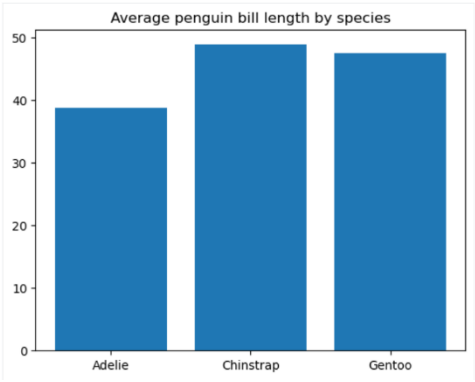
Common Types of Data Plots

There are Four main categories in Data Visualization:

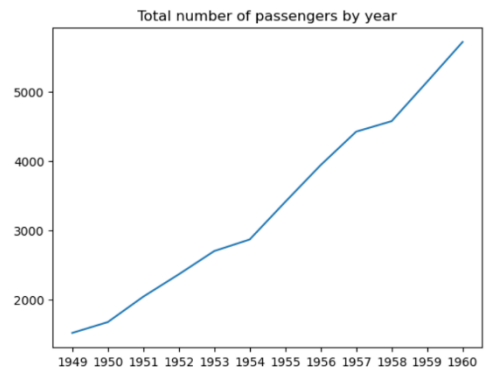
- ✓ Relationships:
    1. Scatter plot
    2. Bubble chart
- ✓ Comparisons:
    1. Bar chart
    2. Line chart
- ✓ Compositions:
    1. Pie chart
    2. Stacked bar chart
    3. Stacked area chart
- ✓ Distributions:
    1. Histogram
    2. Box-plot

Common Types of Data Plots:



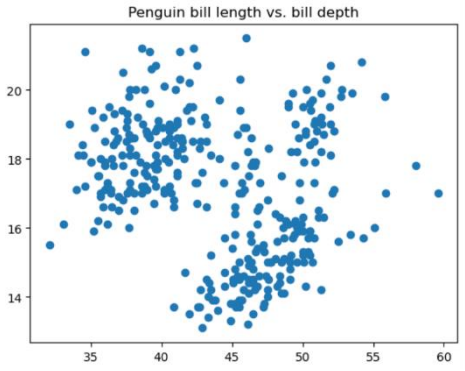
1. Bar chart:

- A bar chart is the most common data visualization for *displaying the numerical values of categorical data to compare various categories between them*.
- The categories are represented by rectangular bars of the same width and with heights (for vertical bar charts) or lengths (for horizontal bar charts) proportional to the numerical values that they correspond to.



2. Line plot:

- A line plot is a type of data chart that shows a progression of a variable from left to right along the x-axis through data points connected by straight line segments.
- Most typically, the change of a variable is plotted over time.

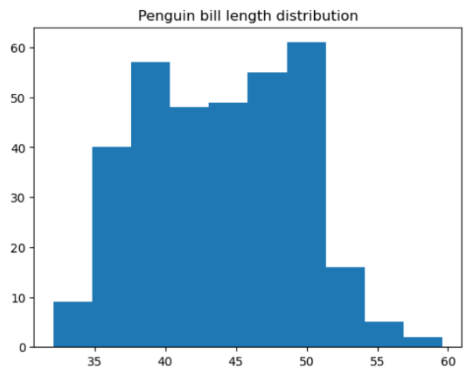


3. Scatter plot:

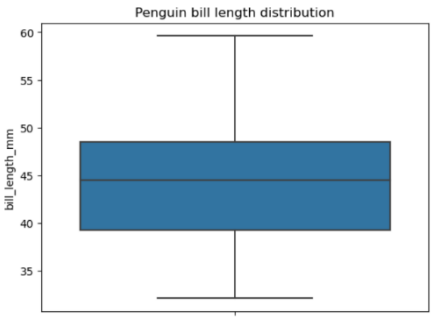
- A scatter plot is a data visualization type that displays the relationships between two variables plotted as data points on the coordinate plane.
- This type of data plot is used to check if the two variables correlate among themselves, how strong this correlation is, and if there are distinct clusters in the data.

4. Histogram:

- A histogram is a type of data plot that represents the frequency distribution of the values of a numerical variable.
- Under the hood, it splits the data into value range groups called bins, counts the number of points related to each bin, and displays each bin as a vertical bar, with the height proportional to the count value for that bin.
- A histogram can be considered as a specific type of bar charts, only that its adjacent bars are attached without gaps, given the continuous nature of bins.

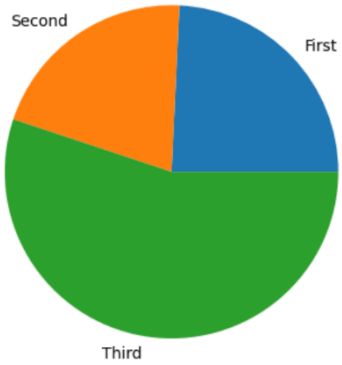


5. Box plot:



- A box plot is a data plot type that shows a set of five descriptive statistics of the data: the minimum and maximum values (excluding the outliers), the median, and the first and third quartiles. Optionally, it can also show the mean value.
- A box plot is the right choice if you're interested only in these statistics, without digging into the real underlying data distribution.

6. Pie chart:

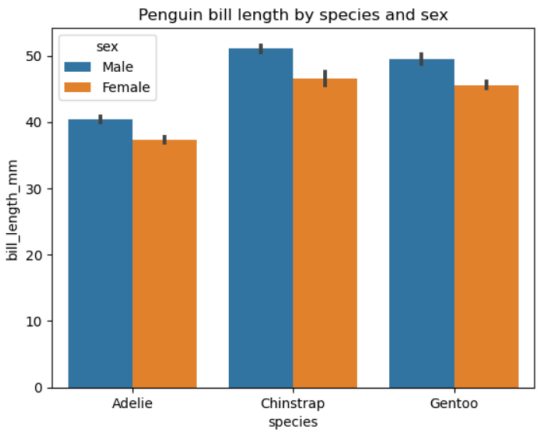


- A pie chart is a type of data visualization represented by a circle divided into sectors, where each sector corresponds to a certain category of the categorical data, and the angle of each sector reflects the proportion of that category as a part of the whole.
- Unlike bar charts, pie charts are supposed to depict the categories that constitute the whole, e.g., passengers of a ship.
  - Angles are more difficult to interpret for the human eye than lengths and often can be misleading.
  - They are less efficient in the case of five or more categories.
  - They can't display more than one set of categorical data. In other words, they can't be grouped, unlike bar charts.
  - They don't easily reveal real values.

Advanced Types of Data Plots:

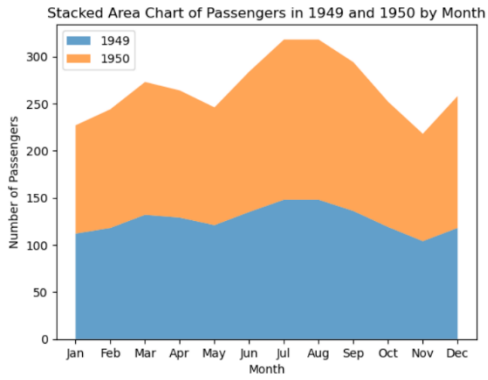
1. Grouped bar chart:

- While a common bar chart is used for displaying the numerical values of a categorical variable by category, a grouped bar chart serves the same purpose but across two categorical variables.
- Graphically, it means that we have several groups of bars, with each group related to a certain category of one variable and each bar of those groups related to a certain category of the second variable.
- Grouped bar charts work best when the second variable has no more than three categories. In the opposite case, they become too crowded and hence less helpful.



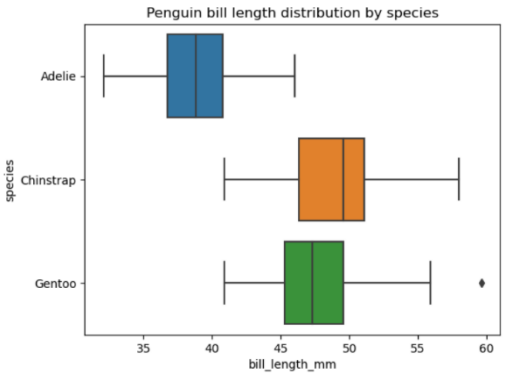
2. Stacked area chart:

- A stacked area chart is an extension of a common area chart (which is simply a line plot with the area below the line colored or filled with a pattern) with multiple areas, each corresponding to a particular variable, stacked on top of each other.
- Such charts are useful when we need to track both the overall progress of a set of variables and the individual contribution of each variable to this progress.
- Like line plots, stacked area charts usually reflect the change of variables over time.



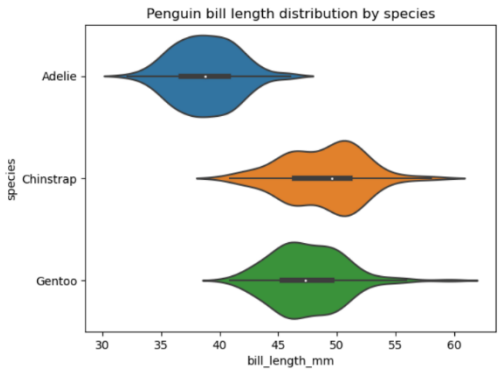
3. Multiple box plot:

- In the section on Common Types of Data Plots, we defined a box pot as a data visualization type that shows a set of five descriptive statistics of the data. Sometimes, we may want to display and compare these statistics separately for each category of a categorical variable.



4. Violin plot:

- A violin plot is similar to a box plot and displays the same overall statistics of the data, except that it also displays the distribution shape for that data.
- Like with box plots, we can create a single violin plot for the data in interest or, more often, multiple violin plots, each for a separate category of a categorical variable.



5. Heatmap:

- A heatmap is a table-style data visualization type where each numeric data point is depicted based on a selected color scale and according to the data point's magnitude within the dataset. The main idea behind these plots is to illustrate potential hot and cold spots of the data that may require special attention.
- In many cases, the data needs some preprocessing before creating a heatmap for them. This usually implies data cleaning and normalization.