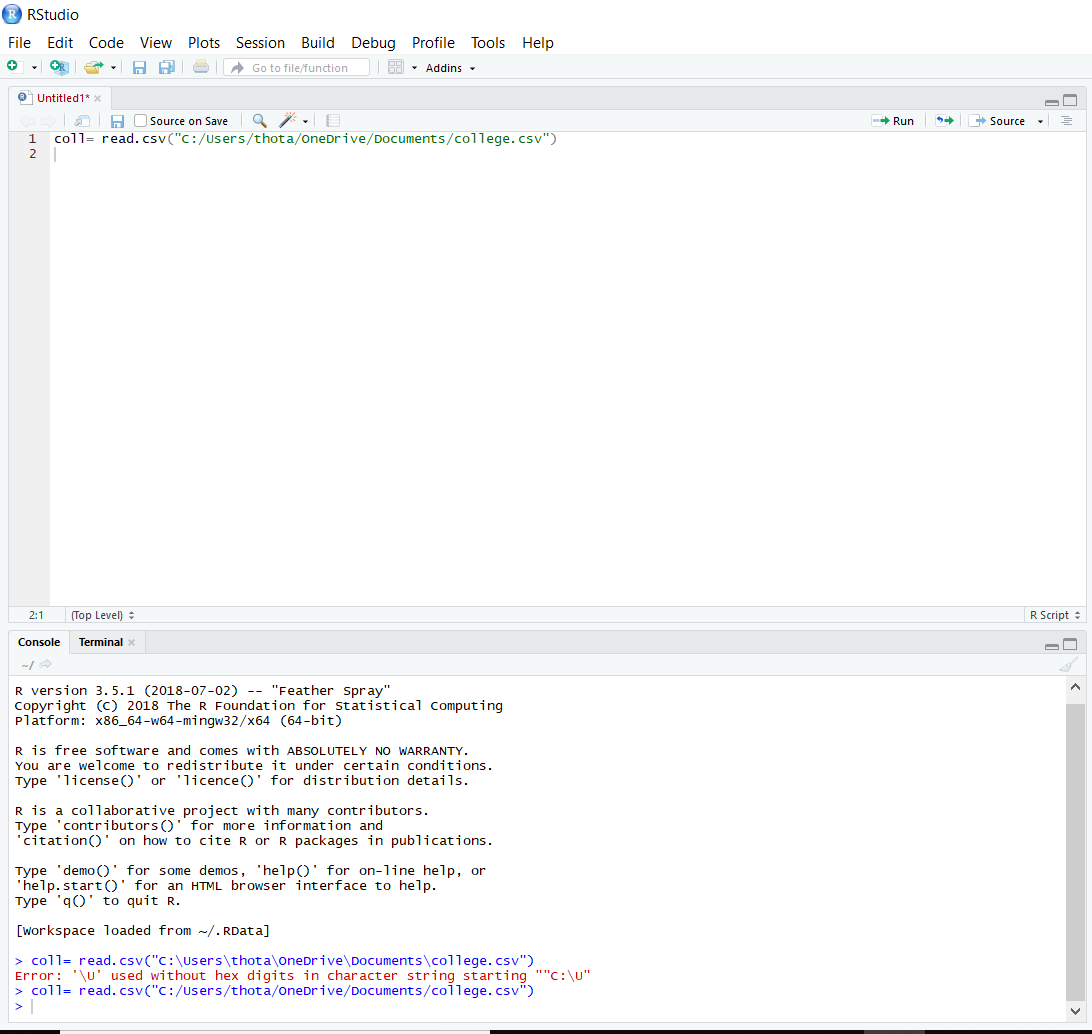
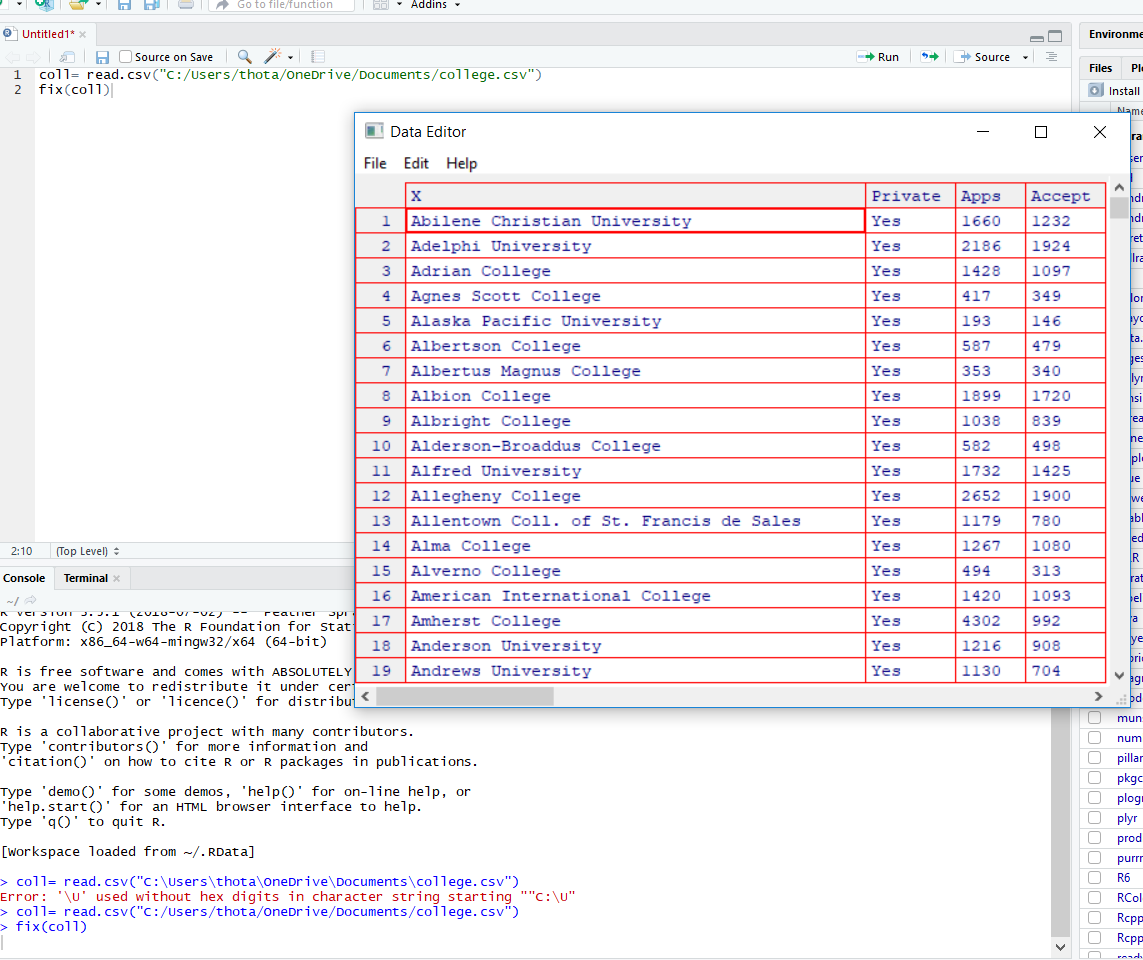
ISL LAB ASSIGNMENT-1

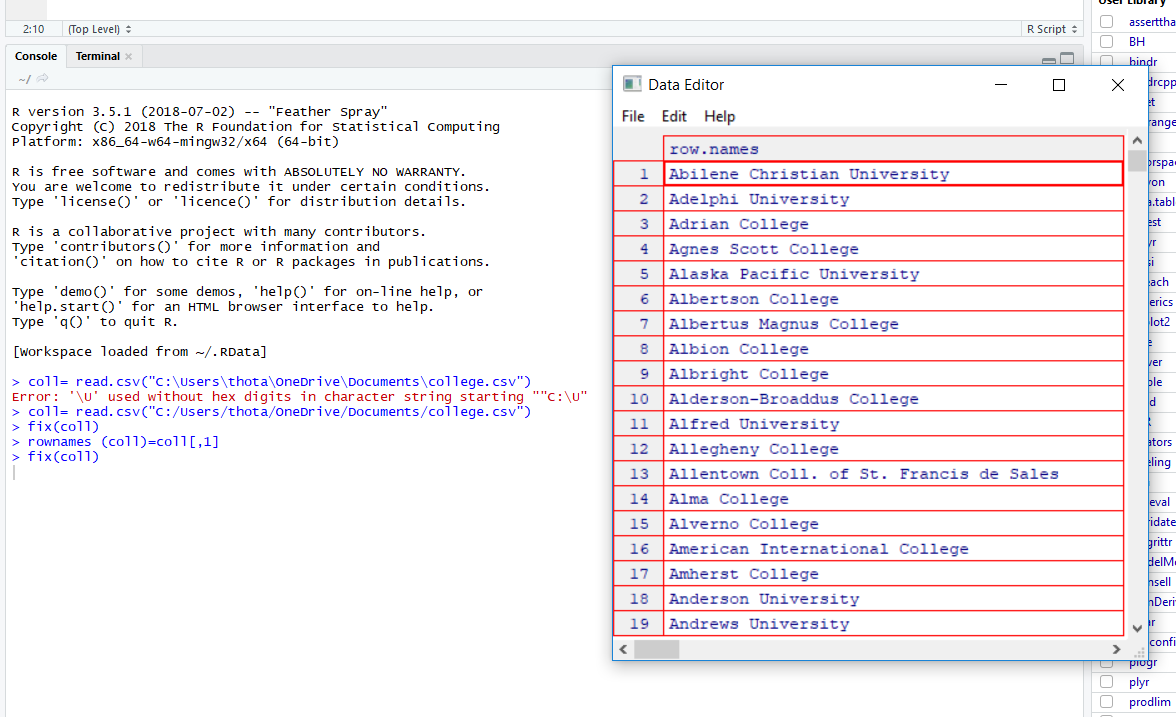
Name: Tejaswi Ayyadapu

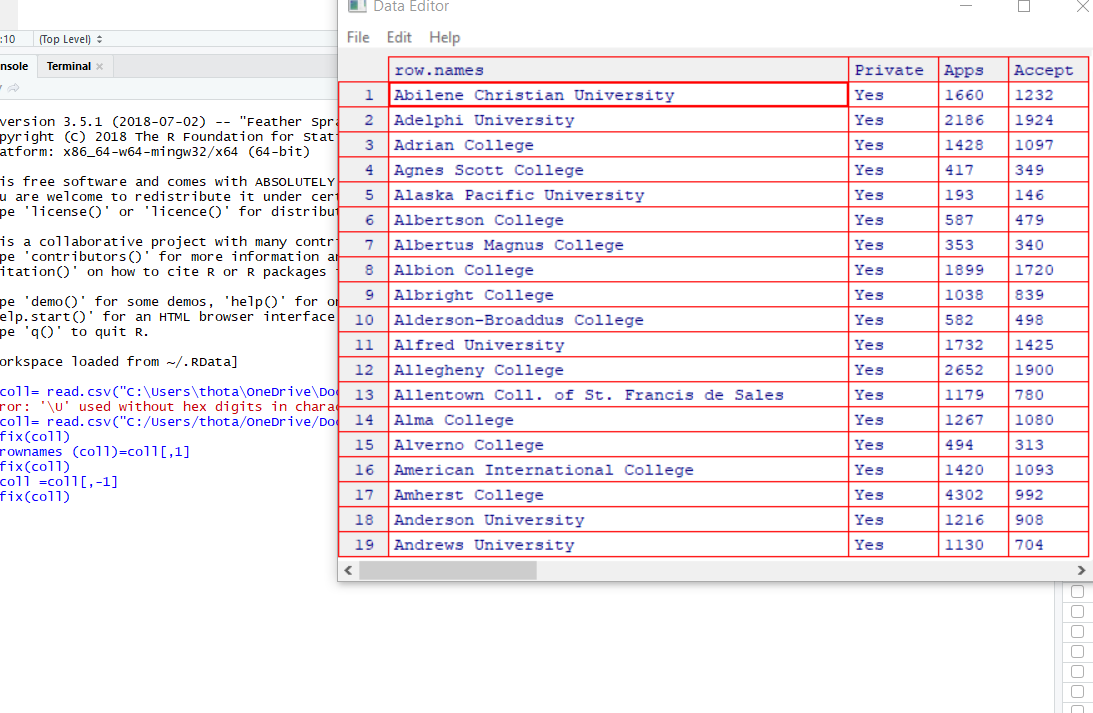
ID:1628799

1a) 

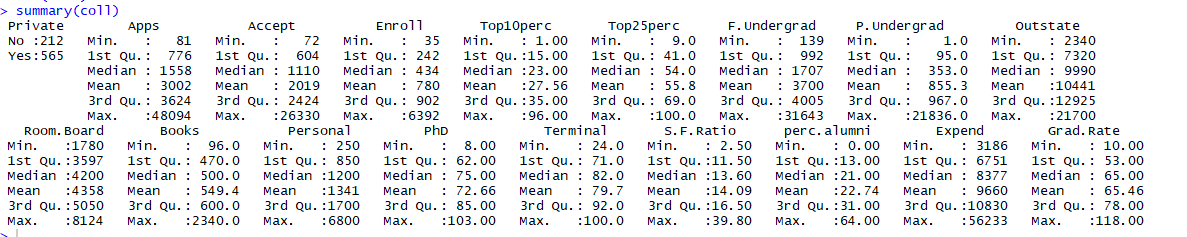
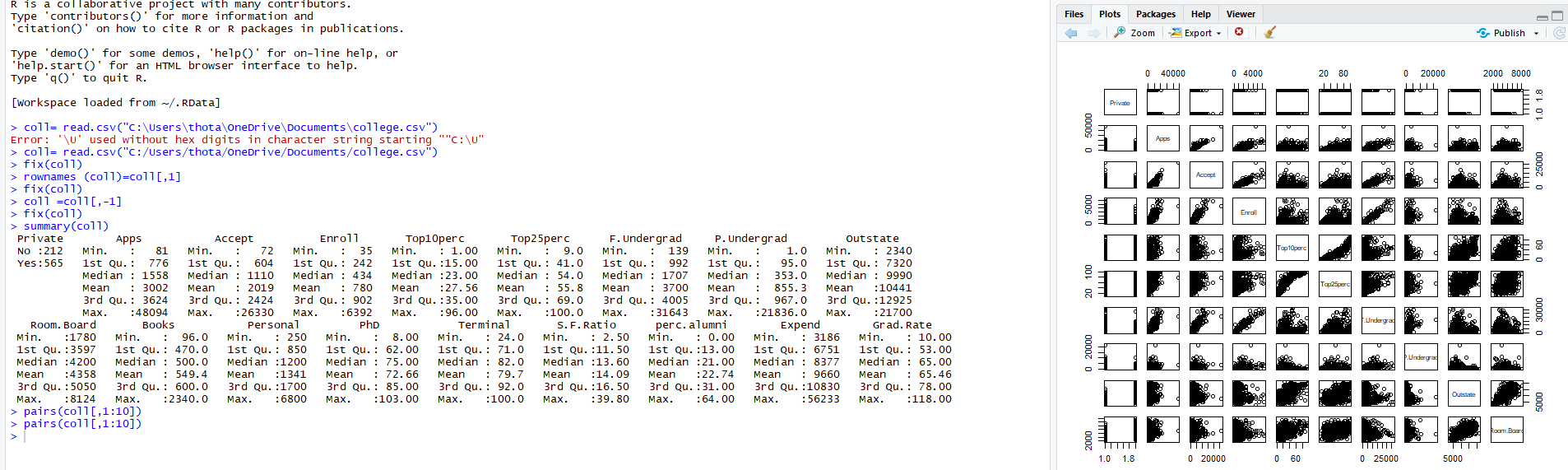
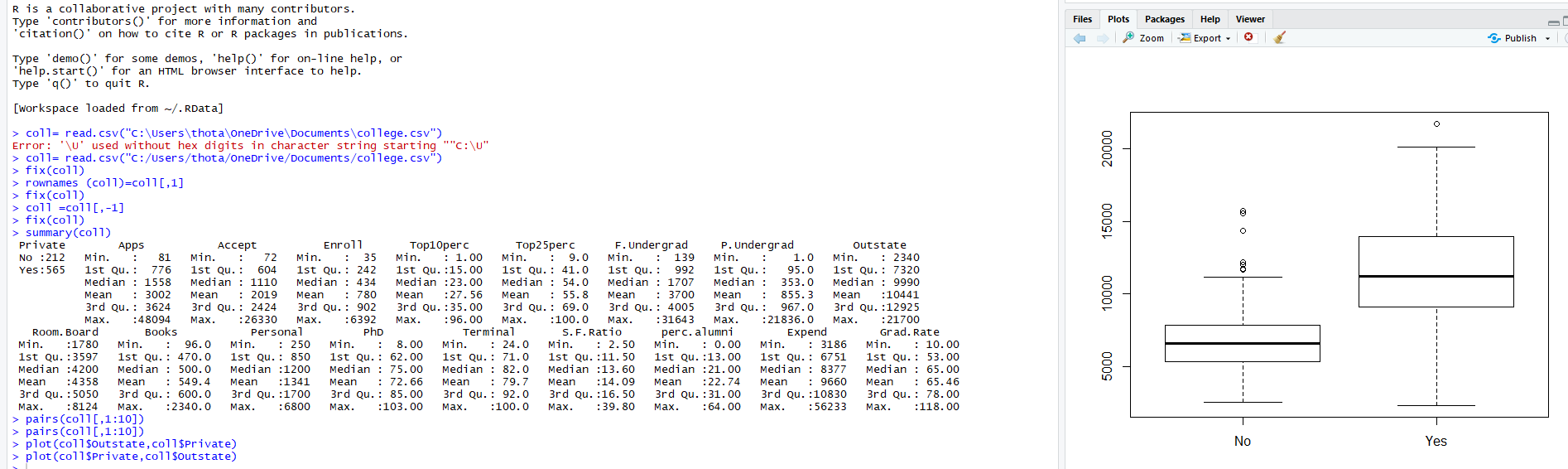
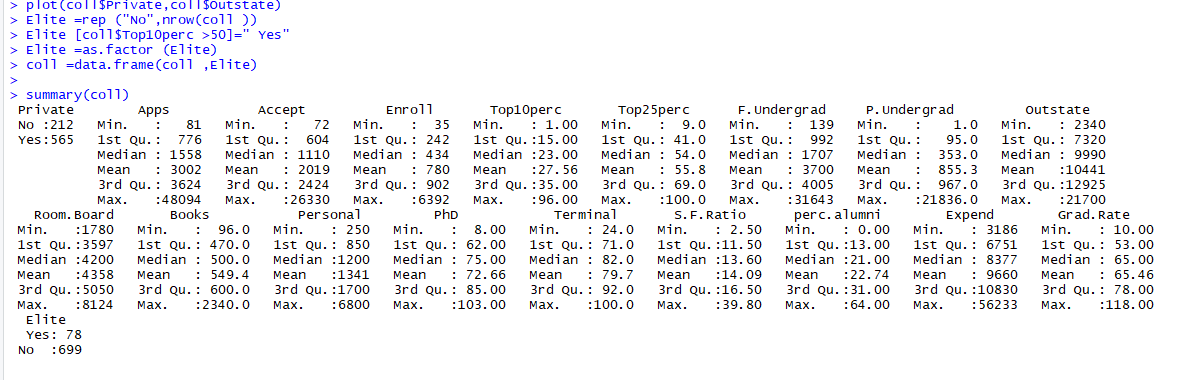
1b)

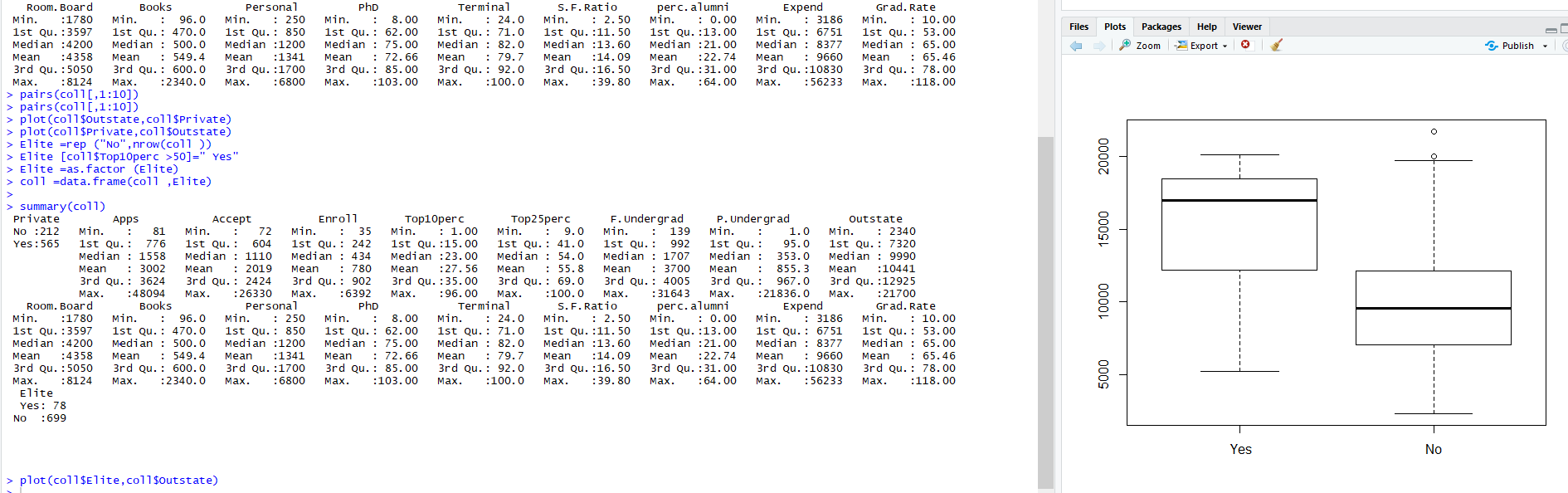


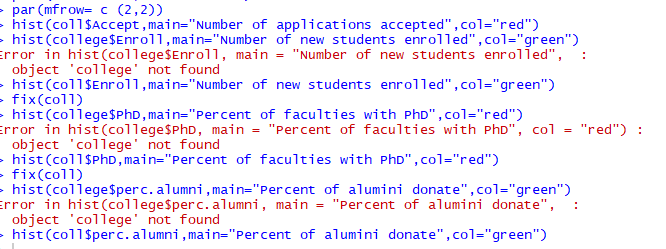


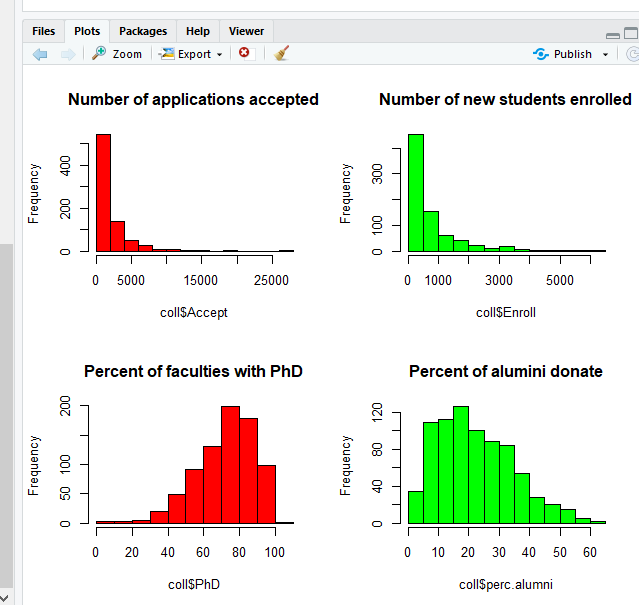


1c)

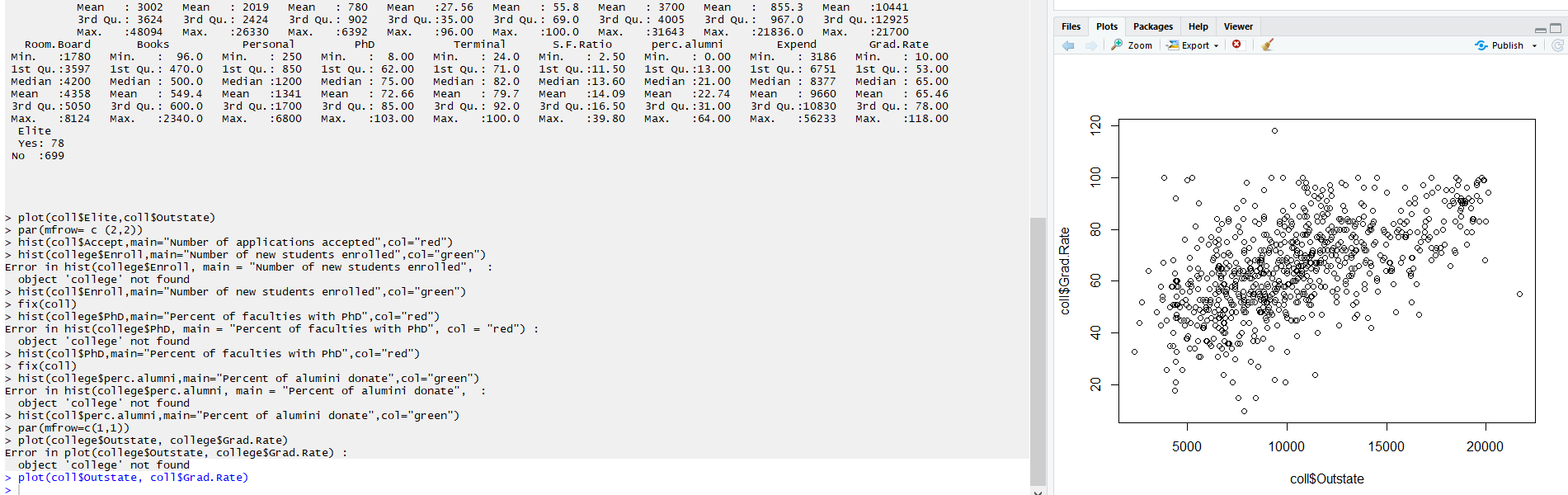
1. 
2. 
3. 
4. 



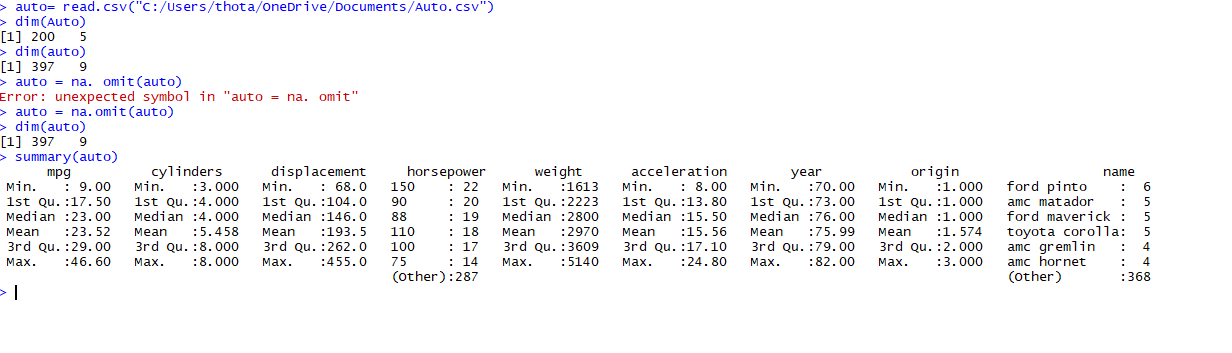




1. High tuition correlates to high graduation rate.



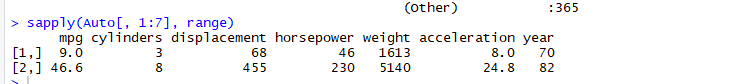
2)



1. Quantitative variables: mpg, cylinders, displacement, horsepower, weight, acceleration. (#7)

Qualitative variables: year, origin, name.

1. range of each quantitative predictor: **sapply(Auto[1:7,range])**

****

1. mean and standard deviation of each quantitative predictor

**sapply(Auto[1:7,mean])**

**sapply(Auto[1:7,sd])**

**A screen shot of a social media post

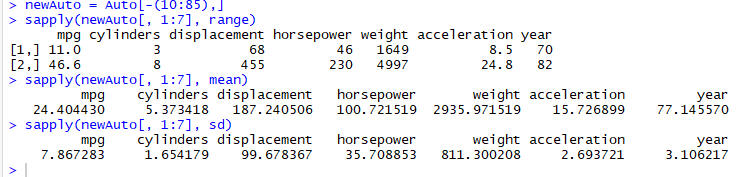
Description generated with very high confidence**

1. **newAuto = Auto[-(10:85),]** /\*remove the 10th through 85th observations.

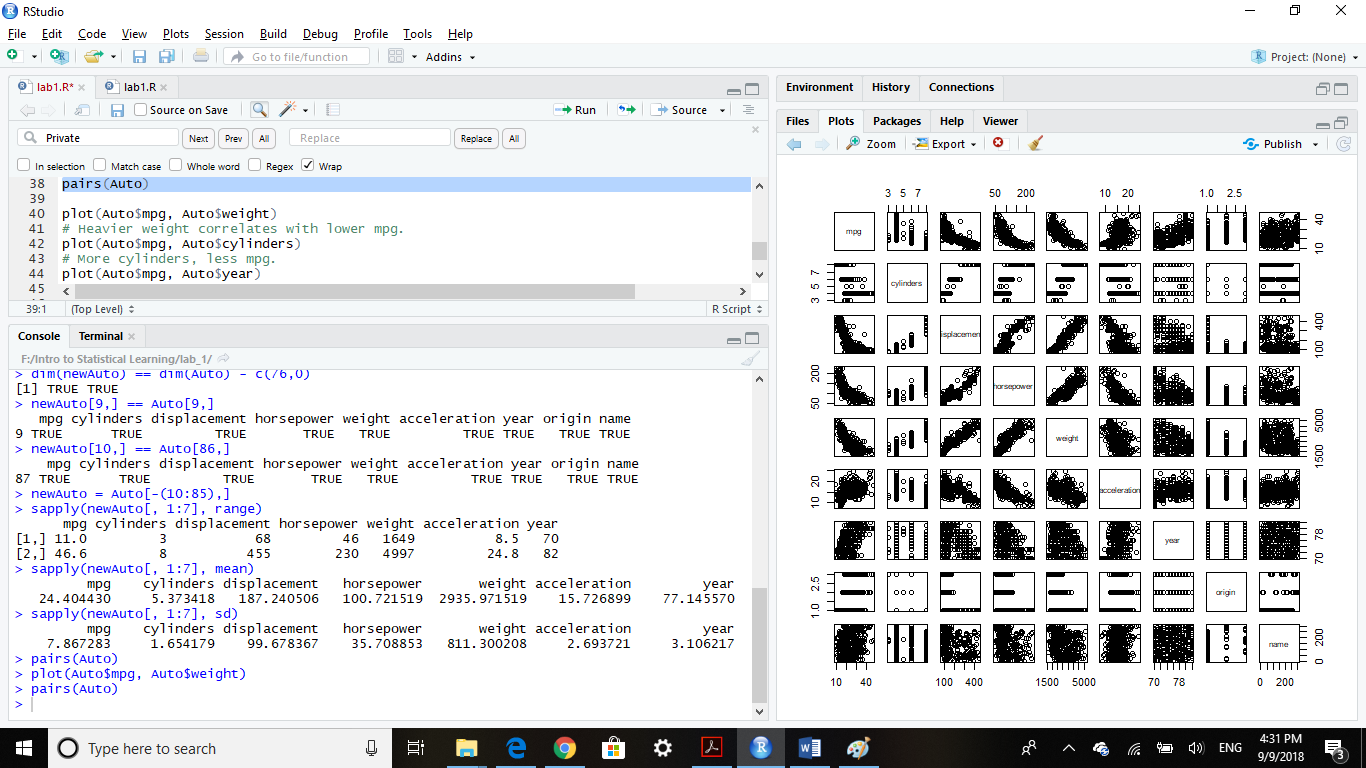
**sapply(Auto[,1:7],range)**

**sapply(Auto[,1:7],mean)**

**sapply(Auto[,1:7],sd)**

****

1. **pairs(Auto)**

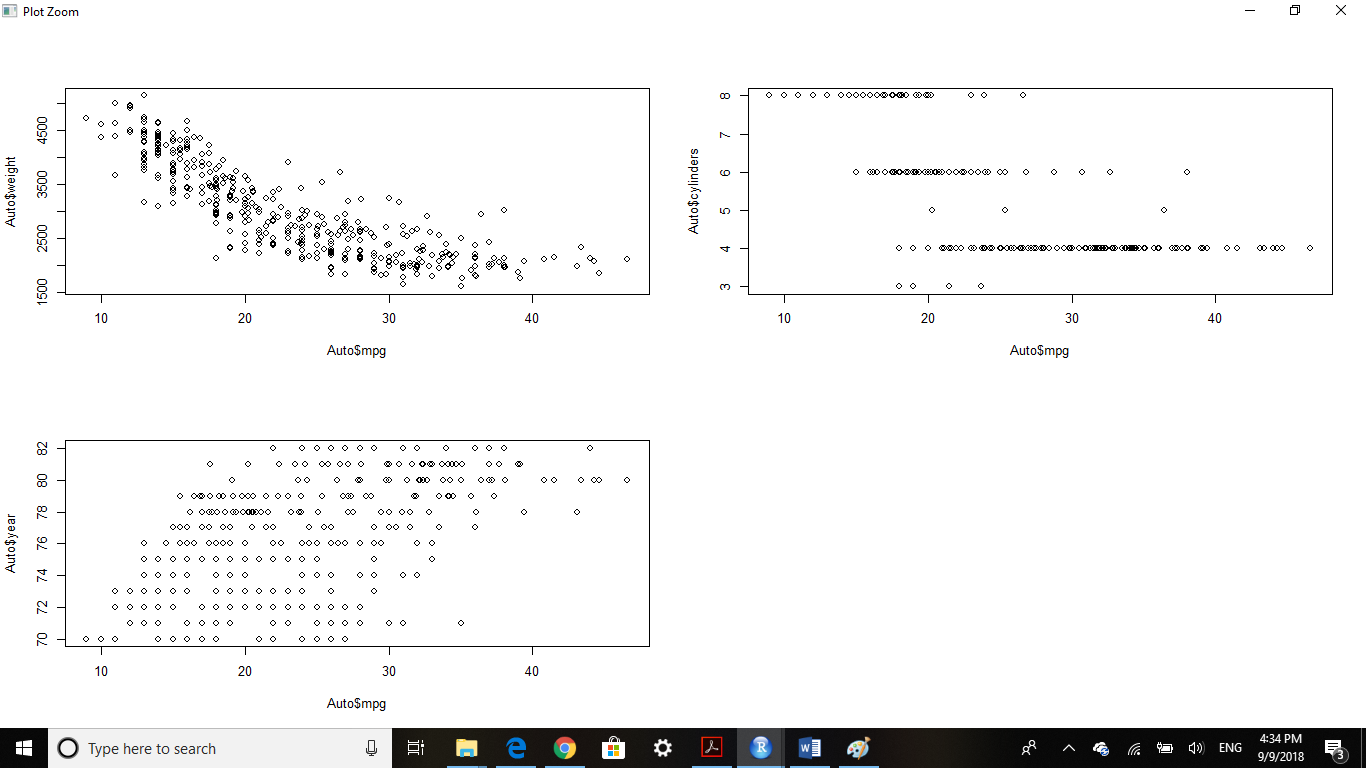


**pairs(Auto)**

**plot(Auto$mpg, Auto$weight) # Heavier weight correlates with lower mpg.**

**plot(Auto$mpg, Auto$cylinders) # More cylinders, less mpg.**

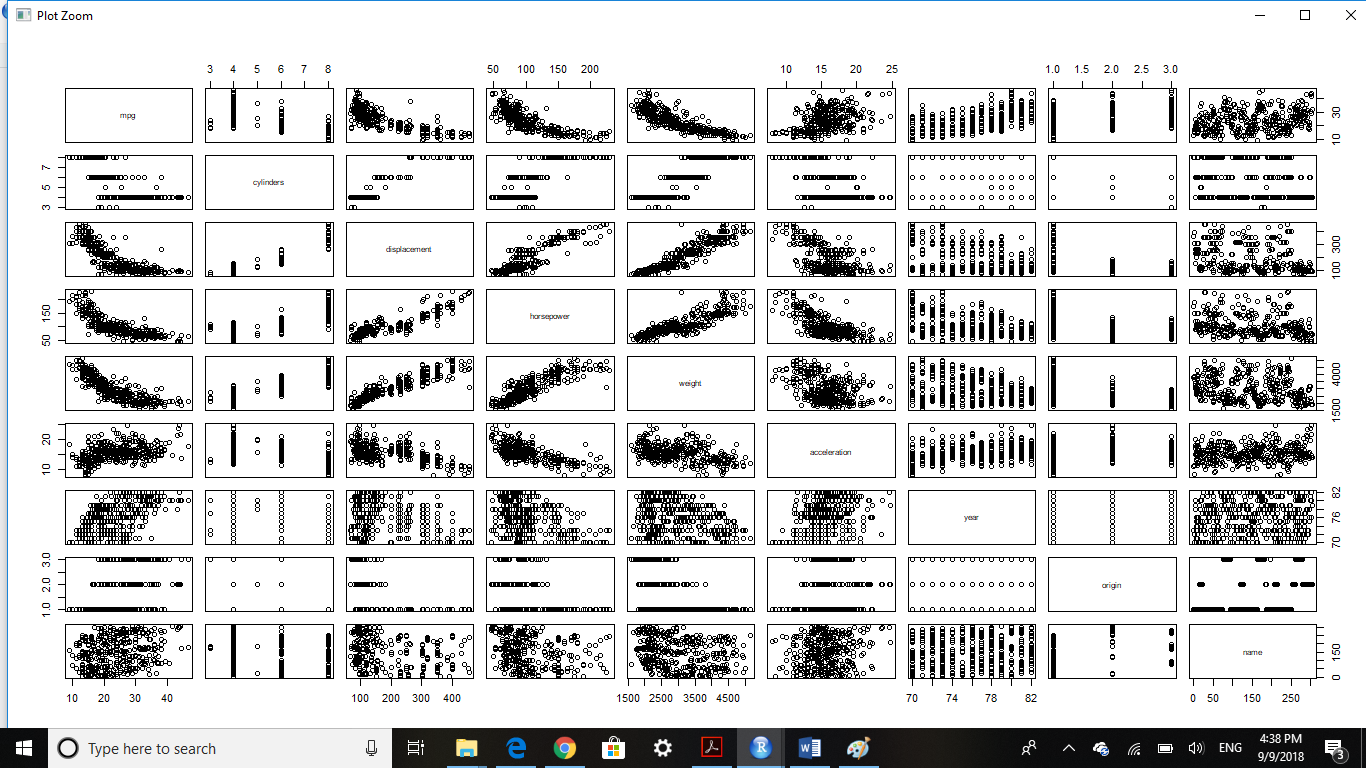
**plot(Auto$mpg, Auto$year) # Cars become more efficient over time.**



1. **pairs(Auto)**

**See descriptions of plots in (e).**

* **All of the predictors show some correlation with mpg. The name predictor has too little observations per name though, so using this as a predictor is likely to result in overfitting the data and will not generalize well.**



**3.**

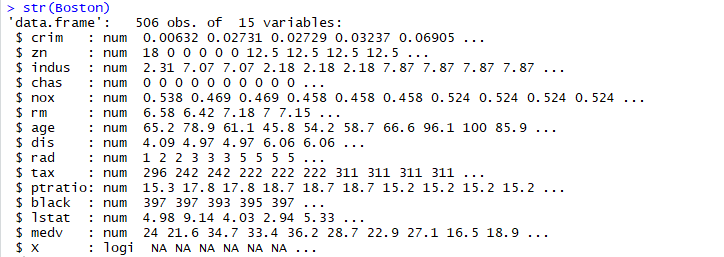
**a. >> Boston = read.csv("F:/Intro to Statistical Learning/lab\_1/data/Bostondata.csv")**

**>> ? Boston #to read about the data set**

**>> Dimensions of Boston:** dim(Boston): [1] 506 15

**>>** 14 features, 506 housing values in Boston data frame.

**>> str(Boston)**

****

**b.>> pairs(Boston)**

# X correlates with: a, b, c

# crim: age, dis, rad, tax, ptratio

# zn: indus, nox, age, lstat

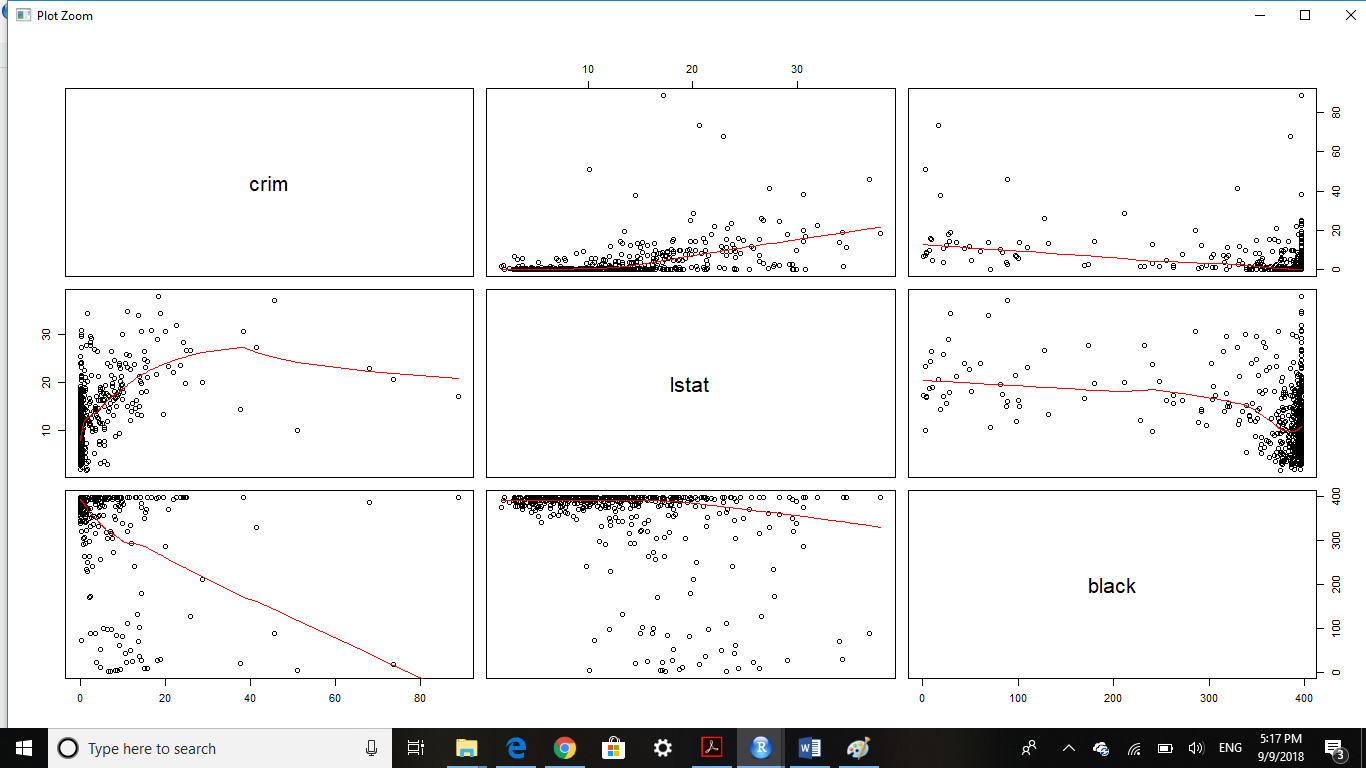
# indus: age, dis

# nox: age, dis

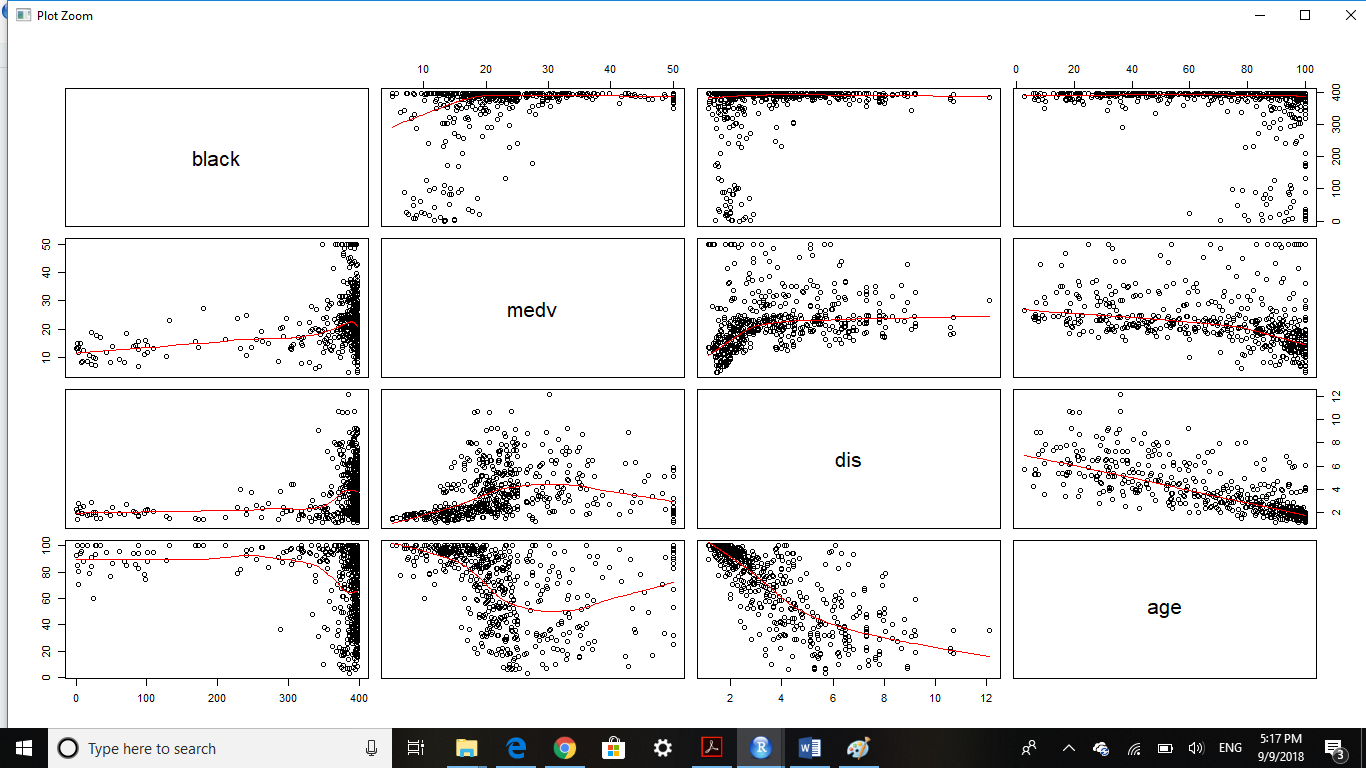
# dis: lstat

# lstat: medv

**>> pairs(~ crim + lstat + black, data = Boston, panel = panel.smooth)**



**>> pairs(~ black + medv + dis + age , data = Boston, panel = panel.smooth)**



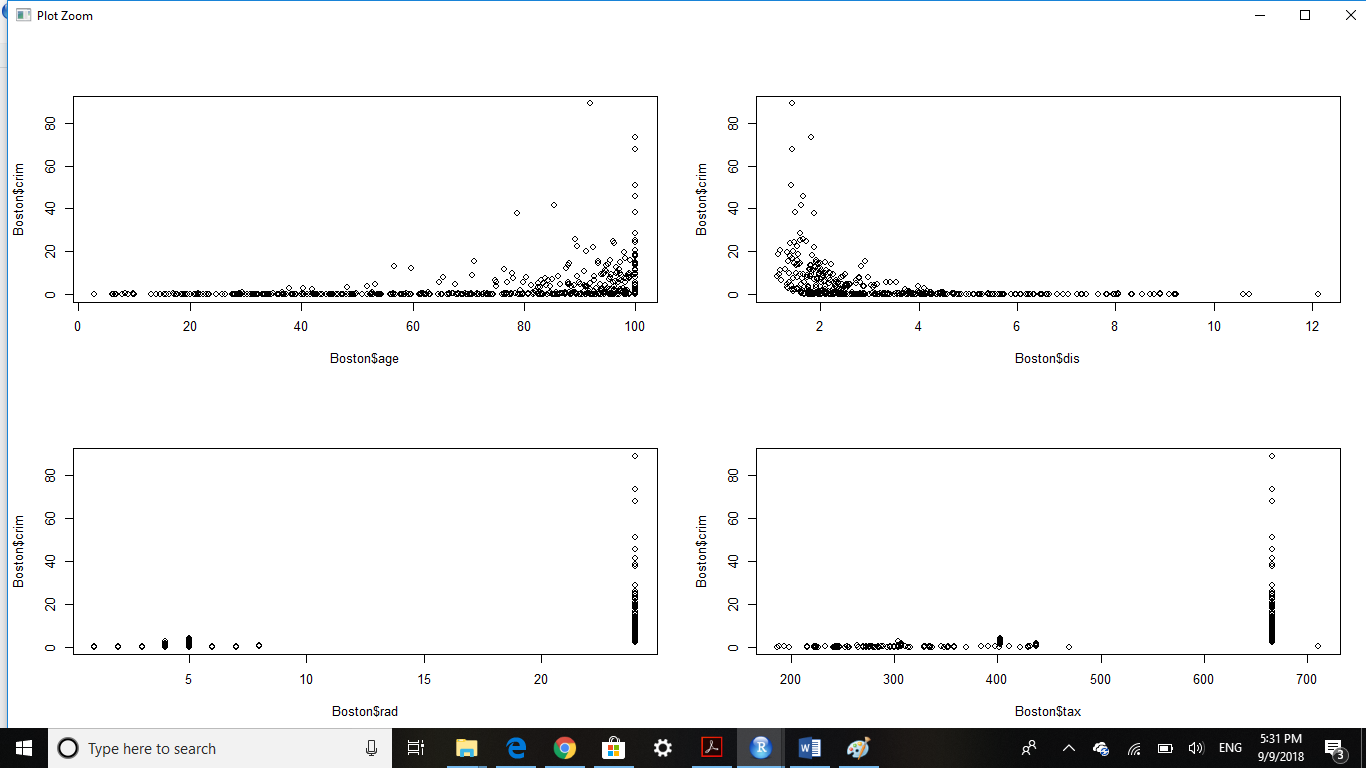
1. **plot(Boston$age, Boston$crim) # Older homes, more crime**

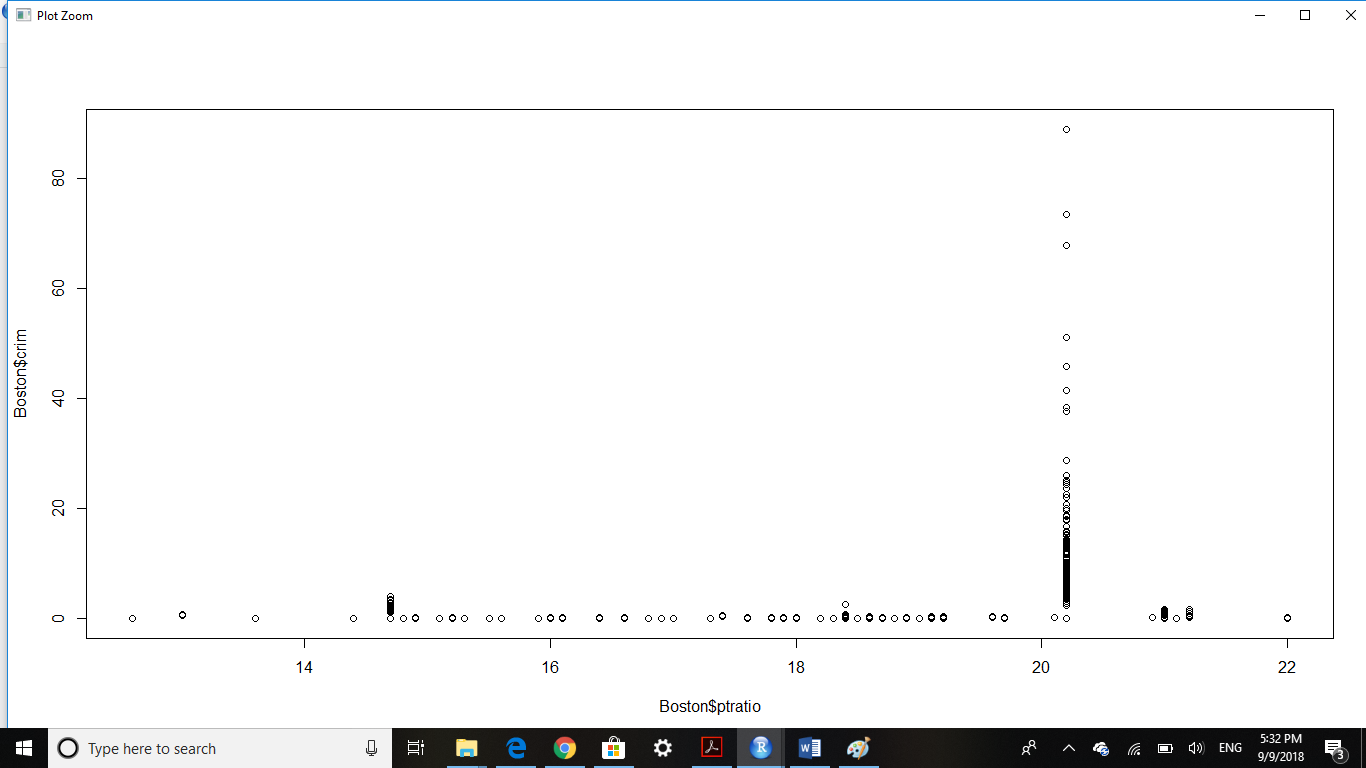
**plot(Boston$dis, Boston$crim) # Closer to work-area, more crime**

**plot(Boston$rad, Boston$crim) # Higher index of accessibility to radial highways, more crime**

**plot(Boston$tax, Boston$crim) # Higher tax rate, more crime**

**plot(Boston$ptratio, Boston$crim) # Higher pupil: teacher ratio, more crime**



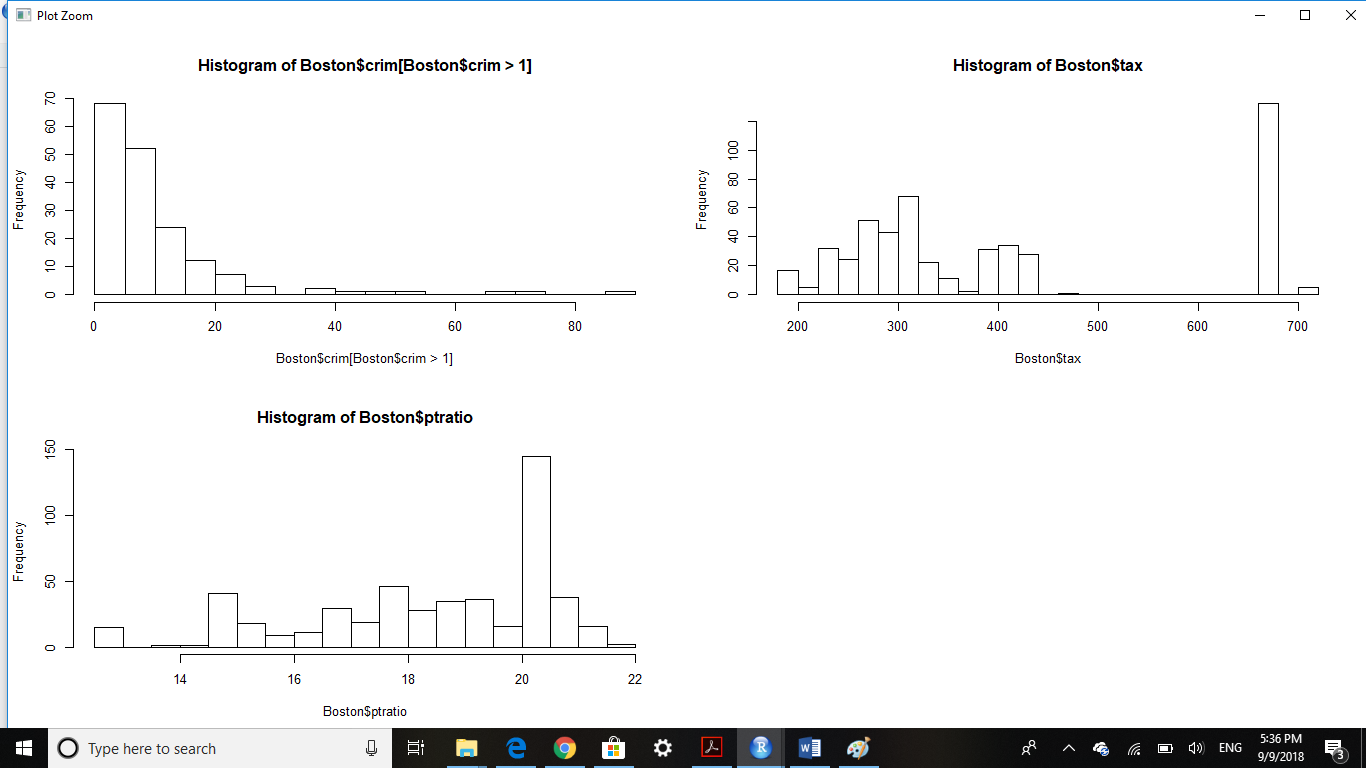


1. **>> par(mfrow=c(2,2))**

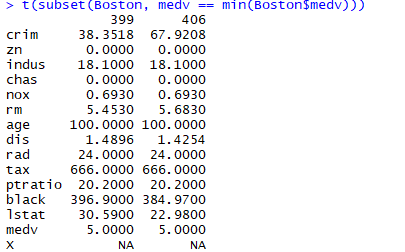
**>> hist(Boston$crim[Boston$crim>1], breaks=25) # most cities have low crime rates, but there is a long tail: 18 suburbs appear to have a crime rate > 20, reaching to above 80**

**>> hist(Boston$tax, breaks=25) # there is a large divide between suburbs with low tax rates a peak at 660-680**

**>> hist(Boston$ptratio, breaks=25) # a skew towards high ratios, but no particularly high ratios**



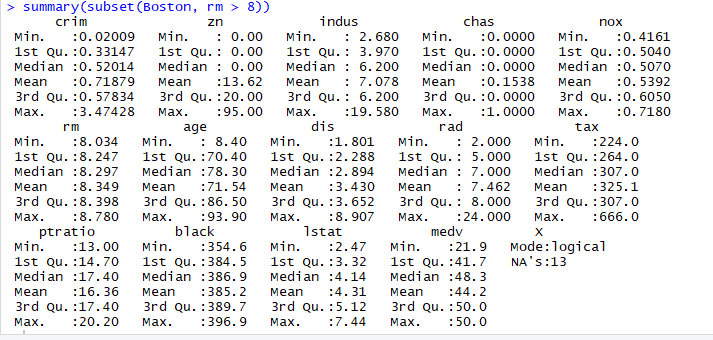
1. **dim(subset(Boston, chas == 1)) -> 35 suburbs**
2. **median(Boston$ptratio) -> median pupil-teacher ratio: 19.05**
3. **t(subset(Boston, medv == min(Boston$medv)))**

****

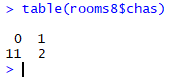
1. **>> table(Boston$rm) # 64 houses with more than 7 rooms**
2. **>> table(Boston$rm) # 13 houses with more than 7 rooms**

**>> rooms8 = (subset(Boston, rm > 8))**

**Summary(rooms8)**

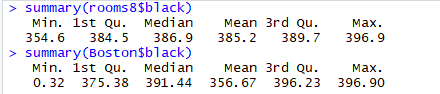
****

**>> table(rooms8$cahs) #** *11 of the houses with 8 rooms are not near Charles river (only 2 are near Charles river*

****

**>> summary(rooms8$black)**

**>> summary(Boston$black)**

****

**All the rooms8 houses black’s population**