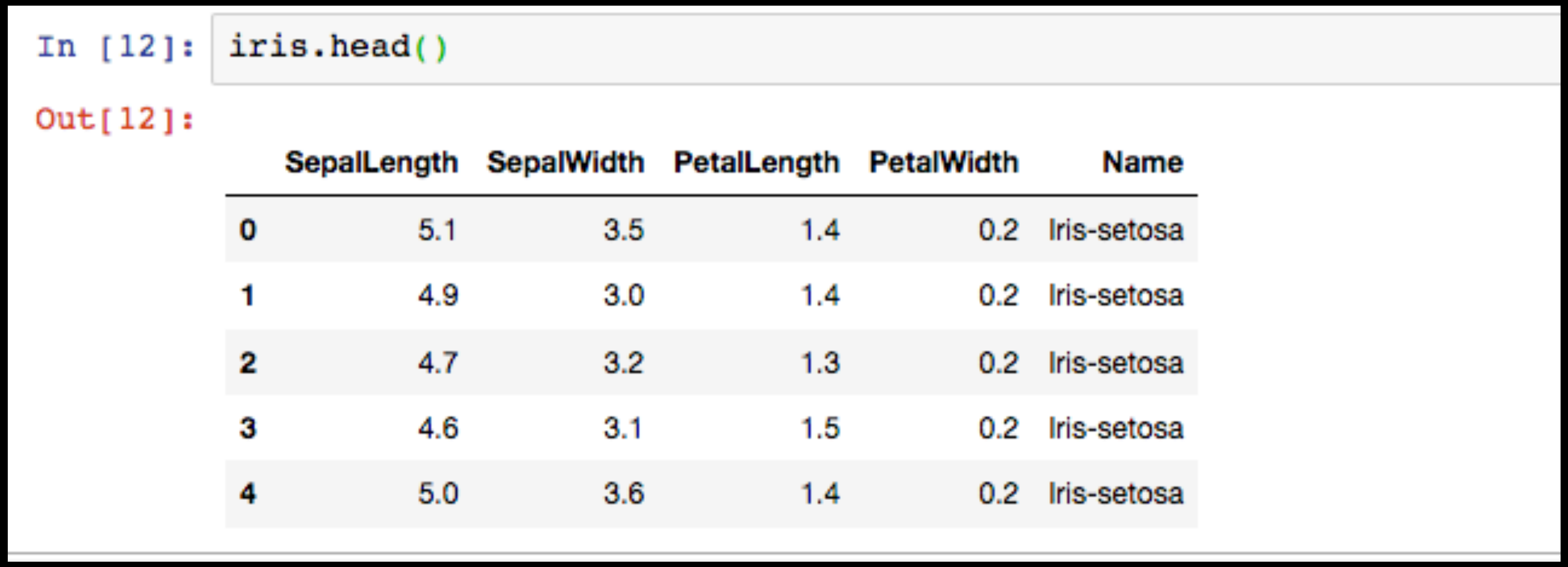
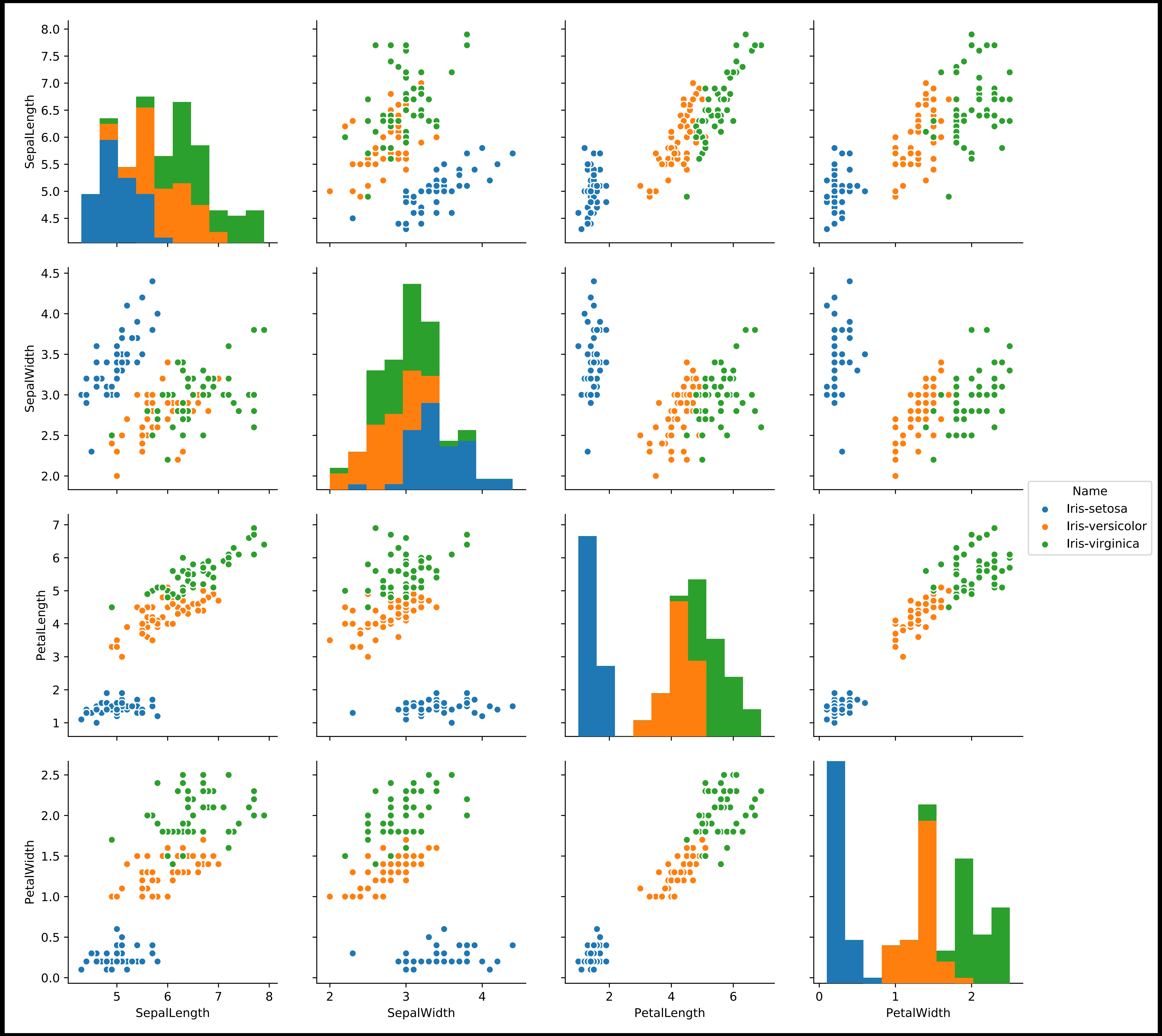
Principle Component Analysis

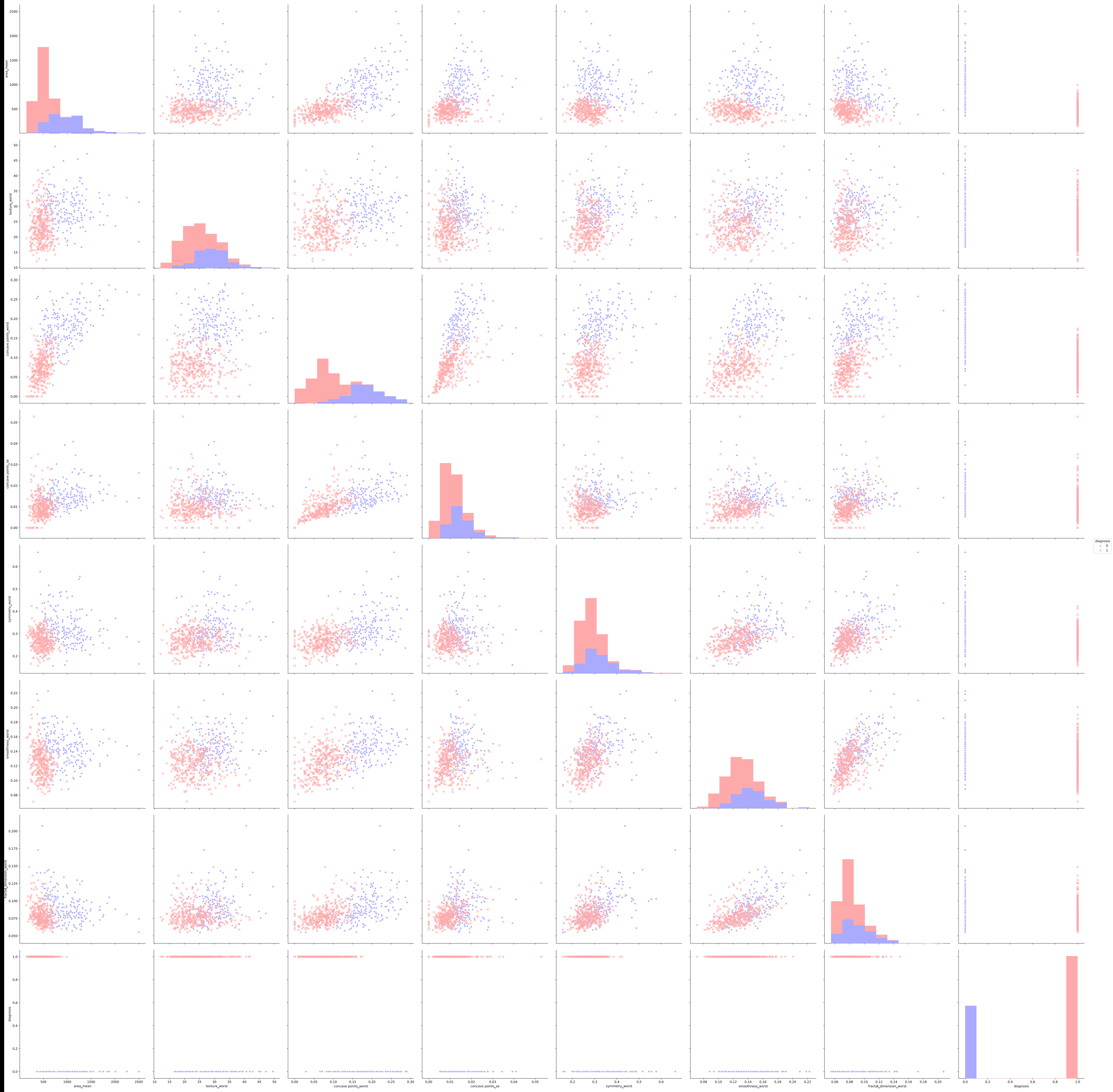


```
In [12]: iris.head()
```

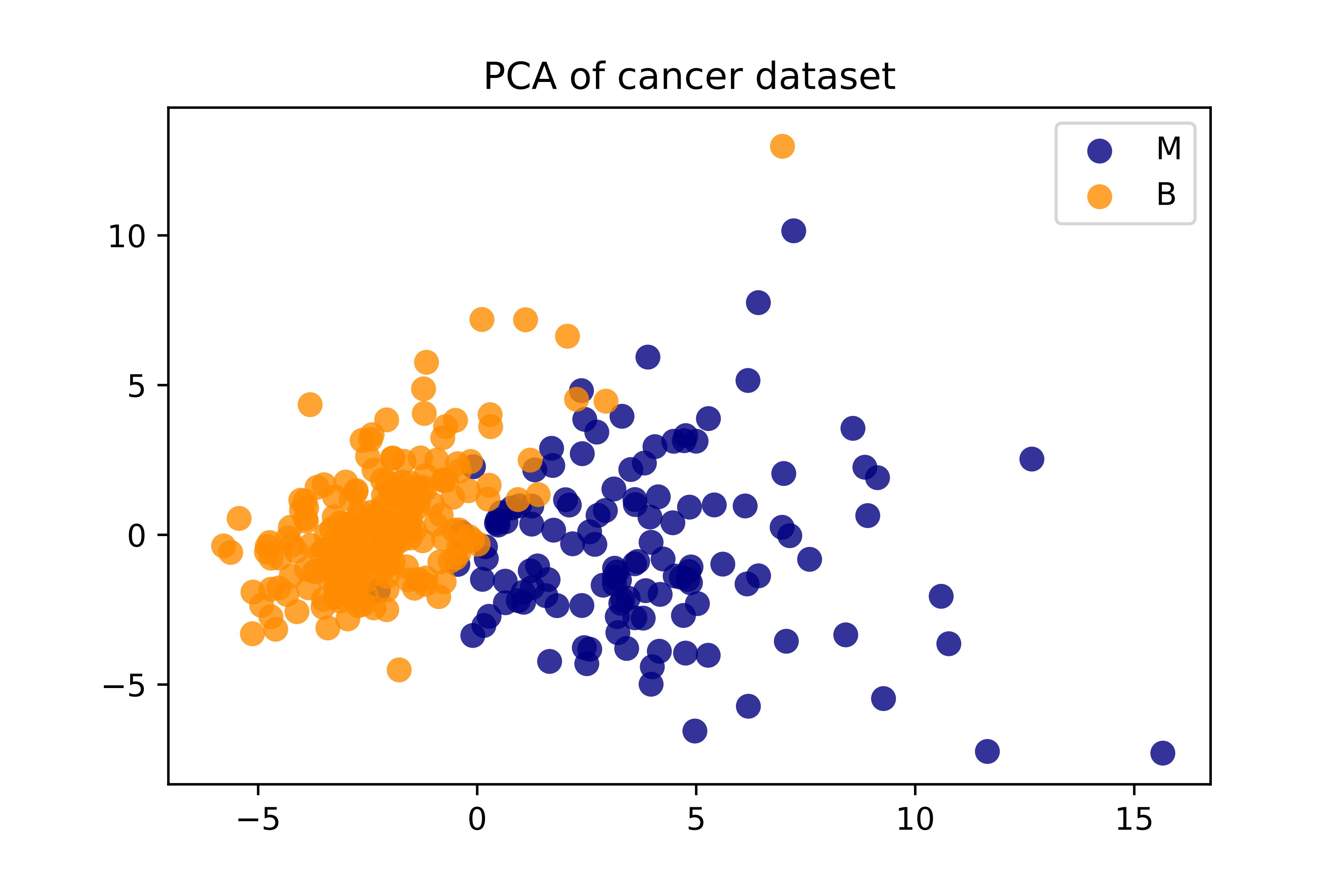
Out[12]:

	SepalLength	SepalWidth	PetalLength	PetalWidth	Name
0	5.1	3.5	1.4	0.2	Iris-setosa
1	4.9	3.0	1.4	0.2	Iris-setosa
2	4.7	3.2	1.3	0.2	Iris-setosa
3	4.6	3.1	1.5	0.2	Iris-setosa
4	5.0	3.6	1.4	0.2	Iris-setosa

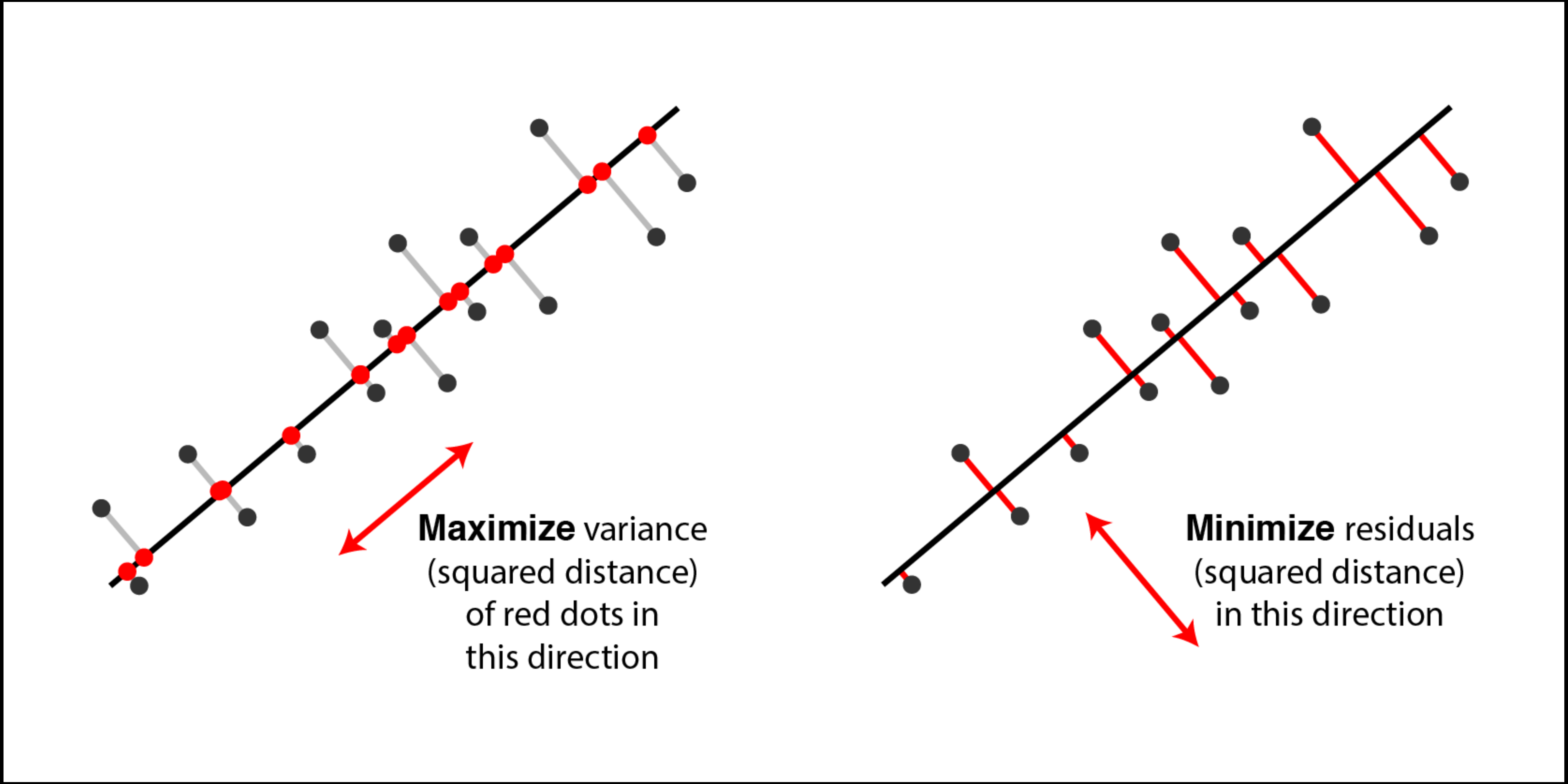




area_mean
texture_worst
concave points_worst
concave points_se
symmetry_worst
smoothness_worst
fractal_dimension_worst
diagnosis
diagnosis
0
1



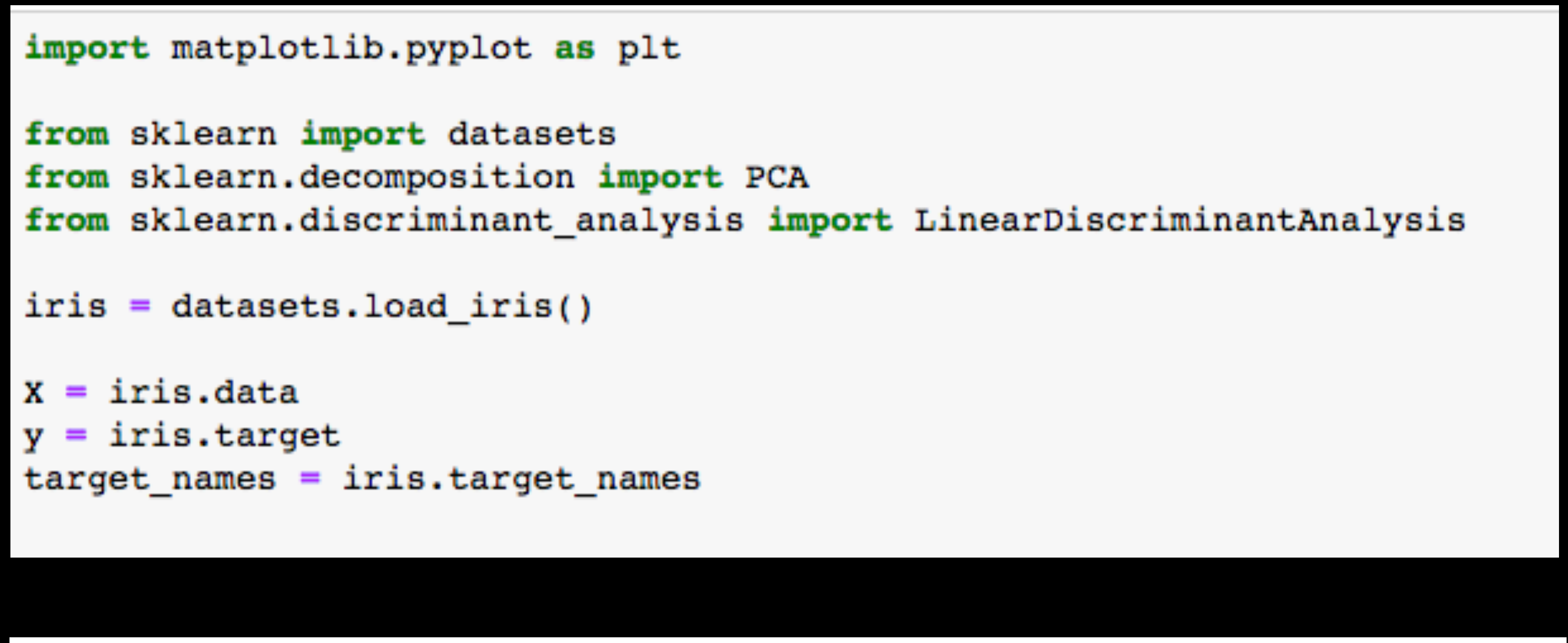
PCA of cancer dataset
M
B
10
5
0
−5
−5
0
5
10
15



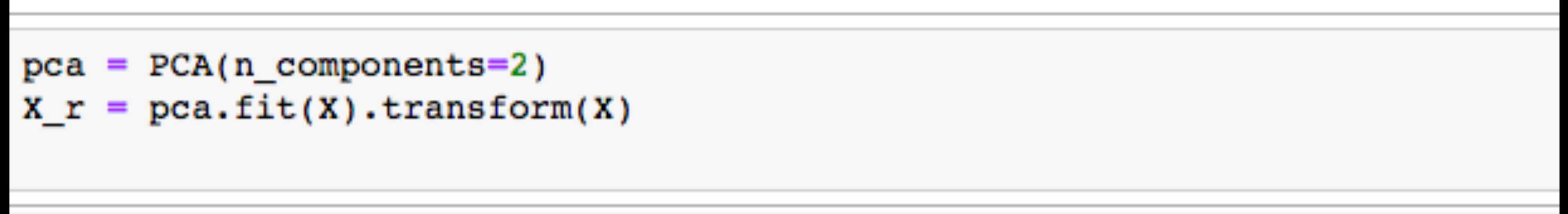
http://alexhwilliams.info/itsneuronalblog/2016/03/27/pca/

PCA algorithm

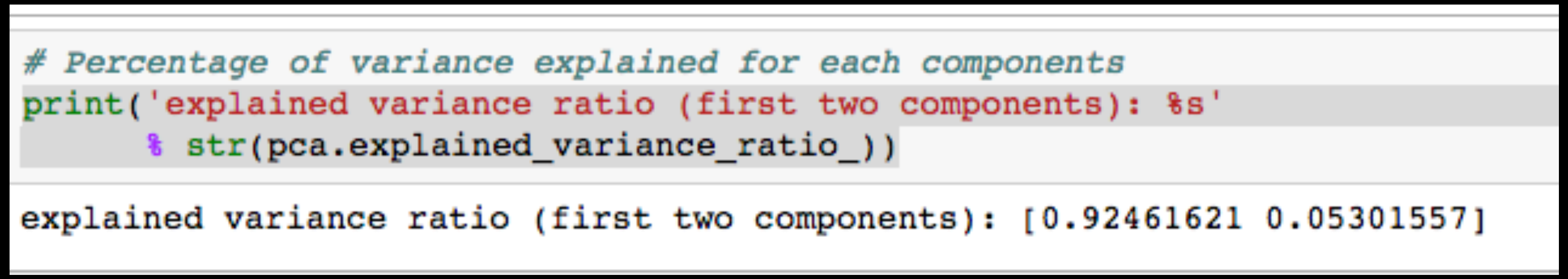
- Calculate covariance matrix of X, Σ
- Calculate the Eigenvectors (U) and Eigenvalues (S) of covariance matrix
- Take first K columns of U to form U_{sub}
- New data given by: $X_{new} = U_{sub}^T X$



```
import matplotlib.pyplot as plt

from sklearn import datasets
from sklearn.decomposition import PCA
from sklearn.discriminant_analysis import LinearDiscriminantAnalysis

iris = datasets.load_iris()

X = iris.data
y = iris.target
target_names = iris.target_names
```

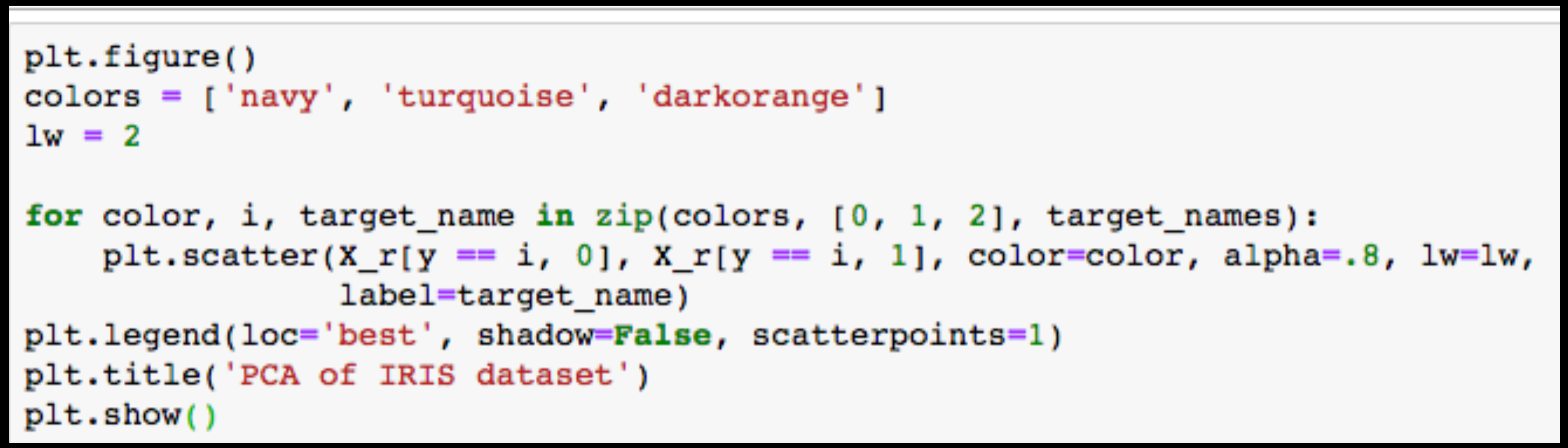


```
pca = PCA(n_components=2)
X_r = pca.fit(X).transform(X)
```

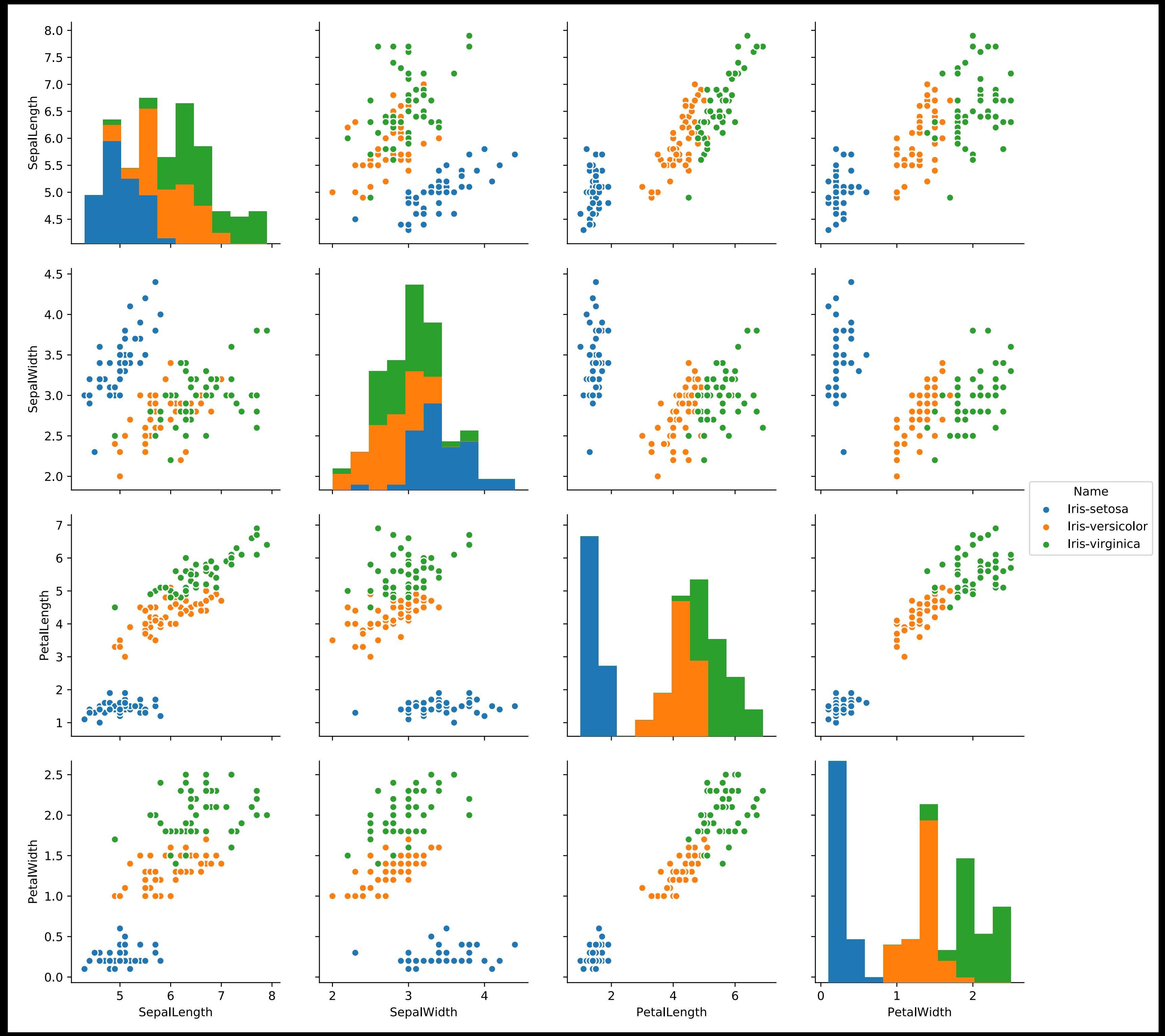


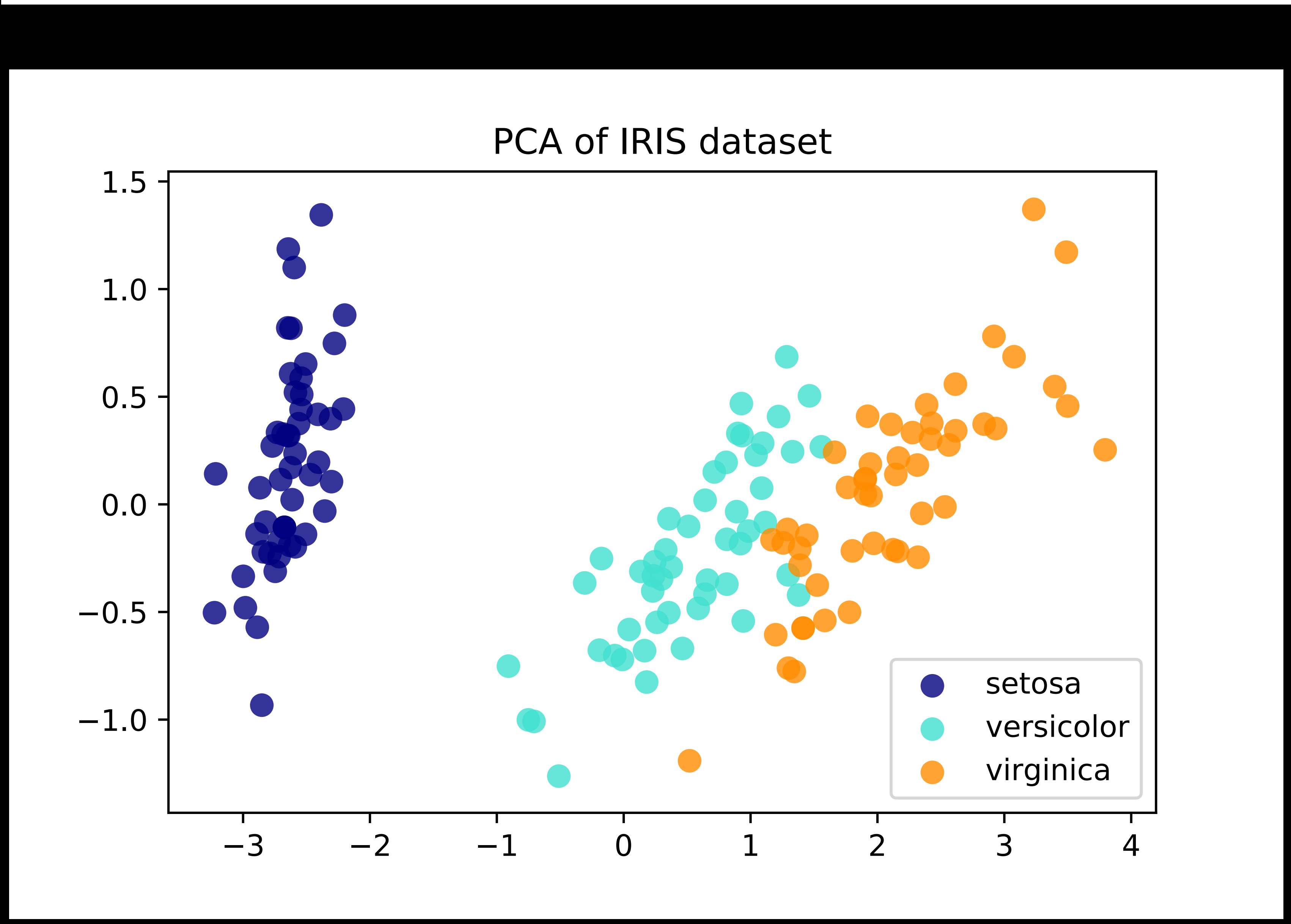
```
# Percentage of variance explained for each components
print('explained variance ratio (first two components): %s'
      % str(pca.explained_variance_ratio_))
```

```
explained variance ratio (first two components): [0.92461621 0.05301557]
```



```
plt.figure()
colors = ['navy', 'turquoise', 'darkorange']
lw = 2

for color, i, target_name in zip(colors, [0, 1, 2], target_names):
    plt.scatter(X_r[y == i, 0], X_r[y == i, 1], color=color, alpha=.8, lw=lw,
                label=target_name)
plt.legend(loc='best', shadow=False, scatterpoints=1)
plt.title('PCA of IRIS dataset')
plt.show()
```


SepalLength
SepalWidth
PetalLength
PetalWidth
SepalLength
SepalWidth
PetalLength
PetalWidth
Name
Iris-setosa
Iris-versicolor
Iris-virginica


PCA of IRIS dataset
setosa
versicolor
virginica