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")[1014:1127]
df
```

Out[2]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	
1014	1014	EAST UTTAR PRADESH	1903	8.2	0.4	1.3	0.7	15.3	71.6	115.3	420.2	258.7	
1015	1015	EAST UTTAR PRADESH	1904	7.3	1.5	8.3	0.4	28.7	148.0	359.4	328.8	95.0	
1016	1016	EAST UTTAR PRADESH	1905	16.8	23.6	20.0	5.4	15.4	17.3	302.4	316.2	169.5	
1017	1017	EAST UTTAR PRADESH	1906	5.7	58.2	6.7	0.0	13.5	140.6	341.2	290.9	135.5	
1018	1018	EAST UTTAR PRADESH	1907	2.1	74.9	19.2	20.3	11.1	48.6	173.5	290.9	25.7	
...	...	...	...	...	...	...	...	...	...	...	...	...	
1122	1122	EAST UTTAR PRADESH	2011	1.0	2.7	1.6	2.9	32.2	163.8	197.9	232.1	146.4	
1123	1123	EAST UTTAR PRADESH	2012	20.3	1.2	3.4	2.8	0.2	18.5	234.2	156.0	164.4	
1124	1124	EAST UTTAR PRADESH	2013	6.1	59.6	2.7	8.7	1.1	309.7	230.0	246.1	78.2	
1125	1125	EAST UTTAR PRADESH	2014	47.4	25.8	15.4	1.7	10.7	47.8	224.5	138.1	106.7	
1126	1126	EAST UTTAR PRADESH	2015	30.0	4.1	48.2	23.2	8.6	95.3	179.0	175.8	21.9	

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
1014	1014	EAST UTTAR PRADESH	1903	8.2	0.4	1.3	0.7	15.3	71.6	115.3	420.2	258.7
1015	1015	EAST UTTAR PRADESH	1904	7.3	1.5	8.3	0.4	28.7	148.0	359.4	328.8	95.0
1016	1016	EAST UTTAR PRADESH	1905	16.8	23.6	20.0	5.4	15.4	17.3	302.4	316.2	169.5
1017	1017	EAST UTTAR PRADESH	1906	5.7	58.2	6.7	0.0	13.5	140.6	341.2	290.9	135.5
1018	1018	EAST UTTAR PRADESH	1907	2.1	74.9	19.2	20.3	11.1	48.6	173.5	290.9	25.7
...	...	...	...	...	...	...	...	...	...	...	...	...
1122	1122	EAST UTTAR PRADESH	2011	1.0	2.7	1.6	2.9	32.2	163.8	197.9	232.1	146.4
1123	1123	EAST UTTAR PRADESH	2012	20.3	1.2	3.4	2.8	0.2	18.5	234.2	156.0	164.4
1124	1124	EAST UTTAR PRADESH	2013	6.1	59.6	2.7	8.7	1.1	309.7	230.0	246.1	78.2
1125	1125	EAST UTTAR PRADESH	2014	47.4	25.8	15.4	1.7	10.7	47.8	224.5	138.1	106.7
1126	1126	EAST UTTAR PRADESH	2015	30.0	4.1	48.2	23.2	8.6	95.3	179.0	175.8	21.9

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, 1014 to 1126
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
```

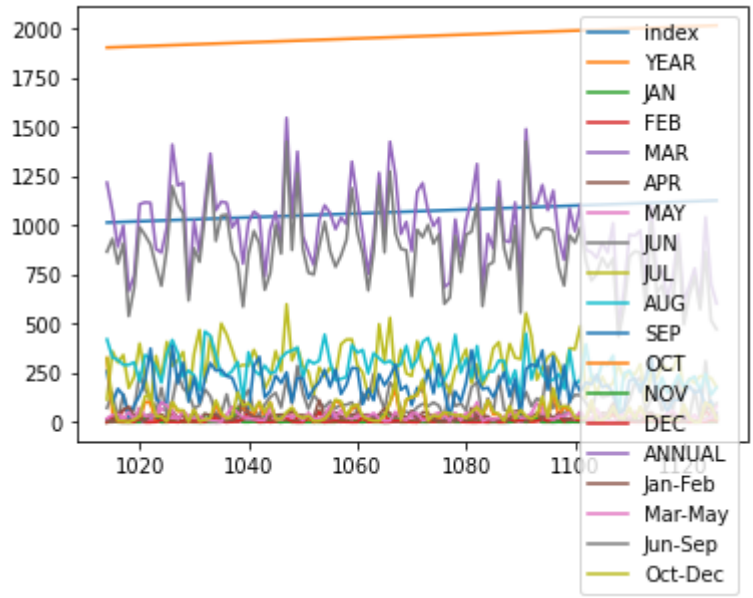
Line Chart

In [6]:

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

Out[6]:

<AxesSubplot:>



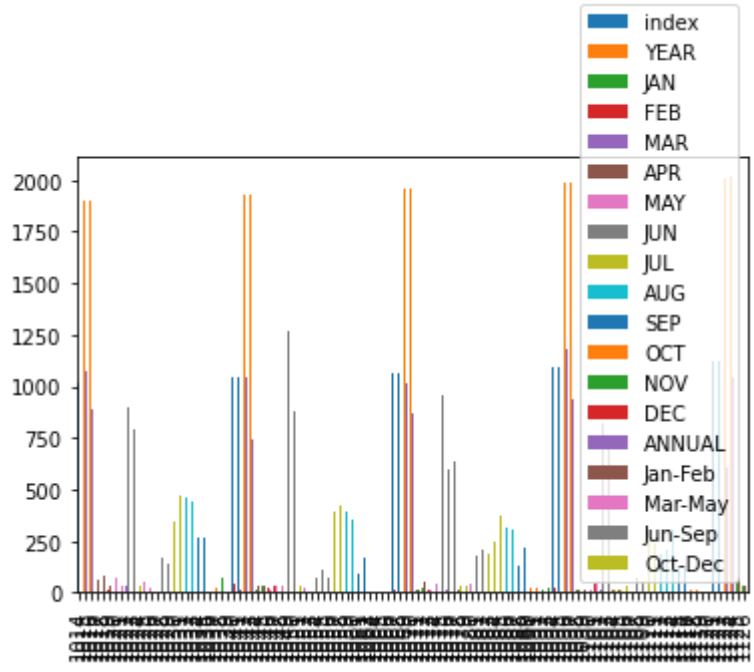
Bar chart

In [7]:

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

Out[7]:

<AxesSubplot:>



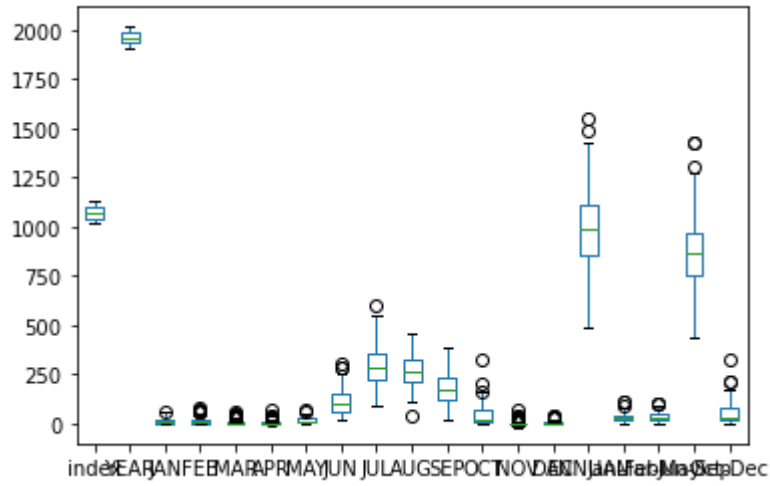
Box chart

In [8]:

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

Out[8]:

<AxesSubplot:>



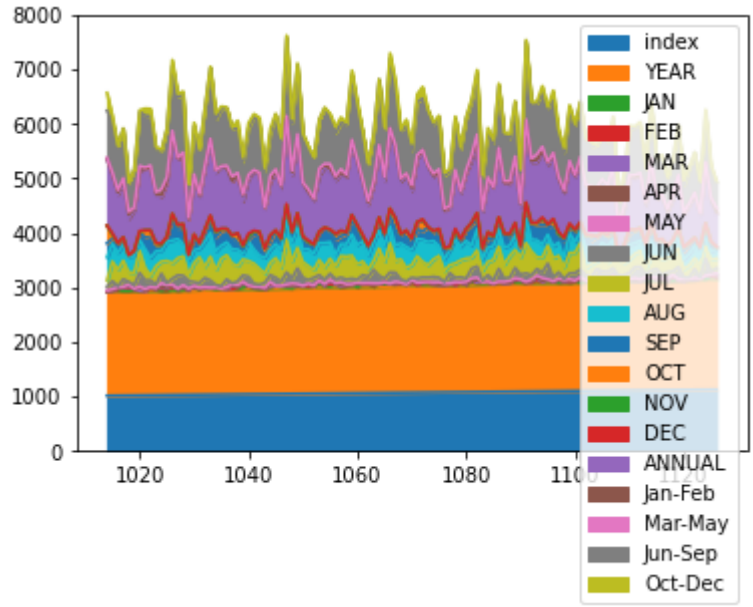
Area Chart

In [9]:

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

Out[9]:

<AxesSubplot:>



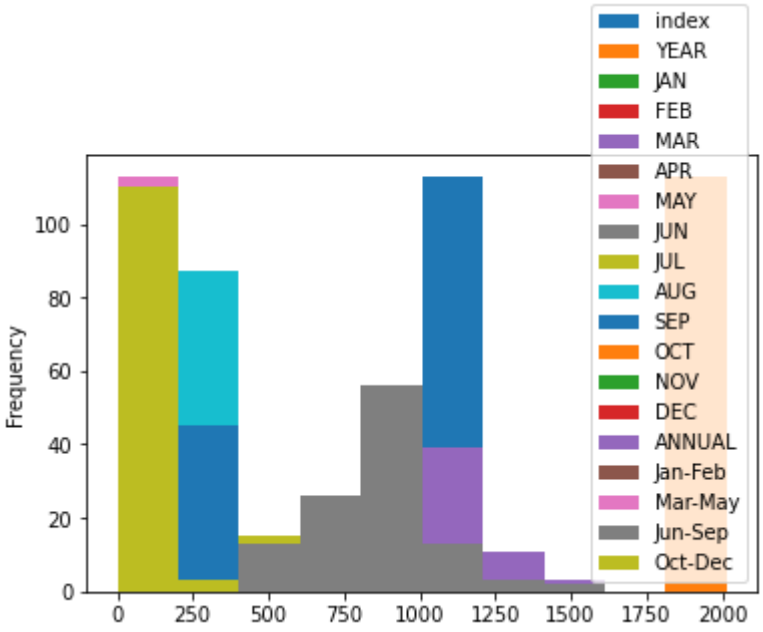
Histogram

In [10]:

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

Out[10]:

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



pie chart

In [11]:

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

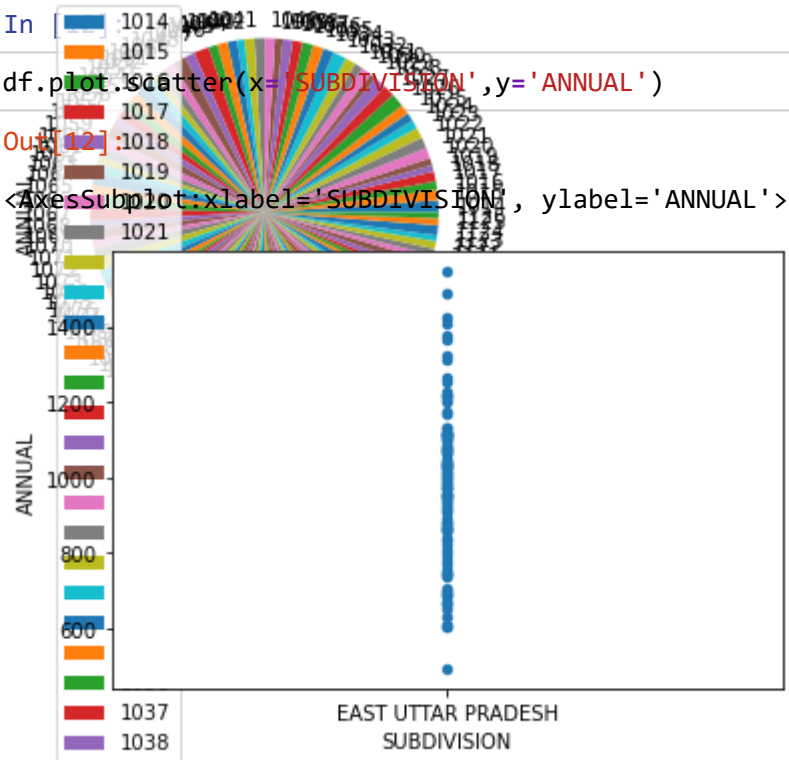
Out[11]:

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

Scatter chart

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

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



In [ ]:
```
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	1070.00000	1959.00000	15.687611	15.857522	8.971681	6.516814	17.206195	1
std	32.76431	32.76431	14.084586	17.308497	11.996573	9.139620	14.210919	
min	1014.00000	1903.00000	0.000000	0.000000	0.000000	0.000000	0.200000	
25%	1042.00000	1931.00000	3.100000	3.200000	0.700000	1.200000	6.500000	
50%	1070.00000	1959.00000	12.500000	8.800000	3.400000	3.900000	14.000000	
75%	1098.00000	1987.00000	24.300000	23.600000	12.300000	8.600000	26.000000	1
max	1126.00000	2015.00000	56.200000	84.300000	63.700000	66.600000	68.700000	3

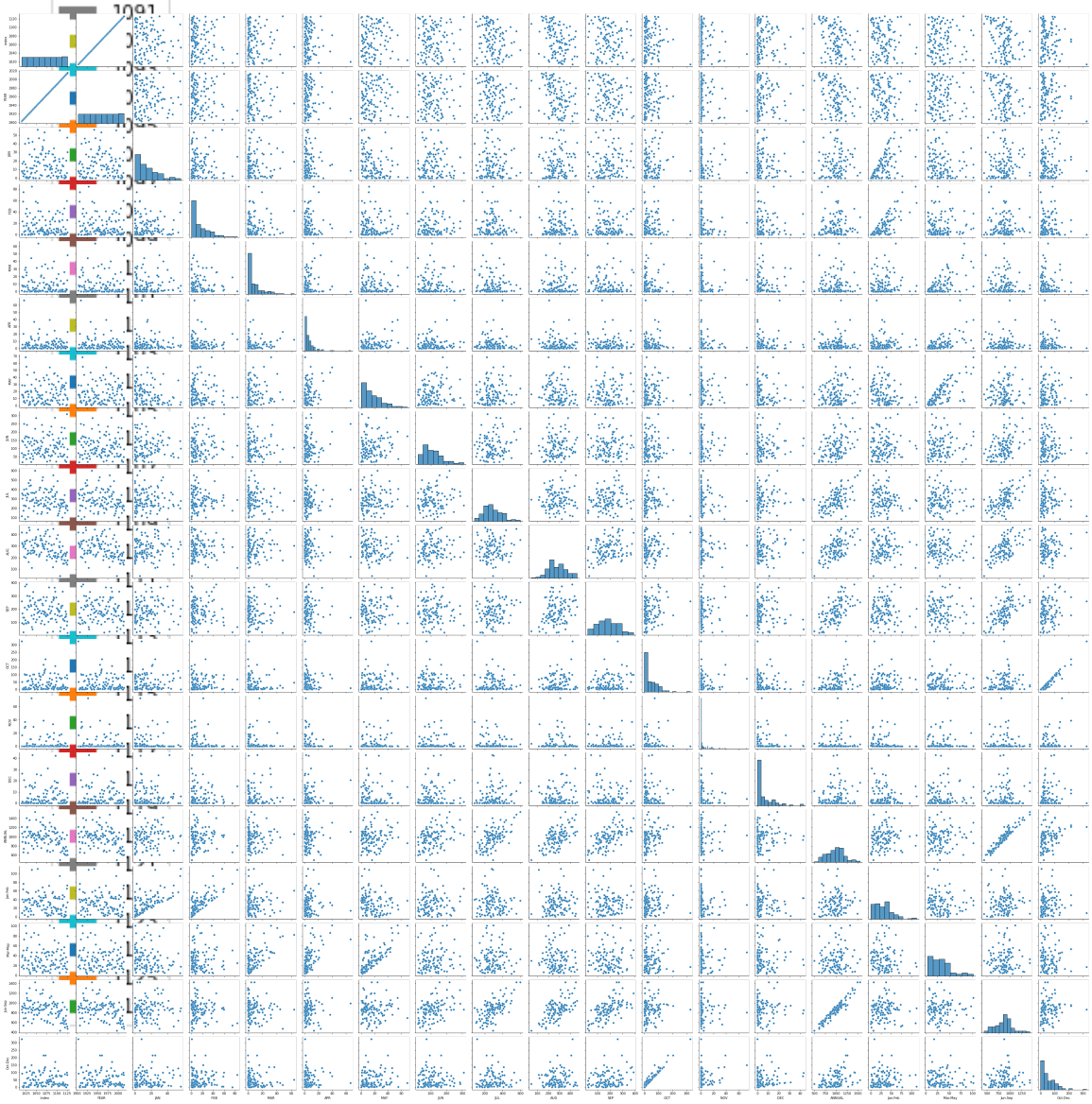
EDA and Visualization

In [14]:

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

Out[14]:

```
<seaborn.axisgrid.PairGrid at 0x1c87b9f5c70>
```

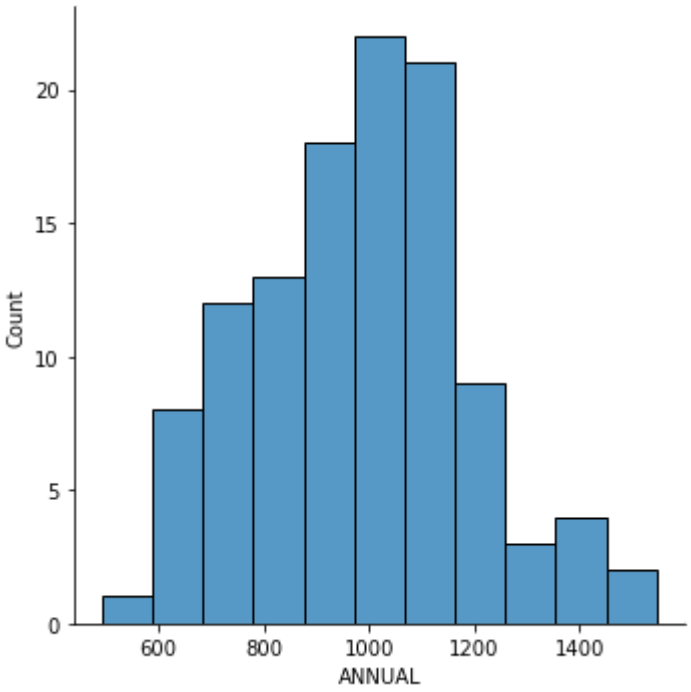


In [15]:

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

Out[15]:

```
<seaborn.axisgrid.FacetGrid at 0x1c80a0be550>
```

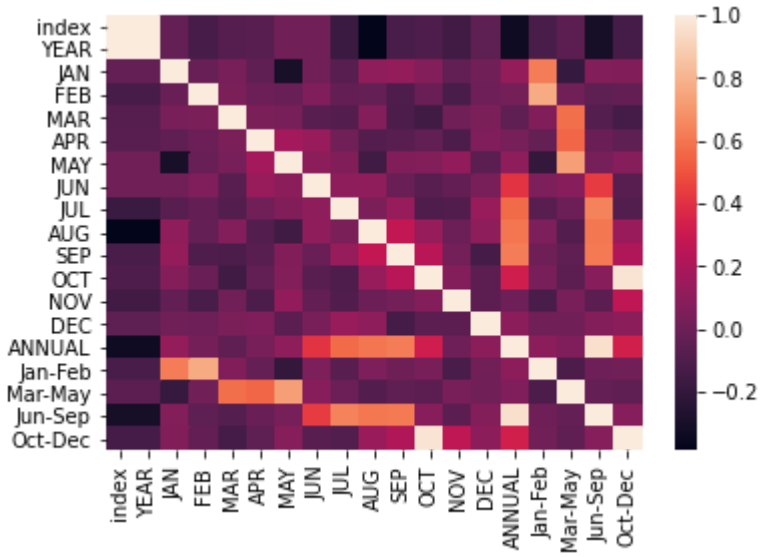


In [16]:

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

Out[16]:

```
<AxesSubplot:>
```



In [ ]: