code <- read.csv(file.choose(), header=TRUE)

summary(code)

attach(code)

> summary(fds[,c("UGPA","Texts")])

1) #in UGPA median is slightly larger than mean,thus it is left skewed,yet could also be seen as almost normal.

#in Texts,median is smaller than mean,thus it is right skewed.

2) par(mfrow=c(2,1))

boxplot(UGPA, horizontal=TRUE)

> boxplot(Texts, horizontal=TRUE)

Yes,it is coincided with the conclusion in a),yet it is also noticeable that

The 2nd one is a strong right skewed since almost all data is in the left

3) par(mfrow=c(1,2))

> hist(Salary)

> qqnorm(Salary)



The Salary is a right skewed and is not normally distributed from the 2nd graph,since it

Is not fitted into a line in the plot.

e)



d) 

The above plots are very much scattered and not fitting into a line.

Not much improvement is made.

e)

major1<-subset(code,UMajor == "Biological Sciences")

> attach(major1)

boxplot(major1,horizontal="TRUE")

##The left one below is for Biological Science

f)



Ignore the left one,the right one is what we want,and the point on top is the outlier where Sebastian sits

g)

cor(cbind(Age,UGPA,FTJobs,Salary,Spending,Adviser,Texts,Wealth),use="pairwise.complete.obs")

FTjob and Salary are most correlated with Wealth

As they each have 0.23644356 and 0.14852203 where have obsolute values bigger than the others

h)

> model1 = lm(Wealth ~ Gender + Employ + Computer)

> summary(model1)

Call:

lm(formula = Wealth ~ Gender + Employ + Computer)

Residuals:

Min 1Q Median 3Q Max

-15.924 -11.696 -4.190 0.353 84.286

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 8.7547 14.2522 0.614 0.543

GenderMale 0.3098 7.1066 0.044 0.965

EmployPart-Time -10.8837 9.7851 -1.112 0.273

EmployUnemployed -13.2070 8.8734 -1.488 0.145

ComputerLaptop 6.9590 13.8690 0.502 0.619

ComputerTablet -8.0645 26.6832 -0.302 0.764

Residual standard error: 22.81 on 38 degrees of freedom

Multiple R-squared: 0.07946, Adjusted R-squared: -0.04167

F-statistic: 0.656 on 5 and 38 DF, p-value: 0.6588

R-sqr =.07946, .07946 of the variation in Wealth is being explained by X ,this is pretty Low and bad

22.81 on 38 degrees of freedom

R-sqr adjusted is -.04167,the proportion of variation explained by the regression, having taken in account for the penalty on the 5 explanatory variables in the model.

i)

Ho: B = 0

H1: B ≠ 0

(4.591– 0)/std

j)

model = lm(Wealth~ Age + GGPA + FTJobs + Salary + Spending+ Adviser + Texts+ Wealthth + Employ + Computer)

Error in eval(expr, envir, enclos) : object 'Adviser' not found

> model = lm(Wealth~ Age + GGPA + FTJobs + Salary + Texts+ Wealthth + Employ + Computer)