PCA

September 30, 2020

```python
[2]: import pandas as pd
     from sklearn.preprocessing import StandardScaler
     import numpy as np
     from mpl_toolkits.mplot3d import Axes3D
     import matplotlib.pyplot as plt
     import matplotlib.tri as mtri
     from sklearn.decomposition import PCA
     import seaborn as sns
     sns.set()
     df=pd.read_csv("https://archive.ics.uci.edu/ml/machine-learning-databases/iris/
      ↪iris.data", names=['sepal length','sepal width','petal length','petal␣
      ↪width','target'])
     print(df)
```

```
     sepal length  sepal width  petal length  petal width          target
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
..            ...          ...           ...          ...             ...
145           6.7          3.0           5.2          2.3  Iris-virginica
146           6.3          2.5           5.0          1.9  Iris-virginica
147           6.5          3.0           5.2          2.0  Iris-virginica
148           6.2          3.4           5.4          2.3  Iris-virginica
149           5.9          3.0           5.1          1.8  Iris-virginica

[150 rows x 5 columns]
```

```python
[3]: features = ['sepal length', 'sepal width', 'petal length', 'petal width']
     # Separating out the features
     x = df.loc[:, features].values
     # Separating out the target
     y = df.loc[:,['target']].values
     # Standardizing the features
     x = StandardScaler().fit_transform(x)
     print(x)
```

```
[[-9.00681170e-01  1.03205722e+00 -1.34127240e+00 -1.31297673e+00]
 [-1.14301691e+00 -1.24957601e-01 -1.34127240e+00 -1.31297673e+00]
 [-1.38535265e+00  3.37848329e-01 -1.39813811e+00 -1.31297673e+00]
 [-1.50652052e+00  1.06445364e-01 -1.28440670e+00 -1.31297673e+00]
 [-1.02184904e+00  1.26346019e+00 -1.34127240e+00 -1.31297673e+00]
 [-5.37177559e-01  1.95766909e+00 -1.17067529e+00 -1.05003079e+00]
 [-1.50652052e+00  8.00654259e-01 -1.34127240e+00 -1.18150376e+00]
 [-1.02184904e+00  8.00654259e-01 -1.28440670e+00 -1.31297673e+00]
 [-1.74885626e+00 -3.56360566e-01 -1.34127240e+00 -1.31297673e+00]
 [-1.14301691e+00  1.06445364e-01 -1.28440670e+00 -1.44444970e+00]
 [-5.37177559e-01  1.49486315e+00 -1.28440670e+00 -1.31297673e+00]
 [-1.26418478e+00  8.00654259e-01 -1.22754100e+00 -1.31297673e+00]
 [-1.26418478e+00 -1.24957601e-01 -1.34127240e+00 -1.44444970e+00]
 [-1.87002413e+00 -1.24957601e-01 -1.51186952e+00 -1.44444970e+00]
 [-5.25060772e-02  2.18907205e+00 -1.45500381e+00 -1.31297673e+00]
 [-1.73673948e-01  3.11468391e+00 -1.28440670e+00 -1.05003079e+00]
 [-5.37177559e-01  1.95766909e+00 -1.39813811e+00 -1.05003079e+00]
 [-9.00681170e-01  1.03205722e+00 -1.34127240e+00 -1.18150376e+00]
 [-1.73673948e-01  1.72626612e+00 -1.17067529e+00 -1.18150376e+00]
 [-9.00681170e-01  1.72626612e+00 -1.28440670e+00 -1.18150376e+00]
 [-5.37177559e-01  8.00654259e-01 -1.17067529e+00 -1.31297673e+00]
 [-9.00681170e-01  1.49486315e+00 -1.28440670e+00 -1.05003079e+00]
 [-1.50652052e+00  1.26346019e+00 -1.56873522e+00 -1.31297673e+00]
 [-9.00681170e-01  5.69251294e-01 -1.17067529e+00 -9.18557817e-01]
 [-1.26418478e+00  8.00654259e-01 -1.05694388e+00 -1.31297673e+00]
 [-1.02184904e+00 -1.24957601e-01 -1.22754100e+00 -1.31297673e+00]
 [-1.02184904e+00  8.00654259e-01 -1.22754100e+00 -1.05003079e+00]
 [-7.79513300e-01  1.03205722e+00 -1.28440670e+00 -1.31297673e+00]
 [-7.79513300e-01  8.00654259e-01 -1.34127240e+00 -1.31297673e+00]
 [-1.38535265e+00  3.37848329e-01 -1.22754100e+00 -1.31297673e+00]
 [-1.26418478e+00  1.06445364e-01 -1.22754100e+00 -1.31297673e+00]
 [-5.37177559e-01  8.00654259e-01 -1.28440670e+00 -1.05003079e+00]
 [-7.79513300e-01  2.42047502e+00 -1.28440670e+00 -1.44444970e+00]
 [-4.16009689e-01  2.65187798e+00 -1.34127240e+00 -1.31297673e+00]
 [-1.14301691e+00  1.06445364e-01 -1.28440670e+00 -1.44444970e+00]
 [-1.02184904e+00  3.37848329e-01 -1.45500381e+00 -1.31297673e+00]
 [-4.16009689e-01  1.03205722e+00 -1.39813811e+00 -1.31297673e+00]
 [-1.14301691e+00  1.06445364e-01 -1.28440670e+00 -1.44444970e+00]
 [-1.74885626e+00 -1.24957601e-01 -1.39813811e+00 -1.31297673e+00]
 [-9.00681170e-01  8.00654259e-01 -1.28440670e+00 -1.31297673e+00]
 [-1.02184904e+00  1.03205722e+00 -1.39813811e+00 -1.18150376e+00]
 [-1.62768839e+00 -1.74477836e+00 -1.39813811e+00 -1.18150376e+00]
 [-1.74885626e+00  3.37848329e-01 -1.39813811e+00 -1.31297673e+00]
 [-1.02184904e+00  1.03205722e+00 -1.22754100e+00 -7.87084847e-01]
 [-9.00681170e-01  1.72626612e+00 -1.05694388e+00 -1.05003079e+00]
 [-1.26418478e+00 -1.24957601e-01 -1.34127240e+00 -1.18150376e+00]
 [-9.00681170e-01  1.72626612e+00 -1.22754100e+00 -1.31297673e+00]
 [-1.50652052e+00  3.37848329e-01 -1.34127240e+00 -1.31297673e+00]
```

```
[-6.58345429e-01  1.49486315e+00 -1.28440670e+00 -1.31297673e+00]
[-1.02184904e+00  5.69251294e-01 -1.34127240e+00 -1.31297673e+00]
[ 1.40150837e+00  3.37848329e-01  5.35295827e-01  2.64698913e-01]
[ 6.74501145e-01  3.37848329e-01  4.21564419e-01  3.96171883e-01]
[ 1.28034050e+00  1.06445364e-01  6.49027235e-01  3.96171883e-01]
[-4.16009689e-01 -1.74477836e+00  1.37235899e-01  1.33225943e-01]
[ 7.95669016e-01 -5.87763531e-01  4.78430123e-01  3.96171883e-01]
[-1.73673948e-01 -5.87763531e-01  4.21564419e-01  1.33225943e-01]
[ 5.53333275e-01  5.69251294e-01  5.35295827e-01  5.27644853e-01]
[-1.14301691e+00 -1.51337539e+00 -2.60824029e-01 -2.61192967e-01]
[ 9.16836886e-01 -3.56360566e-01  4.78430123e-01  1.33225943e-01]
[-7.79513300e-01 -8.19166497e-01  8.03701950e-02  2.64698913e-01]
[-1.02184904e+00 -2.43898725e+00 -1.47092621e-01 -2.61192967e-01]
[ 6.86617933e-02 -1.24957601e-01  2.50967307e-01  3.96171883e-01]
[ 1.89829664e-01 -1.97618132e+00  1.37235899e-01 -2.61192967e-01]
[ 3.10997534e-01 -3.56360566e-01  5.35295827e-01  2.64698913e-01]
[-2.94841818e-01 -3.56360566e-01 -9.02269170e-02  1.33225943e-01]
[ 1.03800476e+00  1.06445364e-01  3.64698715e-01  2.64698913e-01]
[-2.94841818e-01 -1.24957601e-01  4.21564419e-01  3.96171883e-01]
[-5.25060772e-02 -8.19166497e-01  1.94101603e-01 -2.61192967e-01]
[ 4.32165405e-01 -1.97618132e+00  4.21564419e-01  3.96171883e-01]
[-2.94841818e-01 -1.28197243e+00  8.03701950e-02 -1.29719997e-01]
[ 6.86617933e-02  3.37848329e-01  5.92161531e-01  7.90590793e-01]
[ 3.10997534e-01 -5.87763531e-01  1.37235899e-01  1.33225943e-01]
[ 5.53333275e-01 -1.28197243e+00  6.49027235e-01  3.96171883e-01]
[ 3.10997534e-01 -5.87763531e-01  5.35295827e-01  1.75297293e-03]
[ 6.74501145e-01 -3.56360566e-01  3.07833011e-01  1.33225943e-01]
[ 9.16836886e-01 -1.24957601e-01  3.64698715e-01  2.64698913e-01]
[ 1.15917263e+00 -5.87763531e-01  5.92161531e-01  2.64698913e-01]
[ 1.03800476e+00 -1.24957601e-01  7.05892939e-01  6.59117823e-01]
[ 1.89829664e-01 -3.56360566e-01  4.21564419e-01  3.96171883e-01]
[-1.73673948e-01 -1.05056946e+00 -1.47092621e-01 -2.61192967e-01]
[-4.16009689e-01 -1.51337539e+00  2.35044910e-02 -1.29719997e-01]
[-4.16009689e-01 -1.51337539e+00 -3.33612130e-02 -2.61192967e-01]
[-5.25060772e-02 -8.19166497e-01  8.03701950e-02  1.75297293e-03]
[ 1.89829664e-01 -8.19166497e-01  7.62758643e-01  5.27644853e-01]
[-5.37177559e-01 -1.24957601e-01  4.21564419e-01  3.96171883e-01]
[ 1.89829664e-01  8.00654259e-01  4.21564419e-01  5.27644853e-01]
[ 1.03800476e+00  1.06445364e-01  5.35295827e-01  3.96171883e-01]
[ 5.53333275e-01 -1.74477836e+00  3.64698715e-01  1.33225943e-01]
[-2.94841818e-01 -1.24957601e-01  1.94101603e-01  1.33225943e-01]
[-4.16009689e-01 -1.28197243e+00  1.37235899e-01  1.33225943e-01]
[-4.16009689e-01 -1.05056946e+00  3.64698715e-01  1.75297293e-03]
[ 3.10997534e-01 -1.24957601e-01  4.78430123e-01  2.64698913e-01]
[-5.25060772e-02 -1.05056946e+00  1.37235899e-01  1.75297293e-03]
[-1.02184904e+00 -1.74477836e+00 -2.60824029e-01 -2.61192967e-01]
[-2.94841818e-01 -8.19166497e-01  2.50967307e-01  1.33225943e-01]
[-1.73673948e-01 -1.24957601e-01  2.50967307e-01  1.75297293e-03]
```

```
[-1.73673948e-01 -3.56360566e-01  2.50967307e-01  1.33225943e-01]
[ 4.32165405e-01 -3.56360566e-01  3.07833011e-01  1.33225943e-01]
[-9.00681170e-01 -1.28197243e+00 -4.31421141e-01 -1.29719997e-01]
[-1.73673948e-01 -5.87763531e-01  1.94101603e-01  1.33225943e-01]
[ 5.53333275e-01  5.69251294e-01  1.27454998e+00  1.71090158e+00]
[-5.25060772e-02 -8.19166497e-01  7.62758643e-01  9.22063763e-01]
[ 1.52267624e+00 -1.24957601e-01  1.21768427e+00  1.18500970e+00]
[ 5.53333275e-01 -3.56360566e-01  1.04708716e+00  7.90590793e-01]
[ 7.95669016e-01 -1.24957601e-01  1.16081857e+00  1.31648267e+00]
[ 2.12851559e+00 -1.24957601e-01  1.61574420e+00  1.18500970e+00]
[-1.14301691e+00 -1.28197243e+00  4.21564419e-01  6.59117823e-01]
[ 1.76501198e+00 -3.56360566e-01  1.44514709e+00  7.90590793e-01]
[ 1.03800476e+00 -1.28197243e+00  1.16081857e+00  7.90590793e-01]
[ 1.64384411e+00  1.26346019e+00  1.33141568e+00  1.71090158e+00]
[ 7.95669016e-01  3.37848329e-01  7.62758643e-01  1.05353673e+00]
[ 6.74501145e-01 -8.19166497e-01  8.76490051e-01  9.22063763e-01]
[ 1.15917263e+00 -1.24957601e-01  9.90221459e-01  1.18500970e+00]
[-1.73673948e-01 -1.28197243e+00  7.05892939e-01  1.05353673e+00]
[-5.25060772e-02 -5.87763531e-01  7.62758643e-01  1.57942861e+00]
[ 6.74501145e-01  3.37848329e-01  8.76490051e-01  1.44795564e+00]
[ 7.95669016e-01 -1.24957601e-01  9.90221459e-01  7.90590793e-01]
[ 2.24968346e+00  1.72626612e+00  1.67260991e+00  1.31648267e+00]
[ 2.24968346e+00 -1.05056946e+00  1.78634131e+00  1.44795564e+00]
[ 1.89829664e-01 -1.97618132e+00  7.05892939e-01  3.96171883e-01]
[ 1.28034050e+00  3.37848329e-01  1.10395287e+00  1.44795564e+00]
[-2.94841818e-01 -5.87763531e-01  6.49027235e-01  1.05353673e+00]
[ 2.24968346e+00 -5.87763531e-01  1.67260991e+00  1.05353673e+00]
[ 5.53333275e-01 -8.19166497e-01  6.49027235e-01  7.90590793e-01]
[ 1.03800476e+00  5.69251294e-01  1.10395287e+00  1.18500970e+00]
[ 1.64384411e+00  3.37848329e-01  1.27454998e+00  7.90590793e-01]
[ 4.32165405e-01 -5.87763531e-01  5.92161531e-01  7.90590793e-01]
[ 3.10997534e-01 -1.24957601e-01  6.49027235e-01  7.90590793e-01]
[ 6.74501145e-01 -5.87763531e-01  1.04708716e+00  1.18500970e+00]
[ 1.64384411e+00 -1.24957601e-01  1.16081857e+00  5.27644853e-01]
[ 1.88617985e+00 -5.87763531e-01  1.33141568e+00  9.22063763e-01]
[ 2.49201920e+00  1.72626612e+00  1.50201279e+00  1.05353673e+00]
[ 6.74501145e-01 -5.87763531e-01  1.04708716e+00  1.31648267e+00]
[ 5.53333275e-01 -5.87763531e-01  7.62758643e-01  3.96171883e-01]
[ 3.10997534e-01 -1.05056946e+00  1.04708716e+00  2.64698913e-01]
[ 2.24968346e+00 -1.24957601e-01  1.33141568e+00  1.44795564e+00]
[ 5.53333275e-01  8.00654259e-01  1.04708716e+00  1.57942861e+00]
[ 6.74501145e-01  1.06445364e-01  9.90221459e-01  7.90590793e-01]
[ 1.89829664e-01 -1.24957601e-01  5.92161531e-01  7.90590793e-01]
[ 1.28034050e+00  1.06445364e-01  9.33355755e-01  1.18500970e+00]
[ 1.03800476e+00  1.06445364e-01  1.04708716e+00  1.57942861e+00]
[ 1.28034050e+00  1.06445364e-01  7.62758643e-01  1.44795564e+00]
[-5.25060772e-02 -8.19166497e-01  7.62758643e-01  9.22063763e-01]
[ 1.15917263e+00  3.37848329e-01  1.21768427e+00  1.44795564e+00]
```

```
[ 1.03800476e+00  5.69251294e-01  1.10395287e+00  1.71090158e+00]
[ 1.03800476e+00 -1.24957601e-01  8.19624347e-01  1.44795564e+00]
[ 5.53333275e-01 -1.28197243e+00  7.05892939e-01  9.22063763e-01]
[ 7.95669016e-01 -1.24957601e-01  8.19624347e-01  1.05353673e+00]
[ 4.32165405e-01  8.00654259e-01  9.33355755e-01  1.44795564e+00]
[ 6.86617933e-02 -1.24957601e-01  7.62758643e-01  7.90590793e-01]]
```

[4]:
```python
pca = PCA(n_components=4)
principalComponents = pca.fit_transform(x)
principalDf = pd.DataFrame(data = principalComponents
            , columns = ['component1', 'component2','com3','comp4'])
```

[5]:
```python
df.describe()
```

[5]:

	sepal length	sepal width	petal length	petal width
count	150.000000	150.000000	150.000000	150.000000
mean	5.843333	3.054000	3.758667	1.198667
std	0.828066	0.433594	1.764420	0.763161
min	4.300000	2.000000	1.000000	0.100000
25%	5.100000	2.800000	1.600000	0.300000
50%	5.800000	3.000000	4.350000	1.300000
75%	6.400000	3.300000	5.100000	1.800000
max	7.900000	4.400000	6.900000	2.500000

[6]:
```python
plt.figure(figsize = (10, 7))
a = df["sepal length"]

plt.hist(a, bins = 20, color = "green")
plt.title("Sepal Length in cm")
plt.xlabel("Sepal_Length_cm")
plt.ylabel("Count")
```
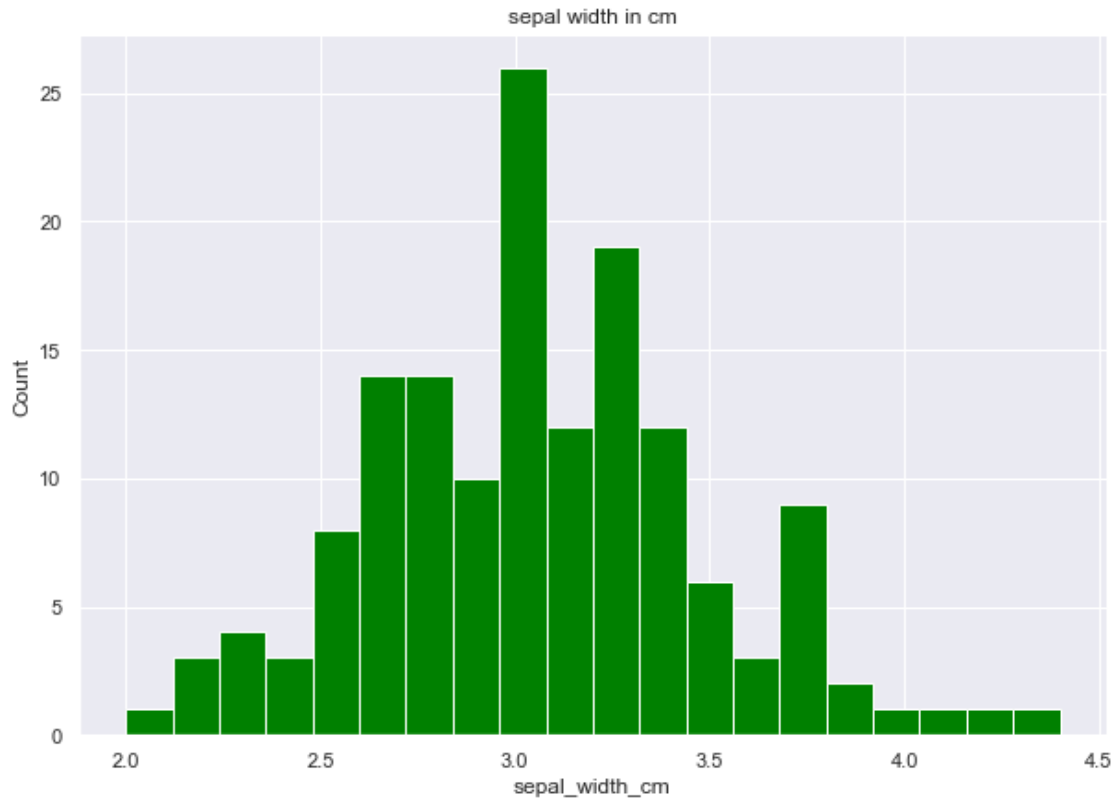
[6]: Text(0, 0.5, 'Count')
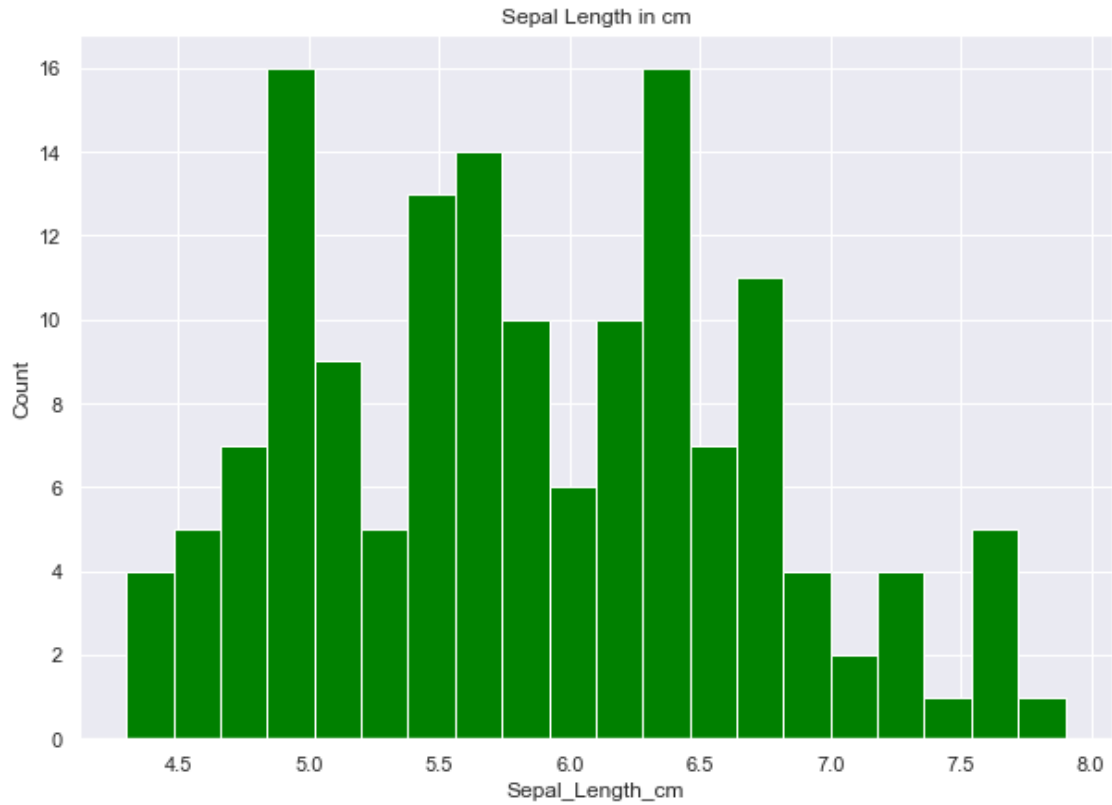
Sepal Length in cm

```
[7]:  plt.figure(figsize = (10, 7))
      b = df["sepal width"]

      plt.hist(b, bins = 20, color = "green")
      plt.title("sepal width in cm")
      plt.xlabel("sepal_width_cm")
      plt.ylabel("Count")
```
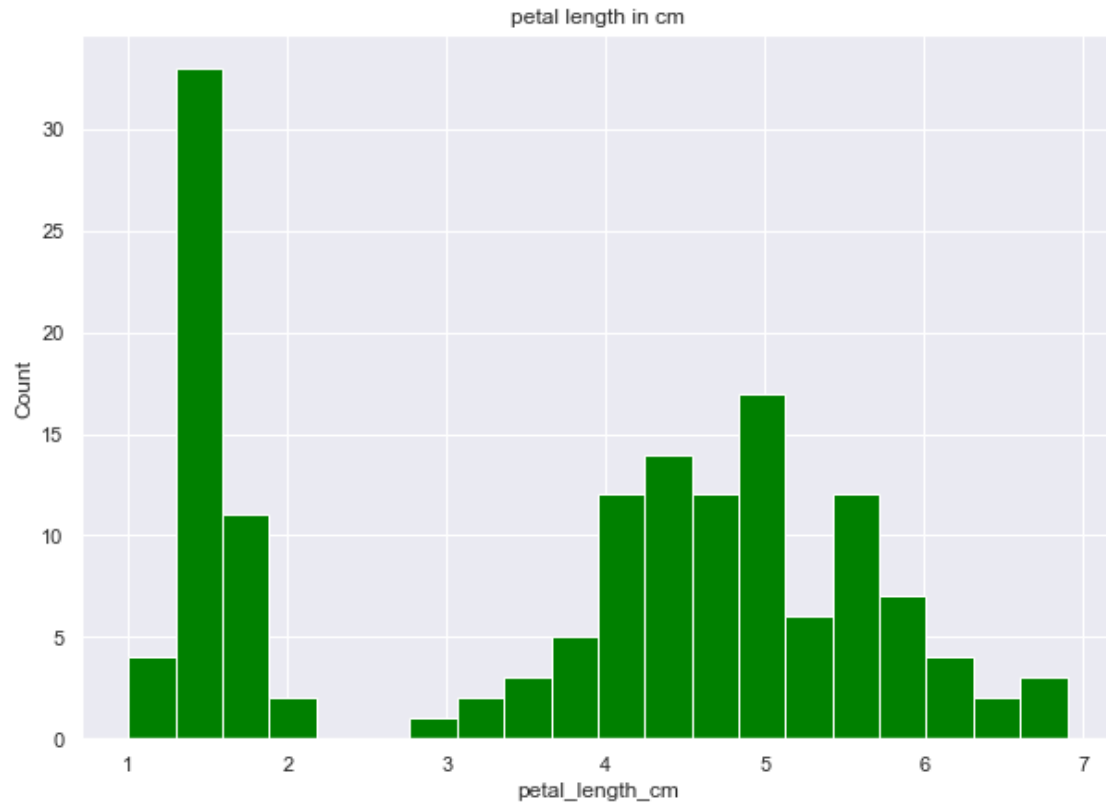
[7]: Text(0, 0.5, 'Count')

sepal width in cm

```
[8]: plt.figure(figsize = (10, 7))
     c = df["petal length"]

     plt.hist(c, bins = 20, color = "green")
     plt.title("petal length in cm")
     plt.xlabel("petal_length_cm")
     plt.ylabel("Count")
```
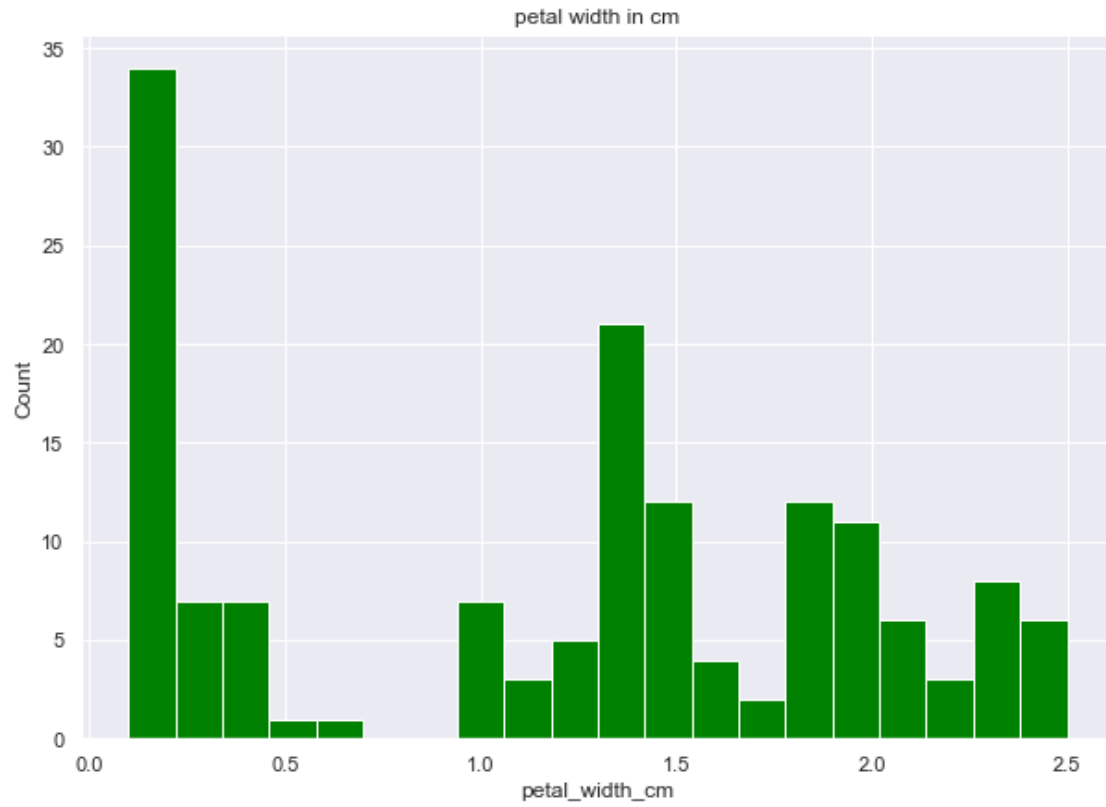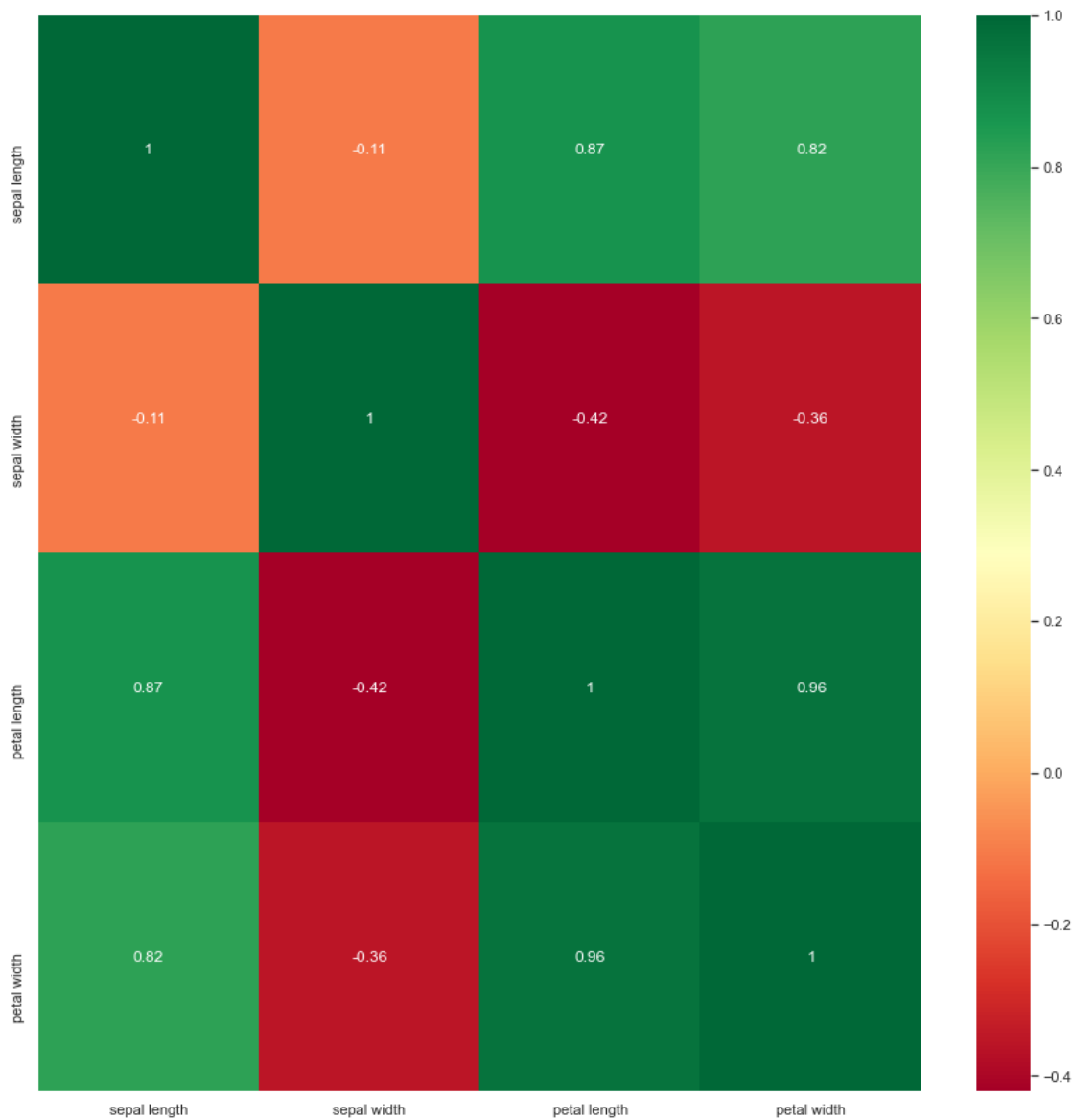
[8]: Text(0, 0.5, 'Count')

petal length in cm

[14]:
```
plt.figure(figsize = (10, 7))
d = df["petal width"]

plt.hist(d, bins = 20, color = "green")
plt.title("petal width in cm")
plt.xlabel("petal_width_cm")
plt.ylabel("Count")
```

[14]: Text(0, 0.5, 'Count')

petal width in cm

```
[15]: plt.figure(figsize=(15,15))
      p=sns.heatmap(df.corr(), annot=True,cmap='RdYlGn')
```

[16]: `print(principalDf)`

```
     component1  component2        com3       comp4
0     -2.264542    0.505704   -0.121943   -0.023073
1     -2.086426   -0.655405   -0.227251   -0.103208
2     -2.367950   -0.318477    0.051480   -0.027825
3     -2.304197   -0.575368    0.098860    0.066311
4     -2.388777    0.674767    0.021428    0.037397
..          ...         ...         ...         ...
145    1.870522    0.382822    0.254532   -0.388890
146    1.558492   -0.905314   -0.025382   -0.221322
147    1.520845    0.266795    0.179277   -0.118903
```

```
148      1.376391      1.016362   0.931405  -0.024146
149      0.959299     -0.022284   0.528794   0.163676
```

```
[150 rows x 4 columns]
```

[17]: `finalDf = pd.concat([principalDf, df[['target']]], axis = 1)`

[18]: `print(finalDf)`

```
     component1 component2   com3   comp4   target
0   -2.264542   0.505704 -0.121943 -0.023073   Iris-setosa
1   -2.086426   -0.655405 -0.227251 -0.103208   Iris-setosa
2   -2.367950   -0.318477   0.051480 -0.027825   Iris-setosa
3   -2.304197   -0.575368   0.098860   0.066311   Iris-setosa
4   -2.388777   0.674767   0.021428   0.037397   Iris-setosa
..   ..
145   1.870522   0.382822   0.254532 -0.388890   Iris-virginica
146   1.558492   -0.905314 -0.025382 -0.221322   Iris-virginica
147   1.520845   0.266795   0.179277 -0.118903   Iris-virginica
148   1.376391   1.016362   0.931405 -0.024146   Iris-virginica
149   0.959299   -0.022284   0.528794   0.163676   Iris-virginica
```
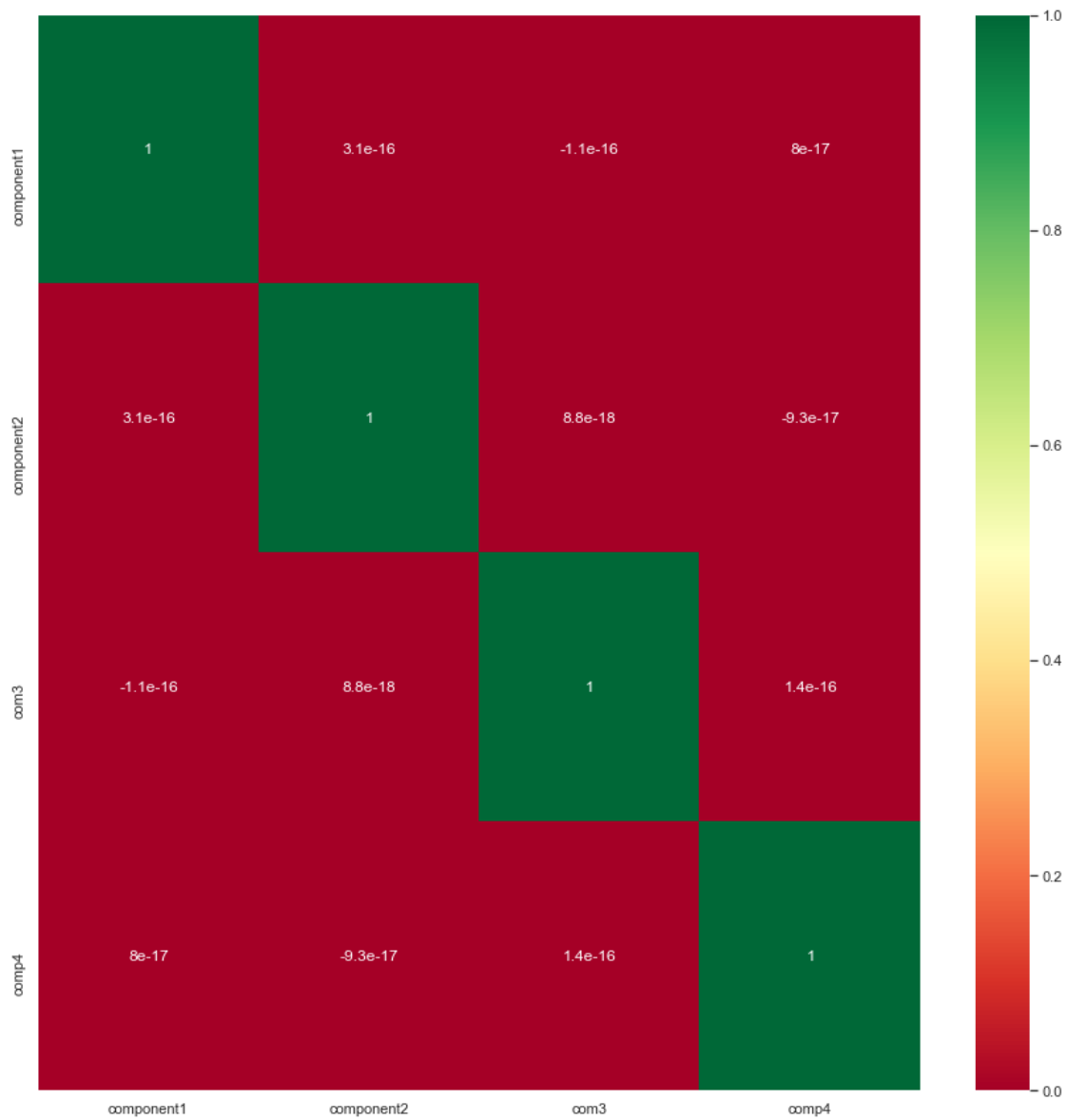
```
[150 rows x 5 columns]
```

[19]:
```python
fig = plt.figure(figsize = (8,8))
ax = fig.add_subplot(1,1,1)
ax.set_xlabel('component1', fontsize = 15)
ax.set_ylabel('component2', fontsize = 15)
ax.set_title('2 component PCA', fontsize = 20)
targets = ['Iris-setosa', 'Iris-versicolor', 'Iris-virginica']
colors = ['r', 'g', 'b']
for target, color in zip(targets,colors):
    indicesToKeep = finalDf['target'] == target
    ax.scatter(finalDf.loc[indicesToKeep, 'component1']
               , finalDf.loc[indicesToKeep, 'component2']
               , c = color
               , s = 50)
ax.legend(targets)
ax.grid()
```

2 component PCA
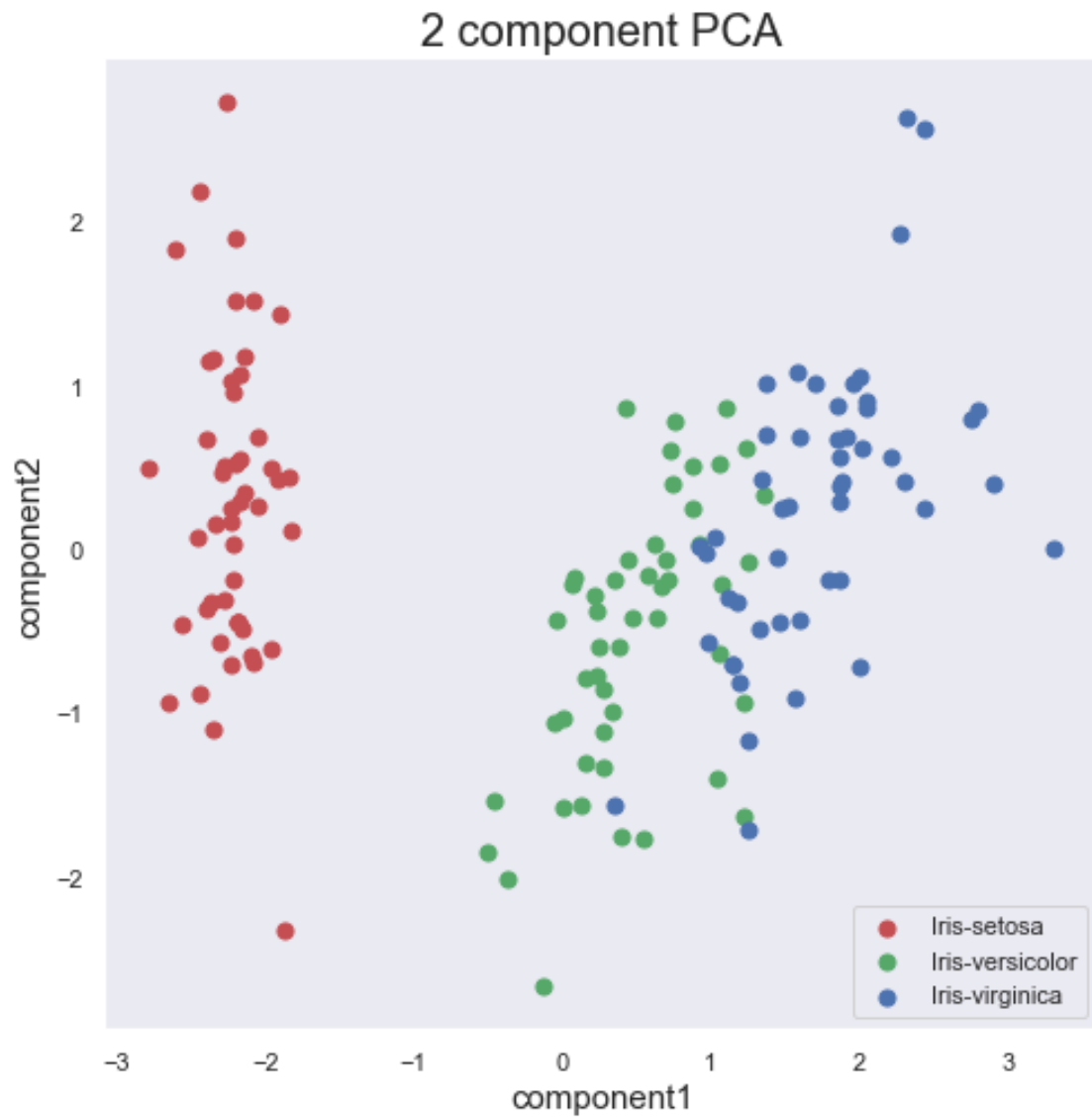

```
[20]: pca.explained_variance_ratio_
```

```
[20]: array([0.72770452, 0.23030523, 0.03683832, 0.00515193])
```

```
[21]: plt.figure(figsize=(15,15))
      p=sns.heatmap(finalDf.corr(), annot=True,cmap='RdYlGn')
```