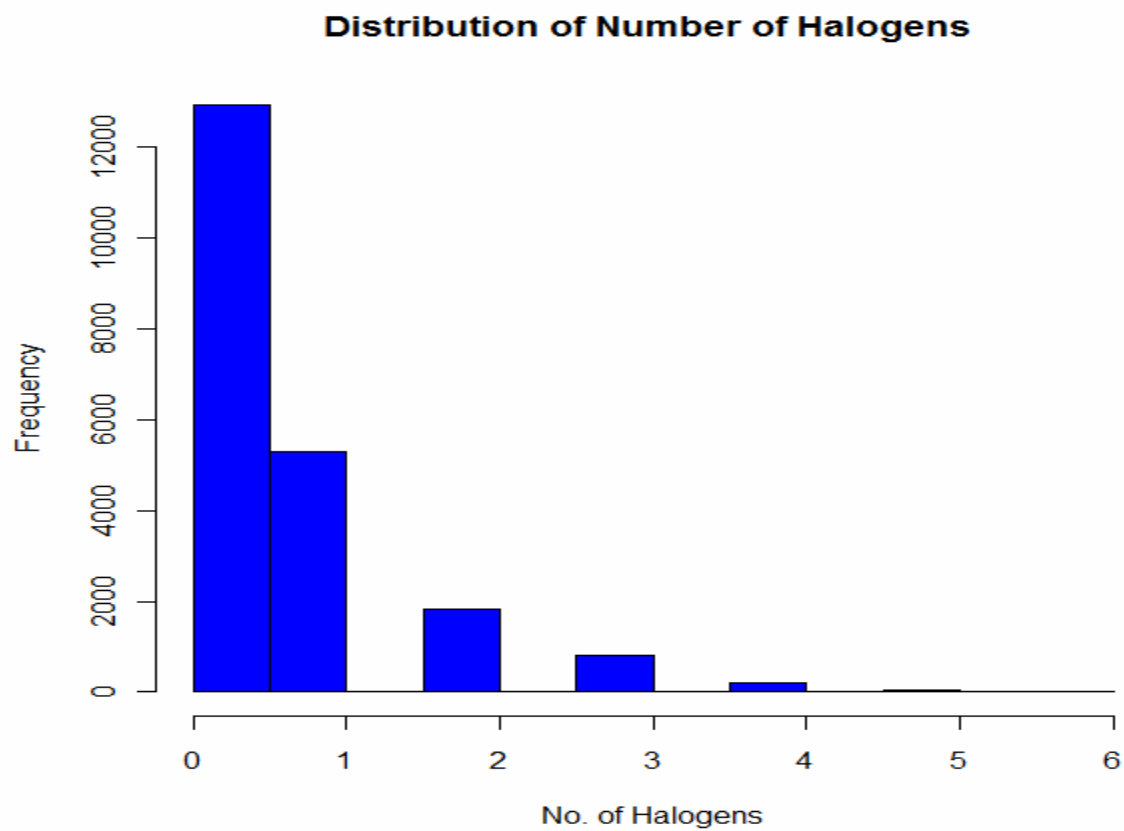
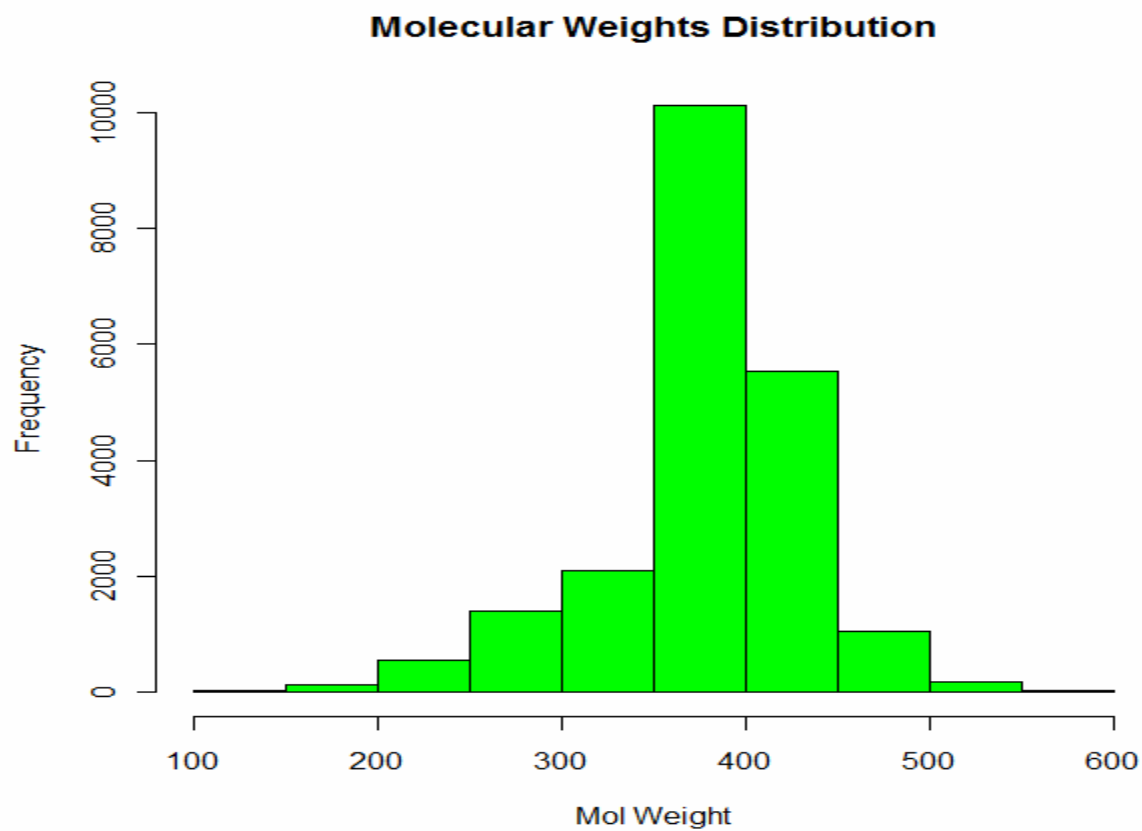
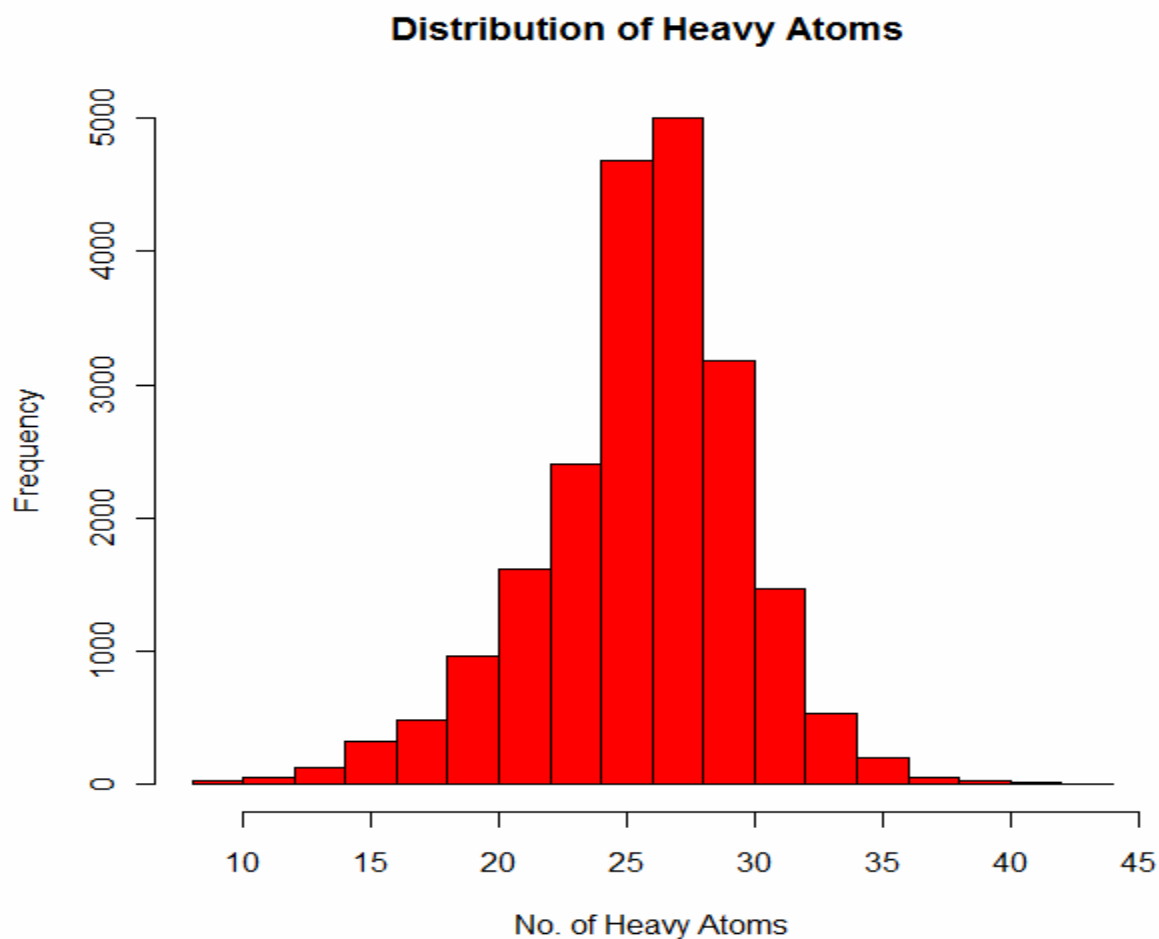


1) SDFFile Parsing Using OEChem toolkit:





1) To get molecular weight of molecules in SDF file **'python oechemmolwt.py inputSDFfile'**

Output file: molweight.txt (having the title and molweight for each molecule)

2) To get Halogens type **'python oechemhalogens.py inputSDFfile'**

Output file: halogens.txt

3) To get Heavy Atoms **'python oechemheavyatoms.py inputSDFfile'**

Output file: heavyatoms.txt

The above histograms are obtained by reading these text files into 'R' and using hist function to plot.

2) Parsing SD File in Python:

The Following program SDparser.py reads through the SD file line by line, looks for atoms in the Atoms Block of the SD file, looks for their corresponding exact masses as given in the exactmass.txt file, computes the Molecular weight of each molecule, also counts number of Carbons, Nitrogens, Oxygens.

To run the program on the cheminfo.informatics.indiana.edu server type:

```
python SDparser.py inputSDFfile
```

Outputs by taking in the compounds.sdf file provided.

1) Average Molecular weight = 375.449878678

2) Average time per molecule = 1.80863614327e-05

3) Average CPU time = 1.60885818388e-05

4) Total Number of Molecules in the file = 21133

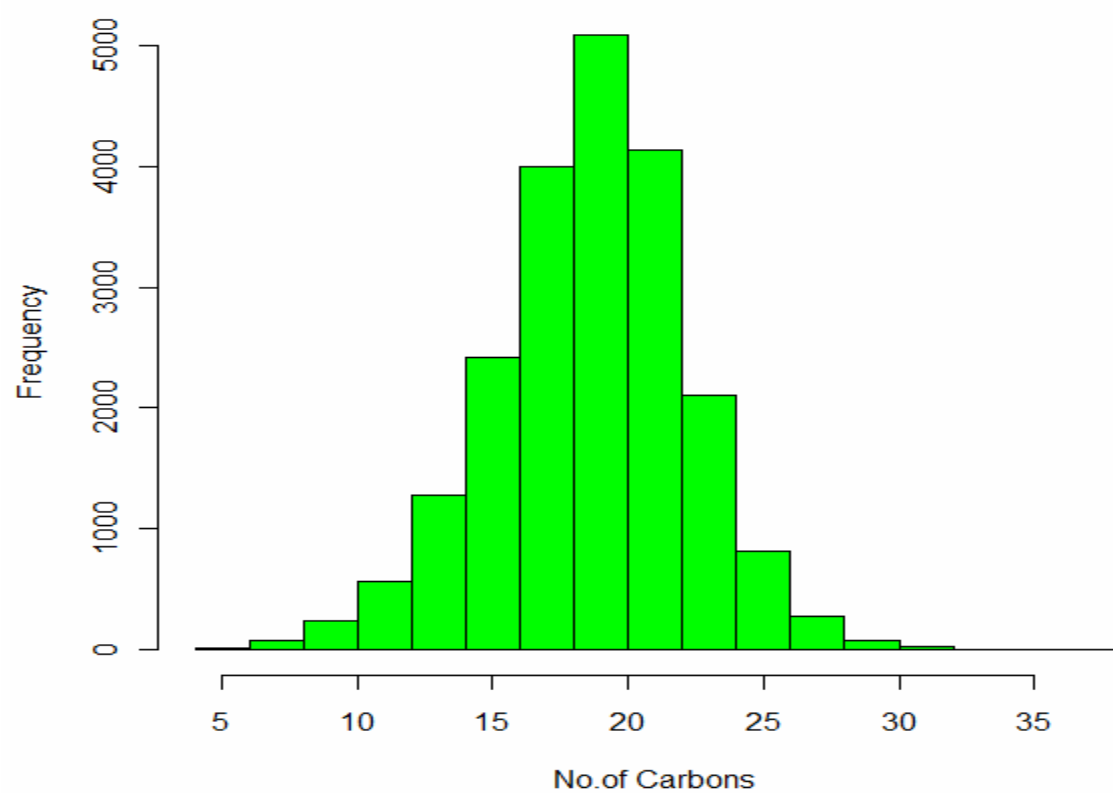
5) Machine Details (performed on cheminfo server)

CPU : 3000.362MHz

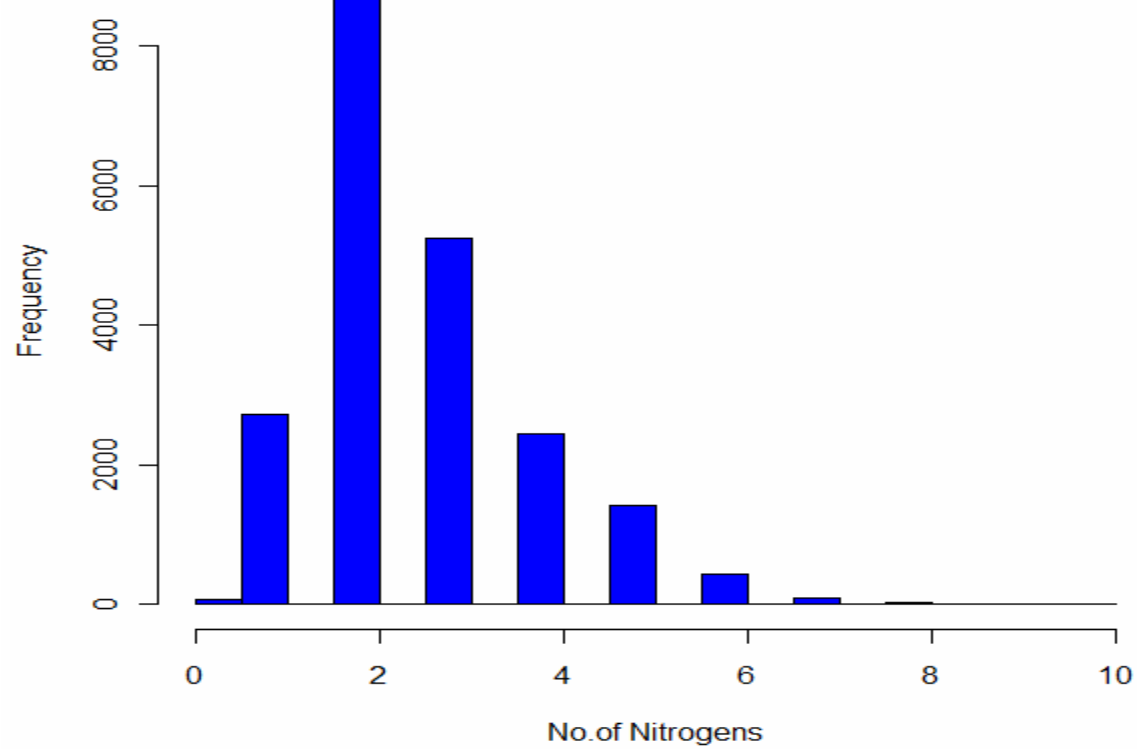
model name : Intel(R) Xeon(TM) CPU 3.00GHz

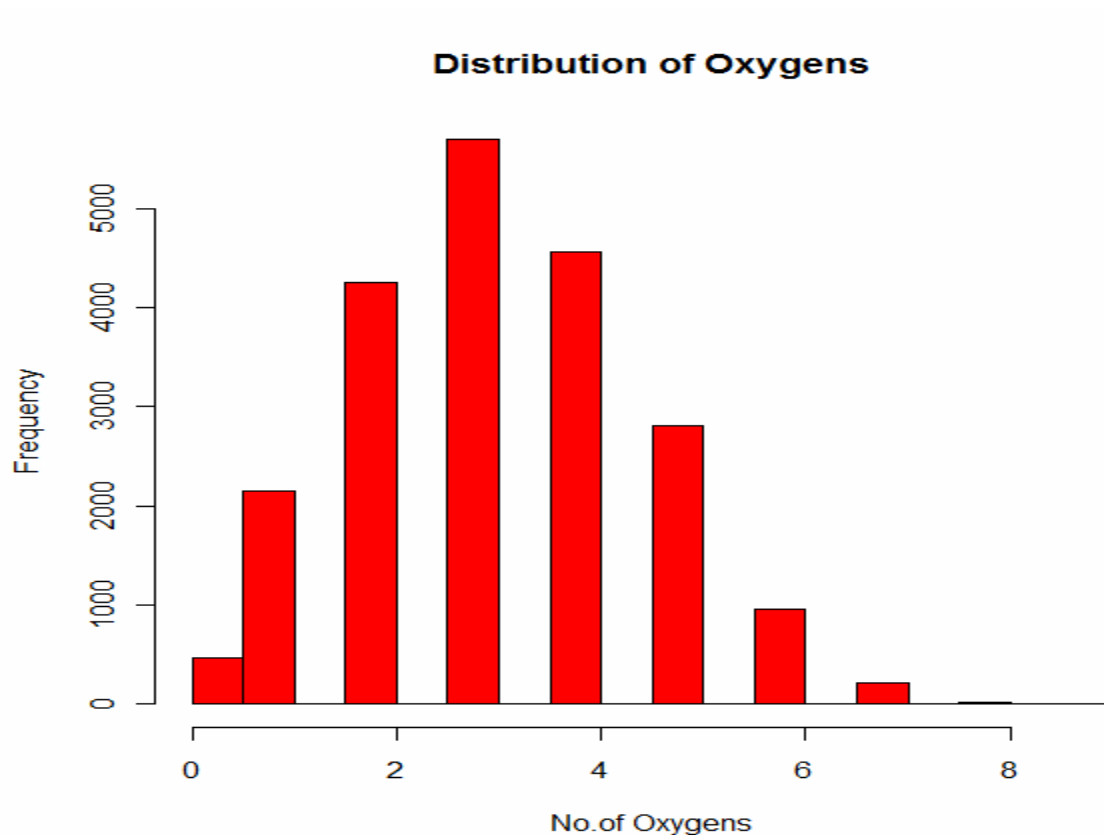
Disk space(from meminfo file):413412kB

Distribution of Carbons



Distribution of Nitrogens





The above plots are obtained by reading into R the Counts.txt file obtained as the output by the SDParser.py program. In the Counts.txt file first column has number of Carbons per molecule, second column has number of Nitrogens per molecule, and third column has number of oxygen atoms per molecule.