Three stages of reproducibility of given data

Collect the data from the given document.

Data Collection and Input

Stage 1

Data Processing

Data Analysis

Stage 2

Stage 3

Store the data in csv file.

Pass this file as input for data processing.

Clean the data by script.

1. Strip the trailing and leading whitespaces in column names.

2. check for the null values if any

3. Rename the column names

Analyse the cleaned data by script.

1. Visualize the grip strength and frailty for given data using heatmap

2. Analyze the resulted graphs and document the results

**Folder Structure in each stage (**Folder Structure has been implemented in Question\_1\_Folders**)**

We create a folder Question\_1\_Folders and in that, we add subfolders data\_raw, data\_clean, results, and src.

**data\_raw** folder contains all the data files and metadata that we have collected during data collection i.e in stage 1.

**data\_clean** folder contains all the data files and metadata that have been cleaned and processed in stage 2 which will be produced by script.

**results** folder contains documentation files that have been generated as a result of the analysis done by the analysis script.

**src** foldercontains scripts that are written to process, clean, analyze, visualize the data.

**Stage 1**

Question\_1\_folders

|— data\_raw

|— raw\_frailty\_data.csv

|—readme.txt (this has metadata)

|— data\_clean

|— results

|— src

**Stage 2**

Question\_1\_folders

|— data\_raw

|— raw\_frailty\_data.csv

|—readme.txt (this has metadata)

|— data\_clean

|— cleaned\_frailty\_data.csv (this file is created by clean script in src folder)

|—readme.txt (this has metadata)

|— results

|— src

|— data\_cleansing.py

**Stage 3**

Question\_1\_folders

|— data\_raw

|— raw\_frailty\_data.csv

|—readme.txt (this has metadata)

|— data\_clean

|— cleaned\_frailty\_data .csv (this file is generated by data\_clean script in src folder)

|—readme.txt (this has metadata)

|— results

|—heatmap.png

|—Question\_1\_reproducibility steps

|— src

|—data\_cleansing.py

|—data\_analysis.py

By analysis of the given data, this a correlation heatmap showing the relationships between the numerical variables (Height, Weight, Age, Grip\_Strength). The colour intensity and annotations within each cell of the heatmap represent the strength and direction of the correlation between the variables. Positive correlations are shown in warmer colours, while negative correlations are shown in cooler colours.

A screenshot of a graph

Description automatically generated