Asssignment completed with Python.

Question 1

1. 2D scatter plots of the four features

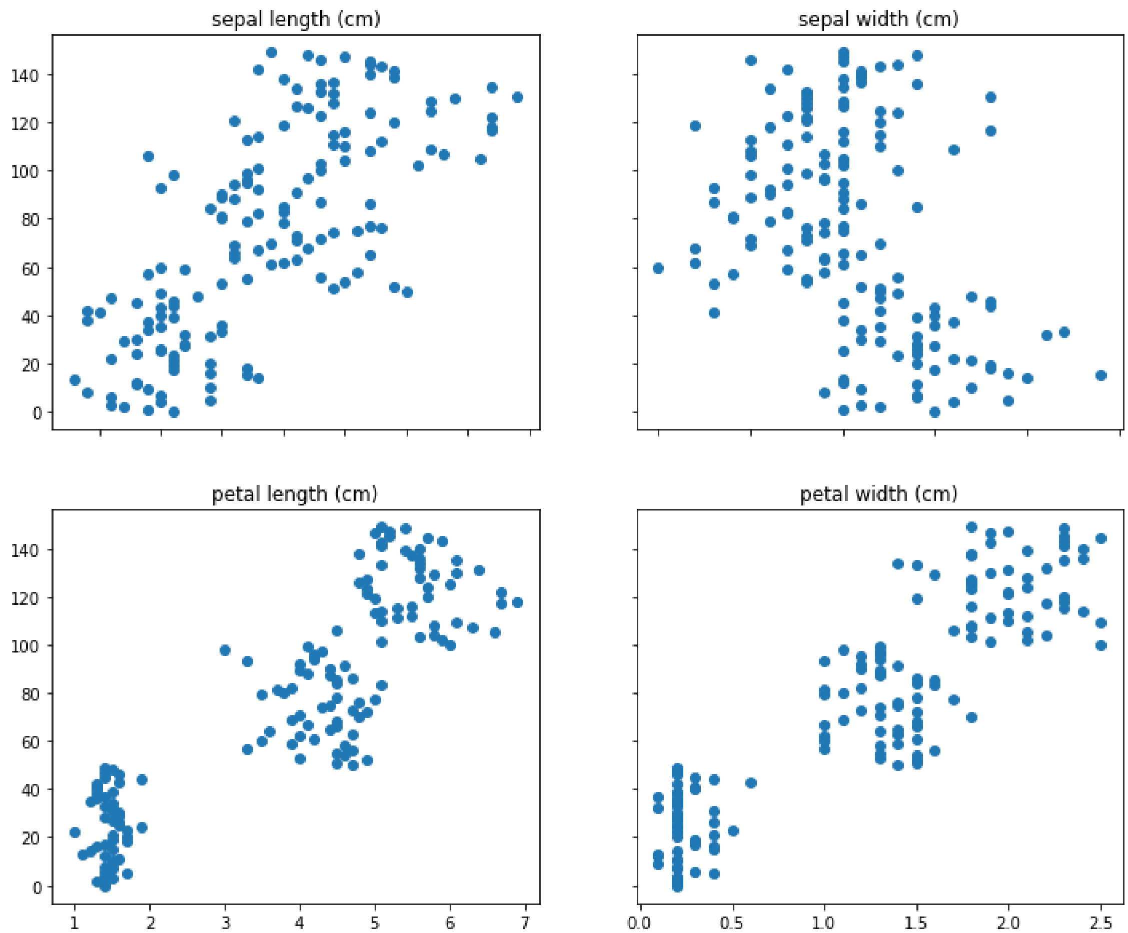

Figure 1: 2D scatter plots of the four features

2. 3D scatter plot of three features: sepal length, sepal width, petal width.

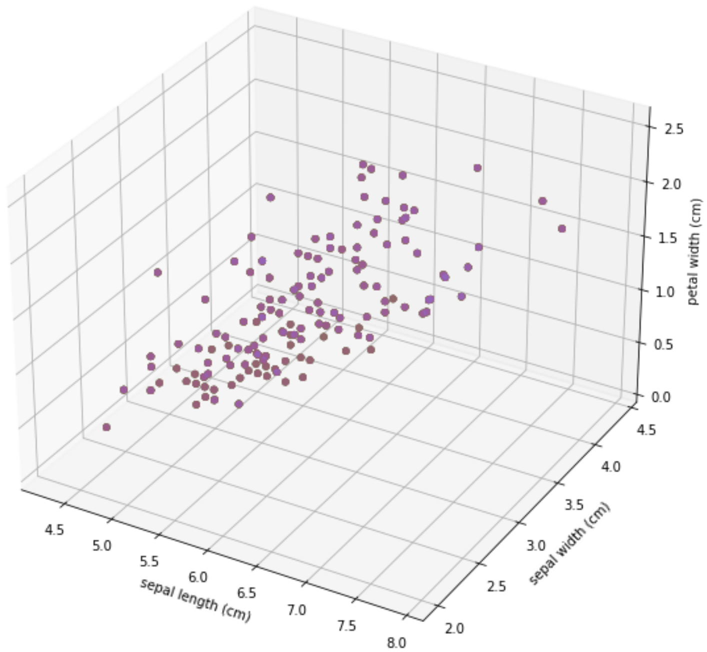

Figure 2: 3D

3. Visualization of the feature matrix (column 1-4) as an image.

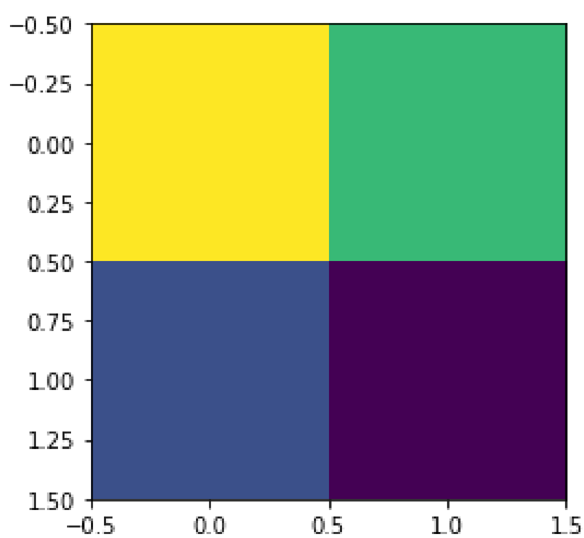

Figure 3: Image

4. For each class, generate histograms for the four features.

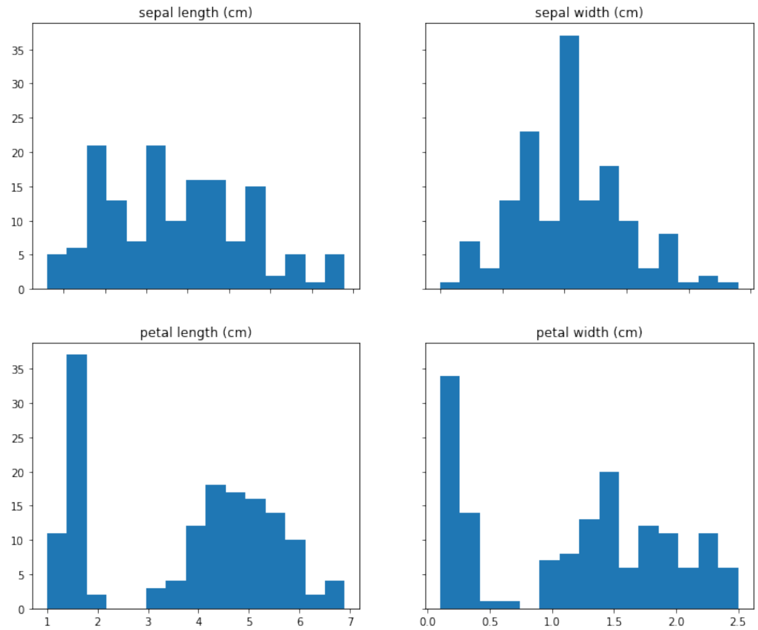

Figure 4: Histogram

5. For each class, generate box-plots of the four features.

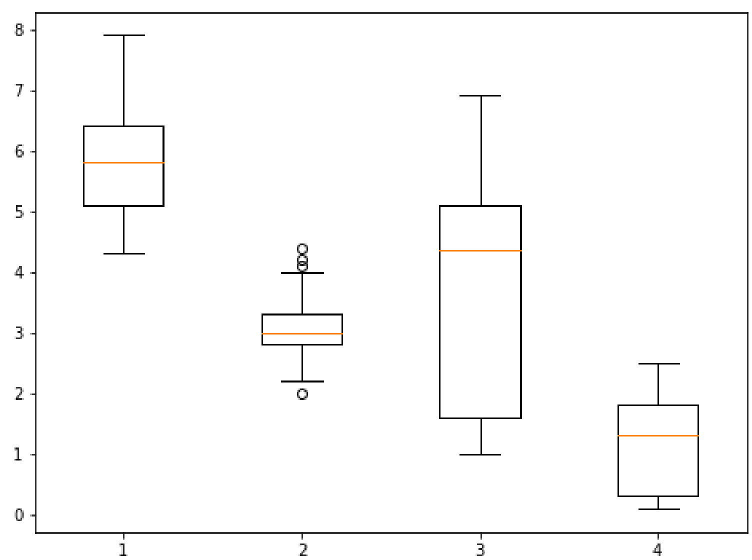

Figure 5: box-plot

6. Calculate the correlation matrix of the four features.

```
x.corr()
```

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)
sepal length (cm)	1.000000	-0.117570	0.871754	0.817941
sepal width (cm)	-0.117570	1.000000	-0.428440	-0.366126
petal length (cm)	0.871754	-0.428440	1.000000	0.962865
petal width (cm)	0.817941	-0.366126	0.962865	1.000000

Figure 6: Correlation matrix of the four features

7. Visualize the correlation matrix as an image.

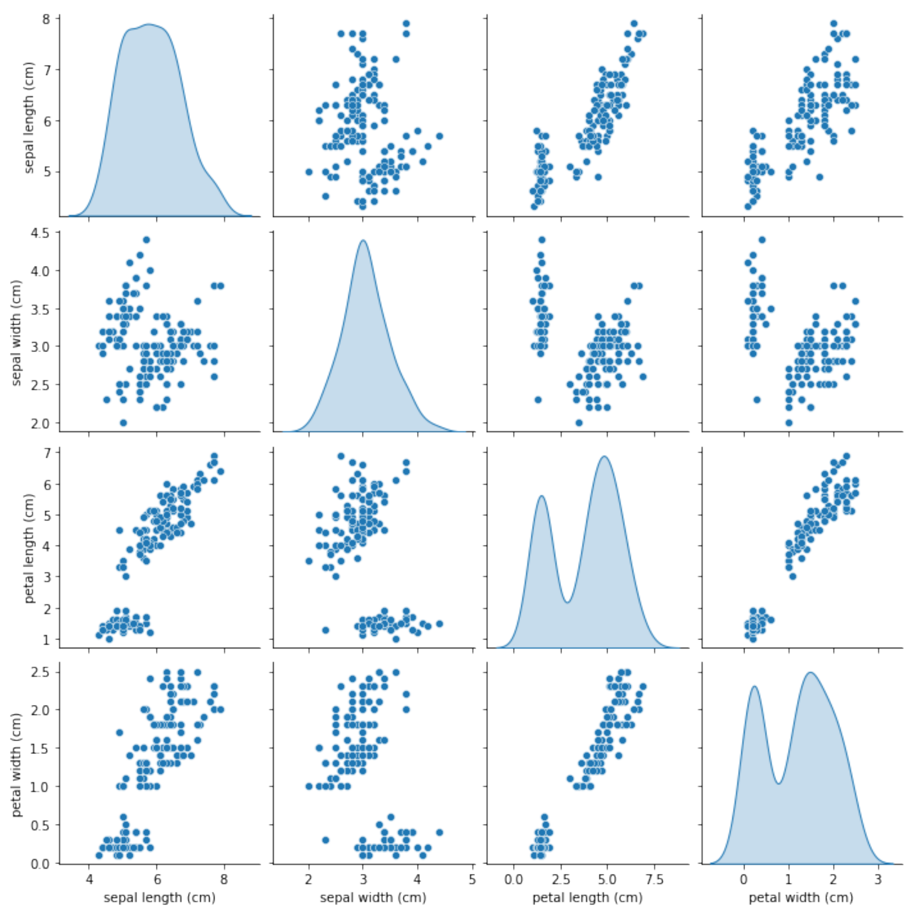

Figure 7: Correlation matrix as an image

8. Create a parallel coordinate plot of the four features.

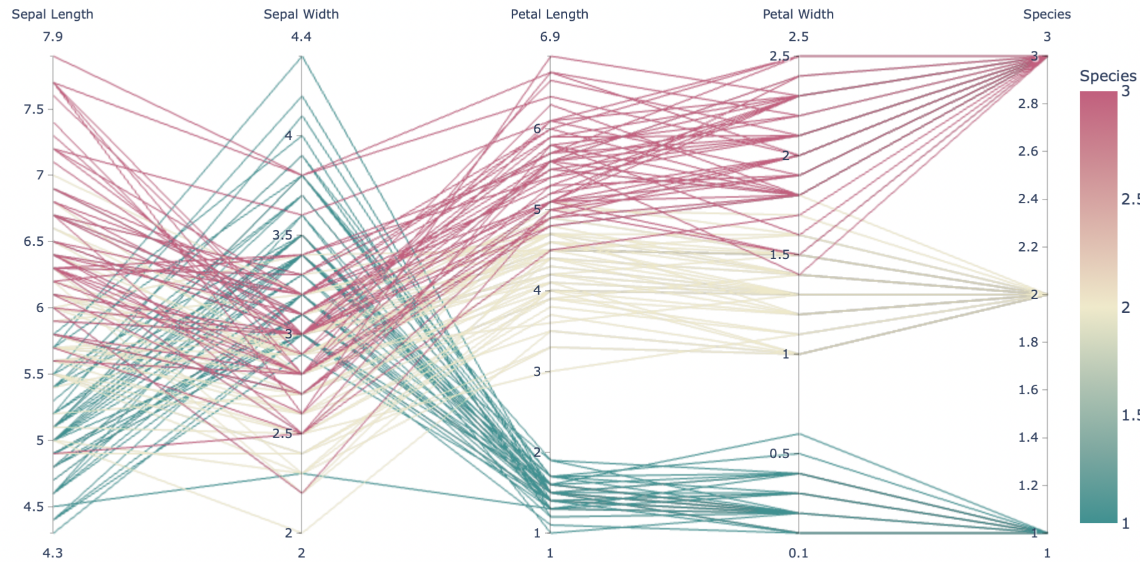

Figure 8: Parallel coordinate plot

Question 2

1. Make a function for Minkowski Distance.

```
def minkowski_distance(vector_a, vector_b, r):
    return (np.sum((abs(vector_a-vector_b)**r),axis=1))**(1/r)
```

Figure 9: Function for Minkowski Distance

2. Make a function for T-statistics Distance.

```
def t_distance(x,y):
        return (abs(np.mean(x)-np.mean(y))*np.sqrt(len(x)))/(np.std(x-y))
```

Figure 10: Function for T-statistics Distance

3. Make a function for Mahalanobis Distance.

```
def mahalanobis(x, vector, cov=None):
        mean = x - vector
        if cov is None:
            cov = np.cov(x.T)
        inverse = np.linalg.inv(cov)
        dot_product = np.dot(mean, inverse)
        mahal = np.dot(dot_product, mean.T)
        return np.sqrt(mahal.diagonal())
```

Figure 11: Function for Mahalanobis Distance

Question 3

1. Calculate Minkowski distances with r = 1, 2, 10, respectively, and plot the obtained distances.

   (a) Minkowski distances with r = 1

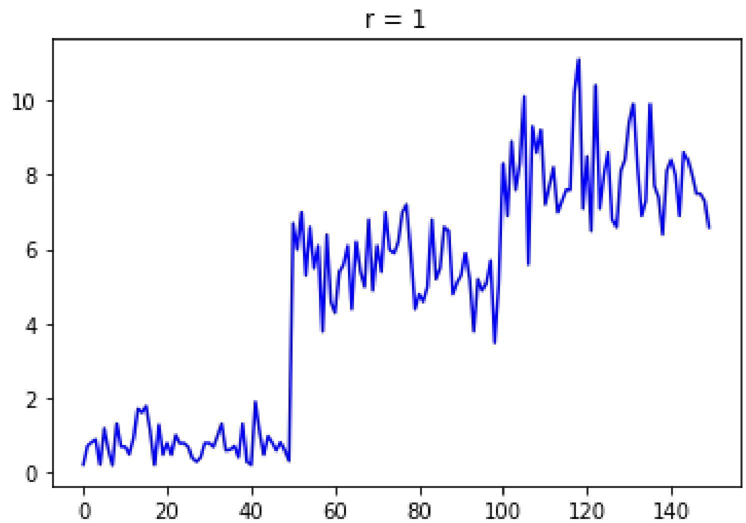

Figure 12: Minkowski distances with r = 1

   (a) Minkowski distances with r = 2

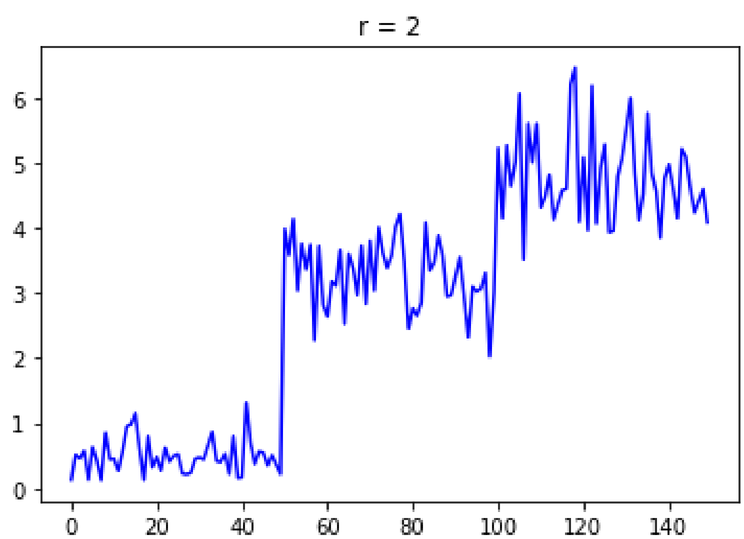

Figure 13: Minkowski distances with r = 2

(a) Minkowski distances with r = 10

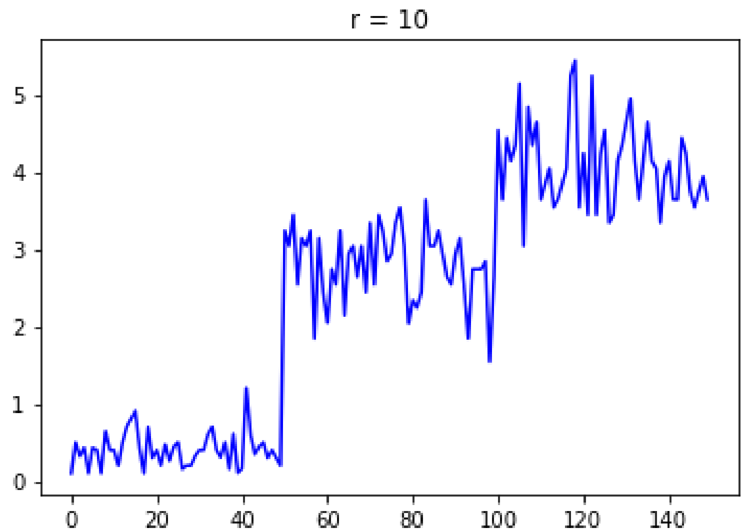

Figure 14: Minkowski distances with r = 10

2. Compare the correlation matrix before and after normalization. If they are the same?

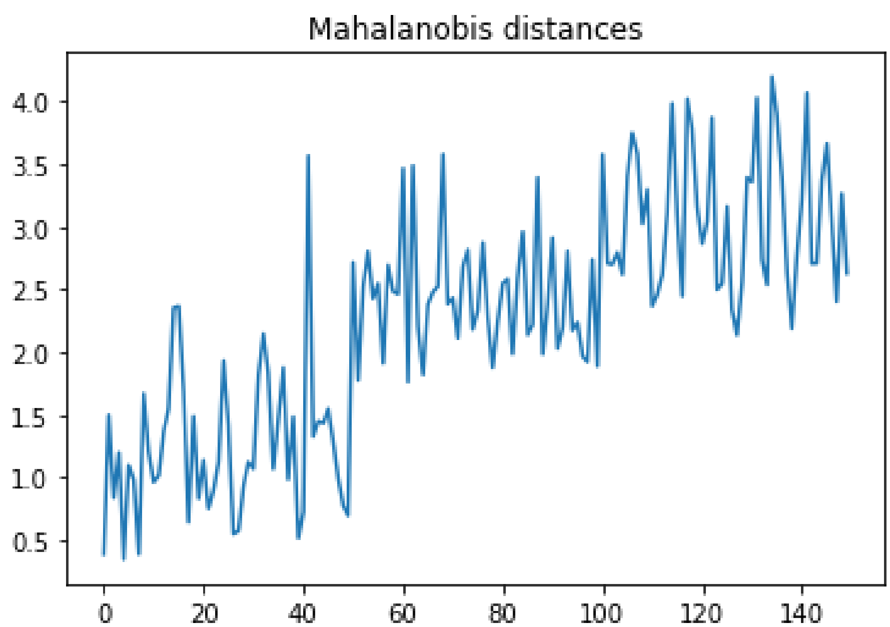

Figure 15: Mahalanobis Distance

Question 4

1. Normalize the feature matrix of the IRIS dataset such that each feature has a mean of 0 and a standard deviation of 1 after normalization.

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)
0	0.803773	0.551609	0.220644	0.031521
1	0.828133	0.507020	0.236609	0.033801
2	0.805333	0.548312	0.222752	0.034269
3	0.800030	0.539151	0.260879	0.034784
4	0.790965	0.569495	0.221470	0.031639
5	0.784175	0.566349	0.246870	0.058087
6	0.780109	0.576603	0.237425	0.050877

Figure 16: Normalized feature matrix of the IRIS dataset

2. Calculate the correlation matrix of the four features after normalization.

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)
sepal length (cm)	1.000000	0.734884	-0.866379	-0.898895
sepal width (cm)	0.734884	1.000000	-0.965108	-0.897886
petal length (cm)	-0.866379	-0.965108	1.000000	0.948639
petal width (cm)	-0.898895	-0.897886	0.948639	1.000000

Figure 17: Correlation matrix of the four features after normalization

3. Make a function for Mahalanobis Distance.

Correlation matrix after normalize data:

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)
sepal length (cm)	1.000000	0.734884	-0.866379	-0.898895
sepal width (cm)	0.734884	1.000000	-0.965108	-0.897886
petal length (cm)	-0.866379	-0.965108	1.000000	0.948639
petal width (cm)	-0.898895	-0.897886	0.948639	1.000000

Correlation matrix before normalize data:

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)
sepal length (cm)	1.000000	-0.117570	0.871754	0.817941
sepal width (cm)	-0.117570	1.000000	-0.428440	-0.366126
petal length (cm)	0.871754	-0.428440	1.000000	0.962865
petal width (cm)	0.817941	-0.366126	0.962865	1.000000

Figure 18: Comparison of correlation of normalized data & original data

Correlation of normalize data and original data aren't same

Question 5

1. Create 2D scatter plots of each pair of the four components

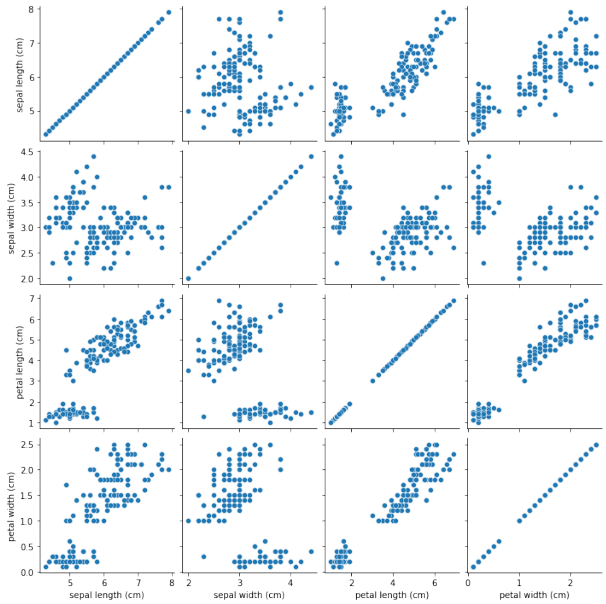

Figure 19: Scatter plots of each pair

2. 3D scatter plot of the first three components

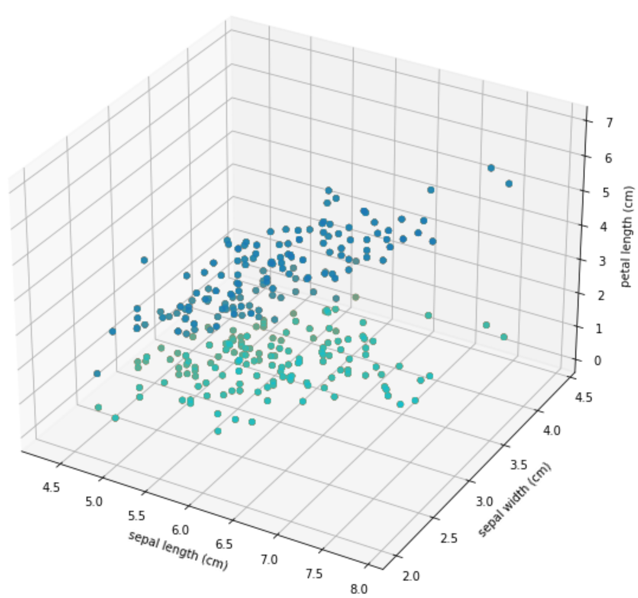

Figure 20: 3D scatter plot

3. Obtain the variance of each component and visualize in a figure plot.

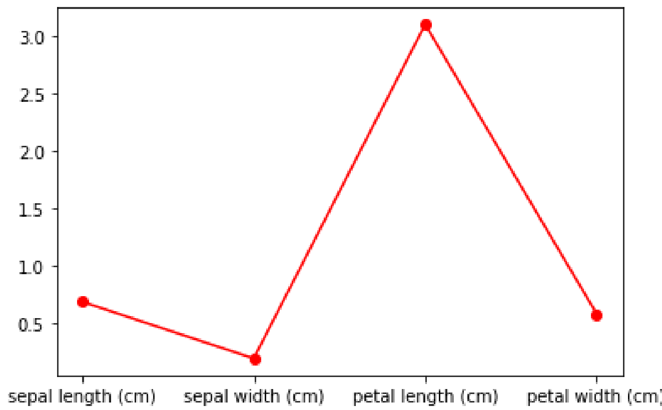

Figure 21: Variance

4. Calculate the correlation matrix of the four components.

	sepal length (cm)	**sepal width (cm)**	**petal length (cm)**	**petal width (cm)**
sepal length (cm)	1.000000	-0.117570	0.871754	0.817941
sepal width (cm)	-0.117570	1.000000	-0.428440	-0.366126
petal length (cm)	0.871754	-0.428440	1.000000	0.962865
petal width (cm)	0.817941	-0.366126	0.962865	1.000000

Figure 22: Correlation matrix of the four components

Question 6

1. Visualize the two time series in one figure plot.

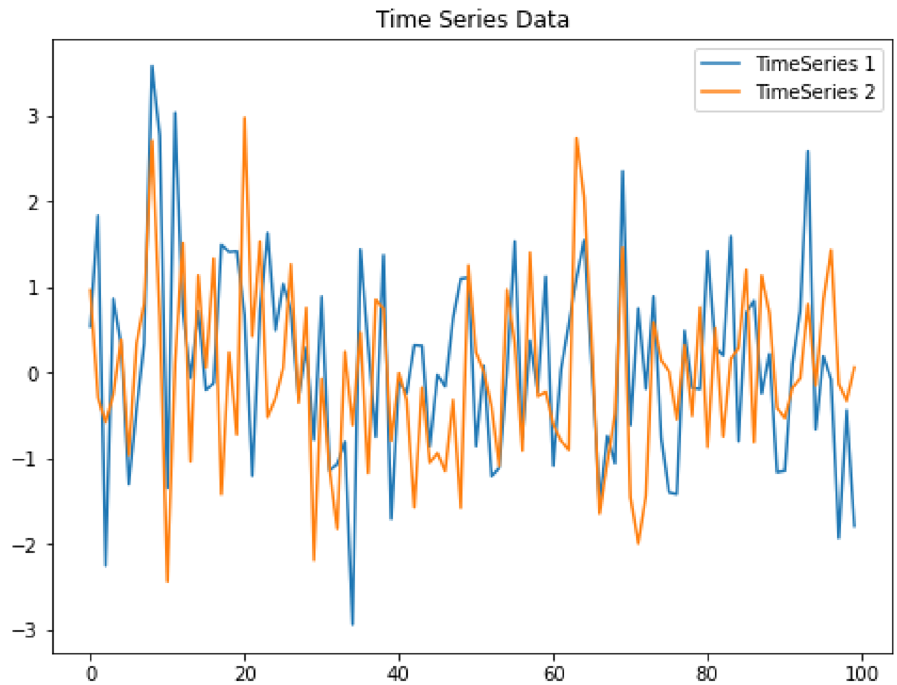

Figure 23: Scatter plots of each pair

2. Calculate the T-statistics distance between the two time series using the function you made in 2.

```
t_distance(time_series['TimeSeries 1'], time_series['TimeSeries 2']) = 1.291758
```

3. Calculate the correlation of the two time series

	TimeSeries 1	TimeSeries 2
TimeSeries 1	1.000000	0.403044
TimeSeries 2	0.403044	1.000000

Figure 24: correlation of the two time series