Exercise 1: Clean Data – Task Outline

Data1.txt is file comprising a table with 6 columns (sample ID, sex, age, height, BMI, and a genotype of a SNP) and 100 rows of samples. With some NA values present. This ‘Dummy’ dataset was cleaned by removing samples with NA values, a clean table was the outputted, in addition to an output table of the removed samples and a calculation of the percentage of samples that remained. There were some NA values in the ‘dummy’ data, which meant some samples failed for ‘some’ measurement due to low quality.