CS 210 Homework 2

1. `rosterSummary.py`: A student roster file `roster1.dat` has these fields:

   ```
   name,major,gpa,credits
   ```

   Write a file `roster1.out` with these new fields:

   ```
   major,avgGpa,avgCredits,count
   ```

   where `avgGpa` is the average GPA for students in that major, `avgCredits` is the average number of credits for students in that major, and `count` is the number of students in that major.

2. `wordSort.py`: Write a program that reads words from a file (filename given as a command-line argument) and prints them in (case insensitive) sorted order.

   For example, if the input file contains:

   ```
   If there's a problem yo I'll solve it
   Check out the hook while my DJ revolves it
   Ice ice baby
   ```

   your program should print (one word per line, compressed here for brevity):

   ```
   a baby Check DJ hook I'll Ice ice If it it my out
   problem revolves solve the there's while yo
   ```

3. `wordCount.py`: Write a program that reads words from a file (filename given as a command-line argument) and prints the number of words.

   Using the same input as problem 2, your program should print 20.

4. `wordCountDistinct.py`: Write a program that reads words from a file (filename given as a command-line argument) and prints the number of distinct words. Words that differ only in case should be considered to be equivalent.

   Using the same input as problem 2, your program should print 18.

5. `studentStats.py`: We can create a numpy array from separate lists by specifying a format for each field:

```
arr = np.zeros(4, dtype={'names': ('name', 'age', 'major', 'gpa'),
                         'formats': ('U50', 'i4', 'U4', 'f8')})

arr['name']  = ['Alice', 'Bob', 'Carol', 'Dennis']
arr['age']   = [21, 25, 18, 29]
arr['major'] = ['CS', 'Math', 'Chem', 'Phys']
arr['gpa']   = [3.8, 3.2, 4.0, 3.5]
```

where `U50` is a string of max length 50, `i4` is a 4-byte integer, `U4` is a string of max length 4, and `f8` is an 8-byte floating-point number. Then `arr` has value

```
array([('Alice', 21, 'CS', 3.8), ('Bob', 25, 'Math', 3.2),
  ('Carol', 18, 'Chem', 4. ), ('Dennis', 29, 'Phys', 3.5)],
  dtype=[('name', '<U50'), ('age', '<i4'), ('major', '<U4'), ('gpa', '<f8')])
```

Then, for example, `arr['name']` will give an array of just the student names.

Write a program that reads the file `roster2.dat` that has this format:

```
name,age,major,gpa
```

Convert this to a numpy array as shown above. Then use that array to compute and print each of the following:

(a) The average GPA of all students

(b) The maximum GPA of students majoring in CS

(c) The number of students with a GPA over 3.5

(d) The average GPA of students who are at least 25 years old

(e) The major that has the highest average GPA among students at most 22 years old

For example, for the four students listed, your program should print:

```
3.625
3.8
2
3.35
Chem
```