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")[112:207]
df
```

Out[2]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SE
112	112	ARUNACHAL PRADESH	1918	10.4	11.0	191.2	144.6	861.1	1609.9	1303.0	692.6	515
113	113	ARUNACHAL PRADESH	1919	34.5	67.8	28.5	256.9	420.6	973.6	999.0	286.7	628
114	114	ARUNACHAL PRADESH	1920	14.0	196.3	605.6	364.7	173.6	840.6	535.4	896.5	376
115	115	ARUNACHAL PRADESH	1921	78.9	54.3	180.3	358.0	598.0	1233.2	1433.0	885.9	603
116	116	ARUNACHAL PRADESH	1922	50.7	59.4	170.4	299.5	350.5	1109.3	918.7	488.3	207
...	...	...	...	...	...	...	...	...	...	...	...	
202	202	ARUNACHAL PRADESH	2011	40.0	51.3	174.5	240.8	219.6	288.4	531.4	277.6	286
203	203	ARUNACHAL PRADESH	2012	57.8	35.8	134.2	403.4	187.4	645.8	638.9	316.0	724
204	204	ARUNACHAL PRADESH	2013	18.5	40.5	115.1	175.1	335.8	290.0	329.6	230.2	316
205	205	ARUNACHAL PRADESH	2014	19.0	101.9	80.3	86.7	299.0	415.8	392.4	599.6	343
206	206	ARUNACHAL PRADESH	2015	30.8	47.5	97.5	287.1	238.9	637.9	329.3	595.5	374

95 rows × 20 columns

Data Cleaning and Preprocessing

In [3]:

```
df.dropna()
```

Out[3]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SE
112	112	ARUNACHAL PRADESH	1918	10.4	11.0	191.2	144.6	861.1	1609.9	1303.0	692.6	515
113	113	ARUNACHAL PRADESH	1919	34.5	67.8	28.5	256.9	420.6	973.6	999.0	286.7	628
114	114	ARUNACHAL PRADESH	1920	14.0	196.3	605.6	364.7	173.6	840.6	535.4	896.5	376
115	115	ARUNACHAL PRADESH	1921	78.9	54.3	180.3	358.0	598.0	1233.2	1433.0	885.9	603
116	116	ARUNACHAL PRADESH	1922	50.7	59.4	170.4	299.5	350.5	1109.3	918.7	488.3	207
...	...	...	...	...	...	...	...	...	...	...	...	
202	202	ARUNACHAL PRADESH	2011	40.0	51.3	174.5	240.8	219.6	288.4	531.4	277.6	286
203	203	ARUNACHAL PRADESH	2012	57.8	35.8	134.2	403.4	187.4	645.8	638.9	316.0	724
204	204	ARUNACHAL PRADESH	2013	18.5	40.5	115.1	175.1	335.8	290.0	329.6	230.2	316
205	205	ARUNACHAL PRADESH	2014	19.0	101.9	80.3	86.7	299.0	415.8	392.4	599.6	343
206	206	ARUNACHAL PRADESH	2015	30.8	47.5	97.5	287.1	238.9	637.9	329.3	595.5	374

91 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: 95 entries, 112 to 206
Data columns (total 20 columns):
 #   Column       Non-Null Count  Dtype  
---  ------       --------------  -----  
 0   index        95 non-null     int64  
 1   SUBDIVISION  95 non-null     object 
 2   YEAR         95 non-null     int64  
 3   JAN          94 non-null     float64
 4   FEB          94 non-null     float64
 5   MAR          94 non-null     float64
 6   APR          95 non-null     float64
 7   MAY          95 non-null     float64
 8   JUN          94 non-null     float64
 9   JUL          95 non-null     float64
 10  AUG          95 non-null     float64
 11  SEP          95 non-null     float64
 12  OCT          94 non-null     float64
 13  NOV          94 non-null     float64
 14  DEC          94 non-null     float64
 15  ANNUAL       91 non-null     float64
 16  Jan-Feb      94 non-null     float64
 17  Mar-May      94 non-null     float64
 18  Jun-Sep      94 non-null     float64
 19  Oct-Dec      93 non-null     float64
dtypes: float64(17), int64(2), object(1)
memory usage: 15.0+ KB
```

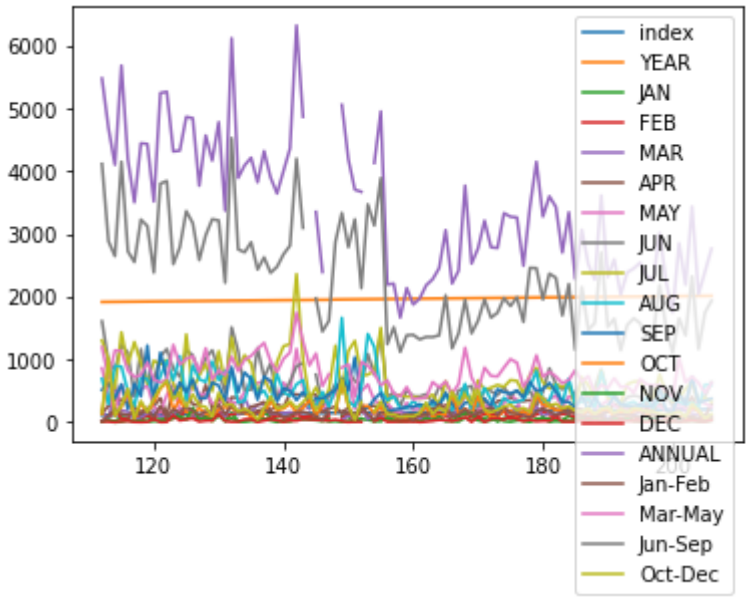
Line Chart

In [6]:

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

Out[6]:

<AxesSubplot:>



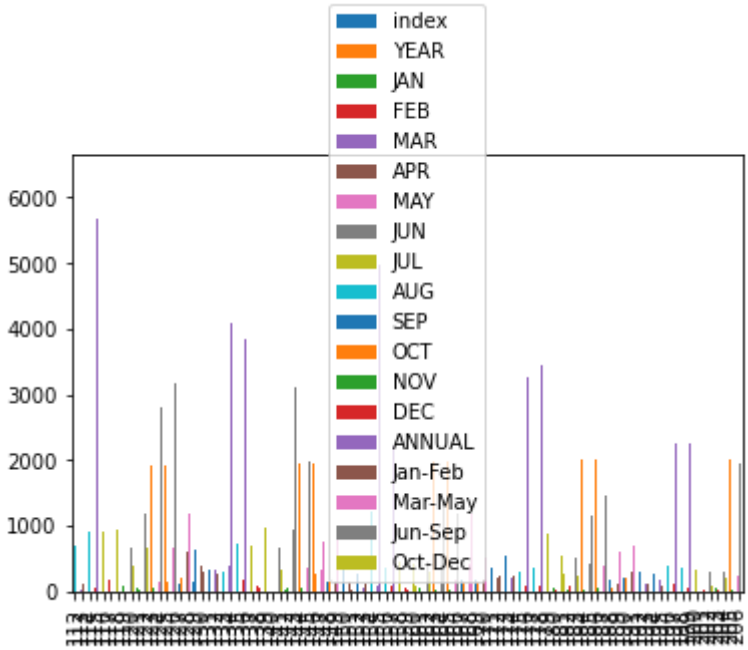
Bar chart

In [7]:

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

Out[7]:

<AxesSubplot:>



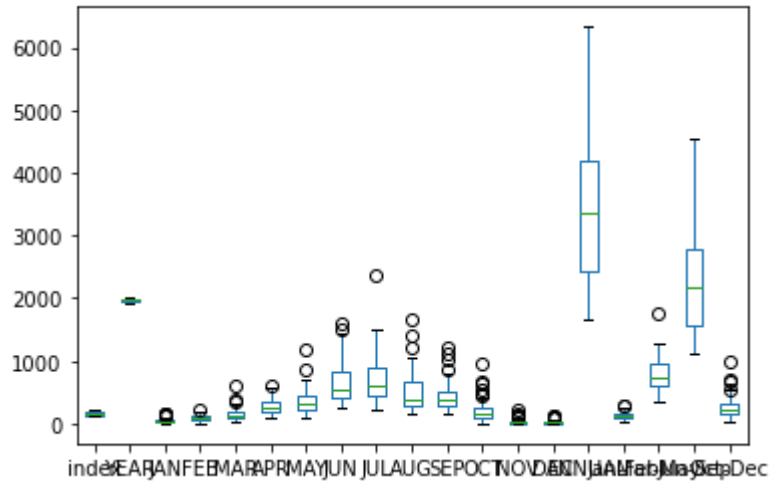
Box chart

In [8]:

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

Out[8]:

<AxesSubplot:>



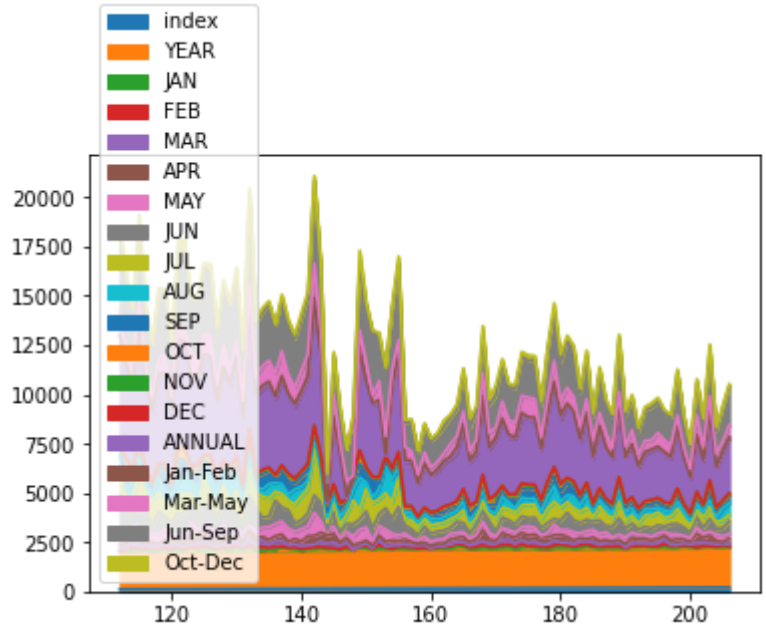
Area Chart

In [9]:

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

Out[9]:

<AxesSubplot:>



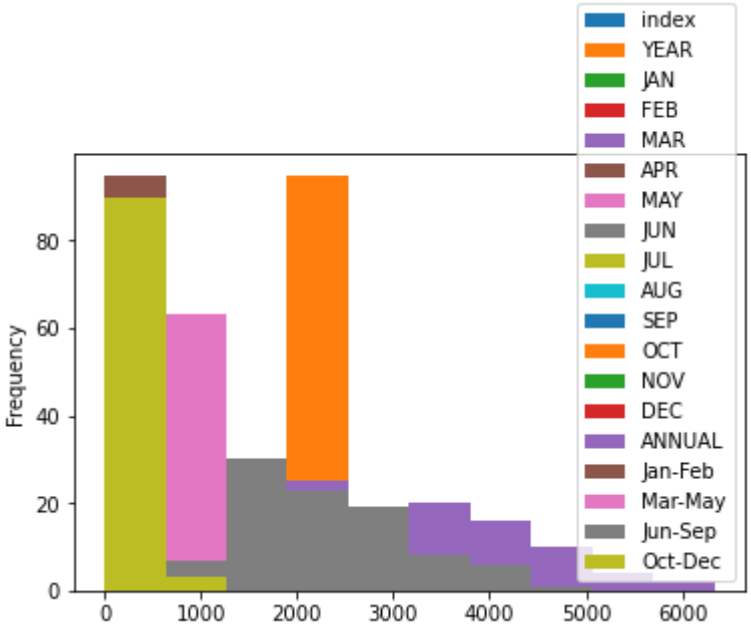
Histogram

In [10]:

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

Out[10]:

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



pie chart

In [11]:

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

Out[11]:

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

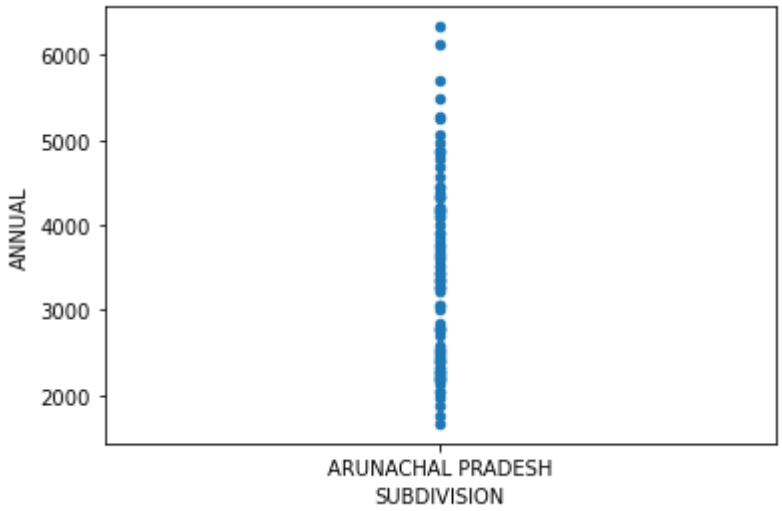
Scatter chart

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	95.000000	95.000000	94.000000	94.000000	94.000000	95.000000	95.000000
mean	159.000000	1966.863158	47.564894	90.562766	154.404255	263.225263	360.465263
std	27.568098	28.808469	34.517412	46.444005	86.217522	115.573282	177.348481
min	112.000000	1918.000000	0.600000	6.100000	28.500000	86.700000	101.800000
25%	135.500000	1941.500000	19.225000	56.950000	102.200000	180.050000	237.000000
50%	159.000000	1968.000000	39.600000	85.150000	140.800000	240.900000	307.200000
75%	182.500000	1991.500000	63.625000	117.275000	188.350000	341.200000	440.350000
max	206.000000	2015.000000	164.500000	208.500000	605.600000	595.100000	1168.600000

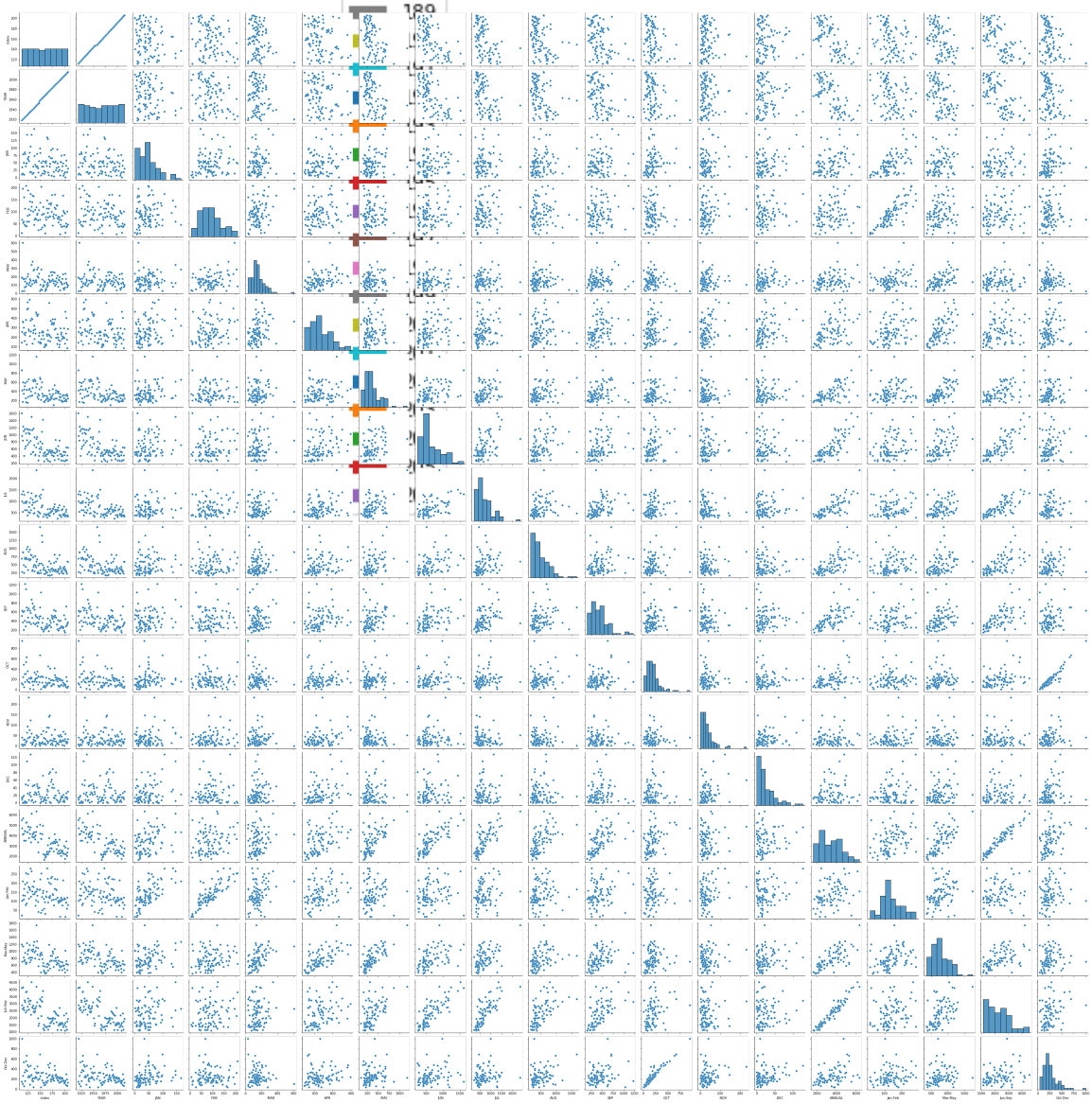
EDA and Visualization

In [14]:

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

Out[14]:

```
<seaborn.axisgrid.PairGrid at 0x2c151275d60>
```

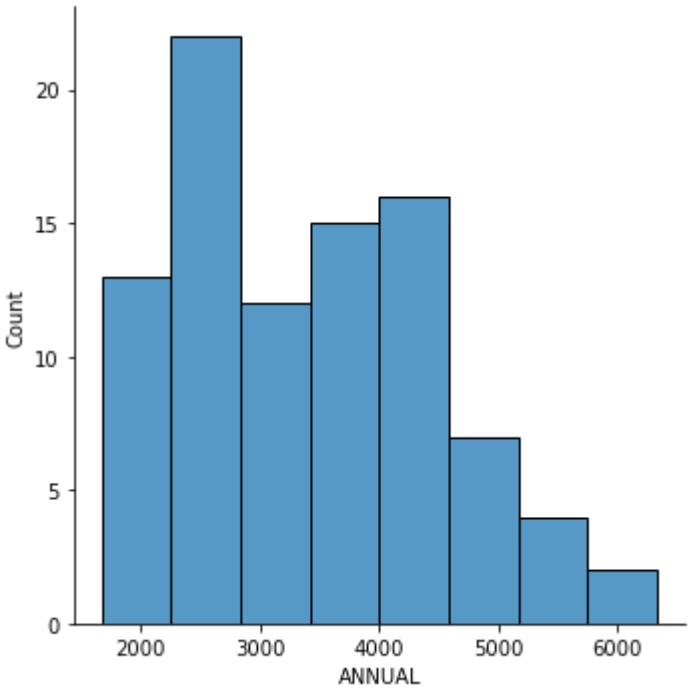


In [15]:

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

Out[15]:

```
<seaborn.axisgrid.FacetGrid at 0x2c1565fbf70>
```

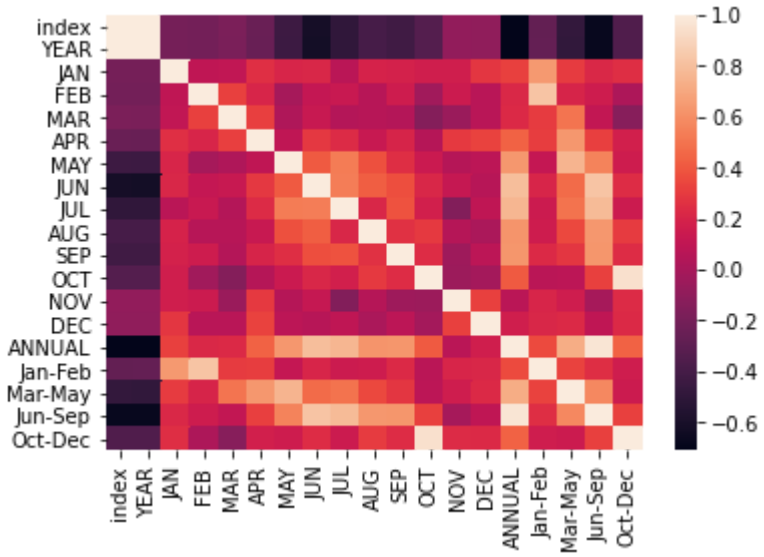


In [16]:

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

Out[16]:

```
<AxesSubplot:>
```



In [ ]: