450128

Homework 2

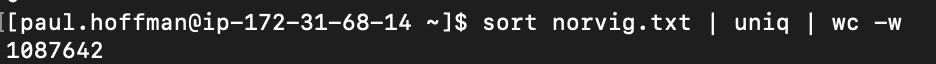
1- Download a text file from the following address:

http://3.236.139.48/dataset/norvig.txt

Text

Description automatically generated

2-  How many unique words are in the file? (number)



3-  Get a 10000 rows sample of this file. Then find what is the most frequent word. How frequent is it (how many times)? Note: You don’t need to do data cleaning!



“the” is the most frequent word

4-  Property changing:

1. Using one command, make a duplicate of norvog.txt and name it bigtext.txt (at the same location).
2. Change the attribute of norvig.txt to be readable and writable only by you



5- Among the codes you run up to this point, find those that you used the cat command.

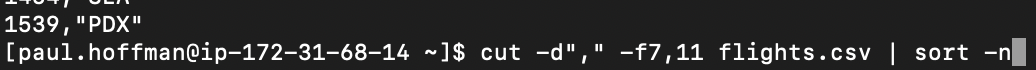
Text

Description automatically generated

Download flights.csv file from the following address and answer the following questions:

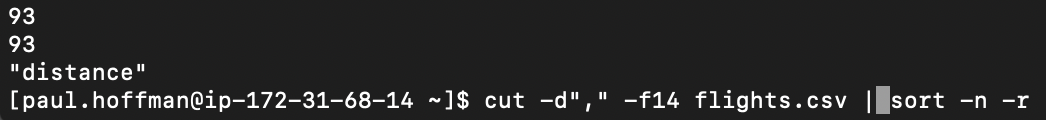
http://3.236.139.48/dataset/flights.csv

6- What is the worst arrival delay in this dataset? What is the origin airport name? You are supposed to write one line code (apparently you need to pipe the commands). Hint: Look at the sort --help and find what you need to add.



Origin airport is PDX

7-  Find the shortest distance of flights in this database. Then, find the airports.





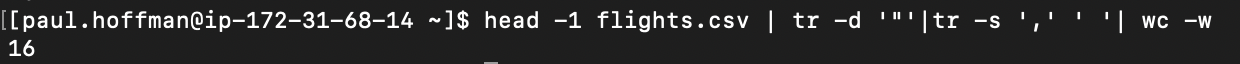
8-  What is the busiest month in this database? (The most frequent month)

Text

Description automatically generated

Month 8 is the busiest which is August.

9-  How many columns does this file have?



10-  Create a 50 random rows sample file from the last 12 columns (columns 5-16) and name it flights\_sample.csv. Hint: Look at the cut --help and find what you need to add. Remember that your sample file needs the header too.



