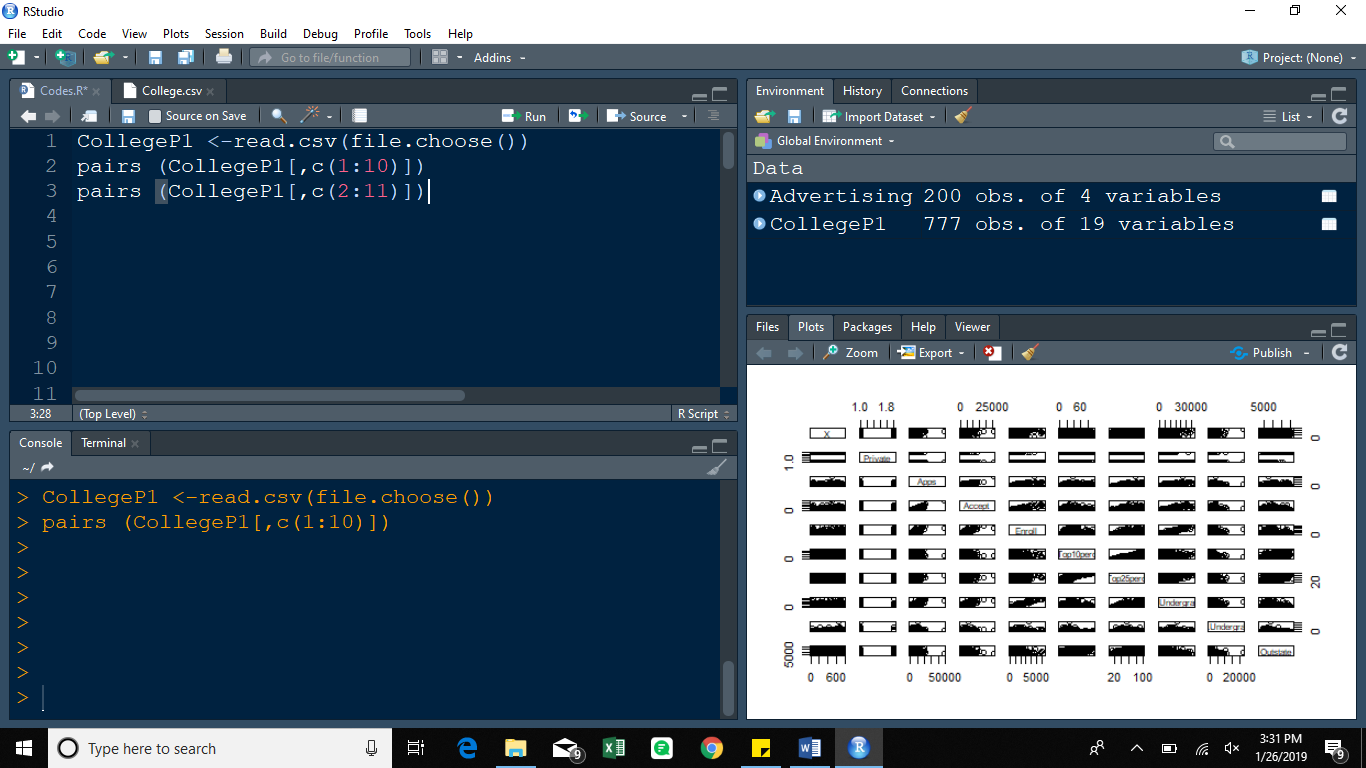
**Problem 1**

1. Use R functions to produce a scatterplot matrix of the first ten columns or variables of the data.

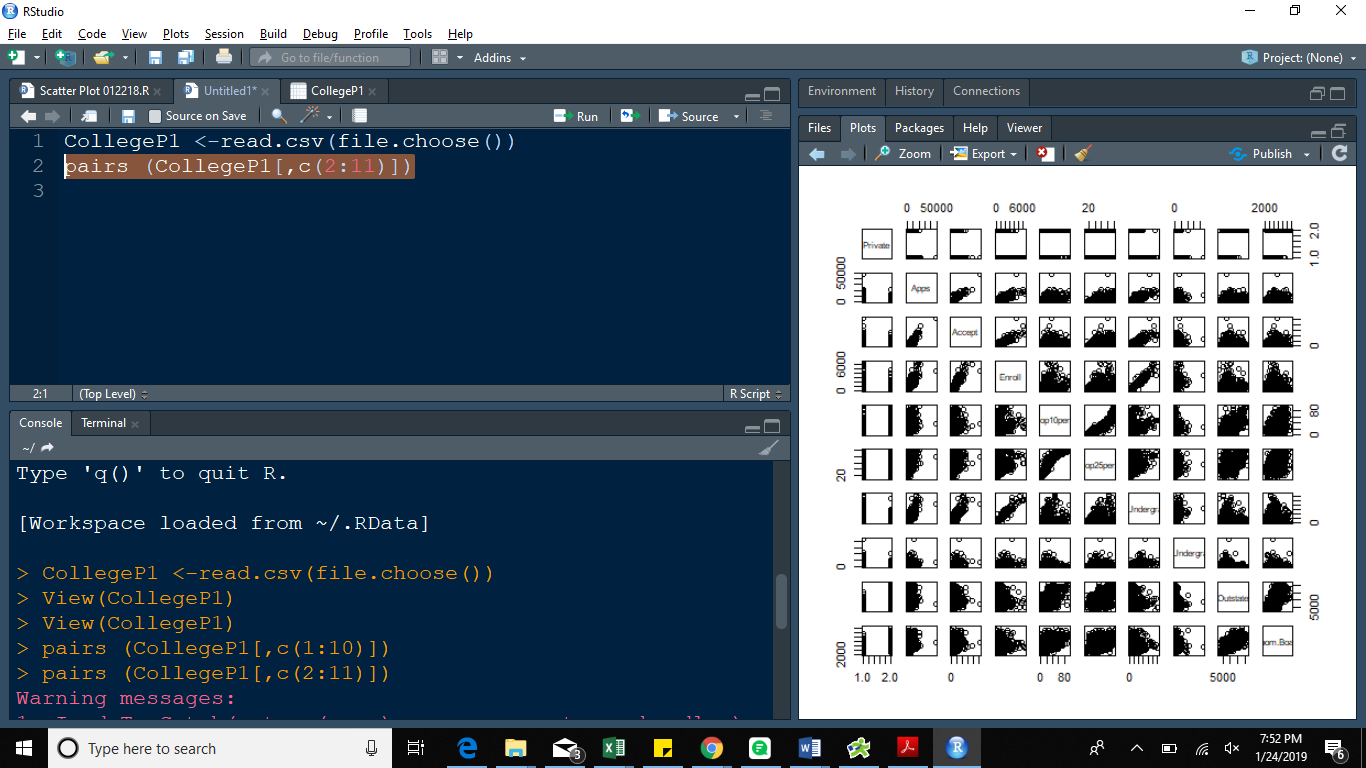
**CollegeP1 <-read.csv(file.choose())** #select the respective file from the folder

**View(CollegeP1)**

**pairs (CollegeP1[,c(1:10)]) #first 10 columns**

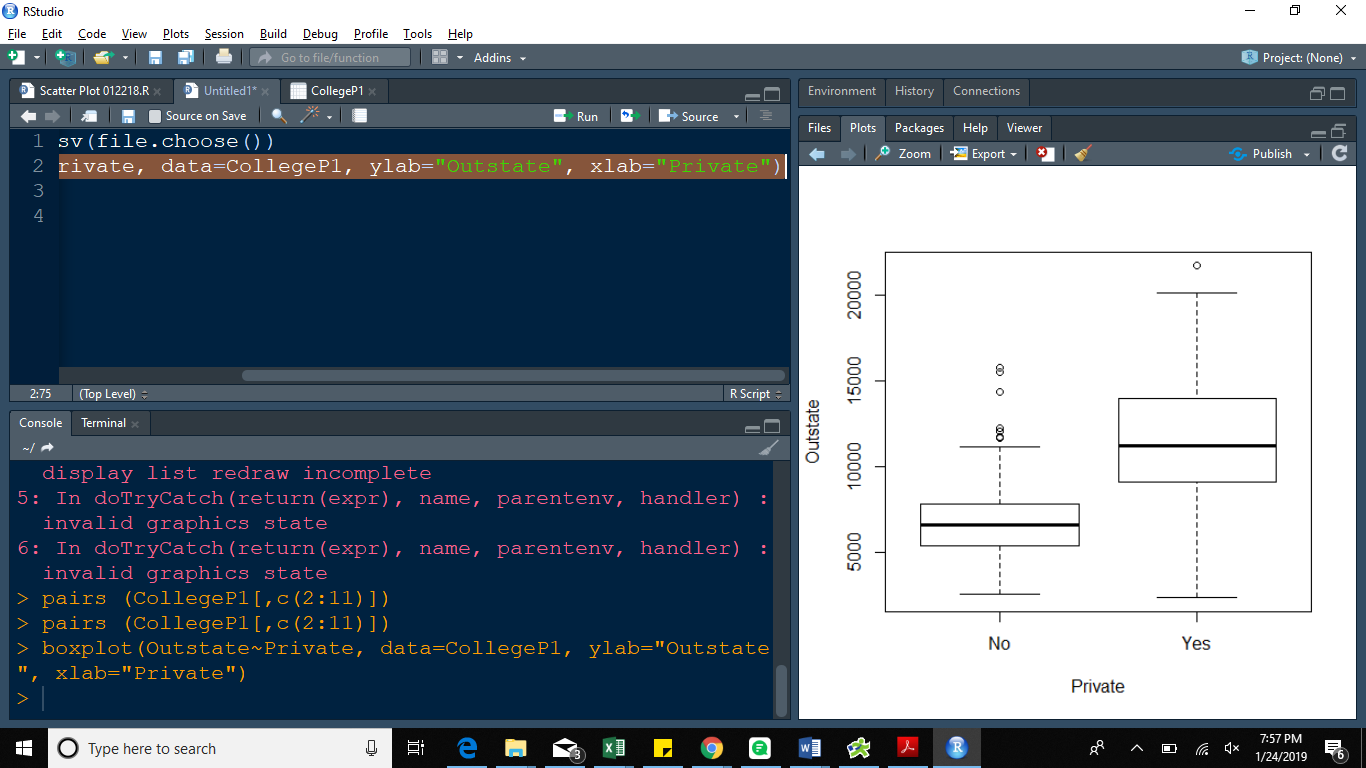


**pairs (CollegeP1[,c(2:11)]) #first 10 variables**



1. Use R functions to produce side-by-side boxplots of Outstate versus Private.

**boxplot(Outstate~Private, data=CollegeP1, ylab="Outstate", xlab="Private")**



1. Use R functions to produce histograms with differing numbers of bins for the first two quantitative variables.

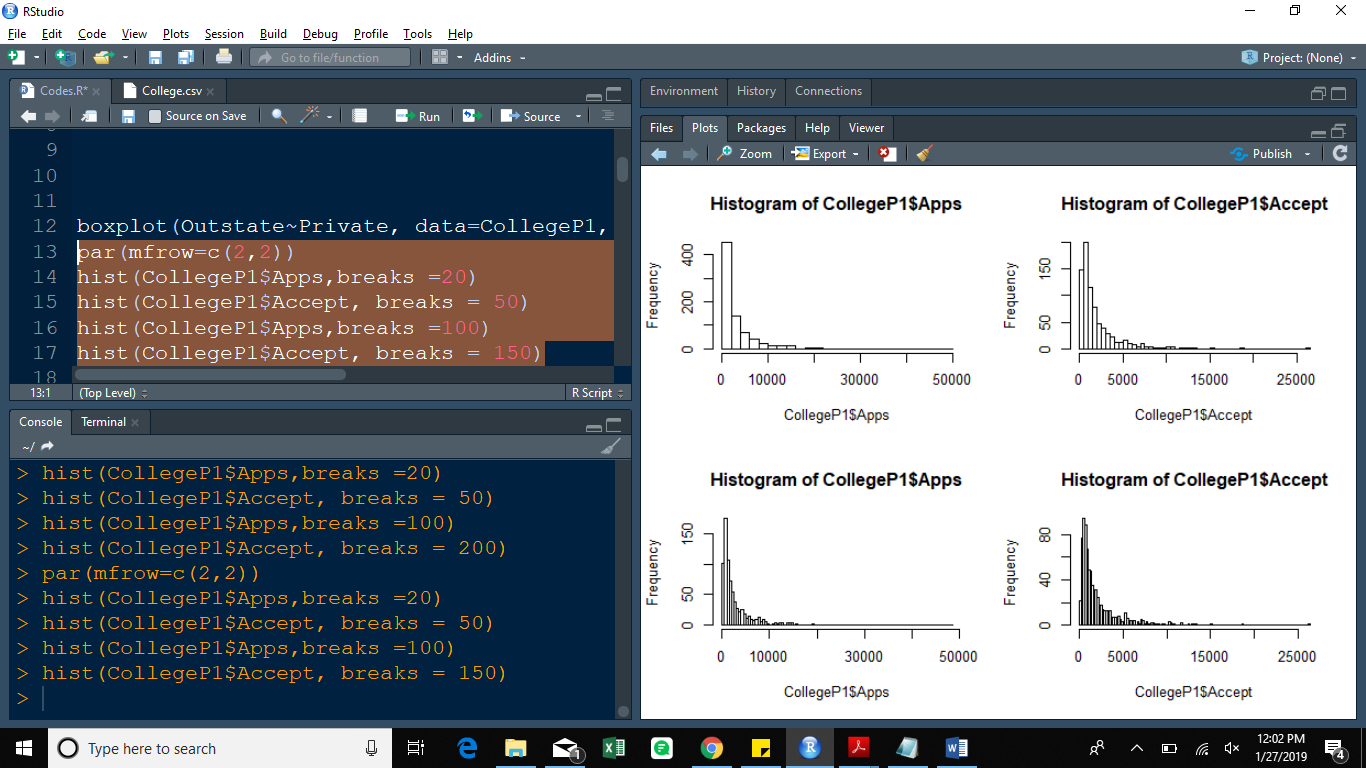
**par(mfrow=c(2,2))**

**hist(CollegeP1$Apps,breaks =20)**

**hist(CollegeP1$Accept, breaks = 50)**

**hist(CollegeP1$Apps,breaks =100)**

**hist(CollegeP1$Accept, breaks = 150)**



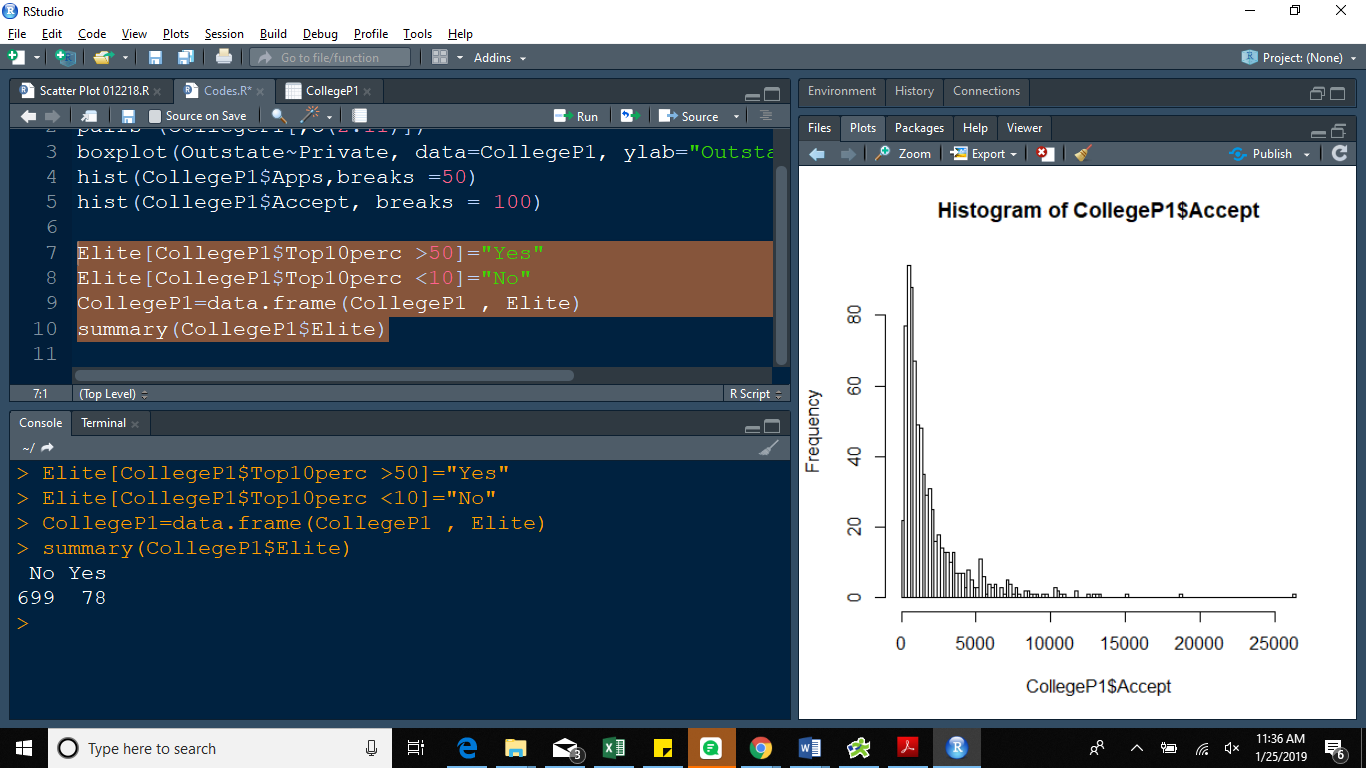
1. Create a new qualitative variable, called Elite, that is based on the Top10perc variable. The value of Elite is Yes if students coming from the top 10% of their high school classes exceeds 50%. Otherwise, the value is No

**Elite[CollegeP1$Top10perc >50]="Yes"**

**Elite[CollegeP1$Top10perc <10]="No"**

**CollegeP1=data.frame(CollegeP1 , Elite)**

**summary(CollegeP1$Elite) # this function lists the number of elite universities**



**Problem 2**

**Auto <-read.csv(file.choose())** #select the respective file from the folder

**View (Auto)**

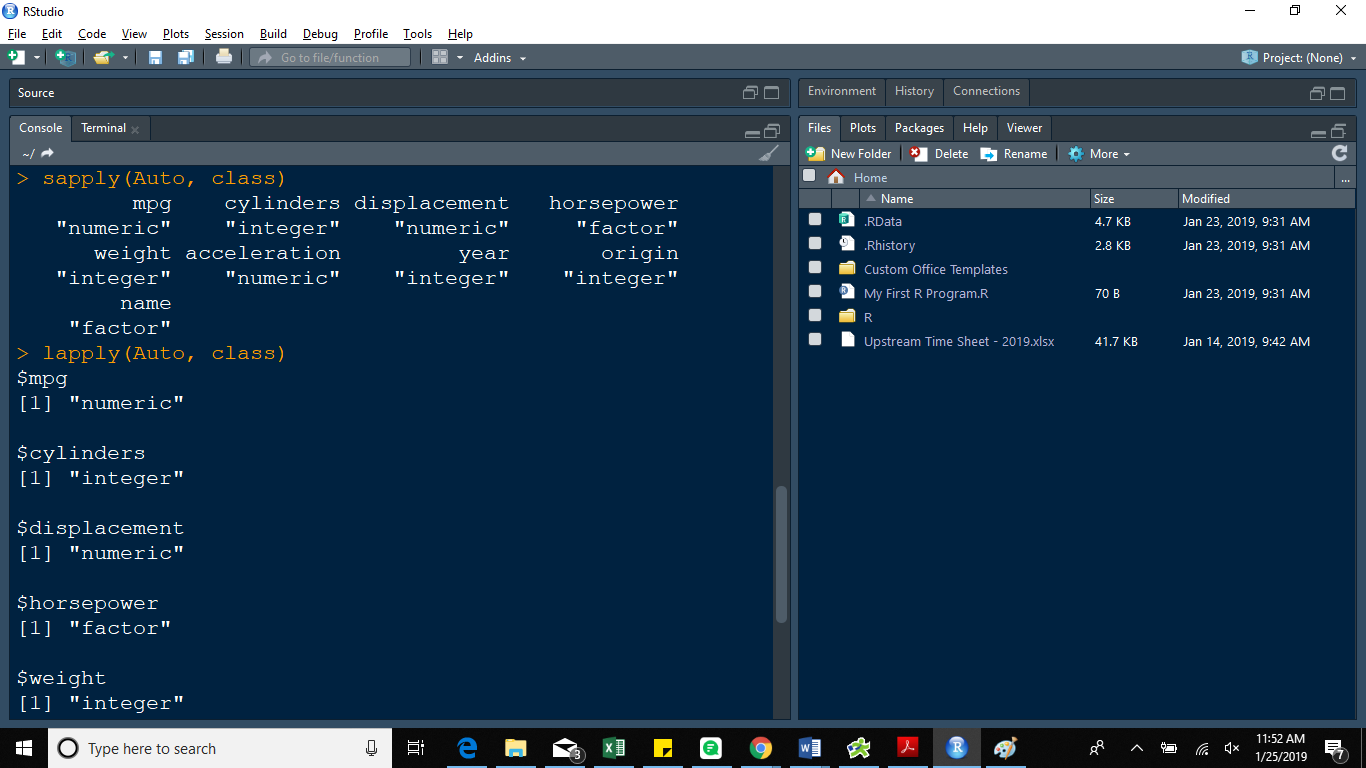
**Auto <-na.omit(Auto)**

1. Which of the predictors are quantitative, and which are qualitative?

**Here two functions can be used**

**sapply(Auto,class) # returns a vector, matrix or to simplify an array**

**lapply(Auto, class) # returns a list**



2. Use R functions to get the range of each quantitative predictor.

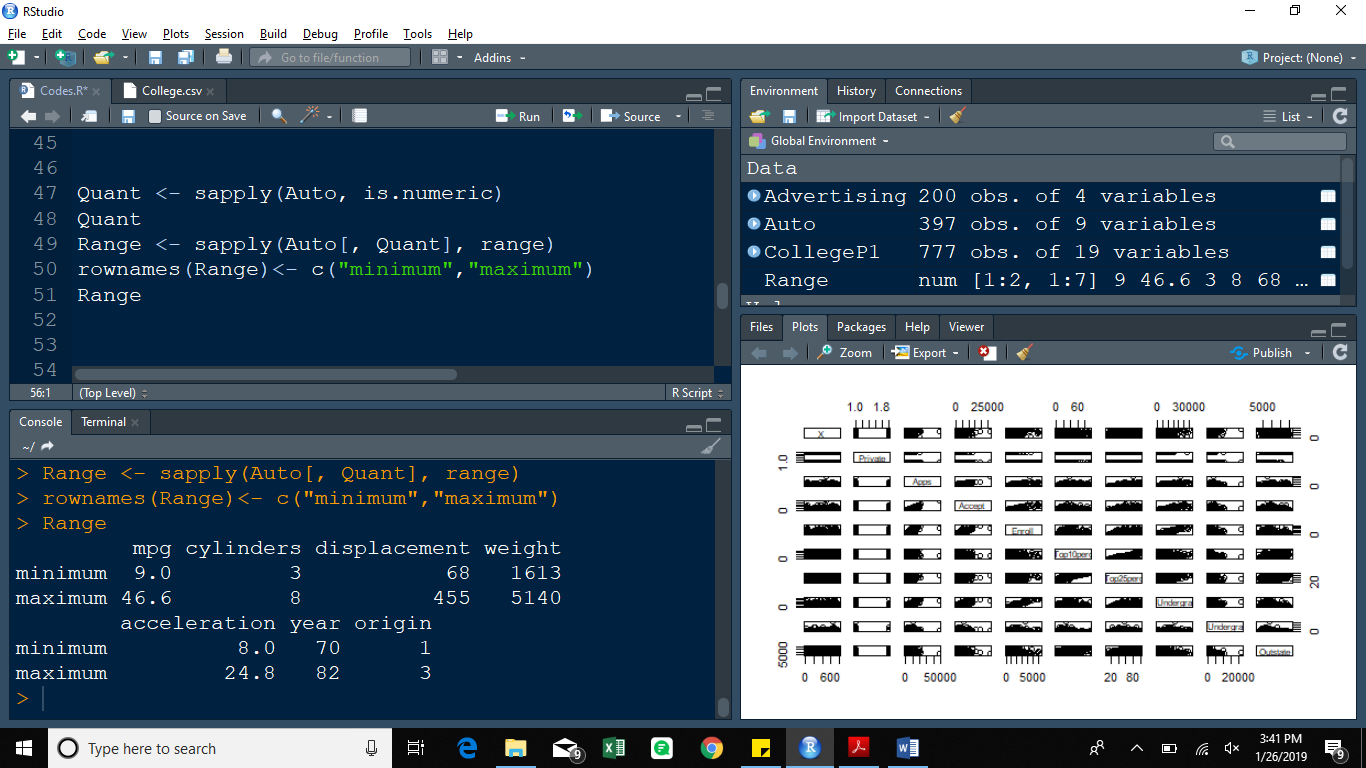
**Quant <- sapply(Auto, is.numeric)**

**Quant**

**Range<-sapply(Auto[, Quant], range)**

**rownames(Range)<- c("minimum","maximum")**

**Range**

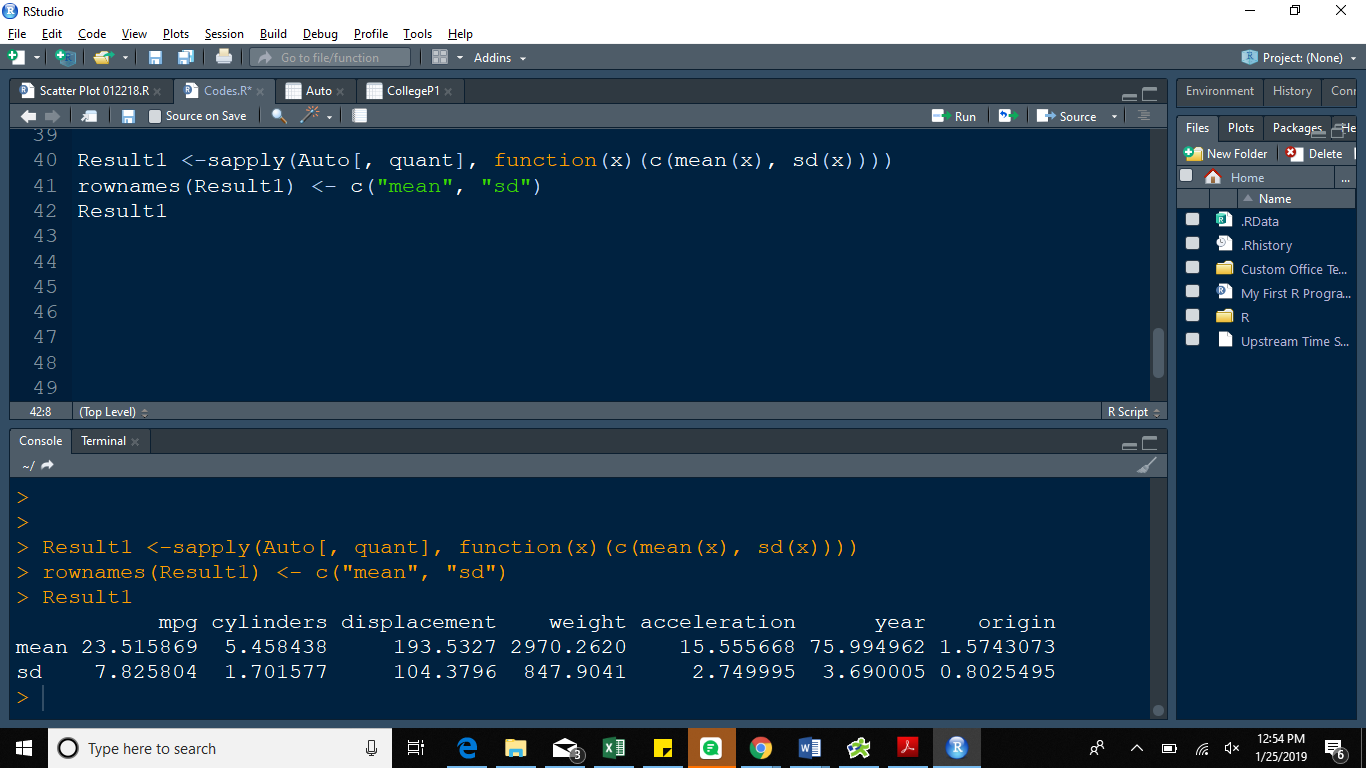


1. Use R functions to find the mean and standard deviation of each quantitative predictor.

**Result1 <-sapply(Auto[, quant], function(x)(c(mean(x), sd(x))))**

**rownames(Result1) <- c("mean", "sd")**

**Result1**



1. Now remove the 10th through 85th observations. What is the range, mean, and standard deviation of each predictor in the subset of the data that remains? Please use R functions to complete the task.

**Result2 <- sapply(Auto[-10:-85, quant], function(x)(c(range(x), mean(x), sd(x))))**

**rownames(Result2) <- c("minimum", "maximum", "mean", "sd")**

**Result2**

