**Assignment 1**

Name: Dennish Gade

ID: 16335329

1) Based on the following table, design the three stages of reproducible workflow, includes the work you can do and the folder structure in each stage (reference study case in chapter 3).  (5 points)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Height (Inches) | Weight (Pounds) | Age | Grip strength | Frailty |
| 65.8 | 112 | 30 | 30 | N |
| 71.5 | 136 | 19 | 31 | N |
| 69.4 | 153 | 45 | 29 | N |
| 68.2 | 142 | 22 | 28 | Y |
| 67.8 | 144 | 29 | 24 | Y |
| 68.7 | 123 | 50 | 26 | N |
| 69.8 | 141 | 51 | 22 | Y |
| 70.1 | 136 | 23 | 20 | Y |
| 67.9 | 112 | 17 | 19 | N |
| 66.8 | 120 | 39 | 31 | N |

**Stage 1: Data cleaning**

**Work to be done:**

* Check for missing data.
* Check for outliers and anomalies.
* Remove duplicates.
* Standardize variable names and units.

**Folder structure:**

* **Raw data:** contains the original dataset.
* **Cleaned data** contains the cleaned dataset
* **Scripts:** contains the scripts used for data cleaning, such as "clean\_data.R" or "data\_cleaning.ipynb"
* **Documentation:** contains any documentation related to data cleaning, such as a data dictionary or a codebook

**Stage 2: Analysis**

**Work to be done:**

* Exploratory data analysis (EDA)
* Statistical modeling
* Hypothesis testing
* Visualization

**Folder structure:**

* **Cleaned data** contains the cleaned dataset
* **Results:** contains the output of the analysis, such as tables, figures, and statistical summaries
* **Scripts:** contains the scripts used for analysis, such as "eda.R" or "modeling.ipynb"
* **Documentation:** contains any documentation related to the analysis, such as a methods section or a statistical analysis plan

**Stage 3: Reporting**

**Work to be done:**

* Write up results.
* Create final figures and tables.
* Format document

**Folder structure:**

* **Results:** contains the final output of the analysis, such as tables, figures, and statistical summaries
* **Scripts:** contains the scripts used for generating the final output, such as "create\_figures.R" or "format\_document.ipynb"
* **Documentation:** contains any documentation related to the reporting, such as a results section or a final report. This could also include any supporting files, such as figures or tables used in the report.

Note: the specific file organization will depend on the project requirements and tools used.

Top of Form

Bottom of Form