1. Discussion and algorithm

1.1 Handle IOError when opening the file. Read in the first line. To identify what columns contain the key words of “Assignment” and so on, my program reads in the first line of the file and find the key word in each string of the string list converted from the first line. The index where the string contains the key word is added to an integer list which stores the indexes corresponding the key word. Thus, the index of data of specific groups are stored in these lists.

1.2 In order to print correctly the first headline, use rstrip() function. However, it is not necessary if no print out result of the headline is wanted, since the find function of the string works properly whether there is a line break character in the string.

1.3 Builds four integer variables to store the sum of target scores of the four group according the step1.2 and the data in the second line.

1.4 Use for loop to read the file line by line, then convert the whole line into a list of string by split() function. Use four for loops to calculate the sum point of each group according the index of 1.1 step. In each for loop, handle ValueError for missing data or other data that cannot convert into float point number.

1.5 After calculating the sum of each group separately as well as the grade for each student, build a dictionary to store the student name and group scores of the student. Then add the dictionary of each student from each line data to the list of dictionary, which stores all the students’ scores.

1.6 Get each student’s grade from the list of dictionary with the assigned key of ‘Grade’. Use hist and plot functions to perform the statistical analysis.

2. Key features

2.1 Identify different data groups using the find function to the headline, and thus if the order of different groups with file changes, the implementation can still handle it properly.

2.2 When handing each student data, handle the ValueError for different groups separately, which is flexible if one wants different dealing method to data error(e.g. missing data) in different data groups.

2.3 With error handling for each element in each line, the implementation prints out where, exactly which column on which line, the data is missing or other value error occurring. In this way, one can check and make up for the value error very easily.

3. Code

```
from matplotlib import pyplot
import matplotlib.pyplot as plt
try:
    f=open('CEE_220_AlternativeList.txt', 'r')
except IOError:
    print("could not open the file")

line=f.readline()#takes one line at one time
#line=line.rstrip()
headline=line.split('\t')
#print headline
```

```
(AS, LB, MT, FI)=[],[],[],[]
for i in range(1,len(headline)):
    if headline[i].find('Assignment')>=0:
        AS.append(i)
    elif headline[i].find('Lab')>=0:
        LB.append(i)
    elif headline[i].find('Midterm')>=0:
        MT.append(i)
    elif headline[i].find('Final')>=0:
        FI.append(i)
studentresult=[]
line=f.readline()
targetline=line.split('\t')

(AStarget, LBtarget, MTtarget, FItarget)=0,0,0,0
for i in AS:
    AStarget+=int(targetline[i])
for i in LB:
    LBtarget+=int(targetline[i])
for i in MT:
    MTtarget+=int(targetline[i])
for i in FI:
    FItarget+=int(targetline[i])

for line in f:
    l=line.split('\t')
    (ASsum, LBsum, MTsum, FIsum)=0,0,0,0

    for j in AS:
        try:
            ASsum+=float(l[j])
        except ValueError:
            print 'missing data or data error for column', j, 'of', l[0]
    for j in LB:
        try:
            LBsum+=float(l[j])
        except ValueError:
            print 'missing data or data error for column', j, 'of', l[0]
    for j in MT:
        try:
            MTsum+=float(l[j])
        except ValueError:
            print 'missing data or data error for column', j, 'of', l[0]
```

```
        for j in FI:
            try:
                FIsum+=float(l[j])
            except ValueError:
                print 'missing data or data error for column', j, 'of', l[0]

weightedscore=ASsum/AStarget*0.25+LBsum/LBtarget*0.05+MTsum/MTtarget*0.40+FIsum/FIt
arget*0.30
        grade=(weightedscore-0.2)/0.20
        if grade<0.7 :
            grade=0.0
        elif grade>4.0:
            grade=4.0
        grade=round(grade,1)
        result = dict(student=l[0], AssignmentSum=ASsum, labSum=LBsum, MidtermSum=MTsum,
FinalSum=FIsum, Weightedscore=weightedscore, Grade=grade)
        studentresult.append(result)
f.close()
g=[]
for i in studentresult:
        g.append(i.get('Grade'))

n,bins, patches = plt.hist(g, 40, range=(0,4))
plt.xlabel('assigned numeric grade')
plt.ylabel('number of students')
plt.title('Grade distribution for CEE200')
plt.savefig('myreport.png')
plt.show()
```

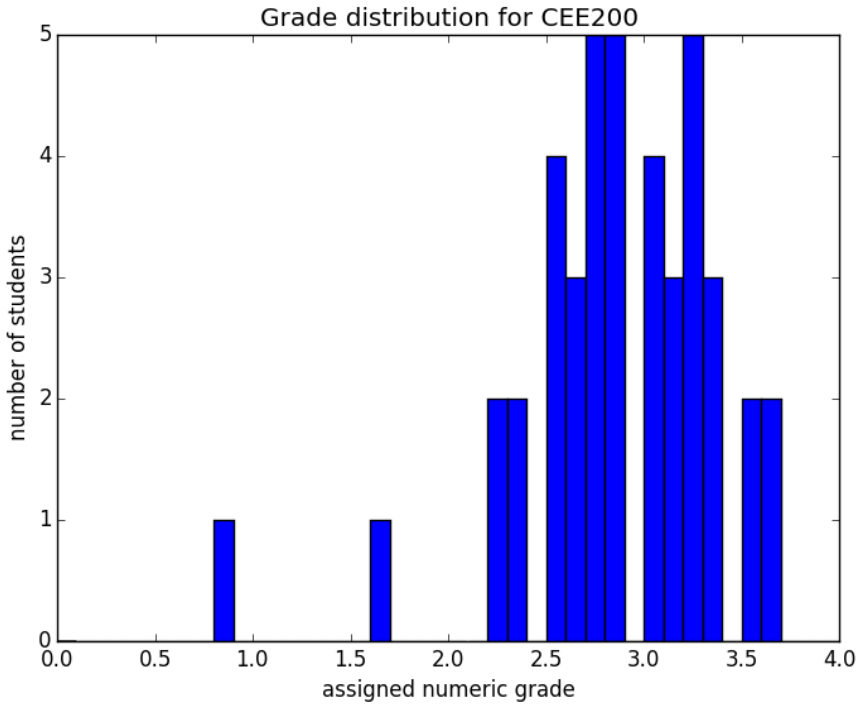
4. Results

4.1 Result on the 'CEE_220_Scores.txt' file

4.1.1 The print out contents tell all the positions of ValueError:

```
missing data or data error for column 3 of "Queen, Elizabeth"
missing data or data error for column 5 of "Queen, Elizabeth"
missing data or data error for column 9 of "Queen, Elizabeth"
missing data or data error for column 22 of "Queen, Elizabeth"
missing data or data error for column 12 of "Paul, Newman"
missing data or data error for column 10 of "Hans, Schmidt"
missing data or data error for column 11 of "Carolle, King"
missing data or data error for column 16 of "Carolle, King"
missing data or data error for column 18 of "Carolle, King"
missing data or data error for column 18 of "Ginghis, Khan"
missing data or data error for column 10 of "Lt, Ohura"
```

4.1.2 The statistical analysis result



4.2 Result on the 'CEE_220_AlternativeList'.txt file.

4.2.1 The print out contents tell all the positions of ValueError:

missing data or data error for column 10 of F4248
missing data or data error for column 24 of F4248
missing data or data error for column 23 of F4248
missing data or data error for column 5 of F4276
missing data or data error for column 24 of F427C
missing data or data error for column 24 of F4280
missing data or data error for column 24 of F4289
missing data or data error for column 24 of F4290
missing data or data error for column 24 of F429E
missing data or data error for column 24 of F42A0
missing data or data error for column 24 of F42A6
missing data or data error for column 24 of F42B0
missing data or data error for column 23 of F42B0
missing data or data error for column 24 of F42B5
missing data or data error for column 24 of F42ED
missing data or data error for column 24 of F42F4
missing data or data error for column 24 of F42F5
missing data or data error for column 24 of F42FB
missing data or data error for column 3 of F4303
missing data or data error for column 24 of F4304
missing data or data error for column 3 of F4305
missing data or data error for column 24 of F4309

missing data or data error for column 24 of F430E
missing data or data error for column 5 of F4311
missing data or data error for column 24 of F4320
missing data or data error for column 20 of F4324
missing data or data error for column 3 of F4325
missing data or data error for column 24 of F4328
missing data or data error for column 23 of F4328

4.2.2 The statistical analysis result