Take Home Exam 1

Due Sunday, September 18th, 2016

STAT 4013: Statistical Methods I (FALL 2016)

Directions: This paper is to be the cover page of your exam. All questions are to be typed and attached to the end of this document. You must submit it to D2L **Dropbox** no later **than September 18th, 2016** at **11:59 pm**. This portion of the exam is to test your knowledge of both statistical concepts learned thus far and how to effectively use statistical software (in this case R). All work submitted must be your own. You may use anything at your disposal except another human being. Solutions should be written following all rules of English, and points will be deducted as needed**. Graphics should appear within the solutions, and an appendix with just your R code should be attached at the end.**

**You are allowed only one submission to Dropbox, so make sure your answers are as complete as you want them to be before submitting because there is no going back!!!!**

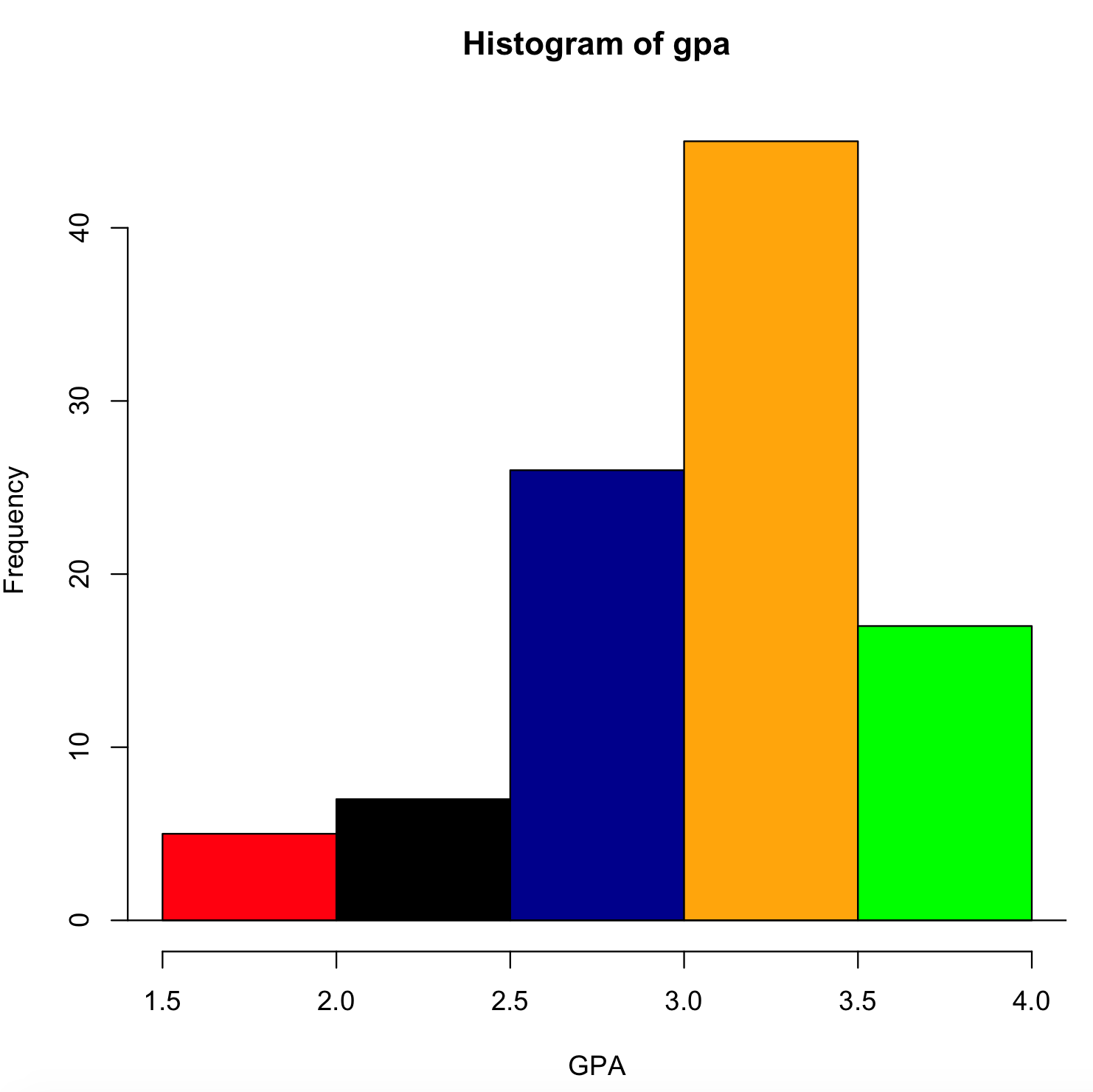
**Failing to submit your code will result in a reduction of your grade by 20 points.**

***Questions:***

Your best friend just started their second statistics course, and they are gung-ho about jumping right in to the project being assigned to them. Your friend collects this wonderful data on grades in other students’ first ever statistics courses; however, it has been a while and she needs help with starting out. Answer the following questions to help her start her research project.

The data can be found at <http://rfs.kvasaheim.com/datafiles.php>. The file name is statgrades.

1. She first wants to investigate the College GPA of the students that she collected data on.
   1. Create an appropriate graph including labels and colors if necessary. (5 points)
   2. Fully describe what the graphic tells us. (3 points)
2. Your friend is beyond thankful for your help so far, but now she wonders if the college a student in is determined by gender or not.
   1. Create an appropriate **relative frequency** graph for these two variables including: labels, colors, and a legend if necessary. (5 points)
   2. Fully describe what the graphic tells us. (3 points)
3. Next, your friend wants to explore the grade a student earned in their first statistics course.
   1. Create an appropriate graph including labels and colors. (5 points)
   2. Fully describe what the graphic tells us. (3 points)
   3. Compute the 5-Number Summary. (5 points)
   4. Are there any outliers in the data set? Explain. Use appropriate mathematics in R to support your answer. (4 points)
4. Finally, your friend wants to explore an association between College GPA and SAT Math Score.
   1. Create an appropriate graph including labels. (5 points)
   2. Does there appear to be a linear association between the two variables? Explain. (3 points)
   3. Compute a statistic that will either support or refute your claim. Interpret it. (5 points)
   4. Does the statistic match your initial guess? If not, what could cause this? (4 points)



The above graph is histogram graph, it usually used to graph between quantitative variables. Here we are computing the College GPAs.

These are the following R commands used:

statgrades = read.csv("/Users/thotasairam/Documents/Fall 2016/STAT 4013/statgrades.csv")

attach(statgrades)

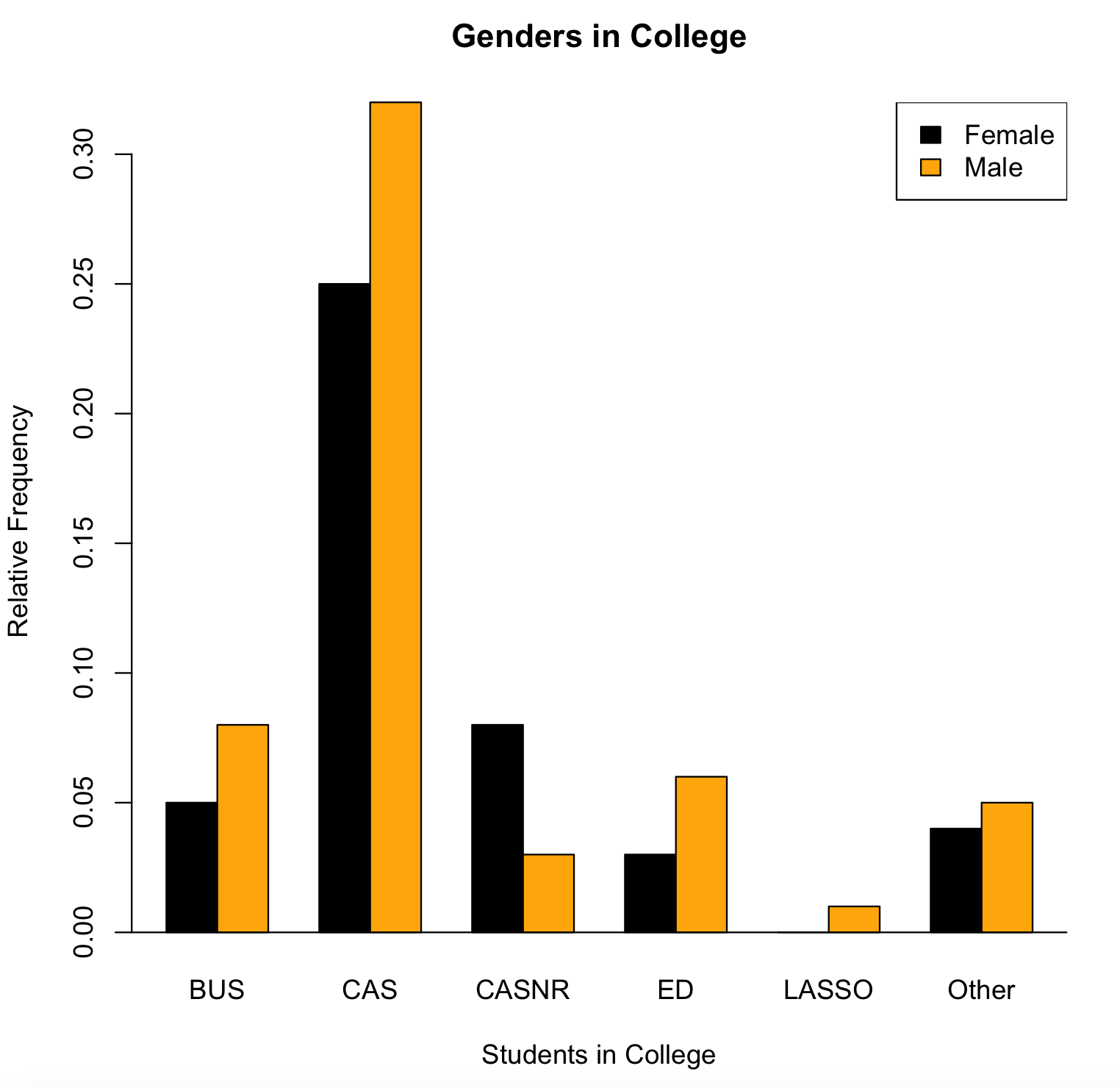
names(statgrades)

hist(gpa, xlab= "GPA", col = c("red", "black", "darkblue","orange","green"))

abline(h=0)

1. The is a unimodal graph, and it is left skewed. Also, the mean is 3.25 approximately.

2)



Since it is a relative frequency between to qualitative variables, we need to use a clustered bar graphs. The R code is down below:

statgrades = read.csv("/Users/thotasairam/Documents/Fall 2016/STAT 4013/statgrades.csv")

attach(statgrades)

names(statgrades)

counts = table(gender,college)

counts.relative = counts / sum(counts)

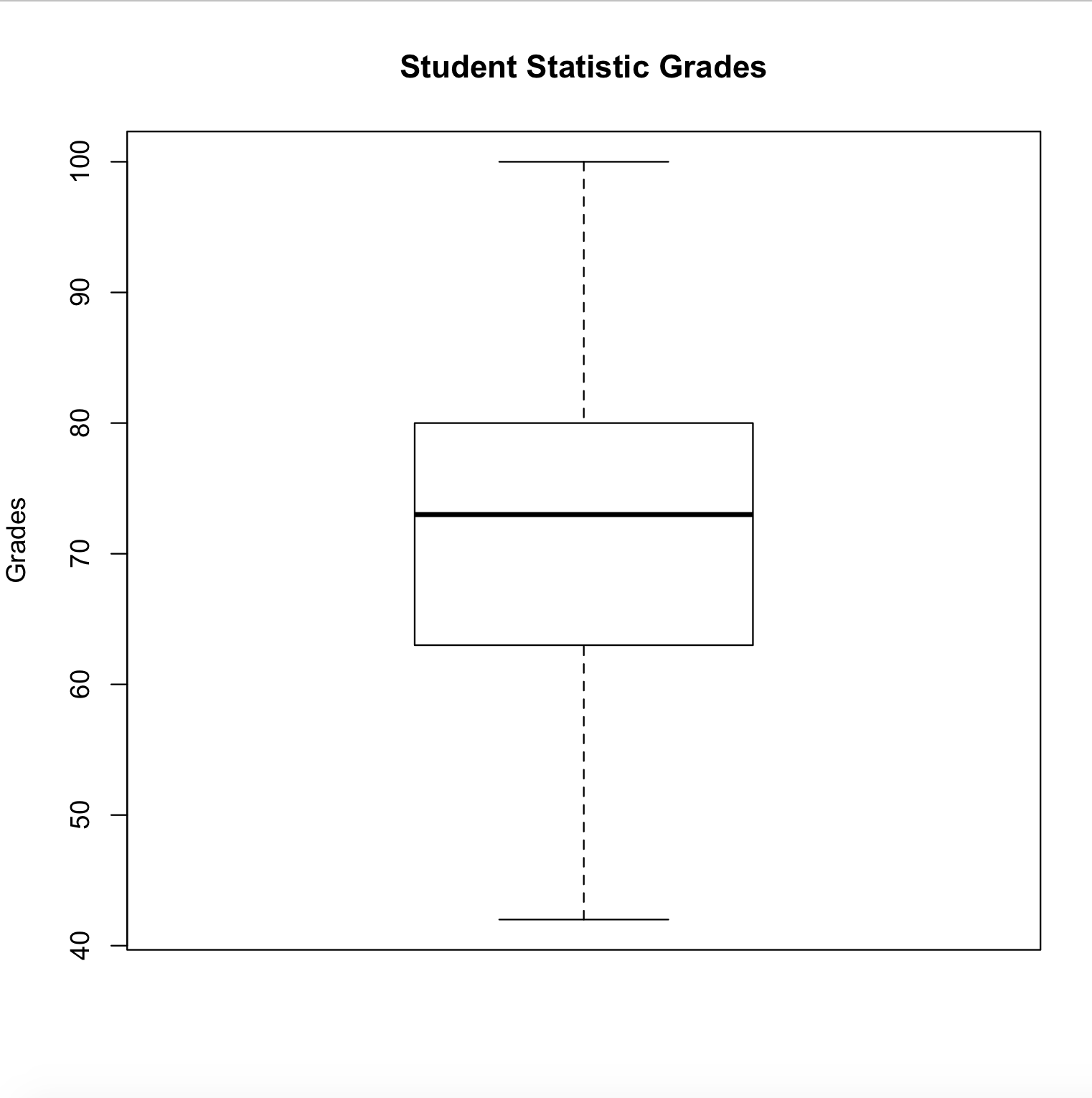
barplot(counts.relative, beside = table(gender), main = "Genders in College", xlab = "Students in College", ylab = "Relative Frequency", col = c("black", "orange") )

abline(h=0)

legend("topright", legend = c("Female", "Male"), fill = c("black", "orange"))

1. The ratio between Men and Women percentage in different colleges are shown in the graph. Overall, men are admitted more than women do. And its unimodal, and right skewed bar graph.

3)

a.

This is R code used for the bax plot above:

boxplot(grade, main= "Student Statistic Grades", ylab= "Grades")

abline(h=0)

b.

This graph looks symmetric, i.e. mound shaped. And, The grades are evenly distributed, with mean approximately equal to 73. And more than 95% of grades lies between 65-80 approximately.

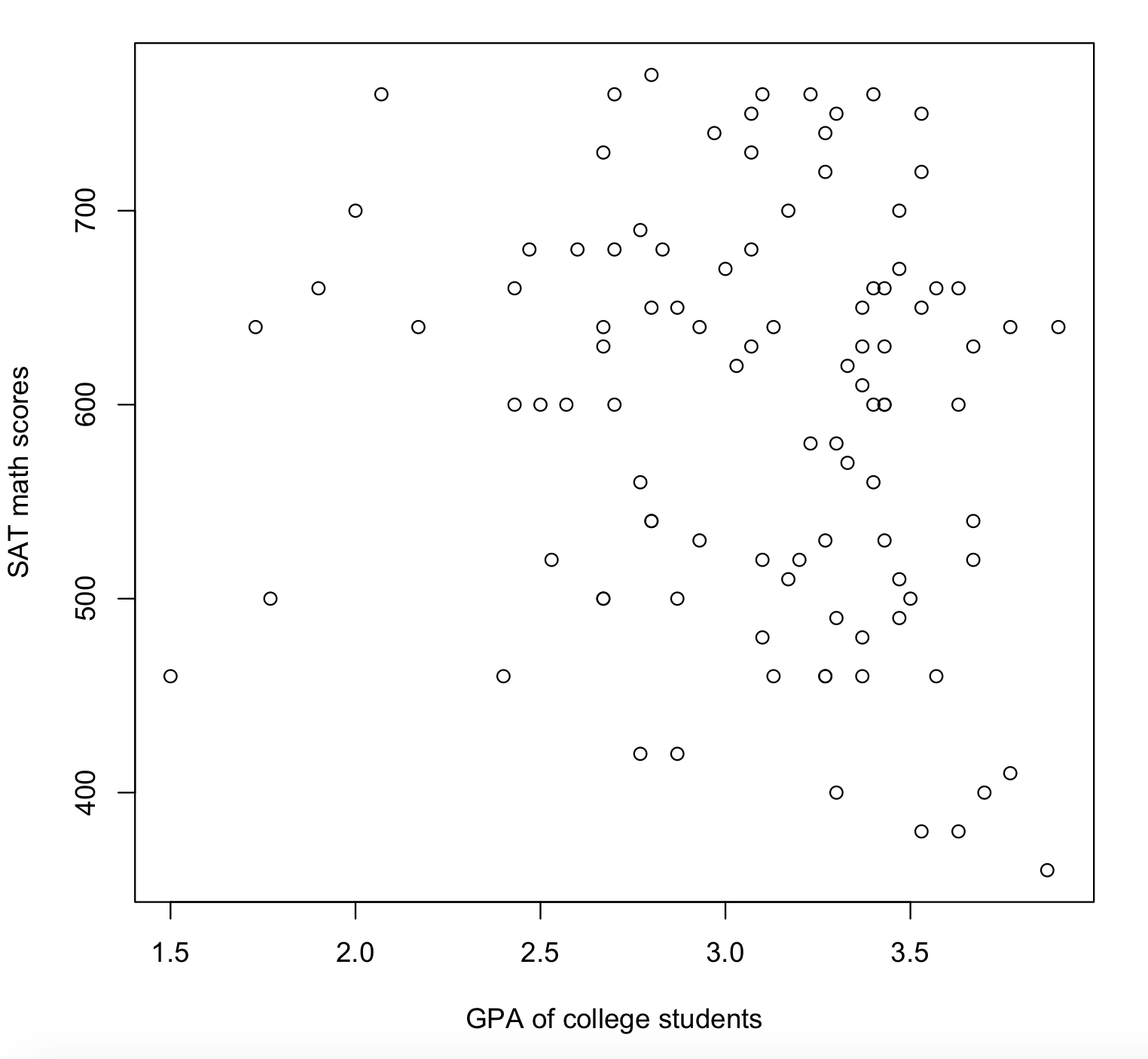
1. 5 number summary is (42,63,73,80,100) as (min,Q1,median,Q3,max). And the R command used for computing this is

summary(grade)

1. There are no outliers in this graph. As I mentioned earlier the graph is is uniform and mound shaped.

4)﷒

a.



The above scatter plot is between the college GPA and and SAT math scores. This is the R code for the above graph

plot(gpa, SATmath, xlab = "GPA of college students", ylab= "SAT math scores")

b.

There is a linear relationship between the GPA and sat math scores but r < 0. Because the graph has negative slope and is linear with some outliers.

c.

The correlation is approximately equal to 0.54. And the result refutes my prediction. I think the reason is because of outliers. The r code I used is

cor(grade,SATmath)

d.

As I mention earlier, the statistic did not match my prediction. And the main reason behind this is outliers. It is an example of Anscombe’s Quartet models.