```
import numpy as np
import pandas as pd
```

```
from google.colab import drive
drive.mount('/content/drive')
```

```
Go to this URL in a browser: https://accounts.google.com/o/oauth2/auth?client_id=9473

Enter your authorization code:
··········
Mounted at /content/drive
```

```
data = pd.read_csv('/content/drive/My Drive/3novext/vehicle (1).csv')
```

```
data
```

	compactness	circularity	distance_circularity	radius_ratio	pr.axis_aspect_rat
0	95	48.0	83.0	178.0	7
1	91	41.0	84.0	141.0	5
2	104	50.0	106.0	209.0	6
3	93	41.0	82.0	159.0	6
4	85	44.0	70.0	205.0	10
...	...	...	...	...	
841	93	39.0	87.0	183.0	6
842	89	46.0	84.0	163.0	6
843	106	54.0	101.0	222.0	6
844	86	36.0	78.0	146.0	5
845	85	36.0	66.0	123.0	5

846 rows × 19 columns

```
%matplotlib inline
import numpy as np
import pandas as pd
import seaborn as sns
from scipy import stats
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.preprocessing import LabelEncoder
from sklearn.preprocessing import StandardScaler
from sklearn.decomposition import PCA
```

```
data.shape
```

```
(846, 19)
```

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 846 entries, 0 to 845
Data columns (total 19 columns):
compactness                  846 non-null int64
circularity                  841 non-null float64
distance_circularity         842 non-null float64
radius_ratio                 840 non-null float64
pr.axis_aspect_ratio         844 non-null float64
max.length_aspect_ratio      846 non-null int64
scatter_ratio                845 non-null float64
elongatedness                845 non-null float64
pr.axis_rectangularity       843 non-null float64
max.length_rectangularity    846 non-null int64
scaled_variance              843 non-null float64
scaled_variance.1            844 non-null float64
scaled_radius_of_gyration    844 non-null float64
scaled_radius_of_gyration.1  842 non-null float64
skewness_about               840 non-null float64
skewness_about.1             845 non-null float64
skewness_about.2             845 non-null float64
hollows_ratio                846 non-null int64
class                        846 non-null object
dtypes: float64(14), int64(4), object(1)
memory usage: 125.7+ KB
```

```
data.head()
```

	compactness	circularity	distance_circularity	radius_ratio	pr.axis_aspect_ratio
0	95	48.0	83.0	178.0	72.0
1	91	41.0	84.0	141.0	57.0
2	104	50.0	106.0	209.0	66.0
3	93	41.0	82.0	159.0	63.0
4	85	44.0	70.0	205.0	103.0

```
data['class']=(LabelEncoder().fit_transform(data['class']))
```

```
data
```

	compactness	circularity	distance_circularity	radius_ratio	pr.axis_aspect_rat
0	95	48.0	83.0	178.0	7
1	91	41.0	84.0	141.0	5
2	104	50.0	106.0	209.0	6
3	93	41.0	82.0	159.0	6
4	85	44.0	70.0	205.0	10
...	...	...	...	...	
841	93	39.0	87.0	183.0	6
842	89	46.0	84.0	163.0	6
843	106	54.0	101.0	222.0	6
844	86	36.0	78.0	146.0	5
845	85	36.0	66.0	123.0	5

846 rows × 19 columns

```
for cols in data.columns:
    if (data[cols].dtypes == np.float64):
        data[cols] = data[cols].fillna(data[cols].median())
        data[cols] = data[cols].astype('int64')
    if (data[cols].dtypes == np.int64):
        data[cols] = data[cols].fillna(data[cols].median())
```

```
data.isna().sum()
```

```
compactness                  0
circularity                  0
distance_circularity         0
radius_ratio                 0
pr.axis_aspect_ratio         0
max.length_aspect_ratio      0
scatter_ratio                0
elongatedness                0
pr.axis_rectangularity       0
max.length_rectangularity    0
scaled_variance              0
scaled_variance.1            0
scaled_radius_of_gyration    0
scaled_radius_of_gyration.1  0
skewness_about               0
skewness_about.1             0
skewness_about.2             0
hollows_ratio                0
class                        0
dtype: int64
```

```
data.isnull()
```

	compactness	circularity	distance_circularity	radius_ratio	pr.axis_aspect_rati
0	False	False	False	False	Fals
1	False	False	False	False	Fals
2	False	False	False	False	Fals
3	False	False	False	False	Fals
4	False	False	False	False	Fals
...	...	...	...	...	.
841	False	False	False	False	Fals
842	False	False	False	False	Fals
843	False	False	False	False	Fals
844	False	False	False	False	Fals
845	False	False	False	False	Fals

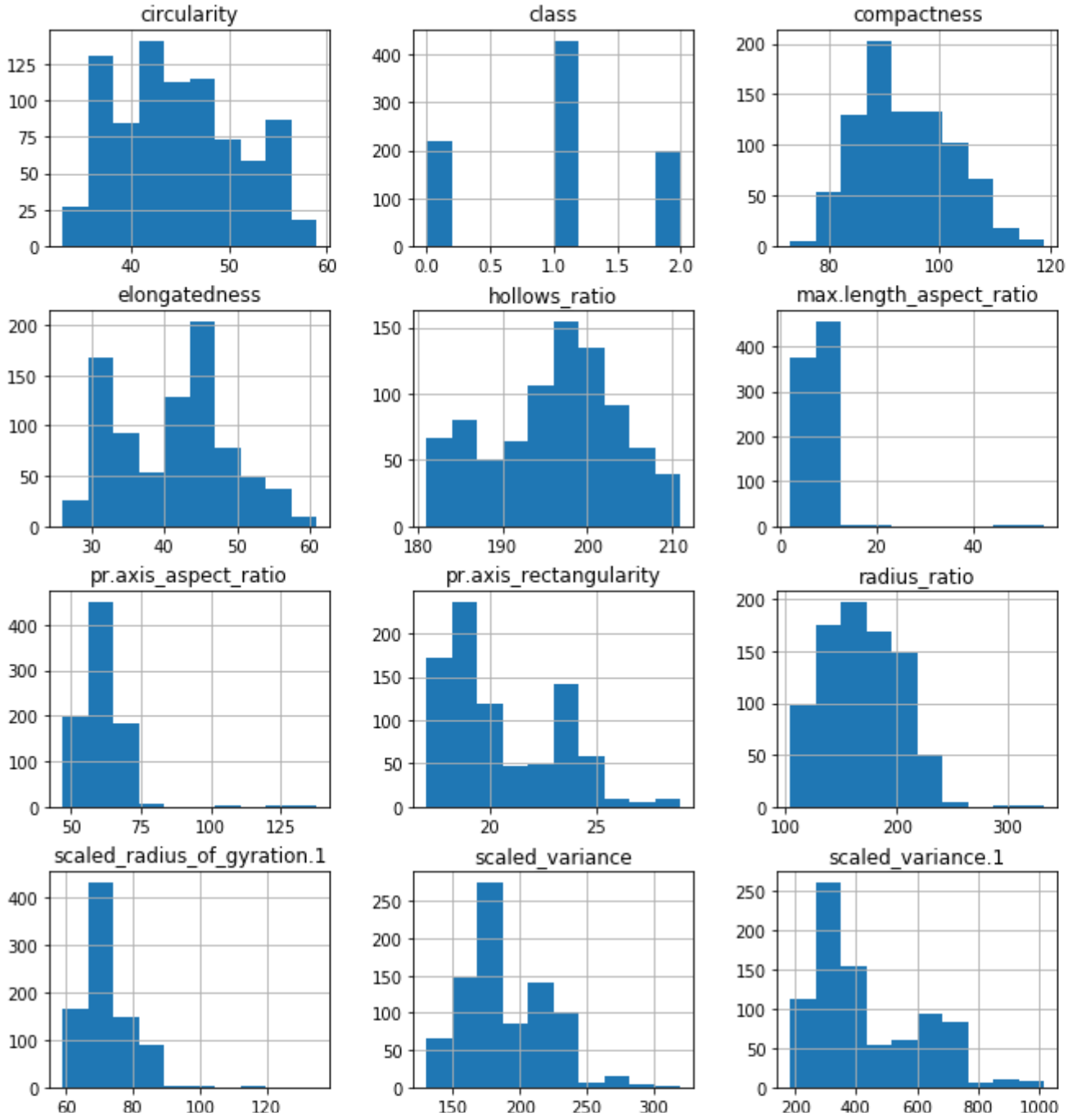
346 rows × 19 columns

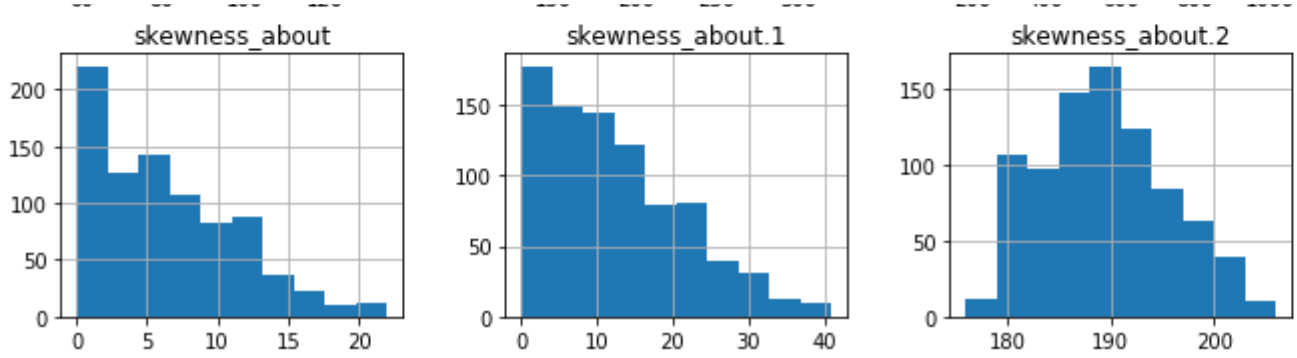
```
data.describe().T
```

	count	mean	std	min	25%	50%	75%
compactness	846.0	93.678487	8.234474	73.0	87.00	93.0	100.00
circularity	846.0	44.823877	6.134272	33.0	40.00	44.0	49.00
distance_circularity	846.0	82.100473	15.741569	40.0	70.00	80.0	98.00
radius_ratio	846.0	168.874704	33.401356	104.0	141.00	167.0	195.00
pr.axis_aspect_ratio	846.0	61.677305	7.882188	47.0	57.00	61.0	65.00
max.length_aspect_ratio	846.0	8.567376	4.601217	2.0	7.00	8.0	10.00
scatter_ratio	846.0	168.887707	33.197710	112.0	147.00	157.0	198.00
elongatedness	846.0	40.936170	7.811882	26.0	33.00	43.0	46.00
pr.axis_rectangularity	846.0	20.580378	2.588558	17.0	19.00	20.0	23.00
max.length_rectangularity	846.0	147.998818	14.515652	118.0	137.00	146.0	159.00
scaled_variance	846.0	188.596927	31.360427	130.0	167.00	179.0	217.00
scaled_variance.1	846.0	439.313239	176.496851	184.0	318.25	363.0	586.75
scaled_radius_of_gyration	846.0	174.705674	32.546330	109.0	149.00	173.0	198.00
scaled_radius_of_gyration.1	846.0	72.440898	7.469112	59.0	67.00	71.0	75.00
skewness_about	846.0	6.361702	4.903244	0.0	2.00	6.0	9.00
skewness_about.1	846.0	12.600473	8.930962	0.0	5.00	11.0	19.00
skewness_about.2	846.0	188.918440	6.152247	176.0	184.00	188.0	193.00
hollows_ratio	846.0	195.632388	7.438797	181.0	190.25	197.0	201.00
class	846.0	0.977541	0.702130	0.0	0.00	1.0	1.00

```
data.hist(figsize=(15,15))
```

```
array([[<matplotlib.axes._subplots.AxesSubplot object at 0x7f2b39cf7978>,
        <matplotlib.axes._subplots.AxesSubplot object at 0x7f2b39c47518>,
        <matplotlib.axes._subplots.AxesSubplot object at 0x7f2b39c77668>,
        <matplotlib.axes._subplots.AxesSubplot object at 0x7f2b39c267b8>],
       [<matplotlib.axes._subplots.AxesSubplot object at 0x7f2b39bdb908>,
        <matplotlib.axes._subplots.AxesSubplot object at 0x7f2b39b8da58>,
        <matplotlib.axes._subplots.AxesSubplot object at 0x7f2b39b42ba8>,
        <matplotlib.axes._subplots.AxesSubplot object at 0x7f2b39b74d30>],
       [<matplotlib.axes._subplots.AxesSubplot object at 0x7f2b39b74d68>,
        <matplotlib.axes._subplots.AxesSubplot object at 0x7f2b39adcf98>,
        <matplotlib.axes._subplots.AxesSubplot object at 0x7f2b39a9e128>,
        <matplotlib.axes._subplots.AxesSubplot object at 0x7f2b39a51278>],
       [<matplotlib.axes._subplots.AxesSubplot object at 0x7f2b39a043c8>,
        <matplotlib.axes._subplots.AxesSubplot object at 0x7f2b39a368d0>,
        <matplotlib.axes._subplots.AxesSubplot object at 0x7f2b399e7e80>,
        <matplotlib.axes._subplots.AxesSubplot object at 0x7f2b399a3470>],
       [<matplotlib.axes._subplots.AxesSubplot object at 0x7f2b39956a20>,
        <matplotlib.axes._subplots.AxesSubplot object at 0x7f2b39905fd0>,
        <matplotlib.axes._subplots.AxesSubplot object at 0x7f2b398c45c0>,
        <matplotlib.axes._subplots.AxesSubplot object at 0x7f2b398f2b70>]],
      dtype=object)
```

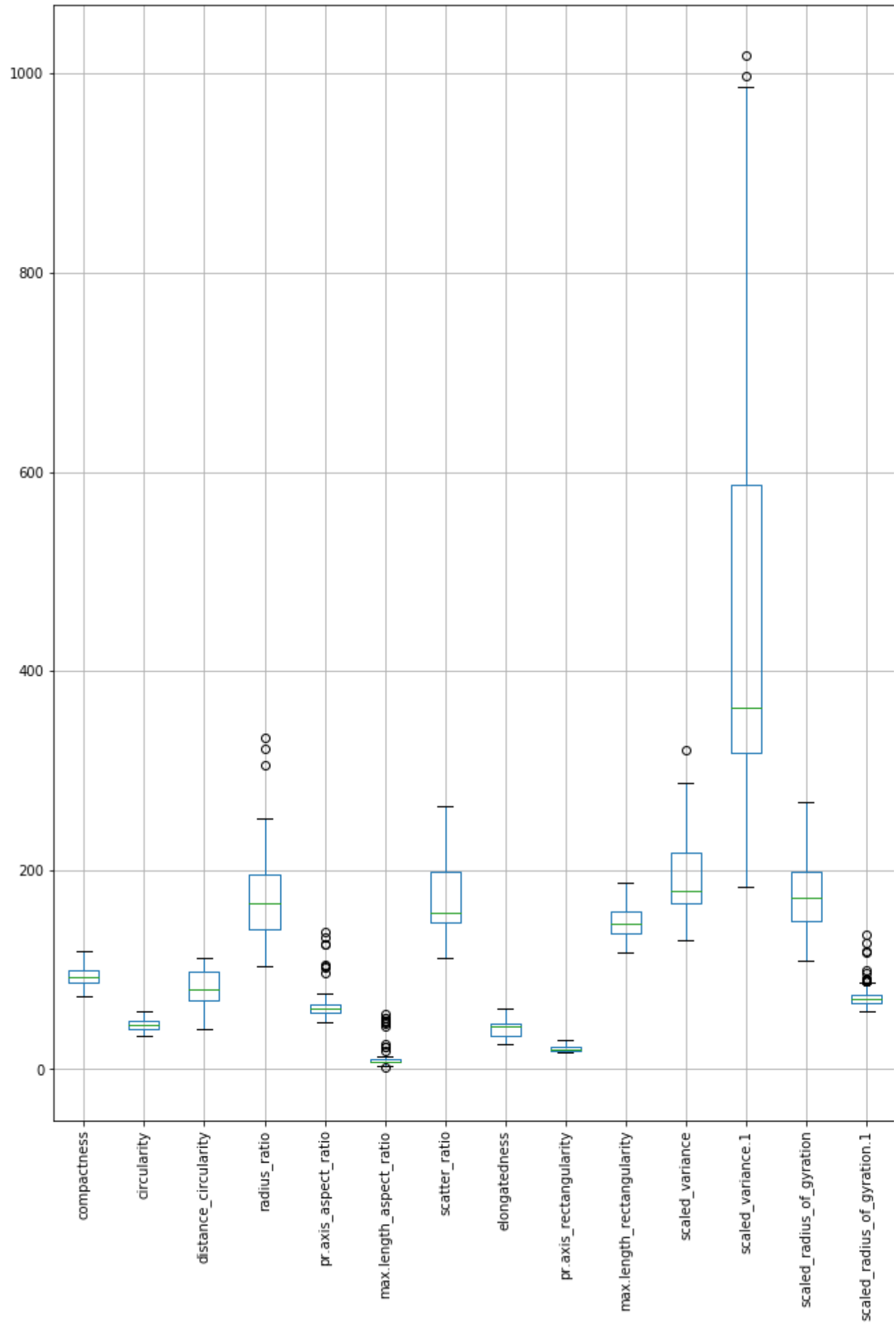




1. Data pre-processing - Understand the data and treat missing values (Use box plot), outliers (5 points)

```
data.boxplot(figsize=(15,15),rot=90,)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f2b3915de80>
```




```
#here we can see these are having outliers
```

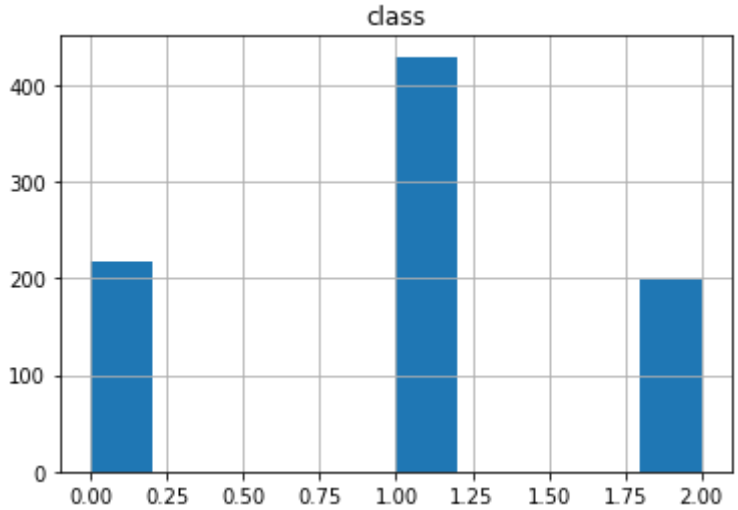
```
#   1. radius_ratio
 #  2. pr.axis_aspect_ratio
  # 3. max.length_aspect_ratio
   #4. scaled_variance
   #5. scaled_variance.1
   #6. scaled_radius_of_gyration.1
   #7. skewness_about
   #8. skewness_about.1
```

```
data.groupby('class').count()
```

	x.length_aspect_ratio	scatter_ratio	elongatedness	pr.axis_rectangularity	max.lengt
	218	218	218	218	
	429	429	429	429	
	199	199	199	199	

```
data.hist('class')
```

```
array([[<matplotlib.axes._subplots.AxesSubplot object at 0x7f2b375897b8>]],
      dtype=object)
```




```
#here class 0 is 218, class 1 is  429  class2 is 199
```

```
data.fillna(data.median(), inplace=True)
```

```
#removing na values with meadian
```

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 846 entries, 0 to 845
Data columns (total 19 columns):
compactness                  846 non-null int64
circularity                  846 non-null int64
distance_circularity         846 non-null int64
radius_ratio                 846 non-null int64
pr.axis_aspect_ratio         846 non-null int64
max.length_aspect_ratio      846 non-null int64
scatter_ratio                846 non-null int64
elongatedness                846 non-null int64
pr.axis_rectangularity       846 non-null int64
max.length_rectangularity    846 non-null int64
scaled_variance              846 non-null int64
scaled_variance.1            846 non-null int64
scaled_radius_of_gyration    846 non-null int64
scaled_radius_of_gyration.1  846 non-null int64
skewness_about               846 non-null int64
skewness_about.1             846 non-null int64
skewness_about.2             846 non-null int64
hollows_ratio                846 non-null int64
class                        846 non-null int64
dtypes: int64(19)
memory usage: 125.7 KB
```

```
data_z=data
data_z.head()
```

	compactness	circularity	distance_circularity	radius_ratio	pr.axis_aspect_ratio
0	95	48	83	178	72
1	91	41	84	141	57
2	104	50	106	209	66
3	93	41	82	159	63
4	85	44	70	205	103

```
def replace(x):
    median, std = x.median(), x.std()
    outliers = (x - median).abs() > 3*std
    x[outliers] = x.median()
    return x
data_c = data_z.apply(lambda x:x.transform(replace))
data_c.head()
```

	compactness	circularity	distance_circularity	radius_ratio	pr.axis_aspect_ratio
0	95	48	83	178	72
1	91	41	84	141	57
2	104	50	106	209	66
3	93	41	82	159	63
4	85	44	70	205	61

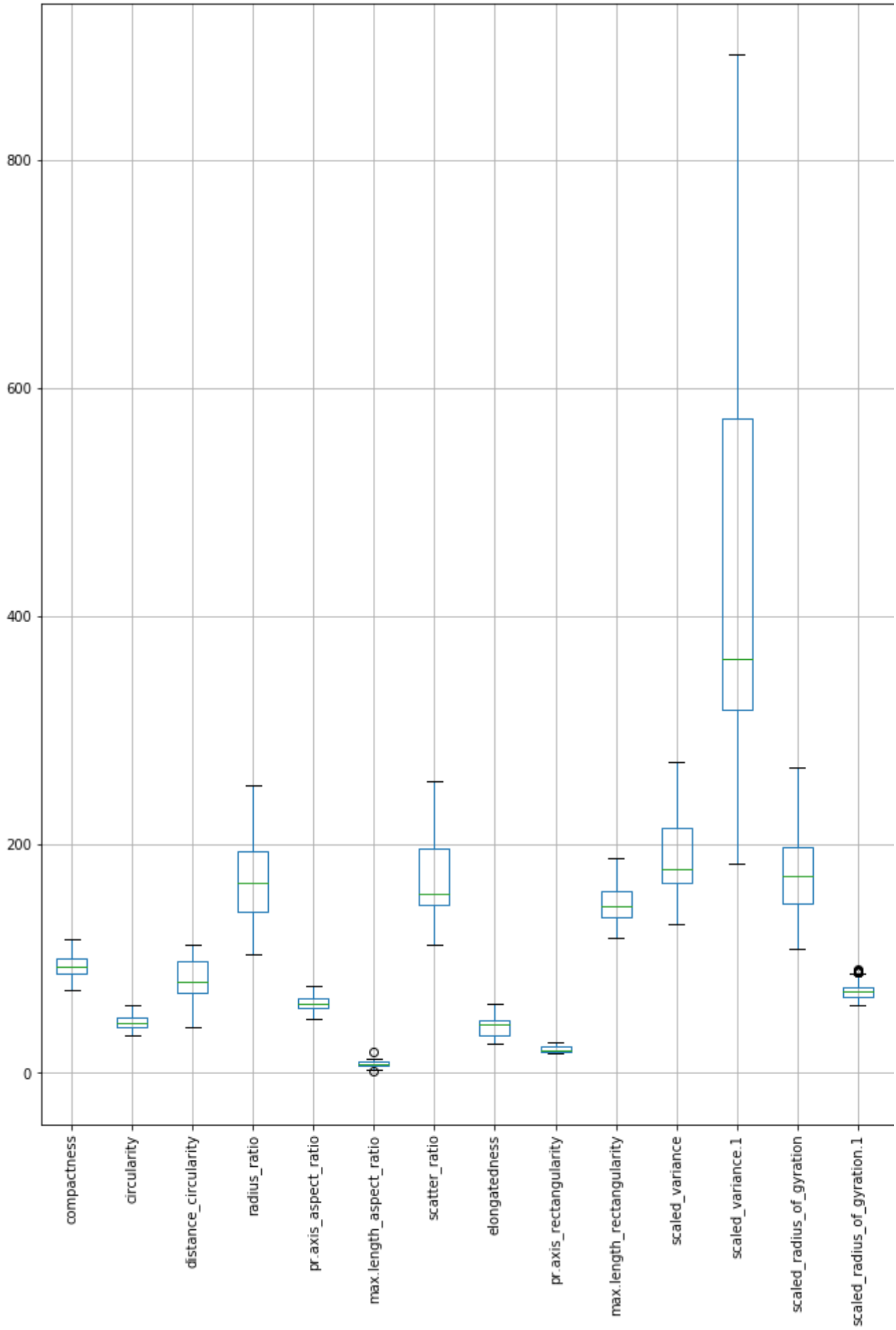
```
data_c.columns=data.columns
```

```
data_c.head()
```

	compactness	circularity	distance_circularity	radius_ratio	pr.axis_aspect_ratio
0	95	48	83	178	72
1	91	41	84	141	57
2	104	50	106	209	66
3	93	41	82	159	63
4	85	44	70	205	61

```
data_c.boxplot(figsize=(15,15),rot=90)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f2b374bfac8>
```




```
#2. Understanding the attributes - Find relationship between different attributes (Indepen
```

```
X = data_c[data_c.columns[1:-1]]
y = data_c["class"]

from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.30, random_state=1)

print(X_train.shape)
print(X_test.shape)
print(y_train.shape)
print(y_test.shape)
```

```
(592, 17)
(254, 17)
(592,)
(254,)
```

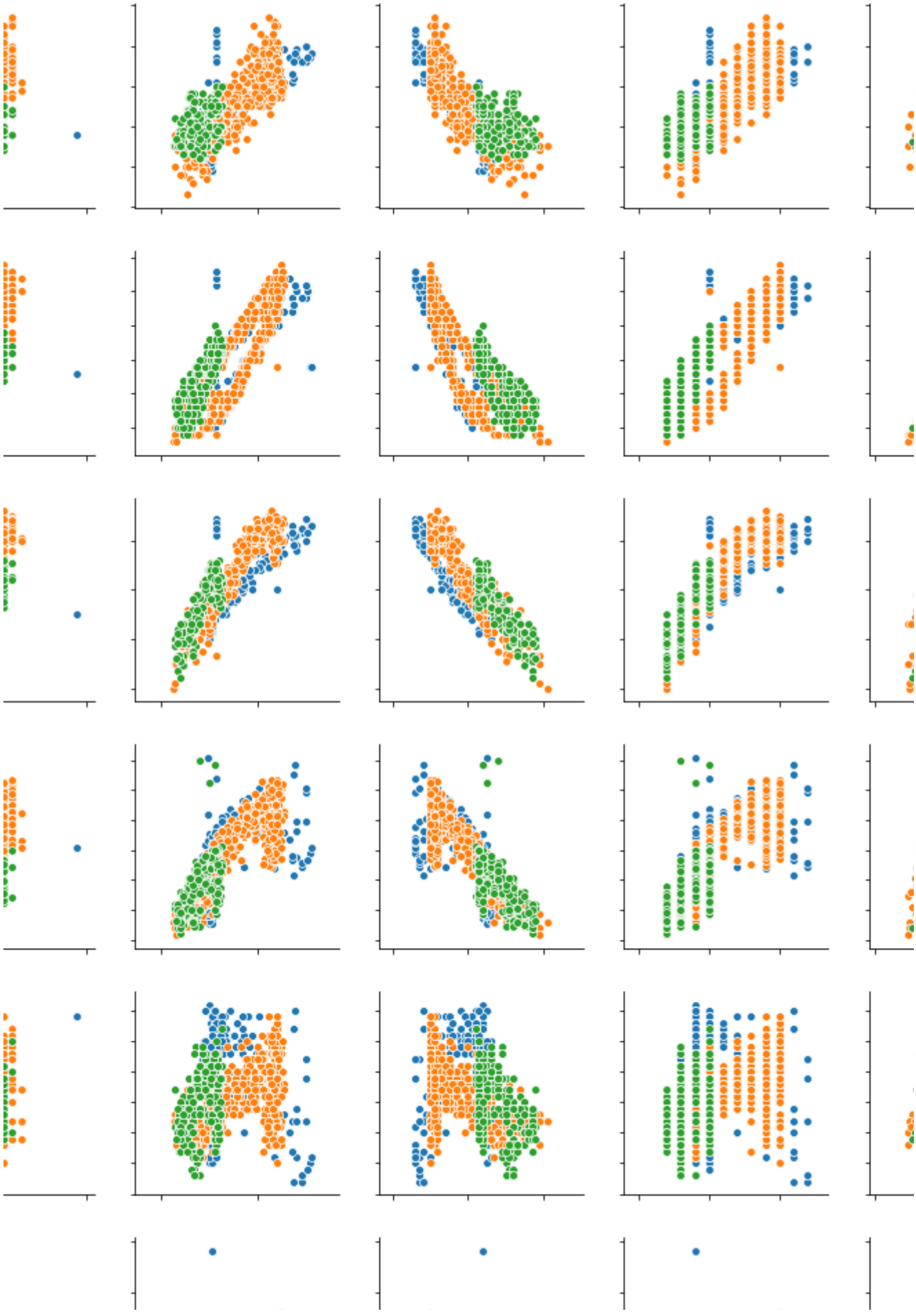
```
train_corr=X_train.corr().style.background_gradient()
train_corr
```

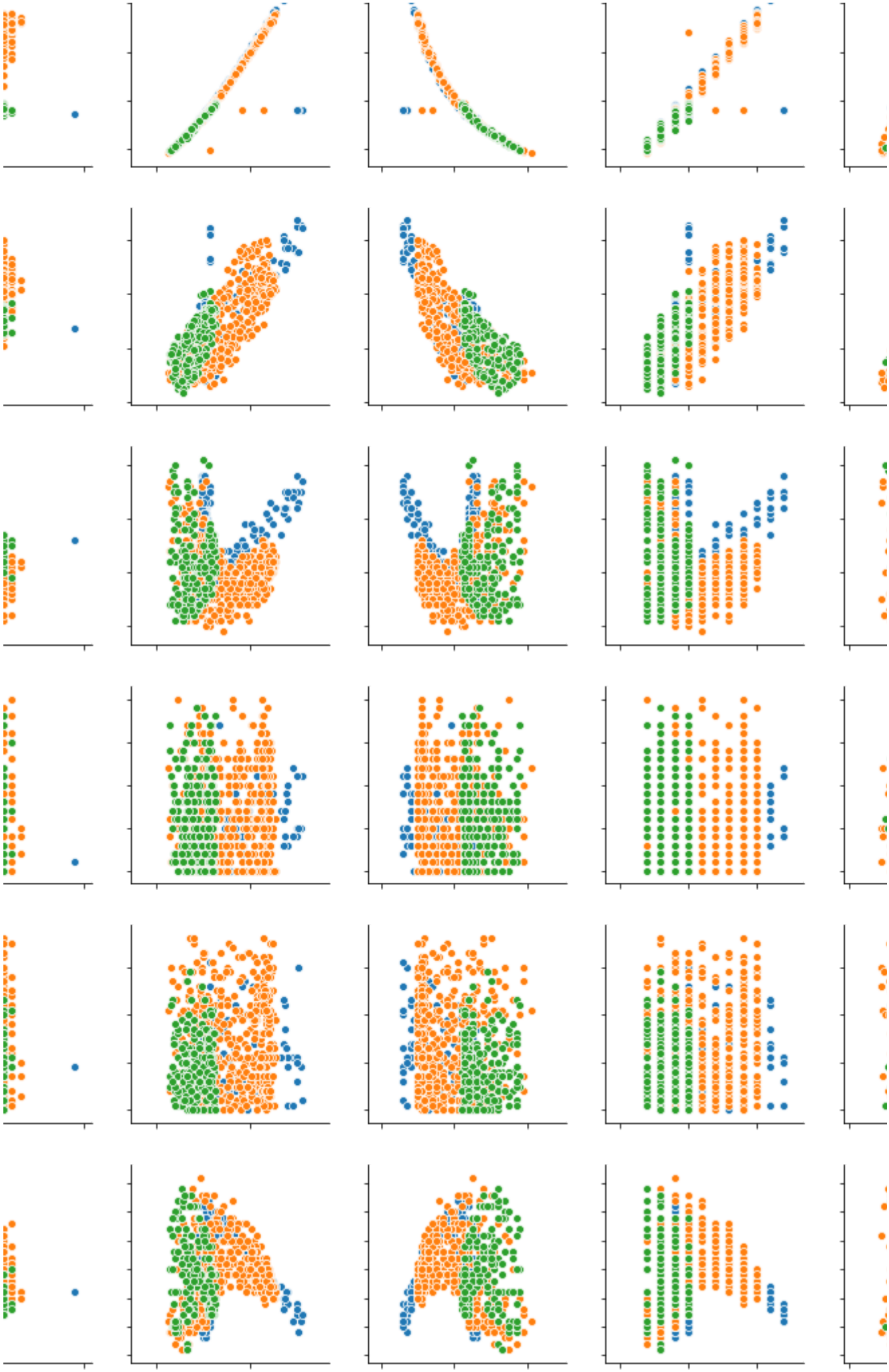
	circularity	distance_circularity	radius_ratio	pr.axis_aspect_ratio	ma
circularity	1	0.793099	0.652182	0.221139	0.5
distance_circularity	0.793099	1	0.808928	0.260097	0.6
radius_ratio	0.652182	0.808928	1	0.65978	0.4
pr.axis_aspect_ratio	0.221139	0.260097	0.65978	1	0.1
max.length_aspect_ratio	0.524865	0.629202	0.451437	0.167362	1
scatter_ratio	0.819132	0.891685	0.784896	0.23268	0.4
elongatedness	-0.823255	-0.910627	-0.837132	-0.303455	-0.
pr.axis_rectangularity	0.82952	0.881099	0.773479	0.231232	0.5
max.length_rectangularity	0.958813	0.776585	0.58817	0.157883	0.6
scaled_variance	0.78582	0.855851	0.798985	0.267762	0.4
scaled_variance.1	0.807244	0.864274	0.77248	0.248935	0.5
scaled_radius_of_gyration	0.923045	0.706822	0.56355	0.162586	0.3
scaled_radius_of_gyration.1	0.0506523	-0.255104	-0.40835	-0.353909	-0.
skewness_about	0.144418	0.0955134	0.0432275	-0.0283747	0.0
skewness_about.1	-0.00445781	0.253311	0.171447	-0.0272137	0.1
skewness_about.2	-0.0997845	0.158252	0.406147	0.41034	0.1
hollows_ratio	0.0447984	0.337156	0.489663	0.431854	0.4

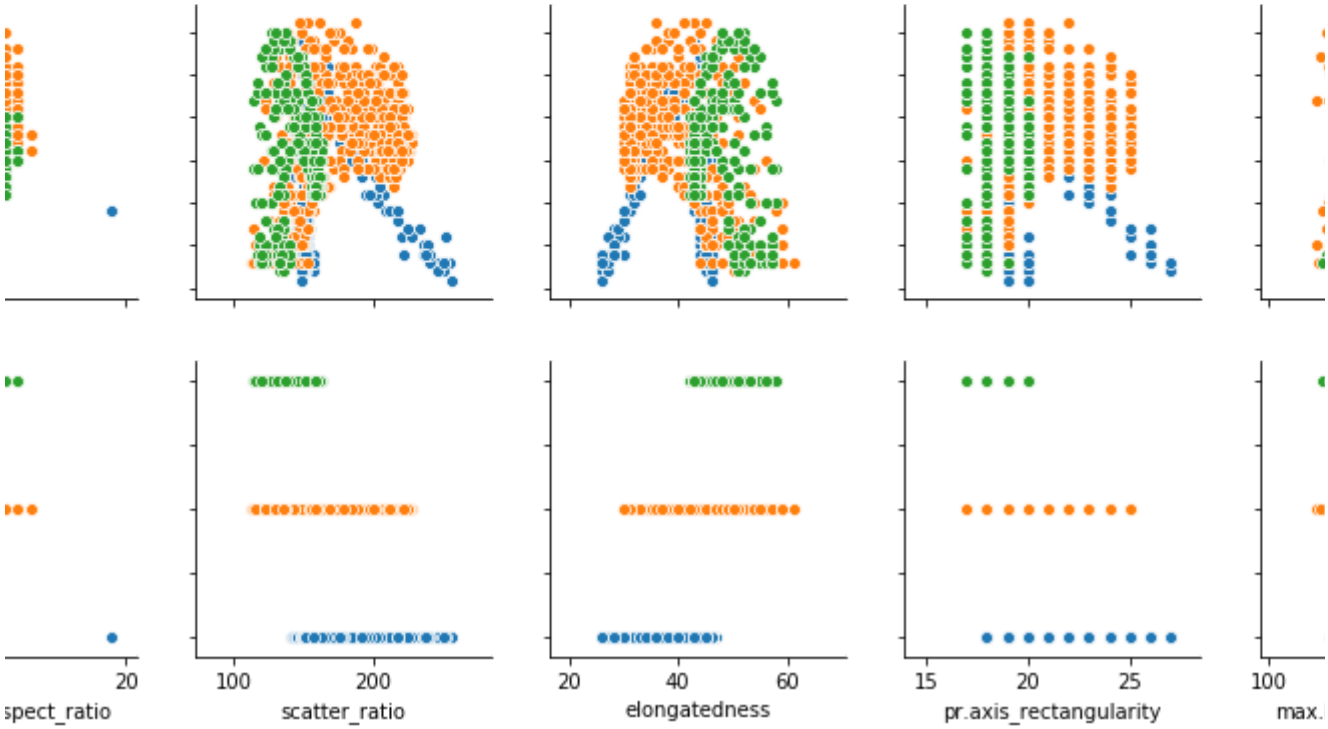
```
sns.pairplot(data=data_c,hue='class')
```

```
e

_scalars
```








```
sc = StandardScaler()


X_train_std = sc.fit_transform(X_train)  # scale training and test data independently to p
X_test_std = sc.fit_transform(X_test)


cov_matrix = np.cov(X_train_std.T)

print('Covariance Matrix \n%s', cov_matrix)
```

```
Covariance Matrix
%s [[ 1.00169205  0.79444139  0.65328602  0.22151356  0.52575315  0.82051801
  -0.82464792  0.83092334  0.96043508  0.7871496   0.80861038  0.92460661
   0.050738    0.14466196 -0.00446536 -0.09995329  0.04487424]
 [ 0.79444139  1.00169205  0.81029723  0.26053716  0.63026623  0.89319384
  -0.91216733  0.88259005  0.77789928  0.85729881  0.86573654  0.70801813
  -0.25553609  0.09567503  0.25373959  0.15851979  0.33772611]
 [ 0.65328602  0.81029723  1.00169205  0.66089651  0.45220119  0.78622451
  -0.83854887  0.77478728  0.5891655   0.80033657  0.77378732  0.56450366
  -0.40904045  0.04330061  0.17173707  0.4068339   0.49049122]
 [ 0.22151356  0.26053716  0.66089651  1.00169205  0.16764489  0.23307354
  -0.3039684   0.23162286  0.15815049  0.26821479  0.24935617  0.1628616
  -0.35450832 -0.02842276 -0.02725977  0.41103437  0.43258482]
 [ 0.52575315  0.63026623  0.45220119  0.16764489  1.00169205  0.48833365
  -0.46811998  0.50764678  0.61231774  0.45453104  0.50805029  0.35299492
  -0.37264761  0.04750333  0.12795609  0.10146741  0.41406278]
 [ 0.82051801  0.89319384  0.78622451  0.23307354  0.48833365  1.00169205
  -0.95301181  0.97571768  0.79303894  0.9321061   0.95473187  0.7707133
  -0.04622871  0.07232571  0.18300995  0.03573341  0.1537995 ]
 [-0.82464792 -0.91216733 -0.83854887 -0.3039684  -0.46811998 -0.95301181
   1.00169205 -0.92762871 -0.78098218 -0.91930965 -0.9163748  -0.77506793
   0.09150698 -0.04777072 -0.17072325 -0.11439267 -0.21233895]
 [ 0.83092334  0.88259005  0.77478728  0.23162286  0.50764678  0.97571768
  -0.92762871  1.00169205  0.79225055  0.9296956   0.96282923  0.75237624
  -0.05285631  0.07892733  0.19889694  0.02415209  0.15519941]
 [ 0.96043508  0.77789928  0.5891655   0.15815049  0.61231774  0.79303894
  -0.78098218  0.79225055  1.00169205  0.72709125  0.7665788   0.86833362
   0.03378529  0.13523682  0.00493555 -0.09871304  0.07329861]
 [ 0.7871496   0.85729881  0.80033657  0.26821479  0.45453104  0.9321061
  -0.91930965  0.9296956   0.72709125  1.00169205  0.94534085  0.72709031
  -0.08577374  0.0334783   0.17941008  0.07216444  0.16449536]
 [ 0.80861038  0.86573654  0.77378732  0.24935617  0.50805029  0.95473187
  -0.9163748   0.96282923  0.7665788   0.94534085  1.00169205  0.71898623
  -0.09367325  0.07063485  0.19205003  0.07164213  0.19686906]
 [ 0.92460661  0.70801813  0.56450366  0.1628616   0.35299492  0.7707133
  -0.77506793  0.75237624  0.86833362  0.72709031  0.71898623  1.00169205
   0.22332673  0.17835774 -0.05408045 -0.23296074 -0.13400811]
 [ 0.050738   -0.25553609 -0.40904045 -0.35450832 -0.37264761 -0.04622871
   0.09150698 -0.05285631  0.03378529 -0.08577374 -0.09367325  0.22332673
   1.00169205 -0.01427516 -0.13182476 -0.84444728 -0.91878821]
 [ 0.14466196  0.09567503  0.04330061 -0.02842276  0.04750333  0.07232571
  -0.04777072  0.07892733  0.13523682  0.0334783   0.07063485  0.17835774
  -0.01427516  1.00169205 -0.04379829  0.06526551  0.03082379]
 [-0.00446536  0.25373959  0.17173707 -0.02725977  0.12795609  0.18300995
  -0.17072325  0.19889694  0.00493555  0.17941008  0.19205003 -0.05408045
  -0.13182476 -0.04379829  1.00169205  0.06909116  0.19009038]
 [-0.09995329  0.15851979  0.4068339   0.41103437  0.10146741  0.03573341
  -0.11439267  0.02415209 -0.09871304  0.07216444  0.07164213 -0.23296074
  -0.84444728  0.06526551  0.06909116  1.00169205  0.8938146 ]
 [ 0.04487424  0.33772611  0.49049122  0.43258482  0.41406278  0.1537995
  -0.21233895  0.15519941  0.07329861  0.16449536  0.19686906 -0.13400811
  -0.91878821  0.03082379  0.19009038  0.8938146   1.00169205]]
```

```
eig_vals, eig_vecs = np.linalg.eig(cov_matrix)
print('Eigen Vectors \n%s', eig_vecs)
print('\n Eigen Values \n%s', eig_vals)
```

```
Eigen Vectors
%s [[-2.98229293e-01  1.47513885e-01  1.71833249e-01  4.47964556e-02
   1.19816367e-01 -8.93897643e-02  3.73969950e-01 -1.50407264e-01
  -1.82959097e-01 -5.64118632e-02 -5.38123538e-01  2.11581901e-01
  -6.96922418e-02 -4.33939542e-01 -2.46294539e-01 -2.02473090e-01
   8.44164360e-02]
 [-3.13984503e-01 -3.63557422e-02 -9.88403043e-02  9.92572013e-02
   3.45196371e-02  2.43525895e-02 -2.42064469e-02  5.74289544e-01
   2.05878014e-01 -2.30792739e-01 -2.05690061e-02  9.91116145e-02
  -5.88669414e-01  8.22378518e-02 -5.60616004e-02  1.48812884e-01
   2.52609076e-01]
 [-2.87443344e-01 -1.84789439e-01  4.87863303e-02 -2.33875794e-01
  -1.70791911e-01 -1.06862840e-01 -8.32545037e-02  3.10426652e-01
  -4.73349658e-02  7.35862653e-02  1.29859837e-02  3.29065612e-01
   5.19852011e-01  2.50232942e-01 -5.23863759e-02 -3.37223653e-01
   3.41804796e-01]
 [-1.18372913e-01 -2.60982708e-01  2.83078234e-01 -5.10733244e-01
  -2.01943819e-01 -6.14248112e-01 -1.15889499e-01 -1.44238690e-01
   1.52957173e-02 -9.25412057e-02  5.63792637e-03 -9.35799527e-02
  -2.26410035e-01 -6.30125826e-02  2.58485198e-02  1.79915950e-01
  -1.36821451e-01]
 [-2.02470016e-01 -1.07749831e-01 -2.84159160e-02  3.46933964e-01
   5.84558288e-01 -3.89704333e-01 -4.00287172e-01 -3.27191095e-02
   1.25342211e-01  2.99357537e-01 -3.60913516e-02 -9.36510076e-02
   1.45308715e-01 -3.65101713e-02 -1.46534201e-01  1.01000636e-01
   4.27468154e-02]
 [-3.18865961e-01  6.66656139e-02 -9.07683366e-02 -3.29568873e-02
  -8.30962311e-02  1.66282640e-01 -1.52570786e-01 -7.00425098e-02
   1.65103070e-01 -2.38235013e-01 -5.93207092e-01 -2.28524749e-01
   2.70754052e-01  2.38345618e-01  3.11972152e-01  2.97597806e-01
  -1.32950447e-01]
 [ 3.19222926e-01 -2.87141292e-02  5.57060405e-02  8.95991525e-02
   9.22649623e-02 -1.41940033e-01  7.13962646e-03 -2.28973188e-01
  -4.00133972e-01 -1.44933347e-01 -1.70662929e-01 -2.95561369e-02
  -1.06229735e-01  3.19317290e-01  1.16314224e-01  1.86077764e-01
   6.57637489e-01]
 [-3.17742103e-01  6.60012369e-02 -1.00974288e-01 -1.26787149e-02
  -7.16418905e-02  1.35286615e-01 -1.84894004e-01 -2.31145329e-01
  -1.25508730e-02 -3.81874080e-01  4.13687644e-01 -1.41862571e-01
   1.94304711e-01 -4.93453491e-01  1.43261386e-02  1.50404647e-01
   3.63253346e-01]
 [-2.88002227e-01  1.38939514e-01  1.49942310e-01  1.33422176e-01
   2.38356649e-01 -1.20771688e-01  3.66504783e-01 -2.81055591e-01
   1.23580660e-01 -1.66121730e-01  3.36625532e-01  4.29227636e-01
   1.63503184e-02  3.26393995e-01  3.00727507e-01  1.19756249e-01
  -1.44041572e-01]
 [-3.10416308e-01  4.20515636e-02 -1.01595465e-01 -9.54007261e-02
  -1.03708407e-01  1.90895933e-01 -2.08324046e-01 -7.81790267e-02
  -3.93281602e-01  5.75095963e-01  1.27298947e-02  1.88828815e-01
  -2.57384565e-01 -1.14808637e-01  4.18255977e-01  9.94487185e-02
   1.62105457e-02]
 [-3.15067911e-01  4.00116042e-02 -1.00209794e-01 -2.25969541e-02
  -7.06101401e-02  1.70577564e-01 -2.28009789e-01 -3.85924727e-01
  -1.73859133e-01 -8.64838813e-02  7.16283313e-02 -1.25973679e-01
  -2.48150127e-01  4.37312879e-01 -4.99582828e-01 -2.81368559e-01
  -1.15286811e-01]
 [-2.69107375e-01  2.32795376e-01  2.08748339e-01 -8.14899194e-03
  -8.27639094e-03 -6.69268297e-02  4.35576787e-01  2.33194984e-01
  -1.84356920e-01  2.58145653e-01  1.82870099e-01 -6.26562001e-01
   1.08921749e-01  1.43434977e-01 -5.92895395e-02  1.11084089e-01
   7.84218782e-02]
```