0ufpaixx8

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_coastal␣
 ↪karnataka.csv")
df
```

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

[2]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	\
0	3542	COASTAL KARNATAKA	1901	1.8	0.6	10.7	52.4	81.6	960.9	
1	3543	COASTAL KARNATAKA	1902	3.2	0.3	4.9	10.2	54.6	698.4	
2	3544	COASTAL KARNATAKA	1903	0.7	0.0	0.0	4.1	202.8	536.5	
3	3545	COASTAL KARNATAKA	1904	2.4	0.0	4.8	23.7	93.2	1108.2	
4	3546	COASTAL KARNATAKA	1905	0.0	0.2	0.0	6.4	83.1	767.3	
..	…	…	…	…	…	…	…	…	…	
110	3652	COASTAL KARNATAKA	2011	4.8	3.8	8.7	66.1	49.3	1018.4	
111	3653	COASTAL KARNATAKA	2012	NaN	11.4	5.1	77.0	22.9	650.9	
112	3654	COASTAL KARNATAKA	2013	2.4	19.6	19.0	28.5	100.4	1153.0	
113	3655	COASTAL KARNATAKA	2014	0.0	0.3	1.9	40.5	181.9	507.0	
114	3656	COASTAL KARNATAKA	2015	1.4	1.0	32.3	72.2	150.3	735.3	

	JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL	Jan-Feb	Mar-May	\
0	991.2	606.4	108.0	120.5	104.9	17.8	3056.9	2.4	144.7	
1	1401.6	454.2	708.4	180.4	50.8	132.2	3699.2	3.5	69.7	
2	1405.5	593.8	304.4	185.0	79.3	5.3	3317.4	0.7	206.9	
3	1070.0	465.6	245.3	127.2	0.7	0.0	3141.1	2.4	121.7	
4	777.3	586.9	172.9	222.2	36.1	0.0	2652.3	0.2	89.4	
..	…	…	…	…	…	…	…	…	…	

```
110  1080.5   861.3  545.2  178.8   81.5   10.2  3908.6      8.6    124.1
111   754.6  1027.6  382.0  115.1   68.0    3.6     NaN      NaN    104.9
112  1515.3   680.2  379.1  265.1   56.9   10.0  4229.4     21.9    147.9
113  1155.4  1121.0  379.3  226.4   40.0   30.8  3684.4      0.3    224.3
114   930.9   575.2  260.3  208.5  124.2   14.3  3106.0      2.4    254.8

     Jun-Sep  Oct-Dec
0     2666.6    243.3
1     3262.6    363.5
2     2840.1    269.6
3     2889.1    127.9
4     2304.3    258.4
..       ...      ...
110   3505.4    270.5
111   2815.1    186.7
112   3727.6    332.0
113   3162.6    297.2
114   2501.8    347.1

[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()
```

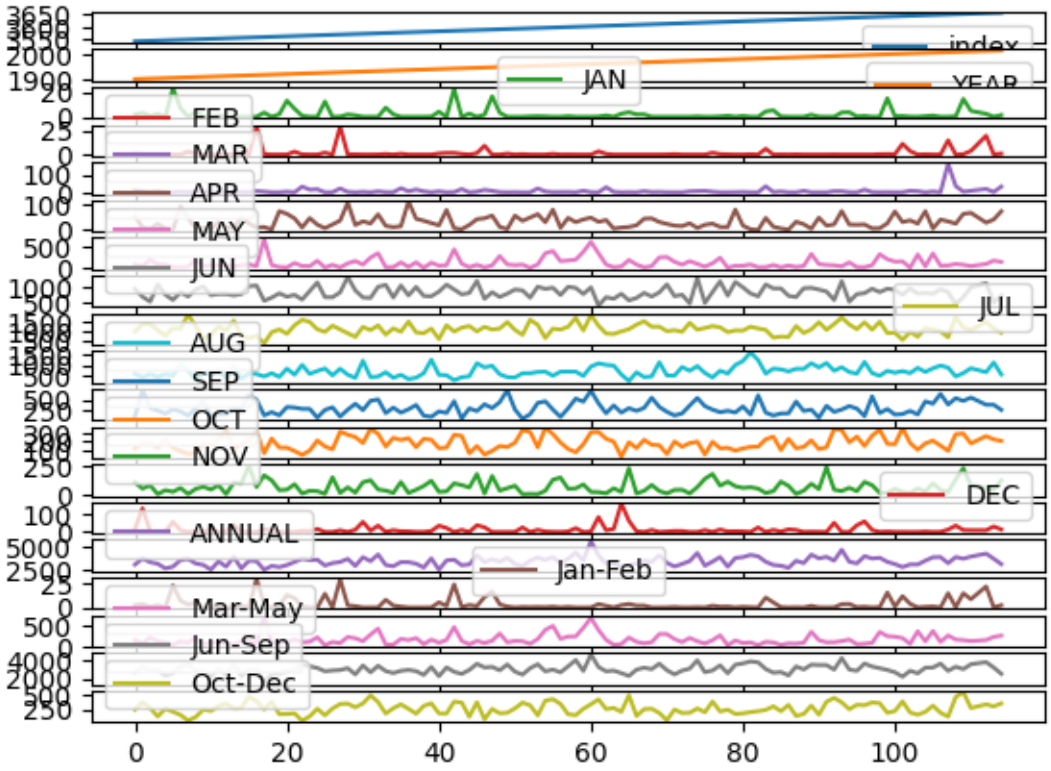
```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 114 entries, 0 to 114
Data columns (total 20 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   index        114 non-null    int64
 1   SUBDIVISION  114 non-null    object
 2   YEAR         114 non-null    int64
 3   JAN          114 non-null    float64
 4   FEB          114 non-null    float64
 5   MAR          114 non-null    float64
 6   APR          114 non-null    float64
```

```
 7   MAY          114 non-null    float64
 8   JUN          114 non-null    float64
 9   JUL          114 non-null    float64
 10  AUG          114 non-null    float64
 11  SEP          114 non-null    float64
 12  OCT          114 non-null    float64
 13  NOV          114 non-null    float64
 14  DEC          114 non-null    float64
 15  ANNUAL       114 non-null    float64
 16  Jan-Feb      114 non-null    float64
 17  Mar-May      114 non-null    float64
 18  Jun-Sep      114 non-null    float64
 19  Oct-Dec      114 non-null    float64
dtypes: float64(17), int64(2), object(1)
memory usage: 18.7+ KB
```

4 Line chart

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

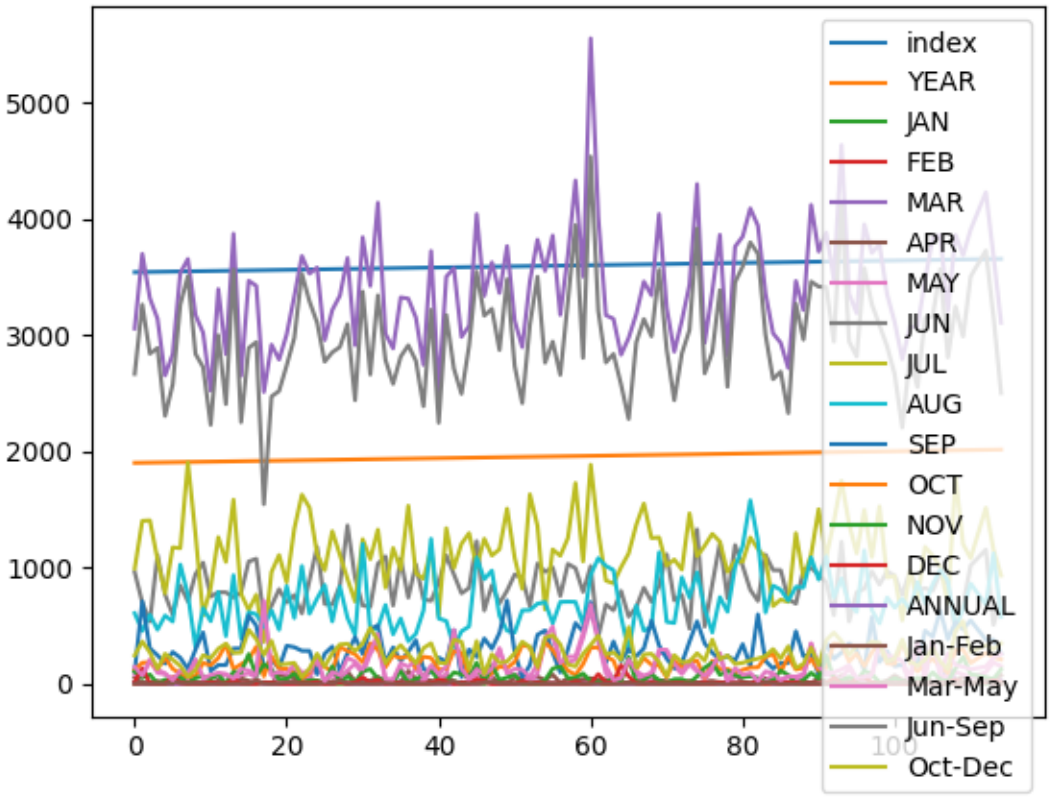
[6]:
```
array([<Axes: >, <Axes: >, <Axes: >, <Axes: >, <Axes: >, <Axes: >,
       <Axes: >, <Axes: >, <Axes: >, <Axes: >, <Axes: >, <Axes: >,
       <Axes: >, <Axes: >, <Axes: >, <Axes: >, <Axes: >, <Axes: >,
       <Axes: >], dtype=object)
```



5 Line chart

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

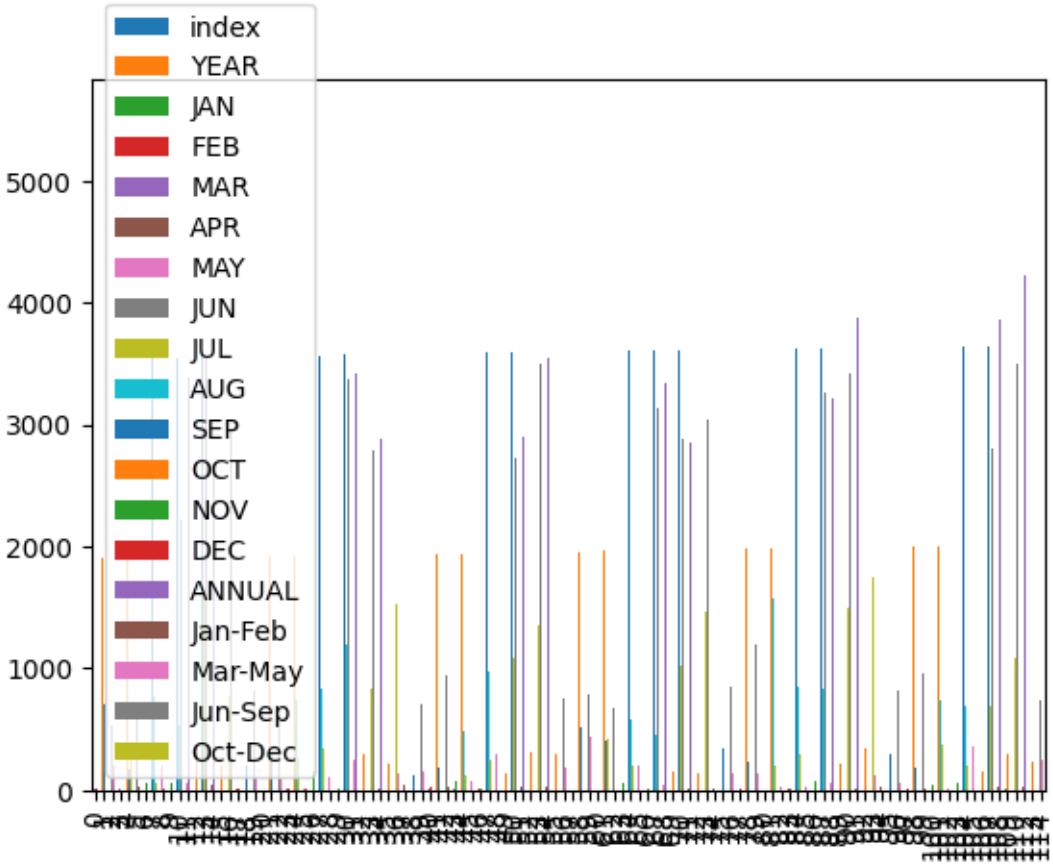
[7]: <Axes: >



6 Bar chart

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

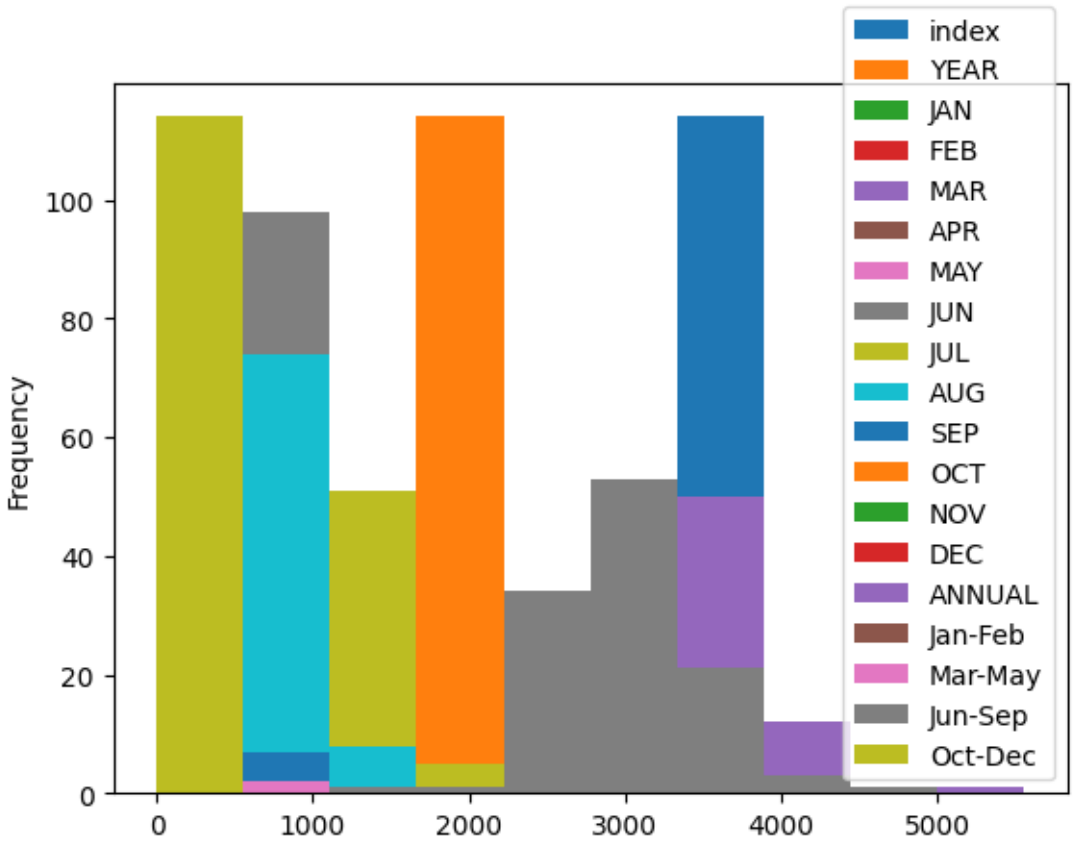
[8]: <Axes: >



7 Histogram

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

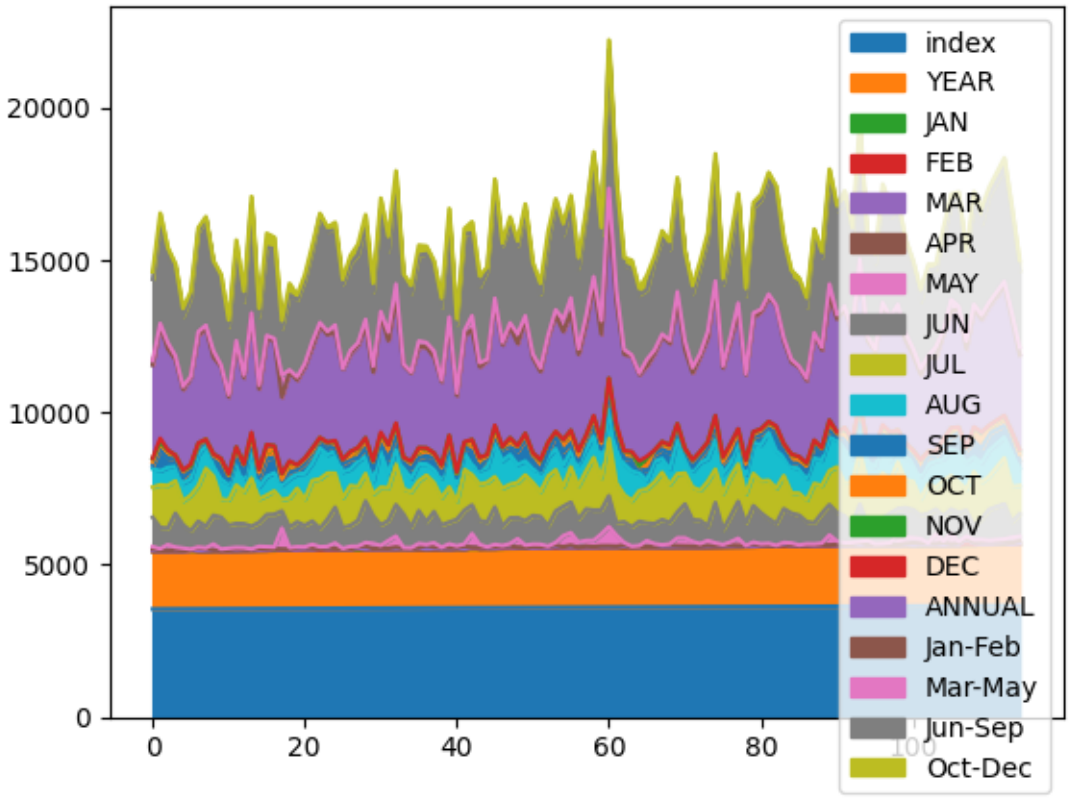
[9]: <Axes: ylabel='Frequency'>



8 Area chart

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

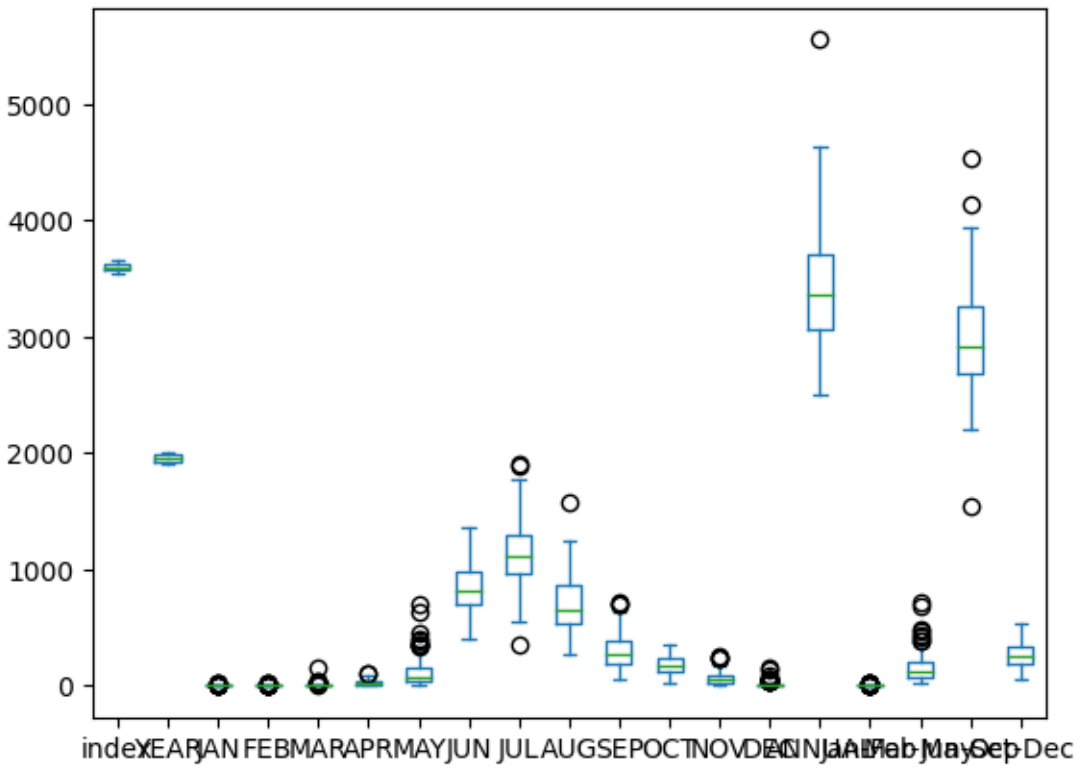
[10]: <Axes: >



9 Box chart

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

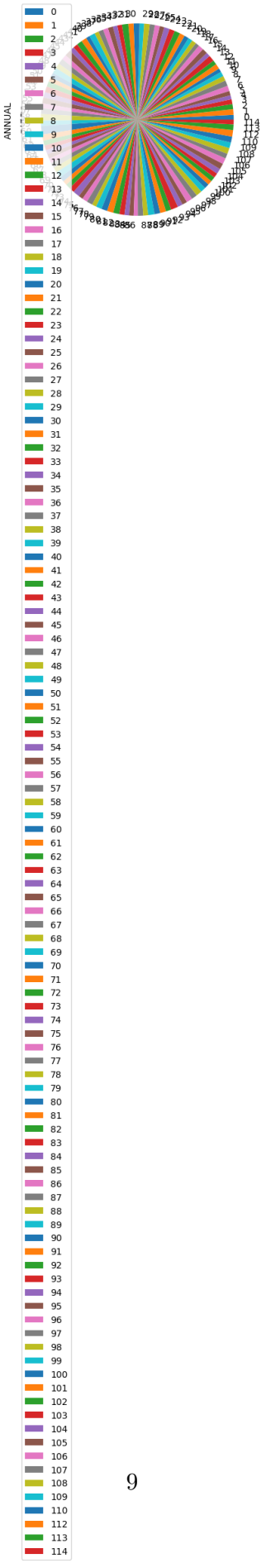
[11]: <Axes: >



10 Pie chart

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

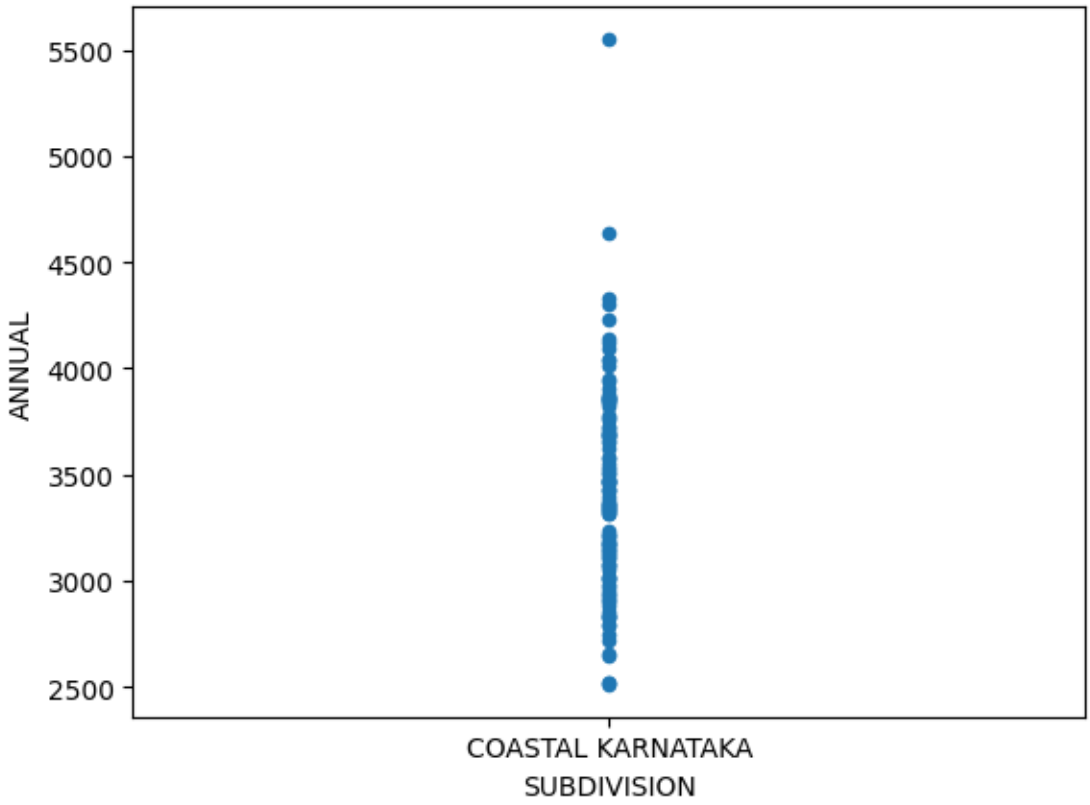
[12]:
```
<Axes: ylabel='ANNUAL'>
```


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
112
113
114

11 Scatter chart

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

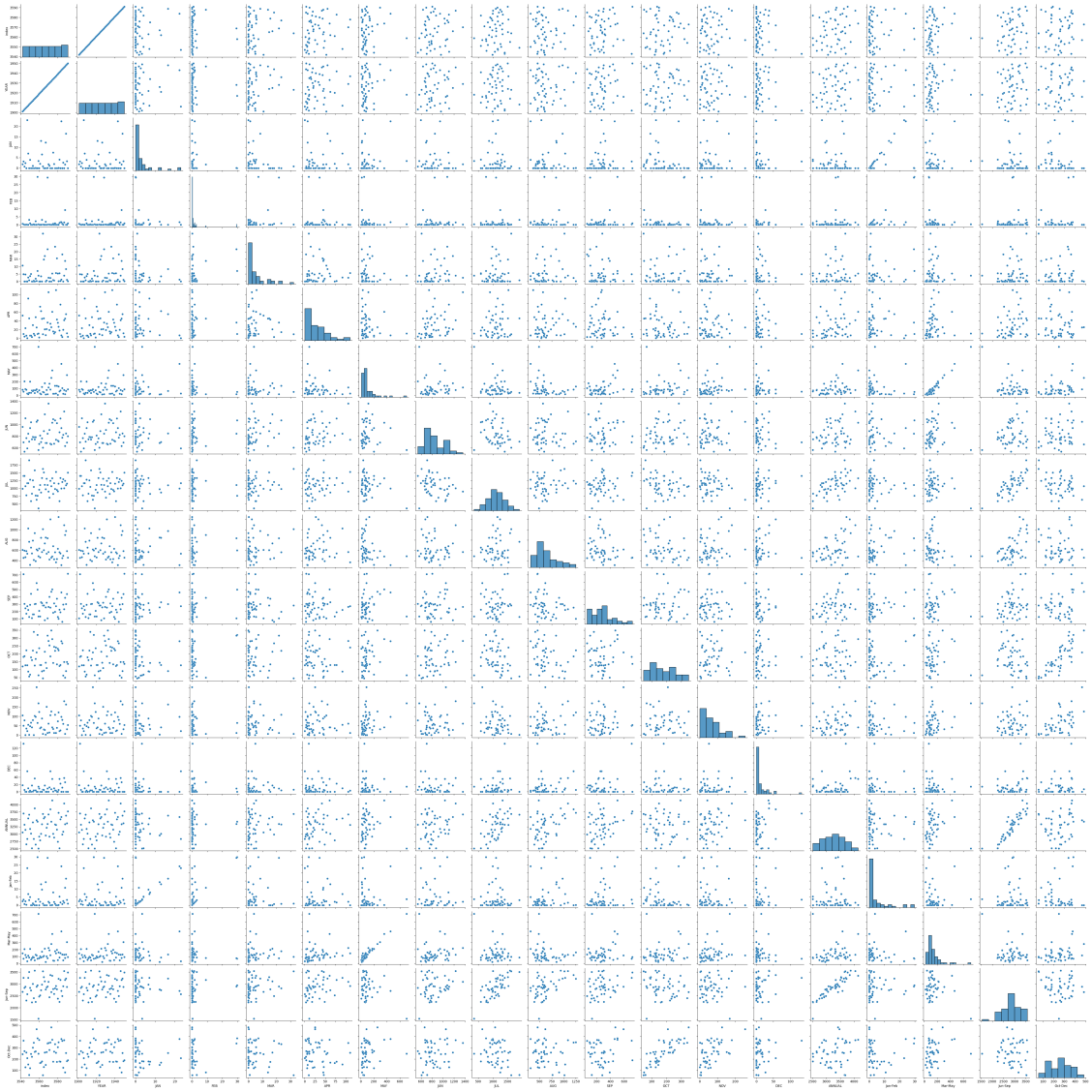
[13]:
```
<Axes: xlabel='SUBDIVISION', ylabel='ANNUAL'>
```



12 Seaborn

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

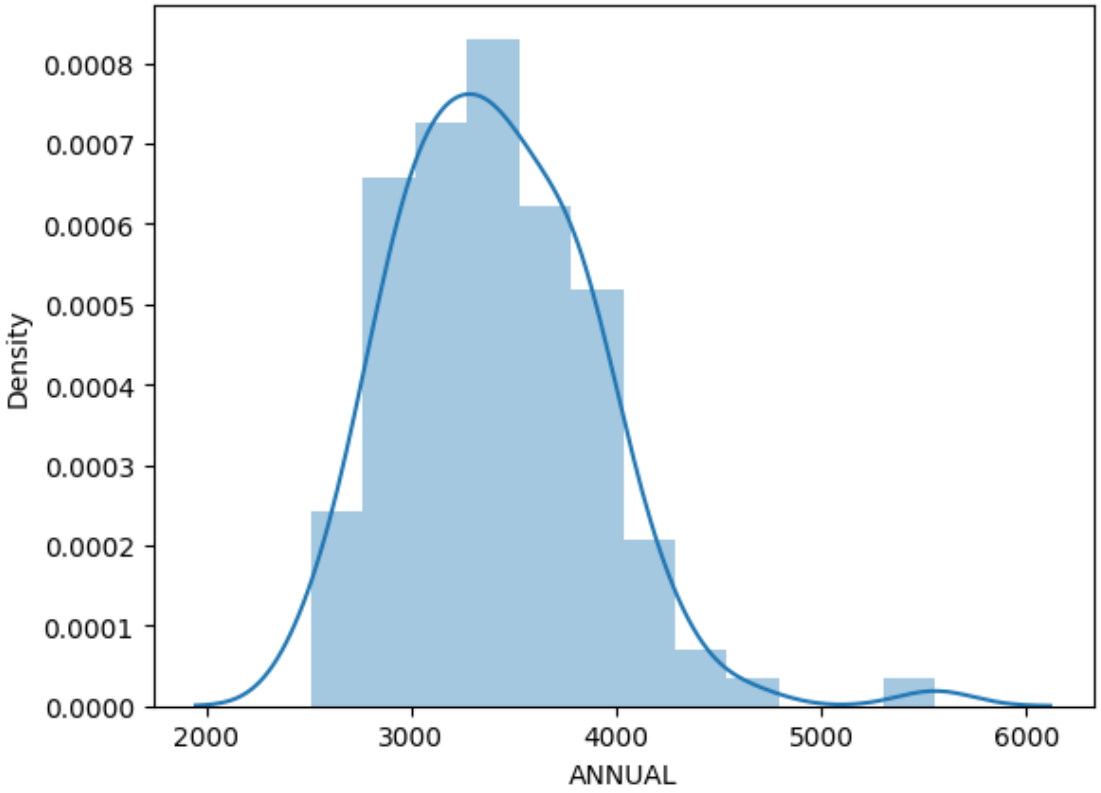
[14]:
```
<seaborn.axisgrid.PairGrid at 0x7e6d1efcc760>
```



[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'])
```

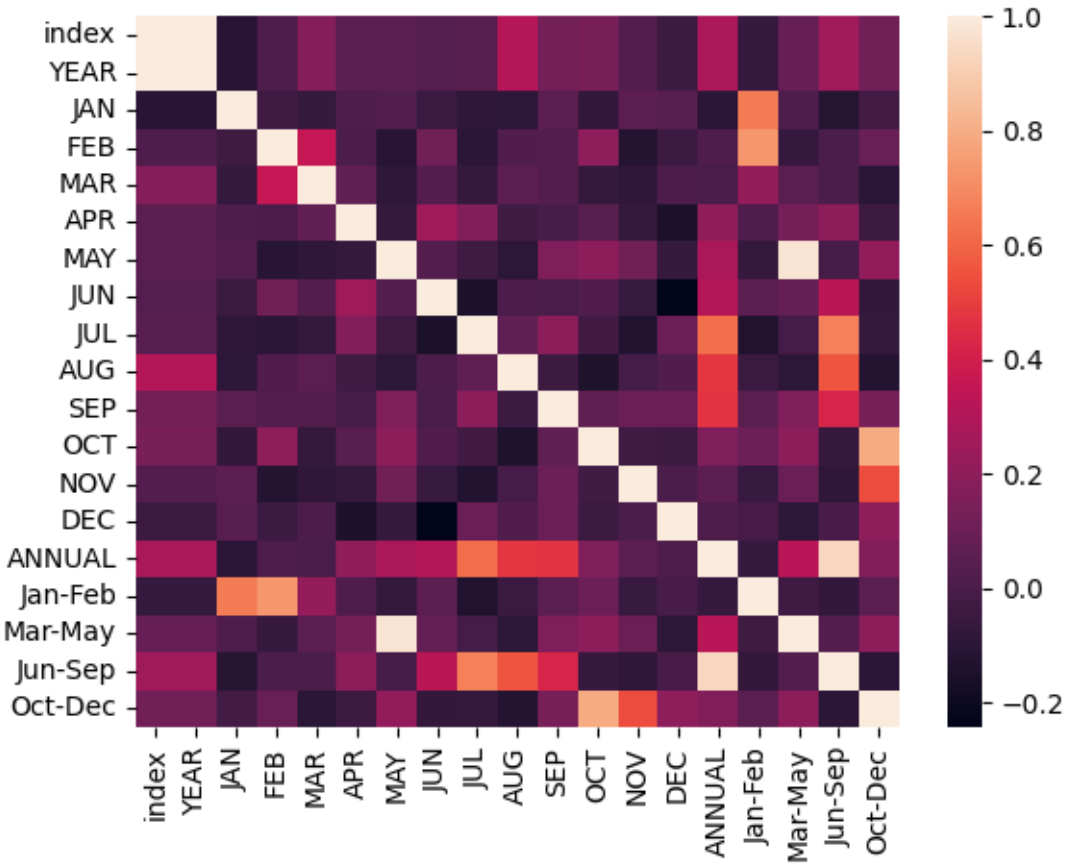
[15]: <Axes: xlabel='ANNUAL', ylabel='Density'>



[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