

```
In [1]: import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

```
In [2]: crop = pd.read_csv('Dataset/Crop.csv')
crop.head(5)
```

```
Out[2]:   N   P   K  temperature  humidity    ph  rainfall  label
0  90  42  43    20.879744  82.002744  6.502985  202.935536  rice
1  85  58  41    21.770462  80.319644  7.038096  226.655537  rice
2  60  55  44    23.004459  82.320763  7.840207  263.964248  rice
3  74  35  40    26.491096  80.158363  6.980401  242.864034  rice
4  78  42  42    20.130175  81.604873  7.628473  262.717340  rice
```

```
In [3]: crop.describe()
```

```
Out[3]:      N          P          K  temperature  humidity    ph  rainfall
count  2200.000000  2200.000000  2200.000000  2200.000000  2200.000000  2200.000000  2200.000000
mean   50.551818  53.362727  48.149091  25.616244  71.481779  6.469480  103.463655
std    36.917334  32.985883  50.647931  5.063749  22.263812  0.773938  54.958389
min    0.000000  5.000000  5.000000  8.825675  14.258040  3.504752  20.211267
25%   21.000000  28.000000  20.000000  22.769375  60.261953  5.971693  64.551686
50%   37.000000  51.000000  32.000000  25.598693  80.473146  6.425045  94.867624
75%   84.250000  68.000000  49.000000  28.561654  89.948771  6.923643  124.267508
max   140.000000  145.000000  205.000000  43.675493  99.981876  9.935091  298.560117
```

```
In [4]: crop.shape
```

```
Out[4]: (2200, 8)
```

```
In [4]: crop.dtypes
```

```
Out[4]: N          int64
P          int64
K          int64
temperature  float64
humidity    float64
ph          float64
rainfall    float64
label       object
dtype: object
```

```
In [6]: crop['label'] = crop['label'].dropna()
crop.isna().sum()
```

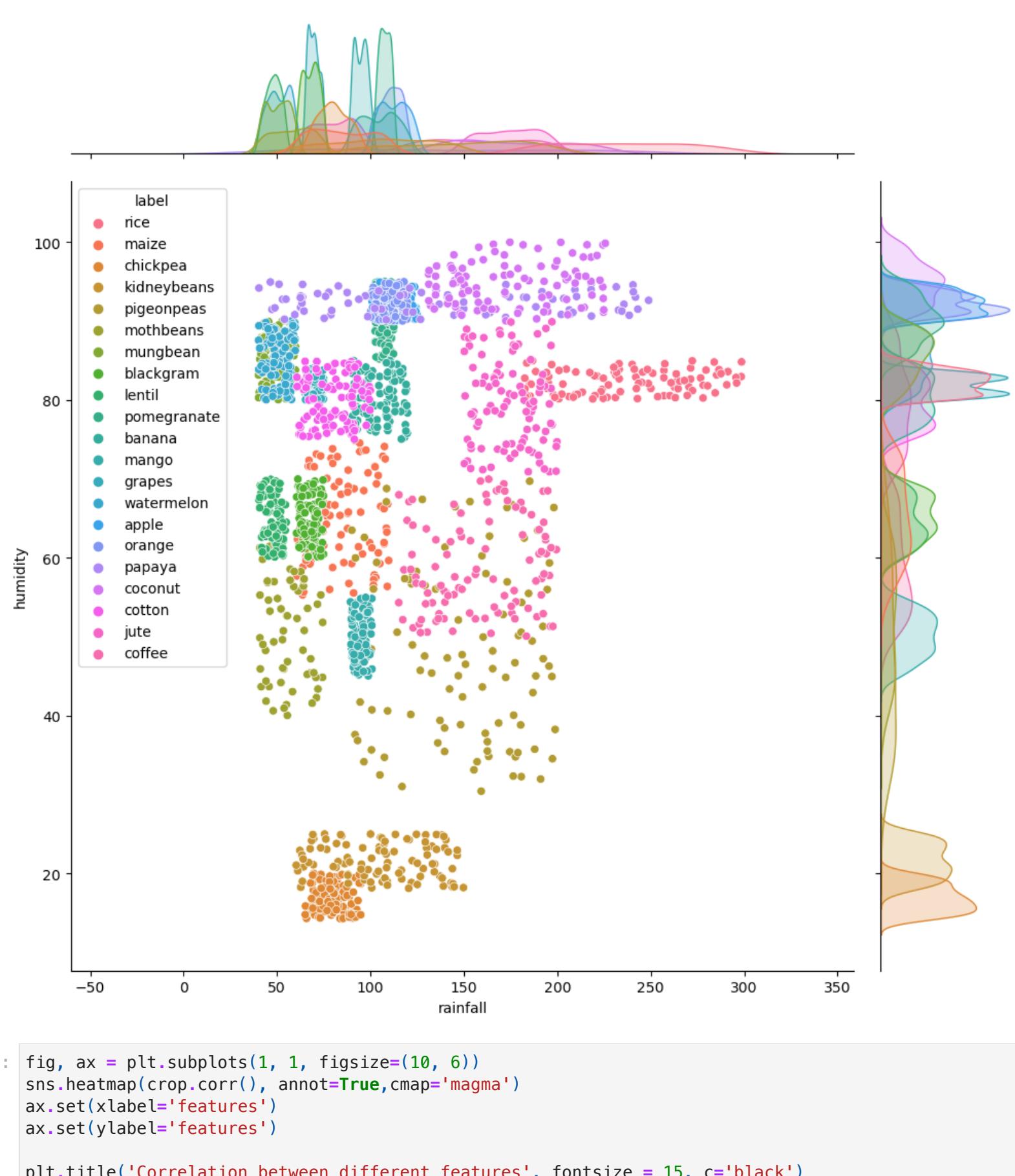
```
Out[6]: N          0
P          0
K          0
temperature 0
humidity    0
ph          0
rainfall    0
label       0
dtype: int64
```

```
In [7]: import seaborn as sns
sns.pairplot(crop,hue = 'label')
```

Intel MKL WARNING: Support of Intel(R) Streaming SIMD Extensions 4.2 (Intel(R) SSE4.2) enabled only on processors has been deprecated. Intel oneAPI Math Kernel Library 2025.0 will require Intel(R) Advanced Vector Extensions (Intel(R) AVX) instructions.

/Users/kevilm/opt/anaconda3/lib/python3.8/site-packages/seaborn/axisgrid.py:118: UserWarning: The figure layout has changed to tight
self._figure.tight_layout(*args, **kwargs)

```
Out[7]: <seaborn.axisgrid.PairGrid at 0x7f7d1006c3a0>
```



```
In [8]: sns.jointplot(x="rainfall",y="humidity",data=crop[(crop['temperature']<40) & (crop['rainfall']>40)],height=10,hue="label")
```

```
Out[8]: <seaborn.axisgrid.JointGrid at 0x7f7d309b5be0>
```



```
In [14]: fig, ax = plt.subplots(1, 1, figsize=(10, 6))
sns.heatmap(crop.corr(), annot=True,cmap='magma')
ax.set(xlabel='features')
ax.set(ylabel='features')

plt.title('Correlation between different features', fontsize = 15, c='black')
plt.show()
```

/tmp/ipykernel_12557/849159460.py:2: FutureWarning: The default value of numeric_only in DataFrame corr is deprecated. In a future version, it will default to False. Select only valid columns or specify the value of numeric_only to silence this warning.

```
sns.heatmap(crop.corr(), annot=True,cmap='magma')
```

Correlation between different features

