**Assignment objectives:**

1. Parse a set of files (column tab-delimited text files) in a directory.
2. Summarize column 8 to obtain –

Number of rows containing ‘intergenic’

Number of rows containing ‘intron’ and so on…

Output format –

File1 File2 File3 File4 File5 (File name)

non-coding N1 N2 N3 N4 N5 etc

Intergenic M1 M2 M3 M4 M5 etc

intron

exon

promoter-TSS

TTS

5’ UTR

3’ UTR

1. For a given gene name (column 16) supplied as an argument,

Find lines in these files containing this gene name and print the same.

Output format –

Filename Rows containing the gene name

File1 Rows

File2 Rows etc

Hint: Dictionary data structure might be a valuable tool here.