Homework assignment: Multiplet sequencing datasets

Attached in the folder are two files

1. Count\_matrix.csv
2. Peak\_names\_out.csv

In this dataset, human (hg19) and mount (mm10) cells were mixed together and a single cell sequencing experiment was run. The goal of this experiment was to determine how often droplets are loaded with more than one cell. Each barcode, or column, in the “count matrix” represents a droplet which was loaded with cells. Each row, whose names are denoted by “peak\_names\_out.csv”, are genes identified by RNA sequencing.

Homework questions:

1. Summarize this dataset. How many droplets can you identify that are loaded with more than one cell? Distinguish between Human/Human, Mouse/Human, or Mouse/Mouse multiplets.
2. Write an R script which takes the two files attached as arguments and returns summary plots as a pdf file and a csv report for the number of droplets with 1 or 2+ cells identified.
3. Provide proper documentation, consider common user errors and provide meaningful error messages in return.