CITS2401 Computer Analysis and Visualisation
Semester 2, 2016
Assignment 2

Set date: 04 August 2016
Due date: **Friday 5:00 pm, 07 October 2016**
Total marks: 50

Assignment submission instructions:

- Submit 2 matlab files (Grading.m and PlotGrades.m) via the CITS2401 page on LMS (http://www.lms.uwa.edu.au).
- Put the files in a folder and rename it under filename: *YourSurname_Studentnumber*, e.g. Smith_902787889. Zip the folder using Windows Zipping/WinZip software. **[Do not use 7-zip or winrar or any other zipping tools]**.
  - If you are using windows platform in University labs then put the required file in a folder. Right click on the folder and hover/click on "*Send To"* option, then click on "*Compressed (zipped) folder*". This will create a zip file. You may need to rename it as stated above.
  - Submission guidelines must be strictly followed otherwise the grade may result in zero or penalized.
- You must name your functions and variables in the exact manner as mentioned in this assignment. Any variation will lead to penalized marks.

Plagiarism

All work submitted should be your own. I am sure that you agree that this is for your own good!! If you do not agree, please note that we have ways to detect plagiarism in code. Incidences of plagiarism will be taken seriously and will involve follow-up with the Head of School and consequences to academic results. . You can read the University policy on Academic Conduct here.

Question

Consider you have successfully completed the CITS-2401 unit and the Head of the School (HoS) has decided to hire your services to compile the results of all students in the School of Computer Science. You have been provided with an Excel file e.g. **"data.xlsx"** which contains the students' details and their marks in the courses.

The first entry of the excel file (in cell A1 of the excel file) gives the *Number of Students 'nS'* whose record is stored in the excel file. Similarly, the second entry is the *Number of Units 'nU'* which each student has taken. The subsequent entries for each student are repeated as follows

- Student Name
- Student Number
- Marks for Unit 1
- Marks for Unit 2
- And so on till Marks of unit *'nU'*

Your code must perform the following tasks

Task-A: (25 marks)

Your first task is to write a function ***Grading.m***. The first line of this function must be as follows:

function [nS, nU, Result, ImpUnitsMks, TopSt, LastSt, Dist] = Grading(filename)

Details of the inputs and outputs are as under:

Input:

1. **filename:** The name of the XLS (or xlsx) file that contains the data. As a sample to check your code, an excel file with the name 'data.xls' is provided but the file name can be changed by the user. Your function should read the excel file provided as an input to the function to extract the data.

Outputs:

1. **nS**: The number of students whose record is provided in the excel file. Remember this is the first entry of the data file.
2. **nU**: The number of units for each student whose record is provided in the excel file. Remember that this is the second entry of the data file.
3. **Result**: This is a structure containing four fields namely "StudentName", "StudentNumber", "Marks" and "Grade". The size of this structure will depend on the number of students. The fields "StudentName" and "StudentNumber" contain the name of the student and his/her student number respectively. The field "Marks" must contain a vector of marks obtained by the student in all units in the order i.e. Unit 1 to Unit nU followed by the average marks ($AvgMks_i$) of the student as described by equation (1).

$$AvgMks_i = \frac{\sum_{j=1}^{nU} Marks_j}{nU} \qquad (1)$$

where $\boldsymbol{AvgMks_i}$ denotes the average marks of the i^{th} student and $\boldsymbol{Marks_j}$ are the marks obtained by the same student in the j^{th} unit. The field "Grade" of the structure contains the grade obtained by the student. The grade is calculated as follows:

Let $AvgMks_i$ be the average marks obtained by each student in all his/her units. Let *MeanofAvg* be the average of $AvgMks_i$ and *SDofAvg* be the standard deviation of $AvgMks_i$. The *MeanofAvg* is defined in equation (2)

$$MeanofAvg = \frac{(\sum_{i=1}^{nS} AvgMks_i)}{nS} \qquad (2)$$

A student can obtain the following grades:

- '**F**'. Fail if the student fails in 3 or more individual units irrespective of his/her result in other units OR if $AvgMks < (MeanofAvg - 1.5\ SDofAvg)$. A student is considered failing in a particular Unit if s/he obtains less than 40 marks in that unit.
- '**P**'. Pass if *AvgMks* fall between *(MeanofAvg – 1.5 SDofAvg)* and *(MeanofAvg – 0.5 SDofAvg)* with both limits inclusive.
- '**CR**'. Credit Pass if *AvgMks* fall between (*MeanofAvg – 0.5 SDofAvg)* and *(MeanofAvg + 0.5 SDofAvg)* with upper limit inclusive.
- '**D**'. Distinction if *AvgMks* fall between (*MeanofAvg +0.5 SDofAvg)*and *(MeanofAvg +1.5 SDofAvg)* with right hand limit inclusive.
- '**HD**'. High Distinction if $AvgMks > (MeanofAvg + 1.5\ SDofAvg)$

The variable Result containing the structure for the sample provided will look like this:

```
>> Result
Result =
1x156 struct array with fields:
    StudentName
    StudentNumber
    Marks
    Grade
```

And the structure for 156th students shows following details:

```
>> Result(156)
ans =
    StudentName: 'S156'
    StudentNumber: 20646624
    Marks: [93 32 100 50 40 63]
    Grade: 'D'
```

4. **ImpUnitsMks.** The HoS considers the first three units to be very important. This variable is a matrix that contains the marks of first three units of each student. The size of the matrix is '*nS* x *3*'.

5. **TopSt, LastSt.** This variable contains the index numbers of the top four and last four students in the three important units only. You should find the average marks of each student in the three important units and then find the top and bottom four students. The index numbers should be in descending order with respect to the average marks of each student in the three important units.

6. **Dist.** This variable is a 4 x 4 matrix that contains the distance between each of the top four students and each of the last four students in the space of important units. Each element Dist(m,n) will contain distance between student *m* among the top four studens and student *n* among the last four students. For instance, Dist(2,3) will contain the distance between 2nd student in variable TopSt and 3rd student in variable LastSt. If x_1, y_1, z_1 are the marks obtained by the first student in the important units u1, u2 and u3 respectively and x_2, y_2, z_2 are the marks obtained by the second student in the important units u1, u2 and u3 respectively, then their distance in the space of important units is defined as :

$$d = \sqrt{((x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2}$$

(Hint: You may solve this equation using Matrix Operations taught in lecture – 8)

In summary the outputs of your function for the provided data in excel file will look like this:

```
[Result,ImpUnitsMks,TopSt,LastSt,nS,nU,Dist]=Grading(filename)
Result =
1x156 struct array with fields:
    StudentName
    StudentNumber
    Marks
    Grade
ImpUnitsMks =
   100    17    31
    72    79    18
   100    56    13
    52    97    92
    54    76    57
```

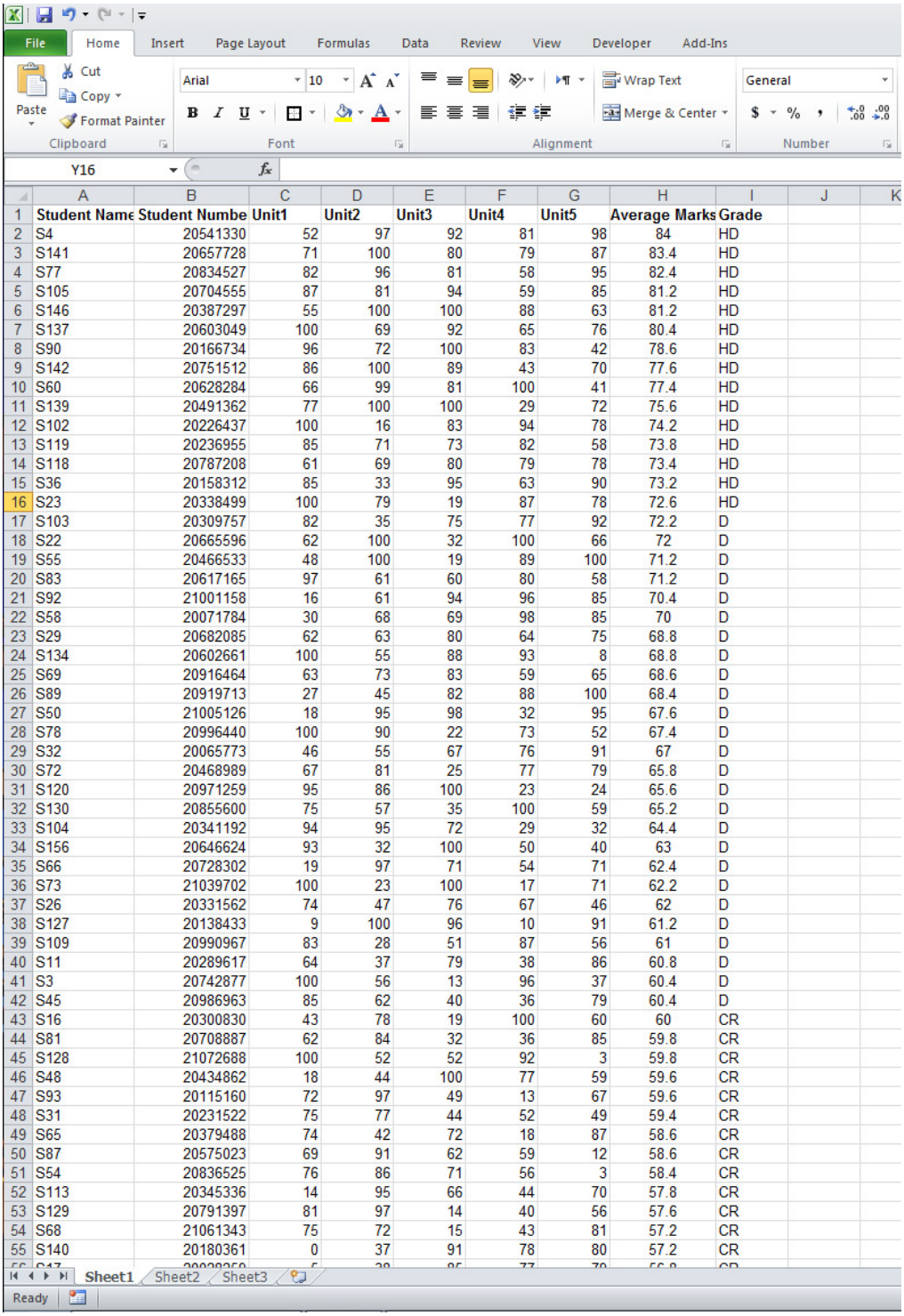
```
    91    58    14
    (Another 150 rows which have not been pasted here due to lack of space.)
TopSt =
   120
   139
   142
    90
LastSt =
   115
    37
    10
    94
nS =
   156
nU =
     5
Dist =
    128.43    128.45    133.55     136.9
    127.05    128.83    132.08    136.18
    128.89    127.71    131.98    134.32
    120.71    120.85    126.57    130.26
```

Task-B: X-Marks (5 marks)

Your code needs to generate an excel file which present the details of the students and their marks in the format presented below and under the following headings:

Student Name, Student Number, Unit1, Unit2, ….., AvgMks, Grade

Sort the students in the descending order of their Average Marks. Write this formatted and sorted data to an Excel file named as **"Result.xlsx".** The code for generating this output file must be written inside the ***Grading.m*** function. Below is a screen shot of this file containing first few rows for the provided sample data.



Student Name	Student Numbe	Unit1	Unit2	Unit3	Unit4	Unit5	Average Marks	Grade
S4	20541330	52	97	92	81	98	84	HD
S141	20657728	71	100	80	79	87	83.4	HD
S77	20834527	82	96	81	58	95	82.4	HD
S105	20704555	87	81	94	59	85	81.2	HD
S146	20387297	55	100	100	88	63	81.2	HD
S137	20603049	100	69	92	65	76	80.4	HD
S90	20166734	96	72	100	83	42	78.6	HD
S142	20751512	86	100	89	43	70	77.6	HD
S60	20628284	66	99	81	100	41	77.4	HD
S139	20491362	77	100	100	29	72	75.6	HD
S102	20226437	100	16	83	94	78	74.2	HD
S119	20236955	85	71	73	82	58	73.8	HD
S118	20787208	61	69	80	79	78	73.4	HD
S36	20158312	85	33	95	63	90	73.2	HD
S23	20338499	100	79	19	87	78	72.6	HD
S103	20309757	82	35	75	77	92	72.2	D
S22	20665596	62	100	32	100	66	72	D
S55	20466533	48	100	19	89	100	71.2	D
S83	20617165	97	61	60	80	58	71.2	D
S92	21001158	16	61	94	96	85	70.4	D
S58	20071784	30	68	69	98	85	70	D
S29	20682085	62	63	80	64	75	68.8	D
S134	20602661	100	55	88	93	8	68.8	D
S69	20916464	63	73	83	59	65	68.6	D
S89	20919713	27	45	82	88	100	68.4	D
S50	21005126	18	95	98	32	95	67.6	D
S78	20996440	100	90	22	73	52	67.4	D
S32	20065773	46	55	67	76	91	67	D
S72	20468989	67	81	25	77	79	65.8	D
S120	20971259	95	86	100	23	24	65.6	D
S130	20855600	75	57	35	100	59	65.2	D
S104	20341192	94	95	72	29	32	64.4	D
S156	20646624	93	32	100	50	40	63	D
S66	20728302	19	97	71	54	71	62.4	D
S73	21039702	100	23	100	17	71	62.2	D
S26	20331562	74	47	76	67	46	62	D
S127	20138433	9	100	96	10	91	61.2	D
S109	20990967	83	28	51	87	56	61	D
S11	20289617	64	37	79	38	86	60.8	D
S3	20742877	100	56	13	96	37	60.4	D
S45	20986963	85	62	40	36	79	60.4	D
S16	20300830	43	78	19	100	60	60	CR
S81	20708887	62	84	32	36	85	59.8	CR
S128	21072688	100	52	52	92	3	59.8	CR
S48	20434862	18	44	100	77	59	59.6	CR
S93	20115160	72	97	49	13	67	59.6	CR
S31	20231522	75	77	44	52	49	59.4	CR
S65	20379488	74	42	72	18	87	58.6	CR
S87	20575023	69	91	62	59	12	58.6	CR
S54	20836525	76	86	71	56	3	58.4	CR
S113	20345336	14	95	66	44	70	57.8	CR
S129	20791397	81	97	14	40	56	57.6	CR
S68	21061343	75	72	15	43	81	57.2	CR
S140	20180361	0	37	91	78	80	57.2	CR

Task-C: X-Marks (20 marks)

The HoS wants to visualise the results that you have just compiled. You are required to write another function which generates three plots. The first line of the function must be:

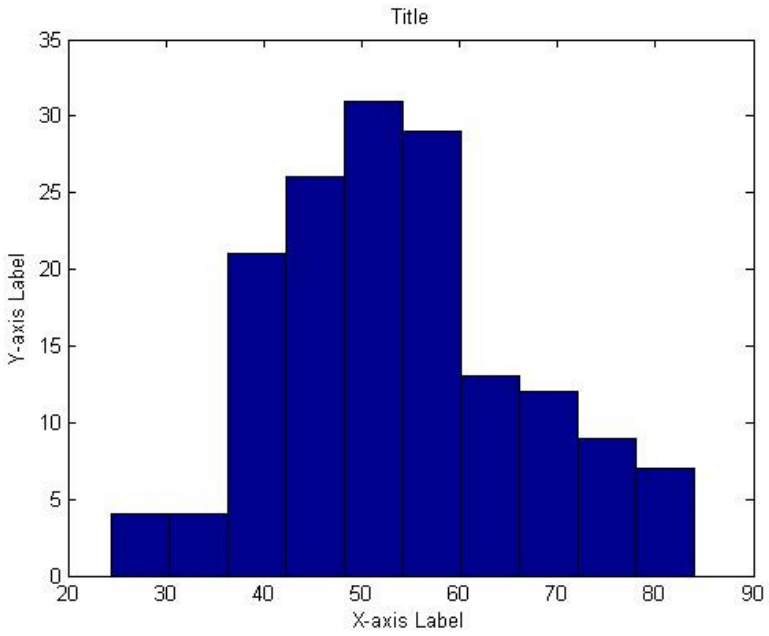
function PlotGrades(filename)

The input is the same filename as described in Task-A. Inside this function you should call the function ***Grading.m*** that you created in Task-A and generate the following plots using only the outputs of ***Grading.m.*** *Do not copy paste the code from Task-A or perform the calculations again in this function.*

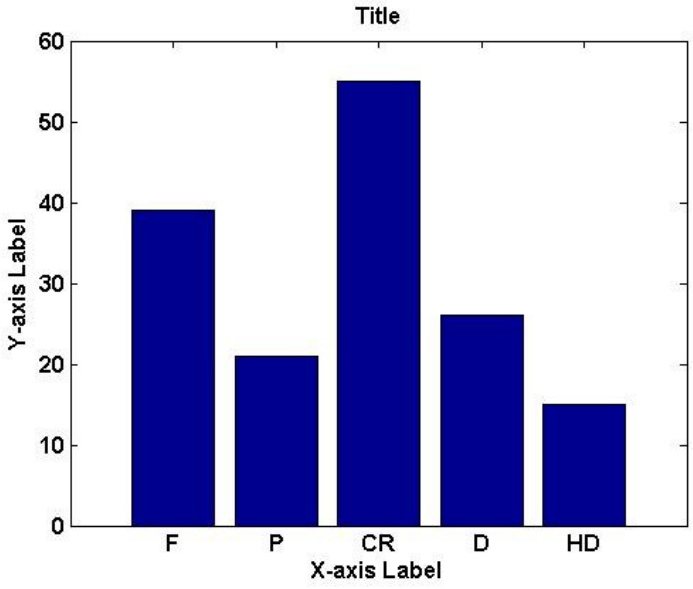
- Plot-1. The histogram of $AvgMks_i$ for all students.
- Plot-2. The histogram of the grades obtained by the students.
- Plot-3. Considering the space of the marks of the Important Units, draw a 3D scatter plot of the position of each student in this space. Mark the top four students in green colour and the last four in red colour.

Remember the characteristics of good plotting taught to you in Lecture-7. The screen shots of the plots for the provided sample data are below. You will need to insert the right title, y-axis label and x-axis label for each plot.

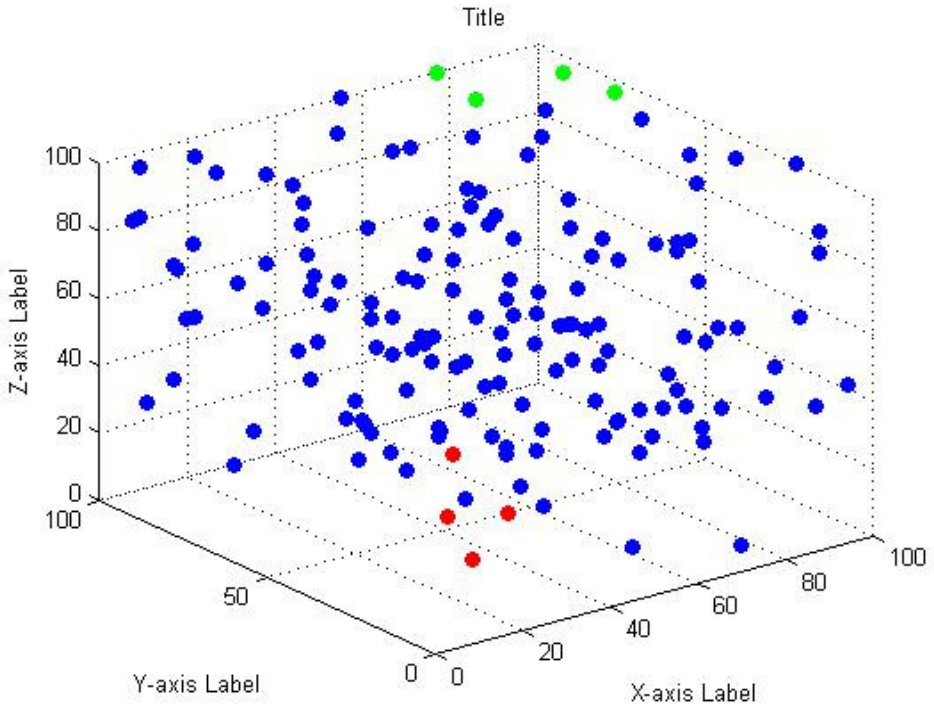
PLOT-1



PLOT-2



PLOT-3



Note: Your code must be generic and must work with any provided excel file containing the data in the above mentioned format. Your code will be tested with a DIFFERENT data file and NOT with the same provided sample excel file.