Seaborn: Statistical Data Visualization

by Mrinal Kanti Roy (Hellol)

In [132…
```
import seaborn as sns
import matplotlib.pyplot as plt
```

In [133…
```
sns.get_dataset_names()
```

Out[133…
```
['anagrams',
 'anscombe',
 'attention',
 'brain_networks',
 'car_crashes',
 'diamonds',
 'dots',
 'dowjones',
 'exercise',
 'flights',
 'fmri',
 'geyser',
 'glue',
 'healthexp',
 'iris',
 'mpg',
 'penguins',
 'planets',
 'seaice',
 'taxis',
 'tips',
 'titanic']
```

Loading the datasets

In [134…
```
tips = sns.load_dataset("tips")
iris = sns.load_dataset("iris")
titanic = sns.load_dataset("titanic")
planet = sns.load_dataset("planets")
```

Checking the Tips dataset

In [135…
```
tips
```

Out[135…

	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
...	...	...	...	...	...	...	...
239	29.03	5.92	Male	No	Sat	Dinner	3
240	27.18	2.00	Female	Yes	Sat	Dinner	2
241	22.67	2.00	Male	Yes	Sat	Dinner	2
242	17.82	1.75	Male	No	Sat	Dinner	2
243	18.78	3.00	Female	No	Thur	Dinner	2

244 rows × 7 columns

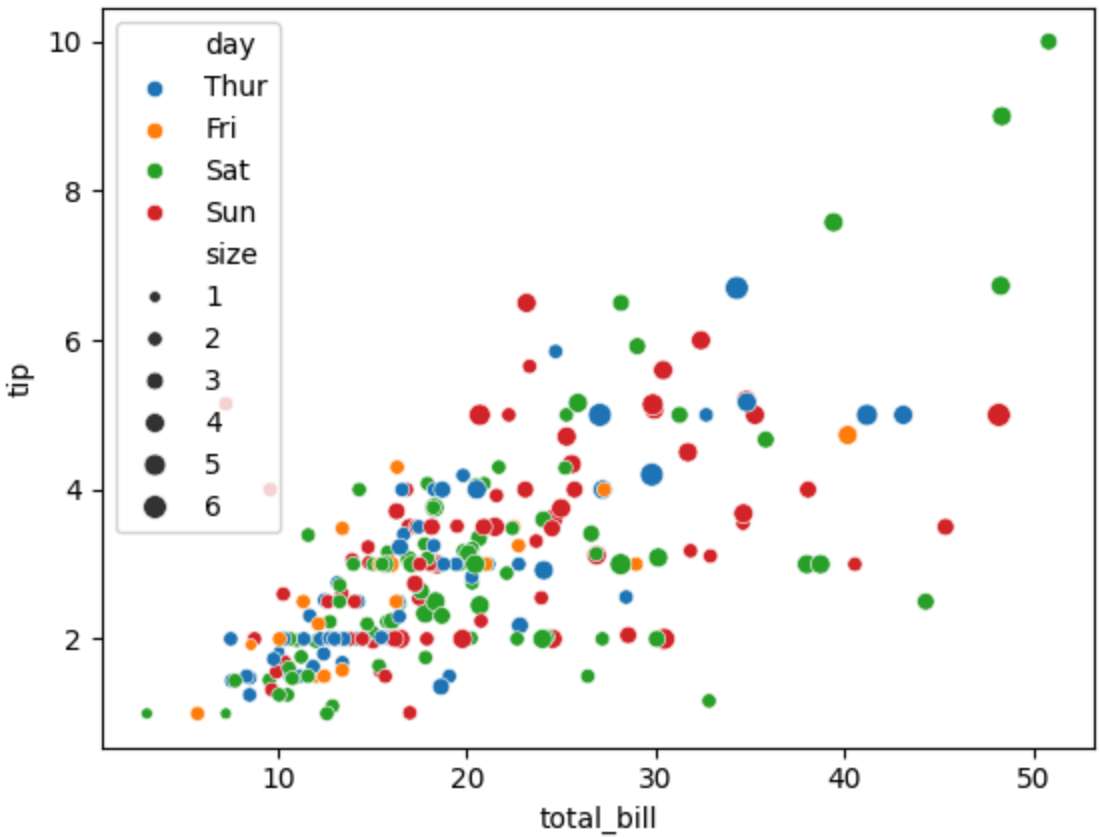
Scatter Plot

In [136…

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

Out[136…

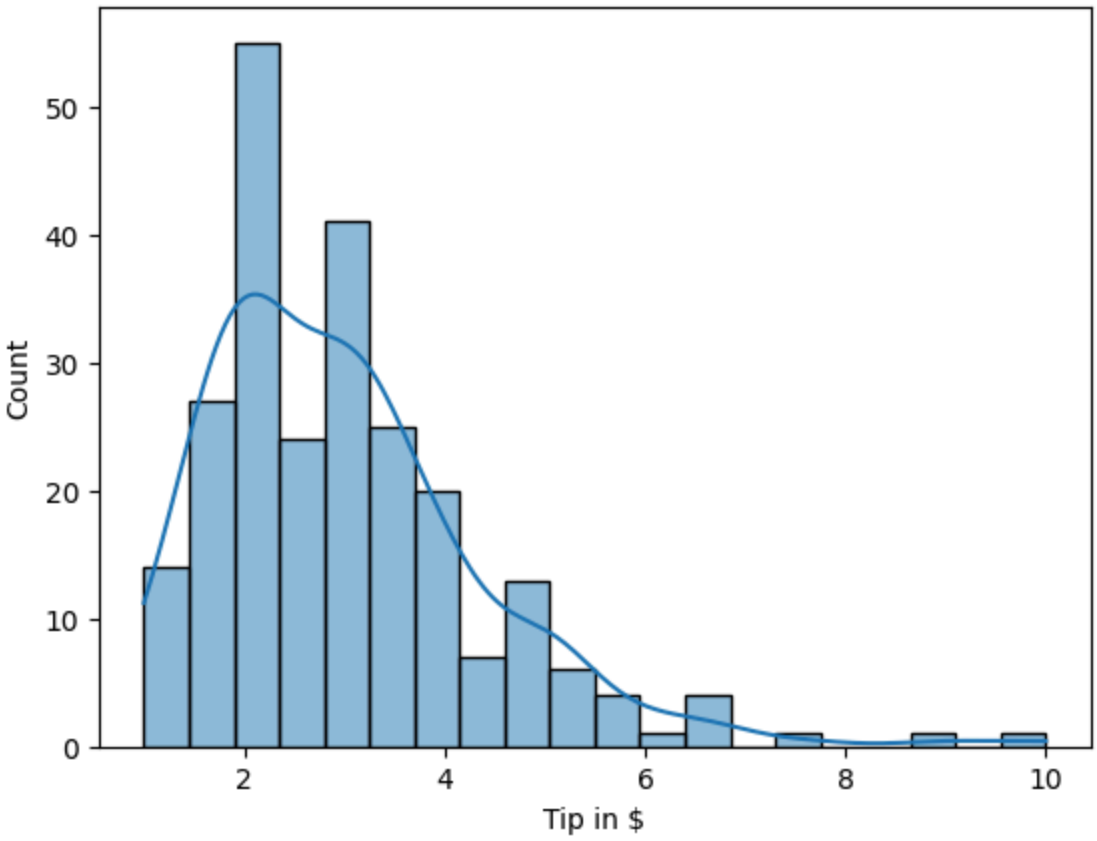
```
<Axes: xlabel='total_bill', ylabel='tip'>
```



Histograms & Distribution Plot

In [137…
```
sns.histplot(tips['tip'], kde=True, bins = 20)
plt.xlabel("Tip in $")
```

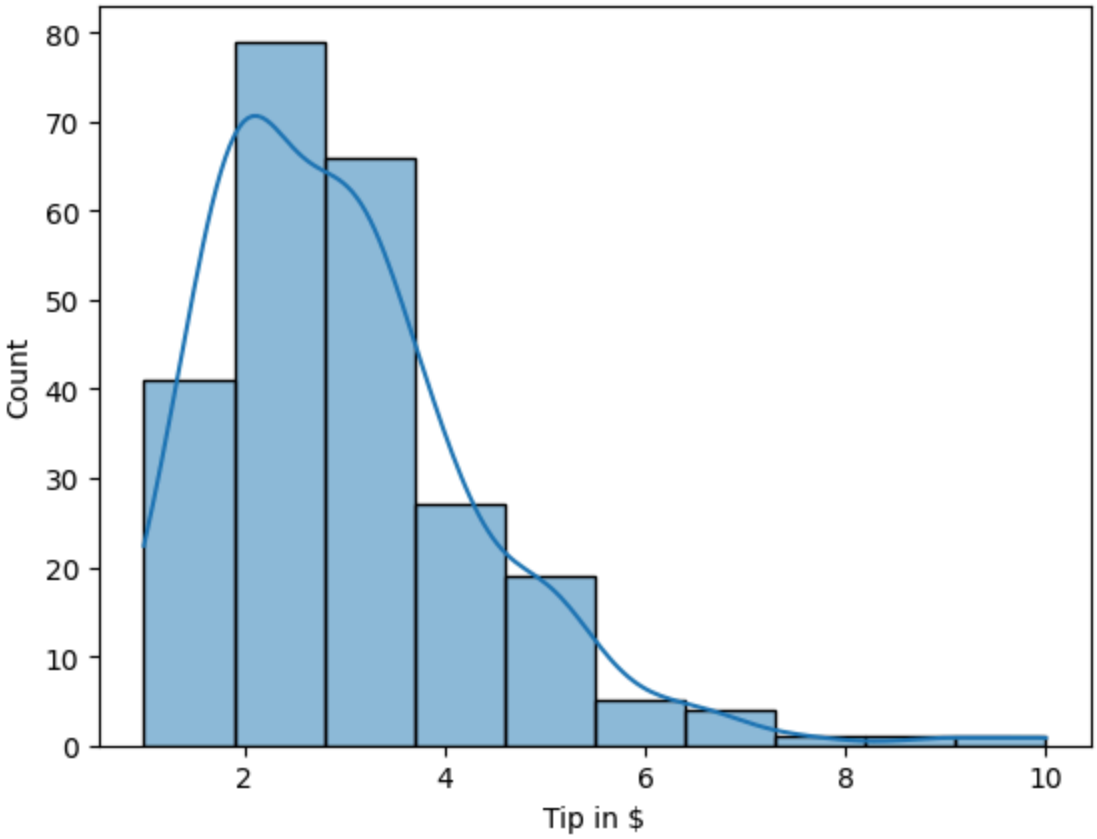
Out[137…
```
Text(0.5, 0, 'Tip in $')
```



Playing with bins size

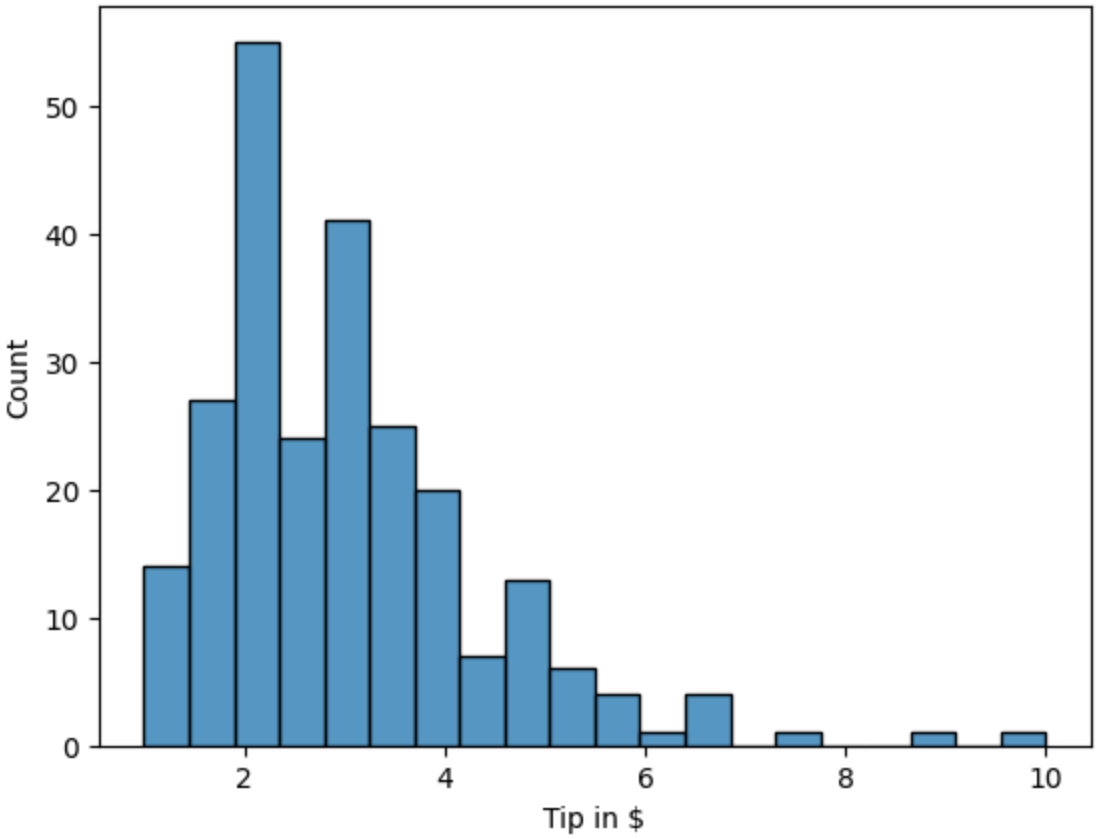
In [138…
```
sns.histplot(tips['tip'], kde=True, bins = 10)
plt.xlabel("Tip in $")
```

Out[138…
```
Text(0.5, 0, 'Tip in $')
```



In [139…
```
sns.histplot(tips['tip'], bins = 20)
plt.xlabel("Tip in $")
```

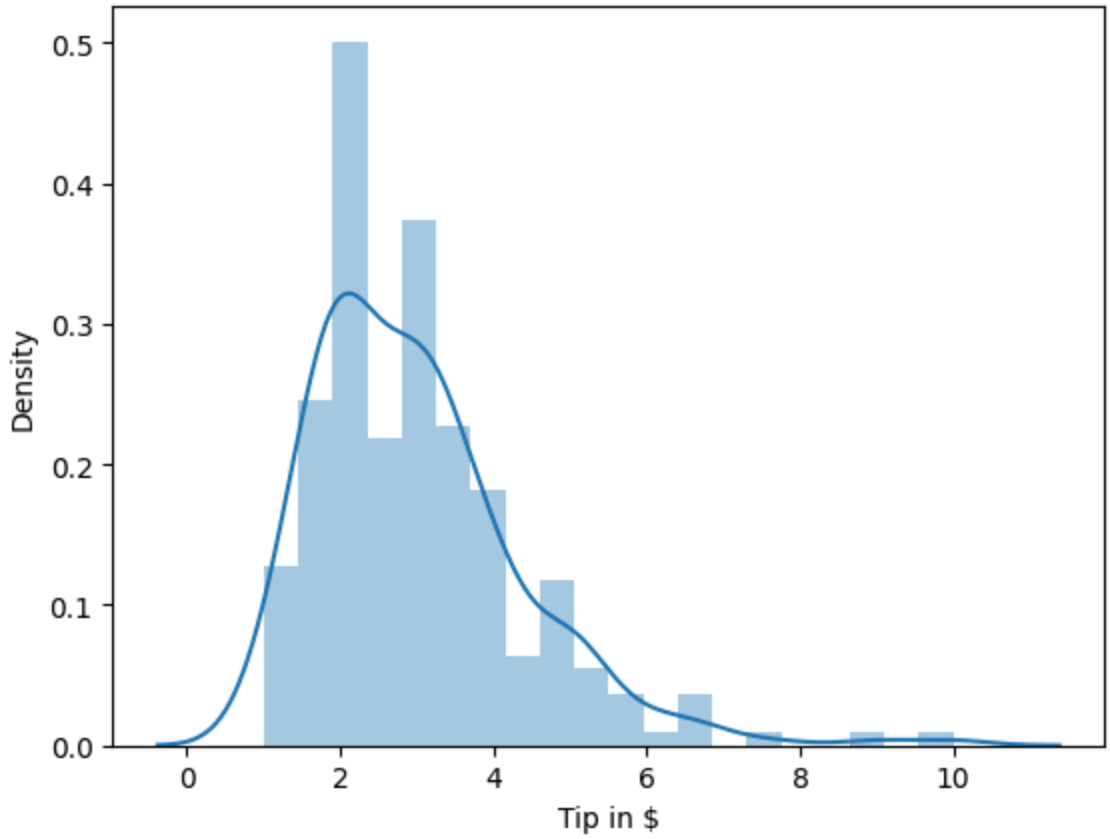
Out[139…
```
Text(0.5, 0, 'Tip in $')
```



In [140…
```
sns.distplot(tips['tip'], kde=True, bins = 20)
plt.xlabel("Tip in $")
```

```
<ipython-input-140-c4b08133469e>:1: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot(tips['tip'], kde=True, bins = 20)
```

Out[140… Text(0.5, 0, 'Tip in $')

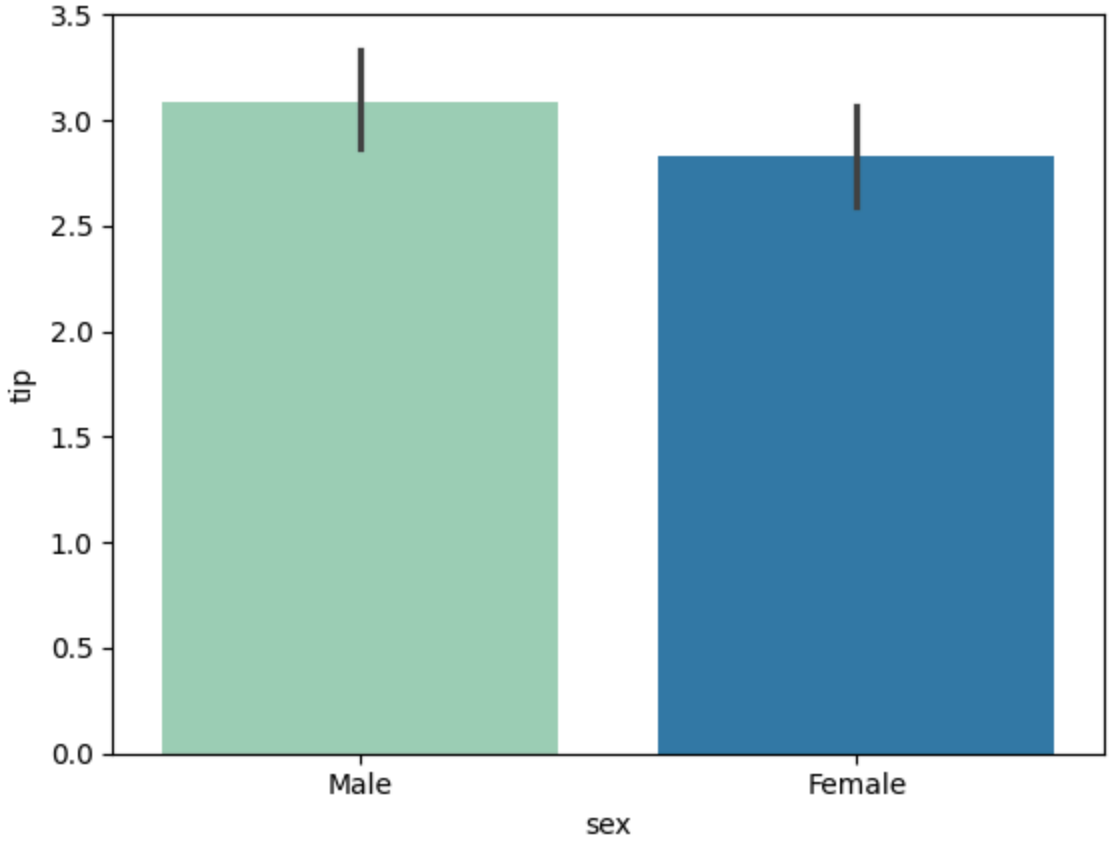


Bar Plot

In [141…
```
sns.barplot(x="sex", y= "tip", data = tips, palette ="YlGnBu")
```

```
<ipython-input-141-7fc6d016a85f>:1: FutureWarning:

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

  sns.barplot(x="sex", y= "tip", data = tips, palette ="YlGnBu")
```

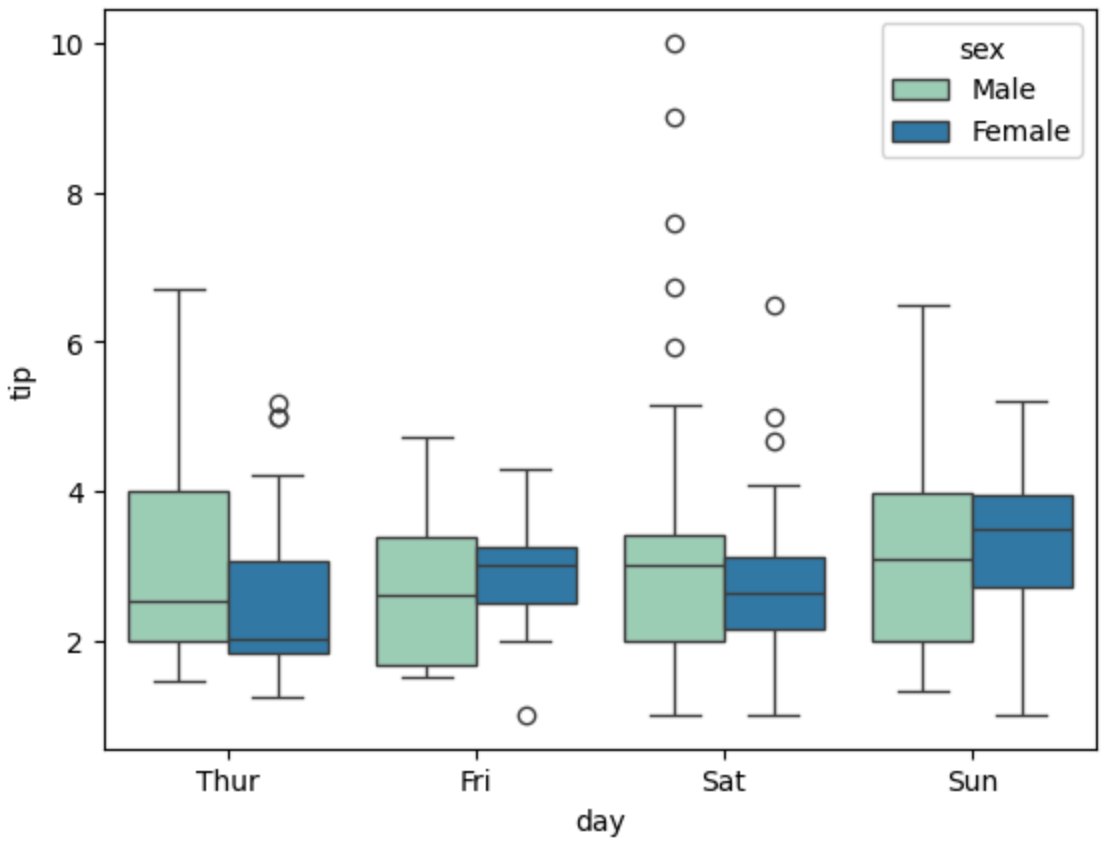
Out[141… <Axes: xlabel='sex', ylabel='tip'>



Box Plot

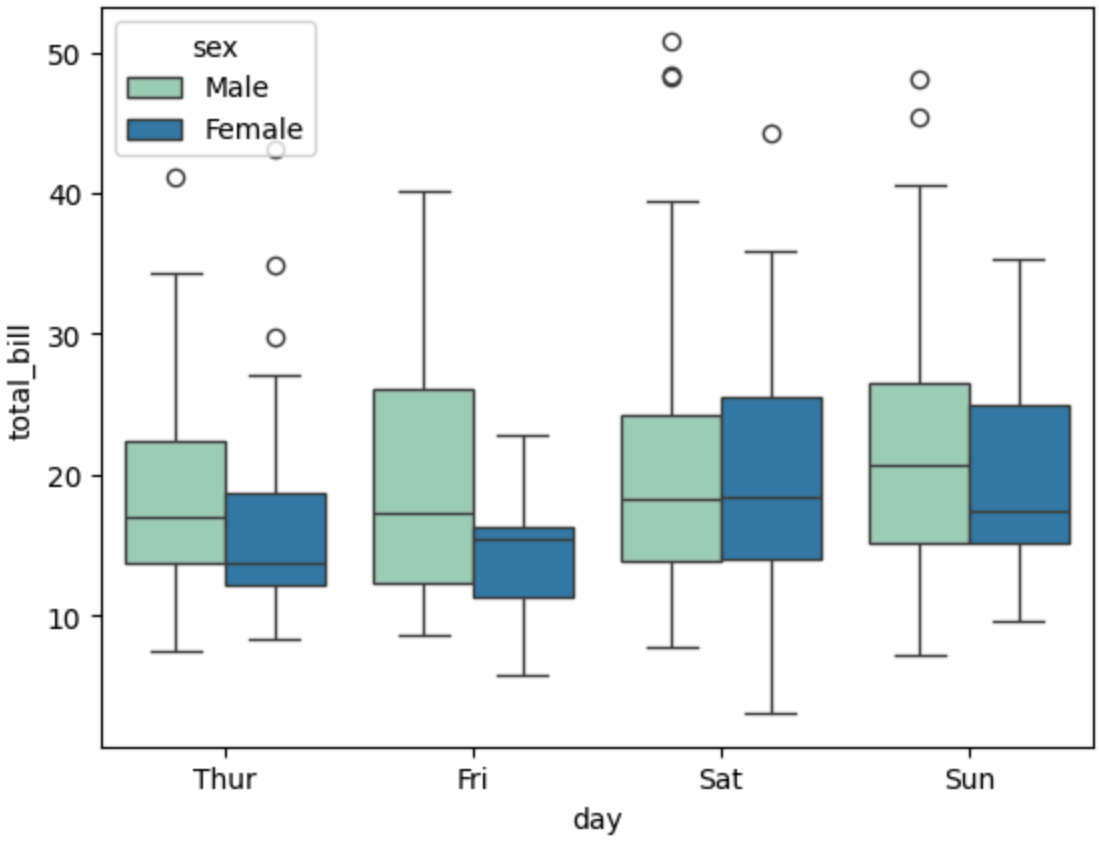
In [142…
```
sns.boxplot(x="day", y="tip", data=tips, hue = "sex", palette = "YlGnBu")
```

Out[142…
```
<Axes: xlabel='day', ylabel='tip'>
```



In [143…
```
sns.boxplot(x="day", y="total_bill", data=tips,hue = "sex", palette = "YlGnBu")
```

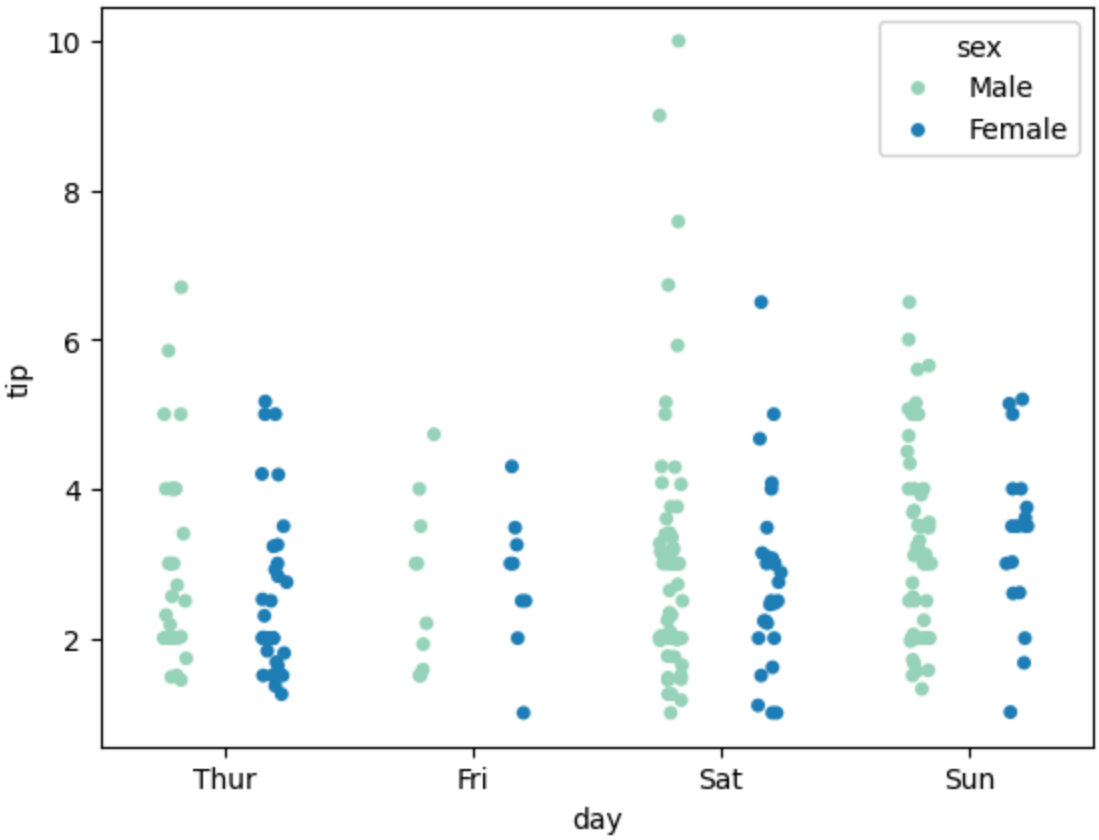
Out[143… <Axes: xlabel='day', ylabel='total_bill'>



Strip Plot

In [144…
```
sns.stripplot(x="day", y="tip", data=tips, palette = "YlGnBu", hue="sex", dodge = True)
```

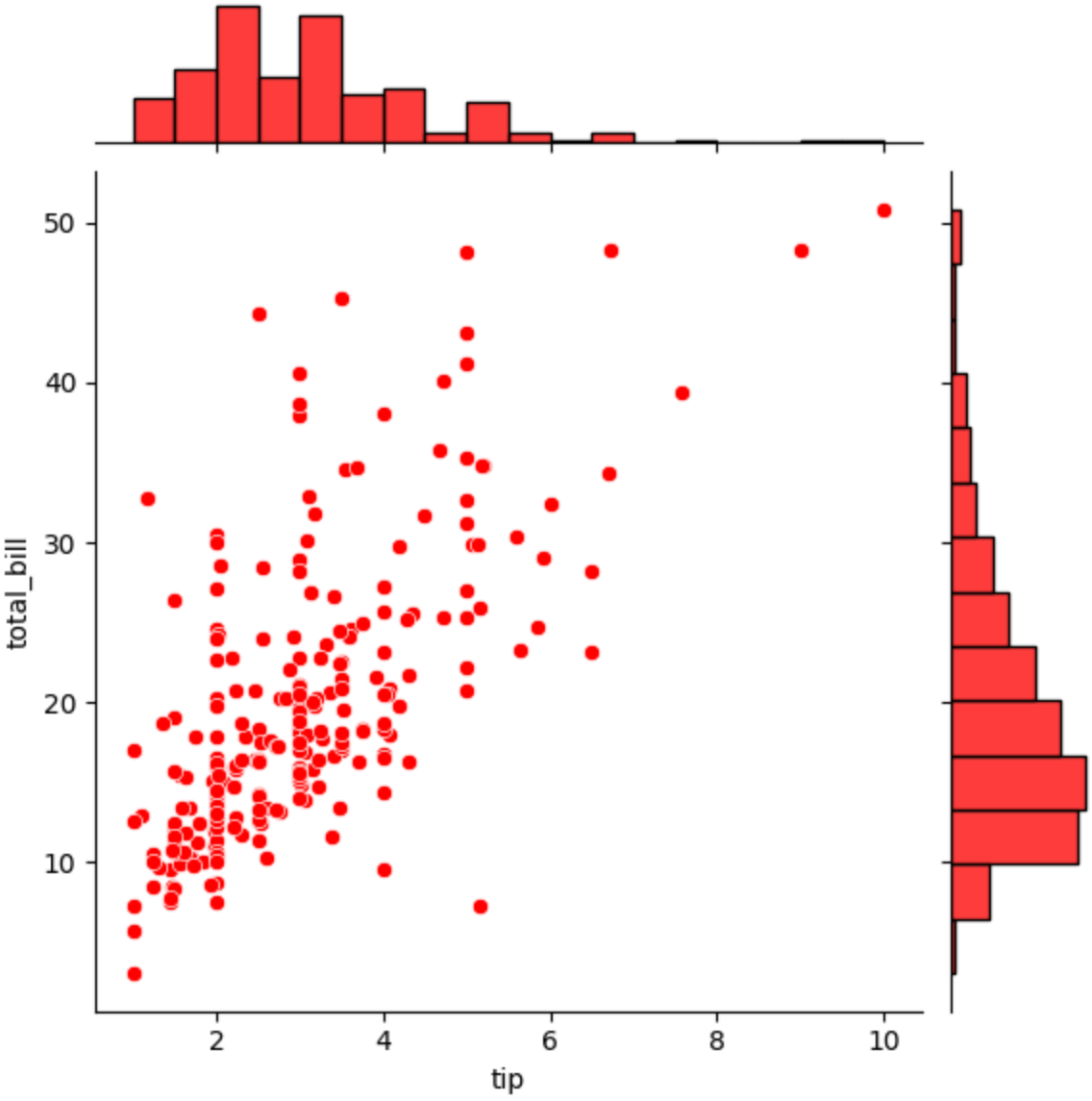
Out[144… <Axes: xlabel='day', ylabel='tip'>



Joint Plot

In [145…
```
sns.jointplot(x="tip", y="total_bill", data=tips, kind="scatter", color = "r")
```

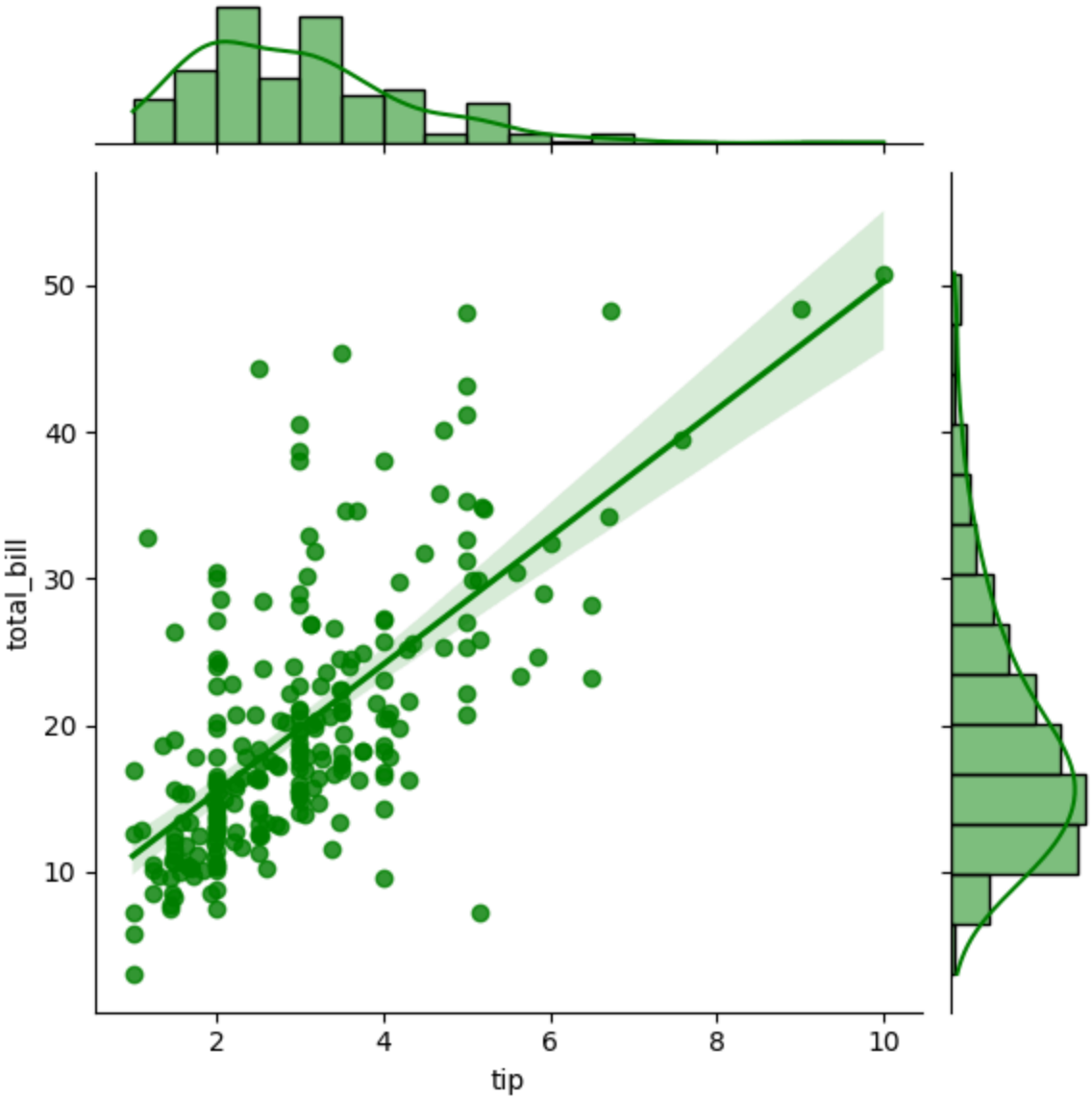
Out[145…
```
<seaborn.axisgrid.JointGrid at 0x79cb02ffe1a0>
```



Joint Plot with Regression

In [146…
```
sns.jointplot(x="tip", y="total_bill", data=tips, kind="reg", color = "g")
```

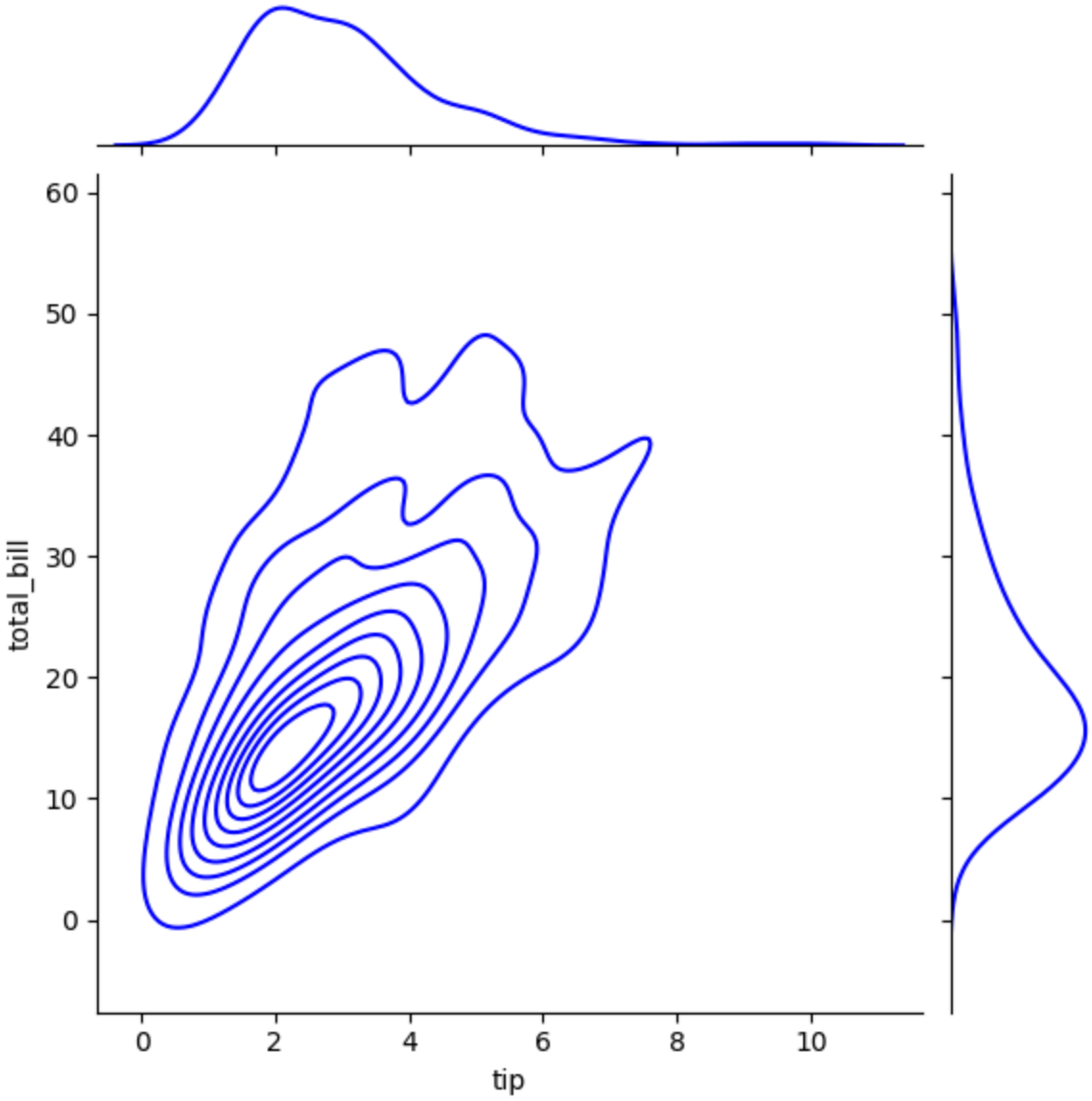
Out[146…
```
<seaborn.axisgrid.JointGrid at 0x79cb02f623b0>
```



Joint Plot with kde without and with shading

In [147…
```
sns.jointplot(x="tip", y="total_bill", data=tips, kind="kde", color = "b")
```

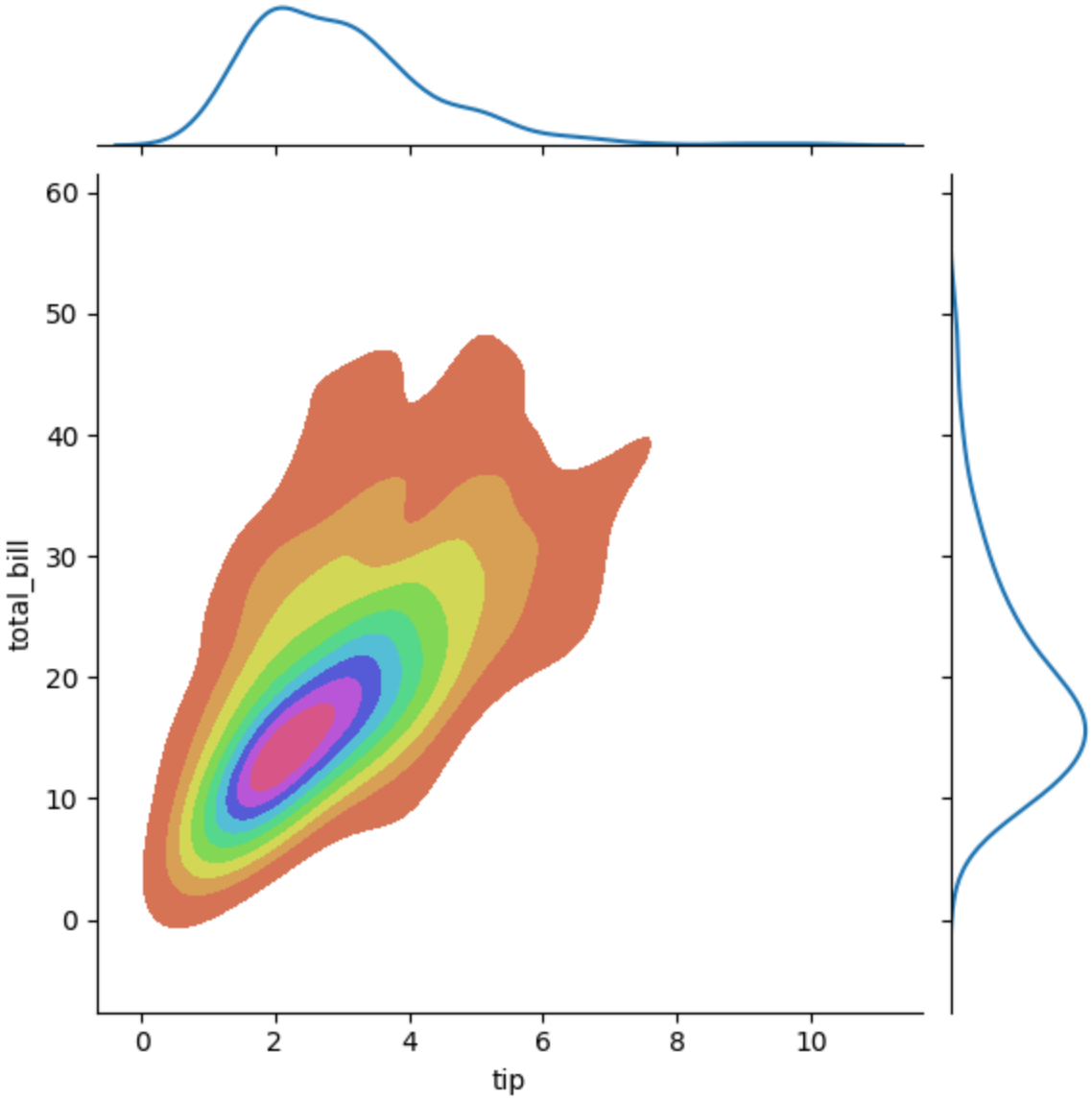
Out[147…
```
<seaborn.axisgrid.JointGrid at 0x79cb02cd2590>
```



In [148…
```
sns.jointplot(x="tip", y="total_bill", data=tips, kind="kde", shade = "True", cmap = "hls")
```

```
/usr/local/lib/python3.10/dist-packages/seaborn/axisgrid.py:1832: FutureWarning:

`shade` is now deprecated in favor of `fill`; setting `fill=True`.
This will become an error in seaborn v0.14.0; please update your code.

  func(x=self.x, y=self.y, **kwargs)
```

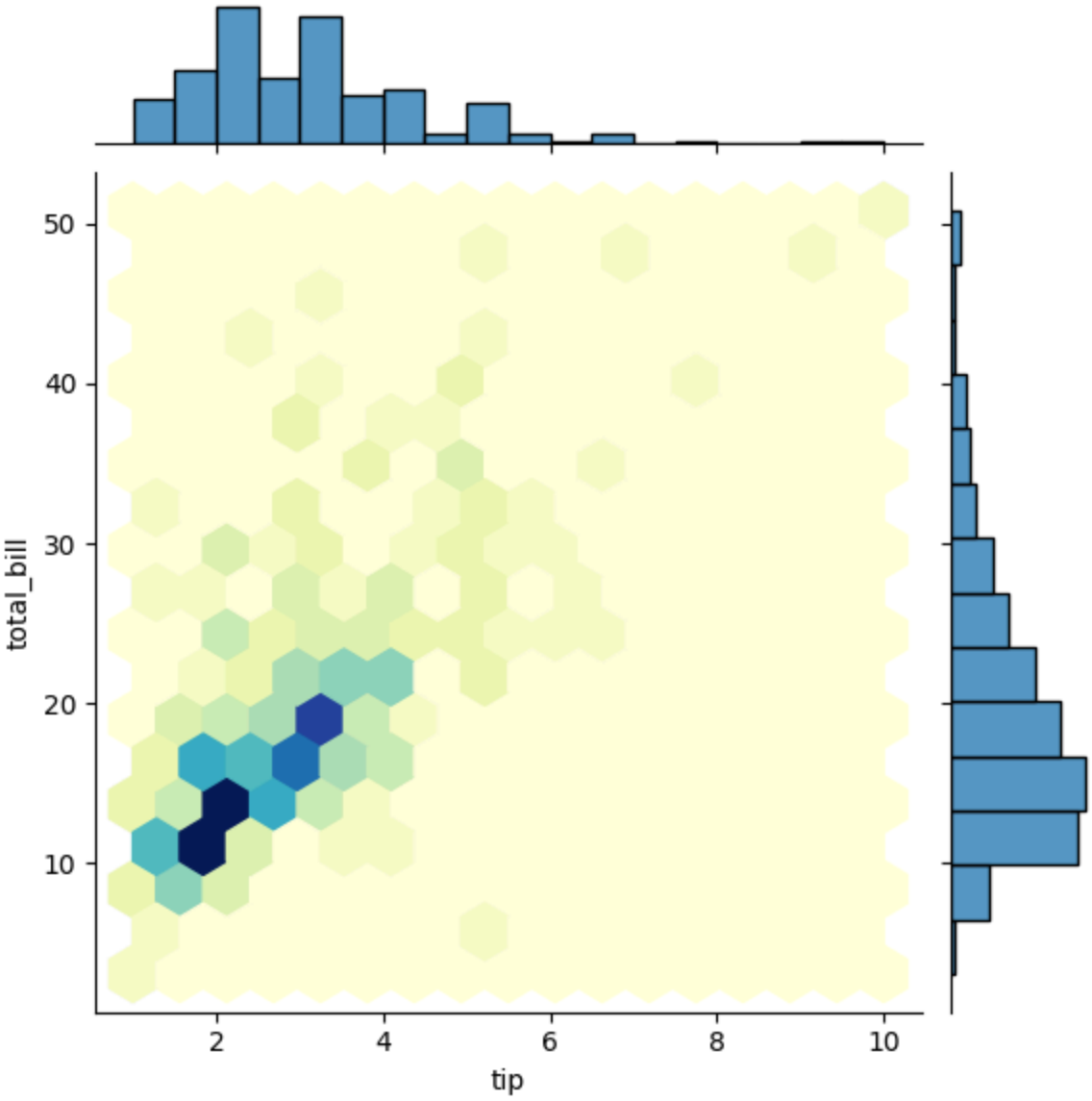
Out[148…
```
<seaborn.axisgrid.JointGrid at 0x79cb02c69930>
```



Hex type joint plot

In [149…
```
sns.jointplot(x="tip", y="total_bill", data=tips, kind="hex", cmap="YlGnBu")
```

Out[149…
```
<seaborn.axisgrid.JointGrid at 0x79cb029a23b0>
```



Checking the titanic dataset

In [150…
```
titanic.head()
```

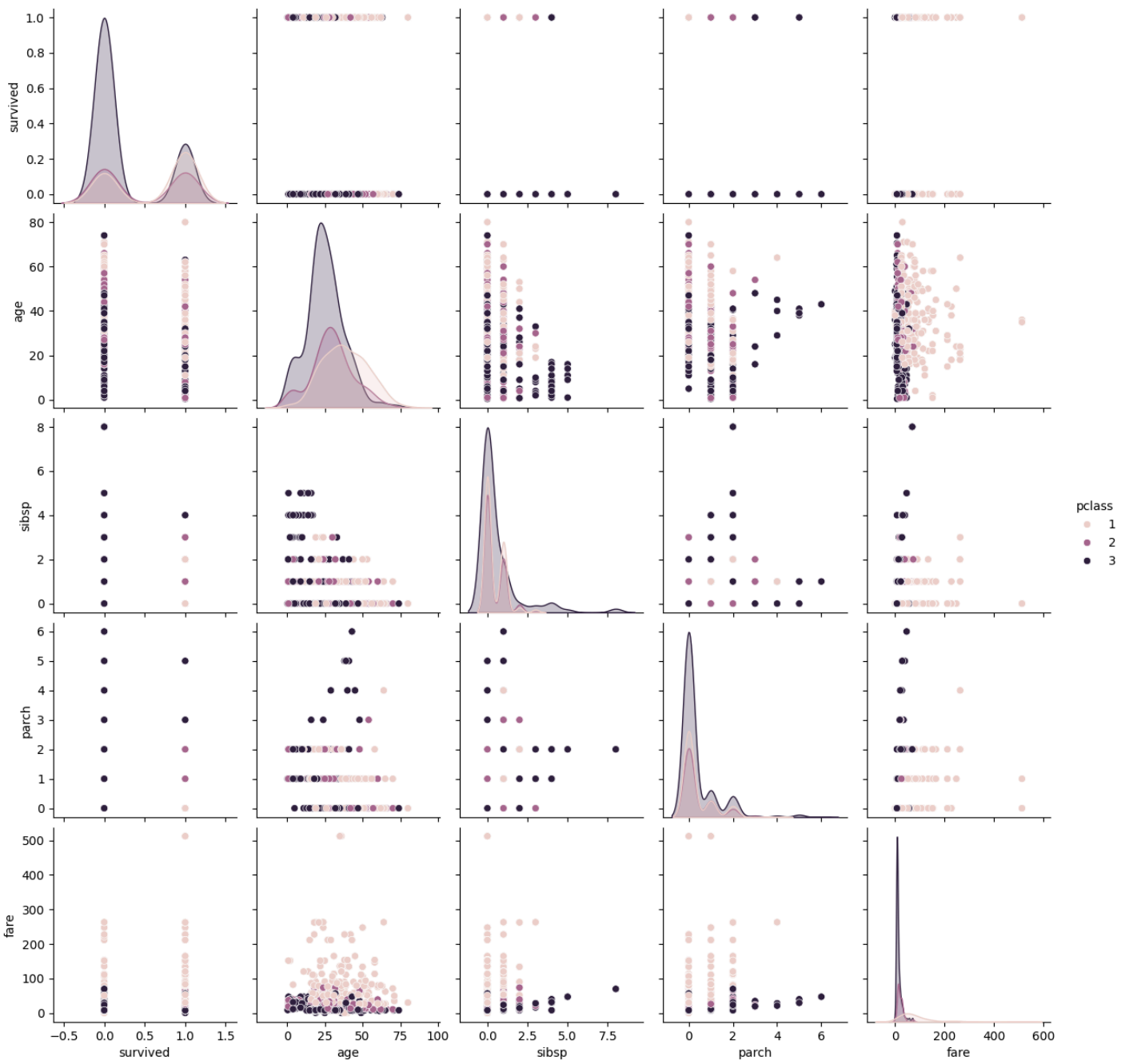
Out[150…

	survived	pclass	sex	age	sibsp	parch	fare	embarked	class	who	adult_male	deck	eml
0	0	3	male	22.0	1	0	7.2500	S	Third	man	True	NaN	Sou
1	1	1	female	38.0	1	0	71.2833	C	First	woman	False	C	
2	1	3	female	26.0	0	0	7.9250	S	Third	woman	False	NaN	Sou
3	1	1	female	35.0	1	0	53.1000	S	First	woman	False	C	Sou
4	0	3	male	35.0	0	0	8.0500	S	Third	man	True	NaN	Sou

Pair plot on titanic dataset

In [151…
```
sns.pairplot(titanic.select_dtypes(["number"]), hue ="pclass")
```

Out[151…
```
<seaborn.axisgrid.PairGrid at 0x79cb028e3c70>
```



Preprocessing with label encoder

In [152…

```
from sklearn import preprocessing

le = preprocessing.LabelEncoder()
columns = ["adult_male", "sex","alone"]
for col in columns:
    titanic[col] = le.fit_transform(titanic[col])
titanic.head()
```

Out[152...

	survived	pclass	sex	age	sibsp	parch	fare	embarked	class	who	adult_male	deck	embarl
0	0	3	1	22.0	1	0	7.2500	S	Third	man	1	NaN	Southa
1	1	1	0	38.0	1	0	71.2833	C	First	woman	0	C	Che
2	1	3	0	26.0	0	0	7.9250	S	Third	woman	0	NaN	Southa
3	1	1	0	35.0	1	0	53.1000	S	First	woman	0	C	Southa
4	0	3	1	35.0	0	0	8.0500	S	Third	man	1	NaN	Southa

Considering only numerical values

In [153...

```
titanicnum = titanic.select_dtypes(["number"])
titanicnum.head()
```

Out[153...

	survived	pclass	sex	age	sibsp	parch	fare	adult_male	alone
0	0	3	1	22.0	1	0	7.2500	1	0
1	1	1	0	38.0	1	0	71.2833	0	0
2	1	3	0	26.0	0	0	7.9250	0	1
3	1	1	0	35.0	1	0	53.1000	0	0
4	0	3	1	35.0	0	0	8.0500	1	1

Correlation of different factors

In [154...

```
titanicnum.corr()
```

Out[154...

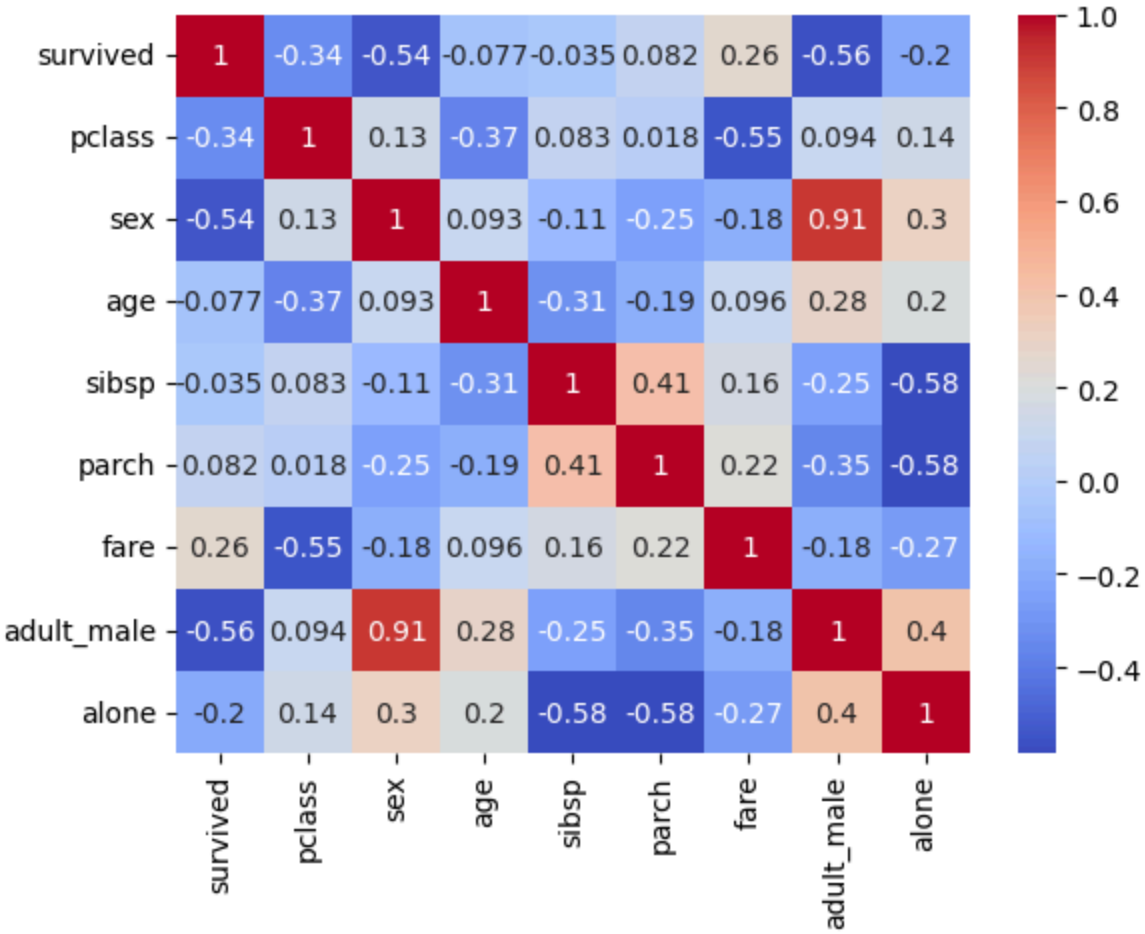
	survived	pclass	sex	age	sibsp	parch	fare	adult_male	alo
survived	1.000000	-0.338481	-0.543351	-0.077221	-0.035322	0.081629	0.257307	-0.557080	-0.203
pclass	-0.338481	1.000000	0.131900	-0.369226	0.083081	0.018443	-0.549500	0.094035	0.135
sex	-0.543351	0.131900	1.000000	0.093254	-0.114631	-0.245489	-0.182333	0.908578	0.303
age	-0.077221	-0.369226	0.093254	1.000000	-0.308247	-0.189119	0.096067	0.280328	0.198
sibsp	-0.035322	0.083081	-0.114631	-0.308247	1.000000	0.414838	0.159651	-0.253586	-0.584
parch	0.081629	0.018443	-0.245489	-0.189119	0.414838	1.000000	0.216225	-0.349943	-0.583
fare	0.257307	-0.549500	-0.182333	0.096067	0.159651	0.216225	1.000000	-0.182024	-0.271
adult_male	-0.557080	0.094035	0.908578	0.280328	-0.253586	-0.349943	-0.182024	1.000000	0.404
alone	-0.203367	0.135207	0.303646	0.198270	-0.584471	-0.583398	-0.271832	0.404744	1.000

Heatmap on titanic dataset

In [155...

```
sns.heatmap(titanicnum.corr(), annot = True, cmap = "coolwarm")
```

Out[155… <Axes: >



**Iris dataset & clustermap **

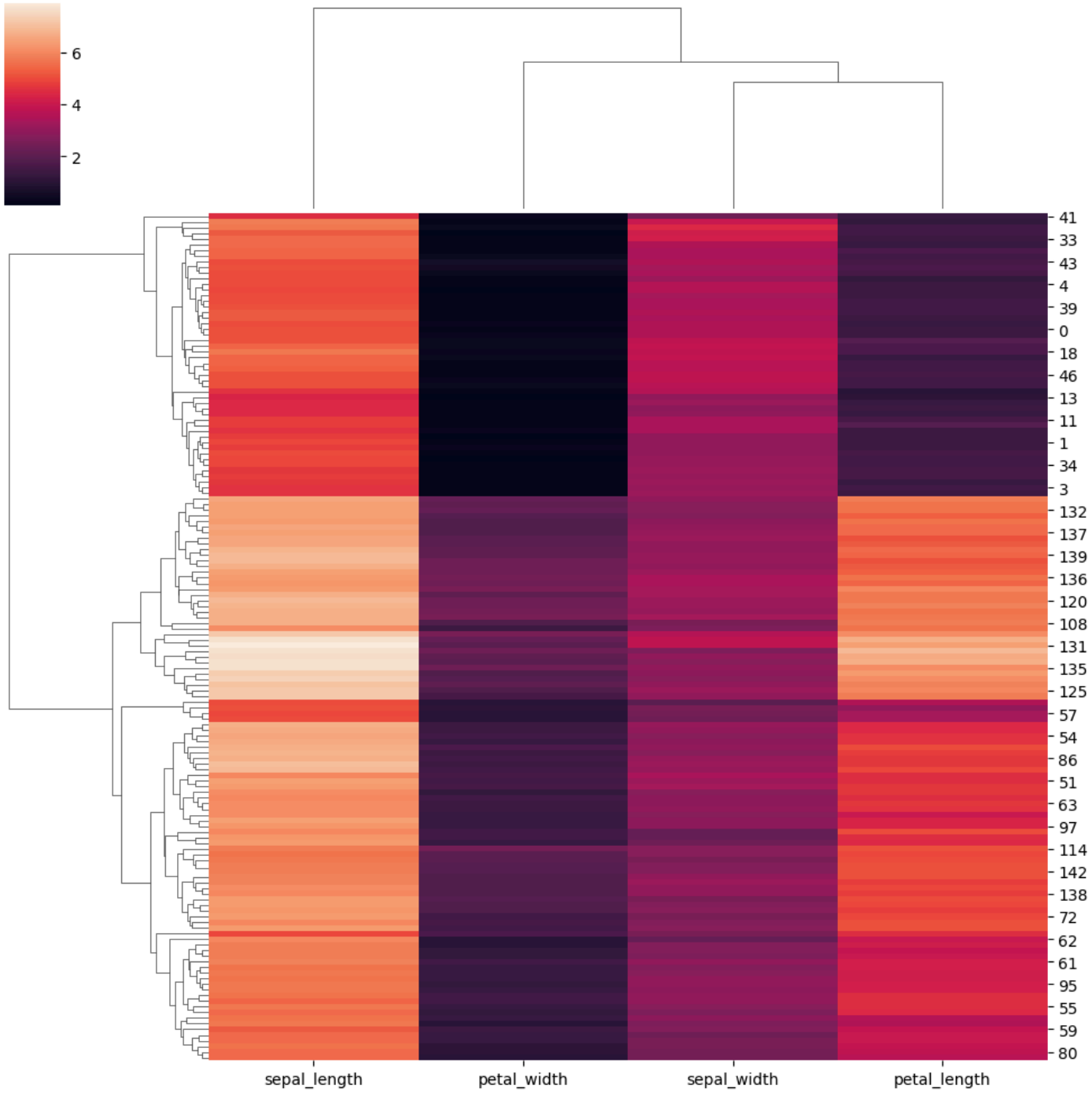
In [156… `iris.head()`

Out[156…

	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

In [157… `sns.clustermap(iris.drop("species", axis = 1))`

Out[157… `<seaborn.matrix.ClusterGrid at 0x79cb028e19f0>`


6
4
2
41
33
43
4
39
0
18
46
13
11
1
34
3
132
137
139
136
120
108
131
135
125
57
54
86
51
63
97
114
142
138
72
62
61
95
55
59
80
sepal_length
petal_width
sepal_width
petal_length