Exercise 1

a) Basic descriptive statistics for Sepal Length for three species

Variable: Sepal Length (Sepal Length (mm))

Moments			
N	150	Sum Weights	150
Mean	58.4333333	Sum Observations	8765
Std Deviation	8.28066128	Variance	68.5693512
Skewness	0.31491096	Kurtosis	-0.552064
Uncorrected SS	522385	Corrected SS	10216.8333
Coeff Variation	14.171126	Std Error Mean	0.67611316

Basic Statistical Measures			
Location		Variability	
Mean	58.43333	Std Deviation	8.28066
Median	58.00000	Variance	68.56935
Mode	50.00000	Range	36.00000
		Interquartile Range	13.00000

Looking at Sepal Length for all the three species, the skew is 0.315 and the mean and median are similar. The mean and standard deviation will work fine for describing typical values and spread. The average Sepal Length is estimated to be about 58.43333 with roughly 95% of the observed Sepal Length for all the three species falling between 41.9 and 75.

b) From these three species which is Setosa, Virginica and Versicolor. We can observe two similarities which they have a very small and similar positive skew around 0.1, so these species for Sepal Length seem to be symmetric. Not only that, all of three species of Sepal Length's mean are similar but the mean is negligible larger than median, consistent with the small positive skew. The differences are they have different mean value which are 50.06 for Setosa, 65.88 for Virginica and 59.36 for Versicolor. Virginica has the largest mean value in Sepal Length while the Setosa has the smallest mean value.

Variable: Sepal Length (Sepal Length (mm)) (Setosa)

Moments			
N	50	Sum Weights	50
Mean	50.06	Sum Observations	2503
Std Deviation	3.52489687	Variance	12.424898
Skewness	0.12008699	Kurtosis	-0.2526888

Uncorrected SS	125909	Corrected SS	608.82
Coeff Variation	7.04134413	Std Error Mean	0.4984957

Basic Statistical Measures			
Location		**Variability**	
Mean	50.06000	Std Deviation	3.52490
Median	50.00000	Variance	12.42490
Mode	50.00000	Range	15.00000
		Interquartile Range	4.00000

Note: The mode displayed is the smallest of 2 modes with a count of 8.

Looking at Sepal Length for the species for Setosa, the skew is 0.12 which is a very small value, and the mean and median are similar. The descriptive statistics suggest the mean of Sepal Length for the species of Setosa was around 50.06, and we might expect about 95% of the observed Sepal Length for the species of Setosa to between 7 below and 7 above the mean value.

Variable: Sepal Length   (Sepal Length (mm)) (Virginica)

Moments			
N	50	Sum Weights	50
Mean	65.88	Sum Observations	3294
Std Deviation	6.35879593	Variance	40.4342857
Skewness	0.11801512	Kurtosis	0.03290442
Uncorrected SS	218990	Corrected SS	1981.28
Coeff Variation	9.65208854	Std Error Mean	0.89926954

Basic Statistical Measures			
Location		**Variability**	
Mean	65.88000	Std Deviation	6.35880
Median	65.00000	Variance	40.43429
Mode	63.00000	Range	30.00000
		Interquartile Range	7.00000

Looking at Sepal Length for the species for Virginica, the skew is 0.12 which is a very small value, and the mean and median are similar. The descriptive statistics suggest the mean of Sepal Length for the species of Virginica was around 65.88, and we might expect about 95% of the observed Sepal Length for the species of Virginica to between 12.7 below and 12.7 above the mean value.

Variable: Sepal Length   (Sepal Length (mm))(Versicolor)

Moments			
N	50	Sum Weights	50
Mean	59.36	Sum Observations	2968
Std Deviation	5.16171147	Variance	26.6432653
Skewness	0.10537762	Kurtosis	-0.5330095
Uncorrected SS	177486	Corrected SS	1305.52
Coeff Variation	8.69560558	Std Error Mean	0.72997624

Basic Statistical Measures			
Location		**Variability**	
Mean	59.36000	Std Deviation	5.16171
Median	59.00000	Variance	26.64327
Mode	55.00000	Range	21.00000
		Interquartile Range	7.00000

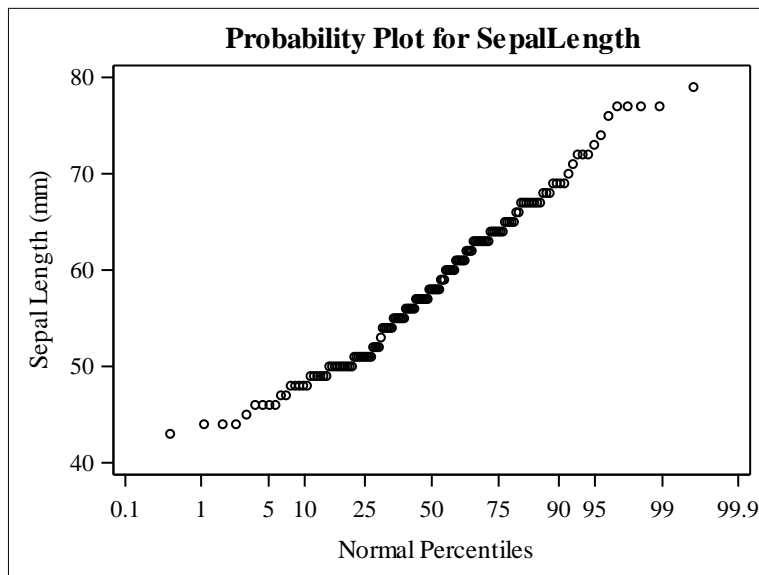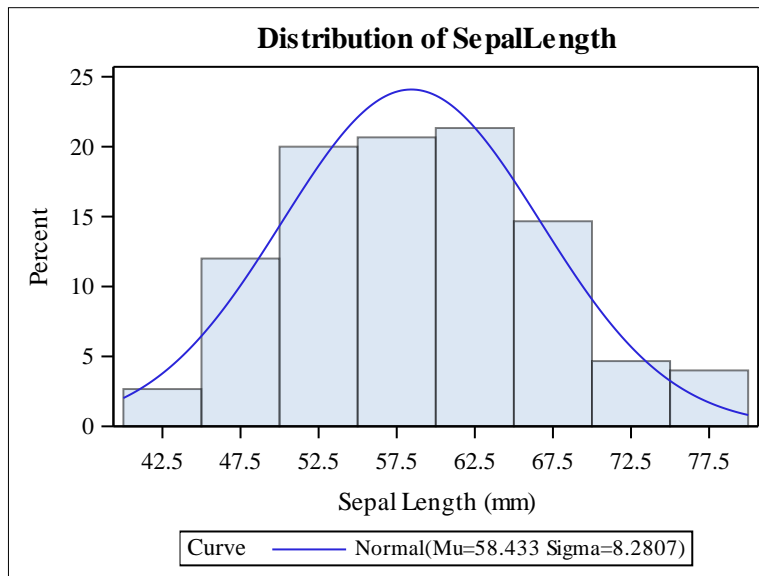***Note: The mode displayed is the smallest of 3 modes with a count of 5.***

Looking at Sepal Length for the species for Versicolor, the skew is 0.11 which is a very small value, and the mean and median are similar. The descriptive statistics suggest the mean of Sepal Length for the species of Versicolor was around 59.36, and we might expect about 95% of the observed Sepal Length for the species of Versicolor to between 10.3 below and 10.3 above the mean value.

Exercise 2

a) The normality tests all strongly reject for the Sepal Length for all the three species. The p-values for all the of the hypothesis are very small, indicating the sample is very far from what would be expected from a normal distribution. The histogram looks like most of the data are in the middle instead of a bell shape.
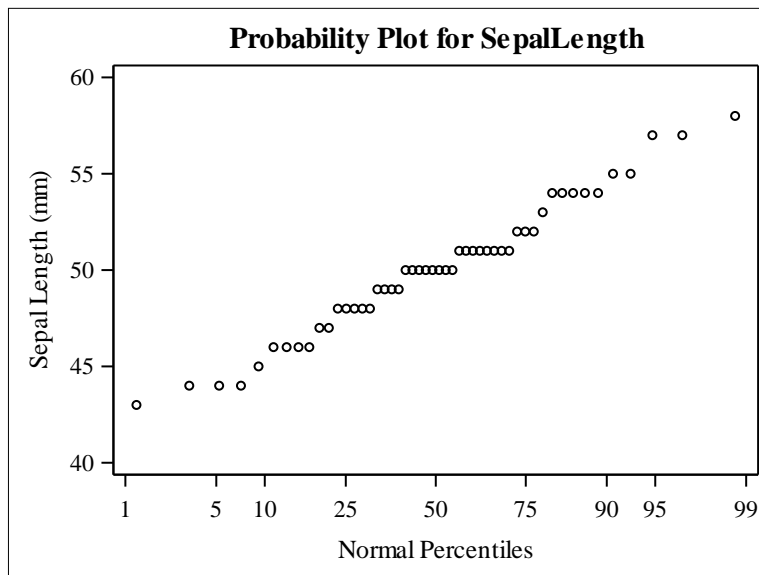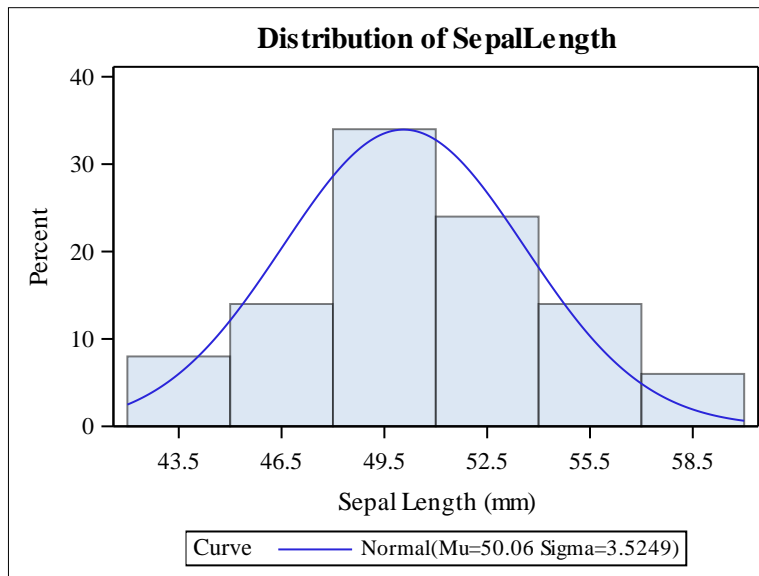
Variable: Sepal Length   (Sepal Length (mm))

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.97609	Pr < W	0.0102
Kolmogorov-Smirnov	D	0.088654	Pr > D	<0.0100
Cramer-von Mises	W-Sq	0.127398	Pr > W-Sq	0.0479
Anderson-Darling	A-Sq	0.889199	Pr > A-Sq	0.0231

Distribution of SepalLength

Curve —— Normal(Mu=58.433 Sigma=8.2807)

Probability Plot for SepalLength

b)However , Sepal Length for the species of Setosa show a different trend from the overall three species of the Sepal Length. The histogram does show like a bell shape and the probability plot appears to be very straight, indicating the sample is not far from normal. The normality tests also fail to reject an assumption of normality. Tests relying on normality assumptions will be fine for Sepal Length for the species of Setosa.

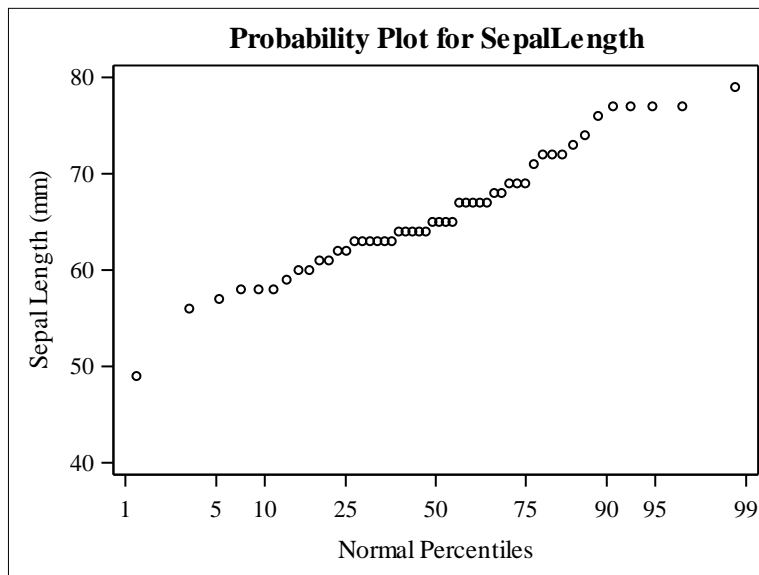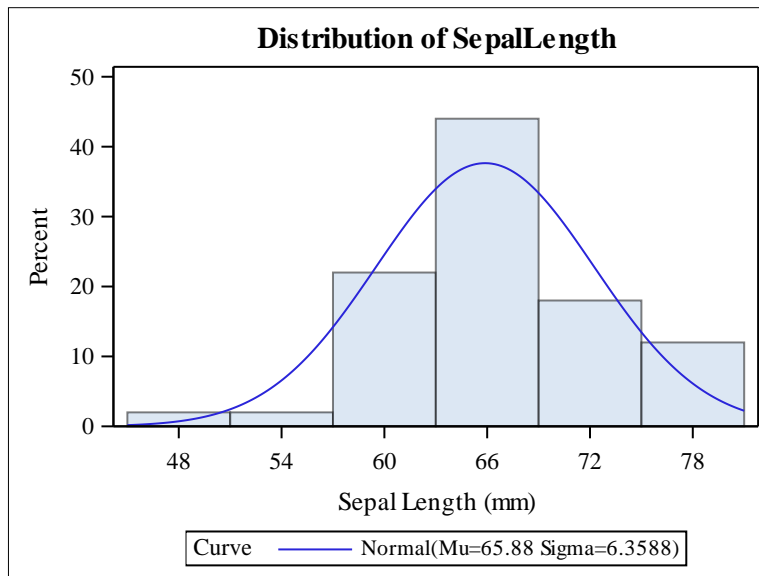*Variable: Sepal Length (Sepal Length (mm))(Setosa)*

Tests for Normality				
Test	**Statistic**		**p Value**	
Shapiro-Wilk	**W**	0.977699	**Pr < W**	0.4595
Kolmogorov-Smirnov	**D**	0.11486	**Pr > D**	0.0962
Cramer-von Mises	**W-Sq**	0.071753	**Pr > W-Sq**	>0.2500
Anderson-Darling	**A-Sq**	0.407986	**Pr > A-Sq**	>0.2500

Distribution of SepalLength

Curve —— Normal(Mu=50.06 Sigma=3.5249)

Probability Plot for SepalLength

Sepal Length for the species of Virginica show a different trend from the overall three species of the Sepal Length. The histogram does not show like a bell shape suggesting the lack of bell shape is more of the histogram's binning, but the probability plot appears to be very straight, indicating the sample is not far from normal. The normality tests also fail to reject an assumption of normality. Tests relying on normality assumptions will be fine for Sepal Length for the species of Virginica.
.

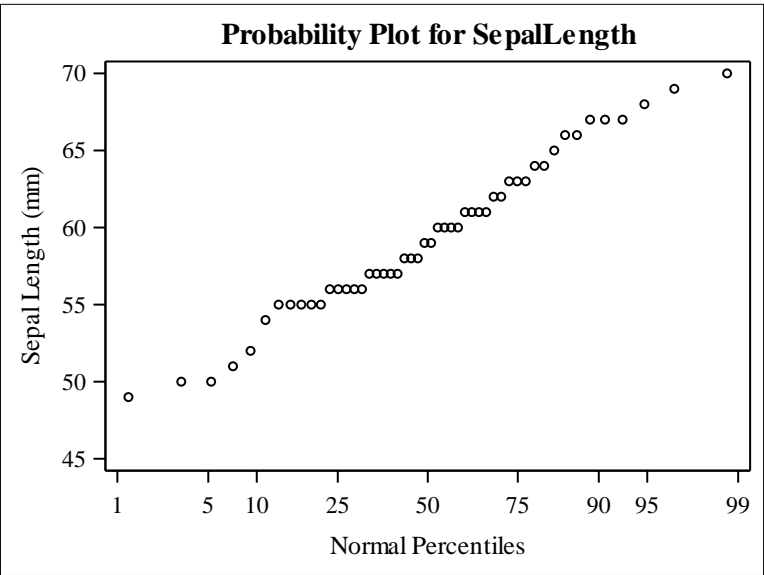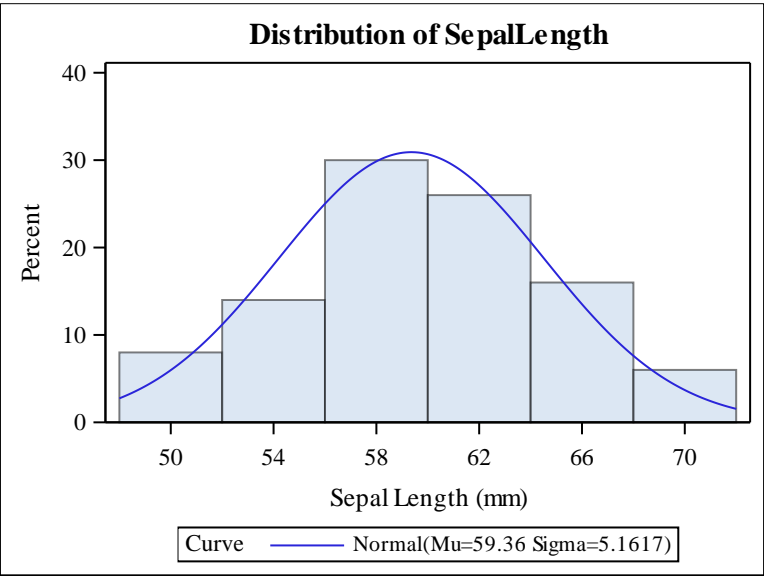*Variable: Sepal Length   (Sepal Length (mm))(Virginica)*

Tests for Normality				
Test	**Statistic**		**p Value**	
Shapiro-Wilk	**W**	0.971179	**Pr < W**	0.2583
Kolmogorov-Smirnov	**D**	0.115034	**Pr > D**	0.0953
Cramer-von Mises	**W-Sq**	0.089467	**Pr > W-Sq**	0.1538
Anderson-Darling	**A-Sq**	0.551641	**Pr > A-Sq**	0.1506

Distribution of SepalLength

Curve ——— Normal(Mu=65.88 Sigma=6.3588)

Probability Plot for SepalLength

Sepal Length for the species of Versicolor show a different trend from the overall three species of the Sepal Length. The histogram does show like a bell shape and the probability plot appears to be very straight, indicating the sample is not far from normal. The normality tests also fail to reject an assumption of normality. Tests relying on normality assumptions will be fine for Sepal Length for the species of Versicolor.

Variable: Sepal Length (Sepal Length (mm))(Versicolor)

Tests for Normality				
Test	**Statistic**		**p Value**	
Shapiro-Wilk	**W**	0.977836	**Pr < W**	0.4647
Kolmogorov-Smirnov	**D**	0.096241	**Pr > D**	>0.1500
Cramer-von Mises	**W-Sq**	0.057273	**Pr > W-Sq**	>0.2500
Anderson-Darling	**A-Sq**	0.360841	**Pr > A-Sq**	>0.2500

Distribution of SepalLength

Curve ——— Normal(Mu=59.36 Sigma=5.1617)

Probability Plot for SepalLength

Exercise 3

a) The Sepal Length were right-skewed. Normality cannot be assumed, so the t test is out. Symmetry cannot be assumed because of the longtail, so the signed rank test should not be used. This leaves the sign test, which only assumed the observations came from the same distribution and are not correlated.

The sign test has a probability of 0.0798 for two-sided test but we need to divide by two to get the p-value for one-sided test which is 0.0399 which is lower than 0.05, so the null is rejected in favor of the alternative that the true median of Sepal Length that is less than 60.0.

Variable: Sepal Length   (Sepal Length (mm))

Tests for Location: Mu0=60				
Test	**Statistic**		**p Value**	
Student's t	t	-2.31717	**Pr > |t|**	0.0219
Sign	M	-11	**Pr >= |M|**	0.0798
Signed Rank	S	-1238.5	**Pr >= |S|**	0.0129

Tests for Normality				
Test	**Statistic**		**p Value**	
Shapiro-Wilk	W	0.97609	**Pr < W**	0.0102
Kolmogorov-Smirnov	D	0.088654	**Pr > D**	<0.0100
Cramer-von Mises	W-Sq	0.127398	**Pr > W-Sq**	0.0479
Anderson-Darling	A-Sq	0.889199	**Pr > A-Sq**	0.0231

b) In the case of the Versicolor and Setosa, the folded F test is highly significant, and we need unequal variance assumption. With the unequal variance assumption being acceptable, the Satterthwaite test and confidence intervals should be used. The Satterthwaite test is highly significant, indicating that there is significantly difference between the Sepal Length of Setosa and the Sepal Length of Versicolor. The estimated difference is -9.3, and the confidence interval is -11.0571 to -7.5429.

*Variable: Sepal Length   (Sepal Length (mm))(Setosa and Versicolor)*

Species	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
Setosa		50.0600	49.0582	51.0618	3.5249	2.9445	4.3925
Versicolor		59.3600	57.8931	60.8269	5.1617	4.3118	6.4322
Diff (1-2)	**Pooled**	-9.3000	-11.0542	-7.5458	4.4197	3.8782	5.1385
Diff (1-2)	**Satterthwaite**	-9.3000	-11.0571	-7.5429			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	98	-10.52	<.0001
Satterthwaite	Unequal	86.538	-10.52	<.0001

| **Equality of Variances**				
Method	**Num DF**	**Den DF**	**F Value**	**Pr > F**
Folded F	49	49	2.14	0.0087

Exercise 4

a) From the scatterplot matrix depicting Sepal Length for all three species, it's evident that there exists a robust positive linear correlation between Petal Length and Petal Width. This conclusion is drawn from the observation that most data points conform closely to a straight-line pattern. Additionally, we observe other significant positive correlations in the scatter plot matrix, notably between Sepal Length and Petal Length, as well as between Sepal Length and Petal Width. In these cases, a substantial portion of the data points align with a linear trend. However, it's worth noting that certain plots within the scatter plot matrix exhibit greater dispersion, indicating weaker linear relationships.

Variable: Sepal Length   (Sepal Length (mm))
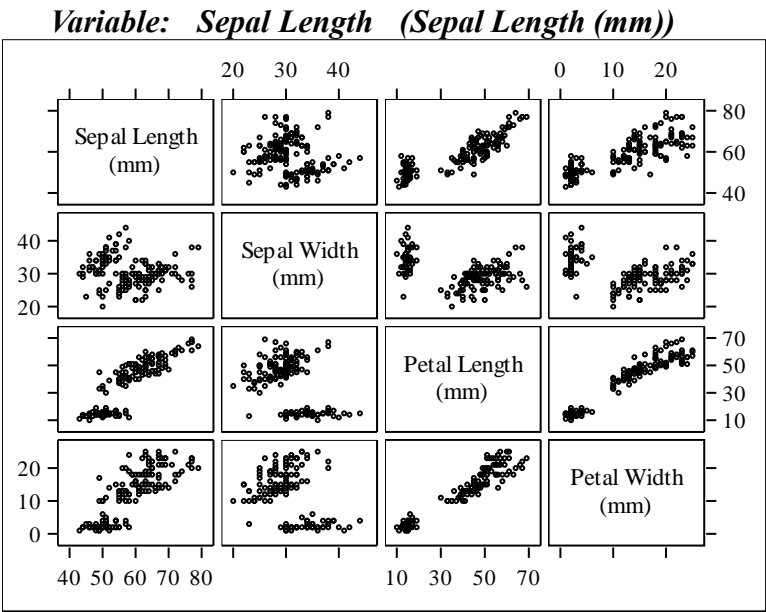
b)From the scatterplot matrix depicting Sepal Length for Setosa, it's evident that there exists a positive linear correlation between Sepal Length and Sepal Width. This conclusion is drawn from the observation that most data points conform closely to a straight-line pattern. However, it's worth noting that certain plots within the scatter plot matrix exhibit greater dispersion, indicating weaker linear relationships. Some of the graphs indicate that while one of the variables is remain constant, but the other variable is varied which is the graph of Sepal Length and Petal Width and it is a different trend from part a.

Variable: Sepal Length   (Sepal Length (mm))(Setosa)
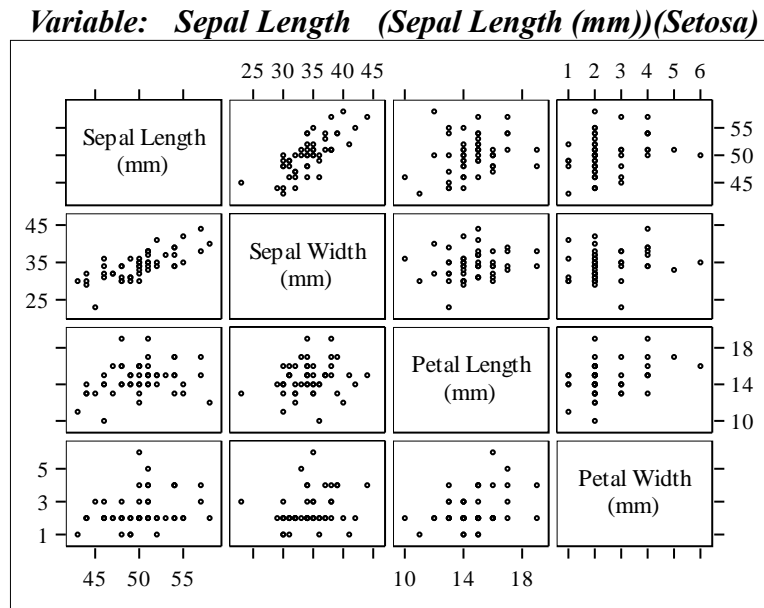
From the scatterplot matrix depicting Sepal Length for Virginica, it's evident that there exists a positive linear correlation between Sepal Length and Petal Length which is a different linear trend from Setosa. This conclusion is drawn from the observation that most data points conform closely to a straight-line pattern. However, it's worth noting that certain plots within the scatter plot matrix exhibit greater dispersion, indicating weaker linear relationships. Some of the graphs indicate that while one of the variables is remain constant, but the other variable is varied which is the graph of Sepal Length and Petal Width and it is a different trend from part a.

Variable: Sepal Length   (Sepal Length (mm))(Virginica)
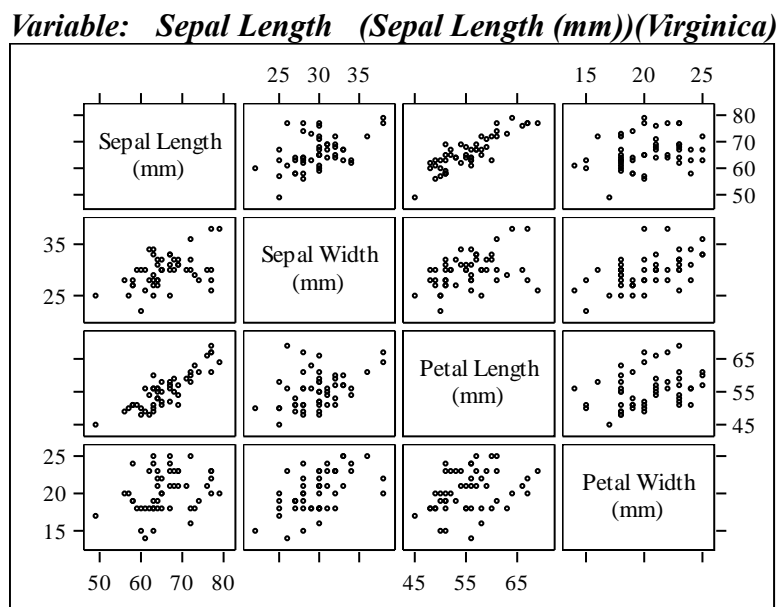
From the scatterplot matrix depicting Sepal Length for Versicolor, it's evident that there exists a positive linear correlation between Sepal Length and Petal Length same as Virginica. This conclusion is drawn from the observation that most data points conform closely to a straight-line pattern. Some of the graphs indicate that while one of the variables is remain constant, but the other variable is varied which is the graph of Sepal Length and Petal Width and it is a different trend from part a.

Variable: Sepal Length   (Sepal Length (mm))(Versicolor)