DSO 545: HW 2
Due February 22

Submission instructions:

1. One submission per group
2. Submit only R code:
   Names (Official Last name Official First name):

   # Case 1:

   #1.
   Code

   #2.
   Code

   #3.
   Code

   ….

   #22.

   # Case 2:

   #1.
   Code

   #2.
   Code

   #3.
   Code

   ….

   #39.

Case 1: Assigning grades

Description: Marshall School of Business offers a wide variety of graduate business and related majors:

Big Data Analytics	BDA
Business Statistics	BS
Accounting	AC
Marketing	MK
Management	MG
Business Administration	BA
Information systems	IS
Finance	FN
Operations management	OM

You are a grader for DSO 545, an elective class in data analysis and modeling in USC's MBA program. At the end of the semester your task is to assign final grades and perform grade analysis based on the following guidelines. The syllabus of the course indicates the following: the course grades are based on the four highest quizzes (the lowest of the five should be dropped), two individual projects, a team project, a midterm and a final exam.

The final grade will be based on the following weightings:

Assignment	Weight in %
Quizzes	12
Projects	20
Team Project	15
Midterm Exam	20
Final Exam	33

The letter grade will be assigned as follows:

Course total weighted score range	Course letter grade	Numerical value of course grade
94 to 100	A	4
92 to 94	A-	3.7
90 to 92	B+	3.3
84 to 90	B	3
82 to 84	B-	2.7
80 to 82	C+	2.3
74 to 80	C	2
72 to 74	C-	1.7
70 to 72	D+	1.3
64 to 70	D	1
62 to 64	D-	0.7
below 62	F	0

Data: Gradebook data from blackboard was downloaded and is stored in **Gradebook.txt**. The above tables are in

GradebookAdditionalInfo1.txt, GradebookAdditionalInfo2.txt, GradebookAdditionalInfo3.txt,

GradebookAdditionalInfo4.txt. Note that all assignment and exam scores in the gradebook are based on a 0 to 100% scale.

Questions:

1. Which assignment has the lowest average?

2. What is the corresponding lowest average?

3. Create a tibble that displays lowest score of all assignments for each major.

4. Which assignment has the highest median? (median = middle of a sorted list from smallest to largest)

5. What is the corresponding highest median?

6. Create a tibble that displays highest score of all assignments for each major.

7. Which quiz has largest number of zeros?

8. What is the largest corresponding number of zeros?

9. Calculate average quiz score for each student. How many students have average quiz score 80 or more but less than 90?

10. Create a tibble that displays average of each assignment for each major.

11. On a 0 to 100 percent scale calculate the average final grade in the course. (Hint: do not forget that the final grade is the weighted average!)

12. Which major corresponds to the highest number of students who performed excellently?

Additional information:

Students are usually concerned with GPA as it relates to their performance over a period of time in several courses; however, professors must be concerned with the course GPA because sometimes a particular course is curved. On a 0 to 4 scale calculate the average GPA for the course. Marshall School of Business standard specifies that the curve on a 4.0 scale be 3.0 give or take 0.1.

13. On a 0 to 4 scale calculate the average final grade in the course.

14. On a 0 to 4 scale calculate the average final grade in the course for each major separetely. Which major corresponds to the highest average?

15. Does the course need to be curved?

16. If instead of dropping the lowest quiz score, a professor decides to drop two lowest quiz scores. On a 0 to 4 scale calculate the new average final grade in the course. (**drop two lowest quiz scores only for this question, for all other questions drop only the lowest quiz score**)

Additional information:

The following defines qualitative performance in the course: "Poor Performance" for those students who have grades F, D-, D or D+. "Below Average" for those students who have grades C-, C, or C+. "Good for those students who have grades B-, B, or B+. "Excellent" for those students who have grades A, or A-.

17. What is the distribution in percentage of the above 4 qualitative performance descriptors?

18. Which major corresponds to the lowest percentage of Below Average performance in the course?

19. What the corresponding percentage value?

20. Students who receive 0 points on quizzes, projects, or exams are more likely to receive a "Poor Performance". Among students receiving a "Poor Performance", what is the average number of quizzes, projects, or exams with a 0 point value?

21. Mirzagaliyeva Shakhizada (ID 20006187) this semester took the following courses: 2 unit course, BUAD 525, from Prof. Plotts and received D+; 4 unit course, DSO 524, from Prof. Ku and received A-; 6 unit course, GSBA 545, from Prof. Porter and received C+ and 3 unit course, DSO 545, from Prof. Gabrys for which you need to determine the grade. Calculate the 4.0 scale GPA for Shakhizada.

22. Due to a very low average, Marshall has decided to inflate ("curve") everybody's total weighted score by the percentage specified in the below table

	A	A-	B+	B	B-	C+	C	C-	D+	D	D-	F
BDA	0.0	1.5	2.1	4.3	4.0	12.0	8.0	10.0	4.0	11.0	11.0	12.0
BS	0.0	1.8	2.3	2.8	4.0	4.0	9.0	8.0	4.0	11.0	3.0	14.0
AC	0.0	1.4	2.0	3.9	7.0	6.0	3.0	8.5	8.0	11.0	2.0	14.0
MK	0.0	1.3	2.7	4.1	7.0	4.0	6.0	8.0	5.0	2.0	4.0	14.0
MG	0.0	1.2	2.6	2.9	8.0	9.0	4.0	5.0	7.0	6.0	3.0	10.0
BA	0.0	1.6	2.5	3.6	6.0	9.0	9.0	9.0	8.0	10.0	8.0	13.0
IS	0.0	1.7	2.5	2.8	4.0	7.0	8.0	3.0	6.0	7.0	9.0	7.0
FN	0.0	1.9	2.4	4.7	7.0	8.0	9.0	5.0	3.0	4.0	9.0	12.0
OM	0.0	0.9	1.9	2.5	3.0	3.0	5.0	9.0	8.0	8.0	9.0	11.0

On a 0 to 4 scale calculate and report the new "inflated" average final grade in the course. (Numbers in the table are in %, i.e. BDA & A- = 1.5% = 0.015)

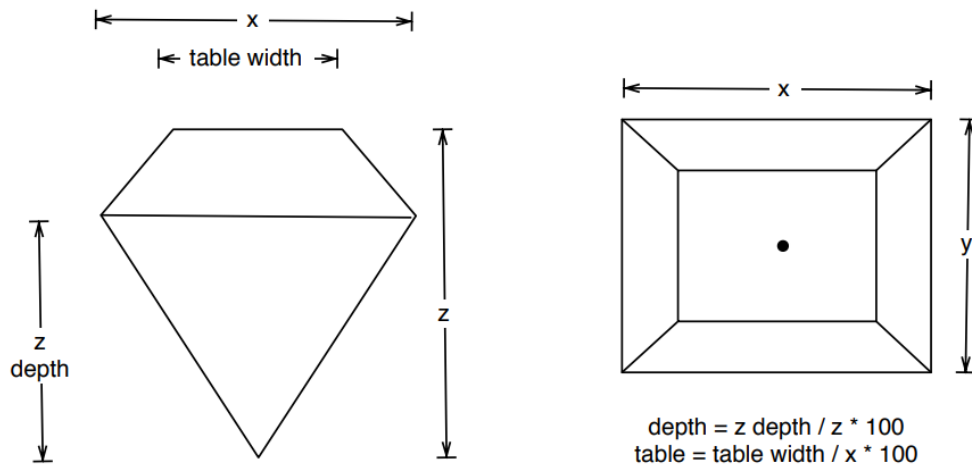# Case 2: Prices of round cut diamonds

Description:

A dataset containing the prices and other attributes of almost 54,000 diamonds.

The variables are described in the table below:

Data: Columns are in bold

price	
	price in US dollars ($326–$18,823)
carat	
	weight of the diamond (0.2–5.01)
cut	
	quality of the cut ("Fair, Good, Very Good, Premium, Ideal)
color	
	diamond colour, from J (worst) to D (best)
clarity	

	a measurement of how clear the diamond is (I1 (worst), SI1, SI2, VS1, VS2, VVS1, VVS2, IF (best))
x	
	length in mm (0–10.74)
y	
	width in mm (0–58.9)
z	
	depth in mm (0–31.8)

depth = z depth / z * 100
table = table width / x * 100

How the variables x, y, z, table and depth are measured.

Instructions: You will submit R code on blackboard that helps answer the questions

Questions:

1. Import data in R and call it diamonds

2. What is the data structure of diamonds?

3. What are the names of columns?

4. Is the above data structure a vector?

5. Is the above data structure a data frame?

6. What type of data does the object that represents column names (answer to your previous question) contain?

7. How big is the table, i.e. how many rows and columns does the table have?

8. Are there any missing values in the data table? How many?

9. How many factor variables does the table have?

10. How many factor variables are there in the table?

11. In this case all factor variables should be ordinal. Based on the variable definitions and descriptions in the above table, convert all factor variables to ordinal factor variables.

12. Compactly display the Structure of the data table

13. Summarize each column

14. How many premium and very good diamonds fall in the price range between mean and median price of all diamonds? Your code should be general enough and return correct answer.

15. Create a scaterplot of price versus weight. Y axis label = Price, X axis label = Weight, Graph title = "Price versus Weight"

16. Build a regression model to predict Price from weight and call it M

17. What is R^2 in %? Based

18. Recreate a scatter plot of price versus weight and overlay a regression line. Regression line should be dashed, thickness=2, and color=red. Add a legend. Hint: you may look at the syntax for legend (there are examples at the bottom of documentation(see help))

19. Upload dplyr and ggplot2 libraries

20. Convert diamonds into a tbl_df format and call it diamonds

21. Create a histogram of price, overlay its density curve and change the default grey background to a white background color histogram bars lighblue, density curve red and make it of thickness of 2

22. Recreate Histogram in 21 by mapping cut to color and facet on cut

23. Recreate Histogram in 21 by mapping color to color and facet on color

24. Recreate Histogram in 21 by mapping clarity to color and facet on clarity and

25. Recreate Histogram in 21 by mapping cut to color and facet on cut and color :

26. Create a scatterplot of price versus weight and color and map cut to color. Call this object g

27. Make a faceted plot of the above scatterplot g by faceting on clarity

28. Make a faceted plot(on a grid) of the above scatterplot g by faceting on clarity and color

29. Create side by side boxplots diamond carats by cut and add the title Boxplot of diamond carats by cut

30. Create a tibble that contains only carat and price (use pipe)

31. Create a tibble that contains columns that start with letter c (use pipe)

32. Create a tibble that contains columns that end with letter e (use pipe)

33. Create a tibble that contains columns that have letter l (use pipe)

34. Create a tibble that contains all columns but carat and price (use pipe)

35. Create a tibble for Ideal diamonds that contains only carat and price (use pipe)

36. Create a tibble for Ideal and Premium diamonds with carat and price being removed and added an additional column price per carat called PricePerCarat (use pipe)

37. Create a tibble that contains mean and median of price and carat: MeanPrice and MeanCarat (use pipe)

38. Create a tibble that shows average price and average price per carat for each category of quality of the cut and call it MeanPrice and MeanPricePerCarat (use pipe)

39. Break down the levels of cut using color variable and create a tibble that displays means both prices for each combination of cut and color and sort the rows in MeanPricePerCarat descending order (use pipe)