In [1]:
```
import pandas as pd
import datasist as ds
```

In [5]:
```
df = pd.read_csv('C:/Users/sundara.rao.ext/Desktop/SUNDAR/R/Social Prachar Mat
erial/Codes With Examples - Python/iris.csv')
```

In [6]:
```
ds.structdata.describe(df)
```

First five data points

	sepal length	sepal width	petal length	petal width	iris
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

Last five data points

	sepal length	sepal width	petal length	petal width	iris
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

Shape of data set: (150, 5)

Size of data set: 750

Data Types
Note: All Non-numerical features are identified as objects in pandas

	Data Type
sepal length	float64
sepal width	float64
petal length	float64
petal width	float64
iris	object

Numerical Features in Data set
['sepal length', 'sepal width', 'petal length', 'petal width']

Statistical Description of Columns

	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

Categorical Features in Data set

['iris']

Unique class Count of Categorical features

	Feature	Unique Count
0	iris	3

Missing Values in Data

	features	missing_counts	missing_percent
0	sepal length	0	0.0
1	sepal width	0	0.0
2	petal length	0	0.0
3	petal width	0	0.0
4	iris	0	0.0

In [7]:
```
#Gets all categorical features
cats = ds.structdata.get_cat_feats(df)
cats
```

Out[7]: ['iris']

In [8]:
```
#Gets all numeric features
nums = ds.structdata.get_num_feats(df)
nums
```

Out[8]: ['sepal length', 'sepal width', 'petal length', 'petal width']

In [9]:
```
#Gets the unique count of all categorical features
ds.structdata.get_unique_counts(df)
```

Out[9]:

	Feature	Unique Count
0	iris	3

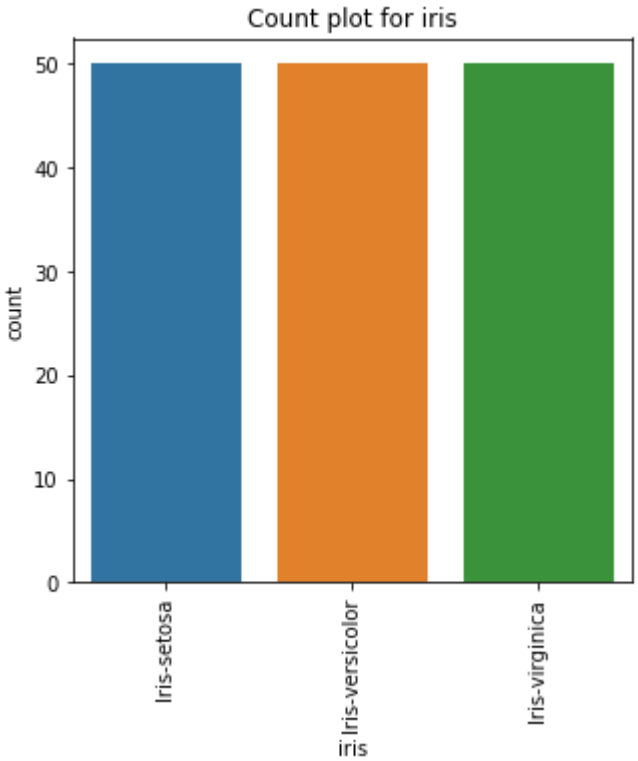
In [10]:
```
#Display missing values in data set
ds.structdata.display_missing(df)
```

Out[10]:

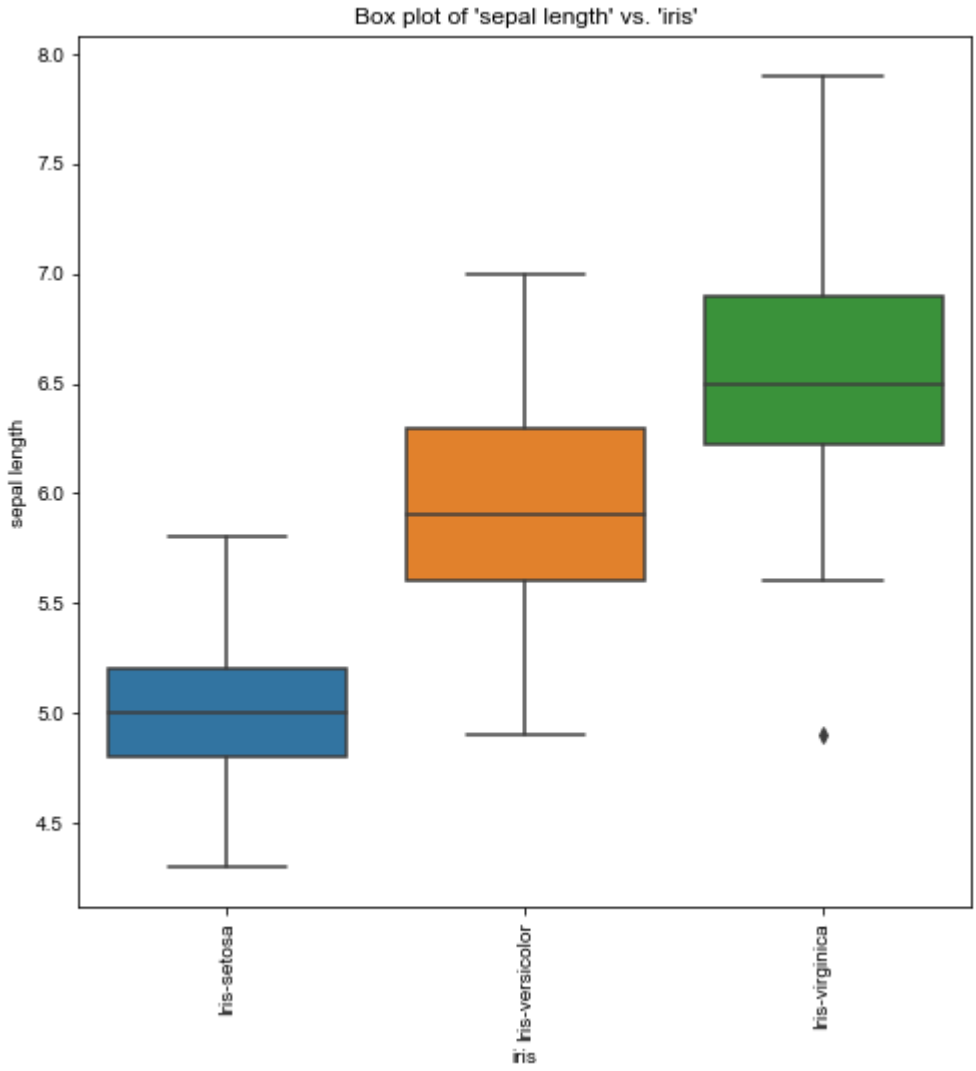
	features	missing_counts	missing_percent
0	sepal length	0	0.0
1	sepal width	0	0.0
2	petal length	0	0.0
3	petal width	0	0.0
4	iris	0	0.0

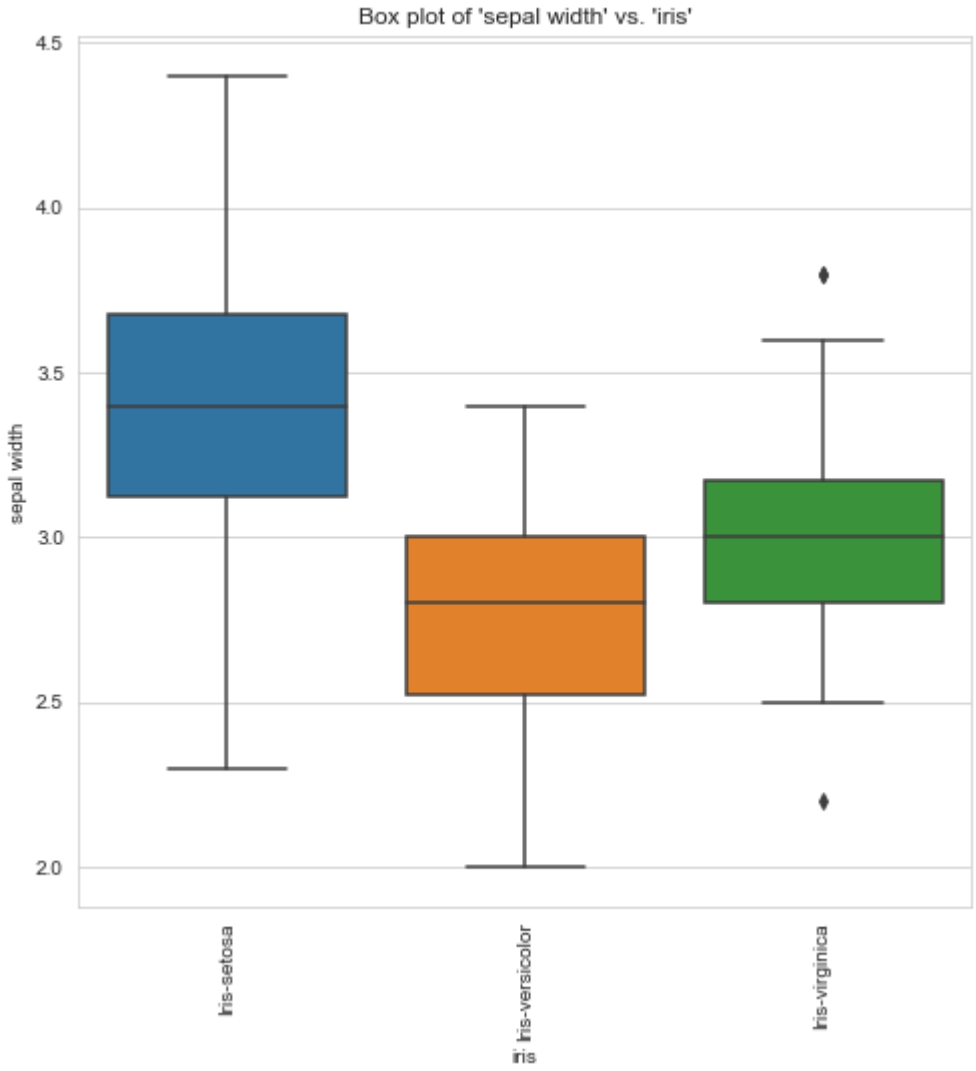
VISUALIZATION

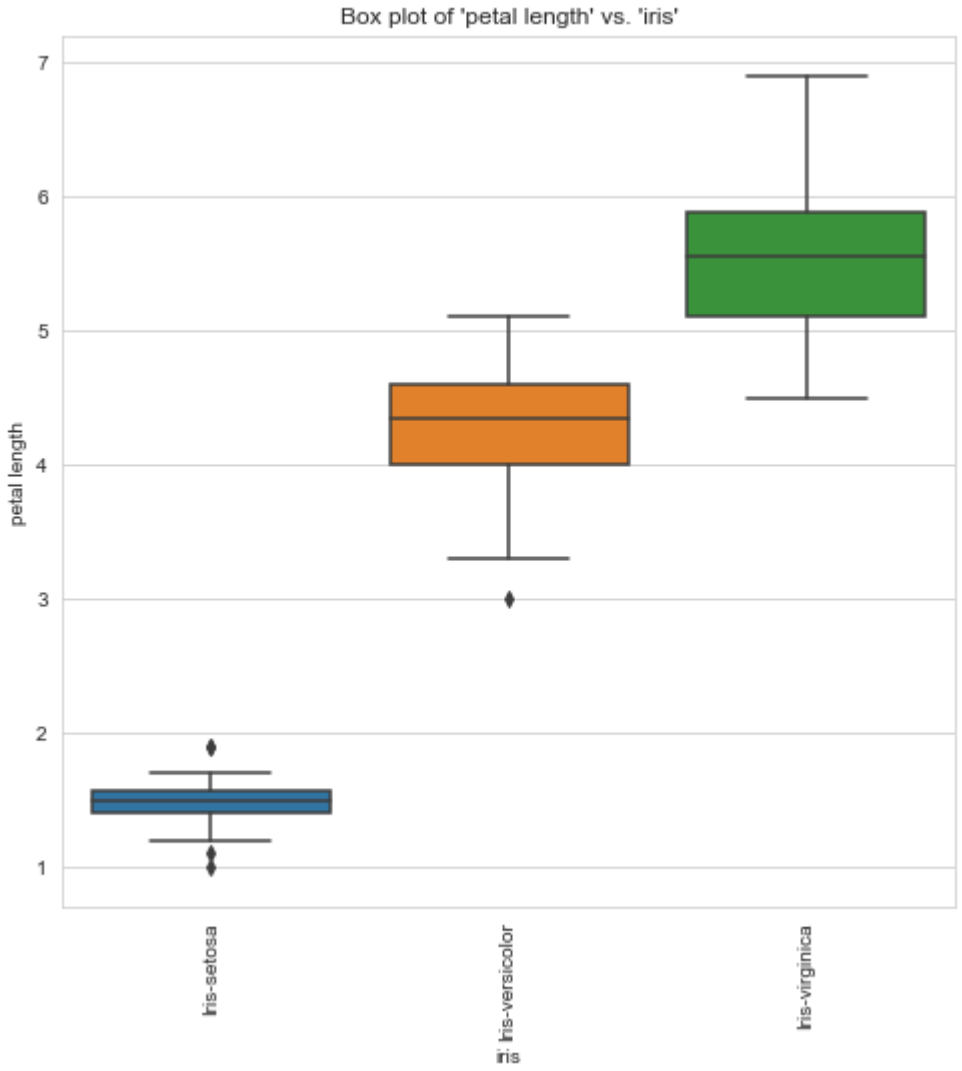
In [11]:
```
ds.visualizations.countplot(df)
```

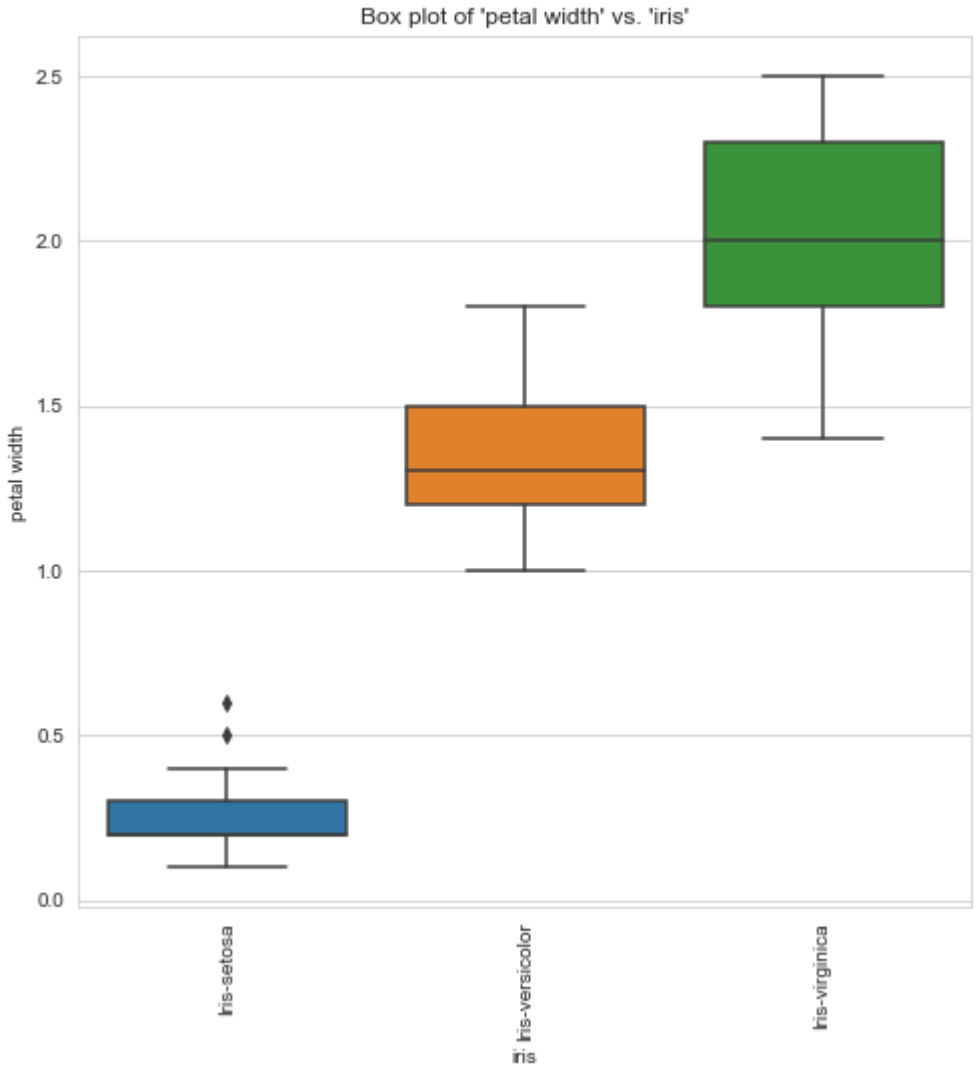


In [13]:
```
ds.visualizations.boxplot(data=df, target='iris')
```

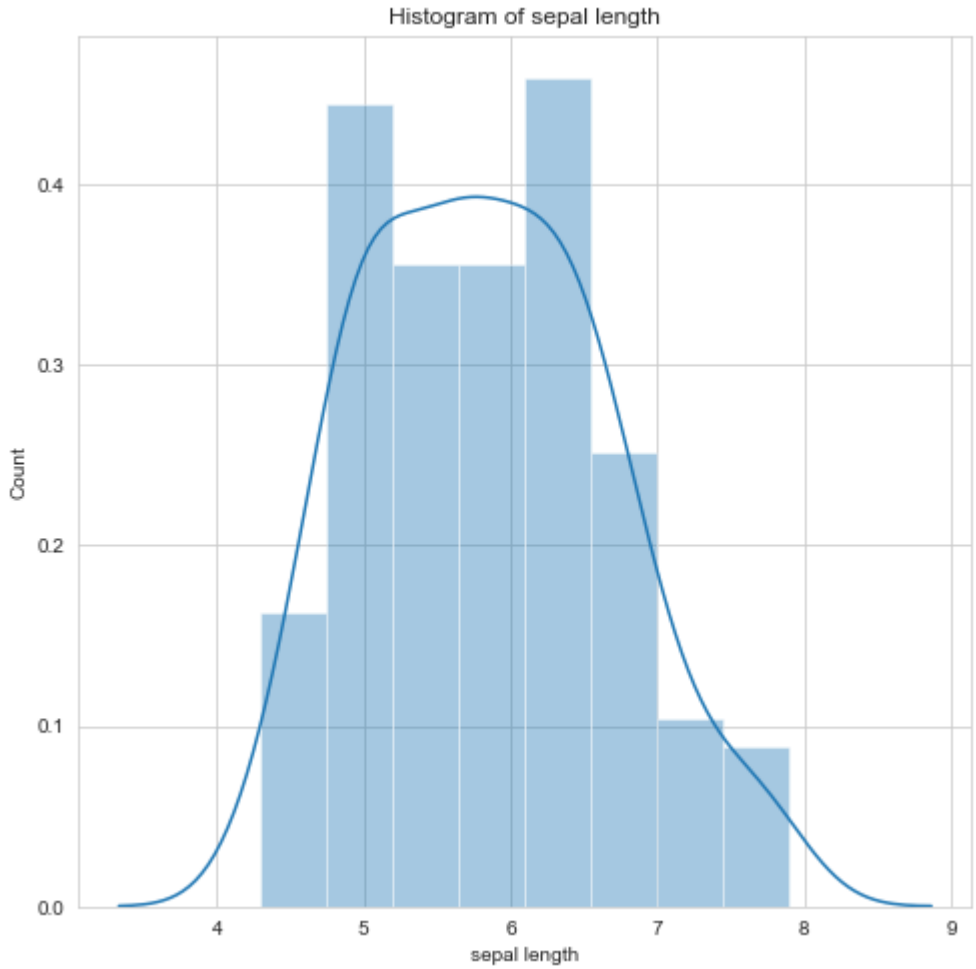

Box plot of 'sepal length' vs. 'iris'
sepal length
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
Iris-setosa
Iris-versicolor
Iris-virginica
iris

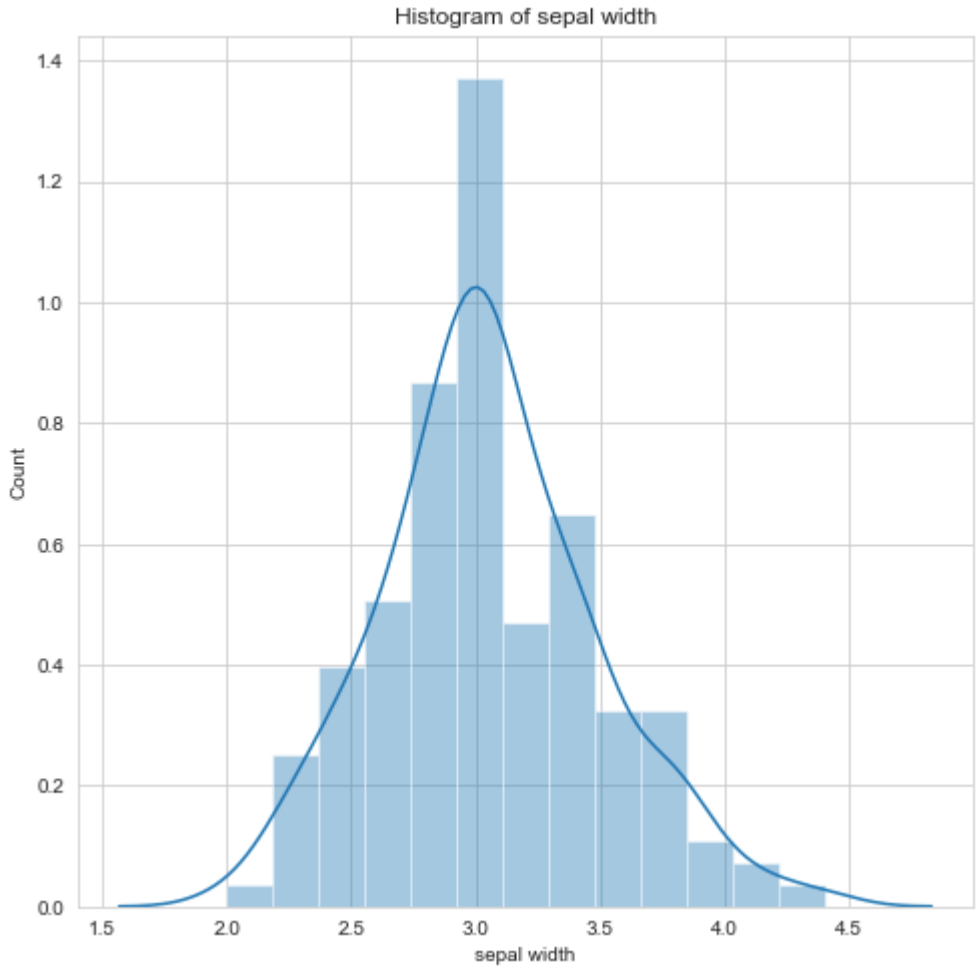

Box plot of 'sepal width' vs. 'iris'
sepal width
4.5
4.0
3.5
3.0
2.5
2.0
Iris-setosa
Iris-versicolor
Iris-virginica
iris

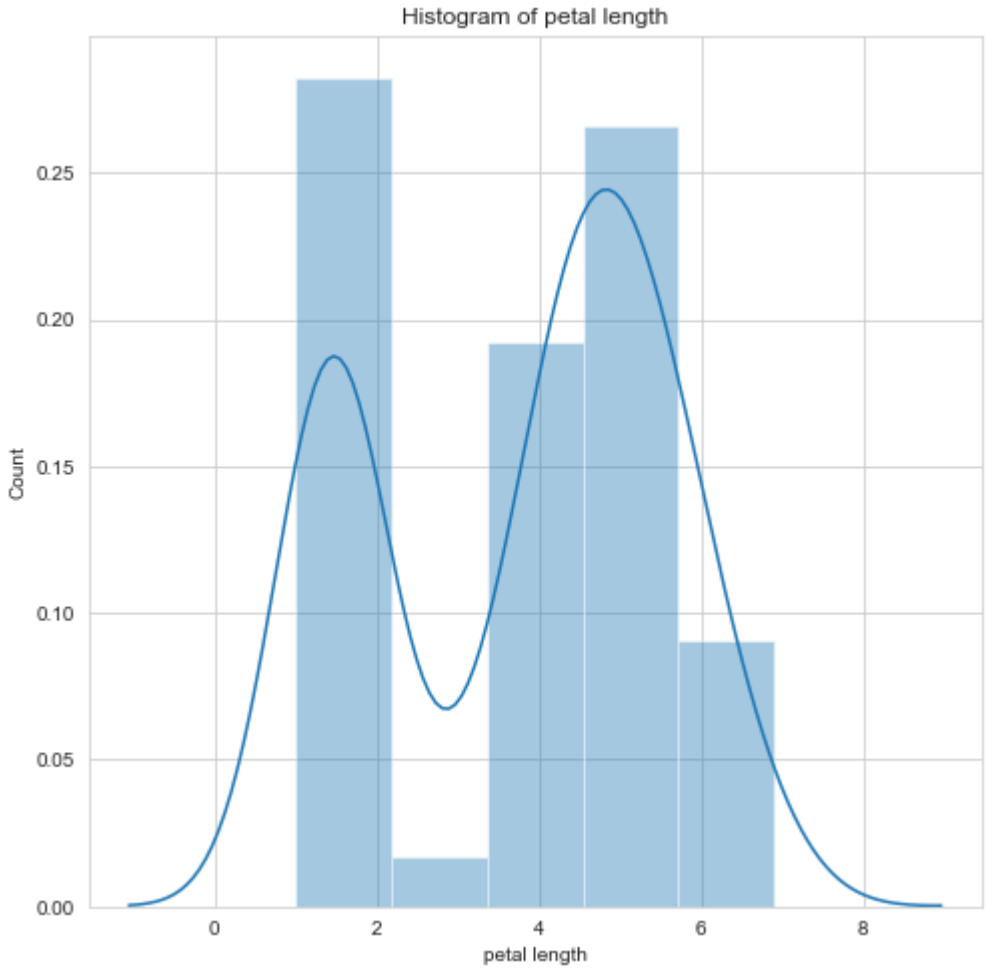

Box plot of 'petal length' vs. 'iris'
petal length
7
6
5
4
3
2
1
Iris-setosa
Iris-versicolor
Iris-virginica
iris

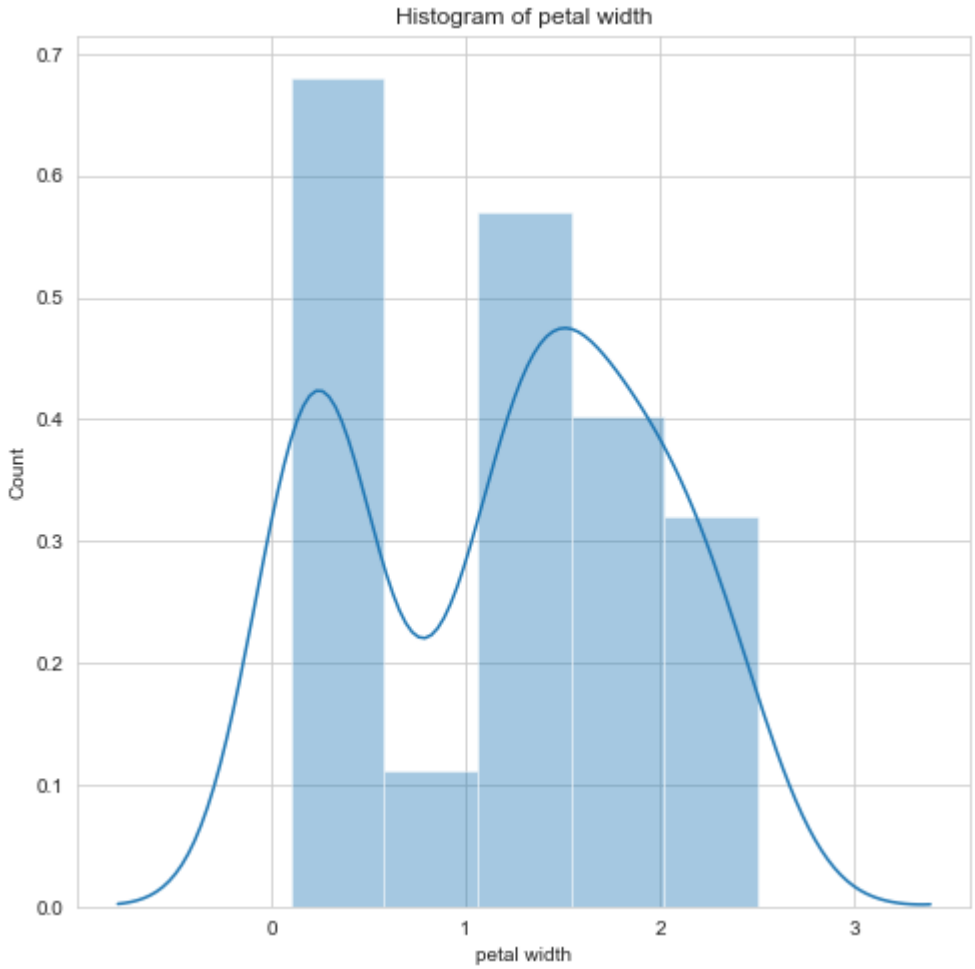

Box plot of 'petal width' vs. 'iris'
petal width
2.5
2.0
1.5
1.0
0.5
0.0
Iris-setosa
Iris-versicolor
Iris-virginica
iris

In [14]:
```
ds.visualizations.histogram(df)
```

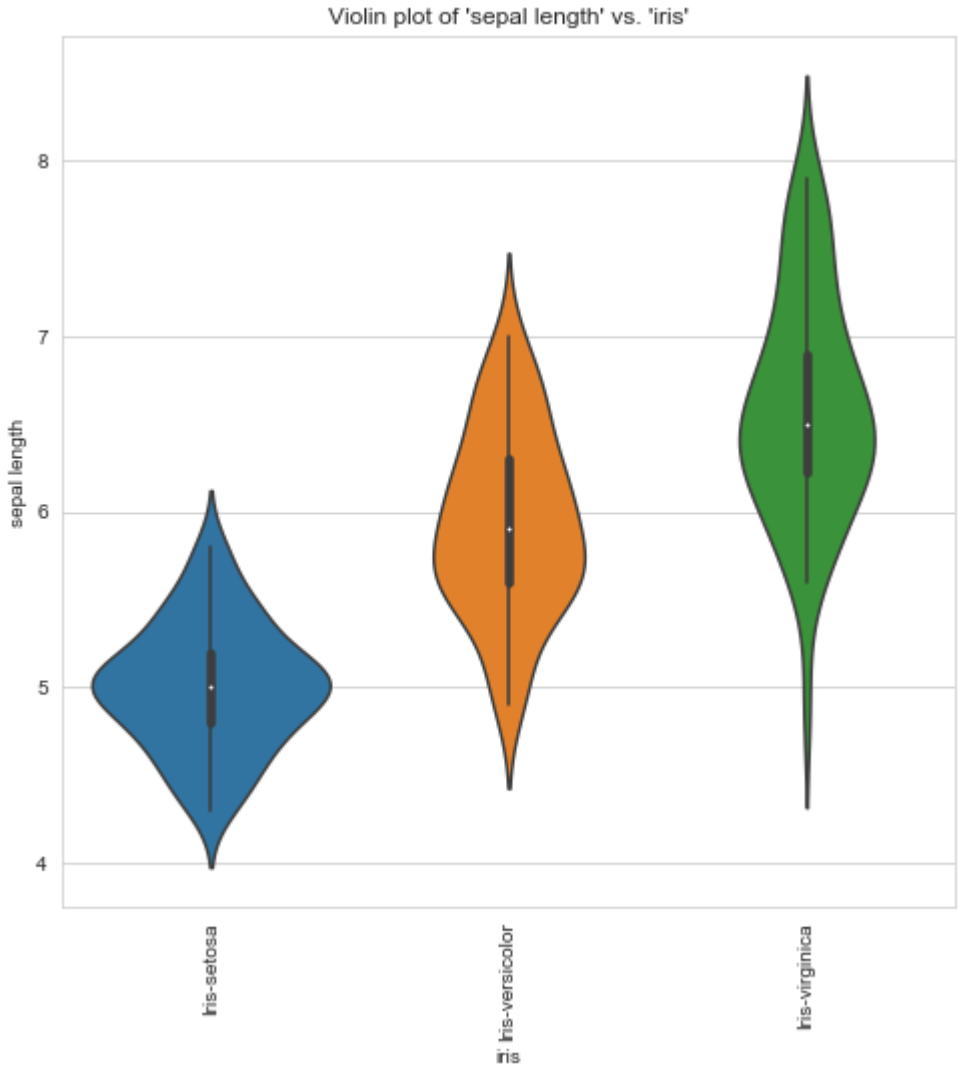

Histogram of sepal length
Count
0.0
0.1
0.2
0.3
0.4
4
5
6
7
8
9
sepal length

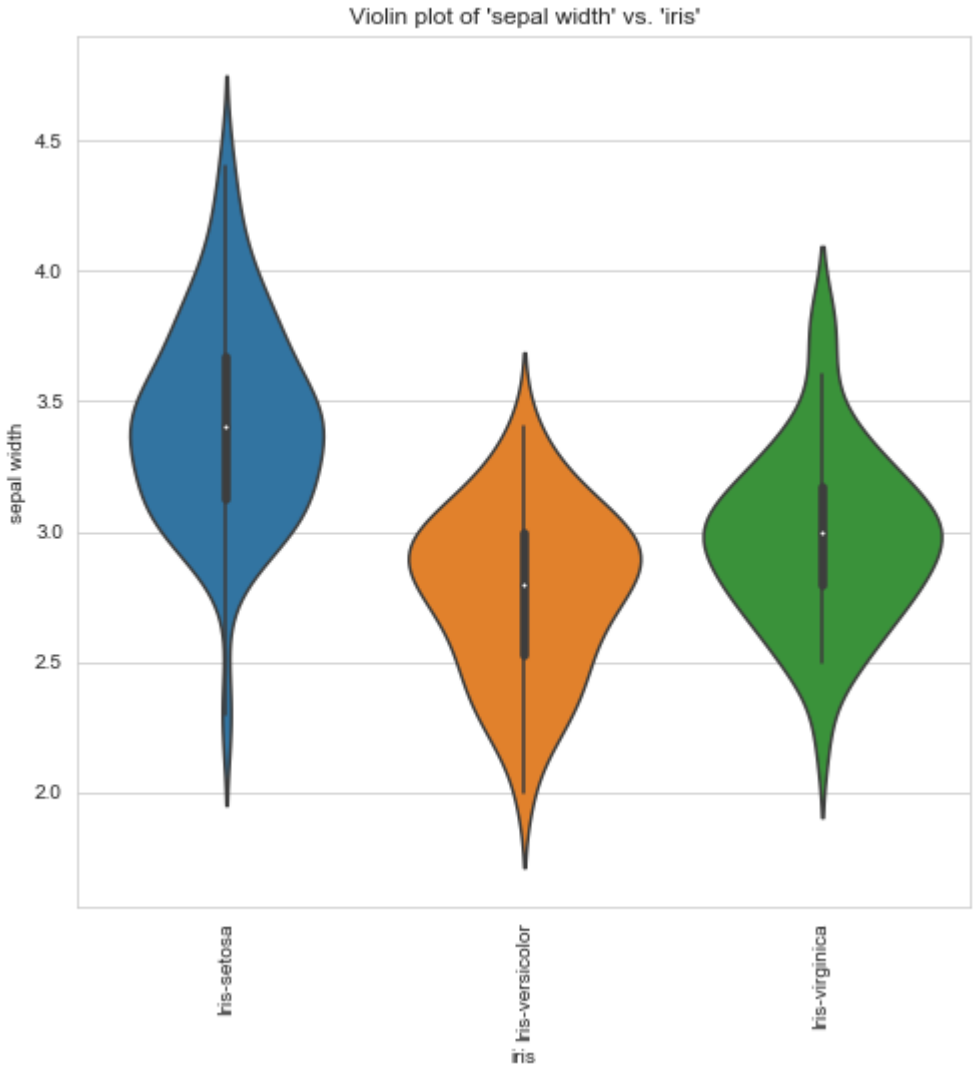

Histogram of sepal width
Count
sepal width
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.5
2.0
2.5
3.0
3.5
4.0
4.5

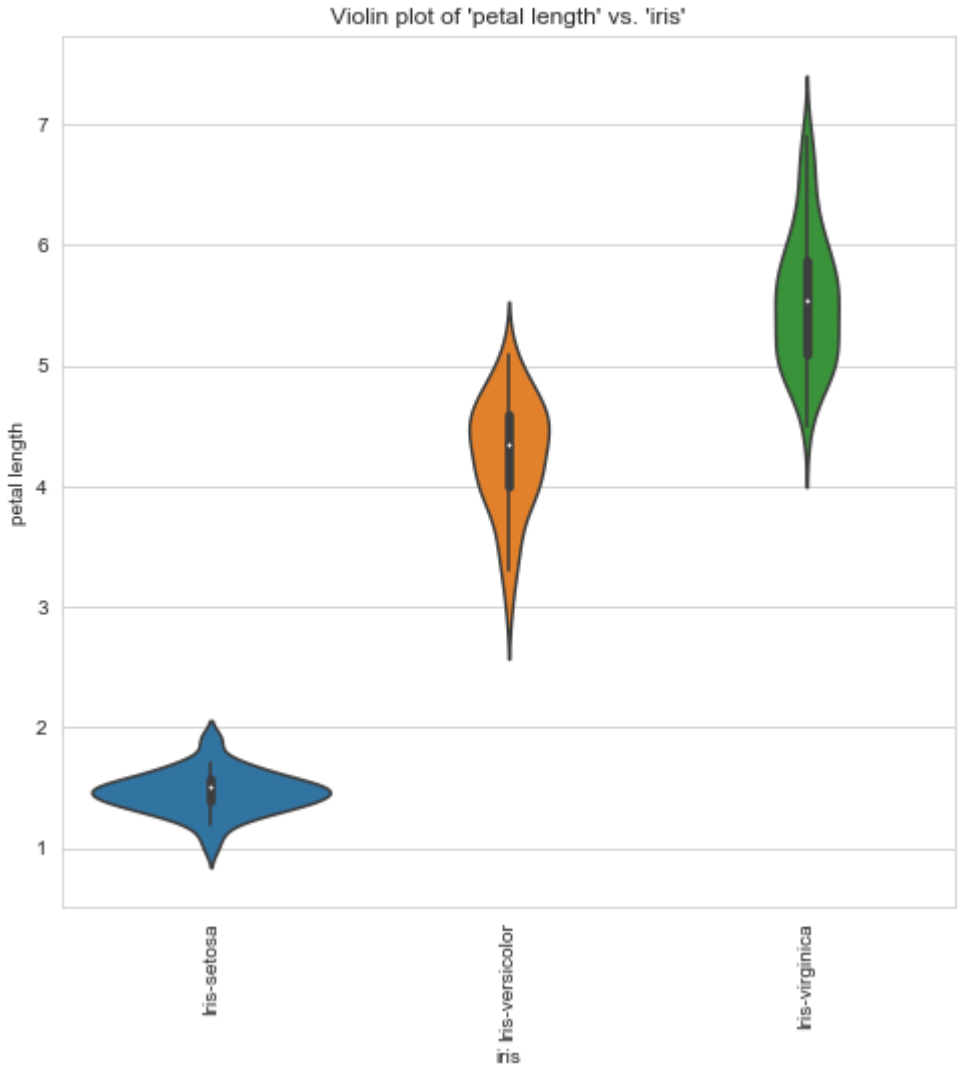

Histogram of petal length
Count
0.25
0.20
0.15
0.10
0.05
0.00
0
2
4
6
8
petal length

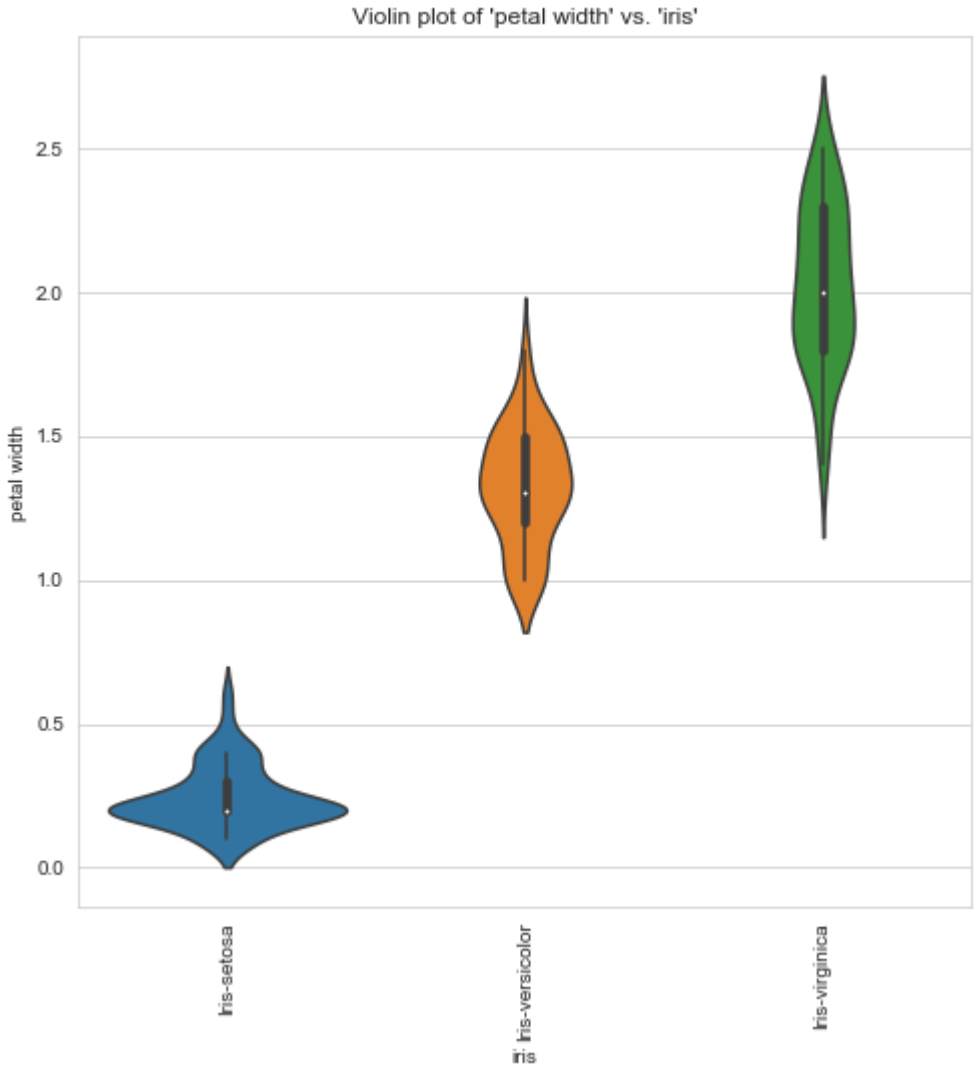

Histogram of petal width
Count
petal width
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0
1
2
3

In [16]:
```
ds.visualizations.violinplot(data=df, target='iris')
```

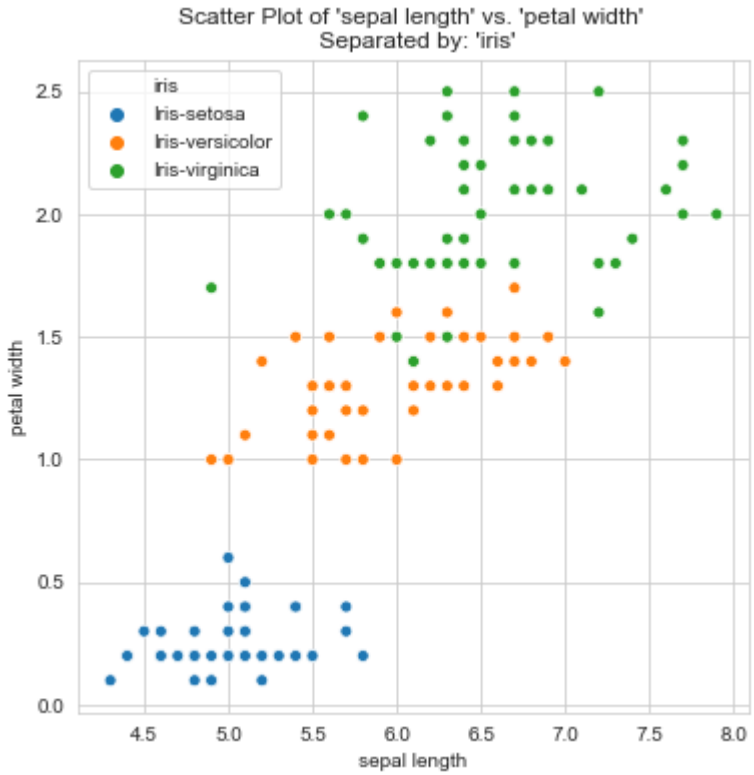

Violin plot of 'sepal length' vs. 'iris'
sepal length
8
7
6
5
4
Iris-setosa
Iris-versicolor
Iris-virginica
iris

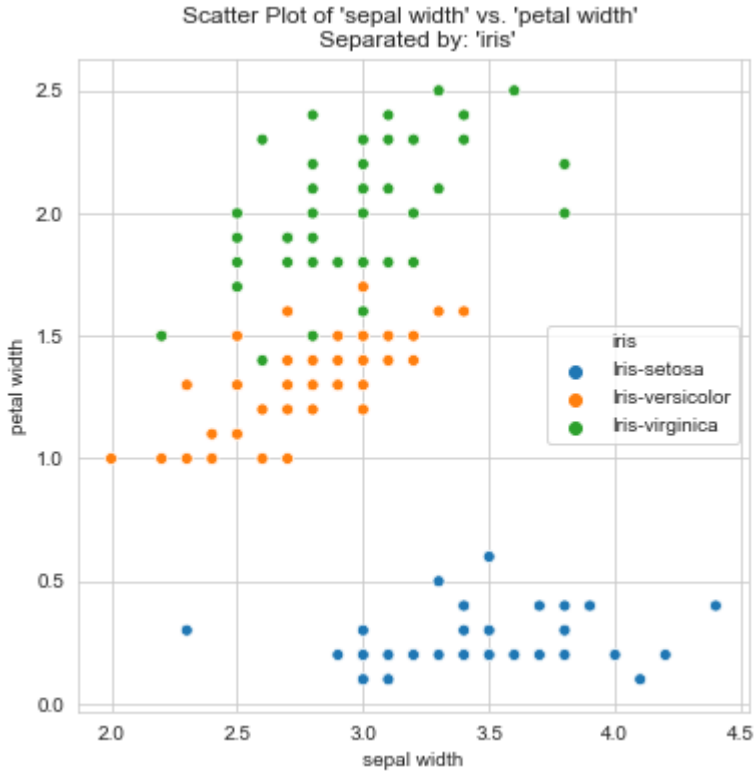

Violin plot of 'sepal width' vs. 'iris'
sepal width
4.5
4.0
3.5
3.0
2.5
2.0
Iris-setosa
Iris-versicolor
Iris-virginica
iris

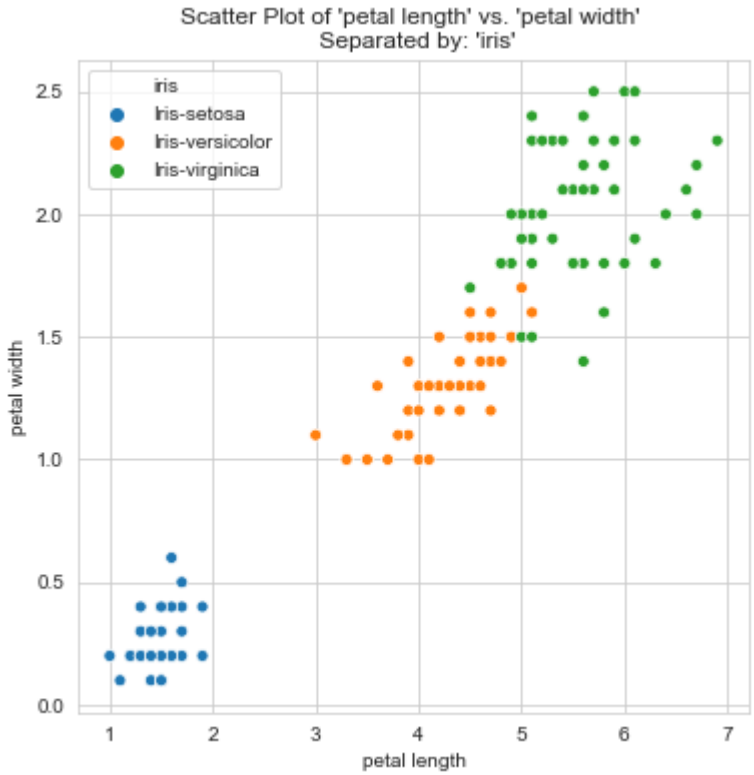

Violin plot of 'petal length' vs. 'iris'
petal length
7
6
5
4
3
2
1
Iris-setosa
Iris-versicolor
Iris-virginica
iris

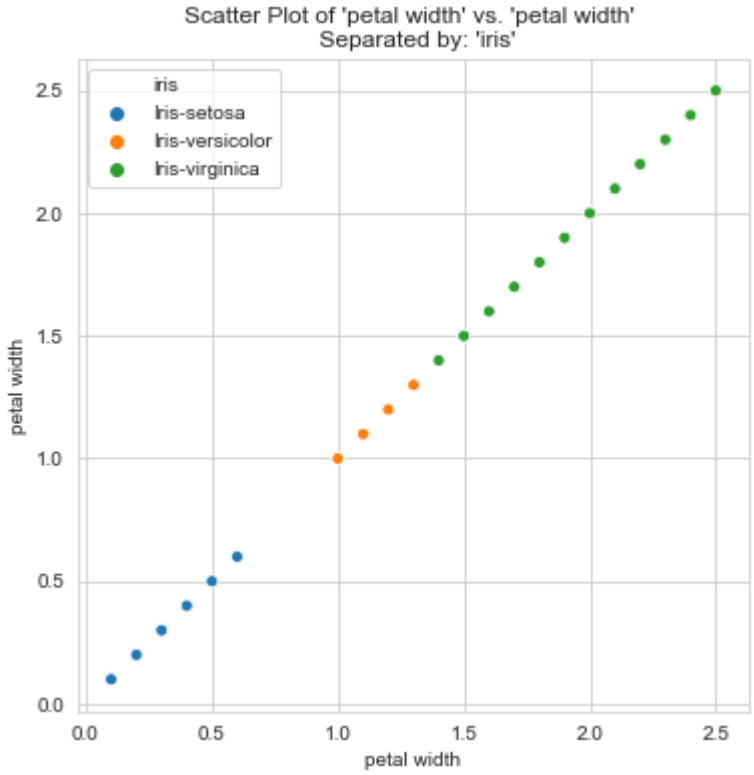

Violin plot of 'petal width' vs. 'iris'
petal width
2.5
2.0
1.5
1.0
0.5
0.0
Iris-setosa
Iris-versicolor
Iris-virginica
iris

In [18]:
```
ds.visualizations.scatterplot(data=df, target='petal width',separate_by='iris'
, fig_size=(6,6))
```


Scatter Plot of 'sepal length' vs. 'petal width'
Separated by: 'iris'
iris
Iris-setosa
Iris-versicolor
Iris-virginica
petal width
sepal length
0.0
0.5
1.0
1.5
2.0
2.5
4.5
5.0
5.5
6.0
6.5
7.0
7.5
8.0


Scatter Plot of 'sepal width' vs. 'petal width'
Separated by: 'iris'
iris
Iris-setosa
Iris-versicolor
Iris-virginica
petal width
sepal width
0.0
0.5
1.0
1.5
2.0
2.5
2.0
2.5
3.0
3.5
4.0
4.5


Scatter Plot of 'petal length' vs. 'petal width'
Separated by: 'iris'
iris
Iris-setosa
Iris-versicolor
Iris-virginica
petal width
petal length
0.0
0.5
1.0
1.5
2.0
2.5
1
2
3
4
5
6
7


Scatter Plot of 'petal width' vs. 'petal width'
Separated by: 'iris'
iris
Iris-setosa
Iris-versicolor
Iris-virginica
petal width
petal width
0.0
0.5
1.0
1.5
2.0
2.5

In [ ]: