Import Libraries

In [1]:

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

In [2]:

```
df=pd.read_csv(r"c:\Users\user\Downloads\FP2_RainFall\rainfall.csv")[2164:2277]
df
```

Out[2]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP
2164	2164	EAST MADHYA PRADESH	1903	5.6	2.9	0.3	0.9	37.5	67.5	261.4	366.7	257.4
2165	2165	EAST MADHYA PRADESH	1904	2.0	15.3	48.2	0.0	8.6	109.9	443.2	316.6	135.6
2166	2166	EAST MADHYA PRADESH	1905	15.9	8.0	14.3	12.3	10.2	34.4	292.4	243.3	250.9
2167	2167	EAST MADHYA PRADESH	1906	12.1	28.3	30.8	0.0	3.5	226.1	444.2	190.1	417.0
2168	2168	EAST MADHYA PRADESH	1907	7.0	103.1	4.5	30.5	5.1	90.9	221.9	512.3	20.1
...	...	...	...	...	...	...	...	...	...	...	...	...
2272	2272	EAST MADHYA PRADESH	2011	0.6	1.9	0.3	7.1	4.7	332.5	323.6	326.9	276.5
2273	2273	EAST MADHYA PRADESH	2012	39.4	0.7	0.6	1.1	1.2	67.8	398.9	351.7	172.6
2274	2274	EAST MADHYA PRADESH	2013	2.0	43.4	14.1	9.5	0.3	311.9	456.2	480.8	78.0
2275	2275	EAST MADHYA PRADESH	2014	32.1	49.7	17.8	5.1	2.5	91.8	283.4	231.8	139.6
2276	2276	EAST MADHYA PRADESH	2015	37.3	11.0	73.4	25.8	6.3	139.2	262.2	272.1	71.6

113 rows × 20 columns

Data Cleaning and Preprocessing

In [3]:

```
df.dropna()
```

Out[3]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP
2164	2164	EAST MADHYA PRADESH	1903	5.6	2.9	0.3	0.9	37.5	67.5	261.4	366.7	257.4
2165	2165	EAST MADHYA PRADESH	1904	2.0	15.3	48.2	0.0	8.6	109.9	443.2	316.6	135.6
2166	2166	EAST MADHYA PRADESH	1905	15.9	8.0	14.3	12.3	10.2	34.4	292.4	243.3	250.9
2167	2167	EAST MADHYA PRADESH	1906	12.1	28.3	30.8	0.0	3.5	226.1	444.2	190.1	417.0
2168	2168	EAST MADHYA PRADESH	1907	7.0	103.1	4.5	30.5	5.1	90.9	221.9	512.3	20.1
...	...	...	...	...	...	...	...	...	...	...	...	...
2272	2272	EAST MADHYA PRADESH	2011	0.6	1.9	0.3	7.1	4.7	332.5	323.6	326.9	276.5
2273	2273	EAST MADHYA PRADESH	2012	39.4	0.7	0.6	1.1	1.2	67.8	398.9	351.7	172.6
2274	2274	EAST MADHYA PRADESH	2013	2.0	43.4	14.1	9.5	0.3	311.9	456.2	480.8	78.0
2275	2275	EAST MADHYA PRADESH	2014	32.1	49.7	17.8	5.1	2.5	91.8	283.4	231.8	139.6
2276	2276	EAST MADHYA PRADESH	2015	37.3	11.0	73.4	25.8	6.3	139.2	262.2	272.1	71.6

113 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-Fe
b',
       'Mar-May', 'Jun-Sep', 'Oct-Dec'],
      dtype='object')
```

In [5]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 113 entries, 2164 to 2276
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
```

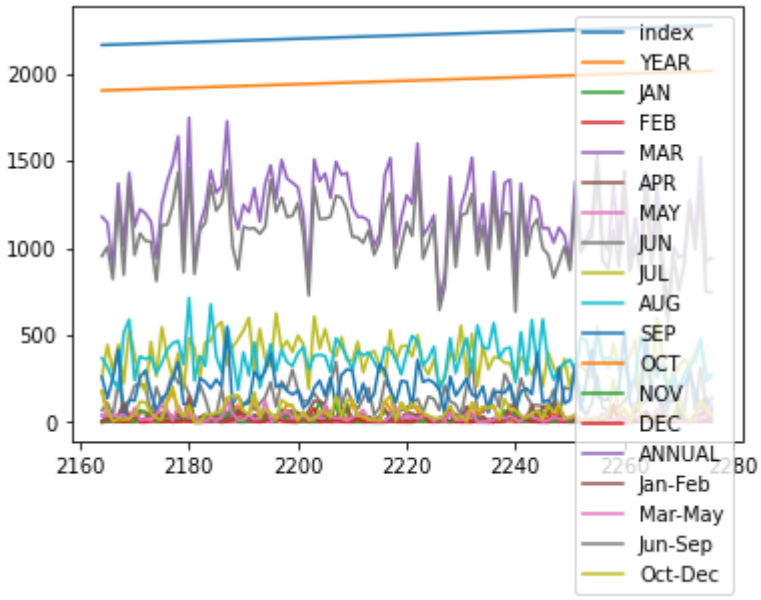
Line Chart

In [6]:

```
df.plot.line()
```

Out[6]:

<AxesSubplot:>



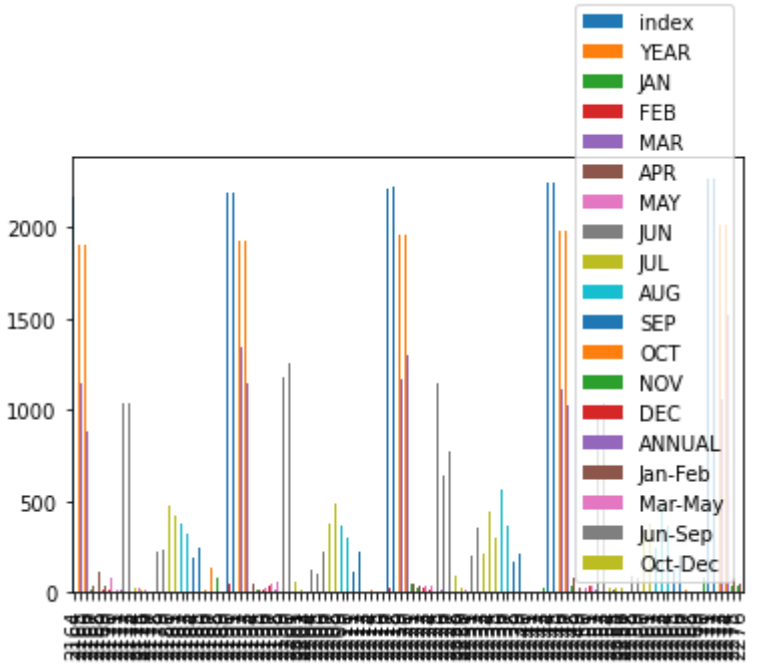
Bar chart

In [7]:

```
df.plot.bar()
```

Out[7]:

<AxesSubplot:>



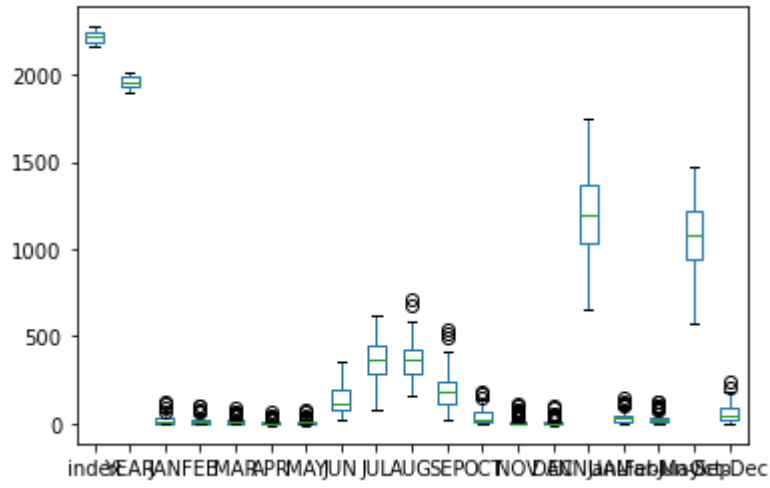
Box chart

In [8]:

```
df.plot.box()
```

Out[8]:

<AxesSubplot:>



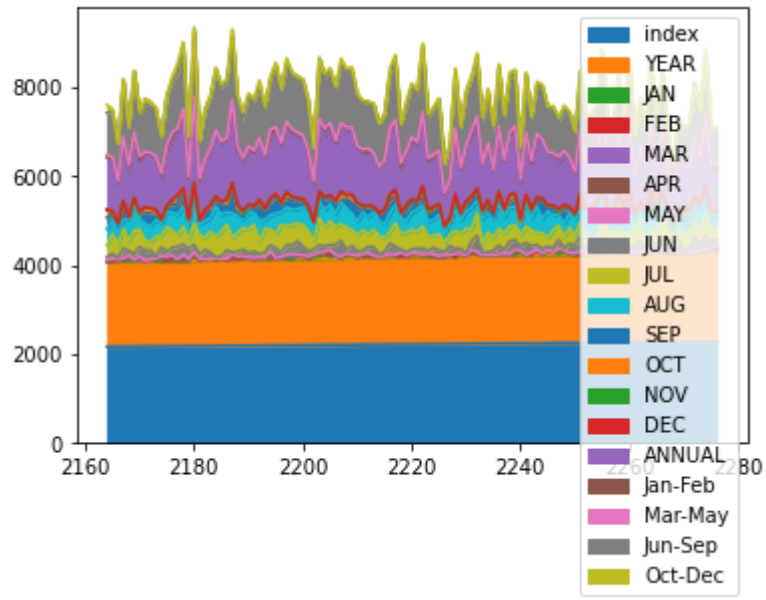
Area Chart

In [9]:

```
df.plot.area()
```

Out[9]:

<AxesSubplot:>



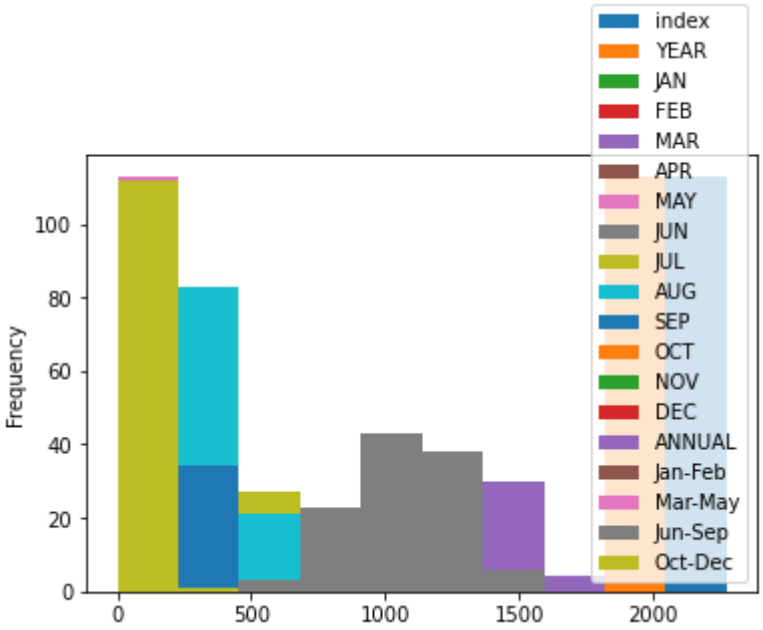
Histogram

In [10]:

```
df.plot.hist()
```

Out[10]:

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



pie chart

In [11]:

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

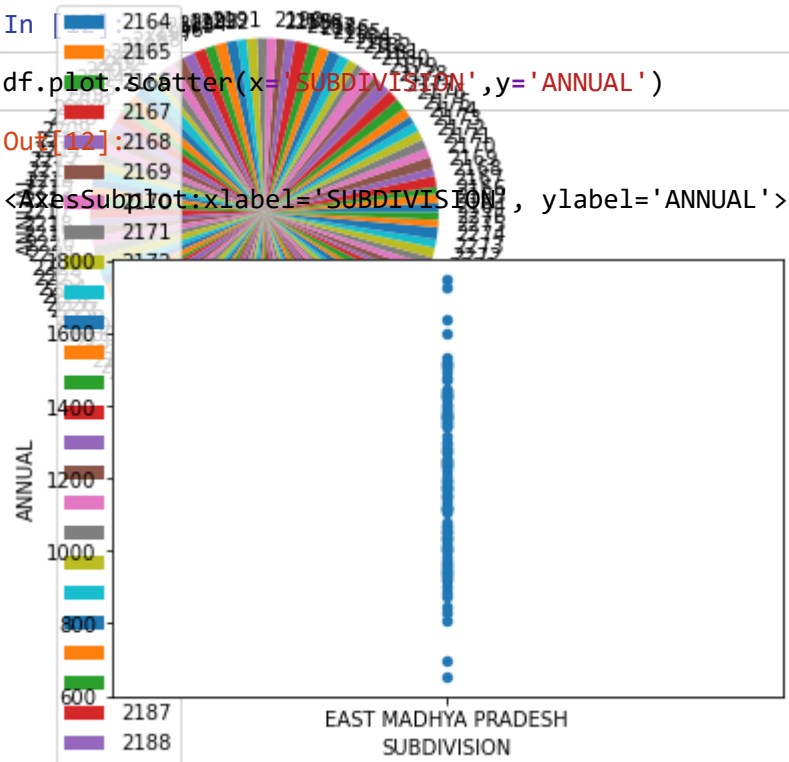
Out[11]:

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

Scatter chart

In [12]:
```
df.plot.scatter(x='SUBDIVISION',y='ANNUAL')
```

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



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

Out[13]:

	index	YEAR	JAN	FEB	MAR	APR	MAY	
count	113.00000	113.00000	113.000000	113.000000	113.000000	113.000000	113.000000	1
mean	2220.00000	1959.00000	19.184071	18.608850	13.739823	7.189381	9.358407	1
std	32.76431	32.76431	22.343114	20.879414	17.460392	10.555714	12.233843	
min	2164.00000	1903.00000	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	2192.00000	1931.00000	2.100000	3.500000	1.200000	1.300000	2.100000	
50%	2220.00000	1959.00000	12.200000	11.300000	8.000000	3.100000	5.100000	1
75%	2248.00000	1987.00000	29.600000	27.100000	18.800000	8.300000	10.600000	1
max	2276.00000	2015.00000	120.700000	103.100000	87.300000	72.400000	74.200000	3

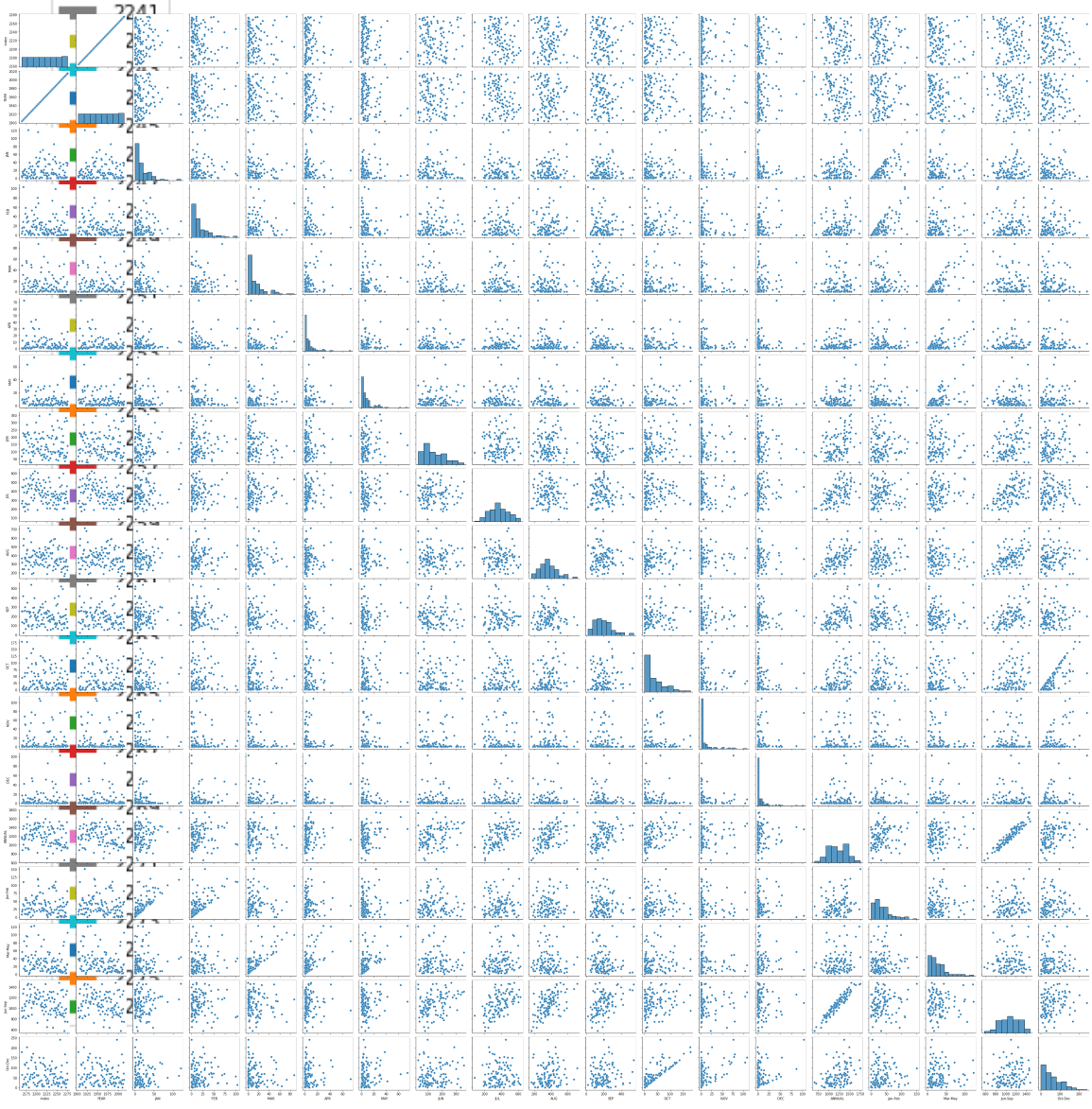
EDA and Visualization

In [14]:

```
sns.pairplot(df)
```

Out[14]:

```
<seaborn.axisgrid.PairGrid at 0x1883657d970>
```

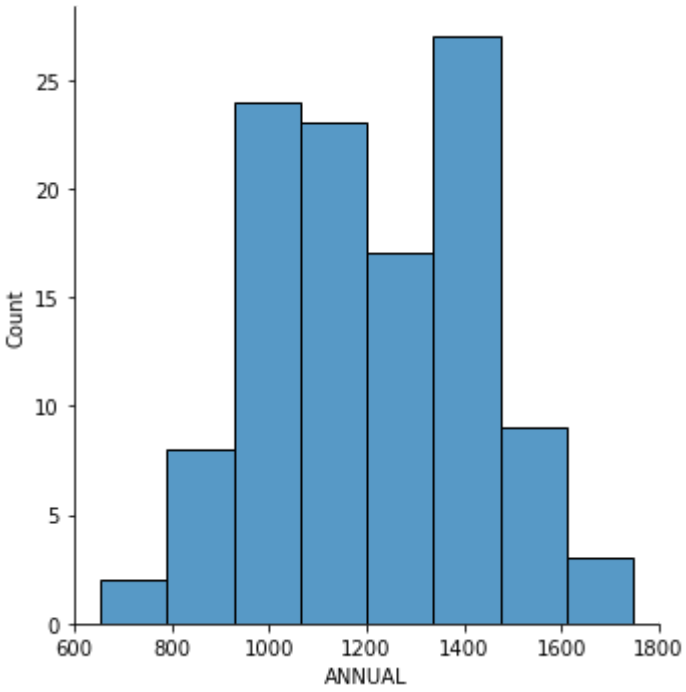


In [15]:

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

Out[15]:

```
<seaborn.axisgrid.FacetGrid at 0x188455d8e80>
```

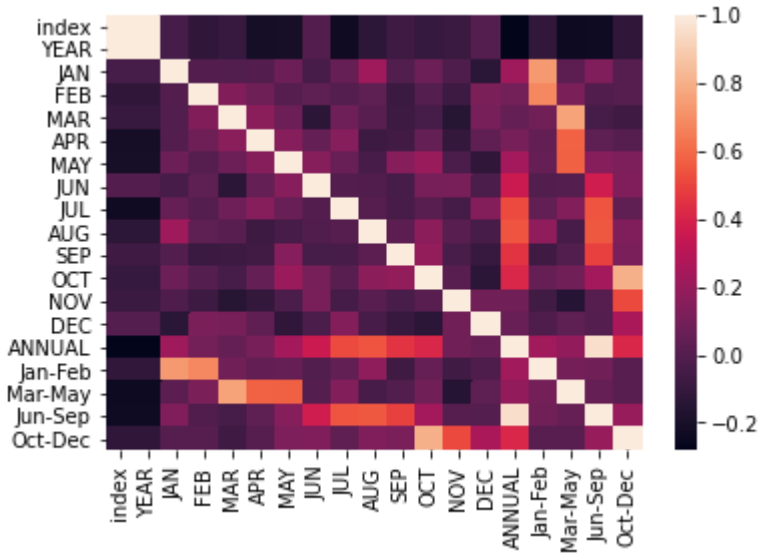


In [16]:

```
sns.heatmap(df.corr())
```

Out[16]:

```
<AxesSubplot:>
```



In [ ]:

```

```