Bar Graph

A **bar graph** is a graphical representation of information. It uses bars that extend to different heights to depict value.

Types of Bar Graphs

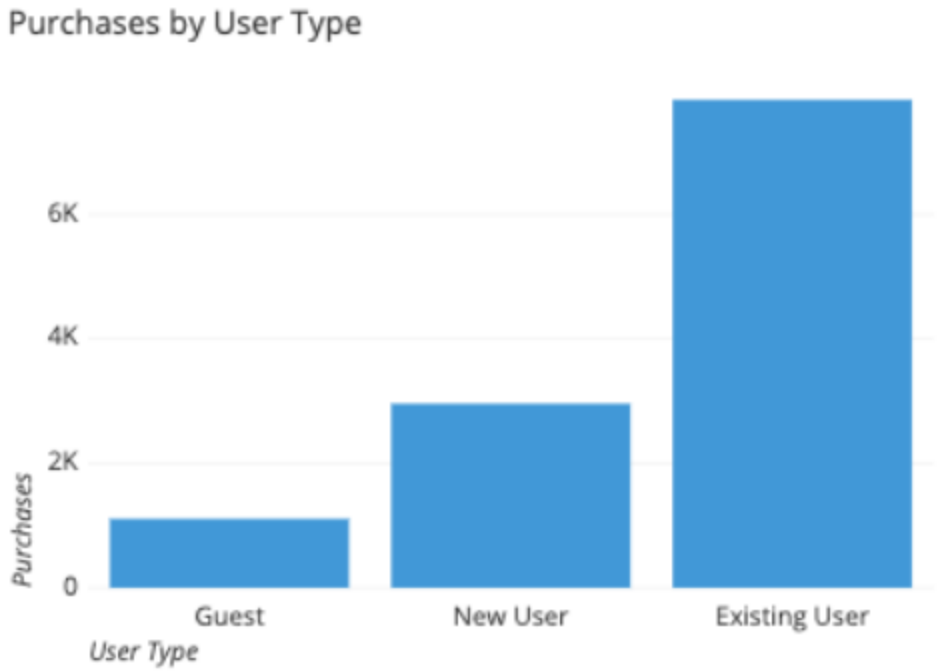
- **Vertical Bars**
- **Horizontal Bars**
- **Grouped Bars** (multiple bars that compare values in a category)
- **Stacked Bars** (bars containing multiple types of information)

Uses of Bar Graphs

Bar graphs are commonly used in business and financial analysis to display often complicated data. They can convey information quickly and effectively.

Example in Finance

In the financial industry, a **volume chart** is a commonly used vertical bar graph.



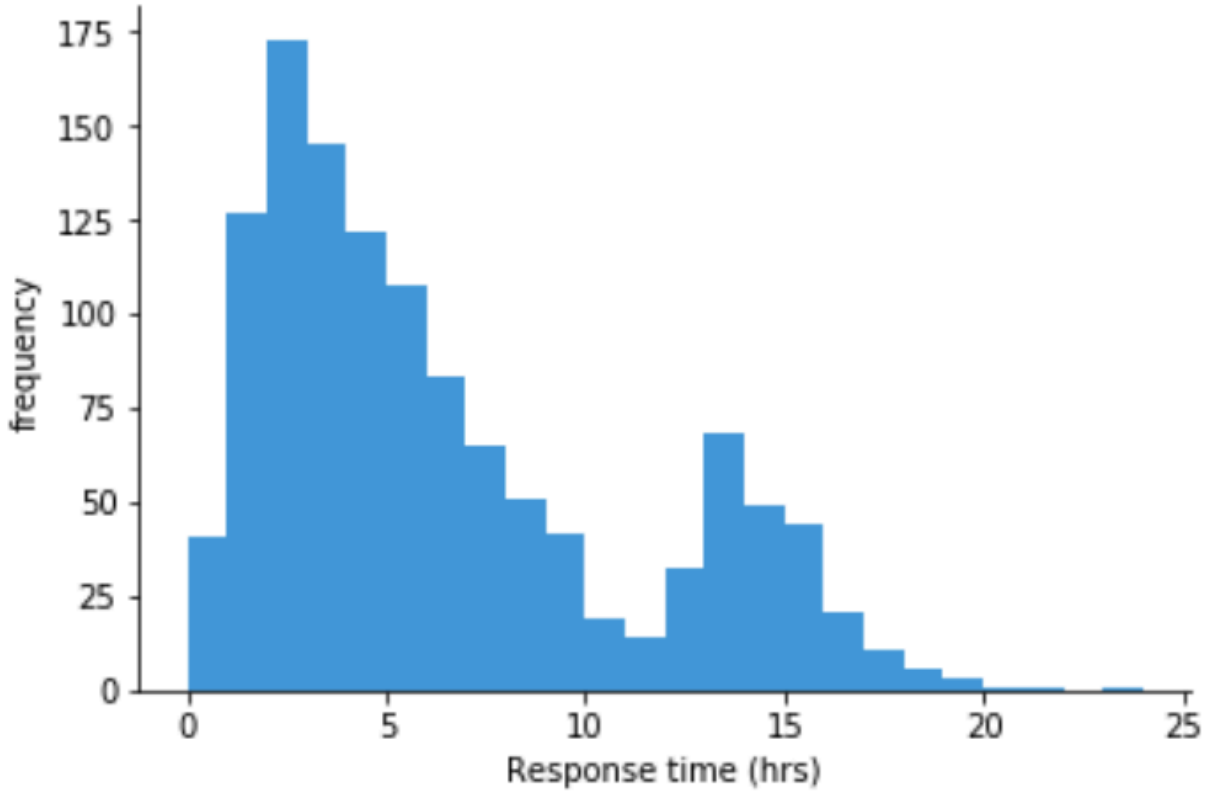
Bar Graph Properties

Certain aspects of a bar graph separate them from other types of graphs and charts.

- The bars on a bar graph have equal width and interval spacing.
- Bars can run vertically or horizontally.
- Bars share the same starting point or base. In other words, all bars will start at the bottom of the graph and extend upward (vertically) or they'll start at the side of the graph and extend across (horizontally).
- The y-axis of a bar graph is the side or vertical axis.
- The x-axis of a bar graph is the bottom or horizontal axis.
- Data value is defined on the y-axis; data type is defined on the x-axis.
- Bar height or extension corresponds to the value of data.
- The higher or longer a bar, the greater the value.
- If colors are used, a bar graph may include a legend that defines them.

Histogram

A **histogram** is a chart that plots the distribution of a numeric variable's values as a series of bars. Each bar typically covers a range of numeric values called a **bin** or **class**; a bar's height indicates the frequency of data points with a value within the corresponding bin.

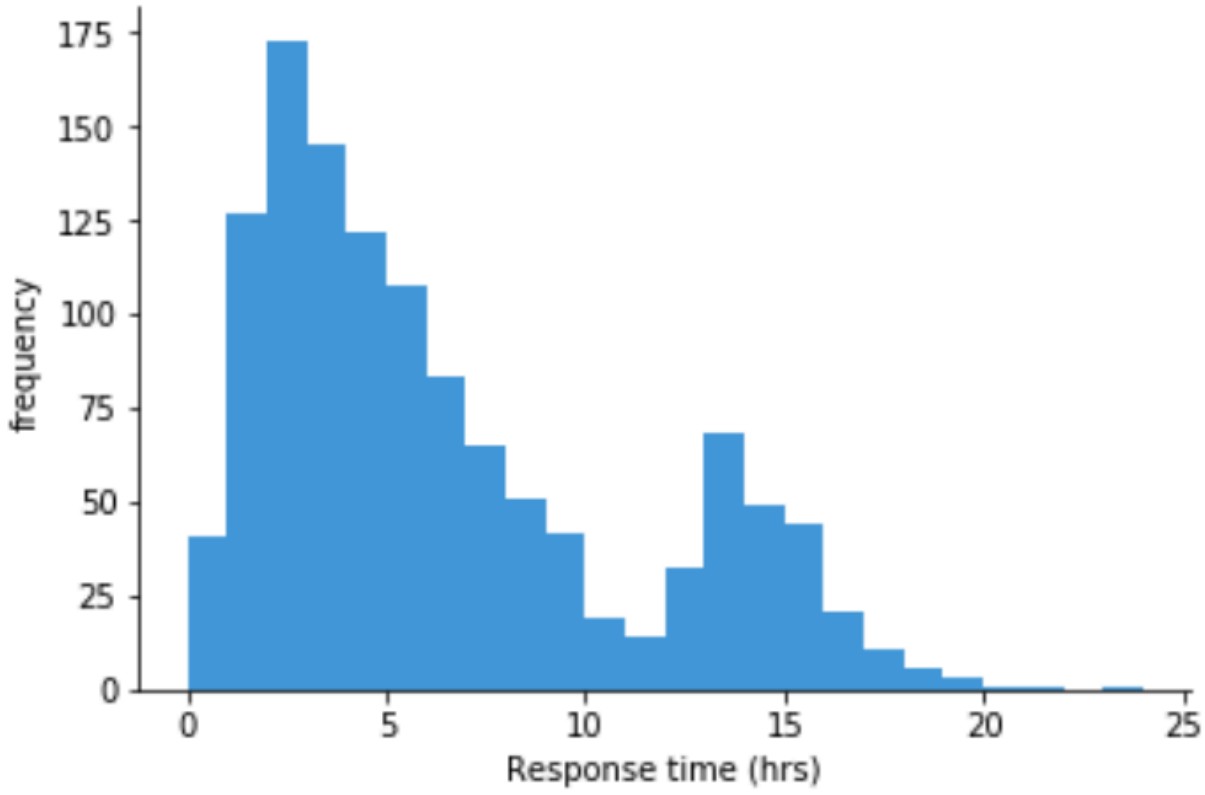


Example

The histogram above shows a frequency distribution for time to response for tickets sent into a fictional support system. Each bar covers one hour of time, and the height indicates the number of tickets in each time range. We can see that the largest frequency of responses was in the **2-3 hour range**, with a longer tail to the right than to the left. There's also a smaller hill whose peak (mode) is at the **13-14 hour range**. If we only looked at numeric statistics like **mean** and **standard deviation**, we might miss the fact that there were these two peaks that contributed to the overall statistics.

When You Should Use a Histogram

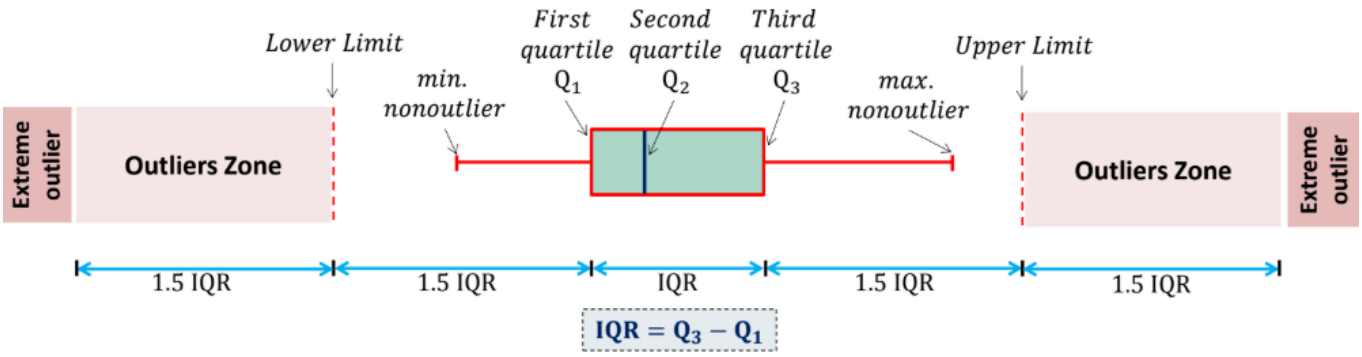
Histograms are good for showing general **distributional features** of dataset variables. You can see roughly where the **peaks** of the distribution are, whether the distribution is **skewed or symmetric**, and if there are any **outliers**.



Box PLot

- A box and whisker plot—also called a box plot—displays the five-number summary of a set of data. The five-number summary is the minimum, first quartile, median, third quartile, and maximum.

- In a box plot, we draw a box from the first quartile to the third quartile.
  A vertical line goes through the box at the median. The whiskers go from each quartile to the minimum or maximum.



Use-Cases of Box Plot

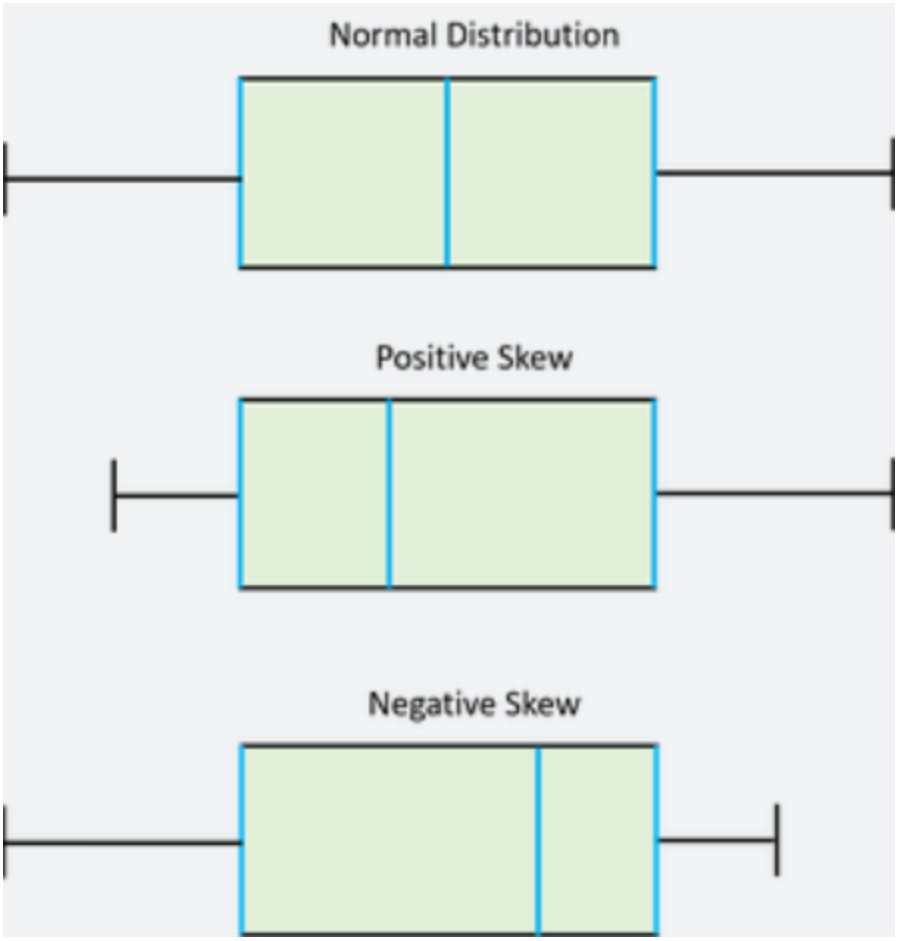
Box plots provide a visual summary of the data with which we can quickly identify:

- The average value of the data.
- How dispersed the data is.
- Whether the data is skewed or not (skewness).
- **The Median** gives you the average value of the data.
- **Box Plots show Skewness** of the data - The dispersion or spread of data can be visualized by the minimum and maximum values, which are found at the end of the whiskers.
- **The Box plot helps identify Outliers** - Outliers are points that are numerically distant from the rest of the data.

a) If the Median is at the center of the Box and the whiskers are almost the same on both ends, then the data is **Normally Distributed**.

b) If the Median lies closer to the **First Quartile** and the whisker at the lower end is shorter, then it has a **Positive Skew (Right Skew)**.

c) If the Median lies closer to the **Third Quartile** and the whisker at the upper end is shorter, then it has a **Negative Skew (Left Skew)**.



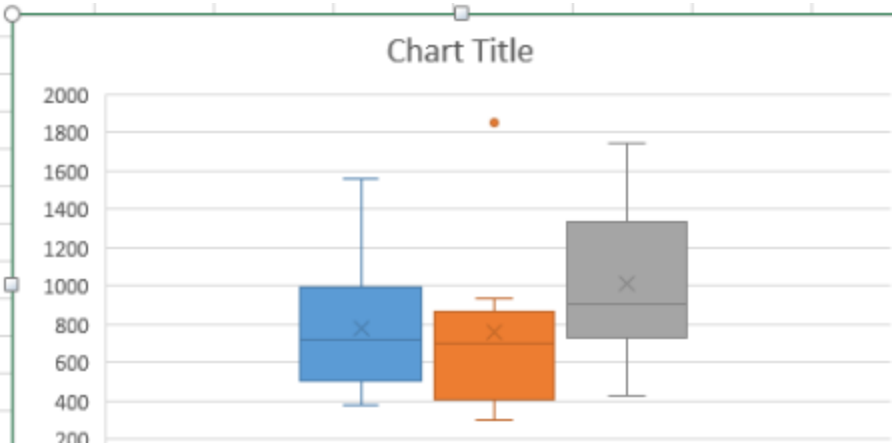
What information cannot be found in a box plot?

- Information that is missing in a box plot includes the **detailed shape of the distribution**. It is quite difficult to find the **mean** as it is a visual representation of the data.

Is a Box Plot vertical or horizontal?

- A Box Plot can be drawn **either horizontally or vertically**. It depends on the estimate **L-estimators, range, mid-range, and trimean**.

Brand A	Brand B	Brand C
1020	840	1430
1560	940	1750
560	780	870
780	650	920
990	720	1300
670	430	890
510	1850	740
490	300	720
380	360	430
880	690	1050



Stem and Leaf

What are Stem and Leaf Plots?

A **stem and leaf plot** is a graphical representation used to organize and display quantitative data in a semi-tabular form. It helps in visualizing the **distribution** of the data set and retains the **original data values**, making it easy to identify the **shape, central tendency, and variability** of the data.

A stem and leaf plot splits each data point into a **"stem"** and a **"leaf."**

- The **"stem"** represents the leading digits.
- The **"leaf"** represents the trailing digit.

This separation makes it easy to organize data and see patterns.

Key Features of Stem and Leaf Plots

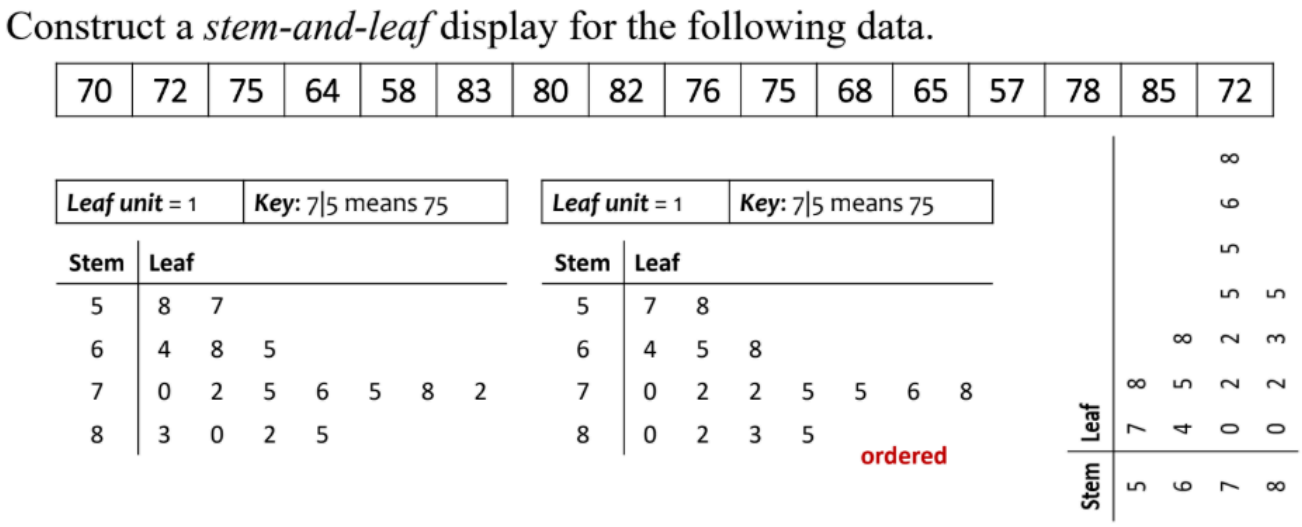
Some of the key features of **stem and leaf plots** are:

- **Numerical Data Representation:**
  - Stem and leaf plots display numerical data by splitting each data point into:
    - A **"stem"** (usually the first digit or digits).
    - A **"leaf"** (usually the last digit).
- **Vertical Listing:**
  - The stems are listed **vertically**, and the leaves are written next to their corresponding stem in numerical order.
- **Key for Interpretation:**
  - A **key** is provided to explain what the stem and leaf represent for that particular plot.
- **Data Distribution Insights:**
  - Stem and leaf plots help visualize the **shape, spread, and central tendency** of a data distribution.
  - They can highlight **outliers** and help identify the **mode**.
- **Retention of Original Data Values:**
  - Unlike histograms, stem and leaf plots **retain the original data values**, rather than just showing summary information.
- **Preservation of Scaling:**
  - Stems with no leaves are still included to **preserve the horizontal axis scaling** and highlight gaps in the data.

How to Create Stem and Leaf Plots?

To create the **stem and leaf plot** of any data set, follow these steps:

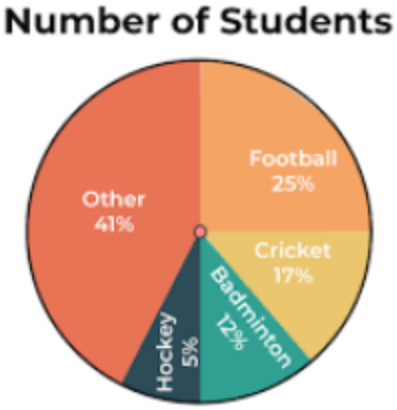
Step	Description
Step 1	Arrange your data set in **ascending order**.
Step 2	Identify the **"stems"** by separating the leading digits of the numbers.
Example	If your data includes numbers like **23, 25, and 27**, the stem would be **2**.
Step 3	Identify the **"leaves"** by taking the trailing digits of the numbers.
Example	In the number **23**, the leaf would be **3**.
Step 4	Write down the **stems** in a **vertical column**.
Step 5	Next to each stem, write down the corresponding **leaves in ascending order**.



Pie Chart

A **pie chart** is a popular and visually intuitive tool used in data representation, making complex information easier to understand at a glance.

This circular graph divides data into **slices**, each representing a proportion of the whole, allowing for a clear comparison of different categories. This makes it easier to digest complex information through a straightforward, intuitive format.



Violin Plot

A **violin plot** depicts distributions of numeric data for one or more groups using **density curves**. The width of each curve corresponds with the approximate frequency of data points in each region.

Densities are frequently accompanied by an **overlaid chart type**, such as a **box plot**, to provide additional information.