HOME WORK ONE Qian Yu

ST635 Intermediate Statistical Modeling for Business Fall 2017

Section 2.4:  Problem 9, Problem 10

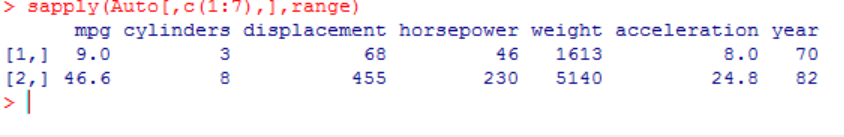
9. This exercise involves the Auto data set studied in the lab. Make sure that the missing values have been removed from the data.

(a) Which of the predictors are quantitative, and which are qualitative?

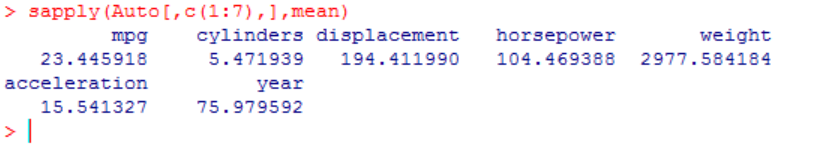
**The webpage has definition of each predictors (> ? Auto), since quantitative variables take on numerical values, In contrast, qualitative variables take on values in one of K different classes, or categories.**

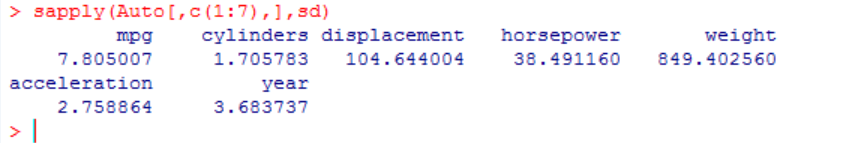
**Therefore, origin is qualitative due to description (1. American, 2. European, 3. Japanese); name is also qualitative due to its type is ‘factor’ not ‘num’(> str(Auto)); other seven predictors are quantitative.**

(b) What is the range of each quantitative predictor? You can answer this using the range() function. range()

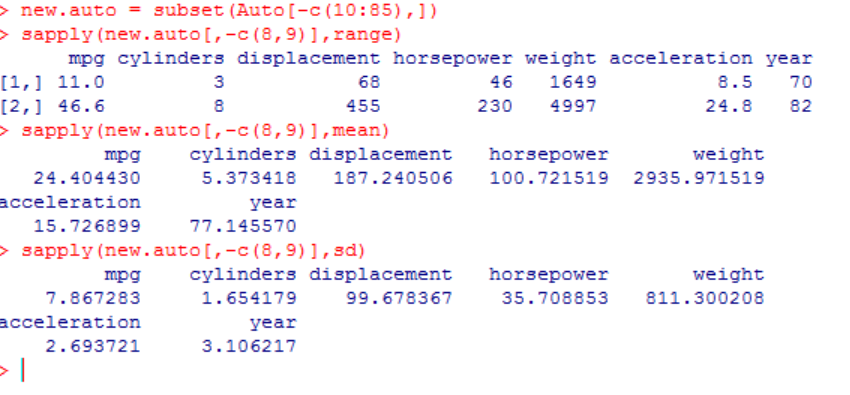


(c) What is the mean and standard deviation of each quantitative predictor?





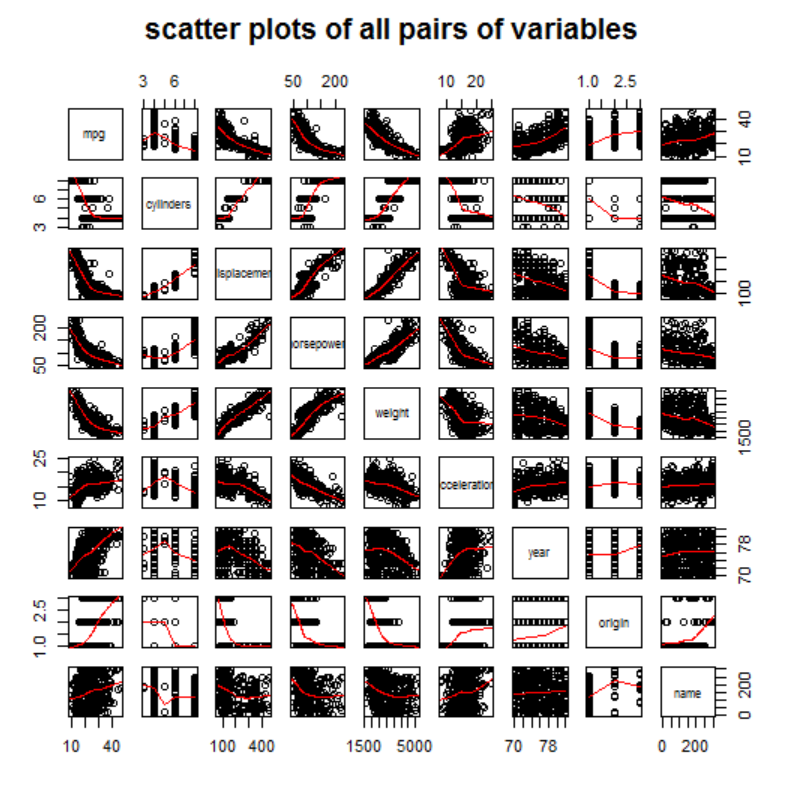
(d) Now remove the 10th through 85th observations. What is the range, mean, and standard deviation of each predictor in the subset of the data that remains?



(e) Using the full data set, investigate the predictors graphically, using scatterplots or other tools of your choice. Create some plots highlighting the relationships among the predictors. Comment on your findings.

**When miles per gallon goes up the inch of Engine displacement goes down. The Engine horsepower, Vehicle weight also have negative relationship with mile per gallon.**

**For Vehicle name, it is vary related with other predictors, except cylinders, the common type of cylinders is 4, 6, and 8.**



(f) Suppose that we wish to predict gas mileage (mpg) on the basis of the other variables. Do your plots suggest that any of the other variables might be useful in predicting mpg? Justify your answer.

**Yes, origin also would be decent predictors of mpg, besides what the answer for (e).**

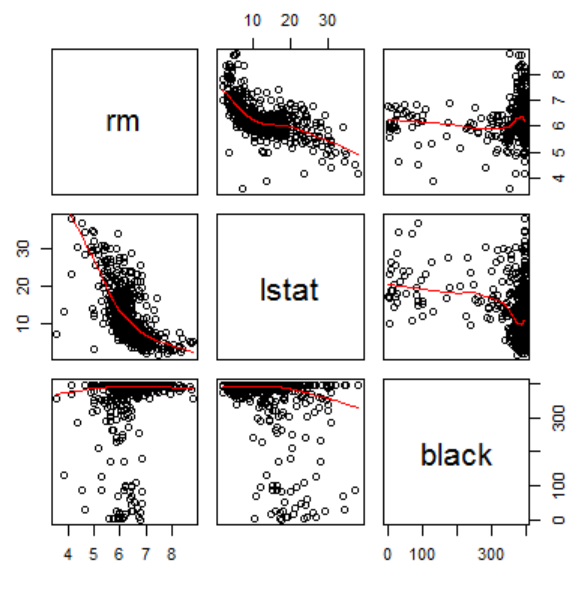
**10. This exercise involves the Boston housing data set.**

(a) To begin, load in the Boston data set. The Boston data set is part of the MASS library in R. > library(MASS) Now the data set is contained in the object Boston. > Boston Read about the data set: > ?Boston How many rows are in this data set? How many columns? What do the rows and columns represent?

**The Boston data frame has 506 rows and 14 columns. ( > ? Boston)**

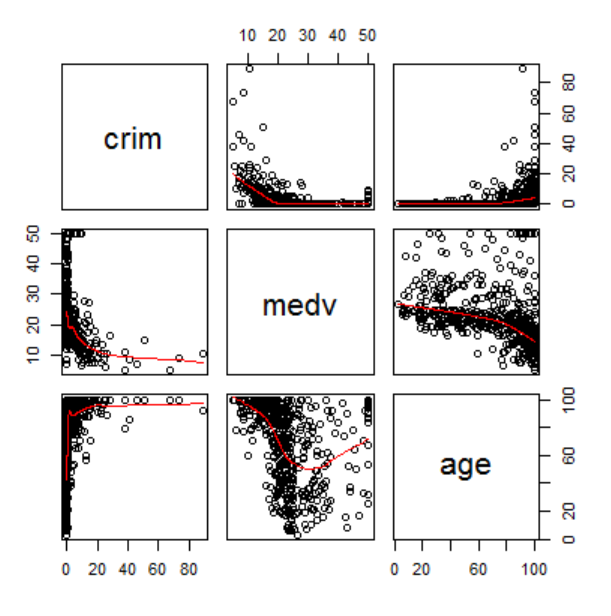
**Columns represent the variables that might effecting house values in suburbs of Boston, while and rows are observations of every houses in the sample of suburbs of Boston.**

(b) Make some pairwise scatterplots of the predictors (columns) in this data set. Describe your findings.



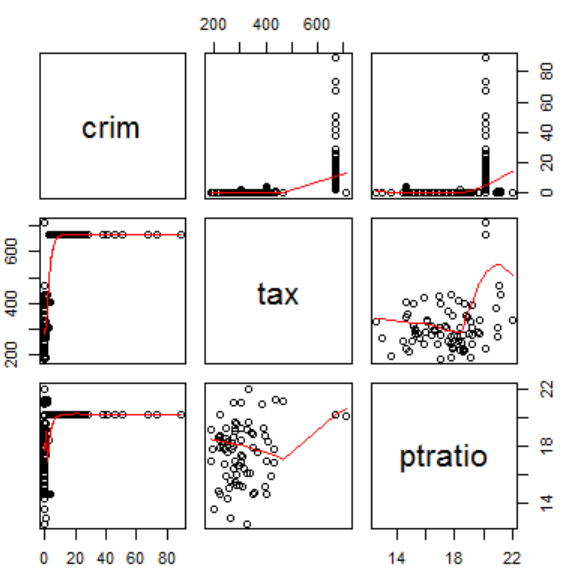
**The more of average number of rooms per dwelling, the less of lower status of the population**

(c) Are any of the predictors associated with per capita crime rate? If so, explain the relationship.



**Per capita crime rate by town goes down and stay very low when median value of owner-occupied homes is higher than 20,000.**

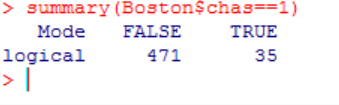
(d) Do any of the suburbs of Boston appear to have particularly high crime rates? Tax rates? Pupil-teacher ratios? Comment on the range of each predictor.



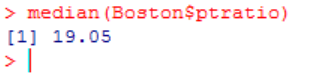
**Per capita crime rate by town keeps in low until the tax full-value property-tax rate per \$10,000 more than 650; per capita crime rate by town keeps in low except when pupil-teacher ratio by town is exactly equal to 20.**

(e) How many of the suburbs in this data set bound the Charles river?

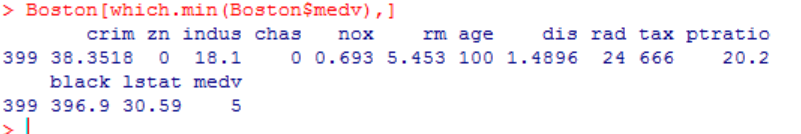
**There are 471 of the suburbs in this data set bound the Charles river**

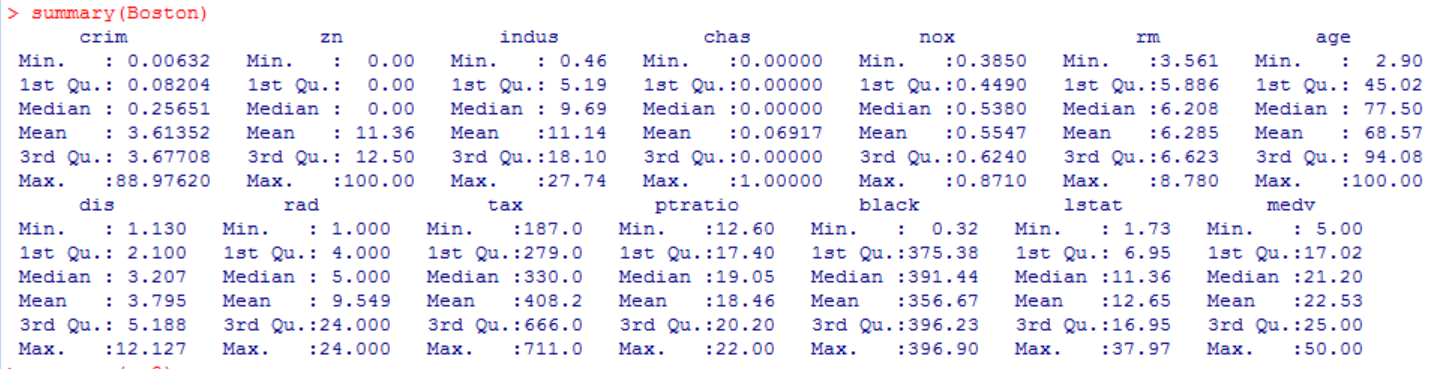


(f) What is the median pupil-teacher ratio among the towns in this data set?



(g) Which suburb of Boston has lowest median value of owneroccupied homes? What are the values of the other predictors for that suburb, and how do those values compare to the overall ranges for those predictors? Comment on your findings.

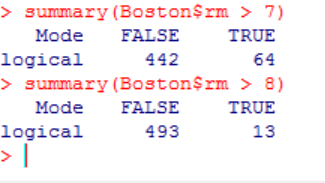




**Most of predictors are in higher side compare with the total sample, while there are only two predictors, which are Zn,Dis . It has low proportion of residential land zoned for lots over 25,000 sq.ft; and low weighted mean of distances to five Boston employment centres.**

(h) In this data set, how many of the suburbs average more than seven rooms per dwelling? More than eight rooms per dwelling? Comment on the suburbs that average more than eight rooms per dwelling.

**Rm represents as average number of rooms per dwelling. There are 64 of the suburbs average more than seven rooms per dwelling, and 13 of them are more than eight rooms per dwelling.**



**Compare each columns’ characteristics of the suburbs that average more than eight rooms per dwelling with them of all the suburbs. I found there are extremely high  proportion of blacks by town when the suburbs more than eight rooms; also, small percentage of lower status of the population when the suburbs more than eight room.**

**> rm8=Boston [Boston$rm > 8, ] VS > summary(Boston)**

**> summary(rm8)**

