

Import Libraries

In [1]:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

In [2]:

```
d=pd.read_csv(r"C:\Users\user\Downloads\FP2_RainFall\rain.csv")[2739:2852]
d
```

Out[2]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP
2739	2739	MATATHWADA	1903	2.6	0.8	0.0	1.7	58.3	104.4	264.2	281.9	173.3
2740	2740	MATATHWADA	1904	0.0	0.9	12.1	0.3	7.2	79.2	118.4	57.3	339.0
2741	2741	MATATHWADA	1905	1.3	2.0	0.0	6.6	4.8	84.6	94.8	137.6	157.8
2742	2742	MATATHWADA	1906	19.8	0.0	0.1	0.0	0.2	220.6	254.9	156.9	82.1
2743	2743	MATATHWADA	1907	1.0	6.6	1.1	61.3	0.0	90.4	198.9	155.2	34.7
...
2847	2847	MATATHWADA	2011	0.0	3.8	0.7	3.5	3.1	79.2	230.1	228.5	90.0
2848	2848	MATATHWADA	2012	0.0	0.0	0.0	0.6	2.3	72.2	161.1	101.4	120.0
2849	2849	MATATHWADA	2013	1.5	9.4	2.6	7.9	6.4	160.9	293.4	136.9	154.1
2850	2850	MATATHWADA	2014	1.4	13.4	79.0	11.9	7.0	30.4	105.0	178.9	84.5
2851	2851	MATATHWADA	2015	10.1	1.6	32.0	39.6	12.3	118.3	27.4	112.2	154.3

113 rows × 20 columns



Data Cleaning and preprocessing

In [3]:

```
d.dropna()
```

Out[3]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP
2739	2739	MATATHWADA	1903	2.6	0.8	0.0	1.7	58.3	104.4	264.2	281.9	173.3
2740	2740	MATATHWADA	1904	0.0	0.9	12.1	0.3	7.2	79.2	118.4	57.3	339.0
2741	2741	MATATHWADA	1905	1.3	2.0	0.0	6.6	4.8	84.6	94.8	137.6	157.8
2742	2742	MATATHWADA	1906	19.8	0.0	0.1	0.0	0.2	220.6	254.9	156.9	82.1
2743	2743	MATATHWADA	1907	1.0	6.6	1.1	61.3	0.0	90.4	198.9	155.2	34.7
...
2847	2847	MATATHWADA	2011	0.0	3.8	0.7	3.5	3.1	79.2	230.1	228.5	90.0
2848	2848	MATATHWADA	2012	0.0	0.0	0.0	0.6	2.3	72.2	161.1	101.4	120.0
2849	2849	MATATHWADA	2013	1.5	9.4	2.6	7.9	6.4	160.9	293.4	136.9	154.1
2850	2850	MATATHWADA	2014	1.4	13.4	79.0	11.9	7.0	30.4	105.0	178.9	84.5
2851	2851	MATATHWADA	2015	10.1	1.6	32.0	39.6	12.3	118.3	27.4	112.2	154.3

113 rows × 20 columns

In [4]:

```
d.columns
```

Out[4]:

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

In [5]:

```
d.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 113 entries, 2739 to 2851
Data columns (total 20 columns):
 #   Column          Non-Null Count  Dtype
---  -
 0   index           113 non-null    int64
 1   SUBDIVISION     113 non-null    object
 2   YEAR            113 non-null    int64
 3   JAN             113 non-null    float64
 4   FEB             113 non-null    float64
 5   MAR             113 non-null    float64
 6   APR             113 non-null    float64
 7   MAY             113 non-null    float64
 8   JUN             113 non-null    float64
 9   JUL             113 non-null    float64
10  AUG             113 non-null    float64
11  SEP             113 non-null    float64
12  OCT             113 non-null    float64
13  NOV             113 non-null    float64
14  DEC             113 non-null    float64
15  ANNUAL          113 non-null    float64
16  Jan-Feb         113 non-null    float64
17  Mar-May         113 non-null    float64
18  Jun-Sep         113 non-null    float64
19  Oct-Dec         113 non-null    float64
dtypes: float64(17), int64(2), object(1)
memory usage: 17.8+ KB
```

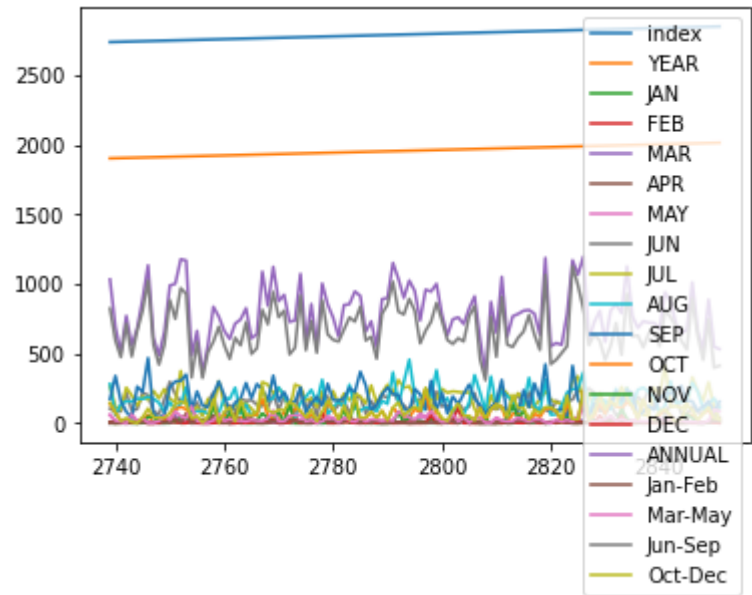
Line Chart

In [6]:

```
d.plot.line()
```

Out[6]:

<AxesSubplot:>



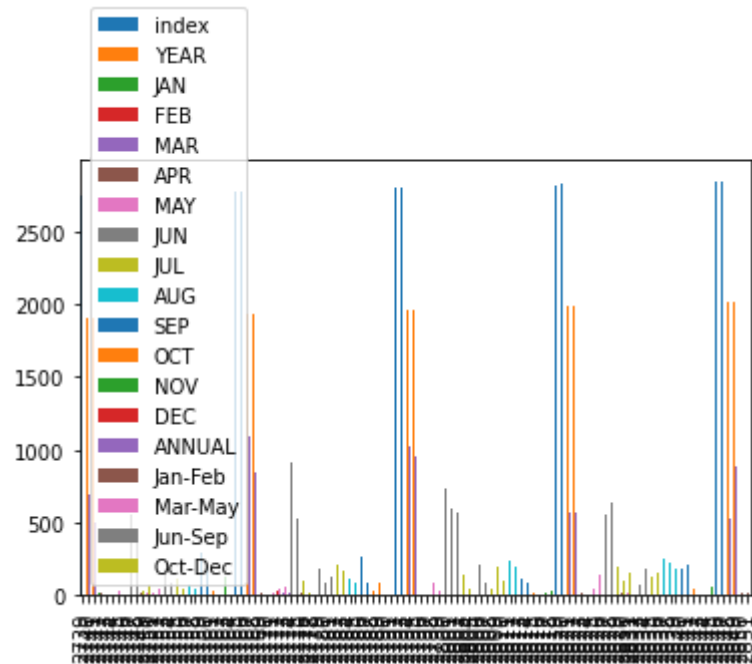
Bar Chart

In [7]:

```
d.plot.bar()
```

Out[7]:

<AxesSubplot:>



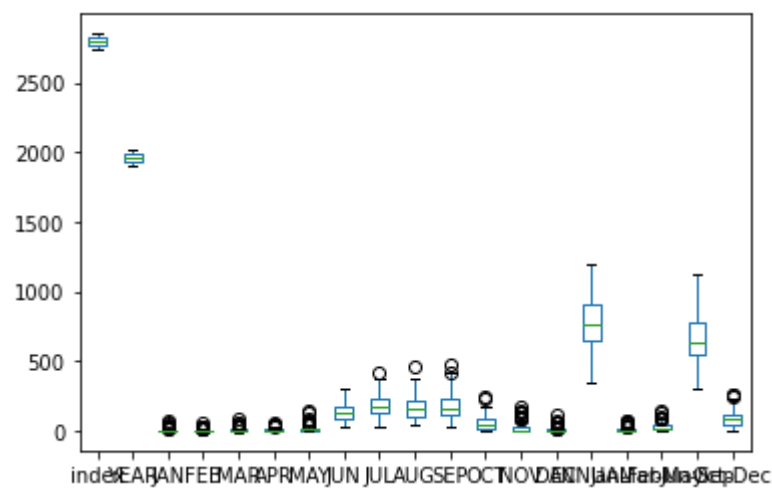
Box Chart

In [8]:

```
d.plot.box()
```

Out[8]:

<AxesSubplot:>



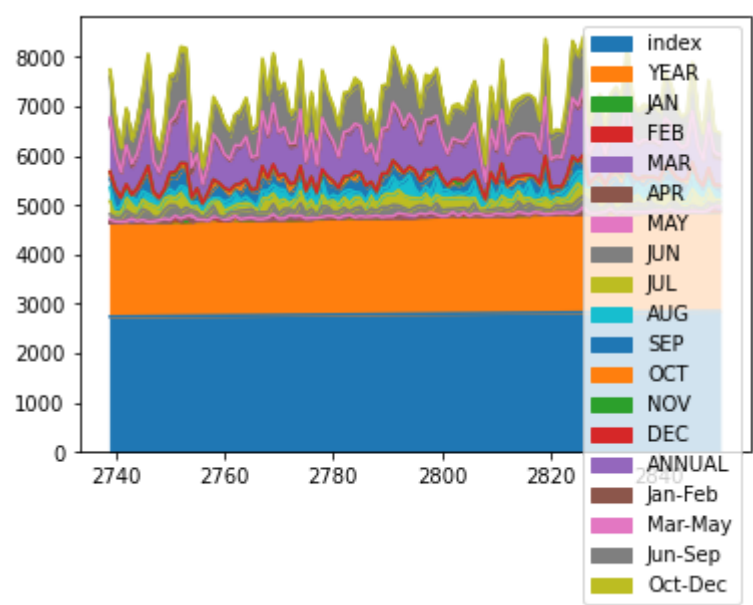
Area Chart

In [9]:

```
d.plot.area()
```

Out[9]:

<AxesSubplot:>



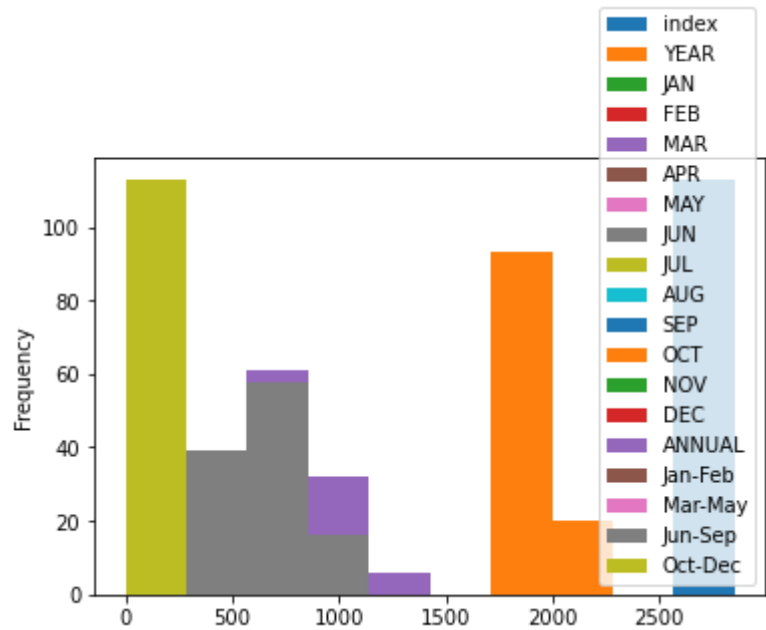
Histogram

In [10]:

```
d.plot.hist()
```

Out[10]:

<AxesSubplot:ylabel='Frequency'>



Pie Chart

In [11]:

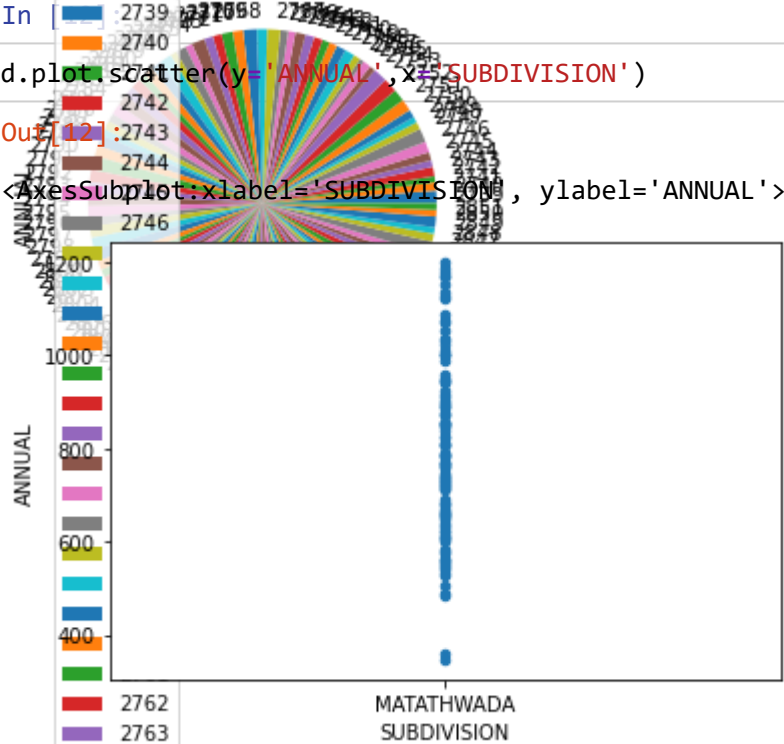
```
d.plot.pie(y='ANNUAL')
```

Out[11]:

```
<AxesSubplot:ylabel='ANNUAL'>
```

Scatter Chart


```
In [12]: d.plot.scatter(y='ANNUAL',x='SUBDIVISION')
Out[12]: <AxesSubplot: xlabel='SUBDIVISION', ylabel='ANNUAL'>
```

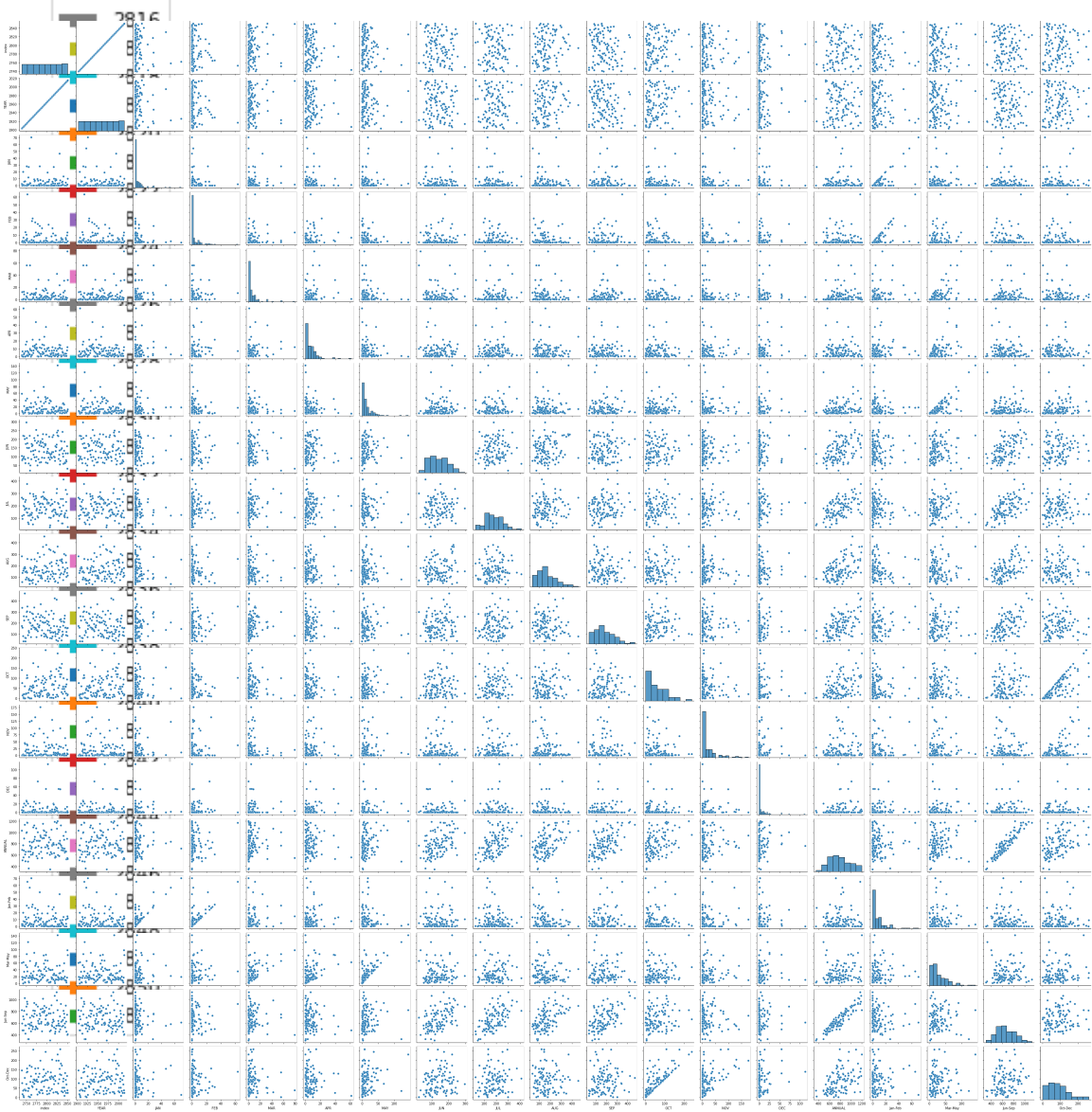


```
In [13]: d.describe()
```

	index	YEAR	JAN	FEB	MAR	APR	MAY	
count	113.00000	113.00000	113.00000	113.00000	113.00000	113.00000	113.00000	11
mean	2795.00000	1959.00000	4.938053	4.49292	6.943363	7.236283	15.682301	13
std	32.76431	32.76431	10.684922	8.79344	12.414176	9.343323	21.952096	5
min	2739.00000	1903.00000	0.000000	0.00000	0.000000	0.000000	0.000000	2
25%	2767.00000	1931.00000	0.000000	0.00000	0.200000	1.500000	2.300000	9
50%	2795.00000	1959.00000	0.700000	0.70000	2.600000	4.400000	8.000000	13
75%	2823.00000	1987.00000	5.500000	5.00000	7.800000	10.500000	19.100000	17
max	2851.00000	2015.00000	70.400000	63.50000	79.000000	61.300000	142.100000	29

EDA AND VISUALIZATION

```
In [24]: sns.pairplot(d)
Out[14]: <seaborn.axisgrid.PairGrid at 0x212d91f3a00>
```

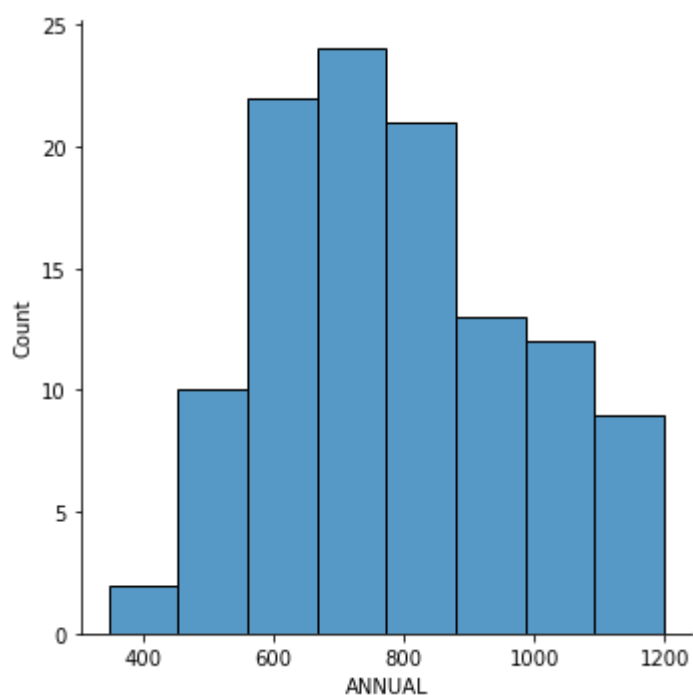


In [15]:

```
sns.displot(d['ANNUAL'])
```

Out[15]:

```
<seaborn.axisgrid.FacetGrid at 0x212e3802760>
```

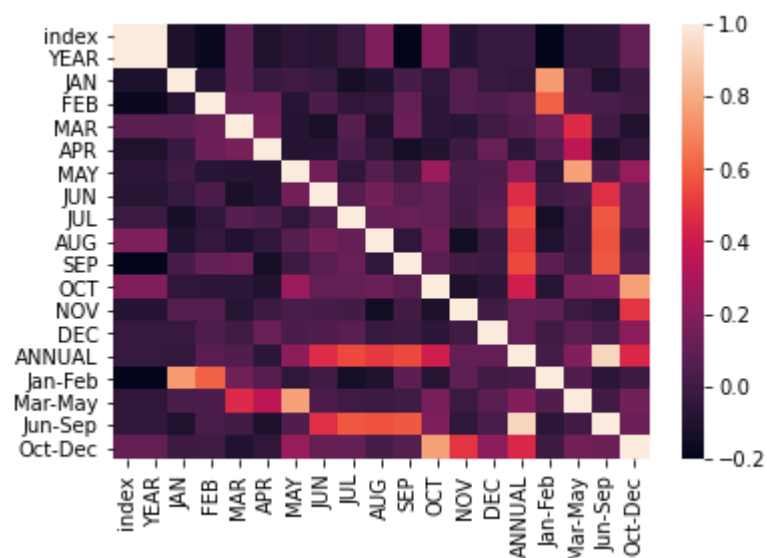


In [16]:

```
sns.heatmap(d.corr())
```

Out[16]:

```
<AxesSubplot:>
```



In []:

