**A Practical Guide to Statistical Data Types and Visualization Techniques with Real-World Examples**

In today's data-driven world, effective data visualization is key to understanding complex information quickly and easily. Whether you're a business analyst, a researcher, or a student, mastering the art of creating multiple charts and graphs can significantly enhance your ability to communicate insights clearly and persuasively. In this guide, we'll explore the principles and techniques behind creating professional-quality visualizations in simple language.

**Introduction Data Visualization**

Data visualization is the graphical representation of information and data. It uses visual elements like charts, graphs, and maps to help viewers understand trends, patterns, and relationships in the data. By presenting complex data in a visual format, data visualization makes it easier to interpret and analyze large datasets. It enables decision-makers to gain insights, identify outliers, and communicate findings effectively. Through visualization, data becomes more accessible, allowing for better-informed decisions and actionable insights.

**What is Data?**

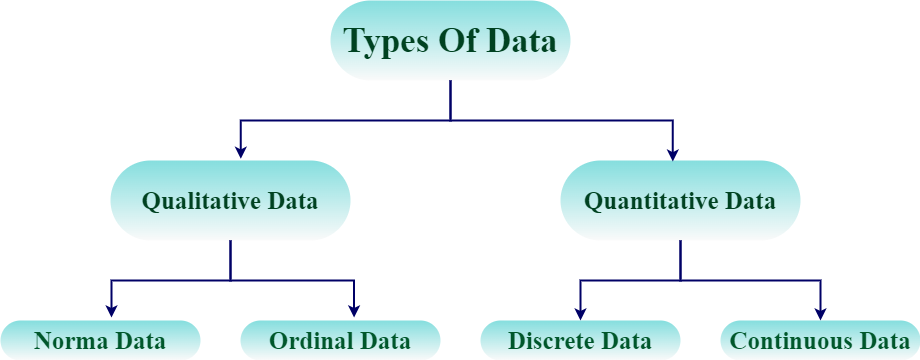
Data is defined as the collection of numbers, characters, images, and others that can arranged in some manner to form meaningful information. In statistics, the data is mainly the collection of numbers that is first studied then analyzed and presented in some way that we can get some meaningful insight from that data.

## What are Types of Data in Statistics?

The data in statistics is classified into four categories:

1. Nominal data
2. Ordinal data
3. Discrete data
4. Continuous data

In statistics, there are four main types of data: nominal, ordinal, interval, and ratio. These types of data are used to describe the nature of the data being collected or analyzed, and they help determine the appropriate statistical tests to use. In this essay, we will explore each type of data in detail, providing examples along the way.



## Qualitative Data (Categorical Data)

As the name suggest Qualitative Data tells the features of the data in the statistics. Qualitative Data is also called Categorical Data and its categories the data into various categories. Qualitative data includes data such as gender of people, their family name and others in sample of population data.

Qualitative data is further categorized into two categories that includes,

* Nominal Data
* Ordinal Data

### **Nominal Data**

Nominal data is a type of data that consists of categories or names that cannot be ordered or ranked. Nominal data is often used to categorize observations into groups, and the groups are not comparable. In other words, nominal data has no inherent order or ranking. Examples of nominal data include gender (Male or female), race (White, Black, Asian), religion (Hinuduism, Christianity, Islam, Judaism), and blood type (A, B, AB, O).

### **Ordinal Data**

Ordinal data is a type of data that consists of categories that can be ordered or ranked. However, the distance between categories is not necessarily equal. Ordinal data is often used to measure subjective attributes or opinions, where there is a natural order to the responses. Examples of ordinal data include education level (Elementary, Middle, High School, College), job position (Manager, Supervisor, Employee), etc.

## Quantitative Data (Numerical Data)

Quantitative Data is the type of the data that represents the numerical value of the data. They are also called the Numerical Data. This data type is used to represent the height, weight, length and other things of the data. Quantitative data is further classified into two categories that are,

* Discrete Data
* Continuous Data

### **Discrete Data**

Discrete data type is a type of data in statistics that only uses Discrete Value or Single Values. These data types have values that can be easily counted as whole numbers. The example of the discreate data types are,

* Height of Students in a class
* Marks of the students in a class test
* Weight of different members of a family, etc.

### **Continuous Data**

Continuous data is the type of the quantitative data that represent the data in a continuous range. The variable in the data set can have any value between the range of the data set. Examples of the continuous data types are,

* Temperature Range
* Salary range of Workers in a Factory, etc.

**Types of Chart And Uses**

**Practical Guide to Bar Charts: Visualizing Categorical Data with Real-World Examples**

**Type of data**:

Categorical, quantitative

**When To Use it:**

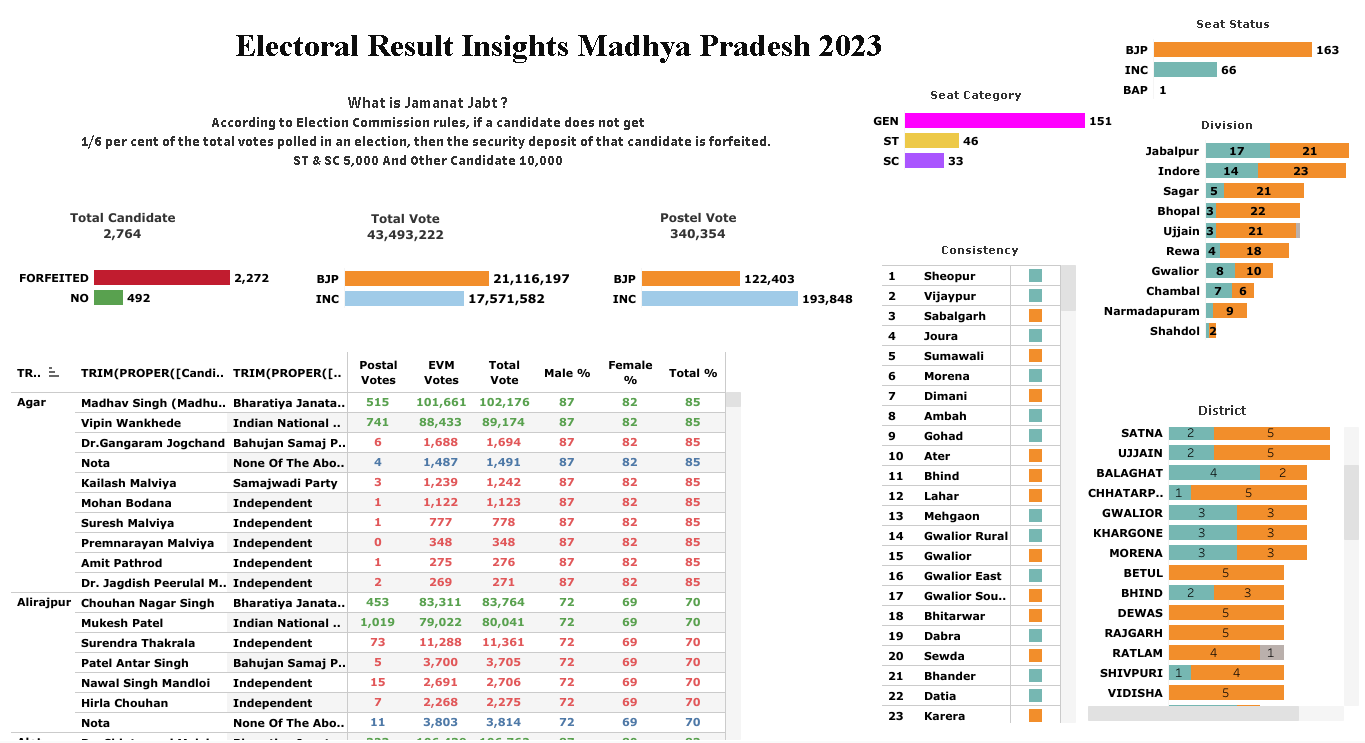
Use a bar Chart to compare data across categories.

**What it shows:**

Bar charts display data using rectangular bars, with the length of the bar representing the value. The bars can be horizontal or vertical.

**When to avoid it:**

Avoid using a bar chart when there are too many categories or if the data is continuous.



**Insights and Findings :**

**Bar graph is a visual representation of data using bars of different heights or lengths to show the frequency or distribution of categorical variables. Each bar represents a category, and the height or length of the bar corresponds to the value of that category.**

We utilized bar graphs in Tableau to visualize the Election dataset because they allow us to compare and analyze various aspects of the election data easily. By representing different variables such as total candidates, forfeited candidates, total votes, postal votes, and voting share gender-wise through bar graphs, we were able to gain valuable insights into the election process. These visualizations help in understanding the distribution and trends within the dataset, making it easier to draw conclusions and make informed decisions.

**A Practical Guide to Line Charts: Trend Analysis Techniques with Real-World Use Cases**

**Type of data:**

Continuous, time-series

**When to use it:**

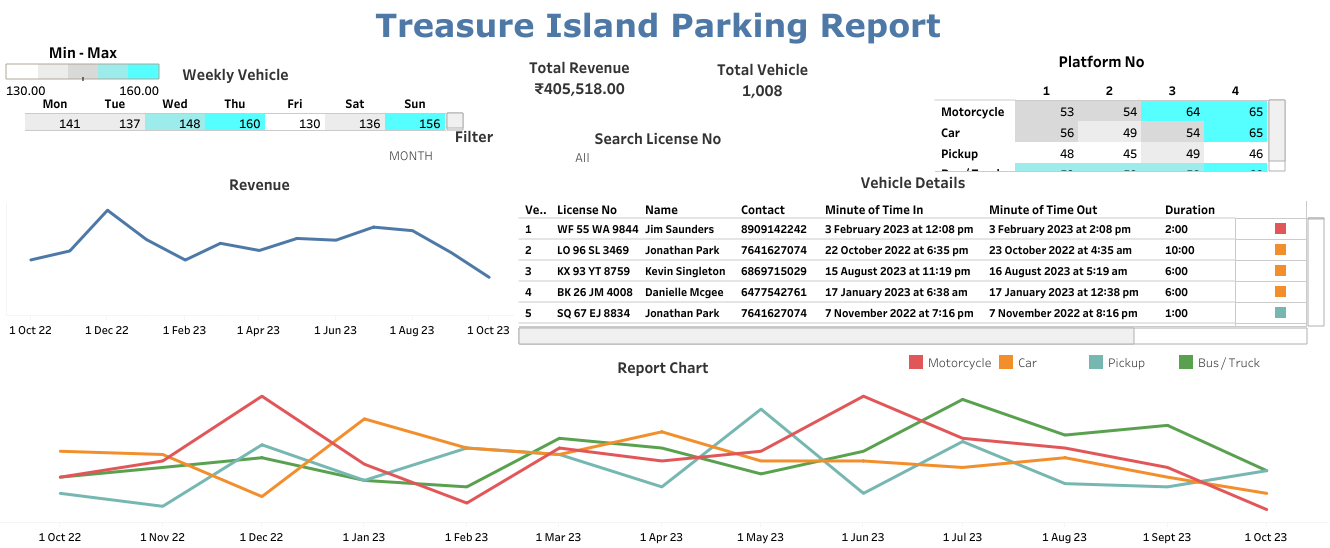
Use a line chart to show trends over time.

**What it shows:**

Line charts plot data points connected by lines. The X-axis usually represents time, and the Y-axis represents the value.

**When to avoid it:**

Only use a line chart when there is a logical order or relationship between data points.



**Insights and Findings :**

We created a line graph for the Treasure Island Parking Report. This report contains time-series data showing the number of motorcycles, cars, pickups, and buses/trucks over different time intervals like weeks, months, quarters, and years. We used different colors for each vehicle type: red for motorcycles, orange for cars, sky blue for pickups, and green for buses/trucks. By analyzing this graph, we can easily see trends and determine which type of vehicle experiences the most traffic during different time periods.

**Insights:**

1. Wednesdays, Thursdays, and Sundays experience significantly higher vehicle traffic.
2. Fridays have the lowest number of vehicles compared to other days.
3. Platforms 3 and 4 receive the highest number of vehicles.
4. Pick-up vehicles are the least frequent compared to all other vehicle types.

**A Practical Guide to Pie and Donut Charts: Representing Proportions with Real-World Examples**

**Type of data:**

Categorical, proportional

**When to use it**:

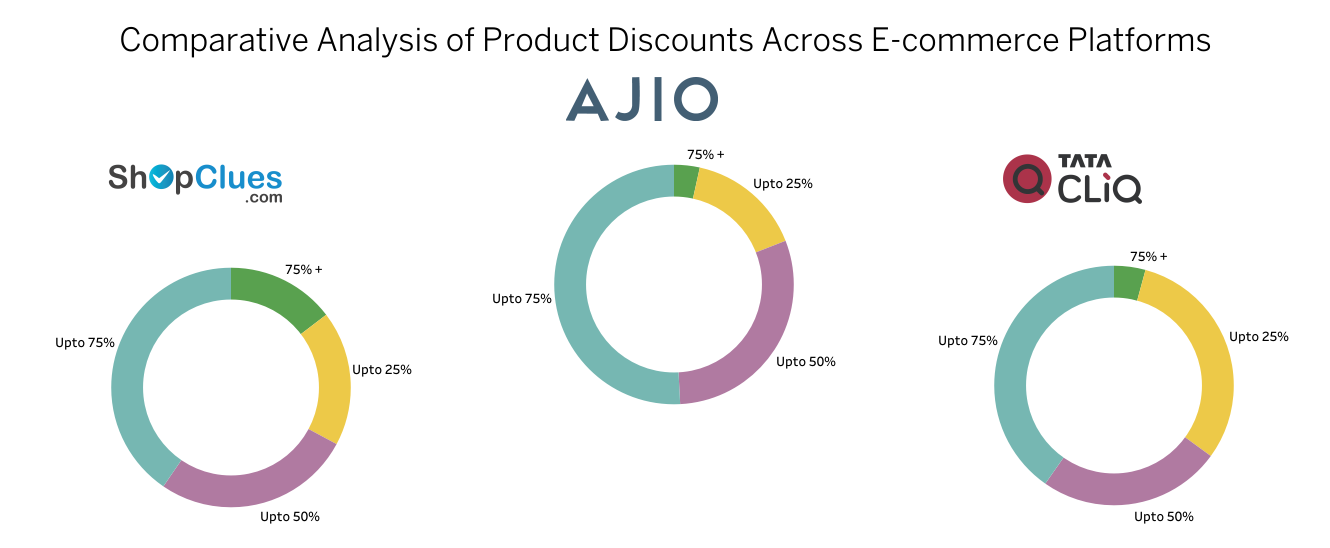
Use a donut chart to show the proportion of each category.

**What it shows:**

Donut charts represent data as slices of a circle, each representing a percentage of the total.

**When to avoid it:**

Avoid using donut charts when there are too many categories or comparing data across groups.



**Insights and Findings :**

We obtained data by scraping e-commerce websites to conduct a comparative study on which website offers the highest discounts based on product price categories. We used a donut chart for this analysis, focusing on three websites: Shopclues, Ajio, and TataCliq. We categorized discounts into four categories: Up to 25%, Up to 50%, Up to 75%, and Up to 75% +. Each category is represented in different colors. Through this analysis, we aim to determine which website offers the most discounts across various price categories.

**Insights:**

1. Tata Cliq has the highest number of products in the Up to 25% discount category.
2. In the Up to 75% discount category, Ajio has some products, but TataCliq and Shopclues have very few products.
3. All websites have the same product categories in the Up to 50% discount range.
4. All websites have the same product categories in the Up to 75% discount range.

**A Practical Guide to Area Charts: Combining Trends and Totals in Real-World Visualizations**

**Type of data:**

Continuous, time-series

**When to use it:**

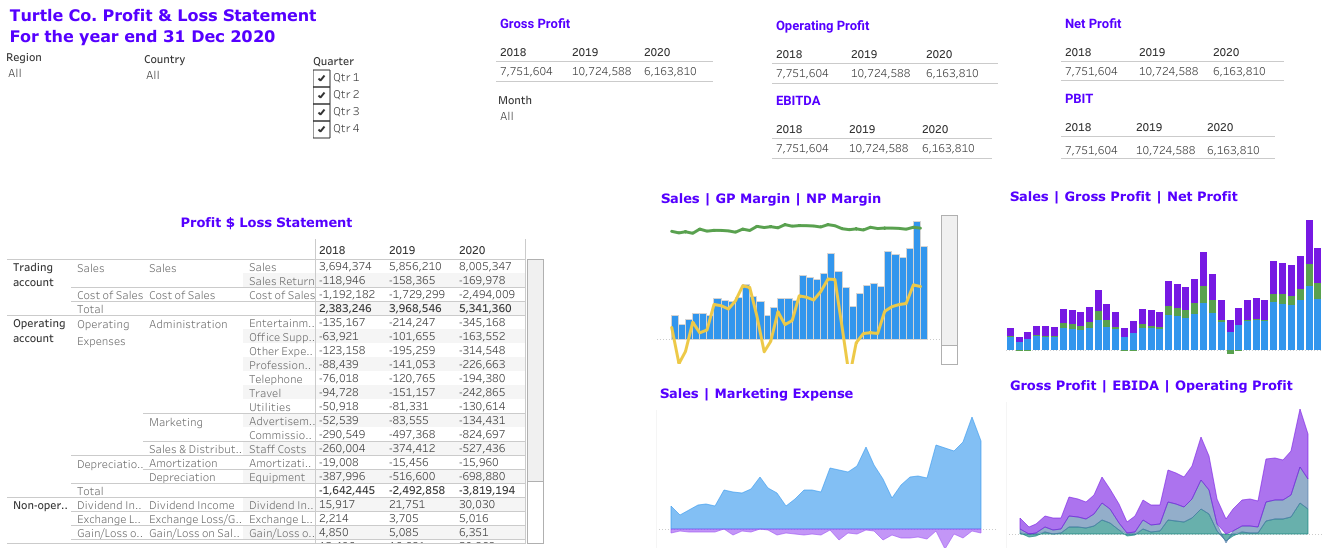
Use an area chart to show the volume or magnitude of data over time.

**What it shows:**

Area charts are similar to line charts, but the area between the line and the X-axis is filled, emphasizing the volume or magnitude.

**When to avoid it:**

Avoid using an area chart with multiple data series with overlapping areas, as it can be confusing.



**Insights and Findings :**

We conducted an analysis of Turtle & Company's financial accounting data for the last three years. Our focus was on understanding sales and marketing expenses. For this, we utilized an area chart, distinguishing between sales and marketing expenses by using different colors. Additionally, to assess gross profit, operating profit, and EBIDA (Earning Before Interest Taxes Depreciation & Amortization), we employed another area chart. Through these visualizations, we aimed to gain insights into the financial performance of Turtle & Company over the specified period.

**Insights:**

1. Marketing expenses remained relatively low compared to total sales.
2. The gross profit margin looks favorable.
3. However, despite the good gross profit, the operating profit was significantly lower, primarily due to EBITDA considerations.

**A Practical Guide to Bubble Charts: Visualizing Multi-Dimensional Data with Real-World Insights**

**Type of data:**

Continuous, multivariate

**When to use it:**

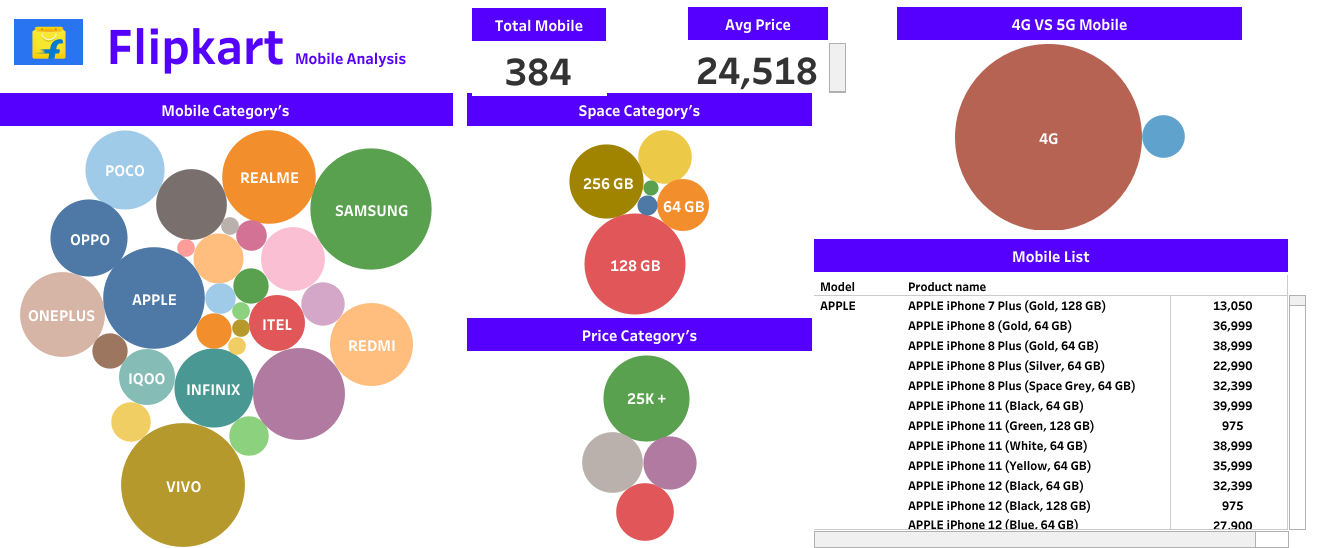
Use a bubble chart to display the relationship between three variables.

**What it shows:**

Bubble charts are a variation of scatterplots, with the size of the bubbles representing the third variable.

**When to avoid it:**

Don't use a bubble chart when the size of the bubbles is not meaningful or when comparing multiple categories.



**Insights and Findings :**

We conducted web scraping of mobile phone data from Flipkart's website, aiming to determine which category of mobile phones is most commonly used in the market. Additionally, we wanted to identify which type of mobile phones, 4G or 5G, dominates the market. To analyze the space category, we used a Bubble chart, where the size of each bubble represents the prominence of a particular category. This allowed us to visually discern trends and patterns in the data, facilitating our understanding of the mobile phone market landscape.

**Insights:**

1. On the website, you'll find the most mobile phones in the Samsung, Vivo, and Realme categories.
2. The highest number of mobile phones priced over 25k are available on the website.
3. When it comes to the space category on the website, you'll find the most mobile phones with 128 GB.
4. Comparing 4G and 5G mobile phones, the website offers more options in the 4G category.

**A Practical Guide to Heatmaps: Uncovering Patterns Through Color-Coded Data with Real Examples**

**Type of data:**

Continuous, multivariate

**When to use it:**

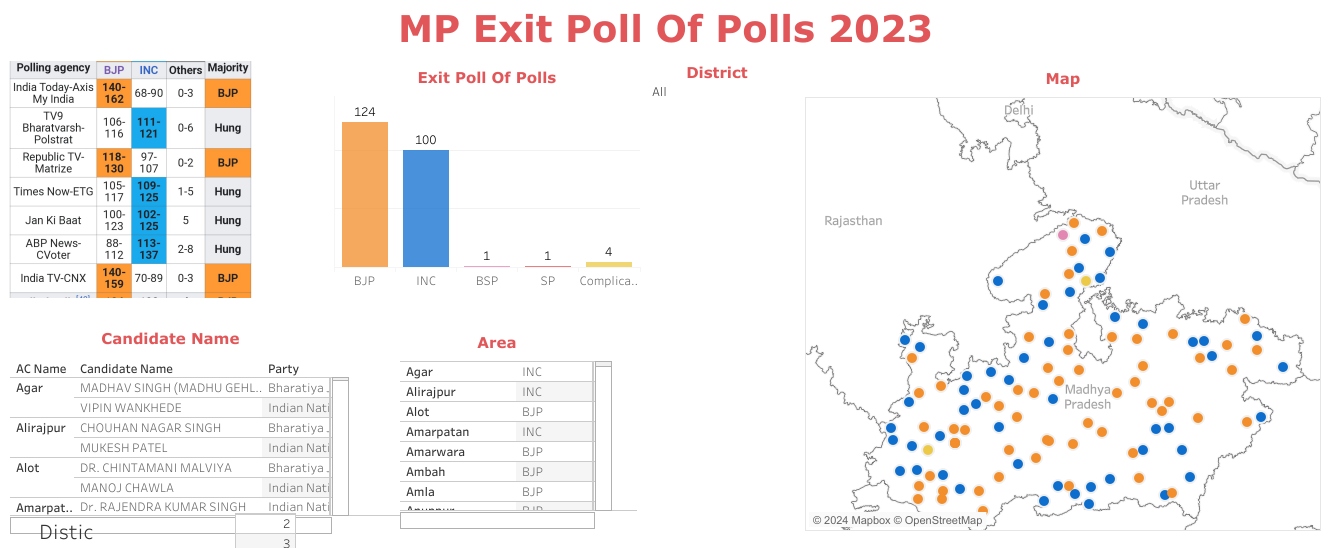
Use a heatmap to display the relationship between two variables using color intensity.

**What it shows:**

Heatmaps use a color scale to represent the value of each cell in a matrix, with one variable on the X- axis and the other on the Y-axis. Darker colors indicate higher values, while lighter colors represent lower values.

**When to avoid it:**

Don't use a heatmap when the relationship between variables is irrelevant, when the data is categorical, or when comparing multiple categories.



**Insights and Findings :**

To display the exit poll results for the Madhya Pradesh Assembly Election 2023 from multiple channels, we utilized a treemap. We represented trends using colors, with the Bharatiya Janata Party (BJP) shown in orange and the Congress party in sky blue. For channels predicting more than 116 seats for a party (the majority required), we highlighted those colors. This visualization method allowed us to easily compare and understand the projected outcomes across different channels.

**Insights:**

1. According to reports from IndiaToday - Axis My India, Republic Tv - Matrize, and India Tv - CNX News Channel, the Bharatiya Janata Party (BJP) has secured a significant majority.
2. Other news channels have indicated a hung assembly, where both the BJP and Congress are projected to win the same number of seats.

**A Practical Guide to Radar (Spider) Charts: Comparing Multiple Variables with Real-World Data**

**Type of data:**

Continuous, multivariate

**When to use it:**

Use a radar chart to display the performance or characteristics of different categories

across multiple dimensions.

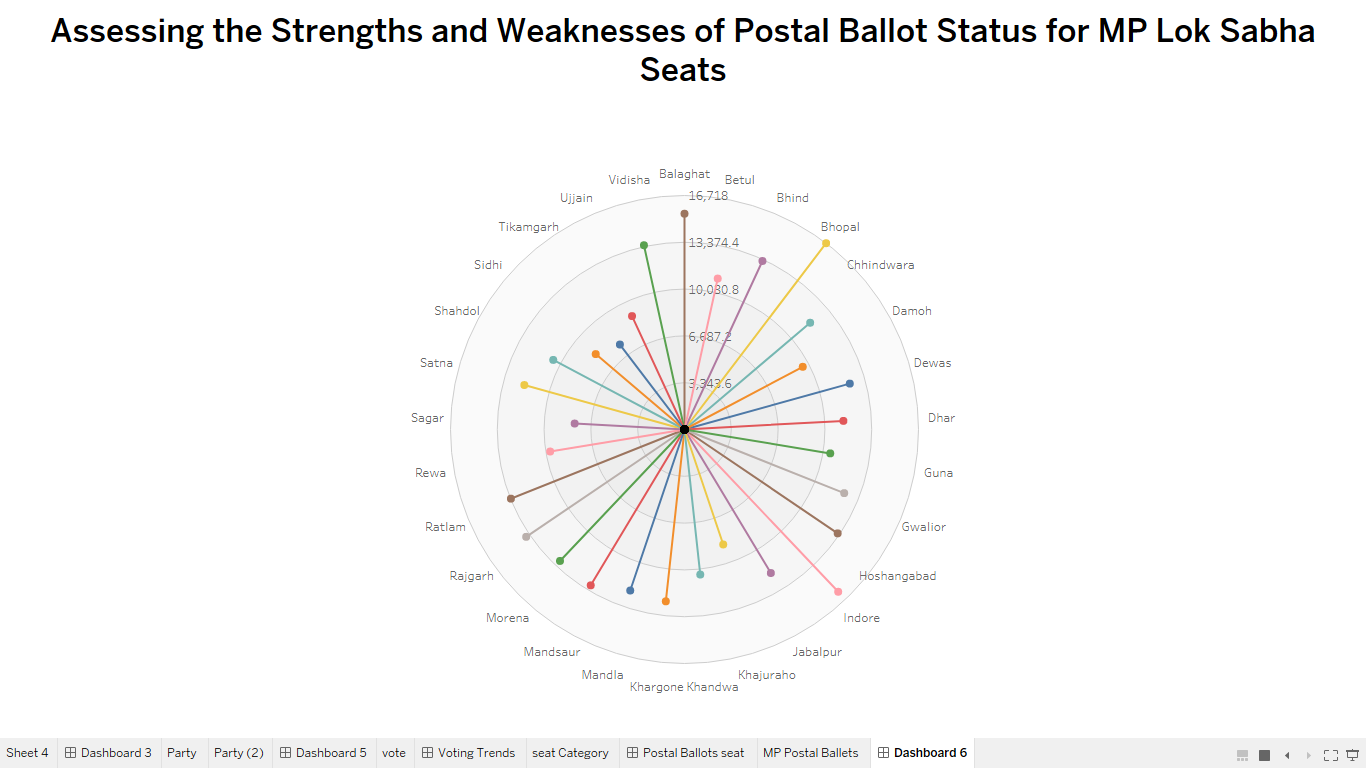
.

**What it shows:**

Radar charts use a circular layout with multiple axes, each representing a dimension. Data points are plotted on each axis and connected to form a shape.

**When to avoid it:**

Don't use a radar chart when there are only a few dimensions or when comparing data across groups.



**Insights and Findings :**

Taking into account the Lok Sabha 2024 elections, we wanted to determine which parliamentary constituency in Madhya Pradesh has the highest number of postal ballots. To achieve this, we examined the results of the Madhya Pradesh Assembly Election 2023. We used a radar chart to visualize the total number of postal ballots for each assembly constituency that falls under the Lok Sabha constituency. This helped us identify which seat has the highest number of postal ballots.

**Insights:**

Through the radar chart, we identified strengths and weaknesses regarding postal ballots in Lok Sabha seats:

1. Bhopal, Indore, and Balaghat have more than 16,000 postal ballots, indicating a significant strength.
2. Hoshangabad, Gwalior, Rajgarh, Ratlam, Vidisha, and Bhind have around 13,000 postal ballots, indicating a moderate strength.
3. Sidhi, Tikamgarh, Khajuraho, and Dewas have fewer than 6,000 postal ballots, indicating a weakness.
4. Other Lok Sabha seats have around 10,000 postal ballots, indicating a moderate strength.

**A Practical Guide to Box Plots: Displaying Distribution and Outliers with Real-World Examples**

**Type of data:**

Continuous, univariate

**When to use it:**

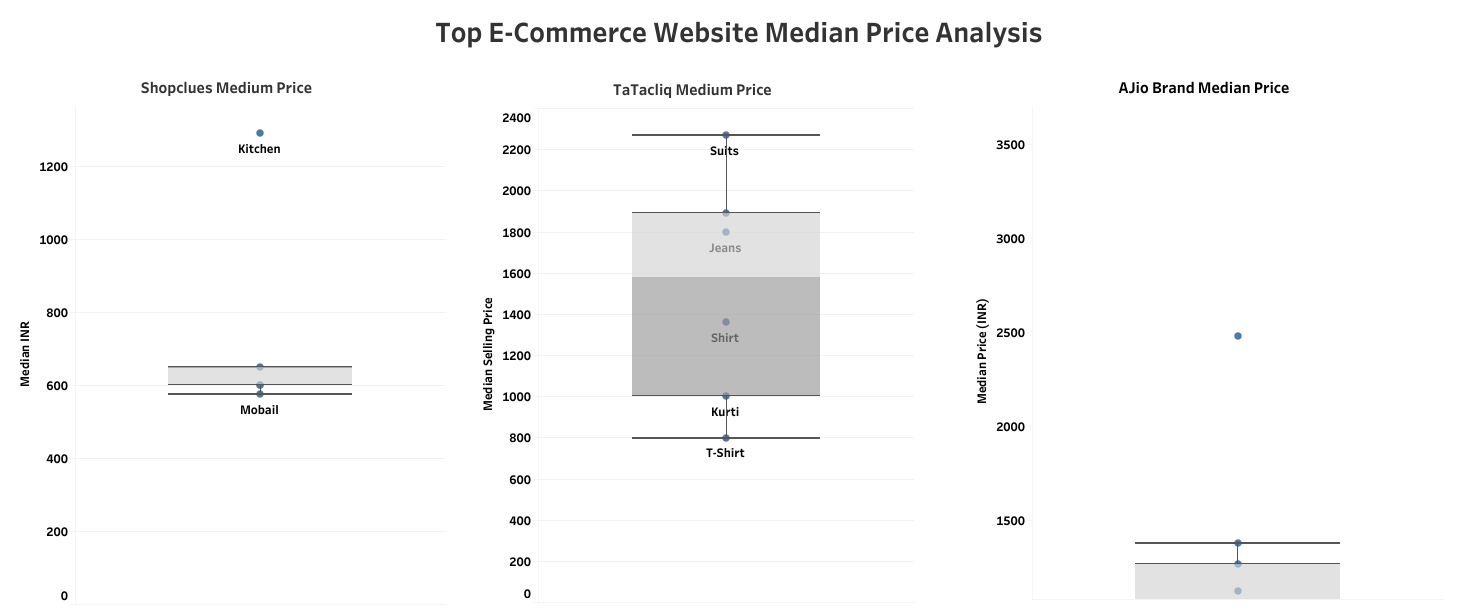
Use a box plot to display the distribution of data and detect outliers.

**What it shows:**

Box plots use a rectangular box to represent the interquartile range (IQR) and whiskers to. show the range of data. The median is represented by a line inside the box, and outliers are plotted as individual points.

**When to avoid it:**

Avoid using box plots when the data is categorical or when showing trends over time.



**Insights and Findings :**

We conducted web scraping of data from the top e-commerce websites. Our goal was to determine the median price of products based on their categories and identify products within a specific price range as outliers. To achieve this, we utilized a box plot, which helps visualize outliers in a dataset. For example, if the prices in a product category are 100, 150, 200, 311, 425, 400, and 18000, with a median price of 425, the product priced at 18000 would be considered an outlier as it falls outside the range of the median price.

**Insights:**

1. In the Kitchen category of Shopclues, the price is significantly higher than the Third Quartile.
2. In Tatacliq, the prices of suits and sarees are outside the Second Quartile.
3. In Ajio, the price of boots exceeds the Third Quartile.

**A Practical Guide to Stacked Bar Charts: Layered Category Visualization with Real Use Cases**

**Type of data:**

Categorical, quantitative

**When to use it:**

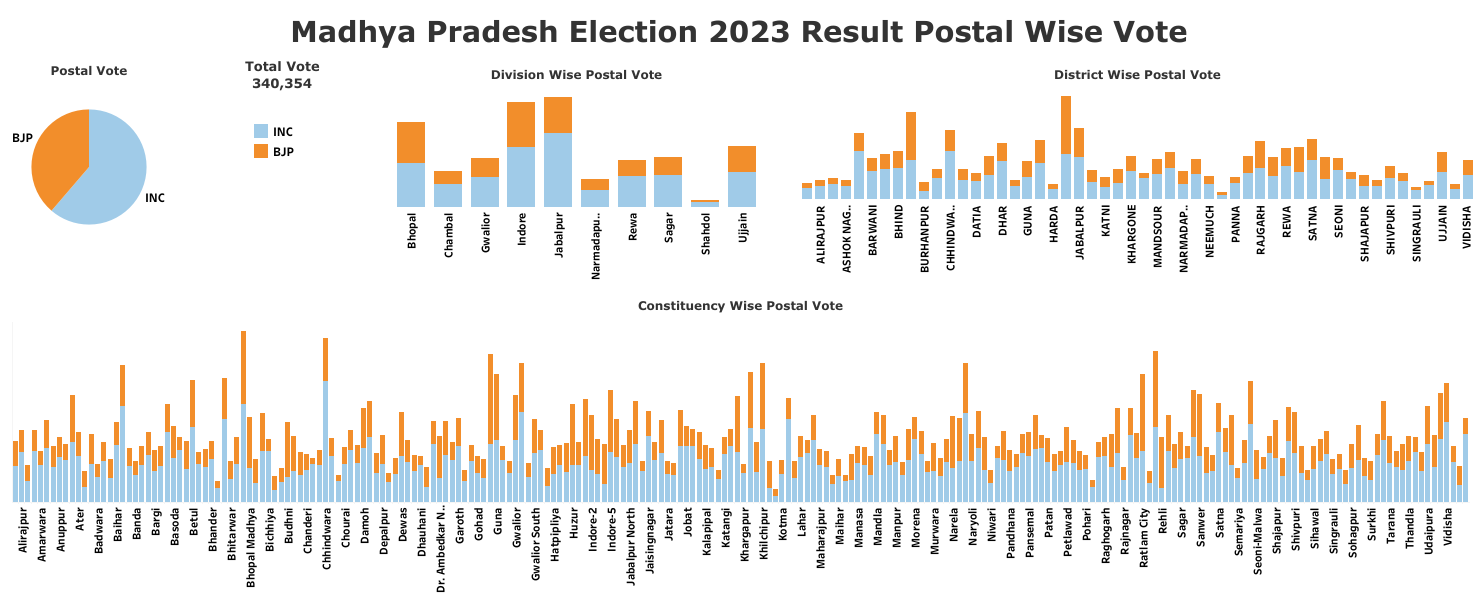
Use a stacked bar chart to compare data across categories and show the composition of each category.

**What it shows:**

Stacked bar charts display data using stacked rectangular bars, with the length of each segment representing the value. The total length of the bar represents the sum of all values in a category.

**When to avoid it:**

Don't use a stacked bar chart when there are too many categories or if the data is continuous.

****

**Insights and Findings :**

We wanted to determine the number of postal ballots for each seat in the Madhya Pradesh Assembly Election 2023 to understand which seats had more votes for the Congress and BJP. To achieve this, we utilized a stacked bar graph. In this graph, we represented the BJP votes in orange and the Congress votes in sky blue. This visualization method helped us compare and analyze the postal ballot counts for each seat, distinguishing between the two major parties.

**Insights:**

Through the stacked bar graph, it's evident that the Congress is represented in sky blue, and its length is greater, indicating it received 56% of the postal ballots. In contrast, the BJP's length is shorter, signifying it received 34% of the votes. Other parties have much smaller lengths, indicating they received very few votes.

**A Practical Guide to Gantt Charts: Visualizing Project Timelines with Real-World Applications**

**Type of data:**

Time-based, project management

**When to use it:**

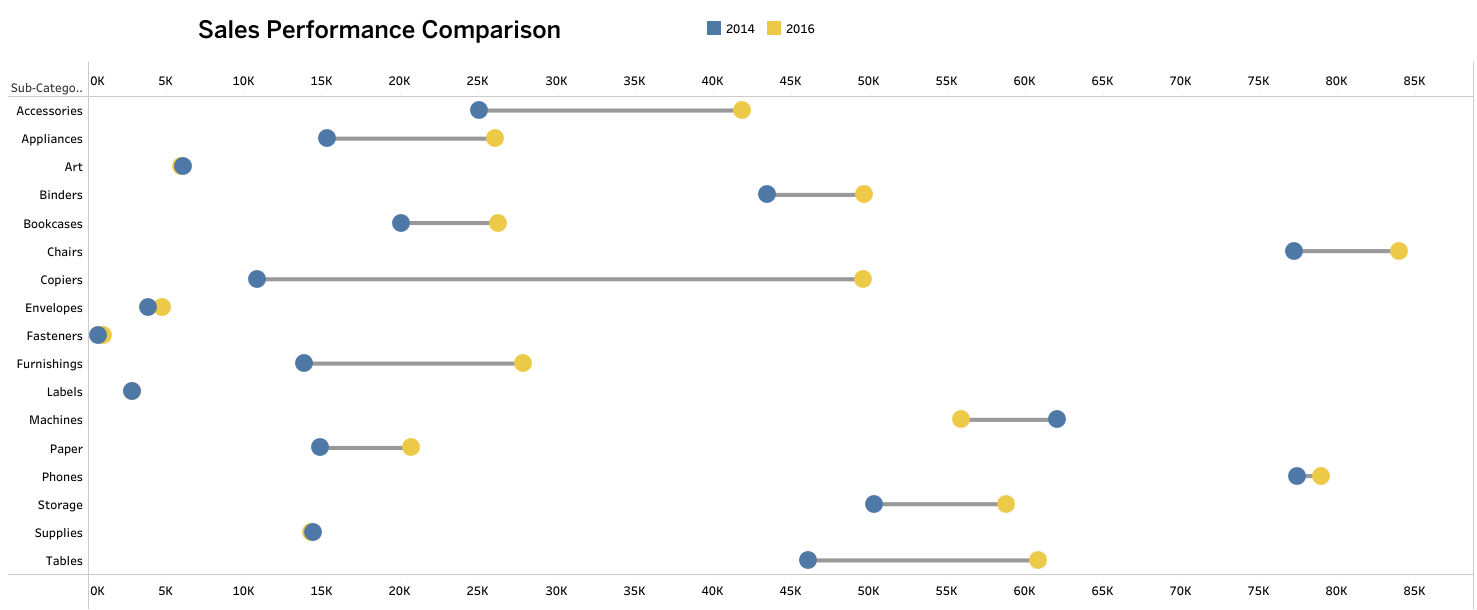
Use a Gantt chart to visualize project schedules, tasks, and Milestones.

**What it shows:**

Gantt charts use horizontal bars to represent tasks, with the length of the bar indicating the duration. The X-axis represents time, and tasks can be grouped by category.

**When to avoid it:**

Avoid using Gantt charts when the data is not time-based or when showing relationships between variables.

****

**Insights and Findings :**

We needed to compare the performance of sales data for two years to determine which year had the highest sales and whether sales increased or decreased in the other year. To accomplish this, we used a Gantt chart. In this chart, we represented the sales of 2016 in yellow and the sales of 2014 in blue. The line between these two colors compares the sales performance, indicating whether sales increased or decreased in the other year.

**Insights:**

1. When it comes to the Copiers product category, there was a significant increase in sales in 2016 compared to 2014.
2. For the Accessories, Appliances, and Suppliers categories, their sales increased compared to 2014.
3. There isn't much difference in the sales of other products.
4. Sales decreased in the labels and Machine categories in 2016 compared to 2014.

**A Practical Guide to Waterfall Charts: Stepwise Financial Impact Visualization with Examples**

**Type of data:**

Quantitative, sequential

**When to use it**:

Use a waterfall chart to visualize the cumulative effect of sequential data, such as financial or inventory changes.

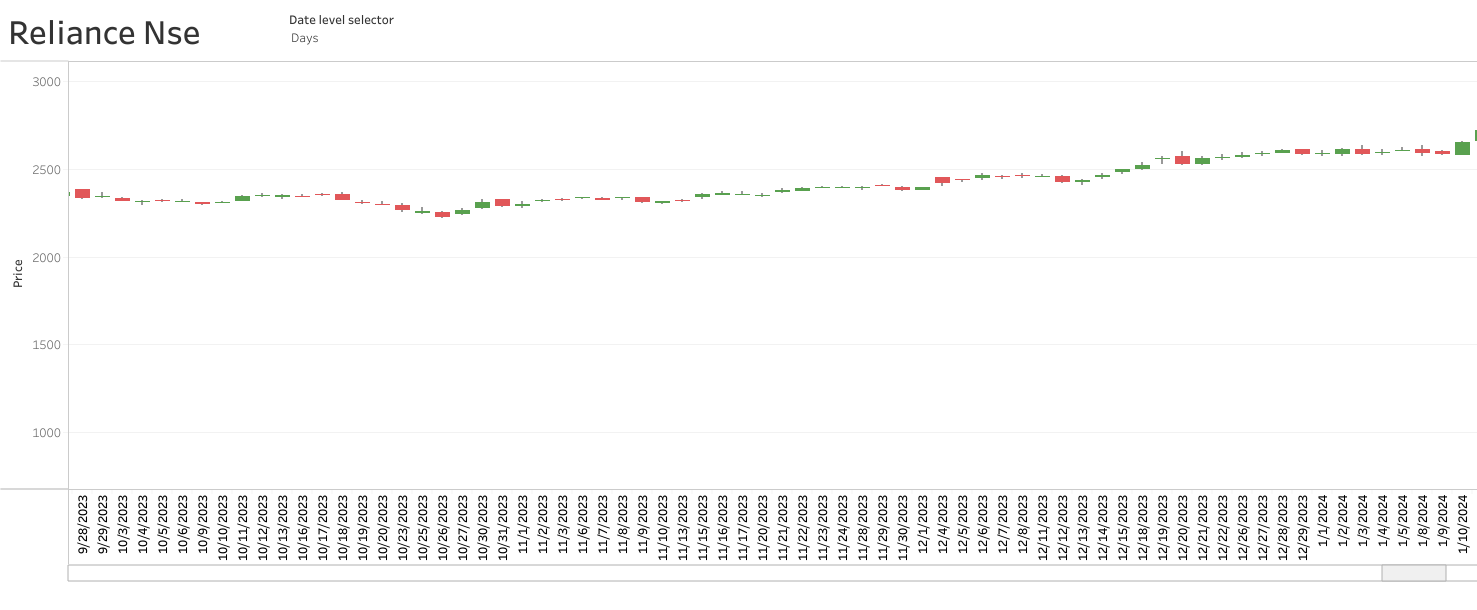
**What it shows:**

Waterfall charts use vertical bars to represent the value of each step, with the bars connected to show the cumulative effect. Positive and negative values can be represented using different colors.

**When to avoid it:**

Don't use a waterfall chart when there is no logical order or

relationship between data points.

****

**Insights and Findings :**

We used a waterfall chart to analyze the time series data of Reliance's NSE stock market. In this chart, when the stock price increases, it's represented in green, and when it decreases, it's shown in red. This visualization method allows us to easily identify the dates when the stock price either increased or decreased.

**Insights:**

1. When the chart displays green, it means the stock price increased on that date.
2. When the chart displays red, it indicates a decrease in the stock price on that date.
3. The trend line suggests that the stock price is likely to increase in the future.

**A Practical Guide to Funnel Charts: Analyzing Conversion Flows with Real-World Data**

**Type of data:**

Categorical, process stages

**When to use it:**

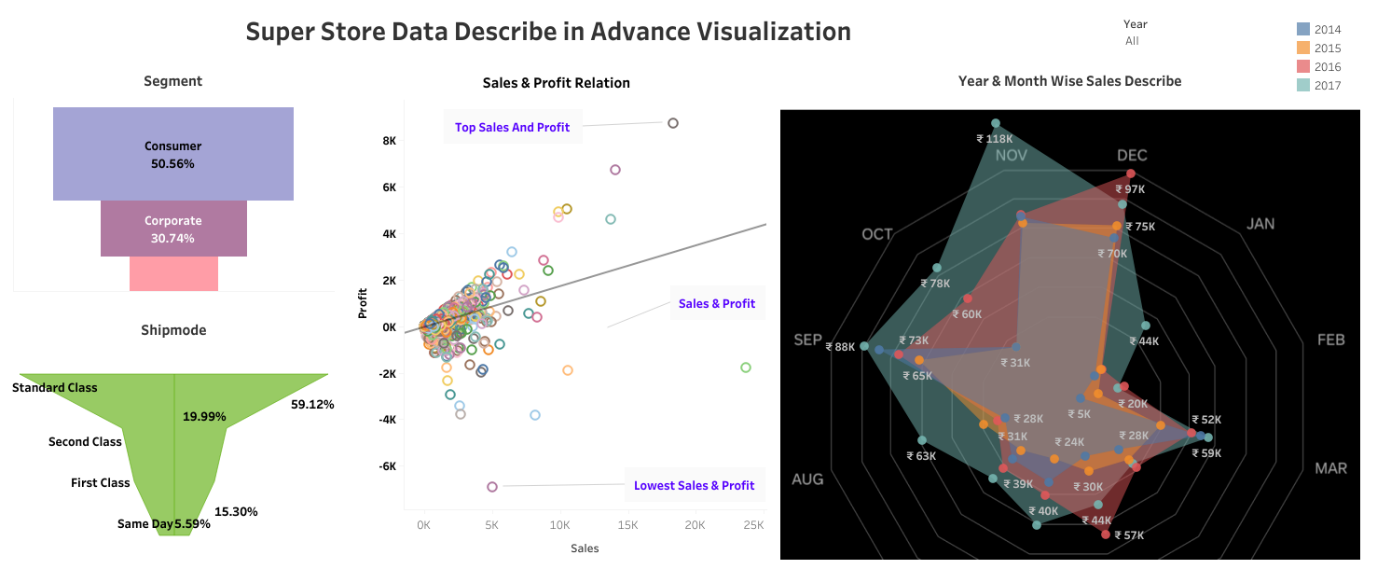
Use a funnel chart to visualize the stages of a process, such as sales, conversions, or customer journeys.

**What it shows:**

Funnel charts use a series of decreasing trapezoids to represent each stage in a process, with the width of each trapezoid proportional to the number of items at that stage.

**When to avoid it:**

Avoid using funnel charts when there is no straightforward process, or the data is continuous.

****

**Insights and Findings :**

To determine the sales in each product category and compare them with competition from other product categories, we used a funnel chart. This chart allows us to view each category as a percentage of total sales, helping us understand which product has competed with others and how much sales each category has garnered.

**Insights:**

1. Consumer sales account for 50%, making it the highest-selling category, placing it at the first position.
2. Corporate sales are moderate, comprising 30% of total sales, positioning it at the second spot in comparison to all categories.
3. Home office sales are the lowest, representing only 18% of total sales, making it the last category in comparison to all others.

**A Practical Guide to Rounded Edges Bar Charts: Stylish Category Comparison with Real Examples**

**Type of data:**

Quantitative, sequential

**When to use it**:

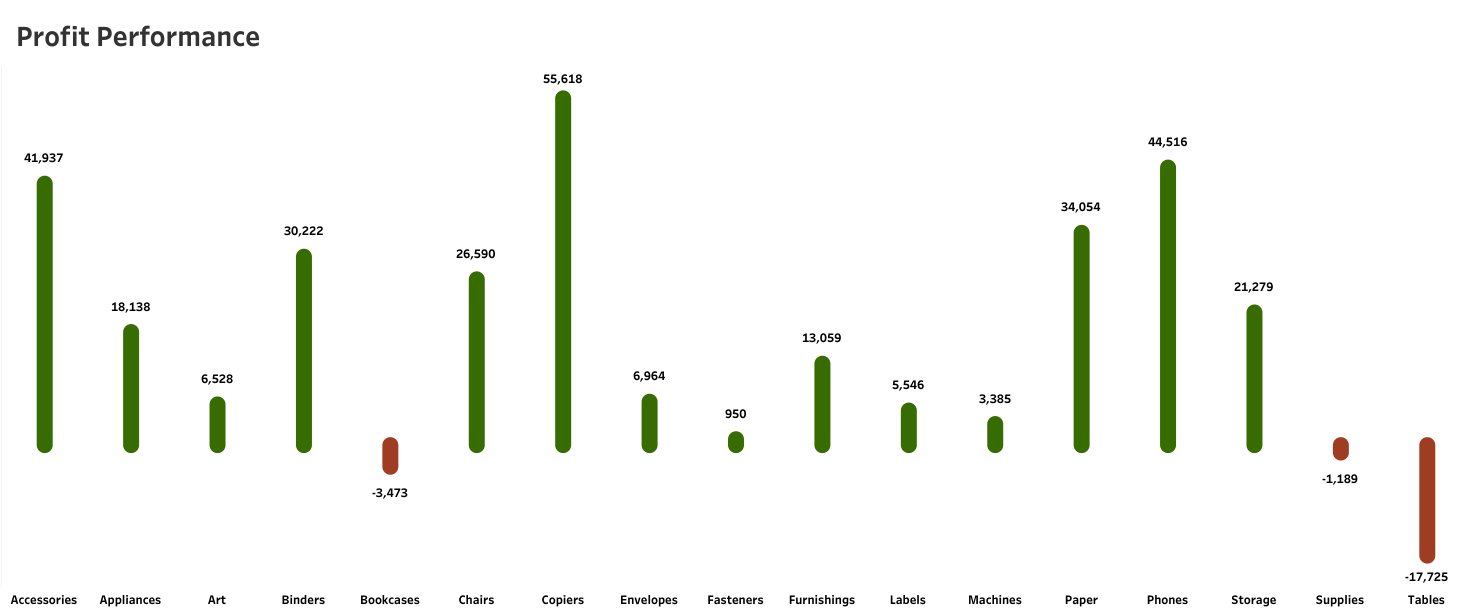
Use a waterfall chart to visualize the cumulative effect of sequential data, such as financial or inventory changes.

**What it shows:**

Waterfall charts use vertical bars to represent the value of each step, with the bars connected to show the cumulative effect. Positive and negative values can be represented using different colors.

**When to avoid it:**

Don't use a waterfall chart when there is no logical order or relationship between data points.

****

**Insights and Findings :**

We wanted to assess the profit performance of all product categories in our business to identify which products are yielding the highest profits, which ones have average profitability, and identify any products that are incurring losses. To achieve this, we utilized Rounded Edge Bar charts. In these charts, products with longer bars represented in green indicate the highest profitability, while products with shorter bars represented in red indicate losses. This visualization method allows us to easily identify the most profitable products, those with average profitability, and any products experiencing losses.

**Insights:**

1. Copiers are generating the highest profits.
2. Accessories, Binders, Paper, Phones, Storage, and other products are yielding average profits.
3. Products such as Art, Fasteners, Labels, and Machines are resulting in very low profits.
4. Both Bookcases and Supplies are causing losses.
5. Tables are resulting in significant losses.

**A Practical Guide to Target Bar Graphs: Comparing Actual vs Target Values with Real Insights**

**Type of data**:

Categorical, quantitative

**When To Use it:**

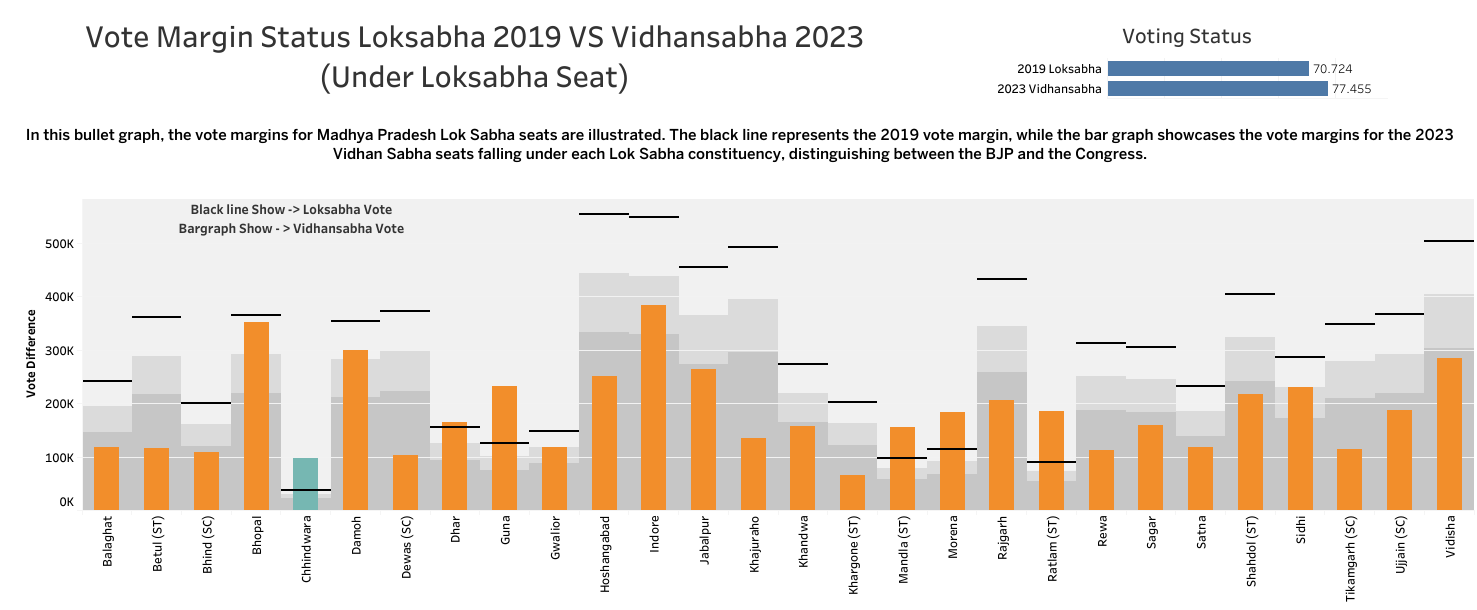
Use a bar Chart to compare data across categories.

**What it shows:**

Bar charts display data using rectangular bars, with the length of the bar representing the value. The bars can be horizontal or vertical.

**When to avoid it:**

Avoid using a bar chart when there are too many categories or if the data is continuous.

****

**Insights and Findings :**

To analyse the trends of LokSabha seats for the 2024 elections, we needed to determine the vote differences in each constituency. For this, we compared the Lok Sabha 2019 election results with the 2023 Madhya Pradesh Vidhan Sabha election results, as these constituencies fall under Lok Sabha seats. We calculated the total votes for each seat in both elections and then compared them to the total votes in the 2019 Lok Sabha results to identify whether there was an increase or decrease in the votes for both BJP and Congress between the two elections. To simplify this analysis, we utilized a target bar graph, where the target line represented the 2019 Lok Sabha results and the bars represented the 2023 Vidhan Sabha results. This enabled us to easily determine whether the vote margin increased or decreased for each seat in comparison to 2019.

**Insights:**

1. Seats such as Betul, Dewas, Hoshangabad, Jabalpur, Khajuraho, Rajgarh, Tikamgarh, and Vidisha are under BJP's control, but in comparison to the 2023 Vidhan Sabha elections, there has been a significant increase in Congress votes, with BJP's votes halving. This indicates a potential shift in these seats for the upcoming 2024 Lok Sabha elections.
2. Seats like Balaghat, Bhind, Indore, Khandwa, Khargone, Rewa, Sagar, Satna, Shahdol, and Ujjain are dominated by BJP. While Congress votes have increased moderately compared to the 2019 Lok Sabha elections, BJP's vote count remains stable.
3. In seats such as Bhopal, Damoh, Dhar, Gwalior, and Sidhi, BJP holds control. However, there hasn't been any significant change in BJP's vote count between the 2019 Lok Sabha elections and the 2023 Vidhan Sabha elections.
4. Chhindwara seat has been under Congress's control, and there has been an increase in Congress's vote margin between the 2019 Lok Sabha elections and the 2023 Vidhan Sabha elections.
5. Seats like Guna, Mandla, Morena, and Ratlam are dominated by BJP. Observing the margin between the 2019 elections and the 2023 Vidhan Sabha elections, BJP's margin has increased, indicating a potential stronghold for BJP in these seats.