The three stages of a reproducible workflow are:

Data collection

Data Cleaning

Data Analysis

**Data collection :**

the process of gathering, recording, and organizing data from various sources. This step is crucial because the quality of the data and the accuracy of its representation can impact the outcomes of the subsequent data analysis.

Data collection can involve various methods such as surveys, experiments, observations, and web scraping. In addition, data can be collected from various sources such as databases, spreadsheets, social media platforms, and sensors.

**Data Cleaning:**

Data cleaning is the process of identifying and correcting or removing inaccurate, incomplete, or irrelevant data from a dataset. It is an essential step in data analysis because the quality of the data can have a significant impact on the results of the analysis.

Data cleaning involves several tasks, including: Removing duplicate data, Handling missing data, Correcting inconsistencies, Handling outliers, Standardizing data

Overall, data cleaning ensures that the data is accurate, complete, and consistent, which leads to more reliable and trustworthy results in subsequent data analysis.

**Data Analysis and Modeling:**

Data analysis is the process of inspecting, cleaning, transforming, and modeling data in order to extract useful information that can be used for making decisions, identifying patterns, and discovering insights. It involves using various analytical and statistical techniques to organize, understand, and interpret data.

The first visualization is a histogram that shows the distribution of grip strength in the given dataset. The x-axis represents the grip strength values and the y-axis represents the frequency of those values. The histogram is divided into 10 bins, each representing a range of grip strength values. This visualization helps to understand the spread of grip strength values in the dataset and the most common grip strength values.

The second visualization is a box plot that shows the relationship between age and grip strength. The x-axis represents the age of the subjects in the dataset, and the y-axis represents the grip strength values. Each box in the plot represents the distribution of grip strength values for a particular age group. The box shows the median grip strength value for that age group, and the vertical lines extending from the box show the range of grip strength values. The visualization helps to identify any trends or patterns in the grip strength values with respect to age.