Types of Graphs and When to Use Them

1. Histogram

Definition:

A histogram is a graphical representation of the distribution of numerical data. It groups data into bins (intervals) and displays the frequency of each bin as bars.

When to Use:

- To show the distribution of continuous data.
- To visualize frequency of occurrences in a dataset.
- To detect patterns such as skewness or normality.

Example:

---

2. Stem and Leaf Plot

Definition:

A stem-and-leaf plot is a method of displaying numerical data where each number is split into a "stem" (initial digit) and a "leaf" (remaining digits).

When to Use:

- To show the shape of data distribution while keeping the actual data values.
- To quickly identify the median, mode, and range.

Example:

```
Stem | Leaf
  3  | 5 8
  4  | 2 7 9
  5  | 1 3 6 7
```

---

3. Box and Whisker Plot (Box Plot)

Definition:

A box plot is a way of summarizing a dataset using five summary statistics: minimum, first quartile (Q1), median, third quartile (Q3), and maximum.

When to Use:

- To show the spread and skewness of data.

- To compare distributions across different groups.

Example:

```
       |-----|=======|-----|
  Min    Q1    Median   Q3   Max
```

---

4. Dot Plot

Definition:

A dot plot displays each data point as a dot along an axis, showing the frequency of values.

When to Use:

- To show small datasets.
- To compare distributions of different categories.

Example:

```
Data Value:  1  2  3  4  5
Frequency:   ●  ●●  ●●●  ●●  ●
```

---

5. Pie Chart

Definition:

A pie chart represents categorical data as slices of a circle, showing proportions.

When to Use:

- To show percentage or proportion of categories.
- To compare parts of a whole.

Example:

(Each slice represents a category proportionally.)

---

6. Bar Chart

Definition:

A bar chart represents categorical data with rectangular bars, where the length of each bar corresponds to the frequency or value of the category.

When to Use:

- To compare values between different categories.
- To show trends over time when bars are grouped.

Example:

```
Category  |  ████
Category  |  ███████
Category  |  ██████
```

---

7. Symmetry and Skewness

Definition:

Skewness measures the asymmetry of a distribution:

- **Symmetric**: Data is evenly distributed around the center.
- **Right-skewed (positive skew)**: A longer tail on the right side.
- **Left-skewed (negative skew)**: A longer tail on the left side.

Example:

```
Symmetric:     📊
Right-Skewed: 📊→
Left-Skewed:  ←📊
```

8. Heatmap

Definition:

A heatmap is a graphical representation of data where values in a matrix are represented as colors. It uses color gradients to show variations in data intensity.

When to Use:

- To visualize relationships between two variables (e.g., correlation matrices).
- To display patterns in large datasets (e.g., user activity over days and hours).
- To compare values across categories using color intensity.

Example:

Key Notes:

- Uses color gradients (e.g., red for high values, blue for low values).
- Ideal for matrices (e.g., gene expression data, website click rates).
- **Heatmap:** Analyzing user engagement on a website by hour/day.

---

9. Violin Plot

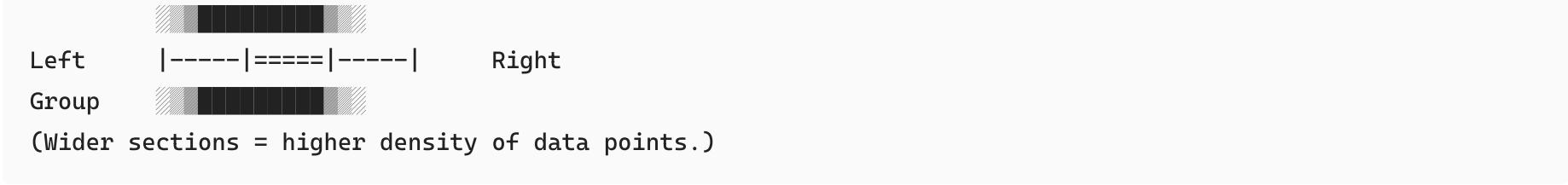
Definition:

A violin plot combines features of a box plot and a kernel density plot. It shows the distribution of numerical data, including the median, quartiles, and density.

When to Use:

- To visualize the distribution and probability density of data.
- To compare distributions across multiple groups/categories.
- To identify bimodality or outliers.

Example:



Key Notes:

- Shows the full distribution of data (unlike box plots, which only show summary statistics).
- Width indicates data density (thicker = more data points).
- **Violin Plot:** Comparing exam scores between different student groups.
