jgvrqscrl

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_kerala.csv")
df
```

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

[2]:
```
     index SUBDIVISION  YEAR   JAN   FEB   MAR    APR    MAY     JUN     JUL  \
0     3887      KERALA  1901  28.7  44.7  51.6  160.0  174.7   824.6   743.0
1     3888      KERALA  1902   6.7   2.6  57.3   83.9  134.5   390.9  1205.0
2     3889      KERALA  1903   3.2  18.6   3.1   83.6  249.7   558.6  1022.5
3     3890      KERALA  1904  23.7   3.0  32.2   71.5  235.7  1098.2   725.5
4     3891      KERALA  1905   1.2  22.3   9.4  105.9  263.3   850.2   520.5
..     …          …    …    …    …    …     …      …      …       …
110   3997      KERALA  2011  20.5  45.7  24.1  165.2  124.2   788.5   536.8
111   3998      KERALA  2012   7.4  11.0  21.0  171.1   95.3   430.3   362.6
112   3999      KERALA  2013   3.9  40.1  49.9   49.3  119.3  1042.7   830.2
113   4000      KERALA  2014   4.6  10.3  17.9   95.7  251.0   454.4   677.8
114   4001      KERALA  2015   3.1   5.8  50.1  214.1  201.8   563.6   406.0

       AUG    SEP    OCT    NOV    DEC  ANNUAL  Jan-Feb  Mar-May  Jun-Sep  \
0    357.5  197.7  266.9  350.8   48.4  3248.6     73.4    386.2   2122.8
1    315.8  491.6  358.4  158.3  121.5  3326.6      9.3    275.7   2403.4
2    420.2  341.8  354.1  157.0   59.0  3271.2     21.7    336.3   2343.0
3    351.8  222.7  328.1   33.9    3.3  3129.7     26.7    339.4   2398.2
4    293.6  217.2  383.5   74.4    0.2  2741.6     23.4    378.5   1881.5
..     …      …      …      …      …       …       …        …        …
110  492.7  391.2  227.2  169.7   49.5  3035.1     66.2    313.5   2209.1
```

```
111  501.6  241.1  187.5  112.9    9.4  2151.1     18.3    287.4   1535.6
112  369.7  318.6  259.9  154.9   17.0  3255.4     43.9    218.5   2561.2
113  733.9  298.8  355.5   99.5   47.2  3046.4     14.9    364.5   2164.8
114  252.2  292.9  308.1  223.6   79.4  2600.6      8.9    465.9   1514.7

     Oct-Dec
0      666.1
1      638.2
2      570.1
3      365.3
4      458.1
..       …
110    446.3
111    309.8
112    431.8
113    502.1
114    611.1

[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')
```

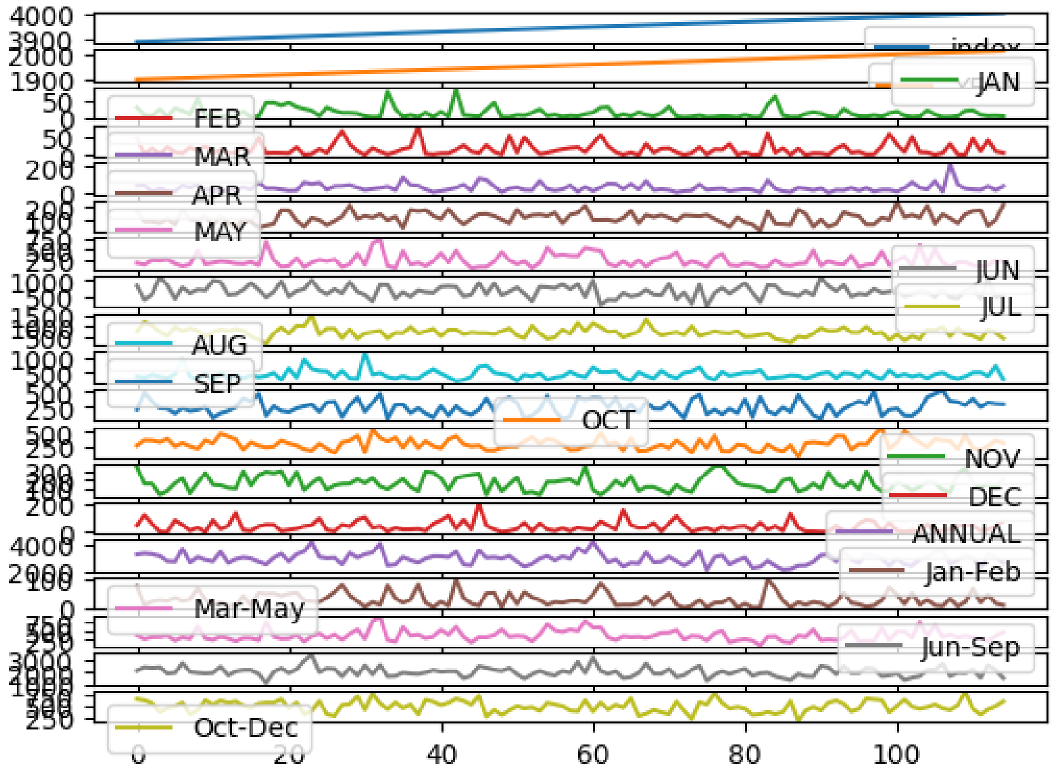
[5]:
```
df.info()
```

```
<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)
```

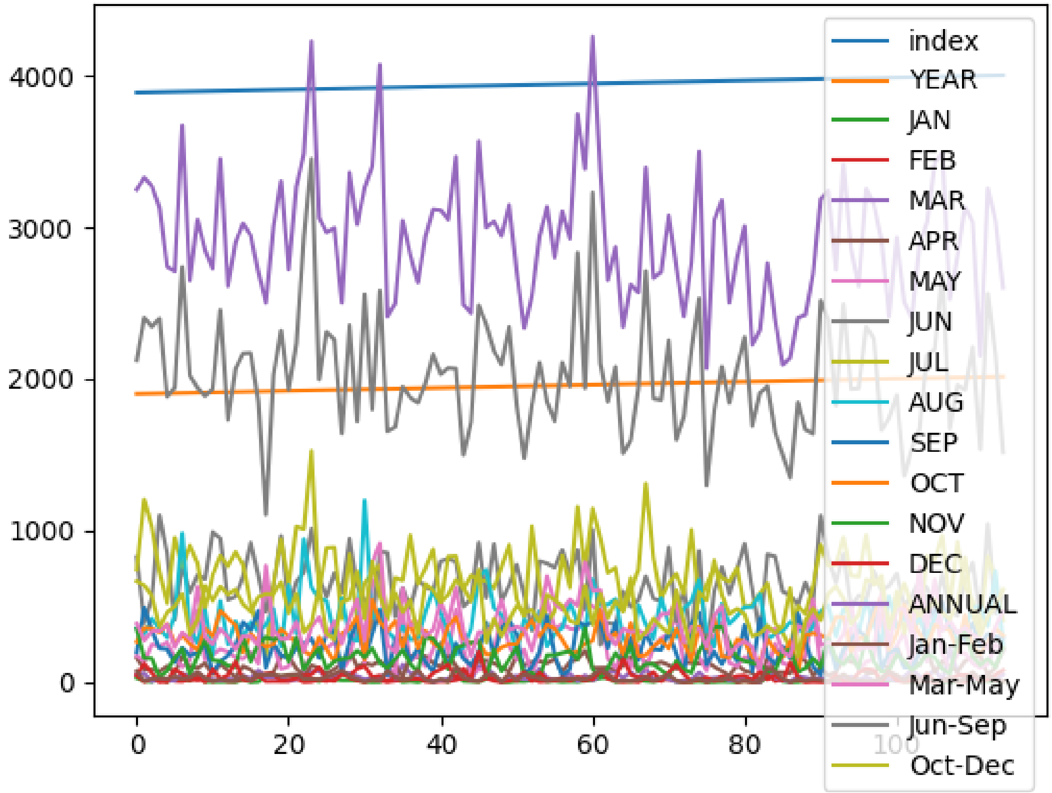
[6]:
```
array([<Axes: >, <Axes: >, <Axes: >, <Axes: >, <Axes: >, <Axes: >,
       <Axes: >, <Axes: >, <Axes: >, <Axes: >, <Axes: >, <Axes: >,
       <Axes: >, <Axes: >, <Axes: >, <Axes: >, <Axes: >, <Axes: >,
       <Axes: >], dtype=object)
```



5 Line chart

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

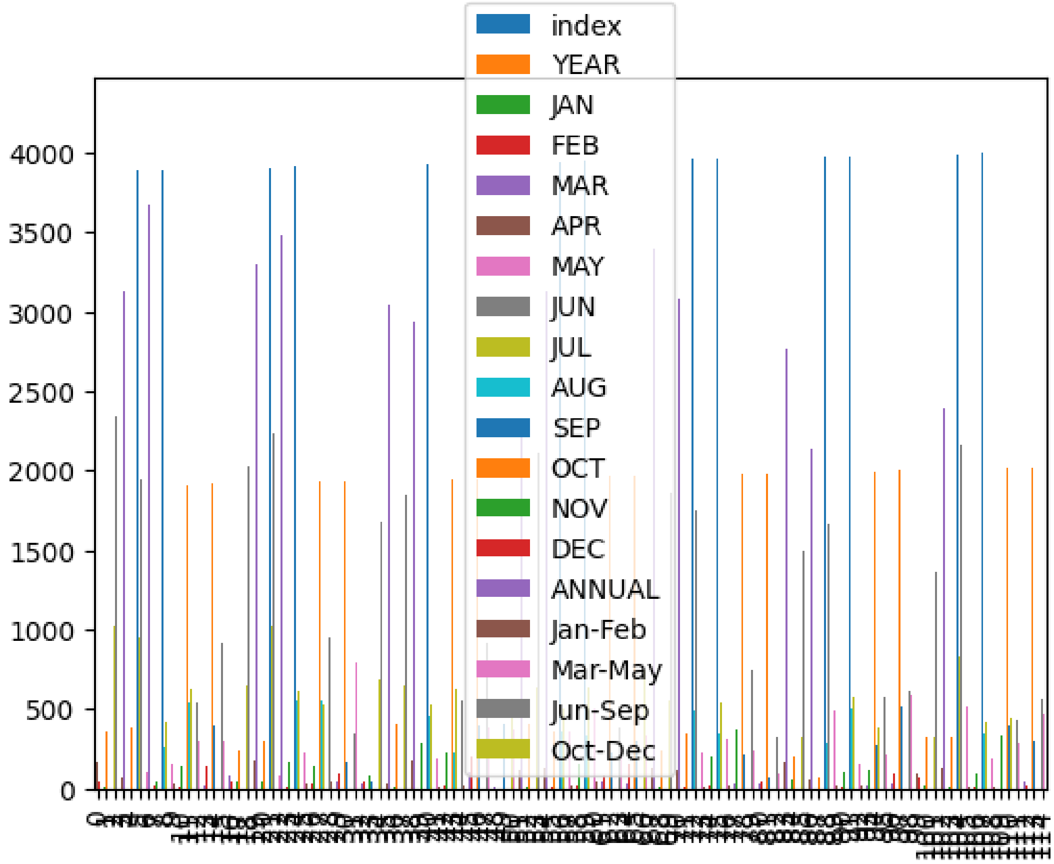
[7]:
```
<Axes: >
```



6 Bar chart

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

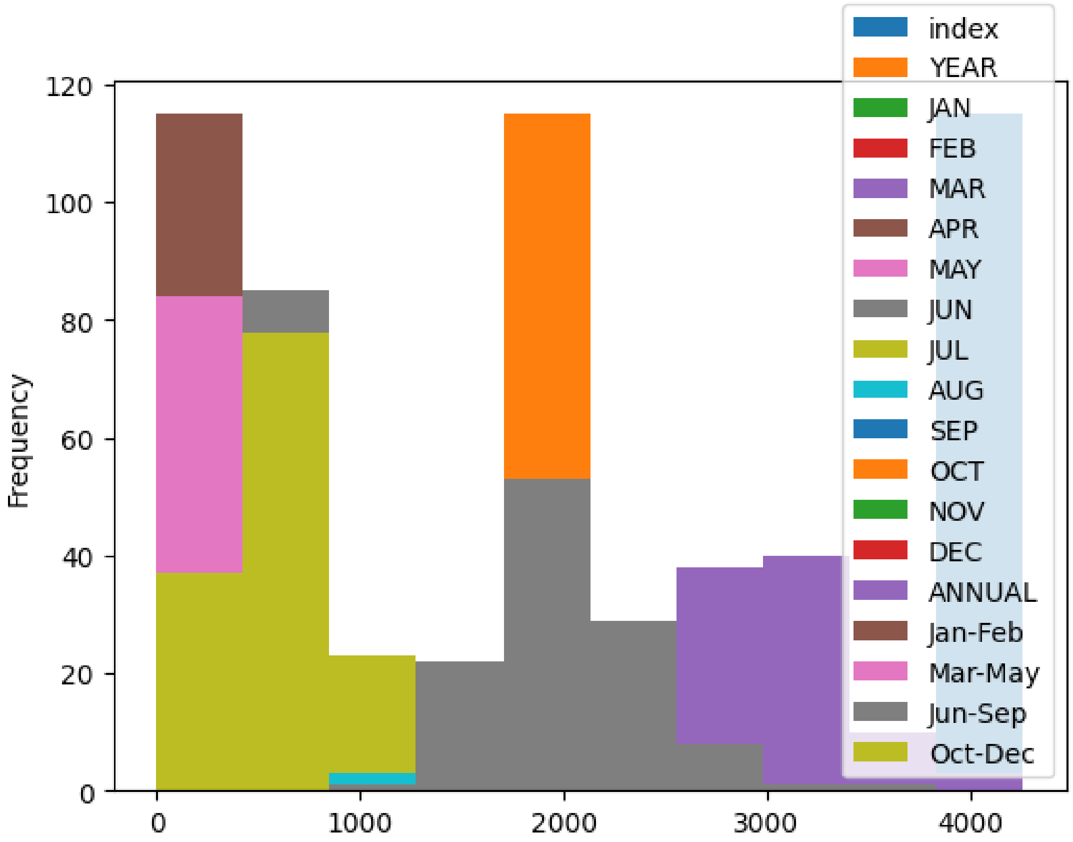
[8]:
```
<Axes: >
```



7 Histogram

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

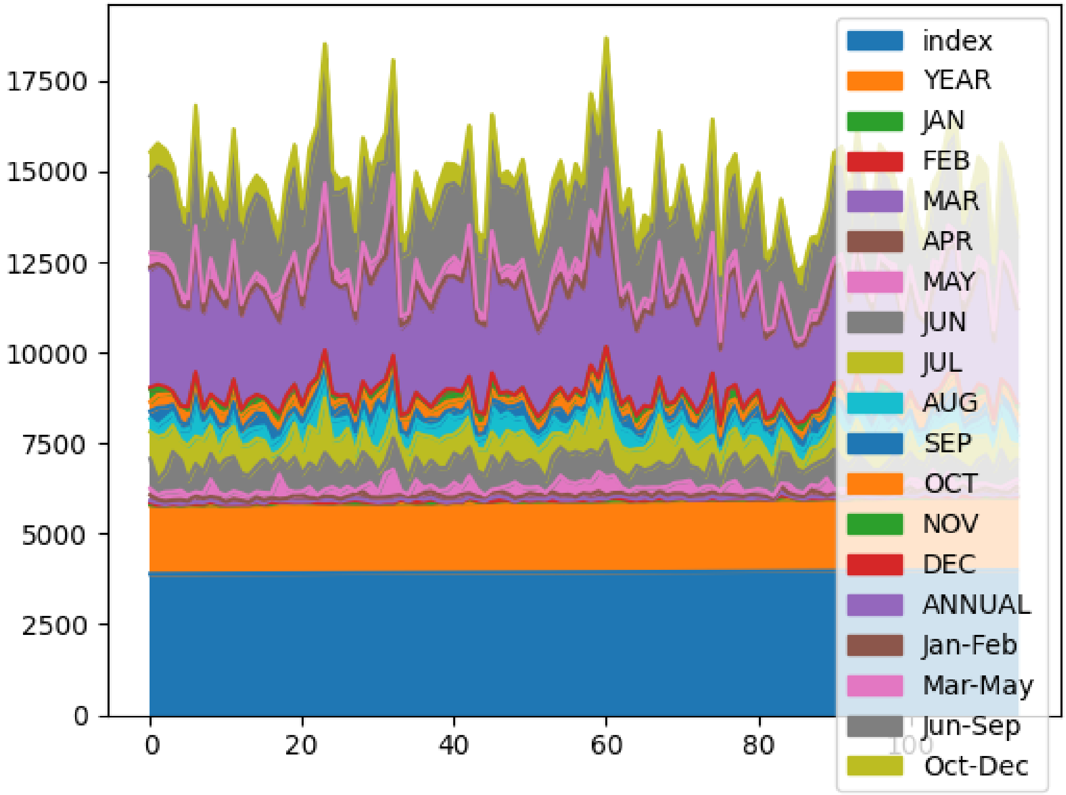
[9]:
```
<Axes: ylabel='Frequency'>
```



8 Area chart

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

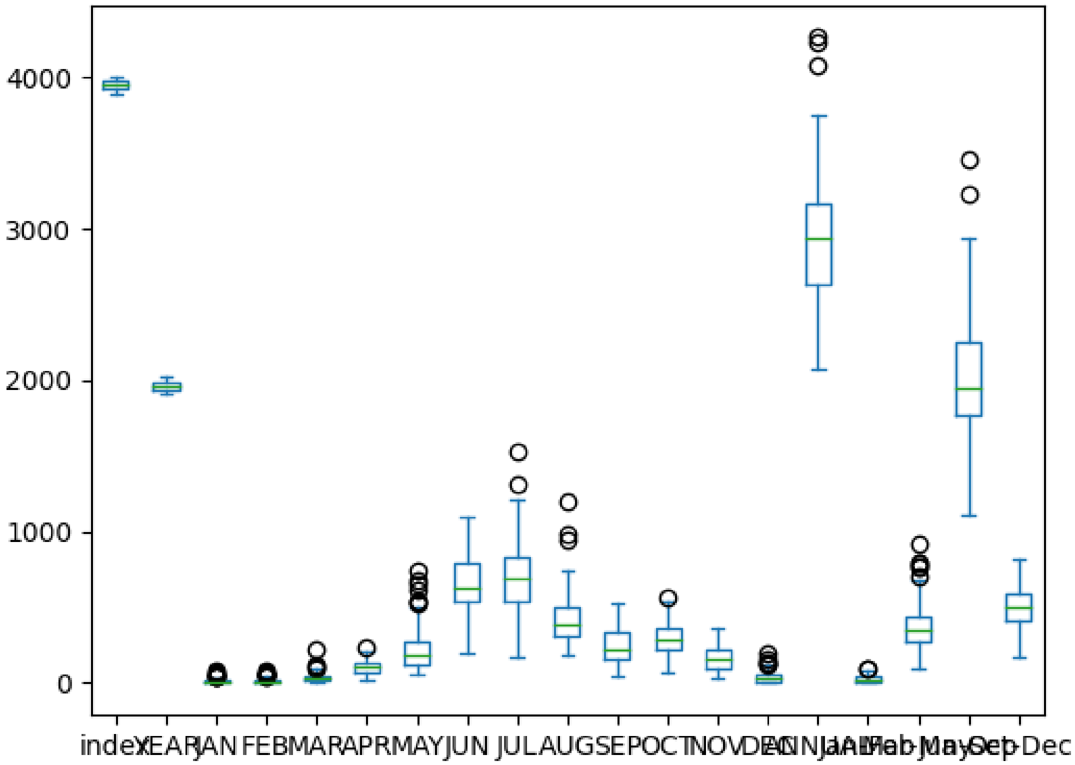
```
[10]: <Axes: >
```



9 Box chart

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

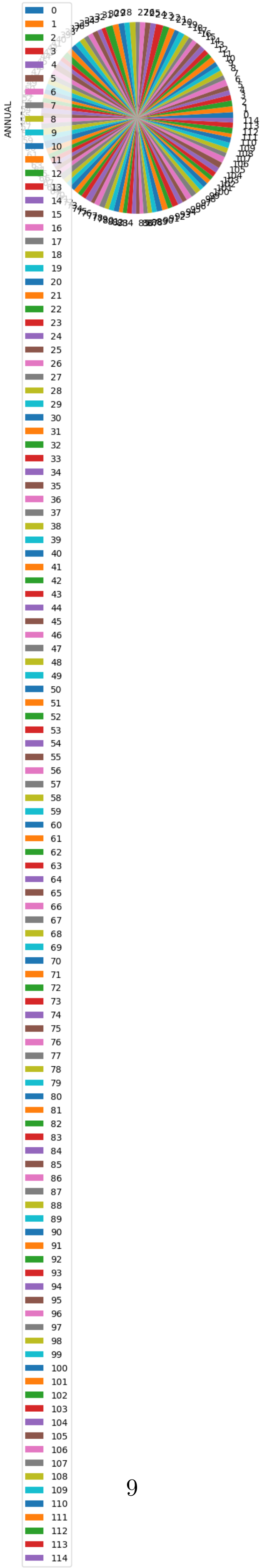
```
[11]: <Axes: >
```



10 Pie chart

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

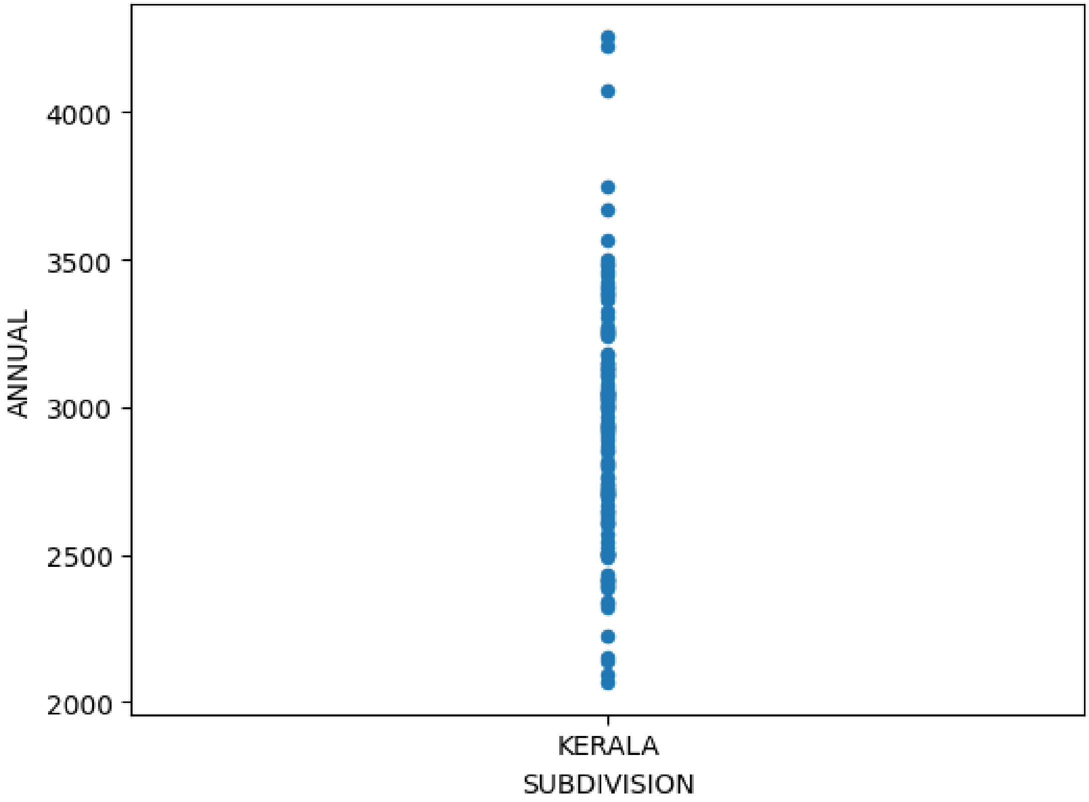
[12]:
```
<Axes: ylabel='ANNUAL'>
```


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')
```

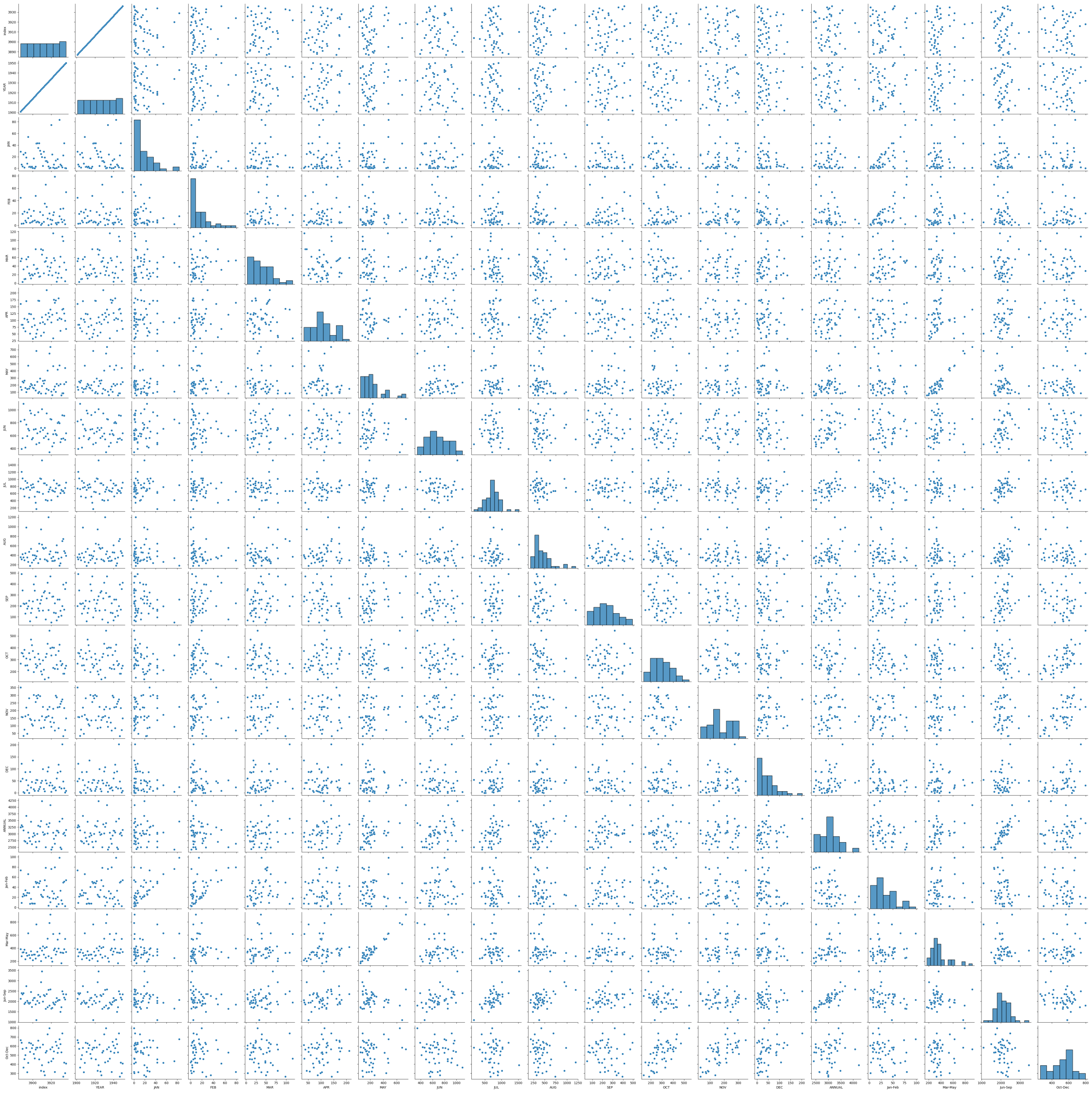
[13]:
```
<Axes: xlabel='SUBDIVISION', ylabel='ANNUAL'>
```



12 Seaborn

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

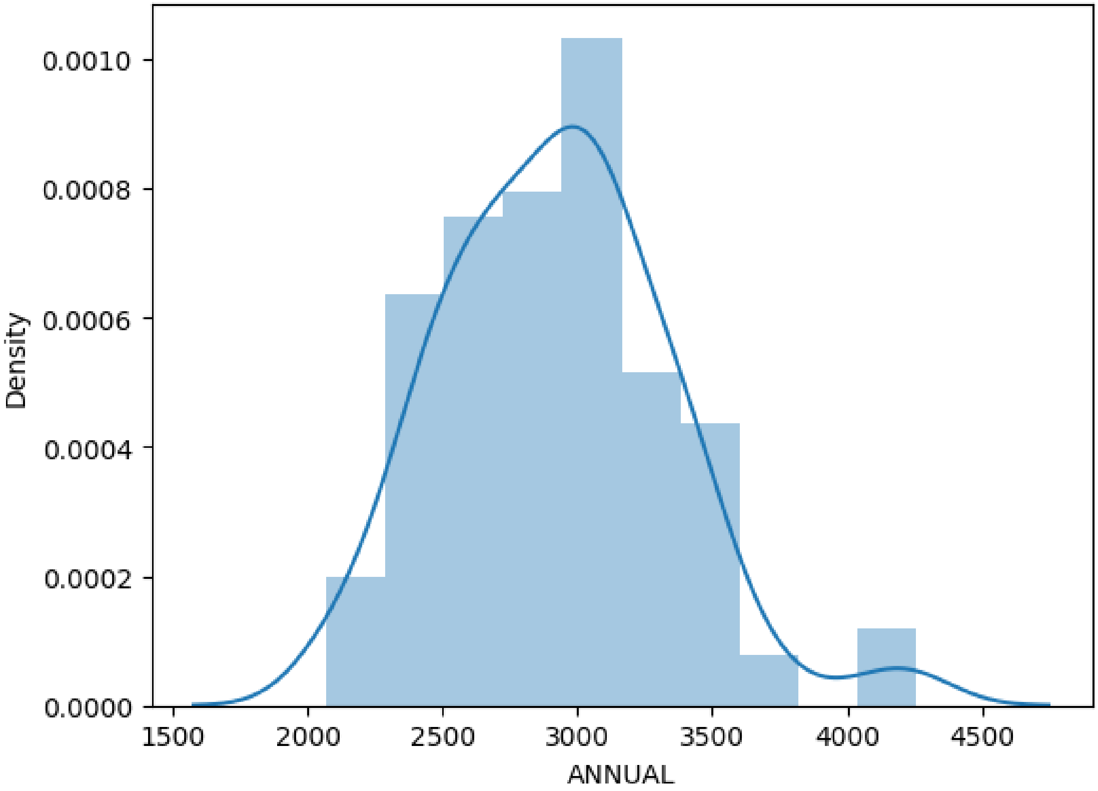
[14]:
```
<seaborn.axisgrid.PairGrid at 0x7c1a2bdc6a40>
```



[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'])
```

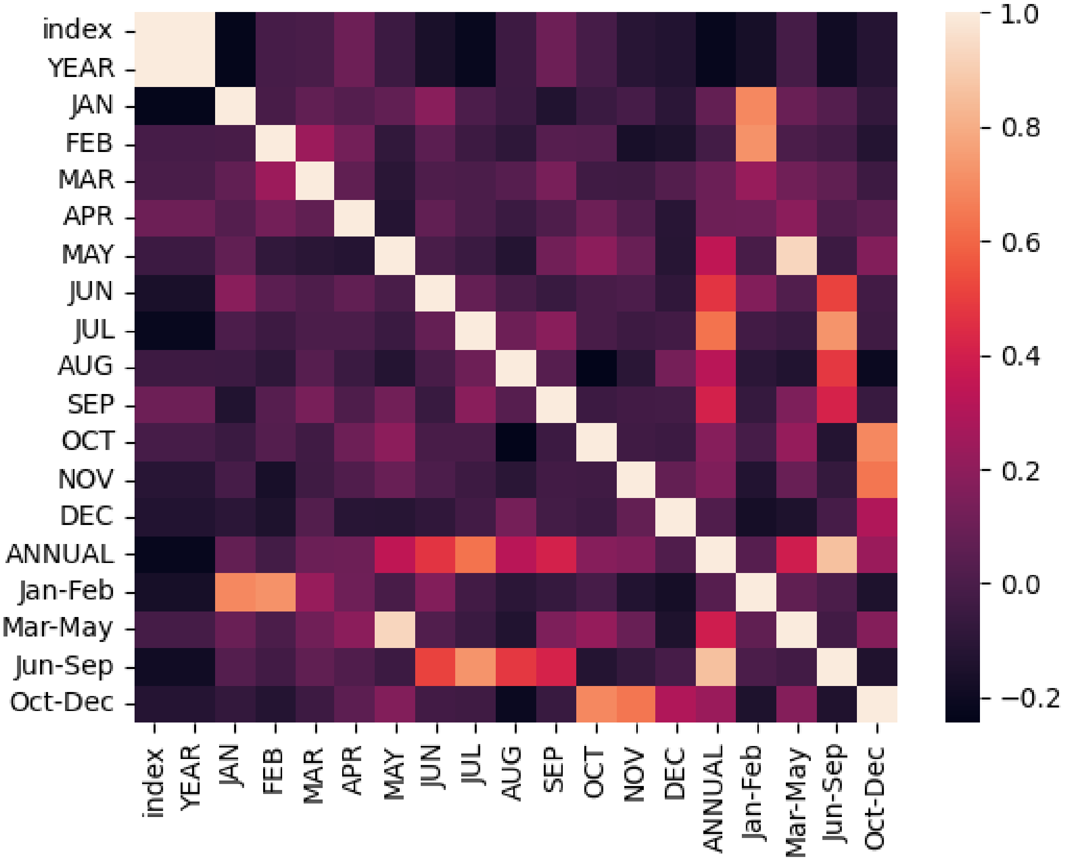
```
[15]: <Axes: xlabel='ANNUAL', ylabel='Density'>
```



```
[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