Eric Paulz

CPSC 3220-001

**Project 1**

1. Please provide command-line code and full-sentence English interpretation of the results for the following:
   1. Identify 4 image file formats and 4 source code formats present within fs1.zip and fs2.zip.
      1. *fs1:* **grep -E '\*\.jpg$|\*\.JPG$' tape\* | wc -l**
         1. grep attempts to match the pattern given by the regular expression. -E allows me to extend the regular expression to look for many patterns instead of just one. Grep’s target, tape\*, tells it to search through all tapes in the file.
         2. wc -l prints the number of lines of output, i.e. the number of files matching the pattern.
         3. Files:
            1. Image: .jpg, .ppm, .gif, .tiff
            2. Source: .c, .cpp, .html, .java
      2. *fs2:* **grep -E '\*\.jpg$|\*\.JPG$' ls-redaction.txt | wc -l**
         1. Same command and explanation as from fs1, except grep’s target is now ls-redaction.txt instead of tape\*.
         2. Files:
            1. Image: .jpg, .ppm, .gif, .tiff
            2. Source: .c, .cpp, .html, .java

Note\* - I matched both lowercase and uppercase file extensions because in Linux they are considered the same, so I didn’t want them to be counted as different by grep.

* 1. How many files appear cumulatively? How many from 1990-1995? How many from pre-1990?
     1. *fs1:*

2666(.jpg)+72(.ppm)+13722(.gif)+1646(.tiff)+15214(.c)+168(.cpp)+3456(.html)+353(.java)

= 37297 total files.

**grep -E ‘regex for all extensions’ tape\* | awk ‘$7 >= 1990 && $7 <= 1995’ | wc -l**

Piping to awk grabs the 7th column (year) and only returns rows where the year is within the specified range. 36928 files from 1990-1995 and 368 files from pre-1990.

* + 1. *fs2:*