The three stages of reproducible workflow are as follows:

1. Data input/collection
2. Data processing
3. Data analysis

For the given table, my workflow is as mentioned below:

**1.Data input/collection**

A screenshot of a computer

Description automatically generated

I made the given table into a **raw*.csv*** file and then load that file in a python environment using ***pandas*** library.

Text

Description automatically generated

**2. Data processing**

Here the predictors are converted into numeric data type. Since ***Frailty*** is the target variable, we do not need to convert its data type and let it remain as it is.

Text

Description automatically generated

After processing, I save the processed dataset in processed.csv file under processed\_data folder.



**3. Data analysis**

Graphical user interface, text, application

Description automatically generated

Chart, box and whisker chart

Description automatically generated

This visualization represents the relationship of ***height, weight*** and ***grip strength*** with ***age*** of each observation.

Text

Description automatically generated

The folder structure looks like this after ***data analysis*** stage.