# Data Visualization I

MSBA7001 Business Intelligence and Analytics

HKU Business School
The University of Hong Kong

Instructor: Dr. DING Chao

# Agenda

- Tableau Overview
- Basic Charts
  - Change Over Time
  - Part-to-whole
  - Correlation
  - Distribution
- Dashboard & Story

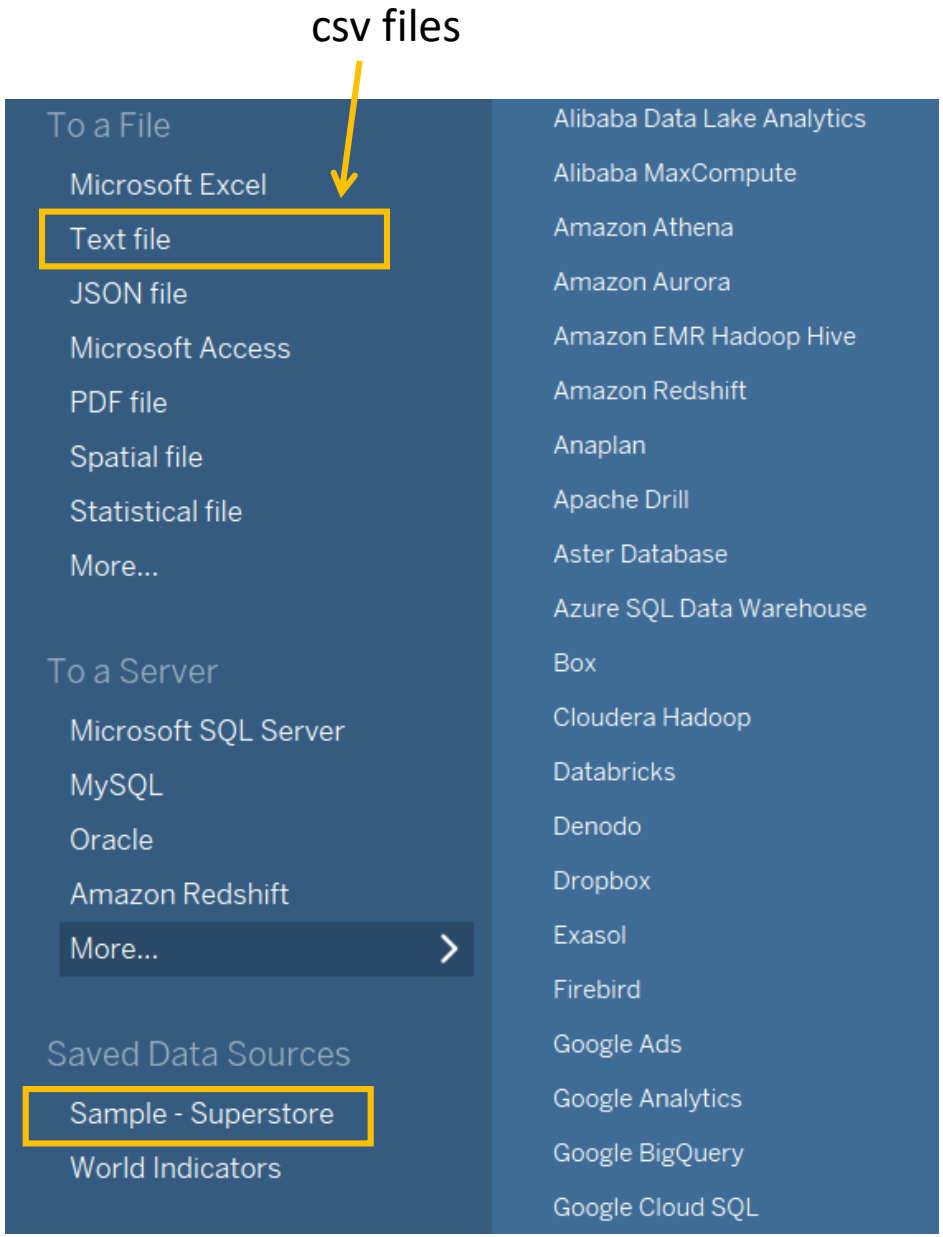
# Tableau Overview

# Why Tableau?

- Tableau is a very effective tool to create interactive data visualizations very quickly.
- It is very simple and user-friendly.
- Moreover, users can perform basic calculations and run some simple stats in Tableau itself.
- More than 50,000 customer accounts and growing.

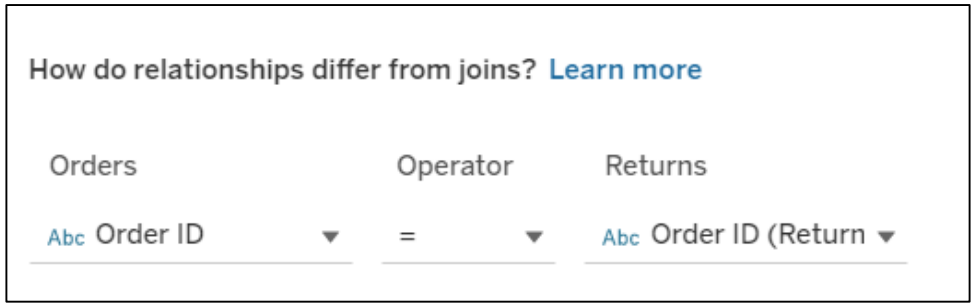
# Connecting to Data

- Tableau supports a lot of data files such as excel, csv, database, spatial, …
- The one we will use is "Sample - Superstore"
- It was created to train tactics, data visualization strategy, and design.
- You may connect to multiple data sources at the same time.

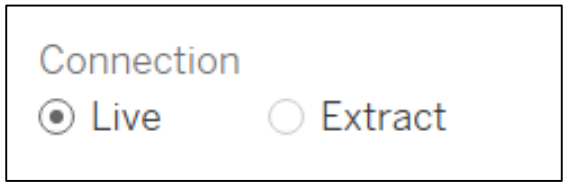


# Connections

- You may join multiple sheets from the same file together.
- Or, join sheets from different files.




- Connecting live leaves the data in the database or file.
- The other option is to extract the data into Tableau's high performance in-memory data engine.



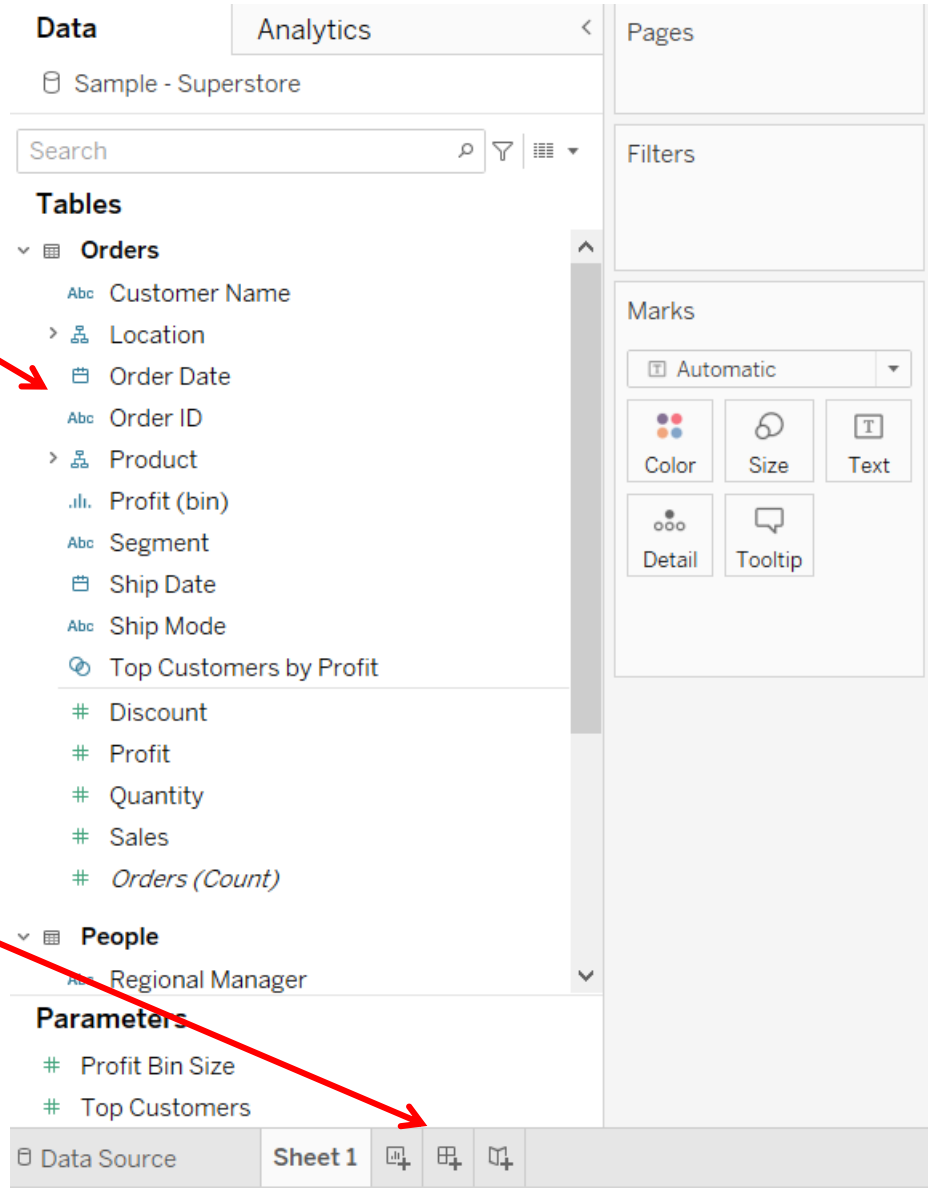
# Data Types

- After connecting data, the first thing to do is check data types.
- Common data types in Tableau.

| Icon | Data Type |
| --- | --- |
| Order Date | Date Values |
| Ship Date | Date and Time Values |
| State | Geographical Values |
| Boolean Values | Boolean Values |
| Region | String Values |
| Profit Ratio | Numerical Values |

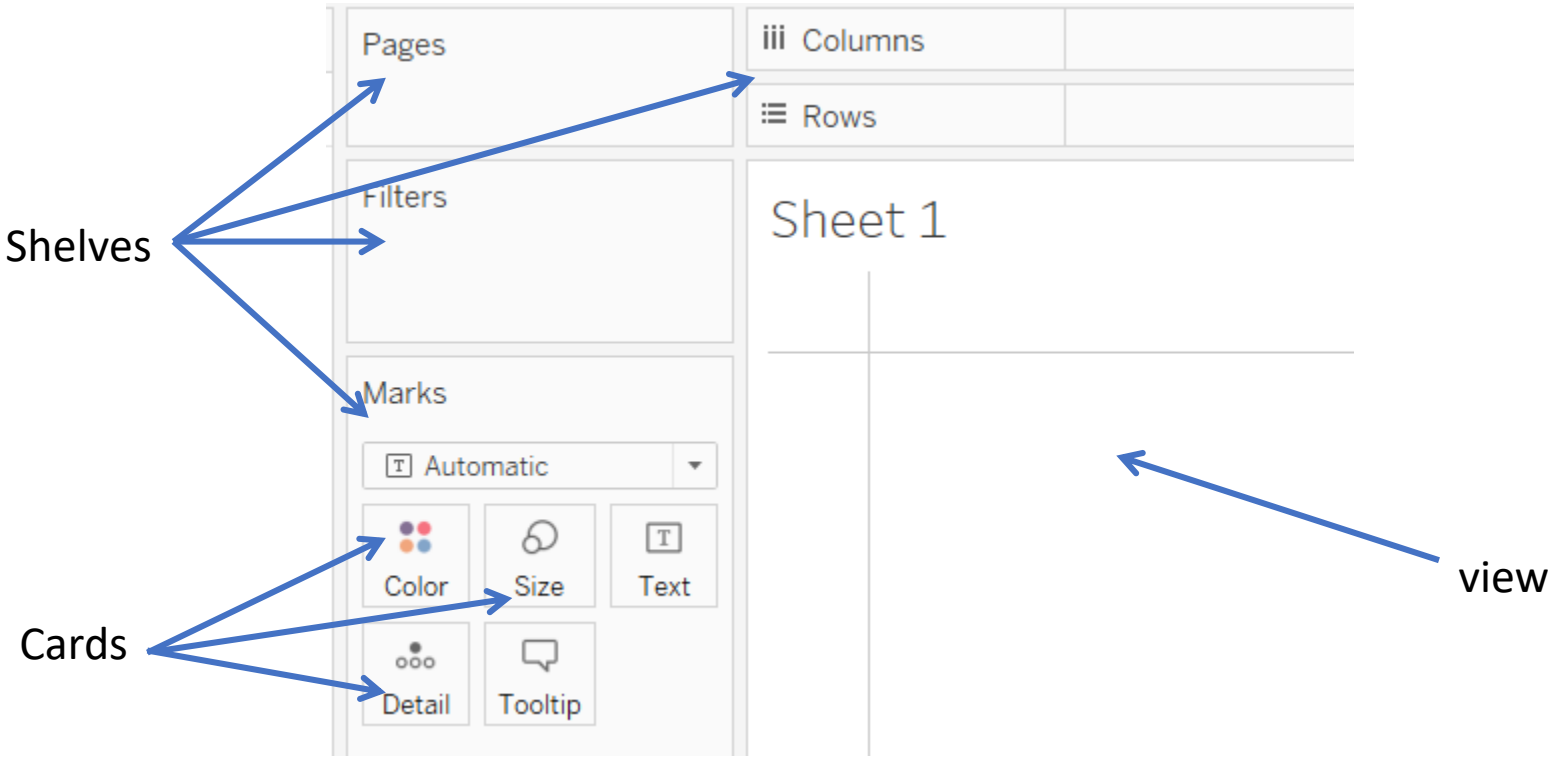
# User Interface

- On the left of the screen is the data pane. The fields from that data source are listed below, broken out into dimensions and measures.
- New sheet tabs are found at the bottom. We can create sheets, dashboards, and stories with these tabs.



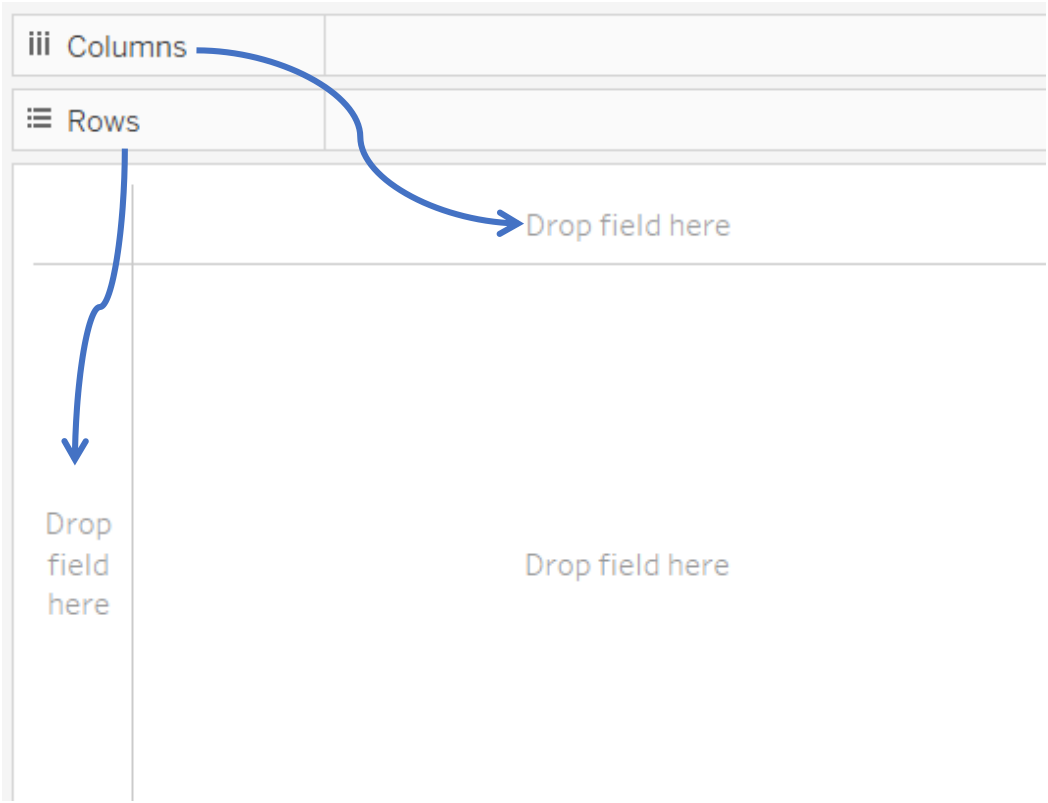
# Shelves & Cards

- A view can be built by dragging and dropping fields from the data pane into the canvas directly, or onto the shelves.



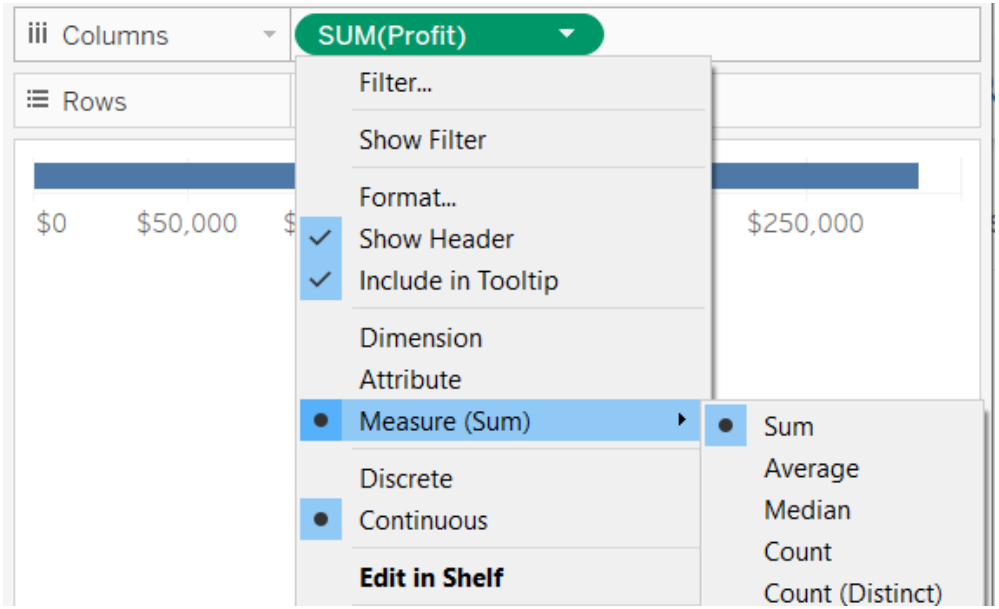
# Columns & Rows Shelves

- The Columns shelf creates the columns of a chart, while the Rows shelf creates the rows of a chart.



# Dimensions & Measures

- Dimensions contain qualitative values (such as names, dates, or geographical data).
- Measures contain numeric, quantitative values that you can measure.
- By default, measures are aggregated when dragged in the view.



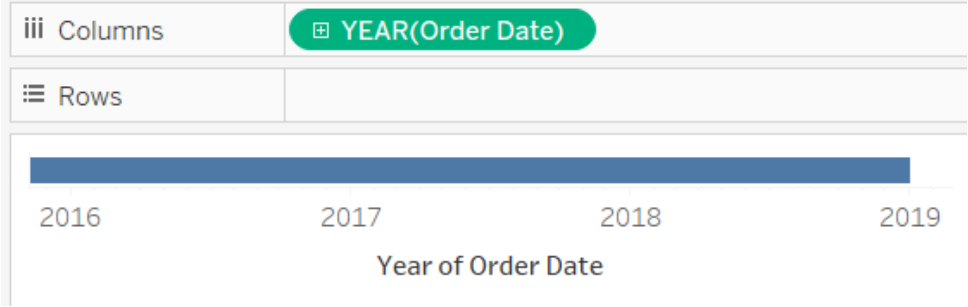
# Continuous vs Discrete

- Discrete fields are represented in **blue**, continuous fields in **green**.

| | |
|---|---|
| **Discrete Dimensions** |   |
| **Continuous Dimensions** (dimensions with a data type of String or Boolean cannot be continuous) |   |
| **Discrete Measures** |   |
| **Continuous Measures** |   |

# Continuous vs Discrete

- Continuous field values are treated as an infinite range. Generally, continuous fields add axes to the view. Measures are generally considered continuous.




- Discrete values are treated as finite. Generally, discrete fields add headers to the view. Dimensions are generally considered discrete.

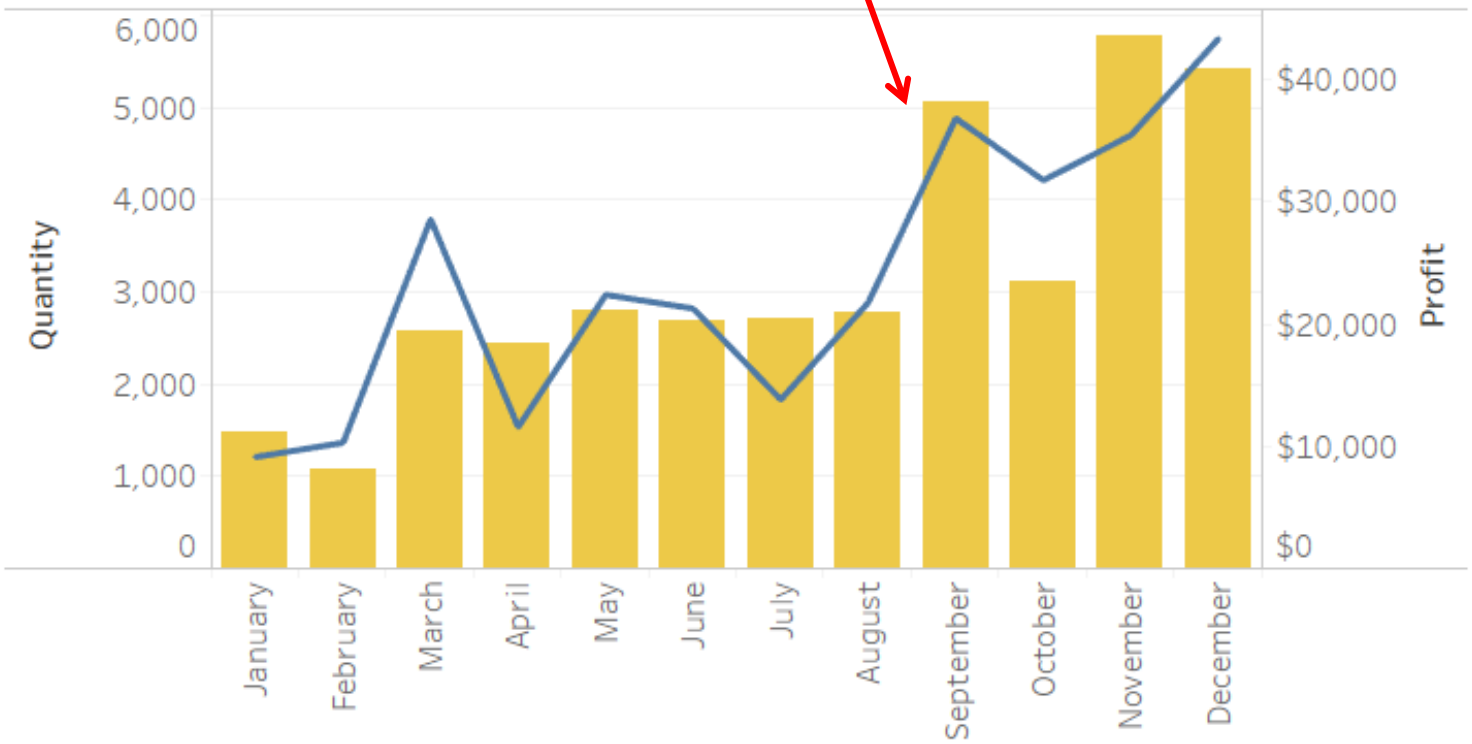
# Creating Tables

- To create a table, we need to add table headers, not axes.
- Drag dimensions to create headers, then drag measures to aggregate.

| Category | Sub-Catego.. | Sales | Profit |
|---|---|---|---|
| Furniture | Bookcases | $114,880 | -$3,473 |
| | Chairs | $328,449 | $26,590 |
| | Furnishings | $91,705 | $13,059 |
| | Tables | $206,966 | -$17,725 |
| Office Supplies | Appliances | $107,532 | $18,138 |
| | Art | $27,119 | $6,528 |
| | Binders | $203,413 | $30,222 |
| | Envelopes | $16,476 | $6,964 |
| | Fasteners | $3,024 | $950 |
| | Labels | $12,486 | $5,546 |
| | Paper | $78,479 | $34,054 |
| | Storage | $223,844 | $21,279 |
| | Supplies | $46,674 | -$1,189 |
| Technology | Accessories | $167,380 | $41,937 |
| | Copiers | $149,528 | $55,618 |
| | Machines | $189,239 | $3,385 |
| | Phones | $330,007 | $44,516 |

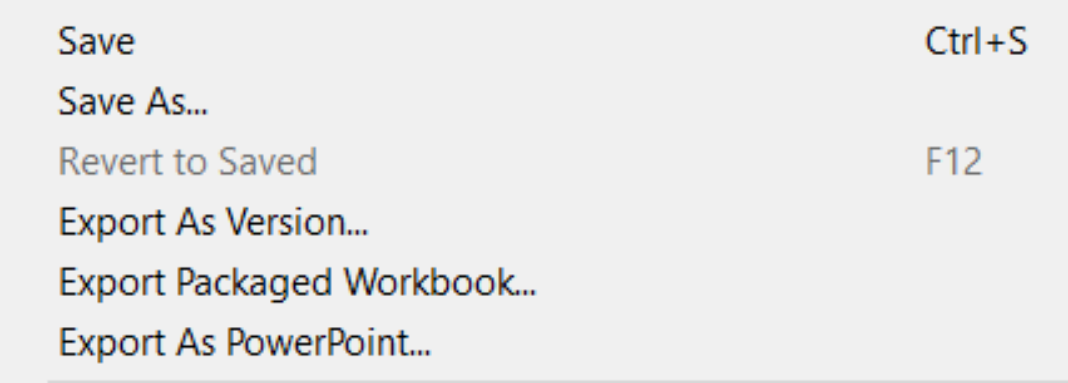
# Creating Charts

- To create a chart, we need fields to create axes and/or headers, and fields to create marks (line, bar, circle, square, text, etc.).



# Saving Charts

- Save as a regular Workbook without the source data.
- Export in a different version, say, Tableau 2018.
- Save as a Packaged Workbook which includes the source data.



# Change Over Time

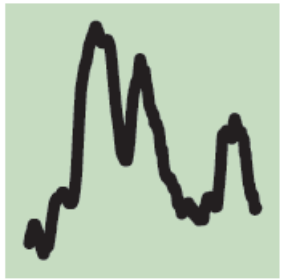
# Change Over Time

- Give emphasis to changing trends.
  - Share price movements
  - Economic time series

**Column**

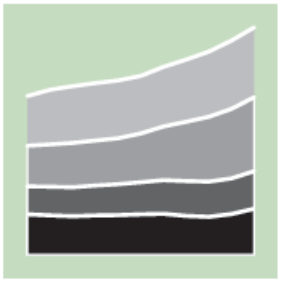
Columns work well for showing change over time - but usually best with only one series of data at a time.

**Line**



The standard way to show a changing time series. If data are irregular, consider markers to represent data points.

**Area chart**

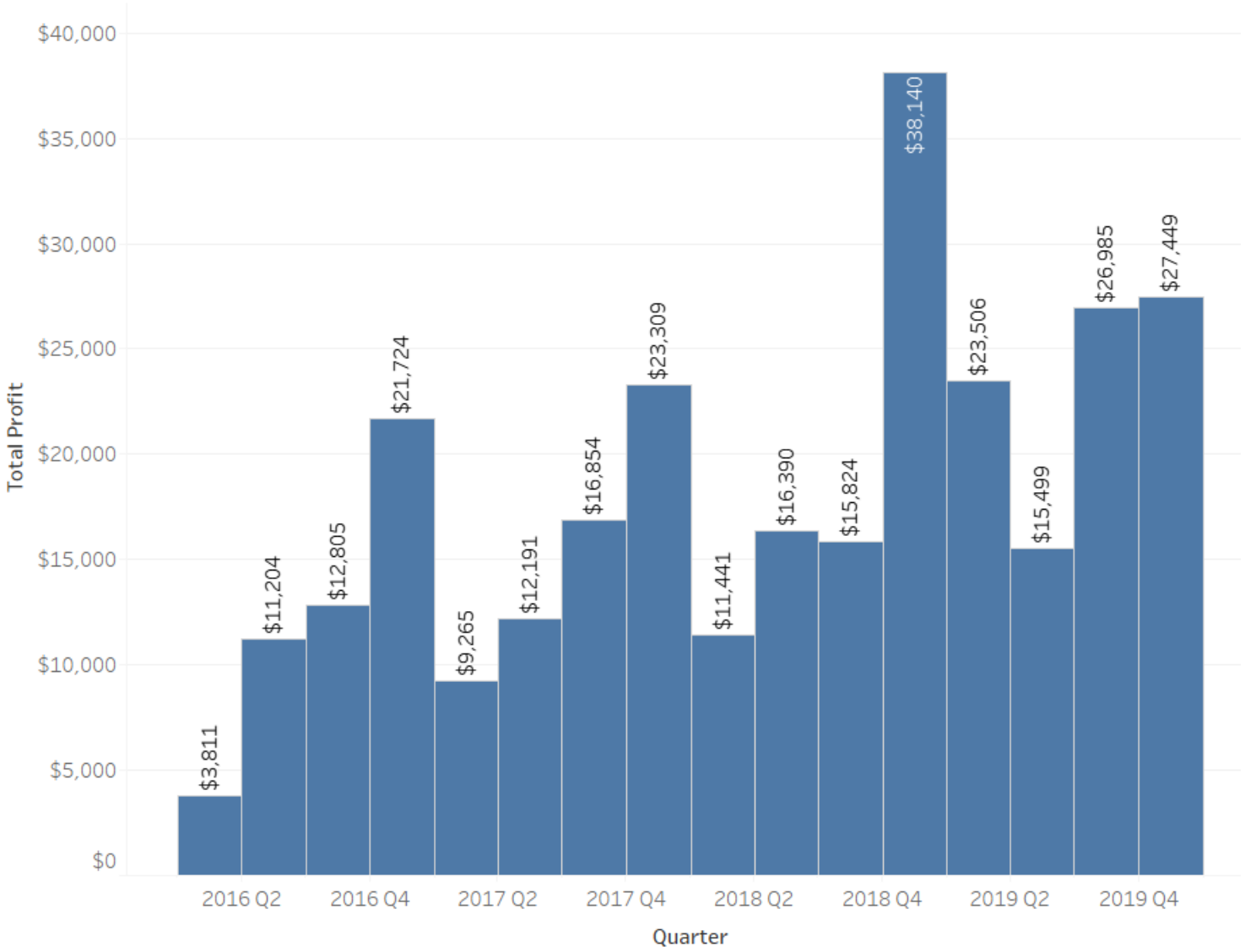


Use with care – these are good at showing changes to total, but seeing change in components can be very difficult.

Source: FT graphic

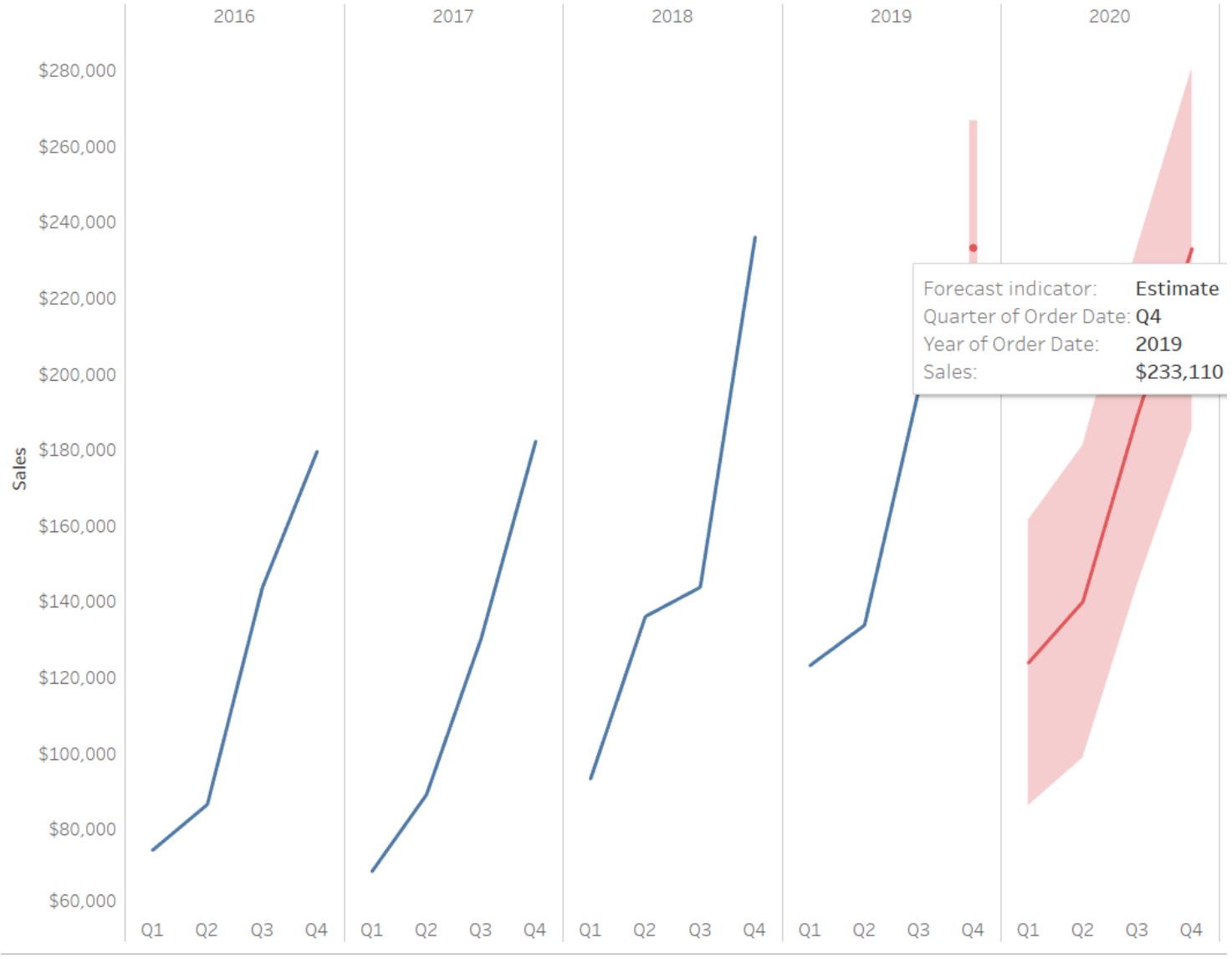
# Bar Chart

## 1. Total Profit by Quarter



# Line Chart

## 2. Forecast Sales for 2020



# Part-to-whole

# Part-to-whole

- Show how a single entity can be broken down into its component elements.
  - Company structures
  - National election results

**Stacked column**

A simple way of showing part-to-whole relationships but can be difficult to read with more than a few components.

**Pie**

A common way of showing part-to-whole data – but be aware that it's difficult to accurately compare the size of the segments.

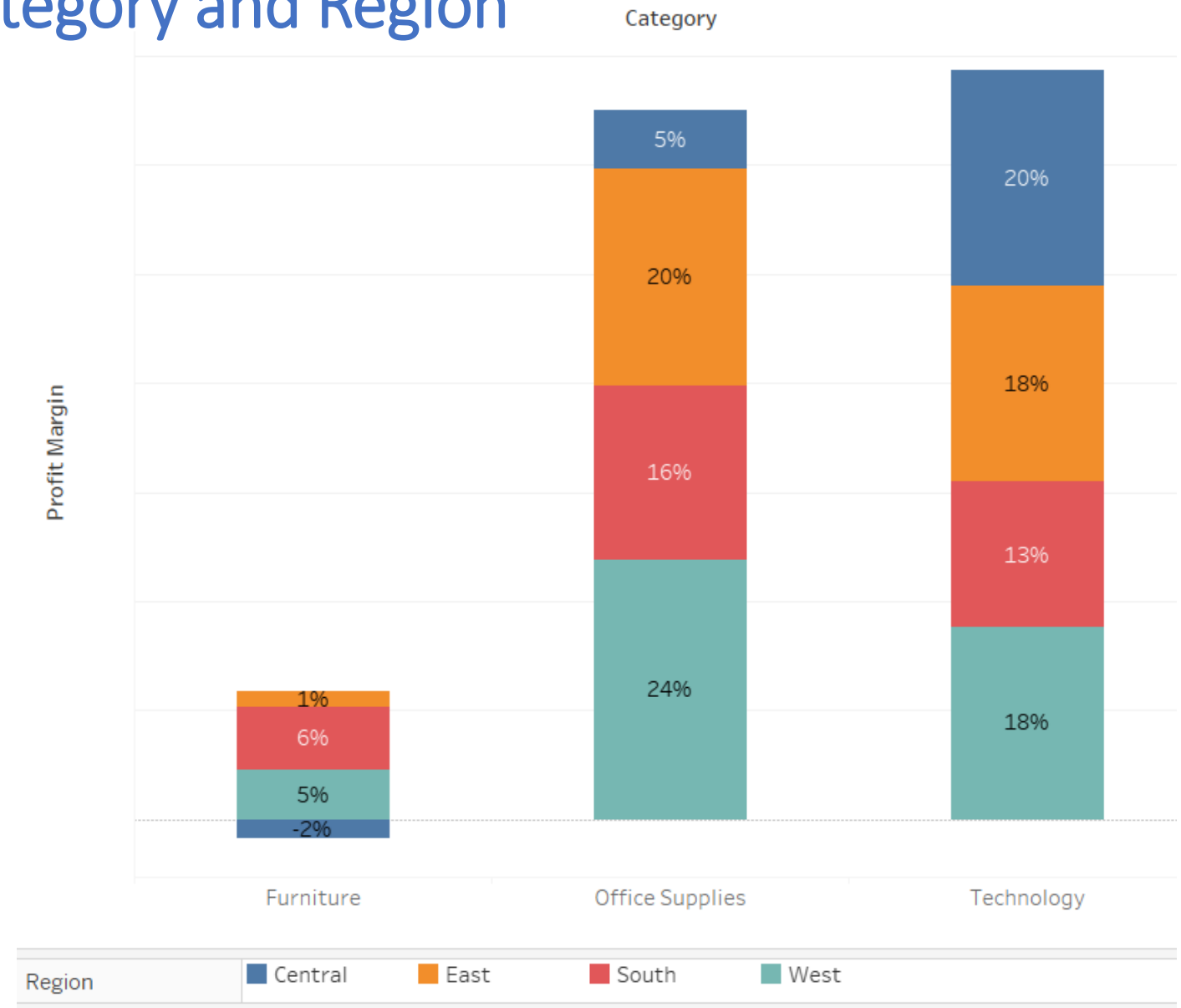
**Treemap**

Use for hierarchical part-to-whole relationships; can be difficult to read when there are many small segments.

Source: FT graphic

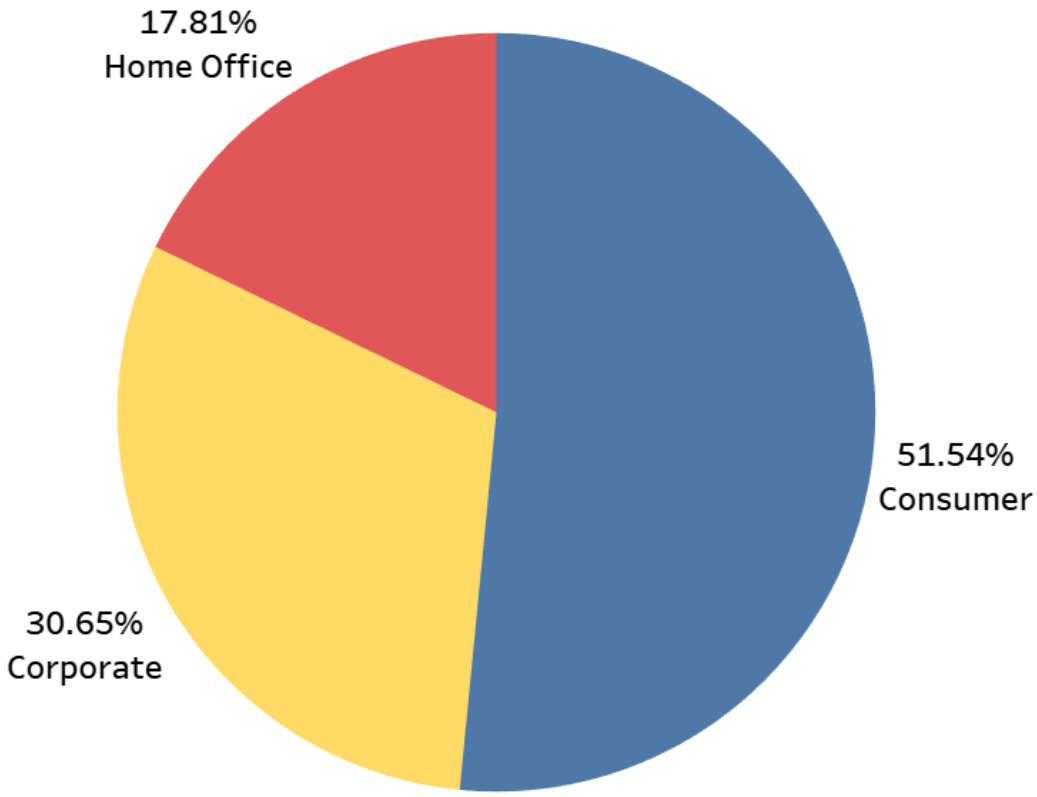
# Stacked Bar Chart

## 3. Average Profit Margin Across Product Category and Region



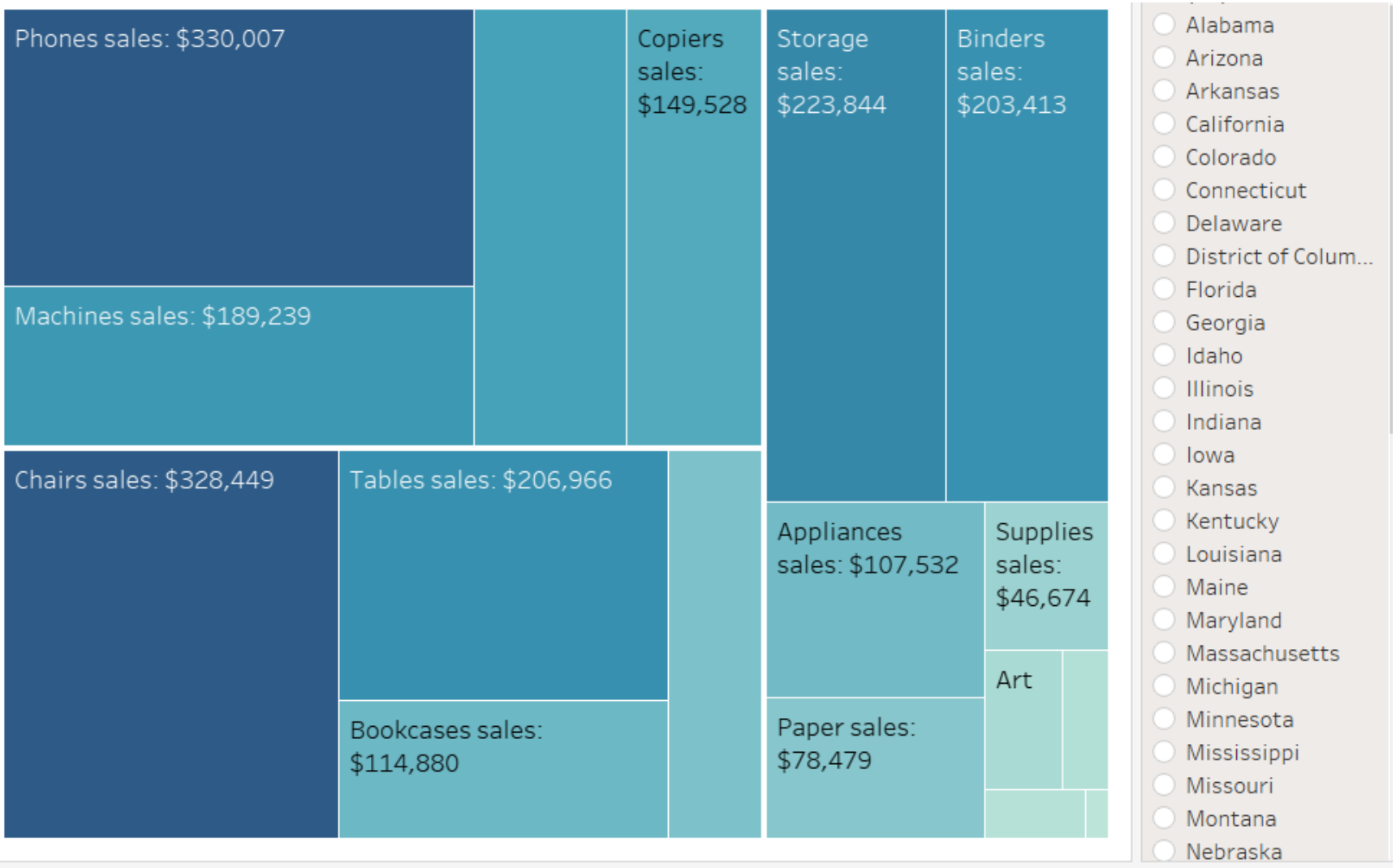
# Pie Chart

## 4. Demand Across Consumer Segment



# Treemap

## 5. Sales Across Sub-Category Filter by State

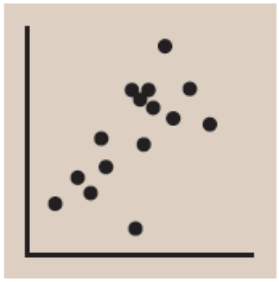


# Correlation

# Correlation

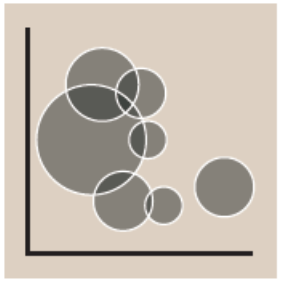
- Show the relationship between two or more variables.
  - Inflation & unemployment
  - Income & life expectancy

**Scatterplot**



The standard way to show the relationship between two continuous variables, each of which has its own axis.

**Bubble**



Like a scatterplot, but adds additional detail by sizing the circles according to a third variable.

**Line + Column**

A good way of showing the relationship between an amount (columns) and a rate (line).

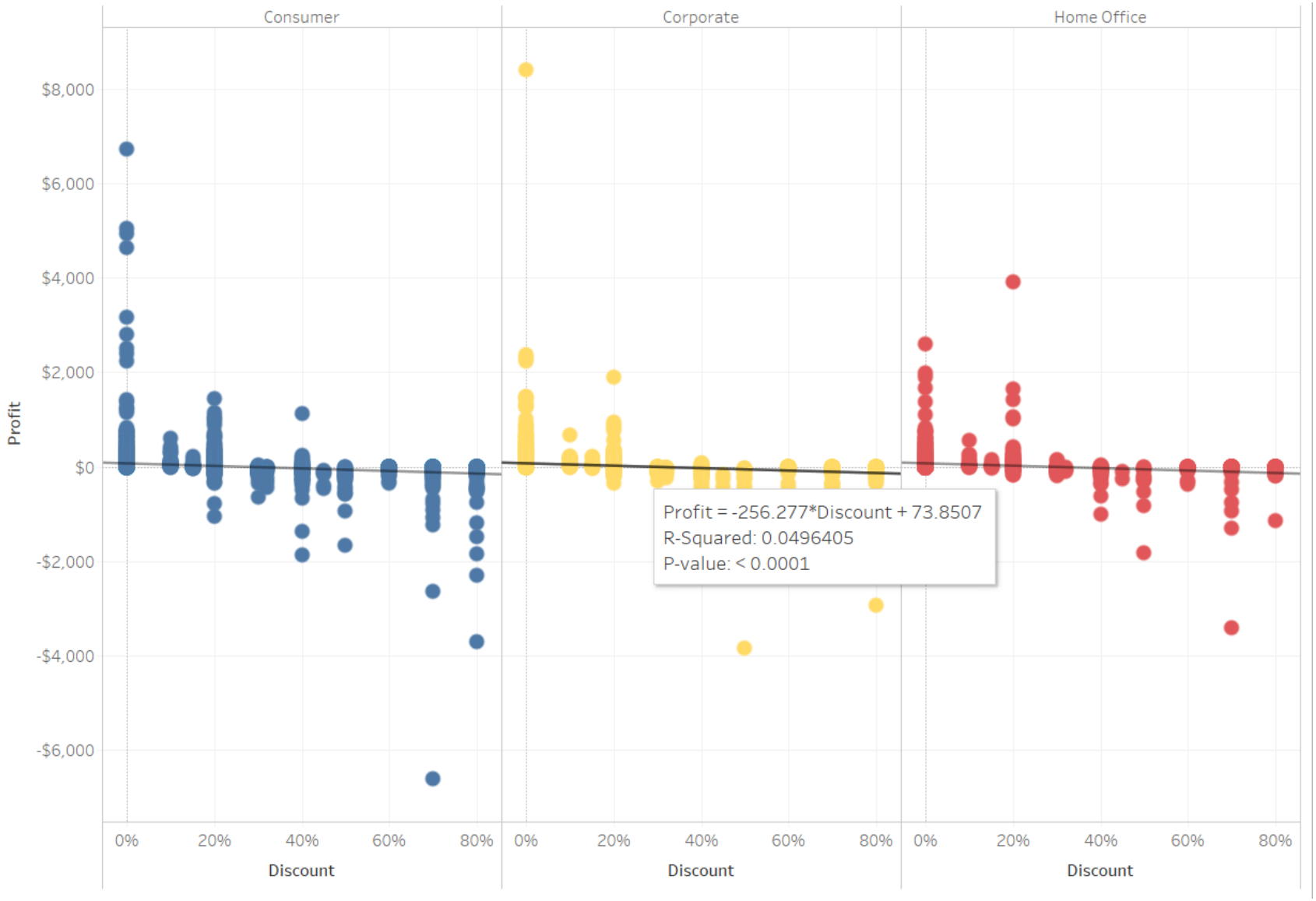
**XY heatmap**

A good way of showing the patterns between 2 categories of data, less good at showing fine differences in amounts.

Source: FT graphic

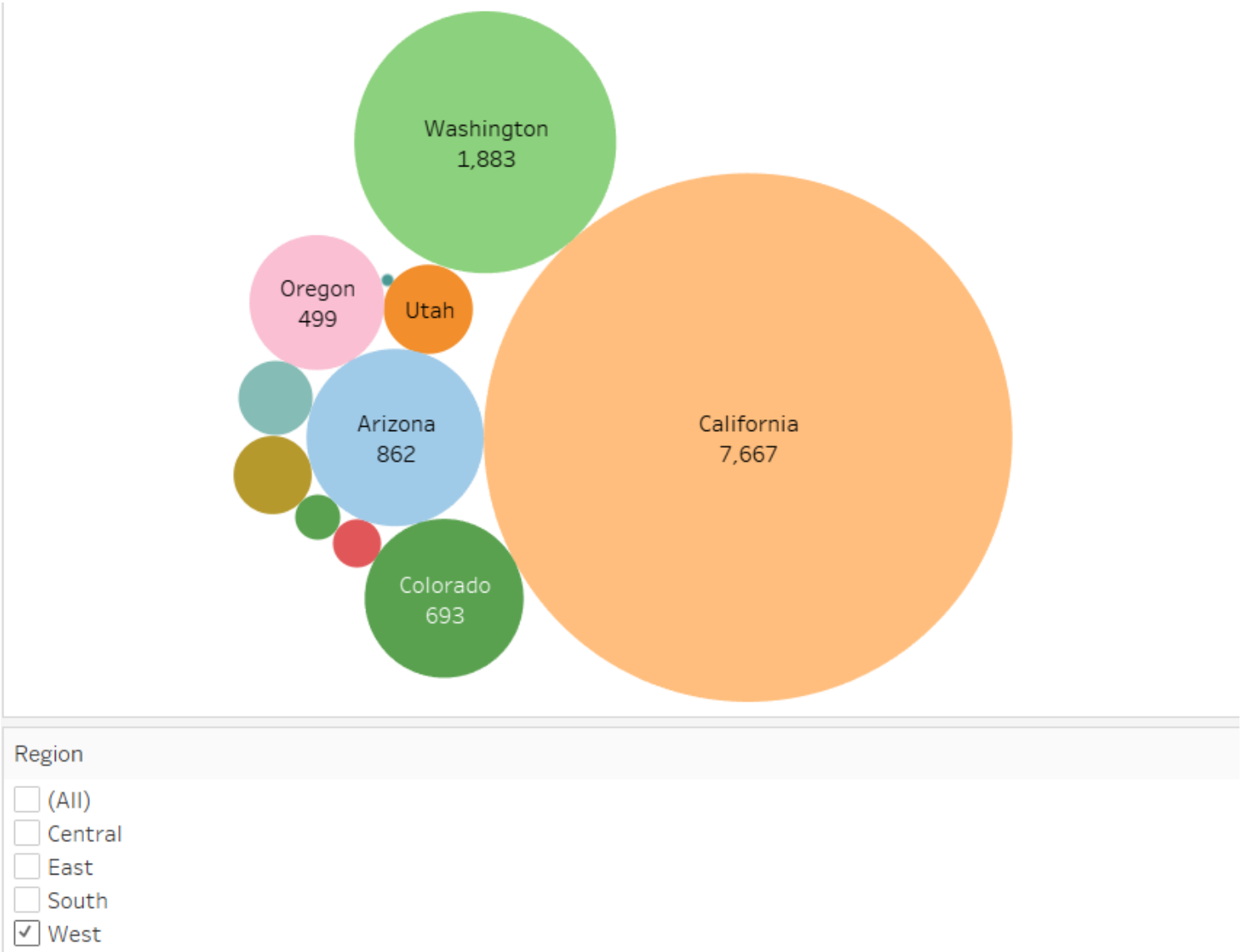
# Scatterplot

## 6. Discount vs. Profit Across Consumer Segment



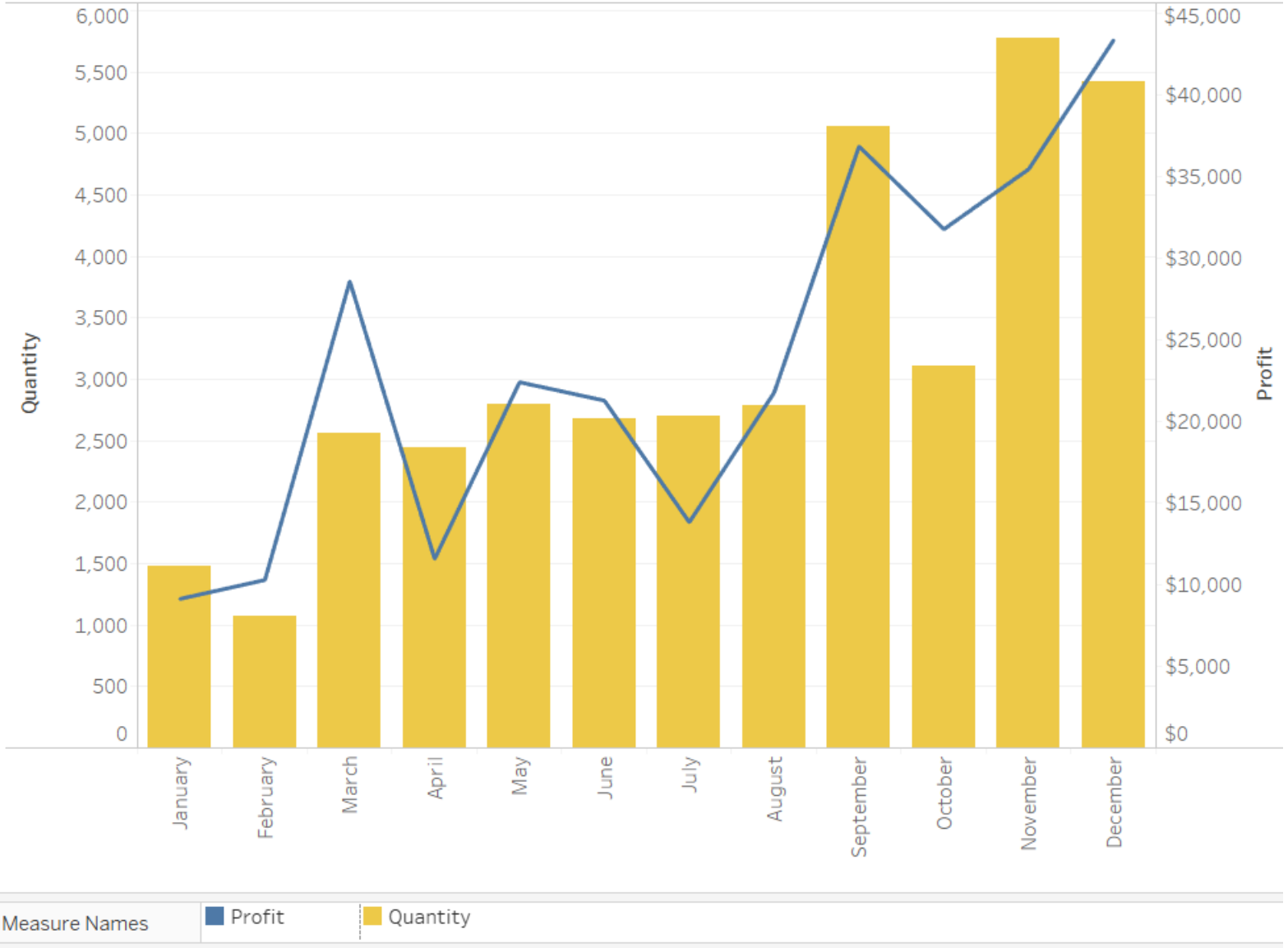
# Bubble

## 7. Demand Across State



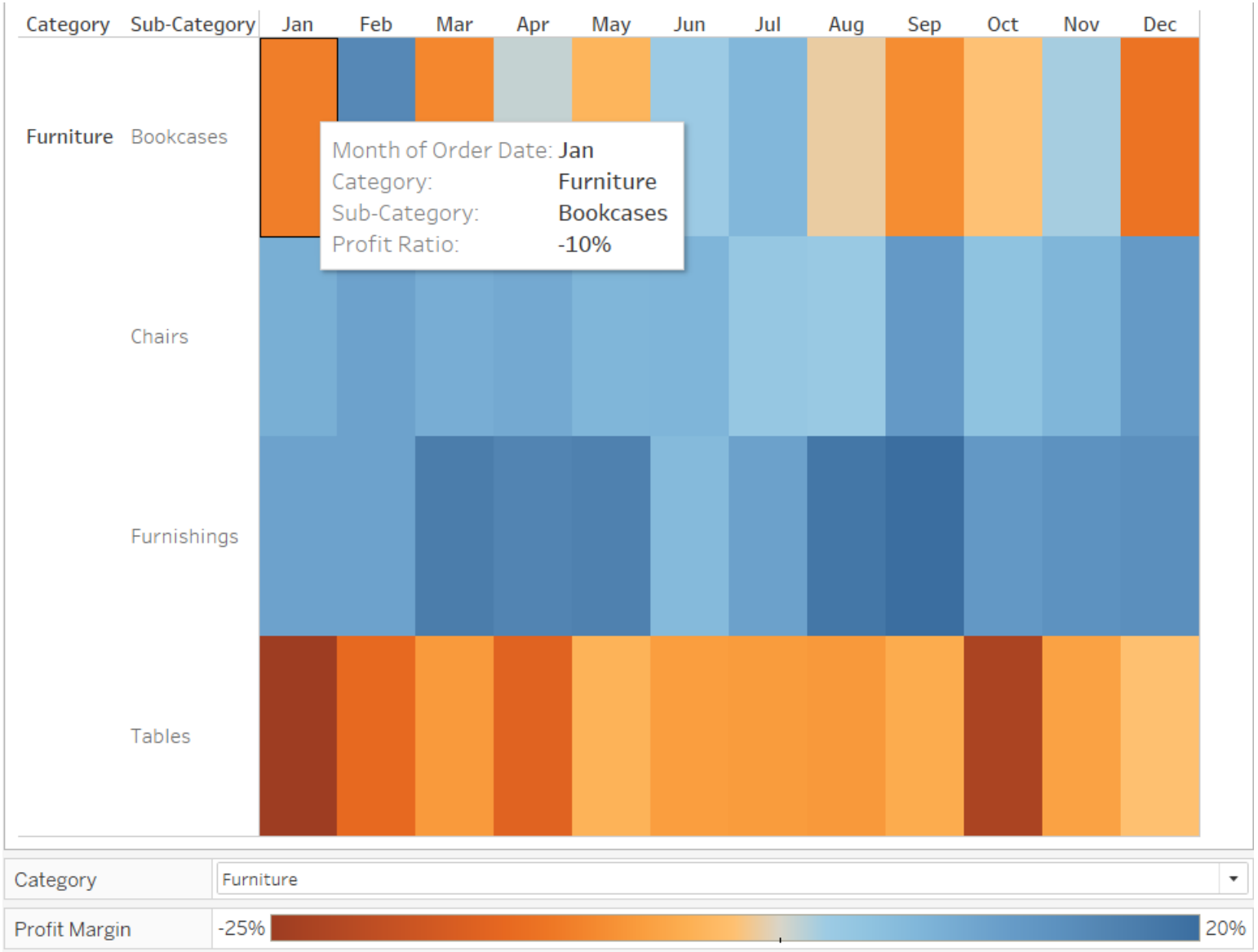
# Dual Axis

## 8. Demand vs. Profit



# Heatmap

## 9. Profit Margin Across Sub-Category & Months

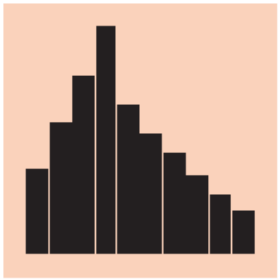


# Distribution

# Distribution

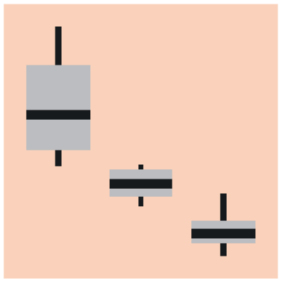
- Show values in a dataset and how often they occur.
  - Income distribution
  - Population (age/sex) distribution

**Histogram**



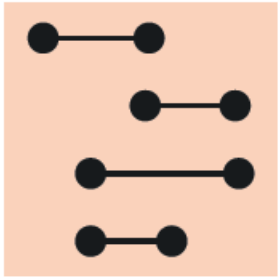
The standard way to show a statistical distribution - keep the gaps between columns small to highlight the 'shape' of the data.

**Boxplot**



Summarise multiple distributions by showing the median (centre) and range of the data

**Dot plot**

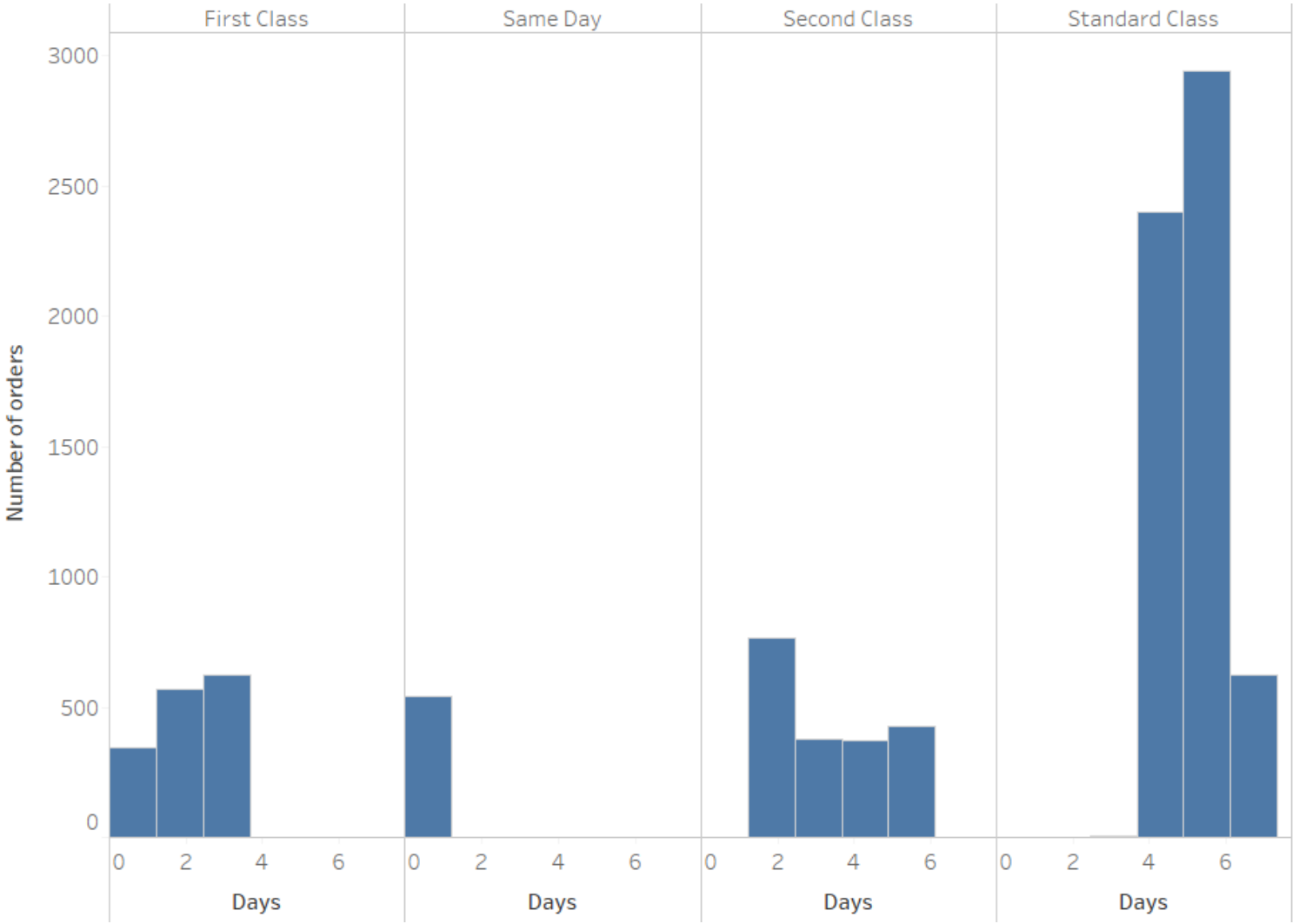


A simple way of showing the change or range (min/max) of data across multiple categories.

Source: FT graphic

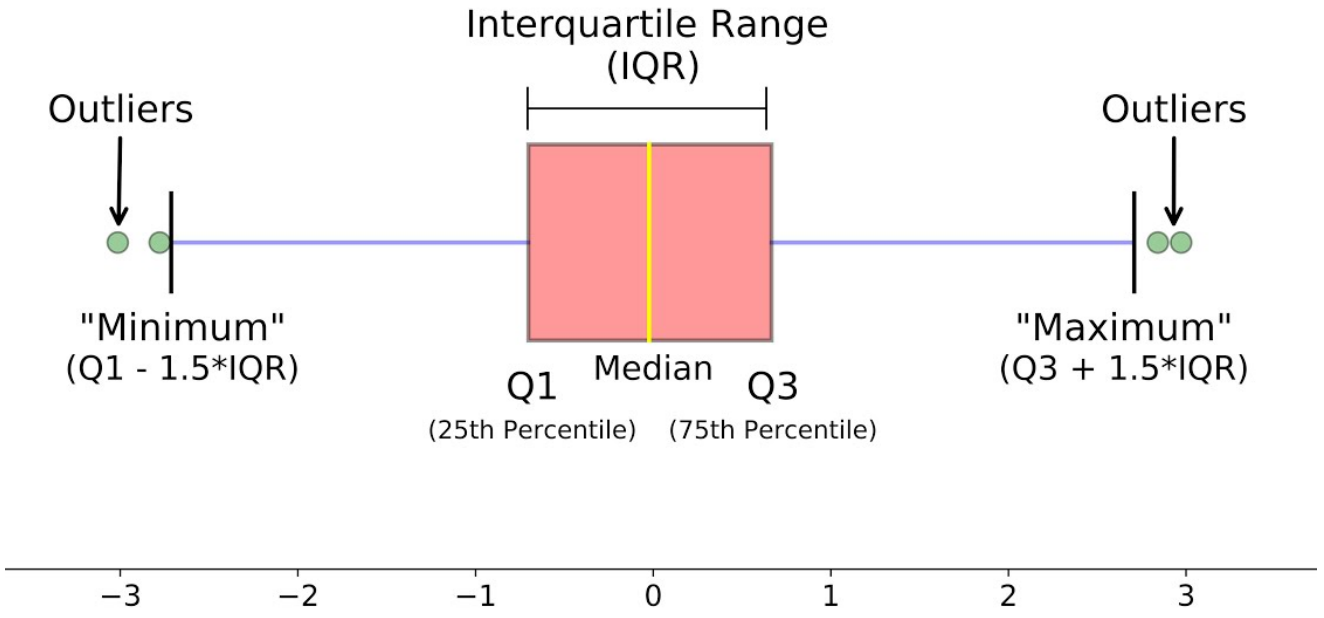
# Histogram

## 10. Distribution of Shipment Time Across Mode



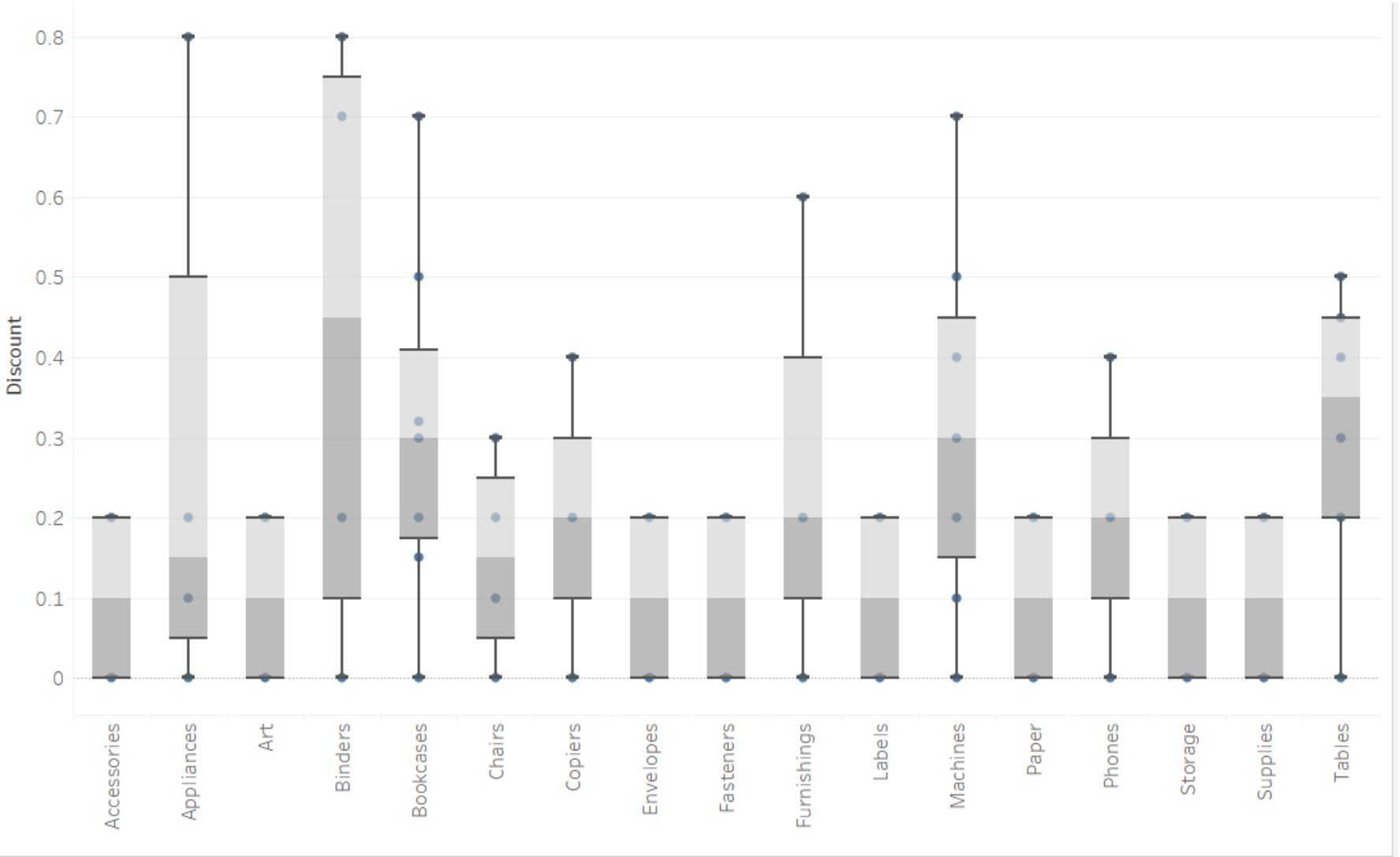
# Boxplots

- Boxplots visually show the distribution of numerical data and skewness.
- They present five statistics: min, first quartile (Q1), median (Q2), third quartile (Q3), max.
- They can help spot outliers.



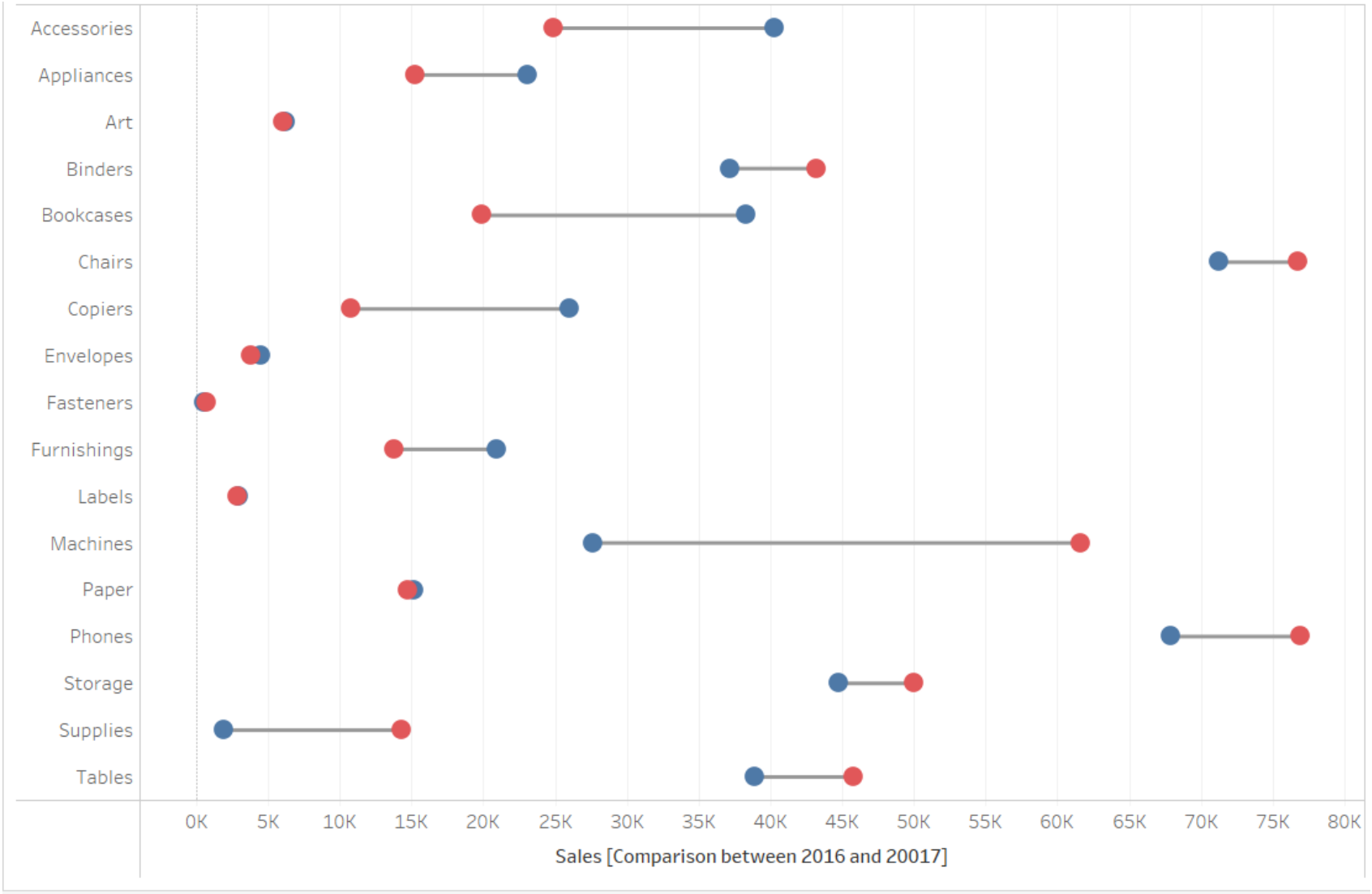
# Boxplot

## 11. Distribution of Discount Across Sub-Category



# Dot Plot (Dumbbell Chart)

## 12. Compare Sales Between 2016 and 2017



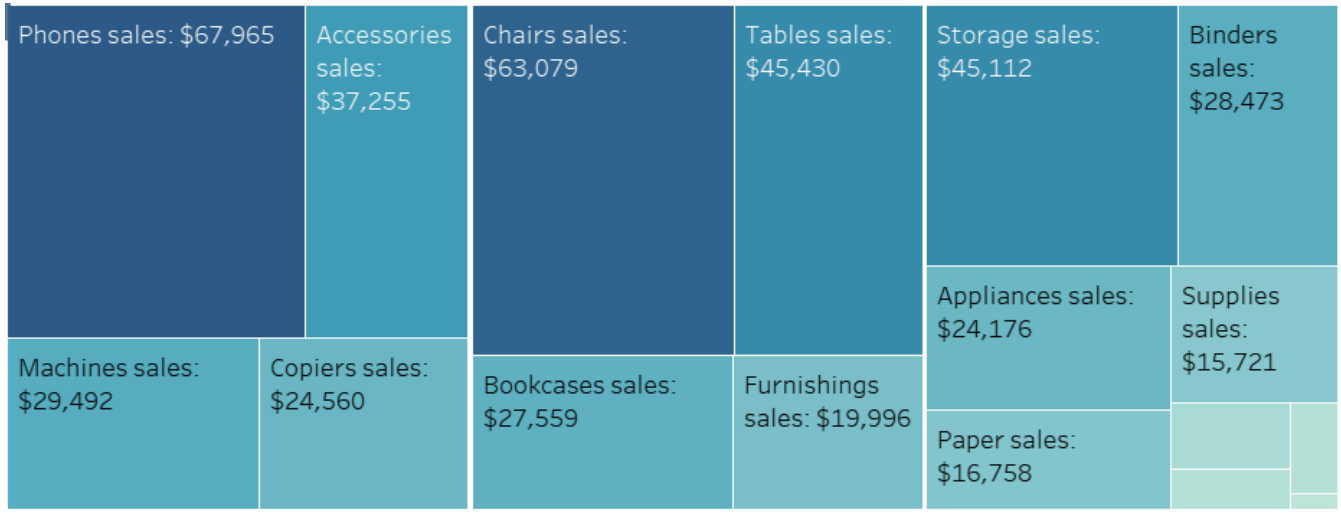
# Dashboard & Story

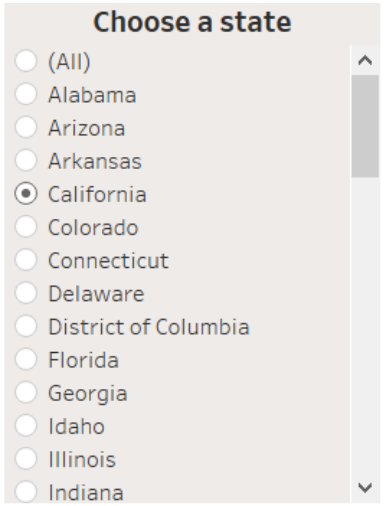
# Dashboard

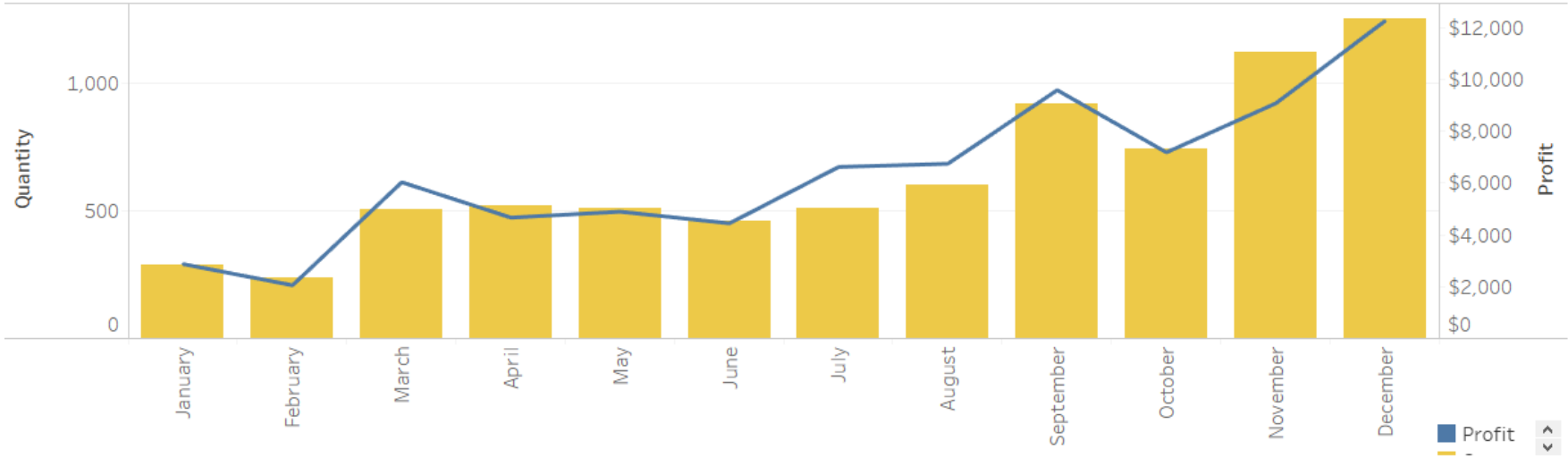
- A dashboard is a collection of several views, letting you compare a variety of data simultaneously.
- For example, if you have a set of views that you review every day, you can create a dashboard that displays all the views at once, rather than navigate to separate worksheets.
- A well-designed dashboard can align your organization's efforts, help uncover key insights, and speed up decision-making.

# Dashboard

## 13. Demand, Sales, and Profit Filter by State



# Story

- A story is a sequence of visualizations that work together to convey information.
- Each individual visualization in a story is called a story point.
- You can create stories to tell a data narrative, provide context, demonstrate how decisions relate to outcomes, or to simply make a compelling case.
- Think of it as presentation slides.