Colab Link:
https://colab.research.google.com/drive/1SQO41fMKlvXbmVSv4g445kDn6YP2Sp9h?usp=sharing

+ Code + Text

```
!pip install seaborn
```

```
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-w
Requirement already satisfied: seaborn in /usr/local/lib/python3.7/dist-packag
Requirement already satisfied: numpy>=1.15 in /usr/local/lib/python3.7/dist-pa
Requirement already satisfied: pandas>=0.23 in /usr/local/lib/python3.7/dist-p
Requirement already satisfied: scipy>=1.0 in /usr/local/lib/python3.7/dist-pac
Requirement already satisfied: matplotlib>=2.2 in /usr/local/lib/python3.7/dis
Requirement already satisfied: python-dateutil>=2.1 in /usr/local/lib/python3.
Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.7/dist-p
Requirement already satisfied: pyparsing!=2.0.4,!=2.1.2,!=2.1.6,>=2.0.1 in /us
Requirement already satisfied: kiwisolver>=1.0.1 in /usr/local/lib/python3.7/d
Requirement already satisfied: typing-extensions in /usr/local/lib/python3.7/d
Requirement already satisfied: pytz>=2017.3 in /usr/local/lib/python3.7/dist-p
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.7/dist-packa
```

```
import seaborn as sns
```

```
import matplotlib.pyplot as plt
import numpy as np
```

```
iris = sns.load_dataset("iris")
```

```
tips = sns.load_dataset("tips")
```

```
iris
```

Saving...

	sepal_length	sepal_width	petal_length	petal_width	species
0	5.1	3.5	1.4	0.2	setosa

```
type(iris)
```

```
pandas.core.frame.DataFrame
```

3	4.6	3.1	1.5	0.2	setosa

```
iris.head()
```

	sepal_length	sepal_width	petal_length	petal_width	species
0	5.1	3.5	1.4	0.2	setosa
1	4.9	3.0	1.4	0.2	setosa
2	4.7	3.2	1.3	0.2	setosa
3	4.6	3.1	1.5	0.2	setosa
4	5.0	3.6	1.4	0.2	setosa

```
iris.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
```

```
iris.describe()
```

	sepal_length	sepal_width	petal_length	petal_width
count	150.000000	150.000000	150.000000	150.000000
mean	5.843333	3.057333	3.758000	1.199333
std	0.828066	0.435866	1.765298	0.762238
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

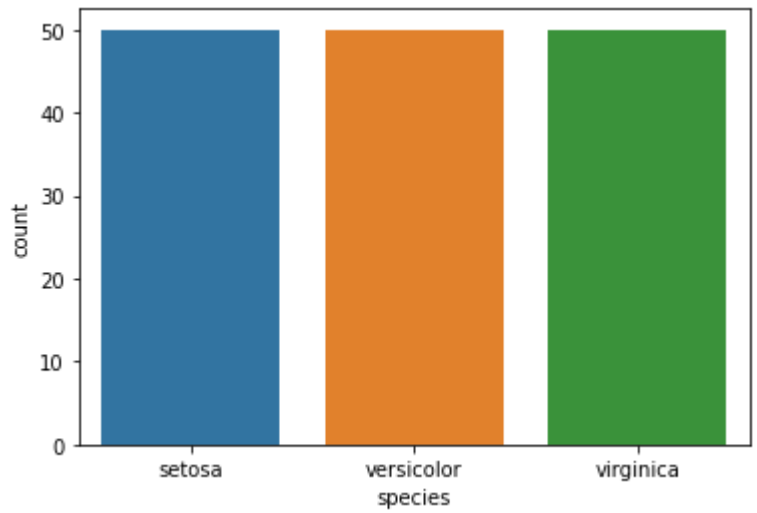
```
iris["species"].unique()
```

```
array(['setosa', 'versicolor', 'virginica'], dtype=object)
```

```
iris["species"].value_counts()
```

```
setosa        50
versicolor    50
virginica     50
Name: species, dtype: int64
```

```
sns.countplot(x="species", data=iris)
plt.show()
```



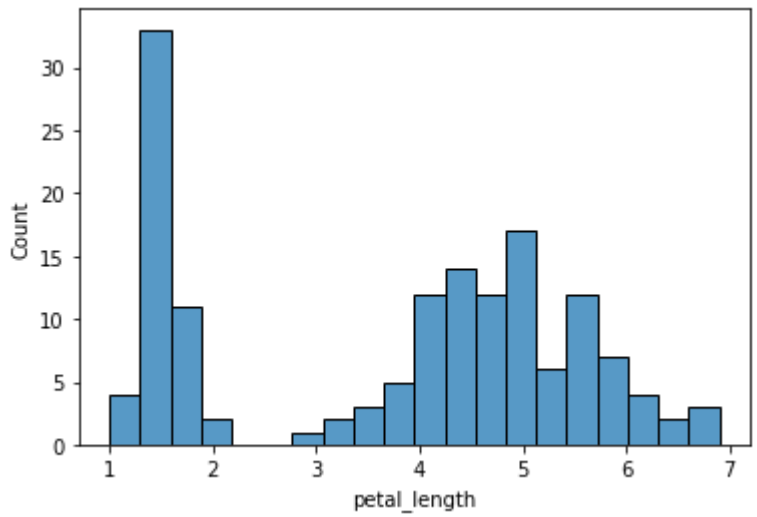

```
n_setosa = iris[iris['species'] == 'setosa'].shape[0]
n_versicolor = iris[iris['species'] == 'versicolor'].shape[0]
n_virginica = iris[iris['species'] == 'virginica'].shape[0]

data = [n_setosa, n_versicolor, n_virginica]
labels=['setosa','versicolor','virginica']

plt.pie(data,
        labels=labels,
        autopct='%.0f%%') # To show the portions in %ages

plt.show()
```

```
# 1V - Numerical
```


```
iris["petal_length"]
```

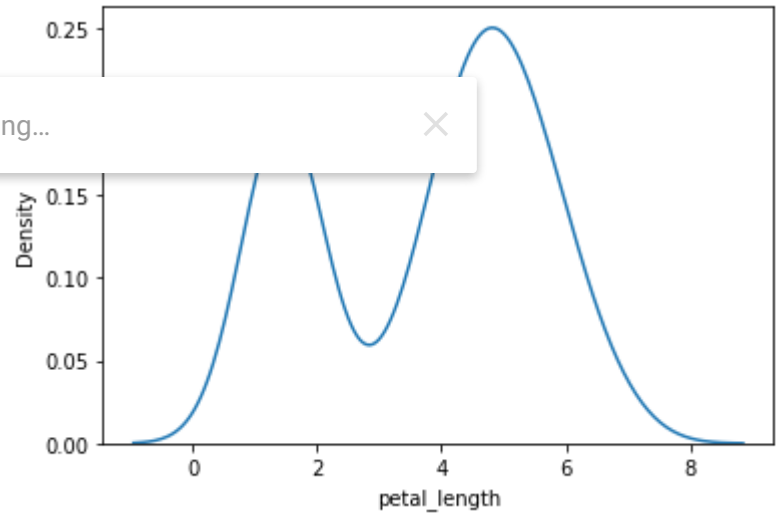
```
0      1.4
1      1.4
2      1.3
3      1.5
4      1.4
      ...
145    5.2
146    5.0
147    5.2
148    5.4
149    5.1
Name: petal_length, Length: 150, dtype: float64
```

```
sns.histplot(data=iris["petal_length"], bins=20)
plt.show()
```



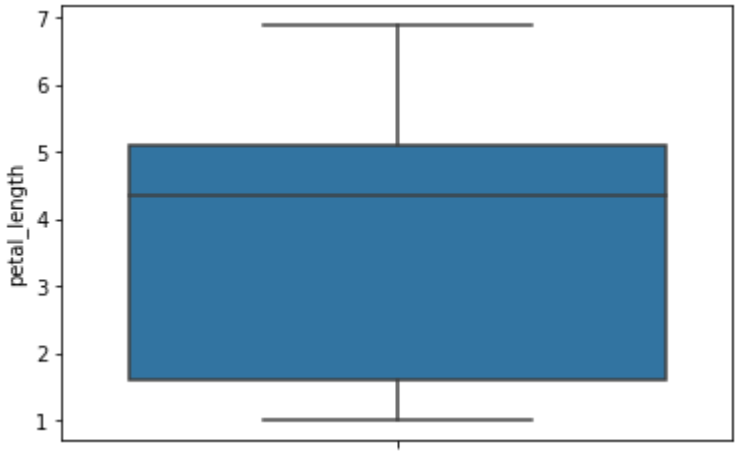

```
sns.kdeplot(iris["petal_length"])
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f9f1298d210>
```



```
sns.boxplot(y=iris["petal_length"])
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f9f128c9890>
```



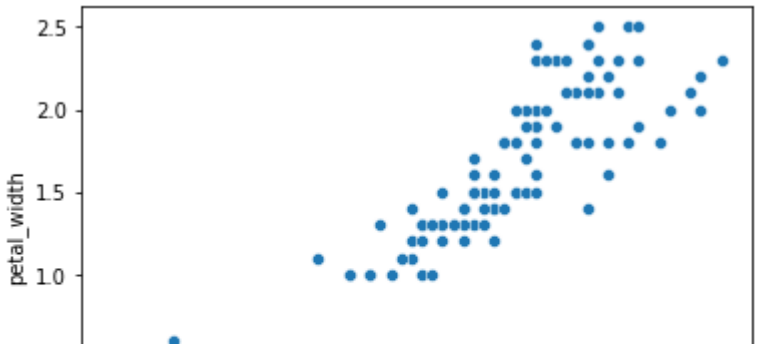
Summarizing Univariate analysis

- Case 1: Numerical variable
  - Type of plots
    - Histogram
    - KDE
    - Box plot
- Case 2: Categorical variable
  - Type of plots
    - Barchart/Count plot
    - Pie chart

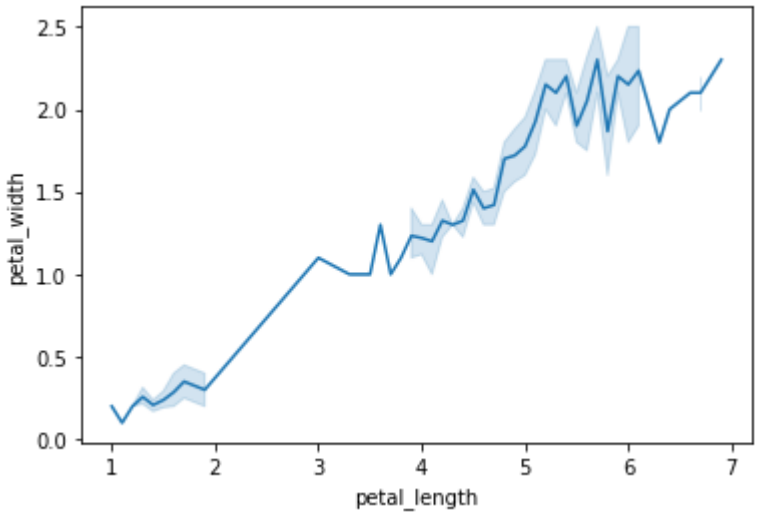
Bivariate Plotting

1. NN
2. NC
3. CC


```
sns.scatterplot(x=iris["petal_length"], y=iris["petal_width"])
plt.show()
#sns.scatterplot(x="petal_length", y="petal_width", data=iris)
```

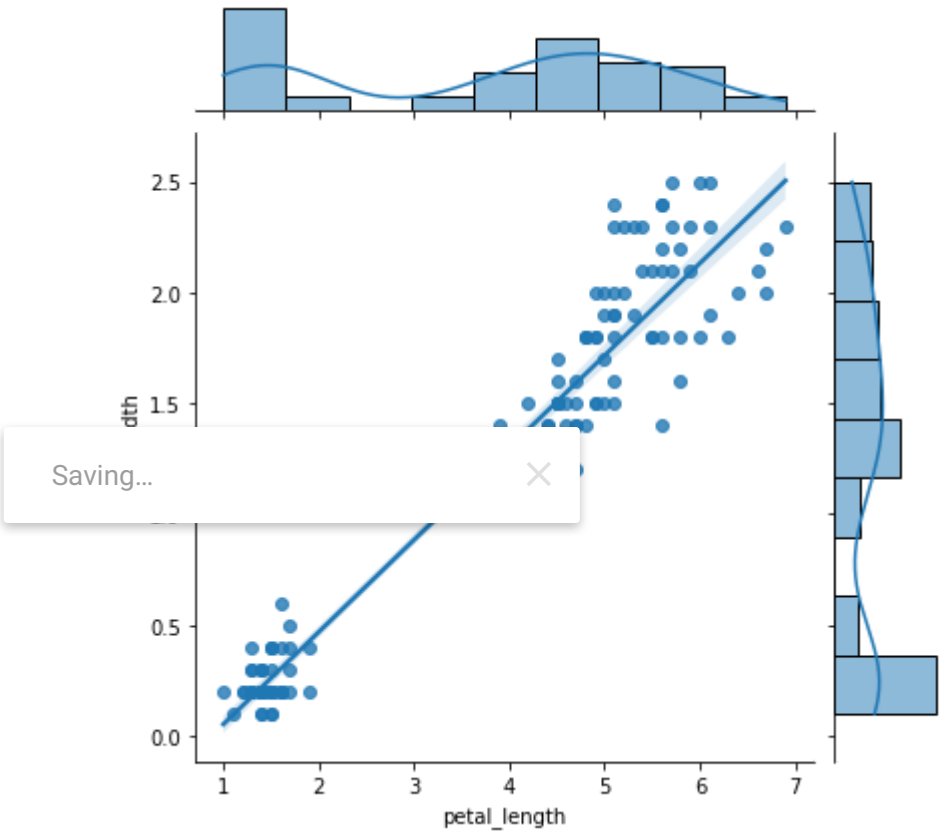



```
sns.lineplot(x=iris["petal_length"], y=iris["petal_width"])
plt.show()
```



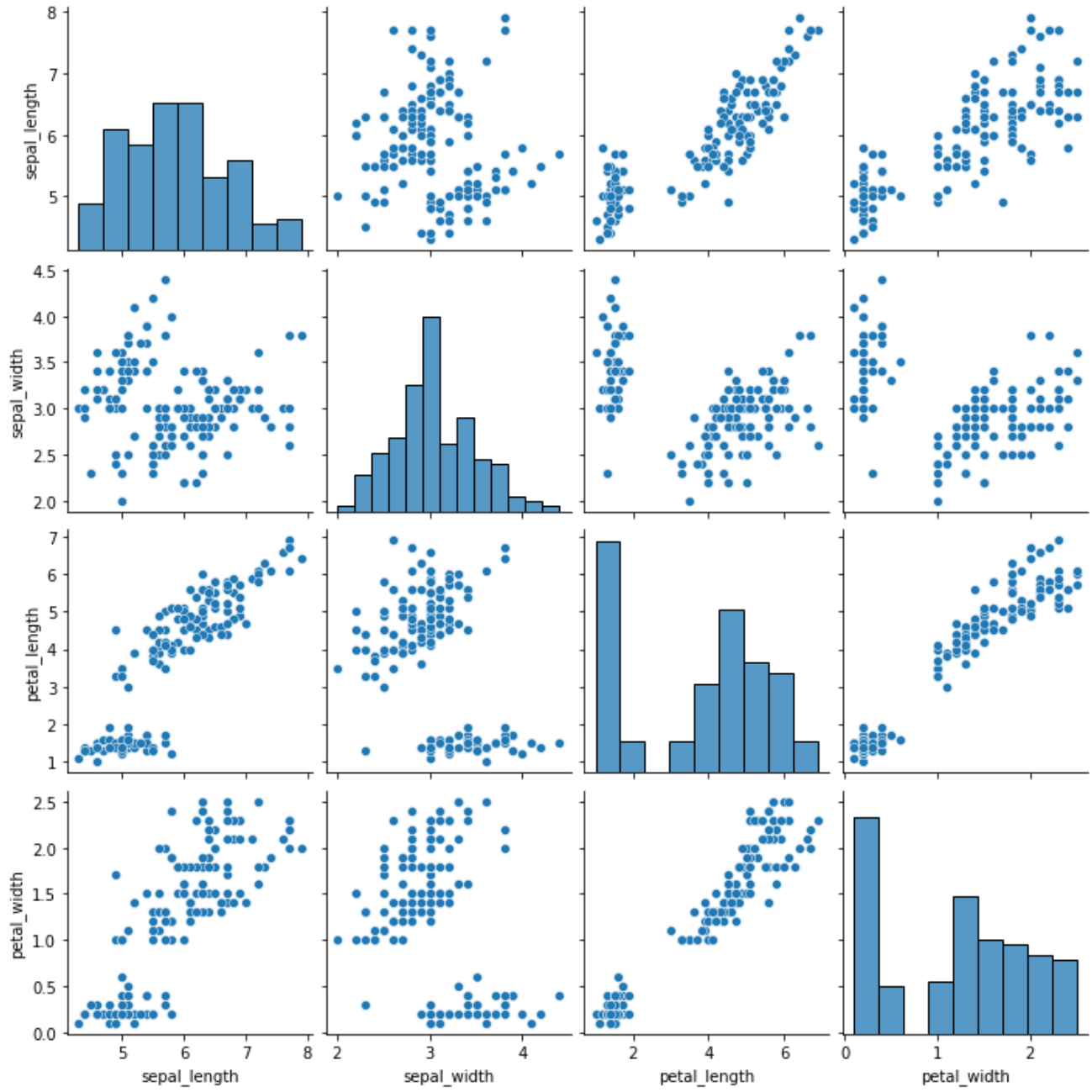

```
sns.jointplot(x="petal_length", y="petal_width", data=iris, kind="reg")
```

```
<seaborn.axisgrid.JointGrid at 0x7f9f0fd21c90>
```



```
sns.pairplot(data=iris)
```

```
<seaborn.axisgrid.PairGrid at 0x7f9f129bcf90>
```

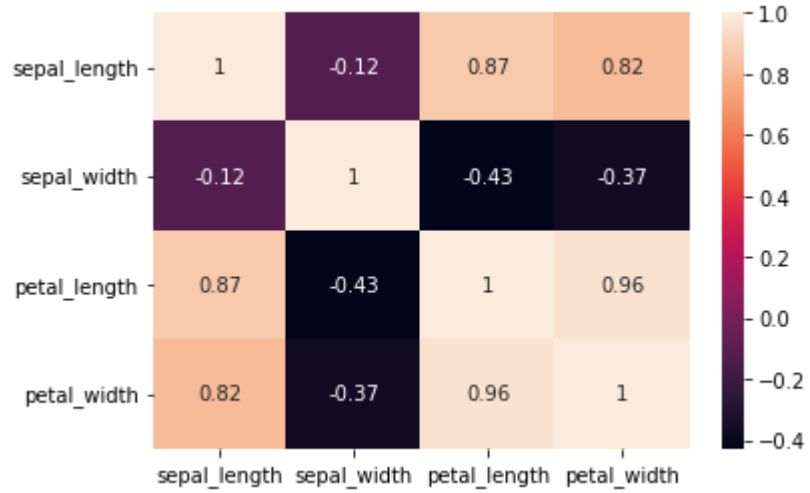



```
iris.corr()
```


		epal_width	petal_length	petal_width
sepal_length	1.000000	-0.117570	0.871754	0.817941
sepal_width	-0.117570	1.000000	-0.428440	-0.366126
petal_length	0.871754	-0.428440	1.000000	0.962865
petal_width	0.817941	-0.366126	0.962865	1.000000

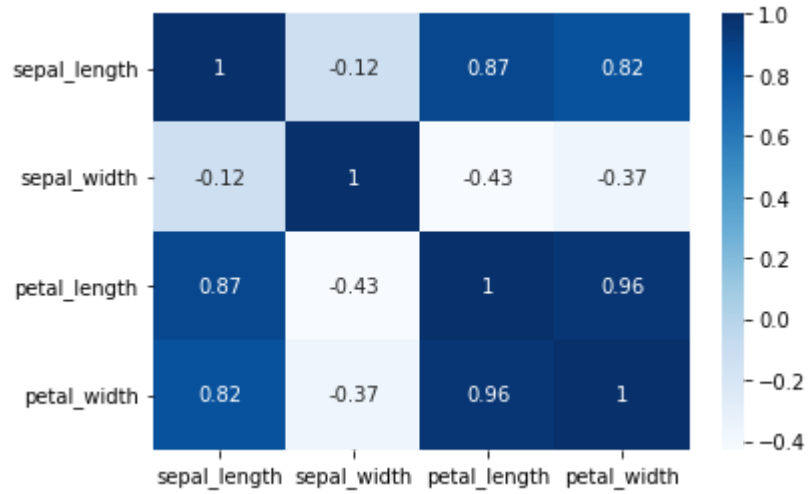
```
sns.heatmap(data=iris.corr(), annot=True)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f9f126e8e90>
```




```
sns.heatmap(data=iris.corr(), annot=True, cmap="Blues")
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f9f0dabe6d0>
```

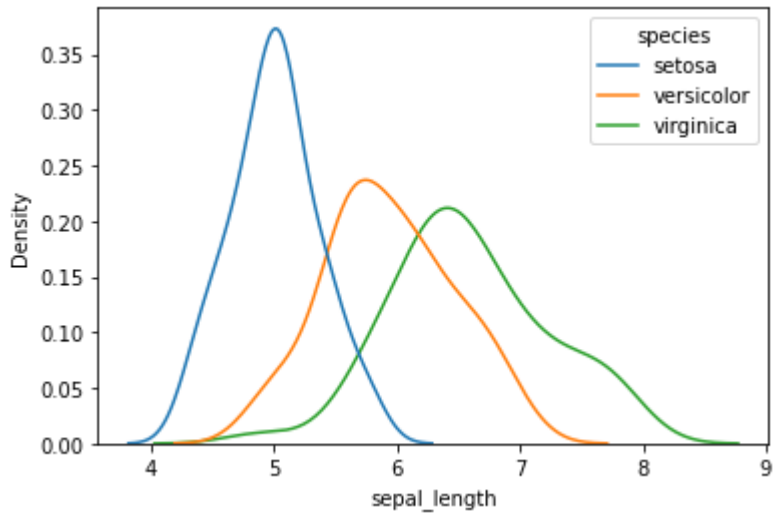



```
# case1: NN ---> Scatter Plot, Line Plot, Joint Plot
# Multiple NN pairs ---> Pairplot, corr --> heatmap


# case2: NC


sns.boxplot(x="species", y="sepal_length", data=iris)
plt.show()
```

Saving...

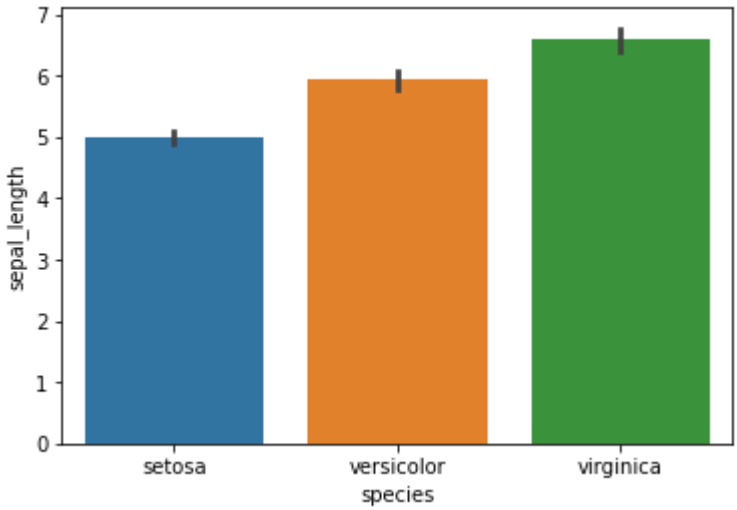

```
sns.kdeplot(x="sepal_length", data=iris, hue="species")
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f9f0d92dfd0>
```

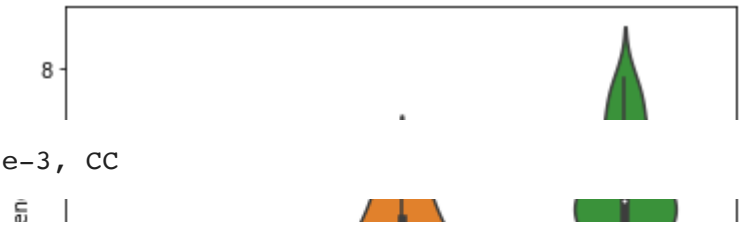



```
sns.barplot(x="species", y="sepal_length", data=iris, estimator=np.mean)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f9f0d5886d0>
```




```
sns.violinplot(x="species", y="sepal_length", data=iris)
plt.show()
```




```
# case-3, CC
```

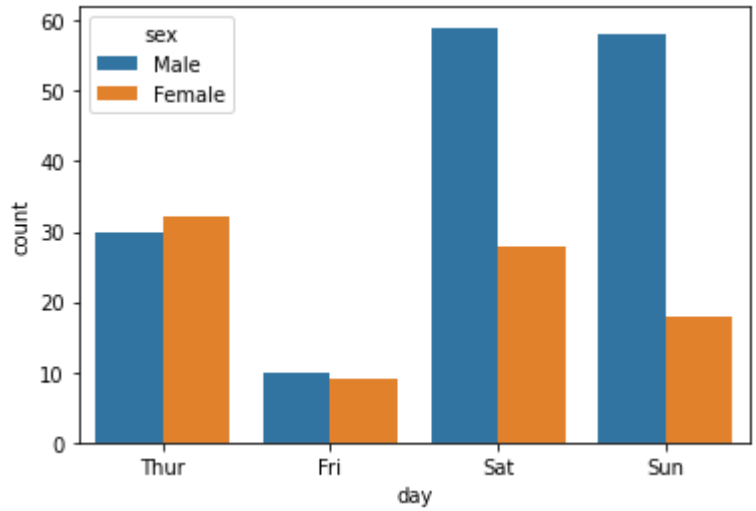


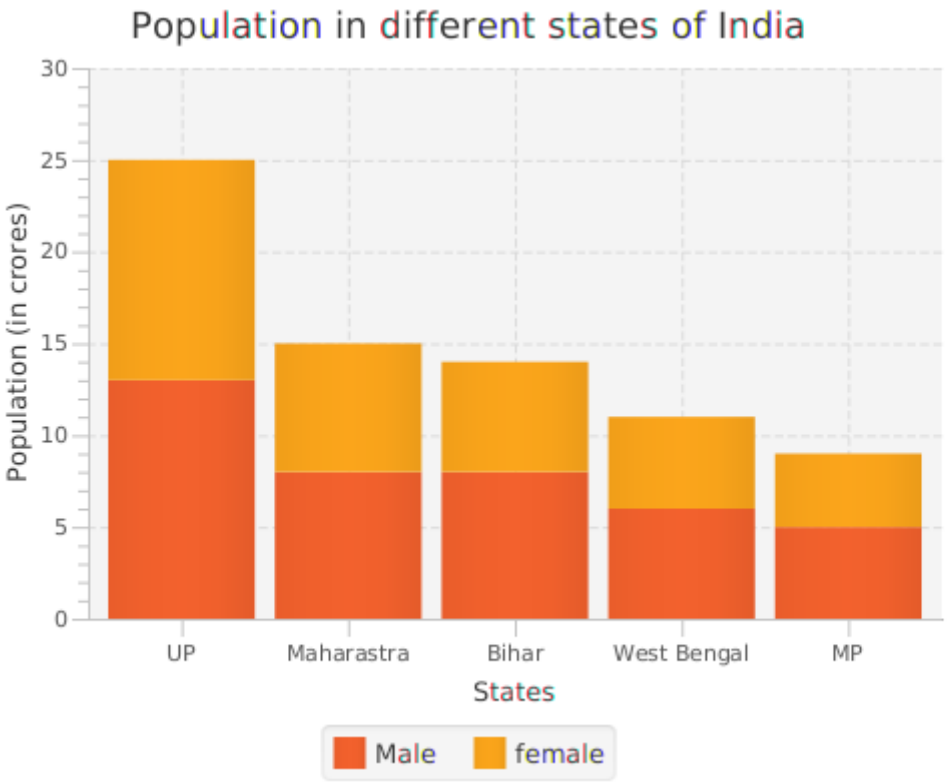
```
tips.head()
```

	total_bill	tip	sex	smoker	day	time	size
0	16.99	1.01	Female	No	Sun	Dinner	2
1	10.34	1.66	Male	No	Sun	Dinner	3
2	21.01	3.50	Male	No	Sun	Dinner	3
3	23.68	3.31	Male	No	Sun	Dinner	2
4	24.59	3.61	Female	No	Sun	Dinner	4

```
sns.countplot(data=tips, x="day", hue="sex", dodge=True)
```

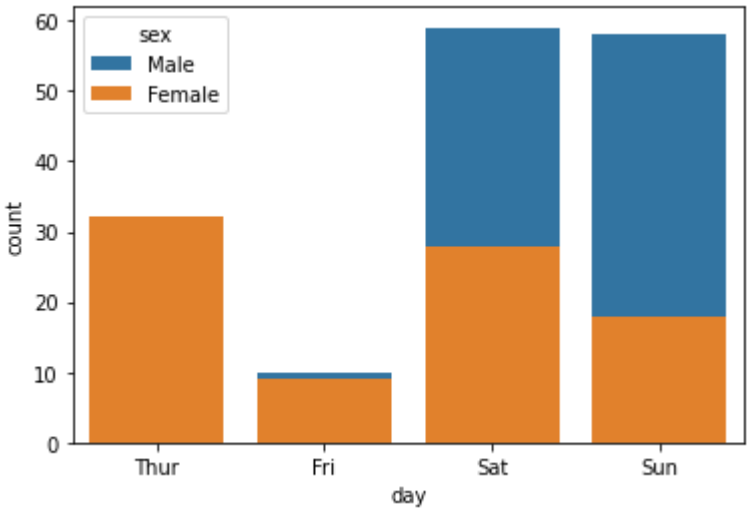
```
<matplotlib.axes._subplots.AxesSubplot at 0x7f9f0d2f9510>
```






```
sns.countplot(data=tips, x="day", hue="sex", dodge=False)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f9f0d97f890>
```




```
df_plot = tips.groupby(['smoker', 'time']).size().reset_index().pivot(columns='smok
                                                              True)
```

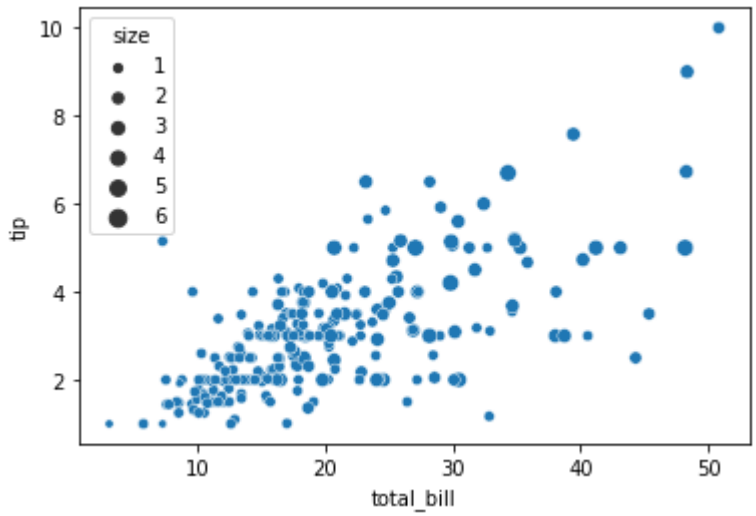
smoker	Yes	No
time		
Lunch	23	45
Dinner	70	106

Multivariate

1. NNN
2. NNC
3. CCN
4. CCC

```
sns.scatterplot(x="total_bill", y="tip", size="size", data=tips)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f9f0d0b94d0>
```



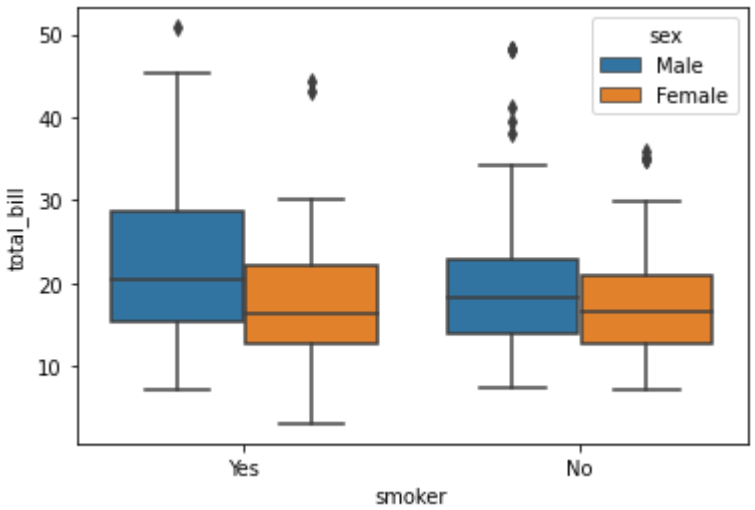

```
y="petal_width", hue="species", data=iris)
```

Saving...

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f9f0d00ee10>
```

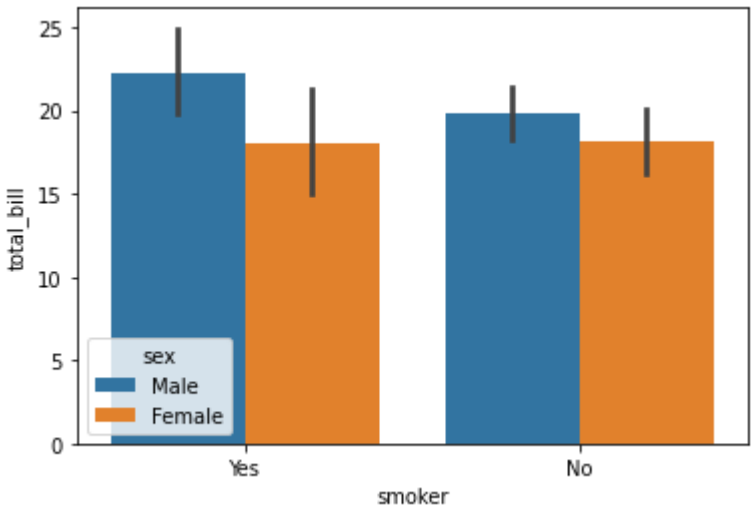

```
sns.boxplot(x="smoker", hue="sex", data=tips, y="total_bill", dodge=True)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f9f0cf91290>
```




```
sns.barplot(x="smoker", hue="sex", data=tips, y="total_bill", dodge=True, estimator
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f9f0cecc210>
```



Saving...

✓ 0s completed at 23:31

Saving...