**1. Without using a text editor examine the contents of the file HapMap.fas.txt**

###- How many lines does this file contain?

command: wc –l HapMap.fas.txt

output: 378672 lines

###- How many characters? (Hint: check out the options of wc)

command: wc –c HapMap.fas.txt

output: 15503584 characters

###- What is the first line of this file? (Hint: read the man page of head)

command: head HapMap.fas.txt

output: >TP1\_query\_64

CGAAAAAAAAAAAAAAAAAAAAAGACGCTACACAATGTCTACTGCTGAGTAATTGACAGTTTCT

###- What are the last 3 lines? (Hint: read the man page of tail)

command: tail HapMap.fas.txt

output: >TP95868\_hit\_64

CGATTTTTTTTTTTCGTTTGTTTGTTTGTTTGTTTTTTTTTTTTTTTCGTAAACCTACTTAGGC

>TP95869\_query\_64

CGATTTTTTTTTTTTATCAAAAATGTTATTTTTCTTATTTTGCCGCTGTCTTAAATGGCCCGCT

>TP95869\_hit\_64

CGATTTTTTTTTTTTATCAAAAATGTTATTTTTCTTATTTTGCCGCTGTCTTAAATGGCCCTCT

###- How many sequences are in the file? (Hint: use grep)

command: grep –c '[AGCT]$' HapMap.fas.txt

output: 189336

**2. Using the file, grep\_data.txt, write a series of grep statements that do the following:**

###Print all lines that contain a phone number with an extension (the letter x or X followed by four digits).

grep '[Xx]\d\d\d\d$' grep\_data.txt

###Print all lines that begin with three digits followed by a blank using the repetition specifier ({})

grep '[0-9][0-9][0-9]\s' grep\_data.txt

###Print all lines containing a vowel (a, e, i, o, or u) followed by a single character followed by the same vowel again. Thus, it will find “eve” or “adam” but not “vera”. Hint: \( and \).

grep '\([a][^aeiou][a]\|[e][^aeiou][e]\|[i][^aeiou][i]\|[o][^aeiou][o]\|[u][^aeiou][u]\)' grep\_data.txt

###Print all lines that do not begin with a capital S.

grep '^[^S]' grep\_data.txt

**Write grep statements that use command-line options along with the pattern to do the following:**

###Print all lines that contain CA in either uppercase or lowercase.

grep '[Cc][Aa]' grep\_data.txt

###Print all lines that contain an email address (they have an @ in them), preceded by the line number.

grep -n @ grep\_data.txt

###Print all lines that do not contain the word Sep. (including the period).

grep -v ^'Sep.' grep\_data.txt

###Print all lines that contain the word de as a whole word.

grep -w 'de' grep\_data.txt