Importing Libraries

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

Importing Datasets

In [2]:
```
df=pd.read_csv(r'C:\Users\user\Downloads\Rainfall\LAKSHADWEEP.csv')
df
```

Out[2]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	O
0	4003	LAKSHADWEEP	1902	99.3	9.6	32.6	40.4	179.1	374.2	413.3	170.0	214.3	38
1	4004	LAKSHADWEEP	1903	63.5	95.0	0.0	29.5	144.1	212.4	261.8	202.0	292.1	7
2	4005	LAKSHADWEEP	1904	0.0	0.0	13.5	13.2	143.3	261.3	256.0	38.9	219.9	15
3	4006	LAKSHADWEEP	1905	62.4	0.0	0.0	0.0	166.7	400.7	68.7	377.5	107.5	23
4	4007	LAKSHADWEEP	1906	17.8	0.0	24.4	33.8	213.0	465.0	348.6	260.5	25.9	25
...	...	...	...	...	...	...	...	...	...	...	...	...	
108	4111	LAKSHADWEEP	2011	5.1	2.8	3.1	85.9	107.2	153.6	350.2	254.0	255.2	11
109	4112	LAKSHADWEEP	2012	19.2	0.1	1.6	76.8	21.2	327.0	231.5	381.2	179.8	14
110	4113	LAKSHADWEEP	2013	26.2	34.4	37.5	5.3	88.3	426.2	296.4	154.4	180.0	7
111	4114	LAKSHADWEEP	2014	53.2	16.1	4.4	14.9	57.4	244.1	116.1	466.1	132.2	16
112	4115	LAKSHADWEEP	2015	2.2	0.5	3.7	87.1	133.1	296.6	257.5	146.4	160.4	16

113 rows × 20 columns

Data Cleaning and Data Preprocessing

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

Out[3]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	O
0	4003	LAKSHADWEEP	1902	99.3	9.6	32.6	40.4	179.1	374.2	413.3	170.0	214.3	3
2	4005	LAKSHADWEEP	1904	0.0	0.0	13.5	13.2	143.3	261.3	256.0	38.9	219.9	1
3	4006	LAKSHADWEEP	1905	62.4	0.0	0.0	0.0	166.7	400.7	68.7	377.5	107.5	2
4	4007	LAKSHADWEEP	1906	17.8	0.0	24.4	33.8	213.0	465.0	348.6	260.5	25.9	2
5	4008	LAKSHADWEEP	1907	60.6	49.3	0.0	123.5	77.0	241.1	199.5	165.6	25.8	1
...	...	...	...	...	...	...	...	...	...	...	...	...	
108	4111	LAKSHADWEEP	2011	5.1	2.8	3.1	85.9	107.2	153.6	350.2	254.0	255.2	1
109	4112	LAKSHADWEEP	2012	19.2	0.1	1.6	76.8	21.2	327.0	231.5	381.2	179.8	1
110	4113	LAKSHADWEEP	2013	26.2	34.4	37.5	5.3	88.3	426.2	296.4	154.4	180.0	
111	4114	LAKSHADWEEP	2014	53.2	16.1	4.4	14.9	57.4	244.1	116.1	466.1	132.2	1
112	4115	LAKSHADWEEP	2015	2.2	0.5	3.7	87.1	133.1	296.6	257.5	146.4	160.4	1

102 rows × 20 columns

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

Out[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')
```

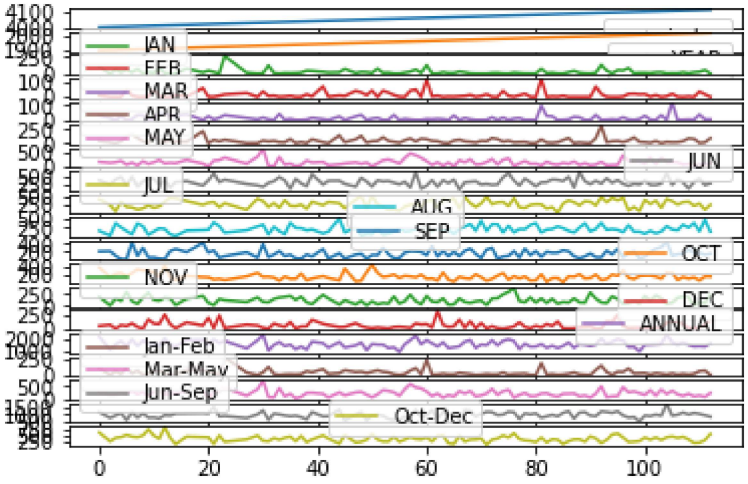
In [5]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 102 entries, 0 to 112
Data columns (total 20 columns):
 #   Column       Non-Null Count  Dtype  
---  ------       --------------  -----  
 0   index        102 non-null    int64  
 1   SUBDIVISION  102 non-null    object 
 2   YEAR         102 non-null    int64  
 3   JAN          102 non-null    float64
 4   FEB          102 non-null    float64
 5   MAR          102 non-null    float64
 6   APR          102 non-null    float64
 7   MAY          102 non-null    float64
 8   JUN          102 non-null    float64
 9   JUL          102 non-null    float64
 10  AUG          102 non-null    float64
 11  SEP          102 non-null    float64
 12  OCT          102 non-null    float64
 13  NOV          102 non-null    float64
 14  DEC          102 non-null    float64
 15  ANNUAL       102 non-null    float64
 16  Jan-Feb      102 non-null    float64
 17  Mar-May      102 non-null    float64
 18  Jun-Sep      102 non-null    float64
 19  Oct-Dec      102 non-null    float64
dtypes: float64(17), int64(2), object(1)
memory usage: 16.7+ KB
```

Line Chart

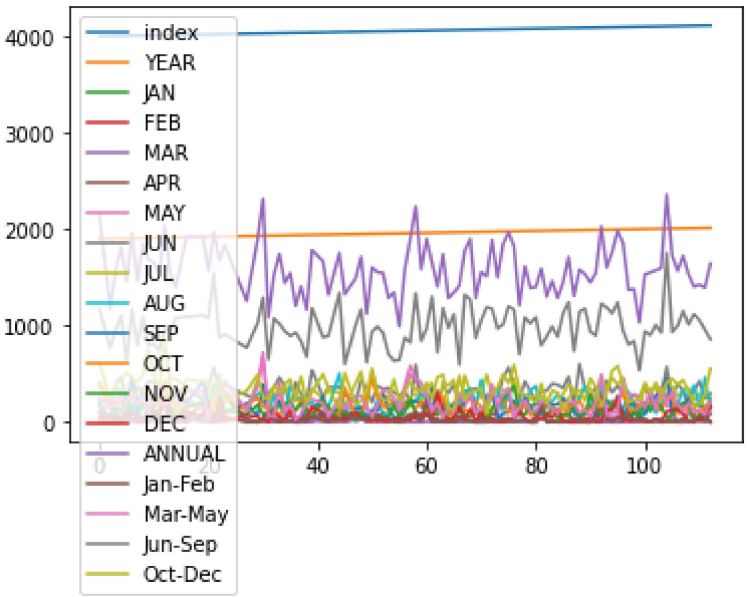
In [6]:
```
df.plot.line(subplots=True)
```

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



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

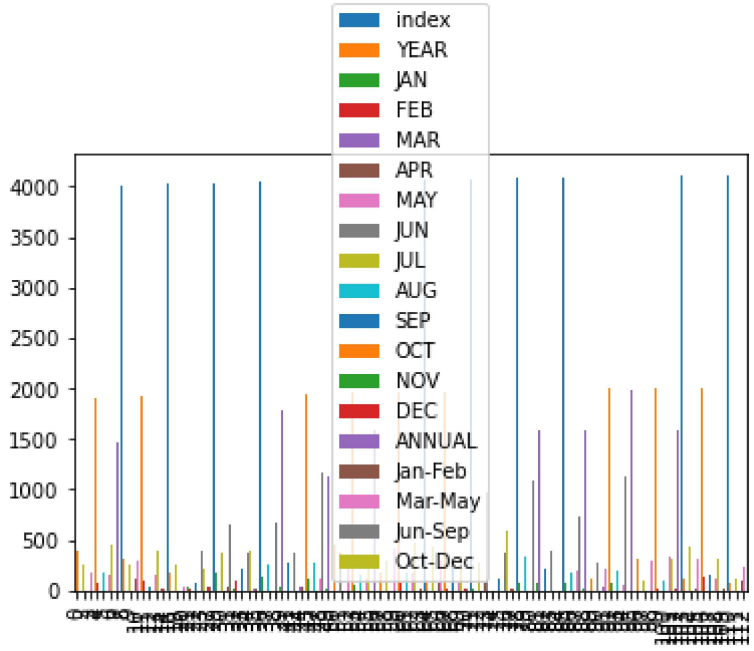
Out[7]:
```
<AxesSubplot:>
```



Bar Chart

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

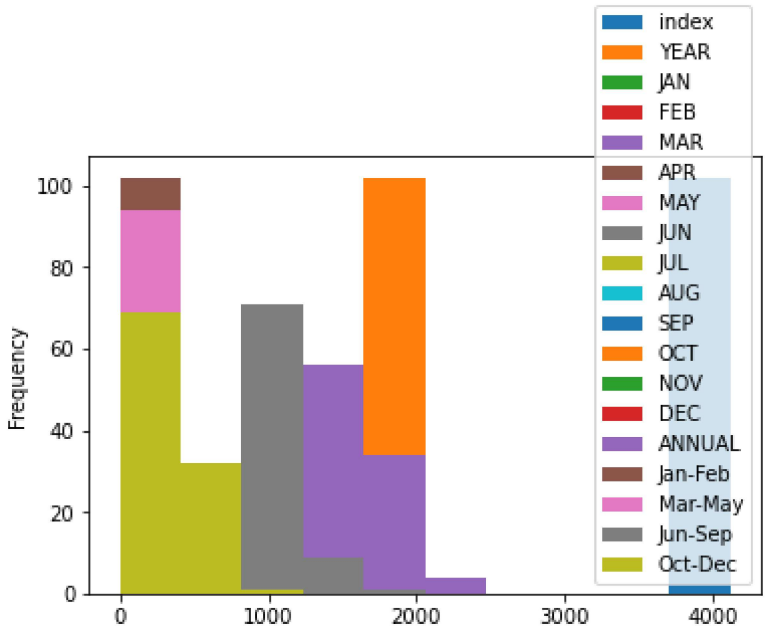
Out[8]:
```
<AxesSubplot:>
```



Histogram

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

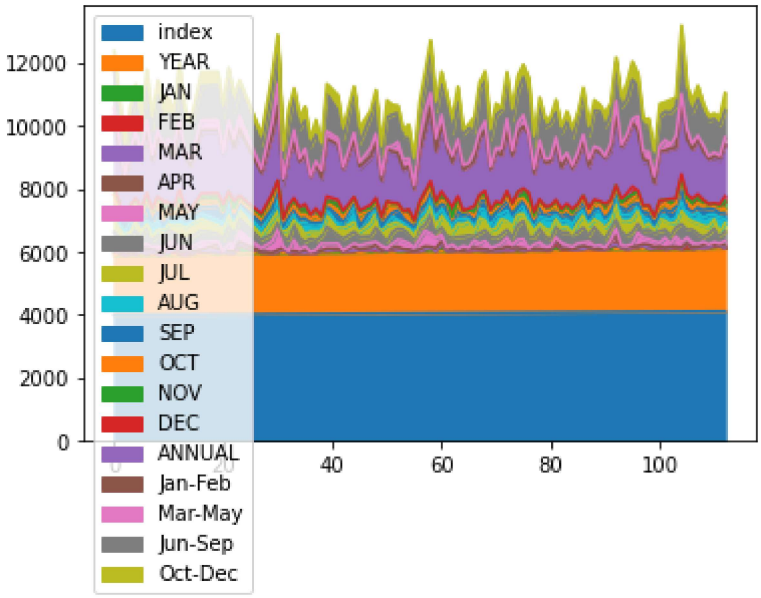
Out[9]:
```
<AxesSubplot:ylabel='Frequency'>
```



Area Chart

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

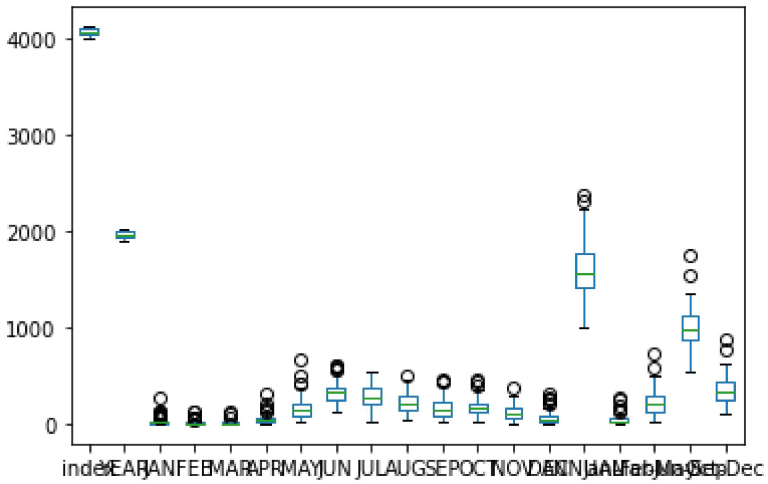
Out[10]: <AxesSubplot:>



Box Chart

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

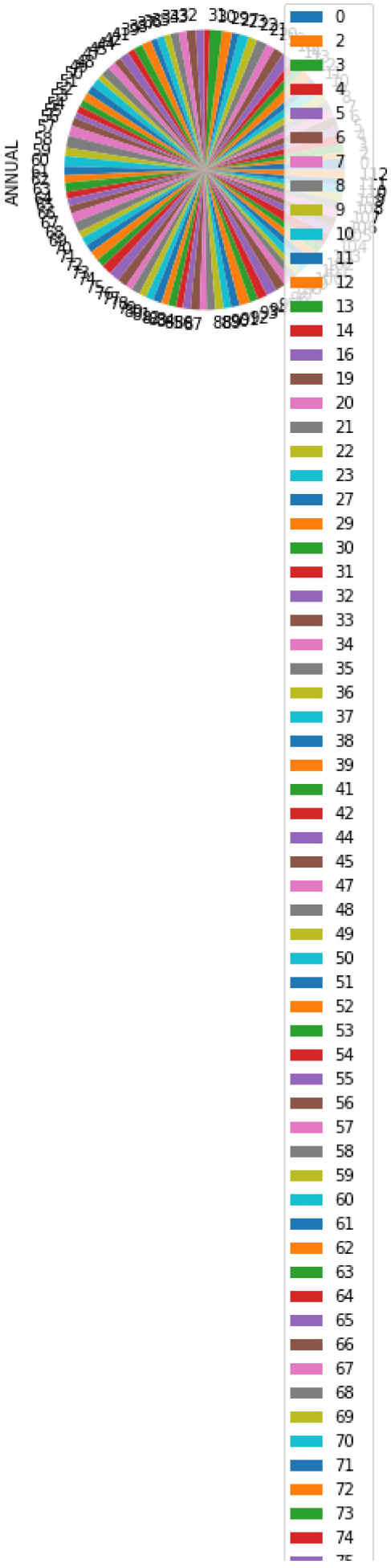
Out[11]: <AxesSubplot:>

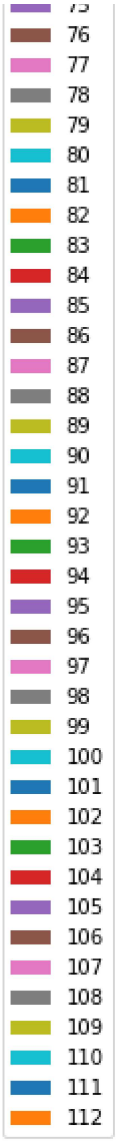


Pie Chart

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

Out[12]:
```
<AxesSubplot:ylabel='ANNUAL'>
```

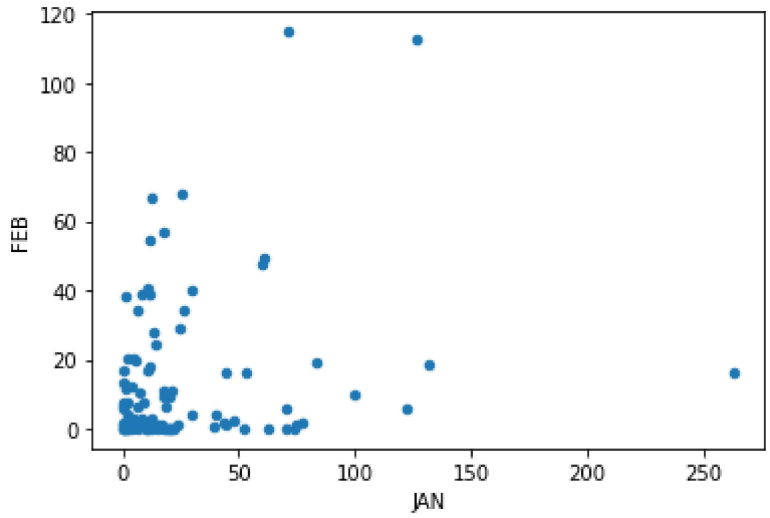

ANNUAL
0
2
3
4
5
6
7
8
9
10
11
12
13
14
16
19
20
21
22
23
27
29
30
31
32
33
34
35
36
37
38
39
41
42
44
45
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74



Scatter Plot

In [13]:
```
df.plot.scatter(x='JAN',y='FEB')
```

Out[13]: <AxesSubplot:xlabel='JAN', ylabel='FEB'>



In [14]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 102 entries, 0 to 112
Data columns (total 20 columns):
 #   Column       Non-Null Count  Dtype  
---  ------       --------------  -----  
 0   index        102 non-null    int64  
 1   SUBDIVISION  102 non-null    object 
 2   YEAR         102 non-null    int64  
 3   JAN          102 non-null    float64
 4   FEB          102 non-null    float64
 5   MAR          102 non-null    float64
 6   APR          102 non-null    float64
 7   MAY          102 non-null    float64
 8   JUN          102 non-null    float64
 9   JUL          102 non-null    float64
 10  AUG          102 non-null    float64
 11  SEP          102 non-null    float64
 12  OCT          102 non-null    float64
 13  NOV          102 non-null    float64
 14  DEC          102 non-null    float64
 15  ANNUAL       102 non-null    float64
 16  Jan-Feb      102 non-null    float64
 17  Mar-May      102 non-null    float64
 18  Jun-Sep      102 non-null    float64
 19  Oct-Dec      102 non-null    float64
dtypes: float64(17), int64(2), object(1)
memory usage: 16.7+ KB
```

In [15]:
```
df.describe()
```

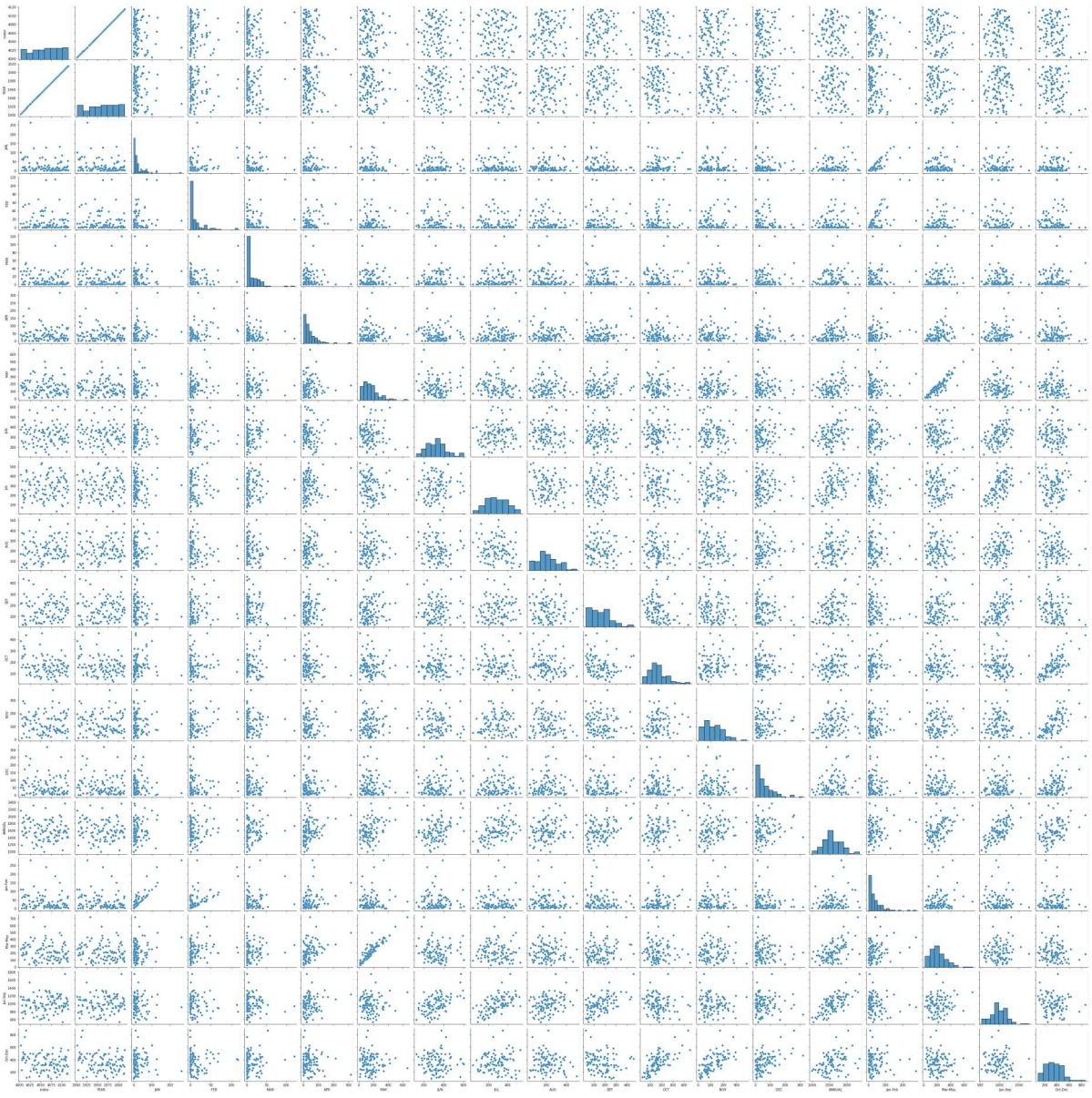
Out[15]:

	index	YEAR	JAN	FEB	MAR	APR	MAY	
count	102.000000	102.000000	102.000000	102.000000	102.000000	102.000000	102.000000	10
mean	4062.264706	1962.127451	25.350980	13.053922	13.141176	43.504902	160.910784	32
std	32.591053	32.812601	37.411685	21.356125	19.139278	47.793316	111.155138	10
min	4003.000000	1902.000000	0.000000	0.000000	0.000000	0.000000	13.500000	12
25%	4036.250000	1936.250000	3.850000	0.400000	0.375000	13.625000	82.300000	25
50%	4064.500000	1964.500000	12.050000	3.650000	5.150000	32.450000	143.050000	32
75%	4089.750000	1989.750000	26.000000	16.725000	20.725000	59.550000	204.600000	37
max	4115.000000	2015.000000	262.800000	114.900000	120.700000	315.400000	660.800000	60

EDA And Visualization

In [16]:
```
sns.pairplot(df)
```

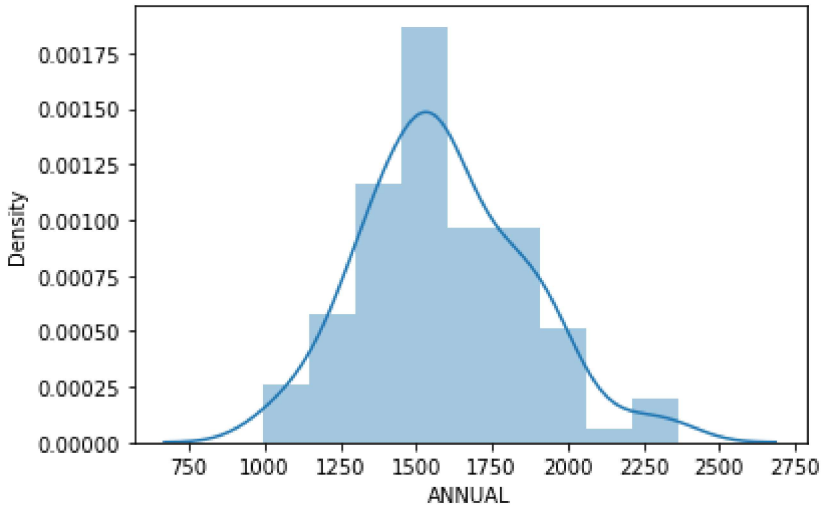
Out[16]:
```
<seaborn.axisgrid.PairGrid at 0x289c107ae50>
```



In [17]:
```
sns.distplot(df['ANNUAL'])
```

```
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\distributions.py:2557: Fut
ureWarning: `distplot` is a deprecated function and will be removed in a futu
re version. Please adapt your code to use either `displot` (a figure-level fu
nction with similar flexibility) or `histplot` (an axes-level function for hi
stograms).
  warnings.warn(msg, FutureWarning)
```

Out[17]: <AxesSubplot:xlabel='ANNUAL', ylabel='Density'>



In [18]:
```
sns.heatmap(df.corr())
```

Out[18]: <AxesSubplot:>