# Name: Aniruddh Kulkarni

# Roll No: I081

# Subject: Machine Learning

# Practical: Practical 2

# Date: 17-12-21

```
In [3]:   # Lab assignment 2
          # Data Visualisation
          # Go to Kaggle and download the wheat seeds dataset
```

```
In [4]:   import numpy as np
          import pandas as pd
          import matplotlib.pyplot as plt
          import seaborn as sns
```

```
In [5]:   df = pd.read_csv('seeds.csv')
```

```
In [6]:   df.head()
```

Out[6]:

| | Area | Perimeter | Compactness | Kernel.Length | Kernel.Width | Asymmetry.Coeff | Kernel.G |
|---|---|---|---|---|---|---|---|
| 0 | 15.26 | 14.84 | 0.8710 | 5.763 | 3.312 | 2.221 | |
| 1 | 14.88 | 14.57 | 0.8811 | 5.554 | 3.333 | 1.018 | |
| 2 | 14.29 | 14.09 | 0.9050 | 5.291 | 3.337 | 2.699 | |
| 3 | 13.84 | 13.94 | 0.8955 | 5.324 | 3.379 | 2.259 | |
| 4 | 16.14 | 14.99 | 0.9034 | 5.658 | 3.562 | 1.355 | |

```
In [7]:   df["Type"].value_counts()
```

```
Out[7]:   2    68
          1    66
          3    65
          Name: Type, dtype: int64
```

```
In [8]:   df.describe()
```

Out[8]:

| | Area | Perimeter | Compactness | Kernel.Length | Kernel.Width | Asymmetry.Coe |
|---|---|---|---|---|---|---|
| count | 199.000000 | 199.000000 | 199.000000 | 199.000000 | 199.000000 | 199.00000 |
| mean | 14.918744 | 14.595829 | 0.870811 | 5.643151 | 3.265533 | 3.69921 |
| std | 2.919976 | 1.310445 | 0.023320 | 0.443593 | 0.378322 | 1.47110 |
| min | 10.590000 | 12.410000 | 0.808100 | 4.899000 | 2.630000 | 0.76510 |
| 25% | 12.330000 | 13.470000 | 0.857100 | 5.267000 | 2.954500 | 2.57000 |
| 50% | 14.430000 | 14.370000 | 0.873400 | 5.541000 | 3.245000 | 3.63100 |
| 75% | 17.455000 | 15.805000 | 0.886800 | 6.002000 | 3.564500 | 4.79900 |
| max | 21.180000 | 17.250000 | 0.918300 | 6.675000 | 4.033000 | 8.31500 |

In [9]:
```python
# matplotlib
# plot
df.plot(kind = 'scatter', x = "Kernel.Length", y = "Kernel.Width")
```
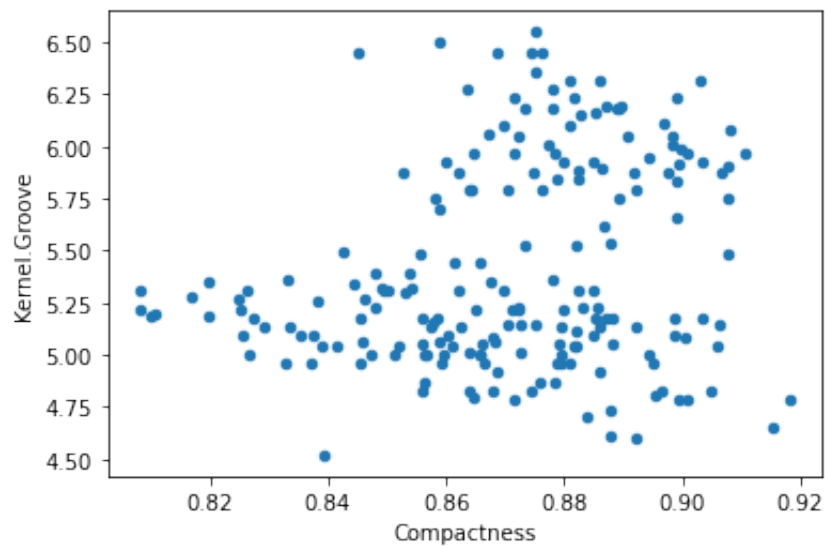
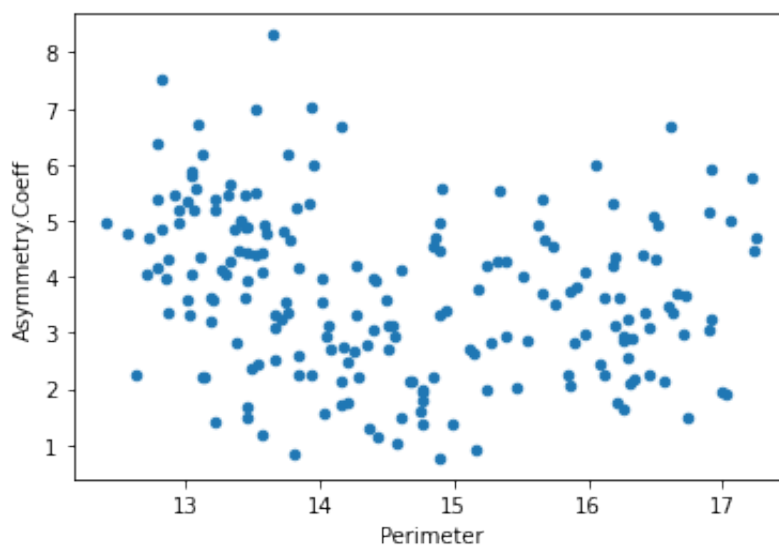Out[9]: `<AxesSubplot:xlabel='Kernel.Length', ylabel='Kernel.Width'>`
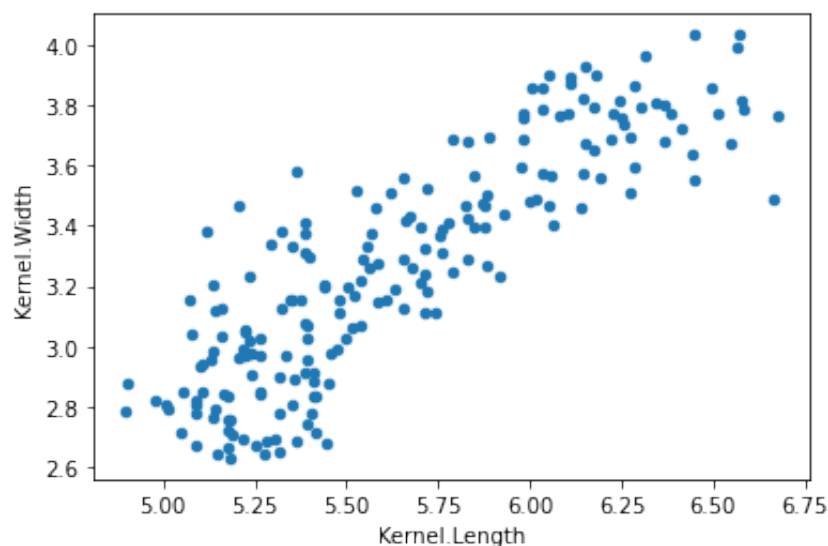


In [10]:
```python
df.plot(kind = 'scatter', x = 'Compactness', y = 'Kernel.Groove')
```

Out[10]: `<AxesSubplot:xlabel='Compactness', ylabel='Kernel.Groove'>`



In [11]: 
```python
df.plot(kind = 'scatter', x = 'Perimeter', y = 'Asymmetry.Coeff')
```

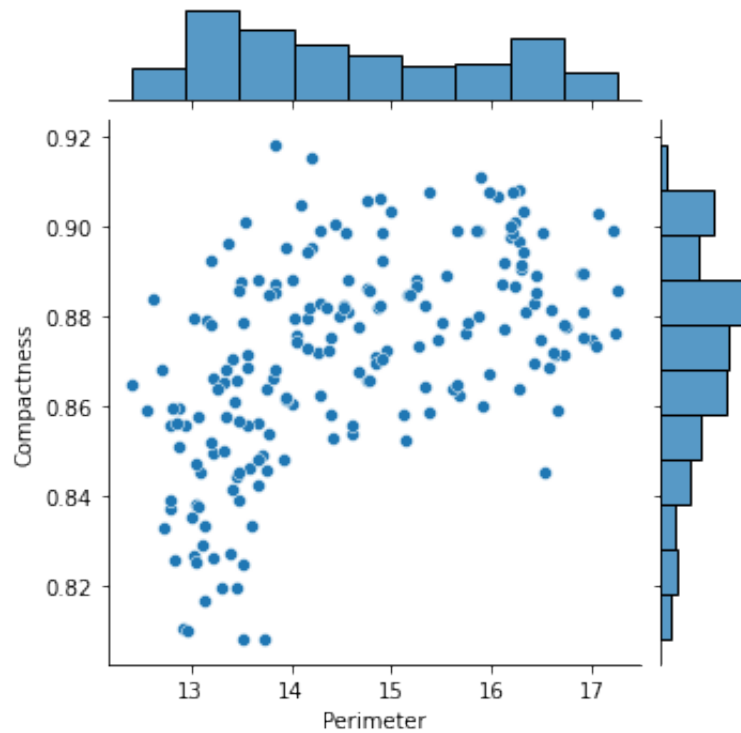Out[11]: `<AxesSubplot:xlabel='Perimeter', ylabel='Asymmetry.Coeff'>`



In [12]: 
```python
# use of seaborn library
sns.jointplot(x = 'Perimeter', y = 'Compactness', data = df, size = 5)
```

```
/opt/anaconda3/lib/python3.8/site-packages/seaborn/axisgrid.py:2015: UserWa
rning: The `size` parameter has been renamed to `height`; please update you
r code.
  warnings.warn(msg, UserWarning)
```

Out[12]: `<seaborn.axisgrid.JointGrid at 0x7fada1acef40>`
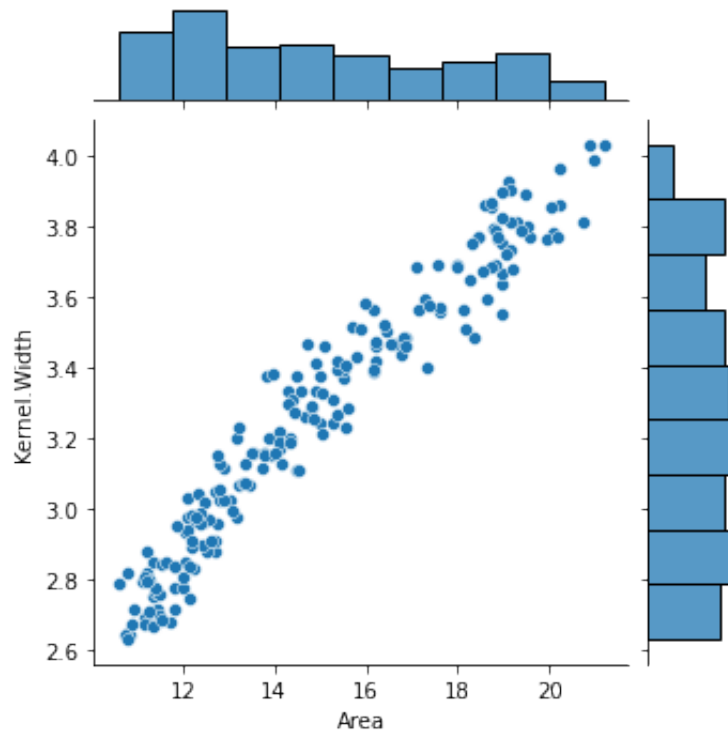


In [13]: 
```
sns.jointplot(x = 'Area', y = 'Kernel.Width', data = df, size = 5)
```

```
/opt/anaconda3/lib/python3.8/site-packages/seaborn/axisgrid.py:2015: UserWa
rning: The `size` parameter has been renamed to `height`; please update you
r code.
  warnings.warn(msg, UserWarning)
```

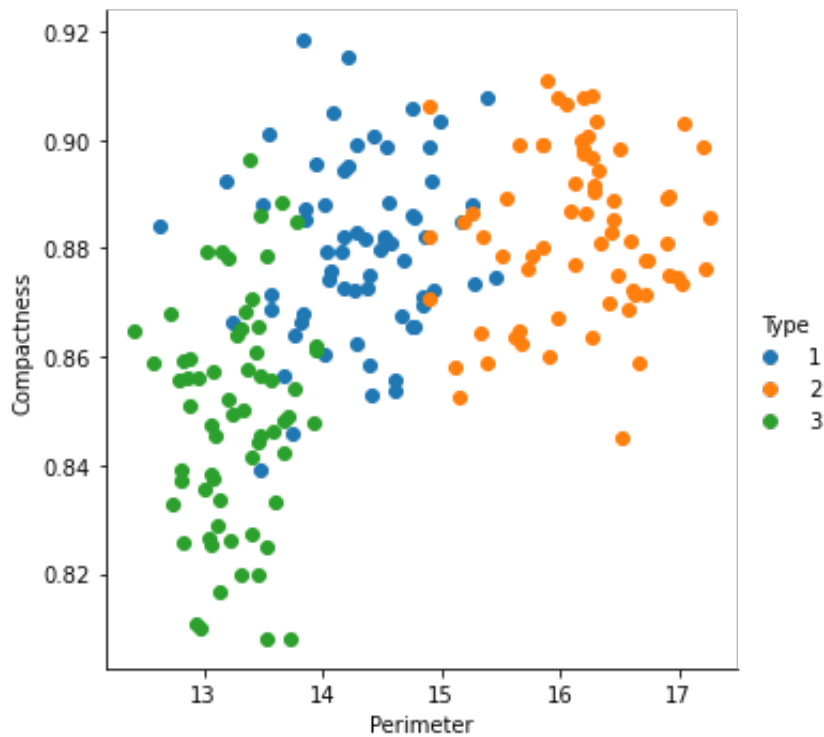Out[13]: `<seaborn.axisgrid.JointGrid at 0x7fada1b6a040>`



In [14]:
```python
# plot three types of seeds with three different colours
sns.FacetGrid(df, hue = 'Type', size = 5).map(plt.scatter, "Perimeter", "C
```

```
/opt/anaconda3/lib/python3.8/site-packages/seaborn/axisgrid.py:316: UserWar
ning: The `size` parameter has been renamed to `height`; please update your
code.
  warnings.warn(msg, UserWarning)
```

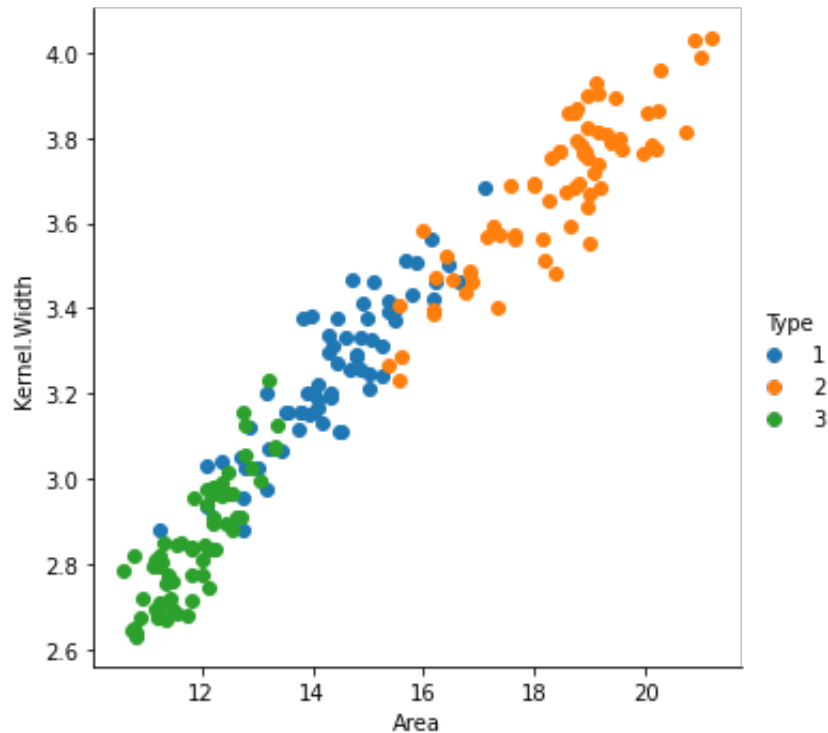Out[14]: `<seaborn.axisgrid.FacetGrid at 0x7fadb22456a0>`



In [15]: 
```python
sns.FacetGrid(df, hue = 'Type', size = 5).map(plt.scatter, "Area", "Kernel
```

```
/opt/anaconda3/lib/python3.8/site-packages/seaborn/axisgrid.py:316: UserWar
ning: The `size` parameter has been renamed to `height`; please update your
code.
  warnings.warn(msg, UserWarning)
```

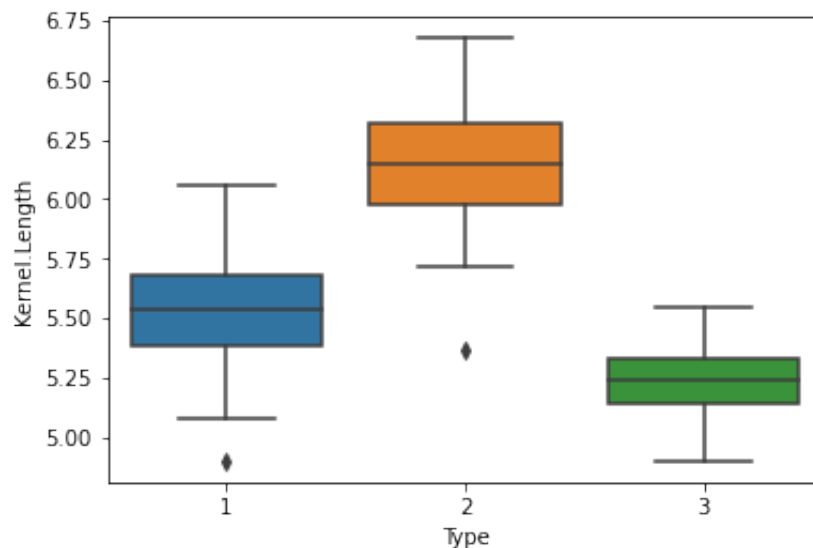Out[15]: <seaborn.axisgrid.FacetGrid at 0x7fadb22ee100>



In [16]:
```python
# boxplot is used to identify the outliers
sns.boxplot(x = "Type", y = "Compactness", data = df)
```

Out[16]: <AxesSubplot:xlabel='Type', ylabel='Compactness'>
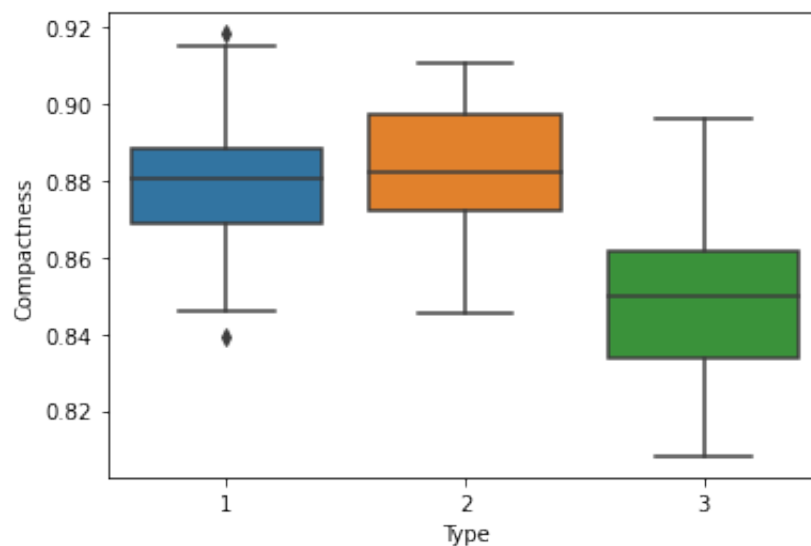


In [17]:
```python
sns.boxplot(x = 'Type', y = 'Kernel.Length', data = df)
```
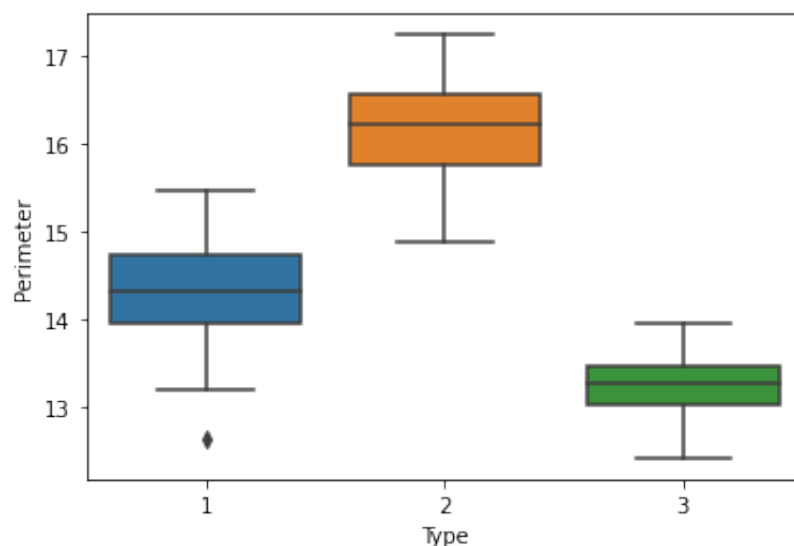
Out[17]:   `<AxesSubplot:xlabel='Type', ylabel='Kernel.Length'>`



In [19]:
```python
# here boxplot gives the range of points which we are looking for and the
sns.boxplot(x = 'Type', y = 'Perimeter', data = df)
```
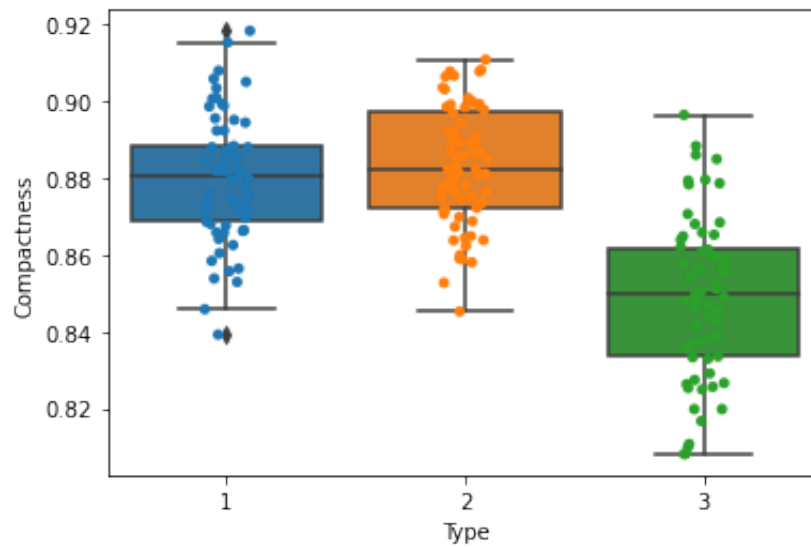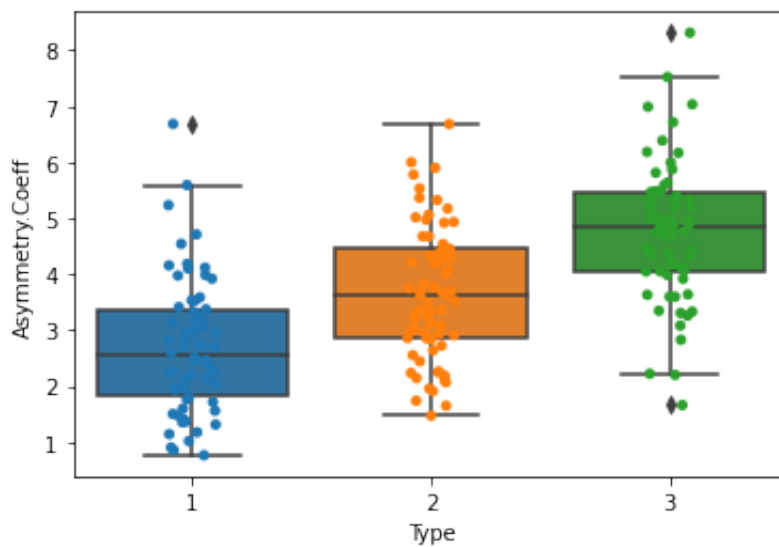
Out[19]:   `<AxesSubplot:xlabel='Type', ylabel='Perimeter'>`



In [20]:
```python
# with the help of stripplot we are superimposing our coloured dots on the
ax = sns.boxplot(x = 'Type', y = 'Compactness', data = df)
ax = sns.stripplot(x = 'Type', y = 'Compactness', data = df, jitter = True
```
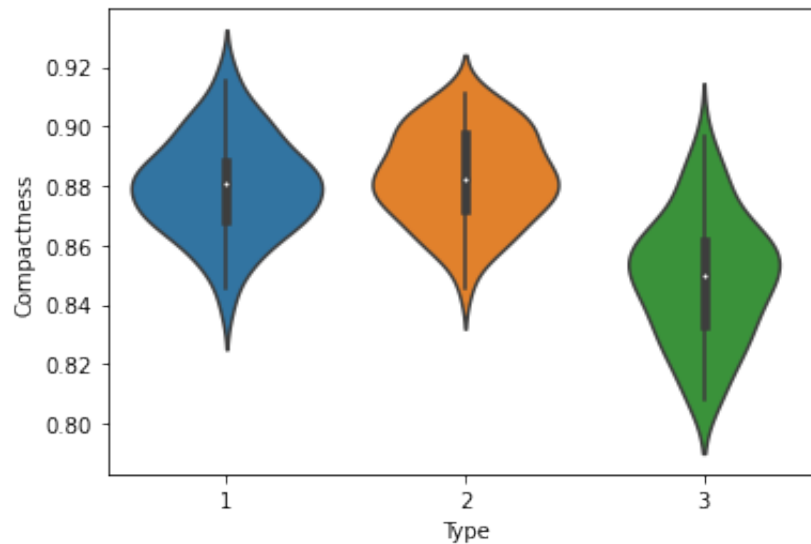
```
In [21]:   bx = sns.boxplot(x = 'Type', y = 'Asymmetry.Coeff', data = df)
           bx = sns.stripplot(x = 'Type', y = 'Asymmetry.Coeff', data = df, jitter = 
```



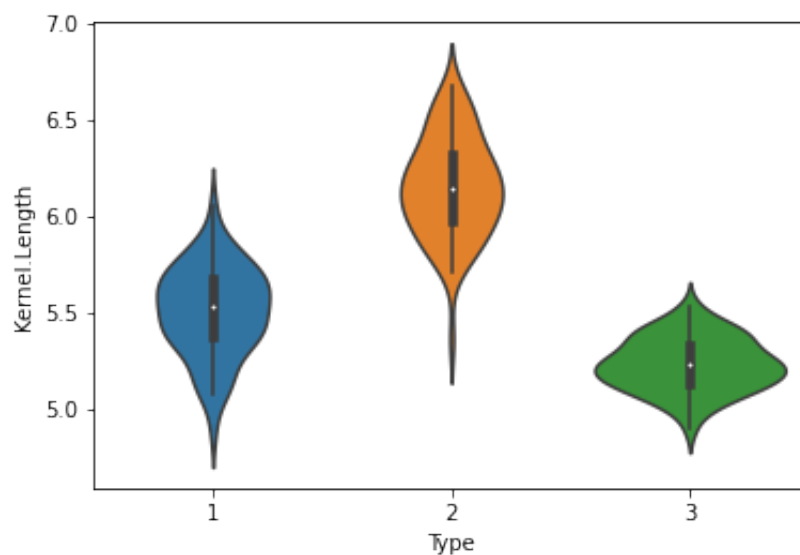```
In [22]:   # violinplot is one more type of visualisation to help us find the outlier
           sns.violinplot(x = 'Type', y = 'Compactness', data = df, size = 5)
```

Out[22]:  `<AxesSubplot:xlabel='Type', ylabel='Compactness'>`



In [25]:
```python
sns.violinplot(x = 'Type', y = 'Kernel.Length', data = df, size = 10)
```

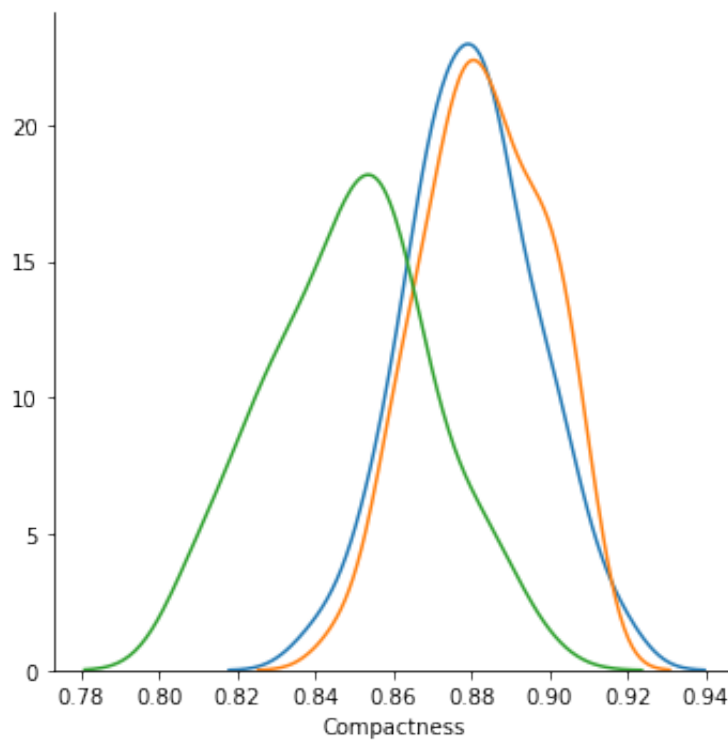Out[25]:  `<AxesSubplot:xlabel='Type', ylabel='Kernel.Length'>`



In [26]:
```python
# kdeplot
sns.FacetGrid(df, hue = 'Type', size = 5).map(sns.kdeplot, "Compactness").
```

```
/opt/anaconda3/lib/python3.8/site-packages/seaborn/axisgrid.py:316: UserWar
ning: The `size` parameter has been renamed to `height`; please update your
code.
  warnings.warn(msg, UserWarning)
```

Out[26]: `<bound method Grid.add_legend of <seaborn.axisgrid.FacetGrid object at 0x7f adb27f30d0>>`



In [27]:
```python
sns.FacetGrid(df, hue = 'Type', size = 5).map(sns.kdeplot, "Kernel.Length"
```
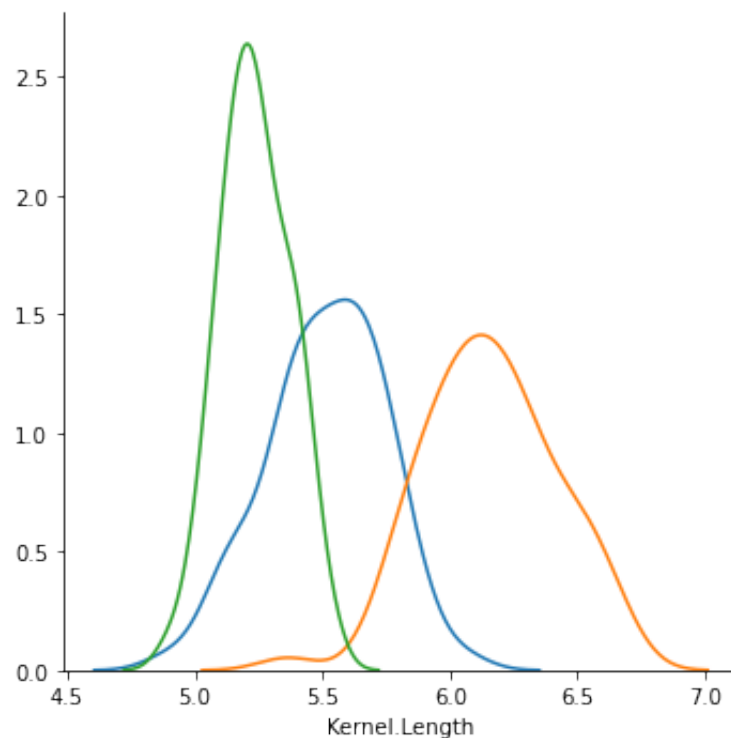
```
/opt/anaconda3/lib/python3.8/site-packages/seaborn/axisgrid.py:316: UserWar
ning: The `size` parameter has been renamed to `height`; please update your
code.
  warnings.warn(msg, UserWarning)
```

Out[27]: `<bound method Grid.add_legend of <seaborn.axisgrid.FacetGrid object at 0x7f`
`ada1cd16a0>>`



In [28]:
```python
# pairplot to be used when we'll implement multivariate regression
sns.pairplot(df, hue = "Type", size = 3)
```
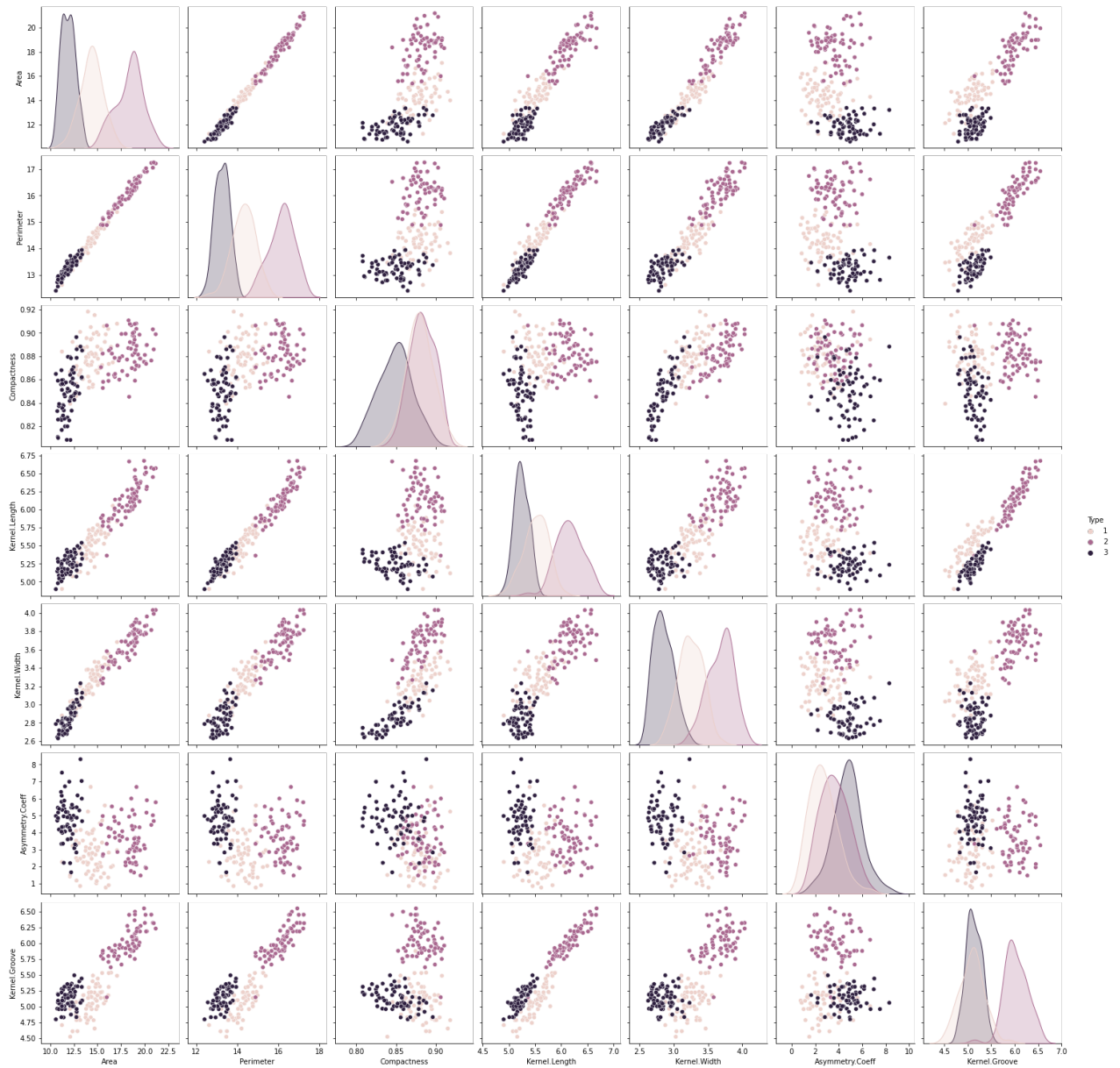
```
/opt/anaconda3/lib/python3.8/site-packages/seaborn/axisgrid.py:1912: UserWa
rning: The `size` parameter has been renamed to `height`; please update you
r code.
  warnings.warn(msg, UserWarning)
```
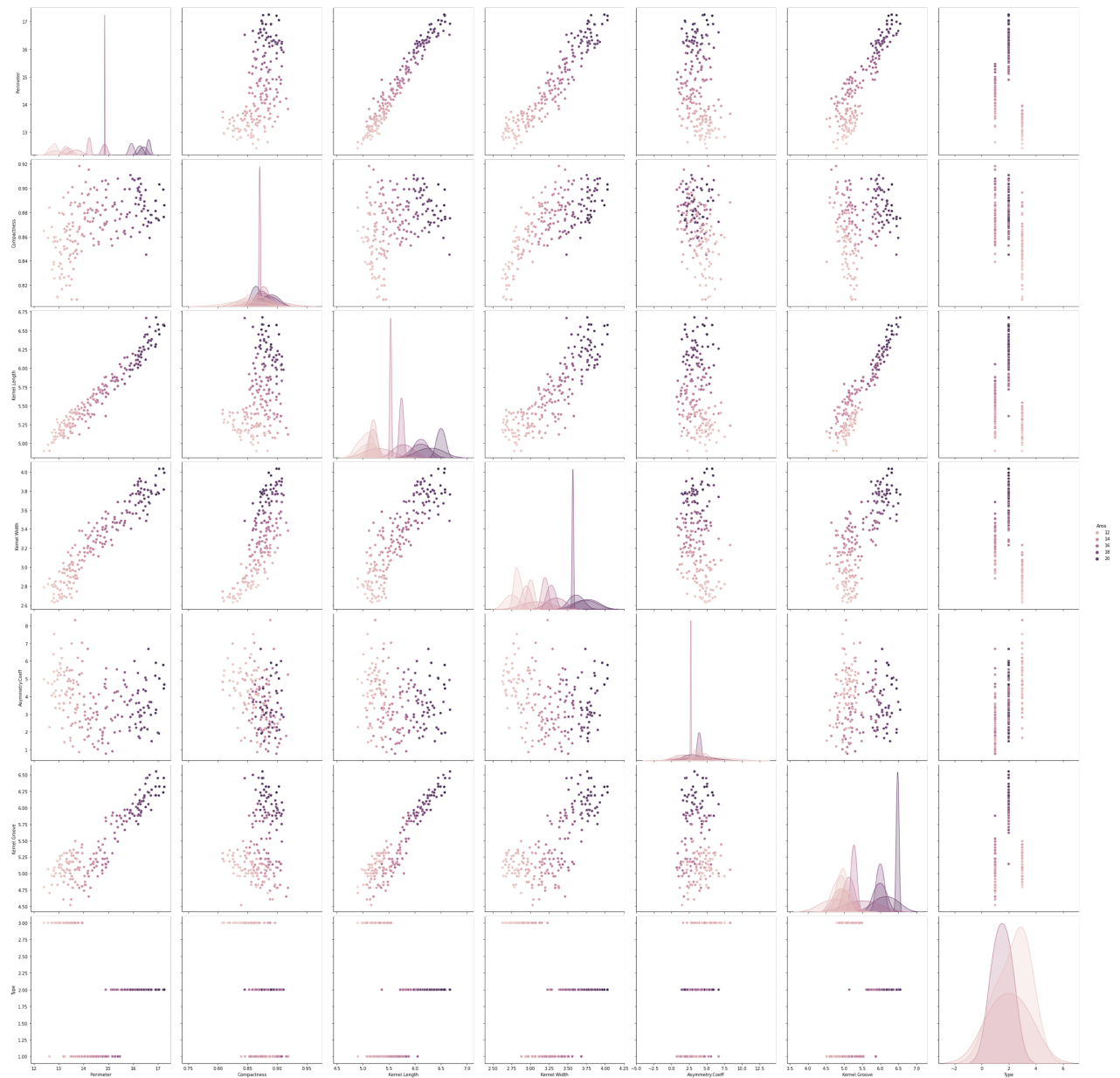
Out[28]: `<seaborn.axisgrid.PairGrid at 0x7fadb27e3dc0>`

```
In [41]:  import warnings
          warnings.filterwarnings("ignore")
          sns.pairplot(df , hue = 'Area', size = 5)
```
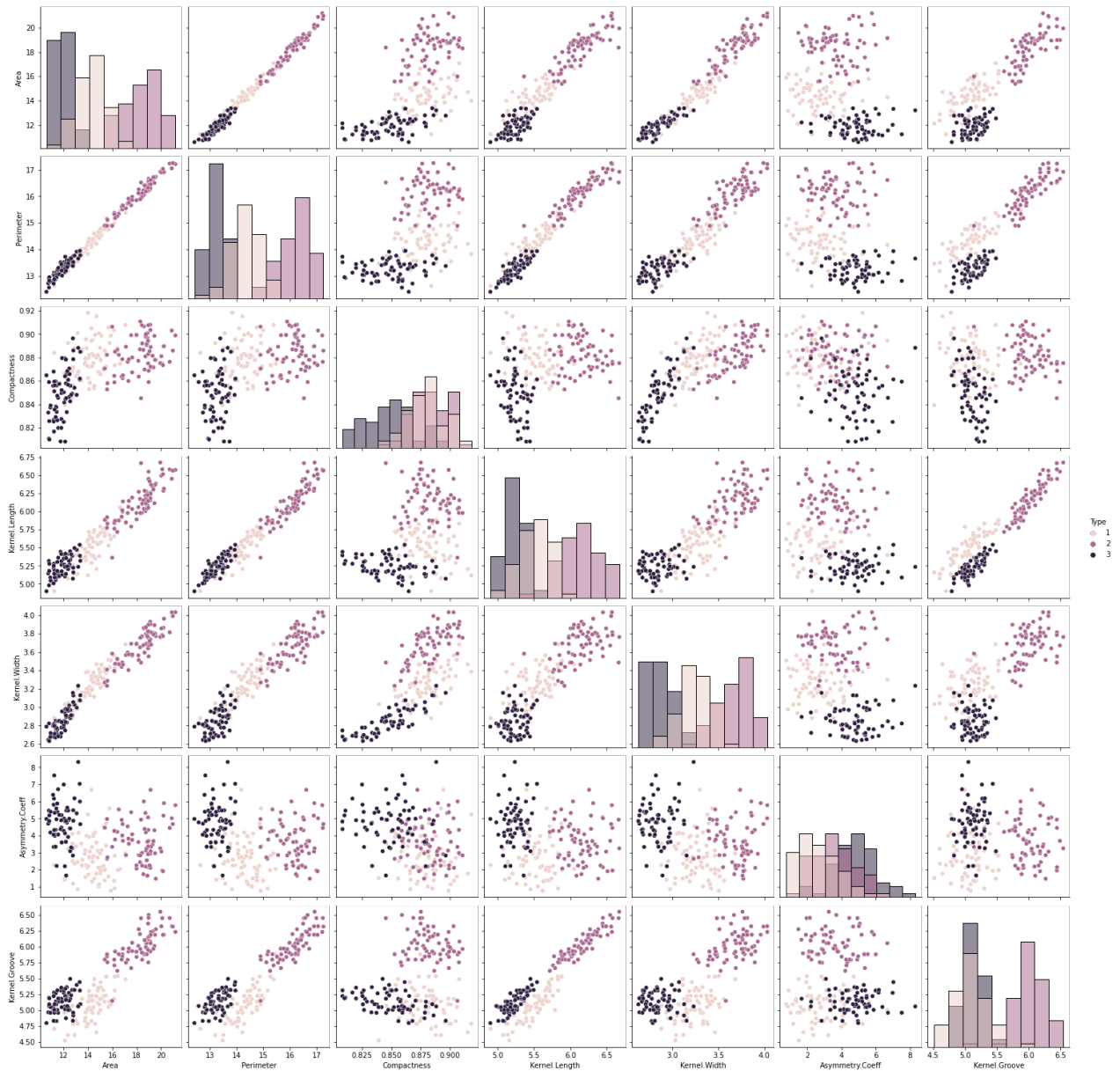
Out[41]:    `<seaborn.axisgrid.PairGrid at 0x7fadd0ee0ca0>`



In [39]:
```python
sns.pairplot(df, hue = "Type", size = 3, diag_kind = 'hist')
```

```
/opt/anaconda3/lib/python3.8/site-packages/seaborn/axisgrid.py:1912: UserWa
rning: The `size` parameter has been renamed to `height`; please update you
r code.
  warnings.warn(msg, UserWarning)
```

Out[39]:    `<seaborn.axisgrid.PairGrid at 0x7fadb4594b20>`



In [40]:    `sns.pairplot(df, hue = 'Area', size = 3, diag_kind = 'hist')`

```
/opt/anaconda3/lib/python3.8/site-packages/seaborn/axisgrid.py:1912: UserWa
rning: The `size` parameter has been renamed to `height`; please update you
r code.
  warnings.warn(msg, UserWarning)
```

Out[40]:    <seaborn.axisgrid.PairGrid at 0x7fadc169f3a0>