Eric Randolph

MBE 635

Homework Week 1

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

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

Erics-MacBook-Air:week 2 school$ wc -l HapMap.fas.txt

378672 HapMap.fas.txt

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

Erics-MacBook-Air:week 2 school$ wc HapMap.fas.txt

378672 378672 15503584 HapMap.fas.txt

Line count. Word count. Character count.

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

Erics-MacBook-Air:week 2 school$ head -1 HapMap.fas.txt

>TP1\_query\_64

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

Erics-MacBook-Air:week 2 school$ tail -3 HapMap.fas.txt

CGATTTTTTTTTTTTATCAAAAATGTTATTTTTCTTATTTTGCCGCTGTCTTAAATGGCCCGCT

>TP95869\_hit\_64

CGATTTTTTTTTTTTATCAAAAATGTTATTTTTCTTATTTTGCCGCTGTCTTAAATGGCCCTCT

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

Erics-MacBook-Air:week 2 school$ grep -c ATGC\* HapMap.fas.txt

125432

**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).**

Erics-MacBook-Air:week 2 school$ grep [xX][0-9][0-9][0-9][0-9]$ grep\_data.txt

714-555-7870 X7310

714-555-7870 x7309

714-555-5350 x2134

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

Erics-Air:week 2 school$ ggrep -P ^[0-9]{3}'\s' grep\_data.txt

401 S. Palm St.

440 Briercrest Ave.

**###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: \( and \).**

Erics-Air:week 2 school$ ggrep -P '([aeiou]).\1' grep\_data.txt

Steve Marshal

Sonora High School

Carl Hohl (aka Krazy Rabbit)

Steve Lawson

Rosemead High School

Rosemead, CA 91770

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

Erics-Air:week 2 school$ ggrep -v ^S grep\_data.txt

Esperanza High School

1830 N. Kellog Dr.

Anaheim, CA 92807-1281

714-555-7870 X7310

aztecwrestling@example.com

Brian Fortenbaugh

714-555-7870 x7309

401 S. Palm St.

La Habra, CA 90631

Carl Hohl (aka Krazy Rabbit)

562-555-9800

Oct. 1, 2013

Lakewood High School

440 Briercrest Ave.

Lakewood, CA 90713-2013

Andy Miramontes

562-555-1281

Oct. 8, 2013

North Torrance High School

2013 W. 182nd

Torrance, CA 90504

Don Henderson

310-555-4412

Oct. 15, 2013

El Dorado High School

1651 N. Valencia Ave.

Placentia, CA 90631

714-555-5350 x2134

Lawsonhawk@example.com

Nov. 12, 2013

Rosemead High School

9063 E. Sepulveda Dr.

Rosemead, CA 91770

Daren de Heras

daren103@example.com

Erics-Air:week 2 school$

**###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.**

Erics-Air:week 2 school$ ggrep -i CA grep\_data.txt

Anaheim, CA 92807-1281

La Habra, CA 90631

Carl Hohl (aka Krazy Rabbit)

Lakewood, CA 90713-2013

Torrance, CA 90504

Placentia, CA 90631

Rosemead, CA 91770

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

Erics-Air:week 2 school$ ggrep -n @ grep\_data.txt

7:aztecwrestling@example.com

38:Lawsonhawk@example.com

45:daren103@example.com

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

Erics-Air:week 2 school$ ggrep -vi SEP. grep\_data.txt

Esperanza High School

1830 N. Kellog Dr.

Anaheim, CA 92807-1281

Steve Marshal

714-555-7870 X7310

aztecwrestling@example.com

Brian Fortenbaugh

714-555-7870 x7309

Sonora High School

401 S. Palm St.

La Habra, CA 90631

Carl Hohl (aka Krazy Rabbit)

562-555-9800

Oct. 1, 2013

Lakewood High School

440 Briercrest Ave.

Lakewood, CA 90713-2013

Andy Miramontes

562-555-1281

Oct. 8, 2013

North Torrance High School

2013 W. 182nd

Torrance, CA 90504

Don Henderson

310-555-4412

Oct. 15, 2013

El Dorado High School

1651 N. Valencia Ave.

Placentia, CA 90631

Steve Lawson

714-555-5350 x2134

Lawsonhawk@example.com

Nov. 12, 2013

Rosemead High School

Rosemead, CA 91770

Daren de Heras

daren103@example.com

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

Erics-Air:week 2 school$ ggrep -w de grep\_data.txt

Daren de Heras