[1]:
```
import matplotlib.pyplot as plt
import seaborn as sns
import pandas as pd
import numpy as np


%matplotlib inline
```

[3]:
```
df=pd.read_csv("IRIS.csv")
print(df)
```

```
     sepal_length  sepal_width  petal_length  petal_width         species
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]
```

[4]:
```
df.head()
```

[4]:

	sepal_length	sepal_width	petal_length	petal_width	species
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

[5]:
```
df.shape
```

[5]:
```
(150, 5)
```

[6]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 150 entries, 0 to 149
Data columns (total 5 columns):
 #   Column        Non-Null Count  Dtype  
---  ------        --------------  -----  
 0   sepal_length  150 non-null    float64
 1   sepal_width   150 non-null    float64
 2   petal_length  150 non-null    float64
 3   petal_width   150 non-null    float64
 4   species       150 non-null    object 
dtypes: float64(4), object(1)
memory usage: 6.0+ KB
```

[7]:
```
df.describe()
```

[7]:

	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

[8]:
```
df.isnull()
```

[8]:

	sepal_length	sepal_width	petal_length	petal_width	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

[9]:
```
df.isnull().sum()
```

[9]:
```
sepal_length    0
sepal_width     0
petal_length    0
petal_width     0
species         0
dtype: int64
```

[10]:
```
df.duplicated()
```

[10]:
```
0      False
1      False
2      False
3      False
4      False
       ...
145    False
146    False
147    False
148    False
149    False
Length: 150, dtype: bool
```

jupyter IRIS Last Checkpoint: 23 minutes ago

File Edit View Run Kernel Settings Help

Trusted

Code

JupyterLab Python 3 (ipykernel)

[13]:
```
df.columns.tolist()
```

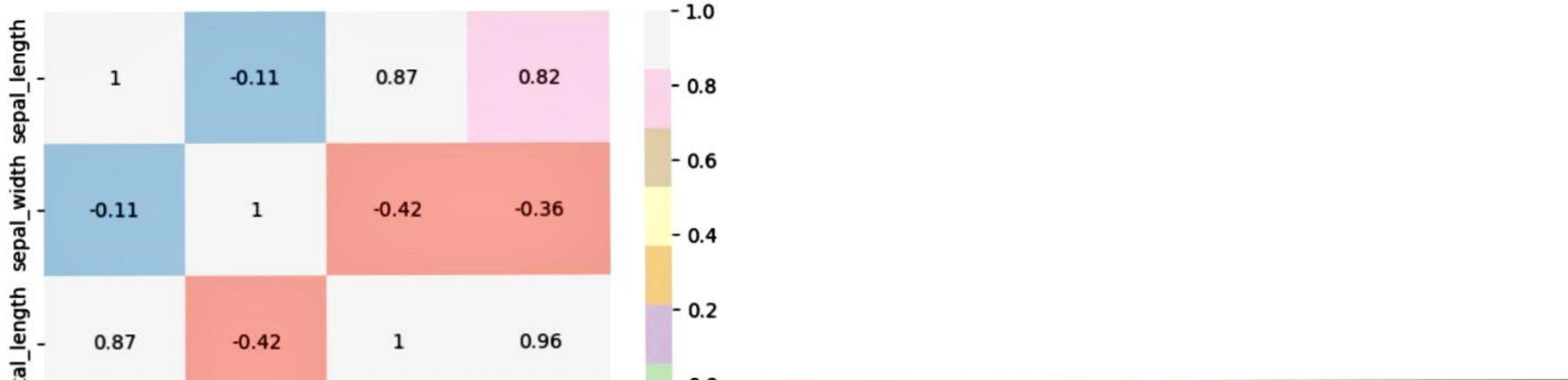
[13]:
```
['sepal_length', 'sepal_width', 'petal_length', 'petal_width', 'species']
```

[28]:
```
numeric_df = df.select_dtypes(include=[np.number])

# Compute the correlation matrix
corr_matrix = numeric_df.corr()
print(corr_matrix)
```

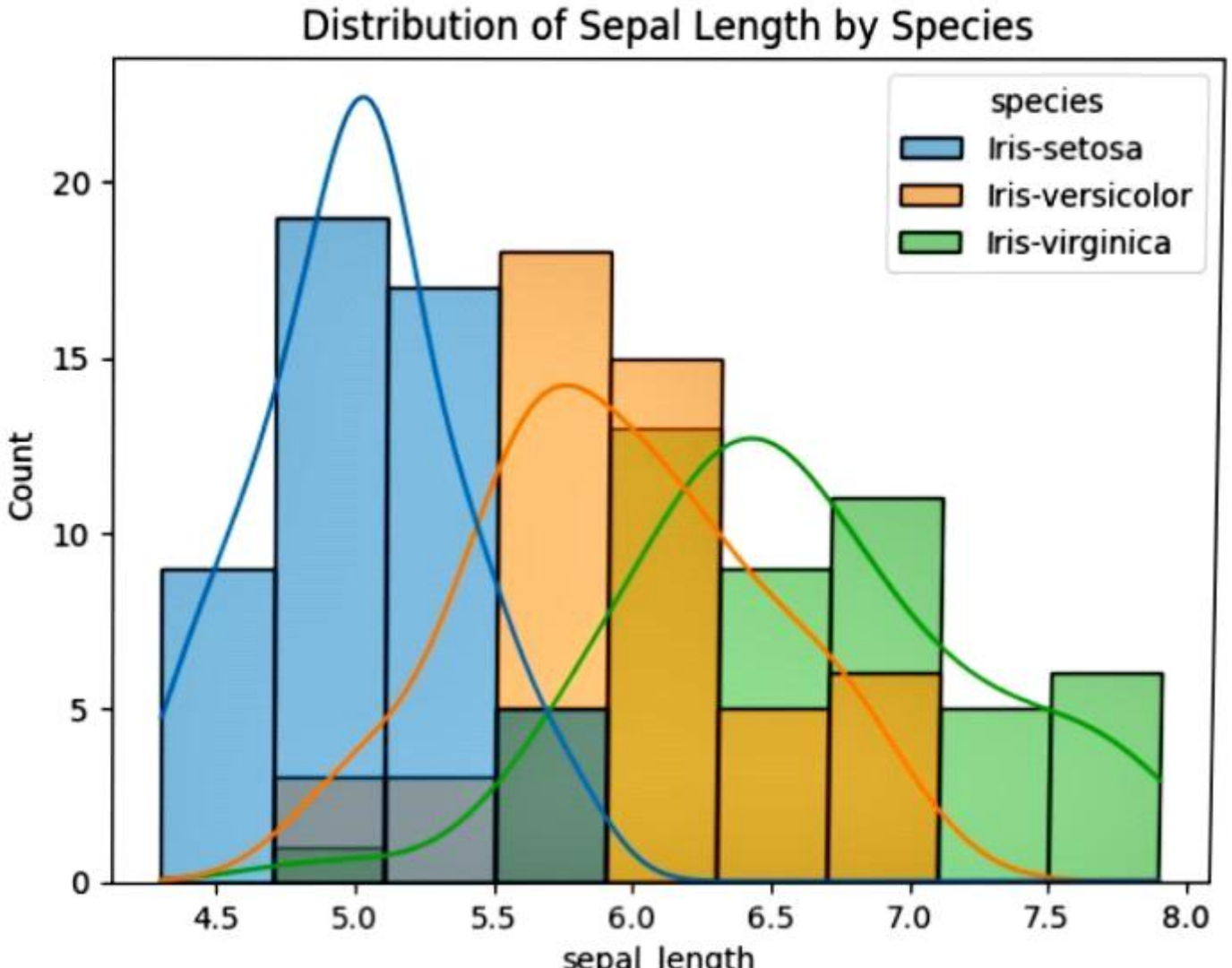
	sepal_length	sepal_width	petal_length	petal_width
sepal_length	1.000000	-0.109369	0.871754	0.817954
sepal_width	-0.109369	1.000000	-0.420516	-0.356544
petal_length	0.871754	-0.420516	1.000000	0.962757
petal_width	0.817954	-0.356544	0.962757	1.000000

[29]:
```
sns.heatmap(corr_matrix, annot=True, cmap="Pastel1")
```

[29]:
```
<Axes: >
```

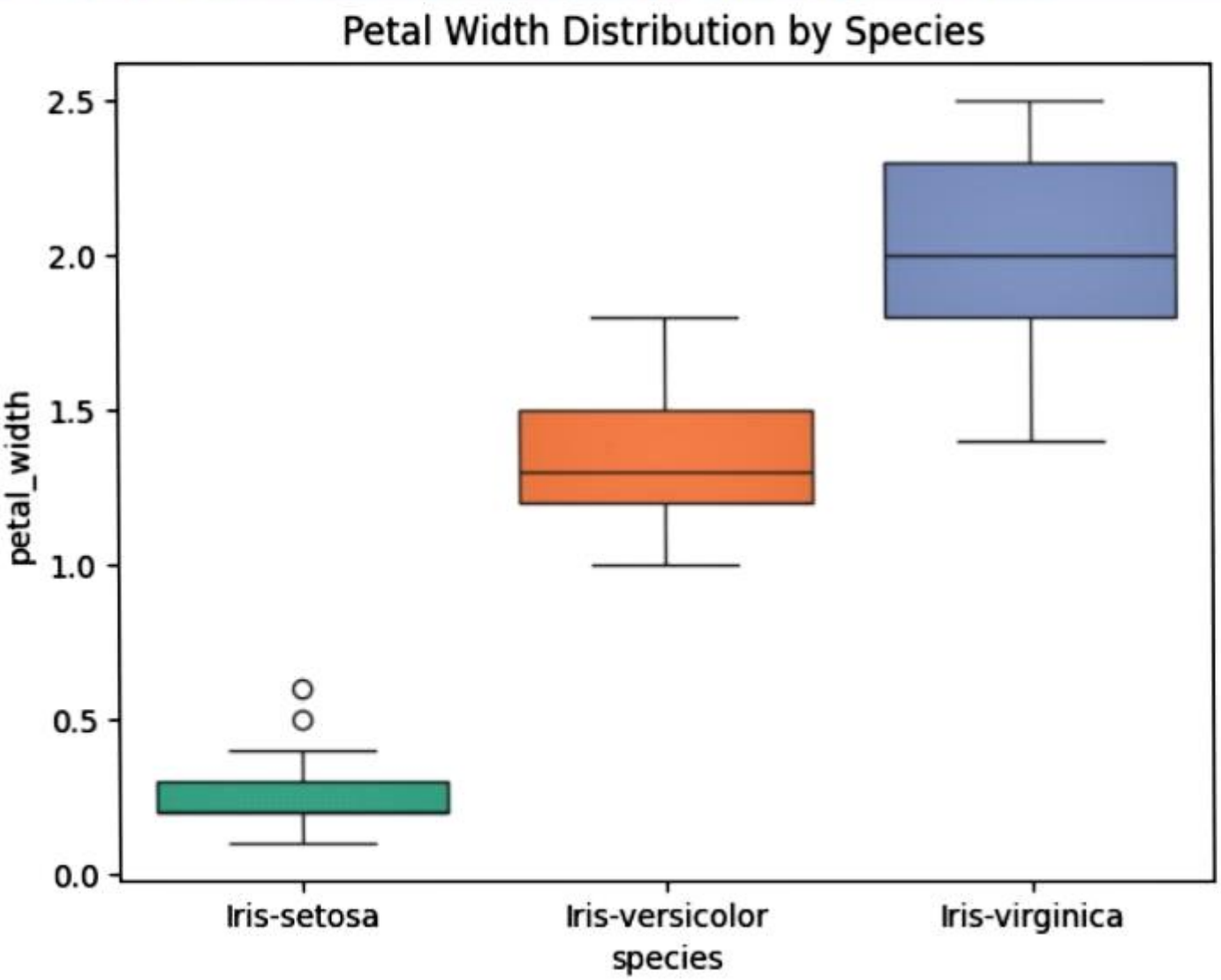


[30]:
```
sns.histplot(data=df, x='sepal_length', hue='species',kde=True)
plt.title('Distribution of Sepal Length by Species')
plt.show()
```



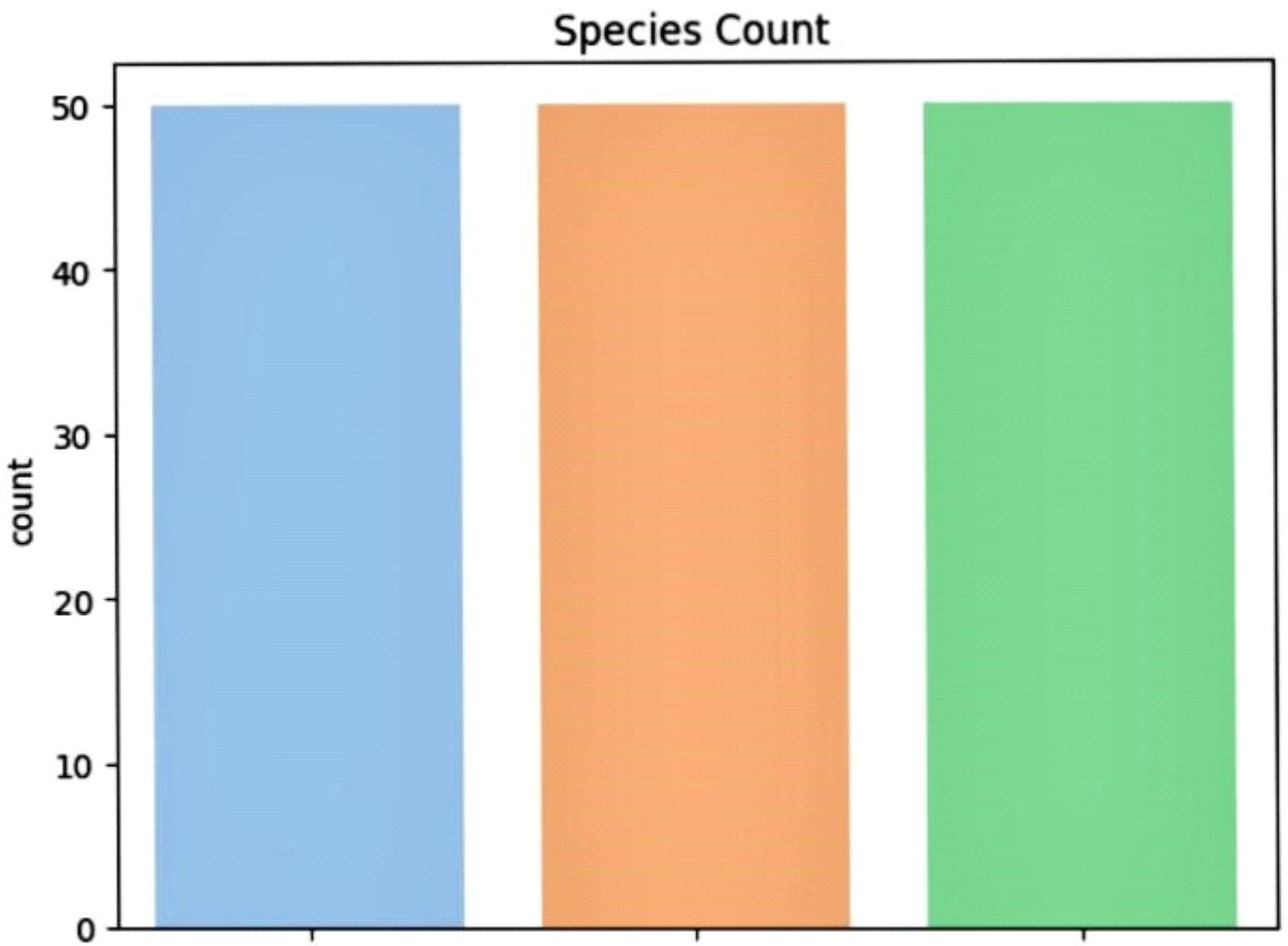
[31]:
```
sns.boxplot(data=df, x='species', y='petal_width',palette='Set2')
plt.title('Petal Width Distribution by Species')
plt.show()
```

```
C:\Users\user\AppData\Local\Temp\ipykernel_332\752195063.py:1: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable to
ame effect.

  sns.boxplot(data=df, x='species', y='petal_width',palette='Set2')
```



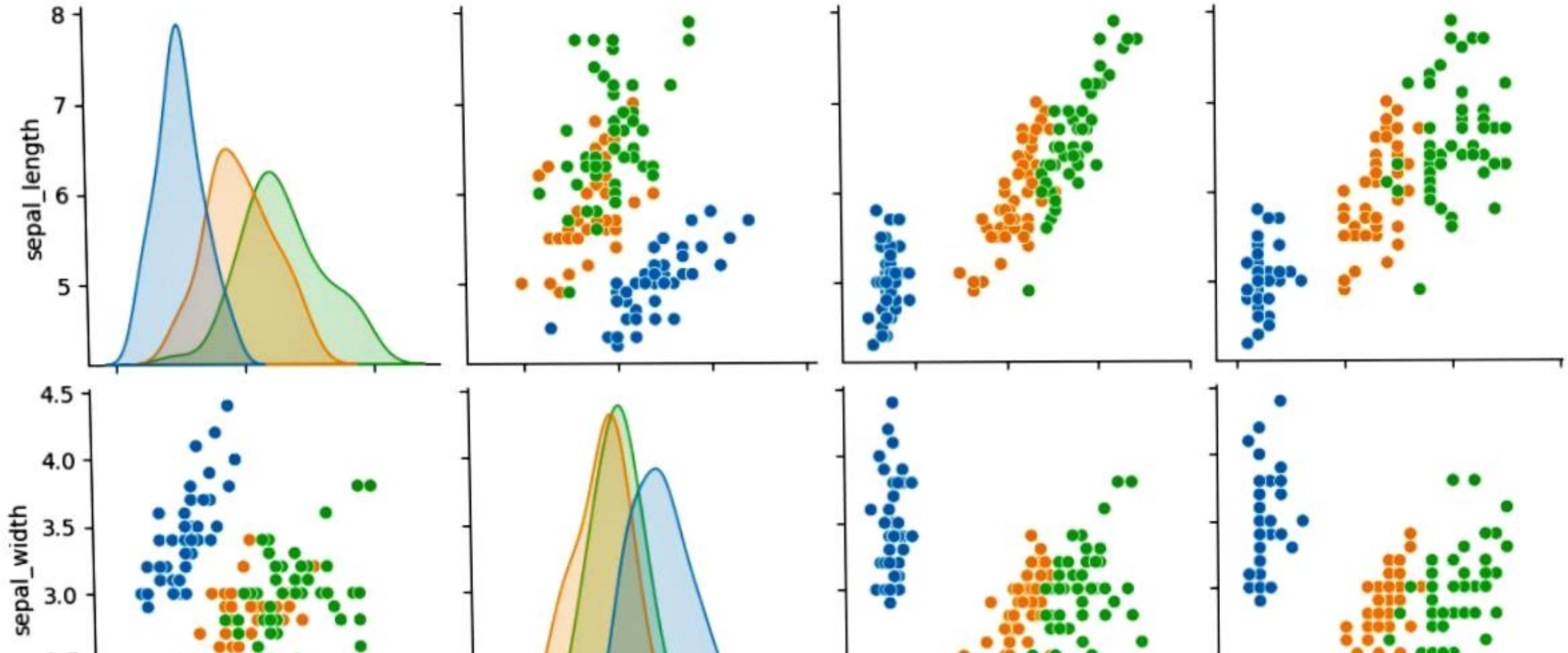
[32]:
```
sns.countplot(data=df, x='species',palette='pastel')
plt.title('Species Count')
plt.xticks(rotation=45)
plt.show()
```

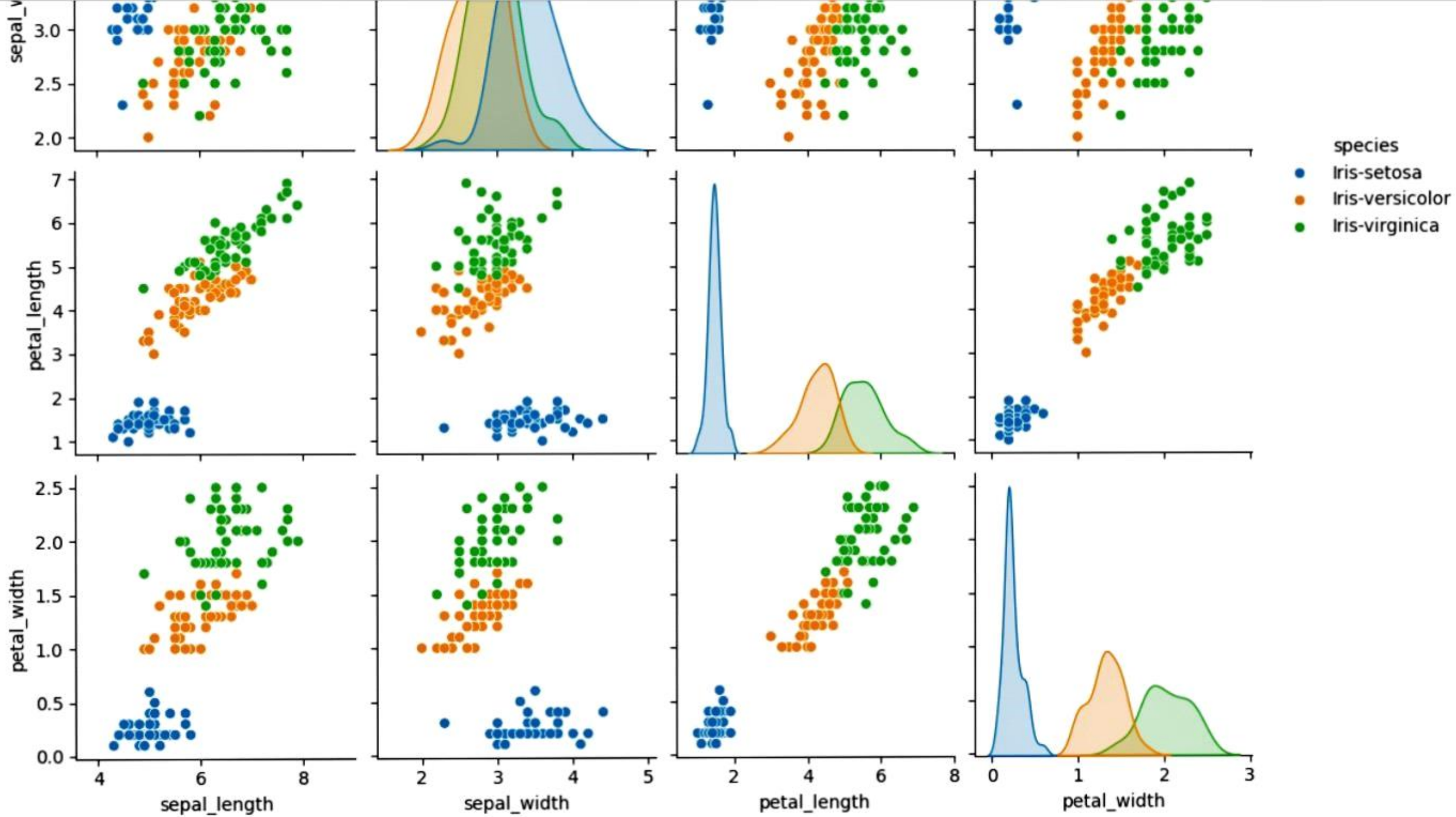
```
C:\Users\user\AppData\Local\Temp\ipykernel_332\1319106465.py:1: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable to `hue` an
ame effect.

  sns.countplot(data=df, x='species',palette='pastel')
```



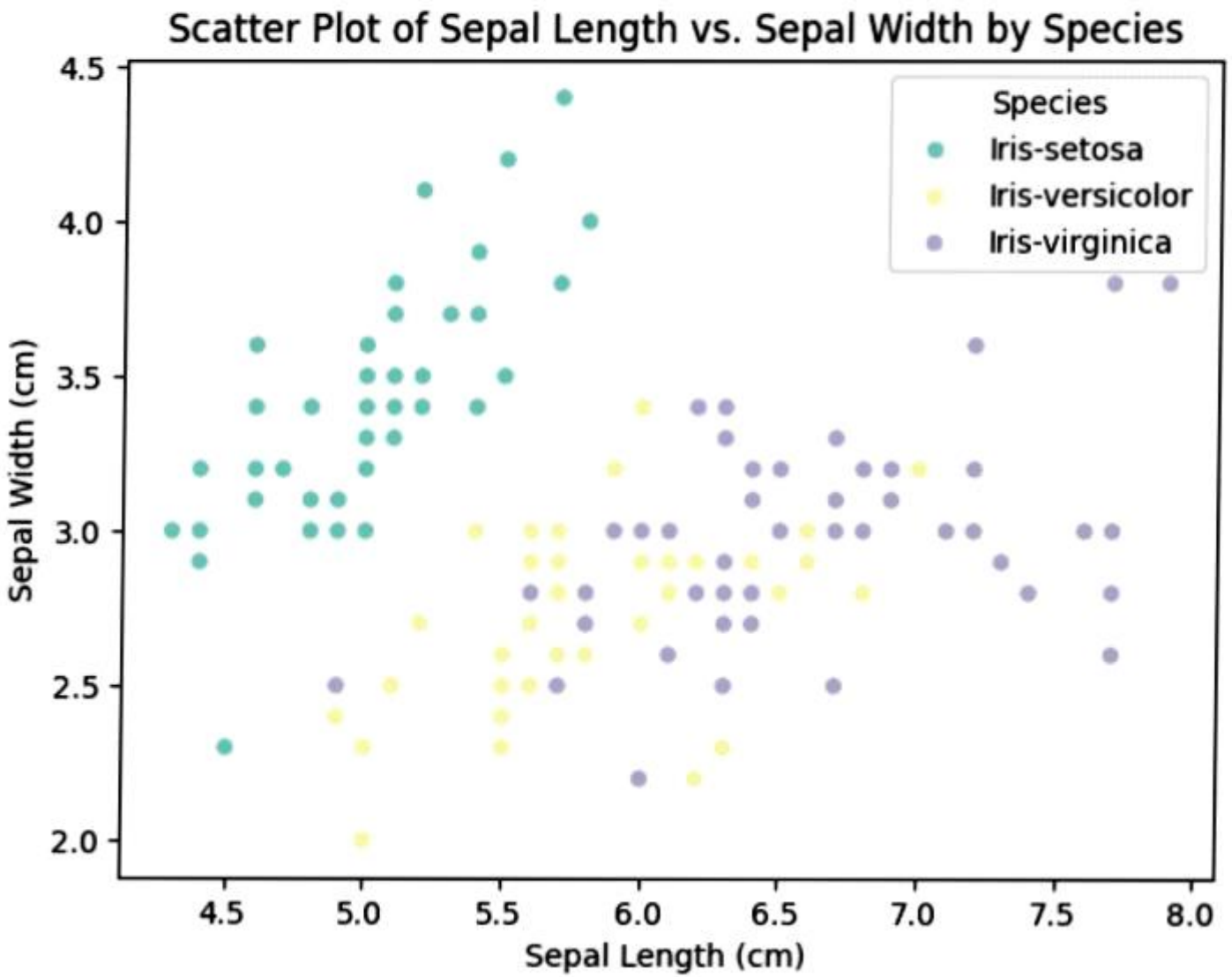
[*]:
```
g = sns.PairGrid(df, hue='species')
g.map_upper(sns.scatterplot)
g.map_diag(sns.histplot)
g.map_lower(sns.kdeplot)
g.add_legend()
plt.suptitle('PairGrid of Iris Dataset')
plt.show()
```

[34]:
```
sns.pairplot(df, hue='species')
plt.show()
```

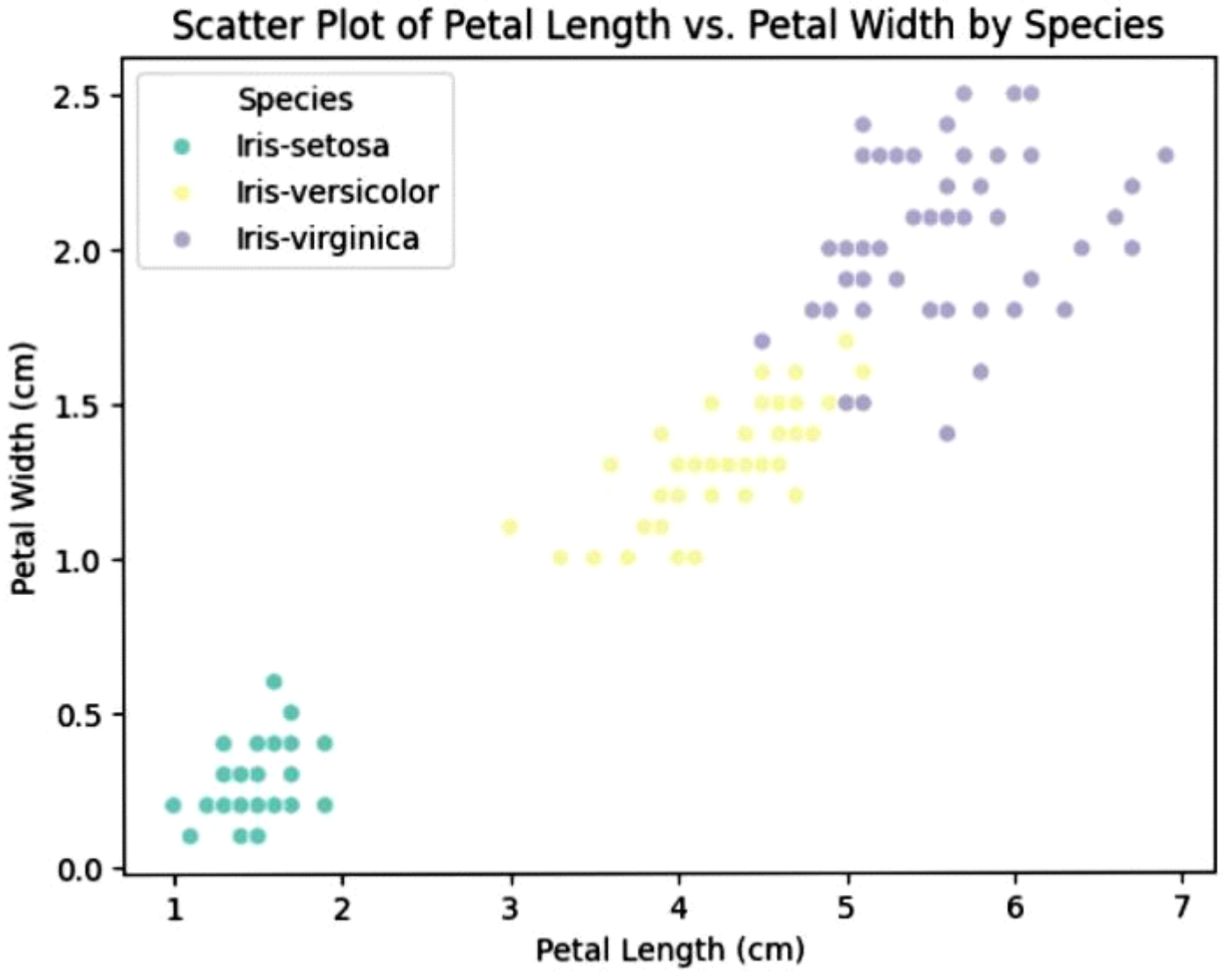



sepal_length
sepal_width
petal_length
petal_width
species
Iris-setosa
Iris-versicolor
Iris-virginica

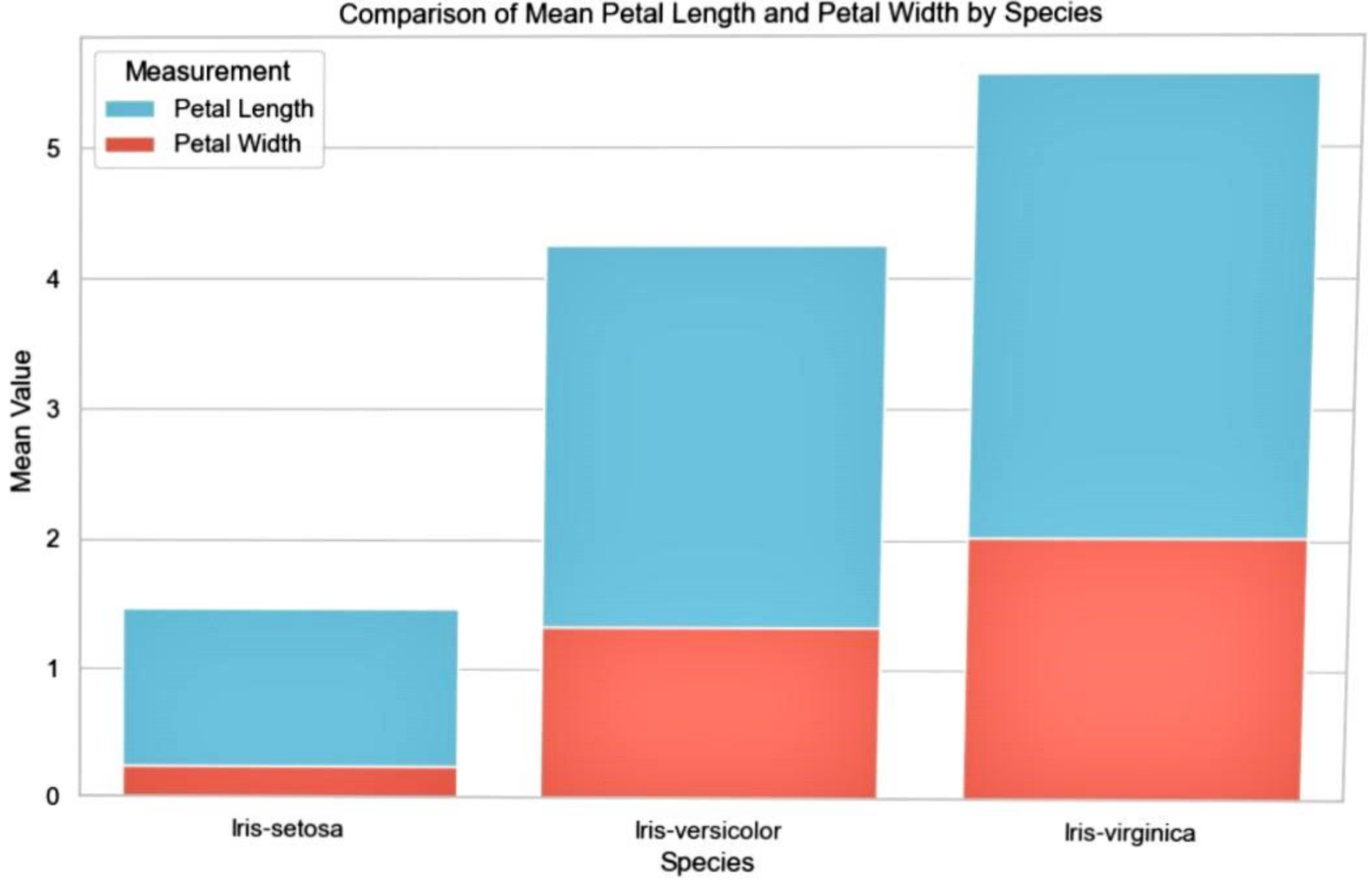
35]:
```
sns.scatterplot(data=df, x='sepal_length', y='sepal_width', hue='species', palette='Set3')
plt.title('Scatter Plot of Sepal Length vs. Sepal Width by Species')
plt.xlabel('Sepal Length (cm)')
plt.ylabel('Sepal Width (cm)')
plt.legend(title='Species')
plt.show()
```



[36]:
```
sns.scatterplot(data=df, x='petal_length', y='petal_width', hue='species', palette='Set3')
plt.title('Scatter Plot of Petal Length vs. Petal Width by Species')
plt.xlabel('Petal Length (cm)')
plt.ylabel('Petal Width (cm)')
plt.legend(title='Species')
plt.show()
```

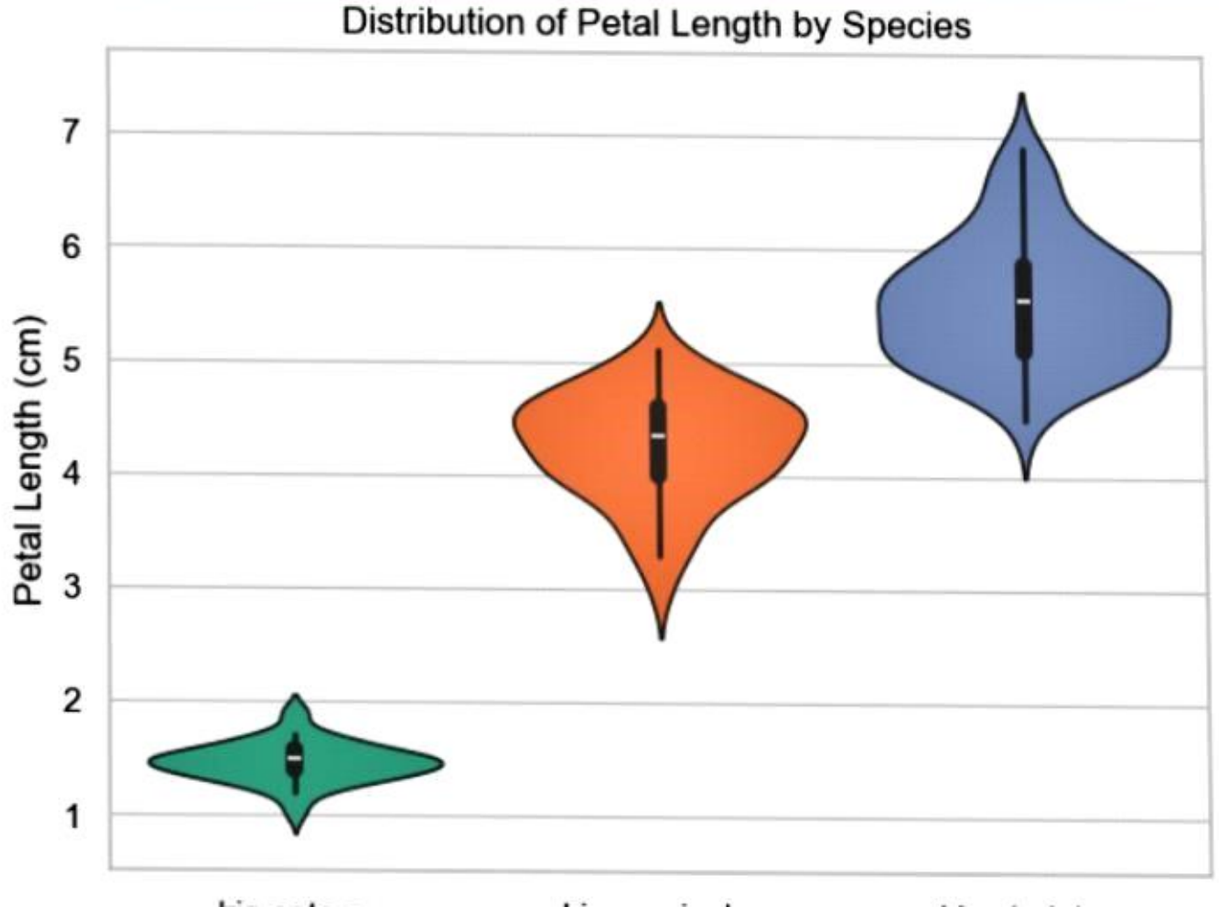


```
]: sns.set(style="whitegrid")
   plt.figure(figsize=(10, 6))

   sns.barplot(data=df, x='species', y='petal_length', estimator=np.mean, ci=None, color='skyblue', label='Petal Length')
   sns.barplot(data=df, x='species', y='petal_width', estimator=np.mean, ci=None, color='salmon', label='Petal Width')

   plt.title('Comparison of Mean Petal Length and Petal Width by Species')
   plt.xlabel('Species')
   plt.ylabel('Mean Value')
   plt.legend(title='Measurement')

   plt.show()
```

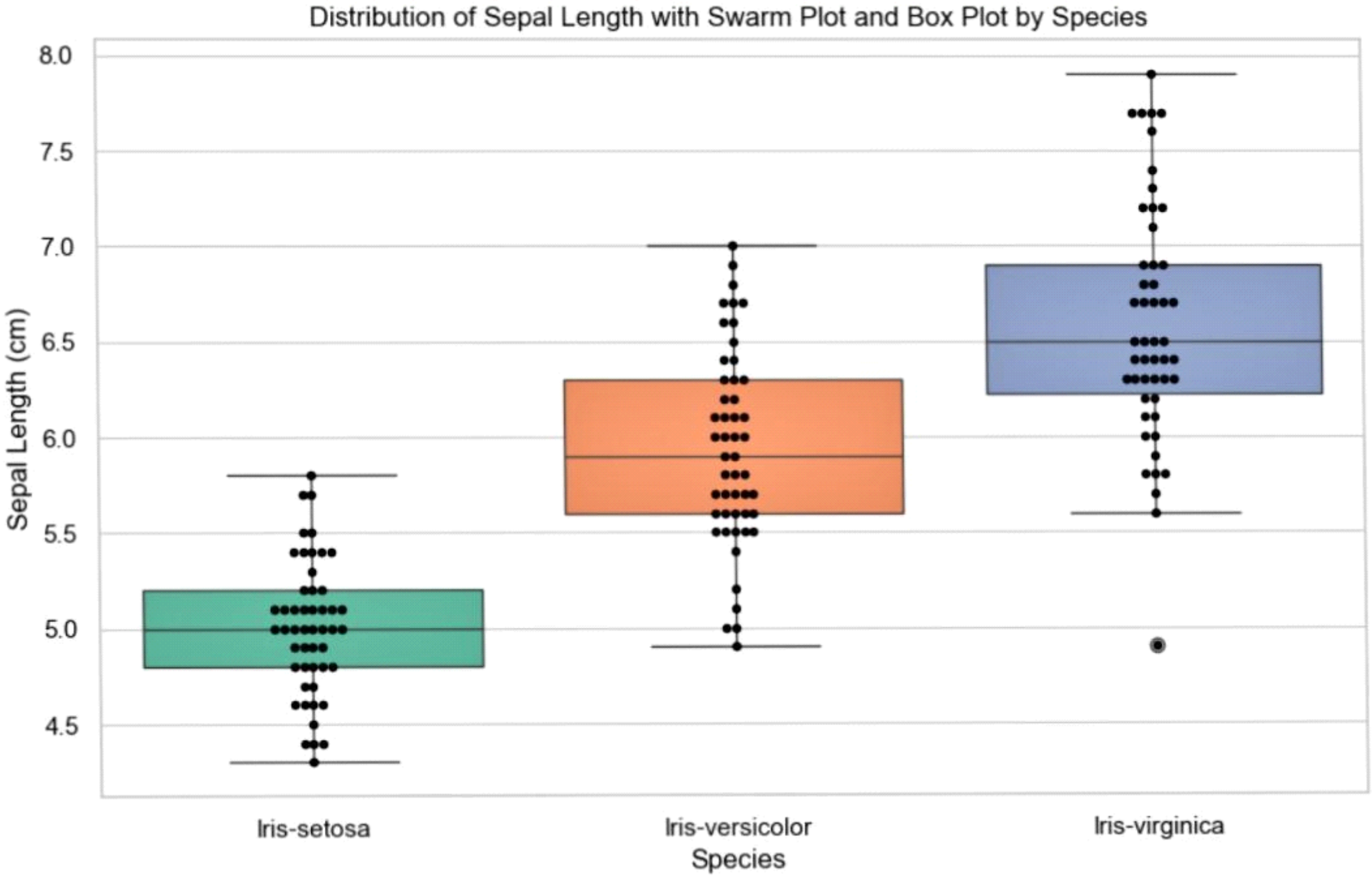

Comparison of Mean Petal Length and Petal Width by Species
Measurement
Petal Length
Petal Width
Mean Value
0
1
2
3
4
5
Iris-setosa
Iris-versicolor
Iris-virginica
Species

[38]:
```
sns.violinplot(data=df, x='species', y='petal_length', palette='Set2')
plt.title('Distribution of Petal Length by Species')
plt.xlabel('Species')
plt.ylabel('Petal Length (cm)')
plt.show()
```

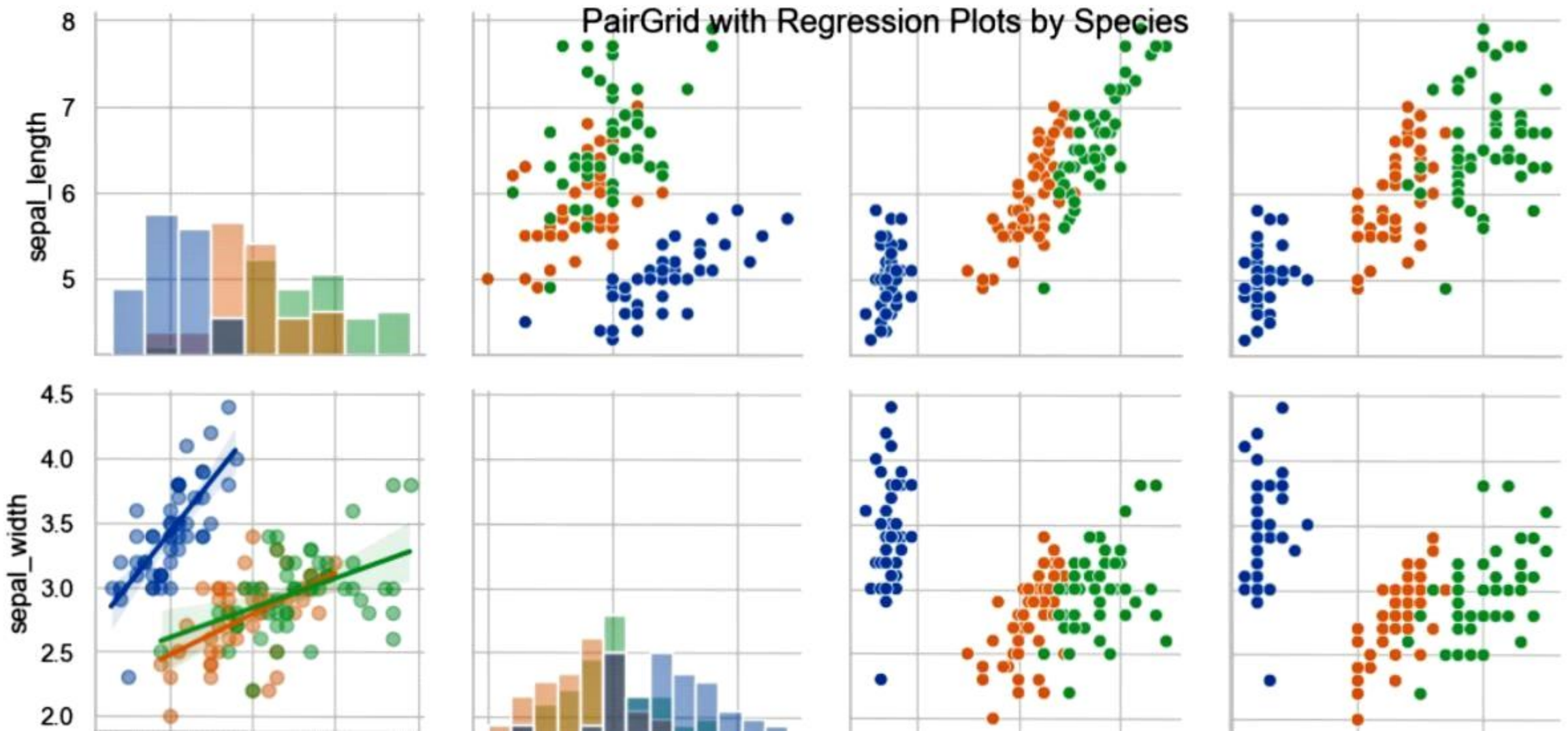
```
C:\Users\user\AppData\Local\Temp\ipykernel_332\3660947341.py:1: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable to `hue` and set `legend=False` fo
ame effect.

  sns.violinplot(data=df, x='species', y='petal_length', palette='Set2')
```

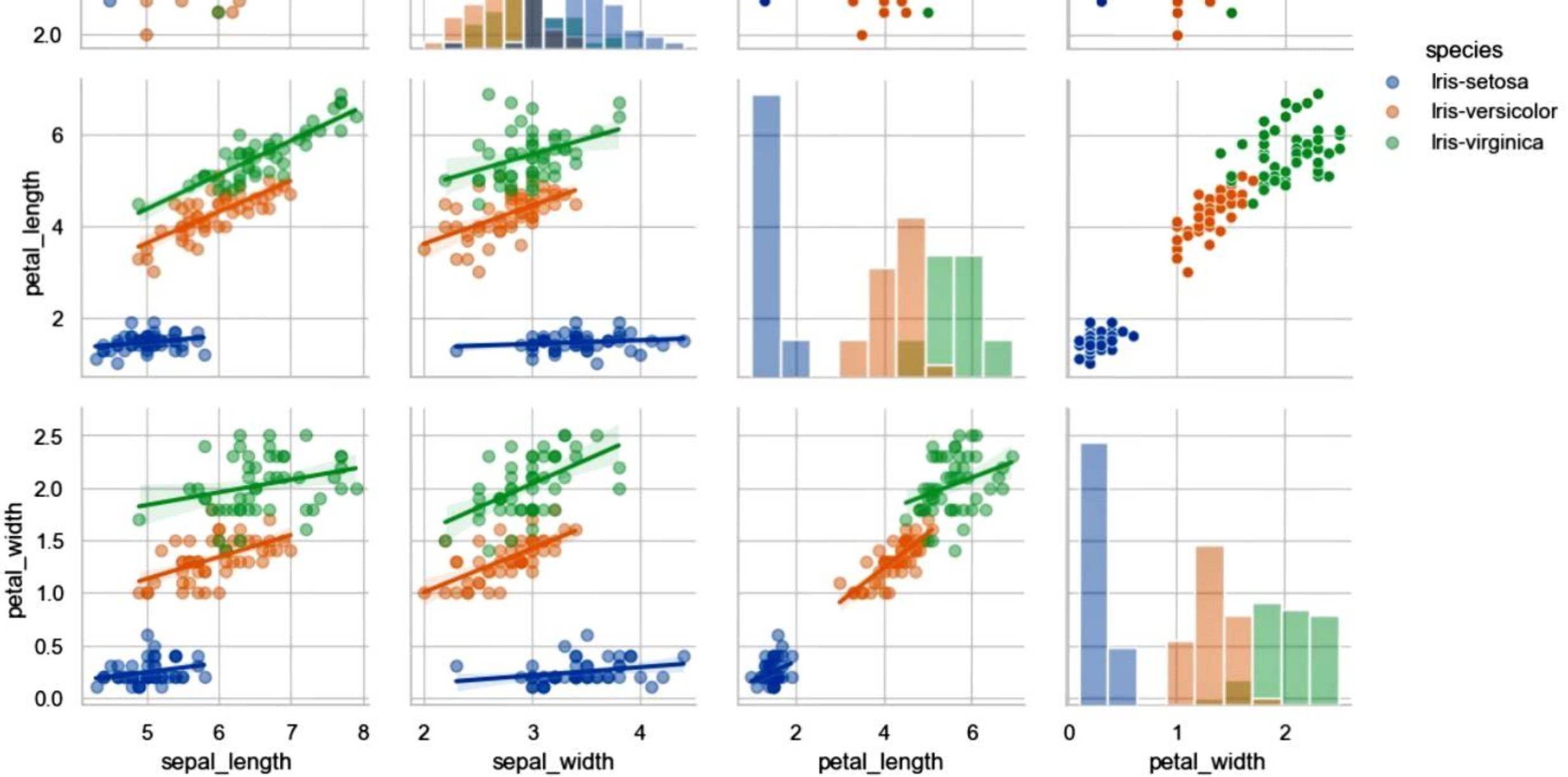


```
plt.figure(figsize=(10, 6))
sns.boxplot(data=df, x='species', y='sepal_length', palette='Set2')
sns.swarmplot(data=df, x='species', y='sepal_length', color='black', size=4)
plt.title('Distribution of Sepal Length with Swarm Plot and Box Plot by Species')
plt.xlabel('Species')
plt.ylabel('Sepal Length (cm)')
plt.show()
```


Distribution of Sepal Length with Swarm Plot and Box Plot by Species
Sepal Length (cm)
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
Iris-setosa
Iris-versicolor
Iris-virginica
Species

[43]:
```
g = sns.PairGrid(df, hue='species')
g.map_upper(sns.scatterplot)
g.map_diag(sns.histplot)
g.map_lower(sns.regplot, scatter_kws={'alpha': 0.5})
g.add_legend()
plt.suptitle('PairGrid with Regression Plots by Species')
plt.show()
```




species
Iris-setosa
Iris-versicolor
Iris-virginica
sepal_length
sepal_width
petal_length
petal_width