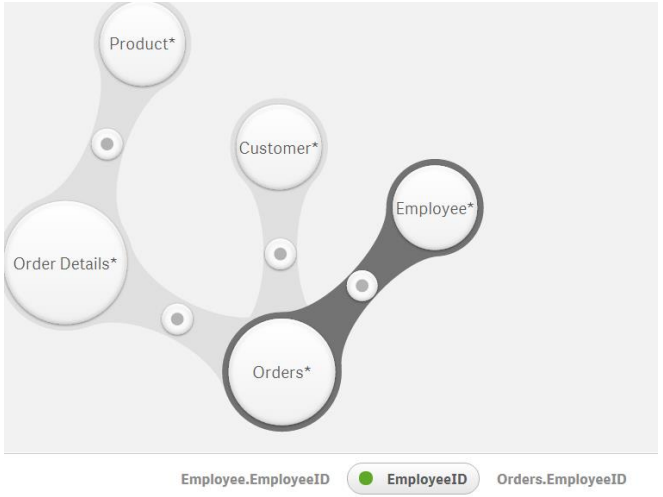
Exercise 2 Solution
Data Import

3. Add association to your data: establish associations between the imported tables

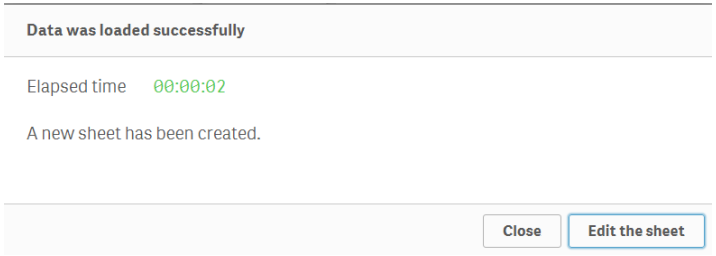
Tips: use recommended associations and apply all







4. Load Data

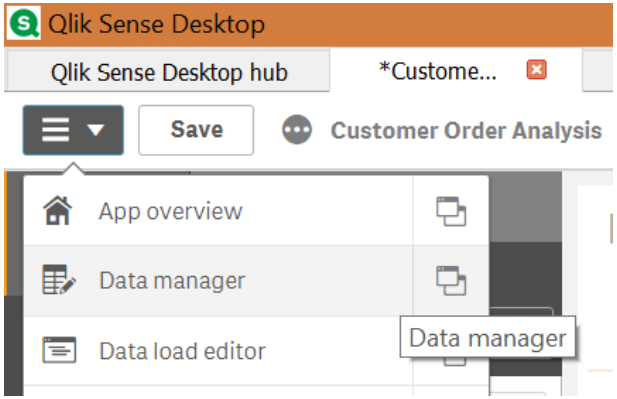


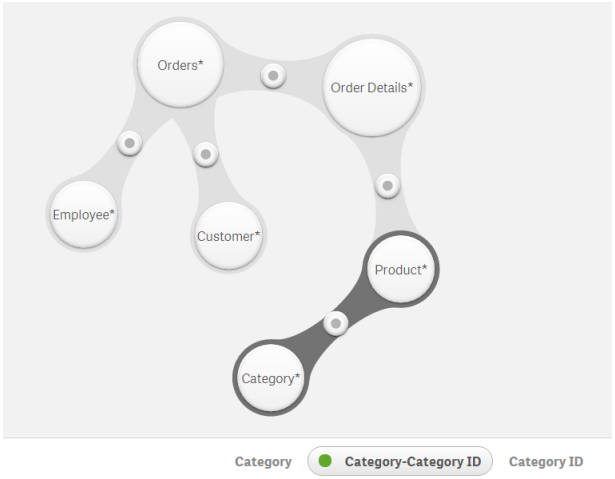
Exercise 3
Modify Data Model

1. **Add new data to your current app**: Go to ‘Data Manager’ and click ‘add data’. Import *Product Category mapping data*.xlsx.
2. **Add association to your data:** add the association between the new mapping data and the existing data
3. **Load data**
4. **Review your data model:** go to ‘data model view’

Exercise 3 Solution
Modify Data Model

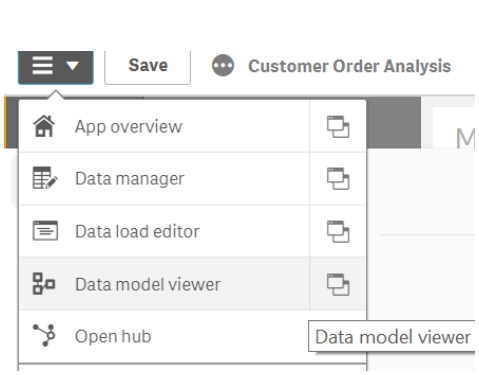
1. **Add new data to your current app**: Go to ‘Data Manager’ and click ‘add data’. Import *Product Category mapping data*.xlsx.
2. **Add association to your data:** add the association between the new mapping data and the existing data
3. **Load data**

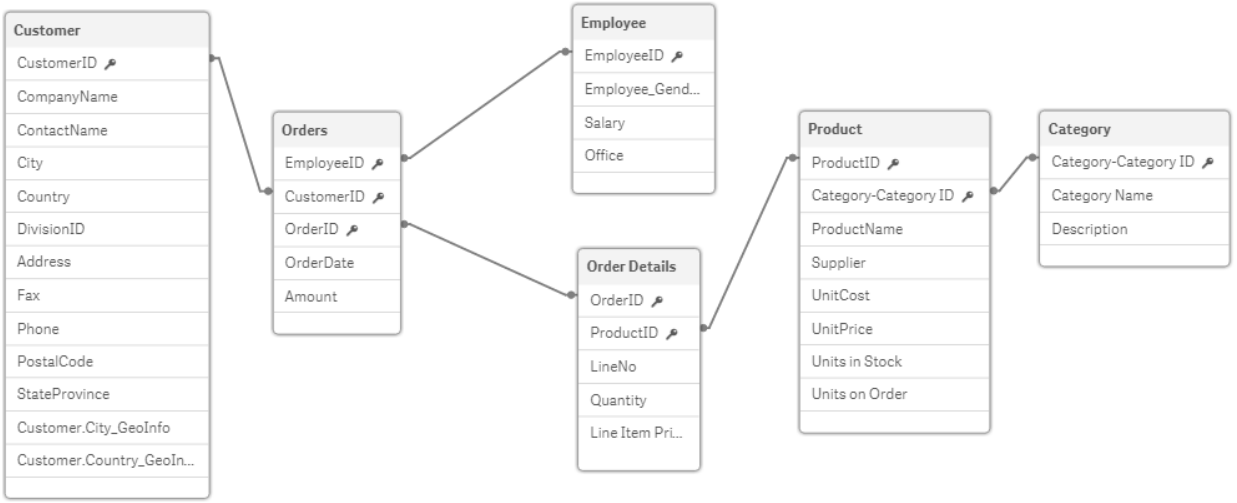




Exercise 3 Solution
Modify Data Model (continued)

4. Review your data model



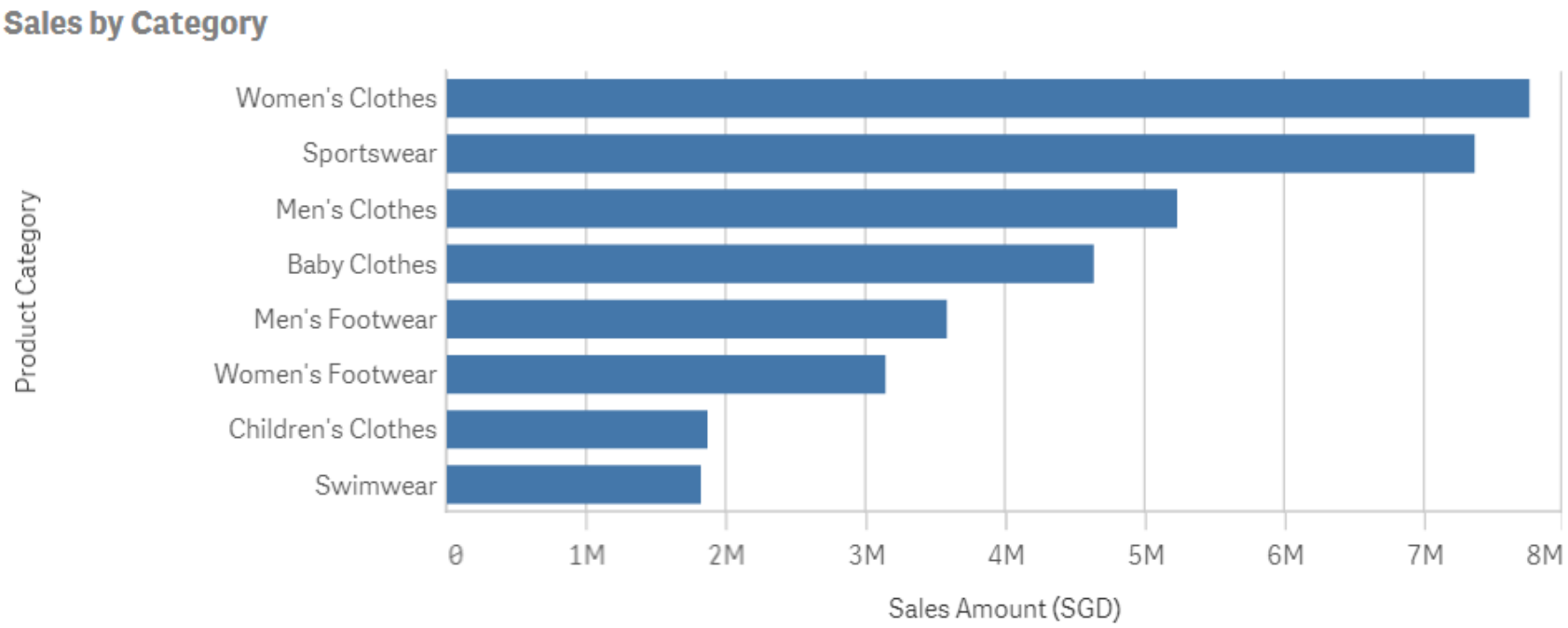


Review: How to review the application?

Exercise 4
Build a new sheet

1. **Create new sheet**: Click ‘create new sheet’ and edit the sheet title ‘Sales by Category’
2. **Select the required fields:** drag and drop fields *category name* and *amount.*
3. **Create a horizontal bar chart:** go to ‘appearance’ – ‘presentation’
4. **Rename axis label:** rename x-axis as ‘Sales Amount (SGD)’ and rename y-axis as ‘Product Category’
5. **Sort by sales amount descending:** go to ‘sorting’

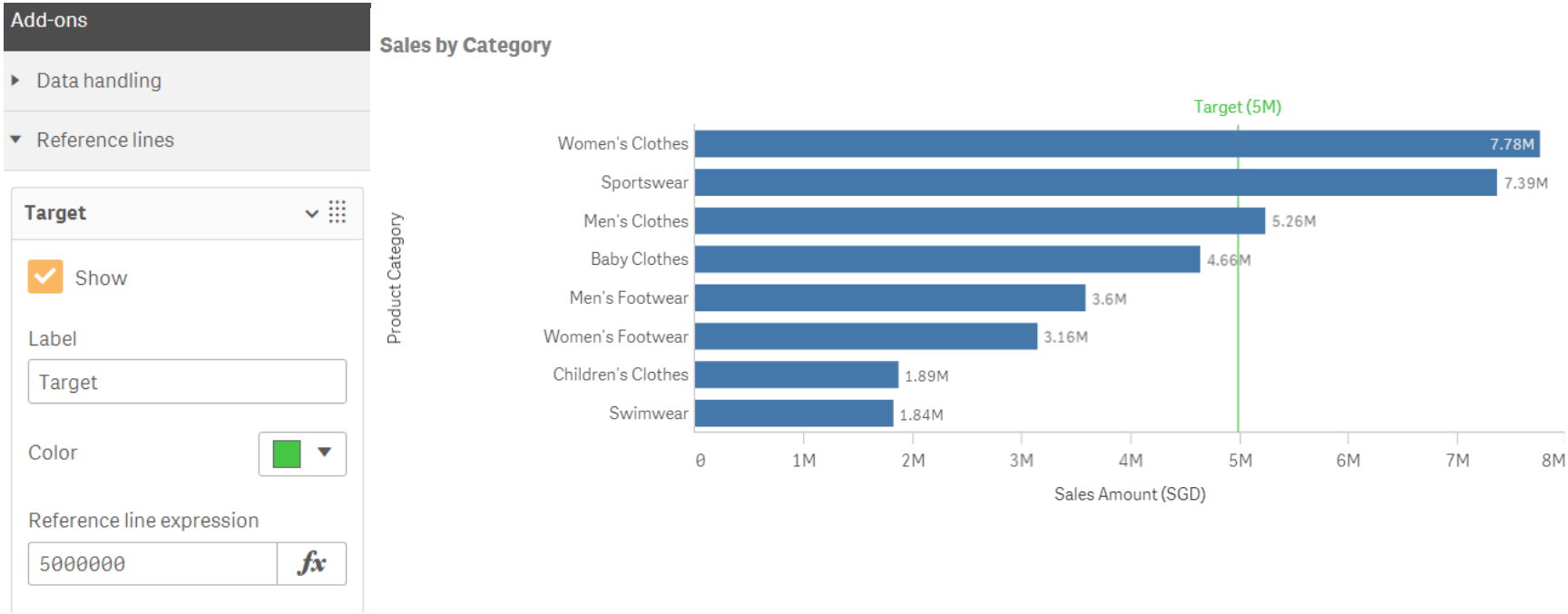
Exercise 4 Solution
Build a new sheet



Exercise 5
Add a reference line

1. **Mark a target sales with value of 5 million SGD:** go to ‘Add-ons’ – ‘Reference lines’
2. **Remove grid line in the pane:** to avoid visual confusion with the new reference line. Go to ‘appearance’ – ‘presentation’
3. **Show value on the bar:** go to ‘appearance’ – ‘presentation’ – ‘value labels’

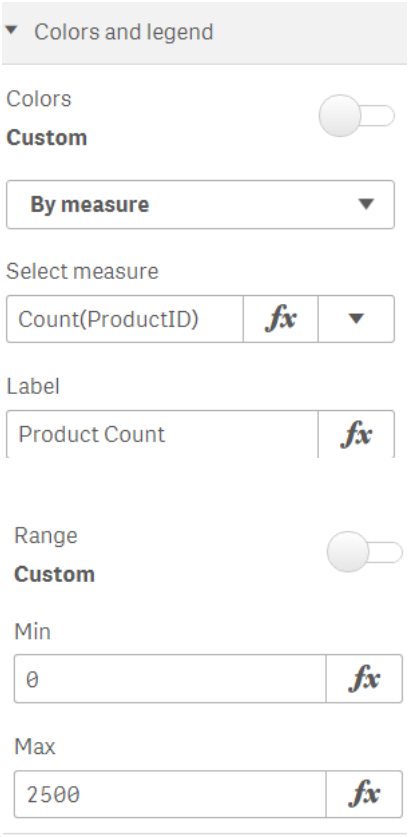
Exercise 5
Add a reference line (continued)

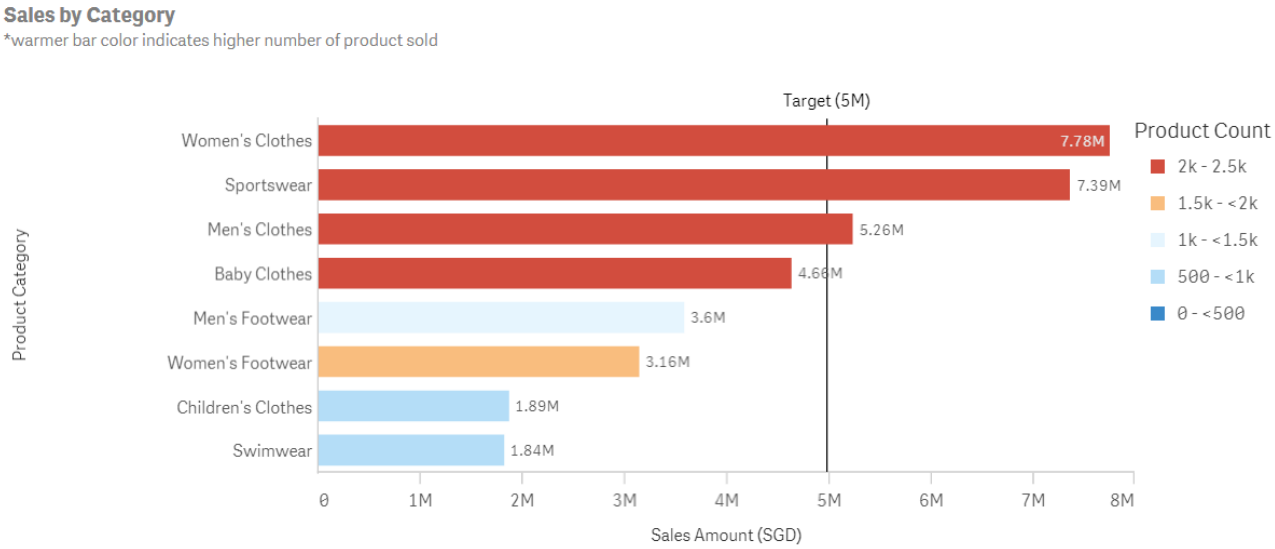


Exercise 6
Add Color and Legend

1. **Add the count of product to color legend:** go to ‘appearance’ – ‘color and legend’
2. **Adjust the range:** set custom range from 0 to 2500

Exercise 6 Solution
Add Color and Legend

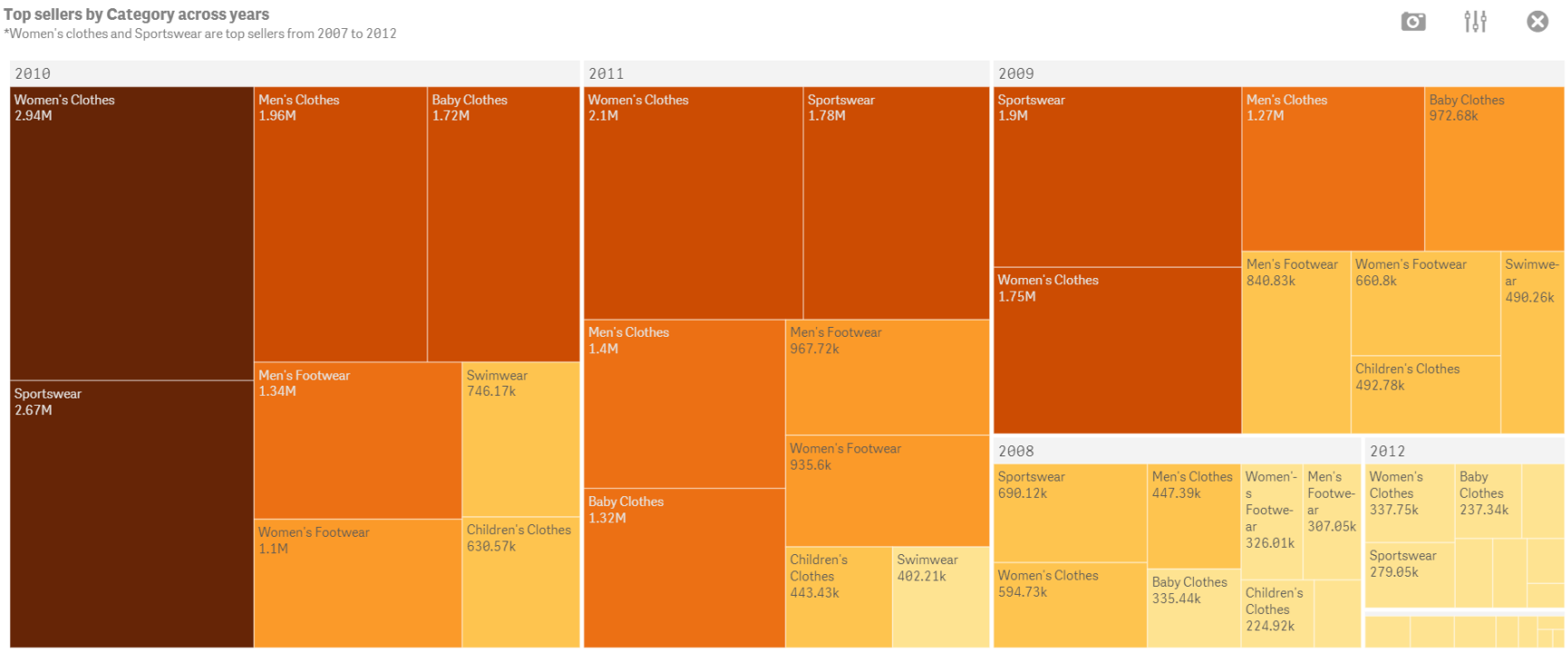




Exercise 7
Add a group dimension

1. **Create a new sheet:** set title as ‘Top sellers by category across years’
2. **Add the required fields:** drag and drop fields *order year, category name* and *amount* into the chart and create a treemap
3. **Set order year as group dimension:** go to ‘data’ – ‘dimension’
4. **Add color legend:** to highlight the higher sales amount

Exercise 7 Solution
Add a group dimension



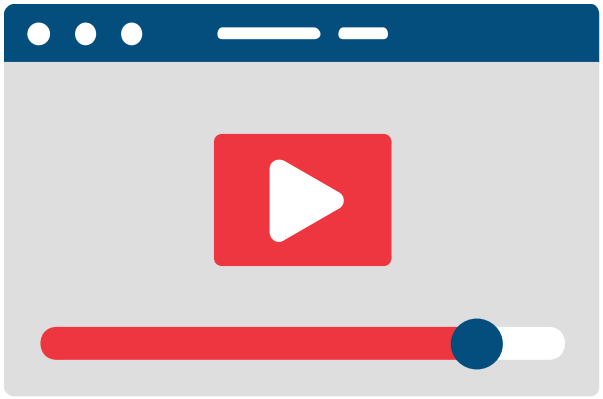
Emerging trends in data visualization

Emerging trends in data visualization

- Increasingly complex data sources
- Data visualization is not just for data scientists
- **Social-first data visualizations**
  - Demo: https://22570l2e793j2oo9c81ug2nh-wpengine.netdna-ssl.com/wp-content/uploads/2017/11/climate-change-data-visualization.gif
- The “interactive map” is becoming a standard medium for data visualizations
  - Demo: https://labs.enigma.com/sanctions-tracker/
- Incorporation of stories into data visualization

Emerging trends in data visualization (continued)

- Mobile-first visualizations
- Data journalism is going mainstream
- Artificial intelligence
- Video: https://www.youtube.com/watch?v=XXXrmO-GIsY&list=PL_qx68DwhYA8g_GffE7fOQ0VPg9A3oSfZ&index=2



Emerging trend for data visualization (Videos)

2019 BI Trend:

https://www.youtube.com/watch?v=vyWbDCn_ncA&list=PL_qx68DwhYA8g_GffE7fOQ0VPg9A3oSfZ

End of Day