xbikehdf6

August 4, 2023

1 20104169 - SUMESH R

2 Importing Libraries

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

[2]:
```
from google.colab import drive
drive.mount('/content/drive')
df=pd.read_csv("/content/drive/MyDrive/mydatasets/rainfall/rainfall_madhya␣
 ↪maharashtra.csv")
df
```

```
Mounted at /content/drive
```

[2]:
```
     index         SUBDIVISION  YEAR   JAN  FEB   MAR   APR   MAY    JUN  \
0     2622  MADHYA MAHARASHTRA  1901  18.8  0.6   7.7  36.6  30.4  107.7
1     2623  MADHYA MAHARASHTRA  1902   7.8  0.0   0.1   5.0   9.8  102.6
2     2624  MADHYA MAHARASHTRA  1903   7.6  0.0   0.0   3.2  77.2   86.3
3     2625  MADHYA MAHARASHTRA  1904   0.4  4.7   1.7   3.0  18.7  114.6
4     2626  MADHYA MAHARASHTRA  1905   0.0  1.2   0.0   2.3  23.6   65.0
..     …                   …    …     …   …     …     …     …      …
110   2732  MADHYA MAHARASHTRA  2011   0.0  0.3   0.3   5.0   2.9  133.3
111   2733  MADHYA MAHARASHTRA  2012   0.0  0.0   0.0   3.0   1.4   67.9
112   2734  MADHYA MAHARASHTRA  2013   0.1  5.3   0.8   5.7   6.0  212.4
113   2735  MADHYA MAHARASHTRA  2014   3.1  6.2  24.4   7.5  29.8   44.0
114   2736  MADHYA MAHARASHTRA  2015   1.4  0.8  41.2   9.6  24.4  177.0

       JUL    AUG    SEP   OCT   NOV   DEC  ANNUAL  Jan-Feb  Mar-May  Jun-Sep  \
0    215.9  194.1   83.7  68.7   4.4   0.5   769.0     19.4     74.7    601.4
1    210.9  114.5  169.5  60.4  40.5  62.9   784.0      7.8     14.9    597.5
2    281.8  155.5  142.3  74.2   7.6   2.2   837.9      7.6     80.4    665.9
3    126.5   59.5  183.0  91.1   0.0   0.4   603.5      5.1     23.4    483.6
4    252.8   79.0   52.6  52.9   8.3   0.0   537.8      1.2     25.9    449.5
..     …      …      …     …     …     …       …        …        …        …
```

```
110  261.4  238.1  148.4  62.8   0.0   0.0   852.6       0.3       8.2    781.3
111  203.0  187.8  129.5  95.2   2.2   0.0   689.8       0.0       4.4    588.1
112  311.8  147.0  210.3  57.8   4.0   1.3   962.4       5.3      12.4    881.5
113  277.9  240.3  120.4  38.5  32.8  13.1   838.0       9.3      61.7    682.6
114  111.7   67.2  146.6  48.3  16.2   0.1   644.5       2.2      75.3    502.5

     Oct-Dec
0       73.5
1      163.8
2       84.1
3       91.4
4       61.2
..       …
110     62.8
111     97.3
112     63.1
113     84.4
114     64.5

[115 rows x 20 columns]
```

3 Data Cleaning and Data Preprocessing

```
[3]: df=df.dropna()
```

```
[4]: df.columns
```

```
[4]: Index(['index', 'SUBDIVISION', 'YEAR', 'JAN', 'FEB', 'MAR', 'APR', 'MAY',
            'JUN', 'JUL', 'AUG', 'SEP', 'OCT', 'NOV', 'DEC', 'ANNUAL', 'Jan-Feb',
            'Mar-May', 'Jun-Sep', 'Oct-Dec'],
           dtype='object')
```

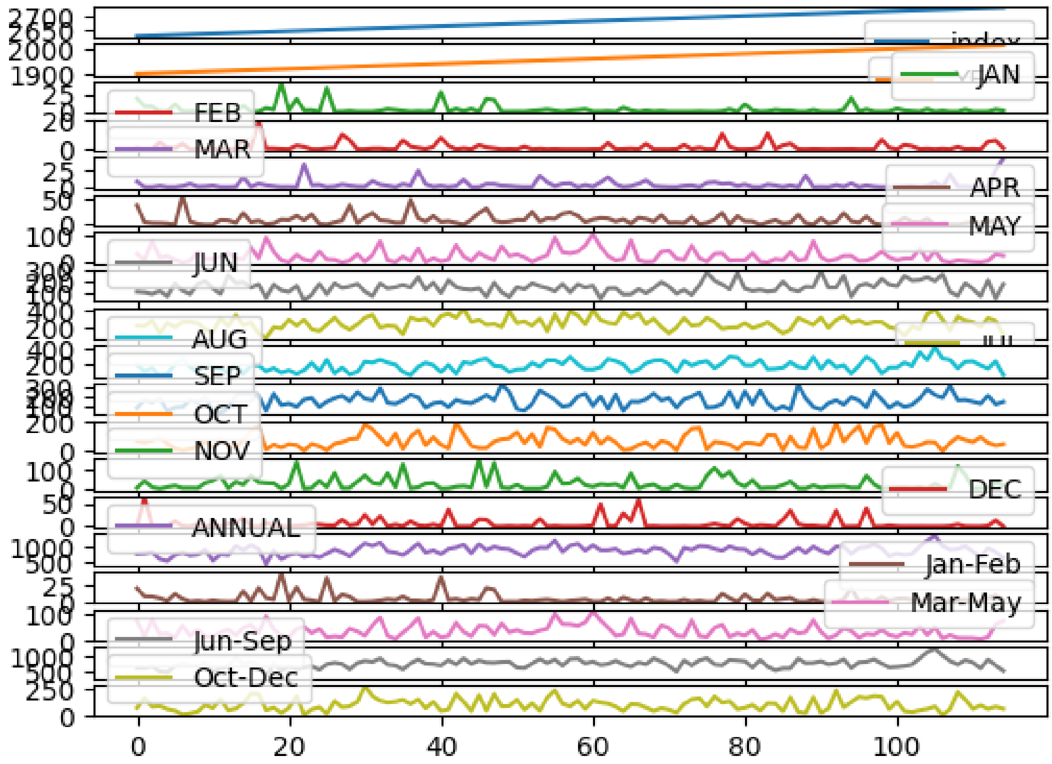
```
[5]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 115 entries, 0 to 114
Data columns (total 20 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   index        115 non-null    int64
 1   SUBDIVISION  115 non-null    object
 2   YEAR         115 non-null    int64
 3   JAN          115 non-null    float64
 4   FEB          115 non-null    float64
 5   MAR          115 non-null    float64
 6   APR          115 non-null    float64
 7   MAY         115 non-null    float64
 8   JUN         115 non-null    float64
 9   JUL         115 non-null    float64
 10  AUG         115 non-null    float64
 11  SEP         115 non-null    float64
 12  OCT         115 non-null    float64
 13  NOV         115 non-null    float64
 14  DEC         115 non-null    float64
 15  ANNUAL      115 non-null    float64
 16  Jan-Feb     115 non-null    float64
 17  Mar-May     115 non-null    float64
 18  Jun-Sep     115 non-null    float64
 19  Oct-Dec     115 non-null    float64
dtypes: float64(17), int64(2), object(1)
memory usage: 18.1+ KB
```

4 Line chart

[6]:
```
df.plot.line(subplots=True)
```

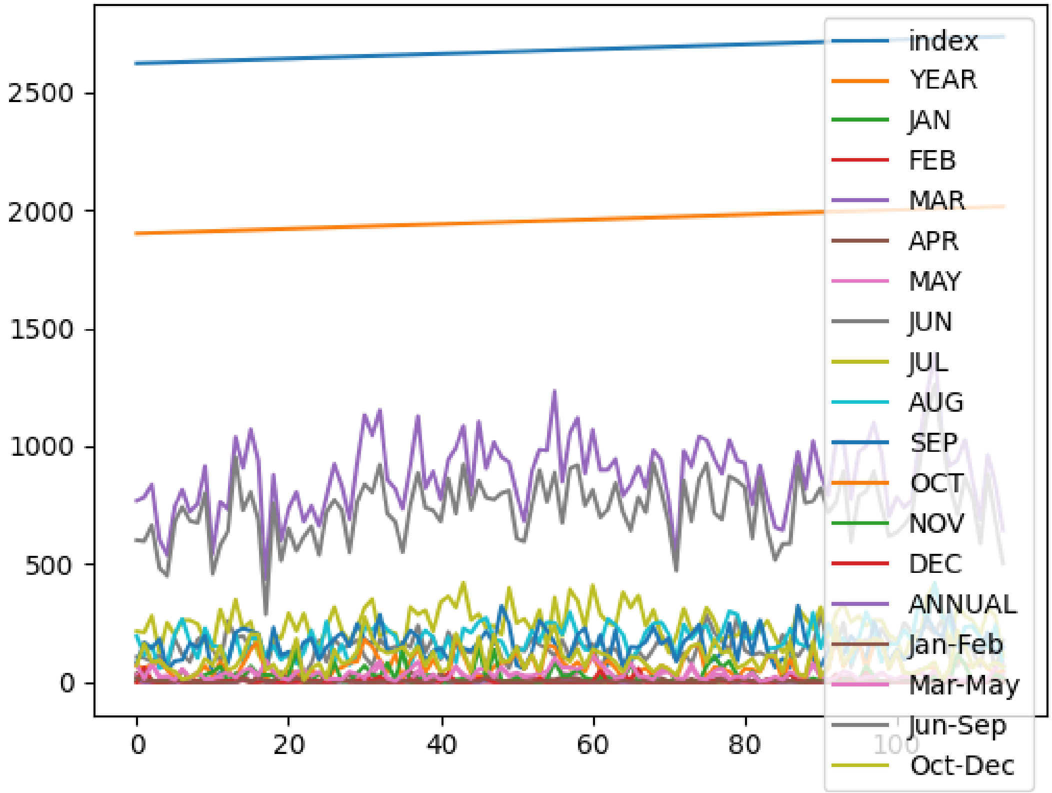
[6]:
```
array([<Axes: >, <Axes: >, <Axes: >, <Axes: >, <Axes: >, <Axes: >,
       <Axes: >, <Axes: >, <Axes: >, <Axes: >, <Axes: >, <Axes: >,
       <Axes: >, <Axes: >, <Axes: >, <Axes: >, <Axes: >, <Axes: >,
       <Axes: >], dtype=object)
```



5 Line chart

[7]:
```
df.plot.line()
```

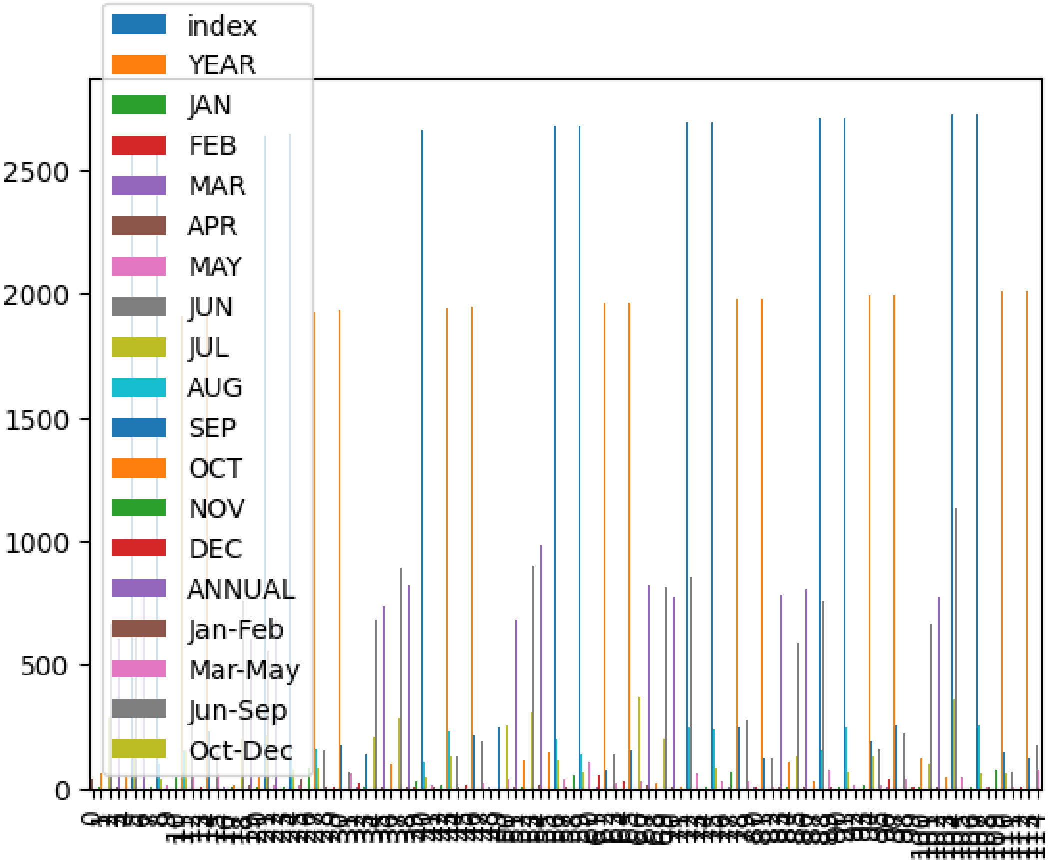
[7]:
```
<Axes: >
```



6 Bar chart

[8]:
```
df.plot.bar()
```

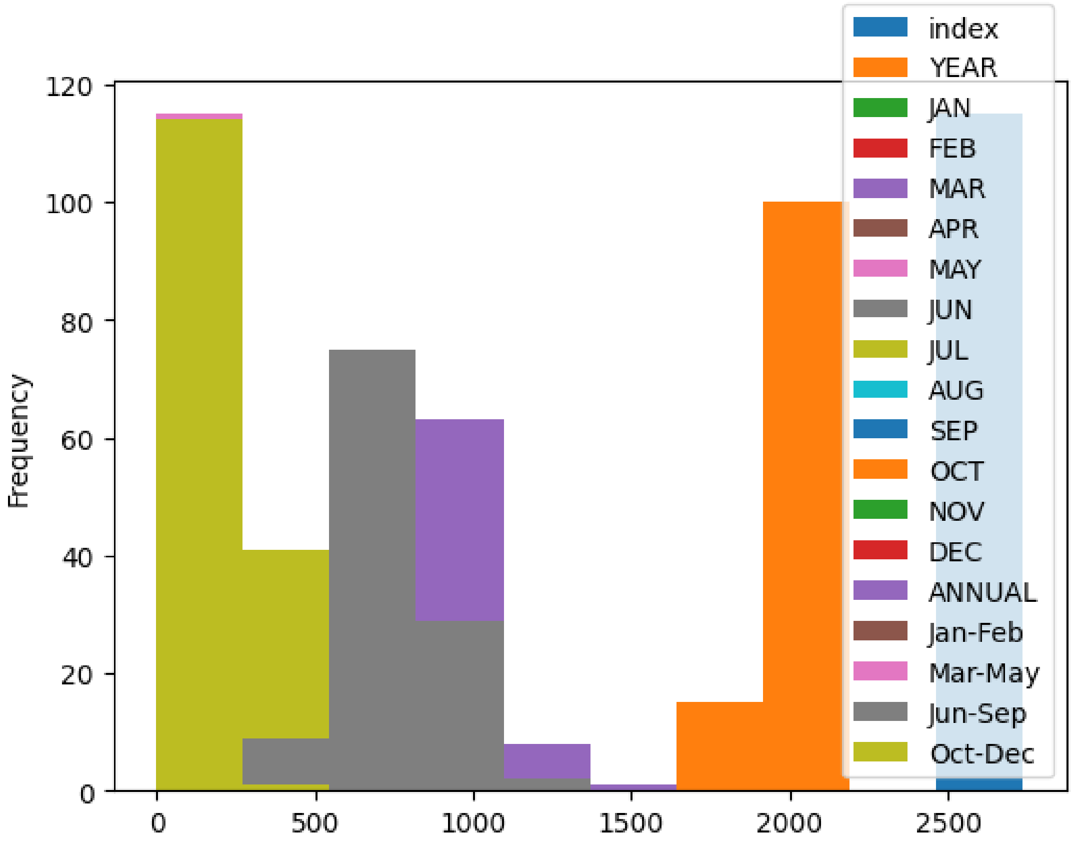
[8]:
```
<Axes: >
```



7 Histogram

```
[9]: df.plot.hist()
```

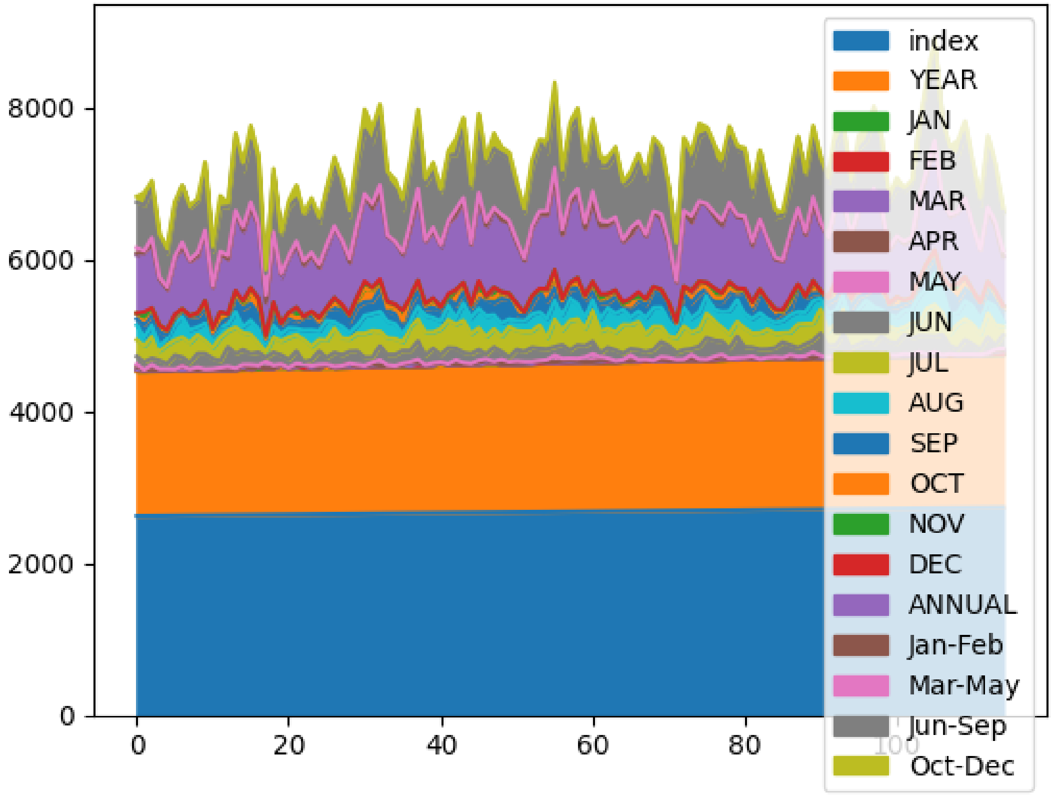
```
[9]: <Axes: ylabel='Frequency'>
```



8 Area chart

```
[10]: df.plot.area()
```

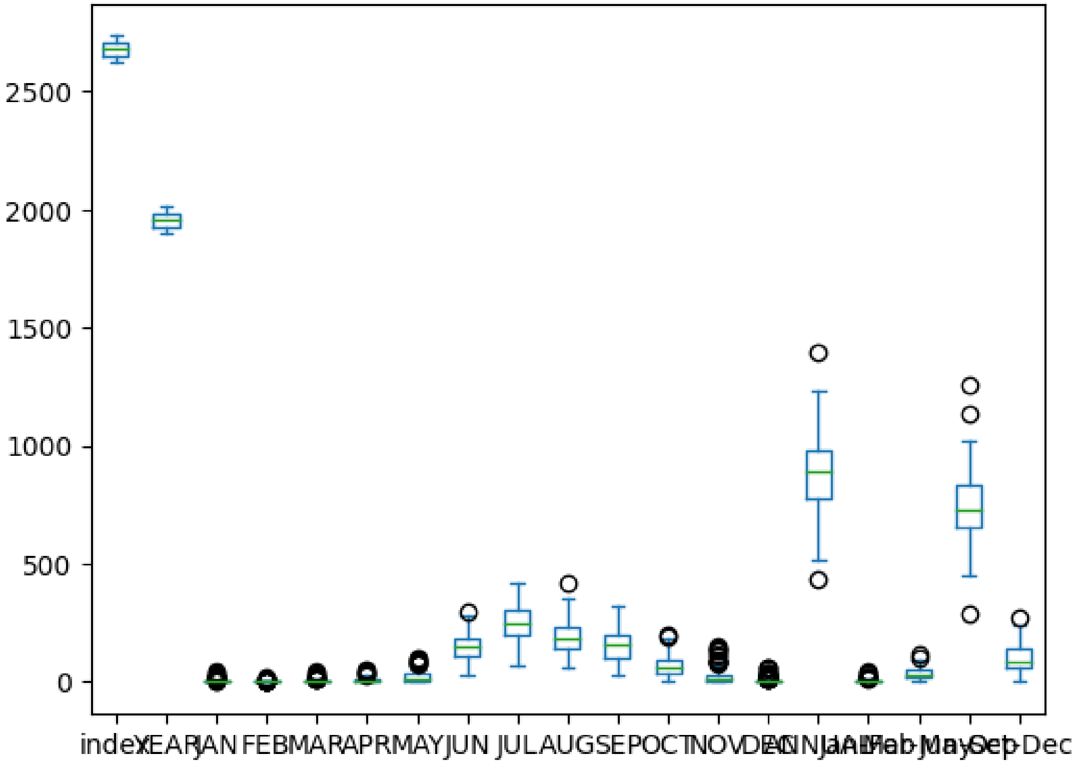
```
[10]: <Axes: >
```



9 Box chart

```
[11]: df.plot.box()
```

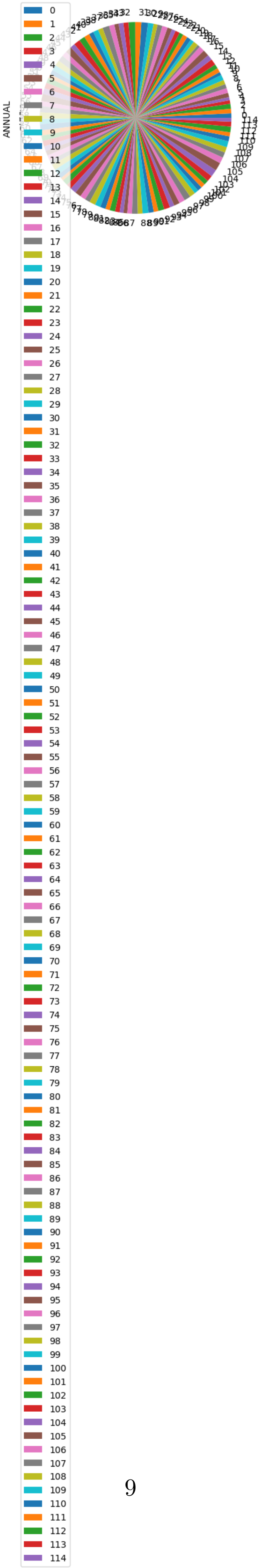
```
[11]: <Axes: >
```



10 Pie chart

```
[12]: df.plot.pie(y='ANNUAL' )
```

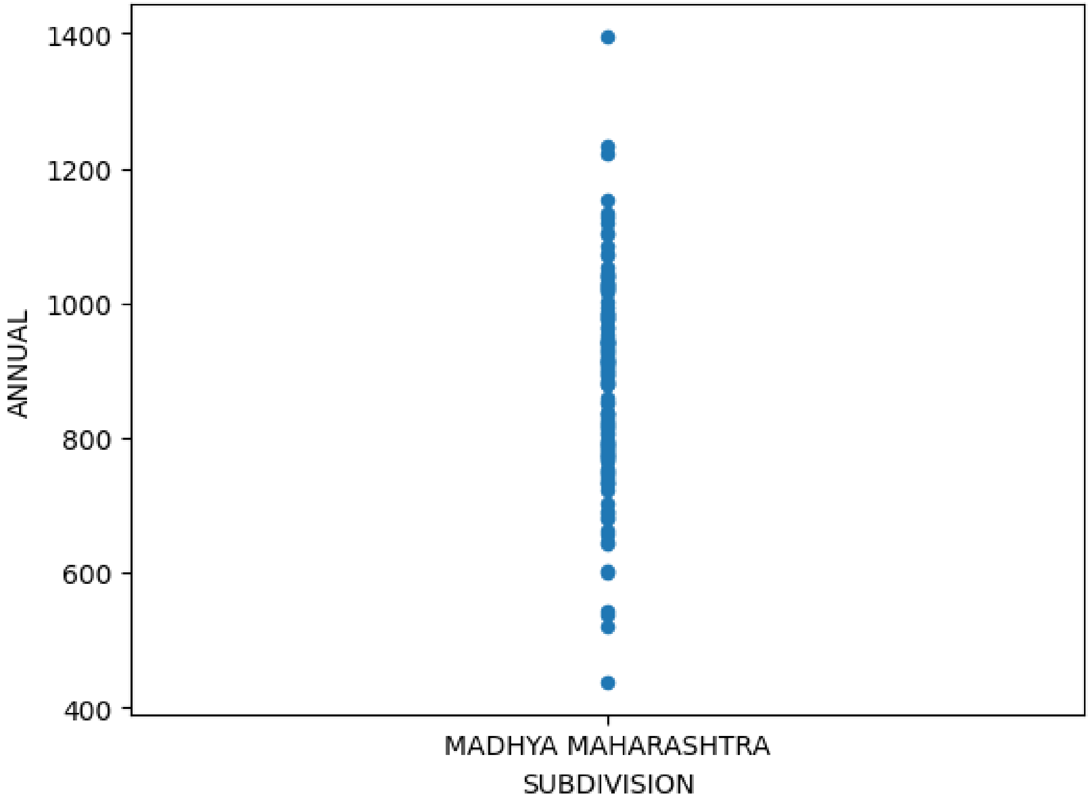
```
[12]: <Axes: ylabel='ANNUAL'>
```


ANNUAL
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114

11 Scatter chart

```
[13]: df.plot.scatter(x='SUBDIVISION' ,y='ANNUAL')
```

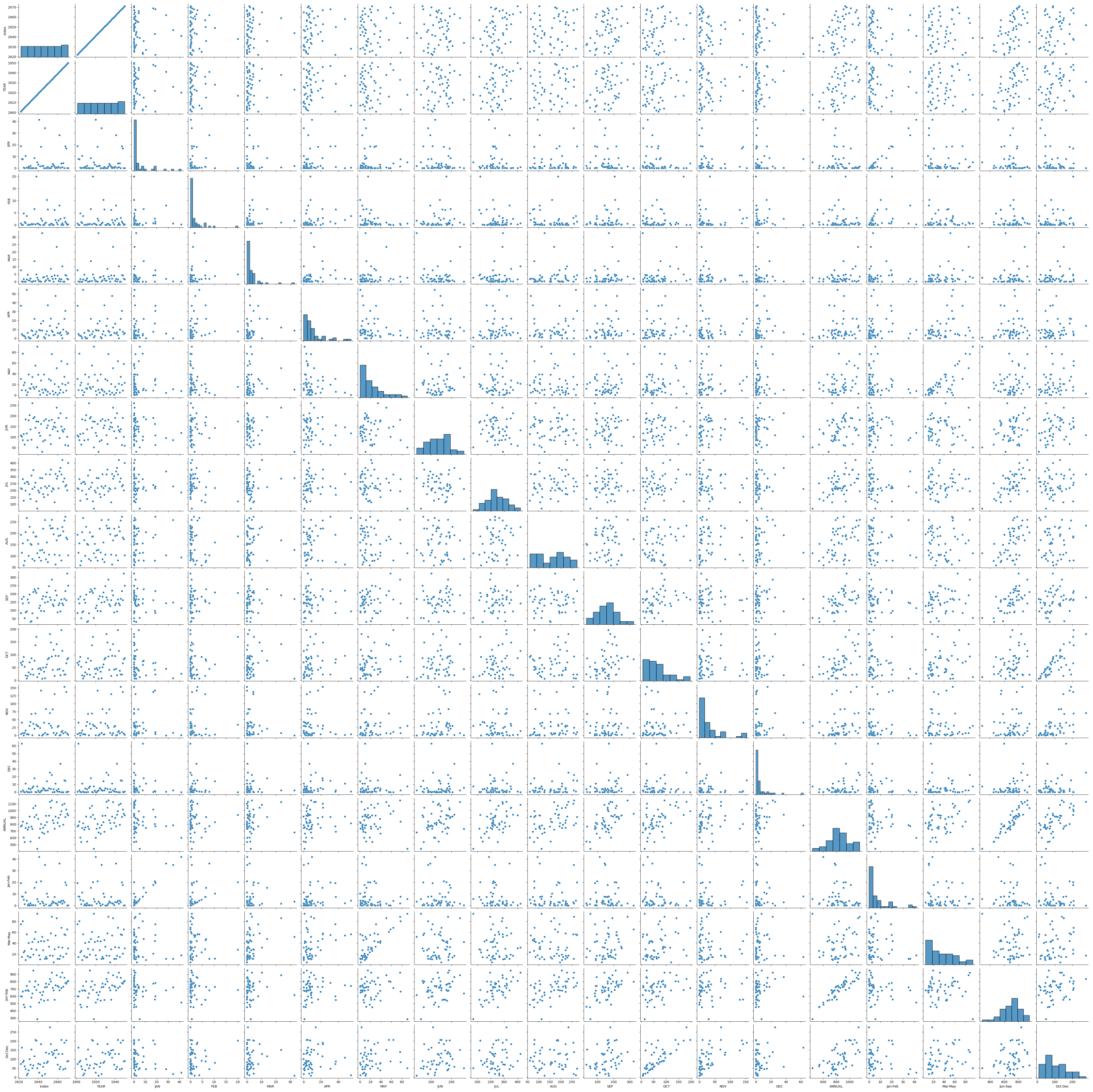
```
[13]: <Axes: xlabel='SUBDIVISION', ylabel='ANNUAL'>
```



12 Seaborn

```
[14]: sns.pairplot(df[0:50])
```

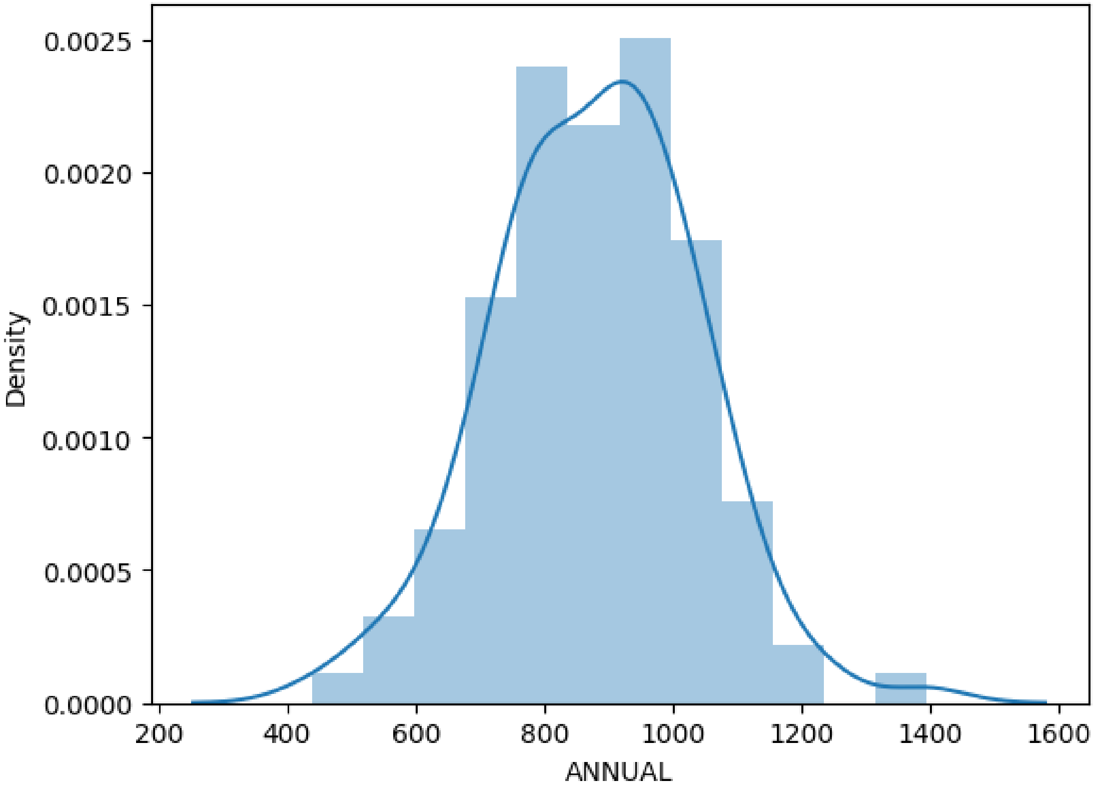
```
[14]: <seaborn.axisgrid.PairGrid at 0x7a7a278fffd0>
```



```
[15]: sns.distplot(df['ANNUAL'])
```

```
<ipython-input-15-5daa97052ca5>: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(df['ANNUAL'])
```

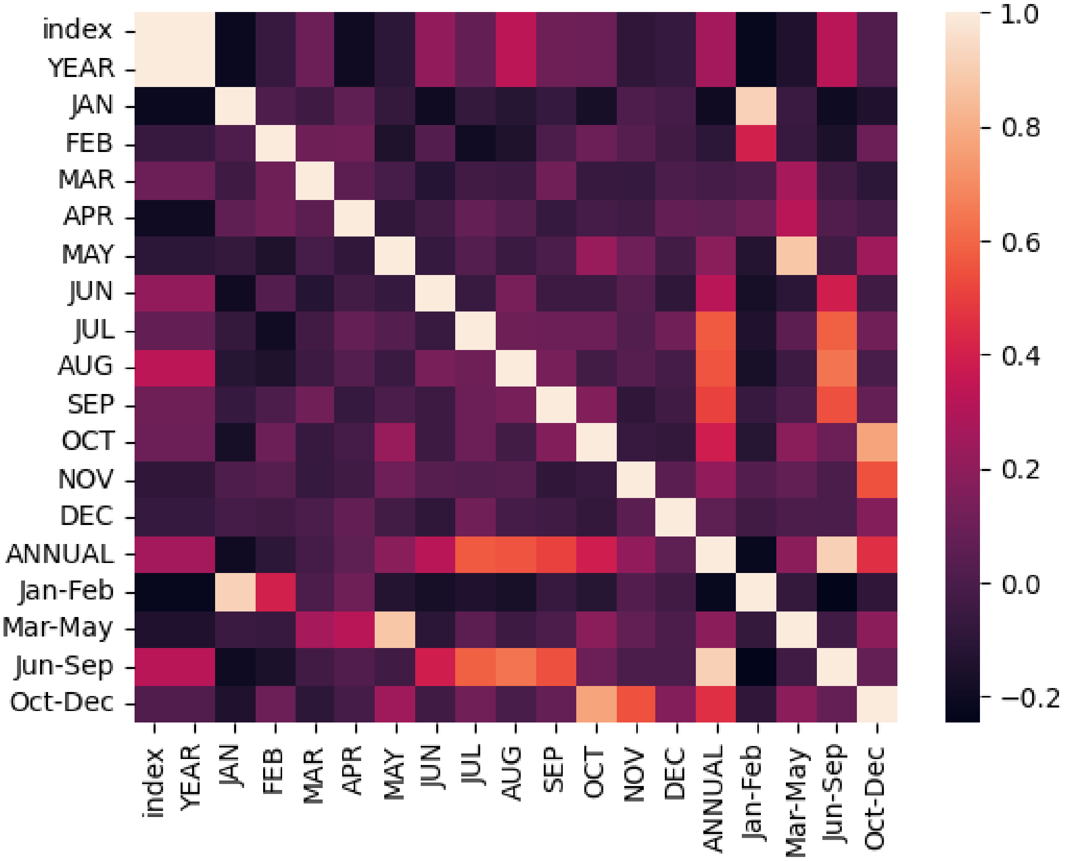
```
[15]: <Axes: xlabel='ANNUAL', ylabel='Density'>
```



```
[16]: sns.heatmap(df.corr())
```

```
<ipython-input-16-aa4f4450a243>:1: 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(df.corr())
```

```
[16]: <Axes: >
```


index
YEAR
JAN
FEB
MAR
APR
MAY
JUN
JUL
AUG
SEP
OCT
NOV
DEC
ANNUAL
Jan-Feb
Mar-May
Jun-Sep
Oct-Dec
1.0
0.8
0.6
0.4
0.2
0.0
−0.2