NetID: au198
Aizhan Uteubayeva

Analysis Logic
Homework 4: Differential Gene Expression (t-test)
HIDS-7003-01

Overview:
- Clinical data: 40 patients // rows. Confirmed de-identification
- Gene expression data in log2 scale: includes gene annotation, has 80 rows that need to be transposed

Goal:

The goal is to compare 20 patients that relapsed vs. 20 patients that have the recurrence of colorectal cancer.

Clinical data:
- 40 rows that correspond to 40 patients
- The target attribute is the "RECURRENCE_ANY" column
- The subset of clinical data should consist of 20 patients, and the patient IDs are unique

Gene expression file:
- Genes are not in rows, the file needs to be transposed
- There are now 80 rows// patients, that will be columns after transposition
- The IDs do not match clinical data
- The gene expression file will be used to subset the patients of interest

Sanity check:

The data needs to match, be a numeric matrix and have no extra rows// junk.