In [44]:

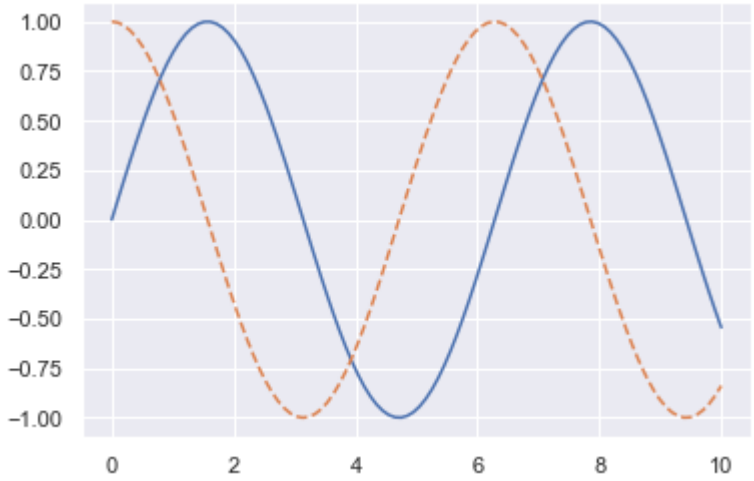
```
# maha ebrahim mohammed
# 4051350
# IA8G
# lab 5: Data Exploratory Analysis

import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
sns.set(color_codes=True)
```

In [45]:

```
x = np.linspace(0, 10, 100)

fig = plt.figure()
plt.plot(x, np.sin(x), '-')
plt.plot(x, np.cos(x), '--')
```

Out[45]:

```
[<matplotlib.lines.Line2D at 0x238a462c400>]
```



In [47]:

```
from sklearn import datasets
iris = datasets.load_iris()
df = pd.read_csv(iris.filename)
df.head(5)
```

Out[47]:

	150	4	setosa	versicolor	virginica
0	5.1	3.5	1.4	0.2	0
1	4.9	3.0	1.4	0.2	0
2	4.7	3.2	1.3	0.2	0
3	4.6	3.1	1.5	0.2	0
4	5.0	3.6	1.4	0.2	0

In [48]:

```
df_with_filepath = pd.read_csv('C:\\anaconda3\\lib\\site-packages\\sklearn\\datasets\\data\
df_with_filepath.head()
```

Out[48]:

	5.1	3.5	1.4	0.2	0
0	4.9	3.0	1.4	0.2	0
1	4.7	3.2	1.3	0.2	0
2	4.6	3.1	1.5	0.2	0
3	5.0	3.6	1.4	0.2	0
4	5.4	3.9	1.7	0.4	0

In [14]:

```
df_with_filepath.head()
```

Out[14]:

	4.7	3.2	1.3	0.2	0
0	4.6	3.1	1.5	0.2	0
1	5.0	3.6	1.4	0.2	0
2	5.4	3.9	1.7	0.4	0
3	4.6	3.4	1.4	0.3	0
4	5.0	3.4	1.5	0.2	0

In [18]:

```
iris.feature_names
```

Out[18]:

```
['sepal length (cm)',
 'sepal width (cm)',
 'petal length (cm)',
 'petal width (cm)']
```

In [37]:

```
df_with_columns = pd.read_csv('C:\\anaconda3\\lib\\site-packages\\sklearn\\datasets\\data\\
df_with_columns.head()
```

Out[37]:

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)	species
0	5.1	3.5	1.4	0.2	0
1	4.9	3.0	1.4	0.2	0
2	4.7	3.2	1.3	0.2	0
3	4.6	3.1	1.5	0.2	0
4	5.0	3.6	1.4	0.2	0

In [31]:

```
df = pd.read_csv(iris.filename, names = iris.feature_names)
df
```

Out[31]:

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)
150.0	4.0	setosa	versicolor	virginica
5.1	3.5	1.4	0.2	0
4.9	3.0	1.4	0.2	0
4.7	3.2	1.3	0.2	0
4.6	3.1	1.5	0.2	0
...	...	...	...	...
6.7	3.0	5.2	2.3	2
6.3	2.5	5.0	1.9	2
6.5	3.0	5.2	2.0	2
6.2	3.4	5.4	2.3	2
5.9	3.0	5.1	1.8	2

151 rows × 4 columns

In [32]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Float64Index: 151 entries, 150.0 to 5.9
Data columns (total 4 columns):
 #   Column             Non-Null Count  Dtype  
---  ------             --------------  -----  
 0   sepal length (cm)  151 non-null    float64
 1   sepal width (cm)   151 non-null    object 
 2   petal length (cm)  151 non-null    object 
 3   petal width (cm)   151 non-null    object 
dtypes: float64(1), object(3)
memory usage: 5.9+ KB
```

In [38]:

```
print(df_with_columns.shape)
print(df_with_columns.columns)
```

```
(150, 5)
Index(['sepal length (cm)', 'sepal width (cm)', 'petal length (cm)',
       'petal width (cm)', 'species'],
      dtype='object')
```

In [39]:

```
df_with_columns.isnull()
```

Out[39]:

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)	species
0	False	False	False	False	False
1	False	False	False	False	False
2	False	False	False	False	False
3	False	False	False	False	False
4	False	False	False	False	False
...	...	...	...	...	...
145	False	False	False	False	False
146	False	False	False	False	False
147	False	False	False	False	False
148	False	False	False	False	False
149	False	False	False	False	False

150 rows × 5 columns

In [40]:

```
print(df_with_columns["species"].value_counts())
```

```
0    50
1    50
2    50
Name: species, dtype: int64
```

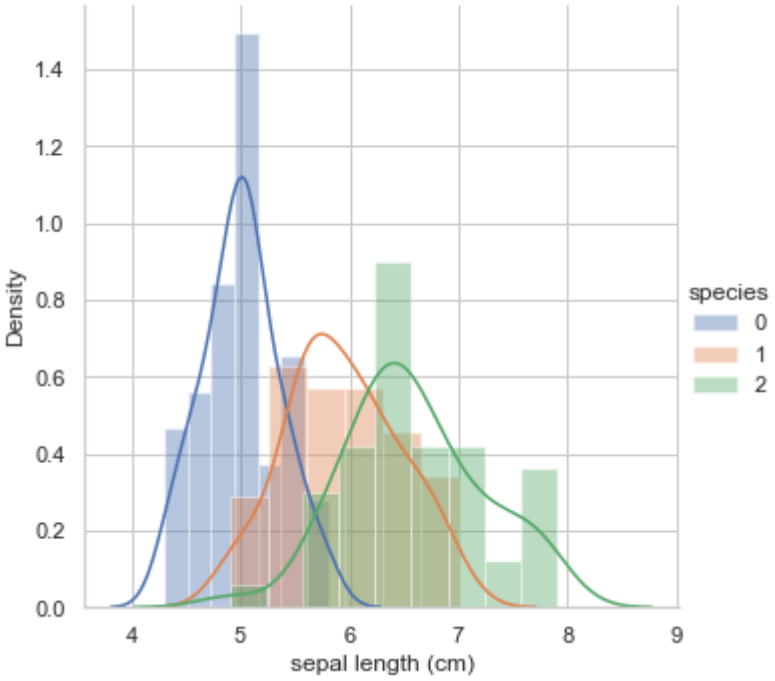
In [41]:

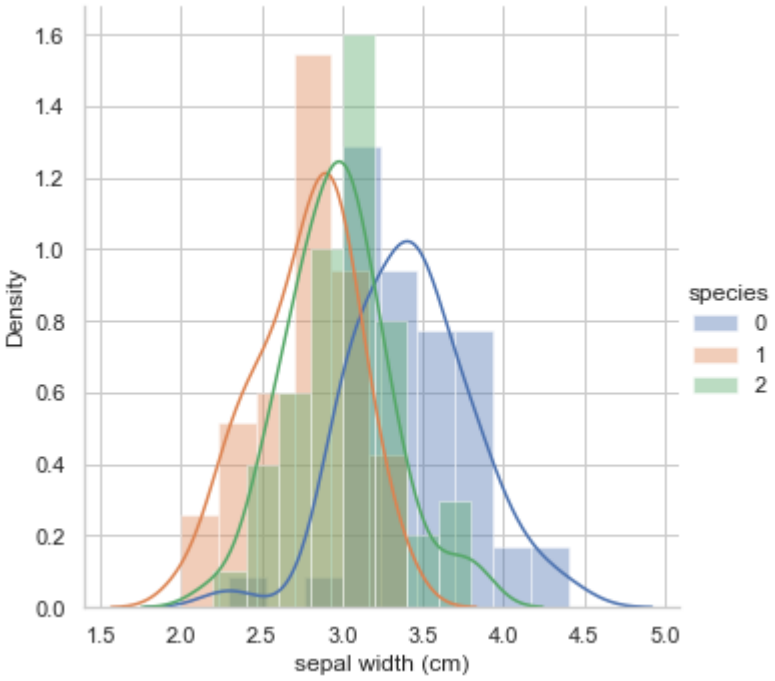
```
print(df_with_columns.describe())
```

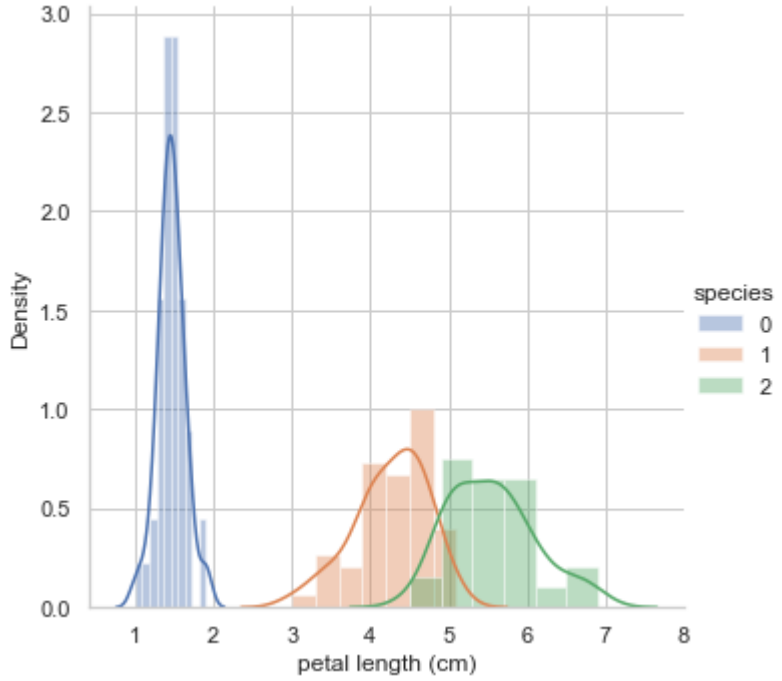
```
       sepal length (cm)  sepal width (cm)  petal length (cm)  \
count         150.000000        150.000000         150.000000
mean            5.843333          3.057333           3.758000
std             0.828066          0.435866           1.765298
min             4.300000          2.000000           1.000000
25%             5.100000          2.800000           1.600000
50%             5.800000          3.000000           4.350000
75%             6.400000          3.300000           5.100000
max             7.900000          4.400000           6.900000

       petal width (cm)     species
count        150.000000  150.000000
mean           1.199333    1.000000
std            0.762238    0.819232
min            0.100000    0.000000
25%            0.300000    0.000000
50%            1.300000    1.000000
75%            1.800000    2.000000
max            2.500000    2.000000
```

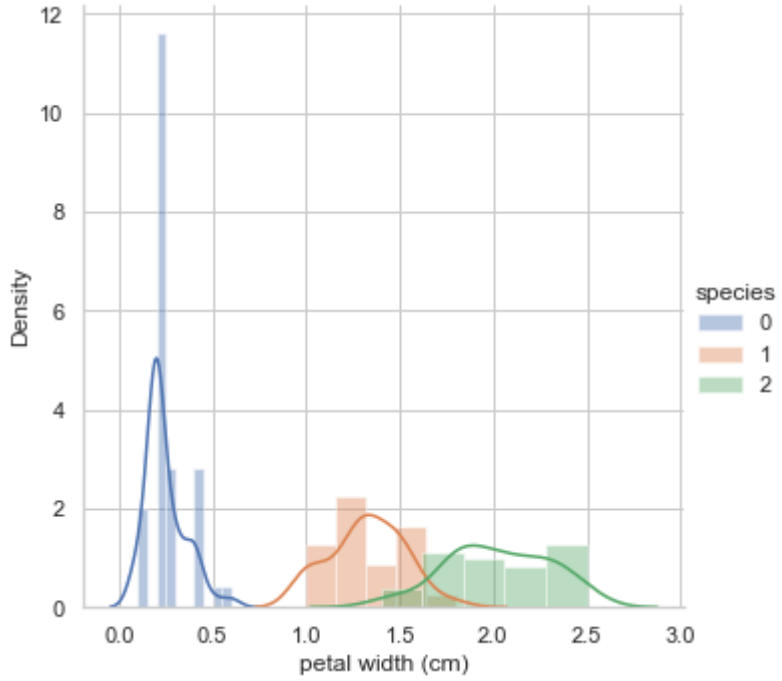
In [57]:

```
import warnings
warnings.simplefilter(action='ignore', category=FutureWarning)

for ojha, feature in enumerate(list(df_with_columns.columns)[:-1]):
    fg = sns.FacetGrid(df_with_columns, hue ='species', height=5)
    fg.map(sns.distplot, feature).add_legend()
plt.show()
```



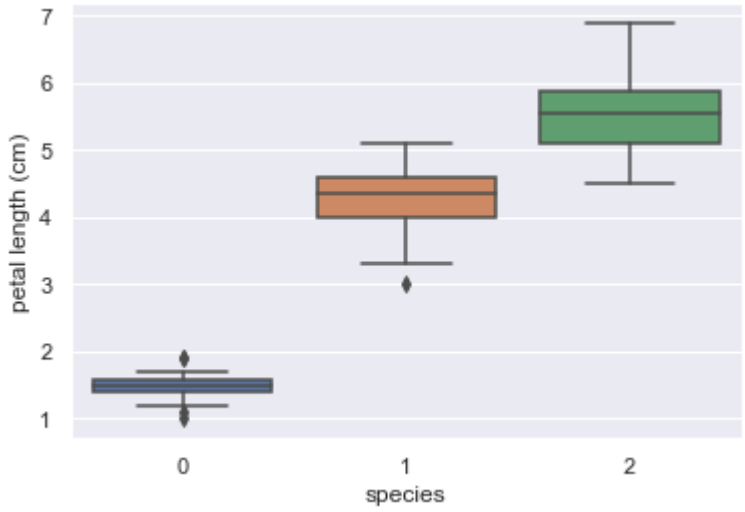



Density
petal length (cm)
species
0
1
2


Density
petal width (cm)
species
0
1
2

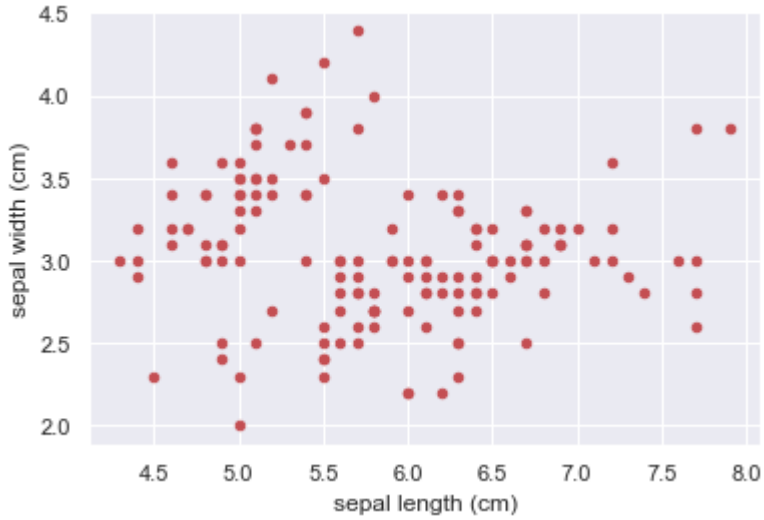
In [52]:

```
sns.boxplot(x='species',y='petal length (cm)', data=df_with_columns)
plt.show()
```



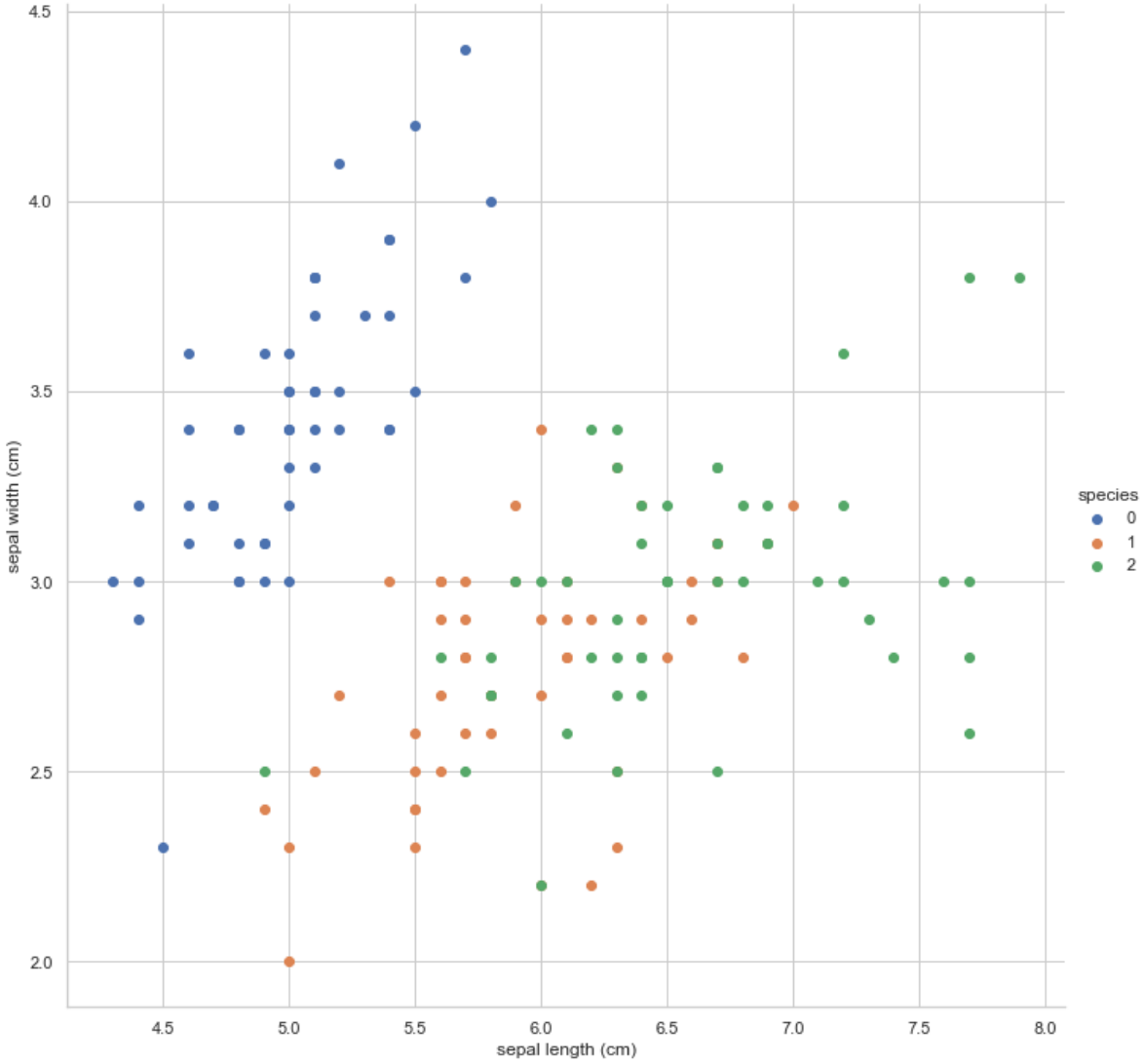
In [54]:

```
df_with_columns.plot(kind= 'scatter' ,x= 'sepal length (cm)', y= 'sepal width (cm)', c = 'r
plt.show()
```



In [55]:

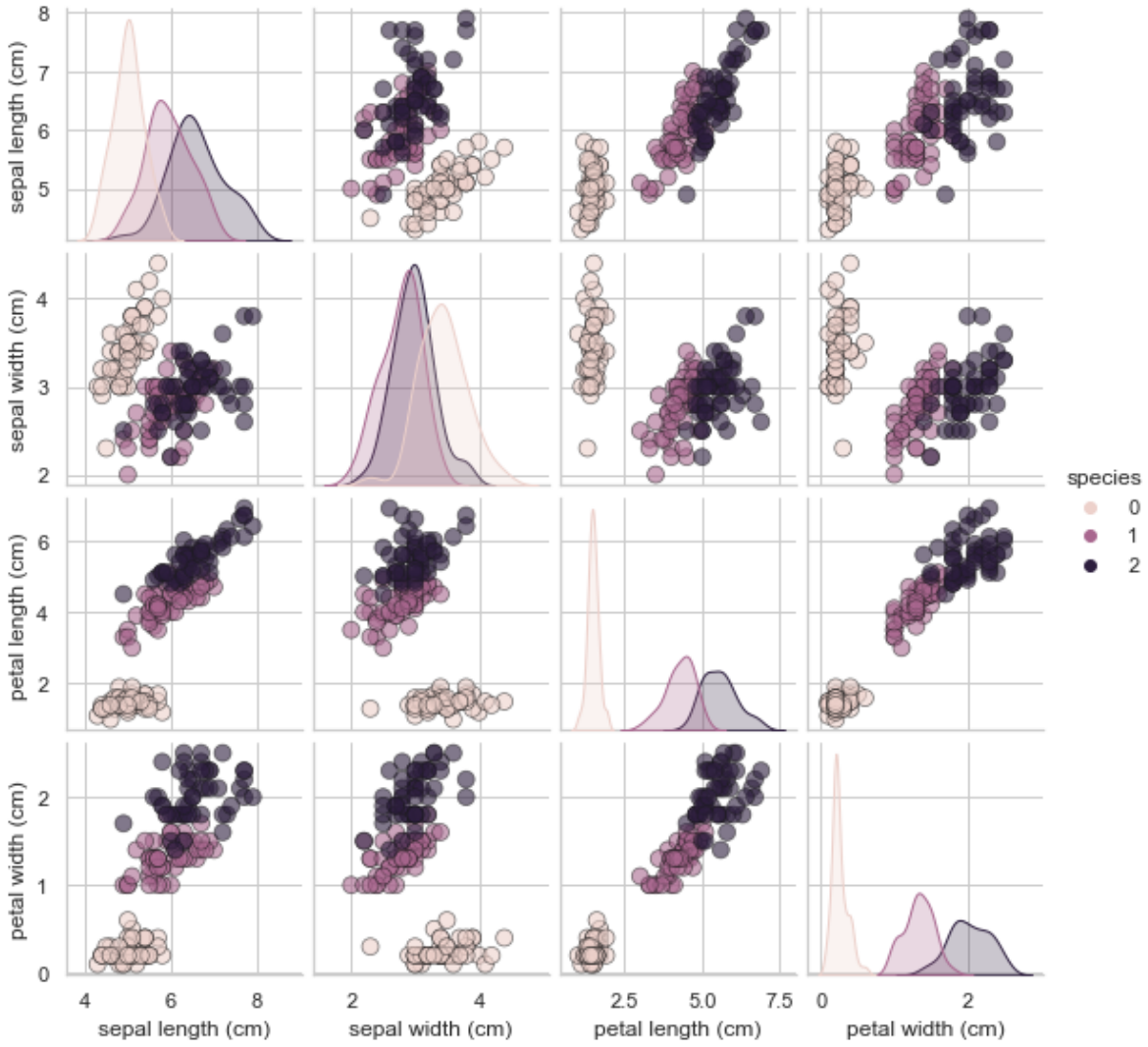
```
sns.set_style("whitegrid");
sns.FacetGrid(df_with_columns, hue="species", height=10) \
    .map(plt.scatter, "sepal length (cm)", "sepal width (cm)") \
    .add_legend();
plt.show();
```



In [56]:

```
sns.set_style("whitegrid");
sns.pairplot(df_with_columns,vars = iris.feature_names, hue = 'species', diag_kind='kde',
             plot_kws = {'alpha': 0.6, 's': 80, 'edgecolor': 'k'}, size = 2)
plt.show()
```

```
C:\anaconda3\lib\site-packages\seaborn\axisgrid.py:2076: UserWarning: The `s
ize` parameter has been renamed to `height`; please update your code.
  warnings.warn(msg, UserWarning)
```



In [ ]: