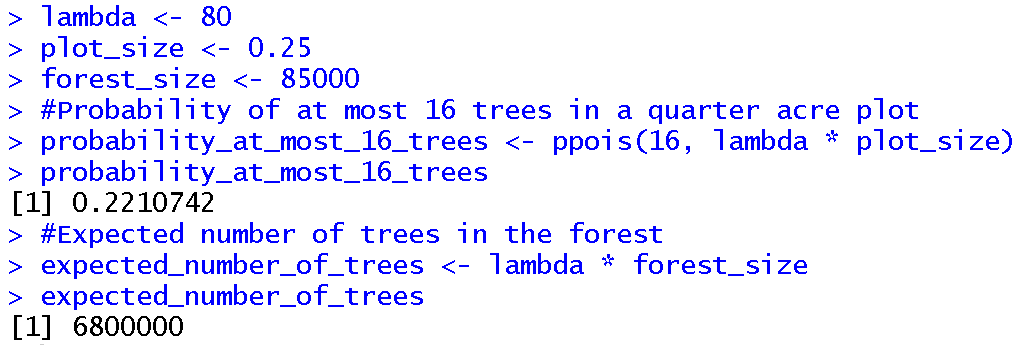
# #Problem 1

**Suppose that trees are distributed in a forest according to a two-dimensional Poisson process with lambda the expected number of trees per acre given by 80.**

**(a)(6pts) What is the probability that in a certain quarter acre plot there will be at most 16 trees?**

**(b) (4pts) If the forest covers 85,000 acres what is the expected number of trees in the forest?**

**Answer:**



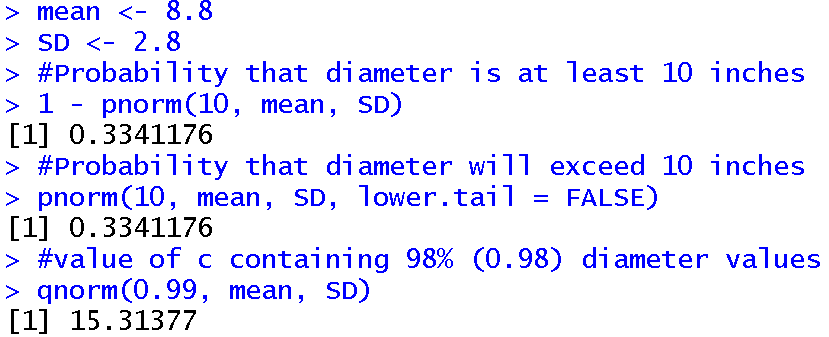
# #Problem 2

**Suppose the diameter at breast height (in.) of a certain type of tree is distributed normally with a mean of 8.8 and a standard deviation of 2.8 as suggested in Aedo-Ortiz, D. M., Olsen, E. D., & Kellogg, L. D. (1997). Simulating a harvester-forwarder softwood thinning: a software evaluation. Forest Products Journal, 47(5), 36.**

**(a) (4pts) What is the probability that the diameter of a randomly selected tree will be at least 10 inches? Will exceed 10 inches.**

**(b) (6pts) What is the value of c so that the interval (8.8-c, 8.8+c) contains 98% of all diameter values**

**Answer:**



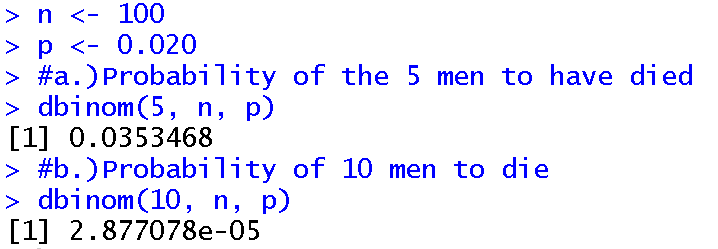
# #Problem 3

**Suppose that a group of 100 men aged 60-64 received a new flu vaccine in 1986 and that 5 of them died within the next year. Assume that a 1986 mortality table tells us that men aged 60-64 have an approximate probability of death in the next year of .020.**

**(a)(6pts) How unusal is it for the 5 men to have died. [Hint: Be careful to review my discussions on assessing extreme event. Also, use the binomial model.]**

**(b) (4pts) Suppose 10 men had died. Should we be concerned?**

**Answer:**



Interpretation:

# #Problem 4

**The incidence of a certain type of chromosome defect in the U.S. adult male population is beleived to be1 in 75. A random sample of 800 individuals in U.S. pemal institutions reveals 16 who have such defects.Can it be concluded that the incidence rate of this defect among prisoners differs from the preseumed rate of the entire adult population.**

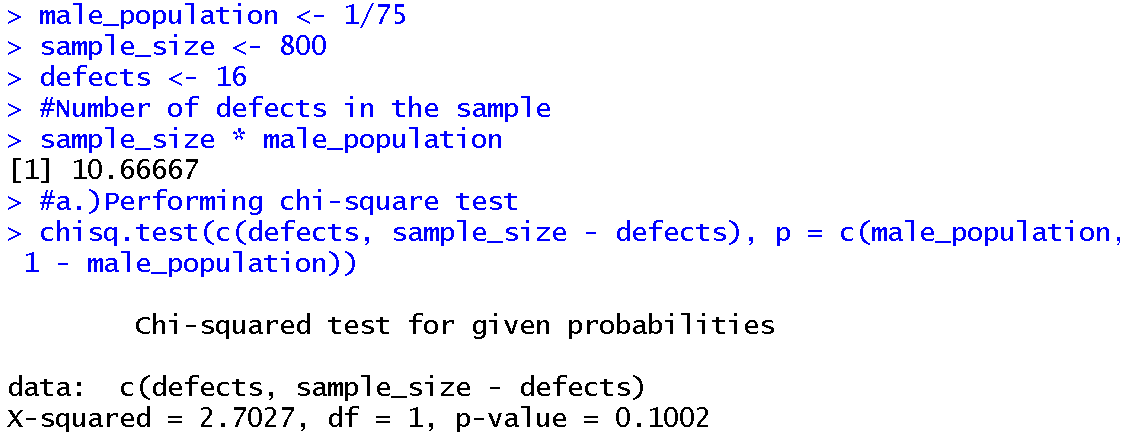
**(a)(6pts)State and test the relevant hypothesis using =.05. What type of error might you have made in reaching a conclusion?**

**(b) (4pts) What P-value is associated with this test. Based on this P-value could be rejected at a significance level of .2**

**Answers:**

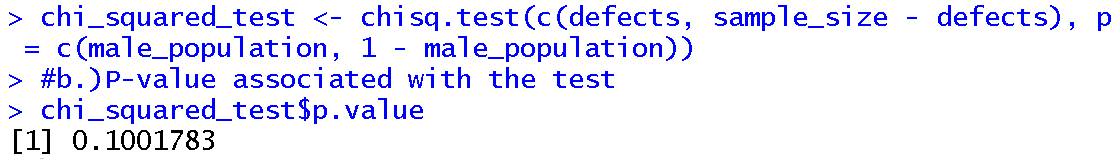
1. Null hypothesis – The incidence rate among prisoners is the same as the presumed rate in the general population

Alternate hypothesis - The incidence rate among prisoners is not the same as the presumed rate in the general population



Rejection of null hypothesis at a significant level of 0.05:

Base on the chi-squared test, there was no evidence to reject the null hypothesis at a significant level of 0.05.



Rejection of null hypothesis at a significant level of 0.2:

Since the p-value obtained is less than the 0.2, the null hypothesis was rejected at a significant level of 0.2.

# #Problem 5

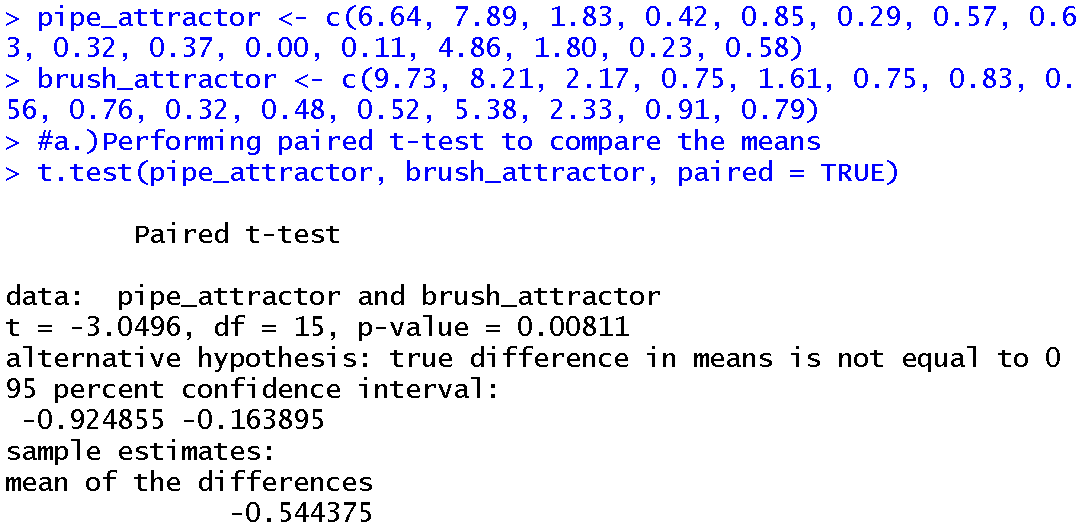
**Two different fish attractors were compared during 16 time periods spanning 4 years. (Wilbur, R. L. (1978). Two types of fish attractors compared in Lake Tohopekaliga, Florida. Transactions of the American Fisheries Society, 107(5), 689-695.) I present the data as a set of ordered pairs where in each case the first entry is the “pipe attractor” and the second entry is the “brush attractor”**

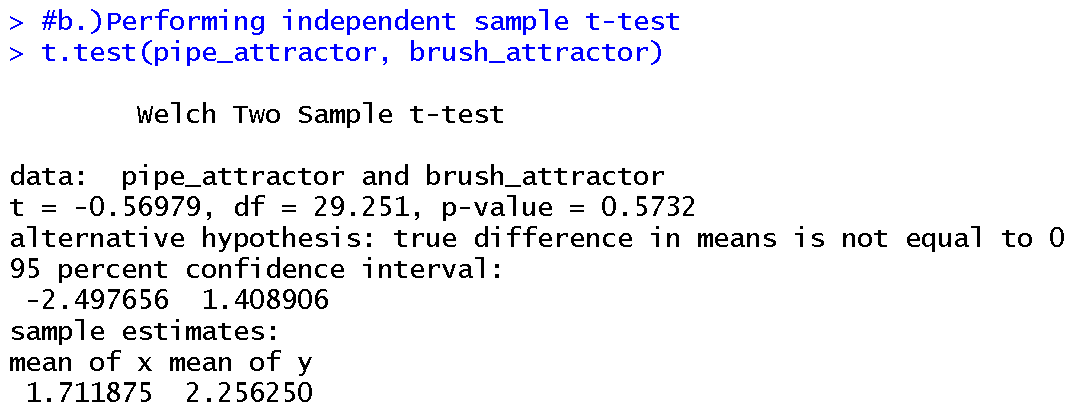
**{(6.64,9.73), (7.89,8.21),(1.83,2.17),(.42,.75),(.85,1.61),(.29,.75),(.57,.83),(.63,.56),(.32,.76),(.37,.32),(.00,.48),(.11.,.52),(4.86,5.38),(1.80,2.33),(.23,.91,),(.58,.79)}**

**(a)(4 pts) Perform the apprpriate paired parameteric t-based tests to compare the means.**

**(b)(6 pts) What happens if the two independent samples t-test is used? Make sure to perform all appropriate tests.**

**Answer:**





The paired t-test compares the means of the two attractors within each time period, while the independent samples t-test treats the data as two independent groups.

# #Problem 6

**The following article (Thomas, H. V., & Simmons, E. L. L. A. M. A. E. (1969). Histamine content in sputum from allergic and nonallergic individuals. Journal of applied physiology, 26(6), 793-797) reports the following data on sputum histamine levels (g/g dry weight of sputum) for a sample of 9 individuals classified as allergics and 13 individuals classified as nonallergics.**

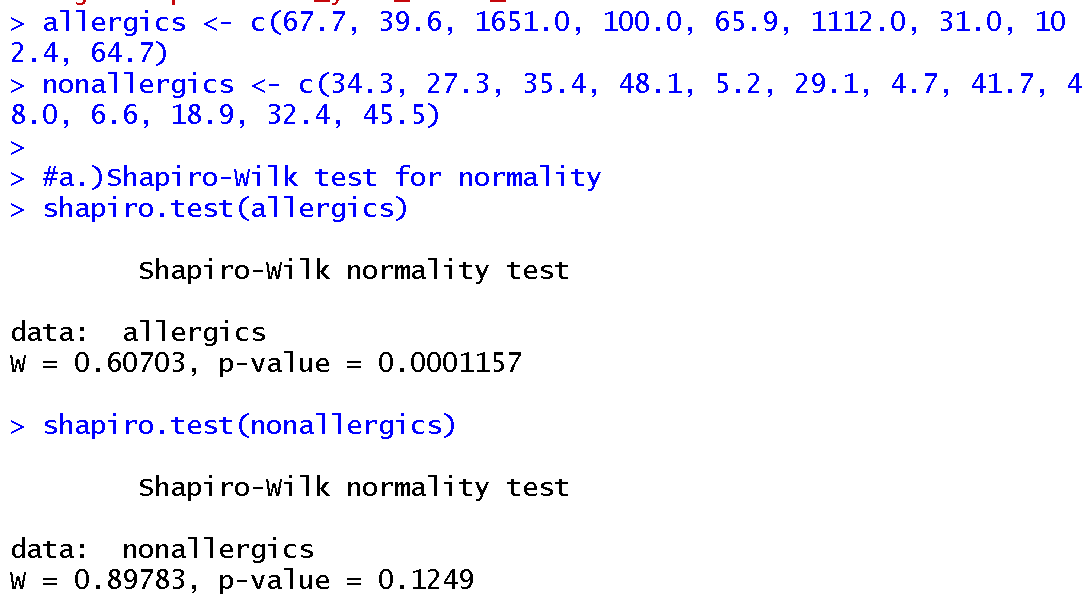
**Allergics = {67.7, 39.6,1651.0, 100.0, 65.9,1112.0,31.0, 102.4,64.7}**

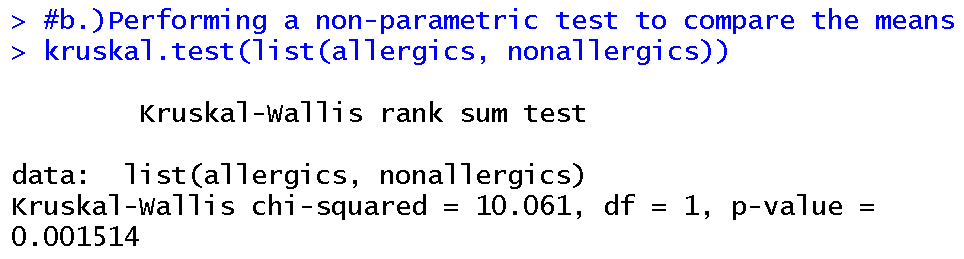
**Nonallergics = {34.3, 27.3, 35.4, 48.1, 5.2, 29.1, 4.7, 41.7, 48.0, 6.6, 18.9, 32.4, 45.5}**

**(a)(4 pts) Ascertain the normality of this data using the Shapiro Wilks test.**

**(b) (6pts) You should have seen that one of these failed the test. Perform the apprpriate nonparametric test to compare the means. Does the P-value suggest that we can reject at the .01 level?**

**Answers:**





Since the p-value from Kruskal-wallis test is 0.001514 which is less than 0.01, this test suggests rejection at the 0.01 level.

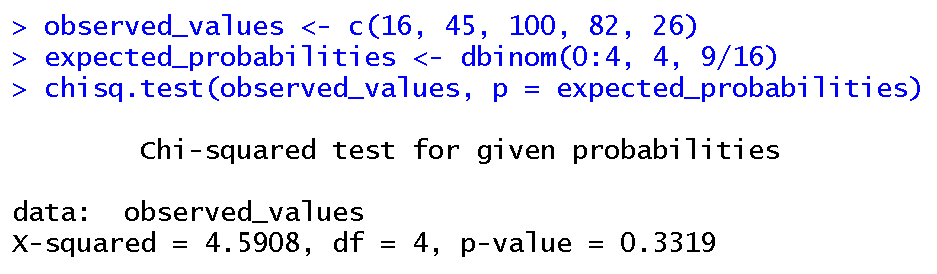
# #Problem 7

**(10 pts) In a well-known genetics article (Yule, G. U. (1923). The progeny, in generations F 12 to F 17 of a cross between a yellow-wrinkled and a green-round seeded pea; a report on data afforded by experiments initiated by the late AD Darbishire, MA, in 1905, and conducted by him until his death in 1915. Journal of Genetics, 13(3), 255-331) the early statistician G. U. Yule dnalyzed data from crossing garden peas. The dominant alleles were Y = yellow color and R = round shape resulting in a double dominant YR.Yule examined 269 4 seed podsresulting from a dihybrid cross and counted the number of YR seeds in each pod.Letting X represent the number of seeds in a randomly selected pods possible X values are 0, 1, 2, 3, and 4 which are identified with cells 1, 2, 3, 4, 5 of a 1 x 5 table. So a pod with X = 4 would have its count values in the 5th entry of the table.The hypothesis that the Mendelian laws are operative and that genetypes of individual seeds within a pod are independent of one another implies that X has a binomial distribution with n=4 and p=9/16. The null hypothesis is that the Medelian lawa hold true and the alternative hypothesis is that they do not. The entry in the table (the observed seed counts) reading from the first to the fifth are {16,45,100,82,26}.The probabilities under Ho are given by the values of the binomial distribution with n=4 and p=9/16 as the number of seeds i obtains values in the set {0,1,2,3,4}. Notice that the observed seed count for the 0th case is 16, the 1 case 45, etc. Use the R function chisq.test to test H0. [Hint - You will have to setup a vector of the observed values and a vector of the expected probabilities. Use the values as detailed above.] Do we reject the null hypothesis at the level of .1?**

Answer:

Null Hypothesis (H0) – The Mendelian laws are operative and that genetypes of individual seeds within a pod are independent of one another

Alternate Hypothesis (H1) - The Mendelian laws are not operative and that genetypes of individual seeds within a pod are not independent of one another



Since the p-value from the chi-square test is 0.3319, which is more than 0.1, we do not reject the null hypothesis at significant level of 0.1

# #Problem 8

**An article Annest, J. L., Pirkle, J. L., Makuc, D., Neese, J. W., Bayse, D. D., & Kovar, M. G. (1983). Chronological trend in blood lead levels between 1976 and 1980.New England Journal of Medicine, 308(23), 1373-1377 gives the following data on y = average blood lead level of white children age 6 months to 5 years and x = amount of lead used in gasoline production (in 1000 tons) for ten 6-month periods.x = {48,59,79,80,95,95,97,102,102,107} y={9.3,11.0,12.8,14.1,13.6,13.8,14.6,14.6,16.0,18.2}.**

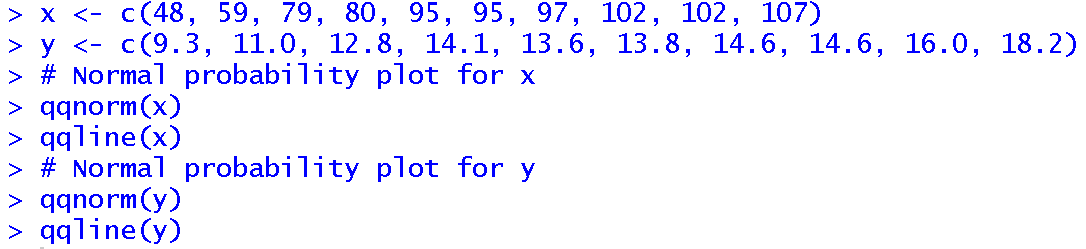
**(a) (3 pts) Construct normal probability plots for x and y. Might it be reasonable to assume that they come from a bivariate normal distribution?**

**(b) (3 pts) Examine a scatterplot of y versus x. Does it look like there may be some correlation structure?**

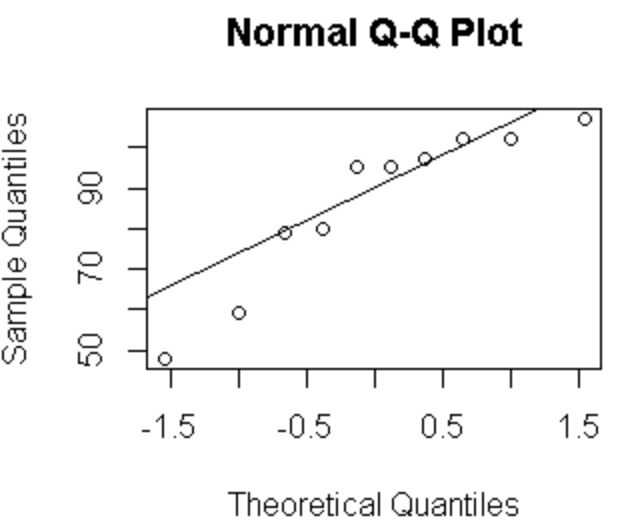
**(c) (4 pts) Does the data provide statistically significant information that there is a linear relationship between blood lead level and the amount of lead used in gasoline production? [Hint: Use an level of .05 and the R function cor.test]**

**Answer:**

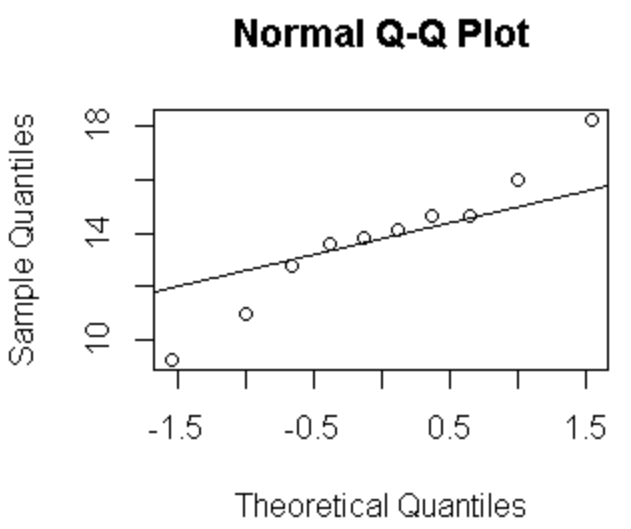
**(a.)**



Normal probability plot for x:

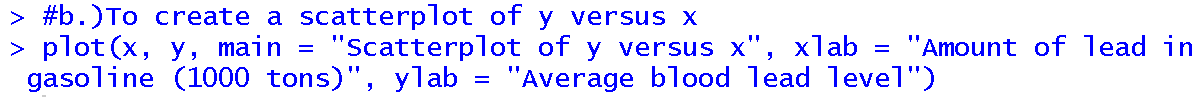


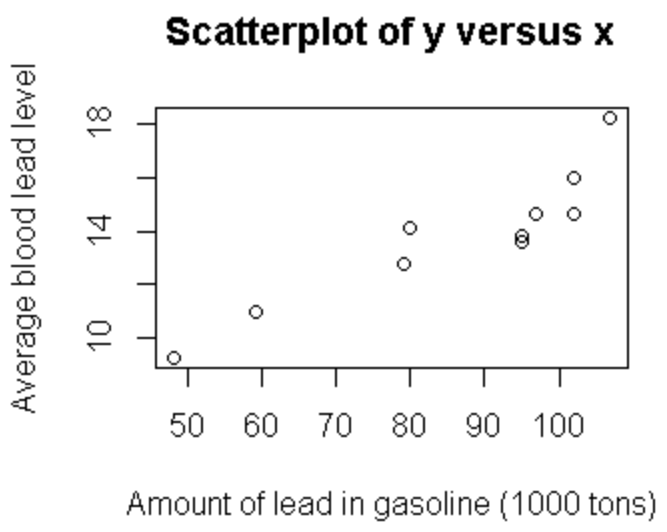
Normal Probability plot for y:



The normal probability plot is a graphical method to assess if a dataset follows a normal distribution. If the points approximately fall along the straight reference line, it suggests that the data may come from a normal distribution.

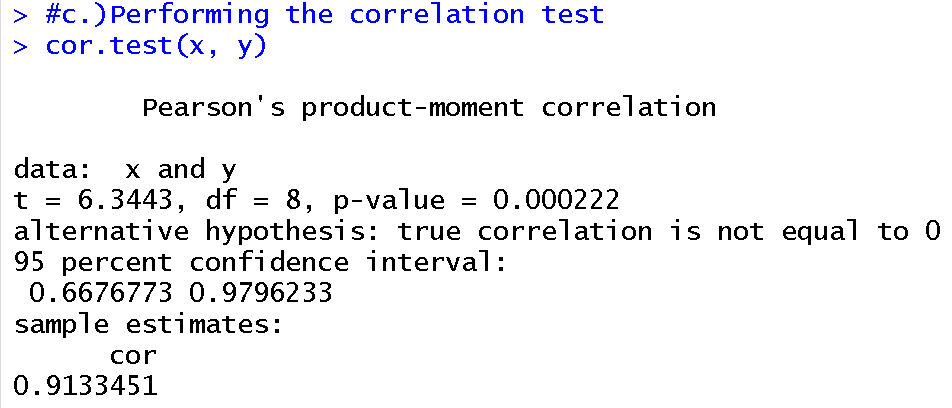
**(b.)**





Interpretation –

**(c.)**

****

Since the p-value from the correlation test is 0.000222, which is less than 0.05, the data provides statistically significant information of a linear relationship at a significance level of 0.05

# #Problem 9

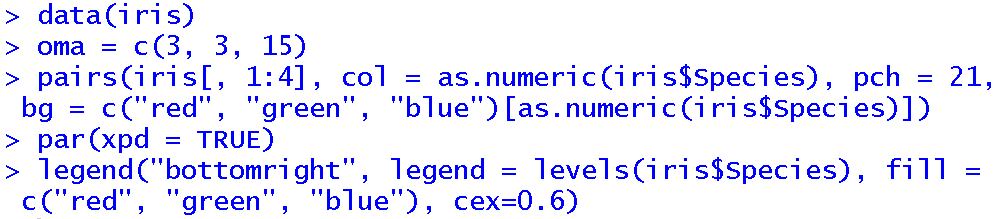
**Consider the famous Fisher Iris data.**

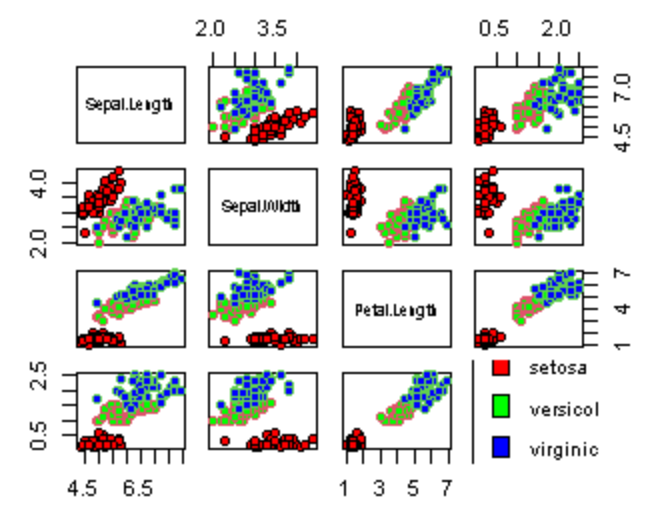
**(a)(5pts) Make a pairs plot of this data coloring the observations based on the species designation using a red, green, blue color scheme and a plot symbol of pch = 21.[Hint: You will need to adjust the figure margin using oma=c(3,3,15) when you call pairs and also call par(xpd=TRUE) prior to calling legend. Position your legend in the bottomright of the plot]**

**(b)(5pts)Repeat the exercise of part (a) except produce a parallel coordinates plot.**

**Answer:**

**(a.)**





**(b.)**

# #Problem 10

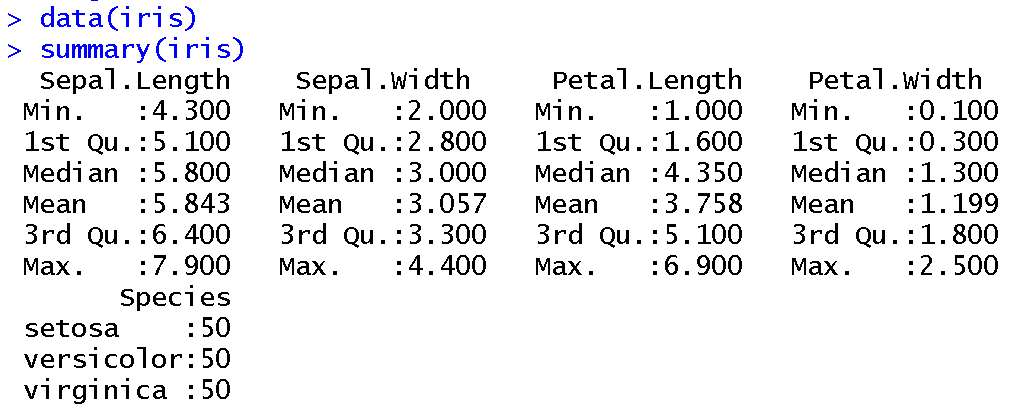
**Let’s return to the Fisher Iris data.**

**(a) (5pts) Use the summary function to explore this data.**

**(b) (5pts) Create a 2 x 2 matrix of box plots of the Fisher iris data with the sepal width and sepal length in the first row and the petal width and petal length in the second row.**

**Answer:**

**(a.)**



**(b.)**

