Professor Notes About the "Univariate EDA - Quant" Homework

- Question 1 asks for a "univariate EDA" – Note that this does not mean to just make a histogram and summary statistics. Performing a univariate EDA means addressing shape, outliers, center, and dispersion from looking at a histogram (or boxplot) and summary statistics. You MUST remember to explain why you chose to use the mean/sd or median/IQR to address shape/center. Though there are no outliers in this example, your description of the outlier should be specific (i.e., say that "there is an outlier at 71" not that "there is an outlier in the 70-80 range.")
- In question 1 the shape, outliers, center, and dispersion are specifically listed, even if there was no outliers, for both class types. Also note that I clearly indicated why I chose to use the mean/sd (because of symmetry and no outliers) or median/IQR (skew or the occurrence of outliers).
- Note how each table and figure is labeled and referred to in the answers.
- You should use no more than one more decimal place than what was recorded for the quantitative variable. GPA is usually recorded to two decimal places so you should use three decimals in the results from `Summarize()`.
- Note that the three individuals on the far right for the math class histogram should not be considered as outliers. The main reason is that outliers should be a very few number of individuals. The two bars on the far right contain 3 of the 22 (14%) math classes. This is simply too many to call outliers. Those individuals should be considered as a separate group (unlikely in this case) or as a simple continuation of the right tail (more likely).

Math Class Grades

1. The distribution of GPA for the math classes is slightly right-skewed with no obvious outliers (Figure 1). The median for the math courses is 2.30 with an IQR from a Q1 of 2.165 to a Q3 of 2.445 (Table 1). The median and IQR were used because of the skewness in the data.

   The distribution of GPA for the other classes is left-skewed with no obvious outliers (Figure 1). The median for the other courses is 2.54 with an IQR from a Q1 of 2.440 to a Q3 of 2.620 (Table 1). The median and IQR were used because of the skewness in the data.

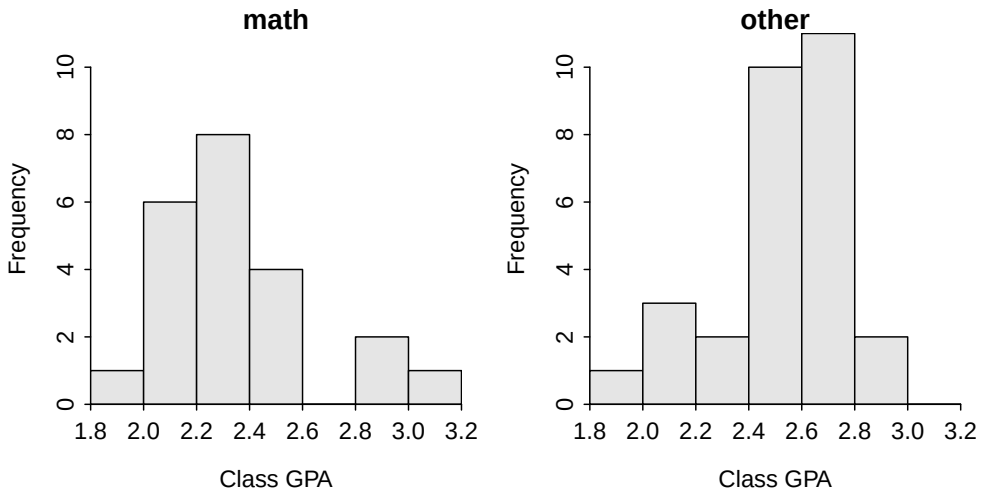

Figure 1. Histograms of class gpa for math (left) and other (right) courses at the University of North Carolina.

Table 1. Descriptive statistics for class gpa of math (left) and other (right) courses at the University of North Carolina

	class.type	n	mean	sd	min	Q1	median	Q3	max
1	math	22	2.353	0.301	1.90	2.165	2.30	2.445	3.02
2	other	29	2.508	0.235	1.96	2.440	2.54	2.620	2.95

2. The two most outstanding differences between the GPA in math and other courses is that the GPA in the math courses is slighty lower (lower median GPA) and slighly more dispersed (more variable as shown by a greater IQR).

R Appendix

```
library(NCStats)
setwd('C:/aaaWork/Books/IntroStats/HW/')
d <- read.csv("UNCgrades.csv")
str(d)
hist(gpa~class.type,data=d,xlab="Class GPA")
```