## Different types of graphs and charts

### Line Chart

- **When to Use**: Line charts are used to track changes over short and long periods of time. They are suitable for showing trends and comparing changes over time for one or more groups.
- **Best Practices**: It is recommended to have a zero baseline, and the range from the smallest to the largest values should take up about 70 to 80 percent of the available vertical space.

### Area Graph

- **When to Use**: Area graphs are used to display how values or multiple values develop over time, highlight the magnitude of a change, and show large differences between values.
- **Best Practices**: Avoid displaying more than four categories and use transparent colors to avoid obscuring data in the background.

### Bar Graph

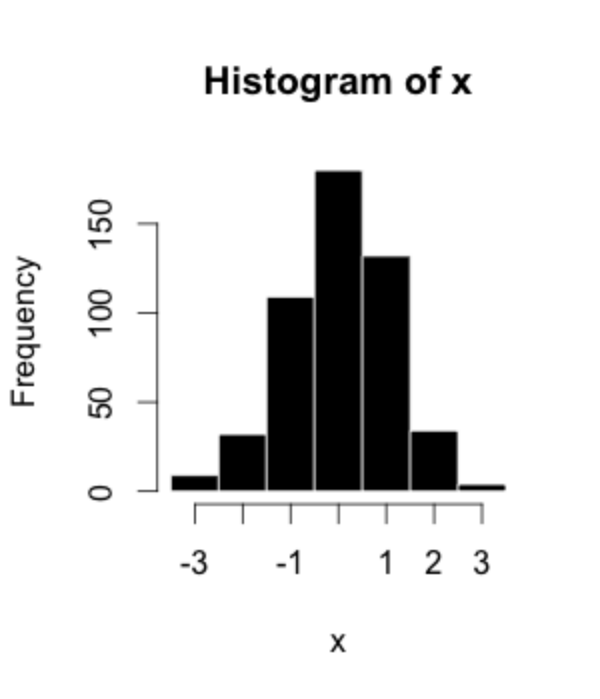
- **When to Use**: Bar graphs are used to compare things between different groups or to track changes over time. They are best for measuring larger changes over time and for showing numbers that are independent of each other.
- **Best Practices**: Useful for comparing discrete categories and for displaying changes over time.

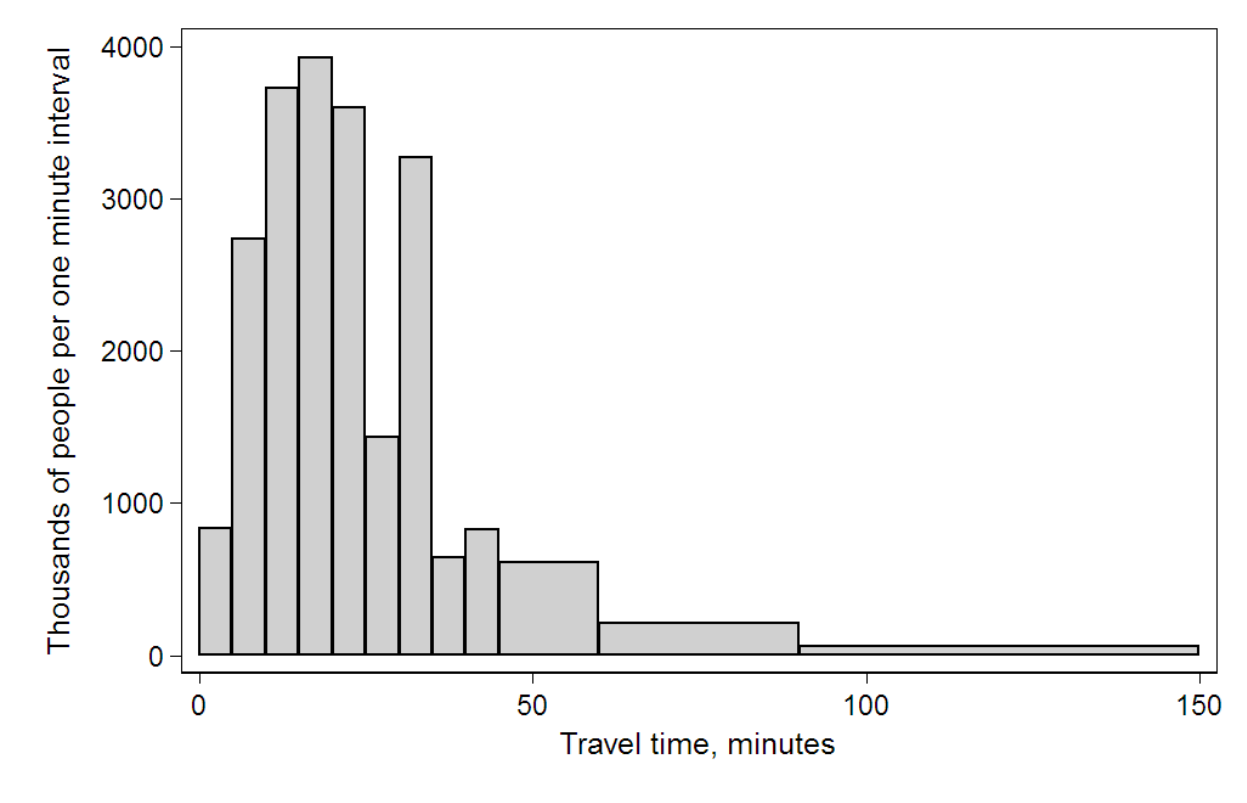
### Pie Chart

- **When to Use**: Pie charts are best used when comparing parts of a whole. They do not show changes over time and are suitable for illustrating the composition of a whole.

### X-Y Plot (Scatter Plot)

- **When to Use**: Scatter plots chart two unique variables to show the relationship between them. They are used to display the correlation or relationship between two variables.



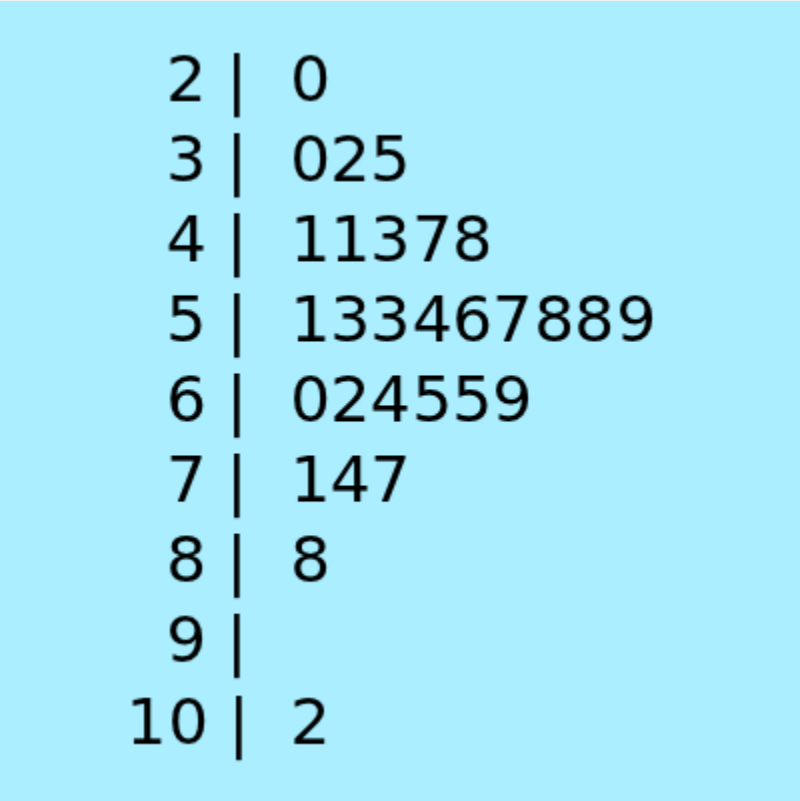

A histogram is a graphical representation of data that uses bars of different heights to display the frequency of occurrences within specific ranges. Unlike a bar chart, which represents groups defined by a categorical variable, a histogram groups numbers into ranges based on continuous and quantitative

variables. The height of each bar in a histogram indicates how many data points fall into each range, and the ranges are decided by the creator of the histogram.

For example, if you were measuring the height of trees in an orchard, you could group the heights into ranges, such as 100-150 cm, 150-200 cm, and so on. Each bar in the histogram would then represent the number of trees within each height range.

Histograms are particularly useful for recognizing and analyzing patterns in continuous data that may not be immediately apparent from just looking at the raw data or calculating averages and medians. They can also help in identifying outliers and understanding the shape and spread of the data.

It's important to note that the height of the bars in a histogram does not directly indicate the frequency of occurrences within each individual bin. Instead, it is the area of the bar that represents the frequency of occurrences for each bin. This distinction is important, especially when dealing with histograms that have equally spaced bins.




A Stem and Leaf Plot is a special table where each data value is split into a "stem" (the first digit or digits) and a "leaf" (usually the last digit). The "stem" values are listed down, and the "leaf" values go right (or left) from the stem values. The "stem" is used to group the scores, and each "leaf" shows the individual scores within each group. Stem and Leaf Plots are particularly advantageous as they group the data and display all the original data, making it easier to observe the frequency of different types of values. Stem and Leaf Plots are especially useful for identifying the actual observations within a specific

interval, which may not be apparent from other types of data representations. They are a valuable tool for organizing and visualizing numerical data, providing insights into the distribution and characteristics of the data being analyzed.

To make a box and whisker plot, follow these general steps:

1. **Organize the Data**: Start by organizing the numbers in your data set from least to greatest.
2. **Find the Median**: Determine the median of the data set. If there is an odd number of values, the median is the middle number. If there is an even number of values, the median is the average of the two middle numbers.
3. **Identify Quartiles**: Find the first quartile, which is the median of the lower half of the data set, and the third quartile, which is the median of the upper half of the data set.
4. **Draw the Plot Line**: Draw a horizontal line and mark the quartiles and the median on it.
5. **Create the Box**: Connect the quartiles and median with horizontal lines to form a box.
6. **Mark the Whiskers**: Draw lines (whiskers) from each end of the box to the minimum and maximum values of the data set.
7. **Identify Outliers**: Mark any outliers, which are data points that fall significantly outside the overall pattern of the data.

Box and whisker plots are a valuable tool for visually displaying the median, lower and upper quartiles, and lower and upper extremes of a set of data. They help to see the variance of data and can be very helpful in understanding the distribution and characteristics of the data being analyzed.

To make a dot plot, follow these general steps:

1. **Organize the Data**: Start by organizing the numerical data you want to represent in the dot plot.
2. **Create the Axis**: Draw a horizontal axis and mark the numerical values along it.
3. **Place the Dots**: For each data point, place a dot above the corresponding value on the axis. If there are multiple data points with the same value, stack the dots vertically.
4. **Label the Axis**: Ensure that the numerical values on the axis are clearly labeled to indicate the meaning of the dots.

Dot plots are a simple yet effective way to display the distribution of numerical data. They are particularly useful for representing small amounts of data and for comparing the frequency of different values.

## What is a Pie Chart?

A pie chart is a graphical representation technique that displays data in a circular-shaped graph. It is a composite static chart that works best with few variables. Each category of data is represented as a slice of the pie, and the size of each slice is directly proportional to the number of data points that

belong to that category. The entire pie represents the whole data set, with each slice representing a fraction or proportionate part of the whole. The slices of the pie are often represented with percentages, signifying their contribution to the whole. This makes it easier to visualize and understand the proportionate parts or composition of a data set. Pie charts were invented by William Playfair in 1801 and are widely used in various sectors due to the advantages they offer in visually representing data .

It's important to ensure that the categories in a pie chart do not overlap, as this would result in a misleading representation. For example, if a chart represents the breeds of dogs owned by U.S. pet owners, a husky-German shepherd mix should not be placed in two categories, as this would create larger pieces of the pie for those categories

.

## What is a Bar Chart?

A bar chart, also known as a bar graph, is a chart or graph that presents categorical data with rectangular bars, with heights or lengths proportional to the values they represent. The bars can be plotted vertically or horizontally, with a vertical bar chart sometimes called a column chart. A bar graph shows comparisons among discrete categories, with one axis of the chart showing the specific categories being compared, and the other axis representing a measured value. Some bar graphs present bars clustered in groups of more than one, showing the values of more than one measured variable. Bar charts provide a visual presentation of categorical data, which is a grouping of data into discrete groups, such as months of the year, age group, shoe sizes, and animals. These categories are usually qualitative. In a column (vertical) bar chart, categories appear along the horizontal axis, and the height of the bar corresponds to the value of each category. Bar charts have a discrete domain of values and are widely used for comparing data across different categories.

Bar charts are a fundamental chart type and a useful tool for exploring and understanding data. They are particularly effective for comparing different categories and understanding the distribution of values within each category.

## Symmetry and Skewness (1.8)

Symmetry and Skewness (1.8)

Symmetry and skewness are important concepts in statistics that describe the shape and distribution of data.

**Symmetry**:

- A distribution is said to be symmetrical when the distribution on either side of the mean is a mirror image of the other. In a symmetrical distribution, mean = median = mode

- .
- Symmetry implies that the data is evenly distributed around the mean, with equal tails on both sides of the distribution.

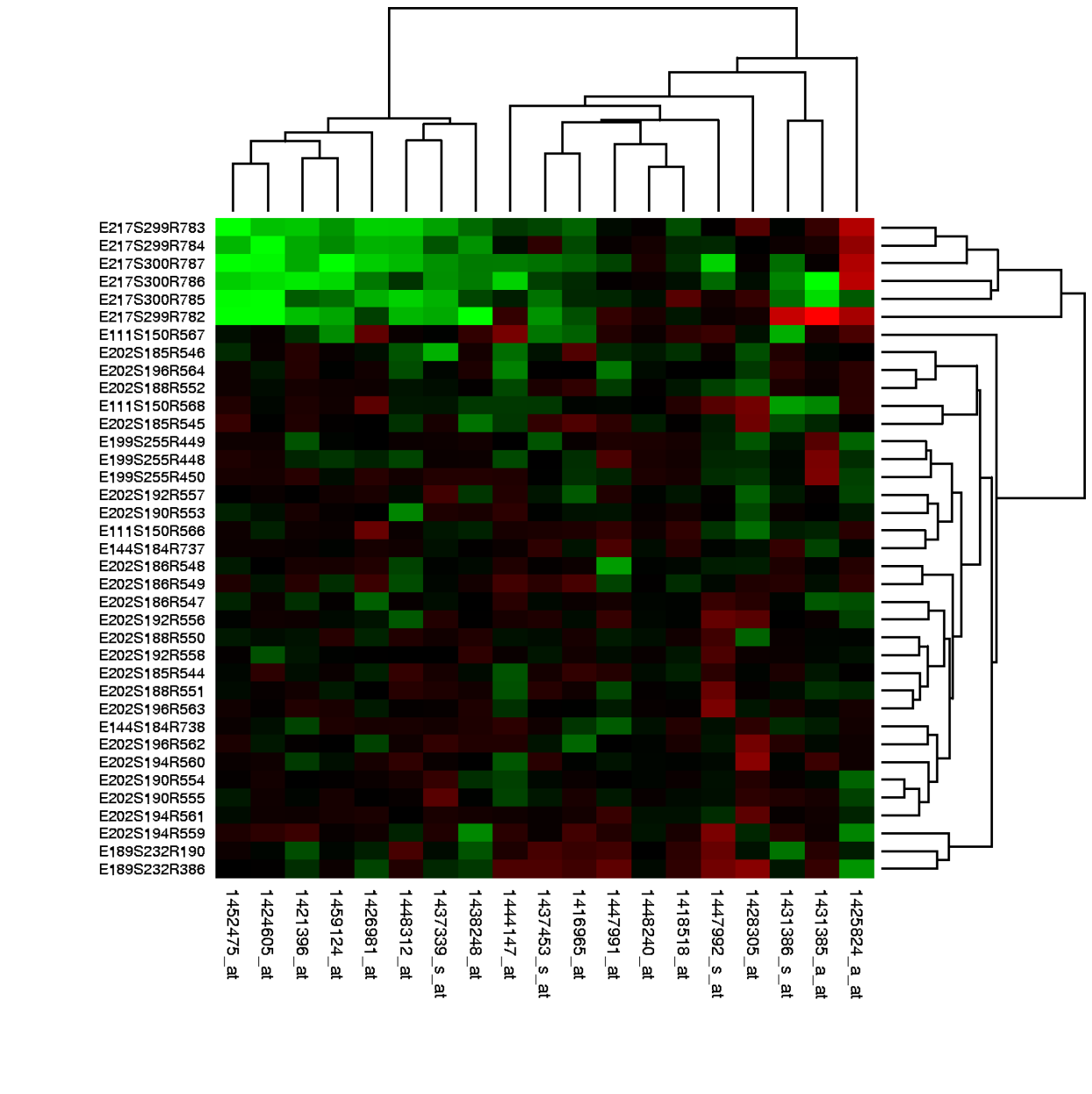
**Skewness**:

- If a distribution is non-symmetrical, it is said to be skewed. Skewness can be negative or positive.
- A positively skewed distribution has a long tail on the right side, which means that there will be frequent small losses and few large gains. Here the mean > median > mode. The extreme values affect the mean the most, pulling it to the right

- .

- A negatively skewed distribution has a long tail on the left side, which means that there will be frequent small gains and few large losses. Here the mean < median < mode. The extreme values affect the mean the most, pulling it to the left .

## What is a Heatmap?




A heatmap, also known as a heat map, is a two-dimensional representation of data in which various values are depicted by colors. It provides a visual summary of information across two axes, allowing users to quickly grasp the most important or relevant data points. Heatmaps use different colors or shades of the same color to represent different values and to communicate the relationships that may exist between the variables plotted on the x-axis and y-axis. Usually, a darker color or shade represents a higher or greater quantity of the value being represented in the heatmap. Heatmaps are widely used in various forms of analytics, such as showing user behavior on specific web pages, displaying the results of eye-tracking tests, and providing data visualization to track the efficiency of websites. They are a powerful way to understand what

users do on a website, including where they click, how far they scroll, and what they engage with, among other insights. Heatmaps are a valuable tool for businesses to discover actionable insights and identify user behavior, making it easier to understand complex data at a glance.

## Violin Plot [Simply explained]

A **violin plot** is a method of visualizing the distribution of numerical data. It is a combination of a box plot and a density plot, providing a more comprehensive representation of the data's distribution and probability density. The violin plot is particularly useful for comparing the distribution of data between different groups or categories.

Here's a breakdown of the key components and functions of a violin plot:

- **Description**: The violin plot is a rotated and mirrored representation of the density plot, with the width of each "violin" corresponding to the approximate frequency of data points in each region. The central white dot represents the median value, while the thick black bar in the center represents the interquartile range. The thin black lines extending from the bar represent the upper and lower adjacent values in the data.
- **Function**: The violin plot is used to visualize the distribution, patterns, and ranges of numeric data for one or more groups.
- **Advantages**: Unlike traditional box plots, violin plots provide a deeper understanding of the distribution of the data, allowing for the visualization of nuances that may not be perceptible in a box plot. They are particularly useful for identifying multimodal distributions and variations in the data.
- **Overlay with Additional Chart Type**: Violin plots are often rendered with another overlaid chart type, commonly a box plot, to facilitate precise comparisons of density curves between groups.