# Statistics Part 3

# Types of Graphs and to Use Them

Graphs help visualize complex data and come in different types, each suited for specific kinds of information representation.

**Bar Graph:** Show the number in categories

**Circle Graph:** Compare parts of the data to the whole

**Double Bar Graph:** Compare two or more sets of data

**Box Whiskers Plot:** Show measures of variation

**Histogram:** Show frequency of data divided into intervals

**Line Graph:** Show change over time

**Line Plot:** Show frequency data on a number line

# Histogram

A **histogram** is a graphical representation of data distribution. It consists of **bars** where each bar represents a range of values and the height of the bar indicates the frequency of data points within that range. Histograms display continuous data, making them useful for analyzing patterns such as skewness, central tendency, and variability.

**How to Make a Histogram:**

1. **Collect Data** – Gather numerical data that you want to analyze.
2. **Determine Bins** – Divide the data range into equal intervals.
3. **Count Frequencies** – Calculate how many data points fall into each bin.
4. **Draw Axes** – The x-axis represents the data intervals, and the y-axis represents frequency.
5. **Plot Bars** – Draw bars for each bin, ensuring their heights correspond to the frequency.

# Stem-and-Leaf Plot

A **stem-and-leaf plot** is a method of organizing numerical data to display its distribution while preserving the original values. It separates each data point into a **stem** (the leading digits) and a **leaf** (the last digit), making it useful for quickly identifying patterns, central tendency, and outliers. Stem-and-leaf plots retain the exact data points, providing a more detailed view of the dataset.

How to Make a Stem-and-Leaf Plot:

1. **Collect Data** – Gather and arrange numerical data in ascending order.
2. **Determine Stems** – Identify the leading digits (for 47, the stem is 4).
3. **Determine Leaves** – Identify the last digit (for 47, the leaf is 7).
4. **Organize Data** – Write stems in a column and place corresponding leaves beside them.

# Dot Plot

A **dot plot** is a simple way to display numerical data using dots to represent individual values. It is useful for small datasets and helps visualize **frequency, clusters, gaps, and outliers** in a distribution. Each dot represents a data point, and dots are stacked above a number line to indicate how often a value appears.

**How to Make a Dot Plot:**

1. **Collect Data** – Gather numerical data and organize it in ascending order.
2. **Draw a Number Line** – Choose a scale that covers the data range.
3. **Plot Dots** – Place a dot above each corresponding value on the number line. If a value repeats, stack dots vertically.

# Pie Chart

A **pie chart** is a circular graph used to represent **proportions** of a whole. Each **slice** corresponds to a category, with its size indicating the relative frequency or percentage of that category within the dataset. Pie charts are useful for visualizing **part-to-whole relationships**.

**How to Make a Pie Chart:**

1. **Collect Data** – Identify categories and their corresponding values or percentages.
2. **Calculate Percentages** – Convert each category's value into a percentage of the total.
3. **Draw a Circle** – Represent the whole dataset as a circle.
4. **Divide the Circle** – Create slices where each angle is proportional to its percentage

# Bar Chart

A **bar chart** is a graphical representation that uses **rectangular bars** to compare different categories or groups. The length or height of each bar corresponds to the value or frequency of the category it represents. Bar charts are ideal for visualizing differences among discrete items and are widely used in fields like business, education, and data analysis.

**How to Make a Bar Chart:**

1. **Collect Data** – Identify the categories and their corresponding numerical values.
2. **Choose a Scale** – Decide on a scale for the axis that will represent the values.
3. **Draw the Axes** – The x-axis typically represents the categories, while the y-axis shows the numerical values.
4. **Plot Bars** – Draw bars for each category, ensuring their height or length accurately reflects the data values.

# Heatmap & Violin Plot

A **heatmap** is a data visualization technique that uses **color gradients** to represent numerical values in a matrix or grid format. It is commonly used to analyze patterns, correlations, or density in large datasets. Darker or more intense colors typically indicate higher values, while lighter colors represent lower values.

**How to Make a Heatmap**:

1. **Collect Data** – Use a dataset with numerical values arranged in a matrix format.
2. **Define Color Scale** – Choose a color gradient where intensity reflects value magnitude.
3. **Plot the Grid** – Represent each data point as a colored cell in a grid.

A **violin plot** is a combination of a **box plot** and a **density plot**, used to visualize the **distribution, spread, and probability density** of a dataset. It is especially useful for comparing multiple distributions and identifying skewness or multimodal patterns.

1. **How to Make a Violin Plot:**
2. **Collect Data** – Gather numerical data for one or more categories.
3. **Calculate Distribution** – Determine the density of the data across its range.
4. **Create a Symmetric Shape** – Reflect the density on both sides of a vertical axis.

# Symmetry & Skewness

**Symmetry**: A dataset is symmetric if its left and right sides are mirror images when split at the center. The **mean and median** are nearly equal
**Skewness**: Measures data asymmetry.

- **Positive Skew (Right-Skewed)**: Tail extends right; mean is greater than the median**.**
- **Negative Skew (Left-Skewed)**: Tail extends left; mean is less than the median