Date 06-07-2021

Distplots Exercise

In [1]:
```
import pandas as pd
import numpy as np
import plotly.offline as pyo
import plotly.figure_factory as ff
```

In [2]:
```
iris_data_csv = pd.read_csv("iris.csv", usecols = ["petal_length", "class"])
iris_data_csv
```

Out[2]:

	petal_length	class
0	1.4	Iris-setosa
1	1.4	Iris-setosa
2	1.3	Iris-setosa
3	1.5	Iris-setosa
4	1.4	Iris-setosa
...	...	...
145	5.2	Iris-virginica
146	5.0	Iris-virginica
147	5.2	Iris-virginica
148	5.4	Iris-virginica
149	5.1	Iris-virginica

150 rows × 2 columns

In [3]:
```
iris_data_csv.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 150 entries, 0 to 149
Data columns (total 2 columns):
 #   Column        Non-Null Count  Dtype  
---  ------        --------------  -----  
 0   petal_length  150 non-null    float64
 1   class         150 non-null    object 
dtypes: float64(1), object(1)
memory usage: 2.5+ KB
```

In [4]:
```
iris_data_csv.describe()
```

Out[4]:

	petal_length
count	150.000000
mean	3.758667
std	1.764420
min	1.000000
25%	1.600000
50%	4.350000
75%	5.100000
max	6.900000

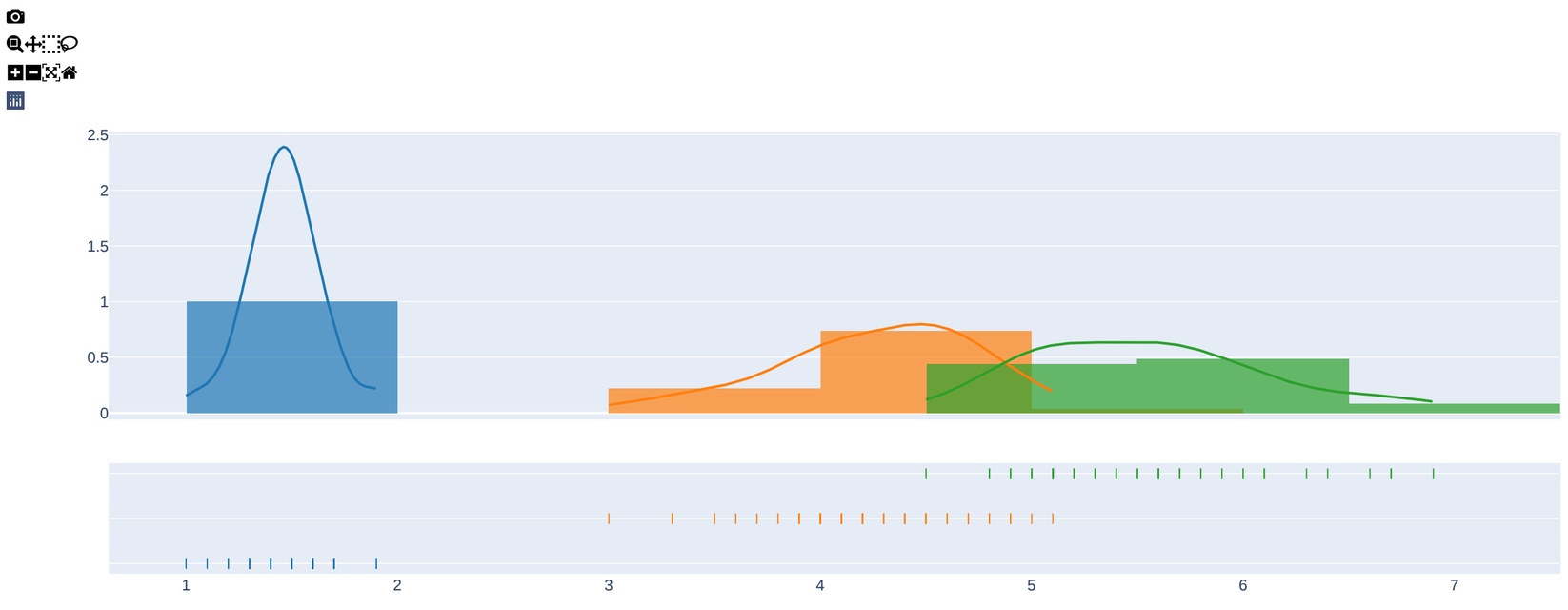
In [5]:
```
data_to_plot_1_Iris_setosa = iris_data_csv[iris_data_csv['class']=='Iris-setosa']['petal_length']
data_to_plot_2_Iris_versicolor = iris_data_csv[iris_data_csv['class']=='Iris-versicolor']['petal_length']
data_to_plot_3_Iris_virginica = iris_data_csv[iris_data_csv['class']=='Iris-virginica']['petal_length']
```

In [6]:
```
hist_data = [data_to_plot_1_Iris_setosa, data_to_plot_2_Iris_versicolor, data_to_plot_3_Iris_virginica]
```

In [7]:
```
group_labels = ["Iris Setosa", "Iris Versicolor", "Iris Virginica"]
```

In [8]:
```
fig = ff.create_distplot(hist_data,
                         group_labels)
```

In [9]:
```
pyo.iplot(fig)
```



In [10]:
```
pyo.plot(fig, filename = "tutorial_20 (Distplots Exercise){Graph}.html")
```

Out[10]:
```
'tutorial_20 (Distplots Exercise){Graph}.html'
```

The Instructor Solution is given below

In [11]:
```
#######
# Objective: Using the iris dataset, develop a Distplot
# that compares the petal lengths of each class.
# File: '../data/iris.csv'
# Fields: 'sepal_length','sepal_width','petal_length','petal_width','class'
# Classes: 'Iris-setosa','Iris-versicolor','Iris-virginica'
######

# Perform imports here:
import plotly.offline as pyo
import plotly.figure_factory as ff
import pandas as pd

# create a DataFrame from the .csv file:
df = pd.read_csv('iris.csv')

# Define the traces
trace0 = df[df['class']=='Iris-setosa']['petal_length']
trace1 = df[df['class']=='Iris-versicolor']['petal_length']
trace2 = df[df['class']=='Iris-virginica']['petal_length']

# Define a data variable
hist_data = [trace0, trace1, trace2]
group_labels = ['Iris Setosa','Iris Versicolor','Iris Virginica']

# Create a fig from data and layout, and plot the fig
fig = ff.create_distplot(hist_data, group_labels)
pyo.iplot(fig)

########
# Great! This shows that if given a flower with a petal length
# between 1-2cm, it is almost certainly an Iris Setosa!
######
```