Import Libraries

In [1]:

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

In [2]:

```python
df=pd.read_csv(r"c:\Users\user\Downloads\FP2_RainFall\rainfall.csv")[2279:2392]
df
```

Out[2]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	
2279	2279	GUJARAT REGION	1903	0.3	0.1	1.4	0.0	12.3	30.1	452.9	202.0	183.2	
2280	2280	GUJARAT REGION	1904	0.8	10.6	16.8	0.2	3.9	48.3	194.8	71.8	138.0	
2281	2281	GUJARAT REGION	1905	0.1	0.7	1.1	0.3	0.0	20.1	668.3	37.9	81.3	
2282	2282	GUJARAT REGION	1906	0.0	12.7	0.0	0.0	0.1	177.5	311.5	247.1	134.8	
2283	2283	GUJARAT REGION	1907	0.3	14.6	0.2	2.1	0.4	72.4	325.0	564.4	17.5	
...	...	...	...	...	...	...	...	...	...	...	...	...	
2387	2387	GUJARAT REGION	2011	0.0	0.2	0.0	0.0	0.0	16.3	259.2	451.7	162.5	
2388	2388	GUJARAT REGION	2012	0.1	0.0	0.0	0.0	0.0	34.4	178.2	230.3	263.8	
2389	2389	GUJARAT REGION	2013	0.0	0.9	0.1	4.6	0.0	155.7	405.4	211.1	287.3	
2390	2390	GUJARAT REGION	2014	5.7	0.1	0.2	1.0	1.3	11.6	307.5	138.6	235.1	
2391	2391	GUJARAT REGION	2015	1.8	0.0	6.1	5.5	0.9	120.7	354.7	37.4	93.4	

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	
2279	2279	GUJARAT REGION	1903	0.3	0.1	1.4	0.0	12.3	30.1	452.9	202.0	183.2	
2280	2280	GUJARAT REGION	1904	0.8	10.6	16.8	0.2	3.9	48.3	194.8	71.8	138.0	
2281	2281	GUJARAT REGION	1905	0.1	0.7	1.1	0.3	0.0	20.1	668.3	37.9	81.3	
2282	2282	GUJARAT REGION	1906	0.0	12.7	0.0	0.0	0.1	177.5	311.5	247.1	134.8	
2283	2283	GUJARAT REGION	1907	0.3	14.6	0.2	2.1	0.4	72.4	325.0	564.4	17.5	
...	...	...	...	...	...	...	...	...	...	...	...	...	
2387	2387	GUJARAT REGION	2011	0.0	0.2	0.0	0.0	0.0	16.3	259.2	451.7	162.5	
2388	2388	GUJARAT REGION	2012	0.1	0.0	0.0	0.0	0.0	34.4	178.2	230.3	263.8	
2389	2389	GUJARAT REGION	2013	0.0	0.9	0.1	4.6	0.0	155.7	405.4	211.1	287.3	
2390	2390	GUJARAT REGION	2014	5.7	0.1	0.2	1.0	1.3	11.6	307.5	138.6	235.1	
2391	2391	GUJARAT REGION	2015	1.8	0.0	6.1	5.5	0.9	120.7	354.7	37.4	93.4	

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, 2279 to 2391
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
```
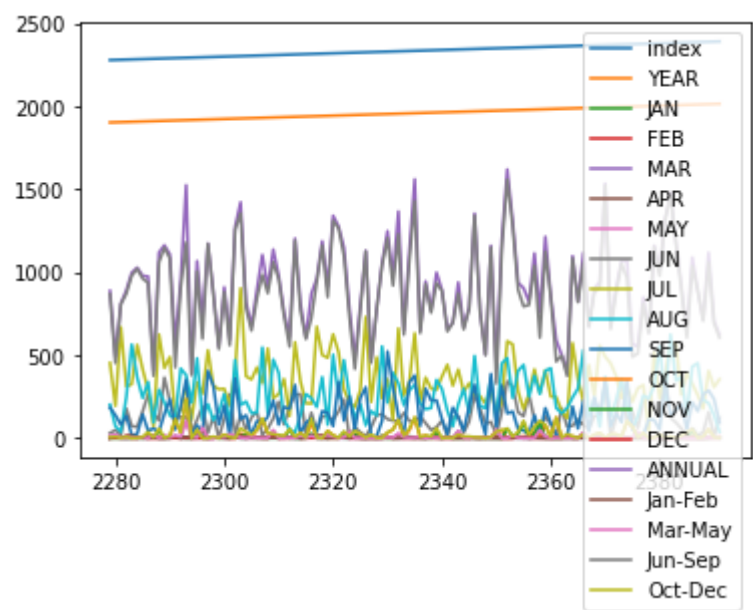
Line Chart

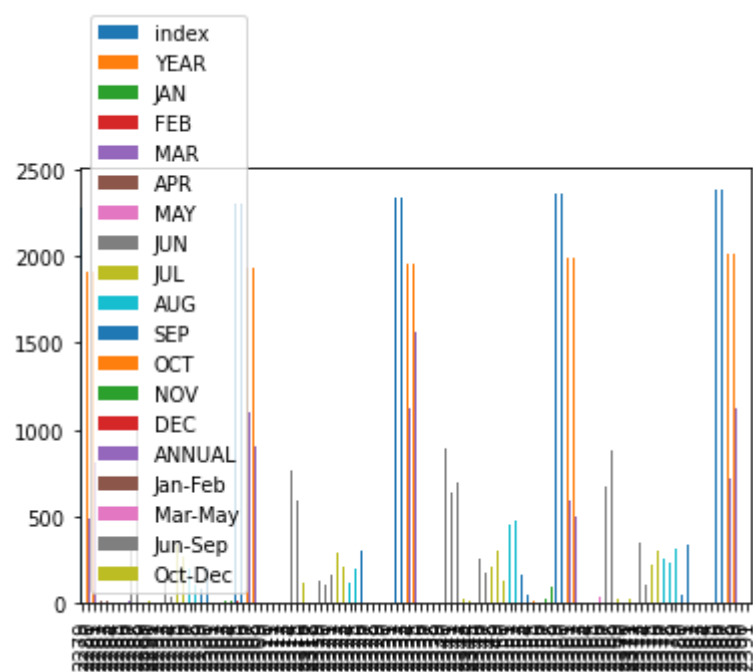In [6]:

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

Out[6]:

<AxesSubplot:>

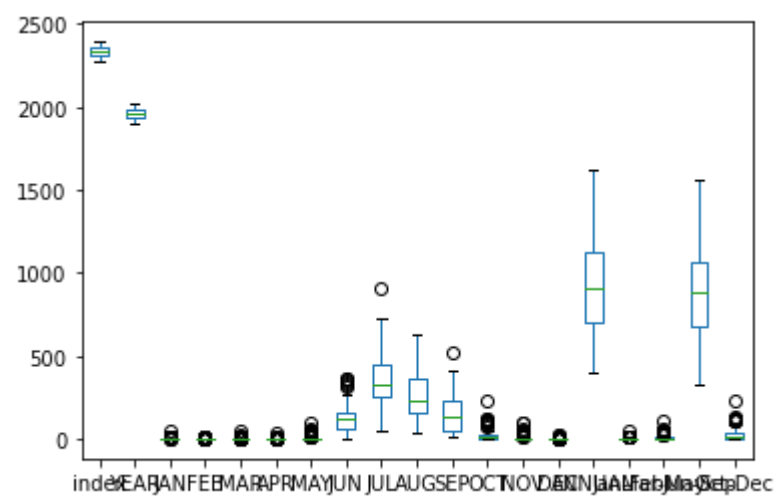# Bar chart

In [7]:

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

Out[7]:

<AxesSubplot:>

Box chart
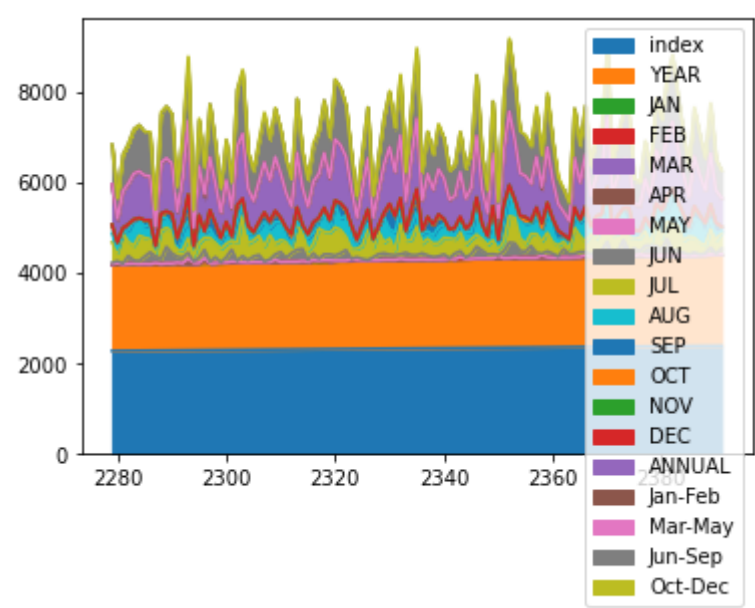
In [8]:

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

Out[8]:

`<AxesSubplot:>`

Area Chart

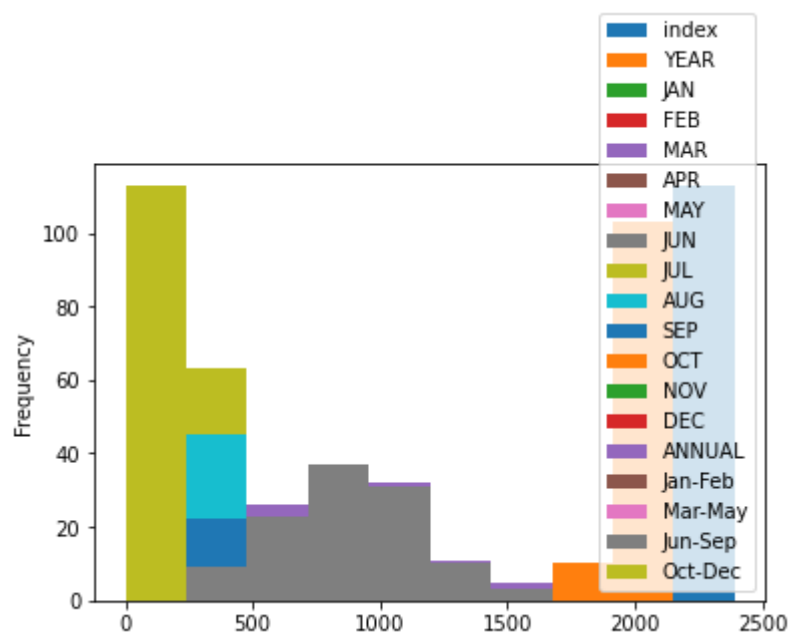In [9]:

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

Out[9]:

`<AxesSubplot:>`

Histogram

In [10]:

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

Out[10]:

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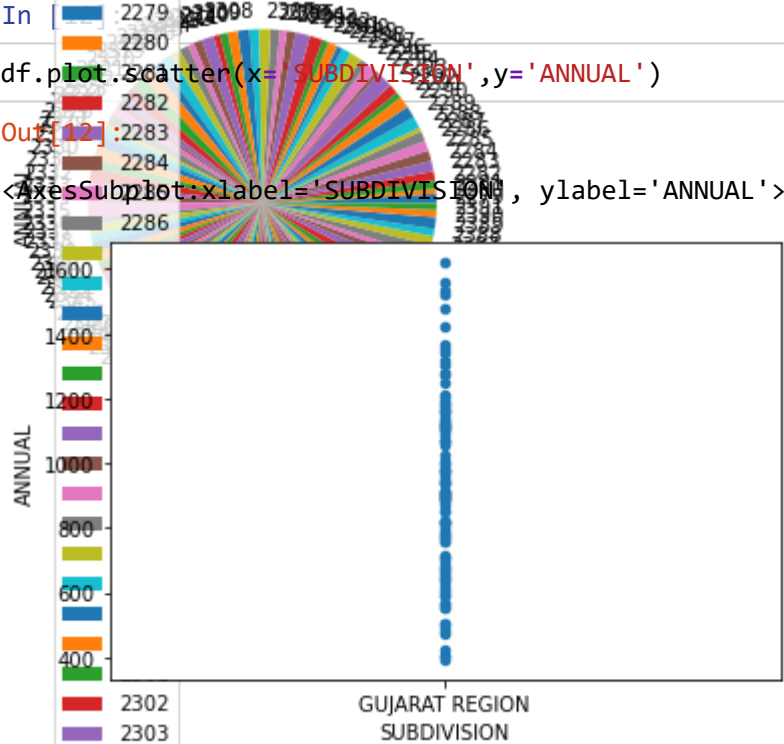

pie chart

In [11]:

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

Out[11]:

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

Scatter chart

In [ ]:

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

Out[12]:

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

In [ ]:

```python
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	2335.00000	1959.00000	1.746018	1.212389	1.237168	1.116814	5.841593	1
std	32.76431	32.76431	4.795181	2.891772	4.824865	4.015214	14.097826	
min	2279.00000	1903.00000	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	2307.00000	1931.00000	0.000000	0.000000	0.000000	0.000000	0.100000	
50%	2335.00000	1959.00000	0.100000	0.000000	0.000000	0.000000	0.900000	1
75%	2363.00000	1987.00000	1.400000	0.700000	0.200000	0.700000	4.000000	1
max	2391.00000	2015.00000	44.100000	14.600000	42.100000	40.400000	98.300000	3

EDA and Visualization

In [14]:

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

Out[14]:

```
<seaborn.axisgrid.PairGrid at 0x2d7777d4cd0>
```
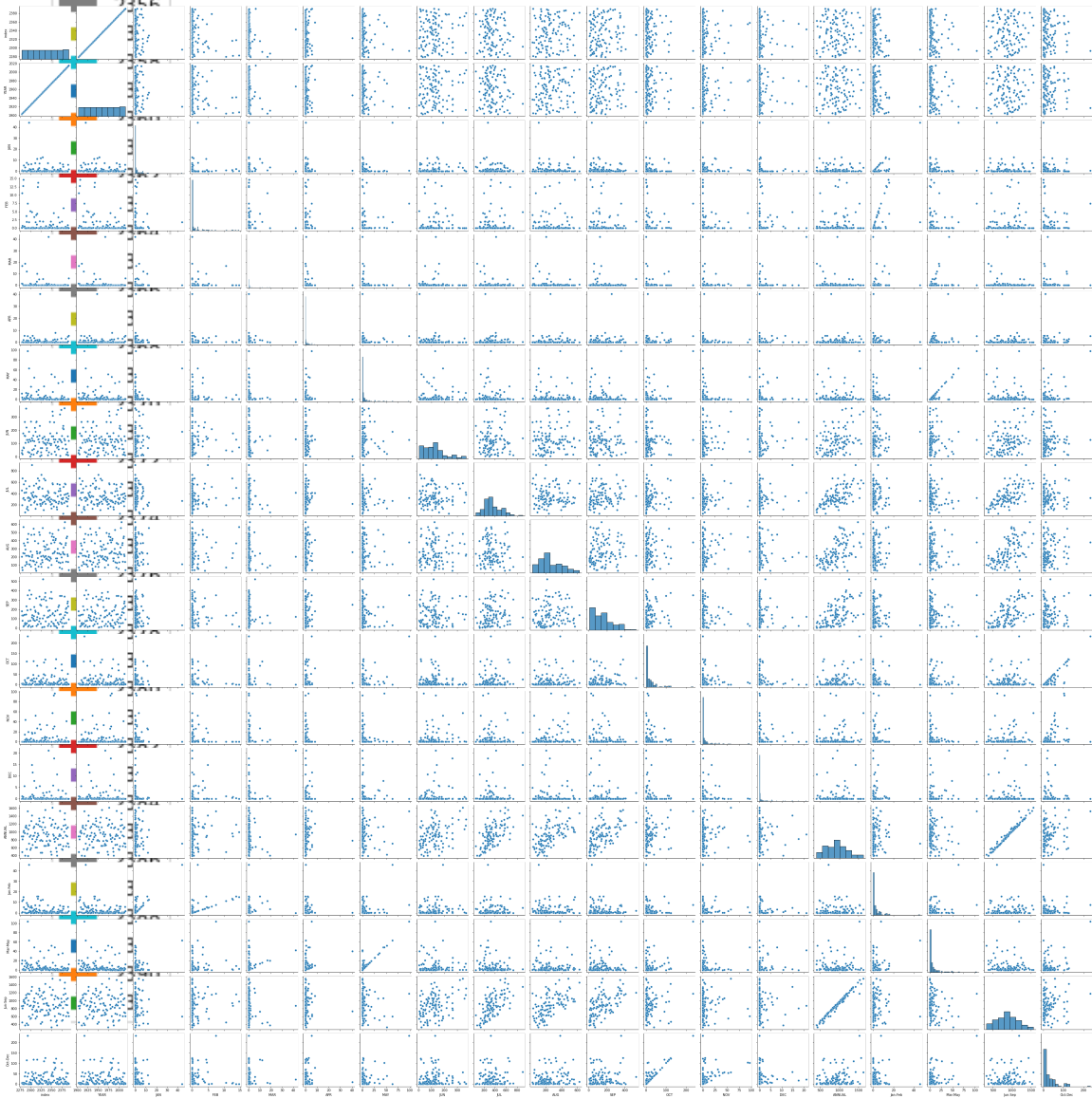
In [15]:

```python
sns.displot(df['ANNUAL'])
```
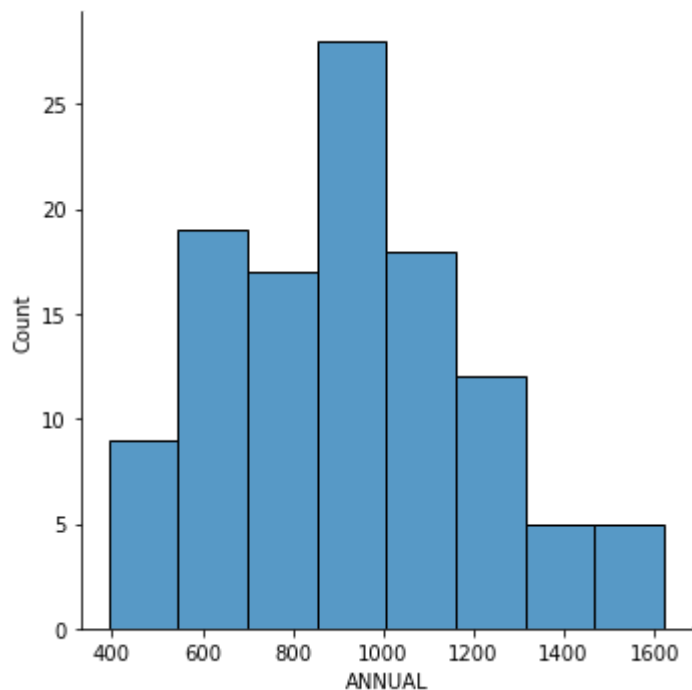
Out[15]:

```
<seaborn.axisgrid.FacetGrid at 0x2d7065b30d0>
```
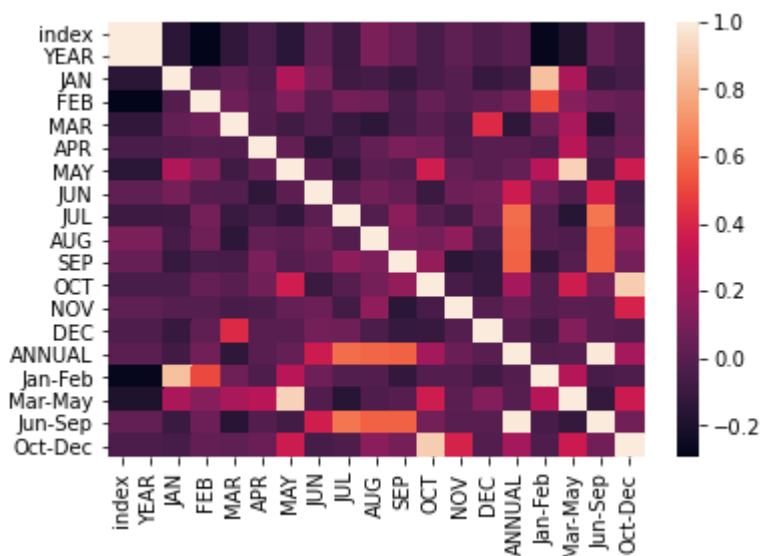

In [16]:

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

Out[16]:

```
<AxesSubplot:>
```


In [ ]: