**Generic Pipeline of Data Visualisation**

The generic pipeline of data visualisation typically includes five key stages: **Gathering, Processing, Analysing, Presenting, and Preserving**. Below is a detailed description of each stage, along with examples of activities and tools used at each step.

**1. Gathering**

**Description**: This stage involves collecting raw data from various sources. The data can come from surveys, sensors, databases, APIs, or web scraping.

* **Activities**:
  + Collecting survey responses.
  + Extracting data from APIs (e.g., REST APIs).
  + Scraping data from websites.
  + Importing datasets from CSV files or databases.
* **Example Tools**:
  + Python libraries like requests or BeautifulSoup for web scraping.
  + APIs such as Twitter API for social media data.
  + SQL for querying relational databases.

**2. Processing**

**Description**: In this stage, the raw data is cleaned and transformed into a usable format. This includes handling missing values, removing duplicates, and standardizing formats.

* **Activities**:
  + Cleaning erroneous or incomplete data entries.
  + Normalizing or standardizing numerical values.
  + Converting unstructured data (e.g., JSON) into tabular formats.
  + Merging datasets from multiple sources.
* **Example Tools**:
  + OpenRefine for cleaning and transforming data.
  + Python libraries like pandas for data manipulation.
  + ETL (Extract, Transform, Load) tools like Talend or Apache Nifi.

**3. Analysing**

**Description**: This stage involves applying statistical methods or algorithms to extract insights and patterns from the processed data.

* **Activities**:
  + Performing descriptive statistics (e.g., mean, median).
  + Conducting regression analysis or clustering.
  + Building predictive models using machine learning algorithms.
  + Identifying trends or correlations in the dataset.
* **Example Tools**:
  + Python libraries like scikit-learn or statsmodels for analysis.
  + R for statistical computing and visualisation.
  + Tableau Prep for exploratory analysis.

**4. Presenting**

**Description**: At this stage, insights are communicated effectively through visualisations or reports. The focus is on clarity and audience comprehension.

* **Activities**:
  + Creating charts (e.g., bar charts, scatter plots) to represent findings.
  + Designing dashboards for interactive exploration of results.
  + Writing reports summarizing key insights.
* **Example Tools**:
  + Tableau or Power BI for creating dashboards.
  + Matplotlib and Seaborn in Python for static visualisations.
  + D3.js for custom web-based interactive visualisations.

**5. Preserving**

**Description**: This final stage ensures that the processed and analysed data, along with its insights, are stored securely for future use or reference.

* **Activities**:
  + Archiving cleaned datasets in databases or cloud storage.
  + Documenting metadata to describe the dataset’s structure and context.
  + Implementing secure access controls to protect sensitive information.
* **Example Tools**:
  + Amazon S3 or Google Cloud Storage for archiving datasets.
  + PostgreSQL or MongoDB for storing structured/unstructured data.
  + GitHub repositories for version control of scripts and documentation.

Summary Table

| **Stage** | **Activities** | **Example Tools** |
| --- | --- | --- |
| Gathering | Collecting raw data | REST APIs, SQL, Python (requests) |
| Processing | Cleaning and transforming data | OpenRefine, pandas, Talend |
| Analysing | Statistical analysis and pattern detection | scikit-learn, R, Tableau Prep |
| Presenting | Creating visualisations/reports | Tableau, Power BI, Matplotlib |
| Preserving | Archiving datasets and metadata | Amazon S3, PostgreSQL |

This pipeline ensures a structured approach to managing data throughout its lifecycle while enabling effective communication of insights.