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")[2509:2622]
d
```

Out[2]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP
2625	2625	MADHYA MAHARASHTRA	1904	0.4	4.7	1.7	3.0	18.7	114.6	126.5	59.5	183.0
2626	2626	MADHYA MAHARASHTRA	1905	0.0	1.2	0.0	2.3	23.6	65.0	252.8	79.0	52.6
2627	2627	MADHYA MAHARASHTRA	1906	10.5	0.8	0.0	0.1	9.3	184.8	199.3	205.0	88.8
2628	2628	MADHYA MAHARASHTRA	1907	0.5	3.7	1.5	54.5	0.6	118.0	262.3	267.8	94.1
2629	2629	MADHYA MAHARASHTRA	1908	0.3	0.0	4.7	6.3	5.8	85.1	263.0	169.9	166.6
...	...	...	...	...	...	...	...	...	...	...	...	...
2731	2731	MADHYA MAHARASHTRA	2010	2.9	0.1	0.9	2.3	5.4	185.6	280.9	233.2	165.6
2732	2732	MADHYA MAHARASHTRA	2011	0.0	0.3	0.3	5.0	2.9	133.3	261.4	238.1	148.4
2733	2733	MADHYA MAHARASHTRA	2012	0.0	0.0	0.0	3.0	1.4	67.9	203.0	187.8	129.5
2734	2734	MADHYA MAHARASHTRA	2013	0.1	5.3	0.8	5.7	6.0	212.4	311.8	147.0	210.3
2735	2735	MADHYA MAHARASHTRA	2014	3.1	6.2	24.4	7.5	29.8	44.0	277.9	240.3	120.4

111 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
2625	2625	MADHYA MAHARASHTRA	1904	0.4	4.7	1.7	3.0	18.7	114.6	126.5	59.5	183.0
2626	2626	MADHYA MAHARASHTRA	1905	0.0	1.2	0.0	2.3	23.6	65.0	252.8	79.0	52.6
2627	2627	MADHYA MAHARASHTRA	1906	10.5	0.8	0.0	0.1	9.3	184.8	199.3	205.0	88.8
2628	2628	MADHYA MAHARASHTRA	1907	0.5	3.7	1.5	54.5	0.6	118.0	262.3	267.8	94.1
2629	2629	MADHYA MAHARASHTRA	1908	0.3	0.0	4.7	6.3	5.8	85.1	263.0	169.9	166.6
...	...	...	...	...	...	...	...	...	...	...	...	...
2731	2731	MADHYA MAHARASHTRA	2010	2.9	0.1	0.9	2.3	5.4	185.6	280.9	233.2	165.6
2732	2732	MADHYA MAHARASHTRA	2011	0.0	0.3	0.3	5.0	2.9	133.3	261.4	238.1	148.4
2733	2733	MADHYA MAHARASHTRA	2012	0.0	0.0	0.0	3.0	1.4	67.9	203.0	187.8	129.5
2734	2734	MADHYA MAHARASHTRA	2013	0.1	5.3	0.8	5.7	6.0	212.4	311.8	147.0	210.3
2735	2735	MADHYA MAHARASHTRA	2014	3.1	6.2	24.4	7.5	29.8	44.0	277.9	240.3	120.4

111 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: 111 entries, 2625 to 2735
Data columns (total 20 columns):
 #   Column       Non-Null Count  Dtype  
---  ------       --------------  -----  
 0   index        111 non-null    int64  
 1   SUBDIVISION  111 non-null    object 
 2   YEAR         111 non-null    int64  
 3   JAN          111 non-null    float64
 4   FEB          111 non-null    float64
 5   MAR          111 non-null    float64
 6   APR          111 non-null    float64
 7   MAY          111 non-null    float64
 8   JUN          111 non-null    float64
 9   JUL          111 non-null    float64
 10  AUG          111 non-null    float64
 11  SEP          111 non-null    float64
 12  OCT          111 non-null    float64
 13  NOV          111 non-null    float64
 14  DEC          111 non-null    float64
 15  ANNUAL       111 non-null    float64
 16  Jan-Feb      111 non-null    float64
 17  Mar-May      111 non-null    float64
 18  Jun-Sep      111 non-null    float64
 19  Oct-Dec      111 non-null    float64
dtypes: float64(17), int64(2), object(1)
memory usage: 17.5+ KB
```

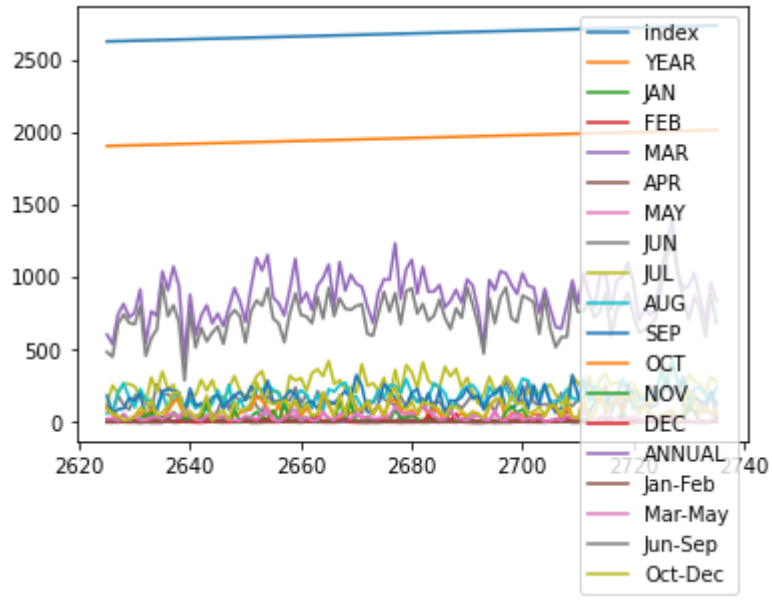
Line Chart

In [6]:

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

Out[6]:

<AxesSubplot:>



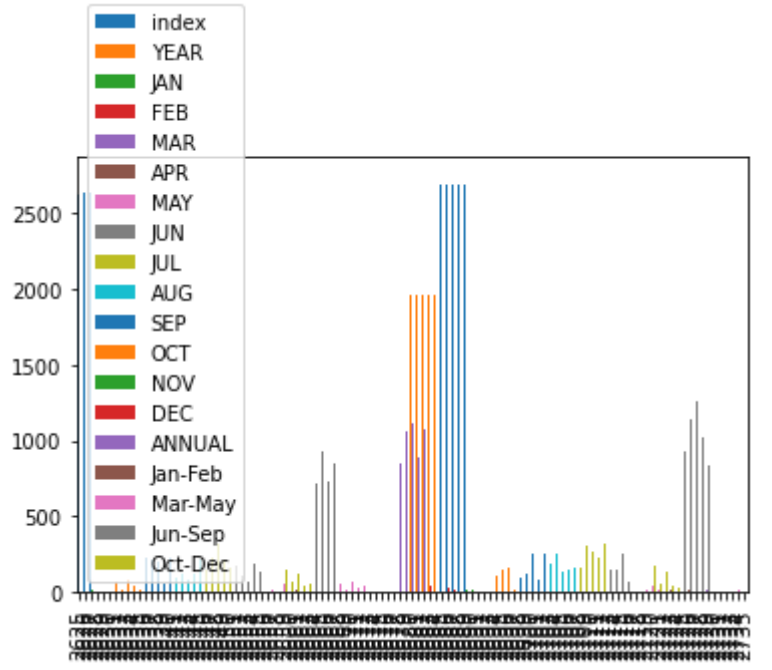
Bar Chart

In [7]:

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

Out[7]:

<AxesSubplot:>



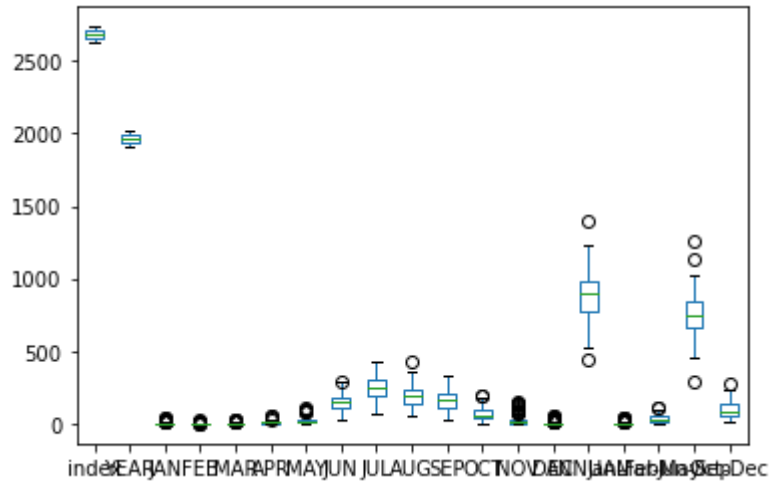
Box Chart

In [8]:

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

Out[8]:

<AxesSubplot:>



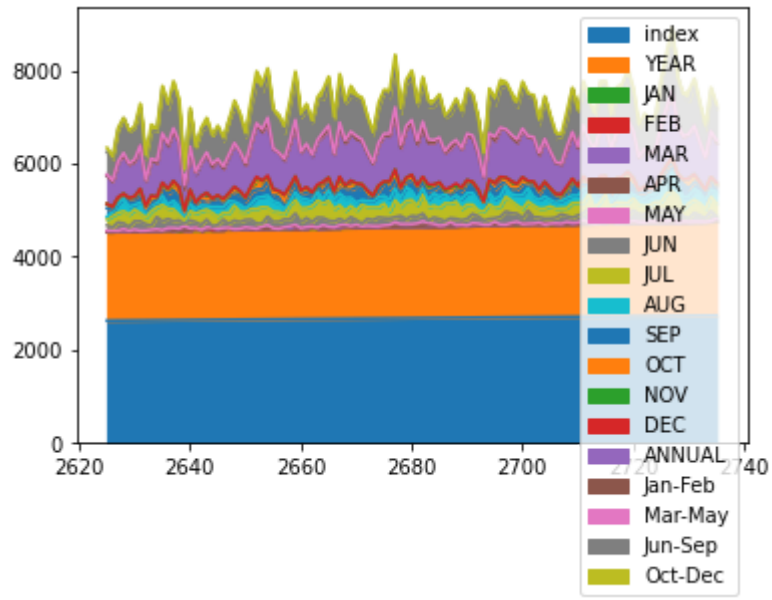
Area Chart

In [9]:

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

Out[9]:

<AxesSubplot:>



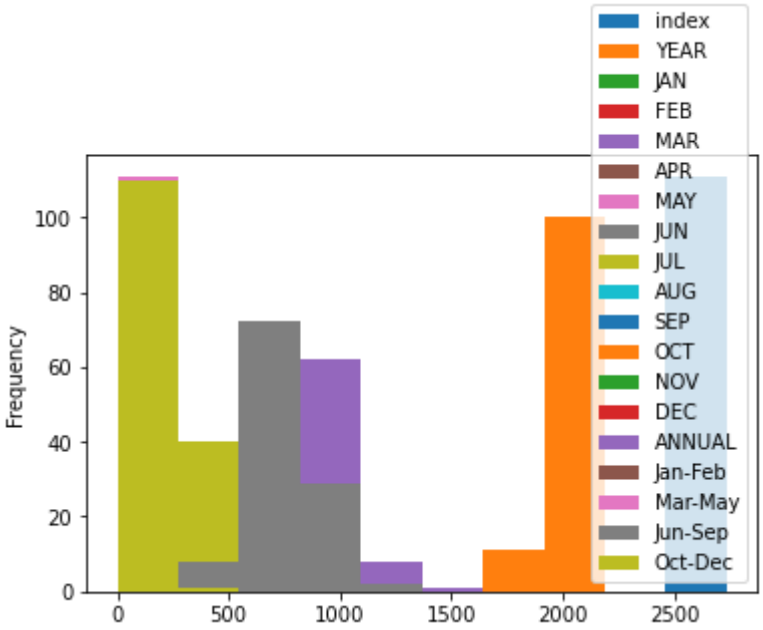
Histogram

In [10]:

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

Out[10]:

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



Pie Chart

In [11]:

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

Out[11]:

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

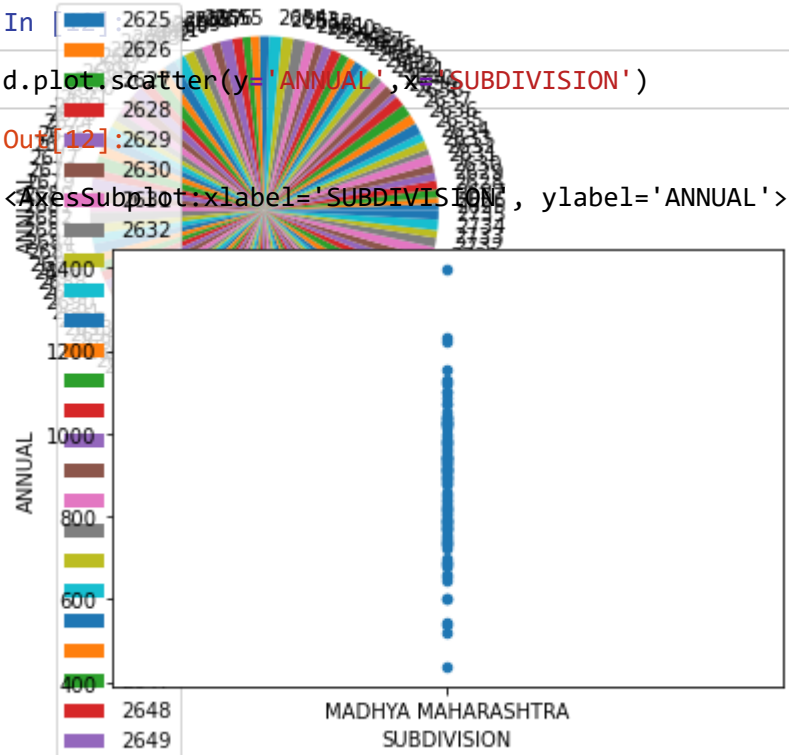
Scatter Chart

In [ ]:

```
d.plot.scatter(y='ANNUAL',x='SUBDIVISION')
```

Out[12]:

```
<AxesSubplot:xlabel='SUBDIVISION', ylabel='ANNUAL'>
```



In [ ]:

```
d.describe()
```

Out[13]:

	index	YEAR	JAN	FEB	MAR	APR	MAY
count	111.000000	111.000000	111.000000	111.000000	111.000000	111.000000	111.000000
mean	2680.000000	1959.000000	2.844144	1.508108	3.284685	8.986486	22.492793
std	32.186954	32.186954	6.583653	2.959089	5.410095	9.179042	22.123832
min	2625.000000	1904.000000	0.000000	0.000000	0.000000	0.000000	0.300000
25%	2652.500000	1931.500000	0.000000	0.000000	0.250000	3.200000	7.050000
50%	2680.000000	1959.000000	0.600000	0.200000	1.500000	6.300000	15.100000
75%	2707.500000	1986.500000	2.450000	1.700000	4.100000	12.100000	33.000000
max	2735.000000	2014.000000	41.500000	20.000000	32.700000	54.500000	104.200000

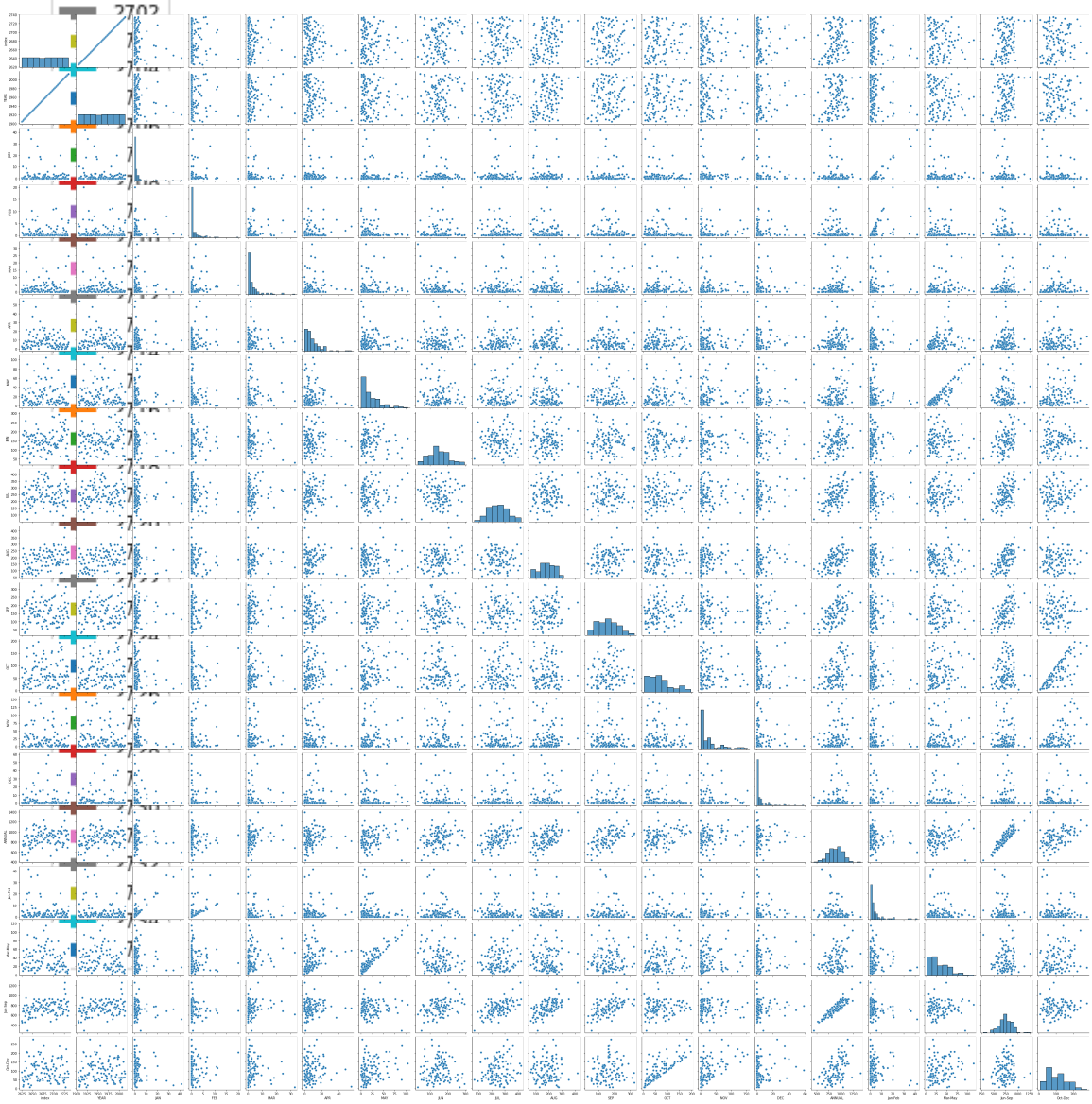
EDA AND VISUALIZATION

In [14]:

```
sns.pairplot(d)
```

Out[14]:

```
<seaborn.axisgrid.PairGrid at 0x15bf523bb20>
```

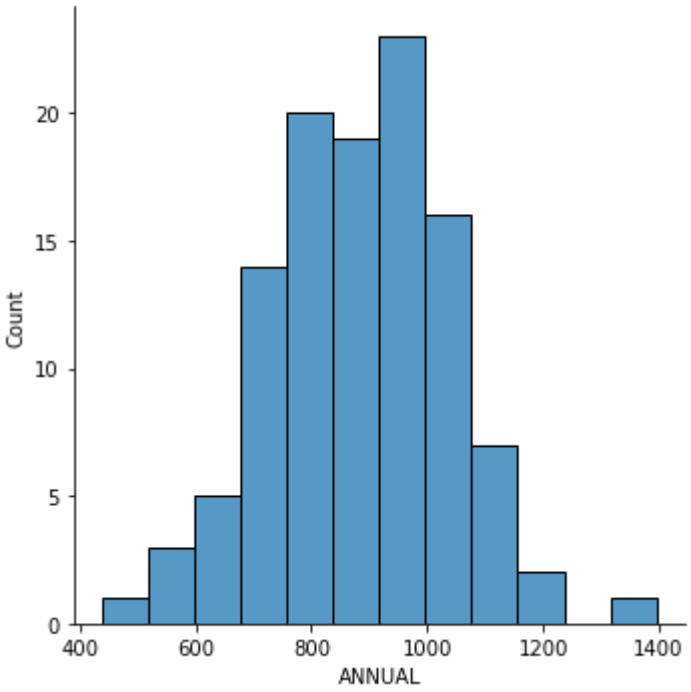


In [15]:

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

Out[15]:

```
<seaborn.axisgrid.FacetGrid at 0x15b85c30f10>
```

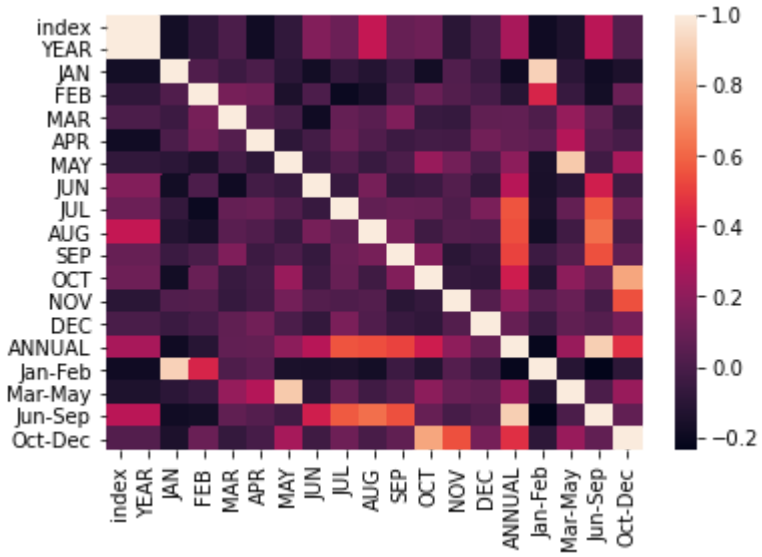


In [16]:

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

Out[16]:

```
<AxesSubplot:>
```



In [ ]: