

```
In [ ]: # packages for data analysis
import numpy as np
import pandas as pd

from sklearn import svm

# visual your data
import matplotlib.pyplot as plt
import seaborn as sns; sns.set(font_scale=1.2)

%matplotlib inline
```

```
In [7]: CancerTypes = pd.read_csv('Cancer_dataset.csv')
print(CancerTypes.head())
```

	Patient_ID	Age	Gender	Protein1	Protein2	Protein3	Protein4	\
0	TCGA-D8-A1XD	36	FEMALE	0.080353	0.42638	0.54715	0.273680	
1	TCGA-EW-A10X	43	FEMALE	-0.420320	0.57807	0.61447	-0.031505	
2	TCGA-A8-A079	69	FEMALE	0.213980	1.31140	-0.32747	-0.234260	
3	TCGA-D8-A1XR	56	FEMALE	0.345090	-0.21147	-0.19304	0.124270	
4	TCGA-BH-A0BF	56	FEMALE	0.221550	1.90680	0.52045	-0.311990	

	Tumour_Stage	Histology	ER status	PR status	HER2 status	\
0	III	Infiltrating Ductal Carcinoma	Positive	Positive	Negative	
1	II	Mucinous Carcinoma	Positive	Positive	Negative	
2	III	Infiltrating Ductal Carcinoma	Positive	Positive	Negative	
3	II	Infiltrating Ductal Carcinoma	Positive	Positive	Negative	
4	II	Infiltrating Ductal Carcinoma	Positive	Positive	Negative	

	Surgery_type	Date_of_Surgery	Date_of_Last_Visit	\
0	Modified Radical Mastectomy	15-Jan-17	19-Jun-17	
1	Lumpectomy	26-Apr-17	9-Nov-18	
2	Other	8-Sep-17	9-Jun-18	
3	Modified Radical Mastectomy	25-Jan-17	12-Jul-17	
4	Other	6-May-17	27-Jun-19	

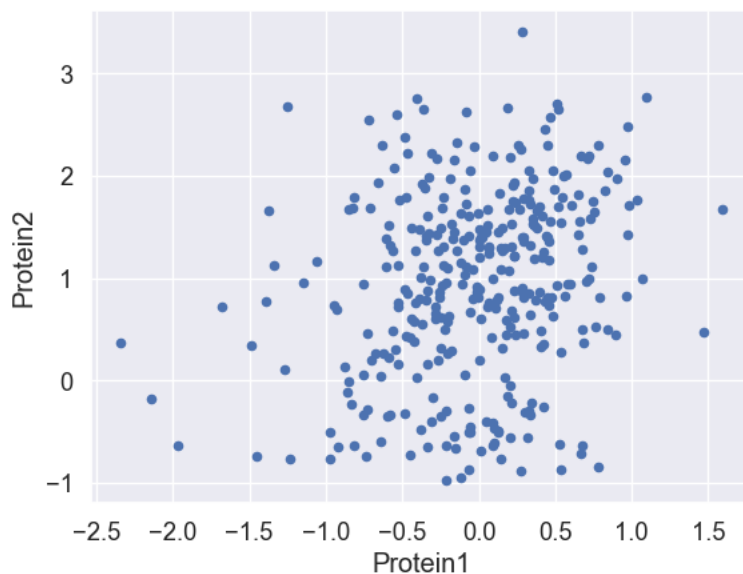
	Patient_Status
0	Alive
1	Dead
2	Alive
3	Alive
4	Dead

```
In [20]: import pandas as pd
import matplotlib.pyplot as plt
```

```
In [22]: df = pd.read_csv('Cancer_dataset.csv')
```

```
In [23]: df.plot(kind = 'scatter', x = 'Protein1', y = 'Protein2')
plt.show()
```

c argument looks like a single numeric RGB or RGBA sequence, which should be avoided as value-mapping will have precedence in case its length matches with *x* & *y*. Please use the *color* keyword-argument or provide a 2D array with a single row if you intend to specify the same RGB or RGBA value for all points.



In []: ▶