

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")[3084:3197]
df
```

Out[2]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP
3084	3084	COASTAL ANDHRA PRADESH	1903	0.8	13.3	0.2	6.2	73.4	154.0	248.6	258.0	216.5
3085	3085	COASTAL ANDHRA PRADESH	1904	1.3	0.0	5.4	3.0	136.3	107.8	120.2	117.7	116.8
3086	3086	COASTAL ANDHRA PRADESH	1905	1.1	16.7	68.0	37.0	68.8	84.4	64.6	210.8	170.2
3087	3087	COASTAL ANDHRA PRADESH	1906	3.9	23.5	9.9	2.3	11.0	252.6	155.8	241.1	126.9
3088	3088	COASTAL ANDHRA PRADESH	1907	0.6	3.0	38.9	80.9	18.9	224.5	119.7	118.2	94.7
...
3192	3192	COASTAL ANDHRA PRADESH	2011	0.0	17.9	0.9	62.3	67.9	86.8	196.0	215.8	129.7
3193	3193	COASTAL ANDHRA PRADESH	2012	37.6	0.0	2.7	24.0	39.3	95.4	221.9	221.2	246.5
3194	3194	COASTAL ANDHRA PRADESH	2013	2.0	29.6	0.2	48.0	28.2	127.5	162.4	123.1	132.0
3195	3195	COASTAL ANDHRA PRADESH	2014	0.4	1.2	9.1	6.0	112.9	45.7	151.8	177.8	144.5
3196	3196	COASTAL ANDHRA PRADESH	2015	2.0	0.6	5.5	32.3	34.1	283.8	116.0	192.0	201.8

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
3084	3084	COASTAL ANDHRA PRADESH	1903	0.8	13.3	0.2	6.2	73.4	154.0	248.6	258.0	216.5
3085	3085	COASTAL ANDHRA PRADESH	1904	1.3	0.0	5.4	3.0	136.3	107.8	120.2	117.7	116.8
3086	3086	COASTAL ANDHRA PRADESH	1905	1.1	16.7	68.0	37.0	68.8	84.4	64.6	210.8	170.2
3087	3087	COASTAL ANDHRA PRADESH	1906	3.9	23.5	9.9	2.3	11.0	252.6	155.8	241.1	126.9
3088	3088	COASTAL ANDHRA PRADESH	1907	0.6	3.0	38.9	80.9	18.9	224.5	119.7	118.2	94.7
...
3192	3192	COASTAL ANDHRA PRADESH	2011	0.0	17.9	0.9	62.3	67.9	86.8	196.0	215.8	129.7
3193	3193	COASTAL ANDHRA PRADESH	2012	37.6	0.0	2.7	24.0	39.3	95.4	221.9	221.2	246.5
3194	3194	COASTAL ANDHRA PRADESH	2013	2.0	29.6	0.2	48.0	28.2	127.5	162.4	123.1	132.0
3195	3195	COASTAL ANDHRA PRADESH	2014	0.4	1.2	9.1	6.0	112.9	45.7	151.8	177.8	144.5
3196	3196	COASTAL ANDHRA PRADESH	2015	2.0	0.6	5.5	32.3	34.1	283.8	116.0	192.0	201.8

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-Feb',  
      'Mar-May', 'Jun-Sep', 'Oct-Dec'],  
      dtype='object')
```

In [5]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 113 entries, 3084 to 3196
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
```

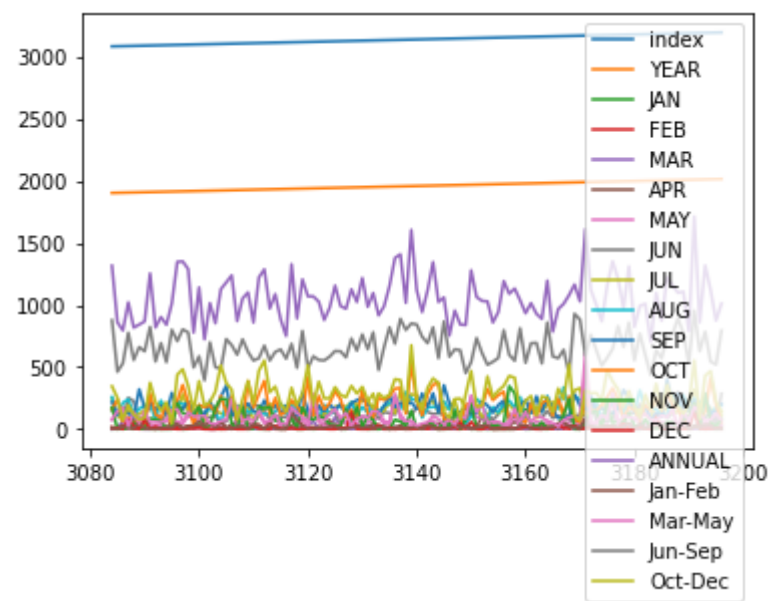
Line Chart

In [6]:

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

Out[6]:

<AxesSubplot:>



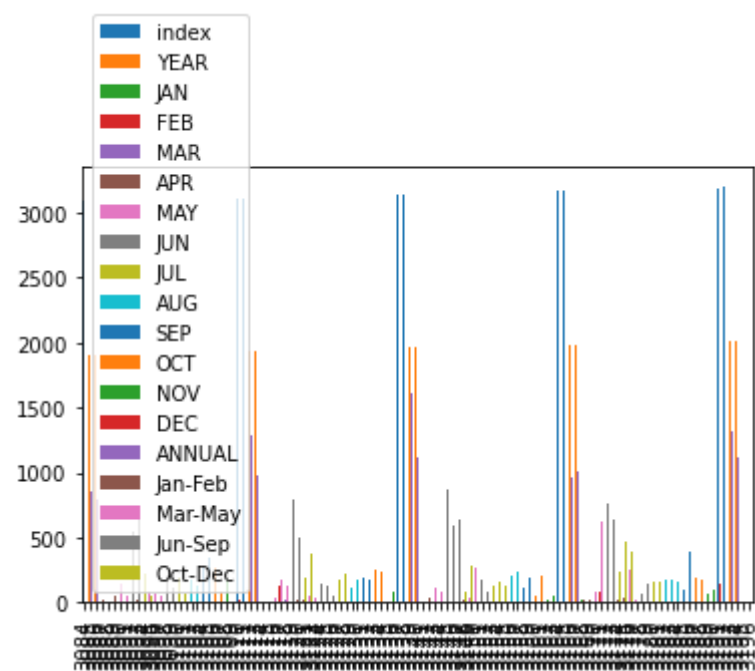
Bar chart

In [7]:

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

Out[7]:

<AxesSubplot:>



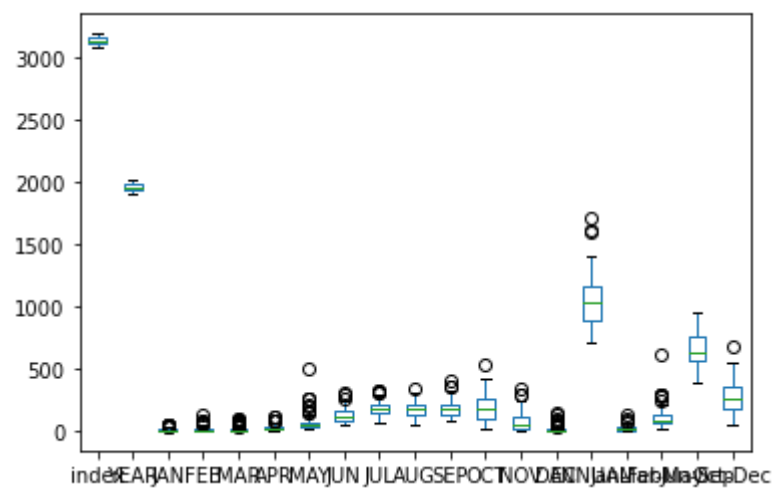
Box chart

In [8]:

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

Out[8]:

<AxesSubplot:>



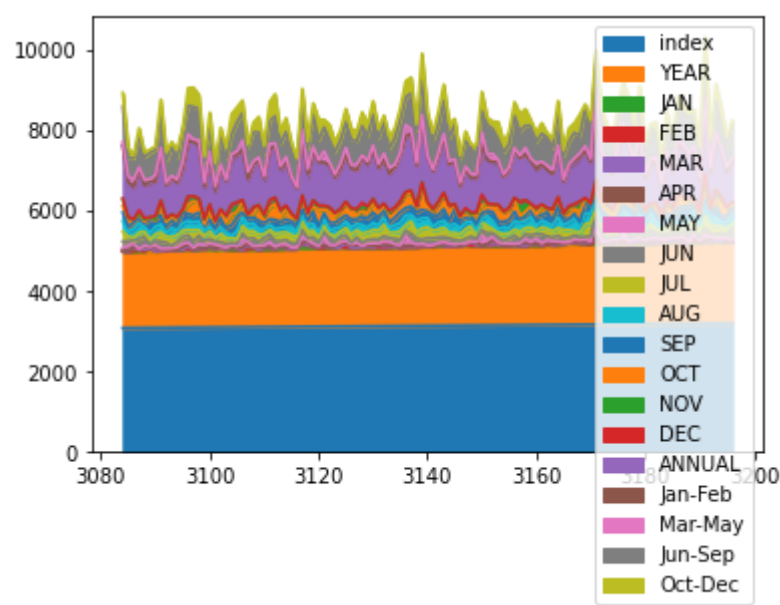
Area Chart

In [9]:

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

Out[9]:

<AxesSubplot:>



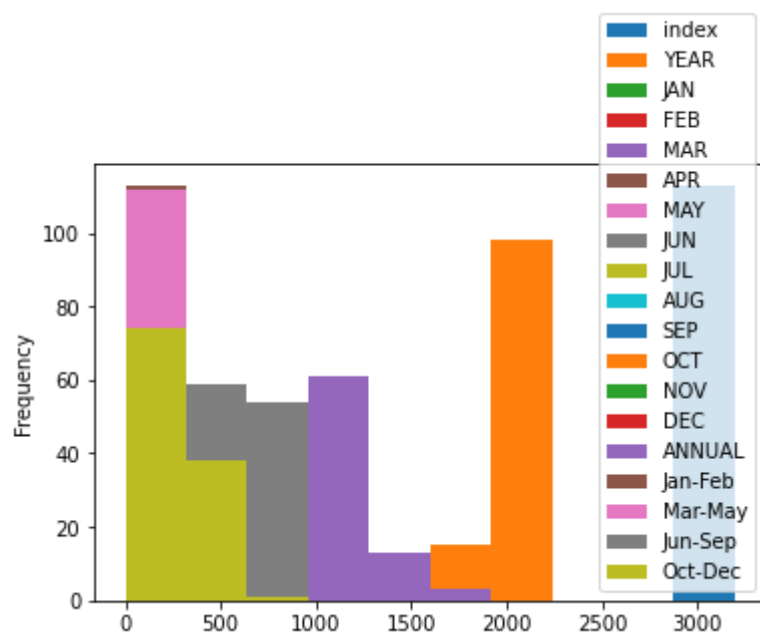
Histogram

In [10]:

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

Out[10]:

<AxesSubplot:ylabel='Frequency'>



pie chart

In [11]:

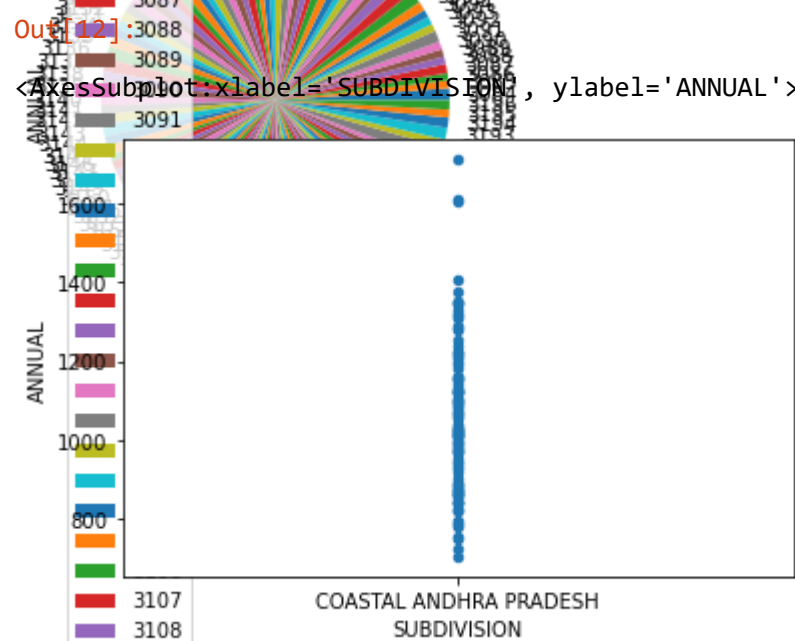
```
df.plot.pie(y='ANNUAL')
```

Out[11]:

<AxesSubplot:ylabel='ANNUAL'>

Scatter chart


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



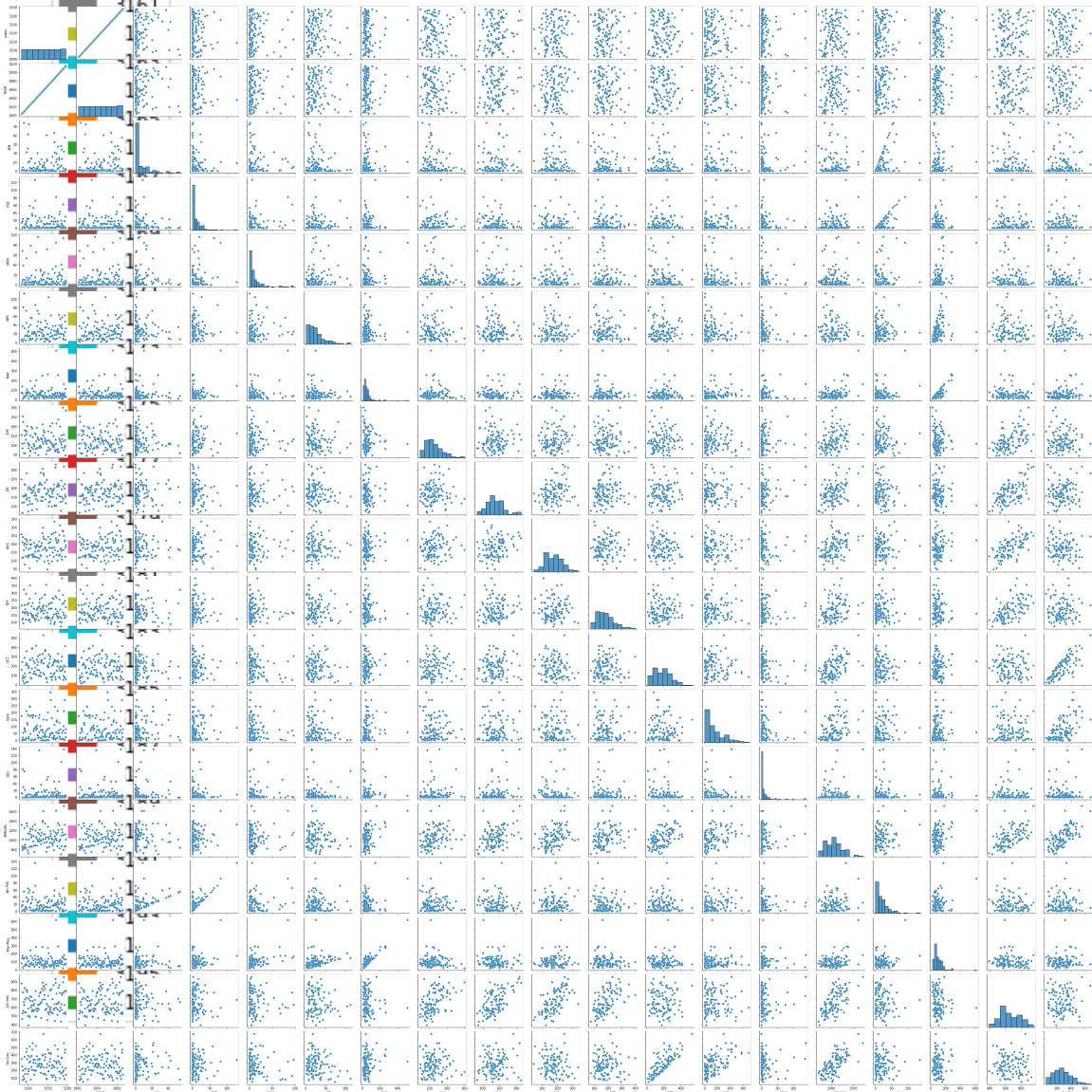
```
In [ ]: df.describe()
```

Out[13]:

	index	YEAR	JAN	FEB	MAR	APR	MAY	
count	113	113.00000	113.000000	113.000000	113.000000	113.000000	113.000000	1
mean	140	1959.00000	7.431858	12.436283	13.367257	26.748673	62.715929	1
std	32.7	32.76431	11.566201	19.079706	20.181770	21.497833	64.240054	
min	3084	1903.00000	0.000000	0.000000	0.000000	1.100000	10.500000	
25%	112	1931.00000	0.200000	0.500000	1.500000	12.900000	31.200000	
50%	140	1959.00000	2.000000	5.100000	5.700000	20.800000	44.400000	1
75%	168	1987.00000	10.300000	17.400000	14.900000	35.300000	70.000000	1
max	196	2015.00000	54.100000	127.100000	96.600000	112.200000	507.700000	3

EDA and Visualization

```
In [ ]: sns.pairplot(df)
Out[14]: <seaborn.axisgrid.PairGrid at 0x185cbf00dc0>
```

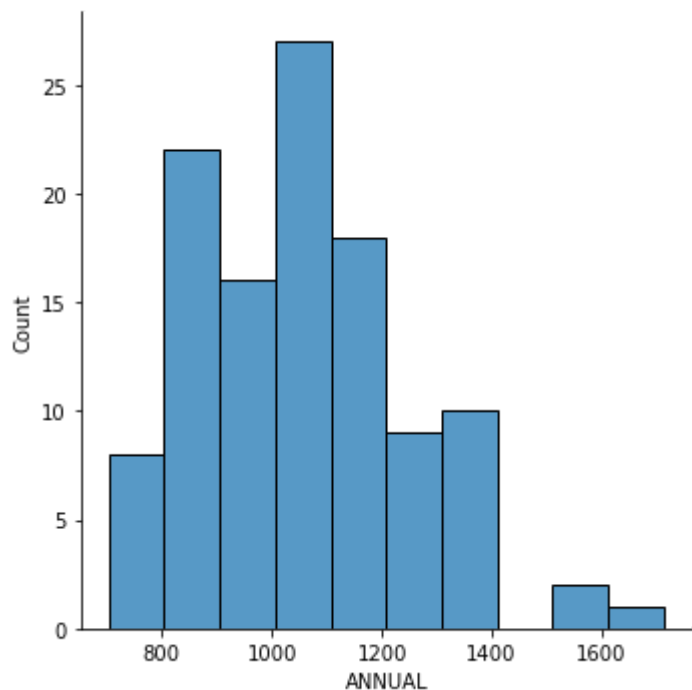


In [15]:

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

Out[15]:

```
<seaborn.axisgrid.FacetGrid at 0x185d835e970>
```

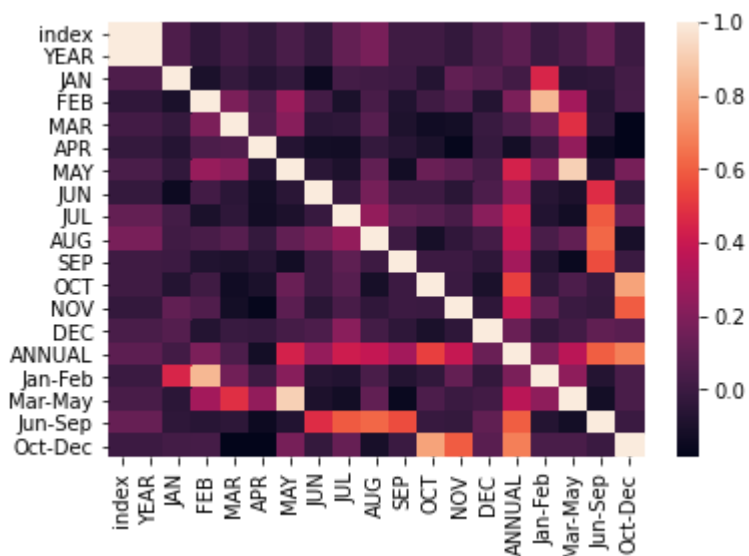


In [16]:

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

Out[16]:

```
<AxesSubplot:>
```



In []:

